Computer-Assisted Synthetic Analysis. Methods for Machine Generation of Synthetic Intermediates Involving Multistep Look-Ahead

E. J. Corey,* W. Jeffrey Howe, and David A. Pensak

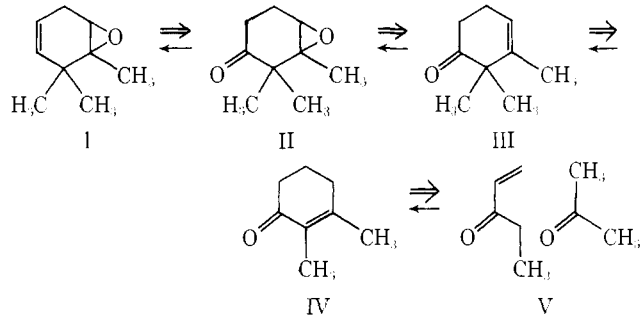
Contribution from the Department of Chemistry, Harvard University, Cambridge, Massachusetts 02138. *Received May 4, 1974*

Abstract: A major extension has been made to the computer program (LHASA) being developed to assist chemists in synthetic analysis. This extension allows the automatic processing of a target molecular structure in the antithetic (retrosynthetic) direction with a depth of search corresponding to as many as 15 steps to determine whether a goal of utilizing a specific important ring-forming process can be attained. The method is described for the Diels–Alder transform (retroreaction), the first specific case to be implemented. The search for applicable subgoals leading to the application of the Diels–Alder transform is directed by a binary search pattern represented in the form of a data table. This table is constructed from a "chemical flow chart" which contains information regarding the modification of structural features by stepwise application of transforms to produce the kind of structural unit required for a valid direct application of the Diels–Alder transform. The data table is derived by transcription of the chart into a special computer-readable language resembling "chemical english" which is easily intelligible to a chemist. The binary search technique, which is applied exhaustively to generate as many synthetic pathways as possible, is very effective. An illustration is given and a brief discussion is presented of other uses of this technique in synthetic analysis by machine. The feasibility of binary-search look-ahead of up to 20 steps for the purpose of applying important transforms or strategies has been clearly demonstrated. Finally, procedures are described for machine perception of stereorelationships in organic molecules so that synthetically significant stereochemical information can be made available to the program.

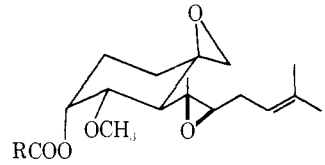
For several years a program of research has been pursued at Harvard with the twin objectives of formalizing the methods and theory of organic synthetic analysis and of developing an interactive computer program to aid the chemist in generating a broad range of hypothetical solutions to a specific synthetic problem. Since the most recent publications regarding this work,[1,2] substantial advances have been made in the capabilities of the computer program (termed LHASA). This paper describes certain of these advances which are of chemical interest and which have been implemented for some time now on both the PDP-1 and PDP-10 machines (Digital Equipment Corp.) at the Harvard Center for Research in Computing Technology.[3]

As indicated previously,[1,2] LHASA generates "trees" of synthetic intermediates from a target molecule by analysis in the antithetic (retrosynthetic) direction. The conversion of a structure in this "synthetic tree" to another structure, which corresponds to a synthetic precursor, is accomplished by the application of a "transform" (retroreaction).[1,2] Transforms which serve to simplify a structure are applied directly whenever possible. If such direct application of a certain simplifying transform is not allowable, it is often desirable to invoke certain nonsimplifying transforms (subgoals) that can generate structures which permit the simplifying transform to function effectively.[4] Straightforward procedures have previously been outlined[1,2] by which one or two such subgoals can be found to clear the way for simplifying transforms such as those of the disconnective two-group[2] type. The major purpose of this paper is to describe a method which can lead to the successful application of simplifying transforms even when the prerequisite is a sequence of *many* subgoals. Such a multistep look-ahead capability is extremely important in synthetic analysis involving complex polycyclic structures. Commonly, simplification of the network is *not* possible by either *direct* application of ring transforms[1,2] or *indirect* application involving a small number of subgoals. It is often necessary to extend the search for applicability over as many as *10 or 15* steps. This is partly due to the limited number of general ring transforms available for each ring size and partly due to the great structural variety encountered in organic molecules. An elementary example will demonstrate this point.

The oxidocyclohexane derivative I cannot be synthesized *directly* using any of the important six-membered ring forming reactions, for example, Diels–Alder addition, cation–olefin cyclization, Birch reduction, Robinson annulation, Dieckmann cyclization, or internal nucleophilic displacement. Consequently, a method of synthesis *cannot* be found by a simple matching process in this case, even with one or two subgoals. To apply the Robinson annulation, for example, I must first be modified to II and III, by successive functional group interchanges[1,2] and III must be dem-



ethylated. Three subgoals are required before the Robinson ring transform can be brought to bear on the problem. The synthesis of I outlined above can be derived by the use of a strategy which has been utilized[5] with much success by synthetic chemists. This strategy consists of several elements: (1) selecting a certain ring in a complex structure, (2) examining that ring for features which suggest that a particular ring transform may be applicable, (3) taking as an objective the application of such a transform, (4) identifying obstacles to such application, and (5) testing a series of approaches *via* intermediate structures (or subgoals) to clear away obstacles and to establish the required substructure for direct application of the ring transform. The recently described synthesis of fumagillin (VI) was devised using just such a strategy, in this case built around the application



VI

of the Diels–Alder transform.[6] Many, many other examples could be given. It is clear that the synthetic power of certain key ring-forming reactions is so important that an effective synthetic plan may be built around the use of such a reaction as a dominating concept.

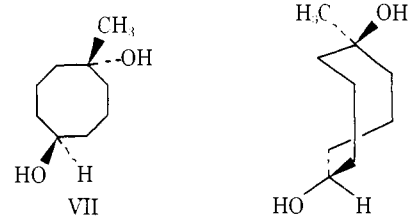
In this paper we focus on one of the most powerful of all ring transforms, the Diels-Alder process, to illustrate the implementation of this ring transform search strategy in LHASA. The search method described below for the Diels–Alder transform can be and has been extended with appropriate modification to the other ring transforms currently available to synthetic chemistry. The following sections deal with the techniques which have been developed for the perception of stereochemistry in LHASA,[7] for the implementation of an extended subgoal search, for the formulation and handling of the required data tables, and for the communication of the analytical process and results to the chemist.

Perception of Stereochemical Information. Due to the critical role played by stereochemistry in synthetic analysis (including processes of ring formation), the perception of the following structural features is required: (1) cis and trans relationships between ring substituents, (2) cis and trans relationships around olefinic units, (3) the "*E*" or "*Z*" character of asymmetrically substituted olefins, (4) the absolute configuration of tetrahedral chiral centers, (5) the axial or equatorial nature of substituents on certain cyclic systems, and (6) cis or trans fusion of rings. Before discussing the perception of stereochemistry as developed for LHASA, a brief outline of the method for input of stereochemical data is appropriate.

Commonly, stereochemical information in two-dimensional structural diagrams is conveyed by the use of wedge-shaped (flared) bonds and dotted (dashed) bonds, to represent bonds projecting above and below the plane of the page, respectively. The central role of computer graphics[8] for communication of all structural information between the chemist and the computer has made this standard convention particularly attractive for the designation of stereochemistry in LHASA. Within the graphical input displays of LHASA are a collection of control words or "buttons" [9] which allow the chemist to input structural information, including stereochemistry. To designate bonds projecting below the plane, for example, the user depresses the stylus[9] on a particular button (labeled DOTTED) and then points to the desired bonds. When selected by the stylus those bonds change from solid lines to dashed lines. Wedge-shaped bonds are specified in a similar manner (using a button labeled WEDGE). However, because of the directionality implied by a wedge-shaped bond, the chemist must point to the atoms at the ends of the desired bond, rather than to the bond itself. The second atom selected becomes the wide or "upper" end of the wedge.

Using a third button, the chemist then points to those chiral centers in the molecule at which stereochemistry has already been indicated by dashed or wedge-shaped bonds. A dot appears at each of the atoms selected. As well as informing LHASA of those atoms in the structure which are chiral, this acts as a double check to ensure that the configuration about the chiral atoms has been unambiguously defined. The dot will *not* appear, for example, if there are not four explicit atoms attached to the chiral carbon, or if the user has neglected to specify one or more of the attached bonds as dashed or wedge shaped.

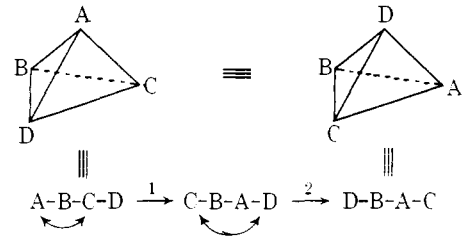
The perception[10] of cis or trans relationships about a ring or double bond by machine requires a more subtle process than the simple comparison of dashed/wedged bonds to the various substituents. The core of the problem is illustrated by a consideration of structures VII and VIII. Although



VII VIII

perception by machine of the trans hydroxyl arrangement in structure VII could be accomplished simply by comparing the wedged or dashed character of the two C–O bonds, examination of the alternate representation VIII reveals that this naive approach fails. By contrast, a trained chemist can correctly identify stereorelationships both in form VII and in form VIII by a process which must be more complicated. A general procedure for the perception of cis and trans relationships between ring substituents has been implemented in LHASA,[3a] which is unaffected by the orientation of the structure and which, therefore, imposes no artificial constraints on the way in which the chemist may draw the target.

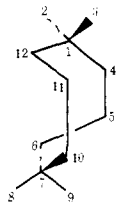
(1) In the first step, a linear representation is assigned to each chiral center in the molecule. This linear representation[11] is of the form "A–B–C–D," where A, B, C, and D are



the atoms at the points of the tetrahedron defined by the attachments to the chiral carbon. The first character, A, is considered to be *above* the plane of the other three atoms, which appear in the linear representation in *clockwise* order. It can be seen by inspection of the diagram below that a linear representation of a tetrahedron can be transformed into any of the other 11 equivalent representations of the same tetrahedron simply by interchanging pairs of atoms in the linear string an *even* number of times. If the configuration at a chiral center has been defined by the chemist with a wedge-shaped bond projecting above the plane, the atom attached to that bond is placed in the leftmost position of the linear representation, and the clockwise order of the other three attached atoms is determined from their graphical scope coordinates. These are positioned in the linear string in this clockwise order. If the defining bond projects "down" (an inverted wedge or dashed bond), the order of the other three atoms in the string must be reversed. Thus, the relative positions of the three rightmost atoms in the string will always represent a *clockwise* ordering of the three atoms when viewing *down* the indicator bond *toward* the stereocenter.

(2) Pairs of stereocenters on rings are then selected for comparison. The cyclic structure IX will be used to illustrate the following steps. Stereocenters 1 and 7 in this struc-

ture could be described by the linear representations 1 = 3-2-4-12 and 7 = 10-6-9-8.



IX

(3) Using the equivalency rule, *one* of the linear representations is then transformed to an equivalent so that pairs of atoms in each of the two paths linking the stereocenters appear in corresponding positions in the linear representations. For example, in structure IX one of the two paths between the stereocenters contains atoms 4 and 6, while the other contains atoms 12 and 10. To place 4 and 6 in corresponding positions, 7 = 10-6-9-8 is transformed to 7 = 10-8-6-9. To place 12 and 10 in corresponding positions, this is further transformed to 7 = 8-9-6-10. Comparison of the two representations indicates that the atom pairs are now aligned.

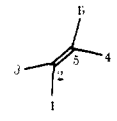
$$1 = 3\text{-}2\text{-}\boxed{4}\text{-}\boxed{12}$$
$$7 = 8\text{-}9\text{-}\boxed{6}\text{-}\boxed{10}$$

(4) In the final step, the two remaining pairs of atoms are examined. Adjacent atoms are always trans. Thus, atoms 2 and 9 are trans, as are 3 and 8. After the derived relationships are stored, the process is repeated with other pairs of stereocenters on the ring under study and for all rings in the molecule.

$1 = \boxed{3}\text{-}\boxed{2}\text{-}4\text{-}12$ $\quad$ $1 = \overset{\times}{3}\text{-}\overset{\times}{2}\text{-}4\text{-}12$
$7 = \boxed{8}\text{-}\boxed{9}\text{-}6\text{-}10$ $\quad$ $7 = \overset{\times}{8}\text{-}\overset{\times}{9}\text{-}6\text{-}10$

trans pairs $\quad$ cis pairs

The computation of cis and trans relationships between olefin substituents is accomplished by LHASA in a relatively straightforward manner. Linear representations are formed by calculating from scope coordinates the clockwise order of the three attachments to each of the olefinic atoms. In structure X above, the olefinic atoms may be represented as 2 = 5-1-3 and 5 = 6-4-2. These representations can be transformed into equivalent representations by applying the same operation used earlier for tetrahedral transformations.[12] After transforming one of the representations to an equivalent so that the olefinic atoms appear in corresponding positions in both strings,



X

$$2 = \boxed{5}\text{-}1\text{-}3$$
$$5 = 6\text{-}4\text{-}2 \longrightarrow 5 = \boxed{2}\text{-}6\text{-}4$$

the positions of the other two pairs are examined. Again, adjacent atoms are always trans.[13]

$2 = 5\text{-}\boxed{1}\text{-}\boxed{3}$ $\quad$ $2 = 5\text{-}\overset{\times}{1}\text{-}\overset{\times}{3}$
$5 = 2\text{-}\boxed{6}\text{-}\boxed{4}$ $\quad$ $5 = 2\text{-}\overset{\times}{6}\text{-}\overset{\times}{4}$

trans pairs $\quad$ cis pairs

LHASA also has the capability to differentiate between "*E*" and "*Z*"[14] isomers of asymmetrically substituted olefins. Once the "heaviest" substituent at each end of the olefin has been established by an atom-by-atom examining technique, LHASA assigns an *E* or *Z* descriptor based upon the cis and trans information perceived earlier. Similarly, the absolute configuration (*R* or *S*) of chiral carbon atoms is determined by an atom-by-atom "weighting" of each of the four attachments to the stereocenter, followed by a comparison of the ordered result with the already perceived linear representation of the stereochemistry at the chiral atom. The *R,S* perception techniques used in LHASA do not cover all conceivable stereochemical situations but are simple, fast, and adequate for synthetic analysis and capable of extension.

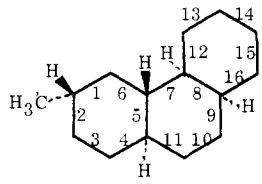
In the analysis of many polycyclic compounds, especially those which contain trans-fused cyclohexane rings, it is often necessary to consider the axial or equatorial character of ring substituents during the transform evaluation process.[2] LHASA's perception module determines such characteristics automatically, starting with the assignment of axial or equatorial orientation to a specific reference substituent. The orientation of that substituent may be input by the chemist or determined by LHASA through a simple calculation employing *A* values[15] of all groups attached to a nonbridged, saturated cyclohexane ring.

The final task of the stereochemical perception routine involves the identification of the stereochemical nature of all ring fusion bonds, employing the cis and trans relationships derived according to the procedure outlined above. Binary sets[10] (Figure 1) are created which denote cis fusion bonds, trans fusions, and ring fusions at which stereochemistry has not been defined. Information of this type is of considerable value in ring transform analyses (*vide infra*) such as the Diels–Alder transform, for example, where the direction of analysis followed by the program is dependent upon, among other things, the presence and stereochemistry of a ring fusion on the ring being studied.

All of the higher level stereochemical information developed by LHASA from the basic graphical representation of the molecule is stored either in "binary sets"[10] (Figure 1) or in "linked list" format[10] (Figure 2) for rapid access and use by other program modules. Until recently these data have been used primarily during the transform selection and evaluation processes,[2] especially in the ring-based analyses to be discussed shortly. It also has provided the necessary basis for the implementation of a variety of synthetically effective multistep stereochemical strategies.

Subgoal Generation by Binary Search. The implementation of the ring transform-oriented search strategies mentioned above requires procedures for the identification of obstacles to the application of the preselected ring transform and for the generation of subgoal sequences which operate to clear away such obstacles. This search for subgoals can be effectively handled by a question and answer process in the form of a binary decision pattern. Typically, an entry in the search pattern poses a question (*e.g.*, "endocyclic C=C present?," in the case of the Diels–Alder transform search), which leads to one follow-up question if the answer is yes or a different follow-up question if the answer is no, and the process is continued. Each element in the search pattern refers to the presence or absence of some structural feature on or near the ring under study or the applicability of some subgoal transform to modify an obstructing feature. Resulting from this form of search is one or more pathways of subgoals which lead from the target to an intermediate containing all the structural features required for the direct application of a particular ring transform.

The search pattern can be conveniently represented in the



FUSION = 000 010 010 000 000 ,,,,,

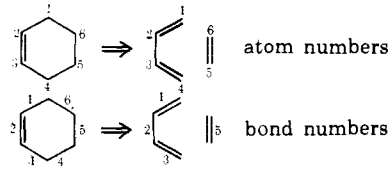
TFUSION = 000 010 000 000 000 ,,,,,

CFUSION = 000 000 010 000 000 ,,,,,

Figure 1. Examples of binary sets representing fusion bonds (FUSION), trans fusion bonds (TFUSION), and cis fusion bonds (CFUSION). Bonds are numbered according to the order in which they are input to the program. In a "set," the ith binary digit is a "1" if the ith bond (or atom) has the named property of the set. For example, in the set called TFUSION, bit 5 is a "1," indicating that bond 5 is a trans fusion bond. The utility of set representations and manipulations is discussed in ref 10.

form of a chemical flow chart. The ring chemistry flow charts utilized in LHASA have been developed by chemists from a very careful consideration and organization of the knowledge currently available. It is the function of these charts to direct a rigorous and exhaustive search to find as many ways as possible in which a particular transform can be used to disconnect[16] a particular ring. Figure 3 and Charts II–XI contain simplified flow chart representations of the search pattern for the Diels–Alder transform, one of several ring transforms that have been implemented in LHASA.

Structure of the Flow Charts. The Diels–Alder flow charts are used in the following way. A six-membered ring which appears in the target structure is assigned the temporary numbering scheme pictured below, such that atoms



1–4 represent the portion of the ring which will become the diene when and if the ring is finally disconnected. Entries in the charts refer to specific atoms or bonds according to these assigned numbers. The analysis of the ring is started at the top of Chart I (Figure 3) and the search proceeds down through that chart in a fashion which is determined by the answers to the structural queries contained in the chart. If during the search it is established that an obstacle to the application of the Diels–Alder is present, one of three courses is taken.

(1) If the obstacle is an undesirable characteristic that is a property of the ring as a whole, the ring is *rejected* from further study, and the next six-membered ring in the molecule is tested. In the present versions of the Diels–Alder search, rings that are aromatic or heterocyclic are rejected. Such "blockages" in the search pattern are merely situations in which further processing may not be warranted in view of the complexity of the needed modification.

(2) If a block occurs which is not serious enough to warrant complete rejection of the ring under study, the ring is *reoriented*. A new "orientation" of the ring is generated by reassigning the temporary atom and bond numbers one unit away from their previous assignments. The analysis is then restarted at the top of Chart I with the ring in its new "orientation." Since a ring can be reoriented six times before

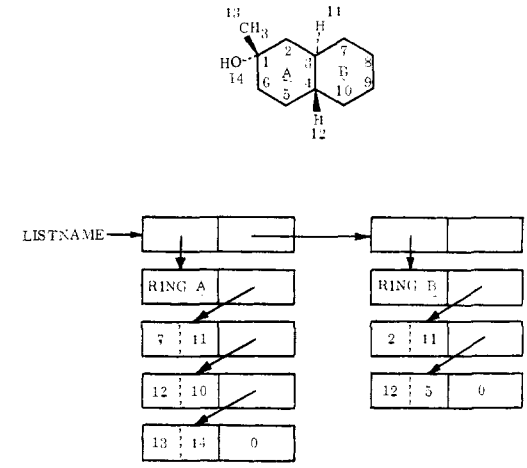
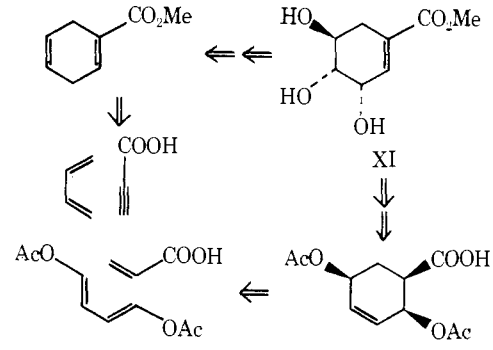


Figure 2. Example of singly linked list structure representing cis and trans relationships between ring substituents. All atoms appearing in the same data field of a sublist are cis. For example, in the left-hand data field of the "ring A" sublist, atoms 7, 12, and 13 are stored as cis. Similarly, in the right-hand field, atoms 11, 10, and 14 are cis. See ref 10 for a more detailed discussion of list storage and processing techniques.

duplication occurs, one ring can give rise to six different sequences leading to ring disconnection by the Diels–Alder transform. The benefits that may derive from such an exhaustive examination of a single ring are clearly illustrated in two published syntheses of methyl shikimate (XI).[17,18] It



is interesting to note that although both approaches employ the Diels–Alder reaction in the critical ring-forming step, different pairs of atoms were chosen to represent the dienophile portion of the ring. That is, the analysis of the synthetic problem was developed through two different ring "orientations."

(3) In most situations where an obstacle is detected, the flow charts direct the search to one or more subgoal steps intended to remove or modify the offending substructure. It is through a series of subgoals generated in this manner that chemical adjustments are made to clear the way for the direct application of the Diels–Alder transform. Once the disconnection has been performed, the ring is reoriented and the analysis is restarted at the top of Chart I (Figure 3). After all six orientations of a ring have been examined,[19] a new six-membered ring is selected and the procedure is repeated.

The search pattern for the Diels–Alder transform is represented in 11 flow charts, presented in simplified form in Figure 3 and Charts II–XI. Each chart is responsible for a very specific set of tasks in the analysis of a ring.

One of the principal tasks of the procedure expressed in Chart I (Figure 3), for example, is to detect at the earliest possible stage a variety of structural obstacles which could cause the ring to be rejected or reoriented, thereby minimizing unproductive analysis (level A of Chart I). Level B of

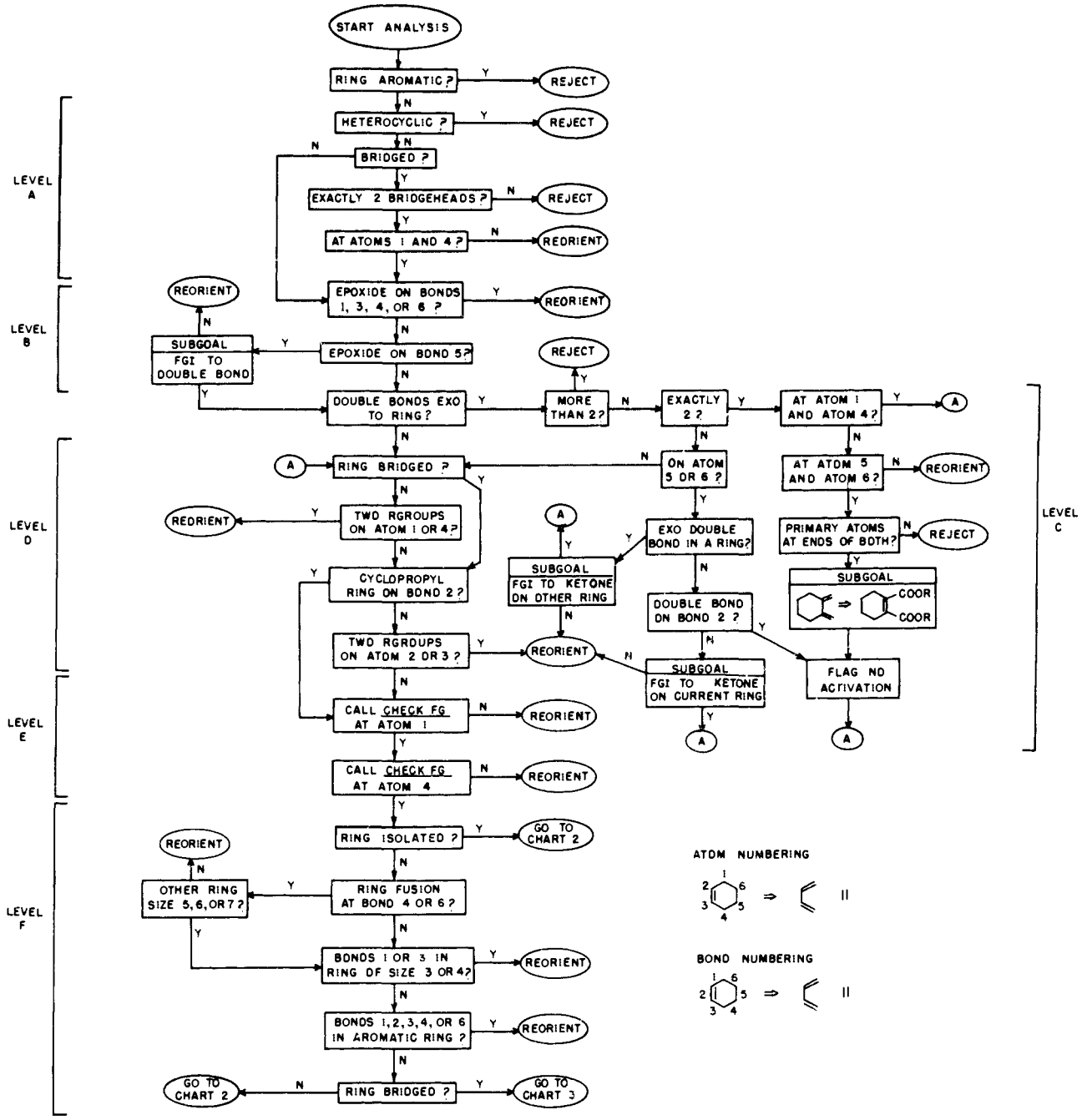
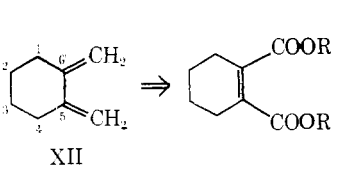


Figure 3. Chart I: A simplified version of the first flow chart to be entered in a Diels–Alder subgoal search. The chart is entered at the top and the flow of control is determined by the answers to the questions posed in the boxes (Y = yes, N = no). If the analysis of a ring proceeds successfully through the chart, control of the search is passed to Chart II or Chart III where the subgoal search is continued.

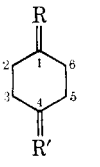
this chart deals with the occurrence of a 1,2-oxido group at various positions around the ring, accepting for further consideration only two of the six possibilities: (1) an epoxide function located on the two carbons of the ring which may become the dienophile,[20] or (2) an oxygen function bridging the two inner carbons of the diene moiety.

In level C the search pattern is concerned with double bonds which are exo to the ring under current study. The occurrence of more than two such bonds causes the rejection of the ring from further study. If exactly two double bonds are attached to the ring, all but two of the possible substitution patterns are discarded. Those two are indicated in structures XII and XIII.

Structures related to XII are dealt with immediately by the operation of a subgoal which produces an intermediate which could be expected to result from the addition of di-



XII

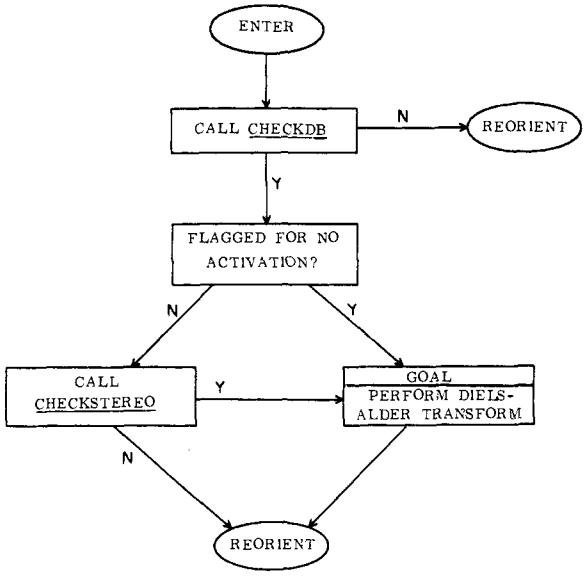


XIII

(R, R′ = alkyl)

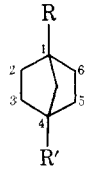
methyl acetylenedicarboxylate to a diene. An "activation flag" is then recorded to indicate that the dienophile portion of the ring need not undergo any further activation before the Diels–Alder disconnection is finally applied. This record, which is used in several locations in the search pattern, is a useful technique for passing information along through the charts for access at a later point (*e.g.*, in Chart II).

Chart II[a]

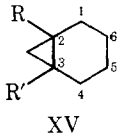


[a] Control of an analysis enters this chart from either Chart I or Chart III. It is in this chart that the Diels–Alder disconnection is finally applied.

Rings which contain a high degree of carbon substitution on the diene portion (atoms 1 to 4) are identified in level D and reoriented. Exceptions to this test are indicated by structures XIV and XV below. Structures of type XV are



XIV



XV

accessible by addition of methylene or "carbene" to the double bond between carbons 2 and 3 which is required for the eventual operation of the Diels–Alder transform (see Chart VI).

Situations frequently arise in the search pattern in which the same series of questions must be asked about different atoms or bonds in the ring. To save space, the series of questions is included in the charts only once and is assigned a name. In level E of Chart I, for example, it is necessary to check for the occurrence of obstructing functionality on atoms 1 and 4. This series of questions is represented in a separate flow chart, called CHECKFG (Chart VIII). When CHECKFG is "called" for atom 1, the search pattern is directed to the CHECKFG chart. The analysis continues through that chart and a series of subgoals is generated, if necessary, to adjust the functionality at atom 1. Control is then returned to level E of Chart I with either a success or a fail indicator. If the call to CHECKFG is successful (*i.e.*, if the functionality can be modified successfully), the process is repeated for atom 4. Since this is analogous to the programming concept of subroutining, those charts that are accessed by name and return eventually to their calling location are termed "subroutines." Charts IV through XI are subroutine charts.

The final task of Chart I involves an examination of ring fusions at various locations around the ring under study (level F). The analysis is then continued in Chart II or Chart III. The functions of the remaining ten flow charts will be discussed briefly.

Chart II and Chart III. These charts direct the analysis to the major subroutine charts CHECKDB and CHECKSTEREO. A test is performed in Chart II to determine if it is necessary to provide additional activation of the dienophile portion of the ring.[21] It is in Chart II that the operation of the Diels–Alder transform is finally invoked. Except for Chart III none of the other charts may terminate an analysis by a reorient or reject exit, since they are all subroutine charts. Control of the analysis in Charts IV–XI always returns to Chart II eventually. Rings which are bridged are first subjected to a screening operation in Chart III; control is then passed to Chart II.

Chart III. For bridged rings

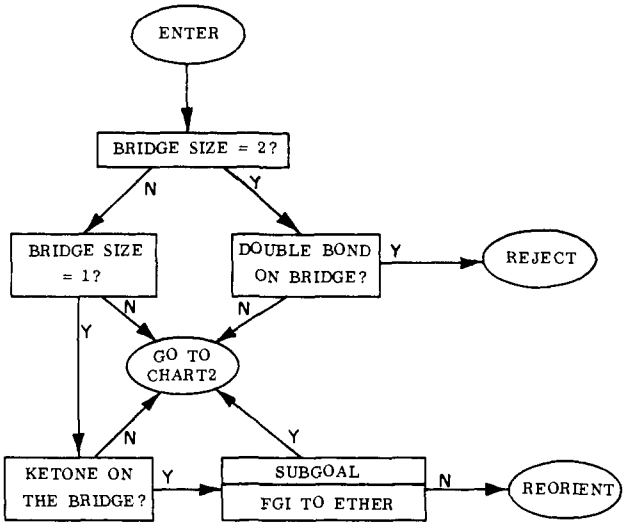
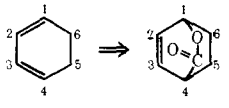


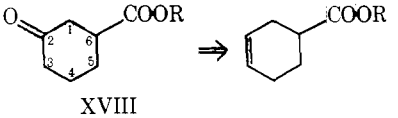
Chart IV. The CHECKDB subroutine is given the task of inserting a double bond at the bond 2 position in the ring. This may be accomplished by a variety of methods, including functional group interchange, functional group addition, or double bond migration from bond 1 or bond 3. If the ring already contains a conjugated diene system between atoms 1 and 4 as in XVI, a carbon dioxide cycloelimination subgoal is applied (XVII) to locate a double bond in the required position.



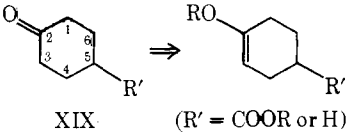
XVI XVII

Chart V. CHECKSTEREO begins by modifying obstructing functionality on the dienophile portion of the ring, if necessary. The stereochemistry of the dienophile substituents is examined and adjusted to provide the isomer that would be expected to result from endo addition of the dienophile to the diene. If bond 5 is a ring fusion then the size of the fused ring must be considered as well as the stereochemistry of the ring juncture.[22] It is also the function of CHECKSTEREO to provide sufficient activation of the dienophile portion of the ring.

Chart VI. PUTDB is called at several points within the CHECKDB chart to insert a double bond on bond 2 by either FGI or FGA subgoals. If a ketone is identified on atom 2, for example, and atom 6 contains a carboxylic function (XVIII), an FGI is invoked to convert the ketone to a dou-

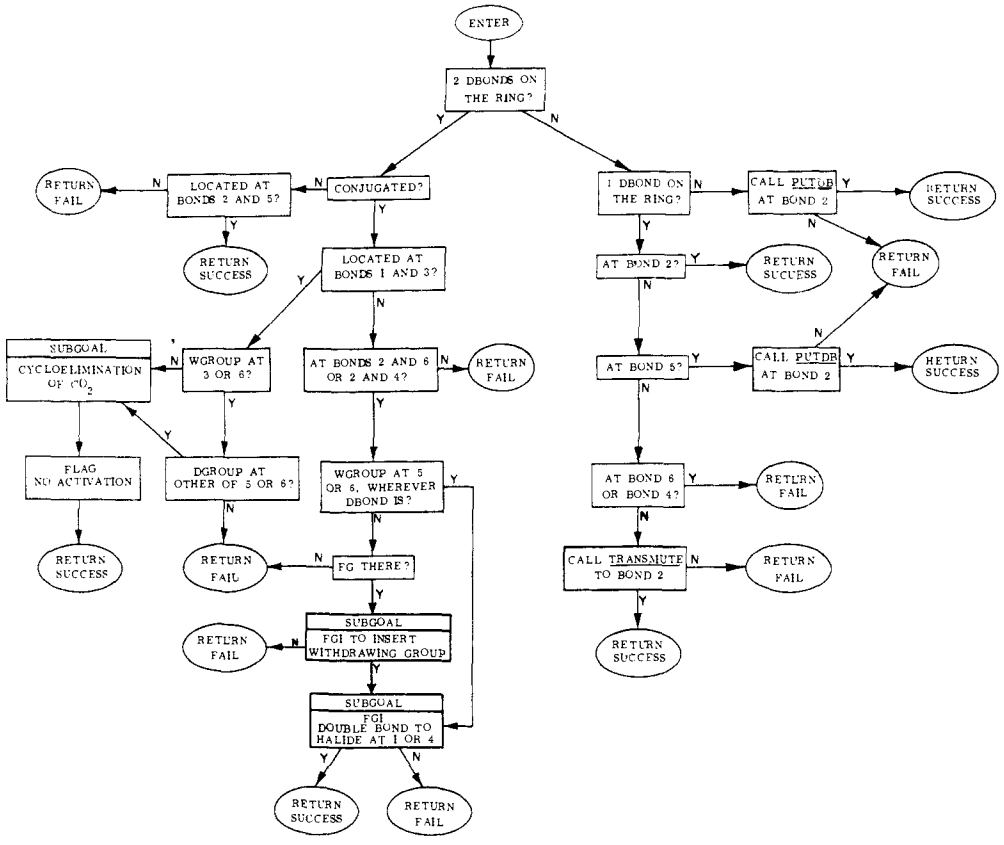


XVIII



XIX (R' = COOR or H)

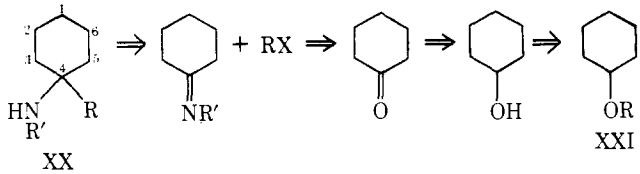
Chart IV, Subroutine CHECKDB



ble bond. The carboxylic group allows oxygen functionalization at the desired location (atom 2) *via* an intermediate halolactonization step. Such assistance is not possible for structures of type XIX, so the ketone is converted to an enol ether.

Chart VII, PUTW attempts to activate the dienophile portion (atoms 5 and 6) by the application of FGI or FGA subgoals. Electron withdrawing groups are produced α to atoms 5 or 6 according to the stereochemistry specified in the various PUTW calling locations in CHECKSTEREO.

Chart VIII, The subroutine chart which analyzes functionality at atoms 1 and 4 is called CHECKFG. A successful return to its calling location in Chart I indicates that the functional groups on atoms 1 or 4 are acceptable or that acceptable functionality has been produced by a series of FGI subgoals. Sequences involving up to four subgoal steps may be generated (XX to XXI).



XX XXI

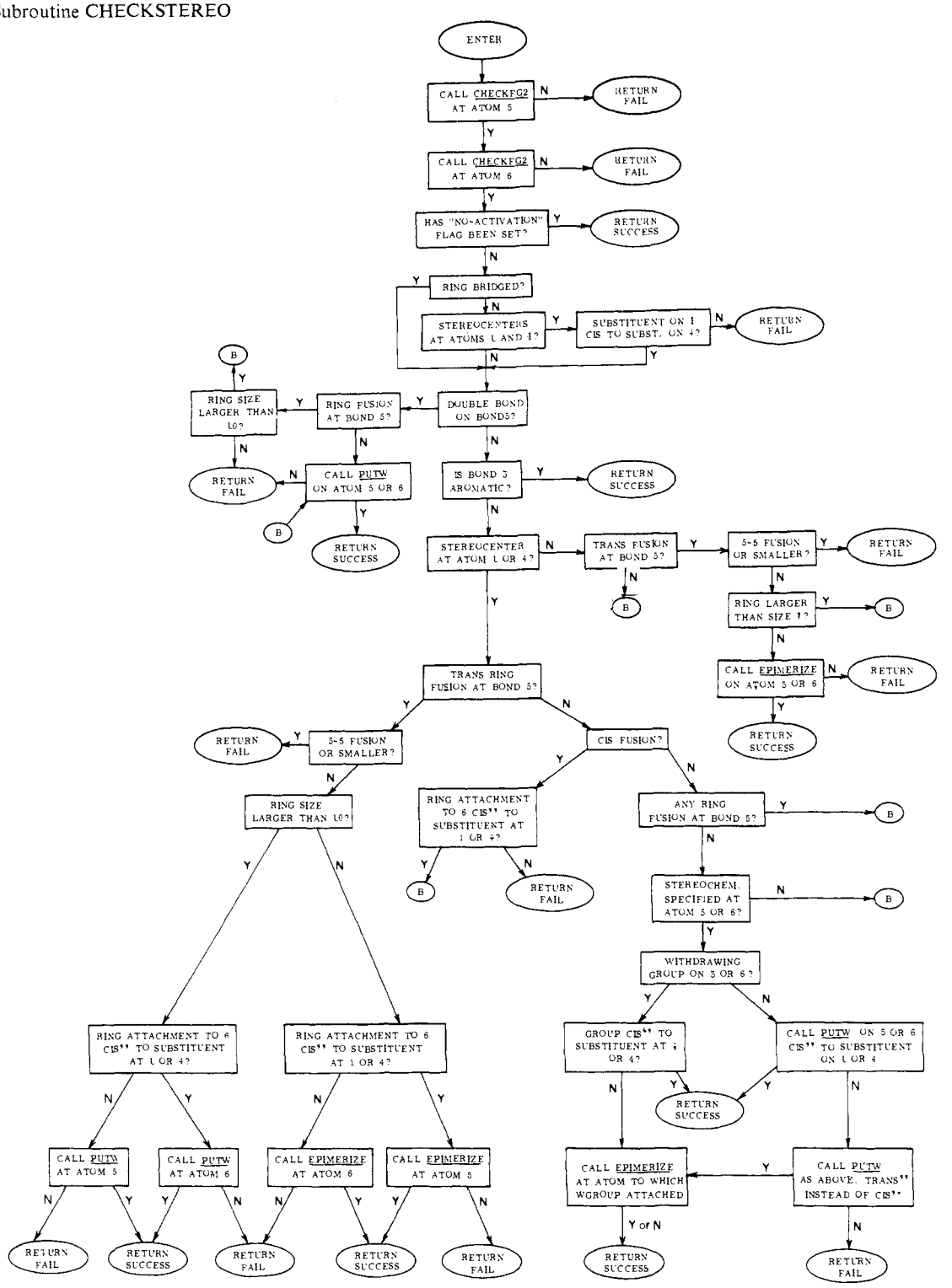
Chart IX and Chart X, CHECKFG2 and EPIMERIZE are subroutines called from CHECKSTEREO, the former to remove obstructing functionality from the dienophile portion of the ring, and the latter to determine if it is possible to invert the configuration of a chiral atom by an epimerization step. EPIMERIZE will also insert a withdrawing group (by a call to PUTW) α to the atom under study, to assist in the epimerization process.

Chart XI, TRANSMUTE is called by CHECKDB in its attempt to situate a double bond at the bond 2 position. This is accomplished by moving a double bond from bond 1 or bond 3, in the retro version of a $\beta,\gamma \rightarrow \alpha,\beta$ (conjugated) double bond transposition.

The flow charts represent a first generation search pattern for the Diels–Alder transform; a more sophisticated version of the tables is currently under development. This will include a capability for modifying the structure to obtain a symmetrical diene or dienophile, especially in cases where position specificity of the Diels–Alder process is unfavorable or doubtful. In connection with this extension, data for the evaluation of positional orientation (regiospecificity) in the Diels–Alder addition need to be included.[23] A much more powerful multistep FGI capability which has recently been developed will be described elsewhere.

Method of Implementation of Ring Transforms, The chemical flow charts described above are a convenient representation of search patterns to apply ring transforms. The technique of using an "interpretive data table" has been imployed no less successfully in LHASA for representation of the information contained in the flow charts. As outlined previously,[2] with each of the individual transforms in the two-group, one-group, FGI, and FGA categories contained in the LHASA data base is associated a collection of queries concerned with structural features which bear on the proper functioning of that transform. These queries are written in a language based on "chemical english" which has been designed to be computer-readable and intelligible to a chemist with little or no programming experience. Each of the queries or "qualifiers" [2] in the data table entry for a transform of above-mentioned types contains information, based on current knowledge of reaction scope, to allow an evaluation of the structural features in a target structure so as to determine whether they can be expected to help or hinder the progress of the corresponding reaction in the synthetic direction. Each qualifier specifies a numerical value which is to be added to or subtracted from a basic transform rating if a structural unit indicated in the qualifier is present in the molecule. After all qualifiers in a transform table entry have been checked, the resulting modified rating is used to determine whether or not the transform will be al-

Chart V. Subroutine CHECKSTEREO



[a] Double asterisks (**) which appear in this chart indicate that the opposite stereorelationship to that specified is to be used if the ring is bridged (*e.g.*, trans instead of cis).

lowed to operate on the target molecule. By this means the applicability of candidate transforms is assessed by LHASA.

Since the most recent publication regarding this data table language (seen in prototype form in ref 2), substantial advances have been made in its flexibility, the most notable extension being in the area of ring transform search patterns. As originally developed for the *group-oriented chemistry packages*, the "chemical english" data table language was concerned exclusively with modifications to the transform ratings, according to the presence or absence of particular structural features. The more complex nature of the queries and commands implied in the *ring transform search patterns* required considerable enhancement of the language capabilities. A sample of the current Diels–Alder table which has been written in this language is presented in Table I. The top half of Table I illustrates a portion of Chart I (Figure 3), while the bottom half is a section of the CHECKSTEREO subroutine. As can be seen, the Diels–Alder table representation which is "read" by the computer is very close to a direct mapping of the search pattern presented in the flow charts.

Chart VI, Subroutine PUTDB

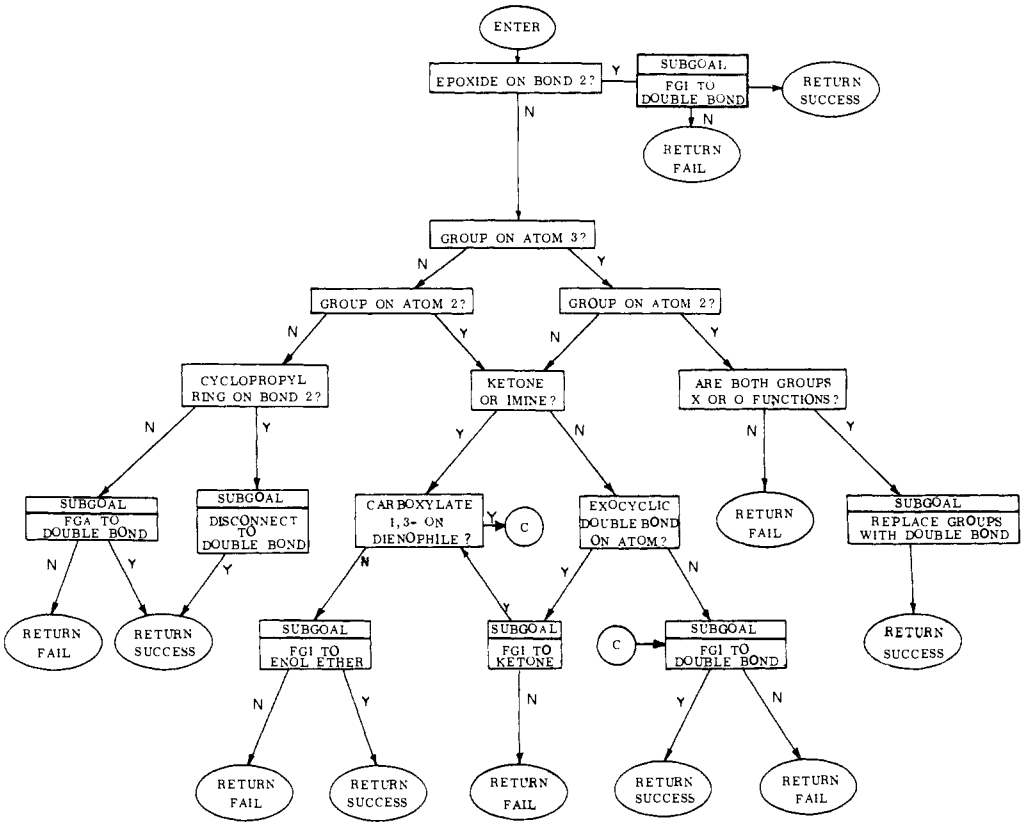


Chart VII, Subroutine PUTW

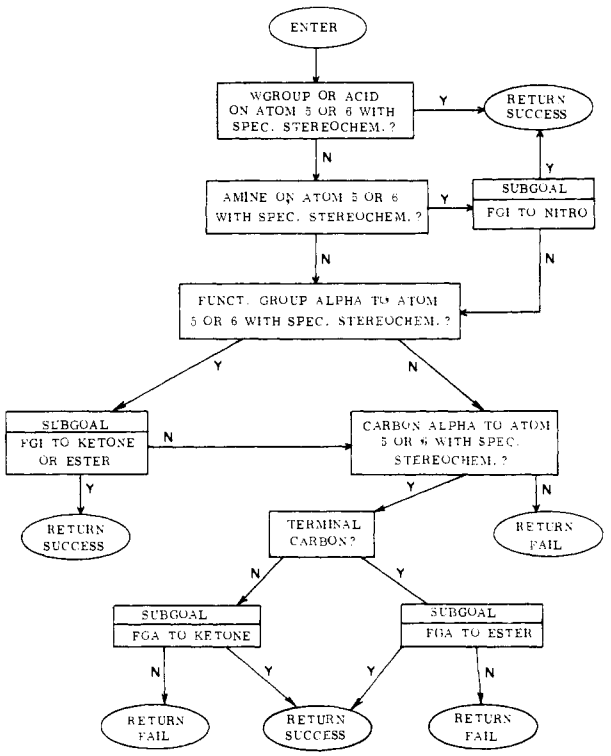
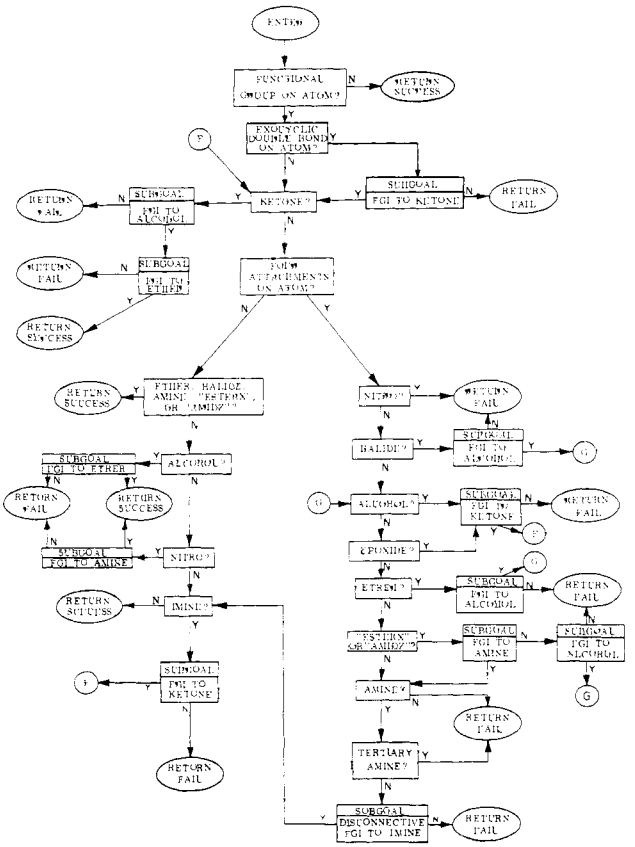


Chart VIII, Subroutine CHECKFG



The portion of the Diels–Alder table pictured in Table I is presented merely to illustrate the technique by which the information presented in the flow charts is represented in machine-readable form. A more detailed discussion of the full capabilities of the language, together with a technical documentation of the table language interpreter, will be presented elsewhere.[3a,b,24] However, several points of a general nature can be made. The table language has been designed to minimize the effort required in the transcription of information from flow chart form to "machine-readable" form. It permits ring reorienting and ring rejection operations (Table I, lines 1 and 5) just as in the chemical flow charts, as well as the use of subroutine-calling (Table I,

Table I. Portions of LHASA's Current Diels–Alder Table

START:	IF THE RING IS AROMATIC THEN REJECT
	IF THE RING IS HETEROCYCLIC THEN REJECT
	IF THE RING IS NOT BRIDGED THEN GO TO A9
	IF THERE ARE NOT TWO BRIDGEHEADS ON*THE*RING THEN REJECT
	IF CARBON*1 IS NOT A BRIDGEHEAD OR CARBON*4 IS NOT A BRIDGEHEAD THEN REORIENT
A9:	IF THERE IS AN EPOXIDE ON BOND*1 OR AN EPOXIDE ON BOND*3 THEN REORIENT
	IF THERE IS AN EPOXIDE ON BOND*4 OR AN EPOXIDE ON BOND*6 THEN REORIENT
	IF THERE IS NOT AN EPOXIDE ON BOND*5 THEN GO TO Q2
	EXCHANGE IT FOR AN OLEFIN . . . *this is an FGI subgoal call*
	IF UNSUCCESSFUL THEN REORIENT
Q2:	IF THERE ARE MORE THAN TWO OLEFINS ALPHA*TO*RING THEN REJECT
	IF THERE ARE NOT TWO OLEFINS ALPHA*TO*RING THEN GO TO A1
	. . . *continued*
CHECKSTEREO:	CALL CHECKFG2 AT CARBON*5 AND IF UNSUCCESSFUL THEN RETURN FAIL
	CALL CHECKFG2 AT CARBON*6 AND IF UNSUCCESSFUL THEN RETURN FAIL
	IF THE FLAG*IS*SET THEN RETURN SUCCESS
	IF THE RING IS BRIDGED THEN GO TO E1
	IF THERE IS NOT A STEREO CENTER AT CARBON*1 OR NOT A STEREO CENTER AT CARBON*4 THEN GO TO E1
	IF THE ATOM ALPHA TO CARBON*1 IS NOT CIS TO THE ATOM ALPHA TO CARBON*4 THEN RETURN FAIL
E1:	IF BOND*5 IS AROMATIC THEN RETURN SUCCESS
	IF BOND*5 IS A SINGLE*BOND THEN GO TO E3
	IF BOND*5 IS A FUSION*BOND AND BOND*5 IS NOT IN A RING*OF*SIZE*LARGER*THAN*10 THEN RETURN FAIL
E2:	IF THERE IS A WITHDRAWING GROUP ON CARBON*5 OR A WITHDRAWING GROUP ON CARBON*6 THEN RETURN SUCCESS
	CALL PUTW AT ANY*STEREO AND GO TO F9
	. . . *continued*

Chart IX. Subroutine CHECKFG2

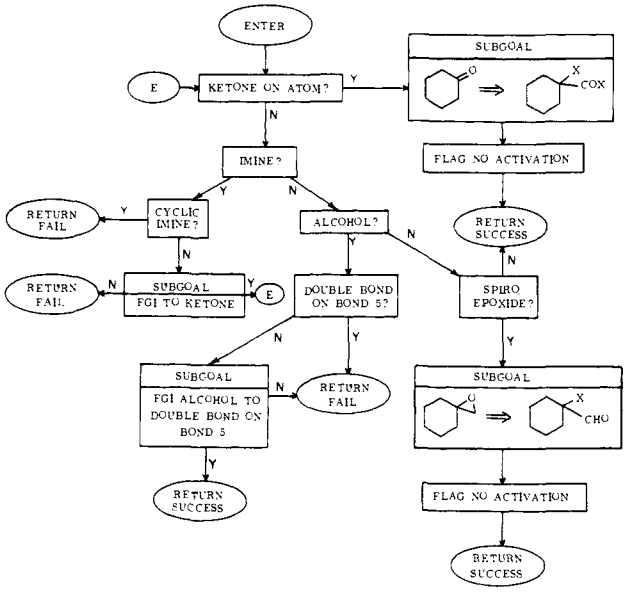
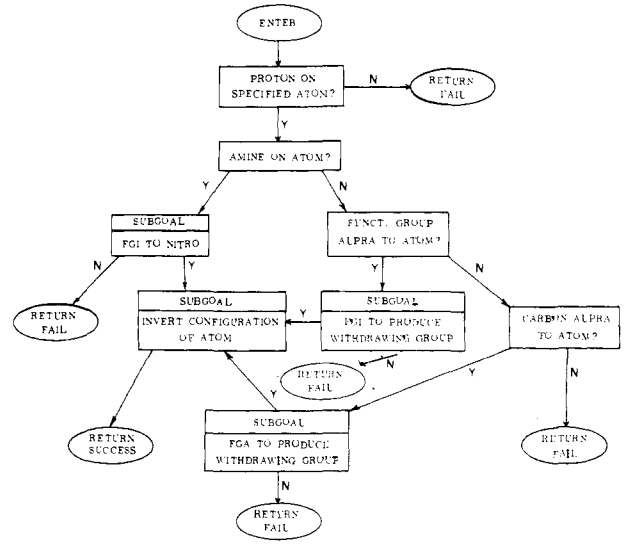


Chart X. Subroutine EPIMERIZE

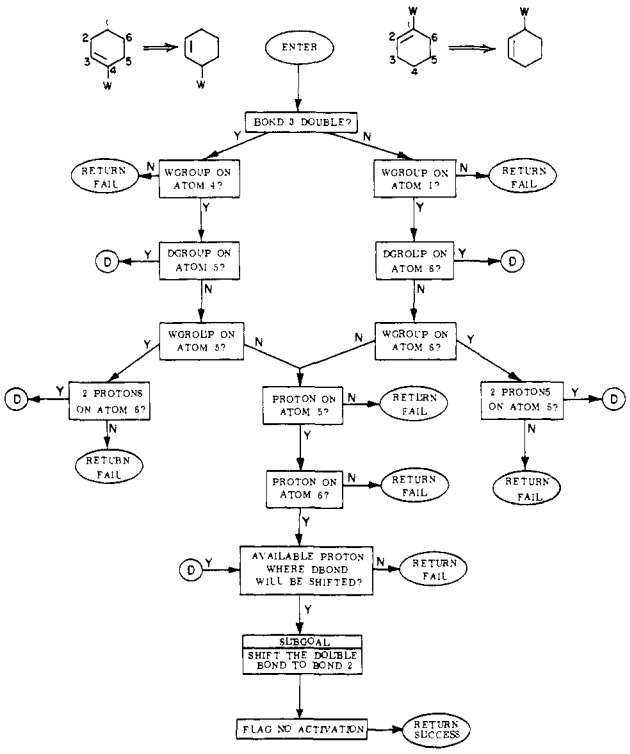


lines 13 and 14) and subgoal-requesting statements (Table I, line 9).

The vocabulary of the language can be subdivided conveniently into six syntactic units (see Appendix).[24b] The first unit contains terms which act either as the "target" of the sentence, identifying features that are to be searched for (*e.g.*, ACETAL), or as commands (*e.g.*, CALL or REJECT). Members of the second syntactic class are called "reaction condition specifiers" which can be used in the chemical data tables to describe the reagents employed in the reaction.[25] The third unit contains location specifiers, or "locants," which permit the specification of a site or location in the structure under analysis for the performance of some action or query. The remaining three categories include terms which allow structural modifications and rating changes, numerical specifications, and the elaboration of sentences to enhance intelligibility.

The ring chemistry tables of LHASA are not executed directly by the computer but are examined as *data* by another program module called an "interpreter." When an analysis is performed by LHASA, each table statement is examined in turn by the interpreter, to decode the sentence by means of a prearranged numerical value assigned to each of the "english" words in the table vocabulary (Appendix). In addition LHASA performs internally various perception operations, subgoal references, subroutine calls, and assorted other specifications that appear in the symbolic version of the "chemical english" table. Modifying, updating, or improving the capabilities of a ring transform analysis is therefore reduced to the simple task of rewriting appropriate table statements. No modifications to the interpreter or other program modules are required.[26] Additionally, no "programming" is involved in the addition of new search patterns to LHASA's ring transform library since the interpretive table language permits rapid and direct transcription of the information contained in the flow charts into a

Chart XI, Subroutine TRANSMUTE



form that is "readable" by the computer. The time-consuming factor in the addition of new ring transforms to LHASA is the detailed study and analysis that is necessary for the development of an effective search pattern for a very broad range of organic structures.

Form and Output of Computer Analysis. One of LHASA's graphical displays presents to the chemist a "menu" of approximately 2 dozen analysis options. Associated with each of these options is an internal analysis module which draws upon transforms in one or more of the five categories defined earlier,[1,2] and determines the applicability of these transforms to the target structure, according to a variety of strategies. One of the search options, for example, attempts to simplify the target molecule by the application of any of the 140 two-group transforms[2] in the transform library to cleave strategically located cyclic bonds. When necessary, this module will also invoke several levels of intermediate nonsimplifying steps (FGI or FGA subgoals) in order to facilitate the eventual operation of a small subset of major simplifying transforms.

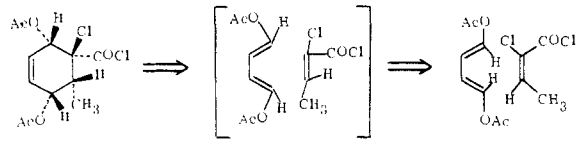
To initiate the analysis of a molecule by a ring transform package, the user simply selects from the graphical menu the desired option.[27] In response, the following actions are taken by the program: (1) a processing phase is entered in which a large number of synthetically significant structural features (including stereochemistry) are perceived[10] for use later in the analysis, (2) a ring of size appropriate to the ring transform specified by the user is selected, (3) a temporary numbering scheme is assigned to the atoms and bonds in the ring (such as the scheme presented above for the Diels–Alder charts), and (4) the search pattern for the selected ring transform is entered.

In the Diels–Alder search pattern, for example, the interpreter module enters the portion of the table pictured at the top of Table I, which corresponds to Chart I (Figure 3). "Reorient" and "reject" exits are handled by the program exactly as described earlier; if the condition described in a "reject" sentence is found to be true (*e.g.*, IF THE RING IS AROMATIC THEN REJECT; line 1 of Table I), the interpretation procedure is terminated, a new six-membered ring is selected, and the interpretation procedure is restarted at line 1. Similarly, when a "reorient" exit is taken, the interpretation procedure is halted, a new numbering scheme for the current ring is generated, and the table is entered again at the top.

The process of *subgoal evaluation* plays a major role in the development of valid subgoal sequences which lead to the final application of the Diels–Alder disconnection, since it is during this evaluation procedure that it is determined whether an individual *synthetic* reaction (corresponding to an antithetic subgoal request) is chemically feasible. If the evaluation is favorable, the transform corresponding to the subgoal is validated and the retrosynthetic sequence generated from the current target molecule is extended. This subgoal evaluation process has been discussed in detail in an earlier publication[2] and will be dealt with here only briefly. Line 9 of Table I illustrates a simple subgoal call (EXCHANGE IT FOR AN OLEFIN) corresponding to a required conversion of a 1,2-oxido function conjunct with bond 5 to a double bond by functional group interchange. At this point the sequential interpretation of the Diels–Alder table is suspended while another program module evaluates the applicability of the requested subgoal. A library of functional group interchange transforms is consulted to determine (a) that the EPOXIDE $\rightleftarrows$ DOUBLE BOND FGI is "known" to LHASA, and (b) that the epoxide function could be produced *synthetically* in reasonable yield from a double bond at the bond 5 position, that is, that the FGI is valid in the *current molecular context.* Associated with each FGI record in the FGI transform library is a basic numerical "rating" for the transform and a collection of queries whose function it is to modify this rating based on the presence or absence of structural features which would be expected to affect the course of the reaction in the synthetic direction. In the EPOXIDE $\rightleftarrows$ DOUBLE BOND library entry, for example, one of the associated qualifiers causes a large decrement of the transform rating if other double bonds are present in the molecule. This reflects the additional difficulty that would be encountered in *selective* epoxidation of the desired double bond. These structural queries are represented in the transform library in another form of LHASA's "chemical english" interpretive language, a prototype version of which has been described previously.[2] After all of the structural queries associated with a particular transform have been examined by the language interpreter module, the cumulative rating is used to determine whether the FGI transform is accepted or rejected. If the evaluation of the subgoal is successful,[28] the requested modification is applied to the target and the new offspring structure is displayed to the user. While the chemist is examining this intermediate, control is returned to the interpreter module which resumes the search where it left off (*i.e.*, line 9 of Table I). Typically, the next table statement poses a question about the result of the subgoal attempt and, in this case, causes the ring to be reoriented if the requested interconversion was not successfully executed (line 10 of Table I).

As each new intermediate which results from the application of a subgoal is displayed, the "synthesis tree" [8] which appears on a second cathode ray tube display is updated to reflect the progress of the analysis.[29] Eventually, the pathway of subgoals leads from the target compound to an intermediate which contains all the structural features required for the direct application of the Diels–Alder transform. The program is then instructed by an appropriate section of the search table to perform the Diels–Alder disconnection of the current ring (see Chart II) and the disconnected structure is displayed to the chemist. In the process of generating a structure for output, the program must not

only make and break bonds around the ring but also define the stereochemistry of the diene and dienophile reaction components. Examination of structure XXIII, for example,



XXII XXIII XXIV

which results from the direct disconnection of XXII, clearly illustrates the necessity of an intervening step which relocates the ring substituents to reflect the true stereochemistry at the terminal carbons of the diene and about the dienophilic double bond. LHASA performs this step internally and displays only XXIV to the chemist. Future generations of offspring structures are then formed in consonance with the correct stereochemistry of the disconnected offspring of XXII.

After a disconnected structure has been generated, the analysis is restarted on the current target[30] with a new orientation of the ring under study. After LHASA has attempted the analysis for all six orientations of the ring, a new candidate ring is selected and the interpreter module repeats the process from the beginning of the interpretive search table. When all the candidate rings in the molecule have been tested in this fashion, a message ("YOUR MOVE") is flashed on the display, indicating that the current Diels–Alder search pattern has been completed. The chemist may then direct further analysis.

Even though a six-membered ring may give rise to six different Diels–Alder sequences, this rarely happens. Typically, between zero and three complete sequences are generated for each cyclohexane ring in the molecule. The development of many pathways is interrupted by a "block" or an obstructing feature which cannot be modified or circumvented by any subgoal process available to LHASA. In such situations the data table forces a "reorient" exit to be taken so that a new ring orientation may be tested. The (current) intermediate being displayed is erased and the branch of the synthesis tree which represents the partially developed pathway is automatically truncated.

The time required for an exhaustive analysis by a ring transform search pattern depends on a number of factors, including the complexity of the individual rings, the number of candidate rings in the structure, and, most notably, the length of time spent by the chemist in the examination of each of the intermediates in the pathways that are generated. When a subgoal is applied to a structure and the new offspring is displayed, the search process is continued automatically. But when the *next* offspring structure in the sequence has been generated,[31] LHASA waits before displaying the new structure until the user depresses the stylus on the tablet to indicate that he has finished examining the preceding intermediate. Thus, despite the unusual complexity of the search that is being performed by LHASA, the chemist very rarely has to wait to see a new structure. The generation of new intermediates appears to the user to be almost instantaneous.

The length of the pathways generated by this technique (or "look-ahead") depends on the structure under analysis. On the average between four and six intermediate subgoal steps are required for the successful application of the Diels–Alder disconnection. However, synthetic sequences of 15 or more steps have occasionally been generated by LHASA in situations which have required an unusual degree of modification of the candidate ring. Sequences of this length typically contain one or more subsequences in which a *single* functional group must undergo several successive FGI's to produce functionality appropriate to the requirements of the Diels–Alder transform.

Figure 4 presents a collection of relatively straightforward synthetic pathways generated by the current Diels–Alder analysis package. The retrosynthetic sequence A ⇒ I illustrates one of several paths developed by LHASA when structure A is processed as a target. One notable feature of this sequence is that the length of the pathway is due in large part to the A ⇒ E subsequence in which a single amino group is modified by a series of FGI subgoals to produce a methoxy function which fits more closely the requirements of the Diels–Alder disconnection.

When structure C is processed as a target, several pathways not obtainable[32] from the structure A target level are developed, two of which are pictured in Figure 4 (C ⇒ L and C ⇒ S). When D is processed as a target, several more sequences appear, one of which (D ⇒ P) involves an interesting *internal* Diels–Alder addition as the critical ring-forming step.

The four pathways presented in Figure 4 have been selected to illustrate the effectiveness with which a *single* ring transform may be brought to bear on a cyclic system through exhaustive examination of all ring orientations according to LHASA's subgoal search procedure. To facilitate manual comparison of the chemical flow charts (Figure 3 and Charts II–XI) with the subgoal steps illustrated in Figure 4, each subgoal process has been labeled. The term which appears above a double arrow identifies the chemical table subroutine in which the subgoal was applied. In parentheses below the double arrows are listed, in order of nesting, the table subroutines from which the active subroutine was called.

Discussion

The analytical technique described above permits a multistep antithetic look-ahead to ascertain whether a given target structure can be dissected to simple precursors by use of the *Diels–Alder* transform. There are a number of other ring transforms applicable to the disconnection of six-membered carbocyclic rings which correspond to sufficiently useful synthetic ring-forming processes to justify being taken as goals for the same kind of multistep, binary search look-ahead. The list of such transforms includes the following important members: Diels–Alder, Robinson annulation, Birch, internal cation–olefin addition*, Dieckmann*, internal aldol*, internal Michael*, internal nucleophilic (SN2) displacement*, internal Friedel–Crafts*, acyloin*, internal pinacol*, electrocyclic closure of 1,3,5-hexatrienes, internal diazo–carbonyl addition*, internal diazo–olefin addition*, internal nitrone–olefin addition*, replacement of B by C in cyclic boranes* (asterisk indicates that the transform is also applicable to other ring sizes). The addition of all these data tables to LHASA to direct the search to apply these transforms should present no serious difficulties, although great care and skill must be exercised in devising the required "chemical" flow charts. Ideally, each data table should contain information which leads automatically to the calculation of a limit on the depth of the search to be carried out for the corresponding transform.[33] In general, a target structure may possess features which are suggestive that a particular transform can be applied advantageously or suggestive that a particular transform is inappropriate. Therefore, the depth of the search (zero to n) to be carried out will be a function of the transform in question and the number and nature of the beneficial or adverse features in the target. These same factors, and others (*e.g.*, the nature of the network containing the ring in question, the presence of interfering functionality elsewhere in the molecule) can also be utilized to generate a priority list for the ordered se-

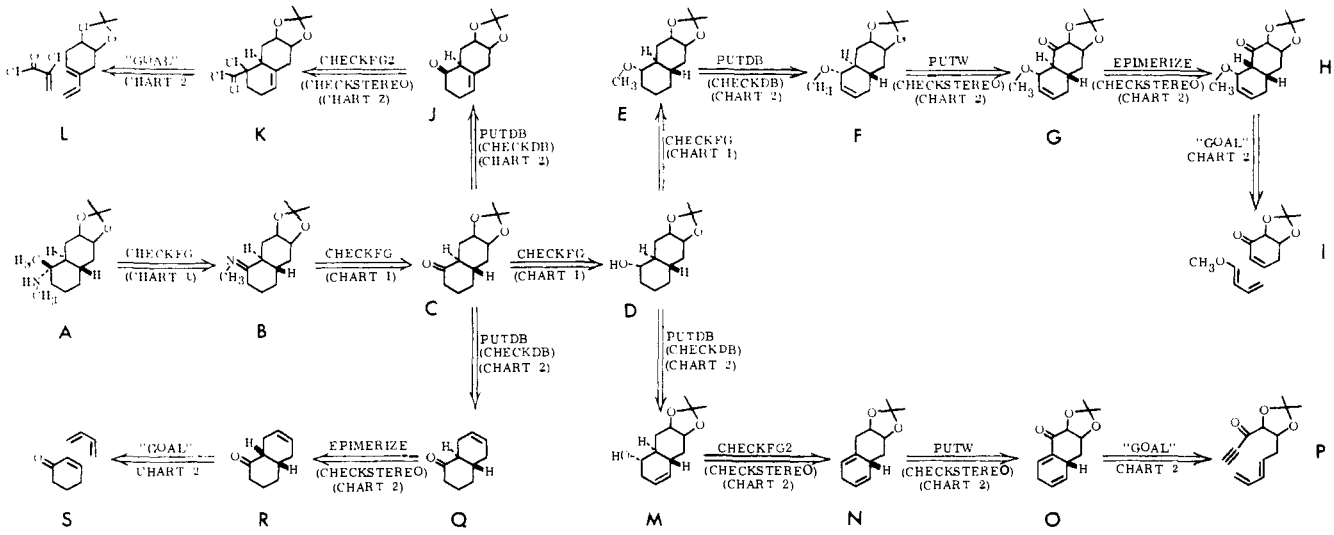
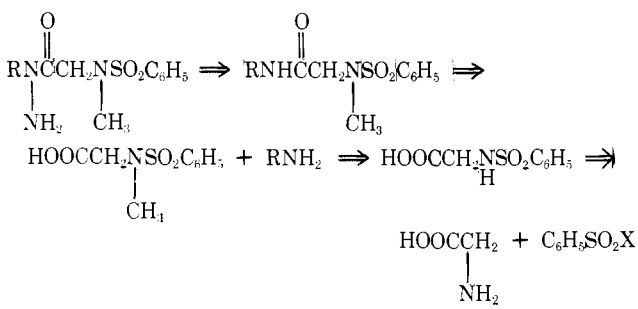


Figure 4. Sample output from LHASA's Diels–Alder analysis package. The words appearing above the double arrows indicate the location in the search pattern in which the individual subgoals were invoked. Parenthesized names appearing below the double arrows indicate the degree of subroutine nesting at the time the subgoal was performed. For example, the FGA subgoal applied to structure F to produce G was invoked by PUTW, which was called as a subroutine by CHECKSTEREO, which was called as a subroutine by the Chart II portion of the search pattern.

lection of ring transforms for application to a given target or to a particular ring within a polycyclic target structure.

Clearly, the same task will have to be undertaken for other ring sizes, 3–5 and 7–*n*. At this stage, very effective search tables have been written for certain other ring transforms, *e.g.*, the Simmons–Smith process for generating three-membered rings, and the Robinson annulation for six-membered rings. For purposes of synthetic analysis involving *olefinic* structures, double bonds may be considered as two-membered rings and search tables similar to those for conventional rings can be utilized to find appropriate transforms. This is especially important in the construction of double bonds with control of stereochemistry. Further, the inclusion of aromatic and heterocyclic rings will have to be undertaken. The great magnitude of this task is obvious, as is the fact that its accomplishment will require the collaboration of many groups of chemists, each concentrating in its field of expertise, to generate rational and sophisticated chemical flow charts.[34] Such undertakings are not less challenging and rewarding than the conception and execution of a specific synthesis.

Many applications of the binary search multistep look-ahead technique can be foreseen in areas other than antithetic ring disconnection. Binary search methods can lead to the dissemblage of collections or ensembles of *functional groups* to produce simple, "core" functional groups. As an example, consider the following antithetic sequence which can easily be generated by a binary search directed toward the goal of producing the "core" amino function, $\rightarrow CNH_2$. There are many problems in which such functional group simplification represents an important key to effective synthetic analysis.



Other areas for effective use of binary search methods include: (1) removal of chiral centers by application of sequences of transforms which include those corresponding to stereoselective or neighboring group reactions, (2) generation of easily constructed rings by *antithetic connections* involving appendages or side chains, (3) modification of ring size by ring contraction or expansion processes,[35] and (4) multistep look-ahead to effect "strategic" bond disconnections.[1,36] Such binary search operations can even be "nested" together in a hierarchical arrangement. The search for removal of chiral centers could be included as a "chemical subroutine" in certain of the ring transform search tables, for instance, since the strategy for antithetic removal of chiral centers from rings prior to the application of ring transforms is frequently highly successful.

The combination of the techniques now used in LHASA and a straightforward enlargement of the data base can lead to a level of performance in synthetic analysis which matches even the most optimistic views of just a few years ago.

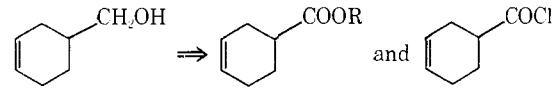
Acknowledgments. We are grateful to the National Institutes of Health for financial assistance. We are also indebted to other members of the LHASA project for helpful discussions and suggestions, especially H. W. Orf, R. D. Cramer, III, K. C. Chu, and W. L. Jorgensen.

Supplementary Material Available. Appendix 2, a listing of computer language, will appear following these pages in the microfilm edition of this volume of the journal. Photocopies of the supplementary material from this paper only or microfiche (105 × 148 mm, 24× reduction, negatives) containing all of the supplementary material for the papers in this issue may be obtained from the Journals Department, American Chemical Society, 1155 16th St., N.W., Washington, D.C. 20036. Remit check or money order for $3.00 for photocopy or $2.00 for microfiche, referring to code number JACS-74-7724.

References and Notes

(1) E. J. Corey, *Quart. Rev., Chem. Soc.*, **25**, 455 (1971) (most recent review).
(2) E. J. Corey, R. D. Cramer, III, and W. J. Howe, *J. Amer. Chem. Soc.*, **94**, 440 (1972).
(3) Two recent doctoral dissertations may be consulted for details and supplementary information not included herein: (a) W. J. Howe, Ph.D. Thesis, Harvard University, 1972, and (b) D. A. Pensak, Ph.D. Thesis, Harvard University, 1973.
(4) The intermediate structures can be regarded as "subgoals" on the way to the "goal" of applying the simplifying transforms.

(5) Consciously or subconsciously!
(6) E. J. Corey and B. B. Snider, *J. Amer. Chem. Soc.*, **94**, 2549 (1972).
(7) Such stereochemical information is, of course, of general use throughout the program.
(8) E. J. Corey, W. T. Wipke, R. D. Cramer, III, and W. J. Howe, *J. Amer. Chem. Soc.*, **94**, 421 (1972).
(9) LHASA's facility for graphical input of structural information has been previously described in detail (ref 8). It is based on the use of a stylus/tablet device in conjunction with cathode ray tubes (CRT's). The chemist "draws" the target structure with the stylus on the horizontal surface of the tablet and the structure appears on the CRT. Also appearing on the displays are a collection of graphical "buttons" which may be activated with the stylus to perform a variety of structural modifications. Although the display formats of the current versions of LHASA-1 (which is implemented on a PDP-1 computer) and LHASA-10 (implemented on a POP-10 computer) differ from the displays shown in ref 8, the basic mode of interaction remains unchanged.
(10) Techniques for the machine perception of structural features such as rings, functionality, and a variety of other synthetically significant data have already been discussed: E. J. Corey, W. T. Wipke, R. D. Cramer, III, and W. J. Howe, *J. Amer. Chem. Soc.*, **94**, 431 (1972).
(11) See also (a) D. C. Garwood and D. J. Cram, *J. Amer. Chem. Soc.*, **92**, 4575 (1970); (b) A. E. Petrarca, M. F. Lynch, and J. E. Rush, *J. Chem. Doc.*, **7**, 154 (1967), for other applications of such linear representations.
(12) This is equivalent to "rotating" the linear representations. Thus, 5-1-3 becomes 3-5-1 which becomes 1-3-5.
(13) Using a simple extension of this approach devised by another member of the LHASA research group, W. L. Jorgensen, cis and trans relationships are correctly perceived by the program even when only one substituent has been explicitly defined at each end of the double bond (*i.e.*, when hydrogen substitution is assumed).
(14) "IUPAC Tentative Rules for the Nomenclature of Organic Chemistry. Section E. Fundamental Stereochemistry," *J. Org. Chem.*, **35**, 2849 (1970).
(15) E. L. Eliel, "Stereochemistry of Carbon Compounds," McGraw-Hill, New York, N.Y., 1962, p 234.
(16) In the retrosynthetic direction.
(17) R. Grewe and I. Hinrichs, *Chem. Ber.*, **97**, 443 (1964).
(18) (a) R. McCrindle, K. H. Overton, and R. A. Raphael, *J. Chem. Soc.*, 1560 (1960); (b) E. E. Smissman, J. T. Suh, M. Oxman, and R. Daniels, *J. Amer. Chem. Soc.*, **81**, 2909 (1959).
(19) Although in principle it is possible that the analysis of one ring will give rise to sequences leading to the application of six different Diels–Alder disconnections, it is also possible that the analysis of some or all of the orientations will be blocked. Frequently, two or three complete sequences are generated for each ring using the current versions of the Diels–Alder search charts.
(20) An FGI subgoal is invoked to deoxygenate that epoxide to form the double bond of a structure which would result from the Diels–Alder addition of an *acetylenic* dienophile to a diene.
(21) One of the functions of the search pattern is to create in the retrosynthetic direction functionality which "activates" the dienophile portion of the ring (atoms 5 and 6), so that the Diels–Alder disconnection produces a dienophile which could be expected to react with the diene in the synthetic direction. The dienophile is generally activated by the insertion of one or more electron withdrawing groups, although in some situations this is not necessary (see Chart IV). The "flag" mentioned above is used to indicate that the activation step may be bypassed.
(22) If the current ring is trans fused to another cyclohexane ring, for example, the stereochemistry at one of the common atoms must be inverted to produce a cis fused system. Otherwise, the Diels–Alder disconnection would generate as the dienophile a trans cyclohexene. With some larger ring systems the trans double bond may be accommodated in the ring so the inversion step is unnecessary.
(23) A very compact method for evaluation of regioselectivity in the Diels–Alder addition has been developed by Dr. D. A. Pensak as a part of this project.
(24) (a) E. J. Corey, D. A. Pensak, and W. J. Howe, to be submitted for publication and ref 3a. (b) The terms utilized in the table language are included in the Appendix which appears as supplementary material in the microfilm version of this paper. See paragraph at end of paper regarding microfilm material.
(25) "Condition statements" prove to be very useful in the evaluation of relative functional group reactivities. See E. J. Corey, D. A. Pensak, and H. W. Orf, to be submitted for publication.
(26) This also applies to the data table entries for transforms in the other four transform categories.
(27) Current (early 1974) versions of LHASA require manual specification of the desired ring transform package. However, as LHASA is expanded to include additional ring transform search patterns, provision for automatic selection and application of the most appropriate ring chemistry module(s) will be made.
(28) *I.e.*, if the rating which results from the evaluation process lies *above* a set threshold value.
(29) Frequently, it is possible to provide multiple solutions to a single subgoal. An electron-withdrawing group can be produced from the hydroxyl group in A below, for example, by conversion of the hydroxyl to either a carboxylic ester or an acid chloride group. Both of these solutions are



A

presented to the chemist, and a branch point appears on the synthesis tree. The analysis continues with the first generated of these intermediates until the Diels–Alder disconnection is finally applied. LHASA then "backtracks" to the most recent branch point in the tree and resumes the analysis at the point in the table where the branch was originally formed. Thus, complete sequences are generated below each of the branches. To simplify the following discussion, however, it will be assumed that only one solution is allowed for each subgoal request.
(30) The "current target" is the tree node upon which the Diels–Alder search pattern was invoked by the user, which is not necessarily the overall target of the analysis session. The chemist may apply any of LHASA's analysis options to any of the intermediates in the synthesis tree.
(31) Typically, 2 or 3 sec elapse between successive subgoal generation calls while the Diels–Alder table is being processed further by the interpreter. This is not noticed by the user, who on the average will spend 5–10 sec examining each intermediate.
(32) The fact that all the sequences pictured in Figure 4 are not generated directly from A is a consequence of limitations in the search pattern with regard to subgoal generation. Having recognized such limitations by the study of examples such as that of target structure A, it becomes obvious and desirable to expand the tables to permit generation of the entire sequence directly and automatically from A.
(33) The most obvious use of this depth limit, or "merit value," would be to terminate the processing of any branch or pathway which has still not arrived at a structure to which the ring disconnection can be directly applied when the limiting search depth has been reached. A much more subtle and perhaps more interesting use of merit values would involve "skipping over" blockages in the subgoal search in situations where a high merit value had been assigned. For example, if a high merit value had been assigned by the program to the eventual opening of a certain ring by a Diels–Alder process, then even if one of the requested subgoal steps could *not* be performed (according to information available to the program), it would be worthwhile to display the *expected* offspring and continue the analysis on that structure. After completion of the sequence the blocked step would be displayed for the chemist's evaluation.
(34) It is the experience of the authors that the design of a chemical flow chart can very effectively be carried out by a small group (two–six) of individuals working part-time, both individually and as a group, over a period of several weeks. Among the benefits which accrue to team members from such an effort are: (1) considerable enhancement in an individual's general problem-solving ability, (2) the generation of interesting and significant new ideas for research, (3) greatly improved pedagogic skills in areas of chemical synthesis.
(35) The antithetic modification of five- or seven-membered rings to six-membered rings, for example, can reveal simple, valid, and even superior synthetic routes. Similarly four-, eight- and higher-membered ring structures are often more readily available by ring expansion or contraction than by direct ring formation.
(36) This strategy has recently been implemented in LHASA and involves the use of multistep sequences of FGI and FGA subgoals which lead to the application of one or more of a very small subset of powerful one- and two-group transforms to disconnect strategically located bonds.