Pay-as-You-Go Searching
A Review of FTIRsearch.com

James W. Rydzak and Paul N. Bourassa

Searching spectral libraries of known compounds is one of the most important tools in spectral interpretation.The problem limiting the use of this tool has always been the expense involved in obtaining a spectral collection large enough to give a decent probability of a match. This past spring, that limitation was eliminated for infrared and Raman spectroscopy when some well-known providers of instrumentation, spectral libraries, and software announced (for lack of a better term) "pay-as-you-go searching." Recognizing the tremendous marketing leap of faith it must have taken to offer this new service in opposition to its normal offerings, these providers should also receive high praise for significantly advancing the field of spectral interpretation by making their huge libraries available to a much larger group of scientists.

One such service was developed as a collaborative effort between Galactic Industries Corporation (Salem, NH) and Nicolet Instrument Corporation (Madison, WI). FTIRsearch.com is a real-time, Internet-based, spectral search service that is easy to use and that has a number of useful and informative extras.

The first informative extra is the search demo section where users can use the provided spectra to see how the service works, as well as using their own spectra. The only limitation is that the set of spectra in the library being searched is very small compared to service's actual libraries. It is, however, well worth playing with before proceeding on to the real thing.

The fee structure for the service is a simple credit-based system. Packages of credits are available for as few as four spectral searches for a starter package. Text searching (looking through library names to find a spectrum) uses fewer credits, so the same starter package could be used for 10 text searches.The charging system is set up to only deduct credits for the initial search, allowing for free refinements to the search parameters. By allowing users to look at their last spectral search, the service provides the capability of logging off and then coming back for additional refinements to the search at a later time.

The procedure for using the service is so simple that it necessitates a reminder of all of the careful steps that should be taken before the unknown spectrum is presented to the searching service. Things like obtaining as much information about the sample as possible, sample preparation (the key to world-class spectroscopy), and spectrum preparation should all be used to maximize the probability of success. The service's knowledge base provides some good information in this regard.

All that is needed for the service is an Internet connection and a standard Java 1 compatible browser. We found that the latest versions of Internet Explorer and America Online worked, but our versions of Netscape seemed to have some problems. Passing this information along to the service was simple and appreciated. Because security is a concern with this type of service, FTIRsearch.com uses the latest Secure Sockets Layer protocol encryption systems to maintain secure links between browser and server. However, with the dearth of interpretationalists, security breaches seem rather like the dog that finally caught the car it was chasing: now what do you do with it? However, as many Webmasters know, the threat to the FTIRsearch.com site itself is very real; therefore, Galactic uses an extensive firewall system where the site is hosted.

With our browser firmly connected to the Internet, we linked up with FTIRsearch.com (we supposed that FTIRandRAMANsearch.com was just too long of a name). After the member's log-in (setting yourself up as a member is a simple process of filling out online forms), all we had to do was upload (send them) the file of the unknown spectrum. To start, we chose some of our Challenge FT-IR spectra from previous articles in the "Spectral Interpretation" series. We didn't even have to worry about what instrument format the file was in because FTIRsearch.com reads the broadest range of formats we have seen.

Within seconds of sending our file, the spectrum of our Challenge appeared on the screen. The first choice we were presented with was the libraries we wished to search. Forensic, industry specific, polymer, Sigma/Aldrich condensed phase, and vapor phase are available. We chose the default, which is all the libraries, covering 71,000 FT-IR spectra of known substances. The database also has 16,000 Raman spectra available for searching.

Two varieties of searching are available. The first is a simple spectrum search, where all the parameters are defaulted to their standard settings, letting users search without further decisions. The advanced spectral search allows users to choose the searching algorithm, auto-baseline correction or not, and a region of the spectrum you wish to exclude. We chose the simple spectral search.

The Challenge FT-IR spectra we selected were reference spectra from other providers of spectral libraries. In each case, the correct identification was returned in less than half a minute with hit quality indices in the 0.9 range. Particularly impressive was the speed with which we could zoom in and stack or overlay the spectra of the unknown and library hits, which are color-coded for easy identification. An information button next to each library hit gives the full spectrum of the library compound along with valuable information like Chemical Abstract Service number, formula, weight, other physical properties, and spectrum references. Although we could not find any structures, the information provided in this section is also valuable in helping identify the unknown in situations where the spectral matches are very close for a number of library compounds.

With these successes in hand, we proceed on to more formidable searching.

FT-IR - Comparison of Transmission to ATR recorded spectra

Acetaminophen, or 4-acetamidophenol, is a handy compound to have around the lab and is used extensively. It is a wavelength reference material for Raman spectrometers in ASTM E-1840-96 (Standard Guide for Raman Shift Standards for Spectrometer Calibration) and is useful in this regard to test your instrument's performance if you routinely measure crystalline materials. It also is the main ingredient in any number of pain relievers. It is in many ways a useful compound. It was used here to test the FTIR search capabilities in both the infrared and Raman.

In Figure 1, the 4-acetamidophenol spectra were recorded as a KBr pellet (blue) and using a Durascope diamond ATR (DATR) accessory (SensIR Technologies, LLC, Danbury, CT). FTIRsearch had no trouble identifying the compound. In both cases the acetaminophen was the top hit in most of the search options (see Table I). Note that the hit quality index (HQI) improved significantly in the correlation search and absolute value search, and to a lesser degree in the least squares search.

Figure 1. 4-acetamidophenol spectra as (a) KBr pellet and (b) diamond ATR.
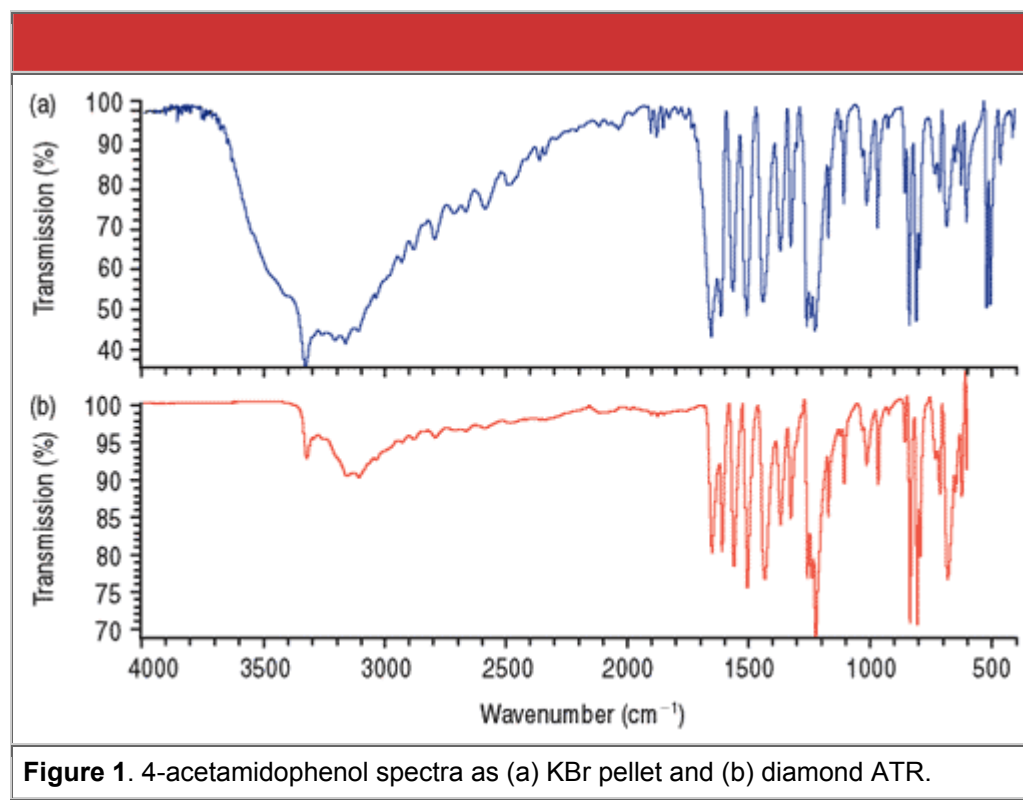
In Table I, we see that the correct match for the DATR spectra was the first hit using five of the seven algorithms, which is as good or better than the KBr pellet spectra (in which four out of the seven algorithms gave the correct match). This comparison indicates that the use of ATR-measured spectra, which may cause slight wavenumber shifts

(because of the ATR effect) compared with transmission spectra, does not seriously affect the ability to get good match results.

Table I. Search results from KBr and DATR spectra of 4-acetamidophenol.				
Search algorithm	**KBr pellet**		**DATR**	
	Match number	**HQI**	**Match number**	**HQI**
Correlation	1	0.725	1	0.583
Least squares	2	0.968	NM	-
Absolute value	11	0.852	NM	-
Euclidean distance	2	0.752	1	0.679
NM indicates no match in the top 25 hit list.				

Raman - Search an unknown contaminate spectrum

The typical use of a search library is the identification of a totally unknown sample. We were faced with identifying several small crystals from a sample. We recorded an FT-Raman spectrum of the sample in a capillary tube. We eliminated any known components in the system by running the reference spectra and comparing our unknown to these reference spectra. No matches were found. At this point, the library search was the best chance. Figure 2 shows the results of the FTIRsearch, which identified our compound as EDTA. This was the fourth hit in the correlation or basic search but was the top hit from both the first derivative correlation and first derivative least square algorithm searches. It is useful to test all the algorithms if the match is not perfect. A close inspection of the reference spectra confirmed the match.
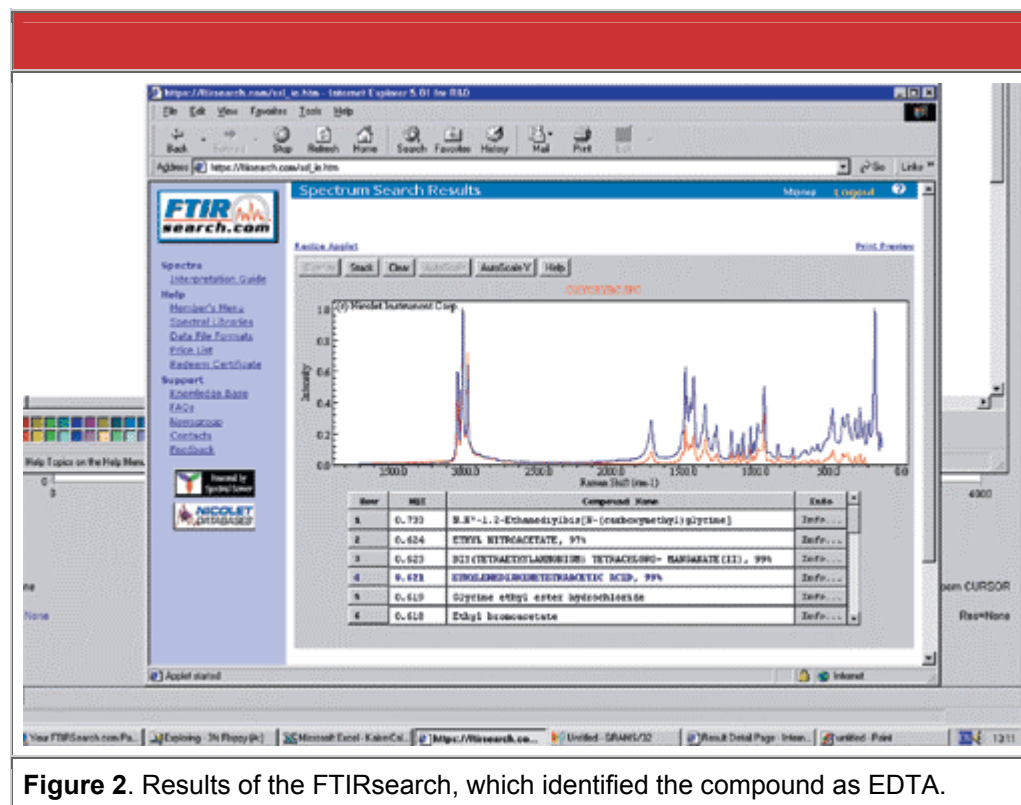

Figure 2. Results of the FTIRsearch, which identified the compound as EDTA.

Raman - Effect of Signal-to-Noise on a search

Returning to acetaminophen, four FT- Raman spectra were recorded at different signal-to-noise levels. The signal-to-noise was measured using the strongest band in the infrared spectrum divided by the noise between 2100-2000 cm^{-1}. The signal-to-noise levels varied from about 84/1 to 5/1.

Figure 3 displays the Raman spectra at the four signal-to-noise levels. Because this compound has many strong vibrations, the matches were number one on the hit list in all cases. Note that in the first derivative correlation search of 4-acetamidophenol in the Raman, there was an improved HQI until the signal-to-noise level deteriorated to the 5/1 level. In both the least square and absolute value searches, even with this significant noise seen in the spectrum, the HQI improves as you use the first derivative search algorithms. This type of improvement would seem to generally hold true especially for spectra having bands that are generally more sharply defined and often less overlapped, like Raman spectra.

Figure 3. Raman spectra of 4-acetamidophenol measured at four different signal-to-noise levels: (a) 5/1, (b) 32/1, (c) 50/1, and (d) 84/1.
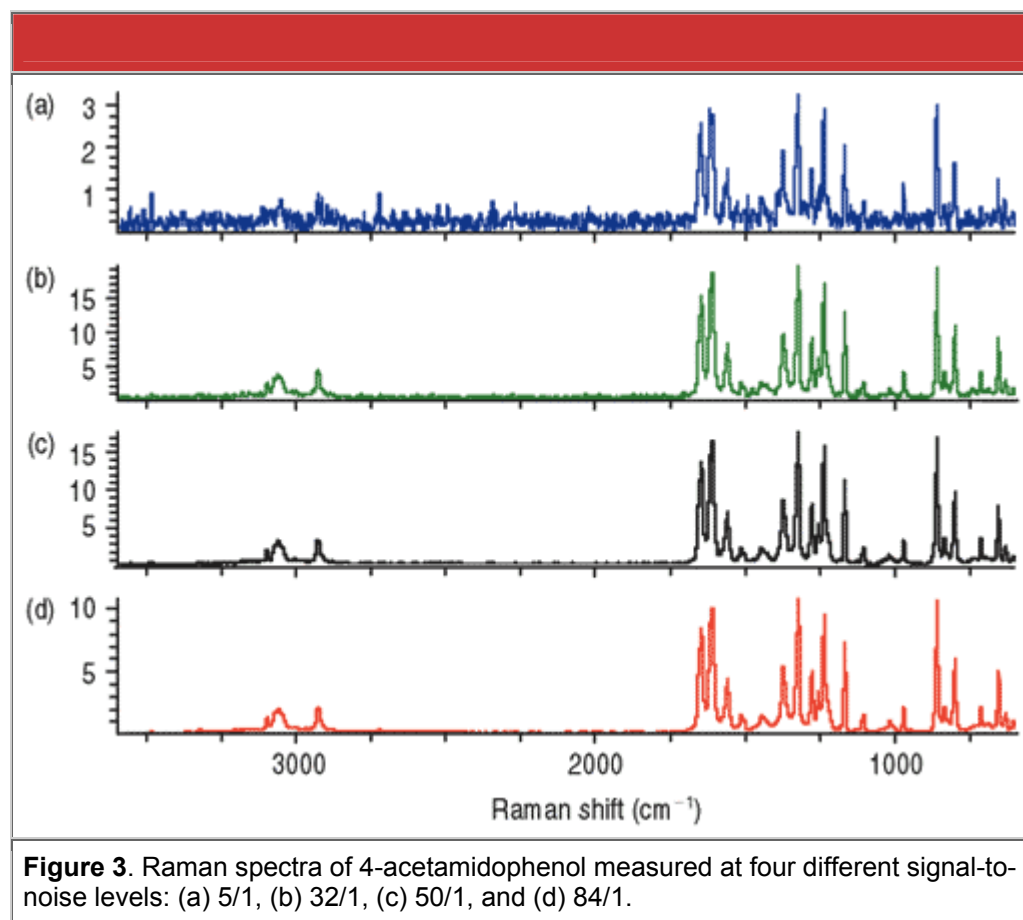
Mixture Spectra - Raman 20% 4-acetamidophenol in Hexane

The searching of pure compound spectra against the database works well, depending on whether the exact match of compound or similar families of compounds are present in the database. Often, when we search an unknown compound we do not know whether the compound is a mixture or a pure compound. In this case the compound was an 80:20 mixture that was 20% 4-acetamidophenol by weight in hexane. This spectrum was measured using a dispersive Raman system using a fiberoptic probe. Again the idea of the measurement was to test the wavelength performance of the Raman spectrometer using a slurry of 4-acetamidophenol in hexane. Figure 4 displays the spectrum of the slurry in hexane and the result of subtracting the hexane from the slurry and the pure hexane spectrum. Table III displays the results of the search of the 80:20 hexane:4-acetamidophenol and the search of the spectrum resulting from subtracting hexane from the mixture.

Table II. Search results from Raman spectra of 4-acetamidophenol measured at four signal-to-noise levels.

Search algorithm	84/1		50/1		32/1		5/1	
	Match #	HQI	Match #	HQI	Match #	HQI	Match #	HQI
Correlation	1	0.952	1	0.958	1	0.948	1	0.798
Least squares	1	0.999	1	0.999	1	0.999	1	0.992
Absolute value	1	0.986	1	0.986	1	0.984	1	0.928
First derivative absolute value	1	0.994	1	0.993	1	0.990	1	0.955
Euclidean distance	1	0.866	1	0.875	1	0.860	1	0.683
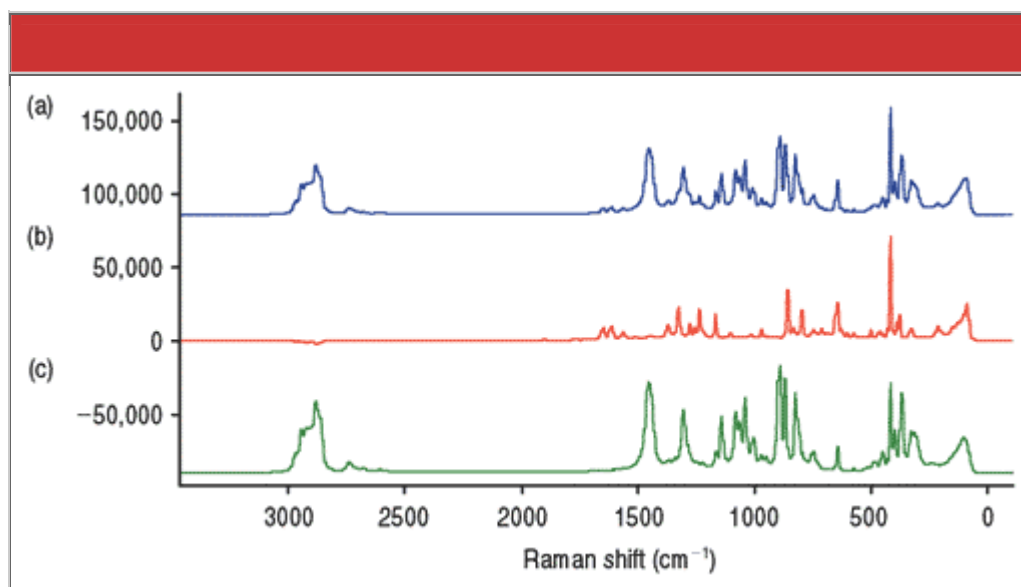
Figure 4. Raman spectra: (a) 80:20 mixture of hexane:4-acetamidophenol, (b) difference spectrum of 80:20 mixture spectrum minus hexane, and (c) hexane spectrum.

The results of the search indicate that presence of the 4-acetamidophenol in the slurry significantly interfered with the simple correlation, least squares, absolute value, and Euclidean searches' ability to identify the main component, hexane. The use of the first derivative version of the searches generally came up with the right compound (hexane) as the first hit. None of the top 25 hits identified the 4-acetamidophenol from the mixture.

To find what was in the solution mix, coauthor Jim Rydzak next subtracted the hexane spectrum from the hexane/4-acetamidophenol mixture spectrum and searched this resulting difference spectrum. This strategy was a good one, as shown in Table III. The 4-acetamidophenol came up as a number-one match on the hit list. As is always the case, the actual spectrum needs to be compared with the spectra displayed from the search to ensure that a good match has been made. When the acetaminophen spectrum from the library was compared with the unknown spectrum, a band was noted at about 418 cm^{-1}. On inspection this was found to be a spurious interference band associated with certain Raman detectors and cosmic events. When this band was eliminated from the spectrum and the resulting spectrum was searched again, the matches were much better, as seen in Table III.

Table III. Result of search of 80:20 mixture of hexane:4-acetamidophenol and the slurry spectrum minus the hexane spectrum.

Search algorithm	80% Hexane		Slurry minus Hexane		Diff. spectrum corrected	
	Match #	HQI	Match #	HQI	Match #	HQI
Correlation	NM	-	1	0.386	1	0.632
Least squares	NM	-	NM	-	3	0.992
Absolute value	NM	-	NM	-	NM	-
Euclidean distance	NM	-	1	0.487	7	0.568

NM indicates no match in the top 25 hit list.

Mixture Spectrum - Identify Major sweetener Component

Another example of a mixture spectrum is a sample of Sweet'N Low sweetener. It lists as ingredients dextrose, calcium saccharin, cream of tartar, and calcium silicate. A particular sample was isolated and recorded using an infrared microscope in transmission mode. The spectrum of this Sweet'N Low particle is seen in Figure 5. Table IV shows the results of this search.
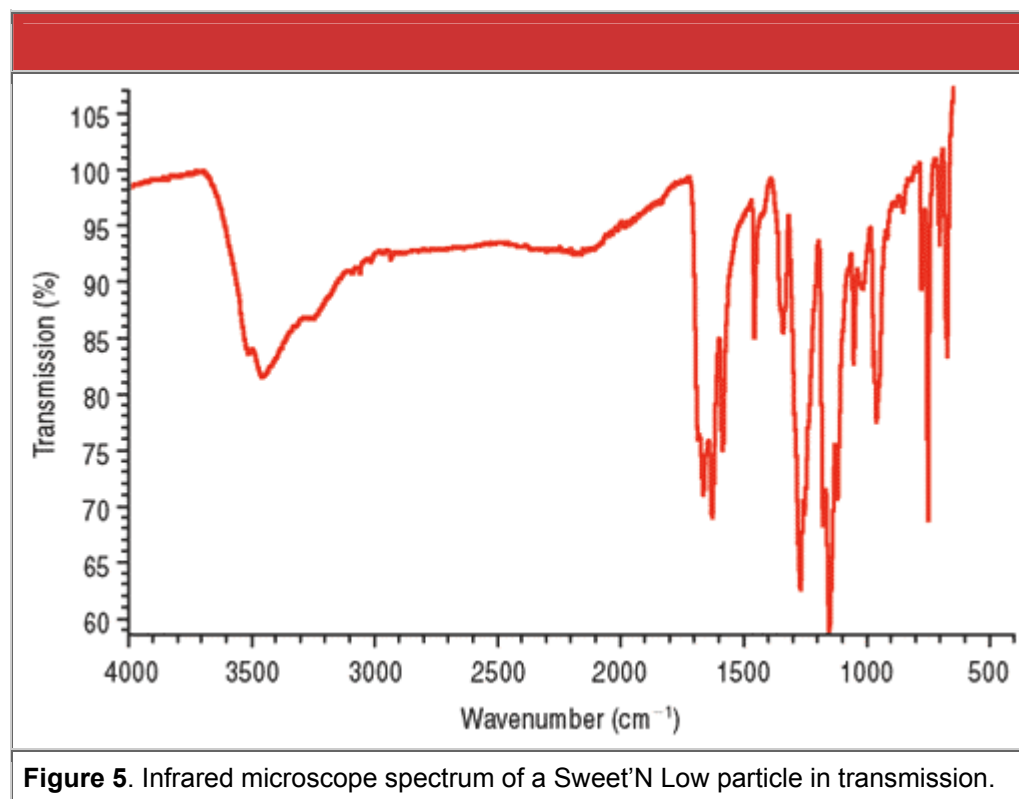

Figure 5. Infrared microscope spectrum of a Sweet'N Low particle in transmission.

| Table IV. Search results for infrared microscope spectrum of Sweet'N Low particle in transmission mode.				
Search algorithm	**Saccharin Na salt**		**Calcium salt**	
	Match #	**HQI**	**Match #**	**HQI**
Correlation	1	0.631	NM	-
Least squares	1	0.988	NM	-
Absolute value	1	0.915	NM	-
Euclidean distance	2	0.715	NM	-
NM indicates no match in the top 25 hit list.				

This search is the best example of the need to look at spectra to determine the best match. In addition to the sodium and calcium salts of saccharin, the library contained a second spectrum of sodium that had large C-H stretching bands as well as a fourth spectrum of sodium saccharin dihydrate (see Figures 6 and 7). Just identifying the compound as sodium saccharin would have been a mistake. After examining all the spectra, no single spectrum would account for all the bands in our spectrum of Sweet'N Low. This would probably indicate that we still have a mixture and would necessitate additional sample preparation. Accurate results for this kind of search would be difficult to conclude definitively.
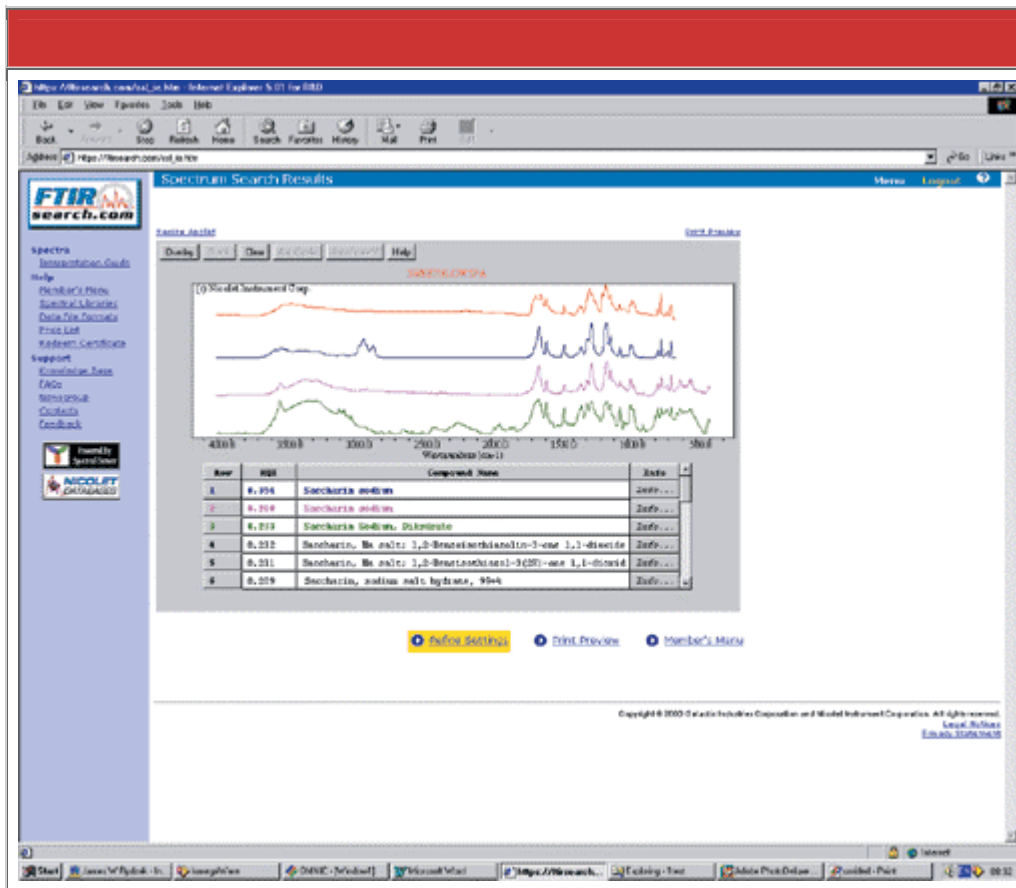
Figure 6. Screen capture of search result for Sweet'N Low using the first derivative correlation search.
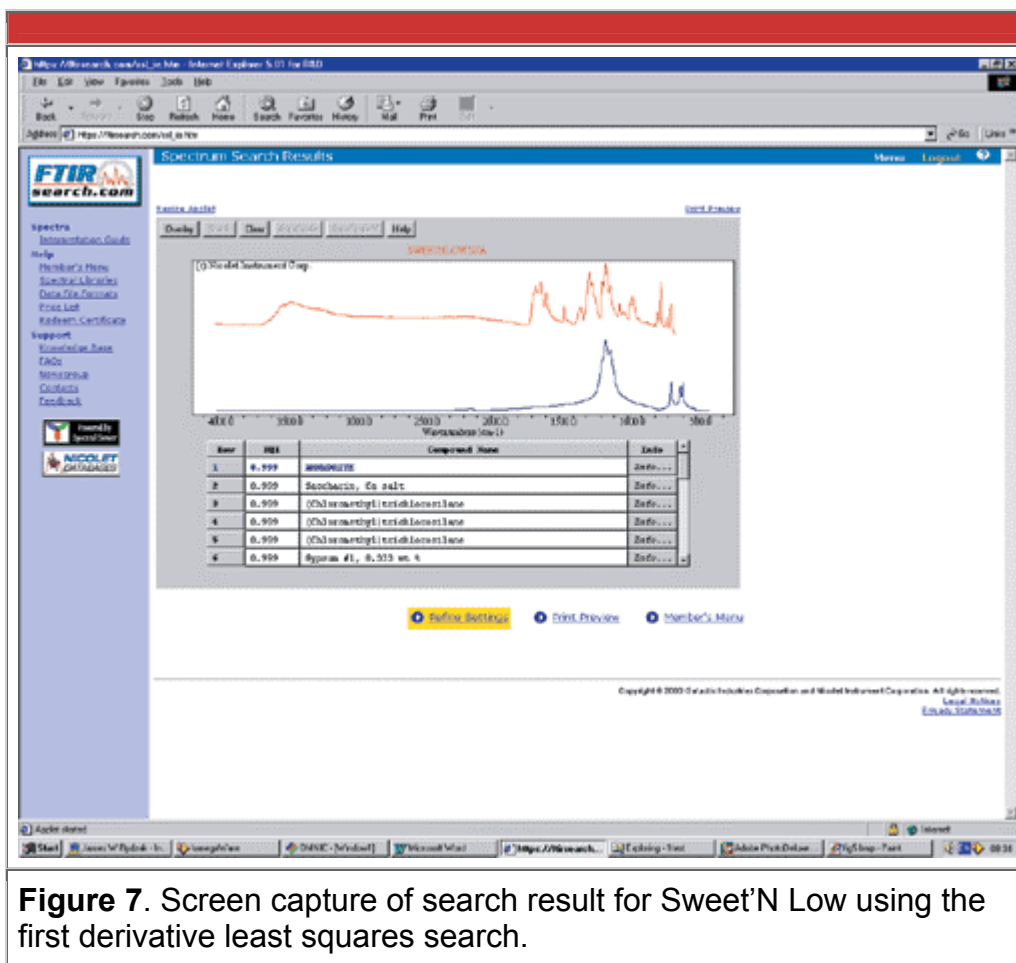
Figure 7. Screen capture of search result for Sweet'N Low using the first derivative least squares search.

Summary

We found this new service to be a very useful tool, particularly for those who only occasionally have the need (or budget) for spectral searching. A crossover obviously exists between the cost of this service and obtaining your own libraries for off-line searching. Even so, this service could be used as an excellent cost justification basis for purchasing off-line libraries.

These experiments do show the need to actively evaluate unknowns beyond the capability of any search. One advantage to the fee structure of this service is that it allows you to refine the search parameters without additional charges; however, refining the unknown spectral data will incur charges. So it is preferable to do as much spectral refinement as possible and practical - such as removal of solvent contributions or separation of mixtures - before spectral searching begins.

Two additional useful features of this service are the spectral interpretation guide and the service's newsletter. The spectral interpretation guide requires the addition of a free plug-in program that can be easily obtained on the Internet. A very useful addition to the interpretation guide that we would like to see made would be the ability to overlay functional group templates with our unknown spectra to visually aid the function group interpretation. The newsletter is sent to the members of FTIRsearch.com and keeps them current on new features of the service and, we hope, tips from other users. One thing we would like to see is an inclusion of the search parameters in the results print-out of the search. This would be especially useful in any retained hard copies of the results. FTIRsearch.com actively solicits input from its users and has already added a number of valuable features based on this exchange of ideas, such as print preview.

There is no question that the Internet age has come to analytical chemistry in general and spectroscopy specifically. By making spectral searching services available to a much broader range of scientists, the field of interpretation has received a tremendous boost. Our caps are off in salute to Galactic Industries Corporation and Nicolet Instrument Corporation for establishing this service to enhance the art and science of interpretation.

Acknowledgments

James W. Rydzak is employed by SmithKline Beecham Pharmaceuticals (King of Prussia, PA) and can be contacted at james_w_rydzak@sbphrd.com. Rydzak is a member of Spectroscopy's Process Analysis Advisory Panel and is active in the use of vibrational spectroscopy in R&D, QA/QC, and process applications. "Spectral Interpretation" Editor **Paul N. Bourassa** is a member of Spectroscopy's Editorial Advisory Board and vice-president of Blue Moon, Inc., a spectroscopy and analytical chemistry consulting service specializing in environmental monitoring. He can be reached by e-mail at paulnb@aol.com.