

GRAPH III: A Computer Program for Calculating Molecular Connectivity Indices on Microcomputers

Aleksandar Sabljic* and Davor Horvatic

Institute Ruđer Bošković, P.O. Box 1016, 41001 Zagreb, Republic of Croatia

Received July 28, 1992

GRAPH III, a program for calculating molecular connectivity indices on microcomputers, is presented. In addition, the methods of calculation for all types of molecular connectivity indices are described. A possible physical interpretation of molecular connectivity indices is also discussed. Finally, a broad application of molecular connectivity indices in estimating various properties of organic chemicals is indicated, with particular emphasis on their application in evaluating environmental fate and effects of commercial chemicals.

1. INTRODUCTION

The quantitative structure-activity relationships (QSAR) approach is based on the assumption, which should be intuitive, that the structure of a molecule, geometric and electronic, must contain the features responsible for its physical, chemical, and biological properties and is based on the possibility to represent a molecule by numerical descriptor(s). The simplest way to numerically represent a molecule's structure is to assign to the structure a number or a set of numbers, termed indices. Indices generated by the application of chemical graph theory¹ are known as topological indices.¹⁻⁴ The most successful of all topological indices today is the molecular connectivity index or a system of molecular connectivity indices. Their numerous applications in various areas of physics, chemistry, biology, pharmacology (drug design), and environmental sciences outnumber all other existing topological indices, whose number is approaching 100. There are two major reasons for being so: first, these indices are based on sound chemical, structural, and mathematical grounds; and second, they were developed with the idea to parallel important physicochemical properties like boiling points, mobility on chromatographic column, enthalpies of formation, and total molecular surface areas.⁵

The following nomenclature is used to designate the molecular connectivity indices. χ is used to represent the index itself. Two superscripts and one subscript are used to specify a particular index. The left-side superscript (zero or positive integers) is used to designate the order of index. The right-side superscript (letter *v*) differentiates between valence- and nonvalence-type indices. The right-side subscript (P, C, PC, or CH) specifies the subclass of molecular connectivity index which may be path, cluster, path/cluster, and chain-type index. If no subscript is indicated, path-type index is assumed.

2. CALCULATION OF MOLECULAR CONNECTIVITY INDICES

2.1. Path-Type MCIs. The concept of molecular connectivity index (MCI) (originally called branching index) was introduced by Randić.⁵ The information used in the calculation of molecular connectivity indices are the number and type of atoms and bonds as well as the numbers of all and valence electrons. Those data are readily available for all chemicals, synthesized or hypothetical, from their structural formulas and periodic table of elements. All molecular connectivity indices are calculated only for the non-hydrogen part of the molecule. Each non-hydrogen atom is described

by its atomic δ value, which is equal to the number of adjacent non-hydrogen atoms. For example, the first-order ${}^1\chi$ molecular connectivity index is calculated from the atomic δ values by eq 1 where *i* and *j* correspond to the pairs of adjacent

$${}^1\chi = \sum (\delta_i \delta_j)^{-0.5} \quad (1)$$

non-hydrogen atoms and the summation is over all bonds between non-hydrogen atoms. The first-order molecular connectivity index was used very extensively in various QSAR studies, and it has been also found that this particular index correlates extremely well with the molecular surface area.⁶

In the valence approximation, non-hydrogen atoms are described by their atomic valence δ^v values which are calculated from their electron configuration by the following equation:

$$\delta^v = (Z^v - h) / (Z - Z^v - 1) \quad (2)$$

where Z^v is the number of valence electrons in the atom, Z is its atomic number, and h is the number of hydrogen atoms bound to the same atom. The most common valence δ^v values are listed in Table I. In addition, the calculation of the atomic δ^v values is illustrated in Figure 1, by a step-by-step procedure, for several atoms or groups containing an atom from the second or third row of periodic table of elements. By analogy with eq 1, the first-order valence molecular connectivity index (${}^1\chi^v$) is then calculated from the atomic δ^v values by eq 3.

$${}^1\chi^v = \sum (\delta_i^v \delta_j^v)^{-0.5} \quad (3)$$

The zero-order (${}^0\chi$) and second-order (${}^2\chi$) molecular connectivity indices are the closest members to the ${}^1\chi$ index described above. The ${}^0\chi$ and ${}^2\chi$ indices are calculated from the same input information (atomic δ values) by eqs 4 and 5, respectively,

$${}^0\chi = \sum (\delta_i)^{-0.5} \quad (4)$$

$${}^2\chi = \sum (\delta_i \delta_j \delta_k)^{-0.5} \quad (5)$$

where *i*, *j*, and *k* correspond to three consecutive non-hydrogen atoms, and the summations are over all non-hydrogen atoms and over all pairs of adjacent bonds between non-hydrogen atoms, respectively. Their valence analogs are defined identically as the first-order valence molecular connectivity index.

For molecular connectivity indices with order higher than 2, it is also necessary to specify the subclass of index. There

Table I. Formula for Calculating Valence δ^V Values and List of Calculated δ^V Values for Non-Hydrogen Atoms That Appear Most Frequently in Organic Chemicals

$\delta^V = \frac{Z^V - h}{Z - Z^V - 1}$					
Z^V - the number of valence electrons Z - the number of all electrons h - the number of hydrogen atoms					
atom	δ^V	atom	δ^V	atom	δ^V
$-\text{CH}_3$	1	$-\text{NH}_2$	3	$-\text{F}$	7
$-\text{CH}_2-$	2	>NH	4	$-\text{Cl}$	0.78
$=\text{CH}_2$	2	$=\text{NH}$		$-\text{Br}$	0.26
$\begin{array}{c} \\ -\text{CH} \\ \end{array}$	3	$-\text{N}-$	5	$-\text{I}$	0.16
$=\text{CH}-$	3	$=\text{N}-$	5	$-\text{SH}$	0.56
$\equiv\text{CH}$	3	$\equiv\text{N}$	5	$-\text{S}-$	0.67
$\begin{array}{c} \\ -\text{C}- \\ \end{array}$	4	$-\text{OH}$	5	$-\text{S}-\text{S}-$	0.89
$=\text{C}<$	4	$-\text{O}-$	6	$-\text{SO}-$	1.33
$=\text{C}=\text{}$	4	$=\text{O}$	6	$-\text{SO}_2-$	2.67
$\equiv\text{C}-$	4			$\begin{array}{c} \\ -\text{P}- \\ \end{array}$	0.56
				$\begin{array}{c} \\ -\text{P}-\text{O} \\ \end{array}$	2.22

are four subclasses of higher order indices: path, cluster, path/cluster, and chain. Those subclasses are defined by the type of structural subunit they are describing and over which the summation is to be taken when the respective indices are calculated. Naturally, the valence counterparts of all four subclasses of higher order indices can be easily defined by the analogy described above for the first-order valence molecular connectivity index.

The higher order path-type indices describe more and more (as the order of index goes higher) local structural properties and possibly long range interactions. For example, we have defined⁷ one analogous index, the valence subgraph index (χ^V_s), which was found to be very successful in describing chromatographic behavior of dichloroalkanes. The idea was to define the quantitative measure for the topological relation between the two chlorine atoms in substituted alkanes. Thus, a simple, local (specific) index was defined to describe the path (subgraph) between the two chlorine atoms. Each non-hydrogen atom on this path was described by its valence δ value, δ^V . The valence subgraph index was then calculated in accordance with the following expression:

$$\chi^V_s = (\delta_1^V \delta_2^V \delta_3^V \dots \delta_n^V) \quad (6)$$

where n is equal to the number of non-hydrogen atom, including both chlorine atoms, in a particular path (subgraph).

2.2. Cluster and Path/Cluster MCIs. The main characteristic of cluster-type indices is that all bonds are connected to the common, central atom (star-type structure). The third-

THE SECOND ROW ELEMENTS

Methyl group ($-\text{CH}_3$)	$\delta^V = (4 - 3)/(6 - 4 - 1) = 1$
Imino group ($-\text{NH}-$)	$\delta^V = (5 - 1)/(7 - 5 - 1) = 4$
Ether oxygen ($-\text{O}-$)	$\delta^V = (6 - 0)/(8 - 6 - 1) = 6$
Fluoro substituent ($-\text{F}$)	$\delta^V = (7 - 0)/(9 - 7 - 1) = 7$

THE THIRD ROW ELEMENTS

Phosphino group ($-\text{PH}_2$)	$\delta^V = (5 - 2)/(15 - 5 - 1) = 3/9 = 0.33$
Thioether ($-\text{S}-$)	$\delta^V = (6 - 0)/(16 - 6 - 1) = 6/9 = 0.67$
Chloro substituent ($-\text{Cl}$)	$\delta^V = (7 - 0)/(17 - 7 - 1) = 7/9 = 0.78$

Figure 1. Step-by-step procedure which illustrates how the valence δ^V values are calculated for several atoms or groups containing an atom from the second or third row of the periodic table of elements.

order cluster molecular connectivity index ($^3\chi_C$) is the first, simplest member of the cluster-type indices where three bonds are joined with the common central atom. The simplest chemicals structure it refers to is the non-hydrogen part of *tert*-butane. This index is then calculated by eq 7 where $i, j,$

$$^3\chi_C = \sum (\delta_i \delta_j \delta_k \delta_l)^{-0.5} \quad (7)$$

k , and l correspond to the individual non-hydrogen atoms that form the subgraph, and the summation is over all *tert*-butane-type subgraph in a molecule. For the cluster-type indices, orders higher than fourth do not have much chemical and structural sense for organic chemicals.

The fourth-order path/cluster molecular connectivity index ($^4\chi_{PC}$) is the first, simplest member of the path/cluster-type indices. It refers to subgraphs consisting of four adjacent bonds between non-hydrogen atoms, three of which are joined to the same non-hydrogen atom. Structurally (chemically) this subgraph corresponds to the non-hydrogen part of isopentane. This index is then calculated by eq 8 where $i, j,$

$$^4\chi_{PC} = \sum (\delta_i \delta_j \delta_k \delta_l \delta_m)^{-0.5} \quad (8)$$

k, l , and m correspond to the individual non-hydrogen atoms that form the subgraph, and the summation is over all isopentane-type subgraphs in a molecule. For the path/cluster-type indices, orders higher than sixth do not have much chemical and structural sense either. In addition, it becomes very difficult to understand what is the structural and physical meaning of higher order path/cluster-type indices.

The cluster and path/cluster indices describe mainly local structural properties, mainly the extend or degree of branching in a molecule. They are highly sensitive to changes in branching, and their value rapidly increases with the degree of branching. As such they may be very useful as steric descriptors. The second structural property described by the $^4\chi_{PC}$ index is the substitution pattern on the benzene ring. The value of $^4\chi_{PC}$ index increases sharply with the degree of substitution, while in the isomeric classes of substituted benzenes it increases with the proximity of substituents.

2.3. Chain-Type MCIs. The chain-type molecular connectivity indices describe the type of rings that are present in a molecule as well as the substitution patterns on those rings. Thus, the chain-type indices also describe more local-type properties. The lowest order for the chain-type index is third-order, and it goes up to the largest ring in any particular molecule. In this class of molecular connectivity indices the most interesting and commonly used are the sixth- (${}^6\chi_{CH}$) and seventh-order (${}^7\chi_{CH}$) chain-type indices since they are related to benzene rings. The ${}^7\chi_{CH}$ index corresponds to monosubstituted benzene rings.

2.4. MCIs Calculation by Microcomputers. All described molecular connectivity indices can be easily calculated by the GRAPH III program^{2,6,8} for microcomputers. There are two versions of the GRAPH III program: one is for the IBM PC (or PS/2) or compatible microcomputers, and the other is for all types of Apple Macintosh microcomputers. Hardware and software requirements for the PC version of GRAPH III are 256K of RAM memory, one double sided/double density disk drive, and a PC-DOS or MS-DOS operating system version 2.1 or higher. The Macintosh hardware requirements for this program are 512K of RAM memory and one 3.5-in. disk drive. GRAPH III can calculate molecular connectivity indices for molecules with 35 non-hydrogen atoms or less. It is user friendly and very simple to operate. At each stage, GRAPH III asks for the specific information it needs to continue with the calculation of connectivity indexes. It will ask for the type of data and explain the format in which data should be entered. The Apple Macintosh version of GRAPH III makes use of the tools provided by the Macintosh-like windows, menus, mouse, editors, etc. GRAPH III has been used for almost 10 years at 50 sites all around the world. Its current users are at universities, government agencies, and large chemical companies in the U.S., Canada, Belgium, France, Germany, Holland, Italy, U.K., Australia, etc.

A large database of molecular connectivity indices is also available.⁸ It contains molecular connectivity indices, up to sixth order, for more than 3000 chemicals, and the database is updated annually. This database contains many classes of chemicals as separate files. Each file contains at least 30 chemicals, and some contain several 100 chemicals. The database files are simple text files (ASCII files) which can easily be examined and modified by a file editor or a word processor and can be used by statistical packages, database systems, spreadsheets, and graphical programs. It is simple to transfer files between different computers.

3. APPLICATION OF MOLECULAR CONNECTIVITY INDICES

There are numerous successful applications of molecular connectivity indices in predicting and estimating various physical properties,^{1-5,7,9} chemical reactivity,^{1,3,10} biological activities,^{1-3,11-18} and environmental fate^{2,6,11,12,19-21} for many classes of organic chemicals.

Evaluation of humane exposure and risk assessment of commercial chemicals relies heavily on our ability to estimate correctly their environmental properties. Reliable quantitative models have been developed that can estimate Henry's Law constants,²² soil sorption coefficients,^{6,23} association with dissolved marine humic substances,²⁴ bioconcentration factors in aquatic organisms²⁵ and vegetation,²⁶ biodegradation rates,²⁷ and acute toxicity²⁸ for a variety of commercial chemicals, synthesized and hypothetical, within the experimental errors. Input data for those quantitative models are various molecular connectivity indices and other topological features.

Recently, models based on the molecular connectivity indices have also been developed for accurate prediction and evaluation of 1-octanol/water partition coefficients ($\log K_{ow}$) of highly hydrophobic chemicals: chlorinated benzenes,²⁹ chlorinated biphenyls (PCBs),²⁹ the polynuclear aromatic hydrocarbons (PAHs),³⁰ and their derivatives.³¹ Developed models enable the K_{ow} data to be predicted within a factor of 2 or better, the probability being above 95%. Thus, the developed models can be used as a reliable and inexpensive estimation method for predicting and evaluating K_{ow} data for several classes of important hydrophobic chemicals in line with the requirements of the OECD guidelines. This result demonstrates that topological quantities derived from molecular structure may confidently replace $\log K_{ow}$ data in many correlation (QSAR) models used to estimate various physical, chemical, biological, and environmental properties of hydrophobic chemicals. Several general and practical advantages of the molecular topology approach, like the speed by which topological quantities can be calculated, their nonempirical nature, and their ability to describe global as well as local and specific structural properties, will make the adjusted correlation (QSAR) models even more useful and widely applicable.

REFERENCES AND NOTES

- (1) Trinajstić, N. *Chemical Graph Theory*, 2nd ed.; CRC Press: Boca Raton, FL, 1992.
- (2) Sabljic, A.; Trinajstić, N. Quantitative Structure-Activity Relationships: The Role of Topological Indices. *Acta Pharm. Jugosl.* **1981**, *31*, 189-214.
- (3) Kier, L. B.; Hall, L. H. *Molecular Connectivity in Chemistry and Drug Research*; Academic Press: New York, 1976.
- (4) Seybold, P. G.; May, M.; Bagal, U. A. Molecular Structure-Property Relationship. *J. Chem. Educ.* **1987**, *64*, 575-581.
- (5) Randić, M. On Characterization of Molecular Branching. *J. Am. Chem. Soc.* **1975**, *97*, 6609-6615.
- (6) Sabljic, A. On the Prediction of Soil Sorption Coefficients of Organic Pollutants from Molecular Structure: Application of Molecular Topology Model. *Environ. Sci. Technol.* **1987**, *21*, 358-366.
- (7) Sabljic, A. Calculation of Retention Indices by Molecular Topology: Chlorinated Alkanes. *J. Chromatogr.* **1984**, *314*, 1-12.
- (8) GRAPH III program and molecular connectivity indices database are distributed by MacAda, Apple Center Ljubljana, Parmova 41, 61000 Ljubljana, Slovenia. Fax, (38-61) 328-887.
- (9) Needham, D. E.; Wei, I.-Ch.; Seybold, P. G. Molecular Modeling of the Physical Properties of the Alkanes. *J. Am. Chem. Soc.* **1988**, *110*, 4186-4194.
- (10) Hanson, M. P.; Rouvray, D. H. Novel Applications of Topological Indices. 2. Prediction of the Threshold Soot Index for Hydrocarbon Fuels. *J. Phys. Chem.* **1987**, *91*, 2981-2985.
- (11) Sabljic, A. In *Physical Property Prediction in Organic Chemistry*; Jochum, C., Hicks, M. G., Sunkel, J., Eds.; Springer Verlag: Heidelberg, 1988; p 336.
- (12) Sabljic, A. Chemical Topology and Ecotoxicology. *Sci. Total Environ.* **1991**, *109/110*, 197-220.
- (13) Sabljic, A.; Protić-Sabljić, M. Quantitative Structure-Activity Study on the Mechanism of Inhibition of Microsomal p-Hydroxylation of Aniline by Alcohols. Role of Steric Factors. *Mol. Pharmacol.* **1983**, *23*, 213-218.
- (14) Sabljic, A.; Šoškić, M. Inhibition of Hill Reaction by 2-Azido-s-triazine Derivatives: QSAR Study with Molecular Connectivity Indices. *Z. Naturforsch.* **1989**, *44C*, 255-261.
- (15) Rouvray, D. H. The Prediction of Biological Activity Using Molecular Connectivity Indices. *Acta Pharm. Jugosl.* **1986**, *36*, 239-252.
- (16) Rouvray, D. H. The Modeling of Chemical Phenomena Using Topological Indices. *J. Comput. Chem.* **1987**, *8*, 470-480.
- (17) Rouvray, D. H. The Limits of Applicability of Topological Indices. *J. Mol. Struct. (THEOCHEM)* **1989**, *185*, 187-201.
- (18) Balaban, A. T.; Motoc, I.; Bonchev, D.; Mekenyan, O. Topological Indices for Structure-Activity Correlations. *Top. Curr. Chem.* **1983**, *114*, 21-55.
- (19) Basak, S. C.; Monsrud, L. J.; Rosen, M. E.; Frane, C. M.; Magnuson, V. R. A Comparative Study of Lipophilicity and Topological Indices in Biological Correlation. *Acta Pharm. Jugosl.* **1986**, *36*, 81-95.
- (20) Boethling, R. S. Application of Molecular Topology to Quantitative Structure-Biodegradation Relationships. *Environ. Toxicol. Chem.* **1986**, *5*, 797-806.

- (21) Newsted, J. L.; Giesy, J. P. Predictive Models for Photoinduced Acute Toxicity of Polycyclic Aromatic Hydrocarbons to *Daphnia magna*. *Environ. Toxicol. Chem.* **1987**, *6*, 445–461.
- (22) Sabljic, A.; Güsten, H. Predicting Henry's Law Constants for the Polychlorinated Biphenyls. *Chemosphere* **1989**, *19*, 1503–1511.
- (23) Sabljic, A. Quantitative Modelling of Soil Sorption for Xenobiotic Chemicals. *Environ. Health Perspect.* **1989**, *83*, 179–190.
- (24) Sabljic, A.; Lara, R.; Ernst, W. Modelling Association of Highly Chlorinated PCBs With Marine Humic Substances. *Chemosphere* **1989**, *19*, 1665–1676.
- (25) Sabljic, A. The Prediction of Fish Bioconcentration Factors of Organic Pollutants from the Molecular Connectivity Model. *Z. Gesamte Hyg. Ihre Grenzgeb.* **1987**, *33*, 493–496.
- (26) Sabljic, A.; Güsten, H.; Schönherr, J.; Riederer, M. Modelling Plant Uptake of Airborne Organic Chemicals. Plant Cuticle/Water Partitioning and Molecular Connectivity. *Environ. Sci. Technol.* **1990**, *24*, 1321–1326.
- (27) Boethling, R. S.; Sabljic, A. Screening-Level Model for Aerobic Biodegradability Based on a Survey of Expert Knowledge. *Environ. Sci. Technol.* **1989**, *23*, 672–679.
- (28) Protić, M.; Sabljic, A. Quantitative Structure–Activity Relationships of Acute Toxicity of Commercial Chemicals on Fathead Minnows: Effect of Molecular Size. *Aquat. Toxicol.* **1989**, *14*, 47–64.
- (29) Sabljic, A.; Horvatić, D.; Güsten, H.; Hermens, J.; Opperhuizen, A. A Novel Efficient and Accurate Method to Calculate *n*-Octanol/Water Partition Coefficients of Highly Hydrophobic Chemicals. *Sci. Total Environ.* **1993**, in press.
- (30) Sabljic, A.; Horvatić, D.; Güsten, H. Modelling *n*-Octanol/Water Partition Coefficient by Molecular Topology: Polycyclic Aromatic Hydrocarbons and Their Alkyl Derivatives. *Chemosphere* **1991**, *23*, 199–213.
- (31) Güsten, H.; Sabljic, A. Predicting *n*-Octanol/Water Partition Coefficient by Molecular Topology: Heterocyclic and Substituted Polynuclear Aromatic Hydrocarbons. *Polycyclic Aromat. Compd.* **1993**, in press.