Bull. Environm. Contam. Toxicol. 28, 373–378 (1982)

Structure-Toxicity Relationships of Selected Nitrogenous Heterocyclic Compounds. III. Relations Using Molecular Connectivity*

T. Wayne Schultz,[1] Lemont B. Kier,[2] and Lowell H. Hall[3]

[1]*Biology Division, Oak Ridge National Laboratory, Oak Ridge, TN 37830,* [2]*Department of Pharmaceutical Chemistry, Medical College of Virginia, Richmond, VA 23298,* [3]*Department of Chemistry, Eastern Nazarene College, Quincy, MA 02170*

The present life style enjoyed by Western man is directly related to progress in industry and agriculture and, consequently, heavily dependent upon xenobiotic compounds. The high living standards afforded by such chemicals, however, may be offset by the pernicious effects they manifest upon entering the environment. Because potential risks exist for every chemical synthesized, toxicity testing is required. The time and cost of generating toxicity data needed for even the most basic evaluation of any large number of industrial chemicals, such as subsets of the 45,000 compounds listed in the Toxic Substance Control Act registry, is rapidly becoming prohibitive. It is this need for a time and cost effective means of screening or ranking chemicals that has been the catalyst for the recent interest in quantitative structure activity relationship (QSAR) in ecotoxicology. QSAR is being demonstrated as a powerful research tool that permits ranking of industrial chemicals on the basis of meaningful correlations between their structure and biological activity.

Description of chemical structure can be broken down into two major categories, each leading to a series of molecular descriptors useful for formulating QSAR. The first general class of molecular descriptors are free-energy related physiochemical properties. Of main interest among these properties are the substituent constants, especially the Hammett electronic parameter σ, the Taft steric parameter E_S, and the hydrophobic parameter log P or π (HANSCH 1969, HANSCH and FUJITA 1964, HANSCH and LEO 1979).

The second broad class of molecular descriptors is based on quantitation of chemical structure that is encoded so the observed biological activity is expressed as a function of descriptors assigned to substituent groups and/or portions of the parent molecule. The latter class includes quantum mechanics, molecular orbital theory, and the less rigorous method of molecular connectivity (KIER and HALL 1976). The method of molecular connectivity generates a series of numerical indexes associated with one, two, or multiple bond fragments by quantitating certain aspects of chemical structure such as number

*Research supported jointly by the Environmental Protection Agency under Interagency Agreement 79-D-X0533, and the Office of Health and Environmental Research, U.S. Department of Energy, under contract W-7405-eng-26 with the Union Carbide Corporation.

0007-4861/82/0028-0373 $01.20

of atoms, branching, unsaturation, and heteroatom content. These indexes are calculated from values associated with non-hydrogen atoms in a molecule and are counts of electrons other than those bonding to hydrogen.

The objective of this investigation is to examine the QSAR between cellular response and molecular connectivity indexes for a series of 24 mono- nd dinitrogen heterocyclic compounds that increase in ring attachment and methyl substitution and that have possible isomeric differences.

MATERIALS AND METHODS

Biological Activity

The biological activity data used in this study was obtained from the studies of SCHULTZ et al. (1980) and SCHULTZ and CAJINA-QUEZADA (1982) on selected mono- and dinitrogenous heterocyclic compounds, respectively. The biological descriptor, log biological response (log BR), is defined as the reciprocal of the IGC50. The IGC50, the concentration (mmol/L) required to inhibit the growth of axenic cultures of the common freshwater ciliate Tetrahymena pyriformis strain GL-C by 50%, were calculated from concentration/absorbance least-squares linear regression analyses COOLEY et al. (1972).

Molecular Descriptor

The molecular connectivity index used as the molecular descriptor in this study was the $^{1}\chi^{v}$ (one-chi-v) index. χ is a weighted count of molecular substructure. The $^{1}\chi$ term is a summation of all the edges in the molecular structure weighted by the reciprocal square root valencies and thus is an index of molecular branching. The $^{1}\chi^{v}$ index is a nonempirical refinement of the $^{1}\chi$ that allows for valence assigned to heteroatoms in the hydrogen-suppressed substructure and thus reduces redundancy associated with isomers and analogs. The $^{1}\chi^{v}$ values reported here were determined by a computer program from input based on connection tables and atom identifiers, such as atomic number, number of valence electrons, and number of hydrogen atoms bonded to the given atom (KIER & HALL 1976, KIER 1980).

QSAR

The quantitative structure-activity relationships determined from this study were generated by the least-squares method of linear regression where Y was the log BR and X was the $^{1}\chi^{v}$ index.

RESULTS

The Chemical Abstract Service registry number, $^{1}\chi^{v}$ molecular connectivity index, log BR observed, log BR predicted, and residual value for each test chemical are compared in Table 1. The nitrogen-containing heterocycles examined in this investigation may be subdivided into two groups: (1) six-membered

TABLE 1

Molecular connectivity index and log biological responses for 24 nitrogen heterocyclic compounds

No.	Heterocycle	CAS number	$^{1}\chi^{v}$	Log BR (observed)	Log BR (predicted)[a]	Residual value
1	Pyridine	110-86-1	1.850	-1.19	-1.28	0.09
2	3-Methylpyridine	108-44-1	2.260	-1.02	-0.91	-0.11
3	2,6-Dimethylpyridine	108-48-5	2.691	-0.81	-0.52	-0.29
4	Pyrazine	290-37-9	1.699	-1.82	-1.42	-0.40
5	2-Methylpyrazine	109-08-0	2.120	-1.09	-1.04	-0.05
6	2,3-Dimethylpyrazine	5910-89-4	2.547	-0.87	-0.65	-0.22
7	Quinoline	91-22-5	3.264	0.01	0.00	0.01
8	2-Methylquinoline	91-63-4	3.685	0.47	0.39	0.08
9	2,6-Dimethylquinoline	877-43-0	4.096	0.68	0.76	-0.08
10	Quinoxaline	91-19-0	3.124	-0.30	-0.12	-0.18
11	2-Methylquinoxaline	7251-61-8	3.545	0.02	0.26	-0.24
12	2,3-Dimethylquinoxaline	2379-55-7	3.972	0.25	0.19	0.06
13	Acridine	260-94-6	4.679	1.40	1.29	0.11
14	Phenazine	92-82-0	4.549	1.40	1.18	0.23
15	Pyrimidine	289-95-2	1.699	-1.75	-1.42	-0.33
16	Pyridazine	289-80-5	1.716	-1.41	-1.41	0.00
17	Phthalazine	253-52-1	3.121	-0.34	-0.13	-0.21
18	Quinazoline	253-82-7	3.114	-0.29	-0.13	-0.16
19	Pyrrole	109-97-7	1.577	-1.11	-1.53	0.42
20	Indole	120-72-9	2.988	0.21	-0.25	0.46
21	1,2-Dimethylindole	875-79-6	3.799	0.84	0.49	0.35
22	Carbazole	86-74-8	4.405	0.91	1.04	-0.13
23	Pyrazole	288-13-1	1.437	-1.71	-1.66	-0.05
24	Imidazole	288-32-4	1.427	-1.00	-1.67	0.67

[a]Based on Equation 3

ring heterocyclics, compounds 1-18; and (2) five-membered ring heterocyclics, compounds 19-24. Plots of log BR vs $^1\chi^v$ reveal strong positive linear relationships (Fig. 1). Least-squares linear regression analyses of these data for the six-membered ring, five-membered ring, and combined compounds yield model 1, 2, and 3, respectively:

$$\log BR = 1.002\ ^1\chi^v - 3.334 \quad r = 0.988 \quad s = 0.15 \quad (1)$$

$$\log BR = 0.816\ ^1\chi^v - 2.437 \quad r = 0.966 \quad s = 0.32 \quad (2)$$

and

$$\log BR = 0.911\ ^1\chi^v - 2.969 \quad r = 0.962 \quad s = 0.27 \quad (3)$$

Statistical analysis indicates that Equation 3 describes 93% of the variation in the observed toxic response. The standard deviation of 0.27 log units is probably close to the experimental limit of accuracy of the system. The dissection of the data into the two subsets above leads to rather similar models (Equations 1 and 2). Thus, a common mechanism of action is implied.

Only one compound, imidazole, lies outside two standard deviations. It is worth noting that imidazole has the smallest $^1\chi^v$ value of the test heterocyclics. If imidazole is omitted, the relationship of $^1\chi^v$ to log BR for the remaining 23 compounds improves to r = 0.974, s = 0.23.

DISCUSSION

Recent QSAR between molecular descriptors and ecotoxicological data has centered on using physicochemical properties especially hydrophobic parameters. Such relationships have been described by KOPPERMAN et al. (1974) using a variety of phenols, ZITKO et al. (1976) with alkyl dinitrophenols, SCHULTZ et al. (1978) for benzenes, phenols, anilines, and nitrogen heterocyclics, and KÖNEMANN (1980) with chloro aromatics and alphatics, alcohols, and glycols. Similarly strong linear correlations between the log BR data used in this study and log P values has been demonstrated by SCHULTZ and CAJINA-QUEZADA (1982) and SCHULTZ et al. (1980).

Finding close relationships between log BR and the $^1\chi^v$ index (see Equations 1–3) makes it possible to describe some of the influences molecular structure have on toxicity. This dependence of log BR on $^1\chi^v$ indicates that with increasing number of atoms, toxicity levels also increase: compare compounds 1, 7, and 13; 4, 10 and 14; 19, 20, and 22 in Table 1. Similarly, the $^1\chi^v$ index encodes the information that successive methylation of a ring system increases toxicity: compare compounds 1, 2 and 3; 4, 5, and 6; 7, 8, and 9; 10, 11, and 12. However, when additional nitrogen atoms are introduced into molecules of the same size, the toxicity decreases: compare compounds 1 and 4, 7 and 10, 19 and 23. An exception to the latter case is noted with the tricyclic compounds numbers 13 and 14 where neither $^1\chi^v$ nor log BR vary much with the substitution of a second nitrogen atom.

Recently SCHULTZ and CAJINA-QUEZADA (1982) demonstrated log P dependent QSAR using this same toxicity data set. A

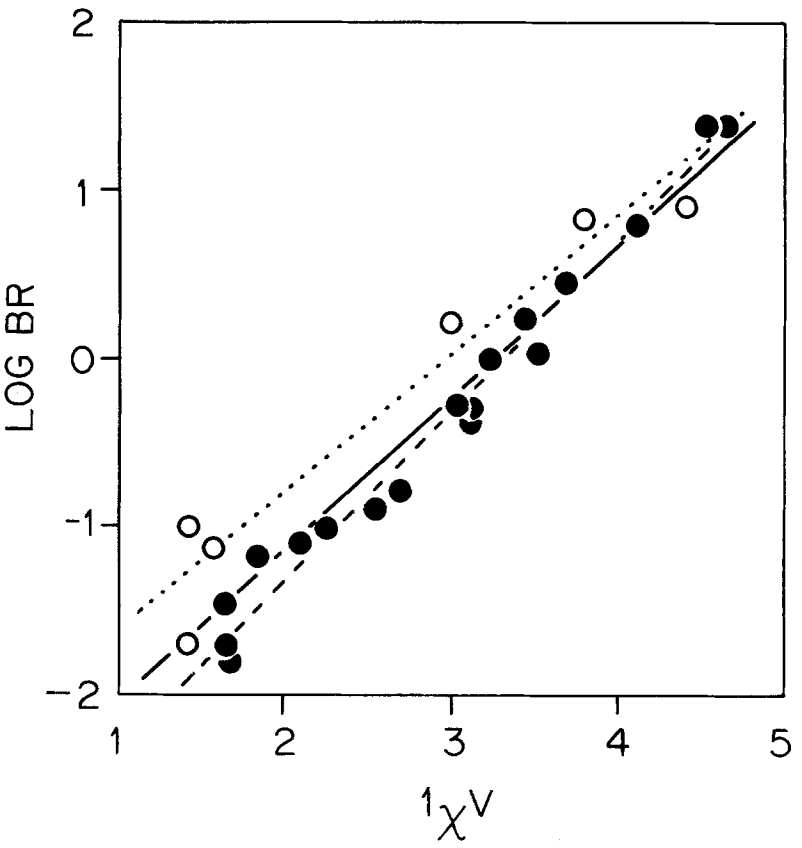

Fig. 1. Least-squares linear regression of log biological response (Y) vs $^1\chi^V$ (X) for six-membered ring heterocyclic compounds (closed circles) and five-membered ring heterocyclic compound (open circles). Dashed line equation 1; dotted line equation 2; solid line equation 3.

question of particular interest was how the present study compares with this previous work. Both studies demonstrate that toxicity increases with an increase in the number of atoms and degree of methylation per compound and that toxicity decreases with an increase in nitrogen substitution. However, the present study gives an improved correlation coefficient, r = 0.962 vs r = 0.898.

REFERENCES

COOLEY, N.R., J.M. KELTNER, and J. FORESTER: J. Protozool. 19, 636 (1972).

HANSCH, C.: Acc. Chem. Res. 2, 232 (1969).

HANSCH, C., and T. FUJITA: J. Am. Chem. Soc. 86, 1616 (1964).

HANSCH, C., and A. LEO: Substituent Constants for Correlation Analysis in Chemistry and Biology. New York: Wiley (1979).

KIER, L.B.: Molecular Connectivity as a Description of Structure for SAR Analysis. In S. H. Yalkowsky (ed.): Physical Chemical Properties of Drugs, p. 277 New York: Marcel Dekker (1980).

KIER, L.B., and L.H. HALL: Molecular Connectivity in Chemistry and Drug Research. New York: Academic Press (1976).

KONEMANN, H.: Ecotoxicol. Environ. Safety 4, 415 (1980).

KOPPERMAN, H.L., R.M. Carlson, and R. Caple: Chem.-Biol. Interact. 9, 245 (1974).

SCHULTZ, T.W., and M. CAJINA-QUEZADA: Arch. Environ. Contam. Toxicol., in press (1982).

SCHULTZ, T.W., M. CAJINA-QUEZADA, and J.N. DUMONT: Arch. Environ. Contam. Toxicol. 9, 591 (1980).

SCHULTZ, T.W., L.M. KYTE, and J.N. DUMONT: Arch. Environ. Contam. Toxicol. 7, 457 (1978).

ZITKO, V., D.W. MCLEESE, W.G. CARSON, and H.E. WELCH: Bull. Environ. Toxicol. 16, 508 (1976).