An Algorithm for the Perception of Synthetically Important Rings

JOHANN GASTEIGER* and CLEMENS JOCHUM

Institut für Organische Chemie, Technische Universität München, D-8046 Garching, West Germany

An algorithm for finding the smallest set of smallest rings (SSSR) is presented. The algorithm is highly efficient and performs well even in those cases where other approaches fail. The SSSR offers particular advantages for chemical purposes. It is best to construct synthetically important rings from the SSSR. Considering all rings up to a certain size as synthetically important is cumbersome and inappropriate.

I. INTRODUCTION

Rings are of fundamental importance in the perception of chemical structures. This is already reflected in the fact that the nomenclature of organic molecules pays tribute to the occurrence of rings. The presence of rings may be of profound influence on the physical or chemical properties of organic compounds. Cyclic structures can be much higher in energy than their open-chain analogs, a phenomenon attributed to ring strain. Further, aromaticity of molecules is necessarily tied to the occurrence of rings of particular sizes. These effects, and the steric restrictions and proximities imposed on a cyclic structure, may result in reactivities which are many orders of magnitude different from comparable open-chain molecules. In the synthesis of cyclic structures the construction of the cyclic skeleton is usually the most difficult task and, therefore, the greater part of synthetic efforts is dedicated to this problem.

Traditionally, the perception of rings in molecules was delegated to the human pattern recognizer. This is successful with many structures. But for more complex molecules this procedure may be quite laborious and the set of rings found can depend on the human individual and the way the structure is drawn in two dimensions.

In the representation and manipulation of chemical structures by computers an automatic procedure for the recognition of rings is indispensable. As we were developing computer programs for the solution of chemical problems we were also faced with the task of recognizing rings in organic molecules. This is particularly true for our synthetic design program, EROS,[1] and for the automatic estimation of ring strain energies.[2]

As the constitution of a chemical compound can be viewed as a graph, graph theoretical concepts can be applied to the detection of rings. Graph theorists have developed algorithms for the recognition of cycles in graphs, not only in problems of chemical constitution but also in problems of electrical engineering and operations research. There have been various approaches to find cycles in directed and undirected graphs. We restrict our considerations to undirected graphs since we are primarily concerned with the detection of rings in chemical structures. Although we will speak mostly of chemical compounds, our algorithm can be applied to any undirected graph.

The problem involves not only the task of developing an algorithm for the detection of rings but also the definition of a suitable set of rings. For, if all rings are taken into account, an unreasonably high number of rings may result.

For example, for prismane one obtains 14 rings, and for cubane a total of 28 different rings can be found.

On the other hand, the minimum number of rings required to describe a ring system (the smallest set; Frèrejacque[3] number) is given by the equation

$$\text{no. of rings} = \text{no. of bonds} - \text{no. of atoms} + 1$$

Thus, a definition for a suitable set of rings in a molecule or graph has to be given.

II. SELECTION OF A SET OF RINGS

Three different basic approaches to this problem can be found in the literature.

A Fundamental Set of Cycles of a Graph. A vector $v_i = (v_{ij}, \ldots, v_{ik})$ can be assigned to each cycle or ring r_i of a graph or molecule with k edges:

$$v_{ij} = 1 \text{ if } j \text{ edge of ring } r_i$$
$$= 0 \text{ otherwise}$$

These vectors v_i ($i = 1, 2, \ldots$) generate a vector space of rings. Let n be the minimum number of rings required to describe the ring system (Frèrejacque number[3]) and $(r_1, \ldots, r_n$ be a set of rings whose corresponding vectors $v_1, \ldots, v_n$ are linearly independent. Then the vectors $v_1, \ldots, v_n$ build a basis of the vector space and the set of rings $r_1, \ldots, r_n$ is called a fundamental set of rings. Note, that there can exist several fundamental sets of rings.

Several authors[4-7] present algorithms for the generation of a spanning tree of a given graph and the subsequent construction of a fundamental set of cycles. Depending on the numbering of the atoms of a molecule, different fundamental sets of cycles can be obtained. Welch[6] and Gibbs[7] give additional algorithms for the construction of all rings of a given graph (which can be quite a substantial number) by linear combination of the rings of the fundamental set. Gibbs could show that Welch's algorithm contains an invalid assumption, which has the consequence that occasionally not all rings are found. Such a fundamental set, however, does not necessarily contain only the smallest possible rings. For example, see bicyclo[2.1.0]pentane (**1**). The rings r_1 and r_2 represented

1
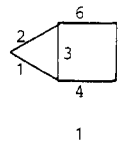
by $v_1 = (1,1,1,0,0,0)$ and $v_2 = (1,1,0,1,1,1)$ give a basis of **1**, but the four-membered ring which is contained in this basis is of more significance than the five-membered one. The two smallest rings r_1 and r_3, however, are given by v_1 and $v_3 = (0,0,1,1,1,1)$.

The Smallest Set of Smallest Rings (SSSR). Plotkin[8] gives a defintion of a particular fundamental set of cycles called the Smallest Set of Smallest Rings (SSSR). In his *procedure P* he constructs the SSSR from the complete set of rings:

All rings are ordered according to their ring size. The smallest ring is always assigned to the SSSR. Subsequent rings are only assigned if they are linearly independent of rings already assigned. As the SSSR is a fundamental set of rings, the order of the SSSR is given by the Frèrejacque number.

Two algorithms[8,9] have been developed which are intended to directly find the SSSR. They both have limitations for certain complex ring systems. Their efficiencies cannot be

compared exactly since the authors give only scant information about the computational results.

A Set of Rings Selected on Heuristics. Different definitions of chemically relevant subsets of the set of all rings of a molecule have been given.

Fugmann, Dölling, and Nickelsen[10] select a set of so-called "fundamental rings" (not to be confused with a fundamental set of cycles), which are all smallest rings of a molecule (i.e., rings, where a ring path is a shortest connection between two ring atoms and, in addition, certain bridged rings). The rules for defining these "fundamental rings" are rather complicated and time consuming to apply.[11b] Many paths between pairs of atoms have to be considered; e.g., for cubane (2) 384 paths have to be searched.

2

Corey and Wipke[11a] give a definition of a "maximum proper covering set of rings". This set is not always a fundamental set of rings since the "maximum proper covering set of rings" can include linearly dependent rings. Both authors later independently tried to improve their ring perception by developing new algorithms.[11b,14]

Corey and Petersson[11b] choose a particular fundamental set of cycles and include all additional rings with six or fewer atoms. Their particular fundamental set of cycles is the SSSR defined by Plotkin[8] for the first time.

Bersohn[12] and Esack[13] have used the same set as Corey[11b] but give differing algorithms.

Wipke and Dyott[14] present a review of the principles involved in ring perception and the attempts made by other authors. They improve on some of the previous algorithms[6,7] by making use of bit operations. As their basic "Welch - Assembly - Gibbs" algorithm is written in Fortran, they have to use assembler routines for implementing bit operations. As a new step they separate molecules comprised of several ring systems into the individual assemblies. The principles involved in their algorithms can be found in the preceding literature.[11b] For their final set of rings they choose the SSSR (which they call reduced basis set) and all other rings with eight or fewer atoms.

All approaches[11-14] mentioned in this paragraph, except Fugmann's,[10] have to generate *all* rings of a given system (which can amount to quite a substantial number) before selecting their particular set.

For further comparison the different sets of heuristically chosen rings are given for cubane (2). The SSSR consists of five four-membered rings. (The same number is considered in the nomenclature of this compound: pentacyclo-[4.2.0.0^{2,5}.0^{3,8}.0^{4,7}]octane.)

Fugmann's[10] algorithm detects six four-membered rings, Corey,[11b] Bersohn,[12] and Esack[13] find, in addition, 16 six-membered rings, and, finally, Wipke[14] obtains all 28 rings, by also including six eight-membered rings.

III. DISCUSSION

It is essential for the perception of rings in organic molecules to obtain a fundamental set of cycles. All algorithms initially aim at that objective. From among the fundamental sets an SSSR offers particular advantages for chemical purposes.

First, it can be noted that in the nomenclature of polycyclic compounds, in essence, the SSSR is used. This is true for the IUPAC nomenclature, the Ring Index,[15] and the Wiswesser line notation.[16]

Further, it is clear that ring strain is predominantly inherent in the smallest rings. Envelopes of rings are not appropriate

for the estimation of ring strain energies. Thus, in decalin, the ring strain of an enveloping ten-membered ring is not encountered. When estimating the ring strain of bicyclo-[2.1.0]pentane (1) by an additivity scheme as the sum of the strain energies of the rings in the SSSR, i.e., of a three- and four-membered ring, one obtains 54.0 kcal/mol which is only 1.3 kcal/mol short of the actual ring strain energy.[2] On the other hand, the sum of the strain energies of a three- and five-membered ring, which also comprise a fundamental set of rings (see above), gives 33.7 kcal/mol, a totally inadequate estimate.

For recognizing aromatic compounds, again, the SSSR is the best choice among the fundamental sets. For indan (3), one of the fundamental sets consists of a five- and nine-membered ring. This set does not contain the aromatic
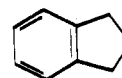

3

six-membered ring. The SSSR, on the other hand, consists of the five- and six-membered rings. Thus, the SSSR is a better fundamental set for the perception of aromatic rings.

Further, in synthetic planning, we are faced with the task of defining "synthetically important" rings. It is clear that for polycyclic compounds a large extent of the synthetic efforts has to be directed to the synthesis of the smallest rings. This means that the SSSR is again of interest when searching for synthetically important rings.

The question is now: Are there additional rings outside of the smallest set of smallest rings which are of synthetic interest? Clearly the answer is yes. The most obvious example is the norbornane skeleton. Many syntheses giving access to norbornene compounds make use of the powerful Diels–Alder reaction to synthesize this skeleton.
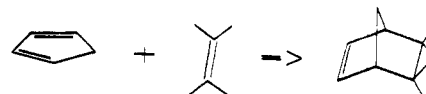
Therefore, in the synthetic analysis of a norbornane or norbornene compound the detection of a six-membered ring seems essential.

Recognizing this and similar situations has made various authors include all rings with up to six[11b-13] or even eight[14] atoms into their lists of synthetically important rings. Whereas this procedure allows finding a six-membered ring in the norbornane skeleton, many rings have to be carried and considered which are *not* synthetically important. Consider, for example, the cubane skeleton (2). There exist 12 six-membered rings of type 4 and four six-membered rings of type 5.
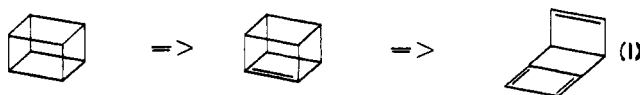

4   5

All these 16 six-membered rings are not important when seeking a synthesis *directly* leading to cubane. It cannot reasonably be expected that the Diels–Alder reaction could be the final step in constructing the cubane skeleton (eq 1).


6

The final step in the synthesis of the cubane skeleton will always be a reaction applicable to the generation of at least one four-membered ring. This does not mean that one of the six-membered rings in the cubane system cannot be generated by a Diels–Alder reaction. But this Diels–Alder reaction could only result in an intermediate not having the entire cubane skeleton. In fact, the first synthesis[17] of a cubane derivate did construct such a six-membered ring by a Diels–Alder reaction (see eq 2), but it would not have been necessary to detect that
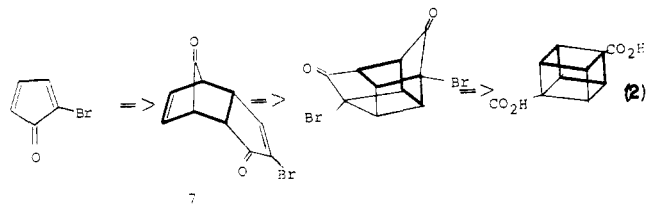
six-membered ring until we had arrived at the intermediate **7** in a synthetic analysis starting from cubane. With this intermediate **7** the problem of synthetic analysis is reduced to the norbornene problem.

To further clarify the situation, let us study all bicyclic ring systems where there exists a six-membered ring which is not part of the SSSR.

**1. The Bicyclo [3.1.0] hexane (8) Skeleton.** Cyclization
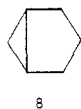
reactions which are peculiar to a six-membered ring, e.g., the Diels–Alder reaction, are of no use in constructing this skeleton. Again it does not mean that in some previous steps in a multistep synthesis of a bicyclo[3.1.0]hexane skeleton a Diels–Alder reaction or another reaction which proceeds with particular ease to a six-membered ring could not have been employed, but it would have resulted in an intermediate which contained the six-membered ring in the SSSR. Thus, a synthesis of the bicyclo[3.1.0]hexane derivative **9** did use a cyclization reaction to a six-membered ring, but the final step in the synthesis of this skeleton did not construct a six-membered ring but the three-membered one[18] (see eq 3).
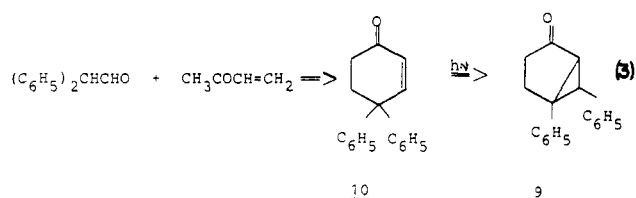
Therefore, in planning a synthesis of **9** the detection of the three-membered ring generates in the retrosynthetic approach the precursor **10**. From **10** it is the six-membered ring which allows finding the appropriate precursors. In both molecules, **9** and **10**, the synthetically important rings are contained in the respective SSSR.

**2. The Bicyclo[2.2.0]hexane (11) Skeleton.** Here, the two
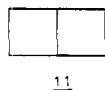
four-membered rings are of prevailing synthetic importance (eq 4). For **12** it is the detection of the cyclobutene entity which leads through a retroelectrocyclic step (path A) to the 1,3-butadiene substructure of 1,3-cyclohexadiene. Formally, one could think of obtaining that skeleton by a Diels–Alder type reaction (path B). Reactions of the type B are known.[19]
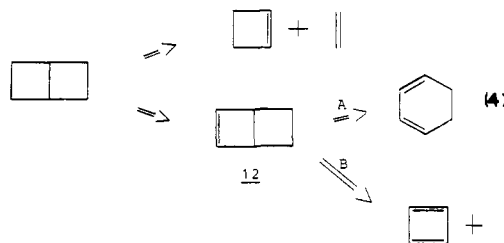
But in the case of cubane the same reaction (path B) must fail because of the geometric restraints in the necessary precursor **6** (see eq 1). We therefore believe that a careful examination of the additional rings present around the bicyclo[2.2.0]hexane entity must decide whether the Diels–Alder reaction can be applied. This analysis can be based on the SSSR. Obviously such an analysis is more rewarding than finding 12 rings of type **4** and applying the retro-Diels–Alder reaction yielding 12 precursors of type **6** which are too strained to react to **2**.

3. The Bicyclo[2.2.1]heptane Skeleton. As already
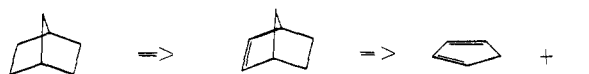
mentioned, the Diels–Alder reaction is a very powerful synthetic tool for constructing this skeleton. This skeleton presumably was the principal cause for Corey and Peterson[11] to include all six-membered rings in their set of rings. We suggest looking for that skeleton directly rather than blindly taking any six-membered rings. The SSSR of this skeleton contains two five-membered rings. Thus, this skeleton can be found by searching for combinations of two five-membered rings which have three atoms in common. These search operations can be performed with the information contained in an SSSR (see below). Along similar lines it can be demonstrated that for the planning of syntheses for the bicyclo[1.1.0]butane and bicyclo[2.1.0]pentane (**1**) skeleton the SSSR is sufficient.

We emphasize that the perception of rings and the application of information about rings are two conceptually entirely different tasks. For the perception of rings a fundamental set is asked for; from among these the SSSR offers the greatest advantages. For many applications (nomenclature, ring strain energies etc.) the SSSR is sufficient to work with. With other applications (e.g., synthetic strategy) a more careful analysis of the entire ring system is asked for. This task cannot be accomplished by an arbitrary selection of all rings up to a certain size.

IV. DEFINITIONS

Before we present our algorithm for detecting the SSSR some frequently used definitions are given:

1. A *bond* is given by a pair (a_i, a_j) of connected atoms a_i and a_j (the bond order is irrelevant when searching for rings).

2. A *path* is defined as a series of bonds (a_1, a_2), (a_2, a_3), . . ., (a_{k-2}, a_{k-1}), (a_{k-1}, a_k).

3. The *distance* $d_{aa}(a_i, a_j)$ between two atoms a_i and a_j of a molecule M is defined as the minimal number of bonds that connect a_i and a_j.

4. The distance $d_{ab}(a_k, (a_i, a_j))$ between the atom a_k and the bond (a_i, a_j) is defined as the smaller of the two distances between atoms $d_{aa}(a_k, a_i)$ and $d_{aa}(a_k, a_j)$.

5. The kth *sphere* $S_k(a_i)$ of an atom $a_i \in M$ is the set of atoms which have distances k to a_i, i.e.

$$S_k(a_i): = \{a_j | a_j \in M, d(a_i, a_j) = k\}$$
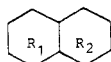
Example :
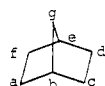
13

$$d(a,g) = 2$$
$$S_2(a) = \{c,e,g\}$$

Let R be a ring system $\{R_1, \ldots, R_k\}$ composed of rings $R_1, \ldots, R_k$.

7. SREL is defined as the sum of the number of atoms in every ring of the ring system:

$$SREL(R): = \sum_{i=1}^{k} |R_i|$$

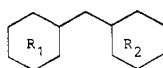8. SEL is defined as the sum of all atoms that belong to at least one ring.

Example:

14

$$R = \{R_1, R_2\} \qquad R_1 = R_2 = 6$$
$$SREL(R) = 12 \qquad SEL(R) = 10$$

15

$$SREL(\underline{15}) = SEL(\underline{15}) = 12$$

9. The *complexity* CR(R) of a ring system R is defined as the ratio of SREL to SEL, i.e.

$$CR(R): = SREL(R) / SEL(R)$$

Example: Let R(**13**) be $\{(a,b,g,e,f),(b,c,d,e,g)\}$
Then $CR(\mathbf{13}) = 10/7 = 1.43$
For cubane (**2**) let $R = \{4,4,4,4,4\}$
Then $CR(\mathbf{2}) = 20/8 = 2.5$

V. THE ALGORITHM

The algorithm is implemented in two stages: a basic ring-finding algorithm and a procedure for generating the SSSR.

The Basic Algorithm. The first step is the generation of a particular spanning tree. An arbitrary reference atom is chosen as the root of the tree. Then the distance of all the other atoms to this reference atom is determined. Atoms having the same distance to the reference atom define a neighbor sphere. For each atom within a certain neighbor sphere only one bond from the set of bonds of this atom to the sphere of lower order is flagged.

The remaining bonds are ring closure bonds and are ordered according to their distance to the reference atom. From each ring closure bond a back tracing to spheres of lower order, i.e., in the direction to the reference atom, is started to find the other atoms of the ring which includes the ring closure bond. This is accomplished by scanning the two branches starting from the ends of the ring closure bond to lower spheres for atoms which they have in common. There is at least one such atom, the reference atom. The first common atom which is detected by this procedure is the common atom which lies closest to the ring closure bond. The atoms of the two paths
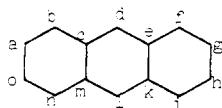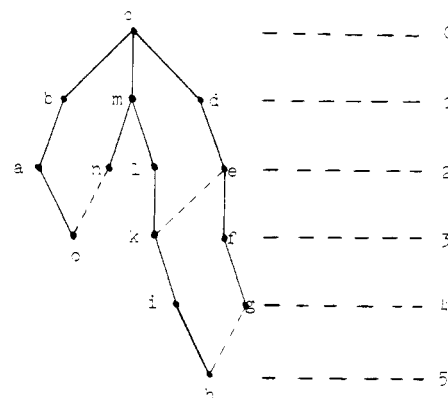

16

Table I

corresponding to the shortest distance of that common atom to the atoms of the ring closure bond are the other ring atoms; see, for example, the anthracene skeleton (**16**),
c is arbitrarily chosen as reference atom. Then the following tree is obtained:

The ring closure bonds can be ordered in the sequence (o,n), (k,l), (h,g); see Table I. Note, that the ring found is not bound to be the smallest ring containing the ring closure bond. There might be a smaller ring within the sphere of the ring closure bond or in the direction to spheres of higher order. On the other hand, this procedure assures the detection of a fundamental set (see below).

The basic strategy can be summarized in the following concise algorithm:

Step 1: Choose an arbitrary reference atom a_1 and generate a spanning tree starting from this atom.

Step 2: Order the ring closure bonds according to their distance to a_1.

Step 3: Take a ring closure bond (a_k, a_l) not yet considered. If there is no such bond stop. The set of ring atoms belonging to this bond may be R_j. Set $R_j = \{a_k, a_l\}$.

Step 4: Generate the two branches starting from a_k and a_l, respectively, proceeding to lower spheres, until a common atom is found.

Step 5: Take this common atom a_{min} with the minimal distance to a_k and a_l and set $R_j = R_j \cup \{a_{min}\}$.

Step 6: Trace back from a_{min} to a_k and a_l, respectively, on two paths corresponding to the shortest distances and add all atoms on these paths to R_j.

Step 7: Go to step 3.

Since this algorithm basically uses the concept of a spanning tree it must necessarily detect the same number of rings as

there are in a fundamental set. Additionally, the algorithm always generates a fundamental set of rings.

Proof: With each new ring at least one new bond, the ring closure bond is added to the bonds of the rings already detected. Thus, the dimensions of the ring vector space is enlarged by 1 with each additional ring. Therefore all vectors corresponding to these rings must be linearly independent. Thus, we see that both conditions of a fundamental set of rings are satisfied: linear independence of the corresponding vectors and the right order of the set.

The set of rings found with the basic algorithm is in most cases an SSSR. Even in rather complicated structures as **25** and **26** in Table II where Zamora's algorithm[9] fails, an SSSR can be found with this basic algorithm.

However, there are cases where some of the rings found in the fundamental set are larger than those in the SSSR. This can happen when a comparatively small ring of the SSSR lies completely in one neighbor sphere of a particular reference atom. See, for example, the bullvalene skeleton (**17**). Ring
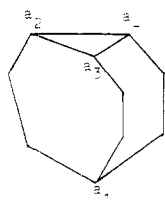
$\{a_2,a_3,a_4\}$ lies completely in $S_3(a_1)$. The basic algorithm would detect three seven–membered rings. SSSR = (3,7,7).

The SSSR Algorithm. To generate the SSSR even for such rare cases as structure **17**, the following algorithm was constructed.

Step 1: If there are no rings stop.

Step 2: Choose an arbitrary reference atom.

Step 3: Apply the basic algorithm.

Step 4: Determine the complexity CR of the ring system.

Step 5: If CR ≤ 1.5 stop; else go to step 6.

Step 6: Choose another reference atom a_i which satisfies the following conditions:

(a) a_i is an element of a smallest ring, which does not contain previous reference atoms. If this is not possible take an arbitrary new reference atom.

(b) a_i has from all the atoms in this ring the largest number of ring atoms as neighbors.

Step 7: Apply the basic algorithm with a_i as reference atom. If there have been three calls of the basic algorithm go to step 9.

Step 8: If CR > 2 go to step 6.

Step 9: Choose an SSSR from the sets of rings detected after two or three calls of the basic algorithm and stop.

Step 9 is executed by procedure[8] P mentioned in part II. There, procedure P is only given to define exactly the SSSR. No corresponding algorithm was developed as the set of all rings would have to be scanned. We, however, generally have to consider only a subset of all rings. Thus, an implementation of procedure P appeared feasible. In addition, choosing a binary Gauss Elimination algorithm for detecting linear independence ensures high performance in speed and storage.

We have not yet encountered any example where our algorithm did not correctly determine the SSSR. But as we do not use the set of all rings we concede that for very complex graphs our fundamental set could be different from an SSSR. We believe that these complex graphs will hardly be en-

countered with chemical structures.

In any case, however, our SSSR algorithm could be easily extended to calling the basic algorithm again for ring systems of high complexity.

Additional Features. The detection of a fundamental set of cycles and the elaboration of the SSSR have been separated conceptually in our algorithm. In the process of extracting the SSSR from among several fundamental sets of rings, additional rings are recognized. Usually these rings are discarded. But as these rings do not constitute enveloping rings they can be of interest for certain applications; thus, they are retained and stored.

For example, for cubane (**2**) the SSSR comprises only five four-membered rings. But as the complexity of this ring system is rather high (20/8 = 2.5), the basic algorithm is called three times, each time with a different reference atom. Each call provides a fundamental set of five four-membered rings. As the reference atoms change, these five rings are not all the same ones. Thus, in the sorting of the rings all six four-membered rings are detected although only five are retained in the SSSR.

When estimating the ring strain energy of cubane it is advantageous to consider all six four-membered rings. The ring strain energy of cubane (**2**) amounts to 162.7 kcal/mol. The ring strain energies of six four-membered rings add up to 6 × 26.5 = 159.0 kcal/mol which is a rather good estimate for the experimental value.[2]

Until now we have started with an arbitrary reference atom. We are exploring whether the choice of a particular initial reference atom improves on computation times and convergence to the SSSR. The values of the complexities of the ring systems which decide on additional calls of the basic algorithm are also further scrutinized. Until now we have always obtained the SSSR. Therefore we believe that the boundaries for the complexities will only be changed when further improvements in execution times result.
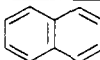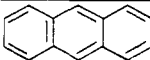
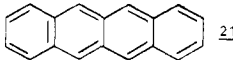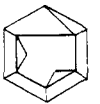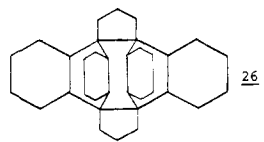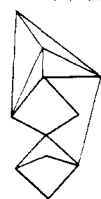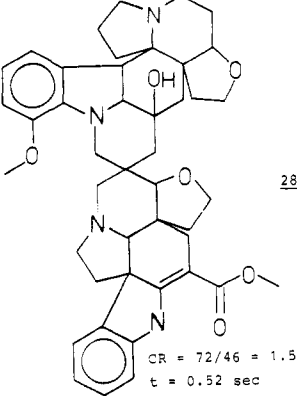## VI. IMPLEMENTATION AND RESULTS
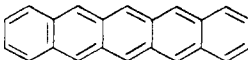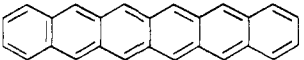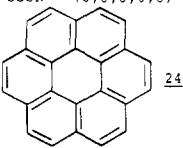
Our algorithm was implemented in PL/1. This language provides several features which make it especially useful for our purposes. First, it has powerful bit-handling capabilities. Many operations in the algorithm, e.g., the storage of rings, the back-tracing of branches, and the Gauss elimination procedure, are performed on bit vectors. When coding an algorithm in Fortran, a language which is poorly designed for bit operations, recourse has to be taken to assembler routines.[14] This makes the entire algorithm dependent on specific machine types.

As another advantage, PL/1 allows dynamic allocation of space. Therefore no restrictions on the number of atoms in a molecule or the number of rings in the system have to be imposed. And finally, PL/1 permits the implementation of recursive procedures which save program code and perform dynamic allocation of space, too. Recursive procedures are used for the determination of neighbor spheres and for the back-tracing procedure.

Molecules are represented by compact connection tables allowing direct access to each atom and bond of a molecule. The entire procedure consists of 318 PL/1 statements. Storage requirements are small as most vectors allocated during execution of the program are packed (unaligned) bit vectors. The procedure is routinely used in our synthetic design program EROS.[1]

The program is implemented on an IBM 360/91 and an AMDAHL 470 V6. Execution times given were obtained with the machine code of a PL/1 F-compiler and refer to an AMDAHL 470 V6. Table II presents some examples together with values for the complexity of the ring systems and the computation times.

Table II

_18_
CR = 6/6 = 1
t = 0.03 sec
SSSR = (6)

_19_
CR = 12/10 = 1.2
t = 0.05 sec
SSSR = (6,6)

_20_
CR = 18/14 = 1.29
t = 0.1 sec
SSSR = (6,6,6)

_21_
CR = 24/18 = 1.33
t = 0.13 sec
SSSR = (6,6,6,6)

_22_
CR = 30/22 = 1.36
t = 0.24 sec
SSSR = (6,6,6,6,6)

_23_
CR = 36/26 = 1.38
t = 0.3 sec
SSSR = (6,6,6,6,6,6)

_24_
CR = 42/24 = 1.75
t = 0.22 sec
SSSR = (6,6,6,6,6,6,6)

_25_
CR = 34/13 = 2.62
t = 0.17 sec
SSSR = (3,3,3,4,4,4,4,4,5)

_26_
CR = 54/34 = 1.59
t = 0.35 sec
SSSR = (5,5,6,6,6,6,6,6,8)

_27_
CR = 29/11 = 2.64
t = 0.12 sec
SSSR = (3,3,3,3,4,4,4,5)

_28_
CR = 72/46 = 1.57
t = 0.52 sec
SSSR = (5,5,5,5,5,5,
6,6,6,6,6,6,6)

Naturally, computation times increase with the size of the ring system. The complexity of a ring system is not of paramount importance as a comparison of coronene with pentacene and hexacene shows. Although the number of carbon atoms of coronene is inbetween the respective numbers of pentacene and hexacene and the complexity of the ring system is much higher, the rings in coronene are found in less time. On the other hand, the largest distance between two atoms in pentacene and hexacene is higher than in coronene. Apparently, it is the determination of the neighbor spheres which consumes much of the execution time.

In our program for canonical numbering of the atoms of a molecule,[20] the neighbor spheres are also determined. This information can be transferred to the ring-finding algorithm decreasing considerably execution times.

VII. CONCLUSION

It is clear that a complete analysis of a ring system asks for a fundamental set of rings. From among these, a smallest set of smallest rings offers particular advantages for chemical purposes. For many applications an SSSR is sufficient. In synthetic analysis some additional rings have to be considered. As the type of these additional rings is dependent on the topology of the entire ring system it is best to generate these rings directly from the SSSR. To aim at these rings by carrying along and considering all rings up to a certain size is too unspecific. Too many rings have to be analyzed which do not bear any significance in synthetic planning but lead to inaccessible precursors. We present an algorithm for detecting the SSSR which, in most cases, provides these rings in one cycle. For more complex cases several calls of the basic algorithm assure the finding of the SSSR. In any case, it is not necessary to first generate *all* rings and then extract the SSSR as many authors do.

We also present an exact definition of the complexity of a ring system which appears to have merits for other purposes, too.

REFERENCES AND NOTES

(1) J. Gasteiger and C. Jochum, *Top. Curr. Chem.*, **74**, 93 (1978).
(2) J. Gasteiger and O. Dammer, *Tetrahedron*, **34**, 2939 (1978).
(3) M. Frèrejacque, *Bull. Soc. Chim. Fr.*, **6**, 1008–1011 (1939).
(4) C. C. Gotlieb and D. G. Corneil, *Commun. ACM*, **10**, 780–783 (1967).
(5) K. Paton, *Commun. ACM*, **12**, 514–518 (1969)
(6) J. T. Welch, *J. Assoc. Comput. Mach.*, **13**, 205–210 (1966).
(7) N. E. Gibbs, *J. Assoc. Comput. Mach.*, **16**, 564–568 (1969).
(8) M. Plotkin, *J. Chem. Doc.*, **11**, 60–63 (1971).
(9) A. Zamora, *J. Chem. Inf. Comput. Sci.*, **16**, 40–43 (1976).
(10) R. Fugmann, U. Dölling, and H. Nickelsen, *Angew. Chem., Int. Ed. Engl.*, **6**, 723 (1967).
(11) (a) E. J. Corey and W. T. Wipke, *Science*, **166**, 178 (1969); (b) E. J. Corey and G. A. Petersson, *J. Am. Chem. Soc.*, **94**, 460–465 (1972).
(12) M. Bersohn, *J. Chem. Soc., Perkin Trans. 1*, 1239–1241 (1973).
(13) A. Esack, *J. Chem. Soc. Perkin Trans. 1*, 1120–1129 (1975).
(14) W. T. Wipke and T. M. Dyott, *J. Chem. Inf. Comput. Sci.*, **15**, 140–147 (1975).
(15) A. M. Patterson, L. T. Capell, and D. F. Walker, "The Ring Index", 2nd ed, American Chemical Society, Washington, D.C., 1960.
(16) E. G. Smith, "The Wiswesser Line-Formula Chemical Notation", McGraw-Hill, New York, N.Y., 1968.
(17) P. E. Eaton and T. W. Cole, Jr., *J. Am. Chem. Soc.*, **86**, 963 (1964).
(18) H. E. Zimmerman, R. D. Rieke, and J. R. Scheffer, *J. Am. Chem. Soc.*, **89**, 2033–2047 (1967).
(19) L. Watts, J. D. Fitzpatrick, and R. Pettit, *J. Am. Chem. Soc.*, **88**, 623 (1966).
(20) C. Jochum and J. Gasteiger, *J. Chem. Inf. Comput. Sci.*, **17**, 113–117 (1977).