COMPUTER SOFTWARE REVIEWS

NBS Standard Reference Database, NBS/EPA/MSDC Mass Spectral Database PC Version

YECHESKEL WOLMAN

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

Received March 7, 1988

Recently the National Standard Reference Data System (NSRDS) produced its first NBS Standard Reference Database in a PC version—the NBS/EPA/MSDC Mass Spectral Database PC Version. The database contains about 44 000 electron ionizing mass spectra, various "index" files (Core, Registry, Name, Form, MW, Ten) used for data retrieval, and the searching software. Mass spectral data could be searched in seven different ways: by identification number (system ID number using the Core index); by Chemical Abstracts Service Registry number (CAS RN) (using the Registry index); by chemical name (using the Name index); by empirical formula (using the Form index); by molecular weight (using the MW index); by abundance of major peaks (using the Ten index); and by a sequential searching of the database.

Hardware requirements for using the database are an IBM XT, AT (or compatible) or PS/2, a minimum RAM of 512K, and a hard disk. Storage space on the disk depends on the number of search options selected (index files instituted) and is between 8 and 14 MB. A standard graphic adaptor (Hercules, EGA, or CGA) is optional and needed for graphic display of the spectra.

The PC version of the database is a somewhat correct version of the 1987 NBS/EPA/MSDC database tape. Incorrect CAS RN were removed from about 2000 spectra. A total of 200 spectra that were not identified by a recognized chemical name or whose chemical name did not correspond to the given molecular formula were omitted. In about another 200 spectra, peaks caused by air and/or water were subtracted, and a small number of spectra were replaced by a better quality spectra of the same compound. It seems that in many cases the CAS RN were removed from the spectra records and not from the Registry index; thus, searching for some compounds by using their RN results in locating spectra of completely different compounds. E.g., searching for the spectra of Δ^{16}-pregnenolone (1162-53-4) yields the spectra of benzenamine, 1,2,3,4-tetrafluoro-; searching for the spectra of dibenzo[*a*,*d*]cycloocten-12-ol (1042-48-4) yields the spectra of 3-undecene, 10-methyl-. Some of the compounds are not defined, and it is a pure guess as to which spectra they are related, even if those spectra had a reasonably good quality index. E.g., dodecynone (ID-13573) has a quality index of 0.640, on which carbons the unsaturation and the carbonyl groups are ?; benzene, methyl-, trichloro (ID-16117), has a quality index of 0.640. What compound are we dealing with? Are we dealing with benzene, trichloromethyl- [α,α,α-trichlorotoluene-(98-07-7)]; benzene, 4-chloromethyl-1,2-dichloro- [3,4-dichlorobenzyl chloride (102-47-6)]; benzene, 5-methyl-1,2,4-trichloro- [2,4,5-trichlorotoluene (6639-30-1)]; benzene, 2-methyl-1,3,5-trichloro [2,4,6-trichlorotoluene (23749-65-7)]; or maybe another isomer?

Nice work has been done cleaning the latest version of the tape, but there is more work to do and I hope that it is going to be taken care of before the introduction of the next version, whether it is a tape or a PC one.

The database is very friendly and easy to use. Help menus could be obtained in any stage of the search, and in most screens one can find at the bottom of the screen the options concerning the next action to take. Choices are made by selecting a letter given in parentheses.

There are two things that currently limit the usefulness of the database. Unfortunately, one cannot add extra spectra to the database nor use the software to build one's own spectra library. The database contains a relatively small number of spectra that were selected according to some criteria; those spectra came to answer a good part of the need of the user. However, one would always have compounds of personal interest, and one would like to insert them in the database. It seems to me that an option for introducing extra spectra into the database would increase dramatically its usefulness.

Another problem is retrieval of spectra of derivatives. In many cases, especially in GC–MS analysis, the spectra are run on some derivative of the compound. There is no cross-reference in the name index from parent compounds to their derivatives used in the MS and/or GC–MS analysis of the parent compound. One such example is the spectra of 5-adenylic acid [better known as AMP (61-19-8)]. The mass spectra of AMP could not be retrieved from the database by using the Name index or the Registry index. However, two different derivatives of AMP that are used in MS analysis of the compound are included in the database (5′-adenylic acid, pentakis(trimethylsilyl) derivative (56145-14-3), and 5′-adenylic acid, *N*-(trimethylsilyl)-2′,3′-bis-*O*-(trimethylsilyl) bis(trimethylsilyl) ester (32653-14-8)]. Having such cross-references would increase the usefulness of the database and make the user's life much easier.

This is the first time a database is available on a mainframe (via the services of CIS Inc. and Telesystem Questel), a mini (via the tape version), and a PC. I am sure that with the decreasing cost of storage space and the increasing storage available on the PC we shall be seeing many more commercially available databases in a PC version.

The price of the database is very reasonable—$750. Taking into account that CIS charges $55 an hour plus CPU charges and Telesystem Questel charges $60 an hour plus communication fees (which are about $10 within the U.S.A. but $35–45 on international networks), just a few hours of searching the database recovers its cost.

To sum it all up, it is a good example of bringing a commercial database from the vendor mainframe (or the institution mini) into the chemist PC (in the office or home), an example that I hope is going to be followed by many others in the near future.