Theor Chim Acta (1988) 73:425–435

Theoretica Chimica Acta

Molecules with holes*

George G. Hall

Division of Molecular Engineering, Kyoto University, Kyoto 606, Japan

(Received May 6, revised June 30, Accepted July 6, 1987)

The *bual* and *laub* operations are defined and shown to be a useful means of analysing and coordinating previous descriptions of polyhexes. Some implications of a polyhex having a hole are explored. The position of such molecules in the Dias periodic table is argued. Formulae for moments of the Hückel energy in terms of graphical invariants are derived. The change in π-electron energy when a hole is formed is calculated for some molecules.

Key words: Graph theory — Polycyclic hydrocarbons — Holes in molecules

1. Introduction

As the size of a polycyclic hydrocarbon molecule increases so it becomes capable of exhibiting greater complexity. Since the number of isomers also increases it becomes necessary to devise methods of characterizing these molecules in a meaningful way. The periodic table of Dias [1] is one very helpful device for this purpose. He has also shown [2] that only a small fraction of the possible isomers has ever been synthesised. This suggests that theoretical methods of distinguishing isomers and of suggesting which molecules would be of particular interest to prepare would be important. Even Hückel theory can serve for this purpose.

Since its introduction in 1968 by Balaban and Harary [3] (but see also Smith [4]) the concept of the *characteristic graph* has been used as a means of describing large polycylic hydrocarbons. For further references see the recent review by Balaban [5]. The term is cumbersome and other authors have called it the dual

---

* Dedicated to Professor J. Koutecký on the occasion of his 65th birthday

or the dualist graph. The author has argued[1] in favour of the term *bual* [6] which is distinctive and meaningful. Section 2 of this paper enlarges the concept of the bual to make it a more useful tool of analysis and codification.

The idea of a conjugated hydrocarbon molecule containing a hole seems to be due to McWeeny [7] but the first synthesis, of Kekulene, was by Diederich and Staab [8] 27 years later. There has been little discussion of the properties to be expected of these molecules although some interesting comments on the metal complexes formed when an inner CH is replaced by N have been made by Balaban [5]. In particular most theoretical discussion of polycyclic hydrocarbons has ignored the possibility of holes. Many formulae need to be amended to allow for this and Sect. 4 and 5 give some examples. Before undertaking the preparation of molecules with holes it will be of importance to know what novel properties are to be expected. Sections 6 begins this more general discussion by considering the loss of π-electron energy when a hole is created.

2. *Bual* and *laub*

It is convenient to begin by defining rather carefully a certain class P of polycyclic molecule (P for polyhex) and restrict most of the discussion to this class. A member M of P will be a flat conjugated hydrocarbon molecule or radical such that:

(a) the C framework of M consists of connected regular hexagonal rings of the same size;
(b) every CC bond is an edge of at least one hexagon;
(c) hexagons are connected only by sharing edges.

The condition (b) excludes molecules with polyene side chains and (c) excludes bridges (cut edges) such as the central bond in biphenyl. The restriction to flat molecules and to regular hexagons is intended to emphasize that this is a theoretical model describing the π-electron system. Before comparisons can be made with experimental results the effects of overcrowding and of bond length differences need to be considered. For some of these molecules these effects are very significant. A discussion of "flat" in this connection has been given by Elk [9].

When a graph, G, consists of regular hexagons of the same size its *bual* is defined by the following construction. A vertex of the *bual* $B(G)$ is placed at the centre of every hexagon of G and an edge of B connects vertices whose hexagons share

---

[1] The arguments against these terms are: (1) The *bual* differs from the dual in ignoring the outside region. (2) The *bual* has a geometric meaning as well as a graphical one. Its angles are essential for its use in counting isomers. (3) Sometimes both concepts are needed. Having two terms makes such a discussion much easier to understand. *Bual* can be considered as a bastard *dual* to acknowledge its origin but emphasise its distinctness. It is, of course, the name of a sweet wine

a common edge. The edge of B bisects this edge of G orthogonally. Thus the *bual* is partly a geometric concept and partly a graph theoretic one. This definition can be generalized to square graphs G by replacing hexagon by square throughout. Below it is applied to triangular graphs with equilateral triangle replacing hexagon. It is clear from this that every benzenoid molecule has a unique *bual* but a bridge in M will lead to isolated vertices in B and polyene sections of M will be ignored. The condition, that rings should be connected but only by sharing edges, in the definition of P ensures that the *bual* $B(M)$, for M in P, will have no isolated vertex and that $B(M)$ will be a connected graph.

A molecule of class P will have a *hole* when there are at least two CH bonds which are not on the outside of the molecule. The shape of the hole is limited by the definition of P to being a closed polygon formed from hexagonal sections. This means that its boundary will be the same as the boundary of some molecule also in P and this gives a convenient label for the hole. The number of holes in the molecules will be called the *genus* of the molecule. Balaban and Harary [3] have used the term *coronafusenes* for molecules with holes. In forming the *bual* of a molecule with a hole, the hole will not be a hexagon and so will not have a vertex. If follows from the definition of P that the molecular hole of M will be surrounded by complete hexagons so the corresponding boundary of $B(M)$ will be continuous and will outline the same holes, i.e. the genus of $B(M)$ will be the same as that of M. The genus of M will be denoted by $g(M)$. In his paper Dias [1] denoted it as c.

From any given *bual* B a unique molecular graph can be constructued by placing a regular hexagon around every vertex of B so that its edges bisect orthogonally the edges of B joined to that vertex and coincide with the edges of the hexagons centred on connecting vertices. For M in P this operation will reproduce the original M from $B(M)$. Since this is the reverse of the bual operation it can be called the *laub* $L(B)$ of the bual B. For a benzenoid molecule not in P there is still a *bual* but, if it contains a bridge, B will have several disconnected parts and its *laub* cannot reproduce the original molecule since bonds not in a hexagon are not recreated. L has no prescription for relating disconnected portions of B. Similarly polyene side chains are not represented in the *bual* and cannot be produced by the *laub*. The significance of the class of polyhexes P is that, since $L(B(M)) = M$, there is a 1:1 relation between M and $B(M)$. This relation is essential to the use that various authors have made of the *bual* as a means to enumerate these hydrocarbons.

The *bual* $B(M)$ will have vertices and edges and these may form rings. If so, these will be part of the equilateral triangular lattice, which is the dual of the original hexagonal lattice. The *bual* operation can be applied in turn to $B(M)$ with the triangles replacing the hexagons. In every triangle a vertex of the *dibual* is placed and connecting edges bisect shared edges of the triangles. The *dibual* is then part of the dual of the triangular lartice, which is the original hexagonal lattice. For the graphical dual, the repetition of the dual operation must bring back the original exactly. This is the reason for the name dual. In fact the *dibual* will be contained in the original molecular graph. Since the *bual* has no vertex outside the molecule and so no edges crossing the molecular boundary its *dibual* has the boundary stripped off leaving what Dias calls the excised internal structure [10] of the molecule. Thus, by subtraction, the *dibual* also gives a rigorous definition of the boundary. Since the *bual* has fewer vertices its repeated operation will lead to a vanishing result. When m is the smallest integer for which $B^m(M) = \emptyset$ then m can be called the level of the molecule M. The catacondensed molecules are conveniently defined as molecules in P with level 2, i.e. vanishing *dibuals*. Because M is in P, $B(M)$ will be a connected graph but, in general, $B^2(M)$ will not. It may also have polyene portions either as side chains or bridges. If M contains polyacene "side chains" or "bridges" then the *bual* has linear sections which are lost in the *dibual*. Many molecules may share the same *dibual*.

The *laub* operation can also be repeated. The *dilaub* of M, $L^2(M)$, is a graph with extra hexagonal rings added to every bond on the boundary of M. It represents a growth process. For the reasons above, the *dilaub* will not generally reproduce M from $B^2(M)$. It will produce a part of M with the less rounded parts removed. It could also be called the *circum* operation since a molecule such as circumnaphthalene (ovalene) is constructured from naphthalene in this way. Since $L^2(M)$ may fill up a hole in M the genus of $L^2(M)$ is not generally equal to that of M.

There is a useful analogy between these operations and the operations of differentiation and integration. The repeated *bual* loses more and more of the original graph just as repeated differentiation loses more and more of a polynomial function. To recover the polynomial requires constants of integration to be added at each integration. The *laub* would generally require analogous "constants of integration" to become the inverse operation to the bual. For the class P the inverse relation of $B = B(M)$ and $M = L(B)$ is rather special.

3. The Dias periodic table

Dias [1] has constructed his periodic table as a means of organizing the polyhexes into useful groups. Isomers are classed together and the chemical formula determines the group. For a molecule with fomula C_nH_h he defines its column number d_s in the table by

$$d_s = (3h - 14 - n)/2 \tag{1}$$

and its row number by N_{Ic} where

$$N_{Ic} = n - 2h + 6. \tag{2}$$

The first row in the table consists of the *catacondensed* molecules. In subsequent rows the molecule most to the left is purely *pericondensed*. Properties relating molecules in the same row or the same column have been discussed.

The significance of this table is made more apparent by using the *bual*. The number, $r(M)$, of benzene rings in the molecule M, i.e. the number of vertices in the *bual*, is an obvious variable which can be deduced from the graph easily. The number of triangles in the *bual* , $R(B)$, is also readily deduced from the graph. These two variables give another interpretation of the table since R is an alternative definition of one of Dias' variables and the other is simply related to it and r:

$$N_{Ic} = R, \; d_s = r - R - 2. \tag{3}$$

Thus the top row has $R = 0$ and its columns, moving to the right, correspond to increasing numbers of benzene rings. The triangles in the *bual* can point up or down and it is readily shown that, for M to be a molecule rather than a radical, the number up must match the number down. Thus the second row of the table has $R = 2$. The later rows have R, the number of triangles, increasing by 2 each time. The *buals* of the molecules along a row have the same fixed number of triangles (R) but add a vertex and an edge for every column. The isomers correspond to the various ways of doing this.

Unfortunately, these formulae assume that the molecule has no holes. Molecules with holes can be brought into the table in various ways. They could be classed according to their chemical formula but this would bring together molecules with different chemical behaviour and some of the advantage of the table would be lost. The table could be made three dimensional with the genus as the new variable but this seems unnecessary. It seems more logical to class them with other molecules having the same number of rings i.e. r and R. If the molecule $M = C_nH_h$ has $N(M)$ edges (CC bonds) and genus $g(M)$ then the Euler relation between these is

$$r + g = N - n + 1. \tag{4}$$

This equation means that if N, n and r are taken as variables then g is not needed explicitly. Since every C is connected to three atoms and every edge (CC bond) connects two C and there are h CH bonds we have

$$3n = 2N + h \tag{5}$$

so that, in terms of the formula and g, the number of rings is

$$r = (n - h)/2 + 1 - g. \tag{6}$$

Similarly, the *bual* has r vertices, R triangles and E edges so its Euler equation gives

$$R + g = E - r + 1. \tag{7}$$

An edge of B bisects an internal edge of M so $(E + N)$ counts the internal edges twice and the external edges once. Since hexagons have six edges this must add to six times the number of hexagons

i.e.

$$E + N = 6r. \tag{8}$$

With these equations the number of triangles is

$$R = n - 2h + 6(1 - g) \tag{9}$$

so that Dias' relations (1) and (2) (but see also [1]) are generalized to

$$d_s = (3h - n)/2 - 7 + 5g, \; N_{Ic} = n - 2h + 6(1 - g). \qquad (10)$$

Thus the variable g, or the equivalent use of N, n and r, is needed for a more complete classification.

4. Holes

Almost all discussions of benzenoid hydrocarbons have assumed that the molecule has only one boundary - on the outside of the molecule. The nature of the boundary of a hole now needs to be investigated.

The interior boundary of a molecule has some properties which are quite different from those of the outside boundary. For example on the outer boundary the numbers of C atoms with degree two, v_2,[2] and of degree three, v_3, satisfy the relation (see Sachs [10] for a geometric proof using the shape of the polygon)

$$v_2 - v_3 = 6. \qquad (11)$$

For each inner boundary the atoms satisfy

$$v_3 - v_2 = 6. \qquad (12)$$

The proof is exactly the same. The difference arises because the angle of $\pi/3$ inside the polygon is found at a (2) atom on the outside boundary and at a (3) atom on the inside boundary and there is a similar interchange for the $-\pi/3$ angle. This is sufficient to show that holes can introduce new features into molecules. In terms of $g(M)$ the correct generalization of Sachs' equation is:

$$v_3 - v_2 = 6(g(M) - 1), \qquad (13)$$

where v_2 and v_3 now refer to the whole molecular boundary.

5. Graphical invariants

In a recent paper [12] it has been shown that some moments of the energy in the Hückel theory, i.e.

$$\mu_n = \operatorname{tr} A^n, \qquad (14)$$

where A is the Hückel matrix, can be conveniently expressed in terms of a limited number of graphical invariants e.g. $\mu_2 = 2N$. The simplest invariants of M are $n(M)$, $N(M)$ and $r(M)$ but others are required for higher moments. The next simplest is the number of bays $b(M)$, which is defined as the number of (3, 3) bonds on the boundary of M (Note that this definition is different from that given by Balaban [5]). This definition will now be extended to include all (3, 3) bonds in the boundaries inside as well as outside.

---

[2] It is convenient to use the notation (n) for a C atom of degree n, (m, n) for a CC bond with atoms of degree m and n at its ends and $\#$ as a number operator so that, e.g. on the boundary $v_2 = \#(2)$, $v_3 = \#(3)$

The fourth moment μ_4 is calculated from A^2 by squaring each element and adding. The diagonal elements of $\boldsymbol{A}^2$ are the degrees of the atoms and $h = \#(2)$ and $(n-h) = \#(3)$. The off-diagonal elements are 1 for second neighbours in M and 0 otherwise. As was shown in [11], the number of open paths of length two is equal to the number of subtended angles at the different C atoms. Each (3) atom has three paths with it as centre and each (2) atom has one. Since each can be traversed in either order these contribute twice to μ_4. Thus

$$\mu_4 = 9(n-h) + 4h + 2(3(n-h)+h) = 15n - 9h = 18N - 12n. \tag{15}$$

This is exactly the same relation as before [12] and suggests that these invariants are appropriate ones for this purpose. In particular the genus, the rings and the bays are not required for this moment.

In order to calculate the sixth moment, the number of bonds in various situations is needed. We define $c = \#(2,2)$, $f = \#(2,3)$ and $l = \#(3,3)$. Since the total number of bonds is N then

$$N = l + c + f. \tag{16}$$

The (2) atoms have two connecting CC bonds and these must be either (2, 2) or (2, 3) so, to agree with the number counted by c and f,

$$2h = 2c + f. \tag{17}$$

Equations (16) and (17) determine c and f in terms of the other variables

$$c = 2h + l - N, \quad f = 2N - 2l - 2h. \tag{18}$$

From the discussion above, the number of internal (3, 3) bonds equals E, the number of edges in B, since each bisects the other. The external (3, 3) bonds are the bays. So the variable l is the sum of these variables

$$l = E + b = 6r - N + b. \tag{19}$$

In [12] it was shown that it was better to select l as the basic variable instead of b since this extended the scope of the equations. For polyhexes either b or l can be used and b is more readily derived from the graph of M.

The calculation of μ_6 now proceeds by using the matrix $\boldsymbol{B}$ which is defined [13] by

$$\boldsymbol{A} = \begin{pmatrix} 0 & \boldsymbol{B} \\ \boldsymbol{B}^T & 0 \end{pmatrix} \tag{20}$$

since the polyhexes are alternant. The trace becomes

$$\mu_6 = 2 \operatorname{tr} (\boldsymbol{B}\boldsymbol{B}^T\boldsymbol{B})^2 \tag{21}$$

so every element of the product is squared and summed. The elements are classified in terms of the elements of $\boldsymbol{B}$. Each of the (3, 3) bonds has an element of 5, each (2, 3) has 4 and each (2, 2) has 3. Elements with 2 represent the two paths of length three around the hexagons and each hexagon has three. Elements

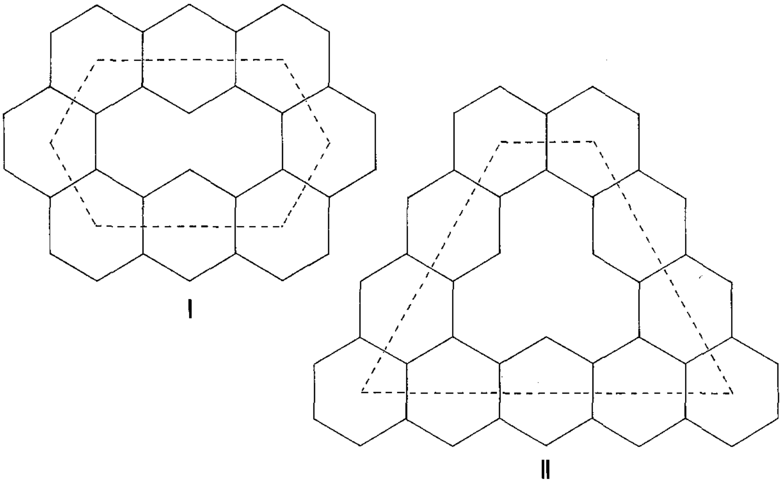

Fig. 1. Polyhexes and holes. *Buals* shown *dotted*

of l occur across each bay and also as zig-zag paths whose central bonds are (3, 3), which have two, and (2, 3), which have one. Thus

$$\begin{aligned}\mu_6/2 &= 25l+16f+9c+4(3r)+(b+2l+f)\\ &= 25(l)+16(2N-2l-2h)+9(2h+l-N)+12r\\ &\quad +(b+2l+(2N-2l-2h))\\ &= 58N-48n+3l+6r\\ &= 55N-48n+24r+3b. \qquad (22)\end{aligned}$$

This has exactly the same form as before [12] and this justifies the choice of variable. The meaning has, however, been more extended. Although the genus does not appear it is involved in r, N and n through Eq. (4). By keeping this form the result is homogeneous in the variables whereas using g would introduce constants into the equations.

6. The formation of holes

As a contribution to the systematic study of holes in large polyhex, some calculations have been made, using Hückel theory, on the change in π-electon energies when a hole is formed. Since the σ-electron energies are also changed, these are not expected to reflect experimental energies of formation. They are merely indications about how the new structure modifies the molecule.

Figure 1 shows two molecules with holes. I has a hole whose boundary is that of naphthalene. This is the simplest molecule with a hole since it has only one

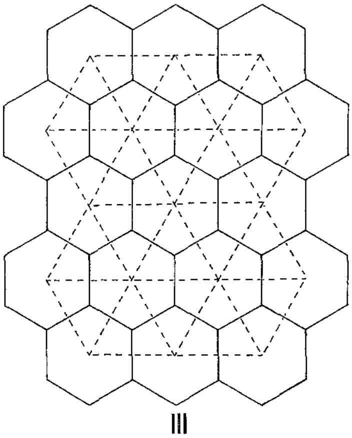

Fig. 2. Circum-perylene $C_{50}H_{18}$ (III). *Bual* shown *dotted*

missing bond. It should be noted that any molecule with a catacondensed hole will have internal CH bonds which interfere with each other strongly and must distort the structure. II has a hole like the phenalenyl radical and has a missing C atom as well as three missing CC bonds. Since these are relatively small molecules and may not give typical results, a large molecule III (Fig. 2) has also been considered and holes of various sizes and shapes formed in it. Figure 3 gives the *buals* of these modified molecules. For example, A has one naphthalene hole while B has two and the energy to create these was found. Table 1 shows the calculated results. The numbers are in units of β.

Table 1 suggests that the π-electron cost of removing one bond from a molecule is nearly constant and requires 0.64. Removing one C with its three bonds is more expensive and in the range 2.4–2.8. Removing two bonds to create a phenanthrene hole costs more than two isolated naphthalene holes and the anthracene hole is even more expensive. On the other hand, as the HOMO energy shows, the anthracene hole in circumperylene leads to a rather reactive molecule with an exceptionally low energy gap.

Table 1. HMO energy differences (units of β)

Molecule			Hole shape	Loss of energy	Total π-energy	e_{HOMO}
I		$C_{32}H_{16}$	Naphthalene	0.6397	45.8578	0.4016
II		$C_{42}H_{20}$	Phenalenyl	2.4110	59.7192	0.2984[a]
III	A	$C_{50}H_{20}$	1st naphth.	0.6411 (A-III)	73.0963	0.3436
	B	$C_{50}H_{22}$	2nd naphth.	0.6310 (B-A)	72.4623	0.3912
	C	$C_{49}H_{21}$	Phenalenyl	2.7630 (C-III)	70.9744	0.4229[a]
	D	$C_{48}H_{24}$	Perylene	4.9150 (D-III)	68.8223	0.4141
	E	$C_{50}H_{22}$	Phenanthrene	1.2990 (E-III)	72.4384	0.2299
	F	$C_{50}H_{22}$	Anthracene	1.5480 (F-III)	72.1894	0.1064

[a] Next to the non-bonding orbital

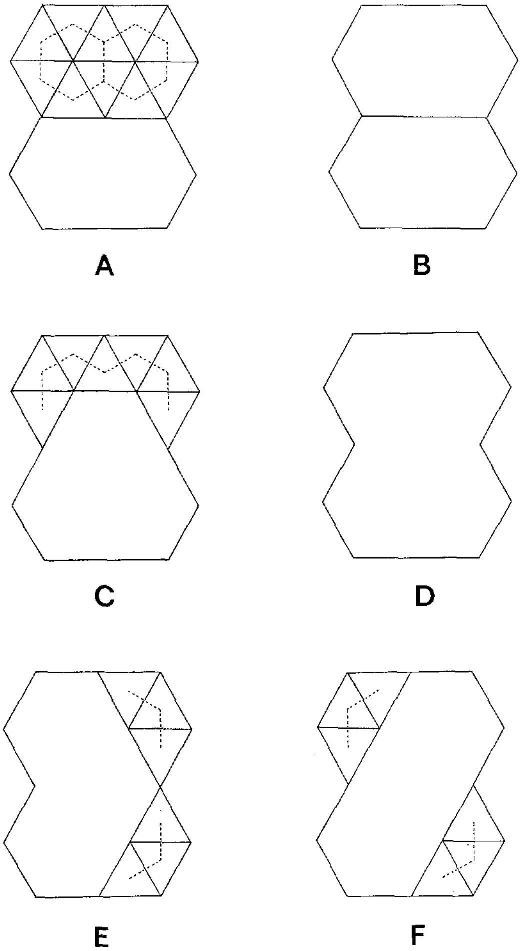

Fig. 3. *Buals* of circum-perylene with various shaped holes. The *dibual* shown *dotted*

Appendix

The formlae derived in Sect. 5 enable us to give an estimate of the spectral radius of the matrix $\boldsymbol{A}$ in terms of graphical invariants. The spectral radius has been discussed recently by Cioslowski [14]. This radius is also the magnitude of the largest eigenvalue of $\boldsymbol{A}$. An estimate of this eigenvalue was given earlier [15] as

$$e_1 \geqslant 2\left(\sum_{ij\,\text{bonded}} d_i d_j\right)\Big/\sum_i d_i^2 = W/S \tag{23}$$

where d_i is the degree of atom i. These sums can be evaluated using the relations above

$$\begin{aligned} S &= 4h + 9(n-h) = 9n - 5h \\ &= 10N - 6n, \end{aligned} \tag{24}$$

$$\begin{aligned} W &= 2(9l + 6f + 4c) \\ &= 24n - 16h + 2l = 30N - 24n + 12r + 2b. \end{aligned} \tag{25}$$

The estimate W/S shows clearly how this quantity is dominated by the short range topology of the molecule. Although this estimate is not quite as accurate as that of Cioslowski it is simpler to use and appreciate. This ratio was originally used as a measure of branching. It is interesting that it involves the number of rings and bays.

References

1. Dias JR (1982) J Chem Inf Comput Sci 22:15, 139
2. Dias JR (1987) Handbook of polycyclic hydrocarbons. Elsevier, Amsterdam
3. Balaban AT, Harary F (1968) Tetrahedron 24:2405
4. Smith FT (1961) J Chem Phys 34:793
5. Balaban AT (1982) Pure and App Chem 54:1075
6. Hall GG (1973) Int J Math Ed Sci Tech 4:233
7. McWeeny R (1951) Proc Phys Soc A64:261, 921
8. Diederich F, Staab HA (1978) Angew Chem 90:383
9. Elk SB (1985) J Chem Inf Comput Sci 25:17
10. Dias JR (1984) Can J Chem 62:2914
11. Sachs H (1983) Combinatorica 4:89
12. Hall GG (1986) Theor Chim Acta 70:323
13. Hall GG (1952) Proc Roy Soc A213:102
14. Cioslowski J (1987) Theor Chim Acta 70:443
15. Hall GG (1977) Mol Phys 33:551