CROATICA CHEMICA ACTA CCACAA 57 (1) 21—45 (1984)

CCA-1419

YU ISSN 0011-1643
UDC 547.53
Original Scientific Paper

Molecular Topology and Molecular Properties I. »Sequence« Labels of Cata-Condensed Benzenoid Hydrocarbons Containing up to Five Rings*

Sherif El-Basil

Faculty of Pharmacy, Kasr El-Aini Street,
Cairo, Egypt

Received March 25, 1983

An algorithm is described for the systematic numbering of benzenoid hydrocarbons. The sum of decimal contributions of all vertices (carbon atoms) in the molecular graph is called the hydrocarbon label, L. These labels, which are based on purely structural elements order a set of isomeric benzenoid hydrocarbons according to their properties including resonance and total pi-energies, ionization potentials, heats of atomization, energies of lowest vacant and lowest filled molecular orbitals and polarographic half-wave reduction potentials. Such hydrocarbon labels might easily be computed without resort to a computer.

INTRODUCTION

Several years ago Randić[1] described a method whereby the vertices (atoms) in a molecular graph are numbered in a special way so that the rows of their adjacency matrix yield, when combined, a minimum and unique label characteristic of the graph. Randić's main objective was to recognize molecular topology by associating it with some particular label that is invariant of the graph under consideration. Some problems, however, arose[2,3] due to possibility of existence of local minima when applying one of the suggested algorithms to certain graphs (but these can be avoided by using alternative algorithm). In another development Randić[4] was able to order alkanes by adopting a particular canonical labelling of atoms associated with the smallest binary code for the structure. Another application of canonical labeling of atoms is to study symmetry properties of graphs.[5—7]

In this work a systematic method is described for labelling molecular graphs of benzenoid hydrocarbons, BH's, following a particular sequence (whence the subtitle of the manuscript). The procedure, when carried out according to the prescribed algorithm, is error-free. A flow chart of the procedure of assigning numbers to the vertices of graphs shown in Fi-

---

* This paper is dedicated to Professor A. T. Balaban.

gure 4 is given in Figure 6. The adjacency matrices of the resulting labeled graphs lead to particular codes that characterize their topologies. The sum of the components of the code is called the label of the graph. The assumption is made that two (or more) graphs will not yield identical codes unless they are isomorphic. The importance of molecular topology in determining molecular and even some biological properties[8–10] justifies this paper. Furthermore, the reported relation between hydrocarbon graph-labels and molecular properties represents an additional illustration of the link between VB and MO theories[11,12] (graph labels are closely related to VB quantities while many molecular properties can be discussed within the MO model).

DEFINITIONS

Before describing our method of assigning numbers to the vertices of a polyhex[13] graph, certain definitions will be stated which are necessary for writing the procedure and for discussion.

(1) Balaban-Harary Dualist,[14] D (or Simply Dualist)

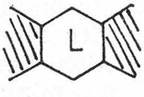
A dualist results when the hexagons of a polyhex are replaced by vertices and then adjacent vertices corresponding to rings which had a common edge in the polyhex are connected.

(2) A Dualist Vertex, v(d)

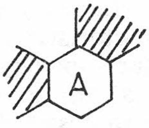
A vertex of a dualist (and not of the hydrocarbon: the latter will simply be termed, vertex).

(3) The L, A-sequence[15]

The nonterminal hexagons in a nonbranched cata-condensed system might be annelated in just two ways, viz.

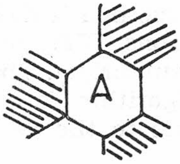


L-mode

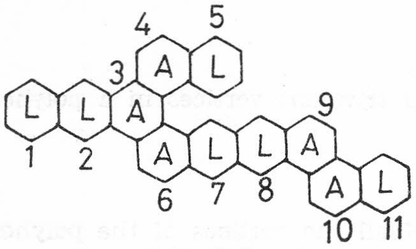


A-mode

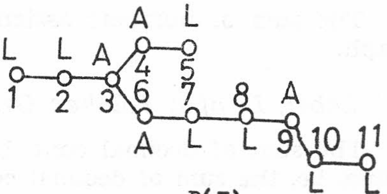
Furthermore, an additional convention will be adopted that a hexagon which is fused simultaneously to three other hexagons (i.e. a branched hexagon) will be assigned the symbol A, viz.,



Therefore every polyhex graph of a cata-condensed BH might be associated with an ordered i-tuple of symbols L, A (i = number of hexagons), the so-called L, A sequence.[15] By definition the terminal rings will be given the symbol L. As an illustration G(I) has the following L,A symbols assigned to its rings:



G(I)



D(I)

(4) Induced Subgraph[16,17] *(or simply subgraph)*

Three types of induced subgraphs[16,17] might be partitioned out of the total molecular graph of the BH, viz.,

$$< g_1 > = < L^{s_1} A > \tag{1}$$

$$< g_2 > = < A\, L^{s_2} A > \tag{2}$$

$$< g_3 > = < A\, L^{s_3} > \tag{3}$$

where L^{s_1}, e.g. means L repeated s_1 times. The values of s_1 and $s_3 = 0$, while s_2 may or may not be zero depending on the topology of G. E.g. *G(I)* has the following induced subgraphs.

$< g_1 > \supset < L^2A >$
$< g_2 > \supset < A^2 >, < A^2 >, < AL^2A >, < A^2 >$
$< g_3 > \supset < AL >, < AL >.$

which, when represented as labels of rings (or corresponding dualist vertices), become:

$< g_1 > \supset < 1, 2, 3) >$
$< g_2 > \supset < 3, 4 >; < 3, 6 >; < 6, 7, 8, 9 >; < 9, 10 >$
$< g_3 > \supset < 4, 5 >; < 10, 11 >$

The order of an induced subgraph will be defined here as the number of its dualist vertices (corresponding to rings defined by eqns. 1—3). Thus the phenanthrene dualist, for example, is partitioned into two induced subgraphs each of order two, *D(I)* has one induced subgraph of order 3, five induced subgraphs of order two each and one induced subgraph of order 4.

(5) Valency (Degree) of a Subgraph

The highest degree given to one of its dualist vertices. Thus the phenanthrene dualist, for example, contains two bivalent subgraphs; while the triphenylene dualist has three trivalent subgraphs. Benzanthracene contains two bivalent subgraphs (one of order two and the other of order three).

(6) Trivalent Sum, $S(3)$

The sum of numbers assigned to all trivalent vertices in a polyhex graph.

(7) Label, L, of a Polyhex Graph

The sum of decimal contributions of all the vertices of the polyhex graph. i.e. the sum of decimal contributions of the rows of the adjacency matrix that corresponds to the particular labeling (numbering) of atoms (vertices). Thus a divalent vertex, for examples, has two adjacent neighbours. Let the numbers given to them be k and j, then the decimal contribution of such a divalent vertex is $2^{n-k} + 2^{n-j}$ where n is the total number of vertices in the polyhex graph. The approach is easily extended to a trivalent vertex. In general, L can be expressed as

$$L = \sum_{\text{all v}} \sum_{j} 2^{n-j} \tag{4}$$

where the first sum is taken oven all vertices and the second over all adjacent vertices around each vertex. (Tables 1 and 2 present decimal contributions of the hydrocarbons shown in Figure 4 according to the labeling shown in the figure). Eqn. (4) might be expressed as $L = \sum_{j=1}^{n} d_j$ where d_j is the decimal contribution of the jth row.

(8) $v_y(x)$

A symbol given to a vertex, the degree of which is x and which is assigned the number y. In this part of the work, since we will be dealing only with polyhex graphs, x might take only the values of 2 or 3 while the number of y is defined in the interval $1 \leqq y \leqq n$.

(9) $v(\ \ )$

A vertex whose valency is not assigned.

(10) Unassigned Vertex

A vertex which has not been assigned a number.

(11) Kink Vertex

A dualist vertex defining the intersection of two induced subgraphs of a dualist, one of which is terminal. Kink vertices are drawn as solid circles in Figure 4.

(12) Arranged Matrix[18]

An adjacency matrix whose decimal (or binary) values of its rows form an ascending sequence i.e. $d_j < d_{j+1}$, $(1 \leqq j \leqq n)$ where d_j is the numerical (decimal or binary) contribution of the jth row.

(13) $v_{\min}(3)$

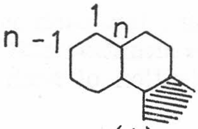
The smallest number assigned to a trivalent vertex of the polyhex graph. Such numbers are encircled in the graphs shown in Figure 4. These numbers turn out to have significance in relation to the molecular properties associated with the polyhex graph (see later).

THE NUMBERING PROCEDURE

We aim at a procedure yielding vertex numberings which result in an arranged matrix such that, when read sequentially, the rows of the matrix give (after addition) the smallest label. The following procedure might be adopted for graphs with vertices of degrees 2 and 3 only. It cannot be applied to vertex-transitive graphs (i.e. graphs with all vertices equivalent). Consider a graph with n vertices:

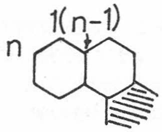
(1) The Numbers 1, n and (n—1)

Assign the number one to one of its *divalent* vertices, $v_1(2)$, the two adjacent vertices on both sides of it are then given the numbers n and n—1. There are two factors which decide the position of v_n (to the left or to the right of v_1): (i) The degree of the vertex adjacent to v_n, and (ii) the degree of v_n itself. The »best« situation would be that v_n is a trivalent vertex followed by a divalent vertex as exemplified.

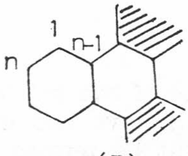


(A)

not

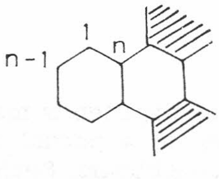


If, however, a $v(3)$ adjacent to v_1 has a neighbour another $v(3)$, one examines the other side of vertices adjacent to v_1. If there are two successive $v(2)$'s adjacent to v_1, one assigns the number n to the vertex adjacent to v_1. This situation is illustrated below:

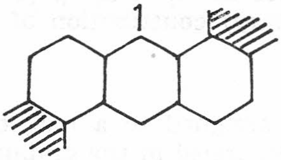


(B)

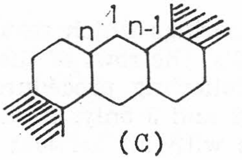
not



A third possibility would be that the two adjacent vertices on both sides of v_1 are trivalent, viz.,

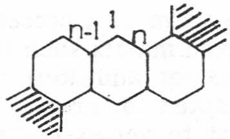


One then examines the environments on both sides of v_1: the number n is assigned to a $v(3)$ that is flanked by v_2's (if any), so, for the case above we would have:

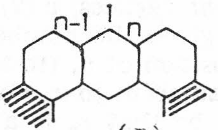


(C)

not

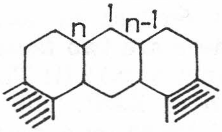


When both v's adjacent to v_1 are trivalent, each being followed by $v(2)$, one attempts both possibilities, e.g.

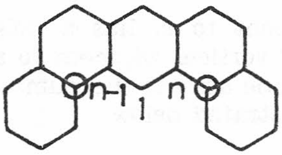


(D)

and



Finally if both v's surrounding v_1 are trivalent, but each is also flanked by a $v(3)$, one gets »locked-up« in a »dead-end« situation and must attempt to start the sequence at another v_1. An illustration of such a partial sequence is shown below:



(Ø)

A dead-end situation is manifested by the presence of two $v(3)$'s defining the termini of a partial sequence such that both have two unassigned adjacent neighbours. Such trivalent vertices might be termed dead-end vertices, $v(\emptyset)$. As a general rule an assignment of the type Ø should never be used.

(2) The Numbers 2 and 3

Under the »best conditions« the lower number, 2, is assigned to a divalent vertex adjacent to v_n. If the vertex adjacent to v_n is trivalent, but the one adjacent to v_{n-1} is divalent, the number 2 is given to the latter. If, however, both vertices adjacent to v_n and v_{n-1} are trivalent, one checks for the possibility of a dead-end situation: if both $v(3)$'s adjacent to v_n and v_{n-1} have two unassigned adjacent vertices, one abandons the possibility of starting the cycle at $v_1(2)$ and must look for another $v(2)$. If, however, one of the two $v(3)$'s flanked by v_n or v_{n-1} has two assigned neighbours, it is given the number 2. If both $v(3)$'s adjacent to v_n andv_{n-1} have two assigned neighbours, the number 2 is given to $v(3)$ adjacent to v_n. (This last possibility cannot happen at this stage of the sequence in polyhex graphs).

(3) The Numbers (n—2) and (n—3)

These are automatically assigned to vertices adjacent to v_2 and v_3 respectively. For example a possible sequence would be:

$$\ldots v_a(x)v_{n-3}(\quad)v_3(2)v_{n-1}(2)v_1(2)v_n(3)v_2(2)v_{n-2}(\quad)v_b(y)\ldots$$

(4) The Numbers 4 and 5

The places of 4 and 5 depend on the values of x and y, i.e. on the degrees of the v's adjacent to v_{n-2} and v_{n-3}. We might list the following possibilities:

(a) $x = y = 2 \longrightarrow b = 4, a = 5$

(b) $x = 2$ $(y = 3)$ or $y = 2$ $(x = 3)$ $\longrightarrow$ the lower number, 4, is given to the divalent vertex, then the environment of $v(3)$ is examined: if two assigned adjacent vertices are present, it is given the number 5, if it turns out to be a »dead-end« $v(3)$ the sequence is propagated from the other side by assigning $(n\text{-}4)$ next to $v_4(2)$.

(c) $x = y = 3$

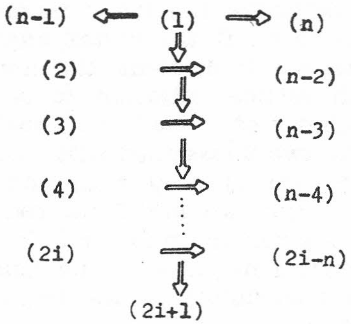
If both (x) and (y) are dead-end vertices one starts the cycle at another $v(2)$, but if one possesses two assigned neighbours, the number 4 is assigned to it. If, finally, both (x) and (y) have two assigned neighbours, then $b = 4$, $a = 5$ i.e. the smaller number is assigned to the v carrying the largest number.

(5) Numbers (n-4) and (n-5)

These are automatically assigned to v's adjacent to v_4 and v_5 respectively.

(6) The process is continued until all vertices are visited.

The sequence of numbers takes the following form:



where i is the number of hexagons (benzene rings) in the polyhex graph (of cata-condensed BH); so the last number given to a vertex is related to the number of benzene rings in the total molecular graph.

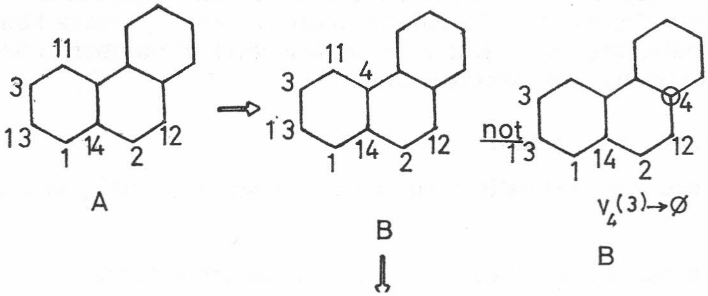
(7) The process is attempted with all other bivalent vertices and trivalent sums, *S*(3)'s computed in each case.

(8) The sequence(s) that leads (lead) to the largest S(3) is (are) choosen and its (their) adjacency matrix (matrices) checked for being arranged.[18] The presence of disorder (i.e. non-arranged matrix) might be remedied by the appropriate exchange of numbers (see later for examples and illustrations).
It must be emphasized that a non-arranged matrix is rejected even if it leads to a smaller L than the arranged matrix (e.g. numberings *D* and *E* below). The whole procedure of labelling is explained in the flow chart shown in Figure 6.

Some Illustrations

I. Phenanthrene

This graph has four *v*(3)'s and ten *v*(2)'s. The latter may be divided into two equal sets of equivalent *v*(2)'s. So we have five possibilities. Figure 1. shows the sequence resulting when vertex number one (in the IUPAC system) is attempted:



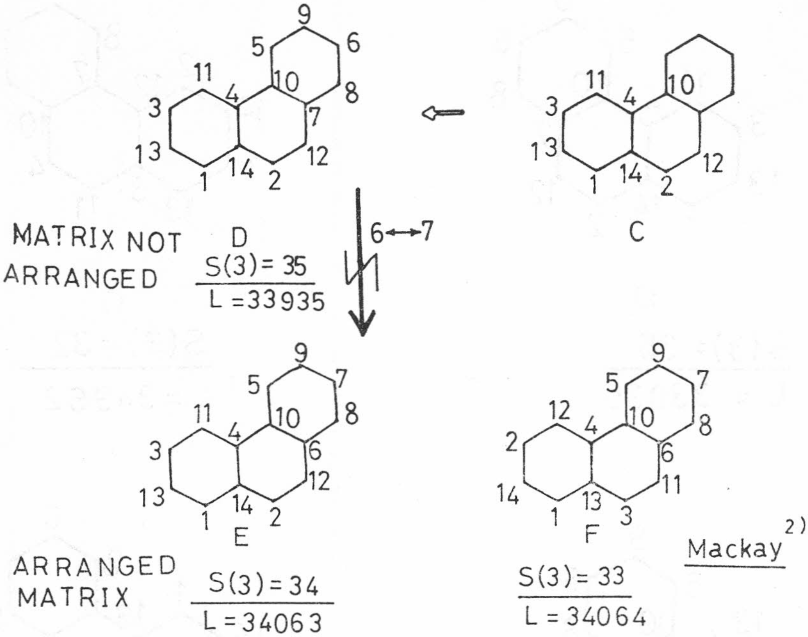
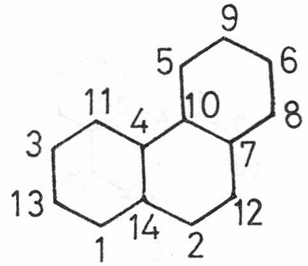


Figure 1. Labelling sequence for the phenanthrene graph when starting with number one at v_1 according to IUPAC rules. Observe that 7 (=2(3)+1) is the last number to be assigned, which must then be exchanged with 6 to yield an arranged matrix *(E)*. Label F is the sequence proposed by Makay[2]

There is no difficulty in assigning the first seven numbers, step A, (Obviously by the »first« seven numbers we do not mean the numbers 1—7).

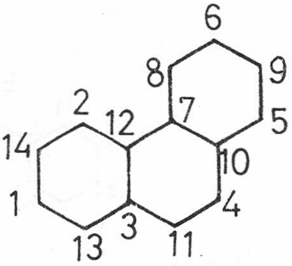
Assignment of the number 4 (Step B) might require a moment of thought: considering the two available vacant vertices, we see that the $v(3)$ adjacent to $v_{12}(2)$ has two unassigned adjacent vertices, so it is abandoned, and thus we are left with the $v(3)$ adjacent to $v_{11}(2)$ which has two assigned adjacent vertices, so it is given the number 4. Assigning n—4(=10) is a routine task, since it should be adjacent to v_4. Step D is based on the fact that divalent vertices take priority over trivalent ones. Sequence D, however, leads to a nonarranged matrix. This becomes clear when we list decimal contributions from the rows successively; these are: {3, 5, 10, 25, 48, *96*, *84*, 384, 768, 1664, 3072, 4224, 10240, 13312}, (leading to an L = 33935). The above string of d_j's does not define order[18] because $d_6 > d_7$ yet other d_j's define the relation $d_j < d_{j+1}$ (d_j = decimal contribution of the jth row). Therefore labelling *D* should be arranged by exchanging 6 → 7 which leads to *E*. The latter is an arranged matrix, leading to the following numerical contributions of the individual rows: {3, 5, 10, 25, 48, 84, 96, 384, 640, 1792, 3072, 4352, 10240, 13312}, which is now ordered. The label from *E* is 34063. Although *L(D)* < *L(E)*, it is rejected beca-



D

S(3) = 35

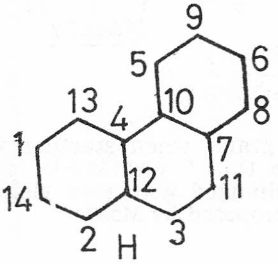
L = 33935



G

S(3) = 32

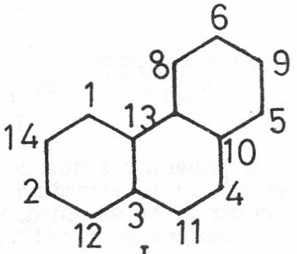
L = 34962



H

S(3) = 33

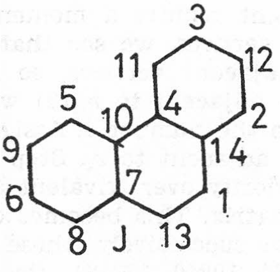
L = 33938



I

S(3) = 32

L = 34960



J

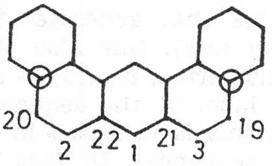
S(3) = 35

L = 33935

Figure 2. Labellings resulting from starting with number one at each of the five non-equivalent vertices of phenanthrene. Values of *S*(3)'s and *L*'s are given. Observe that *D* and *J* lead to identical *S*(3) and *L* yet their adjacency matrices are not identical

use D is not an arranged matrix. It is interesting to observe that sequence E is the one originally assigned by Randić.[1] The one proposed by Mackay[2] (sequence F), however, leads to a larger label! (in decimal). Thus F leads to $\{3, 5, 10, 22, 48, 88, 96, 384, 640, 1792, 2304, 5120, 11264, 12288\}$ giving a label of 34064. Figure 2 shows sequences resulting from starting with all non-equivalent divalent vertices of phenanthrene together with their labels and trivalent sums. Such quantities will be discussed later in relation to RE's of the hydrocarbons.

II. Another example of using the prescribed algorithm that illustrates a vertex leading to a »dead end« is probably desirable. Consider the dibenzanthracene graph shown below. The sequence leads to the appearance of two trivalent vertices each with two unassigned adjacent neighbours (encircled). One cannot proceed any further with the remaining numbers and another divalent vertex must be attempted.



III. A third example illustrates how one assigns priorities to a pair of available divalent or trivalent vertices:

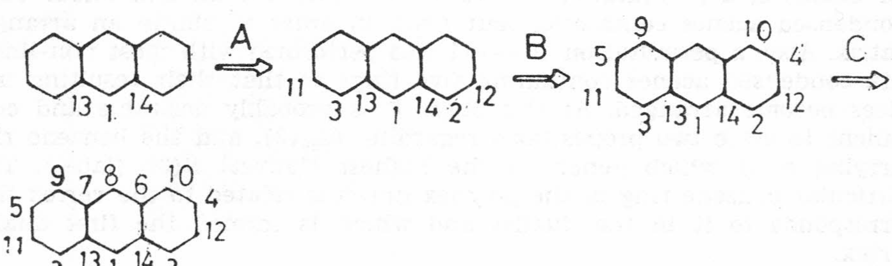


Figure 3. Sequence of numbers assigned to anthracene starting with the number one as shown

Figure 3 shows how one arrives at the anthracene sequence of numbers. One observes that the resulting sequence is identical to the one given earlier by Randić[1].

In step A one assigns the numbers 2 and 3. Since both vertices adjacent to $v_{14}(3)$ and $v_{13}(3)$ are divalent, priorities are given according to numbers assigned in adjacent places: Thus 2 is given to the vertex adjacent to $v_{14}(3)$. Numbers 12 (i.e. 14-2), and 11(14-3), are automatically assigned to v's adjacent to v_2 and v_3. Again the choice in step B is decided so that the lower of the two available numbers (4,5) is assigned to the $v(2)$ adjacent to the v carrying the largest number.

Thus, 4 is assigned to the vertex adjacent to v_{13} while 5 is located next to v_{12}. The number 10(= 14-4) is immediately assigned adjacent to v_4

and 9(= 14-5) is similarly given to the vertex adjacent to v_5. In step C we assign the remaining three numbers 6,7 and 8: the numbers 6 and 14-6(= 8) are assigned to vertices that immediately flank v_{10} (rather than v_9). The remaining number, 7, is then given to the remaining vertex (observe that 7, which is the last number to be assigned is $2i+1$, $i=3$ for anthracene).

RESULTS AND DISCUSSION

Figure 2 shows numbers assigned to phenanthrene starting from v_1 at each of its five non-equivalent divalent vertices. One observes that there is a relation between trivalent sums, $S(3)$ values, and code sums i.e. labels, L's, (sums of decimal contributions from each atom): the larger the value of $S(3)$, the lower is the label. Figure 4 shows sequences assigned to cata-condensed BH's which are based on assigning the number 1 to the bivalent vertex that generates the highest $S(3)$ (followed by arranging the matrix if any). One also observes that the minimum number given to a trivalent vertex, $v_{\text{min}}.(3)$ is significant: the higher this number, the lower is the label of the sequence; and the higher is the corresponding $S(3)$. Again this is reflected in Figure 2 for the five non-equivalent phenanthrene sequences. In fact labels D and J (Figure 2) are identical and so are their $S(3)$ and $v_{\text{min}}(3)$ values, yet their adjacency matrices are not. One should recall that once the highest $S(3)$ is obtained (based on flow chart instructions), the matrix is then checked for order. For example, a permutation $8 \longleftrightarrow 9$ is necessary for all non-linear cata-condensed acenes containing four rings in order to obtain an arranged matrix. Also, a permutation $10 \longleftrightarrow 11$ was performed with most non-linear cata-condensed acenes containing five rings so that their resulting matrices become arranged. At this stage it is probably desirable and convenient to state two propositions regarding $v_{\text{min}}(3)$, and the benzene ring carrying $v_1(2)$ which generates the highest (lowest) $S(3)$ (label). This particular benzene ring in the polyhex graph is related to the vertex that corresponds to it in the dualist and which is termed the first dualist vertex.

Proposition 1

a) for a linear acene containing i rings:

$$v_{\text{min}}(3) = i+3$$

b) For a non-linear acene $v_{\text{min}}(3) = 2+$ the order of the longest terminal subgraph of the dualist of degree ≤ 2 (Terminal subgraphs are defined by $<g_1>$ and/or $<g_3>$ of eqns. 1 and 3 respectively). A few examples would probably illustrate this proposition, especially part b): Figure 4 shows all the hydrocarbons studied in this work where divalent terminal dualist subgraphs of highest order are heavily drawn and values of $v_{\text{min}}(3)$ are encircled around appropriate vertices. In triphenylene, *9*, the order of the terminal dualist subgraph of valency ≤ 2 is 1, so $v_{\text{min}}(3) = 2+1 = 3$ (Triphenylene is an exception in the sense that its dualist contains no bivalent vertices, so our terminal $<g>$ is just a

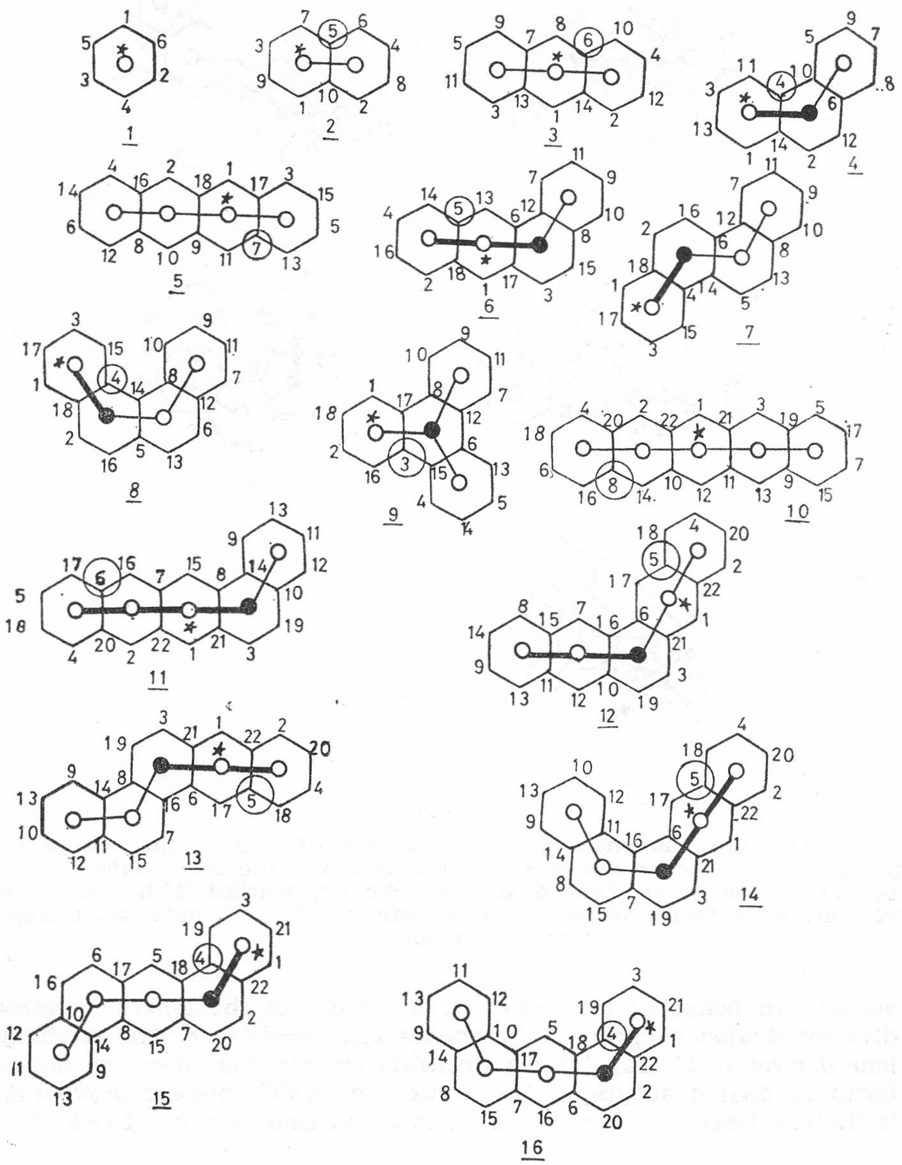


Figure 4. continued on the next page.

Figure 4. contin.

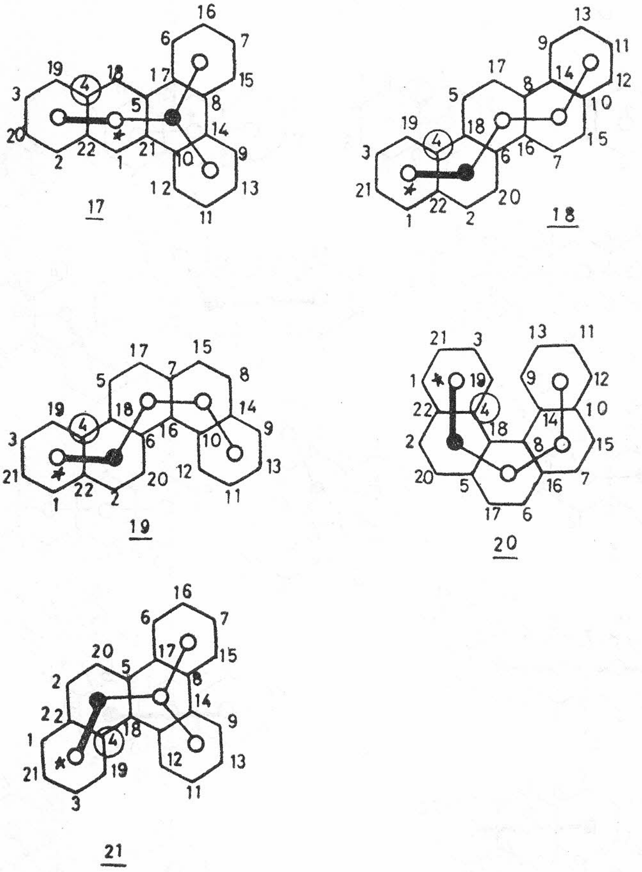


Figure 4. Labellings assigned to cata-condensed BH's containing up to five rings leading to arranged matrices with largest $S(3)$'s and lowest labels. Values of $v_{\min}(3)$ are encircled. First dualist vertices are marked with stars. Kink vertices are solid dualist-vertices. Longest terminal divalent induced subgraphs are heavily outlined

vertex). In benz [*a*] tetracene, 11, the order of the longest terminal divalent dualist subgraph is 4, whence $v_{\min}(3)=2+4=6$. The triphenylene derivative *17* may further illustrate the method: here the longest terminal dualist subgraph, the vertices of which possess degrees $\leqq 2$, is the one heavily drawn. Its order is 2 and therefore $v_{\min}(3)=2+2=4$.

Proposition 2

The »first dualist vertex« is located on the longest terminal dualist subgraph adjacent to the kink vertex. When no kink vertices are available one is dealing with a linear acene. For a linear acene containing i rings its first dualist vertex will be located on the (i/2)th position. Such first dual vertices are marked with an asterisk in Figure 4 where the corresponding kink vertices are the solid circles. Locating such vertices makes the task of obtaining the label of a particular BH very simple, since it reduces the number of choices by factors of 8 for naphthalene, 10 for anthracene, 12 for tetracene and so on. Naturally, symmetry reduces the job further. Once we locate $v_1(2)$ the ordering of a set of cata-condensed BH's according to several of their molecular properties (such as those listed in Table V) becomes quite easy.

Table I contains decimal contributions, their sums (i.e. labels), Randić's[19] RE values of the five cata-condensed BH's containing four rings.

TABLE I

Decimal Contributions of i *th Row* d_i, *($1 \leqq i \leqq 18$), Their Sums (Labels,* L *Values) Together with Randić's RE's (ev)*[17]*, for Four-ring Catacondensed Benzenoid Hydrocarbons*

d_i i					
1	3	3	3	3	3
2	5	5	5	5	5
3	10	10	10	10	14
4	20	20	25	25	24
5	40	49	48	52	48
6	80	98	84	96	104
7	162	192	192	192	192
8	324	384	352	336	322
9	385	324	384	384	384
10	1536	1536	1536	1536	1536
11	2560	2560	2560	2560	2560
12	5120	7168	7168	7168	7168
13	10240	12288	9216	12288	12288
14	20480	24576	28672	25600	24576
15	40960	33792	49152	49152	53248
16	82944	81920	69632	73728	98304
17	165888	167936	163840	163840	164864
18	*197120*	*204800*	*212992*	*212992*	*196608*
L =	527877	537665	545871	549967	562248
RRE =	1.822	2.330	2.540	2.540	2.723

Table II contains the corresponding information for the twelve cata-condensed BH's containing five rings.

A first look at the values of RE's and L's reveals that the latter index is able to reproduce the order of the former among a group of isomeric BH's excepting isoconjugates[19]: the larger the value of L the more stable is the BH. At this point it may be convenient to distinguish a »cisoid« isoconjugate from its »transoid« counterpart, based on the »configurations« of their dualists. E.g. *8* and *7* (Figure 4) are examples of such cases, respectively. Similarly *16* and *15* would also constitute such pairs. In some instances there might be more than one cisoid counterpart for a particular dualist. A more descriptive nomenclature wolud then be desirable, thus *18* may be called all-transoid *20* all-cisoid while *19* cis-

-transoid (or trans-cisoid). It is observed that conjugated circuits counts[19] do not distinguish pairs of isoconjugate hydrocarbons. (Nor, in fact, available values of Dewar-Dellano RE's[20]). This observation is in good approximation to quantum-chemical results. Our L labels, however, do distinguish isoconjugate pairs: in general structures possessing cisoid dualists have higher L values than the corresponding transoid counterparts. E.g. $L_8 > L_7$; $L_{16} > L_{15}$; $L_{19} > L_{18}$; $L_{20} > L_{18}$; $L_{20} > L_{19}$ (Figure 4, Tables I, II). Discrepancies, however, arise with branched analogues, e.g. $L_{17} > L_{16}$ while RE(17) < RE(16). Table III lists Randić's RE's per pi-electron, RE/e, and a function defined by $(L/2^n)$ for cata-condensed non-linear acenes containing 3—5 hexagons. The factor of 2^n »normalizes« the L values to a scale that is independent of n. One observes that the order of RE/e is reproduced by the function $(L/2^n)$ when the cisoid dualists are excluded from the hierarchy. Figure 5 plots the two functions for all dualists exepting cisoid ones. Table III also gives the corresponding values of $S(3)$ and $v_{\min}(3)$. A clear trend is shown by the latter value: The whole hierarchy spans 3 units starting at a value of 6 for benz[*a*] tetracene and terminating at a value of 3 for triphenylene. The former

TABLE II

Decimal Contributions of d_i*,*$(1 \leqq i \leqq 22)$*, Their Labels,* L*, and RRE's (ev.) for Five-ring Catacondensed Hydrocarbons*

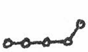
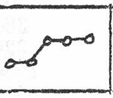
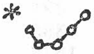
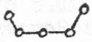
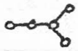
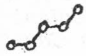
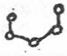
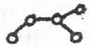
d_i i					* 	
1	3	3	3	3	3	3
2	5	5	5	5	5	5
3	10	10	10	10	10	10
4	20	20	20	20	20	25
5	40	48	49	49	49	48
6	80	100	98	78	98	96
7	160	193	192	192	200	148
8	324	386	384	328	384	416
9	648	768	768	768	768	768
10	1281	1288	1096	1408	1344	1344
11	1538	1536	1664	1536	1536	1536
12	6144	6144	6144	6144	6144	6144
13	10240	10240	10240	10240	10240	10240
14	20480	28672	24576	28672	28672	28672
15	40960	49152	34816	36864	49152	49152
16	81920	98304	102400	114688	102400	69632
17	163840	196608	196608	196608	196608	182992
18	327680	393216	393216	393216	393216	425984
19	663552	528384	528384	540672	557056	786432
20	1327104	1376250	1310720	1310720	1310720	1081344
21	2623488	2637824	2686976	2686976	2686976	2621440
22	*3149824*	*3178496*	*3276800*	*3276800*	*3276800*	*3407872*
$L =$	8419341	8507653	8575169	8606017	*8622401	8674303
RRE =	1.884	2.543	2.705	2.860	2.860	3.046

TABLE II (Cont.)

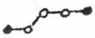
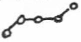
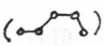
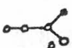
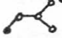
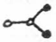
d_i / i	* 			* 	* 	
1	3	3	3	3	3	3
2	5	5	5	5	5	5
3	10	12	10	10	10	10
4	25	25	25	25	25	25
5	48	50	48	48	50	52
6	84	96	84	84	96	96
7	224	192	192	224	192	192
8	384	416	352	384	336	416
9	768	768	768	768	768	768
10	1312	1282	1408	1344	1408	1296
11	1536	1536	1536	1536	1536	1536
12	6144	6144	5144	6144	6144	6144
13	10240	10240	10240	12040	10240	10240
14	28672	28672	28672	28672	28672	28672
15	49152	49152	36864	49152	36864	49152
16	98304	98304	114688	102400	114688	98303
17	167936	212992	147456	163840	196608	212992
18	458752	393216	458752	458752	409600	397312
19	786432	786432	786432	786432	786432	786432
20	1114112	1572864	1114112	1114112	1179648	1179648
21	2621440	2232320	2621440	2621440	2621440	2621440
22	*3407872*	*3407872*	*3407872*	*3407872*	*3407872*	*3407872*
$L =$	*8753455	(8802593)	8737103	*8753487	*8802637	8802607
RRE =	3.046	3.120	3.130	3.130	3.130	3.210

---

* Cisoid duals are not to be compared with transoid counterparts or branched isomers (whose L values are given in parantheses).

TABLE III

Resonance Energy per pi-Electron (in ev. Using Method of Randić[19]), RE/e, Values of Codes Divided by 2^n i.e. $(L/2^n)$, n Being the Number of Rows of the Corresponding Adjacency Matrix, Trivalent Sums, S(3), and Values of $v_{min}(3)$ for Catacondensed Non-Linear Acenes Containing Three to Five Benzene Rings

Dualist		RE/e	$(L/2^n)$	$S(3)^*$	$v_{min}(3)$
11		0.1156	2.0284	108	6
12		0.1241	2.0445	106	5
6		0.1294	2.0510	66	5
13		0.1300	2.0518	103	5
14		0.1300	(2.0557)	102	5
15		0.1385	2.0681	101	4
16		0.1385	(2.0869)	99	4
4		0.14107	2.0790	34	4
7		0.1411	2.0823	62	4
8		0.1411	(2.0979)	61	4
18		0.1417	(2.0831)	98	4
19		0.1417	(2.0870)	97	4
20		0.1417	2.0987	98	4
17		0.1418	2.0987	101	4
21		0.1459	2.0987	98	4
9		0.15127	2.1448	61	3

* Computed on the basis of number assignments shown in Figure 4 Cisoid graphs are shown in parentheses.

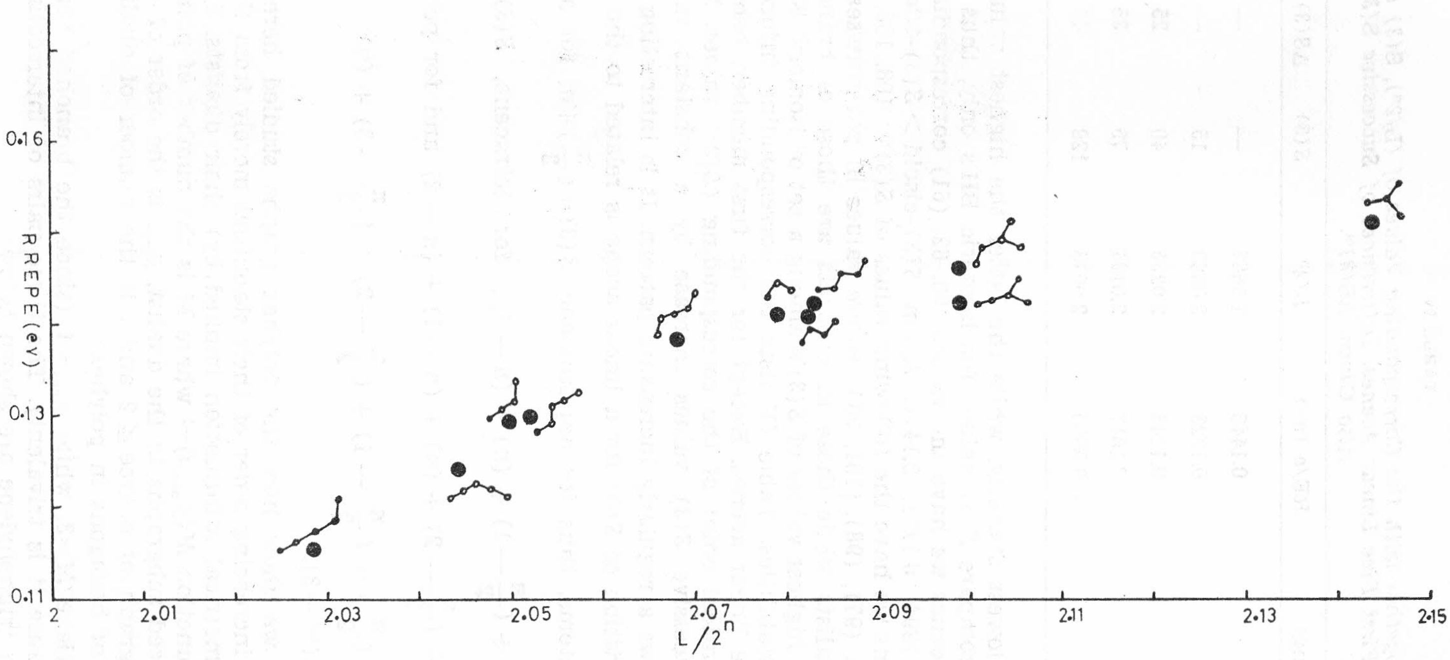


Figure 5. A plot of resonance energies per pi electron,[19] RE/e, (eV), vs. ($L/2^n$) for BH's. The resulting »limited« hierarchy starts with benz [*a*] tetracene and terminates at triphenylene

TABLE IV

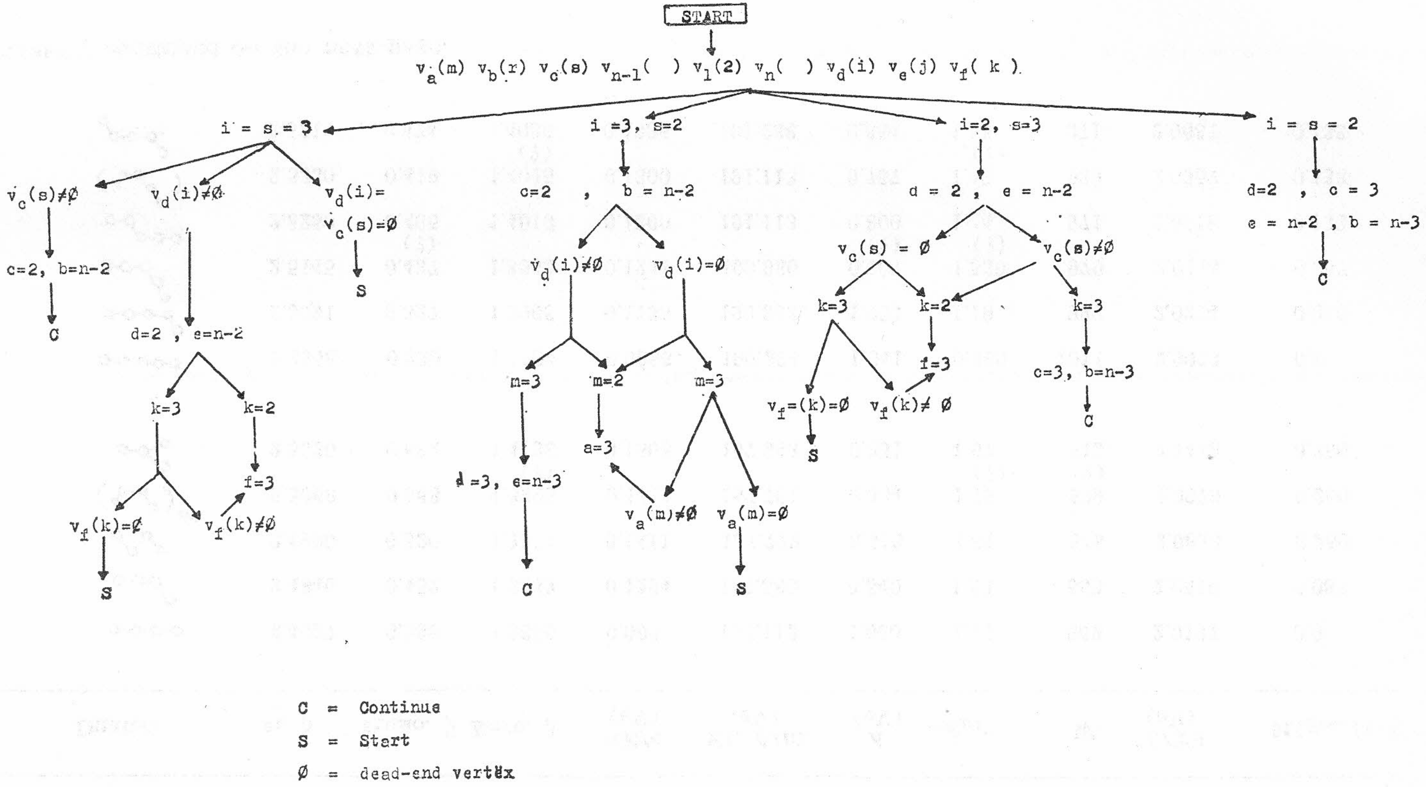
RE/e, (ev) Together with the Corresponding Values of ($L/2^n$), S(3) and $v_{min.}(3)$ for the First Five Linear Acenes. Differences of Successive S(3)'s are Also Given, ΔS(3)'s.

Number of rings	RE/e (ev)	$L/2^n$	$S(3)$	$\Delta S(3)$	$v_{min}(3)$
1	0.1448	1.9687	—	—	—
2	0.1323	2.0303	15	—	5
3	0.1143	2.0235	40	25	6
4	0.1012	2.0137	75	35	7
5	0.0856	2.0073	120	45	8

defines the lowest stability, while the latter the highest value of RE/e. One might compare $S(3)$ values for isomeric BH's only, thus for 4-ring non-linear acenes we have the values: 66, 62, (61) corresponding to RE/e values of 0.1294, 0.1411, 0.1411. Again $S(3)$-cisoid $> S(3)$-transoid. For 5-ring systems we have the following values of $S(3)$'s: 108, 106, 103, (102), 101, (99), 98, (97), (98), {101, 98}, where values in parentheses are those of cisoid dualists while those in brackets are those of branched BH's. Graphs with highest values of $S(3)$'s among a set of isomeric BH's possess the lowest stabilities. Table IV lists the corresponding information for the first five linear acenes. Except for the first member, benzene, RE/e have an identical order of the corresponding ($L/2^n$) values. Differences between successive $S(3)$ values increase by a constant value; while $v_{min}(3)$'s shows a regularly increasing pattern. It is interesting to observe that the partition of $S(3)$ for a linear acene is related to the number of its carbon atoms, thus for naphthalene, $S(3) = (\frac{n}{2}) + n$, for anthracene, $S(3) = (\frac{n}{2}) + (\frac{n}{2} - 1) + (n) + (n - 1)$; for tetracene, $S(3) = (\frac{n}{2}) + (\frac{n}{2} - 1) + (\frac{n}{2} - 2) + (n) + (n - 1) + (n - 2)$ and for pentacene we have $S(3) = (\frac{n}{2}) + (\frac{n}{2} - 1) + (\frac{n}{2} - 2) + (\frac{n}{2} - 3) + (n) + (n - 1) + (n - 2) + (n - 3)$.

In Table V we show how the polyhex graphs studied here might be arranged in increasing order of their stabilities merely from the topology of (and geometrical information implied by) their dualists. For this we define the function $M(g_{max}.i)^{-1}$ where M is the number of pairs of intersecting induced subgraphs in the dualist, g_{max} is the order of the longest induced subgraph of degree ≤ 2 and i is the number of dualist vertices (= number of hexagons in polyhex).

For triphenylene $M=3$, while $g_{max}=1$ (since the branched vertex is not counted because it is trivalent). The three pairs of intersecting induced subgraphs in triphenylene are shown below:



FLOW CHART

Figure 6. Flow-chart leading to sequence labels of vertices of polyhex graphs

TABLE V

Molecular Properties of BH's Including: Eigenvalue of Lowest Occupied,[23] ε_1, *Lowest Unoccupied,*[22] $\varepsilon_{\mathrm{LUMO}}$ *MO's (Both in Units of B), Resonance Energies per Electron,*[19] *RE/e(eV), Total pi-Energies per Election,*[22] $E\pi/e$ *(in B Units), Heats of Atomization,*[22] *Ht Atm. (eV), Ionization Potentials,*[22] *A, (eV), and Polarographic Half-wave Reduction Potentials*[22] $E_{1/2}^{\mathrm{red}}$ *(eV). The Functions* $(L/2^n)$ *and* $M(g_{\max}i)-1$ *and Wiener Numbers,*[23]*W, of BH's are Listed.*

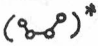
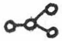
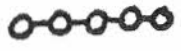
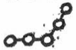
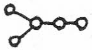
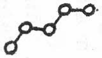
No	Dualist	ε_1, β	$\varepsilon_{\mathrm{LUMO}}$, β	$E\pi/e$, β	RE/e (eV)	Ht. Atm. (eV)	*A* (eV)	$-E_{1/2}^{\mathrm{red}}$	*W*	$(L/2^n)$ (eV)	$M(g_{\max}i)-1$
5		2.4667	0.295	1.3850	0.099	157.112	1.060	1.14	569	2.0137	0.0
6		2.4846	0.452	1.3889	0.1294	157.580	0.640	1.53	553	2.0510	0.083
7		2.4990	0.520	1.3994	0.1411	157.772	0.516	1.81	545	2.0823	0.250
8	**	2.5069	0.568	1.3993 (?)	0.1411	157.781	0.384	1.75 (?)	529 (?)	2.0979	0.250
9		2.5320	0.684	1.4138	0.1505	157.943	0.251	1.97	512	2.1448	0.750
10		2.4955	0.220	1.3884	0.0856	190.294	1.341	0.860	1011	2.0073	0.0
11		2.5091	0.327	1.3966	0.1156	190.853	1.003	1.19	987	2.0284	0.050
12		2.5145	0.437 (?)	1.3983	0.1241	190.980	0.707 (?)	1.530 (?)	979	2.0445	0.067
13		2.5250	0.405	1.4018	0.1300	191.113	0.800	1.44	971	2.0518	0.133
14		2.5330	0.419	1.4015 (?)	0.1300	191.113	0.757	1.40 (?)	939	2.0557	0.133
15		2.5211	0.474	1.4036	0.1385	191.238	0.654	1.55	971	2.0681	0.133

TABLE V continued on the next page.

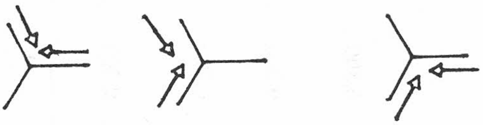
TABLE V contin.

18		2.5347	0.502	1.4036	0.1417	191.361	0.500 (?)	1.79	963	2.0831	0.300
19	()	2.5420	0.550	1.4063	0.1423	191.361	0.591	1.73	931	2.0870	0.300
20	()	2.5510	0.535 (?)	1.4062 (?)	0.1423				899	2.0987	0.300
17		2.5588	0.499 (?)	1.4064	0.1418 (?)	191.348 (?)	0.568 (?)	1.54 (?)	907	2.0987	0.300

---

* Numbers correspond to polyhex graphs shown in Figure 4.
Question marks indicate discrepancies.

** Cisoid dualists are placed in parentheses. They often show discrepancies.



Other examples might be inferred from Table V. However the function $M(g_{\max}i)^{-1}$ is independent of configuration and therefore does not distinguish cisoid from transoid dualists. Nevertheless the function is an illustration of data reduction where the dualist replaces the molecular graph of the hydrocarbon. Data reduction is getting to be quite essential in complex data-analyses[21]. Table V also includes other molecular properties[22] including total pi-energies per electron, eigenvalues of lowest vacant and lowest occupied MO's, ionization potentials of the hydrocarbons, their heats of atomization and polarographic half-wave reduction potentials. It also includes a structural property, the Wiener numbers[23] of the topological matrices describing polyhex graphs. We observe from Table V that the order of molecular properties follows the orders of the functions $L/2^n$ of the hydrocarbon or $M(g_{\max}i)^{-1}$ of dualists for a set of isomeric polyhexes. We may conclude that, despite a few discrepancies, our purely structural label, L, succeeds in predicting the order of molecular and electrochemical properties of a set of isomeric cata-condensed BH's and that further work in this direction seems worth pursuing.

Acknowledgements. — I thank both referees for critical remarks.

REFERENCES

1. M. Randić, *J. Chem. Phys.* **60** (1974) 3920.
2. A.L., Mackay, *J. Chem. Phys.* **62** (1975) 308.
3. M. Randić, *J. Chem. Phys.* **62** (1975) 309.
4. M. Randić, *J. Chem. Inf. Comput. Sci.* **17** (1977) 171.
5. M. Randić, *Chem. Phys. Letters,* **42** (1976) 283.
6. M. Randić, *Internat. J. Quantum Chem.* **15** (1979) 663.
7. M. Randić, *Croat. Chem. Acta,* **49** (1977) 643.
8. W.C. Herndon, *Internat. J. Quantum Chem. Quantum Biol. Symp.* **1** (1974) 123.
9. L.B. Kier, L.H. Hall, W.J. Murray and M. Randić, *J. Pharm. Sci.* **64** (1975) 1971.
10. D. Bonchev and N. Trinajstić, *J. Chem. Phys.* **67** (1977) 4517.
11. M.J.S. Dewar and H.C. Longuet-Hnggins, *Proc. Roy. Soc.* **A-214** (1952) 482.
12. D. Cvetković, I. Gutman and N. Trinajstić, *J. Chem. Phys.* **61** (1974) 2700.
13. The term *polyhex* graph, used to describe the molecular graph of a BH, is introduced in: N. Ohkami, A. Motoyama, T. Yamaguchi, H. Hosoya and I. Gutman, *Tetrahedron,* **37** (1981) 1113.
14. A.T. Balaban and F. Harary, *Tetrahedron* **24** (1968) 2505, A.T. Balaban, *Rev. Roum. Chim* **22** (1977) 45.
15. I. Gutman, *Theoret. Chim. Acta (Berlin)* **45** (1977) 309.
16. F. Harary, *Graph Theory,* Addison-Wesley, Reading (1969) Chapter 8.
17. S. El-Basil, *Croat. Chem. Acta* (1983) **57** in press.
18. The contributions (decimal or binary) of the successive rows of an arranged matrix are said to define an ordered string, See e.g. Ju. A. Schreider, *Equality, Resemblance and Order,* Mir Publishers Moscow (1975) p. 122—123.

19. M. Randić, *Chem. Phys. Letters* **38** (1976) 68; *Tetrahedron* **33** (1977) 1905; *J. Amer. Chem. Soc.* **99** (1977) 444; *Internat. J. Quantum Chem.* **17** (1980) 549.
20. M.J.S. Dewar and C. DeLlano, *J. Amer. Chem. Soc.* **91** (1969) 789.
21. M. Randić, and C.L. Wilkins, *J. Phys. Chem.* **83** (1979) 1537; A.T. Balaban, *Pure and Appl. Chem.* **54** (1982) 1075.
22. Molecular properties are taken from: M. J. S. Dewar, J. A. Hashmall and N. Trinajstić, *J. Amer. Chem. Soc.* **92** (1970) 5555.
23. Wiener numbers are taken from: D. Bonchev, Ov. Mekenyan and N. Trinajstić, *Internat. J. Quantum Chem.* **17** (1980) 845.

SAŽETAK

Molekularna topologija i molekularna svojstva. I. Sekvencijske oznake kata-kondenziranih benzenoidnih ugljikovodika koji sadrže do pet prstenova

S. El-Basil

Opisan je algoritam za sistematsko numeriranje benzenoidnih ugljikovodika. Suma decimalnih doprinosa svih čvorova (ugljikovih atoma) u molekularnom grafu naziva se oznakom ugljikovodika. Ove oznake, zasnovane isključivo na strukturnim elementima, uređuju skup izomernih benzenoidnih ugljikovodika prema njihovim svojstvima uključujući: rezonantne i ukupne pi-energije, ionizacijske potencijale, topline atomiziranja, energije najniže prazne i najniže popunjene molekularne orbitale i polarografske poluvalne redukcijske potencijale. Ove oznake ugljikovodika mogu se lako izračunati bez upotrebe računala.