KnowItAll Software Training

Mass Spectrometry Search

WILEY

Mass Spectrometry Spectral Searching

How to Perform a Mass Spectrometry Spectral Search

Purpose

This exercise demonstrates how to perform a Mass Spectrometry spectral search.

Objectives

This exercise will teach you:

- How to perform a straightforward search
- How to perform a reverse search
- How to complete a mixture analysis
- How to perform an Adaptive search (similar compound search)
- How to perform simultaneous multiple MS spectra search

Background

Spectral searching against reference databases is frequently used in the analysis of unknown compounds. KnowItAll has full-featured MS spectrum comparison tools for this purpose.

Training Files Used in This Lesson

In C:\Users\Public\Documents\Wiley\KnowItAll\Samples folder

- \MS\1,1,1-Trichlorobutane - Adaptive Search demo
- \MS\2-Hydroxybenzoic acid
- \Mixture Analysis\MS Examples\MS Mixture of Two 1
- \Mixture Analysis\MS Examples\MS Mixture of Two 2
- \Mixture Analysis\MS Examples\MS Mixture of Three
- \GC-MS\Barbiturate GC-MS.d
- \Mixture Analysis\MS Examples\Components.SDBX

KnowItAll Applications Used

- SearchIt™
- MineIt™

Algorithms

Research article

Journal of MASS SPECTROMETRY

Received: 5 October 2014 Revised: 16 February 2015 Accepted: 5 March 2015 Published online in Wiley Online Library

(wileyonlinelibrary.com) DOI 10.1002/jms.3591

Evaluation of mass spectral library search algorithms implemented in commercial software

Andrey Samokhin,[a]* Ksenia Sotnezova,[a] Vitaly Lashin[b] and Igor Revelsky[a]

MS SEARCH

Composite algorithm

$$SI = \frac{N_U \cdot \left[\frac{\left(\sum W_L \cdot W_U\right)^2}{\sum W_L^2 \sum W_U^2}\right] + \left[\sum \left(\frac{R_U}{R_L}\right)^n\right]}{N_U + N_{U\&L}}$$

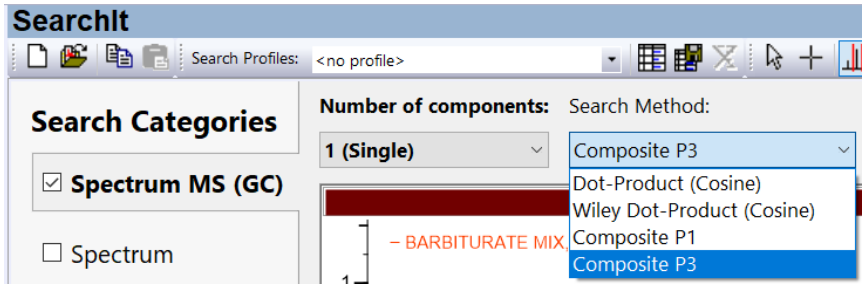
Spectrum search type – identity (normal)
Presearch – default
Included Libs – MainLib
Apply limits – unchecked
Use constraints – unchecked

Dot-product algorithm

$$SI = \frac{\left(\sum W_L \cdot W_U\right)^2}{\sum W_L^2 \cdot \sum W_U^2}$$

Spectrum search type – similarity (simple)
Presearch – default
Included Libs – MainLib
Apply limits – unchecked
Use constraints – unchecked

A. Samokhin, K. Sotnezova, V. Lashin, I. Revelsky. Evaluation of mass spectral library search algorithms implemented in commercial software. *Journal of Mass Spectrometry.* 2015, 50, 820-825.



Search Methods:

- Dot-Product (Cosine): second equation in above graph
- Wiley Dot-Product (Cosine): the Finnigan algorithm that verified at least 12 of the largest 16 peaks and the base peak match before continuing with the dot-product calculation
- Composite P1: first equation in above graph
- Composite P3: first equation in above graph
  P1 and P3 are different by the power applied to the weighted intensity of peak.

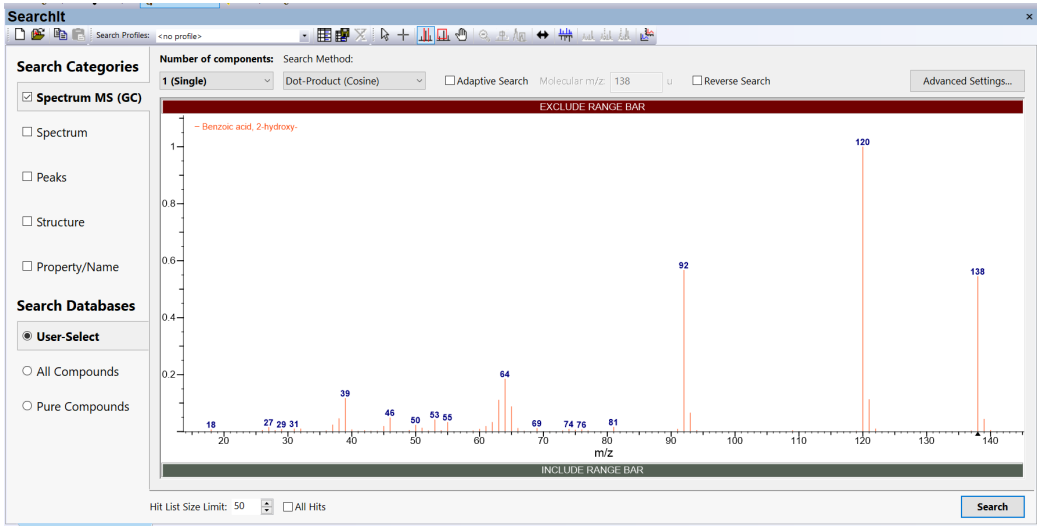
Optimized Corrections:

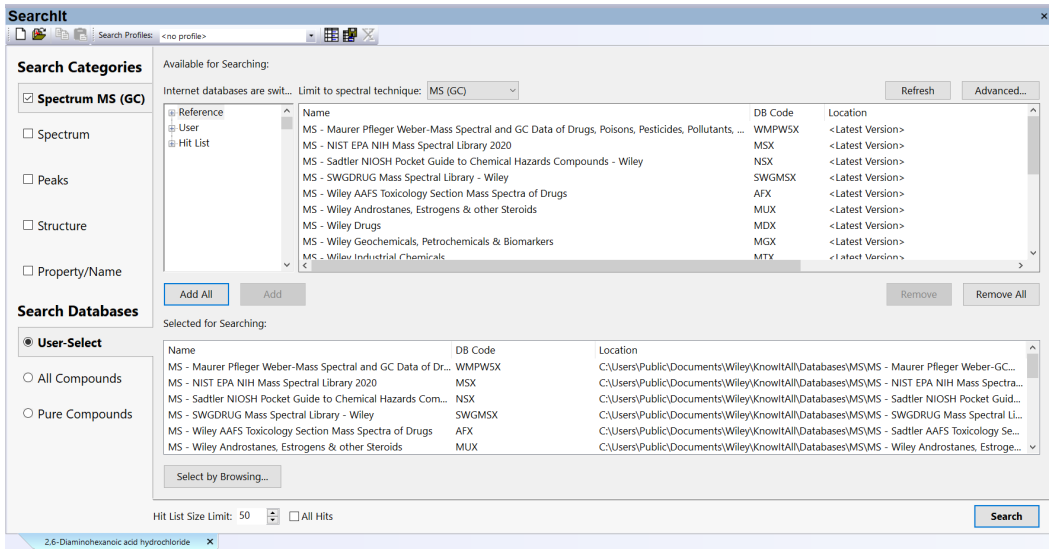
- Mass Defection is the difference between a compound's exact mass and its nominal mass. It is automatically applied for a MS search. Examples:
  - For hydrocarbon compounds, when *m/z* value is over 500, use 0.99888.
  - For polybrominated compounds, when there are more than 5 Br atoms and *m/z* value is over 800, use 1.00087.
- Spectral Skewing is caused by analyte's concentration changing during scan. A linear compensation factor (positive for ascending or negative for descending slope) is calculated for each search match. This factor corrects the intensities as follows:

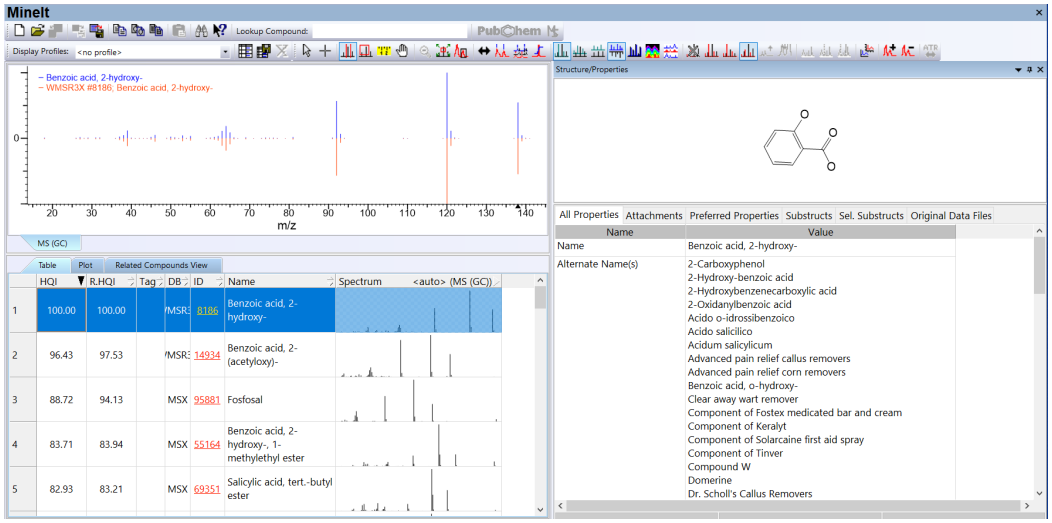
  I(corrected)=I * f * m,

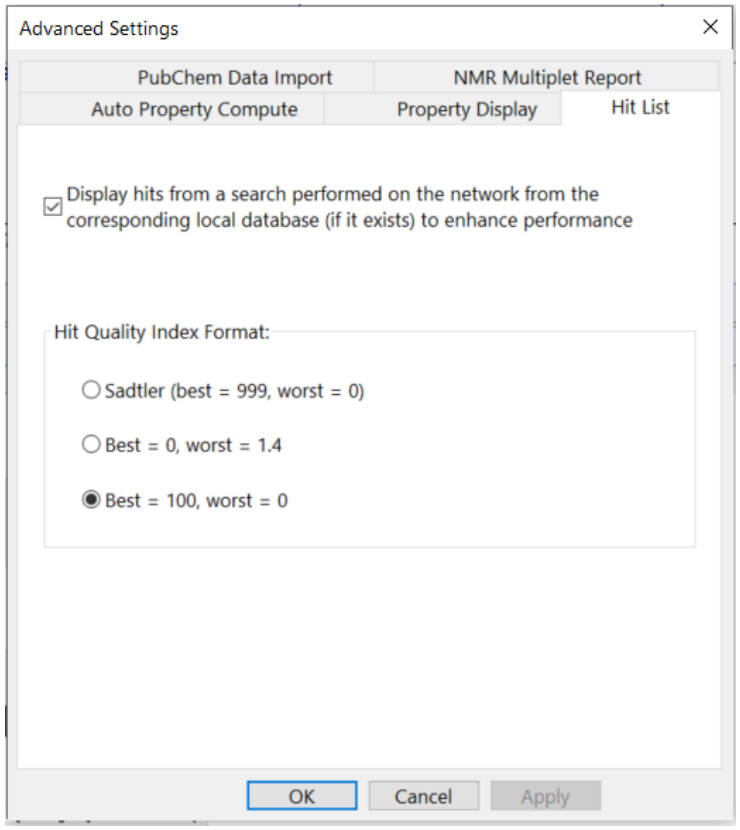
  where I=intensity, f=OC factor, and m=*m/z* value

Straightforward Search

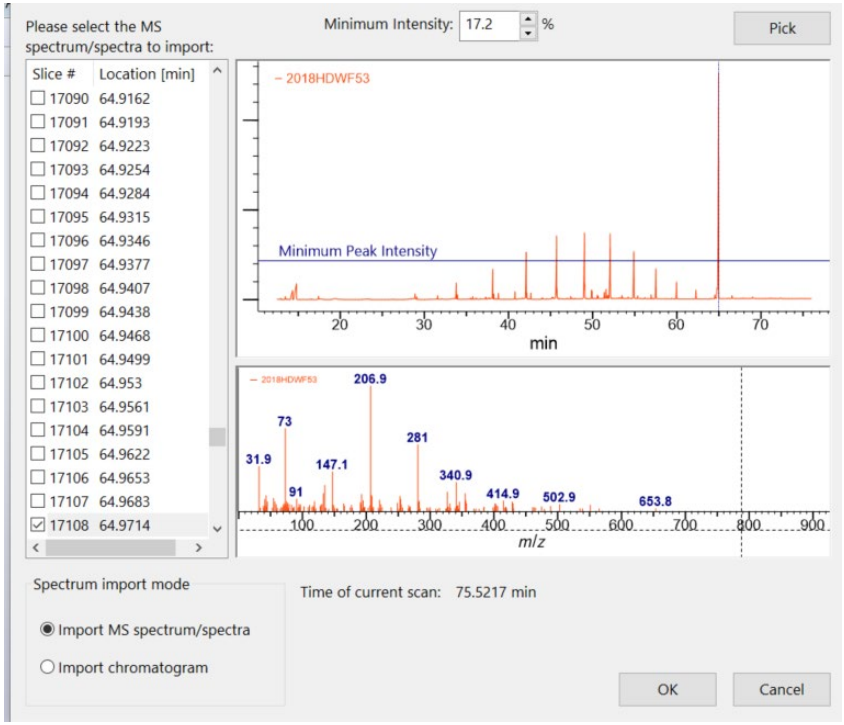
	Action	Result
1	Open the KnowItAll **SearchIt** application by clicking its icon in the KnowItAll application. Check **Spectrum** and navigate to C:\Users\Public\Documents\Wiley\KnowItAll\Samples\MS folder Select the file **2-Hydroxybenzoic acid**. Click **Open**. Ensure **Adaptive Search** or **Reverse Search** is unchecked.	

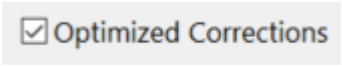
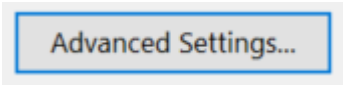
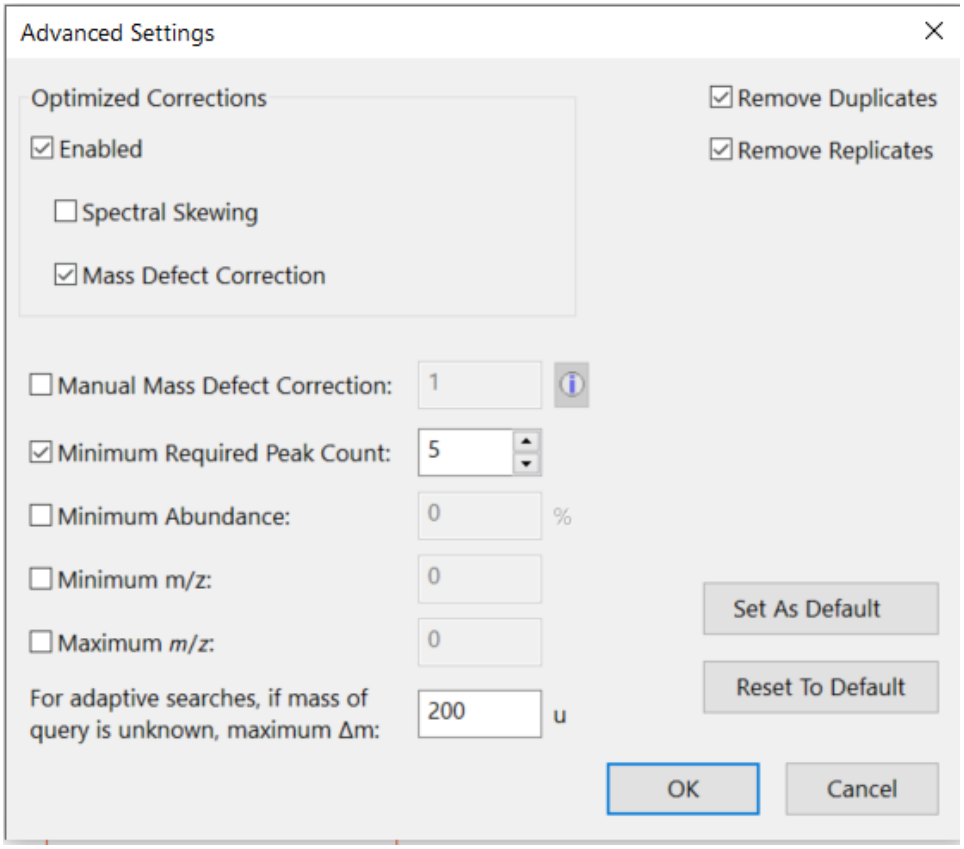
	Action	Result
2	Click **User-Select.** Use **Remove All** to clean selected databases. Use **Add All** to add all MS (GC) databases. Click **Search**.	

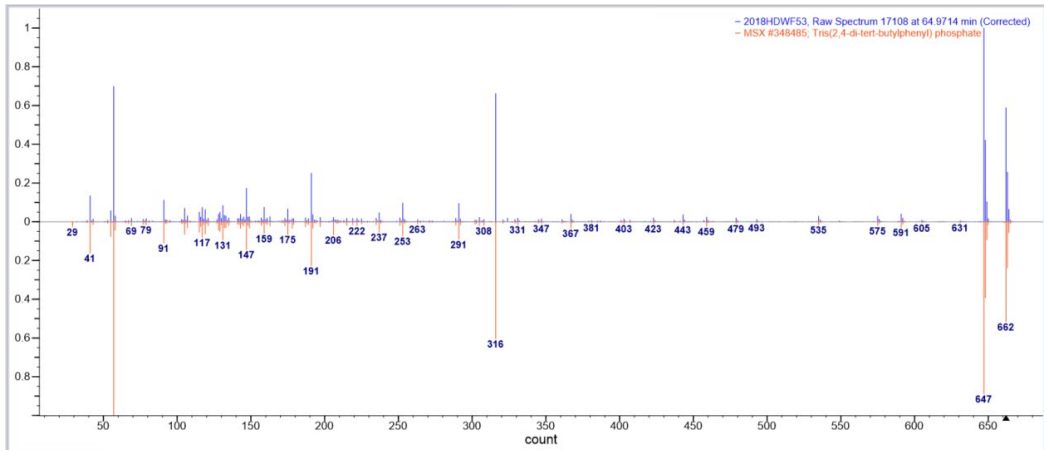
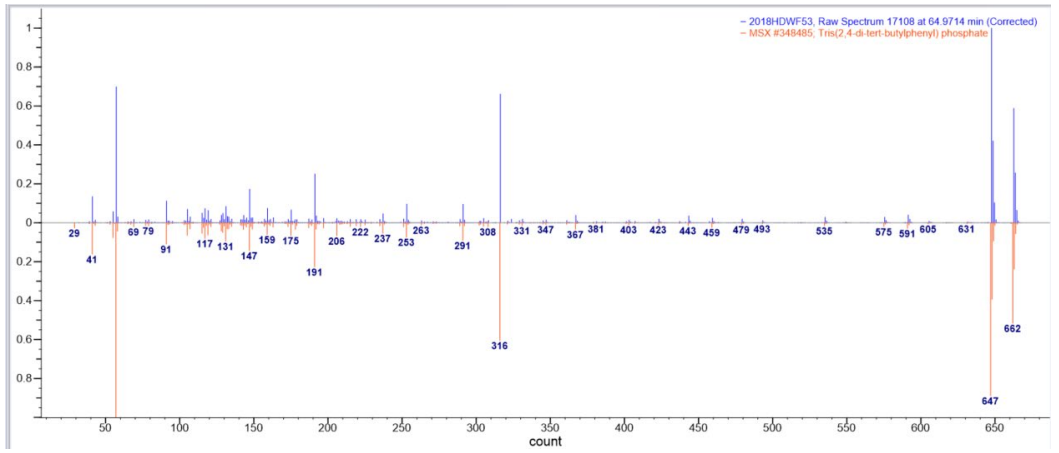
	Action	Result
3	Click the **Butterfly view** icon  to place the unknown and reference spectrum in the opposite Y-direction.	A target from a database is found.  Hits are initially sorted by the **Hit Quality Index (HQI)**. **Reverse Hit Quality Index (R.HQI)** is also calculated for each reference spectrum.

	Action	Result
4	**HQI** (Hit Quality Index)	The **HQI** value measures how close the reference spectrum is to that of the query. The default scale of **HQI** is 0-100. You can change the scale from **MineIt** > **File** > **Preferences** > **Hit List.**  Note: **Reverse Search** ignores peaks that are in unknown but not in reference. One scenario is that the unknown spectrum might be a mixture and the reference spectrum might be a component.

Optimized Corrections

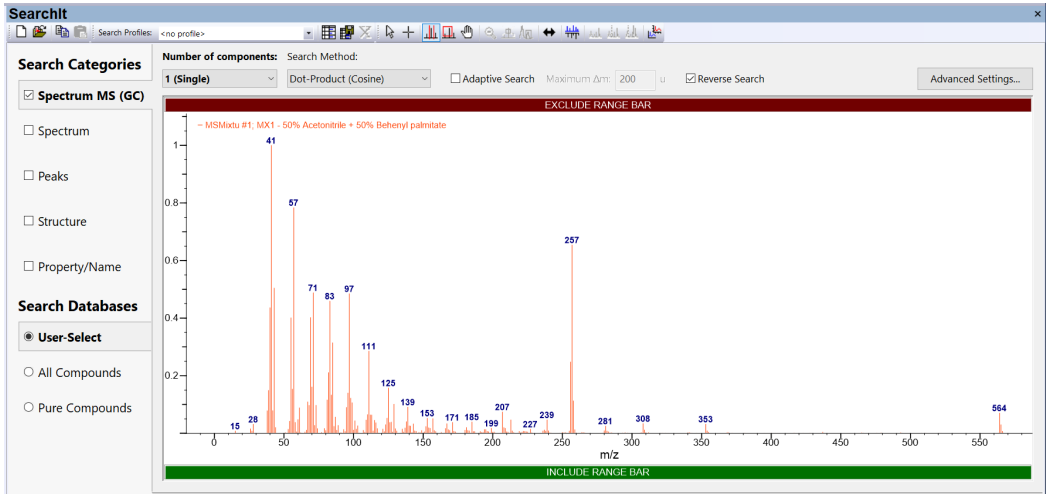
	Action	Result
1	Go back to **SearchIt.** Click  to start a new search. Click **Spectrum** and navigate to C:\Users\Public\Documents\Wiley\KnowItAll\Samples\MS Select the file **Mass defect correction example** Click **Open**. Click **OK** (use the default MS 17108)	

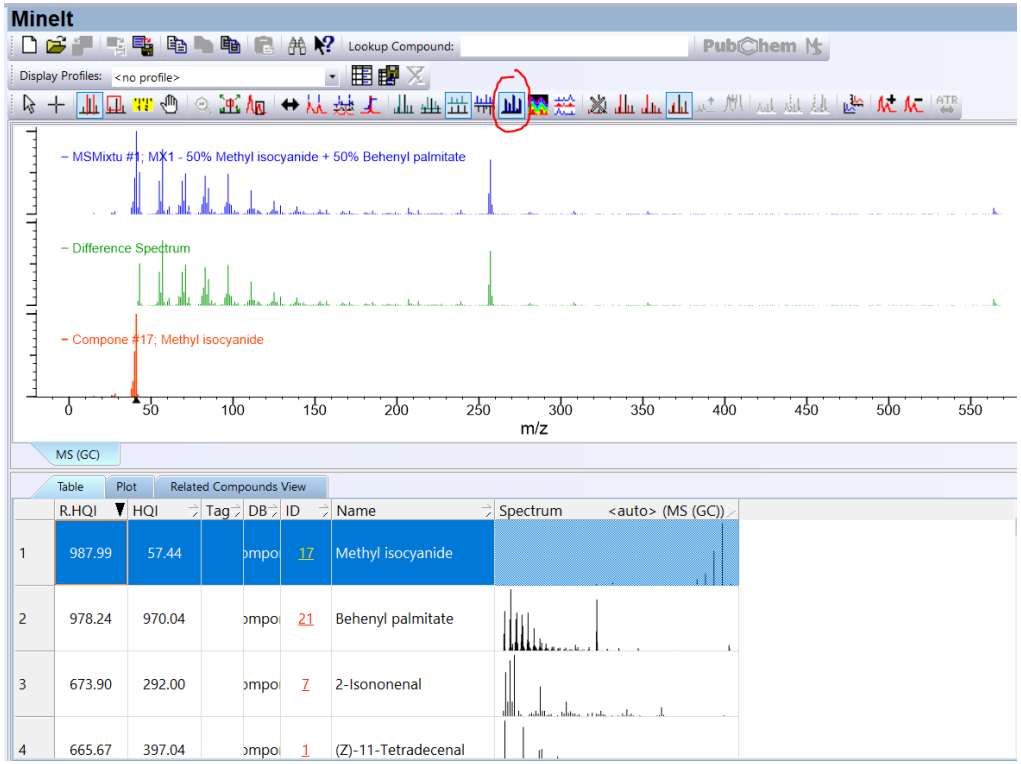
	Action	Result
2	Optimized Corrections for MS spectrum is default to be checked.  Click Advanced Setting button  Click **OK** to close this dialog Click the **Search** button	One can see that Mass Defection (the difference between a compound's exact mass and its nominal mass) is automatically applied. Many other criterion, such as Spectral Skewing (analyte's concentration changing during scan), can be applied by users choice to improve a search. 

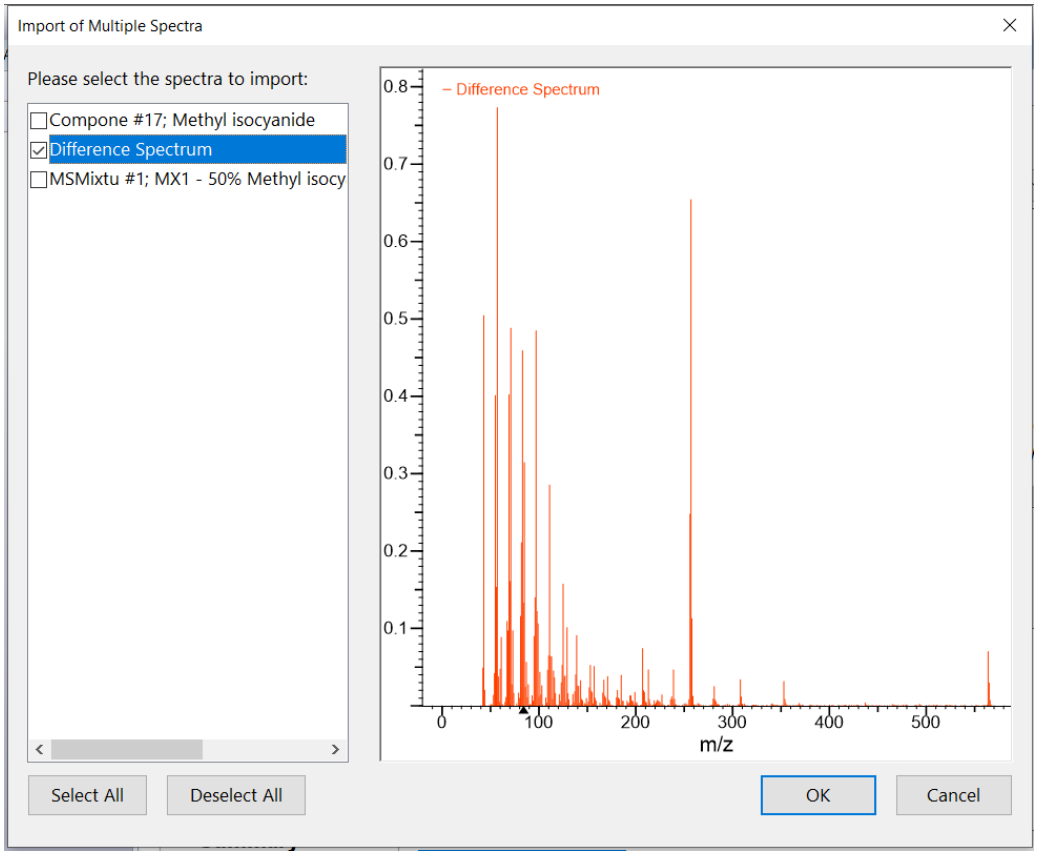
	Action	Result
3		 We can see analyte MS and reference MS are perfectly aligned at m/z 662.
4	Go back to **SearchIt** Check off **Optimized Corrections** **Search** again	 At high m/z range, analyte peaks and reference peaks are not aligned nicely.

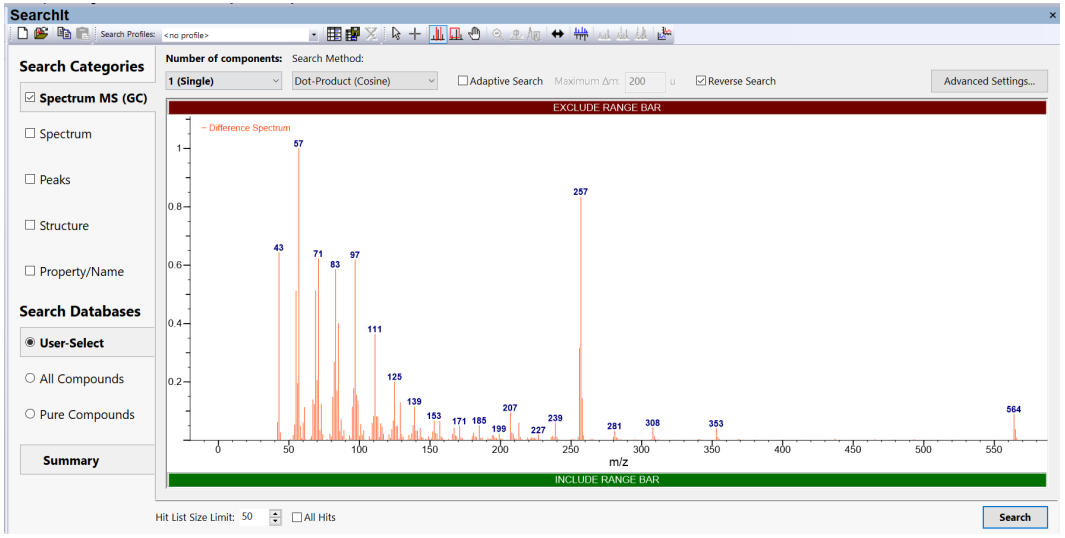
Reverse Search

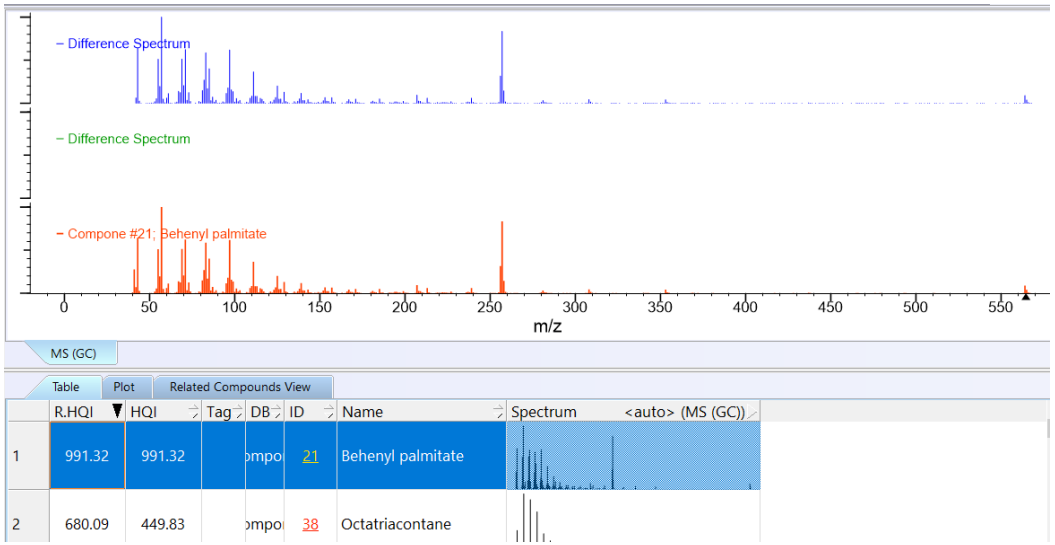
This search ignores peaks that are in unknown but not in reference

	Action	Result
1	Go back to **SearchIt** Click  to start a new search. Click **Spectrum** and navigate to C:\Users\Public\Documents\Wiley\KnowItAll\Samples\Mixture Analysis\MS Examples folder Select **MS Mixture of Two 1** to open Check **Reverse Search**.	
2	Click **User-Select** button Use button **Remove All** to clean current database selection Then, use **Select by Browsing** button to add the example database: C:\Users\Public\Documents\Wiley\KnowItAll\Samples\Mixture Analysis\MS Examples\Components.SDBX **Search**	

	Action	Result
3	In **MineIt**, make sure that **Subtract View** (circled) is selected. In the **spectrumPane**, the first row is the unknown, the last row is the reference spectrum, and the middle row is the difference between the two.	 The hits are initially sorted by the **Reverse Search HQI (R.HQI)**.

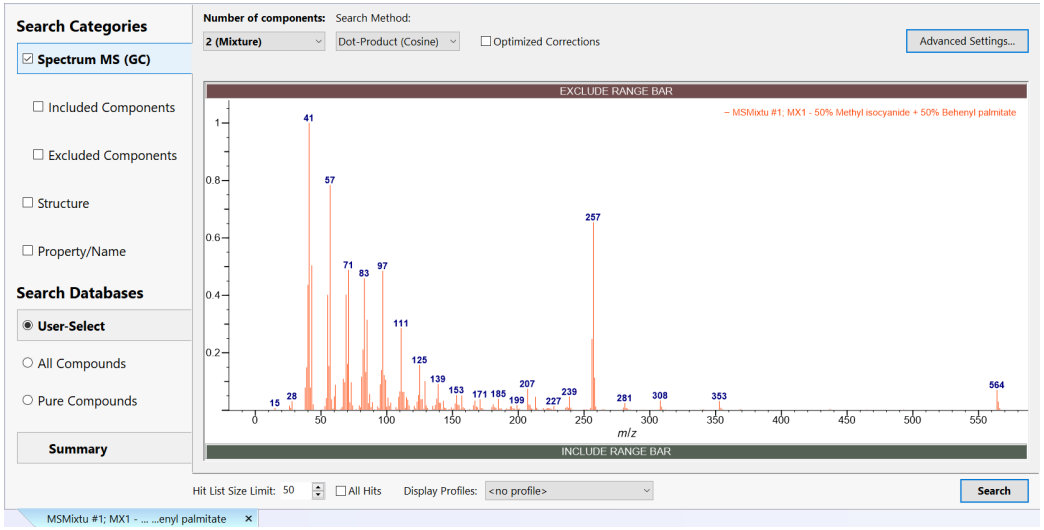
	Action	Result
4	Transfer spectra in **spectrumPane** to **SearchIt** using the Transfer to: menu at the top of the application. Select to transfer only the **Difference Spectrum**. Click **OK**. When prompted by **SearchIt**, choose **Start a new search**.	

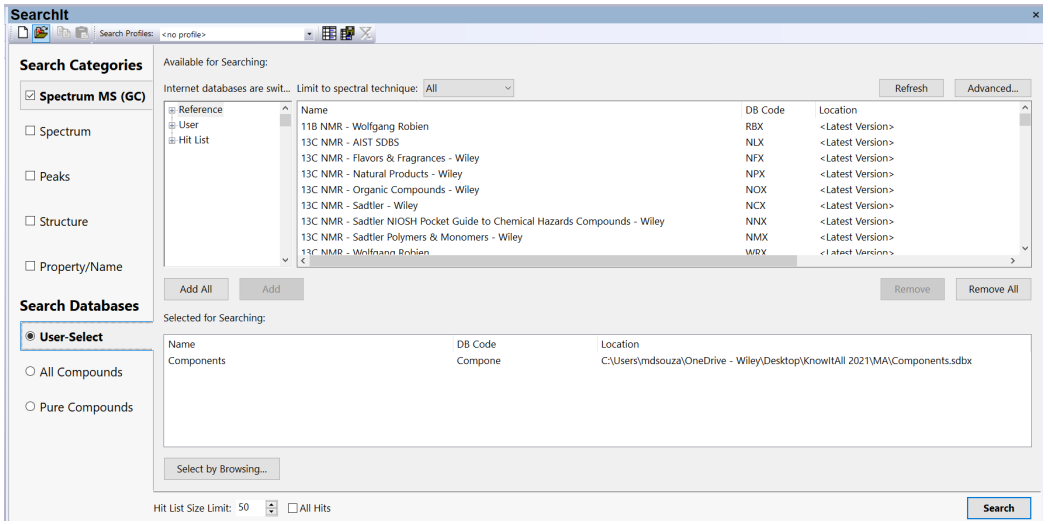
	Action	Result
5	Click **Search**.	

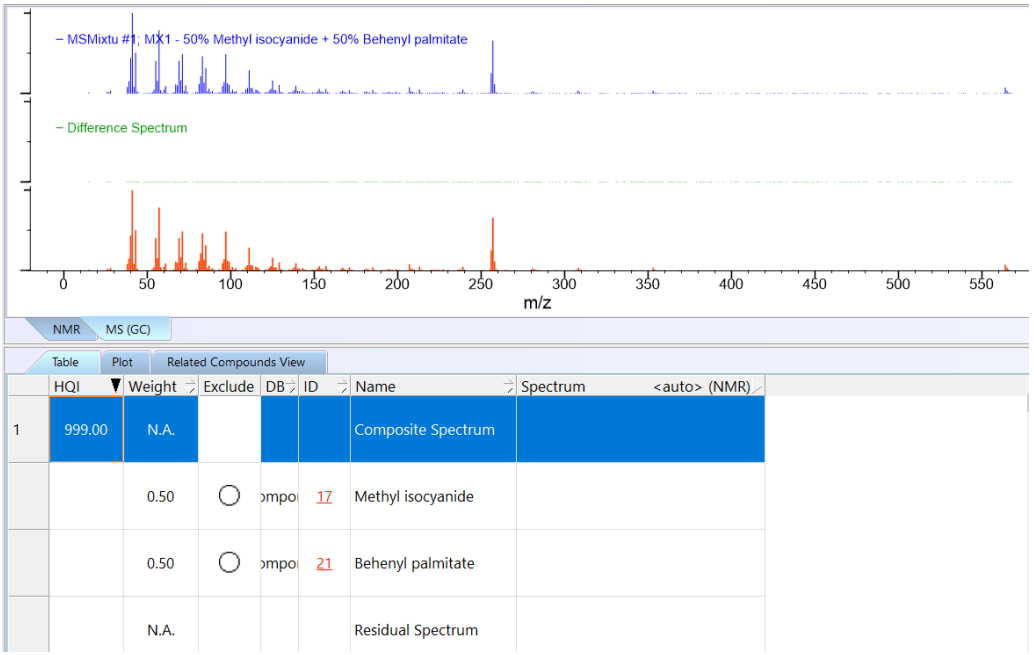
	Action	Result
5		 A second component is identified. Notice that the **Difference Spectrum** is empty, which means there are no more components.

Mixture Analysis

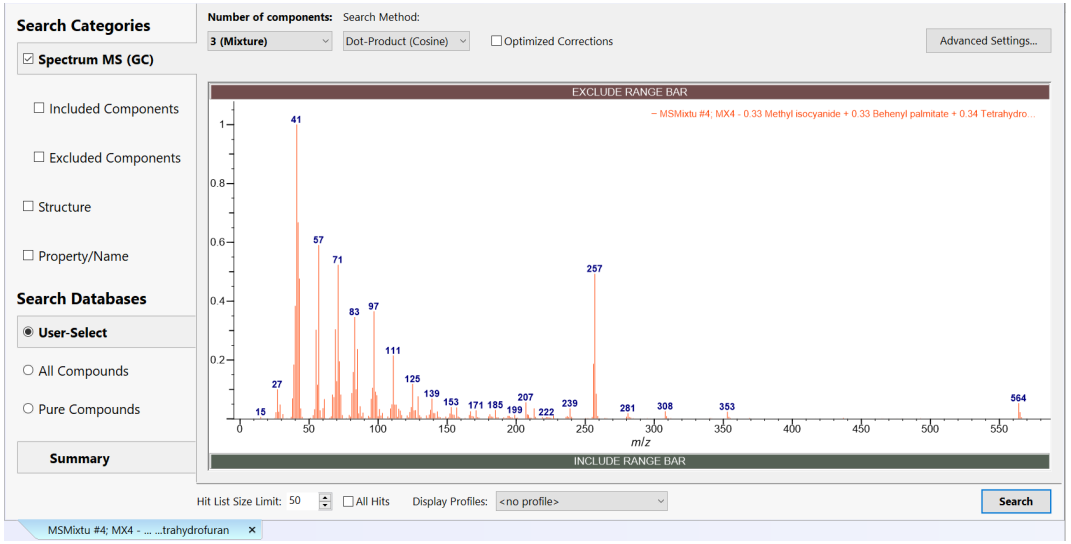
A mixture wherein two-component MS spectra have limited overlap, and one of them has a large MS range

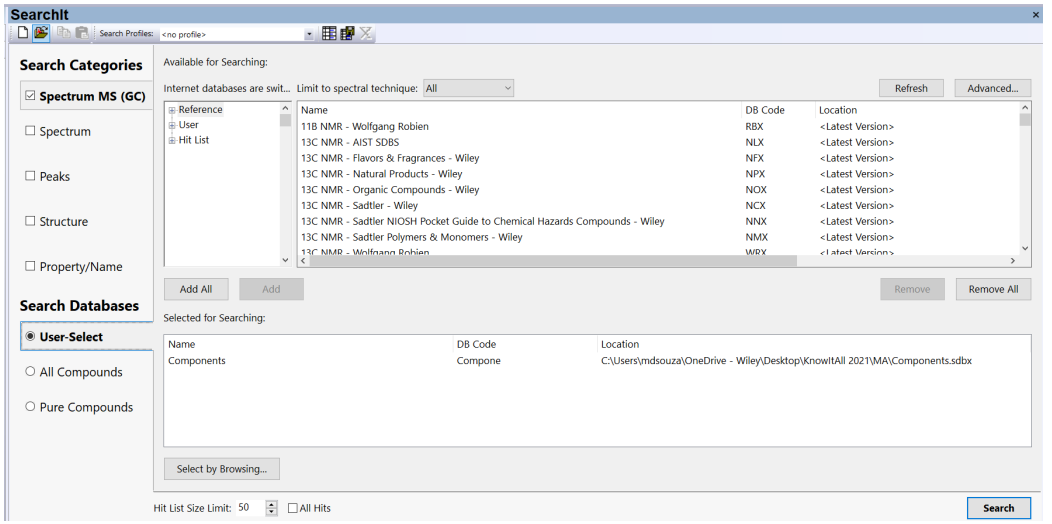
	Action	Result
1	Go back to **SearchIt**. Start a new search by clicking  . Select **MS Mixture of Two 1**. Ensure **Reverse Search** or **Adaptive Search** is unchecked. Use the drop-down menu to set **the Number of components** to **2**.	

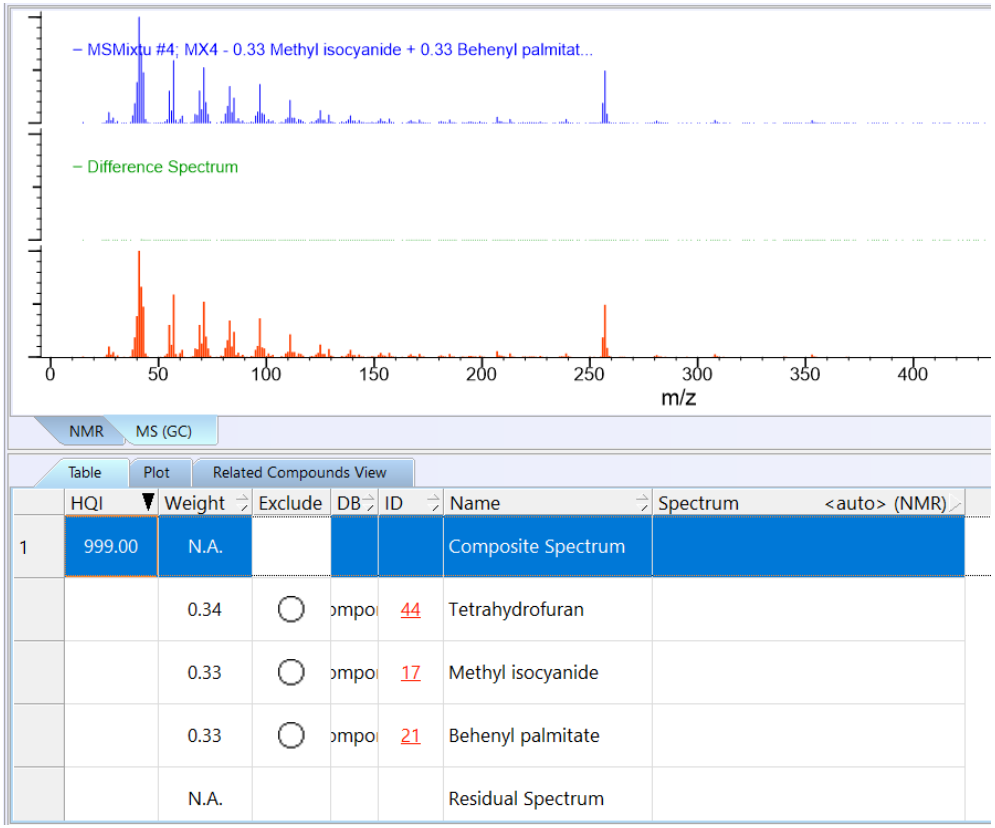
	Action	Result
2	Click on **User-Select**. Make sure **Components** database is selected to use. Click **Search**.	

	Action	Result
		 KnowItAll presents a 2-component search result. The top spectrum is unknown; the bottom spectrum is the composite spectrum of 2 components. The middle spectrum is the difference between the two. In this case, it is next to nothing, indicating that there are no more residual peaks. **Note**: This is an example wherein the two-component MS spectra have limited overlap, but one of them has large MS range.

A more complex example

	Action	Result
1	Go back to **SearchIt**. Start a new search by clicking  . Select the file **MS Mixture of Three**. Use the drop-down menu to set the **Number of components** to **3 (Mixture)**.	

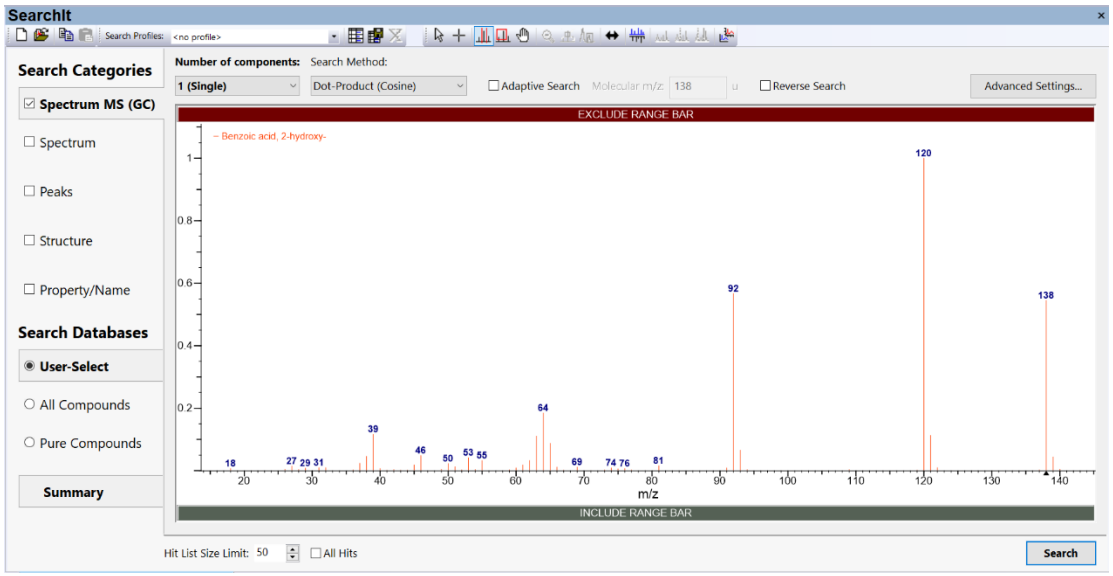
	Action	Result
2	Click on **User-Select**. Make sure the **Components** database is selected for use. Click **Search**.	

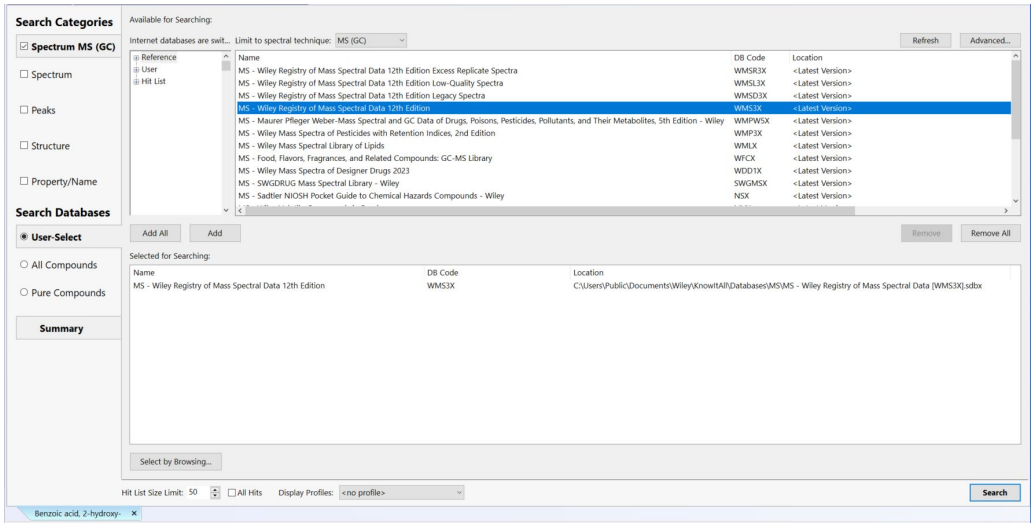
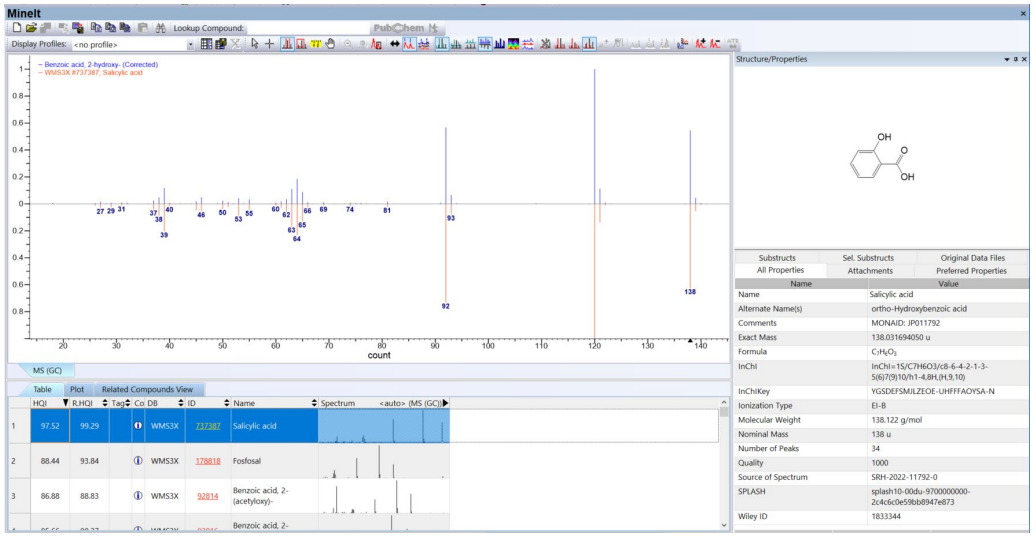
	Action	Result
		 KnowItAll presents a 3-component search result. The top spectrum is unknown; the bottom spectrum is the composite spectrum of 3 components. The middle spectrum is the difference between the two. In this case, it is next to nothing, indicating that there are no more residual peaks. This process accomplishes many steps in one. It also avoids negative peaks from spectral subtraction.

Adaptive Search

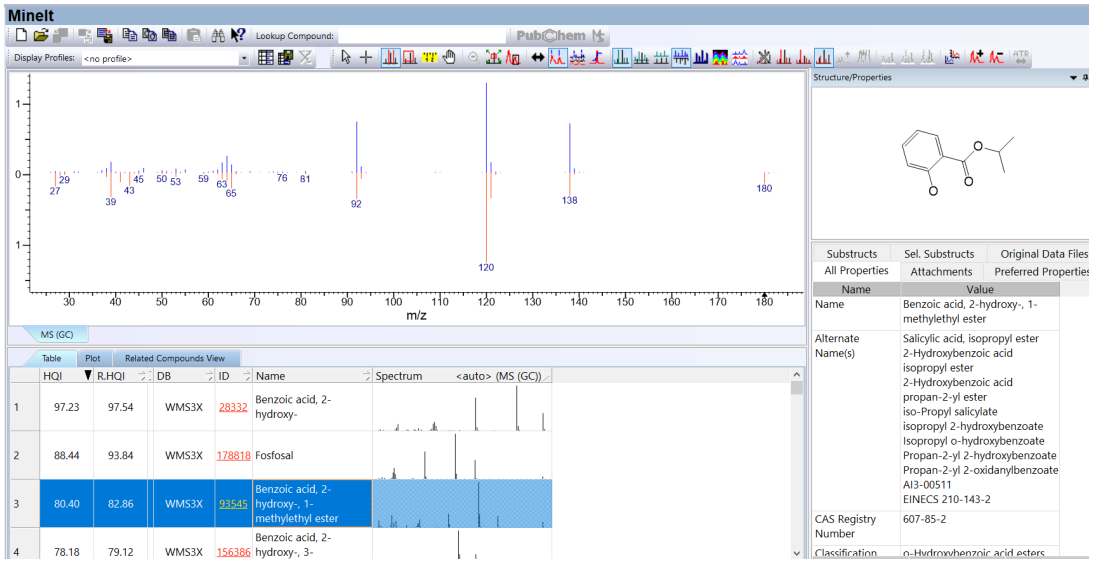
This search finds similar compounds where a fragment group can be present or missing compared to the unknown. The presence/absence of a fragment causes some peak positions in reference MS differing to that of unknown by a delta mass (Δm). KnowItAll shifts some peaks by the Δm to achieve a better matching score. Because of the better matching score, similar compounds come atop of the hit list. To clearly mark the shifts done by Adaptive Search, dotted lines are used to show reference spectrum before and after shifting in the pop-up window when you click on the (i) button in a hit.

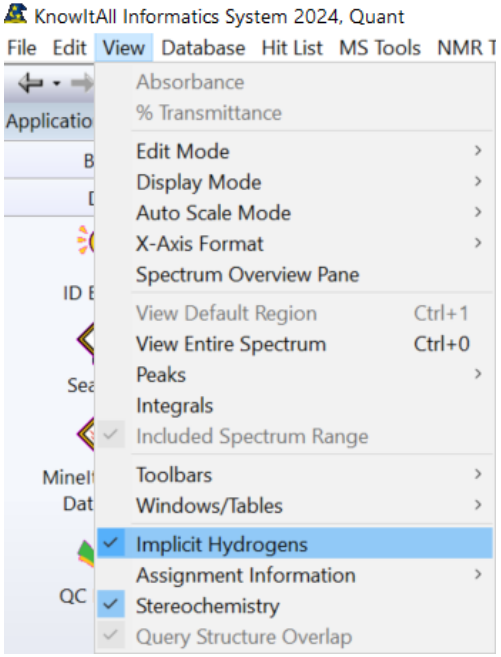
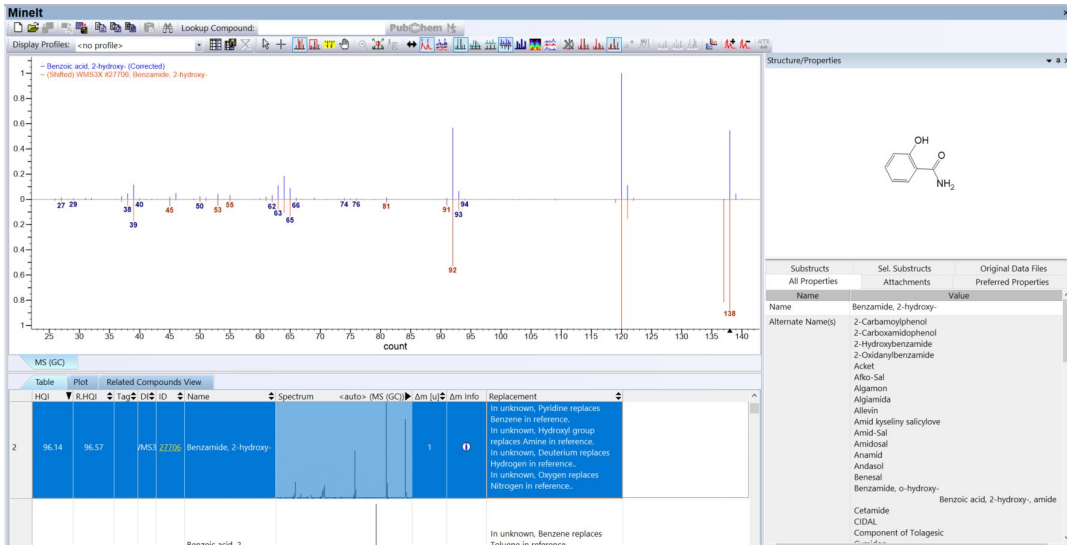
Example wherein exact mass is in a spectrum

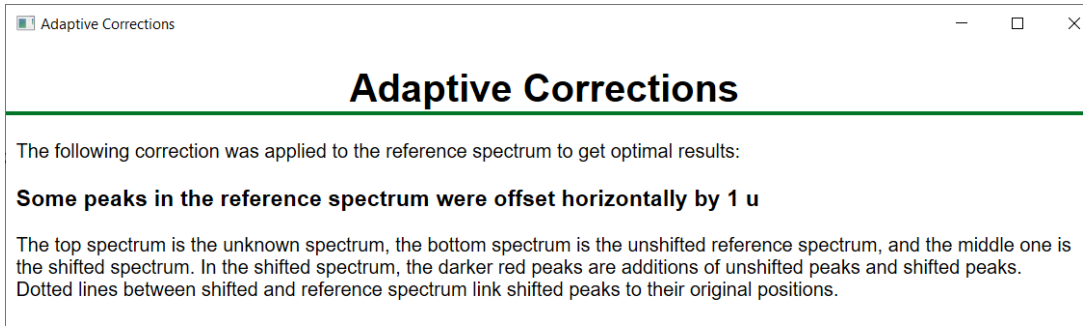
	Action	Result
1	Start a new search by clicking  . Click **Spectrum** to navigate to C:\Users\Public\Documents\Wiley\KnowItAll\Samples\MS folder Select the file **2-Hydroxybenzoic acid**. Use the drop-down menu to set the **Number of components** to **1 (Single)**. Ensure **Adaptive Search** or **Reverse Search** is unchecked.	 This spectrum file contains a molecular ion mass of 138. KnowItAll will use this value for the Adaptive Search.

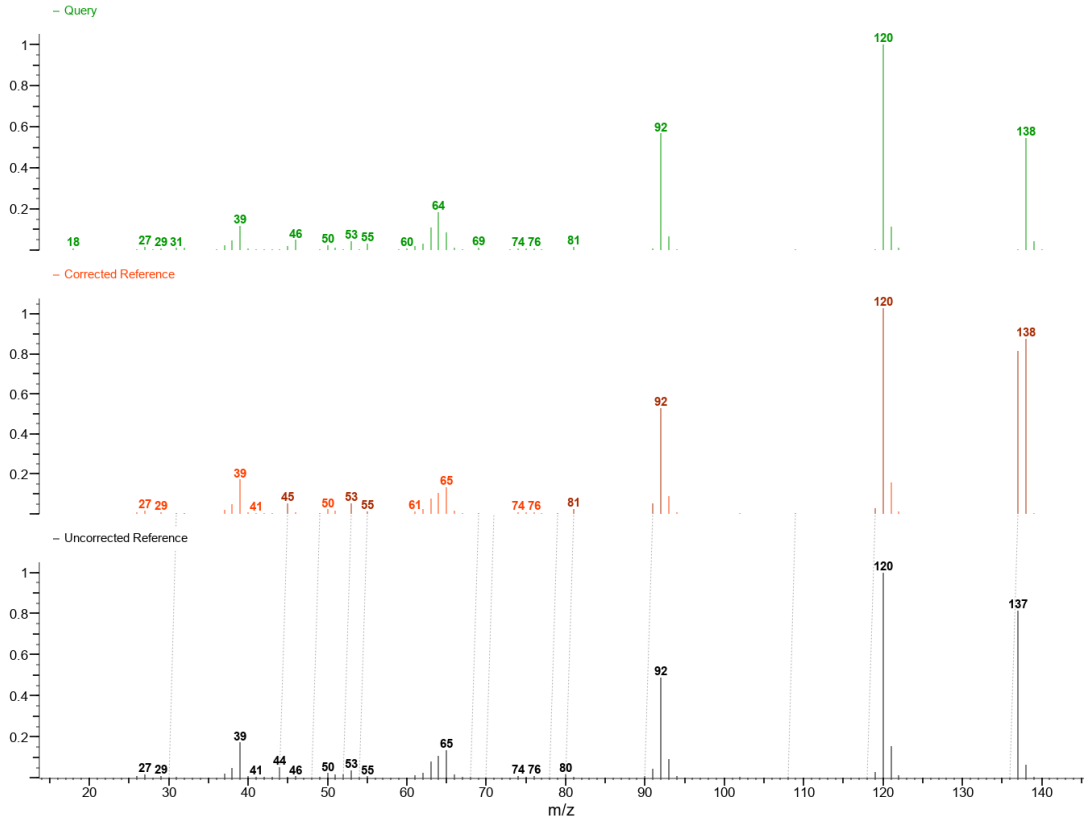
	Action	Result
2	Click **User-Select** button Use **Remove All** to clean selected databases. Use **Add** to add MS (GC) databases by codes: **WMS3X**. Click **Search**.	
3	Click the **Butterfly view** icon  .	

WILEY

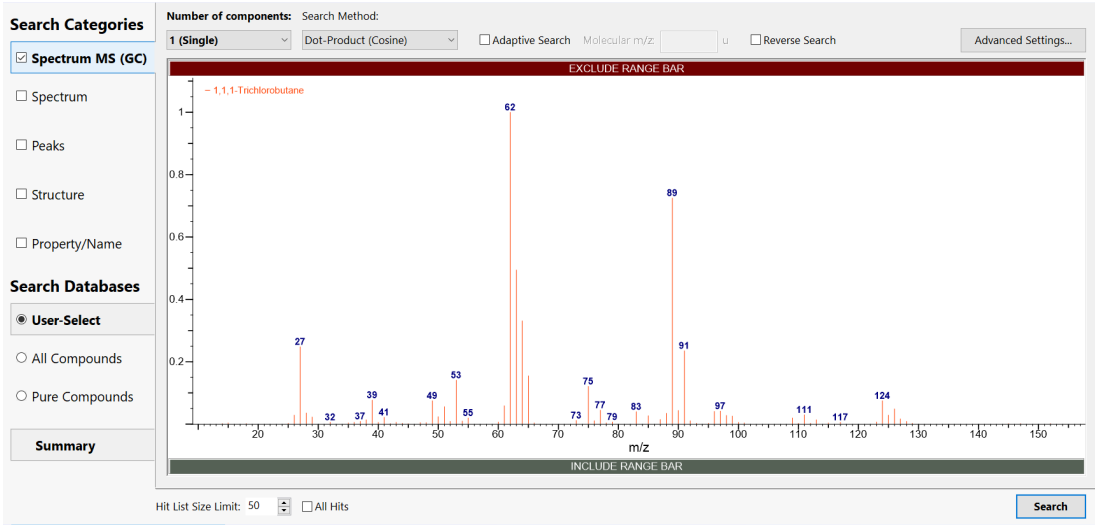
	Action	Result
4	Highlight the 2nd then the 3rd hit.	 Assume that we do not have the first hit in our database. The HQI values for the 2nd and 3rd hits are not totally convincing that they are good matches.
5	Go back to **SearchIt**. Click **Spectrum MS(GC)** Check **Adaptive Search**. Click **Search**.	 • If a spectrum file contains the molecular ion mass, it will be displayed in the “Molecular *m/z*:” box, and a solid triangle marks this position in the spectrumPane. • You can type in an appropriate value as well. This value is used to assist Adaptive Search. • Or, you can ask KnowItAll to estimate a molecular mass for you by clicking the bulb icon. • However, if molecular ion mass is unknown, this box would be empty. KnowItAll Adapative Search will estimate this value from the input unknown MS spectrum.

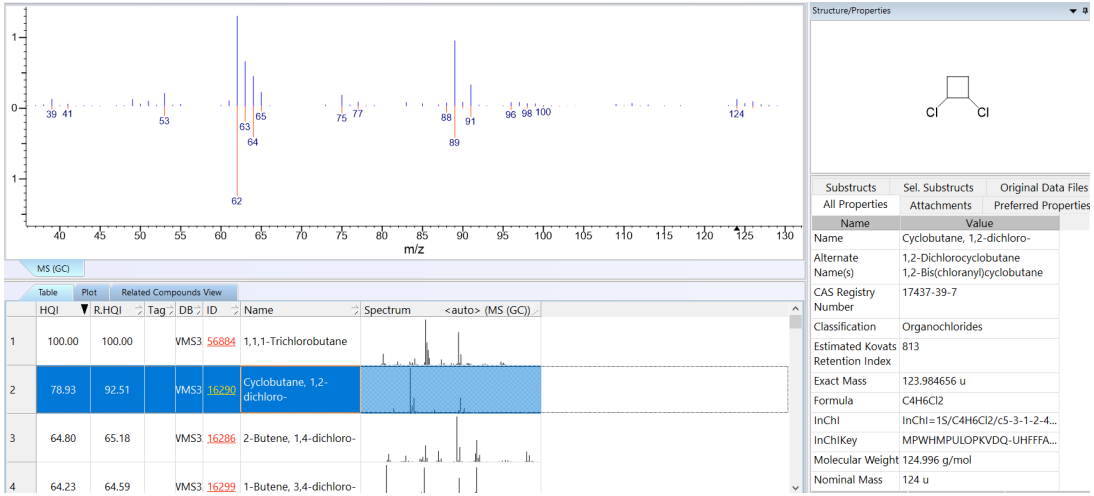
	Action	Result
6	Ignore the 1st hit and move to the 2nd hit. Make sure KnowItAll displays implicit Hydrogens, one need to go to **View** menu and check **Implicit Hydrogens** 	 The result looks good. **Δm** of 1 suggests a few possibilities wherein the unknown's molecular ion mass (m/z) is 1 unit more than that of the reference. One of them is "**In unknown, Oxygen replaces Nitrogen in reference**." Columns of interest: • **Δm**: the molecular ion mass of unknown minus that of the reference • **Δm Info**: details on reference peak shifts to make the match • **Replacement**: suggestions of ***possible*** group exchange which would cause the difference between unknown and reference molecular ion masses

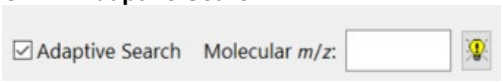
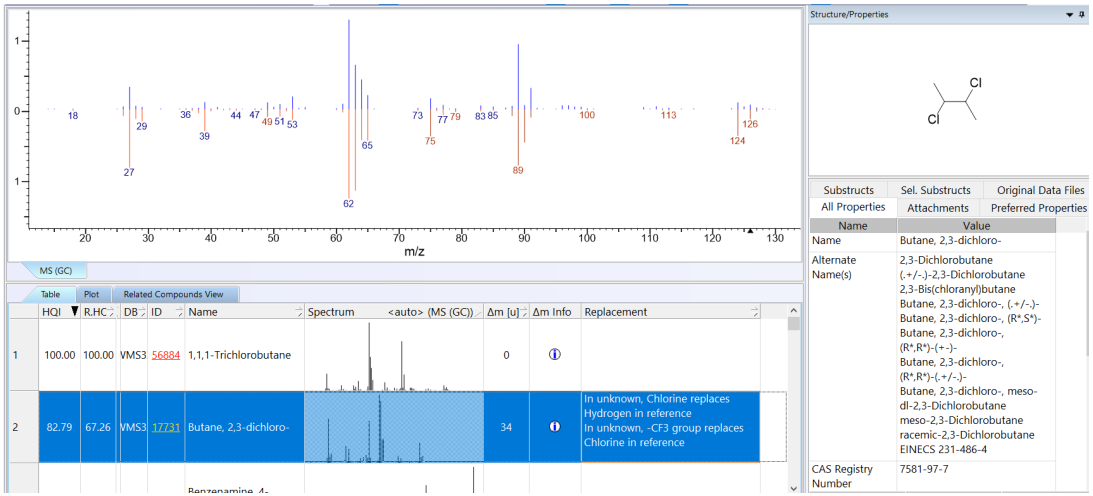
	Action	Result
7	Click **(i)** in the **Δm Info** column. This brings up the **Adaptive Corrections** html page which explains how selective peaks have been shifted to obtain good **Hit Quality Index (HQI)**.	

	Action	Result
8		 • The top spectrum is the unknown • The bottom spectrum is the unshifted reference • The middle “spectrum” contains shifted peaks • Dotted lines indicate what peaks have been moved by an Δm • In the middle “spectrum,” darker red peaks are the additions of existing peaks and moved ones

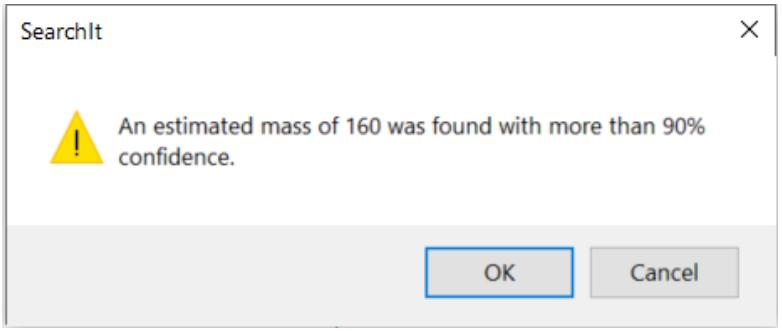
Example wherein exact mass is NOT in a spectrum

	Action	Result
1	Start a new search by clicking  . Click **Spectrum** and navigate to C:\Users\Public\Documents\Wiley\KnowItAll\Samples\MS folder. Select file **1,1,1-Trichlorobutane**. Use the drop-down menu to set the **Number of components** to **1 (Single)**. Ensure **Adaptive Search** or **Reverse Search** is unchecked. Click **User-Select**. Use **Remove All** to clean selected databases. Use **Add** to add MS (GC) databases by codes: **WMS3X.** Click **Search**.	 As we can see, this spectrum does not contain molecular ion mass. KnowItAll will estimate this value and use it for **Adaptive Search**.

	Action	Result
2	Click the **Butterfly view** icon  . Go to the 2nd hit since the 1st hit is the exact match (pretend that we do not have the exact match in our databases).	 We are not sure by looking at the HQI that we have a hit that is structurally similar to our unknown.

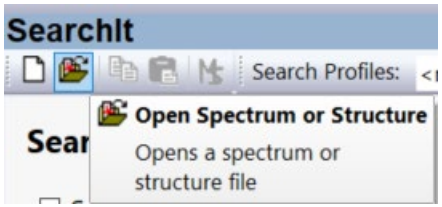
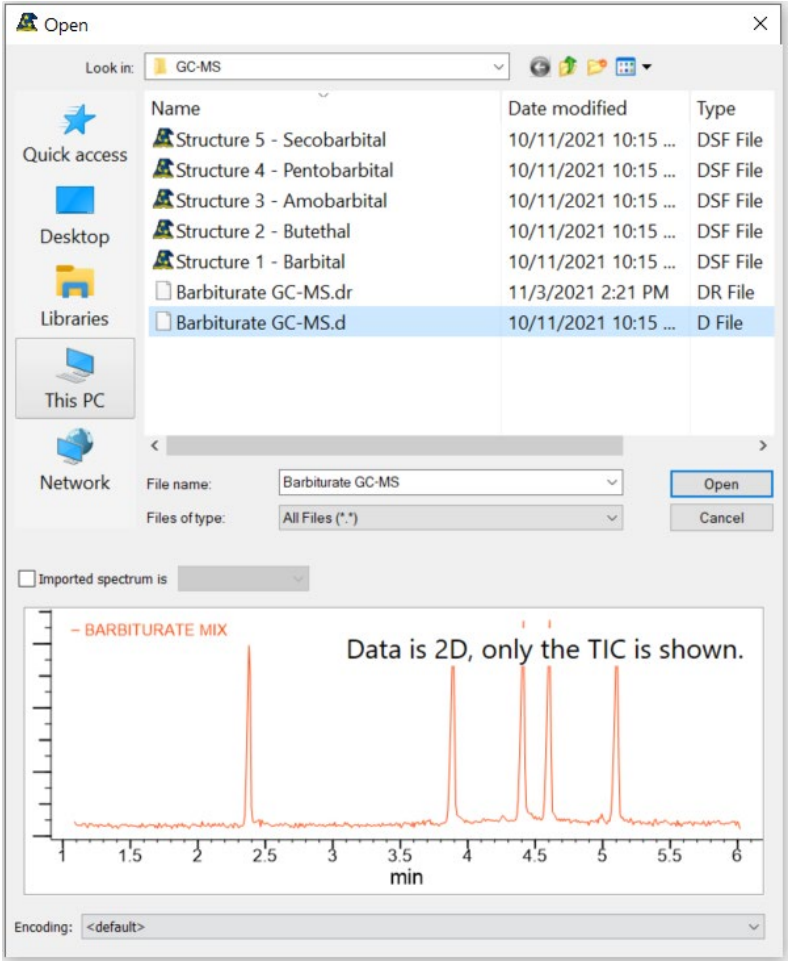
	Action	Result
3	Go back to **SearchIt** Check **Adaptive Search**  **Search** In **MineIt,** Click the **Butterfly view** icon  . Go to the 2nd hit (1st being the unknown which is already in a database).	 One of the suggestions in the **Replacement** cell makes sense. The unknown should have 1 more Chlorine for a delta mass of 34.

Alternatively, one can use the molecular ion mass estimation icon to estimate the molecular ion mass:

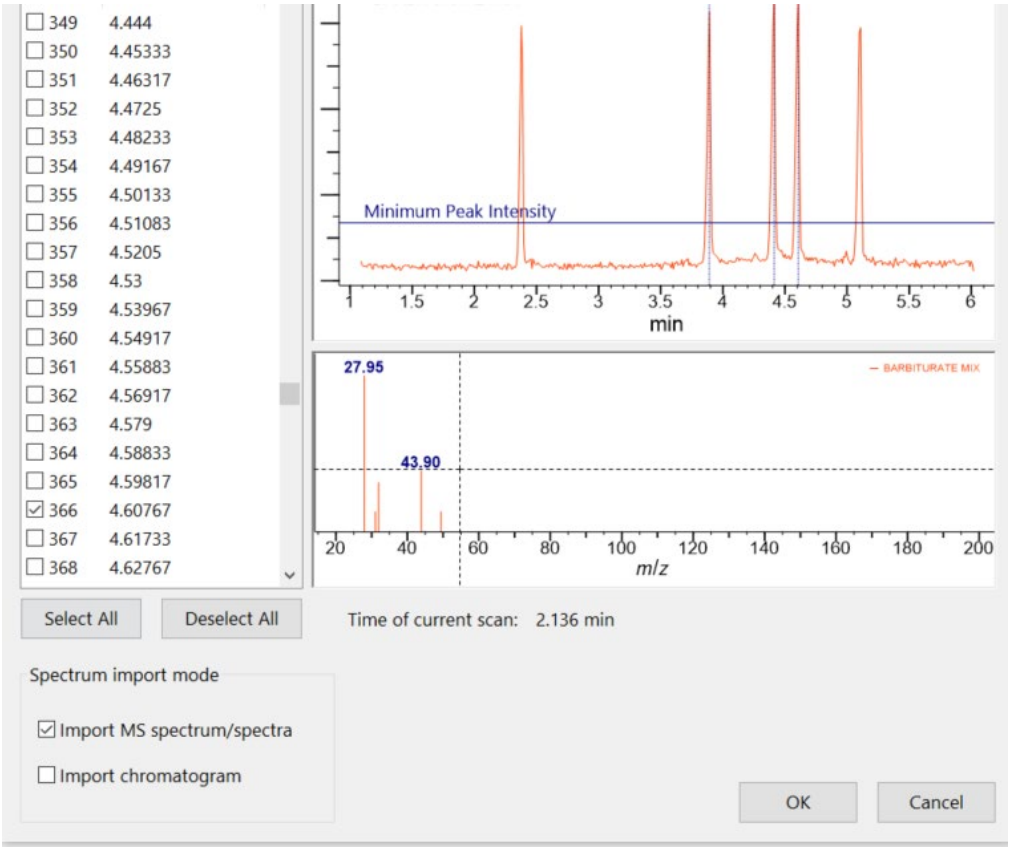
	Action	Result
1	Go Back to **SearchIt**. Click the icon Let the calculation complete	
3	Click **OK** button	The estimated molecular ion mass value fills the box 

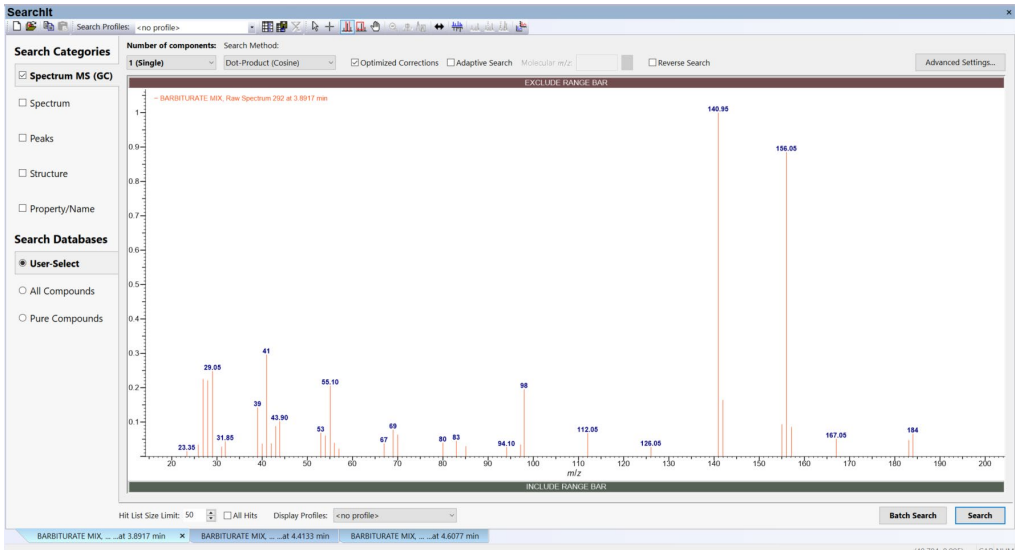
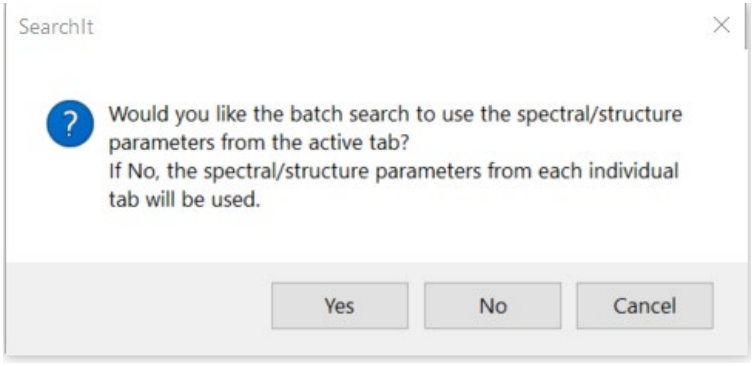
NOTE: KnowItAll runs an adaptive search of the unknown against all of our MS databases. The resulting hit list is then analyzed in steps. The mass of every hit list entry is calculated as the nominal mass of the compound in the database record plus the Δm found for the match. Matches of equal mass are then grouped together into clusters. The higher the found HQI, the higher the score for an individual match. Scores for clusters are then calculated as a combination of individual match scores with additional information such as the number of entries in the cluster and the separation from the next best cluster. The cluster with the best score determines the found mass. As side information of this procedure, information on confidence that the found mass will be correct is reported.
The confidence values found by the algorithm were determined by running statistics with thousands of very diverse compound spectra run against our MS data.

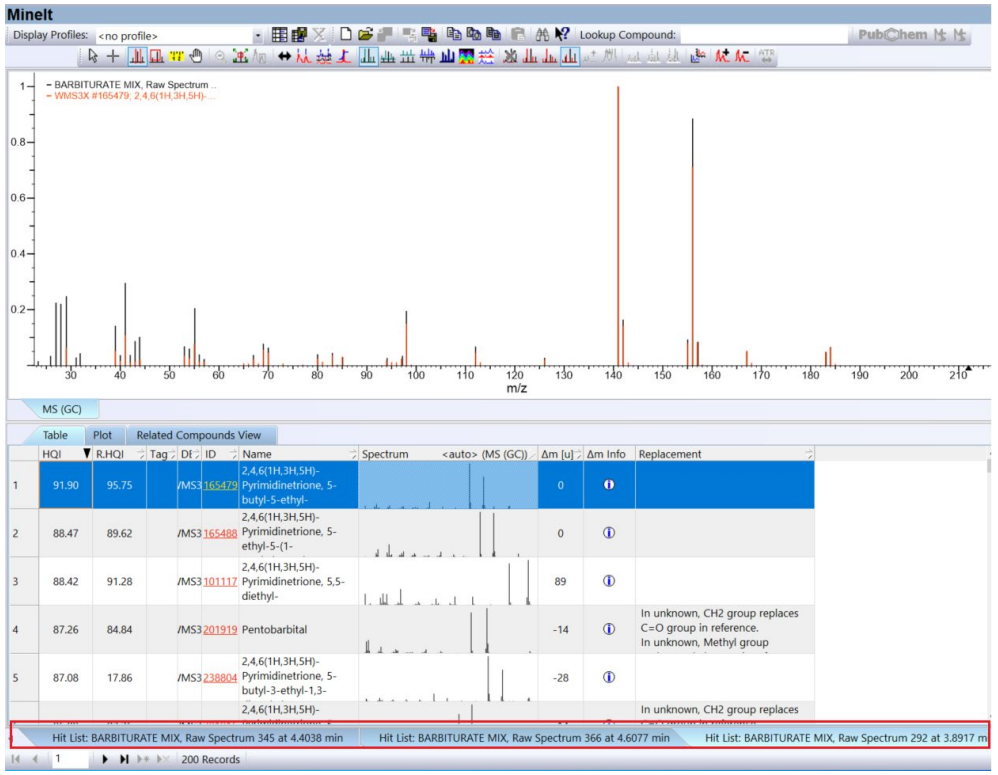
Simultaneous Multiple MS Spectra Search

	Action	Result
1	In **SearchIt**, select **Open Spectrum or Structure**  Navigate to C:\Users\Public\Documents\Wiley\KnowItAll\Samples\GC-MS\Barbiturate GC-MS.d Click **Open** in previous dialog	

WILEY

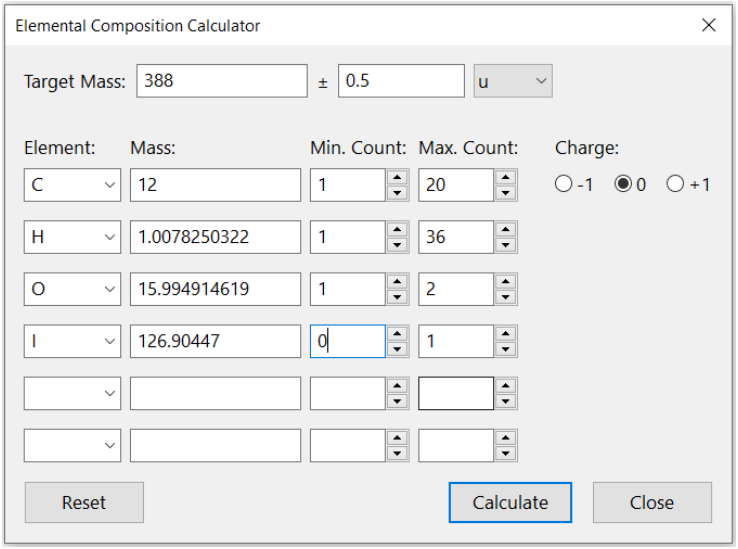
	Action	Result
2	Pick multiple MS spectrum in the **MS Spectra Scan Selection** dialog For example, 292, 346, 366 Click **OK** button	

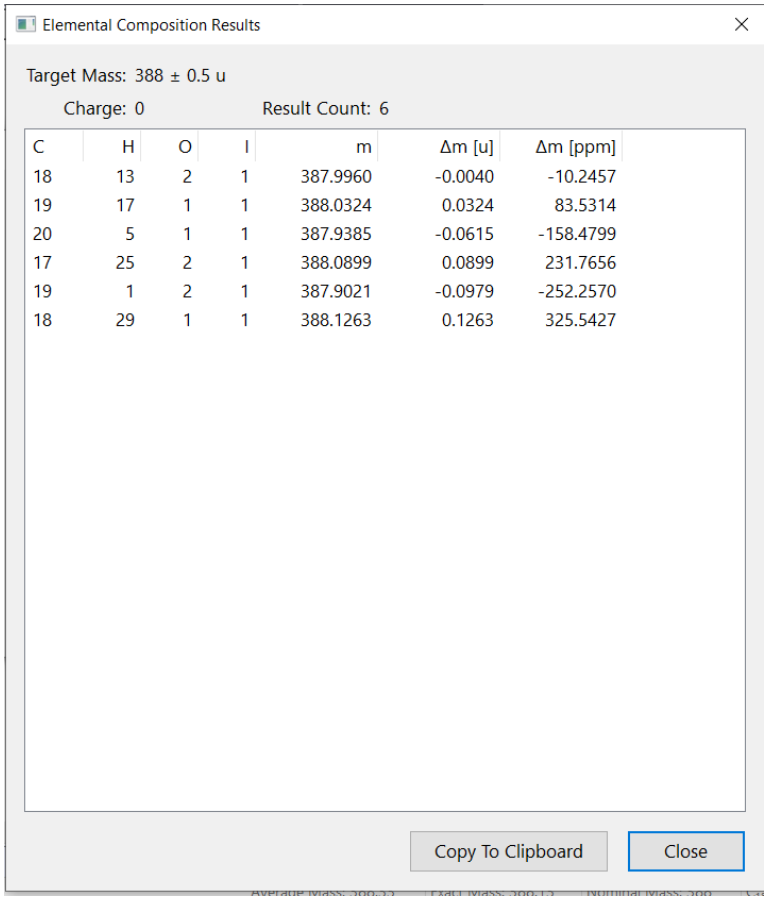
	Action	Result
3	Multiple search tabs are created	
4	One can either search these MS spectra one by one by clicking the **Search** button in each tab Or use **Do Batch Search** button and reply **Yes** when prompted	
5		Multiple hits are transferred to **MineIt** if **Do Batch Search** was selected

	Action	Result
		

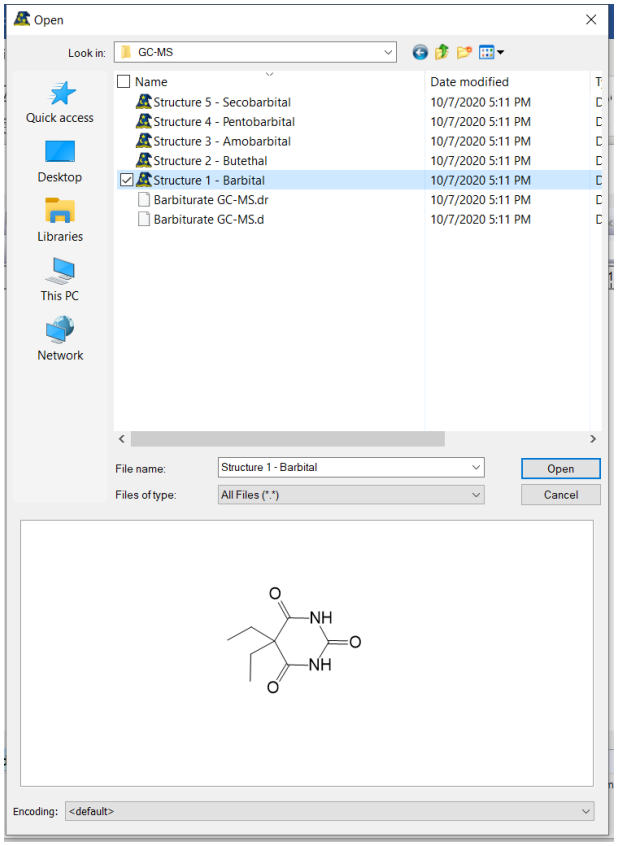
Other Tools for Mass Spectrometry

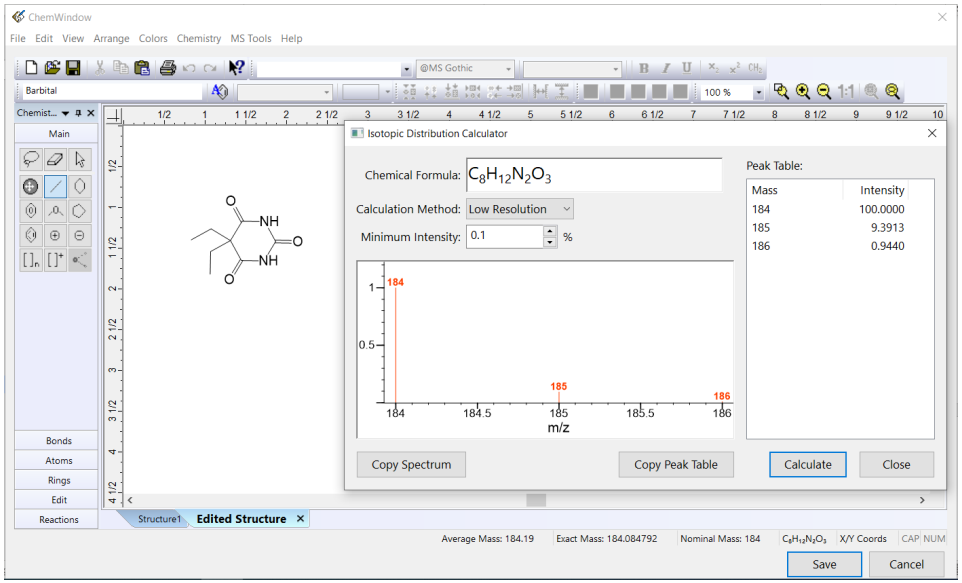
Elemental Composition

	Action	Result
1	From the **Menu,** navigate to **MS Tools > Calculate Elemental Composition**.	 You can fill in this dialog with target mass and elemental information.

	Action	Result
2	Click **Calculate**.	 KnowItAll provides combinations of these elements.

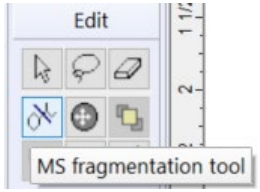
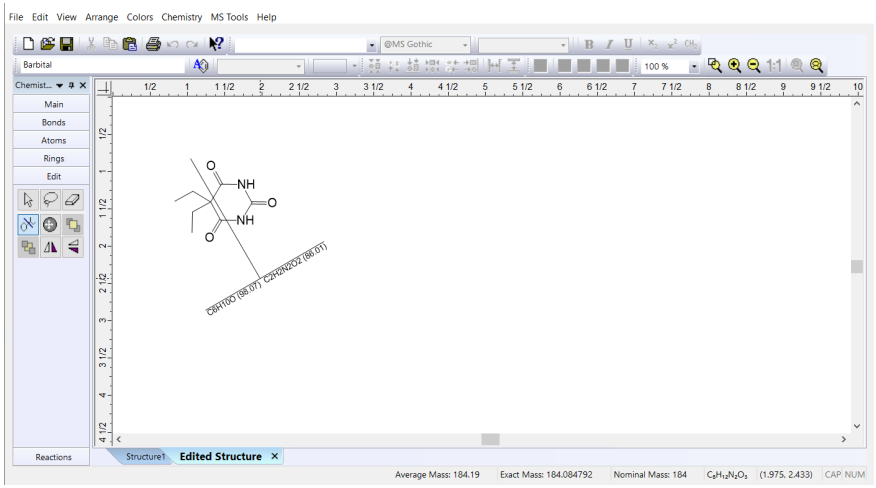
Isotopic Distribution

	Action	Result
1	Double-click the structure pane to bring up ChemWindow or use the Transfer to: bar and select **ChemWindow** at the top of the application. Click **File > Open**. Select a structure from the **Samples > GC-MS** folder.	

	Action	Result
2	Navigate to **MS Tools** > **Calculate Isotopic Distribution**. Click **Calculate**.	

Molecular Fragmentation

In ChemWindow, you can use the MS fragmentation tools to view possible fragments and corresponding masses.

	Action	Result
1	In the **Edit** toolbar on the left, click on the **MS fragmentation tool**.  Point to a position, then drag a line—this is the fragmentation line.	

Solid Triangle Marks Nominal Mass

- In the spectrum display, KnowItAll marks the nominal mass of the structure that corresponds to the spectrum. This mass is shown as a black triangle.
- When importing a spectrum from a data file, a number of import formats define fields for molecular *m/z* (also called precursor *m/z* or base peak *m/z* in some cases) and the charge of the molecular ion. To convert from molecular *m/z* to exact mass, the following formulae are used:
  - For positive charges:
    - M_{exact} = (Mz – M(H) + M(e)) * charge, where M(H) is the mass of a hydrogen atom, and M(e) is the mass of an electron.
  - For negative charges:
    - M_{exact} = (Mz - M(e)) * (-charge).
  - If no charge is defined, a default charge of +1 is assumed.
- If a data file does not have the molecular *m/z* field defined, the exact mass is calculated from the formula field, if available.