Representation of the Molecular Topology of Cyclical Structures by Means of Cycle Graphs. 1. Extraction of Topological Properties

Gonzalo Cerruela García, Irene Luque Ruiz,* and Miguel Ángel Gómez-Nieto

Department of Computing and Numerical Analysis, University of Córdoba, Campus Universitario de Rabanales, Building C2, Plant 3, E-14071 Córdoba, Spain

Received November 10, 2003

In this paper, a new representation model using the existing cycles in the topological structure of the molecules is proposed. Extracting all cycles of a molecule, its topological structure can be represented by means of a weighted, colored, and nondirected graph named "cycle graph", where the nodes represent the cycles in the molecule and the edges the common nodes among those cycles. In this paper, the capacity of cycle graph for the extraction of topological descriptors contributing appropriate measures of complexity, cyclicity, and symmetry of cyclical systems is presented.

1. INTRODUCTION

The representation of the molecular structure of chemical compounds using graphs, named molecular graphs, has been used for the last 50 years in the development of research that embraces many areas of chemistry, biology, pharmacy, etc. Using molecular graphs to represent the topology of the molecules, hundreds of invariants or topological indexes have been described in order to determine certain properties or characteristics of these graphs and, consequently, of the corresponding chemical compounds.[1−3]

Generally, these topological indexes are based on the extraction of some property from the molecular graph that is correlated to some measurement of the distances existing among the graph nodes (minimum, maximum, reciprocal distance, etc.) and the weight or color of the node (atomic weight, electronegativity, etc.), using a suitable representation model of the molecular graph (adjacency, connections, Detour matrix, etc.).[4−7]

Works in QSPR and QSAR make use of these topological descriptors for the extraction of physical, chemical, biological, or topological properties, to describe high correlations for prediction of properties with a reduced set of descriptors.[8,9] Although many of the topological descriptors have been described and proved, and many of them are correlated, it is necessary to continue working in this field due to the following reasons.[8,10]

1. There is improvement in the regressions proposed for the extraction of properties of chemical compounds; researchers wish to enlarge them to larger families of compounds.

2. Depending on which property is studied, it is necessary to use or to define an appropriate descriptor by means of which that property can be correlated,

3. There are topological properties, intrinsically ambiguous, that require the definition of an appropriate descriptor for their prediction.

Measures of topological properties, such as cyclicity, shape, symmetry, complexity, size, etc., have been broadly studied over recent years, and topological descriptors have been proposed for their characterization, due to the influence of these characteristics on molecular properties.[10−17]

The problem in the characterization of these properties may be their own concepts or definitions: What do we understand by shape, complexity, or cyclicity? They are abstract characteristics of the molecular structure that, depending on the paradigm or concept used, will obtain different measures.

In this paper we pay attention to the study of these properties or topological characteristics of cyclical structures, due to the importance of these substructures in the properties of the molecules. So, we consider the information corresponding to all the present cycles in a cyclical structure, and as other authors have proposed,[18−20] we use this information for the construction of an isomorphic graph (cycle graph) from a molecular graph.

Given a graph in which the nodes represent each one of the cycles of the cyclical structure, and the edges the relationships among the cycles, cycle graph (CG), it is feasible that proposing topological descriptors allows the extraction of structural properties of the cyclic compounds, and their application to QSPR problems. In the following paper of the series, we will present the capacity of a cycle graph in clustering and screening processes of chemical databases.[21]

The manuscript has been organized in the following way: in section 2 the representation model of the cyclical structures by means of a cycle graph (CG) is described, as well as other kinds of representation based on the present cycles of the molecular structure. In section 3 a series of topological invariants are proposed, which will be later validated in section 4 using different test sets of cyclical structures. Finally, the results are commented on, and future works in this line are presented.

2. REPRESENTATION OF CYCLICAL STRUCTURES THROUGH CYCLE GRAPHS

A molecular graph is a nondirected graph $G = (N, E, F)$, where N is the set of nodes representing the atoms, E is the set of edges representing the relationships among atoms, and

* Corresponding author phone: +34-957-212082; fax: +34-957-21863; e-mail: ma1lurui@uco.es.

10.1021/ci034256a CCC: $27.50
Published on Web 02/10/2004

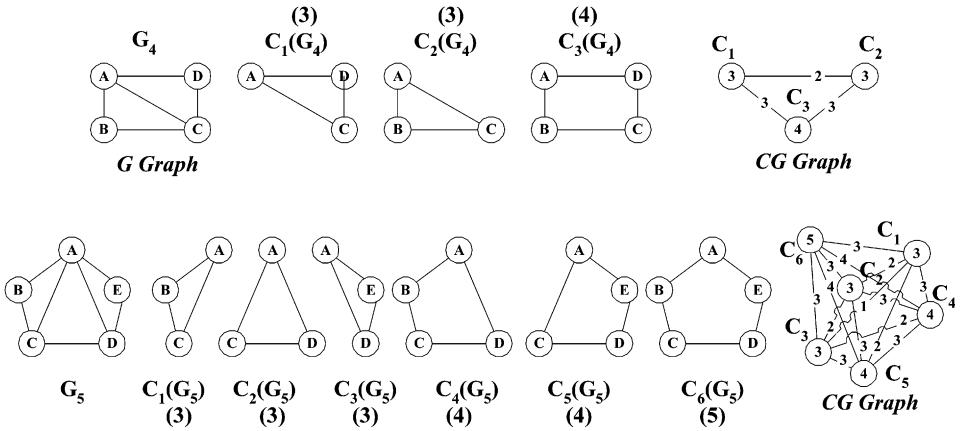


Figure 1. Example of CG graphs for two cyclical structures.

F is a function that represents the connection type (edge in the graph) among the atoms of the molecular structure. When all the nodes of the *G* graph at least belong to a cycle, *G* represents a cyclical structure. In this paper, we only consider this kind of graph.

It is usual to represent a *G* graph by means of another *G′* graph, generally simpler than *G* (smaller number of elements) and containing a subset of the information existing in *G*. In the reduction process of *G*, certain properties will be hidden in *G′*. *G′* is a homomorphic representation of *G* that can be conveniently used for the characterization of *G*.[18]

Dury[19] proposes the reduction of a molecular graph by means of the representation of its cyclical component (GSCE), graph of smallest cycles and edges, and its acyclic component (GAS), graph of acyclic subtrees. The GSCE component is built from the relationships among the set of smallest cycles and edges (SSCE)[19] of the *G* graph, which, depending on the type of graph, can be equal to or different from the SSSR[22] and ESER.[23]

The GAS component is simply obtained by representing as only one node each independent acyclic chain (not connected to each other) present in the *G* graph. From these two components (GSCE and GAS), a new reduced graph RG is obtained that is efficiently used as selection criterion for recovery and search processes in databases of organic compounds.[19]

It is evident that the construction of the RG graph causes a high loss of information corresponding to the cyclicity of the *G* graph, since only a subset of the present cycles in *G* is considered.

However, it is possible to avoid this lack of information using all the present cycles in *G* graph for the construction of the homomorphic *G′* graph and, later, to define topological invariants on the *G′* graph allowing it to derive suitably topological properties of *G* graph.

So, given a cyclical structure represented by a molecular graph *G*, all the present cycles in the graph can be extracted efficiently using the algorithms proposed by the authors,[24,25] and knowing the set of cycles *C*, a cycle graph CG can be built with the following characteristics:

□ The CG graph is a nondirected graph, as the *G* graph.

□ As many nodes as elements are present in the *C* set composing the CG graph.

□ The CG graph is a colored graph; each node is identified by the size of the cycle (nodes number of each cycle in the *G* graph).

□ The edges of the CG graph are labeled, and they represent the relationships among the cycles. Given two nodes $i, j \in$ CG, that represent cycles C_i and C_j present in *G*, the edge that relates both nodes is labeled by the number of common nodes to C_i and C_j, that is, $C_i \cap C_j$.

□ The CG graph can be represented using a symmetrical matrix M_{CG} whose elements $M(i, i)$ take a value equal to the nodes number of *G* present in C_i cycle, and any element $M(i, j)$ is equal to the number of common nodes between the C_i and C_j cycles in the *G* graph.

Figure 1 shows an example of the CG graph for two cyclical structures. As can be observed, the nodes number of the CG graph can be higher (G_5), lower (G_4), or equal to the nodes number of the *G* graph.

For complex cyclical structures with a high number of interrelated cycles (K_n graphs), the CG graph is more complex than the *G* graph. This complexity allows us to consider, in the proposed model CG, that the nodes (atoms) that intervene in more cycles in the *G* graph have a higher influence on the characteristics and properties of the corresponding molecular structure.

2.1. Equivalent Cycle Graphs. A CG graph represents each one of the present cycles in a cyclical structure and the structural relationships (nodes or common atoms) existing among the cycles. It is evident that, given *C* sets of cycles of *G*, a series of subsets CE_i can be extracted whose elements have the same behavior, which is denominated as classes of equivalence of cycles.

The concept and use of equivalence classes in graph theory, and mainly in its applications to molecular graphs, is very extended, its efficiency in many cases being demonstrated.[26−28]

In the representation model proposed, an equivalence class of the CG graph is considered when it is composed for the following:

1. nodes with the same color in the CG graph, therefore representing cycles composed by same number of atoms;

2. nodes of the CG graph participating in the same number and type of relationships with the remaining nodes.

So, we can define an equivalent cycle graph (ECG) as a graph whose nodes represent the equivalence classes of the CG graph, and whose edges represent the relationship among nodes: number of common nodes among equivalent cycles.

Shown in the paper (i.e., Tables 1 and 3) are the characteristics of ECG graphs obtained from the CG graph. The use of ECG graphs introduces a simplification in the

Table 1. Study of the G_{5XX} Graphs in a Cyclical Graph Representation[a]

	Graphs	N	Σ	Γ	E	W	w	Δ	Cycles		
	CG Graphs										
G_{501}	G_{501}	1	4	0.000	1	0	0.0	0.0	5/1		
	G_{511}	3	7	2.720	1	3	9.0	2.0	3/1	4/1	5/1
G_{511}	G_{521}	6	11	14.157	1	15	45.0	10.0	3/3	4/2	5/1
	G_{522}	7	16	17.738	2	21	68.5	13.1	3/2	4/3	5/2
G_{521}	G_{531}	12	23	58.241	2	66	203.0	43.2	3/5	4/5	5/2
	G_{532}	13	30	66.895	4	78	312.0	39.0	3/4	4/5	5/4
G_{522}	G_{541}	22	49	197.487	6	231	924.0	115.5	3/7	4/9	5/6
	G_{551}	37	88	553.837	12	666	2664.0	333.0	3/10	4/15	5/12
	ECG Graphs										
G_{501}	G_{501}	1	4	0.000	1	0	0.0	0.0	5/1		
	G_{511}	3	7	2.720	1	3	9.0	2.0	3/1	4/1	5/1
G_{521}	G_{521}	4	8	4.882	1	6	18.0	4.0	3/2	4/1	5/1
	G_{522}	4	9	4.965	1	6	19.5	3.8	3/1	4/2	5/1
G_{522}	G_{531}	6	11	14.157	1	15	47.5	9.6	3/3	4/2	5/1
	G_{532}	4	9	4.965	1	6	24.0	3.0	3/1	4/2	5/1
	G_{541}	5	10	8.334	1	10	40.0	5.0	3/2	4/2	5/1
	G_{551}	3	7	2.267	1	3	12.0	1.5	3/1	4/1	5/1
	SCG Graphs										
G_{531}	G_{501}	1	4	0.000	1	0	0.0	0.0	5/1		
	G_{511}	2	3	1.207	1	1	3.0	0.7	3/1	4/1	
	G_{521}	3	3	4.000	3	3	9.0	2.0	3/3		
G_{532}	G_{522}	3	4	3.414	1	3	10.5	1.8	3/2	4/1	
	G_{531}	4	4	7.000	4	6	19.5	3.8	3/4		
	G_{532}	4	4	8.000	4	6	24.0	3.0	3/4		
G_{541}	G_{541}	5	5	12.000	5	10	40.0	5.0	3/5		
	G_{551}	6	6	18.000	6	15	60.0	7.5	3/6		
	ICG Graphs										
G_{551}	G_{501}	1	4	0.000	1	0	0.0	0.0	5/1		
	G_{511}	2	3	1.207	1	1	3.0	0.7	3/1	4/1	
	G_{521}	3	3	4.000	3	3	9.0	2.0	3/3		
	G_{522}	5	8	9.855	3	10	33.5	6.1	3/2	4/3	
	G_{531}	5	5	12.000	5	10	31.5	6.4	3/5		
	G_{532}	5	6	12.828	1	10	40.0	5.0	3/4	4/1	
	G_{541}	7	7	27.000	7	21	84.0	10.5	3/7		
	G_{551}	10	10	60.000	10	45	180.0	22.5	3/10		

[a] *N*, nodes number; Σ, global complexity; Γ, relational complexity; *E*, encircling number; *W*, Wiener index; *w*, detour index; Δ, Randic index. Last columns inform of the size/number of the cycles corresponding to the different models.

representation model (a generalization from the point of view of abstraction[29]), and its efficiency will be discussed in this paper.

2.2. Other Representations Based on Cycles. The representation of the molecular structure by means of a CG (or ECG) graph based on the present cycles in the molecule is an open model that allows consideration of any cycles set of the molecular graph *G*. In our paper we have proposed the use of all the cycles (CG graph) and the equivalent cycles (ECG graph); however, other authors have used the SSSR set or other more reduced cycles set for the study of molecules containing cycles.[30−32]

In order to validate our work and to demonstrate that these reduced sets on occasions are not appropriate, two other homomorphic representations of the molecular *G* graph have been built based on these sets:

- SSSR cycle graph (SCG), in which only the cycles belonging to the SSSR set (smallest set of smallest rings) are considered;
- lineally independent cycle graph (ICG), in which case all lineally independent cycles present in *G* are considered. Two cycles C_1 and C_2 are lineally independent if:
  - ○ $C_1 - (C_1 \cap C_2) \neq \emptyset$ or $C_2 - (C_1 \cap C_2) \neq \emptyset$;
  - ○ $C_1 - (C_1 \cap C_2) = C_2 - (C_1 \cap C_2) = \emptyset$, but the order of the nodes involved in C_1 and C_2 graphs is different.

As shown in Figure 2, each one of the representations based on cycles described (CG, ECG, SCG, and ICG) considers different sets of cycles present in the molecular graph. This information is stored in the corresponding *M* matrices whose values are also shown in Figure 2 for some graph example.

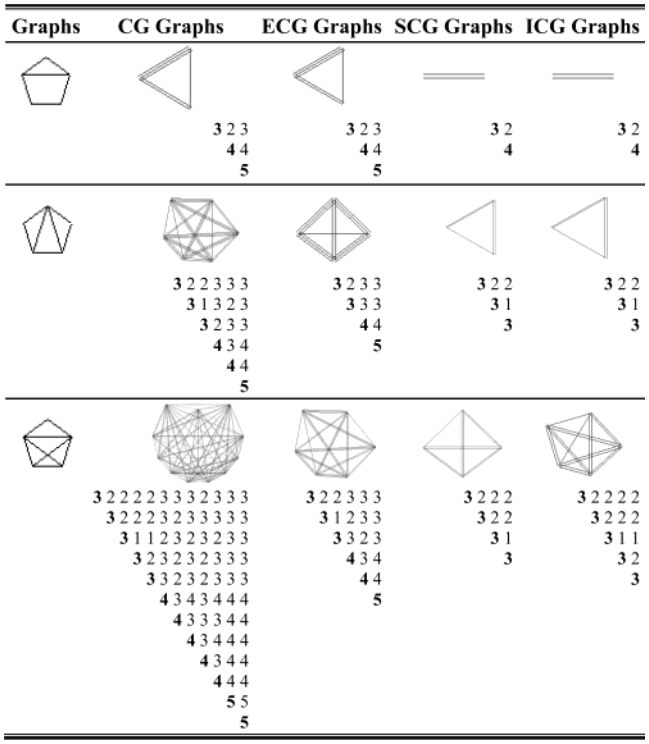


Figure 2. Graphs based on cycles of the different models studied for some example cyclical structures.

The complexity of the graphs based on cycles depends on the number and characteristic of the present cycles in the molecular graph. Evidently, the graphs ECG, SCG, and ICG are always less complex than the CG graph, which is more complex than the molecular graph for the majority of cyclical structures.

The ICG graphs have a close or higher complexity than the SCG graphs. This is because the ICG model considers a higher number of cycles than the SSSR model. In all the cases the cycles considered in the ICG model include the SSSR set and other sets of elementary cycles such as ESER and SSCE, which are equal to the SSSR for many cyclical structures. So, the study of the SCG and ICG models allows us to interpolate our results to other models based on sets of elementary cycles.

3. TOPOLOGICAL INVARIANTS BASED ON CYCLE GRAPHS

The use of topological properties for the obtaining of correlations with physical, chemical, and biological properties is the basis of the research in QSPR. This research line considers that the characteristics of the molecular structure, such as size, shape, complexity, cyclicity, grade of flexibility, presence of chains, heteroatoms, etc., determine in higher or smaller measure the molecular properties. However, some of these characteristics or topological properties are difficult to calculate, simply because they are difficult to define and to quantify. Recent research has shown the complexity in the definition and quantification of structural characteristics,

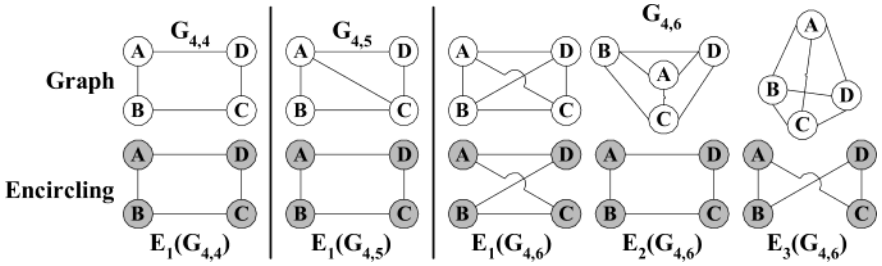


Figure 3. Encirclings for different $G_{4,X}$ graphs.

such as complexity, shape, symmetry, etc., by attempting to get these measurements to satisfy the following requirements:[9,10,13]

1. There is a dependency on the atomic composition and existing relationships in the molecule, that is size or number of atoms, cycles, and chains number, heteroatoms, and unsaturations present, etc.

2. A mathematical formulation allows its calculation and reproduction by means of an algorithm.

3. The ordination of a family of chemical compounds (molecular structures) is allowed in a consistent and logical order.

4. The invariants or structural characteristic defined should be able to be applied to relatively wide families of chemical compounds and not only to small groups of molecular structures with very specific characteristics.

Due to the immense diversity of existing molecular structures (and those that will be discovered), this last characteristic is the most complex. Recently, sets of invariants have been proposed in order to obtain measurements of complexity, symmetry, shape, etc., of chemical structures and have tried to correlate these invariants with some properties of chemical compounds, obtaining very good results with very defined families, although when the set of compounds used in the regression is extended, the correlations diminish.

Maybe the problem, like other authors propose, is due to trying to force the obtaining of lineal regressions that give place to poor correlations, instead of using a wider number of invariants, and to obtain multivariate or quadratic regressions.[5]

On the basis of the information supported by the cycle graphs CG and ECG, in this paper we propose a set of invariants that can be applied to the obtaining of these characteristics for cyclical uncolored structures. The extension of the measures of these characteristics to any structure is the object of our research, and for that, as we have argued previously, it is convenient to make use of other invariants described in the bibliography, whose results will be the subject of a future publication.

3.1. Cardinality of the Cyclical Structure. It is evident that the number and composition of the cycles in a cyclical structure can be a suitable measurement to express characteristics of the molecular structure, as well as of the complexity or cyclicity.

Traditionally, different subsets of the total cycles present in a cyclical structure have been used in order to characterize these structures (SSSR, ESER, SSCE, etc.). The use of these subsets, from the point of view of the abstraction, supposes a generalization and, therefore, the lack of information in the representation of the cyclical structure, and still more, knowing that the SSSR, for example, is not a unique set. Since with the algorithms[24,25] proposed by the authors the calculation of all the present cycles in the cyclical structure has an acceptable computational cost, it is possible to know the cardinality (total number of present cycles) and composition (size and nodes) of all cycles.

Knowing all the present cycles of a *G* graph we can build the cycle graph (CG) that is represented by means of *M* matrix; a symmetrical matrix where the elements $M(i, i)$ store the size of the C_i cycle, and the elements $M(i, j)$ store the common nodes between the cycles C_i and C_j.[21,24]

3.2. The Encircling and Components of the Cyclical Structure. The encircling and perimeter of a graph has been broadly tried in many publications,[11,33−35] being patent that when the *G* graph is not planar, the concept of encircling is difficult to justify. Lipkus[35] defines the perimeter of a cyclic system (*P*) as the edge number of the encircling, calculating it in the way $P = 2E - S$, where *E* is the number of edges in the system, and *S* is the number of total edges of the SSSR set.

In a cyclical structure's context, a shape measure (a concept also broadly discussed) can be given in function of the encircling, which is necessary to define in the context of our model based on the use of the CG and ECG graphs.

Thus, we consider encircling of a *G* graph as a Hamiltonian cycle,[36] that is, a cycle formed by all the nodes (not all the edges) present in the graph. All encircling can be easily obtained by analyzing the main diagonal of *M* matrix, being formed by those cycles, elements $M(i, i)$, whose value is equal to the node number of the cyclical structure *G*.

According to the theorem of Euler,[36] a graph is plane if it satisfies the following expressions (for $n \geq 3$):

$$e \leq 3n - 6 \text{ (if } G \text{ contains cycles of size 3)} \quad (1)$$

$$e \leq 2n - 4 \text{ (if } G \text{ does not contain cycles of size 3)} \quad (2)$$

where *n* is the nodes number and *e* is the number of edges.

According to the theorem of Kuratowski,[36] a graph is plane if and only if it does not contain homomorphic subgraphs to K_5 and $K_{3,3}$.

The encircling number of a graph is suitable to give us a relative measurement of the shape and symmetry of the graphs. So, when a graph has only one encircling, only one Hamiltonian cycle, it satisfies the Euler and Kuratowski theorems, and therefore the graph is planar (the molecular structure that it represents is plane). Graphs with more than one encircling can be planar or not, giving the number of encircling a measure of the cyclical structure shape of the graph.

Figure 3 shows different $G_{4,X}$ graphs. All these graphs are planar. $G_{4,4}$ and $G_{4,5}$ only have one encircling, but $G_{4,6}$ has

three encirclings, giving this information an idea of the most complex shape of this graph (as it is represented in 3D in Figure 3).

The distribution of cycles in the cyclical structure can be used as a measure of the shape as well as the complexity and cyclicity of the cyclical systems. In a cyclical structure the set of all cycles can be distributed in the following ways:

(a) Two or more cycles have two or more common nodes, indicating at least the presence of a cycle that embraces all the nodes of these cycles. Traditionally, this set of cycles is named fused cycles.

(b) Two or more cycles share only one node, and therefore, there do not exist cycles containing all nodes.

We can define a component of the cyclical structure as a set of cycles whose encircling only maintains one and only one common node with any other cycle of the cyclical structure not belonging to the set defined by the encircling.

3.3. Complexity of the Cyclical Structure. It is accepted that the complexity of a molecular structure is directly related to the existence, number, and size of the cycles in the structure (besides other factors).[9,10]

Knowing all cycles C_i of the G graph, the number of existing cycles of a given size is also known. From this information represented in the M matrix, two new parameters can be proposed related to the complexity (and cyclicity) of the cyclical structure.

Global complexity (Σ) of the cyclical structure, a measurement of the existence and composition of the cycles, is defined as:

$$\Sigma = \sum_{i=1}^{i=n} 2^{M(i,i)-3} \tag{3}$$

where $M(i, i)$ stores the size of the cycle i, and n is the number of cycles. The exponent $M(i, i) - 3$ is because the minimum size of a cycle is 3.

The global complexity contributes an idea of the complexity of the cyclic system. As the size and number of cycles increase, Σ also increases. However, this value does not include the relationships among the cycles.

Thus, we define the relational complexity (Γ) in the following way:

$$\Gamma = \sum_{i=1}^{i=n} \sum_{\substack{j=1 \\ j\neq i}}^{j=n} \frac{2^{M(i,i)-3/M(i,j)}}{M(i,j)} \tag{4}$$

where $M(i, i)$ stores the size of the cycle i, $M(i, j)$ stores the common nodes between the cycles i and j, and n is the number of cycles. Γ takes into account the relationships among the present cycles in the graph, and it is calculated through a simple inspection of the M matrix.

With the objective of comparing the complexity values among different cyclical structures, throughout the manuscript we will use the normalized measure of the complexity, which is obtained by dividing the value of Σ for the corresponding value for the K_n graph with the same number of nodes, that we will represent for Σ^+.

3.4. Minimum and Maximum Distances between Cycles. Given two cycles $C_i(n_i, e_i)$ and $C_j(n_j, e_j)$, the minimum distance (D_m) between two cycles is defined as the minimum number of edges that is necessary to travel from some n_i node to any n_j node for all $n_i \neq n_j$. Evidently, $D_m(C_i, C_j) = D_m(C_j, C_i)$.

As in the cyclical system, the relationships among the cycles are very diverse (fused cycles, spiro cycles, embracing cycles), so it is considered that $D_m = 1$, when C_i and C_j have common edges.

Given two cycles $C_i(n_i, e_i)$ and $C_j(n_j, e_j)$, the maximum distance (D_M) between both cycles is defined as the minimum distance among a node $n_i \in C_i$ to any node $n_j \in C_j$, for that n_i node that maximizes the distance between the C_i and C_j for everything n_j node. Evidently, $D_M(C_i, C_j) \neq D_M(C_j, C_i)$.

The calculation of the maximum distances is done as follows:

- The maximum distances between the C_i and C_j cycles are calculated.
- The $n_i \in C_i$ node that maximizes the maximum distance is obtained.
- The minimum distance of n_i node is calculated to any $n_j \in C_j$ node.

From this information can be built the minimum (D_m) and maximum (D_M) distance matrices, and as we will describe in section 4, invariants can be extracted to be applied to determine structural characteristics of the cyclical systems.

4. ANALYSIS OF THE RESULTS

For a consistent set of cyclical structures with different characteristics the information corresponding to all the cycles has been extracted. With this information, the graphs of cycles have been built for all the cycles (CG), the equivalent cycles (ECG), the cycles of the SSSR set (SCG), and the lineally independent cycles, LICS (ICG). Thus, we calculate the corresponding M matrices and the following invariants:[1,3,7]

- encircling or Hamiltonian cycle;
- global complexity, Σ_{CG}, Σ_{ECG}, Σ_{SCG}, Σ_{ICG};
- relational complexity, Γ_{CG}, Γ_{ECG}, Γ_{SCG}, Γ_{ICG};
- Wiener index, W_{CG}, W_{ECG}, W_{SCG}, W_{ICG}, obtained from the corresponding Minimum distance matrices;
- Detour index, w_{CG}, w_{ECG}, w_{SCG}, w_{ICG}, obtained from the corresponding Maximum distance matrices;
- Randic index, Δ_{CG}, Δ_{ECG}, Δ_{SCG}, Δ_{ICG}, obtained from the corresponding matrices.

So, the minimum, maximum, and ratio distances matrices are built (for each model, CG, ECG, SCG, and ICG), and the corresponding indexes are obtained as follows:

$$W_{\text{model}} = \frac{1}{2}\sum_{i=1}^{i=n}\sum_{j=1}^{j=n} D(i,j) \tag{5}$$

$$w_{\text{model}} = \sum_{i=1}^{i=n}\sum_{j=1}^{j=n} DD(i,j) \tag{6}$$

$$\Delta_{\text{model}} = \sum_{i=1}^{i=n}\sum_{j=1}^{j=n} \frac{D(i,j)}{DD(i,j)} \tag{7}$$

where n is the number of cycles corresponding to the considered model, D is the symmetric minimum distances matrix in which $D(i, j) = D_m(C_i, C_j)$, and DD is the maximum distances matrix in which $DD(i, j) = D_M(C_i, C_j)$.

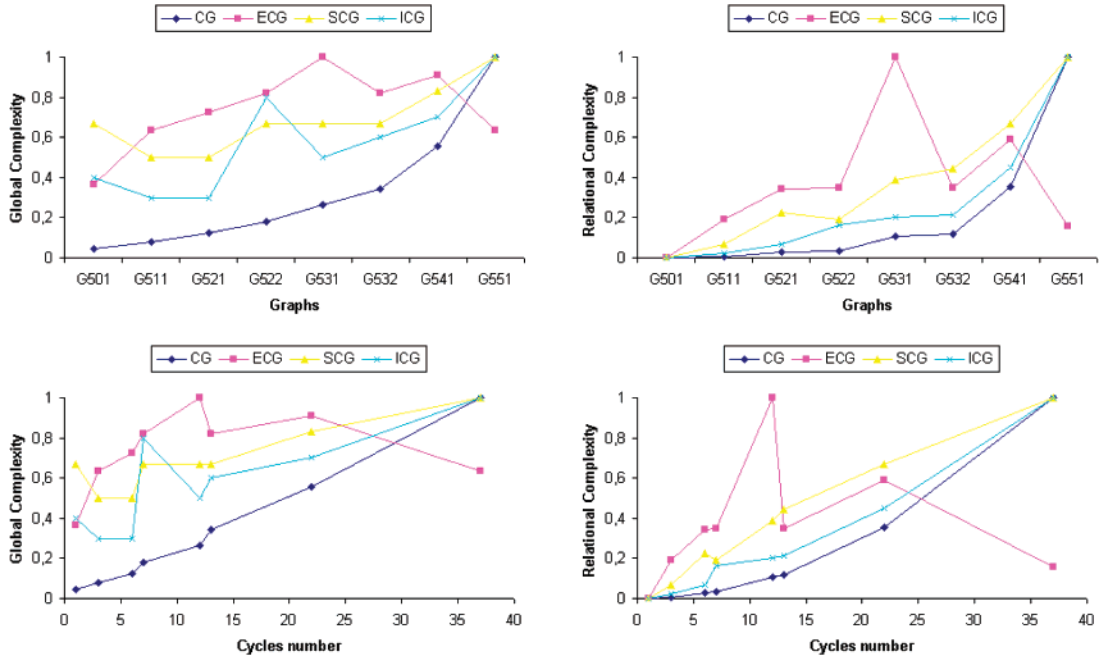


Figure 4. Behavior of proposed topological indexes for G_{5XX} graphs (all the values have been normalized in the interval [0, 1]).

In this section, we will analyze the behavior of these descriptors in order to characterize and to analyze the topological properties of cyclical structures.

4.1.The Complexity of the Simple Cyclical Systems. It is evident that the number and size of cycles in a graph determine the cyclical complexity of the graph. If we observe Table 1 and Figure 4, we can appreciate that the measure of the global complexity (Σ_{CG}) is suitable to determine the cyclical graph complexity. When the number and size of the cycles increases, Σ_{CG} also increases. This behavior is also observed for the relational complexity (Γ_{CG}) and W_{CG}, w_{CG}, and Δ_{CG}.

Σ_{CG} has a lineal behavior ($y = -1.27 + 2.36x$, $r^2 = 0.9932$), also the normalized measurement of cyclicity $\Sigma_{CG}/\Sigma_{CG}^{+}$, with regard to the number of nodes of CG graph (number of total cycles in the original graph) increasing with the graph cyclicity (complexity).

In Figure 4 it is appreciated that Γ_{CG} behaves in a similar way to Σ_{CG}, although it is lightly affected by the existing relationships among cycles (i.e., next values for the G_{531} and G_{532}). On the other hand, ECG graphs allow the extraction of added information on complexity and cyclicity (see Table 1 and Figure 4). Σ_{ECG} value for G_{531} graph is higher than for the remaining graphs, because this graph contains more nodes and cycles than the remaining ECG_{5XX} graphs, since the G_{531} graph has the higher number of equivalence classes.

Therefore, what conclusion can be extracted from this behavior? The direct derivation of this information is that the use of the Σ_{ECG} contributes to a direct measure of the symmetry related to the complexity/cyclicity of the molecular structures. When Σ_{ECG} increases, this complexity also increases. Graphs with high values of Σ_{ECG} mean that these graphs have more equivalence classes of cycles and therefore fewer cycles with the same characteristics. We can ague that in cyclical structures with low values of Σ_{ECG} few diversities of cycles exist and therefore these structures have more symmetry.

This fact became patent for the global complexity values shown in Table 1 and Figure 4. For instance, G_{531} graph has fewer cycles (12) than other graphs as G_{532} (13), G_{541} (22), and G_{551} (37) presenting a higher value of Σ_{CG}. So, G_{531} has a lower cyclicity and complexity than the other graphs. However, G_{531} has a higher value of equivalence classes of cycles (cycles in the graph are very different) presenting a higher value of Σ_{ECG} and, therefore, a higher diversity of cycles, that is, a lesser symmetry.

Γ_{ECG} behaves in a similar way to Σ_{ECG}, but even contributing to a better discrimination. However, the SCG representation model is not suitable for obtaining a measure of the complexity of the cyclical structures, being unable to discriminate different structures and presenting a high degeneration in the values of the proposed descriptors. ICG model behaves similarly (see Figure 4). This fact confirms that the use of a reduced set of cycles (SSSR, LICS, etc.) is not adequate for the calculation of topological descriptors, due to lack of information in considering only one subset of all the cycles.

In order to compare our descriptors with previous ones proposed, in Table 2 are shown the values for some descriptors extracted from the bibliography[12,13,36,37] corresponding to the same set of G_{5XX} graphs, showing similar behavior although certain discrepancies exist. Some descriptors of Table 2 (e.g., W, κ) are not able to discriminate all G_{5XX} graphs, for instance G_{521} and G_{522} graphs, and G_{531} and G_{532} graphs.

Also, we can observe in Figure 5 and Table 2 that the order proposed by some of these descriptors undergoes alterations with the order determined for Σ_{CG}, as is shown by the G_{531} and G_{532} graphs. G_{532} has one cycle more than G_{531} and, therefore, more relationships with the remaining cycles that make all the indexes described in this paper show a higher complexity in G_{532}, as the γ descriptor shows. Notice that G_{532} presents 2 encircling more than G_{531} which induces

Table 2. Topological Index Corresponding to the Graphs of Table 1[a]

graph	twc	*W*	p2	γ	*k*	TC	AV	Cx	λ_1	wcx	*D*/DD	$\Sigma_{CG}/\Sigma_{CG}^{+}$	trees
G_{501}	75	15	5	0	0.2	160	25.000	15.0	5.000	2.00000	4.58	0.045	5
G_{511}	150	14	9	0.32	0.4	482	31.500	89.5	18.750	2.48119	3.92	0.080	11
G_{521}	269	13	14	0.48	0.6	1316	38.500	168.5	23.750	2.93543	3.58	0.125	21
G_{522}	247	13	13	0.56	0.6	1278	38.000	141.5	22.250	2.85577	3.42	0.182	24
G_{531}	421	12	19	0.72	0.8	3290	45.500	240.0	26.500	3.32340	3.08	0.261	40
G_{532}	390	12	18	0.76	0.8	3216	45.000	166.5	18.750	3.23607	3.00	0.341	45
G_{541}	600	11	24	0.88	1.0	7806	52.500	235.5	31.750	3.64575	2.75	0.557	75
G_{551}	850	10	30	1.00	1.2	18180	60.000	170.0	12.000	4.00000	2.50	1.000	125

[a] Extracted from refs 12, 13. Abbreviations are as follows: twc, total walk count; *W*, Wiener index; p2, number of paths of length 2; γ, Randic complexity; κ, rough complexity; Bonchev's TC (sum of all total adjacencies of all connected subgraphs), AV, and Cx, complexity measures proposed by Randic; λ_1, the first (largest) Eigenvalue of adjacency matrix; wcx, walk complexity; *D*/DD, distance/detour quotient index; $\Sigma_{CG}/\Sigma_{CG}^{+}$, normalized global complexity; trees, number of spanning trees.

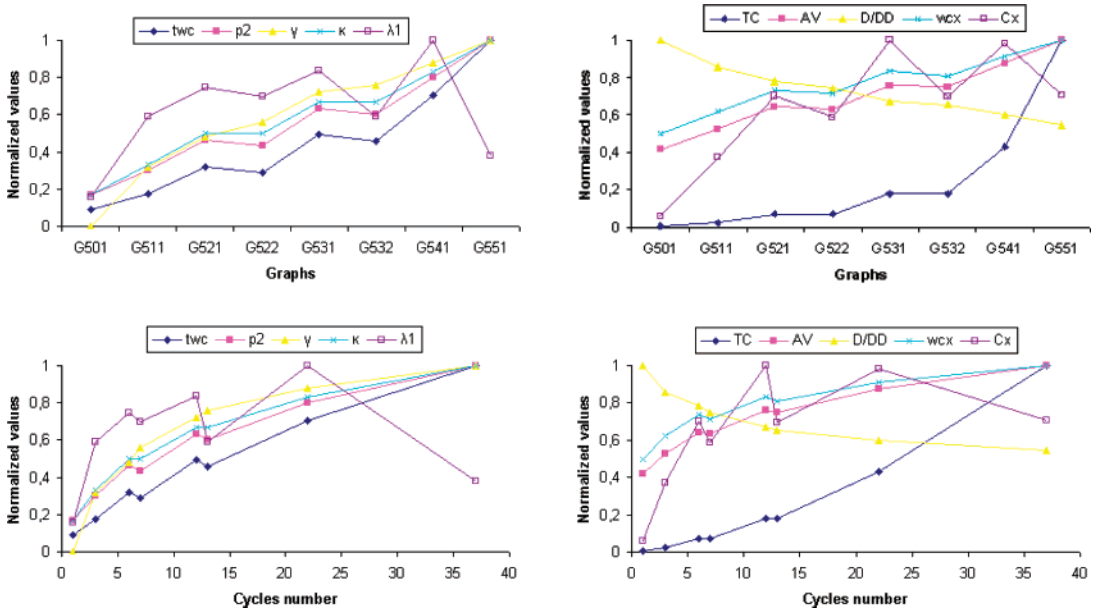


Figure 5. Behavior of topological indexes of Table 2 for G_{5XX} graphs (all the values have been normalized in the interval [0, 1]).

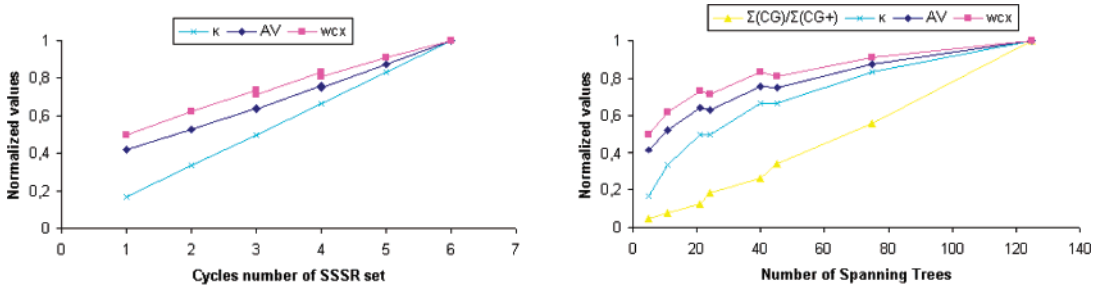


Figure 6. Behavior of topological indexes studied for G_{5XX} graphs with regard to number of cycles of SSSR set and number of spanning trees (all the values have been normalized in the interval [0, 1]).

higher graph cyclicity. To this aspect we will return later on in this paper.

The different behavior of these descriptors to contribute a measure of the cyclicity and complexity of the cyclical structures can be attributed, among other reasons, to the following.

• Some descriptors of Table 2 show a poor lineal behavior regarding the total number of cycles (see Figure 5), while for other descriptors there does not exist a clear relationship (e.g., *Cx*, λ_1). This fact due to these descriptors does not take directly into account the cyclicity of the structures, but only measures of distances among nodes, which are highly influenced by the topology (e.g., twc, wcx, *D*/DD, etc.).

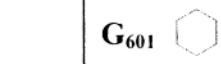
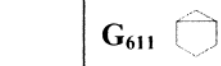
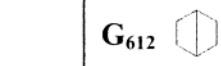
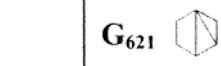
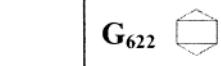
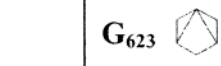
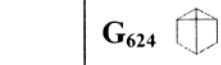
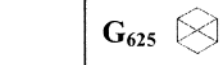
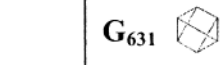
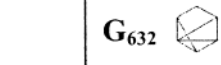
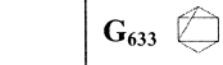
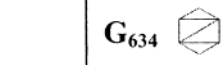
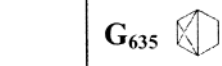
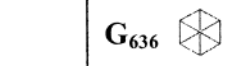
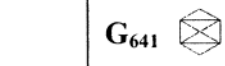
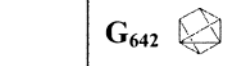
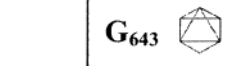
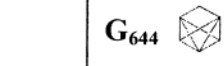
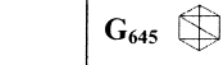
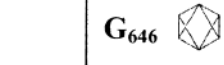
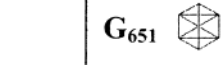
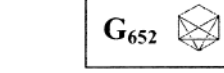
• Other descriptors take into account aspects related to the diversity/symmetry of the graph structure, as λ_1 or *Cx* that consider in its calculation the equivalence classes, presenting for this reason a close behavior to Γ_{ECG}.

• And other descriptors are directly influenced by the SSSR set presenting a close behavior to the SCG and ICG models (e.g., wcx, κ, and AV).

These considerations are shown in Figure 6 where we can observe the lineal behavior of wcx, κ, and AV with regard to the cardinality of SSSR set.

It is also interesting to highlight the linear behavior ($y = 0.008x - 0.0222$, $r^2 = 0.9946$) of $\Sigma_{CG}/\Sigma_{CG}^{+}$ (also Σ_{CG}) with regard to the number of spanning trees, in contrast with the

Table 3. Study of the G_{6XX} Graphs for the CG and ECG Models

	Graphs	N	Σ	Γ	E	W	w	Δ	Cycles (size/number)			
	CG Graphs											
G_{601} 	G_{601}	1	8	0.000	1	0	0.0	0.0	6/1			
	G_{611}	3	13	3.067	1	3	10.0	1.8	3/1	5/1	6/1	
G_{611} 	G_{612}	3	12	2.850	1	3	9.0	2.0	4/2	6/1		
	G_{621}	6	18	16.093	1	15	48.0	9.5	3/2	4/2	5/1	6/1
G_{612} 	G_{622}	6	20	13.000	1	16	47.0	10.3	3/2	4/1	5/2	6/1
	G_{623}	6	20	15.000	1	15	52.5	8.8	3/2	4/1	5/2	6/1
G_{621} 	G_{624}	7	25	18.409	1	21	86.5	10.3	3/1	4/2	5/3	6/1
	G_{625}	7	26	18.345	2	21	84.0	10.5	4/5	6/2		
G_{622} 	G_{631}	14	56	75.642	3	92	455.0	36.8	3/2	4/3	5/6	6/3
	G_{632}	13	49	71.440	3	78	390.0	31.2	3/3	4/3	5/4	6/3
G_{623} 	G_{633}	12	41	59.825	2	67	224.0	40.9	3/3	4/3	5/4	6/2
	G_{634}	10	26	46.113	1	46	140.0	29.8	3/4	4/3	5/2	6/1
G_{624} 	G_{635}	12	36	64.390	2	66	208.0	42.7	3/4	4/4	5/2	6/2
	G_{636}	15	66	86.404	6	105	420.0	52.5	4/9	6/6		
G_{625} 	G_{641}	19	52	160.794	2	172	711.5	83.1	3/6	4/7	5/4	6/2
	G_{642}	23	88	223.395	5	254	1265.0	101.6	3/4	4/6	5/8	6/5
G_{631} 	G_{643}	21	69	189.634	3	211	1024.5	86.9	3/5	4/6	5/7	6/3
	G_{644}	22	68	210.146	2	231	924.0	115.5	3/4	4/8	5/8	6/2
G_{632} 	G_{645}	23	88	223.395	5	254	1265.0	101.6	3/4	4/6	5/8	6/5
	G_{646}	22	80	204.129	4	233	733.0	150.0	3/4	4/6	5/8	6/4
G_{633} 	G_{651}	39	156	641.512	10	742	3705.0	296.8	3/6	4/11	5/12	6/10
	G_{652}	37	127	588.731	6	666	3330.0	266.4	3/7	4/12	5/12	6/6
	ECG Graphs											
G_{634} 	G_{601}	1	8	0.000	1	0	0.0	0.0	6/1			
	G_{611}	3	13	3.067	1	3	10.0	1.8	3/1	5/1	6/1	
G_{635} 	G_{612}	2	10	0.718	1	1	3.0	0.7	4/1	6/1		
	G_{621}	6	18	16.093	1	15	48.0	9.5	3/2	4/2	5/1	6/1
G_{636} 	G_{622}	4	15	5.643	1	6	19.5	3.8	3/1	4/1	5/1	6/1
	G_{623}	4	15	5.005	1	6	21.0	3.5	3/1	4/1	5/1	6/1
G_{641} 	G_{624}	5	19	8.728	1	10	41.5	4.9	3/1	4/1	5/2	6/1
	G_{625}	3	12	2.275	1	3	12.0	1.5	4/2	6/1		
G_{642} 	G_{631}	4	15	5.643	1	6	30.0	2.4	3/1	4/1	5/1	6/1
	G_{632}	8	26	27.571	1	28	140.0	11.2	3/2	4/2	5/3	6/1
G_{643} 	G_{633}	8	23	28.234	1	29	100.0	17.0	3/3	4/2	5/2	6/1
	G_{634}	6	18	15.674	1	15	48.0	9.5	3/2	4/2	5/1	6/1
G_{644} 	G_{635}	6	18	16.093	1	15	50.0	9.3	3/2	4/2	5/1	6/1
	G_{636}	2	10	0.718	1	1	4.0	0.5	4/1	6/1		
G_{645} 	G_{641}	6	18	16.220	1	15	65.0	7.0	3/2	4/2	5/1	6/1
	G_{642}	10	30	44.111	1	45	225.0	18.0	3/2	4/4	5/3	6/1
G_{646} 	G_{643}	12	32	67.195	1	67	318.0	28.0	3/4	4/4	5/3	6/1
	G_{644}	9	27	31.107	1	36	144.0	18.0	3/1	4/5	5/2	6/1
G_{651} 	G_{645}	10	30	45.317	1	45	225.0	18.0	3/2	4/4	5/3	6/1
	G_{646}	5	17	7.967	1	10	34.0	6.1	3/1	4/2	5/1	6/1
G_{652} 	G_{651}	8	24	27.558	1	28	140.0	11.2	3/2	4/3	5/2	6/1
	G_{652}	8	24	26.352	1	28	140.0	11.2	3/2	4/3	5/2	6/1

nonlinear behavior of the descriptors of Table 2 (see Figure 6).

However, despite the appropriate behavior of the proposed descriptors to contribute a measure of the complexity/

Table 4. Study of Proposed Descriptors for Fused Systems

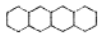
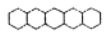
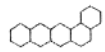
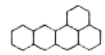
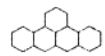
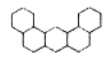
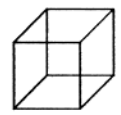
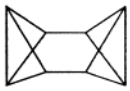
	Graphs	N	Σ	Γ	E	W	w	Δ	Cycles (size/number)					
	CG Graphs													
FG600 	FG600	1	8	0.000	1	0	0.0	0.0	6/1					
	FG601	3	144	4.048	1	3	15.0	1.2	6/2	10/1				
FG601 	FG602	6	2328	26.229	1	17	113.0	4.6	6/3	10/2	14/1			
	FG603	6	2328	26.229	1	16	112.0	4.3	6/3	10/2	14/1			
FG602 	FG604	7	920	25.135	1	21	189.0	4.7	6/3	10/3	12/1			
	FG605	10	37280	116.769	1	57	453.0	11.2	6/4	10/3	14/2	18/1		
FG603 	FG606	14	7840	116.619	3	92	1099.5	15.2	6/4	10/5	12/2	14/3		
	FG607	10	37280	116.769	1	54	447.5	10.7	6/4	10/3	14/2	18/1		
FG604 	FG608	12	13344	105.132	1	70	806.5	11.8	6/4	10/4	12/1	14/2	16/1	
	FG609	10	37280	116.769	1	51	442.0	10.3	6/4	10/3	14/2	18/1		
FG605 	FG610	10	37280	116.769	1	52	442.0	10.5	6/4	10/3	14/2	18/1		
	FG611	11	39328	149.728	1	58	624.0	10.2	6/4	10/3	14/3	18/1		
FG606 	FG612	15	596520	464.992	1	147	1329.0	22.7	6/5	10/4	14/3	18/2	22/1	
	FG613	15	596520	464.992	1	141	1321.0	22.0	6/5	10/4	14/3	18/2	22/1	
FG607 	FG614	18	212136	381.140	1	179	2338.0	24.4	6/5	10/5	12/1	14/3	16/1	18/2	20/1
	FG615	19	189608	397.987	1	189	2335.0	27.7	6/5	10/5	12/1	14/4	16/2	18/1	20/1
FG608 	FG616	15	596520	464.992	1	135	1313.0	21.3	6/5	10/4	14/3	18/2	22/1	
	FG617	15	596520	464.992	1	136	1314.0	21.4	6/5	10/4	14/3	18/2	22/1	
FG609 	FG618	22	118568	420.805	3	244	3613.0	31.5	6/5	10/6	12/2	14/5	16/1	18/3	
	FG619	18	212136	381.140	1	172	1913.5	28.5	6/5	10/5	12/1	14/3	16/1	18/2	20/1
FG610 	FG620	18	212136	381.140	1	174	2327.5	23.9	6/5	10/5	12/1	14/3	16/1	18/2	20/1
	FG621	19	189608	397.987	1	187	2740.0	23.7	6/5	10/5	12/1	14/4	16/2	18/1	20/1
FG611 	FG622	17	631336	633.879	1	155	1743.0	24.1	6/5	10/4	14/4	18/3	22/1	
	FG623	23	96040	431.583	2	261	3793.0	34.9	6/5	10/6	12/2	14/6	16/2	18/2	
FG612 	FG624	22	75560	357.960	1	242	3529.5	31.6	6/5	10/6	12/2	14/4	16/4	18/1	
	FG625	26	74152	457.797	1	331	4903.0	43.9	6/5	10/7	12/3	14/7	16/3	18/1	
	ECG Graphs													
FG613 	FG600	1	8	0.000	1	0	0.0	0.0	6/1					
	FG601	2	136	0.610	1	1	5.0	0.4	6/1	10/1				
FG614 	FG602	4	2192	12.545	1	6	45.0	1.6	6/2	10/1	14/1			
	FG603	4	2192	6.084	1	6	43.5	1.7	6/2	10/1	14/1			
FG615 	FG604	3	648	2.022	1	3	27.0	0.7	6/1	10/1	12/1			
	FG605	6	35088	15.564	1	19	148.0	3.8	6/2	10/2	14/1	18/1		
FG616 	FG606	7	4880	18.479	2	21	255.0	3.5	6/2	10/2	12/1	14/2		
	FG607	6	35088	42.214	1	15	145.0	3.1	6/2	10/2	14/1	18/1		
FG617 	FG608	9	11160	57.188	1	38	433.0	6.5	6/3	10/3	12/1	14/1	16/1	
	FG609	6	35088	22.044	1	16	145.0	3.3	6/2	10/2	14/1	18/1		
FG618 	FG610	6	35088	42.214	1	15	145.0	3.1	6/2	10/2	14/1	18/1		
	FG611	5	34960	15.621	1	10	114.0	1.8	6/2	10/1	14/1	18/1		
FG619 	FG612	9	561432	126.241	1	42	447.0	6.7	6/3	10/2	14/2	18/1	22/1	
	FG613	9	561432	217.634	1	39	442.0	6.4	6/3	10/2	14/2	18/1	22/1	
FG620 	FG614	14	177184	233.852	1	107	1356.0	14.9	6/4	10/4	12/1	14/2	16/1	18/1	20/1
	FG615	12	177048	101.016	1	70	902.0	10.3	6/3	10/3	12/1	14/2	16/1	18/1	20/1
FG621 	FG616	9	561432	55.528	1	45	446.0	7.0	6/3	10/2	14/2	18/1	22/1	
	FG617	9	561432	55.528	1	45	446.5	7.0	6/3	10/2	14/2	18/1	22/1	
FG622 	FG618	16	81440	228.409	2	129	1870.0	16.7	6/4	10/4	12/2	14/3	16/1	18/2	
	FG619	14	177184	233.852	1	102	1152.0	16.7	6/4	10/4	12/1	14/2	16/1	18/1	20/1
FG623 	FG620	14	177184	233.852	1	104	1348.5	14.6	6/4	10/4	12/1	14/2	16/1	18/1	20/1
	FG621	12	177048	101.016	1	70	1049.0	8.9	6/3	10/3	12/1	14/2	16/1	18/1	20/1
FG624 	FG622	13	596384	283.107	1	89	1000.5	13.8	6/4	10/3	14/3	18/2	22/1	
	FG623	17	85024	211.813	2	141	2032.0	18.9	6/4	10/4	12/1	14/5	16/1	18/2	
FG625 	FG624	8	43792	26.177	1	28	428.5	3.7	6/2	10/2	12/1	14/1	16/1	18/1	
	FG625	16	58904	155.376	1	120	1814.0	15.9	6/3	10/4	12/2	14/4	16/2	18/1	

Table 5. Study of the Symmetric Characteristics for Cubane, Octabisvalene, and Cuneane[a]



Cubane



Octabisvalene

Cuneane

Graphs	N	Σ	Γ	E	W	w	Δ	Cycles (size / number)					
CG Graphs													
Cubane	28	332	315.941	6	381	2268.0	126.999	4/6	6/6	8/6			
Octabisvalene	22	296	211.025	4	242	1450.0	81.450	3/4	4/2	6/4	7/8	8/4	
Cuneane	26	302	275.915	3	330	2275.0	94.286	3/2	4/2	5/4	6/7	7/8	8/3
ECG Graphs													
Cubane	4	50	4.737	1	6.0	36.0	2.000	4/1	6/2	8/1			
Octabisvalene	5	59	9.488	1	10.0	70.0	2.857	3/1	4/1	6/1	7/1	8/1	
Cuneane	11	115	48.530	1	56.0	385.0	16.000	3/1	4/1	5/2	6/3	7/3	8/1
SCG Graphs													
Cubane	5	10	11.314	5	12.0	60.0	4.000	4/5					
Octabisvalene	5	6	3.507	1	18.0	66.0	5.736	3/4	4/1				
Cuneane	5	10	10.243	1	13.0	70.0	3.714	3/2	4/2	5/1			
ICG Graphs													
Cubane	10	44	48.095	4	48.0	270.0	16.000	4/6	6/4				
Octabisvalene	9	38	42.588	4	44.0	226.0	14.679	3/4	4/1	6/4			
Cuneane	8	30	34.597	2	31.0	196.0	8.857	3/2	4/2	5/2	6/2		

Other Descriptors	twc	κ	W	p2	AV	Complexity (Cx)	wcx
Cubane	13116	0.4375	48	24	81.0	10.125	1639.5
Octabisvalene	13116	0.4375	50	54	79.5	19.875	3279.0
Cuneane	13116	0.4375	46	24	82.5	31.125	4918.5

[a] twc, total walk count; κ: Kappa index; *W*, Wiener index; p2, number of paths of length 2; AV and Cx, complexity measures proposed by Randic; wcx, walk complexity.

cyclicity, they are not able to completely discriminate the cyclical structures.

Thus, we can appreciate that Σ_{CG} is not able to discriminate graphs with the same number of cycles with equal size, but this is distributed in the molecular structure in a different way (interesting for some type of applications as clustering of chemical databases). It is the case of G_{622} and G_{623}, or G_{642}, and G_{645} graphs shown in Table 3.

However, since Γ_{CG} considers the number of common nodes among cycles, it is suitable for the discrimination of graphs with equal numbers of cycles of the same size but differently related as shown in Table 3 for the G_{6XX} graphs. For instance, while Σ_{CG} gives equal values for G_{622} and G_{623} graphs, Γ_{CG} differentiates the complexity of these graphs due to the different relationships among the cycles. Randic[12] also provides different values of *D*/DD for both graphs, although in an inverse order to Γ_{CG}.

As is also shown in Table 3, Σ_{ECG} provides equal values for graphs in which Σ_{CG} has equal (e.g., G_{622} and G_{623}) or different values (e.g., G_{634}, G_{635}, G_{641}, or G_{612}, G_{636}), because these graphs present the same set of equivalent cycles, although different types of relationships have different values of Γ_{ECG}.

The higher graph discrimination is obtained using the w_{CG} descriptor based on maximum distance, allowing order of the graphs according to cyclicity/complexity, because although all graphs have the same number of nodes, the number and size of cycles is different, the maximum distances among them being different.

4.2. Characteristics of Fused Systems. Table 4 shows the values of the proposed descriptors for some fused systems. As can be observed, Σ_{CG} increases with the number and size of cycles in the graph, giving a measurement of the graph cyclicity/complexity. It is clear that Σ_{CG} provides equal value for structures having the same number of rings, with the same number of relationships among them, but with different space orientations (for instance: FG_{602} and FG_{603}, or FG_{605}, FG_{607}, FG_{609}, FG_{6010}, or FG_{616} and FG_{617} among others). This fact allows us to think of the use of this complexity measure in clustering processes of chemical databases in which similar structures, independently of the relative rings orientation, are assigned to the same cluster that can be suitably used in screening processes based on similarity characteristics.

However, for structures with the same elementary number of cycles, for example FG_{602} or FG_{603} and FG_{604} (with three cycles of size equal to 6), Σ_{CG} diminishes when the connectivity among the elementary cycles increases. This behavior occurs when increasing the connectivity among cycles, the number of nodes in the structure diminishes, and, although the number of cycles increases, these cycles have a smaller size (see Table 4).

This fact can been observed in Table 4. For instance, the sequences $\Sigma_{CG}(FG_{605}) > \Sigma_{CG}(FG_{608}) > \Sigma_{CG}(FG_{606})$ or $\Sigma_{CG}(FG_{623}) > \Sigma_{CG}(FG_{624}) > \Sigma_{CG}(FG_{625})$. This index behavior is shown when structures are compared as in FG_{614} and FG_{615}, where the position of the cycle centered in the FG_{615} structure gives place to a smaller value of Σ_{CG} in FG_{614}. This behavior of Σ_{CG} shows its capacity to be used as a measure of the complexity/cyclicity for cyclical structures.

The discrimination capacity of Γ_{CG} is similar to Σ_{CG}, although Γ_{CG} alters the ordination for some structures considerably. For example $\Sigma_{CG}(FG_{608}) > \Sigma_{CG}(FG_{606})$, while $\Gamma_{CG}(FG_{608}) < \Gamma_{CG}(FG_{606})$, or $\Sigma_{CG}(FG_{624}) > \Sigma_{CG}(FG_{625})$, but $\Gamma_{CG}(FG_{625}) > \Gamma_{CG}(FG_{624})$, since Γ_{CG} considers the cycles' relationships.

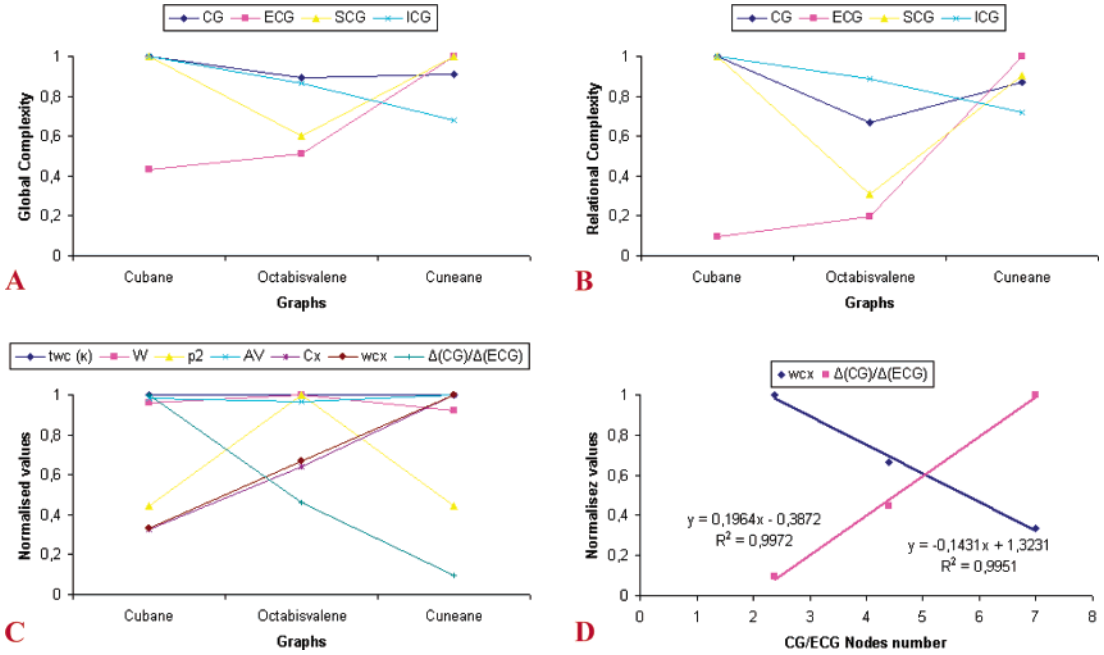


Figure 7. Behavior of proposed and bibliographic topological descriptor for cubane, octabisvalene, and cuneane (all the values have been normalized in the interval [0, 1]).

Table 4 shows a close behavior for Σ_{ECG} with Σ_{CG}. However, Γ_{ECG} introduces a higher discrimination. For example, see the following:

- Γ_{ECG} discriminates having the same values for Σ_{CG}, Γ_{CG}, and Σ_{ECG} (e.g., the structures FG_{605} and FG_{607}).
- Structures with an equal number of elemental cycles but different orientation are partially discriminated and ordered (e.g., FG_{605}, FG_{609}, FG_{607}, or FG_{616}, FG_{612}, FG_{613}).

Although Γ_{ECG} is also able to discriminate structures with a different number of elementary cycles maintaining different relationships, it is unable to fully discriminate isomeric structures with different cycles' orientation. For example: (FG_{607} and FG_{610} or FG_{614}, FG_{619}, FG_{620}).

Although not presented in this paper (for reasons of space), results corresponding to the Σ and Γ descriptors based on the SSSR and LICS cycle graphs (SCG, ICG) present a degenerated behavior.

Although Σ and Γ have the advantage of their efficient calculation, since once the present cycles in the *G* graph are known, the calculation is based on the inspection of the *M* matrix, and these descriptors present an appropriate behavior for simple cyclical structures, as has been proven previously, they present inconveniences that are not able to fully discriminate more complex cyclical structures as the fused systems.

But, is this behavior really a serious inconvenience? Are the representations CG and ECG based on the cycles inadequate? Evidently not, for example, see the following:

- The descriptors based on the CG graph discriminate cyclical structures with different complexity and relationships among the elementary cycles (see Table 4).
- Γ_{ECG} discriminates some isomeric dispositions for cyclical structures of the same complexity and relationships among the elementary cycles (see Table 4).

This behavior can be used suitably in clustering and screening processes in chemical databases, in which it is convenient to aggregate and/or to recover structures with these characteristics.

However, on other occasions and/or for other applications it is necessary to obtain a complete discrimination among isomeric structures, for example among the following graphs: FG_{607} and FG_{610} or FG_{616} and FG_{617} (see Table 4).

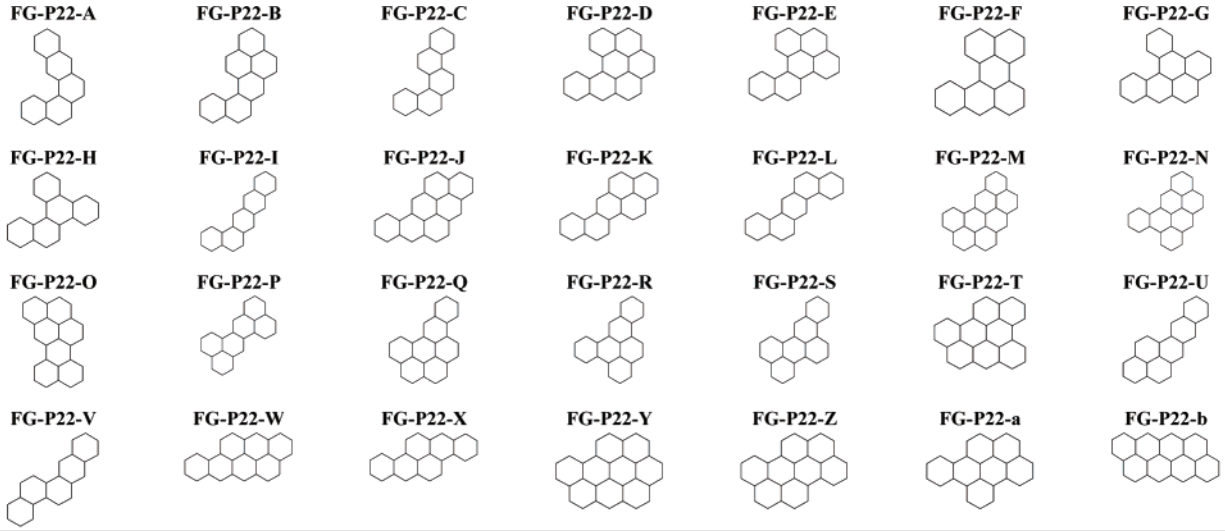
The complete discrimination of the structures is obtained by means of the use of the descriptors based on distances over the CG graph (see Table 4), because the measurement of the distances takes into account the different cycles' orientation in the structure.

4.3. Shape and Symmetry. The shape and symmetry of the molecular structures are other characteristics that are difficult to define and, therefore, to measure and to quantify. An abstract measure of the molecular shape can be considered in function of its "resemblance" to patterns of shapes as stars, rings, spheres, etc., that, occasionally, permit us to relate this characteristic indirectly with the cyclicity and, therefore, with the complexity of the structure, and directly with the molecular symmetry.[16,33]

The results previously presented in Table 4 for fused systems allow us to consider that the proposed descriptors can be used as a measure of the molecular shape, since these descriptors present different values for structures with equal numbers of elementary cycles but with different relationships. For example when Γ_{ECG} diminishes, it increases the grouping of cycles, and therefore, the molecular shape becomes more spherical (less lineal). So, Γ_{ECG} is dependent on the complexity as well as the symmetry of the molecule, since this descriptor is based on the equivalent cycles.

Besides, the use of descriptors based on maximum/minimum distances contributes to a finer measure of the symmetry, due to these descriptors discriminating among isomeric structures in which the elementary cycles have different spatial disposition.

This capacity of the proposed descriptors to give a measurement of molecular shape and symmetry can be clearly appreciated in Table 5 for a classical example used in the bibliography,[13] due to these structures having close values for topological indexes usually used for obtaining



Graphs

H^R	B	b	E	H	I	j	K	G	M	O	Y	D	T	a	S	R	P	F	N	V	Q	U	Z	A	W	X	L	C
AH^R	B	E	H	I	J	K	G	M	O	Y	D	T	a	S	R	P	b	F	N	V	Q	U	Z	A	W	X	L	C

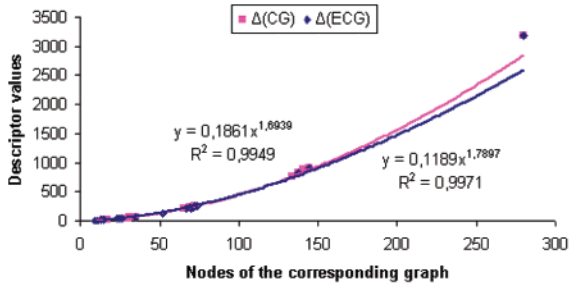
CG Graphs

Nodes number	A	C	I	L	V	H	P	B	K	U	E	F	S	X	G	R	O	J	W	D	Q	N	a	b	M	Z	T	Y
Σ	A	C	I	L	V	H	F	S	P	G	R	E	X	B	K	U	O	N	J	W	a	D	Q	b	T	M	Z	Y
Γ	A	C	I	L	V	H	P	F	S	B	K	U	E	X	G	R	O	J	W	N	D	Q	a	b	M	Z	T	Y
W	C	A	V	L	I	H	P	B	K	U	F	E	S	X	G	R	O	J	W	D	Q	N	a	b	M	Z	T	Y
w	L	A	C	I	V	H	P	B	K	U	S	E	F	X	R	G	O	W	J	Q	D	a	N	b	M	Z	T	Y
Δ	C	A	V	L	I	H	P	K	U	X	F	B	E	S	G	R	O	J	W	D	Q	N	a	b	M	Z	T	Y

ECG Graphs

Nodes number	A	C	I	L	V	P	H	X	B	K	U	E	F	S	G	R	O	J	W	D	Q	b	N	a	Z	M	T	Y
Σ	A	C	I	L	V	H	P	F	S	X	B	K	U	E	G	R	O	N	J	W	a	D	Q	b	Z	T	M	Y
Γ	P	L	V	X	A	C	I	H	F	S	E	B	K	U	G	R	O	b	Z	J	W	N	D	Q	a	M	T	Y
W	V	I	A	C	L	P	H	X	B	K	U	E	F	S	G	R	O	J	W	D	Q	b	N	a	Z	M	T	Y
w	A	C	I	L	V	H	P	X	B	K	U	E	S	F	R	G	O	W	J	Q	D	a	N	b	Z	M	T	Y
Δ	V	I	P	A	C	L	X	H	K	U	B	F	E	S	G	R	O	b	J	W	D	Q	N	Z	a	M	T	Y

Δ(CG)/Δ(ECG)	D	G	J	M	N	Q	R	T	W	Y	a	O	F	S	E	U	H	B	K	C	L	A	V	I	Z	b	X	P



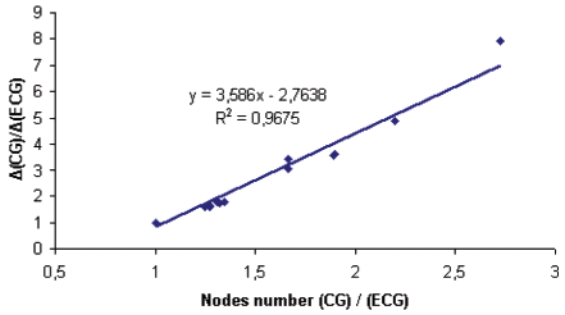


Figure 8. Set of cyclical structures with a perimeter equal to 22.

measures of shape and symmetry (Kappa and Kier indexes). The graphical representation of data represented in Table 5 are shown in Figure 7 where for a better comparison of different descriptors all the values have been normalized in the interval [0, 1].

As is observed in the Table 5 and Figure 7, descriptors based on CG graph (Figure 7A,B) have a close behavior to other descriptors proposed in the bibliography[12,13] (Figure 7C), contributing to measure of the cyclicity/complexity of the molecular structures, but they are not suitable for measuring the symmetry, while other descriptors are not suitable to differentiate these structures (e.g., twc, W, κ).

The reasonable behavior of the descriptors based on the ICG graphs (Figure 7A) is due to the number and composition of the lineally independent cycles present in these structures, very different from the SSSR set that presents an anomalous behavior, as is appreciated in Table 5.

However, the descriptors based on the ECG graph (Figure 7A,B) present a logical behavior to measure the symmetry, at same wcx and Complexity (Cx) descriptors (see Table 5). The values of Σ_{ECG}, Γ_{ECG}, W_{ECG}, w_{ECG}, and Δ_{ECG} increase in the sequence cubane, octabisvalene, cuneane, and with a higher power discriminatory to those descriptors proposed in other papers, contributing to a suitable measurement of

Table 6. Topological Invariants Extracted from the Cyclical Systems of Figure 5[a]

	GG Model							ECG Model								
graphs	*N*	Σ	Γ	*E*	*W*	*w*	Δ	*N*	Σ	Γ	*E*	*W*	*w*	Δ	H^R	AH^R
FG-P22-A	15	596520	465.0	1	132	1319.0	20.532	9	561432	188.8	1	41	440.5	6.662	285	17.820
FG-P22−B	31	1890224	1445.0	3	515	8373.5	60.242	23	1296040	845.4	2	287	4472.5	34.144	229	14.310
FG-P22−C	15	596520	465.0	1	129	1320.5	20.158	9	561432	188.8	1	41	440.5	6.662	301	18.810
FG-P22-D	69	13212984	6074.2	1	2413	48838.0	232.184	69	13212984	6074.2	1	2413	48838.0	232.184	253	15.810
FG-P22-E	33	1507248	1466.4	2	566	9567.0	64.131	25	1332392	799.1	2	325	5366.0	37.387	242	15.130
FG-P22−F	33	1071024	1218.8	1	563	9972.0	59.874	26	894760	771.8	1	348	6104.5	37.288	265	16.560
FG-P22-G	35	1419184	1584.1	2	624	10507.0	70.858	35	1419184	1584.1	2	624	10507.0	70.858	246	15.380
FG-P22−H	17	631336	633.9	1	150	1744.5	23.348	13	596384	283.1	1	85	1001.0	13.244	242	15.130
FG-P22−I	15	596520	465.0	1	141	1321.0	21.988	9	561432	217.6	1	39	442.0	6.368	242	15.130
FG-P22-J	67	9000248	5301.0	2	2298	46958.5	216.674	67	9000248	5301.0	2	2298	46958.5	216.674	242	15.130
FG-P22−K	31	1890224	1445.0	3	520	8609.5	58.318	23	1296040	845.4	2	292	4637.5	33.021	242	15.130
FG-P22-L	15	596520	465.0	1	136	1314.0	21.389	9	561432	89.6	1	43	442.5	6.950	298	18.630
FG-P22-M	137	51999424	19485.0	3	9469	216514.5	814.740	137	51999424	19485.0	3	9469	216514.5	814.740	246	15.380
FG-P22−N	71	8289592	5722.7	2	2549	52695.0	240.599	71	8289592	5722.7	2	2549	52695.0	240.599	274	17.130
FG-P22−O	65	6440248	4408.7	1	2169	44501.0	202.748	52	5379248	2852.8	1	1385	28401.5	129.325	246	15.380
FG-P22−P	30	1206192	1110.3	1	495	8241.5	52.676	11	699280	65.9	1	61	1028.0	6.644	262	16.380
FG-P22-Q	69	13212984	6074.2	1	2421	47926.0	237.107	69	13212984	6074.2	1	2421	47926.0	237.107	278	17.380
FG-P22−R	35	1419184	1584.1	2	627	10272.5	72.715	35	1419184	1584.1	2	627	10272.5	72.715	261	16.310
FG-P22−S	33	1071024	1218.8	1	569	9000.0	67.268	26	894760	771.8	1	352	5534.5	41.628	254	15.880
FG-P22-T	145	50504384	22011.8	3	10551	241494.0	913.596	145	50504384	22011.8	3	10551	241494.0	913.596	253	15.810
FG-P22−U	31	1890224	1445.0	3	529	8675.0	59.049	23	1296040	845.4	2	297	4664.0	33.502	278	17.380
FG-P22−V	15	596520	465.0	1	135	1329.0	20.759	9	561432	157.6	1	37	448.0	6.052	274	17.130
FG-P22−W	67	9000248	5301.0	2	2302	46879.5	217.411	67	9000248	5301.0	2	2302	46879.5	217.411	289	18.060
FG-P22-X	33	1802160	1534.5	3	573	10250.0	59.490	15	1129368	183.2	2	115	2009.5	12.135	290	18.130
FG-P22−Y	280	378810440	78912.0	1	39309	963525.5	3191.622	280	378810440	78912.0	1	39309	963525.5	3191.622	246	15.380
FG-P22-Z	140	55851712	21150.0	4	9865	218514.0	879.219	74	28762528	4877.5	2	2731	60653.0	243.424	278	17.380
FG-P22-a	72	12862776	6545.0	1	2616	52279.0	255.872	72	12862776	6545.0	1	2616	52279.0	255.872	253	15.810
FG-P22-b	133	44182208	17723.6	1	8963	206386.0	762.712	70	26843552	4794.9	1	2461	56819.5	209.291	240	16.380

[a] H^R, sum of Hamming distance for path 1−16; AH^R, average of Hamming distance for path 1−16. Extracted from ref 17.

the molecular symmetry (see Table 5).

There exists lineal behavior of Δ_{CG}/Δ_{ECG} with the quotient between nodes number of CG and ECG graphs (total number of cycles and equivalent cycles of *G* graph). This lineal relationship is only observed for the wcx descriptor, although in reverse order (see Figure 7D).

This fact is due to ECG representation including the symmetry of the structure, and therefore, the ratio between CG and ECG descriptors gives a relative measure of the shape/symmetry of the graphs. This behavior is also observed for the fused systems described in the previous section.

Since the model based on the classes of equivalent cycles (ECG) allows us to obtain measures related to the shape and symmetry of cyclical structures and the use of Δ_{CG}/Δ_{ECG} descriptor allows us to order cyclical structures based on this measure, we have carried out the validation of our descriptors with a set of benzenoids used by Randic[17] to propose a relative chirality measure of the molecules when they are embedded in a plane.

Although it is not precise to speak about chirality of molecules in a 2D space, our objective consists of characterizing and ordering isomeric structures which share equal sets of elemental cycles but are distributed in a different way, in order to enlarge our analyses of results described in previous sections.

In Figure 8, 28 structures with equal perimeter value are shown. Calculating the sum of Hamming distance and the average of Hamming distance for paths from 1 to 16, Randic[17] obtains some indexes, by means of which he proposes an ordination of these molecules. In Table 6 the descriptors proposed for these structures using the CG and ECG models and the values of the sum of the Hamming distance (H^R) for paths from 1 to 16 and the average value of the Hamming distance (AH^R) for paths equal to 16 obtained by Randic[17] are shown. And, Figure 8 shows how these structures are ordered according to the considered descriptor, where the order should be interpreted reading the rows (corresponding to each descriptor) from left to right (upward order). The structures that present equal value of the descriptor have been colored, not being colored when the value is different.

We can observe (see Table 6) that that H^R and AH^R are highly correlated through the function $y = 15.999x - 0.1716$, $r^2 = 0.9985$. So, Randic proposes a partial ordination of the structures using the values sequence of AH^R for each path (1−16).

As Table 6 shows, some graphs have equal value for the proposed indexes for CG and ECG models. That is due to these graphs not having equivalent cycles, and therefore, both representations are identical (e.g., D, G, J, M, N, Q, R, T, W, Y).

The complexity of cyclical structures depends on the number, size, and relationships among the cycles. Although the 28 structures in the test have a perimeter of 22, these structures have a very different number of cycles (see Table 6). The global complexity Σ_{CG} allows ordering these structures based on their cyclicity, and the relational complexity Γ_{CG} as a function of its complexity. But, as we commented previously, these descriptors are not fully useful to differentiate isomeric with the same number of elemental cycles but with different space disposition (e.g., A,C, I, L, V).

However, descriptors based on distances W_{CG}, w_{CG}, and Δ_{CG} allow us to discriminate all graphs although in a slightly different order (see Figure 8). Since these structures have the same perimeter (22), and cycles with the same size (6),

a high correlation among these descriptors based on distances is appreciated (e.g., W_{CG} vs w_{CG}, $y = 24.407x - 6601.5$, $r^2 = 0.9991$), behavior also observed for the graphs of Table 5. All these descriptors have an exponential behavior with the nodes number of the corresponding graph (cycles number) as it is shown in Figure 8 for Δ_{CG} and Δ_{ECG}.

However, as we mentioned previously it is possible to characterize and to order these structures correctly making use of the ratio among descriptors based on the distances (Δ_{CG}/Δ_{ECG}). As observed in Figure 8 there exists a lineal dependency of Δ_{CG}/Δ_{ECG} (also W_{CG}/W_{ECG} and w_{CG}/w_{ECG}) with the ratio between the nodes number (cycles number) of CG and ECG graphs. This fact has been commented on previously for the other graph sequences studied in this paper.

Thanks to this behavior of the proposed descriptors, Figure 8 shows the order obtained for the cyclical system using the ratio of Δ descriptors, where isomeric structures are ordered appropriately (e.g., C, L, A, V, I) in the same order as that proposed by Randic[17] and in accordance with the orientation of the cycles and the number of Kekule structures.

CONCLUSIONS AND REMARKS

In this paper we have proposed a new representation model of the molecular topology of the chemical compounds based on the use of all the present cycles in the molecule. The construction of graphs based on all the cycles and graphs based on the equivalent cycles has been proven appropriate for the extraction of topological invariants oriented to provide measures of topological parameters such as complexity, cyclicity, shape, and symmetry, since these measures are dependent on the presence, type, and distribution of the cycles in the molecular structure.

The topological invariants proposed are very easy to calculate, having obtained all cycles of the molecular graph, by means of simple inspections of the matrix that represents the cycle graph (or equivalent cycles).

The behavior of the descriptors proposed for the measure of the topological characteristics studied has proven that it corroborates the results of some invariants described in the bibliography and improves others.

The proposed representation models and the developed invariants present an appropriate behavior with the different types of cyclical systems, allowing us to consider their application to other problems in computational chemistry as clustering of large databases.[21]

Although in this paper only cyclical systems have been considered, in future papers we will extend the models proposed to molecular structures with presence of noncyclical chains.[21]

REFERENCES AND NOTES

(1) Randic, M. Topological Indices. In *Encyclopedia of Computational Chemistry*; Schleyer, P., Allinger, N. L., Clark, T., Gasteiger, J., Kollman, P. A., Schaefer, H. F., Schreiner, P. R., Eds.; John Wiley & Sons: Chichester, 1998; pp 3018−3032.

(2) Balaban, A. T. Historical Developments of Topological Indices. In *Topological Indices and Related Descriptors in QSAR and QSPR*; Devillers, J., Balaban, A. T., Eds.; Gordon and Breach: Amsterdam, The Netherlands, 1999; pp 403−453.

(3) Todeschini, R.; Consonni, V. *The Handbook of Molecular Descriptors* Methods and Principles in Medicinal Chemistry Series; Mannhold, R., Kubinyi, H., Timmerman, H., Eds.; Wiley-VCH: New York, 2000; Vol. 11, p 680.

(4) Rucker, G.; Rucker, C. On Topological Indices. Boiling Points. and Cycloalkanes. *J. Chem. Inf. Comput. Sci.* **1999**, *39* (2), 788−802.

(5) Randic, M.; Basak, S. Optimal Molecular Descriptors Based on Weighted Path Numbers. *J. Chem. Inf. Comput. Sci.* **2001**, *41* (3), 261−266.

(6) Randic, M.; Pompe, M. The Variable Molecular Descriptors Based on Distance Related Matrices. *J. Chem. Inf. Comput. Sci.* **2001**, *41* (3), 575−581.

(7) Rucker, G.; Rucker, C. Symmetry-Aided Computation of the Detour Matrix and Detour Index. *J. Chem. Inf. Comput. Sci.* **1998**, *38* (4), 710−714.

(8) Randic, M.; Basak, S. Multiple Regression Analysis with Optimal Molecular Descriptors. *SAR QSAR Environ. Res.* **2000**, *11*, 1−23.

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

(10) Rucker; G.; Rucker, C. Substructure, Subgraph and Walk Counts as Measures of the Complexity of Graphs and Molecules. *J. Chem. Inf. Comput. Sci.* **2001**, *41* (6), 1457−1462.

(11) Pisanski, T.; Plavsic, D.; Randic. M. On Numerical Characterization of Cyclicity. *J. Chem. Inf. Comput. Sci.* **2000**, *40* (3), 520−523.

(12) Randic, M. On Characterization of Cyclic Structures. *J. Chem. Inf. Comput. Sci.* **1997**, *37* (6), 1063−1071.

(13) Rucker, G.; Rucker, C. Walk Counts, Labyrinthicity, and Complexity of Acyclic and Cyclic Graphs. *J. Chem. Inf. Comput. Sci.* **2000**, *40* (1), 99−106.

(14) Rucker, G.; Rucker, C. Substructure, Subgraph, and Walk Counts as Measures of the Complexity of Graphs and Molecules. *J. Chem. Inf. Comput. Sci.* **2001**, *41* (6), 1457−1462.

(15) 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.

(16) Randic, M. Novel Shape Descriptors for Molecular Graphs. *J. Chem. Inf. Comput. Sci.* **2001**, *41* (3), 607−613.

(17) Randic, M.; Razinger, M. Molecular Shapes and Chirality. *J Chem. Inf. Comput. Sci.* **1996**, *36*, 429−441.

(18) Gillet, V. J.; Downs, G. M.; Ling, A.; Lynch, M. F.; Venkataran, P.; Wood, J. V.; Dehielfsen, W. Computer Storage and retrieval of Generic Chemical structures in Patents. 8. Reduced Chemical Graph and Their Applications in Generic Chemical Structure Retrieval. *J. Chem. Inf. Comput. Sci.* **1987**, *27*, 126−137.

(19) Dury, L.; Latour, T.; Leherte, L.; Barberis, F.; Vercauteren, D. P. A new graph Descriptor for Molecules Containing Cycles. Application as Screening Criterion for Searching Molecular Structures within Large Databases of Organic Compounds. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1437−1445.

(20) Xiu, Y.; Johnson, M. Using Molecular Equivalence Numbers to Visually Explore Structural Features that Distinguish Chemical Libraries. *J. Chem. Inf. Comput. Sci.* **2002**, *42* (4), 912−926.

(21) Luque Ruiz, I.; Cerruela García, G.; Gómez-Nieto, M. A. Representation of the Molecular Topology of Cyclical Structures by means of Cycle Graphs: 2. Application to Clustering of Chemical Databases. *J. Chem. Inf. Comput. Sci*., submitted for publication.

(22) Figueras, J. Ring Perception Using Breadth-First Search. *J. Chem Inf. Comput. Sci.* **1996**, *36*, 986−991.

(23) Fujita, S. A. New Algorithm for Selection of Synthetically Important Rings. *J. Chem Inf. Comput. Sci.* **1988**, *28*, 22−26.

(24) Cerruela García, G.; Luque Ruiz, I.; Gómez-Nieto, M. A. Cyclical Conjunction: An Efficient Operator for Extraction all Cycles in Graphs. *J. Chem. Inf. Comput. Sci.* **2002**, *42* (6), 1415−1424.

(25) Cerruela García, G.; Luque Ruiz, I.; Gómez-Nieto, M. A. Parallel Algorithms for Graph Cycle Extraction Using the Cyclical Conjunction Operator. *J. Chem. Inf. Comput. Sci.* **2002**, *42* (6), 1398−1406.

(26) Rucker, G.; Rucker, C. Computer Perception of Constitutional (Topological) Symmetry: TOPSYM, a Fast Algorithm for Partitioning Atoms and Pairwise Relations among Atoms into Equivalence Classes. *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 187−191.

(27) Xu, Y. J.; Johnson, M. A. Algorithm for Naming Molecular equivalence Class Represented by Labeled Pseudographs. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 181

(28) Alpin, J.; Mubarakzianow, R. The bases of Weighted Graphs. *Discrete Math.* **1997**, 1−11.

(29) Booch, G. *The Object-Oriented Analysis and Design with Applications*; Addison-Wesley: Reading, MA, 1994.

(30) Hong, H.; Xin, X. ESSESA, an Expert System for Structure Elucidation from Spectral Analysis. Part II. Novel Algorithm of Perception of the Linear Independent Smallest Set of Smallest Rings. *Anal. Chim. Acta* **1992**, *262* (1), 179−191.

(31) Figueras, J. Ring Perception Using Breadth-First Search. *J. Chem Inf. Comput. Sci.* **1996**, *36*, 986−991.

(32) McGregor, M. J.; Pallai, P. V. Clustering of large databases of compounds: using MDL Keys as structural descriptors. *J. Chem. Inf. Comput. Sci*. **1997**, *37*, 443−448.
(33) Kier, L. B. Indexes of Molecular Shape from Chemical Graphs. In *Computational Chemical Graph Theory*; Rouvray, D. H., Ed.; Nova Science Publishers: New York, 1990; pp 151−174.
(34) Randic, M.; Razinger, M. Molecular Shape and Chirality. *J. Chem. Inf. Comput. Sci*. **1996**, *36* (3), 429−441.
(35) Lipkus, A. H. Exploring Chemical Rings in a Simple Topological-Descriptor Space. *J. Chem. Inf. Comput. Sci*. **2001**, *41*, 430−438.
(36) Gross, J.; Yellen, J. *Graph Theory and Its Applications*. CRC Press Series: Boca Raton, FL, 1999.
(37) Bonchev, D. Novel Indices for Topological Complexity of Molecules. *SAR QSAR Environ. Res.* **1997**, *7*, 23−43.
(38) 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: Langhorne, PA, 1999; pp 361−401.

CI034256A