

Having your cake and eating it?— a new concept in reference spectroscopic database distribution

Antony N. Davies

ISAS, Institute of Spectrochemistry and Applied Spectroscopy, Bunsen-Kirchhoff-Str.11, Postfach 10 31 52, 44013 Dortmund, Germany

There has been a revolution in spectroscopic database delivery and, surprise, surprise, the manufacturers know it. For the first time in the history of reference database delivery a manufacturer will provide a customer with their entire database of reference spectra to “have and to hold” on their own computer system and only charge for the actual database hits the user finds as appropriate AFTER they have searched the database.

This is a radical new departure from the tried and trusted approach of buying the database most appropriate to your field of work (e.g. I’m looking into polymers and their additives by infrared spectroscopy so I will buy a polymer and additive database—simple really!). The concept is also an interesting alternative to the Internet web-based spectroscopy services recently highlighted in this column. This idea could also find favour in an industrial environment where no analytical data of any kind is allowed to be transferred across the Internet, ruling out the use of web-based client-server applications.

With this new system, christened rather strangely “HaveItAll™ IR”, from the Bio-Rad Sadtler Division, a single CD is distributed containing all of the reference spectra in their collection. The software will allow complete spectral searching through all of the databases in their collection as well as including your own internal reference databases and any extra hit lists which you have purchased during previous sessions with the software.

The user interface is quite familiar and the software automatically locates other products from this manufacturer to link them into this new system. The database searches run without having to have purchased the data through which the search is carried out, but you initially are presented with a variable length hit list which will, however,

only contain up to a maximum of 50 spectra from the databases you have not already purchased previously.

Installation

The version which I have tested for most of the run up to this article is somewhat different to what I believe will be the final release in that the databases and the software were delivered on two separate CDs and to achieve a faster search time I had to copy the databases onto the hard disk by hand. This is now carried out automatically provided you don’t “un-click” the option box on installation!

You will need the best part of 362 MB free to install the software package with the databases to your hard disk. In my version I was able to move the databases away from the program directory to a data disk once the installation was complete to save space on my system hard disk. The only change needed was simple and required the inclusion of the path to the location of the databases in either the library or the search menus. The installation now includes over 180,659 spectra in various databases and subsets of databases as well as apparently around 105,000 associated chemical structures. It should be noted that duplicates are included, although I have not found out just how many spectra are duplicates.

Spectral searching

Once installed, it is possible to load your unknown spectrum from a wide variety of formats. I was pleased to see that by using the JCAMP-DX file type I was automatically presented with the text from the JCAMP-DX field ##TITLE= as an annotation to my unknown in the spectrum display window (see Figure 1). The software allows

you to search all the newly-installed databases without having to register but initially all that you will see is the reference database alphanumeric code and un-annotated spectrum of the hits matching your unknown (Figure 1).

In order to view the full information on each spectrum in the hit list, you will have to register with the manufacturers and purchase a number of “keys” to unlock the identities of the substances in the hit list (see Figure 2). The default view in Figure 2 is of course a stunningly uninformative example as we were looking for a white industrial residue and the associated data only revealed a generic name for a possible identification of the major component! However, if an extra window “Properties” is opened some of the associated information can be displayed which annotates this database entry (see, for example, Figure 5 below).

As in the case in point the query sample was not part of a regular analysis and so our standard databases didn’t contain the substance we were searching for. This is, however, where the advantage of this concept starts to pay off. By using this system we were able to check first if any suitable candidates were proposed by the full spectral search before using one of our pre-paid hit list keys to reveal the results. Had we seen nothing in the hit list of any interest we could have tried a different search strategy or simply abandoned the search without incurring additional costs.

Once the purchased “key” has been used all the spectra in the unlocked hit list are then yours to use in future searches. Unfortunately one thing that annoyed me in the version I was testing involved the “hiding” of the information already purchased in previous hit lists. If I included an unlocked hit list in the new spectral search the data

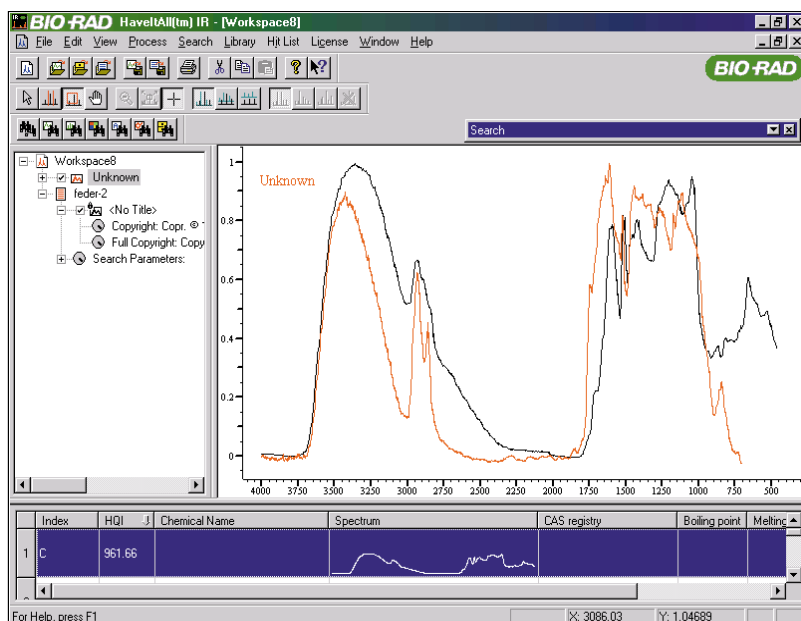


Figure 1. Spectral search showing the hit list (bottom window) revealing only the database code (under the heading INDEX, the Hit Quality Index (HQI) and a thumbprint of the spectrum.

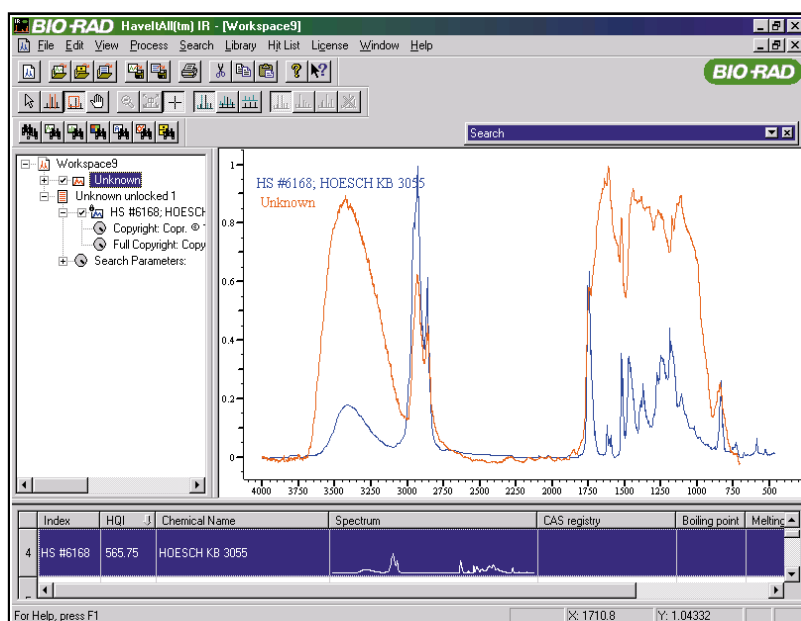


Figure 2. Having used a pre-purchased key to “unlock” the hit list you now have access to more relevant information than the database code!

was re-hidden in the new results list meaning that if I wished to see the full sample relevant data I would have to use another of my pre-paid keys to “unlock” the data I had already paid for once before! It is, however, always possible to compare the new hit list with the previously unlocked hit list by hand without being required to use up a pre-paid key.

The usual warnings apply here as to any other database search system such as to remember that often the most important hit in your hit list is often well away from the first in the list, so

take time to look down the hit list for better results. One additional tip which I can give is that although you now have all the reference databases at your fingertips don't (as I did at first like a fool!) automatically opt for searching all the databases at the same time. Although this might be an option of last resort it is far more efficient to spend a few minutes selecting the databases to be searched, through intelligently knowing what you do about your unknown sample! Another improvement which can be easily implemented is to split up the different

categories of reference databases such as Raman, Gas Phase, Pure Compounds etc. into separate directories. This saves time when selecting which databases you wish to search through as you can use the library paths option to restrict the databases offered up for selection. Currently all 89 databases have to be individually selected if you do not wish to search through all databases. I hope that in future different set-ups can be saved for different types of analytical problem.

Other search options

The system allows searching of the structural information stored with the databases and if the manufacturer's chemical drawing package hasn't been purchased for the system currently in use, then a cut-down version is supplied called the Query Editor. This has the one major drawback of not importing standard molfiles as the full version is capable of, but being restricted to importing ChemWindow (*.cwg) files (Figure 3). It is also incapable of saving the .cwg files it generates which is also restrictive.

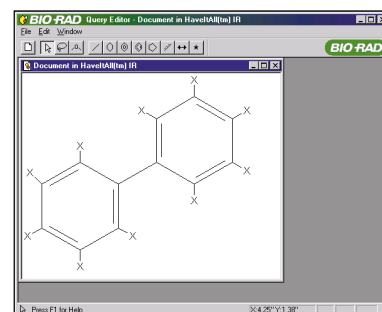


Figure 3. A cut-down version of the ChemWindow package is supplied as standard.

I found this degree of reduced functionality against the full version of the structure package unnecessary, annoying and rather petty as the structures I wished to search then had to be all re-entered by hand! (maybe I'm just getting lazy!).

However, the transfer from the drawing package to the structure search window works well enough (Figure 4). In the structure search displayed I wished to find all polychlorinated biphenyl (PCB) entries in the databases and so used the “X” substitution for “any halogen”.

The structure search results are presented in a similar manner to the results of a spectral search and I have additionally opened the “Properties” and

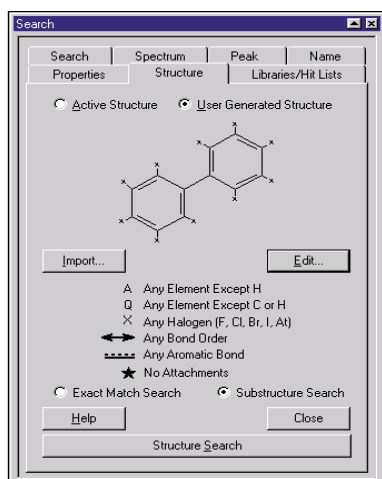


Figure 4. Structure transferred from the drawing package to the structure search window ready for searching.

“Structure” windows to get a clearer idea of what results the substructure search had thrown up (Figure 5).

It should be noticed here that the complete information on the items populating the hit list is available without having used any pre-paid keys. This is because structure searching is treated as a “look-up” operation. “Look-ups” are any searches of the database apart from full spectral or peak searches and these yield a maximum hit list size of five hits but then with the full information content on the hits. “Combined Searches” are exactly what they say, combining any of the single search options. It is possible to purchase only look-up functionality ruling out the ability to reveal full spectra search data.

Another search option is through Property Search where a vast number of different property fields are available, such as simple data entries like melting points and boiling points to more obscure information such as turbidity and UV sensitivity! I have not managed to find out just how highly populated these data fields are, but they could certainly be useful in a regulated environment for storing organisation internal reference data.

Name searching and peak searching are also available, although I would have liked to have seen a more intuitive link between the mouse position in the spectrum plot and the peak list. Although a peak list is generated automatically from the active spectrum when this type of search is opted for, further peaks force the user to type them in by hand! There doesn't appear to be any direct link between the mouse position and the peak list. This may be possible but I hadn't found out by the time this article needed to go to

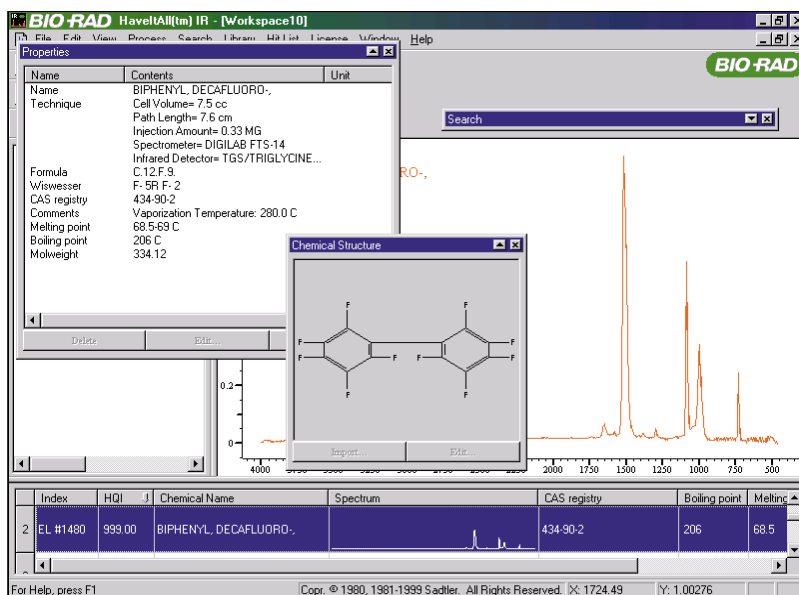


Figure 5. Results of a structure search showing the full associated information.

press. This doesn't mean that such a function doesn't exist and if somebody has the time to read through the 164 page manual and finds out how to do this please let me know! A quick conversation with Jeff Seymour from Sadtler had already revealed such a short-cut which I had completely overlooked when deciding on spectral ranges to be searched (Figure 6).

The last word

I think the database distributors deserve to be congratulated on a radical new concept, which certainly opens up a whole new range of opportunities for the analytical spectroscopist. It could be of particular use in an industrial envi-

ronment where the need to upload your spectroscopic data over the Internet in order to access the large databases becoming available on the Web may well be barred.

For academics the look-up only option mentioned above might be a more affordable option to the full version with pre-paid hit list keys. Whether either concept has a long-term future will depend on the pricing policy, which is out of the remit of this column to comment on. The manufacturers believe that the system could lead to increased purchase of the full databases as users find themselves constantly using their pre-paid hit list keys to open hit lists containing spectra from the same databases.

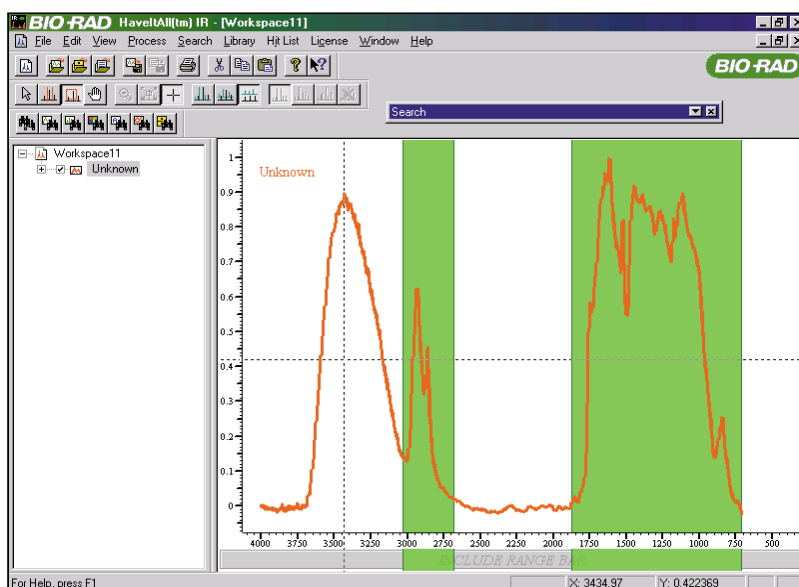


Figure 6. Select the spectral range of interest for the database search using the mouse.