Steps toward the Automatic Compilation of Synthetic Organic Reactions†

MALCOLM BERSOHN* and KEILLER MACKAY

Department of Chemistry, University of Toronto, Toronto, Canada M5S 1A1

Received October 5, 1978

A projected system is described which will collect relevant information on synthetic reactions from the English language journals on tapes and translate the information into a standardized form for use in a synthesis generating program. The preliminary implementation of various modules of this system is described. Much use is made of the stereotyped form and content characteristic of the experimental sections of papers and patents describing synthetic organic reactions.

INTRODUCTION

Summarization of a Document. The classical challenge to artificial intelligence research is to write a program that will read a paper and summarize its contents. A subset of this problem is to write a program, or, more accurately, build a system of programs which will inspect a document, determine whether its content applies directly to synthetic organic chemistry, and, if so, assimilate the essential part of its contents in condensed form into a database of synthetic reaction descriptions. In this paper we describe the beginnings of such an effort.

Synthesis-Generating Programs and Synthetic Reaction Information Retrieval Systems. A synthetic reaction information retrieval system consists of a file, or preferably a database, of synthetic chemical reactions together with programs that can query the system for information about synthetic reactions. The querying can be on line or off line. In contrast, a program that generates synthetic routes is basically off line, since usually at least several minutes of CPU time and much more than that in elapsed clock time are required to generate suitable syntheses.

The contrast between these two kinds of systems, one for retrieving information and the other for solving problems, is typical of many areas. In operations research, for example, one might retrieve the average return on investment historically achieved in a certain kind of expenditure. Alternatively one might use a program to produce the optimum mix of investments utilizing a large quantity of various data. The former kind of system is considered very desirable by the decision maker. The latter system, the total problem solver, is often suspect because it is hard to put the consideration of all variables for every kind of encountered situation into a program.

In a synthesis problem we typically have a goal substance and one or more acceptable starting substances. The acceptable starting substances may be listed specifically or else a definition of acceptability for a starting substance may be provided, so that when the program has progressed back from the goal molecule to a substance satisfying the definition, a synthesis will have been generated. If the length of a suitable synthetic pathway happens to be only one step, then the problem of generating a suitable route of synthesis reduces to retrieving information about a particular reaction. For this reason we may infer that the file of synthetic reactions can be the same for both systems.

In referring to this matter, Gund et al.[1] point out that there are hundreds of ways of reducing a ketone to a secondary alcohol. Each of these conditions is presumably most appropriate for particular combinations of functional groups elsewhere in the molecule. To make our terminology clearer, let us say that the reduction of a ketone to an alcohol is a synthetic reaction, but the reduction of the ketone to an alcohol with a particular set of experimental conditions is a "reaction variant". Variants of the same reaction are alike in that they bring about the same essential transformation in the generalized substructure of interest. They differ in the references cited and the experimental conditions. It is quite plain that the chemist retrieving information wants to know about some reaction variants in particular, not about all of them, because he has a particular substance in mind, with particular sensitivities to reaction conditions. The synthesis program must also choose between reaction variants. The reaction variant chosen will be that one out of those that are compatible with the other functional groups of the molecule that has the highest possible yield.

A synthesis program "understands" and directly uses the code describing the reactions; in contrast, the querying programs that interact with chemists on line must translate the code into natural language.

The contrasts between the two types of system seem minor compared to the fact that both require the building of a chemical reaction database. For both systems the automation of the building of the reaction database is a worthwhile goal. We can go further and say that there seems to be no reason why they cannot be one and the same database. The synthesis generating program should make direct contact with the physical records describing the reactions. Speed is very important. This is shown by the fact that our synthesis program simulates more than 250 reactions per second on the IBM 370/165 computer. In the course of doing this simulation it performs about 2000 retrievals of reaction records in this same second. The chemist user of a reaction database would be happy with one such record with a response time of a second. Therefore, it is feasible and proper to interface between the user and the ultimate physical records a command language and a data description language and then "access methods" for handling the physical records.[2]

Further, the database for direct use by chemists will need inverted lists, so that it can respond efficiently to queries such as: list the reactions that use zinc catalysis, list the reactions that occur in 90% yield, list the reactions of nitroso compounds, etc. The most advanced example of a chemical reaction database is the IDC system, reported by Fugmann.[3] The most extensive publically available synthetic reaction database is that of Derwent Publications Ltd., described by A. Finch in this symposium.

The publication in 1963 by Vleduts,[4] the organizer of this symposium, of the first paper outlining the potential use of a computer in generating syntheses aroused the interest of an eminent synthetic organic chemist. After years of urging by the latter, early in 1969 the senior author began to write a synthesis-generating program, embodying some of the important ideas of Vleduts.[5] The program is now built, in the sense that it manipulates stereochemistry, chirality, etc., 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.

0095-2338/79/1619-0137$01.00/0

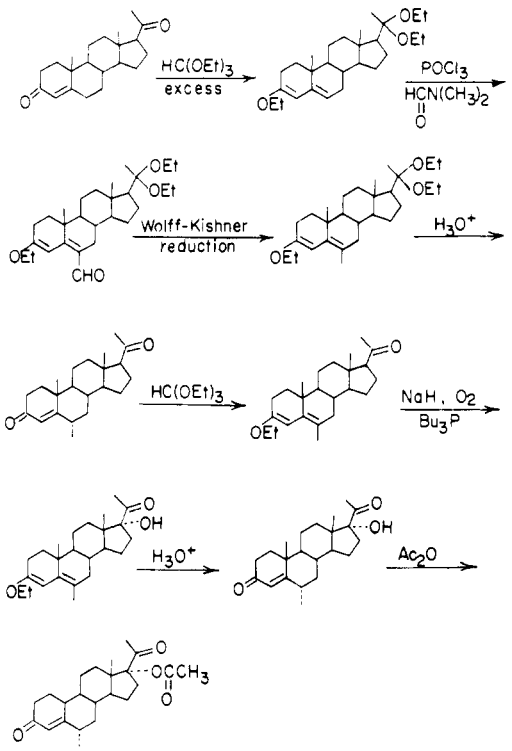


Figure 1. A suggested eight-step synthesis of Depo-Provera, proposed by the synthesis generating computer program.

generates many-step syntheses. A possibly very useful example, the eight-step conversion of progesterone to the birth-control substance, Depo–Provera, is shown in Figure 1.

As of this writing the synthesis-generating program has more than 20 000 lines, but most of this is the thinking part consisting of executable instructions; the data part has only a little over 300 synthetic reactions, which includes roughly 500 variants. We thus have the formidable problem of catching up to and thereafter keeping up with the literature on synthetic reactions.

The idea of using a computer program or a set of computer programs to accomplish this task seems visionary in 1978, but the idea of writing a synthesis-generating program also must have seemed visionary in 1963. The purpose of this paper is to outline a system of programs which will convert literature data to usable parts of the reaction database in our program. The system is by no means built, but we have beginning implementations of certain key aspects of the system. They are described here in the expectation of useful interaction with others who have somewhat similar problems.

The Task of Compilation. The work "compilation" in the title of this paper has two possible meanings. On one hand, searching of the literature and gathering of similar data together in one place is the process of compilation as used in library work. On the other hand, the translation of verbal data to a lower level form more directly usable by a computer is compilation in the computer sense. Our projected system will perform both these functions.

Some major steps in the compilation of synthetic reactions are as follows: (1) selecting a relevant document, (2) identifying in the document IUPAC names of the reactant(s) and product(s), (3) converting these names into connection tables, (4) deducing the unnamed reactants, and (5) finding the essential change in the reaction by using the intersection of reactants for different examples of the reaction to determine the most general form for the reactant(s) and proceeding similarly for the products. The other examples of the reaction may appear in the same document or in previously analyzed documents. The last steps are (6) the indexing of the reaction according to the substructure produced and the substructure(s) destroyed and (7) the conversion of the essential reaction change into bond replacements and the combination of the bond replacement information with yields, reaction conditions, and references to produce the reaction representation in a format usable by the synthesis program. If the reaction has been encountered before, there is an important step which consists of refining the specification of the substructure that is produced and the substructure(s) of the reactant that are destroyed.

Table I

acetalization	hydroboration
alkylation	lactonization
amination	methylation
cleavage	methylenation
coupling	methoxylation
deoxygenation	oxygenation
expoxidation	reduction
formylation	ring expansion
halogenation	silylation
homologation	thienylation

SELECTING A RELEVANT DOCUMENT

Relevance of a chemical document to organic synthesis is determined by the presence of certain keywords or key phrases in the title. The pharases collected so far are as follows:

1. synthesis of
2. a phrase of the type ⟨compound type⟩ from ⟨compound type⟩, e.g., "hydroxy-ketones from ketones" or "nitroalkanes from amines"
3. a phrase of the type ⟨process⟩ of ⟨compound type⟩, e.g., "alkylation of ketones" or "cyclization of, -dialkenes"
4. a phrase of this type "a ⟨process⟩ reagent", or "a ⟨process⟩ catalyst", e.g., "a dehydrogenation catalyst" or "a hydroboration reagent"
5. conversion of ⟨compound type⟩ to ⟨compound type⟩
6. addition of ⟨compound type⟩ to ⟨compound type⟩
7. preparation of ⟨compound type⟩
8. route to ⟨compound type⟩
9. reagent for ⟨process⟩.

Table I gives some further examples of the processes (or reaction types) to be recognized.

If the phrases "kinetics of" or "rates of" precede the process name, the document is discarded. Presumably physical-organic chemical studies are done on reactions whose steric course and synthetic scope are already known. This first section has not yet been programmed, but it is clearly the easiest one to code. The foregoing list of key-phrase types and modifing conditions is preliminary and will eventually be amplified many fold.

IDENTIFYING NAMES OF REACTANTS AND PRODUCTS

In the usual present-day format of experimental description, the IUPAC name of the product of the reaction appears in the title of a paragraph. The paragraph title is detected by the fact that it consists of a name preceded by a period and one or more spaces and immediately followed by a period. The IUPAC name of the reactant(s) are the first such names to be found after the title. The common organic solvents will be recognized as such and discarded. Occasionally the reactants are only indicated by number, and in this case a backward scan will have to be performed to find the first-named compound associated with this number. Usually they are to be found in the headings of previous paragraphs.

The foregoing procedure will often be inadequate for finding all the reactants because they are an inorganic solvent, such as water, or they are formed in the course of the reaction, such as HCN. The method our program uses for finding the missing reactant will be described later.

Table II. Preliminary Connection Table Obtained from the Hydrocarbon Root[a]

row no.	at. no.	degree of unsatn	no. of attached H atoms	row nos. of neighboring atoms	
1	6	0	2	2	
2	6	0	2	1	3
3	6	0	2	2	4
4	6	0	2	3	5
5	6	0	2	4	6
6	6	0	2	5	

$-CH_2-CH_2-CH_2-CH_2-CH_2-CH_2-$

[a] The values for the "degree of unsaturation" have the following meaning: 0, saturated: 1, aromatic: 2, carbon doubly bonded to another carbon atom: 4, an atom doubly bonded to another atom, wherein one of the two atoms is a heteroatom: 8, triply bonded atom.

Table III. Preliminary Connection Table, Cyclized Because of the Prefix "Cyclo"

row no.	at. no.	degree of unsatn	no. of attached H atoms	row nos. of neighboring atoms	
1	6	0	2	2	6
2	6	0	2	1	3
3	6	0	2	2	4
4	6	0	2	3	5
5	6	0	2	4	6
6	6	0	2	5	1

$-CH_2-CH_2-CH_2-CH_2-CH_2-CH_2-$

CONVERSION OF IUPAC NAMES TO CONNECTION TABLES

Vander Stouw and his group[6,7] at Chemical Abstracts Service have produced a huge program, in IBM 370 assembly language, which translates *Chemical Abstracts* names for compounds to connection tables conveying the structure of the named compound.[6,7] Because of uncertainty as to whether we can acquire this program for our system, we are writing a program, in COBOL, that performs this conversion. The general scheme is as follows. The scan is from the end of the name to the front. The last alphabetic string in the name is compared with the members of a list of suffixes; if it does not correspond to any suffix, a systematic search for possible prefixes and suffixes inside this string is instituted and any such are removed. Thus, cyclohexanol is converted to "hex". Having found the root, e.g., "hex", the correct number of atoms corresponding to the root is retrieved from a table. Then a connection table is generated in which all rows are the same; i.e., they represent methylene carbon atoms. The number of rows is that retrieved after finding the root, in this case six (cf. Table II).

Analysis of the Prefix. Expressions such as 2-aza or 3-oxo, etc., trigger alterations of the atomic number column and the number of attached H atoms column for the identical row. If the prefix "cyclo" is found, the first and last atoms are connected, producing the connection table of Table III.

If the prefix "cyclo" is absent, a hydrogen atom is added to each of the terminal atoms and we obtain the connection table of hexane (Table IV).

The names of substituents are found before the name of the main structure. If the name indicates a substituent at atom number 3, then in row 3 of the preliminary connection table the number of hydrogen atoms is reduced by one and the row number of the connecting atom is added to the list of neighbor atoms in row 3. A connection table for the substituent is added, the row numbers to start with seven, all of the atoms to be methylene carbons at first. The number of hydrogen atoms in row 7 is reduced by one and the number 6 is added to the list of the neighbor atoms in row 6. Thus for 3-ethylhexane, we obtain the eight rows of Table V.

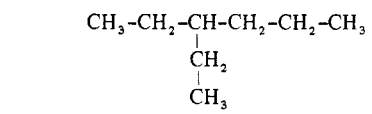
Table IV. Preliminary Connection Table, without Substituents

row no.	at. no.	degree of unsatn	no. of attached H atoms	row nos. of neighboring atoms	
1	6	0	3	2	
2	6	0	2	1	3
3	6	0	2	2	4
4	6	0	2	3	5
5	6	0	2	4	6
6	6	0	3	5	

$CH_3-CH_2-CH_2-CH_2-CH_2-CH_3$

Table V. Preliminary Connection Table, with Substituent

row no.	at. no.	degree of unsatn	no. of attached H atoms	row nos. of neighboring atoms		
1	6	0	3	2		
2	6	0	2	1	3	
3	6	0	1	2	4	7
4	6	0	2	3	5	
5	6	0	2	4	6	
6	6	0	3	5		
7	6	0	2	8	3	
8	6	0	3	7		



Numbers within parentheses must be adjusted by adding the number of atoms in the root structures. Thus the bromine atom in 3-(2-bromoethyl)hexane is given the row number nine.

The foregoing simple examples illustrate our general approach. The program is in COBOL and is still in the early stages.

DEDUCING THE UNNAMED REACTANTS

If the reaction is of the type A + X → B, where X is not specifically named as an organic reactant, either because it is inorganic, such as water or molecular oxygen, etc. or because it can also serve as a solvent, then we are required to find X. We do this by a COBOL program which maps A onto B. The program finds the largest common fragment of the structures of A and B. We will say that an atom of A corresponds exactly to an atom of B if the atomic number, degree of unsaturation, number of attached hydrogen atoms and size of rings in which the atom is a member all agree exactly. If an atom of A agrees with an atom of B only in atomic number, we will say that the two atoms correspond partially. Consider the reaction of cyanohydrin formation: $CH_3C(=O)CH_3 + X \rightarrow CH_3C(CN)OHCH_3$. The four nonhydrogenic atoms of acetone map at least partially onto four corresponding atoms of the cyanohydrin. The CN group and a hydrogen atom of the product have no correspondents in acetone. We link the atoms with no correspondents and obtain HCN as the missing reactant. This is the general procedure.

FINDING THE ESSENTIAL CHANGE IN THE REACTION

Classification of the Reaction. Classification is effectively the same as indexing. This process was done manually in Theilheimer's work. Automation of the process has been investigated for years by Lynch and his group.[8-11] In our projected system we will use our synthetic substructure discovery program, part of our synthesis-generating program. The substructure discovery program will list the synthetic substructures of the product and those of the reactants. A

comparison of the two lists shows what new substructure(s) have been produced and what old substructures have been modified. In our substructure discovery program, each substructure has a unique number which is computed by the program, so that the reaction can then be classified in terms of the product substructure primarily and the reactant substructure(s) secondarily. In our project we are only interested in such classification insofar as it provides us with an index for categorizing the reaction so that it can be stored in the right place in the reaction database and the program can ascertain whether the reaction is new or is a new variant or is another instance of a known variant of a known reaction.

In our synthesis program a reaction must be described so that a section of our program will know exactly where to add hydrogen atoms, replace bonds, etc. Merely indexing the reaction is insufficient. The atoms of the product substructure are given a canonical numbering and any extra atoms which are necessary in the reactants but are absent in the product are also given canonical "extra" numbers. A reaction description in our synthesis generating program then reads like "replace a bond from substructure atom 4 to a hydrogen atom by a bond from substructure atom 4 to extra atom number 1", etc.[12]

Finding the Reacting Atoms. When our program finds the largest common fragment of reactant molecule(s) and product, it discovers that some of the atoms correspond exactly and others do not. The latter are the reacting atoms. They will ordinarily differ in the number of hydrogen atoms or the number of neighbors. This is because in the course of the reaction they have lost or gained neighbors and/or hydrogen atoms. (In epimerizations, only the chirality is changed.) The reacting atoms must be identified by number in the particular molecules and subsequently in the substructures with their generalized numbering.

For our present purposes it is by no means adequate to discard whatever atoms do not react. This would provide a means of classifying the reaction à la Theilheimer,[13] but it would not give a recipe for converting an arbitrary product molecule connection table containing the true substructure of interest into the corresponding reactant connecton table(s). Consequently, we have to be concerned with atoms that do not participate in the reaction but facilitate it or direct it. They can be found by obtaining the intersection of all reactants in every example of the reaction. This is apparent in Figure 2, where the intersection of different cases of the reaction shows that all that is necessary is C-C(=O)-CH + CHI. The program that finds the intersection of reactants and products of different example of the reaction is just a slight modification of our COBOL program that maps reactant atoms onto product atoms in one example of the reaction.

If the variant of this ketone alkylation is the Stork reaction (enamine alkylation), then $CH_2C(=O)-C-C$ will not be a possible product substructure since in the Stork variant the carbon to be alkylated cannot contain more hydrogen atoms than the other α carbon atom. If the Stork variant is being considered, the experimental conditions are duly noted, i.e., the presence of a secondary amine and subsequently of aqueous acetic acid. These conditions and the journal name, date, and page number are tied together in the description of the reaction variant. Another variant would indicate the presence of very strong base, such as butyllithium. In the latter variant $CH_2C(=O)-C-C$ is an acceptable product substructure.

Each synthetic reaction example found in the literature after indexing will be compared with such an indexed reaction in the reaction database. If the intersection of the product in the particular case with the general product substructure conveyed by the index shows that part of the general substructure is unnecessary, then the description of the reaction

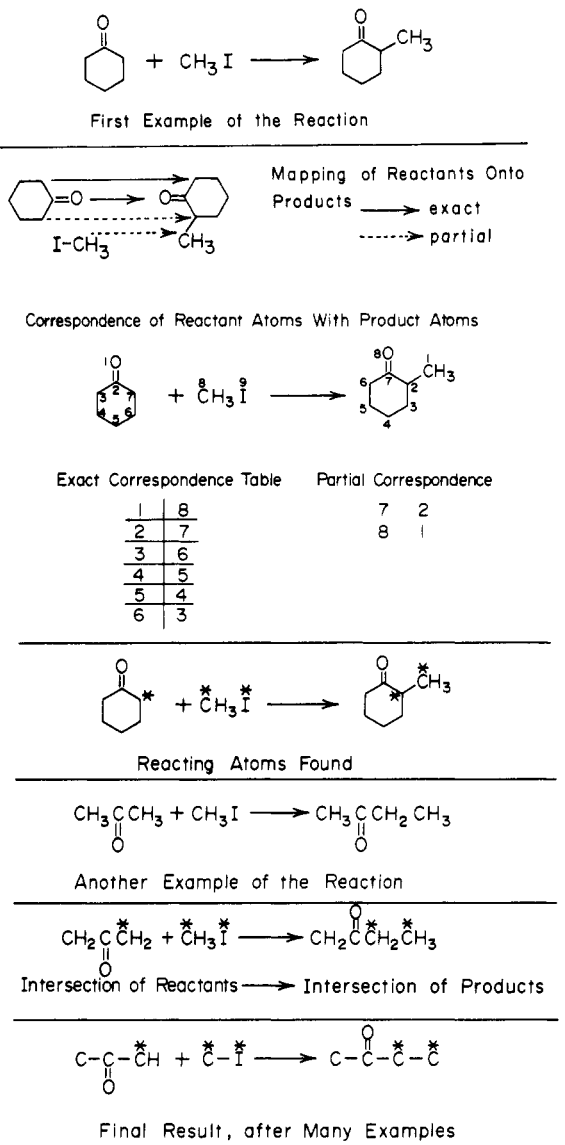


Figure 2. An outline of the operation of the programs that analyze a reaction by first mapping reactants onto products and then finding the intersection of all reactants and mapping it onto the intersection of all products for several examples of the reaction.

must be revised. For example a ketone reaction might have been done first on a β-keto ester and only a case where the ester group was absent would rule out the ester group as a necessary part of the product substructure. Two cases are said to be examples of the same reaction if they produce the same product functionality and destroy the same reaction functionality.

In the experimental part of papers and patents the yield is indicated as such by the word yield so it is simple for the program to note the figure. The experimental conditions are a much more challenging problem. At the least the program will have to know how to calculate the molarity of added reagent, to determine, for example, whether the reaction conditions were strong acid, weak acid, etc. Further, the words "added", "stirred overnight", etc. will have to cause specific responses in the reading program.

It is apparent that the process is quite complicated and for many years after starting up the system an editorial review of the efforts of the machine abstracter will be required.

The conditions under which a reaction should be avoided are not so simple to find since they are usually not reported in the experimental section of the paper but in the discussion section. Some of this difficulty can be circumvented because

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