

## On Canonical Numbering of Carbon Atoms in Fullerenes: C<sub>60</sub> Buckminsterfullerene\*

Dejan Plavšić,<sup>a,\*\*</sup> Damir Vukičević,<sup>b,\*\*</sup> and Milan Randić<sup>c,\*\*</sup>

<sup>a</sup>The Ruđer Bošković Institute, P.O. Box 180, HR-10002 Zagreb, Croatia

<sup>b</sup>Department of Mathematics, University of Split, Nikole Tesle 12, 21000 Split, Croatia

<sup>c</sup>National Institute of Chemistry, P.O. Box 3430, 1001 Ljubljana, Slovenia

RECEIVED FEBRUARY 14, 2005; REVISED MAY 3, 2005; ACCEPTED MAY 3, 2005

*Key words*  
fullerenes  
buckminsterfullerene  
canonical labeling  
canonical adjacency matrix

Numbering of atoms in relatively large molecules, such as fullerenes appears for most part to be arbitrary or based on *ad hoc* schemes. We argue in favor of the use of a particular canonical labeling of atoms in molecules based on the smallest possible binary molecular code obtained from the adjacency matrix when its rows are read from left to right and from top to bottom. The approach has been illustrated with buckminsterfullerene. We have outlined advantages of the approach and have shown that finding canonical labels is practical even in the case of large regular graphs.

### INTRODUCTION

As is well-known numbering of atoms in molecules is an important subject that still, from time to time, is revisited. There are two basically different approaches that we may label as »traditional« and »computational.« The former follows conventional agreements approved by influential international bodies, like IUPAC,<sup>1</sup> or dominant abstract services, like CAS,<sup>2</sup> which can be characterized by rules that are plausible but lacking mathematical content. In contrast »computational« approaches are characterized by algorithms that are mathematically well-founded, and tend to be susceptible to computer processing.<sup>3</sup> The outcomes of some of these graph-theoretically based approaches could be less user-friendly, but they have some other advantages. For example, they facilitate enumeration of isomers, as described by Balaban and Harary,<sup>4</sup> or

lead to systematic classification and nomenclature as outlined by Balaban and Schleyer on diamond hydrocarbons.<sup>5</sup> As already mentioned, in contrast to the traditional numbering of carbon atoms when non-traditional numberings, like the canonical labeling to be considered in this article, are adopted, they may show a »chaotic« distribution of labels, as illustrated in Figure 1 with a set of molecular graphs of smaller organic compounds. This chaotic appearance of labels may have been one of the reasons for the slow acceptance of canonical labeling of atoms in molecules. However, as we will argue in this contribution, when one considers the benefits and the advantages of mathematically based labels for atoms in comparison with traditional approach, which is further burdened with plethora of trivial names used in chemistry, one cannot escape conclusion that the future is in favor of the nov-

\* Dedicated to Dr. Edward Kirby on the occasion of his 70<sup>th</sup> birthday.

\*\* Authors to whom correspondence should be addressed. (M.R: Permanent address: Kingman Rd. Ames, IA50014, USA; E-mail: dplavsic@irb.hr, vukicevi@pmfst.hr, mrandic@msn.com)

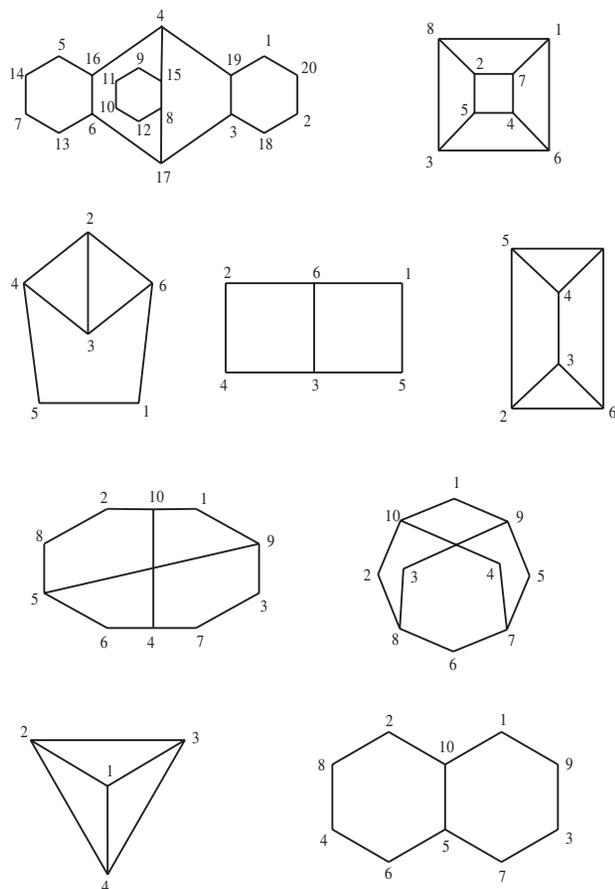


Figure 1. Illustration of the canonical labeling of the vertices of molecular graphs of organic compounds listed in Table I.

elty. In Table I we show a selection of trivial names and the corresponding Chemical Abstracts names used in organic chemistry for some smaller cyclic and polycyclic compounds. We direct readers to the book of Hellwinkel<sup>6</sup> on systematic nomenclature of organic chemistry where one can find numerous additional examples. To make things worse, not only that some chemical names tend to be revised from time to time but the leading institutions like CAS and IUPAC have adopted for a selection of compounds different names.<sup>6</sup> Moreover, the complexity of rules for naming compounds often results in highly cumbersome names. For example, lovastatin, a potent inhibitor of a rate controlling enzyme in cholesterol biosynthe-

TABLE I. A selection of trivial names used in organic chemistry and the corresponding Chemical Abstracts names

Trivial name	Chemical Abstracts name
Triptycene	9,10-Dihydro-9,10-o-benzoanthracene
Cubane	Pentacyclo[4.2.0.0.2 <sup>5</sup> .0 <sup>3,8</sup> .0 <sup>4,7</sup> ]octane
Benzvalene	Tricyclo[3.1.0.0 <sup>2,6</sup> ]hex-3-ene
Dewar benzene	Bicycle[2.2.0]hexa-2,5-diene
Prizmane	Tetracyclo[2.2.0.0.2 <sup>6</sup> .0 <sup>3,5</sup> ]hexane
Twistane	Tricyclo[4.4.0.0 <sup>3,8</sup> ]decane
Adamantane	Tricyclo[3.3.1.1 <sup>3,7</sup> ]decane
Tetrahydrane	Tricyclo[1.1.0.0 <sup>2,4</sup> ]butane
Decalin	Bicycle[4.4.0]decane

sis, has Chemical Abstracts name reading as follows: (2*S*)-2-Methylbutanoic acid (1*S*,3*R*,7*S*,8*S*,8*aR*)-1,2,3,7,8,8*a*-hexahydro-3,7-dimethyl-8-[2-(2*R*,4*R*)-tetrahydro-4-hydroxy-6-oxo-2*H*-pyran-2-yl][ethyl]-1-naphthalenyl] ester. In Table II we give a few additional examples that should suffice to counterbalance opposition that some may have against mathematical approaches to chemical nomenclature. The illustrations listed were found on the first few pages of the Merck Index,<sup>7</sup> which shows that considerable number of compounds in the Merck Index (which lists over 10,000 compounds) is associated with lengthy names that require 3–4 lines to be fully listed. In addition, there are compounds for which yet official names have not been proposed (or could not be proposed?). Just to illustrate the point consider abamectin, which belongs to antiparasitic antibiotics, the first substance listed in the Merck Index. Looking at Table II one cannot escape recalling a witty comment of Sir Winston Churchill regarding the work of committees, when stating that dromedary is a horse designed by a committee!

## CANONICAL LABELING OF ATOMS IN MOLECULES

There are alternative canonical rules for labeling atoms in molecules, some being restricted to subclass of structures (such as acyclic systems) and other being more general. We will consider here a particular canonical num-

TABLE II. Some organic compounds described by their trivial names, empirical formulas, and cumbersome Chemical Abstracts names

Trivial name and empirical formula	Chemical Abstracts name
Abietic Acid C <sub>20</sub> H <sub>30</sub> O <sub>2</sub>	[1 <i>R</i> -(1 <i>α</i> , 4 <i>aβ</i> ,4 <i>bα</i> ,10 <i>aα</i> )]-1,2,3,4,4 <i>a</i> ,4 <i>b</i> ,5,6,10,10 <i>a</i> -Decahydro-1,4 <i>a</i> -dimethyl-7-(1-methylethyl)-1-phenanthrenecarboxylic acid
Absinthin C <sub>30</sub> H <sub>40</sub> O <sub>6</sub>	[3 <i>S</i> -(3 <i>α</i> ,3 <i>aα</i> ,6 <i>β</i> ,6 <i>aα</i> ,6 <i>bβ</i> ,7 <i>α</i> ,7 <i>aβ</i> ,8 <i>α</i> ,10 <i>aβ</i> ,11 <i>β</i> ,13 <i>aα</i> ,13 <i>bα</i> ,13 <i>cβ</i> ,14 <i>bβ</i> )]-3,3 <i>a</i> ,4,5,6,6 <i>a</i> ,6 <i>b</i> ,7,7 <i>a</i> ,8,9,10,10 <i>a</i> ,13 <i>a</i> ,13 <i>c</i> ,14 <i>b</i> -Hexadecahydro-6,8-dihydroxy-3,6,8,11,14,15-hexamethyl-2 <i>H</i> -7,13 <i>b</i> -ethenopentaleno-[1'',2'':6,7;5'',4'':6',7']dicyclohepta[1,2-b:1',2'-b']difuran-2,12-(11 <i>H</i> )-dione
Acarbose C <sub>25</sub> H <sub>43</sub> NO <sub>18</sub>	<i>O</i> -4,6-Dideoxy-4-[[[1 <i>S</i> -(1 <i>α</i> ,4 <i>α</i> ,5 <i>β</i> ,6 <i>α</i> )]-4,5,6-trihydroxy-3-(hydroxymethyl)-2-cyclohexen-1-yl]amino]- <i>α</i> -D-glucopyranosyl-(1→4)- <i>O</i> - <i>α</i> -D-glucopyranosyl-(1→4)-D-glucose

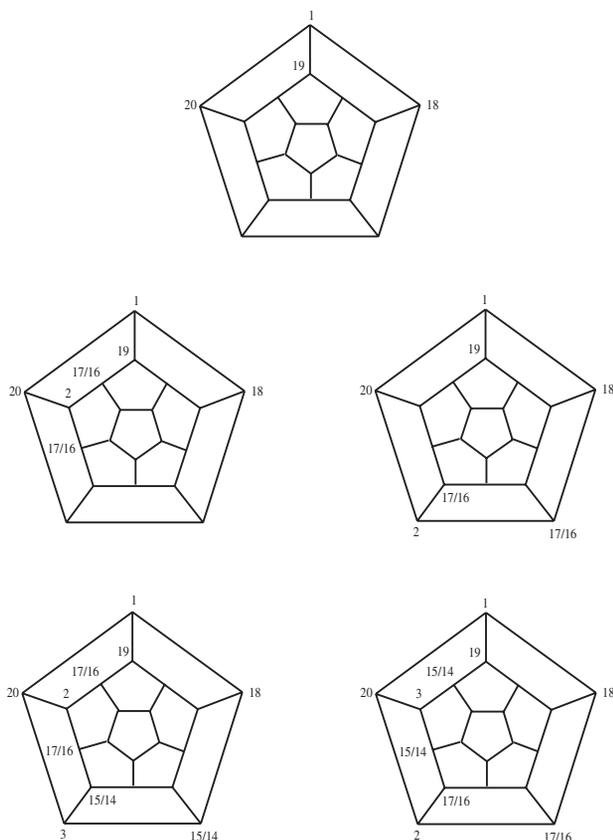


Figure 2. The initial steps in the assignment of the canonical labels to the vertices of the graph representing dodecahedron.

bering of atoms in a molecule that results in a unique adjacency matrix, called the canonical adjacency matrix, whose rows when viewed as binary numbers produce the smallest such numbers possible. The approach has been introduced by one of the present authors some time ago,<sup>8</sup> and was used to investigate symmetry properties of smaller molecules<sup>9</sup> as well as rather involved graphs associated with degenerate rearrangements.<sup>10</sup> The concept of canonical labels has also been used for orderly generation of classes of graphs, such as cubic graphs.<sup>11</sup> The approach will be exemplified by a graph representing dodecahedron (Figure 2), which can be viewed as the graph of the smallest fullerene.

We want to assign labels 1–20 to the twenty vertices of the graph, such that its adjacency matrix results in the smallest binary number possible when its rows are read from left to right and from top to bottom. Consequently, the first row of canonical adjacency matrix has to be the smallest possible binary number, which for trivalent graph on 20 vertices is: 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1. This means that the vertex with label 1 (the first row of the matrix) is adjacent to the vertices with labels 18, 19, and 20 (the last three columns), which is illustrated at the top of Figure 2. In other words, the vertex having the smallest label has as the nearest neighbors the vertices with the greatest labels. Of course, because all the verti-

ces of dodecahedron are equivalent we could have put the label 1 at 20 different sites, one of which is illustrated. Equally the label 20 we could have put on three locations next to vertex 1, one of which is illustrated in Figure 2. Finally, the label 19 has two alternative locations, one of which is illustrated in Figure 2. This determined the site for the label 18 as the remaining »vacant« site next to the vertex labeled with label 1. Hence, all in all there are  $20 \times 3 \times 2 = 120$  possibilities already in the first step in the assignment of labels, but all of these 120 possibilities produce the same entries for the first row of the adjacency matrix.

We continue with considering the location of label 2. The vertex labeled with 2 has to have as few unlabeled nearest neighbors as possible and they have to be labeled with the greatest of the remaining unused labels. These two requirements assure that the second row in the adjacency matrix is the smallest binary number among the binary numbers that can result from labeling these vertices with unused labels. Clearly, we have to place label 2 next to label 20. The remaining two neighbors of the vertex 2 will have labels 17 and 16, but in order not to proliferate the number of combinations to be considered we will postpone full assignment and will only »reserve« labels 17 and 16 for the sites next to the vertex 2, as illustrated in the middle of Figure 2, in which two alternative locations for vertex 2 are shown. In the next step, illustrated at the bottom of Figure 2, we have placed label 3 to the other neighbor of vertex 20, being the optimal site to produce the smallest binary number possible for the third row. One should realize that a graph with 20 vertices has a total of 20! possible assignments of labels to the vertices (a horrendous number  $2.43290 \times 10^{18}$ ), which even when symmetry is taken into account amount to a number of the order of  $10^{14}$ . Yet as we will see, with the adopted postponement of determining the sites for some of the greater labels, it takes about dozen steps to arrive at the unique canonical labels for the vertices. We carry on the assigning of canonical labels to the unlabeled vertices of the graph with label 4, as illustrated in Figure 3 at the top. The neighbors of vertex 4 are the vertices labeled with 12, 13, and 19 as shown in Figure 3. At this stage in the graph there is a single vertex, which has already two labeled neighbors which should be labeled with 5. Its unlabeled neighbor has to be labeled with 11. By placing label 5 we can immediately resolve the ambiguities concerning the neighbors of vertex 2 and 3, the labels 13/12, and 17/16 that were introduced earlier but not resolved. As is shown in the middle of Figure 3 the canonical labels of the nearest neighbors of vertex 5 are 11, 13, and 17. We still have unresolved ambiguity concerning labels 15 and 14, which are the greatest labels having »vacant« neighbor site, which allow placing label 6 in two alternative sites, as illustrated at the bottom of Figure 3. Once label 6 is assigned, in order that the

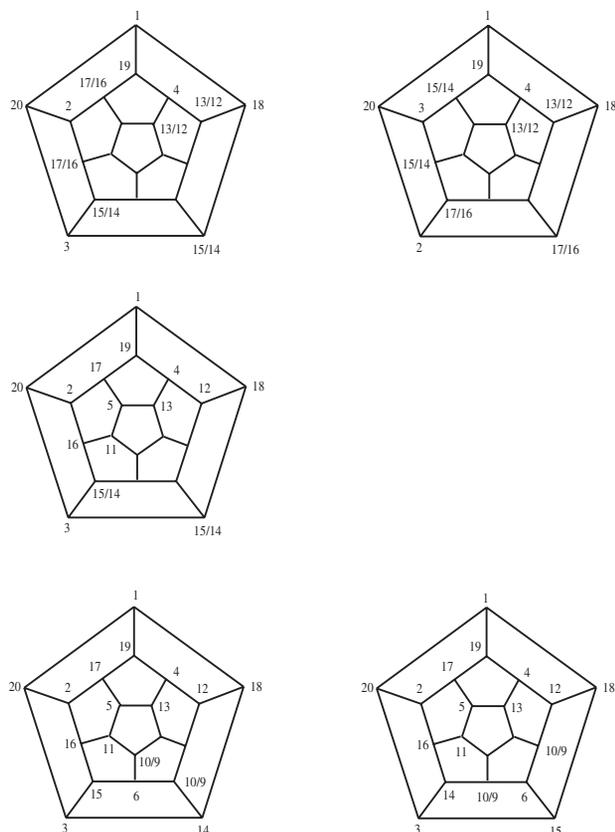


Figure 3. The intermediate steps in the assignment of the canonical labels to the vertices of the graph representing dodecahedron.

sixth row of the adjacency matrix produces the smallest binary number possible we have to assign label 15 to its neighbor. The next smallest labels 7 and 8 produce two alternative assignments shown in Figure 4, which produce different matrices. The labeling shown at the right corresponds to the smallest binary code since the vertex labeled with 7 has neighbors labeled with 8, 10, and 13 while the assignment at the left assigns to the very same vertices labels 8, 10, and 11, which clearly produces a larger binary number for the seventh row. As we have seen the search for canonical labels associated with the canonical adjacency matrix is practical, that is it can be accomplished in short time, even for sizable structures, such as dodecahedron. Observe that although at some

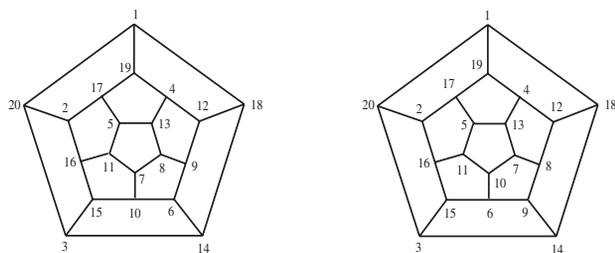


Figure 4. The final stage of the search for the canonical labels of the vertices of the graph representing dodecahedron. The structure at the right has the correct assignment of canonical labels.

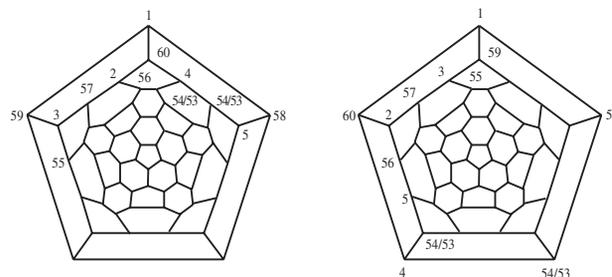


Figure 5. The initial steps in the search for the canonical labels of the vertices of the graph representing buckminsterfullerene.

steps we had to consider more than one possible assignment after the assignment of the initial labels 1, 20, and 19 (associated with 120 possible alternative selections) there was no additional degree of freedom to introduce additional alternative combinations.

### CANONICAL LABELING OF ATOMS, BONDS, AND RINGS IN BUCKMINSTERFULLERENE

Buckminsterfullerene,  $C_{60}$ , has a total of  $60!$  different assignments of labels to its atoms, but it may be surprising that, when one takes into account its symmetry, finding the canonical labels for its 60 atoms is even simpler than was the case of dodecahedron with only 20 vertices. As one can see from Figure 5, where the initial steps in the search for the canonical labels in  $C_{60}$  is illustrated, after placing label 1 to any of the 60 symmetry equivalent vertices we have two symmetry non-equivalent sites for label 60, which are shown at the top of Figure 5. Following with the assignment of labels 2, 3, and 4 we find out that the smallest value for the fifth row of the adjacency matrix comes when label 5 is placed between ver-

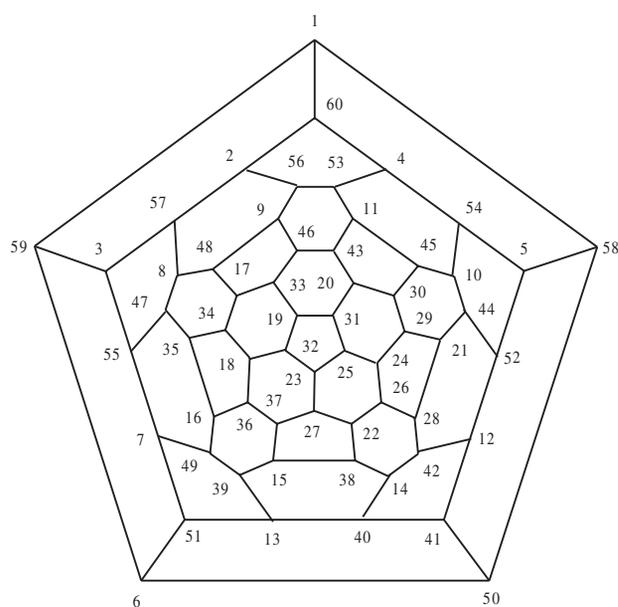


Figure 6. Canonically labeled vertices of the graph representing buckminsterfullerene.

tices with labels 54 and 58 (as shown at the left  $C_{60}$  diagram) rather than between 54 and 56 (as shown at the right  $C_{60}$  diagram). After selecting the site for label 5 the labels for all yet unlabeled sites are predetermined and as a result we obtain canonical labels shown in Figure 6. Canonically labeled atoms of buckminsterfullerene allow a straightforward assignment of unique labels to all the 90 CC bonds as well as to all the 12 pentagonal and 20 hexagonal faces as will be described below. Again we admit that the final labels for CC bonds, pentagonal and hexagonal faces do not appear »orderly« but one has to recognize that there is no »simple« labeling in so large systems that will appear simple. If one succeeded with some rules to set labels for one region that appear orderly that same approach will produce scattered labels for other local regions. So it is out view that what is more important than aiming at some locally pleasing and »simple« arrangement of labels is to have simple rules for assignment of labels, rather than simple distribution of labels.

### Labeling of CC bonds

The canonical labeling of CC bonds in buckminsterfullerene is based on the fact that unique labels have been assigned to the atoms. We start with carbon atom labeled with 1 and assign to its three CC bonds labels 1, 2, 3 such that bond with the highest other label becomes CC(1), next highest is CC(2), and CC(3) is associated with the smallest of the triplet numbers. Hence, CC(1) is bond C(1)-C(60), CC(2) is bond C(1)-C(59) and CC(3) is bond C(1)-C(58). We continue now with the three CC bonds incident to carbon atom 2, which give: CC(4) being bond C(2)-C(60); CC(5) being bond C(2)-C(57), and finally CC(6) being bond C(2)-C(56), and so on. In Table III we list the proposed canonical labels for all the 90 CC bonds of buckminsterfullerene. The way how we constructed the list is that the first entry of bond  $i-j$  is greater than the second entry. As each carbon atom in buckminsterfullerene is incident with three CC bonds

TABLE III. Canonical labels for all the 90 CC bonds in buckminsterfullerene

Canonical label of a bond	Canonical labels of atoms making up the bond	Canonical label of a bond	Canonical labels of atoms making up the bond	Canonical label of a bond	Canonical labels of atoms making up the bond
CC(1)	1-60	CC(31)	11-53	CC(61)	21-44
CC(2)	1-59	CC(32)	11-45	CC(72)	21-29
CC(3)	1-58	CC(33)	11-43	CC(63)	21-28
CC(4)	2-60	CC(34)	12-52	CC(64)	22-38
CC(5)	2-57	CC(35)	12-42	CC(65)	22-27
CC(6)	2-56	CC(36)	12-41	CC(66)	22-26
CC(7)	3-59	CC(37)	13-51	CC(67)	23-32
CC(8)	3-57	CC(38)	13-40	CC(68)	23-27
CC(9)	3-55	CC(39)	13-39	CC(69)	23-25
CC(10)	4-60	CC(40)	14-42	CC(70)	24-29
CC(11)	4-54	CC(41)	14-40	CC(71)	24-26
CC(12)	4-53	CC(42)	14-38	CC(72)	24-25
CC(13)	5-58	CC(43)	15-39	CC(73)	25-31
CC(14)	5-54	CC(44)	15-38	CC(74)	26-28
CC(15)	5-52	CC(45)	15-37	CC(75)	27-37
CC(16)	6-59	CC(46)	16-49	CC(76)	28-42
CC(17)	6-52	CC(47)	16-36	CC(77)	29-30
CC(18)	6-50	CC(48)	16-35	CC(78)	30-45
CC(19)	7-55	CC(49)	17-48	CC(79)	33-46
CC(20)	7-51	CC(50)	17-34	CC(80)	34-35
CC(21)	7-49	CC(51)	17-33	CC(81)	35-47
CC(22)	8-57	CC(52)	18-36	CC(82)	36-37
CC(23)	8-48	CC(53)	18-34	CC(83)	39-49
CC(24)	8-47	CC(54)	18-32	CC(84)	40-41
CC(25)	9-56	CC(55)	19-33	CC(85)	41-50
CC(26)	9-48	CC(56)	19-32	CC(86)	43-46
CC(27)	9-46	CC(57)	19-31	CC(87)	44-52
CC(28)	10-54	CC(58)	20-43	CC(88)	47-55
CC(29)	10-45	CC(59)	20-31	CC(89)	50-58
CC(30)	10-44	CC(60)	20-30	CC(90)	53-56

TABLE IV. Lexicographical ordering of pentagonal rings in buckminsterfullerene

Canonical label of pentagonal ring	Canonical labels of atoms making up the ring
1	1, 6, 50, 58, 59
2	2, 4, 53, 56, 60
3	3, 8, 47, 55, 57
4	5, 10, 44, 52, 54
5	7, 13, 39, 49, 51
6	9, 17, 33, 46, 48
7	11, 20, 30, 43, 45
8	12, 14, 40, 41, 42
9	15, 22, 27, 37, 38
10	16, 18, 34, 35, 36
11	19, 23, 25, 31, 32
12	21, 24, 26, 28, 29

each atomic label appears three times in Table III. Observe that approximately for the two thirds of the list, till we come to carbon atom 25, the occurrence of labels, which are ordered lexicographically, successively repeat the first atom of a bond three times. This is because for atoms with label 24 or less all three canonical neighbors have larger labels. However, with carbon atom 25 and after, we see that some of its neighbors have smaller labels, and hence the corresponding CC bonds have already been listed. Thus, for example, because carbon atom 25 has as its neighbors 23, 24 and 31 we have to add to the list only the bond C(25)-C(31), the other two bonds have already been listed. By looking at the first entries in the last column one can detect that numbers 31, 32, 37, 38, 42, *etc.* do not occur. This is because for these atoms all three nearest neighbors have smaller canonical labels, and hence bonds involving these atoms have already been listed.

TABLE V. Lexicographical ordering of hexagonal rings in buckminsterfullerene

Canonical label of hexagonal ring	Canonical labels of atoms making up the ring	Canonical label of hexagonal ring	Canonical labels of atoms making up the ring
1	1, 2, 3, 57, 59, 60	11	10, 21, 29, 30, 44, 45
2	1, 4, 5, 54, 58, 60	12	12, 21, 28, 42, 44, 52
3	2, 8, 9, 48, 56, 57	13	13, 14, 15, 38, 39, 40
4	3, 6, 7, 51, 55, 59	14	14, 22, 26, 28, 38, 42
5	4, 10, 11, 45, 53, 54	15	15, 16, 36, 37, 39, 49
6	5, 12, 41, 50, 52, 58	16	17, 18, 19, 32, 33, 34
7	6, 13, 40, 41, 50, 51	17	18, 23, 27, 32, 36, 37
8	7, 16, 35, 47, 49, 55	18	19, 20, 31, 33, 43, 46
9	8, 17, 34, 35, 47, 48	19	20, 24, 25, 29, 30, 31
10	9, 11, 43, 46, 53, 56	20	22, 23, 24, 25, 26, 27

We can use canonical labels of atoms to arrive at canonical labels for rings. To wit, we simply assign to each ring the ordered set of labels of atoms making up the ring. Thus, each pentagonal ring will be described by means of five numbers while each hexagonal ring will be described by means of six numbers.

#### *Labeling of pentagonal rings*

In Table IV we have listed for the 12 pentagonal rings labels of carbon atoms forming the rings. Since the rings are disjoint none of the numbers appearing in the lists occurs twice. The lexicographical ordering assigns to each pentagonal ring one of labels between 1 and 12.

#### *Labeling of hexagonal rings*

In Table V we have listed for the 20 hexagonal rings labels of carbon atoms making up the rings. As each carbon atom is common to two hexagonal rings its label occurs twice in the list. The lexicographical ordering assigns to each hexagonal ring one of labels between 1 and 20.

## DISCUSSION

Those who are familiar with various problems accompanying graph constructions, graph enumerations, graph isomorphism and graph automorphism problems hardly need arguments and justification for the effort here undertaken to arrive at canonical labels for fullerenes. To some, however, the development of a computer program (to be briefly outlined in the Appendix) for the assignment of canonical labels for fullerenes may appear as being of marginal importance. We would like to argue just the opposite: Having available the computer program is an important advancement for documentation of information on fullerenes, which we will justify in this section.

Hitherto the overall situation, when discussing or reporting information about fullerenes, is characterized by

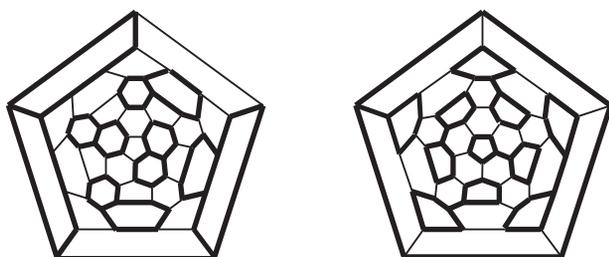


Figure 7. Symmetric decompositions of buckminsterfullerene obtained under action of  $C_{3i}$  (according to Babić and Trinajstić).

the use of graphical representations (such as simple 2-D projections, Schlegel diagrams or 3-D stereo images), which are accompanied by arbitrary labels. Very helpful in that respect is availability of the atlas of fullerenes by Fowler and Manolopoulos.<sup>12</sup> Nevertheless, papers dealing with fullerenes typically will display fullerene structures and often specific results pictorially. See for example the work of Laidboeur *et al.*<sup>13</sup> on the determination of topological equivalent classes of atoms and bonds in  $C_{20}$ – $C_{60}$ , in which all twenty fullerenes have been depicted and all atoms labeled, or see the works of Babić and Trinajstić on assembling fullerenes from identical fragments<sup>14</sup> and particularly on symmetric decomposition of buckminsterfullerene into identical monomeric units<sup>15</sup> where results are displayed on 15 molecular diagrams of buckminsterfullerene. The question that we would like to raise is: Could these authors (and hundred others that have not been mentioned) present their results without using graphical display of fullerene diagrams? The same question, of course, applies also to the present authors regarding their report on the complexity of fullerenes<sup>16</sup> and very recent article which displays 158 symmetry non-equivalent Kekulé valence structures of buckminsterfullerene.<sup>17</sup> The answer is: No. All the mentioned contributions and hundred others not mentioned could not be reported in a simple manner without presenting the results graphically. We have to emphasize »simple manner« because it is always possible to adopt an *ad hoc* numbering of vertices and continue using such in listing connectivity or bonds of interest. But part of the problem is how to communicate the particular numbering adopted? In some cases one can use spiral string of numbers, where it applies, and give some other specifications, which for a most general structures are likely to produce cumbersome and arbitrary nomenclature, that others may not be willing to adopt. If one accepts canonical labels of atoms that generates the smallest canonical adjacency matrix, and follows with lexicographic ordering of pentagonal and hexagonal rings as proposed in this work one could »transmit« results of works on fullerenes without graphical supplementary information at all. Now this is not to say that graphical displays have no merits, on the contrary, they are very valuable and should accompany such work. However, if we want the same in-

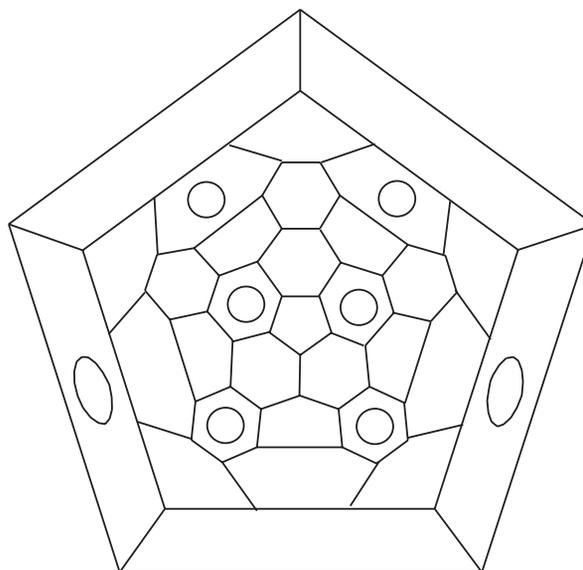


Figure 8. One of the five symmetry related Clar structures of buckminsterfullerene.

formation to store in computer for future search and comparison, then clearly our approach shows its merits.

To illustrate our point we reproduce in Figure 7 two of the 15 symmetric decompositions of buckminsterfullerene obtained under action of  $C_{3i}$ .<sup>14</sup> The following list of hexagonal rings, pentagonal rings, and CC bonds carries the same information:

$$H_1 + H_4; H_5 + H_{10}; H_6 + H_{12}; H_9 + H_{16}; H_{13} + H_{15}; \\ H_{19} + H_{20}$$

$$P_1 + P_3 + CC_7; P_2 + P_7 + CC_{31}; P_4 + P_8 + CC_{34}; \\ P_5 + P_9 + CC_{43}; P_6 + P_{10} + CC_{50}; P_{11} + P_{12} + CC_{72}$$

The subscripts in  $CC_i$ ,  $P_j$ , and  $H_k$  indicate the canonical labels of CC bonds, pentagonal rings, and hexagonal rings as listed in Tables III–V, respectively, and the sign + indicates union of sets.

Similarly, we can indicate which eight (disjoint) benzene rings form the five Clar structures of buckminsterfullerene<sup>18,19</sup> without using molecular diagram of Figure 8, which shows one of the five symmetry related Clar structure with inscribed p-electron sextets. Here is such list:

Clar structure 1:

$$H_1 + H_5 + H_7 + H_9 + H_{12} + H_{15} + H_{18} + H_{20}$$

Clar structure 2:

$$H_1 + H_6 + H_8 + H_{10} + H_{11} + H_{13} + H_{16} + H_{20}$$

Clar structure 3:

$$H_2 + H_3 + H_7 + H_8 + H_{11} + H_{14} + H_{17} + H_{18}$$

Clar structure 4:

$$H_2 + H_4 + H_9 + H_{10} + H_{12} + H_{13} + H_{17} + H_{19}$$

Clar structure 5:

$$H_3 + H_4 + H_5 + H_6 + H_{14} + H_{15} + H_{16} + H_{19}$$

Observe that canonical labels induce ordering, thus allowing unique labels for the five Clar structures (based on lexicographical ordering of the constituting sets of hexagons).

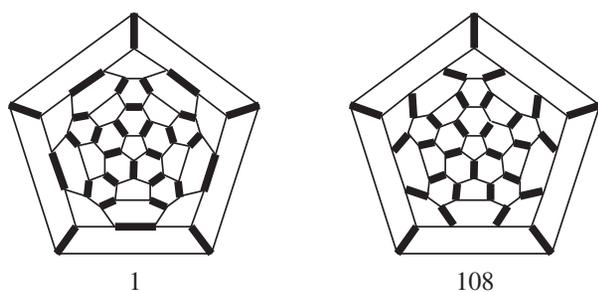


Figure 9. Two symmetry non-equivalent Kekulé valence structures of buckminsterfullerene. The labels 1 and 108 correspond to labels of Kekulé valence structures used in Atlas of Kekulé valence structures of buckminsterfullerene (Ref. 17).

TABLE VI. Canonical labels of the 30 CC double bonds of the Kekulé valence structures of Figure 9. The labels 1 and 108 correspond to labels of Kekulé valence structures used in Atlas of Kekulé valence structures of buckminsterfullerene (Ref. 17)

Label of a structure	Canonical labels of CC double bonds
1	1, 5, 7, 11, 13, 17, 19, 23, 25, 29, 31, 34, 38, 42, 43, 46, 50, 54, 55, 59, 61, 66, 68, 72, 76, 77, 81, 82, 85, 86
108	1, 6, 7, 12, 13, 17, 21, 22, 27, 28, 33, 35, 39, 41, 45, 47, 49, 54, 55, 59, 62, 64, 68, 72, 74, 78, 80, 85, 87, 88

As the final illustration of use of the canonical labels introduced here for  $C_{60}$  we will outline their use for digital transmission of pictographic information (in the form of listing relevant components). In Table VI we have listed CC double bonds for two Kekulé valence structures shown in Figure 9. The first structure (at left in Figure 9) is the dominant Kekulé valence structure with all CC double bonds being exocyclic to the pentagonal rings. This structure has twenty (the maximal number possible) of the smallest and most important conjugated circuits  $R_1$ . This structure is also known as Fries valence structure,<sup>20</sup> being the structure with the maximal number of benzene rings with three double bonds. Conjugated circuits<sup>21</sup> are defined as those circuits within individual Kekulé valence structure that have regular alternation of CC double and CC single bonds. They play the key role in evaluation of the relative contributions of the individual Kekulé valence structures to the molecular resonance energy (RE). The second structure is one of the least important Kekulé valence structure of buckminsterfullerene, which as one can see from Figure 9, has no smaller conjugated circuits  $R_1$ ,  $R_2$ , or  $R_3$ , which make positive contribution to molecular RE. Such structures are also known as anti-Fries valence structures.

Observe the possibility of lexicographical ordering of Kekulé valence structures, based on the list of canonical labels for double CC bonds. It has yet to be seen if such ordering will or will not correspond to the current sorting of Kekulé valence structures based on their innate degree of freedom.<sup>22</sup> However, even if it does not, it will for needs of chemical documentation supply unique labels for 158 symmetry non-equivalent Kekulé valence structures of buckminsterfullerene.

*Acknowledgement.* – This work was supported in part by the Ministry of Science, Education and Sports of the Republic of Croatia and Croatian-Slovenian project »Application of Methods of Discrete Mathematics in Chemistry and Biology«.

## REFERENCES

- (a) IUPAC: *Definitive Rules for Nomenclature of Inorganic Chemistry 1957*, Butterworth, London 1959; (b) IUPAC: *Nomenclature of Inorganic Boron Compounds*, *Pure Appl. Chem.* **30** (1972) 681–710.
- (a) ACS Nomenclature, Spelling, and Pronunciation Committee, The Designation of »Extra« Hydrogen in Naming Cyclic Compounds, *Chem. Eng. News* **27** (1949) 1303–1304; (b) ACS Nomenclature, Spelling, and Pronunciation Committee, Halogenated Derivatives of Hydrocarbons, *Chem. Eng. News* **30** (1952) 4513–4514; (c) ACS Division of Polymer Chemistry Nomenclature Committee, A Structure-Based Nomenclature for Linear Polymers, *Macromolecules* **1** (1968) 193–198.
- (a) A. T. Balaban, *Tetrahedron* **25** (1969) 2949–2956; (b) C. Jochum and J. Gasteiger, *J. Chem. Inf. Comput. Sci.* **17** (1977) 113–117; (c) N. Lozac'h, A. L. Goodson, and W. H. Powell, *Angew. Chem., Int. Ed. Engl.* **18** (1979) 887–966; (d) A. L. Goodson, *J. Chem. Inf. Comput. Sci.* **20** (1980) 167–172; (e) A. L. Goodson, *J. Chem. Inf. Comput. Sci.* **20** (1980) 172–176; (f) A. L. Goodson, *Croat. Chem. Acta* **56** (1983) 315–324; (g) R. C. Read, *J. Chem. Inf. Comput. Sci.* **23** (1983) 135–149; (g) W. C. Herndon and J. E. Leonard, *Inorg. Chem.* **22** (1983) 554–557; (h) J. B. Hendrickson and A. G. Toczko, *J. Chem. Inf. Comput. Sci.* **23** (1983) 171–177; (i) R. C. Read, *J. Chem. Inf. Comput. Sci.* **25** (1985) 135–149; (j) N. Lozac'h, *J. Chem. Inf. Comput. Sci.* **25** (1985) 180–185; (k) M. Randić, *J. Chem. Inf. Comput. Sci.* **26** (1986) 136–148; (l) M. Randić, *Croat. Chem. Acta* **59** (1986) 327–343; (m) M. Randić, S. Nikolić, and N. Trinajstić, *J. Mol. Struct. (THEOCHEM)* **165** (1988) 213–228; (n) S. Nikolić and N. Trinajstić, *Croat. Chem. Acta* **63** (1990) 155–170; (o) E. C. Kirby, *J. Math. Chem.* **11** (1991) 187–197; (p) M. Randić, S. Nikolić, and N. Trinajstić, *J. Chem. Inf. Comput. Sci.* **35** (1995) 357–365.
- A. T. Balaban and F. Harary, *Tetrahedron* **24** (1968) 2505–2516.
- A. T. Balaban and P. v. Ragué Schleyer, *Tetrahedron* **34** (1978) 3599–3609.
- D. Hellwinkel, *Systematic Nomenclature of Organic Chemistry*, Springer-Verlag, Berlin, 2001.
- The Merck Index, An Encyclopedia of Chemicals, Drugs, and Biologicals*, 13th edition, Merck & Co., Inc., Whitehouse Station, NJ, 2001.

8. M. Randić, (a) *J. Chem. Phys.* **60** (1974) 3920–3928; (b) *J. Chem. Phys.* **62** (1975) 309–310; (c) *J. Chem. Inf. Comput. Sci.* **17** (1977) 171–180.
9. M. Randić, G. M. Brissey, and C. L. Wilkins, *J. Chem. Inf. Comput. Sci.* **21** (1981) 52–59.
10. (a) M. Randić, *Chem. Phys. Lett.* **42** (1976) 283–287; (b) M. Randić, *Croat. Chem. Acta* **49** (1977) 643–655; (c) M. Randić, *Int. J. Quantum Chem.* **15** (1979) 663–682; (d) M. Randić and V. Katović, *Int. J. Quantum Chem.* **21** (1982) 647–663; (e) M. Randić, *J. Comput. Chem.* **4** (1983) 73–83; (f) M. Randić, V. Katović, and N. Trinajstić, *Stud. Phys. Theor. Chem.* **23** (1983) 399–408; (g) M. Randić, D. O. Oakland, and D. J. Klein, *J. Comput. Chem.* **7** (1986) 35–54.
11. M. Randić, *Acta Crystallogr., Sect. A* **34** (1978) 275–282.
12. P. W. Fowler and D. E. Manolopoulos, *An Atlas of Fullerenes*, Oxford University Press, London, 1995.
13. T. Laidboeur, D. Cabrol-Bass, and O. Ivanciuc, *J. Chem. Inf. Comput. Sci.* **36** (1996) 811–821.
14. (a) D. Babić and N. Trinajstić, *Fullerene Sci. Technol.* **2** (1994) 343–356; (b) D. Babić and N. Trinajstić, *MATCH-Commun. Math. Comput. Chem.* **33** (1996) 17–24.
15. J. Cioslowski, *Electronic Structure Calculations on Fullerenes and Their Derivatives*, Oxford University Press, London, 1995.
16. M. Randić, X. Guo, D. Plavšić, and A. T. Balaban, *On the Complexity of Fullerenes and Nanotubes*, in: D. Bonchev and D. H. Rouvray (Eds.), *Complexity in Chemistry, Biology and Ecology*, Springer, New York, (in press).
17. D. Vukičević, H. W. Kroto, and M. Randić, *Croat. Chem. Acta* **78** (2005) 223–234.
18. S. El-Basil, *J. Mol. Struct. (THEOCHEM)* **531** (2000) 9–21.
19. M. Randić, *Chem. Rev.* **103** (2003) 3449–3605.
20. (a) K. Fries, *Justus Liebigs Ann. Chem.* **454** (1927) 121; (b) K. Fries, R. Walter, and K. Schilling, *Justus Liebigs Ann. Chem.* **516** (1935) 245.
21. M. Randić, (a) *Chem. Phys. Lett.* **38** (1976) 68–70; (b) *J. Am. Chem. Soc.* **99** (1977) 444–450; (c) *Tetrahedron*, **33** (1977) 1905–1920.
22. M. Randić and T. Pisanski, *Rep. Mol. Theory* **1** (1990) 107–114.
23. D. L. Kreher and D. R. Stinson, *Combinatorial Algorithms Generation, Enumeration and Search*, CRC Press, Boca Raton, 1999.

---

## SAŽETAK

### O kanonskom označavanju ugljikovih atoma u fullerenima: C<sub>60</sub> buckminsterfulleren

Dejan Plavšić, Damir Vukičević i Milan Randić

Označavanje atoma u velikim i složenim molekulama vrlo je često proizvoljno ili je zasnovano na *ad hoc* shemama. Preporuča se novo i sustavno označavanje atoma u molekulama koje se temelji na najmanjem mogućem binarnom molekularnom kodu. Čitanje matrice susjedstva grafa molekule po retcima definira binarni molekularni kod: najmanji među njima se zove kanonski kod. Na primjeru molekule buckminsterfullerena pokazano je kako se može odrediti ovaj kod. Istaknute su prednosti novoga načina označavanja atoma u velikim molekulama a posebice za one prikazane regularnim grafovima.

## Appendix

Let vertices of a graph be initially arbitrary labeled by  $1, 2, \dots, v(G)$ , where  $v(G)$  is the number of vertices of the graph. We propose an algorithm which successively chooses values of  $P'(1), P'(2), P'(3), \dots$ , in the restrictive way ( $P'(i) = j$  means that we label with  $i$  the vertex which was initially labeled with  $j$ ). Suppose that values of  $P'(1), P'(2), \dots, P'(k)$ , are already chosen. Our aim is to (restrictively) determine the possible candidates for the value of  $P'(k+1)$  and proceed in the similar manner. The first permutation produced by this algorithm is stored in the array  $P$ . After that each of the permutations produced by the algorithm is compared to the one stored in  $P$  and if it is found that the considered new permutation is »better«, then it is stored in  $P$ . At the end  $P$  contains the canonical permutation.

More formally we use the following recursive algorithm (presented in the pseudo-code):

*rec(k)*

1) If  $x = n$  do

1.1) If this is the first produced permutation then  
 $P = P'$

1.2) Else if  $P'$  produces »better« labeling then put  
 $P = P'$

2) Else

2.1) Determine the set *Cand* of candidates for  
 $P'(k+1)$

2.2) For each  $c \in \text{Cand}$  then put  $P'(k+1) = c$   
and call *rec(k+1)*

The most important line of this algorithm is the line 2.1 in which one needs to determine the set *Cand*. It is determined in the four steps:

Step 1: To each vertex  $v \in V_0$ , where  $V_0$  denotes the set of vertices that are not already stored in the permutation, we assign the vector  $a_v = (i_1, \dots, i_k)$  such that  $i_q = 1$  if  $v$  and  $P(q)$  are adjacent and 0 otherwise.

Step 2: Vectors  $a_v, v \in V_0$ , are sorted according to their lexicographical order and ranked. The vertices corresponding to the smallest vectors have rank 1, those immediately after them have rank 2 and so on. The maximal rank assigned in this way is denoted by *mr*.

Step 3: To each vertex  $v \in V_0$  of rank 1 we assign a vector  $b_v$  with *mr* entries in such a way that the  $i$ -th component of the vector  $b_v$  is the number of neighbors of  $v$  that have rank  $i$ .

Step 4: From the set of vectors with rank 1, those that are minimal (according to the lexicographical order) are extracted and the vertices corresponding to them form the set *Cand*.

A more complicated and a bit faster algorithm than our algorithm is proposed in Ref. 23. The algorithm is based on the successive refinement of the partition of vertices and in each step it takes into account the refinement obtained by the previous step of the algorithm.