KnowItAll Software Training

Predicted IR

Use Predicted IR Spectra in KnowItAll

How to Use Predicted IR spectra in KnowItAll SearchIt and ID Expert

Purpose

These exercises demonstrate how to use Predicted IR spectra in KnowItAll ID Expert and SearchIt

Objectives

These exercises will teach you:

- How to use Predicted IR spectra in KnowItAll SearchIt
- How to use Predicted IR spectra in KnowItAll ID Expert

Background

Wiley has the largest quantity of experimental IR spectra, which was originally the Sadtler IR collection. Even with the continued development in the chemical industry, there is still insufficient progress to cover the newly-discovered chemical space. Sample collection is also challenging, time consuming and costly. Wiley's Library of Predicted Infrared Spectra is an attempt to increase the catalog coverage of Wiley's IR chemical space. This does not mean to increase the chemical space, but to increase the amount of coverage within the bounds of our current libraries' chemical space. With the development of computer modeling technology, the feasibility of using computed IR spectra to compensate for the lack of samples was investigated.

This library can be helpful when used as a tool for searching unknown compounds, illuminating the composition of an unknown spectrum, or accurately predicting the compound's spectrum for the associated structure and functional groups. Due to the predicted nature of this data, it should be mentioned that the results are not perfect. Wiley acknowledges that some results may not be entirely correct in order to convey that this library should be used as a tool to help the user classify components that are in the spectrum to characterize their unknown spectrum.

Training Files Used in This Lesson

- 5MEODIPT.SPC
- 4-(Pyridin-3-yl)-2-2,6,2-terpyridine.irf

The training files used are for example purposes only. The user should utilize their own IR spectra when following this training.

KnowItAll Applications Used

- KnowItAll SearchIt
- KnowItAll ID Expert™

KnowItAll IR and Raman Search Algorithms

A background in the algorithms used by KnowItAll will be beneficial. For IR and Raman spectral comparison, KnowItAll uses the following algorithms:

Correlation

This is the default algorithm for searching in KnowItAll and it conforms to the industry standard for correlation algorithms. The Correlation algorithm is similar to the Euclidean Distance algorithm. The difference between the two is in the way the spectra are treated before the comparison. Each spectrum is mean centered prior to performing the dot product normalization. This approach can improve search results for noisy spectra and spectra that have baseline issues, particularly with a baseline offset that is the result of a negative spike or chemical noise. It is slightly more time-consuming than the Euclidean Distance algorithm. The search speed is slower because each spectrum in the database must be mean centered and then normalized prior to the comparison. The search results that are obtained with the Correlation algorithm are spectrally similar to the unknown, even if the unknown compound is not in the database. The Correlation algorithm is heavily weighted by peak areas. Broad features are weighted much more strongly than sharp features. The algorithm is most tolerant to peak shifts and non-linearities in relative band intensities.

Correlation (Classic)

The Correlation algorithm that was found in all versions of KnowItAll prior to KnowItAll 2020 is similar to the Euclidean Distance algorithm. However, it did not conform to the industry standard for correlation algorithms. Beginning with KnowItAll 2020, the Correlation algorithm does conform to the industry standard and it is the default algorithm used for searching in KnowItAll. To provide backward compatibility for customers who want to reproduce prior search results, the previous correlation algorithm is now provided as Correlation (Classic).

Euclidean Distance

The Euclidean Distance algorithm measures the point-to-point differences between a pair of spectra. The results that are obtained with the Euclidean Distance algorithm are spectrally similar to the unknown, even if the unknown compound is not in the database. This algorithm, however, can yield degraded search results when the unknown spectrum has a sloping or offset baseline. The Euclidean Distance algorithm is heavily weighted by peak areas. Broad features are weighted much more strongly than sharp features. This algorithm is most tolerant to peak shifts and non-linearities in relative band intensities.

First Derivative Euclidean Distance

Use this algorithm to reduce the effects of baseline slope or offset in the unknown. Although search speed is slightly slower than with the Euclidean Distance algorithm, the First Derivative Euclidean Distance sometimes gives improved search results, especially when the unknown spectrum is a mixture of two or more compounds. The First Derivative Euclidean Distance algorithm is heavily weighted by changes in slope. Sharp features are weighted much more strongly than broad features. The algorithm is also very sensitive to peak shifts. Small shifts can make the algorithm miss a similar result.

Second Derivative Euclidean Distance Use the Second Derivative Euclidean Distance algorithm to compare the second derivative of a reference spectrum to that of the query spectrum.

Optimized Corrections: A Breakthrough Technology for Spectral Searching

Spectral searching is one of the most important tools researchers use to classify or identify materials, yet it continues to be plagued by errors and imperfections. During a spectral search, a sample spectrum is compared to a database of reference spectra. To ensure that an optimal match is found in the database, spectra can be adjusted to compensate for differences between spectra caused by variability in instruments, accessories, environmental conditions, and other factors.

According to ASTM's guide on spectral searching[1], various algorithms and manual methods exist to adjust spectra to get reasonable match scores when two compared spectra of the same compound differ for various reasons. While these methods may work in selected cases, subtle discrepancies such as a shift of the X-axis are very hard to identify and correct manually. The inflexible mathematical algorithms typically employed do not compensate for these types of errors in spectra that are flawed.

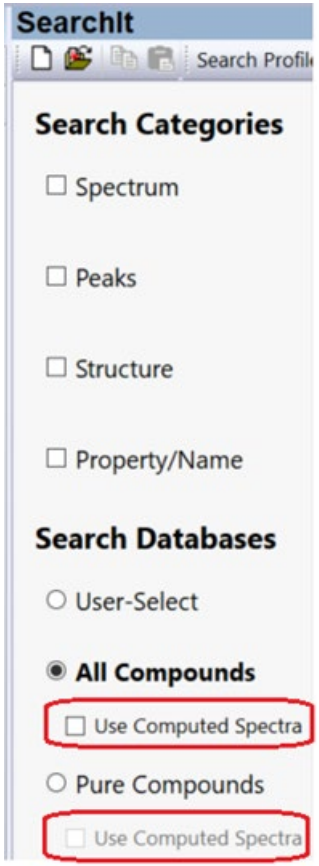
Manual corrections can be made by expert spectroscopists, but those less experienced in spectroscopy are often unaware of how to perform the necessary corrections on their sample spectrum to achieve the best search result. To address this growing concern, Wiley has introduced Optimized Corrections, a breakthrough patented technology that performs a computationally complex set of multiple corrections on query and reference spectra in a search to find the optimal match between the query and each individual reference spectrum. This training guide will demonstrate how the Optimized Corrections technology yields better matches between query and reference spectra than can be attained using rigid search algorithms alone or with manual methods to optimize spectra for searching.

Optimized Corrections consider the full spectrum during a selected range(s) search.

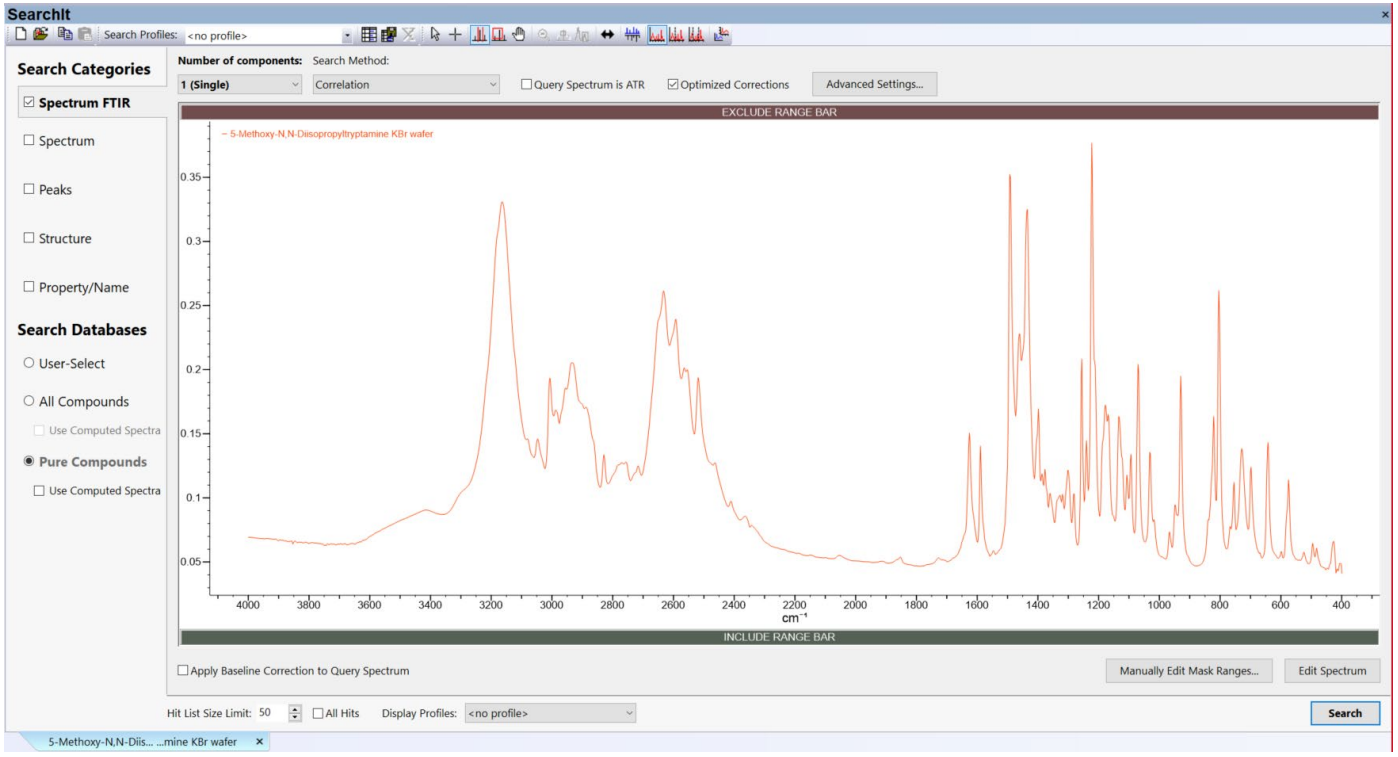
[1] E2310-04 - Standard Guide for Use of Spectral Searching by Curve Matching Algorithms with Data Recorded Using Mid-Infrared Spectroscopy, 2009. ASTM International Web Site. http://www.astm.org/Standards/E2310.htm (accessed March 4, 2015).

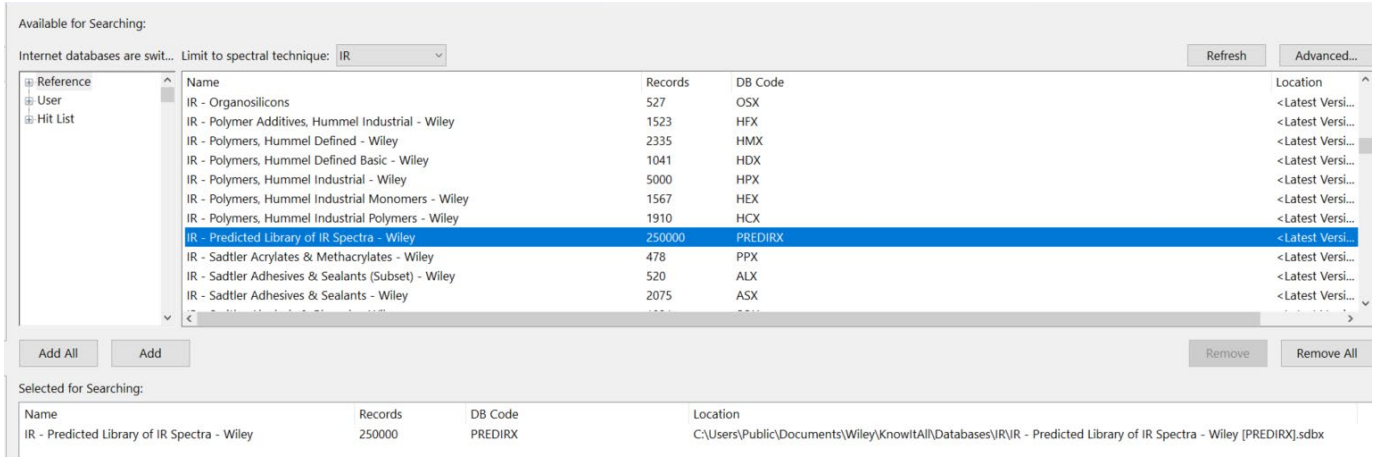
SearchIt

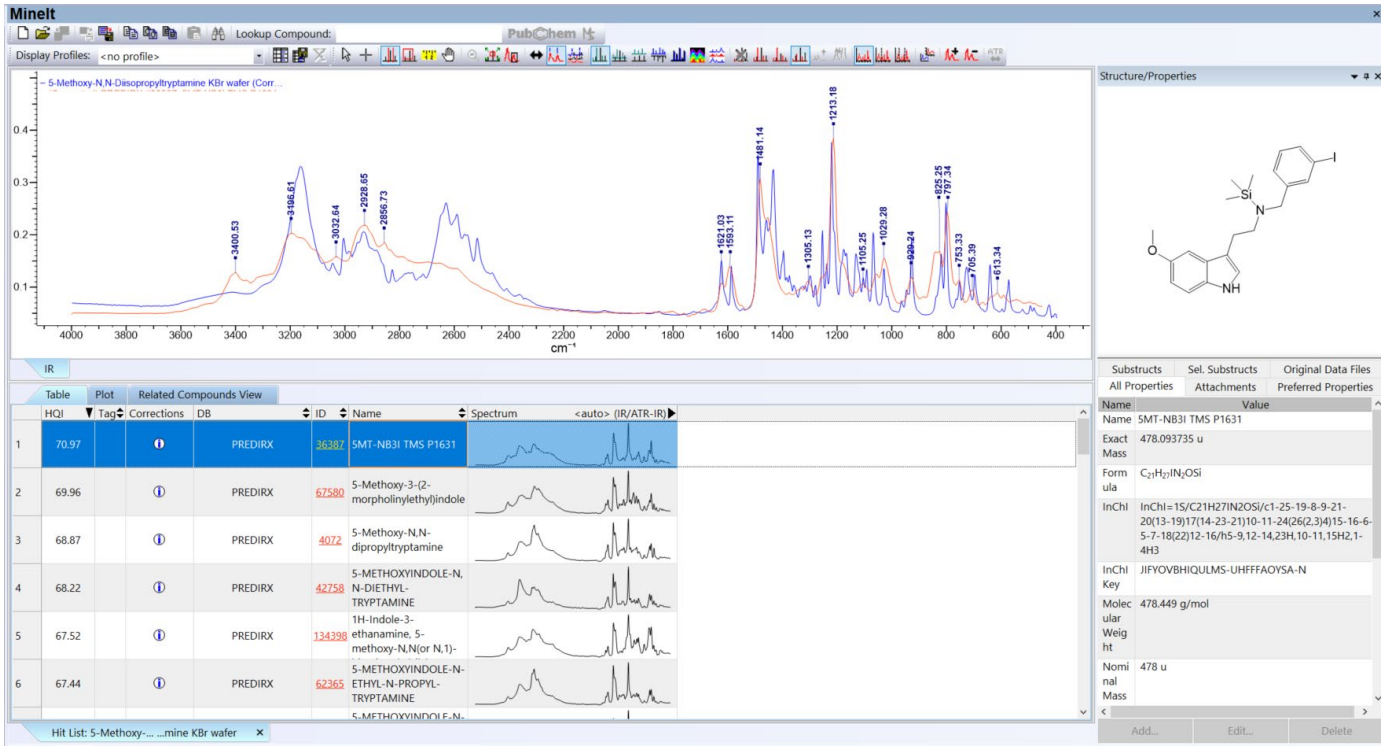
When a license contains the subscription of predicted IR spectra, it can be accessed in the **SearchIt** application **> Databases > All Compounds** or **Databases > Pure Compounds** by selecting the **Use Computed Spectra** option:



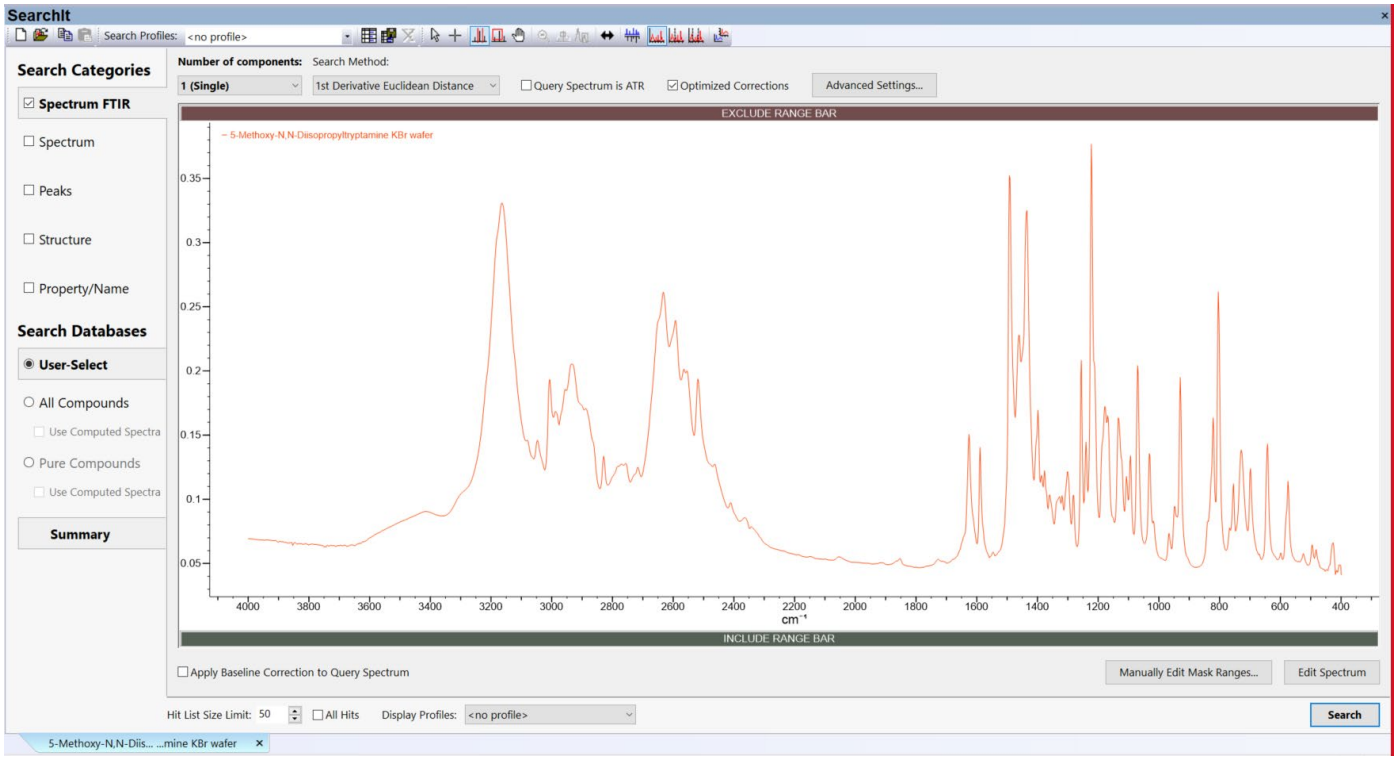
Example 1 - 5MEODIPT.SPC, Correlation Algorithm

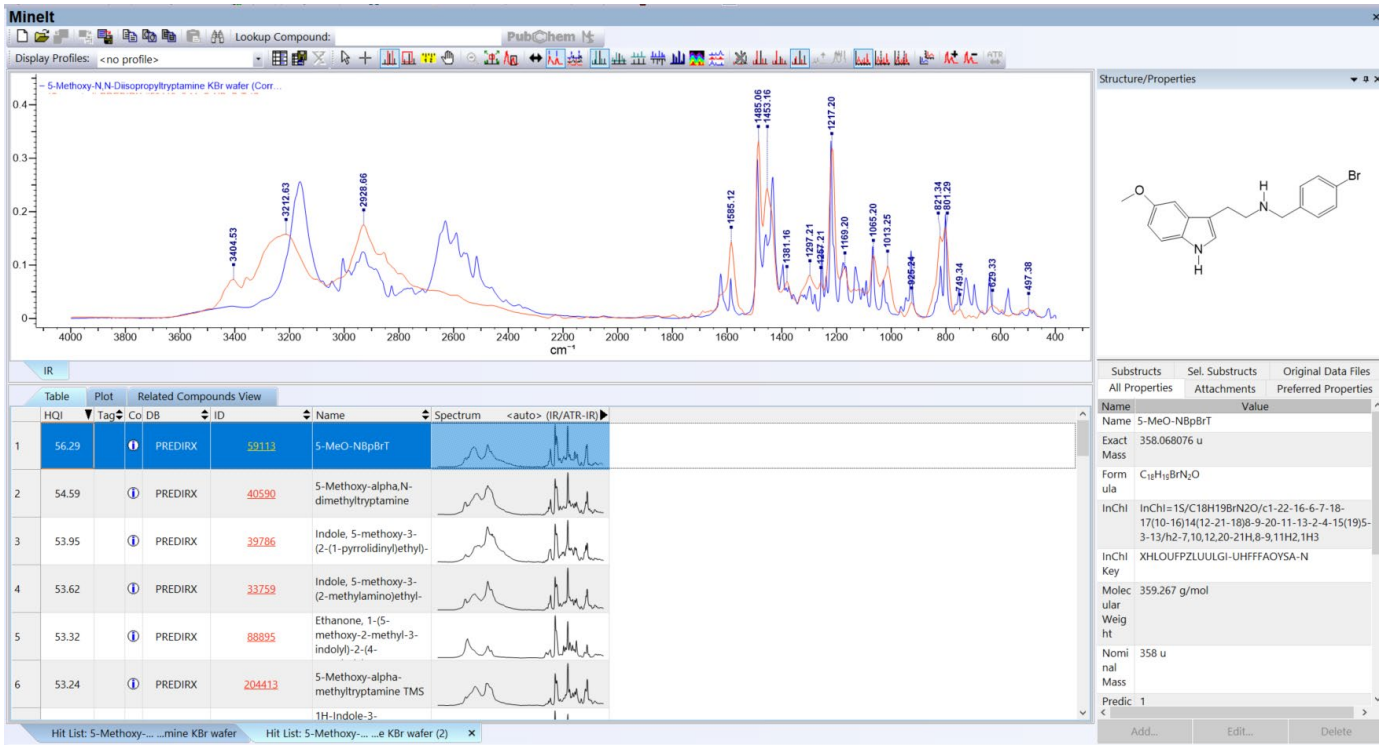
	Action	Result
1	In the **SearchIt** application, click **Open Spectrum or Structure.**	
2	Navigate to find **5MEODIPT.SPC** in the **Predicted IR Example Spectra** folder. (Note that the training files used in this document are for example purposes only. The user should utilize their own IR spectra when following this training.) **Open**	

	Action	Result
3	Click on **User-Select** Click **Remove All** button Limit spectra technique to IR Select the **IR – Predicted Library of IR Spectra – Wiley** and click **Add.** **Search**	

	Action	Result
4	Locate the hit list to view predicted results.	 The similarity in the backbone across the top 10 hits indicates that KnowItAll matched a class of compound.

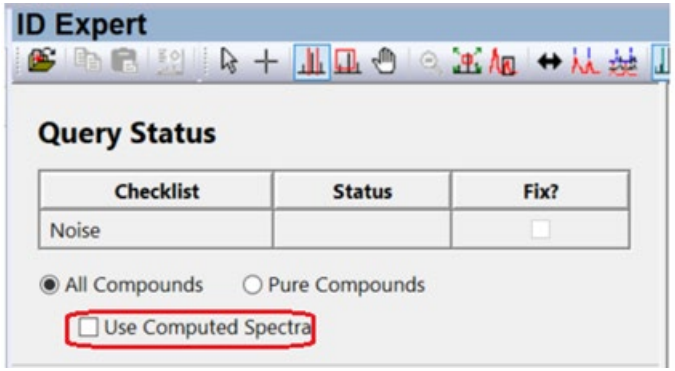
Example 2 - 5MEODIPT.SPC, First Derivative Algorithm

	Action	Result
1	(This is a continuation of Example 1.) Go back to the **SearchIt** application. Reopen the file by clicking **Open Spectrum or Structure > 5MEODIPT.SPA > Open > Start a new search.** Switch **Search Method** to **1st Derivative Euclidean Distance.** Click **Search.**	

	Action	Result
2	Locate the hit list to view predicted results.	 The similarity in the backbone across the top 10 hits indicates that KnowItAll matched a class of compound.

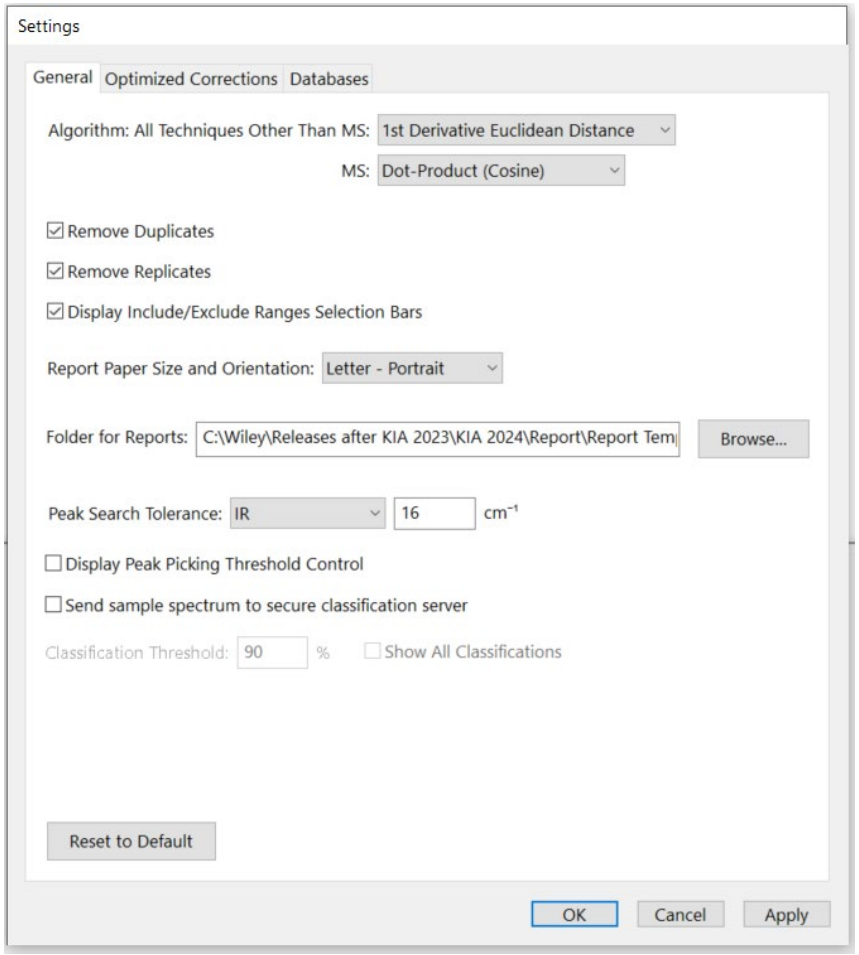
ID Expert

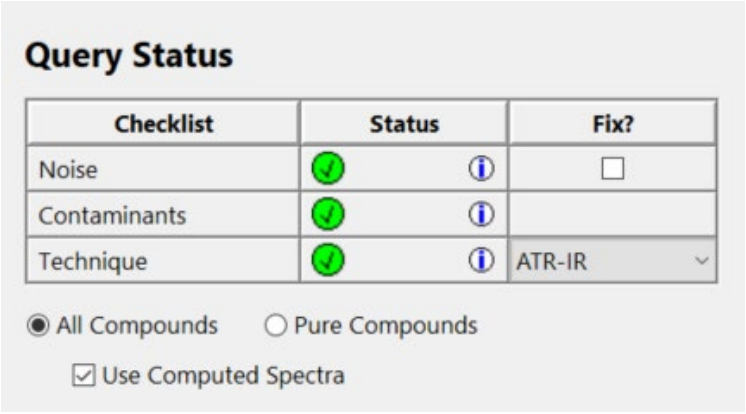
When a license contains the subscription of predicted IR spectra, **ID Expert** interface **Query Status** shows **Use Computed Spectra** option:

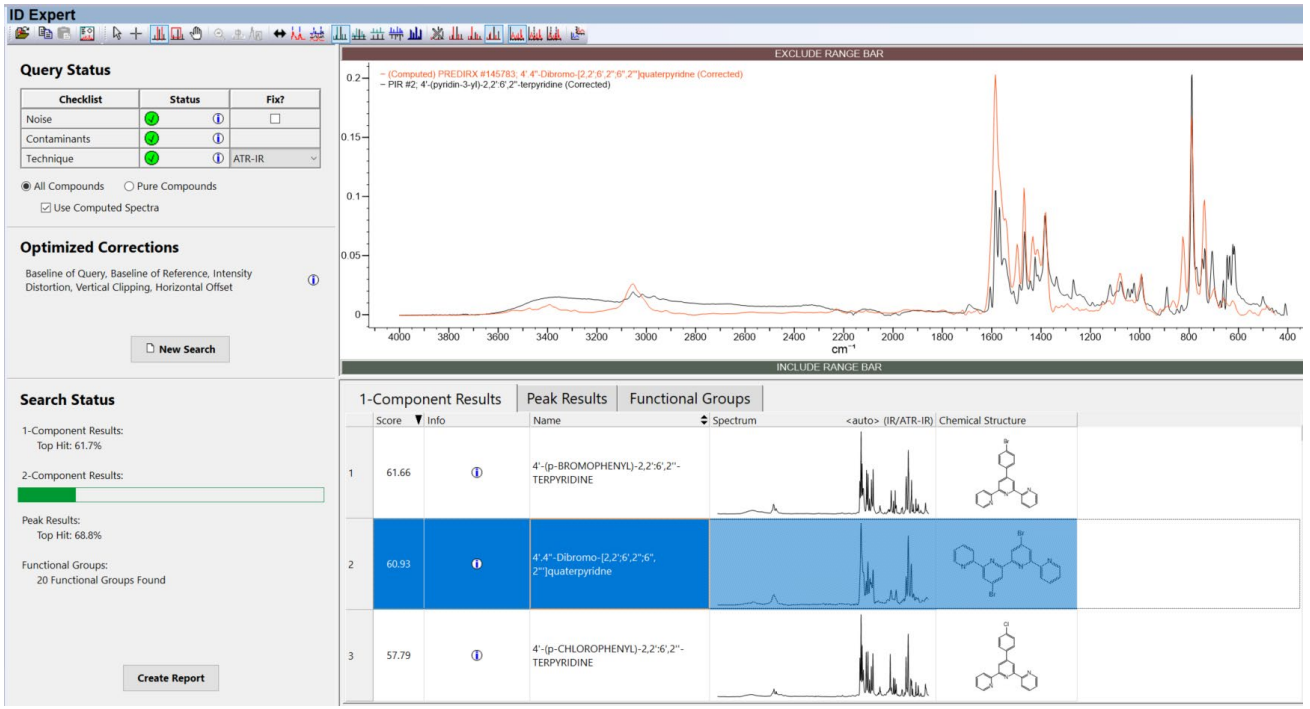


Example 3 - 4-(Pyridin-3-yl)-2-2,6,2-terpyridine.irf

	Action	Result
1	Navigate to the **Data** toolbox and open the **ID Expert** application by clicking the **ID Expert** icon. Alternatively, if the desktop (standalone) application is installed, **ID Expert** can be opened directly by double-clicking on the desktop icon.	The application opens and a Windows **Open** dialog box displays.

	Action	Result
2	Close the **Open** dialog box, then choose **File > Settings**. Set **Algorithm: All Techniques Other Than MS: 1st Derivative Euclidean Distance** **OK**	The **Settings** form opens. 

	Action	Result
3	**New Search** Open **4-(Pyridin-3-yl)-2-2,6,2-terpyridine.irf** in the **Predicted IR Example Spectra** folder (Note that the training files used in this document are for example purposes only. The user should utilize their own IR spectra when following this training.) **Open**	
4	Set **Query Status Technique** to **ATR-IR.** Check **Use Computed Spectra**	

	Action	Result
5	View search results in the Component Table.	 This search included experimental spectra and predicted ones. As we can see, both arrives to similar chemical class.