Computer-Assisted Synthetic Analysis. Performance of Tactical Combinations of Transforms

Alan K. Long* and John C. Kappos†

Department of Chemistry, Harvard University, Cambridge, Massachusetts 02138

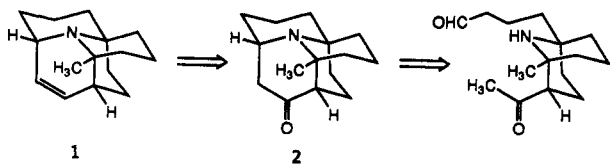
Received October 12, 1993⊛

The LHASA program for computer-assisted synthetic analysis has been expanded to include a module which executes "tactical combinations" (TCs) of transforms. The new module employs a readily modifiable knowledge base of TCs written in a two-dimensional pattern language. TCs are selected *via* a mapping operation which establishes correspondence between the target structure and the retron-substructure patterns of TCs. Successfully mapped TCs are screened to ensure that they fulfill the tactical constraint specified by the chemist at the outset of the retrosynthetic analysis. Transforms are identified as candidates for a TC step by comparison of the differences between retron and precursor patterns for the TC step with the changes effected by LHASA transforms. Sample antithetic analyses are included, and future extensions to TC capabilities are discussed.

I. INTRODUCTION

The computer program LHASA, currently entering its third decade of development, is designed to produce retrosynthetic solutions to complex synthetic problems.[1] The overarching goal of the retrosynthetic analysis is to simplify the target back to readily available starting materials. The program receives a target molecule as input and generates hypothetical precursors from it by the application of theoretical retroreactions, or "transforms". Since the direction of analysis is opposite that of laboratory synthesis, the program must recognize molecular substructures corresponding to reaction products and apply reaction mechanisms in the reverse sense. Therefore, a transform entry must contain a description of the minimum molecular substructure required for its application, the keying element or "retron", and a specification of the atom and bond changes required to generate the precursors or reactants. In fact, each LHASA knowledgebase entry contains additional information, including a substrate-independent "utility rating"[2] and a series of computer-readable questions, or "qualifiers", written in the chemical English language CHMTRN,[3] which modify the utility rating according to the applicability of the transform to the particular target structure being analyzed.

The ultimate goal of retrosynthetic simplification is achieved by applying "goal" transforms to the target. Goal transforms are characterized by their ability to disconnect retrosynthetically carbon–carbon or carbon–heteroatom bonds, reconnect long appendages (especially those containing stereocenters), or effect molecular rearrangements. In addition to goal transforms, there exists a body of transforms which are distinguished as nonsimplifying, since they merely manipulate functional groups, usually without disconnecting carbon–carbon bonds. These "subgoal" transforms are crucial, however, in that they can be used to establish the retron of a goal transform when only a part of the goal transform keying element is present in a target. For example, structure **1** contains only a partial retron for a common variant of the Mannich transform, a transform keyed by a tertiary amine and a ketone separated by a path of two bonds. However, a nonsimplifying step, the Shapiro transform, can be employed to convert the target structure olefin to a ketone retrosynthetically. The precursor **2** now contains the full retron for the Mannich transform and is greatly simplified by this goal transform application.



1 2

A LHASA processing session begins with the graphical entry of an organic molecule to be analyzed. The program then enters a perception phase, analogous to the perception a chemist would conduct upon first encountering a target molecule. Particular account is taken of synthetically significant molecular features: functional groups, rings, ring fusions and bridges, stereocenters, etc. These factors are especially important since they define molecular substructures, the basis for keying transforms, and because they establish a vocabulary through which the program can ask further questions to ascertain the applicability of individual transforms to the target structure.

After perception, the user must select a strategy and substrategy, or tactic, to constrain the retrosynthetic search.[4] For instance, if the chemist knows of a particularly appropriate starting material, the starting structure can be entered, and the program will limit the analysis to routes which have as an eventual precursor the manually designated starting material.[5] (The program is also capable of selecting its own starting materials from a library file.) Alternatively, the analysis can be constrained by pursuing the application of an extremely simplifying transform (e.g., the Diels–Alder transform)[6] or the retrosynthetic disconnection of a manually-designated or machine-perceived strategic bond.[7] If the chemist does not wish to bias the analysis, s/he may elect to process in "unconstrained" mode. In this mode, goal transforms are applied "opportunistically" to the target. In other words, all goal transforms which are directly or almost directly keyed by substructures in the target are attempted. These strategies and tactics employed by LHASA correspond to the most effective problem-solving approaches known to synthetic chemists and thus allow the program to emulate human analysis of synthetic problems.

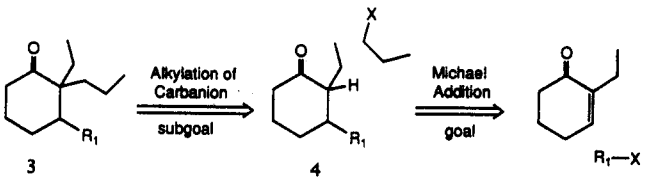
Finally, the program accesses the transform knowledge base to select those goal transforms which simplify the target in the fashion prescribed by the tactical constraint. The goal transforms attempted are directly keyed or partially keyed by the target structure. Short sequences of subgoal transforms

† Recipient of a Fannie and John Hertz Predoctoral Fellowship.
⊛ Abstract published in *Advance ACS Abstracts,* June 1, 1994.

0095-2338/94/1634-0915$04.50/0

are employed to rectify mismatches when a partial match is found between the target and goal transform retron.

We have demonstrated that the LHASA program, which employs the above-described techniques, can design provocative, nonobvious routes to complex molecules. The program, however, is not yet so powerful as to be an indispensible tool to synthetic chemists. One of the primary strengths of the LHASA program which enables it to produce nonobvious retrosynthetic routes is its ability to generate multistep subgoal sequences for mismatch rectification.[8] In assessing the weaknesses of the program, however, we have identified the generation of subgoal sequences as a capability which still requires improvement. There are two limitations which hinder subgoal generation. First, subgoal transforms in the LHASA knowledge base are cross-referenced only by retron and precursor functional groups. In other words, the change effected by a subgoal transform is describable to LHASA only in terms of the functional groups which are modified by the process. For example, the FGI (functional group interchange) of an alcohol to a ketone effects a change which can be accurately specified by considering only the subject and object functional groups. There are numerous transforms, however, which are useful as subgoals but which effect changes not readily delineated by this simple functional group-based scheme. For example, if the retrosynthetic goal for target 3

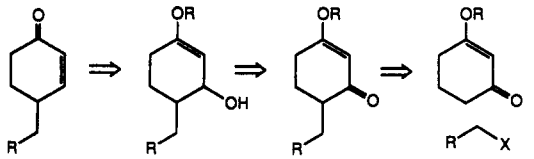


3 4

is to remove appendage R1 and the Michael Addition transform is recognized as a means to this end, then the Alkylation of Carbanion transform would be needed as a subgoal to generate the full retron 4 for the Michael Addition transform. The Alkylation of Carbanion transform, however, effects changes that defy description in terms of functional group differences and thus is not available to LHASA for subgoal application.

The second difficulty in subgoal generation regards a combinatorial explosion of potential subgoal routes when transforms are selected for a lengthy subgoal sequence. Since starting material structures typically contain a significant fraction of the target molecule and the retrons of extremely powerful transforms are usually quite complex, mismatch rectification is often required at more than one site in the target, and a long subgoal sequence can result. If the transforms comprising the sequence are selected from a large knowledge base, many combinations of transforms will, in theory, effect the necessary subgoal change. In the current version of LHASA, which employs only the 200 subgoal transforms classifiable as FGIs (functional group interchanges), FGAs (functional group additions), and FGRs (functional group removals), as many as 50 two-, three-, and four-step sequences can be generated to satisfy a subgoal request. Though only one of these routes is selected for transform execution and displayed to the chemist, much time is often spent determining which sequences to execute and then trying several of them until one succeeds.

To remedy LHASA's inability to execute non-functional group-based subgoals, a new retron and precursor cross-referencing system was devised. This system, which is based on two-dimensional patterns for retron and precursor substructures, is the subject of a separate article.[9] A solution to the difficulties posed by the proliferation of subgoal routes was conceived by recognizing that an entire sequence of transforms could be considered as a single unit of retrosynthetic simplification. This new unit is termed the "tactical combination", or "TC".[10] TCs are characterizable as multistep transform sequences that act as units to achieve specific retrosynthetic objectives, e.g., ring appendage disconnection, functional group transposition, etc. For instance, the Oxyallylic Cation Hydrolysis transform followed by the Hydride Reduction transform and finally by the Alkylation of Carbanion transform form a TC for ring appendage disconnection:[11]



TCs are expected to be useful in eliminating the combinatorial profusion of subgoal routes that can be formulated to rectify a mismatch by greatly reducing the number of retrosynthetic units that must be chosen to bridge the "chemical distance" between a target and the goal transform retron. For example, if an eight-step sequence of subgoals is required, there would be far fewer combinations of four-step TCs than of single-step transforms to bridge this distance.

While searching the chemical literature to gather examples of TCs, it was found that many TCs are not merely subgoal routes but effect powerful retrosynthetic simplification, and as such should be considered goals in themselves. It thus seemed a logical extension to incorporate TC execution into the LHASA goal transform hierarchy, thereby enabling unconstrained and tactically-guided application of goal TCs. Two benefits accrue from this employment of TCs. First, units of simplification more powerful than ordinary goal transforms become available to opportunistic and topological LHASA analyses. Second, TCs lend their built-in depth to these generally breadth-first processing modes. The implementation of tactical combinations as goal sequences is described in this article.

II. DESIGN CRITERIA

A system for machine execution of TCs requires two general facilities: a method of keying TCs and a mechanism for selecting transforms to be executed for each TC step. Considerable planning preceded the development of these facilities, with three principles guiding the design of the package.

First, a general keying mechanism able to accommodate complex substructures was required. Since TCs may consist of many transform steps, several functional groups and even complex cyclic substructures may be required in the retron. The traditional LHASA transform-keying mechanism specifies either a single functional group and a path of bonds extending from it or a pair of functional groups and the bonds between them.[12] A maximum of only two functional groups, however, can be stipulated, and this mechanism was therefore insufficient. The alternative method of keying transforms in LHASA based on a one-dimensional pattern language[13] is capable of handling complicated substructures. Efficient performance of TCs, however, would also require a keying mechanism that specified precursor as well as retron substructures so that consistency checks could be performed between TC steps. Since the one-dimensional pattern language

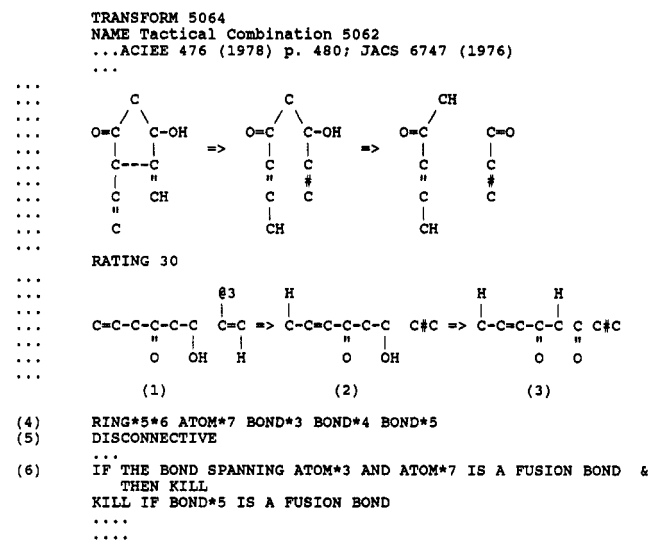


Figure 1. Sample tactical combination knowledge-base entry for a TC consisting of the Ene transform followed by the Aldol transform. (1) Retron 2-D pattern and keying element for the TC. ((2) and (3)) Precursor patterns which specify substructures to be generated by each TC step. (4) Keyword(s) for ring set(s) used to specify rings of variable size. (5) Keyword(s) for special set(s) used for tactical screening of TCs. (6) CHMTRN qualifiers prohibit TC application when molecular features not ruled out by the pattern are encountered.

was not designed to allow specification of precursor substructures, an alternative but equally general pattern-based keying scheme was required.

Second, the method of selecting transforms for each TC step must permit identification of transforms located anywhere in the LHASA knowledge base. A simple approach would be to include in the TC knowledge-base entry a list of transforms to be executed for each TC step. Such a "hard-coded" system, however, would incur enormous overhead in updating the TC knowledge base when new transforms were added to the transform knowledge base. A more general approach, in which the transform knowledge base can be searched for candidate transforms during the LHASA analysis, was required. Optimally, this system would recognize atom and bond differences between the substructure patterns for each TC step. The transform knowledge base could then be searched for entries that effect similar molecular changes. Such a general system would ensure that all appropriate transforms were found for each TC step.

Finally, since TCs will eventually be used as subgoal sequences, they must contain information on the final precursor structure, in addition to the retron and intermediate precursor structures. Since substructure patterns are required for the retron keying element and for the precursor of each step, a final precursor pattern would follow as the source of necessary information. The overall changes effected by a TC would then be recognizable in terms of the atom and bond differences between the retron and final precursor substructure patterns.

III. TWO-DIMENSIONAL PATTERN LANGUAGE

To accomplish TC keying and transform selection for each TC step, a system of two-dimensional (2-D) patterns was developed.[9] An example of a TC knowledge-base entry which employs the 2-D pattern language appears in Figure 1. In this example, the minimum substructure which permits entry into the TC is detailed in the first 2-D pattern (1). This retron pattern is followed by other 2-D patterns ((2) and (3)) which define the precursors for each step. Each precursor pattern represents a substructure which can be generated by application of a single transform to the previous substructure. The patterns are separated from each other by retrosynthetic arrows (⟹) which lie on the "path", or main line, of the pattern. The 2-D pattern language thus provides a flexible keying mechanism which allows specification of certain-sized rings, functional groups, and atom paths to describe substructures of arbitrary complexity.

At the same time, analysis of any pair of consecutive patterns reveals the atom and bond differences between those patterns. These differences can be used as a description of the change required to accomplish that step of the TC. Each transform in the LHASA knowledge base similarly contains a 2-D pattern which describes the atom and bond changes they effect. Thus, by comparing the description of changes required for a TC step to the changes effected by transforms in the knowledge base, candidate transforms for that step can be identified.

IV. IMPLEMENTATION

A. Flow of Control. The executive subroutine TCEXEC guides the performance of tactical combinations in LHASA. Upon specification of a processing tactic, the program branches to the transform executives GPAIR, GSING, and PEXEC and finally to TCEXEC.[14] A simplified flow chart for the TCEXEC subroutine is given in Figure 2. The general procedure for TC execution involves (1) selection of a TC which is both appropriate for the target structure being analyzed and for the processing tactic guiding the retrosynthetic search, (2) identification of candidate transforms for each TC step, and (3) execution of candidate transforms with eventual generation of precursors containing the requested substructures.[15] These steps will now be described in detail.

B. TC Selection. The TC selection process occurs in three phases. First, a "mapping", or correspondence, must be established between the retron 2-D pattern of a TC and the target structure being processed. The mapping phase begins with the selection of one target structure atom to map to one of the pattern path atoms. These atoms are usually selected on the basis of functional group correspondence. For instance, if there is a ketone in both the target and the pattern, the ketone origin in the target will be chosen as the "anchor atom" to begin the mapping. The complete mapping is then grown out from this anchor position by the module PATRNMAP, which is called from TCEXEC. PATRNMAP completes the mapping by considering correspondence of atom type between structure atoms alpha to the last mapped target atom and pattern atoms alpha to the last mapped path atom. For each pattern path atom there may be several target atoms which map. The first candidate is used for the current mapping while the rest are saved for a backtracking phase. After the first mapping is complete, alternative candidates are considered for each path atom, starting from the last path position mapped. This backtracking process, which ensures that all possible matches are found, is illustrated in Figure 3.

The mapping phase concludes with a call to subroutine CHKRNG, which ensures that pattern path atoms constrained to certain-sized rings have been mapped to structure atoms which have the required ring character. The allowed ring sizes are specified in the TC knowledge-base entry as special sets which appear after the 2-D pattern ((4) in Figure 1). Although rings may be coded explicitly in the 2-D pattern language, explicit rings can be of only one size. Many TCs, however, operate on structures with rings of variable size. Thus, the RING*X*Y label followed by a list of pattern path atoms and bonds constrained to a ring of size X through Y is used to specify variable ring sizes. This system permits

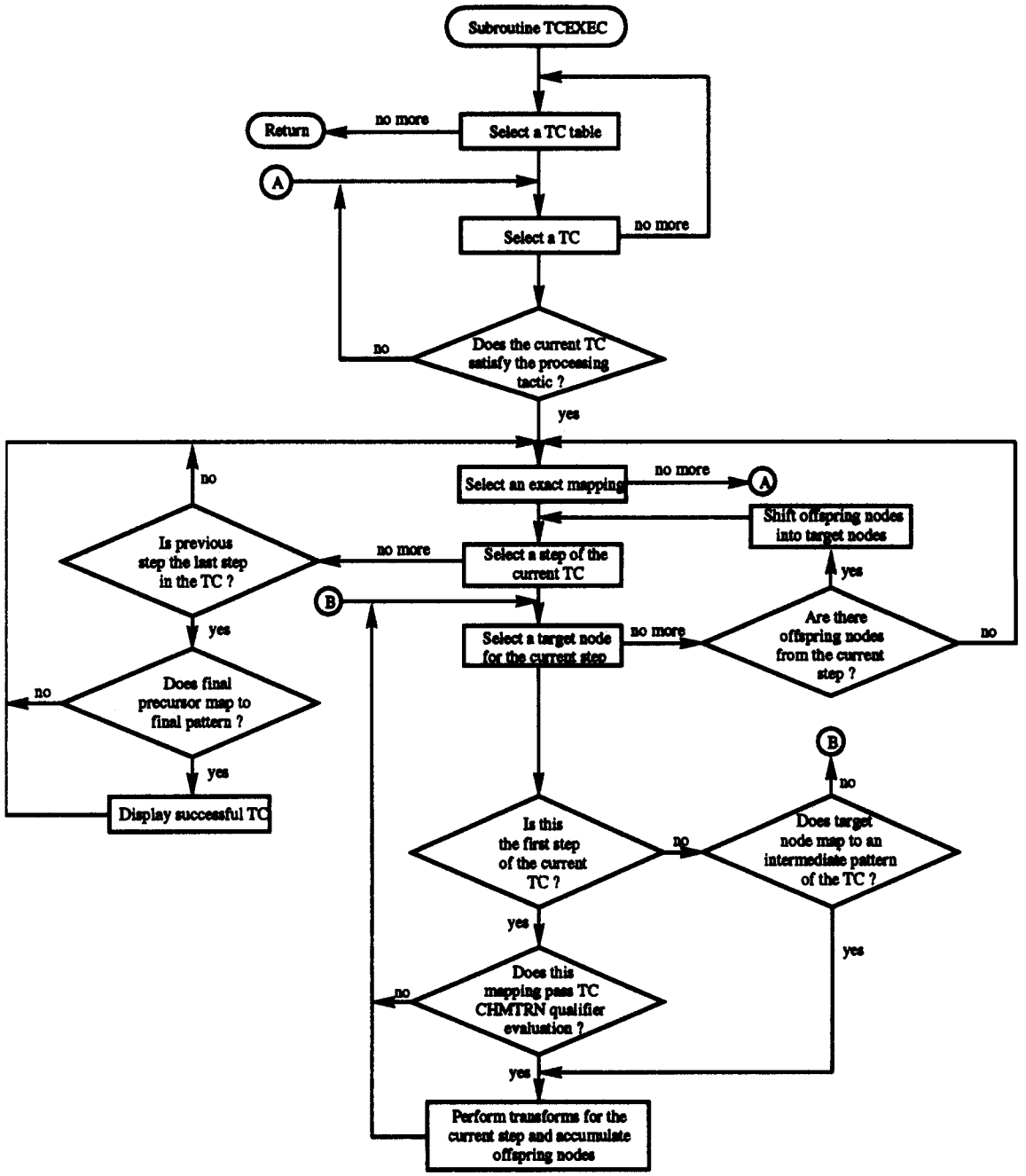


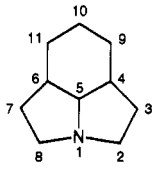
Figure 2. Simplified flow chart of subroutine TCEXEC.

screening of TCs based on cyclic character during the mapping phase.

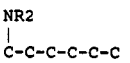
Second, each mapping must be analyzed to ensure that it will accomplish the processing tactic[16] specified by the chemist. As with the transform executives, TCEXEC supports a range of tactical limitations on its selection of TCs. For example, screening TCs for the *C–C disconnective* or *unconstrained* tactics is accomplished by allowing only TCs marked with the DISCONNECTIVE cross-referencing label ((5) in Figure 1) or TCs not marked with the NON*OPPORTUNISTIC label, respectively. Similarly, processing under the *diastereoselective* or *enantioselective* tactics, or with stereocenters designated as "preserved",[17] relies on INVERTS*STEREO, RETAINS*STEREO, RACEMIZES*STEREO, and REMOVES*STEREO cross-referencing labels for TC selection.

The other goal transform executives (GPAIR, GSING, and PEXEC) in LHASA[14] use BROKEN*BOND cross-referencing labels when screening for topological tactics. TCEXEC, on the other hand, accumulates in a set all bonds disconnected during a TC sequence, using subroutine TC_BREAK_STRAT. For example, in Figure 1, the first two patterns indicate that the bond spanning path atoms 3 and 7 is disconnected, while the second pair of patterns signals disconnection of the fifth path bond. The target structure bonds corresponding to the broken pattern bonds are compared to the set of strategic bonds chosen by the processing tactic (e.g., cyclic strategic, polyfused strategic, ring appendages, or manually designated strategic bonds) to determine whether the current TC satisfies the tactic.

Third, each mapping that passes tactical screening enters the TC for evaluation of CHMTRN qualifiers ((6) in Figure 1).[3] These qualifiers fine-tune the selection process by posing questions which reflect the scope and limitations of the TC under consideration. In general, TC qualifiers ensure the presence or absence of molecular features that can be specified neither in the 2-D pattern nor in special sets. For instance, these qualifiers might check that interfering molecular features



Target Structure



Retron 2-D Pattern

Mappings identified:

Path Positions

[1]	[2]	[3]	[4]	[5]	[6]	
2	3	4	5	6	7	Alternative at position 6: 11
2	3	4	5	6	11	Alternative at position 4: 9
2	3	4	9	10	11	Alternative at position 1: 5
5	4	9	10	11	6	Alternative at position 2: 6
5	6	11	10	9	4	Alternative at position 1: 8
8	7	6	5	4	3	Alternative at position 6: 9
8	7	6	5	4	9	Alternative at position 4: 11
8	7	6	11	10	9	Alternative at position 4: 11

Figure 3. Backtracking phase of the mapping process. If structure atoms 2, 5, and 8 are mapped in turn to the anchor atom of the 2-D pattern (the first path atom), then eight mappings are possible. After the first mapping (2, 3, 4, 5, 6, 7) is grown, alternatives are considered for the last position mapped (i.e. [6]). When alternatives for this position are exhausted, the next-to-last position is permuted. This procedure considers all correlations between pattern path atoms and target structure atoms in order to discover all mappings.

are absent or that certain ring sizes, fusions, or bridges are present. Once qualifier evaluation is successfully completed, the TC has been fully keyed and is ready to be executed.

C. Transform Selection for Each TC Step. The mechanism by which transforms are selected for a TC step relies heavily on the ability to perceive differences between 2-D patterns, as described in section III. A brief discussion of the aspects pertinent to TC processing follows. When a tactical combinations knowledge-base table is compiled offline, the atom and bond changes between each pair of consecutive patterns are perceived. These molecular differences are converted to numerical values, which are then canonicalized to a single integer, or "change code". This integer code uniquely describes the changes required by that TC step, independent of the order in which the pattern is drawn, the position of substituents, etc. After compilation, a binary output file is produced which includes information gleaned from each TC pattern and the change code for each TC step. A similar process is followed when a transform knowledge-base table is compiled offline. Since each transform contains a 2-D pattern depicting its retron and precursor substructures, a change code can be generated for each transform and placed in the binary output file from the CHMTRN compiler. Comparison of a change code for a TC step with the codes in the transform knowledge base provides the mechanism for identifying candidate transforms for TCs. This mechanism ensures that new transforms will be accessible to TCs without recompilation of any TC tables.

Upon entering TCEXEC, a call is issued to the routine FILL_CODE_ARRAY. FILL_CODE_ARRAY opens the 2-D pattern binary output file from each transform table, transcribes the change codes for all transforms into an array, and sorts the array[18] into ascending order. Then, after a TC has been fully keyed and is being executed, the change code for a step is unpacked from the TC binary file and a call to the routine FIND_TFS is issued to locate candidate transforms. FIND_TFS performs a binary search in the transform code array for transforms whose change codes match that of the current TC step. The set of candidate transforms is returned to TCEXEC and TCEXEC proceeds to execute them.

D. Executing Tactical Combinations. The TC execution process occurs in three phases. First, TCEXEC calls transform executives to perform candidate transforms for a TC step. GPAIR, GSING, and PEXEC are called in a mode which instructs them to perform only requested transforms. To execute functional group interchange (FGI) and functional group addition (FGA) transforms, the subgoal executives SNGFGI and SNGFGA are employed. Unlike goal transform executives, subgoal executives cannot be called to execute particular transforms. Instead, subgoal executives require information on the retron and precursor functional groups that are to be interchanged or the precursor functional group that is to be added. Thus, two subroutines, TCFGI and TCFGA, were written to act as interfaces between TCEXEC and the subgoal executives. These interface routines examine the 2-D pattern of a TC step to determine which precursor functional groups do not appear in the retron and therefore must have arisen from the FGI or FGA. On the basis of this 2-D pattern analysis, TCFGI and TCFGA are able to formulate calling arguments to SNGFGI and SNGFGA.

The second phase of TC execution consists of an analysis of the precursor nodes returned from the transform executives. The nodes generated from the execution of each TC step are accumulated in a binary set, and each node must be checked to ensure that the requested precursor was generated before the next step is attempted. This checking is accomplished by performing the 2-D pattern mapping operation between each precursor structure and the appropriate precursor pattern of the TC. Furthermore, duplicate nodes must be eliminated, since their continued processing would slow execution without contributing information to the analysis. Duplicate precursors are weeded by the subroutine TC_WEED_DUPS, which prunes a group of identical precursors, retaining only the one resulting from the highest-rated transform.

Finally, after execution of a TC has succeeded in generating final precursors, the routes must be examined for similarity. For instance, when executing the first step of the TC depicted in Figure 1, the program will generate two precursors from the Ene transform, differing only in olefin stereochemistry. Since these structures are not identical, each will be retained and processed to complete the TC. Two very similar sequences will result, and to avoid cluttering the retrosynthetic tree with close-relative routes, one route must be pruned. The subroutine TC_WEED_RELS is called by TCEXEC to accomplish this pruning. TC_WEED_RELS analyzes every pair of sequences in turn to compare the transforms used for each step. At the same time, this subroutine accumulates the total rating for each sequence. If two sequences were found to employ the same transforms for each step, the lower-rated route is pruned.

V. RESULTS

Although originally conceived for LHASA as directed subgoal sequences, TCs have proved useful as goal sequences for opportunistic and tactically-guided processing. Presently, operating with a knowledge base of 450 TCs, the program can produce intellectually provocative routes to complex molecules. In many cases, the machine-generated routes are comparable, on a theoretical plane, to published syntheses. A sample LHASA analysis employing TCs is included in Figure 4. The target, Homogynolide-B (**5**), is a member of the Bakkane family of hydrindane natural products which has recently yielded to total synthesis.[19] The TC package in LHASA suggests a number of routes to Homogynolide-B which are worthy of merit as alternatives to the published synthesis.

VI. CONCLUSION

With the facility for executing tactical combinations now in place, it is interesting to contemplate possible uses of TCs

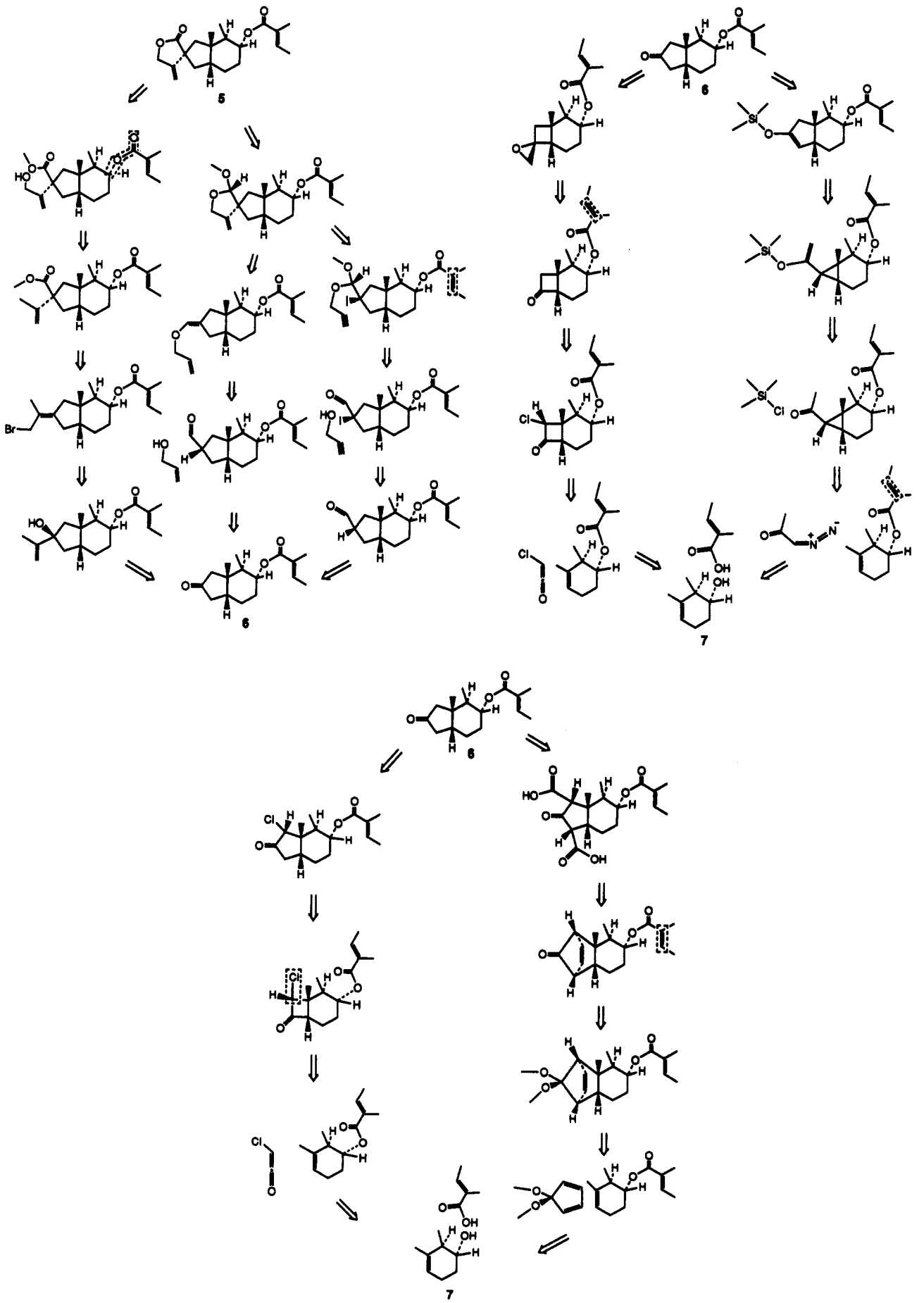


Figure 4. Sample retrosynthetic pathways generated for Homogynolide-B (5) by the TC package in LHASA.

and extensions to the current capabilities. The ability of TCs to effect powerful retrosynthetic simplification and the present use of TCs as goal sequences accessed opportunistically or by tactical guidance suggest that TCs should have access to considerable subgoal power. Furthermore, since TCs are based on the 2-D pattern system of encoding substructures and describing differences between substructures, a method of subgoal transform selection based on 2-D patterns is a logical extension. In fact, one can envision a system which perceives the atom and bond differences between the target structure entered by the user and the retron pattern of a goal TC. A change code generated from this perception would describe the necessary subgoal step and permit identification of candidate subgoal transforms. An obvious advantage of this system over functional group-based subgoal systems[12] is that subgoal transforms could be selected from anywhere in the LHASA transform knowledge base. Thus, the artificial distinction between "goal" and "subgoal" transforms would vanish. Such a system has been implemented in LHASA for single-step rectification of mismatches between target structures and TC retrons.[9] Extension to multiple subgoal steps, a considerably more complicated problem, has also been accomplished and will be reported shortly.

Beyond the use of TCs as goal sequences, TCs are expected to be useful as subgoal sequences, since they achieve retrosynthetic objectives in such a directed fashion. Since TCs are coded with a precursor pattern for each step, the overall atom and bond changes effected by a TC may be perceived by comparing its first and last 2-D patterns. The resulting change code for the entire sequence can then be used when attempting to identify appropriate subgoal TCs.

Finally, continued expansion of the LHASA TC knowledge base is anticipated. The 450 TCs compiled thus far represent only a fraction of the TCs that can be derived from the chemical literature. In fact, a more complete tabulation could eventually result in a dictionary of TCs. Such a lexicon would prove invaluable to synthetic chemists designing routes to complex molecules.

ACKNOWLEDGMENT

We are grateful to the National Institutes of Health for financial assistance. The coding contributions of previous members of the LHASA group, especially Stewart Rubenstein, Gary Walker, and Martin Ott, are greatly appreciated.

REFERENCES AND NOTES

(1) (a) LHASA is an acronym for Logic and Heuristics Applied to Synthetic Analysis. (b) The program is implemented on Digital Equipment Corp.'s VAX computers under the VMS operating system and on a variety of computers running Unix. (c) Long, A. K.; Rubenstein, S. D.; and Joncas, L. J. A Computer Program for Organic Synthesis. *Chem. Eng. News* **1983**, *61* (19), 22. Corey, E. J.; Long, A. K.; and Rubenstein, S. D. Computer-Assisted Analysis in Organic Synthesis. *Science* **1985**, *228*, 408, and references cited therein.

(2) Corey, E. J.; Long, A. K.; Lotto, G. I.; Rubenstein, S. D. Computer-Assisted Synthetic Analysis. Quantitative Assessment of Transform Utilities. *Recl. Trav. Chim. Pays-Bas* **1992**, *111*, 304.

(3) For a detailed discussion of the CHMTRN language, see: Orf, H. W. Ph.D. Thesis, Harvard University, 1976.

(4) LHASA also has a new mode of operation in which the program uses information gained during its perception phase to suggest strategies and tactics to the user.

(5) Johnson, A. P.; Marshall, C.; Judson, P. N. Starting Material Oriented Retrosynthetic Analysis in the LHASA Program. General Description. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 411. Johnson, A. P.; Marshall, C. Starting Material Oriented Retrosynthetic Analysis in the LHASA Program. Mapping the SM and Target Structures. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 418. Johnson, A. P.; Marshall, C. Starting Material Oriented Retrosynthetic Analysis in the LHASA Program. Heuristic Estimation of Synthetic Proximity. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 426.

(6) Corey, E. J.; Howe, W. J.; Pensak, D. A. Computer-Assisted Synthetic Analysis. Methods for Machine Generation of Synthetic Intermediates Involving Multistep Look-Ahead. *J. Am. Chem. Soc.* **1974**, *96*, 7724.

(7) Corey, E. J.; Howe, W. J.; Orf, H. W.; Pensak, D. A.; Petersson, G. General Methods of Synthetic Analysis. Strategic Bond Disconnections for Bridged Polycyclic Structures. *J. Am. Chem. Soc.* **1975**, *97*, 6116.

(8) Corey, E. J.; Jorgensen, W. L. Computer-Assisted Synthetic Analysis. Generation of Synthetic Sequences Involving Sequential Functional Group Interchanges. *J. Am. Chem. Soc.* **1976**, *98*, 203.

(9) Long, A. K.; Kappos, J. C.; Rubenstein, S. D.; Walker, G. E. Computer-Assisted Synthetic Analysis. A Generalized Procedure for Subgoal Transform Selection Based on a Two-Dimensional Pattern Language. *J. Chem. Inf. Comput. Sci.*, following paper in this issue.

(10) Corey, E. J.; Cheng, X.-M. *The Logic of Chemical Synthesis*; Wiley Interscience: New York, 1989; p 31. Long, A. K.; Kappos, J. C.; Nevalainen, V.; Stamp, L. An Extension to the Theory of Retrosynthetic Analysis: "Tactical Combinations" of Transforms. Submitted for publication.

(11) Stork, G.; Danheiser, R. L. The Regiospecific Alkylation of Cyclic β Diketone Enol Ethers. A General Synthesis of 4-Alkylcyclohexenones. *J. Org. Chem.* **1973**, *38*, 1775. Pirrung, M. C. Total Synthesis of (±)-Isocomene and Related Studies. *J. Am. Chem. Soc.* **1981**, *103*, 82. Pirrung, M. C. Total Synthesis of (±)-Isocomene. *J. Am. Chem. Soc.* **1979**, *101*, 7130.

(12) Corey, E. J.; Cramer, R. D., III; Howe, W. J. Computer-Assisted Synthetic Analysis for Complex Molecules. Methods and Procedures for Machine Generation of Synthetic Intermediates. *J. Am. Chem. Soc.* **1972**, *94*, 440.

(13) For a detailed discussion of the one-dimensional pattern language, see: Hoyle, P. L. M. Ph.D. Thesis, University of Leeds, 1986.

(14) GPAIR, GSING, and PEXEC are the executive subroutines which guide the performance of transforms keyed by pairs of functional groups, single functional groups, and one-dimensional patterns, respectively. The order in which these subroutines are executed varies with the target structure being processed. For example, if the chemist enters a target which contains only aromatic or heteroaromatic rings, PEXEC will be executed first. TCEXEC is always executed last, regardless of the target structure.

(15) Tactical combinations processing is guided largely by code written in FORTRAN and C. There are approximately 3100 executable lines of code organized into 32 subroutines which are dedicated to TC processing.

(16) Each retrosynthetic "strategy" in LHASA provides a number of alternative substrategies, or "tactics", to the user. Under the "short-range" strategy, tactics are *unconstrained*, *C–C disconnective*, *nondisconnective*, *reconnective*, *carboaromatic*, and *heteroaromatic*. Under the "topological" strategy, tactics are *cyclic strategic*, *polyfused strategic*, *acyclic strategic*, *strategic bond pairs*, *ring appendages*, and *manual designation*. Transform entries in the knowledge base are labeled, or cross-referenced, according to the tactics that can use them. For example, only transforms labeled DISCONNECTIVE are available to the *C–C disconnective* tactic.

(17) "Preserved stereocenters" are those whose configurations must be kept unchanged during transform execution, either because they are part of a fragment that exists in a designated starting material or because the chemist has specifically selected them for preservation.

(18) The procedure used for sorting the array of change codes is based on the Shell-Metzner algorithm, "A Comparison of Sorts". *Creative Comput.* **1976**, *2*.

(19) Greene, A. E.; Coelho, F.; Depres, J.-P.; Brocksom, T. J. Direct Approach to the Bakkanes: A Synthesis of (±)-Homogynolide-B. *Tetrahedron Lett.* **1989**, *30*, 565.