Applications of Artificial Intelligence for Chemical Inference. XVI. Computer Generation of Vertex-Graphs and Ring Systems[1,2]

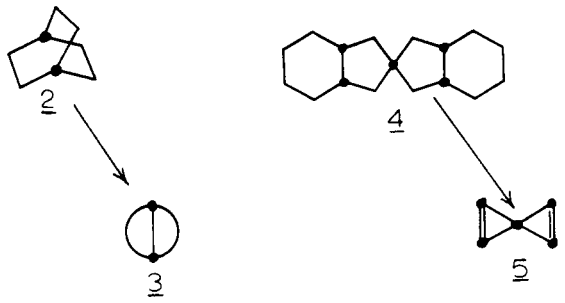
RAYMOND E. CARHART, DENNIS H. SMITH,* HAROLD BROWN, and NATESA S. SRIDHARAN†

The Departments of Computer Science and Chemistry, Stanford University, Stanford, California 94305

Received January 23, 1975

A method for construction of vertex graphs and utilization of these graphs in the construction of ring systems is described. The method allows construction of vertex-graphs containing vertices of any prescribed degree, thus extending previous tabulations and removing the restriction of degree ≤4. Expansions of the vertex-graphs via constructive graph labeling provide exhaustive compilations of possible ring systems. These compilations can also be restricted, using constraints supplied at the discretion of the chemist, to particular types of ring systems.

Vertex-graphs were originally proposed by Lederberg as the foundations upon which cyclic molecules can be constructed.[3] Vertex-graphs consist of the vertices ("nodes") which represent the points of ring fusions in purely cyclic molecules,[4-7] together with the paths ("edges") which interconnect these nodes.[8] Because vertices, or nodes, of degree 2 cannot participate in such ring fusions,[9] they are, by convention, ignored in representations of vertex-graphs. Thus, vertex-graphs contain only vertices of degree 3 or higher. As examples, consider the molecules bicyclo[2.2.2]octane (**2**) and 2,2′-spirobi-2*H*-indene (**4**).



Deletion of the secondary nodes of **2** and **4** and retention of the nodes of degree 3 or 4 (boldface) and the paths which interconnect them yield vertex-graphs **3** and **5**, respectively. Note that the spiro center of **4** is included in the vertex-graph **5**.[4,5] Additional examples of vertex-graphs and descriptions of the method by which vertex-graphs are utilized for exhaustive construction of isomers are available.[6,7,10]

Existing tabulations, however, of vertex-graphs[3,11-14] and the related "multigraphs"[11-14] and "general graphs"[11-14] are complete only for trivalent graphs (all nodes of degree 3) through ten nodes (see below). These tabulations do contain some sets of vertex-graphs possessing nodes of degree 4 (e.g., **5**), but such sets are greatly limited compared to those which can be derived by the procedures discussed in the subsequent section. No tabulation contains nodes of degree greater than 4. More general applications of our structure generator[6,7,10,15] require a more extensive set of such graphs, and not only to accommodate atoms of potential degree (in ring systems, not considering acyclic attachments) greater than 4. An approach to computer-assisted structure elucidation,[15] in which structures can be constructed under constraints, needs to deal with nodes of arbitrary, potentially high, degree. We describe in the subsequent section a method for automatic construction of vertex-graphs which possess nodes of arbitrarily high degree.

Ring systems are a cornerstone of both our conception and our representation of molecules which contain both cyclic and acyclic parts. For example, the Chemical Abstracts Service Registry System is based partly on an initial classification of molecules by ring systems, i.e., molecules stripped of acyclic components.[16] Important synthetic and analytical problems depend, for their successful solution, on specification of exhaustive lists of related, usually isomeric, ring systems. For example, both manual and computer-based methods have been applied recently to a class of such problems, interconversion of isomers under constraints derived from knowledge of reaction mechanism and chemical stability. Whitlock and Siefken[17] have explored the rearrangement pathways to adamantane, and Engler et al.[18] have analyzed the energetics of these potential pathways. Collins and Johnson[19] have studied rearrangements of bicyclo[2.2.1]heptyl carbocations under different assumptions as to the allowed rearrangement pathways. The manual method of Whitlock and Siefken has recently been emulated in a computer program applied to the analysis of diamantane formation from various pentacyclotetradecanes.[20] These methods[19,20] do not construct isomers whose structures are inaccessible by the allowed rearrangement pathways.

We describe in the subsequent section a method for exhaustive generation of ring systems, or generation under constraints, with prospective avoidance of duplicates. This method, which utilizes constructive graph labeling[10] of either the vertex-graphs or cyclic skeletons[7] derived therefrom, can specify a variety of types of ring systems, including not only those which are isomeric carbocyclic skeletons, but also isomeric skeletons comprised of any number of atoms of any atom type. This method can be used to construct structures (nonisomeric) which share the same basic ring system but differ in the numbers of various types of atoms of which the ring system is composed.

METHOD

Construction of Vertex-Graphs. Balaban[11-14] has suggested methods for construction of graphs which can contain nodes of degree 2 through 4 and also single bond bridges and possibly loops.

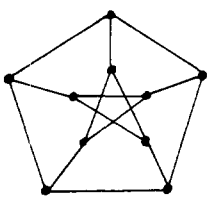
We, however, seek the set of vertex-graphs which conform to Lederberg's original definition.[3] These graphs are precisely those graphs which are: (1) "bridge"-free, i.e., cannot be cut into two parts by scission of an edge[21] (biphenyl, for example, contains a "bridge," the bond connecting the two phenyl rings; this is the graph theoretical

* Author to whom correspondence should be addressed.
† Present address: Department of Computer Science, Rutgers University, New Brunswick, N.J. 08901.

definition of "bridge" and will hereafter be enclosed in quotes to distinguish it from the word bridge as used in chemistry with respect to bridged ring systems), (2) loop-free, and (3) contain no nodes of degree 2 or less. These graphs are sufficient for our needs because the structure generator has the ability to construct "bridged" structures from "bridge"-free components and to add loops together with bivalent nodes where necessary.[4-7,10]

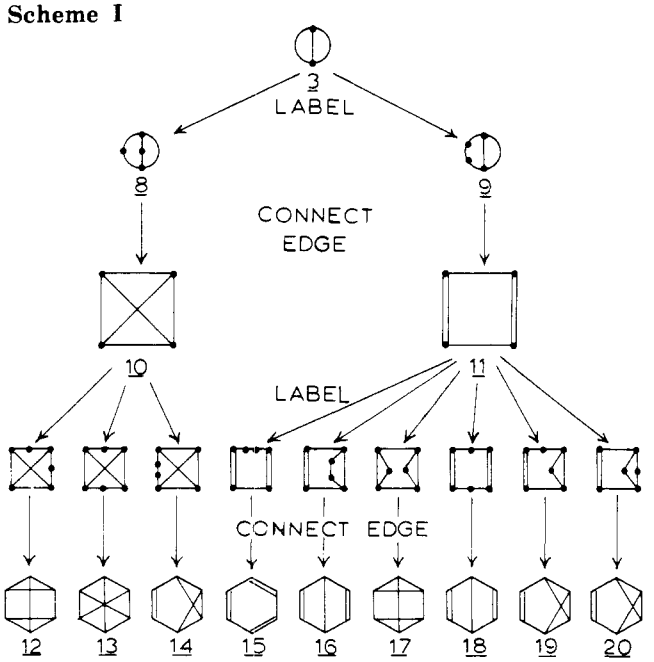
The method described below is implemented in a computer program (see Experimental Section) so that vertex-graphs are constructed automatically.

Construction of Trivalent Vertex-Graphs. The method for construction of graphs possessing nodes of degree greater than 3 depends on having a complete set of trivalent vertex-graphs[3] (all nodes of degree 3). Because questions have been raised concerning the completeness of Lederberg's original tabulation (e.g., Balaban[7] has pointed out that the Peterson graph, **7**, was not included in the tabulation[3] of the trivalent graphs of ten nodes), we first constructed the set of trivalent graphs, starting with the single trivalent graph of two nodes, **3**.

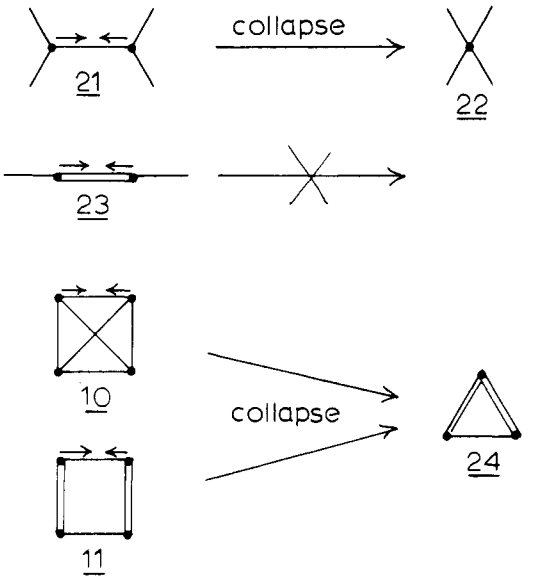


Construction of the set of trivalent vertex-graphs containing $n + 2$ nodes proceeds from the set of trivalent vertex-graphs of n nodes (n must be an even integer because all nodes are of odd degree) by using the labeling algorithm[10] to place two secondary nodes onto the edges of each n node vertex-graph in all topologically distinct ways. These new nodes are then connected with an additional edge (or bond, if you wish to view the nodes as atoms). Duplicate graphs are not totally avoided by this procedure. The labeling procedure provides that no duplicate graphs of $n + 2$ nodes will arise from labeling the graphs of n nodes. But when the new edges are added, duplicates can arise. The procedure is illustrated in Scheme I, which traces the method from graph **3** through the construction of the trivalent graphs of six nodes **12–16**.

Scheme I



Scheme II



There are two unique ways of labeling[10] graph **3** with two secondary nodes, yielding **8** and **9** (Scheme I). Connection of these nodes with an additional edge yields the complete set of trivalent graphs of four nodes, **10** and **11**. Labeling of **10** and **11** with two additional secondary nodes yields three results for **10** and six results for **11**. Connection of the pair of secondary nodes with an additional edge yields **12–14**, derived from **10**, and **15–20**, derived from **11**. Note that **16** ≡ **18**, and **19** ≡ **20**, even though the pairs of duplicates arise from unique *labelings* of **11**. Addition of the new edge produced duplicates. Comparing graphs derived from **10** and **11**, **12** ≡ **17** and **14** ≡ **19** ≡ **20**. Nonisomorphic graphs **12–16** constitute the set of five trivalent vertex-graphs of six nodes. Duplicates are removed automatically subsequent to the construction process.[22-24]

The method outlined in Scheme I was used to construct trivalent vertex-graphs of 4, 6, 8, 10, and 12 nodes. Results are discussed in the subsequent section. A proof of the exhaustiveness of the method is given in Appendix A.

Construction of Vertex-Graphs Containing Nodes of Degree ≥4. A modified method of "node collapse"[22] was used to construct vertex-graphs possessing nodes of degree ≥4 in a stepwise manner beginning with the set of trivalent graphs constructed above. The method of node collapse is exemplified in Scheme II.

The pair of tertiary nodes (**21**, Scheme II) upon removal of the edge between the nodes followed by a merging of them ("node-collapse") yields a single quaternary node, **22**. The pair of nodes to be collapsed must be adjacent. Multiply connected nodes, e.g., **23**, are not collapsed as the result is meaningless to this method. Thus, the method requires identification of "legal" pairs of nodes which can be collapsed, where legality is defined by the following criteria: (1) the nodes must have the proper degree, e.g., both must be tertiary to yield a quaternary node (**21** → **22**); (2) the nodes must be adjacent; (3) the nodes must be singly connected; and (4) for efficiency, symmetrically equivalent pairs of nodes should not be collapsed as this would yield duplication. Again, the labeling algorithm[10] serves nicely to help satisfy these criteria.

For each member of the appropriate set of vertex-graphs, edges (implicitly satisfying criterion 2) which satisfy criteria 1 and 3 are chosen. A single edge is chosen from this set in all unique ways using the labeling algorithm. This satisfies criterion 4. Every labeling then suffers collapse of the nodes at the endpoints of the chosen edge to yield the desired set of graphs. For example, vertex-graphs possessing two tertiary and one quaternary nodes, [2 1][25] graphs, are

Table I. The Number of Trivalent Vertex-Graphs of Two through Twelve Nodes

Nodes (or vertices)	No. of graphs[a]	Nodes (or vertices)	No. of graphs[a]
2	1*	8	16*
4	2*	10	66*
6	5*	12	365

[a] The numbers which are flagged with an asterisk have been verified by comparison with ref 3 (except for **7**) and 11.

derived from collapsing a pair of tertiary nodes in the trivalent graphs of four nodes (**10** and **11**, Scheme II). Of course, duplication is not completely avoided as, for example, the single representative of [2 1] graphs, **24**, is constructed from both **10** and **11**. Note, however, that use of the labeling algorithm and edge selection to satisfy criteria 1–4, above, drastically reduces the problem of duplication. Thus, prior to collapse, the program has identified that there is only one unique pair of nodes to collapse in **10**, and also only one in **11**. Duplicates which remain are removed using the procedure outlined in ref 23.

This method is implemented in a completely general way, permitting pairwise identification, labeling, and collapse of any adjacent pair of nodes to yield a new set of graphs. Collapse of two adjacent nodes, one of degree m, the other of degree p (m need not equal p), yields a single node of degree $m + p - 2$. For example, from the [12] vertex-graphs we obtain the [10 1] set. Then, we can collapse unique tertiary/quaternary pairs in each representative of [10 1] to obtain the set of [9 0 1], then pairs of trivalent nodes to obtain the sequence [9 0 1] → [7 1 1] → [5 2 1], and so forth. Alternatively, [5 2 1] can be obtained by the path of generation [10 1] → [8 2] → [6 3] → [5 2 1]. Duplication can, in some cases, be kept to a minimum by simultaneous collapse of several pairs of nodes, the corresponding edges having been selected by the labeling algorithm.[22] However, the results presented subsequently in Table II were determined using the method described above.

In this way we constructed all possible sets of vertex-graphs which can be derived from trivalent vertex-graphs possessing up to 12 nodes. A proof that this method is exhaustive and independent of the path of generation is given in Appendix B. Results are summarized in the subsequent section.

Construction of Ring Systems. Ring systems are systems which possess nodes of degree 2 and higher, but do not contain "bridges" or acyclic parts. Thus ring systems (devoid for the moment of atom names) differ from vertex-graphs only in that the former may contain any number of nodes of degree 2 and may possess the types of spiro centers discussed in ref 4. The structure generator[7] contains the requisite functions for expansion of the vertex-graphs by constructive graph labeling of secondary nodes on the edges of vertex-graphs and by addition of such spiro centers. In this way, ring systems based on different sets of vertex-graphs and possessing different numbers of secondary nodes can be constructed exhaustively or under constraints.[15] Duplicates are avoided prospectively[7,10] so that testing for duplicates (as was required for construction of vertex-graphs) is not necessary.

If one assumes that all nodes in such ring systems are carbon atoms, then the ring systems constructed as just outlined above (with no nodes of degree >4) represent carbocyclic skeletons. If, however, one desires ring systems comprised of any given variety of heteroatoms, then again constructive graph labeling with atom names, this time applied to the nodes of the graphs rather than the edges, yields the desired results. The subsequent section outlines several applications of this method to construction of various types of ring systems.

RESULTS AND DISCUSSION

Vertex-Graphs. The numbers of trivalent vertex-graphs are given in Table I and the numbers of vertex-graphs possessing nodes of higher degree are given in Table II.[26] The results for trivalent graphs confirm the existing tabulations[3,11] of trivalent graphs through [10], considering also the Peterson graph (**7**) as a supplement to ref 3. However, we obtain 365 [12] graphs, rather than 360 as reported by Lederberg[3] and 362 reported by Balaban.[13] The three missing graphs are additional nonplanar,[3,13] or gauche, graphs.

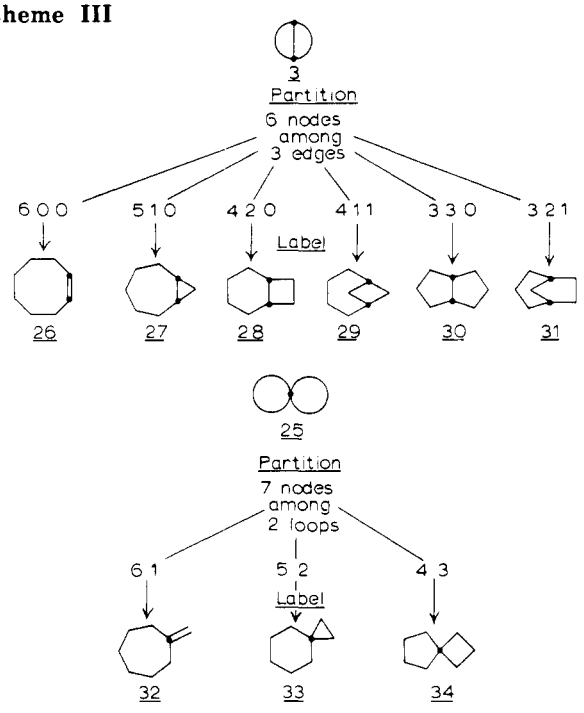
Where our results can be compared to those of Balaban[14]

Table II. The Number of Vertex-Graphs Which Possess at Least One Node (Vertex) of Degree ≥4

Vertex-graph[25]	No.[a]	Vertex-graph	No.	Vertex-graph	No.	Vertex graph	No.	Vertex-graph	No.
[2 1]	1*	[3 1 1]	15	[1 2 1 1]	73	[0 1 0 2]	1	[0 2 0 0 0 1]	1
[4 1]	5*	[3 2 1]	128	[2 0 0 1]	1	[0 2 0 2]	8	[0 3 0 0 0 1]	2
[6 1]	26	[3 3 1]	1024	[2 1 0 1]	4	[0 0 0 3]	1	[0 1 0 1 0 1]	1
[8 1]	180	[5 0 1]	12	[2 2 0 1]	28	[1 1 0 0 1]	1	[1 0 1 0 0 1]	1
[10 1]	1491	[5 1 1]	178	[2 3 0 1]	178	[1 2 0 0 1]	4	[1 1 1 0 0 1]	6
[0 2]	1*	[5 2 1]	2473	[2 0 2 1]	39	[1 3 0 0 1]	21	[1 0 0 0 1 1]	1
[2 2]	4*	[7 0 1]	81	[3 0 1 1]	17	[1 0 2 0 1]	6	[2 1 0 0 0 1]	2
[4 2]	34	[7 1 1]	2167	[3 1 1 1]	291	[1 0 0 1 1]	1	[2 2 0 0 0 1]	18
[6 2]	297	[9 0 1]	665	[4 0 0 1]	5	[1 1 0 1 1]	9	[2 0 0 1 0 1]	4
[8 2]	3143	[0 0 2]	1	[4 1 0 1]	52	[2 0 1 0 1]	4	[3 0 1 0 0 1]	13
[0 3]	1*	[0 1 2]	1	[4 2 0 1]	604	[2 1 1 0 1]	50	[4 0 0 0 0 1]	3
[2 3]	12*	[0 2 2]	7	[5 0 1 1]	238	[3 0 0 1 1]	17	[4 1 0 0 0 1]	37
[4 3]	186	[0 3 2]	25	[6 0 0 1]	34	[3 0 0 0 1]	1	[6 0 0 0 0 1]	25
[6 3]	2902	[2 0 2]	5	[6 1 0 1]	669	[3 1 0 0 1]	12	[0 0 2 0 0 1]	1
[0 4]	3*	[2 1 2]	33	[8 0 0 1]	259	[3 2 0 0 1]	113	[0 0 0 0 0 2]	1
[2 4]	54	[2 2 2]	324	[0 2 0 1]	1	[4 0 1 0 1]	59	[0 1 1 0 0 0 1]	1
[4 4]	1318	[4 0 2]	45	[0 3 0 1]	3	[5 0 0 0 1]	10	[1 2 0 0 0 0 1]	2
[0 5]	6*	[4 1 2]	717	[0 4 0 1]	12	[5 1 0 0 1]	168	[1 0 0 1 0 0 1]	1
[2 5]	232	[6 0 2]	471	[0 0 2 1]	1	[7 0 0 0 1]	82	[2 0 1 0 0 0 1]	2
[0 6]	19	[1 0 3]	3	[0 1 2 1]	8	[0 1 1 0 1]	1	[3 0 0 0 0 0 1]	1
[1 1 1]	1	[1 1 3]	31	[0 0 0 2]	1	[0 2 1 0 1]	6	[3 1 0 0 0 0 1]	6
[1 2 1]	5	[3 0 3]	73	[1 0 1 2]	6	[0 0 1 1 1]	1	[5 0 0 0 0 0 1]	6
[1 3 1]	23	[0 0 4]	4	[2 0 0 2]	5	[0 0 0 0 2]	1	[2 1 0 0 0 0 0 1]	1
[1 4 1]	123	[1 0 1 1]	1	[2 1 0 2]	35	[0 1 0 0 2]	1	[4 0 0 0 0 0 0 1]	1
[3 0 1]	2	[1 1 1 1]	9	[4 0 0 2]	51	[2 0 0 0 2]	5		

[a] An asterisk indicates agreement with Balaban.[11]

Scheme III



for vertex-graphs containing nodes of degree 4, we are in complete agreement as to the total number of vertex-graphs and identity of each graph. This, however, is only a small subset of the total results, Table II. In some cases, we have constructed a given set of vertex-graphs by using different paths of generation. For example, [2 1 0 2] can be derived from [4 0 0 2] by collapse of two tertiary nodes, or from [2 3 0 1] by collapse of two quaternary nodes. Each such case has resulted in construction of the same set of graphs. For several sets of graphs possessing a single node of maximum degree we have verified the identities of the graphs using an independent generation scheme (not presented in this paper) and have found agreement in all cases. We are reasonably confident that the program used accurately reflects the proven algorithms (see Appendixes A and B).

The vertex-graphs, by themselves, do not serve to reveal much new chemistry. Balaban[11-14] has indicated the correspondence of some of his constructs to valence isomers, and the correspondence of trivalent graphs to [*n*]annulenes.[11] The graphs containing nodes of higher valence as such do not represent chemical classes of compounds as interesting as the annulenes, assuming some association of atom names with nodes of appropriate valence. These graphs are primarily useful for constructive expansion to ring systems possessing some number of secondary atoms.

Ring Systems. Vertex-graphs possessing nodes of degree greater than 4 are used in constructive procedures[27] for structure generation employing substructures possessing arbitrary free valence.[15] The number of organic, covalently bonded compounds possessing elements of valence greater than 4 is small compared to the total number of compounds. The number where such an element possesses a *degree* greater than 4 is smaller still. For example, phosphorus is frequently formally pentavalent, but many organophosphorus derivatives are compounds with only three or four (non-hydrogen atom) bonds to phosphorus. Ring systems based in part on nodes of valence greater than 4 have analogs in structures of inorganic or organometallic compounds, for example, cobalt(III) acetylacetonate, iron(II) complexes, and so forth.

We wish, however, in the remainder of this report, to focus attention on ring systems composed of more common elements, which involve almost exclusively ring systems based on vertex-graphs containing nodes of degree 3 and 4.

Table III. The Number of Bicyclic Ring Systems[a] for Three through Ten Carbon Atoms

No. of C atoms	Edge-fused	Bridged	Total[b]	Spiro
3	0	0	0	0
4	1	0	1	0
5	1	1	2	1
6	2	1	3	1
7	2	2	4	2
8	3	3	6	2
9	3	4	7	3
10	4	5	9	3

[a] Excluding structures with multiple bonds. [b] Excluding spiro forms and structures with multiple bonds.

Carbocyclic Ring Systems. Carbocyclic ring systems are constructed by labeling edges of the vertex-graphs of proper degree of unsaturation with the desired number of secondary nodes or by labeling vertex-graphs of lower unsaturation with both loops (which become spiro rings) and bivalent nodes. The only complication in this procedure is that multiple bonds are considered by the program as small rings (which results implicitly from our definition of a ring system[21]). Chemists would not regard cyclohexene as a bicyclic system, but the program does. Exclusion of undesired structures can now be accomplished[15] by constraining the structure generation process.

As an example, consider the problem of specification of possible bicyclic ring systems. Graph **3** and the "daisy" (**25**), a special case graph, are the only graphs with degree of unsaturation equal to 2 which are considered by the program in construction of bicyclics. The former graph (**3**) yields what are usually referred to as bicyclic systems; graph **25** yields spiro systems. The procedure is outlined in Scheme III for bicyclic systems containing eight carbon atoms. Each possible partition of nodes among edges yields only one labeling in the cases of both **3** and **25**, as all edges of **3** are equivalent, as are all edges of **25**.[7,10] If the nodes are only of one type, carbon, then **26–34** can be regarded as the desired carbocyclic systems. Thus from **3** we obtain cyclooctene, **26**, three edge-fused systems **27**, **28** and **30**, and two bridged systems, **29** and **31**. The system of constraints allows us to classify these systems automatically. From **25** we obtain **32**, with an exocyclic double bond, and the two spiro systems **33** and **34**.

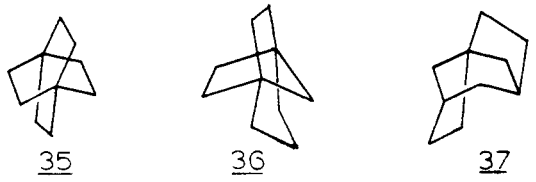
The number of bicyclic ring systems for various numbers of carbon atoms, excluding multiple bonds, are given in Table III. Tricyclic ring systems comprised of carbon atoms are constructed in a similar fashion, using catalog entries [0 2], [2 1], [4 0] and **3** plus an additional loop containing bivalents or a single ring plus two loops with bivalents. Recently, a manual approach to the specification of tricyclic ring systems has appeared.[28] This approach uses a method based on bivalent-free graphs including those which contain "bridges" and loops. A series of rules is included which effectively allows manual graph labeling to be carried out to expand the graphs into complete structures. This approach is, however, restricted to tricyclic ring systems. Our results (see below) are in agreement with the limited example presented.[28] The number of tricyclic ring systems for up to ten carbon atoms is presented in Table IV. Again, systems with multiple bonds are not included. The tabulation is broken down into one category which includes edge-fused and/or bridged systems, and a second category of spiro forms.

The results for C_{10} are particularly interesting, as this group of ring systems contains adamantane and related systems which have been the topics of considerable earlier work.[17,18,28] Using the constraints mentioned by Whitlock and Siefken, i.e., no multiple bonds and no three- or four-membered rings,[17] we obtain in addition to the 16 structures reported[17] five additional structures, **35–39**. **35** and **36**

Table IV. Possible Tricyclic Ring Systems with up to Ten Carbon Atoms

No. of C atoms	Bridged and/or edge-fused	Spiro[a]	Total
3	0	0	0
4	1	0	1
5	3	0	3
6	9	1	10
7	20	5	25
8	40	13	53
9	70	27	97
10	121	49	170

[a] The spiro category includes tricyclic ring systems with one or two spiro fusions.



35 36 37

38 39

are derived from the single vertex-graph of two quaternary nodes, **40,** by labeling the four edges of **40** with 2,2,2,2 and 3,2,2,1 secondary nodes, respectively (other labelings yield

40

multiple bonds or three- or four-membered rings). Structures **37–39,** together with structures **3** and **4** of ref 17, make up the set of five graphs derived from the vertex-graph of one quadrivalent node and two trivalent nodes (**24**). These extra structures represent added complexities to the interconversion maps presented previously.[17,18] Structures **35–39** appear to be highly strained systems. **37–39,** however, do not seem less plausible than other isomers (structures **3** and **4** of ref 17) involving a quaternary center, which, interestingly, have strain energies (see structures **6** and **8** of ref 18) comparable to other more "reasonable" structures. **35** and **36** do not appear to be known compounds. (They can be viewed as homologs of members of the propellane series, specifically of [3.2.1]propellane[29] and [2.2.2]propellane.[30]) The presence of an additional two-carbon bridge in place of the single bond bridge in the last two compounds may result in structures with decreased stability.

Higher *n*-cyclic ring systems are constructed in a similar way. Each increase in degree of unsaturation and in number of atoms, of course, results in a greater variety of vertex-graphs which must be expanded. Thus, the number of possible ring systems increases extremely rapidly. For example, 5291 tetracyclic, C_{12} ring systems have been reported,[31] and an estimated 40,000 pentacyclic, C_{14} ring systems exist.[20]

Heterocyclic Ring Systems. There are two avenues open to construction of heterocyclic ring systems depending on the particular classes of ring systems one desires. Either one wants all (or all, under some constraints) ring systems which can be constructed from a given number of carbons and/or heteroatoms or one wishes to examine ring systems based on a given skeleton, *i.e.*, ring systems formed by substituting a given set of atom names in all ways on the nodes of a given ring skeleton. The former can be accomplished using the structure generator; the latter is a node labeling problem.[10] A brief example of both methods of construction is given below to indicate the scope of possible heterocyclic ring systems.

Table V. Some Examples of the Number of Ten-Atom Bicyclic Ring Systems[a]

Composition	No. of ring systems	Composition	No. of ring systems
C_{10}	9	$C_3N_1O_1$	274 (excluding *N*-oxides)
C_9O_1	36		
C_8O_2	124 (98 excluding peroxides)	$C_8S_1O_1$	212
		$C_8S_1N_1$	274
C_9N_1	45	C_9Si_1	45
C_8N_2	195	C_8Si_2	195

[a] Excluding spiro forms and systems with multiple bonds.

Table VI. Examples of the Number of Heterocyclic Ring Systems Based on the Decalin Skeleton (**41**) and the Naphthalene Skeleton (**42**)

Set of Atom Labels	Number of Ring Systems
Decalin	
9 C's, 1 O	2
8 C's, 2 O's	10 (8 excluding peroxides)
8 C's, 1 S, 1 O	14
7 C's, 2 S's, 1 O	42
Naphthalene	
7 C's, 3 N's	14
6 C's, 4 N's	22
8 C's, 2 Si's	15
7 C's, 3 Si's	32
6 C's, 4 Si's	60

The number of bicyclic ring systems involving heteroatoms is presented in Table V. This tabulation is for ten-atom ring systems and includes only multiple bond-free systems which contain no spiro fusions. The number of possible ring systems possessing heteroatoms increases rapidly with increasing numbers and degree of heteroatoms.[32]

The results in Table V were produced by constructing ring systems as outlined in Scheme III. In this case, however, the skeletons resulting from edge labeling (*e.g.*, **26–31,** Scheme III) must be further labeled with atom names. With only one atom type, e.g., carbon, only one structure results from this labeling as discussed previously. With a variety of atom types, several structures may result from the labeling for each of the skeletons **26–31.** Bivalent atoms (O, S) may occupy any node position of degree 2; trivalent atoms (N) may occupy any node position of degree 2 or 3, and so forth.

Examples of the number of ring systems based on a given ring skeleton are given in Table VI for heterocycles based on the decalin (**41**) or naphthalene (**42**) skeleton. These results were produced by labeling the nodes in the appropriate skeleton with the indicated set of labels (atom names), then counting the results. Other computational methods exist for determining the numbers, but not the identities of each structure.[33] The differences in results for cases such as labeling of naphthalene with 7 C's and 3 N's vs. 7 C's and 3 Si's (14 vs. 32, Table VI) result because a tetravalent silicon atom can occupy all node positions including the quaternary ring junctions, whereas trivalent nitrogen cannot.

41 42

CONCLUSIONS

We have briefly outlined a method and some results for

construction of vertex-graphs and ring systems. This method can be used to help define the scope of such sets of graphs and ring systems in organic chemistry. One could use the vertex-graphs and labelings thereof as a means of classification of ring systems which might have certain advantages over other methods. Although we have no plans to do this ourselves, we are willing to discuss these aspects with interested persons.

The program used in this study is available to a collaborative community of users via a nationwide computer network. For information about gaining access to the program, write to the authors or to Professor Joshua Lederberg, SUMEX Project, Department of Genetics, Stanford Medical School, Stanford University, Stanford, California 94305.

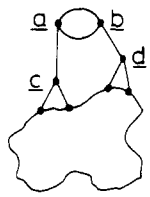
APPENDIX A

We show here that every trivalent vertex-graph **V** with $n > 2$ nodes arises by our method of adding two secondary nodes and an edge between them in all distinct ways to the trivalent vertex-graphs with $n - 2$ nodes. We do this by proving that there is at least one edge $e = (a,b)$ in **V** such that deleting e and "erasing" the resultant bivalent nodes a and b yield a trivalent vertex-graph **V'** with $n - 2$ nodes, and, hence, **V** arises by applying our method to **V'**.

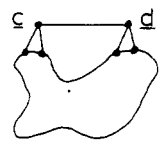
The proof is based on a theorem due to Halin.[34] *Theorem* (Halin, 1969): Every finite, simple, m-connected graph G either contains a node of valence m or at least one edge e such that the graph obtained from G by the deletion of e remains m-connected.

Here, a graph **G** is called simple if it has no loops and no multiple edges and m-connected if there are at least m node-disjoint paths between any two distinct vertices in **G.** (A standard reference for basic graph-theoretical terminology and results is Berge's "Graphs and Hypergraphs."[35])

Because Halin's theorem deals only with simple graphs, we must first consider the cases of trivalent vertex-graphs containing multiple (double) edges. If a trivalent vertex-graph **V,** with $n > 2$ nodes, has a double edge between two nodes a and b (**43**), then the deletion of one of these edges

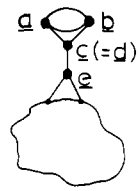


43



44

and the erasure of a and b produces a loop-free, "bridge"-free graph (vertex-graph) **V'** of $n - 2$ trivalent nodes (**44**). Indeed, the newly created edge (c,d) of **V'** cannot be a loop, for this would imply $c = d$. If this were true, then **V** would contain the subgraph **45,** and the edge (c,e) would be a



45

"bridge" violating the initial assumption that **V** is a vertex-graph. Neither can (c,d) of **V'** (**44**) be a "bridge" for this would imply that both (a,c) and (b,d) are "bridges" in **V** (**43**). All other edges in **V'** other than (c,d) are also edges in **V,** and are thus neither loops nor "bridges."

We now consider the case where **V** is a simple trivalent vertex-graph, and note that the edge deletion–node erasure procedure can yield no loops (the creation of a loop requires the presence of one of the two subgraphs **43** or **45,** both of which contain multiple bonds).

We observe that a connected, trivalent graph is 2-connected, if and only if it has no "bridges." This observation follows immediately from the characterization of a 2-connected graph as a connected graph with no cut nodes. Thus, if **V** is a simple, trivalent vertex-graph, the hypothesis of Halin's theorem is satisfied, and, since **V** has no bivalent nodes, it must have at least one edge with the desired property.

Hence, all trivalent vertex-graph with n nodes will be produced by applying our method to the trivalent vertex-graphs with $n - 2$ nodes.

APPENDIX B

The following theorem shows that given two integers m_1 and m_2 with $m_2 \geq m_1 > 2$, every $[n_3, n_4, \ldots, n_k]$ vertex-graph with $k \geq m_1 + m_2 - 2$ and $n_{m_1+m_2-2} > 0$, arises by applying our method of node collapsing to the $[n_3, \ldots, n_{m_1} + 1, \ldots, n_{m_2} + 1, \ldots, n_{m_1+m_2-2} - 1, \ldots, n_k]$ vertex-graphs. Recall that a $[n_3, n_4, \ldots, n_k]$ vertex-graph is a vertex-graph with n_i nodes of valence i, $3 \leq i \leq k$.

Theorem. Let **V** be a vertex-graph and x a node of **V** of valence $m \geq 4$, and let m_1 and m_2 be integers such that $m_2 \geq m_1 > 2$ and $m_1 + m_2 - 2 = m$. Then there is a partition of the edges of **V** with endpoint x into two classes A_1 and A_2 with $|A_1| = m_1 - 1$ and $|A_2| = m_2 - 1$ such that the graph **V',** obtained from **V** by replacing x by an edge $r = (a_1,a_2)$ with the edges in A_1 connected to a_1 and those in A_2 connected to a_2, is also a vertex-graph.

Proof. For any partition A_1,A_2 such that both A_1 and A_2 contain at least two edges, the graph **V'** is connected, loop-free, and has no nodes of valence less than 3. Hence, we need only show that for some partition of this type, **V'** is "bridge"-free. Note that a connected graph is "bridge"-free if and only if every edge lies on an elementary cycle, i.e., a cycle which passes through no node twice.[35]

We form a partition A_1,A_2 of the edges of **V** with endpoint x as follows. Consider the connected components of **V** obtained by deleting the node x and its incident edges from **V.** Since **V** is "bridge"-free, there are at least two edges in **V** from each of these components to x. If there is only one component, we assign any $m_1 - 1$ of the edges in **V** with endpoint x to A_1 and the remaining of these edges to A_2. If there are two or more components, say $\mathbf{C}_1, \mathbf{C}_2, \ldots, \mathbf{C}_d$, we assign one of the edges in **V** going from $\mathbf{C}_1$ to x to A_1 and one of these edges to A_2. The remaining edges in **V** going to x we assign to A_1 and A_2 in any manner such that $|A_1| = m_1 - 1$ and $|A_2| = m_2 - 1$. We claim that for this partition, **V'** is "bridge"-free.

Let $s = (c,d)$ be any edge in **V'**.

Case 1. s is also an edge in **V,** say in the component $\mathbf{C}_i$. Since **V** is "bridge"-free, s must lie on some elementary cycle γ in **V.** If γ does not use the node, x, then γ is also an elementary cycle in **V'.** If γ passes through x, then γ must have precisely two edges with endpoint x, say $t = (e,x)$ and $u = (f,x)$. Now one of the following four-edge sets must be edges in **V':** $\{(e,a_1), (f,a_2)\}$, $\{(e,a_2), (f,a_1)\}$, $\{(e,a_1), (f,a_1)\}$, $\{(e,a_2), (f,a_2)\}$. If, for example, the first edge set is in **V',** then replacing the edges t and u in γ by the edges (e,a_1), (a_1,a_2), and (a_2,f) results in an elementary cycle in **V'** containing s. A similar argument holds if one of the other edge sets is in **V'**.

Case 2. s is not an edge in **V.** In this case s is either (a_1,a_2) or of the form (a_1,e) or (a_2,e). Assume that $s = (a_1,e)$. Let $\mathbf{C}_i$ be the component containing e. There must be at least one node f in $\mathbf{C}_i$ such that either the edge $(f,a_1) \neq s$ is in **V'** or the edge $(f,a_2) \neq s$ is in **V'.** Note that it is possible that $(f = e)$. Since $\mathbf{C}_i$ is connected, there is a simple path γ (i.e., γ uses no node twice) in $\mathbf{C}_i$ going from e to f (if $e = f$, then γ is the null path). The path γ with either

the edge set $\{(f,a_1), s\}$ or the edge set $\{(f,a_2), (a_1,a_2), s\}$ added is an elementary cycle in $\mathbf{V}'$ containing s. A similar argument holds if $s = (a_2,e)$. If $s = (a_1,a_2)$, then there must be two nodes e and f in $\mathbf{C}_1$ such that (a_1,e) and (a_2,f) are edges in $\mathbf{V}'$ (again, it is possible that $(e = f)$). Then any simple path in $\mathbf{C}_1$ from e to f with the edge set $\{(a_1,e), s, (a_2,f)\}$ added is an elementary cycle in $\mathbf{V}'$ containing s. Hence, every edge s in $\mathbf{V}'$ lies on some elementary cycle in $\mathbf{V}'$, and $\mathbf{V}'$ is "bridge"-free and thus is a vertex-graph. The vertex-graph $\mathbf{V}$ arises by collapsing the edge (a_1,a_2) in $\mathbf{V}'$.

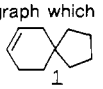
LITERATURE CITED

(1) For Part XV, see D. H. Smith, "Applications of Artificial Intelligence for Chemical Inference. XV. Constructive Graph Labeling Applied to Chemical Problems. Chlorinated Hydrocarbons", *Anal. Chem.*, in press.

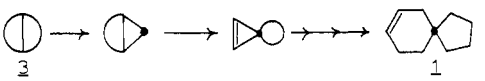
(2) The work was supported by the National Institutes of Health, Grant Nos. RR 00612-05A1 and RR 00785-02 (for the Stanford University Medical Experimental computer facility, "SUMEX") and (to H. Brown) NSF Grant No. DCR74-12646.

(3) Lederberg, J., "DENDRAL. A System for Computer Construction, Enumeration and Notation of Organic Molecules as Tree Structures and Cyclic Graphs, Part II. Topology of Cyclic Graphs", NASA CR-68898. Reprints are available from J. Lederberg, Department of Genetics, School of Medicine, Stanford University, Stanford, Calif. 94305.

(4) Certain ring systems which contain spiro fusions have no direct correspondence with a vertex-graph which contain the spiro center, for example, structure **1**. **1** is constructed from vertex-graph **3** (see text) using loops, *i.e.*,



1



3 1

Such ring systems are characterized by having only secondary nodes in one of the spiro fused rings.

(5) A *loop* is an edge which connects a node to itself. In the process of ascertaining the vertex-group of a molecule, removal of bivalent nodes in ring systems such as **1** may leave one or more loops. The loops are removed. Because this may yield new bivalent nodes, the process is repeated until no bivalents or loops remain, yielding the underlying vertex-graph (see the reverse of the flow from **3** → **1** in ref 4).

(6) Smith, D. H., Masinter, L. M., and Sridharan, N. S., "Heuristic DENDRAL. Analysis of Molecular Structure," in "Computer Representation and Manipulation of Chemical Information," W. T. Wipke, S. Heller, R. Feldmann, and E. Hyde, Ed., Wiley, New York, N.Y., 1974, p 287.

(7) Masinter, L. M., Sridharan, N. S., Lederberg, J., and Smith, D. H., "Applications of Artificial Intelligence for Chemical Inference XII. Exhaustive Generation of Cyclic and Acyclic Isomers," *J. Am. Chem. Soc.*, **96,** 7702 (1974).

(8) In the structure generator[6,7] and in this paper, multiple bonds are considered to be small rings.

(9) The degenerate case of a single ring comprised solely of secondary nodes, e.g., cyclohexane, is accommodated as a special case by the procedures of structure generation.

(10) Masinter, L. M., Sridharan, N. S., Carhart, R. E., and Smith, D. H., "Applications of Artificial Intelligence for Chemical Inference XIII. Labeling of Objects Having Symmetry," *J. Am. Chem. Soc.*, **96,** 7714 (1974).

(11) Balaban, A. T., "Valence—Isomerism of Cyclopolyenes," *Rev. Roumaine Chim.*, **11,** 1097 (1966); Erratum, *ibid.*, No. 1, 103 (1967).

(12) Balaban, A. T., "Chemical Graphs. VIII. Valence Isomerism and General Cubic Graphs," *Rev. Roumaine Chim.*, **15,** 463 (1970).

(13) Balaban, A. T., "Chemical Graphs. XIV. Valence Isomers of [12]Annulene," *Rev. Roumaine Chim.*, **17,** 865 (1972).

(14) Balaban, A. T., "Chemical Graphs. XVIII. Graphs of Degrees Four or Less, Isomers of Annulenes, and Nomenclature of Bridged Polycyclic Structures," *Rev. Roumaine Chim.*, **18,** 635 (1973).

(15) Carhart, R. E., Smith, D. H., Brown, H., and Djerassi, C., "Applications of Artificial Intelligence for Chemical Inference. XVII. An Approach to Computer-Assisted Elucidation of Molecular Structure," submitted for publication.

(16) See the "Index Guide—Index of Ring Systems," *Chem. Abstr.*, 8th Collective Index.

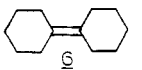
(17) Whitlock, H. W., Jr., and Siefken, M. W., "The Tricyclo[4.4.0.0^8]decane to Adamantane Rearrangement," *J. Am. Chem. Soc.*, **90,** 4929 (1968).

(18) Engler, R. M., Fărcaşiu, M., Sevin, A., Cense, J. M., and Schleyer, P. v. R., "On the Mechanism of Adamantane Rearrangements," *J. Am. Chem. Soc.*, **95,** 5769 (1973).

(19) (a) Collins, C. J., and Johnson, C. K., "A Computer-Assisted Analysis of the Camphor-^{14}C Sulfuric Acid Reaction. Is 3,2-Endo Hydroxyl Shift Necessary?," *J. Am. Chem. Soc.*, **95,** 4766 (1973); (b) Johnson, C. K., and Collins, C. J., "An Algebraic Model for the Rearrangement of 2-Bicyclo[2.2.1]heptyl Cations," *ibid.*, **96,** 2514 (1974); (c) Collins, C. J., Johnson, C. K., and Raaen, V. F., "Molecular Rearrangements XXX. Applications of an Algebraic-Graphical Model for Analyzing Rearrangements of Bicyclo[2.2.1]heptyl Cations," *ibid.*, **96,** 2524 (1974).

(20) Gund, T. M., Schleyer, P. v. R., Gund, P. H., and Wipke, W. T., "Computer-Assisted Graph Theoretical Analysis of Complex Mechanistic Problems in Polycyclic Hydrocarbons. The Mechanism of Diamantane Formation from Various Pentacyclotetradecanes," *J. Am. Chem. Soc.*, **97,** 743 (1975).

(21) Our definition deviates from strict Chemical Abstracts rules on what constitutes a "bridge"-free ring system (see *J. Am. Chem. Soc.*, **82,** 5562 (1960)). Our convention classifies systems which are "bridged" by multiple bonds, e.g., **6,** as "bridge"-free, whereas *Chemical Abstracts* regards **6** as two ring systems. Our convention is that *two* edges must be cut to divide **6** into two parts, considering the double bond as two edges; thus, by strict graph theoretical rules, **6** is "bridge"-free.



6

(22) (a) Sridharan, N. S., "Computer Generation of Vertex-Graphs," Stanford Computer Science Memo, STAN-C5-73-381, Stanford Computer Science Department, Stanford University, Stanford, Calif. 94305; (b) Sridharan, N. S., "Computer Generation of Vertex-Graphs," *Inf. Process. Lett.*, in press.

(23) The present procedure used for the preliminary grouping of a set of graphs prior to graph-matching to eliminate duplicates is a more efficient implementation of the ideas described in ref 22. We presently use the traces of the n powers of the connectivity matrix, where n is the number of nodes, or vertices. These traces group the graphs into their respective "isospectral classes," or graphs with the same characteristic polynomia.[24]

(24) Herndon, W. C., "The Characteristic Polynomial Does Not Uniquely Determine Molecular Topology," *J. Chem. Doc.*, **14,** 150 (1974).

(25) We adopt the convention, for shorthand notation describing the number of nodes of each degree in a vertex-graph, of enclosing a list of nodes of each degree in square brackets [ ]. The order of node degree in the list is tertiary, quaternary, and so forth. Thus, [2 1] refers to vertex-graphs of two tertiary and one quaternary nodes; [0 2 0 2] refers to vertex-graphs possessing two nodes of degree 4 and two nodes of degree 6.

(26) In view of the numbers of some sets of graphs, it is pointless to attempt to draw them all. Alternatively, the connection matrices in compressed form together with instructions on interpretation of the connection matrix format are available from the authors.

(27) Brown, H., "Molecular Structure Elucidation. III," *SIAM, J. Computing*, submitted for publication.

(28) Tanaka, N., Iizuku, T., and Kan, T., "A Graph Theoretical Deriviation of the Isomers of Tricyclic Hydrocarbons," *Chem. Lett.* 539 (1974).

(29) Wiberg, K. B., and Burgmaier, G. J., "Tricyclo[3.2.1.0^{1,5}]octane. A 3,2,1-Propellane," *J. Am. Chem. Soc.*, **94,** 7396 (1972).

(30) Eaton, P. E., and Temme, G. H., III, "The [2.2.2]Propellane System," *J. Am. Chem. Soc.*, **95,** 7508 (1973).

(31) Fărcaşiu, D., Wiskott, E., Osawa, E., Thielecke, W., Engler, E. M., Slutsky, J., Schleyer, P. v. R., and Kent, G. J., "Ethanoadamantane. The Most Stable $C_{12}H_{18}$ Isomer," *J. Am. Chem. Soc.*, **96,** 4669 (1974).

(32) Smith, D. H., unpublished results.

(33) Rouvray, D. H., "The Uses of Graph Theory," *Chem. Brit.*, **10,** 11 (1974), and references cited therein.

(34) Halin, R., "A Theorem on n-connected Graphs," *J. Combinatorial Theory*, **7,** 150 (1969).

(35) Berge, C., "Graphs and Hypergraphs," North-Holland Publishing Co., Amsterdam, 1973.