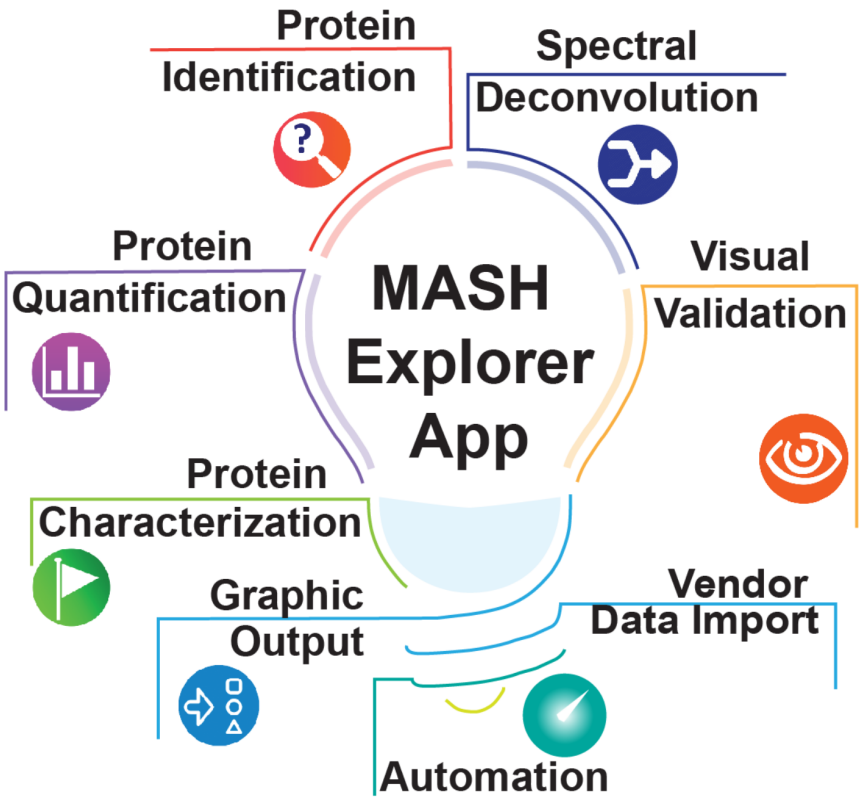
MASH Explorer A Comprehensive and User-Friendly Software Environment for Top-Down Proteomics User Manual

Version 2.2

Ying Ge Research Group - University of Wisconsin, Madison



Contents

1 Overview

MASH Explorer, developed by the Ying Ge research group, is a comprehensive, universal, user-friendly, and freely available software environment for top-down proteomics. MASH Explorer allows data import from various vendor data formats and accommodates several deconvolution and database search pipelines for spectral deconvolution and protein identification. MASH Explorer provides visualization tools for complex top-down mass spectra to validate deconvolution and protein identification results.

This tutorial provides step-by-step instructions for protein identification in discovery mode (large-scale protein identification), targeted mode (comprehensive protein characterization), and visual validation of results.

The development of MASH Explorer is supported by National Institute of Health R01 GM117058, with additional support from R01 GM125085, R01 HL109810, and R01 HL096971. Additional funding support is provided by the Wisconsin Alumni Research Foundation.

2 Interface Layout

The components of the MASH Explorer main window are described below (see Figure 1).

Workflow and Parameters: Users can follow the steps for LC-MS/MS and MS/MS workflow for data processing.

Results View: The mass spectra and chromatogram are shown in this panel.

Mass List: The fragment ion information is presented in this panel.

Experiment List: This panel shows a list of the experiments that are currently loaded into

a

Figure 1: Overview of MASH Explorer software Interface

[a]Note: The interface changed from docking panel to SplitView in version 2.

MASH Explorer, and allows you to switch which experiment is active by double-clicking on the desired experiment. (The active experiment is highlighted in blue.)

LogBook: This panel displays processing and error messages.[1]

Status Bar: The progress of data processing is shown in this panel.

Sequence Table: This panel presents the protein sequence and top-down fragment ion information.

[1]Note: File Import and LogBook are tabs in the same panel of the MASH Explorer App window.

3 Usage Reporting/Privacy

Starting in version V2.0.1, MASH Explorer optionally reports data about your usage of MASH Explorer to the MASH Explorer team. Please refer to the MASH Explorer Privacy Notice for information about what data may be reported and what is done with that data. (The privacy notice is also installed with the other MASH Explorer user documentation.) When you first run a new version of MASH Explorer (V2.0.1 or later) you will see the Version Changed dialog (Figure 2). This asks you to review your privacy settings.

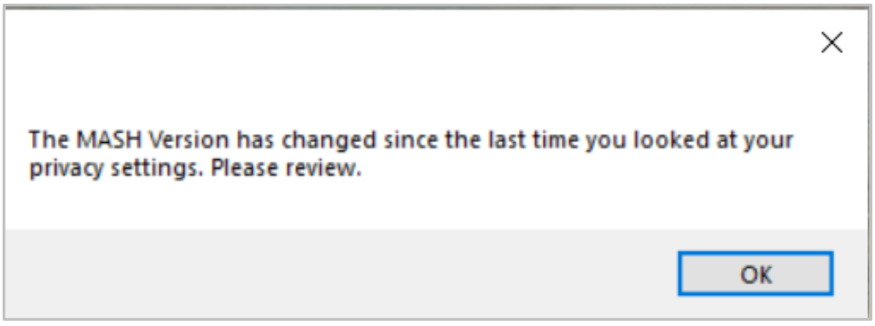


Figure 2: Version Changed Dialog

When you click "OK" in the Version Changed dialog, you will see the Privacy Dialog (Figure 3). This dialog allows you to set the level of data reported by MASH Explorer (note that this setting is separate for each individual MASH Explorer user, as long as you are logged onto a computer under different user names). You can choose reporting levels of "None", "Minimum", "Medium", "Maximum", and "Custom". All choices except "Custom" automatically set whether to report various types of data (except for exceptions). "Custom" allows to you choose whether to report each type of data.

After you have selected your privacy settings, you will see the Privacy Settings Confirmation dialog (Figure 4). This dialog reminds you that you can change your privacy settings at any time by selecting the Configuration option in the Tools menu (see Section

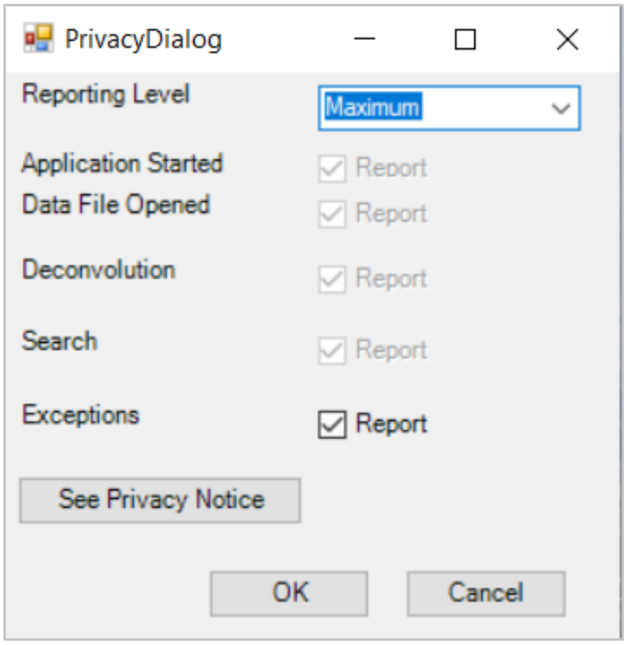


Figure 3: Privacy Settings Dialog

9.4).

Figure 4: Privacy Settings Confirmation Dialog

4 Discovery Mode: Protein Identification and Characterization from LC-MS/MS data

Our team designed the workflow termed "Discovery Mode" to handle batch LC-MS/MS analysis. This workflow includes data import, data processing including deconvolution and database search, and finally data validation for protein identification.

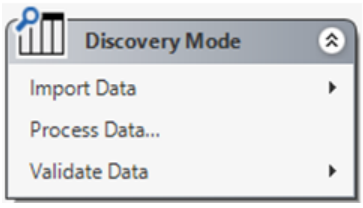


Figure 5: Discovery Mode Button Layout

4.1 Import Data

In the "Discovery Mode" section under the *Workflow and Parameters* panel, Import Data will direct users to load experimental data to the MASH Explorer software. Users can follow the steps below for data import:

1. Click on Import Data in the "Discovery Mode" tab under the *Workflow and Parameters* panel to add a data set (Figure 6).[2]

2. Click on the option that represents the raw data file format that will be uploaded (Figure 6). The current version of MASH Explorer software supports Thermo (.raw), Bruker (.baf), Bruker data output in ASCII format (.ascii), Waters (.raw), and universal file formats (.mzXML, .MGF, and .mzML).

[2]The user can also choose to drag and drop the data file to import data into the software.

3. Select the LC-MS/MS data files to be analyzed. The *Status Bar* will update the data importing progress (Figure 7).

4. The current version of MASH Explorer allows users to have multiple files open at the same time. Users can open the data files one after another.

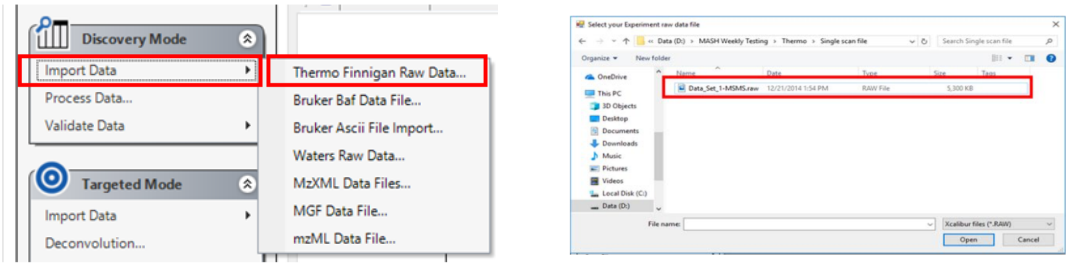


Figure 6: Import Data function and file selection

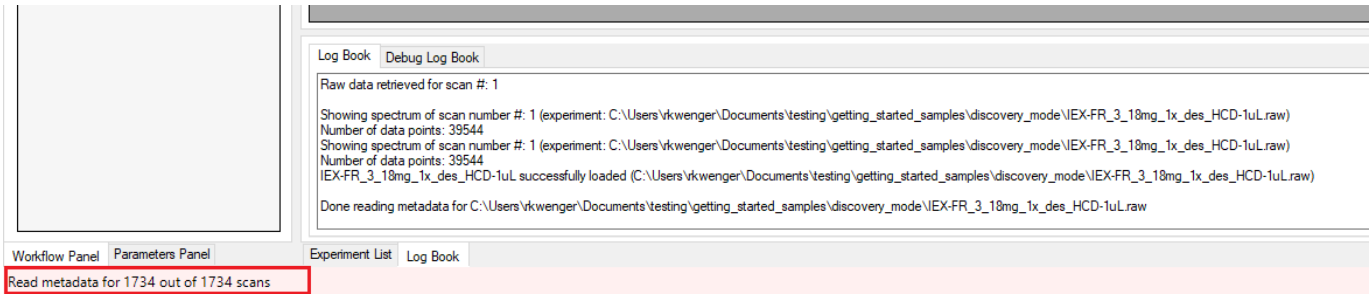


Figure 7: *Status Bar* panel updates for data import

Users can view the chromatogram of the imported data and navigate different scans. The chromatogram can be visualized in the "Chromatogram" tab in the *Results View* panel (Figure 8).

Users can navigate to different scans in three ways:

Figure 8: Viewing chromatogram and individual scan

- By scrolling the scan list and selecting the desired scan or scans.
- By double-clicking in the chromatogram view - this will select the scan closest to the double-click location (you probably want to zoom in first). (Note that the scan list will be automatically scrolled to show the selected scan.)
- By entering a scan number in the Scan# text box (or typing the up or down arrow key with input focus in the Scan# text box).

(In the first two cases, you need to select the "Spectrum" tab to see the relevant spectrum; in the third case, the spectrum will automatically be shown if you are viewing the chromatogram.) (See Figure 8)

The scan list contains all the scans including MS1 and MS2 (MS/MS). Each MS2 scan has the precursor m/z. Retrieval of charge state and activation of the precursor ions will be dependent on the raw data file format (Figure 6). Raw data from Thermo will likely contain information regarding the charge state and activation of the precursor ions. However, Bruker raw data does not provide this information. The user can manually specify the MS level of a scan by double-clicking on that scan. Doing this will show a

dialog in which the MS level can be entered.

When the mouse cursor hovers over either a mass spectrum or chromatogram, the mouse location will be shown in the relevant data units (m/z and intensity for a mass spectrum and retention time (RT) and intensity for a chromatogram) (Figure 8, Purple box).

4.2 Process Data

Data processing in top-down proteomics mainly involves centroiding, deconvolution, and database search. In the Discovery Mode, these data processing tasks are bundled together in the Process Wizard. Users can click on Process Data and the Process Wizard will be opened (Figure 9, left).[3] Select a deconvolution method, a database search method, and upload a targeted database file in .FASTA format (Figure 9, left). Users can click on the Advanced tab to change general settings and algorithm parameters for each algorithm (Figure 9, right). Specific parameter descriptions for each algorithm can be found on the respective algorithm websites. Most algorithms have options for users to change the parameters such as the maximum and minimum values for fragment ion charge and mass. Additionally, users have the option to input fixed and variable post-translational modifications. For search optimization, some algorithms such as TopPIC allow users to input the number of threads the algorithm can use. These parameters are accessible under the Advanced → Search tab. Click on the Start button to begin deconvolution and database search. MASH Explorer will attempt to detect the activation type from the data file. If the activation type is not detected, CID will be executed by default.

If you change any of the parameters, you can click Save Normal to make the new

[3]Please refer to the Search Algorithm Setup Guide included in the MASH Explorer software installation regarding the installation of each deconvolution and search software.

values the defaults for future workflows. Reset to Defaults will revert all the changed parameters to factory defaults.

For more information about running workflows, see section 6 below.

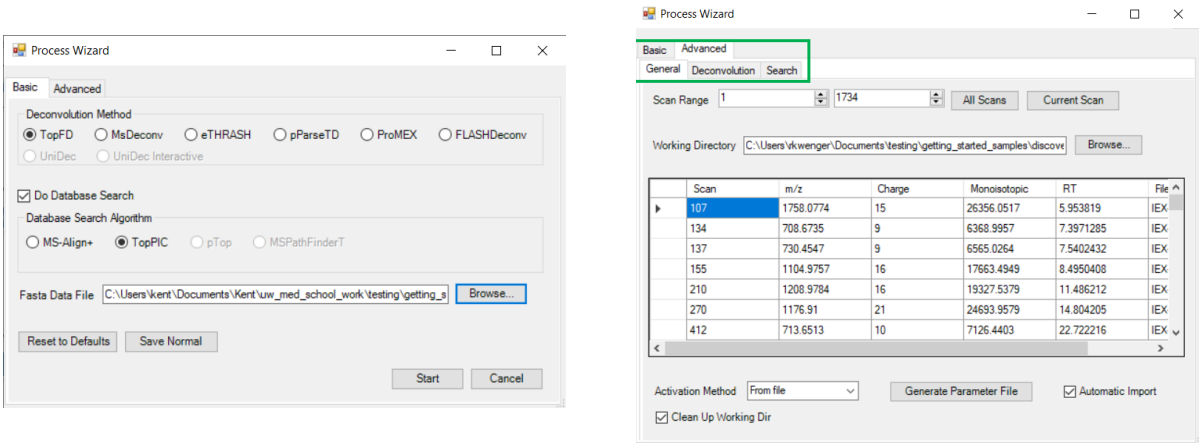


Figure 9: Basic and Advanced settings for Process Wizard in Discovery Mode

4.3 Validate Data

Workflow results can be imported by using *Validate Data* → *Import Mass List* to import deconvoluted mass list and *Validate Data* → *Import Database Search* to import protein search results (Figure 10). Users can import the protein search results. By clicking on entries in the Protein Search Results, the corresponding protein sequence will display in the *Sequence Table* panel (Figure 11). A detailed description of functions in the *Sequence Table* such as adding modifications and displaying fragment ions can be found in Section 7.2 Sequence Table of the user manual. Users can also import the mass list which contained fragment ion information which will display in the *Mass List* panel (Figure 11). A detailed description of functions in the *Mass List* such as correction of fragment ion charge state

and monoisotopic mass is provided in Section 7.1 Mass List and Results View of the user manual.

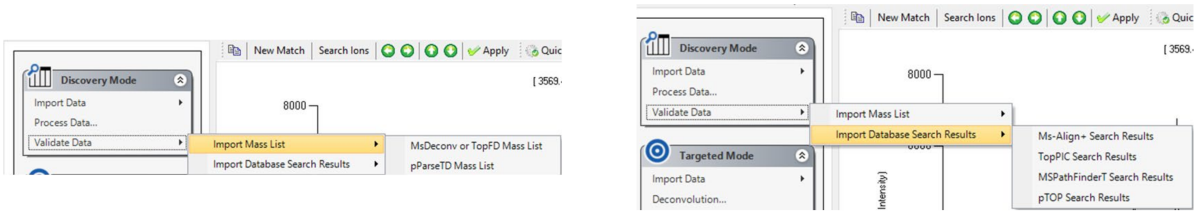


Figure 10: Menus for Validate Data

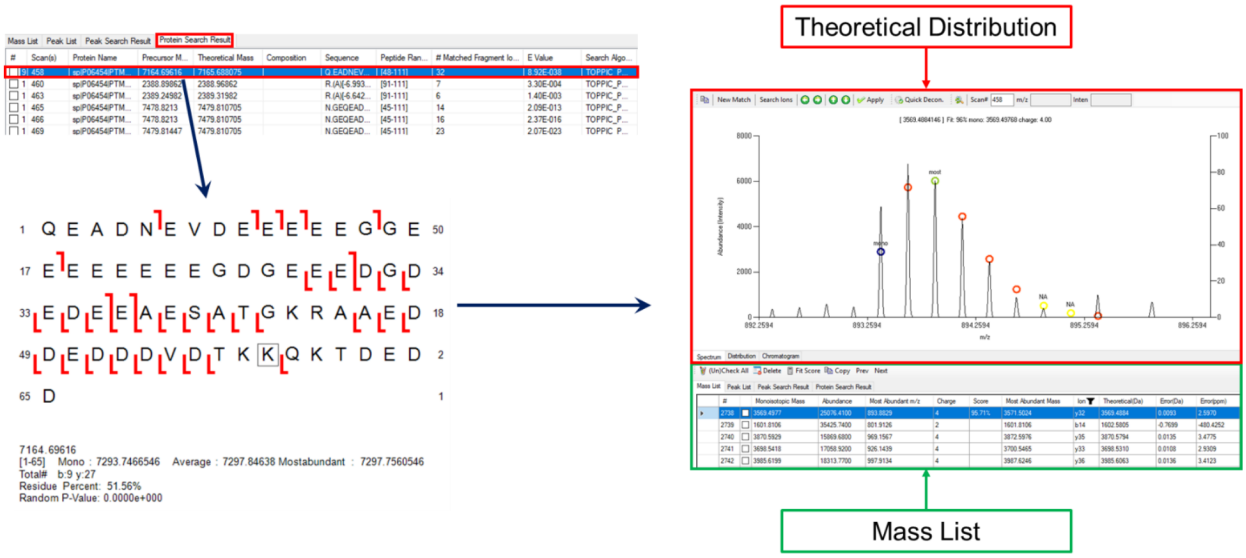


Figure 11: Analysis of protein sequence and fragment ion

Automatic Import

For users' convenience, the mass list and protein sequence after database search is automatically imported. However, when users run batch analysis, the automatic import function will stack one after another. Therefore, it is recommended to turn off the automatic

import function when running batch LC/MS-MS analysis. To disable automatic import, users can uncheck the automatic import checkbox in the Advanced/General tab (shown in the right half of Figure 9, checkbox at the bottom right)

5 Targeted Mode: Protein Identification and Characterization from MS/MS data

Contrary to "Discovery Mode", our team developed "Targeted Mode" workflow for analysis including comprehensive protein characterization. This workflow includes data import, spectral deconvolution to identification of fragment ions, database search based on identified ions, and finally protein characterization. This workflow aims to perform identification of fragmentation ions that describe the protein sequence and identification and localization of post-translational modifications of a target protein sequence.

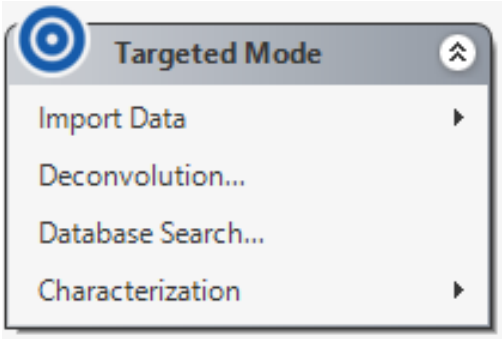


Figure 12: Targeted Mode button layout

5.1 Import Data

Similar to the Import Data in "Discovery Mode", Import Data in "Targeted Mode" will direct users to import data into the software. Users can follow these steps for data import:

1. Click on Import Data in the "Targeted Mode" tab under the Workflow and Parameters panel to add a data set (Figure 13).[4]

2. Click on the option that represents the raw data file format to be uploaded (Figure 13). The current version of MASH Explorer software supports Thermo (.raw), Bruker (.baf), Bruker data output in ASCII format (.ascii), and universal file formats (.mzXML, .MGF, and .mzML).

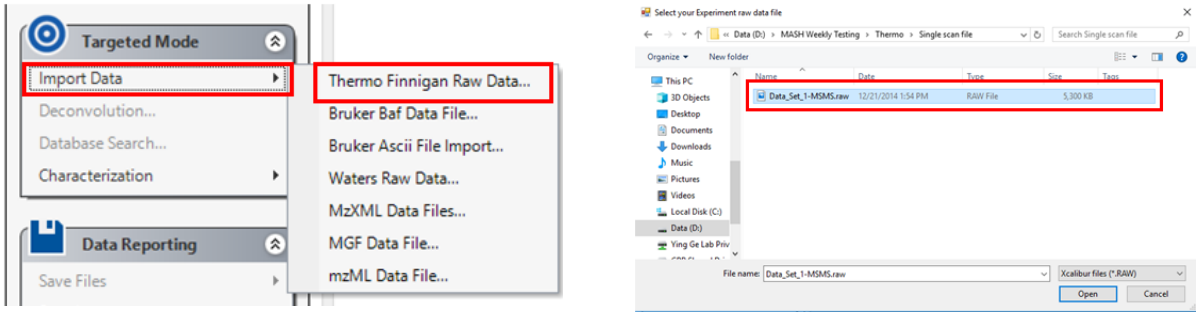


Figure 13: Data Import for Targeted Mode

Experimental data acquired from different instruments should be handled differently. Broadly speaking, there are two types of instruments: those that perform spectral averaging/summing during acquisition and those that perform averaging/summing post-acquisition.

1. Examples for the first category include FT-ICR type instruments including Thermo Finnigan Ultra and Bruker SolariX. The raw data normally has one single scan and can be processed directly by MASH Explorer software.

[4]The user can also choose to drag and drop the data file to import into the software.

2. Examples for the second category include instruments that perform liquid chromatography (LC)-(MS/MS). For example, users inject a solution containing a single protein, and use the LC to deliver the solution to the mass spectrometer. The resulting raw data has multiple scans while each scan contains fragment ion of the same protein. This type of data requires further processing prior to data import to MASH Explorer software.

   (a) Thermo instruments: Xcalibur software can perform spectral averaging natively and the resulting spectrum can be exported as RAW format.

   (b) Bruker instruments: DataAnalysis can perform spectral averaging natively and the resulting spectrum can be output as ASCII format.

   (c) Other instruments and general method: ProteoWizard has a filter function "Scan Summing" (Figure 14) for spectral averaging, and the summed spectrum can be exported in mzXML format. Ideally, the output mzXML file has only one scan.

5.2 Deconvolution

Spectral deconvolution is the process of identifying fragment ion peaks and calculating their charge states, monoisotopic mass, most abundant mass, intensities, and other parameters. Our software currently offers five tools for spectral deconvolution in Targeted Mode: enhanced-THRASH (eTHRASH), TopFD, MS-Deconv, FLASHDeconv, and UniDec. Each of these algorithms can be automatically run by the MASH Explorer software using the Process Wizard (Figure 15). Starting the deconvolution processes will initiate the Workflow Manager. When the calculation is finished, the calculated mass

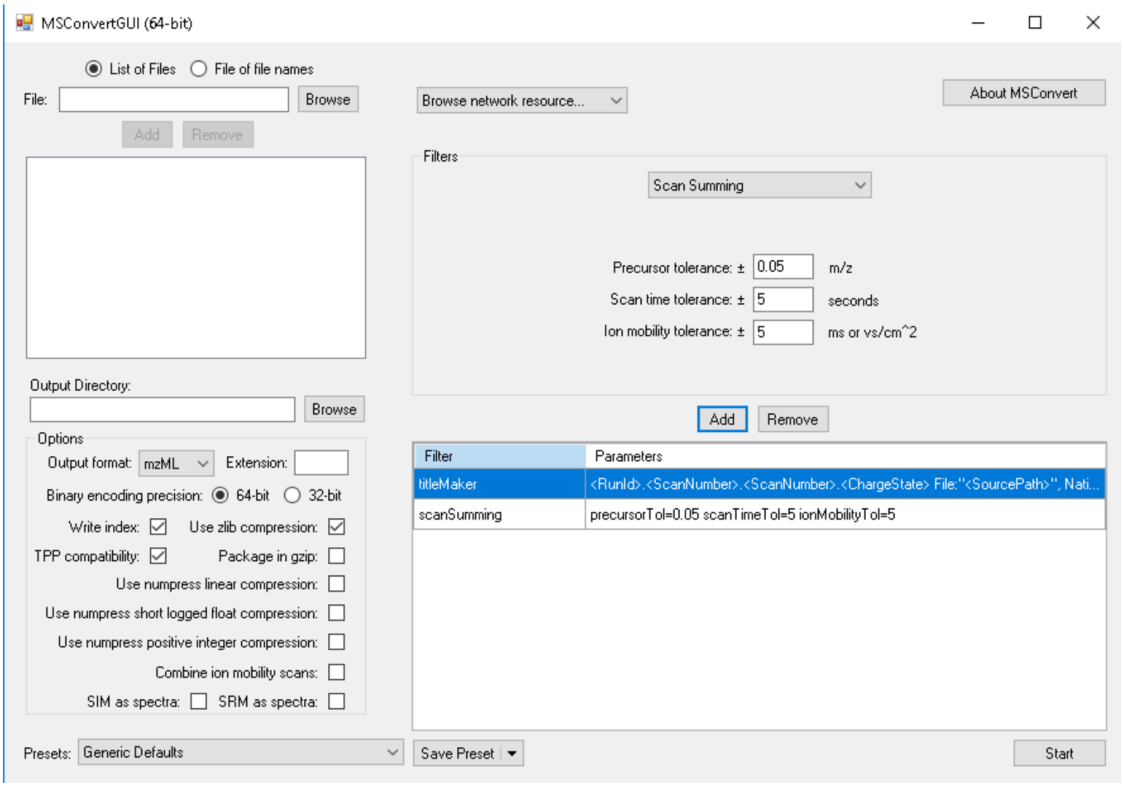


Figure 14: Scan Summing

list from each algorithm will be automatically imported and displayed in the Mass List panel (Figure 15). Users can click on each fragment ion entry for viewing. The detailed functions of the Mass List are described in Section 7.1 *Mass List* and Results View of the user manual.

Notes on spectral deconvolution

1. The current version of MASH Explorer supports TopFD, MS-Deconv, eTHRASH, FLASHDeconv and UniDec for spectral deconvolution in Targeted Mode.

2. eTHRASH, TopFD and MS-Deconv support MS1 deconvolution. MS-Deconv requires users to input the correct charge state of the precursor ion. You can input the

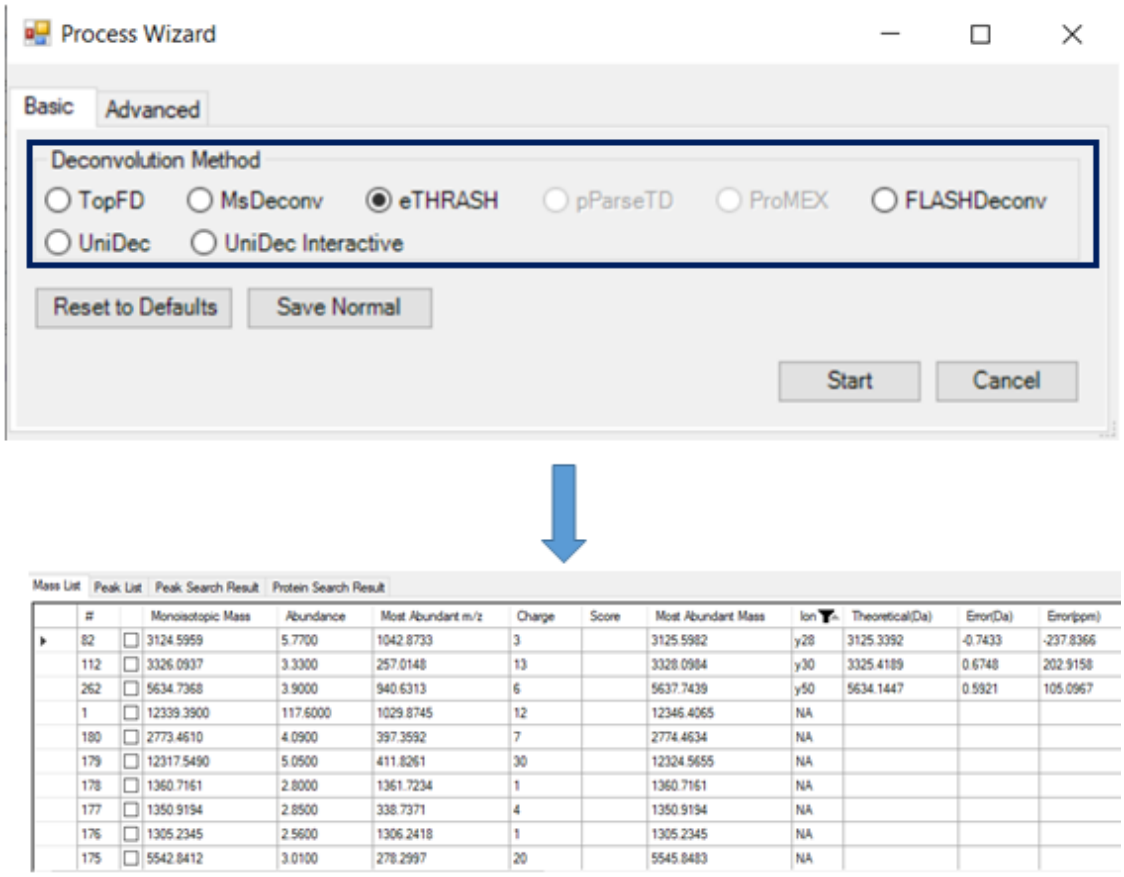


Figure 15: Basic interface of deconvolution-mode Process Wizard and populated Mass List

ion charge state as shown in Figure 16.

3. UniDec can be run in interactive or non-interactive mode. In non-interactive mode, UniDec works like the other deconvolution tools – you start the workflow and MASH Explorer runs UniDec automatically. In interactive mode, MASH Explorer launches UniDec with the relevant data already loaded. You can then manually process the data in UniDec, and when you exit UniDec, the processed data is loaded back into MASH Explorer.

4. The *Quick Decon.* button in the *Results View* can be used as a quick tool for spectral deconvolution. It supports both TopFD and eTHRASH, and the parameters can be configured under *Tools* → *Quick Decon. Parameters*. Details can be found in Section 9.4.

5. The current version of MASH Explorer cannot process MS/MS scans without an MS2 designation. Users can manually edit the MS level of a scan by double-clicking on that scan in the chromatogram view.

For more information about running workflows, see section 6 below.

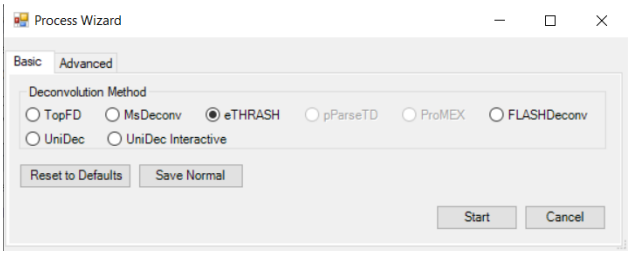


Figure 16: Process Wizard for Deconvolution

5.3 Database Search

Our workflow has incorporated algorithms that use fragment ion (i.e. MS2) information for protein identification. When the fragment ion selection is completed, the fragment ions can be searched against the database for protein identification. Users can choose one of the supported database search algorithms including MS-Align+ and TopPIC, and upload the target database (Figure 17). Click "Start" to begin the database search process and similar to other processes, the Workflow Manager will be invoked to handle algorithm execution.

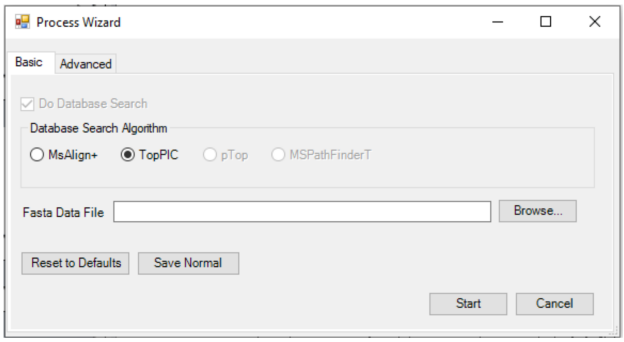


Figure 17: Basic Interface of Process Wizard in Search mode

For the Targeted Mode, it is recommended for the users to provide as much information as possible, as extraction of information from the MS-MS files is harder compared to the LC/MS-MS data. In the Advanced settings, users can enter the charge state of the precursor and the fragmentation activation method (Figure 18). This information will help substantially in file parsing processes prior to database search.

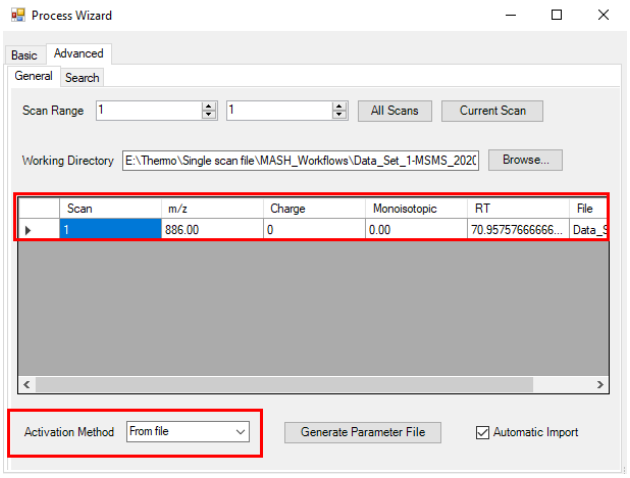


Figure 18: Advanced settings for search-mode Process Wizard

Notes on database search

1. The current version of MASH Explorer supports MS-Align+ and TopPIC for database search in Targeted Mode.

2. It is recommended that users input the charge state of the precursor ion and the activation method of the file in the specified locations in the Process Wizard shown in Figure 18. This information will help substantially in successful file parsing processes.

For more information about running workflows, see section 6 below.

5.4 Characterization

The Characterization menu includes *Import Database Search Results* that imports the sequence from the database search and *Paste Your Sequence* which allows users to copy the sequence to the *Sequence Table* (Figure 19). A detailed description of the Sequence Table panel is in Section 7.2. For database search list import, please refer to Section 4.3.

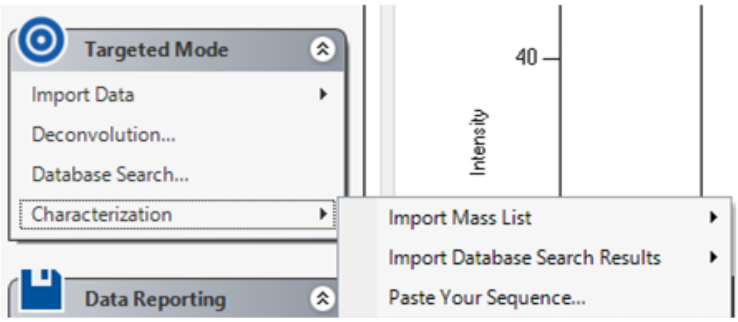


Figure 19: Menu for Characterization

For manually pasting a protein sequence, click on *Paste Your Sequence* and the *Paste Sequence Information* window will be opened. Paste the sequence and click OK and the

pasted sequence will be displayed in the *Sequence Table* (Figure 20).

Figure 20: Pasting a protein sequence

6 Workflows

Note: the information in this section applies to both Discovery Mode and Targeted Mode workflows.

Each time you run a deconvolution and/or search workflow, MASH Explorer will generate a workflow-specific folder containing all files related to that workflow (Figure 21). The workflow-specific folder will be within the MASH_Workflows folder (which is in the folder containing the experimental data file). *Do not open or delete any files in this folder before the workflow has completed.* The processing folder will include the name of the data file, deconvolution algorithm, database search algorithm, and the time the workflow started.

To save disk space, as of version 2.1, temporary files in the workflow folder (such as the copy of the experiment file) are removed when a workflow finishes successfully.

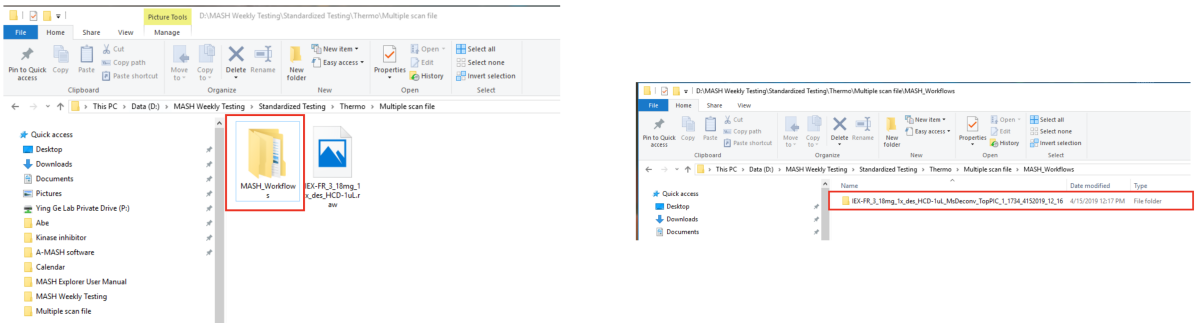


Figure 21: Folder created for data processing

(If the workflow does not finish successfully, the temporary files are left in place to aid debugging.) The workflow folder cleanup can be disabled by unchecking the "Clean Up Working Dir" checkbox in the Advanced/General tab of the Process Wizard (see Figure 9).

When the deconvolution/search workflow starts, the Workflow Manager will be opened (Figure 22).

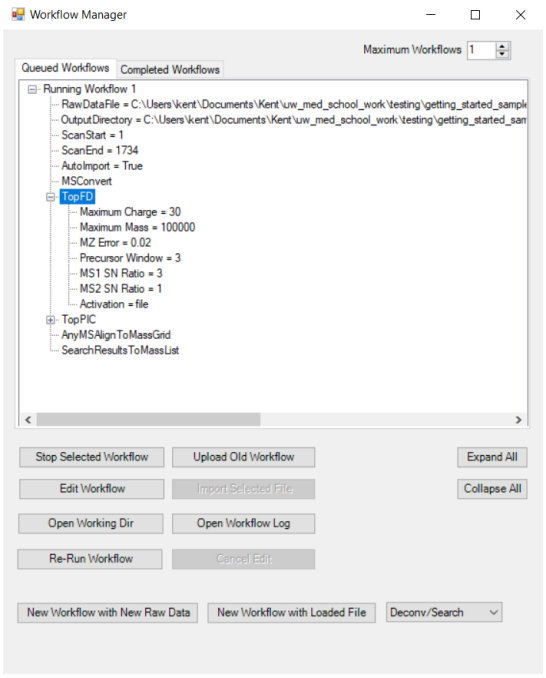


Figure 22: Interface for Workflow Manager (workflow running)

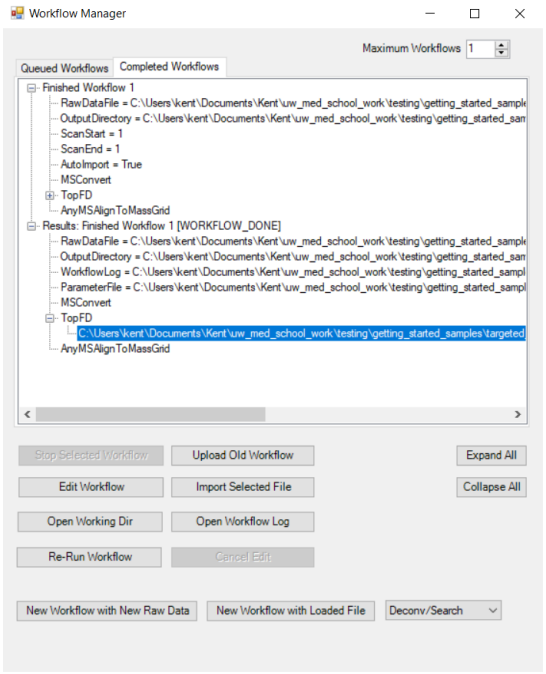


Figure 23: Interface for Workflow Manager (workflow completed)

Note that not all Workflow Manager functionality is available at all times, depending on which state various workflows are in, and which, if any, workflow is selected. The buttons for unavailable functions are now disabled.

Available at any time:

- *Upload Old Workflow* will show a dialog allowing you to select the working directory of a previous workflow (folders such as those shown in Figure 21). You can also upload an old workflow by dragging and dropping the parameter file (`Workflow_Parameters.txt`) on the MASH Explorer main window (not on the Workflow Manager).

- *New Workflow with New Raw Data* will create a workflow with experimental data that is not currently open in MASH Explorer.

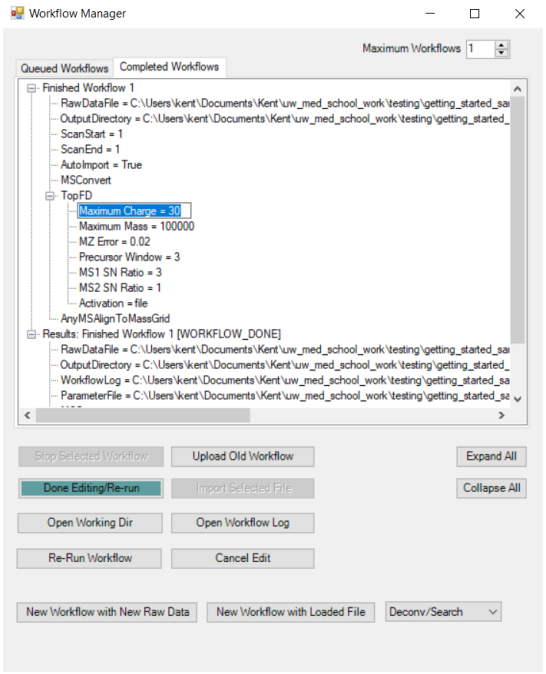


Figure 24: Interface for Workflow Manager (editing workflow)

- *New Workflow with Loaded File* will create a new workflow with experimental data that is currently open in MASH Explorer.

Available if any workflow is selected:

- *Open Working Dir* will open the working directory of the selected workflow in File Explorer.
- *Open Workflow Log* will open the workflow log of the selected workflow.
- *Re-Run Workflow* will re-run the selected workflow without any changes (most useful for re-running workflows that you have uploaded).

Available if a queued or completed workflow is selected:

- *Edit Workflow* allows users to modify parameters for pending and completed workflows (editing a workflow leaves the original workflow unchanged and creates a new workflow with the modified parameters).
  - Click on the *Edit Workflow* button to enable editing of a workflow. Users can edit both pending and completed (but not running) workflows.
  - Double-click on a parameter to change the value (you can change multiple parameters sequentially).
  - Click *Done Editing/Re-run* to queue a new workflow with updated parameters.
  - Click *Cancel Edit* if you have made changes but then deicde you do not want to create and run a new workflow.

Available if a queued or running workflow is selected (see Figure 22):

- *Stop Selected Workflow* will delete the pending task in the queue (stopping it if it's currently running).

Available if a results file from a completed workflow is selected (see Figure 23):

- *Import Selected File* will import the selected deconvolution or database search results file.

Available if a workflow is being edited (see Figure 24):

- *Done Editing/Re-Run* queues a new workflow with the edited parameters.
- *Cancel Edit* exits edit mode without creating a new workflow (discards any edits).

Workflow progress can be monitored in both the *Status Bar* and *LogBook*. Users can click on the *Completed Workflows* tab to view a finished workflow. If the Workflow Manager is closed, it can be reopened through Tools → Workflow Manager.

Previously run deconvolution and database search results can be imported for analysis. Using the *Upload Old Workflow* button, select the folder containing the relevant workflow (or drag and drop the `Workflow_Parameters.txt` file on the MASH Explorer main window). The uploaded workflow will appear in the *Completed Workflows* tab. After expanding the deconvolution and database search nodes, the deconvolution and database search result files can be located. You can import these files by selecting the file and clicking *Import Selected File* (Figure 25).

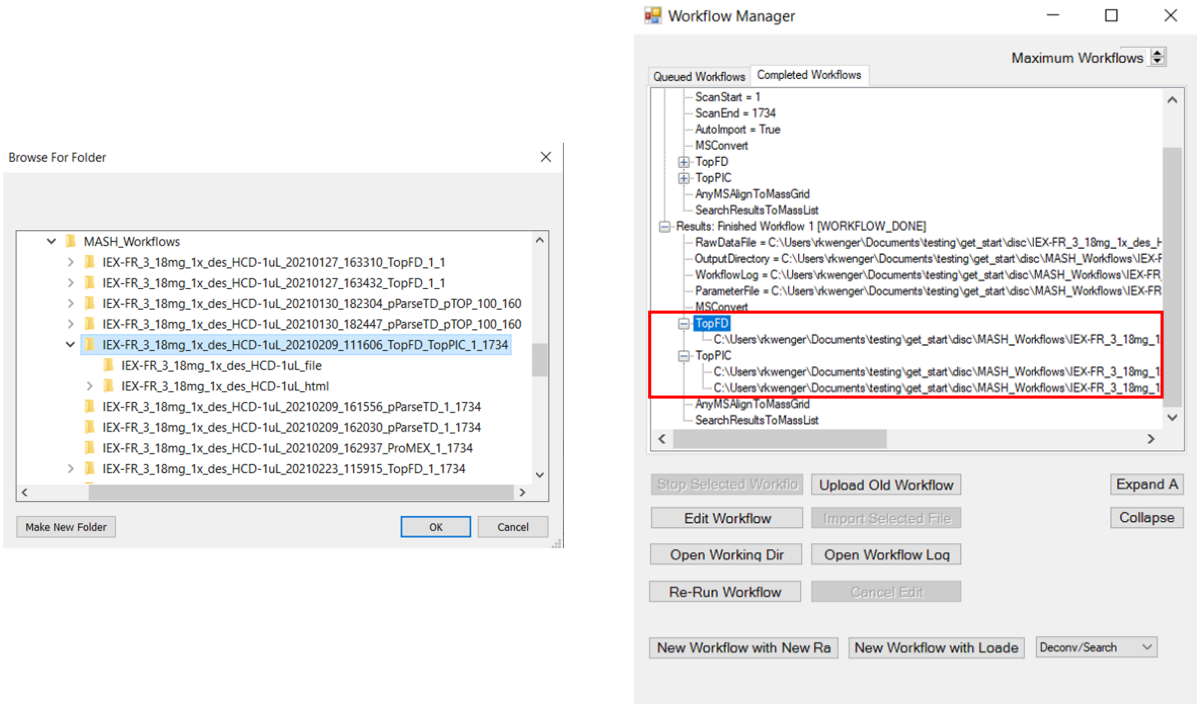


Figure 25: Uploading completed workflows and importing data

7 Data Reporting

This section of the manual introduces the functions in *Data Reporting* including *Save Files* and *Save Images*. These functions allow users to save the verified fragment ion list, enabling users to perform further processing of spectrum and sequence table images using other professional graphic software.

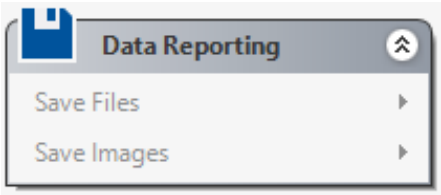


Figure 26: Data Reporting button layout

7.1 Save Files

Users can save their mass list and sequence table information in the xml format which is associated with the original raw data (Figure 27). The xml file is dependent on its original raw data. It is recommended that the users place the saved xml file in the same folder as the raw data.

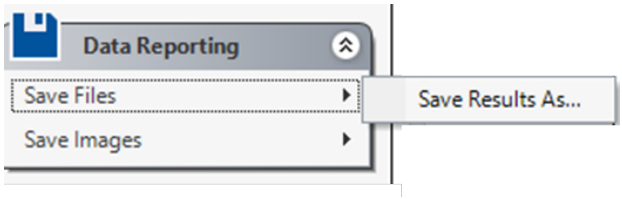


Figure 27: Save Files

To re-open the saved data, go to *File* → *Open Saved Results*. Since the xml file needs its raw data for reference, the users will be prompted to locate the raw data if the raw data is not found within the same folder.

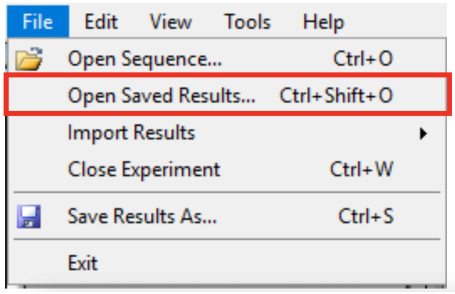


Figure 28: Open Saved Results

A saved results file can also be opened by dragging and dropping it onto the main MASH Explorer window.

7.2 Save Images

A spectrum displayed in the *Results View* panel and sequence information shown in *Sequence Table* panel can be exported as a Microsoft Office object (Figure 29). This function enables users to further edit the images including changing the font, font size of the text and line width and color of the fragment symbols. The copied image can be transferred to Adobe Photoshop, Illustrator, or Microsoft PowerPoint for further processing. Elements of the graph can be ungrouped for further modifications.

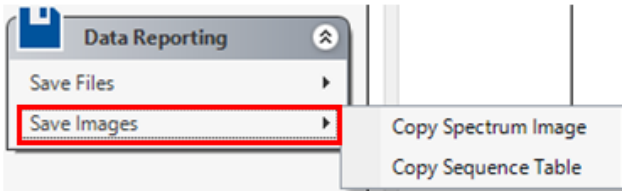


Figure 29: Menu to Save Images

In *Save Images, Copy Spectrum Image* will copy the spectrum in the *Results View* panel (Figure 30).

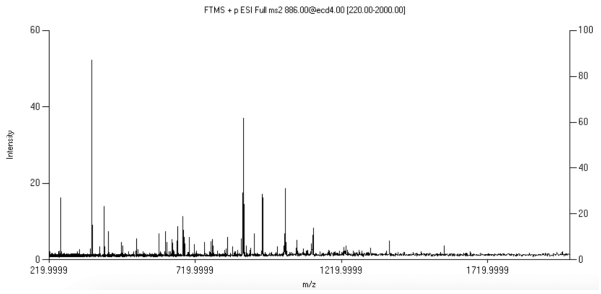


Figure 30: Spectrum Image

In *Save Images, Copy Sequence Table* will copy the spectrum in the *Sequence Table* panel (Figure 31).

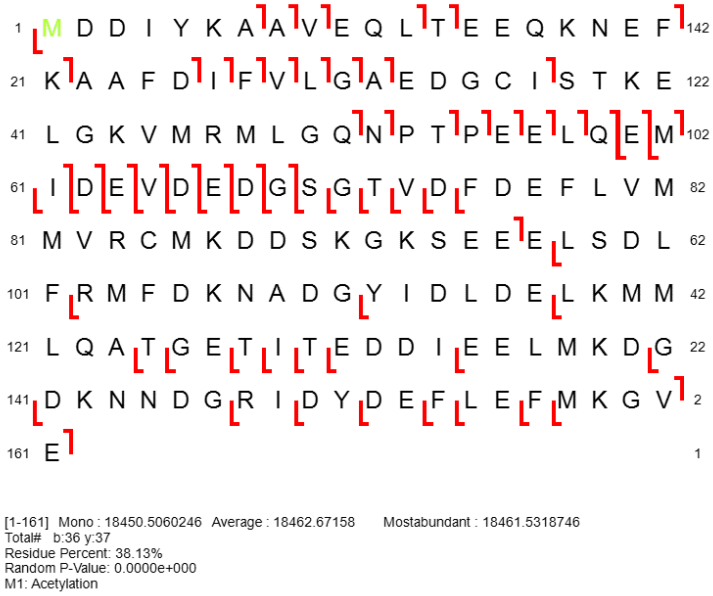


Figure 31: Sequence Table Image

8 Visual Validation of the Computational Output

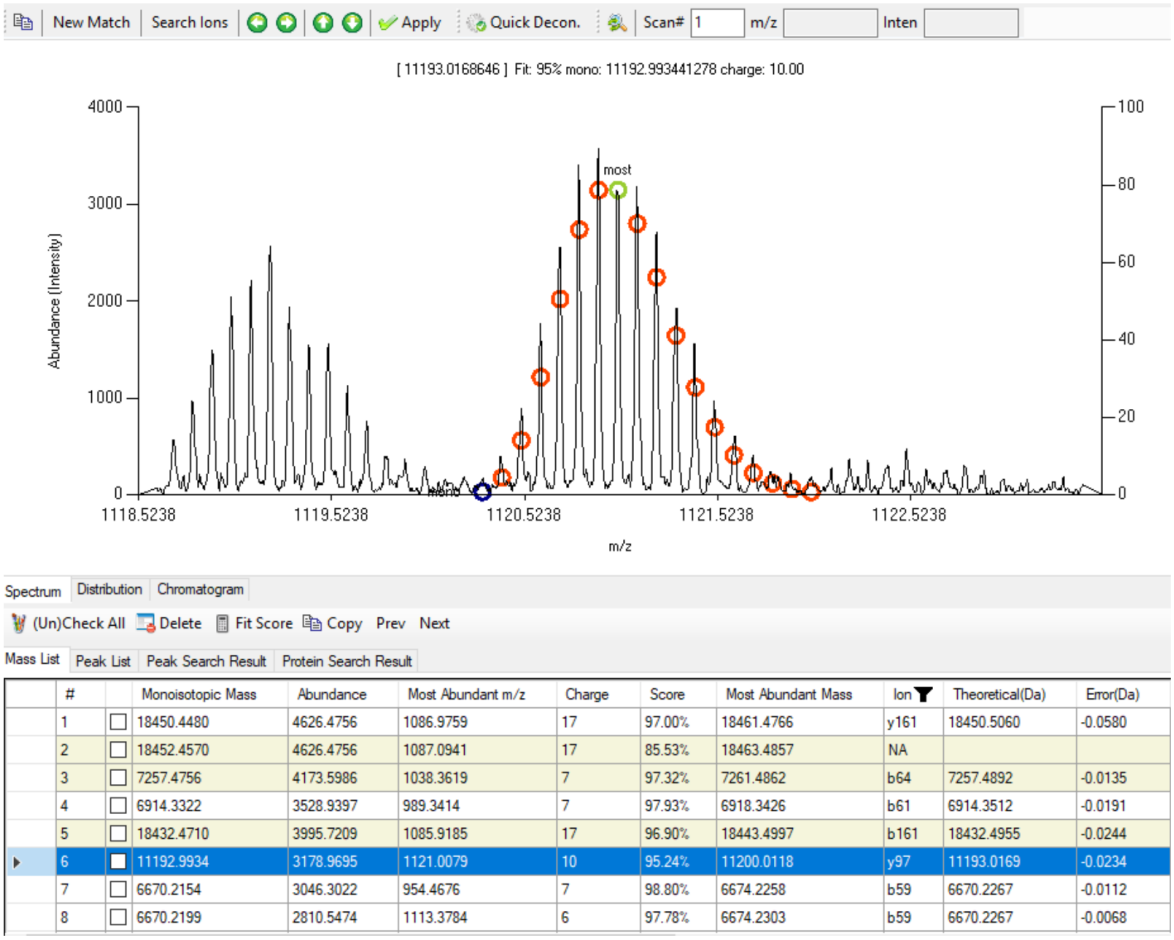
This section of the manual will introduce the function in *Mass List* and *Results View*, *Sequence Table* and *Ion Finder*. These tools are essential for the protein characterization using MASH Explorer software.

8.1 Mass List and Results View

The *Mass List* provides an intuitive display combining the theoretical ion distribution and the actual experimental spectra shown in *Results View*. After deconvolution, users can click on each entry in the mass list and the spectrum will be zoomed in to the location of the targeted ion. The red dots represent the theoretical ion distribution. Previously examined ions will have a light-yellow background in the Mass List. The up/down keys will change selected ion in the *Mass List*.

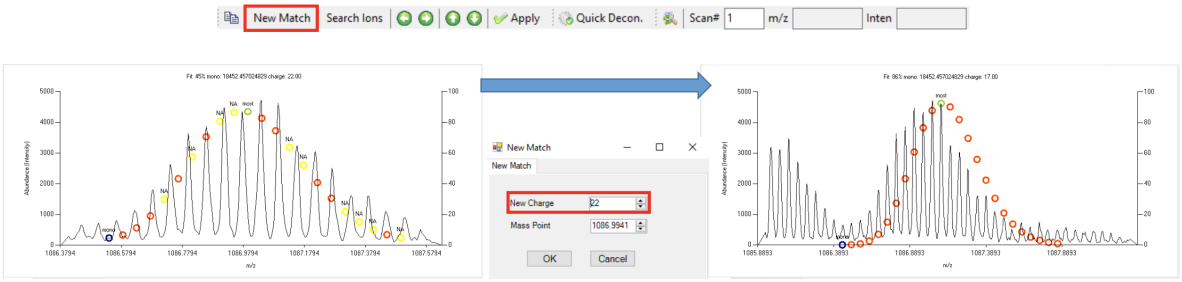
Users can change the following properties of the ion distribution to achieve better theoretical ion fitting to the experimental data:

1. Adjust the charge state of the theoretical distribution



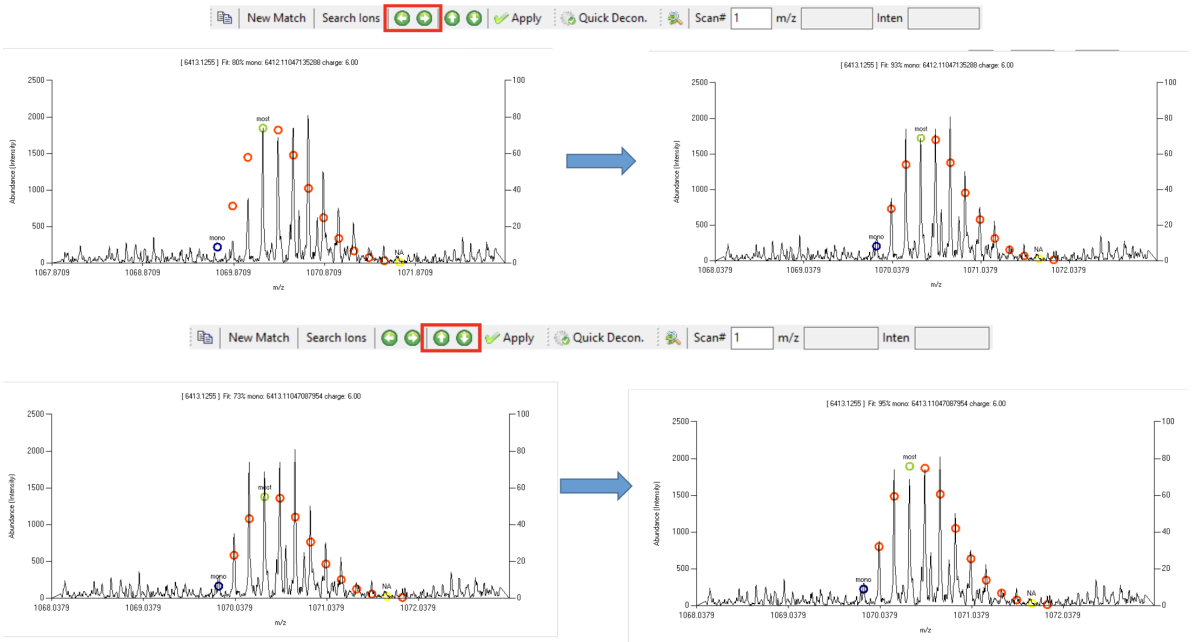
#		Monoisotopic Mass	Abundance	Most Abundant m/z	Charge	Score	Most Abundant Mass	Ion	Theoretical(Da)	Error(Da)
1	☐	18450.4480	4626.4756	1086.9759	17	97.00%	18461.4766	y161	18450.5060	-0.0580
2	☐	18452.4570	4626.4756	1087.0941	17	85.53%	18463.4857	NA		
3	☐	7257.4756	4173.5986	1038.3619	7	97.32%	7261.4862	b64	7257.4892	-0.0135
4	☐	6914.3322	3528.9397	989.3414	7	97.93%	6918.3426	b61	6914.3512	-0.0191
5	☐	18432.4710	3995.7209	1085.9185	17	96.90%	18443.4997	b161	18432.4955	-0.0244
6	☐	11192.9934	3178.9695	1121.0079	10	95.24%	11200.0118	y97	11193.0169	-0.0234
7	☐	6670.2154	3046.3022	954.4676	7	98.80%	6674.2258	b59	6670.2267	-0.0112
8	☐	6670.2199	2810.5474	1113.3784	6	97.78%	6674.2303	b59	6670.2267	-0.0068

Figure 32: Result View and Mass List interactive display



2. Add or subtract one isotopic unit of the theoretical ion distribution

3. Increase or decrease the signal intensities



Clicking *Apply* will save the fragment ion adjustments.

The mass list also allows user to sort the items in different orders, such as by monoisotopic mass, abundance, m/z, charge, score, most abundant m/z, ion designation, and so on (Figure 33).

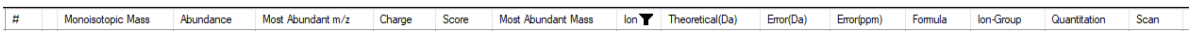


Figure 33: *Mass List* column items

The tool bar in the *Mass List* panel includes buttons such as (Un)Check All, Delete, Fit Score, Copy, Prev, and Next (Figure 34). These buttons interact with the checked item(s) in the *Mass List*.

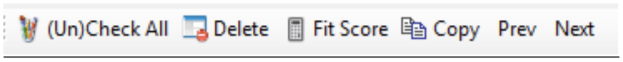


Figure 34: *Mass List* Tool Bar

- The *(Un)Check All* button will check or uncheck all the boxes in the *Mass List*.

- The *Delete* button will delete all items that are checked.
- The *Fit Score* button will calculate all the Score column in the mass list. The Score column describes the fitting between the experimental and theoretical isotopic distribution
- The *Copy* button will copy all items that are checked. The copied content can be pasted in processing software such as Notepad and Microsoft Excel.
- *Prev* and *Next* will move the previous or next scan for which *Mass List* entries exit.

8.2 Sequence Table

Once the mass list is verified by the users, the mass list can be matched with the theoretical fragment ions list generated by the user imported sequence to show fragmentation mapping of proteins (Figure 35). The sequence can be imported via *Characterization* → *Paste Your Sequence* in Targeted Mode, or the *Modify Sequence* button in the *Sequence Table* panel (Figure 35).

In the Modifications section, users can perform several tasks. To add a modification, the user needs to click on a specific amino acid. Table 1 shows the color of different instances.

	Not Selected	Selected
w/o PTM	Black	Red
w/ PTM	Green	Blue

Table 1: Color code for amino acid displayed in the *Sequence Table*

There are preset modifications with mass changes such as phosphorylation (+79.96 Da)

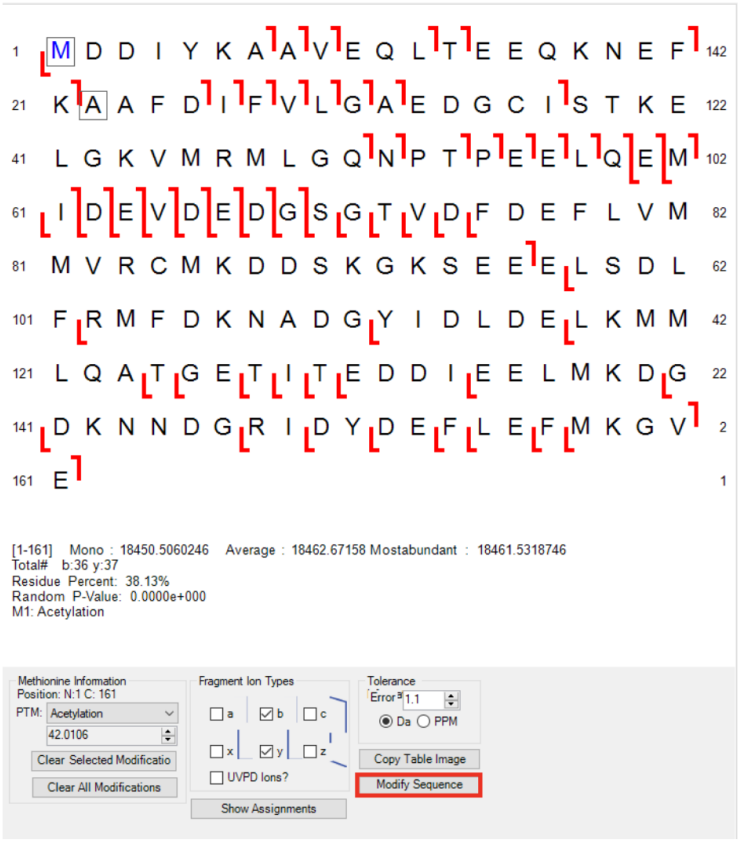


Figure 35: *Sequence Table* panel

and methylation (+14.01 Da). Users can also choose Custom to provide a custom molecular weight change specific for the analysis. A modification can be removed by Clear Selected Modification and all the modifications can be removed by Clear All Modifications.

In the Fragment Ion Types section, users can map the fragment ion in the *Mass List* to the sequence in the *Sequence Table* by selecting fragment ion types and clicking Show Assignments. The lines between two amino acid codes suggested fragment ion detection. We currently support a/x, b/y and c/z fragment ion types (Figure 35). Additionally, the current version supports common UVPD ions such as a+1, x+1, and y-1.

In the Tolerance Information section, the fragment ion mapping can be manipulated

by different tolerance values in both ppm and Da units.

Users can copy the image shown in Sequence Table by Copy Table Image. The copied image can be pasted in either Photoshop or Microsoft PowerPoint for further processing.

8.3 Ion Finder

The Ion Finder tool allows users to perform targeted search for fragment ions against a user imported sequence. In the Theoretical Data of the *Sequence Table* panel, the monoisotopic mass of different ion types is calculated based on the sequence with modifications provided in the Sequence table. Double-clicking a selected entry will open up the Ion Finder tool for the specific fragment number (Figure 36).

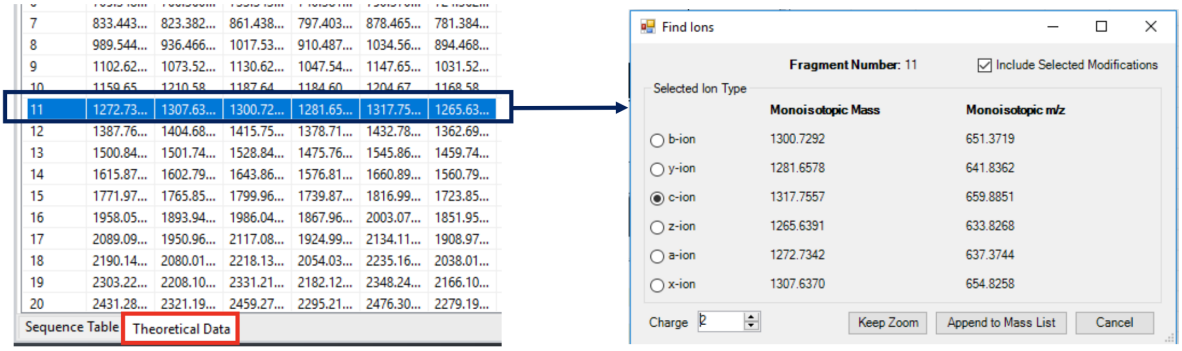


Figure 36: Ion Finder tool

Users can select the fragment ion type for their experiment by clicking at selected radio buttons. Users will also be able to choose to include or exclude selected modification implemented in the sequence. The Ion Finder tool allows users to navigate different charge states of the fragment ion to examine possible experimental fragment ions by zooming at the targeted m/z region calculated based on the monoisotopic mass and charge state of the fragment ion. The spectrum will change accordingly with changes to parameters such as selected ion type, modifications, and charge state (Figure 37), with theoretical value

shown as red circles. Users can stay zoomed- in at the current spectral location by clicking Keep Zoom. If the users find a real ion, clicking Append to Mass List will add the ion information, including modification information, as a new entry at the bottom of the *Mass List* (Figure 37).

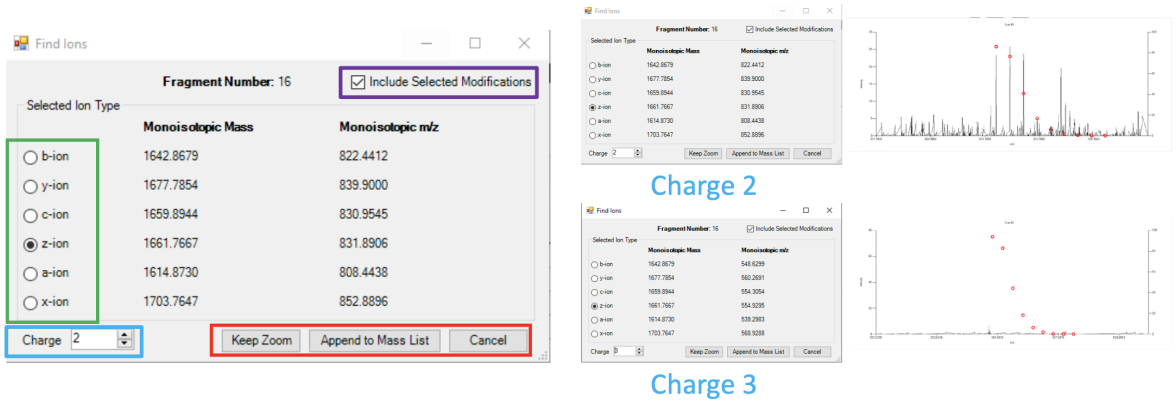


Figure 37: Functions for Ion Finder Tools

9 Tool Bar

The tool bar allows users to access some of the basic functions of the software.

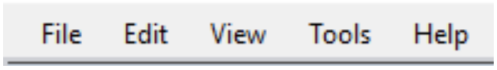


Figure 38: Tool Bar button layout

9.1 File

Items under File include Open Sequence, Open Saved Results, Import Results, Close Experiment, Save Results As, and Exit (Figure 39).

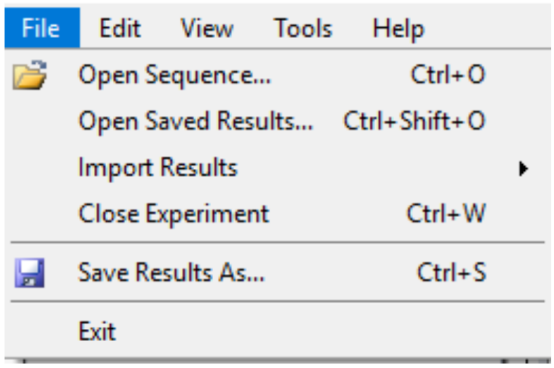


Figure 39: Menus for File option

- Open Sequence allows users to import a .fasta sequence file. This function will be deactivated in future versions.
- The Open Saved Results function provides users a method to open saved .xml results. This function is related to Save Results As function described below. (Note that results files can also be opened by dragging and dropping them on the main MASH Explorer window.)
- The Import Results item has options for users to import both an ion mass list (in .msalign format) and a protein search result (in related database search algorithm format).
- Close Experiment will close the current imported data set.
- Save Results As will output an xml file that includes the information from the *Mass List* and *Sequence Table*. The xml file cannot be opened by itself as it is dependent on the corresponding raw data. If a raw data file is not found, the program will prompt the user to specify the raw data for reference.
- Exit will close the software application.

9.2 Edit

Items under Edit include Copy Checked Mass(es), Copy Sequence Table, Copy Spectrum Image, and (Un)Check All Masses (Figure 40).

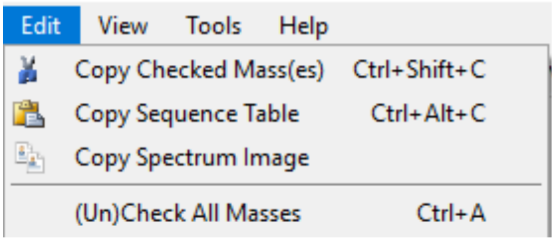


Figure 40: Menus for Edit option

- Copy Checked Mass(es) will copy the selected items in the *Mass List* (i.e. items with a checkmark in the checkbox). The copied item can be pasted in other documents such as Notepad or Microsoft Excel for further editing.
- Copy Sequence Table will copy the image in the *Sequence Table*. This function is the same as previously shown in Section 5.2 Save Images.
- Copy Spectrum Image will copy the image in the *Spectrum View*. This function is the same as previously shown in Section 5.2 Save Images.
- (Un)Check All Masses will select or deselect all the entries in the *Mass List*.

9.3 View

Items under View include Results View, MassList Toolbar, Status Bar, Sequence Table, Fitting Curve, and Fragment Labels (Figure 41).

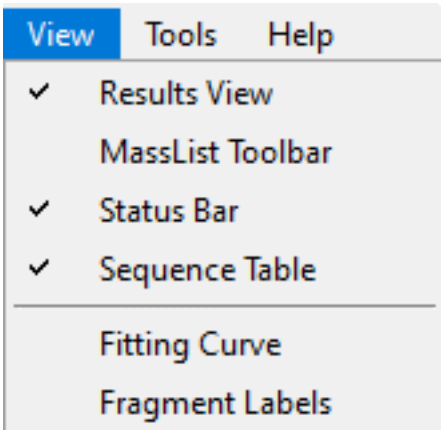


Figure 41: Menu for View option

A checkmark next to each item indicates that the corresponding GUI element is visible. For detailed information regarding the location of the Results View, MassList Toolbar, Status Bar, Sequence Table panels, refer to Section 2. These items allow users to reopen a specific panel if it has been closed.

Fitting Curve and Fragment Labels are not panels in the GUI. Rather, they are options to show additional information in the Results View when a Mass List item is selected.

9.4 Tools

Items under Tools include Workflow Manager, Quick Decon. Parameters, Paste Sequence Configuration and View Log File (Figure 42).

- The *Workflow Manager* is a dialog that controls the data analysis workflow for top-down proteomics. This tool is activated when a workflow such as deconvolution or database search is activated. If the user closes a Workflow Manager, it can be reopened using this menu selection.

- *Quick Decon. Parameters* allows users to change parameters for the Quick Deconvo-

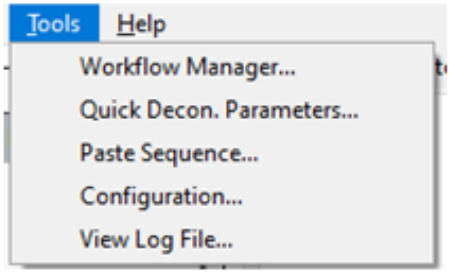


Figure 42: Menu for Tools option

lution function.

- *Paste Sequence* changes the sequence displayed in the Sequence Table panel. The function of the button is the same as Modify Sequence in the Sequence Table and Paste Your Sequence under Characterization within the Targeted Mode workflow.

- The *Configuration* dialog allows users to configure both directory of the deconvolution and database search and protein modifications.

  – *File Paths* tab (Figure 43). For more information regarding the deconvolution and database search, please refer to the Search Algorithm Setup Guide included in the MASH Explorer software installation regarding installation of each deconvolution and algorithm. Users can use Find option to attempt to automatically locate the relevant directory. If the file is not found, the Browse option will permit users to manually specify the directory of the selected software. Users can click the Download button to access the software download page for a given software app. Green indicates that the target file can be located, while red indicates that the target file is not found in the specified directory.

  – *Modifications* tab (Figure 44). Users can also add new protein modifications

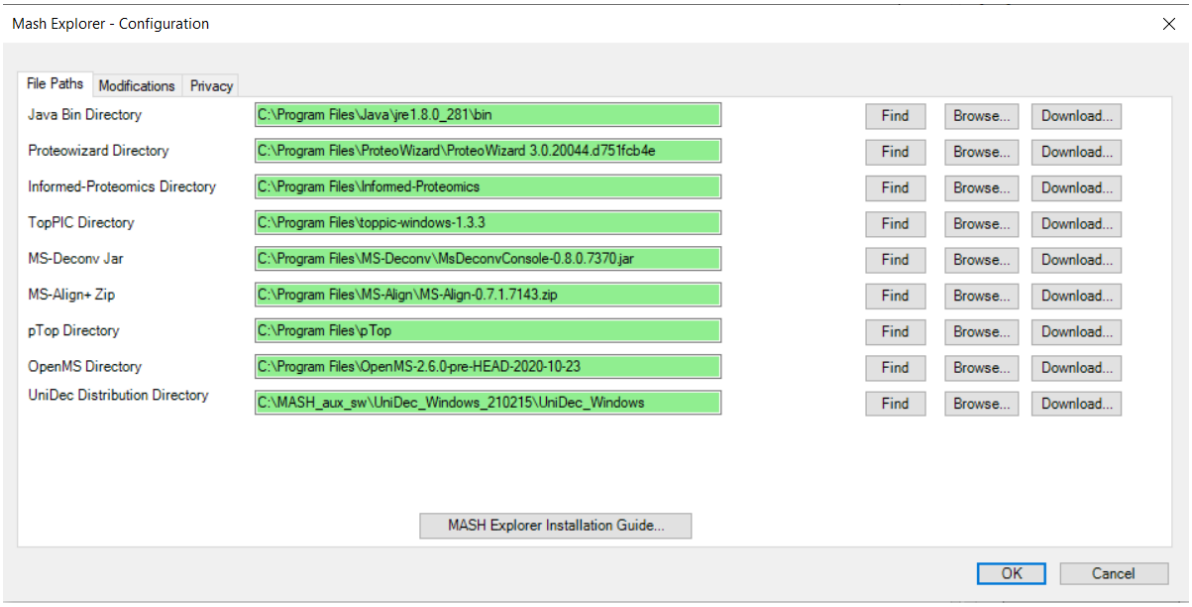


Figure 43: MASH Explorer - Configuration for supporting software

or edit existing modifications (Figure 44). The name, monoisotopic mass and average mass can be input. Press "OK" to save the modifications, and the modifications will be added in the PTM list in the *Sequence Table* panel.

– *Privacy* tab (Figure 45). Users can view and change their privacy/data reporting settings at any time. The options available in the Privacy tab of the Configuration dialog are the same as those available in the Privacy Settings dialog (see Section 7 for details).

- *View Log File* brings up a text editor window with the MASH Explorer log file.

9.5 Help

Items under Help are Getting Started, User Manual, License, and About MASH Explorer (Figure 46).

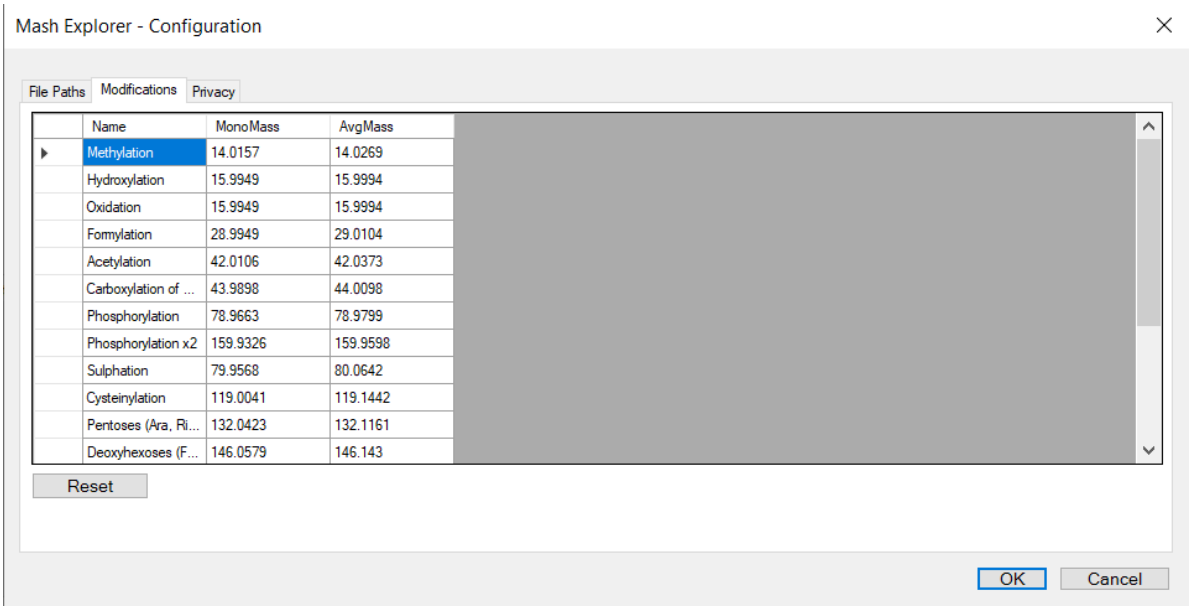


Figure 44: MASH Explorer - Configuration for modifications

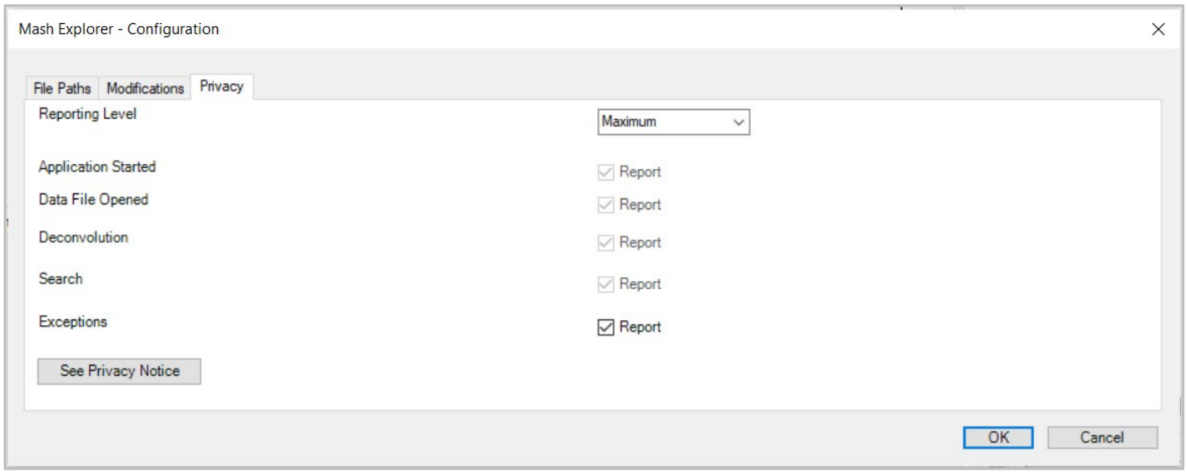


Figure 45: MASH Explorer - Configuration for privacy settings

- Getting Started will open the following website webpage: http://labs.wisc.edu/gelab/MASH_Explorer/GettingStarted_MASHExplorer.htm, which introduces the basic functions in MASH Explorer software.

- User Manual will open the MASH Explorer user manual in your web browser.

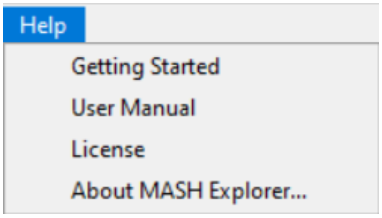


Figure 46: Menu for Help option

- License will open the MASH Explorer license agreement in your web browser.
- About MASH Explorer provides information about the MASH Explorer software, contact information, and references for the supporting software.

Users can also contact MASH team support at mash-support@g-groups.wisc.edu

10 Configuration and Log Files

The MASH Explorer application uses a number of configuration files, and a log file, to store information. These files are found in the folder <user>\AppData\Local\MASH_Explorer. For example, if your user name is johndoe, your configuration and log files would be found in the folder C:\Users\johndoe\AppData\Local\MASH_Explorer if your home folder is in the usual location.

10.1 Configuration Files

When you first start MASH Explorer, it will automatically create a number of configuration files in the MASH_Explorer folder. These configuration files are used to preserve your user preferences across runs of the software. You can change configuration values via the **Tools** → **Configuration** menu option. The Process Wizard also allows you to change

the configuration for both Quick Decon and "normal" deconvolution/search processing. (If necessary, the configuration files can also be manually edited with a tool such as Notepad; however, this should only be done in the case of faulty configuration that prevents MASH Explorer from running, because it is possible to produce an invalid configuration file if you manually edit it.)

Currently MASH Explorer uses the following configuration files:

- `MASH_Explorer_FilePaths.txt`
- `MASH_Explorer_NormalDeconSearch.txt`
- `MASH_Explorer_QuickDecon.txt`

In the future the number of configuration files will increase as we increase the flexibility of the MASH Explorer software.

10.2 Log File

MASH Explorer will create a log file called `MASH_Explorer_LogBook.txt` when it runs. Each time you run MASH Explorer, the log file will be appended to - it will never be overwritten. (If you are concerned about the log file getting too big, you can manually delete or rename it.)

The MASH Explorer log file contains a record of the actions you perform each time you run the application.

When you start MASH Explorer, the log file will contain three lines like this:

```
--------------------------------
2019-05-28 11:26:03:  Log file opened/created
2019-05-28 11:26:03:  MASH Explorer version 1.0.0.20579
```

This tells you when MASH Explorer was started, and what the version of the software is.

The log file also contains some information about the versions of other software packages used by MASH Explorer – for example:

```
2019-05-28 11:27:00:  ProteoWizard version:  3.0.19014
```

The log file contains information about all of the processing you do while running MASH Explorer. This can be useful if you need to go back and figure out what processing you did on a given data set.

The log file can also be useful if you run into any problems with MASH Explorer - if you need to ask us for help with a problem, please preserve the relevant log file so that you can send it to us if necessary.