a few people, however, who recognize a need to use the system but, for reasons of their own, choose to have a search intermediary do it for them or at least work with them on a graphical search session. At this point, therefore, it appears that the role of the search specialist has evolved away from searches originated by individual users (except in a few cases) who now can and should run their own searches, toward a function which satisfies information needs at project team and management levels.

REFERENCES AND NOTES

(1) For a representative sample of applications of interactive graphics for structure entry, see the following articles and references cited therein: (a) Corey, E. J.; Wipke, W. T.; Cramer, R. D.; Howe, W. J. "Computer-Assisted Synthetic Analysis. Facile Man-Machine Communication of Chemical Structure by Interactive Computer Graphics". *J. Am. Chem. Soc.* **1972**, *94*, 421. (b) "Computer Assisted Organic Synthesis"; Wipke, W. T., Howe, W. J., Eds.; American Chemical Society: Washington, DC, 1977; ACS Symp. Ser. No. 61, Chapters 1, 5, 8. (c) Salatin, T. D.; Jorgensen, W. L. "Computer-Assisted Mechanistic Evaluation of Organic Reactions. 1. Overview". *J. Org. Chem.* **1980**, *45*, 2043. (d) "Computer Representation and Manipulation of Chemical Information"; Wipke, W. T., Heller, S. R., Feldmann, R. J., Hyde, E., Eds.; Wiley: New York, 1974; Chapters 3, 7. (e) Howe, W. J.; Hagadone, T. R. "Progress Toward an On-Line Chemical and Biological Information System at The Upjohn Company"; American Chemical Society: Washington, DC, 1978; ACS Symp. Ser. No. 84, Chapter 8. (f) Howe, W. J.; Hagadone, T. R. "Chemical Substructure Searching"; Proceedings of the Manufacturing Chemists Association Meeting, Arlington VA, Aug 1977; Manufacturing Chemists Association: Washington, DC, 1977. (g) Saxberg, B. E. H.; Blom, D. S.; Kowalski, B. R. "A Flexible Interactive Graphics System for Searching Atom Connectivity Matrices". *J. Chem. Inf. Comput. Sci.* **1976**, *16*, 233. (h) Blake, J. E.; Farmer, N. A.; Haines, R. C. "An Interactive Computer Graphics System for Processing Chemical Structure Diagrams". *Ibid.* **1977**, *17*, 223. (i) Dyott, T. M.; Stuper, A. J.; Zander, G. S. "Moly—An Interactive System for Molecular Analysis". *Ibid.* **1980**, *20*, 28.

(2) For a representative sample of the relatively few applications of interactive graphics for substructure query entry, see ref 3 and the following sources and references cited therein. (a) Dyott, T. M., et al. "An Integrated System for Conducting Chemical and Biological Searches"; American Chemical Society: Washington, DC, 1978; ACS Symp. Ser. No. 84, Chapter 11. (b) Blower, P. E.; Peercy, R. R.; Wade, L. G. "Design Your Own Information Service"; Chemical Abstracts Service, Columbus, OH, 1980; CAS Report 9. (c) The MACCS System, Molecular Design Ltd., Hayward, CA.

(3) "Substructure Searching of Large Chemical Files"; McNulty, P. J., Smith, R. B., Eds.; Proceedings of the Manufacturing Chemists Association Meeting, Arlington, VA, Aug 1977; Manufacturing Chemists Association: Washington, DC, 1977.

(4) COUSIN stands for "CompOUnd Search INformation system" and is a trademark of the Upjohn Company.

(5) Hagadone, T. R.; Howe, W. J., manuscript in preparation.

(6) Examples of Markush notations are seen in the following references: Kaback, S. M. "Chemical Structure Searching in Derwent's World Patent index". *J. Chem. Inf. Comput. Sci.* **1980**, *20*, 1. Sneed, H. M. S.; Turnipseed, J. H.; Turpin, R. A., Jr. "A Line Formula Notation System for Markush Structures". *J. Chem. Doc.* **1968**, *8*, 173.

(7) Brown, H. D., et al. "The Computer-Based Chemical Structure Information System of Merck Sharp and Dohme Research Laboratories". *J. Chem. Inf. Comput. Sci.* **1976**, *16*, 5.

A Periodic Table for Polycyclic Aromatic Hydrocarbons. Isomer Enumeration of Fused Polycyclic Aromatic Hydrocarbons.† 1

JERRY RAY DIAS

Department of Chemistry, University of Missouri, Kansas City, Missouri 64110

Received June 26, 1981

A comprehensive framework for the enumeration of all the isomers of even-carbon, fused polycyclic aromatic hydrocarbons (PAH) containing only hexagonal rings is defined. The basis for this framework is the molecular formula in contrast to the number of hexagonal rings. From the molecular formula, one can compute the number of rings from $r = (1/2)(N_c + 2 - N_H)$ and the total number of edges (σ C–C bonds) from $q = (1/2)(3N_c - N_H)$. These and other relationships are derived for the first time. Criteria for stipulating whether some specific molecular formula can be represented by a PAH containing only hexagonal rings is presented. An approach for possible computer enumeration of PAH isomers is proposed. The maximum number of five-membered rings that a structure corresponding to a PAH6 formula can contain in addition to hexagonal rings is given by $r_{5max} \leq N_c - 2N_H + 6$.

Very few attempts have been made to systematically enumerate all possible polycyclic aromatic hydrocarbons (PAH; PNA for polynuclear aromatic compounds has been synonymously used).[1-4] In this paper, the scope and framework for achieving this goal is defined. The basis for this framework is the molecular formula in contrast to the number of hexagonal rings.[2] It is believed that this approach will be applicable to computer methodology. Also, a number of graph theoretical observations and theorems relevant to PAH's are espoused for the first time, and criteria for determining whether some specific formula can be represented by a PAH within the specified constraints are presented.

RESULTS AND DISCUSSION

Totally fused (condensed) PAH's having even-carbon nonradical systems possessing exclusively hexagonal rings will be discussed first (designated PAH6). Then an examination of analogue PAH systems containing pentagonal rings will be briefly pursued. Aromatic hydrocarbon systems that are not

Table I. Glossary of Terms

d_i	degree of vertex i of a graph
d_s	tree disconnections (of internal graph edges)
N_c	total number of carbon atoms in a PAH
N_H	total number of hydrogen atoms in a PAH
N_{Ic}	number of internal carbon atoms in a PAH having a degree of 3
N_{Pc}	number of peripheral carbon atoms in a PAH having a degree of 3
PAH6	fused polycyclic aromatic hydrocarbons possessing exclusively hexagonal rings
p_3	number of graph points (vertexes) having a degree of 3
q	number of graph edges (lines or C–C bonds)
q_I	number of internal graph edges
q_P	number of peripheral graph edges
r	number of rings
r_{5max}	maximum number of pentagonal rings

† This work was presented as a special seminar to the research group of Professor Nenad Trinajstić at the Rudger Bošković Institute in Zagreb, Yugoslavia, on May 21, 1981.

0095-2338/82/1622-0015$01.25/0

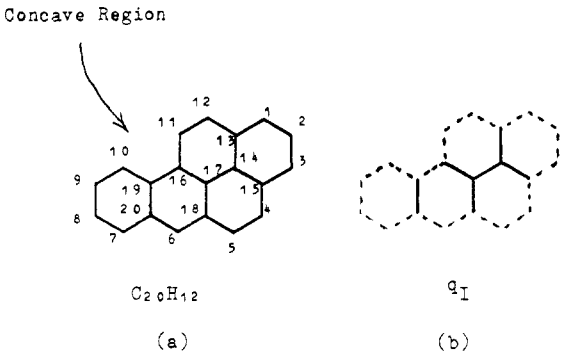
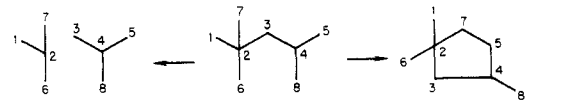


Figure 1. Graph of (a) benzo[*a*]pyrene and (b) its internal edges, q_I.

totally fused (e.g., α-phenylnaphthalene) or are even polyradicals (e.g., coeranthrene) or a ring size different from a hexagon or pentagon (e.g., tetraphenylene) will not be considered.

Terminology. A glossary of terms used in this text is presented in Table I. The meaning of these terms may be illustrated by the molecular graph (only the σ C–C bonds will be shown throughout this manuscript) of benzo[*a*]pyrene ($C_{20}H_{12}$), a well-known carcinogen, shown in Figure 1a. Vertexes 1–12 have a degree of 2 (i.e., have two incident lines) and are counted by N_H. Similarly, vertexes 13–20 have a degree of 3 (p_3) where the vertexes 13, 15, 16, 18–20 are counted by N_{Pc} and vertexes 14 and 17 by N_{Ic}. The total number of vertexes 1–20 is counted by N_c and the total number of edges (bonds) by q. Figure 1b presents the internal edges, q_I, as bold lines and collectively is called a broken tree with a single disconnection (d_s) between points 16 and 19.

Relationships and Equations Applicable to PAH. Table II presents a summary of equations that are applicable to totally fused PAH's and, in general, to graphs composed of a web of hexagons. Equation 1 states that the total number of vertexes (N_c) is equal to the sum of the vertexes of degree 3 (N_{Pc} and N_{Ic}) and vertexes of degree 2 (N_H). Equation 3 relates the total number of edges (q) to the sum of the internal (q_I) and peripheral (q_P) edges. It is believed that eq 2, 4, and 5 are empirical relations observed herein for the first time. Specialization of well-known theorems by Euler[5] lead to eq 6 and 8. If $G(p,q)$ is an unbroken tree graph of p vertexes connected by q lines (e.g., the carbon skeleton of isooctane has $p = 8$ and $q = 7$), then $p = q + 1$. Defining d_s as being the number of disconnections (positive integers 1, 2, 3, ···) or connections (negative intergers −1, −2, −3, ···) associated with the removal or addition of one or more lines (bonds), respectively, while retaining a constant number of vertexes (carbon atoms) results in eq 7 for unbroken and broken trees which has been specialized for the internal edges of fused PAH's; Figure 1b is a broken tree with a single disconnection. The carbon skeleton of isooctane can be used to illustrate the $p = q + 1 + d_s$ equation. The unbroken tree of isooctane may undergo a single disconnection ($d_s = 1$) between vertexes 2 and 3 or a single connection ($d_s = -1$) between vertexes 5 and 7 (cyclization) as shown by the following:



Formula Table for PAH6. A comprehensive summary of all possible PAH6 formulas for compounds containing exclusively fused hexagonal rings is provided by Table III. In Table III every compound in a particular column has the same d_s value, and every compound in the same row has the same

Table II. Equations Applicable to Fused PAH's

(1) $N_c = N_{Pc} + N_{Ic} + N_H$
(2) $N_{Pc} = N_H - 6$ for fused PAH's except circulenes
$N_{Pc} = N_H$ for monocirculenes only
(3) $q = q_I + q_p$
(4) $q_p = q - q_I = 2N_H - 6$ for fused PAH's except circulenes
$q_p = 2N_H$ for monocirculenes only
(5) $d_s = N_{Pc} - r$
(6) $2q = \Sigma d_1 = \Sigma d_2 + \Sigma d_3 = 2N_H + 3(N_c - N_H) = 3N_c - N_H$
(7) $p_3 = q_I + 1 + d_s = N_c - N_H = N_{Pc} + N_{Ic}$
(8) faces − edges + vertexes = 2
$(r + 1) - q + p = 2$ or $r = q + 1 - N_c =$
$(1/2)(N_c + 2 - N_H)$
(9) For the column of Table III starting with $C_{10}H_8$
$C_{10+3(r-2)}H_{8+(r-2)}$ and $\lim_{r\to\infty} N_c/N_H =$

$$\lim_{r\to\infty} \frac{10 + 3(r-2)}{8 + (r-2)} = 3$$

(10) For the row of Table III starting with $C_{10}H_8$
$C_{10+4(r-2)}H_{8+2(r-2)}$ and $\lim_{r\to\infty} N_c/N_H =$

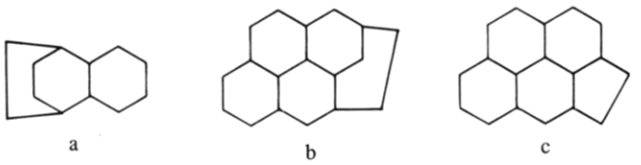
$$\lim_{r\to\infty} \frac{10 + 4(r-2)}{8 + 2(r-2)} = 2$$

(11) $r_{smax} \leq N_{Ic} = N_C - 2N_H + 6$

N_{Ic} value. Member compounds in the same column with the same d_s value belong to the same column series ($N_c = 3N_H - 14 - 2d_s$, where d_s is specified), and member compounds in the same row with the same N_{Ic} value belong to the same row series ($N_c = 2N_H - 6 + N_{Ic}$, where N_{Ic} is specified). For example, all PAH6 isomers of the formulas $C_{16}H_{10}$ and $C_{20}H_{12}$ have the same value of $N_{Ic} = 2$, and similarly, all PAH6 isomers of the formulas $C_{20}H_{12}$ and $C_{26}H_{14}$ have the same value of $d_s = 1$. Table III extends infinitely in three directions—horizonally to the right, vertically to the bottom, and in a slanting direction to the left. The ratio of N_c/N_H for any column approaches 3 as the number of rings increases and is illustrated by eq 9 for the column starting with $C_{10}H_8$, the $N_c = 3N_H - 14$ column series. Every formula in the column beginning with $C_{54}H_{18}$ belongs to the $N_c = 3N_H$ column series, and thus $N_c/N_H = 3$ for every member in this column. Columns to the left of the $N_c = 3N_H$ series have N_c/N_H ratios that approach 3 from larger values and columns to the right from smaller values. Similarly, the ratio of N_c/N_H for any row approaches 2 as the number of rings increases and is illustrated by eq 10 for the row starting with $C_{10}H_8$, the $N_c = 2N_H - 6$ row series. Every formula in the row beginning with $C_{24}H_{12}$ belongs to the $N_c = 2N_H$ row series, and thus $N_c/N_H = 2$ for every member in this series. Rows above the $N_c = 2N_H$ series have N_c/N_H ratios that approach 2 from smaller values and the rows below from larger values.

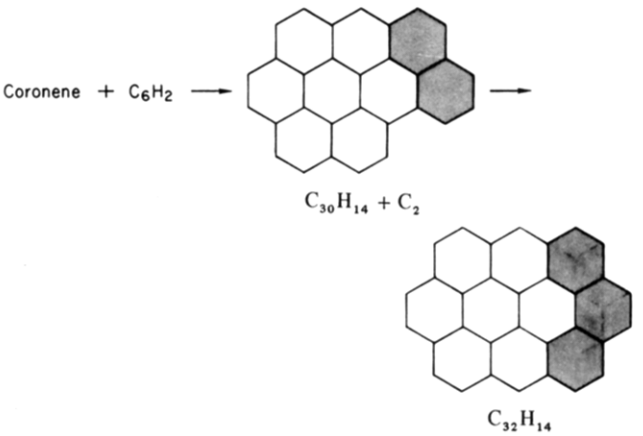
The triangular quadrant to the upper left of Table III is the domain of forbidden graphs for PAH6. In the first row (the $N_c = 2N_H - 6$ row series), a possible formula of C_6H_6 immediately preceding $C_{10}H_8$ is excluded since benzene is not a *fused* PAH. In the second row, no $C_{12}H_8$ compound is possible which would have $N_{Ic} = 2$, $d_s = -1$, and possess only reasonable hexagonal rings; thus compound a has the formula $C_{12}H_8$ and $d_s = -1$ but possesses a sterically forbidden hexagonal ring. A third row compound having the formula $C_{18}H_{10}$ must have $N_{Ic} = 4$ and $d_s = -1$; structures b and c fit this formula, but b contains a sterically forbidden hexagonal ring and $d_s = 0$ and c contains a pentagonal ring and $d_s = 0$. Also, structures a–c do not have consistent N_{Ic} values. This row-by-row analysis could be continued, but the above examples should be ample.

A procedure for determining whether a formula can be represented by a PAH6 is provided by Table IV. It is obvious from Table III that no row series exists above the $N_c = 2N_H$

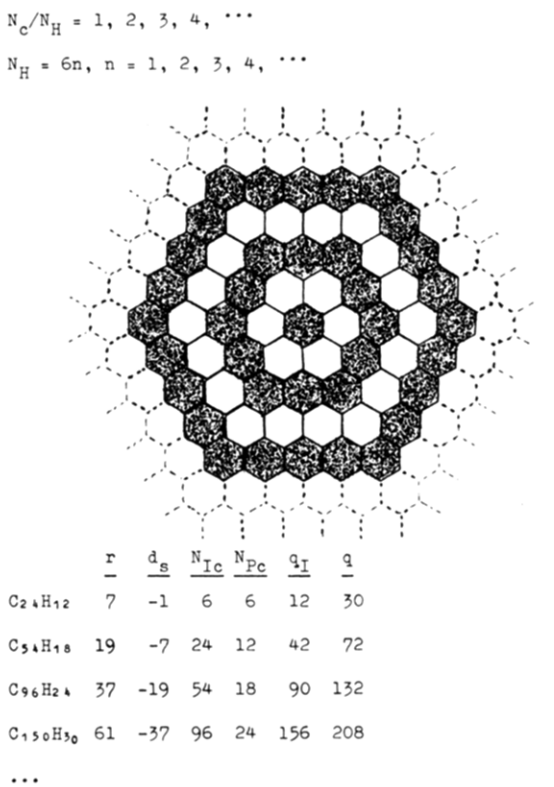


– 6 series, and thus any formula which has a composition such that $N_c < 2N_H - 6$ cannot be a PAH6. Furthermore, the upper left triangular region in Table III is another domain of forbidden formulas for PAH6 that are determined by steps 2 and 3 in Table IV using the general formula for the diagonal members of Table III of $C_{24+6m+2n}H_{12+2m}$. For example, consider the formulas $C_{20}H_{14}$ and $C_{48}H_{16}$. Per Table IV, the first formula has a C and H composition such that $N_c < 2N_H - 6$ ($20 < 28 - 6$) and, therefore, cannot be represented by a PAH6. In the latter formula $N_c = 48$ is greater than $2N_H - 6 = 26$ and must be evaluated as per steps 2 and 3 of Table IV. Step 2: m is computed to be 2 by equating $16 = 12 + 2m$ and n is computed to be 6 by equating $24 + 6m + 2n = 48$. Step 3: the corresponding value for m from the list where $n = 6$ is 3 which is greater than the value of 2 obtained above. Thus, the latter formula of $C_{48}H_{16}$ cannot be represented by PAH6.

Coronene Series. The coronene one isomer series is presented by the successive circular rows of fused hexagons in Figure 2. The formulas for this series fall on the diagonal of Table III and represent limiting cases for those respective rows. Benzene might be regarded as the first member of this series where $N_c/N_H = 1$, but it is not a fused PAH. Coronene ($C_{24}H_{12}$) itself has six fused hexagons around a benzene system (first unshaded ring of hexagons around a benzene center in Figure 2); this compound will be designated coronene-7 to emphasize that it is composed of seven hexagonal rings. The next member ($C_{54}H_{18}$) of this series will be named coronene-19 (first shaded ring of 12 hexagons around a coronene-7 system in Figure 2). The $C_{96}H_{24}$ member will be called coronene-37, and so on. Figure 2 summarizes all the pertinent graphical data associated with first several members of this series. Note that the successive d_s values correspond to the number of rings in the prior member of the series. The formulas on the left-hand diagonal of Table III between the various numbers of the coronene series (Figure 2) are derived by successive attachment of C═CH—C═CH—C═C and C═C units (the latter to concave regions). For example, attachment of a C═CH—C═CH—C═C unit to coronene-7 leads to $C_{30}H_{14}$ which gives $C_{32}H_{14}$ upon attachment of a C═C unit (vida infra) to a concave region (defined in Figure 1). This is illustrated by the following:



Circulenes. A monocirculene is defined as a PAH that possesses a "hole" larger than a hexagon which contains two or more hydrogens. Figure 3 gives a number of examples of monocirculenes. Many of these monocirculenes, no doubt,



	r	d_s	N_{Ic}	N_{Pc}	q_I	q
$C_{24}H_{12}$	7	-1	6	6	12	30
$C_{54}H_{18}$	19	-7	24	12	42	72
$C_{96}H_{24}$	37	-19	54	18	90	132
$C_{150}H_{30}$	61	-37	96	24	156	208
···						

Figure 2. Coronene one isomer series.

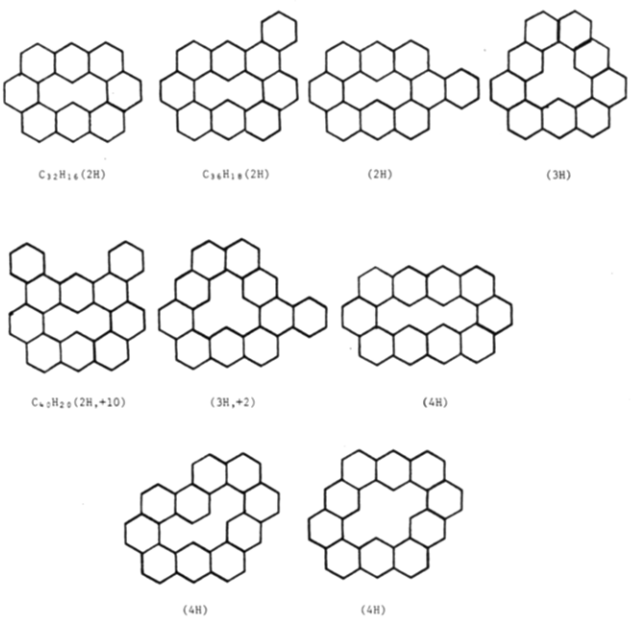


Figure 3. Circulenes.

contain severe steric interaction of the internal hydrogens. However, the last (lower right hand) monocirculene in Figure 3 has been synthesized,[6] and the steric interaction of the internal hydrogens in many of them can be reduced by twisting the PAH structure in a fashion similar to the way hexahelicene and heptahelicene[7] are twisted. Regardless of the probable stability, and therefore existence, of some of these monocirculenes, some interesting graph theoretical relationships will become evident in the following discussion. Note that the formulas for monocirculenes first appear in Table III in the $N_c = 2N_H$ row series and $N_c = 3N_H - 16$ column series. Thus, if a PAH6 formula has a C/H composition such that $3N_H - 16 < N_c < 2N_H$, then this formula cannot be represented by a monocirculene structure; i.e., monocirculenes can only exist

Table III. All the Formulas Possible for Polycyclic Aromatic Hydrocarbons Having Nonradical Hexagonal Rings—PAH Periodic Table

$N_c = 3N_H - 14 - 2d_s$

... $N_c = 3N_H + 60$	...	$N_c = 3N_H + 26$	$N_c = 3N_H + 24$	$N_c = 3N_H + 22$	$N_c = 3N_H + 20$	$N_c = 3N_H + 18$	$N_c = 3N_H + 16$	$N_c = 3N_H + 14$	$N_c = 3N_H + 12$	$N_c = 3N_H + 10$	$N_c = 3N_H + 8$		series $N_c(2N_H - 6 + N_{IC})$
											$C_{74}H_{22}$	...	$2N_H + 30$
										$C_{76}H_{22}$	$C_{80}H_{24}$	...	$2N_H + 32$
									$C_{78}H_{22}$	$C_{82}H_{24}$	$C_{86}H_{26}$	...	$2N_H + 34$
								$C_{80}H_{22}$	$C_{84}H_{24}$	$C_{88}H_{26}$	...		$2N_H + 38$
							$C_{88}H_{24}$	$C_{86}H_{24}$	$C_{90}H_{26}$	...			$2N_H + 40$
						$C_{90}H_{24}$	$C_{94}H_{26}$	$C_{92}H_{26}$	...				$2N_H + 42$
					$C_{92}H_{24}$	$C_{96}H_{26}$	$C_{100}H_{28}$	...					$2N_H + 44$
				$C_{94}H_{24}$	$C_{98}H_{26}$	$C_{102}H_{28}$	...						$2N_H + 46$
			$C_{96}H_{24}$	$C_{100}H_{26}$	$C_{104}H_{28}$	...							$2N_H + 48$
		$C_{104}H_{26}$	$C_{102}H_{26}$	$C_{106}H_{28}$	...								$2N_H + 50$
		$C_{110}H_{28}$	$C_{108}H_{28}$	...									$2N_H + 52$
		$C_{116}H_{30}$	$C_{114}H_{30}$										$2N_H + 54$
		...	$C_{120}H_{32}$										$2N_H + 56$
			$C_{126}H_{34}$										$2N_H + 58$
			$C_{132}H_{36}$										$2N_H + 60$
			...										
													...
$C_{150}H_{30}$													$2N_H + 90$
... ...													...

$N_c = 3N_H - 14 - 2d_s$

$N_c = 3N_H + 6$	$N_c = 3N_H + 4$	$N_c = 3N_H + 2$	$N_c = 3N_H$	$N_c = 3N_H - 2$	$N_c = 3N_H - 4$	$N_c = 3N_H - 6$	$N_c = 3N_H - 8$	$N_c = 3N_H - 10$	$N_c = 3N_H - 12$	$N_c = 3N_H - 14$	$N_c = 3N_H - 16$		series $N_c(2N_H - 6 + N_{Ic})$
										$C_{10}H_8$	$C_{14}H_{10}$	...	$2N_H - 6$
										$C_{16}H_{10}$	$C_{20}H_{12}$	...	$2N_H - 4$
										$C_{22}H_{12}$	$C_{26}H_{14}$	...	$2N_H - 2$
									$C_{24}H_{12}$	$C_{28}H_{14}$	$C_{32}H_{16}$	...	$2N_H$
									$C_{30}H_{14}$	$C_{34}H_{16}$	$C_{38}H_{18}$	...	$2N_H + 2$
								$C_{32}H_{14}$	$C_{36}H_{16}$	$C_{40}H_{18}$	$C_{44}H_{20}$	...	$2N_H + 4$
							$C_{40}H_{16}$	$C_{38}H_{16}$	$C_{42}H_{18}$	$C_{46}H_{20}$	$C_{50}H_{22}$	...	$2N_H + 6$
						$C_{42}H_{16}$	$C_{46}H_{18}$	$C_{44}H_{18}$	...	$C_{52}H_{22}$	$C_{56}H_{24}$	...	$2N_H + 8$
					$C_{50}H_{18}$	$C_{48}H_{18}$	$C_{52}H_{20}$	$C_{50}H_{20}$		$C_{58}H_{24}$	$C_{62}H_{26}$	...	$2N_H + 10$
				$C_{52}H_{18}$	$C_{56}H_{20}$	$C_{54}H_{20}$	...	...		$C_{64}H_{26}$	$C_{68}H_{28}$	...	$2N_H + 12$
			$C_{54}H_{18}$	$C_{58}H_{20}$	$C_{62}H_{22}$	...				$C_{70}H_{28}$	$C_{74}H_{30}$	...	$2N_H + 14$
		$C_{62}H_{20}$	$C_{60}H_{20}$	$C_{64}H_{22}$	...					$C_{76}H_{30}$	$C_{80}H_{32}$	...	$2N_H + 16$
	$C_{64}H_{20}$	$C_{68}H_{22}$	$C_{66}H_{22}$	$C_{70}H_{24}$						$C_{82}H_{32}$	$C_{86}H_{34}$	...	$2N_H + 18$
$C_{66}H_{20}$	$C_{70}H_{22}$	$C_{74}H_{24}$	$C_{72}H_{24}$	$C_{76}H_{26}$						$C_{88}H_{34}$	$C_{92}H_{36}$	...	$2N_H + 20$
$C_{72}H_{22}$	$C_{76}H_{24}$	...	$C_{78}H_{26}$	...						$C_{94}H_{36}$	$C_{98}H_{38}$	...	$2N_H + 22$
$C_{78}H_{24}$	...		$C_{84}H_{28}$							$C_{100}H_{38}$	$C_{104}H_{40}$	...	$2N_H + 24$
...			$C_{90}H_{30}$							$C_{106}H_{40}$	$C_{110}H_{42}$	...	$2N_H + 26$
			...							$C_{112}H_{42}$	$C_{116}H_{44}$	...	$2N_H + 28$
										...	...		$2N_H + 30$
													$2N_H + 32$
													$2N_H + 34$
													$2N_H + 36$
													$2N_H + 38$
													$2N_H + 40$

Table IV. Procedure for Determining Whether a Formula Can Be Represented by a PAH Composed Only of Nonradical Hexagonal Rings[a]

n	m	n	m	n	m	n	m	n	m	n	m	n	m
0	0	14	6	28	8	42	10	56	12	70	14	84	15
1	1	15	6	29	8	43	10	57	12	71	14	85	15
2	2	16	6	30	9	44	11	58	12	72	14	86	15
3	2	17	6	31	9	45	11	59	12	73	14	87	15
4	3	18	6	32	9	46	11	60	12	74	14	88	15
5	3	19	7	33	9	47	11	61	13	75	14	89	15
6	3	20	7	34	9	48	11	62	13	76	14	90	15
7	4	21	7	35	9	49	11	63	13	77	14	...	
8	4	22	7	36	9	50	11	64	13	78	14		
9	4	23	7	37	10	51	11	65	13	79	14		
10	5	24	8	38	10	52	12	66	13	80	15		
11	5	25	8	39	10	53	12	67	13	81	15		
12	5	26	8	40	10	54	12	68	13	82	15		
13	5	27	8	41	10	55	12	69	13	83	15		

[a] (1) If a formula possesses a C and H composition such that $N_c < 2N_H - 6$, then the formula cannot belong to a fused PAH6 containing exclusively hexagonal rings. If the formula passes this test go to step 2. (2) Using this general formula $C_{24+6m+2n}H_{12+2m}$, determine the values of m and n from the formula being tested. If m or n has a negative value the formula cannot be represented by a fused PAH6 structure. If m and n have positive values, go to step 3. (3) Using the following list find the value of n determined in step 2. If the corresponding m value in the following list is larger than the m value obtained in step 2, then the formula being tested cannot be represented by a fused PAH containing exclusively hexagonal rings.

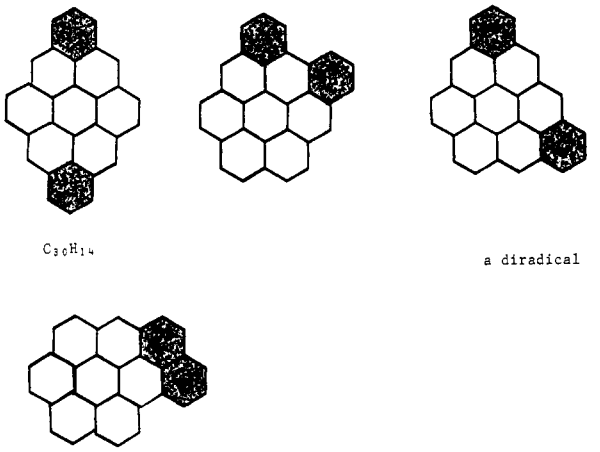
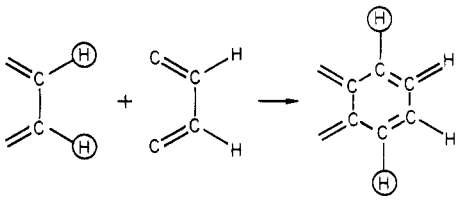


Figure 4. Isomers of $C_{30}H_{14}$ generated from coronene ($C_{24}H_{12}$) by attachment of two C═CH—C units or one C═CH—C═CH—C═C unit.

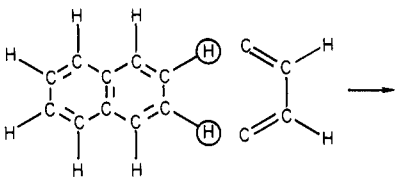
in the formula quadrant of Table III subtended by and including the $N_c = 2N_H$ row series and $N_c = 3N_H - 16$ column series. All the formulas expressed in Table II are applicable to this class of PAH6 with the modifications of 2 and 4 that are noted. In counting the number of rings possessed by a monocirculene, the ring associated with its "hole" is also included. In Figure 3, the number of internal hydrogens and related isomers are itemized by the respective numbers in parentheses under the representative structure. The monocirculene holes containing four internal hydrogens in Figure 3 correspond to excised PAH structures which have $N_{Pc} = 4$. The total number of monocirculene isomers associated with the formulas $C_{32}H_{16}$, $C_{36}H_{18}$, $C_{40}H_{20}$, $C_{44}H_{22}$, and $C_{48}H_{24}$ was determined to be 1, 3, 17, 68, and 322, respectively.

Polycirculenes have two or more holes and have formulas that first appear in the $N_c = 3N_H - 16$ column series. For example, the first dicirculene (has two holes corresponding to excised naphathenes) appears at $C_{56}H_{24}$ in the $N_c = 2N_H + 8$ row series, and the first tricirculene (has three holes corresponding to excised naphthalenes) appears at $C_{80}H_{32}$ in the $N_c = 2N_H + 16$ row series. In general, $N_{Pc} = N_H - 6 + 6c$ and $q_P = 2N_H - 6 + 6c$ where c equals the number of holes.

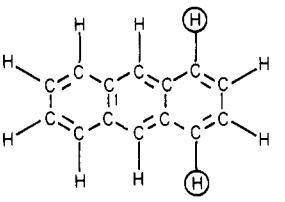
PAH Isomer Enumeration. Currently, efforts to computer enumerate all the structures associated with a particular PAH6 formula are under way.[4] Another approach could incorporate the following postulates which have been tested (but not proved) by a noncomputer method for sample formulas in Table V. Postulate 1: Tables III and IV provide the comprehensive formula framework comprising all possible fused PAH's possessing only hexagonal rings. Postulate 2: All possible isomers are recursively enumerated by adding C═CH—CH═C units to the immediately preceding row series homologue and two C═CH—C and C═CH—C═CH—C═C units to the immediately preceding column homologues. Since there are no column homologues above in any member of the $N_c = 2N_H - 6$ row series, it is easy to demonstrate that the following process can be applied to all isomers of the immediately preceding formula to produce all the possible isomers of the next formula. For example, naphthalene



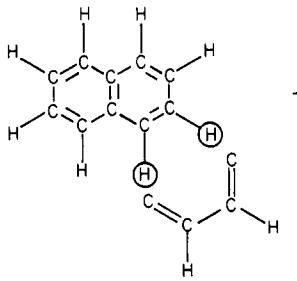
($C_{10}H_8$) gives both possible isomers, anthracene and phenanthrene, of $C_{14}H_{10}$ by this process. This procedure was used

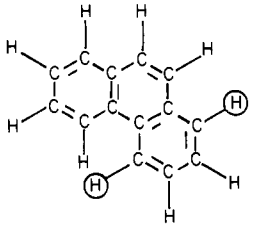


naphthalene ($C_{10}H_8$) + C_4H_2



anthracene ($C_{14}H_{10}$)

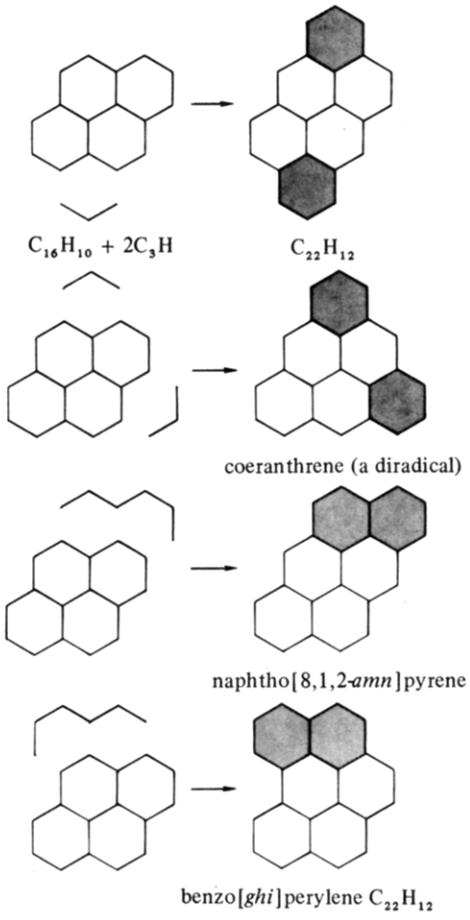




phenanthrene ($C_{14}H_{10}$)

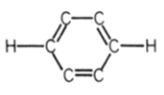
to generate all the cata-condensed (the $N_c = 2N_H - 6$ row series) isomers previously enumerated.[2] Additional straightforward examples of this procedure for recursively enumerating all possible isomers of a specific PAH6 formula is provided by the formulas $C_{16}H_{10}$, $C_{22}H_{12}$, and $C_{30}H_{14}$ which do not have any immediately preceding formula row homologues. Attachment of either two C—CH═C units or one C═CH—C═CH—C═C unit to naphthalene ($C_{10}H_{18}$) leads to only one isomer for the formula $C_{16}H_{10}$ which is pyrene. In this procedure attachment of different units to the same precursor structure or attachment of the same unit to different precursor structures often lead to repetition of a structure. Attachment of these same units to pyrene ($C_{16}H_{10}$) lead to

only two isomers corresponding to the formula of $C_{22}H_{12}$.

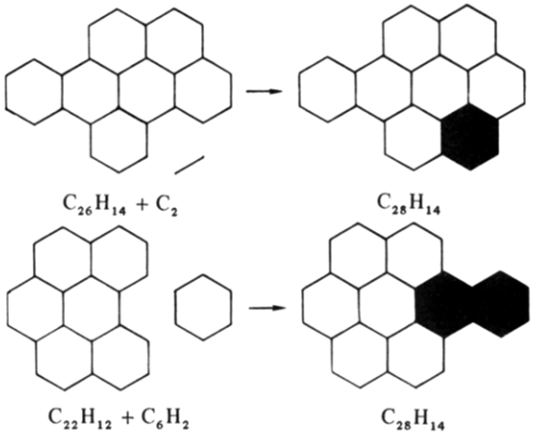


Note that generation of naphtho[8,1,2-*amn*]pyrene is repeated and that coeranthrene is excluded because it is a diradical structure. Application of this procedure to coronene-7 ($C_{24}H_{12}$) leads to the structures of the formula $C_{30}H_{14}$ summarized in Figure 4; excluding the diradical this gives three isomers.

More complicated examples of PAH6 isomer enumeration by this method are provided by the formulas $C_{26}H_{14}$ (nine isomers) and $C_{28}H_{14}$ (eight isomers). The various attachments of C=CH—CH=C units to the two isomers of $C_{22}H_{12}$ gives five different structures of formula $C_{26}H_{14}$, and the various attachments of two C—CH=C units or one C=CH—C=CH—C=C unit to the three different isomers of $C_{20}H_{12}$ gives eight different structures of formula $C_{26}H_{14}$. There are four common structures among the two sets; thus there are nine different isomers of the formula $C_{26}H_{14}$. The attachment of a C=CH—CH=C unit to coronene-7 ($C_{24}H_{12}$) gives only one structure of the formula $C_{28}H_{14}$, and the various attachments of two C—CH=C units or one C=CH—C=CH—C=C unit to the two isomers of $C_{22}H_{12}$ give seven different structures having the formula of $C_{28}H_{14}$. These eight different isomeric structures for $C_{28}H_{14}$ can be partially repeated by attachment of C=C units to the concave regions of the upper diagonally located isomers having the same number of hydrogens (14) or attachment of

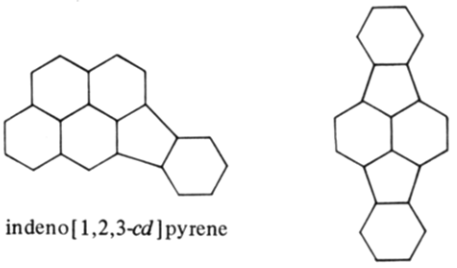


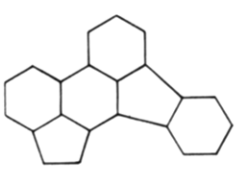
units to concave regions of the upper column isomers of $C_{22}H_{12}$, but these additional units produced no new structures in this example or any of the other test enumeration in Table V. This is illustrated by the following respective examples:



Although these examples do not provide unequivocal proof for the above postulates and the associated systematic procedure for enumerating the PAH6 isomers, they do, at the very least, provide an intuitive starting point for computer enumeration. Table V also summarizes all the graph-theoretical parameters discussed thus far for a limited number of formulas. The values in parentheses starting at the $N_c = 2N_H$ row series and the $N_c = 3N_H - 16$ column series are for the monocirculenes. For the monocirculenes in the $N_c = 2N_H$ row, the defined $N_{Ic} = 0$, but note that $N_{Ic} + N_{Pc} = p_3 = \text{constant}$. Also, the total number of rings (r) and edges (q) remain unaltered.

PAH Containing Pentagonal Rings. Fluoranthene ($C_{16}H_{10}$) is an isomer of pyrene having a pentagonal ring. Since the number of third degree vertexes ($p_3 = N_{Ic} + N_{Pc} = \text{constant}$) and the total number of edges (i.e., σ C–C bonds) must be constant, in going from pyrene to fluoranthene one six-membered ring becomes a five-membered ring while one internal carbon atom becomes a peripheral carbon with a corresponding increase in disconnections per $d_s = N_{Pc} - r$ and decrease in the number of internal edges per $q_I + 1 + d_s = N_{Pc} + N_{Ic} = \text{constant}$. Thus eq 2 must be modified to $N_{Pc} = N_H - 6 + r_5$ and eq 4 to $q_P = 2N_H - 6 + r_5$ where r_5 equals the number of five-membered rings possessed by the PAH. Since the cata-condensed PAH6 ($N_c = 2N_H - 6$ row series) have no internal third-degree carbon atoms, no PAH having five-membered rings can exist for the formulas having a C/H composition such that $N_c = 2N_H - 6$. In the $N_c = 2N_H - 2$ row series if up to four of the four internal third-degree carbon atoms can be converted to peripheral third-degree carbon atoms, then up to four six-membered rings can be converted to four five-membered rings. Benzo[*ghi*]perylene ($C_{22}H_{12}$) and its PAH6 isomer can be converted to the following structures possessing five-membered rings.



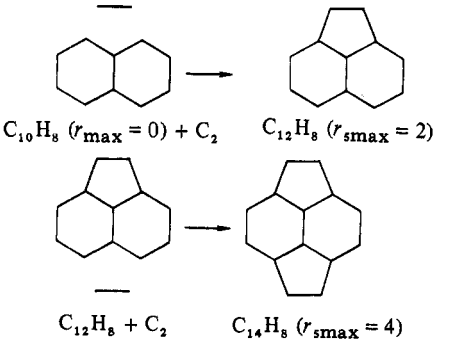


In general, if all of the internal third-degree carbons can be

Table V. Number of Isomers and Parameter Summary for Various PAH's

row series								
$N_c = 2N_H - 6$	$C_{10}H_8$	$C_{14}H_{10}$	$C_{18}H_{12}$	$C_{22}H_{14}$	$C_{26}H_{16}$	$C_{30}H_{18}$	$C_{34}H_{20}$	...
isomers	1	2	5	12	37	123	446	...
rings (r)	2	3	4	5	6	7	8	...
d_s	0	1	2	3	4	5	6	...
N_{Ic}	0	0	0	0	0	0	0	...
N_{Pc}	2	4	6	8	10	12	14	...
q_I	1	2	3	4	5	6	7	...
q	11	16	21	26	31	36	41	...
q_P	10	14	18	22	26	30	34	...
$N_c = 2N_H - 4$	$C_{16}H_{10}$	$C_{20}H_{12}$	$C_{24}H_{14}$	$C_{28}H_{16}$	$C_{32}H_{18}$	$C_{36}H_{20}$	$C_{40}H_{22}$	...
isomers	1	3	13	59				
rings (r)	4	5	6	7	8	9	10	...
d_s	0	1	2	3	4	5	6	...
N_{Ic}	2	2	2	2	2	2	2	...
N_{Pc}	4	6	8	10	12	14	16	...
q_I	5	6	7	8	9	10	11	...
q	19	24	29	34	39	44	49	...
q_p	14	18	22	26	30	34	38	...
$N_c = 2N_H - 2$	$C_{22}H_{12}$	$C_{26}H_{14}$	$C_{30}H_{16}$	$C_{34}H_{18}$	$C_{38}H_{20}$	$C_{42}H_{22}$	$C_{46}H_{24}$	...
isomers	2	9						
rings (r)	6	7	8	9	10	11	12	...
d_s	0	1	2	3	4	5	6	...
N_{Ic}	4	4	4	4	4	4	4	...
N_{Pc}	6	8	10	12	14	16	18	...
q_I	9	10	11	12	13	14	15	...
q	27	32	37	42	47	52	57	...
q_P	18	22	26	30	34	38	42	...
$N_c = 2N_H$	$C_{24}H_{12}$	$C_{28}H_{14}$	$C_{32}H_{16}$	$C_{36}H_{18}$	$C_{40}H_{20}$	$C_{44}H_{22}$	$C_{48}H_{24}$	$C_{52}H_{26}$	...
isomers	1	8						
rings (r)	7	8	9	10	11	12	13	14	...
d_s	−1	0	1 (7)	2 (8)	3 (9)	4 (10)	5 (11)	6 (12)	...
N_{Ic}	6	6	6 (0)	6 (0)	6 (0)	6 (0)	6 (0)	6 (0)	...
N_{Pc}	6	8	10 (16)	12 (18)	14 (20)	16 (22)	18 (24)	20 (26)	...
q_I	12	13	14 (8)	15 (9)	16 (10)	17 (11)	18 (12)	19 (13)	...
q	30	35	40	45	50	55	60	65	...
q_P	18	22	26 (32)	30 (36)	34 (40)	38 (44)	42 (48)	46 (52)	...
$N_c = 2N_H + 2$	$C_{30}H_{14}$	$C_{34}H_{16}$	$C_{38}H_{18}$	$C_{42}H_{20}$	$C_{46}H_{22}$	$C_{50}H_{24}$	$C_{54}H_{26}$	$C_{58}H_{28}$	...
isomers	3	29							
rings (r)	9	10	11	12	13	14	15	16	...
d_s	−1	0	1 (7)	2 (8)	3 (9)	4 (10)	5 (11)	6 (12)	...
N_{Ic}	8	8	8 (2)	8 (2)	8 (2)	8 (2)	8 (2)	8 (2)	...
N_{Pc}	8	10	12 (18)	14 (20)	16 (22)	18 (24)	20 (26)	22 (28)	...
q_I	16	17	18 (12)	19 (13)	20 (14)	21 (15)	22 (16)	23 (17)	...
q	38	43	48	53	58	63	68	73	...
q_P	22	26	30 (36)	34 (40)	38 (44)	42 (48)	46 (52)	50 (56)	...

converted to peripheral third-degree carbon atoms (i.e., $r_{5max} = N_{Ic}$ for a PAH6) with the corresponding simultaneous conversion of six-membered rings to five-membered rings, then the maximum number of five-membered rings (r_{5max}) that a corresponding PAH6 formula can possess is given by eq 11. Examination of the formulas along the diagonal of Table III provides one with additional insight concerning PAH's containing five-membered rings. Extension of the $N_c = 2N_H - 4$, $2N_H - 2$, $2N_H + 2$, $2N_H + 6$, and $2N_H + 12$ row series into the domain of forbidden graphs for PAH6 gives the formulas of $C_{12}H_8$, $C_{18}H_{10}$, $C_{26}H_{12}$, $C_{34}H_{14}$, and $C_{44}H_{16}$, respectively. The structures of these formulas can be derived by attachment of C=C units to the nonconcave regions of the diagonally located PAH6 having the same number of hydrogens and is illustrated with naphthalene ($C_{10}H_8$) below; structure c is derived from pyrene by this process. Thus all the structures corresponding to these formulas must have at least one pentagonal ring. Attachment of another C=C unit to the above formulas lead to $C_{14}H_8$, $C_{20}H_{10}$, $C_{28}H_{12}$, $C_{36}H_{14}$, and $C_{46}H_{16}$ each which have associated PAH structures with at least two pentagonal rings. Note that eq 11 still gives the maximum number of five-member rings possible for all the above formulas having at least one or two pentagonal rings.



CONCLUSION

This work has unified the formulas and related structures of fused PAH's into an overall systematic framework. Specifically, Table III places all the PAH6 formulas into row and column series in which the corresponding structures have the

same number of internal third-degree carbon atoms (N_{Ic}) and disconnections (d_s) of the internal bonds, respectively. Graph theoretical concepts of vertex degree (degree of carbon atom), edges (σ C–C bonds), and trees (formed by the internal edges of PAH) have been uniquely employed in this description of the structure-formula related properties of PAH's.

Graph-theoretically related formulas have been derived or empirically noted for the first time for PAH's. From the formula of a PAH, the number of rings ($r = (1/2)[N_c + 2 - N_H]$) and σ C–C bonds ($q = (1/2)[3N_c - N_H]$) can be determined. Also, the range of possible peripheral and internal carbon atoms (N_{Pc} and N_{Ic}), peripheral and internal σ bonds (q_p and q_I), and the number of five-membered rings (r_5) that are associated with the various related structures of a specific PAH formula can be directly determined.

Finally, a method for determining whether a specific C/H composition can be represented by a PAH6 structure has been presented. A systematic approach for the computer enumeration of all the structures of PAH6 has been proposed. Although, eq 2, 4, and 5 are presented without proofs (in the mathematical sense), it is believed that they will have general application in mathematical graph theory, and additional results emanating from their utility will be forthcoming.

ACKNOWLEDGMENT

I am grateful to the Fulbright Commission and the University of Missouri for providing me with a one-semester sabbatical leave to the University of Ljubljana in Yugoslavia where this work was culminated.

REFERENCES AND NOTES

(1) Balaban, A. T. *Rev. Roum. Chim.* **1966**, *11*, 1097.
(2) Balaban, A. T; Harary, F. *Tetrahedron* **1968**, *24*, 2505.
(3) Balaban, A. T. "Chemical Application of Graph Theory"; Academic Press: New York, 1976; Chapter 5.
(4) Trinajstić, N., work in progress.
(5) (a) Harary, F. "A Seminar on Graph Theory"; Holt, Rinehart, and Winston: San Francisco, CA, 1967. (b) "Graph Theory"; Addison-Wesley Publishing Co.: Reading, MA, 1969.
(6) Jenny, W.; Peter, R. *Angew. Chem.* **1965**, *77*, 1027.
(7) Flammang-Barbieux, M.; Nasielski, J.; Martin, R. H. *Tetrahedron Lett.* **1967**, 743.

The Chemical Abstracts Service Chemical Registry System. 9. Input Structure Conventions

ANNE W. RYAN* and ROBERT E. STOBAUGH
Chemical Abstracts Service, Columbus, Ohio 43210

Received September 8, 1981

The Chemical Abstracts Service (CAS) Chemical Registry System is a computer-based system that uniquely identifies chemical substances on the basis of composition and structure. The structure conventions for substances registered in the Chemical Registry System have been described. The types of substances are fully defined organic and inorganic compounds, ions, polymers, coordination compounds, alloys, mixtures, and certain partially defined substances.

INTRODUCTION

Chemical Abstracts Service, in its mission of covering the world's primary chemical and chemical engineering literature, has from the beginning been involved with documenting and indexing information about chemical substances. A particular problem has been the identification of reoccurrences of substances in scientific literature. Such identification is necessary to assure consistency in the indexes to *Chemical Abstracts*. These indexes are based on a systematic chemical nomenclature, and the technique used for many years to assure consistency was to name each indexed substance every time it was cited and to file the name alphabetically in an index with its citations. This resulted in a large amount of redundant effort in repeated renamings. To remedy this situation, the CAS Chemical Registry System was developed in the early 1960s. The main objective was the establishment of a computer-based system to identify chemical substances uniquely on the basis of molecular structure. The chemical substances involved comprise the whole of chemistry in that all classes are handled—organic and inorganic, completely and partly described. The design, content, function, and some special features of the Chemical Registry System, including both machine and manual registration processes, have been described in detail in previous papers.[1-8] In the development of the Registry System, the structural diagram rather than nomenclature was chosen as the means for recording substances because in actual usage names vary greatly in consistency, standardization, systematization and the language involved. On the other hand, structural representation is standardized to such an extent that structural diagrams are virtually the same on a worldwide basis, even when different languages and alphabets are involved. The techniques and mechanics of typed structure input have been described,[5] and in previous Registry papers[1-8] structure conventions have been alluded to and illustrated. However, these conventions have not been treated specifically, and it is the purpose of this paper to do so.

Since the Chemical Structure File has become available as the basis for the new service CAS ONLINE, the knowledge of how structures are represented for input will enable users to phrase queries in a way to take maximum advantage of particular structural characteristics. While the majority of use of CAS ONLINE can be effective without knowledge of many of the details described here, more retrieval selectivity is possible with more intellectual specificity. Also, in order for users to take full advantage of the CAS Private Registry Service, structural information should be represented in conformance with conventions established for the Chemical Registry System. General structuring conventions will be presented first, followed by those for specialized classes.

GENERAL STRUCTURING CONVENTIONS

The conventions used in preparing structural diagrams, especially of organic compounds, are based on those exemplified in chemical literature. The primary aim has been to allow as much flexibility as possible, keeping the number of arbitrary conventions to a minimum. In cases where more than one method of representation is in current use, a single method has been chosen, for example, for salts, polymers, coordination compounds, and incompletely defined substances. Where it has been possible to recognize certain variations in representation within the registration programs, it has been done, especially for alternating single–double bond and tautomeric

0095-2338/82/1622-0022$01.25/0