J. Phys. Chem. A **2002,** *106,* 9085−9091

Physicochemical Interpretation of Molecular Connectivity Indices

Ernesto Estrada[†]

Faculty of Organic Chemistry, Department of Organic Chemistry, University of Santiago de Compostela, 15706, Santiago de Compostela, Spain

Received: June 5, 2002; In Final Form: July 26, 2002

Molecular connectivity indices are identified as components of the molecular accessibility. The first- and second-order connectivity indices represent molecular accessibility areas and volumes, respectively, whereas higher order indices represent magnitudes in higher dimensional spaces. In identifying accessibility perimeters, we recognized the atom degrees as a measure of the accessibility perimeter of the corresponding atom. The Randić and connectivity indices are identified as the two components of the molecular accessibility area. The accessibility perimeter is computed here from the van der Waals and covalent radii of the atoms and the overlapping angle between the van der Waals circumferences of bonded atoms. The description of the accessibility area in terms of the first-order Randić and connectivity indices accounts for the success of these descriptors in correlating different physicochemical and biological properties because they are a measure of the extension of intermolecular interactions. A theoretical justification for the selection of the exponent in the Randić invariant is provided by the relation between the valence degree and the accessibility perimeters calculated in this work.

1. Introduction

The history of molecular connectivity indices started when Milan Randić introduced the "molecular branching" index in 1975.[1] In this seminal work Randić introduced a graph-theoretical invariant (the Randić invariant) based on the sum of weighted contributions from all bonds in the molecule. These bond weights were determined as the product of vertex degrees of the atoms defining the bond raised to the exponent −0.5. Subsequently, in 1976 Kier, Hall, Murray, and Randić extended the branching index to a series of molecular descriptors accounting for contributions coming from paths, clusters, path-clusters, and chains.[2] The completion of the molecular connectivity descriptors occurs with the introduction of heteroatom differentiation by Kier and Hall in 1976.[3]

The molecular connectivity indices have been widely used in physical chemistry[4−14] as well as for predicting pharmacological and toxicological properties of organic, inorganic and organometallic compounds.[15−18] This large series of applications of molecular connectivity indices reaffirm them as one of the most useful set of molecular descriptors in structure−property−activity relationship studies (QSPR and QSAR). Despite this huge number of applications, a "physical interpretation" of the connectivity indices has remained elusive. However, some attempted interpretation in terms of quantum chemical concepts have been done by Zefirov[19] and Gálvez,[20] and on the basis of intermolecular encounters by Kier and Hall.[21,22] On the other hand, successful attempts to improve the quality of these descriptors have been proposed. They include changing the exponent in the Randić invariant,[23−27] the variable connectivity index introduced by Randić,[28−30] and the long-range connectivity index introduced by Estrada.[13] The most recent attempt for improving and interpreting topological indices (TIs) was introduced by Estrada[31] as generalized TIs, which include the Randić index, together with several other TIs in the same graph invariant.

This apparent divorce between the success in describing the molecular structure and the elusive interpretation of these indices has been the source of certain criticism for using molecular connectivity in QSPR/QSAR studies. The current work represents an effort to solve the "mystery" of why molecular connectivity is so successful in describing molecular structure and consequently in predicting properties that depends on it.

2. Connectivity Index

The basis for the connectivity indices is the branching index introduced by Randić,[1] which is defined as the sum of certain bond contributions calculated from the vertex degrees, δ_i, of the hydrogen suppressed molecular graph.

The values of δ_i represent the number of skeletal neighbors of the atom i and can be obtained as the number of electrons in sigma orbitals, σ_i, minus the number of hydrogen atoms h_i bonded to atom i. To account for heteroatom differentiation Kier and Hall introduced the valence degree of an atom, δ_i^V.[3] For second row atoms in covalent molecules, the valence degree for atom i is given as

$$\delta_i^V = Z_i^V - h_i = \sigma_i + \pi_i + n_i - h_i \qquad (1)$$

where Z_i^V is the number of valence electrons, π_i is the number of electrons in π-orbitals, and n_i is the number of electrons in lone pair orbitals. This is a simple counting algorithm in which the number of electrons in different orbitals are taken into account. Later we will give a graph-theoretical representation for the case of lone-pair electrons in the development of the current approach. To consider the screening effects produced by electrons in inner shells of atoms beyond the second row of

---

[†] Current address: SEAC, Unilever R&D, Colworth House, Sharnbrook, Bedford MK44 1LQ, U.K. E-mail: estrada66@yahoo.com. Fax: +44 1234 264722.

10.1021/jp026238m CCC: $22.00

the periodic table, Kier and Hall proposed to calculate the valence degree of any atom as follows:[15]

$$\delta_i^{\mathrm{V}} = (Z_i^{\mathrm{V}} - h_i)/(Z_i - Z_i^{\mathrm{V}} - 1) \qquad (2)$$

The bond contributions to the connectivity indices, named C_{ij} by Kier and Hall,[32] are calculated as

$$C_{ij} = (\delta_i \delta_j)^{-0.5} \quad \text{and} \quad C_{ij}^{\mathrm{V}} = (\delta_i^{\mathrm{V}} \delta_j^{\mathrm{V}})^{-0.5} \qquad (3)$$

These contributions can be considered as the simplest ones coming from different subgraphs in the molecular graphs. That is, they are contributions coming from first-order subgraphs: graph edges or bonds. Thus, the Randić index, ${}^1\chi$, and the valence connectivity index, ${}^1\chi^{\mathrm{V}}$, are first-order indices, which are defined as

$$ {}^1\chi = \sum C_{ij} = \sum_l (\delta_i \delta_j)_l^{-0.5} \quad \text{and} \quad {}^1\chi^{\mathrm{V}} = \sum C_{ij}^{\mathrm{V}} = \sum_l (\delta_i^{\mathrm{V}} \delta_j^{\mathrm{V}})_l^{-0.5} \qquad (4)$$

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

3. Extended Connectivity Indices

Molecular connectivity indices have been generalized to a series of "higher order" or extended connectivity indices, ${}^m\chi_{\mathrm{t}}$ and ${}^m\chi_{\mathrm{t}}^{\mathrm{V}}$.[9] The exponent m is known as the order of the connectivity index and represents the number of bonds in a subgraph of type t. In this way the indices ${}^1\chi$ and ${}^1\chi^{\mathrm{V}}$ are the simplest examples of this set of descriptors, which are calculated for paths of length 1, i.e., bonds. For the order $m > 2$ there are several types of subgraphs: paths (t = p), clusters (t = C), path-clusters (t = pC), and chains (t = Ch), in the terminology of Kier and Hall. In the case of the third order, there is no path-cluster type.

In this case the indices are calculated by defining the terms ${}^mC_k$ as the product of reciprocal square roots of the vertex or valence degrees in the corresponding subgraph:

$${}^mC_k = \prod_i (\delta_i)_k^{-0.5} \quad \text{and} \quad {}^mC_k^{\mathrm{V}} = \prod_i (\delta_i^{\mathrm{V}})_k^{-0.5} \qquad (5)$$

The index i runs over the $m + 1$ values of δ in the subgraph having m bonds, except for the chain type in which m equals the number of atoms in the ring.

The extended connectivity indices are defined as

$${}^m\chi_{\mathrm{t}} = \sum_k {}^mC_k \quad \text{and} \quad {}^m\chi_{\mathrm{t}}^{\mathrm{V}} = \sum_k {}^mC_k^{\mathrm{V}} \qquad (6)$$

The summation is carried out over the total number of subgraphs of order m and type t in the graph.

4. Atomic Intermolecular Accessibility

Kier and Hall have recently interpreted the molecular connectivity in terms of intermolecular accessibility starting from the interpretation of the bond contributions C_{ij}.[21,22] Thus, they have concluded 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 interpret the connectivity indices in terms of molecular structural features. In doing so, we propose to rewrite the first-order connectivity indices in the following equivalent form:

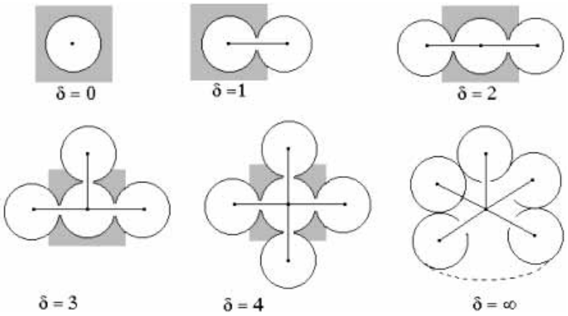

Figure 1. Representation of the external accessibility of an atom as a function of its vertex degree.

$${}^1\chi = \sum_l [(\delta_i^{-0.5})(\delta_j^{-0.5})]_l \quad \text{and} \quad {}^1\chi^{\mathrm{V}} = \sum_l [(\delta_i^{\mathrm{V}})^{-0.5} (\delta_j^{\mathrm{V}})^{-0.5}]_l \qquad (7)$$

For the sake of simplicity we start our interpretation by finding the structural meaning of the terms $\delta_i^{-0.5}$ as the most basic structural element of the connectivity indices. Next we extend our approach to molecules containing heteroatoms and to higher order indices.

Here we represent molecules as molecular graphs, i.e., 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 the vertices are covered by circumferences of radius r in such a way that adjacent circumferences overlap each other. These circumferences 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 circumference covering this vertex is in contact with the external environment. In other words, we can say that this circumference is accessible in the total length of its perimeter. However, a circumference covering a vertex of degree 1 is accessible from the environment from all points except from the place where both circumferences are overlapped, as illustrated in Figure 1. The necessity for the overlapping of the circumferences arises from the fact that in case the circumferences do not overlap they are accessible from the entire environment; that is, the vertices are isolated (disconnected to each other). When circumferences overlap, the accessibility to one of these circumferences from outside is the perimeter of the circumference minus the section overlapped by the adjacent circumference. This accessibility is reduced if we consider a vertex of degree 2. In this case the accessibility to the circumference from the exterior is the perimeter of the circumference minus the sum of overlaps of both adjacent circumferences. 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 call the length of the arc that is accessible from outside the corresponding atom the *atomic accessibility perimeter*.

It is straightforward to realize that the atomic accessibility perimeter of atom i (Acc(i)) is proportional to the inverse of the vertex degree of the corresponding atom ($1/\delta$) (for vertices with degree different from zero, i.e., vertices in connected graphs). This proportionality in maintained by using the inverse of the square root of the vertex degree, which is the term used in the connectivity indices (see further for a theoretical justification of this selection):

$$\mathrm{Acc}(i) = \frac{\alpha}{\sqrt{\delta_i}} \tag{8}$$

where α is a proportionality constant.

On the other hand, we can consider that this accessibility is proportional to the portion of the perimeter that is exposed to the environment, that is, the difference between the perimeter of the circumference surrounding the atom, L, minus the arc, I, which is interpenetrated with the circumference of the neighbor atom:

$$\mathrm{Acc}(i) = \beta(L - I) \tag{9}$$

where β is another proportionality constant, which is the atomic accessibility when the difference between the perimeter of the circumference and the overlap arc is exactly 1 pm. Here we will always consider the value of $\beta = 1$ in all the calculations.

The combination of eqs 8 and 9 means that the inverse of the square root of the vertex degree can be understood as a measure of the perimeter of an atom that is accessible from outside when no heteroatoms or multiple bonds are taken into account. That is, it can be considered as a component of the *relative atomic accessibility perimeter* (RAP). Here relative means that it is related to the carbon atom with degree 1, i.e., $\delta^{-0.5} = 1$.

5. Relative Area of Molecular Accessibility

As we see 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 bond terms is carried out for a molecule, in which the bond terms are $C_{ij} = (\delta_i{}^{-0.5})(\delta_j{}^{-0.5})$, according to the nomenclature used by Kier and Hall.[32] 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). In Figure 2 we represent in the form of rectangles the different bonds present in alkanes whose areas correspond to the values of the RBAs for such bonds. These areas are expressed in squared Randić (R^2).

In closing, the Randić index, which is the sum of all relative bond accessibility areas in the molecule, is the *relative molecular accessibility area* when neither heteroatoms nor multiple bonds are considered. In more appropriate terms we can speak about the Randić index as a "contribution" to the relative molecular accessibility area. In such a way, we represent these areas for the isomers of pentane in Figure 3. These areas, given in R^2, represent the total areas that 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, the boiling points of alkanes, we can see that with the increment of the value of the relative molecular accessible area, that is of the Randić index, the number of intermolecular interactions increases producing an increment in the boiling point and explaining "physically" the correlation of this index with such an experimental property. It also explains the success of this index in describing solubility, partition coefficients, or the interaction of drugs with biological receptors.

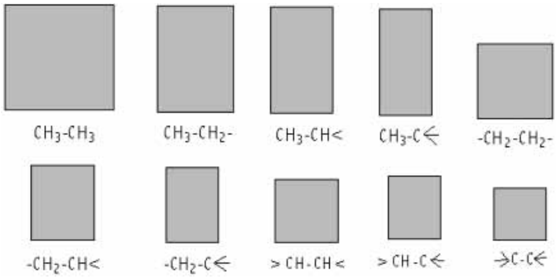

Figure 2. Relative bond areas of external accessibility for the different bond types in alkanes.

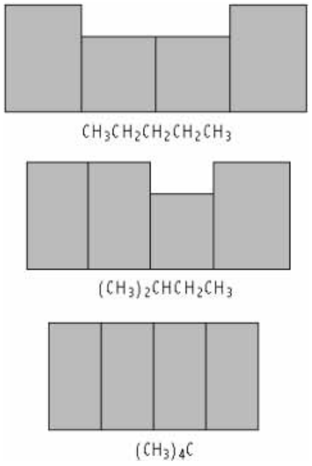

Figure 3. Representation of the relative molecular accessibility areas for pentane isomers.

6. Molecules Containing Heteroatoms and Multiple Bonds

To understand the nature of connectivity indices in molecules with heteroatoms, we have to provide the circumferences around the vertices in molecular graphs with a physical meaning. Continuing with the idea of atomic accessibility, we have to consider different radii for the circumferences surrounding different heteroatoms. One parameter that reflects very well these differences is the van der Waals radius. It is established from the contact distances between nonbonding atoms in touching the corresponding atom. We first will analyze the variations of this parameter in one row of the periodic table. That is, we will consider the differences among C, N, O, and F. According to Bondi,[33] the van der Waals radii (r_{vdW}) for these atoms are, in pm, 170, 155, 152, and 147.[34] Thus, the van der Waals radii decrease with an increase in the number of electrons in the valence shell. To represent the electrons in the lone pairs in heteroatoms using a graph-theoretical approach, we will consider graphs having loops. A loop is an edge that starts and ends at the same vertex. Thus, a loop is doubly incident to the corresponding vertex, which means that it increases the vertex degree by 2. A graphical representation of the previously considered atoms is given in Figure 4. These types of graphs are known in graph theory as pseudographs.

If we take the van der Waals radii as the radii of the circumferences surrounding the vertices, we have identified the perimeter of these circumferences, L, and the length of overlap arc, I, as

$$L = 2\pi \mathrm{r}_{\mathrm{vdW}} \qquad \text{and} \qquad I = \theta_{\mathrm{vdW}} \tag{10}$$

where θ is the angle of overlap between two adjacent circumferences (see further). Thus, the atomic accessibility given by eq 9 is now as follows:

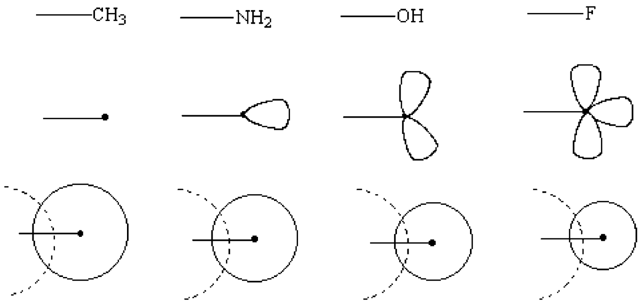

Figure 4. Schematic representations of some first-row atoms illustrating graph-theoretical loops and van der Waals circumferences.

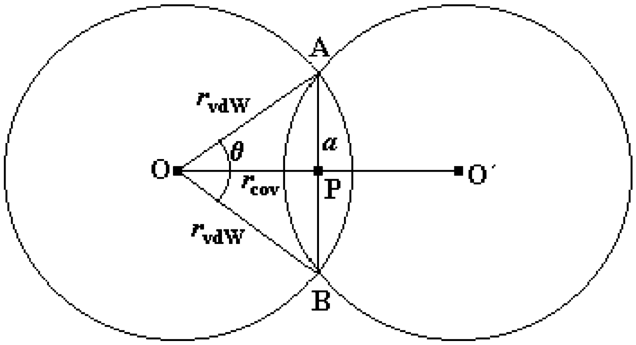

Figure 5. Interpenetration of the van der Waals circumferences of two neighbor atoms.

$$\mathrm{Acc}(i) = (L - I) = r_{\mathrm{vdW}}(2\pi - \theta) \tag{11}$$

In Figure 5 we illustrate the situation in which a circumference with van der Waals radius r_{vdW} and center at O is overlapped with another circumference with center at O’. The angle θ is formed by the points AOB. The distance OO’ is the corresponding bond distance. Thus, the distance OP is the covalent radius of the corresponding atom O.

The angle θ is then obtained as

$$\theta = 2\cos^{-1}[(r_{\mathrm{vdW}}{}^{2} + r_{\mathrm{cov}}{}^{2})/2r_{\mathrm{vdW}}{}^{2}] \tag{12}$$

When atoms in different rows are analyzed, we have to consider the screening effect produced by the electrons in inner shells for the valence shell electrons. This effect is considered in the valence connectivity index by dividing by the term $Z_i - Z_i^{\mathrm{V}} - 1$ in eq 2. This “correction” produces an increment in the value of $(\delta_i^{\mathrm{V}})^{-0.5}$ as we move down in a column of the periodic table. The same variation is observed for the van der Waals radii.

When multiple bonds are analyzed, a similar situation occurs. In this case the variation in the atomic accessibility is accounted for by the change in the covalent radius of atoms supporting multiple bonds. For instance, the covalent radius for the carbon with $\mathrm{sp_3}$ hybridization, e.g., in $\mathrm{CH_3}$, is 77 pm, and the same for a $\mathrm{sp_2}$ carbon, e.g., $=\mathrm{CH_2}$, is 60 pm. This change in the covalent radius is reflected in the values of the overlapping angle and consequently in the values of the atomic accessibility. In the upper part of Table 1 we give the values of the atomic accessibility for different atoms having only one neighbor bonded to them (same connectivity). As can be seen, the values of the atomic accessibility for atoms with the same number of neighbors, $\mathrm{Acc}(i) = (L - I)$, change in the same direction as the values of $(\delta_i^{\mathrm{V}})^{-0.5}$. In Figure 6 we plot the values of the atomic accessibility versus the values of (δ_i^{V}), where we can see the proportionality between both parameters. It can also be observed that $\mathrm{Acc}(i) = (L - I)$ depends on a negative potency of the (δ_i^{V}) values, as can be analyzed later.

TABLE 1: Values of van der Waals and Covalent Radii, Interpenetration Angle between van der Waals Circumferences, Accessibility Perimeter ($L - I$) and Valence Degrees for Different Atoms

atom	r_{vdW} (pm)	r_{cov} (pm)	θ (radians)	$L - I$ (pm)	(δ^{v})
$-\mathrm{CH_3}$	170	77	1.8481	753.9	1.000
$=\mathrm{CH_2}$	170	70	1.9099	743.5	2.000
$\equiv\mathrm{CH}$	170	60	1.9473	737.1	3.000
$-\mathrm{NH_2}$	155	75	1.8115	693.1	3.000
$=\mathrm{NH}$	155	62	1.9041	678.8	4.000
$\equiv\mathrm{N}$	155	55	1.9457	672.3	5.000
$-\mathrm{OH}$	152	73	1.8160	679.0	5.000
$=\mathrm{O}$	152	60	1.9093	664.8	6.000
$-\mathrm{SH}$	180	102	1.6986	825.2	0.556
$-\mathrm{F}$	147	71	1.8127	657.2	7
$-\mathrm{Cl}$	175	99	1.6999	802.1	0.778
$-\mathrm{Br}$	185	114	1.6190	862.9	0.259
$-\mathrm{I}$	198	133	1.5177	943.6	0.156
$-\mathrm{CH_2}-$	170	77	1.8481	439.8	2
$=\mathrm{CH}-$	170	73.5[a]	1.8790	429.3	3
$\equiv\mathrm{C}-$	170	68.5[a]	1.8977	422.9	4
$=\mathrm{C}=$	170	70[a]	1.9099	418.8	4
$-\mathrm{NH}-$	155	75	1.8115	412.3	4
$=\mathrm{N}-$	155	68.5[a]	1.8578	397.9	5
$-\mathrm{O}-$	152	73	1.8160	403.0	6
$-\mathrm{S}-$	180	102	1.6986	519.5	0.667
$-\mathrm{CH}<$	170	77	1.8481	125.6	3
$=\mathrm{C}<$	170	74.7[a]	1.8790	115.2	4
$-\mathrm{N}<$	155	75	1.8115	131.6	5

[a] Averaged values of the covalent radii for the different hybridizations in the atom.

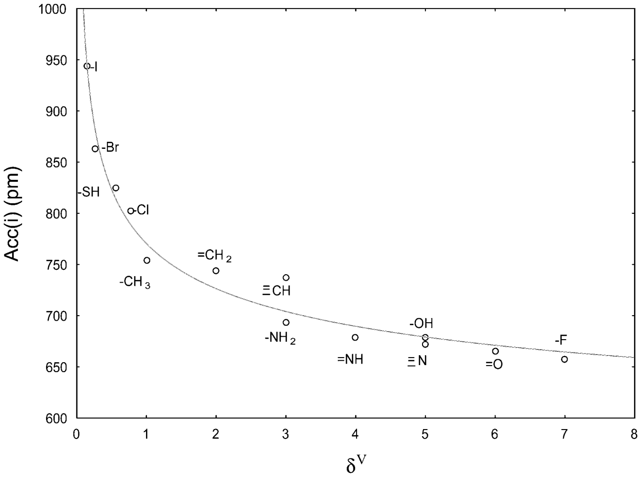

Figure 6. Relation between the atomic accessibility perimeter and the valence degree of different heteroatoms bound to one neighbor atom.

Consequently, for atomic groups having the same number of neighbors (the same connectivity in terms of simple values of δ_i) the atomic accessibility is proportional to the values of (δ_i^{V}) raised to a negative power (here we will use the value of -0.5 as in the connectivity index and later we will analyze this value in more detail):

$$\mathrm{Acc}(i) = \alpha'(\delta_i^{\mathrm{V}})^{-0.5} \tag{13}$$

A new question arises when atoms or atomic groups with different numbers of neighbors are considered together in the current analysis. In this case, the perimeter of the atom in question is overlapped by more than one circumference coming from the neighbor atoms. Consequently, the atomic accessibility is calculated by resting to the perimeter of the atom the lengths

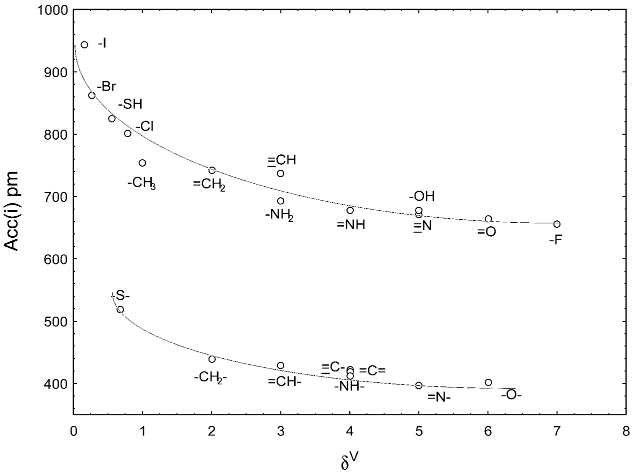

Figure 7. Relation between the atomic accessibility perimeter and the valence degree of different heteroatoms bound to one or two neighbor atoms.

of the arcs overlapping it: Acc(i) = ($L - \delta I$), where δ accounts for the number of neighbors bonded to the corresponding atom. For instance, if we consider the $-CH_2-$ group, the value of Acc(i) will be obtained as the difference between the perimeter of the C atom calculated from the van der Waals radius and 2 times the length of the arc I calculated from the angle of overlapping (eq 12).

In Table 1 we give the values of the atomic accessibilities for atoms with different connectivities. As can be seen, the accessibilities of atoms bonded to two or three neighbors are significantly lower than those having only one neighbor. However, the same kind of proportionality exists between the atomic accessibility and the values of (δ_i^V) for the different groups of atoms with connectivity one, two, or three. In Figure 7 we represent these proportionalities for atomic groups with one and two neighbors.

This means that the atomic accessibility is composed of two different components: one is related to the changes in accessibility due to different heteroatom nature as well as hybridization (multiple bonds), and the other component of the atomic accessibility is related to the number of neighbors bonded to the corresponding atom. In fact, if we consider the 24 atomic groups in Table 1 together, we can find an excellent linear correlation between this property and both the values of $(\delta_i^V)^{-0.5}$ and $(\delta_i)^{-0.5}$ ($R = 0.978$):

$$\mathrm{Acc}(i) = 127.7(\delta_i^{V})^{-0.5} + 1149.3(\delta_i)^{-0.5} - 510.6 \quad (14)$$

This correlation is significantly improved by changing the values of the exponent of the δ values. For instance, $R = 0.9834$ is obtained with -0.3 as exponent (see further analysis of the exponent in the Randić invariant). The previous results mean that the inverse of the square root of the valence degree is a measure of the atomic accessibility when the number of neighbors bonded to the atoms is the same and it can be considered as in the case of the vertex degree as a component of the *relative atomic accessibility*.

It is now straightforward to realize, in a way similar to that for the Randić index, that the valence connectivity index, $^1\chi^V$, is the other component of the *relative accessibility area* of a molecule. It takes into account the differences that arise from the variations in atomic radii for different heteroatoms or due to multiple bonds. This combination of Randić index and molecular connectivity index for explaining the molecular relative accessibility area is probably the main cause for finding these two indices together in many QSAR/QSPR correlations, despite the intercorrelation of both indices in many cases. That is, *they both ($^1\chi$ and $^1\chi^V$) are needed to account for molecular accessibility area*.

7. Interpretation of Extended Connectivity Indices

In light of the current approach, the values of the inverse square root of the vertex degree are interpreted as the length of the arc in the van der Waals circumference accessible from outside. In other words, $(\delta_i)^{-0.5}$ and $(\delta_i^V)^{-0.5}$ values have length units and the first-order connectivity indices are interpreted as components of the accessibility molecular areas. In agreement with this interpretation we are going to interpret the structural meaning of the extended ("higher order") connectivity indices.

In the following we will make all developments in terms of the nonvalence connectivity indices. The same expressions are applicable to the valence connectivity indices by including the superindex V to the values of δ_i and to the corresponding indices.

The connectivity index of path type of second order, $^2\chi_p$ is the product of the values of $(\delta_i)^{-0.5}$ for the atoms in a path of length 2 in the molecule. That is,

$$^2\chi_p = \sum_r [(\delta_i)^{-0.5}(\delta_j)^{-0.5}(\delta_k)^{-0.5}]_r \quad (15)$$

where the summation is carried out for all paths of length 2 in the molecule.

The multiplication of these three accessibility perimeters, $C_{ijk} = (\delta_i)^{-0.5}(\delta_j)^{-0.5}(\delta_k)^{-0.5}$, expressed in length units, gives as a result a volume, expressed in cubic length units, i.e., cubic Randić (R^3). That is, this term represents the volume that is accessible from outside to three adjacent atoms in a path of length 2 in the molecule. Thus, the second-order connectivity index, $^2\chi_p$, which is the sum of volumes C_{ijk}, is interpreted as a component of the *molecular accessibility volume* coming from contributions of paths of length 2 in the molecule. These connectivity indices can be graphically represented as cubes in a three-dimensional space.

In a similar way we can interpret the "higher" order connectivity indices as hypervolumes of order 4, 5, and so on in a molecule coming from contributions of subgraphs of different types t and lengths 3, 4, 5, etc. The main difference between these indices and those before analyzed is that they cannot be represented graphically as geometrical figures, squares, and cubes, as the previous ones.

8. Exponent in the Randić Invariant

The Randić invariant is expressed through eq 4 in which the exponent (−0.5) plays an important role. This exponent was selected by Randić to differentiate the branching of alkanes.[1] In 1988 Randić et al.[23] studied the variation of this exponent as a way to improve the quality of the "branching index" in describing the boiling point of alkanes. As a result of this work the authors show that the change of the exponent from −1/2 to −1/3 improves significantly the linear regression model to describe the cited property of alkanes. The change of this exponent in this work was carried out empirically by following as the main objective the improvement of a QSPR model. However, in 1995 Estrada[24] showed that the use of the exponent −1/2 is not suitable for distinguishing certain pairs of isomers. In fact, for regular graphs, i.e., those having all vertices with the same degree, the Randić index is reduced to $N/2$, where N is the number of vertices (atoms) in the graph. In consequence, cyclohexane (a 2-regular graph) and prismane (a 3-regular graph), which have the same number of vertices, have exactly the same value of the first-order connectivity index. The same happens to cyclooctane and cubane and several other pairs of isomers represented by regular graphs. The solution proposed by Estrada in this work was simply to change the exponent from $-1/2$ to $-1/3$.[24] In more recent works, Amić et al.[25,26] have optimized the value of the exponent for describing different physicochemical properties of organic compounds. In all these studies the optimal values of the exponent in the Randić invariant for connectivity indices have been values of the type $-1/q$. More recently, Gutman has proved that the use of an exponent of -1 does not reflect the extent of branching of the carbon-atom skeleton of organic molecules.[27]

In light of the current approach, the values of $(1/\delta^V)$ are interpreted as a component of the atomic accessibility perimeter. The values for this parameter calculated from the van der Waals and the covalent radii are given in Table 1 for a series of atoms. If we plot these values versus the corresponding values of δ^V as in Figure 6, we observe that the dependence of both variables is inverse, i.e., the greater the accessibility the lower is the valence degree of the atom. However, this dependence is not linear but has the appearance of a model in which the valence degrees are raised to $(-1/q)$. We have used an optimization algorithm for finding the value of q in a model that relates the accessibility to the valence degree for atoms with only one neighbor: $\mathrm{Acc}(i) = a(\delta_i^{\mathrm{V}})^b + c$, where a, b, and c are fit parameters. Using a least-squares algorithm we find that

$$\mathrm{Acc}(i) = 254.8(\delta_i^{\mathrm{V}})^{-0.272} + 515.2 \qquad (16)$$

with a correlation coefficient of 0.985 (the correlation is that observed in Figure 6). The value of the exponent is exactly in the range of optimal values found by several authors using different empirical or mathematical approaches for improving the quality of the connectivity indices. For instance, Amić et al.[26] studied the variation of the exponent in the connectivity index from -1.5 to 0 and determined that the optimal values for describing the solubility of aliphatic alcohols is in the range from -0.6 to -0.2. Randić found the value of $-1/3$, which is very close to the value of -0.272 found here, as the best for describing alkane boiling points. Our findings not only confirm our interpretation of the connectivity indices in terms of accessibility molecular areas but also give a theoretical justification to the use of exponents of the type $-1/q$ in the Randić invariant.

9. Conclusions

According to Heisenberg, phenomenological physics consists of "inventing formulas which seem to reproduce the experiments".[35] This was exactly the approach used first by Randić and later by Kier and Hall in developing the molecular connectivity indices.[1-3] They invented a graph-theoretical invariant based on a molecular graph that has been widely useful in reproducing experimental physicochemical and biological properties. However, the connectivity indices represent a description of a physical reality: the molecular structure. Consequently, the Randić invariant and connectivity indices can be derived from a large number of different physical viewpoints and widely different mathematical formulations.

In fact, we have made here a reformulation of the connectivity indices by assuming a physical basis and keeping approximately the same mathematical formulation used by the original authors. However, we have proved that the first-order connectivity indices can be derived from a different mathematical (matrix−vector) formulation using a common graph-theoretical invariant for several different topological indices.[31]

We have proved here that the connectivity indices represent a good measure of the molecular area that is accessible from outside the molecule. Accordingly, this accessibility area is of great value in describing physicochemical properties of molecules as a measure of the extension of intermolecular interactions. We think that this discovery of the physical meaning of connectivity indices will increase the interest in the applications of these indices to describe physical and biological properties of molecules as well as to provide a physical place for topological indices among the pool of molecular descriptors.

A final lesson that we have to learn from the interpretation of connectivity indices is the following. Because molecules are "physical" objects, i.e., objects in a real world, topological indices are mathematical representations of a physical reality. Consequently, they necessarily have to have a physical meaning. That is, they are not mere mathematical objects, in a similar way that the variation of the distance with respect to time is not a mere mathematical object called "derivative" but a description of a physical reality (the velocity) represented in the language of mathematics. Here, graphs are the mathematical objects (as derivatives in the example of velocity), molecules are the physical reality (as the velocity), and topological indices are the mathematical representation of this reality (as the first derivative of distance with respect to time is to velocity). So, it will not be surprising to find a physical basis for most of the topological indices in the future as we found here for connectivity indices.

Acknowledgment. Thanks are given to the Xunta de Galicia (PR405A 2001/65-0) and Ministerio de Educación y Cultura, Spain (project PM99-0125), for partial financial support. Comments of both reviewers contributed to a significant improvement of this material.

References and Notes

(1) Randić, M. *J. Am. Chem. Soc.* **1975**, *97*, 6609.
(2) Kier, L. B.; Hall, L. H.; Murray, W. J.; Randić, M. *J. Pharm. Sci.* **1975**, *64*, 1971.
(3) Kier, L. B.; Hall, L. H. *J. Pharm. Sci.* **1976**, *65*, 1806.
(4) Randić, M.; Trinajstić, N. *J. Mol. Struct. (THEOCHEM)* **1993**, *284*, 209.
(5) Pogliani, L. *J. Phys. Chem.* **1993**, *97*, 9029.
(6) Pogliani, L. *J. Phys. Chem.* **1994**, *98*, 1494.
(7) Pogliani, L. *J. Phys. Chem.* **1995**, *99*, 925.
(8) Pogliani, L. *J. Phys. Chem.* **1996**, *100*, 18065.
(9) Pogliani, L. *J. Phys. Chem. A* **1999**, *103*, 1598.
(10) Pogliani, L. *Chem. Rev.* **2000**, *100*, 3827.
(11) Pogliani, L. *J. Phys. Chem. A* **2000**, *104*, 9029.
(12) Gálvez, J.; García-Domenech, R.; De Julián-Ortiz, J. V. In *Fundamentals of Molecular Similarity*; Carbó-Dorca, R., Ed.; Kluwer Academic: New York, 2001; p 11.
(13) Estrada, E. *Chem. Phys. Lett.* **1999**, *312*, 556.
(14) Randić, M. *J. Mol. Graphics Modell.* **2001**, *20*, 19.
(15) Kier, L. B.; Hall, L. H. *Molecular Connectivity in Chemistry and Drug Research*; Academic Press: New York, 1976.
(16) Kier, L. B.; Hall, L. H. *Molecular Connectivity in Structure−Activity Analysis*; Research Studies Press: Letchworth, 1986.
(17) Hall, L. H.; Kier, L. B. In *Topological Indices and Related Descriptors in QSAR and QSPR*; Devillers, J., Balaban, A. T., Eds.; Gordon and Breach: Amsterdam, 2001; p 307.
(18) Hall, L. H.; Kier, L. B. *J. Mol. Graphics Modell.* **2001**, *20*, 4.
(19) Stankevich, V.; Skvortsova, M. I.; Zefirov, N. S. *J. Mol. Struct. (THEOCHEM)* **1995**, *342*, 173.
(20) Galvez, J. *J. Mol. Struct. (THEOCHEM)* **1998**, *429*, 255.
(21) Kier, L. B.; Hall, L. H. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 792.
(22) Kier, L. B.; Hall, L. H. *J. Mol. Graphics Modell.* **2001**, *20*, 76.
(23) Randić, M.; Hansen, P. J.; Jurs, P. C. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 60.
(24) Estrada, E. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 1022.

(25) Amić, D.; Beslo, D.; Lucic, B.; Nikolić, S.; Trinajstić, N. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 819.

(26) Nikolić, S.; Trinajstić, N.; Amić, D.; Beslo, D.; Basak, S. C. In *QSPR/QSAR Studies by Molecular Descriptors*; Diudea, M., Ed.; Nova Science: New York, 2001; p 63.

(27) Gutman, I.; Lepovic, L.; Vidovic, D.; Clark, L. H. *Indian J. Chem. A* **2002**, *41A*, 457.

(28) Randić, M.; Mills. D.; Basak, S. C. *Int. J. Quantum Chem.* **2000**, *80*, 1199.

(29) Randić, M.; Plavsić, D.; Ler, N. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 657.

(30) Randić, M.; Basak, S. C. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 614.

(31) Estrada, E. *Chem. Phys. Lett.* **2001**, *336*, 248.

(32) Hall, L. H.; Kier, L. B.; Murray, W. J. *J. Pharm. Sci.* **1975**, *64*, 1974.

(33) Bondi, A. *J. Phys. Chem.* **1964**, *68*, 441.

(34) The experimental values of the van der Waals (vdW) radii depend on the property observed and the model used to interpret them. Consequently, different values of vdW radii are found for the same atom in the literature.

(35) Heinsenberg, W. In *From a Life of Physics*; World Scientific: Singapore, 1989; p 31.