

The relation between molecular connectivity and gas chromatographic retention data

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 recently developed topological index, molecular connectivity (χ), has been correlated using multiple regression analysis with gas chromatographic retention time (Rt) for various series of compounds. For saturated and unsaturated aliphatic hydrocarbons, and aliphatic aldehydes, good correlation with log Rt was achieved by one-parameter linear equations in terms of the first-order connectivity index (${}^1\chi$). The equation for aliphatic alcohols required an additional parameter, the valence connectivity (${}^v\chi$) for satisfactory correlation. One-parameter equations using ${}^1\chi$ also gave good correlation with log Rt for three series of drug molecules, amphetamines, barbiturates and phenothiazines. In all cases the equations generated gave good agreement between calculated and observed log Rt values.

The correlation of both chemical and biological properties with chemical structure is well-known. Perhaps the best examples of such quantitative structure-activity relations (QSAR) are the linear free energy-related (LFER) models of Hammett (1940) and Hansch & Fujita (1964). The LFER model generally involves a congeneric series of molecules whose reactivity is correlated with physicochemical parameters rather than directly with molecular structure. An alternative to this approach is the use of a topologically-derived index, where molecular topology may be described as a structural representation of the molecule in terms of its bonding pattern.

Of the various molecular topological indices that have been proposed, that of Randic (1975) is based on a mathematical ordering of the degree of molecular branching within an alkane series, and has the advantage of being readily calculable. The Randic scheme has been developed by Kier, Hall & others (1975), who have defined a parameter, molecular connectivity (χ), which quantifies the degree of molecular branching for a wide variety of molecules.

Molecular connectivity has been successfully correlated, using multiple regression analysis, with a number of properties which appear to be dependent on molecular topology. These included solubility and boiling point (Hall, Kier & Murray, 1975), density (Kier, Murray & others, 1976) and partition coefficient (Murray, Hall & Kier, 1975). Since the latter case involved distribution of a compound between two liquid phases, it appeared of interest to examine the utility of χ in correlating molecular structure with gas chromatographic retention data, where g.c.

retention involves an equilibrium between a liquid and a gas phase. Initial work (Millership & Woolfson, 1977) indicated that a linear relation existed between the common logarithm of retention time (log Rt) and χ . In the present study a linear relation is confirmed between log Rt and the connectivity index for relatively simple series of straight and branched chain alkanes/alkenes, alcohols and aldehydes. The technique is then extended to encompass g.c. data from complex series of drug molecules.

CALCULATION OF MOLECULAR CONNECTIVITY PARAMETERS

Calculations of molecular connectivity made follow the method described by Kier & Hall (1976). The general form of the index, the first-order connectivity term (${}^1\chi$), 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 subgraphs (where a subgraph consists, in this case, of a bond between two atoms, i and j), ${}^1\chi$ is calculated by equation 1.

$${}^1\chi = \sum_s^n (\delta_i \delta_j)^{-1} \quad \dots \quad (1)$$

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

$$\delta_i^v = Z_1^v - h_i \quad \dots \quad (2)$$

where δ_i^v is the vertex valency, Z_1^v is the number of valence electrons and h_i is the number of hydrogen atoms that are suppressed.

Where hetero-atoms were present in the molecular structure, the valence connectivity (${}^v\chi$) has on

* Correspondence.

occasions been used (Kier & Hall, 1976). This is calculated similarly to ${}^1\chi$, except that the vertex valency of the hetero-atom was found from equation 2.

RESULTS AND DISCUSSION

The first group of compounds investigated was eighteen saturated aliphatic hydrocarbons. These hydrocarbons were both straight and branched chain compounds containing up to eight carbon atoms. The gas chromatographic retention data used was that of Csicsery & Pines (1962) (column—33% dimethylsulpholane on firebrick 110/120 mesh, carrier gas—helium, temp. 43°). Table 1 lists the relative retention

Table 1. Comparison of observed and calculated retention parameters and molecular connectivity (χ) for saturated aliphatic hydrocarbons.

	Log Rt (obs)	Log Rt (calc)	Δ Log Rt	χ
Ethane	0.3617	0.4600	0.0983	1.000
Propane	0.7853	0.7912	0.0059	1.414
Isobutane	1.0414	1.0456	0.0042	1.732
n-Butane	1.2148	1.1912	-0.0236	1.914
Neopentane	1.2305	1.2600	0.0295	2.000
Isopentane	1.5052	1.4760	-0.0292	2.270
n-Pentane	1.6128	1.5912	-0.0216	2.414
2,2-Dimethylbutane	1.7324	1.7088	-0.0236	2.561
2-Methylpentane	1.8692	1.8760	0.0068	2.770
2,3-Dimethylbutane	1.8797	1.7744	-0.1053	2.643
3-Methylpentane	1.9425	1.9056	-0.0369	2.807
n-Hexane	2.000	1.9912	-0.0088	2.914
3,3-Dimethylpentane	2.2305	2.1568	-0.0737	3.121
2,2,4-Trimethylpentane	2.3054	2.3936	0.0882	3.417
2,3-Dimethylpentane	2.2833	2.2048	-0.0785	3.181
n-Heptane	2.3766	2.3912	0.0146	3.414
2,5-Dimethylhexane	2.4487	2.5608	0.1121	3.626
n-Octane	2.7497	2.7912	0.0415	3.914

times (with respect to n-hexane), log relative retention times (the log of {relative retention time \times 100} was used to simplify calculation) and the molecular connectivities of the eighteen hydrocarbons. Regression analysis of the data in Table 1 showed a poor correlation between relative retention time and molecular connectivity, as shown in equation 3.

$$Rt = 1.465 {}^1\chi - 2.615 \dots \quad (3)$$

$$n = 18, r = 0.829, s = 0.804$$

Correlation of log Rt with molecular connectivity was far superior and the standard error showed a marked reduction (eqn 4).

$$\log Rt = 0.800 {}^1\chi - 0.340 \dots \quad (4)$$

$$n = 18, r = 0.996, s = 0.061$$

Equation 4 was used to calculate log Rt values of the 18 hydrocarbons and the results of these calculations, together with the observed results, are shown in Table 1. Values for Δ log Rt show that the

observed and calculated results differ by less than 5% in most cases.

The second set of compounds investigated consisted of 28 monounsaturated hydrocarbons. The retention data used were those of Csicsery & Pines (1962). Once again, excellent correlation was found between log Rt and the connectivity index as shown by equation 5.

$$\log Rt = 0.767 {}^1\chi + 0.284 \dots \quad (5)$$

$$n = 28, r = 0.969, s = 0.139$$

The gas chromatographic retention data for a series of aliphatic alcohols were obtained in our own laboratory (column—15% Carbowax 20M on Chromasorb W 80/100 mesh, carrier gas—nitrogen, temp. 100°). Poor correlation was obtained using both log Rt and Rt with ${}^1\chi$ and the valence connectivity (${}^v\chi$), respectively. However, a combination of these connectivity parameters correlated well with log Rt (eqn 6).

$$\log Rt = 2.370 {}^1\chi - 2.109 {}^v\chi - 1.086 \quad (6)$$

$$n = 10, r = 0.991, s = 0.029$$

The need for a two-parameter equation led to the postulation that inaccuracies in our determination of retention data may have arisen. However, investigations of retention data from literature sources produced correlations which closely paralleled our own.

Finally, a group of 12 saturated aliphatic aldehydes and ketones was studied. Gas chromatographic data was that of Ralls (1960) (column—30% Carbowax 1540 on firebrick, mesh 80/100, carrier gas—helium, temp. 90°). Good correlation was obtained between log Rt and valence connectivity, as shown in equation 7.

$$\log Rt = 0.356 {}^v\chi + 0.376 \dots \quad (7)$$

$$n = 12, r = 0.942, s = 0.070$$

Having established correlations for these relatively simple molecular series, the method was extended to more complex drug molecules. Firstly, nine amphetamine derivatives were investigated. The g.c. data used was that of Beckett, Tucker & Moffat (1967) (column—2.5% SE-30 on Chromasorb G A/W DMCS treated, 80/100 mesh, carrier gas—nitrogen, temp. 120°). The regression equation of best fit was equation 8.

$$\log Rt = 0.503 {}^1\chi - 1.673 \dots \quad (8)$$

$$n = 9, r = 0.998, s = 0.093$$

Secondly, a group of barbiturates were studied. The retention data used was that of Parker & Kirk (1961) (column—5% SE-30 on firebrick 100/120 mesh, carrier gas—argon, temp. 180°). Equation 9 gives the result of the correlation.

$$\log Rt = 0.243 \chi - 0.483 \quad \dots (9)$$

$n = 13, r = 0.916, s = 0.057$

Finally, equation 10 shows the resulting correlation for a group of phenothiazine derivatives, data of Parker, Fonton & Kirk (1962) (column—0.05% SE-30 on glass microbeads, 80/100 mesh, carrier gas—argon, temp. 165°).

$$\log Rt = 0.454 \chi + 0.562 \quad \dots (10)$$

$n = 10, r = 0.980, s = 0.078$

The linear relations between $\log Rt$ and number of carbon atoms for straight-chain aliphatic hydrocarbons is well-established. When data for branched chain saturated hydrocarbons are included, this linear relation no longer holds. Representation of molecular structure by the connectivity index, however,

has been shown to result in linear correlations with $\log Rt$ for these hydrocarbons (eqn 4) and for a wide range of molecular series (eqns 5 to 10). This correlation does not, of course, imply that any direct physical significance can be attributed to the connectivity index in accounting for the various interactions between solute and stationary phase. Nevertheless, the utility of χ in correlation of g.c. data is clearly demonstrated.

Recently, Michotte & Massart (1977) have obtained poor correlation with χ and g.c. retention indices for alcohols, ethers, esters and ketones. The present study indicates that, by using the logarithm of retention time, successful correlation with χ is achieved. The linear relation established enabled prediction of $\log Rt$ for compounds not present in the original series. As an example, equation 4 predicts $\log Rt$ for 2,4-dimethylpentane as 2.16, which is in good agreement with the observed value of 2.08 (Csicsery & Pines, 1962). Conversely, having determined $\log Rt$ for an unknown member of a given series, it may be possible to use the predicted value of χ as an aid to structure determination.

REFERENCES

- BECKETT, A. H., TUCKER, G. T. & MOFFAT, A. C. (1967). *J. Pharm. Pharmac.*, **19**, 273–294.
- CSICSERY, S. M. & PINES, H. (1962). *J. Chromat.*, **9**, 34–43.
- HALL, L. H., KIER, L. B. & MURRAY, W. J. (1975). *J. pharm. Sci.*, **64**, 1974–1977.
- HAMMETT, L. P. (1940). *Physical Organic Chemistry*. New York: McGraw-Hill.
- HANSCH, C. & FUJITA, T. (1964). *J. Am. chem. Soc.*, **86**, 1616–1626.
- KIER, L. B. & HALL, L. H. (1976). *Molecular Connectivity in Chemistry and Drug Research*, London: Academic Press.
- KIER, L. B., HALL, L. H., MURRAY, W. J. & RANDIC, M. (1975). *J. pharm. Sci.*, **64**, 1971–1974.
- KIER, L. B., MURRAY, W. J., RANDIC, M. & HALL, L. H. (1976). *Ibid.*, **65**, 1226–1230.
- MICHOTTE, Y. & MASSART, D. L. (1977). *Ibid.*, **66**, 1630–1632.
- MILLERSHIP, J. S. & WOOLFSON, A. D. (1977). *J. Pharm. Pharmac.*, **29**, Suppl., 75P.
- MURRAY, W. J., HALL, L. H. & KIER, L. B. (1975). *J. pharm. Sci.*, **64**, 1978–1981.
- PARKER, K. D., FONTON, C. R. & KIRK, P. L. (1962). *Analyt. Chem.*, **34**, 757–760.
- PARKER, K. D. & KIRK, P. L. (1961). *Ibid.*, **33**, 1378–1381.
- RALLS, J. W. (1960). *Ibid.*, **32**, 332–336.
- RANDIC, M. (1975). *J. Am. Chem. Soc.*, **97**, 6609–6615.