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.

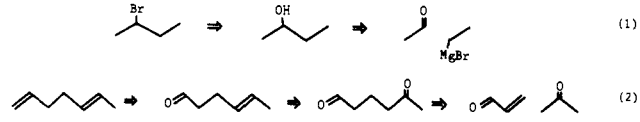
Received May 7, 1975

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

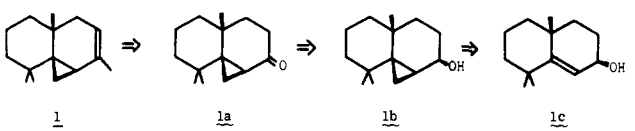
evaluation of the feasibility of each antithetic step.[3] The present paper describes another significant development which has enabled the performance of a sequential *series* of functional group interchanges[4] (FGI's) under the guidance of specific goals such as the application of an important simplifying transform.

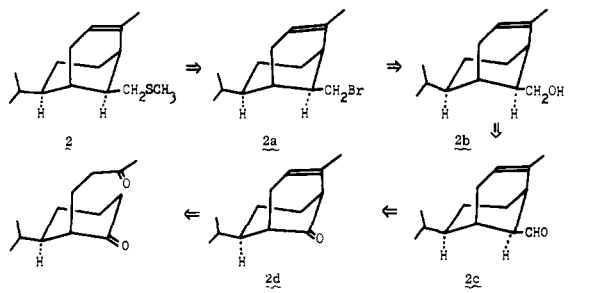
For several years, LHASA has had the ability to perform a single FGI (eq 1) or two parallel FGI's (eq 2) to permit



the use of simplifying one-group and two-group transforms.[4,5] In addition, the program can direct the generation of a series of subgoals (FGI's and functional group additions)[4] to set up an important ring forming process, e.g., Diels–Alder addition or Robinson annulation, when the subgoal sequence is explicitly stated in the data table for the ring transform.[1]

The importance of a procedure leading to the performance of a series of FGI's is apparent from an examination of the literature of multistep synthesis which shows that syntheses commonly involve such sequential transformations. For example, a synthetic route to the sesquiterpene thujopsene (**1**) involved the antithetic sequence: olefin **1** to ketone **1a** to alcohol **1b**, to permit the stereospecific hydroxyl directed Simmons–Smith transform[6] leading to a key precursor **1c**. As another example, consider the target structure **2** from which a sequence of four FGI's produces the ketone **2d** which then can be disconnected by aldol and Michael two-group transforms operating on a strategic bond[2] to afford simple precursors.[7] Clearly, any program intended to design sophisticated syntheses must include an effective procedure for the application of sequential FGI steps as subgoals leading to the operation of structurally simplifying transforms. The design of a computational procedure to carry out this task is complicated by the potentially large number of conceivable sequential routes between one type of functional group and another and by the need for some evaluation of the relative synthetic merit of alternate sequences. The procedure is also required to be flexible in its ability to generate different routes for the same overall FG interchange in order to reflect differences in molecular environments. Furthermore, the data base for the program must be independent of the methods used to generate the sequences and be readily expandable so that the rapid growth of experimental techniques for functional group conversions may be accommodated.[8]





This paper deals in detail with the methods that are used to generate sequential FGI's in LHASA. The procedures are currently being used, as described below, to produce FGI sequences to a maximum depth of four FGI's, although the capacity to create sequences of any depth would be a straightforward extension. The discussion will be restricted to the generation of a single series of FGI's. Subgoal modes such as parallel double sequential FGI[9] and combination sequential FGI–FG addition require minor modifications to the corresponding procedures involving single FGI's that have previously existed in LHASA and that have been discussed.[2,4]

Desirable Features and Restrictions for FGI Sequences. There are two important issues that are crucial to the development of a useful method for generating sequential FGI's. First, when should a sequential FGI sequence be invoked? Secondly, how can the best antithetic routes be selected without creating and analyzing a burdensome number of intermediates? The first problem is easily handled by limiting the use of sequential FGI to cases where the sequence will lead to a significant simplification in the target molecule such as by a strategic bond disconnection[2] or the application of an important ring-forming transform.[1] This is discussed further in the last section of this paper.

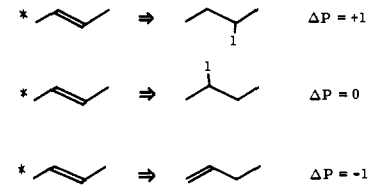
The second question requires a careful analysis of the elements that yield potentially fruitful FGI sequences and the development of rules to help eliminate less productive routes. To begin, some definitions are helpful in this regard. A sequential FGI sequence may be schematized as shown

$$A \overset{P_1}{\Rightarrow} B \overset{P_2}{\Rightarrow} C \overset{P_3}{\Rightarrow} D \overset{P_4}{\Rightarrow} E$$

where A–E represent functional groups with A being the "subject" or target functional group. E is then the "object" group and B–D are "intermediate" groups. The "intermediate transforms" are indicated as

$$\overset{P_1}{\Rightarrow} B, \overset{P_2}{\Rightarrow} C, \overset{P_3}{\Rightarrow} D, \text{ and } \overset{P_4}{\Rightarrow} E$$

The use of the P_n's is necessary to reflect that, in LHASA, FGI transforms have been divided into three categories depending on the location of the object group relative to the subject group.[4] The division is reviewed below



where the object group is placed either one bond farther from the starred atom than the subject group, at the same distance, or one bond nearer. These possibilities are referred to as path change (or ΔP) +1, 0, and −1 type FGI's, respectively.

Since sequential FGI sequences with up to four FGI's are considered by LHASA, it is theoretically possible to have net path changes for a sequence from −4 to +4 bonds. In practice, sequences with $|\Delta P| > 1$ are rarely observed, so the sequential FGI sequences have been restricted to this limit. Also, few reasonable sequences can be imagined where there is a cancellation of path changes, e.g., a net $\Delta P = 0$ FGI formed from a sequence of two FGI's having ΔP's of +1 and −1. So, if a sequential FGI sequence of depth N is desired with a $\Delta P = M$ (−1, 0, +1), it is composed of $N - 1$ steps with $\Delta P = 0$ (by far the most abundant type of FGI) and 1 step with $\Delta P = M$. For $M = \pm 1$, there are, of course, N possible positions for the $\Delta P = M$ step, which are all equally valid.

Several more general rules can be made concerning FGI sequences that help formulate the computational requirements.

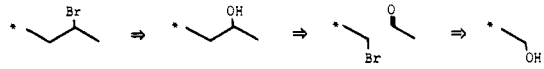
Rule 1. The subject group cannot also be an intermediate group before a path change, i.e., before a $\Delta P = \pm 1$ FGI. The following sequence

$$A \overset{0}{\Rightarrow} B \overset{0}{\Rightarrow} A \overset{0}{\Rightarrow} C$$

where the numbers above the double arrows refer to the ΔP for each step, should obviously be rejected since the first two steps are futile. On the other hand, the sequence

$$A \overset{0}{\Rightarrow} B \overset{-1}{\Rightarrow} A \overset{0}{\Rightarrow} B$$

is perfectly reasonable due to the path change before the recurrence of group A. An example is presented below involving a disconnective Grignard FGI with $\Delta P = -1$.[10] This rule is intuitively obvious to a chemist but, nevertheless, requires definition to be used in a computational procedure.



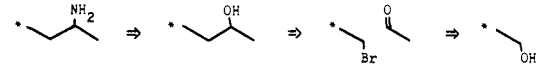
Rule 2. The object group cannot also be an intermediate group except before a path change. Thus, a sequence such as

$$A \overset{0}{\Rightarrow} B \overset{0}{\Rightarrow} C \overset{0}{\Rightarrow} B$$

should not be considered while

$$A \overset{0}{\Rightarrow} B \overset{-1}{\Rightarrow} C \overset{0}{\Rightarrow} B$$

is viable as exemplified by



Rule 3. A longer sequence cannot contain all intermediate transforms of a shorter sequence. This rule is more abstract but serves to eliminate sequences with unnecessary steps. Thus, if the two-step sequence

$$A \overset{n}{\Rightarrow} B \overset{m}{\Rightarrow} C$$

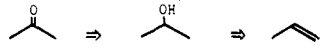
were possible, then the sequence

$$A \overset{n}{\Rightarrow} B \overset{0}{\Rightarrow} D \overset{m}{\Rightarrow} C$$

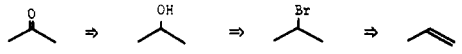
would be ignored, since it contains the same two intermediate steps

$$\overset{n}{\Rightarrow} B \text{ and } \overset{m}{\Rightarrow} C$$

as the shorter sequence. An example is



disallowing



This rule is somewhat restrictive because it is conceivable that

$$B \overset{m}{\Rightarrow} C$$

may not be as synthetically facile as proceeding through the extra intermediate in

$$B \overset{0}{\Rightarrow} D \overset{m}{\Rightarrow} C$$

Although several alternatives to the rule were considered, none were as satisfactory.

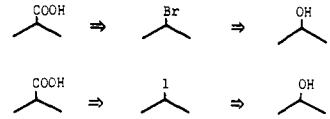
A few other conditions may be defined which serve to increase the efficiency of the search for sequential FGI routes.

Rule 4. Only one representative sequence is chosen whenever a set of sequences are found which have effectively

```
TRANSFORM 269
NAME DIBORANE HYDRATION OF OLEFIN
. . . HO-C1-C2 ⇒ C=C  (ANTI-MARKOVNIKOV)
. . . PATH CHANGE 0
RATING 40
SUBJECT GROUP IS ALCOHOL
OBJECT GROUP IS OLEFIN
MEDIUM*UTILITY
. . .
KILL IF BOND*1 IS A BOND*TO*A*BRIDGEHEAD     . . . BREDT'S RULE
KILL IF NO HYDROGEN ON CARBON*2     . . . VALENCE PROBLEM
KILL IF MULTIPLY*BONDED ATOM ONPATH     . . . E.G., ALLENE CREATED
IF CARBON*2 IS NOT IN A RING*OF*SIZE*3*OR*4 THEN GO TO BLOCK1
    KILL IF CARBON*2 IS A FUSION ATOM     . . . STRAINED SMALL RING
BLOCK1 ADD 15 IF ALPHA TO CARBON*1 OFFPATH IS AROMATIC
. . . .
    REMOVE THIS GROUP     . . . THESE ARE COMMANDS USED TO CREATE
    DOUBLE BOND*1     . . . THE PRECURSOR FROM THE PARENT
. . . .
KILL IF PLUS CHARGE BETTER ON CARBON*1 THAN ON CARBON*2 . . .ANTI-MARKOVNIKOV
CONDITIONS BORANE AND PEROXIDE/ALK
```

Figure 1. Sample transform entry for the FGI data table.

equivalent but not identical *intermediate* groups. This rule is necessary to avoid the redundancy of sequences that differ only trivially, e.g., one having a bromide as an intermediate and the other having an iodide intermediate in the same step.



Currently, three families of closely related functional groups have been placed in corresponding sets of effectively equivalent groups: (1) chloride, bromide, and iodide; (2) acid and ester; and (3) primary, secondary, and tertiary amine. A representative or preferred intermediate group is selected to represent each set. The choices made by the program are iodide, ester, and primary amine for the reason that these representatives are the most likely to enable the program to find further valid synthetic precursors. Of course, the choice of an optimal representative for a specific molecular situation is best left to the chemist and is easily made.

Rule 5. Sequences that cannot meet carbon connectivity (carbon skeleton) requirements are not evaluated. The function of this rule may be clarified by a simple example. The FGI transform ($\Delta P = 0$) alcohol ⇒ acid is allowed only for primary alcohols and not, e.g., for secondary alcohols. The rule requires the differentiation of FGI sequence *generation* and *evaluation*. Sequence generation, which is discussed fully in the next section, yields the theoretical routes between the subject group and the object group. It is based on the information stored in the FGI data table[4] which contains an entry for each FGI known to LHASA including the name of the transform, the subject and object groups, and the chemical qualifiers that are read when the transform is *evaluated*.[1,3] A sample entry is shown in Figure 1. During sequence generation, the only information that is used from the FGI data table is the set of possible object groups for a given subject with the specified path change. If a generated route does not satisfy the rules and conditions that are being delineated in this section, it does not receive evaluation. During evaluation the chemical qualifiers for each transform are read and, if they are passed, the precursor is displayed to the chemist.[11] It is clearly important to limit the display of offspring to possibly fruitful routes.

Rule 6. A nonstrategic ring bond cannot be broken during an FGI sequence used to set up the breaking of a strategic

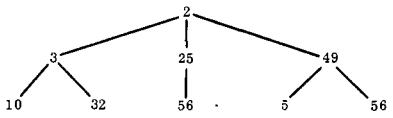


Figure 2. Sample FGI sequence tree.

bond. This rule only applies to FGI sequences generated when strategic bond disconnections are taken as a goal.[2] If a sequential FGI sequence were to disconnect a nonstrategic ring bond, the strategic bonds in the precursor would be different from those in the target molecule[2] and the intent of the original goal would thereby be defeated. However, FGI sequences that only break strategic ring bonds are permitted since they might provide effective synthetic routes to the target molecule. Therefore, a check is necessary during FGI sequence generation to make sure that any disconnective FGI's in a route do not require the breaking of a nonstrategic ring bond.

FGI Sequence Generation. Given the guidelines of the preceding section, attention can now be turned to the detailed methods for generating sequential FGI pathways. The algorithm used in LHASA is straightforward. To begin, it is necessary to have sets[12,13] of object groups for each subject group at each path change (0, ±1). LHASA recognizes 64 different functional groups and assigns each type a number 1 to 64. Thus, two 32-bit computer words are needed to indicate the set of object groups for a particular subject. That is, the *i*th bit in the two-word set is 1 if the *i*th functional group type is an object. These sets are created by reading the subject group and object group entries in the FGI data table (see, for example, Figure 1). The sets are contained in an array[12] and indexed according to the numerical type of the subject functional group and path change.

A request for a sequential FGI must include the types of the subject and object groups and a parameter indicating the desired path change. If the subject group type is 2 and the depth of the routes to be considered is 2, a sequence tree such as in Figure 2 might be constructed. The first level below the tree top contains the possible objects for the subject group and the second level the object group sets for the groups on the first level. The possible object groups that can be reached in 2 steps, therefore, are the groups at the base of the tree. If the path change for the sequence is 0, then only one such tree can be constructed. However, if the path change is ±1, two trees need to be constructed for this example. One tree would be based on the first step being $\Delta P = \pm 1$ transforms and the second step being $\Delta P = 0$ transforms. The second tree would have the $\Delta P = 0$ step first and the $\Delta P = \pm 1$ step at the second level. If the object group for the example in Figure 2 is 56, then two possible sequential FGI routes could be easily found with 25 and 49 as the intermediate group types.

The procedure is formalized in the flow chart presented in Figure 3. It includes the locations in the processing sequence where the restrictions of the last section are heeded and it assumes a fixed depth of search. Since FGI sequences of depth 2, 3, and 4 are considered by LHASA, the flow chart is passed through three times for each request.[14] The shortest sequences are generated first because longer sequences that contain the shorter ones must be rejected (Rule 3 of the preceding section). In addition, for (requests of) $\Delta P = \pm 1$ sequences, the flow chart is passed through N times at each depth N to consider the different placements of the $\Delta P = \pm 1$ step.

The numbered points in the flow chart are referenced by their appearance in brackets in the following discussion of

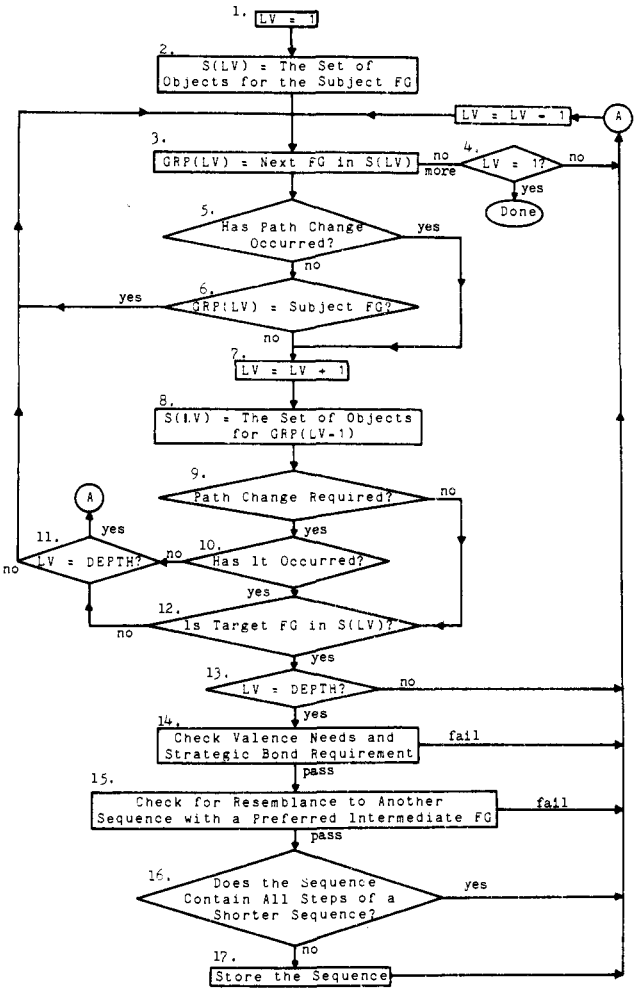


Figure 3. Flow chart for FGI sequence generation. Numbered symbols are used for reference in the text.

the sequence generation. LV keeps track of the processing level in the sequence tree. Initially, LV is 1 [1] and S(1) is set equal to the set of object groups for the subject of the sequence [2]. When the FGI request is for $\Delta P = \pm 1$ sequences, the choice of S(LV) at points [2] and [8] depends on which level has been designated to have the path change. The group types in S(LV) are accessed one at a time and placed in GRP(LV) [3]. If there are no more groups in S(LV) and LV = 1, the processing is complete for this pass through the flow chart [4].

GRP(LV) is always the current subject group so it must satisfy Rule 1 above; i.e., if a path change has not occurred [5], GRP(LV) and the subject group of the request cannot be of the same type [6]. If this condition is passed, LV is incremented [7] and S(LV) becomes the object groups for the current subject [8]. Basically, this procedure [6-8] is repeated until either the maximum depth or the target group is reached.

At this point, a clearer discussion may be achieved by reference to the example in Figure 2. S(1) is the set (3,25,49); GRP(1) is 3; and S(2) is (10,32) assuming all tests have been passed. Letting the example have $\Delta P = 0$, question [9] in the flow chart causes transfer to point [12] where the occurrence of the target group is sought in S(2). Since the target for the example was stated as 56, LV must be compared with the depth of the search [11]. They both equal 2 so it is necessary to backtrack by decrementing LV[A] and getting the next candidate for GRP(1), namely 25 [3]. For a tree with greater depth [11], additional

growth downward would have been required at this point. Returning to the example, S(2) now becomes [8] the set (56) which contains the target [12] and, since LV equals the required depth, the sequence is complete [13]. For the sequence to be accepted [17], it must satisfy the valence test [14] (Rule 5 above), the strategic bond requirement (Rule 6), and the close-relative condition [15] (Rule 4), and it cannot contain all steps of a shorter sequence [16] (Rule 3).

Thus, the pattern of growth through the sequence tree is to go down each branch to the bottom of the tree starting with the farthest right branch and proceeding left (see Figure 2).

Rule 2 from the preceding section is checked at points [9-13]. Specifically, if a path change is required ($\Delta P = \pm 1$) [9] and the processing is past the designated level for the path change [10], the sequence is rejected for cases where the target group is encountered before the bottom of the tree [12-13]. Similarly, this section of the flow chart also causes the rejection of sequences with $\Delta P = 0$ in which the target group is encountered too early.

This brief description has touched upon the main features of the sequence generation. Greater understanding can be gained by manually processing deeper sequence trees using the flow chart. It should be noted that other procedures for the sequence generation can be devised. A particularly fast method would be to use object sets for the subject groups and subject sets for the object groups concurrently. All two-step routes could then be obtained by noting which intermediate groups are in the object set for the subject group as well as the subject set for the target group. Computationally this only requires the performance of one Boolean "and" operation on the two sets. Longer sequences could also be readily generated by extending the technique. This alternate procedure for sequence generation was not implemented due to its added storage requirements.

Sequence Selection, Evaluation, and Display. Although the sequence generation is central to producing sequential FGI routes, it is necessary to have an executive routine in LHASA to oversee their performance. One obvious function of the executive is to decide which sequences are to be attempted. For example, if a good two-step sequence is possible, there is little point in displaying a four-step sequence that includes synthetic steps of questionable utility.

Another function of the executive is to select the object groups (antithetic targets) for the sequential FGI sequences. Specifically, there may be a variety of functional groups that key a particular transform, e.g., aldehyde, ketone, ester, nitro, sulfoxide, etc., for a Michael transform, so a choice of objects is necessary because it would be burdensome and redundant to display FGI routes leading to all the possibilities.

Both of these duties are handled by ranking the sequences at the generation stage. Each sequence that is generated is given a numerical ranking that reflects the length of the sequence and the general utility of each transform in the sequence. Three classifications of transform utility are currently being employed: low, medium, and high. The classification of each transform is included in its FGI data table entry (see Figure 1). Low utility transforms are often characterized by severe reaction conditions, fair-to-poor yields or a lack of stereospecificity. Examples in this category are the Hunsdieker decarboxylation and acid-catalyzed dehydration of an alcohol. At the opposite extreme, transforms with the high utility designation are more synthetically reliable, e.g., hydrolysis of esters, reduction of nitriles, and the Wittig reaction.

The numerical ranking is then determined by assigning 3 points for each step in the FGI sequence and an additional 4 points for each medium utility transform and 6 points for each low utility transform. High utility steps do not add to the point total. So, a three-step sequence with one high, one medium, and one low utility transform receives a ranking of $(3 \times 3) + (1 \times 4) + (1 \times 6) = 19$ points. Clearly, sequences with the lowest point totals are anticipated to be the most fruitful for subsequent evaluation. Also, the empirical point assignments imply a purely empirical equivalence between a sequence with a low utility step and one with two extra high utility steps.

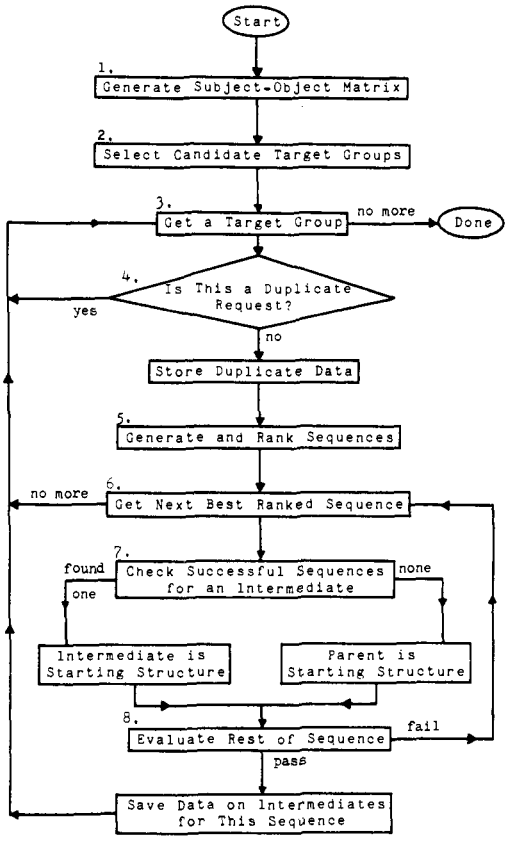


Figure 4. Flow chart for the sequential FGI executive routine.

For each request of sequential FGI's, the possible object groups are screened by selecting the two objects with the sequences that received the smallest ratings. The sequences for these two objects then receive full evaluation in order of ranking. The procedure is shown in Figure 4. Several other duties of the executive routine are also indicated in the flow chart. The following description is intended to clarify these points. The numerical headings below correspond to the numbered locations in Figure 4.

1. The first responsibility of the sequential FGI executive is to create the array of object sets for the subject groups in the FGI data table. As discussed in the last section, this array is central to the sequence generation. The formation of the array only occurs once during any session using LHASA. Any additions to the FGI data table that occur between sessions are, therefore, automatically reflected in the array and resultant sequential FGI sequences.

2. Next, the two antithetic target groups for the request are selected from the possibilities that are input. This requires the generation and ranking of all FGI sequences for the possible targets. The two target groups with the best ranked sequences are retained. To save space, the sequences are not kept during this procedure so they must be regenerated for the two targets at a later point. This is not a significant inconvenience because the sequence generation is exceedingly fast.

3. The two chosen targets are processed one at a time.

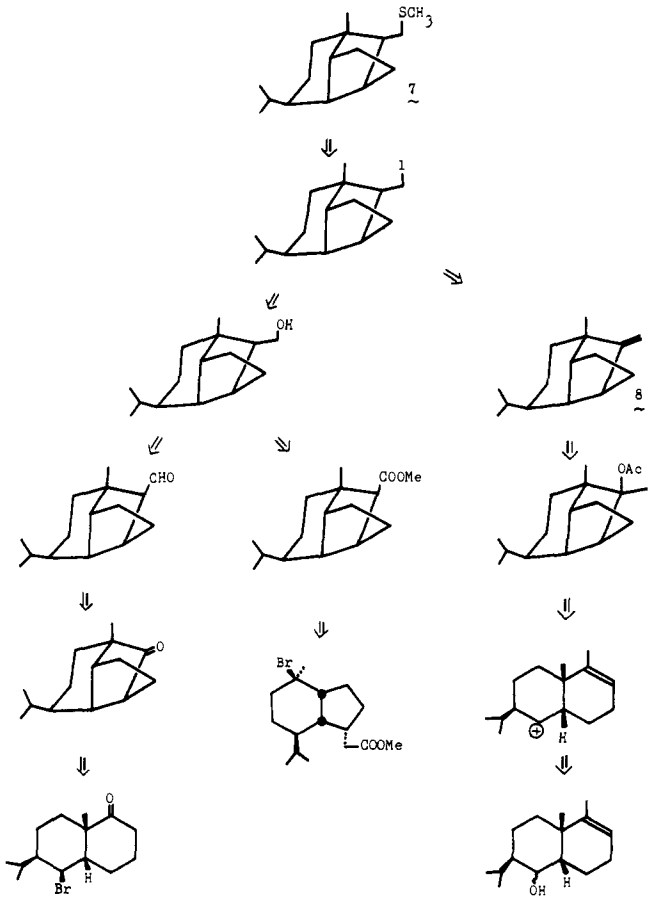


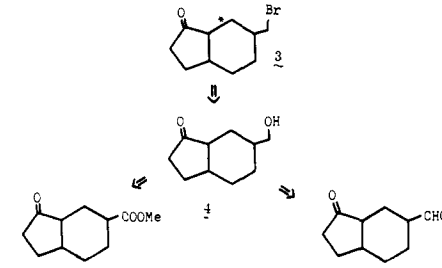
Figure 5. LHASA-generated synthetic routes to sativene (8) and its analogue (7).

After the second target is finished, control returns to the chemistry package that requested the sequential FGI's. The next section describes the strategies that have this power.

4. It is possible that the overall FGI request has been made previously. If this is the case, the request is now ignored. This situation arises because different transforms that have the ability to invoke sequential FGI sequences are sometimes keyed by similar functional groups, e.g., various condensation reactions keyed by withdrawing groups. If the request is not a duplicate, sufficient information to describe the request is stored in a list structure[12,15] so it may be compared against future requests. The stored information includes the subject and object groups for the request, the location of the subject group in the parent compound, and the path change parameter.

5. Next, the sequences for the target group under analysis are regenerated and ranked.

6. The sequences are considered one at a time in order of ranking. Currently, the first sequence that is successfully evaluated is the only one that is displayed for *each* target group. Thus, a request generally yields two sequential FGI routes unless it is a duplicate request. For example, a request to use a Michael transform to break the starred bond in 3 causes the program to generate two sequential FGI routes with ester and aldehyde as their target groups. Although more routes could be easily evaluated and displayed, the number of intermediates differing only slightly soon becomes unmanageable. The routes that are displayed are sufficiently representative. Clearly, it is better to have many simplifying transforms invoke a few sequential FGI sequences than to have only a few transforms invoke many sequences.



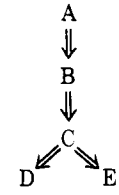
7. In order to streamline the processing, the previously successful sequences are checked for intermediate structures that may be used as a starting point for processing the current sequence. Thus, if a successful sequence

$$A \overset{m}{\Rightarrow} B \overset{n}{\Rightarrow} C \overset{1}{\Rightarrow} D$$

has already been displayed and a new sequence

$$A \overset{m}{\Rightarrow} B \overset{n}{\Rightarrow} C \overset{k}{\Rightarrow} E$$

is being attempted, intermediate C becomes the starting point for the new sequence. If the new sequence is successful, a branch is created in the synthesis tree[11] at the point corresponding to structure C.

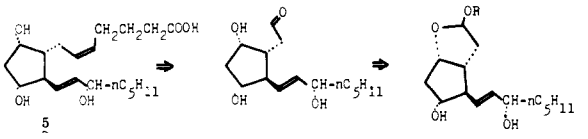


A specific example is shown for the two routes leading to 3 above where 4 is a common intermediate. If an intermediate starting structure is not found in previous sequences, the processing of the current sequence must begin with the parent, e.g., A. Naturally, if there is a choice of intermediate starting points, the one farthest along in the current sequence is chosen, e.g., C rather than B.

8. The remaining steps in the current sequence then receive evaluation. The chemical qualifiers for each transform (see Figure 1) are read and the feasibility of the process is determined. If the qualifiers cause any step to be rejected, the entire sequence must be abandoned and a new one attempted. If the sequence is completely successful, information concerning it must be stored to permit the selection of intermediate starting points for future sequences. The necessary data are similar to those used to avoid duplicate requests and are again placed in list format.[12] The successful sequence is then displayed and the next target group begins analysis.

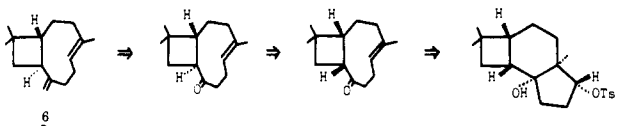
Use of Sequential FGI in LHASA. Two conditions must be satisfied for the program to request a sequential FGI series. First, the series must lead to the achievement of an important antithetic goal. Secondly, the means of executing the goal must have high synthetic merit.

Automatic processing modes may currently be invoked by the LHASA user corresponding to the following antithetic goals: strategic bond disconnections,[2] application of important ring forming transforms (Diels–Alder, Simmons–Smith, etc.),[1] appendage disconnections and reconnections,[16] and syntheses of medium ring compounds (size 8–14) via reconnective transforms.[17] The syntheses of **2** and thujopsene which are mentioned above serve to illustrate the strategic bond and ring transform goals. The synthesis of prostaglandin $F_{2\alpha}$ (**5**) is a convenient example of the appendage strategies[17] because the retrosynthetic sequence corresponds to an appendage disconnection followed by an appendage reconnection to the acetal.[18] This example indicates the stereochemical benefit that is often derived from



5

the use of reconnective transforms. A reconnective approach to medium ring syntheses is particularly attractive due to the stereochemical control and inherent entropic advantage over the traditional disconnective methods.[19] The internal double elimination (fragmentation) utilized in the synthesis of caryophyllene (**6**) exemplifies a practical, reconnective medium ring synthesis.[20]



6

These goals may all be approached through sequential FGI sequences, if the transform used to achieve the goal has widespread synthetic utility. In the case of the important ring-forming transforms this requirement is implicitly satisfied.[1] However, only 7 of the more than 200 transforms in the two-group class[5] have been given the ability to invoke sequential FGI sequences for the purpose of disconnecting strategic ring bonds[2] or appendage bonds.[17] These transforms correspond to the aldol, Michael, and Mannich reactions and the malonic ester, acetoacetic ester, Claisen, and Dieckmann condensations. In the one-group class, the transforms that have been given this status are the Wittig and Grignard reactions, cation-olefin addition, alkylation α to electron withdrawing groups, and malonic ester and acetoacetic ester condensations followed by decarboxylation. The only reconnective transform that may be set up by a sequential FGI sequence is the internal double elimination (fragmentation) process. Restricting the number of transforms that have the power to cause sequential FGI's is important for two reasons. First, it precludes the addition of large numbers of questionable synthetic routes to the synthesis tree, and second, the routes that are found are much more likely to result in the generation of interesting and useful synthetic pathways.

In closing, a synthetic analysis is presented based on the target molecule **7**, an analogue of the sesquiterpene sativene (**8**).[21] Several routes to **7** and **8** that were generated by LHASA are outlined in Figure 5. The goal in this case is the disconnection of strategic bonds by one-group transforms. As illustrated, the sequential FGI sequences lead to strategic bond disconnections via α alkylations and a cation-olefin reaction which yield fused, bicyclic precursors.

Acknowledgment. We are grateful to the National Institutes of Health for financial assistance. The assistance of Mr. Harry Orf and Dr. John W. Vinson is greatly appreciated.

References and Notes

(1) E. J. Corey, W. J. Howe, and D. A. Pensak, *J. Am. Chem. Soc.*, **96**, 7724 (1974).
(2) E. J. Corey, W. J. Howe, H. W. Orf, D. A. Pensak, and G. Petersson, *J. Am. Chem. Soc.*, **97**, 6116 (1975).
(3) E. J. Corey, H. W. Orf, and D. A. Pensak, *J. Am. Chem. Soc.*, the following paper in this issue.
(4) E. J. Corey, R. D. Cramer, III, and W. J. Howe, *J. Am. Chem. Soc.*, **94**, 440 (1972).
(5) A "one-group" transform is keyed by the presence of a *single* functional group in a target molecule; e.g., a Grignard transform is keyed by a hydroxyl group. "Two-group" transforms are keyed by the occurrence of a pair of functional groups in the target separated by a particular number of bonds; e.g., a Michael transform may be keyed by a 1,5-dicarbonyl unit.
(6) W. G. Dauben and A. C. Ashcraft, *J. Am. Chem. Soc.*, **85**, 3673 (1963).
(7) Cf. E. J. Corey and S. Nozoe, *J. Am. Chem. Soc.*, **85**, 3527 (1963).
(8) I. T. Harrison and S. Harrison, "Compendium of Organic Synthetic Methods", Wiley-Interscience, New York, N.Y., 1971; C. A. Buehler and D. E. Pearson, "Survey of Organic Syntheses", Wiley-Interscience, New York, N.Y., 1970; S. R. Sandler and W. Karo, "Organic Functional Group Preparations", Academic Press, New York, N.Y., 1968–1971.
(9) The synthesis of barrelene is an elegant example leading to a double Diels–Alder transform: H. E. Zimmerman, G. L. Grunewald, R. M. Paufler, and M. A. Sherwin, *J. Am. Chem. Soc.*, **91**, 2330 (1969).
(10) The Grignard reagent is represented as the bromide from which it is derived.
(11) E. J. Corey and W. T. Wipke, *Science*, **166**, 178 (1969); E. J. Corey, W. T. Wipke, R. D. Cramer, III, and W. J. Howe, *J. Am. Chem. Soc.*, **94**, 421 (1972).
(12) A. T. Berztiss, "Data Structures, Theory and Practice", Academic Press, New York, N.Y., 1971.
(13) E. J. Corey, W. T. Wipke, R. D. Cramer, III, and W. J. Howe, *J. Am. Chem. Soc.*, **94**, 431 (1972).
(14) When there is a one-step FGI possible to satisfy the FGI request, sequential FGI is not attempted.
(15) 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 and used for the storage of other information.
(16) A reconnective transform corresponds to a synthetic reaction that fragments a ring, e.g., ozonolysis of a cycloalkene or hydrolysis of a lactone.
(17) Details on the use of appendage disconnections and reconnections and medium ring reconnections as synthetic strategies is described in the preceding paper (E. J. Corey and W. L. Jorgensen). See also E. J. Corey, *Q. Rev., Chem. Soc.*, **25**, 455 (1971).
(18) E. J. Corey and R. Noyori, *Tetrahedron Lett.*, 307 (1970).
(19) For reviews, see (a) L. I. Belen'kii, *Russ. Chem. Rev.*, **33**, 551 (1964); (b) M. F. Semmelhack, *Org. React.*, **19**, 115 (1972).
(20) E. J. Corey, R. B. Mitra, and H. Uda, *J. Am. Chem. Soc.*, **86**, 485 (1964).
(21) Cf. J. E. McMurry, *J. Am. Chem. Soc.*, **90**, 6821 (1968).