

# **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
- Minelt

## 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.

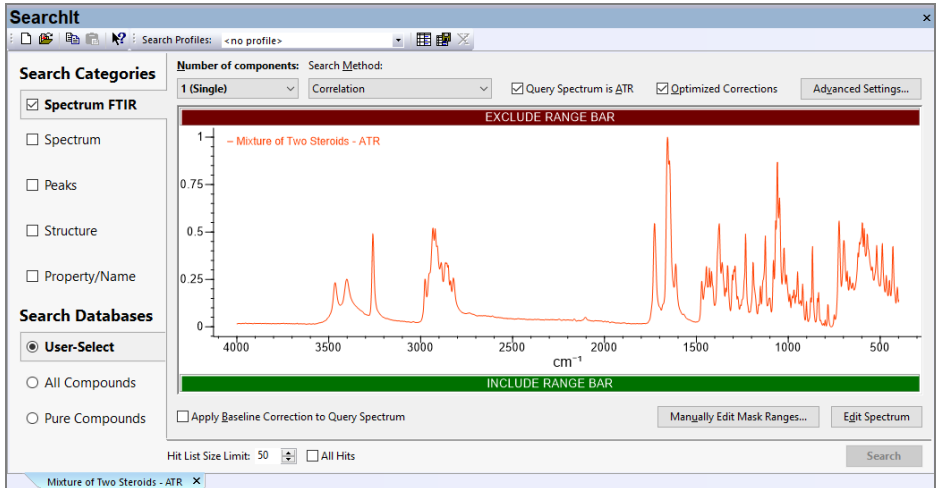
According to ASTM's guide on spectral searching<sup>1</sup>, various algorithms and manual methods exist to adjust spectra to get reasonable match scores when two compared spectra of the same compound differ for various reasons. While these methods may work in selected cases, subtle discrepancies such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

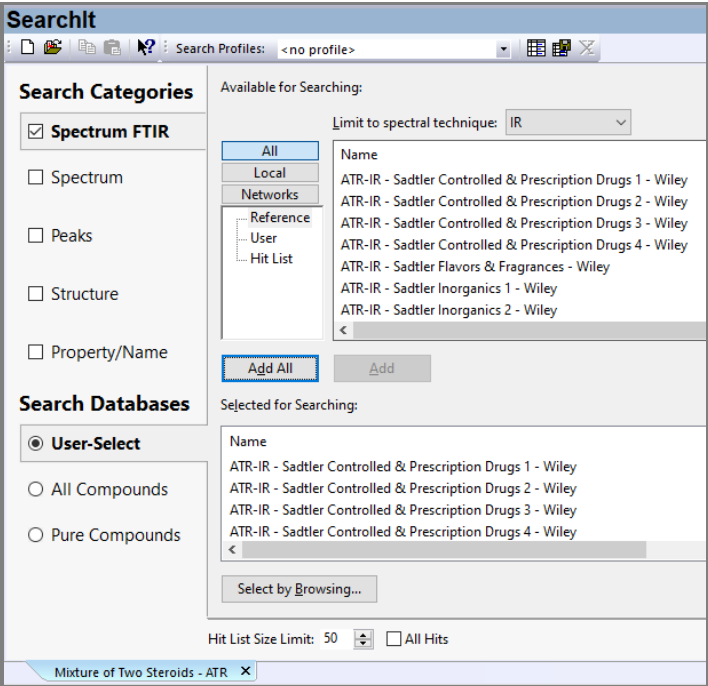
Manual corrections can be made by expert spectroscopists, but those less experienced in spectroscopy are often unaware of how to perform the necessary corrections on their sample spectrum to achieve the best search result. To address this growing concern, Wiley has introduced Optimized Corrections, a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

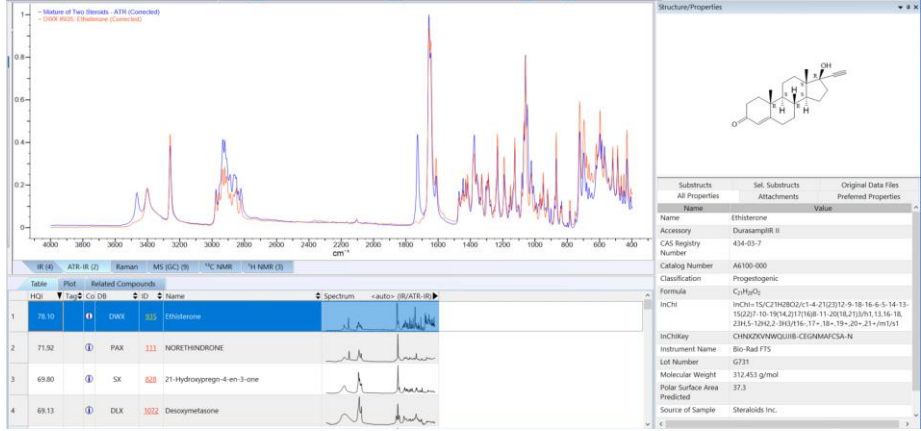
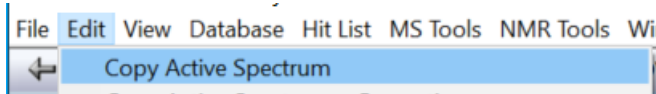
Optimized Corrections consider the full spectrum during a selected range(s) search.

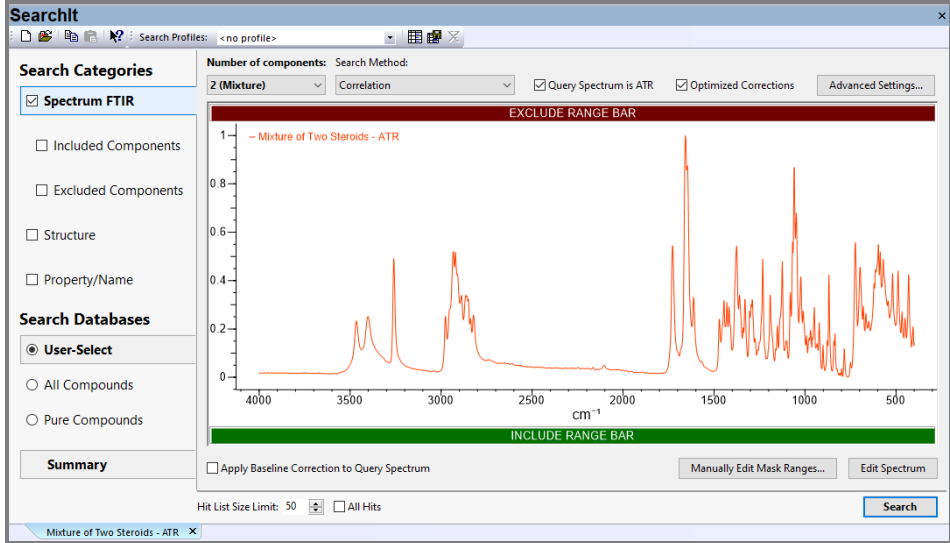
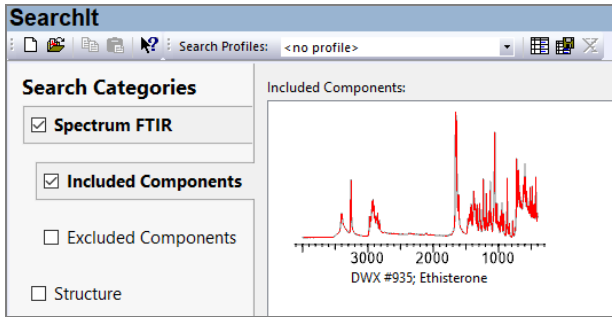
<sup>1</sup> E2310-04 - Standard Guide for Use of Spectral Searching by Curve Matching Algorithms with Data Recorded Using Mid-Infrared Spectroscopy, 2009. ASTM International Web Site. <http://www.astm.org/Standards/E2310.htm> (accessed March 4, 2015).

## A typical mixture analysis workflow

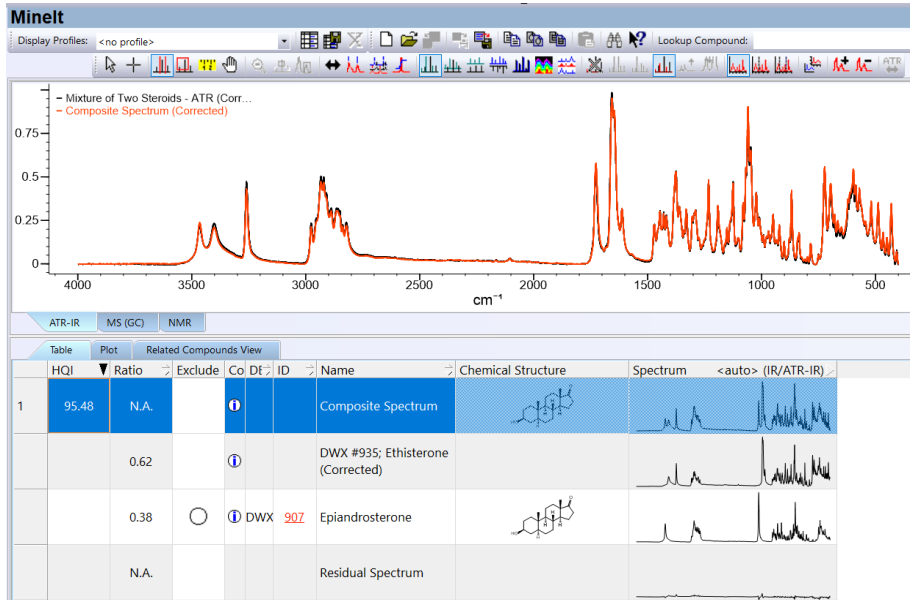





















	Action	Result
1	<p>Start <b>KnowItAll</b></p> <p>Navigate to the <b>Data</b> toolbox and open the <b>SearchIt</b> application.</p> <p>Check <b>Spectrum</b>, and in the resulting <b>Open</b> dialog box, navigate to <b>C:\Users\Public\Documents\Wiley\KnowItAll\Samples\Mixture Analysis\IR Examples</b></p> <p>Open <b>Mixture of Two Steroids - ATR-IR</b>.</p> <p>Set <b>Search Method</b> to <b>Correlation</b>.</p>	 <p>The screenshot displays the SearchIt application window. The 'Search Categories' section on the left has 'Spectrum FTIR' checked. The 'Search Method' is set to 'Correlation'. The main plot area shows an FTIR spectrum with a y-axis from 0 to 1 and an x-axis from 4000 to 500 cm<sup>-1</sup>. The spectrum is labeled 'Mixture of Two Steroids - ATR'. There are 'EXCLUDE RANGE BAR' and 'INCLUDE RANGE BAR' labels above and below the plot area. The 'Hit List Size Limit' is set to 50. The 'Search' button is visible at the bottom right of the plot area.</p>

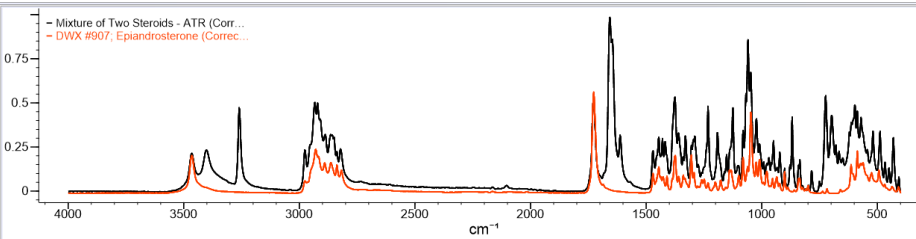







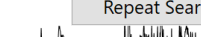







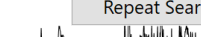







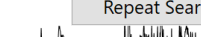
	Action	Result
2	<p>Click on <b>User-Select</b> under <b>Search Databases</b>.</p> <p>Set <b>Limit to spectral technique</b> to <b>IR</b>.</p> <p>Click <b>Add All</b> at the bottom of <b>Available for searching</b> menu.</p>	 <p>The screenshot shows the SearchIt application window. On the left, under 'Search Categories', 'Spectrum FTIR' is checked. Under 'Search Databases', 'User-Select' is selected. The 'Available for Searching' panel on the right has 'Limit to spectral technique' set to 'IR'. A list of databases is shown, and the 'Add All' button is highlighted. Below it, the 'Selected for Searching' panel shows a list of search results.</p>
3	Click <b>Search</b> .	One component search result returns to the <b>Minelt</b> application.

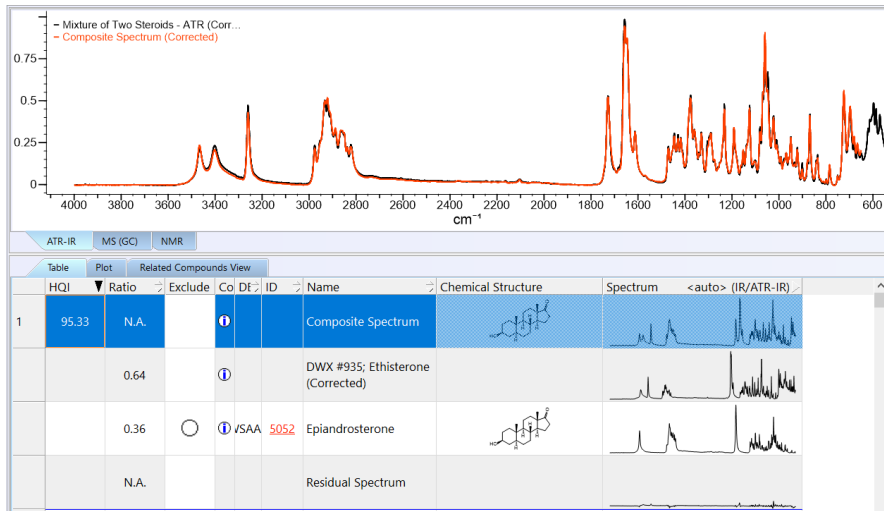


















	Action	Result
		
4	<p>Highlight the first hit.</p> <p>From the <b>Edit</b> menu, select <b>Copy Active Spectrum</b>.</p>	

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



	Action	Result																																																							
7	<p>Click the <b>Spectrum FTIR</b> button to bring up the query spectrum.</p> <p>Click <b>Search</b>.</p>	 <p>The screenshot shows the Minelt software interface. At the top, there's a toolbar with various icons and a 'Lookup Compound:' field. Below that is a plot of an FTIR spectrum. The x-axis is labeled 'cm<sup>-1</sup>' and ranges from 4000 to 500. The y-axis ranges from 0 to 0.75. Two traces are shown: a black trace for 'Mixture of Two Steroids - ATR (Corr...)' and a red trace for 'Composite Spectrum (Corrected)'. Below the plot is a navigation bar with 'ATR-IR', 'MS (GC)', and 'NMR' tabs. Underneath is a table with columns: 'Table', 'Plot', 'Related Compounds View', 'HQI', 'Ratio', 'Exclude', 'Co. DF', 'ID', 'Name', 'Chemical Structure', and 'Spectrum &lt;auto&gt; (IR/ATR-IR)'. The table contains four rows:</p> <table border="1" data-bbox="695 695 1591 950"> <thead> <tr> <th>Table</th> <th>Plot</th> <th>Related Compounds View</th> <th>HQI</th> <th>Ratio</th> <th>Exclude</th> <th>Co. DF</th> <th>ID</th> <th>Name</th> <th>Chemical Structure</th> <th>Spectrum &lt;auto&gt; (IR/ATR-IR)</th> </tr> </thead> <tbody> <tr> <td>1</td> <td></td> <td></td> <td>95.48</td> <td>N.A.</td> <td></td> <td></td> <td></td> <td>Composite Spectrum</td> <td></td> <td></td> </tr> <tr> <td></td> <td></td> <td></td> <td></td> <td>0.62</td> <td></td> <td></td> <td></td> <td>DWX #935; Ethisterone (Corrected)</td> <td></td> <td></td> </tr> <tr> <td></td> <td></td> <td></td> <td></td> <td>0.38</td> <td></td> <td></td> <td>DWX 90Z</td> <td>Epiandrosterone</td> <td></td> <td></td> </tr> <tr> <td></td> <td></td> <td></td> <td></td> <td>N.A.</td> <td></td> <td></td> <td></td> <td>Residual Spectrum</td> <td></td> <td></td> </tr> </tbody> </table> <p>Now we have a good two-component match.</p> <p><b>Notes:</b> 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 <b>Weight</b> value for each component spectrum indicates how much it contributes to the composite spectrum.</p>	Table	Plot	Related Compounds View	HQI	Ratio	Exclude	Co. DF	ID	Name	Chemical Structure	Spectrum <auto> (IR/ATR-IR)	1			95.48	N.A.				Composite Spectrum							0.62				DWX #935; Ethisterone (Corrected)							0.38			DWX 90Z	Epiandrosterone							N.A.				Residual Spectrum		
Table	Plot	Related Compounds View	HQI	Ratio	Exclude	Co. DF	ID	Name	Chemical Structure	Spectrum <auto> (IR/ATR-IR)																																															
1			95.48	N.A.				Composite Spectrum																																																	
				0.62				DWX #935; Ethisterone (Corrected)																																																	
				0.38			DWX 90Z	Epiandrosterone																																																	
				N.A.				Residual Spectrum																																																	

	Action	Result																																																																								
8	<p>The circle in the <b>Exclude</b> column can be checked to exclude a component from consideration.</p> <p>Check to exclude <b>DWX 907</b>. The <b>Repeat Search</b> button shows up for user to repeat Mixture Analysis without considering DWX 907.</p>	 <table border="1" data-bbox="695 594 1604 932"> <thead> <tr> <th colspan="9">ATR-IR</th> </tr> <tr> <th colspan="9">Table Plot Related Compounds View</th> </tr> <tr> <th>HQI</th> <th>Ratio</th> <th>Exclude</th> <th>Co. DF</th> <th>ID</th> <th>Name</th> <th>Chemical Structure</th> <th>Spectrum</th> <th>&lt;auto&gt; (IR/ATR-IR)</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>95.48</td> <td>N.A.</td> <td></td> <td></td> <td>Composite Spectrum</td> <td></td> <td></td> <td></td> </tr> <tr> <td></td> <td>0.62</td> <td></td> <td></td> <td></td> <td>DWX #935; Ethisterone (Corrected)</td> <td></td> <td></td> <td></td> </tr> <tr> <td></td> <td>0.38</td> <td><input checked="" type="checkbox"/></td> <td></td> <td></td> <td>DWX 907 Epiandrosterone</td> <td></td> <td></td> <td></td> </tr> <tr> <td></td> <td>N.A.</td> <td></td> <td></td> <td></td> <td>Residual Spectrum</td> <td></td> <td></td> <td></td> </tr> <tr> <td>2</td> <td>95.48</td> <td>N.A.</td> <td></td> <td></td> <td>Composite Spectrum</td> <td></td> <td></td> <td></td> </tr> </tbody> </table> <p><b>Note:</b> In KnowItAll 2023 release, "Weight" column is labeled as "Ratio" as this value is the ratio of a component spectrum curve.</p>	ATR-IR									Table Plot Related Compounds View									HQI	Ratio	Exclude	Co. DF	ID	Name	Chemical Structure	Spectrum	<auto> (IR/ATR-IR)	1	95.48	N.A.			Composite Spectrum					0.62				DWX #935; Ethisterone (Corrected)					0.38	<input checked="" type="checkbox"/>			DWX 907 Epiandrosterone					N.A.				Residual Spectrum				2	95.48	N.A.			Composite Spectrum			
ATR-IR																																																																										
Table Plot Related Compounds View																																																																										
HQI	Ratio	Exclude	Co. DF	ID	Name	Chemical Structure	Spectrum	<auto> (IR/ATR-IR)																																																																		
1	95.48	N.A.			Composite Spectrum																																																																					
	0.62				DWX #935; Ethisterone (Corrected)																																																																					
	0.38	<input checked="" type="checkbox"/>			DWX 907 Epiandrosterone																																																																					
	N.A.				Residual Spectrum																																																																					
2	95.48	N.A.			Composite Spectrum																																																																					

	Action	Result																																								
9	Click <b>Repeat Search</b> .	<p>The new result does not contain excluded DWX 907.</p>  <table border="1" data-bbox="688 662 1566 899"> <thead> <tr> <th>HQI</th> <th>Ratio</th> <th>Exclude</th> <th>Co. DF</th> <th>ID</th> <th>Name</th> <th>Chemical Structure</th> <th>Spectrum</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>95.33</td> <td>N.A.</td> <td></td> <td></td> <td>Composite Spectrum</td> <td></td> <td></td> </tr> <tr> <td></td> <td>0.64</td> <td></td> <td></td> <td></td> <td>DWX #935; Ethisterone (Corrected)</td> <td></td> <td></td> </tr> <tr> <td></td> <td>0.36</td> <td></td> <td></td> <td>VSAA 5052</td> <td>Epiandrosterone</td> <td></td> <td></td> </tr> <tr> <td></td> <td>N.A.</td> <td></td> <td></td> <td></td> <td>Residual Spectrum</td> <td></td> <td></td> </tr> </tbody> </table>	HQI	Ratio	Exclude	Co. DF	ID	Name	Chemical Structure	Spectrum	1	95.33	N.A.			Composite Spectrum				0.64				DWX #935; Ethisterone (Corrected)				0.36			VSAA 5052	Epiandrosterone				N.A.				Residual Spectrum		
HQI	Ratio	Exclude	Co. DF	ID	Name	Chemical Structure	Spectrum																																			
1	95.33	N.A.			Composite Spectrum																																					
	0.64				DWX #935; Ethisterone (Corrected)																																					
	0.36			VSAA 5052	Epiandrosterone																																					
	N.A.				Residual Spectrum																																					

**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.

**Number of components:**

3 (Mixture) ▼

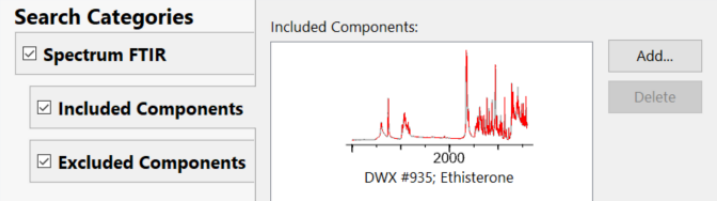
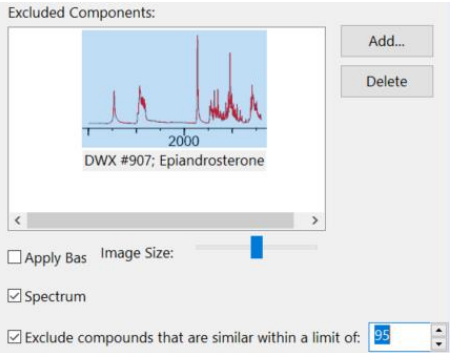
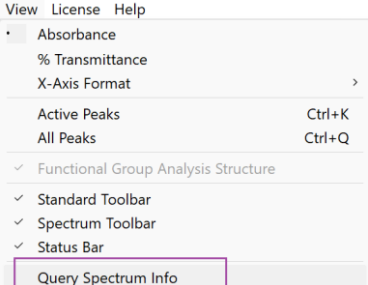
1 (Single)

2 (Mixture)

3 (Mixture)

4 (Mixture)

5 (Mixture)

<ul style="list-style-type: none"> <li>You can add <b>Included</b> or <b>Excluded</b> components from files by clicking the <b>Add</b> button.</li> </ul>	
<ul style="list-style-type: none"> <li>You can exclude spectrally similar records.</li> </ul>	
<p>One can view the metadata of query spectrum by</p> <ul style="list-style-type: none"> <li><b>View &gt; Query Spectrum Info</b> in <b>SearchIt</b> and</li> <li><b>View &gt; Windows/Tables &gt; Query Spectrum Info</b> in a <b>Minelt</b> hit list.</li> </ul>	

View Database Hit List MS Tools NMR Tools Window License Help

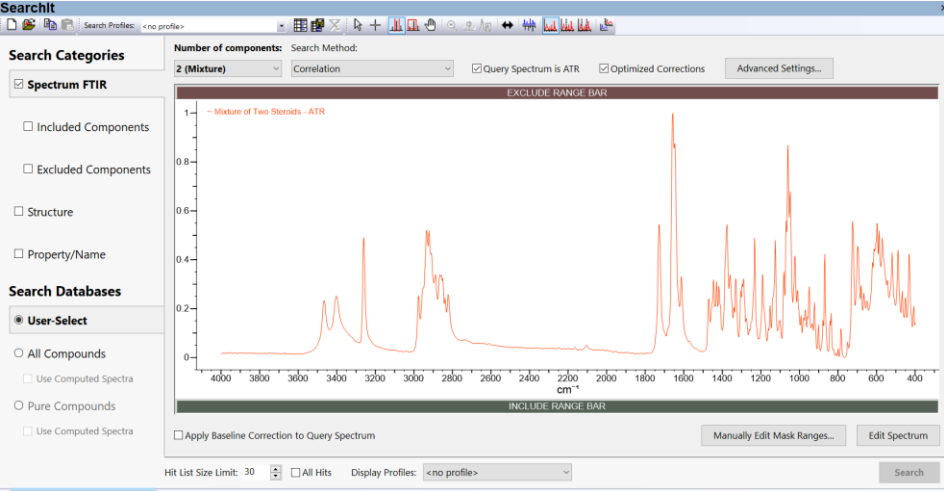
- Absorbance
- % Transmittance
- Edit Mode >
- Display Mode >
- Auto Scale Mode >
- X-Axis Format >
- Spectrum Overview Pane
- View Default Region Ctrl+1
- View Entire Spectrum Ctrl+0
- Peaks >
- Integrals
- ✓ Included Spectrum Range
- Toolbars >
- Windows/Tables >
  - ✓ Spectrum Pane Alt+6
  - ✓ Structure/Properties Table Alt+3
  - Peak Table Alt+2
  - AUC/Integration Table
  - Coupling Data Table
  - Search Parameters Table Alt+5
  - Search Overview Pane Alt+7
  - Query Spectrum Info
- Implicit Hydrogens >
- Assignment Information >
- ✓ Stereochemistry
- ✓ Query Structure Overlap

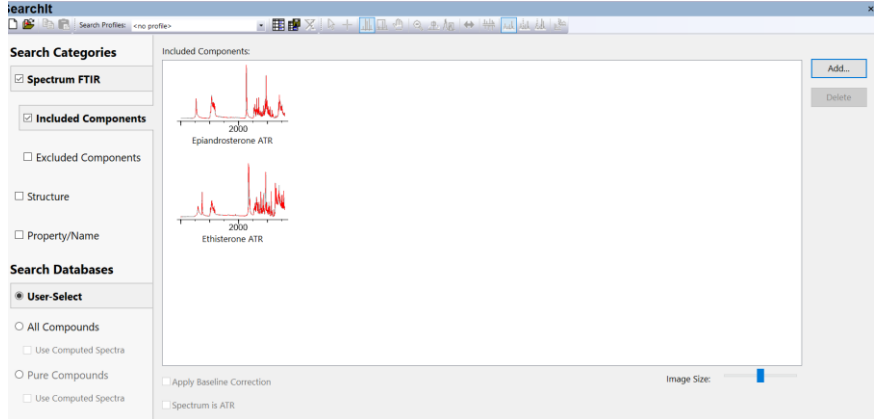
Expert SearchIt QC Expert ProcessIt

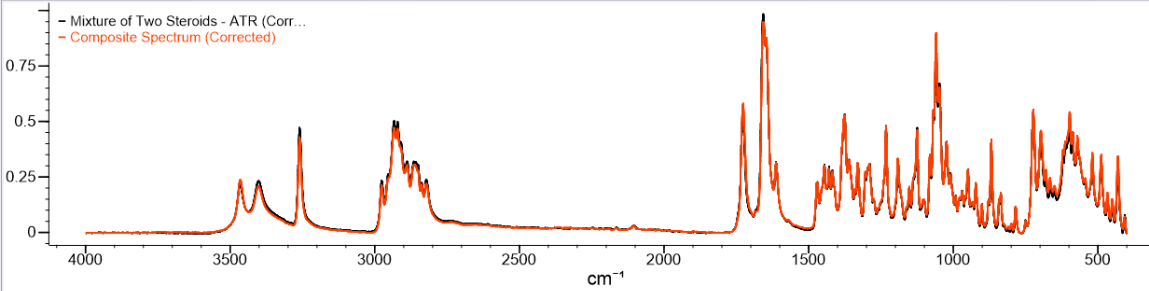
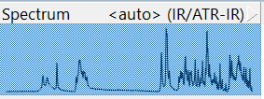



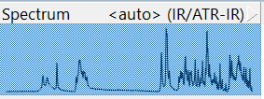



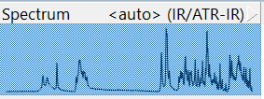


Lookup Compound:

4000 3800 3600

## Use Included Components

	Action	Result
1	<p>In the <b>Data</b> toolbox, open the <b>SearchIt</b> application.</p> <p>Check <b>Spectrum</b>.</p> <p>In the resulting <b>Open</b> dialog box, navigate to <b>C:\Users\Public\Documents\Wiley\KnowItAll\Samples\Mixture Analysis\IR Examples</b> folder.</p> <p>Open <b>Mixture of Two Steroids - ATR-IR</b>.</p> <p>Set <b>Search Method</b> to <b>Correlation</b>.</p> <p>Set <b>Number of components</b> to <b>2 (Mixture)</b>.</p>	 <p>The screenshot displays the SearchIt application window. On the left, the 'Search Categories' panel is active, with 'Spectrum FTIR' checked. Below it, 'Included Components' is also checked. The 'Search Databases' section shows 'User-Select' as the active database. The main area shows the 'Number of components' set to '2 (Mixture)' and the 'Search Method' set to 'Correlation'. The 'Query Spectrum is ATR' checkbox is checked. The central plot shows the FTIR spectrum of the 'Mixture of Two Steroids - ATR' sample, with the x-axis labeled 'cm⁻¹' ranging from 4000 to 400. The y-axis represents intensity, ranging from 0 to 1.0. The plot shows a complex spectrum with several sharp peaks, particularly in the fingerprint region (below 1500 cm⁻¹). The plot is titled 'Mixture of Two Steroids - ATR'. At the bottom of the window, there are options for 'Hit List Size Limit: 30', 'All Hits', 'Display Profile: &lt;no profile&gt;', and a 'Search' button.</p>

	Action	Result
2	<p>Check <b>Included Components</b>.</p> <p>Add the following files from C:\Users\Public\Documents\Wiley\KnowItAll\Samples\Mixture Analysis\IR Examples\Components folder:</p> <ul style="list-style-type: none"><li>• Epiandrosterone ATR-IR</li><li>• Ethisterone ATR-IR</li></ul> <p><b>Note:</b> Use the Ctrl key to select multiple files in the <b>Open</b> dialog box.</p>	 <p>The screenshot shows the iearchit software interface. On the left, there are search categories: <b>Spectrum FTIR</b> (checked), <b>Included Components</b> (checked), <b>Excluded Components</b> (unchecked), <b>Structure</b> (unchecked), and <b>Property/Name</b> (unchecked). Below these are search databases: <b>User-Select</b> (selected), <b>All Compounds</b> (radio button), <b>Use Computed Spectra</b> (checkbox), <b>Pure Compounds</b> (radio button), and <b>Use Computed Spectra</b> (checkbox). At the bottom, there are checkboxes for <b>Apply Baseline Correction</b> and <b>Spectrum is ATR</b>. The main area displays two IR spectra: <b>Epiandrosterone ATR</b> and <b>Ethisterone ATR</b>, both showing characteristic absorption bands. An <b>Add...</b> button is visible on the right side of the main area.</p>
3	<p>Click <b>Search</b>.</p> <p><b>Note:</b> Toggle between different spectra display modes by selecting available options from <b>View &gt; Display Mode</b>.</p>	

Action	Result																																																			
	 <p data-bbox="695 609 1837 641">ATR-IR MS (GC) NMR</p> <table border="1" data-bbox="695 641 1837 982"> <thead> <tr> <th colspan="2">Table</th> <th colspan="2">Plot</th> <th colspan="2">Related Compounds View</th> </tr> <tr> <th>HQI</th> <th>Ratio</th> <th>Exclude</th> <th>Co</th> <th>DF</th> <th>ID</th> <th>Name</th> <th>Spectrum</th> <th>&lt;auto&gt; (IR/ATR-IR)</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>95.48</td> <td>N.A.</td> <td></td> <td></td> <td></td> <td>Composite Spectrum</td> <td></td> <td></td> </tr> <tr> <td></td> <td>0.62</td> <td></td> <td></td> <td></td> <td></td> <td>Ethisterone ATR</td> <td></td> <td></td> </tr> <tr> <td></td> <td>0.38</td> <td></td> <td></td> <td></td> <td></td> <td>Epiandrosterone ATR</td> <td></td> <td></td> </tr> <tr> <td></td> <td>N.A.</td> <td></td> <td></td> <td></td> <td></td> <td>Residual Spectrum</td> <td></td> <td></td> </tr> </tbody> </table> <p data-bbox="695 990 1890 1047">In this case, KnowItAll simply returns possible combinations of two provided components without going through a database search.</p>	Table		Plot		Related Compounds View		HQI	Ratio	Exclude	Co	DF	ID	Name	Spectrum	<auto> (IR/ATR-IR)	1	95.48	N.A.				Composite Spectrum				0.62					Ethisterone ATR				0.38					Epiandrosterone ATR				N.A.					Residual Spectrum		
Table		Plot		Related Compounds View																																																
HQI	Ratio	Exclude	Co	DF	ID	Name	Spectrum	<auto> (IR/ATR-IR)																																												
1	95.48	N.A.				Composite Spectrum																																														
	0.62					Ethisterone ATR																																														
	0.38					Epiandrosterone ATR																																														
	N.A.					Residual Spectrum	