AN IMPROVED SYSTEM FOR THE ENUMERATION AND DESCRIPTION OF RING SYSTEMS

BY JOHN A. SILK
Imperial Chemical Industries Ltd., Jealott's Hill Research Station,
Bracknell, Berkshire, England

The methods which have been developed in recent years for the systematic description and enumeration of ring structures fall mainly into two categories. In one class are those[1,2,3] based essentially on tracing a continuous path through the atoms forming a ring system, and in the other are those[4,5,6] in which individual rings are built up one by one. While the former methods appear, at least in some instances, to be simpler to apply, the reconstruction of a conventional ring-diagram from the linear description of it is less straightforward, because the individual rings comprising a system appear only by forming cross-linkages in the described path. Methods in the second category undoubtedly are easier to interpret, and this is because they follow more closely the usual way of regarding ring systems, i.e., in terms of individual, small rings fused together in various ways.

The most completely developed method of this type has been that of Taylor-Patterson-Dyson (T-P-D),[6] and this paper presents modifications of their system which yield improved enumeration patterns, show the characteristic types of ring fusion clearly, and enable molecular formulas to be calculated from the systematic expressions without any additional information. It will be apparent that the application of these methods lies mainly in the field of chemical notation, although they could also provide a basis for classifying ring systems, and a procedure for converting ciphers to systematic names has been described.[7]

The T-P-D method requires that enumeration shall commence in the largest ring at its point of fusion with the next senior ring, and proceed in the manner which gives the lowest sequence of numbers to the ring-fusion points. In symmetrical systems with all rings of equal size the basic pattern of enumeration is, therefore, a spiral one; enumeration starts in the center ring and moves outward on successive orbits until the whole structure has been embraced (Fig. 1). In many other cases, however, in which the system is asymmetric, or the central ring is not the largest one, enumeration has to commence off-center, and this results (i) in

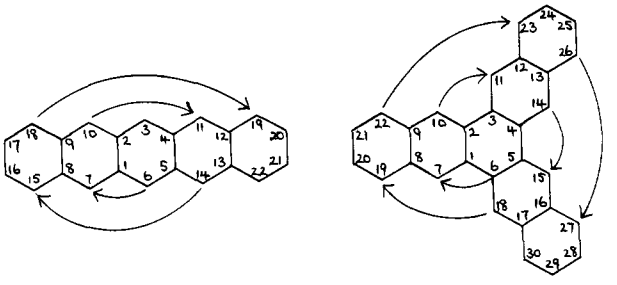


Fig. 1.—The spiral pattern of enumeration in symmetrical systems with like rings.

related systems, which differ only in sizes and/or sequences of rings, having quite different patterns of enumeration, and (ii) in a greater number of orbits of the generating ring being required to complete the enumeration. (In Fig. 2, left-hand column, compare A, B and C; D and E; F and G.) It has now been found that, if all systems are to be described with maximum economy, i.e., in terms of the fewest and smallest orbits, and if related systems are to be enumerated similarly, the enumeration must always start in the most central ring (regardless of size), and further, that it must proceed in the same direction throughout and in the manner which gives not the lowest sequence of locants, but the lowest total. The validity of these findings as general principles has been reinforced strongly by the exploration of the new method explained below for describing and enumerating peri-fused systems; the satisfactory application of this method makes it essential for all the component rings of a system to be enumerated in the same direction (clockwise or anticlockwise) and in the manner which gives the lowest total of locants.

In connection with these modifications it may be noted that in the Wiswesser system,[2] which uses a continuous-path method, the corresponding rule is also a "lowest total" one, and mathematical support has been claimed for this. Furthermore, if one draws an analogy with electron orbits in atoms, it may be seen that the description of a system in terms requiring more

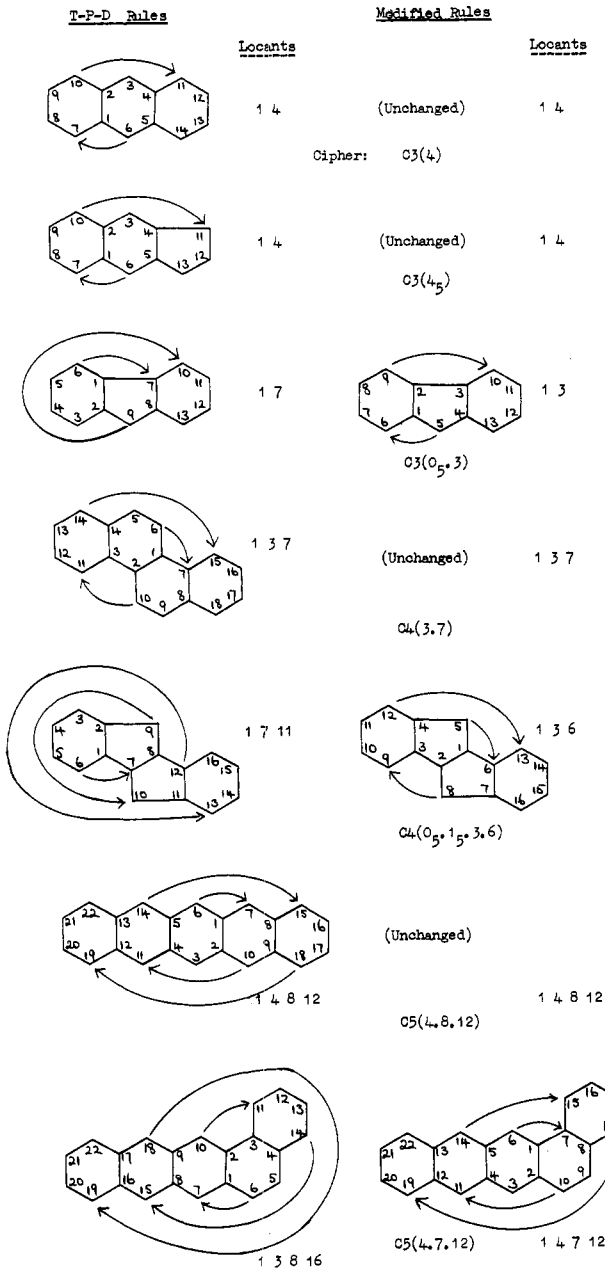


Fig. 2.--Comparison of enumerations of other typical systems.

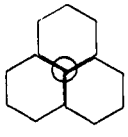
than the minimum number of orbits about the starting ring corresponds to the description of a state of higher energy. Only the lowest total rule enables all systems to be described in their ground states.

The modifications to which these rules give rise are illustrated in Fig. 2, right-hand column. It can be seen that the enumeration of C now closely resembles that of A and B and that fewer orbits of the generating ring have to be made. Similar considerations apply to E and G, in relation to D and F, respectively.

The notation accompanying the examples also differs slightly from the T-P-D one. Cn denotes a system of n fused rings (cycles), and the fusion-point locants are cited after this in parentheses. Since the first locant is 1 in all cases, this normally may be omitted and only the subsequent ones quoted. In addition, since six-membered, aromatic rings are by far the commonest, it may be understood that all are of this type in the absence of other indications. Successive locants are separated by stops. Other sizes of ring may be shown by subscripts to the appropriate locants, or by :n if it is desired to avoid the use of subscripts (as on punched cards). When the first ring is not six-membered, locant 0 is introduced to carry the ring-size number, and for the second ring locant 1 may be used to carry a subscript (as in these examples), or :n may be used on its own.

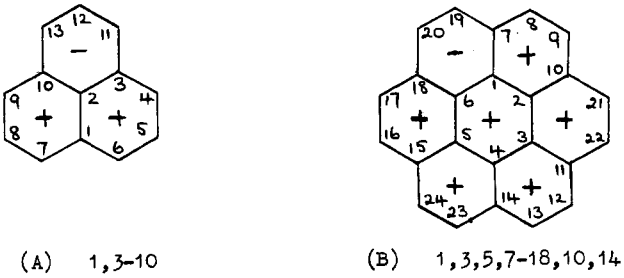
RETICULAR SYSTEMS

"Reticular" is used here in the limited sense in which it was defined by Taylor[4] and was used in my earlier system of notation.[5] Basically, it denotes the peri-fused type of system, the characteristic unit of which is



In this the encircled central atom, which may conveniently be referred to as the Y- atom, is common to a group of three rings, and the three atoms joined to this are each common to one of the three pairs of rings.

With the T-P-D rules systems of this type often require a hyphenated locant to show a link between non-consecutive positions, and a reversal in the direction of enumeration pattern in some rings, which destroys the symmetry of the enumeration pattern (3A,3B). From the resulting expressions this characteristic and fundamental type of fusion cannot be recognized readily, nor molecular formulas calculated directly.



(A) 1,3-10 (B) 1,3,5,7-18,10,14

Fig. 3.--T-P-D enumeration of reticular systems.

The nucleus of a group of rings in reticular fusion is the Y-atom, and the stages by which the simplest reticular structure should be built up systematically are shown in (4A). To the starting ring is added, firstly, a three-atom, reticular adduct; this does not in itself complete the second ring, but it forms the bridge between the starting ring and the next ring-adduct, which starts from the 2-position. This second adduct, which completes the structure, contains the normal number of four atoms, because it joins directly onto its parent ring (at the 3-position).

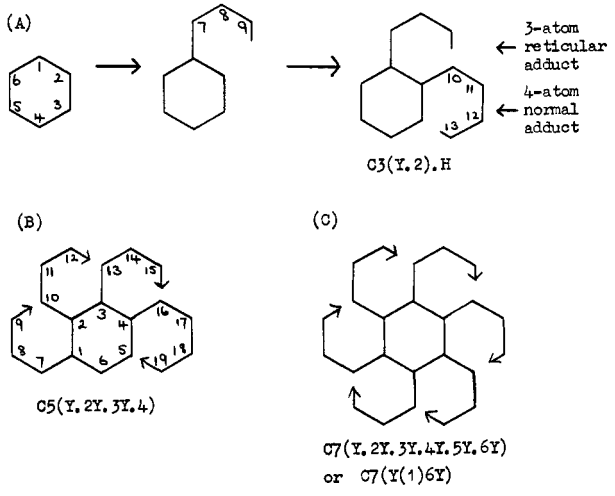


Fig. 4.—Analysis of reticular systems.

(4B) shows that several reticular adducts, each of three atoms, can be added in succession, and, as in (4A), the structure is completed by a normal, four-atom adduct. (4C) shows the addition of a complete cycle of reticular adducts, giving coronene. In this case the sixth reticular adduct completes the structure by bridging the starting ring and the first adduct, so that there is no four-atom adduct. This analysis clearly displays the perfect symmetry of the coronene structure, and it enables the perimeter of this system to be enumerated in one regular and continuous sequence. By the T-P-D method, on the other hand, coronene is built up in two unequal stages via triphenylene, and, because the direction of enumeration is reversed in one ring, the symmetry is destroyed (3B).

Reticular fusion can be shown simply by inserting Y after the appropriate locants. Thus, for (4A) we may write C3(Y.2).H to denote that the adduct starting at the 1-position is in reticular fusion with the adduct starting at the 2-position (H specifies the extra hydrogen). Enumeration continues smoothly from the reticular adduct to the normal one. Position 10 might be regarded as belonging equally to the second or to the third ring, but systematically it is correct to associate it with the latter. In the examples this analysis is shown by the positioning of the numbers.

Further examples of reticular systems are shown in Fig. 5. The Y-atoms are encircled in these, and it can be seen that they are always at a position one higher than the locant with which they are used. Since a reticular adduct contains one atom less than a normal adduct, each Y-symbol requires CH to be subtracted from the general formula $C_{4n+2}H_{2n+4}$ for a system of n six-membered, edge-fused, aromatic rings.

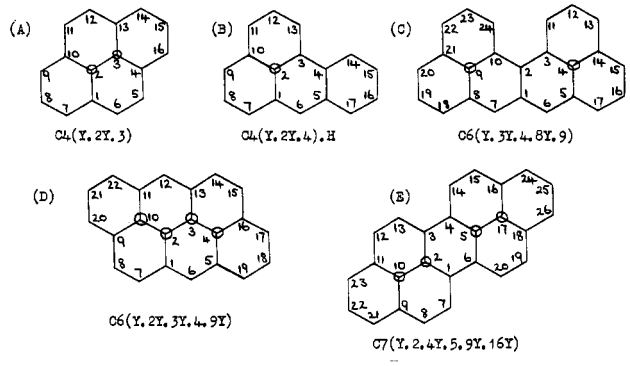


Fig. 5.—Simple reticular systems.

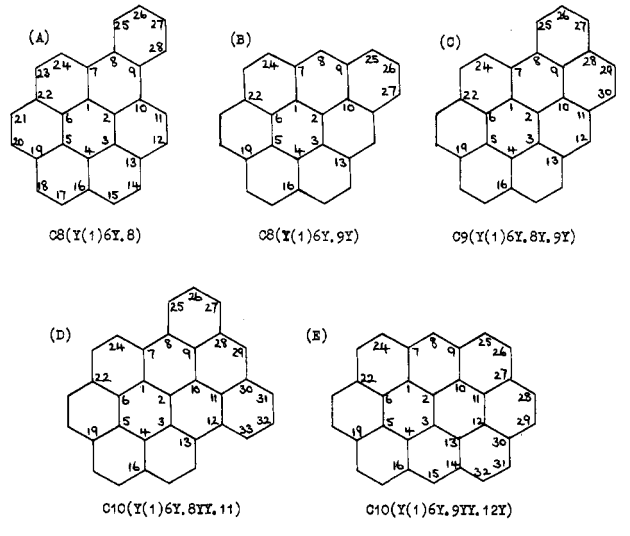


Fig. 6.--Complex reticular systems.

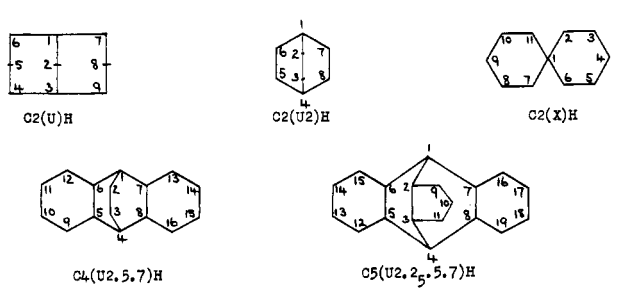


Fig. 7.--Bridged and spiro systems.

Comparatively few known ring systems exhibit a greater degree of condensation than coronene, but analysis in terms of their Y-atoms enables them all to be enumerated and described systematically with remarkable simplicity. There are two different sites at which further rings can be added to coronene. At one, for example, position 8, a normal adduct is required, and at the other, for example, position 9, a reticular one is needed, while when two are added in succession both are reticular. This is shown in Fig. 8 (A,B,C), and it may be noted that, because rings are enumerated in the same direction, it would be incorrect in the second case for the adduct to span positions 8 and 24, instead of 9 and 11.

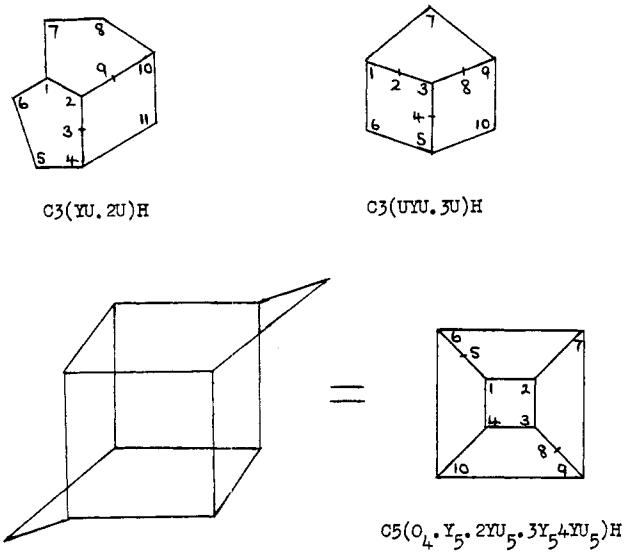


Fig. 8.—Combined reticular and bridge fusion.

The addition of three successive adducts leads to a new, higher degree of condensation, in which the middle ring is formed by the addition of only two atoms (Fig. 8D,E). Analysis of the general situation shows that whereas the conversion of benzene to coronene requires six single reticular adducts, each of three atoms, the conversion of coronene to "coronocoronene" requires six pairs of adducts, each pair consisting of five (3 + 2) atoms. The symbol YY may be used to signify addition of this two-ring, five-atom adduct. Similarly, YYY would denote a three ring-adduct of 3 + 2 + 2 atoms, if systems of this complexity became known. When less than a complete cycle of these adducts is required, the last one in a series necessarily contains one or two additional atoms, and it will, therefore, be ciphered in a separate operation. YY requires the subtraction of C_3H_3 from the formula calculated for a simple ortho-fused system.

OTHER TYPES OF RING FUSION

The description of bridged and spiro systems also benefits from the introduction of special symbols to signify these characteristic types of fusion. When fusion of two rings takes place along two faces instead of only one face, U is placed after the locant to denote the occurrence of a one-atom bridge. Similarly, U2, U3 and U4 are used for bridges of two, three and four atoms, respectively. For spiro fusion, which involves only one atom instead of two, X is placed after the locant. It is clear that Un requires the subtraction of n carbons in calculations of molecular formulae, and X the addition of one carbon in relation to the formula for ortho-fused systems. Systems of these types (Fig. 7) are mainly saturated ones, and in the ciphers the addition of H immediately after the closing parenthesis denotes complete hydrogenation.

Combined reticular and bridged fusion is encountered occasionally, as in hexamine (cf. 8B). Since reticular fusion involves two other rings, U precedes Y when it refers to the first of these rings, and U follows Y when it refers to the second. In order to recognize reticular fusion, it is helpful to draw out such systems in a plane form, avoiding crossed bonds as far as possible. In the rare cases where a bridge spans two or more rings and cannot, therefore, be treated as a reticular system, its termination must be shown by citing the appropriate locant, and these are now the only situations which require a pair of locants to be joined by a hyphen.

When the application of the rules of the lowest total, and then the lowest sequence, of locants still leaves a choice, the preferred expression is the one giving the lowest locant to the largest ring, and then to the senior type of fusion (i) reticular, (ii) normal (simple ortho-type), (iii) spiro, and (iv) bridge.

REVERSAL OF ENUMERATION

The obvious disadvantage of both the T-P-D methods and the present developments of them is that they assign the low numbers to the fused positions, so that even the simplest systems, like naphthalene and indene, are enumerated differently from established practice. The fused positions are, moreover, the ones least likely to be substituted. It is proposed, therefore, that for the purposes of citing the positions of heteroatoms and substituents the enumeration of all systems, as obtained by the methods described, should be reversed from beginning to end. The lower-numbered positions then become predominantly the unshared atoms, as is customary, and the lowest-numbered rings are always the outermost ones. Consequently, for

the commonest type of system, namely, normal fusion with six-membered rings (recognizable by the complete absence of modifying symbols), we know without even working out the ring structure that positions 1, 2, 3 and 4 are in one end-ring and positions 5, 6, 7 and 8 in another end-ring. This results in the unshared atoms in naphthalene, phenanthrene, indene and fluorene retaining their accepted enumerations, and anthracene differing only in the 10-position being renumbered 12--a change which is hardly likely to be misunderstood. Heterocyclic analogs of the two-ring systems are also unchanged, but for three-ring heterosystems no general statement can be made; anthracene, phenanthrene and fluorene are exceptions to the Ring Index rules and so are acridine and carbazole, which are numbered similarly. Some other heterosystems of fluorene structure are numbered like the hydrocarbon, but those resembling anthracene and phenanthrene are generally not.

HETEROCYCLIC SYSTEMS

In structural searches upon collections of organic compounds, it is often desired to identify all systems which contain within them a specific hetero-ring, such as pyridine or thiazole. The association of a heteroatom with a ring of a specific size, and, further, of two or more heteroatoms with the same ring is not always readily deducible from ciphers which specify simply the positions of heteroatoms. It is suggested, although it is not an essential part of the present proposals, that each individual hetero-ring should be described separately, in order to facilitate searches of this type in both printed indexes and punched-card or tape records. In the notation[8] in which the methods of this paper are utilized, D denotes a six-membered hetero-ring and Dn an n-membered hetero-ring, where n is not six. These are followed by the symbols for the hetero-atoms in the standard order, the symbols for like atoms being repeated as necessary; the locants are then cited in the same order. The ciphers for hetero-rings are placed in front of those for the ring system, because this improves the classification produced in cipher indexes; it brings together, for example, all systems containing a thiazine ring. A few simplifications are permissible, (a) in one- and two-ring systems in which the hetero-ring is not six-membered, the specification of ring-size in describing the hetero-ring with Dn makes it unnecessary to repeat this in describing the ring system with Cn, and (b) when a hetero-atom is common to two or more rings the locants in the expressions for rings in which all atoms have previously been cited are omitted. The examples in Fig. 9 illustrate these methods and the reversal of enumeration in specifying hetero-atoms and substituents.

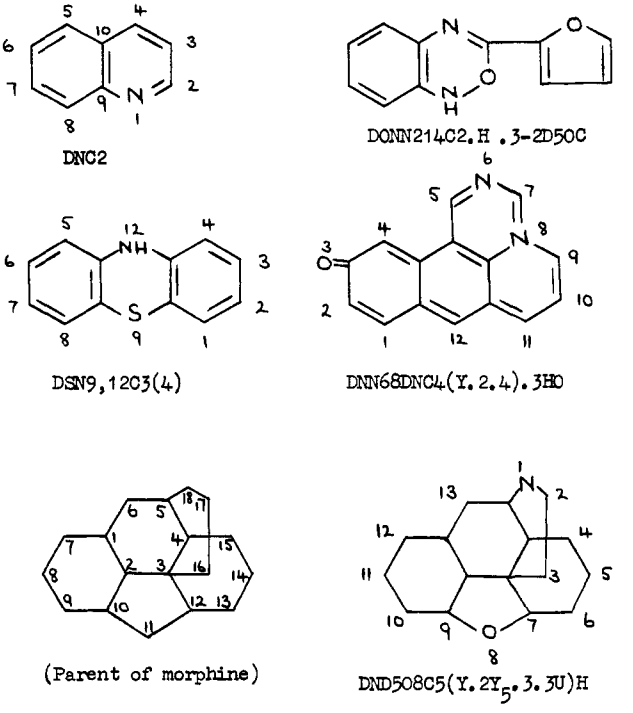


Fig. 9.--Heterocyclic systems and reversal of enumeration.

REFERENCES

(1) M. Gordon, C. E. Kendall and W. H. T. Davison, "Chemical Ciphering," Royal Institute of Chemistry, London, 1948.
(2) W. J. Wiswesser, "A Line-Formula Chemical Notation," T. Y. Crowell Co., New York, N. Y., 1954.
(3) A. P. Terent'ev, A. N. Kost, A. M. Zukerman and V. M. Potapov, Chem. Tech. (Berlin), 7, 579, (1955).
(4) F. L. Taylor, Ind. Eng. Chem., 40, 734 (1948).
(5) J. A. Silk, "A New System of Organic Notation," 1951, distributed privately.
(6) International Union of Pure and Applied Chemistry, "A Proposed International Chemical Notation," Longmans, Green and Co., 1958.
(7) G. M. Dyson, "A New Notation and Enumeration System for Organic Compounds," Longmans, Green and Co., 2nd edition, 1949, Chapter VIII.
(8) J. A. Silk, "A Linear Notation for Organic Compounds," 1961, (copies available on application to the author).