KnowItAll Software Training

Mixture Analysis

Mixture Analysis

How to Analyze Mixture Spectra

Purpose

This exercise demonstrates how to perform a mixture analysis using the KnowItAll Informatics System's SearchIt application.

Objectives

This exercise will teach you:

- How to configure a mixture analysis
- How to interpret the results of a mixture analysis

Background

The spectral analysis of mixtures in experimental data is a challenging task. Manual separation of spectral components, even when they are known in advance, is a tedious job. Attempting to do this analysis in an automated fashion creates a whole new level of challenges.

This chapter introduces how to use the SearchIt application to perform Mixture Analysis.

Training Files Used in This Lesson

C:\Users\Public\Documents\Wiley\KnowItAll\Samples\Mixture Analysis\IR Examples

- Mixture of Two Steroids – ATR-IR.irf

KnowItAll Applications Used

- SearchIt
- MineIt

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.

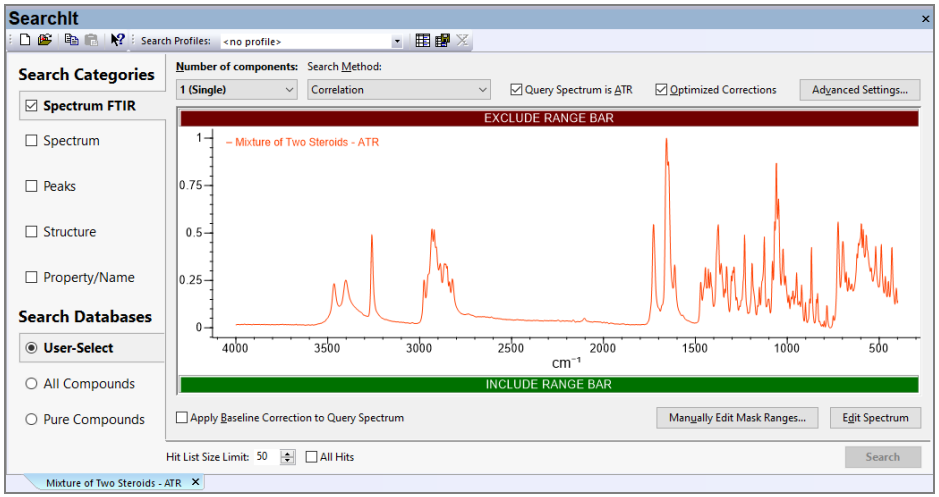
According to ASTM's guide on spectral searching[1], various algorithms and manual methods exist to adjust spectra to get reasonable match scores when two compared spectra of the same compound differ for various reasons. While these methods may work in selected cases, subtle discrepancies such as a shift 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.

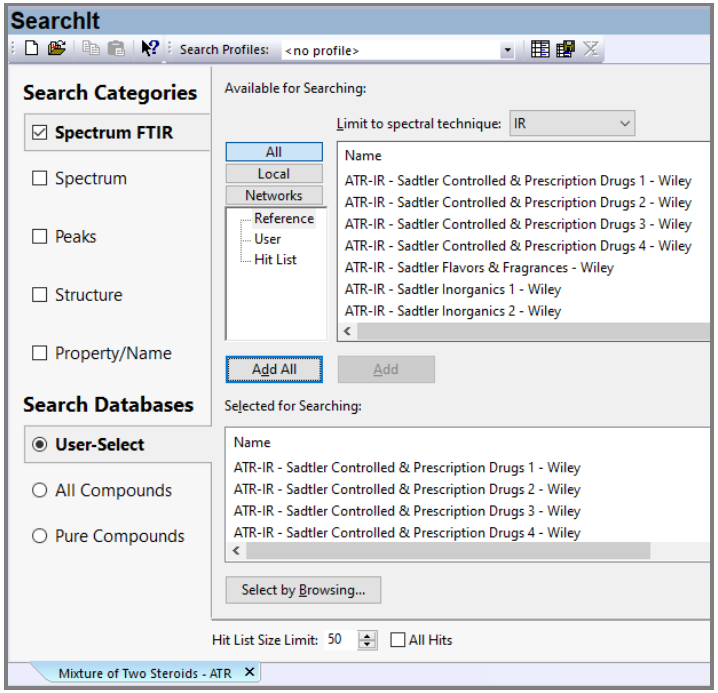
Manual corrections can be made by expert spectroscopists, but those less experienced in spectroscopy are often unaware of how to perform the necessary corrections on their sample spectrum to achieve the best search result. To address this growing concern, Wiley has introduced Optimized Corrections, a 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.

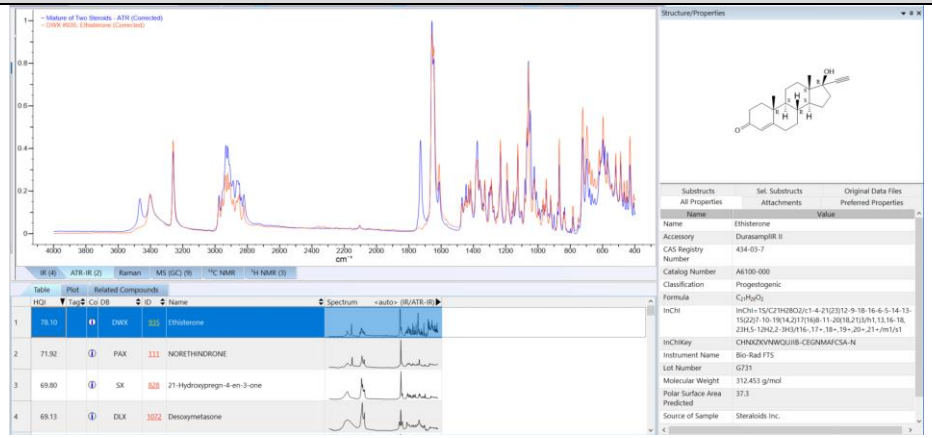
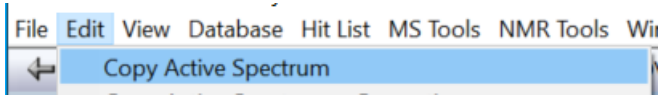
Optimized Corrections consider the full spectrum during a selected range(s) search.

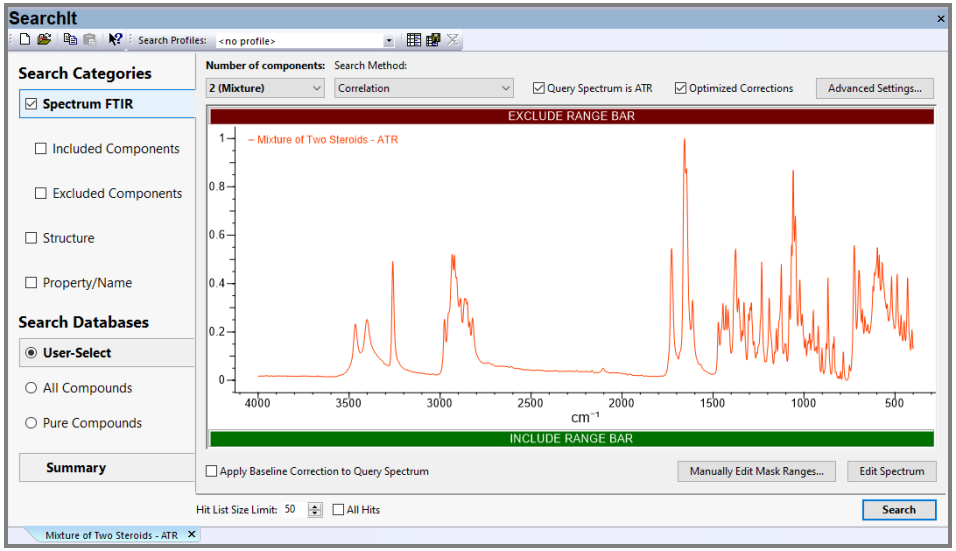
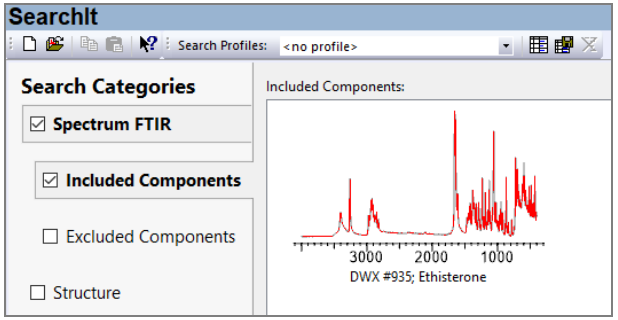
[1] E2310-04 - Standard Guide for Use of Spectral Searching by Curve Matching Algorithms with Data Recorded Using Mid-Infrared Spectroscopy, 2009. ASTM International Web Site. http://www.astm.org/Standards/E2310.htm (accessed March 4, 2015).

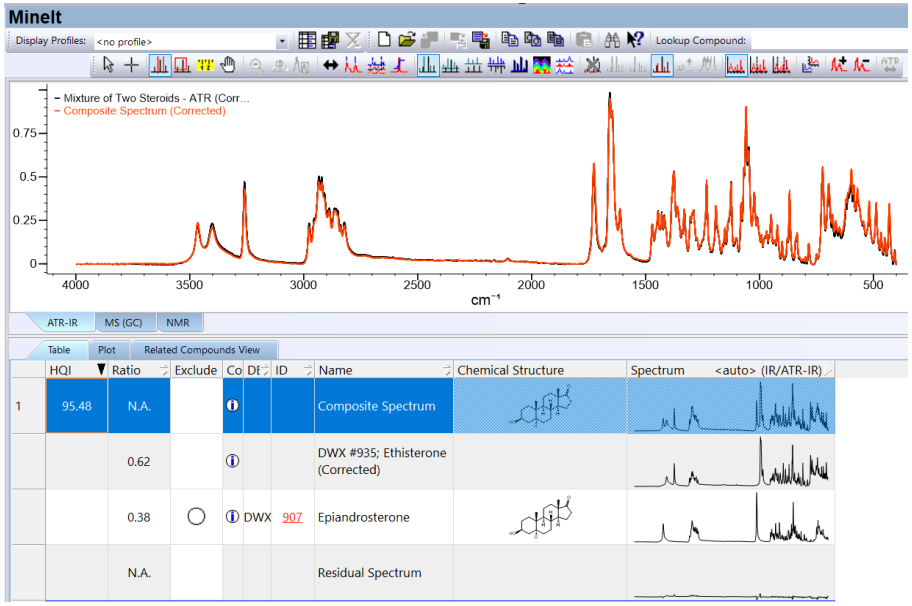
A typical mixture analysis workflow

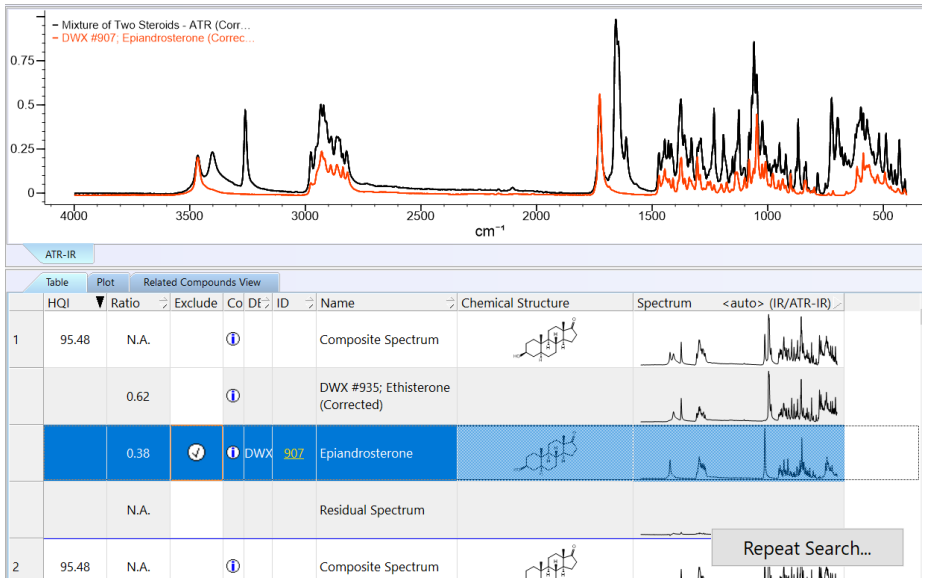
	Action	Result
1	Start **KnowItAll** Navigate to the **Data** toolbox and open the **SearchIt** application. Check **Spectrum**, and in the resulting **Open** dialog box, navigate to **C:\Users\Public\Documents\Wiley\KnowItAll\Samples\Mixture Analysis\IR Examples** Open **Mixture of Two Steroids - ATR-IR**. Set **Search Method** to **Correlation**.	

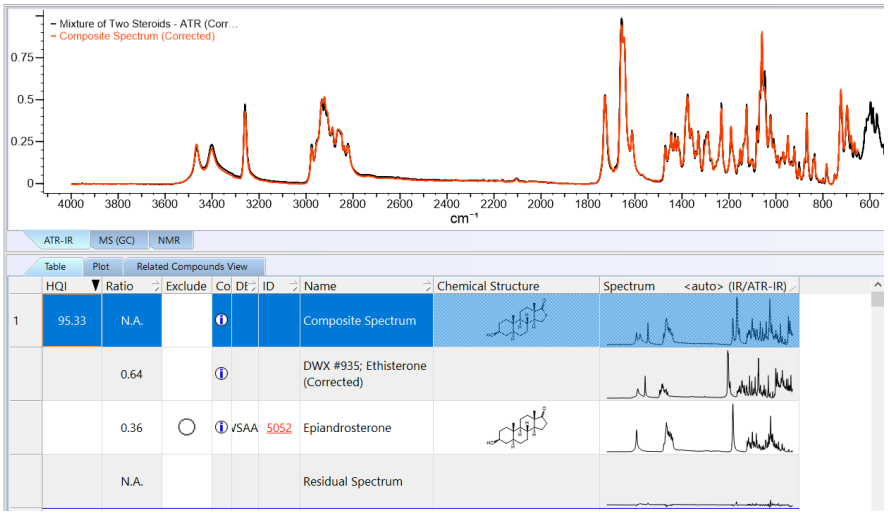
	Action	Result
2	Click on **User-Select** under **Search Databases**. Set **Limit to spectral technique** to **IR**. Click **Add All** at the bottom of **Available for searching** menu.	
3	Click **Search.**	One component search result returns to the **MineIt** application.

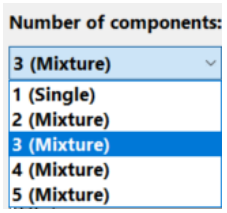
	Action	Result
		
4	Highlight the first hit. From the **Edit** menu, select **Copy Active Spectrum**.	

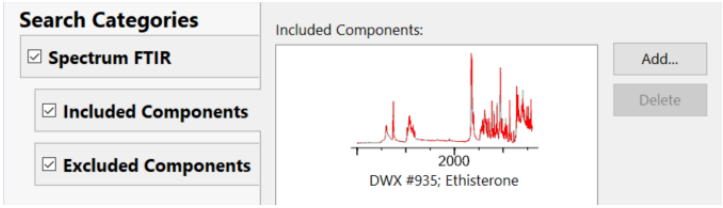
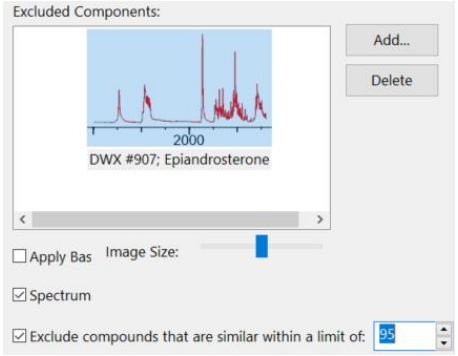
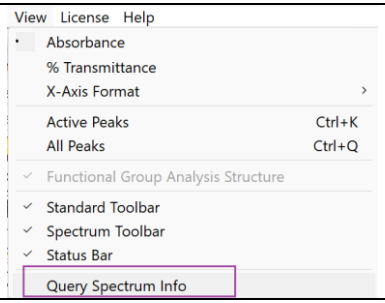
	Action	Result
5	Go back to **SearchIt**, Click the **Spectrum FTIR** button to bring up the query spectrum. Update the **Number of components** to **2 (Mixture)**.	 The **Included Components** and **Excluded Components** checkboxes show up now.
6	Check the **Included Components**. Paste the copied spectrum.	

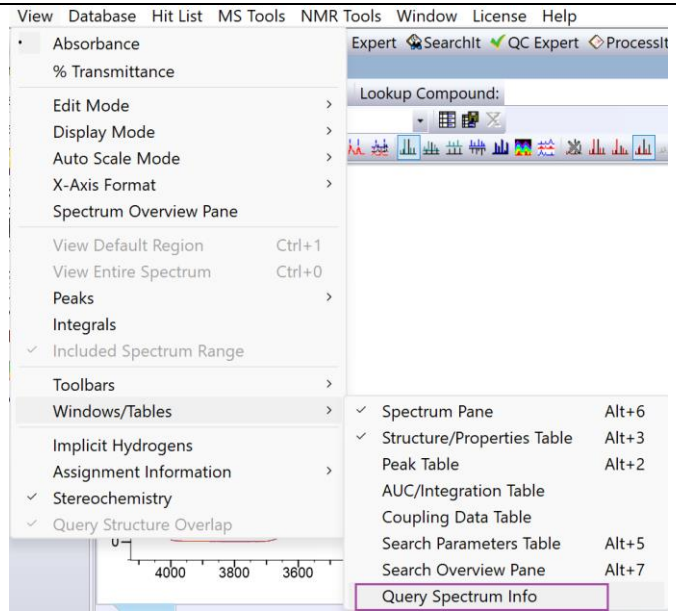
	Action	Result
7	Click the **Spectrum FTIR** button to bring up the query spectrum. Click **Search**.	 Now we have a good two-component match. **Notes**: Each composite spectrum (row 1) is accompanied by the individual component spectra (middle rows) that comprise the composite, as well as the residual spectrum (last row)–the difference between the query spectrum and the composite. The composite spectra are ranked by how closely they resemble the query spectrum. A relatively flat residual spectrum indicates that the software has correctly identified the individual components of the mixture. The **Weight** value for each component spectrum indicates how much it contributes to the composite spectrum.

	Action	Result
8	The circle in the **Exclude** column can be checked to exclude a component from consideration. Check to exclude **DWX 907**.The **Repeat Search** button shows up for user to repeat Mixture Analysis without considering DWX 907.	 **Note**: In KnowItAll 2023 release, “Weight” column is labeled as “Ratio” as this value is the ratio of a component spectrum curve.

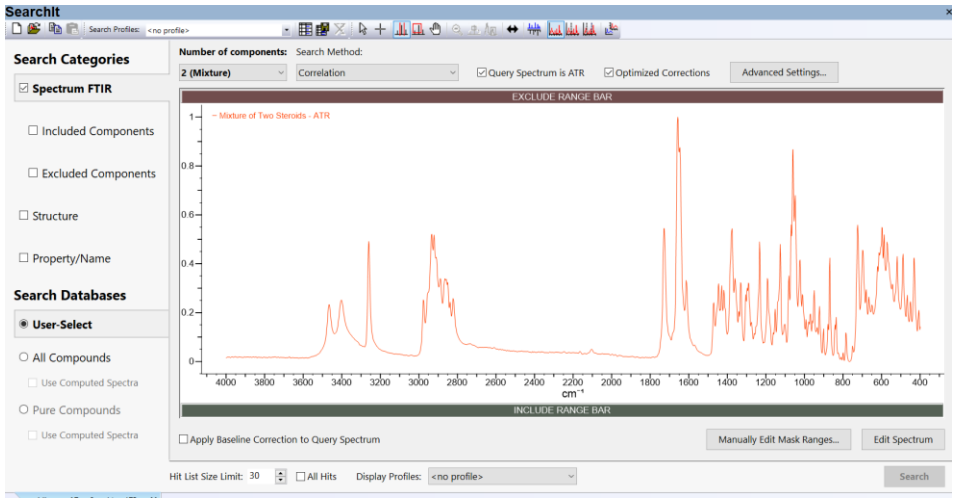
	Action	Result
9	Click **Repeat Search.**	The new result does not contain excluded DWX 907. 

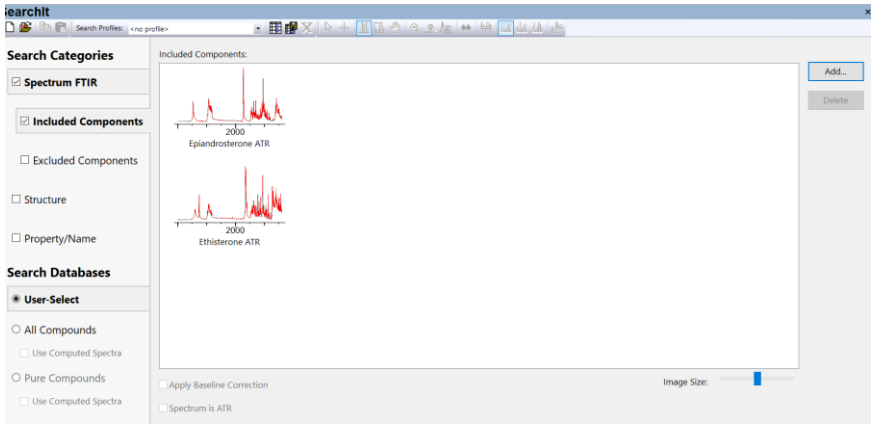
Notes:	
• You do not have to specify **Included** or **Excluded** components to perform a mixture analysis. You can simply open a spectrum in **SearchIt** and specify **Number of components** to a value larger than 1.	

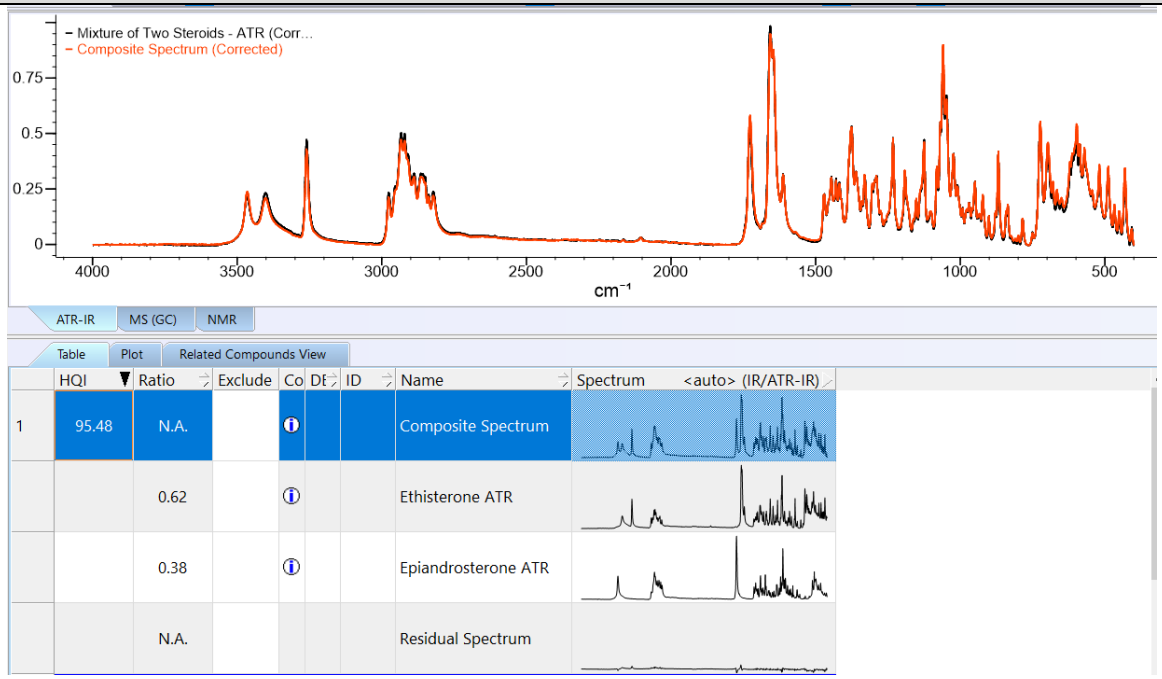
• You can add **Included** or **Excluded** components from files by clicking the **Add** button.	
• You can exclude spectrally similar records.	
One can view the metadata of query spectrum by • **View > Query Spectrum Info** in **SearchIt** and • **View > Windows/Tables > Query Spectrum Info** in a **MineIt** hit list.	



Use Included Components

	Action	Result
1	In the **Data** toolbox, open the **SearchIt** application. Check **Spectrum**. In the resulting **Open** dialog box, navigate to **C:\Users\Public\Documents\Wiley\KnowItAll\Samples\Mixture Analysis\IR Examples** folder. Open **Mixture of Two Steroids - ATR-IR**. Set **Search Method** to **Correlation**. Set **Number of components** to **2 (Mixture)**.	

	Action	Result
2	Check **Included Components**. **Add** the following files from **C:\Users\Public\Documents\Wiley\KnowItAll\Samples\Mixture Analysis\IR Examples\Components** folder: • **Epiandrosterone ATR-IR** • **Ethisterone ATR-IR** **Note**: Use the Ctrl key to select multiple files in the **Open** dialog box.	
3	Click **Search**. **Note**: Toggle between different spectra display modes by selecting available options from **View** > **Display Mode**.	

	Action	Result
		 In this case, KnowItAll simply returns possible combinations of two provided components without going through a database search.