Chemical Society, gratefully acknowledges this support.

REFERENCES AND NOTES

(1) Morgan, H. L. "The Generation of a Unique Machine Description for Chemical Structures—A Technique Developed at Chemical Abstracts Service", *J. Chem. Doc.* **1965**, *5*, 107.
(2) Leiter, Jr., D. P.; Morgan, H. L.; Stobaugh, R. E. "Installation and Operation of a Registry for Chemical Compounds", *J. Chem. Doc.* **1965**, *5*, 238-242.
(3) Dittmar, P. G.; Stobaugh, R. E.; Watson, C. E. "The Chemical Abstracts Service Chemical Registry System. I. General Design", *J. Chem. Inf. Comput. Sci.* **1976**, *16*, 111-121.
(4) Freeland, R. G.; Funk, S. A.; O'Korn, L. J.; Wilson, G. A. "The Chemical Abstracts Service Chemical Registry System. II. Augmented Connectivity Molecular Formula", *J. Chem. Inf. Comput. Sci.* **1979**, *19*, 94-97.
(5) Blackwood, J. E.; Elliott, P. S.; Stobaugh, R. E.; Watson C. E. "The Chemical Abstracts Service Chemical Registry System. III. Stereochemistry", *J. Chem. Inf. Comput. Sci.* **1977**, *17*, 3-8.
(6) Vander Stouw, G. G.; Gustafson, C.; Rule, J. D.; Watson, C. E. "The Chemical Abstracts Service Chemical Registry System. IV. Use of the Registry System to Support the Preparation of Index Nomenclature", *J. Chem. Inf. Comput. Sci.* **1976**, *16*, 213-218.
(7) Zamora, A.; Dayton, D. L. "The Chemical Abstracts Service Chemical Registry System. V. Structure Input and Editing", *J. Chem. Inf. Comput. Sci.* **1976**, *16*, 219-222.
(8) "Inventory Reporting Regulations (40 CFR 710)", *Fed. Regist.* **1977** (Dec 23), *42* (No. 247).

Computer-Assisted Synthetic Analysis at Merck†

PETER GUND, EDWARD J. J. GRABOWSKI, DALE R. HOFF, and GRAHAM M. SMITH*

Merck Sharp & Dohme Research Laboratories, Rahway, New Jersey 07065

JOSEPH D. ANDOSE* and JOSEPH B. RHODES

Management Information Systems, Merck & Co., Inc., Rahway, New Jersey 07065

W. TODD WIPKE

Board of Studies in Chemistry, University of California, Santa Cruz, California 95060

Received January 24, 1980

The Simulation and Evaluation of Chemical Synthesis (SECS) program has been implemented at Merck and has been evaluated by approximately 50 synthetic chemists. The results of this evaluation are summarized, highlighted by several examples of SECS analyses, and future plans for the program at Merck are discussed. The most critical problem which must be solved is that of developing a practical data base of synthetic reactions.

In the decade since the first published description of a computer program capable of deriving synthetic routes to complex molecules,[1] the field has grown into a lively discipline.[2-4] Nevertheless, it cannot yet be said that computer-assisted synthesis is an accepted research tool for many practicing synthetic chemists.

It is easy to see the appeal of computer aids to synthesis. Organic synthesis is complex; in principle, several thousand reactions could be applied in several steps to millions of potential starting materials in order to prepare a desired product.[2,3] The computer has the capability to extend the chemists's perception and memory, and to systematize and organize his synthetic knowledge.

While Merck has for a long time recognized the computer's potential in this area,[5] the decade was half over before the field was sufficiently advanced for an in-house effort to be considered. When a computer-assisted synthesis project was begun, a number of major design decisions were quickly made. To assure data security and to encourage optimal use, we wanted to run the program in-house. We wished to involve the synthetic chemists directly in the analyses, which required interactive time-shared program operation on graphics terminals remote from the computer. This also meant that synthetic analyses would not be run exclusively as a scientific information service. Finally, by using the same graphics terminals for computer-assisted synthesis and for the Merck Molecular Modeling Project,[7] we could make more efficient use of our resources. A brief survey of extant programs indicated that the Simulation and Evaluation of Chemical Synthesis (SECS) program[6,8,9] was most suited to Merck's needs.

Today we are running SECS on our corporate IBM computer, with the program accessed by four graphics terminals in three research laboratories (Rahway, New Jersey; West Point, Pennsylvania; and Montreal, Canada). About 60 chemists have been checked out on running SECS analyses, and the program is used almost daily. About 20 volunteer chemists have contributed chemistry for the program, and two chemists are adding chemistry to the file.

At this stage in the program's development, it is appropriate to report our experiences in implementing, evaluating, and enhancing the SECS program at Merck, and to assess progress toward the ultimate goal of providing a routinely useful aid for the practicing synthetic chemist.

IMPLEMENTATION OF SECS AT MERCK

This section gives a brief discussion of how SECS was converted to run in the Merck environment.[10]

In October of 1974, one of us (W.T.W.) made available version 1.0[11] of the program consisting of approximately 25000 lines of code written for a Digital Equipment Corp. (DEC) PDP-10 computer (ca. 20000 lines of FORTRAN written for

†Presented in part at the Symposium on Computer Assisted Drug Design at the 177th National Meeting of the American Chemical Society, Honolulu, Hawaii, April 1979. A more detailed version of this paper will be published in the Proceedings of this meeting.

0095-2338/80/1620-0088$01.00/0

DEC's F-40 compiler and about 5000 lines of assembly language code written in MACRO-10) together with another 5000 lines of code to support a DEC GT-40 series graphics display terminal (written in MACRO-11 assembly language for a PDP-11 minicomputer).

In addition, there was supplied a preliminary data base of approximately 250 chemical reactions written in the ALCHEM language[8] and treated as data by the program. Our immediate objective was to translate the program into code that would operate on our corporate IBM computer—at that time a model 370/155—under the Time Sharing Option (TSO).[10]

For the purposes of implementation, the program was divided individually into a number of smaller program modules, and then implemented in an order which would permit testing of the program during the intermediate stages of assembly. Where possible, sections were tested independently prior to integrated testing.

Implementation of the program, especially through the intermediate stages, was greatly facilitated by three factors. First, the program contained extensive built-in debugging facilities which could be selectively activated by setting appropriate "debug" switches as part of the input options. Second, a working version of the SECS program, available through a commercial time-sharing service,[12] could be used to compare debug output from our version, allowing the detection of subtle errors introduced during the implementation process. Finally, the program developer (W.T.W.) was available as a project consultant for assisting with difficult problems.

Following implementation of the SECS program and based on our experience in using the program in a production environment (see below), a number of enhancements were incorporated into our version of SECS. These enhancements include a more streamlined (continuous drawing) format for molecule input using a light pen, the option to selectively display at the terminal only those structures which successfully pass evaluation, brightening of lines in the synthesis tree corresponding to structures which have been rated "GOOD", and lastly, a hard copy (Calcomp) summary of the results of an analysis. All of these enhancements have also been incorporated into subsequent versions[11] of the parent SECS program at Santa Cruz by Professor Wipke, although several (such as hard copy) were implemented in a different manner.

EVALUATION OF PROGRAM UTILITY

Once SECS was implemented and running at Merck, it was necessary to involve the chemists in evaluating the program's utility. However, we were concerned that, without being given background information, a chemist might see one or two poor SECS analyses and dismiss the methodology out of hand. In essence, we were asking the chemists to evaluate the utility of an eventual "production" system based on the current developmental system. In order to make such an evaluation validly, we felt the chemists needed some background in how the program actually operated.

Consequently, in the fall of 1976 we organized a course for participating chemists. A total of 25 senior level chemists from six departments in Rahway (Basic Synthetic Research, Process Research, New Leads Discovery, Arthritis & Antiinflammatory Research, Membrane Chemistry, and Radiochemistry) attended. The material (Table I), presented in 15 one-hour lectures, was detailed enough to dispel some of the mystery about how a computer can "perceive" a chemical structure, "remember" chemical reactions, and "suggest" possible precursors. Despite the heavy technical content and the chemists' busy schedules, all participants but one completed the course. One chemist, who was transferred to Merck's West Point, Pa., laboratories commuted one day a week to Rahway in order to complete the course. A spin-off from preparing the lecture materials was a comprehensive SECS user's manual. The first lecture was also published.[3]

Table I. Merck Course on Computer-Assisted Organic Synthetic Analysis

1. Overview of the field
2. Overview of SECS program (W. T. Wipke, 3 h)
3. Program operation; introduction to IBM time-sharing (TSO)
4. Review of typical synthetic analysis
5. Program modules, molecule drawing, structure representation
6. Molecule perception
7. What is a transform?
8. Transform types
9. Detailed walk-through of a coded transform
10. Precursor evaluation, strategy
11. Model building, steric and aromatic effects
12. Writing a transform

In addition to attending the lectures, each participant attended workshops in groups of five and used the program on real synthetic problems. Each chemist performed at least two analyses, for a total of over 50. For perhaps two out of three compounds analyzed, nothing useful was discovered. Occasionally, however, really novel approaches were found to current synthetic objectives.

When the course was finished, the participants were asked to complete questionnaires and write evaluations of the project. We received 22 completed (anonymous) questionnaires and 15 reports. In the fall of 1977, after a terminal was installed in West Point, the course was repeated for 21 Medicinal Chemistry Department chemists. At the conclusion of this course, 16 questionnaires were returned.

The following conclusions emerged from the chemists' evaluations.

1. The methodology offers great promise for aiding synthetic design. The system can improve and extend the chemist's own analytical abilities and stimulate the consideration of new approaches. However, it can neither supplant the synthetic chemist nor relieve him of the task of working out a detailed synthetic plan.

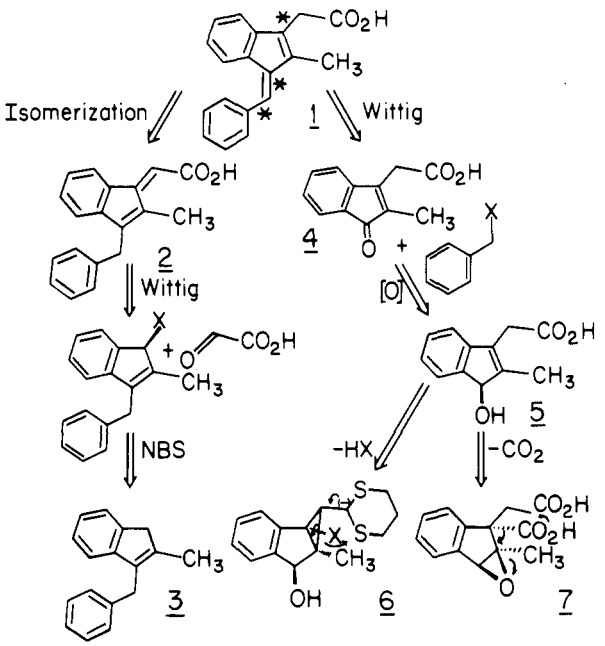
2. System hardware was acceptable. Using the light-pen to "point" to the GT42-GT43 display screen is a convenient input method for the chemist. The IBM time-shared system (TSO) presented some problems for the computer-naive chemist (e.g., cryptic error messages), and often suffered from poor response times; at other times the system was quite satisfactory. Similarly, the SECS program design was considered excellent. There was a strong desire for hard copy summary of analyses (one member of each workshop recorded each analysis; automatic analysis summary on the plotter became available after the Rahway course was over).

3. The version of SECS (1.0) which we implemented[11] was considered to need extensive further development. The chemists found that the preliminary data base needed expansion and revision. This was expected, since the supplied transforms had been written by the system developers to exercise various program features rather than in a systematic attempt to cover the field of synthetic organic chemistry. The chemists also felt that the analysis was too automated; they wanted more control over what chemistry was applied (more recent versions of SECS have high level strategies for this purpose). Many other, relatively minor, complaints and suggestions were made.

4. It was widely believed that such a detailed course was not necessary for learning how to operate the SECS program.

Since the completion of the courses, chemists have continued to participate in the project by improving the data base of reactions (as discussed below) and by using the program.

Scheme I



During 1978, monthly use of the program ranged from a low of 6 to a high of 71 runs. This year we have offered first level SECS analyses as a service, similar to the way Dr. Beryl Dominy has run the program at Pfizer Corp.;[13] this has been well received.

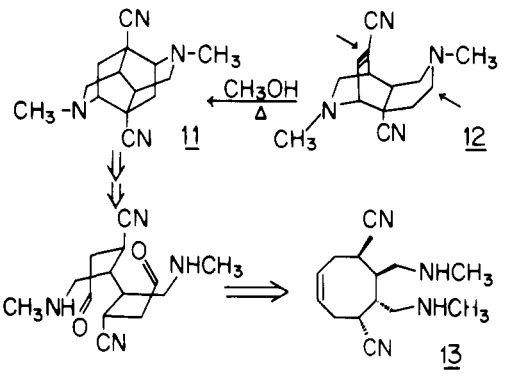
We plan to implement a more up-to-date version of SECS at Merck, and (as discussed below) we are working to improve and expand the program's chemical knowledge. As the program improves, we hope SECS will gain increasing acceptance here as a research tool.

DISCUSSION OF SELECTED ANALYSES

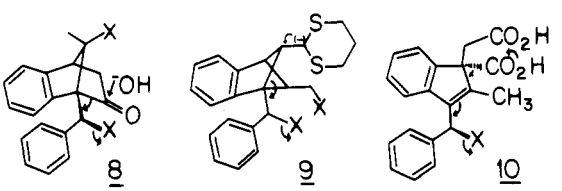
A review of the Merck experience with computer-assisted synthesis would be incomplete without some discussion of actual results which have been generated by SECS. In doing so, one must remember that we are not working with a "finished product", but instead with a developing research tool which is the subject of current research efforts, and that our analyses are based on software which was "state-of-the-art" in 1974.[11] Still, the best way to gain insight into the strengths and weaknesses of the synthesis program at this stage in its development is by considering the results of a few specific analyses. Structure **1** represents the nucleus for the new Merck product Clinoril®, and was our first candidate for a computer synthesis trial. It was an excellent test, because the SECS program was not particularly developed for analyzing structures such as **1**. Furthermore, an extensive body of chemistry had been developed for structures like **1** prior to the SECS analysis, so that the computer suggestions could be evaluated in light of laboratory experience. Finally, **1** remains an excellent target for new synthetic approaches. Part of the analysis developed by the chemists with the aid of the computer is summarized in Scheme I.

During the initial processing of **1** the starred bonds (*) were indicated to be "strategic" bonds.[14] This proved most satisfying since one of the recently reported syntheses of Clinoril is based on the formation of these bonds.[15,16] Valence isomer **2** was not recognized by SECS 1.0 as a precursor of **1**, and had to be processed separately. The combined processing of **1** and **2** did generate the basic elements of the synthesis of Clinoril discussed in ref 15. However, one key point missed by SECS is the acidity of **3** and its amenability to base-catalyzed condensations. SECS did propose **3** as a precursor of **2**, but by way of retro-synthetic Wittig and NBS reactions. The failure of SECS to recognize either the acidity of **3** or the viability of **2** as a precursor to **1** simply reflects deficiencies in the reaction data base at the time of the analysis, as opposed to any defects in the program's overall strategy.

Scheme II



Processing of intermediate **5**, which derived from **1** via retro-synthetic Wittig and oxidation reactions, gave rise to precursors **6** and **7**. A priori, the complex structures **6** and **7** (having four and three chiral centers, respectively) would appear to be poor precursors of the achiral target structure **1**. However, the suggestion of **6** and **7** by the synthesis program incited chemists working on the problem to think about fragmentation processes as a route to **1**. Such processes had not been considered previously because of the planar nature of **1**. As an outgrowth of the SECS analysis of **1**, reactions suggested by the arrows in structures **8**, **9**, and **10** were proposed by the chemists as possible precursors to **1**. Although these proposals did not, in fact, lead to a significant new synthesis of **1**, this example illustrates the program's ability to induce chemists to think about new approaches to familiar synthetic problems.



The potential of computer synthesis for fully and systematically analyzing routes to complex, polycyclic molecules has been noted before.[6] Recently the chemistry of 5,11-diazaditwistanes has been extensively studied in these laboratories.[25] Compound **11**, the key analog providing entry into this series, has been prepared from **12** as indicated in Scheme II. Not surprisingly, this approach was not suggested by SECS; indeed, the necessary hydride-transfer reaction was only discovered when thermolysis of **12** unexpectedly gave rise to **11**. However, an intriguing route via a double Mannich reaction and ozonolysis of an appropriate cyclooctene derivative (**13**) did result (Scheme II). Functional group protection is unavailable in our current version of SECS; consideration of protecting groups makes the analysis of Scheme II somewhat longer.

The two examples shown here, among many that have been performed, are representative of our current program's capabilities. These examples also show the importance of the computer–chemist interaction, with the computer suggesting approaches and the chemist thinking them through and refining them.

FURTHER DEVELOPMENT OF SECS AT MERCK

Our experiences in using the SECS program indicated that further development was necessary for SECS to become

routinely useful to the synthetic chemist. While development of new program features and strategies was deemed essential, these tasks have so far been left mainly in the hands of the SECS program development group at Santa Cruz. We at Merck have concentrated on a different task: expanding and correcting the base of chemical reaction information available to the program, and establishing a system for efficiently writing reaction transforms.

The organization of the reaction information as data, separate from the SECS program, allows transforms to be written, tested, and modified easily. This has enabled us to concentrate on adding and correcting transforms without also having to make changes to the SECS program itself.

Supplied Reactions. The preliminary data base which was received with the SECS 1.0 program, containing approximately 250 transforms, was used unchanged in the initial evaluation phase described above. While many of the supplied transforms gave excellent descriptions of chemical reactions, other transforms were not fully developed and contained errors. Deficiencies in the data base manifested themselves in two ways. Obvious precursors to a target structure were often not generated because a reaction was missing or incorrectly described. Also, unacceptable precursors were generated from reactions which were incorrectly described or insufficiently specific. The problem given highest priority by the participating chemists was the presence of transforms which gave unacceptable results. The next highest priority item was the absence of important reactions. After some discussion a list was drawn up of the 15 least acceptable transforms. The most common features of these transforms were the chemically incorrect application of a reaction, excessive priority for an unimportant reaction, and/or application of a reaction despite an interfering functional group.

Transform Review Workshops. Workshops were organized to deal with the list of worst transforms. Attendance at the workshops, which met for an hour per week, was voluntary and ranged from a low of 3 to a high of 15. The format of the workshops was informal but followed the following outline.

The transform to be reworked was announced in advance, with literature references to the reaction being covered. A person familiar with ALCHEM (a "coder") gave a step-by-step walk-through a transform. The chemical and structural meaning of each statement was explained, triggering a wider discussion of the scope and limitations of the reaction. As new aspects of the reaction were covered, the coder made notes on qualifiers to be added or changed. One key objective of the coder was to prevent detailed discussion of experimental conditions which cannot be represented in a transform. From the notes on the meeting, a corrected transform was written and tested by the coder. Finally, this transform replaced the original. After the 15 transforms had been modified, a noticeable reduction in the number of bad structures was observed in analyses produced by SECS. From this experience two points became clear: the chemists did not want to learn ALCHEM, even though they were willing to volunteer their time and knowledge to improve the chemistry data base; and some other format was required to efficiently collect chemical information.

Reaction Review Seminars. At this point it was decided to change the emphasis from strictly transform correcting to writing new transforms. Reaction review seminars were instituted, in which a volunteer chemist would select any reaction of which he had special knowledge and prepare a seminar on its synthetic utility. The coder attended the seminar and took notes; the chemist also supplied his notes. The coder then wrote a transform or transforms and did preliminary testing. Next the chemist tested the performance of the transforms on target structures of his choosing. Based on the chemist's suggestions, the transform was corrected as often as necessary, and finally added to the data base.

The advantages of this approach were that the speaker selected the reactions and so had more interest in them, and that the seminar served to increase the current awareness of other chemists. The major disadvantage was our difficulty in inducing chemists to volunteer, since preparation for the seminar required a great deal of time. While this format lasted for only a short while, it resulted in significant upgrading of the data base.

Speaker	Resulting Transform(s)
R. A. Firestone	1-3 Dipolar Additions
R. W. Ratcliffe	β-Lactam Preparations
E. D. Thorsett	Ni-Coupling Reactions
E. J. J. Grabowski	Enamine Rearrangement

Current Transform Writing Procedures. At this point we held a project review meeting with the chemists to assess progress and point future directions. The chemists felt they would like to continue to improve the data base, and Dr. Tom Beattie suggested that we send a memo to each Merck chemist asking for volunteers to summarize reactions for transform writing. Response was gratifying; about 60 chemists volunteered to help. To facilitate this approach, we created a "reaction summary form" which was designed to assist the chemist and the coder in focusing on those aspects of a reaction which are important for transform writing. European companies running SECS have independently developed a similar form.[18] To further clarify the type of information required, a seminar was given on how to fill out the form, using an example of a reaction which had been written up as a transform. At the seminar a request was made for volunteers and 19 chemists agreed to participate.

The 15-page summary form consists of several sections, each focusing on a different type of structural feature. The first section requests reaction name, literature references, and a general equation. The next section is for a detailed description of the product substructure. Here the effect of possible substituents on the basic substructure and on reaction yield should be considered. Next the starting materials are similarly described. The following section is concerned with how rings affect the reaction, and whether the reaction is intermolecular, intramolecular, or either. The next section requests information on interfering functional groups, and how unsaturation affects the reaction. The last section concerns experimental detail, which is described as comments in the transform. There is also a request for several test structures and how well the reaction works for their synthesis.

It takes chemists 1–5 days to summarize one reaction; on the average, approximately 1 week is required for a coder to incorporate the information on the form into a working transform. Additional time is required for the chemist to test the transform and for the coder to make corrections. The procedure at this point seems to be accepted by the chemists. One point which is emphasized by the chemists is the need for inclusion of current references for each transform.

This process of collecting chemical information from experienced chemists and incorporating it into transforms has dramatically improved the quality of the analyses produced by SECS. However, there are still gaps in the program's knowledge of most reactions and this is reflected in incorrect applications of some transforms. The result is that the process of transform writing must be interactive and continuing. If a major point has been overlooked or a new fact is learned about a reaction and the deficiency comes to light in an analysis, the transform must be modified and retested. As new facts are incorporated and corrections made based on experience, the program's chemical knowledge can be expected to become continually more sophisticated.

Writing a transform is similar to writing a computer program. When there is a large amount known about a reaction, the result can be a very long and logically complex structure. In order to streamline this procedure, it has been divided into several steps, allowing the structure of the transform to grow in a way which facilitates testing at each stage.

An additional parallel with programming is that everything to be included must be stated explicitly; the SECS program is not able to draw inferences from its input. For example, if a particular reaction involves a carbonium ion, the transform must be written with qualifiers indicating that electron-withdrawing groups (e.g., esters, nitro, etc.) at the reaction center will inhibit the reaction. The requirement that all facts be stated explicitly is partially responsible for the complexity of most transforms. Ideally the analysis of scope and limitations of a reaction is a problem of combination and permutation of groups on the reaction substructures (the minimum set of atoms and bonds needed to key the reaction) and an estimate of how each of the combinations of groups will effect the course of the reaction. Obtaining this latter information is the most important and difficult phase, requiring an experienced chemist. However, considering a long list of possible substrates is not the normal approach used by a chemist and represents one of the barriers to obtaining new transforms.

A final parallel with programming is maintenance of the final product. As new information is discovered regarding a reaction and new references are published, it is necessary to periodically review each transform to be sure that each reflects this new knowledge.

At some point in the writing of a transform the question arises: What level of detail is required in a transform for it to give acceptable results? To consider this question, a second question must be answered; what are the goals of the chemist who is using the program? If he is searching for new ideas in the synthesis of a compound, then the transforms should be written without too much detail. The resulting analysis will be a larger and more general (but not strictly practical) presentation of ideas which would have to be refined further by the chemist. However, if a chemist is looking for a reaction to perform a specific task, he wants to know the known limits of the reaction and the transform should contain more detail. In this case the results will be a more concise analysis showing the application of a few specific transforms. Since a single data base is preferred and since the chemists who use the system have a wide range of requirements, the transforms written at Merck are somewhere between the extremes described and tend toward a more detailed description of the reaction (the chemists seemed more annoyed by generation of inappropriate precursors than by the absence of obvious routes). One problem with attempting to include too much detail in a transform is that there are limits to what can easily be expressed in the ALCHEM language, as discussed below.

Several points have become clear during the two years which this work covered. First and foremost, experienced chemists must be involved in the development of the transform data base—not only to supply the "scope and limitation" information on reactions but also to recommend the types of reactions to be included. It is clear from the current analyses produced by SECS that many more transforms must be added before a reasonably thorough analysis of a structure can be produced. We have also learned that chemists are not generally willing to learn ALCHEM, even enough to understand why a transform is misbehaving; but they will tell us which transforms need further work, and why.

One of the key items which a transform should contain if it is to be truly useful is a list of up-to-date references. This raises the major question of how best to maintain and retrieve the references in a large and growing data base.

Our efforts to write new transforms have increased substantially in the past year with the addition to our group of a synthetic chemist, Dr. Dale Hoff, who can both summarize reactions and code transforms. The current Merck SECS data base contains over 500 transforms.

The ALCHEM language[8] has proved to be general enough and flexible enough to describe every reaction which we have considered, in English-like sentences which are reasonably intelligible to chemists. Nevertheless, the vocabulary of organic chemists is very rich, and some concepts are presently difficult to express in ALCHEM. For example, the chemist fairly easily perceives the effect of heterocyclic functionality remote from a reaction center, but ALCHEM statements to accomplish such perception are quite complex. Besides being less easily comprehended, complex ALCHEM statements generally require extensive testing to verify that they are interpreted by the program as intended. We expect that further development of the ALCHEM vocabulary will make the language even more powerful and flexible.

As the number of transforms in the data base increases, the number of precursors generated must be controlled. The new version of SECS 2.7, currently being implemented at Merck, utilizes strategic options to improve chemist control of an analysis.

CONCLUSION

If a convenient, routinely useful synthetic analysis program is developed, it will be accepted. Although many chemists resisted IR, NMR, and mass spectral techniques when they were introduced, no modern chemist would dream of performing structure determination exclusively by degradation experiments. Complex organic synthesis is difficult, and the chemist can use help. If computer-aided planning can reduce the number of unsuccessful experiments run, research managers will embrace the methodology. While the new version of SECS should overcome some of the current problems, a fully "finished", production system is not yet in hand.

Merck is participating in development of the SECS software in order to gain experience in this area, in order to aid development of a computer program which is useful for pharmaceutical research, and for what it can contribute to Merck's synthetic effort, even at this early stage.

Development of a synthesis may more or less be divided into the "synthetic planning" stage, where a rough synthetic approach is blocked out, and a "reaction retrieval" stage, where analogous sequences and specific reaction conditions are sought. An optimal "synthetic planning" tool is probably not best for the "reaction retrieval" function.[19] SECS is primarily a "synthetic planning" tool. There is a real need for "reaction retrieval" aids, but we suspect that this will have to be met primarily by other systems. Perhaps the new Derwent Reaction Retrieval Service will fill this need.

Merck's main emphasis at present is in building an improved reaction data base. Already it is clear to us that this must be done slowly and carefully, by expert chemists, with continual refinement of transforms based on experience. The total number of synthetic reactions is rather large, and it is not at all clear that one company can commit enough resources to cover much of synthetic chemistry in a reasonable time period. Some sort of exchange of transforms by participating companies, such as is currently underway in Europe,[20] would considerably speed system development.

ACKNOWLEDGMENT

We are grateful to M. J. Pensak for his contributions to an improved molecule input routine and the hard copy summary program; to the many Merck chemists who participated in evaluating and improving the program; and to several far-

sighted administrators who initiated this project and sustained it in its early stages. SECS program development at Santa Cruz has been supported by the National Institutes of Health, Research Resources Grant No. RR-01059, and through computer support from the Stanford University SUMEX-AIM Project Grant No. RR-00785.

REFERENCES AND NOTES

(1) E. J. Corey and W. T. Wipke, *Science*, **166**, 178 (1969).
(2) Review: M. Bersohn and A. Esacks, *Chem. Rev.*, **76**, 269 (1976).
(3) Review: P. Gund, *Annu. Rep. Med. Chem.*, **12**, 288 (1977).
(4) "Computer-Assisted Organic Synthesis", W. T. Wipke and W. J. Howe, Eds., ACS Symposium Series, No. 61, American Chemical Society, Washington, D.C., 1977.
(5) L. H. Sarett, unpublished speech before Synthetic Organic Chemical Manufacturers Association, June 1964; quoted in ref 6.
(6) W. T. Wipke, "Computer Representation and Manipulation of Chemical Information", W. T. Wipke, S. R. Heller, R. J. Feldmann, and E. Hyde, Eds., Wiley, New York, 1974, p 147.
(7) P. Gund, J. D. Andose, and J. B. Rhodes, in Contributed Papers to the 3rd International Conference on Computers in Chemical Research, Education and Technology, E. V. Ludeña and F. Brito, Eds., Centro de Estudios Avanzados del Instituto Venezolano de Investigaciones Cientificas (IVIC), Caracas, 1977, p 226.
(8) Review: W. T. Wipke, H. Braun, G. Smith, F. Choplin, and W. Sieber, in ref 4, p 97.
(9) "Utilization of Stereochemistry and Other Aspects of Computer-Assisted Synthetic Design", T. M. Dyott, Ph.D. Thesis, Princeton University, 1973 (University Microfilms, No. 74-9677).
(10) Our current computer environment consists of an IBM 370/168 computer running under the OS/VS2 MVS-3.7A operating system. We have also run SECS on TSO under the OS-MFT and OS-MVT operating systems. We are using DEC GT-42 and GT-43 graphics display terminals with 16 K words of core memory. Communication with our host computer is at 1200 baud using VADIC 3400 series modems.
(11) Different releases of the SECS program are identified by different version numbers, starting with 1.0. The current release of the program corresponds to version 2.7.
(12) Initially available through First Data Corp., Waltham, Mass., the program is now available on ADP Network Services, Inc., Ann Arbor, Mich.
(13) B. Dominy, "SECS and the Information Scientist", presented at the Science Information Subsection of the Pharmaceutical Manufacturer's Association Meeting, Washington, D.C., March 6, 1977.
(14) E. J. Corey, W. J. Howe, H. W. Orf, D. A. Pensak, and G. Petersson, *J. Am. Chem. Soc.*, **97**, 6116 (1975).
(15) R. F. Shuman, S. H. Pines, W. E. Shearin, R. F. Czaja, N. L. Abramson, and R. Tull, *J. Org. Chem.*, **42**, 1914 (1977).
(16) S. H. Pines, R. F. Czaja, and N. L. Abramson, *J. Org. Chem.*, **40**, 1920 (1975).
(17) J. Ten Broeke, A. W. Douglas, and E. J. J. Grabowski, *J. Org. Chem.*, **41**, 3159 (1976).
(18) H. Bruns, private communication.
(19) P. Gund, J. D. Andose, and J. B. Rhodes, in ref 4, p 179.
(20) H. Bruns, *Naturwissenschaften*, **66**, 197 (1979).

The Evaluation of an Automatically Indexed, Machine-Readable Chemical Reactions File

PETER WILLETT

Postgraduate School of Librarianship and Information Science, University of Sheffield, Western Bank, Sheffield, S10 2TN, United Kingdom

Received April 24, 1979

An automatic system for the analysis and retrieval of chemical reaction information is described which allows searches to be carried out for both reacting and nonreacting substructures in the reactant and product molecules of a chemical reaction. The techniques could be implemented easily in a conventional substructure search system. Applications of chemical reaction files to the use of computer-aided synthesis programs are discussed.

I. INTRODUCTION

Two previous papers in this series have described automatic methods for the characterization of chemical reactions using Wiswesser Line Notations (WLN)[1] and connection tables[2] as the machine-readable structure representations; a comparison of the techniques[3] and an evaluation of a printed index of reactions produced by the WLN approach[4] have also been presented. In this paper, we describe an experimental system for the retrieval of reaction information based, primarily, upon the reaction sites identified by our structure-matching algorithm.[2] The retrieval system described here uses the analyses produced by the structure-matching procedure to allow searches to be made for both reacting and nonreacting substructures in the reactants and products; the WLN analyses allow easy access only to the former type of feature.[3]

II. CREATION OF THE SEARCH FILE

The source file for this work was the 7415 reactions successfully analyzed by the WLN reaction indexing program described earlier.[1] For each such reaction the WLNs of the reactant and product molecules were converted to CROSSBOW connection tables using software provided by ICI Ltd. (Pharmaceuticals Division), and then the tables were written out to tape together with the fragment strings resulting from the WLN analysis, the original WLNs, and the bibliographical reference. In this way, a file of 5226 one-reactant, one-product reactions was obtained for analysis by the structure-matching procedure; it should be noted that the procedure is extensible to more complex transformations by merging the sets of reactant or product structure representations so as to represent a single, discontinuous graph.

Many types of structure search system have been described in the literature.[5,6] We have used a simple sequential organization in which each of the items in the file is characterized by a fragment bitscreen; corresponding query bitstrings are held in core and matched against each of the reactions in the search file in turn. Boolean AND, OR, and NOT logic is available, together with a string search facility for the fragments resulting from the WLN analysis.

The multifarious nature of chemical reaction information and the need to differentiate between reacting and nonreacting substructural features require a variety of modes of access to the data, and we now outline the screening system which has been used to characterize each of the reactions in the file.

An analysis by Clews[7] showed that a file of reaction site residues contained a higher percentage of heteroatoms than the corresponding file of reacting compounds, and thus different sets of screens are required for assignment to the two types of structural feature. We have used atom, bond, ring, and molecular formula screens and as these are to be assigned to both reactant and product reaction sites and parent molecules, a total of 16 different types of screen are used for each reaction; in view of the small number of reactions in the search file, this variety of screen types is probably excessive but the retrieval results described below indicate that the screening

0095-2338/80/1620-0093$01.00/0