that shown in eq 1, so that the reported time constants might refer to some process other than the dissociation of the phosphorane. This possibility can be disposed of on two counts. First, the temperature-dependent NMR spectra were determined in the presence of a quantity of the phosphonium salt equivalent to that of the phosphorane. Since the equilibrium for the reaction of the cation with phenoxide is even more favorable than that with the phosphorane,[2] the rate is also probably much greater. Second, the reaction between methyltriphenoxyphosphonium triflate and phenoxide ion was carried at two different concentrations, 0.314 and 0.153 M, and yielded practically the same rate constants. Such would not have been the case had the dominant reaction been second order in phosphorane, as required by the hypothesis that the hexacovalent anion is involved in the measured rate.

In the accompanying paper,[2] the equilibrium constants have been obtained for the dissociation of three of the phosphoranes. Using these equilibrium constants and the rate data presented here, the conclusion has been drawn that the reaction of these phosphonium salts with phenoxide ion in acetonitrile proceeds with the speed of collision. An extension of these results to other nucleophiles and other solvents is in progress.

Acknowledgment. This work was supported by the National Science Foundation, under Grant No. GP 2098. D.P. gratefully acknowledges a predoctoral fellowship from the National Institutes of Health. I.S. recieved financial support from the Zlato and Joyce Baloković Fund, administered by the Yugoslav Academy of Sciences and Harvard University, while on leave of absence from the Rudjer Bošković Institute of Zagreb. Special thanks go to Mr. Hampar Janjigian, and Dr. Wm. Hull for assistance with the NMR measurements, and to the NSF under Grant No. GP 30965X for assistance in the purchase of the NMR instrument. We are also indebted to Professors J. D. Roberts and Delos DeTar for helpful discussions, and to Charles Lerman for assistance with the computer programming.

References and Notes

(1) Author to whom correspondence should be addressed.
(2) C. L. Lerman and F. H. Westheimer, *J. Am. Chem. Soc.*, preceding paper in this issue.
(3) T. C. Bruice and S. Benkovic, "Bioorganic Mechanism", Vol. II, W. A. Benjamin, New York, N.Y., 1966, Chapter 1.
(4) F. H. Westheimer, *Acc. Chem. Res.*, **1**, 70 (1968).
(5) I. Ugi, D. Marquarding, H. Klusacek, R. Gillespie, and F. Ramirez, *Acc. Chem. Res.*, **4**, 288 (1971).
(6) D. Phillips, Thesis, Harvard University, 1972.
(7) W. Strecker and C. Grossmann, *Chem. Ber.*, **49**, 63 (1916); W. Broeker, *J. Prakt. Chem.*, (2) **118**, 287 (1928).
(8) J. Baddiley, V. M. Clark, J. J. Michalski, and A. R. Todd, *J. Chem. Soc.*, 815 (1949).
(9) A. N. Meldrum and M. M. Patel, *J. Indian Chem. Soc.*, **5**, 91 (1928).
(10) A. Michaelis and W. LaCoste, *Ber.* **18**, 2116 (1885); L. V. Nesterov and A. Ya. Kessel, *J. Gen. Chem. USSR* (*Engl. Transl.*), **36**, 1825 (1966).
(11) G. M. Kosolapoff, "Organophosphorus Compounds", Wiley, New York, N.Y., 1950, p 359.
(12) A. Michaelis and R. Kaehne, *Ber.*, **31**, 1048 (1898); L. V. Nesterov and A. Ya. Kessel, *J. Gen. Chem. of USSR* (*Engl. Transl.*), **36**, 1825 (1966).
(13) R. N. Hazeldine and J. M. Kidd, *J. Chem. Soc.*, 4228 (1954).
(14) A. Michaelis and H. Soden, *Justus Liebigs Ann. Chem.*, **229**, 316 (1885).
(15) *Chem. Abstr.*, **51**, 5720i (1957).
(16) A. E. Arbuzon and L. V. Nesterov, Academy of Sciences USSR, Chemical Science Bulletin, 1954, p 361.
(17) A. Weissberger, E. S. Proskauer, J. A. Riddick, and E. E. Toops, Jr., "Technique of Organic Chemistry", Vol. VII, Interscience, New York, N.Y., 1955, p 365.
(18) A. L. VanGeet, *Anal. Chem.*, **40**, 2227 (1968); **42**, 679 (1970).
(19) Cf. C. K. Tseng, Thesis, Illinois Institute of Technology, 1968.
(20) Both the triflate and the fluoroborate were prepared; they gave identical NMR spectra.
(21) G. Wittig and O. Hesse, *Org. Synth.*, **50**, 66 (1970).
(22) E. Dennis and F. H. Westheimer, *J. Am. Chem. Soc.*, **88**, 3431, 3432 (1966).
(23) R. K. Oran and S. Trippett, *J. Chem. Soc., Trans. 1*, 1300 (1973); S. A. Bone, S. Trippett, M. W. White, and P. J. Whittle, *Tetrahedron Lett.*, 1795 (1974).
(24) J. H. Finley, D. Z. Denney, and D. B. Denney, *J. Am. Chem. Soc.*, **91**, 5826 (1969).
(25) H. S. Gutowsky, D. W. McCall, and C. P. Slichter, *J. Chem. Phys.*, **21**, 279 (1953); J. A. Pople, W. G. Schneider, and H. J. Bernstein, "High Resolution NMR", McGraw-Hill, New York, N.Y., 1959 p 218 ff.

Computer-Assisted Synthetic Analysis. Synthetic Strategies Based on Appendages and the Use of Reconnective Transforms

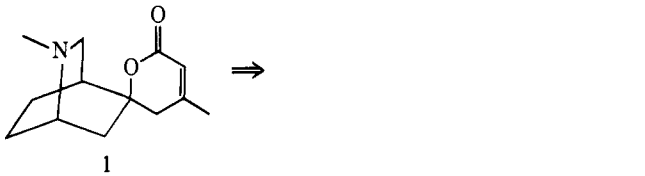
E. J. Corey* and William L. Jorgensen

Contribution from the Department of Chemistry, Harvard University, Cambridge, Massachusetts 02138. *Received May 7, 1975*

Abstract: Computational procedures are developed corresponding to synthetic strategies based on the antithetic (retrosynthetic) disconnection and reconnection of appendages in a target molecule. The reconnective mode is also used to guide the antithetic analyses of target structures containing medium rings. The importance of stereochemical considerations in the reconnective modes is stressed. A definition of "ring" and "branch" appendages is made. In addition, effective methods for the unambiguous recognition of the identicality of two appendages in a molecule are presented. The procedures are completely general since the appendages may include rings and chiral centers. The identicality of different aromatic resonance structures is also recognized. The appendage matching procedure involves a rapid tree search using a "branch atom by branch atom" matching technique. The method described herein is suitable for application to substructure searching in other areas, e.g., in chemical information retrieval systems.

An important aspect of the project at Harvard to devise a program for computer-assisted synthetic analysis has been the development and testing of general synthetic strategies. Many of the strategies currently employed by the program (LHASA) are based on common molecular features. For example, functional groups and rings are the most obvious synthetically significant structural features in a target molecule.[1] They have led to the formulation in LHASA of functional group oriented chemistries[2] and a strategy based on the recognition and antithetic (retrosynthetic) disconnec-

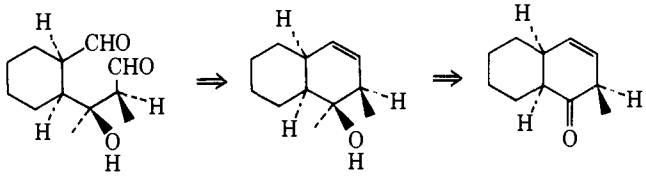
tion of those ring bonds (strategic bonds) whose breaking is anticipated to yield synthetically accessible precursors.[3] In the literature of organic synthesis, one structural feature which has received little formal notice or attention is the unit which has been termed "appendage". Nevertheless, few multistep syntheses are void of processes that involve the disconnection, reconnection, or modification of an appendage in the antithetic sense.[5] A particularly useful antithetic strategy for the analysis of a polycyclic target is to fragment a ring and then disconnect the resultant appendage(s) to yield a structurally simpler precursor. The synthesis of (−)-dioscorine (**1**) by Page and Pinder[6] may be cited as one simple illustration.



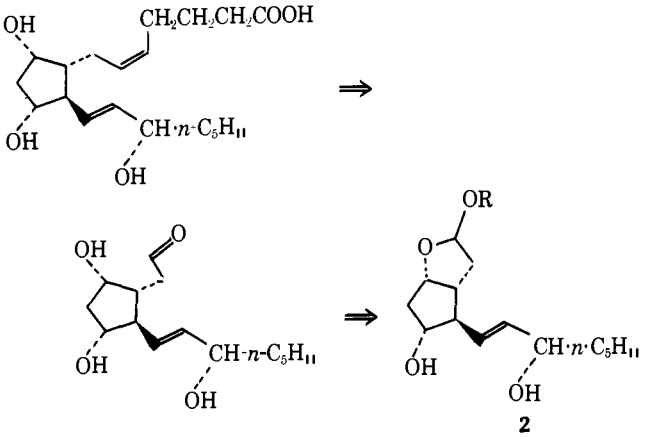
1

Other (and more complex) examples abound in the areas of steroid or terpenoid synthesis.[7]

Reconnective strategies involving appendages are often fruitfully employed in stereospecific syntheses.[4] The rationale in this case reflects the fact that chiral centers are easier to create stereospecifically on a ring than on a chain.



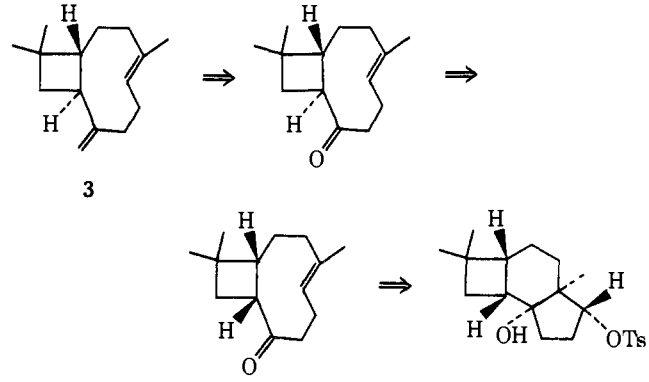
Even when an appendage does not contain stereocenters, its reconnection may have stereochemical merit. The synthesis of prostaglandin $F_{2\alpha}$ is an especially appropriate example since the retrosynthetic sequence corresponds to an appendage disconnection followed by an appendage reconnection forming **2**, the synthesis of which is simplified because of ready access to such cis-fused five-membered rings.[8]



2

It is clearly important for a program designed to generate sophisticated syntheses to have the ability to control the selective disconnection and reconnection of appendages. This capacity has recently been implemented in LHASA and is described in this paper. Before the considerations for the appendage oriented strategies can be described, it is necessary to define what constitutes an appendage, so this must be done at the outset. Further, since it is desirable in an antithetic sense to disconnect simultaneously appendages which are identical, an efficient computational means for recognizing the identity of two molecular substructures will be detailed.

The other area in which reconnective transforms (retroreactions) have become increasingly valuable is in the syntheses of compounds containing medium rings. This approach to medium rings has an inherent entropic advantage over the traditional disconnective methods.[9] The double elimination (fragmentation) in the caryophyllene (**3**) syn-

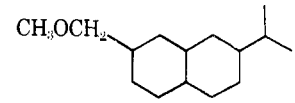


3

thesis[10] exemplifies the stereochemical control that complements the high yields usually found in reconnective medium ring syntheses. The procedures used by LHASA to guide the antithetic analyses for the syntheses of medium ring compounds via reconnective transforms are also presented in this paper.

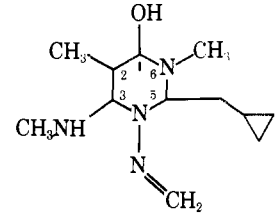
I. Definition of Appendages

Two classes of appendages are readily defined: ring appendages and branch appendages. A ring appendage may be a group of atoms attached to a ring such that the attachment bond is not in a ring itself. In the example



the methoxy methyl and isopropyl units are ring appendages; however, the rings would not be appendages of each other. Some additional restrictions lead to the following prescription for the definition of ring appendages; any atom α to a ring may define an appendage if (1) the bond between the atom and the ring is not in a ring and (2) the atom is a carbon atom or (3) the atom is a hetero atom with at least two carbons attached. In addition, appendages to three-membered rings are ignored due to the dependence of their chemistry on the inherent ring strain.

To illustrate these points, consider the following example.

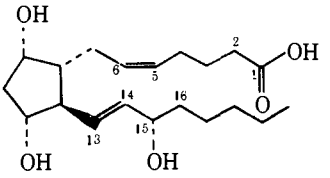


The hydroxyl group is not considered to be an appendage since the oxygen is only attached to one carbon. The groups on atoms 2 and 3 are both appendages since they satisfy rules two and three, respectively. The attachment on atom 4 fails rule three, while the attachments to atoms 5 and 6 both satisfy the second rule. It is noted, however, that the six-

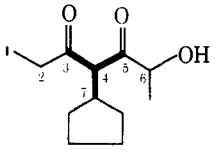
membered ring is not an appendage to the cyclopropyl ring since the latter is ignored. The reason for neglecting the attachments on atoms 1 and 4 is that disconnecting the bonds exo to the ring in these cases would generally correspond to a functional group interchange rather than to a more simplifying single group or group pair disconnection.

The definition of ring appendages is liberal because it is often stereochemically advantageous to reconnect even small ring appendages, e.g., methyls, as illustrated in section IV. The definition of appendages for LHASA is clearly made to reflect the manner in which they are used to guide the antithetic analyses. In the case of branch appendages, i.e., chains on nonring atoms, explicit reconnections are not considered (see section IV). Thus, in attempting to define branch appendages, greater concern is placed on selecting appendages so that their disconnection leads to significant simplification in the target molecule. With this in mind, the definition of branch appendages has been made as follows. For a group of atoms to constitute a branch appendage, it must include at least three carbon atoms. In addition, branch appendages may only originate on a nonring atom that has a total of three or more attachments other than hydrogens including two attachments that qualify as branch appendages. And, finally, to facilitate Wittig-type disconnections, nonterminal double bonds and triple bonds that are on chains are also considered to be the origin bonds of branch appendages.

Appendages are directional and must be represented by a two-atom pair, e.g., (n,m), indicating that the appendage originates on atom n and propagates towards atom m and that the bond between n and m is the origin bond of the appendage. In this notation, the branch appendages for prostaglandin $F_{2\alpha}$ are: (5,6) (6,5), (13,14), (14,13), (15,14), and (15,16).



Therefore, the four branch appendage origin bonds which are candidates for cleavage are C(5)-C(6), C(13)-C(14), C(14)-C(15), and C(15)-C(16). It should be noted that (1,2) is not a branch appendage in this example because C(1) only has this one attachment that could qualify as a branch appendage. Clearly, a disconnection of the C1-C2 bond in $PGF_{2\alpha}$ would be only minimally simplifying. As another example, the following structure may be considered.



The branch appendages in this case are (4,3), (4,5), and (4,7). Thus, the darkened bonds are the origin bonds for the branch appendages. C(3), C(5), and C(6) are not origins of appendages since they each only have one attachment that could be an appendage.

Next, we turn to the determination of identical appendages.

II. Identical Appendage Perception

The recognition of identical substructures in a molecule can be approached in a variety of ways. A method currently popular in chemical information retrieval systems is to compare line-formula notations of the Wiswesser[11] or IUPAC[12]

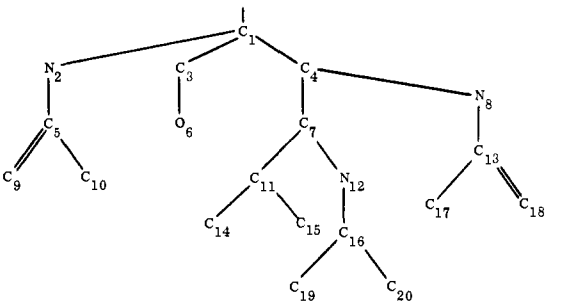


Figure 1.

types. Identicality is readily determined once the task of creating the linear notations for the substructures is accomplished. The computer generation of line formulas from the graphic input of a chemical structure is a research topic of current interest.[13,14] However, the rules for describing cyclic and polycyclic networks are involved and require the perception of numerous, complex subunits, e.g., naphthalenes, pyridines, azulenes, etc. At its current level of development, the line-formula approach is not viable for LHASA due to the imposition of excessive time and space requirements.

A more promising procedure could be based on a set reduction technique.[15] In this scheme, the atoms in the two substructures under comparison are placed in various sets according to their atom types, connectivities, and other characteristics. These sets are then compared pairwise until a one-to-one correspondence is or is not found between the atoms in one unit and those in the other. For example, given that atoms *a*, *b*, *c*, and *d* are in one structure and 1, 2, 3, 4 in another, then if (*a*,*b*,3,4) is the set of carbon atoms and (*a*,*c*,2,3) the set of atoms in double bonds, a one-to-one correspondence between the two subunits is only possible for $a = 3$, and implies $b = 4$, $c = 2$, $d = 1$. If the doubly bonded set were (*a*,*c*,1,2), an inconsistency with the carbon set would be detected and there would be no possibility for identicality. The amount of set construction and manipulation that is necessary becomes cumbersome for large substructures.[15] In addition, extensions would be necessary to handle problems such as the identicality of two resonance structures for the same aromatic system.

The method developed for identifying identical appendages in LHASA may be described as a modified atom-by-atom matching procedure. In this type of approach, the atoms in the two substructures are compared individually until identicality is confirmed or a dead end encountered. For the dead end situation, backtracking to the last decision point is required before comparison can resume along an alternate path. The backtracking aspect of the method is unappealing and time consuming. Therefore, effort was devoted to modify the brute force method[16] to minimize backtracking and produce additional time savings by reducing the number of comparisons during the matching procedure. The results were gratifying since the appendage perception time for even large, highly branched systems is not noticeable to the LHASA user.

Acyclic Substructures. To describe the method, it is instructive to begin with an acyclic example. An acyclic appendage can always be represented by a tree structure as in Figure 1 where hydrogens are implicit. Pursuing the tree analogy, atoms with more than two attachments may be termed branch atoms (BA's), i.e., 1, 4, 5, 7, 11, 13, and 16. A subset of the branch atoms is the set of branch atoms attached to only one other branch atom. These atoms will be called terminal branch atoms (TBA's) since they represent the last level of branching in the tree, i.e., 5, 11, 13, and 16.

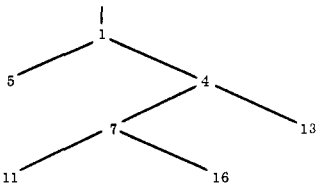


Figure 2.

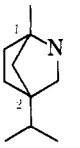
Figure 3.

Finally, sets of identical branch atoms (IDBA's) are defined as branch atoms that are identical with each other with respect to their attachments up to the points where they connect other branch atoms, i.e., (5,13) and (4,7). To form the IDBA sets it is necessary to create a concise numerical description (line string) for the attachments of each BA only up to their connections with other branch atoms.[17] Comparison of the line strings then yields the IDBA sets. Since the attachments must not be branched, the creation of the line string is trivial, as described in a following section Once the BA, TBA, and IDBA sets have been made, the identicality of any acyclic appendages can be determined without further consideration of atoms other than branch atoms. By storing the set of branch atoms adjacent to each BA during the line string generation, the molecular size is, in effect, reduced. For example, the structure in Figure 1 becomes that in Figure 2. The path growing, comparing, and backtracking now involve only branch atoms. The (1,2) appendage may be readily perceived to be identical with the (4,8) because BA's 5 and 13 are identical and terminal.

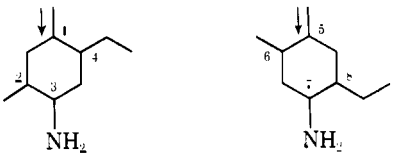
Provision for backtracking must still be provided in the matching program. In Figure 3, the branch atom structure for two appendages has been schematized. Let the IDBA sets be (1,6), (2,3,7,8), (4,10), and (5,9). The BA growing begins with 1 and 6. They are identical so the next level of branch atoms would be retrieved, (2,3) and (7,8). These are processed sequentially which requires 2 and 7 to be compared first. Since they are identical, the next level beneath them is accessed, (4) and (9). A match is not found. Retreat to the last level is, therefore, necessary. The next element in the second set, 8, may then be tried against 2. They are identical and the growing is permitted to resume. If they were not identical, all possibilities would have been exhausted and the appendages could not be identical. Proceeding, the attachments 4 and 10 are also identical and, in addition, terminal. This allows BA's 2,4 and 8,10 to be removed from further consideration. Only 3 and 7 are left on the second level, and they are found to be identical. Since their attachments, 5 and 9, are identical and terminal, the two appendages are recognized as completely identical.

An interesting feature of the branch atom matching procedure is that, once the IDBA sets have been constructed, there is no need to check the actual paths between the branch atoms in acyclic cases. This results because the terminal branch atoms essentially anchor the structures and force the paths to match as long as the branch atoms are identical. If the example in Figure 3 is again considered, it is clear that the terminal branch atoms, 4 and 10, must by definition have only one path to another branch atom. For 4 and 10 to be identical, the line strings,[17] and therefore the paths, to 2 and 8 must be identical. Once this is established, it is evident that, for 2 and 8 to be identical, the paths, 1,2 and 6,8, must also be identical. The progression can continue to the top of any tree and asserts the identicality of the paths concurrently with the branch atom identicality.

Cyclic Substructures. Two modifications are necessary to extend the branch atom matching procedure to perceive identical appendages that include rings. First, since a cyclic appendage does not necessarily contain any terminal branch atoms, the paths between branch atoms in rings must be compared in the two appendages under consideration. This is, however, a simple task because the line strings for the paths are determined in the process of line-string generation. It is emphasized that this checking is only required when both branch atoms are in a ring. In this situation, all linear paths between the two sets of branch atoms must be checked. For example, in the bicyclic system

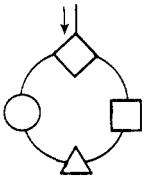


the three paths between 1 and 2 would have to match the paths in the other appendage. If path checking were not done, cases could arise where two dissimilar appendages would be found as identical, such as A and B.



A B

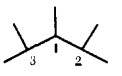
In this example, the IDBA sets are (1,5), (2,6), (3,7), and (4,8) which makes the appendages equivalent to structures that may be schematized as



The problem results from the fact that, when the paths are ignored, the spatial relationship between the branch atoms in the two appendages is the same. However, when the paths from 1 to 2 and from 5 to 6 are compared, it is clear that the appendages are dissimilar.

The second addition is that a record must be kept of the branch atoms that have been grown over. This permits the matching procedure to avoid growing back on itself, i.e., in the example above, the progression would be (1, 2, 3, 4, stop) and not (1, 2, 3, 4, 1, 2, . . .).

Singular Points. One final consideration is required in the event that the two appendages under comparison have atoms in common. If the (2,1) and (3,1) appendages are compared in the following example



growth through branch atom 1 cannot occur in the usual fashion because it is not in any IDBA set. The matching program must, therefore, realize that it should proceed on when any singular BA is reached simultaneously in growing

Table I. Line String Descriptors

Symbol	Descriptor	Symbol	Descriptor
H	1	Triple bond	14
O	2	Ethyl	15
N	3	Propyl	16
C	4	Butyl	17
P	5	Pentyl	18
S	6	Hexyl	19
F	7	*R* center	20
Cl	8	*S* center	21
Br	9	*E* bond	22
I	10	*Z* bond	23
X	11	Cation	24
Branch atom	12	Anion	25
Double bond	13	Radical	26
		Aromatic	27

out both appendages. Alternatively, any branch atom that is not in an IDBA set could be placed in an additional IDBA set in which it is the only element. The latter procedure, however, wastes space and would slow down the matching process for the legitimate IDBA's.

The problem of singular points is unique to programs like LHASA because, for substructure matching in chemical information retrieval systems, one substructure is a reference and is not a part of the other system in which the subunit is sought.

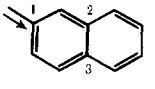
Line-String Generation. In order to recognize subtle molecular features such as chiral centers and stereorelationships, the set of branch atoms must be expanded to include not only atoms with more than two nonhydrogen attachments, but also stereocenters and atoms in multiple bonds. Charged and radical atoms must also be branch atoms, otherwise two appendages that are identical in every respect except for the occurrence of a charged or radical atom in one and not in the other would be recognized as identical. It is then possible to create a compact, but precise, notation for linear chains that permits the rapid comparison of the attachments for two branch atoms. This has been achieved using the numerical descriptors in Table I. The descriptors correspond to the atom types in LHASA with the addition of some special descriptors.

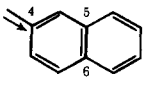
There are six possible descriptors for a branch atom that is not in a multiple bond, 12, 20, 21, 24, 25, and 26. These characters always terminate a line string and are assigned in the following order of importance: stereocenter (most important), charged or radical center, simple branch atom. Thus, after its atom type descriptor, only one special descriptor is assigned to a branch atom.[18]

An atom that is in a multiple bond is described with two characters. The first gives the atom type and the second indicates the bond type as triple (14), *E* (22), *Z* (23), or double of unspecified stereochemistry (13). It is again noted that both atoms in a multiple bond are branch atoms so the multiple bond descriptors are only used in this two character pattern and always terminate line strings.

The descriptors 15–19 are included to shorten the strings by eliminating the repetition of carbon descriptors, e.g., a *n*-pentyl chloride attachment is described as 18,8.

In order to recognize different aromatic resonance structures as identical, it is necessary to use the special descriptor, 27, to indicate the start of a chain of aromatic atoms. Furthermore, doubly bonded aromatic atoms are excluded from the branch atom set unless they have more than two attachments other than hydrogens. In the following structures





the only branch atoms are the numbered centers. The attachments to atoms 2 and 5 would then both have line strings: 27,15,12; 27,4,12; and 27,18,12. This permits branch atoms 2 and 5 to be recognized as identical branch atoms. Similarly, atoms 1 and 4 and atoms 3 and 6 are perceived as identical. The cyclic nature of the structures requires path checking; however, no conflict arises because the line strings are also used to describe the paths. Therefore, without any modifications to the matching procedure, different aromatic resonance structures are recognized as identical.

An added matching difficulty occurs when a hydrogen that is implicit in one appendage is explicit in another, identical appendage. This situation has been efficiently handled by consistently ignoring hydrogens in all aspects of appendage perception. The hydrogen descriptor, 1, is never used and could be replaced to make room for a significant descriptor. Although the need for additional descriptors is not currently apparent, it is convenient to keep the number of descriptors below 32. With this restriction each descriptor only requires five bits of computer storage. This allows a line string of up to six characters to be packed into one, 32-bit computer word.

The Matching Procedure. Before the identicalities of the appendages may be determined, some basic perceptual information must be gathered. The first duty is to find the ring and branch appendage origins according to the rules in section I. The sets of ring and branch appendage origin bonds are created simultaneously. These sets are needed to serve as strategic bonds for appendage disconnections. The sets of branch atoms, terminal branch atoms, and identical branch atoms are determined next by the generation of the line strings for the attachments to the BA's. Finally, the appendage origins may be processed in pairs to establish the identicalities of any appendages.

The information stored during appendage perception by LHASA includes the sets that have been described. In addition, every appendage that is found is represented by a sublist in a linked list.[19] The sublist contains the two-atom appendage origin pair (N,M) and a pointer[19] to the sublist for the next appendage with which it is identical. The last appendage on the list for a set of identical appendages is marked by a zero entry for its pointer. Similarly, an appendage that is not identical with any others also points to zero. The only other information currently in the sublists is an indication of the location and type of the first functional group on the appendage. These data are useful when reconnections of the appendage are considered as described in a following section.

Although the general considerations for the branch-atom-by-branch-atom appendage matching have been provided in preceding sections, the details of the procedure are now presented because the method is complex and central to the topic of substructure searching. A flow chart for the matching algorithm is shown in Figure 4. The explanation below will reference the numbered positions on the chart by indicating the numbers in brackets.

To begin, line strings are created for the two appendages up to the points where they connect with a branch atom [1]. If the initial line strings are not identical, the appendages cannot be identical [A]. If they are identical and no branch atoms are attached, then the line strings completely describe the appendages and they are identical [Z]. In case branch atoms are encountered, the branch atoms are compared in the same fashion with special note being taken in the event that the two branch atoms are the same atom, i.e., a singular point. If the branch atoms are identical and not terminal, further growth is required and begins at [2]. The level of growth or branching is indicated by the counter, *I*,

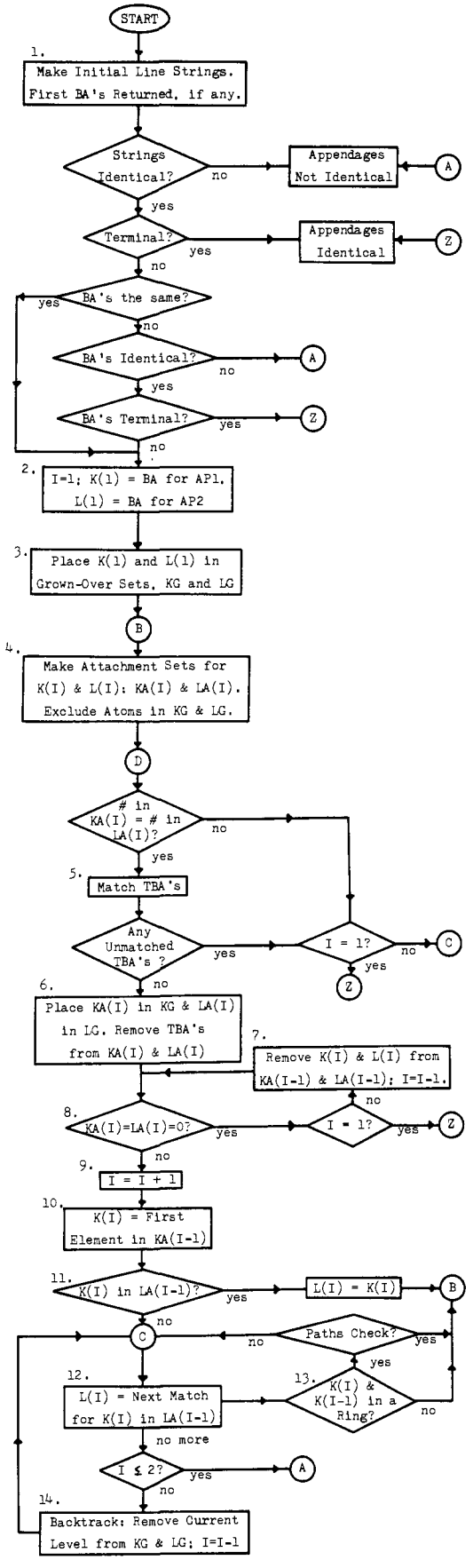


Figure 4. Flow chart for identical appendage perception.

while the branches being matched for the two appendages are recorded in the K and L arrays. To avoid growing over the same branch atom more than once in a cyclic structure,

Table II. Matching for Appendages in Figure 5

Position in flow chart	Level	Action
2	$I = 1$	K(1) = 1, L(1) = 7
3		KG = (1), LG = (7)
4		KA(1) = (2,4), LA(1) = (8, 10)
6		KG = (1,2,4), LG = (7,8,10)
9	$I = 2$	
10		K(2) = 2
12		L(2) = 8
4		KA(2) = (3,5), LA(2) = (9,12)
5		No match for TBA 12
12		L(2) = 10
4		KA(2) = (3,5), LA(2) = (9,11)
6		KG = (1–5), LG = (7–11)
9	$I = 3$	
10		K(3) = 3
12		L(3) = 9
4 via 13		KA(3) = 0, LA(3) = 0
7 via 5,6,8	$I = 2$	KA(2) = (5), LA(2) = (11)
9	$I = 3$	
10		K(3) = 5
12		L(3) = 11
4 via 13		KA(3) = 0, LA(3) = 0
7	$I = 2$	KA(2) = 0, LA(2) = 0
7	$I = 1$	KA(1) = (4), LA(1) = (8)
9	$I = 2$	
10		K(2) = 4
12		L(2) = 8
4		KA(2) = (6), LA(2) = (12)
6		KG = (1–6), LG = (7–12)
		KA(2) = 0, LA(2) = 0
7	$I = 1$	KA(1) = 0, LA(1) = 0
Z		Appendages identical

the sets, KG and LG, are kept for the two appendages, respectively, to record the branch atoms that have been passed over already [3]. The next level of branch atoms for the two appendages is then placed in the KA and LA arrays [4]. Any atoms in KG are excluded from KA(I) and any in LG are excluded from LA(I). For example, in Figure 3 K(1) and L(1) would be 1 and 6, while KA(1) and LA(1) would be (2,3) and (7,8), respectively. Clearly, the number of elements in the KA(I) and LA(I) sets must be the same for a match to be possible. If this is established, any terminal branch atoms in KA and LA may be matched using the IDBA sets [5]. The TBA's must all match or an inconsistency is present. When this test is passed, the KG and LG sets are updated and the matched TBA's are removed from KA(I) and LA(I) [6]. If KA(I) and LA(I) are empty at this point [8], either two subappendages have been fully matched [7] or at the first level, complete identicality is confirmed [Z]. However, if there are still elements in KA(I) and LA(I), another level of growth is required [9]. A new K(I) is taken from the last KA set [10] after the level index has been incremented [9]. If a singular point is detected [11], L(I) is set equal to K(I) and the next sets of attachments retrieved [B]. Otherwise, L(I) is chosen from the preceding LA set and must be an IDBA with K(I) [12]. Path checking is, of course, necessary when the branch atoms are in rings [13] before growth can continue [B]. If the possibilities run out for L(I)'s [12], the program must backtrack to the preceding level [14] to see if all the L(I) candidates were also exhausted there. The backtracking cannot recede past the $I = 2$ level because at the 1 level there is only one choice for K and L. To backtrack, the level counter is decremented and the last KA(I) and LA(I) sets and their TBA's that may have been matched must be removed from KG and LG since the last level was essentially a mistake.

To illustrate the method, Table II follows the processing

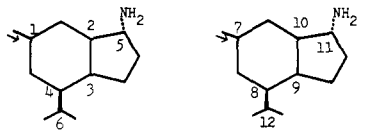
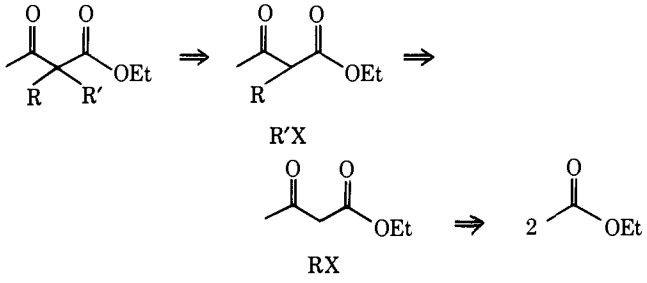


Figure 5.

of the example in Figure 5. The IDBA sets are (1,7), (2,4,8,10), (3,9) (5,11), (6,12), and the set of TBA's is (6,12). The example does not require backtracking, once the numbering reversal for the six-membered rings is overcome. It is interesting to watch the progress of the growing on the rings.

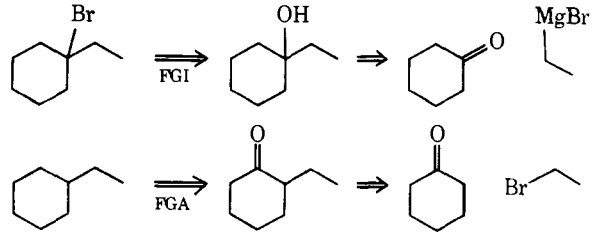
III. Disconnective Appendage Chemistry

It was pointed out above that in antithetic sequences appendage disconnections are often followed by the joining of an appendage to another part of the molecule to form a ring. Appendage disconnections by themselves may also be meritorious. They are, by definition, simplifying, since they reduce the branching in a target molecule. Many reasonable syntheses can, in fact, be devised by applying the simple strategy of disconnecting appendages.



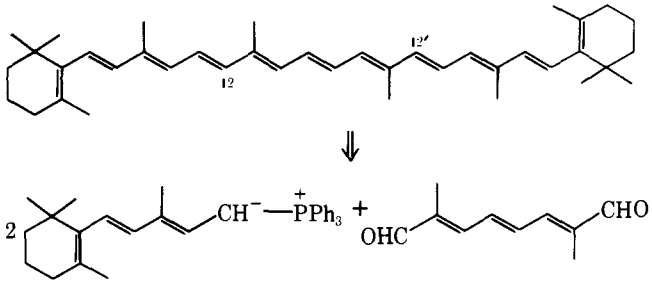
The implementation of an appendage disconnective mode in LHASA was a straightforward process given the strategic bond modes of the one- and two-group chemistry programs[2,20] and the information garnered during appendage perception. Specifically, the sets of ring and branch appendage origin bonds are used as input stategic bonds for the one- and two-group chemistry programs. In their strategic bond modes, the group oriented chemistry programs only evaluate and display transforms that break bonds that have been designated as strategic. This mode was originally developed for the purpose of finding transforms to break strategic ring bonds,[3] i.e., the bonds in a polycyclic target whose breaking is most apt to yield synthetically accessible precursors. Clearly, the same mode can be used to find transforms to break strategic appendage bonds.

The LHASA user may select only ring appendage bonds or only branch appendage bonds or both types to be broken antithetically. The chemist may also request functional group interchange (FGI) and functional group addition (FGA) subgoals[2,21] to assist in breaking any appendage ori-



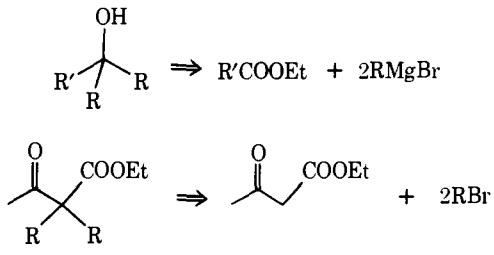
gin bonds that were not broken by a direct one-group or two-group transform. As described in a recent publication,[21] a select number of highly useful one- and two-group transforms may be set up by a sequential series of FGI's consisting of up to four individual synthetic steps. Thus, when the LHASA user feels it is appropriate, the program can automatically generate multistep sequences in order to achieve an important appendage disconnection.

In addition to using the single group and group pair chemistries to disconnect appendages one at a time, the single group program has the capacity to perform simultaneous disconnections of identical appendages. The elegant synthesis of β-carotene[22] by Wittig and Pommer[23] is an example of the utility of multiple disconnections.

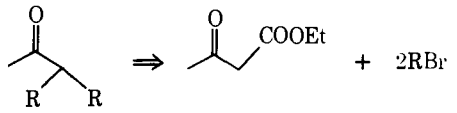


In this case it was necessary to recognize the molecular symmetry or, more specifically, the identicality of the appendages on carbons 12 and 12′ before application of the Wittig disconnection. It is important to perform the identical disconnections simultaneously rather than sequentially. If the disconnections were performed one after the other, the aldehyde generated by the first Wittig transform would appear to be an interfering group for the second disconnection.[24]

Other transforms that may provide appendage disconnections depend critically on the perception of identical appendages. Examples are the addition of organometallic reagents to esters and double alkylations of malonic and ace-



toacetic esters that may or may not be followed by decarboxylation.

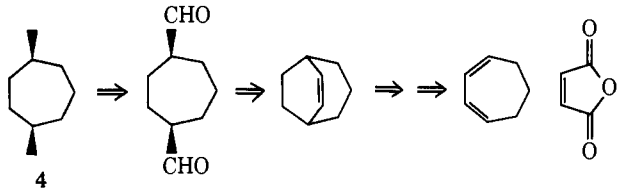


Although the ability to perform appendage disconnections did not require major extensions to the program except for appendage perception, a sophisticated routine had to be developed to oversee the reconnection of appendages and medium ring compounds. This results primarily from the greater concern for stereochemistry and for subgoal performance that is necessary to yield synthetically reasonable sequences containing reconnective transforms. The controlled application of reconnective transforms is the topic of the rest of this paper.

IV. Reconnective Appendage Chemistry

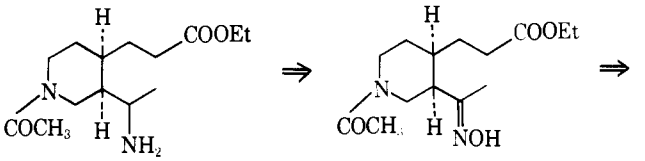
Three classifications may be used to encompass all appendage reconnections: ring appendage to ring appendage (RA–RA), ring appendage to ring (RA–R), and acyclic reconnections. In a RA–RA transform, two ring appendages are connected to each other. They may be on the same ring

initially or on different rings. Besides being useful in placing appendage stereocenters on a ring, RA-RA reconnections secure the stereorelationship of the two appendages in the product. For example, Hendrickson and Boeckman[25] generated the *cis*-methyl appendages in **4** via ozonolysis and re-

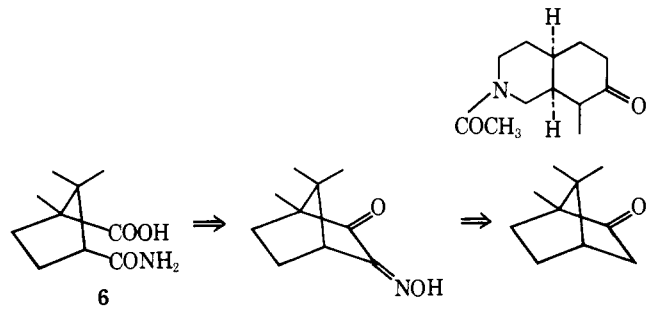


4

duction. Woodward has taken advantage of the stereochemical control of RA-RA transforms on several occasions. His syntheses of a quinine precursor (**5**)[26] and a precursor of the amide **6**, one of the vitamin B_{12} building blocks,[27] illustrate



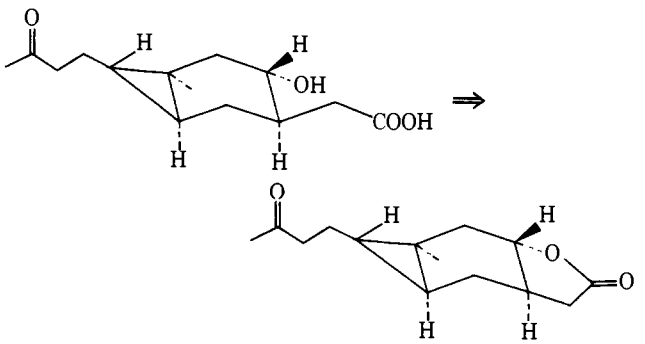
5



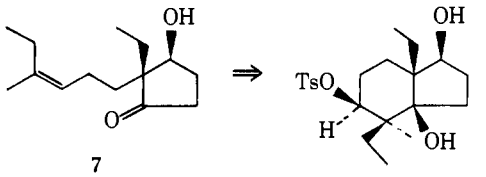
6

the two possible topological outcomes of RA-RA reconnections, namely, fused or bridged ring systems of cyclic order one higher than the target. Since reconnections are not simplifying in this skeletal sense, they must not be applied indiscriminately. Their ability to invoke subgoals, e.g., functional group additions and interchanges, and epimerizations, therefore must be regulated to avoid a proliferation of precursors with only moderate synthetic merit.

In a ring appendage to ring reconnection, an atom on an appendage is reconnected to a ring or heteroatomic functionality directly attached to a ring. In addition to placing stereocenters on a ring, this type of reconnection is often used to establish the stereorelationship of an appendage to other stereocenters on a ring. Examples are the lactone hydrolysis[28]



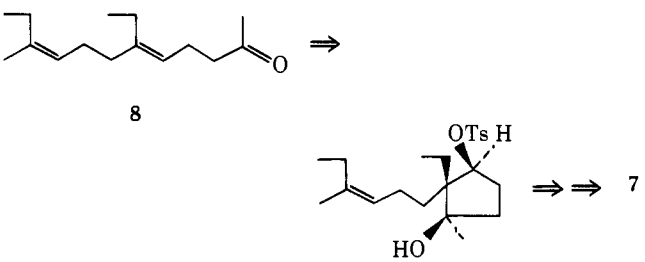
and the prostaglandin synthesis illustrated previously. It should, however, be noted that reconnections may also be used to establish stereochemistry in a resultant appendage. The control of the olefin geometry in the synthesis of the juvenile hormone precursor (**7**)[29] is converted via the double elimination to a problem of controlling the sterochemistry in the bicyclic system.



7

The implicit dependence of RA-RA and RA-R reconnections on appendage pairs and single appendages, respectively, necessitates their independent processing. Furthermore, the subgoal calling restrictions for the two modes must be considered separately. For example, numerous functional group additions can generally be found to permit an appendage to be tied into a ring, while the opportunity for a functional group addition to permit two appendages to be tied together is much less commonplace.

The last type of appendage reconnections may be termed "acyclic" to emphasize that these transforms only directly create a monocyclic system. Reconnections of any acyclic molecule, e.g. **8**, fall into this category.[4,29]



8

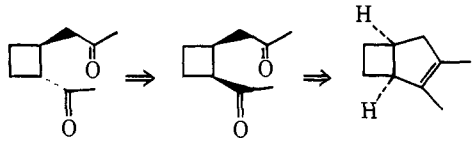
Otherwise the reconnection must connect two atoms on the same ring appendage. Acyclic reconnections will only be permitted when there is strong stereochemical benefit from the transform.

The restrictions and considerations that went into the implementation of the three modes of appendage reconnections in LHASA are described in the following three sections. A unifying theme for reconnective transforms is that they are all currently in the two group class. The examples shown above reflect this fact. Thus, the executive routine for reconnective chemistry only has to interface with the two-group chemistry programs. Single-group chemistry may be ignored since its data base consists solely of disconnective transforms and rearrangements.

The major types of reconnective transforms used by LHASA are illustrated in the Appendix. Although there are only 20 categories represented, there are more than 60 reconnective transform entries in the data table[2] for two-group transforms. The discrepancy primarily reflects the repetition of transform entries for different path lengths separating the keying groups and variations in the keying and leaving groups. Along with each transform type illustration, the transform's subgoal calling capacity is indicated. It is also noted if the transform's functional group addition ability has been restricted as described in the following. Finally, the transform's utility in medium ring syntheses is indicated.

Ring Appendage-Ring Appendage Reconnections. An executive routine has been developed for LHASA to control the administration of reconnective transforms. It consists of two parts corresponding to appendage reconnections and medium ring reconnections. Although the processing for the three types of appendage reconnections, RA-RA, RA-R, and acyclic, is intertwined, it improves clarity to present them individually. To begin, the most complex mode, RA-RA, is described. Its greater intricacy arises from its pair-

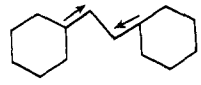
wise nature and the fact that this mode is capable of generating epimerization subgoals when sterically necessary.



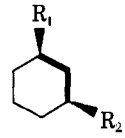
A flow chart for the ring appendage to ring appendage reconnective processor is shown in Figure 6. The following discussion makes reference to the numbered symbols on the chart.

For each unique pair of ring appendages, the processing sequence is followed. N ring appendages yield $N(N-1)/2$ pairwise combinations that are considered. After the first ring appendage is accessed from the appendage list [1], a check is made to determine whether the appendage has any functional groups near enough to its origin to possibly be useful in a reconnection [2-5]. If the appendage does not have any functional group origins five or less atoms from its origin and subgoals have not been requested by the chemist, there is no point in processing the appendage further [4]. The flag, GRPFLG, is set to one when no functional groups are proximate and subgoals are permitted. Its value is saved in GRPFSV [5] so it may be restored [7] after each partner for the appendage is accessed [6].

The two appendages in the pair are not permitted to have any atoms in common [8]. This situation arises when two appendages overlap, e.g., the two appendages formed by a chain between two rings. Any reconnections in this case would properly be described as RA-R or acyclic.

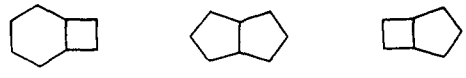


The shortest path between the appendage origins is then grown [9] and the pair's stereorelationship is determined [10], e.g., cis-1,3.



The stereoinformation is crucial for determining the feasibility of reconnections and the necessity for epimerizations. If the appendages are trans and the shortest path between them contains four or more bonds, the pair is rejected [11] due to the need for epimerization and the unlikelihood of reconnecting to a five- or six-membered ring. When the preliminary stereochemical screening is passed, the functional group check is performed on the second appendage [12-14]. If neither appendage has functional groups nearby, multiple subgoals would be required, so a new pair is sought [14]. If only one appendage has candidate functional groups and subgoals have not been specified, the pair must again be rejected [15-16].

When these tests are all passed, epimerizations for trans appendages are considered [17-20]. For trans-1,3 appendages, epimerization is always attempted [17]. For trans-1,2 appendages [19], epimerization is only sterically necessary in cases where the combined sizes of the resultant fused rings total ten or less atoms [20], e.g., 6-4 fusions, 5-5 fusions, etc. The anticipated fusion size is calculated from the functional group proximity information and the size of the ring with the two appendages.



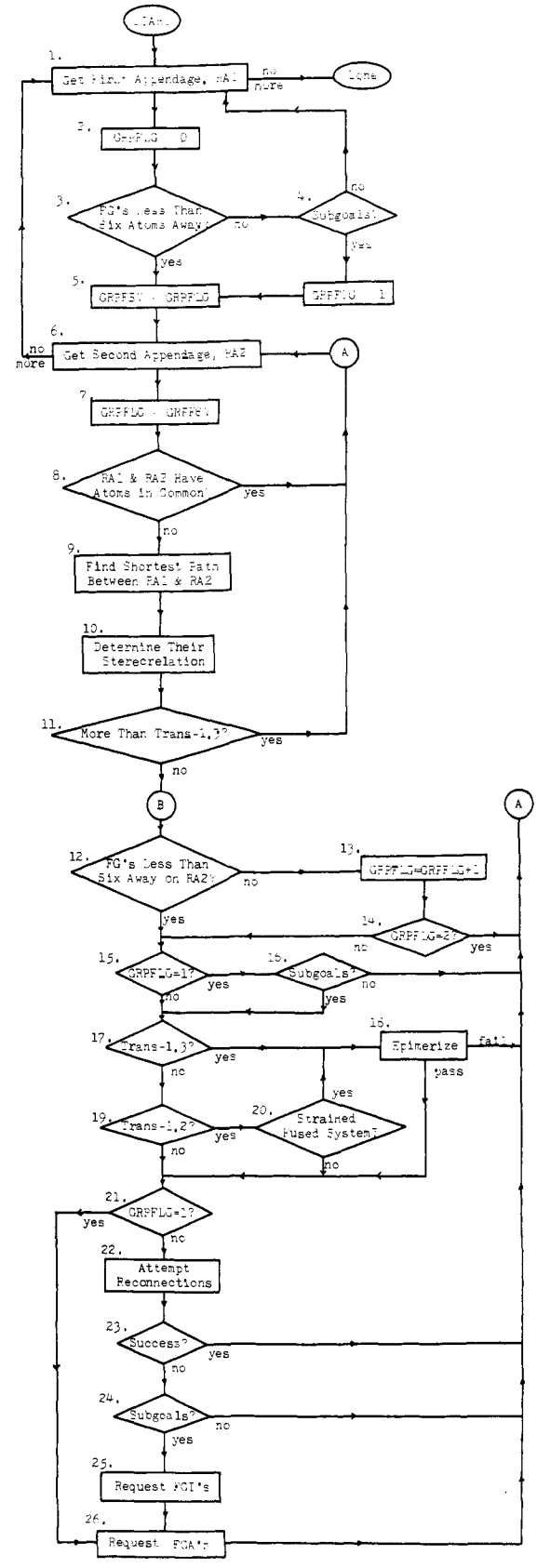
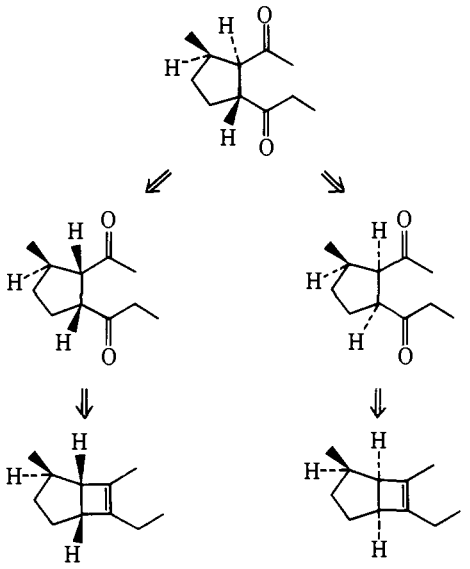


Figure 6. Ring appendage to ring appendage reconnections.

The routine that handles epimerizations determines whether or not an input atom, e.g., a ring appendage origin, is epimerizable. The criterion employed is that the atom must be enolizable. When this condition is met, the epimerized structure is created and displayed. This precursor then

becomes the target to be further processed. Otherwise, a failure indication is returned to the reconnective executive and processing of the appendage pair is aborted [18]. The epimerization processing shown on the flow chart is slightly idealized because provision has been made for epimerizing either or both appendage origins. The following sequence which was generated by LHASA[30] reflects the necessity of being able to epimerize both appendage origins.



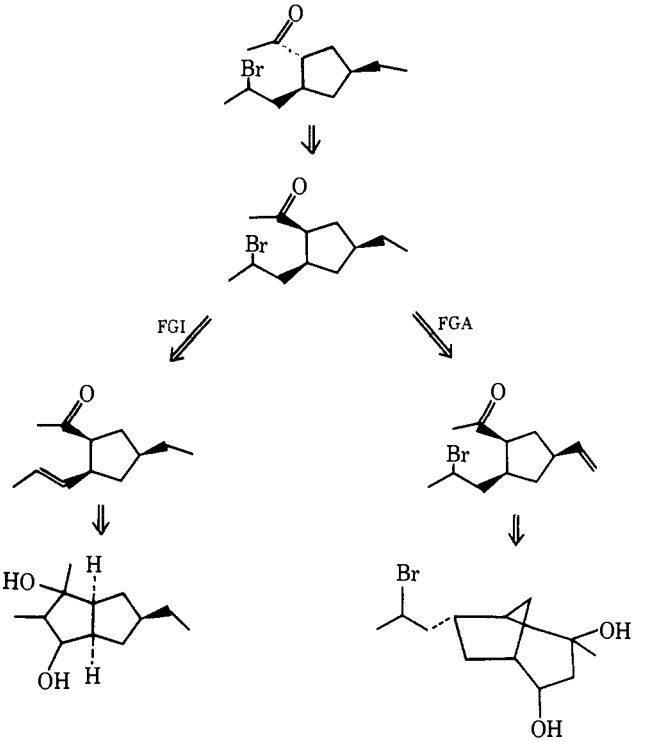
After the epimerization section is passed, GRPFLG is again checked [21]. At this point, if either appendage does not have candidate functional groups, subgoals must have been requested (cf. 15-16), and the only possibilities for reconnections are via functional group additions [26]. Otherwise, direct two-group reconnections of the two appendages are requested [22].

The information about the appendages needed by the two-group chemistry program includes a parameter indicating the reconnective mode, i.e. RA-RA, RA-R, acyclic or medium ring. It also contains two sets that represent the atoms in the two ring appendages for RA-RA reconnections. Reconnective transforms are only performed when one keying functional group is in one set and the other keying group is in the other set. This guarantees the reconnection of the two appendages.

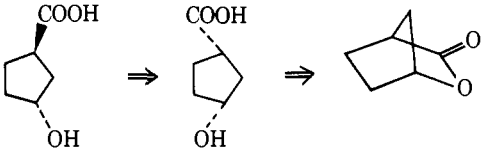
If any direct reconnections of the two appendages are found [23] or none are found and subgoals were not requested [24], the processing of the pair is completed. Otherwise, functional group interchanges and additions are invoked [25-26] to assist in reconnecting the appendages. The entire processing cycle is repeated until all appendage pairs have been analyzed.

The reconnective transforms that are of sufficient utility to empower subgoal requests are indicated in section V. Some of these have restricted functional group addition capacity to reflect possible conflict in the synthetic sequence. An example is ozonolysis of a cycloalkene followed by selective reduction of one of the carbonyl groups. Functional group introduction will only be attempted for these cases when there is considerable potential stereochemical merit. The current condition is that the path between the two functional groups must contain at least two stereocenters, one of which is on an appendage.

To help summarize the RA-RA processor, it is worth noting that most of the checks in the flow chart are designed purely to avoid fruitless two-group chemistry and subgoal requests, particularly senseless epimerizations. In closing, the following excerpts from a reconnective sequence generated by LHASA are shown to illustrate the subgoal capabilities of the RA-RA package. Note that the tosylates implied in the double eliminations are represented as alcohols.



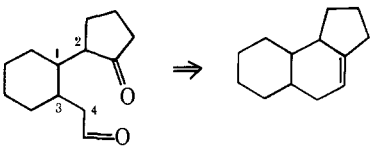
Ring Appendage-Ring Reconnections. RA-R reconnections are more easily administered than RA-RA since only one appendage is involved and its stereorelation to other appendages is not relevant. In particular, epimerizations will not be considered in RA-R mode. The only cases where they might be useful are lactone hydrolyses as in the synthesis of **9**.



9

For this transform, the stereorelationship of the appendage and the ring functional group is important. The majority of transforms that may be used for RA-R reconnections do not, however, have stereochemical restrictions of this nature because they cannot create two chiral centers in the synthetic direction (see the Appendix).

There are cases where the distinction between RA-RA and RA-R reconnections is not obvious. The ozonolysis transform in the example



might be considered a RA-RA reconnection of appendages (1,2) and (3,4) or a RA-R reconnection of the (3,4) appendage. To avoid transform duplication, it is required that, for RA-R reconnections, the path between the keying groups may only contain ring bonds or bonds on the reconnecting appendage. Since the 1-2 bond in the example is neither a ring bond nor on the (3,4) appendage, the transform is only found as the RA-RA reconnection.

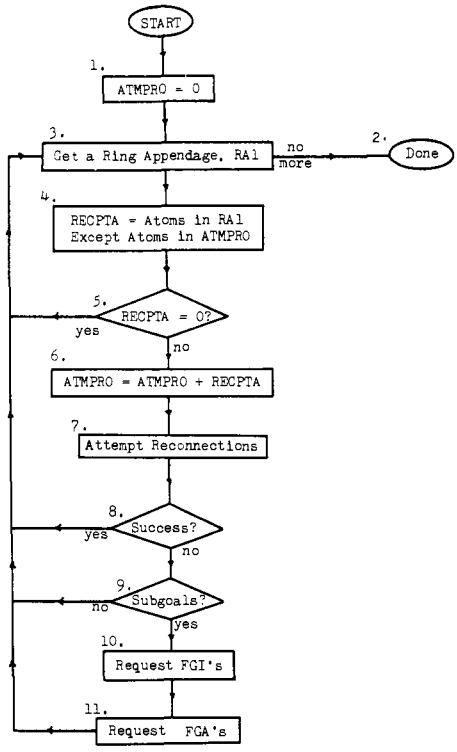
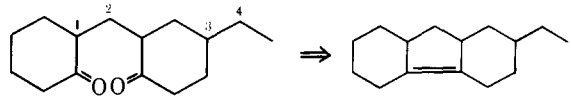


Figure 7. Ring appendage to ring reconnections.

The straightforward processing sequence needed for RA-R reconnections is schematized in Figure 7. The numbers on the chart will be used as reference points in the following discussion.

To begin, the set ATMPRO is initialized [1]. Its purpose is to contain the appendage atoms that have been processed in order to guarantee that no appendage atom is processed more than once. The ring appendages are next accessed one at a time [3]. The atoms in the appendage under consideration are placed in the RECPTA set excluding any atoms that have already been processed [4]. This exclusion is necessary since ring appendages may overlap. For example, the 1-2 appendage in **10** includes the 3-4 appendage. RA-R reconnections for the 3-4 appendage would be contained in the reconnections for the 1-2 appendage. If there are no atoms left in the RECPTA set after the exclusion, a new appendage is sought [5]. This would be the case when the 3-4 appendage in the example is processed.



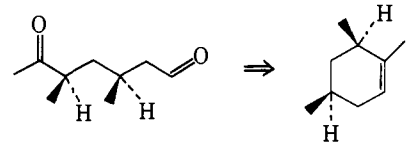
10

Before the transfer to two group chemistry [7], the set of processed atoms is updated to include any atoms from the current appendage [6]. Reconnections are only allowed where one keying group is in the RECPTA (appendage) set and the other keying group is on a ring atom. Note that the ozonolysis transform that yields **10** would be found as an RA-R reconnection for the 1-2 appendage. In this case, the atoms in both rings are in the ring set and the ring with the ethyl appendage is also in the RECPTA set.

The rest of the processing [8-11] is identical with what has already been described for RA-RA reconnections. In fact, the two modes overlap in the reconnective executive routine. The only features unique to RA-R processing are represented by points 1, 4, 5, and 6 on the chart. After an appendage's pairwise reconnections have been performed, its ring reconnections are processed. This occurs before the next appendage's pairwise reconnections are considered. For three ring appendages, A, B, and C, the reconnective processing sequence is, therefore: A-B, A-C, A-ring, B-C, B-ring, C-ring. The acyclic reconnections follow this sequence.

Finally, the restrictions on functional group additions for RA-R reconnections should be mentioned. RA-R reconnections could yield an unmanageable number of FGA requests due to the generally large number of functional group receiving sites available in this mode. In other words, an appendage can be reconnected to numerous ring atoms via an appropriate functional group addition on either the ring atoms or appendage. To regulate this tendency, functional group additions that require removal of a carbon atom in the synthetic direction, e.g., radical decarbonylation, are only allowed to set up reconnections yielding fused ring systems. This restriction reflects several synthetic problems: (1) adding a carbon retrosynthetically increases the complexity of the precursor; (2) functional group additions of this type usually employ severe reaction conditions; and (3) subsequent reconnection to a bridged system is not particularly desirable. Since functional group additions that do not involve carbon removal may circumvent the first two difficulties, they are permitted to lead to bridged systems. In addition, transforms with restricted FGA capability (noted in the Appendix) are not permitted to create bridged precursors unless two or more appendage stereocenters are reconnected. Without these restrictions, the number of precursors with only moderate synthetic accessibility that could be generated in RA-R subgoal mode would be burdensome.

Acyclic Reconnections. After the RA-RA and RA-R reconnections for a target molecule have been completed, one last search for two group transforms is made to yield acyclic reconnections. In this case the reconnective transforms that are considered are required to have the origins of the keying functional groups on nonring atoms. This condition ensures that the acyclic reconnections do not repeat any RA-RA or RA-R reconnections. Finally, an acyclic reconnection is only permitted if it reconnects at least two stereocenters into a ring; i.e., there must be two stereocenters on the path between the keying groups.



This severe stereochemical restriction reflects the fact that reconnections are skeletally nonsimplifying. They should only be applied when there is stereochemical impetus. In the laboratory, reconnections are sometimes useful if they lead to a readily available starting material. Since this type of information is not in LHASA's data base, precursor availability cannot be an incentive to apply any transform.

Subgoals may also be used to facilitate acyclic reconnections subject to the same stereochemical restrictions as for direct acyclic reconnections. Thus, a functional group interchange or addition may be performed if it leads to a reconnection that places two or more stereocenters on a ring.

This completes the description of the appendage chemistry capabilities of LHASA. A final synthetic sequence generated by the program is shown in Figure 8 to illustrate the three types of appendage reconnections.[30] All three modes are invoked simultaneously by the LHASA user. Concurrently, the reconnection of medium rings is also considered as described in the next section.

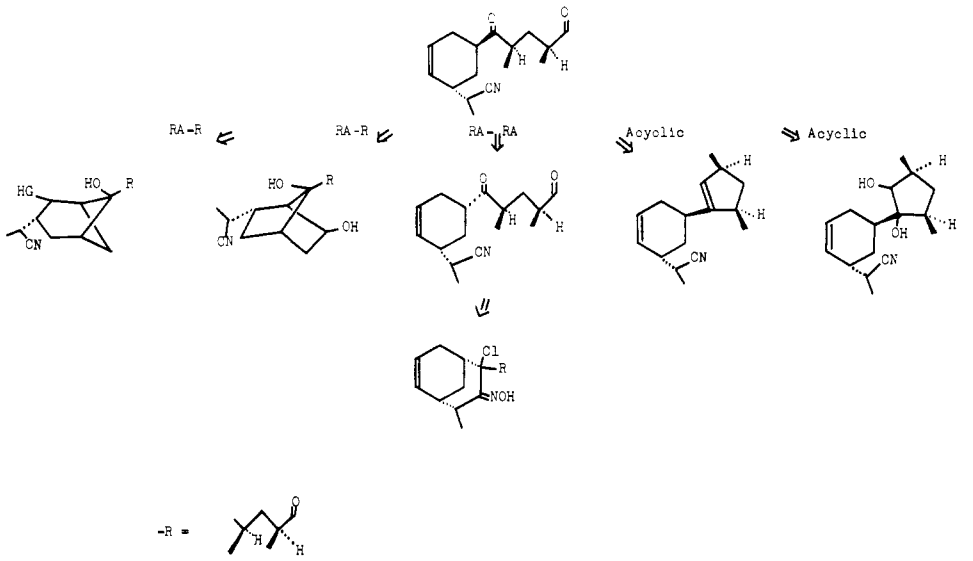
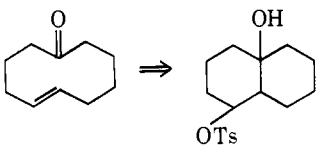


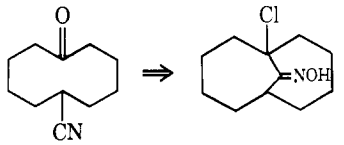
Figure 8. LHASA generated reconnections.

V. Medium Ring Reconnections

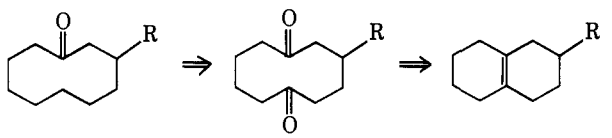
The principal feature of reconnective transforms that may be used in medium ring syntheses is that they create a transannular bond in the antithetic direction.



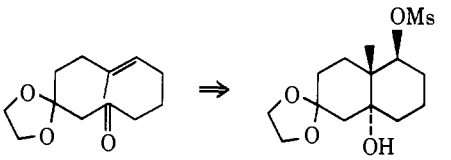
Synthetically, the corresponding reactions are processes that break transannular bonds. The most useful transforms reconnect to *fused* ring systems rather that yielding *bridged* precursors.



The number of appropriate transforms is currently small. In the Appendix, only 7 of the 20 types of reconnective transforms are listed as being at least potentially useful in medium ring syntheses. Of these only three may be readily applied in conjunction with subgoals: double eliminations, cleavage of α,β-epoxy ketones, and dithiane tosylate fragmentations. The other four reactions yield two similar groups in the product. Application of functional group interchanges or additions then implies selective reactions at one functional group in the presence of another similar group, e.g., the ketone reduction in the following example.

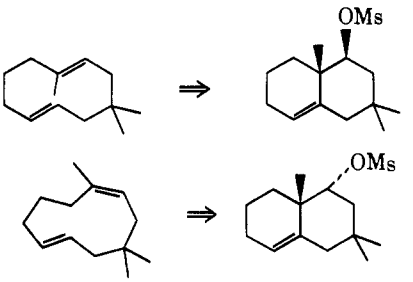


Nevertheless, numerous reconnective medium ring syntheses may be found in the literature. Double eliminations[31] continue to be popular due to their control of olefin geometry in the product, e.g., the caryophyllene synthesis[10] illustrated in the introduction and the synthesis of the trans cyclic olefin **11**.[32]

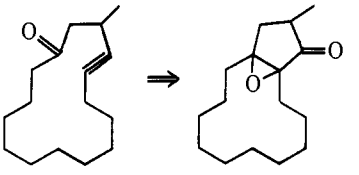


11

The borane mesylate fragmentation developed by Marshall[33] also provides stereoselective routes to medium ring olefins.

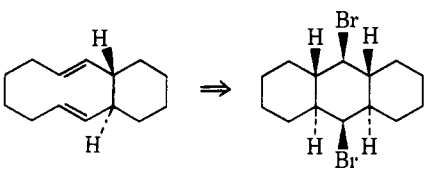


Other reactions that are commonly employed are Eschenmoser's epoxy ketone fragmentation, as in the synthesis of the muscone precursor (**12**)[34]

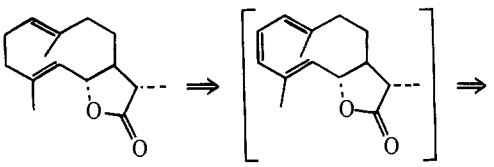


12

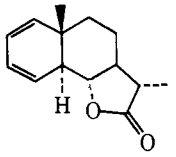
and zinc-catalyzed dihalide eliminations.[35]



Photochemical rearrangements have received limited usage in medium ring syntheses[36] and have not been included in LHASA's data base. An efficient synthesis of dihydrocostunolide (**13**) has, however, involved a photochemical reconnection.[36a]

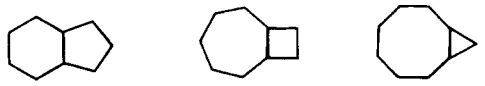


13



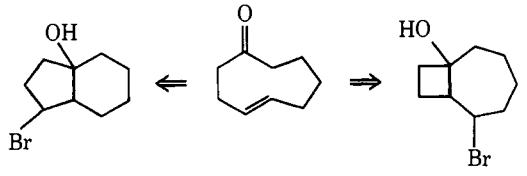
Although reconnections have an obvious entropic advantage, many convenient medium ring syntheses involve disconnective transforms.[9] Particularly notable are Sondheimer's annulene syntheses via oxidative diyne coupling with copper salts,[37] coupling of allylic halides with nickel carbonyl,[38] and diene couplings with zero-valent nickel.[9b,39] The synthesis of medium ring compounds via disconnective transforms is not specifically guided by LHASA; however, many disconnective routes may be found in an opportunistic fashion under normal two group chemistry processing.

Implementation. The major difficulties for implementing a reconnective mode for medium ring compounds in LHASA were to avoid bridged precursors and to yield the proper sizes of fused ring systems. For a nine-membered ring, the latter problem corresponds to reconnecting to 6-5 fused compounds in preference to 7-4 or 8-3 systems.



Both difficulties are handled through the two-group chemistry program.

Bridged precursors are eliminated by rejecting reconnective transforms that do not join two atoms on the medium ring. Fused precursors containing five- and six-membered rings are preferentially generated by regulating the number of ring bonds that must be on the path between the keying groups for the reconnection. Specifically, for rings of size nine or larger, a minimum of three ring bonds must be on the path between the keying groups. Even with this restriction, four-membered rings may be produced by several transforms, e.g., double eliminations.



In this case, reconnection may occur at either end of the olefin to yield a four- or five-membered ring. This is not considered to be a drawback and is occasionally desirable as illustrated:

With these restrictions, the flow chart for the medium ring reconnective mode presented in Figure 9 is essentially self-explanatory. Two points are, however, worth reviewing. First, functional groups either on or α to the medium ring may be used to reconnect it. The transform where an α functional group is essential is the dithiane tosylate fragmentation (see the Appendix). In the flow chart it is shown that checks are made on the number of functional groups on

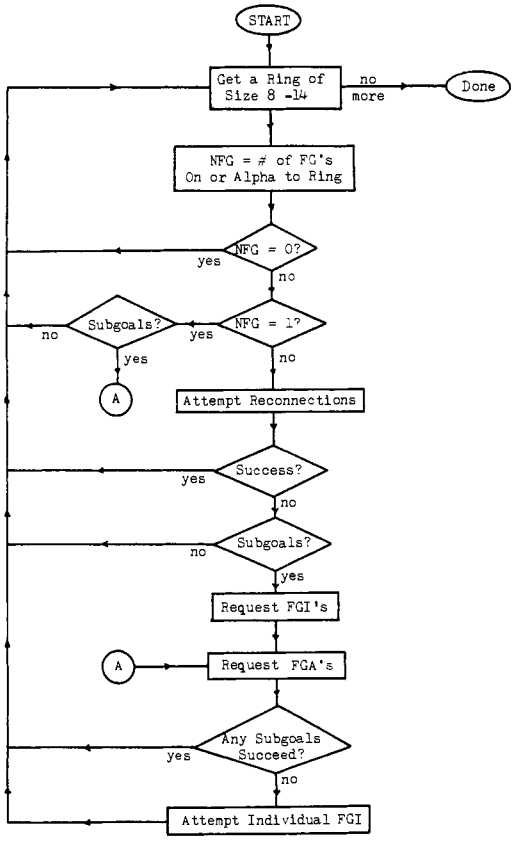
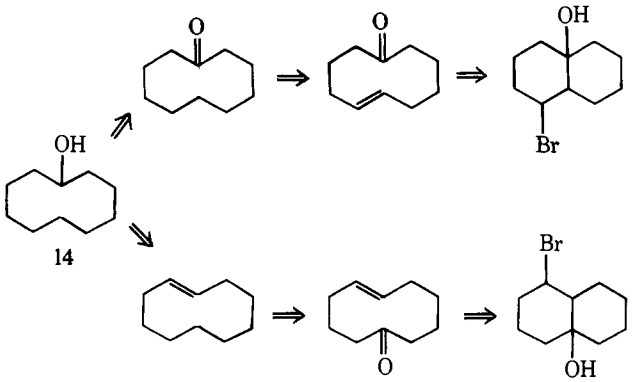


Figure 9. Medium ring reconnections.

or α to the ring to determine the feasibility of reconnections. Similarly, functional group interchange or addition subgoals are considered both on and α to the ring.

The second point is that the medium ring reconnective mode has been provided with an extended subgoal capacity. In the event that no single FGI's or FGA's could be found to set up a reconnection, FGI's are attempted to convert functional groups *on* the ring to C=O or C=C and groups α to the ring to COOR. These are the most fruitful conversions based on the medium ring reconnective transforms that may easily invoke subgoals. The typical case where extended subgoals are generally necessary is when there is only one functional group on or α to the ring and it cannot be used in any medium ring reconnective transform. Thus, for a molecule like **14** reasonable reconnective sequences may still be generated.



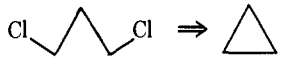
14

Acknowledgments. We are grateful to the National Institutes of Health for financial assistance. The authors are also thankful for assistance from other members of the LHASA group: Dr. W. Jeffrey Howe, Dr. David A. Pensak, Dr. John W. Vinson, and Mr. Harry W. Orf.

Appendix. Types of Reconnective Transforms in LHASA

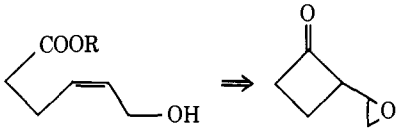
The symbols S, R, and M following a transform name are used to designate the transform as subgoal keying, having restricted functional group addition capacity, and potentially useful in medium ring syntheses, respectively. The symbol X is used to represent a leaving group, e.g., halide, tosylate, etc.

1. Cyclopropyl opening

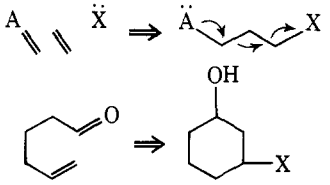


2. Cyclobutene opening

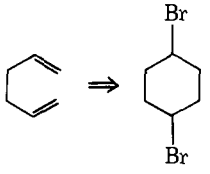
3. Directed cyclobutane opening



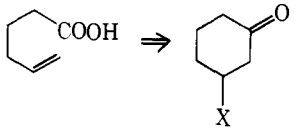
4. Double elimination (S, M), general form



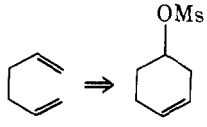
5. Double elimination of a 1,4-dihalide (M)



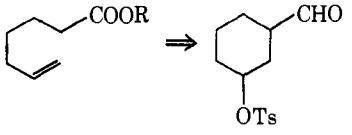
6. Double elimination of a β-X ketone (S)



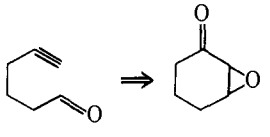
7. Double elimination of a borane mesylate (S, M)



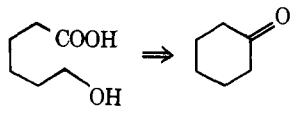
8. Double elimination of a dithiane tosylate (S, M)



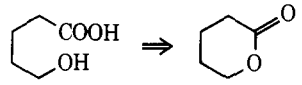
9. Cleavage of an α,β-epoxy ketone (S, M)



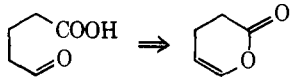
10. Baeyer-Villiger oxidation and hydrolysis (S)



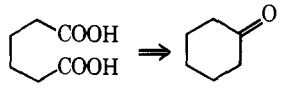
11. Hydrolysis of a lactone (S)



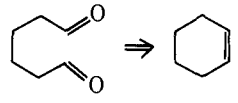
12. Hydrolysis of an unsaturated lactone



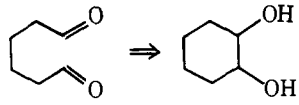
13. Oxidative cleavage of a ketone (S)



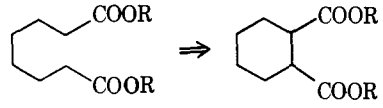
14. Ozonolysis of a cyclic alkene (S, R, M)



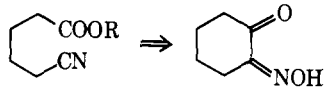
15. Cleavage of a vicinal glycol (M)



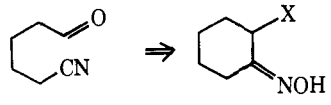
16. Na/NH_3 reductive ring opening



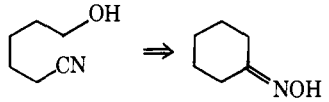
17. Cleavage of a keto oxime (S, R)



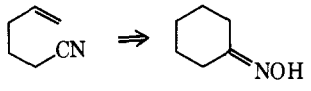
18. Directed cleavage of an oxime (S, R)



19. Cleavage of an oxime and nucleophilic quenching (S, R)



20. Cleavage of an oxime and elimination (S, R)



References and Notes

(1) E. J. Corey, W. T. Wipke, R. D. Cramer, and W. J. Howe, *J. Am. Chem. Soc.*, **94**, 431 (1972); E. J. Corey and G. A. Petersson, *ibid.*, **94**, 460 (1972).
(2) E. J. Corey, R. D. Cramer, and W. J. Howe, *J. Am. Chem. Soc.*, **94**, 440 (1972).
(3) E. J. Corey, W. J. Howe, H. W. Orf, D. A. Pensak, and G. A. Petersson, *J. Am. Chem. Soc.*, **97**,6116 (1975).
(4) E. J. Corey, *Pure Appl. Chem.*, **14**, 19 (1967); *Q. Rev., Chem. Soc.*, **25**, 455 (1971).
(5) A reconnective transform (retro-reaction) creates a ring retrosynthetically. The process corresponds to fragmenting a ring synthetically, e.g., via ozonolysis of a cycloalkene.
(6) C. B. Page and A. R. Pinder, *J. Chem. Soc.*, 4811 (1964).
(7) (a) R. T. Blickenstaff, A. C. Ghosh, and G. C. Wolf, "Total Synthesis of Steroids", Academic Press, New York, N.Y., 1974; (b) J. ApSimon, "Total Synthesis of Natural Products", Vol. 1 and 2, Wiley-Interscience, New York, N.Y., 1973.
(8) E. J. Corey and R. Noyori, *Tetrahedron Lett.*, 307 (1970).
(9) For reviews, see (a) L. I. Belen'kii, *Russ. Chem. Rev.*, **33**, 551 (1964); (b) M. F. Semmelhack, *Org. React.*, 19, 115 (1972).
(10) E. J. Corey, R. B. Mitra, and H. Uda, *J. Am. Chem. Soc.*, **86**, 485 (1964).
(11) E. G. Smith, "The Wiswesser Line-Formula Chemical Notation", McGraw-Hill, New York, N.Y., 1968.
(12) International Union for Pure and Applied Chemistry, "Rules for IUPAC Notation for Organic Compounds", Longmans, Green and Co., London, 1961.
(13) S. R. Heller and D. A. Koniver, *J. Chem. Doc.*, **12**, 55 (1972).
(14) G. A. Miller, *J. Chem. Doc.*, **12**, 60 (1972).
(15) E. H. Sussenguth, Jr., *J. Chem. Doc.*, **5**, 36 (1965).
(16) A brute force approach was described by L. C. Ray and R. A. Kirsch, *Science*, **126**, 814 (1957).
(17) The line strings contain indicators to reflect that a branch atom has been encountered as opposed to a terminal string. See the section entitled "Line String Generation".
(18) For the purposes of LHASA, a charged or radical center is not expected to be an *R* or *S* atom. Clearly, in a more general approach both descriptors could be included in the line string.
(19) An advantage of storing data in lists is that, when the data are no longer needed, the space which they are allocated may be freed for the storage of other information. For further details, see A. T. Berztiss, "Data Structures, Theory and Practice", Academic Press, New York, N.Y., 1971.
(20) The one-group chemistry program controls the application of transforms that are keyed by the presence of a *single* functional group in a target molecule; e.g., a Grignard transform is keyed by a hydroxyl group. Similarly, the two-group chemistry program regulates transforms that are keyed by the occurrence of a pair of functional groups in the target separated by a particular number of bonds; e.g., an aldol transform may be keyed by a β-hydroxy ketone.
(21) E. J. Corey and W. L. Jorgensen, *J. Am. Chem. Soc.*, the following paper in this issue (the second paper in a group of three appearing in this issue).
(22) For a review, see O. Isler and P. Schudel in *Adv. Org. Chem.*, **4**, (1963).
(23) G. Wittig and H. Pommer, German Patent 954,247 (1956).
(24) E. J. Corey, H. W. Orf, and D. A. Pensak, *J. Am. Chem. Soc.*, the third paper in this group of three appearing in this issue.
(25) J. B. Hendrickson and R. K. Boeckman, Jr., *J. Org. Chem.*, **36**, 2315 (1971).
(26) R. B. Woodward and W. von E. Doering, *J. Am. Chem. Soc.*, **67**, 860 (1945).
(27) R. B. Woodward, *Pure Appl. Chem.*, **17**, 519 (1968).
(28) H. Minato and I. Horibe, *Chem. Commun.*, 358 (1967).
(29) R. Zurflüh, E. N. Wall, J. B. Siddall, and J. A. Edwards, *J. Am. Chem. Soc.*, **90**, 6224 (1968). See, also, K. Kondo, A. Negishi, K. Matsui, and D. Tunemoto, *J. Chem. Soc., Chem. Commun.*, 1311 (1972).
(30) The processing mode for the example is reconnective without subgoals. The epimerizations are not considered to be subgoals in this sense.
(31) For reviews, see C. A. Grob and P. W. Schiess, *Angew. Chem., Int. Ed. Engl.*, **6**, 1 (1967); C. A. Grob, *Ibid.*, **8**, 535 (1969).
(32) C. H. Heathcock, R. A. Badger, and J. W. Patterson, Jr., *J. Am. Chem. Soc.*, **89**, 4133 (1967); C. H. Heathcock and R. A. Badger, *Chem. Commun.*, 1510 (1968).
(33) J. A. Marshall, *Rec. Chem. Progr.*, **30**, 1 (1969); *Synthesis*, 229 (1971).
(34) A. Eschenmoser, D. Felix, and G. Ohloff, *Helv. Chim. Acta*, **50**, 708 (1967). See also D. Felix, J. Schreiber, K. Piers, U. Horn, and A. Eschenmoser, *ibid.*, **51**, 1461 (1968).
(35) P. S. Wharton, Y. Sumi, and R. A. Kretchmer, *J. Org. Chem.*, **30**, 234 (1965).
(36) (a) E. J. Corey and A. G. Hortmann, *J. Am. Chem. Soc.*, **85**, 4033 (1963); **87**, 5736 (1965); (b) E. Vogel, W. Grimme, and E. Dinnè, *Tetrahedron Lett.*, 391 (1965); W. G. Dauben and M. S. Kellogg, *J. Am. Chem. Soc.*, **93**, 3805 (1971); W. G. Dauben, R. G. Williams, and R. D. McKelvey, *ibid.*, **95**, 3932 (1973); F. Sondheimer, *Acc. Chem. Res.*, **5**, 81 (1972).
(37) N. Darby, K. Yamamoto, and F. Sondheimer, *J. Am. Chem. Soc.*, **96**, 248 (1974), and references therein.
(38) E. J. Corey and E. Hamanaka, *J. Am. Chem. Soc.*, **86**, 1641 (1964); E. J. Corey and H. A. Kirst, *Ibid.*, **94**, 667 (1972).
(39) G. Wilke, *Angew. Chem., Int. Ed. Engl.*, **2**, 105 (1963); *Pure Appl. Chem.*, **17**, 179 (1968).

Computer-Assisted Synthetic Analysis. Generation of Synthetic Sequences Involving Sequential Functional Group Interchanges

E. J. Corey* and William L. Jorgensen

Contribution from the Department of Chemistry, Harvard University, Cambridge, Massachusetts 02138.

Received May 7, 1975

Abstract: Procedures are described for the generation by computer of a sequential series of functional group interchanges to enable the achievement of important antithetic goals. This capacity, which has been implemented in the Harvard program for computer-assisted synthetic analysis, leads frequently to the automatic design of chemically reasonable multistep sequences. Provision has been made for the overall conversion of one type of functional group to another with up to four individual synthetic steps. However, the methods that are presented may be used to produce sequences of any depth. The potential sequences for converting the functional group in the target molecule (subject group) to the desired functional group in a precursor (object group) are determined by a search procedure through a "sequence tree". The search proceeds directly from the subject group at the tree top to the object group at lower levels in the tree. In order to guide a choice between alternative routes, a rating is derived for each sequence depending on the synthetic utility of each transform in the sequence and the total length of the sequence.

During the last few years the Harvard program for computer-assisted synthetic analysis (LHASA) has been under continuous development resulting in substantial improvement in both scope and sophistication. In recent papers, additions to the program have been described that permit (1) the automatic generation of antithetic (retrosynthetic) sequences of up to 15 steps in a strategy centered about the utilization of important ring-forming transforms (retroreactions);[1] (2) guided antithetic analyses using a synthetic strategy based on the recognition and selective disconnection of those ring bonds (strategic bonds) whose breaking is most apt to yield synthetically accessible precursors;[2] and (3) consideration of the importance of competitive reactions and the necessity for functional group protection during the