On the Use of the Weighted Identification Numbers in the QSAR Study of the Toxicity of Aliphatic Ethers

B. BOGDANOV,* S. NIKOLIĆ, A. SABLJIĆ, AND N. TRINAJSTIĆ
The Rugjer Bošković Institute, P.O.B. 1016, 41001 Zagreb, Croatia, Yugoslavia

S. CARTER
Department of Chemistry, The University of Reading, Reading RG6 2AD, England, United Kingdom

Abstract

We have examined an application of the weighted identification number in the QSAR study of the toxicity of aliphatic ethers on mice. The results obtained are superior to those achieved by the connectivity index.

Introduction

Recently a novel graph-theoretical index, known as the weighted identification (WID) number, has been introduced [1], which appears to be a highly selective structural descriptor. It is also worth noting that the WID number can be computed straightforwardly for any structure [1, 2].

We decided to test the applicability of the WID number in quantitative structure-activity relationships (QSAR) studies. The toxicity of aliphatic ethers was selected for this purpose because the problem has already been treated with the connectivity index of Randić [3] with some success [4]. The connectivity index is so far the most successful graph-theoretical descriptor used in QSAR work [5, 6] and thus we will be able to investigate how the WID number performs in comparison with the connectivity index on the same sample.

People have been interested in the anesthetic activity of aliphatic ethers since the discovery of diethyl ether in 1542 [7]. Interest has been particularly focused on the toxicities of ethers [8]. We will consider the correlation between the WID number and the toxicities (pC) of a set of 21 aliphatic ethers on mice [4, 8] in an attempt to produce a QSAR model of predictive power. We will also carry out calculations with the connectivity index using exactly the same types of regression analyses as those employed for the WID number.

---

* Permanent address: Department of Chemistry, University of Skopje, Skopje, Macedonia, Yugoslavia.

INTERNATIONAL JOURNAL OF QUANTUM CHEMISTRY: QUANTUM BIOLOGY SYMPOSIUM 14, 325–330 (1987)

CCC 0360-8832/87/010325-06 $04.00

The WID Number

In presenting a brief derivation of the WID number we will use graph-theoretical language for convenience [9, 10]. The incentive for the development of the WID number was Randić's molecular identification (ID) number [11] and its successful application in QSAR studies [12, 13]. However, isomeric trees were found with the same ID number [1]. This fact stimulated us to look for a number with much greater selectivity than Randić's ID number, and if such a number could be found, to investigate whether it could be used in QSAR studies. Let $G = (V, E)$ be a graph with the vertex-set $V = V(G)$ and the edge-set $E = E(G)$. Let $V = (v_1, v_2, \ldots, v_N)$ be a labelling of V. The distance between the vertices v_i and v_j is denoted by $d(i,j)$. Note that $d(i,j) = d(j,i)$, and $d(i,i) = 0$. Distances $d(i,j)$ are elements of the distance matrix of G, $\mathbf{D} = \mathbf{D}(G)$ [14–16]. The distance-sum D_i in $\mathbf{D}$ is defined by [17–19]:

$$D_i = \sum_{i=1}^{N} d(i,j); \qquad 1 \le i \le N \tag{1}$$

The distance sum has been used by Seybold [20] as a measure of the compactness or centrality of a particular site in a molecule. The distance sums may be easily obtained with any of several available computer-oriented algorithms for constructing the distance matrix for any structure [21, 22] and they simply represent the sums of the elements in the rows (or columns) of the distance matrix. The weights of edges w_{ij} in G are defined as [3, 18]:

$$w_{ij} = \begin{cases} (D_i D_j)^{-1/2} & \text{if } d(i,j) = 1; 1 \le i \le N, 1 \le j \le N \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

They represent the elements of the matrix of weights, $\mathbf{W} = \mathbf{W}(G)$. Let $w = (v_{i_1}, v_{i_2}, \ldots, v_{i_N})$ be a walk [9] of length k. The weight of this walk is defined by:

$$\prod_{j=1}^{k} (D_{i_j} D_{i_{j+1}})^{-1/2} = \prod_{j=1}^{k} w_{i_j i_{j+1}} \tag{3}$$

The weight of all walks of length k between vertices v_{i_1} and $v_{i_{k+1}}$ is given by:

$$\sum_{i} \prod_{j=1}^{k} w_{i_j i_{j+1}} \tag{4}$$

The entry $w_{ij}^{(k)}$ of $\mathbf{W}^k$ is the sum of all weighted walks of length k. The WID number of G is then defined as follows:

$$\mathrm{WID}(G) = N - (1/N) + (1/N)^2 \cdot \mathrm{ID}^*(G) \tag{5}$$

where:

$$\mathrm{ID}^*(G) = \sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij}^* \tag{6}$$

Note that:

$$(w_{ij}^*)_{1 \le i,j \le N} = W^* \tag{7}$$

and:

$$\mathbf{W}^* = \sum_{k=0}^{N-1} \mathbf{W}^k. \tag{8}$$

Note that because of the way in which the WID number is constructed, its limits are:

$$N \leq \mathrm{WID}(G) \leq N + 1 \tag{9}$$

for each graph G with N vertices.

We have devised a computer program for calculating the WID numbers which starts with the distance matrix and proceeds via D_i, $\mathbf{W}$, $\mathbf{W}^k$, ID* and finally ends with the WID number [23, 24].

Results and Discussion

The toxicity of 21 aliphatic ethers together with their WID numbers and connectivity indices are given in Table I. The WID numbers are calculated as shown in the previous section. The connectivity index χ is calculated by the following formula [3]:

$$\chi = \sum_{\text{bonds}} (m_i \cdot n_j)^{-1/2} \tag{10}$$

where m_i and n_j are the valencies of the endpoints of the bond i–j.

TABLE I. Toxicities of aliphatic ethers ($R_1 - 0 - R_2$) on mice pC and the corresponding WID numbers and connectivity indices.

Ether	pC[a]	WID	χ
Dimethyl	1.43	3.29255	1.414
Methyl ethyl	1.74	4.12444	1.914
Methyl propyl	2.45	5.05815	2.414
Methyl isopropyl	2.26	5.07386	2.270
Methyl butyl	2.70	6.03157	2.914
Methyl isobutyl	2.79	6.03985	2.770
Methyl secbutyl	2.79	6.03723	2.808
Methyl terbutyl	2.79	6.05011	2.561
Methyl pentyl	2.88	7.01917	3.414
Diethyl	2.22	5.05815	2.414
Ethyl propyl	2.60	6.03157	2.914
Ethyl isopropyl	2.60	6.03723	2.700
Ethyl butyl	2.82	7.01917	3.414
Ethyl isobutyl	2.82	7.02315	3.270
Ethyl secbutyl	2.85	7.02154	3.308
Ethyl terbutyl	2.92	7.02712	3.061
Ethyl pentyl	3.00	8.01257	3.914
Ethyl terpentyl	3.15	8.01811	3.621
Dipropyl	2.79	7.01917	3.414
Propyl isopropyl	2.82	7.02154	3.270
Di-isopropyl	2.82	7.02449	3.126

[a] Refs. 4 and 8.

We examined two types of correlations: (a) linear least-squares fit

$$pC = a + b \cdot I \tag{11}$$

and (b) quadratic least-squares fit

$$pC = a + b \cdot I + c \cdot I^2 \tag{12}$$

where I = WID or χ. The results of the above regression analyses are given in Tables II and III.

From the statistical data in Tables II and III, we see that the WID number is superior to the connectivity index. It appears, both from earlier comparative studies [25], and from this work, that of all single graph-theoretical indices used for the correlation with the toxicities of aliphatic ethers, the most promising QSAR model is achieved with the WID number. Even when a polyparametric regression equation with several kinds of graph-theoretical indices is employed, the quality of the QSAR model with only the WID number is unsurpassed.

One possible reason for this good performance of the WID number is discussed here. If we carry out the regression analyses by using only the number of atoms in the ether, the following statistical equations are obtained for the two cases considered above:

$$pC = 1.002 + 0.314 \cdot N; \qquad r = 0.947 \qquad s = 0.133$$
$$F^{1,19} = 166\ r^2 \text{ (adjusted)} = 0.892 \tag{13}$$

TABLE II. Statistical characteristics of a linear relationship between the aliphatic ether toxicities on mice and the WID numbers and connectivity indices.

	Statistical data					
I	a	b	r	s	$F^{1,19}$	r^2(adjusted)
WID	0.602	0.325	0.942	0.139	149	0.881
χ	0.792	0.633	0.909	0.172	91	0.818

TABLE III. Statistical characteristics of a quadratic relationship between the aliphatic ether toxicities and the WID numbers and connectivity indices.

	Statistical data						
I	a	b	c	r	s	$F^{2,18}$	r^2(adjusted)
WID	−1.321	1.020	−0.060	0.976	0.090	181	0.947
χ	−1.019	2.046	−0.261	0.955	0.123	93	0.902

$$pC = 0.027 + 0.767 \cdot N - 0.048 \cdot N^2; \; r = 0.975 \quad s = 0.092$$
$$F^{2,18} = 174 \quad r^2 \text{ (adjusted)} = 0.945 \tag{14}$$

These results are rather nice and one may be tempted to recommend the number of atoms to be used in many QSAR models. However, the number of atoms is a descriptor of low discriminatory power, it cannot, of course, differentiate isomeric molecules.

From the structure of the formula for the WID number [see Eq.(5)], we see that this number is rather closely related to the number of atoms in a molecule. Hence, the WID number could be simply presented as:

$$\text{WID} = N + \text{corr.} \tag{15}$$

where:

$$\text{corr.} = -(1/N) + (1/N)^2 \cdot \text{ID}^*(G) \tag{16}$$

The correction (16) for a great number of chemical graphs is rather small. It will increase with the increasing complexity of a graph. The superiority of the WID over N is clear in the case of isomers. For example, all 366,319 isomers of $C_{20}H_{42}$ alkane are differentiated by the WID number while they all have the same $N = 20$. This sensitivity of the WID number is related to the small correction given in expression (16).

In the past, very discriminative graph-theoretical indices have been found not to be particularly useful in QSAR studies. The large amount of structural information contained in such graph-theoretical indices may obscure those factors that are significant for a particular property that is to be modelled via QSAR technology. A good example to illustrate this point is Balaban's index [18], which is a highly discriminative descriptor which has so far shown little use in QSAR studies [26].

Concluding Remarks

We wish to point out that the WID has many good features for applications to QSAR studies. It is the most discriminative graph-theoretical index that has been found to date. The WID is designed to avoid large structural information which may obliterate its use in constructing QSAR models. Therefore, the WID number shows potential for use in QSAR work. However, more work is needed before the range of its applicability is established. Some research in this direction is already in progress [27].

Acknowledgment

This work has been sponsored in part by the British Council.

Bibliography

[1] K. Szymanski, W. R. Müller, J. V. Knop, and N. Trinajstić, Int. J. Quantum Chem.: Quantum Chem. Symp. **20**, 173 (1986).

[2] N. Trinajstić, S. Nikolić, and D. Horvat, Kem. Ind. (Zagreb), in press.
[3] M. Randić, J. Am. Chem. Soc. **97**, 6609 (1975).
[4] L. B. Kier and L. H. Hall, *Molecular Connectivity in Chemistry and Drug Research* (Academic, New York, 1976), p. 202.
[5] D. H. Rouvray, Sci. Am. **254**, 40 (1986).
[6] L. B. Kier and L. H. Hall, *Molecular Connectivity in Structure-Activity Analyses* (Research Studies Press, Letchworth, 1986).
[7] V. Cordus, Pharmacorum Conficiendorum Ratio, Vulgo Vocant Dispensatorium, Norimbergae, 1546.
[8] D. P. Marsh and C. D. Leake, Anaesthesiology **11**, 455 (1950).
[9] F. Harary, *Graph Theory* (Addison-Wesley, Reading, MA, 1972), second printing.
[10] N. Trinajstić, Chemical Graph Theory (CRC Press, Boca Raton, FL, 1983).
[11] M. Randić, J. Chem. Inf. Comput. Sci. **24**, 164 (1984).
[12] M. Randić, Int. J. Quantum Chem.: Quantum Biol. Symp. **11**, 137 (1984).
[13] M. Randić, in *Molecular Basis of Cancer, Part A: Macromolecular Structure, Carcinogens and Oncogens*, R. Rein, Ed. (Liss, New York, 1985), p. 309.
[14] D. H. Rouvray, in *Chemical Applications of Graph Theory* (Academic, London, 1976), p. 175.
[15] Ref. 10, p. 44.
[16] D. H. Rouvray, in *Mathematics and Computational Concepts in Chemistry*, N. Trinajstić, Ed. (Horwood, Chichester, 1986), p. 295.
[17] M. Randić, Math. Chem. (Mülheim/Ruhr) **7**, 5 (1979).
[18] A. T. Balaban, Chem. Phys. Lett. **89**, 399 (1982).
[19] N. Trinajstić, M. Randić, and D. J. Klein, Acta Pharm. Jugosl. **36**, 267 (1986).
[20] P. G. Seybold, Int. J. Quantum Chem.: Quantum Biol. Symp. **10**, 95 (1983); ibid. **10**, 103 (1983).
[21] M. Bersohn, J. Comput. Chem. **4**, 110 (1983).
[22] W. R. Müller, K. Szymanski, J. V. Knop and N. Trinajstić, J. Comput. Chem. **8**, 170 (1987).
[23] S. Carter, N. Trinajstić and S. Nikolić, Acta Pharm. Jugosl., **37**, 37 (1987).
[24] S. Carter, S. Nikolić, and N. Trinajstić, in preparation.
[25] O. Mekenyan, D. Bonchev, A. Sabljić, and N. Trinajstić, Acta Pharm. Jugosl., **37**, 75 (1987); M. Randić, A. Sabljić, S. Nikolić and N. Trinajstić, Quant. Struct.-Act. Rel., submitted.
[26] A. Sabljić, J. Chromatogr. **314**, 1 (1984); ibid. **319**, 1 (1985).
[27] S. Carter, S. Nikolić, and N. Trinajstić, work in progress.

Received April 30, 1987