Metanomski, W. V. "Is This a Record?". *Chem. Int.* **1987**, *9*(6), 215-216.
Metanomski, W. V. "Morphodelotic (Shape-Descriptive) Names of Chemical Substances". Poster presented at the Chemistry Trivia Session, 194th National Meeting of the American Chemical Society, New Orleans, LA, Aug 31, 1987.
Metanomski, W. V. Review of the book *Literature of Analytical Chemistry: A Scientometric Evaluation*; Braun, T., Bujdoso, E., Schubert, A., Ed.; CRC: Boca Raton, FL, 1987. Submitted Nov 24, 1987, to *Trends Anal. Chem.*
Metanomski, W. V. "Report on the Conference Chemical Structures: The International Language of Chemistry (Noordwijkerhout, The Netherlands, May 31-June 4, 1987)". Submitted Oct 15, 1987, to the *J. Sci. Ind. Res.*
Metanomski, W. V. "Unusual Names Assigned to Chemical Substances". *Chem. Int.* **1987**, *9*, 211-215.
Metanomski, W. V. "What's in the Number and the Name Found in the CAS Database?". Poster presented at the Chemistry Trivia Session, 194th National Meeting of the American Chemical Society, New Orleans, LA, Aug 31, 1987.
Powell, W. H. "Why Continue Systematization of Chemical Nomenclature?". Presented at the Symposium on Chemical Nomenclature, Laboratory of the Government Chemist, London, Nov 12, 1987.
Ramsay, O. B.; Metanomski, W. V. "Organic Functional Group Ubiquity and Importance". Poster presented at the Chemistry Trivia Session, 194th National Meeting of the American Chemical Society, New Orleans, LA, Aug 31, 1987.
Sanderson, J. H.; Mamrack, S.; Brogan, M. "Interface for Electronic Manuscript Input to the ACS Composition System". Presented before the Division of Chemical Information, 194th National Meeting of the American Chemical Society, New Orleans, LA, Sept 2, 1987.
Sloan, T. E. "Nomenclature of Coordination Compounds". In *Comprehensive Coordination Chemistry*; Wilkinson, G., Gillard, R. D., McCleverty, J. A., Eds.; Pergamon: London, 1987; Vol. 1, part 1, Chapter 3, pp 109-134.
Stocker, J. H. "Chemage, a Compendium of Chemical Trivia". Presented at the meeting of the Chicago Section ACS joint with Chicago Chemists Club, Mundelein College, Chicago, IL, Dec 11, 1987.
Stocker, J. H. "Some Fun with Chemistry Trivia". Presented at the luncheon of the Division of Chemical Information, 194th National Meeting of the American Chemical Society, New Orleans, LA, Sept 1, 1987.
Stocker, J. H. "The Chemist's Poetry Corner". Poster presented at the Chemistry Trivia Session, 194th National Meeting of the American Chemical Society, New Orleans, LA, Aug 31, 1987.
Stocker, J. H. "Musichemistry—Treble the Fun!" Poster demonstration presented at the Chemistry Trivia Session, 194th National Meeting of the American Chemical Society, New Orleans, LA, Aug 31, 1987.
Stocker, J. H. "The Living Language of Chemistry". Presented at a seminar of the Department of Chemistry, University of New Orleans, March 23, 1987.
Zaye, D. F. Review of the book *How to Find Chemical Information*, 2nd ed.; Maizell, R. E., Ed.; Wiley-Interscience, New York, 1987. Submitted Nov 6, 1987, to *J. Chem. Inf. Comput. Sci.*

REACCS in the Chemical Development Environment. 1

JOHN E. MILLS,* CYNTHIA A. MARYANOFF, KIRK L. SORGI, LORRAINE SCOTT, and ROBIN STANZIONE

McNeil Pharmaceutical, Spring House, Pennsylvania 19477

Received September 11, 1987

A number of database management systems suitable for the storage of chemical reaction information are currently available. Criteria used at McNeil Pharmaceutical for the selection of a graphics-based database management system for the storage of proprietary information are presented. Applications for the software are discussed.

INTRODUCTION

Information retrieval is a major problem facing the chemical industry. During the time from when a reaction is first attempted to when it is scaled-up for production, many chemists in various departments may have worked on the project. To prevent duplication of expensive experimental work, an efficient method for the storage and retrieval of data is critical.

For example, at McNeil Pharmaceutical, process work on Tolectin (tolmetin sodium) over the past 20 years has produced many volumes of information in the form of Chemical Research Reports, monthly reports, and notebook experimentals. It became obvious during work on this and other projects that data storage and retrieval for formal written reports was satisfactory. However, these reports normally contained only positive results. They frequently excluded experiments that did not give the desired product, or they only briefly summarized such experiments. In addition, these reports did not typically include all reaction variations representing changes in solvents, catalysts, reaction times, and/or reaction temperatures. Consequently, much of the information contained in research notebooks was not readily accessible. In fact, one could access most of this information only with great difficulty. Rather than spending the time necessary to retrieve the original data, chemists repeated experiments. Obviously, some method was necessary to allow for the storage and rapid retrieval of all chemical reaction data generated by Chemical Development and Chemical Research.

Representatives from the Chemical Development and Medicinal Chemistry departments met on several occasions to clearly define the needs for chemical database software at McNeil and to determine how those needs could be met.[1] The various storage systems that they considered were System

0095-2338/88/1628-0153$01.50/0

1022,[2] the Molecular Access System (MACCS),[3] a combination of System 1022 and MACCS, SYNLIB,[4] and REACCS.[3] While System 1022 is easily learned, a text-only database for the storage and retrieval of reaction information would be inefficient. Searching by names and keywords would be too broad in some cases and too narrow in others. The problem of information retrieval would be compounded by the fact that the same keyword may be interpreted differently by different individuals. Structural diagrams have precisely the same meaning for all chemists. The representatives therefore concluded that the graphic storage of molecular structures was essential.

Although certain companies have used MACCS for storage of chemical transformations (e.g., for the storage of biosynthesis–metabolic pathways),[5] this program was not designed for the efficient recovery of reaction information. Datatypes can be designed to link reactants and products, but this method of recovering reaction data contains a number of limitations and may be labor intensive.

Combining MACCS and System 1022 would give both structural search capabilities and efficient storage of reaction text. However, use of this combination would require that functional group transformations, as well as reaction data, be stored in text form in the System 1022 portion of the database. Thus, to enter data, a chemist would have to have a thorough knowledge of the database in order to maintain consistency. Likewise, to prepare efficient search strategies for searches over the database, the user would need a thorough knowledge of the database structure and keywords. In addition, learning the interface between the two programs could be difficult and would probably result in fewer users. One possible approach to the problem of the simultaneous use of two databases would be to commit programming resources to develop a "user friendly" interface in-house. Some companies have used this approach.[6] MACCS-II, in conjunction with its Customization Modules, can simplify interfacing two or more database systems.[3]

With its graphics storage, flexible search modes, multiple levels of security, and ease of use, REACCS appeared to be ideally suited to meet McNeil Pharmaceutical's needs for the storage and retrieval of reaction information. SYNLIB, which also uses graphics for storage and retrieval of reaction data, did not have the flexibility provided by REACCS, nor was it commercially available at the time.

In addition to choosing REACCS for the storage of proprietary data, the group decided to purchase the Theilheimer database from Molecular Design Limited. This database contains a large selection of known chemistry that could provide a starting point for literature searches performed by both Medicinal and Developmental chemists. The Theilheimer database has also since served to stimulate chemists to use REACCS. The information contained in a small but developing in-house REACCS database tended to be too specialized to assist with the large variety of chemistry different groups were exploring; consequently, many chemists felt that the time necessary to learn the program did not warrant the information obtained through the search of REACCS databases.

SECURITY

In addition to the normal system security, REACCS provides for a number of levels of access to any database. Access to a given datatype contained in different databases does not need to be the same for a given user. In fact, one user may have several different levels of access to ensure that information is not accidentally added to or erased from the database. These levels include the ability to read the database, to read and write into the database, or to read, write, and delete information from a database. In addition, individual datatypes may be given different levels of security so that all datatypes are not available to all users. The flexibility and security provided by REACCS was not matched by any of the other chemical database management systems that were available when the group decided to purchase this system.

DATABASE NEEDS

Representatives from the Medicinal Chemistry and Chemical Development departments established that no single database could fit everyone's needs. The following databases were proposed: (1) Chemical Development Proprietary Database—to contain detailed experimentals for work relevant to compounds studied in the Chemical Development department; (2) Literature Database—to contain literature reactions and references of particular interest to chemists at McNeil and to serve as an electronic literature index for any chemist desiring this service; (3) Notebook Database—to contain a minimal amount of information, but useful as an electronic index; (4) Drug Metabolite Database—to contain information on drug metabolites and their distribution in different species. The selection of datatypes for these proprietary databases is the subject of the following paper.[7]

DISCUSSION

The utility of any database depends not only on its content but also on the ease of use and flexibility of the database management system. The fundamentals needed to search a REACCS database can be learned by inexperienced chemists in 1–4 h. REACCS provides a large variety of search capabilities, allowing a novice user to learn efficient search strategies with minimum formal training. Experience provides proficiency in designing search strategies; it is not uncommon for a user to master difficult searches in 10–20 sessions on the program.

Once a database is created, it can be altered; however, the designers should carefully consider the purpose of each database before attempting to create them. Datatypes, needed to achieve the purpose of the database, should be clearly defined. The designers must establish any relationships between datatypes before creating a customized database. A standard database form is provided with REACCS. However, it is not necessarily the most efficient form in which to store data for all needs. As with all databases, the amount of information stored must be balanced against the cost of data entry and the need for all the proposed datatypes. The rate of database growth is highly dependent upon the number of datatypes and the experience of the data entry personnel. The display flexibility provided by REACCS is extremely useful when multiple databases are being used and when the needs of users differ. By selecting different display formats, different users may appear to be using different databases when, in reality, the users have only used an option to display different datatypes.

While McNeil has purchased a commercial literature database, the company recognizes that commercial databases may not contain a complete sampling of the literature in respect to the chemistry being investigated in-house. Creation of an in-house literature database has provided the opportunity for every chemist to submit abstracts of articles pertinent to their own interests. In addition, a special datatype is provided so each chemist may file hard copies of their articles using their own filing system; they may subsequently retrieve the articles through a REACCS search.

As McNeil chemists become familiar with REACCS and the databases it supports, it is becoming obvious that the benefits derived from its use extend beyond those envisioned when the company purchased the software and databases. Since the decision was made to register a complete set of

proprietary data from a given date forward, McNeil's management realized that initially this system would be a drain on resources. With the addition of the Theilheimer and Current Literature File databases, the system has already shown positive results through decreased library search times. In addition, the in-house literature file is proving helpful by showing all users those areas of interest to other chemists at McNeil.

Two other areas, not originally envisioned, in which chemists are applying REACCS are retrieval of articles pertinent to a given area of chemistry and "trouble shooting". Previously, once developmental chemists at McNeil found a leading article, they would request author searches over the CAS databases and cited-article searches over the ISI databases to find as much pertinent information on a reaction as possible. The CLF database on REACCS is becoming one of the primary sources of preliminary reaction citations. Chemists use REACCS to trouble shoot by searching the databases for reactions in which substrates with the same functional groups as those being studied react with the same catalysts or reagents. Occasionally the chemistry obtained in a search of this type can be used to help explain the observed chemistry.

At McNeil, REACCS is rapidly becoming an indispensable tool for our chemists. Although data entry into our proprietary databases began only a few years ago, chemists are already saving time. Often searches reveal chemical reactions that had been optimized earlier and can now shed light on current projects; with REACCS, chemists find this information even though these earlier projects may produce molecules with little obvious similarity to those molecules being prepared in the current projects. As these databases continue to grow, their utility is expected to expand also.

ACKNOWLEDGMENT

We acknowledge all chemists at McNeil Pharmaceutical who helped in the development of our current databases. We are especially grateful to Julie Spink, who did much of the original abstracting for our proprietary database, to William Bullock, who helped in the database creation and data entry, and to Barbara Baughman, for her continuing efforts to keep this database current. Our thanks also to Cynthia Dunn, Leonard Herring, and other members of the Scientific Computer Services Department for their help in bringing this work to fruition.

REFERENCES AND NOTES

(1) The filing system was to be implemented on a DEC-10 computer system.
(2) System 1022 is available from Software House, Cambridge, MA.
(3) MACCS-II and REACCS are available from Molecular Design Ltd., San Leandro, CA.
(4) SYNLIB has been marketed by SmithKline/Beckman, Philadelphia, PA.
(5) Barcza, S.; Kelley, L. A.; Lenz, C. D. "Biosynthesis-Metabolic Pathways Database". Presented at the 192nd National Meeting of the American Chemical Society, Anaheim, CA, Sept 1986.
(6) For examples of software written to interface MACCS with other databases see: (a) Legatt, T.; Saltzman, A. "Development of a Chemical Structure Display System and Its Interface with the Associated Biological Database". *Drug Inf. J.* **1986**, *20*, 51. (b) Adamson, G. W.; Bird, J. M.; Palmer, G.; Warr, W. A. "Use of MACCS within ICI". *J. Chem. Inf. Comput. Sci.* **1985**, *25*, 90.
(7) Mills, J. E.; Maryanoff, C. A.; Sorgi, K.; Stanzione, R.; Scott, L.; Herring, L.; Spink, J.; Baughman, B.; Bullock, W. *J. Chem. Inf. Comput. Sci.* (following paper in this issue).

REACCS in the Chemical Development Environment. 2. Structure and Construction of Proprietary Databases†

JOHN E. MILLS,* CYNTHIA A. MARYANOFF, KIRK L. SORGI, ROBIN STANZIONE, LORRAINE SCOTT, LEONARD HERRING, JULIE SPINK,‡ BARBARA BAUGHMAN, and WILLIAM BULLOCK§

McNeil Pharmaceutical, Spring House, Pennsylvania 19477

Received September 11, 1987

The utility of a chemical database management system is dependent both upon the ease of use of the software and upon the information stored in the database. The process used at McNeil Pharmaceutical to select datatypes for proprietary REACCS databases is discussed. Problems associated with the storage of negative as well as positive results, storage of several structures for a given molecule, and storage of incompletely characterized molecules using REACCS are discussed. Useful techniques to circumvent some potential problems are provided.

INTRODUCTION

The storage and retrieval of chemical reaction information are areas in which dramatic advances have been achieved over the past decade. Although works such as Houben-Weyl's *Methoden der Organische Chemie*, Theilheimer's *Synthetic Methods*, and Beilstein's *Organische Chemie* have been used with limited success for the storage of reaction data, all of these methods suffer from their lack of generality; each is designed for recovery of reaction data through a single specific search strategy. Through its subject index, *Chemical Abstracts* provides somewhat limited search capabilities for reactions and for information on the preparation of a molecule whose structure is known. Advances in computer technology have made the formulation of search queries via graphics input commonplace.[1-5] These advances have simplified the process to such an extent that it is not uncommon for "bench chemists" to learn how to formulate queries using software such as REACCS after a training period of 1-4 h.

IDENTIFICATION OF DATABASES

The need for an efficient means to store and retrieve chemical reaction data has been presented elsewhere.[4,6] This

† Presented in part at the MDL Software User's Group Meeting, San Francisco, CA, April 2, 1987, and at Reaction Indexing Seminar, Saddle Brook, NJ, June 1986.

‡ Present address: Ciba Corning Diagnostics, Medfield, MA 02052.
§ Present address: Department of Chemistry, Emory University, Atlanta, GA 30322.

0095-2338/88/1628-0155$01.50/0