Enumeration and Classification of Benzenoid Hydrocarbons. 2. Symmetry and Regular Hexagonal Benzenoids

J. BRUNVOLL, B. N. CYVIN, and S. J. CYVIN*

Division of Physical Chemistry, The University of Trondheim, N-7034 Trondheim-NTH, Norway

In the computerized enumerations and classifications of benzenoids the following cases are distinguished: (i) specific generation of a class, (ii) recognition of the members of a class, and (iii) determination of the number of benzenoids in a class by exclusion. Enumerated benzenoids for h (the number of hexagons) up to 10 are classified into the eight possible symmetry groups. The procedure is fully computerized. Regular hexagonal benzenoids are treated in detail by means of a special computer program. Some of their topological properties are mentioned. They are enumerated for h up to 55. Both normal and essentially disconnected benzenoids are found among them, but no obvious non-Kekuléans. Concealed non-Kekuléans occur for $h \geq 43$. The 5 existing systems of this type among the 540 regular hexagonal benzenoids with $h = 43$ are shown. Numbers of Kekulé structures are given for all regular hexagonal benzenoids with h up to 37 ($h > 1$). Also the forms of these 191 systems are shown in terms of computer-designed figures.

The computer-aided enumeration and classification of benzenoid hydrocarbons started effectively with the Düsseldorf–Zagreb school (Knop and Trinajstić with collaborators).[1-4] These researchers had been inspired by a work of Balasubramanian et al.[5] Recently the interest in this kind of work has increased substantially, represented by He and He,[6,7] Brunvoll et al.,[8] Gutman,[9] Cioslowski,[10] and Stojmenović et al.[11] In all these works the leading parameter for the benzenoids is h, the number of hexagons. Another principle was used in the enumerations by Dias,[12-18] who put emphasis on the chemical isomerism (C_nH_m). In some works the enumeration in relation to the number of Kekulé structures (K) of the benzenoids was used as the leading motif.[19-22]

In a systematic description of the computer-aided classification of benzenoids we will distinguish the following cases: (i) specific generation, (ii) recognition, and (iii) exclusion. In a standard procedure a broad class of benzenoids is generated, e.g., all benzenoids with a given h. Then a set of benzenoids may be recognized as belonging to a subclass by means of special programs or subroutines in the main program. This is the principle of recognition (ii). Very little work has been done on a conscious employment of the specific generation (i), by which smaller classes within a given h value are generated separately. The principle of exclusion (iii) is applicable when a class of benzenoids in one way or another has been divided into subclasses but so that one subclass remains. Then the number of benzenoids is obviously obtained by subtraction.

In our first computer program for enumeration of benzenoids[8] there are in fact options for specific generation of certain classes. (For precise definitions, see, e.g., ref 8, 19–21, 23, 24 and subsequent sections of the present paper.) (1) The unbranched catacondensed (ub) benzenoids are available separately. (2) The catacondensed (cat) benzenoids in total are also available. (3) Then the branched catacondensed (b) benzenoids are obtained by exclusion (b = cat – ub).[25] (4) The normal (n) benzenoids can be generated specifically. (5) The normal pericondensed (np) systems are either generated specifically or obtained by exclusion (np = n – cat). (6) Finally the total number (tot) of benzenoids for increasing h values are generated. (7) Pericondensed (p) systems are obtained by exclusion (p = tot – cat).

The quantity Δ is an important parameter in benzenoids. It signifies the absolute magnitude of the difference between the number of black and white (or starred and unstarred) vertices.[26] In the mentioned computer program[8] the Δ values are obtained by recognition. In this way the obvious non-Kekuléans[24] (oo) are enumerated. The members of this benzenoid class are defined by $\Delta > 0$. By exclusion (tot – oo), or more directly by recognition, the class of $\Delta = 0$ is determined. This class consists of the normal (n), essentially disconnected (e), and concealed non-Kekuléan[24] (co) benzenoids. The n + e members are Kekuléan; i.e., they possess Kekulé structures. Those termed essentially disconnected (e) have fixed bonds (single or double) in all Kekulé structures; the normal (n) have not. Non-Kekuléan benzenoids (which do not possess Kekulé structures) consist of co ($\Delta = 0$) + oo ($\Delta > 0$). It is known that no concealed non-Kekuléan benzenoids exist for $h < 11$.[24,26]

Symmetry is a fundamental property of molecules in general and benzenoids in particular. It has been pointed out[27] that a benzenoid belongs to one of the eight symmetry groups D_{6h}, C_{6h}, D_{3h}, C_{3h}, D_{2h}, C_{2h}, C_{2v} or C_s. Balaban and Harary[28] were probably the first to make use of symmetry in a classification of benzenoids but considered only the unbranched catacondensed systems in this respect. Their studies were pursued in our previous work.[8] In the present work our original program[8] was extended to recognize the symmetry groups for all symmetrical benzenoids, while the unsymmetrical (C_s) members were determined by exclusion. In this way a complete classification of benzenoids (including the pericondensed and branched catacondensed systems) was undertaken for the first time, and results are reported for h up to 10.

A new feature of the present work is also the design of computer programs for specific generations of benzenoids with a given symmetry. Such programs were produced for all symmetries but C_s. In the present paper we give a description for the hexagonal symmetries (D_{6h} and C_{6h}) and report the results of enumerations for h up to 43.

In papers of the present kind it is important to define precisely the considered classes of molecules. We adhere to the definition adopted by the Düsseldorf–Zagreb school.[1-4] This definition is also adopted in most of the recent work on enumeration of benzenoids.[6-11]

RESULTS AND DISCUSSION

Computer Programming for Recognition of Symmetry. The previously described computer program[8] was supplemented by an additional program in order to recognize the symmetry of a benzenoid. Of particular interest in the previous program description is the generation of 12 positions of a benzenoid to eliminate duplicated systems. Figure 1 shows a benzenoid
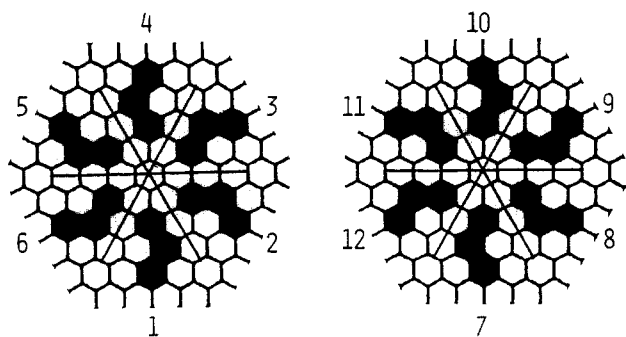
Figure 1. Numbering of the 12 rotated positions of a benzenoid (phenanthrene).

Table I. Identical Positions with Rotation No. 1 (Reference Is Made to the Numbering of Figure 1) of Benzenoids of Different Symmetries

position	D_{6h}	C_{6h}	D_{3h}	C_{3h}	C_{2h}
1	×	×	×	×	×
2	×	×			
3	×	×	×	×	
4	×	×			×
5	×	×	×	×	
6	×	×			
7	×		×		
8	×				
9	×		×		
10	×				
11	×		×		
12	×				

Table II. Number of Catacondensed Benzenoids of Different Symmetries[a]

h		D_{6h}	C_{6h}	D_{3h}	C_{3h}	D_{2h}	C_{2h}	C_{2v}	C_s
1	ub	1	0	0	0	0	0	0	0
2	ub	0	0	0	0	1	0	0	0
3	ub	0	0	0	0	1	0	1	0
4	ub	0	0	0	0	1	1	1	1
	b	0	0	1	0	0	0	0	0
5	ub	0	0	0	0	1	1	4	4
	b	0	0	0	0	0	0	1	1
6	ub	0	0	0	0	1	4	3	16
	b	0	0	0	0	1	0	4	7
7	ub	0	0	0	0	1	4	12	50
	b	0	0	1	1	1	0	4	44
8	ub	0	0	0	0	1	13	10	158
	b	0	0	0	0	1	4	18	206
9	ub	0	0	0	0	1	13	34	472
	b	0	0	0	0	1	4	27	937
10	ub	0	0	0	0	1	39	28	1406
	b	0	0	2	4	3	25	67	3997
11	ub	0	0	0	0	1	39	97	4111
	b	0	0	0	0	4	26	118	16719
12	ub	0	0	0	0	1	116	81	11998
	b	0	0	0	0	4	132	269	unk
13	ub	0	0	0	0	1	115	271	unk
	b	0	0	2	15	4	140	unk	unk
14	ub	0	0	0	0	1	339	226	unk
	b	0	0	0	0	9	unk	unk	unk
15	ub	0	0	0	0	1	336	unk	unk
	b	0	0	0	0	11	unk	unk	unk

[a] Abbreviations: b, branched; ub, unbranched, unk, unknown.

(phenanthrene) in these 12 positions. The present program detects how many of these positions are identical for a given benzenoid. In order to be able to compare the coordinates of two benzenoids, they are all parallel-shifted to come as near as possible to the origin of the coordinate system in the sense of the previous description.[8] More concretely the program finds out how many and which positions are identical with no. 1 (see Figure 1). In the depicted example (phenanthrene with symmetry C_{2v}) two positions are identical, viz., nos. 1 and 10.

Table III. Number of Pericondensed Kekuléan Benzenoids of Different Symmetries[a]

h		D_{6h}	C_{6h}	D_{3h}	C_{3h}	D_{2h}	C_{2h}	C_{2v}	C_s
4	np	0	0	0	0	1	0	0	0
5	np	0	0	0	0	0	0	1	1
	e	0	0	0	0	1	0	0	0
6	np	0	0	0	0	1	2	3	6
	e	0	0	0	0	0	1	0	2
7	np	1	0	0	0	1	0	6	41
	e	0	0	0	0	0	3	6	14
8	np	0	0	0	0	2	11	19	200
	e	0	0	0	0	2	7	2	110
9	np	0	0	0	0	3	3	39	997
	e	0	0	0	0	2	16	29	645
10	np	0	0	1	0	6	52	90	4654
	e	0	0	0	0	1	53	31	3647

[a] Abbreviations: e, essentially disconnected; np, normal pericondensed.

Table IV. Total Number of Benzenoids of Different Symmetries, Classified According to Δ Values

h	Δ	D_{6h}	C_{6h}	D_{3h}	C_{3h}	D_{2h}	C_{2h}	C_{2v}	C_s
1	0	1	0	0	0	0	0	0	0
2	0	0	0	0	0	1	0	0	0
3	0	0	0	0	0	1	0	1	0
	1	0	0	1	0	0	0	0	0
4	0	0	0	1	0	2	1	1	1
	1	0	0	0	0	0	0	0	1
5	0	0	0	0	0	2	1	6	6
	1	0	0	0	0	0	0	3	4
6	0	0	0	0	0	3	7	10	31
	1	0	0	0	1	0	0	1	26
	2	0	0	1	0	0	0	1	0
7	0	1	0	1	1	3	7	28	149
	1	0	0	0	0	0	0	10	124
	2	0	0	0	0	0	0	1	6
8	0	0	0	0	0	6	35	49	674
	1	0	0	0	0	0	0	5	614
	2	0	0	0	0	0	0	7	45
9	0	0	0	0	0	7	36	129	3051
	1	0	0	0	4	0	0	39	2914
	2	0	0	1	1	0	0	9	311
	3	0	0	0	0	0	0	1	2
10	0	0	0	3	4	11	169	216	13704
	1	0	0	0	0	0	0	20	14004
	2	0	0	0	0	0	0	38	1878
	3	0	0	1	1	0	0	0	37

The D_{6h} group is recognized by all 12 positions being identical. Table I shows the situation together with the corresponding analysis for C_{6h}, D_{3h}, C_{3h} and C_{2h}. The D_{2h} group is recognized by four identical positions, whichever they are. The C_{2h} group gives nos. 1 and 4 as identical positions. Also, C_{2v} gives two identical positions only but situated in different places from 1 and 4. Finally, the C_s group is recognized by exclusion.

The analysis may be performed separately for all the options of the program described previously[8,25] pertaining to (i) unbranched catacondensed benzenoids and (ii) all catacondensed, (iii) normal pericondensed, and (iv) all benzenoids (grand total). Results for (v) branched catacondensed benzenoids are then available by subtraction [(ii) − (i)] and for (vi) all normal benzenoids by addition [(ii) + (iii)]. For $h \le 10$ the (vii) Kekuléan and (viii) non-Kekuléan benzenoids are identified by $\Delta = 0$ and $\Delta > 0$, respectively. Then, finally, the number of (ix) essentially disconnected benzenoids is obtained by difference [(vii) − (vi)].

Symmetry Distribution of Benzenoids. The results of the above analysis are compressed into three tables. Table II shows the results for catacondensed benzenoids. For the unbranched systems the analysis confirmed the results of the previous work,[8] which were obtained in an elaborate way without computer aid. All the missing numbers in Table II of ref 8 are supplied as far as the unbranched systems are concerned,
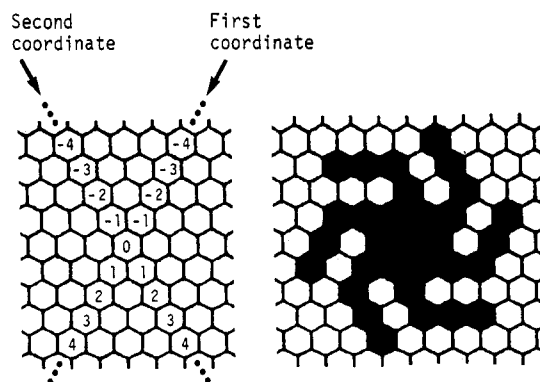
CLASSIFICATION OF BENZENOID HYDROCARBONS

J. Chem. Inf. Comput. Sci., Vol. 27, No. 4, 1987 **173**

Figure 2. Coordinate system and a hexagonal benzenoid with $\eta =$ 5. Its coordinates are (0,1) (1,1) (1,2) (0,3) and (−1,4).

Table V. Number of Hexagonal Benzenoids, Including Their Symmetries

η	h	D_{6h}	C_{6h}	total
1	7	1	0	1
2	13	2	0	2
3	19	2	2	4
4	25	3	8	11
5	31	5	32	37
6	37	8	128	136
7	43	13	527	540
8	49	20	2209	2229
9	55	35	9470	9505

Table VI. Number of Hexagonal Benzenoids with Different Cores: Coronene (c), Circumcoronene (cc), ccc, and cccc

η	h	c	cc	ccc	cccc
1	7	1	0	0	
2	13	2	0	0	
3	19	3	1	0	
4	25	9	2	0	
5	31	31	6	0	
6	37	115	20	1	
7	43	464	73	3	
8	49	1927	293	9	
9	55	8257	1215	33	
10	61	a	5206	126	1
11	67	a	a	519	3
12	73	a	a	2216	12

^{a} Unknown.

Table VII. Number of Hexagonal Benzenoids^a

η	h	n	e	o
1	7	1	0	0
2	13	2	0	0
3	19	4	0	0
4	25	9	2	0
5	31	29	8	0
6	37	91	45	0
7	43	321	214	5

^{a} n, normal; e, essentially disconnected; o, non-Kekuléan (the "neo" classification).

and the analysis is extended to higher h values. Table III shows the results for normal pericondensed and essentially disconnected benzenoids. Finally, we give (Table IV) the symmetry distribution for all benzenoids classified according to their Δ values.

Benzenoids with Hexagonal Symmetry. Some Topological Properties. The (regular) hexagonal (D_{6h} and C_{6h}) benzenoids have been treated particularly by Hosoya[29] and more extensively by Cyvin et al.[30] It has been pointed out that these systems necessarily possess a central hexagon. The total number of hexagons is

$$h = 6\eta + 1$$

where η is a positive integer or zero. For $\eta = 0$ one obtains benzene itself, which is disregarded as a trivial case in the remainder of this work. For all $\eta \geq 1$ the central hexagon is completely surrounded by hexagons in a coronene configuration, which may be referred to as the core. Coronene itself has $\eta = 1$. Circumcoronene has $\eta = 3$ and has of course the coronene core, but in substituted circumcoronenes ($\eta > 3$) it may be natural to consider circumcoronene as the core. Circumcircumcoronene may be considered as the core in some hexagonal benzenoids with $\eta > 6$ and so on.

It is an interesting fact that all hexagonal benzenoids have $\Delta = 0$. Therefore, they are either Kekuléan or concealed non-Kekuléan.[24] Concealed non-Kekuléans are defined as systems with no Kekulé structures, but $\Delta = 0$, where Δ is defined as above and also is equal to the absolute magnitude of the difference between the number of black and white vertices.[24,26]

Computer Programming for Benzenoids with Hexagonal Symmetry. A special program was designed following the same principles as before[8] but with some marked differences.

First, the identification of a hexagonal benzenoid is made by means of the skew coordinate system with allowance for negative numbers (see Figure 2). Conventionally, the coordinates (0,1) are always taken first. In addition, it is sufficient to specify the coordinates for only one of the six arms, as is exemplified in the legend of Figure 2.

Second, new benzenoids are generated by adding six symmetrically equivalent hexagons each time. Again different benzenoids may be used as a starting point, and different options are possible for the types of addition.

Also in this case an analysis of symmetry was performed; here it means only a distinction between D_{6h} and C_{6h}.

A special program was employed to calculate the number of Kekulé structures (K) for every benzenoid. This program basically follows the principles of Brown.[31]

Two options of the program were employed: (i) starting with coronene and adding hexagons in all positions in order to generate all hexagonal benzenoids; (ii) starting with different basic benzenoids and adding hexagons in modes with only one
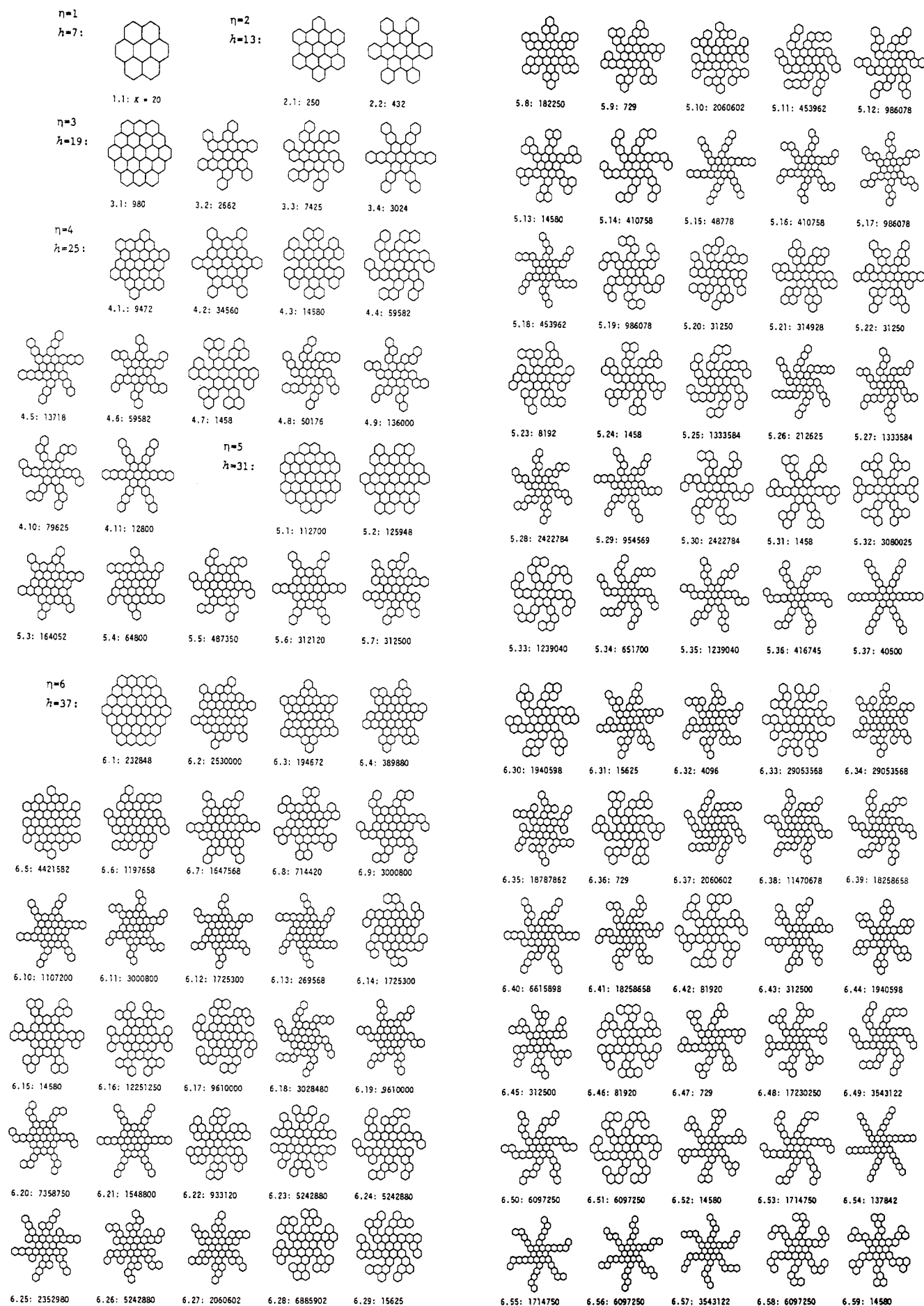
edge contact. The second option was previously described under point ii.[8]

Figure 3 shows the actual forms of the benzenoids completely through $\eta = 6$. For the sake of brevity, only selected representatives from the 540 existing hexagonal benzenoids with $\eta = 7$ are included in the figure. The K numbers are given. The figure simulates the style of Balasubramanian et al.[5] and the Düsseldorf–Zagreb school.[1,4] All the drawings are computer-generated by a screen-printing procedure. Apart from the mathematical and chemical interest, this research has an apparent aesthetic quality.

Enumeration and Classification of Benzenoids with Hexagonal Symmetry. Table V shows the derived numbers of benzenoids, where the classification into the two hexagonal symmetries is included. The analysis proves definitely that there exist two smallest ($h = 19$) benzenoids of symmetry C_{6h},[30] viz., nos. 3.2 and 3.3 of Figure 3. Only one of them (no. 3.3) was identified as the smallest C_{6h} benzenoid in a previous publication.[27]

The benzenoids are generated in such a way that derivatives of coronene, circumcoronene, circumcircumcoronene, etc. follow each other lexigraphically. The appropriate numbers of each kind are collected in Table VI.

A benzenoid is essentially disconnected[23] when it contains edges for which the bonds are fixed (single or double) in all Kekulé structures.[23,32] The two smallest representatives are found for $\eta = 4$ and depicted as nos. 4.3[29,30] and 4.4.[30] The numbers of the hexagonal essentially disconnected benzenoids
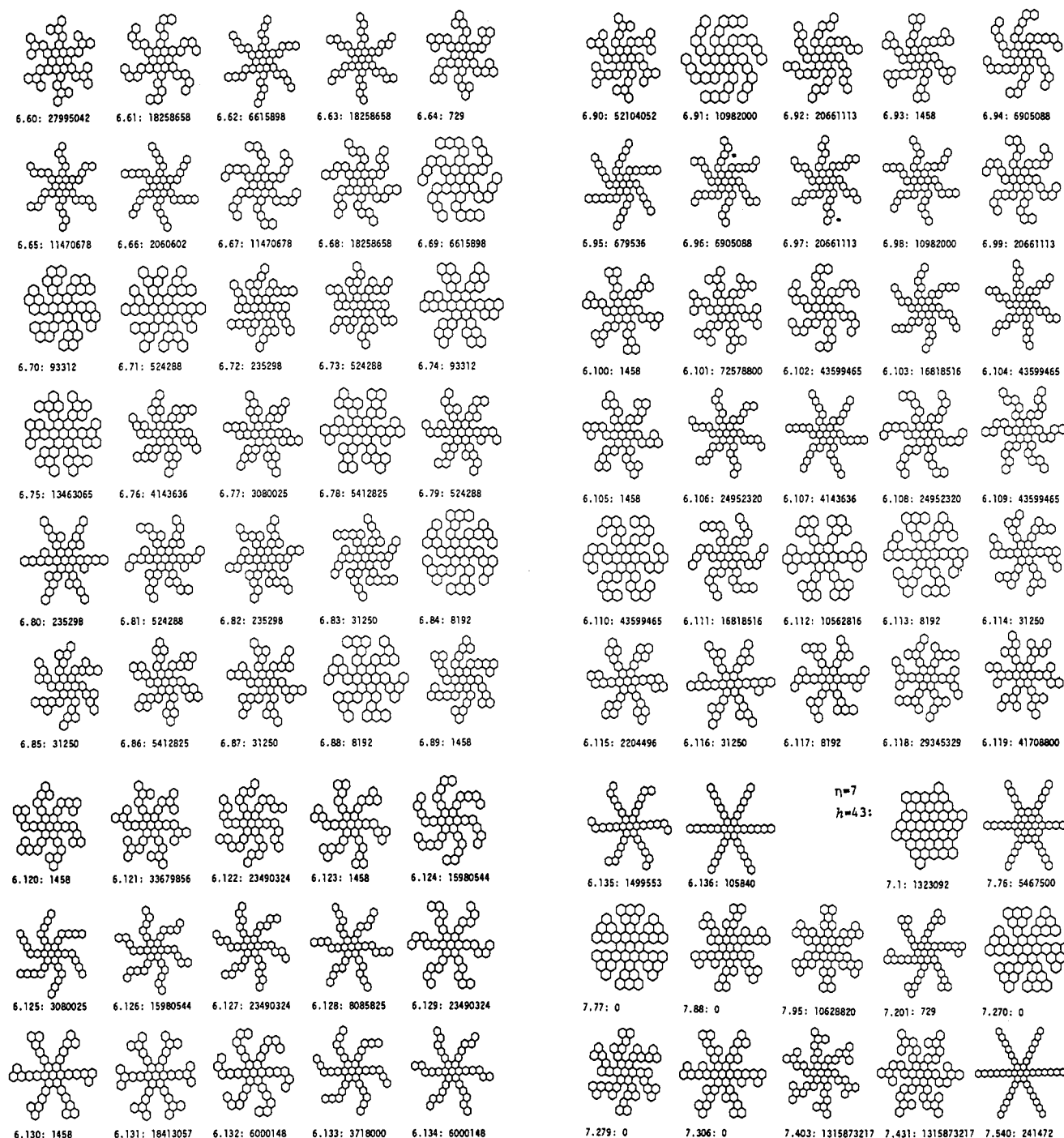
$\eta=1$
$h=7$:

1.1: $X = 20$

$\eta=2$
$h=13$:

2.1: 250

2.2: 432

$\eta=3$
$h=19$:

3.1: 980

3.2: 2662

3.3: 7425

3.4: 3024

$\eta=4$
$h=25$:

4.1.: 9472

4.2: 34560

4.3: 14580

4.4: 59582

4.5: 13718

4.6: 59582

4.7: 1458

4.8: 50176

4.9: 136000

4.10: 79625

4.11: 12800

$\eta=5$
$h=31$:

5.1: 112700

5.2: 125948

5.3: 164052

5.4: 64800

5.5: 487350

5.6: 312120

5.7: 312500

5.8: 182250

5.9: 729

5.10: 2060602

5.11: 453962

5.12: 986078

5.13: 14580

5.14: 410758

5.15: 48778

5.16: 410758

5.17: 986078

5.18: 453962

5.19: 986078

5.20: 31250

5.21: 314928

5.22: 31250

5.23: 8192

5.24: 1458

5.25: 1333584

5.26: 212625

5.27: 1333584

5.28: 2422784

5.29: 954569

5.30: 2422784

5.31: 1458

5.32: 3080025

5.33: 1239040

5.34: 651700

5.35: 1239040

5.36: 416745

5.37: 40500

$\eta=6$
$h=37$:

6.1: 232848

6.2: 2530000

6.3: 194672

6.4: 389880

6.5: 4421582

6.6: 1197658

6.7: 1647568

6.8: 714420

6.9: 3000800

6.10: 1107200

6.11: 3000800

6.12: 1725300

6.13: 269568

6.14: 1725300

6.15: 14580

6.16: 12251250

6.17: 9610000

6.18: 3028480

6.19: 9610000

6.20: 7358750

6.21: 1548800

6.22: 933120

6.23: 5242880

6.24: 5242880

6.25: 2352980

6.26: 5242880

6.27: 2060602

6.28: 6885902

6.29: 15625

6.30: 1940598

6.31: 15625

6.32: 4096

6.33: 29053568

6.34: 29053568

6.35: 18787862

6.36: 729

6.37: 2060602

6.38: 11470678

6.39: 18258658

6.40: 6615898

6.41: 18258658

6.42: 81920

6.43: 312500

6.44: 1940598

6.45: 312500

6.46: 81920

6.47: 729

6.48: 17230250

6.49: 3543122

6.50: 6097250

6.51: 6097250

6.52: 14580

6.53: 1714750

6.54: 137842

6.55: 1714750

6.56: 6097250

6.57: 3543122

6.58: 6097250

6.59: 14580

6.60: 27995042 6.61: 18258658 6.62: 6615898 6.63: 18258658 6.64: 729
6.65: 11470678 6.66: 2060602 6.67: 11470678 6.68: 18258658 6.69: 6615898
6.70: 93312 6.71: 524288 6.72: 235298 6.73: 524288 6.74: 93312
6.75: 13463065 6.76: 4143636 6.77: 3080025 6.78: 5412825 6.79: 524288
6.80: 235298 6.81: 524288 6.82: 235298 6.83: 31250 6.84: 8192
6.85: 31250 6.86: 5412825 6.87: 31250 6.88: 8192 6.89: 1458
6.120: 1458 6.121: 33679856 6.122: 23490324 6.123: 1458 6.124: 15980544
6.125: 3080025 6.126: 15980544 6.127: 23490324 6.128: 8085825 6.129: 23490324
6.130: 1458 6.131: 18413057 6.132: 6000148 6.133: 3718000 6.134: 6000148

6.90: 52104052 6.91: 10982000 6.92: 20661113 6.93: 1458 6.94: 6905088
6.95: 679536 6.96: 6905088 6.97: 20661113 6.98: 10982000 6.99: 20661113
6.100: 1458 6.101: 72578800 6.102: 43599465 6.103: 16818516 6.104: 43599465
6.105: 1458 6.106: 24952320 6.107: 4143636 6.108: 24952320 6.109: 43599465
6.110: 43599465 6.111: 16818516 6.112: 10562816 6.113: 8192 6.114: 31250
6.115: 2204496 6.116: 31250 6.117: 8192 6.118: 29345329 6.119: 41708800
6.135: 1499553 6.136: 105840 n=7 h=43: 7.1: 1323092 7.76: 5467500
7.77: 0 7.88: 0 7.95: 10628820 7.201: 729 7.270: 0
7.279: 0 7.306: 0 7.403: 1315873217 7.431: 1315873217 7.540: 241472

Figure 3. All hexagonal benzenoids with $\eta = 1$–6 and selected representatives of $\eta = 7$.

for $\eta = 5$ and $\eta = 6$ (see Table VII) were easily determined by inspection of the drawings (Figure 3). For $\eta > 6$ the inspection was deemed to be unreliable, and therefore a computer-aided recognition of the essentially disconnected systems was devised. It has been conjectured[20] and supported by a great deal of evidence that any normal benzenoid may be generated by adding one hexagon to a smaller normal benzenoid. Therefore, an algorithm was introduced that tears down a benzenoid hexagon by hexagon, assuring that the process runs through normal benzenoids only. If the process goes successfully down to one single hexagon, it is taken for granted that the original benzenoid is normal. In the opposite case, however, the benzenoid may still be normal because it is possible that a different route of tearing down hexagons could meet with success. Therefore, we have not invented a fully computerized recognition of essentially disconnected benzenoids. The derived set must still be inspected for sorting

out possible normal benzenoids left over, but the number of benzenoids to be inspected will be lowered. In the present case (Table VII) the method eliminated absolutely all normal benzenoids for $\eta \le 6$. For $\eta = 7$ the essentially disconnected and non-Kekuléan benzenoids came together as expected, and only one normal system was left over to be eliminated by inspection. The K values for essentially disconnected benzenoids are known to be composite numbers. A survey of them for $\eta = 4$–6 is given in Table VIII.

A useful classification of benzenoids in general takes into account the categories normal (n), essentially disconnected (e), and non-Kekuléan (o), the "neo" classification[19-21] (cf. also Table VII). In the case of hexagonal benzenoids all non-Kekuléans are concealed (see above). The two classes e and o are treated in the next paragraph in some detail.

Essentially Disconnected Benzenoids with Hexagonal Symmetry. There are relatively few essentially disconnected

176 *J. Chem. Inf. Comput. Sci., Vol. 27, No. 4, 1987*

BRUNVOLL ET AL.

Table VIII. K Numbers for Essentially Disconnected Hexagonal Benzenoids

K		no. of benzenoid with reference to Figure 3
$3^6 = 729$	5.9	6.36, 6.47, 6.64
$2 \times 3^6 = 1458$	4.7 5.24, 5.31	6.89, 6.93, 6.100, 6.105, 6.120, 6.123, 6.130
$4^6 = 4096$		6.32
$2 \times 4^6 = 8192$	5.23	6.84, 6.88, 6.113, 6.117
$20 \times 3^6 = 14580$	4.3 5.13	6.15, 6.52, 6.59
$5^6 = 15625$		6.29, 6.31
$2 \times 5^6 = 31250$	5.20, 5.22	6.83, 6.85, 6.87, 6.114, 6.116
$20 \times 4^6 = 81920$		6.42, 6.46
$2 \times 6^6 = 93312$		6.70, 6.74
$2 \times 7^6 = 235298$		6.72, 6.80, 6.82
$20 \times 5^6 = 312500$	5.7	6.43, 6.45
$2 \times 8^6 = 524288$		6.71, 6.73, 6.79, 6.81
$980 \times 3^6 = 714420$		6.8
$20 \times 6^6 = 933120$		6.22
$20 \times 7^6 = 2352980$		6.25
$20 \times 8^6 = 5242880$		6.23, 6.24, 6.26
$7425 \times 3^6 = 5412825$		6.78

benzenoids with the highest symmetry of D_{6h}. There is one with $\eta = 4$ (no. 4.3 of Figure 3), and one with $\eta = 6$ (no. 6.23), one with $\eta = 7$ (no. 7.95), and there are none with $\eta = 1, 2, 3,$ or 5. The rest of them (cf. Table VII) have the C_{6h} symmetry.

Most of these essentially disconnected benzenoids may be interpreted as consisting of six non-Kekuléan arms added to the core. These arms are obvious non-Kekuléans.[24] For $\eta \geq 7$ another type exists, where the arms are essentially disconnected themselves. Below we show one representative each of these two classes: non-Kekuléan arms added to circumcoronene ($K = 20 \times 3^6$), left, and essentially disconnected arms added to coronene ($K = 2662 \times 3^6$), right.

$K=14580$ $K=1940598$
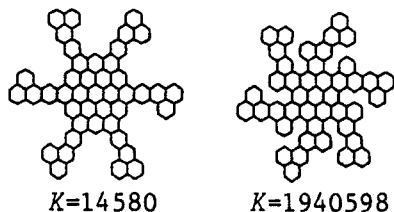
Concealed Non-Kekuléan Benzenoids with Hexagonal Symmetry. One of the most interesting aspects of this research concerns the hexagonal concealed non-Kekuléans. The analysis proved that no such system exists before $h = 43$ ($\eta = 7$), in agreement with Hosoya's[29] supposition. This author depicted the D_{6h} system of this kind (no. 7.77 of Figure 3). Cyvin et al.[30] found three systems of the C_{6h} symmetry (nos. 7.88, 7.270, and 7.306). A fifth system (no. 7.279) was generated in the present work. Simultaneously, it is proved that exactly these five hexagonal non-Kekuléans exist for $\eta = 7$ (cf. Table VII).

Annelated Benzenoids with Hexagonal Symmetry. By virtue of the coronene core all hexagonal benzenoids are pericondensed. Systems where coronene has only catacondensed appendages come closest to the catacondensed systems. They were enumerated with the result given in Figure 4. In our previous work (see Figure 4 of ref 8) we have, inter alia, enumerated the annelations to coronene in general, not only for hexagonal benzenoids. The figures are therefore not directly comparable and, besides, the previous enumeration[8] does not go beyond $h = 12$. Figure 4 includes the enumeration for hexagonal annelation to two additional basic benzenoids, viz., the hexabenzocoronene given as no. 2.1 of Figure 3 and circumcoronene (no. 3.1).

Benzenoids of Hexagonal Symmetry and Maximum Numbers of Kekulé Structures. The last issue concerns the hexagonal benzenoids with maximum K numbers for given η values. For

Number of hexagonal benzenoids with

Catacondensed annelations to:		(a)	(b)	(c)
$n = 2$	$h = 13$	1	-	-
3	19	2	1	-
4	25	4	3	1
5	31	12	9	2
6	37	35	27	6
7	43	116	91	16
8	49	390	301	52
9	55	1368	1055	176
10	61	4856	3722	614
11	67	†	†	2185
12	73	†	†	7914

†Unknown

Figure 4. Number of hexagonal benzenoids with catacondensed appendages to (a) coronene (no. 1.1), (b) hexabenzocoronene A (no. 2.1), and (c) circumcoronene (no. 3.1).

benzenoids in general these systems are found among the highly branched catacondensed benzenoids with many angular annelations (kinks).[19] Therefore, it is not surprising that the systems in question are found among the catacondensed annelations to coronene treated above and those with branched and kinked appendages. We have found specifically $K_{max} = 20, 432, 7425, 136000, 3080025, 72578800,$ and 1315873217 for $\eta = 1, 2, 3, 4, 5, 6,$ and 7, respectively. The actual forms are found in Figure 3 as nos. 1.1, 2.2, 3.3, 4.9, 5.32, and 6.101 along with an isoarithmic pair for $\eta = 7$, viz., 7.403 and 7.431.

CONCLUSION

Benzenoids in general have been enumerated through $h = 10^{1-4}$ and quite recently for $h = 11.$[11] The hexagonal systems are scattered very sparsely among them: among the 179 701 benzenoids with $h = 1-11$ there are only two hexagonal benzenoids. Therefore, there is no real hope to detect interesting features of the hexagonal benzenoids treated as a subset of all benzenoid systems; the present work shows that at least $h = 43$ is required to detect the really interesting features. This goal was reached by blocking out the hexagonal systems or, in other words, by generating them separately. A corresponding analysis of trigonal benzenoids might also be interesting.

ACKNOWLEDGMENT

REFERENCES AND NOTES

(1) Knop, J. V.; Szymanski, K.; Jeričević, Ž.; Trinajstić, N. *J. Comput. Chem.* **1983**, *4*, 23.
(2) Trinajstić, N.; Jeričević, Ž.; Knop, J. V.; Müller, W. R.; Szymanski, K. *Pure Appl. Chem.* **1983**, *55*, 379.
(3) Knop, J. V.; Szymanski, K.; Jeričević, Ž.; Trinajstić, N. *MATCH* **1984**, *16*, 119.
(4) Knop, J. V.; Müller, W. R.; Szymanski, K.; Trinajstić, N. *Computer Generation of Certain Classes of Molecules*; Association of Chemists and Technologists of Croatia: Zagreb, 1985.
(5) Balasubramanian, K.; Kaufman, J. J.; Koski, W. S.; Balaban, A. T. *J. Comput. Chem.* **1980**, *1*, 149.
(6) He, W.; He, W. *Theor. Chim. Acta* **1985**, *68*, 301.
(7) He, W.; He, W. *Tetrahedron* **1986**, *42*, 5291.
(8) Brunvoll, J.; Cyvin, S. J.; Cyvin, B. N. *J. Comput. Chem.* **1987**, *27*, 14.
(9) Gutman, I. *Z. Naturforsch., A: Phys., Phys. Chem., Kosmophys.* **1986**, *41A*, 1089.
(10) Cioslowski, J. *J. Comput. Chem.* (in press).
(11) Stojmenović, I.; Tošić, R.; Doroslovački, R. *Comput. Chem.* (in press).
(12) Dias, J. R. *J. Chem. Inf. Comput. Sci.* **1982**, *22*, 15.
(13) Dias, J. R. *J. Chem. Inf. Comput. Sci.* **1984**, *24*, 124.
(14) Dias, J. R. *Can. J. Chem.* **1984**, *62*, 2914.

(15) Dias, J. R. *Acc. Chem. Res.* **1985**, *18*, 241.
(16) Dias, J. R. *J. Macromol. Sci., Chem.* **1985**, *A22*, 335.
(17) Dias, J. R. *Nouv. J. Chim.* **1985**, *9*, 125.
(18) Dias, J. R. *THEOCHEM* **1986**, *137*, 9.
(19) Cyvin, S. J. *MATCH* **1986**, *20*, 165.
(20) Cyvin, S. J.; Gutman, I. *Z. Naturforsch., A: Phys., Phys. Chem., Kosmophys.* **1986**, *41A*, 1079.
(21) Cyvin, B. N.; Brunvoll, J.; Cyvin, S. J.; Gutman, I. *MATCH* **1986**, *21*, 301.
(22) Cyvin, S. J.; Brunvoll, J.; Cyvin, B. N. *Z. Naturforsch., A: Phys., Phys. Chem., Kosmophys.* **1986**, *41A*, 1429.
(23) Cyvin, S. J.; Cyvin, B. N.; Gutman, I. *Z. Naturforsch., A: Phys., Phys. Chem., Kosmophys.* **1985**, *40A*, 1253.
(24) Cyvin, S. J.; Gutman, I. *THEOCHEM* **1987**, *150*, 157.
(25) In the previous work[8] it was overlooked that the branched catacondensed benzenoids in fact also can be generated specifically: Start from triphenylene and apply additions of hexagons in the position [1], i.e., so that one added hexagon shares exactly one edge with the original benzenoid.
(26) Gutman, I. *Croat. Chem. Acta* **1974**, *46*, 209.
(27) Cyvin, S. J.; Gutman, I. *Comp. Maths. Appl.* **1986**, *12B*, 859.
(28) Balaban, A. T.; Harary, F. *Tetrahedron* **1968**, *24*, 2505.
(29) Hosoya, H. *Comp. Maths. Appl.* **1986**, *12B*, 271.
(30) Cyvin, S. J.; Bergan, J. L.; Cyvin, B. N. *Acta Chim. Hung.* (in press).
(31) Brown, R. L. *J. Comput. Chem.* **1983**, *4*, 556.
(32) Cyvin, S. J.; Gutman, I. *J. Serb. Chem. Soc.* **1985**, *50*, 443.

Computer Storage and Retrieval of Generic Chemical Structures Using Structure Attributes

SOICHI TOKIZANE* and TSUKASA MONJOH

Japan Association for International Chemical Information, Bunkyo-Ku, Tokyo, 113 Japan

HIDEAKI CHIHARA

Faculty of Sciences, Osaka University, Toyonaka, 560 Japan

A method to store and search chemical structure information with generic expressions using digital computers was developed. The method uses specially designed sets of attributes to represent chemical characteristics of both generic and specific expressions. Three types of attributes, ring, chain, and atom attributes, are defined and described. Attributes are expressed as bit-map vectors, which makes them easy to handle by digital computers. A search algorithm and procedure using a modified connectivity stack method to find the match between a query structure and a file structure with generic expressions are described.

INTRODUCTION

Chemical structure searching is a most sophisticated type of information retrieval using computers. One of the biggest and most comprehensive chemical structure search systems, the CAS ONLINE Registry File, is now available commercially worldwide from Chemical Abstracts Service, a division of the American Chemical Society. The Registry File is searchable online by creating a query structure interactively on a graphic terminal and then conducting a search of the file of over 8 000 000 chemical substances collected from journal and patent references since the 1960s. The answer structures can be displayed on the same graphical terminal. Any current implementation of the structure search system can handle so-called specific chemical structures only.

Generic expression of chemical structure has been widely used since the early days to describe a broad range of chemical structures with a concise drawing of the structure containing variables and class expressions. Generic chemical structures are most extensively used in patent literatures, where it is natural to express the scope of their invention as widely as possible, and the use of generic structures (or Markush structures after the name of an American inventor) is generally accepted as a rule. The current status of the handling of generic chemical structures is discussed thoroughly in recent conference proceedings.[1] Several papers were published describing basic concepts and techniques to handle generic chemical structures on a computer. Lynch and his co-workers[2–8,10,11] and Scholley[9] have been working on this topic extensively. A method of handling generic chemical structures was also proposed by Kudo and Chihara.[12]

A typical generic (Markush) chemical structure is shown in Figure 1. There are several characteristics that are used in generic (Markush) chemical structures as seen in Figure 1. They are (1) alternative structure components (as in R1), (2) generic expressions (as in R2), (3) nonfixed attachments

Table I. Types of Chemical Structure Matching by Fisanick

	query structure	file structure
overlap matching criteria (1)	full	specific
overlap matching criteria (2)	full	generic
overlap matching criteria (3)	Markush	specific
overlap matching criteria (4)	Markush	generic
embedment matching criteria (1)	substructure	specific
embedment matching criteria (2)	substructure	generic
embedment/overlap matching criteria	substructure	generic

(as in R3), (4) undefined structure components (as in R3), (5) variable counts (as in R1), and (6) ring node variables (as in X).

Variable structure definitions such as 1, 3, 4, and 5 can be handled without much difficulty if enough computing power is available. Existing systems are already able to handle query structures containing variable components. It is thus possible to extend this capability to handle file structures with variable parts. One approach to this topic is the Colored Complete Graph method described by Kudo and Chihara.[12] On the other hand, storing and retrieving chemical structures with generic groups as seen in the group R2 is a very difficult topic in principle. We will focus on how to store and search structures containing generic groups in this paper.

A generic group may be described by the following components: (1) specific or range of atom count, (2) specific or range of element count, (3) specific or range of element types, (4) specific or range of bond composition (usually other than single bonds), (5) specific or range of branches or substitutions, (6) specific or range of ring size, (7) specific or range of number of component rings, and (8) other textural information.

The above expressions may appear either independently or combined. An example of the generic expression is shown in Figure 2. Numbers in brackets correspond to the group numbers specified above. Unknown data are regarded as