

CCA-1930

YU ISSN 0011-1643

UDC 547.787

*Original Scientific Paper***Compact Molecular Codes for Annulenes, Aza-annulenes, Annulenoannulenes, Aza-annulenoannulenes, Cyclazines and Aza-cyclazines****Sonja Nikolić and Nenad Trinajstić**The Rugjer Bošković Institute, P. O. B. 1016,
41001 Zagreb, Croatia, Yugoslavia*

Received November 22, 1989

The application of the recently introduced concept of compact molecular codes to monocyclic and bicyclic systems such as annulenes and annulenoannulenes is presented. In addition, this concept is reformulated to include heteroatoms. The compact molecular code in a novel formulation is applied to monocyclic, bicyclic and tricyclic heterosystems such as aza-annulenes, aza-annulenoannulenes, cyclazines and aza-cyclazines.

INTRODUCTION

Compact molecular codes are structural codes which are called compact because they use a limited number of digits for encoding molecular structures.¹⁻⁵ They are based on the concept of the N-tuple code⁶ and satisfy all desirable requirements for molecular codes that are proposed by Read⁷ and Randić.¹

The aim of this article is to reformulate the concept of compact molecular codes to include heteroatoms.

For the sake of convenience, all chemical systems in this article will be depicted by the corresponding hydrogen-suppressed graphs. Molecular skeletons will be converted into graphs by replacing atoms with vertices and bonds with edges.⁸ Chemical graph-theoretical language will be used throughout the article.

There is a distinction to be made between a *code* as a concept and set of instructions for encoding and decoding, and a *code value* which is the set of symbols obtained for a particular example. In this report the term code will be used in both these senses.

This article is structured as follows. In the next section a brief description of the concept of compact molecular code is given. The third section contains compact molecular codes for annulenes. In the fourth section the extension of compact molecular codes to heterosystems containing a single heteroatom such as aza-annulenes is described. The fifth section contains com-

* Reported in part at WATOC '87 — *World Congress in Theoretical Organic Chemistry (Budapest: August 12 — 19, 1987)* and *American Chemical Society, 196th National Meeting — Symposium on Computational Graph Theory and Combinatorics (Los Angeles: September 25 — 30, 1988)*.

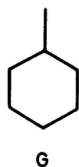
compact molecular codes for annulenoannulenes. The sixth, seventh and eighth sections of the article are devoted to compact molecular codes of aza-annulenoannulenes, cyclazines and aza-cyclazines, respectively, and in them some of ideas developed in the fourth section are further elaborated. The article ends with concluding remarks.

AN OUTLINE OF THE CONCEPT OF COMPACT MOLECULAR CODES

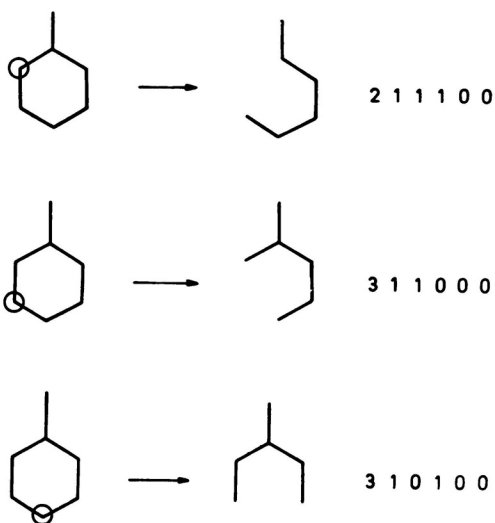
Compact molecular codes are the result of an extension of N-tuple codes to polycyclic systems.¹ The key step is to extract, in an economical way, a suitable acyclic part of a polycyclic system and to produce the N-tuple code for it. Then information on ring closure needs to be combined with the N-tuple code to produce the compact code of a polycyclic system under consideration. Thus, the compact molecular code consists of two parts: N-tuple part and label part. The label part is really a ring closure code, generated using an idea introduced by Herndon.⁹

The problem of finding an optimal way for extracting an underlying tree from a polycyclic system was solved by introducing the concept of subspanning trees (also referred to as spanning subtrees).¹ From now

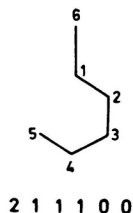
(a) Hydrogen-suppressed graph G representing benzyl



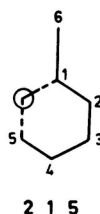
(b) Generation of subtrees of G and their N-tuple codes



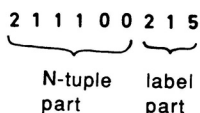
(c) Selected labelled subtree (one with a minimum N-tuple code)



(d) Coding a ring-closure



(e) Compact molecular code for benzyl

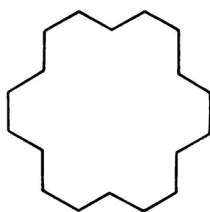


For complex structures the construction of the compact codes becomes rather involved. In order to speed up the search for the code a set of rules was proposed.^{1,3} The rules are ordered in a hierarchical manner, in order of preference.

COMPACT MOLECULAR CODES FOR [N]ANNULENES

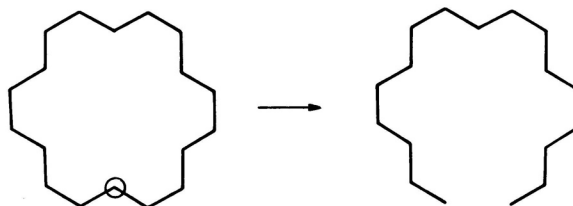
[N]annulenes C_NH_N are monocyclic unsaturated hydrocarbons with interesting properties.¹⁰ Compact molecular codes for [N]annulenes may simply be obtained by following the recipe from the previous section. In Figure 2 as an illustrative example we give the construction of a compact code for [18]annulene.

(a) Hydrogen-suppressed graph G representing [18]annulene

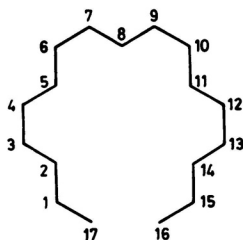


G

(b) Generation of a subtree of G

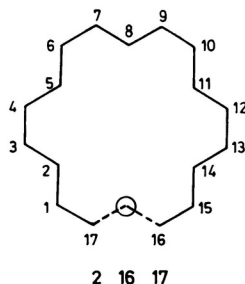


(c) The N-tuple code and labelling of the subtree



2 14x1 0 0

(d) Coding the annulene ring-closure



(e) Compact code for [18]annulene

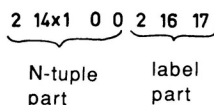


Figure 2. Compact molecular code for [18]annulene

Since [N]annulenes are regular monocyclic structures, their compact codes can be expressed in a general form:

(i) N-tuple code of [N]annulene subtrees

$$2 (N - 4) \times 1 0 0$$

(ii) Ring-closure code of [N]annulenes

$$2 (N - 2) (N - 1)$$

(iii) Compact code for [N]annulenes

$$2 (N - 4) \times 1 0 0 2 (N - 2) (N - 1).$$

The symbol $(N - 4) \times 1$ indicate repetitive occurrences of the digit 1 $(N - 4)$ times in a given compact molecular code.

COMPACT MOLECULAR CODES FOR AZA-ANNULENES

Aza-annulenes are regular monocyclic unsaturated structures containing one or more pyridine-type nitrogens.¹¹ We will consider here only aza-annulenes with a single nitrogen atom.

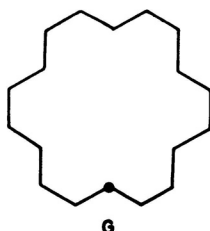
Aza-annulenes will be represented graph-theoretically by weighted cycles.⁸ In other words, the vertex representing the nitrogen atom will be singled out by being depicted as a black point. This vertex will be excised in the first step when a subtree of aza-annulene is being generated. Hence the only difference between compact code of [N]annulene and the corresponding aza-annulene will occur in the ring-closure code. Therefore, it seems that the simplest way to encode the ring-closure in aza-annulenes is by

$$N (N - 2) (N - 1)$$

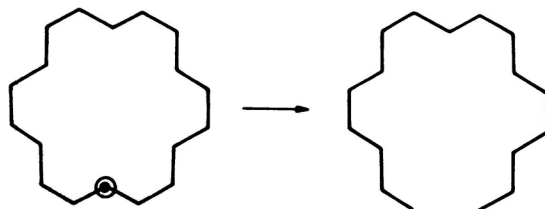
where **N** is the symbol for nitrogen. It is not necessary to indicate the valency of nitrogen because the labels after the symbol will do that impli-

citely. The number of labels after the symbol will specify the valency of the atom it represents. In Figure 3 we give as an illustrative example the construction of the compact code for aza[18]annulene.

(a) Hydrogen-suppressed weighted graph G representing aza[18]annulene

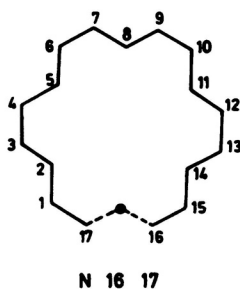


(b) Generation of a subtree of G



(c) The N -tuple code and labelling of the subtree
This step is identical to step (c) in Fig.2

(d) Coding of aza[18]annulene ring-closure



(e) Compact code for aza[18]annulene

2 14x1 0 0 N 16 17

Figure 3. Compact molecular code for aza[18]annulene

In the case of aza-annulenes is also possible to give a generalized compact code. The generalized compact molecular code for aza-annulenes with single nitrogen is given by

$$2(N-4) \times 100N(N-2)(N-1)$$

Similarly, the generalized compact code for [N]annulenes may also be given as

$$2(N-4) \times 100C(N-2)(N-1)$$

where **C** is the symbol for carbon. They differ only in one place, that is in the place occupied by the excised atom which is in the former case nitrogen and in the latter carbon. By introducing symbols instead of numbers for excised atoms into a code, the compact codes are no longer completely numerical but with this innovation the range of their applicability has considerably increased.

COMPACT MOLECULAR CODES FOR ANNULENOANNULENES

Annulenoannulenes are bicyclic unsaturated systems consisting of two annulene subunits glued together through one or more bonds.^{12,13} Compact molecular codes for annulenoannulenes can also be generated by following the procedure outlined in the second section. In Figure 4 we give as an illustrative example the construction of a compact code for [18]annuleno[18]annulene, a representative of [N]annuleno[N]annulene family in which subunits share one bond.

Since [N]annuleno[N']annulenes ($N \geq N'$) are regular structures a generalized compact code for them is also possible. It can be given as the following expression

$$2(N+N'-6) \times 1003(N-2)(N+N'-4)(N+N'-3).$$

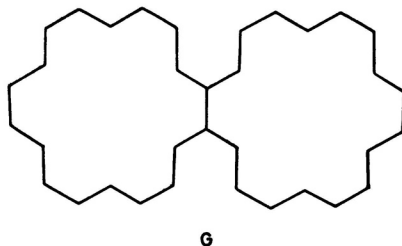
COMPACT MOLECULAR CODES FOR AZA-ANNULENOANNULENES

Aza-annulenoannulenes are hypothetical bicyclic systems consisting of aza-annulene and annulene subunits linked together. Compact molecular codes for aza-annulenoannulenes can be obtained by following ideas developed for aza-annulenes. Since in the case of aza-annulenoannulenes the pyridine-like nitrogen is placed in one of the annulene rings but not in one of the two positions shared by both rings, the simplest way to create the subtree is to erase a vertex in one of these two trivalent positions. The preference is given to the vertex to which the position of the nitrogen atom is closer. The generated subtree is weighted tree.⁸ We set the N-tuple for it in the same way as for the ordinary subtree except that we denote the position occupied by nitrogen as **N**. In Figure 5 we give as an illustrative example the construction of a compact code for aza[18]annuleno[18]annulene.

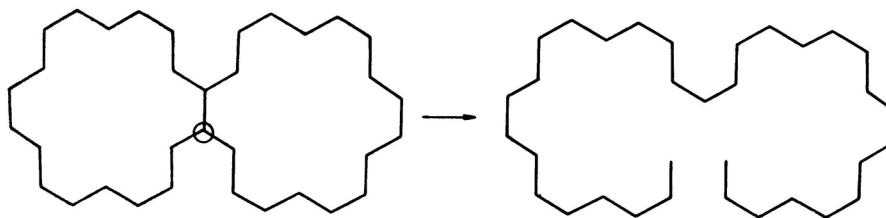
Since aza[N]annuleno[N']annulene ($N \geq N'$) are also regular structures (although not as regular as annulenoannulenes) a generalized form of the compact code for them is possible. It can be given as follows

$$2111\dots N\dots 111111003(N-2)(N+N'-4)(N+N'-3).$$

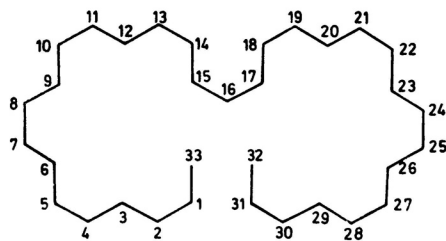
(a) Hydrogen-suppressed graph G representing [18]annuleno[18]annulene



(b) Generation of a subtree of G

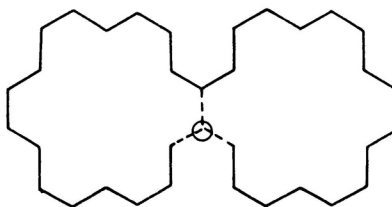


(c) The N-tuple code and labelling of the subtree



2 30x1 0 0

(d) Coding annulenoannulene ring-closure



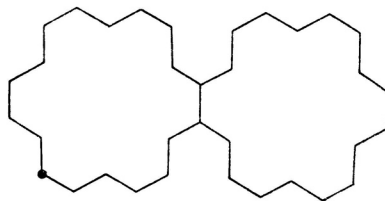
3 16 32 33

(e) Compact code for [18]annuleno[18]annulene

2 30x1 0 0 3 16 32 33

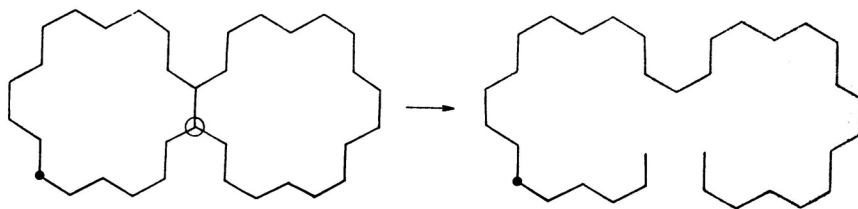
Figure 4. Compact molecular code for [18]annuleno[18]annulene

(a) Hydrogen-suppressed weighted graph G representing aza[18]annuleno[18]annulene

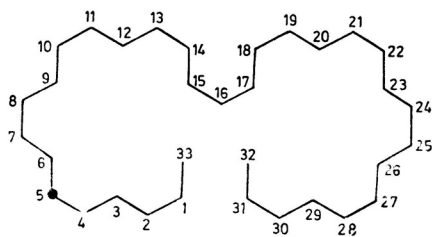


G

(b) Generation of a weighted subtree of G



(c) The N -tuple code and labelling of the weighted subtree



2 3x1 N 25x1 0 0

(d) Coding aza-annulenoannulene ring-closure

The same as step (d) in Fig. 4

(e) Compact code for aza[18]annuleno[18]annulene

2 3x1 N 26x1 0 0 3 16 32 33

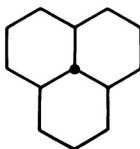
Figure 5. Compact molecular code for aza[18]annuleno[18]annulene

COMPACT MOLECULAR CODES FOR CYCLAZINES

Cyclazines are tricyclic systems which consists of a completely conjugated perimeter of sp^2 hybridized carbon atoms held planar by a central nitrogen atom bearing 2π electrons.¹⁴ In this case the subtree is obtained by removal of two atoms. We first remove the nitrogen atom, which reduces a tricyclic system to a monocyclic (annuleno-like) system. In the next step we remove a carbon atom and obtain the subtree. The selection of this vertex is dictated by the requirement that the code for the closure of the tricycle has the smallest possible labels. Because of this two-step reduction process, the ring-closure code must contain first information on the closure of the monocycle and then on the tricycle. In Figure 6 we give as an example the construction of the compact molecular code for cycl[3.3.3]azine.

Cyclazines are also regular structures and so it is possible to obtain a generalized compact code for them. We denote by N , N' and N'' the sizes of the three rings that make up the cyclazine system (see below structure I). Note that $N \geq N' \geq N''$. This is also based on the requirement that the closure of the tricycle contains the smallest possible labels.

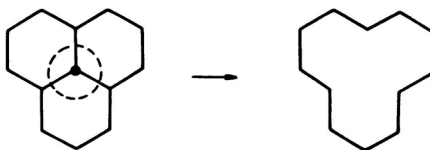
(a) Hydrogen-suppressed weighted graph G representing cycl[3.3.3]azine



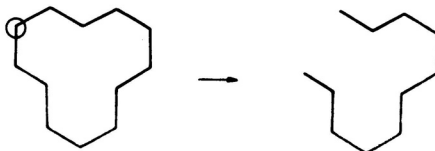
G

(b) Generation of a subtree of G

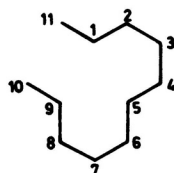
(b. 1) Removal of the weighted vertex



(b. 2) Removal of a convenient carbon atom



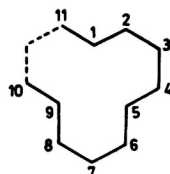
(c) The N-tuple code and labelling of the subtree of G



2 8x1 0 0

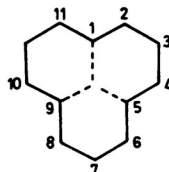
(d) Coding a ring-closure

(d. 1) Closure of a monocycle



2 10 11

(d. 2) Closure of a tricycle



N 1 5 9

(d. 3) Complete code for ring closures

2 10 11 N 1 5 9

(e) Compact code for cycl[3.3.3]azine

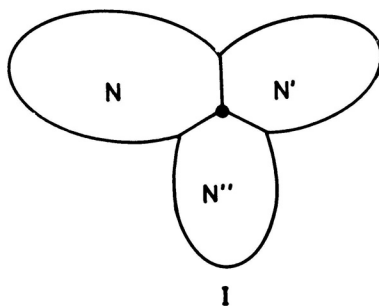
2 8x1 0 0 2 10 11 N 1 5 9

Figure 6. Compact molecular code for cycl[3.3.3]azine

The generalized compact molecular code for cyclazines is given by

$$2(N + N' + N'' - 10) \times 1 0 0 2(N + N' + N'' - 8) (N + N' + N'' - 7)$$

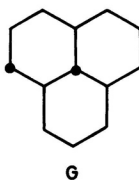
$$N 1 (N' - 1) (N' + N'' - 3)$$



COMPACT MOLECULAR CODES FOR AZA-CYCLAZINES

Aza-cyclazines are tricyclic cyclazine-like systems containing the pyridine-like nitrogen in one of the three rings.¹⁵ Here too the subtree is obtained after removal of two atoms in two steps. First, the nitrogen atom shared by all three rings is removed. This reduces the tricyclic system to the monocyclic (aza-annulene-like) system. In the next step we remove an additional

(a) Hydrogen-suppressed weighted graph G depicting 1-azacycl[3.3.3]azine

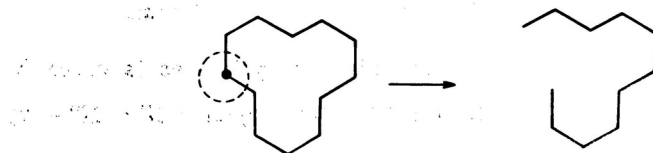


(b) Generation of a subtree of G

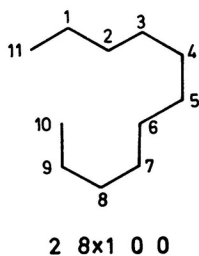
(b. 1) Removal of the central nitrogen atom



(b. 2) Removal of the nitrogen atom from the monocyclic (aza-annulene-like) system

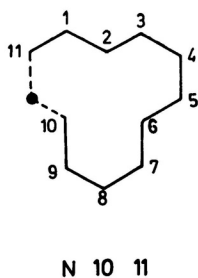


(c) The N-tuple code and labelling of the subtree of G

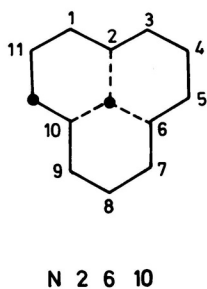


(d) Coding a ring-closure

(d. 1) Closure of a monocycle



(d. 2) Closure of a tricycle



(d. 3) Complete code for ring closures

N 10 11 N 2 6 10

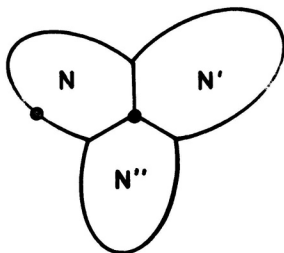
(e) Compact code for 1-azacycl[3.3.3]azine

2 8x1 0 0 N 10 11 N 2 6 10

Figure 7. Compact molecular code for 1-azacycl[3.3.3]azine

(pyridine-like) nitrogen and obtain the subtree. In Figure 7 we give as an example the construction of the compact molecular code for 1-azacycl-[3.3.3]azine.

It is also possible to obtain a generalized compact code for aza-cyclazines. We again denote by N , N' and N'' the sizes of the three rings that constitute a azacyclazine system (see below structure II). Ring labelled by N contains the pyridine-like nitrogen atom. Note that $N' \geq N''$.



II

The generalized compact molecular code for aza-cyclazines is as follows

$$2(N + N' + N'' - 10) \times 100N(N + N' + N'' - 8)$$

$(N + N' + N'' - 7)N$ (labels in this part of a ring-closure code depend on the position of the pyridine-like nitrogen in the ring N).

CONCLUDING REMARKS

In this article the concept of compact molecular codes is extended to some cyclic systems with heteroatom(s). The heteroatom in our case was nitrogen. Introduction of a heteroatom into a code is achieved by using the symbol of the heteroatom as its label. The code still remains compact, but is no longer completely numerical.

The interest in compact molecular codes is understandable since they have many nice features. For example, compact codes are simple to implement and do not require any prior experience, no information has been lost in the coding process, they are applicable to all kinds of hydrocarbon structures and also to many heteroatomic systems,¹⁶ derivation of compact codes implies a definitive labelling of atoms in a structure, etc. To our knowledge compact molecular codes are the simplest codes of general usability.

Acknowledgements. — This work was supported in part by the Croatian Science Fund and the Federal Committee for Development Grant P-339.

REFERENCES

1. M. Randić, *J. Chem. Inf. Comput. Sci.* **26** (1986) 136.
2. M. Randić, *Croat. Chem. Acta* **59** (1986) 327.
3. M. Randić, S. Nikolić, and N. Trinajstić, *J. Mol. Struct. (Theor. Chem)* **165** (1988) 213.

4. M. Randić, *J. Chem. Inf. Comput. Sci.*, in press.
5. J. V. Knop, W. R. Müller, K. Szymanski, S. Nikolić, and N. Trinajstić, in: *Computational Chemical Graph Theory*, edited by D. H. Rouvray, Nova, New York 1990, p. 9.
6. J. V. Knop, W. R. Müller, Ž. Jeričević, and N. Trinajstić, *J. Chem. Inf. Comput. Sci.*, **21** (1981) 91.
7. R. C. Read, *J. Chem. Inf. Comput. Sci.* **23** (1983) 135.
8. N. Trinajstić, *Chemical Graph Theory*, CRC, Boca Raton, FL 1983, Vol. I, Chapter 3.
9. W. C. Herndon, in: *Chemical Applications of Topology and Graph Theory*, edited by R. B. King, Elsevier, Amsterdam 1983, p. 231.
10. F. Sondheimer, *Pure Appl. Chem.* **7** (1963) 363; *Acc. Chem. Res.* **5** (1972) 81.
11. W. Gilb and G. Schröder, *Angew. Chem.* **91** (1979) 332.
12. T. M. Cresp and F. Sondheimer, *J. Amer. Chem. Soc.* **97** (1975) 4412; *ibid.* **99** (1977) 194.
13. M. Nakagawa, *Angew. Chem. Int. Edit. Engl.* **18** (1979) 202.
14. R. J. Windgassen, Jr., W. H. Saunders, Jr., and V. Boekelheide, *J. Amer. Chem. Soc.* **81** (1959) 1459.
15. V. Boekelheide and A. Miller, *J. Org. Chem.* **26** (1961) 431.
16. M. Randić, *Croat. Chem. Acta*, submitted.

SAŽETAK

**Kompaktne molekularne šifre anulena, aza-anulena, anulenoanulena,
aza-anulenoanulena, ciklazina i aza-ciklazina**

S. Nikolić i N. Trinajstić

Nedavno uvedena koncepcija kompaktne molekularne šifre primijenjena je na monocikličke i bicikličke sustave kakvi su anuleni i anulenoanuleni. Ta je koncepcija također proširena na heterocikličke spojeve. Tako je kompaktna molekularna šifra u novoj formulaciji primijenjena na monocikličke, bicikličke i tricikličke heterosustave kao što su aza-anuleni, aza-anulenoanuleni, ciklazini i aza-ciklazini.