

## Determination of Topo-Geometrical Equivalence Classes of Atoms

T. Laidboeur,<sup>†</sup> D. Cabrol-Bass,<sup>\*,†</sup> and O. Ivanciuc<sup>‡</sup>

LARTIC University of Nice-Sophia Antipolis, Parc Valrose 06108 Nice Cedex, France, and  
Department of Organic Chemistry, Faculty of Chemical Technology, University "Politehnica" of Bucharest,  
77206 Bucharest, Romania

Received June 30, 1996<sup>⊗</sup>

In this paper we present a Prolog program which generates a molecular code that identifies the topological and geometrical equivalence classes of atoms in molecules. The initial partitioning of atoms is based on the list of successive extended connectivity and on the chromatism of the molecular graph. The definitive topological classes are obtained by matching all canonical labeled mapping trees exhaustively generated during the construction of a unique one-to-one topological code. Further, the geometric classes of atoms are identified on the basis of the lists of successive interatomic Euclidean distances generated following the molecular topology. The resulting topo-geometrical equivalence classes of atoms allow the construction of a molecular canonical code which maintains the distinction between the topology and the geometry. Thus stereoisomers and conformers share the same topological part of the canonical code but differ by their geometrical parts. The method developed in this work has a wide range of applications in chemical information systems: coding of chemical compounds in three-dimensional databases, searching in molecular databases, generation of substitution and labeled isomers, and the identification of the number of signals and equivalent atoms in NMR and ESR spectra. The algorithm, implemented in a Prolog program, can be used for the automatic recognition of topo-geometrical features of chemical compounds.

### INTRODUCTION

The perception of molecular symmetry is important for the computer-assisted manipulation of chemical structures, such as synthesis planning, structure elucidation, molecular design, database management, and structure-spectra correlation. Many of these problems have been investigated using molecular descriptors based on molecular graphs rather than on a three-dimensional (3-D) representation of the molecular structure. As implied by their name, the molecular graph (topological) descriptors are devoid of spatial characteristics, and their applications are restricted to problems which show little dependence on molecular conformation or to sets of structurally related compounds.<sup>1–5</sup> Based on the good results offered by the topological indices in developing Quantitative Structure–Property Relationships (QSPR) and Quantitative Structure–Activity Relationships (QSAR) models, a novel approach to molecular structure modeling was proposed by characterizing the 3-D chemical structure with numerical descriptors (topographic) derived from the geometry of a molecule.<sup>6–10</sup> The new topographic indices are structural invariants which can discriminate between geometric isomers or different conformations of a molecule.

Molecular graph algorithms form the basis of topological chemical structure manipulations: generation of a molecular code, structure and substructure search, molecular similarity determination, and determination of atoms and bonds equivalence. These fundamental operations are developed using algorithms operating on labeled graphs: graph coding, graph isomorphism, subgraph isomorphism, maximum common subgraph, and graph traversals.

In recent years, more emphasis was placed on 3-D Chemical Information Systems (CIS): molecular modeling,

experimental and computed 3-D structure databases, 3-D similarity, and clustering.<sup>11–13</sup> The development of 3-D CIS offers the opportunity to investigate the relationships between the geometrical arrangement of atoms in a molecule and its physical, chemical, and biological properties.

The manipulation of 3-D molecular structures is computationally much more intensive than the topological analysis; therefore, searching for efficient algorithms is a necessity. However, the two approaches are complementary, and the usefulness of topological information is not ruled out by the availability of geometrical information.

In the present paper we describe a new method for the determination of topo-geometrical equivalence of atoms in 3-D molecular structures; this method is an extension of an algorithm for the determination of topological equivalence classes of atoms. For this extension, the geometry of the molecule must be taken into account explicitly. This information is generally stored as tables of atomic coordinates (internal or Cartesian) or as tables of bond lengths, bond angles, and dihedral angles. In a recent paper<sup>14</sup> Balasubramanian has pointed out the difficulties which originate from using these classical methods of representation and has suggested the use of Euclidean distance matrices (EDMs) which are structurally invariant.<sup>15,16</sup> The EDM contains all information required to characterize the three-dimensional structure of a molecular system.<sup>17</sup>

In a previous paper<sup>18</sup> we presented an algorithm for the determination of topological equivalence classes of atoms and bonds which is based on the matching of canonical labeled mapping trees of the molecular graph. Atoms in the same topological equivalence class are further partitioned into geometrical equivalent classes using as criteria the list of successive interatomic Euclidean distance classes constructed following the molecular graph. The resulting topo-geometrical equivalence classes of atoms allow the con-

<sup>†</sup> LARTIC University of Nice-Sophia Antipolis.

<sup>‡</sup> University "Politehnica" of Bucharest.

<sup>⊗</sup> Abstract published in *Advance ACS Abstracts*, December 15, 1996.

struction of a molecular canonical code which maintains the distinction between the topology and the geometry. Thus stereoisomers and conformers share the same topological part of the canonical code but differs by their geometrical parts.

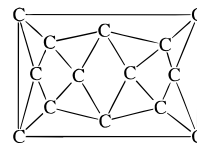
The method developed in this paper has a wide range of applications in chemical information systems: coding of chemical compounds in three-dimensional databases, searching in molecular databases, 3-D similarity and clustering determination, 3-D substructure search, and pharmacophore search. Also, the classes of equivalent atoms generated with the algorithm described in this work allow one to generate substitution and labeled isomers and to determine the number of signals and equivalent atoms in NMR and ESR spectra.

## METHOD

From a chemical point of view, partitioning of atoms in geometrical equivalent classes is meaningful only among atoms belonging to the same topological class. Although it may be possible to formally place atoms in 3D space which are geometrically symmetric but not topologically equivalent, these cases have no chemical significance. Therefore the search of geometrical equivalent atoms is undertaken only after the partitioning of atoms into topological equivalence classes is completed and concerns only atoms in the same topological class.

The method previously used to constitute the topological classes<sup>18</sup> is extended to the determination of geometric classes by using the list of the Euclidean distances between the current atom and all atoms from the same topological distance instead of the value of extended connectivity. For the sake of completeness this topological method will be briefly outlined here before its extension to geometry will be described.

**Determination of Topological Equivalence Classes of Atoms.** The method is made of three steps. The first one consists of an initial partitioning of atoms based on the List of Successive Extended Connectivity (LSEC) which is expanded iteratively. The use of a list instead of a sole extended connectivity value imposes a hierarchy over the classes built at each iteration and avoids oscillations between atoms classes. Therefore the maximum discrimination based on this local invariant is reached in fewer iterations. This partitioning is further refined by taking into consideration the following atom characteristics: atom type, number of attached hydrogen atoms, formal atom charge, and ordered list of bond types originating from this atom. The second step includes the construction of the Canonical Mapping Tree (CMT) of the molecular graph which allows the canonical numbering of atoms. All Canonical Labeled Mapping Trees (CLMT) corresponding to the different equivalent numberings of the structure are generated during this step. This exhaustive search ensures that the complete topological partitioning is obtained, as it was demonstrated in a previous paper.<sup>18</sup> Such a result cannot be guaranteed if one uses only local invariants however discriminant they may be, and although this was pointed out by Trucco<sup>19</sup> and Carhart,<sup>20</sup> some incomplete methods have still been proposed recently.<sup>21,22</sup> Both methods fail in some cases of highly symmetric peri-condensed cycles such as the one presented in Figure 1. However, there exist some algorithms which avoid the problems mentioned by Carhart, either by making a topological classification of the atoms followed by an



**Figure 1.** Example of graph which requires an exhaustive search to determine the topological equivalent classes of atoms.

**Table 1.** Distance Matrix for the Cyclohexane in Chair and Boat Conformation

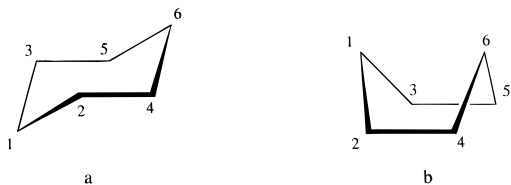
|                       | 1     | 2     | 3     | 4     | 5     | 6     |
|-----------------------|-------|-------|-------|-------|-------|-------|
| a. Chair Conformation |       |       |       |       |       |       |
| 1                     | 0.000 |       |       |       |       |       |
| 2                     | 1.536 | 0.000 |       |       |       |       |
| 3                     | 1.536 | 2.529 | 0.000 |       |       |       |
| 4                     | 2.529 | 1.536 | 2.959 | 0.000 |       |       |
| 5                     | 2.529 | 2.959 | 1.536 | 2.529 | 0.000 |       |
| 6                     | 2.959 | 2.529 | 2.529 | 1.536 | 1.536 | 0.000 |
| b. Boat Conformation  |       |       |       |       |       |       |
| 1                     | 0.000 |       |       |       |       |       |
| 2                     | 1.535 | 0.000 |       |       |       |       |
| 3                     | 1.535 | 2.526 | 0.000 |       |       |       |
| 4                     | 2.560 | 1.540 | 2.959 | 0.000 |       |       |
| 5                     | 2.560 | 2.959 | 1.540 | 2.526 | 0.000 |       |
| 6                     | 2.727 | 2.560 | 2.560 | 1.535 | 1.535 | 0.000 |

exhaustive search of the equivalence relationships, like the HOC (hierarchically ordered extended connectivities) procedures,<sup>23–29</sup> or by making use of highly elaborated atomic descriptors. From the second group we mention here some algorithms which, after extensive tests, gave correct results: the Gasteiger algorithm,<sup>30,31</sup> the partitioning of atoms on the basis of their topological state,<sup>32</sup> the computation of atomic weights<sup>33</sup> by using the layer matrix,<sup>34</sup> or the characterization of the atomic local environment by means of a subspanning tree.<sup>35</sup>

This step of the method is the most demanding one in computer time. However, it can be optimized by pruning the search as soon as the linear code being generated becomes greater than the shortest one currently stored. Tests have shown that, even for highly symmetrical molecules such as fullerenes, the computing time remains reasonable.<sup>18</sup> The final step consists of matching the equivalent CLMTs obtained as a result of the previous step to identify the atoms which occupy the same nodes and building the topological equivalent classes by gathering these atoms in the same class. If the number of such classes equals the number of atoms, all atoms are topologically different, and therefore searching for geometric equivalence is not necessary.

**Determination of Geometrical Classes of Atoms.** The method used afterward for partitioning atoms belonging to the same topological class into geometrical classes will be described using the chair and boat conformations of cyclohexane. To simplify, we will consider only the carbon atoms which are all topologically equivalent. First the matrix **D** of Euclidean geometric distances  $D_{ij}$  is constructed from the internal or Cartesian coordinates. This matrix has the properties that  $D_{ij} = D_{ji}$  for  $i \neq j$  and  $D_{ii} = 0$ . Table 1 (parts a and b) contain the **D** matrices for the chair and boat conformations of cyclohexane, respectively. Due to the symmetry of the **D** matrix, only the lower triangle will be presented.

The distances are rounded off and sorted by ascending order to assign to each pair of atoms Distance Class Numbers

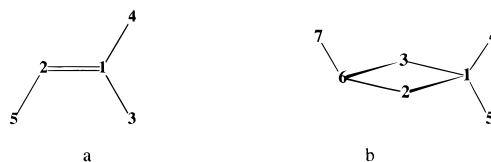
**Figure 2.** Cyclohexane in chair and boat conformations.**Table 2.** Distance Class Matrix for the Cyclohexane in Chair and Boat Conformation

|                       | 1 | 2 | 3 | 4 | 5 | 6 |
|-----------------------|---|---|---|---|---|---|
| a. Chair Conformation |   |   |   |   |   |   |
| 1                     | 0 |   |   |   |   |   |
| 2                     | 1 | 0 |   |   |   |   |
| 3                     | 1 | 2 | 0 |   |   |   |
| 4                     | 2 | 1 | 3 | 0 |   |   |
| 5                     | 2 | 3 | 1 | 2 | 0 |   |
| 6                     | 3 | 2 | 2 | 1 | 1 | 0 |
| b. Boat Conformation  |   |   |   |   |   |   |
| 1                     | 0 |   |   |   |   |   |
| 2                     | 1 | 0 |   |   |   |   |
| 3                     | 1 | 2 | 0 |   |   |   |
| 4                     | 3 | 1 | 5 | 0 |   |   |
| 5                     | 3 | 5 | 1 | 2 | 0 |   |
| 6                     | 4 | 3 | 3 | 1 | 1 | 0 |

(DCNs) which are gathered in the Distance Class Matrix (DCM). Distances can be rounded off to different levels of precision giving some flexibility in what concerns small deviations from the symmetry of a molecule. Table 2 (parts a and b) contain the DCMs obtained for the chair and boat conformations of cyclohexane, respectively, rounding the distances to 0.01 Å. In the case of the boat conformation, although the Euclidean distance matrix contains six different values, only five different class distances are obtained with this level of rounding, and four classes would be obtained if rounding to 0.1 Å. Nevertheless, the correct geometrical classes would be obtained in all cases by application of the method described below.

For each atom from the same topological class, a local geometrical index, the list of successive distance class number (LSDCN), is calculated using a relaxation procedure based on the DCN corresponding to the geometrical distance between the considered atoms and all atoms located at the same topological distance. At each iteration, a new topological environment is taken into consideration. This atomic descriptor is expanded by the ordered list of DCNs of all atoms of the next concentric level. The procedure is repeated until all the topologically equivalent atoms are differentiated or until all atoms of the molecule are taken into account. At this stage, all atoms sharing the same LSDCNs are topologically equivalent.

Table 3 (parts a and b) contain the LSDCNs obtained at each iteration for the cyclohexane chair and boat conformations, respectively. As shown in Table 3 (part a), all the carbon atoms in the chair conformation share the same LSDCNs all along the iterative procedure which ends when all atoms have been considered. On the opposite, as shown in Table 3 (part b), atoms 1 and 6 are differentiated from the four other atoms as early as the second iteration, but the procedure must be pursued because at this stage all atoms are not discriminated. As a consequence, for the chair conformation all atoms are recognized as geometrically equivalent, while for the boat conformation two classes of

**Figure 3.** Examples of molecules with atoms in Z/E positions.**Table 3.** LSDCNs for the Cyclohexane in Chair and Boat Conformation

| atom number           | iteration number |             |                 |
|-----------------------|------------------|-------------|-----------------|
|                       | 1                | 2           | 3               |
| a. Chair Conformation |                  |             |                 |
| 1                     | [1,1]            | [1,1],[2,2] | [1,1],[2,2],[3] |
| 2                     | [1,1]            | [1,1],[2,2] | [1,1],[2,2],[3] |
| 3                     | [1,1]            | [1,1],[2,2] | [1,1],[2,2],[3] |
| 4                     | [1,1]            | [1,1],[2,2] | [1,1],[2,2],[3] |
| 5                     | [1,1]            | [1,1],[2,2] | [1,1],[2,2],[3] |
| 6                     | [1,1]            | [1,1],[2,2] | [1,1],[2,2],[3] |
| b. Boat Conformation  |                  |             |                 |
| 1                     | [1,1]            | [1,1],[3,3] | [1,1],[3,3],[4] |
| 2                     | [1,1]            | [1,1],[2,3] | [1,1],[2,3],[5] |
| 3                     | [1,1]            | [1,1],[2,3] | [1,1],[2,3],[5] |
| 4                     | [1,1]            | [1,1],[2,3] | [1,1],[2,3],[5] |
| 5                     | [1,1]            | [1,1],[2,3] | [1,1],[2,3],[5] |
| 6                     | [1,1]            | [1,1],[3,3] | [1,1],[3,3],[4] |

**Table 4.** DCN Matrix of Molecule of Figure 3 (Parts a and b)

|              | 1 | 2 | 3 | 4 | 5 |
|--------------|---|---|---|---|---|
| a. Figure 3a |   |   |   |   |   |
| 1            | 0 |   |   |   |   |
| 2            | 1 | 0 |   |   |   |
| 3            | 2 | 3 | 0 |   |   |
| 4            | 2 | 3 | 4 | 0 |   |
| 5            | 3 | 2 | 5 | 6 | 0 |
| b. Figure 3b |   |   |   |   |   |
| 1            | 0 |   |   |   |   |
| 2            | 1 | 0 |   |   |   |
| 3            | 1 | 2 | 0 |   |   |
| 4            | 1 | 3 | 3 | 0 |   |
| 5            | 1 | 3 | 3 | 2 | 0 |
| 6            | 2 | 1 | 1 | 4 | 4 |
| 7            | 4 | 3 | 3 | 5 | 6 |

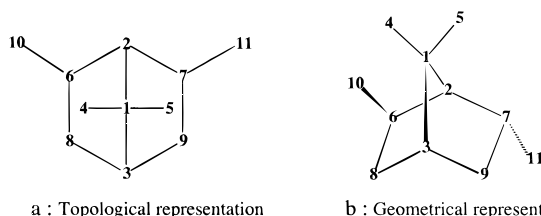
atoms made of the two stern and bow atoms and the four others, respectively, can be identified.

### EXAMPLES OF APPLICATION

The method can be applied to discriminate between all molecules containing topologically equivalent atoms belonging to different 3-D equivalence classes. A few illustrating examples will be given.

Atoms occupying Z or E positions with respect to a substituent on a double bond (Figure 3a) or on a cycle (Figure 3b) are topologically equivalent and are discriminated by their geometry when the LSDCN is expanded for the topologically equivalent atoms.

Table 4 (parts a and b) present the DCN matrices for these molecules rounded to 0.1 Å. The LSDCNs for topologically equivalent atoms: 3-4 in molecule a; 2-3 and 4-5 in molecule b are given in Table 5 (parts a and b, respectively). Examination of these tables shows that atoms 3-4 in molecule a and 4-5 in molecule b are correctly discriminated and that atoms 2-3 in molecule b are geometrically equivalent. Therefore, atoms in Z positions having a smaller LSDCN



a : Topological representation      b : Geometrical representation

**Figure 4.** Example of molecules with endo/exo (syn/anti) positions.**Table 5.** LSDCNs for Atoms 3-4 (Part a) and 2-3 and 4-5 (Part b) of Molecule of Figure 3

| atom number  | iteration number |               |               |                   |
|--------------|------------------|---------------|---------------|-------------------|
|              | 1                | 2             | 3             | 4                 |
| a. Figure 3a |                  |               |               |                   |
| 3            | [1]              | [1],[2,3]     | [1],[2,3],[4] |                   |
| 4            | [1]              | [1],[2,3]     | [1],[2,3],[5] |                   |
| b. Figure 3b |                  |               |               |                   |
| 2            | [1,1]            | [1,1],[2,3,3] |               |                   |
| 3            | [1,1]            | [1,1],[2,3,3] |               |                   |
| 4            | [1]              | [1],[3,3]     | [1],[3,3],[4] | [1],[3,3],[4],[5] |
| 5            | [1]              | [1],[3,3]     | [1],[3,3],[4] | [1],[3,3],[4],[6] |

**Table 6.** LSDCNs for Atoms 4-5, 6-7, 8-9, and 10-11 of Molecule of Figure 4

| atom number | iteration number |             |
|-------------|------------------|-------------|
|             | 1                | 2           |
| 4           | [1]              |             |
| 5           | [2]              |             |
| 6           | [1,1,2]          |             |
| 7           | [1,2,2]          |             |
| 8           | [1,2]            |             |
| 9           | [2,2]            |             |
| 10          | [1]              | [1],[15,16] |
| 11          | [1]              | [1],[13,18] |

compared to atoms in E position must receive a smaller number as shown in Figure 3 (parts a and b).

The method is also applicable, in a similar manner, to atoms in endo/exo (or syn/anti) location as in the molecule presented in Figure 4. Before taking into account the geometry, atoms 4-5, 6-7, 8-9, and 10-11 are pairwise topologically equivalent. But the construction of the LSDCNs for these atoms leads to the discrimination of atoms in endo/exo and syn/anti locations as well as atoms 6-7 and 8-9 as shown in Table 6. Based on the increasing values of these lists a unique one-to-one correspondence between the atom numbering and their 3D position can be achieved as illustrated in Figure 4b. Thus all atoms in this molecule are found to be topo-geometrically different.

The above examples have been given only as illustrations; as a matter of fact the method is general and can be applied

to any class of topologically equivalent atoms in any molecule to derive the topo-geometrical equivalence classes. Starting from the topological and geometrical description of a certain chemical compound, the algorithm can automatically recognize the topo-geometrical features of the molecule and can form the core of an automatic chemical structure perception system.

#### GENERATION OF THE CANONICAL MOLECULAR CODE

The canonical labeled matching trees built by the exhaustive search necessary to constitute the topological equivalence classes of atoms provide the link between the initial arbitrary atom numbering and the canonical one. The linear canonical molecular code is made of FROM, RING CLOSURE, ATOM, and BOND LISTS defined by Morgan.<sup>36</sup> This code contains all the topochromatic information of the molecule. It is completed by a section describing the 3-D geometry. The canonical EDM could be used for this purpose, because it is invariant with respect to the choice of the coordinate system. However the size of the EDM is proportional to the square of  $n$  (the number of atoms). If one keeps only the lower triangle of the matrix the size is  $n(n-1)$ . But an even more serious drawback originates from the fact that stereochemical information is lost in the EDM, thereby it is more convenient to keep atomic coordinates calculated with respect to a canonical system of coordinate. Using the canonical atom numbering, atom 1 is chosen as origin, atoms 1 and 2 define the X axis, atoms 1, 2, and 3 define the XY plane, the Z axis being oriented to lead to a direct coordinate system. The linear geometrical code, LGC, is then simply built by the list of atom coordinates on this system omitting the null values for the three first atoms, leading to a size of  $3n-6$ . Thus, the resulting code maintains a distinction between the topology and the geometry, and as a consequence geometrical isomers and conformers share the same topological part of the canonical code but differ by their geometrical part.

The canonical codes for the molecules given above as examples are given in Table 7.

#### CONCLUSION

The canonical numbering of molecules is by far not a new problem in the field of chemical information. The starting point of the existing methods relies almost universally on the partitioning of atoms in topological equivalence classes. Although several methods have been proposed to avoid exhaustive searches through the molecular graph by using highly discriminating local invariants, to our knowledge even

**Table 7.** Canonical Codes for the Molecules Given in the Text as Examples

| example | canonical molecular code   |
|---------|--|
| 2a      | topological: #FROM>001001002003004 #RING>005006<br>geometrical: #LGC>1.54;-0.55;1.43;2.08;0.80;1.19;0;2.23;1.19;1.54;2.23;1.19   |
| 2b      | topological: #FROM>001001002003004 #RING>005006<br>geometrical: #LGC>1.54;-0.54;1.44;2.13;0.86;1.13;0.05;2.30;1.13;1.05;1.53;2.00  |
| 3a      | topological: #FROM>001001001002 #BOND>2111<br>geometrical: #LGC>1.34;-0.79;1.28;-0.76;-1.31;0;2.20;1.24;0  |
| 3b      | topological: #FROM>001001001002006 #RING>003006<br>geometrical: #LGC>1.57;0;1.57;0;-0.61;-0.61;1.29;-0.61;-0.61;-1.29;1.57;1.57;0;2.20;2.20;1.26   |
| 4       | topological: #FROM>001001001002002003003006006 #RING>006008007009<br>geometrical: #LGC>1.56;-0.02;1.54;-0.68;-0.66;1.22;-0.67;-0.65;-1.23;1.96;0.89;1.20;1.87;0.78;<br>-1.30;0.81;1.91;1.26;0.83;1.92;-1.24;2.29;0.14;2.50;3.31;1.26;-1.54 |

the more recently published methods of this type fail in special cases. Therefore, these methods must be completed by an exhaustive search if one wants to be sure to obtain a correct solution in all cases. The computational cost of this search can be maintained within reasonable limits, even for large regular graphs, by using an efficient initial topological partitioning method and by using optimization techniques to prune the search as soon as possible. Moreover, this unavoidable exhaustive search can be exploited to determine the topological equivalence classes of atoms and bonds, as a byproduct result of the canonical numbering. Atoms belonging to the same class can be further discriminated between different geometric classes in order to determine the so called topo-geometric equivalence classes. In the present paper we have presented an efficient algorithm for the topo-geometrical determination of equivalence classes of the atoms from a 3-D molecular structure. The algorithm was implemented in PROLOG, and we have presented some simple examples of automatic (computer) recognition of the geometrical equivalent atoms. A scheme for completing the molecular topo-chromatic canonical code by a canonical geometric Cartesian coordinates is presented.

The method proposed here has a wide range of applications in chemical information systems: coding of chemical compounds in 3-D databases, searching in molecular databases, 3-D similarity and clustering determination, 3-D substructure search, and pharmacophore search. Also, the classes of equivalent atoms generated with the algorithm described in this work allows one to generate substitution and labeled isomers and to determine the number of signals and equivalent atoms in NMR and ESR spectra. The topo-geometrical code generated with the Prolog program can differentiate between stereoisomers and different conformations of the same molecule; work is in progress along these lines.

## REFERENCES AND NOTES

- Balaban, A. T.; Motoc, I.; Bonchev, D.; Mekenyan, O. Topological Indices for Structure-Activity Correlations. *Top. Curr. Chem.* **1983**, *114*, 21-55.
- Balaban, A. T. Using Real Numbers as Vertex Invariants for Third Generation Topological Indexes. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 23-28.
- Balaban, A. T. Lowering the Intra- and Intermolecular Degeneracy of Topological Invariants. *Croat. Chem. Acta* **1993**, *66*, 447-458.
- Balaban, A. T. Local versus Global (i.e. Atomic versus Molecular) Numerical Modeling of Molecular Graphs. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 398-402.
- Balaban, A. T. Real-Number Local (Atomic) Invariants and Global (Molecular) Topological Indices. *Rev. Roum. Chim.* **1994**, *39*, 245-257.
- Randić, M. On Characterization of Three-Dimensional Structures. *Int. J. Quantum Chem.: Quantum Biol. Symp.* **1988**, *15*, 201-208.
- Randić, M.; Jerman-Blažič, B.; Trinajstić, N. Development of 3-Dimensional Molecular Descriptors. *Comput. Chem.* **1990**, *14*, 237-246.
- Mihalić, Z.; Trinajstić, N. The Algebraic Modelling of Chemical Structures: On the Development of Three-Dimensional Molecular Descriptors. *J. Mol. Struct. (Theochem)* **1991**, *232*, 65-78.
- Diudea, M. V.; Horvath, D.; Graovac, A. Molecular Topology. 15. 3D Distance Matrices and Related Topological Indices. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 129-135.
- Randić, M.; Razingar, M. Molecular Topographic Indices. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 129-135.
- Pepperrell, C. A.; Willett, P. Techniques for the Calculation of Three-Dimensional Structural Similarity Using Inter-Atomic Distances. *J. Comput.-Aided Mol. Design* **1991**, *5*, 455-474.
- Willett, P. A Review of Three-Dimensional Chemical Structure Retrieval Systems. *J. Chemom.* **1992**, *6*, 289-305.
- Bures, M. G.; Martin, Y. C.; Willett, P. Searching Techniques for Databases of Three-Dimensional Chemical Structures. *Top. Stereochem.* **1994**, *21*, 467-511.
- Balasubramanian K. Computer Perception of Molecular Symmetry. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 761-770.
- Nikolić, S.; Trinajstić, N.; Mihalić, Z.; Carter, S. On the Geometric Distance matrix and the Corresponding Structural Invariants of Molecular Systems. *Chem. Phys. Lett.* **1991**, *179*, 21-28.
- Balasubramanian K. Graph Theoretical Perception of Molecular Symmetry. *Chem. Phys. Lett.* **1990**, *169*, 224-230.
- Cripen, G. M.; Havel, T. F. Stable Calculation of Coordinates from Distance Information. *Acta Crystallogr.* **1978**, *A34*, 282-284.
- Laidboeur, T.; Cabrol-Bass, D.; Ivanciuc, O. Determination of Topological Equivalence Classes of Atoms and Bonds in C<sub>20</sub>-C<sub>60</sub> Fullerenes Using a New Prolog Coding Program. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 811-821.
- Trucco, E. A Note on the Information Content of Graphs. *Bull. Math. Biophys.* **1956**, *18*, 123-135.
- Carhart, R. E. Erroneous Claims Concerning the Perception of Topological Symmetry. *J. Chem. Inf. Comput. Sci.* **1978**, *18*, 108-110.
- Hu, C. Y.; Xu, L. Algorithm for Computer Perception of Topological Symmetry. *Anal. Chim. Acta* **1994**, *295*, 127-134.
- Xuixiao, H.; Xinquan, X. ESSESA: An Expert System for Structure Elucidation from Spectral. 4. Canonical Representation of Structures. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 730-734.
- Balaban, A. T.; Mekenyan, O.; Bonchev, D. Unique Description of Chemical Structures Based on Hierarchically Ordered Extended Connectivities (HOC Procedures). I. Algorithms for Finding Graph Orbits and Canonical Numbering of Atoms. *J. Comput. Chem.* **1985**, *6*, 538-551.
- Mekenyan, O.; Bonchev, D.; Balaban, A. T. Unique Description of Chemical Structures on Hierarchically Ordered Extended Connectivities (HOC Procedures). II. Mathematical Proofs for the HOC Algorithm. *J. Comput. Chem.* **1985**, *6*, 552-561.
- Balaban, A. T.; Mekenyan, O.; Bonchev, D. Unique Description of Chemical Structures Based on Hierarchically Ordered Extended Connectivities (HOC Procedures). III. Topological, Chemical, and Stereochemical Coding of Molecular Structure. *J. Comput. Chem.* **1985**, *6*, 562-569.
- Bonchev, D.; Mekenyan, O.; Balaban, A. T. Unique Description of Chemical Structures Based on Hierarchically Ordered Extended Connectivities. IV. Recognition of Graph Isomorphism and Graph Symmetries. *MATCH (Commun. Math. Chem.)* **1985**, *18*, 83-99.
- Mekenyan, O.; Balaban, A. T.; Bonchev, D. Unique Description of Chemical Structures Based on Hierarchically Ordered Extended Connectivities. VII. Condensed Benzenoid Hydrocarbons and Their <sup>1</sup>H NMR Chemical Shifts. *J. Magn. Reson.* **1985**, *63*, 1-13.
- Ralev, N.; Karabunarliev, S.; Mekenyan, O.; Bonchev, D.; Balaban, A. T. Unique Description of Chemical Structures Based on Hierarchically Ordered Extended Connectivities (HOC Procedures). VIII. General Principles for Computer Implementation. *J. Comput. Chem.* **1985**, *6*, 587-591.
- Mekenyan, O.; Bonchev, D.; Balaban, A. T. Hierarchically Ordered Extended Connectivities. Reflection in the <sup>1</sup>H NMR Chemical Shifts of Condensed Benzenoid Hydrocarbons. *Chem. Phys. Lett.* **1984**, *109*, 85-88.
- Jochum, C.; Gasteiger, J. Canonical Numbering and Constitutional Symmetry. *J. Chem. Inf. Comput. Sci.* **1977**, *17*, 113-117.
- Jochum, C.; Gasteiger, J. On the Misinterpretation of Our Algorithm for the Perception of Constitutional Symmetry. *J. Chem. Inf. Comput. Sci.* **1979**, *19*, 49-50.
- Hall, L. H.; Kier, L. B. Determination of Topological Equivalence in Molecular Graphs from the Topological State. *Quant. Struct.-Act. Relat.* **1990**, *9*, 115-131.
- Hu, C. Y.; Xu, L. Algorithm for Computer Perception of Topological Symmetry. *Anal. Chim. Acta* **1994**, *295*, 127-134.
- Diudea, M. V.; Minailiuc, O.; Balaban, A. T. Molecular Topology. IV. Regressive Vertex Degrees (New Graph Invariants) and Derived Topological Indices. *J. Comput. Chem.* **1991**, *12*, 527-535.
- Fan, B. T.; Barbu, A.; Panaye, A.; Doucet, J.-P. Detection of Constitutionally Equivalent Sites from a Connection Table. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 654-659.
- Morgan, H. The Generation of a Unique Machine Description for Chemical Structures. A Technique Developed at Chemical Abstract Service. *J. Chem. Doc.* **1965**, *5*, 107-113.