Current Comments®

EUGENE GARFIELD
INSTITUTE FOR SCIENTIFIC INFORMATION®
3501 MARKET ST., PHILADELPHIA, PA 19104

International Collaboration Helps Establish ISI Reactions Club: *Current Chemical Reactions In-House Database* Is Catalyst to Research Creativity

Number 13 March 30, 1987

Forty years ago as a premedical student I encountered organic chemistry for the first time. Like the thousand other students sitting in that large Berkeley auditorium, I had learned simple inorganic reactions in the freshman chemistry course given by Professor Joel Hildebrand. In that same lecture hall I later learned about the wonders of carbon-containing compounds from Professor Melvin Calvin, who would go on to win the 1961 Nobel Prize in chemistry for his work on photosynthesis.

I recall from those days the endless hours spent memorizing dozens of organic chemical reactions. At the time I wondered why there wasn't an easier way to learn this new subject, which I later came to realize was much like learning a new language.

Three years later I asked this same question regarding the actual preparation of chemical compounds when I worked for Professor Louis P. Hammett at Columbia University. Eventually I came to realize that not only was there a better way but that a lot of others felt as I did. Information-retrieval systems provide a solution to the limitations of human memory.

About 10 years later in 1960, ISI® launched the *Index Chemicus®* (*IC®*) to facilitate locating chemical compounds in the literature. It was only a matter of time before we would tackle the more complex problem of the reactions that lead to these compounds. In 1979 we began to index these reactions on a systematic basis when *Current Chemical Reactions®* (*CCR®*) was born.

Recently we discussed the *Index Chemicus* and *Current Chemical Reactions Personal Databases* available for personal-computer use.[1] For more comprehensive coverage, we have developed an extensive chemical reaction file for use on a mainframe computer. *Current Chemical Reactions In-House Database* is a graphic and textual description of over 5,000 new synthetic methods. Each record is accompanied by detailed bibliographic and experimental data. The advantages of an in-house system, where the software and databases are used on a company's own mainframe computer, are discussed later.

The student of elementary organic chemistry may of course find it mind-boggling to imagine "memorizing" so many reactions, but that is exactly what our new database does for you. We expect this file to grow to at least 50,000 new synthetic methods over the next decade to facilitate the identification of relevant means of synthesizing new or known organic chemical entities.

In writing about databases we tend to lose sight of the creativity involved in chemical synthesis. One might get the impression that chemists are simply chefs executing cookbook recipes found in a large electronic file. But while artificial-intelligence programs have recently been developed in such systems as DENDRAL,[2] LHASA,[3] and SECS,[4] we are indeed far from replacing the human element in the creation of appropriate synthetic methods for new compounds.

After W. Theilheimer started the publication of his annual surveys of synthetic methods in the mid-1940s, the field of chemical

information retrieval attracted a large number of creative people including R. Fugmann, M.F. Lynch, W. Steidle, G. Vladutz, W.T. Wipke, and H.J. Ziegler, to name a few. Another pioneer in this field is Jacques Valls, formerly with Rousell-Uclaf, Paris, but now at the Asian Institute of Technology, Bangkok, Thailand. In what must now seem like an obvious truism, Valls wrote 13 years ago that "I am convinced chemists badly need a reaction documentation [system] and when such a tool is put at their disposal they [will] use it intensively.... Retrieving information on reactions is of major—I would say of vital—importance and at least as necessary as retrieving information on chemical compounds."[5]

Reaction Data Club

Many major chemical and pharmaceutical companies worldwide have joined together to support the *CCR In-House Database*. The members of this Reaction Data Club (RDC) are listed in Table 1. They are a prestigious group of firms. Individuals from each firm will provide editorial direction to the expansion and improvement of this database.

The international cooperative effort by these major companies is not new, but this trend has certainly accelerated. The complexities associated with the numerous chemical products developed yearly prohibit individual companies from supplying all their own information needs. Therefore, for the most complete and cost-effective way for companies to keep abreast of the chemical products developed worldwide, they and their competitors rely on neutral third parties like ISI to satisfy their needs.

It is for this purpose that ISI took the initiative to launch the *Index Chemicus*. Following the same approach, we contacted dozens of companies last year to fund and guide this *CCR* database. Meetings were held in the US and Rome to detail our plans for representatives from the 22 RDC member organizations. We expect membership in the RDC will expand in the future, and this will help control costs as well as enhance the database.

Current Chemical Reactions

Our new in-house database is based on our monthly publication *Current Chemical Reactions*. *CCR* provides a guide to new and newly modified reactions and syntheses reported in the current journal literature. Each entry contains a flowchart of the reaction as well as complete bibliographic information and experimental notations when supplied in the source article.

The synthetic methods reported by the *CCR* are drawn from over 125 "core" organic chemistry and pharmaceutical journals from around the world. Four hundred new synthetic methods are selected each month from this literature based on a strict set of criteria that screen for reactions or synthetic procedures that have not been published earlier; known reactions or synthetic procedures that have been extended for use with different reacting groups; known procedures that have been improved or modified; or the first laboratory or total synthesis of a natural product.

The In-House Advantage

You may wonder why these files were not put up on a vendor system such as Dialog or Datastar as is *SciSearch®*, the online version of the *Science Citation Index®*. We investigated possible vendors, but at the time no vendor had the necessary programs to handle graphics-based reaction files in an online environment. However, the appropriate software for in-house systems was available. Making *CCR* accessible for in-house mainframe use provides distinct advantages. For example, access to the software and databases is direct—it is simpler and faster than dealing with a third-party vendor. In addition, it is possible to combine data created in-house with ISI data so that a single combined search of both can proceed simultaneously. It is also possible to create specialized databases.

It can be significantly less expensive for a company with many employees requiring access to chemical data simply to buy the software and databases for installation in-

Table 1: Members of the Reaction Data Club.

EUROPE
Astra Laekemedal Aktiebolaget
Strangnaesvagan, Sweden

BASF Aktiengesellschaft
Ludwigshafen Rhein
Federal Republic of Germany

Ciba-Geigy Limited
Basel, Switzerland

Hoechst Aktiengesellschaft
Frankfurt am Main
Federal Republic of Germany

F. Hoffmann-La Roche & Co.
Basel, Switzerland

Eli Lilly
Windlesham, Surrey
United Kingdom

E. Merck
Darmstadt
Federal Republic of Germany

Pfizer Limited
Sandwich, Kent
United Kingdom

Rhone-Poulenc
Saint-Fons, France

Schering AG
Berlin, Federal Republic of Germany

UNITED STATES
American Cyanamid Company
Princeton, NJ

American Cyanamid Company
Stamford, CT

American Cyanamid Company
Pearl River, NY

Ciba-Geigy Corporation
Summit, NJ

E.I. du Pont de Nemours & Company, Incorporated
Wilmington, DE

Hoffmann-La Roche, Incorporated
Nutley, NJ

Eastman Kodak Company
Rochester, NY

Eli Lilly & Company
Indianapolis, IN

Monsanto Company
St. Louis, MO

Polaroid Corporation
Cambridge, MA

The Upjohn Company
Kalamazoo, MI

3M (Minnesota Mining & Manufacturing Co.)
St. Paul, MN

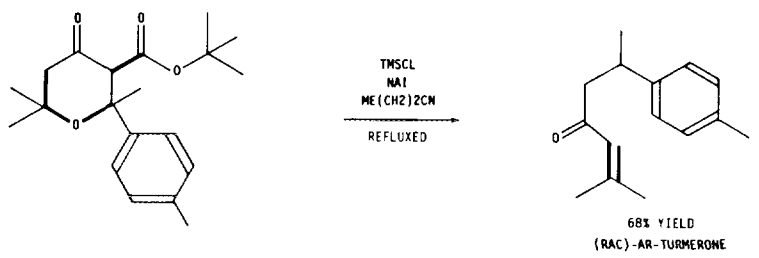
house. Paying a third-party vendor for access to the same information online is expensive if the file is to be used frequently by many users. And there is always a concern about confidentiality. The questions one asks on a public database ought to be protected by privacy laws, but recently certain government agencies have attempted to query online vendors about their customers. The basic position of the information industry, with which I concur, is that this would be a fundamental invasion of our rights under the Constitution. Certain European countries have already enacted strong privacy laws. But apart from all this, those who are conscious of proprietary issues may prefer in-house facilities.

ISI reviewed most of the candidate software packages available for use with the *CCR In-House Database*. We chose the *REaction ACCess System* (*REACCS*) of Molecular Design Limited because we believe it is the system that best meets our needs and those of our clients. Molecular Design Limited is a highly reputable software company that enjoys about a 90 percent share of the chemical software market in the US and abroad. In addition, REACCS can handle and display the types of reaction data in *CCR* and is used by more than 65 companies around the world.

Features

Beginning with data from 1986, the *CCR In-House Database* covers 5,000 new synthetic methods a year. However, since each synthetic method has an average of five in-

Figure 1: Sample reaction record from the *CCR® In-House Database*.



SAKAI T, MIYATA K, ISHIKAWA M, TAKEDA A.
OKAYAMA UNIV, SCH ENGN, DEPT SYN CHEM, OKAYAMA 700, JAPAN.
TETRAHEDRON LETT 26(39),4727-30(1985).

A NOVEL SYNTHETIC METHOD OF DIHYDRO-4-PYRONE DERIVATIVES AND ITS APPLICATION TO THE SYNTHESIS OF AR-ATLANTONE AND (+/-)-AR-TURMERONE.

KEY RXN	RING OPENING

REGNO:	EXTREG:	PATH:	STEP:
4	30716	B1	3 OF 3

dividual steps, we provide a separate record for each individual step of the new synthetic method as well as a step describing the overall reaction. As a result, the database provides access to over 30,000 reaction steps each year.

Viewing an individual step out of context from the entire synthetic method can be confusing. To ensure that you understand the role of each step in the overall synthesis, the system provides a series of indicators. For example, the overall synthetic method may include the synthesis of more than one product and result in various product pathways. The route for each individual step is shown on screen by a pathway indicator. The text also provides a step indicator that tells you the step number of the individual reaction as well as the total number of steps in the overall synthesis.

Often a multistep method has only one step that is considered the new or critical part of the procedure. The text provides a key-reaction indicator to highlight this pivotal step.

In the growing information industry there is a great deal of discussion about *value-added* information. The *CCR In-House Database* provides much more information than that provided by the author in the source article. ISI's staff of professional chemists enhance the appropriate data by highlighting both the reacting sites of the compounds in a reaction as well as any stereochemistry involved if it is not clearly indicated by the authors. On the computer screen these areas are visibly enhanced by color. If you don't have a color graphics monitor, the information is highlighted with bold-faced markers. Figure 1 shows a sample reaction record with the three types of indicators and the reacting sites highlighted. At a glance you can see that this sample reaction is the third of three steps for pathway B.

Access Points

Each new synthetic method can be accessed by chemical structure, and reacting

sites and stereochemistry can be specified. In addition, each record has a large amount of supporting information that can be accessed directly. This textual material includes bibliographic source data, reagents, catalysts, and solvents necessary for the reaction. If mentioned in the source article, the conditions, yields, reaction descriptors, explosive reaction alerts, advantages, and limitations are also included.

Conclusion

Since its introduction in 1979, *Current Chemical Reactions* has been made available in print, on personal databases for use on a personal computer, and now for mainframe use with the *CCR In-House Database*. Suitable for a variety of industries in which synthetic problems are encountered, this database is a cost-effective aid for developing new chemicals and drugs while ensuring that time-consuming duplication is avoided. In the computer world one hears a great deal about computer-aided design (CAD). There is a parallel to chemistry and scientific information. *Index Chemicus* used to be called *Current Abstracts of Chemistry & Index Chemicus* (*CAC/IC*). *CAC* could, however, convey a neat, alternative meaning—computer-aided chemistry, or indeed, computer-aided creativity.

* * * * *

My thanks to Lisa Holland and Daniel Meyer for their help in the preparation of this essay.

REFERENCES

1. **Garfield E.** New compounds and reactions are as close as your PC with the *Index Chemicus* and *Current Chemical Reactions Personal Databases*. *Current Contents* (9):3-6, 2 March 1987.
2. **Buchanan B G, Feigenbaum E A & Lederberg J.** *On generality and problem solving: a case study using the DENDRAL program.* Stanford, CA: Stanford University, August 1970. 50 p. NTIS/AD-715 128.
3. **Corey E J, Hewett A P W, Johnson A P, Long A K, Mulzer J & Orf H W.** Computer-assisted synthetic analysis. *J. Chem. Inform. Comput. Sci.* 20:221-30, 1980.
4. **Gund P, Grabowski E J J, Hoff D R, Smith G M, Andose J D, Rhodes J B & Wipke W T.** Computer-assisted synthetic analysis at Merck. *J. Chem. Inform. Comput. Sci.* 20:88-93, 1980.
5. **Valls J.** Reaction documentation. (Wipke W T, Heller S R, Feldmann R J & Hyde E, eds.) *Computer representation and manipulation of chemical information*. New York: Wiley, 1974. p. 83-103.