A Procedure for Characterization of the Rings of a Molecule

MILAN RANDIĆ*† and CHARLES L. WILKINS*

Department of Chemistry, University of Nebraska—Lincoln, Lincoln, Nebraska 68588

A procedure suited for characterization of rings in a molecule and easily implemented by computer programs that can consider general structure is proposed. The procedure consists of using atomic codes based on enumeration of paths of different length within the molecular skeleton and their subsequent summation for the atoms forming the rings. The proposed characterization provides an illustration of the advantages of treating path sequences rather formally as mathematical entities, to which standard mathematical operations are applicable. Illustrations give a variety of ring codes (sequences) and show how these can be used to recognize equivalent rings in complex polycyclic structures. The procedure can be extended to other kinds of fragments. One such application is illustrated with polyadamantanes. The proposed ring codes need not necessarily be used in the form presented here. Several modifications which may be of interest are indicated.

INTRODUCTION

The importance of recognition of rings in large polycyclic structures is well known and the problem has received considerable attention in both the mathematical and chemical literature. One can delineate three levels of this problem as follows: (1) enumeration of rings of different size; (2) recognition (identification) of all (or selected) rings in a structure; and (3) characterization of all (or selected) rings in a structure. These problems can be ordered in a hierarchical way. For the recognition of rings, it helps if one knows their number. For characterization, it helps if the rings have already been identified. The focus of the present work is the problem of characterization of rings, the subject that has been the least covered in the literature.

The problem of enumeration of rings, it appears, has been of more interest to mathematicians than to chemists. One can appreciate the importance of the problem since it is related to the question of existence of rings of various sizes. For instance, if a graph is to be classified as pancyclic it must possess cycles of all sizes possible. The famous problem of the Hamiltonian circuit requires the existence of a circuit of the largest size (i.e., involving all vertices). The problem of enumeration of rings has been considered by Sachs, among others, who outlined a procedure whereby one can derive the number of circuits of all sizes (except the largest!) by considering the characteristic polynomial of the graph and some derived auxiliary functions.[1] For example, in the case of the Petersen's graph (1), one obtains the following enumerations:

Petersen's Graph	Ring size					
	5	6	7	8	9	10
Number of rings	12	10	–	15	20	?

1
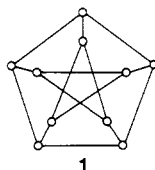
There are no rings of size 10 in the Petersen's graph (i.e., there is no Hamiltonian circuit), but the answer is deliberately indicated with a question mark to emphasize that the mathematical scheme of Sachs (or any other at the present time) cannot answer the question simply by studying selected graph invariants.

The problem of recognition of rings has received considerable attention and here several of the approaches will be mentioned briefly. The problem is of interest for a number

of potential applications such as structure–property correlations, deduction of strategies for synthesis planning, derivation of codes for chemical compounds, and the development of systematic chemical names. For each of these applications, different concepts have developed. In the study of the additivity of molecular resonance energies conjugated circuits have been found crucial.[2] These are circuits in which there is a regular alternation of single and double CC bonds as prescribed by the individual Kekulé valence structures. Hamiltonian circuits emerged as a useful basis for development of the fundamental forms for topological mapping of organic molecules and the subsequent derivation of notation.[3] Such an approach enables a computer program to generate hypotheses about organic molecular structure in an algorithmic and exhaustive way. Chemical information systems tend to use the concept of the smallest set of smallest rings,[4] which in turn can be found once the fundamental set of rings[5] is known. A fundamental set contains all atoms and all bonds and the minimum number of rings, the number being given by the number of cuts[6] necessary to produce a coherent (connected) acyclic structure (a spanning tree). These topics have been covered in the literature, and various algorithms described,[7] compared,[8] and discussed.[9] Finding rings of interest can be ambiguous even for relatively small systems. Conrow[10] quoted dozens of cases of tricyclic and tetracyclic compounds, the names of which have been incorrectly assigned as a consequence of not identifying the relevant naming rings. Meinwald and Crandall[11] pointed out that even the tricyclopentane (2)

2
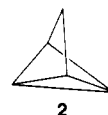
(after tetrahedrane the smallest possible tricyclic system!), for which several synthetic routes have been reported, has been consistently misnamed; and up to the time of their paper, the correct IUPAC name tricyclo[2.1.0.0²,⁵]pentane had never been used. It may be that this is also a sign of the unattractiveness of the IUPAC naming system, which is far from optimal.[12]

Characterization of rings appears to be a problem which has evaded efforts of past investigators, since it is hardly possible that the importance of the problem has not been generally recognized. This paper presents a new method for a characterization of ring systems, but we should indicate clearly that our prime aim is confined to the underlying molecular graphs, not actual molecules. The ultimate utility of this approach in actual applications to molecules has yet to be demonstrated. However, although the utility is conjectural, it seems to us that interested users should have no difficulties

† On leave from Ames Laboratory, DOE, Iowa State University, Ames, Iowa 50011.
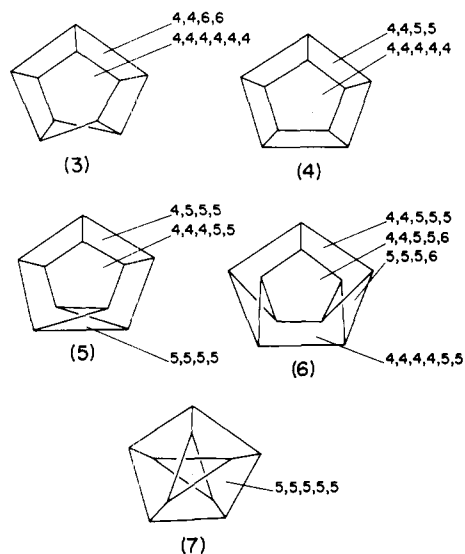
CHARACTERIZATION OF THE RINGS OF A MOLECULE

J. Chem. Inf. Comput. Sci., Vol. 20, No. 1, 1980 **37**

Figure 1. Selected trivalent graphs and ad hoc constructed ring codes signifying the sizes of adjacent rings.

in combining the essentials of their problems and their experience with available coding schemes for real molecules in developing necessary details for their specific use. By characterization here is meant a nonconventional naming of rings based on structural information alone and providing useful discrimination among rings. However, this may not necessarily be a unique characterization. The difficulty of the problem mainly resides in the implied requirement for the *generality* of the solution. For special classes of molecules, it is not difficult to develop ad hoc schemes which will suit a particular purpose. One such example is the ring indexing scheme derived for conjugated circuits and subsequently used for discussion of the relative magnitudes of the ring currents.[2] For example, one may consider polycyclic systems and assign to each ring a sequence of numbers (in an ascending order) which give the sizes of adjacent rings. The results for the five trivalent graphs with 10 vertices[13] are summarized in Figure 1. When dealing with a class of structures of particularly regular forms, such as benzenoid hydrocarbons, one can replace the initial graphs by simpler ones in which a vertex represents a benzene hexagon and a line represents the junction of two hexagons through a common edge. Such simplified pictorial representations of benzenoid hydrocarbons have been used, for example, by Smith[14] for tabulation of specific potentials of conjugated hydrocarbons in a condensed and very clear manner. Balaban and Harary[15] were able to derive *enumerations* of catacondensed benzenoid systems from the analysis of such contracted graphs, which also found application in other graph theoretical descriptions of this particular class of compounds.[16] In such cases, one can simply use atom codes (i.e., the sequences of atom path numbers) to characterize the rings, or any alternative atom characterizations, since the problem has been converted to that of "atom" characterization. Although such possibilities are of considerable interest in particular applications, they clearly cannot apply to the general case.

Ad hoc schemes have their advantages and are certainly legitimate and even desirable. Such methods are, by their nature, unrelated to other available schemes; hence they are likely to lead to novel correlations and can reveal unsuspected connections. A case in point is provided by the concept of conjugated circuits,[2] which has been found not only to lead to circuit additive expressions for molecular resonance energies, but has also led to graph theoretical classifications of nonalternate conjugated hydrocarbons (azulenoid vs. nonazulenoid). Additionally, the circuits considered which suggested

the concept of isoconjugate systems, have clarified the ambiguous and evasive definition of aromaticity and suggested the concept of ring indexes (for the class of conjugated compounds) with the possibility of extending the approach to heteroconjugated systems and systems involving phase changes in their orbital pictures as defined by a Mobius strip.[17] While all these are of considerable interest in the area of organic chemistry of conjugated systems, their limited applicability to other structures is obvious. Structural features of conjugated systems, in particular those requiring several Kekulé valence type structures for their description, are rather unique and special. Connectivity as the fundamental molecular quality is general; hence, it appears that a general approach to ring characterization has to be based on the connectivities of the atoms involved.

OUTLINE OF THE APPROACH

The basis of the present approach is atom codes or atom sequences derived by enumeration of paths of different length in a molecular skeleton. The concept of atom path sequences and their enumeration has been already described.[18] The atomic sequences have been found useful in searches for fragments in a larger structure[19] and ordering of selected atomic data.[20] By *adding* atomic contributions, one derives molecular path sequences. A number of applications of molecular path sequences have been reported. These include ordering of molecules to reveal trends in many standard thermodynamic molecular properties,[21] comparison of molecular structures, and quantification of the degree of structural similarity (or dissimilarity) of molecules.[22] Possible applications of the quantitative similarity test are many and have been shown to be revealing. One finds that among hypothetical structures the naturally occurring skeletal forms show a higher degree of mutual similarity.[22] Furthermore, for structure–activity correlations, one can suggest hierarchical ordering among candidate structures with respect to their potential biomedical activity.[23] This latter technique may serve as a rational alternative to the random searches often employed in drug design and testing efforts.

The notion of the shell of neighbors appears in many applications. It has been used by Lederberg for his dendritic scheme,[24] by Penny for linearization of molecular structure,[25] and by Gordon and Kennedy[26] for ordering and naming acyclic hydrocarbons. It is implied in the Morgan algorithm for numbering of atoms,[27] and has been incorporated in various computer-oriented representations of molecular structure and the study of molecular transformations.[28] The notion of immediate atomic environment has been studied extensively by Dubois and co-workers[29] who developed a rather general scheme, DARC (an acronym for Documentation and Automatization of Researches for Correlations), based on the concept of the chromatic graph. DARC and the auxiliary subschemes are concerned with ordering of an atomic environment and the use of such derived descriptions for specific and definite goals (e.g., the derivation of structure–property correlations). The approach has been recently extended to the study of reaction pathways.[30] Because of both the similarities and the differences from the scheme of Dubois, it is worthwhile to suggest a label for the scheme presented here. Our emphasis is on Rigorous Algebraic Nomenclature Descriptions (RAND) as attributes and, as will shortly be seen, such an orientation and philosophy immediately lead to a general concept of ring codes.

Viewing atom codes algebraically upgrades the nomenclature from a passive description to an operational form. Here, only additions of sequences will be used as operations. Depending on the problem, other mathematical and logistic operations may become of interest. To illustrate the point, note

Table I. Atom Path Sequences or Codes

atom		initial codes	atom	rearranged lexicographic ordering
1	*	1,2,1,2,1	6	1,1,2,1,2
2	*	3,1,2,1	1	1,2,1,2,1
3		2,4,1	7	1,2,1,2,1
4		3,2,2	8	1,2,2,2
5		2,2,1,2	5	2,2,1,2
6	*	1,1,2,1,2	3	2,4,1
7	*	1,2,1,2,1	2	3,1,2,1
8		1,2,2,2	4	3,2,2

that a "scalar product" of a molecular sequence for an alkane with the sequence (1, 2, 3, 4, 5, . . .) will give the total number of paths in the structure, one of the parameters used by Wiener in correlations of thermodynamic properties of alkanes and related structures.[18] Similarly, shifting of path sequences is an operation used in reconstruction of a graph from the list of atomic codes.[18] There are likely to be many additional as yet unused or unanticipated operations that would correspond to meaningful alterations of atomic or molecular path codes.

Atom Codes. In order to illustrate the concept of atom codes, from which later ring codes are to be constructed, consider the simple case of 2,4-dimethylhexane

3
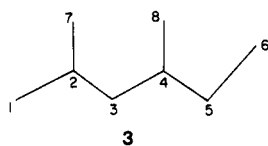
Atom path sequences or codes are listed in Table I for each atom (left) and also as an ordered list. Each sequence indicates for a given atom the number of paths of length one, two, three, and so on. There are several important properties of such atomic sequences (which will also be called atomic codes). These properties have been discussed at some length elsewhere. It, therefore, suffices to list them briefly (with the understanding that molecular graphs of alkanes are implied when acyclic structures are mentioned):

(a) In acyclic alkane structures, atomic paths are equivalent to the atomic neighbor count at the same distance.

(b) It has been conjectured that the list of atom codes uniquely represents a structure. (This has been verified for all acyclic hydrocarbon structures up to 14 vertices, with the maximum valency restricted to four.[31])

(c) There are unique atomic codes which represent molecular connectivity (These have been indicated by an asterisk *.) Unique codes are typified by 1, 2, 1, 2, 1 which appears only in 2,4-dimethylhexane. Other codes, not marked with asterisks, such as 2, 2, 1, 2 in the example, can appear in different molecular graphs.[31] Furthermore, there are certain molecules whose molecular graphs yield no single unique atom code.

(d) The legitimacy of the list of atom codes is an open problem. The problem is that of asserting that the list is genuine and corresponds to a structure. Some necessary conditions have been recognized.[18,32]

(e) Equivalent atoms necessarily have identical codes. The converse is not always true. Hence, the codes alone do not suffice to represent the symmetry of the structures.

(f) Codes allow structure ordering, such as the lexicographic ordering. (Observe that such an ordered list does not need to include atomic labels.)

(g) Reconstruction consists (for hydrocarbons) in establishing adjacency relationships among the codes (i.e., which code is adjacent to which other). An algorithm for reconstruction of acyclic graphs has been discussed.[18] We should note here that the discussion is confined to graphs rather than
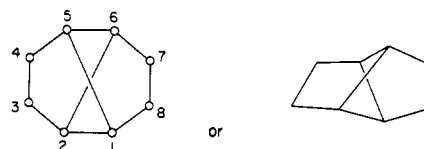

Figure 2. Molecular diagram of tricyclo[3.3.0.0^{2,6}]octane with the numbering of atoms.

molecules. Reconstruction of molecular forms obviously requires information on the identity and location of heteroatoms.

(h) Codes can be truncated. Truncated codes are of interest in searching for regularities in trends in molecular data.

In summary, we would like to reiterate the main points of our approach: (1) formal treatment of sequences of path numbers as operational entities, not merely as notation; (2) construction of additional graph theoretical invariants from atomic codes (such as molecular, fragment, and ring codes); and (3) use of derived graph theoretical quantities in the study of molecular properties.

The last point indicates the primary goal, which, in turn, implies that only nonstructural attributes will be considered. This allows a degree of freedom which applications in many chemical information areas can further exploit by combining some elements of the present analysis with available nomenclature and notation systems.

Ring Codes. The concept of ring codes, or the sequence of ring path numbers, is a natural extension of the concept of atomic codes. Analogously to the construction of molecular codes by summing the contributions of all atoms in a molecule in a single sequence, so ring codes can be derived by summing all contributions of atoms involved in the ring considered. There are some subtle differences which will be discussed shortly. One can view molecular codes, except for the lack of normalization factors, as *average* atomic codes.

Consider the example of 2,4-dimethylhexane. By summing columns of rows 1–8 in the example (using either the initial or ordered codes), so that the sum of the first entries of all atomic sequences gives the first molecular path entry, the sum of the second entries of atomic path sequences gives the second molecular path entry, and so on, we obtain the following sequence: 14, 16, 12, 10, 4. From this we can produce by suitable normalization the following two sequences:

a   7, 8, 6, 5, 2

b   1.75, 2.00, 1.50, 1.25, 0.50

Case a, derived by dividing by 2, lists the sequence of paths of different length in a molecule, provides the molecular codes used in ordering of structures,[21] and facilitates the discussion of molecular skeletal similarity.[22] One has to divide the sum of the corresponding atomic contributions by two, since each path has been counted twice, once for each end atom. Case b follows from the initial sequence by dividing by 8 in this example, which is the number of carbon atoms which contribute to path count. Hence the new sequence represents the *mean* or *average* atomic path sequence. Each entry in the sequence b is the arithmetic mean of the corresponding entries for all carbon atoms in the skeleton. The sequences a and b are trivially related to the un-normalized initial sequence, so any of the three forms can be used, leading to the same conclusions. However, case b involves fractions and is possibly less convenient.

Similarly, the ring codes can be viewed, except for the normalization, as *average* ring-atom codes. In order to illustrate the concept of ring codes, consider tricyclo-[3.3.0.0^{2,6}]octane (Figure 2). First we must derive atom codes. One quickly finds that manual enumeration of paths in polycyclic structures is both tedious and error prone, so that even with molecules as small as tricyclooctanes, the use of computers

CHARACTERIZATION OF THE RINGS OF A MOLECULE

J. Chem. Inf. Comput. Sci., Vol. 20, No. 1, 1980 **39**

Table II. Input and Output of ALLPATH Program for Enumeration of Paths in Tricyclo[3.3.0.0²·⁶]octane[a]

CONNECTION MATRIX

0	1	0	0	1	0	0	1
1	0	1	0	0	1	0	0
0	1	0	1	0	0	0	0
0	0	1	0	1	0	0	0
1	0	0	1	0	1	0	0
0	1	0	0	1	0	1	0
0	0	0	0	0	1	0	1
1	0	0	0	0	0	1	0

1	3	5	7	8	6	4	4
2							
1	3	5	7	8	6	4	4
3							
1	2	3	6	10	8	8	6
4							
1	2	3	6	10	8	8	6
5							
1	3	5	7	8	6	4	4
6							
1	3	5	7	8	6	4	4
7							
1	2	3	6	10	8	8	6
8							
1	2	3	6	10	8	8	6
8	10	16	26	36	28	24	20

[a] The last sequence of path numbers gives the number of paths for a molecule as a whole.

is not only convenient, but essential. In Table II are listed the atomic path numbers produced by the ALLPATH computer program,[33] together with the input information. The connectivity matrix is also given as a check on the input, while the first column in the path listings gives the atom label and an entry "1" (which one can formally view as the number of paths of length zero). If a graph had loops, the first entry would change and, thus, differentiate such atoms (vertices), a flexibility which may be of interest when considering heteroatomic structures. The remaining columns give the number of paths of length one, two, three, and so on. The last row, except for the first entry, is half of the sum of all atomic sequences. The factor 1/2 is introduced, since each path has, in fact, as previously mentioned, been counted twice in the summation, once for each end atom. The total number of paths is 160 + 8 (eight zero-length paths or atoms; this row is the *molecular path code*).

Ring codes are derived directly from the data in Table II in the same fashion but restricting the summation only to atoms forming a ring. Thus:

ring atoms	ring size	ring path sequences
1,2,5,6	4	12,20,28,32,24,16,16
1,2,3,4,5	5	13,21,33,44,34,28,24

There are only two nonequivalent rings in this case if the eight-membered ring which involves all the carbon atoms and whose code becomes identical with the molecular code (except for the factor 2) is not considered. One may prefer to have the sequence numbers give the number of paths of different length, as is the case with the molecular codes. However, the extraction of paths, that is, verifying which paths have been counted twice and which only once (because the end atom does not belong to the ring considered), is considerably more complicated and does not appear warranted at this time. Formally, dividing the path numbers by two, regardless of the character of individual paths, would introduce fractions. Some of the viable alternatives to this procedure will be discussed later,

but first it is appropriate to examine the proposed ring indexes in this most simple version.

One observes that the *collection* of atomic codes for all atoms in a ring contains more information than the *summation* of paths of the same length for atoms in a ring. So one may question this shift from the available collection of ring-atom codes to a single ring code which will be accompanied with some loss of important details concerning the path lengths for each vertex. This will lead to reduced discrimination of nonequivalent rings. On the other hand, it is convenient to have a *single* ring characterization rather than N times as much data for each of the N atoms in a ring, in analogy with a similar dilemma for characterization of a molecule by a single path number sequence instead of more the complete data supplied by the collection of path numbers for all atoms in a molecule. One may view the situation as having an option between the simpler alternative which will occasionally fail to discriminate fully the nonequivalent ring environments and a more complete listing reducing such relatively rare instances to very remote possibilities. One can always in situations of limited discriminatory power invoke the full list of atom codes for ring atoms, should such analysis be warranted. The applications of molecular codes indicated earlier illustrated the utility of the concept, and one should judge the idea of ring codes accordingly, that is, on the basis of the success of their use in applications. Ring codes clearly belong to molecular fragments. However, although there is an apparent need for ring codes or indexes, particularly in searching files or comparing structures, there appears to be limited experimental data relevant for making correlations and other uses of ring indexes. In view of that fact, the use of indexes will probably continue, in the near future, to be limited to chemical documentation and computer processing of structural information and structures (such as in computer-assisted synthesis, generation of hypothetical structures, etc.).

The proposed scheme is conceptually simple. In practice, the enumeration of paths in a graph is of a nonpolynomial character and, hence, imposes some limitations on the size and complexity of graphs to which such a scheme can be applied without additional restrictions. We find that enumeration of paths is practical (with current computer technology) when the total number of paths is about 10^6 or less. This limit encompasses very many graphs of interest in chemistry. In Table III, path numbers are listed for selected complex graphs (shown in Figure 3), in order to indicate the size and the complexity of some graphs which are accessible to the analysis.

It is of interest to examine more closely the derived ring codes of tricyclo[3.3.0.0²·⁶]octane. An isolated four-membered ring has the code 8, 8, 8. By subtracting this from the code found for the four-membered ring in the subject compound, a new code is obtained.

$$
\begin{array}{r}
12,\ 20,\ 28,\ 32,\ 24,\ 16,\ 16 \\
-\quad 8,\ \ 8,\ \ 8,\ \ 0,\ \ 0,\ \ 0,\ \ 0 \\
\hline
4,\ 12,\ 20,\ 32,\ 24,\ 16,\ 16
\end{array}
$$

The simple arithmetic is deliberately shown in order to once again emphasize that the sequences can be viewed formally as mathematical quantities on which selected operations are well defined. The first entry in the new sequence indicates the number of *exocyclic* bonds of the four-membered ring considered. Those bonds can now be identified and a fragment found
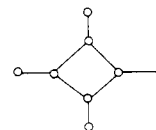
15

Table III. A Selection of Rather Complex Graphs Showing the Total Number of Paths (P) and Path Sequences for a Single Vertex[a]

Molecule	Number of vertices(n)				Total number of paths (P)	
	Atom		Paths of increasing length			
Desargues-Levi graph (10)		n = 20			P = 142290	
	3	6	12	24	84	156
264	456	672	1032	1332	1860	1992
2220	1752	1488	600	228		
Cube-in-cube (11)		n = 16			P = 1450816	
	4	12	36	96	264	624	1536
3072	6240	9840	15912	17472	19200	5712
Dodecahedron (12) n = 20				P = 250740		
	3	6	12	24	42	78	144
240	408	654	936	1272	1626	1818
1806	1614	1140	552	162		
Balaban's double star graph (13)		n = 15			P = 658680	
	4	12	32	88	232	560	1224
2456	4480	7200	9600	9656	6352	2016
Prism-in-prism (14)		n = 12			P = 83352	
	4	12	34	88	212	464	880
1364	1704	1496	688			

[a] For all graphs all vertices are equivalent and have identical path sequences.
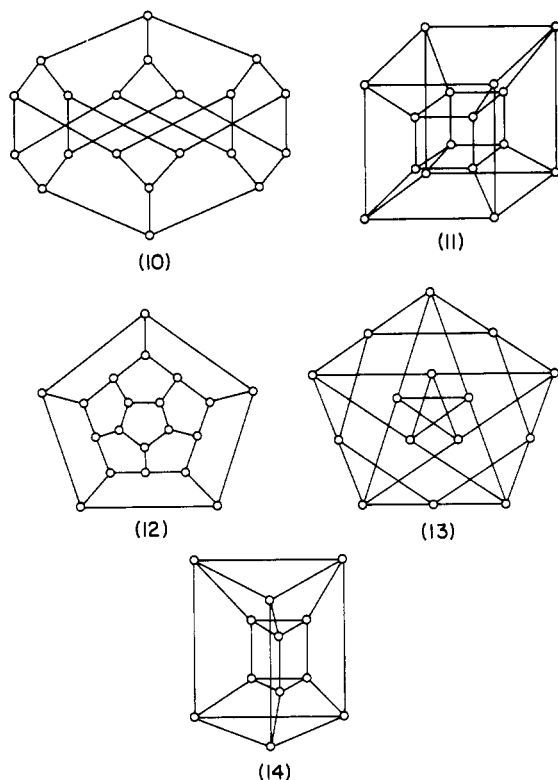

(10)   (11)

(12)   (13)

(14)

Figure 3. A selection of rather complex graphs having from 10^4 to over 10^6 paths illustrating cases which may approach the limit of practicality in enumeration of paths. (For comparison, the cholesterol and morphine structures have only a few thousand paths.)

The fragment extends to the environment of the considered ring at *one* bond length distance. The path enumeration for the ring atoms in the fragment immediately gives 12, 16, 16, 8. The previous step can now be repeated. That is, subtract

the new code from the original code in order to recognize the environment of the ring at a greater distance.

$$
\begin{array}{r}
12, 20, 28, 32, 24, 16, 16 \\
- \ 12, 16, 16, \ \ 8 \\
\hline
0, \ \ 4, 12, 24, 24, 16, 16
\end{array}
$$

The first nonzero entry now indicates the number of paths of length two, exocyclic to the four-membered ring.


16
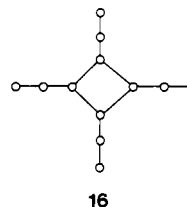
Continuing such analysis, one can expand the radius of the neighborhood of the ring under examination. Obviously the new sequence generated at each step, starting with 4, 4, . . ., can equally well be used as an alternative ring characterization. It can be directly computed by appropriate modification of the existing program, one possibility being that of erasing ring bonds in the structure and counting paths for the residual. However, at this stage, the use of the original sequences is preferred primarily since they allow other sequences to be generated. Of course, the derived new sequences may lack some of the initial information about the structure. For instance, if the four-membered ring is erased, the residual cannot reveal whether the initial ring was four membered or, say, five membered, with an additional atom inserted between atoms 1 and 2 and having no other neighbors.

Illustrations. A number of tricyclic structures, several benzenoid skeletons, a collection of special ring structures, and a few polyadamantanes have been selected in order to illustrate the kind and the type of ring codes that one derives using the present analysis. Obviously, polycyclic structures offer an immense number of possibilities, and one can consider but a fraction of the potentially interesting molecules. However, the examples chosen will clearly indicate that the approach is *general*, that the resulting discrimination is sufficiently large to be of practical interest, and that the derived codes reflect salient structural features. Thus, the method can be expected to be of interest in retrieval work, coding of structures, line notation, and other nomenclature related applications, as well as in correlations with physical and chemical properties if and when these may become available.

Tricyclic Systems. In Figure 4 is a selection of eight- and nine-carbon tricyclic systems having rings of size four and larger. Table IV lists ring sequences found in these compounds arranged according to the size of the ring. The individual faces are indicated by letters, the rings being the periphery of such faces. Label A + B, for example, represents the ring obtained by fusion of rings A and B, which excludes their common part. Only symmetrically nonequivalent rings are included in the table. Since the examples considered represent rather simple tricyclic structures, the symmetries have been recognized by inspection of the molecular graphs. In the few instances where equivalence of rings is not readily apparent, such rings are indicated by the same letter. Equivalent rings necessarily have identical codes, but the converse need not be true. Hence, the ring codes can serve as guidance to the identification of equivalent rings, primarily by excluding a large number of nonequivalent pairs. But, for a rigorous establishment of symmetries, an independent approach has to be used. We find the scheme based on the concept of the smallest binary code[34] to be such an independent useful method.

There are several properties of the ring codes that are immediately obvious. First, the leading terms generally increase

CHARACTERIZATION OF THE RINGS OF A MOLECULE

J. Chem. Inf. Comput. Sci., Vol. 20, No. 1, 1980 **41**
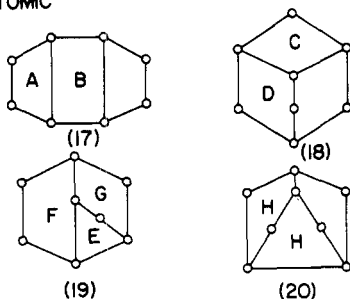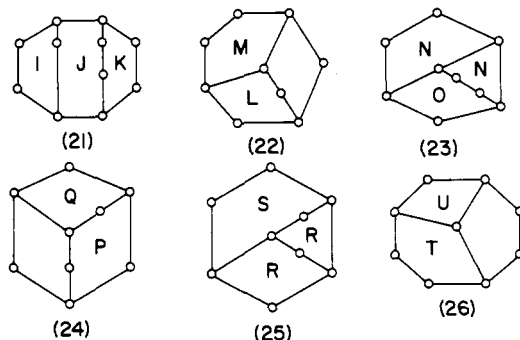
EIGHT ATOMIC

NINE ATOMIC

Figure 4. The eight atomic and nine carbon tricyclic structures examined. The corresponding ring codes are listed in Table IV.

Table IV. Ring Codes for the Selection of Tricyclic Octanes and Tricyclic Nonanes Shown in Figure 4

Four membered rings		Six membered rings	
I	10, 16, 24, 26, 34, 34, 22, 8	P	15, 22, 32, 48, 58, 56, 40, 8
B	10, 16, 26, 28, 32, 20, 14	S	15, 22, 32, 48, 62, 49, 32, 11
C	11, 17, 26, 30, 33, 32, 6	O	15, 23, 35, 51, 54, 43, 35, 21
U	11, 18, 26, 28, 30, 32, 22, 12	C+2D	15, 23, 36, 50, 58, 44, 14
E	11, 18, 27, 32, 20, 25, 14	E+F+G	15, 23, 36, 53, 51, 38, 21
A	12, 20, 28, 24, 24, 12, 12	J+K	16, 24, 34, 46, 48, 40, 26, 10
		2R	16, 24, 34, 50, 58, 44, 30, 16
		2H	16, 24, 36, 46, 42, 36, 16
		2L	16, 25, 35, 51, 52, 41, 31, 21
Five membered rings		A+B	16, 26, 40, 40, 42, 26, 20
K	12, 18, 28, 42, 40, 36, 26, 12		
Q	13, 19, 28, 40, 44, 46, 36, 4		
T	13, 20, 29, 38, 37, 43, 32, 12	**Seven membered rings**	
M,N,L	13, 20, 29, 42, 42, 37, 29, 16	R+S	18, 26, 38, 48, 68, 58, 36, 18
R	13, 20, 29, 42, 25, 41, 27, 14	P+Q	18, 26, 38, 56, 66, 64, 46, 8
F	13, 20, 30, 42, 41, 33, 16	I+J	18, 28, 42, 52, 54, 54, 34, 14
D	13, 20, 30, 42, 43, 39, 11	T+U	18, 28, 42, 53, 54, 60, 43, 18
H	13, 20, 30, 46, 44, 32, 12	2D	18, 28, 42, 58, 62, 62, 16
G	13, 21, 33, 43, 38, 32, 19	C+D	18, 28, 43, 57, 62, 56, 15
J	14, 22, 30, 38, 36, 32, 18, 10		
N+O	14, 22, 30, 42, 42, 32, 22, 16	**Eight membered rings**	
E+G	14, 22, 32, 43, 37, 27, 16	2Q	20, 30, 44, 64, 72, 76, 60, 8
		I+J+K	20, 30, 46, 60, 66, 62, 42, 14
		2T	20, 30, 46, 62, 62, 70, 54, 18
		L+M	20, 31, 47, 67, 70, 59, 51, 27

with the size of the ring, though overlapping with rings of the next higher or smaller size is possible. Second, there are cases of different rings having several leading terms equal. Third, the differences of the individual terms in the sequences for rings of the same size are not excessive, especially if the leading terms are the same. Fourth, the rings of different size, even if having the same beginning for the path sequence, will differ
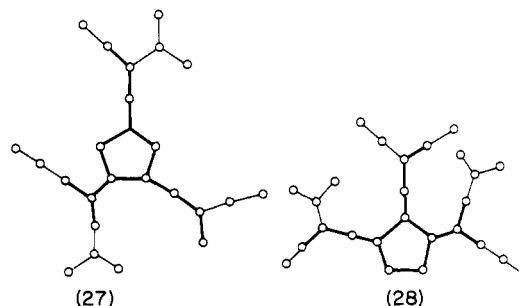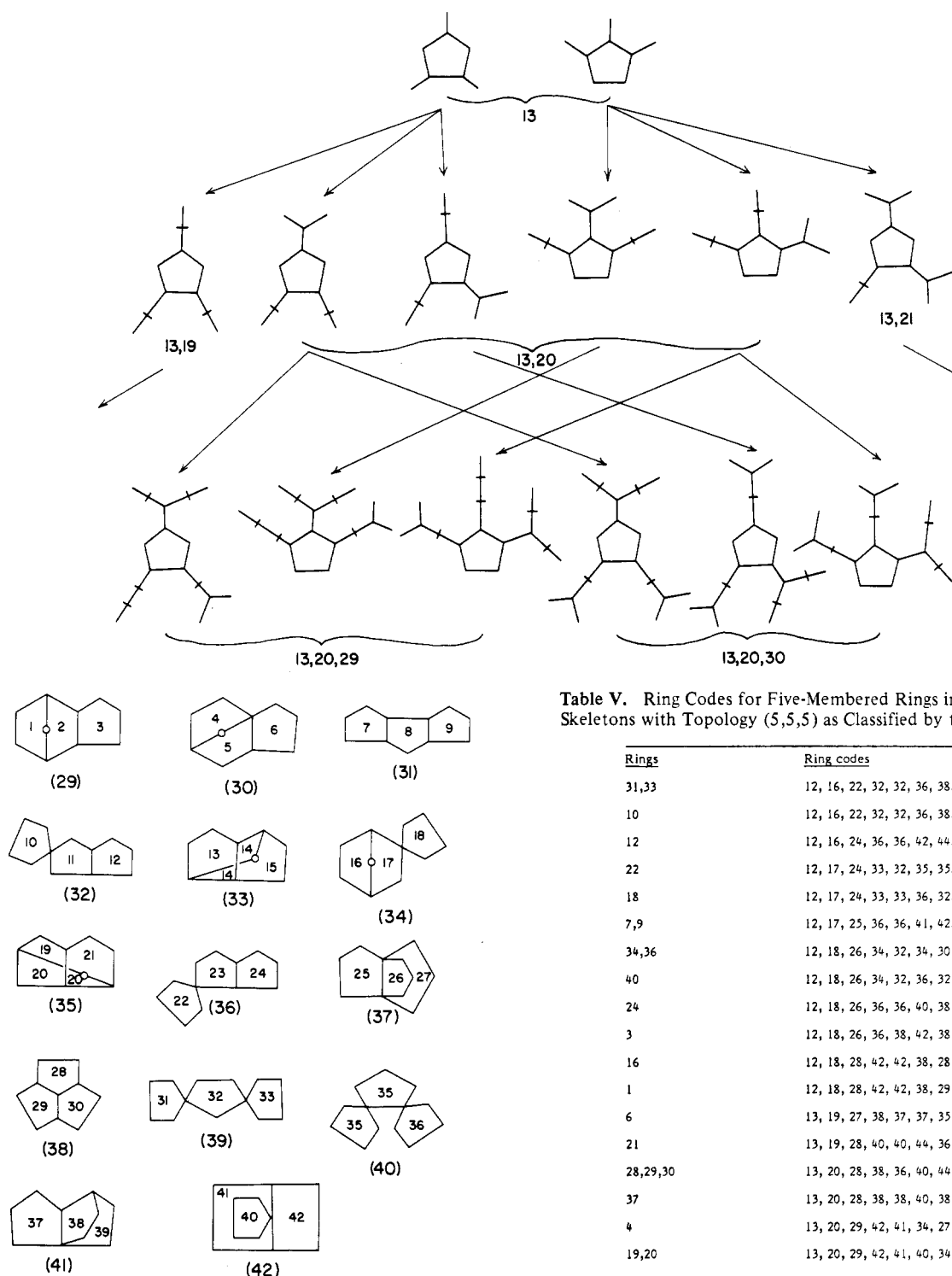

Figure 5. Two five-membered rings have the same initial entries in the path sequences showing the same kind of appended fragments but at different positions.

considerably when paths of longer length are counted. In summary, Table IV confirms the expectation that the suggested ring encoding not only offers useful discrimination but also reflects salient structural features. It is not surprising that the beginning entry may be the same for rings of different size. It has already been shown that the first entry includes the number of exocyclic bonds. Hence, a four-membered ring with four exocyclic bonds will have the same leading term as a five-membered ring with two exocyclic bonds. The cases of rings with several leading terms being equal indicates rings which, within a radius of several bonds, have the same number of exocyclic paths, although the *distribution* of the paths may vary; and the apparent similarity in the rings environments are limited. Consider the case of molecules **18** and **19** and the corresponding ring codes shown below:

molecule	ring code
18	13,20,30,42,41,33,16
19	13,20,30,42,43,39,11

In Figure 5 are depicted the environments for the two rings within the radius of four bonds; and, as an inspection of the Figure 5 shows, the exocyclic fragments are the same but differently distributed. Thick lines indicate the part of the environment which excludes occurrence of the same vertex of the ring. Obviously after some distance paths of longer length will return to the original ring. One may wish to exclude such reentry-type paths in the count. This may be particularly attractive for more complex polycyclic graphs where the number of paths will proliferate fast (e.g., the graphs in Figure 3). The observation that rings of the same size tend to have similar magnitudes for the respective path number reflects that the rings considered generally may be viewed as similar types. Analogously to the measure of similarity among molecules based on molecular codes,[21] one can measure similarity among rings by summing squared differences (i.e., using a metric function) for the corresponding entries in the ring sequences. Rings with the same several leading terms need not necessarily also be the most similar rings when paths of all lengths are considered. The distinction between the two situations is worth emphasizing. The beginnings of sequences indicate the *local* environment of rings, and hence, are of interest when *local* properties dominate. However, a complete sequence is more typical of the *global* ring environment and will better reflect global features (possibly the pattern which would correspond to a Schlegel projection of the structure with the particular ring in a center), hence, will be of more interest when global properties are considered. Thus, care must be taken in the use of truncated codes which are of value in relating step-by-step variations in ring environments and showing possible generic relationships of different ring environments. For the rings of Figure 5 local environments up to the radius of three bonds are depicted in Chart I. One can see from the chart how gradual inclusion of longer paths in the ring codes discriminates among closely related ring environments. The

Chart I. The Generic Relationship of Different Five-Membered Rings Found in Tricyclic Octanes of Figure 4. The truncated codes indicate the rings environment at the radii of one, two, and three bonds.


Figure 6. Tricyclic skeletons classified as (5,5,5) by the Ring Index. The ring labels correspond to the list of ring codes in Table V.

Table V. Ring Codes for Five-Membered Rings in All Molecular Skeletons with Topology (5,5,5) as Classified by the Ring Index[a]

Rings	Ring codes
31,33	12, 16, 22, 32, 32, 36, 38, 32, 20, 12, 4
10	12, 16, 22, 32, 32, 36, 38, 32, 20, 14, 2
12	12, 16, 24, 36, 36, 42, 44, 24, 16, 8
22	12, 17, 24, 33, 32, 35, 35, 30, 20, 12, 6
18	12, 17, 24, 33, 33, 36, 32, 24, 14, 4
7,9	12, 17, 25, 36, 36, 41, 42, 26, 20, 14
34,36	12, 18, 26, 34, 32, 34, 30, 26, 20, 12, 8
40	12, 18, 26, 34, 32, 36, 32, 24, 16, 8
24	12, 18, 26, 36, 36, 40, 38, 24, 16, 8, 4
3	12, 18, 26, 36, 38, 42, 38, 24, 9
16	12, 18, 28, 42, 42, 38, 28, 20, 12, 4
1	12, 18, 28, 42, 42, 38, 29, 26, 12
6	13, 19, 27, 38, 37, 37, 35, 19, 2
21	13, 19, 28, 40, 40, 44, 36, 6
28,29,30	13, 20, 28, 38, 36, 40, 44, 24, 12
37	13, 20, 28, 38, 38, 40, 38, 12, 6
4	13, 20, 29, 42, 41, 34, 27, 21, 8
19,20	13, 20, 29, 42, 41, 40, 34, 11
13, 15	13, 20, 29, 42, 42, 37, 29, 16
23	14, 22, 30, 38, 32, 28, 24, 12, 8, 6, 4
11	14, 22, 30, 38, 32, 26, 24, 12, 8, 8, 4
8	14, 22, 30, 38, 32, 30, 32, 16, 16, 14
32*, 35*	14, 22, 30, 38, 32, 24, 16, 8
25,26,27	14, 22, 30, 38, 36, 36, 36, 12, 8, 4
14	14, 22, 30, 38, 42, 32, 22, 15
17	14, 22, 30, 40, 36, 28, 18, 14, 8, 2
38,39	14, 22, 30, 41, 33, 27, 12, 7
5	14, 22, 30, 41, 39, 26, 17, 6

[a] Asterisks (*) indicate a case of nonequivalent rings having the same ring-path sequence.

diagrams are accompanied by the part of the code, the length of truncation corresponding to the radius of the ring environment shown. Therefore, the number of exocyclic bonds, the number of pairs of adjacent exocyclic bonds, and the number of paths of length three are indicated.

(5,5,5)-Rings from the Ring Index. In order to better illustrate the variations that are possible for ring codes of rings of the same size but in different environment, all topologically distinctive rings classified by the Ring Index[34] as (5,5,5) have been examined. The molecular skeletons from which the rings

are extracted appear in Figure 6. The ring codes have been listed in Table V. The ring codes have been ordered lexicographically, that is, according to the relative magnitudes of the first different entry in each of the respective ring sequences. The previously mentioned regularities for the tricyclic eight- and nine-carbon structures are again visible and require no additional comment. The enlarged collection of ring indexes and structures allows additional observations. Here it is evident that ring codes starting with 12 as the first entry generally belong to peripheral rings (as these generally have fewer exocyclic links), while more centrally located rings have an initial sequence entry of 14. One can also observe that structures having spiro rings tend to have fewer paths. This is most obvious in comparison of paths of intermediate length. The importance of the whole ring sequence, as opposed to the truncated sequence, is well illustrated with the following examples:

| molecule	ring	ring sequence
29	1	12,18,28,42,42,38,28,20,12,4
34	16	12,18,28,42,42,38,29,26,12
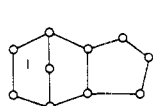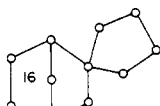
29   **34**

Despite the extended coincidence in many of the leading terms in the sequence, the sequences for the two structures appear less similar than one might expect. The distance of the two rings in the corresponding ring-space (multidimensional space in which the entries of the ring codes are taken as the coordinates of the ring) is 7.28 which gives a quantitative measure of the similarity of the two rings. There are many rings which are more similar among themselves, than the rings 1 and 16 of the indicated substituted norbornane skeleton. For example, rings 10 and 22 in the corresponding spiro environment (shown below) have a similarity given by the distance 5.57. This smaller value corresponds to the points in the ring-space which
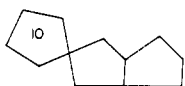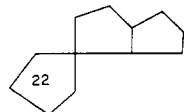

32   36

are closer together and, thus, are defined as more similar to one another than in the previous comparison. Clearly chemical rings 1 and 16 differ, since not only the number of substitutions is important, but also their relative placement. Within the limitations of the ring codes one can still find many useful applications, in particular when examining a large number of ring environments.

Although so far we have not encountered nonequivalent rings that have identical path numbers, such a possibility exists and one should not immediately conclude that rings are equivalent because their sequences are the same. Such a situation should be carefully examined before drawing any conclusion. Clearly, however, nonequivalent sequences indicate nonequivalent rings, and this may be of interest when comparing ring structures. Since visualization of polycyclic complex structures is not always so simple, ring indexes may help in recognizing equivalent rings, or, to put it more correctly, may help in prescreening, since equivalence subsequently has to be verified. For example, in "The Ring Index", we find structures such as:
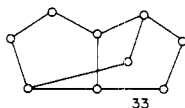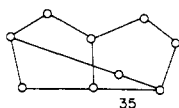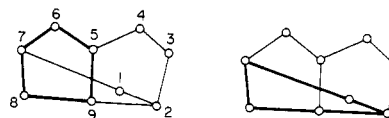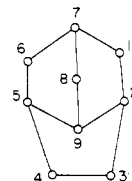

33   35

Because of the particular fashion chosen for drawing the structures it may not immediately be apparent that these molecules have symmetry which causes equivalence of some of the rings. The equivalent rings in structure 35 are shown below:

35a

This would have been obvious if the structure had been drawn in a more symmetrical representation:

35b

The use of ring indexes for recognizing equivalent rings should be tested on large structures, such as the polyadamantanes discussed later, in order to better appreciate this particular application, since the simplicity of the examples shown above does not do proper justice to the importance of such applications.

Finally a pair of nonequivalent rings, rings **32** and **35** in the two dispiro compounds of Figure 6, should be noted. These confirm that the ring indexes are not unique. Such rings are expected to show maximal similarity, and indeed the rings have the same type of ring neighbors, the difference being in the *distribution* of the attached spiro rings. The situation is very much similar to rings with exocyclic bonds or fragments of the same kind but differently distributed (see Figure 5 and Chart I), since the spiro atom behaves like a branching atom of an acyclic pending fragment and can be used only once in path counting. In the case of fused rings, the path can always reenter the ring and thus further differentiates similarly substituted fragments.
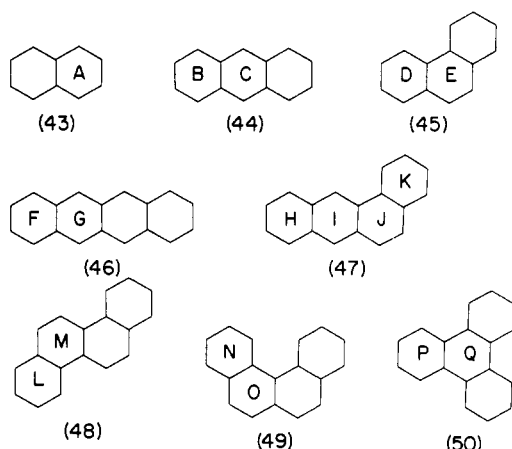
BENZENOID HYDROCARBONS

In the following discussion, molecular graphs of several polycyclic benzenoid hydrocarbons will be considered. No discrimination between bonds will be made, so that, strictly speaking, naphthalene becomes equivalent to decalin, anthracene to a molecule with three fused cyclohexane rings, and so forth. One may distinguish between such cases by introducing appropriate weighting factors,[36] but here the primary interest is to see how different environments in these very regular structures is reflected in the ring codes. Table VI lists truncated ring codes for a selection of benzenoid hydrocarbons. The rings have been identified in Figure 7. As one can see, after a few initial path numbers, the different rings generally lead to different ring sequences, so that for a comparison one may truncate the code, as has been done after the sixth entry. The truncated rings can be lexicographically ordered. Such an order clearly reflects some of the underlying structural conditions. As is expected from the regularities for rings in the previously considered compounds, the path codes for *terminal* rings start with the smallest digits (here 14). The numbers increase as the size of the molecule increases. That also is to be expected and is most clearly seen in the case of the naphthalene, anthracene, tetracene series. Notice that in linearly fused acenes *all* ring path numbers are *even*. The first appearance of an *odd* entry indicates a "kink", that is, an angularly fused ring. The absence of an odd entry, however,

Table VI. Truncated Ring Codes for Benzenoid Hydrocarbons Having Four or Less Fused Benzene Hexagons[a]

Molecule	Ring	Truncated code
Naphthalene (43)	A	14, 18, 22, 34, 30, 20, ...
Anthracene (44)	B	14, 18, 24, 34, 46, 44, ...
	C	16, 24, 32, 40, 48, 40, 36, . . .
Phenanthrene (45)	D	14, 19, 26, 35, 46, 44, ...
	E	16, 24, 32, 40, 48, 40, 38, . . .
Tetracene (46)	F	14, 18, 24, 34, 48, 54, ...
	G	16, 24, 32, 42, 58, 64, ...
Benzanthracene (47)	H	14, 18, 24, 35, 51, 57, ...
	I	16, 25, 36, 49, 64, 64, ...
	J	16, 24, 34, 48, 64, 64, ...
	K	14, 19, 26, 36, 51, 57, ...
Chrysene (48)	L	14, 19, 26, 36, 52, 60, ...
	M	16, 25, 33, 49, 64, 64, ...

[a] Since the aromatic bond has not been discriminated as such, the same codes apply to structures built from cyclohexane rings.

Figure 7. Benzenoid hydrocarbons having four or fewer fused benzene rings. The labels correspond to the ring codes of Table VI.

does not mean no angularly fused ring is present. For example, the central ring of phenanthrene has all entries even (inclusive of paths are not shown here because of truncation) although the molecule is an angularly used catacondensed benzenoid. The more distant the ring from the "kink" site, the later in the sequence the odd entry arises. This particular property may be of considerable interest in searching files and screening structures. For instance, if a list of compounds with a "bay" region, because of their relevance to carcinogenic activity, were desired,[37] one would look for rings with the *second entry being an odd number*, provided the compounds had already been screened for benzenoid topology. Since in most biological and biomedical correlations the size of the molecule is an important factor, one can further eliminate undesirable structures (either too small or too large) by specifying a tolerance level on the path numbers. For instance, by requiring that p_6 (i.e., paths of length six) be larger than 50, one eliminates systems having three fused rings. By careful analysis of the individual entries in a series of related compounds, one can take advantage of the particular encoding of molecular fragments. Here, it is merely observed that such an approach has great *potential*. It is outside the scope of the present paper to elaborate or discuss specific applications. In any case, for the purpose of screening molecules, the ring codes or, more generally, frag-
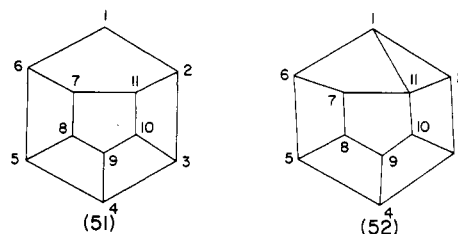
Figure 8. A selection of polycyclic structures. The numbering of atoms corresponds to the ring labels of Table VII.

Table VII. Selected Molecules and Their Ring Codes (Truncated)

Molecule	Ring atoms	Path numbers (truncated)
	2, 3, 8, 9	12, 23, 44, 73, 123, ...
(51)	3, 4, 9, 10	12, 24, 47, 74, 123, ...
	1, 2, 6, 7, 8	14, 26, 50, 80, 136, ...
	7, 8, 9, 10, 11	15, 30, 58, 114, 148, ...
	1, 6, 7	10, 22, 40, 74, 120, 202, . . .
	2, 3, 8, 9	12, 25, 53, 92, 165, ...
(52)	3, 4, 9, 10	12, 24, 50, 92, 172, ...
	1, 2, 7, 8	13, 28, 55, 95, 158, ...
	7, 8, 9, 10, 11	16, 34, 68, 116, 176, ...

ment codes including the atomic codes, may be more useful than the molecular codes as a whole, since in the latter many of the details about the structure are already condensed and are no longer accessible for immediate inspection.

SELECTED SPECIAL CASES

In Figure 8 molecular graphs for selected polycyclic structures are given and in Table VII are the corresponding ring codes for selected rings. The first two graphs illustrate how lexicographic ordering of ring codes serves to point to minor differences among rings in a similar environment. The difference between the two 11-atom structures is the presence (or absence) of a bond between atoms 1 and 11. This makes some of the four-membered rings nonequivalent. Observe now that in case of the structure **51**, the reduced valency of atom 1 makes rings closer to atom 1 have the smaller initial entries. This arises because at an earlier stage in the path counting, the preferred situation (an atom with fewer neighbors) is encountered. For structure **52**, the opposite is the case, since here the unique atom has an increased valency. Thus, rings far removed from that atom will have smaller leading terms in the path enumerations. Hence, by reviewing ring codes, even for rings of the same size and the same number of exo-cyclic bonds, one can infer something about their relative positioning with the respect to a unique atom (which in this case, differs in valency from the remaining atoms). One can use this property even in cases where there are no apparent discriminating features for an atom selected on chemical grounds as unique by special "marking". For instance, one can attach fictitious "ghost atoms" to such an atom, doubling its valency and thus increasing the number of shorter paths lengths for atoms in the neighborhood, with the consequence of discriminating among rings in the vicinity. Such an approach may be particularly attractive for recognizing strategic rings and other special structural features in a larger molecular skeleton.

POLYADAMANTANES

As a final example, several polyadamantanes are considered. The study of adamantane and its higher homologues represent an actively pursued branch of chemistry initiated by Prelog

CHARACTERIZATION OF THE RINGS OF A MOLECULE

J. Chem. Inf. Comput. Sci., Vol. 20, No. 1, 1980 **45**

Table VIII. Path Sequences Corresponding to Adamantane Units in Polyadamantanes[a]

Molecule	Fragment	"Adamantane" - fragment codes (Truncated)
(53)	A	28, 53, 91, 168, 302, 470, ...
	B	31, 60, 110, 193, 336, 502, ...
(54)	C	29, 57, 100, 188, 318, 509, ...
	D	33, 63, 120, 198, 360, 510, ...
(55)	E	29, 55, 99, 177, 312, 480, ...
	F	31, 60, 110, 193, 336, 509, ...

[a] Molecules are shown in Figure 9 where the various fragments A, B, C, etc., are identified.

and Seifert with the first synthesis of the parent compound.[38] As the number of fused adamantane units increases, the diversity in local ring environments also increases. However, rather than continuing with listing ring codes for such compounds, Table VIII lists codes that correspond to an adamantane fragment, as a next obvious larger building block in such compounds. Molecular skeletons are illustrated in Figure 9 with an "exploded" view of the individual adamantane fragments. The convenience of the path codes is obvious here. Furthermore, the concept need not be limited to rings, but applies equally well to any selected parts of a structure considered under the circumstances to be a building block. Most of the properties of ring codes already indicated are also valid here and correspond to fragment codes. An immediate use of such codes is to locate parts of the molecule with similar local environments. While equivalence of various "adamantane" units is not apparent from a limited selection of projections, when considering actual molecular models, visualization of such equivalent rings or fragments becomes easier. Nevertheless, even actual viewing of models may not always reveal equivalence which is always revealed by the equality of ring or fragment codes. Hence, no omissions are possible but verification is required to discriminate among "accidental" degeneracy and actual symmetry degeneracy. In such applications fragment codes will be useful, even though not indispensable. If however, one is interested in *similar* environments (local similarity to a specified distance), viewing molecular models need not help at all and may even be misleading. In this case, ring and fragment codes are essential tools in the search.

CONCLUSION

The outlined concept of ring fragment codes appears to show promise in being conceptually simple, computationally practical, and sufficiently discriminatory in application. The scheme also allows some latitude in modification. It is premature at this point to specify all the details for potential use; and, indeed, optimal solutions may well depend on the purpose of the particular application. Since the concept of atom codes has been extended to include cases of unsaturated systems, aromatic compounds, and some hetero compounds, one anticipates that such extensions to ring codes and other fragment codes should be fairly straightforward. The real limitation of the concept of path-based ring codes is the same as with the atomic codes. That is the exponential growth of the number of paths as the number of vertices and number of edges in a graph increases and can be limiting. In the case of molecular graphs (i.e., graph representing molecular skeletons) this particular limitation may well turn out to be more of an academic nature than a real obstacle, since atomic valencies are limited in the first place. On the other hand, any pragmatic course of action will require truncation of path enumerations, after a certain length (which must be empirically determined
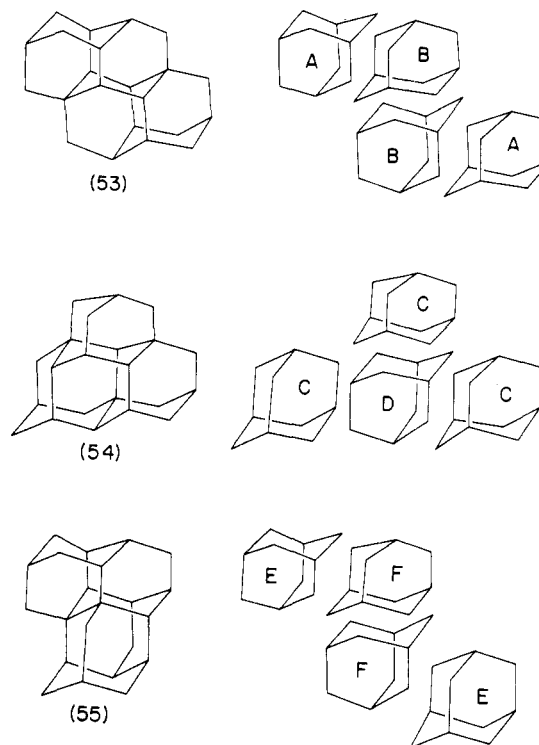

Figure 9. Polycyclic polyadamantanes accompanied with an "exploded" view with identification of the individual "adamantane" units.

for problems of different character). Therefore, the exponential nature of the path enumeration problem is curtailed if not by-passed completely. In the case of more complex chemical graphs, such as the graphs arising in the study of chemical isomerizations and other complex transformations, one may be confronted with the presence of an "explosion" of the number of paths (graphs of Figure 3). The same will also be the case with unusually large molecules (e.g., polymers or proteins). Truncation of path sequences may be necessary in those situations. However, even in such apparently hopeless situations, which would, for instance, include counting of random walks on regular grids, a problem which can currently only be tackled by probabilistic methods,[39] there may well be some possible alternates in the future. Besides the expected availability of computers with parallel data-processing capabilities, one also should expect some progress in analytical representations for the number of paths of different length in selected regular structures. Such expressions have been already derived for complete graphs K_n and complete bipartite graphs $K_{n,n}$. Hence, some highly complex structures may eventually be susceptible to such analytical schemes. It appears, however, that most common chemical applications will rarely encounter the limitations; and when the frequent use of searches is required, it may be more economical to compromise between the extensive path enumerations and truncated path use with occasional mismatches which can be eliminated on other grounds. It appears that the scheme outlined offers enough opportunities for various individual applications, is capable of meeting present needs, and can easily be modified when necessary.

A modification which is of immediate interest is incorporation of heteroatoms. Although heteroatoms are not frequently treated in graph theoretical considerations, graph theory is capable of handling some such situations. Ignoring atom differences makes graph theoretical schemes simple. Observe, however, that the reduced precision in characterization is accompanied by reduced input information, so the loss of discrimination is not inherent in graph theory as such.

The information on heteroatoms corresponds to coloring of vertices and edges in a graph. Such colored graphs would allow a similar analysis of paths of different length for any of the colors of interest. Clearly, with more input information one can only expect further increases in the differentiation of atomic or fragment environments. The cost, of course, is an increase in the bookkeeping of records. Hence, with respect to heteroatoms, one can anticipate their incorporation in the analysis. Stereochemistry, on the contrary, is not suited for straightforward graph theoretical analysis. Strictly speaking, it is beyond graph theory and one would have to adopt some *conventional* descriptors to supplement the partial results that follow from examination of molecular connectivity. We do not want to leave the reader with the impression that graph theory has unlimited potential in chemical analysis. After all, graph theory addresses itself essentially to the combinatorial and topological aspects of molecular structure. Our efforts are to point to rather considerable areas where graph theory has uses, and the concept of ring codes appears to be one such not yet fully explored area. When considering applications to problems where stereochemistry cannot be overlooked and where the positions of heteroatoms are critical, as is the case in structure–activity relations and numerous other correlations involving selected properties, in the absence of reliable alternatives one can use the approach outlined here provided one confines the discussions to structures of similar composition and similar functional groupings combined with similar or related spatial architecture. This may appear rather restrictive but the most interesting studies do involve compounds of similar composition and structure where minor variations in position or the kind of the substituent may dramatically affect properties. In order to differentiate naphthalene and decalin rings we hardly need graph theory!

ACKNOWLEDGMENT

REFERENCES AND NOTES

(1) H. Sachs, *Publ. Math. (Debrecen)*, **11**, 119 (1964).
(2) (a) M. Randić, *Chem. Phys. Lett.*, **38**, 68 (1976); (b) *J. Am. Chem. Soc.*, **99**, 444 (1977); (c) *Tetrahedron*, **33**, 1905 (1977).
(3) J. Lederberg, *Proc. Natl. Acad. Sci.*, **53**, 134 (1965).
(4) A. Zamora, *J. Chem. Inf. Comput. Sci.*, **16**, 40 (1976); M. Plotkin, *J. Chem. Doc.*, **11**, 60 (1971).
(5) (a) C. C. Gotlieb and D. G. Corneil, *Commun. Assoc. Comput. Mach.*, **10**, 780 (1967); (b) K. Paton, *ibid.*, **12**, 514 (1969); (c) J. C. Tiernan, *ibid.*, **13**, 722 (1970); (d) R. Fugmann, U. Dolling, and H. Nickelsen, *Angew. Chem.*, **6**, 723 (1967).
(6) M. Frerejacque, *Bull. Soc. Chim. Fr.* **6**, 1008 (1939). As indicated in ref 5d, there are some undesirable aspects of such determination of the ring number, since one is forced to exclude some rings while retaining others that are in some instances equivalent from the structural point of view.
(7) (a) E. J. Corey and W. T. Wipke, *Science*, **166**, 178 (1969); (b) E. J. Corey and G. A. Peterson, *J. Am. Chem. Soc.*, **94**, 460 (1972); (c) M. Bersohn, *J. Chem., Perkin Trans. 1*, 1239 (1973); (d) A. Esack, *ibid.*, 1120 (1975).
(8) W. T. Wipke and T. M. Dyott, *J. Chem. Inf. Comput. Sci.*, **15**, 140 (1975).
(9) The literature is very informative here. Among various schemes one may mention the "walking" algorithm, which resembles a node-to-node matching of graphs [L. C. Ray and R. A. Kirch, *Science*, **126**, 814 (1957)], use of spanning trees, and considering bonds not involved which contribute to ring closure and thus lead to rings of interest. One finds an equivalent set of rings by applying the logical *exclusive* or operation to a sequence of bonds (see ref 8 for an illustration).
(10) K. Conrow, *J. Chem. Doc.*, **6**, 206 (1966).
(11) J. Meinwald and J. K. Crandall, *J. Am. Chem. Soc.*, **88**, 1330 (Appendix) (1966).
(12) While the original Baeyer system [A. Baeyer, *Ber.*, **33**, 3771 (1900)] was designed for bicyclic structures, the coded part (*x, y, z* in the name bicyclo[*x.y.z*]alkane) had a simple *structural* interpretation, representing the *number* of atoms in each of the three bridges in the structure. The extension to tetracyclic systems introduced atomic *labels* as additional indicators. Hence, such a scheme is no longer purely structural and requires knowledge of the *rules* for numbering. It is possible to extend the Baeyer scheme to tetracyclic structures still without use of atomic *labels* (M. Randić, to be published) and solely using a structural parameters. Examples of special such applications are given by: D. Hellwinkel, *Chem.-Ztg.*, **94**, 715 (1970); N. Tanaka, T. Iizuka, and T. Kan, *Chem. Lett.*, 539 (1974).
(13) M. Randić, *Acta Crystallogr., Sect. A*, **34**, 275 (1978).
(14) F. T. Smith, *J. Chem. Phys.*, **34**, 793 (1961), Table III.
(15) A. T. Balaban and F. Harary, *Tetrahedron*, **24**, 2505 (1968). A. T. Balaban, *MATCH (Inf. Commun. Math. Chem.)*, No. 2, 51 (1976).
(16) O. E. Polansky and D. H. Rouvray, *MATCH (Inf. Commun. Math. Chem.)*, No. 2, 63 (1976).
(17) H. E. Zimmerman and M. Randić, submitted for publication.
(18) M. Randić, Proceedings of the Bremen Conference on Mathematical Structures in Chemistry (to appear in *MATCH (Inf. Commun. Math. Chem.)*).
(19) (a) M. Randić, *J. Chem. Inf. Comput. Sci.*, **18**, 101 (1978); (b) M. Randić and C. L. Wilkins, *ibid.*, **19**, 23 (1979).
(20) M. Randić, *Theor. Chim. Acta*, submitted for publication.
(21) (a) M. Randić and C. L. Wilkins, *Chem. Phys. Lett.*, **63**, 332 (1979); (b) *J. Phys. Chem.*, **83**, 1525 (1979).
(22) M. Randić and C. L. Wilkins, *J. Chem. Inf. Comput. Sci.*, **19**, 31 (1979).
(23) (a) C. L. Wilkins and M. Randić, *Theor. Chim. Acta*, submitted for publication; (b) M. Randić and C. L. Wilkins, *Int. J. Quantum Chem.*, in press.
(24) J. Lederberg, Dendral-64, Interim Report to the National Aeronautics and Space Administration, Cr 68898, Dec. 1965.
(25) R. H. Penny, *J. Chem. Doc.*, **5**, 113 (1965).
(26) M. Gordon and J. W. Kennedy, *J. Chem. Soc., Faraday Trans. 2*, **69**, 484 (1973).
(27) H. L. Morgan, *J. Chem. Doc.*, **5**, 106 (1965).
(28) W. Schubert and I. Ugi, *J. Am. Chem. Soc.*, **100**, 37 (1978).
(29) J. E. Dubois, D. Laurent, and A. Aranda, *J. Chim. Phys.*, **70**, 1608 (1973).
(30) J. E. Dubois in "Chemical Applications of Graph Theory", A. T. Balaban, Ed., Academic Press, London, 1976.
(31) C. A. Shelley and M. Trulson, private communication, 1978. The authors used the molecule assembler in program CASE (Computer-Assisted Structure Elucidation) and verified the conjecture on 159 C_{11} isomers, 355 C_{12} isomers, 802 C_{13} isomers, and finally on 1858 C_{14} isomers.
(32) F. E. Harris, private communication, 1978.
(33) M. Randić, G. M. Brissey, R. B. Spencer, and C. L. Wilkins, *Comput. Chem.*, **3**, 5 (1979).
(34) (a) M. Randić, *J. Chem. Inf. Comput. Sci.*, **17**, 171 (1977); (b) *Chem. Phys. Lett.*, **42**, 283 (1976); (c) *Croat. Chem. Acta*, **49**, 643 (1977); (d) *Int. J. Quant. Chem.*, in press. The approach has now been programmed in Fortran and Basic and will shortly be published: M. Randić, G. M. Brissey, and C. L. Wilkins in preparation. A sample copy of the program is available upon request.
(35) A. M. Patterson, L. T. Capell, and D. F. Walker, "The Ring Index", 2nd ed, American Chemical Society, Washington, D.C., 1960.
(36) M. Randić, G. M. Brissey, R. B. Spencer, and C. L. Wilkins, to be published.
(37) D. M. Jerina and J. W. Daly in "Drug Metabolism", D. V. Parke and R. L. Smith, Eds., Taylor and Francis, London, 1976, p 15. Dr. J. Kaufman reported at the 1979 Sanibel Symposium on Quantum Biology and Pharmacology that a graph theoretical algorithm for finding benzenoids with a "bay" region has been developed.
(38) V. Prelog and R. Seiwerth, *Chem. Ber.*, **74**, 1644 (1941).
(39) D. E. Knuth, *Science*, **194**, 1235 (1976).