

REFINEMENT OF SYSTEMATIC NOMENCLATURE FOR POLYBENZENES
AND ITS EXTENSION TO SYSTEMS OF GENERAL ARENES

Seymour B. Elk
Organisch chemisches Institut
Universitaet Zuerich

on leave from:

Mathematics Department
New Jersey Institute of Technology

(Received: February 19, 1982)

ABSTRACT:

The methods used for the creation of a systematic nomenclature for polybenzenes, described in <1>, are re-examined. The relationship to the presently used IUPAC nomenclature, through the Synthesis Algorithm is mathematically strengthened by the incorporation of graph theoretical analysis techniques, supplementing the more intuitive geometrical foundation using hexagonal tessellations. The inconsistency problems inherent in IUPAC nomenclature are further clarified. The mathematically consistent Analysis Algorithm is greatly expanded in order to include all presently known 'incidence three' compounds; consequently, the entire table of 35 'basic' structures to be used in formulating the internationally accepted name for most ring compounds may be discarded, and replaced with a simple location sequencing scheme. By means of this extended Analysis Algorithm, conjugated ring structures having any size convex rings, either planar or extending into three dimensions (i.e., helicenes and helicenic analogues), may be canonically named. Problems inherent in concave ring systems are described and correction factors introduced in order to assist in canonically naming such structures.

1. INTRODUCTION

In reference <1> three rules for naming polybenzenes were presented. The foundation upon which these rules is built is the hexagonal tessellation of the plane. Contiguous paths and clusters in such a gridwork are then identified and named. In each case the final name consists of three parts:

- a. An ordered sequence of location designators
- b. A Greek prefix, describing the number of rings
- c. The word "benzene", or "benzocluster" for the third rule

Algorithms are required ONLY for part (a).

The first of these algorithms, called the "Synthesis Algorithm", has as its ONLY purpose to, as nearly as possible, literally translate the IUPAC Rules into a working algorithm. However, because there exist inherent inadequacies (which are described below) in some of the pertinent IUPAC Rules, this Synthesis Algorithm is limited in its capabilities and can NOT be applied without creating either ambiguous or contradictory names for various polybenzene systems. In other words, as a consistent, reliable means of producing unambiguous systematic names, this method fails; however, for smaller polybenzenes very few inconsistencies will be encountered.

Because mathematically any algorithm that can give either contradictory or ambiguous names is untenable, a better method is required. Thus, the second algorithm, the

"Analysis Algorithm", was created. This algorithm is based on a different sequencing technique for location designators (than in the Synthesis Algorithm), and requires only lexicographic ordering. In this way it is readily suitable for computer manipulation. Notice that this algorithm contains no convenient pictures and also does not use the "Aufbau" principle; therefore, parents of a given polybenzene are NOT as readily ascertainable as in the more intuitive Synthesis Algorithm.

Although this Analysis Algorithm could be made applicable to any planar fusion of benzene rings using the standard topology technique of introducing a "cut", whenever a multiply-connected set of rings is encountered the computer time involved increases dramatically. Therefore, a third method of systematic naming, called the "Cluster Method", was given. This method first examines each of the six possible orientations that exist, due to the six-fold symmetry of the hexagonal tessellation of the plane. This is then followed by a simple counting technique. Such a system, however, gives very unwieldy names for primarily linear strings of hexagons and is recommended only for clusters and for multiply-connected sets of benzene rings, especially those with high symmetry. Furthermore, because the names produced by the Cluster Method are dependent on pictorial representation, they also suffer from many of the same problems associated with the Synthesis Algorithm.

2. REFINEMENT OF THE SYNTHESIS ALGORITHM FOR POLYBENZENES

Turning now to consideration of the Synthesis Algorithm, the first relevant IUPAC Rule is <2>:

Rule A21.3 - "Ortho-fused" or "ortho- and peri-fused" polycyclic hydrocarbons with maximum number of non-cumulative double bonds which contain at least two rings of five or more members and which have no accepted trivial name such as those of Part .1 of this rule, are named by prefixing to the name of a component ring or ring system (the base component) designations of the other components. The base component should contain as many rings as possible from the beginning of the list of Rule A21.2. The attached components should be as simple as possible.

Since the Synthesis Algorithm uses only one type of module, the benzene ring, rather than 21 (benzene, plus 20 polybenzenes listed in IUPAC's Rule A21.2 -- see Table 1 below), implementation of this algorithm reduces to annelling a single benzene ring onto a previously named polybenzene of $n - 1$ rings. Thus, it follows that the class of new common names needed to be memorized, in order to produce a canonical name for a given compound, is reduced from 20 to ZERO. As an example, the compound which IUPAC has given the name picene (#22 in Table 1) would have the synthesis name: FDB-Pentabenzene.

The Synthesis Algorithm, with some minor revisions from the version presented in <1>, is next re-stated with detailed explanations and comments.*

* In order that the relevant comments be located immediately following the affected part of this algorithm, they are indented one level in the presentation; also authoritative rules are set to the outer margin.

1. Start with the longest, connected, straight chain of benzene rings. Call this chain the "primary chain". For the case that this choice is unambiguous, consider that the molecule lies in a hexagonally tessellated grid. Overlay grid with a Cartesian coordinate system so that this primary chain lies along the x- axis and is bisected by the y-axis.

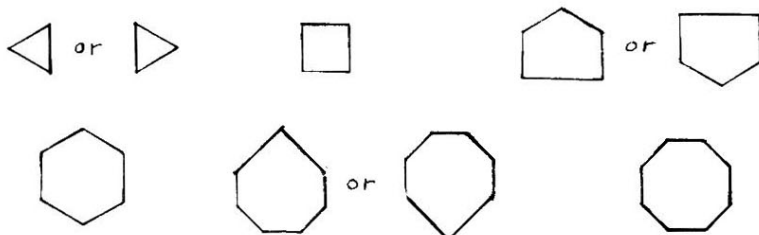
Comment: Because a positive direction has not as yet been allocated to either the x or y-axes, there may exist four possible ways to do this. For those cases wherein there is either horizontal or vertical symmetry, this number is reduced to two. Similarly, for both horizontal and vertical symmetry, the choice of positive is immaterial on both axes.

In case there exists two or more equally long "longest" chains, see below for determination which will be the primary chain.

2. Position unsymmetric forms so that they are "top heavy" to the upper right.

Comment: In most instances, "top heavy" makes a clear demarcation line, and thus a choice has been made as to which of the four possibilities listed in 1 should be selected. This is an outgrowth of IUPAC Rule A-22.1:

For the purposes of enumeration, the individual rings of a polycyclic "ortho-fused" or "ortho- and peri-fused" hydrocarbon system are normally drawn as follows:



and the polycyclic system is oriented so that (a) the greatest number of rings are in a horizontal row and (b) a maximum number of rings are above and to the right of the horizontal row (upper right quadrant). If two or more orientations meet these requirements, the one is chosen which has as few rings as possible in the lower left quadrant.

The system thus oriented is numbered in a clockwise direction commencing with the carbon atom not engaged in ring fusion in the most counter-clockwise position of the uppermost ring, or if there is a choice, of the uppermost ring farthest to the right, and omitting atoms common to two or more rings.

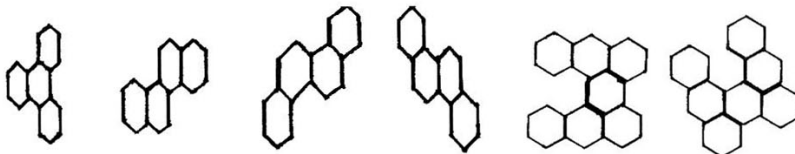
Because a logically consistent algorithm to yield such an orientation will require a sophisticated method of weighting the contributions from each individual ring in order to form a composite picture over the entire plane:

Some criterion to determine comparative "top heaviness", such as the location of the center of gravity, must be employed for large molecules.

Comment: IUPAC's Rules are vague as to the method for determining that "a maximum of

rings are above and to the right of the horizontal row." However, the original article by Patterson <3> on which these IUPAC Rules are based is much more detailed:

Note 11. Only the number of rings and not their size or distance from the center, is taken into account. When the correct orientation is not immediately apparent, bisect the "horizontal row" by a horizontal axis and a vertical axis (as in XXI) and count the rings and fractions of rings in the four quadrants. If there is more than one combination of rings that could serve as the "horizontal row," apply the bisection in the other cases also (as in XXII). Choose the orientation that has as many as possible of the "remaining rings" in the upper right quadrant; if two or more orientations meet this requirement, choose the one of them that has as few rings as possible in the lower left quadrant. In the examples, phenanthrene (XVI) is shown correctly oriented with its single "remaining ring" turned upward to the right; triphenylene (XVII) has one in the upper right quadrant, and one in the lower right; Examples XVII-XX show right and wrong orientations of chrysene; and XXI and XXII show right and wrong orientations in a case that might not be immediately apparent,



XVII XVIII(right) XIX(wrong) XX(wrong) XXI(right) XXII(wrong)

Comment: Contrasting Synthesis Algorithm's Rule 2 to the overly simplistic comparison of a single number used in Patterson's Note 11, observe that intuition, parallel BUT NOT IDENTICAL to that used by Patterson, is introduced. For instance, the concept of balance chosen includes the distance from the

center. Also, if this had been proposed as a widely usable (rather than merely as a historical) algorithm, and if the logical extension of this algorithm to general arenes were to be developed, then size would definitely influence the center of gravity computations and thus have to be included. Note that the recognition and usage of intuition -- to describe the relevant physical principles, as well as to introduce simplification -- is readily subscribed to in this algorithm. This is in contradistinction to the disguised usage of intuition by Patterson, which was never fully legitimized by incorporation of this note by IUPAC. Such an acceptance could have led to a long regress of rules to compensate for the exceptions to the exceptions, as they are discovered.

For the case of two or more equally long chains, use the criterion of "top heavy to the right" to select: 'Which is the primary chain?' and 'What orientation should this primary chain have?'

Comments: (1) Selection is from a set of eight orientations for two such chains in an unsymmetric molecule, 12 orientations for 3 chains, etc. Similarly, molecules containing

symmetries will require examination of fewer orientations.

(2) That one orientation could be more "top heavy"

(Figure 1) and a different orientation have more rings to the right of the y-axis

(Figure 2) has not been

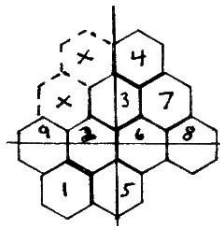


FIGURE 1

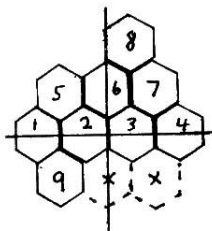


FIGURE 2

examined by either IUPAC or Patterson.

The only additional situation considered is that a choice had not yet been made.

This is covered by the criterion of minimizing the number of rings in the lower left quadrant. In other words, the only comparisons made are between

the number of rings in quadrants I and

III. By this criterion, Figure 1 is the desired orientation.

Next, adding a tenth and eleventh rings at the hexagonally tessellated positions marked by an x in Figures 1 and 2 makes the first and third quadrants congruent in these two figures and thus NO choice yet using IUPAC's Rules. This is despite that Figure 1 is still more top heavy, while Figure 2 has more rings to the right of the y-axis.

(3) Assignment of an orientation (through the primacy of "top" over "right" to the situation in Comment 2 does supply 'desirable' completeness; however, this is entirely counter-productive, and contrary to the spirit, if not the letter, of both the accepted IUPAC Rules and an intuitive sense of balance.

Because of the finite nature of the ring system to be named, it will always be possible to make a selection by adding yet another rule. The result will be that, after some point, these later rules will be "purely curve fitting" (that is, if a large enough polynomial is chosen, any finite data set can be exactly fit by this polynomial). This, however, is pseudo-science in the spirit of "Ptolemaic astronomy". <4> It is for this reason that, whenever approaching such a regress of rules, the Synthesis Algorithm should be abandoned, rather than "corrected".

3. Contiguous to the set of rings being named, there exists a unique set of "boundary" hexagons. All position locations will refer to this one-hexagon thick "envelope". Let position "A" be the boundary hexagon which elongates the primary chain.

Comments: (1) The rings in the primary chain will be named first. This may be seen by beginning with first a one and then a two ring system. Because of the uniqueness of these ring systems, there is no advantage in including a location name for these rings. This may be indicated by an * in each ring. For additional rings in this primary chain, each such ring will be annelled at boundary position "A". For example, a four ring linear chain would be named:

AA-Tetrabenzene. (No location name was included for the first two rings.)

(2) Notice that use of locant numbering was deliberately avoided, since using locants would imply that location "A" be on the upper right ring between locants 1 and 2 -- in conformity with IUPAC Rule A22.1 (see above). This, however, would be internally contradictory, since by Rule 2, location "A" elongates the primary chain. In other words, "A" would have been used in two different, non-compatible ways.

When position "A" could be either an end chain or an internal chain choose the free end of the end chain. In case there are either two free ends or no free ends to this primary chain, choose

position "A" to the left.

(1) Because balance had been directed to the right by Rule 2, the majority of free ends of end chains will lie to the left.

(2) Adding a ring to an end chain of an existent system will, in "nearly all" cases, be an orthoquentlyfusion. Consequently it won't create additional triple points <1> and change the "chemical viability" of the ring system. Adding an internal ring, on the other hand, always increases the number of triple points.

(3) When the total number of triple points is even, the resultant system usually allows for conjugation without problem; however, an odd number of triple points always produces a chemically unattainable system of polybenzenes.

(4) The earlier version of this rule in <1> is not able to account for exotic combinations, such as in Figure 3,

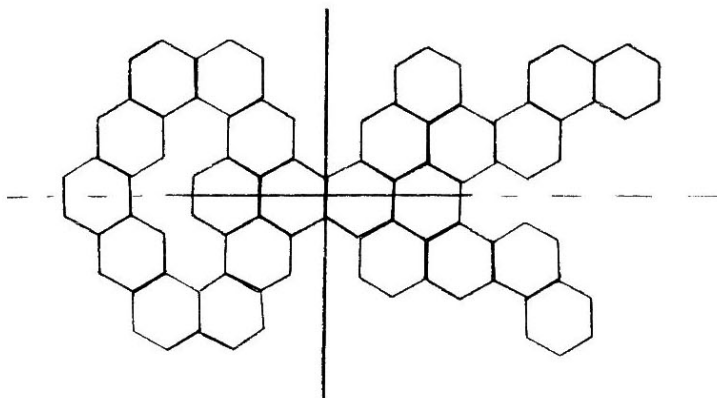


FIGURE 3

wherein the primary chain is "surrounded".

In this case, although the term "end chain" is unambiguous, the intent has been subverted, and such a system has exceeded the practical limits of this Synthesis Algorithm.

4. Position "B" is the first hexagon of the envelope contiguous to the set of rings already named and also to envelope hexagon "A", in a clockwise direction. Similarly, "C" is the next clockwise hexagon of this envelope, etc.

Comments: (1) Because regular hexagons tessellate the plane, there exists exactly one such ring location for any polybenzene. Furthermore, notice that names of neither the edges nor the vertices of the polybenzene itself were needed in the determination of these locations, only names for the hexagonal embedding space. In other words, a locant numbering system was NOT required for unsubstituted polybenzenes.

(2) Although there exists a simple relationship between these envelope hexagons and the edges of a polybenzene, this is not needed for the Synthesis Algorithm. This relationship will be described below, when it is used in extending the Analysis Algorithm to the broader class of arenes.

The last two Rules given in the Synthesis Algorithm <1> may be disregarded: Rule 5 had been included merely to emphasize the non-computer aspects of this algorithm, and, especially in the case of rotational symmetry to keep the location designations low. It was modelled after IUPAC's not assigning locant numbers to carbon atoms that were at a point of ring fusion, and Patterson's supplemental note on the definition of his Class B <5>:

"It would be more consistent to number the systems of Classes B and C alike and to include all ring members in the numbering, but this would run the numbers up very high in

polycyclic aromatic compounds, and for the most part these additional numbers would not be used."

Rule 6 has been included above by the delineation of what is being synthesized, and thus is redundant.

Before turning to the Analysis Algorithm, it is important to notice that if one were to now combine Patterson's Rule 12 (with its explanatory Notes 12 and 13):

Rule 12. Number the oriented formula by beginning with the first free angle of the upper right ring and proceeding clockwise around the entire formula to the beginning, numbering all carbon atoms that are not common to two or more rings.

Note 12. The "first free angle" is the first angle, not also part of another ring, from which one may proceed clockwise around the ring.

Note 13. The "upper right ring" is the highest ring in the formula or, if two or more are equally high, then the ring farthest to the right in the highest row.

with IUPAC's Rule A21.5:

Isomers are distinguished by lettering the peripheral sides of the base components a, b, c, etc., beginning with "a" for side "1,2", "b" for "2,3" (or in certain cases "2,2a") and lettering every side around the periphery....

a different edge labelling sequence would be obtained than the one used in the Synthesis Algorithm presented above. Such a sequence might possibly delay, but could not avoid, some of the inconsistencies mentioned in <1>. The disadvantage of such a sequence, however, is that there is not a fixed generalizable site for a given locant, and thus no standard relation between name and geometry, as exists in the Synthesis Algorithm presented above; e.g., all-A chains are linear, all-B chains form a circular arc, etc.

3. EXTENSIONS OF THE ANALYSIS ALGORITHM

Although the hexagonal tessellation supplies a unique framework into which the presence or absence of a ring may be readily noted, this requirement may be relaxed in order to expand the system of compounds being canonically named to convex modules; viz., internally all-cis conjugated ring systems.

The changes induced by this relaxation are to replace the word "polybenzene" by a different generic word, such as "arene", and to require the naming of a ring size as well as the location of each ring. This results in there being two more prefixes than for the corresponding polybenzene name, since the size of the first two rings must be specified, even though a location has been implied and need not be included. Furthermore, as there is now no desire to group into subsets, the arene name has listed the locants in the reverse order from the one used in the polybenzene name, inasmuch as left to right naming seems more natural. Notice that, in principle, this is still a one-parameter system in which bond saturation has been maintained. Table 1 contains the 35 compounds which IUPAC has chosen as its starting point <6>, along with their arene and polybenzene names, when applicable.

T A B L E 1

| # | IUPAC NAME | ARENE NAME | POLYBENZENE NAME (when applicable) |
|----|-------------------|-------------------------|---------------------------------------|
| 1 | Pentalene | 5,5-Diarene | |
| 2 | Indene | 6,5-Diarene | |
| 3 | Naphthalene | 6,6-Diarene | Dibenzene |
| 4 | Azulene | 7,5-Diarene | |
| 5 | Heptalene | 7,7-Diarene | |
| 6 | Biphenylene | 6,4,6a-Triarene | |
| 7 | as-Indacene | 5,6,5b-Triarene | |
| 8 | s-Indacene | 5,6,5a-Triarene | |
| 9 | Acenaphthalene | 6,6,5c-Triarene | |
| 10 | Fluorene | 6,5,6a-Triarene | |
| 11 | Phenalene | 6,6,6c-Triarene | 1H-c-Tribenzene |
| 12 | Phenanthrene | 6,6,6b-Triarene | b-Tribenzene |
| 13 | Anthracene | 6,6,6a-Triarene | a-Tribenzene |
| 14 | Fluoranthene | 6,5,6a,6c-Tetraarene | |
| 15 | Acephenanthrylene | 6,6,5a,6b-Tetraarene | |
| 16 | Aceanthrylene | 6,6,6a,5c-Tetraarene | |
| 17 | Triphenylene | 6,6,6b,6j-Tetraarene | jb-Tetrabenzene |
| 18 | Pyrene | 6,6,6b,6e-Tetraarene | eb-Tetrabenzene |
| 19 | Chrysene | 6,6,6b,6b-Tetraarene | bb-Tetrabenzene |
| 20 | Naphthacene | 6,6,6a,6a-Tetraarene | aa-Tetrabenzene |
| 21 | Pleiadene | 6,7,6a,6c-Tetraarene | |
| 22 | Picene | 6,6,6b,6b,6d-Pentaarene | dbb-Pentabenzene |
| 23 | Perylene | 6,6,6a,6c,6i-Pentaarene | ica-Pentabenzene |
| 24 | Pentaphene | 6,6,6a,6b,6c-Pentaarene | cba-Pentabenzene |

| | | | |
|----|----------------|---------------------------------|---|
| 25 | Pentacene | 6,6,6a,6a,6a-Pentaarene | aaa-Pentabenzene |
| 26 | Tetraphenylene | 6,8,6a,6d,6l-Pentaarene | |
| 27 | Hexaphene | 6,6,6a,6a,6b,6c-Hexaarene | cbaa-Hexabenzene |
| 28 | Hexacene | 6,6,6a,6a,6a,6a-Hexaarene | 4a-Hexabenzene |
| 29 | Rubicene | 6,5,6a,5a,6a,6d,6k-Heptaarene | |
| 30 | Coronene | 6,6,6a,6c,6e,6i,6k-Heptaarene | kieca-Heptabenzene (2-3-2)-Heptabenzocluster |
| 31 | Trinaphthylene | 6,6,6a,6b,6c,6n,6o-Heptaarene | oncba-Heptabenzene |
| 32 | Heptaphene | 6,6,6a,6a,6b,6c,6d-Heptaarene | dcbaa-Heptabenzene |
| 33 | Heptacene | 6,6,6a,6a,6a,6a,6a-Heptaarene | 5a-Heptabenzene |
| 34 | Pyranthrene | 6,6,6a,6b,6c,6g,6h,6k-Octaarene | khgcba-Octabenzene |
| 35 | Ovalene | - | (3-4-3)-Decabenzocluster |

Some important comments about this table are:

(1) The choice of the name "arene" is probably chemically acceptable, even though it is not desirable, inasmuch as the aromaticity of these compounds is not relevant mathematically. All that is of mathematical significance for the configurations under consideration is: (a) polycyclicality; (b) conjugated single and double bonds; and (c) co-planarity locally. A better generic term might be "polycycloconjugene"; e.g., pentalene would be 5,5-Dicycloconjugene; etc. Throughout the rest of this report, the shorter term "arene" will be used for known aromatic systems only. "Cycloconjugene" will include those

compounds which are not aromatic, such as cyclooctatetrene, or where the aromaticity is not stated.

(2) Indene (#2), fluorene (#10) and phenalene (#11) form a conjugated system ONLY for the free radical. The IUPAC table, therefore, shows one specific hydrogenated compound, with its concomitant break in conjugation.

(3) In 'walking around' a molecule of rubicene (#29), the relaxation of the tessellation requirement is most evident. The greater angles between locants #7 and 8, and between #14 and 1, seem to suggest the possibility of annelling various combinations of conjugated figures between the named locants. Were it not for the specific constraint of a maximum vertex incidence equal to three, this would raise the question 'How many letters should be assigned to the set of edges between locants #7 and 8, and between #14 and 1?' However, by the imposed limit, the answer is unambiguously one. Notice that in any polybenzene this problem never arises as the tessellation grid describes exactly how many additional hexagons can be added.

(4) For both coronene (#30) and pyranthrene (#34) there have been listed both a linear and a cluster polybenzene name. A cluster name could be given for every polybenzene; however, there is no advantage in doing so.

Before examining the extension of the Analysis Algorithm, note that from a purely mathematical perspective, any multiply-connected polybenzene of n rings may be viewed as a ring system containing $n + g$ rings, where g is the genus of the system. For example, a doubly-connected region

would have $n + 1$ rings, etc. Consequently, names for globally-planar, multily-connected regions, such as hfdb-Hexabenzene for coronene should NOT be used. Instead, in conformity with the principles adopted by Chemical Abstracts <7>:

"... coronene is considered to be a seven-ring system instead of a six-ring system even though the six benzene rings will account for all the atoms and valences of the system. The "hole" in the center is considered to be the seventh ring...."
In this way, whenever a name might be interpreted as either a coplanar or non-coplanar union of rings with the corresponding difference in the number of peripheral hydrogen atoms, the non-coplanar configuration will be implied. (Notice that there is no loss of generality as coplanar forms were named as the set of $n + g$ rings.) Therefore, hfdb-Hexabenzene will be the major part of the name for helicene. Because helicene exists as two enantiomorphs, a chiral indicator (R or S) should precede the sequence of location designators.

This technique is adequate PROVIDED all regions of multiple-connection are convex (i.e., all double bonding is cis). This is shown in the nomenclature of the circulenes, which is given below. Concave configurations (e.g., trans double bonding), on the other hand, as in the annulenes, require special treatment.

For simply-connected cycloconjugenes, the extended analysis algorithm becomes:

T A B L E 2

EXTENDED ANALYSIS ALGORITHM FOR NAMING SIMPLY-CONNECTED CYCLOCONJUGENES

1. Find the longest "straight" chain of ortho-fused rings. "Straight" is unambiguous PROVIDED each ring has an even number of edges; also if there are exactly two rings of any size. In the event that only one ring in this chain has an odd number of edges (called an "odd ring"), "straight" shall mean either of the two most remote edges from the ring fusion or the starting ring. For more than one odd ring in this chain, "straight" shall mean an alternation of the odd rings about successively most remote edges from the position of ring fusion.
2. If there is more than one chain of this length, consideration may be limited to those chains having the largest end ring. In case two or more such chains have equal size end rings, compare next to end rings, etc.
3. For all such chains, designate all rings, after the first two, by location "a".
4. In addition to the rings in this longest chain, also designate as position "a" the straight extension of this longest chain away from the larger end ring. If either end ring is odd and there are no other odd rings in this chain, or if this chain is symmetric, there are two possible locations for "a".
5. Form a sequence of "interim" names for a given

structure by designating location "b" to be the edge adjacent to "a", first in a clockwise, then in a counterclockwise, direction. The assignment of "b" creates a unique vertex locant numbering: the vertex common to sides "a" and "b" is named "2", while the preceding vertex (in the direction of "a") is named "1". Walk around the perimeter of the molecule naming the successive vertices (in the positive sense) as "3", "4", etc. Notice that every vertex of the perimeter of the figure is given a locant number.

6. Observe the number of edges incident at each vertex. For every vertex with incidence = 3 the two incident boundary edges will be assigned the same letter name, while for every vertex with incidence = 2 these edges will have sequential letter names.
7. Choose the lexicographically lowest interim name -- reading left to right.

Some important comments about this table are:

- (1) The four ring system shown in Figure 4 is classified as "straight" and is named 6,5,5a,6a-Tetraarene; whereas the molecule shown in Figure 5 is incorrectly drawn. The demand for ring alternation of odd rings causes this molecule to be drawn as shown in Figure 6, with the name 6,5,5a,6b-Tetraarene.
- (2) Because of Rule 4, the larger end ring is always named first.

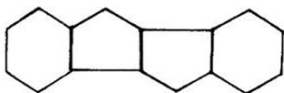


FIGURE 4

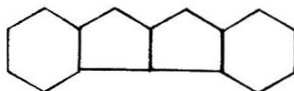


FIGURE 5

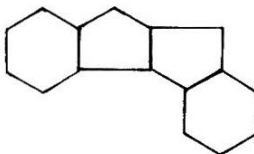


FIGURE 6

(3) Both the larger end ring and the next ring in sequence are denoted ONLY by their size. All other rings are specified by both size and location.

(4) When forming interim names for the polybenzenes, two points of view could have been chosen:

(a) The given configuration is embedded onto the hexagonal tessellation of the plane, as in the Synthesis Algorithm. Next, the hexagons of the boundary "envelope" are named in increasing alphabetic order (with repetition if necessary). See Figure 7.

The first time that any specific hexagon of the given configuration is adjacent to any of the boundary hexagons it is given the next letter as its locant name. For example, the first hexagon not on the starting line is contiguous to boundary hexagon "b" and thus is given the name "c". Notice that in the process of circumscribing the given figure the number of sequential edges that are to be given the same

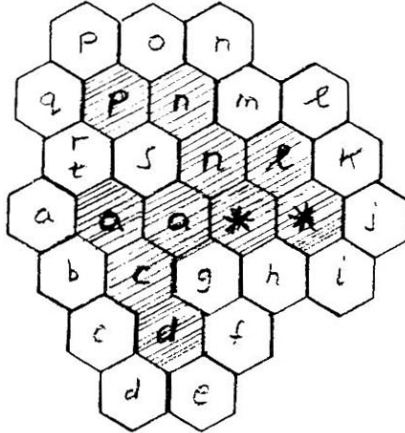


FIGURE 7

letter is clearly indicated (since all belong to the same boundary hexagon; e.g., 5 edges are named "s", 4 are named "g", only 1 named "a", etc.). In other words, this corresponds to a simple geometrical picture.

(b) Starting from location "a" and walking around the perimeter give a common name to both edges adjacent to any vertex whose incidence is exactly three. For example: denote the vertices of edge "a" (See Rule 4 of Table 2) by "1" and "2". In Figure 8 both of these vertices have incidence = 2 and thus the adjacent edges must have different names. (These will subsequently be shown to be "t", "a" and "b" respectively.) Vertex "3", on the other hand, has incidence = 3 and thus both adjacent edges must have the same letter name ("b"). Continuing in this manner

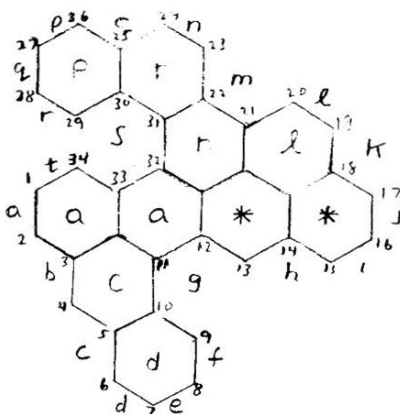


FIGURE 8

for each of the 34 perimeter vertices, the edge incidence is noted. For instance, vertices "10", "11" and "12" all have incidence = 3, but "9" and "13" have incidence = 2; therefore, the portion of the perimeter between "9" and "13" is given a common name -- namely "g". Notice that representations (a) and (b) give an identical name to a configuration, as may be seen by comparing Figures 7 and 8. In other words, the hexagonal tessellation was a pictorial convenience, but not a necessity, for determining how many contiguous edges should be given the same locant name. For the general cycloconjugene, such a grid is either highly irregular or, in some instances, may not even exist; however, the perimeter incidence is clearly determined. For example, measuring counterclockwise along the perimeter of the nine ring compound shown in Figure 9, it may be noted

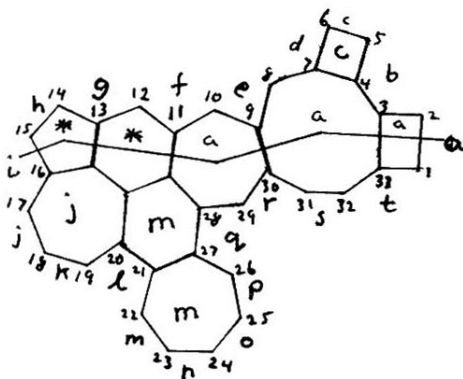


FIGURE 9

that vertices "3" and "4" have incidence = 3, but vertices "2" and "5" have incidence = 2; thus a common name is given to that portion of the perimeter; namely "b". Similarly, the rest of the perimeter can be labelled. Interim names can thus now be determined as they were for polybenzenes. For instance, the interim name for counterclockwise traversing of the configuration given in Figure 9 is:

5,6,7a,9a,4a,4c,7j,6m,7m-Nonacycloconjugene

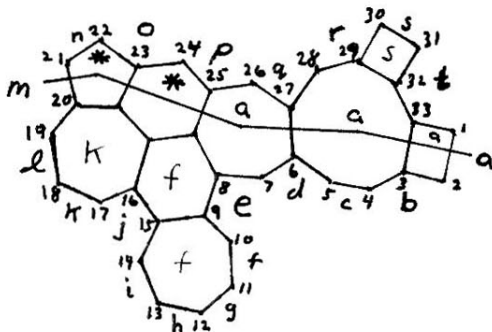


FIGURE 10

locant numbers have been allocated to every vertex that lies on the boundary. Consider next that an additional single bond has been included in Figure 11 between locants #2 and 4; thereby forming a highly strained three membered ring (Figure 12).

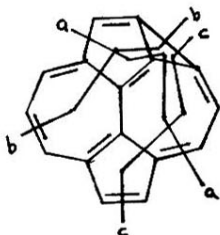


FIGURE 12

Notice that such a system would have retained its conjugation, but would have a different locant numbering, inasmuch as two edges have been replaced by one.

Consider now the different three membered straight chains that can be formed:

- Heptagon (right) - triangle - pentagon (c) -- Path a
 - Heptagon (left) - pentagon - triangle -- Path b
 - Pentagon (h) - heptagon (right) - triangle -- Path c
- (as well as these in reverse order).

All attempts to form a four polygon long chain are frustrated by the demand for straightness. By Rule 3, Path b represents the structure to be used for purposes of nomenclature. This configuration is thus named: 7,5,3a,7b,5e-Pentacycloconjugene. The reversed direction path gives an interim name with higher locant numbers and thus may be discarded.

Another example of the use of this nomenclature is given for various of the circulenes <10>. The desired analytic name and the method of determination for [5], [7]

and [8]-circulene are given in Figures 13 thru 15.

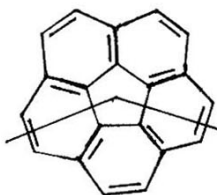


FIGURE 13
6,5,6a,6c,6g,6i-
Hexaarene

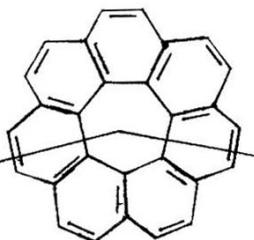


FIGURE 14
6,7,6a,6c,6e,6i,
6k,6m-Octaarene

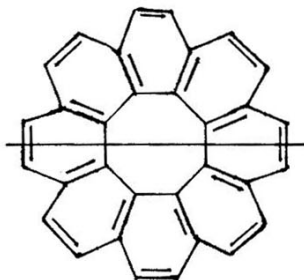


FIGURE 15
6,8,6a,6c,6e,6g,6k,
6m,6o-Nonaarene

Notice that [6]-circulene was not included, as this is simply coronene (See #30 in Table 1). An important observation now is that, unlike the polybenzenes, wherein a highly peri-fused ring system is geometrically desirable, there is no comparable cluster with incidence = 3 of conjugatable modules that would not be very highly strained. Therefore, it does not seem worthwhile developing nomenclature for such a system at this time.

Because it would be desirable to avoid cluster names altogether, the algorithm given in Table 2 should be expanded to accommodate concave (e.g., trans bonded) rings. For such an expansion, Rule 1 of Table 2 requires major revision. First, the term "longest" must be redefined. For instance, in Figure 16, there are three horizontal rows

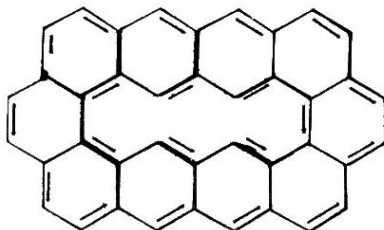


FIGURE 16

containing, from top to bottom: 4 hexagons, 1 tetradecagon and 2 hexagons, 4 hexagons. Using Algorithm 2, this would cause either the top or bottom row to be designated as the longest chain -- which is contrary to chemical considerations; namely, the elimination of a pair of hydrogen atoms on "nearby" carbons of the tetradecagon thereby forming bonds and creating a middle row of five benzene rings greatly increases the stability of the compound. This is easily compensated for by amending Rule 1 to count every such trans bond as a half ring; however, such a correction is adequate ONLY when two adjacent rings share no more than two common edges. (Note that in an all cis configuration the problem is obviated by the definition of "straight".) In Figure 17, on the other hand, the two graph theoretical pictures using the classical concept of different "resonance structures" create a different number of cis and trans bonds. This problem arose because cis bonding is only an approximation to convexity

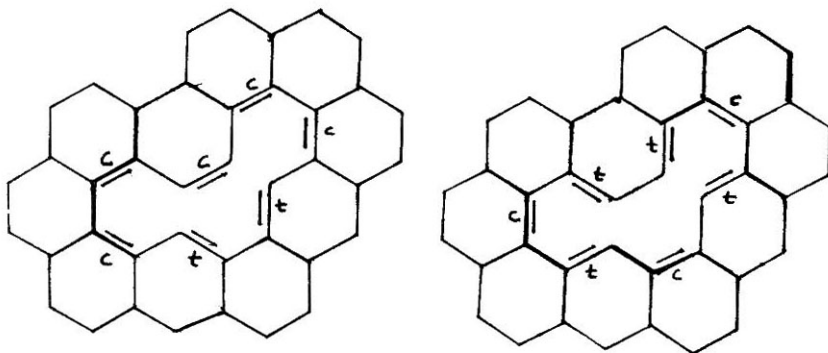


FIGURE 17

mathematically, and the use of a different edge coloring of a graph would be required in order account for such structures. Development of this idea for the very few, if any, chemically realizable structures is not warranted at this time, as it would lengthen the Analysis Algorithm beyond practical limits; however, if such compounds are developed, any systematic nomenclature would be very cumbersome and this development would then be worthwhile.

REFERENCES

1. Elk, Seymour B., "A Nomenclature for Regular Tessellations and its Application to Polycyclic Aromatic Hydrocarbons", MATCH - Communications in Mathematical Chemistry, 8, 121-158, 1980.
2. "Nomenclature of Organic Chemistry: Section A", IUPAC, Pergamon Press, Oxford, England, 1979
3. Patterson, Austin M., "Proposed International Rules for Numbering Organic Ring Systems", JACS, 47, p.547, 1925
4. Encyclopedia Britannica, 15, Chicago, Ill., 1976, p.179
5. Ibid 3, p.554
6. ibid 2, p.23
7. Capell & Walker, "The Ring Index", 2-nd Edition, Supplement 1, American Chemical Society, Washington, D.C., 1962, p.ix
8. Randic, Milan, "Aromaticity and Conjugation", JACS, 99, 1977, 444-50
9. ibid 2, p.28
10. Greenberg, Arthur & Liebman, Joel F., "Strained Organic Molecules", Academic Press, New York, N.Y., 1978, p.153.