Complexity of Molecules[†]

Sonja Nikolić* and Nenad Trinajstić

The Rugjer Bošković Institute, P.O. Box 180, HR-10002 Zagreb, Croatia

Iva Marija Tolić[‡]

Harvard School of Public Health−Physiology Program, Harvard University, 665 Huntington Avenue,
Boston, Massachusetts 02115

Several currently used measures of the complexity of molecules, such as, for example, those by Bertz and Randić or those based on the number of spanning trees, are briefly reviewed. We also proposed as complexity measures the sum of vertex-weights and the sum of edge-weights and their variants related to partition of vertex-weights and edge-weights into classes by their numerical values. The vertex-weights considered are the squares of the vertex-degrees, and the edge-weights are products of vertex-degrees making up the edges. Comparison is made between considered complexity measures for selected molecular graphs. All considered indices increase with increasing size and cyclicity and most with increasing branching. However, they differ regarding the influence of symmetry.

1. INTRODUCTION

The concept of complexity despite being a difficult concept to define[1] is receiving considerable attention, not only in science[2,3] but also in contemporary science fiction.[4] Since the term complexity of molecules and chemical reactions is used in the everyday chemical jargon, it is quite understandable that attempts to quantify the concept of molecular complexity have been the topic of research in chemistry for many years.[5,6] However, one should be aware that there are different kinds of molecular complexity, e.g., elemental complexity, structural complexity, dynamic complexity, etc.[7]

In the present paper our attention will be centered on the two-dimensional or topological complexity. There are many topological complexity measures proposed, and their uses described, in the chemical literature.[8−21] In the next section we will review a few of these and then will propose our approach to two-dimensional molecular complexity. Since the natural tool for studying the two-dimensional complexity of molecules appears to be graph theory, we will use in this paper the language of (chemical) graph theory.[22] Topological complexity of a molecule is characterized by its size (in terms of either the number of atoms or bonds), branching, cyclicity, etc. Symmetry is also an important molecular characteristic and diminishes the structural complexity of a molecule.

2. SOME APPROACHES TO THE COMPLEXITY OF MOLECULES

Bonchev-Trinajstić Approach. The Bonchev-Trinajstić approach was used for studying the concept of molecular branching:[23]

$$BT = n \log_2 n - \sum_l n_l \log_2 n_l \qquad (1)$$

where BT stands for the Bonchev−Trinajstić index. This formula was designed to be used in conjuction with the distance matrix of a molecular graph. If only distances in a molecular graph are considered, then n = the total number of distances and n_l = the number of distances of length l. Note that $n = \sum n_l$. Equation 1 is related to the Mowshowitz equation for measuring the relative complexity of graphs:[24]

$$M = -\sum p_i \log_2 p_i \qquad (2)$$

where $p_i = n_i/n$ is the probability that a given vertex is in the ith class of symmetry-equivalence of vertices. Equations 1 and 2 are based on the Shannon formula[25] for the information content of a system. Rashevsky[26] was the first to apply the Shannon formula to graphs.

Bertz Approach. The Bertz approach is based on the equation[6,8,9,17,27−30]

$$BI = 2n \log_2 n - \sum n_i \log_2 n_i \qquad (3)$$

where BI stands for the Bertz complexity index and n for a selected graph invariant. Bertz found that a good invariant to use is the number of pairs of adjacent edges in a molecular graph G. Thus, n is the total number of pairs of adjacent edges, and n_i is the number of pairs of adjacent edges in the ith class by symmetry. The number of pairs of adjacent edges can be, for example, obtained by counting the edges in the line graph (called by Bertz the bond graph) generated from G.[31,32] It should be noted that the number of pairs of adjacent edges in G is equal to the number of ways the graph representing the carbon skeleton of propane can be cut out of G.

COMPLEXITY OF MOLECULES

J. Chem. Inf. Comput. Sci., Vol. 40, No. 4, 2000 **921**

Equation 3 represents a pragmatic modification of eq 1. Bertz added to eq 1 the term $n \log_2 n$ to prevent the result BI = 0 when all pairs of adjacent edges in a graph are equivalent, as is the case for graphs depicting propane, isobutane, cycloalkanes, benzene, annulenes, etc. The first part of eq 3 reflects structural characteristics of G (size, branching, cyclicity, etc.), while the second part reflects the symmetry of G in terms of equivalent pairs of adjacent edges. Therefore, the Bertz index represents a balance between the structural and symmetry features of a molecular graph. Bertz introduced his approach for use in synthesis planning and analysis of synthetic strategies.[6,8,9,17,27−29]

Approach Based on Spanning Trees. Bonchev and co-workers[33−36] used the number of spanning trees to study the complexity of reaction mechanisms, and Mallion and co-workers used this number as a criterion for the complexity of (poly)cyclic graphs.[37−41] The complexity criterion based on spanning trees can be stated as follows: The most complex structure possesses the highest number of spanning trees. A spanning tree of a graph G is a connected acyclic subgraph containing all the vertexes of G. In the case of trees, the spanning tree is identical to the tree itself. Therefore, the number of spanning trees can only be used as a measure of complexity for monocyclic and polycyclic molecular graphs. A lot of work on the enumeration of spanning trees for various classes of (molecular) graphs has been carried out in the past 25 years.[37−45] The interest in spanning trees is related to several chemical problems, such as the use of spanning trees in chemical kinetics,[33−36] in calculating the magnetic properties of conjugated systems by means of the ring−current model[46−49] in the framework of π-electron molecular orbital theory,[50] and in coding and enumerating polyhex hydrocarbons.[51]

The number of spanning trees can be obtained in several ways.[15] The easiest way is perhaps to use the Laplacian matrix of a graph.[52−54] The Laplacian matrix $\mathbf{L} = \mathbf{L}(G)$ of a graph G is defined as

$$\mathbf{L} = \mathbf{V} - \mathbf{A} \tag{4}$$

where $\mathbf{V}$ is a diagonal matrix whose entries are the vertex-degrees and $\mathbf{A}$ is the adjacency matrix of G. The diagonalization of the Laplacian matrix of G with V vertexes produces V real eigenvalues. The set of Laplacian eigenvalues is usually referred to as the Laplacian spectrum and denoted by $x_i(\mathbf{L})$. The smallest member of the Laplacian spectrum is always zero. This is a consequence of the special structure of the Laplacian matrix.

The number of spanning trees $t(G)$ of G is given by[52]

$$t(G) = (1/V) \prod_{i=2}^{V} x_i(\mathbf{L}) \tag{5}$$

It should be noted that the number of spanning trees is a rather explosive quantity even for not too large graphs. For example, the modestly sized Blanuša graph has 1037136 spanning trees.[15]

Approach Based on the Subgraphs of a Molecular Graph. Bertz et al.[6,17] introduced an approach based on the subgraphs[55] of a molecular graph. They proposed using as the complexity indices the number of kinds of connected subgraphs $N_s(G) = N_s$ and the total number of connected subgraphs $N_t(G) = N_t$ of G. Connected subgraphs are a single vertex (representing the carbon atom of methane), two vertexes connected by an edge (representing the carbon skeleton of ethane), three vertexes connected either as a chain (representing the carbon skeleton of propane) or a three-membered cycle (representing the carbon skeleton of cyclopropane), four vertexes connected either as a chain (representing the carbon skeleton of butane) or a star (representing the carbon skeleton of isobutane), etc. The graph G itself is counted as a subgraph, because in the formal graph theory[55] G is its own subgraph. Bonchev[7,16] also proposed in the same year (1997) as Bertz and Sommer[17] the number of all connected subgraphs as a measure of molecular complexity.

It should be noted that the number of methane subgraphs (N_m) of G is equal to the number of vertexes (V) in G: $N_m = V$. Similarly, the number of ethane subgraphs (N_e) is equal to the number of edges (E) in G: $N_e = E$. A chainlike structure of n-alkane trees allows one to derive formulas for obtaining their N_t and N_s numbers: $N_t = V(V + 1)/2$ and $N_s = V$. An unbranched cyclic structure of cycloalkane graphs leads straightforwardly to $N_t = V^2 + 1$ and $N_s = V + 1$.

It appears that a rather difficult problem of calculating the number of all connected subgraphs is solved. Rücker and Rücker[56] recently designed an efficient computer program for generating and counting all connected subgraphs of a connected undirected simple graph.

Randić Approach. The Randić approach[20] is based on the concept of the augmented vertex-degree.[57] The augmented degree of a given vertex is obtained by adding to its degree a half of the degrees of the nearest neighbors, a quarter of the degrees of the next nearest neighbors, and so on. The Randić complexity index (RCI) is then given by the sum of the augmented vertex-degrees for symmetry nonequivalent vertexes.

Computations using the Randić approach can be carried out by means of the augmented degree matrix $\mathbf{D}'(G) = \mathbf{D}'$ with elements:

$$(\mathbf{D}')_{ij} = \frac{d_j}{2^{l_{ij}}} \tag{6}$$

where d_i is the vertex-degree and $l_{ij} = $ the distance between the vertex i and neighbors j. The augmented degree matrix is a nonsymmetric matrix.

3. ZAGREB INDICES AS POSSIBLE COMPLEXITY MEASURES

Already Bertz and Wright[6] noticed that the so-called Zagreb M_1 index[58,59] has the characteristics of a complexity index. However, we have also independently investigated whether or not M_1 and M_2 indices can serve as the complexity measures.[21] The M_1 and M_2 indices are defined as follows:

$$M_1 = \sum_{vertices} d_i^2 \tag{7}$$

$$M_2 = \sum_{edges} d_i d_j \tag{8}$$

where d_i is the degree of vertex i.

Both M_1 and M_2 indices increase with molecular size. This can be seen from the analytical formulas for (i) n-alkanes, $M_1 = 4(V - 2) + 2$, $M_2 = 4(V - 2)$ for $n > 2$; (ii)

Table 1. Complexity Indices for Lower *n*-Alkanes

indices[a]					for given *n*-alkane tree				
BT	0	2.75	8.76	18.47	32.24	50.37	73.09	100.60	133.08
BI		0	2.00	7.51	12.00	19.22	25.05	33.30	40.00
RCI	1.50	5.25	6.38	11.44	12.10	17.67	18.02	23.82	24.01
N_s	2	3	4	5	6	7	8	9	10
N_t	3	6	10	15	21	28	36	45	55
M_1	2	6	10	14	18	22	26	30	34
M'_1	1.5	5.5	7.5	8.83	9.83	10.63	11.30	11.87	12.37
M_2	1	4	8	12	16	20	24	28	32
M'_2	1	3	7	9	10.33	11.33	12.13	12.80	13.37

[a] Symbols are explained in the text.

Table 2. Isomeric Hexane Trees and Their Complexity Indices

	for given hexane tree				
indices[a]	**A**	**B**	**C**	**D**	**E**
BT	32.24	28.63	27.38	23.48	22.59
BI	12.00	21.22	19.22	21.02	29.79
RCI	12.10	21.94	17.82	9.75	19.50
N_s	6	8	8	7	8
N_t	21	24	25	28	30
M_1	18	20	20	22	24
M'_1	9.83	16.83	16.83	15.58	22.08
M_2	16	18	19	21	22
M'_2	10.33	16.5	15.0	15.25	17.33

[a] Symbols are explained in the text.

cycloalkanes, $M_1 = M_2 = 4V$; (iii) polyacenes, $M_1 = 26R - 2$, $M_2 = 33R - 9$; and (iv) polyphenanthrenes, M_1(polyphenanthrenes) $= M_1$(polyacenes), $M_2 = 34R - 11$, where $V =$ the number of carbon atoms (vertexes) in an *n*-alkane or cycloalkane and $R =$ the number of hexagons in a polyacene or polyphenanthrene. The equality $M_1 = M_2$ for cycloalkanes is the consequence of the fact that in these molecules the number of carbon atoms is equal to the number of carbon bonds. The equality M_1(polyacenes) $= M_1$(polyphenanthrenes) is the consequence of the definition of the M_1 index.

The computation of the M_1 and M_2 indices is shown in Figure 1 for the 2,3-dimethylpentane tree.

Both the M_1 and M_2 indices as defined above depend strongly on the molecular size in terms of the number of atoms and bonds, respectively, and somewhat less on branching. We also introduced variants of the M_1 and M_2 indices, denoted by M_1' and M_2', by partitioning the vertex-weights d_i^2 and edge-weights $d_i d_j$ into classes by their numerical values. The population of each class k is denoted by the number n_k. Then each member of a class is weighted in the following way: The first member is divided by one, the second by two, the third by three, the fourth by four, etc. In the example under the consideration: $M_1' = 1/1 + 1/2 + 1/3 + 1/4 + 9/1 + 9/2 = 15.58$; $M_2' = 3/1 + 3/2 + 3/3 + 3/4 + 9/1 = 15.25$.

The M_1' and M_2' indices for a given molecular graph can be computed using the following expressions:

$$M'_1 = \sum_{k \in vw} k \sum_{i=1}^{n_k} 1/i \qquad (9)$$

$$M'_1 = \sum_{k \in ew} k \sum_{i=1}^{n_k} 1/i \qquad (10)$$

where $vw = \{1, 4, 9, 16\}$ is the set of possible vertex-weights and $ew = \{2, 3, 4, 4, 6, 9, 12, 16\}$ is the set of possible edge-weights. The edge-weight $4 = 1 \times 4$ is to be differentiated from the edge-weight $4 = 2 \times 2$.

The M_1' and M_2' indices can be given in a closed form for homologous series, but this time formulas are much more complicated than in the case of the original M_1 and M_2 indices.

(i) *n*-alkanes

$$M'_1 = 1.5 + 4 \sum_{i=1}^{V-2} 1/i \quad \text{(for } V \geq 2) \qquad (11)$$

$$M'_2 = 3 + 4 \sum_{i=1}^{V-3} 1/i \quad \text{(for } V \geq 3) \qquad (12)$$

(ii) cycloalkanes

$$M'_1 = M'_2 = 4 \sum_{i=1}^{V} 1/i \qquad (13)$$

(iii) polyacenes

$$M'_1 = 4 \sum_{i=1}^{2(R+2)} 1/i + 9 \sum_{i=1}^{2(R-1)} 1/i \qquad (14)$$

$$M'_2 = 49/5 + 6 \sum_{i=1}^{4(R-1)} 1/i + 9 \sum_{i=1}^{R-1} 1/i \qquad (15)$$
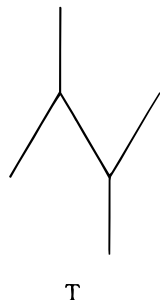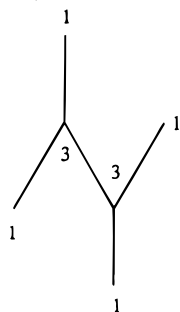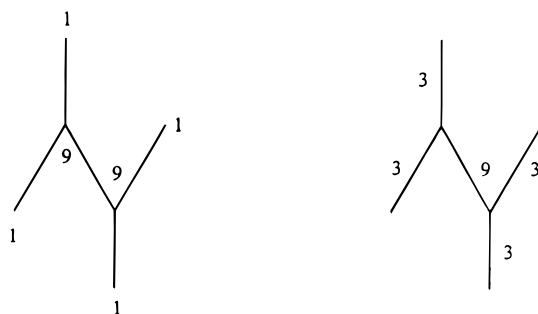
(iv) polyphenanthrenes

$$M'_1(\text{polyphenanthrenes}) = M'_1(\text{polyacenes}) \qquad (16)$$

$$M'_2 = 4 \sum_{i=1}^{R+4} 1/i + 6 \sum_{i=1}^{2R} 1/i + 9 \sum_{i=1}^{2R-3} 1/i \qquad (17)$$

4. COMPARISON BETWEEN REVIEWED COMPLEXITY MEASURES ON SELECTED MOLECULES

We illustrate the application of reviewed complexity indices on several sets of molecular graphs. In Table 1 we give *n*-alkane-trees up to $n = 10$ and their complexity indices.

(i)   2,3-dimethylpentane tree T

(ii)   Vertex-degrees d_i

(iii)   Vertex-weights d_i^2 and edge-weights $d_i d_j$

(iv)   Zagreb indices of T

$M_1 = 22$

$M_2 = 21$

Figure 1. Computing the M_1 and M_2 indices for the 2,3-dimethylpentane tree.

A proposed criterion for the complexity of a homologous series of molecules is that a higher member is expected to be more complex than a lower member of a series.[28] This criterion is fulfilled by all considered complexity indices since they all increase with increasing size of n-alkanes.

The BT index increases because the total number of distances increases. The Bertz index increases with the size of the n-alkane tree because the number of pairs of adjacent edges increases. The Randić index increases in groups of two (propane and butane, pentane and hexane, heptane and octane, etc.) reflecting the number of vertexes in different environments. For example, in propane there are two end-vertexes (one type of vertexes) and one central-vertex (the second type of vertexes). In butane there are again two end-vertexes and two central-vertexes. Thus, propane and butane possess the Randić indices which differ only slightly. The same is encountered for the duo of pentane and hexane; their indices also differ slightly. However, there is a considerable jump in the value of their indices in comparison with the values of the preceding duo of propane and butane. This is so because the duo of pentane and hexane possesses three different types of vertexes and hence the jump in the value of their indices.

The N_s and N_t numbers for n-alkane trees steadily increase with the number of vertexes (see above formulas for computing these numbers for n-alkane-trees). Each n-alkane tree in the series has the N_t number, which is made up from the N_t number of the preceding n-alkane and its own size in terms of the number of vertexes. For example, the N_t number of nonane (45) is equal to N_t(octane) + 9.

The M_1 number increases with the size of the n-alkane in steps of 4. Thus, $M_1(V\text{-alkane}) = M_1((V-1)\text{-alkane}) + 4$. The M_2 number also increases with size in steps of 4; that is, $M_2(V\text{-alkane}) = M_2((V-1)\text{-alkane}) + 4$ (for $V > 2$).

The M'_1 and M'_2 indices also increase with size, but the difference between the M'_1 and M'_2 numbers for higher n-alkane trees is becoming smaller and smaller, yet they do not converge. This is, of course, the consequence of the forms of eqs 11 and 12.

Next we give in Table 2 five isomeric hexane trees and their complexity indices.

Isomeric hexane trees exhibit two structural characteristics: branching[60] and symmetry.[61] The balance between branching and symmetry determines the complexity order of hexane trees. In the following we list the ordering of hexane trees as given by considered complexity indices:

$-$BT	A < B < C < D < E
BI	A < C < D < B < E
RCI	D < A < C < E < B
N_t	A < B < C < D < E
N_s	A < D < B = C = E
M_1	A < B = C < D < E
M'_1	A < D < B = C < E
M_2	A < B < C < D < E
M'_2	A < C < D < B < E

All indices, but the RCI index, agree that the least complex is n-hexane (**A**) and the most complex 2,2-dimethylbutane (**E**). There are various orderings of **B**, **C**, and **D** obtained, but the ordering **B** < **C** < **D** is supported by several indices.

Complexity ordering (which is given by $-$BT, N_t, M_2): **A** < **B** < **C** < **D** < **E** can also be obtained by the recently proposed index called total walk count.[62] If the total walk count is symmetry-modified, the so-called symmetry-modified walk count index, it produces the following complexity ordering: **A** < **D** < **C** < **B** < **E**, which differs from the ordering given by the total walk count since the positions of

Table 3. Complexity Indices for Lower Cycloalkanes

	for given cycloalkane							
indices[a]	△	▢	⬠	⬡	◯	◯	◯	◯
BT	0	5.51	10.00	22.83	33.29	54.61	72.00	102.64
BI	4.75	8.00	11.61	15.51	19.65	24.00	28.53	33.22
RCI	4.00	4.50	5.00	5.25	5.50	5.63	5.75	5.81
$t(G)$	3	4	5	6	7	8	9	10
N_s	4	5	6	7	8	9	10	11
N_t	10	17	26	37	50	65	82	101
$M_1 = M_2$	12	16	20	24	28	32	36	40
$M'_1 = M'_2$	7.33	8.33	9.13	9.80	10.37	10.87	11.32	11.71

[a] Symbols are explained in the text.

B and **C** are inverted in it. No one among the considered indices matches this order.

In Table 3 we give the complexity indices for lower cycloalkanes.

All complexity indices given in Table 3 increase with the size of cycloalkanes.

The Bertz index increases because the number of pairs of adjacent edges increases. The comparison between BI-(cycloalkanes) and BI(n-alkanes) shows that after a certain size ($V = 8$) the value of the BI(cycloalkane) becomes smaller than the value of BI for the corresponding n-alkane. This is so because, for the same size in terms of the number of vertexes, providing that $V \geq 8$, the first part of the formula for the BI index of n-alkanes ($2n \log_2 n$) becomes much bigger than $n \log_2 n$ used for computing the BI index of cycloalkanes and the correction for symmetry $\sum n_i \log_2 n_i$ influences less and less the BI index of larger n-alkanes.

The BT index increases because the total number of distances also increases. However, it does not increase to the same magnitude as that in the corresponding n-alkane since the longest path in an n-alkane vanishes when it closes into the related cycloalkane.

The Randić index increases rather slowly because cycles possess only one symmetry-nonequivalent vertex since all vertexes in a cycle share the same environment; that is, every vertex in a cycle has an identical augmented degree. Because of that, the Randić index for a given cycloalkane is always smaller than that for the corresponding n-alkane. This can be easily seen if we compare the Randić indices for hexane and cyclohexane. While hexane has three symmetry-non-equivalent atoms,

$$\text{RCI(hexane)} = 5 + {}^9/_2 + {}^7/_4 + {}^5/_8 + {}^3/_{16} + {}^1/_{32} = 12.09$$

cyclohexane possesses only one symmetry-nonequivalent atom,

$$\text{RCI(cyclohexane)} = 2 + {}^4/_2 + {}^4/_4 + {}^2/_8 = 5.25$$

The number of spanning trees $t(G)$ and the N_s and N_t numbers increase with the sizes of the cycloalkanes in terms of either the number of vertexes (V) or the number of edges ($E = V$): $t(G) = V$ (or E), $N_s = V + 1$ (or $E + 1$), and $N_t = V^2 + 1$ (or $E^2 + 1$). The N_s and N_t numbers are always bigger for cycloalkanes than for the related n-alkanes: $N_s = V + 1$ (or $E + 1$) and $N_t = V^2 + 1$ (or $(E^2 + 1)$).

The M_1 and M_2 indices are identical for a cycloalkane and increase in units of 4; that is, $M_1(V\text{-cycloalkane}) = M_1((V - 1)\text{-cycloalkane}) + 4$. The difference between the M_1 and M_2 indices for n-alkanes and corresponding cycloalkanes is given by $M_1(V\text{-cycloalkane}) - M_1(V\text{-alkane}) = 6$ for all V and $M_2(V\text{-cycloalkane}) - M_2(V\text{-alkane}) = 8$ for $V \geq 3$.

The M'_1 and M'_2 indices are also identical for cycloalkanes and increase with increasing number of carbon atoms, but slower than in the case of n-alkanes. However, the M'_1 (and consequently M'_2) index for cycloalkanes also does not converge. This is, of course, the consequence of the form of eq 13.

In Table 4 we give all molecular graphs with four vertexes and their complexity indices.

In these graphs there are three structural characteristics present: branching, cyclicity,[63] and symmetry. The balance between these three structural features determines the complexity ordering of all connected molecular graphs with four vertexes. In the following we give orderings given by considered complexity indices:

$$\begin{array}{ll} -\text{BT} & \text{F} < \text{G} < \text{H} = \text{I} < \text{J} < \text{K} \\ \text{BI} & \text{F} < \text{G} < \text{I} < \text{H} < \text{J} < \text{K} \\ \text{RCI} & \text{I} < \text{F} < \text{G} = \text{K} < \text{J} < \text{H} \\ t(G) & \text{F} = \text{G} < \text{H} < \text{I} < \text{J} < \text{K} \\ N_s & \text{F} = \text{G} < \text{I} < \text{H} < \text{J} < \text{K} \\ N_t & \text{F} < \text{G} < \text{I} < \text{H} < \text{J} < \text{K} \\ M_1 & \text{F} < \text{G} < \text{I} < \text{H} < \text{J} < \text{K} \\ M'_1 & \text{F} < \text{I} < \text{G} < \text{H} < \text{K} < \text{J} \\ M_2 & \text{F} < \text{G} < \text{I} < \text{H} < \text{J} < \text{K} \\ M'_2 & \text{G} < \text{F} < \text{I} < \text{H} < \text{J} < \text{K} \end{array}$$

All indices, but RCI and M'$_2$, agree that the least complex graph is **F** (representing butane) and the most complex graph is **K** (representing tetrahedrane). There are various orderings of **G**, **H**, **I**, and **J**. The ordering **G** < **I** < **H** < **J** is supported by the following indices: BI, N_s, N_t, M_1, and M_2. This ordering is also supported by the total walk count index,[62] which also agrees with **F** and **K** being the least and most complex graphs, respectively. The related complexity index, the symmetry-modified walk count index, produces the

COMPLEXITY OF MOLECULES

J. Chem. Inf. Comput. Sci., Vol. 40, No. 4, 2000 **925**

Table 4. All Molecular Graphs with Four Vertices and Their Complexity Indices

indices[a]	for given graph					
	F	**G**	**H**	**I**	**J**	**K**
BT	8.76	6.01	5.51	5.51	3.90	0.00
BI	2.00	4.75	19.22	8.00	36.00	43.02
RCI	6.38	7.50	13.75	4.50	12.00	7.50
$t(G)$	1	1	3	4	8	16
N_s	4	4	7	5	9	10
N_t	10	11	18	17	33	64
M_1	10	12	18	16	26	36
M'_1	7.5	10.83	16	8.33	19.5	18.75
M_2	8	9	19	16	33	54
M'_2	7	5.5	16	8.33	21.5	22.05

[a] Symbols are explained in the text.

following complexity ordering: **I** < **F** < **G** < **K** < **H** < **J**. No one among the considered indices matches this ordering. However, the RCI ordering is similar.

5. CONCLUDING REMARKS

We compared several measures of the topological complexity for selected sets of molecular graphs. In the case of a homologous series all indices increase with size. Thus, they satisfy criterion according to which the higher members are more complex than the lower members of the homologous series of molecules. Most of the complexity indices also agree regarding branching by having larger values for more branched graphs. In the case of cyclicity all considered indices (the BT index with the sign minus) increase with the increasing number of cycles in the graph. However, authors differ in assessing the role of symmetry. Some authors (e.g., Bertz[8] and us) think that the structural characteristics should prevail over the molecular symmetry, but the symmetry must be included in the definition of the complexity index. Others (e.g., Randić)[20] are supporting the view that the symmetry should dominate over other structural features of a molecule. The third group of authors (e.g., Bonchev)[16] design complexity indices free of direct symmetry influence.

ACKNOWLEDGMENT

REFERENCES AND NOTES

(1) Waldrop, M. M. *Complexity*; Touchstone/Simon&Schuster: New York, 1992.

(2) Cowan, G. A., Pines, D., Meltzer, D., Eds. *Complexity, Models and Reality*; Addison-Wesley: Reading, MA, 1994.

(3) Badii, R.; Politi, A. *Complexity: Hierarchical Structures and Scaling in Physics*; Cambridge University Press: Cambridge, U.K., 1997.

(4) Crichton, M. *The Lost World*; Knopf: New York, 1995. This is a sequel to a bestselling novel of the same author: *Jurassic Park*; Knopf: New York, 1990.

(5) Bonchev, D.; Seitz, W. A. The Concept of Complexity in Chemistry. In: *Concepts in Chemistry-A Contemporary Challenge*; Rouvray, D. H., Ed.; Wiley: New York, 1997; pp 353−381.

(6) Bertz, S. H.; Wright, W. F. The Graph Theory Approach to Synthetic Analysis: Definition and Application of Molecular Complexity and Synthetic Complexity. *Graph Theory Notes New York* **1998**, *35*, 32−48.

(7) Bonchev, D. Overall Connectivity and Topological Complexity: A New Tool for QSPR/QSAR. In *Topological Indices and Related Descriptors in QSAR and QSPR*; Devillers, J., Balaban, A. T., Eds.; Gordon and Breach Amsterdam, The Netherlands, 1999; pp 361−401.

(8) Bertz, S. H. The First General Index of Molecular Complexity. *J. Am. Chem. Soc.* **1981**, *103*, 3599−3601.

(9) Bertz, S. H. Convergence, Molecular Complexity and Synthetic Analysis. *J. Am. Chem. Soc.* **1982**, *104*, 5801−5803.

(10) Mekenyan, O.; Bonchev, D.; Trinajstić, N. Structural Complexity and Molecular Properties of Cyclic Systems with Acyclic Branches. *Croat. Chem. Acta* **1983**, *56*, 237−261.

(11) Bonchev, D.; Polansky, O. E. On the Topological Complexity of Chemical Systems. In *Graph Theory and Topology in Chemistry;* King, R. B., Rouvray, D. H., Eds.; Elsevier: Amsterdam, 1987; pp 126−158.

(12) Basak, S. C. Use of Molecular Complexity Indices in Predictive Pharmacology and Toxicology: A QSAR Approach. *Med. Sci. Res.* **1987**, *15*, 605−609.

(13) Mil'man, B. L. A Complexity Measure for Chemical Compounds. *J. Struct. Chem.* **1988**, *29*, 957−960.

(14) Kier, L. B.; Testa, B. Complexity and Emergence in Drug Research. *Adv. Drug Res.* **1995**, *26*, 1−43.

(15) Nikolić, S.; Trinajstić, N.; Jurić, A.; Mihalić, Z.; Krilov, G. Complexity of Some Interesting (Chemical) Graphs. *Croat. Chem. Acta* **1996**, *69*, 883−897.

(16) Bonchev, D. Novel Indices for the Topological Complexity of Molecules. *SAR QSAR Environ. Res.* **1997**, *7*, 23−43.

(17) Bertz, S. H.; Sommer, T. J. Rigorous Mathematical Approaches to Strategic Bonds and Synthetic Analysis Based on Conceptually Simple New Complexity Indices. *Chem. Commun.* **1997**, 2409−2410.

(18) Nikolić, S.; Tolić, I. M.; Trinajstić, N. On the Complexity of Molecular Graphs. *Commun. Math. Comput. Chem.* **1999**, *40*, 187−201.

(19) Basak, S. C. Information Theoretic Indices of Neighborhood Complexity and Their Applications. In *Topological Indices and Related Descriptors in QSAR and QSPR*; Devillers, J., Balaban, A. T., Eds.; Gordon and Breach: Amsterdam, 1999; pp 563−593.

(20) Randić, M. On the Concept of Molecular Complexity. *Croat. Chem. Acta*, in press.

(21) Nikolić, S.; Tolić, I. M.; Trinajstić, N.: Baučić, I. On the Zagreb Indices as Complexity Indices. *Croat. Chem. Acta*, in press.

(22) Trinajstić, N. *Chemical Graph Theory*, 2nd revised ed.; CRC Press: Boca Raton, FL, 1992.

(23) Bonchev, D.; Trinajstić, N. Information Theory, Distance Matrix and Molecular Branching. *J. Chem. Phys.* **1977**, *67*, 4517−4533.

(24) Mowshowitz, A. Entropy of the Complexity of Graphs. 1. An Index of the Relative Complexity of Graphs. *Bull. Math. Biophys.* **1968**, *30*, 175−204.

(25) Shannon, C.; Weaver, W. *Mathematical Theory of Communication*; University of Illinois Press: Urbana, IL, 1949.

(26) Rashevsky, N. Life, Information Theory and Topology. *Bull. Math. Biophys.* **1955**, *27*, 379−423.

(27) Bertz, S. H. On the Complexity of Graphs and Molecules. *Bull. Math. Biol.* **1983**, *45*, 849−855.

(28) Bertz, S. H. A Mathematical Model of Molecular Complexity. In *Graph Theory and Topology in Chemistry*; King, R. B., Ed.; Elsevier: Amsterdam, 1983; pp 206−221.

(29) Bertz, S. H. Branching in Graphs and Molecules. *Discrete Appl. Math.* **1988**, *19*, 65−83.

(30) Bertz, S. H.; Sommer, T. J. Applications of Graph Theory to Synthesis Planning: Complexity, Reflexivity and Vulnerability. *Organic Synthesis: Theory and Applications*, Vol. II.; JAI Press: Greenwich, CT, 1993; pp 67−92.

(31) Bertz, S. H. The Bond Graph. *J. Chem. Soc., Chem. Commun.* **1981**, 818−820.

(32) The line graph $L(G)$ of G can be constructed by representing each edge in G by a vertex and connecting a pair of vertexes with an edge whenever the corresponding edges in G are adjacent.

(33) Bonchev, D.; Temkin, O. N.; Kamenski, D. On the Complexity of Linear Reaction Mechanisms. *React. Kinet. Catal. Lett.* **1980**, *15*, 119–124.

(34) Bonchev, D.; Kamenski, D.; Temkin, O. N. Complexity Index for the Linear Mechanisms of Chemical Reactions. *J. Math. Chem.* **1987**, *1*, 345–388.

(35) Gordeeva, E.; Bonchev, D.; Kamenski, D.; Temkin, O. N. Enumeration, Coding and Complexity of Linear Reactions Mechanisms. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 436–445.

(36) Temkin, O. N.; Zeigarnik, A. V.; Bonchev, D. *Chemical Reaction Networks. A Graph Theoretical Approach*; CRC Press: Boca Raton, FL, 1996; Chapter 6.

(37) Mallion, R. B. On the Number of Spanning Trees in a Molecular Graph. *Chem. Phys. Lett.* **1975**, *36*, 170–174.

(38) Gutman, I.; Mallion, R. B.; Essam, J. W. Counting the Spanning Trees of a Labeled Molecular-Graph. *Mol. Phys.* **1983**, *50*, 859–877.

(39) O'Leary, B.; Mallion, R. B. Counting the Spanning Trees of Labeled, Planar Molecular Graphs Embedded on the Surface on a Sphere. In Rouvray, D. H., Eds.; *Graph Theory and Topology in Chemistry*, King, R. B., Rouvray, D. H., Eds.; Elsevier: Amsterdam, 1987; pp 544–551.

(40) Kirby, E. C.; Mallion, R. B.; Pollak, P. On the Question of Counting Spanning Trees in a Labeled, Nonplanar Graph. *Mol. Phys.* **1994**, *83*, 599–602.

(41) John, P. E.; Mallion, R. B. Calculating the Number of Spanning Trees in a Labeled Planar Molecular Graph Whose Inner Dual Is a Tree. *Int. J. Quantum Chem.* **1996**, *60*, 59–66.

(42) Brown, T. J. N.; Mallion, R. B.; Pollak, P.; de Castro, B. R. M.; Gomes, J. A. N. F. The Number of Spanning Trees of Buckminsterfullerene. *J. Comput. Chem.* **1991**, *12*, 1118–1124.

(43) Gutman, I.; Mallion, R. B. On Spanning Trees in Catacondensed Molecules. *Z. Naturforsch.* **1993**, *48a*, 1026–1030.

(44) John, P. E.; Mallion, *R. B.* An Algorithmic Approach to the Number of Spanning Trees in Buckminsterfullerene. *J. Math. Chem.* **1994**, *15*, 261–271. Erratum: *J. Math. Chem.* **1994**, *16*, 389–390.

(45) John, P. E.; Mallion, R. B.; Gutman, I. An Algorithm for Counting Spanning Trees in Labeled Molecular Graphs Homeomorphic to Cata-Condensed Systems. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 108–112.

(46) Mallion, R. B. Some Graph-Theoretical Aspects of Simple "Ring-Current" Calculations on Conjugated Systems. *Proc. R. Soc. London* **1975**, *A341*, 429–449.

(47) Haig, C. W.; Mallion, R. B. Ring Current Theories in Nuclear Magnetic Resonance. In *Progress in Nuclear Magnetic Resonance Spectroscopy*, Vol. 13; Emsley, J. W., Feeney, J., Sutcliffe, L. H., Eds.; Pergamon Press: Oxford, U.K., 1979/1980; pp 303–344.

(48) Haig, C. W.; Mallion, R. B. Rationalisation of Relative "Ring-Current" Sizes in Polycyclic Conjugated Hydrocarbon. *Croat. Chem. Acta* **1989**, *62*, 1–26.

(49) Gomes, J. A. N. F.; Mallion, R. B. The Concept of Ring Currents. In *Concepts in Chemistry: A Contemporary Challenge*; Rouvray, D. H., Ed.; Wiley: New York, 1997; pp 205–253.

(50) McWeeny, R. Ring Currents and Proton Magnetic Resonance in Aromatic Molecules. *Mol. Phys.* **1958**, *1*, 311–321.

(51) Nikolić, S.; Trinajstić, N.; Knop, J. V.; Müller, W. R.; Szymanski, K. On the Concept of the Weighted Spanning Tree of Dualist. *J. Math. Chem.* **1990**, *4*, 357–374.

(52) Mohar, B. Laplacian Matrices of Graphs. In *MATH/CHEM/COMP 1988*; Graovac, A., Ed.; Elesevier: Amsterdam, 1989; pp 1–8.

(53) Trinajstić, N.; Babić, D.; Nikolić, S.; Plavšić, D.; Amić, D.; Mihalić, Z. The Laplacian Matrix in Chemistry. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 368–376.

(54) Balasubramanian, K. Laplacians of Fullerenes (C_{42}–C_{90}). *J. Phys. Chem.* **1995**, *99*, 6509–6518.

(55) Wilson, R. J. *Introduction to Graph Theory*; Oliver and Boyd: Edinburgh, U.K., 1972.

(56) Rücker, G.; Rücker, Ch. Automatic Enumeration of All Connected Subgraphs. *Commun. Math. Comput. Chem.*, in press.

(57) Randić, M. Augmented Valence-Novel Graph-Theoretical Index with an Apparent Interpretation. *Preprint*.

(58) Gutman, I.; Trinajstić, N. Graph Theory and Molecular Orbitals. Total π-Electron Energy of Alternant Hydrocarbons. *Chem. Phys. Lett.* **1972**, *17*, 535–538.

(59) Gutman, I.; Ruščić, B.; Trinajstić, N.; Wilcox, C. F., Jr. Graph Theory and Molecular Orbitals. XII. Acyclic Polyenes. *J. Chem. Phys.* **1975**, *62*, 3399–3405.

(60) Branching is reflected in the vertex degrees higher than 2; trees with vertex degrees 1 and 2 are not branched.

(61) Symmetry is reflected in the number of symmetry-nonequivalent vertexes.

(62) Rücker, G.; Rücker, Ch. Walk Counts, Labyrinthicity, and Complexity of Acyclic and Cyclic Graphs and Molecules. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 99–106.

(63) One can look upon cyclicity as a kind of structural change resembling branching in that the values of degrees of some vertexes within the cycle increase from 2 to 3 and higher at the sites of bridging cycle. This structural characteristic has been named (Bonchev, D. Kolmogorov's Information, Shannon's Entropy, and Topological Complexity of Molecules. *Bulg. Chem. Comm.* **1995**, *28*, 567–582) an internal branching. Therefore, the degree of a vertex in a cycle can increase either by cyclization (the internal branching) or by attaching it to acyclic (the external branching) or cyclic (spiro-cycles) fragments.