and efficient manner by the mechanistically based program.

V. Conclusion

A program, CAMEO, is being developed to predict the products of organic reactions. Reactant molecules are input via a graphics tablet and CRT, through which all chemist-computer communication takes place. Routines for the perception of sets, rings, functional groups, stereochemistry, aromaticity, and acidities have been written. This information provides the foundation for the internal simulation of basic reaction mechanisms, utilizing general structure-reactivity correlations to control program flow among competing pathways. Products are output on the CRT, where the chemist can select, modify, and resubmit structures. Repetition of the procedures causes multistep reaction sequences to be created, which are recorded in a synthetic tree.

One of the advantages of the program's design is that new, mechanistically sound reactions may be discovered without specific programming. Work is currently taking place to expand both the number of reaction classes treated and the heuristics used in directing flow among the pathways involved.

Acknowledgment. The original equipment purchases for this project were made possible by grants from the Du Pont Young Faculty fund and Research Corporation. Acknowledgment is made to the donors of the Petroleum Research Fund, administered by the American Chemical Society, for partial support of this work. Gratitude is also expressed to the Dreyfus and Sloan Foundations for assistance. We are indebted to Barbara L. Roos-Kozel for developing the ring and stereochemical perception routines in CAMEO. Synthetic consultation with Professor P. L. Fuchs has been most valuable.

Computer-Assisted Synthetic Analysis. Techniques for Efficient Long-Range Retrosynthetic Searches Applied to the Robinson Annulation Process

E. J. Corey,* A. Peter Johnson,[1] and Alan K. Long

Department of Chemistry, Harvard University, Cambridge, Massachusetts 02138

Received November 23, 1979

One of the major fundamental problems in computer-assisted synthetic analysis is the development of methods for conducting goal-driven, deep (or long-range) search. Among the many problems associated with this approach is the avoidance of unfruitful lines of analysis which not only slow the problem-solving process but inundate the user with an unacceptable number of possibilities. Since chemists face similar difficulties during their attempts to plan syntheses, progress with the computer-assisted approach can provide perspective. The newest developments in multistep analysis in the Harvard program LHASA are illustrated for the Robinson annulation search. LHASA treats a target structure systematically, examining each of the 12 possible positions for an α,β-unsaturated ketone in each six-membered ring. Sequences of retrosynthetic steps called "procedures" are applied to remove obstacles blocking performance of the key simplifying disconnection, the Robinson annulation transform. A new technique for preevaluation of these procedures is used to ensure that only the sequences most likely to succeed in the laboratory are actually displayed to the chemist. Several chemical examples are shown.

One of the most powerful strategies available for simplifying the analysis of a complex synthetic problem is the key-reaction-based strategy. This strategy depends on the selection of an important reaction to construct a crucial section of the synthetic target. After application of the key reaction, a series of reactions are chosen which convert the functionality and structural features of the key-reaction product into the corresponding features in the target molecule.

In antithetic (or retrosynthetic) analysis[2] an analogous strategy is correspondingly effective. After selection of a key transform (or retroreaction), structural features in the target molecule are correlated with those required by the transform (that is, the features characteristically produced by the corresponding key reaction). Finally, "subgoal" transforms are selected which remove (antithetically) features deleterious to the application of the key transform and introduce features required by that transform.

This strategy has been employed at two levels in the LHASA program for computer-assisted synthetic analysis.[3] At the first level a number of structurally simplifying transforms have the ability to request other transforms which either exchange one functional group for another (functional group interchange, or FGI)[4] or introduce a desirable functional group (functional group addition, or FGA) and so pave the way for the operation of the main transform. At a higher level, certain powerfully simplifying (or key) transforms, for example, the Diels-Alder,[5] Robinson annulation, and halolactonization[6] transforms, have much more extensive subgoal capabilities. The search procedures driven by these transforms can generate retrosynthetic sequences of up to 20 steps in order to set up the structure required for valid operation of the key transform.

Machine application of these latter transforms uses a data-driven, binary search technique[5] to identify and re-

(1) Department of Organic Chemistry, The University, Leeds LS2 9JT, England.

(2) Corey, E. J. *Q. Rev., Chem. Soc.* **1971**, *25*, 455.

(3) For recent LHASA publications, see: Corey, E. J.; Long, A. K. *J. Org. Chem.* **1978**, *43*, 2208.

(4) Sequential FGI's of up to four steps are currently available in LHASA. See: Corey, E. J.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1976**, *98*, 203.

(5) Corey, E. J.; Howe, W. J.; Pensak, D. A. *J. Am. Chem. Soc.* **1974**, *96*, 7724.

(6) Corey, E. J.; Long, A. K.; Mulzer, J.; Orf, H. W.; Johnson, A. P.; Hewett, A. P. W., in preparation.

0022-3263/80/1945-2051$01.00/0

move obstacles to performance of the key transform. The long-range searches in LHASA have proven effective in generating plausible and often unexpected synthetic pathways. However, when these searches are conducted without a provision for predicting a priori the depth of the search required to find a successful disconnection, considerable machine time can be expended in unfruitful search if an unsuitable target is being analyzed. The problem of unsuccessful search becomes more serious for antithetic sequences involving many steps and assumes major importance as the number of key transforms to be tested is increased.

This article describes a new technique for long-range search which overcomes this disadvantage and illustrates the approach for the specific case of the Robinson annulation transform.[7] The technique is also generally useful for many other key-transform-oriented searches.

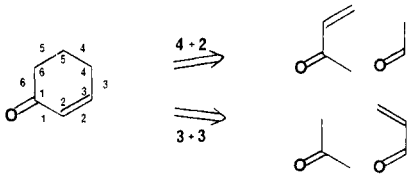
Matching Operations in Retrosynthetic Analysis

When the objective of a long-range antithetic search table is the application of a key simplifying transform, this ultimate aim can be regarded as a "transform goal", or T-goal. An equivalent but alternative approach is to view as the goal of the search the generation of a *specific structure* which allows direct application of the key transform. This type of goal, the "structure goal" or S-goal, may be more convenient than the T-goal in that the structural features required by the key transform are explicitly identified. The analysis then reduces to a matching operation between the structural features of the synthetic target and the structural features of the S-goal.

It is helpful to define here two types of matching operations which are used in retrosynthetic analysis, full structure and substructure. In the "full-structure match" the goal is the antithetic conversion of the target to a specific compound. Full-structure matching is not currently used in LHASA but will clearly become important in an industrial environment where the cheapness and availability of specific starting materials may be a dominating factor in the design of a synthesis. In the "substructure match" the goal is the generation of one or more members of a group of compounds (such as Diels–Alder adducts) with a common subunit, or S-goal. This second type of matching operation is the basis for all the key-transform-based strategies in LHASA.

The Robinson Annulation S-Goal

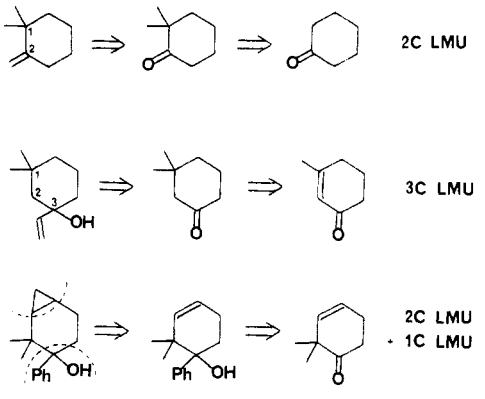
For any substructure matching algorithm it is necessary to define carefully the salient characteristics which the S-goal must have. In the case of the Robinson annulation, the reaction product must possess the following features: (a) a nonaromatic, carbocyclic, six-membered ring; (b) a carbonyl group at atom 1 and a double bond at bond 2 (numbering as shown below); (c) no multiple bond or



three-membered ring on atoms 4, 5, or 6; (d) a hydrogen on atom 6 for the "4 + 2" Robinson and a hydrogen on atom 4 for the "3 + 3" Robinson; (e) no donor groups on atom 5 (they would deactivate the Michael acceptor for both "4 + 2" and "3 + 3" annulations); and (f) no leaving groups on the off-ring carbon atoms adjacent to atoms 2, 4, or 6 (since at various stages of the annulation process carbanions are generated at atoms 2, 4, and 6).

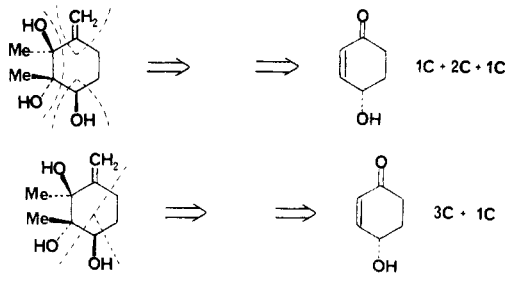
Matching Procedures

The problem of matching a target structure to an S-goal is one that can be solved in many ways. To the synthetic chemist one of the most appealing approaches lies in systematically stripping away interfering structural features which are present in the target, while concurrently introducing the units required for an eventual simplifying disconnection. This strategy, which is strictly geared to the essential structural features of the S-goal, is readily adapted to analysis by computer. A "local" matching operation is carried out in which both target and S-goal are dissected into structural subunits and each subunit of the target is retrosynthetically converted to the corresponding subunit in the S-goal. In this paper these subunits are referred to as "localized matching units" (LMU's). The transforms which carry out the local matching process are keyed by structural features within the LMU and only effect structural changes to carbon atoms within the LMU. Three examples of the LMU matching process are shown below.



Key functional groups in the S-goal play a large part in suggesting LMU assignments. For example, if C=C is present in the S-goal, it is not possible to have an LMU assignment which involves only one of the two carbon atoms of the double bond because it is not possible to effect structural changes at one end and not the other.

In order to match a given target and S-goal it is necessary to assign each of the atoms of the target which differ from the corresponding atoms of the goal to an LMU. As shown below, it is possible to have more than one such complete LMU assignment for any target/S-goal combination.



Chemistry Subroutines

Each different complete LMU assignment represents a new approach for accomplishing the matching operation. Each of the component LMU assignments represents a choice of a particular tactic (or group of related tactics) for performing the matching process on the atoms within the LMU. For example, a 2-C LMU assignment in which

(7) Jung, M. E. *Tetrahedron* **1976**, *32*, 3.

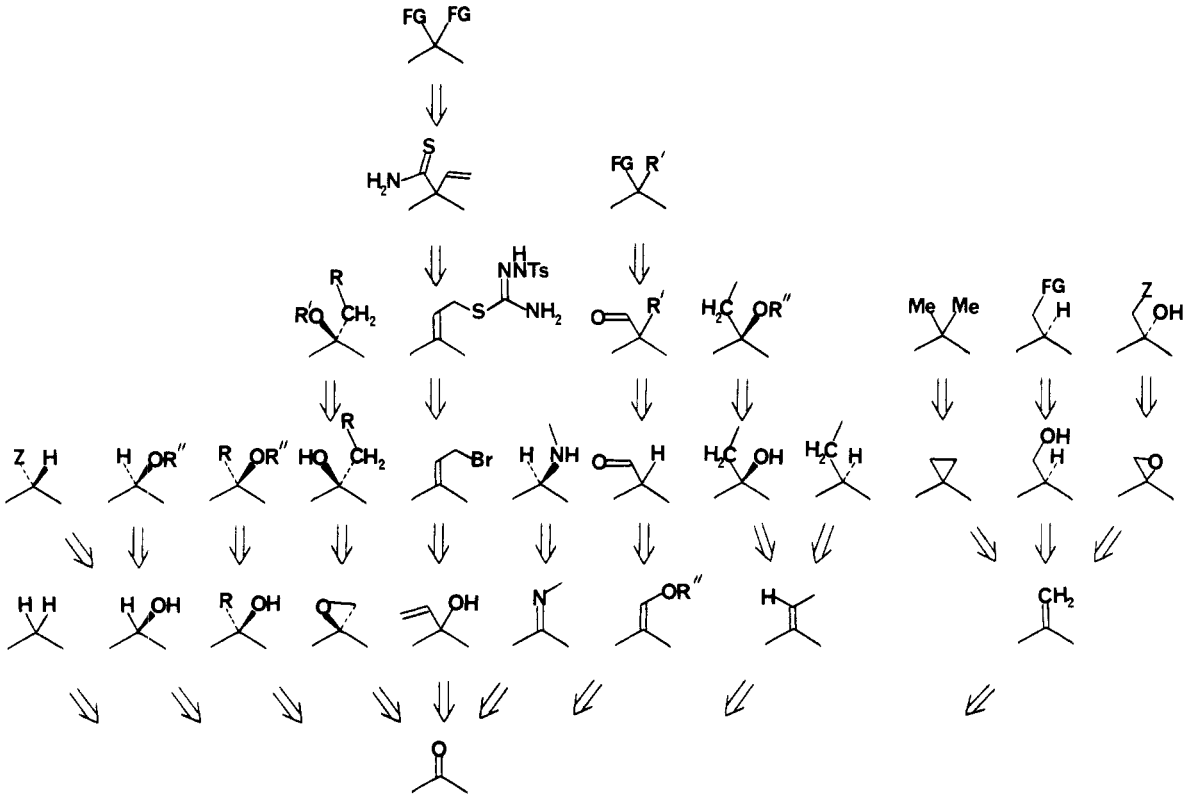


Figure 1. A selection of the transforms available to subroutine GET CO.

the two carbon atoms in the S-goal are joined by a double bond invariably means that the isolated C=C tactic has been chosen to perform the local matching process. Since key local-matching tactics such as the isolated C=C double-bond tactic are likely to be used repeatedly in matching operations, each tactic has been coded as a subroutine in LHASA. Every such "chemistry subroutine" handles a wide variety of LMU's, converting each LMU retrosynthetically to the desired core functionality and often simplifying the target by removal of appendages along the way. Each retrosynthetic step is thoroughly evaluated to ensure that there are no structural features in or around the LMU which would prohibit application of the required transform(s).

The most important chemistry subroutines in the Robinson annulation search table are listed below and described in detail in the text that follows:

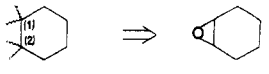
(a) GET CO: generates a carbonyl at a 1C LMU.
(b) GET OH: generates a hydroxyl at a 1C LMU.
(c) GET EPOXIDE: generates an epoxide at a 2C LMU.
(d) GET DB: generates a C=C at a 2C LMU.
(e) DEALKYLATE: removes appendages α to C=O by dealkylation (2C LMU).
(f) GET ENONE: removes appendages from positions α and/or β to C=O by conjugate addition/alkylation (3C LMU).
(g) RING3: removes appendages from positions α and β to C=O by cyclopropanation/alkylation (3C LMU).
(h) CUPRATE: checks appendages for functionality corresponding to a legitimate cuprate reagent.
(i) CLAISEN: performs allylic transpositions.
(j) ROBMCH: performs the actual Robinson annulation disconnection.

The antithetic transformations available to the chemistry subroutines will be described here briefly.[8] Note that each subroutine is called with a local path, which is referred to as "specified atom 1", "specified atom 2", etc.

Subroutine GET CO is the longest of the chemistry subroutines. This routine tries a number of methods for antithetically placing a ketone on specified atom 1, among them organometallic attack, two-group addition to a Wittig product, alkylation of an α-off-ring aldehyde, [2,3] sigmatropic rearrangement to a vinyl dithioacid ester,[9] Simmons–Smith cyclopropanation of an exo methylene, and spiro-epoxide formation and opening. A detailed outline of the transforms which GET CO can call into play is given in Figure 1. The tree structure shown here is quite typical and reflects the fact that while there are many entry points, there is only one successful exit (the subroutine S-goal).

Subroutine GET OH places a hydroxyl on specified atom 1. GET OH is a simple subroutine, since the alternative of calling the powerful GET CO routine and then obtaining a hydroxyl via functional-group interchange exists.

Subroutine GET EPOXIDE tries a number of methods for antithetically obtaining an epoxide between specified atom 1 and specified atom 2. The routine is divided into two sections, one for locked six-membered rings[10] (diaxial opening of the epoxide is assumed) and one for all other types of rings. Subroutine GET DB uses a variety of

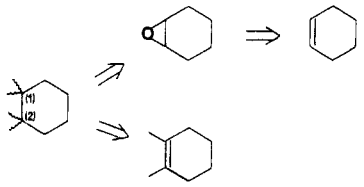


techniques to try to place a double bond between the two specified atoms. First, GET EPOXIDE is called since it is frequently able to remove appendages. If GET EPOXIDE fails, several double-bond-addition transforms are attempted, depending on the substitution at the two specified atoms.

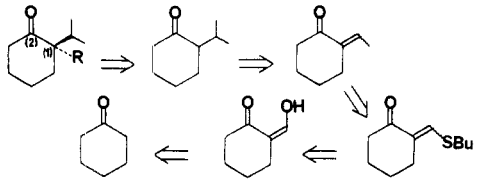
(8) For complete flow charts of the chemistry subroutines, see: Long, A. K. Ph.D. Thesis, Harvard University, 1979, Chapter 7.

(9) Baldwin, J. E.; Walker, J. A. *J. Chem. Soc., Chem. Commun.* **1972**, 354.

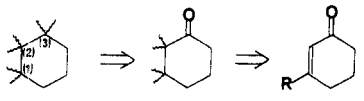
(10) Corey, E. J.; Feiner, N. F. *J. Org. Chem.*, **1980**, *45*, 757, 765.



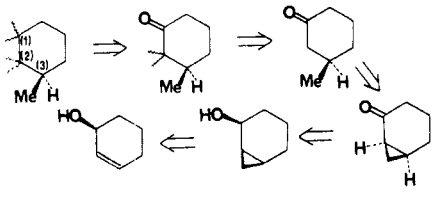
Subroutine DEALKYLATE tries to remove appendages on specified atom 1 via antithetic dealkylation of a ketone at specified atom 2. If specified atom 1 is a quaternary center, one or both appendages may be removed. Sequences involving addition–elimination mechanisms[11] are also included, for example:



Subroutine GET ENONE uses cuprate addition chemistry coupled with alkylation of enolate intermediates to remove appendages in going back to an α,β-unsaturated ketone.

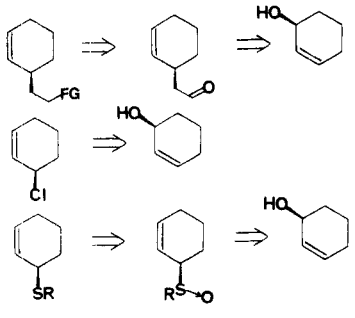


Subroutine RING3 tries to demethylate specified atom 3 by cyclopropane formation at specified atoms 2 and 3. The routine places a ketone on specified atom 1 and uses it to dealkylate specified atom 2, if necessary. Additional interchanges are performed to remove the cyclopropane and antithetically generate an allylic alcohol.



Subroutine CUPRATE sets up functionality on the designated appendage to correspond to a legitimate cuprate reagent.

Subroutine CLAISEN performs three kinds of allylic transpositions (Claisen rearrangement, allylic rearrangement with $SOCl_2$, and [2,3] sigmatropic rearrangement of a sulfenate ester).[12]

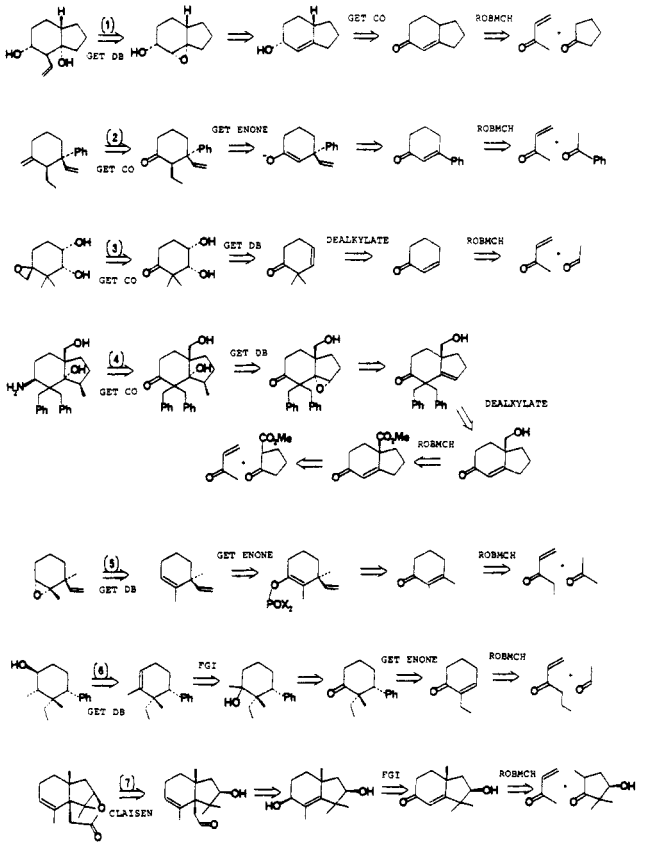


Subroutine ROBMCH performs the retrosynthetic operation corresponding to the Robinson annulation itself.

(11) (a) Coates, R. M.; Sandefur, L. O. *J. Org. Chem.* **1974**, *39*, 275. (b) Corey, E. J.; Chen, R. H. K. *Tetrahedron Lett.* **1973**, 3817. (c) Coates, R. M.; Sowerby, R. L. *J. Am. Chem. Soc.* **1971**, *93*, 1027.

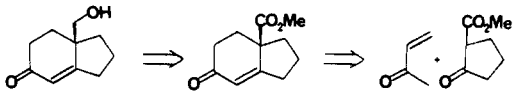
(12) Evans, D. A.; Andrews, G. C.; Sims, C. L. *J. Am. Chem. Soc.* **1971**, *93*, 4956.

Scheme I. Sample Retrosynthetic Analyses Illustrating the Seven Robinson Annulation Procedures[a]



[a] Procedure numbers are indicated in parentheses. Chemistry subroutines responsible for individual transforms are identified above the corresponding retrosynthetic arrows. In all of these sample sequences except for those illustrating procedures 4 and 7, "3 + 3" disconnections would be found by LHASA as well as the "4 + 2" disconnections shown. The "3 + 3" precursors have been omitted for simplicity.

ROBMCH also requests subgoal transforms (FGI's) to convert appropriately placed target functionality to an additional withdrawing group to ensure regioselectivity in the annulation step.



It should be emphasized that chemistry subroutines are linked to specific structural features of subunits in the overall S-goal but are independent of the complete structure of the goal. Thus, the same chemistry subroutines can be used in matching processes involving a wide variety of S-goals such as the Diels–Alder S-goal, the Birch reduction S-goal, etc. Once a sizeable collection of chemistry subroutines has been assembled, devising matching operations relating to different ring-forming reactions becomes a relatively simple process, dependent only on the various ways in which the available chemistry subroutines can be linked together to carry out the whole matching operation.

Matching Procedures for the Robinson Table

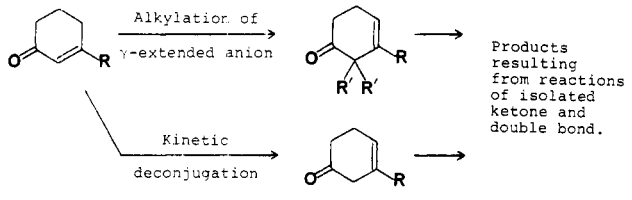
Each combination of chemistry subroutines which is capable of generating the required S-goal from the target is referred to as a substrategy, or "procedure". There are, in general, several such substrategies in a given search strategy. For the Robinson annulation strategy, for ex-

ample, seven procedures have been selected. Each of these procedures is based on some aspect of the chemistry of α,β-unsaturated ketones or the closely related allylic alcohols, as described below. Specific examples of the operation of these matching procedures are shown in Scheme I.

Procedure 1. The most straightforward procedure involves treating atom 1 of the target as an isolated entity that can be antithetically converted to a ketone by the subroutine GET CO. Bond 2 is treated as another isolated entity that can be converted to a double bond by the operation of GET DB.

Procedure 2. This procedure attempts to strip away appendages from atom 3 via transforms corresponding to conjugate addition of an organocuprate or other nucleophile. Any appendages at atom 2 are candidates for removal by the transform corresponding to alkylation of the enolate anion derived from conjugate addition or reduction.

Procedure 3. This procedure is based on the tactic shown below.



Procedure 3 is only allowed to operate if there is no possibility of formation of the alternative γ-extended enolate anion.

Procedure 4. This procedure is similar to procedure 3 except that the alkylation of the alternative γ-extended enolate anion generates an exo double bond. The example in Scheme I also illustrates the function of the subroutine ROBMCH, which removes any remaining barriers to the Robinson annulation and then applies the goal transform.

Procedure 5. In the synthetic direction, this tactic involves an initial conjugate addition or reduction followed by formation of a double bond between atoms 1 and 2 either via an enol phosphate or, if atom 6 does not bear a hydrogen, by conversion of the corresponding ketone to an alcohol followed by elimination.

Procedure 6. This procedure only differs from the previous one in that the double bond is generated between atoms 1 and 6 instead of between atoms 1 and 2. Procedure 6 is only allowed to operate if atom 2 either lacks a hydrogen or is a bridgehead.

Procedure 7. The synthetic sequence corresponding to this procedure involves conversion of the ketone at atom 1 to hydroxyl followed by conversion of the alcohol to a derivative which can undergo a sigmatropic shift to attach a functional group (amine, halide, or sulfoxide) or a new carbon appendage (via Claisen rearrangement) at atom 3. It should be noted that the CLAISEN subroutine is able to convert a variety of appendages antithetically to the aldehyde which keys the Claisen rearrangement transform.

Procedure Selection

For a given six-membered ring, there are 12 alternative paths over which each of the above 7 procedures might operate (6 alternative starting points and two alternative directions around the ring). This represents a total of 84 different procedure/path combinations, each of which might serve to generate the Robinson S-goal. Although in principle a computer could apply each of these alternative procedures to a target molecule and then evaluate the relative effectiveness of all the successful procedures, this would be an expensive and time-consuming process. Obviously some element of look-ahead, by which an a priori estimate can be made as to how effective each procedure will be when applied to the current target, is necessary. With such a method for prior evaluation, procedure selection becomes an automatic process driven by the absence or presence of particular structural features in the target molecule.

The module responsible for this a priori procedure assessment in LHASA is termed Prior Procedure Evaluation (PPE). Since PPE is treated elsewhere in considerable detail,[6] only a brief description will be given here. PPE assigns a rating to each LMU along the path according to the estimated number of retrosynthetic steps necessary to convert the functionality at that LMU to the desired S-goal. First, LMU's are divided into sets labeled "there", "easy", "moderate", "difficult", and "impossible", according to how far they are removed synthetically from the corresponding S-goal functionality. Each of these sets is then given a rating, usually 0, 1, 3, 5, and 100, respectively. Since each LMU $\Rightarrow$ S-goal conversion corresponds to the specific task of a chemistry subroutine, these numbers of estimated steps are called "subroutine ratings". For a given path/procedure combination, a "procedure rating" is calculated by summing the "subroutine ratings" for the chemistry subroutines called by that procedure and any additional steps performed by the procedure itself. In the procedure 7 sequence in Scheme I, for example, the procedure rating is 3: the 2-carbon appendage at the fusion is in the CLAISEN EASY set (a subroutine rating of 1), and two additional steps (the Claisen rearrangement and the functional-group interchange of hydroxyl $\Rightarrow$ carbonyl) are performed in the procedure itself.

If a procedure rating is greater than or equal to 100, either because one of its constituent subroutine ratings was 100 or because of an interfering structural feature not contained in one of the LMU's, PPE discards that particular path/procedure combination and examines the next one. When all the procedures have been rated for all the paths, the 10 lowest-ranked are sorted by rating and attempted in order by the program. As a result of this very efficient prescreen,[13] only the best procedures are actually displayed to the chemist.

Sample Antithetic Analyses

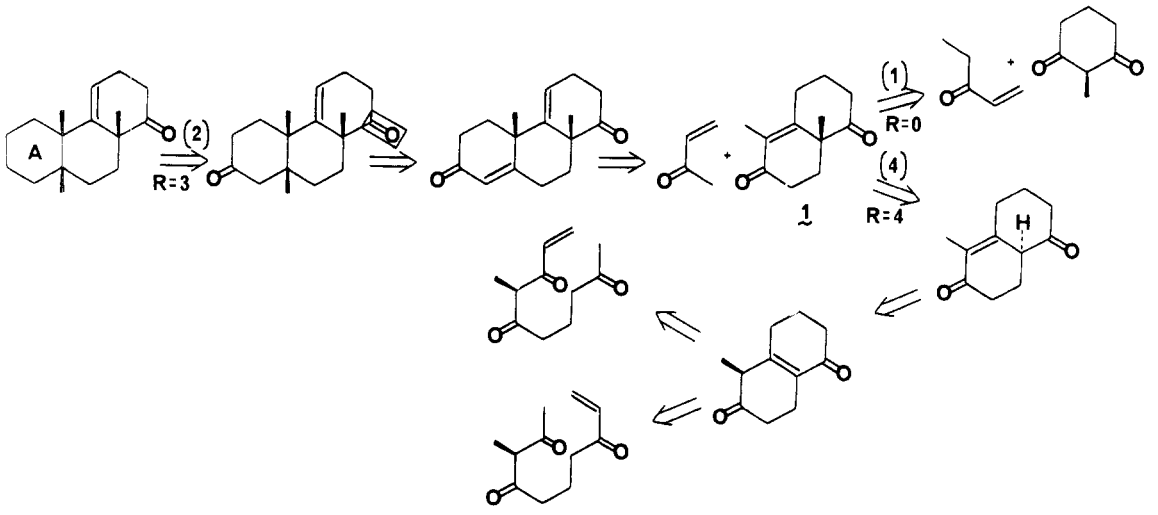
Results from the Robinson annulation module in LHASA have been very promising. Several complete analyses have been performed, both on examples from the literature and on a number of test structures. The program is consistently able to reproduce published synthetic routes and also to suggest novel yet reasonable additional pathways. The sample retrosynthetic analyses in Schemes II and III amply demonstrate the power of the search.

The test structure in Scheme II was designed as a target for synthesis by Robinson annulation. The PPE module found between 11 and 20 possible Robinson syntheses for the A ring alone. Of the ten lowest ranked of these, six sequences passed further evaluation and were displayed to the chemist. It was quite gratifying to see that the first sequence displayed (rating = 3) corresponded to a route which clearly would be effective. When structure 1 was reprocessed using the Robinson annulation option, six sequences were found by PPE, of which two passed further evaluation (see Scheme II). The first of these (rating = 0) is a well-demonstrated and known process, and the

(13) PPE takes an average of less than 20 s of computer time to select the 10 best procedures for each ring.

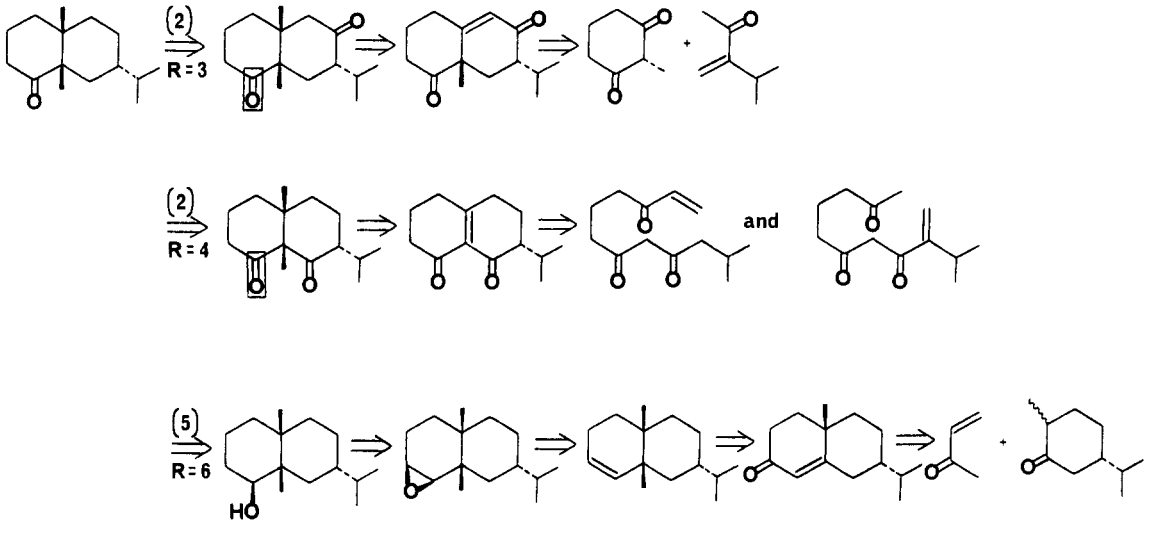
(14) Corey, E. J.; Orf, H. W.; Pensak, D. A. *J. Am. Chem. Soc.* **1976**, *98*, 210.

Scheme II. Sample LHASA-Generated Robinson Annulation Sequences for the A Ring of the Indicated Target[a]



[a] Procedure numbers are shown in parentheses, and PPE ratings are included under the first retrosynthetic arrow for each sequence. The solid box indicates protectable interfering functionality.[14] Structure 1 was reprocessed with the Robinson annulation option.

Scheme III. Sample Sequences Suggested by the LHASA Robinson Annulation Module for the Synthesis of Valeranone[a]



[a] Procedure numbers are shown in parentheses, and PPE ratings are included under the first retrosynthetic arrow for each sequence. Solid boxes indicate protectable interfering functionality.[14]

second (rating = 2) suggests an intriguing pair of intramolecular annulations, little explored to date.

Note that in each of these sequences the PPE rating is one less than the number of synthetic steps. This discrepancy reflects the fact that the actual Robinson annulation step is not included in the rating, as it is common to all Robinson sequences. The correspondence between PPE rating and number of steps is not always so good, especially in targets which require more complicated sequences of subgoal transforms. In general, however, the agreement has been very good, and considerable confidence in the predictive powers of PPE has developed.

Scheme III shows three sequences suggested by LHASA for the synthesis of valeranone (previously prepared by two routes[15]). PPE identified more than 30 Robinson annulation routes to this target, of which 15 passed evaluation and were displayed to the chemist. The sequences shown are among the lowest ranked of these 15. The first route suggested, while quite simple, is an excellent illustration of the power of the Robinson search. By considering both six-membered rings in the target and all the possible Robinson paths around each ring, LHASA has found a route which is short and stereocontrolled, with readily available starting materials and a single (easily solved) problem of functional-group interference. The second route, somewhat less likely to succeed than the first, shows again that intramolecular possibilities should not be overlooked. The third sequence provides an insight into some of the newer stereochemical capabilities in LHASA. In the first step (the alcohol-oxidation transform), the program automatically generates both epimeric alcohols. For the second step, however, the on-line conformational analysis module[10] predicts that only the β epimer could result from diaxial opening of the desired epoxide. Hence, the α alcohol is discarded and only the β epoxide is displayed to the chemist.

Conclusion

Development of efficient yet sophisticated long-range searches is one of the most important goals of computer-assisted synthetic analysis. In complicated multistep

(15) (a) Marshall, J. A.; Bundy, G. L.; Fanta, W. I. *J. Org. Chem.* **1968**, *33*, 3913. (b) Wenkert, E.; Berges, D. A. *J. Am. Chem. Soc.* **1967**, *89*, 2507.

syntheses especially lies the danger that the synthetic chemist can become channeled into one line of thinking and overlook more efficient methods, simply because the number of combinatorial possibilities for synthetic routes is so great. The challenge of designing a computer package to direct a long-range search, then, is to be able to handle a very large diversity of target structures without becoming polarized into overly restrictive antithetic channels. The Robinson annulation package described in this paper represents a highly effective approach to this problem.

It is not possible, and certainly not desirable, for the search-table writer to "lead the computer by the hand" back antithetically from an arbitrary target structure to a preselected key intermediate. The aim of the long-range search package is rather to provide a framework wherein the computer program can use its own subgoal capabilities to arrive at desired precursors in an efficient fashion. It is in such an unbiased reduction of the combinatorial explosion that the computer has an advantage over the chemist and has a potential for making a very positive contribution to the solution of synthetic problems in years to come.

Acknowledgment. This research was carried out over a period of years with considerable help from other members of the LHASA group at Harvard. Dr. Johann Mulzer introduced "Localized Matching Units" and "Chemical Procedures" to LHASA, and Dr. Harry W. Orf and Dr. Alan Hewett greatly extended the capabilities of the CHMTRN language to permit the developments described herein. Other significant contributions were made by Dr. John Vinson and Dr. W. Jeffrey Howe. We are grateful to the National Institutes of Health and the Science Research Council (U.K.) for grants in support of this work.

Molecular Orbital Theory of Supernucleophiles: Complementary Criteria and Supporting Evidence

Walter B. England,* Peter Kovacic,* Stephen M. Hanrahan, and Martin B. Jones

Department of Chemistry and The Laboratory for Surface Studies, The University of Wisconsin—Milwaukee, Milwaukee, Wisconsin 53201

Received February 1, 1979

Molecular orbital (MO) theory is used to show that the two commonly invoked and apparently different electronic structure criteria for supernucleophilic propensity are complementary. The complementary character is realized through the inherent flexibility of MO wave functions. Canonical MO's express the Ingold criteria. Localized MO's express the lone-pair-repulsions criterion. A consequence of the complementarity is that more extensive models of the α effect may be based on either electronic structure criterion. A simple treatment of the electronic structure of supernucleophiles is likewise a consequence of the complementarity. Supernucleophilic propensity may be characterized by the numbers of valence electrons and α atoms that are associated with the nucleophilic moieties. The concepts of enhanced supernucleophilic and moderated nucleophilic propensities are proposed on the basis of electronic structure arguments. It is found that both types of nucleophilicity are observed in potential supernucleophiles. Experimental evidence is presented in support of the proposed degree of nucleophilic character for the dichloroamide anion and trichloroamine.

Canonical molecular orbitals are the usual delocalized molecular orbitals (MO's).[1] Localized MO's are MO counterparts of valence bond (VB) concepts such as lone pairs.[2] Canonical MO's and localized MO's are completely equivalent descriptions because of the well-known arbitrariness of MO wave functions.[3] A consequence of the equivalence is that apparently different valence models may be essentially the same. For example, Walsh's rules[4] in terms of canonical MO's are equivalent to Gillespie's rules[5] in terms of localized MO's.[6]

Two commonly invoked and apparently different criteria for supernucleophilicity have been proposed on the basis of the electronic structure of nucleophilic moieties. The *lone-pair-repulsions criterion*[7,9-11] is based on repulsions between lone-pair electrons on the nucleophile and the α atom. The *Ingold criterion*[12] requires the highest energy occupied MO to be antibonding with a node that is normal to the bond between the nucleophile and the α atom. We show that the two criteria are quantum mechanically equivalent. The lone-pair-repulsions criterion can be expressed with localized MO's, and the Ingold criterion can be expressed with canonical MO's. *Consequently, extensive models of supernucleophilicity, such as reaction schemes and catalytic arguments, may be based on either electronic structure criterion.* This is the spirit of Klopman's approach,[8] which requires the highest occupied orbital to have an especially high energy.

Given the equivalence of the two electronic structure criteria, a systematic analysis of the electronic structure of supernucleophiles can be achieved. Two numbers are necessary to characterize potential supernucleophiles, the

(1) Roothaan, C. C. J. *Rev. Mod. Phys.* **1951**, *23*, 69.
(2) England, W.; Salmon, L. S.; Ruedenberg, K. *Fortschr. Chem. Forsch.* **1971**, *23*, 31.
(3) Fock, V. *Z. Phys.* **1930**, *61*, 126.
(4) Walsh, A. D. *J. Chem. Soc.* **1953**, 2260, 2266, 2288, 2296, 2301, 2306.
(5) Gillespie, R. J. *J. Chem. Educ.* **1963**, *40*, 295; *J. Am. Chem. Soc.* **1960**, *82*, 5978; *J. Chem. Soc.* **1963**, 4672.
(6) Thompson, H. B. *Inorg. Chem.* **1968**, *7*, 604.
(7) Ibne-Rasa, K. M.; Edwards, J. O. *J. Am. Chem. Soc.* **1962**, *84*, 763.
(8) Klopman, G.; Tsuda, K.; Louis, J. B.; Davis, R. E. *Tetrahedron* **1970**, *26*, 4549.
(9) Aubort, J. D.; Hudson, R. F. *J. Chem. Soc. D* **1970**, 937.
(10) Filipinni, F.; Hudson, R. F. *J. Chem. Soc., Chem. Commun.* **1972**, 522.
(11) Aubort, J. D.; Hudson, R. F.; Woodcock, R. C. *Tetrahedron Lett.* **1973**, 2229.
(12) Ingold, C. K. "Structure and Mechanism in Organic Chemistry", 2nd ed.; Cornell University Press: Ithaca, NY, 1969; p 452.

0022-3263/80/1945-2057$01.00/0