CROATICA CHEMICA ACTA CCACAA **60** (2) 269—277 (1987)

YU ISSN 0011-1643
CCA-1724 UDC 543
Original Scientific Paper

The Compact Nomenclature of the Benzenoid Hydrocarbons: A Short Review*

J. Ciosłowski

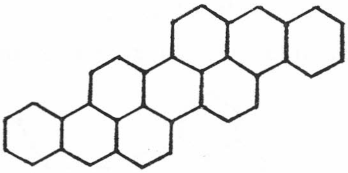
Department of Theoretical Chemistry, Jagiellonian University, 30060 Cracow, Poland

Received September 2, 1985

The short review concerning the recently proposed compact naming (CN) of the benzenoid hydrocarbons is presented. Since this nomenclature allows to construct the dualist of a particular hydrocarbon directly from its name, the connections between the structure of the dualist and various properties of hydrocarbon are discussed. In general, these properties fall into three categories according to their local, nodal or global dependence.

1. INTRODUCTION

At first sight it would seem that there is no reasons to work out a novel system of nomenclature for a class of compounds which just consists of about 500 individual molecules. Nevertheless, in the case of polycyclic benzenoid hydrocarbons this task is clearly justified. That is so because of the fact that these compounds are of a great interest for both theoretical and organic chemists[1,2]. The present situation in the field of the nomenclature of the benzenoid hydrocarbons is very unsatisfactory. In order to realize the typical deficiencies of the currently applied IUPAC nomenclature, let us consider the following example: in Chemical Abstracts[3] the hydrocarbon of the formula:



has an entry as benzo[rst]phenanthro[10,1,2-cde]pentaphene in Vol. 100. The same name is sustained in Vols. 99 and 98. But in Vol. 97 we can find this hydrocarbon under the name of dinaphtho[1,2,3-cd: 1',2',3'-lm] perylene; whereas in Vol. 72 the citations refer to isoviolanthrene. All these names are acceptable within the IUPAC rules of nomenclature.[4]

---

* Reported in part at the IUPAC *International Symposium on Applications of Mathematical Concepts to University* (Dubrovnik, September 2—5, 1985).

This example points out four shortcomings of the presently used naming system:

- i) lenghty names (the names of more than 50 letters are not so rare)
- ii) lack of any clear correspondence between the structural formula and the name of hydrocarbon
- iii) the evident inconsistence in usage of the names, which most certainly leads to mistakes and lost of some information while reviewing the chemical literature
- iv) the prevailing use of traditional names that are easily recognized by people, but explicitely unrecommended for use by digital computers.

To complete the above discussion, we should additionally turn the attention of the reader to the fact, that there are also other systems of nomenclature, namely the Beilstein's[5] and the Clar's[6], each one having the aforementioned drawbacks and each different from the other.

The various concepts of the rational naming of the benzenoid hydrocarbons, which are reported in literature[7,8,9], even appear to be of great importance in the enumeration and generation of polyhexes,[8,9] seem to offer no distinct improvement in the matter of searching for a nomenclature well suited both for human and computer purposes.

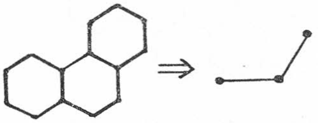
In 1984 we proposed[10] a novel approach to this problem, which is absolved from the hitherto prevailing faults.

The purpose of this paper is to sum up the algorithm, the rules, the advantages and the various applications of the novel compact naming (CN) system, that have been published or are considered in papers being currently in press. The intention of the author is to try to bring about wide utilization of the CN by chemist to simplify their research work on polycyclic hydrocarbons during scanning and storing of chemical data.

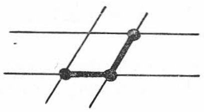
2. THE ALGORITHM[10,11]

The algorithm that enables to assign a compact name of the hydrocarbon to its structural formula is the following:

Step 1: Construct the dualist[2] corresponding to the structural formula:



Step 2: Place the dualist onto the smallest possible part of the paralelogram net and construct the matrix by replacing all the occupied nodes by 1 and all the unoccupied by 0:

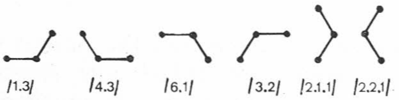


$$\Rightarrow \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$$

Step 3: Convert the matrix into the decimal code treating every its row as a binary number:

$$\begin{pmatrix}0 & 1\\ 1 & 1\end{pmatrix} \Rightarrow /1.3/ \qquad 01_2=1_{10} \quad 11_2=3_{10}$$

Step 4: Repeat the steps 1—3 for all the possible orientations of the structural formula:



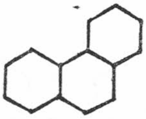
Step 5: Select the code with the shortest length and the smallest numbers, then add the core word defining the number of the benzene rings and the suffix-acene:

(1.3)-triacene

It should be accentuated that the selection rule used currently in the step 5 differs slightly from the one described in the original paper[10]. This makes no difference for the majority of the hydrocarbons' names, however in some cases produces codes containing smaller numbers.

3. THE RULES OF NOMENCLATURE[10]

1. The structural formula having the position corresponding to the selected code and thus to the compact name, is called the normal formula. The same name is assigned also to the decimal code:



/1.3/

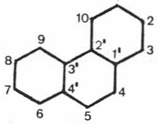
2. The rules for numbering the carbon atoms have been proposed in the original paper. However, we have found them to be rather unexpedient and tedious. So we present here some modified rules, which seem to be quite simple and convenient:

a) the system of carbon atoms is divided into two classes: the atoms linked with the hydrogens and the remaining ones.

b) the numbering of the atoms of the first subsystem begins on the top of the benzene ring lying in the highest row and in the position most to the left; then continues clockwise over all these atoms.

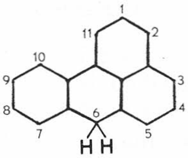
c) the numbering of the atoms of the second subsystem begins on the top of every ring and then proceeds in the clockwise manner. The priority for the rings is the following: the ring in the higher row is prior to the one in the lower row, for rings lying in the same row, the ring lying more to left is prior to the one lying more to the right. The carbon atoms once numbered are omitted. Arabic numbers with primes are assigned to the atoms that belong to the second subsystem and without primes to all the others.

Examples:



3. In the case of the radical molecules the parent system is recognized as that in which the carbon atoms are linked with single hydrogen atoms. The position of the additional hydrogen atoms is marked by the prefix in the compact name.

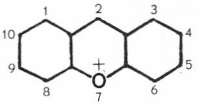
Example:



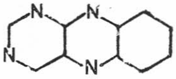
6H-/2.7/-tetracene

4. It is easy and straightforward to extend this nomenclature to the heteroanalogs of the polycyclic hydrocarbons. The position of the heteroatoms is marked by the appropriate prefix.

Example:



7-oxo-/7/-triacenylium cation



1,2,7,9-tetraza-/7/-triacene

4. THE ADVANTAGES AND SHORTCOMINGS OF THE CN SYSTEM

The simple algorithm of the CN enables us to convert the name into the structural formula and vice versa easily and directly. The algorithm allows to treat all the planar benzenoid hydrocarbons in the unique way, producing very short names that contain numeric code which can be utilized for computer data handling.

The shortocomings of the proposed method are the following: first, in the present version of the nomenclature the formally nonplanar molecules (such as helicene) cannot be named; secondly, the drawing of all the 12 possible orientations of the dualist in the parallelogram net can become tedious and error-prone in the case of large systems. The first drawback has been recently recovered by application of some additional rules to divide the nonplanar systems into parts[11]. The second one can be easily disposed when the computer-assisted generation of the normal formula is used. The appropriate package of the FORTRAN subroutines has been written[12]. The program reads the trial code, generates the structural formula and then constructs all the possible orientations in the net. Eventually, the proper code is selected and the compact name, as well as the graph representing the normal formula are printed.

5. THE APPLICATIONS OF THE CN SYSTEM

The formulated method of the CN offers diverse utilities which can be groupped into two general classes:

i) applications in which the possibility of fast construction of the dual graph enters a path to name/properties correlations;

ii) applications to store informations and to generate the structural formulae of benzenoid hydrocarbons.

First, we will focus our attention on the second aspect of the applications. The algorithmic character of the CN enables us to construct simple approaches to generate all possible polyhexes having a given number of the benzenoid rings. One of such algorithms, based on the computer simulated annellation of rings followed by deletion of the degenerated structure has been worked out and programmed on CDC machine.[13] The number of the found benzenoid hydrocarbons, which is presented in Table I agrees with the literature data.[8,9] The details concerning this program as well as timing of the execution will be published elsewhere.

TABLE I

The Number of the Rigolously Planar Polyhexes Generated by RINGGEN (ref. 13)

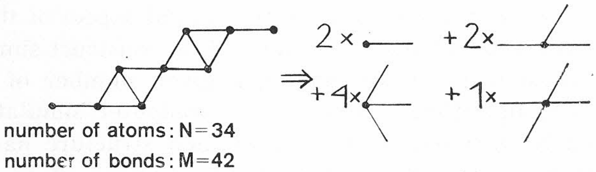
Number of rings	Number of polyhexes		
	with singlet ground state	with radical character	total
1	1	0	1
2	1	0	1
3	2	1	3
4	6	1	7
5	15	7	22
6	51	30	81
7	190	141	331
8	765	671	1436

Recently, in order to make the organic chemists more familiar with the task to catalogue all the known benzenoid hydrocarbons has been undertaken.[11] Over 400 structures have been identified on the basis of the data from Chemical Abstracts[3], Beilstein[5] and Clar's monography.[6] In all the cases the CN has been found to produce the shortest possible codes, which after a short practice are easily recognized even without any pencil and paper. During this work various FORTRAN programs have been written in order to facilitate the scanning of the accumulated data. In the future projects, the data-base for benzenoid hydrocarbons is to be worked out.

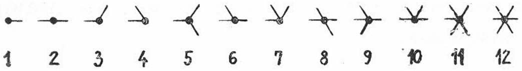
The CN, offering fast construction of the dualists, enters the path to name/properties correlations. Since diverse applications of the dualist have been recently summarized by Balaban[14], we will quote here only the novel relationships between the structure of the dualist and the molecular properties revealed by the recently developped method of nodes.

6. THE METHOD OF NODES[15,16,17]

The dualist obtained from the normal formula of a particular hydrocarbon can be further decomposed into the nodes reflecting the local structure of the benzene ring annellation:



There are only 12 possible kinds of nodes, which are shown in Figure 11.



The next step is a construction of a vector $\vec{n}$ having the components $n_1 \ldots n_{12}$; where n_j is the number of the nodes of the type j. Thus, for example, the vector $\vec{n}$ corresponding to the hydrocarbon from Figure 10 has a form (2, 0, 0, 0, 0, 2, 4, 1, 0, 0, 0, 0).

Then we can introduce the following definitions:

1.If a particular property P can be expressed with a high accuracy by the equation:

$$P \approx f(N, M) \qquad (1)$$

then P is called the locally dependent property.

2. If P can be expressed to a substantial degree of accuracy by the equation:

$$P \approx f(\vec{n}) \qquad (2)$$

then P is called the nodally dependent property, otherwise P is globally dependent.

3. A special case of the nodally dependent properties occurs, when f has a simple form such as:

$$P \approx \vec{g} \cdot \vec{n} \qquad (3)$$

Then P is called nodally additive and $\vec{g}$ is the vector of the nodal increments.

It should be outlined that the classification imposed by the definitions 1—3 is not very strict, because of some discrection of the accuracy required for eqs. 1—3. For instance, the total π-electron energy ($E\pi$) is nodally additive with the mean error of the order of 0.1%,[15] whereas the spectral radius of the adjacency matrix (R) corresponding to the molecular graph of hydrocarbon is nodally dependent with the error of about 0.3%[18]. Some typical examples of the various kinds of properties are presented in Table II.

TABLE II

The Dependence of Various Molecular Properties

Locally dependent properties
- molecular weight
- number of hydrogen atoms
- number of rings

Nodally dependent properties
- total π-electron energy
- spectral radius of the adjacency matrix
- topological resonance energy and number of Kekule structures[a]

Globally dependent properties
- energy of the HOMO orbital and the first ionization potential
- reactivity toward maleic anhydride
- carcinogenic activity

An interesting problem related to the method of nodes is an occurence of so called isonodal hydrocarbons[15] having the same vector $\vec{n}$. This can be treated as a generalization of the idea of the isoarithmic molecules[14]. Of course, it is expected, that all the nodally and locally dependent properties should not differe from one isonodal hydrocarbon to another. This is confirmed by the inspection of the available molecular data. In Table III we present some examples of the isonodal molecules·

7. CONCLUSIONS

The compact naming (CN) system developped for the benzenoid hydrocarbons is found to have a number of advantages over traditional nomenclature. Its use enables to establish a direct linkage between the names and the diverse properties of molecules. The method of nodes is one of the ways to construct such correlations. In the light of the facts presented in this paper, it seems to be highly justified to accept this novel nomenclature as the convenient standard.

---

[a] with the error of about 10%, whereas other nodally dependent properties are calculated with the accuracy of 0.1—0.3%.

TABLE III

Some Isonodal Molecules and Their Properties[a]

Molecule	/7.14/-hexacene	/1.3.7/-hexacene	/1.7.4/-pentacene	/9.7/-pentacene	/3.14/-pentacene
total π-electron energy (E_π)[b]	33.9277	33.9541	30.8805	30.8794	30.8390
topological resonance energy (TRE)[c]	0.844	0.861	0.807	0.806	0.781
spectral radius of the adjacency matrix (R)[b]	2.6035	2.6037	2.5212	2.5219	2.5251
energy of the HOMO orbital[b]	0.3027	0.3420	0.4735	0.4915	0.4048
reactivity toward maleic anhydride[d]	—	—	1.02	1.02	2.58
number of Kekule structures[b]	13	14	12	12	11
carcinogenic activity[e]	68	74	26	4	—

[a] see also Table II.

[b] ref. 19.

[c] ref. 20.

[d] measured as $6 + \log k_2$, where k_2 is the second-order rate constant for the Diels-Alder reaction carried out in boiling 1,2,4-trichlorobenzene (ref. 21).

[e] the Iball index (ref. 22).

Acknowledgment. — The assistance of Dr. A. M. Turek during preparation of this manuscript is gratefully acknowledged.

REFERENCES

1. I. Gutman, *Bull. Soc. Chim. Beograd* **47** (1982) 453.
2. O. E. Polansky and D. H. Rouvray, *Math. Chem.* **2** (1976) 63.
3. *Chemical Abstracts,* American Chemical Society, Columbus.
4. *Nomenclature of Organic Chemistry,* IUPAC, Butterworths, London, 1958. IUPAC *Nomenclature of Organic Chemistry, J. Amer. Chem. Soc.* **82** (1960) 5545.
5. *Beilsteins Handbuch der Organischen Chemie,* Springer Verlag, Heidelberg
6. *Polycyclic Hydrocarbons,* E. Clar, Academic Press, New York / Springer Verlag, Berlin (1964).

7. D. Bonchev, *Pure Appl. Chem.* **55** (1983) 221; S. B. Elk, *MATCH* **8** (1980) 121, **13** (1982) 239; D. Bonchev and A. T. Balaban, *J. Chem. Inf. Comp. Sci.* **21** (1981) 223; H. Wenchen and H. Wenjie, *Theor. Chim. Acta (Berlin)* **68** (1985) 301.
8. A. T. Balaban and F. Harary, *Tetrahedron* **24** (1968) 2505, **25** (1969) 2944.
9. J. V. Knop, K. Szymanski, Z. Jeričević, and N. Trinajstić, *J. Comput. Chem.* **4** (1983) 23.
10. J. Cioslowski and A. M. Turek, *Tetrahedron* **40** (1984) 2161.
11. J. Cioslowski and M. Wala, in preparation.
12. J. Cioslowski and A. M. Turek, *Computers & Chemistry,* in press.
13. J. Cioslowski, program RINGGEN, unpublished results.
14. A. T. Balaban, *J. Mol. Struct. (THEOCHEM)* **120** (1985) 117.
15. J. Cioslowski, *Theor. Chim. Acta (Berlin)* **68** (1985) 315.
16. J. Cioslowski, *Chem. Phys. Lett.* **12** (1985) 234.
17. J. Cioslowski, *MATCH,* in press.
18. J. Cioslowski, *Theor. Chim. Acta (Berlin),* in press; J. Cioslowski, *Chem. Phys. Lett.,* submitted.
19. *Supplemental Tables of Molecular Orbital Calculations,* A. Straitwieser and J. I. Brauman, Pergamon Press, Oxford 1965.
20. I. Gutman, M. Milun, and N. Trinajstić, *J. Amer. Chem. Soc.* **99** (1977) 1692.
21. D. Biermann and W. Schmidt, *Isr. J. Chem.* **20** (1980) 312.
22. L. von Szentpaly, *J. Amer. Chem. Soc.* **106** (1984) 6021 and references cited therein.

SAŽETAK

Kompaktna nomenklatura benzenoidnih ugljikovodika: kratki pregled

J. Ciosłowski

Dan je kratki prikaz nedavno predložene kompaktne nomenklature za benzenoidne ugljikovodike. Ova nomenklatura dopušta izravnu konstrukciju dualista za danu benzenoidnu strukturu.