""""""Current comments

ISI®'s CHEMTRAN™ "Compatibilizes" Files of Encoded Chemical Structures

November 15, 1972

"Compatibilize" and CHEMTRAN are new words. Surprisingly, "debabelize" is not. Unfortunately it not only disparages the useful diversity of chemical "languages" and codes, but also suggests that CHEMTRAN is another chemical language. It isn't. CHEMTRAN aspires to become a universal interpreter of and between various existing systems—hence, it compatibilizes.

The origins of CHEMTRAN go back to my original interest in this problem. 1-2 Recently, this culminated in the successful completion of computer programs which convert Wiswesser Line Notation into so-called "Dokumentationsring" fragment codes by means of computer-generated connectivity tables. 4 These connectivity tables provide the basic intermediate records required not only for the WLN-to-Ring conversion, but also for conversion to any other fragment code, as e.g., the DuPont code used for indexing chemical patents.

CHEMTRAN can also be used to generate other connectivity tables, such as those used by Chemical Abstracts. CHEMTRAN might be compared to a common computer language such as Cobol. Through "compilers" Cobol compatibilizes various computers. Unlike Cobol, interestingly enough, CHEMTRAN does not increase the computer running time required to solve searching problems.

On the contrary, the generation of fragment codes or "screens" simplifies searching enormously, since the individual fragment code, like a standard computer instruction, pre-programs otherwise complex programs.

In a previous series, 6-9 I described why chemists, pharmacologists and others need to search by substructures, fragments, moieties, etc. Space does not permit me to repeat those basic notions here. The diversity of problems requiring such a capability accounts for the incredible variety of systems in the literature. The irresistible intellectual challenge of designing a new chemical code must also be a factor, since in so many cases existing methods would have been adequate to the problem area involved.

Some special codes concentrate on specific structural features, others can be applied only to one type of compound, as e.g., steroids, peptides, etc. One such code included the notion of "electron sinks." Unfortunately, finding encoders capable of identifying these sinks was quite prohibitive.

Millions of compounds have been encoded in various systems. Chemical created for about 2 million compounds. The U.S. Patent Office has large files that use the Hayward notation. The Internationale Dokumenta-

tionsgesellschaft für Chemie uses topological methods as well as a fragment code known as GREMAS. The previously mentioned Dokumentationsring has a huge file of literature encoded in its "Ring" code. ISI's magnetic tape files contain almost one million compounds in the Wiswesser Line Notation. Most drug and chemical companies have large internal files, many in WLN, and numerous files exist at various 2. _____. An algorithm for transacademic and governmental organizations.

Most of these files are publicly available. The very availability of these large manipulable data banks makes the problem of compatibility critical. Acquiring necessary expertise in two or three systems can be enervating and expensive. Further, it is frequently desirable to maintain uniformity in formulating search strategies.

The obvious answer is the ability to convert one file into the language and format of any other, which is what CHEMTRAN is designed to do. Many organizations have studied the interconversion problem, but none has produced a universally acceptable intermediary language. Since every system seems to have some decided advantages, the only practical solution is to provide interconversion so that the user can chose the system that best satisfies his needs.

CHEMTRAN's capabilities were recently described to the representatives of over 25 European and American drug companies in meetings held in Paris and Washington. The outcome of these meetings will be a major modification in our Index Chemicus Registry System tapes, and the availability of 9. The retrieval and disthem to both industrial and academic groups. In addition, licenses for CHEMTRAN will be made available. If you are interested in any of these developments, please let us know.

- 1. Garfield, E. An algorithm for translating chemical names to molecular formulas. (Philadelphia: Institute for Scientific Information, 1961. Originally presented to the Faculty of the Graduate School of Arts and Sciences of the University of Pennsylvania in partial fulfillment of the requirements for the degree of Doctor of Philosophy.)
 - lating chemical names to molecular formulas. J. Chem. Doc. 2:177-179, 1962.
- 3. Steidle, W. Possibilities of mechanical documentation in organic chemistry. Pharm. Ind. 19:88-93, 1957.
- 4. Granito, C.E. CHEMTRAN and the interconversion of chemical substructure systems. Paper presented at the 164th Annual Meeting of the American Chemical Society, Division of Chemical Literature, New York City, August 1972.
- 5. Rasmussen, L.E. & Van Oot, J.G. Operation of DuPont's Central Patent Index. I. Chem. Doc. 9(4): 201-206, 1969.
- 6. Garfield, E. The retrieval and dissemination of chemical information. I. Current Contents No. 28, July 15, 1970, p. 4-5.
- 7. ----- The retrieval and dissemination of chemical information. II. The Wiswesser Line Notation. Current Contents No. 29, July 22, 1970, p. M1-2.
- 8. ---- The retrieval and dissemination of chemical information. III. ICRS. Current Contents No. 30, July 29, 1970, p. M1-2.
- semination of chemical information. IV. ICRS ® Radiical ™ software. Current Contents No. 31, August 5, 1970, p. M1-2.