KnowItAll Software Training

GC-MS Analysis Using KnowItAll MS Expert And SearchIt

WILEY

GC-MS Analysis By Automatic Deconvolution

How to Perform GC-MS Analysis by Automatic Deconvolution

Purpose

These exercises demonstrate how to use KnowItAll MS Expert to automatically deconvolute/analyze GC-MS.

Objectives

These exercises will teach you:

- How to use KnowItAll MS Expert to auto-deconvolute GC-MS data into chemical component MS spectra which are automatically searched against millions of references.
- How to generate reports.

Background

GC-MS data are information rich. Analysis can be time consuming, especially when examining complex analytes. We present a computer system that combines fast, flexible automated deconvolution, automatic database search to identify knowns and unknowns. Novel compounds can be identified, and structural characteristics deduced from applying the MS Adaptive search that uses fragmentation and structural data to propose likely structural details of the unknown.

Training Files Used in This Lesson

- C:\Users\Public\Documents\Wiley\KnowItAll\Samples\MS Expert folder files

KnowItAll Applications Used

- KnowItAll MS Expert

GC-MS Deconvolution Algorithms

Our system follows individual *m/z* values across multiple spectra and extracts a pure spectrum from the data for each individual component while trying to separate components with overlapping *m/z* value peaks. If accurate *m/z* value data are available and the user selects this data to use it instead of unit *m/*z values, the chosen instrument accuracy (automatic, ppm, or fixed value) is used to determine the correct accurate *m/z* values throughout the entire GC-MS analysis. The *m/z* values in the raw data are converted into correct accurate values based on the closest value found by taking the instrument resolution into account. The corrected *m/z* values form the basis of the following deconvolution.

During the deconvolution step, individual *m/z* values are traced across multiple raw spectra, and a component spectrum is extracted while trying to separate components with overlapping *m/z* value peaks. The details of the algorithm are summarized to a large extent by the papers listed below[1-4].

Additional steps are added to automatically detect components with low intensity Reconstructed Total Ion Current (RTIC) chromatographic peaks as long as they can well be separated from neighboring components.

Although the algorithm is very complex, the details of the algorithm are summarized to a large extent by the following papers:

1. S. E. Stein. An Integrated Method for Spectrum Extraction and Compound Identification from Gas Chromatography/Mass Spectrometry Data. *J Am Soc Mass Spectrom* 1999, **10**, 770 –781.
2. R. G. Dromey, M. J. Stefik, T. C. Rindfleisch, A. M. Duffield. Extraction of Mass Spectra Free of Background and Neighboring Component Contributions from Gas Chromatography IMass Spectrometry Data. *Analytical Chemistry*, 1976, **48(9)**, 1368-1375.
3. J. E. Biller, K. Biemann. Reconstructed Mass Spectra, A Novel Approach For The Utilization Of Gas Chromatography-Mass Spectrometer Data. *Analytical Letters* 1974, **7**, 515-28.
4. B. N. Colby. Spectral Deconvolution for Overlapping GC/MS Components. *J Am sot Mass Spectrom* 1992, **3**, 558-562.

MS Spectral Comparison 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

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

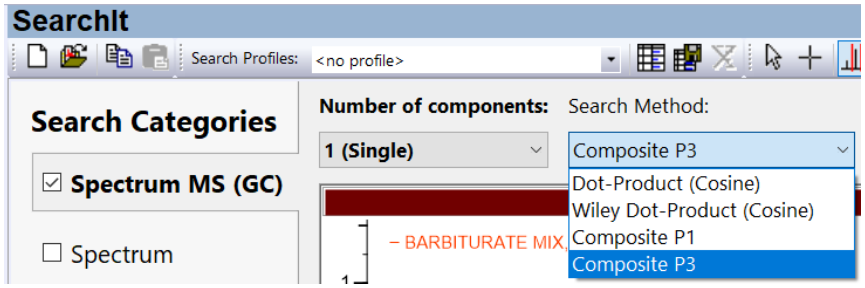
Spectrum search type – identity (normal)
Presearch – default
Included Libs – MainLib
Apply limits – unchecked
Use constraints – unchecked

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

Samokhin, K. Sotnezova, V. Lashinb, I. Revelskya. Evaluation of mass spectral library search algorithms implemented in commercial software. *J. Mass Spectrom.* 2015, **50**, 820-825.

KnowItAll has 4 different algorithms: 

Where

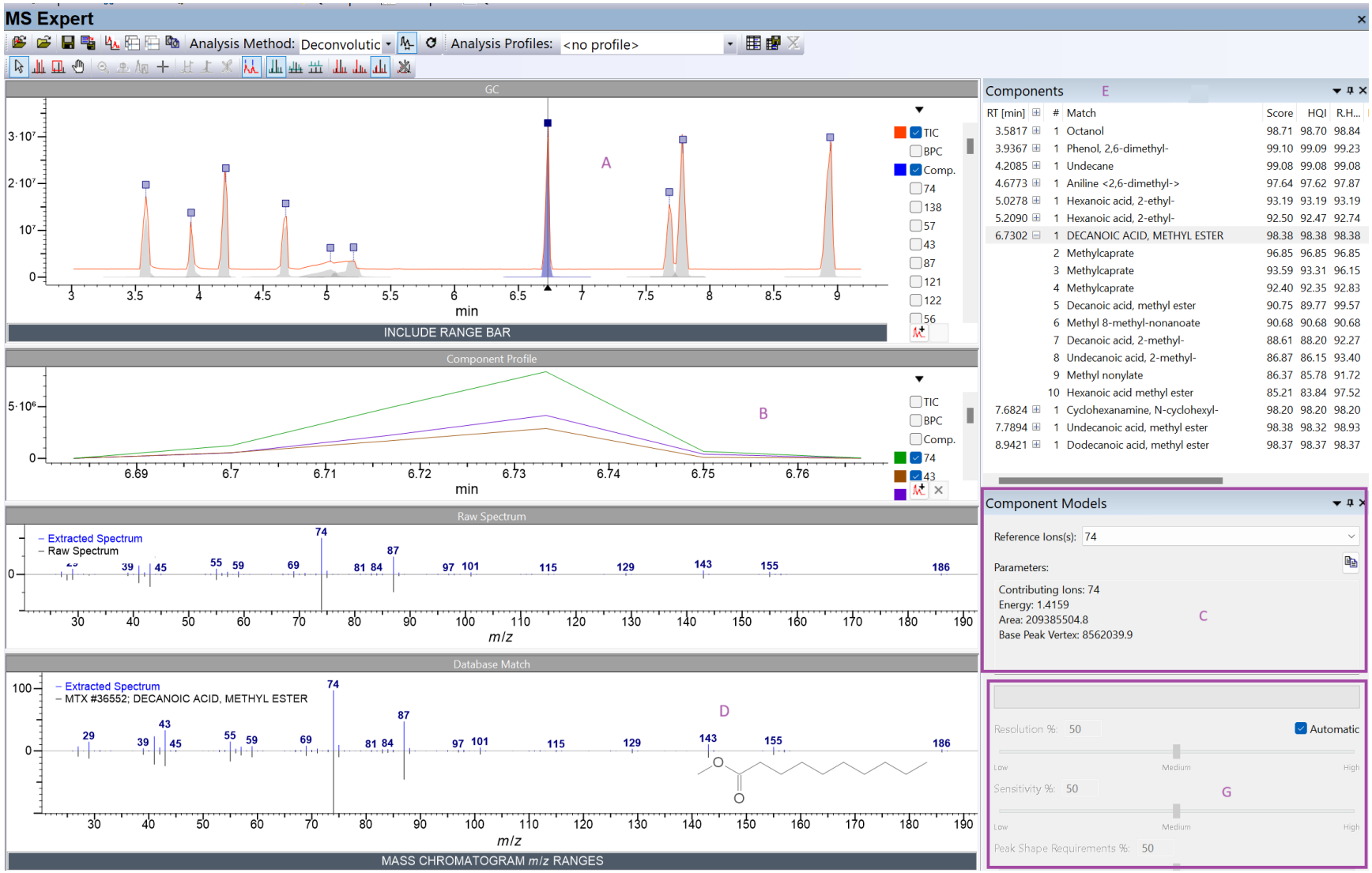
- Dot-Product (Cosine) – second equation in above image
- Wiley Dot-Product (Cosine) (old KnowItAll algorithm) - the old Finnigan algorithm that verified at least 12 of the largest 16 peaks AND the base peak match before continuing with the dot product calculation.
- Composite P1 – first equation in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

Example 1: GC-MS of Unit m/z Values

GUI explanation

The image below shows the deconvoluted GC-MS data of unit m/z values, and the search results for each component within the massive Wiley GC-MS database.

A – GC panel, showing the deconvoluted component peaks. One can use the checkboxes on right side of the GC panel to turn on/off chromatogram(s) display. When the box corresponding to the selected component peak of the chromatogram is selected, the box has a darker color;
B – Component Profile panel, highlights selected ions or components of the chromatogram(s);
C – Component Models panel, identifies the reference ion(s) used to model a component;
D – Database Match panel, shows the extracted spectrum (top) vs reference spectrum (bottom);
E – Components panel, where the Score is the combined spectrum search and reverse search Hit Quality Index (HQI) and, each component's GC area under curve (AUC) value;
G – Panel where parameters can be adjusted in algorithm.

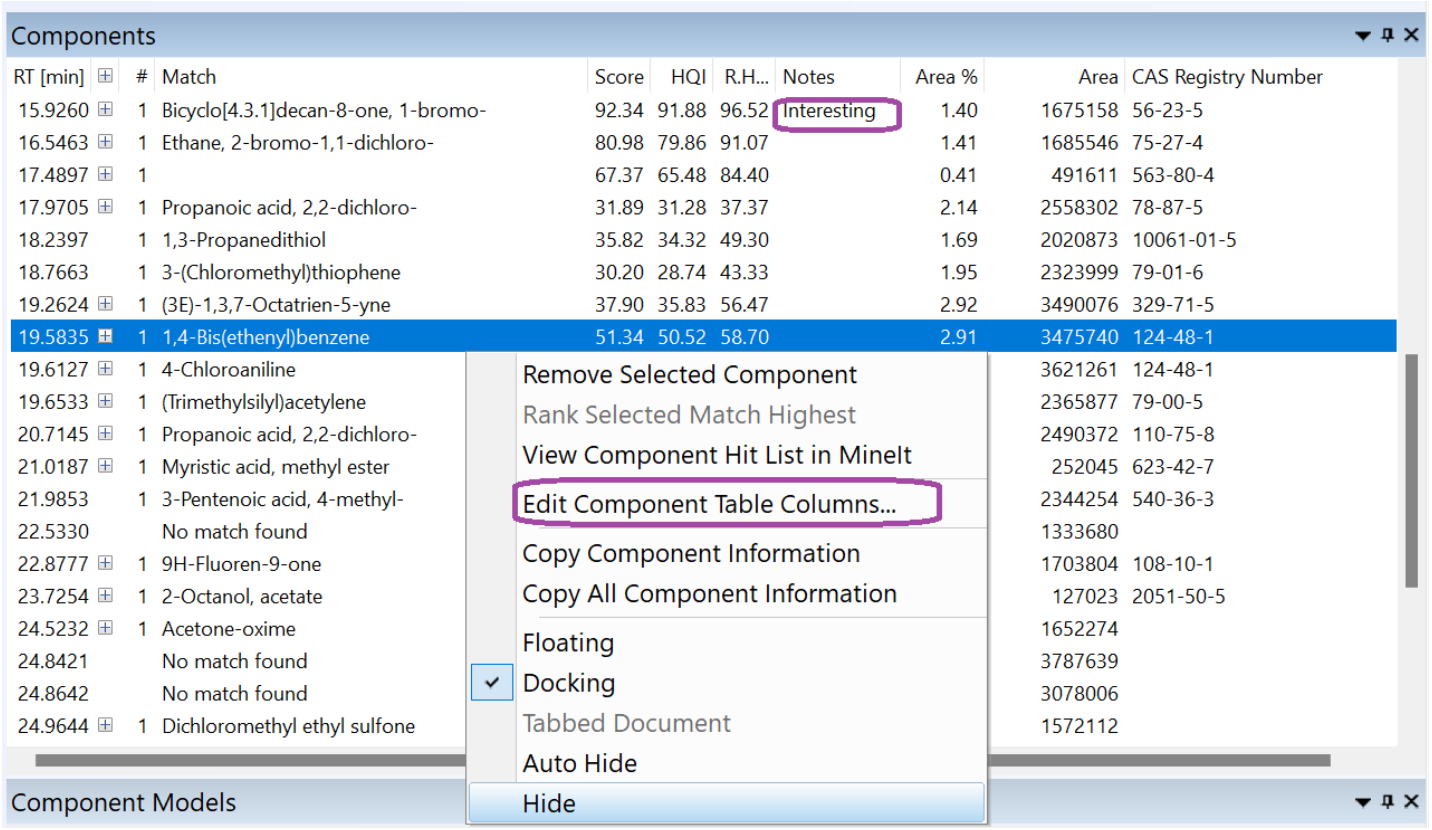


WILEY

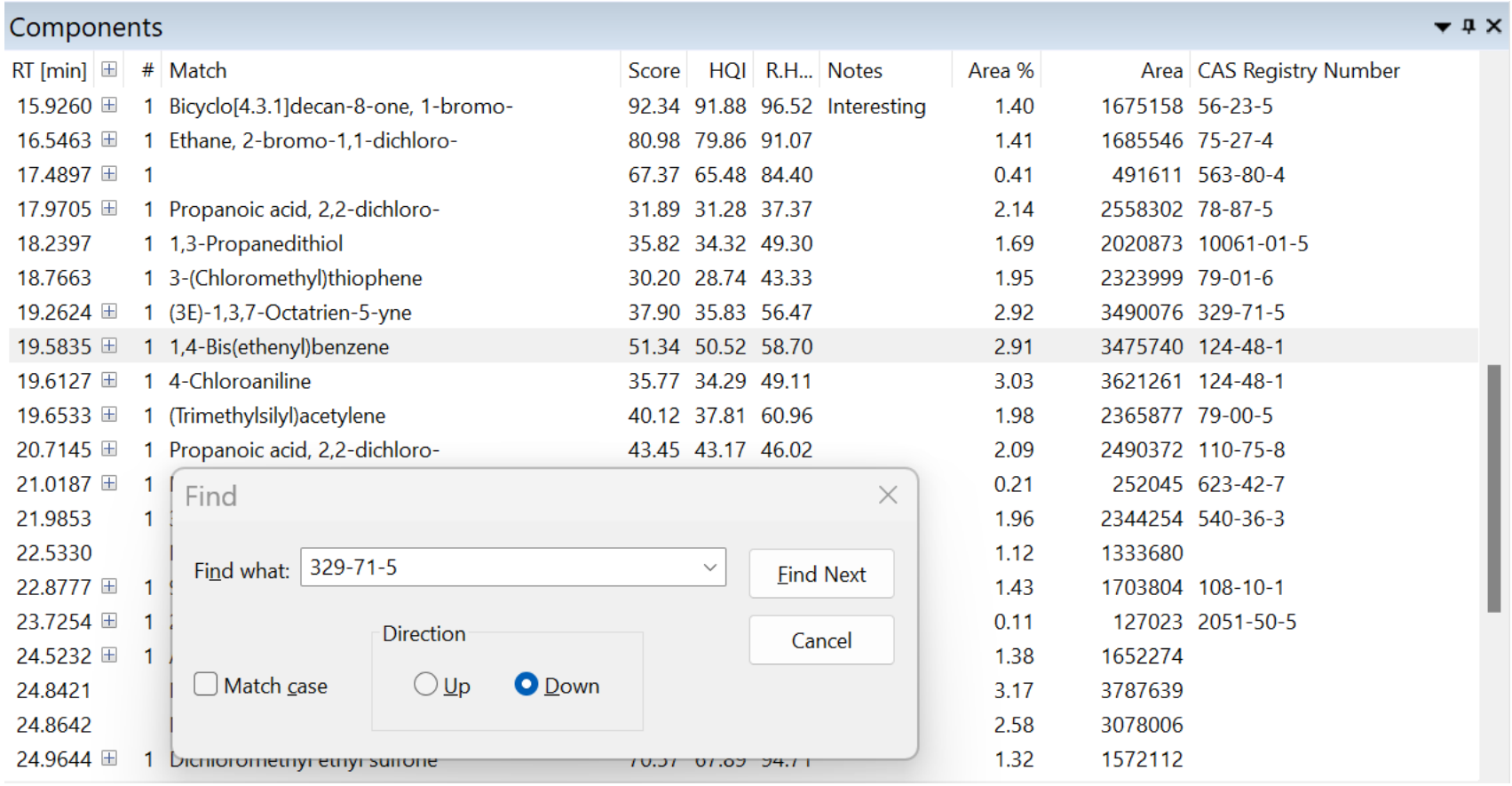
Features of MS Expert

The Component Table (E) contains useful information:

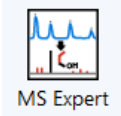
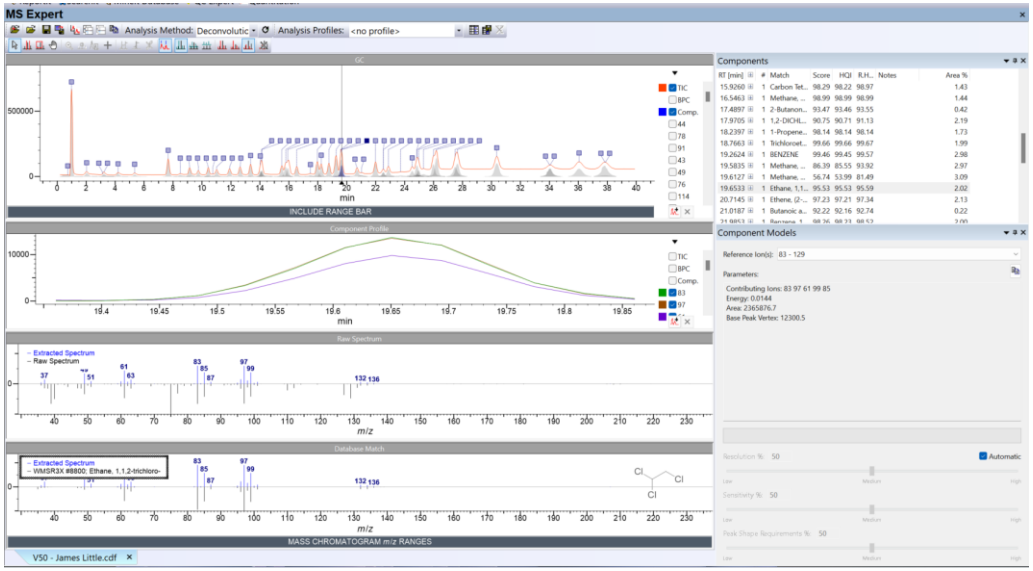
- It shows each Components' estimated area.
- One can add **Notes** to a match. This can be done by clicking on the space below **Notes**, in the row of the matched component of interest.
- Right mouse-click brings up actions that the user can take. Among many, **Edit Component Columns** allows one to rearrange the display.

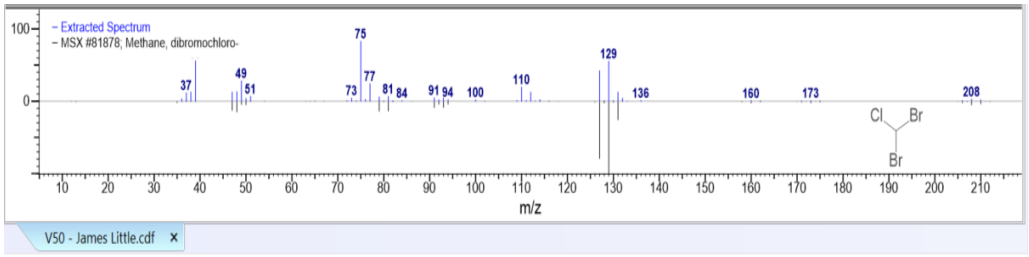
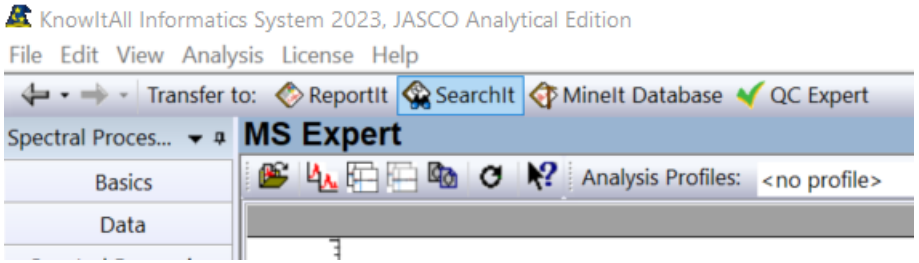


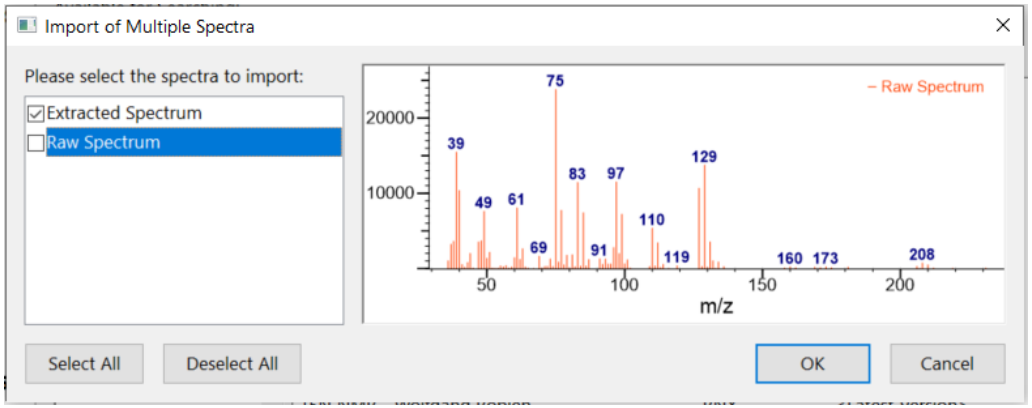
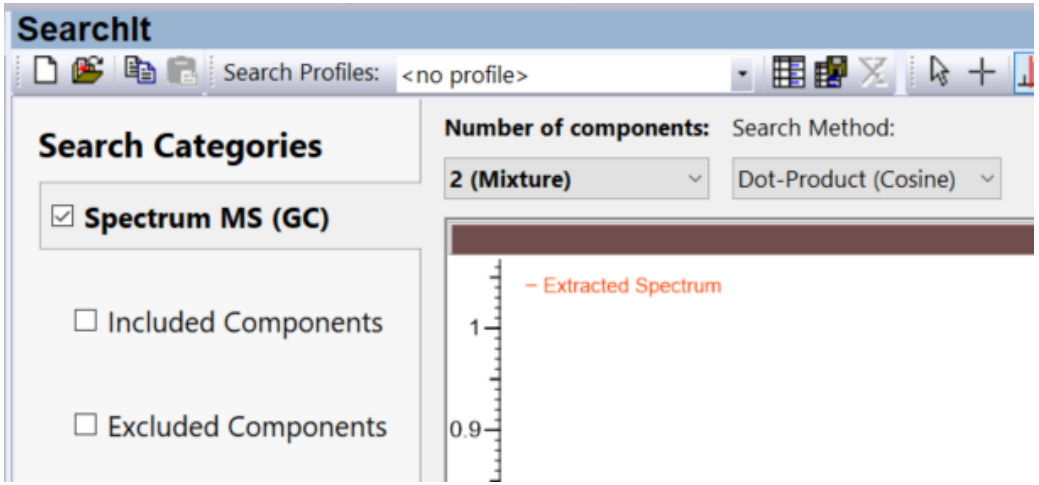
A hidden feature is the **Lookup** of a displayed property by using the **Control + F** action. For example, one can use **Lookup** to find a match to CAS Numbers such as CAS Number 329-7105.

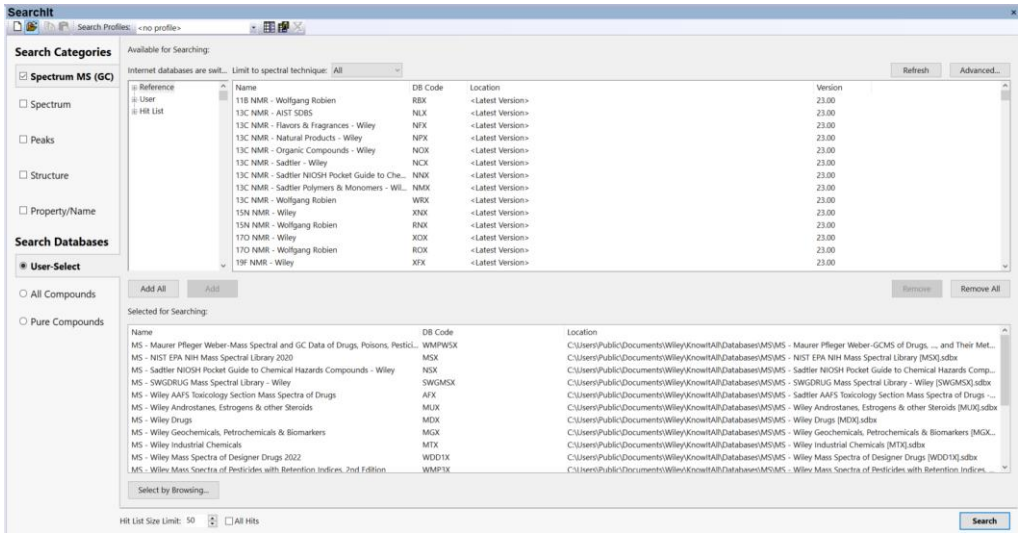


Now, we will go through a complete process.

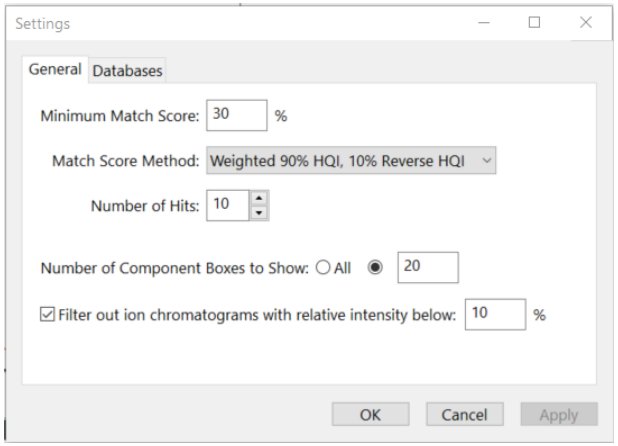
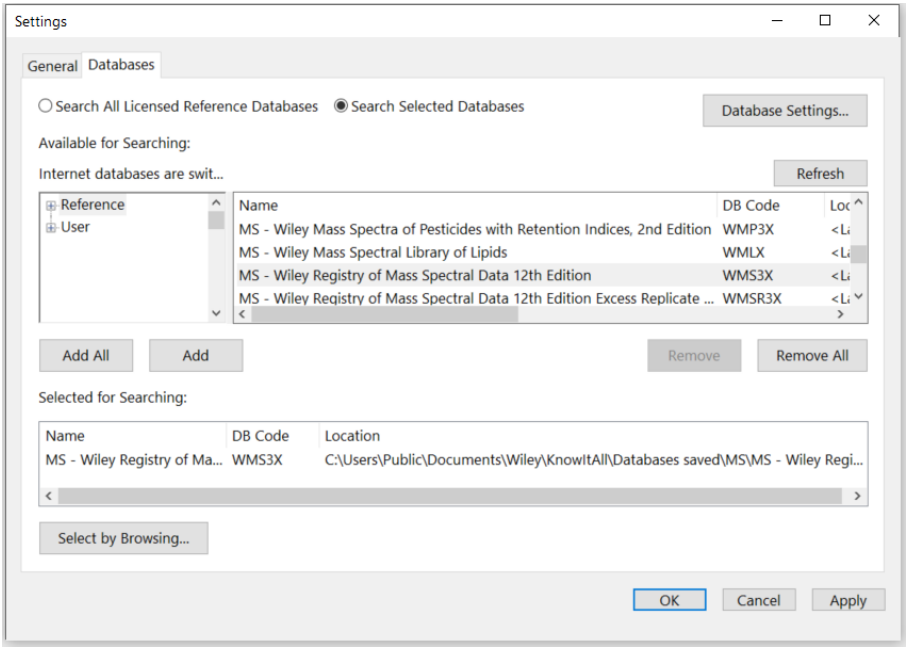
	Action	Result
1	Go to the **MS Expert** application under **Spectral Process** toolbar  Click **Open Raw GC-MS Data File** button Navigate to folder **C:\Users\Public\Documents\Wiley\KnowItAll\Samples\MS Expert\V50 - James Little.cdf** file Go to Component at 19.6533 min	**MS Expert** automatically performs 1. GC deconvolution to component extracted MS 2. Search extracted MS again reference databases 3. Report top hits 

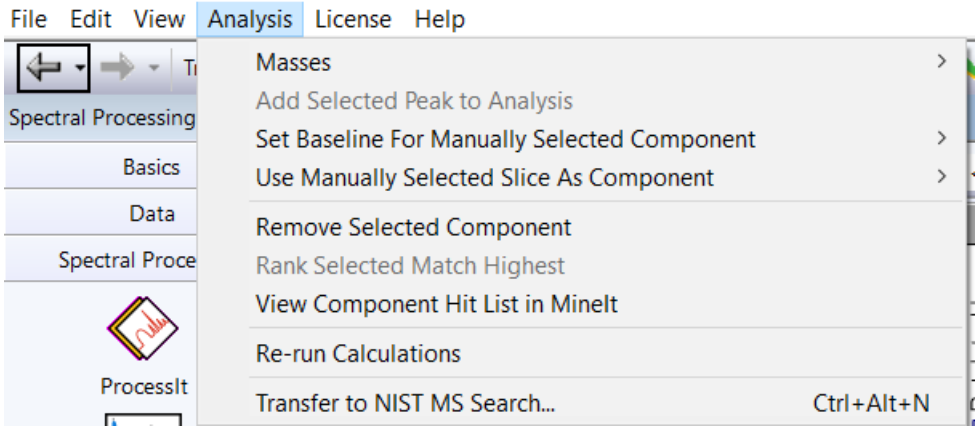
	Action	Result
2	Examine the hit lists for each deconvoluted components in the **Components** table Go to **RT (MIN)** 19.6127, notice the **Score** for this hit is very low, 56.74.	 From the comparison of **Extracted Spectrum** (top) and reference spectrum of methane, dibromochloro- in the bottom pane, we can see that the **Extract Spectrum** contains methane, dibromochloro- as well as other components.
3	Transfer the **Extracted Spectrum** to **SearchIt**	

	Action	Result
4	Only check the **Extracted Spectrum** at prompt Click **OK**	
5	In **SearchIt,** set to search for **2 (Mixture)**	

	Action	Result
6	Click the **User-Select** tab to add all licensed MS databases **Search**	

	Action	Result
7		Mixture analysis returns a very good **Composite Spectrum** match top of the hit list:  The **Composite Spectrum** in the first row is a mix of two components: 1-propene, 1,3-dichloro- (new component) and methane, dibromochloro-. The **Residual Spectrum** in the last row, the difference between **Extracted Spectrum** and **Composite Spectrum** is negligible, indicating no more components remain undetected.

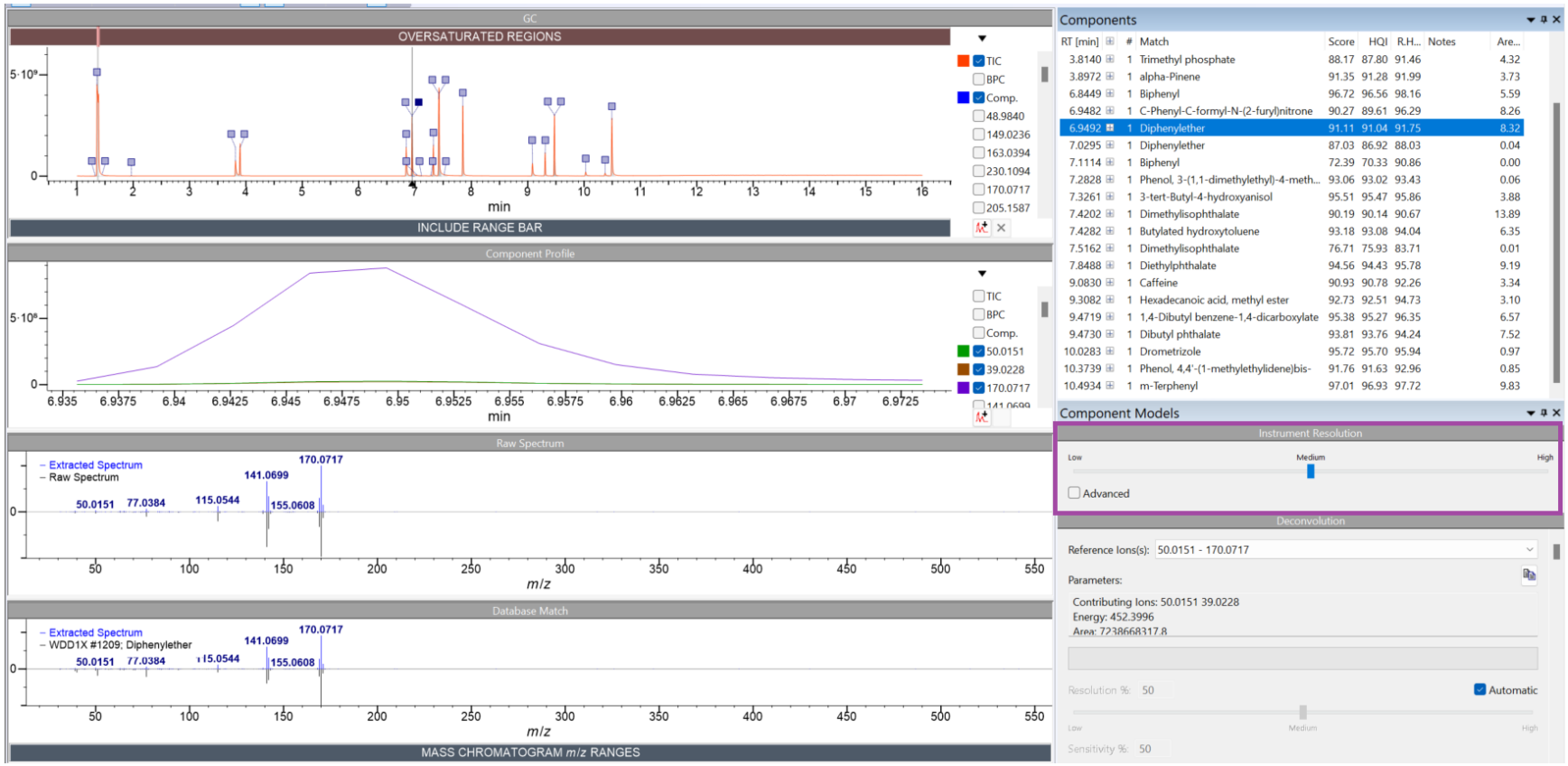
	Action	Result
8	Go back to **MS Expert,** One can go to **File > Settings** menu to adjust component hit list parameters	Hits evaluation parameters:  Reference database choices 

	Action	Result
9	One can click the TIC pane to add peaks manually, then select **Analysis > Add Selected Peak To Analysis** to perform a new analysis One can select **View Component Hit List in MineIt** to see Component table contents in **MineIt**	

Example 2: High-Resolution GC-MS

GUI explanation

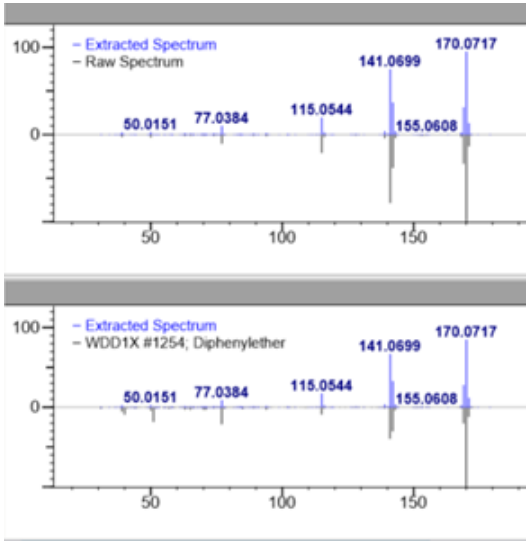
The instrument resolving power is required for the algorithm to automatically calculate the accurate *m/z* value of the data. Therefore, we use what our research considers a reasonable value by default that has a constant value and a variable value depending on mass (ppm). Empirically, this works in most cases. Increasing the *m/z* value accuracy too much incurs the danger of splitting an individual *m/z* value into individual mass spectral peaks that should be considered as just one. Decreasing the *m/z* value accuracy too much may cause individual mass spectral peaks to merge, resulting in incorrectly reported accurate *m/z* values.

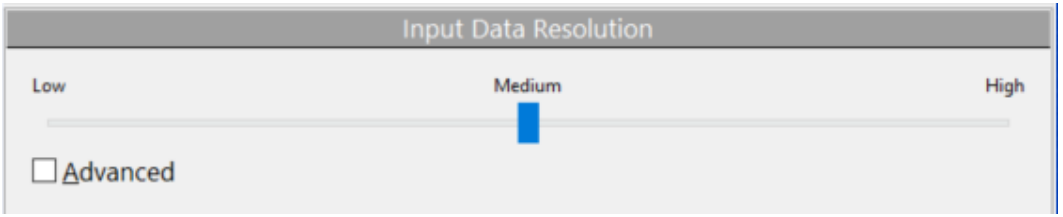
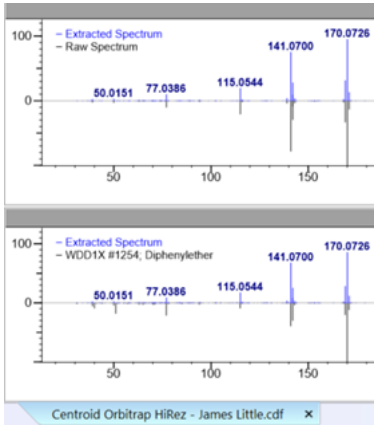


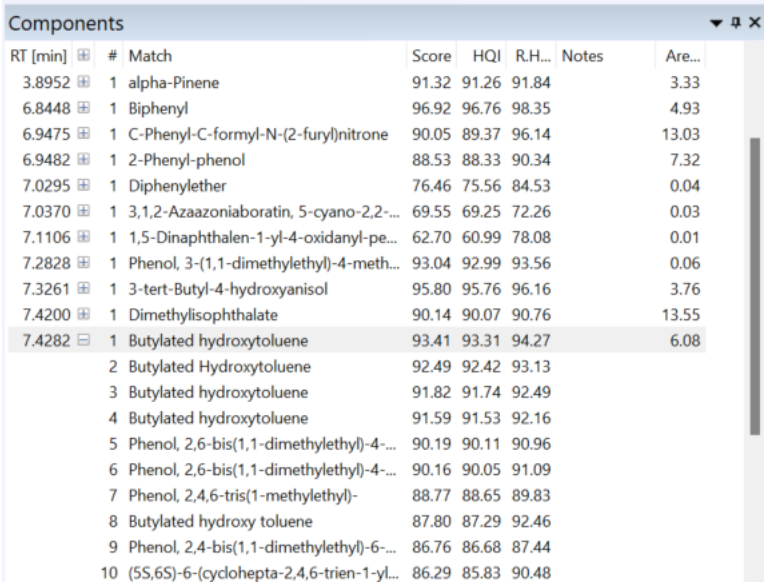
If a user knows the instrument's resolving power, that value should be entered in the **Instrument Resolution** panel, highlighted by the purple box, in above figure by clicking on 'Advanced' and manually entering the value.

It is possible to save the instrument's resolving power as part of a profile, which can then be selected depending on the type of instrument. A user can create several profiles with different resolution settings for different types of data (and instruments).

Hi-Resolution GC-MS Analysis using MS Expert – Exercise 1

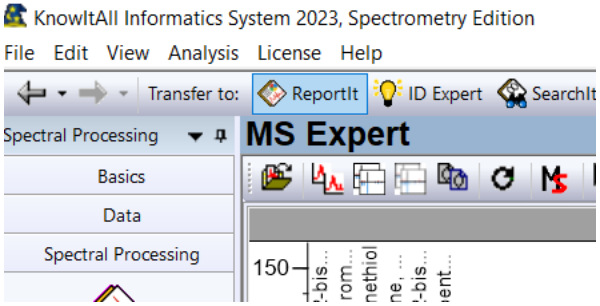
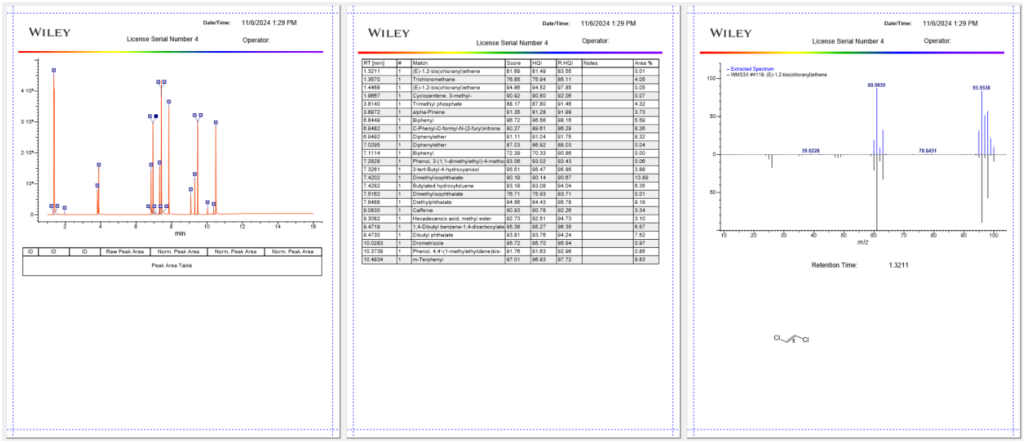
	Action	Result
1	Note: If you have changed your **File > Settings** in previous exercise, you should go back to **File > Settings** to reset for this exercise. Start a new analysis by clicking on the **Open Raw GC-MS Data File** button.  Navigate to the folder: **C:\Users\Public\Documents\Wiley\KnowItAll\Samples\MS Expert** Select the file: **Centroid Orbitrap HiRez - James Little.cdf** *Note: A pop-up window will appear. Click 'OK' to ignore the warning.* Using the **Components** panel, find the **RT (MIN)** 6.9482 and click on that component.	 The m/z value of 170.0717 is slightly off for this compound.

	Action	Result
2	Check **Advanced** in the **Instrument Resolution** panel. Set the instrument resolution at 3 ppm, click **OK.**	**MS Expert** recognizes this is a high-resolution data file and uniquely allows user to fill in resolution information or adjust resolution: 
3		The m/z value is corrected to 170.0726. This is the correct value. This instrument resolution set-up is persistent. 

	Action	Result
4	To expand the component hits of a selected match, click on the (⊞) to the left of the component name. For example, find the component with RT 7.4282. Click the (⊞) icon.	

Components

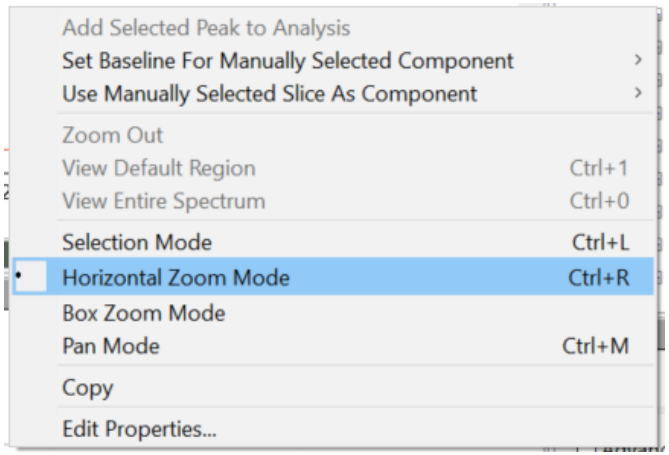
RT [min]	#	Match	Score	HQI	R.H...	Notes	Are...
3.8952	1	alpha-Pinene	91.32	91.26	91.84		3.33
6.8448	1	Biphenyl	96.92	96.76	98.35		4.93
6.9475	1	C-Phenyl-C-formyl-N-(2-furyl)nitrone	90.05	89.37	96.14		13.03
6.9482	1	2-Phenyl-phenol	88.53	88.33	90.34		7.32
7.0295	1	Diphenylether	76.46	75.56	84.53		0.04
7.0370	1	3,1,2-Azaazoniaboratin, 5-cyano-2,2-...	69.55	69.25	72.26		0.03
7.1106	1	1,5-Dinaphthalen-1-yl-4-oxidanyl-pe...	62.70	60.99	78.08		0.01
7.2828	1	Phenol, 3-(1,1-dimethylethyl)-4-meth...	93.04	92.99	93.56		0.06
7.3261	1	3-tert-Butyl-4-hydroxyanisol	95.80	95.76	96.16		3.76
7.4200	1	Dimethylisophthalate	90.14	90.07	90.76		13.55
7.4282	1	Butylated hydroxytoluene	93.41	93.31	94.27		6.08
	2	Butylated Hydroxytoluene	92.49	92.42	93.13		
	3	Butylated hydroxytoluene	91.82	91.74	92.49		
	4	Butylated hydroxytoluene	91.59	91.53	92.16		
	5	Phenol, 2,6-bis(1,1-dimethylethyl)-4-...	90.19	90.11	90.96		
	6	Phenol, 2,6-bis(1,1-dimethylethyl)-4-...	90.16	90.05	91.09		
	7	Phenol, 2,4,6-tris(1-methylethyl)-	88.77	88.65	89.83		
	8	Butylated hydroxy toluene	87.80	87.29	92.46		
	9	Phenol, 2,4-bis(1,1-dimethylethyl)-6-...	86.76	86.68	87.44		
	10	(5S,6S)-6-(cyclohepta-2,4,6-trien-1-yl...	86.29	85.83	90.48		

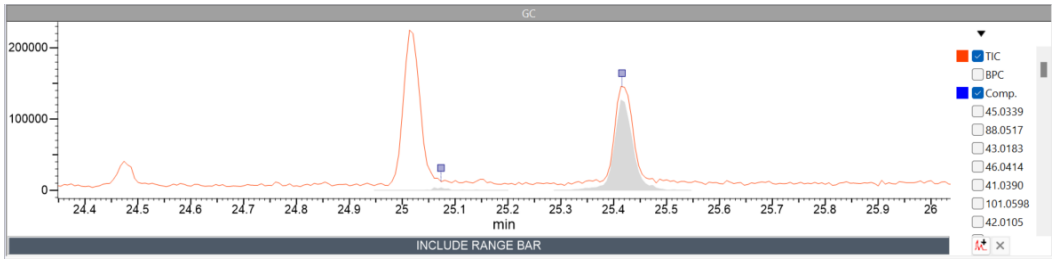
	Action	Result
5	To export results to **ReportIt**, use the **Transfer to:** toolbar. Click on **ReportIt.** A window will pop-up. Select the preferred report template. Click OK.  Select the **MS_Expert_Portrait** template	 Note: if this is the first time using the MS Expert reporting feature, templates will need to be added ahead of this step. The following steps should be performed using **MS Expert.** • File > Edit Report Templates • Choose **Add** button in the pop-up window. • Select the files, MS_Expert Landscape and MS_Expert Portrait. • Click **Open.** • Click **Close** in the Report Templates pop-up window.

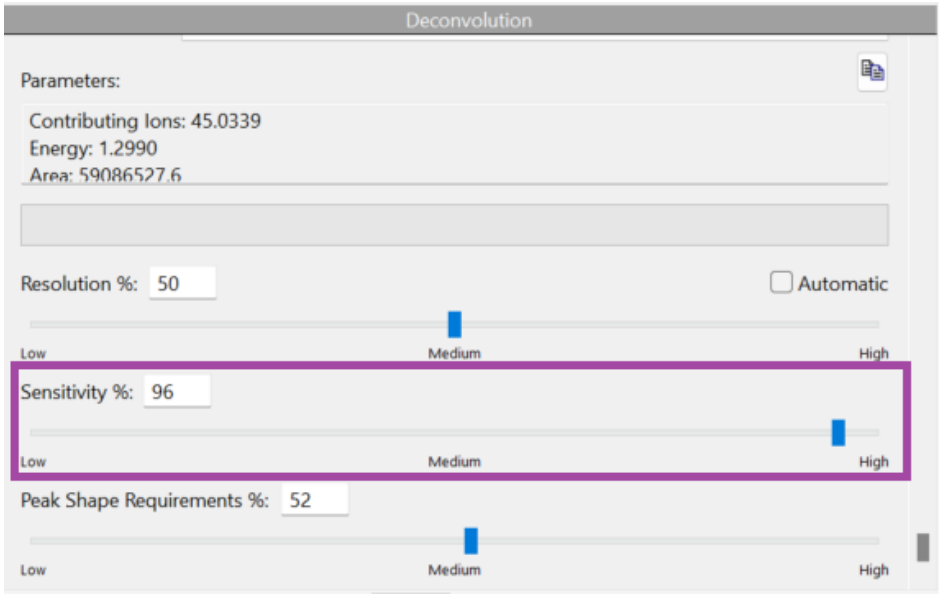
	Action	Result
6	Go back to **MS Expert.** Expand a few components' reference hits, and add some comments Right-mouse click the **Components** table and select **Copy All Component Information**. One can then paste this information to Excel or Word, etc.	

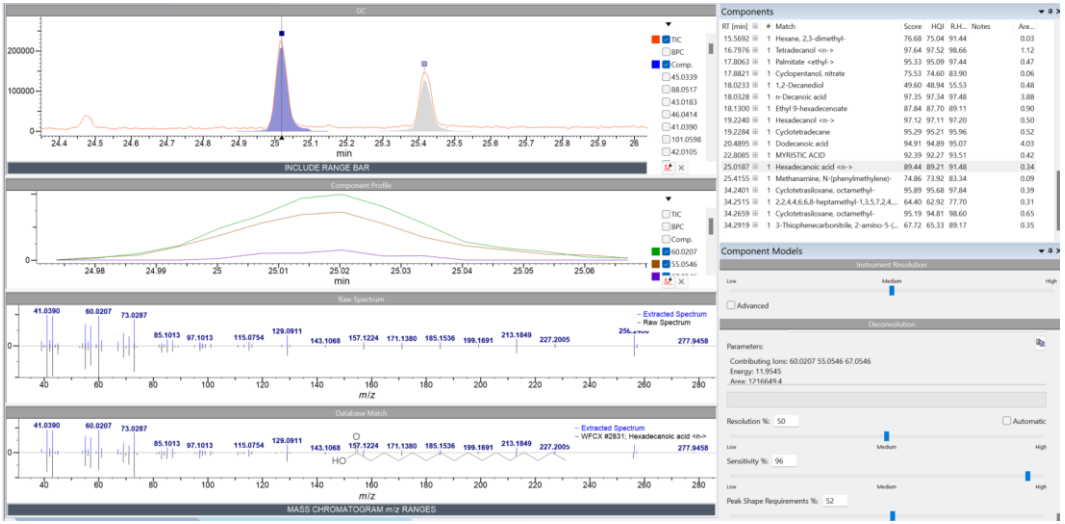
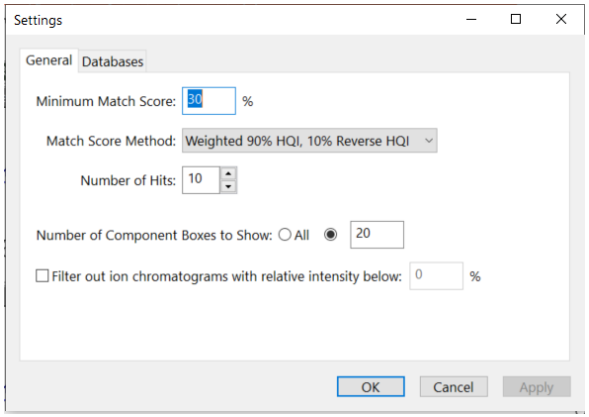
	A	B	C	D	E	F	G	H
1	RT [min]	#	Match	Score	HQI	R.HQI	Notes	Area %
2	1.3211	1	(E)-1,2-bis	81.69	81.49	83.55		0.01
3	1.357	1	Trichlorom	76.85	75.94	85.11		4.05
4	1.4468	1	(E)-1,2-bis	94.86	94.52	97.85		0.05
5	1.9667	1	Cyclopent	90.92	90.8	92.06		0.07
6	3.814	1	Trimethyl p	88.17	87.8	91.46		4.32
7	3.8972	1	alpha-Pine	91.35	91.28	91.99		3.73
8	6.8449	1	Biphenyl	96.72	96.56	98.16		5.59
9	6.9482	1	C-Phenyl-(	90.27	89.61	96.29		8.26
10	6.9492	1	Diphenylet	91.11	91.04	91.75		8.32
11	7.0295	1	Diphenylet	87.03	86.92	88.03	This is righ	0.04
12		2	DIPHENYL	85.7	84.87	93.22		
13		3	C-Phenyl-(	84.56	83.32	95.75		
14		4	2-Phenyl-p	84.5	84.18	87.39		
15		5	Biphenylol	82.6	82.52	83.33		
16		6	8-Methyl-1	81.62	80.31	93.42		
17		7	5-Methyl-1	81.57	80.26	93.36		
18		8	1,2-Dihydr	80.56	79.27	92.21		
19		9	1H,3H-Nap	80.43	80.31	81.46		
20		10	1-Naphtha	80.15	78.86	91.74		
21	7.1114	1	Biphenyl	72.39	70.33	90.86		0
22	7.2828	1	Phenol, 3-(	93.06	93.02	93.43		0.06
23	7.3261	1	3-tert-Buty	95.51	95.47	95.86		3.88
24	7.4202	1	Dimethylis	90.19	90.14	90.67		13.89
25	7.4282	1	Butylated I	93.18	93.08	94.04		6.35
26	7.5162	1	Dimethylis	76.71	75.93	83.71		0.01
27	7.8488	1	Diethylpht	94.56	94.43	95.78		9.19
28	9.083	1	Caffeine	90.93	90.78	92.26		3.34
29	9.3082	1	Hexadecar	92.73	92.51	94.73		3.1
30	9.4719	1	1,4-Dibuty	95.38	95.27	96.35		6.57
31	9.473	1	Dibutyl pht	93.81	93.76	94.24		7.52
32	10.0283	1	Drometrizo	95.72	95.7	95.94		0.97
33	10.3739	1	Phenol, 4,4	91.76	91.63	92.96		0.85
34	10.4934	1	m-Terpher	97.01	96.93	97.72		9.83

Hi-Resolution GC-MS Analysis using MS Expert – Exercise 2

	Action	Result
1	Open another high-resolution GC-MS file from the same folder. Navigate to the folder: **C:\Users\Public\Documents\Wiley\KnowItAll\Samples\MS Expert** Select the file: **OBAN 14 JEOL HiRez - James Little.cdf** Right-mouse-click the **GC** panel. Choose Horizontal Zoom or Box Zoom mode.	

	Action	Result
2	Zoom into region 24 – 26 min (3 tiny peaks). Using Horizontal Zoom; • Using the mouse cursor, click on region of 24 min and drag the cursor to region 26 min. Using Box Zoom mode. • Using the mouse cursor, click on region 24 toward the top of the GC panel. Drag the cursor to the right and down to include region 26min. • This creates a box around the region of interest.	

	Action	Result
3	We can increase the algorithm's Sensitivity to resolve the second issue. In the **Deconvolution** panel, uncheck the box for **Automatic** deconvolution. Move the **Sensitivity** bar to **High.**	

	Action	Result
4		Now the second component is correctly identified as the missing hexadecanoic acid. 
5	Sometimes, there are too many GC peaks crowding the GC panel. To adjust the display to reduce crowding, click **File > Settings** to set preferences:	

How to Perform GC-MS Analysis by Peak Picking

Purpose

These exercises demonstrate how to use KnowItAll MS Expert to analyze GC-MS data by peak picking.

Objectives

These exercises will teach you:

- How to use KnowItAll MS Expert to analyze GC-MS using peak picking.
- How to generate reports.

Background

GC-MS data can be analyzed by peak picking without going through deconvolution. Novel compounds can be identified, and structural characteristics deduced from applying the MS Adaptive search that uses fragmentation and structural data to propose likely structural details of the unknown.

Training Files Used in This Lesson

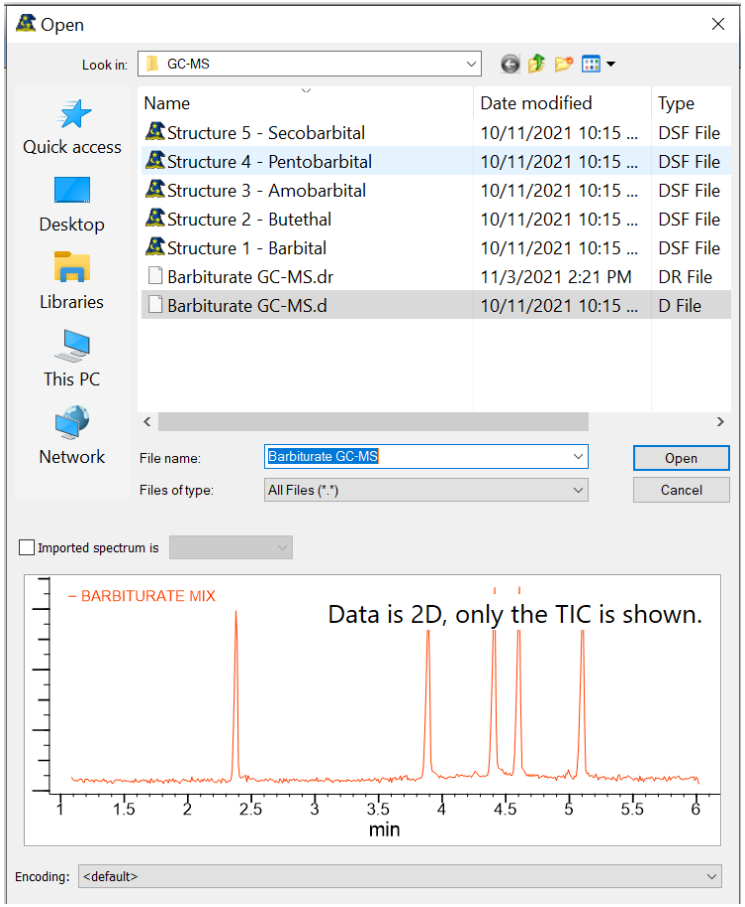
- C:\Users\Public\Documents\Wiley\KnowItAll\Samples\GC-MS\Barbiturate GC-MS.d

KnowItAll Applications Used

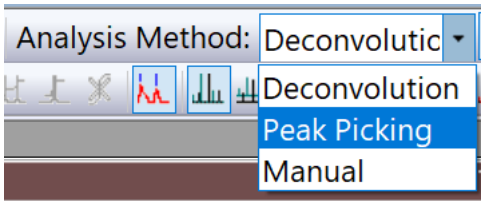
- KnowItAll MS Expert

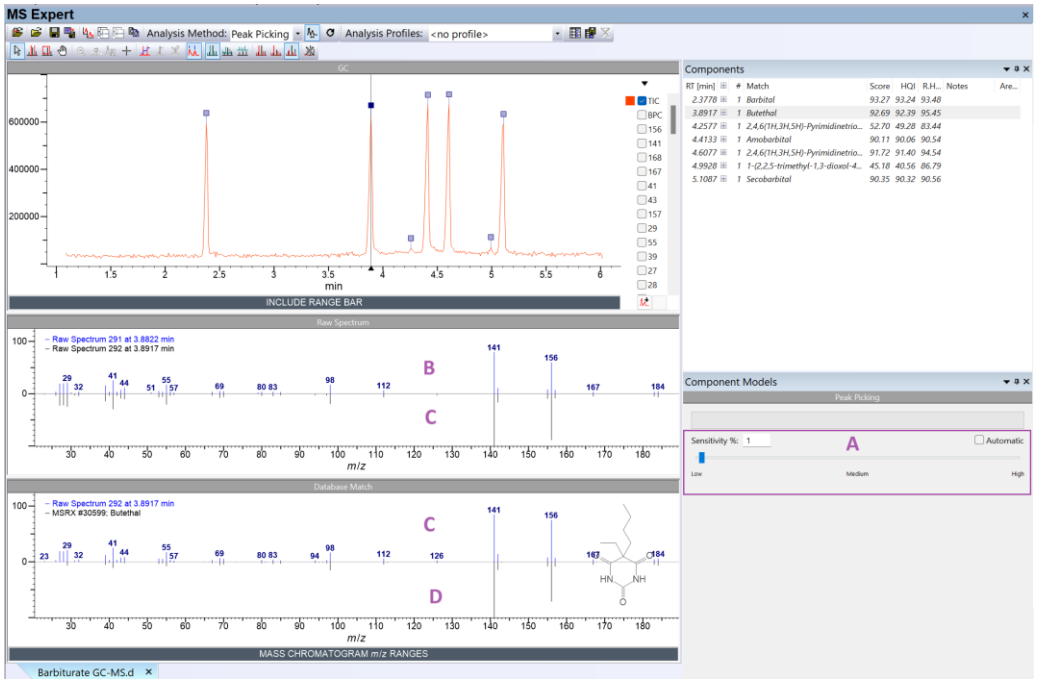
KnowItAll MS Expert defaults to perform deconvolution, however one can just use the **Peak Picking** option to analyze GC-MS data, especially if the GC separates components well enough.

Example: Peak Picking

	Action	Result
1	Click **Open Raw GC-MS Data File** button. Navigate to "**C:\Users\Public\Documents\Wiley\KnowItAll\Samples\GC-MS**" Select the file: **Barbiturate GC-MS.d**	

WILEY

	Action	Result
2	Switch **Analysis Method** to **Peak Picking.**	

	Action	Result
3	Reduce the **Sensitivity** to **Low** so that only main peaks are picked (**A**). Select the peak at 3.8917 min. Hold the mouse cursor over the solid square denoting the selected peak in the **GC** panel. • In the **Raw Spectrum** panel, **B** displays the MS spectrum the of the selected peak • **C**, in both the **GC** and the **Database Match** panels, displays selected peak's MS spectrum; • In the **Database Match** panel, **D** indicates the reference spectrum match to the selected spectrum indicated by **C**.	

Manually Analyze GC-MS

How to manually analyze GC-MS

Purpose

These exercises demonstrate how to use KnowItAll MS Expert to manually analyze GC-MS.

Objectives

These exercises will teach you:

- How to use KnowItAll MS Expert to manually analyze GC-MS.

Background

The KnowItAll MS Expert allows one to examine GC-MS data and, perform spectral subtraction. Matches to the subtracted spectrum are searched for against the reference data.

Training Files Used in This Lesson

- C:\Users\Public\Documents\Wiley\KnowItAll\Samples\GC-MS\Barbiturate GC-MS.d

KnowItAll Applications Used

- KnowItAll MS Expert

Average MS (green bar in GC pane), defined by by clicking the mouse on the bar, drag and drop

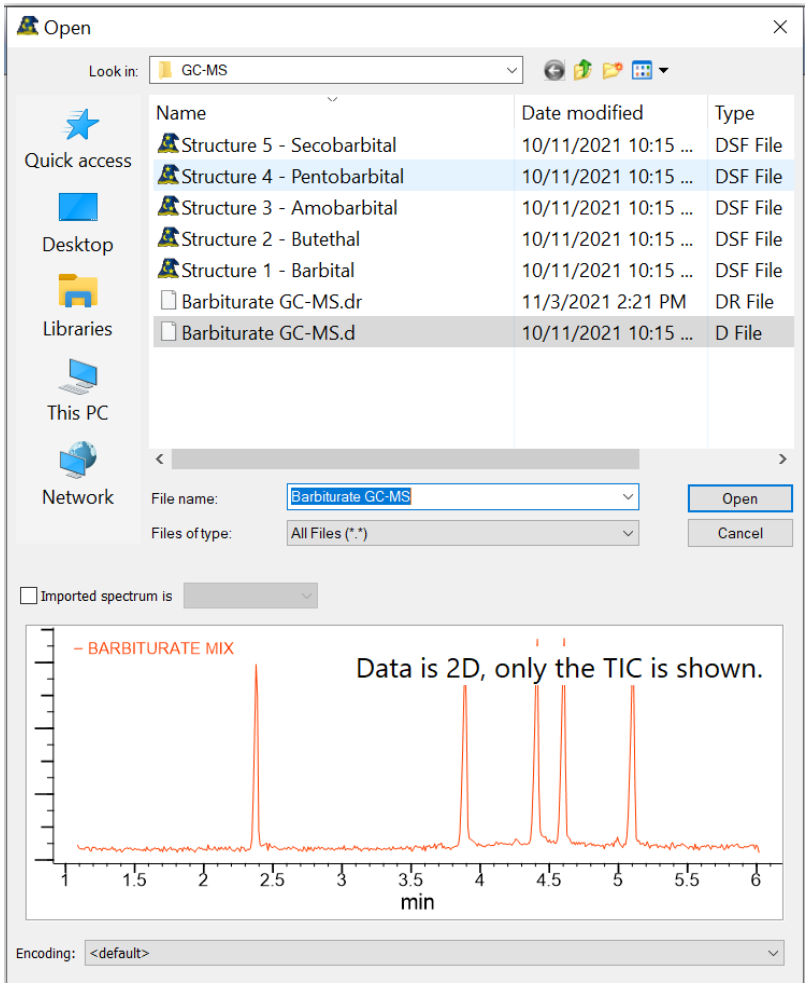
- Average MS (green bar) - background (red bar)

The **Matches** table are database search hit list for the selected MS.

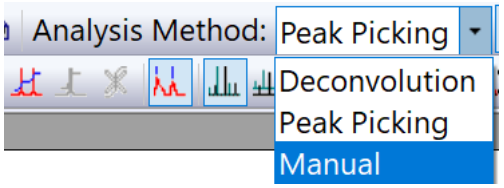
The selected MS can be transferred to **SearchIt** application for spectrum search.

The area under curve (AUC) and peak height values for the ion chromatogram can be calculated by using the **Analysis > Peak Area/Integration** menu-item.

Example: Manual Analysis

	Action	Result
1	In **MS Expert**, click on the **Open Raw GC-MS Data File** button. Navigate to: "**C:\Users\Public\Documents\Wiley\KnowItAll\Samples\GC-MS**" Select the file: **Barbiturate GC-MS.d**	

WILEY

	Action	Result
2	Switch **Analysis Method** to **Manual.**	
3	Use the **Average Range Bar** (A) to define a selected region of the MS chromatogram. • Click and drag the mouse cursor along the **Average Range Bar**. • You select many ranges. Use the **Background Range Bar** (B) to define the selected background region of the chromatogram. • Click and drag the mouse cursor along the **Backrgound Range Bar**. • You select many ranges.	The MS spectrum with the subtracted background (C) and the corresponding matched reference MS spectrum (D) is displayed in the bottom panel.