# KnowItAll Software Training

## Data Mining & Analysis

# Data Mining & Analysis

## Overlap Density Heatmap: A Technology to Analyze Spectral, Chromatographic, and Other Graphical Data

### Purpose

This exercise demonstrates how to use Overlap Density Heatmap for data mining and visualization.

### Objectives

This exercise will teach you:

- How to view and manipulate an Overlap Density Heatmap

### Background

Wiley's patented Overlap Density Heatmap is useful for visual data mining and analysis to assess the similarities and dissimilarities in large amounts of spectral, chromatographic and other graphical data.

This technology allows the visualization of common features of overlapped objects, such as spectra or chromatograms, by color-coding the areas from highest to lowest overlap.

***Training Files Used in This Lesson***

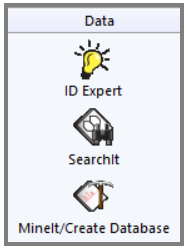
C:\Users\Public\Documents\Wiley\KnowItAll\Samples\IR

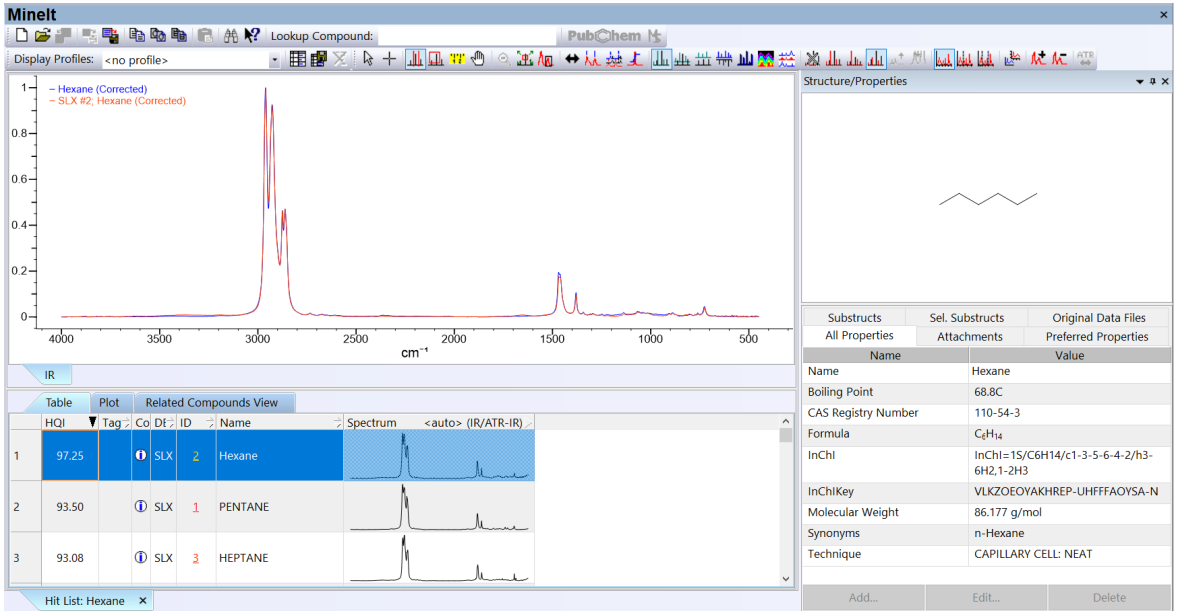
- Hexane.jdx

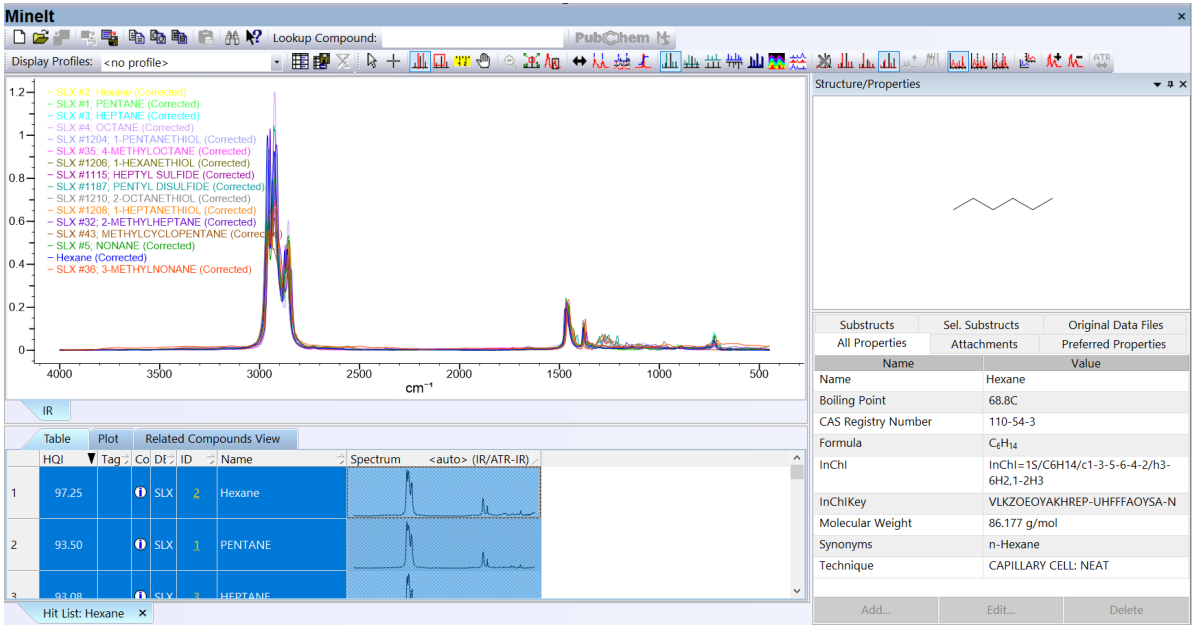
***KnowItAll Applications Used***

- MineIt™

## Open a spectral hit list in the MineIt application

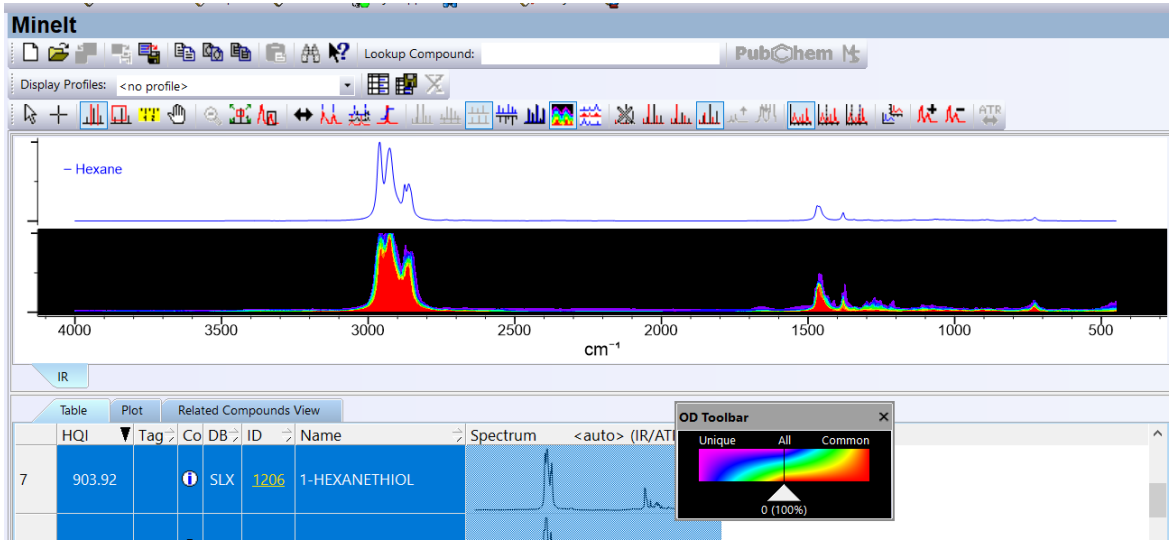
| | Action | Result |
|---|---|---|
| 1 | In the **SearchIt** application, click **User-Select** under **Search Databases**.<br><br>If databases are present in the **Selected for Searching** pane, click **Remove All.**<br><br>Add the **IR - Sadtler Standards (Selected Subset) - Wiley** (DB Code SLX) to the **Selected for Searching** pane. |  |
| 2 | Click **Spectrum** in the **Search Categories** pane.<br><br>Navigate to **C:\Users\Public\Documents\Wiley\KnowItAll\Samples\IR** folder.<br><br>Open **Hexane.jdx**. | |

| | Action | Result |
|---|---|---|
| 3 | Click **Search**. | The results are displayed in the **MineIt** application.  |

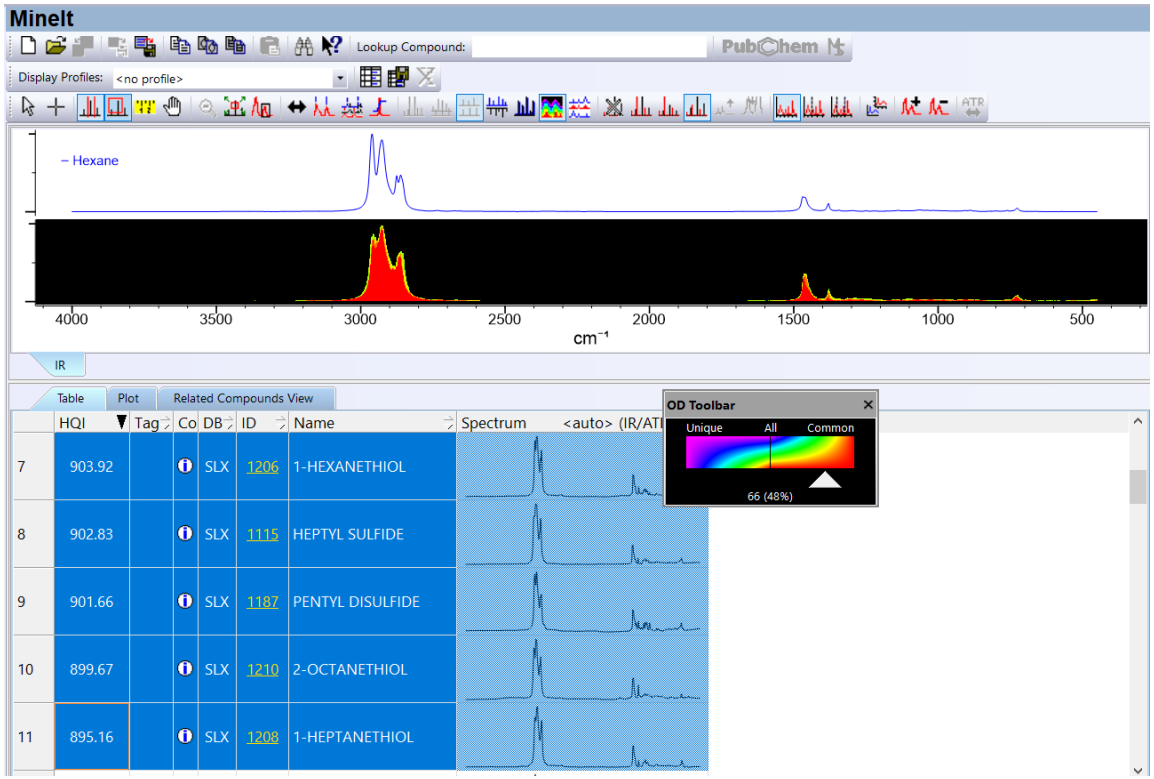
| | Action | Result |
|---|---|---|
| 4 | Select the first 15 entries in the hit list.<br><br>**Note**: Select the first record, hold down the shift key, then scroll to and select the 15th entry. | All the selected hits are displayed in the spectral pane.<br>  |

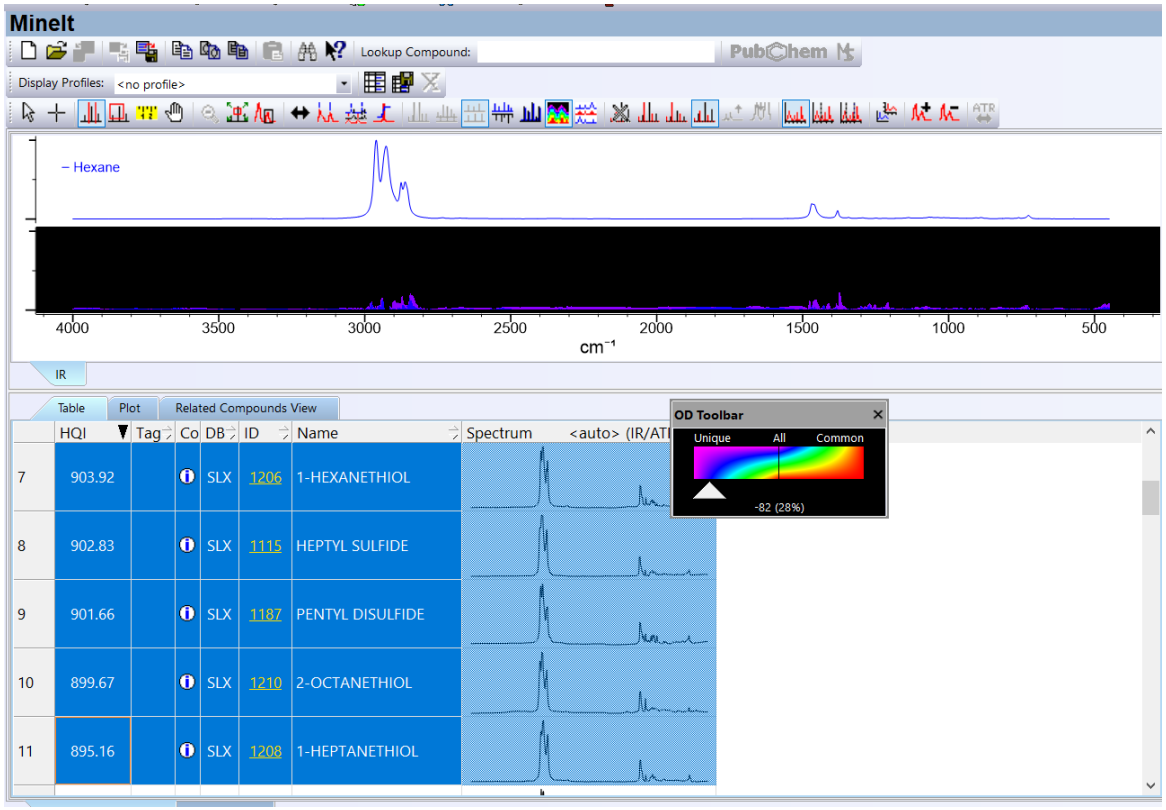
## Change the spectral display to Overlap Density Heatmap

| | Action | Result |
|---|---|---|
| 1 | Click the **Overlap Density (OD) Heatmap** tool in the **Spectral** toolbar. | The spectral display changes, and the **Overlap Density (OD) Toolbar** is added to the display.<br>  <br>The default Overlap Density Heatmap display shows all overlap levels. High levels of overlap are displayed in red; low levels are displayed in violet. The **OD Toolbar** allows you to control the amount of overlap displayed using the slider. In the default Overlap Density Heatmap, the slider is set at OD level = 0, where all colors–representing all overlap density levels–are shown. |

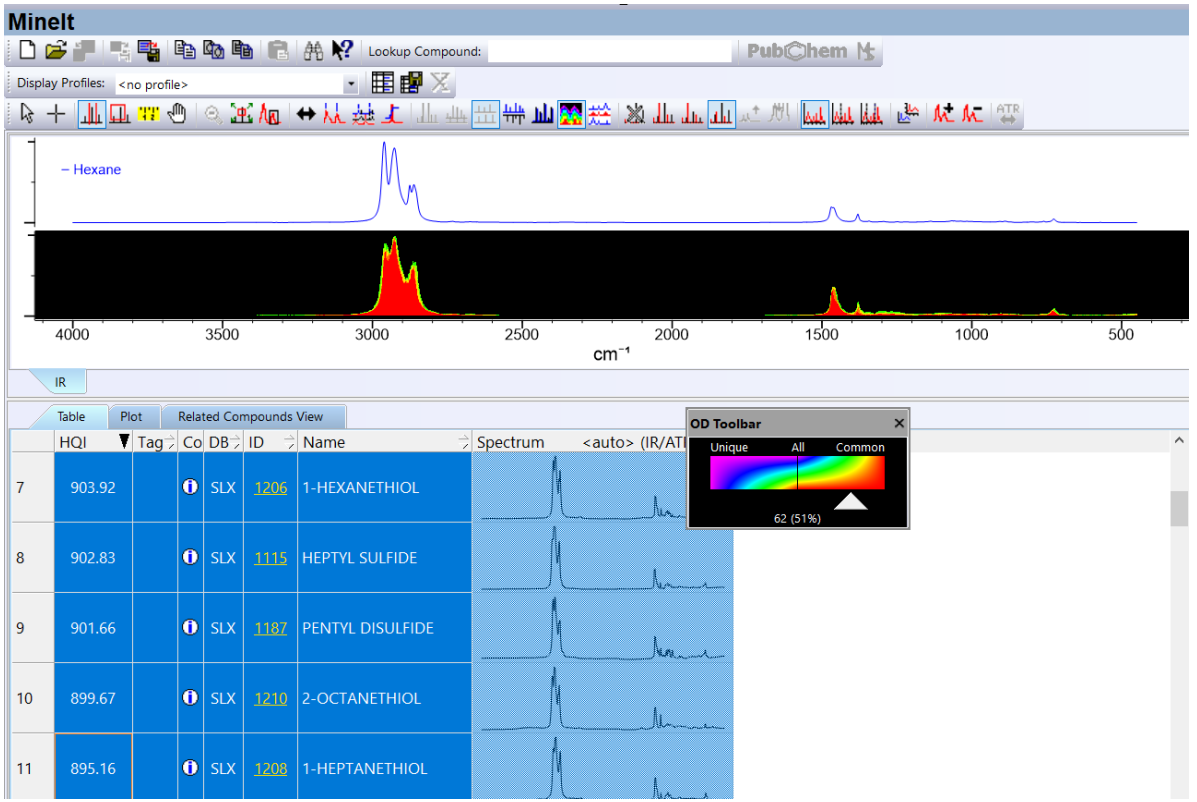
## Manipulate the Overlap Density Heatmap

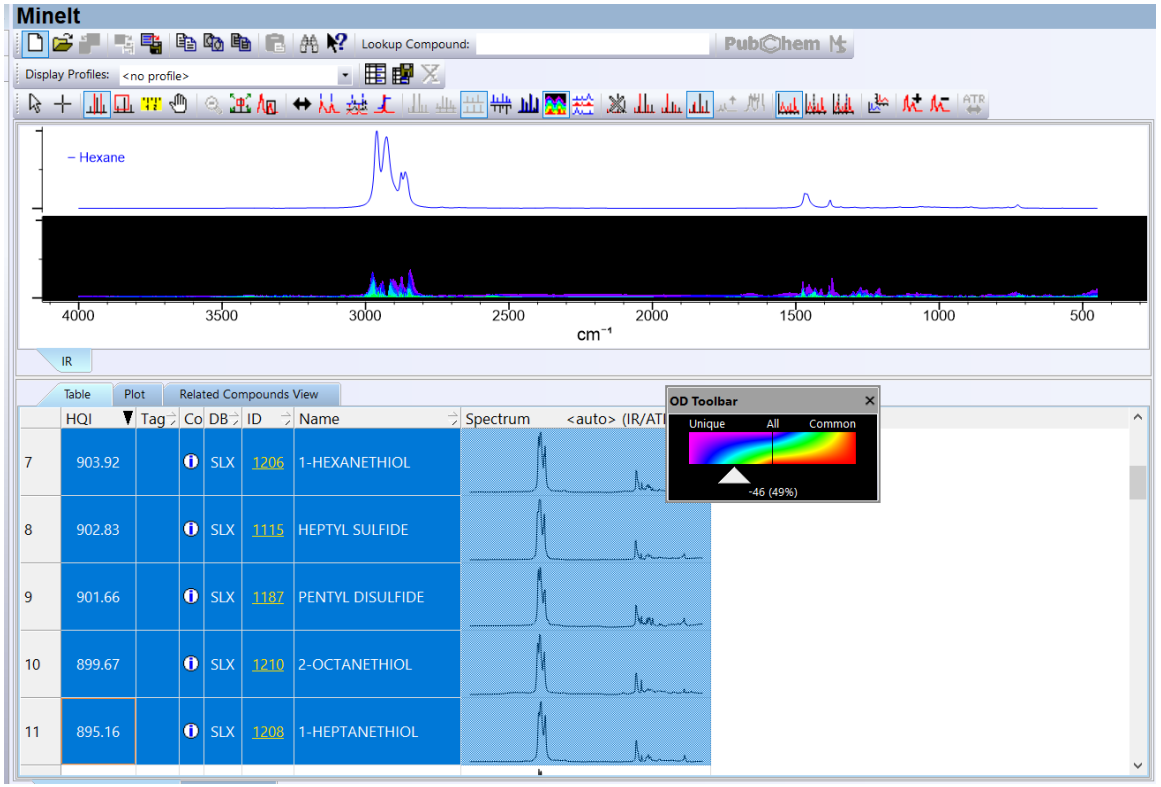
| | Action | Result |
|---|---|---|
| 1 | Move the slider on the **OD Toolbar** to the right. | As the slider moves to the right, the areas of common overlap are shown. Only the areas of most common overlap are shown as the OD Level approaches 100.   |

| | Action | Result |
|---|---|---|
| 2 | Move the slider to the left. | As the OD Level approaches -100, only the areas of lowest overlap–the unique areas–are shown.<br>  <br>As you move the slider to change the OD Level, a second value is shown in parentheses. This value is the percent area under the curve, or %AUC, relative to the value at OD Level = 0. |

| | Action | Result |
|---|---|---|
| 3 | Move the slider to the point where %AUC equals approximately 50%. | The OD Level on the common side–or ODC–required to show 50% of the area that is most in common, is 62.<br>  <br>Another way to express this is to say that $ODC_{50}$ = 62. An alternative is to say that $\%AUC_{42}$ = 50. |

| | Action | Result |
|---|---|---|
| 4 | Next, move the slider to the left to where %AUC equals approximately 50%. | The OD Level on the unique side–or ODU–required to show 50% of the area that is most unique is -46.<br>  <br>Another way to express this is to say that $ODU_{50}$ = -64. An alternative is to say that $\%AUC_{-64}$ = 50. |
| | **TIP** | Use the **Horizontal Zoom** tool  to examine different regions. To do this, select the Horizontal Zoom tool, and then click and drag over an area on the spectrum to zoom in. Use the **View Entire Spectrum** tool  to zoom out. |

# Data Mining & Analysis

## How to Create and Use Overlap Density Consensus Spectra

### Purpose

This exercise demonstrates how to create and use Overlap Density Consensus Spectra in the KnowItAll Informatics System.

### Objectives

This exercise will teach you:

- How to view and manipulate an Overlap Density Consensus spectrum
- How to use an Overlap Density Consensus spectrum in a search

### Background

Wiley's patented Overlap Density Heatmap technology is useful for visual data mining and analysis to assess the similarities and dissimilarities in large amounts of spectral, chromatographic and other graphical data.

By tracing the outline of the highest level of overlap at a given OD Level, it is possible to mathematically construct a composite spectrum by using the maximum spectral y-values at each spectral x-value in the OD Heatmap. This Overlap Density Consensus Spectrum can be used in a spectral search to find similar spectra or can be stored in a database for future use.

***Training Files Used in This Lesson***

C:\Users\Public\Documents\Wiley\KnowItAll\Samples\IR

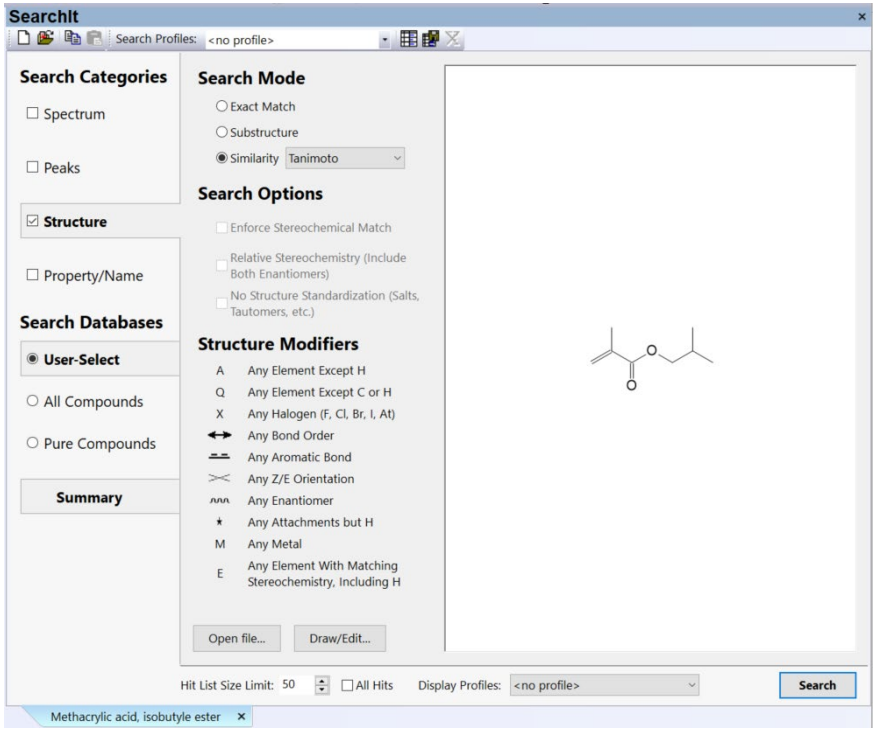
- Propiophenone Query.dsf

***KnowItAll Applications Used***

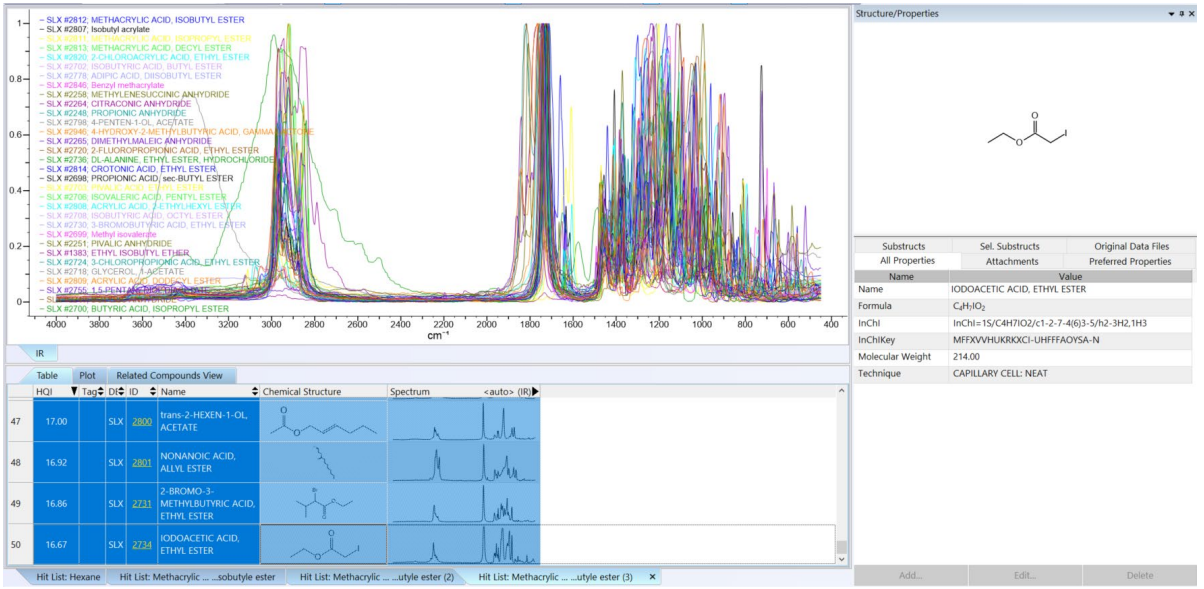
- SearchIt™
- MineIt™

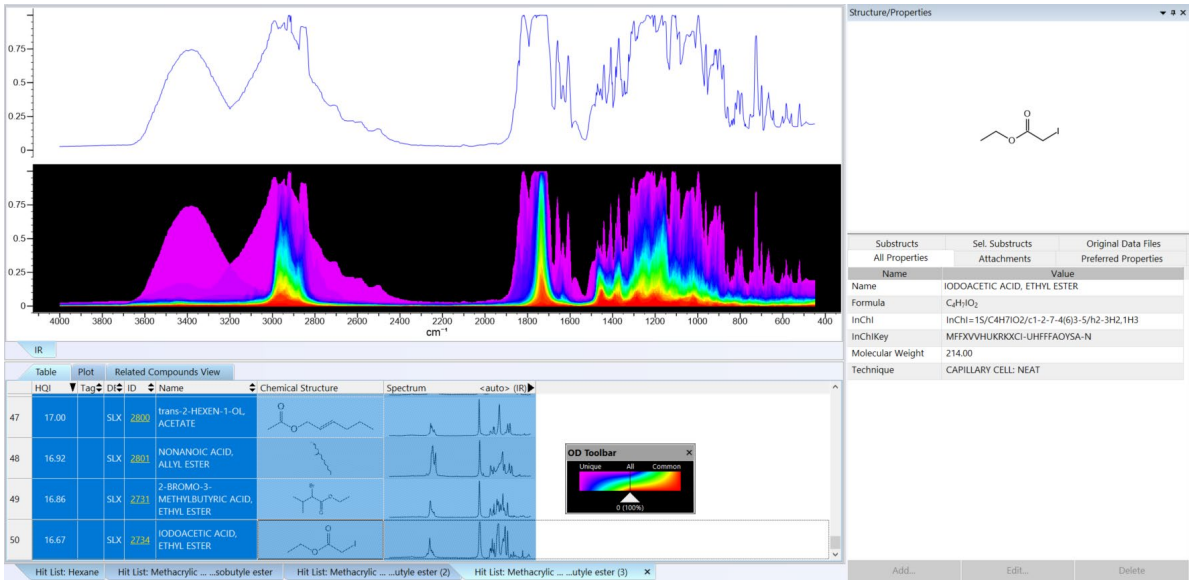
## Perform a substructure search

| | Action | Result |
|---|---|---|
| 1 | In the **SearchIt** application, click **User-select** under **Search Databases**.<br><br>Under **Available for Searching**, select **IR - Sadtler Standards (Selected Subset) - Wiley** (DB Code SLX).<br><br>Click **Add**. | |
| 2 | Click **Structure**.<br><br>Click **Open Spectrum or Structure** icon. |  |

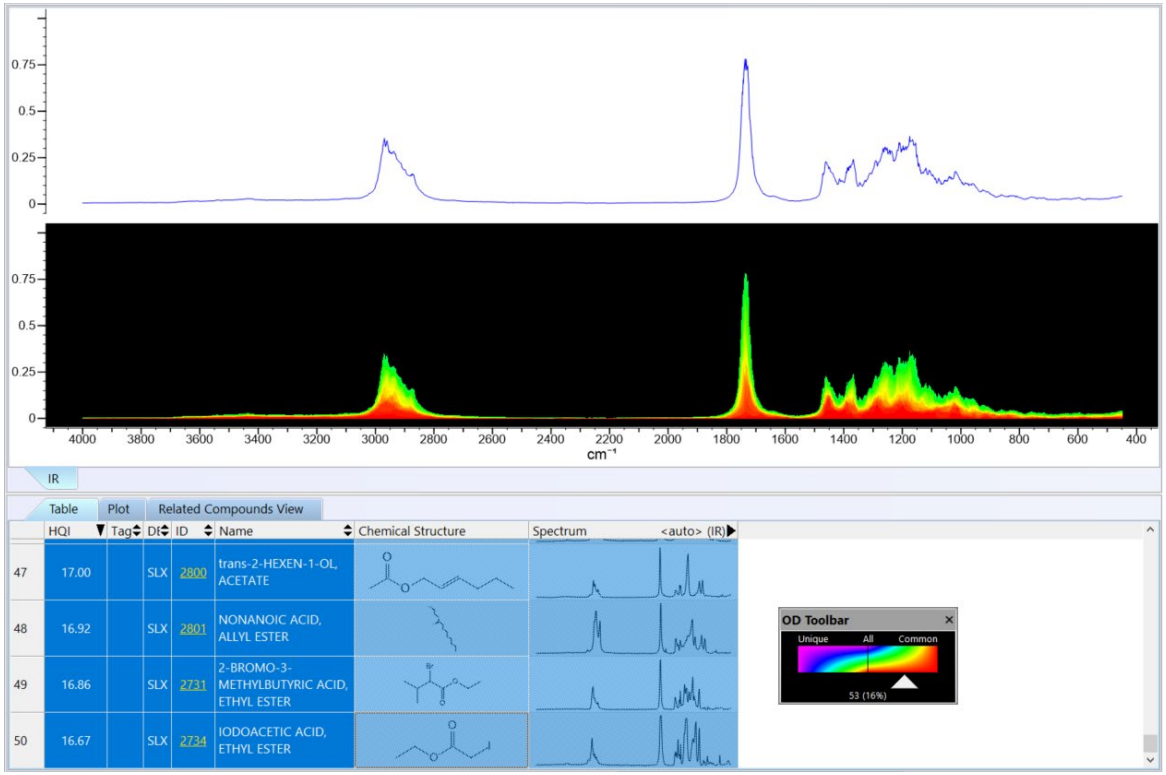
| | Action | Result |
|---|---|---|
| 3 | Navigate to **C:\Users\Public\Documents\Wiley\KnowItAll\Samples\Structues** folder.<br><br>Open **Methacrylic acid, isobutyle ester**. | The structure is displayed in the **Structure** tab. |
| 4 | Make sure **Search Mode** is set to **Similarity,** use default **Tanimoto** algorithm**.**<br><br>Click **Search**. |  |

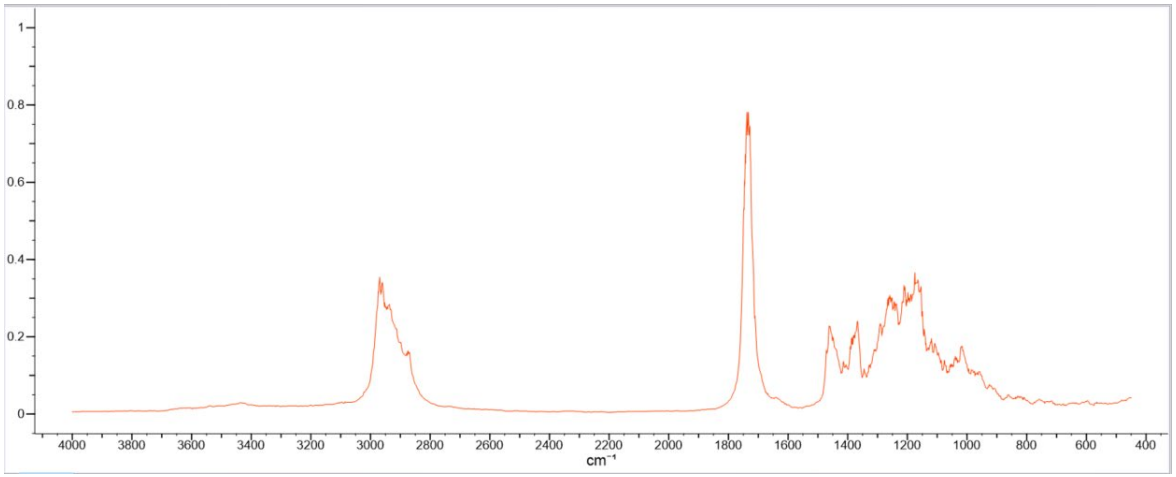
## Examine the results in MineIt

| | Action | Result |
|---|---|---|
| 1 | Select all eleven records in the hit list.<br><br>**Note**: Select the first record, hold down the shift key, then scroll to and select the last record. | All spectra are displayed in the spectral pane.<br>  |
| 2 | Click the **OD Heatmap** toolbar button. | |

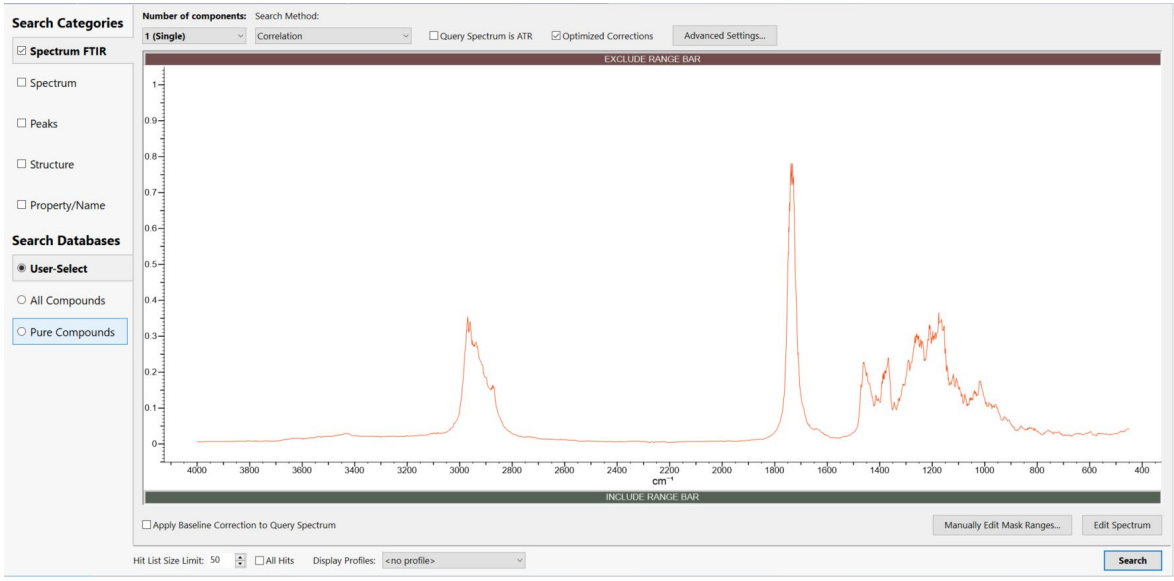
| | Action | Result |
|---|---|---|
| 3 | Click the **OD Consensus Spectrum** toolbar button  . | The consensus spectrum is displayed above the Overlap Density Heatmap.   |

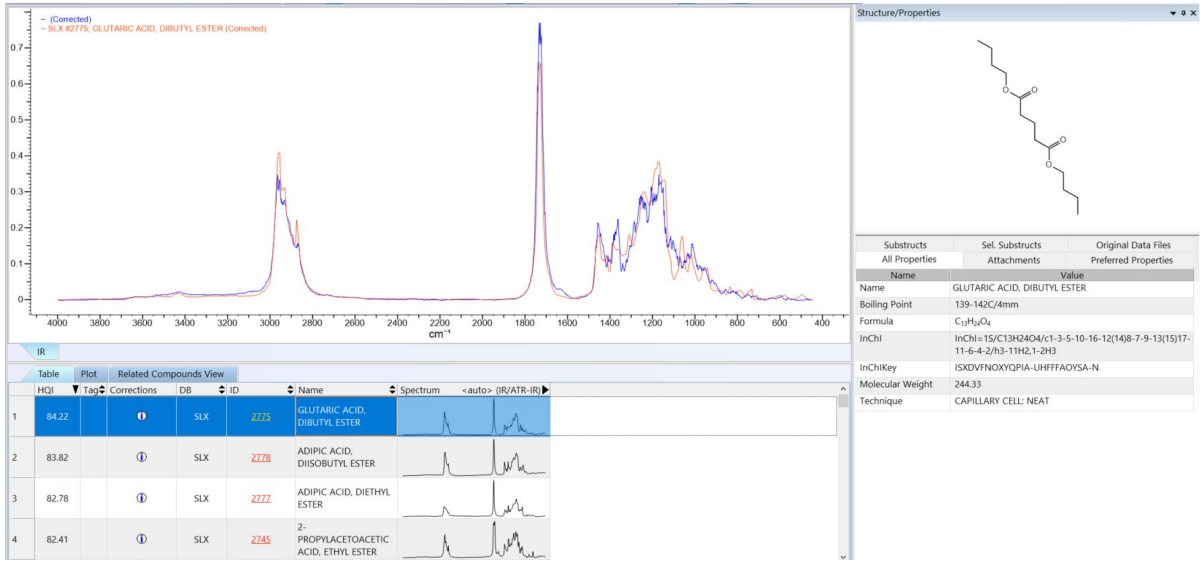
 

| | | |
|---|---|---|
| 4 | Move the OD Toobar to around 50 (16%) |  |

| | Action | Result |
|---|---|---|
| 5 | Click the **OD Heatmap** toolbar button again to turn off this display. | The consensus spectrum remains in the display.<br>  |

## Use the consensus spectrum to perform a search

| | Action | Result |
|---|---|---|
| 1 | Click **SearchIt** in the **Transfer to** bar. | The consensus spectrum is displayed on the **IR Spectrum** search tab.   |

| | | |
|---|---|---|
| 2 | Click the **Search** button.<br><br>**Note**: The IR - Sadtler Standards (Selected Subset) - Wiley database remains in the spectrum pane. | The results are displayed in the **MineIt** application on a new tab.<br>  <br><br>You got a hit list where hits are structurally similar to that of **Methacrylic acid, isobutyle ester** |