On the Calculation of the Molecular Descriptor χ'/χ

Dejan Plavšić,*,† Milan Šoškić,‡ and Nella Lerš†

The Rugjer Bošković Institute, P.O.B. 1016, HR-10001 Zagreb, The Republic of Croatia, and Department of Chemistry, Faculty of Agriculture, University of Zagreb, HR-10000 Zagreb, The Republic of Croatia

Received April 23, 1998

The molecular descriptor χ'/χ was recently introduced by Randić. The explicit formula for calculation of the χ'/χ index is derived. The formula provides an easy method to calculate the χ'/χ index for any connected graph. The advantages of the formula over the algorithm proposed by Randić et al. are pointed out.

1. INTRODUCTION

One of the fundamental concepts in chemistry is the concept of molecular structure.[1−7] The search for regularities in the way in which various molecular properties change, depending on molecular structure, is one of the central themes of contemporary chemistry and particularly the main subject of QSPR (quantitative structure−property relationship) as well as its extension to problems involving pharmacological and biological activity of molecules, QSAR (quantitative structure−activity relationship).[8−18]

Very interesting and efficient approach to QSPR/QSAR is the structure-explicit approach by means of graph theoretical methods.[9−15,18−29] The approach, in the simplest form, is founded on the assumption that a molecule can be represented by a (molecular) graph and characterized by graph invariants as well as that some of them correlate with molecular properties. The characterization of a molecule (or molecular graph) by a single number (topological index, TI) brings about a considerable loss of information concerning the molecular structure. Therefore, an intensive search for novel molecular indices, which would improve the graph-theoretical characterization of molecular structure has been undertaken.[30,31] Randić has recently put forward novel bond-additive molecular indices, TI′/TI, derived from known topological indices, TI.[32,33] The TI′/TI index of a molecular graph G is the sum of graphical bond orders[32,33] of all edges of G calculated by means of the invariant TI. Several of TI′/TI indices were used in modeling molecular properties of heptane and octane isomers.[32−35] The relationship between the P′/P index[34] and the Wiener number[36,37] as well as the relation between the descriptor Z′/Z[32,33] and the Hosoya Z index[37] were studied analytically.[38−40]

The molecular descriptor χ'/χ[32,33] is constructed from the connectivity index χ.[41] In the framework of single variable linear regression model the χ index is not very successful in modeling the van der Waals areas of heptanes ($r = 0.735$; $s = 9.1881$), whereas the molecular descriptor χ'/χ is rather successful ($r = 0.951$; $s = 3.971$).[32] In the modeling a number of molecular properties (van der Waals area was not one of them) of octane isomers the behavior of these two indices is reversed.[35] Such behaving of χ and χ'/χ indicates that these two indices encode to a large extent different parts of the structural information contained in a graph and that the χ'/χ index could be used either alone or in combination with one or more molecular indices in a multiparametric fit of molecular properties. Testing of the χ'/χ index on small sets, like heptanes and octanes, gives us just an indication and by no means the firm conclusion about the value and usability of the χ'/χ index in QSPR/QSAR. To find out the real role and value of χ'/χ in QSPR/QSAR, the index has to be tested on much larger and diverse sets of compounds. Evidently, in reaching this goal a great number of computations has to be performed. Hence, the first and necessary step is to find out an efficient algorithm for calculation of the χ'/χ index. It seems to us that the algorithm recently proposed by Randić et al. in this journal[33] is not the optimal one. The optimal algorithm for calculation of χ'/χ would be an explicit analytical expression "similar" to the expression for the connectivity index. Such a formula would give us an opportunity to calculate the χ'/χ index with a minimum number of steps.

In this article we will present the derivation of the explicit formula for calculation of the χ'/χ index of a molecular graph.

2. DEFINITIONS OF χ, χ_e/χ, AND χ'/χ

χ Index. The connectivity index χ of a molecular graph G, $\chi = \chi(G)$, is defined by[41]

$$\chi = \sum_{e_{ij}} (v_i v_j)^{-1/2} \tag{1}$$

where v_i and v_j are the valences of vertexes i and j, respectively, e_{ij} denotes the edge e connecting vertexes i and j, and the summation goes over all edges of G. Specially, χ of the trivial graph (graph of order one; e.g., hydrogen suppressed graph of methane) is by definition equal to 0. Since χ is bond additive it may be also viewed as a sum of bond weights given in terms of $(v_i v_j)^{-1/2}$.

Graphical Bond Order χ_e/χ. The graphical bond order χ_e/χ of an edge e of a molecular graph G is defined by[32,33]

$$\chi_e/\chi = \frac{\chi(G-e)}{\chi(G)} \tag{2}$$

† The Rugjer Bošković Institute.
‡ Faculty of Agriculture.

S0095-2338(98)00034-1 CCC: $15.00
Published on Web 07/09/1998

where $\chi(G)$ is the χ index of the graph G, and $\chi(G-e)$ is the χ index of the graph $G-e$ obtained from G by erasing the edge e. Since G is connected[19,22] its spanning subgraph $G-e$ is either connected or disconnected depending on the type of G and location of the edge e in G. To wit, $G-e$ is a connected graph if and only if G contains at least one ring and the edge e is one among the edges making up the ring-(s). A disconnected $G-e$ has two components, say G_1 and G_2, and the χ index associated with it is given by the expression[32]

$$\chi(G-e) = \chi(G_1) + \chi(G_2) \tag{3}$$

χ'/χ Index. The χ'/χ index of a molecular graph G is defined by the expression[32,33]

$$\chi'/\chi = \sum_e \chi_e/\chi = \frac{1}{\chi(G)} \sum_e \chi(G-e) \tag{4}$$

The summation goes over all edges of the graph G.

3. DERIVATION OF THE FORMULA

Let G be a molecular graph with N vertexes and R rings and let R equal the cyclomatic number[22] of G. Suppose i and j are two adjacent vertexes of G with valences v_i and v_j, respectively. The removal of the edge e_{ij} from G creates the spanning subgraph $G-e_{ij}$ with $N+R-2$ edges and changes the valences of i and j to $v_i - 1$ and $v_j - 1$, respectively. The valences of all other vertexes remain unchanged because they do not depend on the edge e_{ij}. Therefore, an edge of $G-e_{ij}$ being neither incident with the vertex i nor with the vertex j makes a contribution to $\chi(G-e_{ij})$ identical to the contribution of the corresponding edge of G to $\chi(G)$. In $G-e_{ij}$ there exists $N+R-v_i-v_j$ edges with this property. Their total contribution to $\chi(G-e_{ij})$ can be expressed by means of $\chi(G)$ and the contributions of the edge e_{ij} and all edges adjacent to e_{ij},

$$\chi(G) - (v_i v_j)^{-1/2} - \sum_{\substack{e_{ik} \\ k\neq j}} (v_i v_k)^{-1/2} - \sum_{\substack{e_{jl} \\ l\neq i}} (v_j v_l)^{-1/2}$$

where v_k and v_l denote the valences of vertexes k and l being adjacent to the vertexes i and j, respectively, and the summation in the third (fourth) term goes over all edges of G incident with the vertex i (j) except the edge e_{ij}. If either of the vertexes i and j is a terminal vertex in G, say the vertex i ($v_i = 1$), then there are no vertexes k and edges e_{ik} different than the vertex j and the edge e_{ij}, respectively. Hence, the third term in the expression is equal to 0. If both vertexes i and j are terminal vertexes in G, for instance G is the hydrogen suppressed graph of ethane, then the expression is equal to 0. The remaining $v_i + v_j - 2$ edges of $G-e_{ij}$ are incident either with the vertex i or with the vertex j. Their contributions to $\chi(G-e_{ij})$ are given by the expression

$$\sum_{\substack{e_{ik} \\ k\neq j}} [(v_i-1)v_k]^{-1/2} + \sum_{\substack{e_{jl} \\ l\neq i}} [(v_j-1)v_l]^{-1/2}$$

If either of the vertexes i and j is a terminal vertex in G, say the vertex j, then the erasure of the edge e_{ij} from G transforms the vertex j into the trivial graph and the second term in the expression is equal to 0 by definition. If e.g., G represents the carbon skeleton of ethane, then the expression is equal to 0.

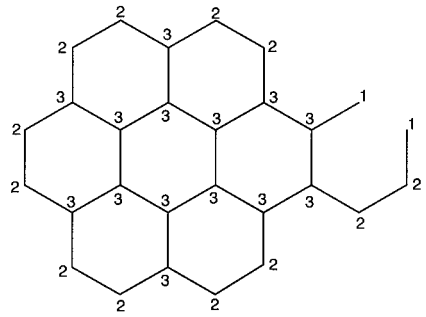

Figure 1. Calculation of the χ'/χ index of the hydrogen suppressed graph of 3-methyl-4-propylcoronene. Numbers at each site represent the corresponding graph-theoretical valences.

In general one can write

$$\chi(G-e_{ij}) = \chi(G) - (v_i v_j)^{-1/2} - \sum_{\substack{e_{ik} \\ k\neq j}} (v_i v_k)^{-1/2} - \sum_{\substack{e_{jl} \\ l\neq i}} (v_j v_l)^{-1/2} + \sum_{\substack{e_{ik} \\ k\neq j}} [(v_i-1)v_k]^{-1/2} + \sum_{\substack{e_{jl} \\ l\neq i}} [(v_j-1)v_l]^{-1/2} \tag{5}$$

Dividing $\chi(G-e_{ij})$ by $\chi(G)$ one gets graphical bond order, $\chi_{e_{ij}}/\chi$, of the edge e_{ij} of G,

$$\chi_{e_{ij}}/\chi = 1 + \frac{1}{\chi(G)}\{-(v_i v_j)^{-1/2} - \sum_{\substack{e_{ik} \\ k\neq j}} (v_i v_k)^{-1/2} - \sum_{\substack{e_{jl} \\ l\neq i}} (v_j v_l)^{-1/2} + \sum_{\substack{e_{ik} \\ k\neq j}} [(v_i-1)v_k]^{-1/2} + \sum_{\substack{e_{jl} \\ l\neq i}} [(v_j-1)v_l]^{-1/2}\} \tag{6}$$

This quantity can be interpreted as a measure of relative "importance" of e_{ij} in G. Equation 6 is just the first step toward the explicit formula for calculation of the χ'/χ index of a molecular graph. To get the explicit formula one has to add up graphical bond orders of all $N+R-1$ edges of G:

$$\chi'/\chi = \sum_{e_{ij}} \left\{ 1 + \frac{1}{\chi(G)}\{-(v_i v_j)^{-1/2} - \sum_{\substack{e_{ik} \\ k\neq j}} (v_i v_k)^{-1/2} - \sum_{\substack{e_{jl} \\ l\neq i}} (v_j v_l)^{-1/2} + \sum_{\substack{e_{ik} \\ k\neq j}} [(v_i-1)v_k]^{-1/2} + \sum_{\substack{e_{jl} \\ l\neq i}} [(v_j-1)v_l]^{-1/2}\} \right\} \tag{7}$$

By noting that

$$\sum_{e_{ij}} [\sum_{\substack{e_{ik} \\ k\neq j}} (v_i v_k)^{-1/2} + \sum_{\substack{e_{jl} \\ l\neq i}} (v_j v_l)^{-1/2}] = \sum_{e_{ij}} (v_i+v_j-2)(v_i v_j)^{-1/2} \tag{8}$$

and

$$\sum_{e_{ij}}\{\sum_{\substack{e_{ik}\\ k\neq j}}[(v_i-1)v_k]^{-1/2}+\sum_{\substack{e_{jl}\\ l\neq i}}[(v_j-1)v_l]^{-1/2}\}=$$

$$\sum_{e_{ij}}\left\{\frac{v_i-1}{[(v_i-1)v_j]^{1/2}}+\frac{v_j-1}{[(v_j-1)v_i]^{1/2}}\right\}=$$

$$\sum_{e_{ij}}\frac{v_i[(v_i-1)v_j]^{1/2}+v_j[(v_j-1)v_i]^{1/2}}{v_iv_j} \quad (9)$$

eq 7 can be transformed into a much simpler form:

$$\chi'/\chi = N+R-1+\frac{1}{\chi(G)}\sum_{e_{ij}}$$

$$\frac{v_i[(v_i-1)v_j]^{1/2}+v_j[(v_j-1)v_i]^{1/2}-(v_i+v_j-1)(v_iv_j)^{1/2}}{(v_iv_j)}$$

$$=\frac{1}{\chi(G)}\sum_{e_{ij}}$$

$$\frac{(N+R-v_i-v_j)(v_iv_j)^{1/2}+v_i[(v_i-1)v_j]^{1/2}+v_j[(v_j-1)v_i]^{1/2}}{v_iv_j} \quad (10)$$

Equation 10 is the sought after formula for calculation of the χ'/χ index of a molecular graph. The formula is well defined not only for a molecular graph with $N \geq 2$ vertexes but also for every connected graph with $N \geq 2$ vertexes.

The derived formula for calculation of the χ'/χ index makes it possible for us to calculate the χ'/χ index directly from G, whereas the algorithm proposed by Randić et al.[33] demands the knowledge of all the spanning subgraphs $G-e$ of G and calculation of the χ index for all of them. The application of the formula is illustrated for the hydrogen suppressed graph of 3-methyl-4-propylcoronene in Figure 1.

Paths. A special case of an acyclic molecular graph is the path graph, P_N. Equation 10 in case of P_N can be reduced to the form

$$\chi'/\chi = \frac{(N-4)(N-3)+4[2^{1/2}(N-3)+1-m]}{N-3+2\,(2^{1/2}+m)} \quad (11)$$

The parameter m takes the following values:

$$m=\begin{cases}3/2-2^{1/2} & \text{for } N=2\\ 0 & \text{for } N>2\end{cases} \quad (12)$$

Cycles. For a cycle C_N (regular graph of degree 2) eq 10 takes a rather simple form

$$\chi'/\chi = N-3+8^{1/2} \quad (13)$$

ACKNOWLEDGMENT

This work was supported by the Ministry of Science and Technology of the Republic of Croatia.

REFERENCES AND NOTES

(1) Turro, N. J. Geometrical and Topological Thinking in Organic Chemistry. *Angew. Chem., Int. Ed. Engl.* **1986**, *25*, 882−901.
(2) Seybold, P. G.; May, M.; Bagal, U. A. Molecular Structure − Property Relationships. *J. Chem. Educ.* **1987**, *64*, 575−581.
(3) Randić, M. The Nature of Chemical Structure. *J. Math. Chem.* **1990**, *4*, 157−189.
(4) Randić, M. Chemical Structure − What is She? *J. Chem. Educ.* **1992**, *69*, 713−718.
(5) Mekenyan, O.; Basak, S. C. Topological Indices and Chemical Reactivity. In *Graph Theoretical Approach to Chemical Reactivity;* Bonchev, D., Mekenyan, O., Eds.; Kluwer: Dordrecht, Netherlands, 1994; Chapter 8, pp 221−239.
(6) Randić, M. On Characterization of Chemical Structure. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 672−687.
(7) Milne, G. W. Mathematics as a Basis for Chemistry. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 639−644.
(8) Hansch, C.; Leo, A.; Hoekman, D. *Exploring QSAR.;* American Chemical Society: Washington, DC., 1995.
(9) Kier, L. B.; Hall, L. M. *Molecular Connectivity in Chemistry and Drug Research;* Academic Press: New York, 1976.
(10) Kier, L. B.; Hall, L. M. *Molecular Connectivity in Structure−Activity Analysis;* Research Studies Press Ltd.: Letchworth, England, 1986.
(11) Trinajstić, N.; Randić, M.; Klein, D. J. On the Quantitative Structure-Activity Relationship in Drug Research. *Acta Pharm. Jugosl.* **1986**, *36*, 267−279.
(12) Needham, D. E.; Wei, I.-C.; Seybold, P. G. Molecular Modeling of the Physical Properties of the Alkanes. *J. Am. Chem. Soc.* **1988**, *110*, 4186−4194.
(13) Basak, S. C.; Niemi, G. J. Predicting Properties of Molecules Using Graph Invariants. *J. Math. Chem.* **1991**, *7*, 243−272.
(14) Katritzky, A. R.; Gordeeva, E. Traditional Topological Indices vs Electronic, Geometrical, and Combined Molecular Descriptors in QSAR/QSPR Research. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 835−857.
(15) Katritzky, A. R.; Lobanov, V. S.; Karelson, M. QSPR: The Correlation and Quantitative Prediction of Chemical and Physical Properties from Structure. *Chem. Soc. Rev.* **1995**, *24*, 279−287.
(16) Karelson, M.; Lobanov, V. S.; Katritzky, A. R. Quantum-Chemical Descriptors in QSAR/QSPR Studies. *Chem. Rev.* **1996**, *96*, 1027−1043.
(17) Katritzky, A. R.; Karelson, M.; Lobanov, V. S. QSPR as a Means of Predicting and Understanding Chemical and Physical Properties in Terms of Structure. *Pure Appl. Chem.* **1997**, *69*, 245−248.
(18) *From Chemical Topology to Three-Dimensional Geometry;* Balaban, A. T., Ed.; Plenum Press: New York, 1997.
(19) Gutman, I.; Polansky, O. *Mathematical Concepts in Organic Chemistry;* Springer-Verlag: Berlin, 1986.
(20) Labanovsky, J. K.; Motoc, I.; Dammkoehler, R. A. The Physical Meaning of Topological Indices. *Comput. Chem.* **1991**, *15*, 47−53.
(21) *Chemical Graph Theory. Introduction and Fundamentals;* Bonchev, D., Rouvray, D. H., Eds.; Abacus Press: Gordon and Breach: New York, 1991.
(22) Trinajstić, N. *Chemical Graph Theory,* 2nd revised ed.; CRC Press: Boca Raton, FL, 1992.
(23) Harary, F. *Graph Theory;* Addison-Wesley: Reading, MA, 1972.
(24) Chartrand, G.: Oellermann, O. R. *Applied and Algorithmic Graph Theory;* McGraw-Hill: New York, 1993.
(25) Rouvray, D. H. Predicting Chemistry from Topology. *Sci. Am.* **1986**, *254*(3), 36−43.
(26) Rouvray, D. H. The Modeling of Chemical Phenomena Using Topological Indices. *J. Comput. Chem.* **1987**, *8*, 470−480.
(27) Rouvray, D. H. The Limits of Applicability of Topological Indices. *J. Mol. Struct. (THEOCHEM)* **1989**, *185*, 187−201.
(28) Randić, M.; Mihalić, Z.; Nikolić, S.; Trinajstić, N. Graph-Theoretical Correlations − Artifacts or Facts? *Croat. Chem. Acta* **1993**, *66*, 411−434.
(29) Rouvray, D. H. A Rationale for the Topological Approach to Chemistry. *J. Mol. Struct. (THEOCHEM)* **1995**, *336*, 101−114.
(30) Randić, M. In Search of Structural Invariants. *J. Math. Chem.* **1992**, *9*, 97−146.
(31) Randić, M.; Trinajstić, N. In Search for Graph Invariants of Chemical Interest. *J. Mol. Struct.* **1993**, *300*, 551−571.
(32) Randić, M. Search for Optimal Molecular Descriptors. *Croat. Chem. Acta* **1991**, *64*, 43−54.
(33) Randić, M.; Mihalić, Z.; Nikolić, S.; Trinajstić, N. Graphical Bond Orders: Novel Structural Descriptors. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 403−409.
(34) Randić, M. Generalized Molecular Descriptors. *J. Math. Chem.* **1991**, *7*, 155−168.
(35) Randić, M. Comparative Regression Analysis. Regressions Based on a Single Descriptor. *Croat. Chem. Acta* **1993**, *66*, 289−312.
(36) Wiener, H. Structural Determination of Paraffin Boiling Points. *J. Am. Chem. Soc.* **1947**, *69*, 17−20.

(37) Hosoya, H. Topological Index. A Newly Proposed Quantity Characterizing the Topological Nature of Structural Isomers of Saturated Hydrocarbons. *Bull. Chem. Soc. Jpn.* **1971**, *44*, 2332−2339.
(38) Plavšić, D.; Šoškić, M.; Landeka, I.; Trinajstić, N. On the Relation between the P′/P Index and the Wiener Number. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 1123−1126.
(39) Plavšić, D.; Šoškić, M.; Landeka, I.; Gutman, I.; Graovac, A. On the Relation between the Path Numbers ^{1}Z, ^{2}Z and the Hosoya Z Index. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 1118−1122.
(40) Gutman, I.; Plavšić, D.; Šoškić, M.; Landeka, I.; Graovac, A. On the Calculation of the Path Numbers ^{1}Z, ^{2}Z and the Hosoya Z Index. *Croat. Chem. Acta* **1997**, *70*, 941−954.
(41) Randić, M. On Characterization of Molecular Branching. *J. Am. Chem. Soc.* **1975**, *97*, 6609−6615.

CI980034E