

- Classification of Chemical Structures", *J. Chem. Inf. Comput. Sci.*, **15**, 55-58 (1975).
- (23) G. W. Adamson and J. A. Bush, "Evaluation of an Empirical Structure-Activity Relationship for Property Prediction in a Structurally Diverse Group of Local Anaesthetics", *J. Chem. Soc., Perkin Trans. 1*, 168-172 (1976).
- (24) G. W. Adamson and D. Bawden, "A Method of Structure-Activity Correlation Using Wiswesser Line Notation", *J. Chem. Inf. Comput. Sci.*, **15**, 215-220 (1975).
- (25) G. W. Adamson and D. Bawden, "An Empirical Method of Structure-Activity Correlation for Polysubstituted Cyclic Compounds Using Wiswesser Line Notation", *J. Chem. Inf. Comput. Sci.*, **16**, 161 (1976).
- (26) O. Exner in "Advances in Linear Free Energy Relationships", N. B. Chapman and J. Shorter, Ed., Plenum Press, London, 1972, pp 44-46.
- (27) A. Albert in "Physical Methods in Heterocyclic Chemistry", A. R. Katritzky, Ed., Academic Press, London: Vol. 1, 1963, p 2; Vol. 3, 1971, p 1.
- (28) T. Fujita, "The Analysis of Physiological Activity of Substituted Phenols with Substituent Constants", *J. Med. Chem.*, **9**, 797-803 (1966).
- (29) J. P. Tollenaere, "Structure-Activity Relationships of Three Groups of Uncouplers of Oxidative Phosphorylation: Salicylanilides, 2-Tri-fluoromethylbenzimidazoles, and Phenols", *J. Med. Chem.*, **16**, 791-796 (1973).
- (30) E. G. Smith and P. A. Baker, "The Wiswesser Line-Formula Chemical Notation", 3rd ed, Chemical Information Management Inc., Cherry Hill, N.J., 1976.
- (31) G. W. Snedecor and W. G. Cochran, "Statistical Methods", 6th ed, Iowa State University Press, Ames, Iowa, 1967.
- (32) "Statistical Analysis Mark 2 Applications Package", ICL Technical Publication 4301, International Computers Ltd., London, 1971.
- (33) J. G. Topliss and R. J. Costello, "Chance Correlations in Structure-Activity Studies Using Multiple Regression Analysis", *J. Med. Chem.*, **15**, 1066-1068 (1972).
- (34) J. E. Dubois, J. J. Aaron, O. Alcais, J. P. Doucet, F. Rothenberg, and R. Ucan, "A Quantitative Study of Substituent Interactions in Aromatic Electrophilic Substitution. 1. Bromination of Polysubstituted Benzenes", *J. Am. Chem. Soc.*, **94**, 6823-6828 (1972).
- (35) B. Roth and J. Z. Strelitz, "The Protonation of 2,4 Diaminopyrimidines. 1. Dissociation Constants and Substituent Effects", *J. Org. Chem.*, **34**, 821-836 (1969).
- (36) P. Willett, M.Sc dissertation, PGSLS, Sheffield University, 1976.
- (37) D. Bawden, Ph.D Thesis (in preparation), Sheffield University, 1976.

On Canonical Numbering of Atoms in a Molecule and Graph Isomorphism†

MILAN RANDIĆ

Energy & Mineral Resources Research Institute, Iowa State University, Ames, Iowa 50011

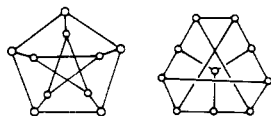
Received November 8, 1976

Use of a canonical numbering based on a particular interpretation of the adjacency matrix of a graph is advocated. The proposed numbering may be used in tests of isomorphism of graphs (molecular skeletons). It also makes possible study of the symmetry properties of graphs and recognition of equivalent vertices. The adjacency matrices based on the proposed canonical numbering may serve as a basis for ordering structures in a sequence. The relative position in the sequence is determined by the magnitude of the associated binary code derived by reading the entries of the adjacency matrix row by row from left and right and from top to bottom. For selected structures, such as isomers of paraffins, the derived ordering parallels certain molecular properties and thereby points to a topological origin of some correlations. This also suggests that the particular numbering involves some inherent features of the connectivity in molecular skeletons and graphs. The particular numbering hence may provide a basis for a systematic nomenclature which does not require supplementary rules. Although ultimately any labeling scheme is arbitrary, it is argued that the scheme proposed has additional properties which will facilitate solving some problems associated with graphs, and hence deserves some attention.

INTRODUCTION

Numbering of the atoms in a molecule and the related problem of numbering of vertices in a graph, apart from their apparent use in chemical documentation and nomenclature, have importance in other areas of science. In chemistry, chemical physics, statistical mechanics, and the theory of disordered structures, graphs provide a convenient representation of the combinatorial possibilities, facilitate visualizing contributing terms in expansions, and may also lead to a structural interpretation of correlation parameters.

One of the most important problems in the study of graphs is that of recognizing identical graphs.¹ It is generally quite difficult and tedious to assert that two graphs have an identical connectivity, as is well illustrated with well-known distinctive representations of the following Petersen graph:



A trial-and-error matching is time consuming and may require checking all $n!$ possibilities. This is essentially implemented

in the node-to-node search,² an early systematic attempt to resolve the problem. The scheme requires extensive book-keeping of the examined possibilities, involves considerable backtracking, and becomes impractical for applications to graphs of medium size and complexity. Alternative schemes considered in the literature attempt to recognize a singular property or a selection of properties that could differentiate between nonisomorphic graphs. Such properties include the graph spectrum,³ distribution of valences of vertices and discrimination of edge types, and examination of subgraphs and their characteristic polynomials.⁴ Sussenguth listed a set of crucial qualities;⁵ however, if one uses any combination of criteria such as mentioned above, one cannot be sure that all pairs of graphs would be differentiated by the criteria. This ultimate uncertainty undermines such efforts, which at best provide a list of necessary conditions for isomorphism.

The problem of graph isomorphism is related to the question of ordering graphs in a sequence. A complete order relation on graphs establishes which of two selected graphs precedes the other. So a search for criteria for ordering of graphs is fundamentally the same problem as the search for isomorphism. It is generally recognized that the use of a standard numbering procedure for vertices makes the problem of establishing isomorphism in graphs trivial and, one could add, makes ordering of graphs straightforward. The problem is in designing such a scheme which will apply to any graph so that the rules give a unique numbering; for a scheme to be practical,

† Portions of this material were presented to the Division of Chemical Information, 172nd National Meeting of the American Chemical Society, San Francisco, Calif., Aug 30, 1976, under the title "Symmetry Properties of Graphs" within the Symposium on Application of Nonnumerical Mathematics in Chemistry.

it should be possible to derive the numbering without screening an excessive number of undesired possibilities. A standard numbering yields a canonical matrix, which permits ordering of graphs. The concept of the canonical matrix is simple, elegant, and potentially useful. The idea has been considered by several people in various contexts, including molecular chemistry. Furthermore, specifically extremal matrices corresponding to the largest or the smallest binary code (when the rows are read from left to right and from top to bottom) have been mentioned.⁶⁻⁸ The procedure for an actual search for the canonical numbering, supplemented with numerous illustrative examples and discussion of the various properties of the accompanying canonical matrices, is described at some length in ref 8. The work, however, appears to have been obscured by deficiency in one of the *two* alternative algorithms outlined there. As has been found later by Mackay,⁹ the algorithm may not guarantee an absolute minimum for the derived binary code. The first alternative procedure suggested that by pairwise permutation of rows and columns one will arrive at the absolute minimum, but as we know now, the search can be locked in a local minimum.¹⁰ The second procedure, which in fact has been used for an actual search for the canonical numbering of cases illustrated in ref 8, starts with a completely unlabeled graph (in contrast to the first approach when one assumes an arbitrary labeling as initial). Then, in order to have in the first row as small a number as the situation allows, one assigns the smallest labels to vertices of the lowest valency, and as their adjacent labels combine the largest available labels. The process is carried out stepwise, one label at a time, since the already assigned labels may dictate the future assignments. This particular algorithm appears to be less suited for computer implementation, certainly less simple than the former, but is actually more efficient and powerful. In this paper we would like to elaborate this particular procedure and provide sufficient information to facilitate programming of the algorithm for computer applications. Moreover, by again addressing ourselves to the question of graph isomorphism, hopefully the message that the problem is solved by canonical numbering may reach those interested.

The scheme for graph isomorphism based on the canonical matrix has been anticipated by various people, in particular by Nagle⁶ who discussed the merits of such a scheme at length. Nagle considered a general linear ordering relation for undirected graphs; though, having in mind specific applications, he advocates introducing restrictions on canonical matrices. Previous work indicated a canonical matrix as a tool for resolving graph isomorphism tests. However, the concept of a canonical matrix has not been developed beyond recognizing its importance, and no plans have been suggested for its use in applications. For a scheme to be practical, one has to be able to obtain the desired numbering in a limited number of steps. Until and unless this is demonstrated, the problem of graph isomorphism remains unsolved.

Application of canonical numbering of vertices (atoms in a molecule) to chemical documentation is so obvious that it will not be elaborated upon. The examples of graphs selected for illustrations in this paper will represent both molecular skeletons and a more general class of chemical graphs associated with representations of various chemical transformations (isomerizations). This may promote chemical applications of graph theory, a field which is still not fully recognized among chemists despite its vigorous development in the last few years. Here we would like to emphasize another aspect of the problem of the numbering of atoms which is very relevant to chemistry. A choice of numbering of vertices is equivalent to a choice of a coordinate system in a mathematical treatment of a problem. Hence, although it cannot affect the

results in an observable fashion, it may *expedite solving the problem*. The choice of the numbering is then of considerable importance, and the existence of a numbering which suits a particular problem better than others should be examined. For molecules, and graphs in general, it is important to be in a position to compare different systems (skeletons), to recognize similarities, to classify different structures, and to place them in order according to a selected structural feature. Once a sequence is derived, one searches for molecular properties which themselves follow a similar sequence. Subsequently, one attempts to find a quantitative correlation between the selected chemical or physical property and the graph theoretical parameter used in ordering the structures. In this way an insight into the combinatorial nature of the particular molecular property is obtained.

In the following sections as shall (1) outline the suggested unique numbering of atoms in a molecule and (2) describe several interesting properties of such canonical numberings.

PRELIMINARIES

In the following we will refer to graphs, vertices, and edges, rather than to the more specific molecular skeletons, atoms, and bonds, respectively. The reason for this is that the approach equally suits other applications of graphs, such as their use in representing chemical transformations, and graphs of interest outside chemistry. Some of the relevant terminology can be found in the available introductory texts¹¹ or recent review articles concerned with application of graphs in chemistry.¹² In considering an analytical representation of a graph, one adopts some labeling of the vertices, using letters and numbers (or both) as convenient. For example:



Analytically, the above labelings may be represented by tables:

	1	2	3	4		a	b	c	d
1	0	1	0	0	a	0	1	0	0
2	1	0	1	1	b	1	0	1	1
3	0	1	0	1	c	0	1	0	1
4	0	1	1	0	d	0	1	1	0

Here the *j*th element in the *i*th row is given by

$$a_{ij} = \begin{cases} 1 & \text{(if } ij \text{ are connected)} \\ 0 & \text{(if } ij \text{ are not connected)} \end{cases}$$

The form of the table clearly depends on the assignment of labels to the individual vertices and an alternative selection will in general produce a different form for the table of the same graph:

	a	b	c	d
a	0	0	0	1
b	0	0	1	1
c	0	1	0	1
d	1	1	1	0

The two labelings are related through a permutation of the letters; symbolically:



$$L' = T L$$

where *L'* and *L* stand for new and old labeling, respectively, and *T* for the transformation. One can now verify that the two tables *A'* and *A* (new and old table, respectively) themselves are related through:

$$A' = T A T^{-1}$$

which is the transformation requirement for matrices. Hence the above tables qualify as matrices and the changes of numberings are equivalent to changes of coordinate systems or basis vectors in other calculations. In graph theory the above matrices are referred to as adjacency matrices.

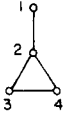
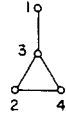
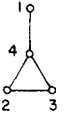
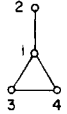
Another concept to be used in the present exposition is that of a *code*, and in particular the *structural code*. We shall illustrate the distinction in the nomenclature of familiar compounds: norbornane and adamantane

		
Systematic name (IUPAC)	Bicyclo[2.2.1]-heptane	Tricyclo[3.3.1.1 ^{3,7}]-decane
Code	2 2 1	3 3 1 1 (3,7)
Type	Structural	Mixed

The systematic names have a pretension to be structural in kind, but this is not always so. In the case of norbornane the digits 2.2.1 refer to *structural information*, viz., the number of carbon atoms in each of the three branches of the molecular skeleton. The same is, however, no longer true with the systematic name of adamantane, since here the digits 3,7 are *labels*, and so the code is label dependent. To encode it one also needs to know the numbering *rules* which apply. We see that adamantane requires *six* data; immediately a question arises of developing a code which will use solely structural information (as in the case of norbornane). Since the skeleton has *six* bridges, a natural extension would suggest the code 1 1 1 1 1 1, which signifies that each bridge of the highly symmetrical adamantane skeleton has only one CH₂ group. Such codes have been used in a classification scheme for corannulenes and in a systematic construction of skeletal forms of the possible isomers of decane.¹³

Let us briefly mention the advantages of structural codes in contrast to codes requiring some conventional agreements for their interpretation. One should bear in mind that codes, conventional, structural, or mixed, *always need some instructions for their interpretation*. The structural codes are unique in that *the instructions themselves are of structural origin*. For the purpose of storing and handling information in documentation, the nature of the code may in some instances be of secondary importance, but even here a structural approach has additional advantages, such as being more suited to a systematic generation of structures of prescribed type, or to cataloguing. The numerical digits in such codes may allow a *meaningful extension to structures when digits are altered*, increased, or decreased. Also the digital basis of the code permits a simple ordering of systems in a sequence which reflects some structural characteristics of such molecules. Furthermore, when codes are free of labels, as is the case with the code 1 1 1 1 1 1 of adamantane, a comparison of files which are based on different numbering schemes is simple. But in addition to such technical advantages, and more important for structural chemistry and chemometrics,¹⁴ a *development of molecular codes of exclusively structural origin is helpful for discerning structural features in various molecular properties*. As already mentioned, such codes facilitate *ordering* of structures which may be expected to *reflect* some structural characteristic. Going from the ordered sequence to a quantitative correlation takes some ingenuity in some instances, while in others the transition may be more straightforward. In all such cases, however, there is a considerable freedom in choice of structural parameters, but since these are nonobservable in nature, their selection is a matter of practical, traditional, and other preferences. Such structural parameters, frequently referred to as *indices* (e.g., ring indices,

Chart I

				...
$\begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix}$...

bond indices), play the role of a dummy variable; they only serve to facilitate a comparison of selected property between different molecules (or molecular fragments).

OUTLINE OF THE SCHEME

A graph with n vertices has $n!$ possible labelings. A few of 24 labelings of the four-vertex graph are indicated together with their adjacency matrices in Chart I. Now each of the matrices can be represented in a single line of digits by sequencing the rows. This process is a natural way to linearize a two-dimensional array and *does not correspond as such to some intrinsic mathematical operation*, for it resembles rather several *physical processes*, such as transmission of a picture at a distance. For example, for the third matrix the result is:

	0	0	0	1	0	0	1	1	0	1	0	1	1	1	0
Row	1st				2nd					3rd					4th

The above is one of the approaches to the linearization of graphs or structures—an important issue for the development of abbreviated line notations such as those of Wiswesser or Hayward.¹⁵ The line form of the adjacency matrix allows sorting of all $n!$ different numberings by considering the line as a single binary number. Such numbers can be ordered according to their magnitude, starting with the smallest binary number. The first member in the sequence has a unique position and is easily recognized. For the moment we will assume that there is just one such number, but even this, as will be seen later, is not an essential requirement. Hence, if we select the smallest binary number as a representation of the graph considered, we have in fact adopted a particular numbering that generates the smallest binary code. *For such a scheme to be practical, one would need an algorithm which produces the numbering desired without screening all $n!$ possibilities*. Initially, it has been suggested that one can arrive at such a numbering by considering permutations of rows and columns, two at the time, starting from an arbitrary numbered graph.⁸ Such a procedure, which has been implemented for a computer by Mackay,⁹ was found useful for application to molecular graphs, but does not guarantee that the final output represents an absolute minimum, since search restricted to permutations of two rows and columns at the time can be locked into a local minimum.⁹ Although the situation is not quite hopeless,¹⁶ it appears that the particular algorithm requires further elaboration. However, a *more efficient procedure for arriving at the desired numbering associated with the smallest binary code consists of the exploitation of its basic property: that it represents the smallest possible sequence of digits*. For this it may be more convenient (though not essential) to view the linear binary codes as a sequence of n (binary) numbers: (n_1, n_2, n_3, n_4) for the considered graph. To have the smallest binary code, obviously n_1 (the first row of the adjacency matrix) has to be as small as possible, because this is the leading part of the code. This means that the row should have as many zeros as possible, and that these should precede digit(s) 1. In terms of numbering this means that the

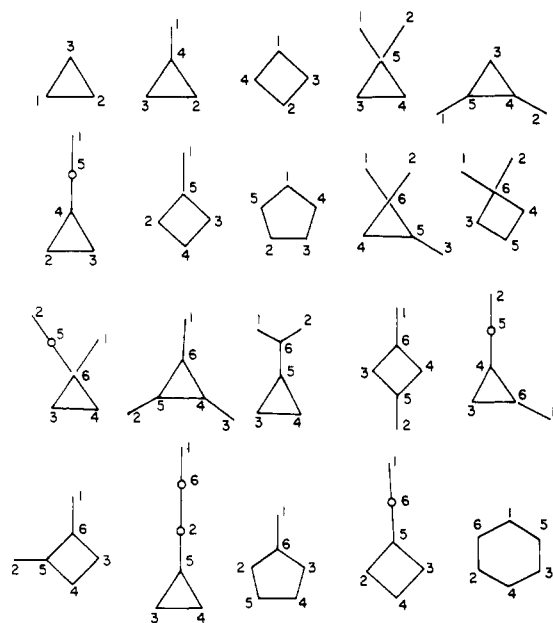


Figure 1. A selection of simple cyclic graphs with six and fewer vertices and associated labeling of vertices producing the smallest binary code for the structures when the adjacency matrix is interpreted as a single binary number by reading its rows from left to right and from top to bottom.

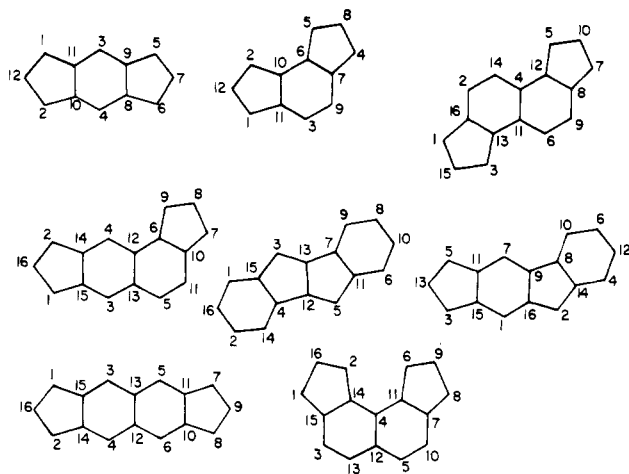


Figure 2. A selection of fused nonalternant systems with the corresponding numbering of atoms which gives the smallest binary code.

smallest label (1) should have as its neighbors the largest labels. Clearly the above requirement on the maximal number of zeros in the first row implies that label 1 can only be associated with vertices of the lowest valency possible. In many situations this will suffice to assign the label uniquely, whereas in others, assignment of adjacent labels having the largest available labels will also be possible. There may be several possibilities compatible with the requirement of a firm assignment of the largest label, or allowing multiple locations for both the smallest and the largest labels (see Figure 1 and 2). This first step discriminates between vertices of different valency and hence reduces the number of possibilities to be examined dramatically for a graph having such vertices. Even in *regular* graphs in which all vertices are of a same valency, and where no apparent assignment of the smallest label is possible at this stage, the total number of possibilities is still not excessive: $(n) \cdot (d!)$, where n is the number of vertices in the graph and d is the degree (valency) of the vertices. This is generally much smaller than $n!$ (for instance, for a cube, it is only 48 compared to $8!$, and for the Petersen graph it is 60 compared to $10!$). In cases where the assignment of the smallest label is not

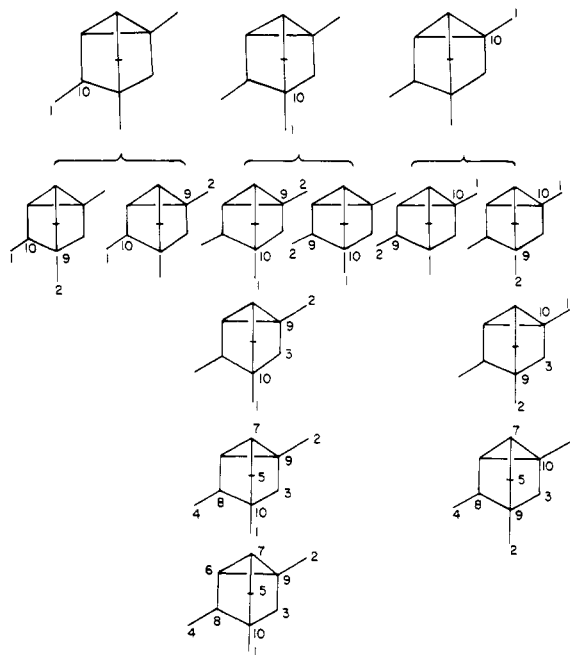


Figure 3. A development of the best numbering for a tricyclic 10-atom skeleton (a derivative of nortricyclene).

unique, one assigns the labels in each of the possible ways, recognizing that the assignments are not final. So, at least formally, label 1 and one or several of the largest labels have been used. The process is now repeated considering the rest of the graph and the remaining labels in the same manner: label 2 is assigned so that the second row of the adjacency matrix, i.e., n_2 , is as small as possible. Hence, one ignores the already assigned sites and examines the remainder of the skeleton for vertices of the lowest valency. If a vertex is adjacent to an already assigned vertex, its *effective* valency, i.e., the number of *new* labels that it requires for a complete assignment of its nearest neighbors, is reduced accordingly. However, among vertices of equal (effective) valency the partially assigned skeleton may dictate the subsequent assignment. With completion of the assignment of the second label, one proceeds in the same way until all vertices have been assigned. Each such step may bring *additional* possibilities to be considered, but equally may *eliminate* some of the previous alternatives from further examination. Generally the procedure is very efficient as the selected examples show.

ILLUSTRATIVE EXAMPLES

Tricyclic Decanes. We shall first consider some tricyclic systems which are easy to determine, yet which provide a sufficient variety of situations encountered in more general applications. In Figure 3, determination of the unique numbering for trimethyltricyclo[2.2.1.0^{2,6}]heptane is illustrated. This is a 10-atom structure with three terminal (methyl) carbons, so initially three distinctive assignments of labels 1 and 10 are possible. Before proceeding to the assignment of the next smallest label, observe that, only in the second and the third cases, the largest label 10 is adjacent to a bridging vertex (having the next smallest valency). So the assignment of labels 2 and 9 to another terminal vertex will yield the optimal result only when labels are associated with the other terminal group adjacent to the bridging vertex. Out of six possible combinations of labels 1 and 2 (and adjacent 10 and 9), only two qualify for further consideration. The next smallest label *will not go to the remaining terminal carbon atom*, as such an assignment requires an additional label for its neighbor (label 8), while if 3 is assigned to the bridge vertex (having the effective valency zero at this stage), a smaller

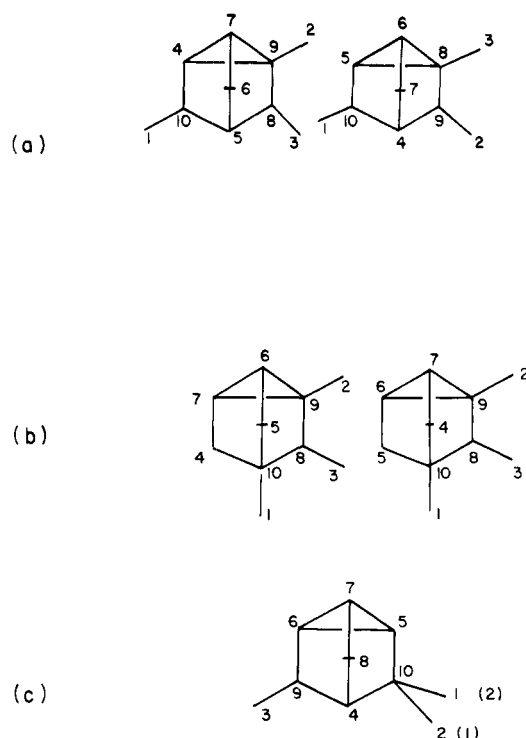


Figure 4. Examples of tricyclic 10-atom skeletons and their numbering: (a) a case of complete assignment which does not resolve which of the structures has the smallest label, but this is easy to verify once the assignments are given. Here left labeling corresponds to the smallest binary code, having a smaller code in the sixth row of adjacency matrix; (b) a case with multiple solution (two structures with different distribution of labels correspond to the smallest binary code) indicating a presence of symmetry; (c) an example in which each step in the assignment is uniquely determined (except for obvious equivalence of vertices 1 and 2).

binary code results. The next step is obvious, the last terminal carbon atom receives label 4 and its adjacent vertex is 8. Continuing with the assignment, the next smallest label 5 has to be attached to the bridge vertex which has not yet been assigned and will have 7 as its remaining neighbor. At this stage, the two possibilities show a difference: on the left (Figure 3), 5 has neighbors 7 and 10, while on the right its neighbors are 7 and 9. The associated binary codes for n_5 (ignoring all preceding zeros) are 1001 and 1010, respectively; only the first alternative can produce the smallest overall binary code for the structure. This completes the assignment for this example since the only unlabeled vertex has to be 6. In this example we have at most considered six possibilities. In fact, with some experience only two would need to be considered instead of 10!

In Figure 4, numberings for several other isomers of the tricyclic $C_{10}H_{16}$ system are shown. In the first example, we end with two complete assignments, and one has to check which of the two represents the solution. This can be done by continuing to examine the immediate environment of the next smallest (assigned) label, and so on until one arrives at a label which in the two numberings have different neighbors. In this example, label 6 has on the left numbering as neighbors 5 and 7, while on the right labeling its neighbors are 5, 7, and 8, showing that the first alternative corresponds to the smallest binary code. It may also happen that two labelings correspond to a same binary code, i.e., lead to an identical adjacency matrix. Such situations point to the presence of equivalent vertices and imply symmetry properties for the graph considered. An example of a multiple solution is also shown in Figure 4. The graph (middle of the figure) as drawn, does not immediately reveal its symmetry, but the final numberings

point to the equivalence of the pair of vertices (4,5) and (6,7). This becomes apparent if one looks at the skeleton as that of substituted nortricyclene. The apparent symmetry of a graph may be used to reduce listing of symmetry equivalent numberings (which can be recovered by applying the symmetry operations of the figure). However, graphs may have symmetry properties which cannot be visualized from a single or a few different diagrammatic representations. The procedure outlined above will generate all the numberings and will enable one to characterize the symmetry properties of the graph considered.

In a search for the numbering, several alternatives have to be pursued until they are discarded as being unacceptable. Efficiency of an algorithm depends on the number of unproductive alternatives which are eliminated from the analysis. The number of possibilities to be examined will vary from case to case. The particular algorithm permits one to attribute to each graph a number indicating the degree of freedom associated with the search of the smallest binary code. It is determined by the number of alternative assignments when a multiple choice is possible. A complete exhaustion of all possibilities involves $n!$ alternatives. The power of an algorithm can be represented by the ratio of all possibilities to the number of those actually processed for a given graph. Different algorithms can then be compared by examining their application on selected graphs. That the particular algorithm described here is very powerful is well illustrated by the last example in Figure 4. The assignment proceeds here without any degree of choice whatsoever, except for an apparent equivalence of vertices 1 and 2. The same intrinsic feature of the algorithm is operative in other instances, not for a graph as a whole, but for some fragments. To appreciate the efficiency of the algorithm, one should not place so much emphasis on the number of alternatives to be examined, but on the rapid elimination of the enormous number of unproductive possibilities. Thus a graph with a single terminal vertex immediately requires the label 1 and label n (n being the number of vertices) to be assigned to the terminal edge, terminal vertex being 1. This assignment excludes any other combination of labels for the terminal edge, the total number of which is $n(n-1)/2$.

Dodecahedron. Let us now consider graphs of relatively high symmetry, those having several or all vertices of the same degree. The presence of several vertices of a same valency does not permit a quick perception of the potentially productive assignments at the initial stage of the search. It is important now to keep records on numerous possibilities for further examination which may make the search tedious. It is possible, however, in such situations also to proceed with relative efficiency and speed if one tentatively does not specify all labels, and refer to some as unknown variables (which can take prescribed integral values only). To illustrate this, consider a Schlegel projection of a regular dodecahedron (Figure 5). There are 20 vertices, each of the same degree (3). We can assign label 1 to any of the 20 vertices, and one particular assignment is shown. Vertices adjacent to 1 must have the largest available labels: 20, 19, and 18, the various permutations of which lead to six possibilities. One of these is shown in Figure 5. Continuing the assignment we should not forget that the considered numbering is only one of 120 possibilities that still need to be examined. If one takes into account the symmetry of the figure (the presence of the fivefold rotational axis), there are only four apparently nonequivalent vertices to be considered for assignment of the label 1. The figure has a plane of symmetry also, so left and right need not be differentiated, which leads to only three nonequivalent assignments for labels 20, 19, and 18. Thus, taking advantage of the apparent symmetry features reduces the search for the numbering to 12 instead of 120 possibilities. Once the

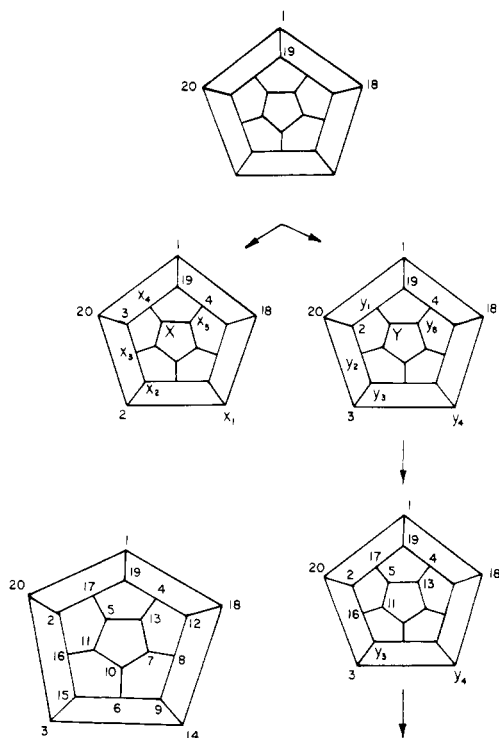


Figure 5. Assignment of labels in dodecahedron is expedited by introduction of unspecified labels x_i and y_i , the values of which are determined at a later stage of the search.

numbering for these 12 situations is found, the remaining numbering can be derived applying symmetry operations of the figure. The symmetry of the graph is actually higher than apparent, and *all* 20 vertices are in fact equivalent (as is obvious since the graph represents a projection of a regular polygon). This permits construction of all numberings of interest from the one shown in Figure 5 by a suitable relabeling of vertices and preserving the adjacency relations. In a more general graph the symmetry may be lacking or not necessarily apparent, and we will proceed as though unaware of equivalence of the vertices. We continue with the assignment of label 2 which, as seen from Figure 5, has two possible sites, both adjacent to the largest available label 20. Its other neighbors should be 17 and 16, each of which has two possible sites. Continuation calls for assignment of label 3, which can be uniquely assigned to the yet unnamed vertex adjacent to 20. The other nearest neighbors of 3 are to be 15 and 14, which again give two alternative possibilities. The next smallest label 4 is again uniquely assigned to the only available site next to vertex 19, the largest unconsumed label, while its neighbors 13 and 12 may take alternative adjacent sites. In this example assignment of each smallest label after 2 was unique, while assignment of the labels adjacent to them (large numbers) doubled the number of possibilities at each step. Instead of keeping track of all these alternatives (many of which may well be discarded at a later stage when found that they do not correspond to the smallest binary code), we can label adjacent vertices as variables x_1, x_2, x_3, \dots and y_1, y_2, y_3, \dots for each of the considered steps and defer the precise assignment of labels, except that labels used should be reserved. For the above example we have:

$$x_1, x_2 = 17 \text{ or } 16 \text{ (same for the pair } y_1, y_2)$$

$$x_3, x_4 = 15 \text{ or } 14 \text{ (same for the pair } y_3, y_4)$$

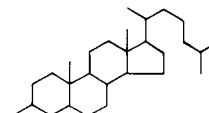
$$x_5, x_6 = 13 \text{ or } 12 \text{ (same for the pair } y_5, y_6)$$

At this stage all assigned vertices (regardless of whether we know the label precisely or not) have had two nonassigned nearest neighbors, except the vertex indicated as X (and Y ,

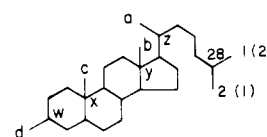
respectively) which has only one nonassigned vertex as the nearest neighbor. Hence, by putting $X = 5$ (and $Y = 5$) we introduce only a *single new* label 11, which makes the fifth row of the adjacency matrix smaller than any other possibility involving *two new* labels, if 5 is to be assigned elsewhere. The neighbors of the fifth vertex are for the left and the right possibilities, respectively: $(x_4, x_5, 11)$ and $(y_1, y_5, 11)$. Since x_4 is smaller than y_1 , the first possibility no longer needs to be considered, because regardless of how we specify the labels, 5 will always have a neighbor which is larger than that for the right assignment. Now for 5 to have the largest neighbors and produce the smallest binary code, y_1 and y_5 must take larger values from the available alternatives, i.e., $y_1 = 17$ and $y_5 = 13$. Then y_2 and y_6 are also determined producing a partial assignment as shown in Figure 5. In the continuation of the search one may ignore the part of the skeleton fully assigned, completion of the assignment is simple, and the result is shown in Figure 5.

Cholesterol. As a final example we consider a relatively large molecule—a steroid skeleton—in order to illustrate that the scheme is not difficult to implement even for sizable structures. We ignore differences among atoms and bond types and merely consider the corresponding molecular graph.

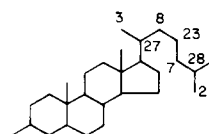
There are 28 atoms in the system, hence in total $28!$ (approximately 3×10^{29}) possible labelings an enormous number to start with:



By examining the skeleton, we find six terminal atoms, and since none of them make a pair with a single bridging vertex in between (as was the case illustrated in Figure 3), we can reserve labels 1–6 and 28–24, in all 11 labels, for the terminal atoms and their nearest neighbors. Labels 1, 2, and 28, however, can be immediately assigned, as shown below. Since *two* terminal carbons are attached to a single branching atom, they both can share the largest available label producing identical smallest possible binary codes for rows 1 and 2. The remaining terminal vertices have been tentatively assigned by letters a – d in an arbitrary mode, while the adjacent branching vertices have labels w, x, y , and z :

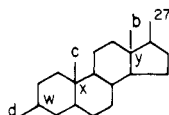


In continuation of the assignment we will disregard the equivalence of atoms 1 and 2. The vertices which are next to be considered for assignment of labels are those of valency 2 (bridging vertices). The smallest *available* label is 7, and it can be definitely assigned to the site next to label 28, having as the other neighbor the largest *available* label 23. With the assignment of label 23 we created a vertex of zero effective valency, the assignment of which does not require *new* labels at this stage. Its neighbors are $(z, 23)$. Hence to make the optimal assignment requires that $z = 27$ (which is the largest *reserved* label available). This completes the assignment of the “tail” of the steroid, since by assigning 27 we have also specified that $a = 3$ as an optimal choice:

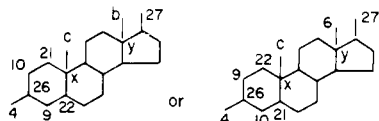


We can now concentrate on the residual structure. Among

the number of remaining bridging vertices, only at w two bridging vertices are adjacent to an assigned vertex:

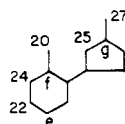


Hence, by assigning the largest available reserved label to w it can be shared by two bridging vertices, ensuring that two vertices of low effective valency will attain the smallest possible corresponding code value. The smallest available labels are 9 and 10, and they will be therefore adjacent to $w = 26$. The other neighbors of 9 and 10 are 22 and 21, respectively. The assignment of $w = 26$ also determines $d = 4$, and so we have the following alternative partial assignments:



It immediately becomes clear that only the left assignment will satisfy the criterion for the desired numbering, since its label 22 can be shared by yet another unassigned vertex of lowest valency, while in the right alternative it is "wasted", being adjacent to already assigned vertices.

An examination of the above partially assigned skeleton still shows several unassigned vertices of degree 2. Of these only two have already one of their neighbors assigned and hence require a *single* additional label. These are the vertices adjacent to y and adjacent to 22, but since y can take larger values (25 or 24), clearly the former is the acceptable solution. This determines the 11th row of the adjacency matrix, 20 being the other neighbor. With this step we also settle assignment of 5, 6, and 24, as the unique solutions for b , c , and x , respectively. The remaining fragment is quite small:



Its completion is not difficult. Among the yet unassigned vertices three require only a single additional label. The largest available number which is now 19 should be adjacent to the smallest available label 12. There are three sites for label 12 to be considered: $e = (22, 19)$, $f = (24, 20, 19)$, and $g = (27, 25, 19)$. Obviously the last alternative provides the smallest binary value. This again confirms that an effective valency, not the factual valency, plays the crucial role in dictating the assignment. Completion of the assignment is now very simple and gives the numbering shown in Figure 6a. The search for the numbering may at first appear lengthy and involved, but it is not difficult even for this relatively large skeleton. One should not overlook that our task consisted of singling out one in almost 10^{30} possibilities.

SOME PROPERTIES OF THE NUMBERING SCHEME

The proposed numbering scheme is intended for several purposes: (1) to facilitate identification of a system and serve as the basis for checking on graph isomorphism; (2) to provide a basis for nomenclature in which numbering of atoms is less arbitrarily decided upon; (3) to provide a linear code for two- or three-dimensional structures; (4) to facilitate a search for hypothetical structures and construct the forms of all structures of prescribed connectivity or prescribed characteristics (size, number or rings, etc); (5) to illuminate some structural features of the particular connectivity, such as presence of equivalent vertices and associated symmetry properties; (6) to assist in

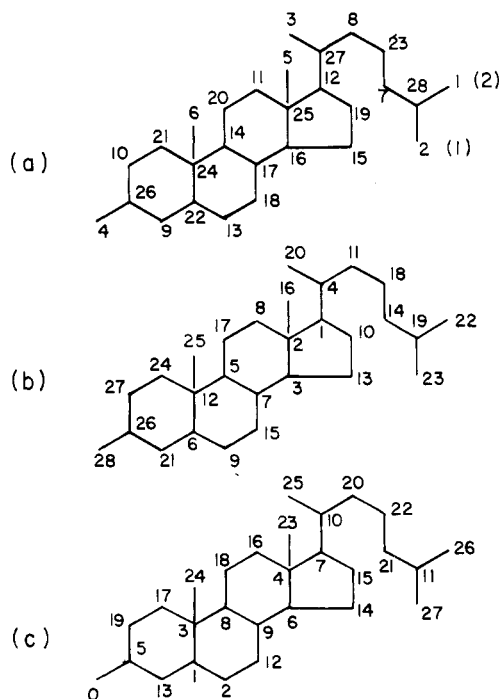
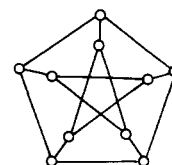


Figure 6. Examples of canonical numbering for skeleton of cholesterol: (a) scheme based on the smallest binary label; (b) numbering derived according to Morgan's algorithm (extended connectivity); and (c) an alternative due to Bersohn and Esack (preferential treatment of functional roles of atoms and groups) [H. L. Morgan, *J. Chem. Doc.*, 5, 107 (1965); M. Bersohn and A. Esack, *Chem. Scr.*, 6, 122 (1974)].

cataloguing and ordering of a set of structures in a sequence which would reflect some inherent qualities of the graphs; and other potential application as recognition of common fragments.

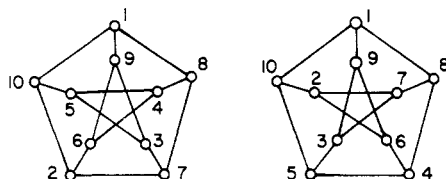
The fact that a unique numbering as such is both possible and practical to derive is sufficient ground for advocating our scheme. The uniqueness relates to the form of the adjacency matrix, while the presence of equivalent vertices leads to alternative numberings, all equally acceptable and interrelated by symmetry considerations. Another asset of our scheme is its notable conceptual simplicity. For those who find this insufficiently compelling, we would like to elaborate on the properties of the numbering scheme briefly mentioned above.

Recognizing Equivalent Vertices. In simple structures frequently it is possible to recognize equivalent vertices merely by inspection. Equivalence of atoms in large polycyclic systems is generally obscured when projections with different degrees of deformation are used. The situation in graphs is even more complex since they generally allow various representations in which equivalent vertices appear in an apparently different environment. There are also graphs for which it is not possible to draw a figure which maximally exposes their symmetry. For instance, the Petersen graph (arising in descriptions of the interrelationships between various isomers having a trigonal-bipyramidal structure) is frequently shown with a pronounced C_5 axis:



It is not apparent from the above representation that *all* vertices are equivalent. By searching for all available numberings, very soon one can discern equivalent vertices, without listing all acceptable solutions. The presence of al-

ternative labels for a vertex immediately indicates equivalence. Consider the final assignment for two alternative solutions of the 120 possibilities:¹⁷



A comparison immediately reveals in a number of instances that a same vertex has in the two alternative assignments different labels, such as pairs (2,5), (3,6), and (4,7). But vertices 2 and 5 (and the same for the pair (4,7)) correspond to an *inner* and an *outer* vertex, and therefore the inner and the outer vertices are *necessarily* equivalent. Notice, that to reach our result does not require all 120 distinctive numberings which this highly symmetrical graph has.

Characterization of the Symmetry. In molecular structures it is not difficult to recognize the symmetry type. This is because of our *prior knowledge of symmetry operations* characterizing the point group. Even for some nonrigid systems¹⁸ and simple graphs,¹⁹ all relevant *symmetry operations* have been recognized. This helps one to derive symmetry properties of such systems. For a general graph, however, the symmetry operations are not so apparent and the existing procedures do not apply. It is known that all permutations of labels which leave the adjacency matrix of a graph invariant form a group. Therefore if we find all distinctive labelings which preserve adjacency of labels, a listing of all symmetry operations of a graph is possible. One such system of labeling corresponds to the smallest binary code, and as we have demonstrated it is possible and practical to derive such numbering. Symmetry operations can then be obtained by considering changes from one particular labeling to another.¹⁷ In particular, it is simple to derive the order of the symmetry group, i.e., the number of distinctive symmetry operations. It equals the number of distinctive labelings preserving the adjacency matrix. One finds this number by keeping a record of the number of alternatives (degrees of freedom) at each successive step in the assignment procedure. For instance in the Petersen graph, in view of the equivalence of all ten vertices, label 1 introduces 10 possibilities. These should be multiplied by 3! for possible permutations of labels 10, 9, 8. Finally, there are two alternative sites for label 2, which makes in all 120 possibilities. Assignment of subsequent labels in this case is unique. In the case of the graph of regular dodecahedron, the first label has 20 possibilities, each to be multiplied again by 3! arising from permuting labels 20, 19, and 18 among themselves. In this case, however, there are no additional degrees of freedom for yet unassigned labels, and hence again the order of the symmetry group is 120.

Ordering of Graphs. The problem of graph isomorphism is related to that of ordering of graphs in a sequence or more general tabular forms. The outcome of an ordering process will depend on criteria adopted, and here there is some room for different approaches. Advantages or disadvantages of various orderings should ultimately be judged from their use in applications.

The ordering of structures is an important step in an intelligent search for correlations between structure and properties. For purposes of documentation it is desirable to have a *complete* ordering, devoid of ambiguities characterizing a *partial* order when two or more systems are not discriminated. The canonical numbering provides a basis for such a complete ordering of structures. No two different structures can have an identical adjacency matrix; if they could, by definition, they represent an identical connectivity. We will

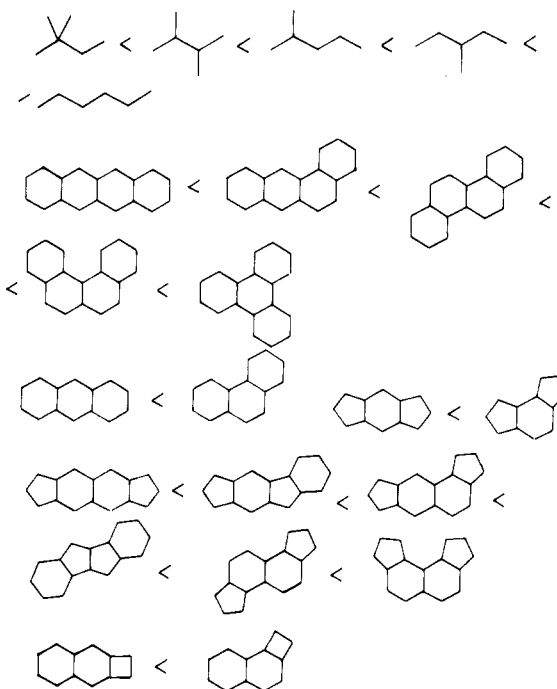
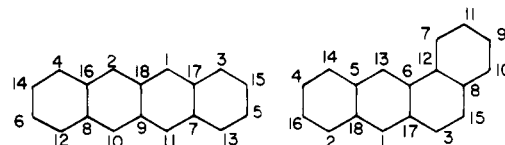


Figure 7. Ordering of various classes of molecular graphs according to the magnitude of the binary label derived from the canonical numbering of vertices.

briefly point out few sequences produced by the numbering scheme discussed here. We grouped structures which have some common features (Figure 7): (a) acyclic structures corresponding to C_6 isomers; (b) fused benzenoid hydrocarbons having four and three benzene rings, respectively; (c) selected nonbenzenoid structures. Each set has been ordered according to the magnitude of the binary code derived from the adjacency matrix. In the first group, which has been discussed at some length elsewhere,²⁰ the ordering reflects the degree of branching of the molecular skeleton; the first member in the sequence is the most branched isomer, and the last is a linear chain. By contrast in the case of benzenoid hydrocarbons, linearly fused benzene rings appear first in the sequence, while angularly fused and "branched" systems appear later in the sequence. In order to see this, we give below the numbering for tetracene and benzantracene:



A comparison of labels show that with label 5 there are differences between tetracene and benzantracene: in the former the neighbors of 5 are (13, 15) while in the latter case the neighbors are (13, 14, 18). Hence the fifth row of the corresponding adjacency matrices differs, the former representing a smaller magnitude. A similar situation appears again among *s*-indacene and *as*-indacene, related higher members of the series, as well as for naphthocyclobutadienes. In some instances such an ordering will parallel the relative molecular stability, aromaticity, and other like properties.

Construction of All Graphs of Prescribed Form. The scheme permits a systematic derivation of all graphs of desired size and valency distribution. The prescribed valency restricts the number of ones in rows of adjacency matrices; the number of vertices gives the size of the matrix. We will illustrate the process on a simple case of construction of all trivalent graphs with six vertices.²¹ In this case we have 6×6 matrices, and clearly the first row in all possible solutions will be:

000111

The second row leads to only two possibilities shown below:

000111 000111

000111 001011

Any other alternative does not correspond to the possible smallest binary code.²¹ When considering the form for the third row, as can be easily verified in this simple example, each of the two possibilities leads to only one acceptable solution. The remaining rows are also uniquely determined so that in this case in all only two distinctive nonisomorphic graphs result:

000111 000111

000111 001011

000111 010101

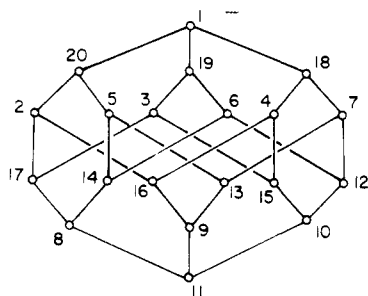
111000 101010

111000 110100

111000 111000

The first adjacency matrix defines the Kuratowski graph $K_{3,3}$, one of the two graphs characterizing nonplanar graphs;^{1,11} the second adjacency matrix defines a graph representing a trigonal prism. In applications to larger graphs care has to be exercised to avoid unnecessary duplication. However, the nature of the scheme permits a simple check for isomorphism, which has been a troublesome problem in some other approaches. The scheme permits a compilation of a complete catalogue which incorporates structures not necessarily similar. For instance, we may lift the restrictions on valencies and construct all graphs having six vertices applying the same procedure in a systematic way.

Properties of Canonical Numberings. The derived numbering in a number of graphs shows regularities which are of interest. For instance, in benzenoid hydrocarbons, shown in Figure 7, and also in other alternants having no pendant bonds, smaller labels are associated with one class of atoms (say starred), and the larger labels belong to the other class (unstarred). As a result the adjacency matrix takes a particular form in which nonzero entries appear only in off-diagonal blocks. Such off-diagonal submatrices summarize all the information in the graph and can be used for an alternative approach to discussion of molecular features.²² The Desargues-Levi graph, the first graph introduced in chemistry which related to molecular *isomerization* rather than to molecular *structure*,²³ is also an alternant graph and shows the same classification of labels among the two kind of vertices present.¹⁷



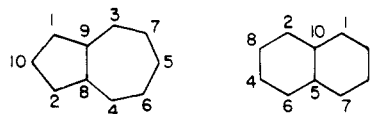
In addition the labels n and $n + 10$ are symmetrically placed with respect to the center of the figure. Since they correspond to enantiomers we see how the canonical numbering preserves, through labeling, the relation between the corresponding structures.

If a system is nonalternant the adjacency matrix corresponding to the canonical numbering will necessarily have nonzero entries also in the diagonal block submatrices. In Figure 8 the situation is illustrated for the case of azulene.

0	0	0	0	0	0	0	0	1	1
0	0	0	0	0	0	0	1	0	1
0	0	0	0	0	0	1	0	1	0
0	0	0	0	0	1	0	1	0	0
0	0	0	0	0	1	1	0	0	N
0	0	0	1	1	0	0	0	0	0
0	0	1	0	1	0	0	0	0	0
0	1	0	1	0	0	0	0	A	0
1	0	1	0	0	0	0	A	0	0
1	1	0	0	N	0	0	0	0	0

Figure 8. The adjacency matrix corresponding to the smallest binary code for (1) azulene ($A = 1, N = 0$); (2) naphthalene ($A = 0, N = 1$); and (3) [10]annulene ($A = 0, N = 0$). The adjacency matrices are symmetrical about the diagonal a_{ij} for $i = j$, so for computer applications only half of the matrix needs to be stored.

Finally, it is instructive to compare labeling in different structures as this may reveal some common features, and even a common fragment. For example consider azulene and naphthalene:



Both systems represent a bridged ten-membered ring. The numbering along the ten-membered ring in both molecules turns out to be identical and is the same as the numbering for a ten-membered ring without any bridges. The two systems differ in the mode of the *bridging*. The sites at which the bridges are inserted are precisely those where the introduction of an extra link in a ten-membered ring will *least* increase the value of the corresponding binary code. A fragment with a common labeling can be found by making a list of vertices and their nearest neighbors. An overlapping in the listing for two molecules would indicate such a fragment. This is illustrated in Figure 9 for a less obvious case of a cube and another caged structure. In this way some common fragments may be detected, but others may not be recognized since in a general case there is not necessarily a correspondence between labels in different graphs. The comparison secures finding the largest fragment with compatible numbering in the two graphs. It has to be seen if this particular aspect can be developed for a more general case. The presence of a large fragment with a same numbering is also of interest for storing the information on such systems. Namely, one can then combine the information on two structures in a single adjacency matrix which will have few variable parameters that require specification only. In Figure 8 the adjacency matrix shown illustrates the situation: by selecting $A = 0, N = 0$, we have the input for [10]annulene; when $A = 1, N = 0$, we have the input for azulene; and when $A = 0, N = 1$, we have input adjacency matrix for naphthalene. Such a combined storage is more economical not only in saving precious memory space but in facilitating certain file searches.

CONCLUDING REMARKS

The current numbering of vertices in a graph can immediately be applied to problems of chemical documentation where the problem of graph isomorphism is important. Equally, the scheme can find its use in computer manipulations of structural data, such as in computer-assisted synthesis and related heuristic schemes, although for such purposes where considerations go beyond the connectivity, the problem of handling of stereoisomerism necessarily appears (and can

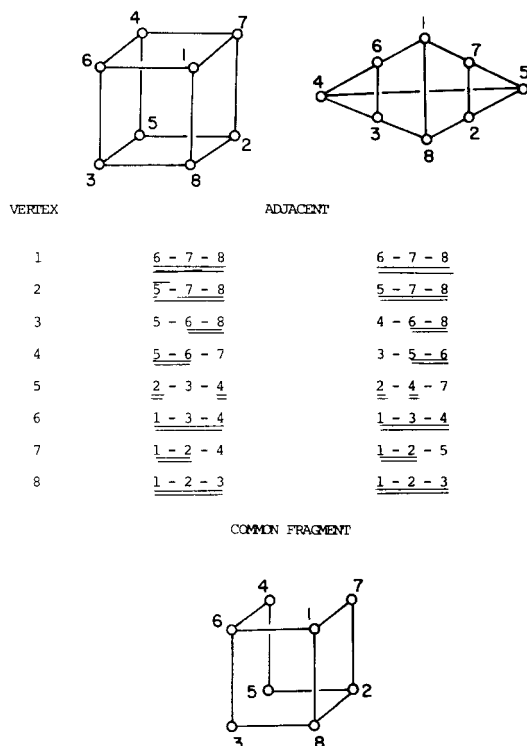


Figure 9. Illustration of a common fragment in two related caged structures. Vertices which have a same label and same environment labeling form a list of the common fragment.

presumably be accomplished along the lines already applied to noncanonical numbering of vertices). For the purpose of chemical nomenclature, a subject which is constantly under revision, existence of unique labels for any structure ensures that no difficulties will arise in the future. If an exotic structural form emerges, which has not been anticipated, and hence need not conform to the existing schemes tailored for prevailing structures, it will not introduce a special consideration.

The apparently awkward distribution of numbers over the whole molecular skeleton may be displeasing to some. What appears to be a convenient form for a computer manipulation is not equally suitable to direct human use, and so the dilemma remains as it is difficult to satisfy two so diverse needs. In evaluating this or some other alternative scheme, one should not deliberate on how it suits small structures, for which many problems of importance, such as isomorphism or fragment recognition, can frequently be quickly resolved intuitively or by trial and error. For large molecular structures these issues become tantalizing and a need for an efficient resolution imperative. Locally grouped labels, such as prescribed by convention, may be more pleasing, but again in large systems at some regions they are bound to introduce abruptness. One should not overlook an interest in locally grouping labels, and perhaps a *compromise* could be worked out by considering an idea of a *dual* labeling system. Namely, once the unique numbering has been found, it is a simple matter to renumber atoms according to some prescribed rules to suit better less general applications. For instance, one may prefer to assign smallest labels to vertices of the highest valency. Since each

class of atoms has already been assigned labels, these can be ordered within each group of atoms of the same valency and then serially renumbered. One may also mention that an alternative numbering based on the largest binary label is possible. Such a scheme would tend to group together labels of approximately the same magnitude; however, for larger systems, such a criterion would also show abrupt changes in some adjacent atoms. It is not immediately clear what is to be gained and what would be lost if such an alternative is pursued, but advantages of the canonical numbering outlined here when contrasted to conventional numberings justify closing the exposition by paraphrasing George Orwell: "All numberings are arbitrary, but some are more arbitrary than others!"

ACKNOWLEDGMENT

The author with to thank Dr. D. H. Rouvray (Mülheim, Germany), Dr. A. Zamora (Columbus, Ohio), and Dr. M. Bersohn (Toronto, Canada) for their interest in the work and useful suggestions which have led to an improved presentation of the material.

REFERENCES AND NOTES

- (1) F. Harary, "Graph Theory", Addison-Wesley, Reading, Mass., 1972.
- (2) L. C. Ray and R. A. Kirch, *Science*, **126**, 814 (1957).
- (3) A. T. Balaban and F. Harary, *J. Chem. Doc.*, **11**, 258 (1971). For discussion of isospectral graphs, see T. Živković, N. Trinajstić, and M. Randić, *Mol. Phys.*, **30**, 517 (1975); M. Randić, N. Trinajstić, and T. Živković, *J. Chem. Soc., Faraday Trans. 2*, **72**, 244 (1976); W. C. Herndon and M. L. Ellzey, Jr., *Tetrahedron*, **31**, 99 (1975), and references in these papers.
- (4) M. Randić, to be submitted for publication.
- (5) E. H. Sussenguth, Jr., *J. Chem. Doc.*, **5**, 36 (1965).
- (6) J. F. Nagle, *J. Math. Phys.*, **7**, 1588 (1966).
- (7) F. Harary in "Combinatorial Theory and its Applications", P. Erdős, A. Renyi, and V. T. Sos, Ed., Janos Bolyai Mathematical Society, Budapest, Hungary, and references mentioned therein.
- (8) M. Randić, *J. Chem. Phys.*, **60**, 3920 (1974).
- (9) A. L. Mackay, *J. Chem. Phys.*, **62**, 308 (1975).
- (10) The existence of local minima and the possibility that pairwise permutations can lead to a dead end is of some interest in contemplations on molecular evolution; see A. L. Mackay, *J. Theor. Biol.*, **54**, 399 (1975).
- (11) Besides ref 1, there are several other available books on graph theory: R. G. Bursaker and T. L. Saaty, "Finite Graphs and Networks", McGraw-Hill, New York, N.Y., 1965; W. T. Tutte, "Connectivity in Graphs", University of Toronto Press, Toronto, Canada, 1966; O. Ore, "Theory of Graphs", American Mathematical Society Colloquium, Publication 38, Providence, R. I. 1962; N. L. Biggs, E. K. Lloyd, and R. J. Wilson, "Graph Theory 1736-1936", Clarendon Press, Oxford, 1976. Most graph theoretical terms are briefly defined in J. W. Essam and M. E. Fisher, *Rev. Mod. Phys.*, **42**, 272 (1970).
- (12) D. H. Rouvray, *RIC Rev.*, **4**, 173 (1971); I. Gutman and N. Trinajstić, *Top. Curr. Chem.*, **42**, 49 (1973); *Croat. Chem. Acta*, **47**, 507 (1975).
- (13) D. Hellwinkel, *Chem.-Ztg.*, **94**, 715 (1970); N. Tanaka, T. Iizuka, and T. Kan, *Chem. Lett.*, 539 (1974).
- (14) B. R. Kowalski, *J. Chem. Inf. Comput. Sci.*, **15**, 201 (1975).
- (15) W. J. Wiswesser, "A Line-Formula Chemical Notation", Crowell, New York, N.Y., 1954; H. W. Hayward, Patent Office Research and Development Report 21, U.S. Patent Office, Washington, D.C.; D. Lefkowitz, *J. Chem. Doc.*, **7**, 186 (1967); P. M. McDonnell and R. F. Pasternack, *ibid.*, **5**, 56 (1965); H. Hiz, *ibid.*, **4**, 173 (1964).
- (16) M. Randić, *J. Chem. Phys.*, **62**, 308 (1975).
- (17) M. Randić, *Chem. Phys. Lett.*, **42**, 283 (1976).
- (18) H. C. Longuet-Higgins, *Mol. Phys.*, **6**, 445 (1963).
- (19) U. Wild, J. Keller, and H. H. Günthard, *Theor. Chim. Acta*, **14**, 384 (1969).
- (20) M. Randić, *J. Am. Chem. Soc.*, **97**, 6609 (1975).
- (21) M. Randić, *Acta Crystallogr.*, in press.
- (22) G. G. Hall, *Int. J. Math. Educ. Sci. Technol.*, **4**, 233 (1973).
- (23) A. T. Balaban, D. Farcasiu, and R. Banica, *Rev. Roum. Chim.*, **11**, 1205 (1966).