

Getting Started with MASH Explorer

Version 2.1

MASH Explorer, developed by the Ying Ge research group, is a comprehensive, universal, and user-friendly, free software environment for top-down proteomics. This tutorial provides a brief introduction to the basic functions of MASH Explorer in both Discovery Mode (LC-MS/MS datasets) and Targeted Mode (single or targeted protein datasets) for top-down proteomics analysis. Download the sample data and follow the instructions below to begin using MASH Explorer. For a detailed description of the functions of MASH Explorer software, please view the *User Manual* included with your software download at “C:\Program Files\Human Proteomics Program\MASH Explorer\doc\MASH_Explorer_User_Manual.pdf”.

Installation:

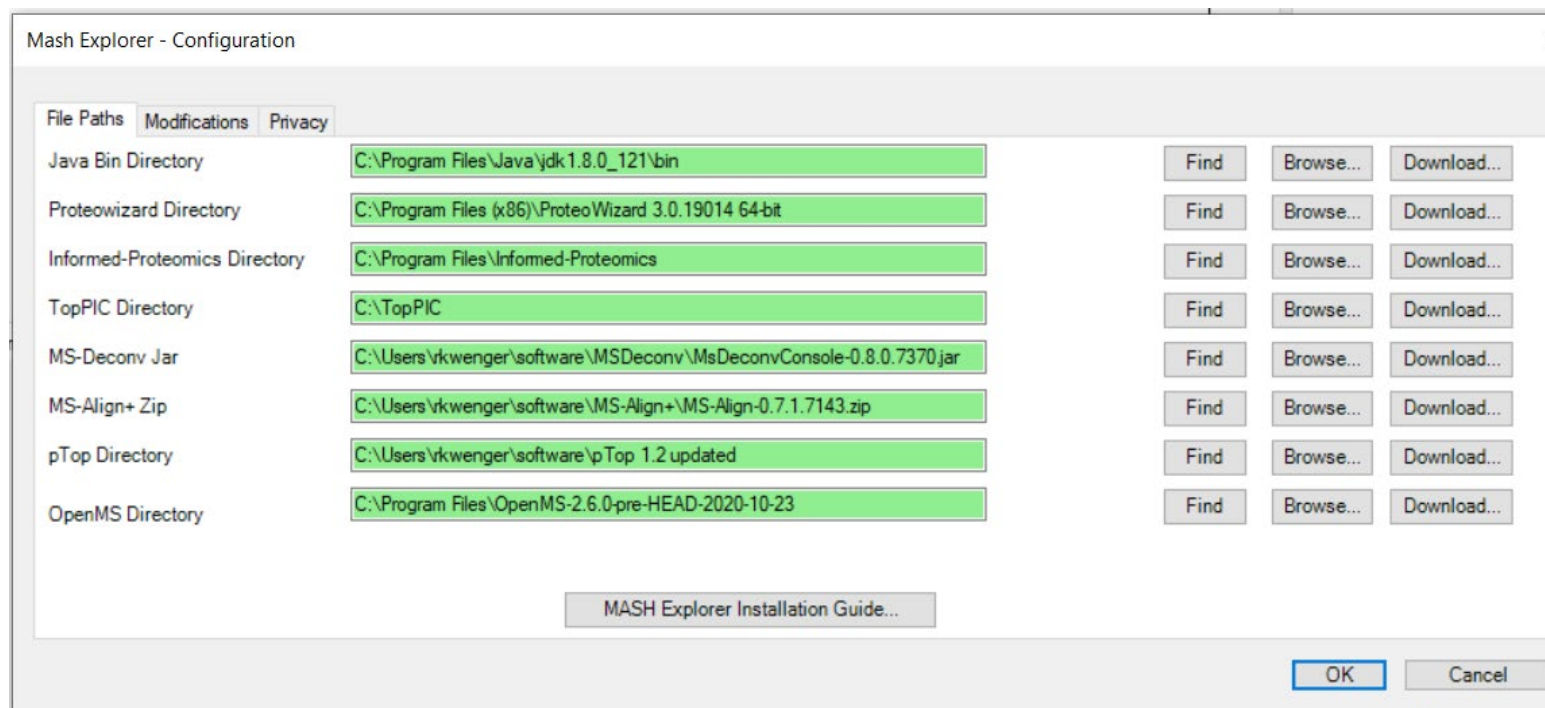
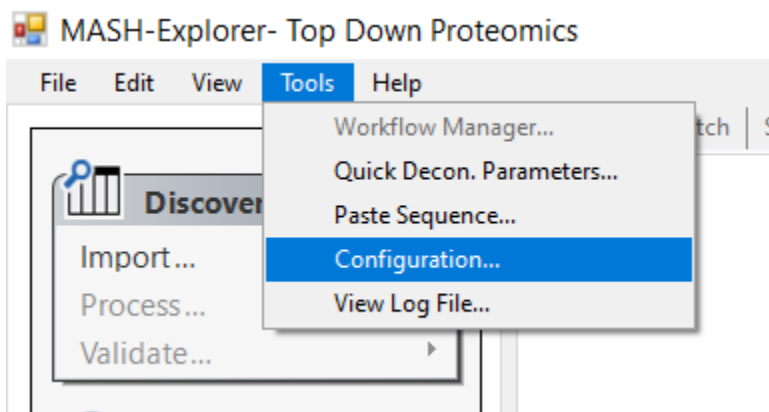
To begin, please install MASH Explorer by referring to the installation guide, available on our [webpage](#).

Privacy/Usage Reporting:

In MASH Explorer V2.0.1 and later, MASH Explorer may report data about your usage of the software to the MASH team. Each time you start a new version of the software, you will be asked to review your privacy/usage reporting settings. The GUI for this should be quite intuitive, but if you want more details please see section 3 of the *User Manual*. Please refer to the MASH Explorer Privacy Notice (https://labs.wisc.edu/gelab/MASH_Explorer/doc/MASH%20Explorer%20Privacy_Notice.pdf) for information about what data may be reported and what is done with that data.

Associated Software and Configuration:

Make sure that all associated software you need (for example, deconvolution and search tools) is installed and properly configured in the Configuration dialog (accessed via the Tools menu).

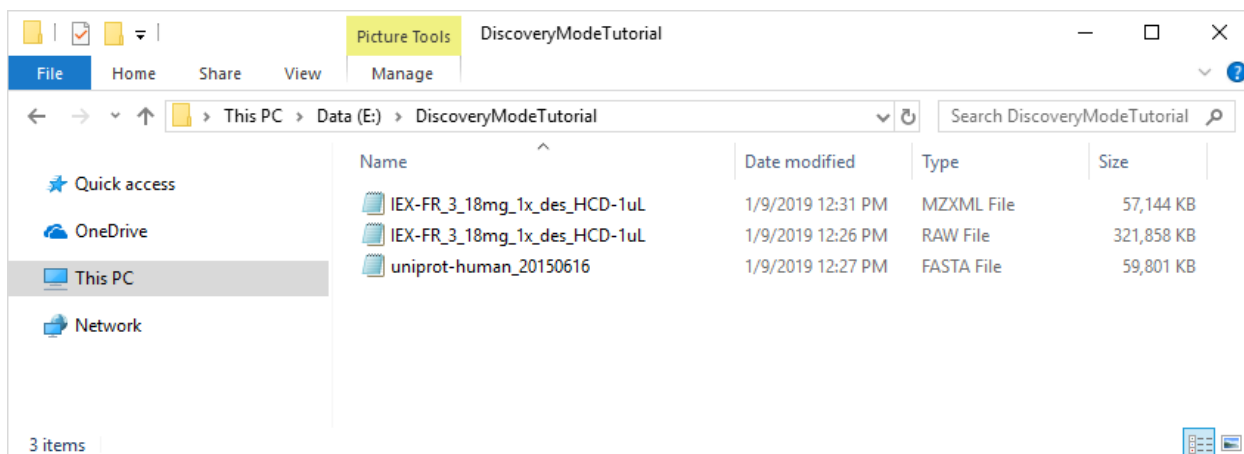


Please note that this tutorial uses TopPIC Suite version 1.3.3. Previous versions of TopPIC Suite are also supported.

Discovery Mode:

Discovery Mode is intended for LC-MS/MS data analysis. Please follow the instructions in User Manual *Section 4 Discovery Mode: Protein Identification and Characterization from LC-MS/MS* if you require further assistance processing the data.

This sample data shows MASH Explorer's capability for LC-MS/MS data analysis, which includes spectral deconvolution and database search. To follow this tutorial, download these files and save them to a new local folder called “DiscoveryModeTutorial” so that your working directory is similar to “E:\DiscoveryModeTutorial” as seen below:

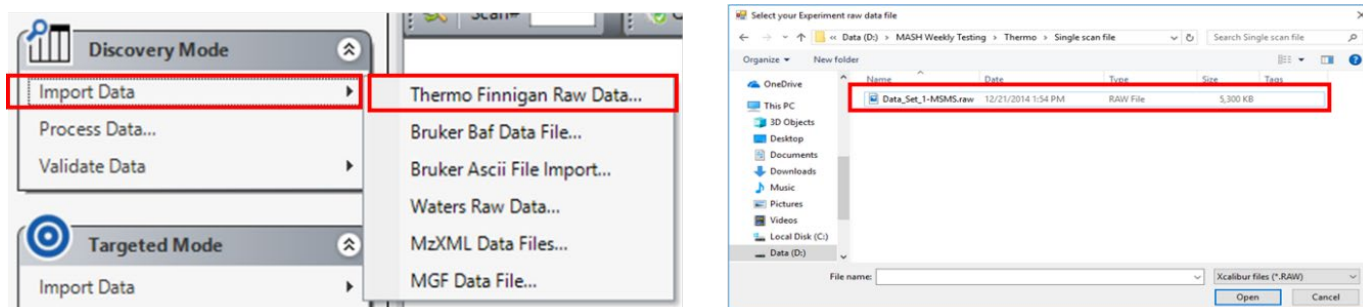


Sample Files: Download these files available on the Google Drive.

- Sample Thermo raw data: [IEX-FR_3_18mg_1x_des_HCD-1uL.raw](#).
- Sample 32-bit mzXML file: [IEX-FR_3_18mg_1x_des_HCD-1uL.mzXML](#)
- FASTA file: [uniprot-human_20150616.fasta](#).

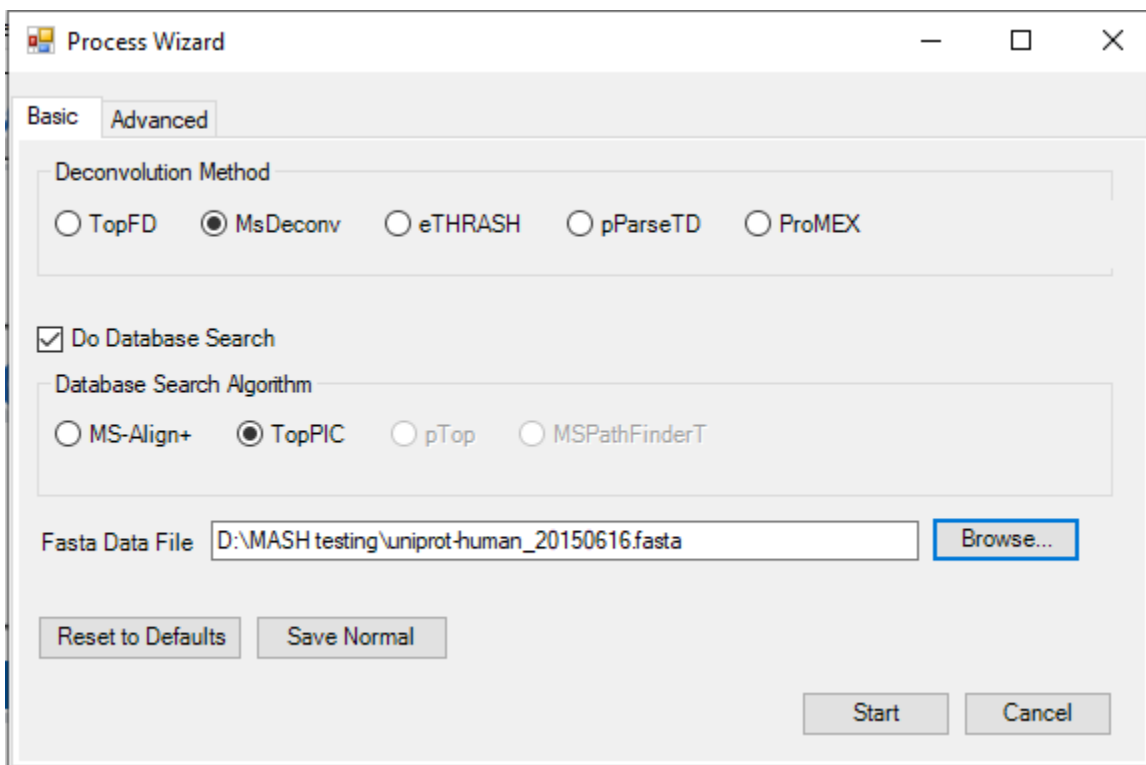
Step 1: Import Data

Use the Import Data function in the Discovery Mode to import the sample Thermo raw data file.

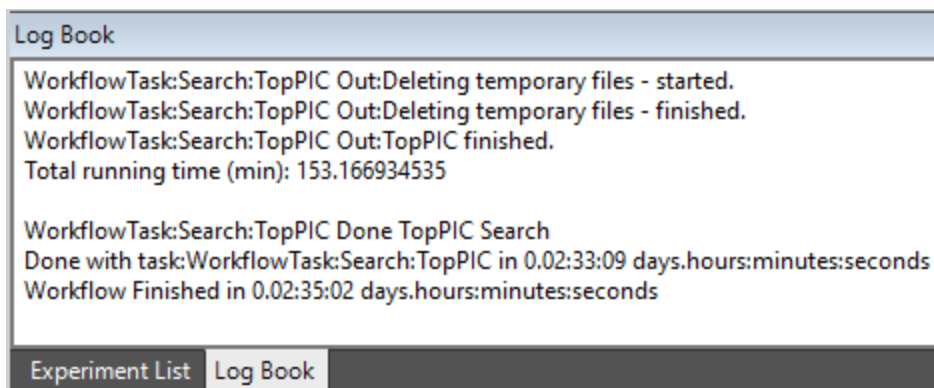


Step 2: Spectral deconvolution and Database Search

Click the Process Data menu item in the Discovery Mode section to open up the Process Wizard dialog. The Process Wizard will setup the series of tasks for selecting, configuring, and running the deconvolution and database search algorithms. Select the MSDeconv deconvolution method and the TopPIC database search algorithm. Next, click on the Browse button and select the sample FASTA file to load into MASH. To set advanced parameters, open the Advanced tab; however, it is not necessary to set these parameters to begin your workflow. Click on the Start button to begin your search.



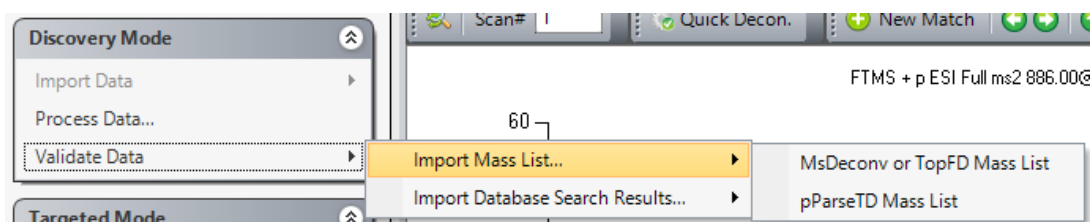
Starting your search will open the Workflow Manager window and the terminal window. The terminal window is where MSDeconv and TopPIC are running. You can minimize this window, but closing this window will stop your workflow. You can close and open the Workflow Manager window as you wish, and you can use the Workflow Manager to check on the status of your run. When your run is completed, you will see the workflow in the Completed Workflows tab of the Workflow Manager and the line “Workflow Finished” in the Log Book, as shown below.



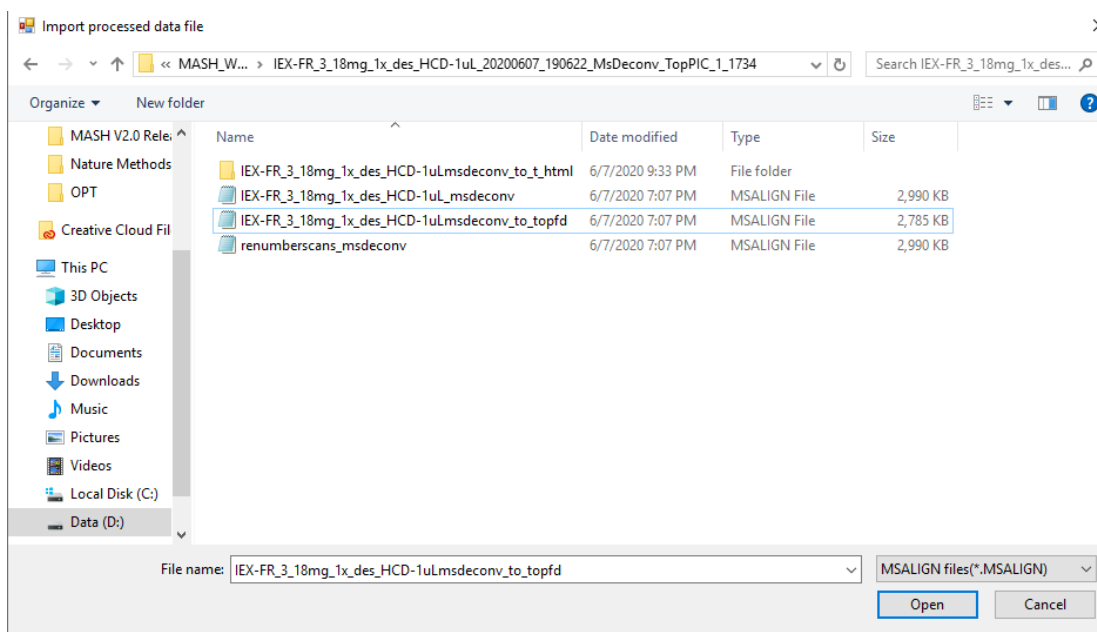
Step 3: Import Results

Your deconvolution and search results will be automatically imported into MASH Explorer if the run was successful. You can find your deconvolution results in the Mass List and your search results in the Protein Search Results List.

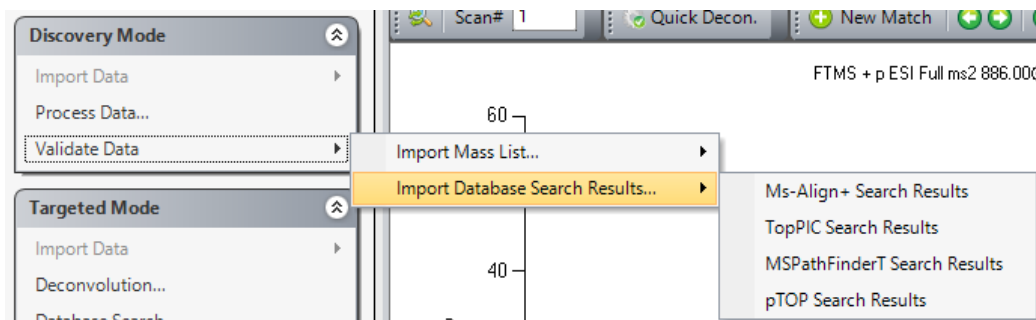
Follow the following steps if you wish to manually import your results. To import the deconvolution mass list for detailed analysis of the data set, select the Validate Data function in the Discover Mode panel, and click on Import Mass List.



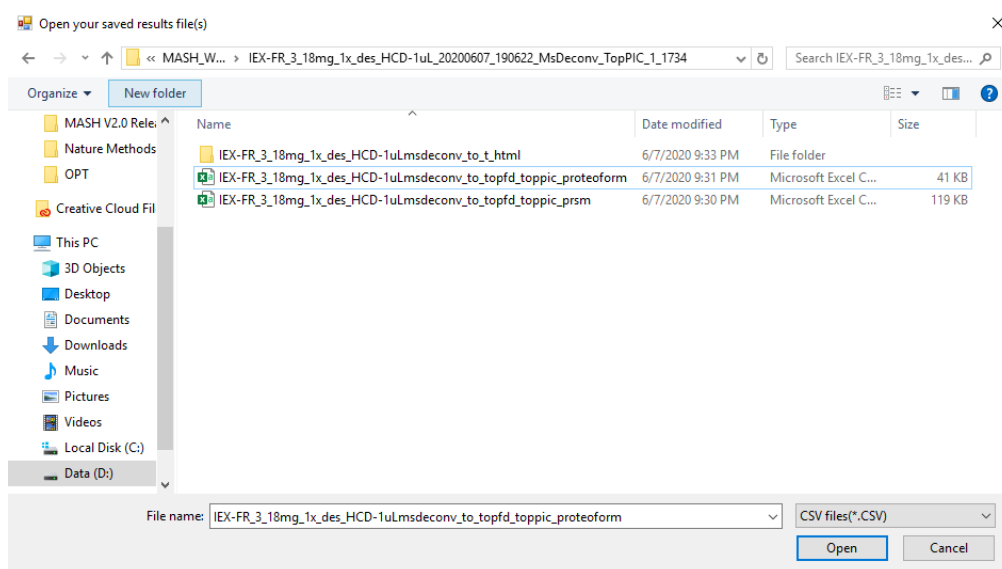
Your saved deconvolution results will be in your working directory under MASH_Workflows. The folder containing the results should look something like “MASH_Workflows\IEX-FR_3_18mg_1x_des_HCD1uL_MsDeconv_TopPIC_1_1734_4112019_11_5”. Double click on the folder and select the file IEX-FR_3_18mg_1x_des_HCD-1uLmsdeconv_to_topfd.msalign as seen below. The loaded results will populate the Mass List. You may wish to select “Overwrite” in order to avoid viewing duplicated results.



To import the database search results for detailed analysis of the data set, select the Validate Data function in the Discovery Mode panel, and click on Import Database Search Results, and select TopPIC Search Results.



Navigate to your working directory and select the following file to import: IEX-FR_3_18mg_1x_des_HCD-1uLmsdeconv_to_topfd_toppic_proteoform.csv. You may wish to select “Overwrite” in order to avoid viewing duplicated results. Search results will be loaded into the Protein Search Results Table.

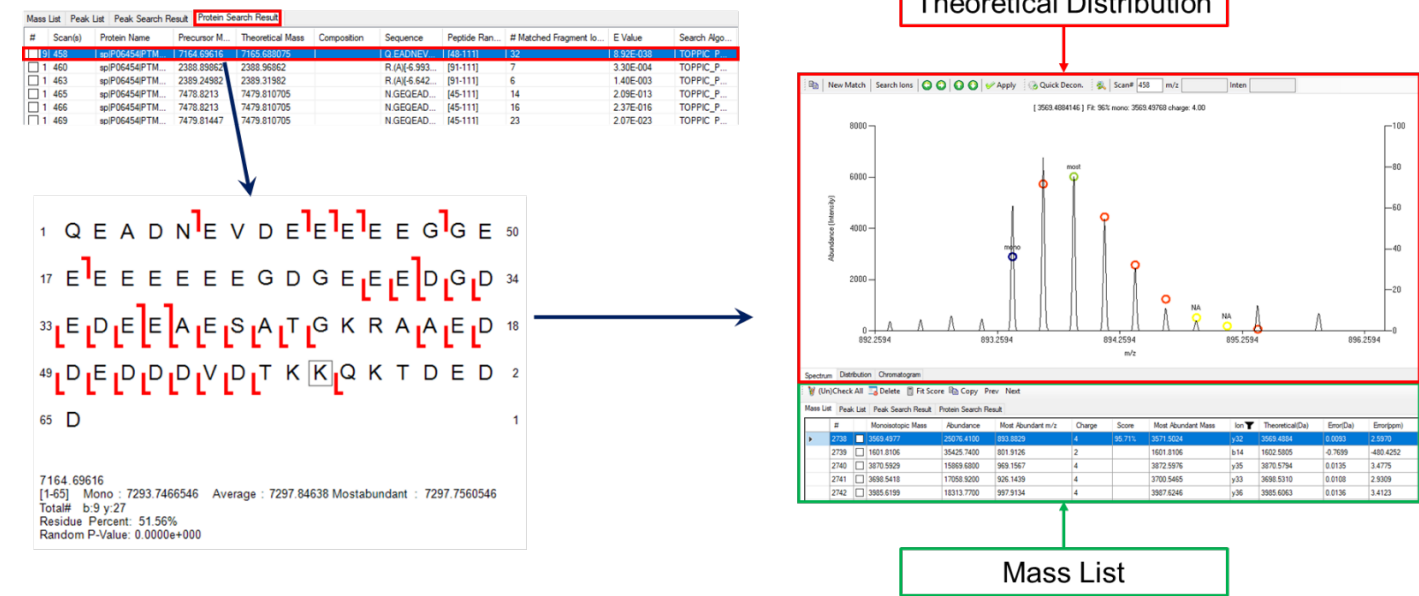


Step 4: Annotate Results

MASH Explorer allows you to view and annotate your imported results with the Mass List and Protein Search Results List, the Results View, and the Sequence Table. For more details on these functions, refer to the *User Manual*.

In the Mass List, click on some of the loaded masses to view the fragment ions of a selected scan in the dataset. Change the scan number in the Chromatogram or at the top of the Results View to view the mass fragments of a different scan. Use the arrows in the results view to alter the theoretical fit, and click the Apply Changes button to save your adjustments.

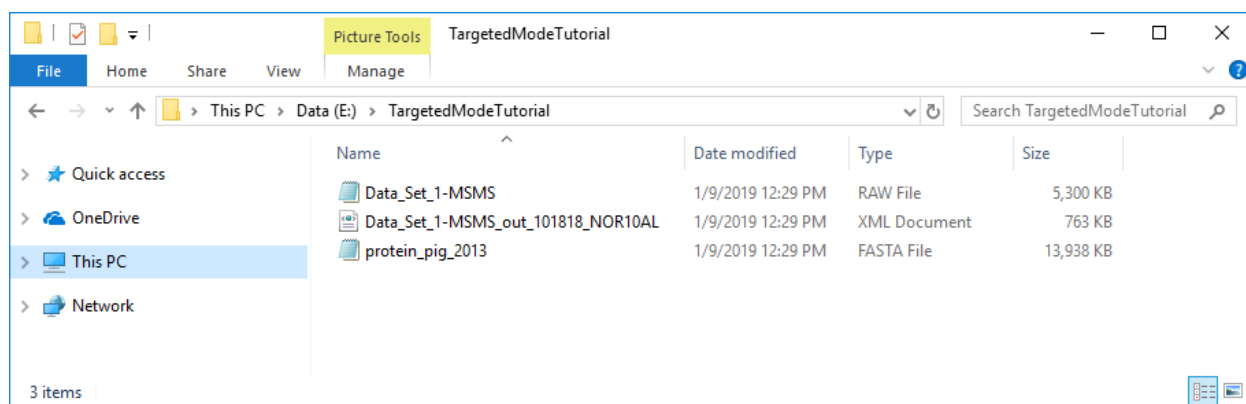
You can click on different entries in the Protein Search Results List. Notice that clicking on each scan updates the sequence in the Sequence Table, as seen below.



Targeted Mode:

Targeted Mode is intended for MS/MS data analysis. Detailed instructions on using this mode can be found in the User Manual at *Section 4 Targeted Mode: Protein Identification and Characterization from MS/MS Data*.

The following sample data demonstrates the MASH Explorer's functionality for single protein characterization. Download the following sample data and save the files to a similar local directory “E:\TargetedModeTutorial” as seen below:

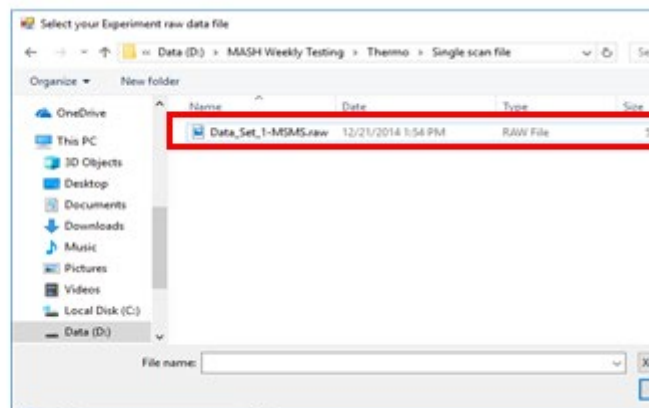
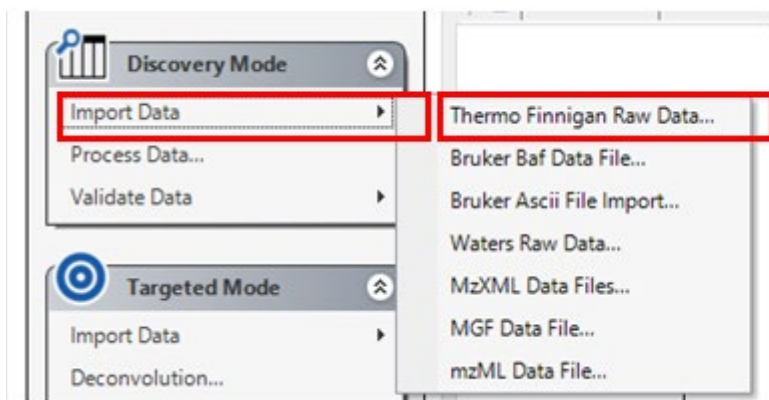


Sample Files: These data are downloaded via Google Drive.

- Sample Thermo raw data: [Data_Set_1-MSMS.raw](#).
- Sample 32-bit mzXML file: [Data_Set_1-MSMS.mzXML](#).
- FASTA file: [protein_pig_2013.fasta](#).
- Saved xml file: [Data_Set_1-MSMS_output.xml](#).

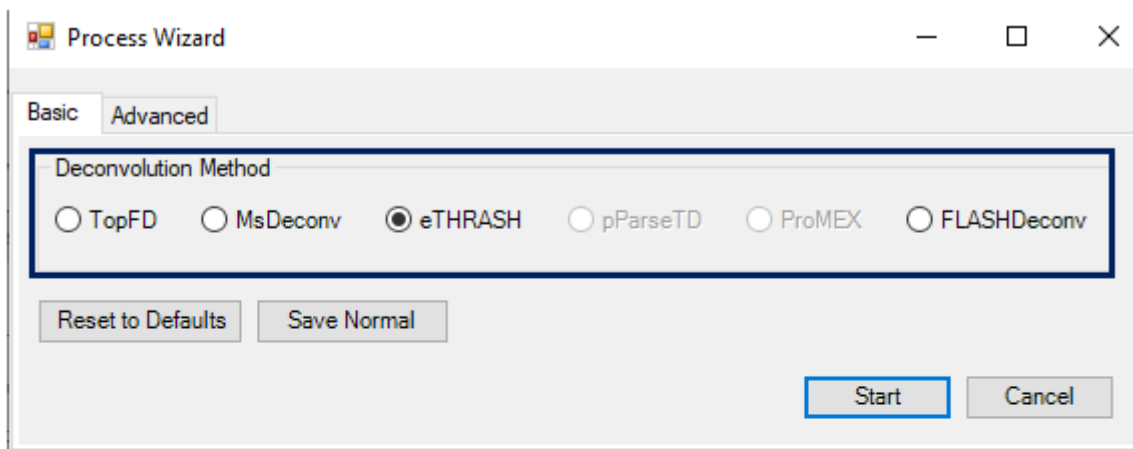
Step 1: Import Data

Use the Import Data function in Targeted Mode to import the sample Thermo data file.



Step 2: Deconvolution

Under the Targeted Mode, Process Wizard will guide users through the deconvolution step. Select the TopFD deconvolution method and click the Start button. You have the option to set more advanced parameters in the Advanced tab, but this is not necessary to start your run. The Workflow Manager and terminal window will open. You can close the Workflow Manager if you wish, but do NOT close the terminal because this is where your computer is running TopFD. When the deconvolution is finished, the Mass List will automatically populate for fragment ion validation, shown in the image below.



Process Wizard

Basic Advanced

Deconvolution Method

☐ TopFD
 ☐ MsDeconv
 ☒ eTHRASH
 ☐ pParseTD
 ☐ ProMEX
 ☐ FLASHDeconv

Reset to Defaults Save Normal

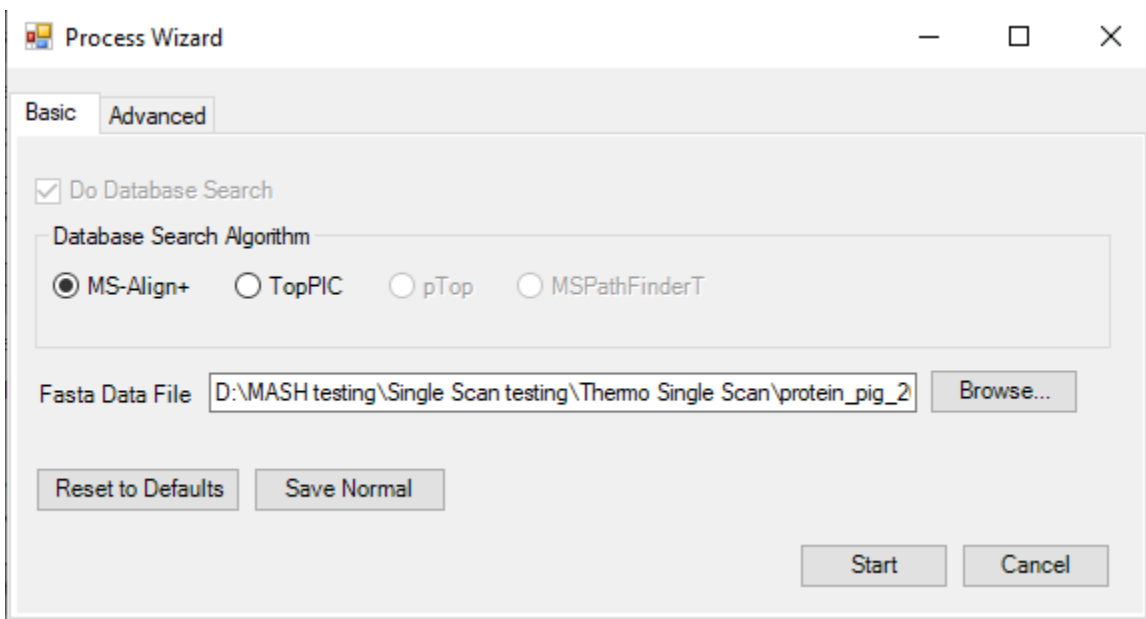
Start Cancel



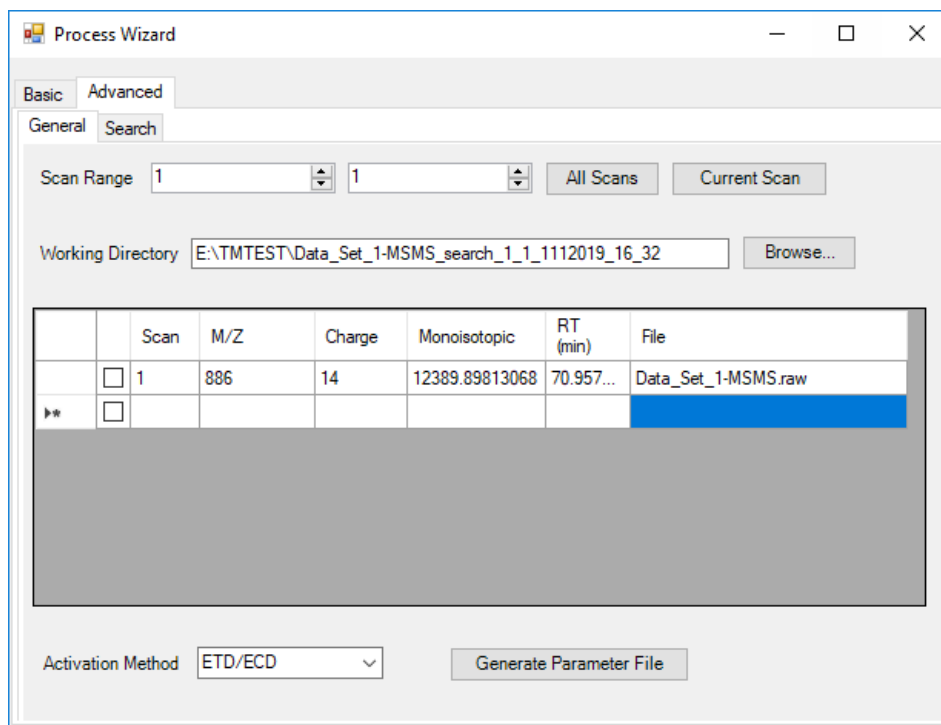
Mass List	Peak List	Peak Search Result	Protein Search Result									
	#		Monoisotopic Mass	Abundance	Most Abundant m/z	Charge	Score	Most Abundant Mass	Ion	Theoretical(Da)	Error(Da)	Error(ppm)
▶	82	<input type="checkbox"/>	3124.5959	5.7700	1042.8733	3		3125.5982	y28	3125.3392	-0.7433	-237.8366
	112	<input type="checkbox"/>	3326.0937	3.3300	257.0148	13		3328.0984	y30	3325.4189	0.6748	202.9158
	262	<input type="checkbox"/>	5634.7368	3.9000	940.6313	6		5637.7439	y50	5634.1447	0.5921	105.0967
	1	<input type="checkbox"/>	12339.3900	117.6000	1029.8745	12		12346.4065	NA			
	180	<input type="checkbox"/>	2773.4610	4.0900	397.3592	7		2774.4634	NA			
	179	<input type="checkbox"/>	12317.5490	5.0500	411.8261	30		12324.5655	NA			
	178	<input type="checkbox"/>	1360.7161	2.8000	1361.7234	1		1360.7161	NA			
	177	<input type="checkbox"/>	1350.9194	2.8500	338.7371	4		1350.9194	NA			
	176	<input type="checkbox"/>	1305.2345	2.5600	1306.2418	1		1305.2345	NA			
	175	<input type="checkbox"/>	5542.8412	3.0100	278.2997	20		5545.8483	NA			

Step 3: Database Search

When the fragment ions are verified, users can use the selected ion list for database search via the Process Wizard. In the Targeted Mode panel, select the Database Search function, which will open the Process Wizard. Select the MsAlign+ database search to run on the selected ion list shown in the Mass List. Click on the Browse button to load the sample FASTA file.

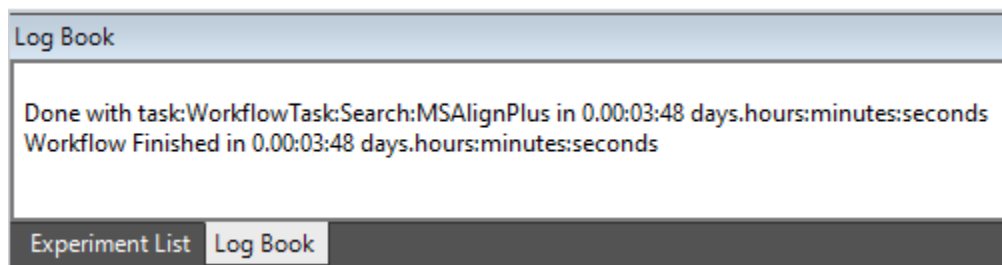


This file requires Advanced Parameters to be set. Please click on the Advanced Parameters tab. Change the charge from 0 to 14 and hit Enter, so that the parameters in the list are exactly as shown below. The monoisotopic mass will automatically be calculated. Ensure that the Activation is set at ETD/ECD. Navigate back to the basic tab and click on the start button to begin your search.



As before, the Workflow Manager and the terminal window will open. You can close the Workflow Manager if you wish, but do NOT close the terminal because this is where your computer is running MS-Align+. When the search is finished, the terminal window will automatically close, the LogBook will show the line “Workflow

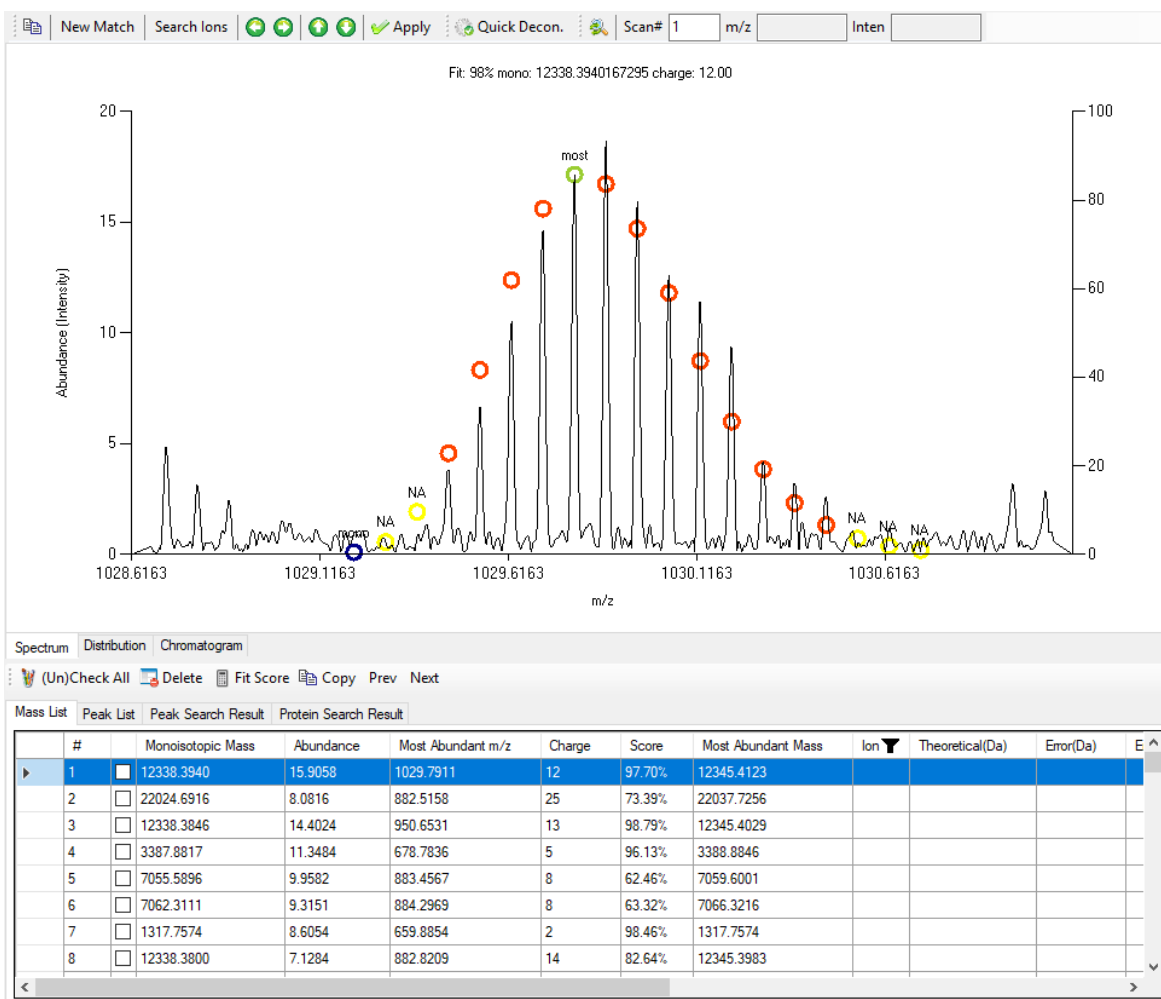
Finished”, as seen below, and the Workflow Manager will move the workflow into the “Completed Workflows” Tab.



Your search results will be automatically imported into MASH Explorer. Look under the Protein Search Results list to view your results. If you would like to manually import your search results, please read Step 3: Import Results from the Discovery Mode and import your search results file with the Targeted Mode Characterization button.

Step 4: Annotate Results

MASH Explorer allows you to view and annotate your imported results with the Mass List and Protein Search Results List, the Results View, and the Sequence Table. For more details on these functions, refer to the *User Manual*.



Your MSAlign+ search results contain the protein sequence of the identified protein, which will be automatically import. Go to the [Protein Search Result](#) list and click on the first result. The Sequence Table will display this sequence. The protein sequence may need to be modified to remove some of MSAlign+'s notation. To do this, please click on the Modify Sequence button under the Sequence Table.

The screenshot shows the 'Modify Sequence' dialog box. It has three main sections: 'Modifications', 'Fragment Ion Types', and 'Tolerance Information'. The 'Modifications' section includes a 'Position' dropdown set to 'Custom', a 'PTM' input field with '0', and buttons for 'Clear Selected Modification' and 'Clear All Modifications'. The 'Fragment Ion Types' section has checkboxes for 'a', 'b', 'c', 'x', 'y', and 'z', with 'a', 'b', and 'c' currently checked. Below these are buttons for 'Show Assignments' and 'Copy Table Image'. The 'Tolerance Information' section has an 'Error' input field set to '1.1', radio buttons for 'Da' (selected) and 'PPM', and a 'Modify Sequence' button.

In the Modify Sequence Dialog, remove the following characters so that the sequence matches the image below: "M.", "[48.68]", and final "." Click OK to display these notation changes in the Sequence Table.

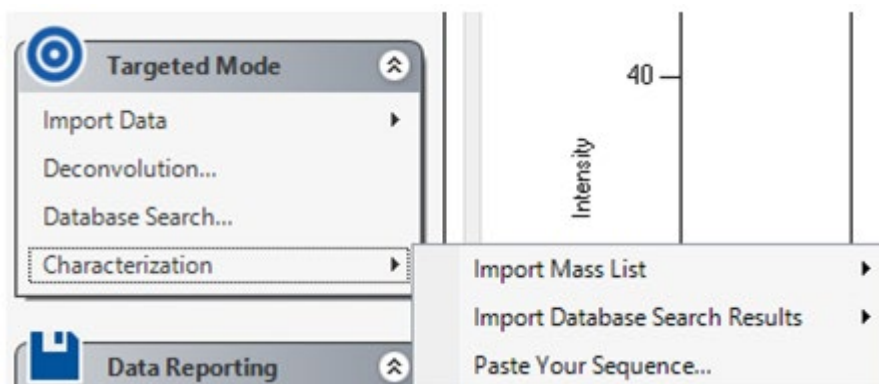
You can also manually type in the following sequence:

PFFDVQKRLGLDLDRWMTIQSAEQPHKIPGRCHAFEKEWIECAHGIGGIRAEKECKIEFDDFVECLLRQ
KTMKRLSAIKRQRDKLIKEGKYTPPPHHLGKEDPRP

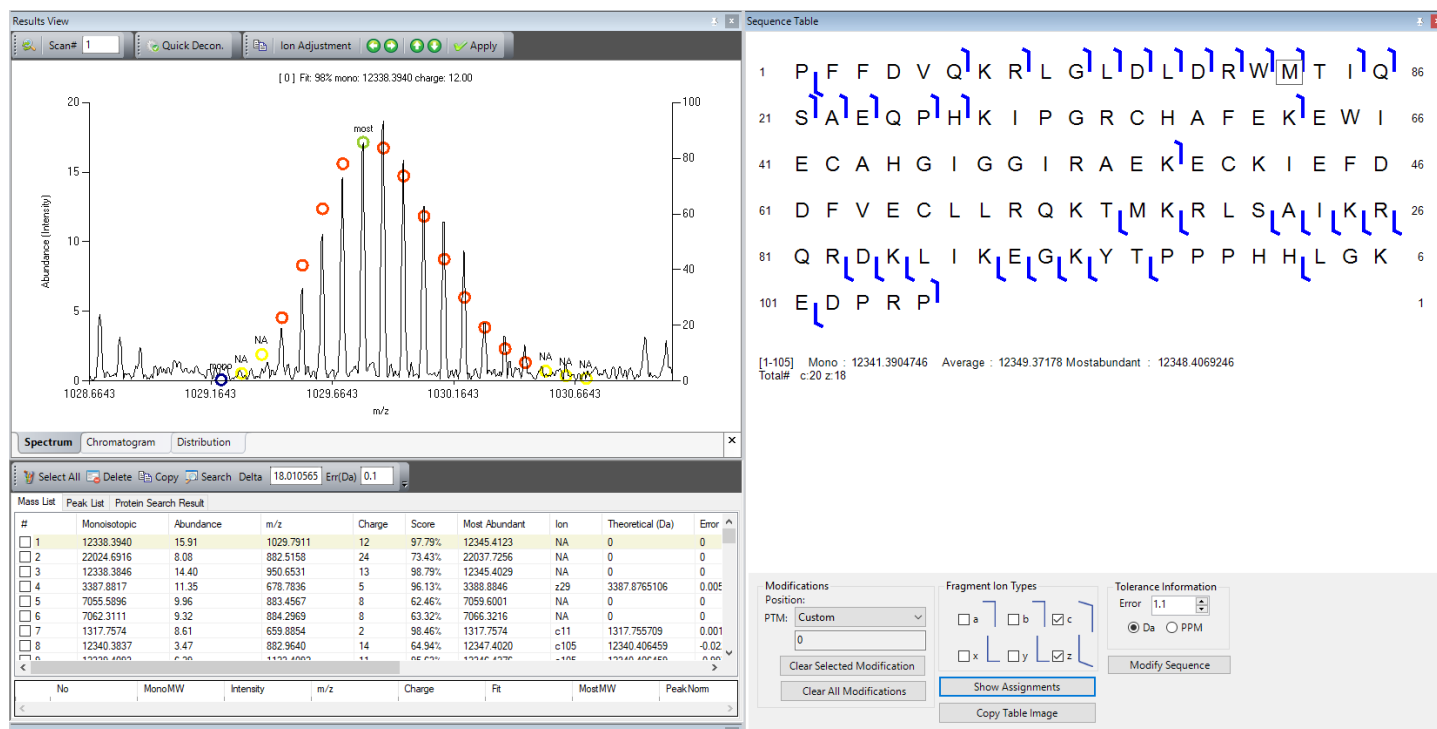
Position	Sequence	Length
1	P F F D V Q K R L G L D L D R W M T I Q	86
21	S A E Q P H K I P G R C H A F E K E W I	66
41	E C A H G I G G I R A E K E C K I E F D	46
61	D F V E C L L R Q K T M K R L S A I K R	26
81	Q R D K L I K E G K Y T P P P H H L G K	6
101	E D P R P	1

[1-105] Mono : 12341.3904746 Average : 12349.37178 Mostabundant : 12348.4069246 Total#

For manually pasting protein sequence, click on Paste Your Sequence and the Paste Sequence Information window will be opened. Paste the target sequence and click OK and the pasted sequence will be displayed in Sequence Table.

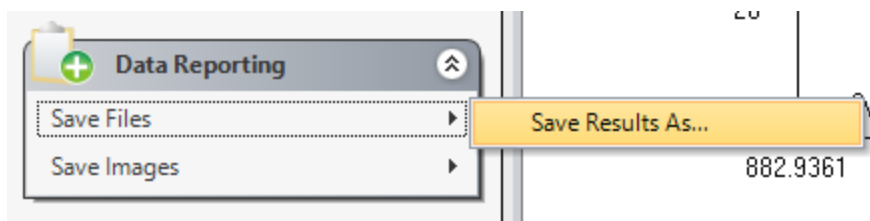


In the Sequence Table, check the c and z ion types and click on the Show Assignments button to view the fragmentation for that peptide sequence. Users can click on each entry in the Mass List to see the profile data.



Step 5: Save Results

Saving annotated results is only available for Targeted Mode. In the Data Reporting panel click Save Files and then Save Results As. Your Mass List and Protein Search Results list will be saved as an XML file. Save this XML file in the same directory as your raw data file.



To load a saved XML file into MASH Explorer, in the top left hand corner go to File and click Open Saved Results.

