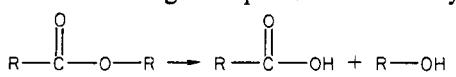
bond indexes are summed over all bonds in a molecule to give the molecular connectivity index, $^1\chi^V$. The reaction connectivity index, $\Delta^1\chi^V$, is $^1\chi^V$(product) minus $^1\chi^V$(reactant).

It is apparent that inclusion or omission of the carbonyl group in both structures of the model is immaterial, since the index value would be the same. It is desirable to include the carbonyl group since the chemist associates this feature with an ester and an acid.

By convention, the bond indexes are computed to four decimal places and summed to three places. The $\Delta^1\chi^V$ values are computed to three places with a plus or minus sign.

It is possible to calculate a reaction index from a general reaction model including both products of ester hydrolysis:



The same index of opposite sign would encode the general reaction of ester formation from alcohol and acid. This single number, computed from the general reaction model, is the code under which information may be stored or retrieved for the event described by the model.

It is not possible to claim that every general reaction will have a unique reaction connectivity index. Our intuition is that redundancies would be rare and that "false drops" in the retrieval process would probably be minimized when terms appropriate to the query were coordinated with the connectivity value.

Variations in the adoption of the general reaction model could be introduced to permit a more detailed refinement of reaction classification. Thus, entry 5 in Table II describes acylation of any amine. The amine is nonspecified by using two R symbols for carbon or hydrogen atoms. A more specific index may be calculated for an amine acylation reaction involving only primary amines. This would have a calculated reaction connectivity index of 0.877. This could index a separate file or could be a subfile under the general acylation index of 0.822.

Another variant might be the use of a reaction connectivity index computed from a general reaction model reflecting two products, as, for example, ester hydrolysis to both acid and alcohol. From such a general reaction model, a $\Delta^1\chi^V$ would be computed.

CONCLUSION

As described above, the value for a specific reaction type can be quickly calculated; it is entirely dependent upon the structure; it is universal for all reactions of the same type; it is sufficiently different from the values of other reactions to provide an unambiguous identification.

We believe the generation of a reaction index file, for use with large chemical information and/or data files, would greatly enhance the ability of the information specialist to refine retrieval. The appropriate value(s) for different reaction(s) could be incorporated as a searchable field in computer-accessible files just as the CAS Registry number has been incorporated.

REFERENCES AND NOTES

(1) Gluck, D. J. "A Chemical Structure Storage and Search System Developed at Du Pont", *J. Chem. Doc.* **1965**, *5*, 43-51.
(2) Beach, A. J.; Dabek, H. F., Jr.; Hosansky, N. L. "Chemical Reactions Information Retrieval from Chemical Abstracts Service Publications and Services", *J. Chem. Inf. Comput. Sci.* **1979**, *19*, 149-55.
(3) Valls, J. In "Computer Representation and Manipulation of Chemical Information"; Wipke, W. T., et al., Ed.; Wiley, New York, 1974.
(4) Lynch, M. F.; Harrison, J. M.; Town, W. G.; Ash, J. E., Eds. "Computer Handling of Chemical Structure Information"; Macdonald and American Elsevier, New York, 1971.
(5) Gelberg, A. J. "Rapid Structure Searches via Permuted Chemical Line Notations. IV. A Reactant Index", *J. Chem. Doc.* **1966**, *6*, 60-1.
(6) Weygand, C. "Organisch-chemische Experimentierkunst", 4th ed.; Barth, Leipzig, 1970.
(7) Theilheimer, W. "Synthetic Methods of Organic Chemistry", 26 volumes; Karger, Basel and New York, 1946-1972.
(8) Vleduts, G. E.; Mishchenko, G. L. *Tr. Vses Konf. Inf. Poisk. Sist. Autom. Obrab. Nauchno-Tekh. Inf., 3rd, 1966*; *Chem. Abstr.* **1969**, *70*, 43991m.
(9) Hudrlik, P. F. "Reaction Index" in "Survey of Organic Synthesis", Buehler, C. A.; Pearson, D. E., Eds.; Wiley: New York, 1977; pp 1001-18.
(10) Hendrickson, J. B. "A Systematic Organization of Synthetic Reactions", *J. Chem. Inf. Comput. Sci.* **1979**, *19*, 129-36.
(11) Ziegler, H. J. "Roche Integrated Reaction System (RIRS). A New Documentation System for Organic Reactions", *J. Chem. Inf. Comput. Sci.* **1979**, *19*, 141-9.
(12) Bersohn, M.; MacKay, K. "Steps toward the Automatic Compilation of Synthetic Organic Reactions", *J. Chem. Inf. Comput. Sci.* **1979**, *19*, 137-41.
(13) Willett, P. "Computer Techniques for the Indexing of Chemical Reaction Information", *J. Chem. Inf. Comput. Sci.* **1979**, *19*, 156-8.
(14) Kier, L. B.; Hall, L. H. "Molecular Connectivity in Chemistry and Drug Research"; Academic Press, New York, 1976.

Computer-Assisted Synthetic Analysis. Long-Range Search Procedures for Antithetic Simplification of Complex Targets by Application of the Halolactonization Transform

E. J. COREY,* ALAN K. LONG, JOHANN MULZER, HARRY W. ORF, A. PETER JOHNSON,† and ALAN P. W. HEWETT

Department of Chemistry, Harvard University, Cambridge, Massachusetts 02138

Received July 23, 1980

A major problem in computer-assisted synthetic analysis is the development of techniques for long-range searches directed toward retrosynthetic simplification of a complex target. The approach used in the Harvard program, LHASA, combines multistep antithetic analysis with an efficient method for prescreening targets to eliminate less promising routes and to limit the number of precursors generated. These techniques are illustrated for the halolactonization search in LHASA, and the method for preevaluation of pathways is described in detail. A number of chemical examples are included.

The most elegant and simplest chemical routes for the synthesis of complex molecules are to be found by approaches which combine several lines of analysis so as to allow the concurrent application of several powerful strategies. This collection of strategies generally includes the overarching and most general principles of synthetic design such as the rigorous application of antithetic (retrosynthetic) search, coupled antithetic–synthetic tree generation, the use of self-reinforcing cycles of perception and analysis, and the characterization of unique features of the problem or target structure. In addition,

† Department of Organic Chemistry, The University, Leeds LS2 9JT, England.

0095-2338/80/1620-0221$01.00/0

strategies that effectively identify problem-solving approaches which are ideally suited for a particular target structure are synergistic to lower level strategies of the hierarchy in that they permit more efficient and more powerful use of those strategies.

We have emphasized previously[1] the importance of the selection of one or more key transforms for possible application to a target structure and the use of the application of such transforms as a goal for long-range antithetic search. Frequently such key transforms correspond to especially powerful synthetic operations, for example, ring-forming reactions (Diels–Alder, Robinson annulation, internal acylation, internal S_N2, etc.), stereospecific reactions (Claisen rearrangement, S_N2, stereocontrolled $sp^2 \rightarrow sp^3$ carbon, ene reactions, etc.), and reactions which are position specific and stereospecific because they are rigidly controlled by mechanism in a way which is invariant with substrate. In this paper we demonstrate the manner in which the transform corresponding to one of these latter reactions, halolactonization, when used as a long-range goal for antithetic search, can lead to major simplifications of the target structure and open the way for application of further simplifying transforms, such as those capable of antithetic disconnection of rings. The combined search to apply two or more key transforms (T goals) synergistically or in a coordinated way is an important objective of this work. In addition to this coupled search for T goals, combinations involving structure-based goals, or S goals, are also important.[1d,2] Examples of such structure-based goals include the retrosynthetic disconnection of certain specific bonds or bond pairs in cyclic structures (strategic bonds)[2] so as to achieve maximal simplification of a network and the identification of structural subunits which can serve as building blocks for the synthesis.

For a variety of reasons we selected the halolactonization transform as the objective of the long-range search procedures which are the concern of this paper. First of all it is a powerfully simplifying transform since its application reduces stereochemical complexity, removes reactive functional groups, and generates readily accessible "core" functionality (C=C and COOH). Secondly, it is mechanism controlled, hence highly predictable as mentioned above; in the synthetic direction halolactonization results in stereodirected delivery of oxygen to carbon by the carboxyl function with concomitant trans addition of halogen. The process is generally position controlled as well, because of the preference for 5 over 6 over smaller or larger membered lactone rings as well as the tendency for Markovnikov addition. Thirdly, the application of the halolactonization transform often removes obstacles to the application of other powerfully simplifying transforms, allowing their use in a way which is at once more straightforward and successful. The application of the halolactonization reaction to the successful solution of a spectrum of complex problems testifies to both the power of the transform itself and the elegance with which it dovetails with other transform-oriented strategies. In our laboratory alone halolactonization has been used for the synthesis of prostaglandins,[3] thromboxanes,[4] gibberellic acid,[5] erythronolides A and B,[6] 5-HPETE,[7] leukotrienes A, C, D, and E,[8] and part of the ansa bridge of the rifamycins.[9,10]

THE SEARCH FOR HALOLACTONIZATIONS BY COMPUTER

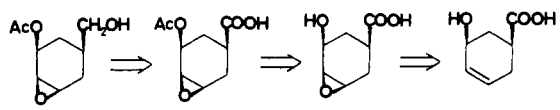
As indicated above, the stereochemical and regiochemical outcome of the halolactonization reaction is quite predictable. However, elaboration of the resulting halo lactones can result in a very wide variety of structures. The main challenge in constructing a search for halolactonizations by computer is to devise an efficient method for converting (retrosynthetically) this variety of structure types to the appropriate halo lactones. Clearly the search must be general enough to accommodate any possible target structure but at the same time specific enough to avoid generating a large number of minimally different antithetic pathways.

In order to obtain maximum benefit from the stereospecificity of the reaction, the search in LHASA has been restricted to those halolactonizations which control stereochemistry in carbocyclic rings of 5 to 7 members. The reaction has limited utility in smaller-ring substrates, and there is considerable stereochemical ambiguity for larger ring sizes. In addition, only halolactonizations which control the relative configurations at three on-ring atoms are taken as goals for the search.

At a preliminary stage in the design of the halolactonization search in LHASA, a number of common synthetic transformations of halo lactones were identified. These transformations resulted in a number of "generic intermediates", as shown in Scheme I for the specific example of γ-lactones formed from 6-membered-ring olefinic acids. Each generic intermediate contains a common substructure which is readily accessible from a halo lactone of the corresponding olefinic acid and represents a point at which several additional synthetic possibilities become available. Since only a limited number of these generic intermediates exist, it was reasonable to construct a retrosynthetic search around them. As mentioned above, structures like these generic intermediates fall into the category of structure goals, or S goals. Ten of these S goals have been selected as key intermediates for the halolactonization search. Accordingly, the halolactonization search table[11] is composed of ten "matching procedures", each of which performs a "substructure match" between the target and the appropriate S goal. The specific tasks of a procedure are twofold: to convert the target structure (antithetically) via a number of subgoal (Functional Group Interchange and Functional Group Addition) steps to the intermediate S goal and to convert the S goal in turn to the desired olefinic acid via the corresponding halo lactone. This latter task is straightforward, since each of the pathways has been worked out in the synthetic direction as described above (see Scheme I). The former task, that of antithetic conversion of the target to the S goal, is considerably more complicated, since a great diversity of possible target structures must be accommodated.

LOCALIZED MATCHING UNITS

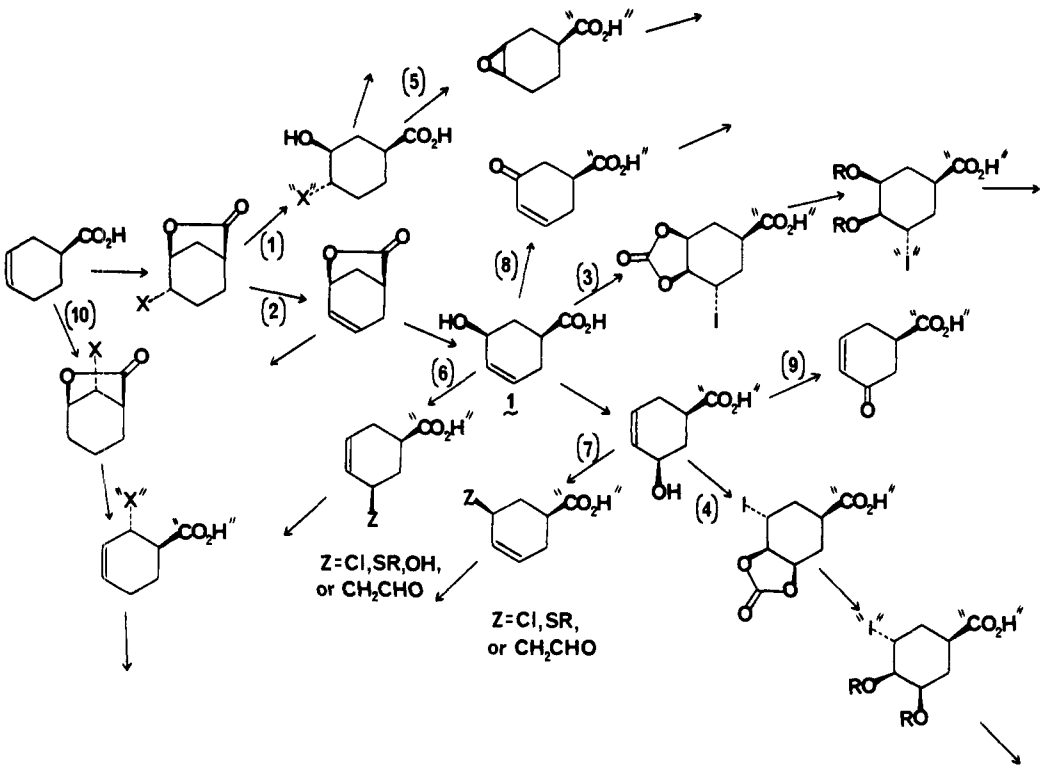
The method devised for antithetically transforming the target to a desired S goal is as follows. First, the salient features of the S goal substructure are identified. In the case of S goal **1** (see Scheme I), for example, the substructure consists of a carboxyl group, a hydroxyl, and a carbon–carbon double bond. Next, these elements of the S goal are correlated with structural subunits in the target molecule. Finally, the transforms necessary to perform the process of antithetically converting these subunits, termed "localized matching units" (LMU's), to the functionality in the S goal are requested and performed:[12]



CHEMISTRY SUBROUTINES

As is evident from the sample S goals in Scheme I, many of them share common structural subunits. Accordingly, the CHMTRN code[11] responsible for performing the antithetic conversions to these subunits has been organized into a number of "chemistry subroutines". Each subroutine is specific for a single type of LMU (one, two, or three carbon) and always results in either a failure return or in generation of the re-

Scheme I. Synthetic Transformations of Generic Intermediates Used in the LHASA Search for γ-Lactones of Six-Membered-Ring Olefinic Acids[a]



[a] Procedure numbers are indicated in parentheses.

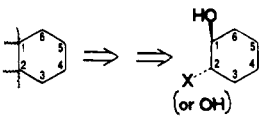
quested S-goal functionality. Chemistry subroutines are independent of the procedure (or other subroutine) which calls them and are thus capable of being used in any search table. For example, many of the chemistry subroutines written for the halolactonization search have also been used in the Robinson Annulation table.[1d]

The most important chemistry subroutines in the halolactonization search table are as follows:[13]

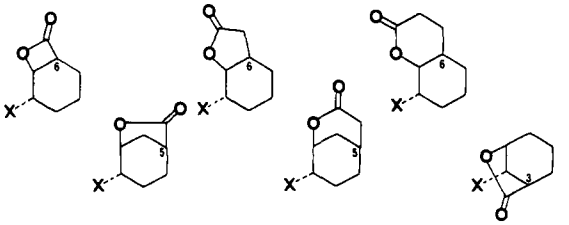
(a) GET*CO: generates a carbonyl at a 1C LMU.
(b) GET*OH: generates a hydroxyl at a 1C LMU.
(c) GET*EPOXIDE: generates an epoxide at a 2C LMU.
(d) GET*DB: generates a C=C at a 2C LMU.
(e) SUPER*GET*DB: generates a C=C at a 2C LMU in cases where GET*DB fails because of regiochemical problems by using neighboring carboxyl groups to assist in double bond addition transforms.
(f) DEALKYLATE: removes appendages α to C=O by dealkylation (2C LMU).
(g) GET*ENONE: removes appendages from positions α and/or β to C=O by conjugate addition/alkylation (3C LMU).
(h) RING3: removes appendages from positions α and β to C=O by cyclopropanation/alkylation (3C LMU).
(i) CUPRATE: checks appendages for functionality corresponding to a legitimate cuprate reagent.
(j) CLAISEN: performs allylic transpositions.
(k) CARBONATE: generates an allylic alcohol by carbonate lactonization (3C or 4C LMU).[14]
(l) GET*ACID: generates the COOH group (1C LMU).
(m) GET*LACTONE: calls GET*OH and GET*ACID, then antithetically reconnects the two groups to a lactone.
(n) DO*HALAC: performs the actual halolactonization or hydroxylactonization transform.

HALOLACTONIZATION PATHS

One of the advantages of using a computer program to assist the chemist in finding halolactonization is that no reasonable possibilities are overlooked. Each five- to seven-membered ring in the target molecule can be treated in turn, and for each ring, all the ring atoms can be examined as hydroxyl LMU's, halide LMU's, etc. To this end, LHASA grows "paths" around the current ring, labeling the atoms in the ring as atom 1, atom 2, etc. For the halolactonization search, atom 1 is always considered to be the LMU corresponding to the hydroxyl of the lactone. Similarly, atom 2 is the LMU for the halide (or the non-lactone hydroxyl in a hydroxylactonization):



Once a path has been selected, LHASA searches for atoms which could serve as suitable "carboxyl equivalents", or COOH LMU's, in the vicinity of atom 1 of the path. More explicitly, valid carboxyl equivalents for the path shown above would be found α to atom 6 off-ring, β to atom 6 off-ring, γ to atom 6 off-ring, α to atom 5 off-ring, β to atom 5 off-ring, and α to atom 3 off-ring, corresponding to the following six halo lactones:



PROCEDURE SELECTION

In principle, it would be possible to have the computer program examine each path/carboxyl equivalent combination for each path around the current ring. However, since there

are 12 paths for a six-membered ring (six "orientations" of the ring and two "directions" around it), 2 "faces" on which the lactone hydroxyl might be placed, 6 possible COOH LMU's for each path/face combination, and 10 procedures, there is a potential for 12 × 2 × 6 × 10 = 1440 halolactonization sequences for a target containing a six-membered ring. Usually this combinatorial problem is obviated by the limited number of COOH LMU's, but for highly functionalized rings with many appendages (the type of target for which the halolactonization search was in fact designed) the large number of sequences found becomes quite burdensome.

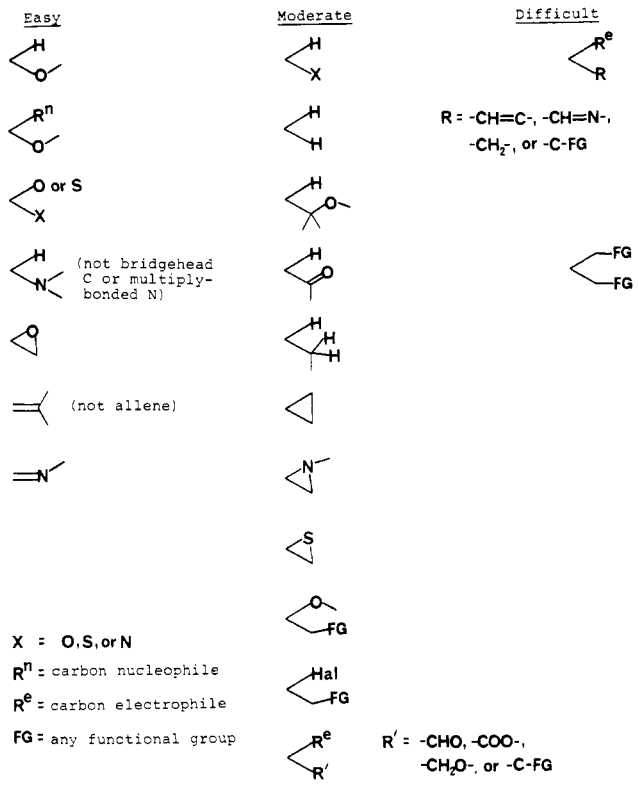
The approach to procedure selection used in LHASA is based on the fact that each of the procedures is simply a collection of chemistry subroutines. Since chemistry subroutines operate on small areas of the target, it is quite easy to make an assessment of how efficient a chemistry subroutine will be in almost any given situation. A "subroutine rating" may be assigned and an overall "procedure rating" calculated by summing the subroutine ratings over the particular path being analyzed. The result of the application of this approach (described in detail below) is that for a given target molecule a numerical rating is computed for each of the available procedure/path combinations, with lower ratings corresponding to more efficient procedures. Once all the procedures have been rated in this fashion, it is possible to execute only those procedures which the a priori evaluation has judged to be particularly suited to the structure of the target molecule. This approach reduces the inherent inefficiency of an exhaustive synthetic analysis by pinpointing those locations where good solutions are likely to be found.

PRIOR PROCEDURE EVALUATION

The above-mentioned a priori evaluation of procedures in LHASA is called Prior Procedure Evaluation (PPE). The PPE routine calculates a numerical rating for a procedure which is roughly equivalent to the number of chemical reactions likely to be involved.[15] Since each procedure rating is simply a sum of ratings for individual chemistry subroutines, the task of PPE is reduced to an assignment of ratings for the chemistry subroutines comprising the procedures. For most chemistry subroutines, the rating is assigned on the basis of an estimate of the facility with which functionality and appendages present in a particular portion of the target can be antithetically converted to the structural subunit which is the goal of the subroutine. This technique is, however, not appropriate to subroutines which operate only at the latest stages of a procedure, at which point the target may have been extensively modified, for example by a functional group transposition. Fortunately the number of steps involved in such subroutines does not vary greatly from one case to another, and it is sufficient to assign a standard rating for all such cases. The details of the subroutine evaluation are described below.

Subroutine Evaluation. (a) Categorization of Localized Matching Units. A table is prepared for each chemistry subroutine listing (i) all the types of structural subunits which the subroutine can handle (take back to its goal) and (ii) structural features whose presence would automatically cause the subroutine to fail. Within each table the structural units to be processed by the subroutine are classified "there", "easy", "moderate", "difficult", or "impossible". In the there category the unit being analyzed corresponds to the object of the subroutine. The easy category is associated with straightforward processes of one to two steps, the moderate category with three to four step sequences, and the difficult category with sequences involving five or more steps in going from the target LMU to the goal of the subroutine. The ranking of subunits used in the table for the GET*CO subroutine is shown in Table I.

Table I. Categorization of Subunits for the GET*CO Subroutine



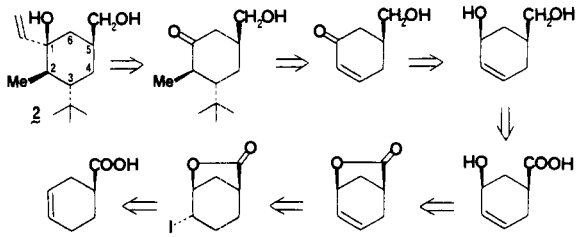
Impossible: Carbon origin is part of an onpath multiple bond or small ring.

Drawing up this kind of table for those chemistry subroutines whose field of operation is a single onpath carbon atom is a comparatively simple process. The analogous tables for subroutines involving two onpath carbon atoms (subroutines GET*DB and GET*EPOXIDE, for example) are much more complicated, because it is the *combination* of features (including stereochemistry) at both onpath atoms which determines the category to which the structural subunit belongs. When the subroutine involves three or more onpath atoms (such as the GET*ENONE subroutine), the number of different combinations is so great that this approach becomes unwieldy and an alternative approach must be used. Fortunately in all such cases it is possible to break down the main structural unit into smaller units and have a separate table for each of these smaller units. Thus the GET*ENONE subroutine gives rise to three tables: ENONB for the β carbon, ENONA for the α carbon in a GET*ENONE process, and GET*CO for the atom which is to become the ketone group. Clearly it is necessary to inspect all three tables in order to estimate how readily a particular three-carbon unit could be retrosynthetically converted to an α,β enone by the GET*ENONE subroutine.

It should be stressed that the rankings in these tables are based on the chemical content of the relevant chemistry subroutines and consequently may have to be updated whenever a chemistry subroutine is substantially modified.

(b) Creation of Categorization Sets. The information derived from the above-mentioned categorization is most conveniently stored in set format. For example, if atom 25 in a target was found to have a substitution pattern corresponding to the "easy" GET*CO category, bit 25 would be turned on in the computer word corresponding to the "easy" GET*CO set.

In principle, all of the necessary sets could be filled by a



PROCRAT = SR*GET*CO 1 + SR*ENONA 2 + SR*ENONB 3 + BOXRAT + #2
= 1 + 1 + 1 + 1 + 2 = 6

Figure 1. Calculation of the procedure rating PROCRAT for processing of target **2** by procedure 8 with the halolactonization path as indicated. See text for further details.

binary search table which looked in turn at each atom or bond, binary search table which looked in turn at each atom or bond, precisely identifying the nature of the particular substructural unit. While straightforward, this serial atom-by-atom (or bond-by-bond) filling of the sets is much less efficient than an alternative method in which all the appropriate atoms in the target are added to a specific set simultaneously. These set operations can be performed in the halolactonization search table itself, thanks to recent extensions to the CHMTRN data table language. The five categories, there, easy, moderate, difficult, and impossible, are handled using four sets for each subroutine. The there set is obtained from basic perception, and after the easy, moderate, and difficult sets are filled, any remaining atoms are taken to be of the impossible type.

Computation of Subroutine Ratings. For any given target it is likely that a number of procedures will call the same chemistry subroutine at a particular location. Since it makes little sense to keep computing the same value for the rating of a particular subroutine at a given location, ratings for each subroutine at all the sites at which it might be called are evaluated and stored. Again, extensions to the CHMTRN language which permit a variety of arithmetic operations to be performed allow these computations to be carried out in the search table itself rather than in FORTRAN code. As a result of similar evaluations of other chemistry subroutines at each of the sites at which each one might be called, a total of 54 subroutine ratings are eventually obtained and stored until needed.

Computation of Procedure Ratings. After all the chemistry subroutines have received ratings, computation of the rating for any given procedure is simply a matter of summing the component subroutine ratings. In target **2** (with the ring path numbered as shown in Figure 1), for example, the procedure rating (PROCRAT) for procedure 8 would be 6, since atom 1 is in the GET*CO*EASY set, atom 2 is in the ENON*ALPHA*EASY set, atom 3 is in the ENON*BETA*EASY set, the carboxyl equivalent at α to atom 5 offpath is in the GET*COOH*EASY set, and there are two additional steps performed in the procedure itself (functional group interchanges of C=O ⇒ secondary OH and of C=C ⇒ I). Note that the lactone hydrolysis step and the actual halolactonization are not counted in the procedure rating. These steps are common to all the halolactonization sequences generated by LHASA except those for targets in which the lactone still exists.

In addition to the numerical rating obtained as described above, the evaluation process for each procedure includes some checks to ensure that there are no structural elements present in the target which would make the procedure inappropriate for that particular target. In most cases these checks are made at atoms outside the field of operation of the relevant chemistry subroutine, and it is for this reason that they cannot be included in the initial subroutine evaluation. For example, procedure 8 is killed by PPE if there is a leaving group on atom 4 or on the atom α on-ring to atom 1, since such a group would interfere with additions to the enone which is the S goal for the procedure.

Procedure Evaluation for Alternative Ring Orientations. For a given ring it makes sense to evaluate all procedures at all possible orientations before actually displaying any sequences to the chemist, since one orientation assignment (path) may prove much more appropriate than any other. Procedure evaluation for other orientations is a relatively simple process since much of the required information has already been obtained during the first evaluation process. In particular, as shown in the example below, the ratings for the chemistry subroutines are simply taken from the values which were previously computed for the same atoms and bonds (which are now numbered differently) in the first path assignment.

orientation 2	orientation 1
(atom 1 was atom 2 in orientation 1)	

for $m = 1$ to $m = 5$ SR*GET*CO (m) = SR*GET*CO ($m + 1$)
SR*GET*CO 6 = SR*GET*CO 1

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/face/procedure combination and examines the next one. When all the procedures have been rated for all the path/face combinations, the 20 lowest ranked are sorted by rating and attempted in order by the program. As a result of this very efficient prescreen,[16] only the best procedures are actually displayed to the chemist.

LIMITATIONS OF PRIOR PROCEDURE EVALUATION

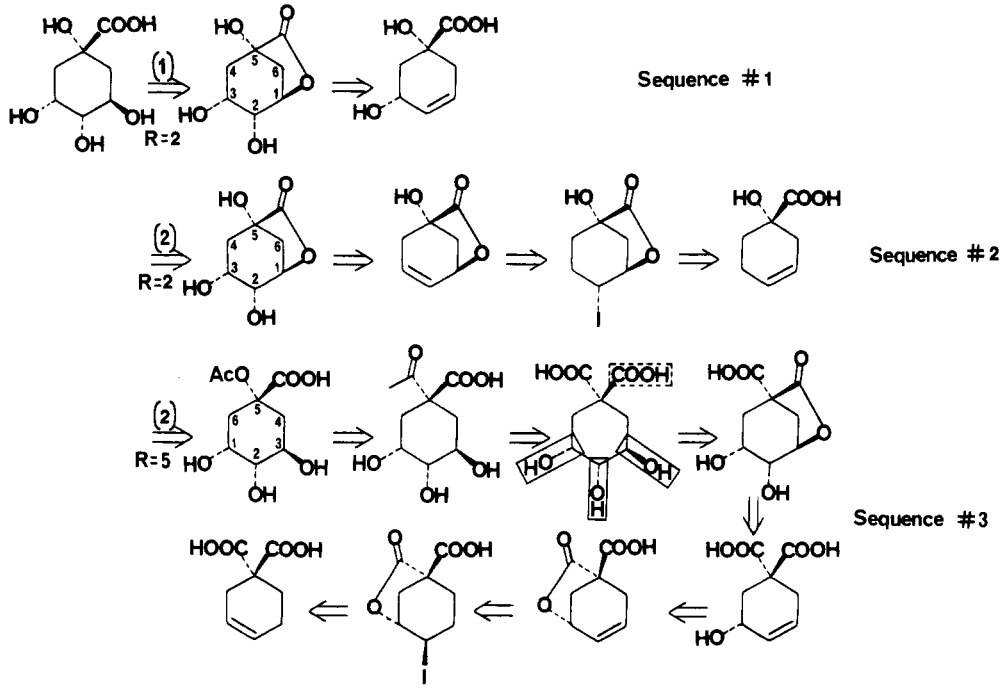
Any procedure which has been categorically rejected by PPE stands no chance of success (if the table has been properly constructed), but many procedures which do receive favorable ratings can eventually fail. This is because PPE takes no account of steric shielding considerations or of most types of interfering functionality. It is obviously not possible by inspection of a target structure to make predictions regarding the preferred steric course of a reaction buried deeply in a retrosynthetic sequence. Similarly, in the case of interfering functionality, it is impractical at the prior evaluation stage to try to take account of interfering functionality which is not present in the target but is generated during processing.

Another limitation of PPE is that the process of substructural unit categorization is rather imprecise in that it is assumed that the most direct route from target to the goal of the subroutine is applicable in all cases. In actual processing a longer but more controlled pathway may be selected, leading PPE to underestimate the length of the route. Thus PPE provides an estimate of the minimum length of each matching procedure, but this estimate is only approximate. Despite these limitations PPE provides a very effective method of controlling the combinatorial explosion which results whenever it is necessary to conduct a deep search over a very wide area as is the case with halolactonization.

ANTITHETIC ANALYSES USING THE HALOLACTONIZATION SEARCH IN LHASA

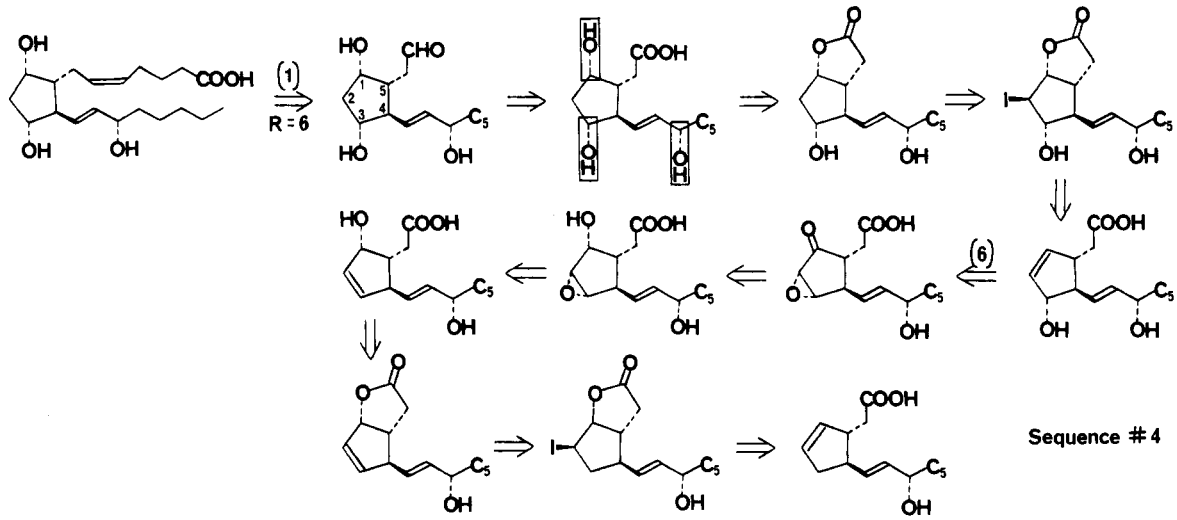
Results from the halolactonization module in LHASA have been very promising. Complete antithetic analyses have been

Scheme II. Selected Antithetic Analyses of Quinic Acid Performed by the Halolactonization Search in LHASA[a].



[a] Halolactonization paths are numbered on the first precursor in each sequence. Procedure numbers are shown in parentheses. Procedure ratings calculated by the Prior Procedure Evaluation module are indicated under the first retrosynthetic arrow for each sequence. Dotted and solid boxes indicate interfering functionality.[24]

Scheme III. One of the Antithetic Routes to PGF2α Generated by the Halolactonization Search in LHASA[a]



[a] The halolactonization path is numbered on the first precursor after the target structure. Procedure numbers are shown in parentheses. A procedure rating of 6 was calculated by Prior Procedure Evaluation, as indicated under the first retrosynthetic arrow. Solid boxes indicate protectable interfering functionality.[24]

obtained for quinic acid, $PGF_{2\alpha}$, dendrobine, and gibberellic acid, and for key intermediates not far removed from erythronolide B, vernolepin, reserpine, and trichoviridin. The sample retrosynthetic analyses shown in Schemes II–IV amply demonstrate the power of the search and the efficiency of Prior Procedure Evaluation.

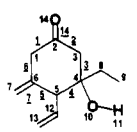
Scheme II shows three of the successful sequences generated for quinic acid. PPE initially passed more than 50 path/face/procedure combinations for this target. Of these, the 20 lowest ranked were sorted according to rating and sent along to the procedure section of the halolactonization search table. Of these 20, 16 passed the evaluation process in the chemistry subroutines and the functional group addition and functional group interchange CHMTRN tables[17] and were displayed to the chemist. It is gratifying to see that the obvious hydroxylactonization, sequence 1, is the first to be displayed. The second sequence to appear, sequence 2, is equally obvious and in fact corresponds to an even more interesting synthetic route. Procedure 1 (sequence 1) treats atoms 1 and 2 of the path as 1C LMU's. In the case of this particular path, it is more advantageous to treat atom 1 as a 1C hydroxyl LMU and atoms 2 and 3 as a 2C C=C LMU, as is done by procedure 2 (sequence 2).

Interesting routes often emerge from consideration of less obvious carboxyl equivalents as well. In sequence 3, the tertiary hydroxyl at atom 5 is used as the COOH equivalent—the hydroxyl may be generated synthetically via a Baeyer–Villiger rearrangement. In the course of execution

```
1.      IF THE CURRENT*RING IS AROMATIC OR: A HETEROCYCLIC &
                RING OR:IF THE CURRENT*RING HAS NO  &
                FUNCTIONAL GROUPS OR: STEREOCENTERS  &
                THEN REJECT
```

2. (The following code fills the GET*CO*EASY set. A chemical example is shown below. Set names like LEAVING GROUPS, MULTIPLY BONDED ATOMS, EXO*TO*RING, etc. which are not included in the list below the structure are LHASA perception sets generated in FORTRAN code and not specific to the particular Prior Procedure Evaluation table. BONDS*TO*ATOMS, ATOMS*TO*ATOMS, and ATOMS*TO*BONDS are growth functions which yield, respectively, the atoms at the ends of bonds in the operand set, the atoms attached to the atoms in the operand set, and the bonds attached to the atoms in the operand set.)

```
        PUT: THE ATOMS IN THE CURRENT*RING INTO RINGATS
        PUT: THE BONDS IN THE CURRENT*RING INTO RINGBDS
        TEMP 1 GETS LEAVING GROUPS MINUS: RINGATS MINUS:  &
                MULTIPLY BONDED ATOMS
        EXODBS GETS THE BONDS*TO*BONDS OF RINGBDS WHICH  &
                ARE ALSO: MULTIPLE BONDS
        CARBON*NUCLEOPHILE GETS CARBON ATOMS WHICH ARE  &
                ALSO: ATOMS*TO*ATOMS OF RINGATS MINUS:  &
                ATOMS*TO*ATOMS OF TEMP 1 MINUS:  &
                BONDS*TO*ATOMS OF EXODBS
        ALPHATS GETS CARBON ATOMS WHICH ARE ALSO:  &
                ATOMS*TO*ATOMS OF RINGATS
        TEMP 2 GETS SECONDARY*CENTERS MINUS:  &
                FUNCTIONAL GROUPS
        METHYL*ENE GETS PRIMARY*CENTERS MINUS METHYL  &
                MINUS: MULTIPLY BONDED ATOMS MINUS:  &
                HEMIACETAL MINUS: ACETAL
        METHYL*ENE GETS METHYL*ENE PLUS TEMP 2
        TEMP 2 GETS ALPHATS WHICH ARE ALSO: METHYL*ENE  &
                MINUS: FUNCTIONAL GROUPS
        TEMP 3 GETS ALPHATS WHICH ARE ALSO:  &
                BONDS*TO*ATOMS OF EXODBS
        TEMP 4 GETS ATOMS*TO*BONDS OF ALPHATS WHICH ARE  &
                ALSO: MULTIPLE BONDS MINUS:  &
                ATOMS*TO*BONDS OF HETERO ATOMS MINUS:  &
                THE BONDS*TO*BONDS OF RINGBDS
        TEMP 4 GETS BONDS*TO*ATOMS OF TEMP 4 WHICH ARE  &
                ALSO: ALPHATS
        TEMP 5 GETS ALPHATS WHICH ARE ALSO: FUNCTIONAL  &
                GROUPS MINUS: TEMP 4
        CARBON*ELECTROPHILE GETS TEMP 2 PLUS: TEMP 3  &
                PLUS: TEMP 5
        TEMP 6 GETS HETERO ATOMS MINUS: MULTIPLY  &
                BONDED ATOMS
        GET*CO*EASY GETS ATOMS*TO*ATOMS OF  &
                CARBON*NUCLEOPHILE PLUS: SECONDARY*CENTERS &
                MINUS: HALIDE WHICH ARE ALSO: RINGATS  &
                WHICH ARE ALSO: ATOMS*TO*ATOMS OF TEMP 6
        GET*CO*EASY GETS GET*CO*EASY PLUS: BONDS*TO*ATOMS  &
                OF EPOXIDE PLUS: BONDS*TO*ATOMS OF EXODBS  &
                WHICH ARE ALSO: RINGATS
```



Set	Set Elements
RINGATS	1 2 3 4 5 6
RINGBDS	1 2 3 4 5 6
EXODBS	7 14
CARBON*NUCLEOPHILE	8 12
ALPHATS	7 8 12
METHYL*ENE	1 3 8
CARBON*ELECTROPHILE	7 8
GET*CO*EASY	4 6

3. (The following code extracts the GET CO subroutine rating from the LMU categorization sets.)

```
CALC    VAR1 = #1
BLOCK1  IF SAVED*ATOM [ VAR1 ] IS NOT A KETONE  &
                THEN GO TO CO1
        SR*GET*CO 0 (VAR1) = #0
        GO TO DE
CO1     IF SAVED*ATOM [ VAR1 ] IS NOT IN GET*CO*EASY  &
                THEN GO TO CO3
        SR*GET*CO 0 (VAR1) = #1
        GO TO DE
CO3     IF SAVED*ATOM [ VAR1 ] IS NOT IN GET*CO*MOD  &
                THEN GO TO CO5
        SR*GET*CO 0 (VAR1) = #3
        GO TO DE
CO5     IF SAVED*ATOM [ VAR1 ] IS NOT IN GET*CO*DIFF  &
                THEN GO TO COKILL
        SR*GET*CO 0 (VAR1) = #5
        GO TO DE
COKILL  SR*GET*CO 0 (VAR1) = #100
DE      VAR1 = VAR1 + #1
        IF (VAR1 :LE: #6) THEN GO TO BLOCK1
```

4. (The LMU's which comprise the 4 carboxyl sets are shown below with their corresponding values of BOXRAT (see paragraph 6).)

BOXRAT	CHMTRN SET	LMU
0	CARBOXYL	-COOR
1	GET*COOH*EASY	$-CH_2OH$, -CHO, -NR2
3	GET*COOH*MOD	-OR, $-CR_2X$, SPIRO EPOXIDE
5	GET*COOH*DIFF	$-CR_3$, =O, $=CR_2$

X = any hetero atom
R = H or alkyl or acyl

5. (The following code calculates a preliminary procedure rating PRELRAT for Procedure 8 by summing the subroutine ratings for the chemistry subroutines called by the procedure. The first few lines of code apply restrictive heuristics to invalidate the procedure if certain adverse structural features are encountered in the target molecule.)

```
PR8     PROCNO = 8
        IF THERE IS A LEAVING GROUP ON  &
                SAVED*ATOM [ RING*SIZE ] OR: SAVED*ATOM 4  &
                THEN GO TO PR9
        IF SAVED*ATOM 3 IS MULTIPLY BONDED THEN GO TO PR9
        IF THERE ARE NOT TWO HYDROGENS ON SAVED*ATOM 3  &
                THEN GO TO PR8A
        IF ALPHA TO SAVED*ATOM 3 HAS A LEAVING GROUP OR:  &
                IS ENOLIZABLE THEN GO TO PR9
PR8A    PRELRAT = SR*GET*CO 1 + SR*ENONA 2 + SR*ENONB 3 + #2
        CALL A*CARBOX*RAT
        CALL B*CARBOX*RAT
PR9...

6.      PROCRAT = PRELRAT + BOXRAT

7.      WHO*WHERE (IX) = DIRECTION + FACE*COUNT * #2  &
                + ORIENTATION * #8 + PROCNO * #64  &
                + LABEL * #4096

8.      SORT IN ASCENDING ORDER [ PROC*PTR ] ELEMENTS  &
                OF THE ARRAYS AT PROC*RATING 1 AND  &
                WHO*WHERE 1

9.      LOOP ON IX VALUES EQUALING FROM (#1) UP*TO  &
                (PROC*PTR)
        IF UNSUCCESSFUL THEN DISCONTINUE
        DIRECTION = WHO*WHERE (IX) :AND: #1
        FACE*COUNT = ( WHO*WHERE (IX) / #2 ) :AND: #1
        ORIENTATION = ( WHO*WHERE (IX) / #8 ) :AND: #7
        PROCNO = ( WHO*WHERE (IX) / #64 ) :AND: #15
        LABEL = WHO*WHERE (IX) / #4096

10.     LOOP ON A PATH OF [ RING*SIZE ] BONDS STARTING  &
                FROM SAVED*ATOM [ ORIENTATION ]  &
                TOWARDS SAVED*BOND [ ORIENTATION -  &
                DIRECTION ] USING ONLY ATOMS IN RINGATS

11.     JUMP USING [ PROCNO ] TO LOCATIONS PROC1 PROC2  &
                PROC3 PROC4 PROC5 PROC6 PROC7 PROC8  &
                PROC9 PROC10
        .
        .
        .
PROC8   WRITE MESSAGE 12
        CALL GET*ENONE AT ATOM*3 AND ATOM*2 AND ATOM*1
        IF SUCCESSFUL THEN GO TO PROC8B
        IF THERE IS A HETERO ATOM ALPHA TO ATOM*3 THEN  &
                GO TO PROC8A
        CALL RING3 AT ATOM*1 AND ATOM*2 AND ATOM*3
        IF UNSUCCESSFUL THEN DISCONTINUE
PROC8B  CALL GET*LACTONE AT ATOM*1 AND SAVED*ATOM 1
        IF UNSUCCESSFUL THEN DISCONTINUE
        IF THERE IS AN OLEFIN ON ATOM*2 THEN EXCHANGE  &
                THE GROUP FOR AN IODIDE DEFINED*CIS TO  &
                THE BETA*FACE
```

(The EXCHANGE command causes the program to identify all possible functional group interchange sequences of up to 4 steps that could convert the functionality currently on specified atom 1 to a ketone. Sequences are ranked according to probable efficiency and then attempted until one succeeds or all fail.)

```
        IF UNSUCCESSFUL THEN DISCONTINUE
        CALL DO*HALAC AT ATOM*2 AND ATOM*1 AND SAVED*ATOM 1
        IF UNSUCCESSFUL THEN DISCONTINUE
        FINISHED
PROC8A...
```

One chemistry subroutine, RING3, is reproduced below in its entirety. RING3 is called with three atom calling arguments which are referred to as SPECIFIED*ATOMs 1, 2, and 3, respectively, within the subroutine. Lines beginning with dots are comments.

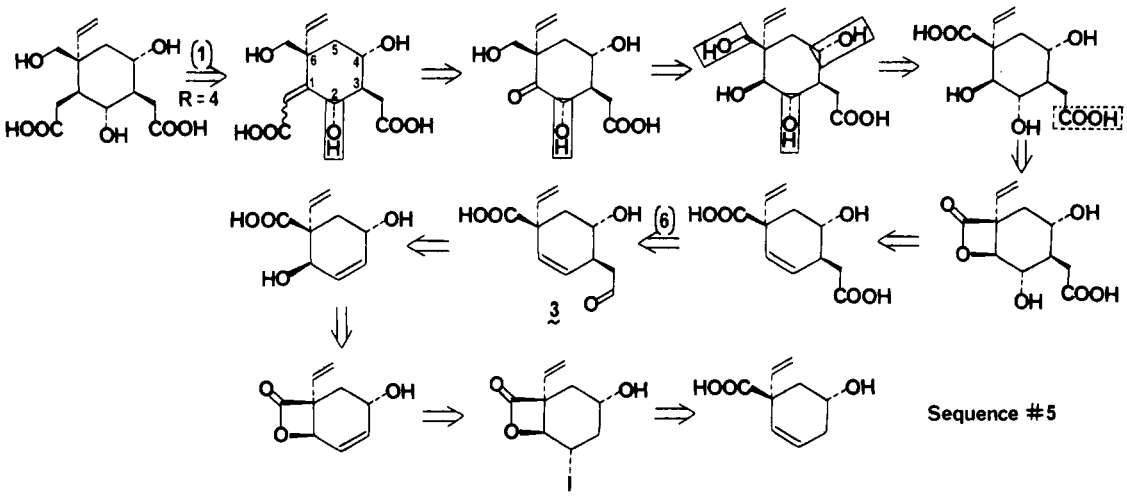
```
...
...Check for a methyl group on SPECIFIED*ATOM 3
...
RING3   IF SPECIFIED*ATOM 3 IS NOT A TERTIARY*CENTER THEN  &
                RETURN FAIL
        IF THERE IS NOT A METHYL GROUP ON  &
                SPECIFIED*ATOM 3 THEN RETURN FAIL
        SAVE AS 1 THE PREVIOUS LOCANT
                ...Label Me as SAVED*ATOM 1
        IF SAVED*ATOM 1 IS NOT CONTAINED IN THE  &
                ALPHA*FACE THEN RETURN FAIL
...
...Try to get a ketone on SPECIFIED*ATOM 1
...
        IF THERE IS A FUNCTIONAL GROUP ON  &
                SPECIFIED*ATOM 1 THEN GO TO RING3A
        IF THERE ARE NOT TWO HYDROGENS ON  &
                SPECIFIED*ATOM 1 THEN RETURN FAIL
        MAKE A KETONE ON SPECIFIED*ATOM 1
```

(The MAKE command causes the program to search its data base for functional group addition transforms that could introduce a ketone on the desired locant.)

```
        IF UNSUCCESSFUL THEN RETURN FAIL
        GO TO RING3B
RING3A  IF THERE IS A KETONE ON SPECIFIED*ATOM 1  &
                THEN GO TO RING3B
        EXCHANGE THE GROUP FOR A KETONE
        IF UNSUCCESSFUL THEN RETURN FAIL
...
...Try to dealkylate SPECIFIED*ATOM 2
...
RING3B  IF THERE ARE TWO HYDROGENS ON SPECIFIED*ATOM 2  &
                THEN GO TO RING3C
        CALL DEALKYLATE AT SPECIFIED*ATOM 2 AND  &
                SPECIFIED*ATOM 1
        IF THERE ARE NOT TWO HYDROGENS ON  &
                SPECIFIED*ATOM 2 THEN RETURN FAIL
...
...Reconnect to the cyclopropane via composite subgoal
...
RING3C  JOIN SPECIFIED*ATOM 2 AND SPECIFIED*ATOM 1
        THE BOND SPANNING SPECIFIED*ATOM 3 AND  &
                SAVED*ATOM 1 IS DEFINED*CIS TO THE  &
                NEWEST*BOND ON SPECIFIED*ATOM 2
        PERFORM THE MECHANISM AND DISPLAY THE PRECURSOR
...
...Try to convert the ketone to an alpha face alcohol
...
        IF THERE IS NOT A KETONE ON SPECIFIED*ATOM 1 THEN  &
                RETURN FAIL
        EXCHANGE THE GROUP FOR AN ALCOHOL DEFINED*CIS TO  &
                THE ALPHA*FACE
        IF UNSUCCESSFUL THEN RETURN FAIL
...
...Perform Simmons-Smith transform via composite subgoal
...
        DELETE SAVED*ATOM 1
        JOIN SPECIFIED*ATOM 2 AND SPECIFIED*ATOM 3
        PERFORM THE MECHANISM AND DISPLAY THE PRECURSOR
        RETURN SUCCESS
```

Figure 2.

Scheme IV. LHASA-Generated Antithetic Routes to a Key Intermediate for the Synthesis of Vernolepin[a]



[a] The halolactonization path is numbered on the first precursor after the target structure. Procedure numbers are shown in parentheses. A procedure rating of 4 was calculated by Prior Procedure Evaluation, as indicated under the first retrosynthetic arrow. Solid and dotted boxes indicated interfering functionality.[24]

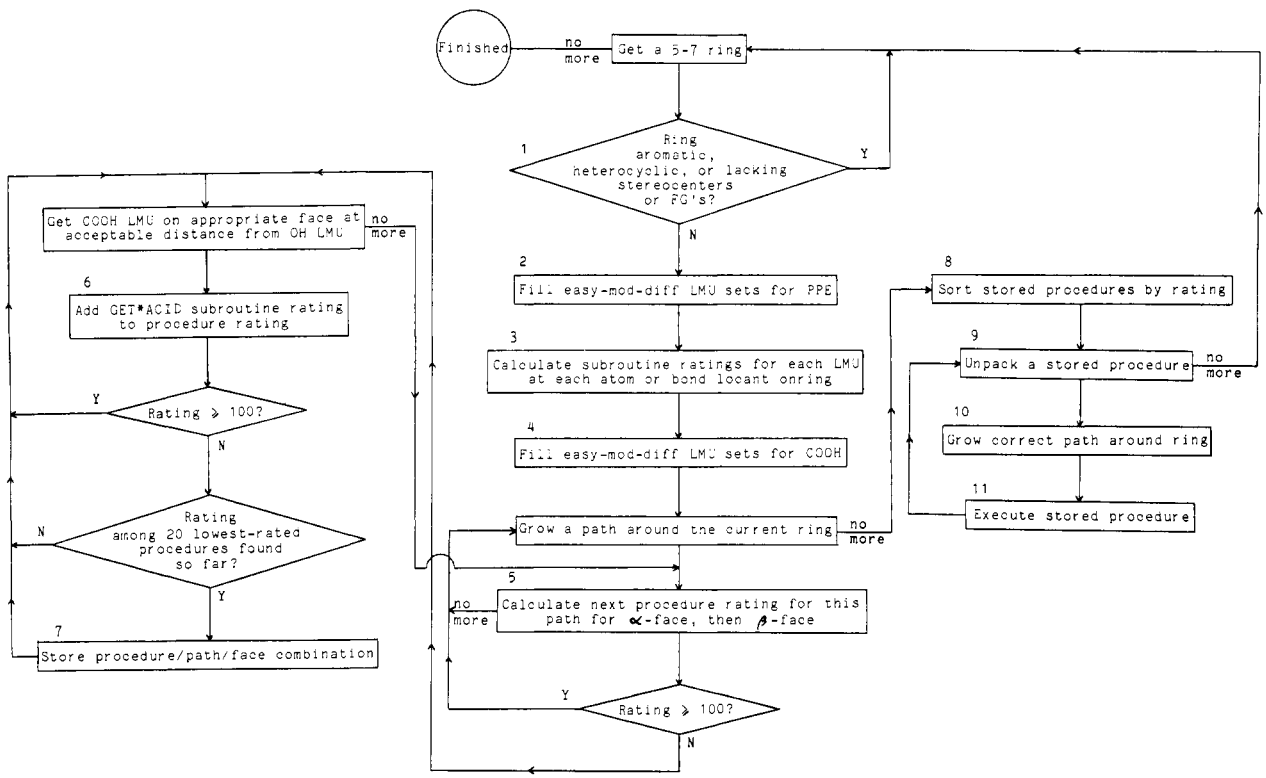
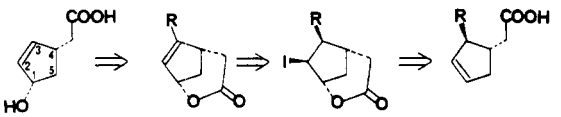


Figure 3. Flowchart for the halolactonization search in LHASA. Numbers on the boxes correspond to numbered paragraphs in Figure 2.

of this procedure, subroutine GET*DB fails to obtain a C=C between atoms 2 and 3 of the path. Fortunately, however, the SUPER*GET*DB subroutine is able to use the COOH present at atom 5 in the target to direct the required trans-hydroxylation transform, paving the way for the dehydroiodination transform and the eventual iodolactonization. While the symmetry of the olefinic acid generated in this sequence is intriguing, certain functional group protection problems might arise in the suggested synthesis. A number of these problems could be avoided by a simple reordering of steps, in which the cis glycol would be protected after the hydroxylactonization reaction and the Baeyer–Villiger sequence performed with the lactone still closed.

Scheme III shows a representative sequence from the halolactonization analysis of $PGF_{2\alpha}$. In the case of this target, regiochemical problems which could not be anticipated by PPE were encountered, and only 5 of the 20 lowest ranked path/face/procedure combinations passed further evaluation. Sequence 4 illustrates the method by which procedure 1 can remove functionality at atom 3 of the path. After generation of the olefinic acid which is the normal end point for procedure 1, processing passes to procedure 6 if there is functionality at atom 3. In this case, the hydroxyl at atom 3 and the atom 1–atom 2 C=C are the exact S goal functionality for procedure 6, and a Wharton rearrangment sequence paves the way for the second iodolactonization. The Wharton rearrangement is especially useful for five-membered-ring targets, since the simpler alternative of treating the allylic alcohol as a path-reversed procedure 2 S goal leads to a halolactonization with regiochemical problems because the C=C of the olefinic acid

is symmetrically disposed with respect to the COOH group:



In cases where there is a functional group at atom 3 of the path, the rating for procedure 1 is incremented by 2 (sequences 1 and 4). This increment reflects not only the added complexity associated with coupling procedure 1 with procedure 6 but also the added efficacy of processing these path/face combinations with procedure 2, in which atom 3 is included in the atom 2–atom 3 C=C LMU. The procedure 6 sequence coupled to sequence 1 was omitted for simplicity.

Coupling procedure 1 with procedure 6 often provides an elegant method for antithetic removal of appendages, as exemplified by sequence 5. In this case the procedure 6 S goal is the olefinic aldehyde **3** which keys a Claisen rearrangement transform. This stereospecific antithetic removal of the two-carbon appendage at atom 3 results in an allylic alcohol which keys the second iodolactonization transform in the standard fashion.

IMPROVEMENTS PLANNED FOR THE HALOLACTONIZATION SEARCH

Despite the obvious sophistication of the sequences in Schemes II–IV, there are a few aspects of the halolactonization search for which improvements are envisioned. The first concerns the handling of stereochemistry. Recent additions to the stereochemical modules have allowed generation of specific epimers when converting an sp^2 center to an sp^3 center antithetically and also the automatic generation of all possible stereoisomers if a specific epimer is not explicitly requested. Another recent improvement is the ability to perform, via a rapid, semiquantitative algorithm, the complete conformational analysis of six-membered rings.[18] While this capability leads to a powerful means of evaluating the stereochemical outcome of many transforms in six-membered rings, prediction of the stereochemical feasibility of transforms in which an sp^3 center is converted to an sp^2 center antithetically[19] is not yet possible. The special module for steric assessment conceived by the LHASA group[20,21] in 1975–1976 will give LHASA greatly expanded capabilities in this area.

Functional group interference is a particularly difficult problem in highly functionalized molecules like those for which the halolactonization search was designed. It is often the case (as, for example, in sequence 3 in Scheme II) that a sequence can be markedly improved by a judicious reordering of its constituent steps. The matching procedure/chemistry subroutine construction of the halolactonization search imposes restrictions on the order in which transforms are suggested by the program, and a postprocessor module which can optimize that order will provide a significant improvement for many of the sequences.

Finally, as outlined in the introductory section, the use of coupled long-range searches involving combined goals or strategies is an important target of our ongoing investigations. Synergistic combinations of two or more T goals will become increasingly important as more long-range searches are added to LHASA.

ACKNOWLEDGMENT

We are grateful to the National Institutes of Health and the Science Research Council (U.K.) for grants in support of this work. J.M. was the recipient of a scholarship from the Deutsche Forschungsgemeinschaft.

APPENDIX: CHMTRN CODE FOR THE HALOLACTONIZATION TABLE

The entire data base of LHASA, including the halolactonization table, is written in the CHMTRN (CHeMistry TRaNslator) language.[22] Although CHMTRN is a computer-readable language,[23] its words and grammar are fashioned after chemical English so that a chemist with no computer experience can quickly learn to read and write it. In spite of its easily comprehensible format, however, CHMTRN is a powerful language. A large number of terms are defined, making it possible to pose a wide variety of chemically significant questions.

The first version of CHMTRN, written in 1971, contained many of the fundamental features of the current version. However, as the LHASA program has grown, the language has evolved accordingly. The implementation of increasingly sophisticated chemical strategies such as halolactonization has required a corresponding increase in the scope and versatility of CHMTRN. Representative of current language features are the ability to call subroutines with multiple atom and bond arguments, the execution of iterative blocks of code (i.e., loops), logical combinations of chemical sentences, definition and manipulation of sets of atoms and bonds, and performance of integer mathematics with user-defined variables and arrays.

Many sophisticated CHMTRN features, including all of those just mentioned, are used in the halolactonization table. To illustrate some of these features, we show in Figure 2 the CHMTRN code corresponding to portions of the main halolactonization search (see Figure 3). Numbers accompanying the CHMTRN code correspond to the numbered boxes in Figure 3.

REFERENCES AND NOTES

(1) (a) Corey, E. J. *Q. Rev., Chem. Soc.* **1971**, *25*, 455. (b) Corey, E. J.; Howe, W. J.; Pensak, D. A. *J. Am. Chem. Soc.* **1974**, *96*, 7724. (c) Corey, E. J.; Long, A. K. *J. Org. Chem.* **1978**, *43*, 2208. (d) Corey, E. J.; Johnson, A. P.; Long, A. K. *Ibid.* **1980**, *45*, 2051.
(2) Corey, E. J.; Howe, W. J.; Orf, H. W.; Pensak, D. A.; Petersson, G. *J. Am. Chem. Soc.* **1975**, *97*, 6116.
(3) (a) Corey, E. J.; Weinshenker, N. M.; Schaaf, T. K.; Huber, W. *J. Am. Chem. Soc.* **1969**, *91*, 5675. (b) Corey, E. J.; Noyori, R. *Tetrahedron Lett.* **1970**, 311. (c) Corey, E. J.; Moinet, G. *J. Am. Chem. Soc.* **1973**, *95*, 6831. (d) Corey, E. J.; Mann, J. *Ibid.* **1973**, *95*, 6832.
(4) Corey, E. J.; Shibasaki, M.; Knolle, J. *Tetrahedron Lett.* **1977**, 1625.
(5) (a) Corey, E. J.; Danheiser, R. L.; Chandrasekaran, S.; Siret, P.; Keck, G. E.; Gras, J.-L. *J. Am. Chem. Soc.* **1978**, *100*, 8031. (b) Corey, E. J.; Danheiser, R. L.; Chandrasekaran, S.; Keck, G. E.; Gopalan, B.; Larsen, S. D.; Siret, P.; Gras, J.-L. *J. Am. Chem. Soc.* **1978**, *100*, 8034.
(6) (a) Corey, E. J.; Trybulski, E. J.; Melvin, L. S.; Nicolaou, K. C.; Secrist, J. A.; Lett, R.; Sheldrake, P. W.; Falck, J. R.; Brunelle, D. J.; Haslanger, M. F.; Kim, S.; Yoo, S. *J. Am. Chem. Soc.* **1978**, *100*, 4618. (b) Corey, E. J.; Kim, S.; Yoo, S.; Nicolaou, K. C.; Melvin, L. S.; Brunelle, D. J.; Falck, J. R.; Trybulski, E. J.; Lett, R.; Sheldrake, P. W. *Ibid.* **1978**, *100*, 4620. (c) Corey, E. J.; Hopkins, P. B.; Kim, S.; Yoo, S.; Nambiar, K. P.; Falck, J. R. *Ibid.* **1979**, *101*, 7131.
(7) Corey, E. J.; Albright, J. O.; Barton, A. E.; Hashimoto, S. *J. Am. Chem. Soc.* **1980**, *102*, 1435.
(8) Corey, E. J.; Barton, A. E.; Clark, D. A. *J. Am. Chem. Soc.* **1980**, *102*, 4278.
(9) Corey, E. J.; Hase, T. *Tetrahedron Lett.* **1979**, 335.
(10) For a recent review of halolactonization see: Dowle, M. D.; Davies, D. I. *Chem. Soc. Rev.* **1979**, 171.
(11) The halolactonization table, like all the long-range search tables in LHASA, is written in CHMTRN (CHeMistry TRaNslator), a powerful "chemical English" language.
(12) The reader should note that other possibilities for dissection of S goal 1 into localized matching units are available. For example, the hydroxyl and C=C functional groups could be regarded as an allylic alcohol (a three-carbon LMU) rather than as two isolated LMU's.
(13) For complete flowcharts of the chemistry subroutines, see ref 17b, Chapter 7.
(14) Julia, S.; Furer, B. *C. R. Hebd. Seances Acad. Sci.* **1963**, *257*, 710.
(15) It is recognized that the number of reactions is a relatively poor indicator of the overall effectiveness of the process. Revised versions of PPE will use more sophisticated measures of effectiveness.
(16) PPE takes an average of 45 s of computer time for each ring, depending on the number of carboxyl equivalents.
(17) (a) Corey, E. J.; Cramer, R. D., III; Howe, W. J. *J. Am. Chem. Soc.* **1972**, *94*, 440. (b) Long, A. K. Ph.D. Thesis, Harvard University, 1979.

(18) Corey, E. J.; Feiner, N. F. J. Org. Chem. 1980, 45, 757, 765.
(19) Wipke, W. T.; Gund, P. J. Am. Chem. Soc. 1974, 96, 299.
(20) Corey, E. J.; Feiner, N. F.; Orf, H. W.; Long, A. K.; Vinson, J. W.; Hewett, A. P. W.; Stolow, R. D., unpublished results.
(21) The steric control module is outlined in ref 17b, Chapter 5, Section VI.
(22) A complete description of the language as of 1976 appears in: H. W. Orf, Ph.D. Thesis, Harvard University, 1976, Appendix C.
(23) In order to store CHMTRN sentences as compactly as possible and to save time as they are being read by the computer, the chemical English sentences are translated into a more compact, computer-readable form by a "translator" program. The translator program, TBLTRN (TaBLe TRaNslator), was written by D. E. Barth at Harvard University in 1971. It is written in FORTRAN and is available as program 10-168 from the DECUS Program Library.
(24) Corey, E. J.; Orf, H. W.; Pensak, D. A. J. Am. Chem. Soc. 1976, 98, 210.

Computer-Aided Spectral Identification of Laser-Induced Plasma Emission

YEDIDYAH LANGSAM,* AARON TENENBAUM, and JOSEPH THURM

Department of Chemistry and Department of Computer & Information Science, Brooklyn College of the City University of New York, Brooklyn, New York 11210

Received February 7, 1980

The deconvolution of the electronic emission spectra produced by a laser-induced plasma may be performed by computer analysis. This paper presents two algorithms for matching such spectra against standard tables. The relative efficiencies of the two algorithms are analyzed.

I. INTRODUCTION

Since its very inception in the early 1960s, laser technology has been applied to a diverse number of fields. Yet it was only recently that the laser's inherent monochromaticity, high intensity, and coherence was utilized in order to directly manipulate the course of chemical reactions.[1-3] In recent years, much effort has been spent in the initiation of chemical reactions by the absorption of resonant laser energy. The technique of Laser-Induced Dielectric Breakdown (LIDB), on the other hand, creates a plasma and subsequent chemical reaction even where a resonant condition does not exist.[2,4-7] Of particular interest is the investigation of the identities and lifetimes of various species found within the laser-induced plasma and their dependence on such parameters as pressure, laser fluence, and companion gases which may be present.

Recent work in our laboratory focused on the LIDB of the metal carbonyls.[8] When a sample of metal carbonyl is irradiated by a CO_2 laser of sufficient fluence (J/cm^2), visible breakdown occurs. A number of features of the laser-induced plasma are easily apparent to the eye. The plasma is an extremely intense blue-white isotropic source of light containing distinct regions. At the center there is an extremely intense core of indeterminably small volume. In this high-energy region only highly ionized atoms and molecules may exist. Surrounding this core is a region of lesser intensity consisting of excited species rather than ions.[8] Electron energies may range from 20 keV in the core to tens of electron volts in the surrounding cloud.[9] As is to be expected, measurements of the spectral distributions of the plasma emission indicate a strong line spectrum superimposed on a continuum with peaks at characteristic wavelengths depending again on the plasma medium as well as the laser energy.

An optical multichannel analyzer was used in order to identify the species produced within a laser-induced dielectric breakdown plasma. This device records the plasma emission, after suitable background subtraction and enhancement by multipulse integration, into 500 channels as a function of wavelength. The channels may be calibrated vs. known Hg-lamp emission lines, and thus the wavelengths corresponding to each of the 500 channels may be calculated. In these experiments, the visible region was scanned in two parts, a "high" region roughly between 4164 and 7164 Å calibrated by means of the Hg lines at 5461 and 5770 Å, and a "low" region corresponding to 3093–6001-Å emission calibrated against the 4350- and 5461-Å Hg lines. Typical emission spectra obtained by using the OMA are presented in Figure 1.

As mentioned earlier, the plasma state is richly populated by a wide range of ionic and neutral, ground and excited, atomic and molecular species. The identification of these plasma states may be accomplished by comparison with all the tabulated emission lines of any species which may conceivably be present. Examination of the thousands of lines tabulated in ref 10–15 reveals that the identification is complicated not only by the many overlapping emission lines but by the line width of the OMA detector (~6 Å/channel). It was therefore thought that the nature of the emitting species could be elucidated by means of a statistical matchup of OMA spectra with known emission lines. While a manual search of this type could be described as tedious if not overwhelming at best, it could easily be done, for the many emission spectra investigated, by means of a computer.

II. METHODOLOGY

The analysis of the OMA output proceeds in three phases. The first phase analyzes the OMA spectra and calculates the wavelength (in angstroms) of each peak observed, by comparison with the mercury-line calibration. These wavelengths, from both the "high" and "low" regions, are sorted in ascending order and stored on a disk.

In each computer run, the OMA spectra are compared with up to five elements or compounds (e.g., O, O_2, CO, etc.). Each of these elements, of course, has a group of subcategories (e.g., OI, OII, $B^1\Sigma^+$, $X^3\Sigma^-$, etc.), each of which has its own characteristic spectrum. Up to five "elemental input decks" containing these spectra are input in the second phase of the processing. Each input deck consists of a card containing the name of the major element or compound (identified by a 1 in column 1), followed by one or more groups of cards representing the subcategories and their spectra. Each of these groups consists of a card containing the name of the subcategory (identified by a 2 in column 3), followed by cards containing the known wavelengths of the spectrum for that subcategory (identified by a 3 in column 5). A sample ele-

0095-2338/80/1620-0230$01.00/0