

Molecular connectivity and gas chromatographic retention parameters

J. S. MILLERSHIP* AND A. D. WOOLFSON

Department of Pharmacy, The Queen's University of Belfast, Medical Biology Centre, 97, Lisburn Road, Belfast BT9 7BL, Northern Ireland, U.K.

The relation between gas chromatographic retention parameters and molecular connectivity has been investigated for several series of compounds including hydrocarbons, compounds containing oxygen functional groups (esters, aldehydes, ketones, alcohols and ethers) and drug molecules. With the oxygenated compounds good correlation was observed with the first order valence connectivity index, whilst for the other groups multiparameter equations were required for satisfactory correlation. The effect of column to column variation within a series was also investigated.

It has previously been demonstrated (Millership & Woolfson 1978) that a linear relation exists between the common logarithm of retention time (log RT) and molecular connectivity when considering series of similar compounds. In the present study this relationship is extended to cover series in which compounds with diverse structures are examined.

Calculation of molecular connectivity parameters

Calculations of molecular connectivity made throughout this paper follow the method of Kier & Hall (1976). The general form of the index, the first-order connectivity term (1X), was found by assigning to each vertex in the molecular graph a value (δ) which is the number of edges (bonds) to that atom, bonds to hydrogen being ignored (suppressed). Thus, for a graph of n edges and s sub-graphs (where a subgraph consists, in this case, of a bond between two atoms, i and j), X is calculated by equation 1.

$$^1X = \sum_{s=1}^n (\delta_i \delta_j)^{-\frac{1}{2}} \quad \dots \quad (1)$$

The zero order, second order and third order connectivity terms were calculated as described by Kier & Hall (1976).

Where unsaturation was present in the molecule, the vertex valencies of the unsaturated carbon atoms were calculated using equation 2.

$$\delta_i^v = Z_i^v - h_i \quad \dots \quad (2)$$

Where δ_i^v is the vertex valency, Z_i^v is the number of valence electrons and h_i is the number of hydrogen atoms that are suppressed. We have also introduced a term DV which we have utilized in the hydro-

carbon correlation. DV is calculated using equation 3.

$$DV = (\delta_i)^{-\frac{1}{2}} + (\delta_j)^{-\frac{1}{2}} \quad \dots \quad (3)$$

Where δ_i and δ_j are the vertex valencies of the carbon atoms involved in multiple bonding. DV for all saturated hydrocarbons is zero.

In cases where hetero-atoms were present in the molecular structure, the valence connectivity (X^v) has on occasions been used (Kier & Hall 1976). This is calculated similarly to 1X , except that the vertex valency of the hetero-atom (N or O) was found from equation 2. The empirically derived values for the halogens (Kier & Hall 1976) were also used.

RESULTS AND DISCUSSION

Using the data of Csicsery & Pines (1962) we have previously demonstrated that a linear relation exists between log RT and molecular connectivity for a series of 18 saturated hydrocarbons and 28 unsaturated hydrocarbons.

It was felt that the proven power of molecular connectivity should allow a successful correlation between the whole group of 46 compounds. The correlation of log RT and molecular connectivity using only the first order connectivity term gave poor correlation as demonstrated by equation 4.

$$\log RT = 0.721^1X + 0.188 \quad \dots \quad (4)$$

$n = 46, r = 0.880, s = 0.290$

Several multiple parameter equations was developed which gave improved correlation and a lower standard error, equation 5 was the equation of best fit incorporating up to second order connectivity terms,

* Correspondence.

$$\log RT = 0.495^{\circ}X + 0.242^{\circ}X - 0.269^{\circ}X + 0.311 DV - 0.681 \quad (5)$$

$n = 46, r = 0.984, s = 0.111$

Table 1 compares the calculated and observed values for log RT of the 46 compounds.

The second series of compounds investigated was a group of aldehydes, ketones, esters, ethers and alcohols. Correlation was investigated using the

Table 1. Comparison of observed and calculated retention parameters for aliphatic hydrocarbons.

Compound	Log R.T. (OBS)	Log R.T. (Calc)*	Δ Log R.T.
Ethane	0.362	0.511	0.149
Propane	0.785	0.811	-0.026
Isobutane	1.041	1.043	-0.002
n-Butane	1.215	1.203	0.012
Neopentane	1.231	1.224	0.007
Isopentane	1.505	1.504	0.001
n-Pentane	1.613	1.579	0.034
2,2-Dimethylbutane	1.732	1.732	0.000
2-Methylpentane	1.869	1.873	-0.004
2,3-Dimethylbutane	1.880	1.841	0.039
3-Methylpentane	1.943	1.953	-0.010
n-Hexane	2.000	1.955	0.045
3,3-Dimethylpentane	2.231	2.229	0.002
2,2,4-Trimethylpentane	2.305	2.385	-0.080
2,3-Dimethylpentane	2.283	2.283	0.000
n-Heptane	2.377	2.331	0.046
2,5-Dimethylhexane	2.499	2.543	0.044
n-Octane	2.750	2.707	0.043
Ethylene	0.447	0.580	-0.133
Propylene	1.009	0.978	0.031
1-Butene	1.408	1.380	0.027
1-Pentene	1.775	1.789	-0.014
2-Methylbut-1 ene	1.848	1.742	0.016
Isobutene	1.433	1.284	0.149
3-Methyl-1-Butene	1.629	1.846	-0.217
3,3-Dimethyl-1-butene	1.820	1.997	-0.177
2-Methyl-2-butene	1.964	1.740	0.224
3-Methyl-1-pentene	2.017	2.133	-0.116
4-Methyl-1-pentene	2.017	2.040	-0.023
2,3-Dimethyl-1-butene	2.093	2.080	0.013
1-Hexene	2.179	2.125	0.054
2-Methyl-1-pentene	2.190	2.107	0.083
2-Ethyl-1-butene	2.255	2.189	0.066
2-Methyl-2-pentene	2.267	2.149	0.188
3,3-Dimethyl-1-pentene	2.305	2.554	-0.249
2,3,3-Trimethyl-1-butene	2.339	2.334	0.005
2,4-Dimethyl-1-pentene	2.335	2.396	-0.061
2,3-Dimethyl-2-butene	2.143	2.112	0.030
2,3-Dimethyl-1-pentene	2.415	2.519	-0.104
2-Methyl-3-ethyl-1-butene	2.449	2.522	-0.073
2,4,4-Trimethyl-1-pentene	2.491	2.616	-0.125
1-Heptene	2.542	2.501	0.059
2,4,4-Trimethyl-2-pentene	2.580	2.544	0.036
2,3-Dimethyl-2-pentene	2.681	2.561	0.120
3-Ethyl-2-hexene	2.935	3.054	-0.119
1-Octene	2.928	2.877	0.051

* Using equation 5.

Column 33% dimethylsulpholane on Firebrick 110/120 mesh.

Carrier gas—helium, temperature 43 °C.

data of McReynolds (1966). The retention data utilized was the Specific Retention Volume (Vg) and the Kovats Retention Index (RI).

Regression analysis was used to investigate the relation between log Vg (and RI) with molecular connectivity for the individual groups of compounds. The initial investigations were carried out on data from an SE 30 column operating at 160 °C. The reason for investigating individual groups of compounds first was that no relationships had previously been established for these retention parameters. Equations 6–9 describe the results for the individual groups of compounds.

Alcohols

$$\log Vg = 0.479^{\circ}X - 0.317 \quad \dots \quad (6)$$

$n = 10, r = 0.997, s = 0.044$

Aldehydes and ketones

$$\log Vg = 0.475^{\circ}X - 0.255 \quad \dots \quad (7)$$

$n = 10, r = 0.996, s = 0.037$

Esters

$$\log Vg = 0.457^{\circ}X - 0.386 \quad \dots \quad (8)$$

$n = 10, r = 0.998, s = 0.025$

Ethers

$$\log Vg = 0.375^{\circ}X - 0.319 \quad \dots \quad (9)$$

$n = 10, r = 0.982, s = 0.115$

SE 30

$$\log Vg = 0.380^{\circ}X - 0.149 \quad \dots \quad (10)$$

$n = 40, r = 0.946, s = 0.153$

Squalene

$$\log Vg = 0.667^{\circ}X + 0.034 \quad \dots \quad (11)$$

$n = 40, r = 0.937, s = 0.221$

Carbowax 20M

$$\log Vg = 0.318^{\circ}X + 0.500 \quad \dots \quad (12)$$

$n = 40, r = 0.666, s = 0.420$

Carbowax 400

$$\log Vg = 0.307^{\circ}X + 0.795 \quad \dots \quad (13)$$

$n = 40, r = 0.579, s = 0.511$

These equations clearly demonstrate excellent correlation between molecular connectivity and log Vg. Correlation between RI and molecular connectivity was equally good. Similarly, if the results for individual compounds are considered on Squalene, Carbowax 20 M or Carbowax 400 excellent correlation between log Vg and molecular connectivity is observed. On all columns the standard error was much greater for the ethers than for any of the other groups.

If we now consider the combined group of 40 compounds on the four different columns (SE 30, Squalene, Carbowax 20 M and Carbowax 400) equations 10–13 were generated.

The correlation on the SE 30 and Squalene columns is reasonable for a diverse range of compounds using a single parameter equation (multiple X terms improve the correlation). The results for the Carbowax 20 M and 400 columns are poor and may arise from differences in the activity of the columns. The SE 30 and Squalene columns are considered non-polar whereas the Carbowax 20 M and 400 columns are polar.

From the data available for the SE30 column the number of compounds considered was increased to 76 (again ethers, alcohols, aldehydes, ketones and esters), equation 14 was thus generated.

$$\log Vg = 0.0405^1X^v - 0.055 \quad (14)$$

$$n = 76, r = 0.953, s = 0.129$$

By means of residual analysis, and using this equation, it was possible to show that the main source of error once again arose from the inclusion of the ethers. If we therefore consider the 60 esters, aldehydes, ketones and alcohols alone equation 15 is generated.

$$\log Vg = 0.454^1X^v - 0.126$$

$$n = 60, r = 0.993, s = 0.052 \dots (15)$$

A computer drawn graph of $\log Vg$ against $^1X^v$ for these 60 compounds is produced in Fig. 1 and Table 2 is a comparison of observed and calculated values.

Michotte & Massart (1977) have investigated the relationship between the first order molecular connectivity index and RI the Kovats Retention Index. Similar variation from column to column was observed. No attempt was made to utilize valence connectivity or higher order terms.

Kier & Hall (1979) have reinvestigated Michotte & Massart's results. They have investigated the relationship between molecular connectivity and gas chromatographic retention indices for individual groups of compounds (e.g. alcohols, ketones, esters and ethers). They suggest that chemical behaviour across classes is dependent on topological and non-topological structural characteristics. They further suggest that correlation of combined groups of compounds would show different but parallel chromatographic behaviour due to the variation in non-topological properties between groups.

Finally we turned our attention to a series of drug molecules. The retention data used were those of Moffat (1975). The retention data are expressed in terms of the Kovats Retention Index (RI) and were determined on an SE 30 column. The Retention Index of 41 drugs (amphetamines, local anaesthetics

Table 2. Comparison of observed and calculated retention parameters for esters, aldehydes, ketones and alcohols.

Compound	Log Vg (OBS)	Log Vg (Calc)*	Δ Log Vg
Methanol	0.204	0.077	0.127
Propanol	0.623	0.565	0.058
Butanol	0.869	0.792	0.077
Hexanol	1.346	1.246	0.100
Octanol	1.781	1.700	0.081
Isopropanol	0.462	0.516	-0.054
Pentanol	1.111	1.019	0.092
2-Methyl-2-butanol	0.826	0.911	-0.085
Ethanol	0.342	0.338	0.004
Heptanol	1.568	1.473	0.095
Acetaldehyde	0.146	0.243	-0.097
Propionaldehyde	0.462	0.487	-0.025
2-Methylbutyraldehyde	0.886	0.901	-0.015
Hexanal	1.187	1.168	0.019
Heptanal	1.389	1.395	-0.006
Acetone	0.447	0.421	0.026
2-Pentanone	0.924	0.902	0.022
3-Pentanone	0.939	0.930	0.009
3-Methyl-2-butanone	0.857	0.849	0.008
3-Hexanone	1.143	1.157	-0.014
Ethyl acetate	0.724	0.738	-0.014
Propyl acetate	0.949	0.965	-0.016
Methyl acetate	0.544	0.471	0.073
Ethyl formate	0.519	0.540	-0.021
Methyl formate	0.279	0.274	0.005
Pentyl acetate	1.413	1.419	-0.006
Butyl acetate	1.185	1.192	-0.007
Isobutyl formate	0.919	0.929	-0.010
Isobutyl acetate	1.104	1.127	-0.023
Hexyl acetate	1.622	1.645	-0.023
Nonanol	1.987	1.927	0.060
3-Heptanol	1.387	1.399	-0.012
2-Methyl-3-pentanol	1.137	1.173	-0.036
2,2-Dimethyl-1-butanol	1.196	1.138	0.058
3-Ethyl-3-pentanol	1.326	1.420	-0.094
Isobutyl alcohol	0.792	0.727	0.064
t-Butyl alcohol	0.531	0.657	-0.126
2-Ethyl-4-methyl-1-pentanol	1.607	1.604	0.003
4-Methyl-1-pentanol	1.265	1.181	0.084
3-Pentanol	0.973	1.004	-0.031
Isovaleraldehyde	0.857	0.876	-0.019
Butyraldehyde	0.708	0.714	-0.006
Valeraldehyde	0.954	0.941	0.013
Isobutyraldehyde	0.653	0.657	-0.004
2-Ethyl-hexanal	1.524	1.599	-0.075
3,3-Dimethyl-2-butanone	0.982	0.987	-0.005
2-Heptanone	1.356	1.356	0.000
4,4-Dimethyl-2-pentanone	1.158	1.196	-0.038
2-Octanone	1.590	1.583	0.007
2-Nonanone	1.801	1.810	-0.009
Propyl formate	0.771	0.767	0.004
Butyl formate	1.013	0.994	0.019
Pentyl formate	1.243	1.221	0.022
Hexyl formate	1.462	1.448	0.014
Isopropyl acetate	0.822	0.918	-0.086
Isopropyl formate	0.672	0.719	-0.047
t-Butyl acetate	0.934	0.988	0.054
Hexyl acetate	1.622	1.646	-0.024
Heptyl acetate	1.826	1.873	-0.047
Isopentyl acetate	1.334	1.290	0.044

* Using equation 15.

Column 20% SE 30 on Celite 545 79.5%.

With Poly-Tergent J-300 (0.5%), temperature 160 °C.

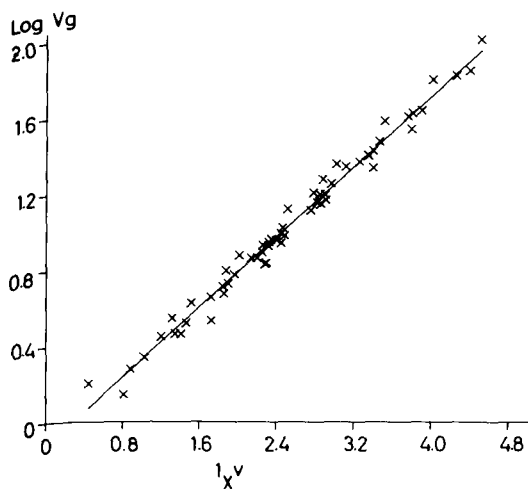


FIG. 1. A computer drawn graph of $\log Vg$ versus $1/X^v$ for the sixty esters, aldehydes, ketones and alcohols.

and morphine-like drugs) was correlated with molecular connectivity. Single parameter equations all produced poor correlation as might be expected for such a diverse and complex group of compounds. Several multi-parameter equations were generated and equation 20 was the superior equation utilizing connectivity indices up to the third order.

$$RI = 151.880^0X - 186.022^1X^v + 235.569^3X_p + 575.211 \quad (20)$$

$n = 41, r = 0.911, s = 193.029$

The equation has a lower correlation coefficient and a poorer standard error than the previous equations. This can be accounted for when the structures of the molecules are considered (see Table 3 for a list of compounds). It would also indicate that the use of up to third order connectivity terms is not providing a suitable description of the series of molecules.

A computer program CFUNC* recently obtained allows the calculation of all connectivity indices up to the sixth order. Using this program a preliminary investigation of the relation between RI and molecular connectivity (incorporating these higher order terms) for the drug molecules was performed. Equation 21 was generated which shows improved correlation and a reduction of 20% in the standard error.

$$RI = 164^0X - 76.5^1X + 31^3pcX + 390 \quad (21)$$

$n = 41, r = 0.941, s = 157$

* Supplied by Professor L. G. Hall

Table 3. Comparison of observed and calculated retention parameters for drug molecules.

Drug	R.I. (OBS)	R.I. (Calc)*	$\Delta R.I.$
Amethacaine	2215	2066	149
Benzocaine	1530	1476	54
Bupivacaine	2270	2257	13
Butacaine	2470	2307	163
Butanilcaine	2010	1929	81
Chloroprocaine	2200	2061	139
Cinchocaine	2690	2557	133
Cyclomethacaine	2225	2520	— 295
Lignocaine	1860	1983	— 123
Mepivacaine	2075	2025	50
Prilocaine	1845	1840	5
Procaine	1995	1913	82
Propoxycaine	2320	2290	30
Amphetamine	1110	1269	— 159
Benzphetamine	1850	1892	— 42
<i>NN</i> -Dimethylamphetamine	1230	1465	— 235
Methylamphetamine	1170	1345	— 175
<i>p</i> -Nitromethylamphetamine	1655	1677	— 22
<i>N</i> -Ethylamphetamine	1290	1428	— 138
<i>N</i> -Propylamphetamine	1330	1504	— 174
Codeine	2385	2381	4
Desomorphine	2275	2166	109
Dextromethorphan	2115	2141	— 26
Dihydrocodeine	2365	2381	— 16
Dihydrocodeinone	2425	2381	44
Dihydromorphine	2440	2300	140
Ethylmorphine	2415	2459	— 44
Hydrocodone	2425	2381	44
Hydromorphone	2490	2301	189
Isoquinoline	1440	1174	266
Levallorphan	2340	2220	120
Levorphanol	2225	2063	162
Methyldihydromorphine	2375	2504	— 129
6-Monoacetylmorphine	2480	2581	— 101
Morpheridine	2500	2515	— 15
Morphine	2435	2300	135
Nalorphine	2570	2458	112
Oxycodone	2425	2592	— 167
Oxymorphone	2520	2512	8
Papaverine	2805	2567	238
Pholcodine	2380	2846	— 466

*Using equation 21.
Column SE 30.

Using this equation Table 3 was compiled which gives a comparison of observed and calculated RI values. A more detailed examination of this series involving the valence connectivity is now under way.

This paper demonstrates that a relationship exists between molecular connectivity and gas chromatographic retention parameters for groups of molecules with diverse structures. In some cases a single parameter equation will carry enough descriptive power to allow a good correlation to be observed. When the drug molecules are considered poorer correlation is observed even using a multiparameter equation.

Considering the mixed group of 40 alcohols, aldehydes, ketones, esters and ethers, good cor-

relation was obtained with the non-polar columns whereas poor correlation was observed with the polar columns. These results are in agreement with Michotte & Massart's work where the best correlation was obtained with the non-polar squalene column. Kier & Hall (1979) pointed out that chromatographic behaviour depended on topological and non-topological structural characteristics. It would therefore appear that on the non-polar columns the non-topological structural characteristics were far less important than on the polar columns. Such properties as volatility and partition coefficient (important to the Retention Index) have been shown to correlate with molecular connectivity (a topological index). These properties may play a major role in the variation of RI on non-polar columns. Properties such as hydrogen bonding may be important in the variation of RI on polar columns and the group variation in hydrogen bonding (and other properties) cannot be explained by molecular connectivity.

It is hoped the relationship between molecular connectivity and Retention Parameters may have

applications as an aid to structure elucidation. Thus if the retention index of an unknown compound is obtained on a column and the equation relating connectivity indices to R.I. is available then the determination of the connectivity index (indices) should be possible. A knowledge of this index would then allow predictions to be made about the possible structures of the compound.

REFERENCES

- Csicsery, S. M., Pines, H. (1962) *J. Chromatogr.* 9: 34-43
- Kier, L. B., Hall, L. H. (1976) *Molecular Connectivity in Chemistry and Drug research*. London, Academic Press, Ltd
- Kier, L. B., Hall, L. H. (1979) *J. Pharm. Sci.* 68: 120-121
- McReynolds, W. O. (1966) *Gas Chromatographic Retention Data*, Preston Technical Abstracts Company
- Michotte, Y., Massart, D. L. (1977) *J. Pharm. Sci.* 66: 1630-1632
- Millership, J. S., Woolfson, A. D. (1978) *J. Pharm. Pharmacol.* 30: 483-485
- Moffat, A. C. (1975) *J. Chromatogr.* 113: 69-94