with the sulfenyl chloride in the absence of methanol. The epimers (at C-6) of **11** and **12** were not observed. However, if formed it is unlikely they would survive the chromatography: see ref 3 and 9.

(11) The 7-chloro esters **17a** and **17b** were made by treatment of the diazo ester **16** with the sulfenyl chloride in the absence of methanol. After extensive chromatography on silica gel, which decomposed most of the material put on the column (see ref 9), a small amount of each epimer was isolated (see Experimental Section).

(12) The diazo ester **16**, made by the method of Wiering and Wynberg (ref 17), is a crystalline compound; however, again (see ref 8) better overall yields of **19** are obtained if **16** is not isolated. The lower yields in this series compared to the previous two in the penicillin series are attributed mainly to the low yield in the preparation of the diazo ester **16**.

(13) The benzhydryl group has not been used to protect the carboxyl in penicillins because the mild acidic conditions required to remove it destroy the penam system. In the absence of an acylamino group at the C-6 position, which readily reacts with the β-lactam ring to form azlactones, penicillins are more acid stable (see ref 2, p 258). It was therefore anticipated that in the case of the α-methoxy thiol penicillanates, the benzhydryl group could be removed under mild acidic conditions. The carboxyl of **12** was smoothly deprotected with trifluoroacetic acid at 0 °C. With **14** and **15**, these same conditions gave complete destruction of the penam system. However, satisfactory results were obtained when TFA was used with methylene chloride as a solvent (see Experimental Section).

(14) J. C. Sheehan, Y. S. Lo, J. Loliger, and C. Podewell, *J. Org. Chem.*, **39**, 1444 (1974).
(15) German Patent 1 224 720; *Chem. Abstr.*, **65**, 12112h (1966).
(16) G. Zumach and E. Kühle, *Angew. Chem., Int. Ed. Engl.*, **9**, 54 (1970).
(17) J. S. Wiering and H. Wynberg, *J. Org. Chem.*, **41**, 1574 (1976).
(18) U.S. Patent 3 129 224; *Chem. Abstr.*, **61**, 5659e (1964).
(19) E. H. W. Bohme, H. E. Applegate, J. B. Ensing, P. T. Funke, M. S. Puar, and J. E. Dolfini, *J. Org. Chem.*, **38**, 230 (1973).

Computer-Assisted Synthetic Analysis. Performance of Long-Range Strategies for Stereoselective Olefin Synthesis

E. J. Corey* and Alan K. Long

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

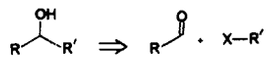
Received December 15, 1977

The Harvard program for computer-assisted synthetic analysis (LHASA) has been expanded to include a module which directs antithetic simplifications of olefinic target molecules. The new module employs a readily modifiable data base of C=C transforms (retroreactions) written in chemical English (CHMTRN). Transforms are selected via a unique set-matching process based on prescreen information extracted both from the target molecule and from the transform entries in the data table. Each transform has access to a considerable amount of subgoal power and is thus capable of generating quite long and sophisticated sequences. Several strategies, corresponding to effective plans for polyene synthesis, have been implemented. A number of sample antithetic analyses are included, and future extensions are discussed.

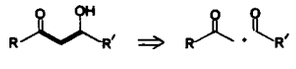
Synthetic methodology for the stereospecific and highly stereoselective construction of carbon–carbon double bonds has expanded dramatically in the last 15 years. Challenges presented by biogenetically interesting isoprenoid molecules, such as squalene[1] and farnesol,[2] and in particular by the insect juvenile hormones,[3] have stimulated development of a large number of versatile techniques for olefin synthesis.[4] To keep pace with these new methods, a special module for olefin synthesis has recently been added to the LHASA[5] computer program. The new package combines stereochemical sophistication[6] with a broad data base of chemical reactions, employing a variety of "strategies" to construct efficient and often elegant routes to polyolefinic molecules.

As previously described,[7] LHASA is an interactive program for synthetic analysis which employs straightforward graphical input and output. The program analyzes an input "target" molecule antithetically, generating a "tree" of potential synthetic precursors. Individual steps in the antithetic analysis correspond to "transforms" (retroreactions) which are chosen, or "keyed," by certain arrangements of functional groups and structural features in the target molecule.

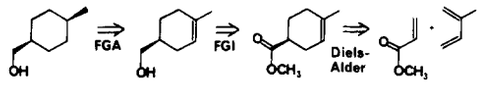
Early work on LHASA divided transforms into two categories, group oriented[8] and substructure oriented.[9] In the former category, an opportunistic, or breadth first, search through the data base selects transforms purely on the basis of arrangements of functionality. A Grignard transform, for instance, is keyed by the presence of a hydroxyl group:



and an Aldol condensation by (among other combinations) a carbonyl group and a hydroxyl separated by a "path" of two bonds:



In the latter category, certain powerful transforms generate antithetic pathways in a depth-first fashion. The existence of an appropriate substructure (for instance a ring of a certain size) is sufficient to key entry into the transform, and the existing functionality is modified as necessary for transform performance:



In this last example, the functionality in the target molecule was not correct for performance of the Diels–Alder disconnection. Accordingly, two nonsimplifying "subgoal" steps, a Functional Group Addition (FGA) of a C=C and a Functional Group Interchange (FGI) of the hydroxyl for the ester group, were performed by the program before the Diels–Alder transform. These steps, like the goal transform, were thoroughly evaluated by the program before display to ensure that they correspond to reasonable synthetic reactions. The subgoal powers of the LHASA program have recently been expanded to include sequential functional group interchange (SEQFGI),[10] double parallel functional group interchange (FGIFGI), and parallel functional group interchange–functional group addition (FGIFGA).

The new package for olefin syntheses combines features of both the group-oriented and substructure-oriented approaches, as described below. Considerable planning preceeded implementation of the module, with six important concepts guiding its development.

First, the data base for the package needed to reflect both the great diversity of new olefin syntheses and the stereochemical specificity of many of these new methods. The efficiency of a simple, functionality based search through all the transforms keyed by the presence of a C=C decreases dramatically with the addition of large numbers of new trans-

0022-3263/78/1943-2208$01.00/0

forms. Accordingly, a fast and economical method for accessing the data base had to be devised.

Almost any recent olefin synthesis illustrates the degree to which functionality around a C=C can be modified after establishment of the C=C stereochemistry. A second requirement for an intelligent antithetic analysis, then, was the power to access several nonsimplifying, or subgoal, steps to transform the target structure to a precursor containing the functionality resulting from the C=C-forming reaction. In this sense each olefin transform would be like a substructure-oriented transform in the traditional group/substructure dichotomy. The C=C, like the six-membered ring in the Diels–Alder example above, would be a "key substructure" signaling entry into a depth-first search leading to the performance of the goal step, in this case the stereospecific C=C transform.

For syntheses of olefins containing more than one C=C, the number of precursor structures generated by a program capable of both broad and deep searches would often be unmanageable. Implementation of heuristically derived "strategies" would allow the program not only to choose individual transforms but actually to plan in advance the order in which to apply them. Strategies such as sequential application of the same transform, simultaneous disconnections at both ends of a chain, and disconnection of a central C=C (corresponding to a convergent synthesis) would limit the number of precursor structures and greatly enhance the sophistication of the resulting routes.

In an antithetic analysis using one of the strategies just mentioned, the "goal" of the analysis (e.g., two sequential applications of the Claisen rearrangement transform) is on a higher level than the goal of applying a single powerful transform. Accordingly, an extension of the normal subgoal structure would be necessary. Group-oriented transforms capable of disconnecting certain "strategic" bonds[11] might be used as subgoal steps, and in fact olefin transforms themselves could be used as subgoals to remove, antithetically, parts of a molecule blocking performance of the transforms required by the higher level strategy.

In order to decide which strategies would be most appropriate for a given target polyene, the program would have to make a C=C/transform match before any transforms were performed and displayed. Not only was a method for "prescreening" C=C's necessary, but the method had to be flexible enough to accommodate the addition of new transforms as well.

Finally, with a continuing commitment to the interactive nature of the LHASA program, the olefin module would allow the chemist to select for further processing precursor structures which appeared especially promising. This type of participation by the chemist would limit structure proliferation while obviating the need for overly restrictive tree-pruning heuristics in the program.

Implementation

Transform Selection.[12] The olefin package focusses on acyclic C=C's possessing *E* or *Z* character. Bonds whose synthesis would not require stereochemical sophistication are used only to key subgoal (FGI) transforms. Each acyclic *E* or *Z* C=C undergoes a special "perception" process prior to transform selection. First, a trisubstituted C=C is labeled as shown in Figure 1, with L, C, and T denoting respectively the "lone," "cis," and "trans" appendages on the olefinic atoms. For tetrasubstituted C=C's a fourth appendage label (X) is added, and for disubstituted bonds either the C or the T appendage disappears. Next, the labeled bond is scanned for structural features, both simple and complex, which are relevant to the transform selection process. Categories of molecular characteristics included correspond to those features which might potentially kill a C=C transform. Simple characteristics, like ring bonds and quaternary centers, are matched with the appendage atom and bond labels (see Figure 1) by FORTRAN code within the olefin package. More specialized features, such as the availability of various appendages from organocuprate reagents, are perceived by use of a binary search table.[9] A sample perception is shown in Figure 1. All of the perception information about a C=C is stored in a **perception set.** Sets are among the most frequently used data structures in LHASA. A set consists of a computer word (or words) in which each bit (valued 0 or 1) is assigned a unique meaning. In the perception set for the C=C in Figure 1, for example, bit 15 is on (= 1) signifying that there is a functional group on L3, bit 41 = 1 indicates a tertiary or quaternary center on C2, etc. One-hundred bits are sufficient to contain all the perception information necessary for prescreening of an arbitrary C=C.

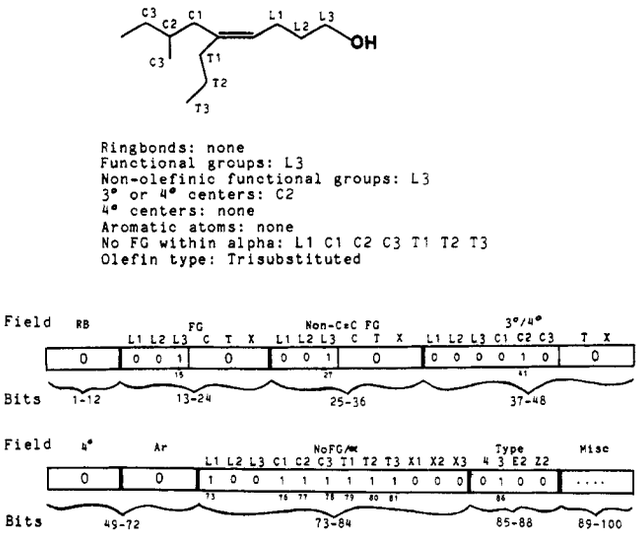


Figure 1. Sample perception set. Field abbreviations are as follows: RB = bonds which are ring bonds, FG = atoms bearing functional groups, Non-C=C FG = atoms bearing functional groups which are not C=C or C≡C, 3°/4° = atoms bonded to three or four other carbons, 4° = atoms bonded to four other carbons, Ar = atoms in aromatic rings, No FG/α = atoms not bearing a functional group whose adjacent atoms do not bear a functional group, Type = C=C substitution type (4 = tetrasubstituted, 3 = trisubstituted, E2 = E disubstituted, Z2 = Z disubstituted), Misc = miscellaneous characteristics (omitted for simplicity). Atom labels X1, X2, and X3 refer to the fourth appendage on a tetrasubstituted C=C.

In order to use the perception set characteristics to select transforms, the program must know which of these characteristics would prohibit application of a given transform. Therefore, each transform has associated with it a **screening set,** again 100 bits long, with bit positions assigned exactly as in the perception sets. An "on" bit (= 1) in the screening set for a transform corresponds to a molecular characteristic which should kill that transform. For easy modification by the chemist, the transform information necessary for creation of the screening sets is included at the top of a transform entry in the chemical English (CHMTRN)[13a] version of the data table under the comment heading "... Kill Specifiers for Prescreen," as shown in the sample transform entry in Figure 2. The CHMTRN assembler, TBLTRN,[13b] constructs the screening sets from the kill specifiers before LHASA is actually run.

After C=C perception is complete, a logical comparison (ANDing) operation is performed between the perception and screening sets, resulting in a set which is the intersection of the two input sets. Only when two corresponding bits are on in the input sets will a "1" appear in that position in the resultant set. Thus, if the resultant set in a comparison is nonzero, a feature was found which would block application of the

```
Transform 410
Name Claisen Rearrangement
...TL 3243(1969), JACS v.92 741,4461,4463(1970),v.95 553(1973)
Rating 30
...Kill Specifiers for Prescreen
Ringbond Trans2, Functional*group Trans1
No*FG*within*alpha Trans3
Olefin*type Tetrasubstituted Z*Disubstituted
Spacing*1*5 Well*defined*T*first Well*defined*L*last
...
If there is a functional group on alpha to atom*3  &
        offpath then go to Block4
If there is a functional group on beta to atom*3  &
        offpath then go to Block3
If there is a functional group on gamma to atom*3  &
        offpath then go to Block4
Block3 Designate the group as group one
    If hetero1*1 is nitrogen then go to Block6
    If the first group is carbonyl or: ester then go to Block2
    Go to Block4
Block6 If the first group is amide*3 then go to Block2
    Exchange the group for an amide*3 and*then go to Block5
Block4 Exchange the group for an ester on beta to atom*3 offpath
Block5 If unsuccessful then reject
Block2 Designate the group on beta to atom*3 offpath as group one
    If alpha to carbon1*1 is the same as alpha to atom*3
    Save as 1 the previous locant
    Kill if there is a hetero atom alpha to saved*atom 1
    Subtract 10 if atom*3 is a quaternary*center
    ....
            If there is a multiple bond in the first group  &
                    then break it
            If first group is carbonyl then join saved*atom 1  &
                    and carbon1*1
            If first group is carboxyl then attach an ether  &
                    to carbon1*1
            Attach an ether to carbon1*1
            Separate saved*atom 1 and atom*3
            Double bond*3
            Single bond*5
            Attach a hydroxyl to atom*5
```

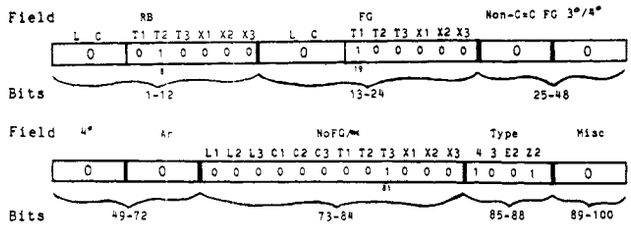
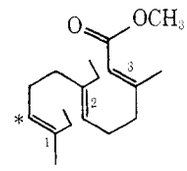


Figure 2. Sample transform entry with screening set for Claisen rearrangement transform. See Figure 1 for explanations of field abbreviations and appendage atom and bond labels.

transform concerned, while a zero resultant set indicates that the transform is appropriate for the C=C being perceived. The set that results from ANDing the perception set in Figure 1 with the screening set in Figure 2 is nonzero, since bit 81 is on in both sets. Thus, the program would not attempt the Claisen rearrangement transform for the target in Figure 1. In this fashion, a complete list of transforms applicable to each C=C is compiled *before* the actual transform entries are evaluated. This list not only serves as an extremely efficient means of transform selection but also provides direct input to the strategy selection process.

Strategy Selection. As mentioned above, the olefin package selects strategies by using heuristics which correspond to the most effective problem-solving techniques employed by a synthetic chemist. Strategy for olefinic target structures with only one C=C is simple. All the appropriate transforms (preceeded by subgoal steps, if necessary) are applied systematically and exhaustively (i.e., opportunistically). For polyenes, a hierarchy of three strategies is available. That which seems most powerful is attempted first, and if it fails, the next, etc. A given polyene strategy operates on a "chain" of C=C's. Strings of consecutive C=C's are identified and ordered according to length in a "chain perception" process which precedes strategy execution.

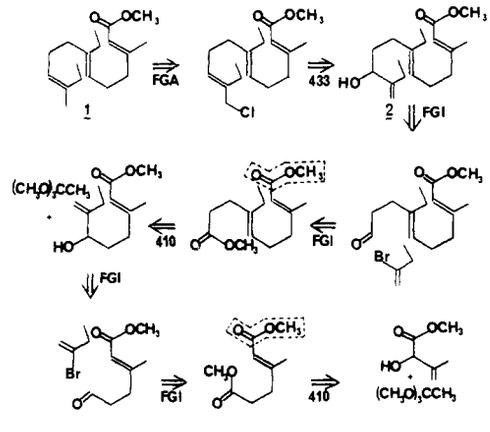
Sequential application of the same reaction is frequently an effective approach to polyene synthesis.[14] The antithetic analogue of this approach is the "sequential disconnection" strategy. This strategy is reserved for dienes and trienes, since a convergent approach is almost always more efficient for higher polyenes. Only certain transforms are given sequential disconnection power. Each such transform has a section in its CHMTRN data table entry giving information on the connectivity and spacing necessary for the sequential disconnection. For example, the specifiers "Spacing*1*5 Well*defined*T-*first Well*defined*L*last" in the Claisen rearrangement transform (see Figure 2) require that the two (or three) C=C's to be sequentially disconnected be in a 1,5 relationship, with the L appendage of one bond connected to the T appendage of the other, and that the bond with the "free" T appendage be disconnected first. (In contrast to the sequential Claisen rearrangement, which requires the well-defined 1,5-diene spacing, certain other transforms are labeled "undefined" with regard to spacing.) Analysis of Cecropia juvenile hormone triene ester **1** provides several antithetic sequential disconnection routes.



1

One of these routes involves disconnection by sequential application of the Claisen rearrangement transform. (While other lines of analysis are also generated by the program, they will not be discussed here in detail.) Perception of the chain containing C=C's 1, 2, and 3 indicates that bond 1 lacks a three-carbon T appendage but that bonds 2 and 3 form an appropriate "subchain" for disconnection by sequential application of the Claisen rearrangement transform. As indicated by the "Well*defined*T*first" specifier in the transform entry, the sequential Claisen rearrangement must *first* operate on the "free" T appendage of the 2,3 subchain. A problem is encountered, however, when the program attempts to exchange (via the CHMTRN command "Exchange the group for an ester on beta to atom*3 offpath") (see Figure 2) the C=C on atom T3 of bond 2 (the starred atom) for an ester, since the simple FGI chemistry does not include any stereospecific olefin syntheses. Bond 1 is thus recognized as a block to the higher-level goal of performing the sequential disconnection of bonds 2 and 3. To remove this block the program disconnects bond 1 opportunistically, using the appropriate olefin transforms, thereby paving the way for the sequential disconnection. Several structures are generated in the opportunistic disconnection of bond 1. At this juncture the program

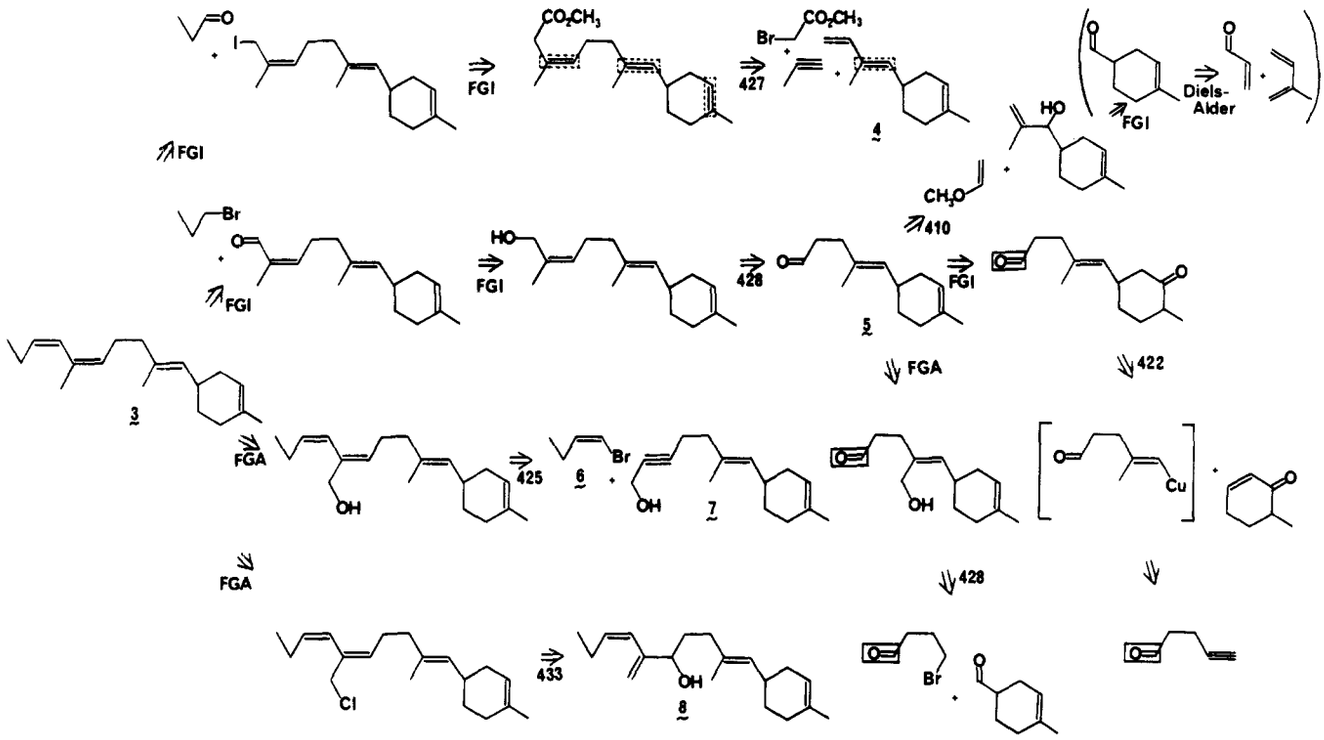
Scheme I.[17] Sequential Application of the Claisen Rearrangement Transform[a]



Sequential Disconnection of Juvenile Hormone Triene Ester

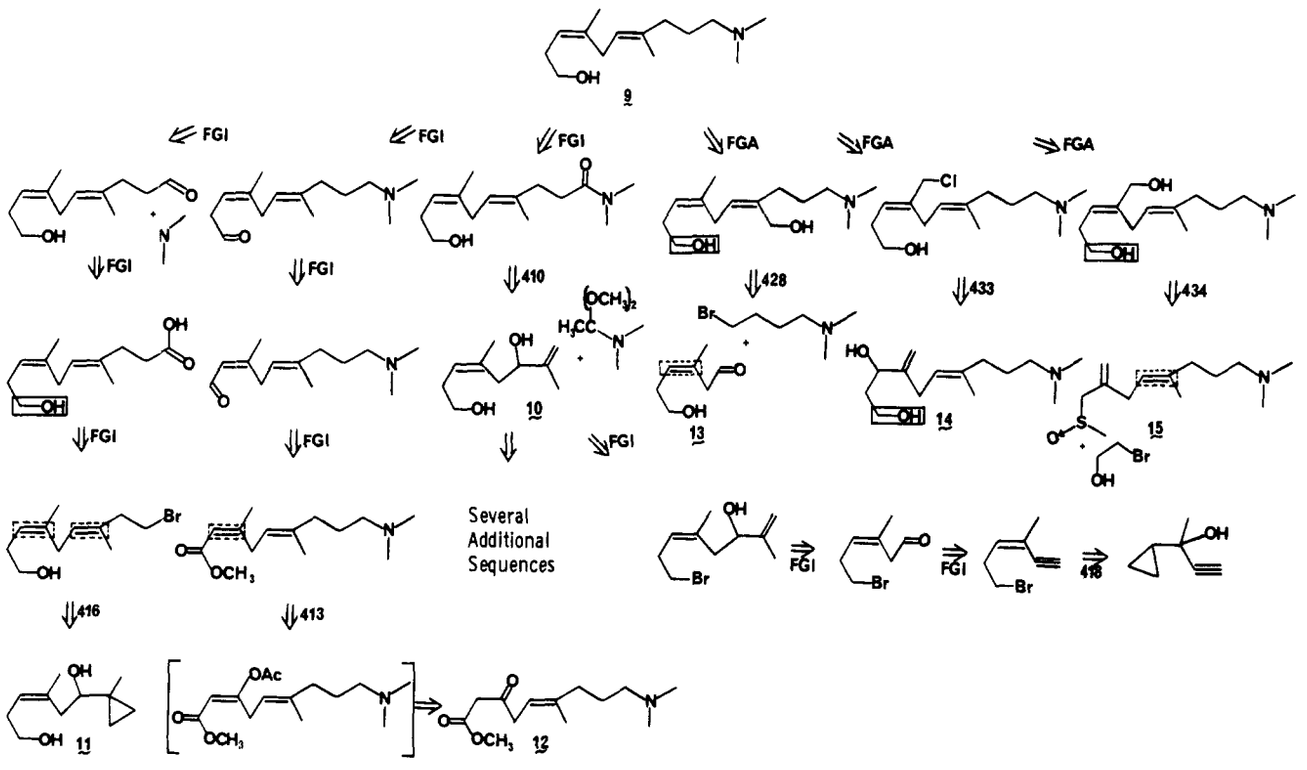
[a] Transform code numbers (see Table I) are as follows: 410 = Claisen rearrangement, 433 = allylic rearrangement with $SOCl_2$.

Scheme II.[17] Convergent Strategy[a]



[a] Transform code numbers (see Table I) are as follows: 410 = Claisen rearrangement, 422 = conjugate addition by vinyl copper reagent, 425 = R_2AlH reduction of propargylic alkoxide, 427 = alkylation of acetylenic borate, 428 = double Wittig with formaldehyde, 433 = allylic rearrangement with $SOCl_2$.

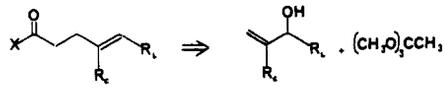
Scheme III.[17] Linear Strategy[a]



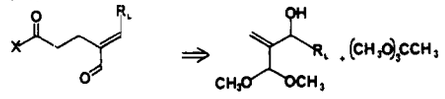
[a] Note that for the $RCH_2Br \rightarrow RCH_2CH_2NMe_2$ conversion leading to structure 11, LHASA finds six chemically reasonable sequences. To avoid node proliferation, only one of these is displayed. Transform code numbers (see Table I) are as follows: 410 = Claisen rearrangement, 413 = 1,4 addition–elimination, 416 = Julia synthesis, 418 = Julia synthesis on tertiary carbinol, 428 = double Wittig with formaldehyde, 433 = allylic rearrangement with $SOCl_2$, 434 = allylic sulfoxide rearrangement.

Table I. The Following Transforms Form the Data Base for the Olefin Package in LHASA

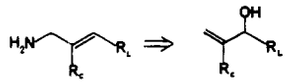
TRANSFORM 410
CLAISEN REARRANGEMENT
TL 3243(1969), JACS 92 741,4461,4463(1970); 95 553(1973)



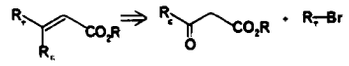
TRANSFORM 411
ACETAL CLAISEN REARRANGEMENT
TL 3469(75)



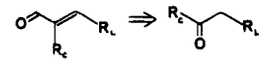
TRANSFORM 412
TRICHLOROACETIMIDATE REARRANGEMENT
JACS 96 597(74), 98 2901(76)



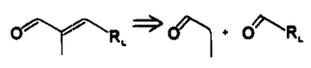
TRANSFORM 413
1,4 ADDITION-ELIMINATION
CHEM LETT 1097(1973), TL 925(1974)



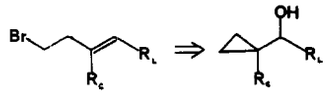
TRANSFORM 414
ALPHA-CHLORO ALDEHYDE ELIMINATION
TL 2465(73)



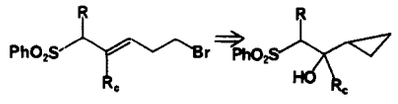
TRANSFORM 415
DIRECTED ALDOL CONDENSATION
NEUERE METHODEN DER PRAP. ORG CHEM, BAND VI, P.42 (70)
ACIE 7 7(68)



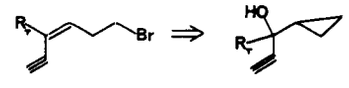
TRANSFORM 416
JULIA SYNTHESIS
JACS 90 2882(1968), TL 111(1973)



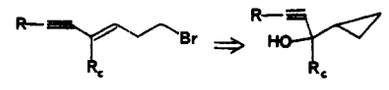
TRANSFORM 417
JULIA SYNTHESIS ON TERTIARY CARBINOL
TL 3445 (1974)



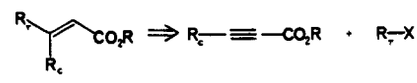
TRANSFORM 418
JULIA SYNTHESIS ON TERTIARY CARBINOL
TET. 28 3739(1972)



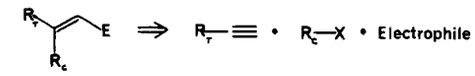
TRANSFORM 419
JULIA SYNTHESIS ON TERTIARY CARBINOL
TL 745(76)



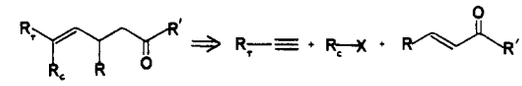
TRANSFORM 420
ORGANOCUPRATE ADDITION TO PROPARGYLIC ESTER
JACS 94 4395(1972), TL 1277,1281(1973); JOC 38 2733(1973)
JACS 97 1197(75), JOC 41 3629(76)



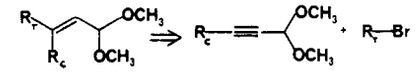
TRANSFORM 421
ELECTROPHILIC ATTACK BY VINYL COPPER COMPLEX
SYN 245(76), TET 32 1675(76), TL 2023(77)
J ORGMET CHEM 40 C49(72), 77 269,281(74)



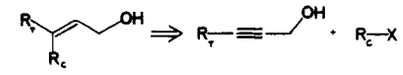
TRANSFORM 422
CONJUGATE ADDITION BY VINYL COPPER REAGENT
JACS 99 253 (77)



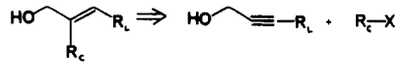
TRANSFORM 423
ORGANOCOPPER ADDITION TO PROPARGYLIC ACETAL
TL 2313 (76)



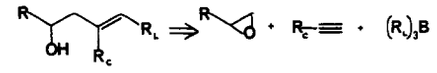
TRANSFORM 424
LAH REDUCTION OF PROPARGYLIC ALCOHOL
JACS 89 4245(1967), JOC 38 2733(1973), TL 1983(1973)



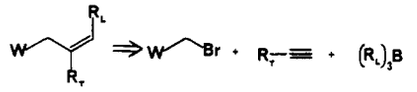
TRANSFORM 425
R2ALH REDUCTION OF PROPARGYLIC ALKOXIDE
JACS 92 6314(1970)



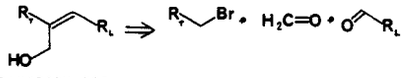
TRANSFORM 426
EPOXIDE OPENING BY ALKYNYL BORATE
TL 2741(1973), TET 30 3037(74), BUT CHEM LETT 397(75)



TRANSFORM 427
ALKYLATION OF ACETYLENIC BORATE
TL 795,4491(73); 3327(75)



TRANSFORM 428
DOUBLE WITTIG WITH FORMALDEHYDE
JACS 92 226,6635-7(1970); TL 3231(77)



TRANSFORM 429
DOUBLE WITTIG WITH TWO ALDEHYDES
JACS 92 226,6635-7(1970); TL 3231(77)

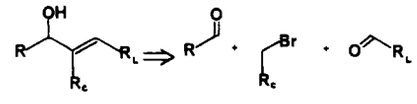
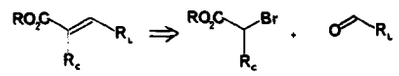
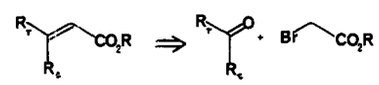


Table I (*continued*)

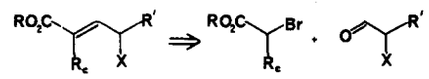
TRANSFORM 430
EMMONS-WADSWORTH-HORNER WITH ALDEHYDE
BCSJ 40 2968(1967), JACS 83 1733(1961), CHEM REV 87(74)



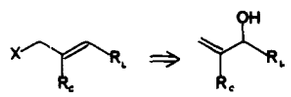
TRANSFORM 431
EMMONS-WADSWORTH-HORNER WITH KETONE
JACS 90 3769(1968), CHEM REV 87(74)



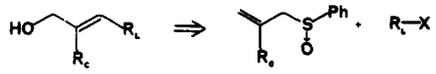
TRANSFORM 432
WITTIG REACTION TO GAMMA-HALO TIGLATE
TL 1679(75), 167(77)



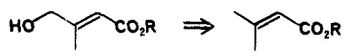
TRANSFORM 433
ALLYLIC REARRANGEMENT WITH SOCL2
BSCF 3568(1969), PNAS 68 1294(1971),
JACS 92 737,4461(1970), 95 7067(1973)



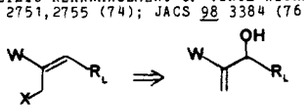
TRANSFORM 434
ALLYLIC SULFOXIDE REARRANGEMENT
CHEM COMM 702 (72), TL 1389(73), ACR 7 147 (74)
JOC 38 2245,2572(73)



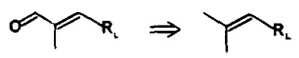
TRANSFORM 435
ALLYLIC SULFOXIDE REARRANGEMENT OF UNSAT. ESTER
TL 4215 (76)



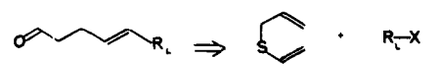
TRANSFORM 436
ALLYLIC REARRANGEMENT OF VINYL-WITHDRAWING GROUP
TL 2751,2755 (74); JACS 98 3384 (76)



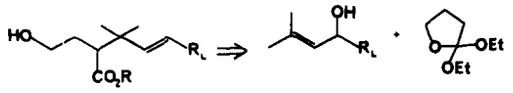
TRANSFORM 437
SELENIUM DIOXIDE OXIDATION
JACS 94 7154(72), 93 4835(71)



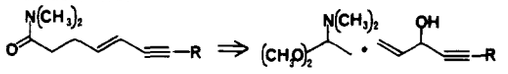
TRANSFORM 438
THIO-CLAISEN REARRANGEMENT
JACS 95 2693 (1973)



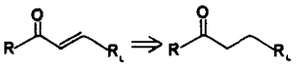
TRANSFORM 439
CYCLIC ORTHOESTER CLAISEN REARRANGEMENT
TL 847 (74)



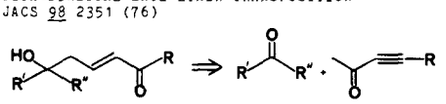
TRANSFORM 440
CLAISEN REARRANGEMENT OF ACETYLENIC ALCOHOL
TL 2607 (76)



TRANSFORM 441
ALPHA-KETO SULFOXIDE ELIMINATION
JACS 98 4887 (76)

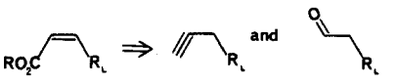


TRANSFORM 442
BETA-DIKETONE ENOL ETHER TRANSPOSITION
JACS 98 2351 (76)

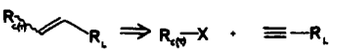


TRANSFORM 443
ALKYLATION OF DIMETHYLTHIO-ALLYLLITHIUM
JACS 93 1724 (1971)

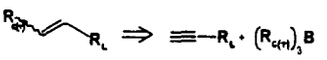
TRANSFORM 444
MODIFIED FAVORSKI REARRANGEMENT
PROC CHEM SOC 148 (64)



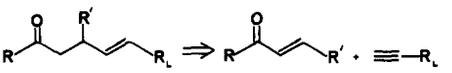
TRANSFORM 445
ACETYLENE REDUCTION
MARCH 361,593; HOUSE 19,91,124,172,205,252; B+P 106,162
ORGANOMET CHEM SYN 1 249(71),CHEM COMM 17,452(76)
TL 1815(72), 1927(75); JACS 96 316(74)
JOC 41 2214,2215,3484(76), 42 579(77)
SYN 525,816(76)



TRANSFORM 446
HYDROBORATION OF TERMINAL ACETYLENE
JACS 95 5786,6456 (1973)

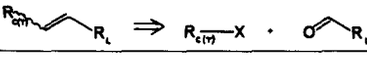


TRANSFORM 447
CONJUGATE ADDITION OF VINYL ALUMINUM COMPLEX
CJC 51 2098 (1973)

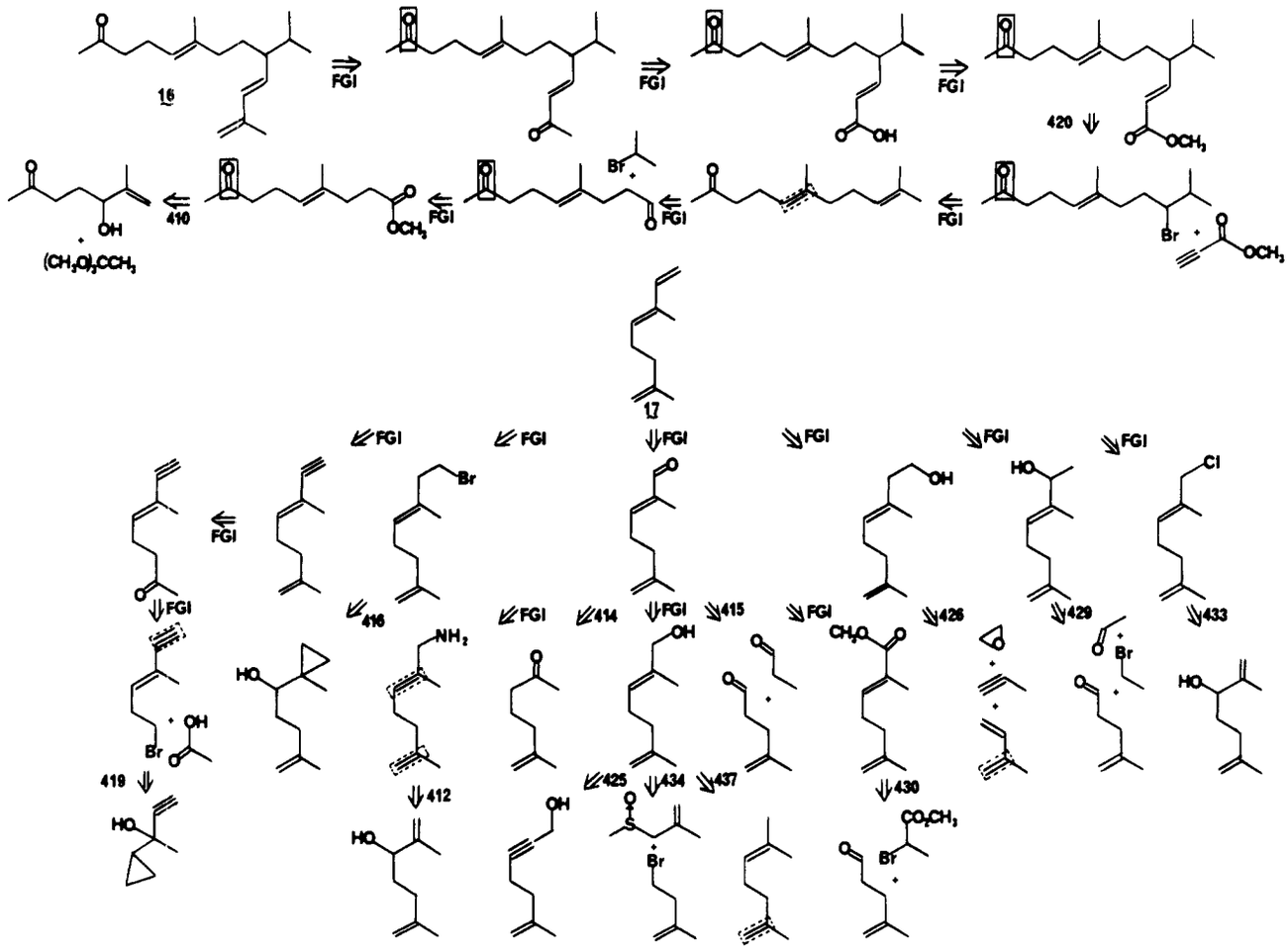


TRANSFORM 448
ALKYLATION OF VINYL CUPRATE
TL 2583 (71), 3461 (76)
J ORGMET CHEM 40 C49 (72), 77 269 (74)

TRANSFORM 449
WITTIG REACTION WITH REACTIVE YLIDE
MARCH 702; HOUSE 682-709; B+P 141
JACS 88 5653(66), 89 2758(67), 91 5675(69)
ACIE 4 689(65), 5 126(65); ANN 708 1(67)



Scheme IV.[17] Additional Examples[a]



[a] Transform code numbers (see Table I) are as follows: 410 = Claisen rearrangement, 412 = trichloroacetimidate rearrangement, 414 = α-chloro aldehyde elimination, 415 = directed aldol condensation, 416 = Julia synthesis, 419 = Julia synthesis on tertiary carbinol, 420 = organocuprate addition to propargylic ester, 425 = R_2AlH reduction of propargylic alkoxide, 426 = epoxide opening by alkynyl borate, 429 = double Wittig with two aldehydes, 430 = Emmons–Wadsworth–Horner with aldehyde, 433 = allylic rearrangement with $SOCl_2$, 434 = allylic sulfoxide rearrangement, 437 = selenium dioxide oxidation.

gives the chemist the opportunity to choose from among these structures (such as **2**) the one(s) he wishes to process further. Intervention by the chemist here allows him to guide the antithetic analysis in the directions which seem most reasonable and thus to limit the proliferation of precursor structures. When **2** is selected, processing continues *within the same strategy* and the antithetic analysis shown in Scheme I results.[15,16]

Convergent syntheses in which a central C=C in a polyene is formed at the end of the route have been used on several occasions.[18] If no sequential disconnections are found for a triene, or if the target has a chain longer than three bonds, the program attempts to disconnect the central C=C(s) (corresponding to a convergent synthesis). In this strategy, a C=C chain is scanned from the center outward in two passes. On the first pass, only disconnections of disubstituted bonds are considered, since the methods for synthesis of disubstituted C=C's are usually more straightforward than those required to make trisubstituted bonds. If no disubstituted C=C disconnections succeed, a second pass is performed in which only trisubstituted C=C transforms are considered. Bonds in the center of a chain are given higher priority than those nearer the ends in order to divide the molecule into fragments of as equal size as possible. Scheme II shows a representative analysis. Processing of **3** in the convergent strategy results in a request for opportunistic disconnection of the central C=C. Four transforms are found in the opportunistic search, and the program asks the user to "Select one or more nodes from among **4, 5, 6, 7,** and **8.**" The antithetic route to **5** seems most reasonable, suffering from no functional group interference problems, using no FGA's, and having the advantage that it disconnects bond 1 as a subgoal. When **5** is selected, three transforms are suggested. It should be emphasized that each fragment from the initial convergent disconnection has access to the entire range of strategies. If structure **5** had been a triene, it would have been considered by the sequential and convergent strategies itself. In this case, **5** had only one *E* or *Z* acyclic C=C, which was disconnected opportunistically.

The third level in the current strategy hierarchy corresponds to a synthesis in which a chain is built up from one end to the other. This "linear disconnection" strategy is used for dienes which cannot be disconnected by sequential application of the same transform and for trienes and higher polyenes for which neither sequential nor convergent disconnection succeeds. Scheme III shows a typical analysis. Again, opportunistic search through the olefin transforms has been used as a subgoal to the higher goal of achieving one or more linear disconnections. First, all transforms for distal bonds (those at either end of a chain) are performed, and the chemist is allowed to choose structures for further processing. When structure **10** is selected, the linear analysis is completed by disconnection of the remaining C=C. The linear strategy, more than either of the others in the hierarchy, leads to a proliferation of precursors. To avoid using simplistic and

overly restrictive heuristics in the program, the chemist is allowed to intervene in two ways. The first, already described, involves the choice of structures for further processing. The second operates as follows. Simple subgoals (FGI, SEQFGI, and FGA) are arranged in a hierarchy with the simplest (FGI) first and the one often requiring the most rigorous conditions (FGA) last. Four passes are made through the olefin transforms appropriate for a particular C=C, first with no subgoals, then with FGI capability, next with FGI and SEQFGI capability, and finally with FGI, SEQFGI, and FGA power. After each of the first three passes, if precursors have been generated in that pass, the chemist is asked if he wants to see deeper subgoals. If he responds "yes," processing continues with the next level of the hierarchy. If he says "no," he is then asked to choose from among the precursor structures which still contain *E* or *Z* acyclic C=C's, and processing continues. In Scheme III, for example, structure **10** results from the FGI level of the hierarchy, and the chemist has the option of seeing disconnections of the second C=C immediately. If more precursors to **9** are desired, the program will generate **11** and **12** at the SEQFGI level of the hierarchy and query the chemist again. If even deeper subgoals are desired, paths leading to **13, 14,** and **15** will be grown. Similarly, in Scheme II, the chemist has the option of selecting structures for further processing before **6, 7,** and **8** (whose routes involve FGA's) are generated.

The simple opportunistic disconnection strategy, normally accessed for monoenes or as a subgoal to one of the three higher-level strategies, has several interactive features of interest. It is possible for the chemist to select for antithetic analysis one or more specific C=C's in a polyene. The user may also choose a specific olefin transform, though this option is used primarily for debugging purposes. In addition, it is possible to override the normal subgoal hierarchy by depressing the "ALL SUBGOALS" button on the menu of processing options.[19] With this option selected, FGI, SEQFGI, and FGA capabilities are all accessed on the first pass through the appropriate transforms.

Additional Examples. The sophistication of any computer program for synthetic problem solving must be judged by the sequences it generates. The sample targets in Scheme IV have both been synthesized in the laboratory.[20] The olefin package in LHASA suggests a route (among others) to phenylsolanone (**16**) which is at least as inventive as the published synthesis. For ocimene (**17**), LHASA finds the published route and a large number of other reasonable pathways as well.

Future Extensions. Several extensions to the olefin package are envisioned for the near future. A strategy which can be highly effective for certain polyenes is one which applies simultaneous disconnections of identically substituted C=C's at opposite ends of a chain.[21] Performance of the goal transform in such a strategy would result in two identical disconnection products and a central fragment. Existing chain perception, transform selection, and identical appendage perception capabilities are sufficiently general to handle most of the problems associated with implementation of this strategy.

Another extension which will make efficient use of existing LHASA modules is the generation of stereoselective routes to epoxides, aziridines, episulfides, and cyclopropyl compounds. Target structures containing these functional group types will be preprocessed in "UNMASKING" mode,[13a,19] an option which accesses transforms capable of removing (antithetically) masked functionality, generating precursors containing only "core" functional groups. These unmasked precursors will then be processed by the olefin package.

Modification of the existing LHASA identical appendage perception modules to recognize near symmetry as well as perfect symmetry will allow convergent disconnections yielding similar, though nonidentical, fragments. In fact, disconnections in such a strategy need not break C=C's. The synthesis of squalene via sulfur-stabilized carbanion chemistry[22] is an excellent example. One possibility for implementation of such a strategy involves trial disconnections between central C=C's followed by identical appendage perception. Appendages would be classified as "potentially identical" on the basis of their carbon skeletons and subgoal chemistry could be requested to rectify differences in functionality.

An important goal in the antithetic simplification of bridged and fused polycyclic target structures is the disconnection of heuristically identified "strategic" bonds.[11] New heuristics could identify bonds blocking the performance of a higher-level strategy, and disconnection of this new type of strategic bond would then be a *subgoal* to that strategy. In addition, standardization of the entire transform-oriented data base (CHMTRN tables) and generalization of the methods for indexing and cross-referencing transforms will allow *all* transforms (not just FGI's and FGA's) to be used either as goals or as subgoals, enormously enhancing the subgoal power of the program.

An extension to the olefin package which is already being implemented involves control of C=C stereochemistry by (antithetic) reconnections to yield rings. Such reconnective chemistry exists in LHASA[23] but is not yet interfaced with the olefin package. The new, expanded reconnective package will interface efficiently with a ring executive capable of choosing among a variety of ring-synthesis strategies.

After several new strategies have been added, considerable effort will be devoted to the problem of selecting among these strategies intelligently. A large number of factors must be taken into account, among them chain length (number of C=C's), adjacent C=C separation (e.g., conjugated; 1,4; 1,5; etc.), C=C connectivity pattern (e.g., L appendage to adjacent T appendage), availability of transforms, and arrangement of nonolefinic functionality. Functional group interference must be accurately assessed along with the potential for both internal and external protection of interfering groups, and an optimal ordering of steps must be chosen.

It is gratifying to see, however, that despite the breadth of areas for future research, a recent review of pheromone synthesis[24] posed no problems which the current olefin package in LHASA did not handle efficiently.

Acknowledgments. We are grateful to the National Institutes of Health for financial assistance. The contributions of the other members of the LHASA group, especially Donald Barth, Harry Orf, and John Vinson, are greatly appreciated.

Appendix

For the transforms that form the data base for the olefin package in LHASA see Table I.

Supplementary Material Available: A complete listing of the CHMTRN version of the transforms (see Appendix) comprising the data table for the olefin package (30 pages). Ordering information is given on any current masthead page.

References and Notes

(1) W. S. Johnson, L. Werthemann, W. R. Bartlett, T. J. Brocksom, T. Li, D. J. Faulkner, and M. R. Peterson, *J. Am. Chem. Soc.*, **92**, 741 (1970); D. J. Faulkner and M. R. Peterson, *ibid.*, **95**, 553 (1973).
(2) E. J. Corey, J. A. Katzenellenbogen, and G. H. Posner, *J. Am. Chem. Soc.*, **89**, 4245 (1967); E. J. Corey and H. Yamamoto, *ibid.*, **92**, 6637 (1970).
(3) For representative references, see E. J. Corey, J. A. Katzenellenbogen, N. W. Gilman, S. A. Roman, and B. W. Erickson, *J. Am. Chem. Soc.*, **90**, 5618 (1968); R. Zurfluh, E. N. Wall, J. B. Siddall, and J. A. Edwards, *ibid.*, **90**, 6224 (1968); W. S. Johnson, T. Li, D. J. Faulkner, and S. F. Campbell, *ibid.*, **90**, 6225 (1968); E. J. Corey and H. Yamamoto, *ibid.*, **92**, 6636 (1970); D. J. Faulkner and M. R. Peterson, *ibid.*, **95**, 553 (1973); P. L. Stotter and R. E. Hornish, *ibid.*, **95**, 4444 (1973).

(4) For reviews of stereoselective olefin synthesis, see D. J. Faulkner, *Synthesis*, 175 (1971); J. Reucroft and P. G. Sammes, *Q. Rev., Chem. Soc.*, **25**, 135 (1971); A. S. Arora and I. K. Ugi in "Methoden der Organischen Chemie (Houben-Weyl)", Vol. V, Part 1b, 4th ed., E. Muller, Ed., Georg Thieme Verlag, Stuttgart, 1972, pp 728–945. For references to newer methods, see A. Marfat, P. G. McQuirk, R. Kramer, and P. Helquist, *J. Am. Chem. Soc.*, **99**, 253 (1977).
(5) LHASA is an acronym for Logic and Heuristics Applied to Synthetic Analysis.
(6) For recent advances in the conformational analysis of six-membered rings, see E. J. Corey and N. F. Feiner, in preparation.
(7) E. J. Corey and D. A. Pensak, *ACS Symp. Ser.* **No. 61**, 1 (1977); and earlier papers, e.g., E. J. Corey, W. T. Wipke, R. D. Cramer III, and W. J. Howe, *J. Am. Chem. Soc.*, **94**, 421 (1972).
(8) E. J. Corey, R. D Cramer III, and W. J. Howe, *J. Am. Chem. Soc.*, **94**, 440 (1972).
(9) E. J. Corey, W. J. Howe, and D. A. Pensak, *J. Am. Chem. Soc.*, **96**, 7724 (1974).
(10) E. J. Corey and W. L. Jorgensen, *J. Am. Chem. Soc.*, **98**, 203 (1976).
(11) E. J. Corey, W. J. Howe, H. W. Orf, D. A. Pensak, and G. Petersson, *J. Am. Chem. Soc.*, **97**, 6116 (1975).
(12) A list of the transforms in the data table for the olefin package is included as Table I.
(13) (a) For a detailed discussion of the CHMTRN language, see H. W. Orf, Ph.D. Thesis, Harvard University, 1976. (b) TBLTRN is a FORTRAN program, written in 1971 at Harvard University by D. E. Barth, available as program 10-168 from the DECUS program library.
(14) (a) E. J. Corey, J. A. Katzenellenbogen, N. W. Gilman, S. A. Roman, and B. W. Erickson, *J. Am. Chem. Soc.*, **90**, 5618 (1970); (b) W. S. Johnson, T. J. Brocksom, P. Loew, D. H. Rich, L. Werthemann, R. A. Arnold, T. Li, and D. J. Faulkner, *ibid.*, **92**, 4463 (1970); (c) D. J. Faulkner and M. R. Peterson, *ibid.*, **95**, 553 (1973); (d) S. B. Bowlus and J. A. Katzenellenbogen, *Tetrahedron Lett.*, 1277 (1973).
(15) For a published synthesis using this strategy, see ref 14b.
(16) Several other routes to **1** are also generated by the program.
(17) Solid and dotted boxes indicate interfering functionality. For further details, see E. J. Corey, H. W. Orf, and D. A. Pensak, *J. Am. Chem. Soc.*, **98**, 203 (1976).
(18) See E. J. Corey and H. Yamamoto, *J. Am. Chem. Soc.*, **92**, 6636, 6637 (1970).
(19) A. K. Long, "LHASA-10 Users' Guide," Harvard University, 1975.
(20) For **16:** E. Demole and P. Enggist, *Helv. Chim. Acta*, **58**, 1602 (1975). For **17:** O. P. Vig, B. Vig, and R. C. Anand, *Indian J. Chem.*, **7**, 1111 (1969).
(21) J. W. Cornforth, R. H. Cornforth, and K. K. Mathew, *J. Chem. Soc.*, 2539 (1959); W. S. Johnson, L. Werthemann, W. R. Bartlett, T. J. Brocksom, T. Li, D. J. Faulkner, and M. R. Peterson, *J. Am. Chem. Soc.*, **92**, 741 (1970); and ref 14c.
(22) J. F. Biellmann and J. B. Ducep, *Tetrahedron Lett.*, 3707 (1969).
(23) E. J. Corey and W. L. Jorgensen, *J. Am. Chem. Soc.*, **98**, 189 (1976).
(24) C. A. Henrick, *Tetrahedron*, **33**, 1845 (1977).

Sulfene and the $^1CH_2/SO_2$ Potential Energy Surface[1a]

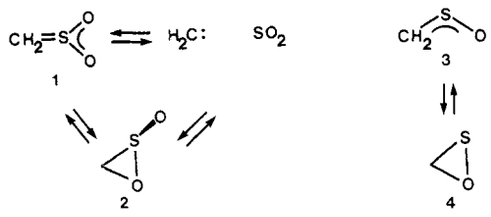
Lars Carlsen[1b] and James P. Snyder*

Department of General and Organic Chemistry, H. C. Ørsted Institute, University of Copenhagen, DK-2100, Copenhagen, Denmark

Received December 12, 1977

Sulfene, $CH_2{=}SO_2$ (1), is a highly reactive intermediate directly observable only at temperatures below −150 °C. Nonetheless it has been implicated in a variety of reactions in the gas phase and in solution. In particular its rearrangement to α-sultine **2**, the cheletropic fragmentation of **2** to $^1CH_2$ and SO_2, and its formation from the latter moieties have drawn attention. Exploration of the potential energy surfaces of these transformations has been undertaken by means of CNDO/B semiempirical calculations and compared with results for the corresponding hydrocarbon systems. Within the CNDO/B framework, the replacement of π-deficient $^1CH_2$ by π-rich SO_2 produces no fundamental mechanistic variations in the $CH_2{=}SO_2$ or α-sultine forming reactions relative to the $CH_2{=}CH_2$ or cyclopropane producing processes. Small but significant differences in the potential energy surfaces are, however, observed. These can be associated with the π-electron distributions of the reacting fragments. The electrocyclic ring closure of sulfene to α-sultine is predicted to follow an "allowed" pathway. It differs from the cyclization of the isoelectronic allyl anion (σ–π correlation) and sulfine (π–π_S correlation) in that a third type of orbital correlation is evident: π–n. High-lying nonbonding oxygen levels introduced by S oxidation are responsible. We suggest the existence of a four-membered ring cyclic sulfoxylate ester 8 on the sulfene potential surface and argue for its intervention in the chemistry of the previously postulated α-sultine. Other sulfene isomers are ruled out as unlikely transients. Finally the cheletropic addition of $^1CH_2$ to SO_2 to give sulfoxylate 8 is identified as an "electron pair excess" pericyclic reaction. The essential "forbiddeness" of this and related reactions is discussed.

Although sulfur trioxide is a stable and familiar laboratory reagent, the carbon analogue, sulfene (1), is a fleeting intermediate directly observable only at temperatures below ca. −150 °C.[2] The behavior of sulfene in solution is characterized by capture of electrophiles at carbon and nucleophiles at sulfur.[2,3] Oligomerization occurs in the absence of external addends.[4] When sulfene is generated in the gas phase at high temperatures, formaldehyde and SO are formed.[4b,5] It has been suggested that intramolecular cyclization to the α-sultine



2 precedes ultimate fragmentation.[5,6] More recently the preparation of transitory **2** at 25 °C has been claimed to result from the addition of methylene, $^1CH_2$, to SO_2[7] and from the peracid oxidation of thiocarbonyl *S*-oxides.[8]

In connection with our continuing fascination with sulfene,[9] sulfine **3**,[10] and oxathiirane **4**,[11] we have investigated the interconversion of the isoelectronic species **1**, **2**, and $^1CH_2/SO_2$ by the CNDO/B procedure.[12] [Throughout this paper, CH_2 is singlet ($^1CH_2$) unless indicated otherwise.] Besides possibly illuminating the above experimental findings, the triad is of interest for two additional reasons. Singlet methylene dimerizes to ethylene[13] and adds to C–C couble bonds to give cyclopropanes.[14] With regard to the first process, a study of the formation of sulfene **1** from CH_2 and SO_2 permits an inquiry into the consequences of the existence of π electrons in one of the combining fragments. A comparison of cyclopropane and α-sultine formation raises the additional question of symmetry and its absence in a pair of model cheletropic reactions. The electrocyclic ring opening of **2** to sulfene addresses the symmetry issue as well. Previous calculations on heteroelectrocylic reactions for oxathiirane **3**[11a] and related systems[15] suggest unexpected electronic features may be associated with the potential energy surface connecting end-point minima.

0022-3263/78/1943-2216$01.00/0