

# Structure Generation from a Gross Formula. 7. Graph Isomorphism: A Consequence of the Vertex Equivalence<sup>1</sup>

Ivan P. Bangov†

Institute of Organic Chemistry, Bulgarian Academy of Sciences, Building 9, Sofia 1113, Bulgaria

Received May 18, 1993\*

The origin of isomorphism in the case of structure generation is examined. It is shown that isomorphism is a consequence of the vertex, hence of bond equivalence. If two equivalent potential extensions are implemented into actual ones in different stages of the generation process, a multiple isomorphism results. A multilevel depth-first procedure for avoiding the generation of duplicated structures is developed. It is based on a two-row matrix representation of the molecular graph. The vertex equivalence is examined at each step of the graph construction, and only non-equivalent extensions are allowed.

## INTRODUCTION

For nearly 10 years we have been developing an approach to structure generation.<sup>1-6</sup> It was noticed that most of the duplications during the graph-construction process are due to permutations between equivalent atoms.<sup>4-6</sup> Moreover, in the course of this development it was found that isomorphism may be a consequence of the vertex equivalence. This paper is an attempt to consider critically and more rigorously the nature of the isomorphism problem in the particular case of graph (structure) generation.

There are several definitions of isomorphism.<sup>7-11</sup> Mathematically it can be expressed by the following relationship:<sup>10,11</sup>

$$A(G_2) = P_{12}^{-1} A(G_1) P_{12} \quad (1)$$

Here  $G_1$  and  $G_2$  are two graphs and  $A(G_1)$ ,  $A(G_2)$  are their adjacency matrices. A graph is usually defined as:

$$G = (V, E) \quad (2)$$

where  $V$  is the set of vertices  $v_i \in V$  and  $E$  is the set of edges. The set  $E$  may be expressed as

$$E = f(V \otimes V) \quad (3)$$

where  $V \otimes V$  is the Cartesian product of all vertices and  $f$  is an operator which selects only some of the  $(v_i, v_j)$  pairs forming the edges  $e_{ij}$ .  $P$  is a permutation matrix, which is orthogonal, hence relation 1 can be rewritten as follows:<sup>11</sup>

$$A(G_2) = P_{12}^T A(G_1) P_{12} \quad (1')$$

Here  $P^T$  is the transposed  $P$ . Relation 1 ((1')), respectively, practically permutes two columns and two rows of the adjacency matrix. For example the two graphs a and b of Figure 1 obtained after permuting vertices 3 and 4 are isomorphic. One can easily see from their adjacency matrix representations in Figure 1 that the latter is obtained from the former by permutations of rows 3 and 4, and columns 3 and 4. One can accomplish  $N!$  such permutations generating  $N!$  isomorphic graphs for each non-isomorphic one ( $N$  is the number of atoms). Thus, if  $M$  non-isomorphic graphs are generated, then isomorphism will increase their number to  $M \times N!$  graphs. Obviously, any generation of such permutations must be avoided in order for the isomorphism redundancy to be reduced.

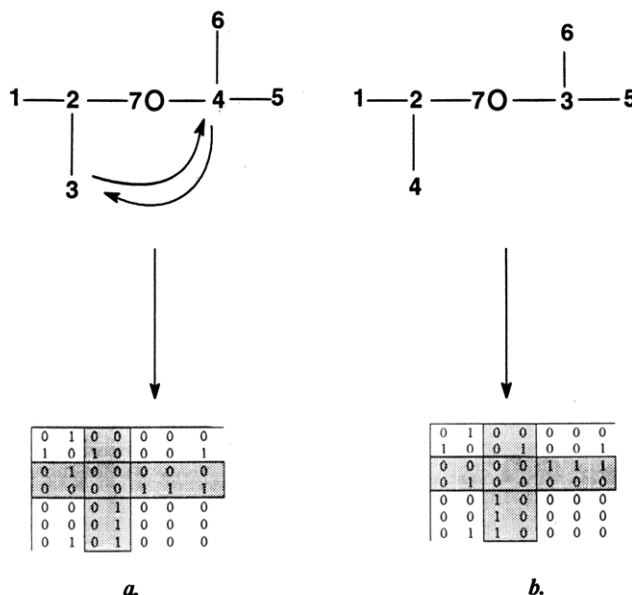


Figure 1. Two isomorphic graphs and their adjacency matrix representations generated from the operation given by expressions 1 and 1'.

However, the use of relations 1 or 1' within a generation procedure does not contribute very much. There is no way to generate a new structure from any such permutation. This is so because relation 1 in fact does not permute the vertices themselves but their labels, which are arbitrary integer numbers. In a recent paper Klin and Zefirov<sup>12</sup> introduce two types of graphs: *numbered* and *abstract*. In a numbered graph  $G$  each vertex  $v_i \in V$  is given a label usually a number from the set  $N = \{1, 2, \dots, n\}$  (the notion numbered graph practically coincides with the well-known notion of *labeled graph* in the cases where the vertices are labeled by numbers). In contrast, the abstract graph is not related to any numbering of its vertices and it uniquely represents the structure of the studied object. The notion of an *unlabeled graph* is related to an abstract graph, but while the former is defined only as a juxtaposition to the numbered graph, the latter is defined as an invariant of the induced action of a symmetry group  $S_n$  of the set  $N$  (for a strict mathematical definition of abstract graph see ref 12). Hence, every numbered graph is a concrete mathematical representation of the abstract graph, and numerous numbered graphs correspond to one abstract graph. Thus, the permutations expressed with relation 1 ((1')), respectively, simply permute the labels but not the vertices (atoms) of the abstract

<sup>†</sup> Presently, a Senior Guest Scientist on a NATO grant at the Université de Nice, Sophia-Antipolis, Parc Valrose 06108, Nice Cedex 2, France.

\* Abstract published in *Advance ACS Abstracts*, February 15, 1994.

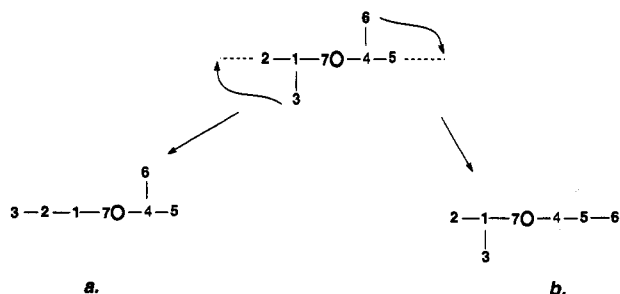


Figure 2. Generation of two isomorphic structures by breaking the bonds 1-3 and 4-6 and creating two new bonds 2-3 and 5-6.

graph. They produce different representations (numbered graphs of the abstract graph) but no new abstract graph results in them. To eliminate the isomorphic structures, one out of all such representations is usually singled out as canonical, the others being discarded. The maximal adjacency matrix<sup>10,13</sup> or the highest connectivity stack<sup>14</sup> structures can be selected to this end. However, this procedure requires additional permutations in order for the canonical representation to be perceived, and it encounters severe problems when treating fragments.

Some isomorphic structures as a and b in Figure 2 cannot be represented by simple relations such as (1) and (1'). They may be considered as being produced by a series of such operations. In order to have a deeper insight in the isomorphism problem, we need another model.

Permutations of the vertices (actually labels) of the numbered graphs preserving the adjacency matrix are considered *automorphic*, and permutations changing it, *isomorphic*. There is one strange contradiction: symmetry operations represented by an *automorphism partitioning* (partitioning into equivalence classes) do not always form an *automorphic group*.<sup>11</sup> This discrepancy can be explained with the fact that whereas an automorphic group is defined on numbered graphs, the automorphism partitioning is carried out on abstract graphs, taking implicitly into account the real topological symmetry of the vertices. In contrast, any arbitrary labeling in a numbered graph, hence the corresponding adjacency matrix, does not take the topological symmetry into account. Consequently, we need a new algebraic representation of the molecular graph which is closer to the abstract than to the numbered graph.

## TWO-ROW MATRIX REPRESENTATION OF THE MOLECULAR GRAPH

Our mathematical representation of the molecular graph is based on *directed* graphs. The directed graph of the structure from Figure 1 is presented in Figure 3a. Here, instead of edges the links between the vertices form *arcs*. Each arc is an ordered pair  $(v_i, v_j)$  of vertices. However, to include the fragments within the generation procedure, bonding sites (BSs) instead of atoms are considered to be elements of the combinatorial process. Whereas the vertices in graph theory are associated with the atoms in chemistry and the arcs with the chemical bonds, the BSs are associated with the atom free valences. Thus, a bond (arc)  $(v_i, v_j)$  is formed by two BSs: the former where the arc is going out of atom  $v_i$ , denoted by an arrow ( $\leftarrow$ ) in Figure 3, and the latter where the arc is going in the atom  $v_j$ , denoted as  $+$ . In previous papers<sup>1,4-6</sup> we called these saturating valences (SVs) and saturation sites (SSs), respectively. Hence, the mathematical representation of this model is carried out by partitioning the BSs into two sets: SVs  $\in V^s$ , and SSs  $\in S^s$ . It is seen from Figure 3a that all

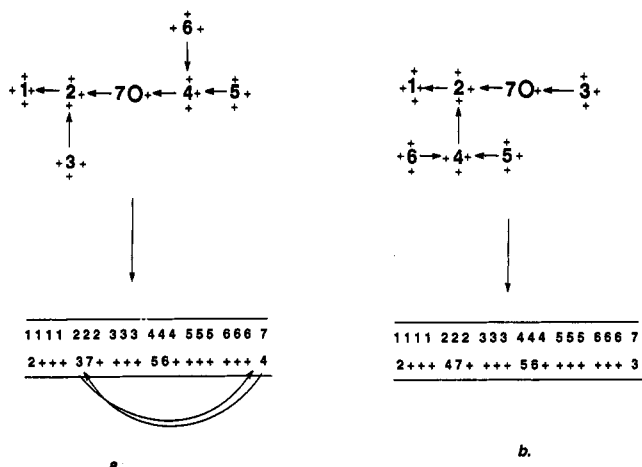


Figure 3. Directed graph and two-row matrix representation of the chemical structure (a) and generation of a new structure from the permutation between the SVs of atoms 3 and 4 (b).

the vertices (atoms) except the first have  $n-1$  SSs and 1 SV. All the BSs of the first atom being SSs. Here  $n$  is the atom valence (the degree of the vertex). In atoms where a cycle closes one more BS is transformed into a SV (see atom 2 from Figures 4 and 5). This partitioning of the BSs into SVs and SSs corresponds to the *father/successor*<sup>15,16</sup> relationship in rooted trees. Each atom, except the first has 1 father and  $n-1$  successor nodes. The first atom (vertex) has no father vertex.

The fragments are treated in a manner similar to the single atoms. Each fragment except the first has 1 SV and  $n-1$  SSs. All first fragment BSs are of SS type. Here  $n$  is the number of the BSs in a given fragment. The only difference is that the BSs in the fragments are not equivalent. This requires all non-equivalent BSs to be considered in turn as once being a SV while the remaining are of SS type.

The two-row matrix representation of the structure from Figure 1 is given in Figure 3a. The first-row elements are the SSs given with their vertex numbers, and the second-row elements are the SVs. Any juxtaposing of a first-row SS to the second-row SV element represents an arc (chemical bond). The BSs which are not saturated are denoted with  $+$  in the second row. The SVs of the hydrogen atoms are not considered, but after a whole complete structure is formed they fill the unsaturated SSs. All the SV transpositions with respect to the SSs produce the complete set of graphs  $\Gamma$ . Accordingly, any permutation between two second-row elements corresponds to a bond break and a creation of a new bond.

One can see that, in contrast to the case in Figure 1, such a permutation between SVs 3 and 4 (between atoms 3 and 4, respectively) leads to a new structure presented in Figure 3b. This difference with relations 1 and 1' could be explained with the fact that by permuting the SVs, which emanate from the atoms, we permute the atoms themselves, not their labels, as in the case of numbered graphs. In the same manner if a SV emanates from a fragment, its permutation with another SV emanating either from an atom or from another fragment leads to permutation of the whole fragment with the corresponding atom (fragment, respectively). It looks as though the SV drags the atom or the whole fragment from which it emanates. Hence, each structure may be represented by a permutation of the second-row (both saturated or unsaturated) elements. If we have  $L$  SSs and  $m$  SVs and free BSs, then  $L!/(L-m)!$  such permutations will be generated in total. Most of them produce disjoint structures, others have no chemical meaning, and a great part of them are isomorphic.

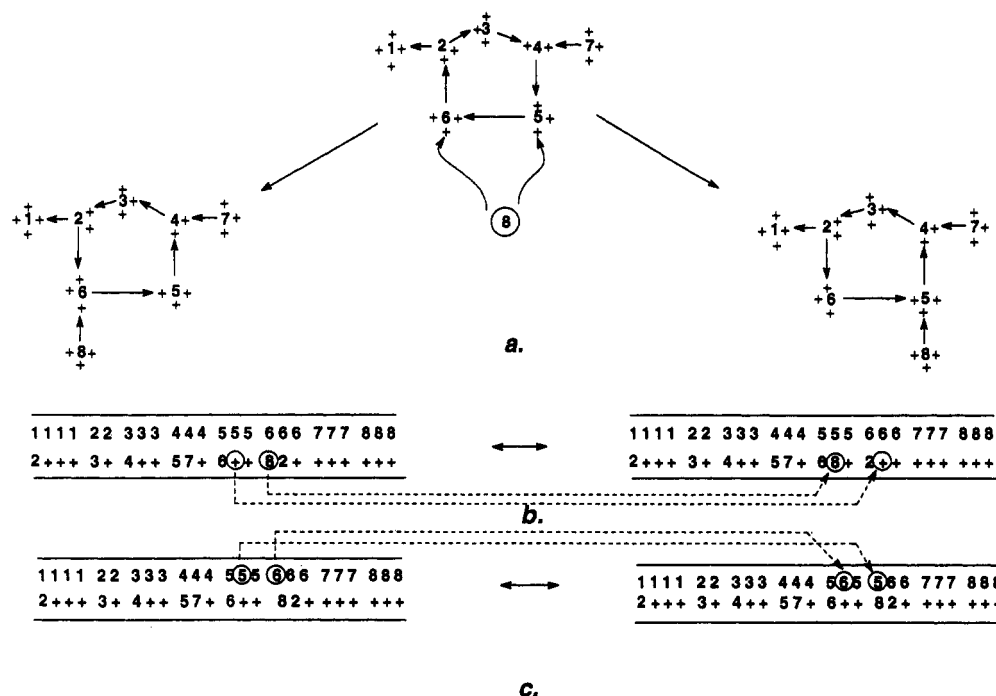


Figure 4. Generation of two isomorphic extensions from two equivalent atom SSs.

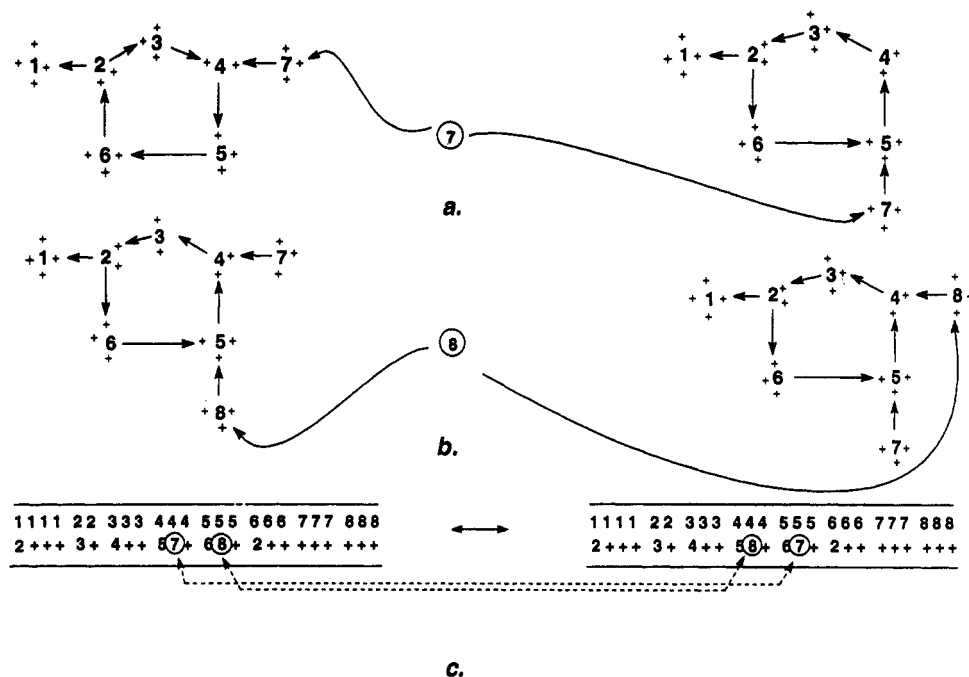


Figure 5. Generation of two isomorphic extensions by permutations of two equivalent atom SVs.

Evidently, some rules concerning the chemical nature and the topological symmetry properties must be applied. But what is more important, the two-row representation gives us a new tool for examining the isomorphism problem in the case of structure generation.

#### TOPOLOGICAL SYMMETRY AND ISOMORPHISM

Topological symmetry is related to automorphism partitioning. This partitioning separates the vertices into *equivalence classes*.<sup>8,11</sup> According to Gutman<sup>9</sup> two vertices are *equivalent* if a permutation exists between these two vertices which is invariant for the edges. Here is a formal definition:

Let  $x_i$  and  $x_j$  be two vertices of the graph  $G = (V, E)$ . If  $\pi$  is a permutational operator and, from  $\pi(x_i) = x_j$ , it follows that  $\pi(E) = E$ , then  $x_i$  and  $x_j$  are equivalent.

Following ref 11 such permutations lead either to isomorphism or to automorphism (which is a special case of isomorphism). Hence, it is clear that the equivalence partitioning of vertices and isomorphism are closely related, and isomorphism can be formally defined as follows:

Let  $G_1 = (V_1, E_1)$  and  $G_2 = (V_2, E_2)$  be two graphs in which  $G_1$  is isomorphic to  $G_2$ ; if for an operation  $\pi(V_1) = V_2$ , it follows that  $\pi(E_1) = E_2$ , i.e.,  $\pi(G_1) = G_2$ .

The equivalence class partitioning resembles the magnetic equivalence in <sup>13</sup>C NMR spectra,<sup>11</sup> provided no accidental signal overlapping is observed. In order to avoid the manipulation of numbered graphs, we label additionally the vertices in a manner similar to the NMR signal assignment. Thus, a representation corresponding to the abstract, rather to the numbered graph is formed. Each vertex  $v_i \in V$  is labeled by

a local index from the set  $R = \{r_1, r_2, \dots, r_n\}$ . Each arc  $(v_i, v_j) \in E$  can be also labeled by a bond index  $b_i$  defined as  $b_i = r_i r_j$ .<sup>5</sup> Let  $S_g$  be a symmetry group of  $V$  and  $S$  is an operator from this group. If the action  $S(v_i) = v_j$  leads to  $r_i = r_j$ , we say that the two vertices  $v_i$  and  $v_j$  are equivalent. In this way all the vertices can be partitioned into *equivalence classes*. Hence, an isomorphism  $\pi(G_1) = G_2$  exists if the following relation holds:  $R_1 = R_2$ . In a recent paper<sup>17</sup> a global charge related topological index (CTI) was devised by us according to the following expression:

$$CTI = \sum_{ij} \frac{r_i r_j}{D_{ij}}$$

Two graphs  $G_1$  and  $G_2$  may be considered isomorphic if  $CTI_1 = CTI_2$ .

Further, we shall use the notion *extension of a graph* as defined by Foulon:<sup>18</sup>  $G_1 = (V_1, E_1)$  is an extension of  $G = (V, E)$  if for  $Y \subset V$  and  $E \subset E_1$ ,  $Y$  is saturated in  $E_1$  and every edge (arc in our case) in  $E - E_1$  has at least one vertex in  $Y$ . Hereafter, we use extensions which differ by one bond  $e_{ji} = E - E_1$ . Moreover, two types of extensions will be defined here: *potential* and *actual*. A potential extension  $G_1 = (V_1, E_1)$  of the graph  $G = (V, E)$  is an extension which contains a vertex  $v_1 = V_1 - V$ , but no bond is formed; i.e.,  $E_1 = E$ . If during the generation process such a bond is created, the extension is transformed into actual. In the same manner, a potential bond (arc)  $e_1^p = (v_0, v_1)^p$  may be defined. During the generation process the latter is transformed into an actual bond (arc)  $e_1^a = (v_0, v_1)^a$ . The difference between  $e_1^p$  and  $e_1^a$  is that the vertices  $v_0$  and  $v_1$  before and after saturation are in different classes. Accordingly their bond indices are different  $b_1^p \neq b_1^a$ . For example, two potential bonds  $e_{5,8}^p = (v_5, v_8)^p$  and  $e_{6,8}^p = (v_6, v_8)^p$  and their implementation in the actual bonds  $e_{5,8}^a = (v_5, v_8)^a$  and  $e_{6,8}^a = (v_6, v_8)^a$  are presented in Figure 4a. This specification is necessary as graph generation is a combinatorial procedure; each new operation depends on the previous ones. Accordingly, a dynamic partitioning of the vertices into equivalence classes is required. After each extension a new automorphism partitioning must be carried out.

Evidently, if two potential bonds (arcs)  $(v_1, v_2)^p$  and  $(v_3, v_4)^p$  are equivalent, then the corresponding actual bonds  $(v_1, v_2)^a$  and  $(v_3, v_4)^a$  will be also equivalent. The assertion that the two potential bonds  $(v_1, v_2)^p$  and  $(v_3, v_4)^p$  are equivalent means that there is a symmetry operation  $\pi$  which when applied to the vertices results in  $\pi(v_1) = v_3$  and  $\pi(v_2) = v_4$ . Evidently, after implementation of these potentials bonds into actual ones these relations remain in force.

Here the following theorem will be proved:

**Theorem 1.** Let  $C_1, C_2, \dots, C_p$  be an equivalence class partitioning. Let  $(v_1, v_2)^p$  and  $(v_3, v_4)^p$  be two potential bonds and let  $v_1, v_3 \in C_i$ , when  $v_2, v_4 \in C_j$ . If  $(v_1, v_2)^p$  has once been implemented as an actual bond  $(v_1, v_2)^a$  and further, during the generation process, the bond  $(v_3, v_4)^p$  is implemented as an actual bond  $(v_3, v_4)^a$ , then  $(m-1)!$ -fold isomorphism results.

**Proof:** The implementation of the  $(v_2, v_2)^p$  bond into actual means that 1 SV saturates 1 SS. Then, during the generation process, the remaining  $m-1$  SVs will produce  $(m-1)!$  permutations resulting in a set  $\Gamma_1$  of  $(m-1)!$  extensions, all having the common bond  $(v_1, v_2)^a$ . In the same way an implementation of the  $(v_3, v_4)^p$  bond into an actual one will produce the set  $\Gamma_2$  of  $(m-1)!$  extensions, with the common bond  $(v_3, v_3)^a$ . Since the sets  $\Gamma_1$  and  $\Gamma_2$  are complete for every initial graph  $G = (V, E)$ , one can always find one pair of

extensions  $G_1 = (V_1, E_1)$  and  $G_2 = (V_2, E_2)$  where  $v_1, v_3 \in V$ ,  $v_2 = V_1 - V$ , and  $v_4 = V_2 - V$ . So as long as  $\pi(v_2) = v_4$  and  $\pi(v_1) = v_3$ , then  $\pi(V_1) = V_2$ , and  $\pi(E_1) = E_2$ ; hence  $\pi(G_1) = G_2$ , and finally this will result in  $(m-1)!$  one-to-one pairs of isomorphic graphs.

Evidently, the number  $(m-1)!$  is valid only within our method and a two-row matrix representation. However, according to theorem 1, in order to avoid isomorphism, the generation procedure must implement only unique potential bonds into actual ones. This means that if a combination of two vertices belonging either to the same or to distinct equivalence classes is implemented once into an actual bond, no other combination of vertices belonging to the same classes must be implemented into an actual bond during the whole generation process. Accordingly, if the set of all bond indices  $b_{ij}$  is generated, during the whole generation process, no two indices will have the same values in this set.

Any perception of equal  $b_{ij}$ 's is not practical. Hence, the generation of duplications may be avoided by taking at each step the topological symmetry of the generated extensions into account. Here we shall consider some cases of isomorphism resulting from topological symmetry.

Let  $G = (V, E)$  be a graph, and its vertices are partitioned into equivalence classes. Let  $G_1 = (V_1, E_1)$  and  $G_2 = (V_2, E_2)$  be two extensions of the graph  $G$  obtained by linking in turn a new vertex  $v_p$  to two vertices  $v_1$  and  $v_2$ . If the vertices  $v_1$  and  $v_2$  belong to the same equivalence class, then the two graphs  $G_1$  and  $G_2$  are isomorphic.

Evidently  $V_1 = V_2$  because they are formed of the  $V$  ( $V \subset V_1$ ,  $V \subset V_2$ ) and the vertex  $v_p$ . The difference between extensions  $G_1$  and  $G_2$  is only between the sets  $E_1$  and  $E_2$ , which differ by the arcs  $(v_p, v_1)$  and  $(v_p, v_2)$ . As long as, the atoms  $v_1$  and  $v_2$  are equivalent, it follow from theorem 1 that  $\pi(E_1) = E_2$ ; hence  $\pi(G_1) = G_2$ .

This is illustrated in Figure 4a with the two extensions formed by vertex 8. Their two-row matrix representations are provided in Figure 4b. The two consecutive saturations of the two equivalent vertices 6 and 5 may be viewed as a permutation between the free valence BS at vertex 5 with the vertex 6 SV. But the same result can be obtained if the two SSs from the equivalent atoms 5 and 6 are permuted, as shown in Figure 4c. It can be easily proved that if we form the extensions  $G_1$  and  $G_2$  by using two equivalent vertices  $v_p$  and  $v_g$  instead of the only vertex  $v_p$  (applied in turn twice), the extensions  $G_1$  and  $G_2$  will be again isomorphic. This is illustrated in Figure 5. Atom 7 SV saturates sequentially atoms 4 and 5 (Figure 5a), and atom 8 SV saturates atoms 5 and 4, respectively (Figure 5b). One can see from Figure 5c that these saturations may be viewed as permutations of the equivalent atom 7 and 8 SVs.

A different way of obtaining isomorphic graphs is given in Figure 2. Two isomorphic structures appear after breaking of the equivalent bonds 9(1,3) (Figure 2a) and (4,6) (Figure 2b) and the subsequent creation of the bonds (2,3) and (5,6), respectively. Note, that the two potential bonds (3,2) and (5,6) have the same vertex partitioning, hence two isomorphic structures result according to theorem 1.

## STRUCTURE GENERATION METHOD

The considerations presented above show that in order to avoid isomorphism the equivalence class partitioning must be perceived after each new extension is created and no repeated implementations of potential equivalent bonds into actual are allowed. Accordingly, the following procedure was developed:<sup>1,5,6</sup>

**Table 1.** Valences and Initial Values  $r_0$  for Different Types and Hybridization States

type and hybridization state	valence	$r_0$
C		
sp <sup>3</sup>	4	4
sp	2	7
sp <sup>2</sup> (olefinic)	3	11
sp <sup>2</sup> (aromatic)	3	13
N		
sp <sup>3</sup>	3	15
sp	1	18
sp <sup>2</sup> (non-aromatic)	2	20
sp <sup>2</sup> (aromatic)	2	20
O		
sp <sup>3</sup>	2	23
sp <sup>2</sup> (non-aromatic)	1	25
sp <sup>2</sup> (aromatic)	1	25
F	1	32
Cl	1	33
Br	1	34
I	1	35
S		
sp <sup>3</sup>	2	28
sp <sup>2</sup> (non-aromatic)	1	29
sp <sup>2</sup> (aromatic)	1	29

The atoms and their SVs and SSs, respectively, are split into several classes, and the SVs are ranked as follows:

- h SVs of heteroatoms
- m SVs of multiple bond atoms
- c SVs of single carbon atoms or carbon atoms forming chemical groups with valences equal to or greater than 2
- l SVs of carbon atoms from univalent fragments and chemical groups

These classes are additionally partitioned into subclasses according to the aforementioned local indices  $r$ . Recently, the construction of new types of local indices have been discussed.<sup>19</sup> Here, a charge related topological index developed by us is employed to this end. It is defined by the following expression:<sup>1,4-6,17,20</sup>

$$r = r_0 + N_H - q_C \quad (4)$$

where  $r_0$  is a constant chosen for each type of atom and hybridization state in such a way that its values for two atoms or two hybridization states of the same atom do not overlap. The  $r_0$  values are provided in Table 1.  $N_H$  is the number of the hydrogen atoms, directly attached to a given atom, and  $q_C$  is the charge density, calculated by a fast procedure (the method of Gasteiger et al.<sup>21</sup> was used to this end).

Finally, the atoms (their SVs, respectively) being of the same equivalence class, are ranked according to their numbering; the smaller the number, the lower the level it occupies. Thus, a hierarchical system of levels is constructed, each SV occupying one level. It is evident from relation 4 that all free atoms of a given type, e.g., carbon atoms, will form a separate class, but taking different levels. The bonded carbon atoms will form other classes depending on the topological symmetry of the substructure formed at a given level. This partitioning is dynamic, and with every new extension at a given level, new equivalent class partitioning and SV ranking for the highest SV levels are carried out. Hence, the structure generation is a recursive depth-first procedure consisting of the following steps: equivalence class atom partitioning, SS-selection, and SS-SV saturation.

**Step 1:** This step is carried out at the beginning of the generation process. The local indices  $r$  are calculated, and an

initial SV-SS partitioning with a subsequent rank of the SV levels is carried out.

**Step 2:** At each level, starting from the zero level (before the first), a SS selection procedure for the next (higher) level is carried out. It produces a list of SSs which will form bonds at the higher level SV-SS saturation (hereafter called the SS-list). The SS selection procedure is subjected to the following rules: (i) *no SV having the same number as the corresponding SV is selected* (atoms cannot be bonded to themselves). (ii) *If the higher level SV is of the same equivalence class (has the same  $r$ ) as the current level SV, then the SS selection starts from the next SS of the last atom saturated at the current level; otherwise it starts from the first free SS of the first atom.* This rule prevents the repeated extensions with two equivalent SV atoms. If the SS selection were from the first atom, permutations between the two equivalent SV atoms will result, as in the case in Figure 5. (iii) *Only one out of all equivalent SSs is selected for the next level SV-SS saturation. In cases of selection of closure bond SSs only one out of all SSs being both equivalent and equidistant (of the same distance from the SV atom) is selected.* This rule avoids any repeated extensions, formed by repeated bonding of the current SV to two equivalent SS atoms, such as those presented in Figure 4. (iv) *If the SV and the SSs are of the same equivalence class, then SSs having a number lower than the SV atom number are selected only for the next SV-SS saturation.* (v) *In the cases where the selection procedure starts from the last saturated atom, and an atom of a given equivalence class appears before this atom, no new SSs of atoms belonging to this class are further selected.* This rule assumes that a bond between a lower level SV (being of the same equivalence class as the current SV) and a SS being before the last saturated atom SS has already been created; any selection of a new SS of the same class will produce repeatedly an equivalent bond, hence an isomorphic extension.

Rules ii and iv lead to an ordering of the equivalent atoms in an ascending order of their numbering. Hence, a canonical form similar to the maximal adjacency matrix<sup>10,13</sup> or highest connectivity stack<sup>14</sup> is naturally formed.

**Step 3:** The SV-SS saturation step leads to formation of a new extension. So as long as this is a depth-first procedure with a backtrack, each time a backtrack is carried out an old SV-SS bond is cut and a new bond is created in the next SV-SS saturation step using a new SS from the SS-list.

The SS-selection and SV-SS saturation steps are illustrated in Figure 6. The levels O<sub>7</sub> and C<sub>2</sub> are not equivalent, hence the SS selection is carried out from the first atom BSs. All SV levels, higher than the O<sub>7</sub> level are mutually equivalent, and the SS selection procedure starts from the last saturated SS, e.g. from atom 2 instead of atom 1 of the C<sub>4</sub> SS selection range in Figure 6.

The generation of four isomers is depicted in Figure 7. Since all carbon atoms are equivalent, oxygen atom 7 SV selects and saturates only the first SS from the first carbon atom according to the rule iii. Two extensions result from the second level SV-SS saturation. More interesting is the SS selection for the SV of atom 4. While a SS is selected from atom 2 of the first structure, no SS is selected from the equivalent atom 3 according to rule iii. Only SSs from atom 4 and SSs from atoms 4 and 5, respectively, are selected for the SV-SS saturation at the next (atom 5 and atom 6) SV levels, following rule ii. Two symmetric structures c and d are formed at the highest level. However, if this was not the highest level, e.g. the gross formula had 7 carbon atoms, then no extensions

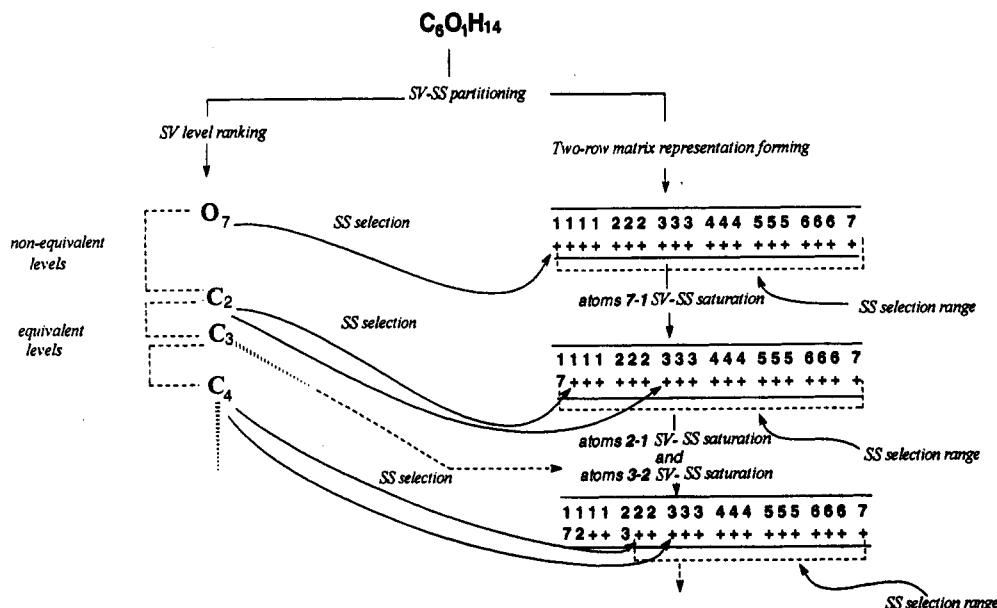


Figure 6. SS-selection and SV-SS saturation steps in the multiple-level depth-first structure generation procedure.

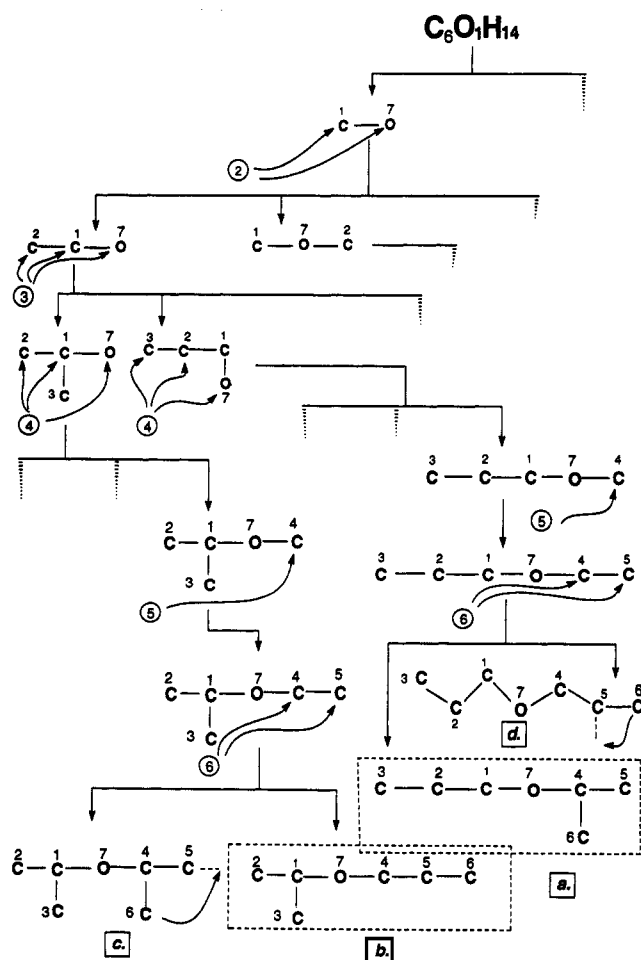


Figure 7. Generation of two isomorphic structures (**a** and **b**) as a consequence of the topological symmetry of the structures **c** and **d**. From these structures would be produced because atoms 4–6 are equivalent to the previous atoms 1–3 (rule V). Thus, Foulon<sup>17</sup> has introduced the notion of *ideal graph*: a graph is *ideal* if the edges which are created in any extension are not equivalent to the initial edges. According to this definition the predecessors of structures **c** and **d** are *ideal graphs* (the new edges 4–6 for **c** and 5–6 for **d** are equivalent to the initial edges 1–3 and 2–3, respectively). Hence, no extensions from them

are generated at the lowest levels of his generation method. Note that the generation method of Foulon is different from ours, but the perception of the topological symmetry leads to the same results.

**Step 4:** After each new extension the local indices,  $r_i$  are recalculated taking into account the new connectivity, and a new equivalence class partitioning is carried out. This leads to a dynamic partitioning of the vertices into equivalence classes, e.g., if an atom in a given level is in an equivalence class, it will be in a different class on the next level. Here, also a topological symmetry check is carried out, and the following rule is established. (vi) *If an SS from the SS-list is of an atom being in the same equivalence class as a previously saturated atom, then this SS is excluded from the SS-list and further no extension results in it.* This situation is illustrated in Figure 7. The two structures **a** and **b** in dashed-line boxes are isomorphic (see also Figure 2). This isomorphism is due to the topological symmetry of structures **c** and **d** generated by the SV-SS saturation of both atoms 6–5 and 6–4, respectively. Since the structure **d** appears after structure **a**, the topological symmetry perception will not contribute here very much. However, in the case of generation of structure **b** after structure **c** atom 5 is equivalent to atom 2; hence to avoid duplications, any SS of atom 5 must be excluded from the SS-list and no further extension carried out.

The procedure presented here allows all potential duplications to be *foreseen* by examining the topological symmetry of the generated extensions and subsequently avoided.

#### ACKNOWLEDGMENT

This work has been supported by the National Scientific Fund at the Ministry of Culture Science and Education-Sofia, Bulgaria (Contract X96). This support is gratefully acknowledged.

#### REFERENCES AND NOTES

- (1) Part 6 in the series: Bangoy, I. P.; Simova, S.; Cabrol-Bass, D.; Laude, I. Computer-Assisted Structure Generation from a Gross Formula. 6. Reducing the Structural Redundancy by Employment of 2D NMR Spectral Information. *J. Chem. Inf. Comput. Sci.*, in press.

- (2) Bangov, I. P. Computer-Assisted Generation of Molecular Structures from a Gross Formula. I. Acyclic Saturated Compounds. *Commun. Math. Chem. (MATCH)* **1983**, *14*, 235–246.
- (3) Bangov, I. P.; Kanev, K. D. Computer-Assisted Structure Generation from a Gross Formula: 2. Multiple Bond Unsaturated and Cyclic Compounds. Employment of the Fragments. *J. Math. Chem.*, **1988**, *2*, 31–48.
- (4) Bangov, I. P. Computer-Assisted Structure generation from a Gross Formula. 3. Alleviation of the Combinatorial Problem, *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 277–289.
- (5) Bangov, I. P. Toward the Solution of the Isomorphism Problem in Generation of Chemical Graphs: Generation of Benzenoid Structures. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 167–173.
- (6) Bangov, I. P. Computer-Assisted Structure Generation from a Gross Formula. 4. Fighting against Isomorphism Decease. *Commun. Math. Chem. (MATCH)* **1992**, *27*, 3–30.
- (7) Sussenguth, E. H. A Graph-Theoretical Algorithm for Matching Chemical Structures. *J. Chem. Doc.*, **1969**, *5*, 36–43.
- (8) Read, R. C.; Corneil, D. G. Graph-Isomorphism Disease *J. Graph Theory* **1977**, *1*, 339–363.
- (9) Polansky, O. E. Elements of Graph Theory for Chemists. In *Mathematical Chemistry*; Bonchev, D., Rouvray, D. H., Eds.; Gordon and Breach: Chichester, U.K., 1991; Vol. 1, p 43.
- (10) Faradjiev, I. A. Constructive Computation of Combinatorial Objects. In *Algorithmic Investigations in Combinatorics*; Faradjiev, I. A., Ed.; Nauka: Moscow, 1978; p 11 (in Russian).
- (11) Razinger, M.; Balasubramanian, K.; Munk, M. E. Graph Automorphism Perception Algorithms in Computer-Enhanced Structure Elucidation. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 197–201.
- (12) Klin, M. H.; Zefirov, N. S. Group Theoretical Approach to the Investigation of Reaction Graphs for Highly Degenerate Rearrangements of Chemical Compounds. II. Fundamental Concepts. *Commun. Math. Chem. (MATCH)* **1993**, *26*, 171–190.
- (13) Randic, M. Recognition of Identical Graphs Representing Molecular Topology. *J. Chem. Phys.* **1974**, *60*, 3920–3928.
- (14) Kudo, Y.; Sasaki, S. Principle for Exhaustive Enumeration of Unique Structures Consistent with Structural Information. *J. Chem. Inf. Comput. Sci.* **1976**, *16*, 43–49.
- (15) Trinajstić, N.; Nikolic, S.; Knop, J. V.; Muller, W. R.; Szymanski, K. In *Computational Chemical Graph Theory*; Mellor, J. M., Ed.; Ellis Horwood: New York, London, 1991; Chapter 3, p 79.
- (16) Contreras, M. L.; Valdivia, R.; Rozas, R. Exhaustive Generation of Organic Isomers. 1. Acyclic Structures. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 223–230.
- (17) Demirev, P. A.; Dyulgerov, A.; Bangov, I. P. CTI: A Novel Charge-Related Topological Index with Low Degeneracy. *J. Math. Chem.* **1991**, *8*, 367–382.
- (18) Foulon, J.-L. On Using Graph-Equivalent Classes for the Structure Elucidation of Large Molecules. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 338–348.
- (19) Filip, P. A.; Balaban, T. S.; Balaban, A. T. A New Approach for Devising Local Graph Invariants: Derived Topological Indices with Low Degeneracy and Good Correlation Ability. *J. Math. Chem.* **1987**, *1*, 61–83.
- (20) Bangov, I. P. Use of the  $^{13}\text{C}$  NMR Chemical Shift/Charge Density Linear Relationship for Recognition and Ranking of Chemical Structures. *Anal. Chim. Acta* **1988**, *209*, 29–43.
- (21) Gasteiger, J.; Marsili, M. Iterative Partial Equalisation of Orbital Electronegativities-A Rapid Access to Atomic Charges. *Tetrahedron* **1980**, *36*, 3219–3228.