

# Prediction of Molecular Volume and Surface of Alkanes by Molecular Topology

Jorge Galvez\*

Molecular Connectivity & Drug Design Research Unit, Department of Physical Chemistry,  
Faculty of Pharmacy, University of Valencia, Av. V.A. Estelles s.n., 46100-Burjasot, Valencia, Spain

Received February 25, 2003

Molecular volume and molecular surface are expressed as a function of topological degree in alkane graphs. This allows not only a straightforward approach to calculate such physicochemical magnitudes but also an interpretation of the role of the local vertex invariant (LOVI) or valence degree,  $\delta$ , as well as the connectivity indices in the prediction of physicochemical properties. The interpretation is based on the concept of *molecular accessibility* (as introduced by Estrada, *J. Phys. Chem. A* **2002**, 106, 9085) for which precise mathematical definitions are provided.

## 1. INTRODUCTION

Topological indices (TI) have deserved more and more attention from chemists along the later years, and the number of papers on the matter have experienced a significant increment. Since the pioneer work from Sylvester, who introduced the topological matrix in 1874<sup>1</sup> and the term *graph* in 1878,<sup>2</sup> and Wiener<sup>3</sup> till the contributions of Balaban,<sup>4</sup> Randić,<sup>5</sup> Kier and Hall,<sup>6</sup> Bonchev,<sup>7</sup> Basak,<sup>8</sup> Pogliani,<sup>9</sup> and others, its applicability to the prediction and, in some cases, interpretation of many physical, chemical, and biological properties has been widely demonstrated. Our group has focused its efforts mainly on their application in drug design, so that many new active compounds within different pharmacological skills have been found out.<sup>10–13</sup>

Nevertheless, one of the most frequent criticism about TI, since their early stages of development, was the absence of a physical or physicochemical interpretation, so that many critics think about molecular topology just like a typical *heuristic* and/or *stochastic* formalism. This is why it is important to attempt to provide an explanation about their physical meaning or their physical applicability.

A good way to make so is exploring the meaning of one of the most widespread TIs, the Randić index, further developed by Kier and Hall; hereinafter we will refer to both as *connectivity indices*: CI. They are based on the notion of valence degree, also known as Local Vertex Invariant, LOVI, and because of their pioneer character as well as their huge applicability, several attempts have been made to explain why they work so well. Some explanations, as for instance Stankevich's,<sup>14</sup> are based upon a quantum mechanical interpretation, whereas others, such as Galvez's,<sup>15</sup> are based just on topological features within a framework of self-consistency and, more recently, Kier and Hall have provided insight from a probabilistic point of view.<sup>16</sup>

Very recently, Estrada<sup>17</sup> has discussed the meaning of molecular connectivity indices and given a possible physicochemical interpretation in terms of loss of *accessibility*.

In this paper, using a Estrada-like approach, we suggest new possible interpretation for both the topological valence

and the connectivity indices in alkanes. Our results are partially coincident with Estrada's but partially divergent as well.

## 2. CONNECTIVITY INDICES

The Randić algorithm considers the molecule as a sum of parts, namely the sum of the inverses of the square roots of the products of the pair of values of vertices degrees,  $\delta_i$ , for each one of the vertices of the graph. The vertex degree (also referred to as topological valence) is just the number of edges on the same vertex at the hydrogen-depleted graph of the molecule. Thus the original Randić index is given by

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

Kier and Hall proposed to assign the  $(\delta_i \delta_j)^{-0.5}$  terms as Cij.

Subsequent development of this algorithm has been carried out elsewhere,<sup>18</sup> so that the extended connectivity indices were raised. The number of edges in a graph is called *order* of the graph. For orders below or equal to 2, there is only one type (t) of graph: The *path* type (t = p), while above 2 it may appear as graphs of the types *cluster* (t = C), *path-cluster* (t = pC), and *chain* (t = Ch), except for those graphs of the third order, for which the pC type is not possible.

For these extended indices, the calculation implies the summation over all the subgraphs of a given type and order, of the inverse of the square roots of the products of valences for every vertex in the graph, i.e.:

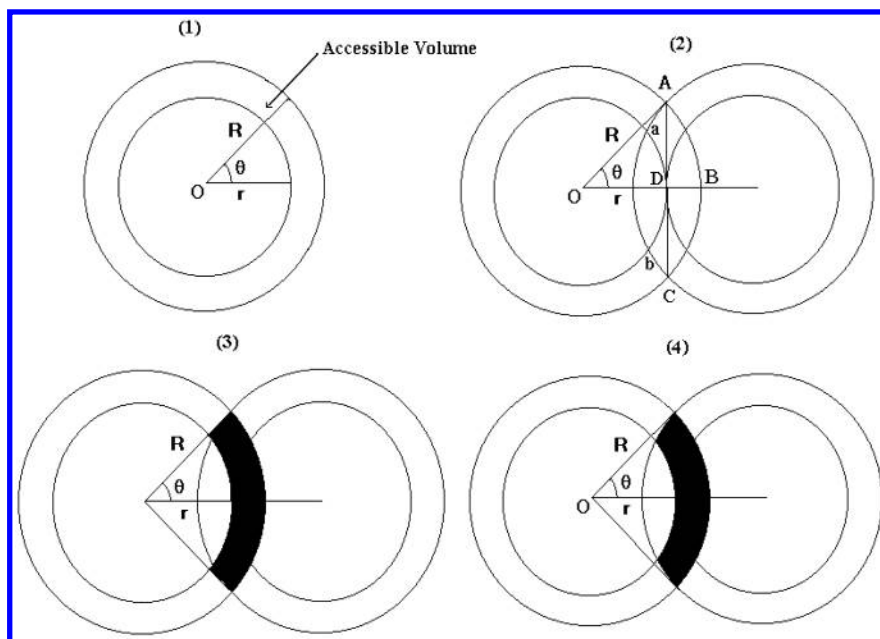
$${}^m\chi_t = \sum_k {}^mC_k \quad {}^mC_k = \prod_k (\delta_i)_k^{-0.5} \quad (2)$$

Valence indices can also be achieved for molecules containing heteroatoms, just assigning a weight to every heteroatom, according to different algorithms.<sup>19</sup>

## 3. THE MEANING OF THE VALENCE DEGREE ( $\Delta$ )

Accordingly to the stated above, it seems reasonable that, prior to the understanding of the meaning of the connectivity

\* Corresponding author fax: +34 6 354 4891; e-mail: jorge.galvez@uv.es.



**Figure 1.** Overlapping  $sp^3$  carbons.  $R$  = van der Waals radius,  $r$  = covalent radius. (1) Isolated carbon (carbon in methane). Look at the “accessible” volume. (2) Two carbons bounded (ethane). (3) Angular loss of accessible volume (darken) for the carbon at the left. (4) Real loss of accessible volume (darken) for the carbon at the left.

indices, we wonder about the meaning of their original bricks: the valence degrees ( $\delta_i$ ).

The usual language in many branches of physical sciences is plenty of expressions such as *physical meaning* or *physicochemical meaning* of this or that mathematical concept. It is beyond a simple semantic issue to wonder whether using these expressions is a real attempt to give a *physical meaning* to what is just a pure mathematical abstraction or it is rather finding a physical *application* of a mathematical concept into a particular physical (“real”) property. Under this point of view, the important question is not about the meaning of  $\delta_i$  but about the sort of physical or physicochemical property for which it is able to achieve information about.

Thus, Estrada’s approach based on the loss of accessibility for a given atom when bound to others is a very good starting point.

Figure 1 shows the different situations for an isolated carbon atom [case (1)] with respect to the atom bound to another [cases (2), (3), and (4)]. Since the accessibility of every atom is limited, it is intuitive that when the atom links into another, there is some loss of accessibility, even if we have not defined so far what we do understand by *accessibility*.

In Figure 1, the atoms are represented by concentric spheres in which  $r$  is the covalent radius and  $R$  is the van der Waals radius. This is a Estrada-like approach, but, unlikely, we do not understand the overlapping resulting in a loss in the *perimeter* accessibility but in a loss of *volume* accessibility. Indeed, for the single atom, the accessible volume, defined as the volume enclosed between the two concentric spheres, is

$$V_0 = \frac{4}{3}\pi(R^3 - r^3) \quad (3)$$

where  $V_0$  is the accessible volume of the isolated carbon atom.

If we do not consider the hydrogens, as usual in graph-theoretical chemistry, then  $r_0$  stands for the covalent radius of the carbon in methane.

When the atom A is bound to another, B, there is a loss of accessible volume, so that the remaining accessible volume is

$$V_a = V_0 - V_0\left(\frac{2\theta}{2\pi}\right) = V_0\left(1 - \frac{\theta}{\pi}\right) \quad (4)$$

where  $V_a$  is the accessible volume of the atom A when bound to B, and  $\theta$  is the angle between the vector radii  $R$  and  $r$  (see Figure 1).

Obviously, if A is bound to  $\delta_i$  other carbons (what is another way to define  $\delta_i$ ) then its accessible volume is

$$V_a = V_0 - \delta_i V_0\left(\frac{2\theta}{2\pi}\right) = V_0\left(1 - \delta_i \frac{\theta}{\pi}\right) \quad (5)$$

#### 4. VALENCE AND VOLUME ACCESSIBILITY

From here it is immediate the definition of volume accessibility,  $A_V$ , as

$$A_V = \frac{V_a}{V_0} = \left(1 - \delta_i \frac{\theta}{\pi}\right) \quad (6)$$

It is interesting to realize that a natural definition of accessibility has been defined from this elemental reasoning.

This definition brings with it a first evident consequence: For  $\delta_i = 0$ , i.e., for an isolated carbon atom (methane)  $A_V = 1$ . This is the top value for volume accessibility. To account for the accessibility in other valences (1, 2, 3, and 4), it is necessary to assign values for  $\theta$ . Arbitrarily, as a first approach we can take the value  $\theta = \pi/4$  (i.e.  $45^\circ$ ). In that case

$$A_\delta = 1 - \frac{\delta}{4} \quad (7)$$

**Table 1.** Values of Valence Degrees and Accessibility for the Five Possible Delta Values<sup>a</sup>

$\delta_i$	$1 - \delta_i/4$	$1 - (\theta_i\delta_i)/\pi$	$1 - (f_i\theta_i\delta_i)/\pi$	$(1 - \delta_i/4)^{0.5}$	$(1/\delta_i)^{0.5}$	$\theta_i$ (°)	$\phi_i$ (°)
0	1	1	1	1		63.95	0
1	0.750	0.646	0.703	0.866	1.000	63.82	30
2	0.500	0.293	0.409	0.707	0.707	63.69	45
3	0.250	-0.058	0.120	0.500	0.577	63.56	60
4	0.000	-0.408	-0.165	0.000	0.500	63.43	90

<sup>a</sup>  $\delta_i$  = valence degree,  $\theta_i$  = covalent-van der Waals vector radii angle,  $\phi_i$  = equivalent angle,  $1 - \delta_i/4$  = valence accessibility or first volume accessibility ( $A_{V_1}$ ),  $1 - (\theta_i\delta_i)/\pi$  = second volume accessibility ( $A_{V_2}$ ),  $1 - (f_i\theta_i\delta_i)/\pi$  = third volume accessibility ( $A_{V_3}$ ),  $(1 - \delta_i/4)^{0.5}$  = surface accessibility ( $A_S$ ),  $(1/\delta_i)^{0.5}$  = basic connectivity input,  $\theta = \arcsin\left(\frac{0.7465 + 0.00354\delta}{1.7}\right)$ .

**Table 2.** Values of Standard Volumes and Surface Areas of a Set of Alkanes as Calculated from HiperChem 7.0 (Chemplus), Together with Their Values of Different Valence, Volume, and Surface Accessibilities<sup>a</sup>

compound	volume <sup>b</sup> (Å <sup>3</sup> )	surface area <sup>b</sup> (Å <sup>2</sup> )	$A_\delta$	$A_{V_1}$	$A_{V_2}$	$A_{S_1}$	$A_{S_2}$
methane	49.026	73.483	1	1	1	1	1
ethane	67.487	91.916	1.5	1.2916	1.4053	1.7321	1.866
propane	85.756	110.519	2	1.5846	1.8147	2.4392	2.7195
<i>n</i> -butane	104.040	129.290	2.5	1.8777	2.2241	3.1463	3.5733
2Mpropane	104.564	123.952	2.5	1.8792	2.2283	3.0981	3.5490
<i>n</i> -pentane	122.690	148.222	3	2.1708	2.6336	3.8534	4.4265
2Mbutane	122.801	144.223	3	2.1722	2.6377	3.8052	4.4025
22MM propane	124.580	138.825	3	2.1752	2.6459	3.4641	4.2323
<i>n</i> -hexane	140.731	166.918	3.5	2.4638	3.0430	4.5605	5.2806
2Mpentane	140.352	159.302	3.5	2.4653	3.0471	4.5123	5.2562
3Mpentane	140.352	159.302	3.5	2.4653	3.0471	4.5123	5.2562
23MMbutane	140.413	154.846	3.5	2.4668	3.0512	4.4641	5.2323
22MMbutane	140.097	151.040	3.5	2.4683	3.0553	4.1712	5.0855
<i>n</i> -heptane	159.508	186.438	4	2.7569	3.4524	5.2676	6.1335
2Mhexane	158.379	178.609	4	2.7584	3.4565	5.2194	6.1095
3Mhexane	159.238	177.279	4	2.7584	3.4565	5.2194	6.1095
23MMpentane	157.767	170.462	4	2.7598	3.4606	5.1712	6.0855
24MMpentane	159.276	173.949	4	2.7598	3.4606	5.1712	6.0855
22MMpentane	158.688	168.892	4	2.7613	3.4648	4.8783	5.9390
33MMpentane	158.688	168.892	4	2.7613	3.4648	4.8783	5.9390
22MM3Mbutane	157.765	163.137	4	2.7628	3.4689	4.8301	5.9150

compound	volume <sup>b</sup> (Å <sup>3</sup> )	surface area <sup>b</sup> (Å <sup>2</sup> )	$^0\chi$	$^1\chi$	$^2\chi$	$^3\chi_p$
methane	49.026	73.483	0	0	0	0
ethane	67.487	91.916	2	1	0	0
propane	85.756	110.519	2.7071	1.4142	0.707	0
<i>n</i> -butane	104.040	129.290	3.4142	1.9142	1	0.5
2Mpropane	104.564	123.952	3.5774	1.7320	1.7320	0
<i>n</i> -pentane	122.690	148.222	4.1213	2.4142	1.3535	0.7071
2Mbutane	122.801	144.223	4.2845	2.2700	1.8020	0.8165
22MM propane	124.580	138.825	4.5000	2.0000	3.0000	0
<i>n</i> -hexane	140.731	166.918	4.8284	2.9142	1.7071	0.9571
2Mpentane	140.352	159.302	4.9916	2.7700	1.8021	0.8165
3Mpentane	140.352	159.302	4.9916	2.808	1.9217	1.3938
23MMbutane	140.413	154.846	5.1547	2.6427	2.4880	1.3333
22MMbutane	140.097	151.040	5.2071	2.6427	2.9142	1.0607
<i>n</i> -heptane	159.508	186.438	5.5355	3.4142	2.0606	1.2071
2Mhexane	158.379	178.609	5.6987	3.2700	2.5360	1.1350
3Mhexane	159.238	177.279	5.6987	3.3080	2.3021	1.4784
23MMpentane	157.767	170.462	5.8618	3.1807	2.6295	1.7820
24MMpentane	159.276	173.949	5.8618	3.1258	3.0234	0.9428
22MMpentane	158.688	168.892	5.9142	3.0607	2.0607	1.2071
33MMpentane	158.688	168.892	5.9142	3.1213	2.8713	1.9142
22MM3Mbutane	157.765	163.137	6.0774	2.9434	3.5207	1.7320

<sup>a</sup> Path connectivity indices between the zero and the third order are also outlined. <sup>b</sup> Assignment of symbols: Standard volume and surface area  $A_\delta = \Sigma[1 - \delta_i/4]$ ,  $A_{V_1} = \Sigma[1 - \delta_i\theta/\pi]$ ,  $A_{V_2} = \Sigma[1 - f\delta_i\theta/\pi]$ ,  $A_{S_1} = \Sigma[\cos \phi_i]$ ,  $A_{S_2} = 1/2\Sigma[1 + \cos \phi_i]$ .

Interestingly, we have reached to the simplest and intuitive notion of accessibility as related to the concept of chemical valence. We may call it as *valence accessibility* ( $A_\delta$ ). Indeed, when  $\delta_i = 0$  (methane), the four carbon valences are free and the accessibility is maximum ( $A_\delta = 1$ ), whereas  $\delta_i = 4$  (the central carbon in neopentane), all the valences are occupied and the accessibility is null.

Table 1 illustrates the values of  $A_\delta$  for each of the five possible values of  $\delta_i$ .

The definition of accessibility derived from eq 7 provides a simple, linearly changing and intuitive measure of accessibility.

To go forward in the understanding of the concept of accessibility, we must turn over the angle  $\theta$ . From Figure 1,

**Table 3.** Comparison between Molecular Volumes and Surface Areas Calculated through Standard Approach and through Graph Theoretical Formalisms Based on Accessibility

compound	standard vol <sup>a</sup> (Å <sup>3</sup> )	vol from eq 10 multiplied by $f_h = 2.45$	vol calcd from eq 17 corrected by (20) and (21)	standard surface area <sup>a</sup> (Å <sup>2</sup> )	surface area calcd from eq 32 <sup>b</sup>
methane	49.026	50.421	50.883	73.483	75.256
ethane	67.487	68.061	68.571	91.916	93.066
propane	85.756	85.784	86.489	110.519	110.267
<i>n</i> -butane	104.040	103.513	104.407	129.290	127.469
2Mpropane	104.564	103.603	104.635	123.952	126.296
<i>n</i> -pentane	122.690	121.243	122.324	148.222	144.670
2Mbutane	122.801	121.329	122.553	144.223	143.498
22MM propane	124.580	121.510	123.009	138.825	135.200
<i>n</i> -hexane	140.731	138.966	140.242	166.918	161.872
2Mpentane	140.352	139.057	140.471	159.302	160.699
3Mpentane	140.352	139.057	140.471	159.302	160.699
23MMbutane	140.413	139.148	140.699	154.846	159.527
22MMbutane	140.097	139.238	140.927	151.040	152.401
<i>n</i> -heptane	159.508	156.697	158.160	186.438	179.073
2Mhexane	158.379	156.788	158.388	178.609	177.901
3Mhexane	159.238	156.788	158.388	177.279	177.901
23MMpentane	157.767	156.871	158.617	170.462	176.728
24MMpentane	159.276	156.871	158.617	173.949	176.728
22MMpentane	158.688	156.962	158.844	168.892	169.603
33MMpentane	158.688	156.962	158.844	168.892	169.603
22MM3Mbutane	157.765	157.055	158.994	163.137	168.431

<sup>a</sup> Values calculated from HiperChem 7.0 (Chemplus). <sup>b</sup> The values are corrected by a hybridization factor  $f_h = 1.35$ .

it is evident that

$$\cos \theta = \frac{r}{R} \quad \text{Thus, } \theta = \arccos\left(\frac{r}{R}\right) \quad (8)$$

Accordingly, we can get a single value for  $\theta$ , taking the standard average values:<sup>22</sup>  $r = 0.77$  Å and  $R = 1.7$  Å, then  $\theta = 1.1007$  rad (about 63°).

Either from eqs 7 or 8 it is to be expected for accessibility to be additive, which means that molecular accessibility,  $A_V$ , must be the sum of atomic accessibilities.  $A_{V_i}$

$$A_V = \sum_{i=1}^n A_{V_i} \quad (9)$$

wherein  $n$  stands for the number of carbons in the alkane.

The values of volume molecular accessibility for a set of alkanes ranging from methane up to the heptane isomers are illustrated in Table 2 for both definitions of accessibility as derived from eqs 6 and 7. For comparison only, Table 2 also includes the values of the reference volumes and surface areas as calculated from HyperChem 7.0 (Chemplus).

## 5. CALCULATION OF MOLECULAR VOLUME

Since volumetric accessibility is a measure of the accessible volume, it is to be expected that molecular volume may be calculated from accessibility. If we consider the contribution of the van der Waals volume as the most important, then the molecular volume is given by

$$V = nV_{VDW} - V_0 \sum_{i=1}^n n_i A_{V_i} \quad (10)$$

where  $n$  is the number of carbon atoms in the alkane,  $V_{VDW}$  is the van der Waals volume (Å<sup>3</sup>),  $n_i$  is the number of carbons with accessibility  $A_{V_i}$ , and  $A_{V_i}$  is the accessibility of the atom  $i$  (either from eqs 7 or 8).

The second term of this equation represents the volume excluded because of the loss of accessibility.

The values of the standard volumes as well as the volumes calculated through eq 10 for the set of alkanes are shown in Table 3. To get a good fitting

between the experimental and the calculated values it is necessary to multiply this last by  $f_h$ , a factor taking into account the  $sp^3$  hybridization which had not been considered so far. The hybridization as well as the presence of hydrogens (both not considered in the present formalism) is encoded within a constant factor of about 2.43.

As outlined in Table 3, the prediction is good enough with respect to the number of carbons, but both definitions of accessibility—based on eqs 6 and 7—fail in the prediction of isomers volume. The reason is simple: Contrary to appearance, both accessibility values are not depending on the degree of branching in the alkane. Indeed, from eq 10 it follows:

$$V = nV_{VDW} - V_0 \theta \left( \frac{1+n}{2} \right) \quad (11)$$

It is easy to demonstrate that the van der Waals volume can be expressed as a function of  $V_0$ , so that the final equation is

$$V = V_0 \left[ \left( \frac{n}{1 - (\cos \theta)^3} \right) - \theta \left( \frac{1+n}{2} \right) \right] \quad (12)$$

Since  $\theta$  and  $V_0$  are constant, the molecular volume is mainly depending on  $n$ .

This is an apparent paradox: Despite the fact that we have introduced the values of the valence degrees into the accessibility equation, molecular volume results to be independent of the  $\delta$  values. Nevertheless this result can be explained considering that the overall sum of  $\delta$  values is just twice the number of edges in the graph,  $m$ , which, in turn, is depending only on the number of vertices,  $n$ . This



explains why it is necessary to transform  $\delta$  into another nonlinear function, for example raising them into a negative potency (usually  $-0.5$ , but also other values)<sup>20,21</sup> to build up topological indices.

The reason for this default is that the  $\theta$  value is not constant, as we have assumed so far. Since  $\theta$  depends on  $r$  and, in turn,  $r$  depends on the valence degree, this correction must be introduced. An experimental fact easily observable is that the bonds lengths increase along with  $\delta$ . We found that a simple linear relation between the bond lengths and  $\delta$  works well in alkanes. Indeed, using the AM1 semiempirical approach and optimizing the geometry through the conjugated gradient method, the so obtained  $r$  values vary linearly along with  $\delta$  as

$$r = r_0 + m\delta \quad (13)$$

$$N = 5, r^2 = 0.9939, F = 31576$$

where  $r_0 = 0.7465 \text{ \AA}$  and  $m = 0.00354 \text{ \AA}$ ,  $r$  is the covalent radii of the carbon atom with  $\delta = \delta_i$ , and  $r_0$  is the covalent radius of the carbon in methane.

The increment of the bond length along with  $\delta$  can be readily figured out from the repulsion among the increasing number of nuclei as well as the repulsion among the likewise increasing number of inner electrons, both, nuclei and electrons, surrounding the carbon atom taken as reference.

Taking  $R = 1.7 \text{ \AA}$ , the angle  $\theta$  is

$$\theta = \arccos\left(\frac{0.7465 + 0.00354\delta}{1.7}\right) \quad (14)$$

Thus, the equations for accessibility and volume are, respectively

$$A_{V_i} = 1 - \frac{\theta_i \delta_i}{\pi} = 1 - \frac{\delta_i}{\pi} \arccos\left(\frac{0.7465 + 0.00354\delta_i}{1.7}\right) \quad (15)$$

$$V = nV_{VDW} - \frac{V_0}{\pi} \sum_{i=1}^n \frac{\delta_i}{\pi} \arccos\left(\frac{0.7465 + 0.00354\delta_i}{1.7}\right) \quad (16)$$

The values of  $\theta$  as well as the correspondent accessibility values obtained from eq 15 are illustrated also in Table 1 (second column). It is interesting to realize that for  $\delta = 3$  and  $\delta = 4$ , the values of accessibility are negative for  $A_{v1}$ . Nevertheless, no one alkane shows negative values of molecular accessibility because the positive inputs of degrees 1 and 2 balance the negative inputs of degrees 3 and 4. This is illustrated in Table 2 (column 3). It is also noteworthy from Table 2, the improvement in the predictive ability of molecular volume from the new model with respect to the former one, so that the tendencies in the volume changes from isomer to isomer are now taken into account.

Table 2 also illustrates the predicted values of volume from eq 3. The results are also in excellent agreement with the experimental values, and it may provide a theoretical explanation of results from other authors, as for instance the excellent fitting obtained by Kier and Hall<sup>23</sup> when correlating Bondi volume with  $\delta$  and  $\delta^v$ . If we express the molecular volume as function of  $V_0$ , the resulting equation is

$$V = V_0 \left[ \sum_i \left( \frac{1}{1 - \cos^3 \theta_i} \right) - \frac{1}{\pi} \sum_i \theta_i \delta_i \right] = V_0 \left[ \sum_i \left( \frac{1}{1 - \cos^3 \theta_i} \right) - \frac{1}{\pi} \sum_i \theta_i \arccos\left(\frac{0.7465 + 0.00354\delta_i}{1.7}\right) \right] \quad (17)$$

Although at this point it is clear that the definition of volume accessibility is a useful tool to define in a precise way the concept of accessibility, and by means of that, predicting molecular volume, we have still to relate this concept with the connectivity indices.

There is a way to make so if we consider that the definition of accessibility we have used so far is not entirely correct. Indeed if we look at Figure 1(4), we immediately realize that the accessible volume between two carbon atoms is not given exactly either by eq 4 nor eq 5. Indeed the real volume excluded is not equal to the angular volume excluded (Figure 1(3)), but it is just the darkened surface in Figure 1(4). Hence, a corrective factor must be introduced in eq 3, so that

$$A_{V_i} = 1 - \frac{f_i \delta_i \theta_i}{\pi} \quad (18)$$

It is not difficult to demonstrate that the real excluded volume (darkened in Figure 1(4)) is just the difference between twice-the-volume of the spherical casket limited by the segment  $AC$  and the van der Waals overlapping arc  $ABC$  and twice-the-volume of the small casket limited by the segment  $ab$ —not drawn in the figure for more clarity—and the covalent overlap arc  $aDb$ .

Thus, the real volume excluded is

$$V_e = \pi(R - r)(R^2 - r^2)\sin^2 \theta \quad (19)$$

Consequently, the accessibility is given by

$$A_{V_i} = 1 - \frac{V_e}{V_0} = 1 - \frac{3(1 - \cos \theta)(1 - \cos^2 \theta)^2 \delta}{4(1 - \cos^3 \theta)} = 1 - \frac{f \delta \theta}{\pi} \quad (20)$$

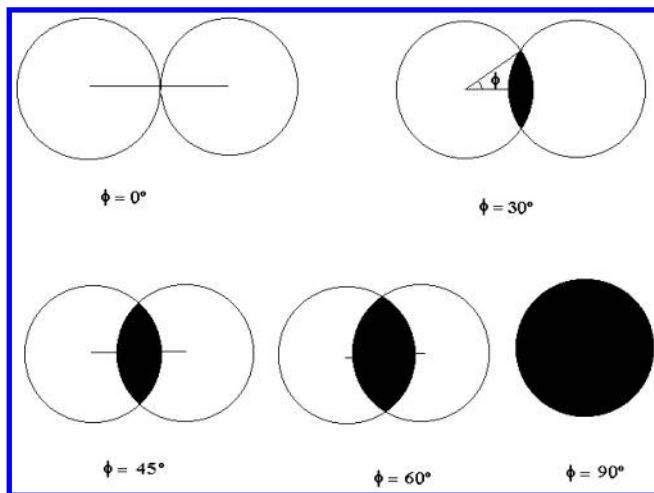
Once more, for further simplicity subindex  $i$  has been removed from eq 20.

From experimental calculations it stands that the last two terms of the above equation fit well to a linear  $f$ - $\delta$  dependence:

$$f = 0.8436 - 0.0041 \delta \quad (21)$$

The values of accessibility,  $A_{v2}$ , according to eq 20 are illustrated in Table 1 for every  $\delta$ . The only negative input is for  $\delta = 4$ . As for  $A_{v1}$ , there are no alkanes showing negative values of the overall molecular accessibility (the sum of eq 20's terms over all the vertices of the graph) because the positive inputs of degrees 1, 2, and 3 balance the negative inputs of degree 4 (see Table 2, sixth column).

The molecular volumes for the same set of alkanes, calculated through eqs 17, 20, and 21, are outlined in Table 3. As may be verified there, this improved calculation of



**Figure 2.** Overlapping zones as a function of angle  $\phi$ . Overlapping is complete (accessibility null) for  $\phi = 90^\circ$ .

accessibility allows for a more accurate prediction of molecular volumes.

## 6. SURFACE ACCESSIBILITY AND RELATION WITH CONNECTIVITY INDICES

Let us get back to the definition of valence accessibility given by eq 7.

$$A_{\delta i} = 1 - \frac{\delta_i}{4}$$

The reason this expression fails in the prediction of the volume of alkane isomers (what forced us to introduce the corrective term  $\theta$  as a function of  $\delta$ ) is closely related to the need to raise the  $\delta$  values to the exponent  $-0.5$  (or any other). From eq 7,  $A_\delta$  can be expressed as

$$A_{\delta i} = \left(1 - \frac{\sqrt{\delta_i}}{2}\right) \left(1 + \frac{\sqrt{\delta_i}}{2}\right) \quad (22)$$

Since for any value of  $\delta$  the term  $(\delta)^{1/2}/2$  is the value of the sinus (or cosine) of a given angle, then we may write it as

$$A_{v_i} = (1 - \sin \phi)(1 + \sin \phi) = 1 - \sin^2 \phi = \cos^2 \phi \quad (23)$$

This is an interesting point of view, since, once more, we are assigning angles to accessibility. What comes next is wondering about the meaning of the  $\phi$  angle. This is not difficult to figure out as long as we take into account the definitions of accessibility we have introduced above. Indeed, for  $\phi = 0^\circ$  ( $\delta = 0$ ),  $r = R$ , i.e., the covalent radius gets its maximum value, whereas for  $\phi = 90^\circ$  ( $\delta = 4$ ),  $r = 0$ . From here it is immediate to assign  $\phi$  as a measure of the covalent accessibility, i.e.:  $\phi = 0^\circ \Rightarrow A_v = 1$  (maximum accessibility, i.e., the corresponding to carbon in methane), whereas  $\phi = 90^\circ \Rightarrow A_v = 0$  (minimum accessibility, i.e., that of the central carbon in neopentane).

According to this interpretation,  $\phi$  would play a similar role as  $\theta$  in the  $sp^3$ - $sp^3$  bonds, i.e.,  $\phi$  represents the overlapping angle between two hypothetical spheres with radius  $= r$ . Figure 2 illustrates the five possible overlapping situations as a function of  $\delta$ .

Despite the failure of  $A_{\delta i}$  in the distinction of isomers' molecular volumes, it is clear that  $A_\delta$  and hence  $\cos^2 \phi$  are a good measure of molecular volume in alkanes (the corrections introduced along this work from the original definition on up, i.e.,  $\theta$  and  $f$ , do not change the concept but just the value of  $\phi$ ). Therefore, if  $\cos^2 \phi$  is a measure of volume, then  $\cos \phi$  must be a measure of molecular surface area. Let us see how we can demonstrate it: First, we may introduce the concept of *surface accessibility*. Following the Estrada's van der Waals-covalent model, it is noteworthy that, while the loss of volume accessibility is depending on the difference ( $R^3 - r^3$ ), the loss of surface accessibility when a carbon atom is bound into another is just depending on  $R^2$ . As may be verified in Figure 1(2), the loss of surface area per atom is just the area of the spherical casket (Asc) resulting from the intersection of the two van der Waals spheres, i.e., the casket defined by the segment AC and the arc ABC:

$$Asc = 2\pi R(R - r) \quad (24)$$

Consequently the accessible surface area stands

$$S_a = 4\pi R^2 - 2\pi R(R - r) \quad (25)$$

where  $S_a$  is the accessible surface area.

Given that the total available surface area,  $S_0$ , is equal to  $4\pi R^2$ , the surface accessibility is

$$A_{S_i} = \frac{S_a}{S_0} = \frac{1}{2} \left(1 + \frac{r}{R}\right) = \frac{1}{2} (1 + \cos \phi_i) \quad (26)$$

The same result can be achieved by expressing the surface accessibility,  $A_{S_i}$ , for a given atom  $i$  as

$$A_{S_i} = \frac{S_a}{S_0} = \frac{V_0}{V_0 S_0} \left(\frac{dV_a}{dr}\right) = \frac{V_0}{S_0} \left(\frac{dA_v}{dr}\right) \quad (27)$$

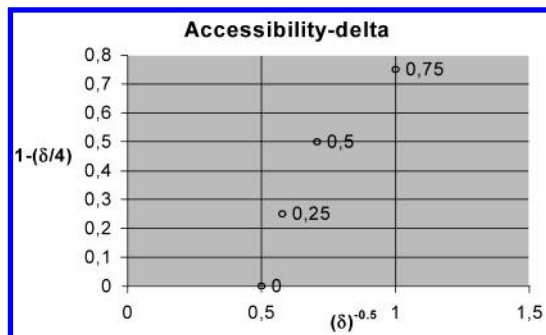
Since both  $V_0$  and  $S_0$  are constants, the surface accessibility is depending on the first derivative of volume accessibility with respect to  $r$ . Thus, it follows that

$$\begin{aligned} \left(\frac{dA_v}{dr}\right) &= \left(\frac{d \cos^2 \phi}{dr}\right) = \left(\frac{d \cos^2 \phi}{d\phi}\right) \left(\frac{d\phi}{dr}\right) = \\ &= (-2 \sin \phi \cos \phi) \left(\frac{-1}{R \sin \phi}\right) = \frac{2 \cos \phi}{R} \quad (28) \end{aligned}$$

Hence

$$A_{S_i} = \frac{V_0}{S_0} \left(\frac{2 \cos \phi_i}{R}\right) \quad (29)$$

Although both expressions of atomic surface accessibility, namely eqs 26 and 29, are equivalent, the first is more elegant in the sense that no physical parameters are needed. Furthermore, the deduction of eq 29 implies the assumption for  $\phi$  to be a continuous variable, what is not real since, as pointed above,  $\phi$  only takes discrete values. Either from eq 26 or 29, it becomes clear that for  $\phi = 0$ , their values become maximum ( $A_{S_1} = A_{S_2} = 1$ ), and consequently the atomic accessible surface area is also maximum, whereas for  $\phi = 90^\circ$  the accessibility (and the accessible surface area as well) is minimum ( $A_{S_1} = 0.5$  and  $A_{S_2} = 0$ ).



**Figure 3.** Plot of valence accessibility  $(1-\delta/4)$  versus  $(1/\delta)^{0.5}$ . Observe the hyperbolic dependence.

Moreover, two similar definitions of molecular surface accessibility arise from eqs 26 and 29:

$$A_{S_1} = \sum_{i=1}^n (\cos \phi_i) \quad (30)$$

$$A_{S_2} = \sum_{i=1}^n (1 + \cos \phi_i) \quad (31)$$

Since  $\cos \phi$  is evaluating atomic surface area, both, (30) and (31) must be a good measure of the molecular accessible surface area.

Table 2 illustrates the outputs of surface accessibility from eqs 30 and 31.

Although molecular surface area,  $S$ , for any alkane can be readily calculated either from  $A_{S_1}$  or  $A_{S_2}$ , our results indicate that  $A_{S_1}$  allows for the best prediction according to

$$S = 4\pi R^2 \sum_i (1 + A_{S_i}) \quad (32)$$

Table 3 (two last columns) shows the comparison between the “experimental” molecular surface area, as calculated from HiperChem 7.0 (Chemplus Software), and the surface area calculated from eq 32.

On the other hand,  $(\delta_i)^{-0.5}$ , what is the basic connectivity input, may be expressed as a function of  $\phi$ :

$$(\delta_i)^{-0.5} = \frac{1}{2\sin \phi_i} \quad (33)$$

From that it is immediate for connectivity indices to be outlined as

$${}^m\chi_i = \sum_{i=1}^{n_m} \prod_i (2\sin \phi) \quad (34)$$

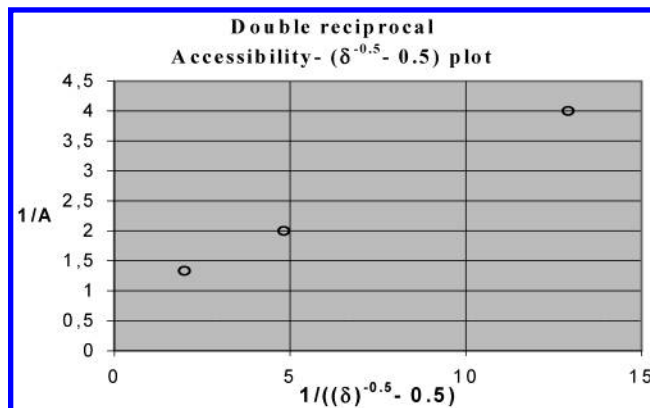
where  $n_m$  is the number of subgraphs of the type  $t$  with  $m$  edges.

Figure 3 shows the plot of  $A_\delta$  versus  $(\delta_i)^{-0.5}$  for four of the five possible  $\delta$  values. Since when  $\delta = 0$ ,  $(\delta_i)^{-0.5}$  tends to infinite and  $A_\delta$  tends to 1, then the dependence is two-constants hyperbolic.

Therefore, a double-reciprocal relation fits well between  $\phi = 30^\circ$  and  $\phi = 90^\circ$ , according to

$$\frac{1}{A_\delta} = \frac{0.244}{(\delta_i)^{-0.5} - 0.5} + 0.845 \quad (35)$$

(See Figure 4).



**Figure 4.** Double-reciprocal plot of accessibility,  $A_\delta$ , versus  $(1/\delta)^{0.5}$ . Observe the linear dependence of the three determinate points.

Several conclusions can be set up from the results in Table 3 and Figures 3 and 4:

1. Both  $(\delta_i)^{-0.5}$  as well as  $(1-\delta/4)$  are a good measure of volume accessibility.
2. Both can be expressed a function of the angle  $\phi$  as  $1/[2 \sin \phi]$  and  $\cos^2 \phi$ , respectively.
3.  $(1-\delta/4)$  is a better descriptor than  $(\delta_i)^{-0.5}$  as long as the isolated carbon (methane) can be described without indetermination. Furthermore,  $(1-\delta/4)$  is more intuitive and it changes linearly between 0 and 1.
4. Since  $\phi$  is a measure of the covalent radii and the covalent volume, both,  $(1-\delta/4)$  and  $(\delta_i)^{-0.5}$  are also a measure of these physicochemical parameters.
5. New topological indices may be derived from the angle  $\phi$  according to eqs 30 and 31. Such indices are a measure of molecular surface area.

## 7. THE MEANING OF THE CONNECTIVITY INDICES

Considering that a precise meaning for the basic connectivity input— $(\delta_i)^{-0.5}$ —has been outlined and, by the way, new definitions of molecular accessibility have been raised, we are now in a better position to provide a physical/physicochemical meaning for the  $\chi$  terms. Perhaps the best way to visualize that meaning is taking an overview in Table 2, whose last columns include the values of  ${}^0\chi$ ,  ${}^1\chi$ ,  ${}^2\chi$ , and  ${}^3\chi_p$  as well as the values of molecular volume and surface area for the set of alkanes.

Moreover, Table 4 illustrates the results of the regression equations obtained by multilinear regression analysis between volume, surface area, and volume/surface area ratio versus the corresponding values of accessibility's and connectivity indices from which we were found to be depending on.

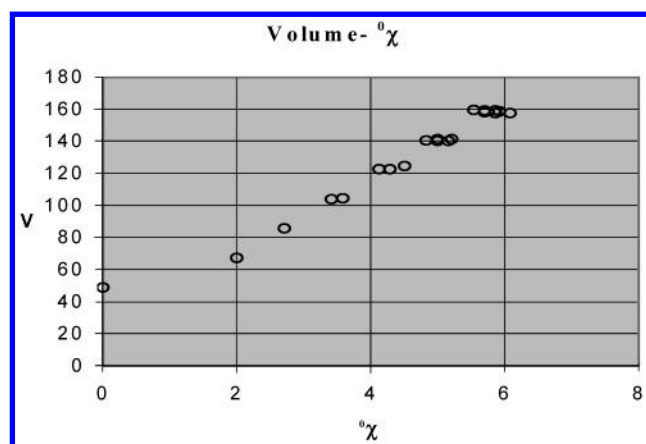
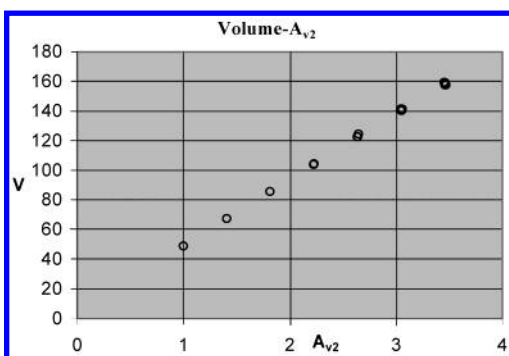
From the results so far outlined, several points stand:

1. Since there is a good correlation between volume accessibility and  $(\delta_i)^{-0.5}$ , which is the basic input for  ${}^0\chi$ , it is to be expected for  ${}^0\chi$  to be a measure of molecular volume. Indeed, from Table 2 it becomes clear that  ${}^0\chi$  is a good measure of molecular volume. It is interesting that molecular volume increases with the branching degree in the same way as  ${}^0\chi$  does; however, the successive definitions of volume accessibility given in this work allow for a better prediction of molecular volume than  ${}^0\chi$ . Just for comparative purposes, Figures 5 and 6 illustrate the plots of molecular volume versus  ${}^0\chi$  and  $A_{v2}$ , respectively. It is noteworthy that the

**Table 4.** Regression Equations for Molecular Volume, Molecular Surface Area, and Volume/Surface Area Ratio<sup>a</sup>

property	regression equation	N	R	S	F
molecular volume	$V = -12.4102 + 62.0990A_\delta$ (36)	21	0.99953	0.7425	37203
molecular volume	$V = 13.1554 + 36.4120A_{V_1}$ (37)	21	0.99972	0.7177	40690
molecular volume	$V = 5.4388 + 44.3279A_{V_2}$ (38)	21	0.99985	0.7007	42695
molecular volume	$V = 35.8930 + 20.7715^0\chi$ (30)	21	0.98652	5.4394	690.43
molecular volume	$V = 23.2094 + 23.2683^0\chi$ (40)	20	0.99219	3.4715	1138
molecular volume	$V = 11.5134 + 37.4001A_\delta - 1.5850^3\chi_p$ (41)	21	0.99992	0.5278	37718
molecular surface area	$S = 50.9288 + 24.3270A_{S_1}$ (42)	21	0.99332	3.4914	1408
molecular surface area	$S = 66.5863 + 33.7143^1\chi$ (43)	21	0.99319	3.5498	1362
volume/surface area ratio	$\frac{V}{S} = 0.7238 + 0.071^2\chi$ (44)	21	0.93766	0.0258	138
volume/surface area ratio	$\frac{V}{S} = 0.7229 + 0.059^2\chi + 0.0028^3\chi_p$ (45)	21	0.95431	0.0228	92

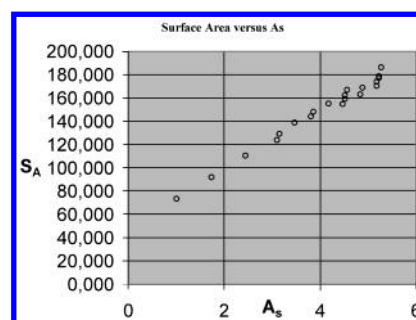
<sup>a</sup>  $V$  = molecular volume ( $\text{\AA}^3$ ).  $S$  = molecular surface area ( $\text{\AA}^2$ ).  $A_\delta$  = valence accessibility.  $A_{V_1}$  = first volume accessibility.  $A_{V_2}$  = second volume accessibility.  $A_{S_1}$  = surface accessibility (eq 30).  $^0\chi, ^1\chi, ^2\chi, ^3\chi_p$  = molecular connectivity terms of the type *path* and orders 0, 1, 2, and 3, respectively.

**Figure 5.** Plot of molecular volume,  $V$ , versus  $^0\chi$ . Observe the degeneracy for those entries showing the same  $\delta$  values as well as methane's *outlier* character.**Figure 6.** Plot of molecular volume,  $V$ , versus  $A_{V_2}$ . Observe that degeneracy has disappeared.

degeneracy outstanding in  $^0\chi$  disappears when using  $A_{V_2}$ . Furthermore, the regression equations in Table 4 indicate a better result for  $A_{V_2}$  than  $^0\chi$ . Compare eqs 38 and 39. Moreover, since methane is a clear *outlier* in the  $^0\chi$  plot, the regression equation just taking out methane was achieved (eq 40), and it still continues showing a worse predictive capability than  $A_{V_2}$ .

Finally, eq 41 shows the best 2-variable equation for molecular volume; it was achieved by combining  $A_\delta$  with  $^3\chi_p$ .

2. Contrary to volume, molecular surface area decreases as branching degree increases, in the same way as  $A_S$  and  $^1\chi$  do. Therefore,  $^1\chi$  is just a measure of molecular surface area. As for volume, the prediction of molecular surface area is better using the novel  $A_S$  index than using the  $^1\chi$ . Figure 7

**Figure 7.** Plot of molecular surface area,  $S_A$ , versus surface accessibility,  $A_S$ .

shows the plot of molecular surface area versus  $A_S$ , while eqs 42 and 43 at Table 4 illustrate the results of the regressions of molecular surface area versus  $A_S$  and  $^1\chi$ , respectively. Despite the lack of selectivity of the  $A_S$  index (which is calculated for each vertex in the graph, i.e., for each atom) as compared with  $^1\chi$  (which is obtained from a couple of adjacent vertices) the first allows for a better prediction of molecular surface area, as deduced from the correlation coefficients (see Table 4 again).

3. Although not shown here, it is to be expected that both the prediction of molecular volume and molecular surface area for alkanes from our novel approach based on the angle  $\phi$  should improve by using the refined definitions of accessibility given by eqs 15 and 20 instead of the intuitive definition in eq 17.

4. The higher order connectivity indices,  $^2\chi$  and  $^3\chi_p$ , encode information over the volume/surface area ratios. This results are coherent with the interpretation provided for  $^0\chi$  and  $^1\chi$ . Indeed, Equations 44 and 45 illustrate the results of the *volume/surface area* ratios for the set of alkanes analyzed here. The best result for the 1-variable equations was for  $^2\chi$ , whereas the joint use of  $^2\chi$  and  $^3\chi_p$  allowed the best two-variable equation.

That means that, accordingly with  $^0\chi$  and  $^1\chi$ 's significance,  $^2\chi$  and  $^3\chi_p$  are a measure of the volume/surface area ratio. At the present this conclusion is established as a matter of fact, but further insight should be performed in the future.

5. From Table 3, it makes clear the good fitting between the *standard* and calculated by our formalism values of volume and surface area, although in some cases there is degeneracy (compare for instance 23MMPentane and 24MMPentane). This is readily explainable because of the nature of the indices used here, such as  $A_{V_2}$  and  $A_S$ , which are related to the single-atomic-valence degree. Of course



this degeneracy will be resolved in the future as long as higher order  $A_V$  and  $A_S$  indices are input, but anyway, this fact does not diminish the interest of the present results.

6. Since many important physicochemical properties, such as density, molar refractivity, chemical reactivity, catalytic effects, etc., are depending either on molecular volume, molecular surface area, or *volume/surface area* ratios, we may conclude that both, the novel indices introduced here as well as the interpretation of the  $\chi$  terms, are to be a significant insight in the prediction and interpretation of many physicochemical experimental properties.

From among the many possible examples stands the Kier's result on alkane density:<sup>24</sup>

$$\rho = 0.7348 + 0.003^3 \chi_p - \frac{0.2927}{^1\chi} \quad (46)$$

The  $^3\chi_p$  index usually appears when correlating density.<sup>25</sup> Increment of both,  $^1\chi$  and  $^3\chi_p$ , leads to an increment of density. The first index encodes information about the molecular surface area, while the second is a measure of the surface area/volume ratio. Indeed, it is to expect a higher density of compounds with higher  $^1\chi$  and  $^3\chi_p$  values.

Of course further examples must be added in the future so that the interpretation outlined here can provided further insight into the meaning of topological indices.

## 8. CONCLUSIONS

In this work we have demonstrated that connectivity indices  $^0\chi$ ,  $^1\chi$ ,  $^2\chi$ , and  $^3\chi_p$  can be interpreted as a measure of molecular volume, surface area, and volume/surface area ratios, respectively. At the same time, a pure topological notion, accessibility, is clearly defined upon an intuitive and simple basis following an Estrada-like approach and it allows the aforementioned interpretation. Furthermore, this makes possible a straightforward prediction and interpretation of the many and important molecular properties related to these physicochemical features. Nevertheless, there is one point to be emphasized: Graph invariants as well as topological indices are pure mathematical objects, so that they do not have really a physical meaning, in the same sense as for example, Pythagoras' Theorem does not store any physical *meaning* but only one (or many) physical *applications*. Hence, although many key physical and geometrical parameters are depending on topological features, the reciprocal does not work, i.e., topological indices do not depend on physical or geometrical parameters and, consequently, *they do not have a physical or geometrical meaning*. The

importance of this assertion is that we can express many experimental or "real" properties of molecules as direct functions of topological indices, so bypassing the time-consuming and bothersome calculations of physical and geometrical parameters, which is necessary under other formalisms, particularly for those based on Quantum Mechanics.

## ACKNOWLEDGMENT

The author acknowledges Generalitat Valenciana (GV99-91-1-12) and the Spanish Ministry of Science and Technology (SAF2000-0223-C03-02) for funding support.

## REFERENCES AND NOTES

- (1) Sylvester, J. J. *Am. J. Math.* **1874**, *1*, 64.
- (2) Sylvester, J. J. *Am. J. Math.* **1878**, *1*, 65.
- (3) Wiener, H. *J. Am. Soc.* **1947**, 69.
- (4) Balaban, A. T.; Motoc, L.; Bonchev, D.; Mekenyan, O. *Top. Curr. Chem.* **1983**, *114*, 21–25.
- (5) Randić, M. *J. Am. Chem. Soc.* **1975**, *97*, 6609.
- (6) Kier, L. B.; Hall, L. H.; Murray, W. J.; Randić, M. *J. Pharm. Sci.* **1975**, *64*, 1971.
- (7) Bonchev, D.; Trinajstić, N. *Intern. J. Quantum Chem. Symp.* **1978**, *12*, 293–303.
- (8) Basak, S. C.; Gute, B. D.; Grunwald, G. D. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 1054–1060.
- (9) Pogliani, L. *J. Phys. Chem.* **1993**, *97*, 9029.
- (10) Galvez, J.; Garcia, R.; Julian-Ortiz, J. V. de; Soler, R. Topological approach to drug design. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 272–284.
- (11) Galvez, J.; Garcia-Domenech, R.; De Gregorio, C.; De Julian-Ortiz, J. V.; Popa, L. *J. Mol. Graphics* **1996**, *14*, 272–276.
- (12) García-Domenech, R.; García-March, F. J.; Soler, R.; Gálvez, J.; Antón-Fos, G. M.; de Julián-Ortiz, J. V. *Quant. Struct.-Act. Relat.* **1996**, *15*, 201–207.
- (13) Julian-Ortiz, J. V. de; Galvez, J.; Muñoz-Collado, C.; Garcia-Domenech, R.; Jimeno-Cardona, C. *J. Med. Chem.* **1999**, *42*, 3308–3314.
- (14) Stankevich, I. V.; Skovortsova, M. I.; Zefirov, N. S. *J. Mol. Struct. (THEOCHEM)* **1995**, *342*, 173–179.
- (15) Gálvez, J. *J. Mol. Struct. (THEOCHEM)* **1998**, *429*, 255.
- (16) Kier, L. B.; Hall, L. H. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 792.
- (17) Estrada, E. *J. Phys. Chem. A* **2002**, *106*, 9085.
- (18) Kier, L. B.; Hall, L. H. *Molecular Connectivity in Structure-Activity Analysis*; Research Studies Press: Letchworth, England, 1986; pp 225–246.
- (19) Kier, L. B.; Hall, L. H. *J. Pharm. Sci.* **1976**, *65*, 1806.
- (20) Estrada, E. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 1022.
- (21) Amic, D.; Beslo, D.; Lucic, B.; Nilolic, S.; Trinajstić, N. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 819.
- (22) Bondi, A. *J. Phys. Chem.* **1964**, *68*, 441.
- (23) Kier, L. B.; Hall, L. H. *J. Pharm. Sci.* **1983**, *72*, 1170.
- (24) Kier, L. B.; Hall, L. H. *Molecular Connectivity in Chemistry and Drug Research*; Academic Press: New York, 1976; p 221.
- (25) Rubio, E. *Determinación de algunas propiedades físicoquímicas de grupos heterogéneos de moléculas orgánicas por el método de conectividad molecular*. Degree Thesis. University of Valencia, 1986; p 193. (Note: This reference is at the reader's disposal under request).

CI034036X