Studies in Reaction Path Synthesis

RAKESH GOVIND

Department of Chemical and Nuclear Engineering
University of Cincinnati
Cincinnati, OH 45221

and

G. J. POWERS

Department of Chemical Engineering
Carnegie-Mellon University
Pittsburgh, PA 15213

The development of processes requires the consideration of a wide variety of reactions which are required to transform raw materials into desired products, solve separation problems, and heat or cool process streams. These reactions have to be evaluated on the basis of raw material costs, kinetics, potential separation difficulties, catalyst costs, safety and reliability, as well as other factors that would arise in their industrial implementation. In this paper, the current status of research in reaction path synthesis has been reviewed. Details have been presented on the REACT program that has been developed for the synthesis of reaction paths for industrial organic chemicals.

SCOPE

The synthesis, analysis, and evaluation of reaction paths is of fundamental importance in the chemical process industries. Research and development chemists and chemical engineers are often confronted with problems which are best solved by utilizing chemical reactions. Over the past ten years several researchers have presented systematic techniques to synthesize reaction paths. In this paper a review of the current status of research on the reaction path synthesis problem with details on the REACT program have been presented. The generation of a particular intermediate by the REACT program involves searching a data base of 200 generalized reactions, the search being driven by the structural features of the parent structure and certain basic strategies. The evaluation of each reaction path is based on raw material costs, thermodynamic feasibility, yield estimates and energy costs. The synthesis tree, generated by the REACT program, for Methyl Crysanthemate, an important natural insecticide, has been presented.

CONCLUSIONS AND SIGNIFICANCE

The REACT program can be used for planning the course of future research in an industrial environment. With escalating costs of petrochemical-based raw materials, future production of chemicals would require implementing reaction paths with very different starting materials, conditions and catalysts. Advance knowledge of these reaction paths is crucial in initiating research for industrial implementation. Generation and preliminary evaluation of reaction paths relevant to such an environment is the basic contribution of this research.

The consideration of a wide variety of paths allows long-range planning for changes in product demand, feedstocks, energy costs, new catalyst developments and improvements in separation technology. Reaction paths are currently generated by experienced chemists and engineers. They appear to use knowledge of similar reaction paths, availability of raw materials, named organic reactions and basic reaction mechanisms to guide their search for new reaction paths. These inventors are remarkably adept when one considers the immense number of paths which could lead to even a simple organic molecule.

The problem is to find relevant paths in this multitude which could potentially lead to economical, flexible and safe processes. These paths could then be subjected to more detailed analysis and experiment to further resolve their potential.

This type of analysis gives a basis on which a research and development program can be planned. The candidate paths may require new catalysts, solvents, reaction conditions, materials of construction or different types of separators. In addition, it might be necessary to investigate new raw material sources or develop markets for new by-products (Govind and Powers, 1976).

Correspondence concerning this paper should be addressed to Rakesh Govind.

0001-1541-81-4509-0429-$2.00.

A number of synthetic-design programs have already emerged. A brief description of each program together with appropriate references has been presented in Table 1. Some of these programs were developed primarily for discovering "elegant" laboratory syntheses of complex organic compounds. This has resulted in combining problem-solving procedures of Artificial Intelligence with strategies in the area of synthetic chemistry.

An area of increasing importance is the task of tailoring synthesis of organic compounds to suit industrial implementation. It requires careful and intensive developmental work, with objectives greatly differing from that of a laboratory synthetic chemist. Most industrial chemicals are relatively simple compounds which need to be synthesized with a good overall yield of the product. Emphasis is placed on solvent costs, by-product and side product control, and energy consumed by the reactions. Reliability of the reactions and toxicity of the compounds involved are just some of the considerations that receive priority in evaluating a synthesis.

Our primary objective in developing a synthesis program was from the viewpoint of generating synthetic routes to industrial chemicals. The program was designed to generate a large number of feasible routes, rather than producing only a few possible synthetic routes. This requires generating most of the

TABLE 1. REVIEW OF CURRENT SYNTHETIC-DESIGN PROGRAMS

Program Name	Senior Author's	Molecular Structure Representation	Reaction Representation	Program Language	Brief Description
—	Malcom Bersohn	Connection table with a canonical numbering of the atoms (Bersohn, 1975)	Reactions stored as operations to be performed on the connection table	IBM 370 assembly language	The program accesses a database of 207 synthetic reactions (1976). Program was designed to run without chemist's intervention. It has generated as many as 12,000 nodes.
SECS	W. T. Wipke	Similar to LHASA	Reactions are represented in an English-like language called ALCHEM (Wipke, 1976)	Written in FORTRAN IV Installed on PDP10, UNIVAC, IBM 370, and Honeywell, Bell Computers	SECS (Simulation & Evaluation of Chemical Syntheses). The program has concentrated on the synthetic aspects of stero-chemistry, steric effects, proximity effects and electronic effects. Endowed with the capability of building a three-dimensional model using energy minimization techniques.
LHASA	E. J. Corey W. T. Wipke	Atoms are numbered in order of their insertion by the chemist. There is an atom-atom connection table and a bond table	Reactions are stored in substructure form with conditions stated in semi-English	FORTRAN	The LHASA (Logic and Heuristics Applied to Synthetic Analysis) program uses the logic centered approach to propose synthetic routes. The program interacts graphically with the chemist (Corey at al., 1969, 1972, 1976)
SYNCHEM	H. Gelernter	Molecular structure is represented in two forms: 1) Wiswessels Line Notation (Smith, 1968); 2) Connection table	Reactions are stored in substructure form. Conditions for avoiding a reaction are given via a bit string	PL/1 IBM 360	The program ran without intervention from a chemist, and used yield information to evaluate the generated possibilities.
REACT	R. Govind G. J. Powers	Molecular structure is represented by a linear list of numbers	Reactions are stored in substructure form using a linear list of numbers	FORTRAN IBM 360/67 AMDAHL	The program was written to develop industrial reaction paths from a Chemical Engineering point of view. The program currently has 500 substructure reactions and runs without chemist's intervention.
CICLOPS, EROS, MATCHEM	Ivar Ugi	Molecule is represented in bond-electron connection matrix form	Reaction is represented by a difference matrix obtained from the molecule representation of the reactants		The program is based on a deductive mathematical approach and has inventive capability. However, it is limited to only constitutional chemistry. The original synthetic design program CICLOPS was modified to EROS (Elaboration of Reactions for Organic Synthesis). MATCHEM is a modular System of programs, whose individual parts may be combined in different ways. (I. Ugi, et.al 1974)
SOS	R. Barone	A connection table based on the charge of each atom and its coordinates	Reactions are stored in substructure form, with conditions similar to LHASA	FORTRAN IBM 11/30	The program was written to generate synthetic routes to organic molecules and particularly to heterocycles (R. Barone, 1977)
MASSO	G. Moreau	Based on Hendrickson's representation of a molecular structure	Half Reactions as defined by Hendrickson (Hendrickson, 1977) and are coded in digit notation		The aim of this program is to suggest ideas for the synthesis of a target molecule. The solutions appear on the CRT screen in the form of formal mechanism, with no evaluation. It cannot build synthetic trees (Moreau, 1977)
—	G. Kaufmann	Similar to the SECS program (G. Kaufmann, 1977)	Based on the SECS program		Based on the SECS program, the program has been developed with the aim of enabling synthetic planning in organophosphorus chemistry.
—	Howard W. Whitlock, Jr., Paul E. Blower, Jr.	Two word package called a "cell" (Blower, 1975)	Symbolic representation of the Reactant and product substructures together with the name of the reaction	LISP	Based on the heuristic programming approach the program generates synthetic pathways for a class of linear organic molecules.

synthesis tree to available starting materials, with emphasis on some of the factors outlined in the following section.

REACTION PATH SYNTHESIS

The challenge in systematic reaction path synthesis is to generate reaction paths and steps which could be viable alternatives in an industrial environment. The reaction paths used in industrial environments have the following characteristics:

Small TARGET Molecules. The basic petrochemical and fine chemical industries commonly deal with fewer than 20 hetero- and carbon atoms.

Multiple Target Molecules (Reaction Networks). The chemical industry is a network of reactions fed by three to five basic

Main Reactions at Other Sites

Parallel	$R_1 + R_2 = P$	:	$R_1 + R_2 = BP_1$	
Serial	$R_1 + P = BP_2$	:	$R_1 + BP_1 = BP_3$	

Other Reactions that Occur at the Same Conditions

$R_1 + R_2$	$R_1 + R_1$	$R_1 + P$
$R_2 + R_2$	$R_2 + P$	$R_1 + BP_1$
$P + P$	$P + BP_1$	etc.

Impurities in Reagents

$I_1 + R_1$	$I_1 + R_2$	$I_1 + P$
$I_1 + I_1$	$I_1 + I_2$	$I_1 + BP_1$
$I_2 + BP_1$	etc.	

Figure 1. Reactions which could occur between the species (R = Reagents, P = Products, I = Impurities, BP = Byproducts) in a reaction mixture.

materials and producing hundreds of target molecules. The paths of each target molecule interact with each other by sharing raw materials and byproducts.

Stoichiometry Must Be Known. The stoichiometry for the paths must be known in detail. This means that all main reaction products must be considered at each reaction step. The fact that 2 mol of NaCl or other simple reaction products are produced during a reaction can have a major impact on the reaction's economics.

Yield. The fraction of the limiting reagent which is transformed into the desired target molecule must be known.

By-Products. The amount and type of the by-products produced are necessary pieces of information. The main reaction by-products and side reaction products are needed for both the yield calculation and the determination of separation difficulty, corrosiveness, safety, etc.

Impurity Reactions. The large-scale production of molecules demands a knowledge of the fate of impurities which enter with the raw materials or are produced in the reaction steps. These impurities can cause considerable pollution and product quality problems. In many processes, more equipment and energy is devoted to controlling the impurities than the main reagents and products. Figure 1 illustrates the type of reactions which could occur between impurities and other species in the reaction mixture.

Reaction Conditions. All of the factors discussed above depend on the reaction conditions. It is necessary to have information on phase(s), solvents, catalysts, temperature, pressure, concentrations, mixing, etc. Can modern theories and applications of reaction path synthesis make any contribution to this problem? Is too much specific data required?

In the following sections, we present our research in computer-assisted reaction path synthesis. The work is compared to the needs of industrial reaction path problems. Finally a program called REACT, which we have developed for reaction path synthesis is described and illustrated.

Representation of Molecules and Reactions

One key element in developing a reaction path synthesis program is the selection of representations of molecules and their reactions which are appropriate for a given problem. Figure 2 conceptually compares the predictive power and generality of three representations. Predictive power is defined as the ability to predict the conditions, stoichiometry, by-products, yield, etc., of a given reaction step in each representation. Generality is the ability to generate "all" possible reactions.

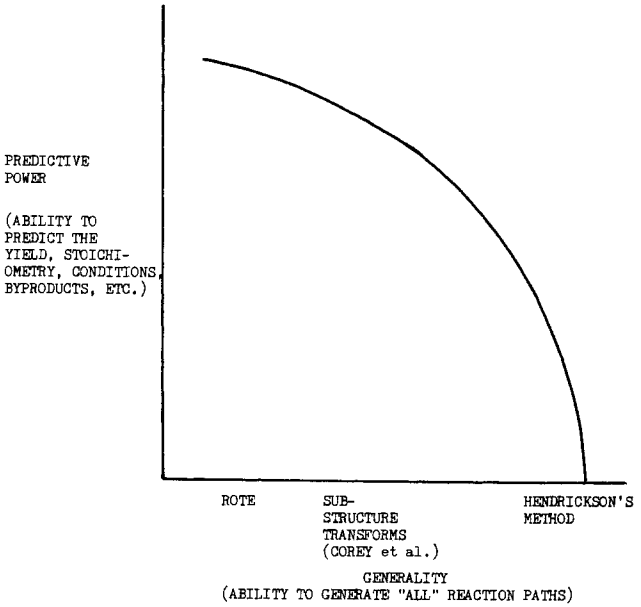


Figure 2. Ability to predict performance vs. generality for three representations of reaction.

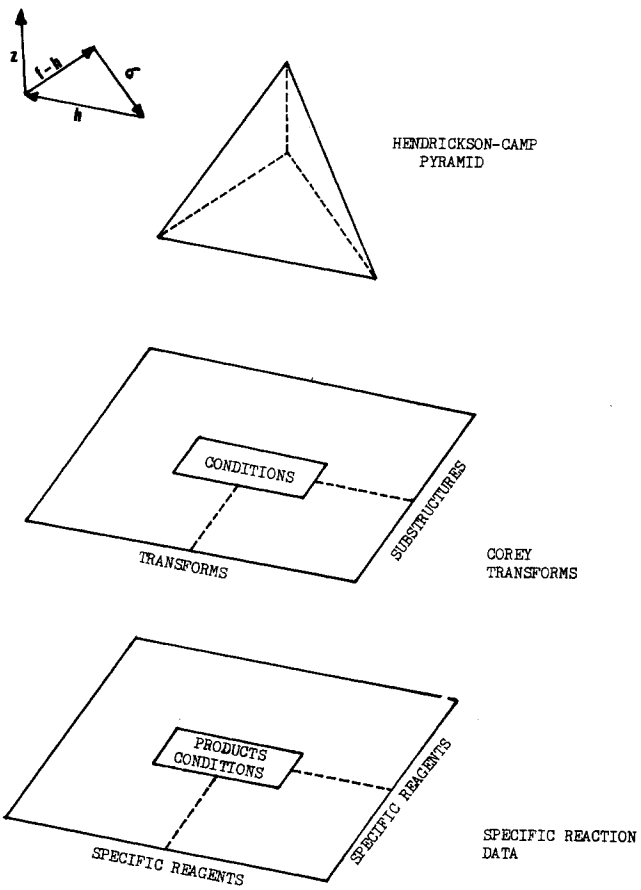


Figure 3. A schematic drawing of three representations of molecules and their reactions.

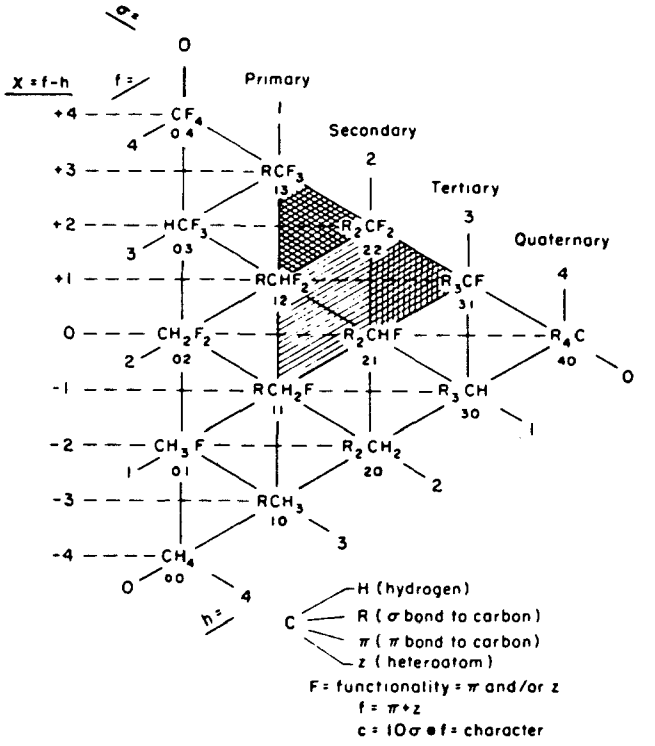


Figure 4. Character triangle for carbon sites and interconversions. The character C, is shown beneath each carbon site. The shaded areas indicate possible double bond sites (C=C)□; possible triple bond sites (C≡C)□; and possible aromatic sites□.

The three representations are shown schematically in Figure 3. The most general types of representations use a mathematical abstraction of a reaction which considers all possible means for making and breaking bonds between atoms. Hendrickson (1976) has investigated a general representation in which molecules are represented by the "character" of the carbon sites which occur within the molecule. The character of a site is a two digit number which indicates whether a carbon atom is attached to other carbon atoms by sigma bonds or to "functional" groups. The character of a site is given by:

$$c = 10\sigma + f \tag{1}$$

where

$$f = \pi + z \tag{2}$$

The functionality of a carbon site is defined as the number of π bonds (π) and heteroatoms (z) attached to the carbon. Valence constraints give

$$\sigma + \pi + h + z \leq 4 \tag{3}$$

where h is the number of bonds to hydrogen or less electronegative atoms.

With this definition of a carbon site, Hendrickson was able to classify all possible carbon sites and the "reactions" which might interconnect them. Figure 4 gives the reaction triangle for this representation.

Camp (Camp, 1974) extended this representation by allowing the functionality (f) to be separated into π and z dimensions. The character of a site then becomes:

$$c = 100\sigma + 10f + z \tag{4}$$

In addition, Camp developed the site characteristics for heteroatoms. These extensions transform Hendrickson's triangle into a pyramid. It is this representation that is shown in Figure 3.

The advantage of this representation is that it is very compact. Hendrickson's triangle has only 15 sites and 70 "reactions". Camp's pyramid has 24 sites and 132 reactions. This representation is also able to generate "all" possible precursors by using the reactions on the pyramid. Each site in the target is developed independently but certain dependencies (e.g., σ bonds between atoms) have to be checked. In essence, this representation involves making (or breaking) every possible bond in the molecule subject to the valency and site constraints.

This generality is obtained with the loss of power to predict the consequences of each reaction step. The reactions are not classified by the exact nature of the functionality. All reactions with a functionality of $f = 1$ (alcohol, ether, olefin) are treated similarly. In addition, some of the reactions are not known to have analogs in nature.

Hendrickson's representation is an example of a generate-and-test problem solving approach in which the generator is uninformed. It simply generates possible reactions, and it is up to the tester (evaluator) to determine the feasibility and value of each step or path.

A reaction path synthesis program was developed based on Camp's generalization of Hendrickson's representation. The program, designated REPAS, was used to generate paths for a wide range of industrial molecules. For a simple molecule, like Acrylic acid, well over 500 reaction paths were generated. The evaluation of these paths was a difficult task and while a few of these paths provided interesting alternatives, there was a general lack of correspondence between these paths and those which experienced chemists and chemical engineers might execute in the laboratory, pilot plant, or industrial process. Our inability to be specific about each reaction step greatly limits the usefulness of this representation. The overwhelming number of combinations of functional groups and reactions severely limits the probability of finding a new and realizable reaction path. A representation of reactions is needed which is perhaps less general but has more specific chemical data and theory in it.

Corey and Wipke (1969) have developed a representation of molecules and reactions based on the combination of atoms in the molecule which change during known reactions. Molecules are represented by linked data lists. The lists contain the complete topological and atom type information for the molecule. Reactions are coded reterosynthetically (called transforms) using substructures which change during the reaction. Such a classification of reactions allows the transform to be applied to a wide range of molecules which might contain these substructures.

Corey et al., subdivided these transforms into classes:

I. Transforms which require in the parent structure a pair of functional groups connected by some atom-bond path.

II. Transforms which require a single functional group.

III. Transforms which depend on ring size with or without functionality.

Each of these may be further classified depending on whether the transform: (a) disconnects the path between functional groups; (b) forms a new path (ring formation, cleavage, etc.); (c) modifies functionality without altering path; (d) effects rearrangement by overall disconnection and formation of two or more bonds.

With each transform, there are checks which prohibit the application of the transform to cases where the reaction would actually be successful or where it would not be competitive with other reactions which might take place under the same conditions. If the local atoms prohibit the particular transform, the reaction is either killed (is no longer considered) or a subgoal is set up to change the atoms which hinder the reaction.

Corey and co-workers have developed a library of over 500 transforms of this type. Most of the checks on the reactions are knowledgeable generalizations of published reviews (March, 1965; House, 1972; Buehler and Pearson, 1970).

In the REACT program, a molecule is represented by a connection table (CT) defined as follows:

TABLE 2. CODE NUMBERS AND CHARACTERS THAT ARE USED BY THE PROGRAM TO CONSTRUCT A CONNECTION TABLE.

Atom/Group/Bond	Code Number	Character
Single Bond	1	—, | , /, \
Double Bond	2	=, ‖
Triple Bond	3	≠, %
Hydrogen	1	H
Carbon	4	C
Nitrogen	5	N
Oxygen	6	O
Halogen	7	X
Aromatic	2	A
Alcohol	8	&
Nitro	9	@
Withdrawing	24	W
General Group R	30	R

$$CT = \begin{cases} \text{atom type/group type, if } i = j \\ \text{bond type, if } i \neq j \end{cases}$$

The atom or group type is a code of numbers which designates the respective atom or group. Table 2 gives the code numbers that are used by the program. The use of a single number to represent a group of atoms provides a powerful way of storing information about the structure. The bond type indicates the bond order—1 for a single bond, 2 for double bonds and 3 for triple bonds.

A connection table contains the relevant information—connectivity, atom and bond types. All relevant information is obtained from the table by subroutines in the program.

Since the connection table is diagonally symmetric, it is stored in computer memory as a list which is the upper triangular of the table. Figure 5 gives examples of this representation for some simple molecules.

In the REACT program, each transform in the database is represented by its relevant substructure which participates in the reaction. A transform is designated by the connection table of the product substructure and the changes that are required to obtain the connection table of the reactant substructure. Figure 6 gives two examples of this representation scheme.

It should be noted that the change matrix is not merely an algebraic difference between the connection table of the reactant and product substructures. It is obtained from consideration of the mechanistic changes that occur in the course of reaction. As shown in Example 1 of Figure 6, the nitrogen atom leaves as ammonia in the reaction. This requires that the atom be disconnected from the product substructure and the two carbons reconnected to form the reactant substructure. These are the changes that are reflected numerically in the change matrix.

The connection table of the product substructure and the change matrix is stored in their upper triangular form as linear lists. These two lists designate a transform structurally in the program database.

In addition to these lists, each transform representation has checks which decide the applicability of the transform. These checks have been basically organized along the lines of Corey's checks discussed before. However, they have been considerably expanded to include special cases which become predominant in a free running synthesis program. The checks, as listed in Table 3, have been generalized. Each transform accesses these checks, the outcome being logically ANDed or ORed depending on the situation. A detailed discussion of how these checks are represented and applied for each transform will be included in the transform selection section.

Each transform in the database also contains side products/side reactants that are evolved or required when the transform is implemented in the laboratory. These side products/reactants are required to stoichiometrically balance the reaction. In addition, they are important when the economics of the reaction is being considered.

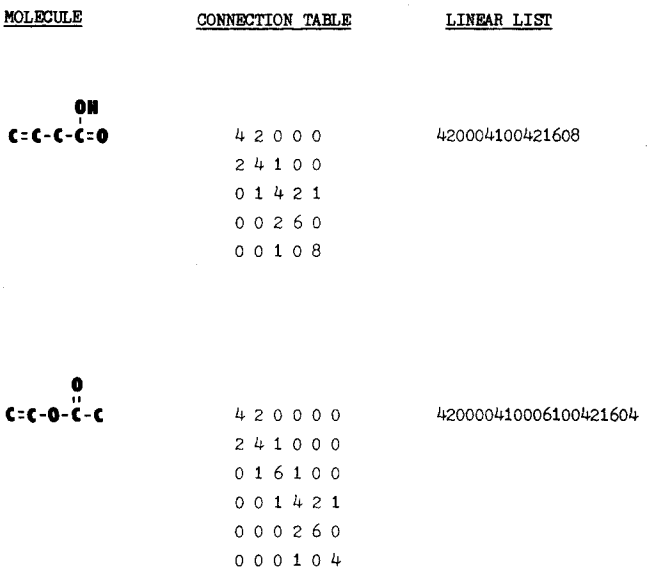


Figure 5. Examples of the connection table representation scheme for some simple molecules.

Normal conditions under which the reaction is conducted in the laboratory/industry are part of the transform database. These conditions give an indication of potential solvents, catalysts or additional chemicals that might be required for the reaction. A list of references where detailed information about the reaction may be obtained are also included in the database. An example of a transform without its checks is given in Figure 7.

Representation of reactions by transforms has considerably more predictive power than the Hendrickson's method. It requires a larger database and the generality of the approach is

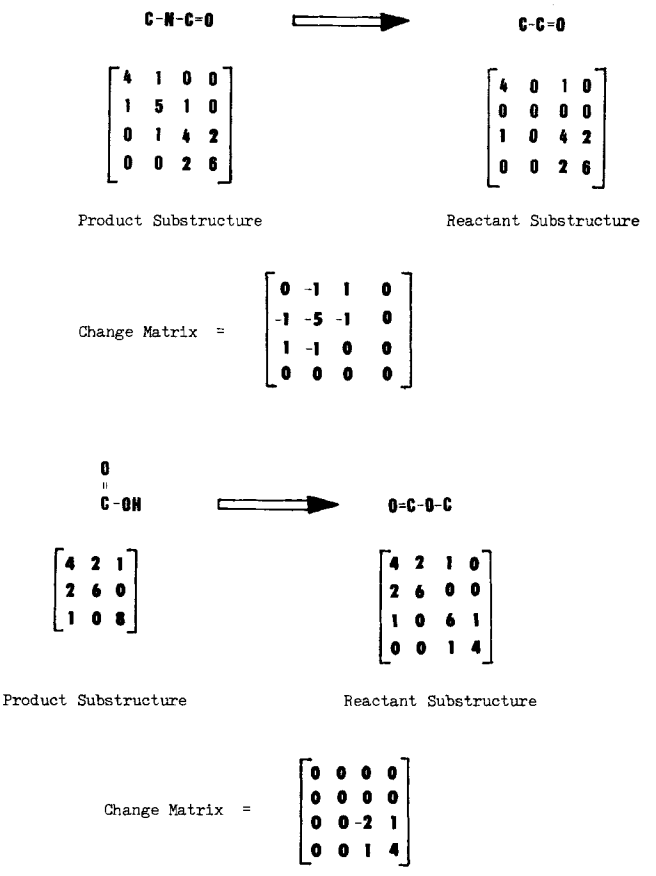


Figure 6. Two examples of reaction representations by their product substructure and change matrices.

TABLE 3. A LIST OF LOGICAL FUNCTIONS USED IN SECONDARY CHECKS

1. SEC 1(A) is true if offpath hetero atom attached to atom A.
2. SEC (A,B) is true if any offpath carbon on atom A has group code B.
3. SEC 3(A,X) is true if SEC. X is true for any offpath carbon neighbor of atom A.
4. SEC 4(A,B) is true if atom A is in a ring of size B.
5. SEC 5(A) is true if no hydrogen on atom A. Atom A is assumed to be carbon.
6. SEC 6(A) is true if atom A is in a ring of any size.
7. SEC 7(A) is true if atom A is part of a functional group.
8. SEC 8(A) is true if there is a leaving group on any offpath carbon attached to atom A. Atom A is assumed to be carbon.
9. SEC 9(A,C) is true if the flagged† (a) substructure at atoms A and C have the symmetry which would result from the aldol condensation of two identical molecules of an aldehyde or a ketone, where A is the carbonyl carbon and C is the alcohol carbon of the product.
10. SEC 10(A,B) is true if bond (A,B) is in a ring.
11. SEC 11(A) is true if atom A is primary.
12. SEC 12(A) is true if atom A is secondary.
13. SEC 13(a) is true if atom A is tertiary.
14. SEC 14(A) is true if atom A is quaternary.
15. SEC 15(A) is true if atom A is part of an offpath multiple bond.
16. SEC 16(A,B) is true if atom A has group code B.
17. SEC 17(A,B) is true if the fragment of molecule M at atom A is identical to the fragment at atom B. No bonds of the current substructure are included in either fragment.
18. SEC 18(A,B) is true if the fragment at any offpath atom alpha to atom B is identical to the fragment at atom B. No atom (except A) or bond of the current substructure are included.
19. SEC 19(A,B) is true if any offpath carbon alpha to atom A is enolizable.
20. SEC 20(A,B) is true if cation better on atom B than on atom A, or it is equally good and there is no match beween the fragments at atoms A and B.
21. SEC 21(A,B) is true if borane addition is better at atom B than atom A, or it is equally good and there is no match between the fragments at atoms A and B.
22. SEC 22(A,X) has value .NOT.SEC.X(A).
23. SEC 23(A) is true if atom A is in a ring of size 5 or 6.
24. SEC 24(A) is true if atom A is in a ring of size 5 to 7.
25. SEC 25(A) is true if MATCHS† (b) (A,ALPHA(B)) is true for every ALPHA (B) with the same number of carbon neighbors as A.
26. SEC 26(A) is true if atom A is alpha to an electron withdrawing group.
27. SEC 27(A) is true if carbon atom A is allylic.
28. SEC 28(A,X) is true of SEC.X is true for any atom which is alpha offpath to atom A.
29. SEC 29(A) is true if atom A is the distinguished†(c) atom of a functional group.
30. SEC 30(A) is true if the number of carbon neighbors atom A is B.
31. SEC 31(A,B) is true if the number of carbon neighbors of atom A is greater than those of atom B.
32. SEC 32(A,B) is true if the number of carbon neighbors of atoms A and B are equal.
33. SEC 33(A,B) is true if the number of carbon neighbors of atom A is less than those of atom B.
34. SEC 34(A,B) is true if the number of carbon neighbors of atom A is greater than that of any atom alpha-offpath to atom B.
35. SEC 35(A,B) is true if the number of carbon neighbors of atoms A and alpha (B) are equal. This function assumes that there is exactly one carbon alpha offpath to atom B.

limited by the transforms and the checks made on each reaction. If the substructure transform is not in the database it will not be included in any reaction path generated using this approach. If the checks are too conservative, the reaction may not be used when it actually might work. It might be possible to systematically relax the local constraints in some future program. These reaction steps would then be the subject of a research and development program to see if they produce quantitative yields.

The most powerful predictor and least general approach is that of simply storing specific instances of known reactions along with the conditions under which the reactions were performed. The particular target molecule is then simply pattern matched with reactions in the database.

```
96.  ALKYLATION WITH ORGANOBORANES GENERATED FROM GRIGNARD REAGENTS.

  W-CH-C   =>   W-CH-X + C-X

  REACTIONS-

  1.                ET2O,REFLUX
      C-X + MG      ----------->  GRIGNARD REAGENT

  2.  GRIGNARD + (9-BBN)-BOET    ------>   ALKYL-B(9-BBN) + MG(OET)2
      REAGENT

  3.                                1.THF,0 C
      ALKYL-B(9-BBN) + W-CHX        --------->   W-CH-C + NAX + (9-BBN)-B-OET
      + SODIUM 2,6-DI-T-BUTYL       2.ETANOL
        PHENOLATE

  REFERENCES-
  1. FIESER V.3,P.234.
  2. H.C.BROWN,H.NAMBU AND M.ROGIC,JACS 91,6852,6854,6855(1969).
  3. H.C.BROWN AND H.NAMBU,JACS,92,5790(1970).
  4. H.C.BROWN AND G.KRAMER J.ORGANOMETALLIC,CHEM.73,1,(1974).
  5. H.C.BROWN AND M.ROGIC JACS,91,4304(1969).
```

Figure 7. An example of a transform datablock.

TABLE 4. LIST OF THE FUNCTIONAL GROUPS RECOGNIZED BY THE PROGRAM.

Acetal	Epoxide
Acetylene	Ester
Acid	Ether
Acid Halide	Halo olefin
Alcohol	Hydrazine
Aldehyde	Hydrazone
Alkyl Halide	Imine
Alkyl Nitro	Ketal
Amide	Ketone
Amine	Nitrile
Anhydride	Nitro Alkyl Halide
Azide	Nitro olefin
Diazo	Olefin
Enamine	Oxime
Enol Ether	

Selection of transforms from the database for a particular target molecule requires perception of key structural features like ring size, groups present, position of groups, etc. These features are accessed when checks are conducted during the process of transform selection. In the following section, molecule perception as programmed in the REACT program is described.

Molecule Perception

The process of perception begins with the input of the target molecule. The molecule is simply typed in at the terminal in the same form as it would be drawn in a chemistry textbook. Special characters are used only when the actual character is not available on the terminal print. The characters used to enter the input molecule have been listed in Table 2.

The target structure is scanned to find the type of atoms, bonds, and the overall connectivity of the molecule. This information is used to create the connection table for the target.

The key structural features that are perceived by the program are: (a) ring size and ring atoms; and (b) type and location of groups. The ring perception module uses a standard loop finding algorithm, starting from a given atom and stepping on the structure using connectivity information. If in this process, the starting atom is encountered, a ring is detected in the structure. A backtrack algorithm is used to assure that all the atomic sites have been visited.

Recognition of groups involves a process of pattern matching of the group substructure with substructures on the target molecule. Table 4 lists all group substructures that are presently recognized by the program. Position of groups on the target molecule are stored for use by the transform selection module of the program.

Transform Selection

Transform selection is a three step process. The first step is to find all possible different ways in which the transform substruc-

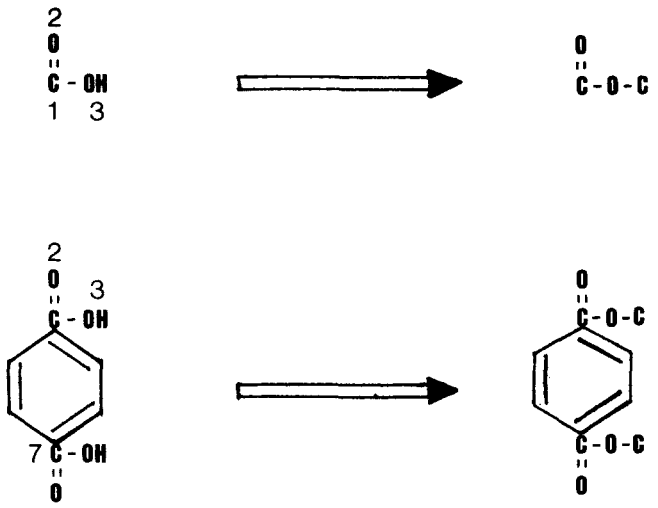


Figure 8. Example of multiple matches of a hydrolysis transform to Pthallic Acid.

ture can be matched with substructures of the target molecule. This list is created using a pattern matching algorithm. Figure 8 gives an example of multiple matches of a hydrolysis transform to Phthallic Acid. In the first match atom 1 of the target molecule matches atom 1 of the transform pattern and so on, while in the second match atom 7 of the target corresponds with the first atom of the transform.

As illustrated by this example, two distinct matches produced by the pattern match may lead to identical reactions. These redundancies arise because the target molecule has an element of symmetry. Detection of symmetry is an important feature in the program and is achieved by matching one appendage of the target molecule to another using the pattern matching algorithm.

The second and third steps are concerned with deciding whether the transform represented by any substructure match can be expected to be successful in a particular situation of a target molecule. In the second step, the program examines the atoms of the substructure and those adjacent to it. Associated with each transform is a list of checks which have to be applied on the target molecule. These checks have been classified on the basis of:

(a) type of functional group to which an atom belongs;
(b) the presence and size of a ring;
(c) the substitution pattern of an atom or pair of adjacent atoms;
(d) the presence of identical molecule fragments.

Specific examples of secondary checks are presented in Tables 5 and 6. Table 5 is a listing of the secondary checks for the Hydroboration-Alkylation transform and Table 6 are those for Allylic oxidation with Selenium-di-oxide.

To facilitate the writing and storage of the secondary checks, the program has a group of 38 logical functions which test for the presence or absence of a structural property relative to an atom or pair of atoms. These functions are listed in Table 3. A secondary check consists of some logical combination of these functions, and is coded as a linear list. A simple interpreter is used to read the numerical strings.

In the third step, the program determines whether the functional groups of the target, which are not contained in the substructure under consideration, are stable to the reaction conditions. For this, the program has a reagent table similar to the one described by Corey et al. (1976). The output from this step is a list of functional groups determined to be unstable to the reaction conditions.

The selected transforms are then applied to the target molecule to generate precursors. However, even with a transform database of modest size, the program will generate a very large tree. Strategies for pruning the tree are essential to control

TABLE 5. SECONDARY CHECKS FOR THE HYDROBORATION-ALKYLATION TRANSFORM. CNUM IS THE NUMBER OF CARBON NEIGHBORS ADJACENT TO ITS ARGUMENT ATOM.

Substructure: W — C_1 — C_2 — C_3

1. Kill, if W is not an ester or a nitrile.
2. Kill, if there is no hydrogen attached to C_1, C_2 or C_3.
3. Kill, if C_1 is part of any functional group and not an alkyl halide.
4. Kill, if C_2 is part of any functional group.
5. Kill, if bond (C_1 — C_2) is in a ring.
6. If C_3 is part of any functional group,
   a. Pass, if C_3 is an ether or a tertiary amine.
   b. Kill.
7. Kill, if there is a hetero atom attached to any carbon alpha to C_3.
8. Pass, if CNUM (C_3) → CNUM (C_2) — 1.
9. Kill, if CNUM (C_3) ← CNUM (C_2) — 1.
10. Kill, if the fragment at C_2 is not identical to the fragment at C_3.

the program's storage requirements. The following section describes methods for strategic development of the synthesis tree and subsequent evaluation of the generated reaction paths.

Strategy of Transform Application

Since we are interested in finding a large number of potentially good routes, there is a need for a high level strategy of

TABLE 6. SECONDARY CHECKS FOR ALLYLIC OXIDATION WITH SELENIUM-DI-OXIDE. (A) CNUM IS THE NUMBER OF CARBON NEIGHBORS ADJACENT TO ITS ARGUMENT ATOM. (B) ALPHA (A) IS THE CARBON ATTACHED TO A WHICH IS NOT PART OF THE SUBSTRUCTURE.

Substructure: C_1 = C_2 — C_3 — OH

1. Kill, if C_3 is not alcohol.
2. Kill, if C_2 is not an olefin.
3. Kill, if C_1 is primary.
4. Kill, if bond (C_2—C_3) is in a ring and C_1 is not in a ring.
5. If C_1 is in a ring,
   A. Kill, if C_2 is not in a ring.
   B. Kill, if C_3 is not in a ring.
6. Kill, if CNUM[a] (C_1) > CNUM (C_2).
7. If CNUM (C_1) < CNUM (C_2),
   A. Pass, if C_1 is in a ring.
   B. If CNUM (C_3) = CNUM (ALPHA[b] (C_2),
      (1) Pass if the fragment at C_3 is identical to the fragment at ALPHA (C_2).
      (2) Kill.
   C. Pass, if C_3 is secondary.
   D. Kill, if ALPHA (C_2) is secondary.
   E. Pass, if C_3 is primary.
   F. Kill.
8. If C_2 is secondary,
   A. If CNUM (C_3) = CNUM (ALPHA (C_1)),
      (1) Pass, if fragment at C_3 is identical to fragment at ALPHA (C_1).
      (2) Kill.
   B. Pass, if C_3 is secondary.
   C. Kill, if ALPHA (C_1) is secondary.
   D. Pass, if C_1 is primary.
   E. Kill.
9. Kill, if C_3 is tertiary.
10. If C_3 is primary,
    A. Kill, if any atom alpha to C_1 or C_2 is secondary.
    B. Pass.
11. If any atom alpha to C_1 is secondary,
    A. Pass, if any fragment alpha to C_1 is identical to the fragment at C_3.
    B. If (C_2—C_3) is a ring bond,
       (1) Kill, if any atom alpha to C_1 is a ring atom and is secondary.
       (2) Pass.
    C. Pass, if any fragment alpha to C_2 is identical to the fragment at C_3.
    D. Kill.
12. Pass, if (C_2—C_3) is a ring bond.
13. Pass, if ALPHA (C_2) is not secondary.
14. Pass, if fragment at ALPHA (C_2) is identical to the fragment at C_3.
15. Kill.

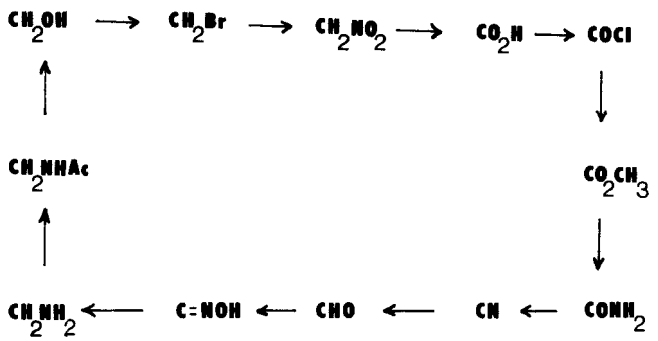


Figure 9. One of many circuits resulting from uncontrolled application of Functional Group Interchange reactions.

transform application. There are two main reasons for obtaining a large synthesis tree. These reasons are:

1. Repetitive application of a group of transforms resulting in circuits of the type illustrated in Figure 9.
2. Inability to apply transforms that are capable of synthesizing the target skeleton in a few reaction steps. This inability stems from the presence of functional groups in the target that interfere with the desired transform.

One possible solution to the problem of a large synthesis tree is to delete paths that are generated by repetitive application of functional group interchange transforms. However, a great deal of time would be wasted in generating such circuits and then finding and deleting them. In the REACT program the following two methods are used for avoiding the generation of these circuits.

The first method is to classify each functional group into one of three classes (Blower and Whitlock, 1976). These classes for each functional group are presented in Table 7. Associated with each functional group x in classes 2 and 3 is a list of pairs (t, y), where y is a functional group in the same class as x, and t is a transform which accomplishes $x \Rightarrow y$. Transforms which appear in one of these lists are not applied automatically when the program is searching for precursors as described in the previous section.

There are several reasons for selecting these particular functional group interchange transforms. Firstly, these transforms are quite general. For example, a primary alcohol can be transformed (retrosynthetic) into an aldehyde via $LiAlH_4$, and this transformation is independent of the particular molecule in which it occurs. In addition, since no change takes place in the carbon skeleton, indiscriminant application of these transforms is undesirable. Secondly, group interchanges between classes are not general, since successful application of the transform relies on symmetry or special structural features in the vicinity of the functional group. For example, the success of the transform CH—COH $\Rightarrow$ C=C depends on the substitution pattern of the carbon atoms or some electronic effect. Thirdly, transforms which create functional groups at unfunctionalized carbons (e.g., $CH_2 \Rightarrow C{=}O$) require special attention and will be treated separately.

The second method used by the program to avoid reaction circuits is to allow the pattern matching routine to make *one* pseudomatch of either of the following three types:

1. A carbon of any class 2 functional groups "matches" a carbon of any other class 2 functional group as long as the carbon substitution is the same. For example, an aldehyde (—CH=O) matches a primary alkyl halide (*—CH_2X) or an ester at the carbonyl carbon (*—CO_2—C), but not a secondary alcohol (*—CHOH) or the sp^3 carbon of an ester derived from a tertiary alcohol (*—C—O—CO);* indicates the atoms which are considered to match the pseudo-match.
2. The carbon atoms of an acetylene group in the transform match the carbon atoms of any class 3 functional group of the target, as long as it contains no tertiary carbons.

TABLE 7. CLASS OF FUNCTIONAL GROUPS RECOGNIZED BY THE PROGRAM

Name	Class
Acetal	2
Acetylene	3
Acid	2
Acid Halide	2
Alcohol	2
Aldehyde	2
Alkyl Halide	2
Alkyl Nitro	2
Amide	2
Amine	2
Anhydride	2
Azide	2
Diazo	2
Enamine	1
Enol Ether	1
Epoxide	3
Ester	2
Ether	2
Hale Olefin	1
Hydrozine	2
Hydrazone	2
Imine	2
Ketal	2
Ketone	2
Nitride	2
Nitro Alkyl Halide	2
Nitro Olefin	1
Olefin	3
Oxime	2

3. An olefin group in the transform matches an epoxide of the target.

The secondary checks involved in transform application operate as before, except that when a secondary check is applied to a substructure which contains an atom involved in a pseudomatch, the substructure's functional groups are considered to be those present in the transform substructure.

When a target substructure-transform substructure match is found which contains a pseudomatch, all feasible transform sequences which convert the given functional group into the desired one, of length one and two, are constructed to the precursor which contains the desired functional group.

Consider the example shown in Figure 10, where the malonic ester synthesis transform is applied to the target designated as A

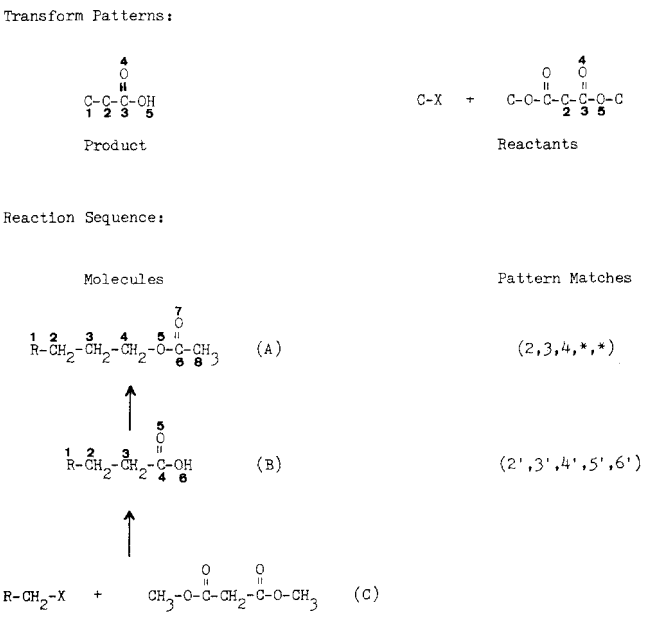


Figure 10. Application of the Malonic Ester Synthesis at level one.

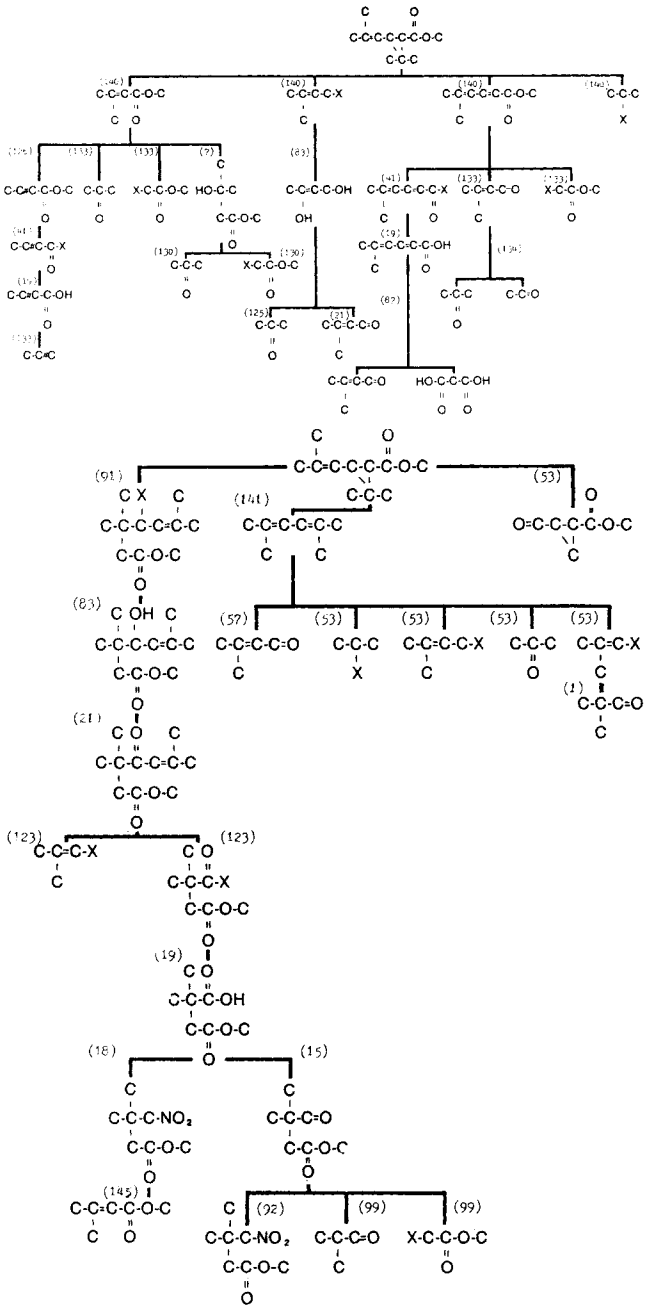


Figure 11. Synthesis Tree for Methyl Crysanthemate generated by the REACT Program.

in the figure. The superscripts appearing in the molecular formulas simply number the atoms. The numbering in the reactant pattern identifies the corresponding atoms in the product pattern. The transform product pattern is "pseudo"-matched with the target (moleculeA) substructure, resulting in a pattern match written as (2, 3, 4, *, *). The entry in the i^{th} position of a match is the number of the atom in the target molecule that matches the i^{th} atom of the transform product pattern.

Once the transform product pattern has been "pseudo"-matched with a substructure in molecule A, the program applied a functional group interchange transform to convert the target substructure to the desired group, resulting in molecule B. The transform product pattern can now be matched completely with a substructure in precursor B resulting in the two precursors designated as C in Figure 10. The exact pattern match is also shown in the figure.

To synthesize the target molecule in a few reaction steps, the program uses blocking groups. As discussed in the transform application section, each transform is applied in a three step process, where in the third step the program determines whether the functional group in the target (other than the substructure under consideration) is stable to the reaction conditions. If functional groups are found that would interfere with the desired transform, blocking is initiated. This is achieved by maintaining a list of transforms that allow functional groups to be interchanged (Lednicer, 1972). For each interfering group, the program applies all potential blocking group transforms.

Once the synthesis tree has been developed, the program estimates the cost of each path in the tree. The cost estimate begins with the cost of the starting materials. The cost estimate for each predecessor of the terminal molecules (starting materials) in the synthesis tree is the cost of the terminal molecule modified by economic factors involved in the given reaction. The problem associated with evaluating these factors is discussed in the following section.

Evaluation of Reaction paths

There are several difficult problems associated with estimating the cost of a reaction path. First, yield prediction inevitably introduces errors. There is considerable fluctuation in the reported yields for reactions which are identical or very similar. The methods for determining yields vary, and frequently yield data is only available for a few simple cases. There is also insufficient data from kinetic studies or competition experiments to allow the prediction of produc ratios for competing reactions.

Secondly, problems of separation, and consequently their costs, are difficult to decide. Whether the separation should be based on chemical or physical properties, what equipment and materials should be used, and how to predict the composition of the reaction mixture are just some of the decisions that have to be made before separation costs can be estimated. Finally, it is difficult to price the safety measures and special handling required in using dangerous materials.

A great deal of information (not normally available at the synthesis stage) is required to compute these costs. One of the major functions of a research and development effort is to generate the knowledge and data necessary to evaluate the above costs, thereby determining the "optimum" reaction path.

In the immediate future, we plan to extend our program in three directions. First, we will develop an evaluation function which would be an improved version of the evaluation function currently used by the program. This function will incorporate a few simple rules for bondset selection based on symmetry considerations and convergent synthesis (Hendrickson, 1976). It will also use empirical estimates of reaction yield, based on literature reports, and use linear free energy relationships to estimate rate constants.

Example. As an example, the output for Methyl Crystanthemate, a precursor of the naturally occurring pyrethrin insecticides, is presented in Figure 11. The output is only a selected portion of the three generated by the REACT program. In several cases, small precursors (e.g., formaldehyde) were not drawn by the program in order to save space. The numbers in parentheses appearing in the Figures refer to the transforms listed in the Appendix.

The synthesis tree shows two routes to the cyclopropane ring via sulfur-ylides. The first reaction is between diphenylsulfonium 3-methyl-2-butenylide, 3-methyl-2-butenoic acid and methyl ester, and the second reaction between diphenylsulfonium isopropylide, 5-methyl-hexa-2, 4-dienoic acid and methyl ester. Syntheses of crysanthemic acid esters using reactions similar to both of these reactions have been reported in the literature (Martel and Huynh, 1967; Corey and Jautelat, 1967). Another route to the cyclopropane ring similar to the internal alkylation of 4-chloro-3,3, 6-trimethyl-5-heptenoic acid, methyl ester, given in the tree, was developed by chemists at ICI (Bader, 1976). The final route to the cyclopropane ring in the tree is the reaction between 2,5-dimethyl-2,4-hexadione and methyl diazoacetate. There are two reports in the literature (Bader, 1976; Julia and Guy-Rouault, 1967) of similar reactions.

ACKNOWLEDGEMENTS

I would like to thank Bill Trosky for assisting in the programming and Paul Blower for his contributions to the project.

APPENDIX 1: List of Transforms Used by the Program in Generating the Synthesis Tree Shown in Figure 11.

1. HALO-OLEFINS VIA PHOSPHORUS PENTAHALIDE

   X—C=C ⇒ O=C—CH

   Reaction

   $$PX5 + \underset{\|}{O}\!\!-\!C\!-\!CH \xrightarrow[REFLUX]{CH2CL2} X\!-\!C\!=\!C + POX3 + HX$$

   References

   1. MARCH P. 677; B&P P.335; FIESER V.1 P.866.
   2. M.S.NEWMAN, F.FRAENKEL AND W.KIRN, JOC.28, 1851(1963).

7. ELIMINATION OF HYDROXYL BETA TO W GROUP

   W—C=C ⇒ W—CH—C—OH

   Reaction

   $$W\!-\!CH\!-\!C\!-\!OH + SOCL2 \xrightarrow[25\ C]{PYRIDINE} W\!-\!C\!=\!C + SO2 + HCL$$

   References

   1. HOUSE P. 673; B&P P.71.

15. OXIDATION WITH ACIDIC PERMANGANATE

    CO2H ⇒ CHO

    Reaction

    $$CHO + KMNO4 + H2SO4 \xrightarrow[H2O]{15\ C} \begin{matrix} CO2H + MNO2 + \\ K2SO4 + H2O \end{matrix}$$

    References

    1. HOUSE P.273;B&P P.762;FIESER V.1, P.943

18. NEF—SCHECHTER REACTION

    $\underset{\|}{O}\!\!-\!C\!-\!OH \Rightarrow CH2NO2$

    Reaction

    $$KMNO4 + CH2NO2 + KOH \xrightarrow[H2O]{25\ C} \begin{matrix} CO2H + KNO3 + KOH + \\ MNO2 + H2O \end{matrix}$$

    References

    1. HOUSE P.285;B&P P.565.
    2. H.SCHECHTER AND F.T.WILLIAMS JOC.27, 3699(1962).

19. FORMATION OF ACID HALIDES

    $\underset{\|}{O}\!\!-\!C\!-\!X \Rightarrow \underset{\|}{O}\!\!-\!C\!-\!OH$

    Reaction

    $$\underset{\|}{O}\!\!-\!C\!-\!OH + SOCL2 \xrightarrow[80\ C]{DMF(CAT.)} \underset{\|}{O}\!\!-\!C\!-\!CL + HCL + SO2$$

    Reference

    1. B&P PP.860,336-7

21. SODIUM BOROHYDRIDE REDUCTION OF CARBONYL COMPOUNDS

    CHOH ⇒ C=O

    Reaction

    $$C\!=\!O + NABH4 \xrightarrow[2.\ HX,H2O]{1.\ ISOPROPANOL,\ 25\ C} CHOH + NACL + B(OH)3$$

    References

    1. HOUSE P.49;B&P P.194;FIESER V.1,P.1049.

27. ACID CATALYZED HYDRATION OF OLEFINS

    HO—C—CH ⇒ C=C

    Reaction

    $$C\!=\!C + H2O \xrightarrow[100\ C]{AQUEOUS\ H2SO4} HC\!-\!C\!-\!OH$$

    Reference

    1. B&P P.185.

41. SCHOTTEN-BAUMANN ESTERIFICATION

    $\underset{\|}{O}\!\!-\!C\!-\!O\!-\!C \Rightarrow \underset{\|}{O}\!\!-\!C\!-\!X + HO\!-\!C$

Reaction

$$\mathrm{C(=O)-X + C-OH + KOH \xrightarrow{H2O} C(=O)-O-C + KX + H2O}$$

Reference

1. B&P P.807.

53. WITTIG REACTION

C=C ⇒ C=O + C−X

Reactions

1. $\mathrm{C-X + PH3P \xrightarrow[25\ C]{BENZENE}}$ PHOSPHONIUM SALT

2. $\mathrm{PHOSPHONIUM\ SALT \xrightarrow[2.\ ADD\ C=O]{1.THF,\ 0\ C} C=C + CH4 + LIX + PH3P=O}$ (25 C)

References

1. HOUSE P.682;B&P P.141;FIESER V.1,P.1238.

57. THERMAL COUPLING OF ORGANOCUPRATES

C—C ⇒ C—X

Reactions

1. $\mathrm{C-X + LI \xrightarrow[0\ C]{ET20,} C-LI + LIX}$

2. $\mathrm{C-LI + CUX \xrightarrow[25\ C]{PH3P,ET20} C-C + CU + LIX}$

References

1. G.H.POSNER ORG.RXNS.22,253(1975)
2. FIESER V.1,P.95.

83. ALKYL HALIDES FROM ALCOHOLS AND PX3

C—X ⇒ C—OH

Reaction

$$\mathrm{C-OH + PX3 \xrightarrow[NO\ SOLVENT]{25\ C} C-X + P(OH)3}$$

References

1. HOUSE P.452;B&P P.332;FIESER V.1,P.873,5.

84. HYDROHALOGENATION OF OLEFINS

X—C—CH ⇒ C=C

Reaction

$$\mathrm{HX + C=C \xrightarrow[25\ C]{HOAC} HC-CX}$$

References

1. B&P P.346; HOUSE,P.446.

87. KNOWEVENAGEL REACTION WITH DECARBOXYLATION

W—C=C ⇒ W—CH—CO2H + C=O

Reaction

$$\mathrm{C=O + W-CH-CO2H \xrightarrow[PYRIDINE;\ REFLUX]{PIPERIDINE,} W-C=C + H2O + CO2}$$

References

1. HOUSE P.646.
2. G.JONES, ORG.RXNS,15,204(1967).

91. ALKYLATION OF CARBON ALPHA TO WITHDRAWING GROUP

W—C—CH ⇒ W—CH + CH—X

Reaction

$$\mathrm{W-CH + LIN(I-PR)2 \xrightarrow[2.ADD\ ENOLATE\ TO\ HALIDE,\ 25\ C]{1.THF,-78\ C} W-C-CH + LIX + (I-PR)2NH}$$

References

1. HOUSE P.546.
2. M.RATHKE & A.LINDERT,JACS,93,2318(71).
3. R.CREGER,JACS,89,2500(1967).

92. NEF-SCHECHTER REACTION CARBONYL COMPOUND

C=O ⇒ CHNO2

Reaction

$$\mathrm{CHNO2 + KMNO4 \xrightarrow[2.STEAM\ DISTILL]{1.0\ C,(.1N)KOH} C=O + KOH + MNO2 + KNO3}$$

References

1. H.SCHECHTER AND F.T.WILLIAMS,JR.,JOC.27,3699(1962).

99. ALKYLATION OF CARBONYL COMPOUNDS VIA ENAMINES

O=C—C—CH ⇒ O=C—CH + CH—X

Reactions

1. $$O{=}C{-}CH + \underset{\text{AMINE}}{\text{CYCLOHEXYL}} \xrightarrow[\substack{\text{AZEOTROPIC}\\\text{REMOVAL OF}\\\text{WATER}}]{\text{BENZENE}} \text{SCHIFF BASE}$$

2. $$\text{SCHIFF BASE} + CHX \xrightarrow[2.HX,H2O,\text{REFLUX}]{1.ETMGX,\text{REFLUX}} O{=}C{-}C{-}CH + \text{ETHANE} + MGX2 + \text{CYCLOHEXYL AMINE}$$

Reference

1. G.STORK AND S.DOWD JACS,85,2178(1963).

113. CARBONATION OF ACETYLIDES

$C^*C{-}CO2H \Rightarrow C^*CH + CO2$ *—TRIPLE BOND

Reaction

$$C^*CH \xrightarrow[3.HCL,H2O]{1.LINH2,ET2O\ 2.DRY\ ICE} C^*C{-}CO2H + NH3 + LICL$$

Reference

1. W.ZIEGENBEIN,"CHEMISTRY OF ACETYLENES",H.VIEHE,ED., M.DEKKER,N.Y.,1969.

123. ACYLATION OF ORGANOCUPRATES

$O{=}C{-}C \Rightarrow O{=}C{-}X + C{-}X$

Reactions

1. $$C{-}X + LI \xrightarrow[2.CUX]{1.0\ C,THF} \text{DIALKYL CUPRATE} + LIX$$

2. $$\underset{\text{CUPRATE}}{\text{DIALKYL}} + O{=}C{-}X \xrightarrow[2.HX,H2O]{1.-78\ C,THF} O{=}C{-}C + LIX + CUX$$

Reference

1. G.H.POSNER,ORG.RXNS.,22,253(1975).

125. DOUBLE WITTIG

$HO{-}C{-}C{=}C \Rightarrow C^*O + C{-}X + C{=}O$ *—TRIPLE BOND

Reactions

1. $$CHX + PH3D \xrightarrow[25\ C]{\text{BENZENE}} \text{PHOSPHONIUM SALT}$$

2. $$\underset{\text{SALT}}{\text{PHOSPHONIUM}} \xrightarrow[\substack{3.BULI,-25\ C\\4.ADD\ CARBONYL\\5.HX,H2O}]{\substack{1.BULI,THF,-78\ C\\2.ADD\ (ENE)CARBONYL}} HO{-}C{-}C{=}C + \text{BUTANE} + LIX + PH3PO$$

Reference

1. E.J.COREY AND H.YAMAMOTO,JACS,92, 226,3523,6636,6637(1970).

126. CONJUGATE ADDITION TO ACETYLENES

$W{-}C{=}C{-}C \Rightarrow W{-}C^*C{+}\ C{-}X$ *—TRIPLE BOND

Reactions

1. $$C{-}X + LI \xrightarrow[2.CUX]{1.0\ C,THF} \text{DIALKYL CUPRATE} + LIX$$

2. $$\underset{\text{CUPRATE}}{\text{DIALKYL}} + C^*C{-}W \xrightarrow[2.HX,H2O]{1.THF,-78\ C} C{-}C^*C{-}W + LIX + CUX$$

References

1. G.H.POSNER ORG.RXNS.,19,1(1972).
2. R.ANDERSON,V.CORBIN,G.GOTTERRELL,C.HENDRIK F.SCHAUB AND J.SIDDELL,JACS 97,1197(1975).

130. REFORMATSKY REACTION

$W{-}C{-}C{-}OH \Rightarrow W{-}C{-}X + C{=}O$

Reaction

$$W{-}C{-}X + C{=}O + ZN \xrightarrow[2.AQUEOUS,HOAC]{1.BENZENE,\text{REFLUX}} W{-}C{-}C{-}OH + ZN(OAC)2$$

References

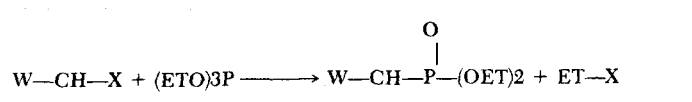
1.HOUSE P.671;B&P P.229.

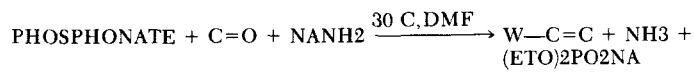
133. WADSWORTH-EMMONS REACTION

$W{-}C{=}C \Rightarrow W{-}CH{-}X + C{=}O$

Reaction

1.

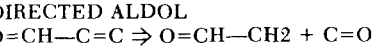


2. 

References

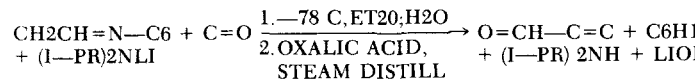
1. HOUSE P.690,701.
2. W.WADSWORTH AND W.EMMONS, JACS 83,1733(1961); ORG.SYN.,45,44(1965).
3. A.BOSE AND R.DAHILL,JOC,30,505(1965).

134. DIRECTED ALDOL



Reactions

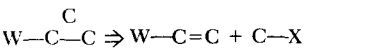
1.

2. 

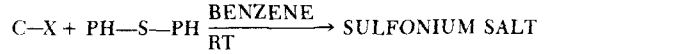
References

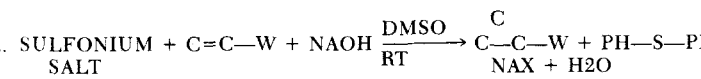
1. G. WITTIG,H.D.FROMMELD AND P.SUCHANEK,ANGEW. CHEM.INT.ED.2,683(1963).
2. G.WITTIG AND H.REIFF,ANGEW.CHEM.INT.ED.1,7(1968).
3. W.NAGATA AND Y.HAYASE,J.CHEM.SOC,(C)1969,460.

140. CYCLOPROPANES VIA SULFUR YLIDS



Reactions

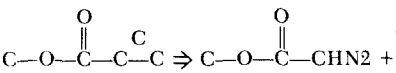
1. 

2. 

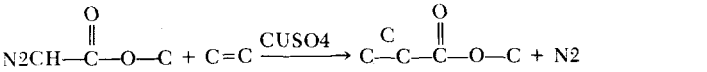
References

1. HOUSE P.709.
2. B.M.TROST AND L.S.MELVIN JR., "SULFUR YLIDES" ACADEMIC PRESS, NEW YORK, 1975.
3. B.M.TROST AND M.J.BOGDANOWICZ,J.AM.CHEM.SOC.95,5298(1973).

141. CYCLOPROPANATION WITH DIAZOACETATE



Reaction

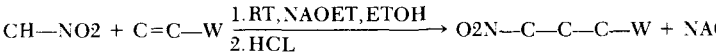


References

1. FIESER P.367.
2. W.KIRMSE, "CARBENE CHEMISTRY"'ACADEMIC PRESS,NEW YORK,1964.

145. MICHAEL REACTION WITH NITROALKANES

Reaction



References

1. HOUSE, P.595.
2. G.N.WALKER, J.ORG.CHEM.30,1416(1965).

LITERATURE CITED

Bader, A., "Recent Advances in Synthetic Pyrethroids," *Aldrichimica Acta*, **9**, 49 (1976).

Barone, R., "Simulation of Organic Synthesis," *Computer-Assisted Design of Organic Synthesis*, Table Ronde Roussel, UCLAF, Paris (1977).

Bersohn, M., and A. Esack, "A Canonical Connection Table Representation of Molecular Structure, *Chemica Scripta*, **6**, 122 (1974).

Blower, P. E., and H. W. Whitlock, Jr., "An Application of Artificial Intelligence to Organic Synthesis," *J. Am. Chem. Soc.*, **98**, 1499 (1976).

Buehler, C. A., and D. E. Pearson, *Summary of Organic Synthesis*, Wiley Interscience, New York, NY (1970).

Camp, D. M., "A Reaction Path Synthesis Strategy for Organic Compounds," M.S. ChE. Thesis, Massachusetts Institute of Technology, Cambridge, MA (1974).

Corey, E. J., and M. Jautelat, "Construction of Ring Systems Containing the gem-Dimethylcyclopropane Unit Using Diphenyl-Sulfonium Isopropylide," *J. Am. Chem. Soc.*, **89**, 3917 (1967).

Corey, E. J., and W. T. Wipke, "Computer-Assisted Design of Complex Organic Syntheses," *Science*, 166 (1969).

Corey, E. J., W. T. Wipke, R. D. Cramer, III, and W. J. Howe, "Techniques for Perception by a Computer of Synthetically Significant Structural Features in Complex Molecules," *J. Am. Chem. Soc.*, **94**, 431 (1972).

Corey, E. J., H. W. Orf, and D. A. Pensak, "Computer-Assisted Synthetic Analysis. The Identification and Protection of Interfering Functionality in Machine-Generated Synthetic Intermediates," *J. Am. Chem. Soc.*, **98**, 210 (1976).

Govind, R. and G. J. Powers, "A Chemical Engineering View of Reaction Path Synthesis in Computer-Assisted Organic Synthesis," ACS Symposium Series, 61 (1976).

Hendrickson, J. B., "A General Protocol for Systematic Synthesis Design," *Topics in Current Chemistry*, 62 (1976).
House, H., *Modern Synthetic Reactions*, Benjamin, Inc., Menlo Park (1972).
Julia, M., and A. Guy-Rouault, "Synthese de Cyclopropanes a' partir de sulfones," Bulletin De la Societe Chimique De France, 1411 (1967).
Kaufman, G., "Computer Design of Synthesis in Organo-phosphorous Chemistry," *Computer-Assisted Design of Organic Synthesis*, Table Ronde Roussel, UCLAF, Paris (1977).
Lednicer, D., "Latent Functionality in Organic Synthesis," *Advances in Organic Chemistry*, 8, 179 (1972).
March, J., *Advanced Organic Chemistry: Reactions, Mechanisms and Structure*, McGraw Hill, New York, NY (1965).
Martel, J., and C. Huynh, "Synthese de Lacide crysanthemique," Bulletin De La Societe Chimique DeFrance, 985 (1967).
Moreau, G., "The Masso Program," *Computer-Assisted Design of Organic Synthesis*, Table Ronde Roussel, UCLAF, Paris (1977).
Smith, E. J. "W. J. Wiswesser's Line Formula Notation," McGraw Hill (1968).
Ugi, I., J. Blair, J. Gasteiger, C. Gillespie and P. Gillespie, "Representation of the Constitutional and Stereochemical Features of Chemical Systems in the Computer-Assisted Design of Syntheses," Tetrahedron, **30**, 1845 (1974).
Wipke, W. T., H. Braun, G. Smith, F. Choplin and W. Sieber, "SECS—Simulation and Evaluation of Chemical Synthesis: Strategy and Planning," *Computer-Assisted Organic Synthesis*, ed., W. T. Wipke and W. J. Howe, ACS Symposium Series, 61 (1976).

Manuscript received January 29, 1980; revision received August 5, and accepted August 14, 1980.

Effect of Flow Direction on Conversion in Isothermal Radial Flow Fixed-Bed Reactors

Singular perturbation solutions are used to derive criteria predicting the influence of change in flow direction on the conversion in an isothermal radial-flow, fixed-bed reactor. It is shown that when the reaction does not involve a change in the number of moles, the outward flow direction is the preferred one for any reaction with a convex rate expression. The inward flow direction is a better choice for any reaction with a concave rate expression. A more intricate behavior occurs for reactions involving a change in volume. Analytical approximations and numerical solutions indicate that the influence of flow direction on the conversion is rather small.

VEMURI BALAKOTAIAH
and
DAN LUSS
Department of Chemical Engineering
University of Houston
Houston, TX 77004

SCOPE

Cylindrical or spherical radial-flow, packed-bed reactors (RFBR) are used in several important industrial processes. Previous studies (Raskin et al., 1968; Hlavacek and Kubicek, 1972; Calo, 1978) pointed out that when an exothermic reaction is carried out in these reactors, the conversion as well as the yield of a desired product depend on the direction of flow. Dudukovic and Lamba (1975) used numerical solutions to show that when an isothermal n-th order reaction ($n \neq 1$) is carried out in a cylindrical RFBR, the conversion depends on the flow direction.

Analysis of the governing steady-state equations indicates that the flow direction affects the conversion only when the axial dispersion is important. When the reaction does not involve a change in volume, the flow direction affects the conversion for all kinetic rate expressions except for a first and a zeroth-order reaction. When the reaction involves a change in volume, the conversion is dependent on the flow direction for all rate expressions except for a zeroth-order reaction.

The purpose of this work is to enhance our understanding of the causes and magnitude of the influence of flow direction on the conversion in a RFBR. The method of matched asymptotic expansions is used to obtain approximate solutions of the governing nonlinear differential equations and to derive criteria predicting the pereferred flow direction for reactions with no change in the number of moles. Numerical simulations are used to find the preferred flow direction for reactions involving a change in volume.

CONCLUSIONS AND SIGNIFICANCE

It is shown that for all isothermal reactions, which do not involve a change in volume, the outward RFBR yields a higher conversion than the inward RFBR for reactions with convex rate expressions, such as n-th order reactions with $n > 1$. The inward RFBR is preferred when the kinetic expression is concave. For large Peclet numbers (small deviation from plug flow), simple criteria are derived predicting the difference in conversion between the two flow directions for any kinetic rate expression when either the Bodenstein number is a constant (Eq. 44) or the radial dispersion coefficient is a constant (Eq. 54). For an n-th order reaction, these expressions become Eqs. 55 and 56, respectively.

It is shown that for a first-order reaction involving an increase in the number of moles, the outward flow gives a higher conversion than the inward flow. The converse is true for a first-order reaction with a decrease in the number of moles.

The above findings lead to the prediction that the outward flow direction is preferred for any reaction with a convex rate expression which involves an increase in volume. Similarly, the inward flow is the preferred direction for any reaction with a

0001-1541-81-4826-0442-$2.00.