CENTENARY LECTURE

Computer-assisted Analysis of Complex Synthetic Problems

By E. J. Corey
DEPARTMENT OF CHEMISTRY, HARVARD UNIVERSITY, CAMBRIDGE, MASSACHUSETTS 02138, U.S.A.

1 Introduction

For the past several years a programme of research has been carried out at Harvard, the major objective of which has been the application of digital computers to assist in the derivation of synthetic routes to complex molecules. A review of this work will be presented here.

A complex synthetic problem for purposes of this discussion can be defined by the following criteria:

1. A solution cannot be obtained simply by analogy with previously solved problems.
2. The starting point(s) or material(s) for the synthesis are not directly apparent.
3. Many possible pathways of synthesis must be examined in order to ensure the selection of the simplest and most useful approach.
4. The structure to be synthesised is itself complex—not merely because of its size but also in terms of the presence of complicating structural features such as functional groups or reactive centres, rings, chiral or geometric stereocentres, destabilising interactions, *etc.*

With this definition it becomes apparent to a synthetic chemist that a project to develop a general problem-solving procedure for use by a computer must be regarded as long range in character. The task is too large to be accomplished in a five- or ten-year period and indeed is unlikely to be complete, in a final sense, in the foreseeable future. As with a 'theory' for complex situations, any general procedure for complex problem solving will be subject to further improvement and hence to an evolutionary course of development. At the outset of the studies described herein, it was by no means certain that meaningful progress could be made even in the development of simple problem-solving techniques and, indeed, there will doubtless continue to be a large group of chemists who take a sceptical view of the whole enterprise. Such scepticism may reasonably be based on the almost incredible diversity of organic structures, the complexities of stereochemistry, and the very large variety and number of chemical processes available to the synthetic chemist (not to mention the remarkably complicated limits on the scope of each process). Further, the need for compromise or 'trade-off'

between 'generality' and 'power' in general problem-solving procedures[1] must be borne in mind. On the other hand, even if the effort to devise an effective problem-solving computer program were to fail utterly, a deeper comprehension of the strategies, principles, and elements of chemical synthesis would be gained, the classification and organisation of basic chemical data according to the requirements of synthesis would be advanced, and new and more powerful methods of teaching chemical synthesis and solving synthetic problems would result. These advances in the understanding and codification of an important area of chemistry can be regarded as a goal of *fundamental* synthetic research, the attainment of which is destined to yield results of considerable value.

Two computers have been used in the studies on machine-assisted synthetic analysis at Harvard, a PDP-1, vintage *ca.* 1960, and a modern PDP-10 (both Digital Equipment Corp.). The newer machine has the capacity to handle properly the very large program which is evolving. This program will be transferable to other machines, since it is being written in the most commonly used higher level language, FORTRAN IV, and designed so as to minimise hardware dependence. In contrast, the older machine has quite limited memory resources and must be programmed in a specialised assembler language (DECAL) which is unique to it. The program currently being used with the PDP-1, designated LHASA-1 (Logic and Heuristics Applied to Synthetic Analysis), lacks a stereochemical capability and is incomplete with regard to chemical program modules. However, it is very useful in the development and testing of new ideas and program modules which will be used eventually in LHASA-10, the PDP-10 program which is expected to become operational in the mid-to-later 1970s. The aspects of computer-assisted synthetic analysis discussed herein, unless otherwise indicated, have been implemented in the LHASA-1 and/or LHASA-10 programs.

An outline of the general approach which has guided the initial phase of program development has been presented previously.[2] This paper can also serve as an introduction to the present Review which will be oriented mainly towards points of chemical interest rather than programming details or computational aspects. In connection with the latter there are a number of printed works[3] which may be used as reference texts in connection with this Review or for the purpose

[1] G. W. Ernst and A. Newell, 'GPS: A Case Study in Generality and Problem Solving', Academic Press, New York, 1969.
[2] E. J. Corey and W. T. Wipke, *Science*, 1969, **166**, 178.
[3] The following references are listed approximately in order of increasing sophistication. (*a*) A. I. Forsyth, T. A. Keenan, E. I. Organick, and W. Sternberg, 'Computer Science, a First Course', J. Wiley and Sons, New York, 1969; (*b*) A. Ralston, 'Introduction to Programming and Computer Science', McGraw-Hill, New York, 1971; (*c*) I. Flores, 'Computer Programming', Prentice-Hall, Englewood Cliffs, N.J., 1966; (*d*) A. T. Berztiss, 'Data Structures, Theory and Practice', Academic Press, New York, 1971; (*e*) D. G. Hays, 'Introduction to Computational Linguistics', Elsevier, New York, 1967; (*f*) P. Wegner, 'Programming Languages, Information Structures and Machine Organization', McGraw-Hill, New York, 1968; (*g*) 'Computers and Thought', ed. E. A. Feigenbaum and J. Feldman, McGraw-Hill, New York, 1963; (*h*) 'Semantic Information Processing', ed. M. Minsky, MIT Press, Cambridge, Mass., 1968.

of gaining a general background in the computational and data processing techniques which are fundamental to this sort of computer application.

One approach[3,4] to the derivation of synthetic pathways to some target structure involves the generation of a set of intermediates which can be converted into that structure by one synthetic step and the iteration of this procedure for each intermediate so generated until a 'tree' of intermediates is developed (Figure 1). This technique forms the basis for the computer programs currently being

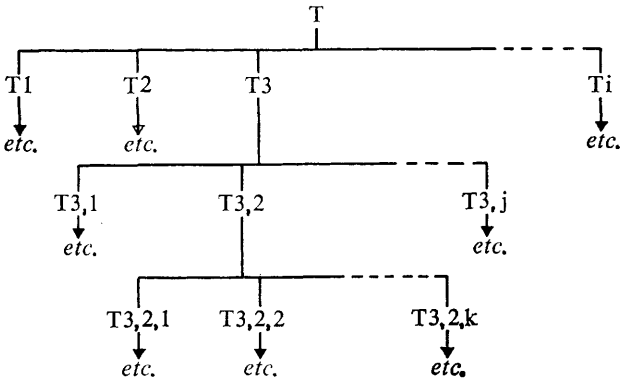


Figure 1 *Synthetic analysis of target T generates a 'tree' of intermediate precursor structures*

used and extended. It involves analytical processes which depend heavily upon the structural features of *reaction products* (as contrasted with starting materials) and the consideration of molecular changes in the *retro-synthetic* sense. In order to avoid confusion, two special terms and a special graphic have been employed to provide a distinction between the nomenclature appropriate to these analyses and that which is conventionally applied to synthesis in the direction of laboratory execution. The terms *antithetic* and *transform* and the 'double-arrow' graphic will be used strictly as is indicated in Scheme 1. It is noteworthy that this usage of the term *transform* has a parallel to the mathematical meaning (a *function*

1. Direction of laboratory execution is '*synthetic*'		1. Direction of computer analysis is '*antithetic*'
2. Represented as (s) $\longrightarrow$	*vs.*	2. Represented as (a) $\Longrightarrow$
3. Process is called a '*reaction*'		3. Process is called a '*transform*'

Scheme 1 *Terminology for chemical structural changes in either of two directions*

[4] E. J. Corey, *Pure Appl. Chem.*, 1967, **14**, 19.

operating on an *argument* to produce a *result*) in the sense that a *transform* operates on a chemical structure to produce a different (transformed) structure.

By way of introduction a précis of the key features of current programs is given in Scheme 2. Communication between man and machine is accomplished graphically by a method first developed for this project in 1967.[3,5] The chemist 'inputs' a structure by drawing a standard two-dimensional structural diagram

1. Graphical Man-Machine Communication
   Structural input: electrostatic (Rand) tablet and stylus.
   Structural output: CRT displays (2), plotter (hard copy).
2. Emphasis on Interactive Relationship between Man and Machine
3. Tabular Internal Representation of Structure
   Atom and bond connection tables (input structure).
   Structure information blocks (each intermediate).
4. Machine-Oriented Perception of Synthetically Significant Structural Features
   Functional groups, rings, stereorelationships, *etc.*
5. Automatic Generation of Synthetic Intermediates
   Direction of analysis: antithetic.
   'Tree' collection of synthetic intermediates with target molecule as parent.
   Target-oriented data files, process (transform) selection and evaluation.
6. Multiple Problem-Solving Strategies
   Heuristic rules; goal and subgoal generation; strategic bond disconnections.

Scheme 2 *A program for computer-assisted synthetic analysis—LHASA (Logic and Heuristics Applied to Synthetic Analysis)*

using an electrostatic tablet (Rand tablet) capable of sensing positions on a 2^{10} by 2^{10} point grid and a pen equipped with a switch which closes and allows communication with the tablet when pressed down. The pen leaves no visible trace on the tablet but creates a display of the structure being drawn (and a tracking cross which locates the pen) on a cathode ray tube. All structural information output from the computer is also displayed on a cathode ray tube as a conventional formula drawing. Further, output structures can be provided as hard copy on paper using a commercial graphics plotter. In this form easy and rapid communication occurs in the natural pictorial language of the chemist in a way which requires neither training nor special skills. Because of the availability of such ready man–machine communication and the desirability of allowing the chemist to influence and direct the analysis of a problem to whatever extent he judges appropriate, the programs for synthetic analysis are designed to be highly interactive. The chemist at each stage has the option to specify, modify, or channel the flow of analysis.

[5] E. J. Corey, W. T. Wipke, R. D. Cramer, and W. J. Howe, *J. Amer. Chem. Soc.*, in press.

As is indicated in Scheme 2, the internal representation of chemical structure within the computer involves connection tables for the atoms and bonds within the structure (together with x, y co-ordinates for graphical display). Structure information blocks which contain data on *changes* in atom and bond tables may also be used to provide information on any 'offspring' structures which result from structural manipulation of an input or target structure. From the atom and bond tables information is extracted which is needed by the program. This process, which is designated 'perception', makes available data on synthetically significant structural features such as functional groups, rings, *etc.* The program provides for automatic selection of transforms, their evaluation, and their application to generate a 'tree' of synthetic intermediates. Transform selection may be guided by the chemist or by a number of strategies which are being added to the program. These strategies, which are allowed to operate independently of one another and which vary greatly in nature, parallel to a certain extent those used by the expert chemist.

Before proceeding further in the discussion of machine program(s) for synthetic analysis, it is instructive to review the most common approach of the chemist to problem solving, and this is outlined in Scheme 3. This approach could be simulated by two computer programs being executed simultaneously and intercommunicating (for example, in a time-sharing environment). One program would operate in the antithetic direction and the other in the synthetic direction. The latter would require as input the target structure or the latest level of intermediates on the antithetic tree and one or more starting structures which would be specified by the chemist or derived by another program. The synthetic and antithetic programs would have to communicate to one another

1. Grow tree in antithetic (retrosynthetic) direction (τ_a) from target (T).
2. At some point associate one or more structures (I_n) at lowest levels of τ_a with possible available starting structures (I_o).
3. Grow tree in synthetic direction (τ_s) from I_o *toward* I_n *as a goal.*
4. Alternately extend τ_a and τ_s using latest intermediates as strategy-providing subgoals (*match phase*).
5. Examine intermediates on a linear path from I_o to T in synthetic direction to *optimise ordering of sequence of I's* and to optimise application of *control elements, i.e.*, activating, deactivating, or stereo-correcting operations.

Scheme 3 *Typical course of problem analysis by a synthetic chemist*

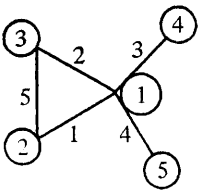
the latest structures generated for their respective trees. The latest intermediates of one program would serve as goals which guide the operation of the other. The realisation of such a scheme is regarded as a major long-term objective of the Harvard program.

2 Internal Representation of Chemical Structure

The utility of connection tables for both bonds and atoms to provide an internal representation of a target structure for synthetic analysis has already been dis-

cussed, and a particular arrangement of these tables has been described. Recently, an improved version of these tables has been devised[6] for LHASA-10 which is illustrated in Table 1 for the specific case of 1,1-dimethylcyclopropane. The atom part of the table (one computer word of 32 bits for each entry) contains for each atom an atom number (sequence of atom input from Rand tablet), the number of attachments, charge, the valence of atom, the atom type (C, N, *etc.*), and a pointer to (relative address of) the first bond entry for this atom. The *x*, *y* co-ordinates of the atom in the external representation are also stored in a

Table 1 *Sample table for* 1,1-*dimethylcyclopropane*



ATNO	NATCH	CHARGE	VALENCE	TYPE	POINTER*
1	4	1	4	4	350
2	2	1	4	4	349
3	2	1	4	4	347
4	1	1	4	4	345
5	1	1	4	4	343

ATNO	BNDNO	BNDTYP	POINTER**	LOCATION
2	5	1	0	341
3	5	1	0	342
1	4	1	0	343
5	4	1	0	344
1	3	1	0	345
4	3	1	344	346
1	2	1	341	347
3	2	1	346	348
1	1	1	342	349
2	1	1	348	350

*First bond entry for this atom. **Bond entry for next attached atom (in number sequence).

second computer word. In the bond part of the connection table, an entry is made twice for each bond between atoms (once for each atom). Each entry (one 32-bit word) for a bond contains the sequence numbers for the attached atoms, the bond number (according to the sequence with which bonds were input from the Rand tablet), the bond type (single, double, *etc.*), and a pointer to (address of) the bond entry for the next attached atom (according to number sequence). Information on up to 64 explicit atoms can be accommodated in

[6] E. J. Corey and D. A. Pensak, to be published.

LHASA-10. Because of the storage of atom and bond data according to sequence, information on each is available without searching. Although information for a structure which has been input is in the form of a connection table, offspring of the parent structure which are generated by LHASA are internally represented by the *physical differences* between the table of the parent and the table that would correspond to the offspring. This approach is very economical in terms of memory. Generation of an offspring structure involves only arithmetic replacement operations, and this constitutes another advantage of the LHASA-10 system over its predecessor.[2]

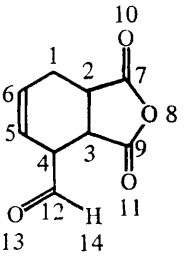
3 Perception

One of the most challenging aspects of developing a program for synthetic analysis by machine is the gathering and storage of the synthetically significant structural information which is required for the parts of the program which select, evaluate, and apply chemistry so as to generate the tree of synthetic intermediates.[2,7] This function, which is essentially a form of perception, is performed by a separate module of the program. The techniques used for machine perception are on the whole very different from those used by a chemist. They are highly formalised in a way which is efficient with regard to machine memory and execution time, and they are applied systematically.

Among the large variety of data generated by the perception module are certain types which are obtained and stored in binary set form. Set information is obtained for atoms or bonds which possess a particular property. Examples of simple atom sets are: BOND1SET (atoms to which at least one single bond is attached), NITROGEN (set of all N atoms), HETERO (set of all N, O, S, P atoms), and RINGSET (set of all atoms in rings). Examples of bond sets are: BOND1 (single bonds), RING5 (bonds in a five-membered ring), RESON (bonds in an aromatic ring), CJBD (multiple bonds in conjugation). Starting with the most basic sets, increasingly complex sets can be constructed by standard set operations. Sets can be manipulated and combined by computer with great facility using basic instructions such as the logical *AND*[8a] or the inclusive *OR*.[8b] In this way sets are obtained which deal with many types of structural features including, for example, those having to do with molecular topology, sites of reactivity, electronic properties, and structural redundancy.[7] Table 2 shows some simple binary sets which are derived by the perception module for a particular structure and indicates the way in which data are 'bit' coded and stored in memory. In LHASA-1 two 18-bit computer words (locations) are used to accommodate each set (up to 36 atoms). The leftmost bit in word 1 corresponds to the first atom (or bond) according to input sequence, and each bit to the right thereafter corresponds to the next atom (or bond) in the sequence.

[7] E. J. Corey, W. T. Wipke, R. D. Cramer and W. J. Howe, *J. Amer. Chem. Soc.*, in press.
[8] (*a*) The result of the *AND* operation between two sets is the intersection of the sets; (*b*) the result of the inclusive *OR* operation on two sets is the union of the sets; (*c*) the exclusive *OR* of two binary sets returns a set whose non-zero elements are present in one but not both of the original sets.

Table 2 *Some sample sets*



SETNAME	WORD1						WORD2
BOND1SET	111	111	111	001	010	000	zero
OXYGEN	000	000	010	110	100	000	zero
BOND2SET (atoms)	000	011	101	111	100	000	zero
BOND2 (bonds)*	000	010	000	011	010	000	zero
RINGSET	111	111	111	000	000	000	zero
JUNCTSET	011	000	000	000	000	000	zero
ALLYLIC	111	100	010	000	010	000	zero

* Bonds also are numbered in the order of drawing.

Illustrations of set operations:

Carbonyl oxygens: BOND2SET *AND* OXYGEN.

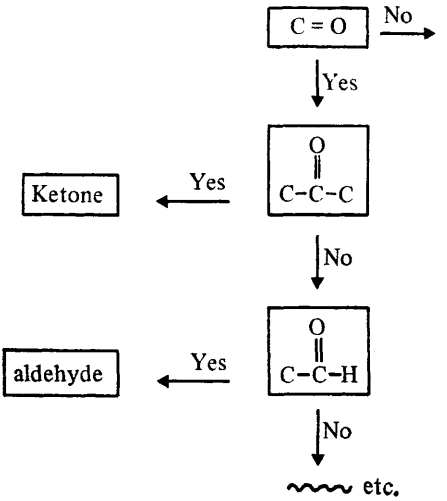
Atoms defining any double bond in a ring: RINGSET *AND* BOND2SET atoms.

The binary digits 0 or 1 indicate, respectively, that the corresponding atom (or bond) is not or is a member of the set in question.

The perception of functional groups has been accomplished by a variety of schemes, and that used by LHASA-1 has previously been outlined.[3,7] In the approach[6] which has been implemented in LHASA-10, all non C—C bonds are encoded (assigned numerical names) and a search is conducted in the order: triple bonds, double bonds (C=O first), then single bonds. The 'recogniser' program then reads a table which contains information on 64 functional groups and which is suitable for binary search. Starting with the bond which determines entry into the table, the group of contiguous bond(s) in the structure is matched against the table. Success or fail pointers then reference the address of the next appropriate table entry. Scheme 4 illustrates this operation diagrammatically for a small portion of the table. The functional groups thus found are stored in list form,[9] each along with the group name, level of reactivity (normal for the group, above- or below-normal), and the atom to which the functional group is attached (group origin).

[9] A list (singly linked) is a collection of elements in memory each of which is composed of two contiguous fields (storage locations). The first field contains datum or a pointer to (address of) a sublist, and the second contains a pointer to the next element on the list. As a data structure the list has the advantage of representing relationships between data as well as storing the data. In addition, the elements in the list need not be stored in contiguous memory locations. See J. M. Foster, 'List Processing', Elsevier, London, 1967, and ref. 3*d*.

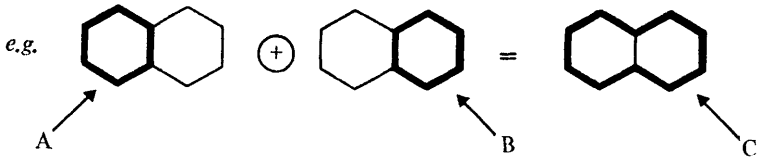
Binary tree search by recognizer:



Scheme 4 *Functional group recognition*

A procedure for the perception of rings has been devised which selects the subset of synthetically significant rings ('synthetic subset') from an *n*-cyclic structure with high computing efficiency even for complex networks.[10] This has now been implemented in both LHASA-10 and LHASA-1 (replacing an algorithm described earlier[2]). Since ring-closure reactions depend on the size of the smallest ring containing the newly formed bond, 'envelope' or 'peripheral' rings must not be present in the synthetic subset of rings. The elimination of peripheral rings can be accomplished by the use of the exclusive *OR* operation on pairs of rings. The algorithm[10] for ring perception which is based on this elimination is summarised in Scheme 6. Definitions for some of the graph theoretic terms used in this summary appear in Scheme 5. The 'synthetic subset' is defined as the set of all minimum spanning rings plus any rings of size $\leqslant 6$.

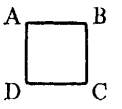
1. *Ring sum* = logical exclusive *OR* of rings ($\oplus$).

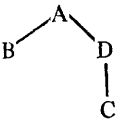


C contains bonds in A or B but *not both*, A $\oplus$ B = C

[10] E. J. Corey and G. A. Petersson, *J. Amer. Chem. Soc.*, in press.

2. *Spanning tree* = acyclic molecular graph corresponding to same cyclic graph.

e.g. for 

, spanning tree = 

3. *Ring-closure bond* = bonds required for conversion of a spanning tree to corresponding cyclic graph.
   e.g. for above example, bond BC.
4. *Fundamental ring* = spanning tree + a ring-closure bond.

Scheme 5 *Perception of rings in polycyclic molecules. Some definitions*

Eliminate acyclic appendage atoms (successively eliminate atoms of connectivity 1); if cyclic order >0:

1. *Grow a spanning tree. Find fundamental rings (FR)*:
   Encountering an atom already in the spanning tree indicates FR.
2. *Remove envelope rings to form reduced basis*:
   For each triplet of rings, R_I, R_J, $R_I \oplus R_J$, retain two smallest rings.
3. *For each ring-closure bond, b_c, find rings containing b_c not larger than FR or largest reduced basis ring:*
   Grow a tree from each end of b_c; a common atom in the two trees indicates a ring; iterate until the smaller of (a) fundamental or (b) the largest reduced basis ring is found.
4. *Order these rings by size and store as bond sets* (**U** FR = {ring bonds}).
5. *Select smallest ring* not in {MSR} with bonds not in **U** MSR and place them in {MSR}; iterate until **U** MSR = **U** FR.

Scheme 6 *Ring perception algorithm*

An indication of the efficiency of the ring perception technique may be found in the performance of LHASA-10 for the case of dodecahedrane, an undecacyclic structure with a total of 1168 possible rings. The synthetic subset which consists of 12 rings (all the five-membered rings) can be found in 0·264 s of PDP-10 time, and only 12 rings need to be grown. The various rings in a synthetic subset are stored in list form,[9] and the atoms and bonds in each ring are stored as sublists of the rings list. This is illustrated by Figure 2 for the specific example of bicyclo[2,1,0]pentane.

Other types of structural information are perceived by the program in considerable number. These include (i) interconnecting paths (*e.g.* between pairs of structural features such as functional groups or asymmetric centres), (ii) appendages on rings or functional group origins, (iii) relative levels of reactivity of each type of functional group in terms of high, low, or normal steric accessibility, electrophilicity, or nucleophilicity, (iv) sensitivity of functional groups to

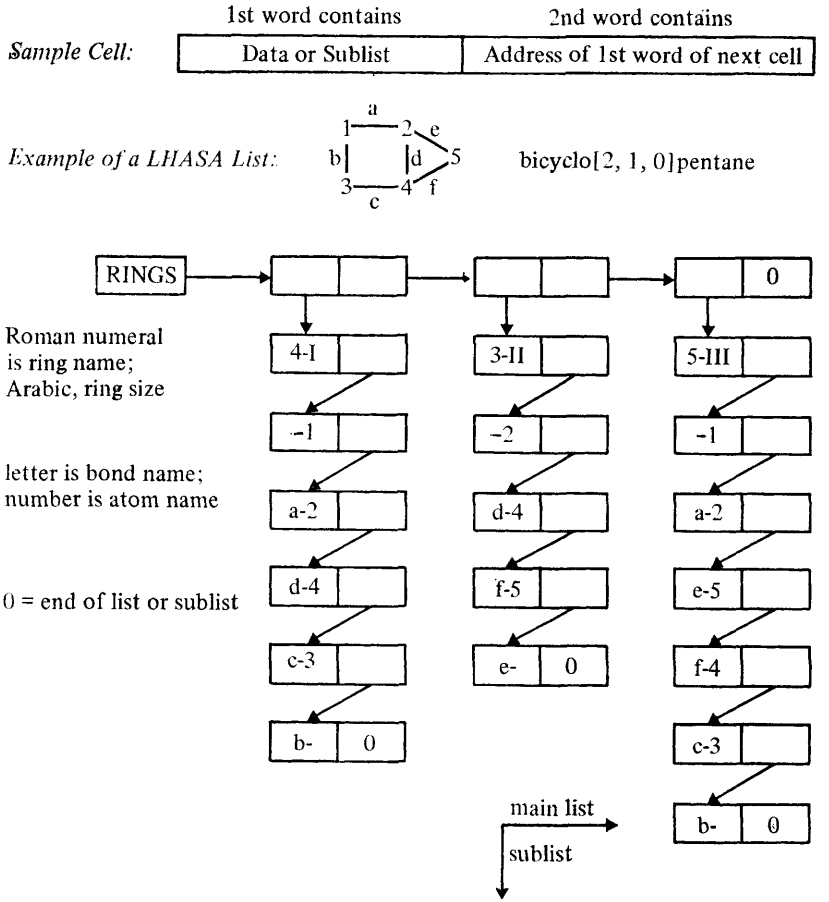


Figure 2 *Linked lists*

various reagents (*e.g.* oxidising, reducing, acid, or base), (v) especially strategic bond disconnections, (vi) aromatic ring systems, (vii) dihydroaromatic ring systems, (viii) asymmetric centres, and (ix) stereorelationships between groups on asymmetric centres. A procedure for the perception of stereochemical features and stereorelationships has already been developed for LHASA,[11] and this will serve as a basis for a stereochemical capability in synthetic analysis.

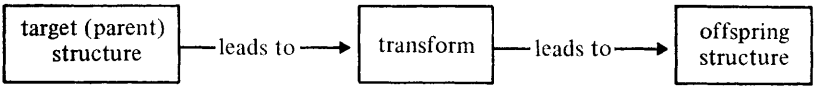
In the presently existing programs the perception of such fundamentally important structural features as are described above occurs prior to transform selection and, indeed, provides the basic information required for transform selection. However, additional perception of a much more varied and context-

[11] E. J. Corey and W. J. Howe, to be published.

dependent sort is needed for evaluation of the suitability of the various theoretically useful transforms. These perceptual processes are carried out by the program at the later stage of transform evaluation, as will be described in the next section. Machine perception therefore plays a key role in strategy selection, transform selection, and transform evaluation.

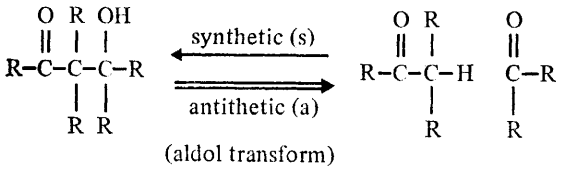
4 Organisation and Utilisation of Chemical Data

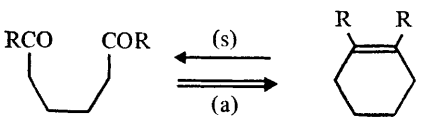
The process of generating an antithetic tree depends upon the recognition of key structural features of a target (or 'parent') structure which signal the applicability of certain transforms to the manipulation of the target structure. Once identified, each of these transforms can then be utilised to derive the structure of the corresponding precursor. The flow of events may thus be regarded as:



The identification of applicable transforms is made on the basis of the target structure and is independent of offspring structure.[12] Clearly then, it is both possible and useful to classify transforms according to the nature of the critical structural features ('synthons')[4] of a target molecule to which the transforms may be keyed. Several important classes of transforms are illustrated by the following entries:[13]

1. Transforms which create two functional groups in the synthetic product (group pair transforms).

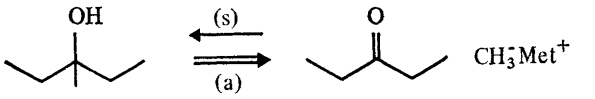




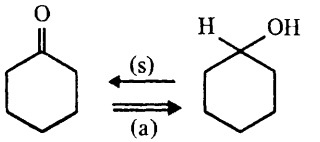
[12] As is indicated in a later section, however, the *evaluation* of the merit of a particular transform in a specific situation is dependent on *both* parent and offspring structures. Nonetheless, since for a particular transform the offspring structure is determined by that of the parent, it is possible to express such an evaluation solely in terms of parent structure.

[13] For a discussion of 'transform'-based data tables and the method by which these are used for the automatic generation of synthetic intermediates by LHASA, see E. J. Corey, R. D. Cramer, and W. J. Howe, *J. Amer. Chem. Soc.*, in press.

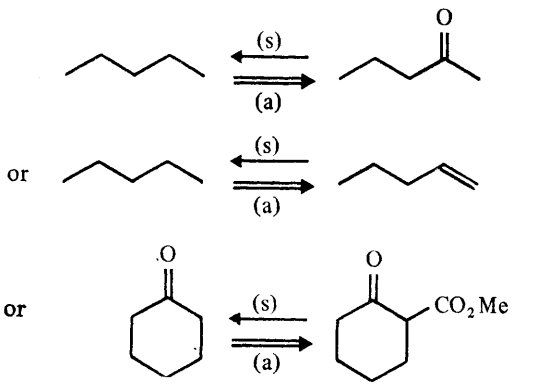
2. Transforms which change one functional group and which also modify the structural skeleton in the synthetic product (single group transforms).



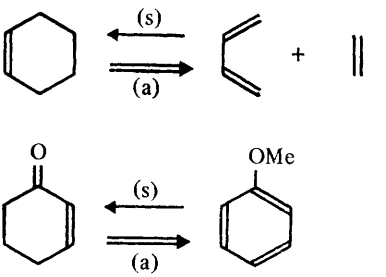
3. Single group transforms which modify only a functional group (functional group interchange or FGI).



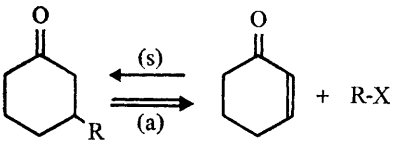
4. Transforms which add a functional group in the antithetic direction (FGA).



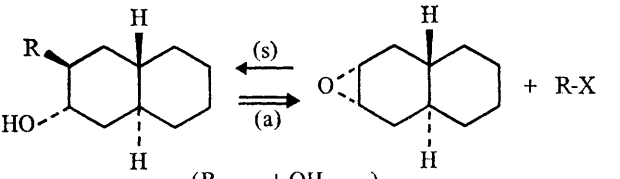
5. Transforms which form **or** modify some particular type of ring (ring or cycle transforms).



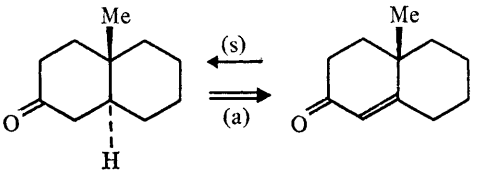
6. General pair transforms (synthetically significant pairs other than FG pairs).
   Examples: FG + appendage, FG + ring fusion, FG + stereocentre.



(C=O + β-R)

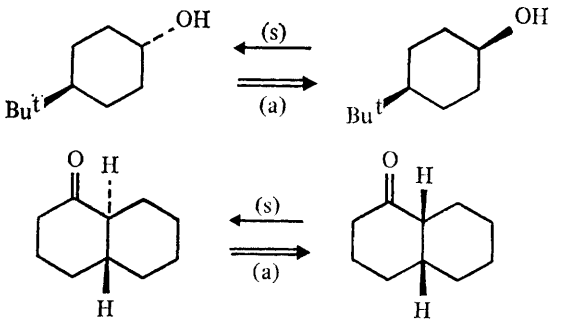


(R_{axial} + OH_{axial})



(C = O + *trans*-decalin)

7. Stereochemical transforms.



Each class of transforms may be subdivided in various ways for purposes of convenient organisation, table searching, or use by a computer program. For example, the functional group pair transform class may be segmented according

to the number of atoms in the path which connects the two functional groups. The 'synthon' for a specific group pair transform, which consists of the group pair and the interconnecting path, likewise can be placed into an appropriate subdivision of the 'two-group synthon class' on the basis of path length. Ring transforms can be subdivided according to ring type—alicyclic, heterocyclic, aromatic, dihydroaromatic, for example—and also ring size. Clearly there are a very large number of important subdivisions in the ring transform class even with only these two distinguishing criteria. Despite the evident proliferation of subdivisions, it is useful to make a number of further distinctions between individual transforms based on other important properties. For example, a stereospecific transform which requires a particular stereochemical arrangement within the target structure should be differentiated from a non-stereospecific transform or even a stereospecific transform for a diastereomeric arrangement. In general, distinctions between transforms may be based on the structural *changes* they effect as well as on the basis of synthon type. Transforms may result in a change of:

(i) molecular skeleton (disconnection, connection, or rearrangement)
(ii) functional groups (addition, removal, interchange)
(iii) stereocentres or stereorelationships (addition, inversion, removal).

The separation of single group transforms into disconnective and FGI classes in the manner indicated above is the result of the consideration of both synthon type and structural change.

On an even more general basis it is important to note that certain transforms simplify molecular structure (in the antithetic direction), whereas others either do not, or actually cause an increase in structural complexity. Transforms of the last two types are obviously useful if their operation results in the generation of structures which then are susceptible to the operation of simplifying transforms. Transforms which simplify molecular structure vary with regard to the degree of simplification which their application produces; those whose operation results in major simplification are clearly more powerful than those which yield only a small decrease in molecular complexity. The Diels–Alder transform is one of the most powerful of all, since its application can result simultaneously in (i) a decrease in the number of rings, (ii) a decrease in the number of asymmetric centres, (iii) disconnection of molecular skeleton to generate two fragments, and (iv) simplification of functionality. (An equivalent statement can, of course, be made concerning the effectiveness of the Diels–Alder *reaction* in *increasing* molecular complexity in the *synthetic direction*.) Pair transforms which disconnect molecular skeleton or remove functionality and/or stereorelationships also are of considerable power, though in general the ring transform group may be regarded as the most powerful. It is evident that one strategy which should be useful in a computer program for synthetic analysis is that of trying to apply the most powerful transforms even though direct application may not be possible for a particular target structure. When the target structure lacks one or more of the features required for the application of a major simplifying transform, the strategy is to define the direct application of that transform as a goal toward

which a number of other steps (*subgoals*) may be tried. These steps will in general involve less powerful or even non-simplifying transforms. Success of the strategy requires the generation from a target structure of a series of intermediate 'subgoal' structures leading finally to a 'goal' structure which allows the effective operation of the major or simplifying transform. This approach is treated in somewhat more detail in the section on strategies which follows. It is a strategy which chemists frequently are influenced by in some measure even though they may not have articulated a completely systematic and formal strategic technique.

The basic organisation of the chemical data tables in LHASA has been formed about the 'transform' as a key element. For *each class* of transform there is a 'data table' which contains an entry for *each transform* within the class. Each table entry contains the following types of information: (i) transform name, (ii) characteristic synthon, (iii) an intrinsic (target-independent) numerical rating of the transform (to be used in transform evaluation), (iv) the bonds within the synthon which are made or broken, (v) a set of conditional statements ('qualifiers') which cause the basic rating to be increased or decreased by certain amounts if certain structural features are present in the particular target structure. These conditional statements, which reflect what is known about the 'scope' of a given transform (or the corresponding reaction), allow the derivation of a rating for a transform as applied to a particular target. This rating is essentially a measure of the probability that the synthetic process corresponding to the transform is a realisable operation.

Within LHASA there exist packages of chemical information which might be termed 'chemistry units', each of which consists of three components.[13]

1. A data table of the type described above which refers to one *class* of transform.
2. A transform-choosing program which matches the features of a target (or 'current target') structure against the data table, evaluates all transforms for which there is a match, and stores all transforms passing evaluation with a rating above a pre-set cut-off value.
3. A transform-executing program which executes stored transforms one-by-one to generate new intermediates in the synthetic tree.

The chemistry unit for the two-group class of transform will now be described to illustrate in somewhat greater detail the operation of that part of LHASA which actually is concerned with the manipulation of chemical structures and the generation of an antithetically directed tree of intermediates.[13] At present there are about 125 entries in the two-group unit in LHASA-1. For each of the $n(n-1)/2$ pairs of functional groups in a molecule of n functional groups, the interconnecting path(s) are determined and the occurrence(s) of full matches with the table entries for each are recorded. Further, for certain of the more powerful transforms in the two-group class, part-matches (one functional group and path, but not the other functional group) are also determined for use later in connection with the application of functional group interchange (FGI) transforms as a subgoal of group-pair transform application. The transform(s)

corresponding to full matches are then evaluated from the 'qualifiers' in the data table to derive a rating for each transform.

The operation of the LHASA scheme for the evaluation of suitability of a transform within the context of a particular target structure is best explained by the use of an example. The aldol transform, being both suitable for this purpose and relatively important within the class, is chosen for the illustration. A very brief summary of the kinds of information found in the table entry for the aldol transform is presented in Scheme 7.[14] As indicated in item 1, the aldol transform is assigned an intrinsic rating of 70 (relative to a cut-off value of −50) and is

Brief summary of table entry:

1. 

   HO at atom (1), W (an electron-withdrawing group) at atom (2)

   2-atom path, bond 1 broken, initial rating 70, try FGI (subgoal flag).

2. '*Standard*' *qualifiers*—statements modifying initial rating according to target structure—combined by inclusive *OR*.

   . . addt 30 / if / grp2 / . . is nitro
   action optype modifier phrase

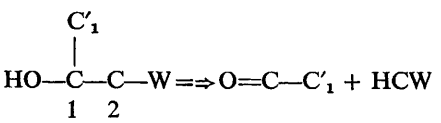
3. *Control phrases* permit qualifiers to interact by modes such as logical *AND*, exclusive *OR*.

4. *Condition statements* pertain to reaction conditions.

Scheme 7 *Table-driven rating of an aldol transform*

designated as a goal in case of a part match *via* the FGI subgoal. The example of a standard qualifier which is given in item 2 instructs the computer to increase the rating by 30 if group 2, the electron-withdrawing (W) group, is nitro. This and other statements in the data tables are written in a 'chemist-oriented' higher level language which is translated by a separate program (compiler) into machine language. The aldol entry in LHASA-1 contains *ca.* 40 standard qualifiers which raise or lower the transform rating by certain amounts. A representative collection of structural features in the target structure which raise or lower the rating is given in Scheme 8. The complete listing of the aldol table entry, which is presented elsewhere,[13] should be consulted for additional detail, including the use of control phrases to permit qualifiers to depend upon certain other qualifiers

[14] This table for the 'aldol transform' corresponds to what would be appropriate for the 'aldol reaction' in the synthetic direction. For the 'retro-aldol transform', which may be defined as the transform corresponding to the 'retro-aldol reaction', an *entirely* different table is required. Also it should be noted that although the cyclic version of the retro-aldol transform is a two-group transform, the non-cyclic version is a one-group transform (connective).



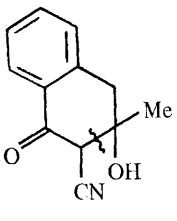
Rating decreased by: Hal, O, or S β to W; W = CONHR, CN, COOR; C=C— on C_1; for each Alk at C_2; W = CH=CHW; presence acid- or base-sensitive groups elsewhere.

Rating increased by: additional W at C_2; W = NO_2; no hydrogens on C'_1.

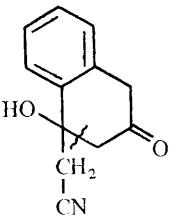
If C_1—C_2 in ring: rating is decreased if ring size other than 5 or 6 and increased for 5 or 6; rating is decreased if W or Alk groups exist at positions (*e.g.* C'_1) which would favour aldol cyclisation to a structure other than target.

Scheme 8 *Rating of an aldol transform*

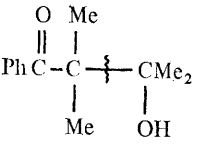
(*i.e.* a 'nesting' of qualifiers). Condition statements play the useful role of allowing the identification of interfering groups remote to the reaction site. The target structures (1)—(4) which are candidates for the aldol transform, since they each possess the required synthon, can be used to exemplify the rating performance of the present version of the aldol table entry. The current LHASA-1 ratings of structures (1)—(4) are, respectively, +300, fail, −30, and fail (cut-off = −50).



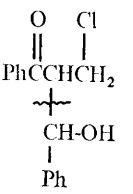
(1)



(2)



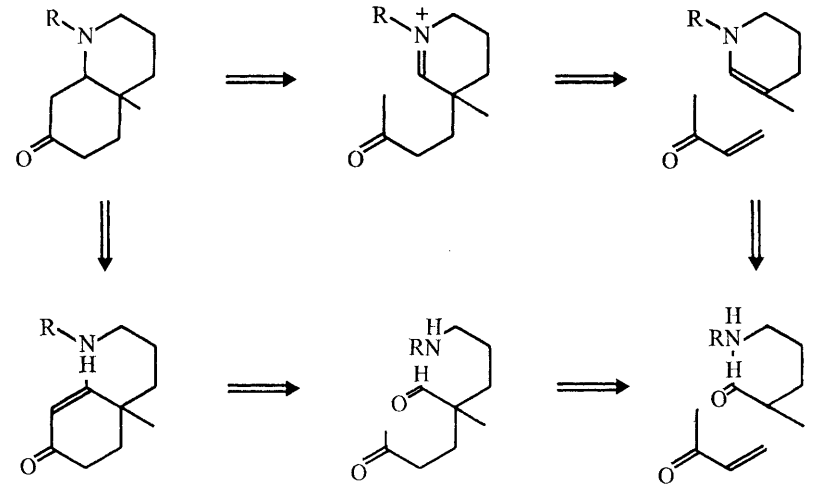
(3)



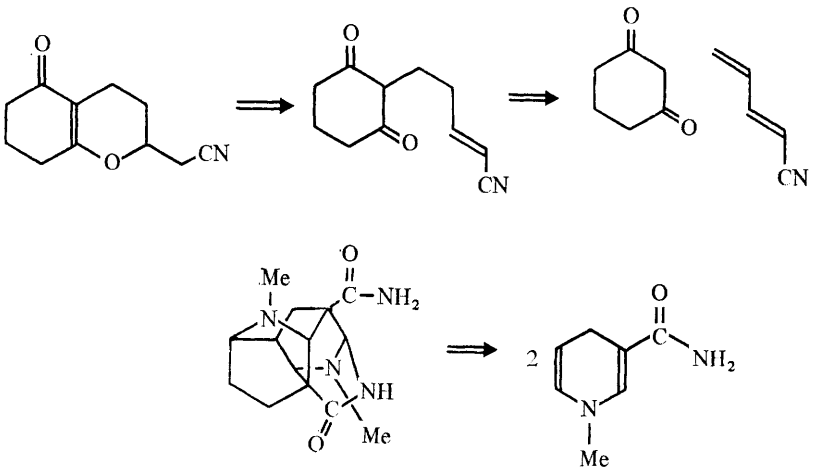
(4)

The ratings are used to circumvent the generation of intermediates corresponding to naïve or highly dubious synthetic processes and also to order the output of structures on a given level of the tree of synthetic intermediates. The intermediates for each level are displayed in order of decreasing rating by machine. The chemist can cause the computer to make any specific further deletions which he deems appropriate by use of the Rand tablet input; further, he can alter the cut-off value of the ratings.

The group-pair transforms are frequently highly effective in the generation of synthetic pathways to a target structure. Many of the published syntheses of alkaloids, for example, can be reproduced by machine using almost exclusively group-pair transforms. The pathways of synthesis shown in Schemes 9 and 10 were derived by computer solely through application of the group-pair chemistry unit.



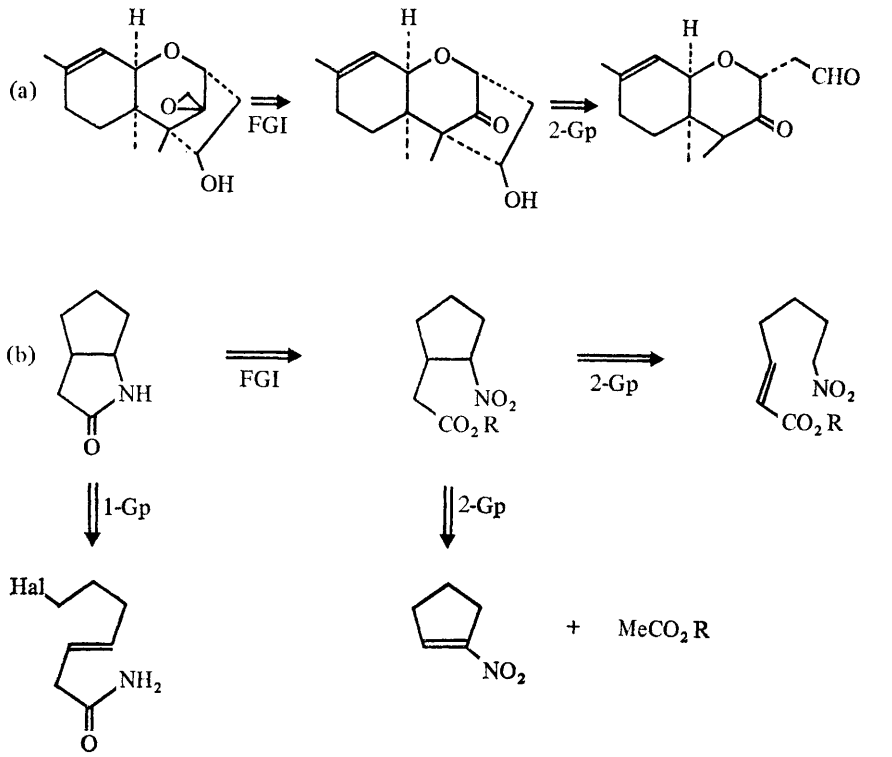
Scheme 9



Scheme 10

When a part match occurs between a specific functional group-pair–path combination in some target structure and a group-pair table entry which corresponds to an important simplifying transform, a request is made to a functional group interchange (FGI) chemistry unit to ascertain whether there exists an FGI transform which can convert the non-matching group into that required

for match to the pair table entry. An FGI transform selecting program scans the FGI data table to find whether the required transform exists and is applicable to the specific target structure. If this subgoal is achieved, FGI transform execution then occurs to generate a new intermediate which is further modified by the action of the pair transform that generated the FGI request initially. Two examples of sequential FGI and group transform application as executed by LHASA-1 are shown in Scheme 11.



Scheme 11 *Application of FGI transforms as a subgoal of pair transforms*

The FGI and functional group addition (FGA) transforms indicated above are applied only as subgoals which allow the utilisation of a simplifying transform, for example, of the pair or ring type. Disconnective single functional group transforms are more versatile. Although they may be used to generate subgoals to satisfy some defined goal, they are also allowed by LHASA to operate *directly* under certain circumstances including cases where the target structure has less than four functional groups or cases where a single group transform effects 'strategic bond disconnection' (*see later*). When the opportunities for the exercise

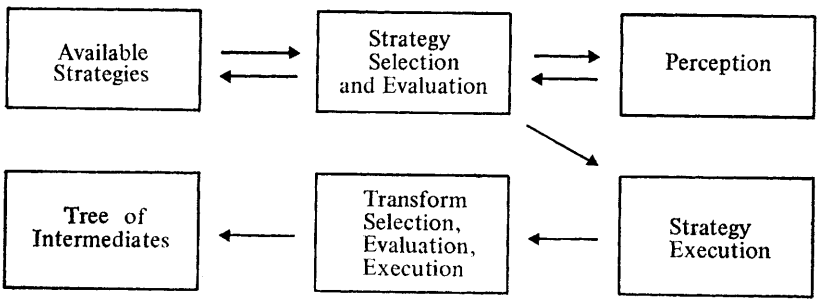
of single group transforms are especially numerous, their use must be controlled by one or more strategies.

The applicability of ring transforms to a particular target structure is normally quite limited (even more so than for group-pair transforms), and there is no difficulty in selecting and executing such transforms where direct matching techniques suffice. However, it is often the case that some ring transform can be applied successfully only after a number of other transform types are utilised to pave the way for a direct match. Some techniques for accomplishing the analysis required for this approach are discussed in the following section on strategies. The use of *general* pair transforms has obvious utility, since it broadens the range of synthons which can be matched against an organised data table and since its application can be accomplished by the same techniques which are used for the group-pair class of transform.

5 Strategy Selection

Probably the most fascinating and exciting area of the theory of synthetic analysis is that which concerns the strategies of synthetic chemistry and their effective use in problem solving or, perhaps more aptly, problem simplification.[4] The creative challenge, the formidable intellectual barriers, and the satisfaction which are associated with the design of a fine synthetic plan are all rooted in the process of devising a good strategy. Although only a brief discussion of this subject can be presented here, a more comprehensive and detailed treatment is planned for a future publication.

Synthetic organic chemistry makes use of a considerable number of different strategies, and it is no mean task to select one which is singularly appropriate to a problem. It is even more difficult to invent a basically new and original strategy in response to a refractory and complex problem. It is not unexpected therefore that the formulation and selection of strategies for a wide range of synthetic problems is the most formidable task in the development of a sophisticated computer program for synthetic analysis. The process of strategy selection and its relationship to synthetic problem solving are outlined in simplified form in Scheme 12. Analysis must start with perception, which in the initial stage is



Scheme 12 *Synthetic problem solving*

systematic and also limited to certain basic features of the target molecule, and with explicit information on the various available strategies. If strategy selection is not to be perception-limited, there must be an additional type of perception based upon (and driven by) the more specialised data required for selection of individual strategies. This situation involving target-driven and strategy-driven perception (effectively a recursive type of perception) is clearly similar to the analog for transform selection and evaluation discussed earlier, although more complex information structures are involved. It is also true that the information content of individual strategies and their evaluation are more complex than is the case for transforms and, of course, our basic understanding of synthetic strategies in generalised form is relatively primitive. The analogy between 'transform' and 'strategy' is, however, useful in helping us to scrutinise the latter, even though it straddles two different levels[4] of problem solving. Like 'transforms', strategies can be allowed to operate independently of one another on a particular target structure; further, the more strategies which are available, the greater will be the power of a general problem-solving procedure. A rational and systematic classification and definition of strategies is fully as essential as in the case of transforms. Finally, just as there is a variation among transforms in their 'power' to reduce molecular complexity, there is an analogous variation of power among strategies. No strategy is universal, some strategies will only rarely be useful, and some strategies are helpful specifically because they remove the obstacles to the application of other strategies. Furthermore, the application of various strategies at more than one level of a synthetic tree means that along one or more vertical pathways in the tree there will be a *de facto* 'nesting' of strategies with those of the upper part of the path being of greater influence and consequence than those applied below.

Many of the most useful synthetic strategies fall into four categories which can be summarised as:

1. Strategies based on particular structural characteristics of a target molecule.
2. Strategies based on the selection of certain key transforms, or more generally, certain chemical information, the application of which becomes a goal.
3. Strategies based on the matched development of 'antithetic' and 'synthetic' trees, or certain assumed starting materials.
4. Strategies based on certain special (but external) circumstances connected with the problem (for example, the desirability of synthesising two or more related compounds *via* a common route or common intermediate).

The last two categories of strategy will not be considered here, since they are outside the scope of the LHASA programs. In the first category are included strategies which deal with the selection of certain structural features for modification because of their reactivity. For example, the presence in a target of one or more functional groups which would be highly sensitive to acids or bases signals the strategy of removal or modification of that group by the application of an appropriate transform, since it would clearly be difficult to carry out a multi-

step synthesis with such a group present. The existence within a target structure of certain functional groups which are very easily interconvertible with others indicates the desirability of considering several 'close relatives' of the target structure (for instance: cyclic ketal *and* ketone; or carboxylic acid *and* carboxylic ester; or lactone *and* hydroxy-acid; or $\alpha\beta$- *and* $\beta\gamma$-enones). In molecules containing a large number of functional groups (especially of the same type, or clustered on contiguous or nearby atoms), the application of transforms leading to reduction in the number of functional groups (for instance, transforms which generate unsaturated or aromatic units by elimination or connective processes) may be heuristically effective in uncovering especially simple synthetic routes. Functional groups which interfere with the operation of important transforms (for example, by their presence within a synthon) and which are detectable by recursive perception are especially deserving subjects for this strategy.

Tricyclic or higher polycyclic ring systems, especially of the bridged ring type, provide an opportunity to apply a network-oriented strategy which also comes under the first category. This strategy depends upon the possibility of identifying certain bond disconnections which are strategic in the sense that they lead to especially simple or accessible ring systems of lower cyclic order. A relatively simple but nonetheless useful algorithm for identifying such strategic bond disconnections for polycarbocyclic systems which has been implemented in LHASA-1[15] is outlined in Scheme 13.[16] This procedure leads to the generation of intermediates in which the following structural features are minimised (m) or avoided (a): (i) appendages (m), (ii) appendages carrying asymmetric centres (a), (iii) rings having greater than six members (m), and (iv) bridged rings (m). Rule 5, which is the only stereochemical provision in the algorithm outlined in Scheme 13, is itself a very powerful strategic guide that is both easy to apply and extremely useful from a chemist's point of view. Surprisingly, this rule and the others given in Scheme 13 have not previously been formulated. An example of

A strategic C—C bond must

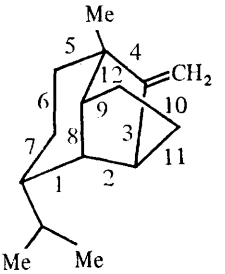
1. be *endo* to a 5-, 6-, or 7-membered ring.
2. be *exo* to a ring larger than 3.
3. be a *perimeter bond.* (The *set of perimeter bonds* is obtained from all pairs of minimum spanning rings by taking the ring sums (logical exclusive *OR*, *XOR*, $R_i \oplus R_j$), or R_iUR_j if R_i and R_j are both smaller than $R_i \oplus R_j$ but $R_i \oplus R_j$ is not larger than 6.)
4. be *endo* to a ring of maximum bridging [*i.e.* ring(s) bridged to max. no. of other rings].
5. not leave stereocentres on side-chains after cleavage.
6. minimise the cyclic order of the largest resulting substructure.

Scheme 13 *Rules for identification of strategic bond disconnections for polycyclic structures*

[15] E. J. Corey and G. A. Petersson, to be published.
[16] For a forerunner of historical interest see E. J. Corey, M. Ohno, P. A. Vatakencherry, and R. B. Mitra, *J. Amer. Chem. Soc.*, 1964, **86**, 478.

the application of the algorithm to the tricyclic molecule sativene[17] is given in Table 3. With a few minor additions this algorithm for strategic bond disconnections can be extended to cover polycyclic systems containing the heteroatoms O, N, and S.

Table 3 *Sample analysis for strategic bond disconnections*



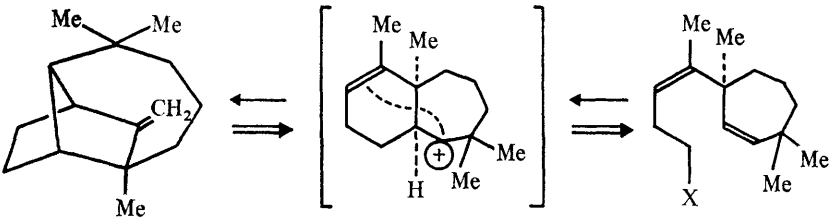
*Rule**	*Bond No.*											
	1	2	3	4	5	6	7	8	9	10	11	12
1	×	×	×	×	×	×	×	×	×	×	×	×
2	×	×	×	×	×			×	×		×	×
3	×	×	×	×	×	×	×		×	×	×	
4		×	×	×				×				×
Strat. Bonds		×	×	×								

* As in Scheme 13.

At this point mention should be made of a most interesting interactive feature which has been incorporated into LHASA-1. Through the Rand tablet and pen device for graphical input together with a graphically displayed control switch accessible on the tablet, the chemist can specify one or more bonds in a target structure as 'strategic' for disconnection. This simple device gives the chemist a powerful and unusual tool for guiding (and experimenting with) the course of synthetic analysis. It can even be used in effect to specify which part of a target structure is the starting point for the synthesis (by the designation of all other bonds as strategic).

The automatic or interactive identification of one or more strategic bond disconnections has an important part in other target-oriented strategies which have been devised. The designation of some bond disconnections as strategic can serve as a guide to enable the selection of one or more single functional group transforms or general pair transforms which break that bond. One of the schemes for synthesis of longifolene as generated by LHASA-1 using this guidance of single group transform selection is illustrated in Scheme 14. If reducible to practice, this would constitute a very simple route to this interesting sesquiterpene

[17] For a simple synthesis of sativene according to the guidance of this strategy, see J. E. McMurry, *J. Amer. Chem. Soc.*, 1968, **90**, 6821.



Scheme 14 *A hypothetical synthetic route to longifolene*

which has already been synthesised by another route[16] which also happens to follow the 'strategic bond' strategy.

Clearly, a designated strategic bond can also be used to direct the *introduction* of a functional group, *i.e.* FGA transform selection, if that bond is not disconnectable by other types of transforms, for example, pair or single group transforms. A less obvious technique based on strategic bond disconnections permits the effective use of double functional group interchange as a subgoal of important group-pair transforms even when there is *no* match between a group-pair in a target structure and a group-pair table entry. The algorithm for accomplishing this double FGI operation for the aldol transform is given in Scheme 15.

1. Strategic bond on path between functional groups G_1 and G_2 in 5- or 6-membered ring?

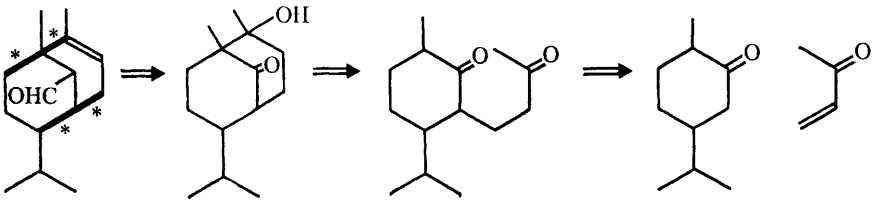
2. If yes, $G_1 \xrightarrow{\text{FGI}}$ OH attached to C(1), where C(1)-C(2) = strategic bond?
3. If yes to (2), $G_2 \xrightarrow{\text{FGI}}$ W attached to C(2)?
4. If no to (2), $G_2 \xrightarrow{\text{FGI}}$ OH attached to C(2)?
5. If yes to (4), $G_1 \xrightarrow{\text{FGI}}$ W attached to C(1)?

Scheme 15 *Double FGI as a subgoal of the aldol transform using an identified strategic bond disconnection*

A specific application of this strategy which leads to an already demonstrated synthesis of the plant toxin helminthosporal is outlined in Scheme 16.[18]

Another set of target-oriented strategies, which in a certain sense represent an opposite of strategic bond disconnections, are those which lead to the goal of applying to a target a connective transform which introduces a bond or a bridge between two of the atoms in the structure. These 'connective' strategies depend for their use on the presence within the target of units such as: (i) asymmetric centres on a functionalised appendage or chain, (ii) a functionalised appendage at an asymmetric centre in a five-, seven-, or higher-membered ring, (iii) two appendages (functionalised or non-functionalised) in proximity and involved in

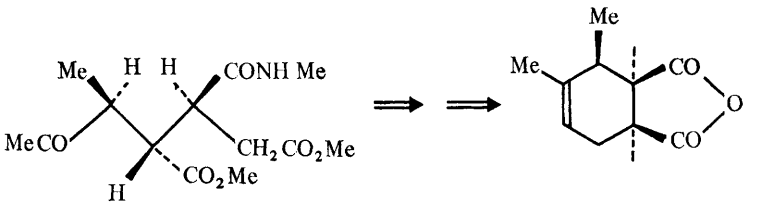
[18] The total synthesis of helminthosporal has been described by E. J. Corey and S. Nozoe, *J. Amer. Chem. Soc.*, 1963, **85**, 3527.



*Strategic bond disconnection

Scheme 16 *Application of double FGI and group-pair transforms to a synthetic scheme for helminthosporal*

non-bonded repulsion, and (iv) a ring of medium size, especially of 8, 9, or 10 members but in certain circumstances also of 7 or 11 members. The major goal of these connective strategies is the generation of six-membered cycles. It will be seen that condition (i) for the application of a strategy of connection is the opposite of rule 5 of Scheme 13 which refers to a disconnective operation. A case which illustrates how effective the connective strategy in situation (i) can be is that of the acyclic ketoamido-diester (5). An outstandingly clever application



(5)

of connection under condition (ii) is found in the recently outlined synthetic approach to vitamin B-12.[19] A fine use of the connective strategy under circumstances of type (iii) is seen in the synthesis of *o*-di-t-butylbenzene.[20] The strategy implied in the application of oxidative C–C cleavage and fragmentation reactions to the synthesis of medium ring compounds falls in the category covered by condition (iv) as, for example in the synthesis of caryophyllene.[21]

Several of the most important target-based strategies are stereochemical in nature. Among the most interesting of these is a strategy being developed for inclusion in LHASA for restricting the order of removal of asymmetric centres from a target structure which contains three or more such centres. Basically the strategy depends upon the perception of stereorelationships between groups

[19] R. B. Woodward, *Pure Appl. Chem.*, 1968, **17**, 519. An especially convenient flow chart of this synthesis appears in the excellent compilation of N. Anand, J. S. Bindra, and R. Ranganathan, 'Art in Organic Synthesis', Holden-Day, San Francisco, 1970.
[20] L. R. C. Barclay, C. E. Milligan, and N. D. Hall, *Canad. J. Chem.*, 1962, **40**, 1664.
[21] E. J. Corey, R. B. Mitra, and H. Uda, *J. Amer. Chem. Soc.*, 1964, **86**, 485.

on asymmetric centres along a stereopath which connects the centres. An algorithm for this strategy is currently under test which directs the preferential removal of terminal or peripheral stereocentres on the stereopath.

A final comment with regard to target-based strategies concerns what might be called *opportunistic* strategies. These are applied whenever some particular structural feature occurs in the target molecule. For example, the group-pair transforms can be utilised opportunistically with considerable effectiveness. The occurrence of group-pair matches is usually sufficiently limited so that the intermediates so generated are not excessively numerous.

At this point it is appropriate to consider the transform-oriented strategies, a class which appears to be of major significance. The algorithm outlined in Scheme 15 for the application of double functional group interchange followed by a disconnective aldol transform in a sense deals with a transform-oriented strategy, although in this case it is one which is keyed by a target-oriented strategy (identification of a strategic bond disconnection). In fact, all transform-oriented strategies depend upon some initial perception of the target structure which serves the function of preselecting one or more transform-oriented strategies for trial. By way of illustration we may consider the preselection and application of transform-oriented strategies for a structure containing a non-aromatic six-membered carbocyclic ring. For this structural unit there are at least four powerful ring transforms which are of sufficient importance to justify trial application even if this would entail a search procedure of some length and complexity which may fail. These are the Diels–Alder, Robinson annulation, Birch reduction, and cation–olefin cyclisation transforms. Each transform which has been identified by preselection is examined separately. In the case of the Diels–Alder transform, an appropriate subclass of the search strategy for this transform is entered according to whether the six-membered ring in question is an isolated ring, part of a fused ring system, part of a bridge ring system, or part of a spiro ring system. Next a rating is derived for the search strategy on the basis of the particular substitution, functionality, and stereochemistry about the six-membered ring in question. This rating determines the maximum depth of search (and number of intermediates) which will be allowed in the attempt to apply the Diels–Alder transform. The search for subgoals follows a binary decision pattern and is table driven.[22] Typically, an entry in the search tree poses a question (*e.g.* endocyclic C=C present?) which leads to one follow-up question if the answer is yes or a different follow-up question if the answer is no, and the process is continued. The elements in the binary search tree refer either to whether some structural feature is present on the six-membered ring in question or whether some transform is applicable.[23] A pathway leading to a success point corresponds to a sequence of transforms which, if applicable, would produce from the target structure an intermediate containing the essential features

[22] E. J. Corey, D. A. Pensak, W. J. Howe and R. D. Cramer, to be published.

[23] Certain parts of the binary decision tree which occur multiply can be handled as subroutines. For example, the subroutine 'epimerise' contains all the necessary decision elements to test for and attempt epimerisation at some centre on the six-membered ring.

required for the operation of the Diels–Alder transform. The accumulated list of transforms on that pathway are then selected for successive application and evaluation. If none of these transforms fail, the goal has been reached of matching the original target to the Diels–Alder transform, which is then applied and evaluated in the normal way.

The above example of a transform-oriented strategy illustrates the current approach of the project at Harvard to one of the most complex and difficult areas of synthetic problem solving. It also provides some grounds for an optimistic view of the eventual possibilities of computer-assisted synthetic analysis as an aid to the chemist and a guide to those who would teach and/or learn this fascinating branch of science. In the final analysis the effectiveness of a general problem-solving computer program must be judged by its performance on a range of specific problems together with the extent to which it includes important chemical information on strategies and transforms. On this basis the performance of LHASA-1 (which is now being equipped with a stereochemical capability) is decidedly encouraging. Among those chemists for whom the program has been demonstrated there is essentially complete agreement on this point and an enthusiasm for the excitement and liveliness which result from the interactiveness and graphical communication which are basic to the functioning system.

Acknowledgment. A major part of the credit for the progress which is outlined in this lecture belongs to a group of unusually able and dedicated research students who have worked on the LHASA and OCSS programs at Harvard. These individuals (in the order of joining the project) are: W. Todd Wipke, W. Jeffrey Howe, Richard D. Cramer, David A. Pensak, Donald E. Barth, and George A. Petersson.

Financial assistance from the National Institutes of Health and the National Science Foundation is also gratefully acknowledged. The Advanced Research Projects Agency provided support in the form of grants to the Harvard Centre for Research in Computer Science and the PDP-1—PDP-10 facility.