Inter*net* Electronic Journal of Molecular Design

July 2002, Volume 1, Number 7, Pages 360–366

Editor: Ovidiu Ivanciuc

Special issue dedicated to Professor Milan Randić on the occasion of the 70th birthday
Part 3

Guest Editor: Mircea V. Diudea

The Structural Interpretation of the Randić Index

Ernesto Estrada

Faculty of Pharmacy, Department of Organic Chemistry, University of Santiago de Compostela, Santiago de Compostela 15706, SPAIN and Universidade Bandeirantes de São Paulo–UNIBAN, CEP: 02071–013, Brazil

Received: February 15, 2002; Revised: June 17, 2002; Accepted: July 3, 2002; Published: July 31, 2002

Citation of the article:
E. Estrada, The Structural Interpretation of the Randić Index, *Internet Electron. J. Mol. Des.* **2002**, *1*, 360–366, http://www.biochempress.com.

Copyright © 2002 BioChem Press

Inter*net* Electronic Journal of Molecular Design

BioChem Press
http://www.biochempress.com

The Structural Interpretation of the Randić Index[#]

Ernesto Estrada*

Faculty of Pharmacy, Department of Organic Chemistry, University of Santiago de Compostela, Santiago de Compostela 15706, SPAIN and Universidade Bandeirantes de São Paulo–UNIBAN, CEP: 02071–013, Brazil

Received: February 15, 2002; Revised: June 17, 2002; Accepted: July 3, 2002; Published: July 31, 2002

Internet Electron. J. Mol. Des. **2002**, *1* (7), 360–366

Abstract

Motivation. The Randić index is one of the most successful molecular descriptors in structure–property and structure–activity relationships studies. In spite of some attempts to interpret the physical meaning of this index it is still not clear what exactly this index measures in a molecular structure. The finding of the structural interpretation of this topological index will open the doors to novel applications as well as to a better understanding of the models developed with its use.

Results. The inverse squared–root of the vertex degree is identified here as a measure of the relative accessible perimeter of an atom from the outside. These perimeters, which have length units, are proposed to be measured in a new unit called the Randić (R). One Randić correspond to the perimeter of a carbon atom with degree one. On this basis, the bond contributions to the Randić index are relative areas of bond accessibility from the environment. Consequently, the Randić index is interpreted as the relative molecular area of external accessibility, that is the sum of relative accessibility bond areas in a molecule.

Conclusions. The results obtained here provide the Randić index with a clear structural interpretation, which is very close to the physical thinking commonly used in chemistry.

Keywords. Topological index; Randić index; molecular connectivity; molecular accessibility; atomic accessibility; accessibility area; molecular graph; structural descriptor.

1 INTRODUCTION

The history of topological indices (TI) can be divided into three parts. The first starting when Wiener [1] introduced his index to account for the changes in the boiling points of alkanes and continuing with the seminal work of Hosoya [2] who introduced the term topological indices for these molecular descriptors. In this stage of research very few TIs where introduced and almost all applications where constrained to physical properties of alkanes.

The second stage of research in TIs started when Milan Randić introduced the molecular branching index in 1975 [3]. After this seminal work, an explosion, in some cases indiscriminate, of these indices was observed in the literature. However, the most important contribution of this stage

[#] Dedicated to Professor Milan Randić on the occasion of the 70th birthday.

* Correspondence author; E–mail: ernesto.estrada@unilever.com.

is probably the great number of applications of TIs in several fields of chemistry [4,5]. These applications where head by the molecular connectivity indices developed and applied by Kier and Hall in the 70's and 80's [6,7]. These TIs are based on the original idea of Randić of molecular branching but extended to account for contributions coming from paths, path–clusters, clusters and chains of different lengths. These indices were also extended to account for heteroatom differentiation in molecular structures [6,7]. This contribution can be considered as a great step ahead in the research of these molecular descriptors. All these applications proved that TIs are very useful to describe the molecular structure in a way that permits the description and prediction of physicochemical, biological and toxicological properties of chemicals.

The third stage of research in TIs probably started in the 90's and it is characterized by a more selective introduction of TIs, the use of certain rules for the definition of such indices and the extension of applications to more complex situations in medicinal chemistry. At the beginning of this decade Randić introduced a series of rules for the definition of TIs [8]. It is true that even today some TIs are proposed without the observation of these rules. However, most of TIs introduced in these years are not trivially related to previous ones, account (or can be extended to account) for heteroatom differentiation, or to higher analogues, etc. A very good example of this kind of indices is the electrotopological state developed by Kier and Hall [9]. From the middle of the 90's the number and quality of applications of TIs in medicinal chemistry was dramatically increased (see [10–12] for recent reviews and examples). Among the most successful TIs in these applications we can mention the molecular connectivity indices [6,7], including the original Randić index [3], the kappa indices of Kier [13], electrotopological state indices [9], Balaban index [14] and the Wiener index [1]. This stage of research is still under progress, and includes the interpretation of TIs in term of physical or structural molecular features as well as the improvement of the existing TIs. In this sense Randić has also played an important and decisive role starting by the proper definition of the term TI [15] and continuing with the introduction of the variable molecular descriptors [16–19]. Concerning the Randić index some attempts of interpretation have been done in terms of quantum chemical concepts by Galvez [20], and on the basis of intermolecular encounters by Kier and Hall [21]. Some ways to improve the quality of this descriptor are the variable connectivity index introduced by Randić [16–18] as well as the long range connectivity index introduced by Estrada [22]. The most recent attempt was made by Estrada [23] in terms of generalized TIs, which include Randić index, together with several other TIs in the same graph invariant. Other approaches to generalized topological indices have been reported in the literature [24,25].

2 THE RANDIĆ INVARIANT

The branching index introduced by Randić is defined as the sum of certain bond contributions calculated from the vertex degree of the hydrogen suppressed molecular graphs. These bond

contributions, named C_{ij} by Kier and Hall [26], are calculated as:

$$C_{ij} = \left(\delta_i \delta_j\right)^{-0.5} \tag{1}$$

where δ_i is the degree of the vertex representing atom *i*, *i.e.*, the number of bonds incident to this atom. Accordingly, the Randić index is defined as [3]:

$$\chi = \sum C_{ij} = \sum \left(\delta_i \delta_j\right)^{-0.5} \tag{2}$$

where the summation is carried out over all the bonds in the molecule.

Kier and Hall have recently interpreted the molecular connectivity in terms of intermolecular accessibility starting from the interpretation of the bond contributions C_{ij} [21]. In consequence, they have stated that [21]: "the molecular connectivity index is the contribution of one molecule to the bimolecular interactions arising from encounters of bonds among two molecules".

We have been inspired in the concept of intermolecular accessibility to make a step forward in interpreting the Randić index in terms of molecular structural features. In doing so, we propose to define the Randić index in the following equivalent from:

$$\chi = \sum \left(\delta_i^{-0.5}\right)\left(\delta_j^{-0.5}\right) \tag{3}$$

Thus, we start our interpretation by finding a structural meaning of the terms $\delta_i^{-0.5}$ as the more basic structural element of the Randić invariant.

3 ATOMIC INTERMOLECULAR ACCESSIBILITY

Here we will represent molecules as molecular graphs, namely as hydrogen depleted graphs in which vertices represent atoms and edges represent covalent bonds. The vertices of a graph are generally represented as points in a plane. Here we will consider that vertices are covered by circles of arbitrary radius *r* in such a way that adjacent circles overlap each other. These circles have perimeters of lengths *L*. If we consider an isolated vertex (see Figure 1) for which the degree is zero it is trivial that the whole perimeter of the circle covering this vertex is in contact with the external environment. In other words, we can say that this circle is accessible in the total length of its perimeter. However, a circle covering a vertex of degree one is accessible from the environment from all points except from the place where both circles are overlapped, as illustrated in Figure 1. The necessity for the overlapping of the circles arises from the fact that in case the circles are not overlapped they are accessible from all the environment, that is the vertices are isolated (disconnected to each other). In this case, the accessibility of this circle from outside is the perimeter of the circle minus the section overlapped by the adjacent circle. This accessibility is reduced if we consider a vertex of degree two. In this case the accessibility to the circle from the exterior is the perimeter of the circle minus the sum of overlapping of both adjacent circles. If we

consider a vertex of infinite degree we see that the accessibility from outside is null as shown in Figure 1. We propose to define the atomic accessibility perimeter as the length of the arch that is accessible from outside the corresponding atom. Here we are not providing the circles covering the vertices of the graph with any physical meaning. In principle, they can be identified with different fields around the atom: van der Waals radius, electrostatic potential, etc.

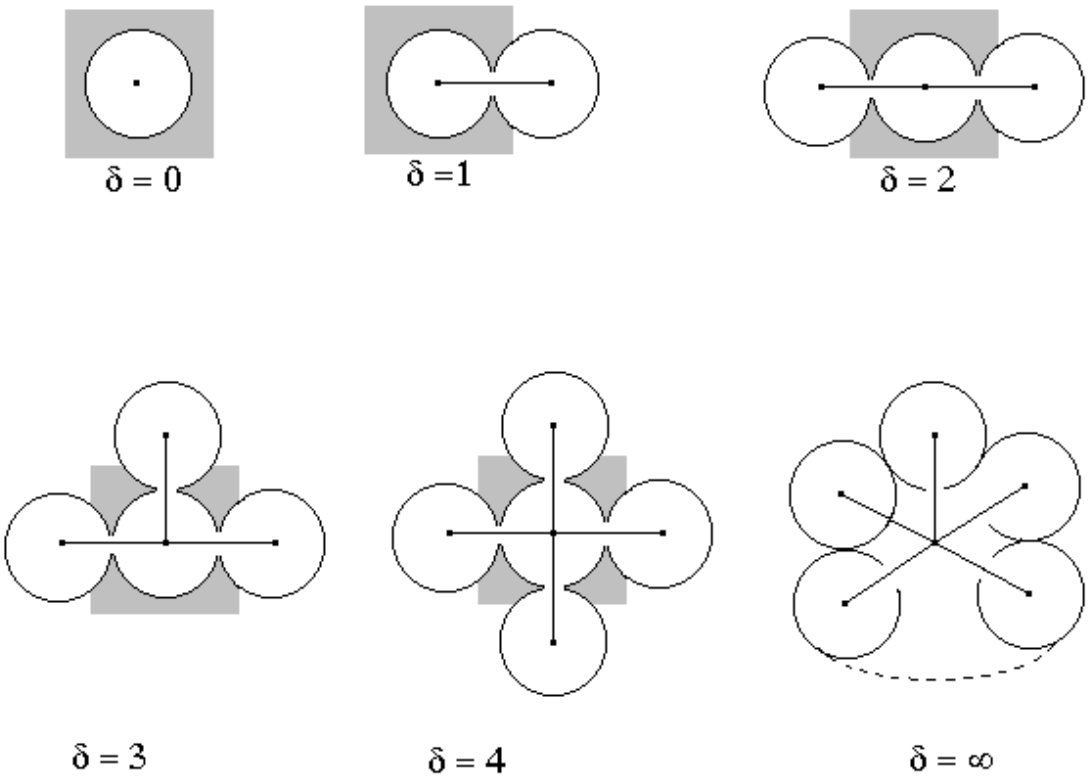

Figure 1. Representation of the external accessibility of an atom as a function of its valence (vertex degree).

It is straightforward to realize that the atomic accessibility perimeter of atom *i*, *Acc*(*i*), is proportional to the inverse of the squared root of the vertex degree (valence) of the corresponding atom (for vertices with degree different from zero, *i.e.*, vertices in connected graphs):

$$Acc(i) = \frac{\alpha}{\sqrt{\delta_i}} \tag{4}$$

If we consider the proportionality constant $\alpha = 1$ we will call relative atomic accessibility perimeter (RAP) to the values of $\delta_i^{-0.5}$.

4 THE RELATIVE AREA OF MOLECULAR ACCESSIBILITY

As we have seen in the previous section the values of $\delta_i^{-0.5}$, which are the basic elements in the definition of the Randić index, can be interpreted in terms of the accessible perimeter of an atom from the environment. In this sense, the values of RAP can be expressed in length units. We

propose to call this length unit the Randić (R). Thus, an atom having relative accessibility of 1 is said to have 1 R (one Randić) of accessibility.

If we take now the Randić invariant we observe that a summation of bonds terms is carried out for a molecule, in which the bond terms are: $C_{ij} = \left(\delta_i^{-0.5}\right)\left(\delta_j^{-0.5}\right)$, according to the nomenclature used by Kier and Hall. However, if we agree that the terms $\delta_i^{-0.5}$ are lengths of accessible perimeters we straightforwardly observe that the terms C_{ij} represent relative bond accessibility areas (RBA). The relative bond areas of the different bonds present in alkanes are given in Table 1. These areas are expressed in squared Randić (R^2).

Table 1. Values of the Relative Bond Areas for Different Bonds in Alkanes

Bond	RBA (R^2)	Bond	RBA (R^2)
CH_3–CH_3	1.0000	CH_2–CH	0.4082
CH_3–CH_2	0.7071	CH_2–C	0.3536
CH_3–CH	0.5774	CH–CH	0.3333
CH_3–C	0.5000	CH–C	0.2887
CH_2–CH_2	0.5000	C–C	0.2500

In closing, the Randić index, which is the sum of all relative accessibility areas in the molecule, is the relative molecular accessibility area RMA expressed in R^2. In such a way, we obtain the following values for the isomers of pentane: n–pentane, 2.4142; 2–methylbutane, 2.2700; 2,2–dimethylbutane; 2.0000. These areas represent the total area which are accessible from the environment surrounding the molecules. This explains why the Randić index has been so successful in modeling very diverse physical and biological properties. If we analyze for instance alkanes boiling temperatures we can see that the increase of the value of relative molecular accessible area, that is of Randić index, the number of intermolecular interactions increase producing an increment in the boiling point and explaining "physically" the correlation of this index with such experimental property. It also explains the success of this index in describing solubility, partition coefficients or the interaction of drugs with biological receptors.

5 CONCLUSIONS

A more general interpretation of molecular connectivity indices as well as other TIs is now under development by this author. However, due to the importance of the Randić index in structure–property–activity studies, we consider that the current work will open the doors of topological molecular descriptors to those that have been skeptic due to a "lack of physical interpretation" for these very useful molecular descriptors.

Acknowledgment

The author acknowledges FAPESP for financial support during his visit to Universidade Bandeirante (UNIBAN) de São Paulo, Brazil. Thanks are given to Prof. J. Quincoces and his team at UNIBAN for their kind hospitality during February 2002.

5 REFERENCES

[1] H. Wiener, Structural Determination of Paraffin Boiling Point, *J. Am. Chem. Soc.* **1947**, *69*, 17–20.

[2] H. Hosoya, Topological Index. A Newly Proposed Quantity Characterizing Topological Nature of Structural Isomers of Saturated Hydrocarbons, *Bull. Chem. Soc. Jpn.* **1971**, *44*, 2332–2339.

[3] M. Randić, On Characterization of Molecular Branching, *J. Am. Chem. Soc.*. **1975**, *97*, 6609–6615.

[4] A. Sabljić and N. Trinajstić, Quantitative Structure–Activity Relationships: The Role of Topological Indices, *Acta Pharm. Jugosl.* **1981**, *31*, 189–214.

[5] P. G. Seybold, M. May, and U. A. Bagal, Molecular structure–Property Relationships, *J. Chem. Educ.* **1987**, *64*, 575–582.

[6] L. B. Kier and L. H. Hall, *Molecular Connectivity in Chemistry and Drug Research*, Academic Press, New York, 1976.

[7] L. B. Kier and L. H. Hall, *Molecular Connectivity in Structure–Activity Analysis*, Research Studies Press, Letchworth, 1986.

[8] M. Randić, Generalized Molecular Descriptors, *J. Math. Chem.* **1991**, *7*, 155–168.

[9] L. B. Kier and L. H. Hall, *Molecular Structure Description: The Electrotopological State*, Academic Press, New York, 1999.

[10] A. T. Balaban, Topological and Stereochemical Molecular Descriptors for Databases Useful in QSAR, Similarity/Dissimilarity and Drug Design, *SAR QSAR Environ. Res.* **1998**, *8*, 1–21.

[11] E. Estrada, Recent Advances on the Role of Topological Indices in Drug Discovery Research, *Curr. Med. Chem.* **2001**, *8*, 1573–1588.

[12] I. Rios–Santamarina, R. García–Domenech, J. Cortijo, P. Santamaria, E. J. Morcillo, and J. Gálvez, Natural Compounds with Bronchodilator Activity Selected by Molecular Topology, *Internet Electron. J. Mol. Des.* **2002**, *1*, 70–79, http://www.biochempress.com.

[13] L. B. Kier, A Shape Index from Molecular Graphs, *Quant. Struct.–Act. Relat.* **1985**, *4*, 109–116.

[14] A. T. Balaban, Highly discriminating distance–based topological index, *Chem. Phys. Lett.* **1982**, *89*, 399–404.

[15] M. Randić, Topological Indices; in: *The Encyclopedia of Computational Chemistry*, Eds. P. v. R. Schleyer, N. L. Allinger, T. Clark, J. Gasteiger, P. A. Kollman, H. F. Schaefer III, and P. R. Schreiner, John Wiley & Sons, Chichester, 1998, Vol. 5, pp. 3018–3032.

[16] M. Randić, D. Mills, and S. C. Basak, On Use of Variable Connectivity Index for Characterization of Amino Acids, *Int. J. Quantum Chem.* **2000**, *80*, 1199–1209.

[17] M. Randić, D. Plavsić, and N. Ler, Variable Connectivity Index for Cycle–Containing Structures, *J. Chem.. Inf. Comput. Sci.* **2001**, *41*, 657–662.

[18] M. Randić and S. C. Basak, On Use of Variable Connectivity Index $^{1}\chi^{f}$ in QSAR: Toxicity of Aliphatic Ethers, *J. Chem.. Inf. Comput. Sci.* **2001**, *41*, 614–618.

[19] M. Randić and M. Pompe, The Variable Molecular Descriptors Based on Distance Related Matrices, *J. Chem.. Inf. Comput. Sci.* **2001**, *41*, 575–581.

[20] J. Galvez, On the Topological interpretation of Electronic and Vibrational Molecular Energies, *J. Mol. Struct. (Theochem)* **1998**, *429*, 255–264.

[21] L. B. Kier and L. H. Hall, Intermolecular Accessibility: The Meaning of Molecular Connectivity, *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 792–795.

[22] E. Estrada, Connectivity Polynomial and Long–Range Contributions in the Molecular Connectivity Model, *Chem. Phys. Lett.* **1999**, *312*, 556–560.

[23] E. Estrada, Generalization of Topological Indices, *Chem. Phys. Lett.* **2000**, *336*, 248–252.

[24] O. Ivanciuc, Design of Topological Indices. Part 14. Distance–Valency Matrices and Structural Descriptors for Vertex– and Edge–Weighted Molecular Graphs, *Rev. Roum. Chim.* **2000**, 45, 587–596.

[25] O. Ivanciuc, Design of Topological Indices. Part 16. Matrix Power operators for Molecular Graphs, *Rev. Roum. Chim.* **2000**, 45, 1027–1044.

[26] L. H. Hall, L. B. Kier, and W. J. Murray. Molecular Connectivity II: Relationship to Water Solubility and Boiling Point, *J. Pharm. Sci.* **1975**, *64*, 1974–1979.

Biographies

Ernesto Estrada is computational chemist at Unilever R&D, Colworth in UK. After obtaining a Ph. D. degree in organic chemistry and mathematical/computational modeling by using graph theoretical tools from the Central University of Las Villas, Cuba, Dr. Estrada undertook postdoctoral research with Professor Galvez at the University of Valencia, Spain and Prof. Avnir at the Hebrew University of Jerusalem, Israel. After that Dr. Estrada moved to Spain as Researcher at the University of Santiago de Compostela from 2000 to 2002. Dr. Estrada main research interests include the development of novel molecular descriptors for using in the development of QSPR, QSAR, drug design, early ADME and toxicity prediction. More recently, he is also interested in the development of macromolecular descriptors for the study of proteins folding degree, protein classification and protein-drug interactions.