Chapman and Hall Dictionary of Organic Compounds on CD-ROM

Loren D. Mendelsohn

Science and Engineering Library, Wayne State University, Detroit, Michigan 48202

One of the more heavily-used compendia of data on organic compounds is Chapman and Hall's *Dictionary of Organic Compounds* (DOC). Originally published in 1934, the fifth edition was published in 1982. Additionally, supplements are published annually. Chapman and Hall also publish similar dictionaries covering specific classes of organic compounds such as alkaloids, natural products, organometallics, and others. The DOC with its first ten supplements contains data on over 145 000 compounds in over 50 000 entries. Obviously, no effort has been made to provide comprehensive coverage; however, the entries included represent most of the more intensively studied organic substances. Each entry includes information such as the CAS registry number, melting points and boiling points of the base compounds and their derivatives, literature references, and (in most cases) structures. The data included are generally reliable, but one reviewer has pointed out several serious errors and inaccuracies in structures, names, and indexing.[1]

A source like DOC easily lends itself to being automated; it has been searchable for many years as an online file through DIALOG, STN, and other vendors. This year, Chapman and Hall released it as a CD-ROM database: the *Chapman and Hall Dictionary of Organic Compounds on CD-ROM* (DOC CD-ROM).[2] The DOC CD-ROM requires an IBM compatible PC with an Intel 80386 or equivalent CPU, 4 MB RAM, 5 MB hard disk space, a mouse, and a CD-ROM drive. VGA color graphics is also recommended. The following software is also required: MS DOS 3.1 or higher, Windows 3.0 or higher, and the MS DOS CD-ROM Extension 2.1 or higher. The computer on which I reviewed the software is a 486SX 25-MHz machine with 4MB RAM and a Hitachi CDR-1503S CD-ROM player, running DOS 5.0 and Windows 3.1.

Installation is very straightforward. After loading the disk in the CD-ROM drive, the user types *WIN d:INSTALL* at the DOS prompt, where *d* is the CD-ROM drive letter. If one is already in Windows, click on *File* in the Program Manager, then double click on *Run*. Type *d:INSTALL* at the command line. The installation software will prompt the user for the name of a directory in which to install the software. The current version of DOC CD-ROM is not network compatible; while installation on a Novell NetWare drive was successful, the software would not run. A network compatible version is, however, under development.

DOC CD-ROM provides two approaches for searching the database: (1) text words and (2) structural formulas. When one starts up the program, it presents the *Application Main Menu*, from which one can choose *Text Search* or *Structure Search*. The Text Search software (Headfast/CD from Head Software International) enables search, retrieval, and display of the text of DOC entries. Additionally, this software can display the structural diagram associated with each entry. The Structure Search software (PsiBase for Windows from Hampden Data Services) enables searching on the structural formulas of any of the substances listed in DOC 5.

When one selects Text Search, the *CHCD Dictionary of Organic Compounds* menu is presented, from which one can choose *Quick Search*, *Form Search*, or *Command Search*. Selecting Quick Search brings up the following selections: *Chemical Name*, *Words in Text*, *Type of Compound*, *CAS Registry No.*, *References*, *Molecular Formula*, *Melting Point*, and *Boiling Point*. These selections represent the major searchable fields. In Quick Search mode, it is only possible to perform searches in one field at a time. Once a field has been selected, one can either type in a search term directly or select a term from a browsing menu. The Form Search option allows one to search on a number of selected fields simultaneously. When selected, it presents a list of all searchable fields. At each field, one may enter a term or terms appropriate to that field. Again, it is also possible to select a term from a browsing menu. Finally, the Command Search option allows free text as well as field-limited searching. It is much more versatile than the other options, allowing recombination of search terms and previous searches. It also makes use of a search command language similar to that used by various online database vendors, providing the ability to limit and display results, to save, retrieve, and execute search strategies, and to perform other similar functions.

All of the text search options use the same special symbols and Boolean operators to structure the search. An operator toolbox, listing all of the Boolean operators, is available under all of the text options and makes selection of the appropriate operator much easier. A special symbol keypad is also available under all of the text options, enabling the selection of Greek letters and other special characters for use in the search string.

The other approach to searching is the Structure Search option. When this option is selected, the program presents a blank window and a palette of chemical graphics tools. It is possible to draw structures freehand or to use any one of a number of templates included with the package. The software "knows" chemistry; it will not permit the generation of impossible structures. It is also possible to define generic structures. For example, one can define a structure with a halogen substituted at any node or set of nodes. Unless otherwise specified, all queries are assumed to be substructure searches, but it is also possible to perform exact searches on a specific structure.

One of the most useful features of DOC CD-ROM is its ability to transfer the results of text mode searches to structure mode and vice versa. Thus, one can perform a text search and switch to structure mode to view or modify the structure. Likewise, one can perform a structure search and switch to text mode to view the DOC entry or to modify the query by using text qualifiers.
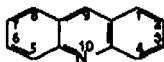
While DOC CD-ROM is a very powerful tool, it has several problems. The software generates windows which are quite different from those generated by most of the other Windows applications. For example, there is no file or edit menu on most of the windows, and the help command is at the left rather than the right of the screen where it is usually located.

COMPUTER SOFTWARE REVIEWS

J. Chem. Inf. Comput. Sci., Vol. 33, No. 5, 1993 **791**

Figure 1. Sample printout.

ALT-F4 does not exit the program as it does with most of the other Windows applications; instead, it backs the user up one screen at a time until the program is finally exited. The procedure for copying text from the entries to the clipboard is counter-intuitive, requiring the user to select *Output* from the menu bar and to choose between *Print* or *Copy to clipboard*. It is extremely difficult to modify the basic configuration, especially for printing. In short, much of the operation of the software runs contrary to what one would expect from a Windows-based software package. Additionally, it was my observation that the browse mode of text searching and structure searching were very slow. Part of the reason for this may be the CD-ROM player which I used, but even so it took nearly 15 min to perform a full substructure search on the anthracene nucleus. One final note on the quality of printing: I printed my search results on an HP LaserJet III with a Postscript cartridge, which ordinarily generates excellent quality text and graphics. Search results printed from the DOC CD-ROM had poor quality graphics and text, as is shown in Figure 1.

The DOC CD-ROM software can be compared with the *Beilstein Current Facts in Chemistry on CD-ROM*.[3,4] While the Current Facts database is larger in both extent and depth than DOC CD-ROM, there are many similarities in the software. For example, they both can use either text or structure to search the database, and they both can use Boolean operators. Field searching is also quite similar, as also is the system for structure searching. In general, the Current Facts software is easier to set up and customize. The procedures for printing and saving results are also quite straightforward, and searches are completed much more rapidly. On the other hand, the DOC software for drawing structures is much easier to use. Also, it is easier to transfer results back and forth between text and structure modes in DOC.

Overall, DOC CD-ROM has the potential to be an extremely useful tool. The entries contain useful information, and the ability to search on both structures and text makes this a very powerful source. Many (but not all) of the errors pointed out

in the review of the print version[1] seem to have been dealt with. Nevertheless, the difficulty of use of the software and the poor quality of the printout are problems which need to be resolved if DOC CD-ROM is to earn a place in the highly competitive market for such database services.

REFERENCES AND NOTES

(1) Smith, P. A. S. Heilbron's Dictionary of Organic Compounds, 5th ed. *J. Am. Chem. Soc.* **1983**, *105*, 6198–6199.

(2) The CD-ROM is available from Chapman and Hall, 2–6 Boundary Row, London SE1 8HN, England. The price is £2,000.00 if the print version is owned or £2,750.00 if it is not. Also available is a subscription to semiannual updates at £495.00 if the print version is owned or £595.00 if it is not. Various special offers are also available for subscriptions to multiple Chapman and Hall CD-ROM databases or for paying subscription fees in advance.

(3) Heller, S. R. The Beilstein Current Facts in Chemistry CD ROM. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 430–432.

(4) Berks, A. H. Beilstein Current Facts in Chemistry on CD-ROM, Version 1.0. *J. Am. Chem. Soc.* **1992**, *114*, 3576–3577.