

## COMPUTER SEARCHINGS OF INFRARED SPECTRA

### Introduction

The purpose of this experiment is learn to apply the tools in Spectra Calc SEARCH to identify unknown compounds, and to use these tools to find relationships between compounds which give similar spectra. This process will provide a better understanding of infrared spectroscopy from a practical application point of view.

There is no other tool in the chemist's arsenal which has such a wide range of applications to compound identification as the infrared spectrometer. It may be applied to organic and inorganic liquids and solids, and when a high resolution instrument is available such as the Midac, gases and vaporized compounds may be analyzed with selectivity.

Infrared microscopes can identify the composition of particles 10 microns in size, and conventional systems can monitor production of tons of product per hour. Infrared telescopes peer over hazardous waste dump sites without setting foot on the property, and readily quantify organic and inorganic molecules in the low ppb levels in air. While the FTIR is capable of performing all of these feats, the data must be interpreted before it is of value. Given that there are millions of compounds in our environment, even a trained eye has difficulty. Computer searching is therefore essential in many of these applications.

Most infrared computer libraries of spectra range from about 100 to a few thousand compounds in size. It is quite probable that an unknown compound will not be present in a library. However, this is not as bad a situation as one might think. All is not lost if the compound is not present in the library if its identity can be narrowed down. In fact, in many cases an answer as simple as "the compound is not organic", or "this is not compound xyz" is all that is needed to solve a problem.

With the help of a library SEARCH, unknowns can be identified or classified to various degrees. Once narrowed down, authentic compounds may be obtained and tested to confirm identity. The tools in Spectra Calc SEARCH make it easy to shorten the list of possible candidates, and to exclude compounds from the list when their spectra do not agree with identified library compounds. Unidentified peaks in unknowns can be isolated and searched separately to indicate the kinds of structures which produce those peaks. Then by comparing the structures which have this peak in common, it may be possible to identify the functionality responsible for the unknown peak. If this process is repeated for all unknown peaks, a better understanding of the unknown compound may be attained. However, we must remember that molecular vibrations and rotations and combinations responsible for infrared absorption are complicated processes. Similar systems may look quite different in certain regions of the infrared if their symmetries are different. Nevertheless, peak classification is a useful tool for compound identification.

## Using Spectra Calc SEARCH

When using SEARCH to identify unknown compounds, 20 possible hits are provided, from which the best hit must be selected. The word "best" means chemically the best, to differentiate this from the SEARCH algorithm's best fit which is general, and may not weigh the molecular structural data in the most appropriate manner. For example, if you believe you have a pure compound which contains a carbonyl group, SEARCH may tell you that a non-carbonyl compound is the best match for your compound if many other regions of the spectrum are similar. In many cases, we will know that the SEARCH choice is wrong. In other situations a compound may be impure or a mixture of two or more compounds. Here the situation gets more complicated. With practice and patience, it is possible to identify components of mixtures by selectively scanning selective spectral regions, and/or by using subtract and repeating the SEARCH. In the latter case, a SEARCH hit is subtracted from the unknown spectrum, and the difference spectrum is re-SEARCHed.

## Acquisition of Unknown Spectra

Obtain the unknown compounds and decide upon the best method to acquire their spectra. With liquids, thin films between two salt plates usually afford good results. For the purposes here, it is recommended that one not use a cavity cell and a diluted solution of samples because some other manipulations may be required to give good solvent subtraction--see discussion at the end of this experiment. If the sample is a solid, grind it well in an agate mortar, mix it with potassium bromide, and press the mixture into a clear pellet. Use about 0.5 mg of sample and the minimum potassium bromide which produces a stable window. It may be necessary to dry this ground mixture to eliminate water.

Begin by acquiring a reference spectrum using the same salt plates or a blank KBr pellet. This will help to cancel adventitious impurities in the KBr or on the salt surface and to keep the baseline close to zero absorbance.

When using sodium chloride windows, zinc selenide, or certain other materials, there is a large signal above the 400 wavenumber cutoff due to sodium chloride (or other media) absorption of radiation. The large absorbance values created in this situation can be rectified by eliminating these regions of the spectrum. See the discussion at the end of this experiment.

It is important to obtain the best possible spectrum from the sample, which may require some purification steps and repeating the analysis more than one time. The concentration of sample in the potassium bromide, or the film thickness of a liquid sample should be such that no absorbance values are above 1 absorbance unit. Below 1 unit most absorbance values will be linear in concentration, and the spectrum envelope will be linearly related to the corresponding library spectrum.

Acquire the reference spectrum as follows:

**F2 C O ENTER ENTER**

Then label and acquire the sample spectrum:

**F2 C N XX ENTER** \*WHERE XX IS A DOS COMPATIBLE  
FILE NAME UP TO EIGHT CHARACTERS  
OR THE COMPOUND IDENTIFICATION  
PROVIDED BY THE INSTRUCTOR\*

**M XXXX... ENTER** \*WHERE XXXX... IS UP TO 70 CHARACTERS  
OF MEMO INFORMATION\*

Once the spectrum is acquired, save it to the default directory and any other directory desired:

**F2 F S ENTER** \*THIS SAVES TO THE DEFAULT DIRECTORY. TO  
SAVE TO ANY OTHER DIRECTORY, SEE  
DIRECTIONS BELOW.

The directory path is changed as follows:

**F2 E D PATHNAME ENTER** \*PATHNAME IS THE COMPLETE PATH  
INCLUDING THE ROOT AND SUBDIRECTORIES  
WITH CORRECT DOS SYNTAX\*

To save a spectrum named 12345 to the C drive in the subdirectory Library type the following:

**F2 F S ENTER C:\library\12345 ENTER** \*CAPITAL AND SMALL LETTERS ARE BOTH  
ACCEPTABLE. THE SPC FILE EXTENSION IS  
ADDED AUTOMATICALLY BY SPECTRA CALC\*

Now enter SEARCH:

**F2 S**

Type **C** for configuration and make sure the parameters here are set as described below  
Type the appropriate letter from the menu bar to select the any item which needs to be corrected.  
The configuration should be:

<b>M</b>	*Max Hits*	<b>20</b>
<b>R</b>	*Re search*	<b>No</b>
<b>T</b>	*Text Search*	<b>No</b>
<b>P</b>	*Peak Search*	<b>No</b>
<b>S</b>	*Spectral Search*	<b>Yes</b>
<b>C</b>	*Custom Data Search*	<b>No</b>

Press the escape key (Esc) to return to the main SEARCH menu. Now select the search limits by typing **U** for UNKNOWN/LIMITS. Using the left and right arrow keys or mouse, select the left region of the spectrum to be near 3600 and press **F9**. Then use the arrow keys or the mouse to select the right spectral limit near 400 and then press **F10**, then **ENTER**. Two vertical yellow lines should now

define the search boundaries on the spectrum. Now type **L** for libraries and select the library from the list available. This is performed by using the up and down arrows to highlight the library. Stop at SRL\_SR or other library which you are instructed to use and press **ENTER** to select. Press **G** for GO and the search commences. While searching, the screen display will give the library name, the number of compounds which have been searched, and a search index number which provides an indication of the goodness of fit between a library spectrum and an unknown. Indexes below 0.2 usually indicate excellent agreement for the region of the search selected.

When the search is completed, the computer will display a list of 20 spectra names in order of hit index number. If desired, the list may be printed by pressing the **F7** key. Press **ENTER** to see the first spectrum. The "hit" spectrum can be printed but not saved to disk. Using the up and down arrows or page up and down keys, one may move through the list of the 20 best hits and compare them with the unknown.

Some spectral regions of the unknown may not match the library compound. It is sometimes found that hits with higher index numbers appear to agree better with the unknown than some with lower index numbers. This is where the chemist's knowledge of infrared spectra can be fruitfully applied. Where are the areas of disagreement? Do these regions suggest the possibility of isomers, logical contaminants (such as a synthetic precursor), dissolved water, oxidation, etc. What is the likelihood that this compound is present in the storeroom? Is the search algorithm appropriately weighing the spectral regions? Thinking carefully through these questions will help decide the best fit based both upon the fitting algorithm and the chemist's knowledge.

A useful tool in SEARCH is the ability to set the search limits to any spectral window of interest, and then repeat a search with a new window. Let us say that a peak is present in the unknown compound, and absent in some of the best matched library spectra. The frequency value of the peak may suggest that an impurity or additive is responsible, or that the unknown is a derivative with new functionality. It is instructive to peruse the library spectra for compounds which also have this unique peak, and then decide based upon pattern recognition, if any of the new hits are related to the unknown. Then taking all the information together, one might have a good idea of the identity of the unknown compound. The entire library may be examined by entering ARCHIVE from the main SEARCH menu, then CHOOSEing the library to peruse, then using DISPLAY to go through all of the spectra in the library.

At least one of the compounds assigned in this experiment is not present in the spectral library. This will probably be evident from the indices found in the search hit list. When the compound(s) don't "fit" well, we will use the limit tool with SEARCH to try to classify the compound and narrow down the list of possible candidates. Begin with the limits set to cover the entire spectral region of the compound. Then move the limits around individual peaks and peak regions and rerun the search. Don't let the search index number dominate your thinking about the quality of the match. You decide what looks best. Pay careful attention to types of compounds whose spectra match with the unknown in the search region and ask what is in common with these compounds. Differentiate between the hydrocarbon region of the spectrum and the fingerprint region. Make notes of the match compound. Try to answer the following questions about the unknown. Some important spectral regions are bracketed to assist in your decision making. What processes are referred to by these bracketed regions?

## Questions

1. What major functional groups are present?
2. Does the compound have a long alkyl side chain? (3000-2700, 740-720)
3. Is the compound aliphatic, aromatic, or both? (3100-2700, 1650-1550)
4. Are there methyl groups in the compound? (3000-2900)
5. Are there methylene groups in the compound? (3000-2900)
6. If the compound is aromatic, is there any indication of the substitution pattern on the aromatic ring? (2000-1670, 900-690)
7. Is the compound an olefin, and is there information on the substitution pattern, stereochemistry, or conjugation of the double bond? (3100-3000, 1700-1550, 1000-650)
8. Is the compound an ester, an acid, an amide, a ketone, or an aldehyde? (1800-1650)
9. What types of compounds have similar peaks to mine, and what peaks does my compound have that similar library compounds do not have?
10. What is the identity of my unknown compound?
11. What three compounds in the library are chemically the closest to the unknown? What chemical features in these three compounds do not agree with the unknown?
13. Based upon your observations, how do you think the SEARCH algorithm works? What might be done to improve this algorithm?

## Removing Large Absorbing Peaks

When using sodium chloride, cavity cells, or in certain other situations, off scale peaks arise due to virtually complete absorption of light. No spectrometer can perform properly under such conditions. These situations manifest themselves by producing very high almost vertical peaks in some situations. The screen window normalization process adjusts these peaks to fill 80% of the spectral window, and the result is that the rest of the spectrum becomes lost in the baseline. This calls for either cropping the window to eliminate the end of the spectrum which is a problem or using ZAP in the arithmetic menu, which can remove any region or regions of a spectrum. When using ZAP, the window spectrum is permanently changed, but the disk spectrum (if it has been saved) stays in the original form.

When using sodium chloride it is often desirable to eliminate the region below 600 wavenumbers.

If this is the case, the spectrum's right limit should be changed to a value which deletes this region. For example, to cut off all frequencies below 600, enter the following commands:

**F2 E L R 600 ENTER Q**

Then press **F7** to save this window size. If the window size is expanded to the default value of 400 wavenumbers for example by pressing **F4**, then pressing **F8** will return it to the 600 wavenumber limit.

When a spectrum window has been reduced in wavenumbers to eliminate undesired features as discussed above, only the screen version of the spectrum is cropped. The truncated data points are still stored in the screen memory. To delete these from the screen memory, run the DOPROGRAM SMALL as discussed below. If you wish to make the disk copy SMALL also, then the spectrum must be re-saved. It is desirable to perform this operation when large absorbance peaks appear at the low frequency end of the spectrum due to the sodium chloride (or other window) cutoff. Otherwise, upon entering SEARCH, the full re-normalized spectrum will appear. To perform the truncation with the program SMALL follow the directions below.

With the spectrum in the active window cropped to the desired region, enter the following:

**F2 A D SMALL ENTER**

Now when you hit **F4** the file will not expand beyond the current window.

To convert the disk version to this small size:

**F2 F S**            \*THIS RE-SAVES THE SPECTRUM WITH THE  
                      SAME FILE NAME\*

ZAP is another way to remove selected regions of a spectrum. ZAP is entered as follows:

**F2 A Z Z** \*NOW SELECT THE REGION TO ZAP WITH THE F9, F10  
                      AND ARROW KEYS AND PRESS **ENTER**\*

It is a good idea to have the spectrum saved on disk before using ZAP. When ZAP is used, the ZAPped region is gone forever.

### **Notes to Instructors**

In the early 1980's, FTIR spectrometers began to appear in university laboratories. The FTIR spectrometer provided another dimension to infrared spectroscopy--digitized data within a computer memory. (Some dispersive instruments also produced digitized data). In addition to providing an entry to invaluable tools and techniques for manipulating spectra, computer infrared data files also provide for computerized searching of infrared spectra. The purpose of this experiment is to introduce to the student the value of this tool, which goes beyond identifying

unknowns from a long list of candidates in the library. It teaches by example how structural perturbations manifest themselves in the spectra.

An added benefit of this experiment is the creation of a library of infrared spectra for the compounds in the laboratory storeroom. This is accomplished by providing each student with at least one unknown compound from the stockroom which is not in the current SEARCH library. If the student produces a good quality spectrum from this unknown sample, the spectrum is then entered into a library or libraries. The library can grow each semester and becomes a useful instructional tool and problem solving tool for the entire department. Unlabeled chemicals can easily be identified if they are in the library. Another benefit is that the spectral data in this library is not copy protected and can be written to disk for further analysis.

The instructions provided the student are for a version of Spectra Calc that used the older collect menu. It may be necessary to slightly modify the collect instructions below if you are using a new version or if you are using Lab Calc. Lab Calc also eliminates the need to hit the F2 key to bring up the menus, so this should be dropped when called for when using Lab Calc. It is also possible to write macros for most of the operations below, and provide the student with the macro names. This is left to your discretion.

### Unknowns

Two unknowns are provided each student from which they must produce good quality spectra. One of the unknowns is selected from the list of compounds present in the current 200 compound library, and the second unknown is not present in the library. Suggest to the student that one or both of the compounds may not be present in the library, to let them decide if the hit is good enough for positive identification. The list of compounds in the library can be printed out as follows:

### **F2 S A C**

Now using the up and down arrow keys select the library SRL\_SR or any other library you wish to print, then choose print.

It is recommended that each unknown be assigned a unique number with less than eight digits for DOS and Spectra Calc compatibility. Part of this number could be a student identification number. If spectra are then saved to a common directory path for construction of the library, duplicate sample names will not overwrite one another. It is also recommended that each student submit a hardcopy of his or her spectra with the laboratory report. It is usually easier to review a spectrum for general accuracy with a hardcopy, and should someone accidentally overwrite a student's spectrum on disk, all is not lost. It is further recommended that each student has his or her own floppy disk and keep all of their work backed up on this disk.

The second unknown compound is selected from the laboratory storeroom and is not included in the list of compounds in the library. Optionally, one can provide the student with a mixture of two compounds, which can be a difficult or simple identification problem depending upon the mixture. A suitable spectrum must be acquired from this sample, which usually means keeping all

absorbance values below 1 absorbance unit. The spectrum is saved to a special directory which will become the source for the library files and should not contain unwanted SPC files. When it comes time to create the library, all spectra are checked for quality and accuracy and if not suitable, they are deleted.

Before creating a library from the student's spectra, the spectrum memo needs to be changed to include the identification of the unknown and other information, such as physical properties, CAS number, etc. The SEARCH utility allows up to 70 characters of information. The student can do this editing after turning in the assignment, or the laboratory instructor can do this before creating the library. To edit the memo:

**F2 F M \*TYPE IN 70 OR FEWER CHARACTERS\* ENTER**

To create the library, follow the steps below (also refer to the instructions in Spectra Calc).

**F2 S A N \*ENTER the Dos compatible library name\* ENTER**

**C \*ENTER a 2 letter code for the library\* ENTER**

**T S ENTER**

**F 400 ENTER \*THIS SETS THE LOWER SPECTRAL LIMIT\***

**L 4400 ENTER \*THIS SETS THE UPPER SPECTRAL LIMIT\***

**P R 4 ENTER \*THIS SETS THE SPECTRAL RESOLUTION\***

**B 16 ENTER \*THIS SETS THE ABSORBANCE RESOLUTION\***

**L yes ENTER \*THIS PERFORMS A TWO POINT BASELINE  
LEVEL\***

Spectra can now be entered into the library one at a time from the default directory using ADD ENTRY in the SEARCH/ARCHIVE menu. Alternatively, the DOPROGRAM BUILDLIB in the ARITHMETIC menu will create a library automatically from the default drive. All SPC files in the default directory will be entered into the library, so make sure unwanted spectra are removed. Also note that the library renames the spectra sequentially. Do not rely on having the original file name for identification. The memo serves this purpose.