forms of monosaccharides and their derivatives", Recommendations 1980. *Eur. J. Biochem.* **1980**, *111*, 295-298.
Joint Commission on Biochemical Nomenclature (JCBN). "Nomenclature of tetrapyrroles," Recommendations 1978. *Eur. J. Biochem.* **1980**, *108*, 1-30.

OTHER

International Union of Pure and Applied Chemistry and International Federation of Clinical Chemistry. Quantities and units in clinical chemistry", Approved recommendations 1978. *J. Clin. Chem. Clin. Biochem.* **1979**, *17* (12), 807-821.
International Union of Pure and Applied Chemistry and International Federation of Clinical Chemistry. "List of quantities in clinical chemistry", Approved recommendation 1978. *J. Clin. Chem. Clin. Biochem.* **1979**, *17* (12), 822-835.
International Union of Pure and Applied Chemistry and International Federation of Clinical Chemistry. "Physico-chemical quantities and units in clinical chemistry", Document stage 2, draft 1 with a view to a recommendation, *Clin. Chim. Acta* **1980**, *108*, 501F-539F.
International Federation of Clinical Chemistry: Committee on Standards. Expert Panel on Nomenclature and Principles of Quality Control in Clinical Chemistry. "Approved recommendation 1978 on quality control in clinical chemistry. Part 1. General principles and terminology", *J. Clin. Chem. Clin. Biochem.* **1980**, *18*, 69-77.

Computer-Assisted Mechanistic Evaluation of Organic Reactions. 2. Perception of Rings, Aromaticity, and Tautomers

BARBARA L. ROOS-KOZEL and WILLIAM L. JORGENSEN*[1]
Department of Chemistry, Purdue University West Lafayette, Indiana 47907

Received August 29, 1980

New algorithms for the perception of rings, aromaticity, and tautomers were developed in conjunction with the CAMEO program for computer-assisted mechanistic evaluation of organic reactions. Briefly, the algorithms contain the following steps: prune the molecule of side chains, find ring(s), subdivide each ring into structural units, sum the units' contributions for the π-electron count, and determine whether tautomerization is possible. A noteworthy aspect of the ring alogrithm is that only rings belonging to the smallest set of smallest rings (SSSR) are found. Results for representative molecules are presented, and rules for determining the aromaticity/tautomer status of rings, especially heterocycles, are discussed.

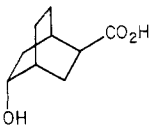
INTRODUCTION

An interactive computer program (CAMEO) is being developed in our laboratory to predict products of organic reactions given starting materials and conditions.[2a] A particularly novel aspect of the program is that mechanistic reasoning is used to predict products rather than relying on data tables for numerous, known, often mechanistically similar transformations.[2b-d] The program has three principal segments: (1) graphics which control the input and output of structures and communication between the user and computer, (2) perception of structural features and reactive sites in the reactants and products, and (3) mechanistic evaluation of potential reaction paths. An overview of CAMEO has been presented previously.[2a] Since substantial innovation has been made in several areas, details on specific procedures will also be presented in a series of papers. The present work focuses on the algorithms that have been developed for the perception of rings, aromaticity, and tautomers. The procedures are general and can be applied to other problems in which the perception of such structural features is required. Listings of the perception routines in CAMEO are available from the authors upon request.

RING PERCEPTION

Background. Enabling a computer to perceive rings is an essential part of any organic synthesis program. The presence of rings restricts conformational freedom, decreasing the feasibility of some intramolecular reactions but enhancing the likelihood of others. Reactivity toward common reagents may be substantially altered; contrast the behavior of cyclopropanone and acetone.

Although one could perceive all possible rings, i.e., all unique closed paths, this is excessive. Only the number which will suffice to define intramolecular relationships must be found. Frērejacque's number,[3] calculated by applying the following equation to a molecule, number of rings = number of bonds - number of atoms + 1, yields the number of rings in the smallest set of smallest rings (SSSR). From the SSSR, one can determine the total ring strain energy, aromaticity, stereochemistry, topology, and the set of synthetically important rings. A recent paper by Gasteiger and Jochim explores the advantages of finding and using the SSSR in detail.[4] It should be noted that envelope rings (rings which can be constructed by the intersections of, and which circumscribe, two SSSR rings, i.e., the six-membered ring in bicyclo[2.2.0]hexane) are not necessary for determining ring strain energy. Difficulties arise when applying this equation to highly symmetrical molecules, for instance the bicyclo[2.2.2]octanes. Following Frērejacque, only two of the three equivalent (if unsubstituted) rings would be perceived. Furthermore, the two rings perceived would vary depending on the order of atom numbers. As unsymmetrical substitution, shown in **1**, can occur, either (1)



1

subsequent processing must be independent of which pair of rings is found, or (2) the ring perception algorithm must find all three rings. The second option was chosen, necessitating the redefinition of the SSSR to include (ignoring substitution) symmetrically equivalent rings.

Previously published programs have constructed the SSSR by several methods. Wipke and Dyott used Welch's first algorithm to find the basis set of rings; then ring assemblies were constructed.[6] Finally all rings were found by means of Gibbs' algorithm. A minimal basis set and a set of "chemically interesting" rings (the basis set plus any ring of ≤ 8 bonds) were then found.[6] Despite the circuitous path, this approach

0095-2338/81/1621-0101$01.25/0

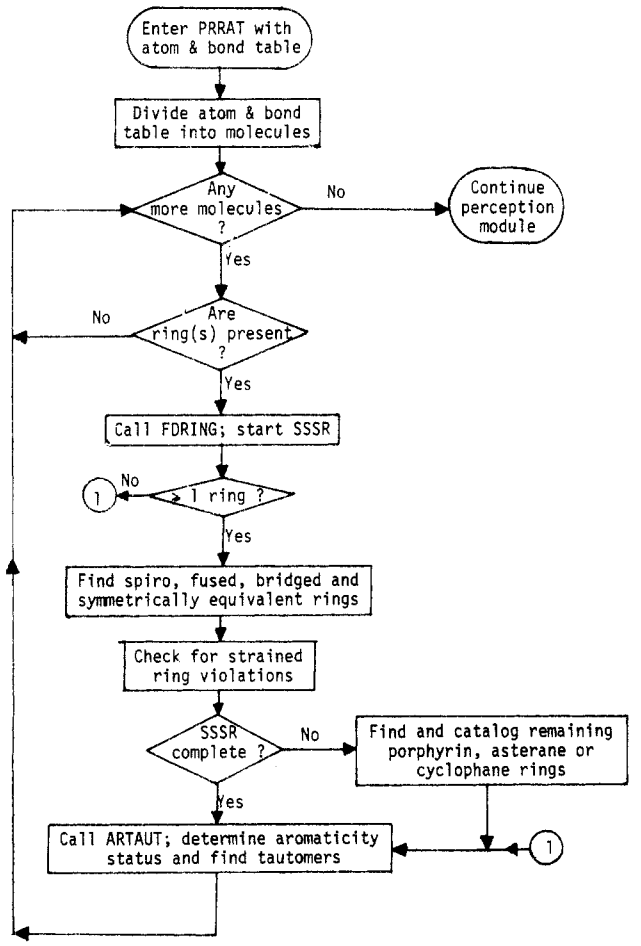


Figure 1. An overview of ring, aromaticity, and tautomer perception.

proved efficient and showed how bit operations could be used to facilitate this task. Zamora used a path-tracing technique to find the smallest ring(s) in which (a) unused atoms occur, or when all atoms in the molecule have been used, (b) unused faces and edges occur. This process continues until the smallest set of rings required to characterize the molecule is found. One disadvantage is that a ring surrounded by other rings may not be found.[7] Gasteiger and Jochim begin by generating a spanning tree from an arbitrary reference atom. Tracing back from common atom(s) along the shortest paths yields a set of rings. Next, the "complexity" is determined. If it is over a specified value the spanning tree process is repeated, using a reference atom calculated to yield additional rings.[4] A number of workers have also used the graph theory approach, wherein rings are treated as cyclic graphs.[2e,5,8-10] As described here an efficient algorithm which generates the SSSR and symmetrically equivalent rings has been implemented in CAMEO. The procedure features the recognition of rings simultaneously with growing paths and trees. If one tree does not yield at least Frêrejacque's number of rings, another tree is grown. This approach has resulted in a concise, efficient program (FDRING) capable of handling a diverse assortment of ring systems. However, it was found to be more expeditious to handle the bridging rings of asteranes, cyclophanes, and porphyrins in PRRAT, which is the executive program for the module and also determines inter-ring relationships. Figure 1 provides an overview of ring/aromaticity/tautomer perception in CAMEO and will serve as the basis for the following detailed discussion.

Overview of FDRING. Ring perception occurs in CAMEO after fundamental atom and bond sets have been created from the atom and bond table for a structure.[2a] Five major steps are performed to perceive the ring(s) in a molecule. The first step

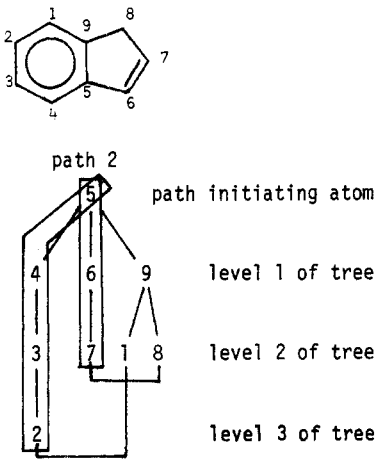
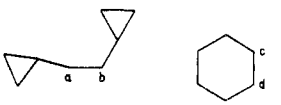


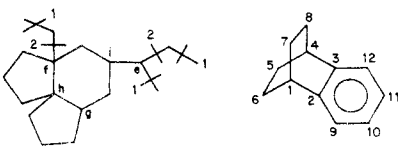
Figure 2. How FDRING finds the rings in indene.

is to remove nonring atoms. This eliminates the growing of useless paths containing acyclic substituents. Next, the branched atoms (atoms connected to >2 nonhydrogen atoms) are located. These will produce the largest number of rings when used as path-initiating atoms (branched atom at origin of paths used to construct rings). Paths are grown from an atom in this set, and at each level of the "tree" (set of paths originating at a given atom) pairs of paths are examined to determine whether or not they constitute a ring (see Figure 2). Finally, a "buried" ring, i.e., the central, completely enclosed ring in a highly branched molecule like strychnine, is searched for. If the SSSR remains incomplete, the next branched atom is used as a path-initiating atom. A detailed explanation of these steps will now be given (see also FDRING flowchart, Figure 3).

Trimming the Molecule. To limit path growth and to decrease the number of path initiating atoms, all nonring atoms (except those in an unbranched chain between two rings) must be removed by FDRING.



Atoms such as a and b are not removed as it is impossible to distinguish them from c and d; all are adjacent to two other atoms in the molecule. Only molecules with no primary nonhydrogen atoms or which consist of a single, unsubstituted ring, i.e., benzene, can be processed directly. Otherwise, as shown below, the exocyclic atoms are successively removed, proceeding from the terminal atoms to the ring atoms. All chains are trimmed simultaneously; the process ceases when the longest chain encounters a ring atom. Whenever a branched atom is found, it must be determined whether it is a ring atom. If the atom in question is adjacent to at least two atoms in the trimmed molecule, it is a ring atom; note the difference between e and f after cut 2. The set of branched atoms[6] is then analyzed to find the set of path-initiating atoms. Ring atoms at the ends of chains (f and i) are scrutinized and if, like atom i, they are *not* ring junction atoms (RJA; the atom located at the intersection of ≥2 rings),



2

they are deleted from the set of potential path initiating atoms. RJAs are subdivided into tertiary and quaternary. Atoms such

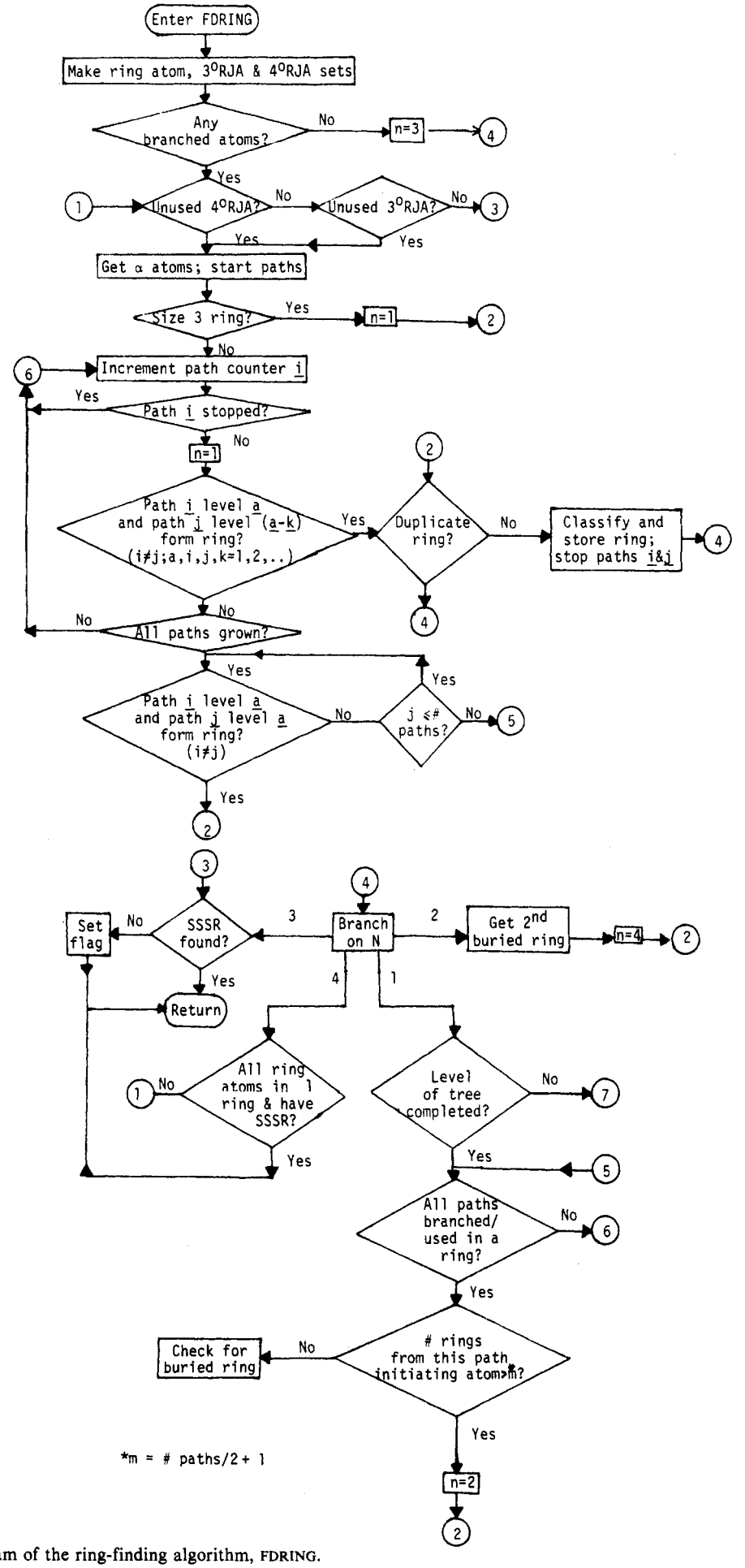


Figure 3. Flow diagram of the ring-finding algorithm, FDRING.

as f, nominally quaternary but only tertiary with respect to the number of adjacent ring atoms, are tertiary, as is g. Only atoms analogous to h reside in the quaternary category.

Path Growing. Quaternary, then tertiary path-initiating atoms are cycled through the path-growing–ring-finding algorithm in numerical order; using the more branched atoms first expedites finding the rings. If a path-initiating atom is already in at least two rings, it is not used unless all the other path-initiating atoms have been used and the SSSR has not been found. This greatly decreases the number of duplicate rings found. Atoms are processed until (a) Frêrejacque's number[3] of rings have been found *and* all the atoms in the trimmed molecule have been placed in a ring or (b) every path-initiating atom has been used. The requirement that all the atoms in the trimmed molecule must be placed in a ring is essential for the correct perception of the rings in molecules like **2**. If a molecule is drawn in such that a bridgehead atom has the lowest number (1 in **2**), paths would first be grown from it, and Frerejacque's number of rings (the three bridged rings) would be found. Unable to exit as atoms 9–12 have not been placed in a ring, the program grows paths from atom 4 and finds the aromatic ring. Asterane-, cyclophane-, and porphyrin-type molecules come under case b; the large circumscribing ring(s) which are needed to complete the SSSR are found in PRRAT by using a special algorithm (vide infra).

Each path being grown from the path-initiating atom is stored as a 32-bit set (CAMEO handles molecules with up to 32 explicit atoms). *Atoms are added to each path until it is used in constructing a ring or until the path branches.* Further path growth would result in picking up envelope rings.

Finding Rings. The presence of ring(s) (consult Figures 2 and 3) is checked for (1) after finding the set of atoms α to the path-initiating atom (three-membered rings), (2) when each path is grown one level (even size rings), and (3) after all the paths have been grown one level, i.e., when the level is complete (odd size rings). Using indene as an example (see Figure 2), when the second level has been grown, path 2 = {5,6,7} and path 2's terminal atom is 7; path 3 = {5,9} and its terminal atoms are 1 and 8. The terminal atoms are used when checking for the presence of a ring; the paths are used to construct the set of atoms constituting a ring. Thus, the intersection of the set of atoms α to the terminal atom of path 2, {6,8}, and the set of terminal atoms of path 3, {1,8}, is nonzero; a ring exists. The ring is the union of path 2 and path 3, {5,6,7,9}, plus the intersection atom, {8}. (If more than one intersection atom is found, the algorithm recycles to construct a ring using the next intersection atom and the same union of paths.) The aromatic ring is found by a similar procedure. Path 1 is grown another level; the terminal atom set is {2}. The intersection of atoms α to {2} and the set of terminal atoms of path 2, {7}, is zero; however, the same operation using path 3 leads to a nonzero set, {1}. As before, path 1 united with path 3 plus {1} yields the ring {1,2,3,4,5,9}. Since these two rings constitute the SSSR, the task is complete, and the second atom in the path-initiating atom set, 9, is not used.

When a ring is found, it is first ascertained whether or not it is a duplicate of a previously found ring. If not, it is stored. *Additionally*, its atoms and bonds are stored according to the size of the ring they belong to, facilitating subsequent queries by CAMEO for information on such items as the number of trans double bonds in rings of size *n*. Two flags are set at this point. One prevents further growth of the paths used to form the ring. The second prevents the algorithm from checking this pair of paths twice, avoiding picking up duplicate rings. This flag uses the bits in a single, 16-bit word. Since the pairs of paths are always examined in the same order (1–2, 1–3, 2–3, etc.), a unique number 1,2,3...) can be assigned to each pair. When the bit corresponding to this number is "on", the appropriate pair is skipped over.

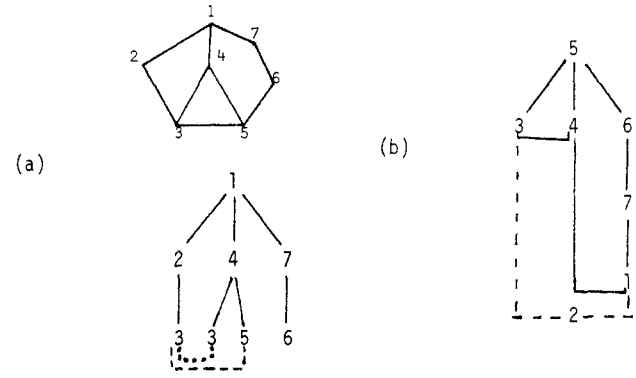


Figure 4. (a) Growing path 2 after a ring has been found using this path, {1,2,3,4}, leads to an envelope ring (---) and a duplicate ring (. . .) (b) Dotted lines show envelope ring found during buried ring check; since a three-membered ring was found from this path-initiating atom, a buried ring check should not have been done.

One advantage of this algorithm is that the same code is used to inspect paths of the same level, and paths of the *n* and $n - m$, $m = 1,2, \ldots$, levels. A second advantage is that the problem of picking up envelope rings is completely avoided by the unique ring-finding process employed. Two features are essential. First, a path must be halted as soon as it branches or is used to form a ring. Failure to do this will result either in a duplicate ring or envelope ring being found (see Figure 4a). Second, a buried ring must not be sought when a three- or four-membered ring has been found, as any buried ring discovered may be an envelope (Figure 4b). A buried ring exists when, in a highly branched molecule, no pair of paths is allowed to grow far enough to meet and form the ring in question. Instead, the paths stop one atom short of the necessary length. If, for a given path-initiating atom, the quantity 1 + (number of paths/2) is greater than the number of rings found from that path-initiating atom, a buried ring is searched for, subject to the above restriction.

For example, in strychnine (see Figure 5) the first path-initiating atom to be used is 7 (the only quaternary one). Two rings are found; $1 + (^4/_2) > 2$, so the buried ring algorithm is entered. For a buried ring which is *not* an envelope ring to exist, one of the atoms not on any of the paths grown from atom 7, {1,2,3,4,10,11,12,13,15,21,22,23,24}, must be (1) α to two terminal atoms and (2) not *both* a secondary atom and in a ring which has been found. Atom 15 meets these requirements; hence, the buried ring is ([(path 2)$\cup$(path 3)] + 15 + [(atoms α to 15)$\cap$(terminal atoms)]). It is possible to have 2 buried rings per path-initiating atom, given the limit of 32 atoms per molecule.

During the perception of strychnine (SSSR = 7), four path-initiating atoms were used to find nine rings, two of which were duplicates of previously found rings. If path-initiating atoms which were in at least two rings at the time they were selected for path growing had not been removed from consideration, nine more duplicate rings would have been found.

When Frêrejacque's number of rings has not been found and all the atoms in the trimmed molecule have not been placed in a ring though paths have been grown from each path-initiating atom, a separate algorithm searches for an asterane (**3**) or cyclophane (**4**) if two rings have been found by FDRING. If four path-initiating atoms are present, the molecule is a cyclophane (defined by CAMEO as a molecule containing two rings linked by two bridges); if greater than four are present, the molecule is an asterane. A porphyrin-like molecule (**5**, **6**) is present if FDRING found more than two rings (see Figure 6). To find the large rings in asterane-type molecules an algorithm proceeds outward from two adjacent vertices on ring 1 until it encounters two adjacent vertices on

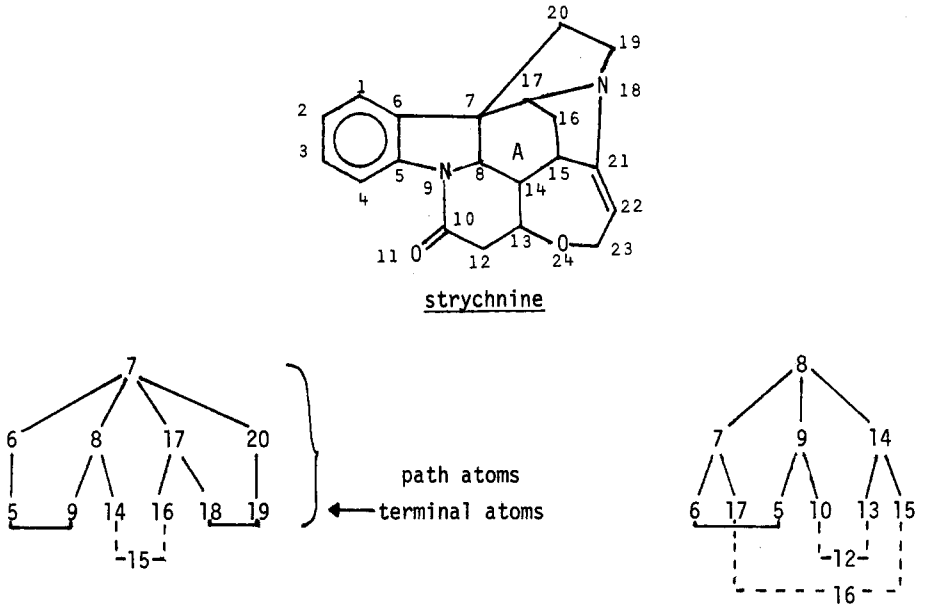
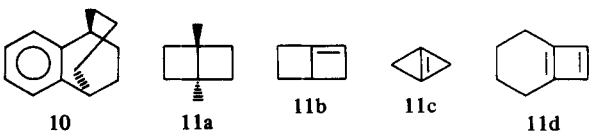


Figure 5. How a buried ring is found, using ring A in strychnine as an example.

the second ring. After finding a ring, the pair of vertices on ring 1 is shifted one atom and the ring-growing process repeated. When ring 1 has been circumnavigated in this fashion, all the rings have been found. A separate algorithm treats paracyclophanes; ortho- and metacyclophanes are handled by FDRING. An orthocyclophane, such as **7**, is a trivial case; it contains fused, not bridged, rings and FDRING handles it. Metacyclophane **8** and mixed-cyclophane **9** (contains one meta and one para linkage) are also handled by FDRING, and the three smallest rings are found (see Figure 6). Paracyclophanes have a different topology, and it seems appropriate to elucidate all six rings. FDRING finds only the two bridged rings due to the ≥2-atom gap separating pairs of terminal path atoms. So, as shown in Figure 6, the two bridges, *n* and *m*, and the four half-rings are combined to yield four more unique rings. Two general classes of porphyrin-type molecules exist. Those analogous to **5** have a single large central ring (heavy line). Alternatively, those resembling **6** have two possible central rings, as *m* and *q* are equivalent segments. These central rings are constructed by manipulating the sets of ring and nonring atoms/bonds.

Determining Inter-Ring Relationships. If two or more rings exist in a given molecule, the number of atoms common to each possible pair of rings is found. This number indicates the relationship between the pair; for example, zero implies they are not adjacent, one that they are spiro. If the pair is bridged (three or more common atoms), and a symmetrically equivalent ring exists which has not yet been picked up, this ring is constructed and processed through the same classification and storage procedure used in FDRING. Three symmetrically equivalent (ignoring ring substituents) rings will exist when two rings of equal size are bridged across their centers, forming an envelope ring of the same size, as in **1**. If FDRING does not use a bridgehead atom (1 or 4 in **2**), to initiate paths, only two of the three rings will be found. This results in a flag being set and the remaining ring being picked up in PRRAT. During extensive testing on a wide variety of natural products and cage compounds, the algorithms have proved completely reliable. Table I lists the rings found for molecules of diverse structural types. In each case the computation time for the ring/aromaticity/tautomer perception was less than 3–4 s on our TI990/10 minicomputer which is generally 50 times slower than a CDC/6400.

Weeding Out Strained Ring Systems. Some screening out of reactants and products with overly strained rings is performed concomitantly with the perception of inter-ring relationships. Information obtained during stereoperception concerning the stereochemistry of ring fusions and bridges is also utilized for this purpose. Currently, occurrence of the following structural features causes a molecule to be rejected: (1) a trans bridge between two rings if both rings are size seven or less, e.g., **10**; (2) a trans fusion bond and less than seven



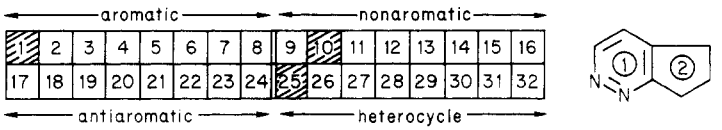


Table I. Rings Found in a Wide Assortment of Molecules

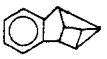
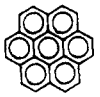
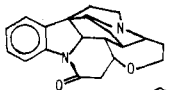
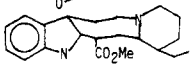
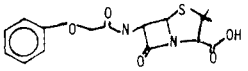
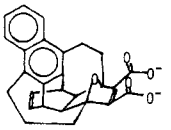
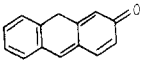
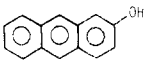
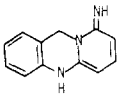
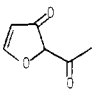
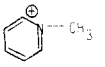
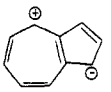
molecule	rings found
	3,4,4,5,6
	6,6,7,7
	6,6,6,6,6,6,6
	5,5,5,6,6,6,7
	5,6,6,8
	3,7,7,7
	4,5,6
	3,4,5,5,5,5
	6,6,12,12,12,12
	5,5,6,6,6,6, 6,7,8

Table II. Aromaticity/Tautomer Classifications of Representative Molecules

structure entered	AR, AA, NA[a]	tautomer?	tautomerized structure
	AR,AR,AR	Yes	
	AR,NA,NA	No	
	AR,AR	No	
	AR	No	
	NA	No	
	AR	Yes	
	AR,AR	No	
	AR,NA	No	
	NA	No	

[a] AR = aromatic, AA = antiaromatic, NA = nonaromatic.

atoms in the envelope, e.g., **11a**; (3) a trans double bond in a ring of size seven or less, which is a modified form of Bredt's rule; (4) a double bond adjacent to a fusion bond and less than seven atoms in the envelope, e.g., **11b**; (5) a double fusion bond and less than six atoms in the envelope, e.g., **11c**; and (6) a cyclobutadiene ring not fused to an aromatic ring, e.g., **11d**. These restrictions are based on synthetic results and the consideration of molecular models. They could be easily modified in the future if they are found to be too restrictive. If the aborted molecule is a reactant, CAMEO returns the initial structure and an error message; if the molecule was a product, it is eliminated from the list of possible products.

AROMATICITY AND TAUTOMER PERCEPTION

Introduction. Classifying a ring as aromatic implies that it possesses exceptional stability and that it only undergoes certain reactions. The program can significantly decrease its options by declaring aromatic rings unreactive toward many reagents. In addition, resonance stabilization energy must be included in any ΔH_{rxn} calculation performed to help assess the probability of a given computer-generated reaction.

The Hückel $(4n + 2)$ rule is the mainstay of aromaticity determination, but it is not infallible. Above ca. 22 π electrons,[11] conjugated rings assume nonaromatic behavior, and only cyclobutadiene or other 4 π-electron rings are truly antiaromatic. Eight π-electron rings such as cyclooctatetraene, oxepin, and azepin assume a nonplanar "tub" conformation. Eight, twelve, . . . π electron rings are therefore placed in the nonaromatic category, as are rings having $(4n + 2)$ π electrons with $n > 5$. Many azoles, six π-electron heterocycles, undergo ring cleavage and other reactions characteristic of nonaromatic compounds.[15,16] Reactivity varies greatly, however, and this class *is* generally more stable than furan, pyrrole, or thiophene.[12] It appears best to retain these in the aromatic category and to deal with their unusual reactivity in the mechanistic portion of CAMEO. The aromaticity rules and tautomerizations described here are based on experimental observations; so as new data become available additions and modifications may be necessary. General principles have been sought that are not unduly restrictive.

Aromaticity Classification. The aromaticity–tautomer program (ARTAUT) uses Hückel's rule (with the modifications noted above) to classify rings as aromatic (AR), antiaromatic (AA), or nonaromatic (NA). A ring containing more than one heteroatom is also noted, due to its unusual reactivity, as mentioned above. This information is stored in one 32-bit word. The word is divided into four blocks of eight bits each; each bit corresponds to one of the eight possible rings allowed in CAMEO. An example is shown below. Bits 1 and 10 are on, indicating that ring **1** is aromatic and ring **2** is nonaromatic. Bit 25 indicates **1** has ≥2 heteroatoms and may show unusual reactivity.

ARTAUT processes rings by summing the contributions from *sets* of atoms, rather than by examining each ring atom in-

Structure	Rings found in FDRING	Rings found in second algorithm
3A (Ast.)*	3,3	6,6
3B (Ast.)	4,4	7,8,8,9
4 (Cyc.)[1]	6,6	{1,2,3,4,7,8,9,10 + n + m} {1,2,3,4,7,10,11,12 + n + m} {1,4,5,6,7,8,9,10 + n + m} {1,4,5,6,7,10,11,12 + n + m}
5 (Por)[2]	5,5,5,6	14 (heavy line)
6 (Por.)[3]	5,6,6	13,13
7 (Cyc.)	4,6,6	none; FDRING handles
8 (Cyc.)	4,5,8	none; FDRING handles
9 (Cyc.)	4,6,9,9[4]	none; FDRING handles

[1] Rings not necessarily the same size; n=m or n≠m.

[2] n=0,1,2,... ; all m ≠ 1 ; q > 1.

[3] n=0,1,2,... ; > 1 m > 2 ; q > 1.

[4] This molecule has two symmetrically equivalent rings.

* Ast. = Asterane ; Cyc. = Cyclophane ; Por. = Porphyrin

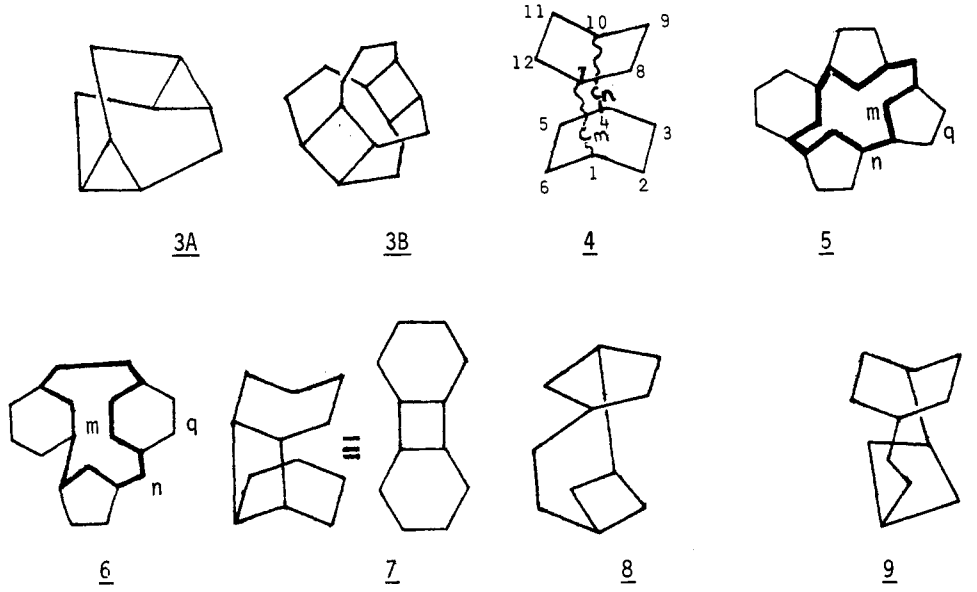
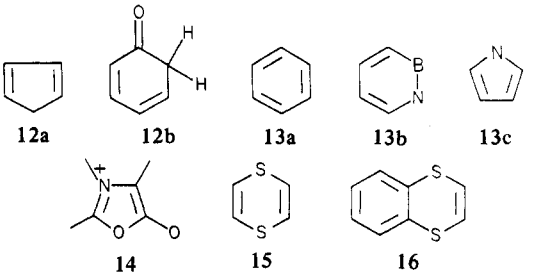


Figure 6. Rings which are found in asteranes, cyclophanes, and porphyrins.

dividually, as was done in LHASA.[12] This method allows extremely rapid processing of simple aromatic and nonaromatic systems. The flowchart is shown in Figure 7.

Preliminary Screening. The first step is to find those rings which are obviously nonaromatic. Such rings have at least one of the following characteristics: (1) no intra-ring double bonds, (2) contain a quaternary atom, (3) contain more than one saturated carbon (cannot tautomerize), (4) contain a monoradical, (5) contain a sulfoxide or sulfone. Secondly, ARTAUT searches for simple aromatic rings, i.e., those with no exocyclic π bonds that are fully conjugated. The conjugating atoms in rings and the number of π electrons they contribute are as follows: (1) cationic carbon and boron, 0; (2) saturated heteroatoms, 2; (3) anionic carbon, 2; (4) radicals not on π bonds, 1; (5) atoms on intra-ring π bonds, 1. The ring under consideration may be classified immediately if (1) it has no exocyclic π bonds *and* (2) all the ring atoms belong to one of the above sets. In this case Hückel's rule as amended above applies. Both conditions are satisfied by **13a–c**; thus, their π-electron count has been completed at this point and they may be classified. If condition 2 is not satisfied, the presence of a nonaromatic ring (**12a**), aromatic tautomer (**12b**), or ring



with exocyclic π bonds (heterocyclic ring in **16**) is indicated. Further study of the ring is required in order to distinguish among these possibilities.

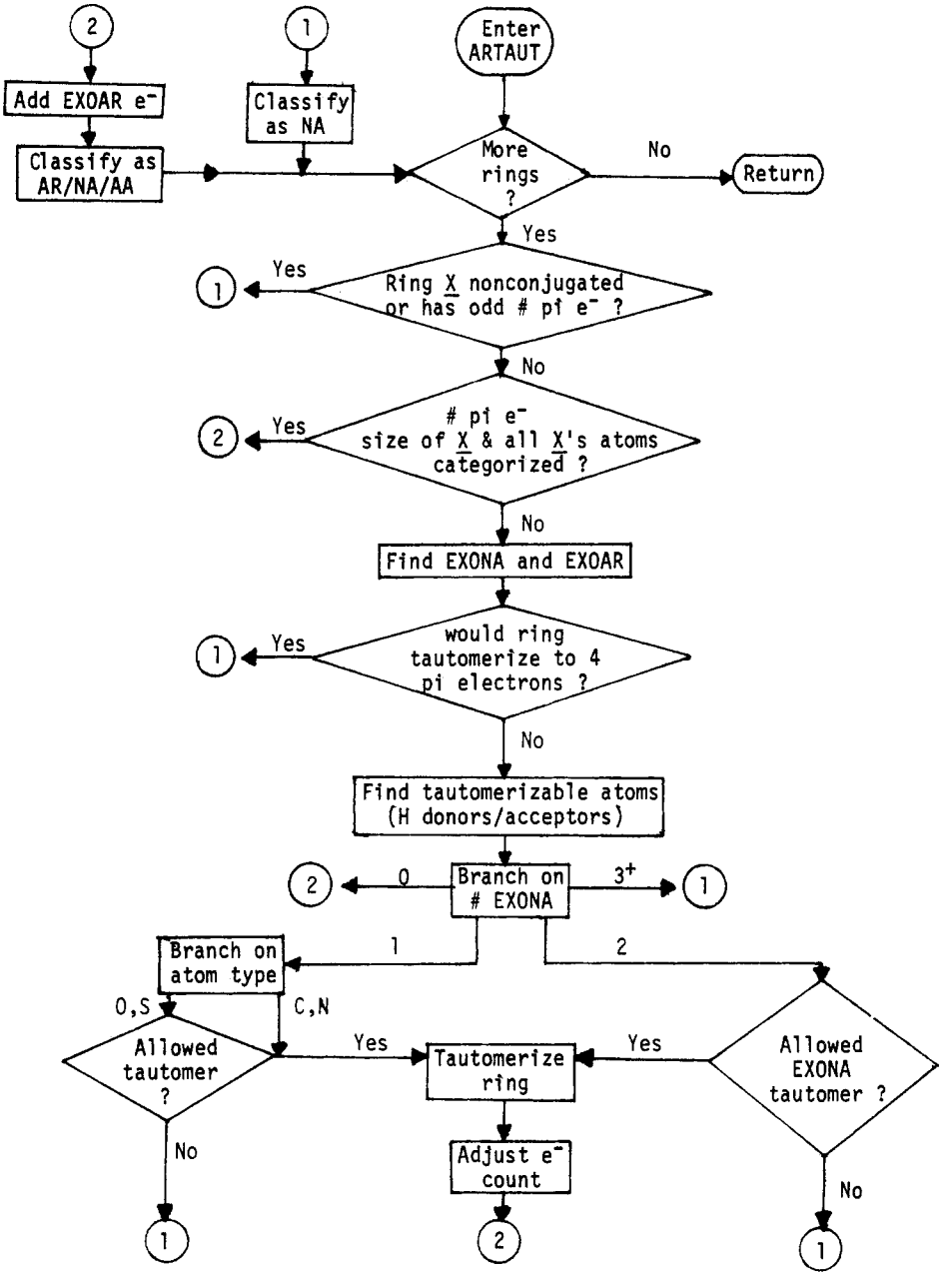


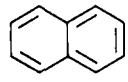
Fig. 7

Figure 7. Flow diagram of the aromaticity/tautomer algorithm, ARTAUT.

Exceptions to this scheme are rare. One is the münchnone **14**, nonaromatic despite the six electron π system.[13] A second arises when a ring containing saturated heteroatoms changes character when fused to an aromatic ring. **15** is nonaromatic, preferring a boat conformation; **16** is aromatic.[14] ARTAUT, which examines each ring independently, would classify the heterocyclic ring as nonaromatic in both cases. A program which also examined ring envelopes would fare no better.

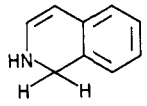
Lastly, ARTAUT checks for the presence of exocyclic π bonds. If there are none, as in **12a**, the molecule is nonaromatic. If some exist, they are divided into two groups. EXOAR are those which are contained in a second completely conjugated ring having 6 or 10 π electrons. (Atoms in rings with 6 or 10 π electrons are collected into a set in PRRAT.) EXONA contains those which are acyclic or in a nonconjugated ring. The quantities of each type of exo π bond and the types of atoms constituting them and adjacent to them determine whether the ring is aromatic or can be tautomerized. These possibilities will now be discussed.

If EXONA is a null set, the ring's status depends on the number of atoms in EXOAR. The possibilities are illustrated by **18–20**. When EXOAR is even and no saturated carbon atoms are present, the ring must be completely conjugated and have either $4n$ or $(4n + 2)$ π electrons. When one saturated carbon atom is present in a ring with ≥ 1 EXOAR, and no EXONA exists, the ring must be nonaromatic, since any attempted tautomerization would destroy the aromaticity of the ring the EXOAR atoms are in. An odd number of EXOAR implies that the π-electron count of the ring is an odd number, since intra-ring π bonds + heteroatoms + anions + diradical electron total is even. (Even if tautomerization were possible, two more electrons would be added, keeping the π-electron count odd.) A ring with an odd number of π electrons can never, of course, be aromatic.



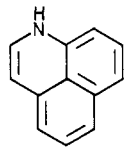
18

even
0
A, AA



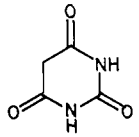
19

even
1
NA



20

odd
0 (no. is irrelevant)
NA



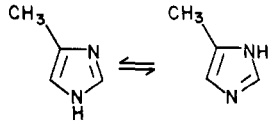
17

no. of EXOAR
no. of saturated carbons
possibilities

AR = aromatic, NA = nonaromatic, AA = antiaromatic

If EXONA contains more than two atoms, as in **17** (barbituric acid), tautomerization is not allowed or observed.[17] When EXONA contains one or two atoms, however, it may be possible to transform the ring into an aromatic tautomer. The remainder of ARTAUT considers this problem.

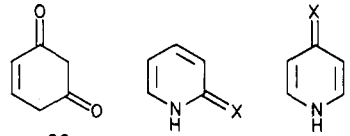
Tautomerization. A "tautomer" is defined as a nonaromatic ring which, after undergoing *one* [1, *m*], $m = 3, 5, 7$, hydrogen shift, has $(4n + 2)$ π electrons. The two rapidly equilibrating forms of 4-methylimidazole (**21**), thus, are not considered tautomers because both forms are aromatic.



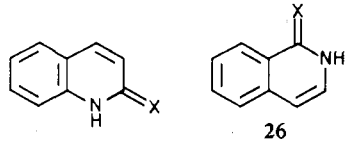
21

Imposing the condition that aromaticity be achieved by one (1, *m*) shift is consistent with experience.[18,21] Very few rings spontaneously undergo multiple tautomerizations; however, **22** is an example.[18] We now consider how the number of EXONA, the types of atoms on the path of the (1, *m*) shift, and the size of the ring determine which tautomerizations are allowed.

Exocyclic Tautomers: One EXONA Atom. All one-step tautomerizations, using a carbon or nitrogen atom bonded to at least one hydrogen, do not necessarily occur. Theoretical[19] and experimental[20] evidence for protic media indicates that the oxo, thione, amino, and methylene tautomers usually predominate, as in **23–26**.[21] Exceptions arise when electron-withdrawing groups are present ortho to the nitrogen atom,[22] if benzo annelation occurs at particular sites,[23,2,4] and if the ring contains multiple heteroatoms.[26] It is impractical

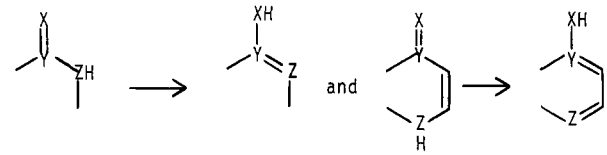


22 23 24



25 26

X = O, S, NH, CH_2



X	Y	Z	Tautomerize
C,O,N,S	C	C	Yes*
C,O,N,S	C	O,S	No
C,N	C	N	Yes
O,S	C	N	No
C,O,N	$N^{\oplus}$	C,N	Yes

*except molecules such as anthrone or phenanthrone.

Figure 8. Summary of the allowed tautomerizations for a six-membered ring with one exo π bond.

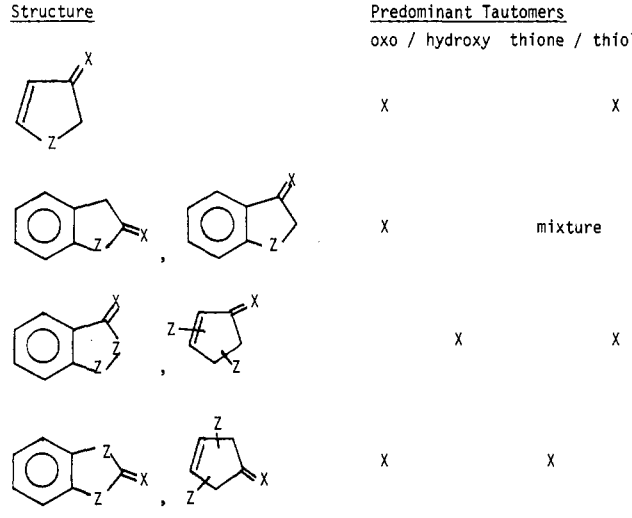
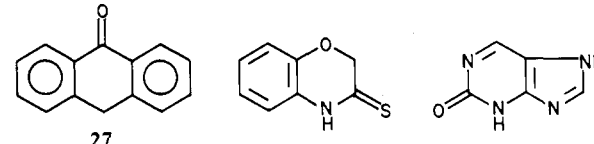
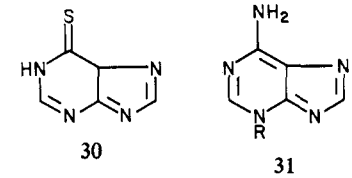


Figure 9. Summary of hydroxy/oxo and thiol/thione tautomerization in five-membered heterocycles with one exo π bond.

at this time for ARTAUT to consider the effects of electron-withdrawing groups, but benzo annelation and multiple heteratoms are dealt with. Ring size must also be considered. Six-membered rings generally fall into case **27**,[18] or one of the



27 28 29

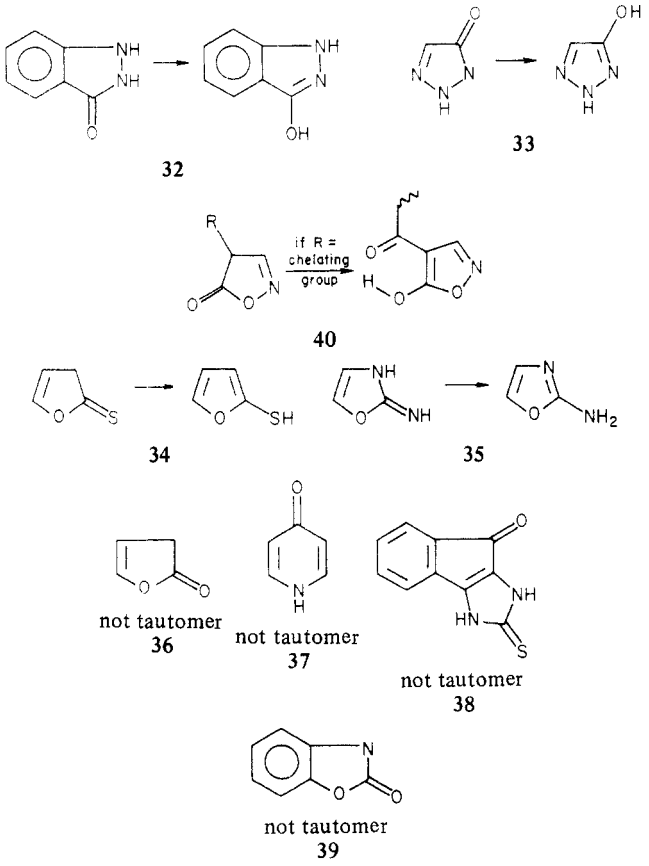


30 31

systems listed in Figure 8, as can be gathered from perusal of Katritzky's review.[21] The generalizations of Figure 8 are violated by **28**,[35] but it would be thrown out as nonaromatic earlier in ARTAUT, as it would tautomerize to an eight π-electron system. Fusing a five-membered ring to the six-membered ring does not change the susceptibility to tautomerization, as **29–31**[36–38] illustrate.

Five-membered rings are more diverse and do not always follow the rules promulgated for six-membered rings. The

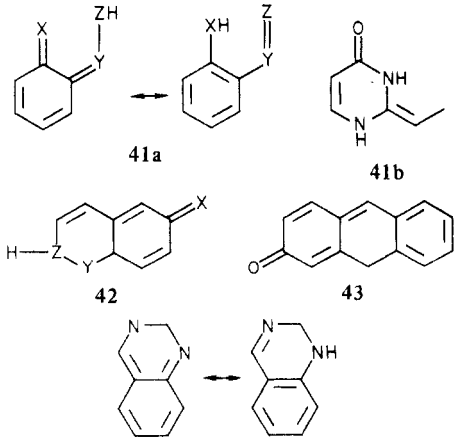
Scheme I



major division results from exo methylene and imine following the rules in Figure 9 and being unaffected by benzo annelation,[31] while oxo and thione tautomers exhibit different behavior (see Figure 9).[25,26,29]

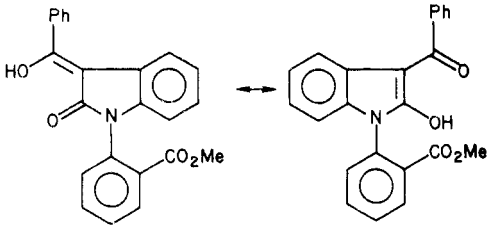
Utilizing the patterns shown in Figure 9, five-membered heterocycles with oxo/thione functionality are processed; **32–39** are correctly perceived and assume the structure indicated. When hydrogen bonding intervenes **40**[32] this scheme may break down (Scheme I).

Exocyclic Tautomers: Two Unlike Hetero-EXONA Atoms or Both EXONA Are Carbon. When EXONA contains two atoms of different types, as in **41–44**, tautomerization is allowed if the EXONA atoms are ortho or para. If they are meta, as in **41b**, tautomerization is impossible. For tautomerization to proceed, condition 1 and either condition 2 or 3 must be met: (1) Mandatory: Z must be bonded to at least one hydrogen, and both Y and Z must not be heteroatoms.[33] (2) If X, Y, and Z are exocyclic, Z must be a heteroatom. (3) If Y and Z are endocyclic (**42**), Z may be carbon if the tautomerization creates at least one aromatic ring. Heterocycles in this category are subject only to the above rules, as further categorization seems unrealistic at this time.

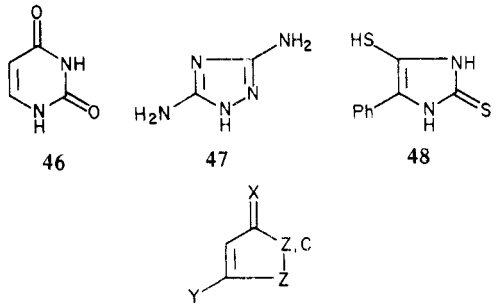


X, Y, Z = O, N, S, X ≠ Y

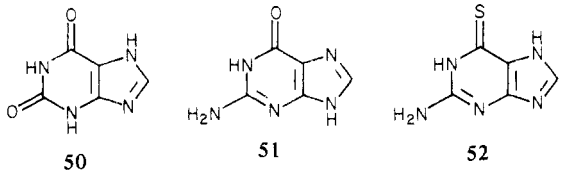
As in the case of **40**, hydrogen bonding can alter the predominating tautomer if the two EXONA are ortho. Thus ARTAUT would incorrectly tautomerize **45a** to **45b**. Elucidating a logical pattern to handle the occurrence and influence of hydrogen-bonded structures is a challenge left for the future.[21]



Exocyclic Tautomers: Two Like Hetero-EXONA. If two identical exocyclic heteroatoms are present, usually the amino–amino, mercaptothione, or dioxo tautomer predominates (**46–48**). If the heteroatoms are not identical, the following two rules apply: (1) having two endocyclic double bonds is generally less favored; (2) the tendency to retain the endocyclic double bond decreases in the order $NH_2 > SH > OH$. For example, the aminothione tautomer would be preferable to the corresponding thiolimine.[42] ARTAUT approximates these rules by tautomerizing only those rings which have nitrogen for *both* exocyclic heteroatoms, when the ring is of the type **49**. Dioxo–thione–imine structures are not tautomerized as a two-step process is required to achieve aromaticity. Since stable dioxo/thione tautomers exist, and may predominate,[43] this rule appears to be satisfactory. Five–six condensed rings follow the same patterns; **50–52**[39-41] exist as such.



X, Y, Z = O, N, S



CONCLUSION

New methods for ring, aromaticity, and tautomer perception have been described. Direct perception of the SSSR, by a tree growing algorithm, has been shown to be feasible. Processing atoms as sets enables ARTAUT to classify aromatics, nonaromatics, and tautomeric structures efficiently. General rules regarding the tautomerization of heterocycles have been presented and their limitations noted. As our predictive ability

in heterocyclic chemistry improves, it will be possible to formulate more precise and rigorous rules.

ACKNOWLEDGMENT

Gratitude is expressed to the National Science Foundation and Research Corporation for support of this work. The assistance of Timothy D. Salatin in developing the CAMEO program is also gratefully acknowledged.

REFERENCES AND NOTES

(1) Camille and Henry Dreyfus Foundation Teacher-Scholar, 1978-1983; Alfred P. Sloan Foundation fellow, 1979-1981.
(2) (a) Salatin, T. D.; Jorgensen, W. L. *J. Org. Chem.* **1980**, *45*, 2043. (b) Corey, E. J.; Wipke, W. T. *Science (Washington, D.C.)* **1969**, *166*, 178. (c) Corey, E. J.; Wipke, W. T.; Cramer, R. D.; Howe, W. J. *J. Am. Chem. Soc.* **1972**, *94*, 421, 431. (d) Wipke, W. T.; Dyott, T. M. *Ibid.* **1974**, *96*, 4825, 4834. (e) Corey, E. J.; Petersson, G. A. *Ibid.* **1972**, *94*, 460.
(3) Frérejacque, M. *Bull. Soc. Chim. Fr.* **1939**, *6*, 1008-1011.
(4) Gasteiger, J.; Jochim, C. *J. Chem. Inf. Comput. Sci.* **1979**, *19*, 43-48.
(5) Plotkin, M. *J. Chem. Doc.* **1971**, *11*, 60-63.
(6) Wipke, W. T.; Dyott, T. M. *J.* Chem. *Inf. Comput. Sci.* **1975**, *15*, 140-147.
(7) Zamora, A. *J. Chem. Inf. Comput. Sci.* **1976**, *16*, 40-43.
(8) Schmidt, B.; Fleischhauer, J. *J. Chem. Inf. Comput. Sci.* **1978**, *18*, 204-206.
(9) Paton, K. *Commun. ACM* **1969**, *12*, 594-598.
(10) Gibbs, N. E. *J. Assoc. Comput. Mach.* **1969**, *16*, 564-568.
(11) Dewar, M. J. S.; Gleicher, P. *J. Am. Chem. Soc.* **1965**, *87*, 685.
(12) Paquette, L. A. "Principles of Modern Heterocyclic Chemistry"; W. A. Benjamin: New York, 1968; pp 184-185.
(13) Forsen, S.; Wilsson, M. "The Chemistry of the Carbonyl Group"; J. Zabicky, Ed., Interscience: New York, 1970; Vol. 2, p 135.
(14) Garratt, P. J. "Aromaticity"; McGraw-Hill: London, **1971**; p 131.
(15) Gilman, H.; Melstrom, D. S. *J. Am. Chem. Soc.* **1948**, *70*, 1655.
(16) Adams, A.; Slack, R. *J. Chem. Soc.* **1959**, 3061; *Chem. Ind. (London)* **1956**, 1232.
(17) Reference 21, p 135.
(18) Reference 13, p 167-183.
(19) Cook, M. J.; Katritzky, A. R.; Linda, P.; Tack, R. D. *J. Chem. Soc. Perkin Trans. 2* **1972**, 1295.
(20) Beak, P.; Covington, J. B. *J. Am. Chem. Soc.* **1978**, *100*, 3961.
(21) Katritzky, A. R.; Boulton, A. J., Eds., "Tautomerism of Heterocycles"; Academic Press: New York, 1976; pp 204-212, 249-250, 388-389, 414-415, 444-445.
(22) Katritzky, A. R. *Chimia* **1970**, *24*, 134.
(23) Reference 12, p 205.
(24) Frank, J.; Katritzky, A. R. J. *Chem.* Soc. *Perkin Trans.* 2 1976, 1428.
(25) Reference 13, pp 195-197.
(26) Reference 21, pp 388-389.
(27) Baba, H.; Takemura, T. *Bull. Chem. Soc. Jpn.* **1964**, *37*, 124. Fieser, L. F. *J. Am. Chem. Soc.* **1931**, *53*, 2329. Clar, E. *Chem. Ber.* **1949**, *82*, 495.
(28) Hornfeldt, A. B. *Arkiv. Kemi* **1968**, *29*, 461.
(29) Reference 21, pp 405-415.
(30) Tanaka, T. *J. Pharm. Soc. Jpn.* **1971**, *91*, 338.
(31) Reference 21, 443-445.
(32) Reference 21, pp 303-304.
(33) This prevents tautomerizing functional groups such as oximes.
(34) Schulenberg, J. W. *J. Am. Chem. Soc.* **1968**, *90*, 7008.
(35) Mazharuddin, M.; Thyagarajan, G. *Tetrahedron* **1969**, *25*, 517.
(36) Mason, S. F. *J. Chem. Soc.* **1954**, 2071.
(37) Lister, J. H. *Adv. Heterocycl. Chem.* **1966**, *6*, 1.
(38) Pal, B. C.; Horton, C. A. *J. Chem. Soc.* **1964**, 400.
(39) Cavalier, L. F.; Fox, J. J.; Stone, A.; Change, N. *J. Am. Chem. Soc.* **1954**, *76*, 1119.
(40) Angell, C. L. *J. Chem. Soc.* **1951**, 504.
(41) Bugg, C. E.; Thewalt, U. *J. Am. Chem. Soc.* **1970**, *93*, 7441.
(42) Reference 21, pp 211, 253, 484-485.
(43) Reference 21, p 250.

CBF—Computer Handling of Chemical and Biological Facts. 2[1]

J. BECKER and D. JUNG*

Electronic Data Processing Department, Boehringer Ingelheim KG, 6507 Ingelheim am Rhein, Federal Republic of Germany

W. KALBFLEISCH and G. OHNACKER

Research Administration Department, Dr. Karl Thomae GmbH, 7950 Biberach an der Riss, Federal Republic of Germany

Received July 29, 1980

CBF[1] is an EDP-supported documentation and retrieval system for structural formulas of defined organic compounds and their biological activities. It fits the needs of a firm concerned with drug research for prevailing unpublished internal data, which the scientists can use for reflections on structure-activity relationships and to search for lead compounds with special activity profiles.

INTRODUCTION

The CBF system has been used successfully for 12 years by various research centers of C. H. Boehringer Sohn, Ingelheim. It was conceived as a data input system for

(1) storing biological screening results from drug research in a computerized data base,

(2) providing printed information about chemical compounds and/or screening results either as a continous service in file-card form or as printouts of results of retrospective searches.

Our files contain connectivity tables of 170 000 chemical structural formulas as well as 260 000 individual results from biological screening tests of 78 000 substances. As could be expected, everyday use of this system over the years has illuminated several features requiring improvement. Innovations were undertaken with respect to the program, and certain alterations to individual elements of the entire system were introduced. The aim of such improvements was to rationalize data input and to expand search capabilities.

INPUT OF CHEMICAL INFORMATION

We are using the methods of machine transformation into a condensed connectivity table to store all structural information unambiguously.[2,3] Together with the connectivity table, a series of screens is machine generated. Structural formulas requiring maximally 256 nodes (nonhydrogen atoms) in the topological list can be currently handled.

As a consequence of topological storage, only unambiguously defined structures can be stored in a retrievable form. Substances whose structures are equivocal can be stored either in

0095-2338/81/1621-0111$01.25/0