Toward the Solution of the Isomorphism Problem in Generation of Chemical Graphs: Generation of Benzenoid Hydrocarbons[†]

IVAN P. BANGOV

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

A novel approach to the treatment of the isomorphism problem in generation of chemical graphs (structures) is discussed. It is based on the presumption that the isomorphic structures are the result of permutations of equivalent substructure units: atoms, fragments, etc. To avoid such permutations, a hierarchical generation scheme has been developed. One of the aims of this paper is to show how this approach can be reduced to the particular case of generation of benzenoid compounds.

INTRODUCTION

Two graphs Γ and Γ^* are considered isomorphic if the following relation holds:[2]

$$A^* = gAg^{-1} \qquad (1)$$

Here A is the adjacency matrix and g a similarity operation leading to simultaneous permutations between two rows and two columns of Γ.

For example, from the two adjacency matrices of the isomorphic graphs a and b in Figure 1, the matrix **b** is obtained from matrix **1** by permuting both rows 1 and 2 and columns 1 and 2.

However, keeping track of all such permutations during the generation process is practically an impossible task. Therefore, a different approach has been exploited in the literature.[2-5] One out of all the structures belonging to an isomorphism class is selected as canonical on the basis of a given canonicity predicate. Further, at each step of the structure growth the structure examines itself against this predicate by checking all generated permutations and only those producing canonical continuations are allowed.

Kudo and Sasaki[3,4] introduced the connectivity stack with the following relationship between the adjacency matrix elements a_{ij} and their position in the stack k:

$$k = i + (j-1)(j-2)/2 \qquad (i < j)$$

Faradjiev[2] has suggested an n^2-dimensional vector (binary code) $(a_{11}, a_{12}, ..., a_{nn})$, where n is the dimension of the matrix A and the entries a_{ij} are its elements taken columnwise. Hence, the structure having either the greatest connectivity stack or the greatest characteristic vector is considered canonical.

A slightly different procedure was suggested by Randić.[5] The rows of a matrix are considered as single names n_1, n_2, ..., n_n. Each matrix is then characterized by a sequence of names, which form its associated name N. Now the problem is finding out the matrix whose associated name represents the smallest binary number among all permutations of columns and rows.

Accordingly, structure a in Figure 1 taken from the set of two structures should be considered canonical. From their characteristic vectors

|011110 100001 100000 100000 100000 010000|

a

|010001 001110 010000 010000 010000 100000|

b

is obvious that vector **a** is lexicographically greater than vector **b**.

These approaches have two substantial drawbacks:

(1) At each step of the generation process all possible permutations of equivalent vertices (in ref 4 they are called segments) must be generated, and the permutation producing the greatest characteristic vector or stack is selected. This leads to an enormous number of permutations for structures having more than 10 atoms, i.e., for most of the real-world problems.

(2) There is no efficient method for handling structural fragments (substructures) within this approach. The fragments are usually perceived by using substructure search procedures after each structure generation.[6] This burdens the generation process with additional computational work.

In a series of papers[1,7-9] we have introduced a new approach whose purpose is to circumvent the drawbacks just mentioned above.

In this paper we summarize the general principals of our approach. Further, the generation of benzenoid structures is discussed. This new development illustrates the flexibility of the approach which allows a reduction of the general scheme to some particular classes of structures.

FUNDAMENTALS OF THE METHOD

The isomorphism problem is usually related to the problem of comparing two graphs in order to establish whether they are identical. It emerges from the necessity that the graph vertices be labeled by numbering them. For a structure of n atoms, we have $n!$ numberings, i.e., we have $n!$ distinct adjacency matrices. The isomorphic graphs are transformed one into another via nonautomorphic permutations of the vertex numbers; the corresponding adjacency matrices must obey relation 1.

However, in the case of structure generation any comparing of the current structure with a list of compiled structures previously generated is impractical, i.e., the generation of isomorphic graphs must be avoided rather than being first generated and then perceived. Additionally, there is another substantial difference: whereas the connectivity within two compared structures in the former case is fixed and only the vertex numbers are permuted, the atoms and groups forming vertices are permuted in the latter, their numbering remaining fixed. All that indicates that in the case of structure generation we have a different paradigm concerning the isomorphism problem.

For example, if one considers the structures from Figure 1 constructed of equivalent carbon atoms labeled, say, as C, all permutations (such as the permutation between atoms 1 and 2) form a symmetry group of automorphic operations while with regard to their numbering they appear nonautomorphic. In both cases they result in isomorphic structures obeying relation 1. Here one can see the origin of the isomorphism problem. In order to represent connectivity algebraically we are forced to assign distinct labels (numbers) to
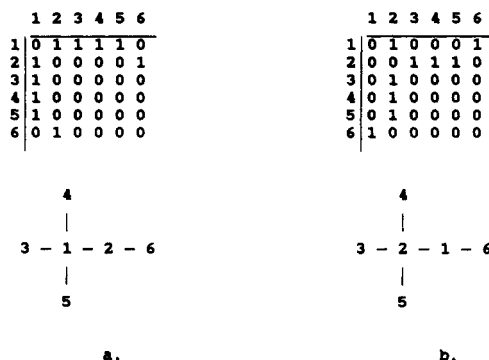
```
    1 2 3 4 5 6                      1 2 3 4 5 6
1 | 0 1 1 1 1 0 |              1 | 0 1 0 0 0 1 |
2 | 1 0 0 0 0 1 |              2 | 0 0 1 1 1 0 |
3 | 1 0 0 0 0 0 |              3 | 0 1 0 0 0 0 |
4 | 1 0 0 0 0 0 |              4 | 0 1 0 0 0 0 |
5 | 1 0 0 0 0 0 |              5 | 0 1 0 0 0 0 |
6 | 0 1 0 0 0 0 |              6 | 1 0 0 0 0 0 |


         4                            4
         |                            |
    3 - 1 - 2 - 6                 3 - 2 - 1 - 6
         |                            |
         5                            5

         a.                           b.
```

Figure 1. Two isomorphic graphs and their adjacency matrices.

```
              5                                          5
              C           permutation                    C
     1    2  |3   4        ( 1  4 )          4    2  |3   1
     C - C - C - C        ----------->       C - C - C - C
              |                                          |
              C                                          C
              6                                          6

                              a.

              5                                          5
              C           permutation                    C
     1    2  |3   4        ( 1  4 )          4    2  |3   1
     N - C - C - C        ----------->       C - C - C - N
              |                                          |
              C                                          C
              6                                          6

                              b.

             5   breaking              5   breaking                5   6
             C     bond                C     bond                  C - C
    1    2  |3   4  (3 6)     1    2  |3   4  6  (4 6)     1    2  |3   4
    C - C - C - C  ------->   C - C - C - C - C  ------->  C - C - C - C
             |       and               |        and
             C    creating             C     creating
             6     bond                      bond
                   (4 6)                      (5 6)

                              c.
```
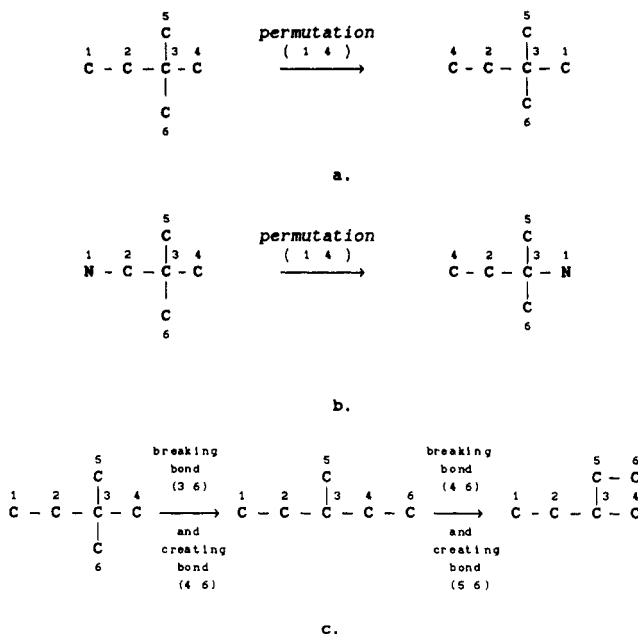
Figure 2. Generation of two isomorphic (case a and third structure of c) and two anisomorphic (case b and second structure of c) structures.

all atoms. However, some of them are equivalent by their chemical nature. Hence their permutations result in isomorphic (duplicate) structures.

In Figure 2 two different ways of generation of isomorphic (case a) and the third structure from panel c and anisomorphic structures (case b) and the first structure from panel c are provided. Case a is topologically analogous to case b as both structures result from permutations of atoms 1 and 4 by conforming to relation 1. However, whereas the generated structure in the former case is isomorphic it is anisomorphic in the latter due to the different equivalence of atoms 1 and 4 (N and C, respectively). In case c the second structure is not isomorphic to the first because it cannot be represented as an automorphic permutation of any two of the carbon atoms, i.e., it does not conform to relation 1. In contrast, the third structure from Figure 2c is isomorphic to the second as it can be regarded as a result of an automorphic permutation between atoms 4 and 5.

Accordingly, we consider isomorphism as a function of automorphism, i.e., isomorphic structures are the result of permutations of equivalent vertices (atoms). Therefore our approach consists of a development of such a generation scheme which inhibits the permutations between equivalent vertices, i.e., all transformations between isomorphic structures such as those provided in Figure 1, and cases a and the third structure of panel c from Figure 2.

Additionally, we tried to extend these considerations to the case of employment of fragments (substructures) so they can
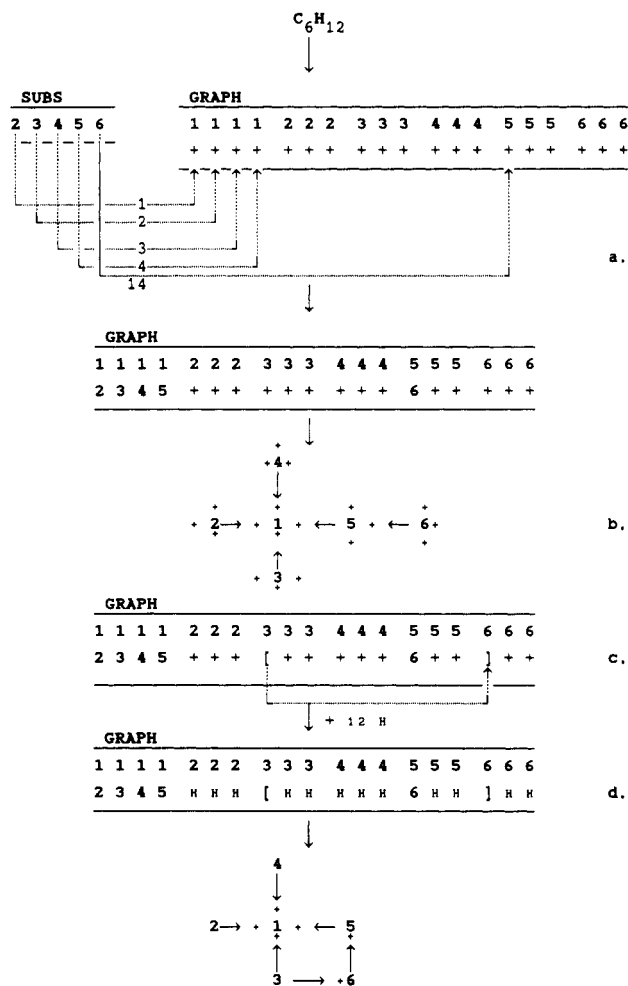

Figure 3. Matrix representation of the generation process.

be treated in the same manner as the single atoms. It is evident that any knowledge about the connectivity within a fragment is a substantial constraint on the generation process.

Structure generation is a combinatorial process of excising old and creating new bonds between bonding sites (BSs) of either single atoms, or atoms of chemical groups and structural fragments. Here, BS is a synonym of free valence taken in its classical meaning. Thus, the number of BSs of a single carbon atom are 4.

In contrast to other methods, our approach employs directed rather than nondirected graphs. A directed graph can be drawn as shown in Figure 3b. Each edge is directed (drawn as an arrow) and formed of two types of BSs: (+) type (denoted with + in Figure 3) and (←) type (denoted as ←). In previous papers[1,9] we called these Saturation Sites (SSs) and Saturating Valences (SVs). Hence, the BSs of a given structure form two sets (V and S) of BSs: SV $\in$ V and SS $\in$ S.

One can see from Figure 3b that each atom except the first provides one SV and $n - 1$ SSs. Here n is the atom valence (the full number of the atoms BSs). All BSs of the first atom are of SS type. Our mathematical representation of the two sets V and S is provided in Figure 3a. The set V is represented by the one-dimensional array SUBS and the set S by the two dimensional array GRAPH. The first row GRAPH elements are the corresponding SSs (given with their atom identifiers and numbers) and the second row elements are denoted by +. Any filling of a second row GRAPH element with a SV from the SUBS matrix (saturation of a SS with a SV) leads to the formation of a new edge (bond).

All $^mP_L = L!/(L - m)!$ permutations of m selected from L SSs without repetition saturated with the m SVs produce a

Table I. Valences, Atom Codes and Initial Values R_o for Different Types and Hybridization States

type and hybridization state	valence	code	R_o
C			
sp³	4	C	4
sp	2	#C	7
sp² (olefic)	3	=C	11
sp² (aromatic)	3		13
N			
sp³	3	N	15
sp²	2	=N	18
sp	1	#N	20
O			
sp³	2	O	23
sp²	1	=O	25
F	1	F	32
S	2	S	28
Cl	1	Cl	33
Br	1	Br	34
I	1	I	35

Figure 4. Generation of permutations between nonequivalent atoms and combinations (between equivalent) atoms.

full set of structures. The generation of such a structure is given in Figure 3a (permutation 1 2 3 4 14).

A cyclic structure having r cycles is generated from the corresponding acyclic structure (spanning tree) by linking r pairs of SSs at a time. The first SS (having always the smaller number) is transformed into SV denoted as] and saturates the second SS. The sign [in the second GRAPH row indicates that the corresponding BS is not currently an SS. The bond formed in such a way we call a *closure bond*. The number r is the corresponding cyclomatic number minus the number of double and minus twice the number of triple bonds. The generation of a closure bond is illustrated in Figure 3c.

Finally all the nonsaturated SSs are saturated with H atoms as shown in Figure 3d.

Fragments are considered "*superatoms*" having n BSs. Hence they are processed in the same way as single atoms. Each fragment except the first provides one SV and $n - 1$ SSs. The only difference is that the BSs are no longer equivalent. Thus, each one BS (nonequivalent of any of the previous BSs) is selected in turn as SV, the other $n - 1$ being selected as SSs.

However, it appears that most of the structures generated through this scheme are isomorphic and some disjointed. In order to treat this redundancy, the following hierarchical generation scheme was adopted:

(1) All the vertices (atoms) are partitioned into automorphism classes: $v_1 \in V_1, v_2 \in V_2, ..., v_n \in V_n$ by assigning to each atom a Local Charge-related Index (LCI):[1,7-10]

$$LCI + R_o - N_H + q \qquad (2)$$

Here R_o is a constant for each atom in a given hybridization state (the corresponding values are presented in Table I), N_H is the number of the attached hydrogen atoms, and q is the charge density of the atom calculated through a fast empirical method. We use the method of Gasteiger et al.[11,12] In the case of fragments, the free valences (BSs) are considered univalent atoms having arbitrarily chosen constants $A = 5.0$, $B = 2.0$, and $C = 0.5$ within the Gasteiger charge density computational scheme.

(2) The SVs are ranked according to the LCIs into a hierarchical system of levels: the greater the LCI of a SV, the lower the level it occupies. The SVs of the same equivalence class are ranked according to their atom numbering: the lower number the related atom has the lower level its SV occupies and vice versa. This procedure we called[9] Hierarchical Selection of Equivalent SVs (HSESV).

(3) A selection of SSs at each current level for saturation with a SV at the higher level is carried out conforming to the following rules:

(i) If the current and the higher levels are of different equivalence classes, then the SS selection starts from the
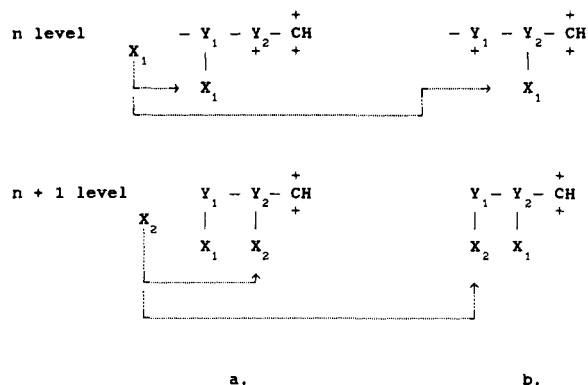
first SS of the first atom and vice versa; if the two levels are of the same equivalence classes, the selection starts from the SS next to the SS just saturated.

(ii) No SS from the atom which the higher level SV belongs is selected (no atom is bonded to itself).

(iii) Only one out of all SSs belonging to a given equivalent class is selected to be saturated with the higher level SV.

The SS selection procedure was named[9] Hierarchical Selection of the SSs (HSSS).

The generation process is a hierarchical depth-first procedure of recursive execution of the steps:

Step 1: LCI determination, carried out at each level.

Step 2: SV partitioning onto different levels, carried out only if the current and the higher levels are of different equivalence classes.

Step 3: SS selection, carried out at each current level for the higher.

Step 4: Saturation of a SS with a SV.

In the case of cyclic structures after a spanning tree is built the following steps are carried out:

Step 5: Selection of the first closure bond atom SS and transforming it into a closure-bond SV (CBSVs). Thus, additional cycle-forming levels are formed. The CBSVs are considered equivalent and their selection obeys the following rules:

(iv) Each higher level CBSV selection starts from a SS next to the one that has been transformed into CBSV at the current level.

(v) Only one SV out of a set of equivalent atoms is selected.

Step 6: Selection of a partner SS (the second atom of the closure-bond SS, (BSS) which obeys the following rules:

(vi) Each higher level CBSS selection starts from a SS next to the one that has formed a closure bond at the current level.

(vii) Only one out of all SSs emanating from a set of equivalent and equidistant atoms (being at the same distance from the CBSV atom) is selected.

If the current level is the highest a saturation of the remaining SSs with hydrogen atoms is carried out, the coded complete structure is recorded in a file. Otherwise, backtracking to the lower level and saturating the next SSs at that level is carried out.

Practically, the rules i and vii determine our approach to the isomorphism problem. Consider the simple hierarchical scheme presented in Figure 4. Each of the atoms Y_1, Y_2, and group CH provide SSs to be selected for the nth level atom X_1 SV saturation. If the selection of the SSs for the higher $n + 1$ level starts from the first atom SS, then a permutation of X_1 and X_2 occurs. In the case of equivalent atoms X_1 and
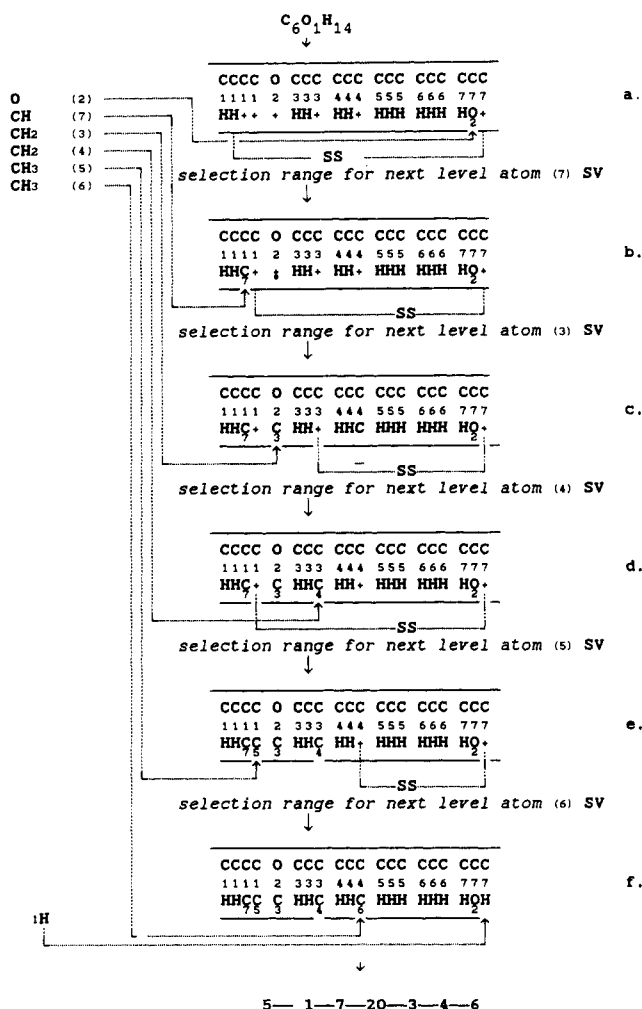
$C_6O_1H_{14}$

```
                    cccc o ccc ccc ccc ccc ccc
 O     (2) ------   1111 2 333 444 555 666 777      a.
 CH    (7) ------   HH++ + HH+ HH+ HHH HHH HO+
 CH2   (3) ------
 CH2   (4) ------          L———— SS ————
 CH3   (5) ------   selection range for next level atom (7) SV
 CH3   (6) ------
                    cccc o ccc ccc ccc ccc ccc
                    1111 2 333 444 555 666 777      b.
                    HHC+ ; HH+ HH+ HHH HHH HO+
                              L————SS————
                    selection range for next level atom (3) SV

                    cccc o ccc ccc ccc ccc ccc
                    1111 2 333 444 555 666 777      c.
                    HHC+ C HH+ HHC HHH HHH HO+
                              L— ——SS—
                    selection range for next level atom (4) SV

                    cccc o ccc ccc ccc ccc ccc
                    1111 2 333 444 555 666 777      d.
                    HHC+ C HHC HH+ HHH HHH HO+
                              L————SS————
                    selection range for next level atom (5) SV

                    cccc o ccc ccc ccc ccc ccc
                    1111 2 333 444 555 666 777      e.
                    HHCC C HHC HH+ HHH HHH HO+
                              L——SS——
                    selection range for next level atom (6) SV

                    cccc o ccc ccc ccc ccc ccc
                    1111 2 333 444 555 666 777      f.
 1H                 HHCC C HHC HHC HHH HHH HOH
                    L————————————————————

                    5— 1—7—20—3—4—6
```

Figure 5. Matrix representation of hierarchical structure generation.

X_2, an isomorphic substructure is formed. To avoid its generation the SS selection in step b for the $n + 1$ level must start from the CH group SS next to the Y_2 atom saturated at the nth level. In contrast, in the case of nonequivalent atoms X_1 and X_2 their permutation is allowed, and the SS selection can start from the Y_1 atom rule i satisfied).

In the case where the two atoms Y_1 and Y_2 are equivalent, only one SS (from atom Y_1) must be selected at the nth level (rule iii) and step b is skipped. Otherwise an isomorphic structure will appear at the $n + 1$ level, nevertheless the equivalence of atoms X_1 and X_2 as this corresponds to a permutation of the two equivalent atoms Y_1 and Y_2.

The hierarchical process of isomer generation (gross formula $C_6O_1H_{14}$) is illustrated in Figure 5. One can see that the SS selection for the second equivalent CH_2 SV (taken from atom 4) at the third level (c) starts from the second free SS instead from the beginning. Furthermore, if both the SV and SS atoms are of the same equivalence class the SS–SV bonding is carried out in such a way that only SSs of atoms having a number lower than the partner SV atom are selected, e.g., a bond 3–4 but never 4–3 may be formed in Figure 5d. Consequently, the equivalent atoms, e.g., the two CH_2 carbon atoms in our case, are arranged according to their numbering from lower to higher numbers. Thus an adjacency matrix having greater characteristic vector or connectivity stack is formed corresponding to a canonical structure.

However, it should be admitted that we were not able to eliminate completely the isomorphic structures although their number is sharply reduced. Only a few isomorphic structures due to mirror symmetry are generated. The reason for this redundancy is still not fully understood. Their discrimination

is carried out by using the total Charge-related Topological Index (CTI):[10]

$$\text{CTI} = \sum\sum(x_i x_j / \mathbf{D}_{ij}) \qquad (3)$$

where $x_i = LCI_i$, $x_j = LCI_j$, and $\mathbf{D}_{ij}$ is the distance matrix element between atoms i and j. This index is calculated at each level, and the isomorphic partial structures are pruned right away after their emerging. Thus, whole branches of the generation tree are cut.

Although it is impossible to provide a rigorous mathematical proof on the reliability of this index, our experience[1,9,10] shows that no false isomorphism prediction is observed so far.

GENERATION OF BENZENOID HYDROCARBONS

One of the aims of this paper is to show how this approach can be reduced to particular problems such as the generation of benzenoid hydrocarbons (polyhexes). These chemicals have interesting features from theoretical, experimental, biological, and pharmacologic points of view. Many among them are cancer-causing, and this property is related to their structure.

Benzenoid compounds are configurations containing a given number of connected hexagons. Because of their π-electron structure, they have graphite-type connectivity, i.e., they are planar. These two constraints are very important for a further expansion of our approach to this particular case.

Apparently, we may employ the general scheme discussed above without imposing any constraints on the generation process. However, this will lead to the generation of a great variety of nonplanar structures which must be subsequently recognized and pruned. On the one hand, this redundancy is a very severe problem; on the other, the perception of non-benzenoid hydrocarbons requires the development of complicated and additional computer time-consuming procedures.

Thus, our generation scheme covers all and only the planar hexagonal graphs, which implies that the generation of helicenes is not considered here. As one will see below this approach does not have the restriction that only simply connected graphs will be considered. Hence, structures such as circulenes may also be produced.

As an initial and basic information for the generation of polyhexes we can use either the number of hexagons or a given molecular formula. Trinajstić et al.[13,14] have reported a generation scheme based on the number of hexagons. It consists of the generation of a boundary code, a sequence of digits representing the outer edges of the benzenoid structure. This representation is not unique, but by cyclic permutations all numerical representations of the polyhex are further generated. The sequence giving the lexicographical maximum is considered *normed representation*. This representation corresponds to the canonical structure approaches of Randić, Kudo and Sasaki, and Faradjiev outlined above. Here again, as in these methods, the generation of all permutations for each polyhex *normed representation* is necessary.

In a series of papers Dias[15-19] has developed a periodic (Mendeleevian-type) table for fused nonradical benzenoid hydrocarbons. The basis of the Dias method is the molecular formula. The method consists of two parts: a periodic table and rules for the enumeration of the benzenoids according to their location in the table. The possible isomers can be recursively enumerated by adding a C=CH-CH=C fragment to the immediately preceding row series homologue and a C=CH-CH=CH-CH=C fragment to the immediately preceding column homologues. However, the implementation of this approach into a efficient computer-generation program requires that some combinatorial and related isomorphism problems also be considered.

The isomorphism problem within those structures is extremely severe since they show an additional source of symmetry (both topological and geometrical equivalence).

GENERATION OF BENZENOID HYDROCARBONS

J. Chem. Inf. Comput. Sci., Vol. 32, No. 2, 1992 **171**

Equivalent aromatic carbon atoms form new types of equivalent configurations—the hexagons. Hence the general rule given above must be modified as follows:

To avoid isomorphism any permutations between both equivalent atoms and equivalent hexagons must be avoided.

The approach presented here starts from a given number of aromatic carbon atoms, i.e., it is closed to the method of Dias. As long as our generation scheme is directed to the solution of structure elucidation problems it was found that the number of aromatic carbon atoms can be more easily derived from the spectral information (e.g., either from the number of ^{13}C NMR signals or from the intensities of the ^{1}H NMR signals in the aromatic region of the corresponding NMR spectra) than from the number of fused benzene rings.

The method is based on the following rules:

(i) The aromatic carbon atoms forming benzenoid structure are a special type of atoms coded as :C. They are considered to have a valence of $n = 3$. The general rules for formation of directed graph structures described above are also valid here: each :C atom except the first provides 1 SV and $n - 1 = 2$ SSs. The first atom provides $n = 3$ SSs. As the edges of directed graphs are directed they can be considered vectors.

(ii) All the SVs are equivalent, and they form a hierarchical set of levels following the rules given in the preceding section: the greater number the SV atom has the higher the level occupying its SV.

(iii) According to our general scheme the generation process is a hierarchical recursive (depth-first) procedure consisting of the steps: selection of SSs for the higher levels and subsequent saturation of a SS with a SV at this level. The latter step, practically, is formation of a bond by attaching a new atom to the structure under construction.

(iv) After each SS-SV saturation, the program checks whether a hexagon is formed (a ring–closure bond is generated) or the construction will proceed with the extension of the current chain. We speak of *ring-closure level* in the former and about *chain-extension level* in the latter case.

(v) After the SS-SV saturation at the highest level (exhausting the last :C atom) a control for the integrity of the generated structure is carried out. If this level ends up with a closure of the last hexagon, i.e., it is a *ring-closure level* then the structure is considered correct and vice versa if the last hexagon cannot be closed then the program backtracks to the lower levels for the next SS hexagon formation.

Further, a more detailed discussion of the method is necessary:

Inasmuch as a directed bond is a vector it can be represented by an ordered pair of atoms (i, j). Each new bond is initiated from the second atom j of a previously generated bond (hereafter we shall call the latter *adjacent bond*). If there is no second atom SS, e.g., at some bonds of bay regions, then a SS from the first bond atom i may be taken. Here, obviously, the direction of the chain growth will be the opposite—from the first to the second atoms.

A new hexagon is fused to a previously generated one by forming a common bond which we call the *bond adjacent to a hexagon*. Hence the chain is built up in such a direction that the hexagon ring is closed around the *hexagon adjacent bond* (i, j). Generally speaking the SS selection is related to the selection of *adjacent bonds*.

Let a bond be formed by saturating a SS with a SV at the current level, i.e., an addition of a new :C atom (c).

The distance matrix and the coordinates of the atom c are calculated, and the program checks whether it forms a planar hexagon with any of the atoms p previously generated by applying the following relations:

$$D_{jc}^{T} + D_{ip}^{T} + 2 = 6 \tag{4}$$

$$D_{cp}^{G} = [(X_c - X_p)^2 + (Y_c - Y_p)^2]^{1/2} = 1.4 \tag{5}$$

Here D_{jc}^{T} is the topological distance between the second atom j of the bond *adjacent to the hexagon* in construction and the current atom c, i.e., the (j, c) element of the distance matrix, and D_{ip}^{T} is the corresponding distance between the first (*current hexagon*) *adjacent bond* atom i and any of the previous atoms p candidates to form closure bonds. Thus, relation 4 indicates that a hexagon might be formed around the selected *hexagon adjacent bond* after the addition of the closure bond (c, p). D_{pc}^{G} from relation 5 is the geometrical distance between these atoms which is 1.4 Å for a benzenoid bond, i.e., the formed hexagon must be planar.

If the current atom c satisfies the relations 4 and 5 a closure bond (c, p) is formed. This is carried out by transforming an atom p SS into a CBSV denoted as] and a SS from atom c is saturated with this CBSV. The corresponding p SS is currently denoted as [, which means that it is excluded from the generation process at the higher levels.

After each SS-SV saturation at the current level, the step of SS selection for the higher level is carried out. This step depends on the result of the previous saturation. If the saturation does not lead to the formation of a hexagon i.e., none of the relations 4 or 5 are satisfied, only one SS is selected for the higher level which leads to a extension of the current chain. In order to preserve the canonicity of the structure, this SS is taken from the atom having the number next to the SS atom currently saturated. For example, the current saturation in Figure 6a is carried out by selecting a SS from atom 13 and saturating it with a SV from atom 12. This leads to a noncanonical structure. In contrast, the selection of a SS in Figure 6b produces a canonical continuation. In case the current level is a ring-closure level, the SS selection for the higher level consists of a selection of all positions from which the construction of a new hexagon-forming chain will be initiated. As stated above this is equivalent to the selection of bonds that this hexagon will be adjacent to. The bond selection conforms to the following rules:

(j) Only bonds having free SSs are selected.

(jj) The bond selection starts from the bond next to the adjacent bond of the hexagon previously generated.

(jjj) If two or more bonds are equivalent then only one of them is selected.

The efficiency of the rules (j–jjj) is illustrated with the three isomorphic structures (d,e,f) in Figure 6. All of them originate from the substructure c. As the hexagon (9–11–12–13–14–8) has been initiated from the second atom 9 of the bond 8–9, then according to jj the only possible adjacent bonds to the next hexagon chain formation might emanate from the bonds 9–10, 10–1, 8–11, 11–12., etc. but never from bonds having both atoms of lower numbering (e.g., such as 2–7, 2–3, 6–1, etc.). Accordingly, the SSs for the higher level will be selected from atoms 10, 11, 12, ..., etc., respectively. This rule ensures that no permutations between hexagons are carried out. As a consequence, the generation of the substructure d in Figure 6 must be avoided.

As the substructure c is highly symmetric, following rule jjj only the bonds 9–10, 9–11, 11–12, and 12–13 (SSs from
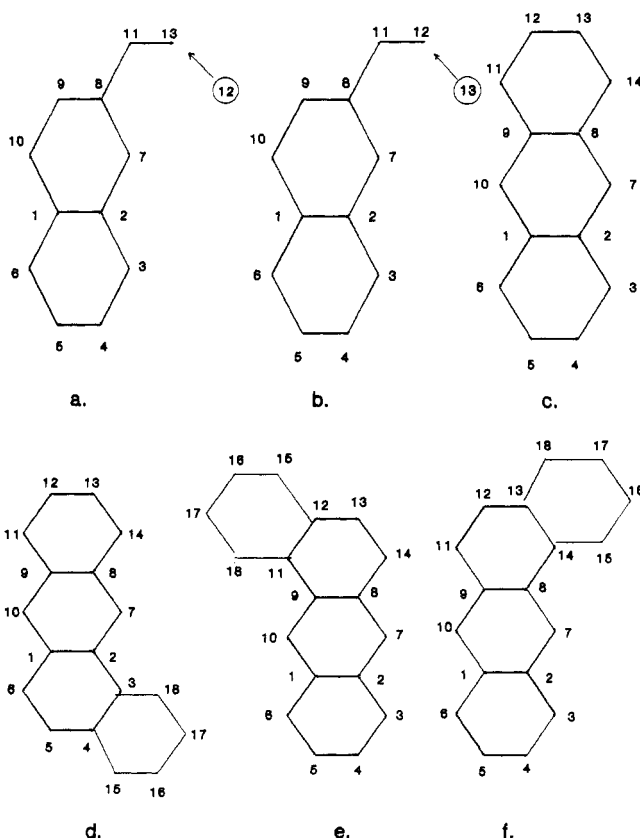
Figure 6. Canonical generation of benzenoid structures.

atoms 10, 11, 12, 13) may be selected as *adjacent* to the next hexagon. For example, only structure e in Figure 6 with last hexagon formed around bond 11–12 will be generated, other isomorphic structures such as f with the hexagon formed

Table II. Generated Benzenoid Structures and Duplications from a Given Gross Formula

gross formula	no. of generated benzenoid structures	no. of hexagons	no. of generated duplications
$C_{18}H_{12}$	5	4	1
$C_{21}H_{13}$	6	5	16
$C_{22}H_{14}$	12	5	6
$C_{25}H_{15}$	24	6	44
$C_{26}H_{14}$	36	6	14

around bond 13–14, which is equivalent to bond 11–12, are skipped. The matrix representations of the examples a–f from Figure 6 are provided in Figure 7.

The equivalence of the bonds obeying rule jjj is estimated by using a newly-devised Bond Charge-related Index (BCI) which for the bond (i, j) is defined as follow:

$$BCI = LCI_i LCI_j \qquad (6)$$

The bonds having the same BCIs are considered equivalent. The BCIs are dynamically evaluated at each *closure-ring* level.

PROGRAM IMPLEMENTATION AND TESTING

A flowchart of the program implementation of this generation scheme is presented in Figure 8. The program is an option of our structure-generation procedure StrGen, which is a part of a structure-elucidation system. It written in Turbo-Pascal 6 for the IBM PC-compatible microcomputers in the range from I8088 to I486.

Some results from this program are provided in Table II. They indicate that although the combinatorial explosion of duplications is already curbed by applying our canonicity rules, some isomorphic structures will appear. We found that they are more in the cases having hexagons around bay regions, e.g., examples 2 and 4 from Table II. This isomorphism is due to a higher-order symmetry which cannot be predicted
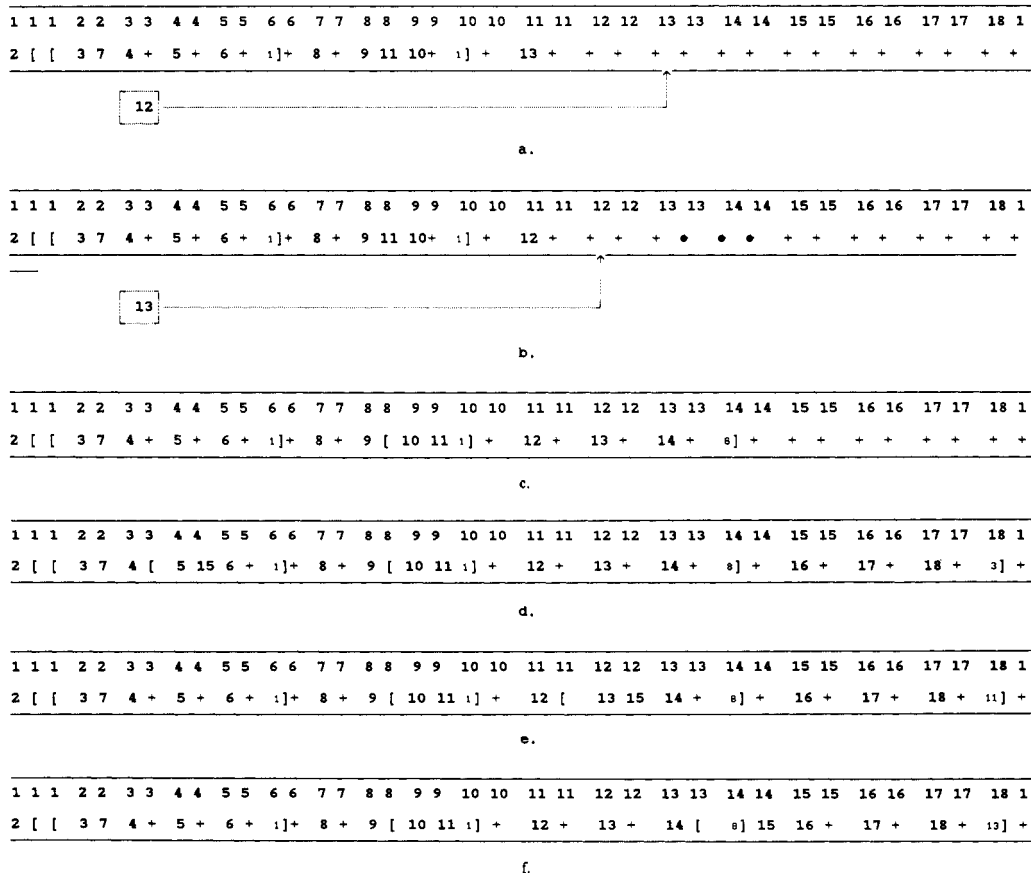

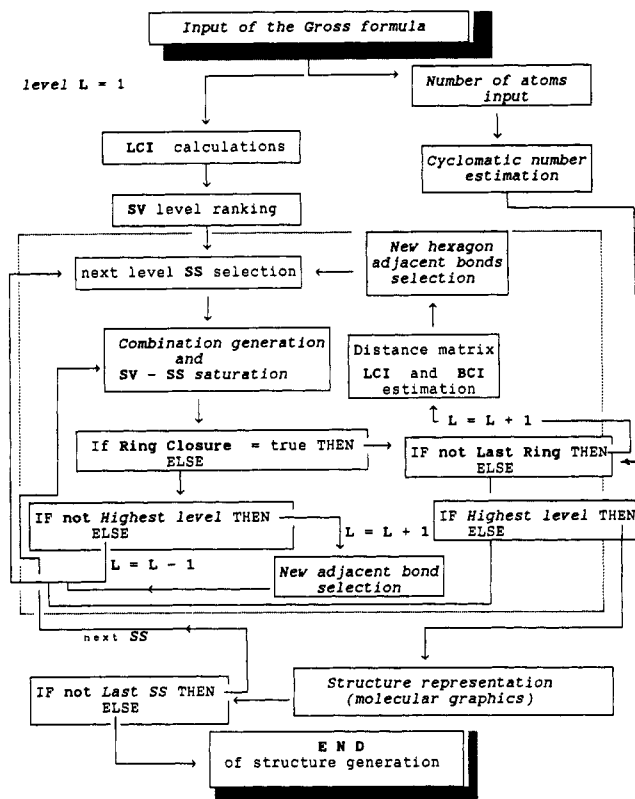**Figure 7.** Matrix representation of the generation process from Figure 6.

GENERATION OF BENZENOID HYDROCARBONS

J. Chem. Inf. Comput. Sci., Vol. 32, No. 2, 1992 **173**

Figure 8. Flowchart of the module for the benzenoid structure generation.
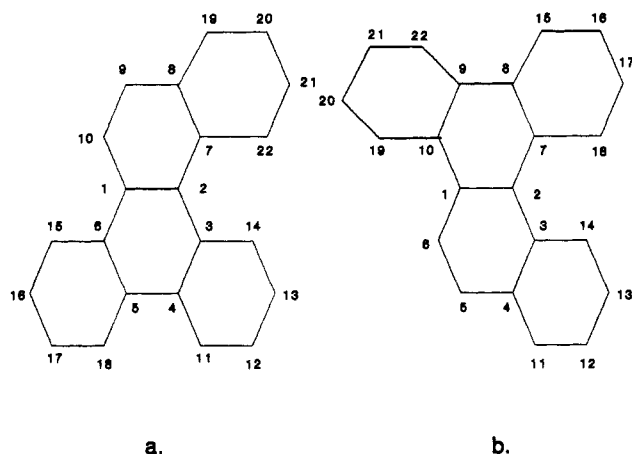

a.   **b.**

Figure 9. Two isomorphic benzenoid structures generated by this approach.

from the current level to next level. Thus, structure b in Figure 9 is generated by conforming to all of the rules given above, but it is isomorphic to structure a and as discussed the latter is eliminated by using the the CTI.

It is clear that this is an atom-by-atom building-up (aufbau) procedure for the generation of benzenoid structures. It may be considered "blind" in the sense that the program cannot determine at a lower level whether the chosen path leads to a complete structure or to a structure not corresponding to the input molecular formula. This produces a number of "error-and-trial" generations. This algorithm could be further modified with the incorporation of the Dias Periodic Table and rules. Then, the generation process will be *guided* and the program more efficient.

REFERENCES

(1) Bangov, I. P. *Commun. Math. Chem.* (*Match*) **1991**, in press.
(2) Faradjiev, I. A. *Algorithmic Investigations in Combinatorics*; Faradjiev, I. A., Ed.; Nauka: Moscow, 1978 (in Russian).
(3) Kudo, Y.; Sasaki, S. Connectivity Stack: A New Format for Representation of Organic Chemical Structures. *J. Chem. Inf. Comput. Sci.* **1974**, *14*, 200–202.
(4) Kudo, Y.; Sasaki, S. Principle for Exhaustive Enumeration of Unique Structures Consistent with Structural Information. *J. Chem. Inf. Comput. Sci.* **1976**, *16*, 43–49.
(5) Randić, M. Recognition of Identical Graphs Representing Molecular Topology. *J. Chem. Phys.* **1974**, *60*, 3920–3928.
(6) Funatsu, K.; Miyabayashi, N.; Sasaki, S. Further Developments of Structure Generation in the Automated Structure Elucidation System CHEMICS. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 18–28.
(7) Bangov, I. P. Computer-Assisted Generation of Molecular Formulas from a Gross Formula. *Commun. Math. Chem.* (*Match*) **1983**, *14*, 235–246.
(8) Bangov, I. P.; Kanev, K. D. Computer-Assisted Structure Generation from a Gross Formula. II. Multiple Bond Unsaturated and Cyclic Compounds. Employment of Fragments. *J. Math. Chem.* **1988**, *2*, 31–48.
(9) 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.
(10) Demirev, P. A.; Dyulgerov, S. A.; Bangov, I. P. *J. Math. Chem.* **1991**, in press.
(11) Gasteiger, J.; Marsili, M. Iterative Partial Equalization of Orbital Electronegativity: A Rapid Access to Atomic Charges. *Tetrahedron* **1980**, *36*, 3219–3222.
(12) Gasteiger, J.; Marsili, M. *Data Processing in Chemistry*; Hippe, Z., Ed.; Studies in Physical and Theoretical Chemistry Series No. 16.
(13) Knop, J. V.; Mueller, W. R.; Jericevič, Z.; Trinajstić, N. Computer Enumeration and Generation of Trees and Rooted Trees. *J. Chem. Inf. Comput. Sci.* **1981**, *21*, 91–99.
(14) Knop, J. V.; Mueller, W. R.; Jericevič, Z.; Trinajstić, N. *Computer Generation of Certain Classes of Molecules*; SKTH-Kemija u Industriji: Zagreb, 1985.
(15) Dias, J. R. A Periodic Table for Polycyclic Hydrocarbons. Isomer Enumeration of Fused Polycyclic Hydrocarbons. 1. *J. Chem. Inf. Comput. Sci.* **1982**, *22*, 15–22.
(16) Dias, J. R. Periodic Table for Polycyclic Hydrocarbons. 2. Polycyclic Aromatic Hydrocarbons Containing Tetragonal, Pentagonal, Heptagonal, and Octagonal Rings. *J. Chem. Inf. Comput. Sci.* **1982**, *22*, 139–152.
(17) Dias, J. R. A periodic table for polycyclic aromatic hydrocarbons. Part 3. Enumeration of all the polycyclic conjugated isomers of pyrene having ring sizes ranging from 3 to 9. *Commun. Math. Chem.* (*Match*) **1983**, *14*, 83–138.
(18) Dias, J. R. Current Status of Isomer Enumeration of Practical Benzenoids. *J. Mol. Struc.* (*Theochem*) **1991**, *76*, 155–190.