Table VI. Final Format of the Data Concerning the Stork Enamine Variant of the Ketone Alkylation Reaction

Reaction number: concatenation of substructure number and an arbitrary unique number 7.257
Address of next variant: address of 8.257 ketone alkylation using strong base
Address of next reaction of this priority that produces this substructure: address of 11.257 reaction of boron alkyl with alkyl vinyl ketone, followed by hydrogen peroxide oxidation
Miscellaneous features of the reaction: the reaction cannot open rings, there are extra atoms present in the reactant(s) but absent in the identified product.
Yield: 60%
Tests on the feasibility of the reaction to produce the particular product at hand
reaction conditions: weak base (pyrrolidine), aqueous weak acid (acetic acid)
substructure limitations: atom 2 must have at least as many hydrogen atoms as atom 4.
Special tasks to be performed on certain atoms of the substructure, e.g., chirality inversion, change of double bond cis–trans relationship: no such tasks in this reaction
Change in central atom functional group labels: atom number 1 of the reacting substructure must receive the label 41, meaning alkyl iodide; these labels facilitate subsequent substructure discovery in the reactant(s)
Bond replacements: the bond from atom 1 to atom 3 is replaced by a bond from atom 1 to the iodine atom, extra atom number 1; the bond from atom 3 to atom 1 is replaced by a bond from atom 3 to a hydrogen atom.
Reference: *J. Am. Chem. Soc.*, **85**, 207 (1963)
Particular substructure of this example, showing the atom numbering 4 3 2 1 C–C–C–C O

the synthesis-generating program knows which functional groups will not survive the various standard reaction conditions, but the behavior of the functional groups in the presence of new reagents often is not known from previous experience. This deficiency will have to be overcome manually, at the time of editorial review, until such time as a program is written to read the discussion section. Extension to the German language will be straightforward for the experimental section but not so for the discussion section.

Table VI shows the format of the expected final result, a description of this reaction variant as it appears in our synthesis-generating program.

Note: After this paper was written, important work by Lynch and Willett appeared which describes major advances in the automatic compilation of synthetic reactions.[14,15]

LITERATURE CITED

(1) P. Gund, J. D. Andose, and J. B. Rhodes, "Computer-Assisted Synthetic Analysis in Drug Research" in "Computer-Assisted Organic Synthesis", W. T. Wipke and W. J. Howe, Eds., ACS Symposium Series, No. 61, American Chemical Society, Washington, D.C., 1977, p 179.
(2) J. Martin, "Computer Data-Base Organization", 2nd ed, Prentice-Hall, Englewood Cliffs, N.J., 1977.
(3) R. Fugmann in "Chemical Information Systems", J. E. Ash and E. Hyde, Eds, Ellis Horwood, Chichester, 1975, pp 195–226.
(4) G. E. Vleduts, *Inf. Storage Retr.*, **3**, 303 (1963), expanded in the book, G. E. Vleduts, and E. A. Geivandov, "Automated Information Systems for Chemistry", Nauka, Moscow, 1974 (in Russian).
(5) M. Bersohn, "Automatic Problem Solving Applied to Synthetic Chemistry", *Bull. Chem. Soc. Jpn.*, **45**, 1897–1903 (1972); M. Bersohn and A. Esack, "Some Early Developments in Programming Synthetic Strategies", *Comput. Chem.*, **2**, 105–111 (1978).
(6) G. G. Vander Stouw, P. M. Elliott, and A. C. Isenberg, "Automatic Conversion of Chemical Substance Names to Atom-Bond Connection Tables", *J. Chem. Doc.*, **14**, 185–193 (1974).
(7) G. G. Vander Stouw, I. Naznitsky, and J. E. Rush, "Procedures for Converting Systematic Names of Organic Compounds into Atom-Bond Connection Tables", *J. Chem. Doc.*, **7**, 165–169 (1967).
(8) J. E. Armitage, J. E. Crowe, P. N. Evans, M. F. Lynch, and J. A. McGuirk, "Documentation of Chemical Reaction by Computer Analysis of Structural Change", *J. Chem. Doc.*, **7**, 209–215 (1967).
(9) J. M. Harrison and M. F. Lynch, "Computer Analysis of Chemical Reactions for Storage and Retrieval", *J. Chem. Soc. C*, 2082 (1970).
(10) M. F. Lynch, P. R. Munn, and J. Radcliffe, "Production of Printed Indexes of Chemical Reactions Using Wiswesser Line Notations", *J. Chem. Inf. Comput. Sci.*, **18**, 94 (1978).
(11) M. Bersohn and A. Esack, "Functional Group Discovery Using the Concept of Central Atoms", *Chem. Scripta*, **9**, 213–217 (1976).
(12) M. Bersohn and A. Esack, "A Computer Representation of Synthetic Organic Reactions", M. Bersohn and A. Esack, *Comput. Chem.*, **1**, 103–107 (1976).
(13) W. Theilheimer, "Synthetic Methods in Organic Chemistry", Interscience, New York, pp 11 246–11 278.
(14) M. F. Lynch and P. Willett, "The Production of Machine-Readable Descriptions of Chemical Reactions Using Wiswesser Line Notations", *J. Chem. Inf. Comput. Sci.*, **18**, 149–154 (1978).
(15) M. F. Lynch and P. Willett, "The Automatic Detection of Chemical Reaction Sites", *J. Chem. Inf. Comput. Sci.*, **18**, 149–154 (1978).

Roche Integrated Reaction System (RIRS). A New Documentation System for Organic Reactions

HERBERT J. ZIEGLER

Scientific Documentation Department, F. Hoffmann-La Roche & Co. A.-G., 4002 Basel, Switzerland

Received February 28, 1979

A new computer-assisted Reaction Documentation System is described which is in use in the research department of Hoffmann-La Roche in Basel. Its purpose is the retrieval of organic reactions according to criteria which are of interest to the synthetic chemist: synthetic pathways, reactivities, types of reactions, reagents, reaction conditions, and side reactions. The following data are stored in searchable form in the computer: starting materials, products, substructures which are involved in the respective reaction, substructures which are characteristic for the respective synthesis, reaction conditions, other important aspects, as well as names of the reactions.

I. INTRODUCTION

In 1965, publication of a reaction documentation system on punched cards as "Reactiones Organicae" was started by Georg Thieme Verlag, Stuttgart.[1-3] This system had all the restrictions and deficiencies inherent in a punched card method. However, making use of the basic ideas of the system and experience with it, and taking advantage of hardware and software of modern data-processing technology, a new system has been developed which avoids the disadvantages of handling punched cards. This computer-assisted retrieval system for synthetic methods is fast, inexpensive, and satisfies user needs to a high degree.

The system contains two independent files. The first, with reactions and procedures reported within Roche, is called the

R-file and is subject to security precautions appropriate to proprietary information. The second, which is called the S-file, includes interesting synthetic reactions from the literature. Classical reactions are chosen from "Reactiones Organicae", and a selection from Theilheimer's "Synthetic Methods", Vol. 1–30, is made by a team of synthetic chemists. Furthermore, the S-file contains all reactions collected in the journal *Synthesis* (*Synthesis Abstracts*), which are selected by Roche Nutley, N.J., from the most recent chemical literature.

In the following paragraphs the new Roche Integrated Reaction System (RIRS) is described and illustrated with examples.

II. OBJECTIVES

The purpose of the Reaction Documentation System is to answer as many pragmatic, practical questions of the synthetic chemist as possible by finding examples of synthetic reactions which are possible solutions to his problem. Therefore, the system should be in a position to answer the following or similar questions:

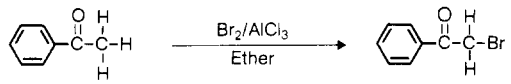
- How do specific compounds react?
- How can specific compounds be synthesized?
- How can specific compounds be converted into other specific compounds?
- How do specific functions or groupings of atoms react?
- How do they react under specific conditions?
- How can certain functional groups be introduced into a molecule and how can the synthesis of specific partial structures be realized?
- How can functional groups be introduced into a molecule in the presence of other functions which should not be altered?
- How can certain functions or groups of atoms be transformed into other specific functions or groups of atoms?
- Which reactions exist that make use of a specific reagent or reaction conditions?

Answers from the Reaction Documentation System should be comprehensive; therefore, all reactions have to be included, the classical ones as well as the most recent ones, published in the latest issues of scientific journals. As an integrated documentation system, it should not only give bibliographic information and abstracts but also provide access to the original text together with the original procedure.

It is, however, not possible to include reactions and methods of very special fields like peptides, carbohydrates, alkaloids, or nucleotides. The system would either be too large and too costly or it would not be comprehensive with respect to these particular fields. The reaction documentation system is not intended to be a method for creative synthesis planning, leaving thinking of the chemist to the computer, though this could be a fascinating project. The purpose of this reaction documentation system is more modest, but it appears to be of immediate interest for the synthetic chemist in his daily work. It is supposed to support the chemist's memory and knowledge and to be able to display solutions to a preparative problem; solutions which are stored in the computer.

III. CONCEPT

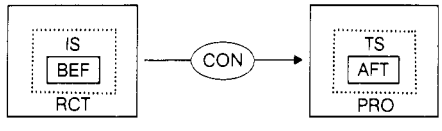
1. Documentation Concept. Each chemical reaction represents a partial change in the structure of one or more compounds under the influence of certain auxiliary materials and specific conditions. For example, a hydrogen atom of the methyl group of acetophenone is replaced by a bromine atom under the influence of bromine in the presence of aluminum chloride in ethereal solution.



Four points of interest result from each reaction, which are the center of all preparative questions:

- the full structures of the compounds, which react or are formed (in our example acetophenone and α-bromoacetophenone)
- those parts of the structure which are actually changed during the reaction, that is which are different before and after the reaction (in our example, –CH before and C–Br after the reaction)
- the reaction conditions which are the cause of these changes (in our example, bromine, aluminum chloride, and ether)
- those parts of the structures which themselves have *not* been changed but which nevertheless might have an important or even decisive influence on the outcome of the reaction or which are associated with the reaction centers for other reasons. These groups form substructures with the reaction centers which are typical for the reaction described (in our example the carbonyl group is added to the reaction center: the reaction is an α-bromination of a ketone)

These four points are also of special emphasis in the reaction documentation system and may generally be formulated by the following scheme:



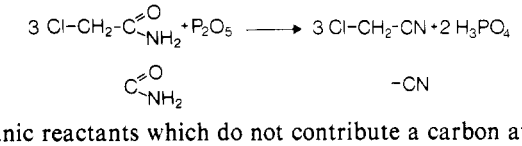
The full structures of the compounds before and after the reaction are called in short form RCT (*reac*tant) and PRO (*pro*duct). The centers which are different *bef*ore and *aft*er the reactions are described as BEF and AFT and the reaction *con*dition is noted as CON. Characteristic groups of atoms before and after the reaction are indicated as *i*nitial *s*ubstructures (IS) and as *t*arget *s*ubstructures (TS).

In addition to this most simple basic scheme various other possibilities exist: several reactants, several products, more than one reaction step, several procedures, and reactions with more than one center in the same molecule. In all these cases, however, the same four points of interest are always present, the only difference being that some of them appear more than once.

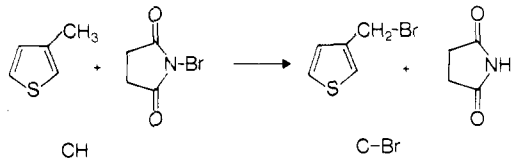
a. Reaction Centers and Reaction Substructures. For the preparative chemist and therefore also for the reaction documentation system, the centers of interest are the reaction centers BEF and AFT as well as initial and target substructures. Very often, the synthetic chemist has the following problems:

- How will he be able to use a well-known reaction (BEF→AFT) in a specific synthetic scheme?
- How can he obtain his target substructure by introduction of a specific function or group of atoms (AFT or target substructure)?
- What reactions are possible with a given grouping of atoms (BEF or initial substructure)?

The *centers of a reaction* (BEF and AFT) can easily be formulated by omitting from the reaction equation all inorganic reactions and products

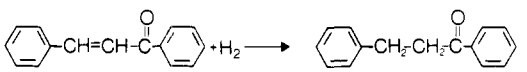


organic reactants which do not contribute a carbon atom to the final structure and preparatively irrelevant organic products



CH C-Br

all unchanged bonds and atoms from the structures of organic reactants and products



C=C CH-CH

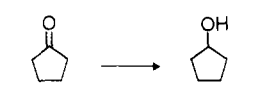
If reactions are recorded and displayed in this manner, the essential points of the course of each reaction can be seen clearly. In many cases this kind of representation will be even more lucid and more differentiated, if in addition to these substructures, certain neighboring atoms are included, which have not participated directly in the reaction, but which form with the actual reaction center a consistent chemical unit and should therefore not be separated. Consequently, esterification of a carboxy group is represented by the reaction centers COOH → COO–C and not O–H → O–C or → O–C, though only an O–C– bond is generated. In this way, esterification of a carboxylic acid can be differentiated from a whole series of other reactions, which also consist of a replacement of an O–H– bond by an O–C– bond:

Esterification of alcohols	C-OH → C-O-C
Esterification of sulfonic acids	SO_2-OH → SO_2-OC
Reductive etherification of carbonyl groups	-CO → CH-O-C
Acetalization of carbonyl groups	-CO → C(-OC)$_2$
Enol-etherification of carbonyl groups	CH-CO → C=C-OC
Enol-esterification of carbonyl groups	CH-CO → C=C-O-CO
Etherification of carboxylic acid amides	CO-N → N=C-OC

The following additional statements improve and differentiate the search possibilities still further:

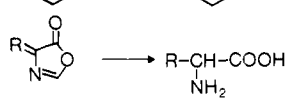
Amines:	PRIM SEC TER QUAT
Alcohols:	PM SC TR EN HT
CC double bonds:	ISOL CONJ
Carbonyl groups:	ALDEHYDE KETONE

And finally, with the help of easily remembered symbols and with numbers from the Ring Index, it can be stated

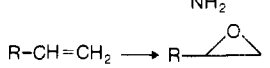


At the Ring BEF: (A)-155

In the Ring BEF: (I)-293

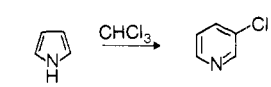


Ring Opening BEF: (O)-119

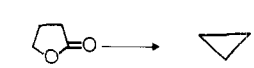


Ring Closure BEF: (R)-12

Ring Substitution BEF: (S)-145 AFT: (S)-149



Ring Expansion BEF: (+)-142 AFT: (+)-277



Ring Contraction BEF: (-)-145 AFT: (-)-14

precisely and in a simple way which rings, ring systems, and changes of rings are involved in the reaction:

Halides are represented in the formulation of the reaction centers by the expression HAL, followed by a statement which indicates the type of halogen atom present. In this way, it is possible to search either for halogen compounds in general or for specific halides, according to choice.

In many cases the reaction centers formulated as just described are sufficiently precise for the description and retrieval of a reaction. However, there exist reactions whose expressions for the reactions centers are very simple and therefore give only little information, e.g., a halogenation, CH → C–HAL, or an alkylation, CH + C–HAL → C–C.

In these and many other cases the special and characteristic features of a reaction are not defined by the reaction centers alone. Therefore, in all these cases additional atoms and bonds are specified which are characteristic for both reactants or products. Thus, these new *reaction substructures* contain in addition to the original reaction centers also the neighboring atoms which are essential for a precise description of the reaction.

In the case of the α-halogenation of a ketone the reaction center, CH → C–HAL, is supplemented by the addition of the keto group to give the substructures CH–CO → HAL–C–CO (target and initial substructure).

Instead of using only the reaction centers CH–HAL → CO for the formation of an α-keto carboxylic ester from an α-bromo carboxylic ester, the product is better represented by the target substructure CO–COO–C.

These examples illustrate that reaction centers and initial or target substructures not only differ in the size of the formula, but also represent an entirely different level of information. Reaction centers show in a formal way the precise changes of the structures connected with the reaction. Initial and target substructures inform us on the use of a reaction for the synthesis of larger, typical structural fragments or of target molecules.

The highly flexible possibility of combining reaction centers and substructures, the readily comprehensible codeless representation, ready adaptability for computer operation ("on line"), and the independence from problems regarding nomenclature are the main features of this reaction documentation. All these facts lead to a highly efficient system which at the same time has a transparent simplicity.

A thesaurus which is periodically updated and printed using an IBM Model 370/158 computer and IBM 3800 laserprinter contains all the formulas occurring in the system of the *reaction centers before*, the *reaction centers after*, the *initial substructures*, and the *target substructures*. They are arranged alphabetically according to the symbols of the elements present. Formulas containing the same elements are distinguished by functional groups. Appropriate cross references are inserted for formulas which contain more than one functional group and could therefore be listed in more than one place. Within each section, formulas are listed in the standard data-processing collating sequence (punctuation symbols, alphabetic characters, numerals). The files and fields [*reaction center before* (*B*), *reaction center after* (*A*), *initial substructure* (*I*), *target substructure* (*T*)] in which the formula has occurred are indicated in tabular form. To know which of the reaction centers and substructures are already in the files and in what connection they have been used is important in itself, and forms a sound basis for the subsequent computer search.

The great value of the thesaurus, however, lies in the fact that it will help the chemist in the formulation of his query. He will compare the target compound that he wants to synthesize with the formula in the relevant sections of the thesaurus. Using the groups of atoms which are present both

Table I

O * 2.1.

H	R	S	
			ONE CARBONYL GROUP
		T	<HETEROCYCLIC|C>-CO
A		T	<3-CYCLOHEXENE>-CO
B A		B A	-CO
B			ANTHRONE
		A	AR-C-CH-CO
		AT	AR-C-CO
		A	AR-C-CO-CH
		A	AR-C=CH-CO
		AT	AR-CH=O
		I	AR-CH2-CO
		B AT	AR-CO
		B	AR-CO-CH
		B	AR-CO-CH2
A			C-<3.5-CYCLOHEXADIENE>=O<1.2>
		B	C-<4-CYCLOHEXENE>=O<1.3>
		A	C-AR-CO-C<1.2>
		T	C-C-C-CO
		T	C-C-C=C-C-CH-CO
		A	C-C-C=C-CO
		A	C-C-CH-CO
A		AT	C-C-CO
		A	C-C-CO-C-C
		A	C-C-CO-CH
		AT	C-C=C-CO
		A	C-C=CH-CO
A		A	C-C=CH-CO-C
A		A	C-CH=O
		T	C-CH2-CO-CH2-C
		T	C-CH2-CO-CH3
A		B A	C-CO
		A	C-CO-C
		A	C-CO-CH
		I	C-CYCLOHEXANE=O<1.4>
		B	C#C-CO-C#C
		A	C=<2.5-CYCLOHEXADIENE>=O<1.4>
		T	C=C<-C-C>-CO
		T	C=C<-C>-CO-CH-CH
		I T	C=C<-CH>-CO
		A	C=C---CO
		A	C=C-C<-C>-CO
		A	C=C-C<-C>2-CO
		A	C=C-C<=C>-CO
A			C=C-C-C-CO
A		A	C=C-C-C=C-CO
		A	C=C-C-CH-CO
A			C=C-C-CH=O
		AT	C=C-CO
		T	C=C-C=C-C=C-CO
B			C=C-C=C-CO
		I	C=C-CH-CH-CO
		I A	C=C-CH=O
B A		BI AT	C=C-CO
A		A	C=C-CO-C
B			C=C-CO-C=C
		B	C=C-CO-CH
		I	C=C-CO-CH<-C>-CH
		I	C=C-CYCLOPROPANE-CO
A			C=C=C-C-CO
		I	C=CH-C-C-CO
		T	C=CH-CH=CH-CO
		I	C=CH-CO
A		A	C=CH-CO-C
		B	CH---CO
		T	CH-C<-C>-CO
		A	CH-C-C-CO
A		A	CH-C-CH=O
		A	CH-C-CO
		A	CH-C-CO-C
		A	CH-C-CO-C-CH
		B AT	CH-C=C-CO
		T	CH-CH-C=C-CO
		I T	CH-CH-CO
B A		BI AT	CH-CO
		B	CH-CO-CH2
		BI AT	CH=C-CO
A		B A	CH=O
B			CH2---CO
		T	CH2-CH2-CO-CH3
B		I T	CH2-CO
		T	CH3-C-CH-CO
		I T	CH3-CO
		I	CYCLOALKANONE
A			CYCLOHEXANONE
		T	CYCLOPROPANE-C=C-CO
A			DECALONE<2>
		B	2-CYCLOHEXENONE
		B	3-CYCLOHEXENONE

in his target molecule and in the reaction centers and substructures of the thesaurus for his computer search, he will find new and often unexpected synthetic possibilities.

To illustrate the preceding paragraph, the section comprising formulas with one carbonyl group is shown in Table I. Column H indicates a special Roche proprietary file, column S the literature file.

Finally, the following important detail should be emphasized. The reaction centers and substructures *after* the reaction can be displayed independently of those present *before* the reaction. This fact allows the following question, which is fundamental for the preparative chemist, to be easily answered: How can the synthesis of a specific function or group of atoms be realized? And the user does not need to know the starting structure and thereby already part of the answer.

b. Other Data. The structural features of the full structures before the reaction (RCT) and equally the full structures after the reaction (PRO) are represented by a series of independent, searchable descriptors. With their help all important information on the target compound can be described by standardized units. This allows searching for reaction conditions which do not affect certain functions or groups of atoms, but convert another specific center BEF in the required center AFT, or more general, reaction conditions which lead to the formation of a specific reaction center AFT or to a specific target substructure.

Special interest is accorded to the reaction conditions. Here again, they are described by standardized terms which give information on reagents, solvents, special experimental features, and auxiliary materials to be used. This allows, on the one hand, a simple description of the facts, on the other, a flexible, differentiated search. Table II shows how the thesaurus for the reagents is organized and how many reagents each section contains.

In addition to these reagents and auxiliary materials 16 other indications on the reaction can be made:

PH CONSTANT	VACUUM
NEUTRAL	LOW PRESSURE
CONTINUOUS PROCESS	HIGH PRESSURE
ION EXCHANGE	IRRADIATION
MOLECULAR SIEVE	ELECTROLYSIS
SOLID PHASE	BIOCATALYSIS
PHASE TRANSFER	PREP. IMPROVEMENT
CHROMATOGRAPHY	KNOWN

Table II

elements	30
inorganic acids	19
organic acids	17
inorganic bases	26
amines	19
hydrides	22
carbonyls	12
organic compounds containing metals of the first 3 groups of the periodic table	37
organic compounds containing other metals	74
metal derivatives of reactants	13
further inorganic reagents	276
further organic reagents	205
	750

Also, the following general remarks on type and character of the reaction can be recorded in a special field:

CC FORMATION	INVERSION
CC CLEAVAGE	*E-Z* INTERCONV.
SALT FORMATION	NO STER. CHANGE
BASE FORMATION	RACEMIZATION
POLYMERIZATION	RESOLUTION
OXIDATION	STER. HINDR. OVERCOME
REDUCTION	PROTECT. GROUP
FUNCT. SELECT	LABELING
REGIOSELECT	PURIFICATION
MULT. CC-BOND SHIFT	X PRODUCTS
REARRANGEMENT	X CENTERS
ENANTIOSELECT	X SUBJECTS
DIASTEREOSELECT	X PROCEDURES
RETENTION	ALTERN. SCHEME

Furthermore, expressions for name reactions, such as "Wittig", "Vilsmeier", and so on, will be stored and, finally, bibliographic data and the name of the first author of the paper or the original document with the procedure will be added. There is no limitation on the number of factors that can be indexed. They can be extended as necessary.

2. Organization and Working Scheme. a. Selection of Reactions. The literature file (S-file) includes reactions and procedures of general interest from the most recent to old classic papers. New literature has been selected since 1975 from the latest issues of the most important scientific journals by Roche Nutley and the editors of *Synthesis*. The articles are abstracted, published in *Synthesis*, and stored in the reaction documentation system.

Reactions from older literature of the period 1946–1975 will be chosen from Vol. 1–30 of Theilheimer's "Synthetic Methods of Organic Chemistry". Reactions from the classical literature will be chosen from the first five volumes of "Organic Syntheses". Preparations for the selection of these reactions have already reached an advanced stage.

It is understandable that one can ask if the selection of reactions as it has been described is a valid or an arbitrary procedure. Behind this question looms the fundamental problem of "completeness" of a reaction file. In order to define a complete, comprehensive reaction documentation system, one should first know unambiguously when a reaction is new. Is this the case only when a new type of structure change has been used? Or also when known reaction conditions and a known type of structural change are combined in a new way? Or even when only slight modifications of the conditions give an improvement of the yield or other sizable advantages? These and similar questions show that the expression "new reaction" is largely a matter of personal judgment. "Completeness" of a reaction file is therefore an illusion. Important, however, is an optimum selection of the reactions to be recorded. And this quantitatively and qualitatively optimum selection can be realized only when it is carried out by a team of excellent chemists who work synthetically and think realistically. We believe we can claim that this is the case in our project.

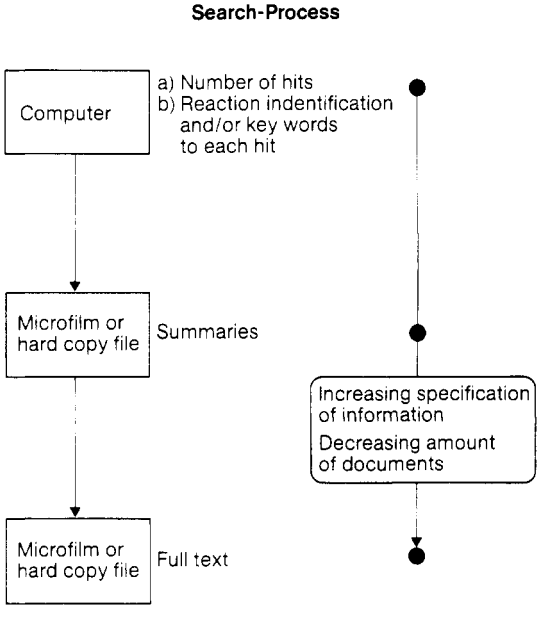


Figure 1. Search process.

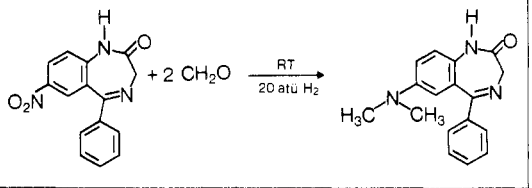
b. Input and Processing Data. Input of data is carried out with a Digital Equipment Corporation VT52 Display Terminal and a PDP 11/34 minicomputer. For processing, storage, and searching, an IBM 370/158 unit is used with the IBM program called document processing system (DPS). This DPS software has been adapted by our Data Processing Department to the special needs of the Reaction Documentation System. Searches are currently carried out in batch mode, but it is planned to switch to on-line searching using the IBM package STAIRS (Storage and Information Retrieval System).

c. Searching. A search is divided into three steps. In the first step, the computer states how many hits it has located and prints a list of the reactions found, with or without the corresponding keywords according to choice. In the second

```
H01-
  -0900/0

RCT2 : C15 H11 N3 O3 * NITRO * KETONE * C-AROMATIC *  HETEROCYCLE * 292 * 1829;  RCT2 : C H2 O * ALDEHYDE *
PRO1 : C17 H17 N3 O * AMINE * TER * KETONE * C-AROMATIC *  HETEROCYCLE * 292 * 1829 *
BEF2 : NO2;  BEF2 : 2 TIMES -CO * ALDEHYDE *
AFT1 : N<-CH>2 * TER *
CON1 : H * NI * RANEY * WATER * HIGH PRESSURE *
REMK : REDUCTION *
```

Figure 2.

Summenformel	Ausbeute		Nummer
$C_{15}H_{11}O_3N_3$	78%	900	6712



Katalysator	Lösungsmittel (Reagenzien)
Ra-Ni	H_2O/CH_2O

Name

Mogadon-Substanz

Figure 3.

step, the chemist will read the summaries of the citations found. He will have these summaries available in the form of cards or microfiches at his work-place, and easily locate them with the identification addresses provided by the computer. In the third and last step he finally consults the full text of the original paper or document, which whenever possible will be available to him at his work-place too. In this three-step process, the precision of the information increases from step to step; the number of useful hits decreases (see Figure 1). The following examples illustrate how a search is carried out.

A question for a search in the special file of the catalyst group reads as follows: How can a nitro group in a compound of the 1,5-benzodiazepine series be converted into a dimethylamino group using formaldehyde? Translated into the language of the system, the question consists of the following statements:

BEF = NO2
AFT = N(-CH)2
RCT = 1829 (Ring Index No. for 1,4-benzodiazepine)

The computer finds one reaction, which is numbered HO1-0900, and it prints the keywords of this reaction (Figure 2). Using the number 0900, the text of the summary card (Figure 3) of the H-file can be studied with a microfilm reader. More detailed information is available from the lab notebook under No. 6712.

In a similar manner the following questions concerning the file of the catalyst group have been answered:

Conversion (by hydrogenation)
- of primary alcohol groups into methyl groups
- of alkyl carboxylates into methyl groups
- of ketones into secondary alcohols
- of oximes into primary amines
- of the benzene ring into the cyclohexyl ring
- of nitrophenols into aminophenols
- of quinones into hydroquinones

Hydrogenation of the hetero ring in quinolines

Debenzylation
- of benzyl phenyl ethers
- of benzylamines
- of benzylamides

```
S01-

76-4800/0

SYNT

RCT2 : AMINE * PRIM * ESTER * T-O *  H2N-CH-COO-C; RCT2 : HAL * ALIPHATIC * BR OR J * POSS BZ-HAL *

PRO1 : AMINE * PRIM * ESTER * T-O *  N-C<-C>-COO-C *

BEF2 : CH;  BEF2 : C-HAL * BR OR J * POSS BZ *

AFT1 : C-C *

CON4 : KETONE * B F3 * DIETHYLETHER * BENZENE;  CON4 :  ORG.LI-AMIDE * THF;  CON4 : CARBOX.ACID WATER * OR
NH2-OH ACET.ACID ETHANOL;  CON4 : H CL WATER *

REMK : ENANTIOSELECT *

NAME :

LITE : YAMADA + AL,J.CHEM.SOC.CHEM.COMMUN.1976,136-137
```

Figure 4.

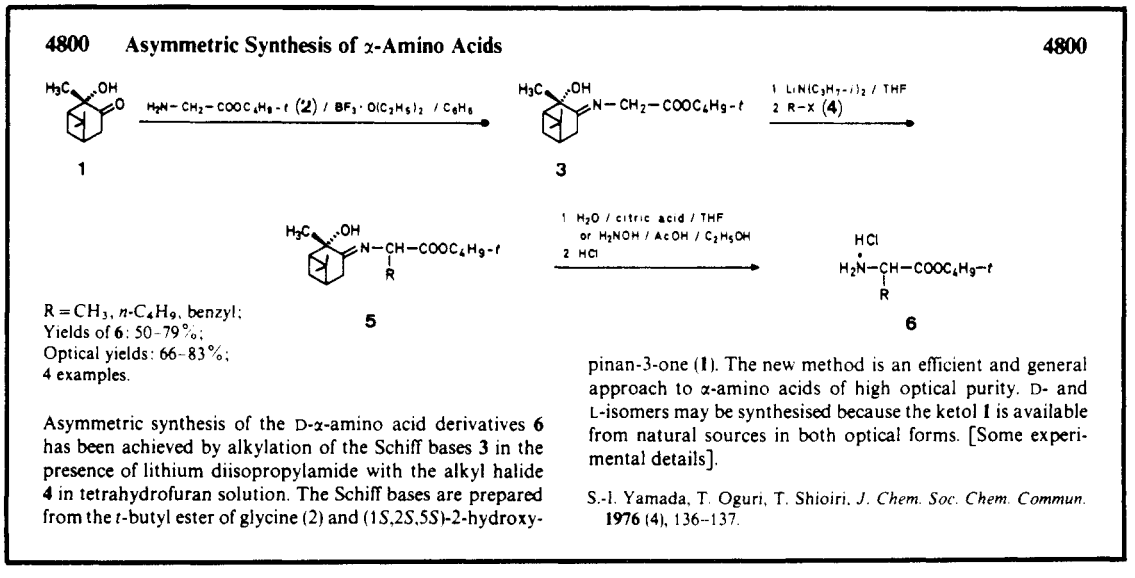
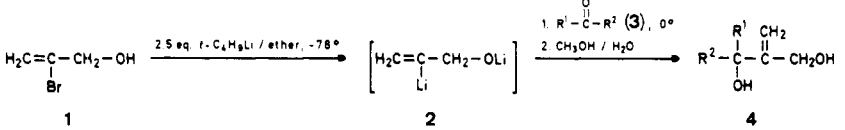


Figure 5.

```
S01-
76-4703/0
SYNT
RCT2 : C3 H5 BR O * ISOL * HAL * ALIPHATIC * BR * ALCOHOL * PM * HAL-C<=C>-C-OH; RCT2 : ALDEHYDE OR
KETONE * POSS ALICYCLE 293 *
PRO1 : ISOL * ALCOHOL * PM * SC OR TR * POSS ALICYCLE 293 * HO-C-C<=C>-C-OH *
BEF2 : C=C-HAL * ISOL * BR; BEF2 : -CO * ALDEHYDE OR KETONE * POSS <A>- 293 * 293 *
AFT1 : C=C-C-OH * ISOL * SC OR TR *
CON1 : BUTYL-LI * DIETHYLETHER *
REMK : CC-FORMATION *
NAME :
LITE : COREY + AL, J.ORG.CHEM.40,2975-2976<1975>
```

Figure 6.

4703 α-Substituted Allyl Alcohols **4703**



1 2 4

R^1 = H, alkyl;
R^2 = alkyl, aryl;
$R^1 + R^2$ = cycloalkyl;
Yields: 65-73%;
4 examples.

Treatment of the readily available 2-bromoallyl alcohol (**1**) with *t*-butyllithium in ether at −78° resulted in the formation of the *O*,2-dilithio derivative of allyl alcohol (**2**), which, upon reaction with carbonyl compounds **3**, gives rise to the α-substituted allyl alcohols **4** in good yields. Compound **2** is thus a useful new synthetic reagent. [Full experimental details].

E. J. Corey, G. N. Widiger, *J. Org. Chem.* **40** (3), 2975–2976 (1975).

Figure 7.

Monoalkylation of amines by alcohols

Selective hydrogenation

- of certain isolated CC double bonds of compounds containing several isolated CC double bonds
- of an isolated CC double bond in unsaturated aldehydes
- of a CC double bond conjugated with an aldehyde group in the presence of an isolated CC double bond

Conversion

- of the carbonyl group of citral into a primary alcohol group by selective hydrogenation
- of nitro groups into amino groups in the presence of aromatic halogens
- of 2-hydroxymethylfuran into 2-methylfuran

Hydrogenation

- in the presence of $CuCrO_3$
- in the presence of Lindlar's catalyst

The search process with the S-file is illustrated by the following four questions.

First Question. How can glycine or a glycine ester be alkylated in the α position with simultaneous formation of optical activity?

Search conditions are:

BEF = CH
AFT = C–C
REMK = ENANTIOSELECT

Inspection of the amino acids section of the thesaurus for reaction centers and reaction substructures results in the following target substructures useful for the formulation of our question:

N–C(–C)–COO–C
H_2N–C(–C)–COO–C

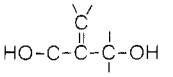
```
S01-
76-4746/0
SYNT
RCT1 : CONJ * N O-FUNCTION * ALICYCLE * 4781 *
PRO1 : KETONE * HETEROCYCLE * ALICYCLE * 4845 *
BEF1 : CC-BOND * N O-FUNCTION<A> * <+>-293 * 4781 *
AFT1 : AMIDE * <+>-355 * 4845 *
CON1 : TOSYL CHLORIDE * PYRIDINE *
REMK : REARRANGEMENT *
NAME :
LITE : BARTON + AL, J.CHEM.SOC.PERK.TRANS.1,1975, 1764-1767
```

Figure 8.

CO–N–C(–C)–COOH
CO–N–C(–C)–COO–C

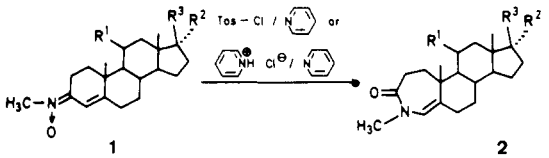
Their combination with the three search conditions mentioned earlier gives the printout in Figure 4. This printout carries the number SO1-4800. It is therefore the reaction in *Synthesis Abstract* 4800. If the chemist has the card edition of this service (Synthesis Abstract Service (SAS), Georg Thieme Verlag, Stuttgart) at his work-place, the card of the file gives the information in Figure 5.

Second Question. How can 1,3-diols be prepared in which both alcohol groups are allylic alcohols? This means a compound with the following partial structure is required:



In the polyalcohols section of the thesaurus of the reaction centers and substructures we find this target substructure spelled as follows: HO–C–C(=C)–C–OH. Use of this search condition gives the printout in Figure 6. From the *Synthesis*

4746 Rearrangement of Nitrones to *N*-Alkyl Amides 4746



1 2

R^1 = H, OH;
R^2 = H, OH, $OCOCH_3$;
R^3 = *n*-C_8H_{17}, $COCH_3$, $COCH_2OH$, OH;
Yields: 43–70%;
4 examples.

A novel rearrangement of *N*-methylnitrones (**1**, prepared from the corresponding ketones and *N*-hydroxymethanaminium chloride in pyridine) is carried out with *p*-toluenesulfonyl chloride or pyridinium chloride in pyridine – either in anhydrous pyridine or in the presence of 15 mol equiv of water – and affords the lactams **2** in high yields. In contrast to the Beckmann rearrangement, this nitrone rearrangement does not depend on stereochemistry; both *syn*- and *anti*-isomers give the same lactam. The new method provides a direct route to *N*-methyl- or *N*-substituted lactams which are less readily available from the Beckmann rearrangement. [Full experimental details].

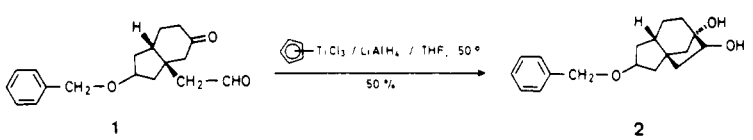
D. H. R. Barton, M. J. Day, R. H. Hesse, M. M. Pechet, *J. Chem. Soc. Perkin Trans.* **1**, 1975 (18), 1764–1767.

Figure 9.

```
S01-

76-4768/0

SYNT

RCT1 : ALDEHYDE KETONE * OR 2 TIMES ALDEHYDE * OR 2 TIMES  KETONE * POSS ALICYCLE 293 OR 1391 OR 1754 *

PRO1 : ALCOHOL * 2 TIMES SC * OR 2 TIMES TR * OR SC TR *  ALICYCLE 49 OR 293 OR 1391 OR 3136 OR 10389 OR
12368 *

BEF1 : CO---CO * ALDEHYDE OR KETONE * POSS <I>-293 *  293 OR 1391 OR 1754 *

AFT1 : CC-BOND * 2 TIMES ALCOHOL<A> * SC OR TR *  <R>-49 * 49 * OR <R>-155 * 1391 OR 10389 OR 12368 *  OR
<R>-293 * 293 OR 3136 *

CON1 : ORG.TI-COMPOUND * LI AL H4 * THF *

REMK : REDUCTION *

NAME :

LITE : COREY + AL, J.ORG.CHEM.41,260-265<1976>
```

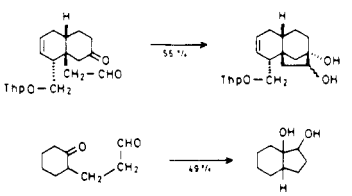
Figure 10.

4768 Cyclic Glycols by Intramolecular Coupling of Dicarbonyl Compounds 4768



1 2

Further examples:



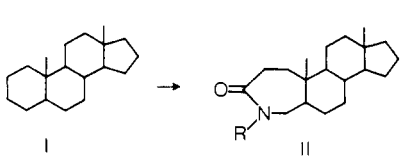
The reduction of cyclopentadienyltitanium(III) chloride with lithium aluminium hydride provides a new reagent which induces the intramolecular coupling of carbonyl compounds (e.g. **1**→**2**) in fair to excellent yields. Optimum reaction conditions involve reduction of 6 equiv. of cyclopentadienyltitanium(III) chloride with 4.5 equiv. of lithium aluminium hydride at 50° in tetrahydrofuran followed by rapid addition of 1 equiv. of the carbonyl compound. High dilution techniques are not necessary to inhibit intermolecular condensation. Other reactant ratios and solvent systems resulted in inferior yields. No general method for effective intramolecular pinacolic reduction has previously been reported. Classical conditions proved totally ineffective for these reactions. Thus, the present method extends considerably the scope and effectiveness of the pinacolic coupling reaction. [Full experimental details].

E. J. Corey, R. L. Danheiser, S. Chandrasekaran, *J. Org. Chem.* **41** (2), 260–265 (1976).

Figure 11.

Abstract card no. 4703, the information in Figure 7 is obtained.

Third Question. Can, and if yes, how can the A ring of a steroid system be enlarged to give a seven-membered lactam ring?



I II

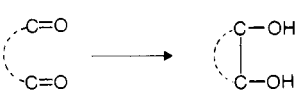
The search conditions have to be formulated in the following manner:

BEF = (+)-293 * 4781 *

AFT = (+)-355 * 4845 *

((+) = symbol for ring enlargement; the numbers correspond to the Patterson Ring Index: 293 for cyclohexane, 355 for diazepine, 4781 for the steroid ring system I, and 4845 for cyclopentanaphthazepine II). The search yields the printout in Figure 8 and the Synthesis Abstract in Figure 9.

Fourth Question. How can a ring closure be accomplished by condensation of the carbonyl groups of a dicarbonyl compound and formation of a 1,2-dihydroxy ring?



The following search conditions will be used:

BFE = CO——CO *

AFT = CC-BOND * 2 TIMES ALCOHOL(A) * (R)-

(The conditions which are described for AFT with standardized terms mean that the ring closure (R-; any ring size possible) is accomplished by formation of a CC-BOND with two alcohol groups (2 TIMES ALCOHOL(A)) on the ring.)

The search gives the printout in Figure 10 and the Synthesis Abstract in Figure 11.

ACKNOWLEDGMENT

Thanks are due to the many colleagues, most of them young chemists, especially to Dr. H. Scherrer, for their collaboration and to Mr. Margadant for writing the successful computer programs.

REFERENCES AND NOTES

(1) H. J. Ziegler, "Reactiones Organicae", A Documentation of Organic Reactions on Slotted Punch Cards, Georg Thieme Verlag, Stuttgart, 1965.
(2) H. J. Ziegler, "A New Information System for Organic Reactions", *J. Chem. Doc.*, **6**, 81-9 (1966).
(3) H. Grünewald, "Besprechung der Reactiones Organicae von H. J. Ziegler", *Angew. Chem.*, **79**, 830-1 (1967).

Chemical Reactions Information Retrieval from Chemical Abstracts Service Publications and Services†

A. J. BEACH, H. F. DABEK, JR., and N. L. HOSANSKY*

Chemical Abstracts Service, P.O. Box 3012, Columbus, Ohio 43210

Received March 20, 1979

Information on chemical reactions in general and on specific reactions such as oxidation may be found by searching CAS publications and services. Using common or author terminology, the searcher relies on the Keyword Index in the weekly issues of *Chemical Abstracts* (CA) for current awareness. Having consulted the cross-references and indexing notes in the Index Guide, the searcher accesses the controlled set of terms in the General Subject Index for in-depth and retrospective searching. In this index, the searcher finds reaction headings and headings related to reactions such as technological processes and classes of compounds. The Index Guide also includes a hierarchy of reaction headings to help the searcher choose the most appropriate heading. Reactants, intermediates, and products found in the Chemical Substance Index provide access to reactions as well. The entries in these indexes lead the searcher to the abstracts for further information concerning the primary documents in which the study was reported. Two examples of searching for reactions are given.

In the introduction to the First Decennial Index of *Chemical Abstracts* (CA), E. J. Crane, the Editor, wrote ". . . it is obvious that this First Collective Subject Index to the journal should be not only accurate and thorough but that it should also be so prepared and arranged that those who use it can find all of the references on the various subjects with certainty and with a minimum of effort". This statement, made in 1919, is still valid today, and the policies implicit in it lead to complete, accurate, consistent, and rational indexing of all that is new and significant in chemistry and chemical engineering. What holds true for chemistry in general holds also for reactions in particular.

The First Decennial Index had many entries for different types of reactions. There were, for example, about 280 entries at **Oxidation**, 190 at **Rearrangements**, and 70 at **Substitution**. The concept of cross-referencing was used: the Beckmann rearrangement was cross-referenced to **Rearrangements**, and at **Condensation** one was additionally directed to such specific headings as **Claisen condensation** and **Friedel-Crafts reaction**.

In addition to reactions, classes of compounds, such as amines, have also been indexed since the beginning of CA. These headings, together with the text that accompanies and expands upon them, provide additional access to reaction information. An entry in the First Decennial Index, for example, reads "**Amines**, addition reaction with diazonium salts". The entries for specific chemical substances also provide useful information about reactions that sometimes cannot be found at other entries. For example, starting compounds and intermediates, which are particularly valuable entries, have been extensively indexed since 1973.

It is obvious from this brief introduction that CAS publications and services can readily answer questions that can be expressed in terms of rather general reactions or classes of compounds. Questions that can be expressed in terms of specific compounds, whether such compounds are reactants, intermediates, or products, can also be answered easily.

Questions expressed in terms of structural moieties or substructures, or in terms of bonds broken and made, on the other hand, can be answered only with difficulty. If the searcher wants to know about reactions that involve a transformation of one substructure to another, the searcher must express that question either at a more specific level by thinking of a specific substance, or at a more general level by fitting it to a reaction heading or a compound class heading. The latter can be used if a substructure fits one or more such headings, but not all substructures correspond to index headings.

SEARCH FOR INFORMATION USING A GENERAL SEARCH QUESTION

Two questions, a general one and a specific one, will be explored to demonstrate how a searcher can use CAS publications and services to retrieve information about chemical reactions. The search strategy illustrated leads the searcher to the Keyword Index, and to the General Subject Index and

† Presented in the symposium, "Retrieval, Analysis, and Indexing of Chemical Reactions", 176th National Meeting of the American Chemical Society, Miami Beach, Fla., Sept 12, 1978.

0095-2338/79/1619-0149$01.00/0