

Spectral Library Searching and FT-IR / FT-Raman Spectroscopies: The Effects of Resolution and Search Algorithms

Michael Fuller, Ph.D. and Robert Rosenthal, Ph.D., Nicolet Spectroscopy Research Center, Madison, WI, USA

KEYWORDS

FT-IR, FT-Raman, spectral library searching, resolution, identification, subtraction, spectral noise

INTRODUCTION

Commercial infrared and Raman spectral libraries are generally only available at 16 cm^{-1} resolution by 8-bit ordinate precision. This is relatively low resolution compared to the typical spectrum acquired at 4 cm^{-1} by 16-bit precision. The low resolution format for libraries was selected to optimize library search speeds and reduce computer hardware requirements. Full 4 cm^{-1} spectral libraries offer a number of advantages including the positive identification of closely related compounds and greatly improved results when using spectral subtractions.

Currently, commercial spectral libraries for infrared and Raman spectral analysis are typically only available at 16 cm^{-1} resolution by 8-bit ordinate precision. Historically, these low resolution library formats were selected to optimize library search speeds and reduce expensive hard disk space requirements. Today, high speed, 66 MHz desktop computers with 500+ Mbytes hard disks are relatively inexpensive compared to the total cost of spectrometer systems. We have recently re-evaluated the importance of full resolution libraries (4 cm^{-1} resolution by 16-bit ordinate precision) as applied to the analysis of real-world samples.

When vibrational spectroscopy libraries were first being developed, researchers concluded that correct search result lists could, in most cases, be obtained using deresolved spectral libraries.¹ As mentioned above, this research and expensive mass storage devices led to deresolved spectral libraries becoming the standard. While it is true that the quality of search results can often be satisfactory using deresolved libraries, using full resolution digital libraries can provide incremental benefits in library searching.

SPECTRAL RESOLUTION & SPECTRAL IDENTIFICATION

When comparing the spectrum of the unknown sample to that of the best matches, small features in the spectra may be important in confirming the identity of an unknown. As an example, the

deresolved (bottom) and full resolution (middle) library spectra are compared to an unknown sample spectrum (top) in Figure 1. The library search results for both libraries correctly identify the unknown as 1-chlorobutane but a visual comparison of the unknown spectrum and the full resolution library spectrum (especially in the region from $1600\text{--}1300\text{ cm}^{-1}$) clearly provides the complete proof of compound identity.

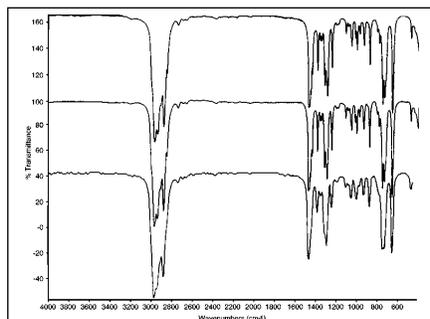


Figure 1: FT-IR spectrum of 1-chlorobutane and search results

Barbiturates are a group of compounds which are substituted derivatives of barbituric acid. Several hundred derivatives have been prepared and approximately fifty have been introduced as therapeutic agents. Chemically these compounds are very similar with only minor differences in the pendant groups attached to the central cyclic malonylurea ring. As a result of this structural similarity, these compounds give rise to very similar infrared spectra.

It is important for forensic labs to be able to positively identify the exact barbiturate that has been submitted for analysis. One way to positively identify a compound is using library searching where the spectrum of the unknown is automatically compared to a library of standard spectra. When digital infrared libraries were first being developed, researchers concluded that correct search matches could, in many cases, be obtained using deresolved spectral libraries (16 cm^{-1} resolution).

While it is true that the quality of search results can often be acceptable using deresolved libraries there are cases where higher resolution libraries are required for a correct match. Samples of secobarbital and sodium secobarbital were analyzed as KBr disks. The absorption bands in the spectrum of the sodium salt are relatively broad and the 16 cm^{-1} deresolved library spectrum provides about as much information as the 4 cm^{-1} spectrum (Figure 2). In contrast the acid form of secobarbital exhibits sharp, characteristic bands and

splitting patterns (Figure 3). In this case, a full resolution library spectrum is essential to properly identify the barbiturate.

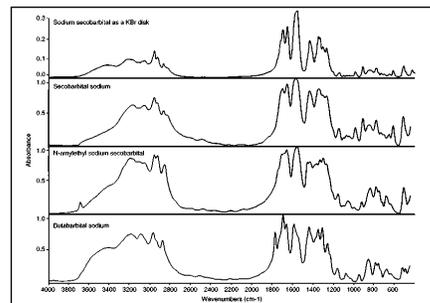


Figure 2: FT-IR spectrum of sodium secobarbital and search results

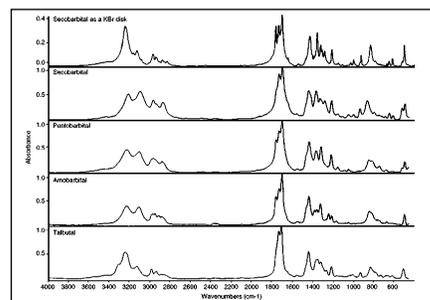


Figure 3: FT-IR spectrum of secobarbital and search results

LIBRARY RESOLUTION & SPECTRAL SUBTRACTION

Often an unknown sample is actually a mixture of compounds. The best library match may correctly identify the major component, but what about the identity of the other components? It is possible to recall a spectrum from the spectral library and to subtract this spectrum from that of the mixture. With deresolved libraries it is necessary to deresolve the unknown spectrum prior to the subtraction process. The deresolution process discards important information that could be used to identify additional components present in the sample. This often limits the usefulness of this subtraction approach.

A sample mixture containing dibutyl, diallyl, diisobutyl and dimethyl phthalates was analyzed. The mixture spectrum and those of the individual components are shown in Figure 4. The first search result identified dibutyl phthalate as the major component. The full resolution library spectrum of this component was subtracted from the mixture. The search results for this difference spectrum indicated the presence of diallyl phthalate and this library spectrum was subtracted from the first difference spectrum. The third search identified diisobutyl phthalate

and after the subtraction of this component the final search identified the last component as dimethyl phthalate. The final difference spectrum and the spectrum of dimethyl phthalate are shown in Figure 5.

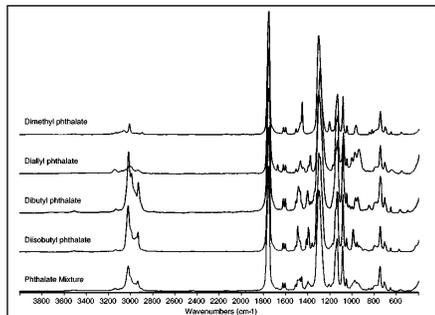


Figure 4: FT-IR absorbance spectra of pure phthalates and mixture

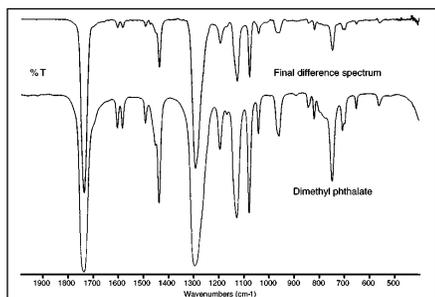


Figure 5: Final FT-IR difference spectrum and reference spectrum of dimethyl phthalate

SEARCH ALGORITHMS & THE EFFECTS OF SPECTRAL NOISE

A number of different search algorithms for the analysis of infrared and Raman spectra have been described in the literature.^{1,2,3} In the absolute value algorithm the spectrum of the unknown is compared point by point to each of the library spectra and the differences are summed across the spectrum. Relative band intensities and positions are very important to the final result. The first derivative algorithm is similar to the absolute value algorithm except the first derivative of the unknown is compared to the derivatives of the library spectra. The derivative approach is not as sensitive to absolute positions and intensities but the derivative step tends to reduce the signal-to-noise.

In the correlation search algorithm the first derivative of the unknown spectrum is correlated with the derivatives of the library spectra. An overall correlation index ranging from 0-100 is calculated. This approach has proven superior to both the standard absolute and derivative approaches.

In the pre-filter algorithm the spectrum is divided into 25 zones and an integrated area calculated for each zone. The magnitude of the area in each zone is used to create a 35 point binary mask for the unknown spectrum. The actual unknown is then only compared with library spectra that have the same binary mask. This optional approach can be used to reduce search times for full resolution libraries.

The effect of noise on search results was studied using a spectrum of 2-bromopropane to which random noise was added. The original spectrum and the spectrum to which noise was added are shown in Figure 6. The Nicolet Aldrich library was searched using absolute value, first derivative and correlation algorithms. Only the correlation algorithm identified the correct compound. The top four correlation indices were 57.89, 39.74, 33.38 and 31.67.

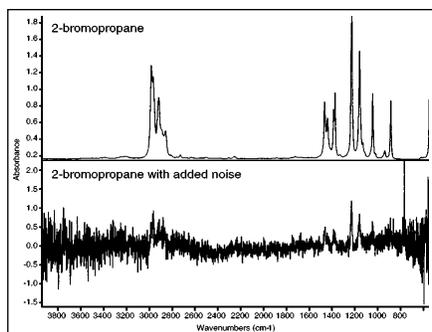


Figure 6: FT-IR spectrum of 2-bromopropane used to test search algorithms

EFFECTS OF BASELINE SHIFTS & SPECTRAL NOISE ON RAMAN DATA

An FT-Raman spectrum of polypropylene was modified by adding noise and a baseline curvature (Figures 7 and 8). To evaluate the search results a variable called the relative match number is defined. The relative match number is the percent difference between the best match and subsequent matches. Ideally, this number would be large for the second match indicating an excellent best match and much worse subsequent match indices. As an example of a good search result, the first match index might be 97 and the second match index 55. The relative match index in this case is $[(97-55)*100]/97$ or 43.

In this experiment, the derivative search algorithm was used. The match results for the original spectrum shows a relative match number of over 40 for the second match index indicating a good first match (see Figure 9). When noise

is added to the spectrum the first relative match number drops to around 10. This demonstrates the adverse effect of excessive noise on the derivative search algorithm approach. The search results for the bad baseline case are quite good as expected for the derivative algorithm which minimizes baseline differences.

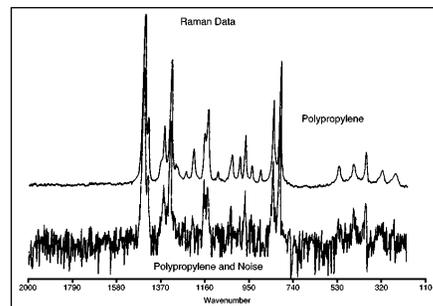


Figure 7: FT-Raman spectrum of polypropylene with and without added noise

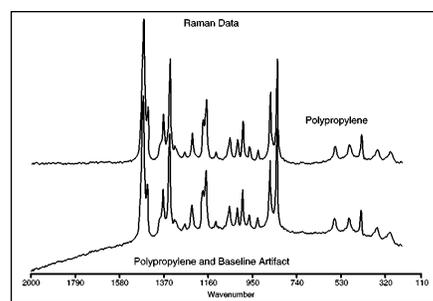


Figure 8: FT-Raman spectrum of polypropylene with and without baseline offset

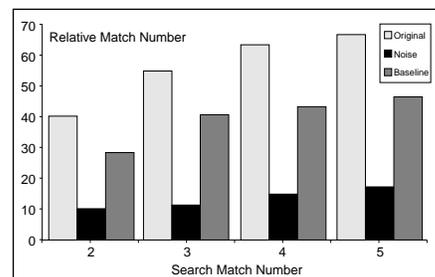


Figure 9: Relative search match values for FT-Raman polypropylene spectra

REFERENCES

- S.L. Lowry, D.A. Huppler and C.R. Anderson, "Database development and search algorithms for automated infrared spectral interpretation," *J. Chemical Info. and Computer Sci.*, Vol. 25, 1985.
- P.B. Harrington and T.L. Isenhower, "Applications of robust eigenvectors to the compression of infrared spectral libraries," *Anal. Chem.*, Vol. 60, number 24, 1988.
- J.W. Sherman, J.A. deHaseth and D.G. Cameron, "A window Fourier domain infrared search system," *Applied Spectroscopy*, Vol. 42, number 8, 1989.

This technical note is for informational purposes only. Nicolet Instrument Corporation makes no warranties, expressed or implied.