Med. Sci. Res., 1988; **16**, 185-186

On the use of ID numbers in drug research: a QSAR study of neuroleptic pharmacophores

Stuart Carter, Nenad Trinajstić and Sonja Nikolić*

Department of Chemistry, The University of Reading, Reading RG6 2AD, England, UK and
**The Rugjer Bosković Institute, P.O.B. 1016, 41001 Zagreb, Croatia, Yugoslavia*

One of the growing interests in QSAR (quantitative structure-activity relationships) studies is the use of non-empirical models which can be used to correlate mathematically derived structural descriptors with the biological activities of families of molecules (or drugs) [1–8]. In this respect, we have recently investigated the application of graph-theoretical (topological) indices [7, 9, 10], originally proposed by Randić [11, 12], in the study of anti-tumour activities of mitindomides [13]. For the selection of drugs studied, the overall agreement between calculated and experimental toxicities was encouraging, bearing in mind the difficulties in obtaining reliable *in vivo* experimental measurements in cancer research. In the light of this initial success, we decided to test further the method and the anti-emetic activities of a neuroleptic family of phenothiazines were selected for this purpose. The present communication is a report on this work.

The index proposed by Randić [11], and used in our earlier work [13], was an identification (ID) number based on path counts in the molecule. The ID number is defined as follows. For any molecular structure S, the mapping (g) from its bond-set, B(S), to the set of real numbers R:

$$g : B(S) \rightarrow R \quad \ldots\ldots (1)$$

is defined by:

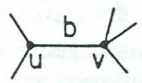
$$g(b) = g(u, v) = (\deg u \cdot \deg v)^{-1/2} \quad \ldots\ldots (2)$$

where b stands for the bond connecting atoms u and v:

$$b = (u, v) \in B(S) \quad \ldots\ldots (3)$$

From the above, it follows that g(b) represents a given bond weight [14]. In (2), deg u denotes the number of nearest neighbours of atom u in S, but in our earlier work [13] we also considered the valencies of atoms in the molecule.

Consider, for example, bond b in the following structure:



The degree of u is 3, and that of v is 4. Hence

$$g(b) = 1/(3 \cdot 4)^{1/2} = 1/(2\sqrt{3})$$

Let $p_l = b_1, b_2, \ldots, b_l$ $(l > 0)$ be a path of length l in S, where path is defined as a sequence of adjacent bonds which does not pass through the same atom more than once. The length l of path p_l is merely the number of bonds comprising the path. The mapping g may be extended (g*) to the set of all paths in S by:

$$g^* = \Pi_{i=1}^{l} g(b_i) \quad \ldots\ldots (4)$$

The ID number of the molecule S is then finally given (according to Randić [11, 12]) by:

$$ID(S) = N + \sum_{p_l} g^*(p_l) \quad \ldots\ldots (5)$$

where N is the number of atoms in the molecule and the summation is taken over all paths in S. For families of related molecules containing a common skeleton structure, the proposed ID is that for which the summation in (5) is taken only over those paths that originate at a skeleton atom. Such paths must then start at a skeleton atom, but thereafter are allowed to traverse both skeleton and substituent atoms, and may terminate at either type of atom. In this way, substituent effects are encompassed in the model.

The bond weights given by (2) do not differentiate between different chemical environments; in particular, there is no scope in the theory to accomodate heteroatoms. Furthermore, bonds of the same type, (e.g., C-C) but in different environments (e.g., in side chains or ring structures) can quite frequently be given the same bond weight by (2). One way to overcome this problem is to define the bond weights as:

$$g(b) = w_{uv} (\deg u \cdot \deg v)^{1/2} \quad \ldots\ldots (6)$$

where w_{uv} is a parameter that can be used in the differentiation between bonds of different types.

In the present work, we used bond weights given by (6), where w_{uv} was chosen so that all different bond types had different values of g(b). The different g(b) are then least-squares iterated so that the differences between calculated and observed activities are minimized. Due to the complexity of the anti-emetic activities of phenothiazines, it is not possible simply to relate it linearly to the mathematically derived

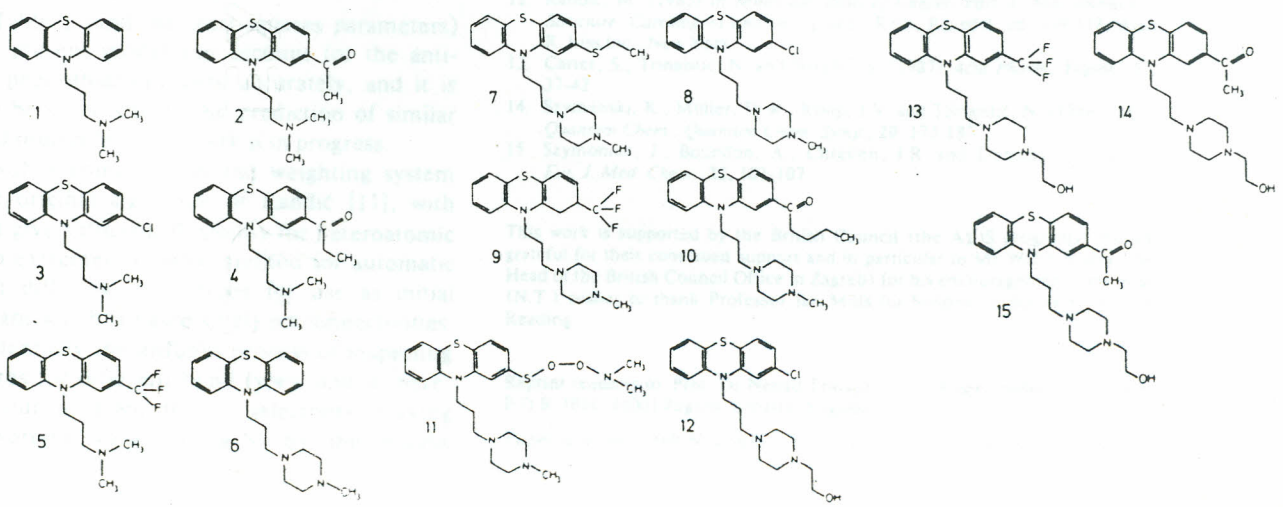


Figure 1: Diagrams of studied molecules.

Table 1: Comparison between calculated and experimental anti-emetic activities of phenothiazines

Molecule[a]		$-\log PD_{50}$ Calculated[b]	$-\log PD_{50}$ Experimental[c]	Error
Promazine	(1)	4.52	4.54	−0.02
Acepromazine	(2)	6.00	6.04	−0.04
Chlorpromazine	(3)	5.70	5.70	0.00
Propionylpromazine	(4)	6.04	5.94	0.10
Triflupromazine	(5)	6.37	6.41	−0.04
Perazine	(6)	5.18	5.16	0.02
Triethylperazine	(7)	6.40	6.40	0.00
Prochlorperazine	(8)	6.33	6.18	0.15
Trifluperazine	(9)	6.98	7.15	−0.17
Butaperazine	(10)	7.30	7.30	0.00
Thioperazine	(11)	8.15	8.15	0.00
Perphenazine	(12)	7.45	7.53	−0.08
Fluphenazine	(13)	8.08	7.86	0.22
Acetophenazine	(14)	7.74	7.70	0.04
Carphenazine	(15)	7.77	7.88	0.11

[a] *These molecules are shown in Figure 1.* [b] *The calculated values are defined by: A = 0.008554; B = 0.759786; C = −0.006937; D = 0.006803; and E = 0.280321 (see expression (7)).* [c] *Experimental values are taken from ref [15].*

ID. However, in order to achieve the desired correlation, it was found necessary to relate the skeleton ID to the biological activity (BA) by a relationship of the form:

$$BA = A + B \cdot (ID) + C \cdot (ID)^2 + D \cdot (ID)^{-1} + E \cdot \ln (ID) \quad (7)$$

This procedure has been applied to 15 phenothiazines which possess anti-emetic activity PD_{50} [15].

A computer program has been written in FORTRAN to determine the paths of all lengths in the molecule. Input to the program consists of atom connectivities which are used to determine the quantity g(b) in equation (6). The molecular ID is then calculated from the paths via equations (4) and (5), which are condensed forms of the general algorithm given by Randić [12]. The program is not, as yet, generally available.

The structures of these molecules are given in Figure 1, and the comparison between observed and calculated activities are given in Table 1. The agreement between experimental anti-emetic activities, $(-\log PD_{50})_{expt}$, and calculated anti-emetic activities $(-\log PD_{50})_{calc}$ was analyzed by the following linear least-squares fit equation:

$$(-\log PD_{50})_{expt} = a\,(-\log PD_{50})_{calc} + b \quad (8)$$

The following least-squares parameters were obtained: a = 0.995, b = 0.036, r = 0.996 (the correlation coefficient), SD = 0.098 (the standard deviation), F = 1601.490 (the F-ratio between the variances of observed and calculated values) and r^2 (adjusted) = 0.991.

Inspection of Table 1 (and the least-squares parameters) suggests that the present model can account for the anti-emetic action of phenothiazines very accurately, and it is hoped that it will be successful in the prediction of similar activities of related molecules. This work is in progress.

We conclude with a comment on the weighting system used in (6). The original algorithm of Randić [11], with $w_{uv} = 1$, does not give sufficient flexibility for heteroatomic systems, but is an extremely reliable method for automatic selection of most differing bond types for use as initial weights in a program which is based solely on connectivities. This reduces substantially the arduous process of inspecting molecular structures for different bond types, and is therefore retained in our program for its selectivity. Having established the various values of g(b) by the Randić algorithm (eq. (6) with $w_{uv} = 1$), bonds which are structurally the same, but are in different heteroatomic environments, are now distinguished by modifying the initial value of w_{uv} for those bonds which contain heteroatoms (1.1 would be a typical value). This procedure serves as an indicator to the least-squares routine to individually refine the g(b) of the two types of bonds. In his initial work, Randić attempted to interpret a particular bond type with a well-defined value of g(b). In our method, this is no longer needed since the minimisation procedure automatically gives the optimum values of g(b) for any particular property of the molecule under consideration.

1. Kier, L.B., Hall, L.H., Murray, W.J. and Randić, M. (1975) *J. Pharm. Sci.*, **64**, 1871-1974
2. Hall, L.H., Kier, L.B. and Murray, W.J. (1975) *J. Pharm. Sci.*, **64**, 1974-1977
3. Murray, W.J., Kier, L.B. and Hall, L.H. (1975) *J. Pharm. Sci.*, **64**, 1978-1981
4. Kier, L.B., Murray, W.J., Randić, M. and Hall, L.H. (1976) *J. Pharm. Sci.*, **65**, 1226-1230
5. Kier, L.B. and Hall, L.H. (1977) *J. Pharm. Sci.*, **65**, 1806-1809
6. Kier, L.B. and Hall, L.H. (1986) *Molecular Connectivity in Structure - Activity Analysis*, Research Studies Press, Chichester
7. Basak, S.C. (1987) *Med. Sci. Res.*, **15**, 605-606
8. Basak, S.C., Frane, C.M., Rosen, M.E. and Magnuson, V.R. (1987) *Med. Sci. Res.*, **15**, 887-888
9. Trinajstić, N. (1983) *Chemical Graph Theory*, Vol. II, pp. 105-140, CRC Press, Boca Raton, FL
10. Rouvray, D.H. (1986) *Sci. Am.*, 255, 40-47
11. Randić, M. (1984) *J. Chem. Inf. Comput. Sci.*, **24**, 164-175
12. Randić, M. (1985) in *Molecular Basis of Cancer, Part A: Macromolecular Structure, Carcinogens and Oncogenes*, (Rein, R., ed.), pp. 309-318, Alan R. Liss Inc., New York
13. Carter, S., Trinajstić, N. and Nikolić, S. (1987) *Acta Pharm. Yugosl.*, 37, 37-42
14. Szymanski, K., Müller, W.R., Knop, J.V. and Trinajstić, N. (1986) *Int. J. Quantum Chem., Quantum Chem. Symp.*, **20**, 173-183
15. Szymoniak, J., Boundon, A., Chretien, J.R. and Dubois, J.-E. (1987) *Eur. J. Med. Chem.*, **22**, 101-107

This work is supported by the British Council (the ALIS program). We are grateful for their continued support and in particular to Mr Walter Curry (the Head of the British Council Office in Zagreb) for his encouragement. One of us (N.T.) wishes to thank Professor Ian Mills for hospitality during his stay in Reading.

Reprint requests to: Prof. Dr Nenad Trinajstić, The Rugjer Bošković Institute, P.O.B. 1016, 41001 Zagreb, Croatia, Yugoslavia.

Paper received: 20th November, 1987; amended 25th January, 1988