KnowItAll Software Training

GC-MS Analysis Using KnowItAll MS Expert, ProcessIt And SearchIt

WILEY

Automatic GC-MS Analysis

How to Use KnowItAll MS Expert to Perform Automatic GC-MS Analysis

Purpose

These exercises demonstrate how to use KnowItAll MS Expert to automatically 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 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 that exist in 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 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 pursued across multiple raw spectra, and a component spectrum is extracted from the data for each individual component while trying to separate components with overlapping *m/z* value peaks. The details of the algorithm are summarized to a large extent by the following papers[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.

The details of the algorithm are summarized to a large extent by the following papers, although we go beyond that.

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, Mark J. Stefik, Thomas C. Rindfleisch, Alan M. Duffield. Extraction of Mass Spectra Free of Background and Neighboring Component Contributions from Gas Chromatography IMass Spectrometry Data. *ANALYTICAL CHEMISTRY*, 1976, **VOL. 48, NO.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. Bruce 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

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

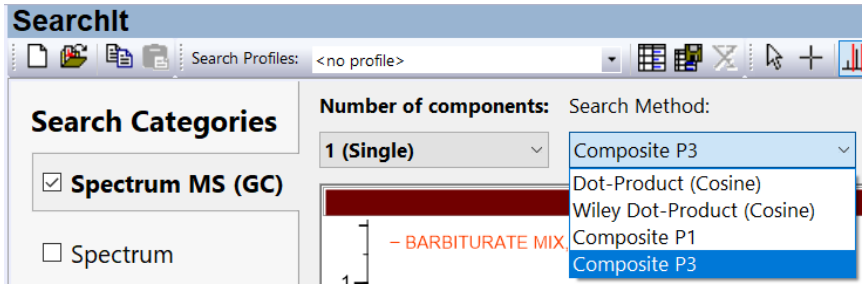
Spectrum search type – identity (normal)
Presearch – default
Included Libs – MainLib
Apply limits – unchecked
Use constraints – unchecked

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.



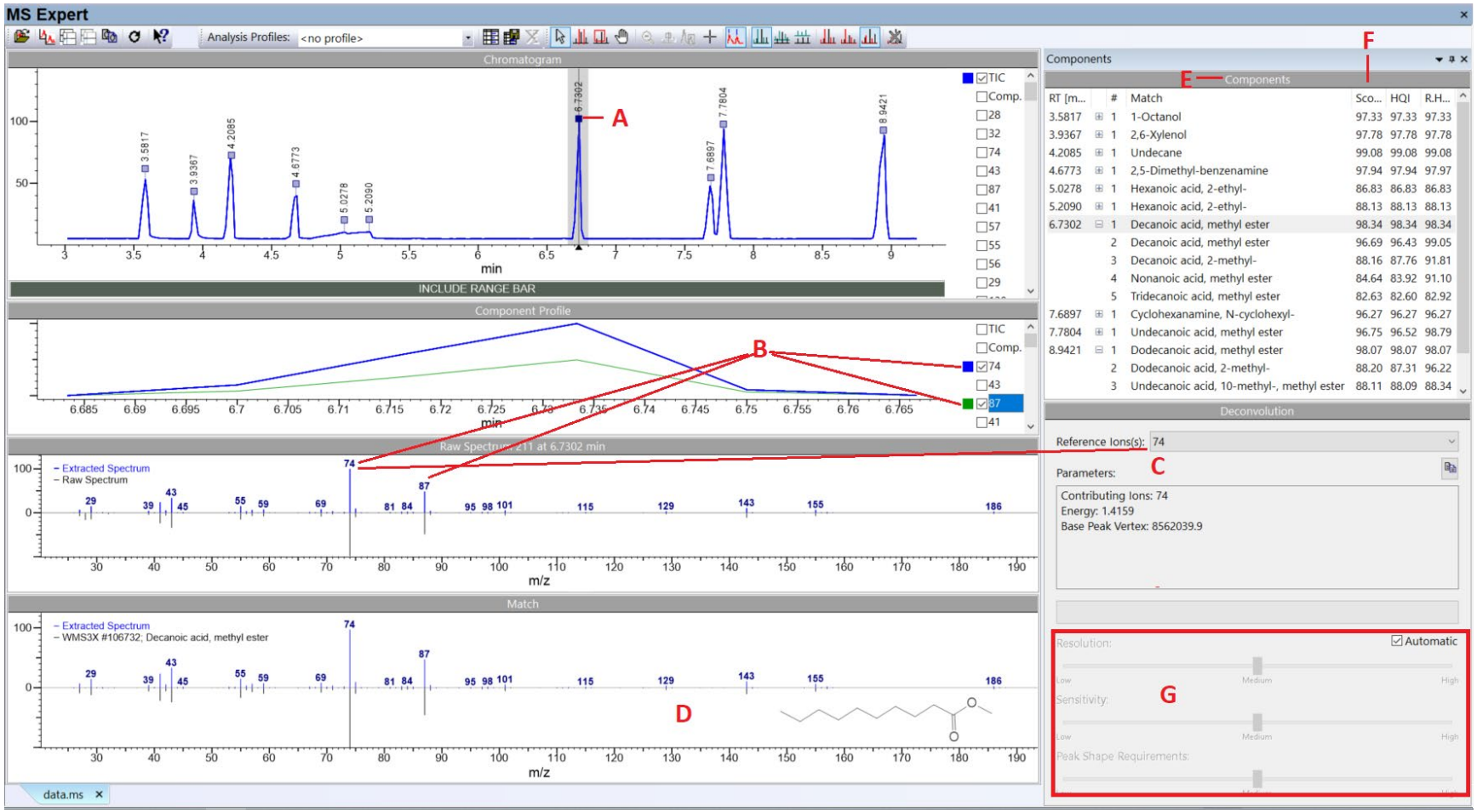
Where

- Dot-Product – second equation in above graph
- Wiley Dot-Product (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 graph
- Composite P3 – first equation in above graph
  P1 and P3 are different by the power applied to the weighted intensity of peak.

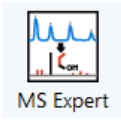
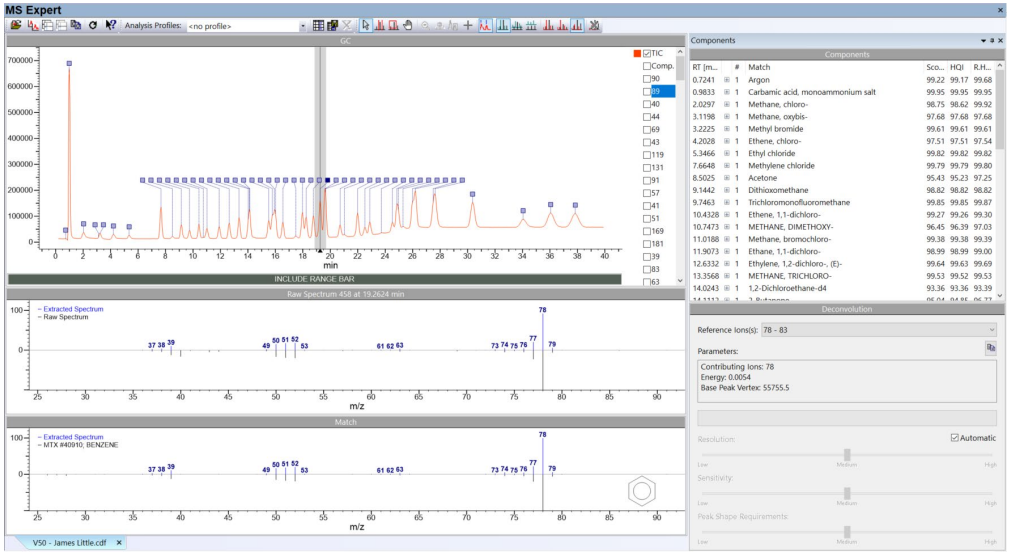
Example 1: GC-MS of Unit m/z Values

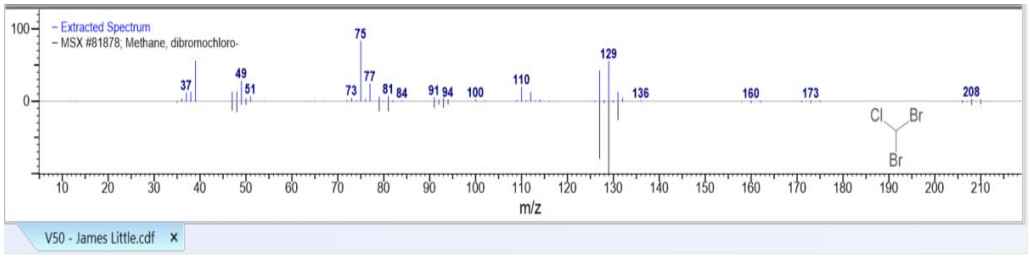
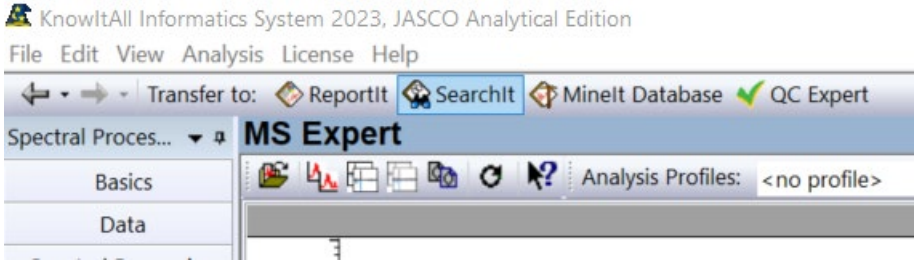
GUI explanation

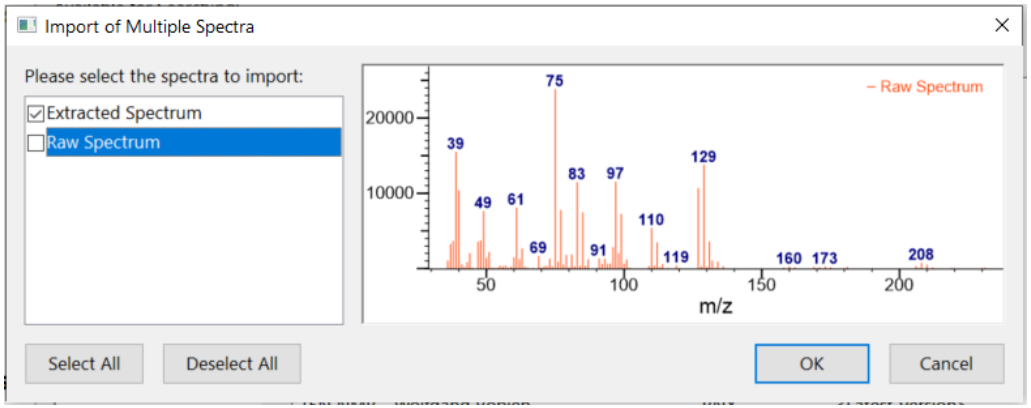
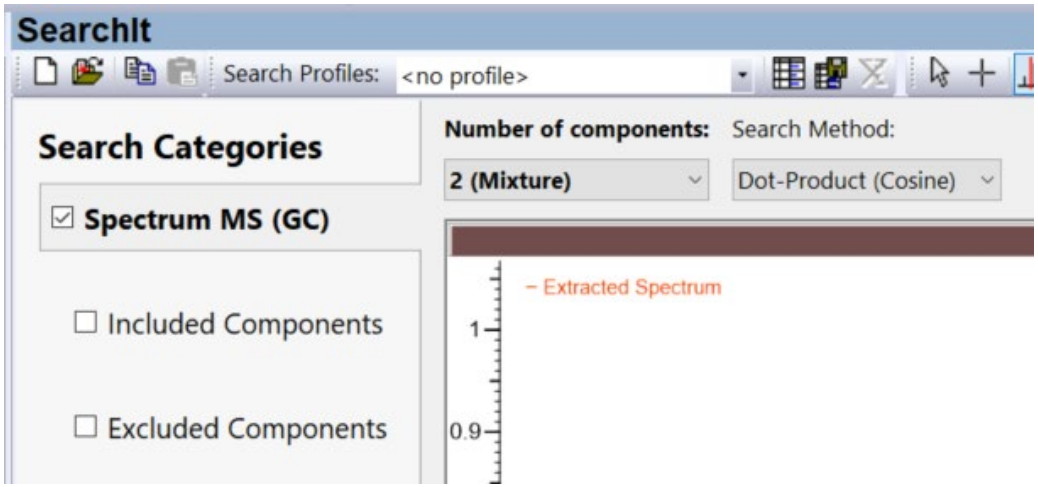
This picture shows the deconvoluted GC-MS data of unit m/z values, and the massive Wiley MS database search results for each component.

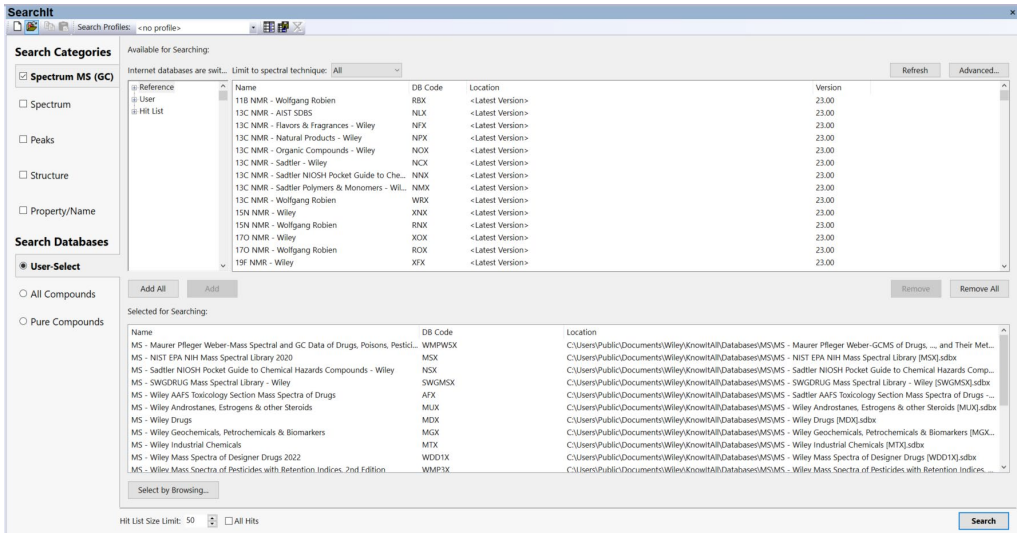


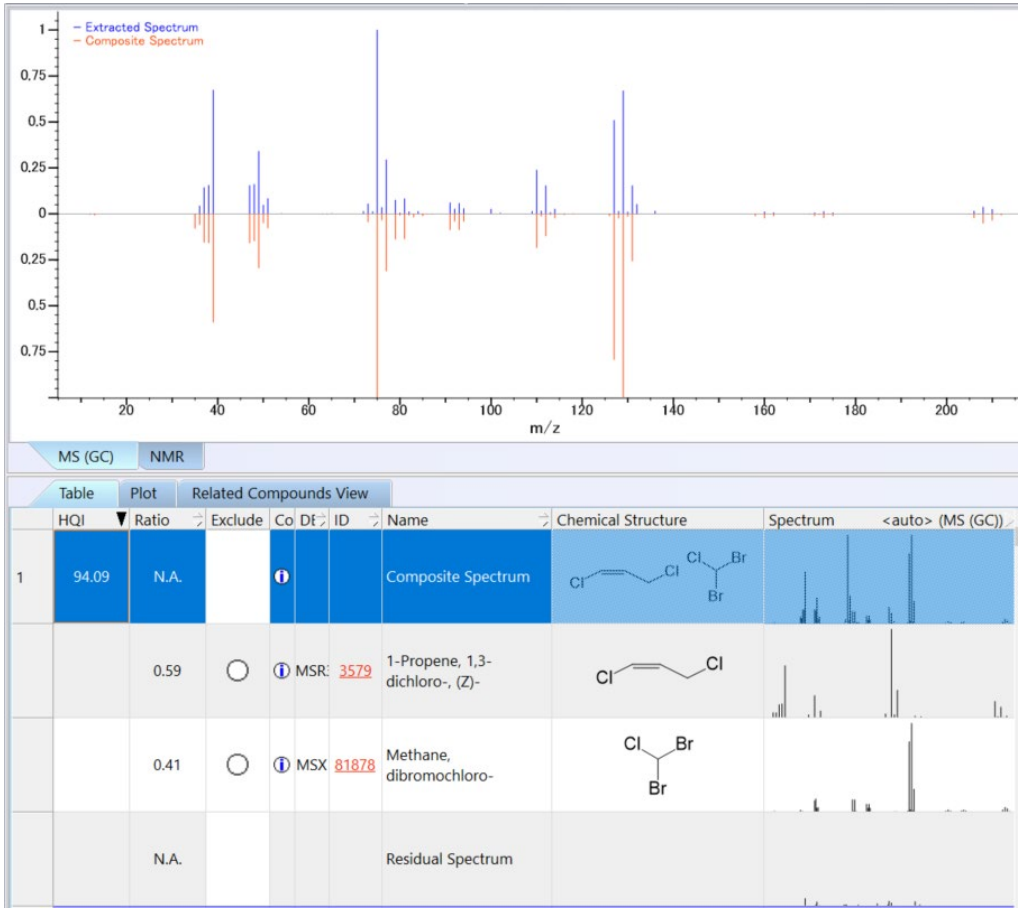
A – deconvoluted GC peak which found matches from reference MS databases; B – selected ions in a component; C – reference ion(s) used to model a component; D – extracted spectrum (top) vs reference spectrum (bottom); E – components table; F – combined spectrum search and reverse search Hit Quality Index (HQI) and, not shown in the above screenshot, each component's GC area under curve (AUC) value; G – Adjustable parameters in algorithm.

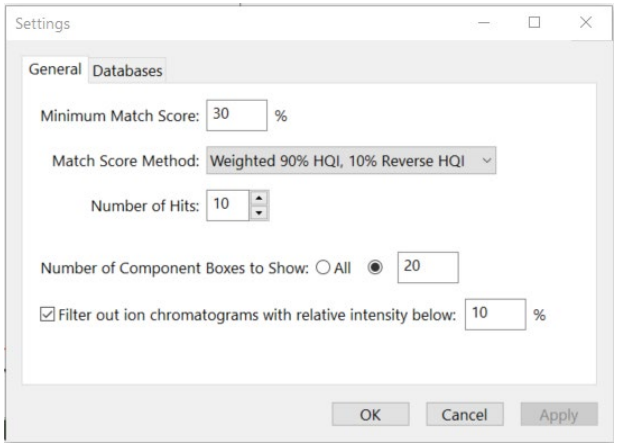
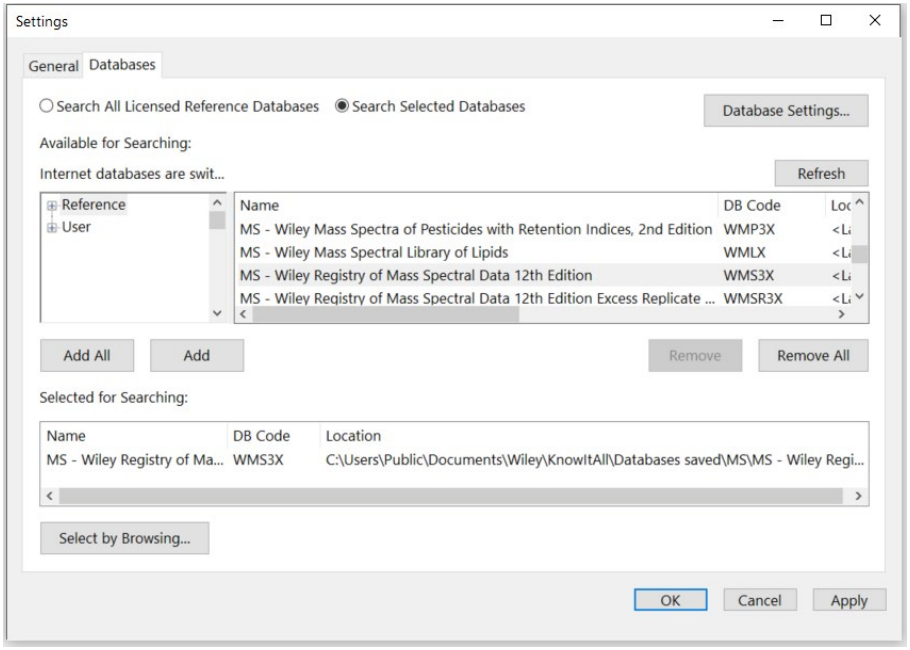
	Action	Result
1	Go to the **MS Expert** application under **Spectral Process** toolbar  Click **Open Data File** button Navigate to folder C:\Users\Public\Documents\Wiley\KnowItAll\Samples\MS Expert V50 - James Little.CDF file ***Note:*** *in 2024 release, one can use button Open Analysis to open an externally saved analysis file.*	**MS Expert** automatically performs 1. GC deconvolution to component extracted MS 2. Search extracted MS again reference databases 3. Report top hits  ***Note:*** *in 2024 release, we add a column "Area%" to the Components table to give user a rough idea for the component percentage of chromatogram peak intensity.*

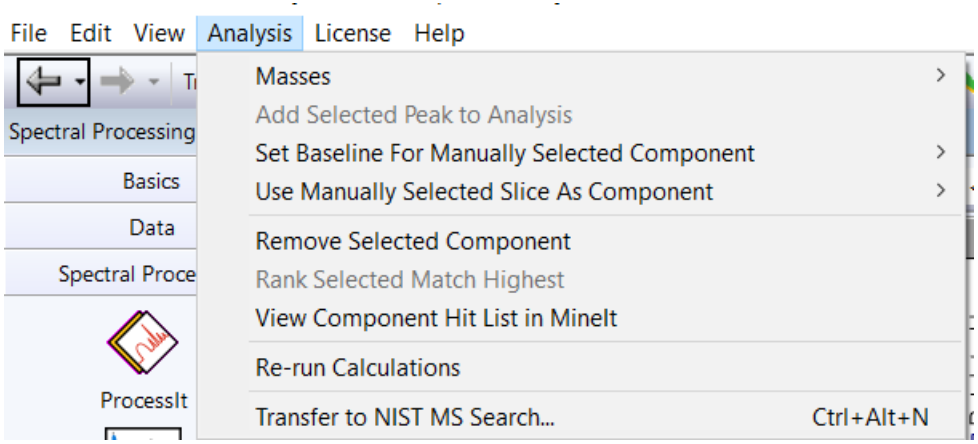
	Action	Result
2	Examine the hit lists for each deconvoluted components in the **Components** table Go to **RT (MIN)** 19.6127, notice the **Score** for this hit is very low, 56.74.	 From the comparison of **Extracted Spectrum** (top) and reference spectrum of methane, dibromochloro- in the bottom pane, 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 **Extract Spectrum** at prompt **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**	

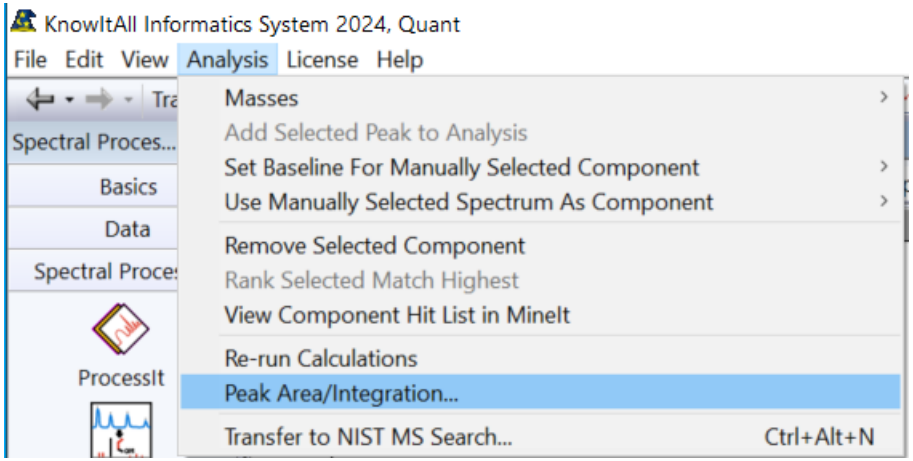
	Action	Result
7		Mixture analysis returns a very good **Composite Spectrum** match top of the hit list:  The **Composite Spectrum** in the first row is a mix of two components: 1-propene, 1,3-dichloro- (new component) and methane, dibromochloro-. The **Residual Spectrum** in the last row, the difference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

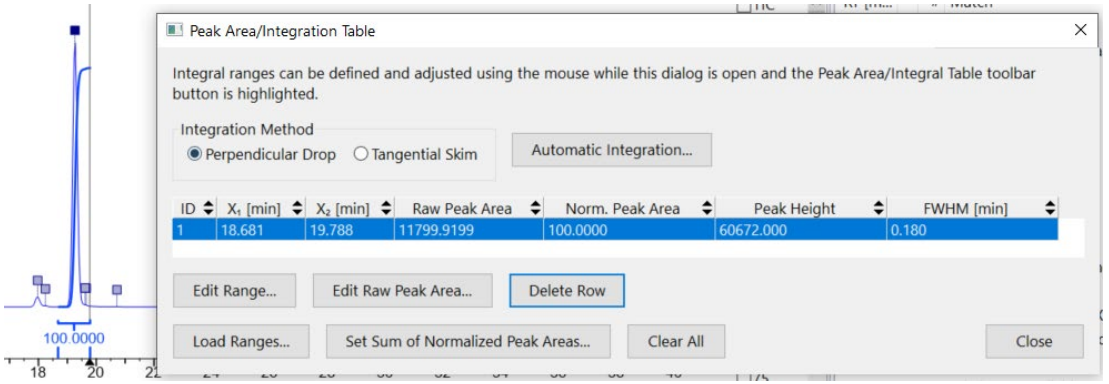
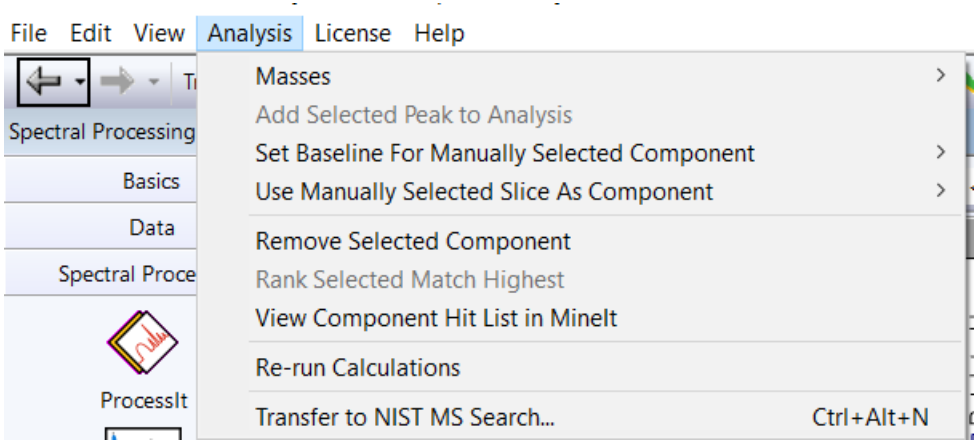
	Action	Result
8	Go back to **MS Expert,** One can go to **File > Settings** menu to adjust component hit list parameters	Hits evaluation parameters:  Reference database choices 

	Action	Result
9	One can click the TIC pane **INCLUDE RANGE BAR** to add peaks manually, then select **Re-run Calculations** to perform a new analysis One can select **View Component Hit List in MineIt** to see Component table contents in **MineIt**	

Example 2: Evaluate Chromatogram Peak Area Under Curve Values

This is a new feature in KnowItAll 2024 release. With this capability, one can get the chromatogram AUC value per component.

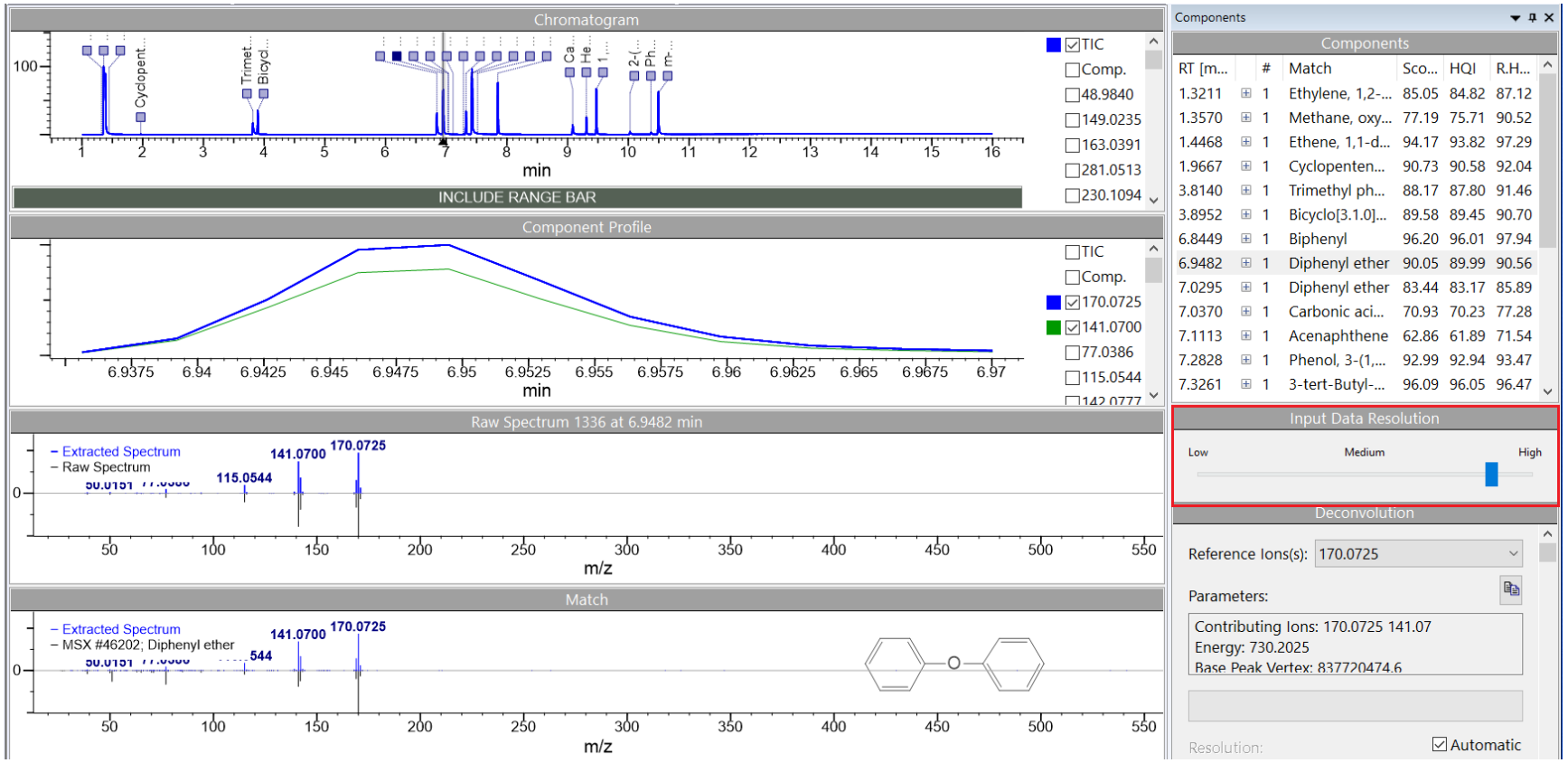
	Action	Result
1	Continue with the above example. Turn off TIC check and turn on the ion m/z value of 78  Select menu-item **Analysis > Peak Area/Integration**	

	Action	Result
2	Mouse over the chromatogram pane, an integration sign ∫ occurs, one can drag and drop the mouse to define a region of AUC calculation.	 Or one can use the pop-up window to auto populate the AUC table. Tangential skim is a method to calculate AUC and peak height for a slopping baseline.
9	One can click the TIC pane **INCLUDE RANGE BAR** to add peaks manually, then select **Re-run Calculations** to perform a new analysis One can select **View Component Hit List in MineIt** to see Component table contents in **MineIt**	

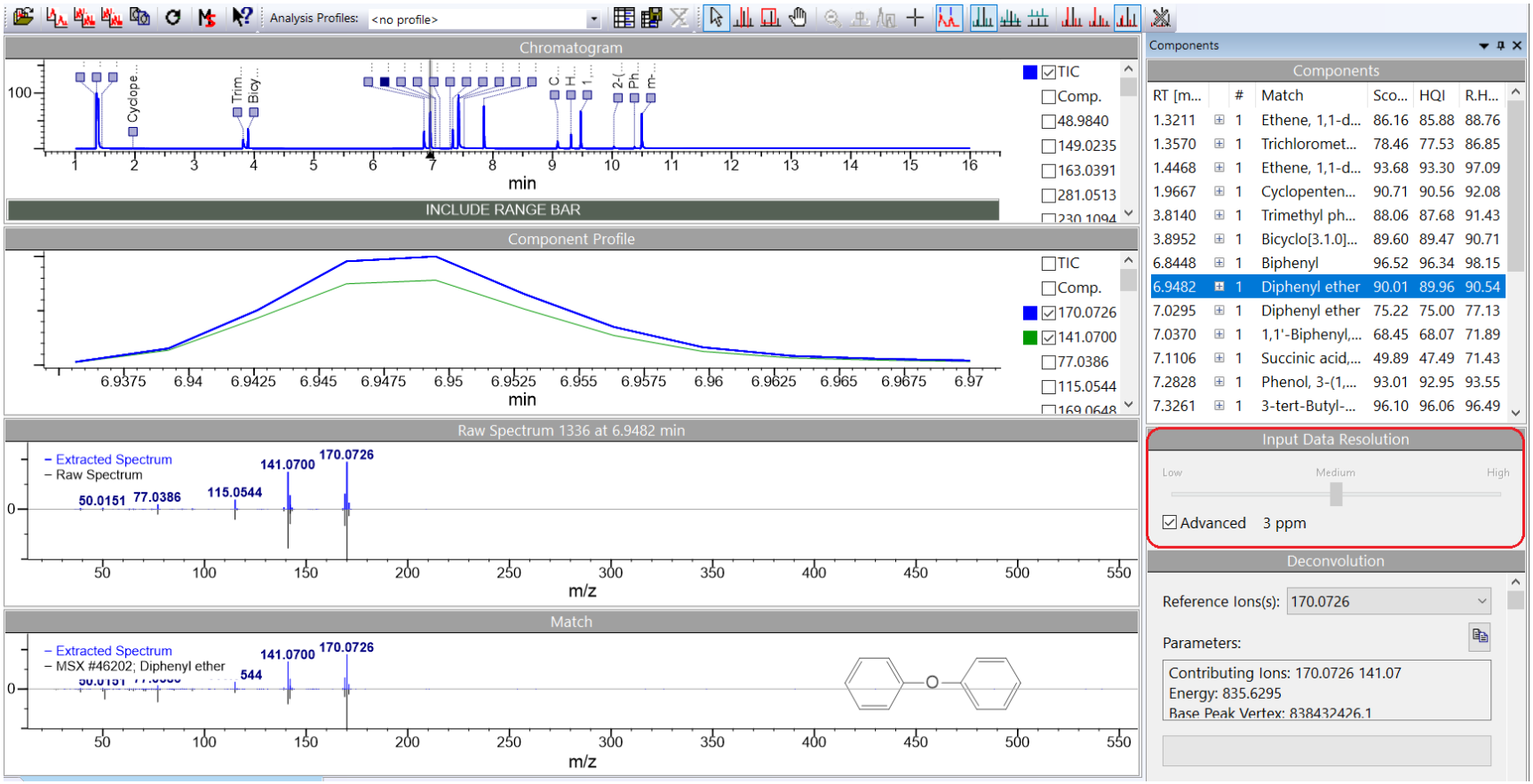
Example 3: High-resolution GC-MS

GUI explanation

For accurate *m/z* value data, without knowing the instrument resolving power, there is no safe algorithm to automatically calculate this value. Therefore, we use what our research considers a reasonable value by default that has a constant component and a variable component depending on mass (ppm). Empirically, this works in most cases. Increasing the *m/z* value accuracy too much incurs the danger of splitting an induvial *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 be merged resulting incorrect accurate *m/z* values to be reported.

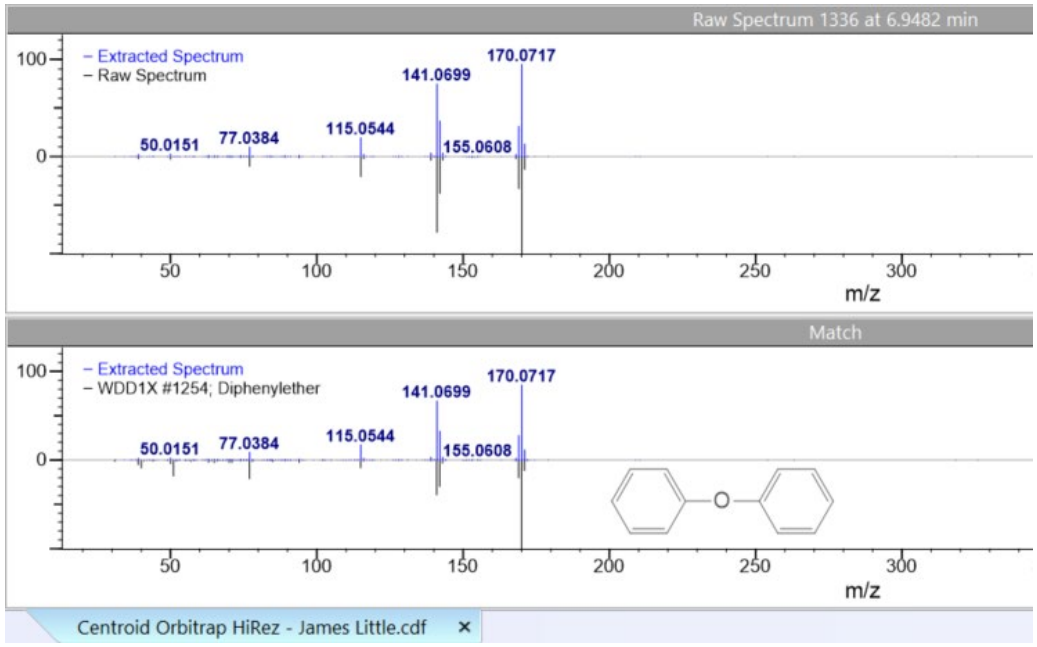


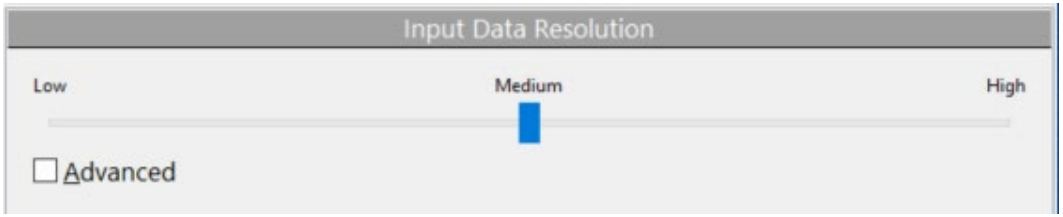
If a user knows the instrument's resolving power, that value should be entered in the highlighted dialog in above figure:

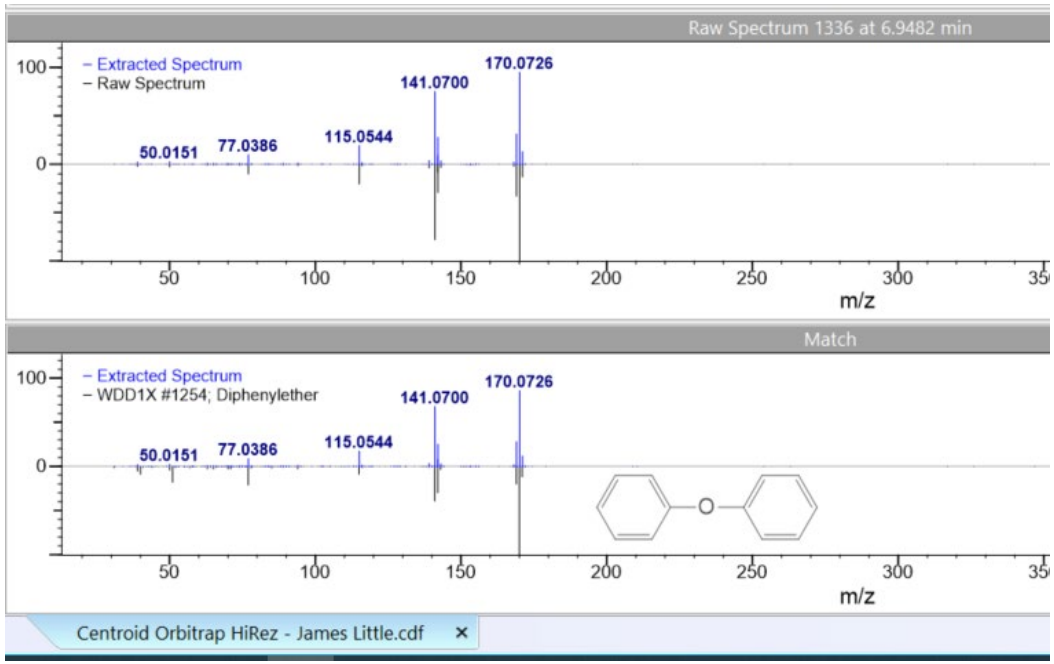


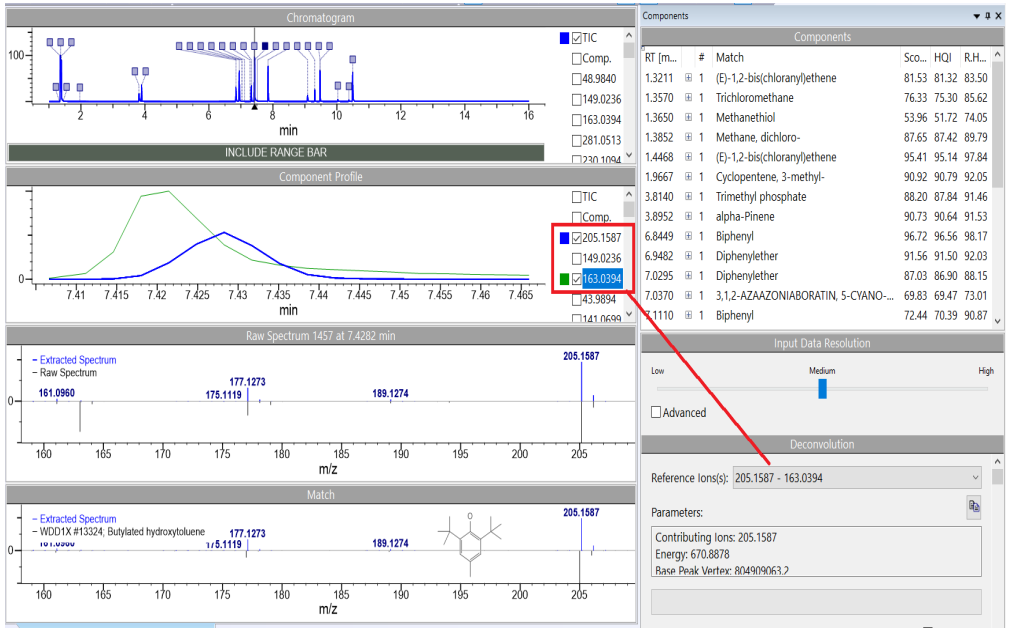
It is possible to save it as part of a profile, which can then be selected depending on the type of instrument. A user can create several profiles with different resolution settings for different types of data (and instruments).

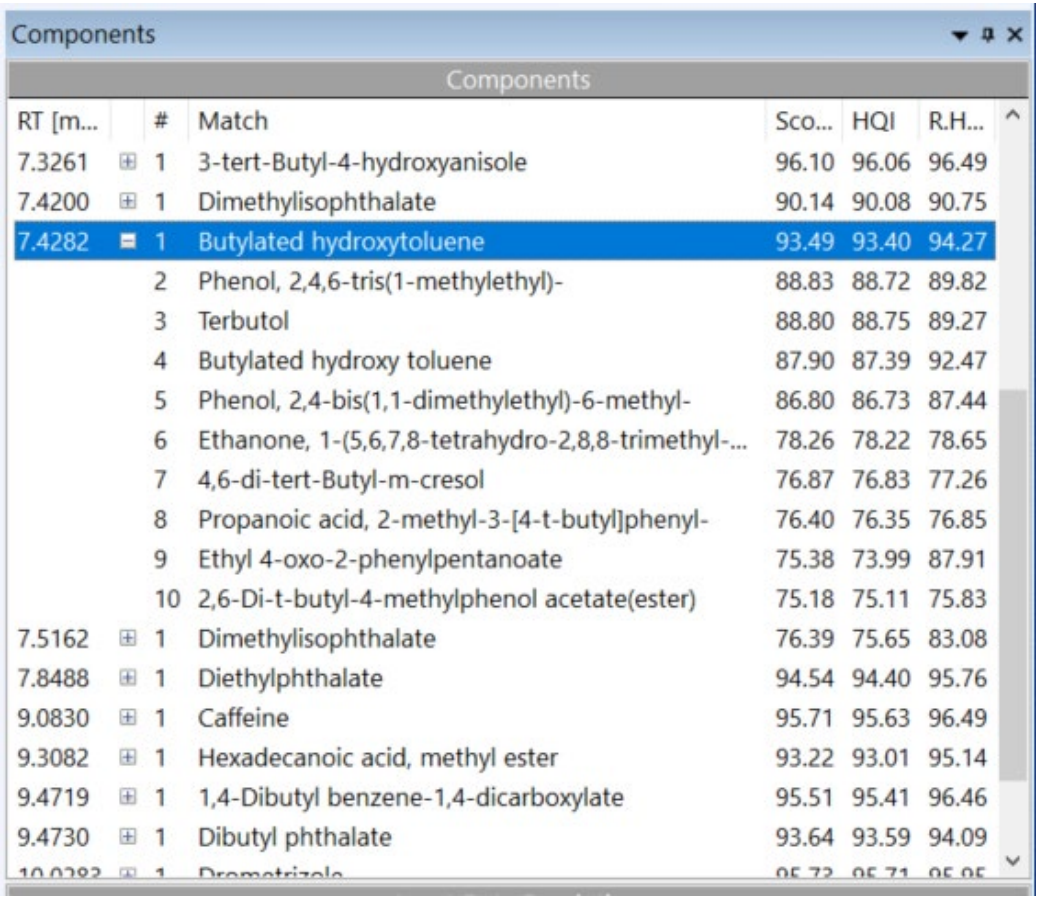
Excise 1

	Action	Result
1	Open a high-resolution GC-MS file in the same folder Centroid Orbitrap HiRez - James Little Navigate to hit list **RT (MIN)** 6.9482	 The m/z value of 170.0717 is slightly off for this compound.

	Action	Result
2	Check **Advanced** Set the instrument resolution at 3 ppm, **OK**	**MS Expert** recognizes this is a high-resolution data file and uniquely allows user to fill in resolution information or adjust resolution: 

	Action	Result
3		 The m/z value of 170.0726 is correct. This instrument resolution set-up is persistent.

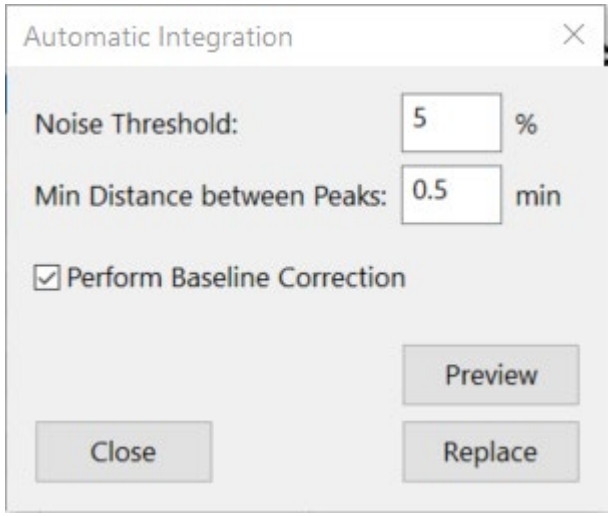
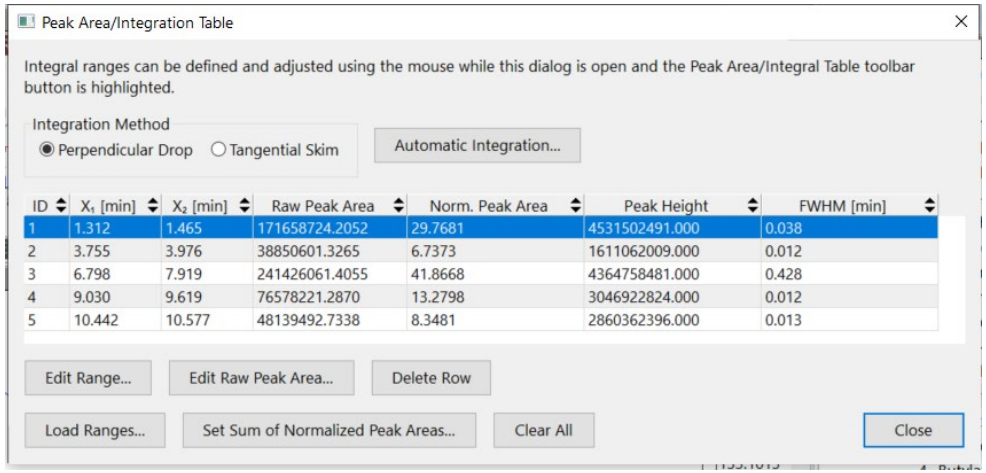
	Action	Result
4	Navigate to component TIC 7.4282	One can see **MS Expert** separates two components characterized by peaking at slightly different TIC values. The m/z values are 205.1587 and 163.0394, respectively. Component 205.1587 is in the extracted spectrum in this case. 

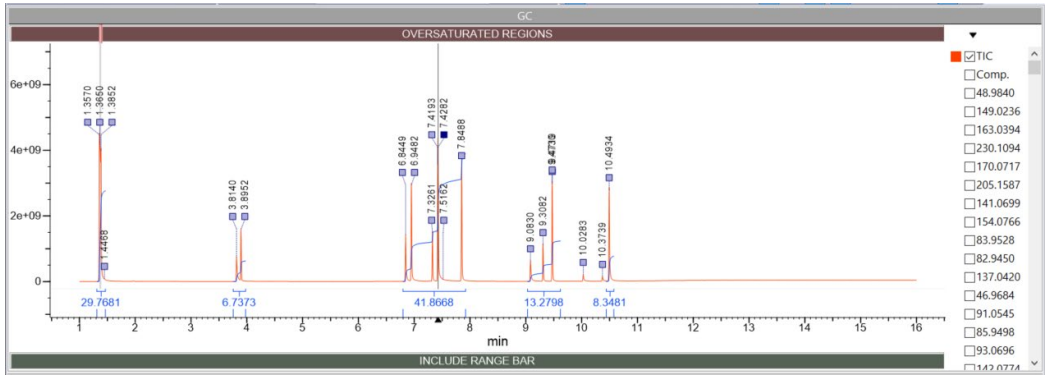
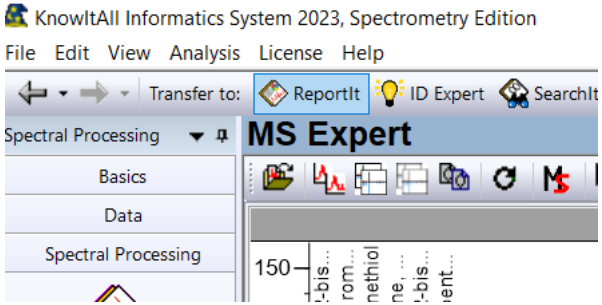
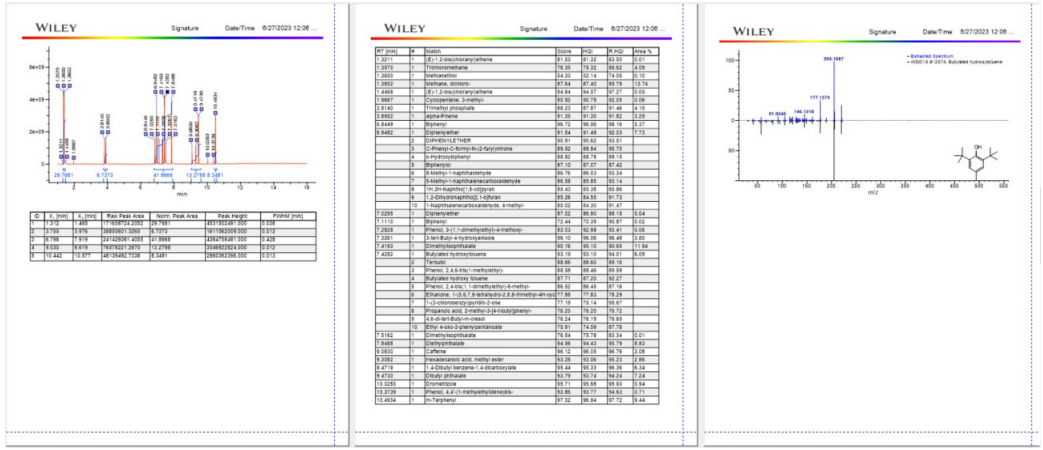
	Action	Result
5	Expand a few component hits, for example, RT 7.4	

Components

Components

RT [m...		#	Match	Sco...	HQI	R.H...
7.3261	⊞	1	3-tert-Butyl-4-hydroxyanisole	96.10	96.06	96.49
7.4200	⊞	1	Dimethylisophthalate	90.14	90.08	90.75
7.4282	⊟	1	Butylated hydroxytoluene	93.49	93.40	94.27
		2	Phenol, 2,4,6-tris(1-methylethyl)-	88.83	88.72	89.82
		3	Terbutol	88.80	88.75	89.27
		4	Butylated hydroxy toluene	87.90	87.39	92.47
		5	Phenol, 2,4-bis(1,1-dimethylethyl)-6-methyl-	86.80	86.73	87.44
		6	Ethanone, 1-(5,6,7,8-tetrahydro-2,8,8-trimethyl-...	78.26	78.22	78.65
		7	4,6-di-tert-Butyl-m-cresol	76.87	76.83	77.26
		8	Propanoic acid, 2-methyl-3-[4-t-butyl]phenyl-	76.40	76.35	76.85
		9	Ethyl 4-oxo-2-phenylpentanoate	75.38	73.99	87.91
		10	2,6-Di-t-butyl-4-methylphenol acetate(ester)	75.18	75.11	75.83
7.5162	⊞	1	Dimethylisophthalate	76.39	75.65	83.08
7.8488	⊞	1	Diethylphthalate	94.54	94.40	95.76
9.0830	⊞	1	Caffeine	95.71	95.63	96.49
9.3082	⊞	1	Hexadecanoic acid, methyl ester	93.22	93.01	95.14
9.4719	⊞	1	1,4-Dibutyl benzene-1,4-dicarboxylate	95.51	95.41	96.46
9.4730	⊞	1	Dibutyl phthalate	93.64	93.59	94.09
10.0283	⊞	1	Prometrizole	95.73	95.71	95.95

	Action	Result
6	Select **Analysis > Peak Area/Integration** Select **Automatic Integration** Set the parameters in the **Automatic Integration** pop-up window	
7	Click **Replace** Click **Close** button	We only use a few peaks to demonstrate this feature. 

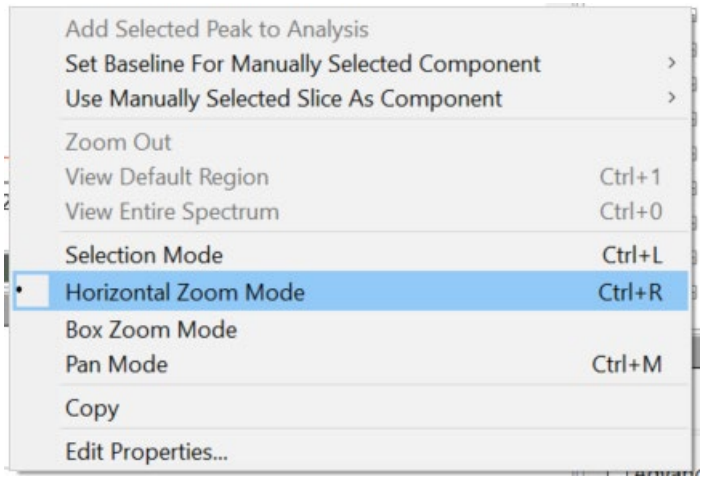
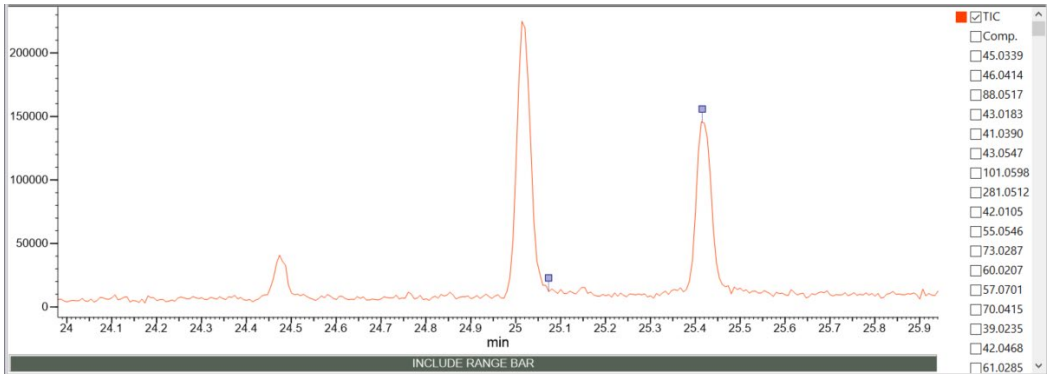
	Action	Result
8	The GC pane displays the peak integration values	
9	Click **Transfer to: ReportIt** 	

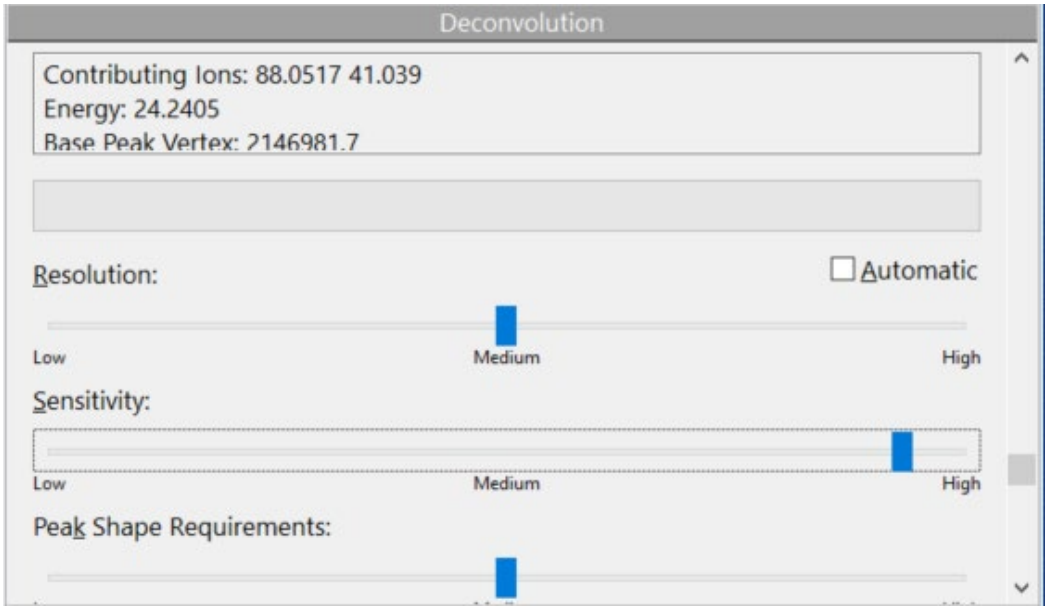
	Action	Result
10	Go back to **MS Expert** Right-mouse click the **Components** table and select **Copy All Component Information**. One can then paste this information to Excel or Word, etc..	

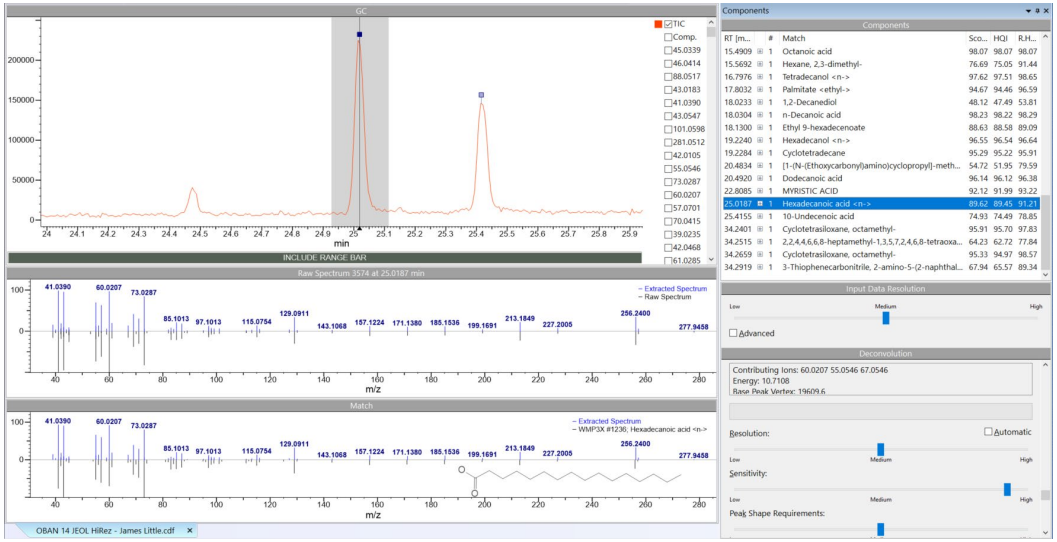
RT [min]	#	Match	Score	HQI	R.HQI	DB	ID
1.3211	1	(E)-1,2-bis(chloranyl)ethene	88.34	88.04	90.99	WMS3X	4144
	2	Ethene, 1,1-dichloro-	86.55	86.26	89.15	MSX	3048
	3	Ethene, 1,1-dichloro-	84.14	82.84	95.9	WMSR3X	1983
	4	2,2,3-Trichloropropionaldehyde	78.8	78.56	80.92	WMS3X	56863
	5	2,2,3-Trichloropropane-1,1-diol	75.53	75.3	77.6	WMS3X	88275
	6	Dihydro-4,5-dichloro-2(3H)furanone	72.69	71.59	82.61	MSX	30922
	7	Propanoic acid, 2,2,3-trichloro-	68.81	68.59	70.83	WMS3X	84684
	8	Propanoic acid, 2,2,3-trichloro-	59.81	57.45	80.97	WMS3X	84685
	9	Chloromethylmethyl sulfide	57.41	57.23	59.07	MSX	3054
	10	Methyl 2,3,3-trichloropropanoate	53.98	53.77	55.85	MSX	64271
1.357	1	Trichloromethane	78.5	77.57	86.84	MSX	9775
	2	Methane, oxybis[dichloro-	78.48	77	91.75	MSX	56380
	3	Methane, trichloro-	78.05	76.69	90.33	WMS3X	13436
	4	Ethane, 2-bromo-1,1-dichloro-	69.34	68	81.41	WMSD3X	330
	5	Ethane, 1,1,2,2-tetrachloro-	68.51	68.31	70.3	WMSR3X	13033
	6	1,1',3-trimethyl-3,3'-biindolin-2-one	67.98	65.35	91.65	WMSD3X	1193
	7	Methane, dichloronitro-	67.11	65.71	79.7	MSX	14617
	8	N,N-Dimethyl-2H-pyran-2-iminium chloride	64.16	62.24	81.49	WMSD3X	219
	9	Ethane, 1,2,2-trichloro-1,1-difluoro-	63.71	62.79	71.95	MSX	42934
	10	N-(Phenyl MIDA boronate-4-yl)-S-methyl-S-phe	63.11	59.87	92.28	WMS3X	576845
1.4468	1	(E)-1,2-bis(chloranyl)ethene	93.9	93.52	97.26	WMS3X	4143
	2	Ethene, 1,1-dichloro-	93.88	93.5	97.28	MSX	3048
	3	(E)-1,2-bis(chloranyl)ethene	93.86	93.48	97.32	WMS3X	4144
	4	Ethene, 1,1-dichloro-	92.61	91.89	99.11	WMSR3X	1983
	5	2,2,3-Trichloropropionaldehyde	81.63	81.34	84.21	WMS3X	56863
	6	Dihydro-4,5-dichloro-2(3H)furanone	81.56	81.04	86.23	MSX	30922
	7	2,2,3-Trichloropropane-1,1-diol	77.78	77.5	80.3	WMS3X	88275
	8	Propanoic acid, 2,2,3-trichloro-	76.16	75.88	78.67	WMS3X	84684
	9	Propanoic acid, 2,2,3-trichloro-	68.17	66.78	80.63	WMS3X	84685
	10	Chloromethylmethyl sulfide	65.61	65.46	66.99	MSX	3054
1.9667	1	Cyclopentene, 3-methyl-	90.7	90.55	92.08	MSX	1382
	2	Cyclopentene, 1-methyl-	90.19	90.07	91.25	MSX	1367
	3	Cyclopentene, 4-methyl-	89.11	88.84	91.55	WMS3X	1919
	4	Cyclopentene, 4-methyl-	89.05	88.94	90.1	MSX	1389
	5	(Z),(Z)-2,4-Hexadiene	87.84	87.69	89.17	MSX	1356

Note: component Area% column is exported but removed from the above Excel display

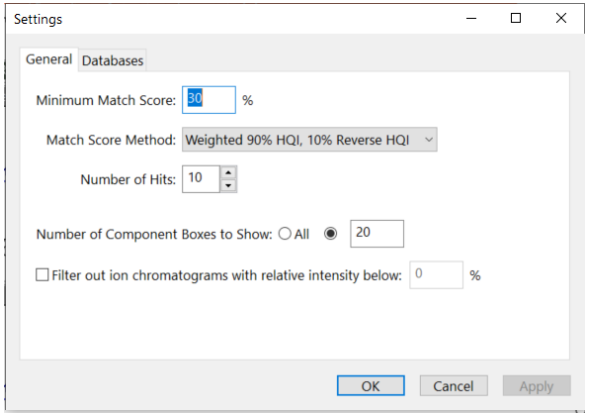
Excise 2

	Action	Result
1	Open another high-resolution GC-MS file in the same folder OBAN 14 JEOL HiRez - James Little Right-mouse-click the **TIC** pane Choose Horizontal Zoom or Box Zoom mode	
2	Zoom to region 24 – 26 min (3 tiny peaks):	 We have two issues here: first, tiny peak was not picked up for analysis; and second, component was assigned to baseline.

	Action	Result
3	We can increase the algorithm Sensitivity to resolve the second issue. Check off **Automatic** under **Deconvolution** and move the **Sensitivity** bar to **High**	

	Action	Result
4		Now the second component is correctly identified as the missing hexadecanoic acid 

Tip: Sometimes, there are too many GC peaks crowding the GC pane. One can control how to display by using **File > Settings** to set preferences:



WILEY

Manual GC-MS Analysis

How to Use KnowItAll ProcessIt Perform Manual GC-MS Analysis

Purpose

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

Objectives

These exercises will teach you:

- How to use KnowItAll ProcessIt.

Background

The KnowItAll ProcessIt software reviews GC-MS data and, allow a user to control spectral subtraction. Subtracted spectrum is searched for matches in reference data.

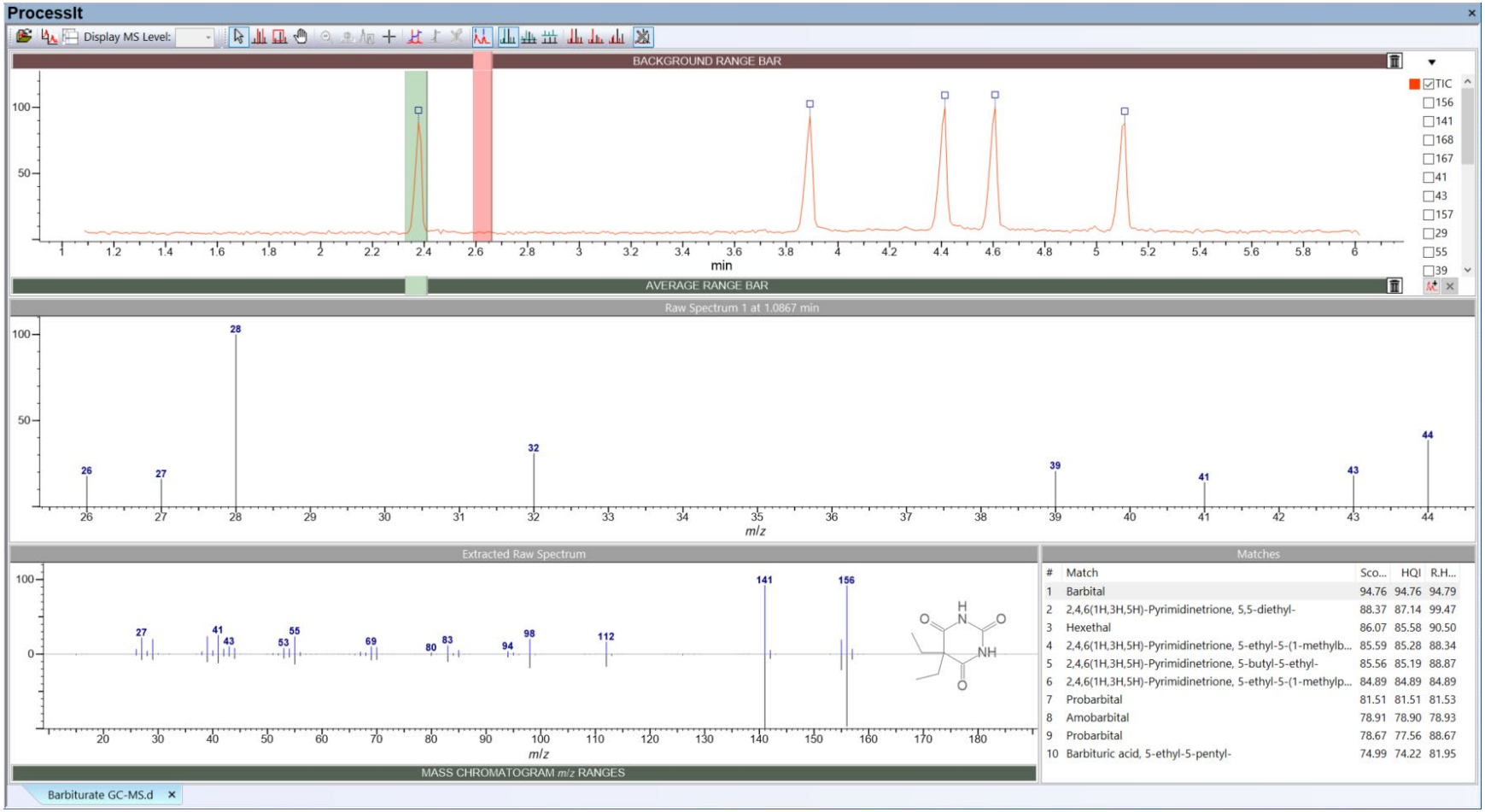
Training Files Used in This Lesson

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

KnowItAll Applications Used

- KnowItAll ProcessIt

GUI explanation



The bottom spectrumPane holds

- Selected MS, defined by left mouse clicking or
- Selected MS – background (red bar in GC pane), defined by clicking the mouse on the bar, drag and drop

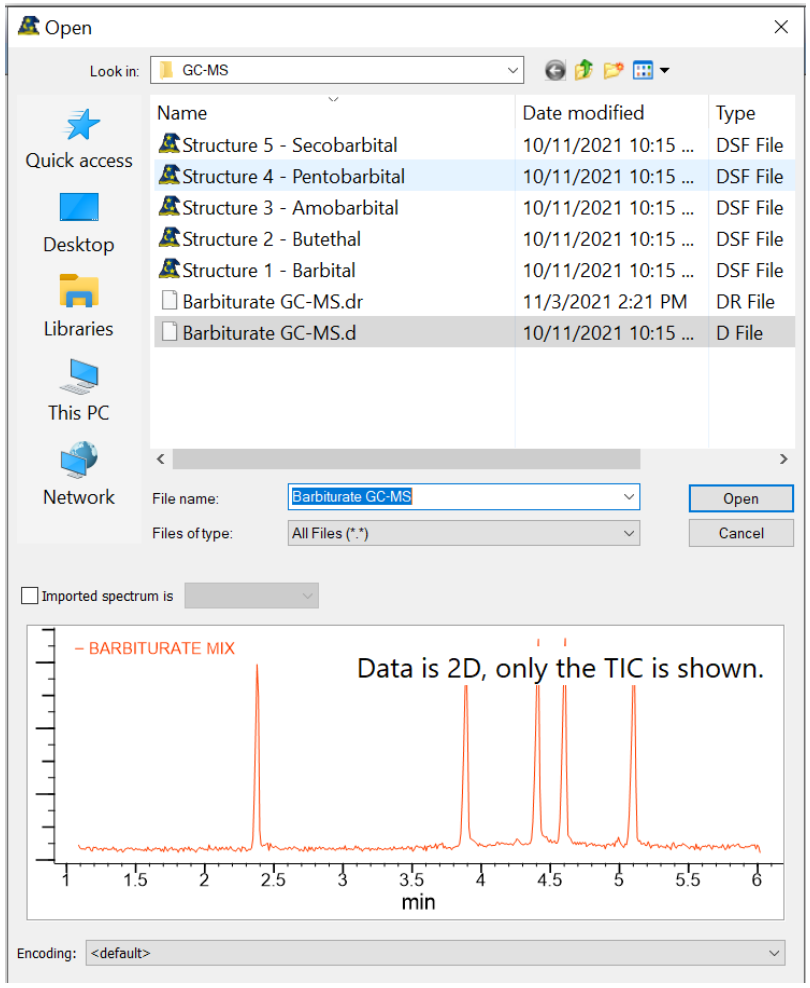
- Average MS (green bar in GC pane), defined by by clicking the mouse on the bar, drag and drop
- Average MS (green bar) - background (red bar)

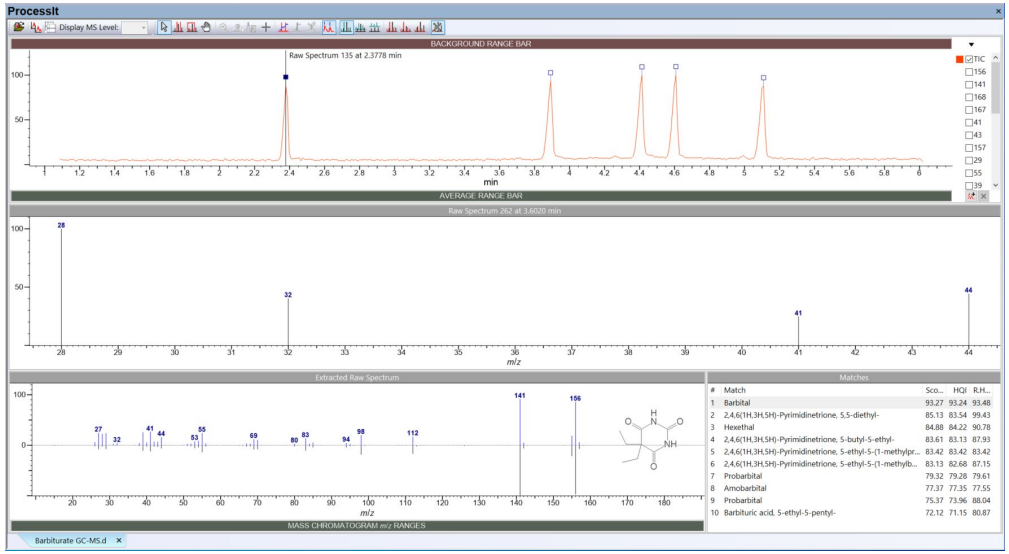
The **Matches** table are database search hit list for the selected MS.

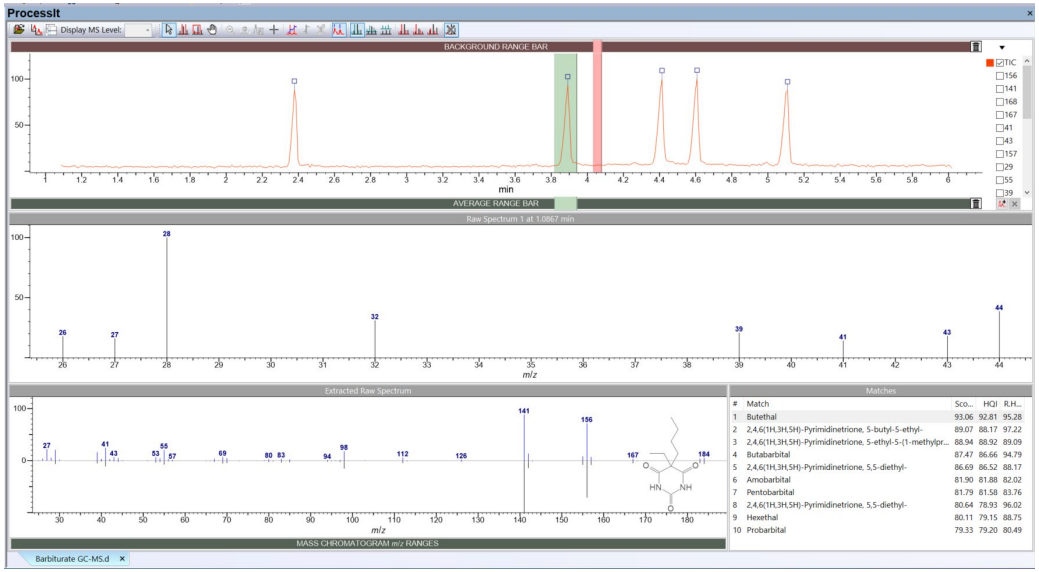
The selected MS can be transferred to **SearchIt** application for spectrum search.

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.

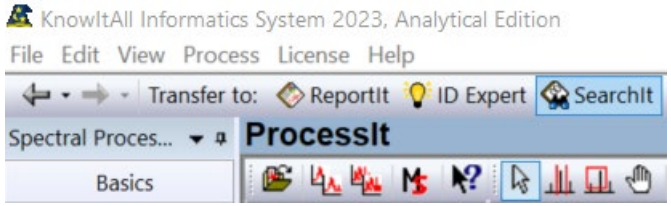
Example: ProcessIt MS

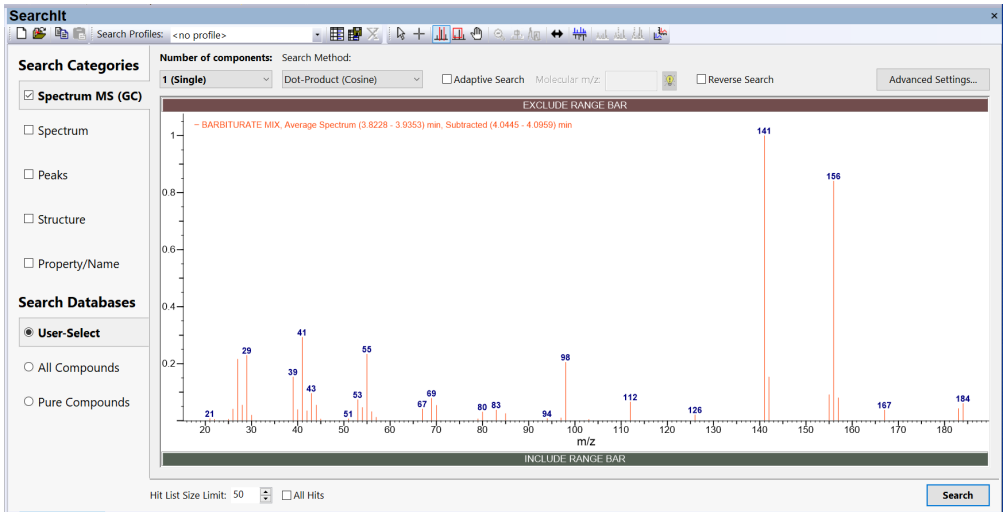
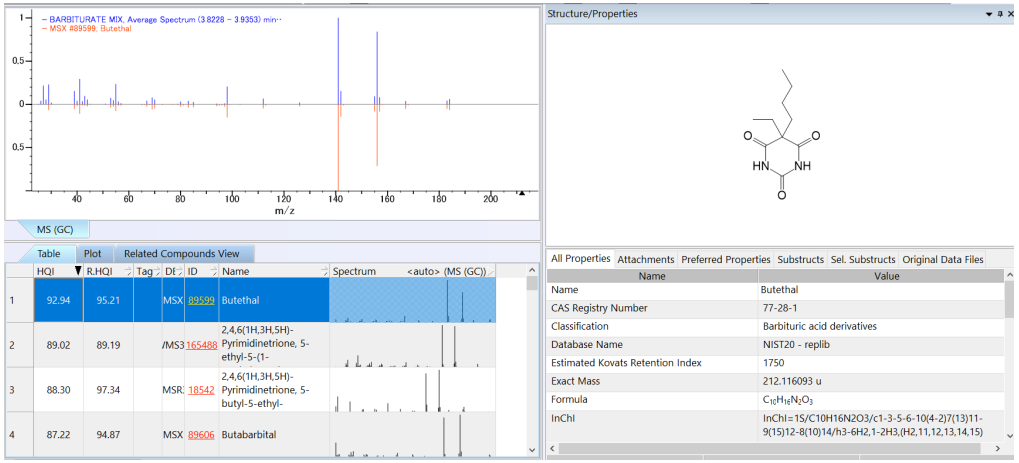
	Action	Result
1	In **ProcessIt**, click **Open Data File** button Navigate to "C:\Users\Public\Documents\Wiley\KnowItAll\Samples\GC-MS" Select Barbiturate GC-MS.d **Open**	

	Action	Result
2	Click a peak in chromatogram pane (top pane)	You would see that the corresponding MS and hit database search results are reflected in the bottom pane: 

	Action	Result
3	Define average **spectrum range** • Click the top (green) bar • Drag and • Drop Define background spectrum **range** • Click the top (red) bar • Drag and Drop	 The database search result is default to display, however, one can turn off this by **Process > Search Extracted MS Spectrum for Database Matches** menu-item.

Example: Transfer to SearchIt

	Action	Result
1	Click **Transfer to: SearchIt**	

	Action	Result
2	In **SearchIt**, one can perform various MS spectral searches. Check off **Adaptive Search** Make sure all licensed databases are selected in **User-Select** **Search**	
3		 The above is an expected hit.