Search Strategies for IR Spectra - Why Bother Looking Beyond the First Hit? It's the First Hit - It Must Be Right

Bio-Rad Laboratories, Inc., Informatics Division, Philadelphia, PA, USA

Spectroscopy 94038

This Application Note explains why you should always look at more than just the first hit in the search result.

If your database has only one spectrum in it, for example, toluene, then no matter what unknown you search against the database the first hit you get will be toluene. This exaggerated example alone clearly illustrates that the first hit is not always the correct answer. So the first rule of thumb is that the first hit is not always the best hit.

The algorithms used in Bio-Rad's KnowItAll® search software for full spectrum searching all yield a Hit Quality Index (HQI). This HQI measurement is an attempt to rank the results of a search according to the algorithm's determination of how well each database spectrum in the "hit list" or result list matches the unknown spectrum. Every entry in the reference database matches the spectrum searched against it to some degree, whether it is a good or bad match. It is important to keep in mind that the algorithm is pattern matching and does not incorporate factors into the HQI ranking that a seasoned spectroscopist might, such as the history of the unknown sample.

So even if the first hit does look good, you should still look at several more hits to improve the confidence level in the first hit. Looking beyond the first hit can help confirm at least the chemical class of the first hit. If you are using a well-represented reference database (either commercial or proprietary) to search against, you should find several examples within the top 10 or 20 hits that represent the same chemical class or functionality of the first hit. And in some cases, you may even find a better match as you look through the other hits.

It is also possible that several very similar spectra could exist in the hit list. Figure 1 shows the top four spectra of a search result. All of the top four hits are carbonates. If you only looked at the first spectrum in the hit list, you might conclude that the unknown was Manganese Carbonate. If you took the time to look at more than just the first hit, you would see that there are several carbonates in the list.

In this example, differences in the baseline or in the amount of water in the KBr pellet can account for more of a difference than the position of the two small bands around 860 and 700 cm^{-1}.

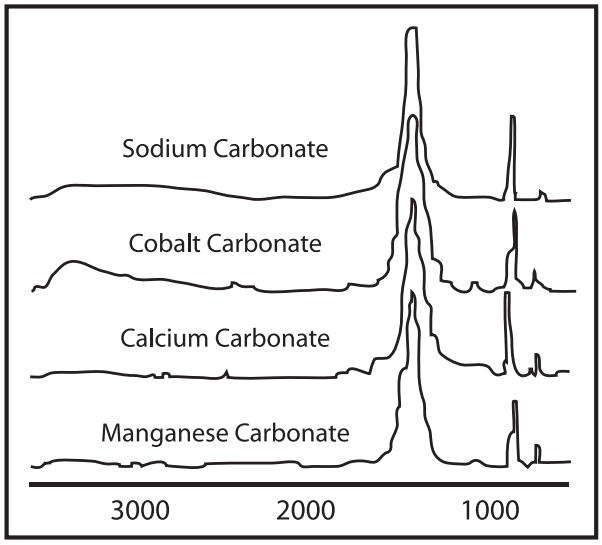

Fig. 1.