A novel nomenclature of polycyclic aromatic hydrocarbons without using graph centre

He Wenchen[1] and He Wenjie[2]

[1]Hebei Chemical Engineering Institute, Shijiazhuang, People's Republic of China
[2]Hebei Academy of Sciences, Shijiazhuang, People's Republic of China

In this paper, we develop a novel adjacency matrix, He-matrix, corresponding to the dualist graph. Without using the graph center concept, we advance a novel nomenclature of polycyclic aromatic hydrocarbons. Further, we derive some distinguishing theorems about PAH molecules and present some results of our automatic derivatization and automatic classification counting of fused PAH molecules.

Key words: PAH characteristic graph — He-matrix — Kekulé structure — Cut set

1. Introduction

The nomenclature of organic compounds is a fascinating field of research [1-5]. In the past years, A. T. Balaban, D. Bonchev, M. Randić and their collaborators have done a lot of very attractive work [6-12]. They developed a universal nomenclature on the basis of the recently developed generalized concept for a graph center (called the nomenclature - using graph center). Maybe, such a nomenclature is more suitable for acyclic structure, but for the case of cyclic structures, it is less suitable, because in the case where the dualist graph ([6]; for a sufficient explanation see below) contains a larger number of vertices, the graph center determination is rather cumbersome and it is not easy to do such a lot of work without the help of computers. In addition, for the automatic derivation and the automatic classification counting the method is more complicated, because after derivation, the graph center is often changed. The redetermination of the graph center requires a lot of work.
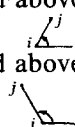
In this paper for the dualist graphs [6, 7, 13–19] of polycyclic aromatic hydrocarbons (PAH), we develop a novel adjacency matrix with some geometric factors (called He-matrix). Using it, we can set up a novel nomenclature of polycyclic aromatic hydrocarbons in which we do not need the concept of the graph center (for the general problem of nomenclature for PAH's without using a graph center, see [14]). An He-matrix not only contains the information about the topological relation in the dualist graph but also the geometric relation of angles between adjacent edges in the graph. Thus using the dualist graph and its corresponding He-matrix, we can discuss some properties of PAH molecules. Therefore, using them, we can carry out the automatic derivation, and the automatic classification counting. In this paper, we introduce some distinguishing theorems about PAH molecules which are derived from the dualist graph and He-matrix, and, further, we present some results of the automatic classification of fused PAH molecules.

In addition, the procedure of rotation and reflection transformation suggested in this paper represents a solution of the equivalence problem of dualist graphs of PAH and, further, an improvement of the research done by O. E. Polansky and D. H. Rouvray [19].

2. Definitions of the characteristic graph and He-matrix

We call the dualist graph of a PAH molecule its characteristic graph. For example, a molecule containing 7 benzene rings is shown in Fig. 1. Its characteristic graph $G(V, E)$ is given in Fig. 2, in which the ordinal numbers of vertices are given by sweeping from the left to the right and from top to bottom.

The so-called He-matrix of a characteristic graph $G(V, E)$ is an $N \times N$ matrix (a_{ij}), where $N = |V|$ (the total number of vertices), v_i $(i = 1, 2, \ldots, N)$ represents the ith vertex, and

$$
a_{ij} = \begin{cases}
0 \text{ (if } i = j, \text{ or } v_i \text{ nadj } v_j) \\
1 \text{ (if } v_i \text{ adj } v_j, \text{ and the angle between } \overline{ij} \text{ and the} \\
\quad \text{positive horizontal direction is } k\pi) \quad i \text{ ——} j \\
2 \text{ (if } v_i \text{ adj } v_j, \text{ and the angle stated above is } k\pi + \pi/3) \\
\qquad\qquad\qquad\qquad i \overset{j}{\diagup} \\
3 \text{ (if } v_i \text{ adj } v_j, \text{ and the angle stated above is } k\pi + 2\pi/3) \\
\qquad\qquad\qquad j \diagdown \; i
\end{cases}
$$

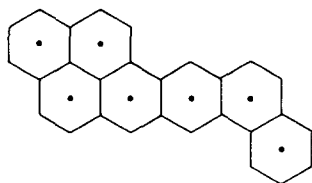(Note: the positive horizontal direction is from the left to the right.)
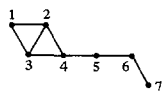

Fig. 1. Molecule graph

Fig. 2. Characteristic graph

Thus, the He-matrix of the PAH molecule in Fig. 2 is

$$(a_{ij}) = \begin{Bmatrix} 0 & 1 & 3 & 0 & 0 & 0 & 0 \\ 1 & 0 & 2 & 3 & 0 & 0 & 0 \\ 3 & 2 & 0 & 1 & 0 & 0 & 0 \\ 0 & 3 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 3 \\ 0 & 0 & 0 & 0 & 0 & 3 & 0 \end{Bmatrix}. \tag{1}$$

Obviously, an He-matrix is a real symmetric matrix. Its diagonal elements are always equal to zero. Apart from the information about the adjacent topological relation of the characteristic graph, an He-matrix also contains the information about the geometric relation of angles between adjacent edges in the graph. And so it contains more information that the common adjacency matrix of the characteristic graph.

3. The nomenclature of fused PAH molecules

For any fused PAH molecule, at first, put one of the edges in its characteristic graph into a horizontal position (as in Fig. 2). Secondly, from top to bottom layer by layer and in every layer from the left to the right one by one, give the natural ordinal numbers $1, 2, 3, \ldots, N$ of vertices. Then from top to bottom write down the total of vertices in every layer except the lowest one and insert a semicolon between every two adjacent layers. In the example shown in Fig. 2, it is $2;4$, called the first section of the name (or the nomenclature series) of the PAH molecule.

After that, turn the characteristic graph through an angle 60° counterclockwise, then from top to bottom, and from the left to the right write down the original ordinal numbers, and insert a semicolon between every two adjacent layers. In our example, it is $6, 7; 5; 2, 4; 1, 3$, called the second section of the name of the PAH molecule. Thus, we obtain one complete name of the PAH molecule. It is

$$2; 4/6, 7; 5; 2, 4; 1, 3,$$

in which the sign "/" is inserted between the two sections. Of course, the name obtained above is not necessarily a *standard name*, because corresponding to the twelve different orientations of the PAH molecular characteristic graph on a plane, there are twelve different names, respectively. We can select those having the minimum layers and compare them according to the "dictionary order", namely: begin at the first ordinal number, compare those ordinal numbers and punctuations as well (define ";">",","), one by one, and choose the minimum. It is the standard nomenclature series of the PAH molecules. In our example, the twelve names of the PAH molecule, corresponding to the twelve different orientations of the graph respectively, are given in Fig. 3. Notice, Fig. 2 and Fig. 3a are
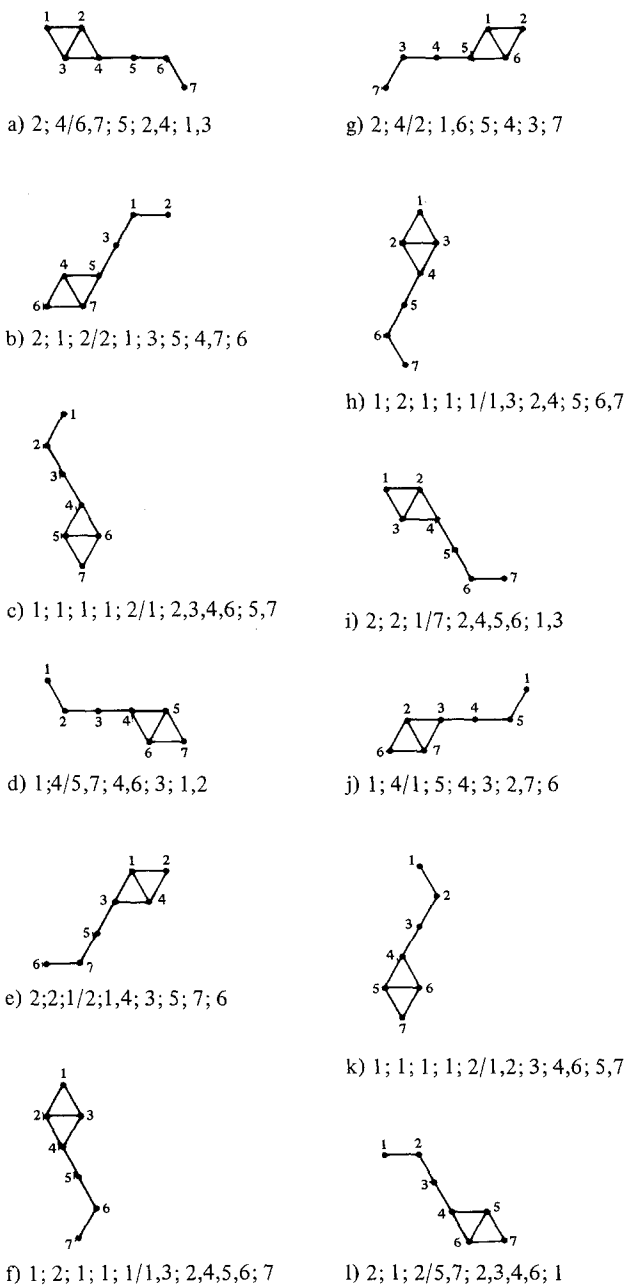
a) 2; 4/6,7; 5; 2,4; 1,3   g) 2; 4/2; 1,6; 5; 4; 3; 7

b) 2; 1; 2/2; 1; 3; 5; 4,7; 6

h) 1; 2; 1; 1; 1/1,3; 2,4; 5; 6,7

c) 1; 1; 1; 1; 2/1; 2,3,4,6; 5,7

i) 2; 2; 1/7; 2,4,5,6; 1,3

d) 1;4/5,7; 4,6; 3; 1,2   j) 1; 4/1; 5; 4; 3; 2,7; 6

e) 2;2;1/2;1,4; 3; 5; 7; 6

k) 1; 1; 1; 1; 2/1,2; 3; 4,6; 5,7

f) 1; 2; 1; 1; 1/1,3; 2,4,5,6; 7   l) 2; 1; 2/5,7; 2,3,4,6; 1

Fig. 3. Twelve graph orientations and nomenclatures

identical. The standard name of the molecule is

$\qquad$ 1; 4/1; 5; 4; 3; 2, 7; 6.

The corresponding orientation of the graph is shown in Fig. 3j.

4. Rotation and reflection

The nomenclature stated above can be carried out by computer. The key steps are to derive all others (they correspond with different orientations of the same graph) from one name, i.e. to carry out the rotation transformation and the reflection transformation.

4.1. Rotation transformation

Take its second section out of a nomenclature series and permute it to give the natural increasing series. In our example, one nomenclature series of the PAH molecule is $2; 4/6, 7; 5; 2, 4; 1, 3$. Then the permutation is

$$\begin{pmatrix} 6, 7, 5, 2, 4, 1, 3 \\ 1, 2, 3, 4, 5, 6, 7 \end{pmatrix}. \tag{2}$$

According to the permutation (2), rearrange the rows and the columns of the original He-matrix (1). Thus, we obtain the matrix (a'_{ij}), from which a new He-matrix (a^*_{ij}) is obtained, where

$$a^*_{ij} = \begin{cases} 0 \ (\text{if } a'_{ij} = 0) \\ a'_{ij} + 1 \ (\text{if } a'_{ij} = 1, 2) \\ 1 \ (\text{if } a'_{ij} = 3). \end{cases} \tag{3}$$

In the preceding example the new He-matrix is

$$(a^*_{ij}) = \begin{cases} 0 & 1 & 2 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 2 & 3 \\ 0 & 0 & 2 & 1 & 0 & 0 & 2 \\ 0 & 0 & 0 & 2 & 0 & 0 & 1 \\ 0 & 0 & 0 & 3 & 2 & 1 & 0 \end{cases}. \tag{4}$$

The corresponding graph orientation (Fig. 3b) is rotated counter-clockwise by 60° from Fig. 3a.

4.2. Reflection transformation

For one orientation of a characteristic graph, from top to bottom layer by layer and in every layer from the right to the left one by one, revert the ordinal numbers of the vertices in the graph. In our example, the orientation of the characteristic graph is shown in Fig. 2 with 2, 4, and 1 vertices in each layer, and the reflection transformation is expressed by the permutation

$$\begin{pmatrix} 1, 2, & 3, 4, 5, 6, & 7 \\ 2, 1, & 6, 5, 4, 3, & 7 \end{pmatrix}. \tag{5}$$

According to the same permutation rule, rearrange the rows and columns of the original He-matrix (1), We obtain the matrix (a'_{ij}). The new He-matrix is (a^*_{ij}), where

$$
a^*_{ij} = \begin{cases} a'_{ij} & \text{(if } a'_{ij} = 0, 1) \\ 5 - a'_{ij} & \text{(if } a'_{ij} = 2, 3). \end{cases}
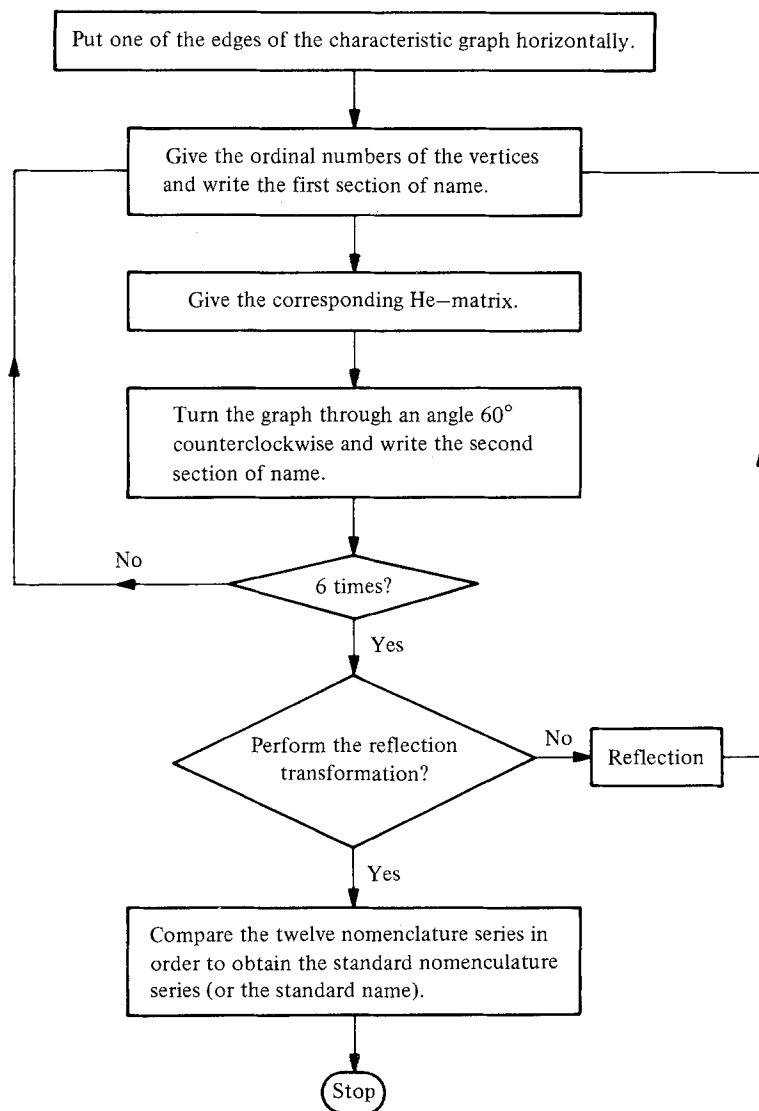\tag{6}
$$

Put one of the edges of the characteristic graph horizontally.

Give the ordinal numbers of the vertices and write the first section of name.

Give the corresponding He—matrix.

Turn the graph through an angle 60° counterclockwise and write the second section of name.

No ← 6 times?

Yes

Perform the reflection transformation? → No → Reflection

Yes

Compare the twelve nomenclature series in order to obtain the standard nomenculature series (or the standard name).

Stop

Fig. 4. Program frame

In the preceding example, the new He-matrix is

$$
(a_{ij}^*) = \begin{cases} \begin{pmatrix} 0 & 1 & 0 & 0 & 2 & 3 & 0 \\ 1 & 0 & 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 2 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 2 & 0 & 0 & 1 & 0 & 1 & 0 \\ 3 & 2 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 & 0 \end{pmatrix} \end{cases}
$$
(7)

The corresponding graph orientation (Fig. 3g) is the reflection of Fig. 3a in a vertical plane, normal to the plane of the molecule.

From our example (Fig. 2) one obtains the orientations shown in Fig. 3a–f by rotations by an angle of 0°, 60°, 120°, 180°, 240° and 300° respectively; the reflection of these orientations in a vertical plane generates the orientations shown in Fig. 3g–l. The program frame of standard nomenclature of a fused PAH molecule is shown in Fig. 4.

5. Distinguishing theorems about the automatic classification

In order to distinguish common and uncommon fused PAH molecules (the former have, the latter do not have, Kekulé structure), we set up the following theorems.

Let U be the total number of upright triangles in the PAH molecular characteristic graph and D the total number of the upset triangles in the same graph (the so-called upright triangle is one whose horizontal side is at the bottom of the triangle, the upset triangle is one whose horizontal side is at the top of the triangle [20, 21]). Then one may state:

Theorem 1. *If* $|U - D| = k$, *then in the* PAH *molecule there are at least k atoms which have* sp^3 *hybridization.*
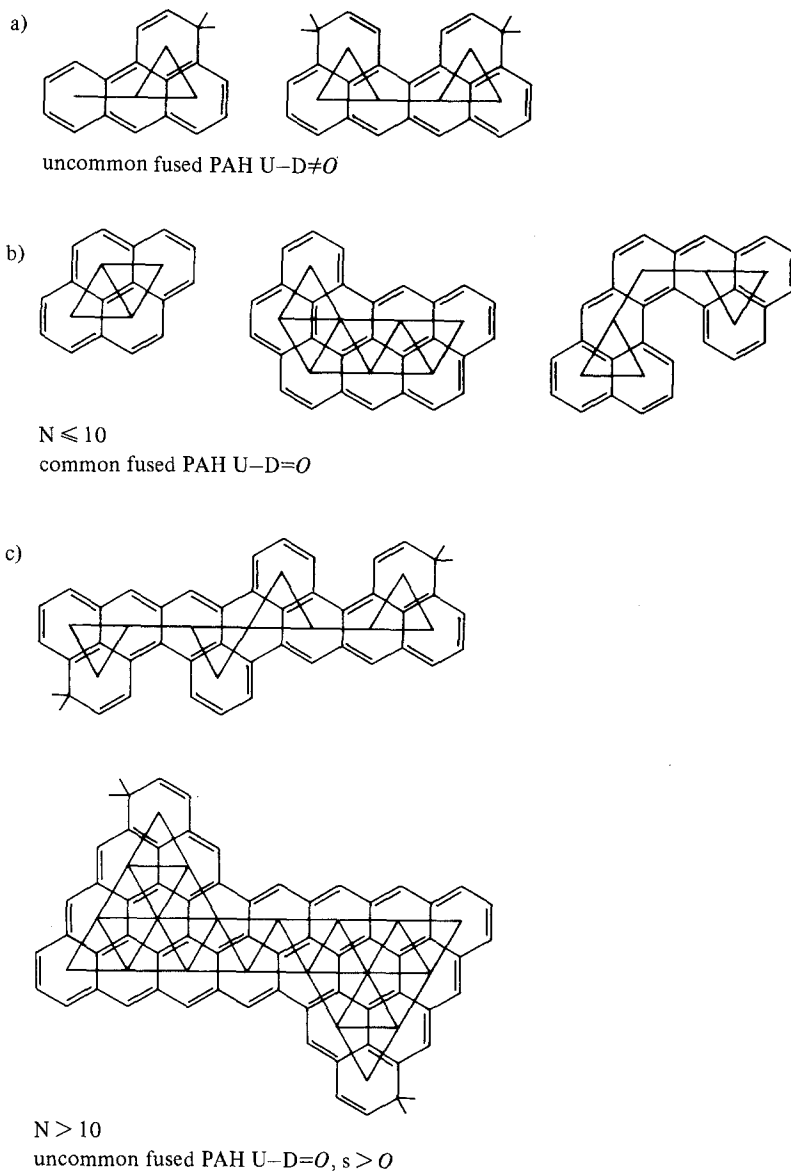
We can prove that $|U - D|$ is equal to the difference of number of starred carbon atoms and non-starred carbon atoms in the PAH molecule graph by using the properties of a pair graph in graph theory [22].

Now we discuss the consequence of Theorem 1:

Let us denote the number of those atoms of a given PAH which have sp^3 structures. Obviously $s \geq k = |U - D|$ follows from Theorem 1 and, hence, one has to consider the following three cases which are illustrated by Fig. 5:

(i) If $k \neq 0$, then $s > 0$ and as consequence of this the PAH considered cannot be common (Fig. 5a).

(ii) If the fused PAH is common, we have $s = 0$ by definition and consequently $U = D$ must hold (Fig. 5b).

a)

uncommon fused PAH U−D≠O

b)

N ⩽ 10
common fused PAH U−D=O

c)

N > 10
uncommon fused PAH U−D=O, s > O

Fig. 5. Common and uncommon PAH

(iii) In constrast to that from $U - D = 0$ no conclusion about s can be drawn and $s > 0$ may be valid (Fig. 5c).

Although we can show that for PAH with ten or less sixmembered rings the case (iii) cannot be realized, a method is needed to decide whether $U - D = 0$ corresponds to case (ii) or (iii). For that purpose we define the cut set and develop Theorem 2.

Definition. If any two connected induced subgraphs G_1 and G_2 [13, 23] of the characteristic graph $G(V, E)$ satisfy:

(A) $V = V(G_1) \cup V(G_2)$ and $E = E(G_1) \cup E(G_2)$;
(B) neither of $E(G_1)$ and $E(G_2)$ is empty;
(C) For every vertex of $V(G_1)$ or $V(G_2)$ if we deleted it then (A) would always be violated.

Then G_1 and G_2 are referred to as a cutting of $G(V, E)$ denoted by $C(G_1, G_2)$; $V' = \{V(G_1) \cap V(G_2)\}$ is called the cut set.

For example, for the characteristic graph shown in Fig. 2 there are only four cuttings shown in Fig. 6a–d by this definition. However, Fig. 6e is not a cutting, because when vertex V_2 is deleted from G_2 of Fig. 6e item (A) is not violated.

Lemma 1. *Any representative point of a ring (sixmembered ring in a* PAH *molecule) containing* sp^3 *structure can be passed along any straight chain in the characteristic graph to any vertex on the chain.*

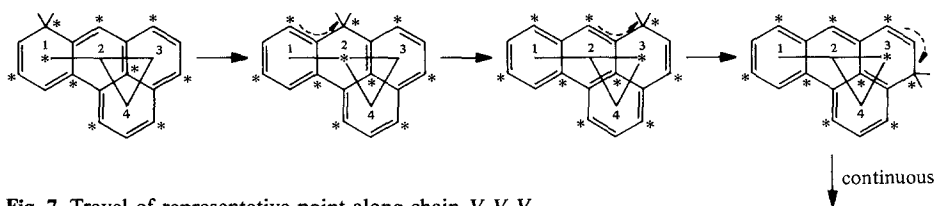It is illustrated by Fig. 7.
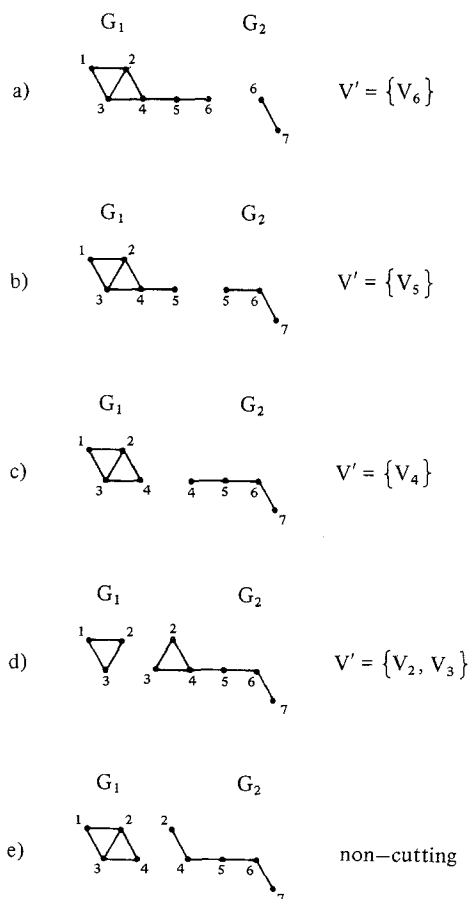
Fig. 6. Cutting and non-cutting

Fig. 7. Travel of representative point along chain $V_1 V_2 V_3$

Lemma 2. *Any representative point of a sixmembered ring containing sp^3 structure can be passed to any vertex in the characteristic graph through any path connecting the two vertices.*

It is not difficult to prove these two lemmata. Here we do not give details.

There are two classes of sp^3 structure point (a starred one, as an example shown in Fig. 7, and non-starred one). And so in those corresponding characteristic graphs there are also two classes of representative points of the ring containing sp^3 structure. (Note: one sixmembered ring containing sp^3 in a RAH graph only have one representative point in the corresponding characteristic graph, by rule). They are also called "starred" and "non-starred", respectively. If one starred and one non-starred representative point travel to meet each other in a characteristic graph, the two sp^3 structures are cancelled out. As an example see Fig. 8.

As stated above, starred and non-starred representative points can travel to meet each other in pair, through non-intersected paths, in order to eliminate sp^3 structures. Hence, using Lemmata 1 and 2, the connectivity of characteristic graphs, and Minger's theorem [13, 23], we can prove the following theorem.

Theorem 2. *Let $G(V, E)$ be the characteristic graph of a* PAH *molecule satisfying $U - D = 0$; if and only if all possible cuttings of $G(V, E)$ satisfy*
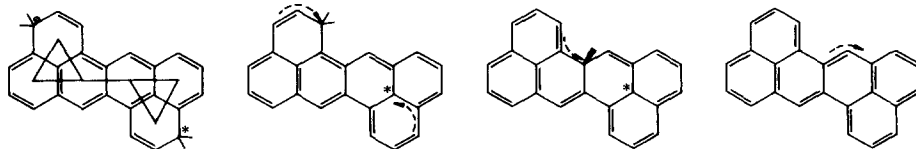
$$|V'| \ge |U_{G_1} - D_{G_1}|, \tag{8}$$

then the PAH *molecule is common, where $|V'|$ is the number of the vertices in the cut set V', U_{G_1} and D_{G_1} represent the number of the upright triangles and the number of the upset triangles in G_1, respectively.*

a) *elimination*

characteristic graph

b) *elimination*

molecule graph

Fig. 8. Travel and elimination of sp^3

By this theorem, the PAH molecules shown in Fig. 2, Fig. 5b and Fig. 8 are common, but those shown in Fig. 5c, although $U - D = 0$, are uncommon.

6. Method of classification

Let V_1, V_2 and V_3 represent the number of vertices of degree 1, 2 and 3 in G, respectively.

The automatic classification of fused PAH molecules is as follows. A fused PAH molecule is

A) a common one, if the conditions of Theorem 2 hold; otherwise it is uncommon;
B) a chain fused molecule (cata-condensed) without branches, if $U + D = 0$, $V_3 = 0$ and $V_1 \neq 0$;
C) a chain fused molecule (cata-condensed) with branches, if $U + D = 0$, $V_3 \neq 0$ and $V_1 > V_3$;
D) a peri-fused or cyclic peri-fused molecule, if $U + D \neq 0$ and $V_1 = 0$;
E) a cyclic fused molecule (corona-condensed), if $U + D = 0$, and $V_1 = 0$;
F) a derived fused molecule from D) and E), then the condition $V_1 = 0$ in D) and E) is changed into $V_1 \neq 0$. For a derived fused molecule from E), an additional condition is $V_1 \leq V_3$.

Using the He-matrix, we can enumerate V_1, V_2, V_3 and U, D by computer. In fact, the number of nonzero elements in the ith row (or column) represents the degree of the ith vertex. And

$$U = \sum_{i=1}^{N-2} u_i, \tag{9}$$

where

$$u_i = \begin{cases} 0 \left(\text{if } \sum_{j=i+1}^{N} a_{ij} = 0, 1, 2, 3, 4 \right) \\ 1 \left(\text{if } \sum_{j=i+1}^{N} a_{ij} = 5, 6 \right) \end{cases} \quad (i = 1, 2, \ldots, N-1), \tag{10}$$

$$D = \sum_{i=1}^{N-2} d_i, \tag{11}$$

where

$$d_i = \begin{cases} 0 \left(\text{if } \sum_{j=i+1}^{N} a_{ij} \neq 4, 6 \right) \\ 1 \left(\text{if } \sum_{j=i+1}^{N} a_{ij} = 4, 6 \right) \end{cases} \quad (i = 1, 2, \ldots, N-1). \tag{12}$$

The proof of (9)–(12) can be easily obtained from the properties of the He-matrix.

Table 1. Enumeration of PAH's

N	Common fused PAH molecules						Uncommon			
	B	C	D	F	E	Total	D	F	Total	Total
1	1	0	0	0	0	1	0	0	0	1
2	1	0	0	0	0	1	0	0	0	1
3	2	0	0	0	0	2	1	0	1	3
4	4	1	1	0	0	6	0	1	1	7
5	10	2	1	2	0	15	1	6	7	22
6	24	12	3	12	0	51	3	27	30	81
7	67	51	6	66	0	190	9	132	141	331
8	182	229	22	331	1	765	27	644	671	1436

7. Discussion

1. Using the He-matrix, we can also perform the automatic derivation of fused PAH molecules. In this paper we will not introduce it in detail.

2. Only using a personal computer, we obtain Table 1. Our data are partly different from those in [6, 7, 24–31]. We will discuss the reason in another paper.

3. This algorithm belongs to the polynomial algorithm (P-algorithm) according to the complicated structure of the algorithm.

4. Of course, there are still some other nomenclature methods without using the concept of graph center, such as to count the number of two-degree-vertices along the rings of the circumference in a PAH molecule graph one by one. But such a method is not suitable for cyclic fused, cyclic peri-fused and their derived fused PAH molecules. We also considered the method to count the vertices along a Hamiltonian cycle [13, 23]. But since many PAH molecule graphs are not Hamiltonians, such a method is not suitable either. After that, we considered the method of superimposing a dualist graph containing N vertices on a hexagonal lattice whose one diagonal line has N or $N + 1$ crosspoints (it is easy to prove that we can do this). Perform the rotation and the reflection to the dualist graph. After every transformation, translate the dualist graph to the top left corner of the lattice and give the coding of the graph. Then compare those codings and give the standard nomenclature. But such a method is too laborious even by computer, and so we gave it up.

5. Like D. Bonchev's work [12], the method in this paper can also be generalized for other kinds of organic compounds.

References

1. Cahn, R. S.: Introduction to chemical nomenclature (3rd ed.) London: Butterworths 1968
2. Dyson, G. W.: A new notation and enumeration system for organic compounds (2nd ed.) London: Longman 1949
3. IUPAC, Nomenclature Rules for Organic Compounds. London: Butterworths 1965

4. Patterson, A. M., Capell, L. T., Walker, D. F.: The ring index (2nd ed.) Washington, DC: American Chemical Society 1960
5. Index Guide Introduction. Chem. Abstr. **76** (1972)
6. Balaban, A. T., Harary, F.: Tetrahedron **24**, 2505–2516 (1968)
7. Balaban, A. T.: Tetrahedron **25**, 2949–2958 (1969)
8. Balaban, A. T. (ed.): Chemical Applications of graph theory. London: Academic Press 1976
9. Bonchev, D., Balaban, A. T., Mekenyan, O.: J. Chem. Inf. Comput. Sci. **20**, 106–113 (1980)
10. Bonchev, D., Balaban, A. T., Randić, M.: Int. J. Quantum Chem. **19**, 61–82 (1981)
11. Bonchev, D., Balaban, A. T.: J. Chem. Inf. Comput. Sci. **21**, 223–229 (1981)
12. Bonchev, D.: Pure and Appl. Chem. **55**, No. 22, pp. 221–228 (1983)
13. Bondy, J. A., Murty, U. S. R.: Graph Theory with Applications. London: Macmillan 1976
14. Cioslowski, J., Turek, A. M.: Tetrahedron **40**, 2161 (1984)
15. Whitney, H.: Am. J. Math. **54**, 150 (1932)
16. Whitney, H.: Trans. Am. Math. Soc. **34**, 339 (1932)
17. Whitney, H.: Fund. Math. **21**, 73 (1933)
18. Polansky, O. E., Rouvray, D. H.: Math. Chem. **2**, 147 (1976)
19. Polansky, O. E., Rouvary, D. H.: Math. Chem. **2**, 63 (1976)
20. Balaban, A. T.: Rev. Roum. Chem. **26**, 407 (1981)
21. Balaban, A. T.: Pure and Applied Chem. **54**, 1075 (1982)
22. Coulson, C. A., Rushbrooke, G. S.: Proc. Cambridge Phil. Soc. **36**, 193 (1940)
23. Harary, F.: Graph Theory. Reading: Addison-Wesley, MA 1972
24. Klarner, D. A.: Fibonacci Quart. **3**, 9 (1965)
25. Klarner, D. A.: Can. J. Math. **19**, 851 (1967)
26. Balaban, A. T.: Math. Chem. **2**, 51 (1976)
27. Harary, F.: Read, R. C.: Proc. Edinburgh Math. Soc. **17**, 1 (1970)
28. Lunnon, W. F.: Graph Theory and Computing, R. C. Read (ed.) New York: Academic Press 1972
29. Dias, J. R.: J. Chem. Inf. Comput. Sci. **22**, 15, 139 (1982)
30. Dias, J. R.: Math. Chem. **14**, 83 (1983)
31. Knop, J. V., Szymanski, K., Jericević, Z., Trinajstić, N.: Math. Chem. **16**, 119 (1984)