under Windows. STN Express is now available under the same user interface. There are various data systems for analytical instruments. Molecular Design Limited is announcing new Windows-based software and soon the Beilstein Institute will also offer its current software and database products under a Windows interface. It will probably take you a day to become familiar with Windows and then the same time as above to learn Winword.

In conclusion Word for Windows is a very powerful, but yet easy to use, word processor. It is particularly well suited for chemists because graphics and text can be combined very easily. The outline feature makes it simple to write scientific reports. A major drawback is that you cannot use it on your old IBM XT. The minimum hardware requirements include a fast AT-type computer, at least 2 MBytes of RAM and a fairly fast graphics card (for working without eye strain at least VGA or higher resolution). Winword should only be used with Windows 3.0 (or higher in the future) since it requires too much DOS memory if run under Windows-286 or Windows-386.

REFERENCES AND NOTES

(1) Wordstar 5.5, Wordstar International Corp., 33 San Pablo Ave., San Rafael, CA 94903.
(2) WordPerfect 5.1, 1555 N. Technology Way, Orem, UT 84057.
(3) MS Word 5.0, Microsoft Corp., One Microsoft Way, Redmond, WA 98052-6399.
(4) Chemtext, Molecular Design Ltd., 2132 Farallon Dr., San Leandro, CA 94577.
(5) Available from Microsoft Corp., One Microsoft Way, Redmond, WA 98052-6399. The price of Winword is $450.00.
(6) HiJaak 1.1, Inset Systems Inc., 71 Commerce Dr., Brookfield, CT 06804.
(7) Software Bridge, Systems Compatibility COrp., 401 N. Wabash, Suite 600, Chicago, IL 60611.
(8) Pagemaker 3.01, Aldus Corp., First Ave. S., Seattle, WA, 98104.
(9) Jochum, C. Pagemaker. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, (1) 40–42.
(10) ChemWindow, SoftShell International Ltd., 2754 Compass Dr., Suite 375, Grand Junction, CO 81506.

The NIST/EPA/MSDC Mass Spectral Database, PC Version 3.0

STEPHEN R. HELLER

Agricultural Research Service, BARC-W, Building 005, 10300 Baltimore Boulevard, Beltsville, Maryland 20705-2350

Received February 2, 1991

In the fall of 1973 the first version of the Mass Spectral Search System (MSSS) was made publicly available to the scientific community on the General Electric Mark III computer network (1). At that time it consisted of slightly under 9000 nonunique spectra and required a large time-sharing system to store the database and programs. Now, some 18 years later, the database has increased to almost 54000 unique spectra, and the entire system can be searched almost as quickly and flexibly on an IBM PC, either on a hard disk or on a CD-ROM, as it was searchable on the 1973 million dollar computer system. The June 1990 release, Version 3.0,[1] is the one reviewed here.

The National Institute of Standards and Technology (NIST—formerly NBS, the National Bureau of Standards) Standard Reference Data Program (SRD) has created an excellent and inexpensive source of mass spectral data and search software for the scientific community. In addition, the Aldrich Chemical Co., as part of their database activities, has created an identical CD-ROM version of the database. The NIST system is well designed to serve the researcher who needs access to a large mass spectral database for spectral identification or as an aid in structure determination of unknowns from mass spectral data. It is also a very useful tool for the classroom, both for courses on mass spectrometry or spectral interpretation, as well as any course in organic chemistry.

The details of the database have been described previously[2-6] and consist of 53994 EI (electron impact) only spectra. Each spectrum has a complete list of masses (m/z values) and intensities, a chemical name, as well as a display option for showing synonyms, molecular formula (and partial formula), molecular weight, CAS Registry Number, source of the spectrum, a Quality Index (QI) assigned to the spectrum,[7] and, when plotted out, a structure diagram (for over 95% of the compounds in the database), but there is no CAS Registry Number or other identifying number on the spectrum, only the name, molecular formula, and molecular weight. Regular updates of the database, which will include additional spectra as well as replacement of existing spectra with higher quality data, addition of missing chemical structures, and improvements to the software, along with new software capabilities, have occurred over the past three years since the first release of this system for the IBM PC.[6] Enhancements included in this latest version are the ability for users to add their own spectra to the system as a separate, searchable, private database, and a graphics display (both on the screen and printed) for 53994 compounds in the database.

The search software is almost as flexible and extensive as the original online system. What few search capabilities are still missing are being worked on, and it would seem that in a short period of time this PC version will exceed the capabilities of the highly versatile online system. The menus are easy to use, and the help messages are indeed helpful. The system has been well designed, and the system designer and programmer, Dr. Steve Stein of NIST, has listened carefully and well to the many suggestions from the hundreds of users who have purchased the system in the past three years. The 29-page manual (which is the same for the hard disk and CD-ROM versions), with two appendices, is quite readable. As seen in Figure 1, which lists all the system options, one can search the database by ID number (an internal numbering system code found on each spectrum), the CAS Registry Number, the chemical name, molecular formula (partial or complete), molecular weight, abundances of 10 major peaks, and by means of a complete sequential search of the entire database. About the only mass spectral search option currently missing is a search by neutral losses. Only the first 35 characters of a chemical name may be used in a search, but this does not seem to be too severe a limitation—how many users for example, can correctly type longer chemical names? The results of a search or spectrum look-up can be very quickly displayed (m/z values and abundances) or plotted. An example of the peaks and intensities in 1-docosanol, acetate, printed out and plotted are shown in Figures 2 and 3, respectively. The plots shown in Figure 3 took about 20 s to

```
Search By:

   D - Database Number
   R - Registry Number
   N - Name
   F - Formula
   W - Molecular Weight
   M - Major Peaks
   A - Any Peaks
   U - User Spectrum
   S - Sequential Method

       Enter Choice W

Enter Molecular Weight 368

To return to the Main Menu, press ENTER with blank field.  F1=Help
```

Figure 1.

```
        Name:   1-Docosanol, acetate
Formula: C24H48O2 ; MW: 368
CAS#: 822264;  EPA# 16366; QI: 769; DB#: 4067
Num Peaks: 184
  12    2;  14    4;  15   20;  16    7;  17   31;
  18  160;  26    5;  27   39;  28   32;  29  140;
  30    5;  31    5;  38    3;  39   37;  40    9;
  41  390;  42  100;  43  999;  44  120;  45    9;
  50    2;  51    2;  52    2;  53   21;  54  100;
  55  580;  56  340;  57  790;  58   49;  59   14;
  60    8;  61  670;  62   17;  63    5;  64    4;
  65    5;  66   14;  67  140;  68  250;  69  620;
  70  280;  71  400;  72   27;  73   40;  74    8;
  75    4;  76    2;  77    5;  78    4;  79   15;
  80   12;  81  110;  82  460;  83  830;  84  200;
  85  250;  86   24;  87   13;  88    5;  91    7;
  92    2;  93    7;  94    7;  95   53;  96  190;
  97  770;  98  200;  99   53; 100    8; 101   12;
 102    5; 103    2; 104    4; 105    4; 106    2;
 107    4; 108    4; 109   35; 110   86; 111  450;
 112  140; 113   57; 114    7; 115    5; 116   49;
 117    5; 119    2; 121    4; 122    5; 123   22;
 124   60; 125  250; 126   99; 127   40; 128    6;
 129    7; 130    8; 131    2; 133    2; 135    4;
 136    3; 137   14; 138   46; 139  140; 140   75;
 141   29; 142    5; 143    2; 144    2; 150    2;
 151    6; 152   34; 153   94; 154   60; 155   22;
 156    4; 157    2; 163    2; 164    2; 165    5;
 166   29; 167   69; 168   49; 169   19; 170    4;
 171    2; 179    4; 180   25; 181   52; 182   42;
 183   14; 184    2; 185    2; 193    2; 194   22;
 195   41; 196   32; 197   11; 199    2; 200    2;
 202    2; 207    4; 208   20; 209   33; 210   31;
 211    9; 212    2; 222   18; 223   27; 224   33;
 225   11; 226    2; 227    2; 236   14; 237   22;
 238   24; 239    9; 250   11; 251   18; 252   20;
 253    6; 255    2; 256    3; 264    9; 265   15;
 266   13; 267    5; 278    6; 279   12; 280   98;
 281   22; 282    5; 295    2; 296    4; 306    6;
 307    4; 308  220; 309   59; 310    9; 350    2;
 353    5; 367    8; 368    7; 369    4;
```

Figure 2.

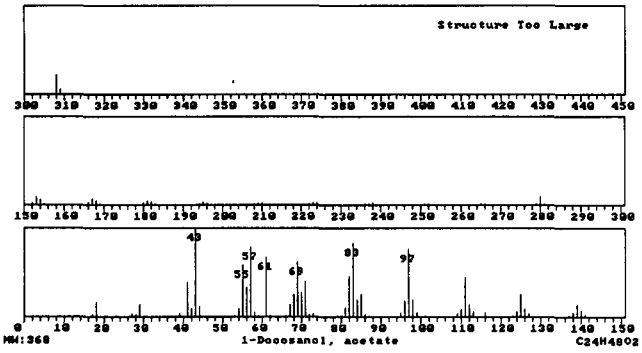

Figure 3.

print on an HP Laser Jet Series II printer. The quality of the algorithm for generating the structure diagrams is generally quite good. Structure diagrams, shown in Figure 4 (as the result of a molecular weight search), show some of the limitations on the system. For example, if there are many atoms in the molecule the diagram does tend to get a bit squashed, but there is a resize option which enlarges any structure to make it more legible. In some cases I found that if a molecule is "perceived" to be too "big" by the computer program it will

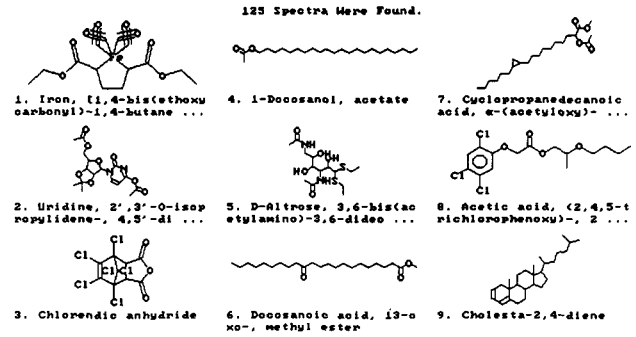

Figure 4.

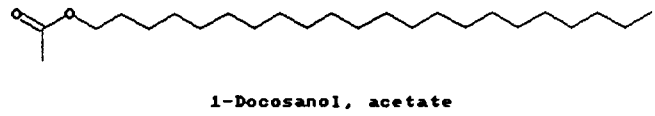

Figure 5.

not be displayed in the plot, as seen in Figure 3. However, as Figure 5 shows, there seems to be no reason why this molecule could not have been drawn in the blank space provided for the structure of Figure 3. Problems like this are few and far between, however, in my testing of the database.

Searches can be done either as a quick look-up in less than 1–2 s on the hard-disk version and not noticeably longer on the CD-ROM version, using the three largest peaks (coupled with up to 10 other peaks) or as a sequential search of the whole database (with filters or screens such as elements present or molecular weight), which usually takes about a minute of elapsed time on a Compaq 386/25 computer and about 2 min on an Epson 286/12 computer. Sequential searching worked well, and no problems were encountered. Molecular formula searches are easy to do. For example, there is no need to worry about order of the atoms, but upper case is de rigor (except for two-letter atomic symbols where the second letter must be lower case). Thus neither c7h4brn or C7H4BRN are found, a search for C7H4BrN succeeds. As mentioned above, the order in which elements are entered in the molecular formula search is unimportant; C7H4NBr or C7H4BrN are both found in the database.

The database and software are provided only on high density $5^1/_4$-in. or $3^1/_2$-in. floppy disks. The number of disks needed in this version was "only" 19. Loading the database and programs are easy, although it does take at least 20 min to transfer all the data onto a hard disk. Thus for the hard-disk version one should have sufficient storage capacity (some 23 Mbyte) on one's hard disk in order that the database can reside on the system. Otherwise, searching will be a considerable inconvenience. The CD-ROM version, which requires only 0.65 Mbyte of storage on the hard disk, is the more convenient for all but the most frequent user. The system has been nicely modularized so that you can choose which system features you want to have available should you not have all of the needed 23 Mbyte of disk storage available. The minimum system needs are about 8 Mbyte of hard-disk space for the disk version and 0.65 Mbyte for the CD-ROM version. The system runs best on an AT class machine, a PS/2 machine, or a 386/486 class machine. The programs require 512K of RAM. EGA, CGA, and Hercules graphics, all of which are optional for the spectrum plotting, can be used. The system is quite flexible in allowing one to specify the disk drive for the data (other than the standard drive C) and allows one to have the database on a different drive than the drive on which the programs are stored. Data can be printed on most any printer, and the system does allow for high-resolution plotting using an HP LaserJet series or equivalent printer. I have installed the CD-ROM and hard-disk versions and tested both systems on

a Compaq 386/25, IBM AT, Epson Equity 286/12, and a number of no-name clones. I had no trouble with any machine I tried. Only my wrist tired after loading all the floppy disks! The system comes with a 29-page manual and a 10-page Appendix A, which consists of a good set of examples. I would rate the technical information and accuracy of the manual as high.

As for searching the database, the system is quite easy to use and rates a good grade in the area of user friendliness. Searches for CAS Registry Numbers, ID numbers, and molecular formula took less than a second of elapsed time. The hyphens in the CAS Registry Numbers are not needed in a search; in fact, if one tries to enter a hyphen, the program will not allow it. I like features like that "know" enough to let you enter something in more than one way and still get the right answer. A molecular weight search took from 1 to 3 s depending on what, if any, constraints were placed on the search. The slowest search in the system, a sequential search of the entire database using 10 peaks, took about 4.5 min with the CD-ROM version using a Epson Equity 286/12 with a 40MB hard disk operating at a clock speed of 12 MHz and only 65 s on a Compaq 386/25 with a 110 MB hard disk.

This latest version of the system comes with the option to create and maintain a private database of mass spectral data, which is the "User Database" option. The user manual describes how this is done quite well and no one should have any trouble creating their own local database. Thus one can search either the master database, a private local database, or both. Of course the private database created will be on your hard disk, not on the CD-ROM.

At present there still are no plans for a Macintosh version of the system. Substructure searching of the entire database on the PC is probably not too far away. All in all, this is a very impressive, well-designed system, and for the low cost, this should be bought by every mass spectrometry or organic chemistry laboratory and organic chemical spectroscopy course. Useful, easy-to-use databases for the PC (and hopefully the Macintosh too) are the wave of the future. See the future now by getting this database and learning how to make effective use of it. It surely will be followed by many others like it in the next few years.

REFERENCES AND NOTES

(1) The database is available either on floppy disks or CD-ROM. The system program written by Dr. Stephen E. Stein, National Bureau of Standards, Office of Standard Reference Data, Building 221, Room A-325, Gaithersburg, MD 20899. The database on floppy disks is available from NIST, OSRD, Physics Building, Room A323, Gaithersburg, MD 20899 for $1050.00. The same database on a CD-ROM is available (catalog no. Z21,399-3) from Aldrich Chemical Co., 1001 West Saint Paul Ave., Milwaukee, WI 53233 also for the same price of $1050.00.

(2) Heller, S. R.; McGuire, J. M.; Budde, W. L. *Environ. Sci. Technol.* **1975**, *9*, 210.

(3) Heller, S. R.; Milne, G. W. A. *EPA/NIH Mass Spectral Data Base*, in five volumes, part of the NBS' National Standard Reference Data Series of Critical Data Compilations (GPO no. SN003-003-01987-9, NSRDS-NBS 63), 4634 pp., U.S. Government Printing Office (1978 and reprinted 1980).

(4) Heller, S. R.; Milne, G. W. A. *EPA/NIH Mass Spectral Data Base, Supplement Number One*, in two volumes, part of the NBS' National Standard Reference Data Series of Critical Data Compilations (GPO no. SN003-003-02268-3, NSRDS-NBS 63, Suppl. 1, 2151 pp, Dec 1980).

(5) Heller, S. R.; Milne, G. W. A.; Gevantman, L. H. *EPA/NIH Mass Spectral Data Base, Supplement Number Two*, in two volumes, part of the NBS' National Standard Reference Data Series of Critical Data Compilations (GPO no. SN 003-003-02268-3, NSRDS-NBS 63, Suppl. 2, 1110 pp, 1983).

(6) Heller, S. R. NBS Mass Spectrometry Database. *J. Am. Chem. Soc.* **1988**, *110*, 3336-3337.

(7) Dillard, J. G.; Heller, S. R.; McLafferty, F. W.; Milne, G. W. A.; Venkataraghavan, R. *Org. Mass Spectrom.* **1981**, *16*, 48-49.

Pro-Search, Pro-Cite, and Biblio-Links

STEPHEN R. HELLER

USDA, ARS, BARC-W, Building 005, 10300 Baltimore Boulevard, Beltsville, Maryland 20705-2350

Received March 13, 1991

More and more chemists are involved in online searching of databases and management of the ensuing results, but if they are like me, they don't enjoy it very much. The collection of programs from Personal Bibliographic Software Inc. (PBS), which are reviewed here, helps to alleviate much of the tedium of database searching and management of results. To accompany the programs, PBS also publishes *Format*, a quarterly newsletter, free of charge.

Pro-Search[1] is a program which acts as a front-end to online services, such as DIALOG and BRS. It simplifies searching for the novice and helps the experienced user. There are a number of other similar programs, such as STN Express and DialogLink, but Pro-Search has the advantage of being one of a series of programs from a single vendor. It suffers from the disadvantage that is tailored to the overall bibliographic market, while a program like STN Express is more oriented to the chemical, scientific, and technical community.

Pro-Cite[2] takes the properly converted search results and converts the bibliographic references into a form that allows the information to be used readily in bibliographies. Finally, Biblio-Links[3,4] is a set of programs that serve as the interface between Pro-Search and Pro-Cite and allow one to convert Pro-Search literature search results from a variety of online systems into the format used by Pro-Cite, after they have been downloaded to the hard disk.

Each of the three programs can be purchased separately for IBM PC or Macintosh computers. Copies of all three were available for my IBM PC, and so this review will cover all three programs, using the STN and DIALOG modules as examples of the Biblio-Links programs. All three programs can be ordered on $3^1/_2$ in. or $5^1/_4$ in. disks. The programs all require varying amounts of storage; with all the options installed, Pro-Search and Pro-Cite, the two largest, each require about 1.3 Mbytes. Each Biblio-Links module requires 0.15 Mbytes of disk storage. The programs require DOS 3.0 or higher and only 320K of RAM. A color monitor is optional. Installation is easy, but the manuals are somewhat confusing. The installation of Pro-Search, for example, is made unnecessarily difficult by the need to skip around the manual depending on a variety of parameters. The first 100 pages of the Pro-Search manual deals with installation and only then does one find the Table of Contents. I installed Version 1.08 (dated October 1989), but if you wish to upgrade to 1.08 starting from 1.03, 1.04, 1.05, or 1.07 (whatever happened to 1.06?), a slightly different procedure must be used. With the advent of easy-to-use install programs such as the one supplied