

PRINCIPLES OF A NOVEL NOMENCLATURE OF ORGANIC COMPOUNDS

Danail Bonchev

Department of Physical Chemistry, Higher School of Chemical Technology,
Burgas 8010, Bulgaria

Abstract - The main ideas of a universal nomenclature and coding of organic compounds are discussed on the basis of the recently developed generalized concept for a graph centre. In addition, a clusterization procedure is used in the most complicated cases. The new system is applicable to any kind of molecules that can be uniquely represented in a plane.

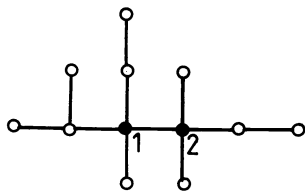
INTRODUCTION

The nomenclature of organic compounds is a fascinating field of research, due to the enormous variety of molecular structures. The contemporary computer-assisted molecular design expands more and more this variety of molecular forms. The pitfalls of the available nomenclatures (Ref. 1-5) are discussed in the literature (Ref. 6-9) and they prompt further efforts in the field. Two such pitfalls should be particularly emphasized: the lack of a general unified foundation for the nomenclatures of distinct classes of compounds, and, secondly, the lack of a general nomenclature which should be equally well suited to man and computer.

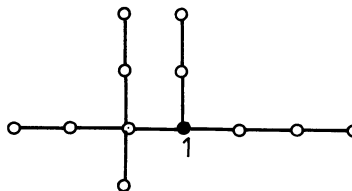
It seems quite natural to search for a more general solution of the nomenclature problems making use of graph theory (Ref. 9,10), as done in part in some recent publications (Ref. 8, 11-13). Structural formulae are known to be in essence chemical graphs (Ref. 14) where atoms are denoted by points (vertices) and bonds are denoted by lines (edges). Stereochemistry can also be taken into consideration using directed graphs, i.e. oriented edges (arcs). The present lecture provides some more general principles of organic compounds classification, nomenclature, and coding, reflecting the major topological features of chemical structures, such as cyclicity and branching, the way of linking two cycles (fusion, spiro-, and bridged linking), and first of all, the centric ordering of rings, atoms, and bonds in molecules. The ordering is carried out on the basis of the generalized concept developed recently for the graph centre (Ref. 15-17), as well as using standard lattices for the unique molecular representation. In what follows this approach is discussed starting with acyclic and going to more and more complicated cases of cyclic structures.

ACYCLIC COMPOUNDS

Alkanes. The molecular graphs of 2,3,4,4,5-pentamethyl-, 3-ethylhexane and 3,6-dimethyl-, 3,4-diethylheptane will be taken as examples illustrating the procedure.



(1)



(2)

According to the classical definition (Ref. 9) both vertices 1 and 2 in (1) should be regarded as central since their maximum distance (the number of bonds to the most distant vertex) is minimal: $d_1^{\max} = d_2^{\max} = 3$. In (2) the centre is a single point, denoted as vertex 1. This is a general result: the acyclic compounds having an even numbered chain have always bicentres while those with an odd-numbered chain have a single central point. Hence, for acyclic compounds we do not need a more general graph-centre definition. It seems quite sufficient to

use the two standard lattices shown below:

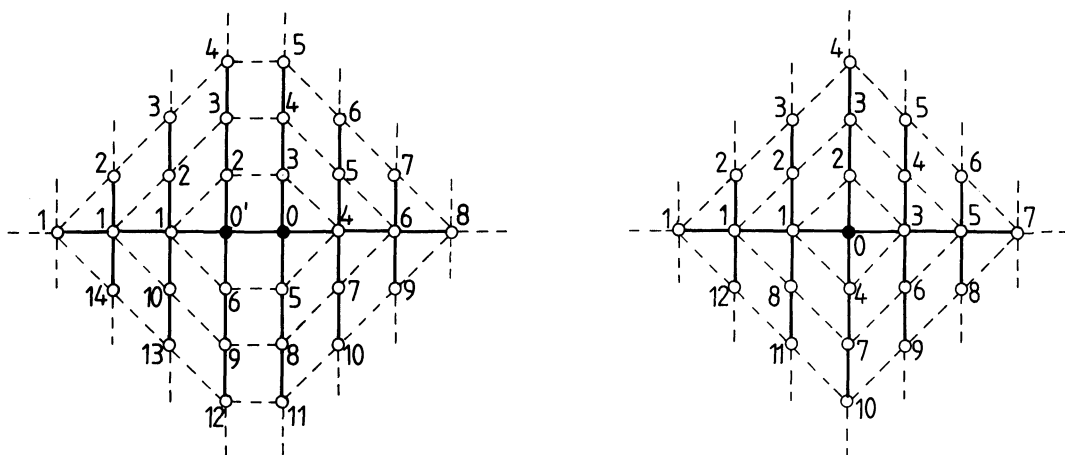
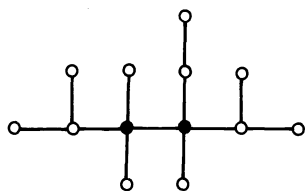
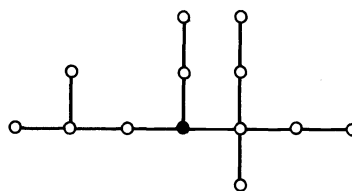


Fig. 1. Two standard lattices used in the nomenclatures of the compounds of even- and odd-number chain, respectively.

The molecular graphs are superimposed on the corresponding standard lattice so as to cover the central points. If several such possibilities exist they are all compared, preferring the one which provides the minimal numerical code. The latter is an ordered sequence of numbers starting from the first shell of neighboring vertices and counting clockwise then, after a slash, continuing with the second, third, etc. shells. Thus, for (1) and (2) one obtains /1-6/1-3,5,6, and /1-3/1-3,5,8/1,3,6,7, respectively. The alternative graphical representations (1') and (2') are shown below to have larger codes:



(1')

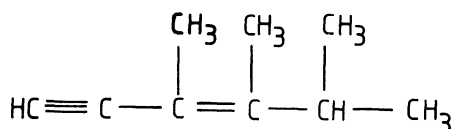


(2')

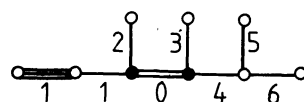
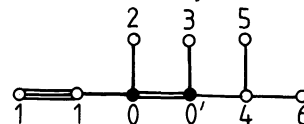
/1-6/1,2,4-6, and /1-3/1,3-6/1,2,5,7, respectively.

Within the approach under consideration the nomenclature and code of compounds are unified facing thus the need for an equally well computer and man-oriented nomenclature. The name of each compound consists of the numerical code, followed by the name, the same as used for alkanes, but denoting the total number of carbon atoms in the molecule. Thus, for (1) and (2) one ultimately arrives to the names: /1-6/1-3,5,6-tridecane and /1-3/1-3,5,8/1,3,6,7-tridecane.

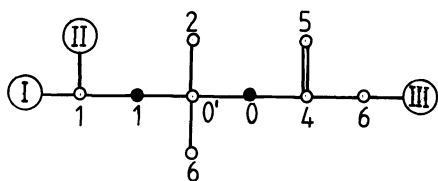
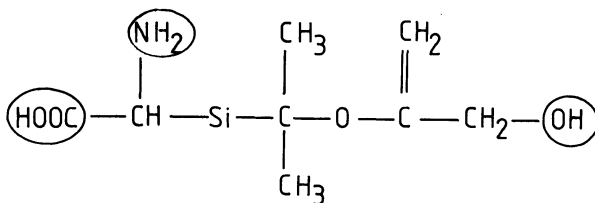
Unsaturated hydrocarbons. When multiple bonds are present, the name is supplemented by a string of symbols expressing the exact position of the double and triple bonds by numbers followed by the suffix -ene, and -ine, respectively. All bonds are canonically numbered by virtue of the standard lattices. An example follows:



/1-4/1,5,6-nonane, 0-ene, //1-ine

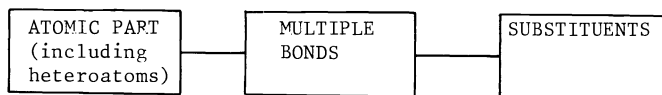


Compounds with heteroatomic chains and derivatives. The presence of a heteroatom in the main chain or in the side branches (but not as part of a functional group in terminal position) is denoted in the code by the atomic symbol of the chemical element, written after the number that shows the exact position of this atom in the standard lattice. In the case where several heteroatoms of the same kind are present their numbers are listed consecutively, writing then the atomic symbol. When several kinds of heteroatoms occur they are listed in lexicographic order. Dealing with substituted acyclic compounds the various functional groups, such as COOH, NH₂, OH, etc., are listed at the end of the compound name using again an alphabetic priority order. Below, for the example, this is the order: COOH < NH₂ < OH.



C-0,0'/1-Si,2,4,6/1,5,6-nonane,//5-ene,//1-COOH,NH₂,6-OH

Here the name "nonane" is used for the total number of non-hydrogen atoms in the non-substituted molecule, i.e. not carbon atoms like Si and O are also included. In this way the full name of an acyclic compound consists of three parts: the main part describing all atoms, and two other parts for the multiple bonds and substituents:

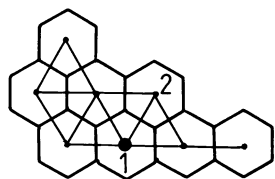


CYCLIC COMPOUNDS

General Remarks. The advantages of the system proposed for designation of organic compounds are more obvious when dealing with complicated cyclic structures. The name is similarly constructed starting with a string of symbols for the cyclic part, including here the available heteroatoms, then listing the multiple bonds, then the side chains and substituents. The complicating factors here are the different kind of linkage between the cycles, as well as the different cycle sizes. For these reasons the centric ordering of the rings, atoms, and bonds in the molecule can be uniquely done for compounds having cycles of the same size (homocyclic compounds), as well as for some relatively simple compounds having cycles of different size (called here heterocyclic compounds, a term used so far for another class of compounds which from the systematical viewpoint should rather be called cyclic heteroatomic compounds). For more complicated cases the centric approach is supplemented by a clusterization procedure which will be discussed later. It seems justified on the other hand to deal separately with the polycyclic compounds whose cycles are joined by the same type of linkage: fusion, spiro- or bridged one. The name of the compound will then be fusane, spirane, and bridgane, respectively. A special name "fuseene" may be reserved for the aromatic compounds having a delocalized π -electron system. Standard lattices (hexagonal, tetragonal, etc) are again used for the unique centric numbering of the rings, atoms, and bonds. The classical graph centre definition is, however, no longer sufficient and the generalized concept (Ref. 15,16), containing more detailed criteria, will be used.

Homocyclic hexofusanes and hexofuseenes. The problem of finding the central ring in these compounds, containing hexagonal fused rings, is solved by making use of the so-called dualist graphs (Ref. 6). Each vertex in these graphs represents the central point of a hexagon and two vertices are connected by a line when the corresponding cycles are fused. This is illustrated below by a hydrogen-suppressed molecular graph and a dualist graph superimposed on it.

According to the classical definition both vertices 1 and 2 in the dualist graph are central since their maximum distance is minimal: $d_1^{\max} = d_2^{\max} = 2$. The second criterion in the generalized graph centre definition (Ref. 15) is the minimum sum of the distances to all other vertices (otherwise, minimum distance rank, d_r). The third criterion requires a less frequent



(5)

occurrence of the largest distance(s). The occurrence numbers are usually expressed as powers to the respective distances:

$$\begin{aligned} \text{vertex 1: } & 1^4 2^3, \quad d_1 = 10 \\ \text{vertex 2: } & 1^3 2^4, \quad d_2^r = 11 \end{aligned}$$

As seen, both criteria 2. and 3 specify vertex 1 as a centre of the molecular dualist graph. In some more complicated cases the graph centre determination requires additional criteria. A computer programme is also available for such purpose.

After finding the central ring the structure is superimposed on the standard hexagonal lattice (Ref. 17) shown in Fig.2. If more than one way of superimposing is possible all ways are compared and the pictorial presentation of the molecule is selected that provides a minimal code. For structure (5) the minimal code is obtained: 1-4/2,3,7, while any other representation of this molecule will provide a large code. E.g. structure (5'):

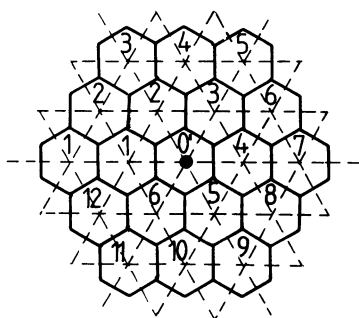
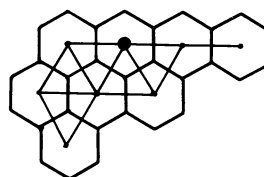


Fig. 2. Standard hexagonal lattice

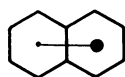


(5')

code:

1,4-6/7,11,12

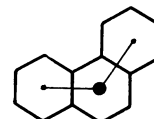
The name of the examined compound will be: 1-4/2,3,7-octahexofusane or, in general, the numericalcode presenting the topology of the cyclic part of the molecule is followed by a combined name containing the number of rings ("octa"), their size ("hexo"), and the type of of the compound (fusane). In case the above hydrogen-supressed molecular graph represents an aromatic compound, the same numerical code will be followed by the name "octahexofuseene". Some examples of well known benzenoid hydrocarbons follow:



1-dihexofuseene



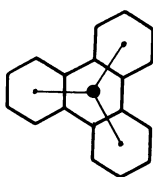
1,4-trihexofuseene



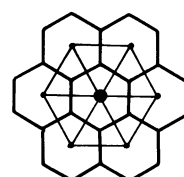
1,3-trihexofuseene



1,2-trihexofuseene



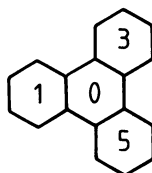
1,3,5-tetrahexofuseene



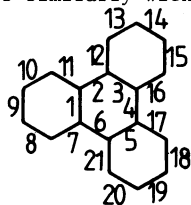
1-6-heptahexofuseene

Homocyclic hexafusanes having double bonds. The name of such a compound has a supplementary part where the numbers of the double bonds are listed in increasing order followed by the suffix -enes. This is possible since the unique numbering of the rings provides a unique (canonical) numbering of the bonds and atoms in the molecules. Both numberings are carried-out clockwise proceeding from that bond which connect rings 0 and 1. After finishing with the numbering in ring 0 one proceeds similarly with the next rings.

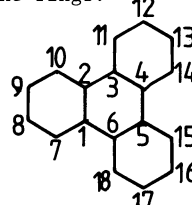
Examples:



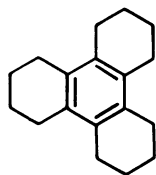
ring numbering



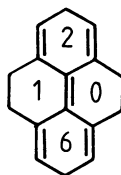
bond numbering



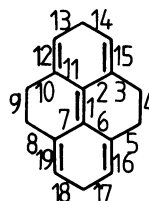
atom numbering



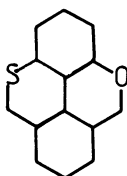
1,3,5-tetrahexofusane,1,3,5-triene



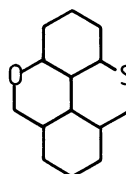
1,2,6-tetrahexofusane, 1,12,15,16,19-pentaene



Homocyclic hexafusanes and fuseenes having heteroatoms and side-chains. Utilizing the atom numbering introduced above all heteroatoms are listed in another supplementary part of the code in a lexicographic order after their respective numbers. If several equivalent graphical representations are possible, the minimal code condition is used again. The same holds for the third supplementary part, specifying in lexicographic order the substituents together with the numbers of the atoms they are connected with.

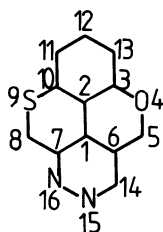


1,2,6-tetrahexofuseene,4-0,9-S

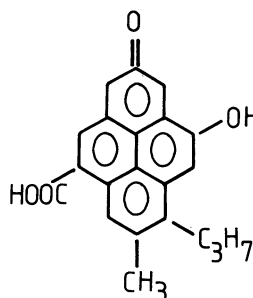


not

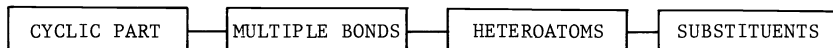
1,2,6-tetrahexofuseene,4-S,9-0



1,2,6-tetrahexofuseene,15,16-N,4-0,9-S

1,2,6-tetrahexofuseene,15-CH₃,14-C₃H₇,8-COOH,12-0,4-OH

In this way the full name of a homocyclic hexafusane having double bonds, heteroatoms, and side chains consists of four parts:



Other homocyclic fusanes. The same nomenclature principles, as those given in Subsections 1 to 4, are applicable to other homocyclic fusanes: tetrofusanes, trifusanes, etc. The standard lattices used in these cases are shown in Fig. 3:

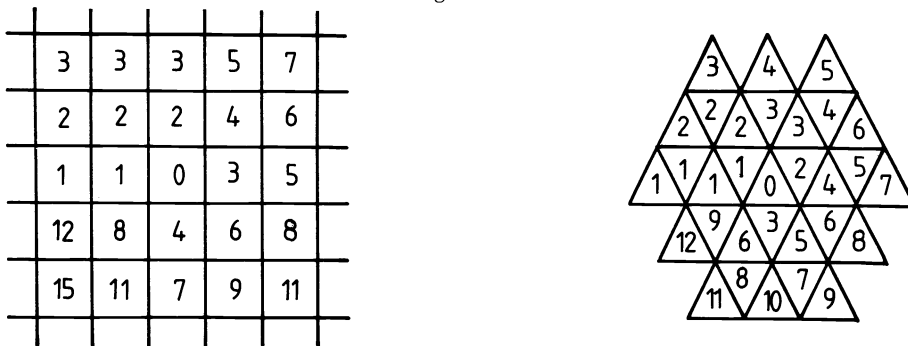


Fig. 3. Tetragonal and trigonal standard lattices and the cycle numbering in the centric shells of rings

Homocyclic spiranes and bridganes. Replacing the name "fusane" for "spirane" or "bridgane" one arrives at the names of homocyclic compounds whose cycles are all spiro-, or bridged linked, respectively. Evidently, the coding and nomenclature principles are again preserved. The standard lattices are more complicated, as shown in Fig. 4.

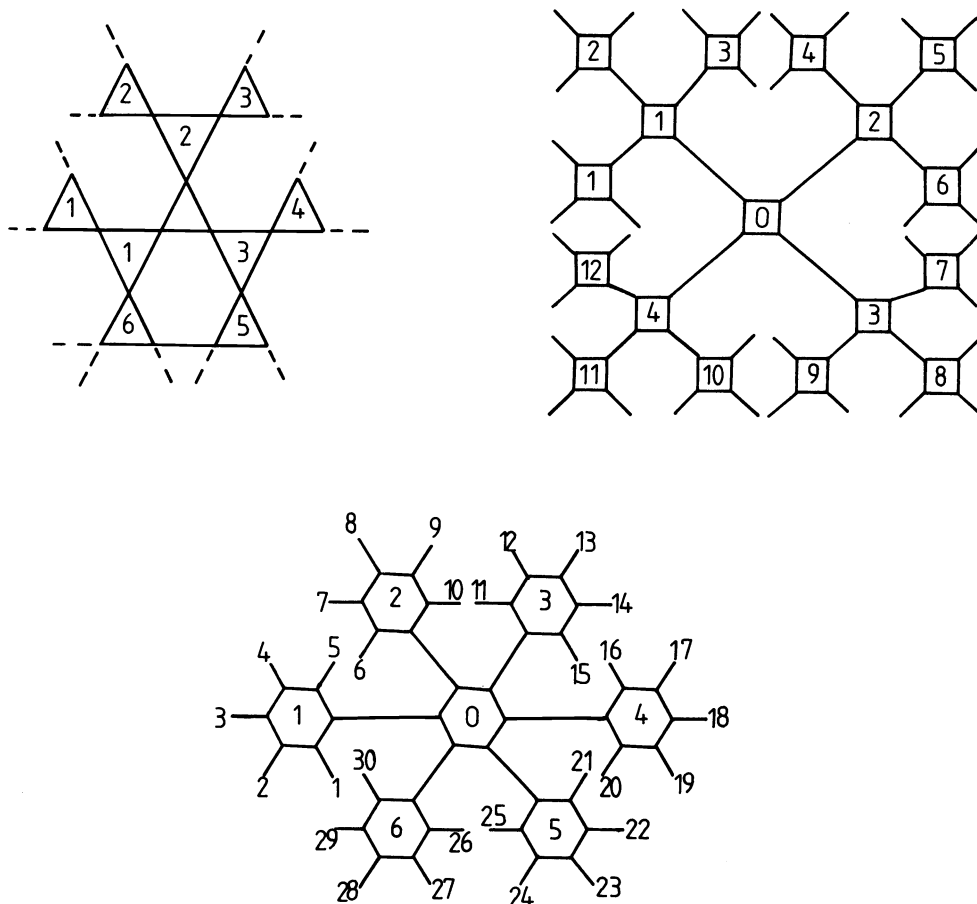
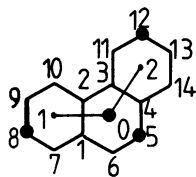


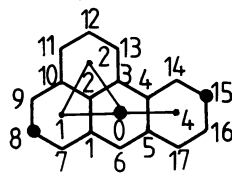
Fig. 4. Standard lattices for some homocyclic spiranes and bridganes.

CYCLIC COMPOUNDS WITH DIFFERENT TYPE OF CYCLE LINKAGE

When in a polycyclic compound the fusion, spiro- and bridged linking occur simultaneously the centric ordering of the cycles is no longer unique. A supplementary concept for the so-called "cluster" graph is introduced here. The points of this graph represent molecular fragments within which the type of cycle linking is the same. The centre of the cluster graph is easily determined obtaining also a centric ordering of the molecular fragments. This ordering holds when describing the full name of the compound, dividing the different fragments by semicolons, and denoting the fragments by Roman numerals. The description of each fragment includes its full name as described in the preceding Section. In addition, the linking with other fragments is pointed out by writing subsequently the number of atom or bond where the linkage occurs, then the number of the fragment, and finally one of the letters f, s, or b, for fusion, spiro-, and bridged linking, respectively. An example is given below where for simplicity no heteroatoms, side chains or multiple bonds are taken into consideration. The latter can be included whenever necessary in the same way, as discussed in the preceding sections.



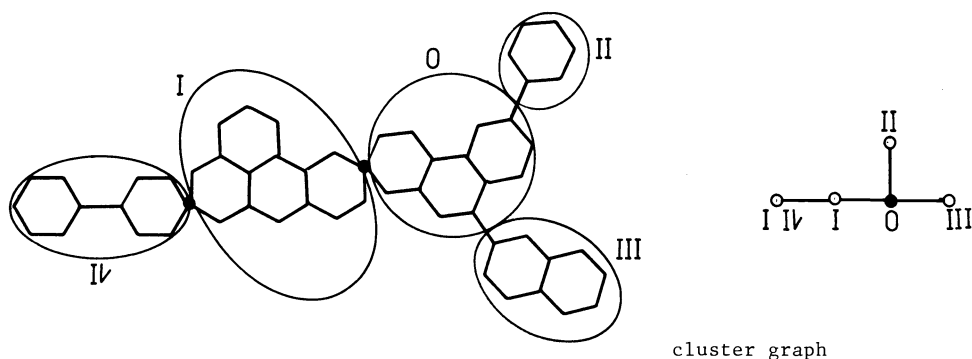
fragment 0
1,3-trihexofusane-
8Is, 12IIb, 5IIIb



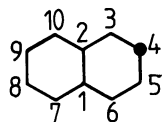
fragment I
1,2,4-tetrahexofusane
15(O)s, 8IVs



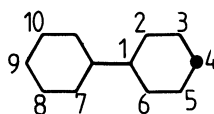
fragment II
0-hexofusane
1(O)b



hydrogen suppressed molecular structure having five fragments



fragment III
1-dihexofusane-4(O)b

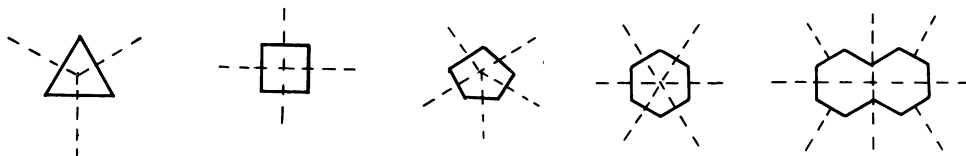


Fragment IV
1-dihexobridgane-4Is

Hence, the full name of the compound is: 2,4-trihexofusane-8Is,12IIb,5IIIb;I,1,2,4-tetrahexofusane-15(O)s,8IVs;II,0-hexofusane-1(O)b;III,1-dihexofusane-4(O)b;IV,1-dihexobridgane-4Is or more general: fragment 0 name + its connections with other fragments; fragment I name + its connections with other fragments; etc.

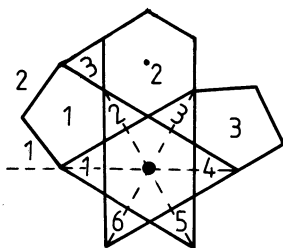
HETEROCYCLIC COMPOUNDS (COMPOUNDS WITH CYCLES OF DIFFERENT SIZE)

Proceeding from the assumption that every cycle, irrespective of its size, can be a central ring in somelattice with axes specified by the condition:

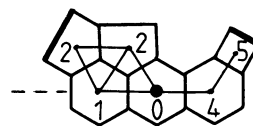


one arrives to the possibility of naming uniquely the simple compounds containing cycles of different size.

Examples:



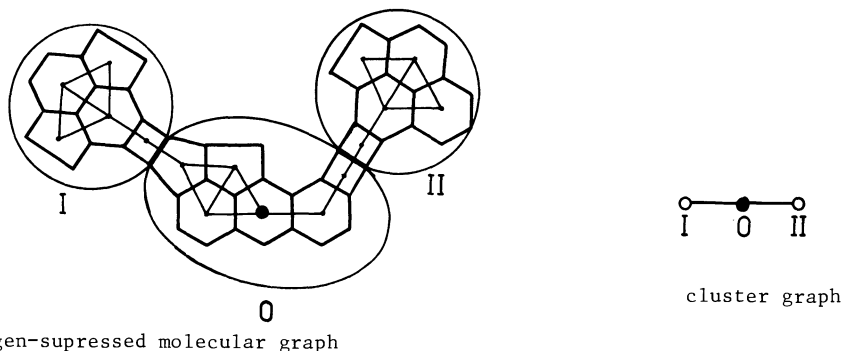
0-hexo/1-6-hexatrio/1,3-dipento,2-hexo/3-trio-fusane



0-hexo/1,4-dihexo,2-pento/2-pento,5-tetro-fusane

In more complicated cases, however, the neighbouring shells of cycles cannot be uniquely constructed. The necessity of a clusterization procedure arises again, in addition to the centric ordering of cycles, atoms, and bonds. All the cycles that cannot be uniquely positioned in one of the shells of neighbouring cycles could in principle be uniquely partitioned into fragments. Then one can proceed as in the preceding Section by constructing the cluster graph, the centric ordering of the fragments, and their subsequent description.

Example:



The full name of the above compound is : O-hexo/1,4-dihexo,2-pento/2-pento,5-tetro-fusane-21If,24If; I,0-hexo/1,2-dihexo,6-pento,4-tetro-fusane-17(O)f; II,0-hexo/1-tetro,3-pento,4,5-dihexo-fusane-8(O)f.

CONCLUDING REMARKS

This lecture presents a new approach towards a unified nomenclature and coding of organic compounds which should be equally well computer- and man-oriented. The two main principles used, namely the centric ordering of the rings, atoms, and bonds, and the fragmentation of the molecule into cluster graph, are of a general topological nature. They cover in principle the treatment of all organic compounds that are planar or can unambiguously be depicted as planar. One might also expect a further extension of the approach to the most complicated cage structures. The canonical centric numbering of the rings, atoms, and bonds is accessible for every chemist. He could also be helped by a programme applicable to the known programmable microcalculators. In conclusion, it should be stated that the present lecture does not deliver ultimate results. It is rather a programme for a series of studies undertaken in collaboration with Prof. A.T. Balaban (Bucharest) by the group of theoretical chemistry in Burgas, Bulgaria.

REFERENCES

1. R.S. Cahn, An Introduction to Chemical Nomenclature, 3rd ed., Butterworths, London(1968).
2. G.M. Dyson, A New Notation and Enumeration System for Organic Compounds, 2nd ed., Longman, London (1949).
3. IUPAC, Nomenclature Rules for Organic Compounds, Butterworths, London (1965).
4. A.M. Patterson, L.T. Capell, and D.F. Walker, The Ring Index, 2nd ed., American Chemical Society, Washington, DC (1960).
5. Index Guide Introduction, Chem. Abstr. 76 (1972).
6. A.T. Balaban and F.Harary, Tetrahedron 24, 2505-2516 (1968).
7. A.T. Balaban, Tetrahedron 25, 2949-2958(1969).
8. R.S. Read and P.S. Milner, Res. Report CORR 78-42, University of Waterloo, Canada(1978).
9. F.Harary, Graph Theory, Addison-Wesley, Reading, MA (1972).
10. R.J. Wilson, Introduction to Graph Theory, Longman, London (1972).
11. J. Lederberg, G.L. Sutherland, B.G. Buchanan, E.A. Feigenbaum, A.V. Robertson, A.M. Duffield, and C. Djerassi, J. Am. Chem. Soc. 91, 2973-2976 (1969).
12. R.C. Read, Res. Report CORR 80-7, University of Waterloo, Canada (1980).
13. N. Lozac'h, L.A. Goodson, and W.H. Powell, Angew. Chem. 91, 951-958 (1979).
14. A.T. Balaban, Ed., Chemical Applications of Graph Theory. Academic Press, London (1976).
15. D. Bonchev, A.T. Balaban, and O. Mekenyan, J. Chem. Inf. Comput. Sci. 20, 106-113(1980).
16. D. Bonchev, A.T. Balaban, and M. Randić, Int. J. Quantum Chem. 19, 61-82 (1981).
17. D. Bonchev and A.T. Balaban, J. Chem. Inf. Comput. Sci. 21, 223-229 (1981).