DISCUSSION

SMILES is believed to represent the best compromise to date between the human and machine aspects of chemical notation. It is easily understood by the chemist because it has a minimum number of simple rules. Computationally, SMILES is interpreted in a very fast, compact manner, thereby satisfying the machine objectives of time and space savings. It is based on a computer approach to language parsing so that the machine part follows algorithms consistent with rigorous hierarchical nomenclature rules. This results in a great improvement in the efficiency of information processing as compared to conventional methods.

In computer terms, SMILES notation represents a tree that can be interpreted in a single pass. The increase in efficiency derives from the language syntax. For example, the bonding arrangements are implied by the position of atoms without having to be defined specifically. Thus, where previously 1000–2000 characters may have been needed to store a connection table, depending on the method used, to describe a compound such as morphine (Figure 1), SMILES stores the same information in 40 characters. It can also be shown that the necessary computer processing time is reduced 100-fold over conventional procedures using connection table format, e.g., those used in CAS or MOL connection tables.

Originally, SMILES was developed to provide a human/machine language interface. Beyond this objective, it has been valuable for the implementation of a wide variety of machine-oriented chemical information functions. Successful applications include data storage, structural display, modeling new structures, and substructure searches and recognition. An example is the use of SMILES in computation of partition coefficients and molecular refractivity in model compounds. These two properties are widely used in biochemical research. The logarithm of the partition coefficient is the hydrophobic parameter in Hammett methodology and has been applied to quantitative structural activity (QSAR).[14] On the basis of structural considerations, fragments of a modeled compound, designated and processed in SMILES terminology, serve to estimate log P accurately.

This paper is intended to be an introduction to SMILES methodology and to cover the fundamental rules needed to enter a structure into the SMILES system. Subsequent publications will present the method of obtaining "unique" SMILES, generating a structural depiction for any SMILES, generating a SMILES-oriented database that retrieves information at a speed independent of how many structures are stored, and SMILES methods for fast and powerful substructure searching.

ACKNOWLEDGMENT

Research on SMILES was initiated by the author at the Environmental Research Laboratory, U.S.E.P.A., Duluth, MN, and was completed at Pomona College. The author thanks Arthur Weininger for assistance in programming the SMILES system and Dr. Joseph L. Weininger for editorial assistance.

REFERENCES AND NOTES

(1) Wiswesser, W. J. *J. Chem. Inf. Comput. Sci.* **1985**, *25*, 258–263.
(2) Morgan, H. L. *J. Chem. Doc.* **1965**, *5*, 107–113.
(3) Dittmar, P. G.; Farmer, N. A.; Fisanick, W.; Haines, R. C.; Mockus, J. *J. Chem. Inf. Comput. Sci.* **1983**, *23*, 93–102.
(4) Balaban, A. T. *J. Chem. Inf. Comput. Sci.* **1985**, *25*, 334–343.
(5) Wipke, W. T.; Rogers D. *J. Chem. Inf. Comput. Sci.* **1984**, *24*, 255–262.
(6) Attias, R. *J. Chem. Inf. Comput. Sci.* **1983**, *23*, 102–108.
(7) Stobaugh, R. E. *J. Chem. Inf. Comput. Sci.* **1985**, *25*, 271–275.
(8) Fugman, R. *J. Chem. Inf. Comput. Sci.* **1985**, *25*, 174–180.
(9) Wiswesser, W. J. *A Line-Formula Chemical Notation*; Crowell: New York, 1954.
(10) *IUPAC Nomenclature of Organic Chemistry*, Sections A–F, and H; Pergamon: Oxford, England, 1979.
(11) Read, R. C. *J. Chem. Inf. Comput. Sci.* **1983**, *23*, 135–149.
(12) Weininger, D.; Weininger, A.; Weininger, J. L. *Chem. Des. Autom. News* **1986**, *1*(8), 2–15.
(13) Weininger, S. J.; Stermitz, F. R. *Organic Chemistry*; Academic: Orlando, FL, 1984.
(14) Leo, A. J. *J. Chem. Soc., Perkin Trans. 2* **1983**, 825–838.

COMPUTER SOFTWARE REVIEWS

Chemtext

YECHESKEL WOLMAN

Department of Organic Chemistry, The Hebrew University of Jerusalem, Jerusalem, Israel

Received September 9, 1987

There has been a need for a good chemical word-processing program that would enable the chemist to create documents which include complex chemical and mathematical formulas, documents in which one would be able to place chemical structures, molecules, and/or reactions directly into the text, manipulate them, store them, and finally print them on various printers.

CHEMTEXT (Version 1.1.) is an excellent answer to the above need. It is part of the Molecular Design Ltd (MDL) Chemist Personal Software Series (CPSS). CHEMTEXT is a word-processing program that consists of a main menu together with four different editors (Document, Molecule, Reaction, and Form editors). In order to use the program one needs an IBM personal computer (PC, PC XT, PC AT) or compatibles with 640K memory, color or monochrome monitor, graphic board (Hercules, IBM CGA, IBM EGA, or compatibles), a mouse (Mouse System or Microsoft Mouse), at least one floppy disk drive, a hard disk (at least 10 MB), and a dot-matrix or a laser printer.

The main menu is the entry point into CHEMTEXT. It is used to insert images into documents from the editors; it is as well a drawing editor on its own. It is used to enhance images

before they are inserted into the various documents, as well as for adding lines, boxes, arrows, etc. to the images using a variety of pen styles and widths. Text in 12 different fonts can be added anywhere on the screen. Complex chemical and mathematical formulas and equations can easily be created by using the Main editor.

The Document editor is a powerful word-processing program that contains a few additional features rarely found in most word processers such as editable macro commands and windows for viewing up to nine documents, any two of which can be seen on the screen at the same time. Boldface, italics, underline, and/or any combination of these, superscripts, and subscripts, as well as nine different fonts or typeset (e.g., Greek, Fraktur, Cyrillic, script, mathematical symbols) can be used with this editor.

The Molecular editor is used to create new molecules or to edit existing ones. Molecules are created by freehand drawing or by using predetermined templets (58 of those are available). One may also save his/her own templets (their number is practically unlimited). Once created or edited, the molecule could be saved as a molfile or placed anywhere in a document. As a matter of fact, this editor is identical with the one in CHEMBASE, which is another component of MDL CPSS.

The Reaction editor is used to create new reactions or edit old ones. Once created or edited, the reaction could be saved as a rxnfile or placed in a document. This editor is identical with the one in CHEMBASE.

The Form editor is used to create new forms for displaying structures and data or to edit old ones. Once created or edited, the form could be saved as a file or placed in a document. Like the other editors, this one is also identical with the one in CHEMBASE.

The program consists of nine disks (eight disks and a backup of the program disk) and a user guide in two parts (the first part, 196 pp, contains installation instruction and a tutorial; the second part, 240 pp, deals with the various components of the program). The manual is very clear and easy to read and use. Besides the manual, the program comes with a very good Quick Guide as well as a Quick Reference that lists and explains the various program commands.

The program seems to be so interesting and the Quick Guide is written in such a way that one tends to try to explore it even before reading the manual, usually with considerable success.

The program includes some tutorial files that could be edited into a whole document containing text, sketches, molecules, etc. The tutorial manual tells the user step by step how to use the program in order to create and print a finished complex document—an excellent idea that makes one immediately familiar with the program even before learning all its commands. Another interesting feature are the detail help screens available anywhere along the program. These help screens are of two versions: one version for the novice user, another for the expert.

All the commands in the Molecular, Reaction, and Form editors as well as those of the main menu are chosen from pull-down menus. The situation is somewhat different in the Document editor, where they are all keyboard commands; however, most but not all are also menu chosen.

Text and images can be imported to CHEMTEXT from other Molecular Design Programs: CHEMBASE, MACCS, MACCS-II, REACCS, and CHEMLAB-II. The program contains some tools for the conversion of RS/1 files and HP-GL plot files into MDL format metafiles, thus enabling their incorporation into CHEMTEXT.

One really cannot see on the screen how the printed page would look. To see it, one has to use the print preview function (e.g., right-hand justification is not shown on the screen; however, it is seen in the page preview).

There is one main feature that is missing in the program, and I hope it would be included in a later version—the undo command. Although in the Form editor there is an undo command that throws away any changes that were made in the form since one entered the editor, it does not restore a structure that was cleared with the Clear Structure command. Indeed, one may cancel a command, but it works only before the command is executed. However, if cancellation is given during execution, whatever had been carried out could not be returned to the "original" situation.

Another problem that limits to some extent the usefulness of CHEMTEXT is that files cannot be transferred to mainframes for storage or printing (when one has a laser printer under a Vax or a Prime) unless CHEMTALK is used, or that dot-matrix or postscript files are used. The problem with these files is their size. A 10-page document is about 400–450 kB—a huge file to transfer or save (although it can be done).

To sum up, we have here a very good chemical word processer, which has been needed for a long time. This program enables the chemist as well as the scientific secretary to compose and print a document containing complex chemical and/or mathematical formulas in a short time and with good quality. The only inhibitory factor for purchasing this package is its price, which is $1500. For universities there is a reduced price of $500. In keeping with the recent enlightened attitude of software companies toward their users, the program is not copy protected.

ChemIntosh and ChemPanion, Organic Structure-Drawing Desk Accessories for the Apple Macintosh†

ALLAN L. SMITH
Chemistry Department, Drexel University, Philadelphia, Pennsylvania 19104

Received October 27, 1987

As chemists began during the past 15 years to prepare manuscripts for publication using computers, the first software to become widely used was word processing. Word processors initially had line- or string-oriented user interfaces, since their progenitors were line-oriented text editors. But much of the chemical literature uses embedded chemical structure drawings that could not be handled with a character-based screen editor. With the advent of bitmapped displays such as the Apple Macintosh, however, screens could be painted pixel by pixel, and the potential existed to produce on a personal computer chemical manuscripts that intermingled text and graphics.

Since the Macintosh has a standard method (the clipboard) of transferring data from one software package or "application" to another, the first approach taken by chemists using Ma-

† Version of software reviewed: 1.2 (both ChemPanion and ChemIntosh). Publisher: Softshell Co., 88 Fox Chapel Road, Henrietta, NY 14467. Hardware requirements: Macintosh 512KE, Macintosh Plus, Macintosh SE, Macintosh II.