HIERARCHICAL PARTIALLY ORDERED SETS BASED ON TOPOLOGICAL COMPLEXITY

Danail Bonchev[a] and Ekaterina Gordeeva[b]

[a]Program for Theory of Complex Natural Systems, Texas A&M University,
Galveston, TX 77553-1675, email bonchevd@aol.com
[b]Molecular Design Limited (MDL), San Leandro, California

Abstract

Topological complexity of molecules was recently quantified by means of the total number of subgraphs, and the overall connectivity (the connectivity of a molecular graph and all of its subgraphs), as well as by the vectors of these descriptors. Hierarchically ordered partial sets of molecules are defined on this basis and their properties are studied in a number of propositions, theorems, and corollaries. The components of molecular topology are thus hierarchically ordered in the sequence: size < multiple bonds, cycles < heteroatoms < vertex degree majorization (branching, cycle redistribution) < centricity, elongation, clustering. The hierarchical ordering of the sets of C7-C10 alkanes and the cyclopentane isomers is compared with their global ordering induced by the numerical values of topological complexity indices. Some conclusions are drawn on the nature of branching and cyclicity of molecules.

1. Introduction.

Partial orderings of sets (posets) in chemistry have a long and distinguished history beginning with Mendeleev's Periodic Table of chemical elements. The review of Klein and Babić [1] is a good introduction to the area. Mention should be made to the pioneering works of Ruch [2,3] and those of Randić [4,5]. Attention may be directed to the idea of constructing *representative* posets by the superposition of several such sets based on essential graph invariants [6,7].

Posets reflect the inherent similarity in the set elements, however, they may also be regarded as resulting from insufficiently discriminating criteria. The question whether every partial ordering can be converted into a complete ordering has been solved and widely discussed in the mathematical literature [8]. Isotonic functions (or "Schur convex") from the poset to the totally ordered real numbers exist at least for all denumerable posets.

Due to the isomerism, chemistry commonly deals with sets of molecular structures which are only partially ordered. Strategies for transforming less ordered sets of chemical structures into more ordered and, possibly, completely ordered sets characterized by quantitative structural descriptors are of certain interest for classification and coding of chemical compounds, as well as in the area of structure-property and structure-activity relationships. An effective approach to the problem is the introduction of hierarchical relationships between the elements of the set. Different types of hierarchy may be used with that purpose. In this paper we deal with hierarchical partial orderings based on topological complexity.

Topological complexity of a chemical compound is a very important component of the general molecular complexity; other contributions emerging from the 3D-geometry, elemental composition, types of chemical bonds, etc.[9-12] The representation of topological complexity has evolved through the years from measures based on the automorphism group of the molecular graph [13-18], through some topological indices sensitive toward branching and cyclicity patterns [19-24], to the recent extensive use of all substructures and their weightings [25-30]. Throughout this paper our analysis of partially ordered sets is based on the concept of topological complexity as overall connectivity, and on the underlying count of substructures of varying size [27,30].

2. Topological Complexity of a Molecule as Presented by the Number of Subgraphs and the Overall Connectivity of Molecular Graph

Let G be a hydrogen-depleted molecular graph with N vertices, E edges, and a total of $K(G)$ subgraphs including G itself . Let also a_i (G) and $a_i(G_e)$ be the degree of vertex i in G and in its subgraph G_e *respectively.* The set of connected subgraphs may be represented by an ordered set $K'(G)$ of terms ^eK each one being the count of e-edge connected subgraphs, and the sum $K(G)$ of these terms may be regarded as an overall measure for the topological complexity of G:

$$K'(G) = K\{^oK,^1K,^2K,... ,^EK\} ; \quad K(G) = \sum_{e=0}^{E} {}^eK \tag{1}$$

The zero-order complexity in (1) is simply the number of graph vertices, and the first-order term ^1K counts the number of bonds. The second-order complexity ^2K is equal to the number of 2-edge subgraphs ("propane" fragments in the language of organic chemistry) widely used in the past as a structural index by Platt [31], as well as Gordon and Scantlebury [32], and as a complexity measure by Bertz [9,15,16]. The third-order term ^3K, the number of three-edge subgraphs, includes graphs standardly denoted as K_3 and $K_{1,3}$, and termed path graphs and stars, respectively. The number of K_3 graphs has been used in structure-property studues by Wiener [33] under the name "polarity number", as a second structural index along with his widely known "Wiener" number.

When every vertex in each of the subgraphs is weighted by its degree a_i (G), the weighted sum of all subgraphs $TC(G)$ is defined as a more sensitive measure of topological complexity than $K(G)$. Being a generalization of the molecular connectivity indices of Randić, Kier and Hall [34-37], $TC(G)$ has the meaning of *overall connectivity* of G. It can also be decomposed into topological complexities of order e, $^eTC(G)$ terms. In ascending order these include ^oTC as the zero-order complexity (vertex connectivity), ^1TC as the first-order complexity (edge connectivity), ^2TC as the second-order complexity (two-edge subgraph connectivity), etc. As an ordered set they form the vector of topological complexity (overall connectivity) $TC'(G)$:

$$TC(G) = \sum_{e=0}^{E} {}^{e}TC(G) = \sum_{e=0}^{E} \sum_{t=1}^{{}^{e}K} \sum_{i=1}^{n_t} a_i(G) \tag{2}$$

$$TC'(G) = TC \{{}^{o}TC, {}^{1}TC, {}^{2}TC,..., {}^{E}TC\} \tag{3}$$

The summing in eq. (2) runs first over all n_t vertices of subgraph G_t, then over all ${}^{e}K$ subgraphs G_t having e edges, and finally over all E classes of subgraphs having a fixed number of e edges. When the vertex degrees $a_i({}^{e}G_t)$ are those from the individual subgraphs ${}^{e}G_t$ having e edges, the second version $TC1(G)$ of the topological complexity index is defined:

$$TC1(G) = \sum_{e=0}^{E} {}^{e}TC1(G) = \sum_{e=0}^{E} \sum_{t=1}^{{}^{e}K} \sum_{i=1}^{n_t} a_i({}^{e}G_t) \tag{2a}$$

$$TC1'(G) = TC1 \{{}^{o}TC1, {}^{1}TC1, {}^{2}TC1,..., {}^{E}TC1\} \tag{3a}$$

The vectors of $K'(G)$, $TC'(G)$, and $TC1'(G)$, which are hierarchically ordered sets of graph invariants characterizing the topological complexity of the individual molecule, provide a natural basis for a hierarchical ordering of a set of molecules. Here, *hierarchical* is regarded in contrast to *global* ordering of molecules, according to their total values of $K(G)$ and $TC(G)$. Our idea of hierarchically ordered overall connectivities (HOOC) may be regarded as an extension of the concept of hierarchically ordered extended connectivities (HOC) by Balaban et al [38,39].

3. Definitions for Hierarchical Partial Orderings Based on Topological Complexities

Let G be a molecular graph. Denote the scalar complexity vectors $K'(G)$ or $TC'(G)$, as defined by eqs. (1-3), by $S'(G)$; denote also the complexity indices $K(G)$ and $TC(G)$ by $S(G)$.

Definition 1: The intramolecular hierarchical order of topological complexities is defined as
$$^{o}S < {}^{1}S < {}^{2}S < ... < {}^{E}S \tag{4}$$

Let graphs G_1 and G_2 represent molecules 1 and 2, respectively. Let also the complexity vectors $S'(G_1)$ and $S'(G_2)$ be specified as

$$S'(G_1) = S\{^oS(G_1),\ ^1S(G_1),\ ^2S(G_1),...,^ES(G_1)\} \tag{5a}$$

$$S'(G_2) = S\{^oS(G_2),\ ^1S(G_2),\ ^2S(G_2),...,^ES(G_2)\} \tag{5b}$$

Then,

$\quad\quad S'(G_1) < S'(G_2)$ if

a) $\quad ^oS(G_1) < {}^oS(G_2)$ or

b) $\quad$ if $^oS(G_1) = {}^oS(G_2)$, and $^1S(G_1) < {}^1S(G_2)$ or

c) $\quad$ if $^oS(G_1) = {}^oS(G_2)$, $^1S(G_1) = {}^1S(G_2)$, and $^2S(G_1) < {}^2S(G_2)$, etc. $\tag{6}$

The comparison between the two sequences (5a) and (5b) continues until the first different pair of corresponding terms is found. If the two sets (5a) and (5b) are identical, there is no seniority relation between the two molecules. The satisfaction of the hierarchical condition (6) may be denoted as

$$G_1 \leq_{S'} G_2 \tag{6a}$$

where the relation $\leq_{S'}$ is a *partial ordering* on S' sequences.

Definition 3: Graphs G_1 and G_2 are ordered by a *complete ordering* relation $\leq_{S'}$ defined such that

$$G_1 \leq_{S'} G_2 \quad \text{iff} \quad S(G_1) \leq S(G_2) \tag{7}$$

4. Theorems for Complexity Factors Influencing the Hierarchical Partial Orderings of Sets of Molecules

4.1. The Size.

We start our analysis with $^oK(G)$, the number of non-hydrogen atoms in the molecule (the number of vertices in the molecular graph). Taken as the first criterion in our hierarchical ordering scheme, it classifies the molecules in any set into classes according to their H-depleted size. From Definitions

1 and 2 (eqs. 4-6) we can formulate

Proposition 1: The molecules having a larger number of atoms are more complex than the ones with a lesser number of atoms,

$$^{o}K_{N} < {}^{o}K_{N-1} \tag{8}$$

The importance of molecular size as complexity factor was analyzed in the studies of Bertz proceeding from his "size term", based on two-edge subgraphs [15], and in his branching Theorem 1 [20].

The classification of larger molecules as more complex than the smaller one seems natural, but perhaps the connectedness is more essential for complexity than size. Shouldn't we regard a smaller molecule with a complicated structure as more complex than a larger molecule with a simple architecture? We point out that both size and connectivity have their importance for structural complexity. The orderings produced relying more heavily on connectivity, like those resulting from the *overall* indices $K(G)$ and $TC(G)$, classify in many cases the smaller molecules with a higher degree of branching or/and cyclicity as more complex than the larger molecules with a limited atom-atom connectedness.

The ranges of values of the two overall indices in alkanes illustrate the point. They are shown in TABLE 1 along with the ranges of a number of physicochemical properties of these compounds. The comparison shows that both types of ordering, the hierarchical which favors the size, and the overall one which gives preference to connectivity, have their analogs in alkane properties. Thus, for molar volume, molecular refraction, and heat of atomization, the ranges corresponding to different numbers of atoms are narrow and clearly separated; it is the number of atoms that matters rather than their connectedness. Conversely, for the heat of vaporization, critical pressure, critical temperature, surface tension, and boiling point (beginning with C9), the span and the overlap of ranges shows that the structural complexity of the highly branched molecules is frequently more important than their size.

TABLE 1. Ranges of values of some physicochemical properties of alkanes and those of the complexity indices K and TC.

Indices and Properties	C6	C7	C8	C9
$K(G)$	21-30	28-49	36-86	45-124
$TC(G)$	100-181	154-369	224-798	312-1303
Boiling point[a], °C	57-69	81-98	99-126	122-151
Heat of vaporization[a], kJ/mol	27.7-31.6	32.0-36.6	35.1-41.5	38.1-46.4
Critical Temperature[a], °C	216-235	247-268	271-305	297-330
Critical pressure[a], atm	29.9-31.0	27.1-30.0	24.5-29.0	22.7-27.0
Surface tension[a], dyn/cm	16.3-18.4	18.0-22.0	18.8-22.0	20.4-23.8
Molar volume[a], cm³/mol	130-133	144-149	157-165	169-181
Molecular refraction[a], cm³/mol	29.8-29.9	34.3-34.6	38.7-39.3	43.0-43.9
Heat of atomization[b], kJ/mol	1797-1801	2078-2082	2358-2362	

[a]Data taken from [40]; [b]Data from [36].

The ranges of values of the two overall indices in alkanes illustrate the point. They are shown in TABLE 1 along with the ranges of a number of physicochemical properties of these compounds. The comparison shows that both types of ordering, the hierarchical which favors the size, and the overall one which gives preference to connectivity, have their analogs in alkane properties. Thus, for molar volume, molecular refraction, and heat of atomization, the ranges corresponding to different numbers of atoms are narrow and clearly separated; it is the number of atoms that matters rather than their connectedness. Conversely, for the heat of vaporization, critical pressure, critical temperature, surface tension, and boiling point (beginning with C9), the span and the overlap of ranges shows that the structural complexity of the highly branched molecules is frequently more important than their size.

4.2. Adjacency

Within each class of molecules with N atoms we continue the hierarchical classification by the first level of molecular connectivity which is measured either by the number of graph edges $E = {}^1K$, the first-order K-index, or by the graph adjacency $A = 2E = {}^\circ TC$, which by definition is the zero-order overall connectivity. Adding one edge to a graph with a fixed number of vertices means a formation of either a cycle, or a multiple bond. Therefore, although this factor does not add to the complexity of acyclic molecules with single bonds, it hierarchically distinguishes them from the large classes of molecules with multiple bonds and those of cyclic molecules. Actually, from a graph-theoretical point of view, the multiple bonds are cycles of two vertices, a fact correctly recognized by the hierarchical ordering based on ${}^1K$ and ${}^\circ TC$. Our Definitions 1 and 2 thus provide

Proposition 2: ${}^1K_N$(simple acyclic graphs) $<$ ${}^1K_N$(multigraphs; cyclic graphs)   (9a)

   ${}^\circ TC_N$(simple acyclic graphs) $<$ ${}^\circ TC_N$(multigraphs; cyclic graphs)   (9b)

Proposition 2 confirms the similar conclusions of Bertz [15, 29] about the larger complexity of multigraphs and cyclic graphs relative to that of acyclic graphs.

The presence of cycle(s) or multiple bond(s) in a molecule increases the number of subgraphs of any number of edges considerably stronger than the presence of a branch. As a result, the hierarchical precedence order induced by the size of the molecule is more frequently reversed by the overall ordering imposed by the $K(G)$ and $TC(G)$ indices. We prove below this assertion for the case of chain molecules with N atoms, as compared to monocyclic molecules with $(N-1)$ atoms.

Theorem 1a: The number of subgraphs in a path graph with N vertices, P_N, is always smaller than that of the monocyclic graph with $(N-1)$ vertices C_{N-1}.

Proof: From the formulas for the index K, previously derived [30],

$$K(P_N) = N(N+1)/2 \quad ; \quad K(C_{N-1}) = N^2 + 1 \tag{10a,10b}$$

one obtains

$$K(C_{N-1}) - K(P_N) = (N^2 - 3N + 4)/2 > 0 \qquad (11)$$

Theorem 1b: The overall connectivity TC of a path graph with N vertices, P_N, is always smaller than that of the monocyclic graph with $(N-1)$ vertices C_{N-1}.

Proof: From the formulas for the TC index , previously derived [30],

$$TC\ (P_N) = N(N-1)(N+4)/3 \quad ; \quad TC\ (C_{N-1}) = N(N^2 + N + 2) \qquad (12)$$

one obtains

$$TC\ (C_{N-1}) - TC\ (P_N) = 2(N-1)(2N^2 - 4N + 6)/3 > 0 \qquad (13)$$

The competition between the size and adjacency factors in molecules with cycles or multiple bonds may be illustrated with several properties of cyclohexane and 1-hexyne molecules [41]. Thus, for the boiling point, a property for which the value ranges of different sets of isomers are well separated up to C9 alkanes, the hexyne with a triple bond and a boiling point of 71.3 °C lies outside the 57- 69 °C range of the C6 alkanes, whereas the value of 83°C for cyclohexane overlaps with the 81-98°C range of C7 alkanes. Even more impressive is the example with molecular volume, a property which depends strongly on the number of atoms. The values 108.7 cm³/mol and 115.7 cm³/mol for cyclohexane and 1-hexyne, respectively, go *below* the range of 116-123 cm³/mol of the three acyclic pentanes.

The competing complexity factors of branching and cyclicity of the molecular skeleton will be briefly analyzed by a comparison of monocyclic graphs and star graphs the ordering of which is shown below to depend on the total number of graph vertices. More specifically, we show that monocycles are more complex than the respective stars.

Theorem 2: The number of subgraphs in 3-arm stars and 4-arm stars is smaller than that of the corresponding monocycles having four and five vertices; the 5-arm star and the hexagonal cycle have the same number of subgraphs; monocyclic graphs having more than six vertices have less subgraphs than the stars with the same number of vertices.

Proof: From (10b), and the previously derived formula (14)

$$K (S_N) = 2^{N-1} + N - 1 \qquad (14)$$

one obtains

$$K(C_N) - K (S_N) = N^2 - N - 2 - 2^{N-1} \qquad (15)$$

wherefrom Theorem 2 immediately follows.

4.3. Heteroatoms

The presence of heteroatoms does not change the total number of subgraphs in a molecular graph, therefore the K index cannot be used as an additional criterion for ordering of heteroatomic molecules having the same number of atoms, cycles, and multiple bonds. However, one may use the overall connectivity indices as recently modified by replacing the vertex degrees a_i in formula (2) by the valence connectivities δ_i of Kier and Hall [42]:

$$TC^v(G) = \sum_{e=0}^{E} {}^tTC^v(G) \overset{E}{\underset{e=0}{\Sigma}} \overset{{}^tK}{\underset{t=1}{\Sigma}} \delta_i(G) \overset{n_i}{\underset{i=1}{\Sigma}} \qquad (16)$$

$$TC^v(G)' = TC^v({}^0TC^v, {}^1TC^v, {}^2TC^v, \ldots, {}^ETC^v) \qquad (16a)$$

For main group elements, δ_i is equal to the number of valence electrons diminished by the number of adjacent H atoms. Thus, for the nitrogen atom in $-NH_2$, $\delta^v = 3$, in $-NH-$ and $=NH$ it is 4; in $=N-$ it is 5. For the oxygen atom in $-OH$, $\delta^v = 5$, whereas in $=O$ and in $-O-$ it is 6, etc.

We may extend our graph theoretical analysis by representing the molecules with heteroatoms by vertex-weighted graphs, considering valence connectivity δ_i as a weight of vertex i.

Definition 4: In sets of graphs having the same number of vertices, multiple edges and cycles, as well as having their vertices weighted by the valence connectivity weights δ_i, graph G_I has a precedence

over (is less complex than) graph G_2,

$$TC'(G_1) < TC'(G_2) \text{ if}$$

a) $^oTC'(G_1) < {}^oTC'(G_2)$ or if

b) $^oTC'(G_1) = {}^oTC'(G_2)$, and $^1TC'(G_1) < {}^1TC'(G_2)$ or if

c) $^oTC'(G_1) = {}^oTC'(G_2)$, $^1TC'(G_1) = {}^1TC'(G_2)$, and $^2TC'(G_1) < {}^2TC'(G_2)$, etc.   (17)

Proposition 3: The overall valence connectivity of a graph, $TC'(G)$ is not smaller than its overall connectivity $TC(G)$:

$$TC(G) \leq TC'(G) \tag{18}$$

Since $\delta_i \geq a_i$, the inequality takes place if at least for one of the G_i vertices i the inequality $\delta_i > a_i$ holds. The condition for partial ordering is thus met.

Notably, the presence of heteroatoms in a molecule affects the zero-order overall valence connectivity $^oTC'(G)$, so that it appears as a complexity factor of importance comparable to that of the multiple bonds and cycles, which affect the zero-order overall connectivity $^oTC(G)$. We may summarize the first part of our analysis of complexity factors that hierarchically order molecular graphs into the precedence relationship

size < multiple bonds, cycles < heteroatoms

We shall turn now to complexity factors on the next hierarchical level determined by the first order overall connectivities ^1TC or/and the second order complexities ^2K.

4.4. Branching

The number of branches is the next important complexity factor which imposes additional ordering in the subclasses of graphs with a fixed number of vertices, multiple edges, cycles, and heteroatoms. The creation of a branch at the cost of the shortening of the longest chain changes the degrees of the

two vertices to which the transferred terminal vertex is attached.
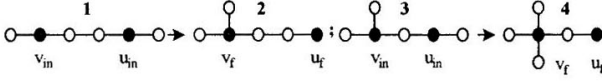
Creating a vertex i of a higher degree increases the number of two-edge subgraphs ^2K. We may summarize several branching patterns, related to changes in the vertex degrees, into

Theorem 3: Any change in the structure of a graph that preserves the number of vertices and edges, and the vertex degrees of all but two vertices u and v, increases the graph complexity by changing

a) the number of two-edge subgraphs by

$$\Delta^2 K = a_v - a_u + 1 \tag{19}$$

b) the first-order overall connectivity index by

$$\Delta^1 TC = 2(a_v - a_u + 1) = 2\Delta^2 K \tag{20}$$

Proof: a) Denoting the initial and final vertex degree of u and v by the subscripts *in* and *f*, respectively, we have :

$$\Delta^2 K = {}^2K_{v,f} + {}^2K_{u,f} - {}^2K_{v,in} - {}^2K_{u,in} \tag{21}$$

The number of two-edge subgraphs ^2K_i which include vertex i is calculated according to the formula [20]

$$^2K_i = a_i (a_i - 1)/2 \tag{22}$$

The only option for u and v to change their degrees is the increase in a_v to be exactly

compensated by the decrease in a_u. For simplicity, we assume that the vertex degrees change by one, because the larger changes can be presented as a total change in two or more consecutive steps of structure modification:

$$a_{v,f} = a_{v,in} + 1 \quad ; \quad a_{u,f} = a_{u,in} - 1 \tag{23}$$

Substituting (23) and (22) into (21) one arrives at eq. (19).

b) The first-order overall connectivity is the sum of vertex degrees over all edges in the graph:

$$^{1}TC = \sum_{\text{all edges}} (a_i + a_j) \tag{24}$$

Since only the vertex degrees of u and v alter, we have

$$\Delta^{1}TC = a_{v,f}^{2} - a_{v,in}^{2} + a_{u,f}^{2} - a_{u,in}^{2} \tag{25}$$

Substituting (23) into (25) one obtains eq. (20).

Corollary 1: When in a graph, a branch is formed at the expense of the shortening of the longest chain or at the expense of the shortening of a cycle: a) the number of two-edge subgraphs increases by one, and the first-order overall connectivity increases by two, when the new branch is attached at a vertex of degree two; b) the increase in these two quantities is twice as much when the new branch is formed at a vertex of degree three.

For acyclic graphs cases (a) and (b) are exemplified by graphs **1**, **2**, and by graphs **3**, **4**, respectively. For cyclic graphs, this corollary is illustrated by graphs **5** → **6** → **7**, and **6** → **8**, respectively.

	5	**6**	**7**	**8**
		6	**7**	**8**
^{2}K	5		24	26
^{1}TC	20	22		
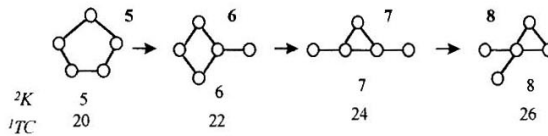
Corollary 2: When in a graph having a fixed number of branches a branch is displaced to a vertex of a higher degree, the number of two-edge subgraphs increases by one, and the first-order overall connectivity increases by two.

This corollary can be traced down from the direct comparison of the cyclic graphs **7 → 8** and the acyclic **2 → 4** ones. We may denote this type of branching pattern as 3,3 → 4,2, in contrast to the 2,2 → 3,1 or 3,2 → 4,1 vertex degree redistributions, the presence of "1" in which indicates the formation of a new branch. The new branching pattern 3,3 → 4,2 may be called after Klein & Babić [1] a "vertex-degree majorization". This pattern of increasing graph complexity appears also in cyclic graphs without any relation to branching; in these graphs it is a pattern of cyclic complexity that may be termed *cycle redistribution.*

Corollary 3: When in a cyclic graph a vertex-degree majorization occurs between two vertices without forming a branch, the graph complexity increases (the number of two-edge subgraphs increases by one, and the first-order overall connectivity increases by two).

We illustrate this pattern of partial orderings in cyclic graphs by graphs **9-10**.

$^{1}TC = 30$ $^{2}K = 10$; $^{1}TC = 32$
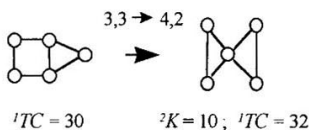
Our analysis of the factors that affect the next hierarchical level of topological complexity is thus completed. We may summarize our findings by expanding the hierarchical complexity relationship established in the previous subsection to

size < multiple bonds, cycles < heteroatoms < branching, cycle redistribution

It should be emphasized that Theorem 3 seems to summarize or generalize a number of previously specified branching patterns, such as branching rules 1-4 and 8 of Bonchev and Trinajstić [19], the

branching extent theorem of Ruch and Gutman [43], some of the branching theorems and analyses of Bertz [20], the generalized branching rules G1-G4 of Bonchev [21], and it is close to the branching extent definition of Klein and Babić [1]. On the other hand, Theorem 3 treats in a similar manner complexity in acyclic and cyclic systems thus creating bridges between the traditional separate treatment of acyclic and cyclic complexity termed branching and cyclicity.

We shall now turn to the more subtle complexity factors discriminated by the next hierarchical levels of the overall connectivities ^eTC and complexities ^eK.

4.5. Subtle Branch Patterns: Centricity, Elongation, and Clustering

To avoid the mixed influence of several factors we consider singly-branched graphs.

Theorem 4: The **shifting of a branch** from a position i to the neighboring more central position j along the longest chain increases: a) the jth order complexity index K by one; b) the $(j-1)$th order overall connectivity TC by two:

$$\text{(a)} \quad \Delta^j K = +1 \quad ; \quad \text{(b)} \quad \Delta^{j-1}TC = +2 \tag{26}$$

$^3K = 4$ $^3K = 5$ $^4K = 6$ $^4K = 7$
$^2TC = 29$ $^2TC = 31$ $^3TC = 56$ $^3TC = 58$
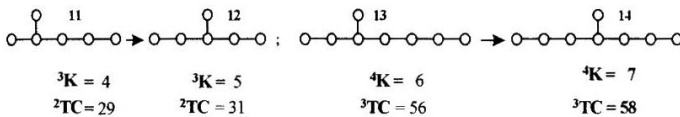
The proof of (26a) follows directly from the fact that the only change in the jth order complexity index ^jK upon the transformation described by Theorem 4 is the emergence of one more subgraph with j edges, namely, the subgraph connecting the terminal vertex of the branch to the closer terminus of the longest chain.

In proving (26b) we first derive by induction the formula for ^jK in singly-branched acyclic graphs having 2j vertices:

$$^{j-1}K_{2j} \text{ (singly-branched)} = j + 2 \tag{27}$$

Next, we prove that j of these $(j + 2)$ subgraphs preserve their $(j-1)$th order overall connectivity, ^{j-1}TC, because the latter is a sum of the same summands in which only the order of summing is changed. Each of the remaining two subgraphs increases its ^{j-1}TC by one: the first one, due to the inclusion of a vertex of degree two instead of the terminal vertex of degree one, and the second one, due to the inclusion of the branched vertex of degree three in one more $(j-1)$th order subgraph - that one starting with vertices $j, j+1$. $\Delta^j TC = +2$ is thus proved.

Corollary 4: The **elongation** of a branch up to half the length of the longest chain in an acyclic graph increases : a) the jth order complexity index ^jK by one; b) the ith order overall connectivity ^iTC by two:

$$\text{(a)} \quad \Delta^jK = +1 \;\; ; \;\; \text{(b)} \quad \Delta^iTC = +2 \tag{28}$$

Relations (28a) follow directly from Theorem 4 when presenting the longest chain of the graph to pass through the branch. This shows that the branch elongation is equivalent to a branch shift to a more central position.
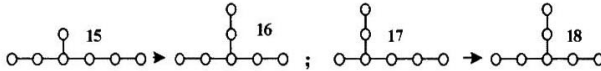
Theorem 5: **Clustering** of vertices of an acyclic graph (shifting them closer to each other), without changing their position with respect to the center of the longest chain, increases : a) the third order complexity index ^3K by one; b) the second order overall connectivity ^2TC by two:

$$\text{(a)} \quad \Delta^3K = +1 \;\; ; \;\; \text{(b)} \quad \Delta^2TC = +2 \tag{29}$$

Proof: The number of subgraphs with three edges increases by one, namely by the new three-edge subgraph that includes the two branches. The overall connectivity of two-edge subgraphs remains the same except for the two species incorporating one of the branches and the root vertex of the second

branch, each of the species increasing one vertex degree from two to three.

$^3K = 8$ $^3K = 9$ $^3K = 11$ $^3K = 12$
$^2TC = 50$ $^2TC = 52$ $^2TC = 64$ $^2TC = 66$
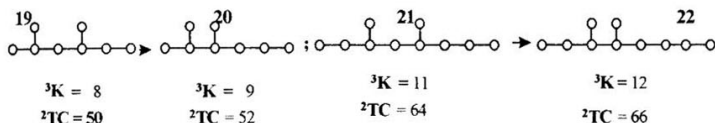
Theorem 5 with its three corollaries confirms the previously specified branching rules 5-7 of Bonchev and Trinajstić [19], based on the Wiener number, and the generalized rule G5 of Bonchev [21], based on the number of self-returning walks. Similar results have been obtained by Bertz [20] by applying the methodology of line graphs of higher order.

This concludes the analysis of the branching patterns that impose additional precedence order within the subclasses of graphs having the same number of vertices, edges, and two-edge subgraphs. More examples and subtle patterns will be analyzed in the next section. All the hierarchically ordered criteria of partial ordering of a set of chemical compounds, according to the topological complexity of the respective molecular graphs, may be summarized as follows:

size < multiple bonds, cycles < heteroatoms < vertex degree majorization (branching,

cycle redistribution) < centricity, elongation, clustering   (29)

The larger variety of finer patterns of cyclic complexity will be not be specified here. In Section 6, we will discuss some of them proceeding from the ordering of the cyclic graphs with five vertices.

5. Hierarchical Versus Global Ordering of C7-C10 Alkanes Based on Topological Complexity

Theorems 1-5 derived in the previous section enable the transformation of the partial alkane orderings, based on a limited set of criteria, into a complete ordering. Our analysis of the C7-C8 alkanes is based on recent data for the complexity indices ^cK and overall connectivities ^cTC [27]. The topological complexities of the C9 alkanes are given in TABLE 2. We shall also use the data for C10 alkanes, which will be published elsewhere [44].

An illustration is presented with the set of nine C7 alkane isomers (FIGURE 1).

$^3K=5, {}^2TC=35$ $^3K=6, {}^2TC=41$ $^3K=8, {}^2TC=52$

$^3K=6, {}^2TC=37$ $^3K=8, {}^2TC=46$ $^3K=10, {}^2TC=56$

$^3K=7, {}^2TC=39$

$^2K=5, {}^1TC=22$ $^2K=6, {}^1TC=24$ $^2K=7, {}^1TC=26$ $^2K=8, {}^1TC=28$ $^2K=9, {}^1TC=30$
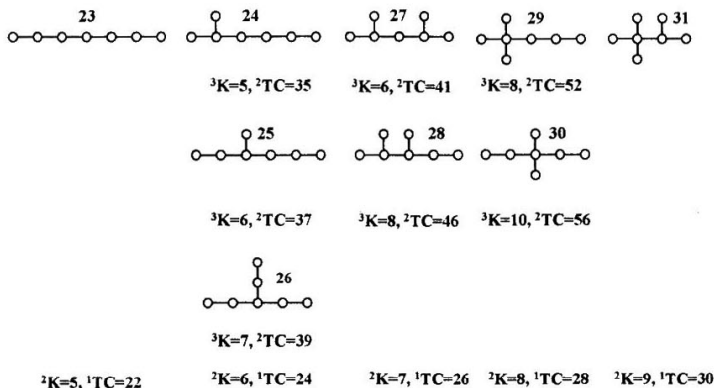
FIGURE 1. Hierarchical ordering of the nine C7 isomers according to the values of topological complexity indices ^cK (second and third order) and ^cTC (first and second order).

As seen from FIGURE 1, due to the fact that for acyclic isomers the number of atoms (^0K) and the number of bonds (^1K) is constant, it is the number of the two-edge subgraphs (the propane fragments count) ^2K, that imposes the isomer hierarchical grouping into five subsets. The hierarchical ordering continues within the subsets with cardinality larger than one, according to the count of the three-edge subgraphs ^3K. Exactly the same ordering is obtained, according to the

values of the first and second order overall connectivities ^1TC and ^2TC. Thus, the use of two consecutive orders of the complexity indices ^2K, ^3K or ^1TC, ^2TC suffice to transform the partial ordering of the C7 isomers into a complete one. The same complete ordering is provided by the overall indices K and TC:

$$K(C7): \quad 28 < 32 < 34 < 36 < 37 < 40 < 41 < 44 < 49$$
$$TC(C7): \quad 154 < 194 < 213 < 234 < 245 < 276 < 286 < 314 < 369$$

It will be shown in this section that in higher alkanes the hierarchical ordering and the ordering

produced by the overall indices deviate, and the number of reorderings increases with the number of carbon atoms. It will also be shown that beginning with C10 the $K'(G)$ vector becomes degenerate and the hierarchical partial ordering of the isomers can be converted into a complete one only with the $TC'(G)$ vector.

TABLE 2. Topological Complexity and Overall Connectivity of C9 Alkanes

#	Structure		^0S	^1S	^2S	^3S	^4S	^5S	^6S	^7S	^8S	Total
1	n-nonane	K	9	8	7	6	5	4	3	2	1	45
		TC	16	30	40	46	48	46	40	30	16	312
2	2-M		9	8	8	7	6	5	4	3	1	51
			16	32	47	55	59	59	55	45	16	384
3	3-M		9	8	8	8	7	6	5	3	1	55
			16	32	49	64	70	72	68	45	16	432
4	4-M		9	8	8	8	8	7	5	3	1	57
			16	32	49	66	81	83	68	45	16	456
5	2,6-MM		9	8	9	8	7	6	6	4	1	58
			16	34	54	64	70	74	84	60	16	472
6	3-E		9	8	8	9	9	8	6	3	1	61
			16	32	51	75	92	96	81	45	16	504
7	4-E		9	8	8	9	10	9	6	3	1	63
			16	32	51	77	103	107	81	45	16	528
8	2,5-MM		9	8	9	9	8	8	7	4	1	63
			16	34	56	73	83	99	97	60	16	534
9	2,2-MM		9	8	10	10	9	8	7	4	1	66
			16	36	64	83	93	99	97	60	16	564
10	2,4-MM		9	8	9	9	10	9	7	4	1	66
			16	34	56	77	104	110	97	60	16	570
11	2,3-MM		9	8	9	10	10	9	7	4	1	67
			16	34	58	85	104	110	97	60	16	580
12	3,5-MM		9	8	9	10	10	10	8	4	1	69
			16	34	58	84	106	124	110	60	16	608
13	2-M, 4-E		9	8	9	10	11	11	8	4	1	71
			16	34	58	86	117	135	110	60	16	632
14	3,4-MM		9	8	9	11	12	11	8	4	1	73
			16	34	60	96	127	135	110	60	16	654
15	3,3-MM		9	8	10	12	12	11	8	4	1	75
			16	36	68	104	127	135	110	60	16	672
16	2-M, 3-E		9	8	9	11	13	12	8	4	1	75
			16	34	60	98	138	146	110	60	16	678

TABLE 2 (Contd.)

# Structure	°S	¹S	²S	³S	⁴S	⁵S	⁶S	⁷S	⁸S	Total
16 2-M, 3-E	9	8	9	11	13	12	8	4	1	75
	16	34	60	98	138	146	110	60	16	678
17 2,2,5-MMM	9	8	11	11	10	11	10	5	1	76
	16	38	71	92	108	141	140	75	16	697
18 4,4-MM	9	8	10	12	14	12	8	4	1	78
	16	36	68	108	148	146	110	60	16	708
19 2,3,5-MMM	9	8	10	11	12	12	10	5	1	78
	16	36	65	96	129	152	140	75	167	725
20 3-M, 4-E	9	8	9	12	14	13	9	4	1	79
	16	34	62	107	150	160	123	60	16	728
21 2,2,4-MMM	9	8	11	12	13	14	11	5	1	84
	16	38	73	105	144	178	153	75	16	798
22 3-M, 3-E	9	8	10	14	16	14	9	4	1	85
	16	36	72	127	171	171	123	60	16	792
23 2,3,4-MMM	9	8	10	13	15	15	11	5	1	87
	16	36	69	117	165	189	153	75	16	836
24 2,4,4-MMM	9	8	11	13	15	15	11	5	1	88
	16	38	75	117	165	189	153	75	16	844
25 2,2,3-MMM	9	8	11	14	16	15	11	5	1	90
	16	38	77	128	175	189	153	75	16	867
26 2,4-MM, 3-E	9	8	10	13	17	16	11	5	1	90
	16	36	69	121	186	200	153	75	16	872
27 3,3-EE	9	8	10	16	19	16	10	4	1	93
	16	36	76	148	204	196	136	60	16	888
28 2,3,3-MMM	9	8	11	15	18	16	11	5	1	94
	16	38	79	140	196	200	153	75	16	913
29 2,2-MM, 3-E	9	8	11	15	19	18	12	5	1	98
	16	38	79	141	211	226	166	75	16	968
30 3,3,4-MMM	9	8	11	16	19	18	12	5	1	99
	16	38	81	149	211	226	166	75	16	978
31 2,3-MM, 3-E	9	8	11	17	21	19	12	5	1	103
	16	38	83	161	232	237	166	75	16	1024
32 2,2,4,4-MMMM	9	8	13	14	17	20	15	6	1	103
	16	42	88	128	198	260	210	90	16	1048
33 2,2,3,4-MMMM	9	8	12	16	20	21	15	6	1	108
	16	40	86	151	229	271	210	90	16	1109
34 2,3,3,4-MMMM	9	8	12	18	23	22	15	6	1	114
	16	40	90	174	260	282	210	90	16	1178
35 2,2,3,3,-MMMM	9	8	13	20	26	25	16	6	1	124
	16	42	100	197	299	320	223	90	16	1303

One can see from TABLE 2 that the overall connectivity TC provides a complete ordering of the 35 C9 isomers. The total number of subgraphs K, however, is degenerate and generates only a partial ordering. Five pairs of isomers were found with the same K values (2,2MM/2,4MM; 3,3MM/2M3E; 4,4MM/2,3,5MMM; 2,2,3MM/2,3MM,3E; and 2,3MM/2,2,4,4MMMM). For the set of 75 C10 isomers the K index is much more degenerate with 57 different values, whereas the TC index again provides a complete ordering of the set.

In TABLE 3 we present the hierarchical ordering of the C7-C9 alkanes according to the set of topological criteria specified in the previous section. The number of atoms (graph vertices) is the first criterion which separates the C7, C8, and C9 sets of isomers. Subsets are formed in columns within each set according to the second-order complexity ^2K. Within each column, the ordering from top to bottom is imposed by the third-order complexity ^3K. This suffices to order all C7 isomers, as already shown in FIGURE 1. In the set of eighteen C8 alkanes two pairs of isomers (3,4MM/2M3E; 3,3MM/2,3,4MMM) have the same ^3K values, and the fourth-order complexities ^4K are used to complete the hierarchical ordering. In the set of thirty five C9 alkanes there are already seven pairs and a triplet of isomers with the same ^3K values. The fourth-order terms ^4K order all seven pairs but the complete discrimination of the triplet of isomers required the fifth-order complexities to be used. The same hierarchical ordering of the C7-C9 alkanes is produced by the first-, second-, and third-order overall complexities ^eTC, the values of ^1TC being shown in parentheses in the headings of TABLE 3. The only case in which ^eK and ^{e-1}TC fail to produce the same ordering deals with the pair of isomers 3,3-dimethylhexane / 2,3,4-trimethylpentane. The number of branches B is also shown for each column to facilitate the branching pattern analysis.

In TABLE 4 we order in the same hierarchical manner the set of seventy five C10 isomers. They are grouped in eight subsets according to the values of their second-order complexities ^2K which range from 8 to 15, as well as by their first-order overall connectivities ranging from 34 to 48. The high cardinality of the subsets produced nine pairs, four triplets, and even five quintets of isomers, the complete hierarchical ordering of which required not only ^3K and ^4K but in ten cases also the fifth-order complexities ^5K. Moreover, the pair of isomeric 4,5-dimethyloctane and 2-methyl,3-ethylheptane was identified to be the first one for which the entire K' vectors 94(10, 9, 10, 12, 14, 14, 12, 8, 4, 1) are the same. However, these isomers are discriminated by the more sensitive fourth-order

Table 3. The ordering of C7- C9 alkane isomers as produced by the hierarchical ($^2K, ^1TC$; $^3K, ^2TC$; etc.) and the overall topological complexities (TC; the consecutive numbering), compared in parentheses with the ordering by the Randić $\lambda\lambda_1$ index [24]

^2K (^1TC)	5(22)	6(24)	7(26)	8(28)	9(30)
$B*$	0	1	2	2	3
1 nC7		2 2M 3 3M 4 3E	5 24MM 6 23MM	7 22MM 8 33MM	9 223MMM

^2K (^1TC)	6(26)	7(28)	8(30)	9(32)	10(34)	12(38)
B	0	1	2	2,3	3	4
1 nC8		2 2M 3 3M 4 4M 6 3E	5 25MM 7 24MM 9(8) 23MM 10 34MM 11 2M3E	8(9) 22MM 12 33MM 14 234MMM 15 3M3E	13 224MMM 16 223MMM 17 233MMM	18 2233MMMM

^2K(^1TC)	7(30)	8(32)	9(34)	10(36)	11(38)	12(40)	13(42)
B	0	1	2	2,3	3	4	4
1 nC9		2 2M 3 3M 4 4M 6 3E 7 4E	5 26MM 8 25MM 10(9)24MM 11 23MM 12 35MM 13 2M4E 14 34MM 16(15)2M3E 20 3M4E	9(10) 22MM 15(16)33MM 18 44MM 23 234MMM 26(24)2M3E4M 27(28) 33EE	17 225MMM 19 235MMM 21 224MMM 24(26)244MMM 25 223MMM 28(27)233MMM 29 22MM3E 30 334MMM 31 23MM3E	33 2234MMMM 34 2334MMMM	32 2244MMMM 35 2233MMMM

*B stands for the number of branches.

Table 4. Ordering of C10 isomers according to the hierarchical (^2K, ^1TC; ^3K, ^{27}TC, etc.) and the overall topological complexities (TC; the consecutive numbering), also compared (in parentheses) with the ordering generated by the Randić $_\partial\lambda_1$ index [22].

$^2K(^1TC)$	8(34)	9(36)	10(38)	11(40)	12(42)	13(44)	14(46)	15(48)
B	0	1	2	2,3	3,4	4	4	5
	1 nC10	2 2M	5(6) 27M	10(11) 22M	20(23) 226M	55(54) 2245M	50 2255M	74 22344M
		3 3M	8 26M	26 246M	29 225M	57 2235M	64 2244M	75 22334M
		4 4M	11(10) 25M	24(25) 236M	36(34) 224M	61(60) 2335M	70 2233M	
		6(5) 5M	14(13) 24M	19(20) 33M	34(35) 255M	66(65) 2234M	72 344M	
		7 3E	15 36M	32(31) 235M	43 22M4E	68 224M3E	73 223M3E	
		9 4E	13(14) 23M	27(28) 44M	42 244M	67(66) 2344M		
		12 4P	16 2M5E	35(33) 245M	40 223M	69 2334M		
		17 35M	39(37) 25M3E	46(47) 335M	71 23M3-iP			
		21(19) 2M4E	41(39) 234M	44(46) 233M				
		18 34M	45(44) 2M3-iP	52(55) 24M4E				
		25(24) 3M5E	33(36) 3M3E	59(58) 2345M				
		23(21) 45M	47(45) 345M	53 334M				
		22 2M3E	37(41) 4M4E	54(52) 22M3E				
		28(27) 4-iP	49(48) 23M4E	63(62) 24M3-iP				
		30(32) 3E4	51(49) 24M3E	56 344M				
		31(30) 3M4E	48(51) 33EE	58(59) 23MM3E				
		38 34EE		60(61) 33MM4E				
				62(63) 34MM3E				
				65(67) 2M33EE				

fourth-order overall connectivity ^4TC. The agreement in the ordering imposed by the two kinds of topological complexities is not as complete as it was for C4-C9; in eight cases the ^eK and ^eTC criteria produce a reversed isomer ordering (2,3MM/2M5E; 3,4MM/3M5E; 2,4,6MMM/2,3,6MMM; 3,3MM/2,3,5MMM; 4M4E/2,3MM4E; 2,2,3MMM/3,3,5MMM; 2,2MM3E/2,4MM3-iP; 2,3MM3E/3,3MM4E).

The **hierarchical ordering** determined by Theorems 1-5 and the global ordering produced by the overall connectivity TC (the complexity index K is not used in the comparison because of its degeneracy at higher numbers of carbon atoms) are the same for alkanes having 4-7 carbons. In C8 alkanes three pairs of isomers are reordered which is 1/3 of the eighteen molecules. In C9 only eleven out of thirty five isomers (31%) are ordered in the same manner, whereas in C10 this number is thirteen out of seventy five (only 17%). The complexity patterns that produce similar hierarchical and global ordering are the number of branches, their length and centricity. Thus, relations that are followed by both orderings are, for example

$$n\text{-}C_n < 2M\text{-}C_{n-1} < 2,3MM\text{-}C_{n-2} < ...$$
$$3M\text{-}C_{n-1} < 3E\text{-}C_{n-2}$$
$$2M\text{-}C_{n-1} < 3M\text{-}C_{n-1} < 4M\text{-}C_{n-1} < ... \text{ or}$$
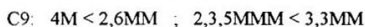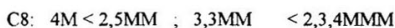$$2,2MM\text{-}C_{n-2} < 3,3MM\text{-}C_{n-2} < 4,4MM\text{-}C_{n-2} < ..., \text{ etc.}$$

Randić suggested several relationships of this type as "cases of no dispute". However, his idea to extend such relationships to orderings influenced by more than one factor is disputable. We will show in the following that the orderings 2,5MM < 2,4MM < 2,3MM < 2,2MM vary with the size of the molecule.

The major difference between the global and the hierarchical ordering is that the latter restricts strongly the competition between the different complexity features. In most cases, the classes of hierarchically ordered isomers are classes that differ by the total number of branches. When in two classes the number of branches is the same, they differ by the number of vertices with the highest degree four (0 and 1 for the dibranched heptanes, 1 and 2 for the four-branched nonanes and decanes). Only a few classes allow competition between the number of branches and the multiple vertex

branching, favoring the latter as more complex; these are the classes having both two and three branches in C8-C10, and the class in C10 having both three and four branches. The central location of vertices, termed vertex centricity [45], is regarded as a complexity factor inferior relative to the number of branches and the presence of multiple vertex branching (see relationship 29).

The global ordering provided by the overall connectivity TC allows considerably higher competition of the complexity factors and the relative importance of these branching patterns varies with the increased size of the molecule. The presence of multiple vertex branching (vertices of degree four) and, particularly, branch centricity are of considerably higher importance and compete even with the number of branches in larger alkane molecules.

The last two trends may be illustrated by comparing the ordering of the monobranched species with the central branch position to that of the 2-,n-1-dibranched species and, correspondingly, by comparing dibranched isomers having a vertex of degree four to tribranched ones with branches attached to different vertices:

$$C8: \quad 4M < 2,5MM \quad ; \quad 3,3MM \quad < 2,3,4MMM$$
$$C9: \quad 4M < 2,6MM \quad ; \quad 2,3,5MMM < 3,3MM$$
$$C10: 2,7MM < 5M \quad ; \quad 2,4,6MMM < 2,3,6MMM < 3,3MM$$

Interestingly, the leading eigenvalue $\lambda\lambda_1$ of the path matrix, introduced by Randić as a very sensitive branching measure [23], does not exhibit these trends. However, the higher complexity of the 3,3MMC7 isomer relative to that of the tribranched 2,3,5-isomer is well reflected by three other topological indices examined by Randić, the Wiener index, the hyper-Wiener index and the largest eigenvaue of the adjacency matrix.

Compare now the relative importance of vertex centrality and multiple vertex branching (vertex degree majorization). Typical examples are the 2,2-dimethylalkanes, the presence of the vertex degree four of the next-to-terminal atom two in which outweighs the branch centricity only in C6 and C7 alkanes. In C8-C10, the increased distance to the graph center displaces the 2,2MM isomer to the area of less complex isomers:

C6: 2,3MM < 2,2MM
C7: 2,4MM < 2,3MM < 2,2MM
C8: 2,5MM < 2,4MM < 2,2MM < 2,3MM
C9: 2,6MM < 2,5MM < 2,2MM < 2,4MM < 2,3MM
C10: 2,7MM < 2,6MM < 2,2MM < 2,5MM < 2,3MM < 2,4MM

This trend of increasing superiority of the vertex centricity factor, as compared to the vertex degree four factor, is mirrored also by the Randić $\lambda\lambda_1$ index, although to a lesser extent. It appears first in C9 alkanes, and the 2,2MM isomer is shifted one place to the right relative to our C9 and C10 sequences.

The C10 alkanes offer also a rare example in which clustering appears as a complexity factor superior over the branch centricity (normally branch centricity is a considerably stronger factor). The TC index of the dimethyl octanes orders them in the sequence shown above to end with the ratio 2,5MM < 2,3MM < 2,4MM. The reason for the reversed order of the 2,4-MM and 2,3-MM isomers is that the more centric position of the branch in position 4 is outweighed by the gain in complexity from having the two branches in the neighboring positions 2 and 3.

6. Hierarchical Versus Global Orderings of Cyclic Graphs Based on Topological Complexity

Traditionally, studies on ordered sets of molecules discuss in detail acyclic molecules while leaving cyclic molecules with less attention. This is partially due to the much higher complexity of cyclic compounds, as well as to the lack of clear criteria for a more detailed graph theoretical analysis. We will approach the problem by applying both the hierarchical and the global ordering of the set of all simple cyclic graphs having five vertices and vertex degrees not exceeding four (FIGURE 2 and TABLE 5. The complete tables containing the complexity index ^eK and the overall connectivity ^eTC for all orders of $e = 0$ to E of these graphs are published elsewhere [30].
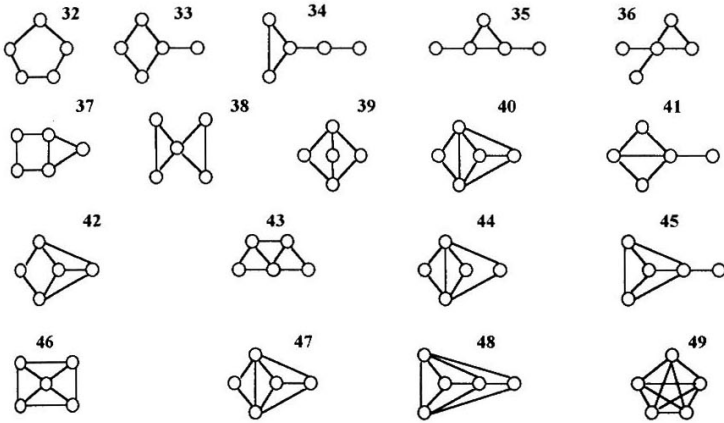
TABLE 5. The ordering of the cyclic graphs **32-49**, as produced by the hierarchical ($^1K,^0TC$; $^2K,$ ^1TC, etc.) and the overall topological complexities (TC, K; the consecutive numbering)

$^1K,^0TC$	5; 10	6; 12	7; 14	8; 16	9; 18	10(20)
Number of cycles	1	2	3	4	5	6

Graph	$^2K,$ ^1TC					
	$K,$ TC	**32** 5; 20	**39** 9; 30	**42** 13; 40	**46** 18; 52	**48** 24; 66	**49** 30; 80
		26; 160	56; 504	111; 1278	230; 3216	477; 7806	973 18180
		34 6; 22	**37** 9; 30	**43** 14; 42	**47** 19; 54		
		29; 190	54; 482	114; 1316	235; 3290		
		33 6; 22	**40** 10;32	**45** 15;44			
		27; 172	57; 522	119;1396			
		35 7; 24	**38** 10;32	**44** 15;44			
		31; 212	55; 492	120;1394			
		36; 8, 26	**41** 11;34				
		33; 230	61; 566				

As seen in TABLE 5, the nondegenerate overall connectivity TC and complexity index K provide a complete ordering of the set of 18 graphs, as expressed by the consecutive numbers of the graphs. The hierarchical ordering of this set starts with the ^1K and ^oTC indices which form six subsets of increasing complexity, each of which has a fixed number of cycles. The vertical ordering within each subset follows the values of the ^2K and ^1TC complexities. The ordering of several pairs of graphs, namely **33, 34; 37, 39; 38, 40;** and **44, 45**, require higher-order complexities to be used. The second-order ^2TC suffices to order them, however, the other complexity criterion ^3K produces a reverse ordering of the first three pairs, and so does ^4K for the pair **42, 43** for which ^3K it is still degenerate. Thus, in sets of cyclic graphs we also can convert the partial hierarchical ordering into a complete one by making use of higher-order complexity terms. The difference relative to the sets of acyclic graphs is in the larger divergence of the hierarchical orderings produced by ^eK and ^{e-1}TC. Generally, cyclic graphs contain more vertices of degree higher than two, and this makes the weight of subgraph connectivity to increase as compared to that of the number of subgraphs. Also, we find in TABLE 5 that the ordering generated by the overall connectivity TC (the consecutive numbering) in six out of the eighteen cases examined deviates from the ^eTC-based hierarchical ordering (the ordering controled by the ^eK deviates only once). In acyclic graphs, the global and the hierarchical orderings coincide up to seven-vertex graphs; we see that in cyclic graphs the divergence between the two orderings starts much earlier. The reason for this different behavior is that cyclic graphs have more edges and, therefore, many more larger subgraphs than the acyclic graphs with the same number of vertices. The global ordering by the K and TC complexities is then determined to a larger extent by the *larger* subgraphs, not by the smallest ones used as criteria in the hierarchical ordering.

In the hierarchy of topological complexity features, the number of cycles is followed by *branching* which provides a further discrimination within the sets of constant number of cycles. The first subset of graphs **32-36** illustrates the trend of increasing complexity in cyclic structures with increasing number of branches: $TC(\mathbf{35, 36}) > TC(\mathbf{33, 34}) > TC(\mathbf{32})$, and a similar relation is valid for K. The other examples of branched cyclic graphs **40, 41, 45** reinforce this observation.

As seen in TABLE 5, the differentiation in complexity caused by the number of branches appears in an order later than that caused by cyclicity, namely, at the first-order TC, and at the second order K. However, as indicated by the *unbranched* graphs **37-39, 42-44,** and **46, 47**, there is another

complexity factor that affects exactly the same order complexites. In identifying this complexity pattern, compare the vertex degree sequences of **42-44**: 33332, 43322, and 44222, as well as those of **46** and **47**, 43333 and 44332, respectively. We find again, as in Section 4.4 and Corollary 6, a pattern of *vertex degree redistribution* with a seniority relationship *3, 3 < 4, 2* that increases topological complexity. This pattern, called after Klein and Babić [1] *vertex degree majorization*, is also responsible for the variation in complexity in the pairs of graphs **40, 41** and **35, 36**. It provides an explanation why graphs **44** and **45** have almost the same *TC* value. The comparison of the vertex degree sequences of these two graphs, 44222 and 43331, respectively, reveals two vertex degree redistributions to occur at a time: *2, 2 < 3, 1* which increases complexity, and *4, 2 > 3, 3* which acts in the opposite direction, and almost entirely compensate the first one.

Importantly, similar patterns of vertex redistribution are associated with the most essential cases of branching, as already discussed for acyclic molecules in Section 4.4. The creation of any new branch at a constant number of graph vertices means just the increase in the degree of one vertex at the expense of the decrease in the degree of another vertex. Therefore, one might consider a possible generalization of the notion of branching. The branching of a molecular skeleton which creates terminal vertices (vertices of degree one) may be regarded as an *external branching,* while the cases of vertex redistribution in cyclic graphs with a constant or null number of external branches may be qualified as *internal branching*, meaning branching of inner vertices. Thus, the cases of distinct first-order *TCs* within the sets of graphs with the same number of vertices and cycles might be treated as different manifestations of this generalized concept of branching. Such ideas agree with Bertz's view of cycles as a kind of branching. Seemingly paradoxical, because cycles are generally believed to have a considerably larger contribution to complexity than branching, such views look acceptable when presenting the branching of a molecule as a result of branching of each of the graph vertices [20].

Another topological feature that influences complexity of cyclic structures, is the manner in which a pair of cycles is connected. We term this feature *cycle connectivity*. As shown in our study on kinetic networks [46-48], complexity of kinetic graphs (measured by the number of spanning trees in the kinetic graph and all of its cyclic subgraphs) tends to increase when going from a pair of cycles attached by a bridge, to those connected by a common vertex, to those sharing a common edge, then two common edges, etc. This topological pattern, if mirrored in a similar manner by our *TC*-index,

would show an increase in complexity in the series $38 \to 37 \to 39$. The inversion of **37** and **38** may be interpreted as an evidence of the inferiority of cycle pair connectivity relative to the pattern of the vertex degree majorization occurring in these two graphs: *3,3* (**37**) < *4,2* (**38**). However, this example may also indicate a potential trend of divergence in the complexity estimates based on the connectivity and on the spanning trees of the cyclic graphs and all of their subgraphs.

7. Concluding Remarks

This study was centered on partial and complete orderings of molecules which result from a set of hierarchical criteria, the hierarchy of the intermolecular order being derived from the intramolecular hierarchical order of topological complexity. Hierarchically ordered sets are less studied than the orderings which result from the numerical values of topological indices and may be termed "global". We have found that the agreement between the two types of ordering weakens with increasing size of the molecules, the divergence in the ordering of acyclic molecules beginning in octane isomers, whereas in cyclic moieties it starts in cyclopentane isomers. The comparative study of the global versus hierarchical orderings sheds some light on the different weight of the contributions coming from small and from large subgraphs. The hierarchical orderings are controlled entirely by the small molecular fragments, therefore these orderings may be considered as "local", in contrast to the large fragments which play a significant role in the global orderings. Related to this, more comparative studies of the global and hierarchical orderings are needed, including a systematic search for all global ordering criteria for which a hierarchical counterpart can be built.

Another problem of importance needs further attention. Inventing more and more sophisticated topological indices we gradually transform partial orderings of sets of compounds into much more discriminated orderings and, for molecules of reasonably small size, into complete orderings. However, molecular properties, with few exceptions for very small molecules, do not follow closely the structural patterns imposed by such ordering criteria. Therefore, one should not dream about finding "the ultimate" ordering criterion which would have no degeneracy, i. e., which would provide a complete ordering only. Molecular descriptors with extremely high ordering potential, if ever found, would be of value only for coding of chemical compounds but not for identifying the patterns in

molecular properties. In contrast, descriptors with a moderate discriminating potential, like the Wiener number, which is highly degenerate in nonanes and decanes, are considerably better in identifying property trends. They are degenerate mainly for isomers which are ordered differently by the different more sophisticated indices. A possible better approach for identifying property trends and for a better property prediction could be the idea of *representative ordering*, obtained by the superimposing of several molecular descriptors constructed from essential graph invariants [6,7], an idea undeservedly forgotten.

References

[1] Klein, D.J. and Babić, D. (1997). *J. Chem. Inf. Comput. Sci. 37*, 656-671.

[2] Ruch, E. (1975). *Theor. Chim. Acta 38*, 167-183.

[3] Ruch, E. and Mead, A. (1976). *Theor. Chim. Acta 41*, 95-117.

[4] Randić, M. and Wilkins, C. L. (1979). *J. Phys. Chem. 83*, 1525-1540.

[5] Randić, M. and Wilkins, C. L. (1979). *Chem. Phys. Lett. 63*, 332-336.

[6] Bonchev, D. and Mekenyan, O. (1984). *J. Chem. Soc. Trans. Faraday* II *80*, 695-712.

[7] Bonchev, D., Kamenska, V., and Mekenyan, O. (1990). *J. Math. Chem. 5*, 43-72.

[8] Trotter, T. (1992). *Combinatorics and Partialy Ordered Sets - Dimension Theory.* The John Hopkins University Press: Baltimore.

[9] Bertz, S. H. (1983). In *Chemical Applications of Topology and Graph Theory.* King, R. B., Ed. Elsevier: Amsterdam, pp. 206-221.

[10] Bonchev, D. and Polansky, O. E. (1987). In *Graph Theory and Topology in Chemistry.* King, R.B. and Rouvray, D. H., Eds. Elsevier: Amsterdam, pp. 126-158.

[11] Bonchev, D. (1990). In *Computational Chemical Graph Theory.* Rouvray, D. H., Ed. Nova: New York, pp. 34-67.

[12] Bonchev, D. and Seitz, W. A. (1996). In *Concepts in Chemistry: Contemporary Challenge.* Rouvray, D. H., Ed. Research Studies Press: Taunton, U. K., pp. 353-381.

[13] Rashevsky, N. (1955). *Bull. Math. Biophys. 17*, 229-235.

[14] Trucco, E. (1956). *Bull Math. Biophys. 18*, 129-135; *18*, 237-253.

[15] Bertz, S. H. (1981). *J. Am. Chem. Soc. 103*, 3599-3601.

[16] Bertz, S. H. (1983). *Bull Math. Biol.*, *45*, 849-855.

[17] Bonchev, D. (1983). *Information Theoretic Characterization of Chemical Structures*. Research Studies Press, Chichester (U. K.), 274 pp.

[18] Bonchev, D., Mekenyan, O., and Trinajstić, N. (1981). *J. Comput. Chem. 2*, 127-148.

[19] Bonchev, D. and Trinajstić, N. (1977). *J. Chem. Phys. 67*, 4517-4533.

[20] Bertz, S. H. (1988). *Discr. Appl. Math. 19*, 65-83 .

[21] Bonchev, D. (1995). *Theochem 336*, 137-156.

[22] Bonchev, D., Mekenyan, O., and Trinajstic, N. (1980). *Int. J. Quantum Chem. 17*, 845-893.

[23] Randić, M. (1998). *J. Math. Chem. 44*, 57-77.

[24] Randić, M. (1997). *J. Chem. Inf. Comput. Sci., 37*, 1063-1071.

[25] Bertz, S.; Herndon, W. C. In *Artificial Intelligence Applications in Chemistry;* American Chemical Society: Washington, D. C., 1986, pp.169-175.

[26] Bone, R. G. A. and Villar, H. O. (1997). *J. Comput. Chem. 18*, 86-107.

[27] Bonchev, D. (1997). *SAR/QSAR Envir. Res. 7*, 23-44.

[28] Bertz, S. H.; Sommer, T. J. (1997). *Chem. Commun.* 2409-2410.

[29] Bertz, S. H. and Wright, W. F. (1998). *Graph Theory Notes New York Acad. Sci. 35*, 32-48.

[30] Bonchev, D. (1999). In *Topological Indices and Related Descriptor*. Devillers, J., Balaban, A. T., Eds. Gordon and Breach: Reading, U. K., pp. 361-401.

[31] Platt, J. R. (1952). *J. Phys. Chem. 56*, 328-336.

[32] Gordon, M. and Scantlebury, G. R. (1964). *Trans. Faraday Soc. 60*, 604-621.

[33] Wiener, H. (1947). *J. Am. Chem. Soc. 69*, 17-20.

[34] Randić, M. (1975). *J. Am. Chem. Soc. 97*, 6609-6615.

[35] Kier, L. B., Hall, L. H., Murray, W. J., and Randić, M. (1975). *J. Pharm. Sci. 64*, 1971-1974.

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

[37] Kier, L. B. and Hall, L. H. *(1986). Molecular Connectivity in Structure-Activity Analysis*, Research Studies Press, Letchworth, U. K., 262 pp .

[38] Mekenyan, O., Bonchev, D., and Balaban, A. T. (1984). *J. Comput. Chem. 5*, 629-639.

[39] Balaban, A. T., Mekenyan, O., and Bonchev, D. (1985). *J. Comput. Chem. 6*, 538-551.

[40] Needham, D. E., Wei, I.-C., and Seybold, P. G. (1988). *J. Am. Chem. Soc. 110*, 4186-4194.

[41] Selected Values of Properties of Hydrocarbons and Related Compounds (1977). American Petroleum Institute. Research Project 44; Pittsburg, Pennsylvania.

[42] Kier, L. B. and Hall, L. H. (1981). *J. Pharm. Sci. 70*, 583-589.

[43] Ruch, E. and Gutman, I. (1979). *J. Comb. Inf. System Sci. 4*, 285-295.

[44] Bonchev, D. (in preparation). In *Mathematical Chemistry Series, Volume 7, Complexity in Chemistry*. Bonchev, D. and Rouvray, D. H., Ed. Gordon & Breach: Reading, U. K.

[45] Bonchev, D. (1989). *Theochem 185*, 155-168.

[46] Bonchev, D., Kamensky, D., and Temkin, O. N. (1987). *J. Math. Chem, 1*, 345-388.

[47] Gordeeva, K., Bonchev, D., Kamenski, D., and Temkin, O. N. (1994). *J. Chem. Inf. Comput. Sci. 34*, 244-247.

[48] Temkin, O. N., Zeigarnik, A. V., and Bonchev, D. (1996). *Chemical Reaction Networks. A Graph Theoretical Approach.* CRC Press: Boca Raton, FL, 286 pp.