MASH Native

User Manual: Version 1.1

Ying Ge Research Group - University of Wisconsin, Madison

Contents

1 Overview

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

This manual describes the functions available within MASH Native for processing native and denatured top-down mass spectrometry data. Video tutorials can be found on YouTube™.

The development of MASH Native 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 Welcome Screen

When you start MASH Native (version 1.1 or later) for the first time, you will be shown the Welcome Screen (see Figure **1**). (Note that if you want to see the Welcome Screen each time you start MASH Native, you can do so by unchecking the "Don't Show Again?" checkbox in the lower left corner. If you want to see the Welcome Screen at any point while running MASH Native, you can do so by selecting *Help* → *Welcome Screen* in the

main toolbar.)

While the Welcome Screen is shown, MASH Native will install the ProteoWizard, UniDec, TopPIC, MS-Deconv and MS-Align+ supporting software (which are now part of the MASH Native .zip file). If you start MASH Native and included supporting software is not installed, the Welcome Screen will be shown even if the "Don't Show Again?" checkbox is checked. You can browse the listed documents while the installation process is occurring.

The components of the MASH Native Welcome Screen are as follows:

- *User Documents:* The user documents include the User Manual, Installation Guide, Getting Started Guide, and YouTube™ Tutorials; each will open in your web browser when selected. The user manual provides you with in-depth information on the capabilities of MASH Native and each aspect of the application. The Installation Guide will walk you through the installation of MASH Native and the included and additional software. The Getting Started Guide introduces the basic functions in MASH Native software and walks you through some example workflows. The YouTube™ Tutorials button will open the Ge Research Group YouTube™ channel which has videos going through the Getting Started and Installation Tutorials.

- *Support Links:* The support links will open in your web browser when selected. The `mash-support` email link will open an email to our support line for any help you may need. The MASH Native contact page will open our contact page for additional help. The MASH Native home page will open the MASH Native website with additional user information including a user forum, user documents, and description of our software.

- *"Don't Show Again?" checkbox:* The "Don't Show Again?" checkbox can be checked or unchecked according to whether to want to see the Welcome Screen each time you start MASH Native.
- *Go! Button:* Once UniDec and TopPIC Suite are finished installing, the Go! button will be activated, and you can open MASH Native.
- *Exit Button:* You can select the Exit button to close the Welcome Screen and exit the MASH Native app.

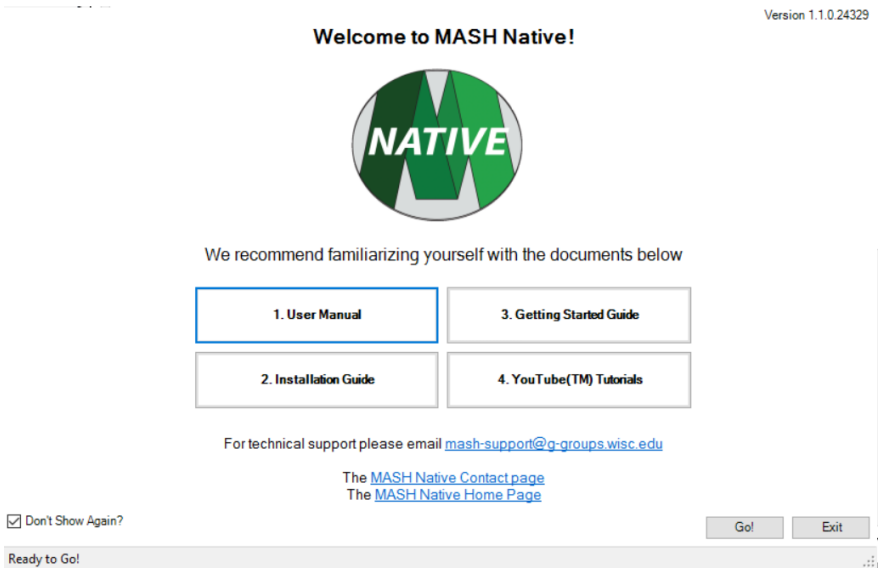


Figure 1: MASH Native Welcome Screen

3 Interface Layout

The components of the MASH Native main window are described below (see Figure **2**).

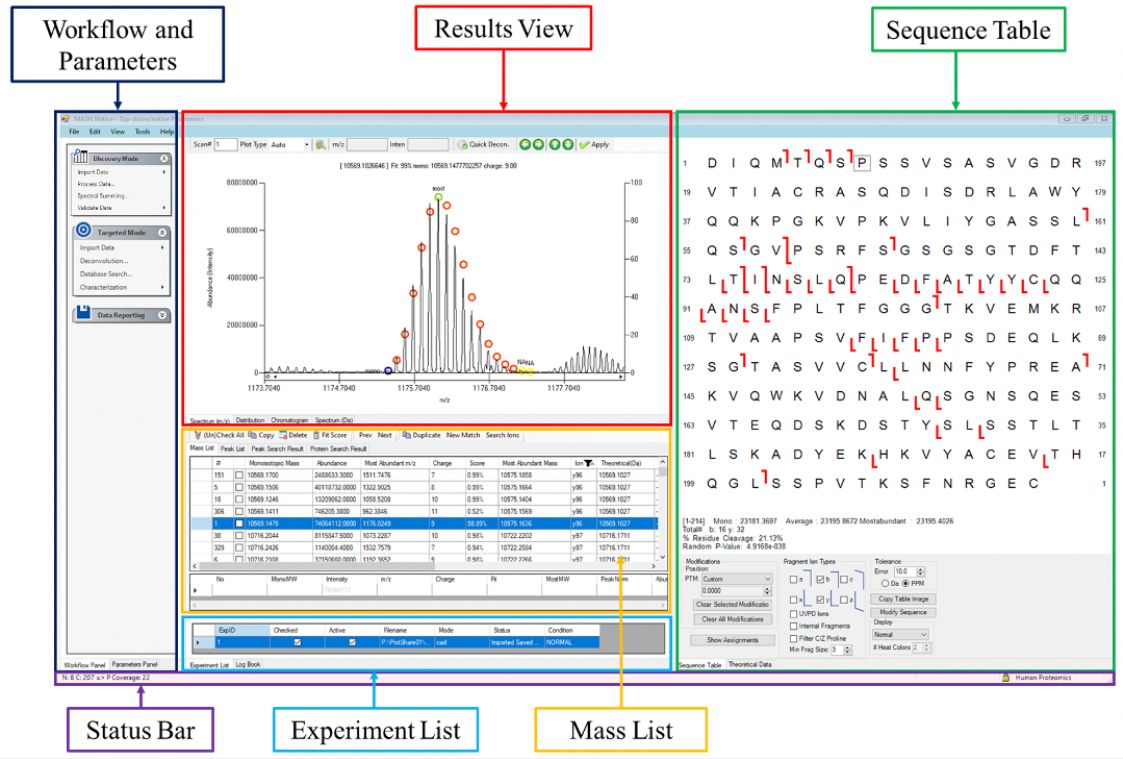


Figure 2: Overview of MASH Native software Interface

- *Workflow and Parameters:* The menus in the Workflow tab allow you to perform data processing for liquid chromatography (LC)-tandem mass spectrometry (MS/MS) and MS/MS analyses. The Parameters tab allows you to view and edit various parameters of the software and data, such as the amino acid sequence.

- *Results View:* The mass spectra and chromatogram of the active experiment 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 MASH Native, and allows you to switch which experiment is active by double-

clicking on the desired experiment. (The active experiment is highlighted in blue.)

- *Log Book:* This panel displays processing status and error messages. (Note that the Experiment List and Log Book are tabs in the same panel of the MASH Native App window.)
- *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.

4 Usage Reporting/Privacy

MASH Native, and MASH Explorer starting with version V2.0.1, optionally report data about your usage of MASH App to the MASH team. Please refer to the MASH Native 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 Native user documentation. When you first run a new version of MASH Native (or MASH Explorer V2.0.1 or later) you will see the Version Changed dialog (Figure **3**). This asks you to review your privacy settings.

When you click "OK" in the Version Changed dialog, you will see the Privacy Dialog (Figure **4**). This dialog allows you to set the level of data reported by MASH Native (note that this setting is separate for each individual MASH Native 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"

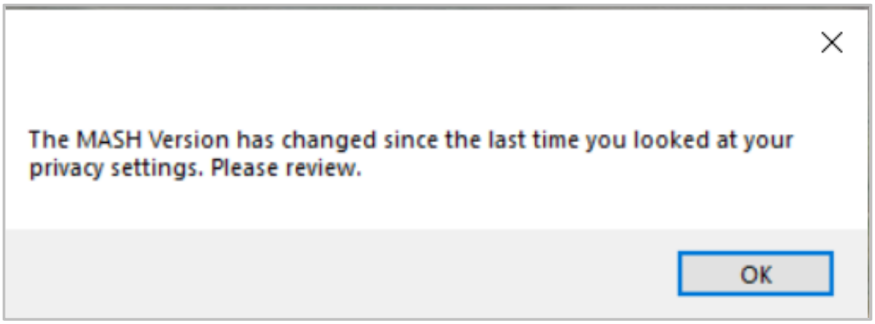


Figure 3: Version Changed Dialog

allows to you choose whether to report each type of data.

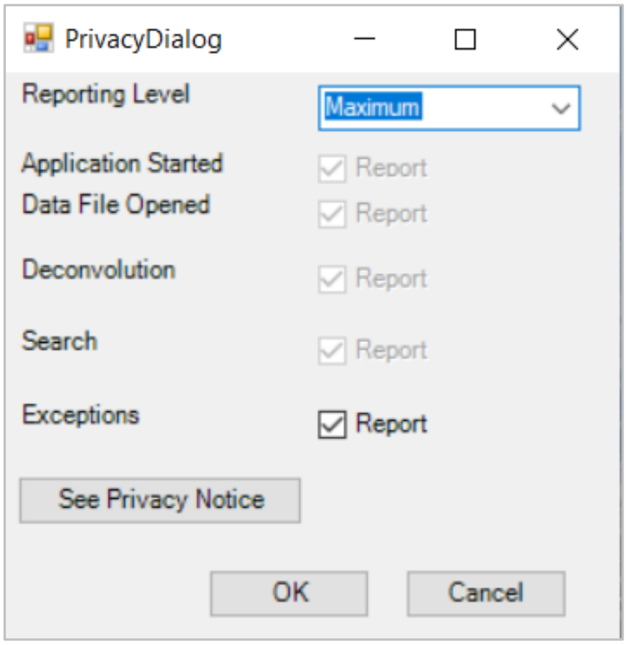


Figure 4: Privacy Settings Dialog

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 **5**).

Figure 5: Privacy Settings Confirmation Dialog

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

Our team designed the workflow termed "Discovery Mode" (see Figure **6**) to handle batch LC-MS/MS analysis. Discovery mode is intended to identify unknown identify unknown proteins analyzed via mass spectrometry under both native and denaturing conditions. This workflow includes data import, data processing including deconvolution and database search, and finally data validation for protein identification.

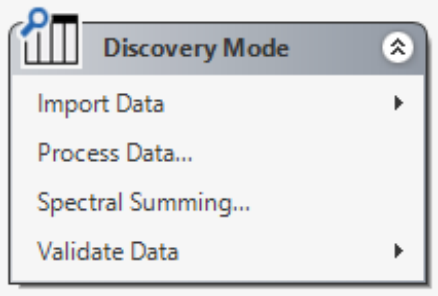


Figure 6: Discovery Mode button layout

5.1 Import Data

In the "Discovery Mode" section of the *Workflow and Parameters* panel, Import Data will direct users to load experimental data to the MASH Native software. You 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 **7**).[1]

2. Click on the option that represents the raw data file format that will be uploaded (Figure **7**). The current version of MASH Native software supports Thermo (.raw), Bruker data directory (.d), Bruker (.baf), Bruker data output in ASCII format (.ascii), and universal file formats (.mzXML, .MGF, and .mzML).

3. Select the LC-MS/MS data files to be analyzed. The *Status Bar* will update the data importing progress (Figure **8**).

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

You can view the chromatogram of the imported data and navigate different scans. By navigating the scans, you can visualize the different m/z spectra of each scan in the “Spectrum (m/z)” tab. The chromatogram can be visualized in the "Chromatogram" tab in the *Results View* panel (Figure **9**).

[1] You can also choose to drag and drop the data file to import data into the software.

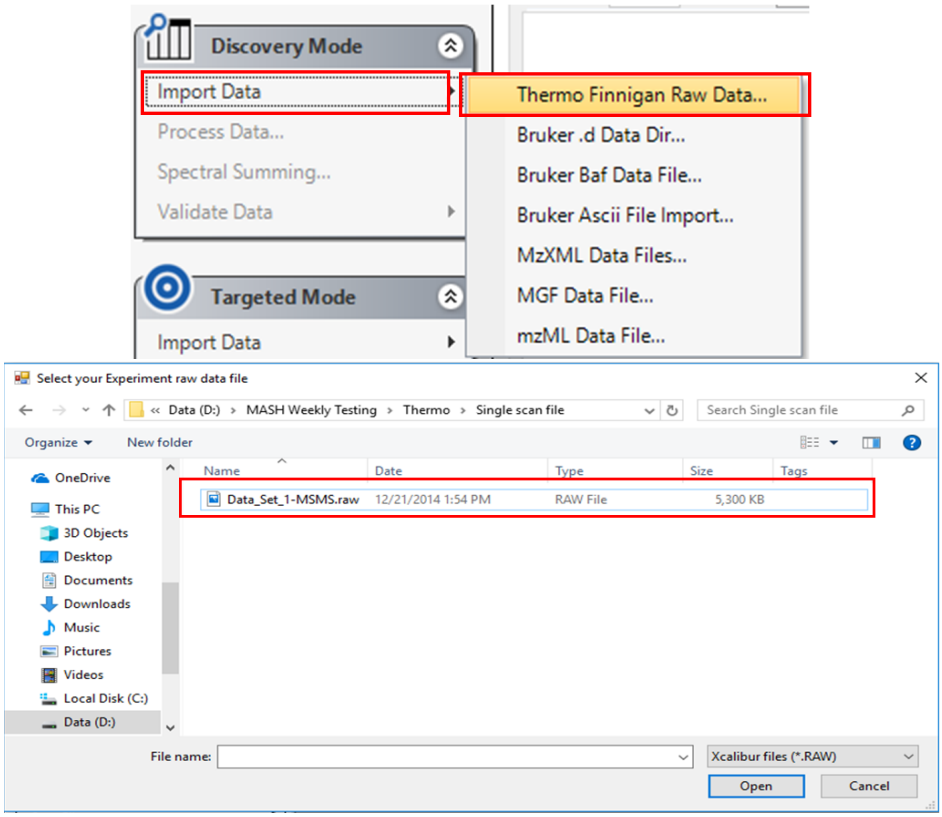


Figure 7: Data Import for Discovery Mode. Seen are (top) the Import Data menu; (bottom) the file selection dialog.

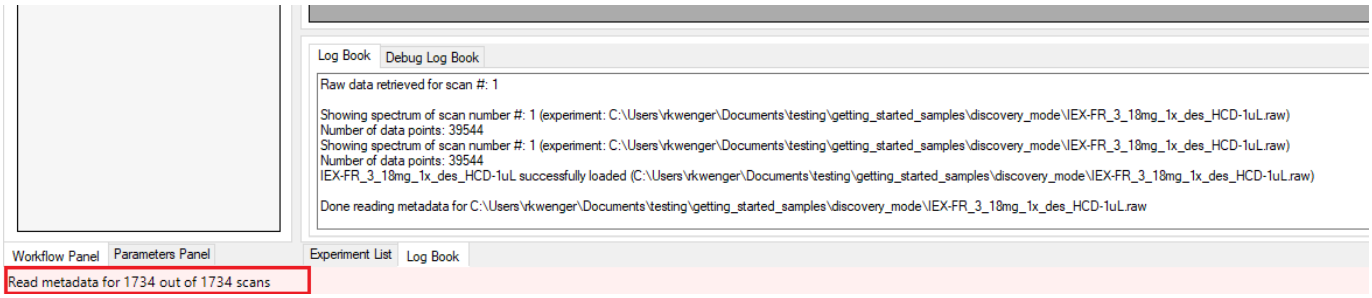


Figure 8: *Status Bar* panel updates for data import

You can navigate to different scans in three ways:

- 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

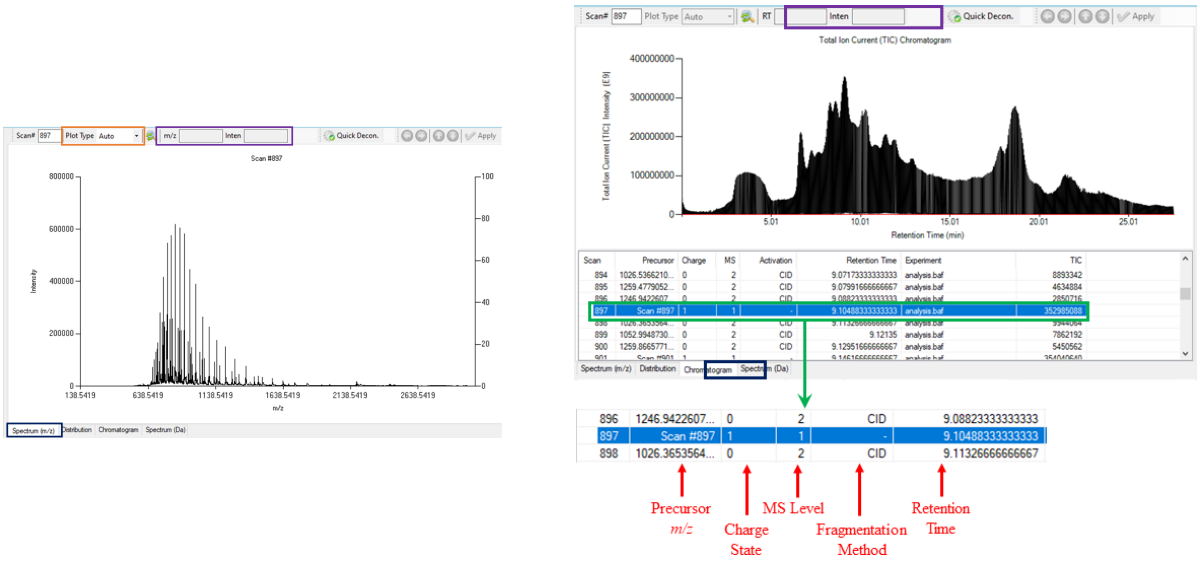


Figure 9: Viewing individual scan and chromatogram

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 **9**)

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 **7**). 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 9 purple box).

In the toolbar, there is a an option to select the Plot Type (Figure **9**, orange box), which determines the way the spectra are plotted to help with imported experiments whose spectra that are either centroided (Centroid) or are a profile scan (Profile). The default is Auto, which means to automatically detect whether the spectra have been centroided or not. Centroid mode shows the m/z spectrum as impulses, whereas Profile mode creates a connected line plot.

5.2 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. Click on *Process Data* and the Process Wizard will be opened (Figure 10 left).[2] Select a deconvolution method, a database search method, and upload a targeted database file in .FASTA format (Figure **10**, left). You can click on the *Advanced* tab to change general settings and algorithm parameters for each algorithm (Figure **10**, 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, you have the option to input fixed and variable post-translational

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

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 Native 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 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 **7** below.

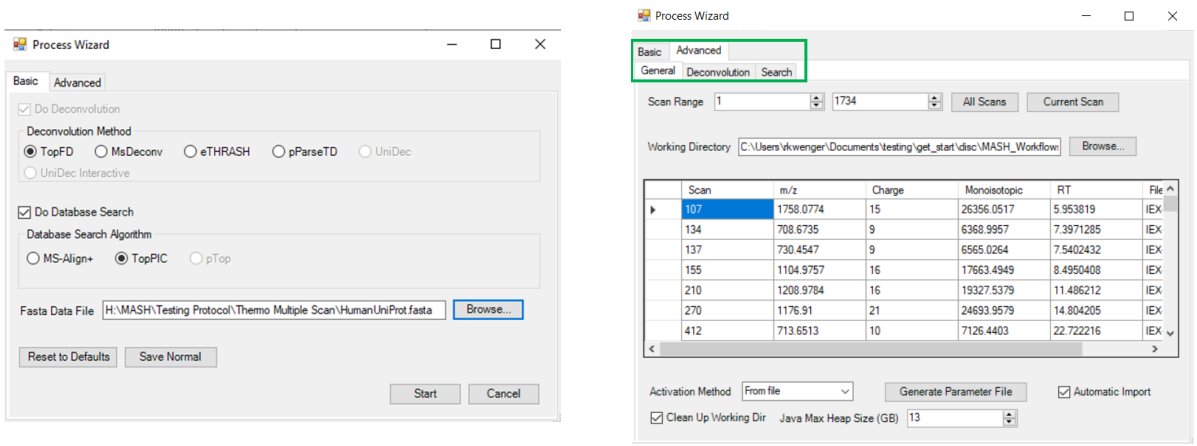


Figure 10: Basic and Advanced settings for Process Wizard in Discovery Mode

5.3 Spectral Summing

Spectral Summing opens the spectral summing tool (Figure **11**), allowing users to sum scans within their original experiment file (you can also sum the results of a summation). The spectral summing process can be initiated in the Discovery Mode panel, or you can

select the spectral summing tool in the "Tools" menu. The summing process generates an mzML file which can then be processed by MASH Native, including deconvolution and searching. The summing tool is designed to be flexible in giving the user control over the scan selection, the summing algorithm to use, and how to handle MS1 scans in the dataset. The main tab contains three sections, one for the strategy used to select scans, one to determine how the MS1 scans are handled, and finally one to determine which Summing Algorithm to use.

- *Selection Strategy* Currently, three options exist for selecting scans. The "All Scans" option selects all of the scans in the experiment. The "Select Ranges" selects all scans in a range provided in the "Range Selection Parameters" on the "Parameters" tab of the Spectra Summing Tool dialog. The "Proteowizard" option selects scans based upon the precursor, scan time, and ion mobility tolerances specified in the "Proteowizard Select Parameters" section on the "Parameters" tab of the Spectra Summing Tool dialog. The "Proteowizard" option will find groups of scans to sum, and perform the summing on each group using the chosen summing routine.

- *Summing Algorthm* The summing tool allows up to four choices of Summing Strategy. UniDec-Interpolate and UniDec-Integrate provide an implementation of spectral summing similar to UniDec's respective summing tool. If the experiment file is an Thermo Raw or Waters file type, the Vendor option is available to use the corresponding vendor's algorithm for summing the spectra. Finally, the "Proteowizard" option uses ProteoWizard's method of performing spectral summing.

- *MS1 Handling* The summing algorithm sums selected MS2 scans, but also allows three options for handling MS1 scans. The "Skip" option skips all MS1 scans and

doesn't save them in the resulting mzML file. The "Copy" option just copies the MS1 scans as they were in the original experiment file, and the "Sum" option will also sum the MS1 scans and put the result in the new mzML file.

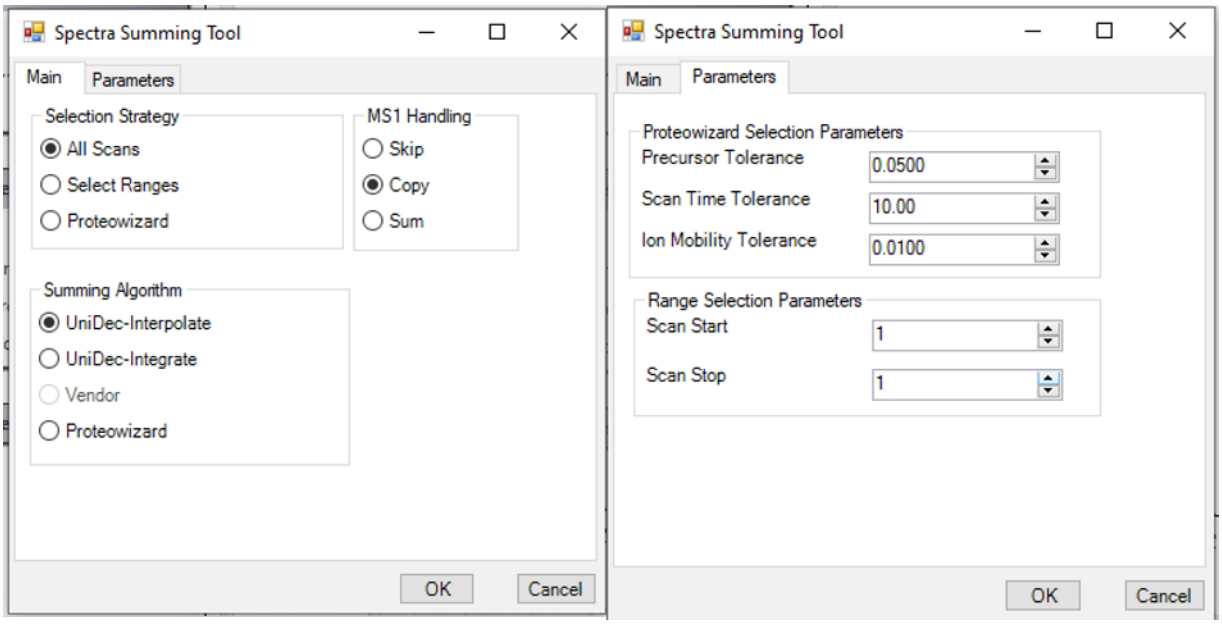


Figure 11: Spectral Summing dialog which allows you to define which scans should be summed, and to set parameters for summing. Seen are the Main (left) and Parameters (right) tabs.

When the options are selected and the "OK" is clicked, the spectral summing is performed on the active experiment file; the results of the summation are saved to an mzML file. This mzML file is automatically loaded into MASH Native after summing has been performed (it shows up as a new experiment in the Experiment List). The spectral summing tool uses the Workflow Manager to perform the summing and the results and logs are saved in a workflow folder for the user to reference.

5.4 Validate Data

Workflow results can be imported by using *Validate Data* → *Import Mass List* to import a deconvoluted mass list and *Validate Data* → *Import Database Search* to import protein search results (Figure **12**). By clicking on entries in the Protein Search Results, the corresponding protein sequence will be displayed in the *Sequence Table* panel (Figure **13**). A detailed description of the functions of the *Sequence Table*, such as adding modifications and displaying fragment ions can be found in Section **9** of this manual. You can also import the mass list, which contains fragment ion information, which will be displayed in the *Mass List* panel (Figure 13). A detailed description of the functions of the *Mass List*, such as correction of fragment ion charge state and monoisotopic mass, is provided in Section 9 of this manual.

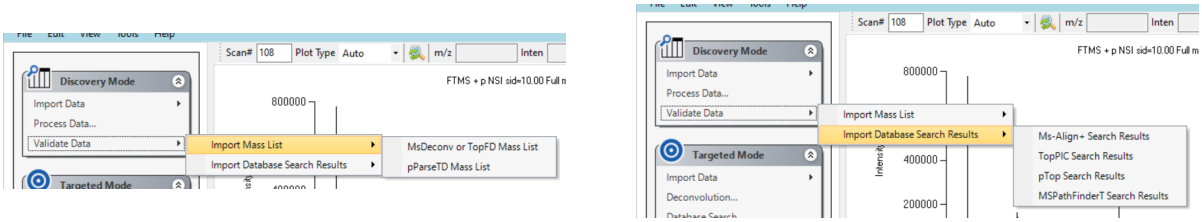


Figure 12: Menus for Validate Data

5.4.1 Automatic Import

For your convenience, the mass list and protein sequence are automatically imported after deconvolution and database searching. However, when you run a 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 analyses (shown on the right of Figure **12**).

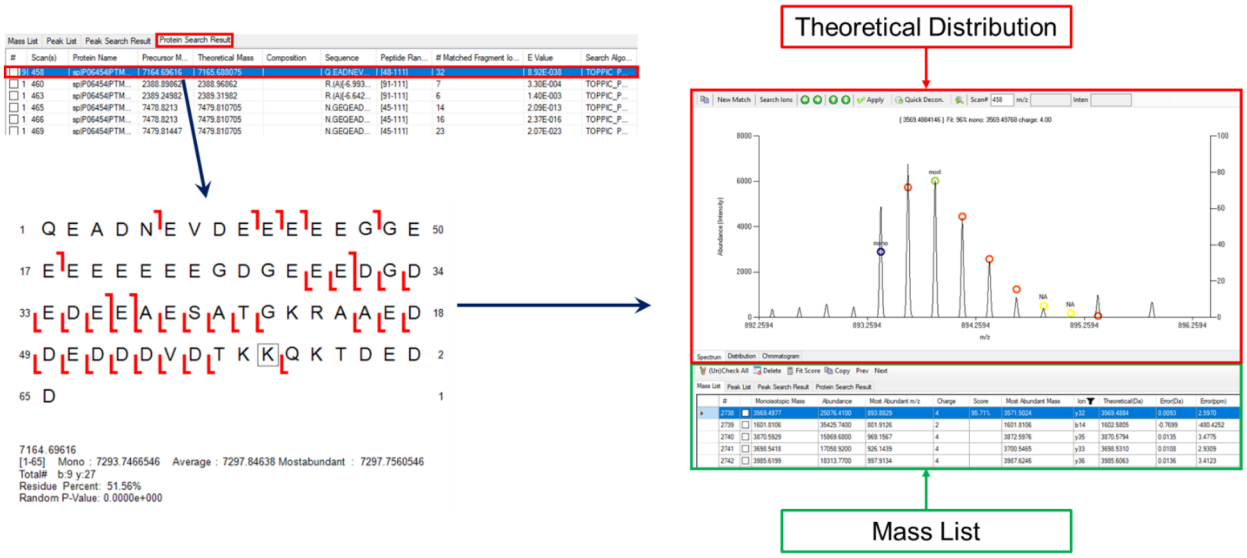


Figure 13: Analysis of protein sequence and fragment ion

To disable automatic import, you can uncheck the automatic import checkbox in the Advanced/General tab (shown in the right half of Figure **12**, the checkbox at the bottom right)

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

In contrast to "Discovery Mode", our team developed the "Targeted Mode" workflow (see Figure **14**) for analysis including comprehensive protein characterization. This workflow includes data import, spectral deconvolution, identification of fragment ions, and database search based on the identified ions. 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.

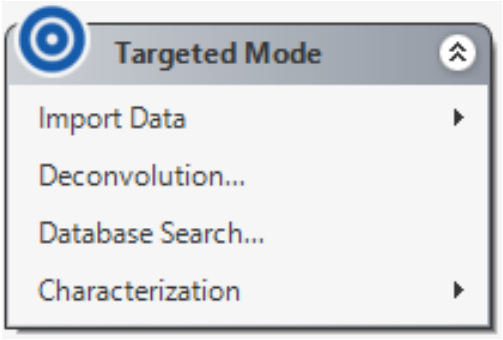


Figure 14: Targeted Mode button layout

6.1 Import Data

Similar to the Import Data in "Discovery Mode", Import Data in "Targeted Mode" will direct you to import data into the software. You 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 **14**).[3]

2. Click on the option that represents the raw data file format to be uploaded (Figure **15**). The current version of MASH Native software supports Thermo (.raw), Bruker data directory (.d), Bruker (.baf), Bruker data output in ASCII format (.ascii), and universal file formats (.mzXML, .MGF, and .mzML).

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.

[3]You can also choose to drag and drop the data file to import into the software.

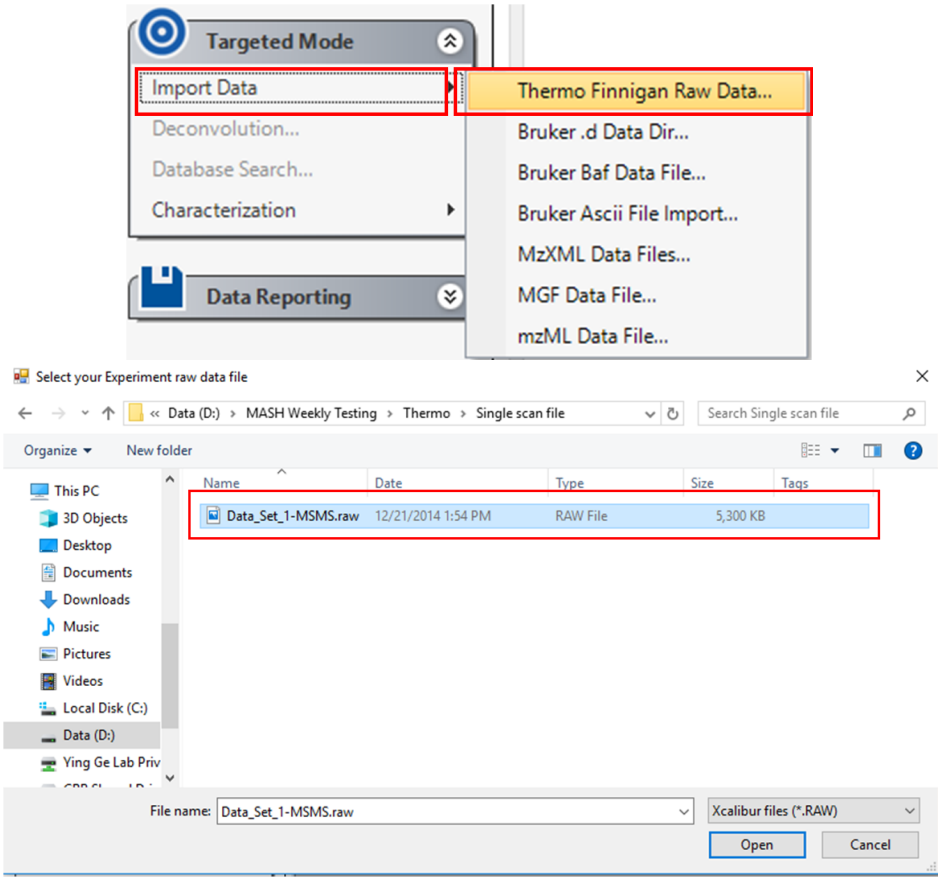


Figure 15: Data Import for Targeted Mode. Seen are (top) the Import Data menu; (bottom) the file selection dialog.

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 Native software.

2. Examples for the second category include instruments that perform LC-MS/MS. For example, users inject a solution containing a single protein and use the LC to deliver the sample to the mass spectrometer. The resulting raw data has multiple scans while each scan contains fragment ions of the same protein. Further pre-processing of the data before importing it into the MASH Native software can aid in analyzing

this type of data.

(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 **16**) for spectral averaging, and the summed spectrum can be exported in mzXML format. Ideally, the output mzXML file has only one scan.

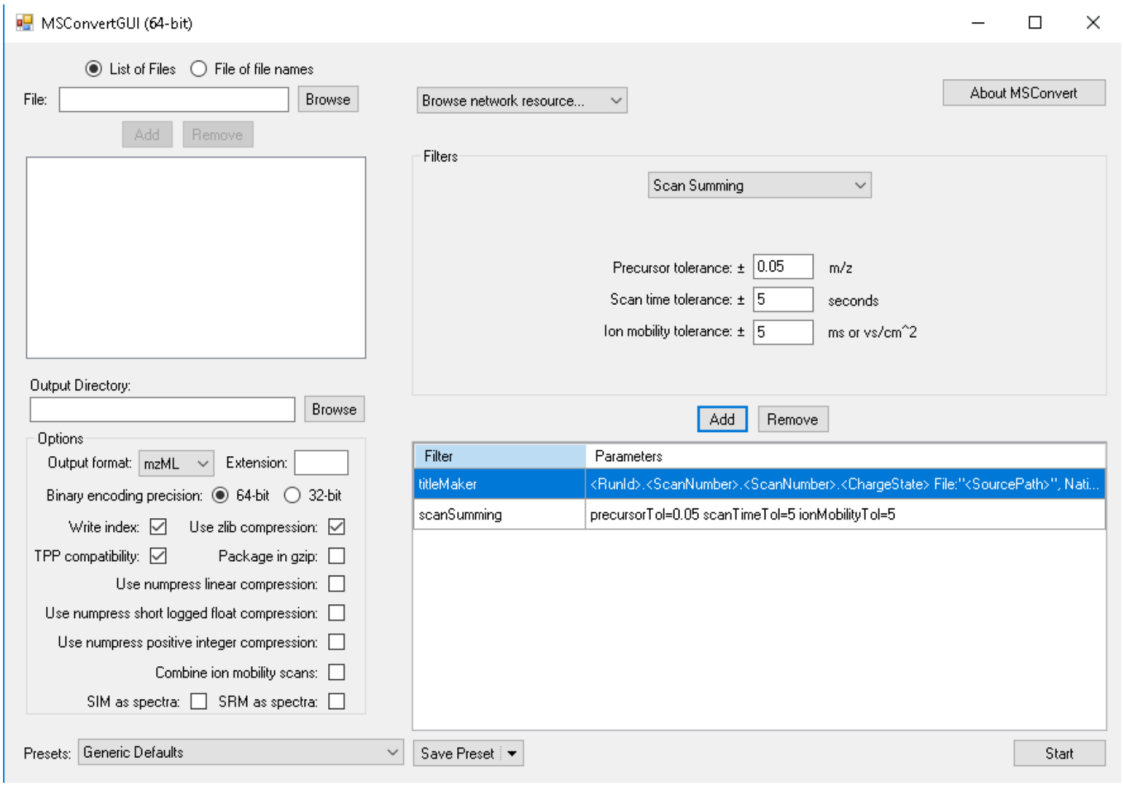


Figure 16: ProteoWizard Scan Summing

6.2 Deconvolution

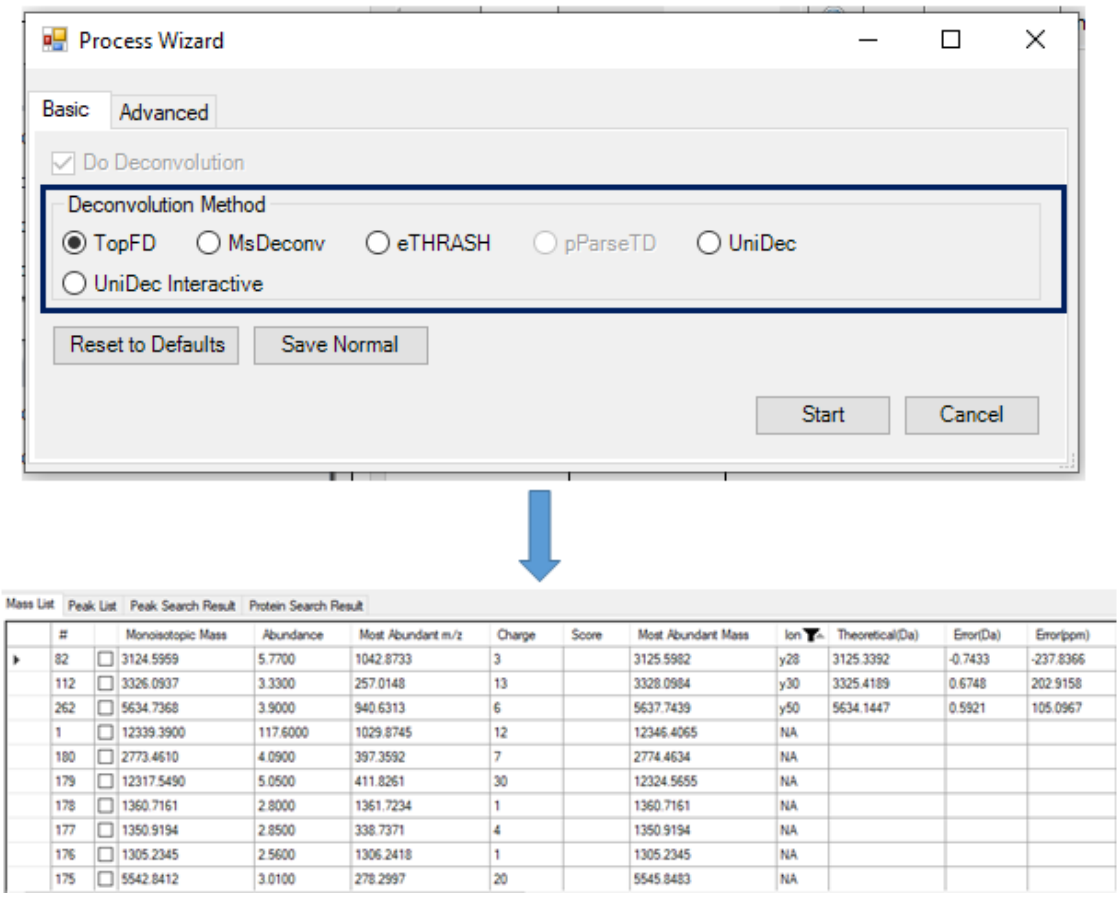
Spectral deconvolution is the process of identifying fragment ion peaks and calculating their charge states, monoisotopic mass, most abundant mass, intensities, and other parameters. Our software currently offers four tools for spectral deconvolution in Targeted Mode: enhanced-THRASH (eTHRASH), TopFD, MS-Deconv, and UniDec. Each of these algorithms can be automatically run by the MASH Native software using the Process Wizard (Figure **??**). Starting the deconvolution processes will open the Workflow Manager. When the calculation is finished, the calculated mass list from each algorithm will be automatically imported and displayed in the Mass List panel (Figure **17**). Users can click on each fragment ion entry for viewing. The detailed functions of the Mass List are described in Section **9**.

6.2.1 Notes on spectral deconvolution

1. The current version of MASH Native 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 ion charge state as shown in Figure **18**.

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 Native runs UniDec automatically. In interactive mode, MASH Native 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



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

Figure 17: Basic interface of deconvolution-mode Process Wizard and populated Mass List

MASH Native.

4. The *Quick Decon.* button in the *Results View* can be used as a quick tool for spectral deconvolution. It supports the same deconvolution options as targeted mode, and the parameters can be configured via *Tools* → *Quick Decon. Parameters*. Details can be found in Section **7**.

5. The current version of MASH Native 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 **7**.

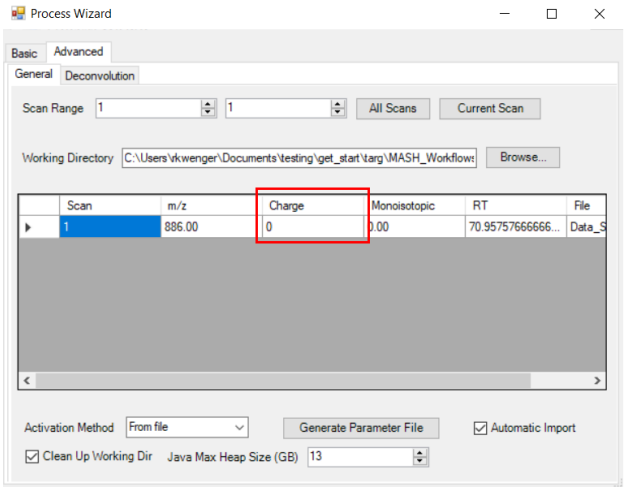


Figure 18: Process Wizard for Deconvolution

6.2.2 UniDec Deconvolution and the Deconvoluted Spectrum on the Mass (Da) Level

After running a UniDec deconvolution through MASH Native or importing a UniDec processed HDF5 file, MASH Native extracts the deconvoluted spectrum with mass (Da) on the x-axis and probability on the y-axis; this is referred to as the Da spectrum. The extracted Da spectrum can be investigated using the "Spectrum (Da)" tab in the Results View (see figure **19**).

Once the Da spectrum is imported, the user can use the zooming tools as with the m/z spectrum. The Da and Prob (probability) toolbar items will also report the mass (Da) and probability values corresponding to the mouse cursor location. Since the Da spectrum is an estimation provided by UniDec, the Plot Type is set to profile mode and cannot be changed. The Da Spectrum can also be copied as an image for pasting into other documents such as PowerPoint using the Edit menu as indicated in figure **20**.

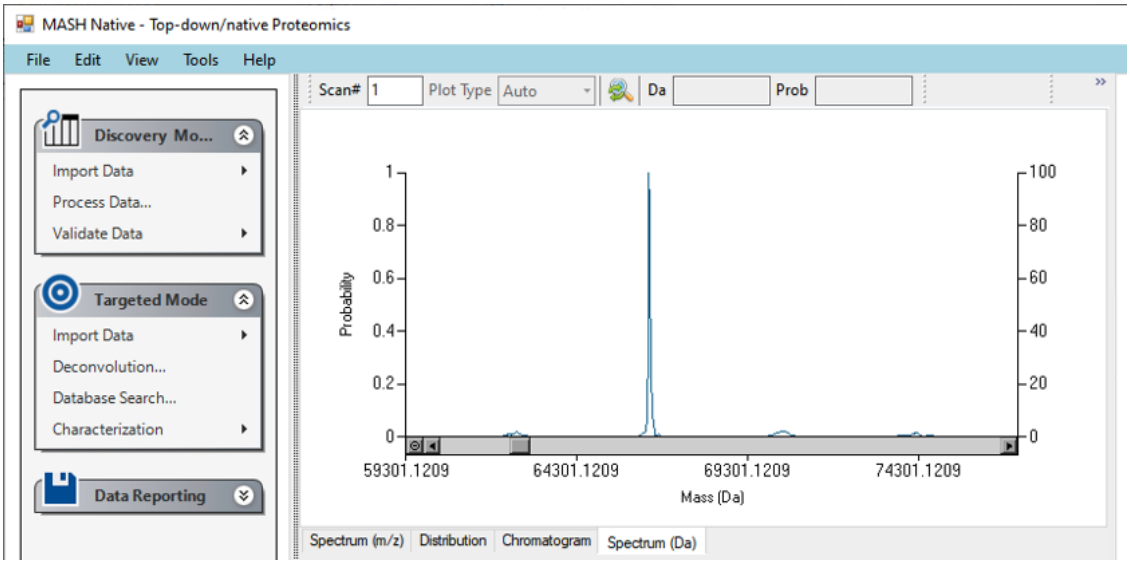


Figure 19: Example Deconvoluted Da Spectrum

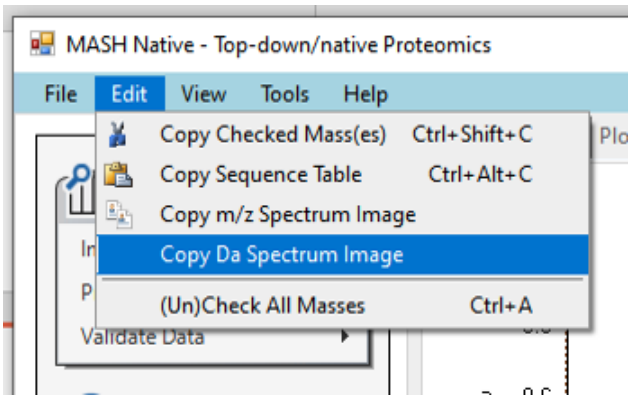


Figure 20: Copying the Da spectrum as an image

UniDec provides many different parameters that can affect deconvolution. Please see the Best Practices for UniDec Deconvolution in MASH documentation for information about setting these parameters as appropiate for the type of spectra you are interested in getting deconvoluted results from.

6.3 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 **21**). Click "Start" to begin the database search process and similar to other processes, the Workflow Manager will be invoked to handle algorithm execution.

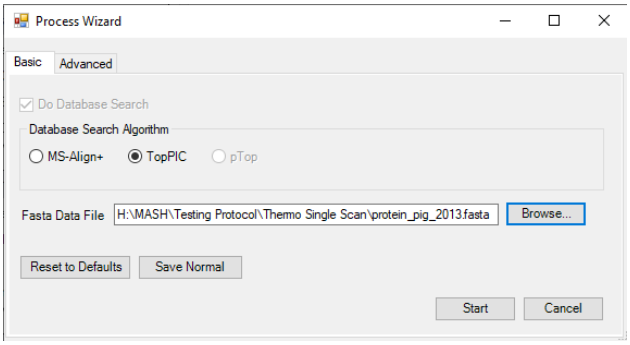


Figure 21: Basic Interface of Process Wizard in Search mode

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 **22**). This information will help substantially in file parsing processes prior to database search.

Notes on database search

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

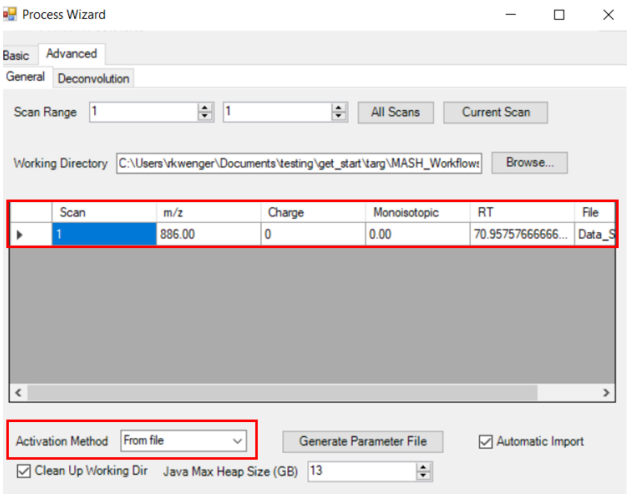


Figure 22: Advanced settings for search-mode Process Wizard

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 **22**. This information will help substantially in successful file parsing processes.

For more information about running workflows, see section **7** below.

6.4 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 **23**). A detailed description of the Sequence Table panel is in Section **9**. For database search list import, please refer to Section **5**.

To manually paste 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 **24**).

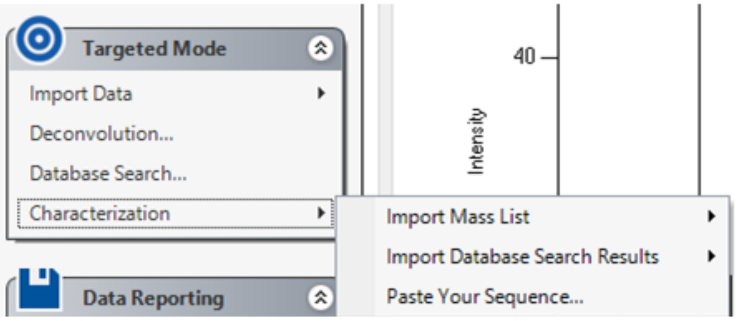


Figure 23: Menu for Characterization

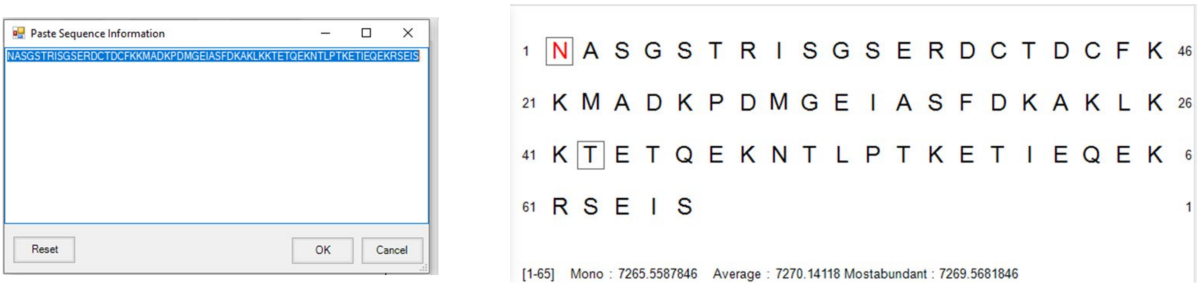


Figure 24: Pasting a protein sequence

7 Workflows

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 Native will generate a workflow-specific folder containing all files related to that workflow (Figure **25**). 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.

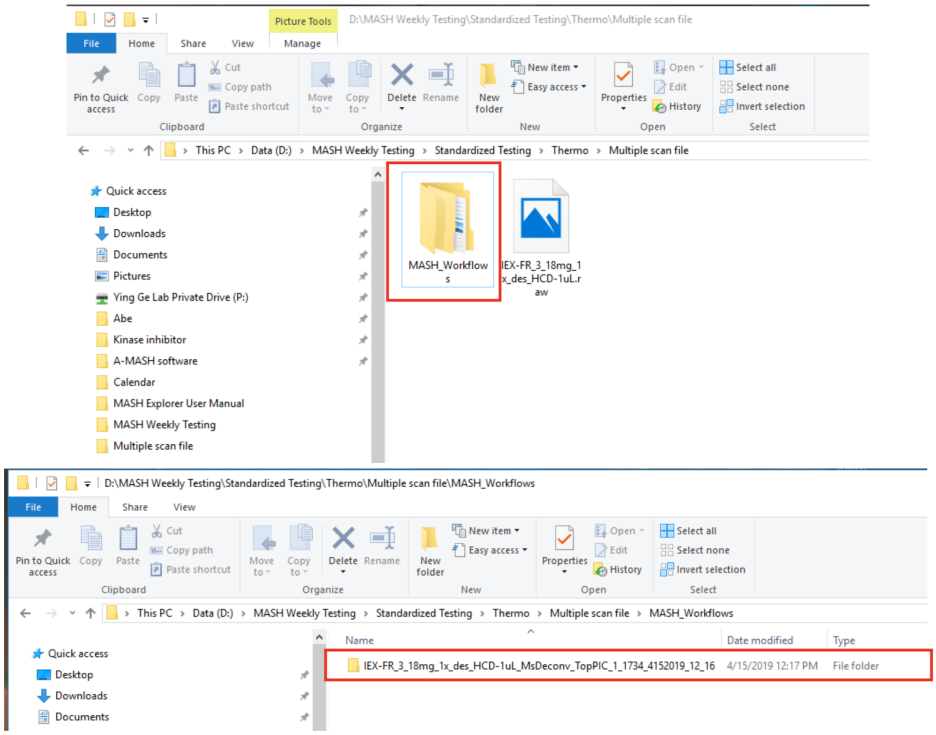


Figure 25: Folder created for data processing

7.1 Workflow Folder Contents and Cleanup

Inside the workflow folder are several files that are generated during processing, including:

- `Workflow_Parameters.txt` – Parameters set for the workflow.
- `Workflow_Timings.tsv` – Record of the time elapsed for each task.
- `Workflow_Log.txt` – Log file of the workflow.

The remaining files are results from the tools selected to be run within the workflow.

To save disk space, as of MASH Explorer V2.1 (and in all versions of MASH Native), temporary files in the workflow folder (such as the copy of the experiment file) are removed when a workflow finishes su ccessfully. (If the workflow does not finish successfully, the temporary files are left in place to aid d ebugging.) 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 **22**).

7.2 Java Heap Space

As of MASH Explorer V2.2.2 (and in all versions of MASH Native), you have the option of setting the Java maximum heap size for workflow tasks (see Figure **26**). Of course, this is only applicable to workflow tasks that run under Java (currently, MSDeconv and MSAlign+). The default maximum is 12 GB.

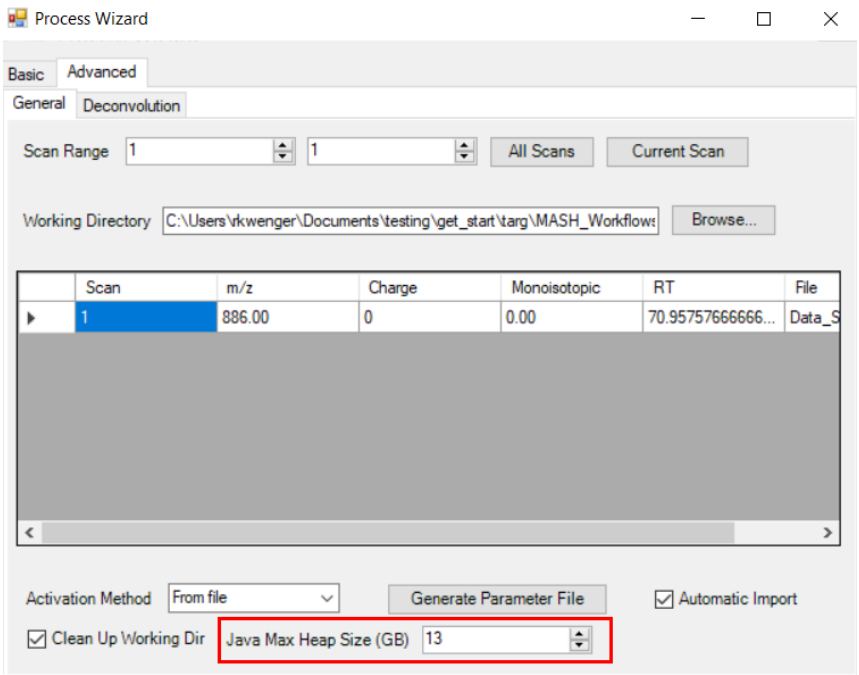


Figure 26: Java heap size setting in Process Wizard

7.3 Workflow Manager

When the deconvolution/search workflow starts, the Workflow Manager will be opened (Figure **27**).

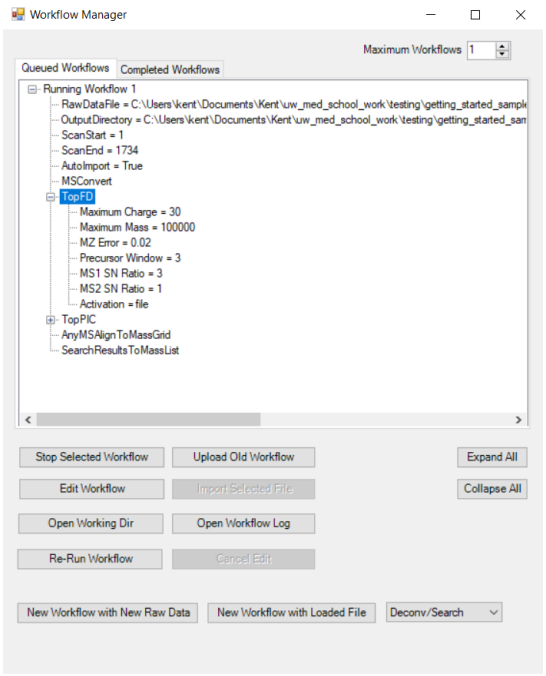


Figure 27: Interface for Workflow Manager (workflow running)

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 **28**). You can also upload an old workflow by dragging and dropping the parameter file (`Workflow_Parameters.txt`) on the MASH Native main window (not on the Work-

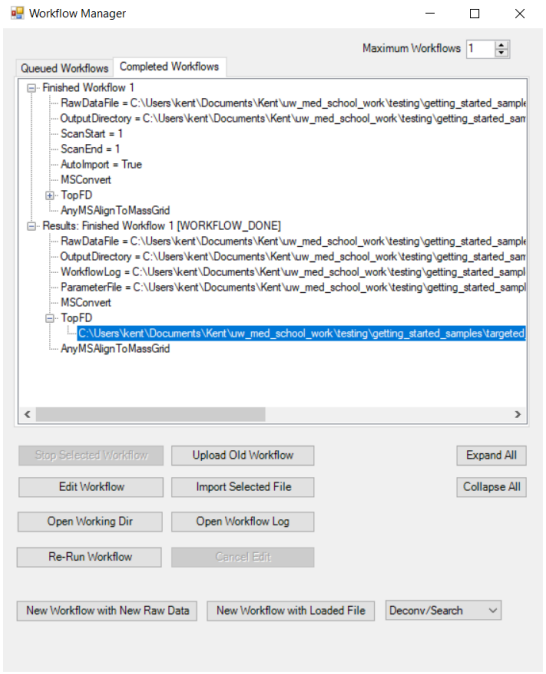


Figure 28: Interface for Workflow Manager (workflow completed)

flow Manager).

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

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.

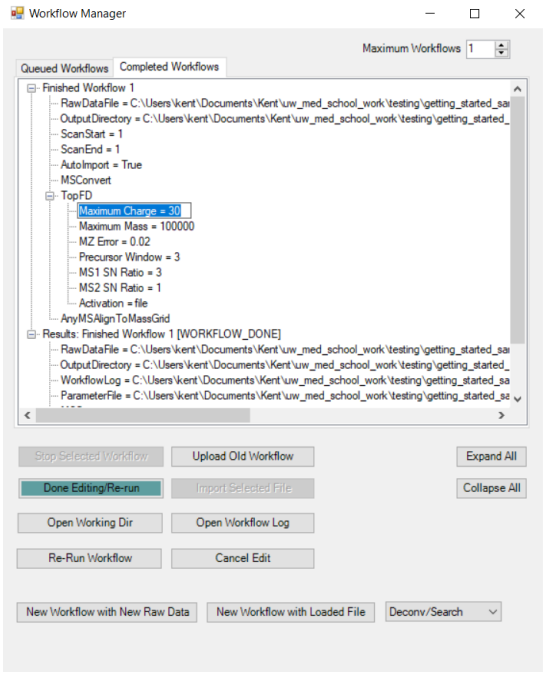


Figure 29: Interface for Workflow Manager (editing workflow)

- *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 decide you do not want to create and run a new workflow.

Available if a queued or running workflow is selected (see Figure **29**):

- *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 **30**):

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

Available if a workflow is being edited (see Figure 29):

- *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 Native 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 **30**).

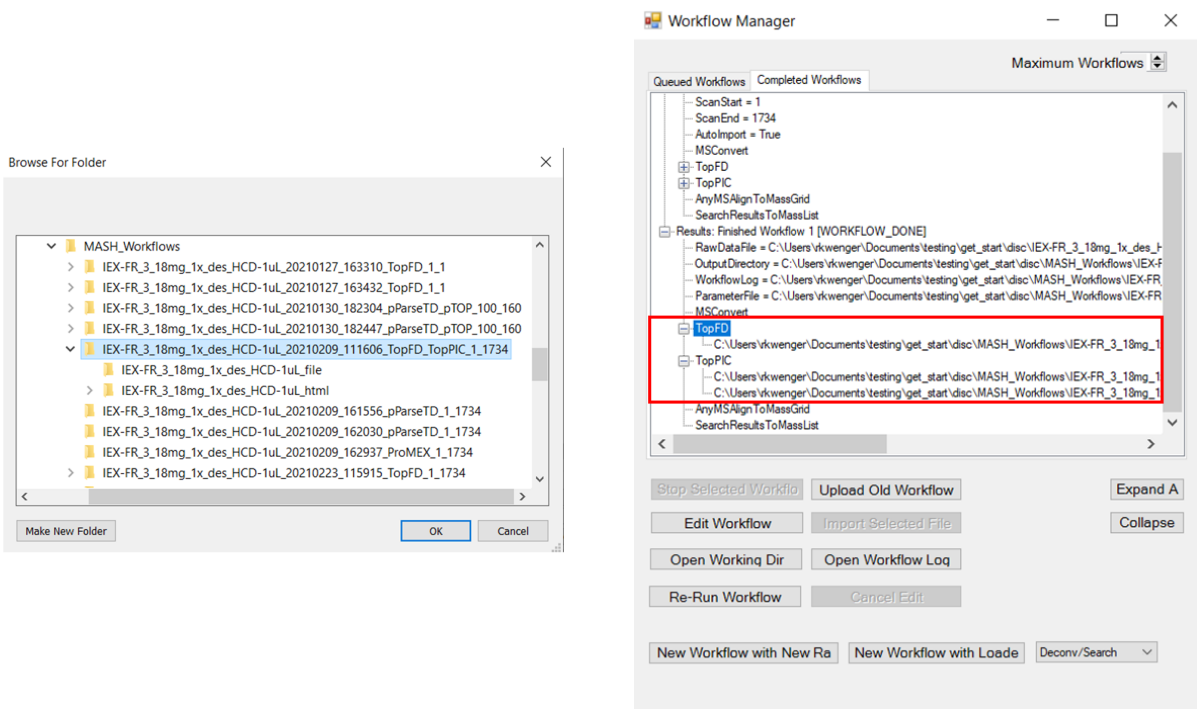


Figure 30: Uploading completed workflows and importing data

8 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.

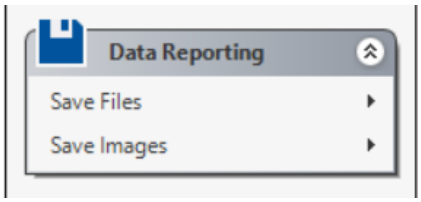


Figure 31: Data Reporting button layout

8.1 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 **31**). 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.

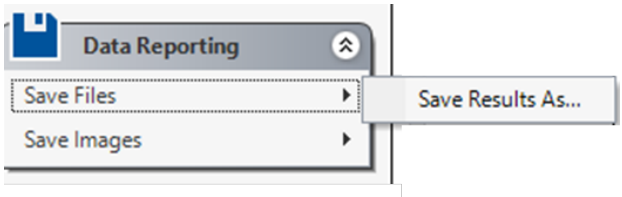


Figure 32: Save Files

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.

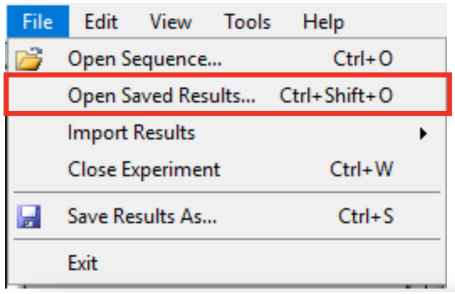


Figure 33: Open Saved Results

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

8.2 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 34. 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.

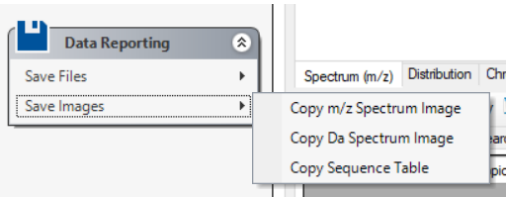


Figure 34: Menu to Save Images

In *Save Images, Copy Spectrum Image* will copy the spectrum in the *Results View* panel (Figure **35**).

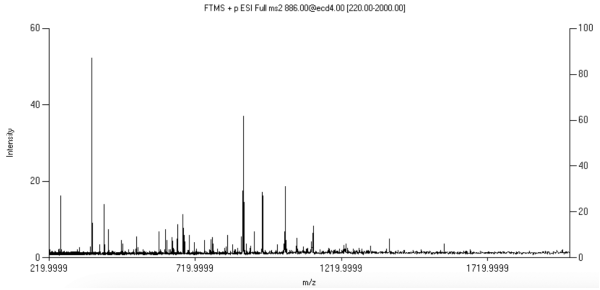


Figure 35: Spectrum Image

In *Save Images, Copy Sequence Table* will copy the spectrum in the *Sequence Table* panel (Figure **36**).

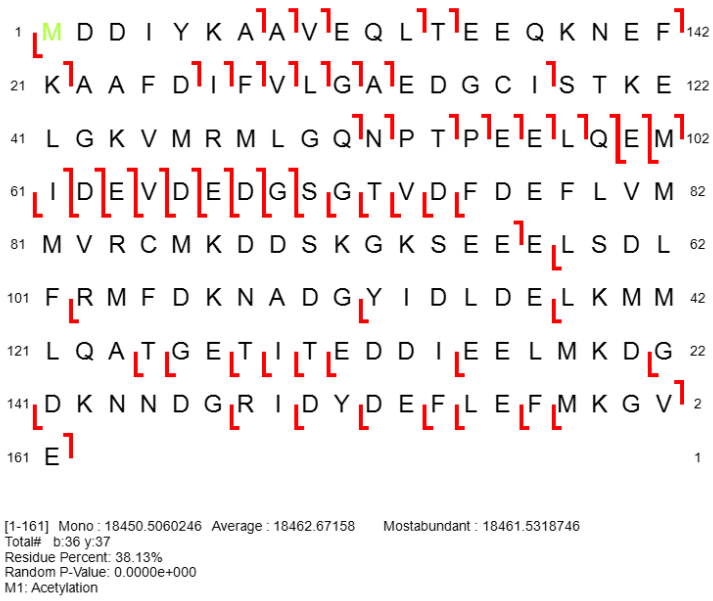


Figure 36: Sequence Table Image

9 Visual Validation of the Computational Output

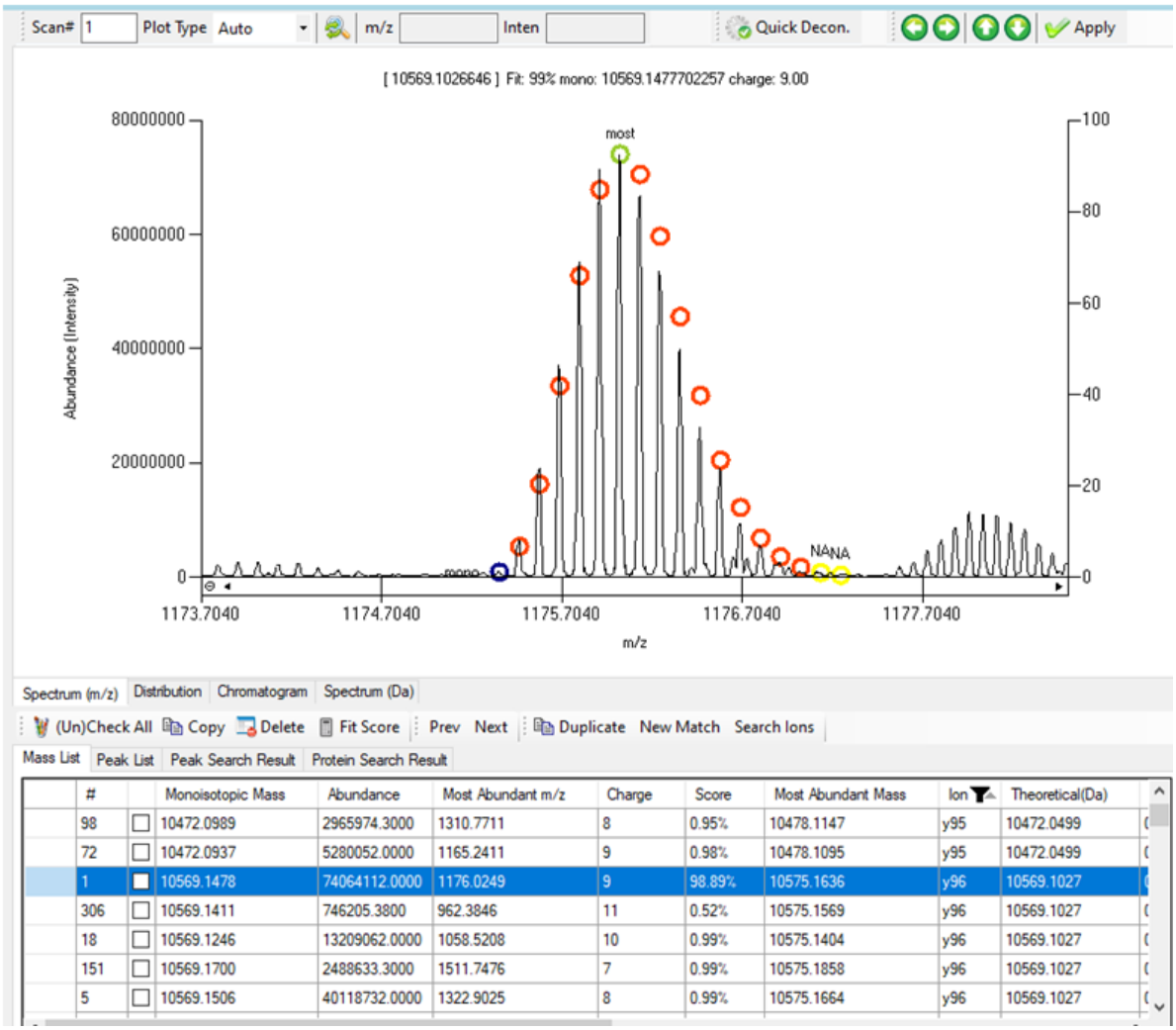
This section of the manual will introduce the function in *Mass List* and *Results View*, *Sequence Table* and *Ion Finder*. These tools are essential for the protein characterization using MASH Native software.

9.1 Mass List and Results View

9.1.1 Viewing and adjusting a Mass List entry

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*.

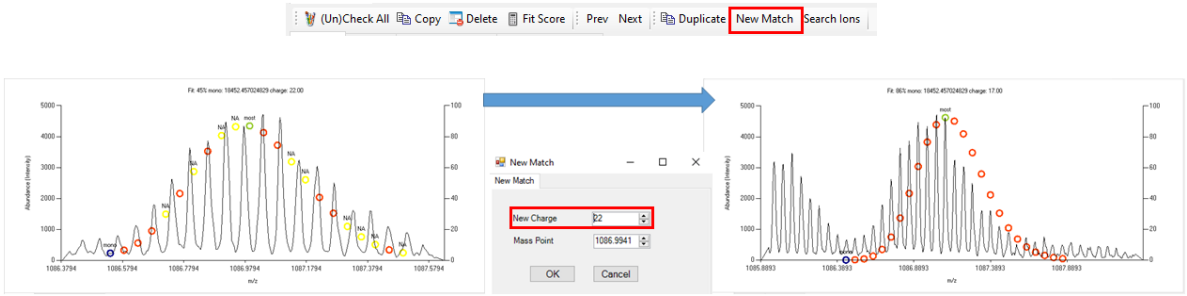


#		Monoisotopic Mass	Abundance	Most Abundant m/z	Charge	Score	Most Abundant Mass	Ion	Theoretical(Da)
98	☐	10472.0989	2965974.3000	1310.7711	8	0.95%	10478.1147	y95	10472.0499
72	☐	10472.0937	5280052.0000	1165.2411	9	0.98%	10478.1095	y95	10472.0499
1	☐	10569.1478	74064112.0000	1176.0249	9	98.89%	10575.1636	y96	10569.1027
306	☐	10569.1411	746205.3800	962.3846	11	0.52%	10575.1569	y96	10569.1027
18	☐	10569.1246	13209062.0000	1058.5208	10	0.99%	10575.1404	y96	10569.1027
151	☐	10569.1700	2488633.3000	1511.7476	7	0.99%	10575.1858	y96	10569.1027
5	☐	10569.1506	40118732.0000	1322.9025	8	0.99%	10575.1664	y96	10569.1027

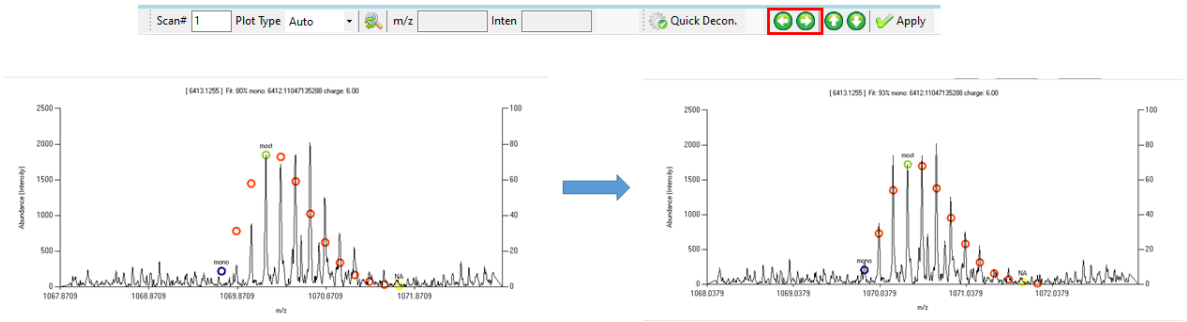
Figure 37: Result View and Mass List interactive display

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

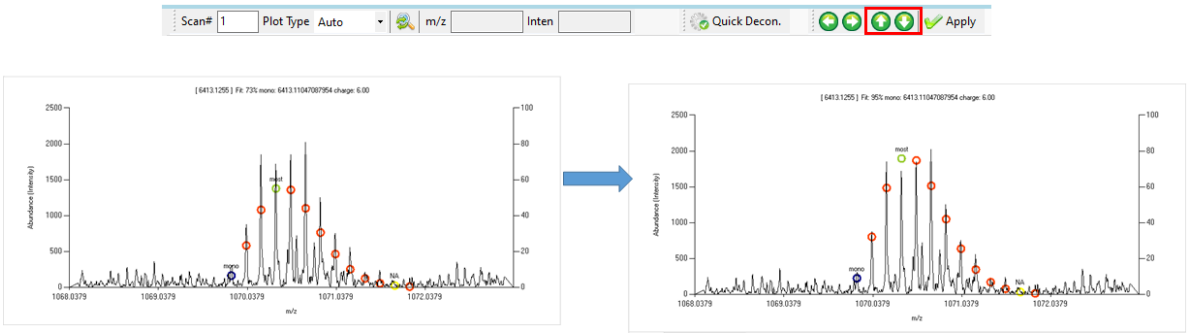
1. Adjust the charge state of the theoretical distribution



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 38).

Figure 38: *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 **39**). These buttons interact with the checked item(s) in the *Mass List*.

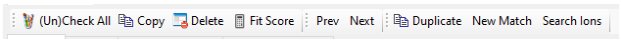


Figure 39: *Mass List* Tool Bar

- 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.

9.1.2 Isotopic Fitting Curve

When viewing an ion from the Mass List, you can superimpose the fitting curve corresponding to the theoretical distribution on the actual spectrum (see figure **40**). To show the isotopic fitting curve, select **View > Isotopic Fitting Curve** in the main toolbar.

If you want to change the color of the isotopic fitting curve, you can do so in the Parameter tab of the Workflow and Parameters panel (see figure **41**).

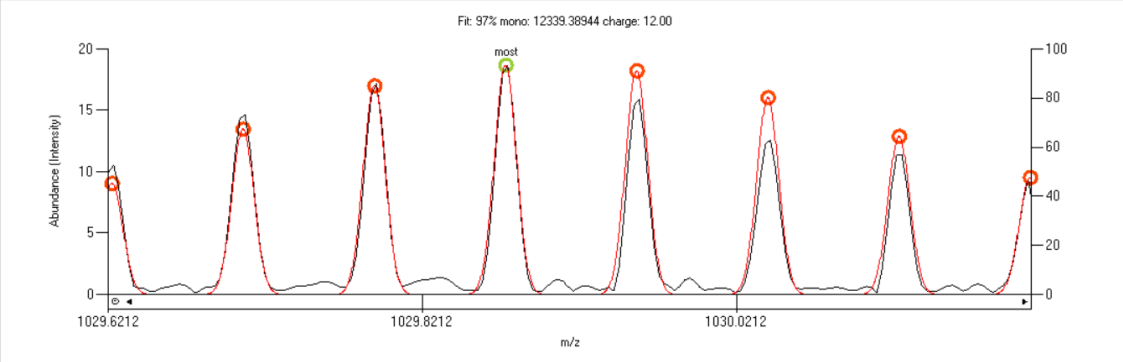


Figure 40: Isotopic fitting curve

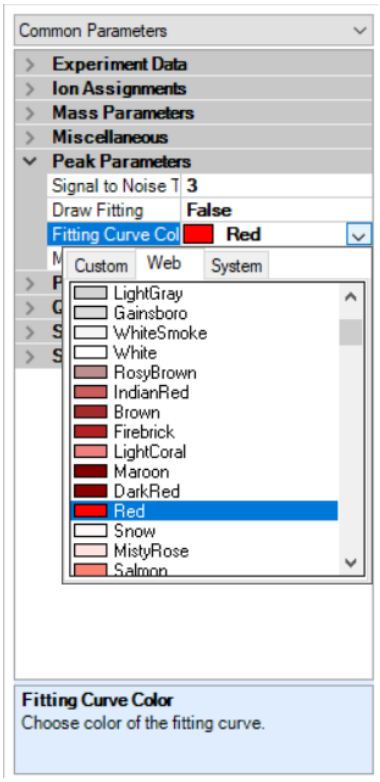


Figure 41: Color selection for isotopic fitting curve

9.1.3 Peak Annotations

When viewing an ion from the Mass List, you can also annotate the peaks with their m/z values (see figure **42**). To show the peak annotations, select **View > Peak Annotations**

in the main toolbar.

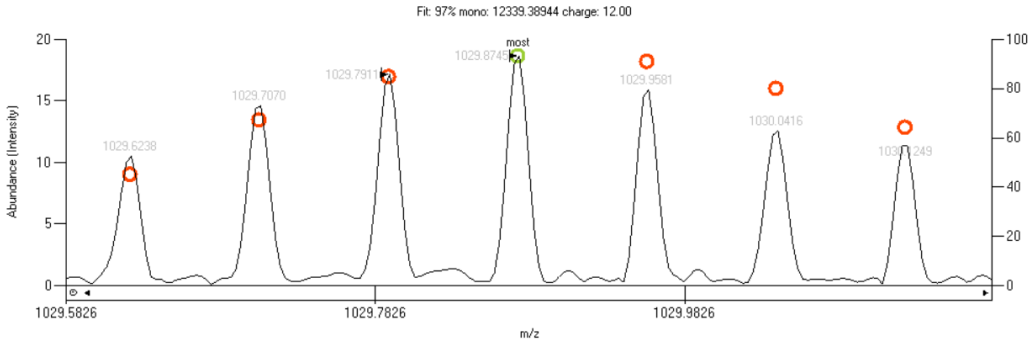


Figure 42: Peak annotations

9.2 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 **43**). The sequence can be imported via *Characterization* → *Paste Your Sequence* in Targeted Mode, or the *Modify Sequence* button in the *Sequence Table* panel (Figure **43**).

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*

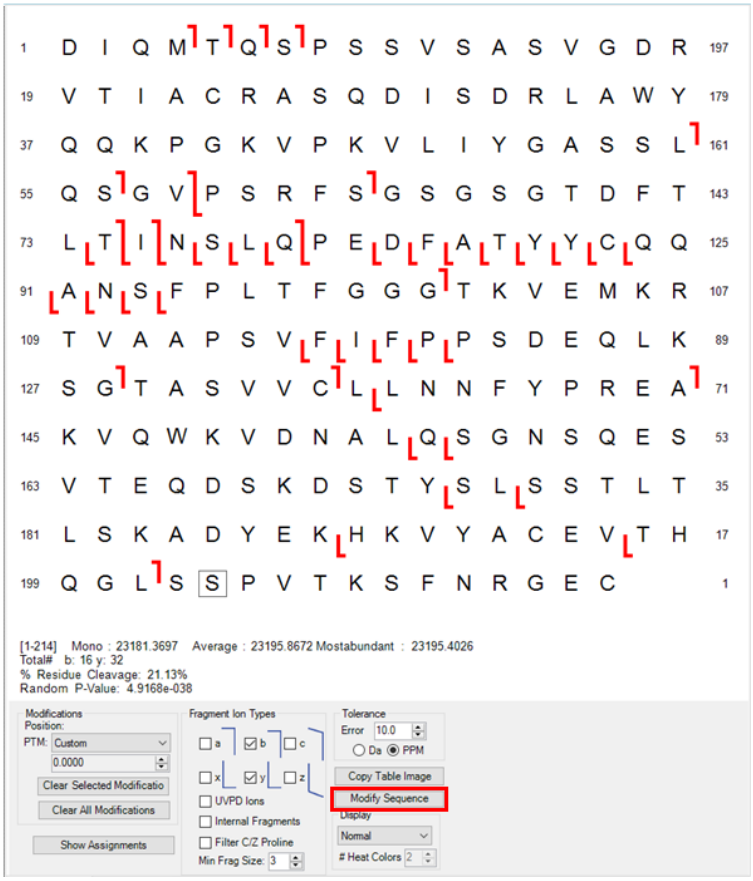


Figure 43: *Sequence Table* panel

There are preset modifications with mass changes such as phosphorylation (+79.96 Da) and methylation (+14.01 Da). You 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 **43**). Additionally, the current version supports common UVPD ions such as a+1, x+1, and y-1. The "Filter C/Z

Proline" checkbox allows users to filter out matches that contain c and z fragmentation at proline residues, which also includes the cz internal fragmentation matches (Section 9.2.1).

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 clicking Copy Table Image. The copied image can be pasted in either Photoshop or Microsoft PowerPoint for further processing.

9.2.1 Internal Fragment Matching

If the "Show Assignments" button is clicked when the "Internal Fragments" checkbox is selected, the internal fragments associated with the selected terminal fragments will be generated. For example, if A, X, and B ion checkboxes are selected, AX and BX internal ion theoretical masses will be generated and included in the matching process. Any matches will be annotated in the mass list. Please note that the ion generation and search generally takes a longer time to compute internal fragment matches than it takes to compute terminal fragment matches only.

In cases where a given observed mass matches multiple possible theoretical ions, MASH Native will choose the observed-theoretical pair based upon the following logic (similar to ClipsMS's biased mode). First, if there are any possible terminal ion matches, the match with the lowest mass error will be selected as the match. Next, if there are only matches to internal fragment ions, the internal fragment ion with the lowest mass error will be selected. The unselected matches are reported in the log book (Figure **44**).

The "Min Frag Size" parameter specifies the minimum number of amino acids for which an internal fragment will be generated to be searched for matches. The default

value is 3, and the maximum possible value for "Min Frag Size" is the length of the peptide sequence or 100, whichever is smaller.

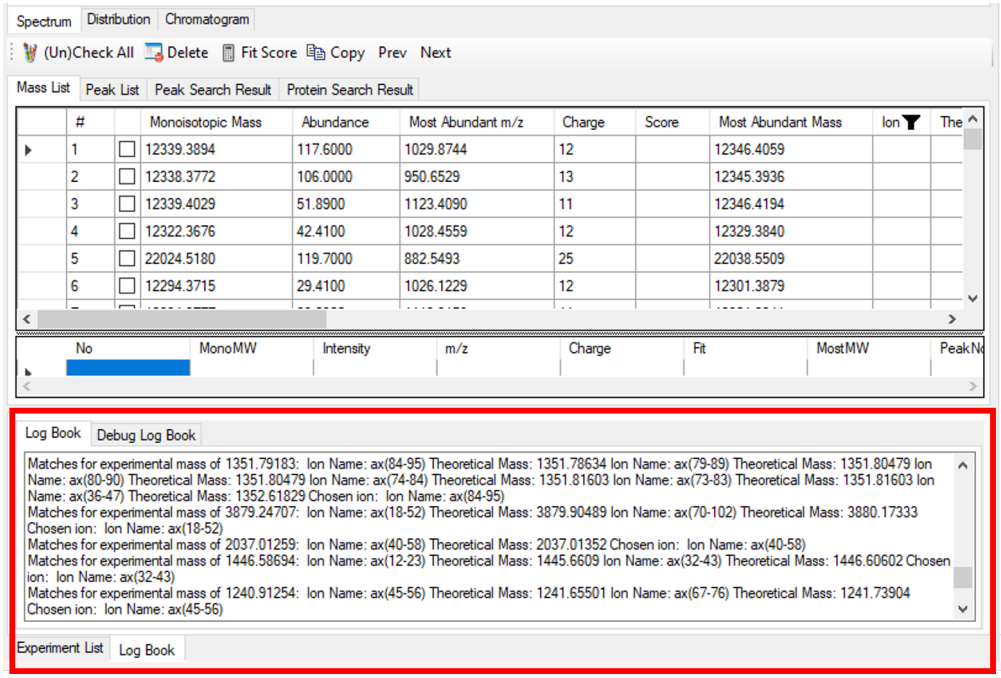


Figure 44: Example logbook entries for reported matches and unmatched theoretical ions

Normal and Heatmap Display Modes

The display mode (Figure **45**) allows the user to toggle between the "Normal" and "Heat Map" mode for annotating the ions on the sequence table. The "Heat Map" mode gives a visual representation of the relative number of theoretical ions matched to the observed ions within the mass list in which each amino acid is present (Figure **45**).

Setting the Number of Colors in Heatmap Mode

When the "Heatmap" display mode is selected, users can select the number of colors for the heatmap display using the enabled "# Heat Colors" parameter (Figure **46**). The

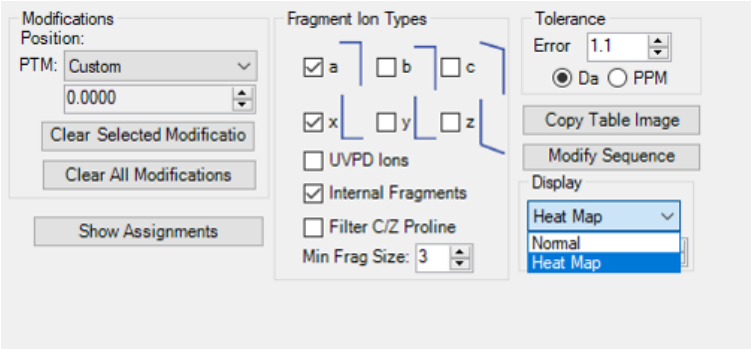


Figure 45: Display Mode Selection for Heatmap

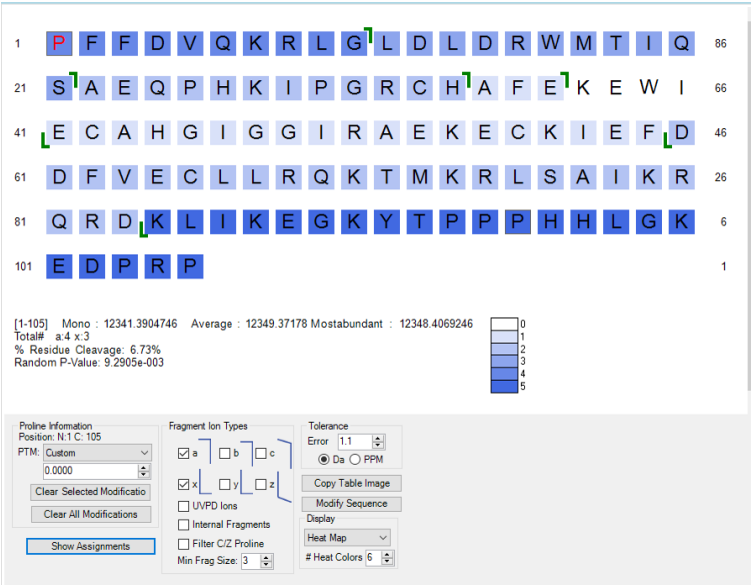


Figure 46: General overview of "Heat Map" display mode

minimum possible value for the number of colors on the heat map is two; the maximum possible value is equivalent to the maximum amino acid coverage. The colors as well as their corresponding coverage values are displayed in the key.

When the "# Heat Colors" parameter is equal the default value of two, the heat map sets the background of the amino acids that include at least one fragment match fragments in the mass list to blue, otherwise the background is set to white (Figure **47**). When "# Heat

Colors" is set to values greater than two, users can see the relative coverage of each amino acid. The heatmap key on the sequence table (Figure **48**) indicates what each of the colors equates to, and users can see the specific coverage of a selected amino acid (see amino acid coverage annotation section). Note that amino acids with one or more matches are always distinguished from those with zero matches.

The heat map display mode can be used without or with the found internal fragment matches (Figures 50 and **51** respectively). Note that the ion markers are always only for the terminal ion matches. Currently, internal ions are only displayed as part of the heatmap.

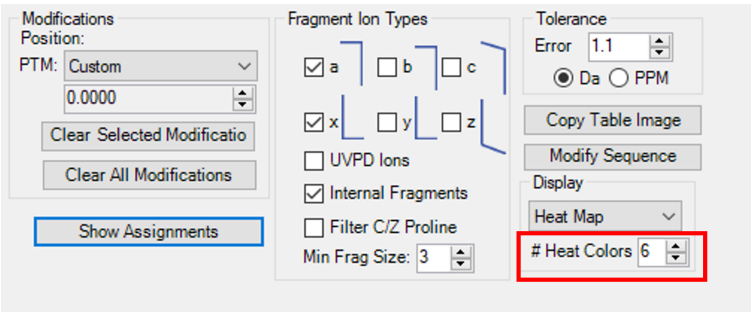


Figure 47: To change the number of colors included in the heat map, change the "# Heat Colors" value in the "Display" section.

9.3 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. Only the ion types selected in the sequence panel will have their theoretical ion types calculated after the "Show Assignments" button is clicked. Currently, only the terminal ion types are supported by the ion finder. Double-clicking a selected entry will open up the Ion Finder tool for the specific fragment number (Figure **52**).

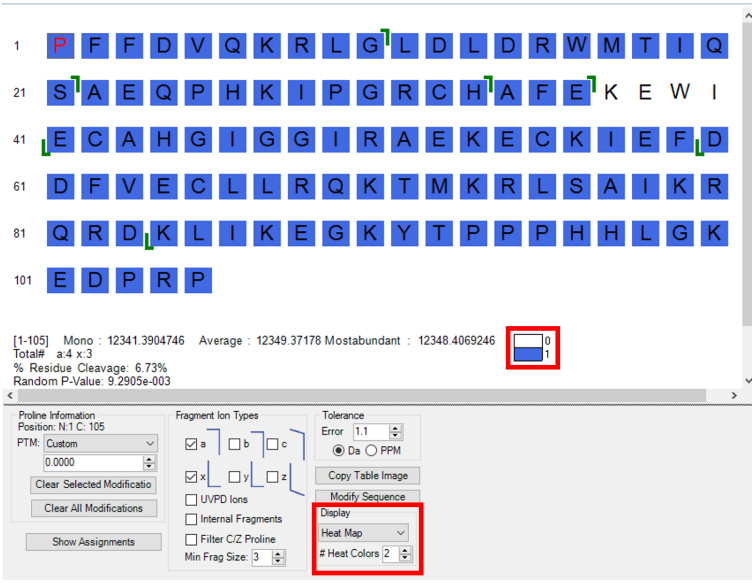


Figure 48: Example Two-color heatmap

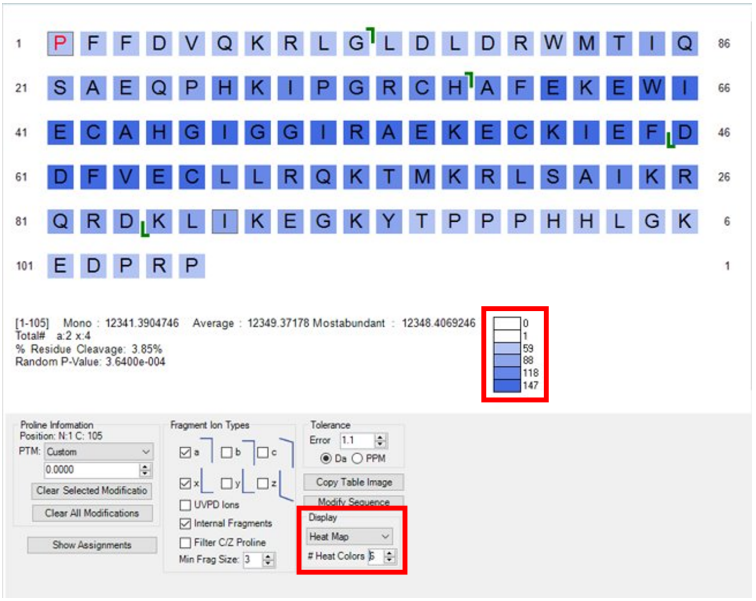


Figure 49: Example heatmap with more than 2 colors

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

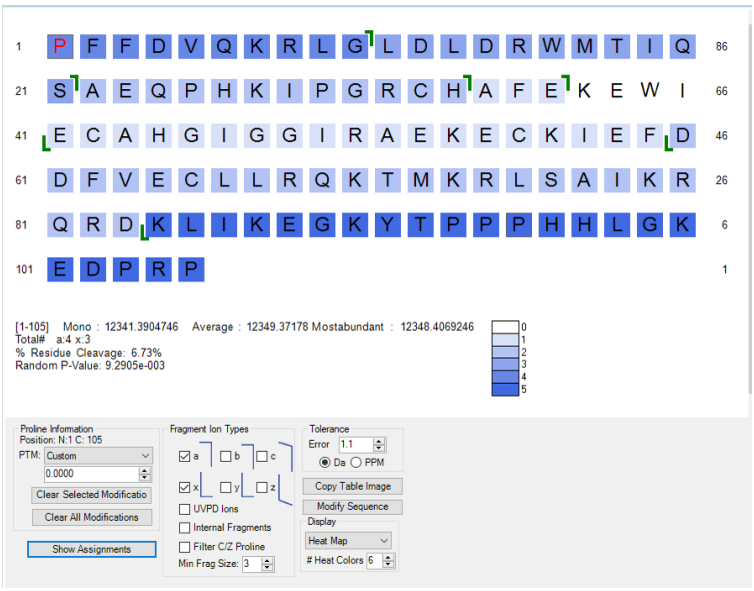


Figure 50: The heat map when only terminal fragments are selected. Note that the amino acids in the center of the peptide have relatively little coverage.

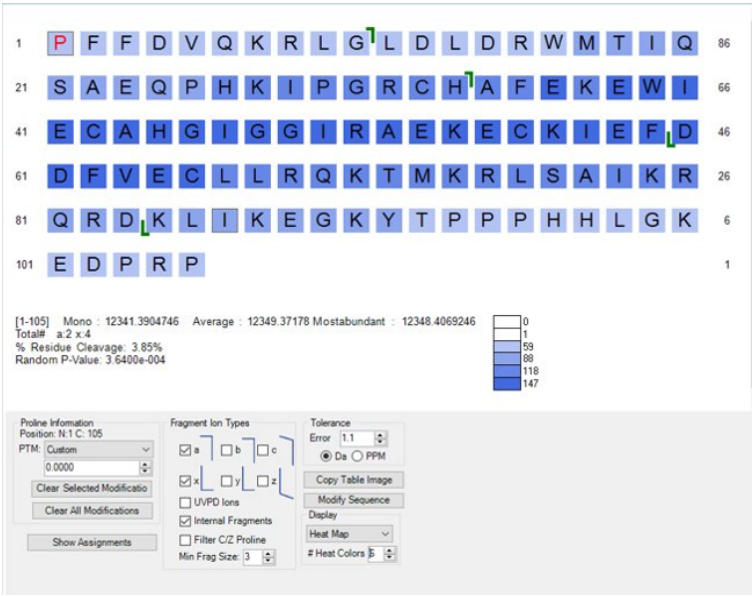


Figure 51: The heat map when both terminal fragments and internal fragments are selected for the same data set, deconvolution, and terminal ion selection as above. Note that the amino acids in the center of the peptide have relatively higher coverage, and the key values have changed to account for the increase in the number of matched fragments.

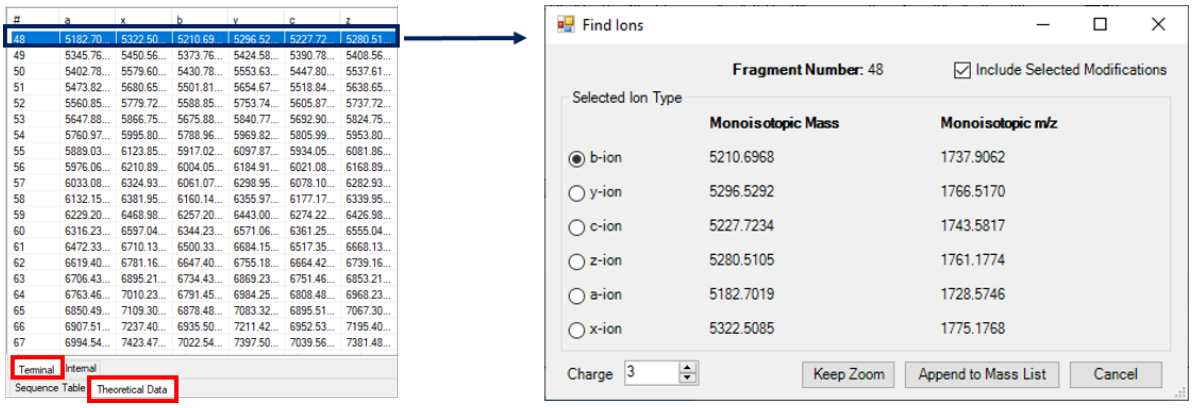


Figure 52: Ion Finder tool

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 **53**), 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 53).

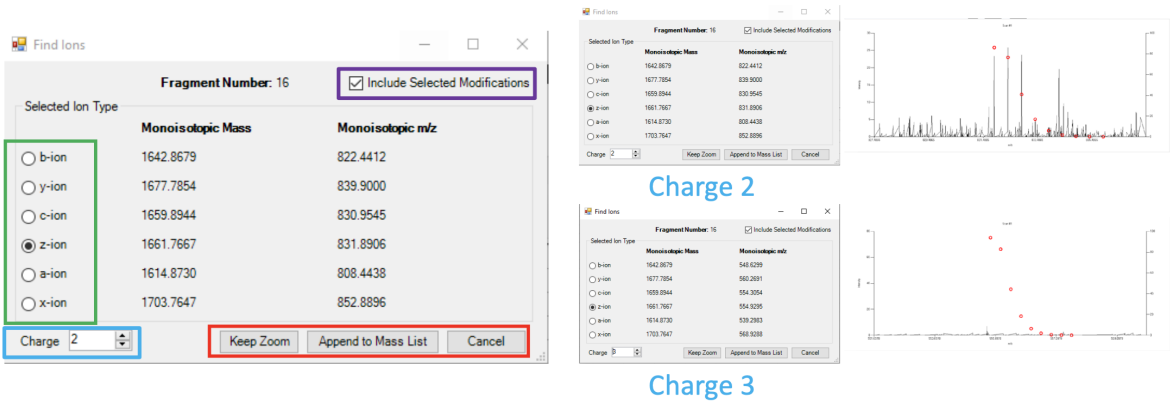


Figure 53: Functions for Ion Finder Tools

10 Tool Bar

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

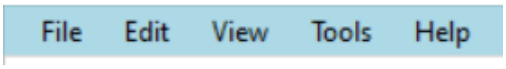


Figure 54: Tool Bar button layout

10.1 File

Items under File are Open Sequence, Open Saved Results, Import Results, Close Ex-periment, Save Results As, and Exit (Figure **55**).

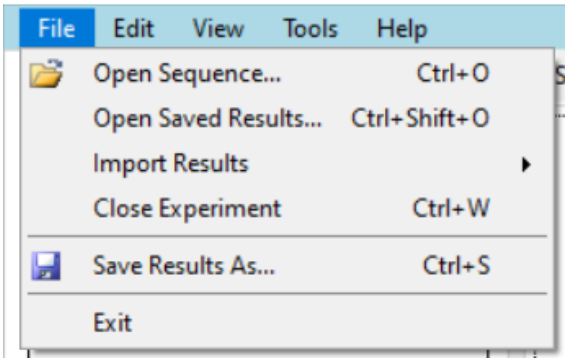


Figure 55: Menus for File option

- 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 Native 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.

10.2 Edit

Items under Edit are Copy Checked Mass(es), Copy Sequence Table, Copy Spectrum Image, and (Un)Check All Masses (Figure **56**).

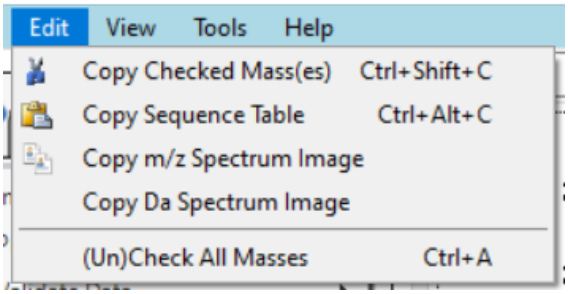


Figure 56: Menus for Edit option

- 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*.

10.3 View

Items under View are Spectrum Plot Type, Results View, MassList Toolbar, Status Bar, Sequence Table, Fitting Curve, and Fragment Labels (Figure **57**).

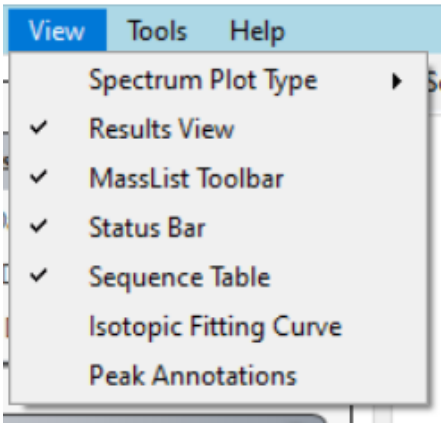


Figure 57: Menu for View option

For all items except Spectrum Plot Type, 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 7. 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.

Spectrum Plot Type allows you to select between Auto, Profile, and Centroid plot types. The default is Auto, which means to automatically detect whether the spectra have been centroided or not. Centroid mode shows the m/z spectrum as impulses, whereas Profile mode creates a connected line plot.

10.4 Tools

Items under Tools are Workflow Manager, Quick Decon. Parameters, Paste Sequence, Spectral Summing, Configuration, View Log File, and Open App Data (Figure **58**).

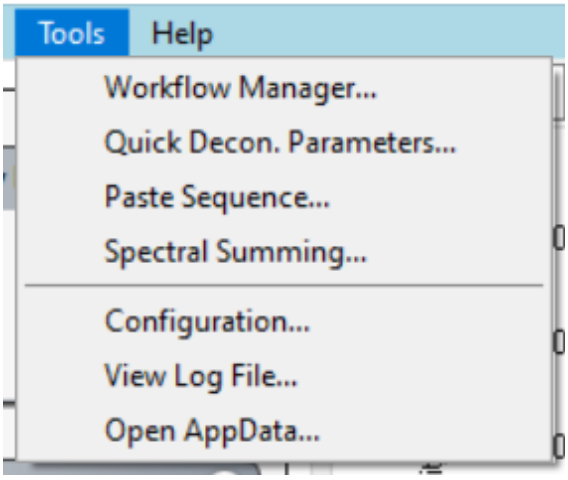


Figure 58: Menu for Tools option

- 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 Deconvolution 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.

- *Spectral Summing* See Section **5** for information on the spectral summing functionality.

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

  - *File Paths* tab (Figure **59**). For more information regarding the deconvolution and database search, please refer to the Search Algorithm Setup Guide included in the MASH Native 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 **60**). Users can also add new protein modifications or edit existing modifications (Figure **60**). 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 **61**). 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

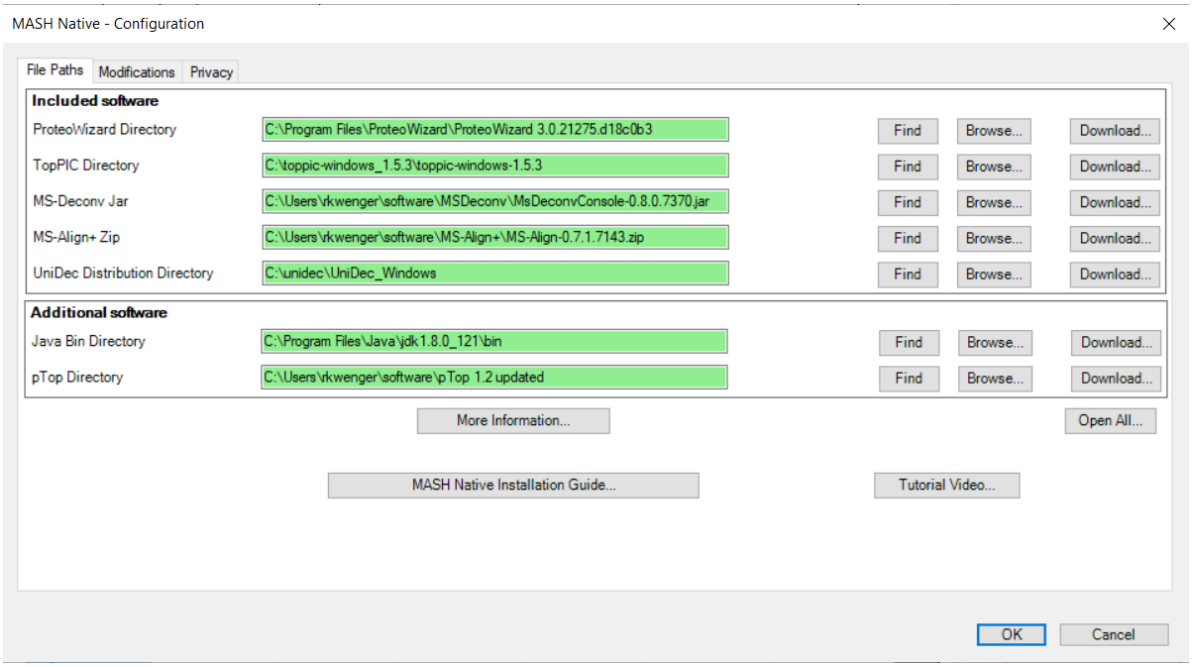


Figure 59: MASH Native - Configuration for supporting software

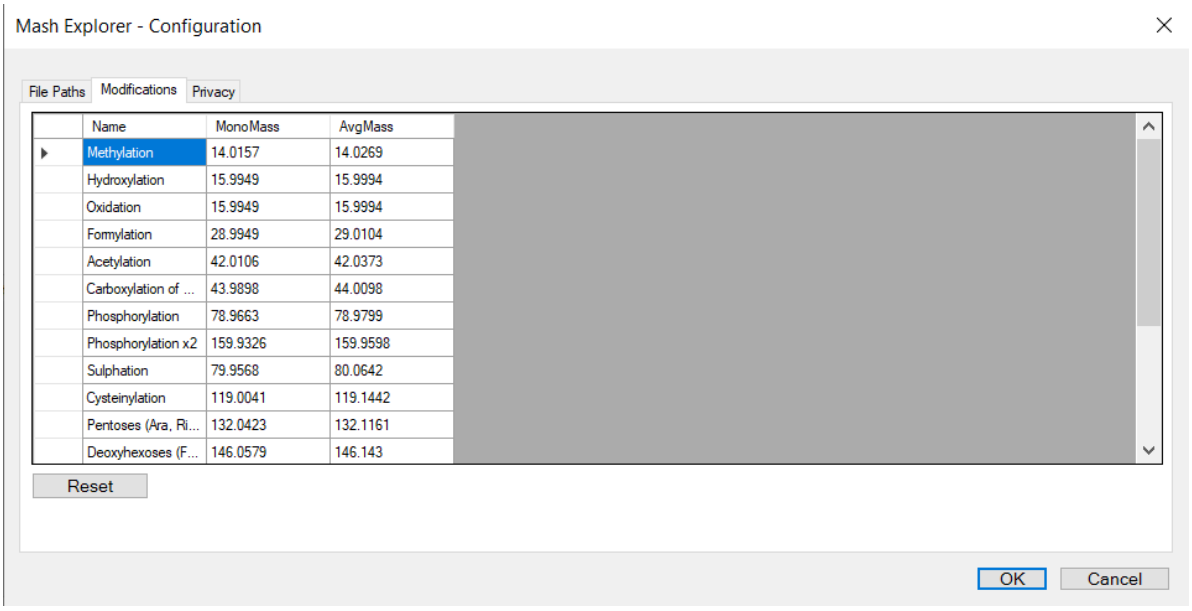


Figure 60: MASH Native - Configuration for modifications

Section **??** for details).

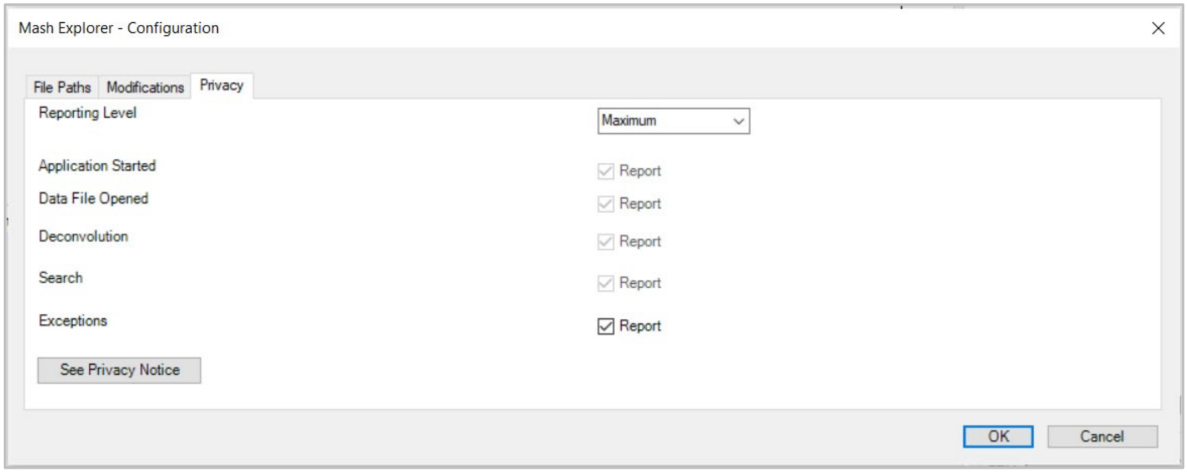


Figure 61: MASH Native - Configuration for privacy settings

- *View Log File* brings up a text editor window with the MASH Native log file.
- *Open AppData* opens the MASH Native AppData folder, which contains MASH Native configuration files and the MASH Native log file.

10.5 Help

Items under Help are Documents Folder, Getting Started, User Manual, License, Re-lease Notes, and About MASH Native (Figure **62**).

- *Documents Folder* opens the folder containing the MASH Native user documentation.
- *Tutorial Videos* opens the Ying Ge Research Group YouTube™ channel, which contains tutorial videos explaining a number of aspects of the MASH Native software.
- *Installation Guide* opens the MASH Native Installation Guide in your browser.

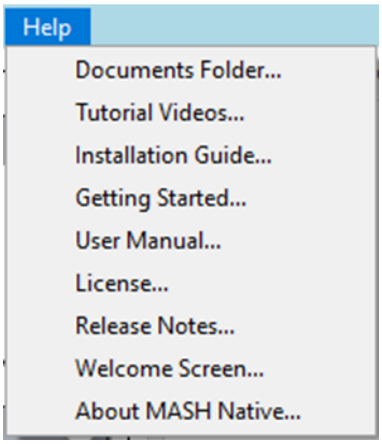


Figure 62: Menu for Help option

- *Getting Started* will open the MASH Native getting started guide: https://labs.wisc.edu/gelab/MASH_Explorer/doc/Native_GettingStarted.pdf, which introduces the basic functions in MASH Native software.
- *User Manual* will open the MASH Native user manual in Acrobat Reader (if possible) or in your web browser.
- *License* will open the MASH Native license agreement in Acrobat Reader (if possible) or in your web browser.
- *Release Notes* opens the release notes document. This provides information about the changes in each version of MASH Explorer and MASH Native.
- *Welcome Screen* shows the Welcome Screen described in section **2**.
- *About MASH Native* provides information about the MASH Native software, contact information, and references for the supporting software.

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

11 Parameters Tab

The parameters tab is in the bottom left corner of the MASH window (see figure **63**) and allows you to change sequence table and software parameters easily. These parameters can be accessed and changed at any time.

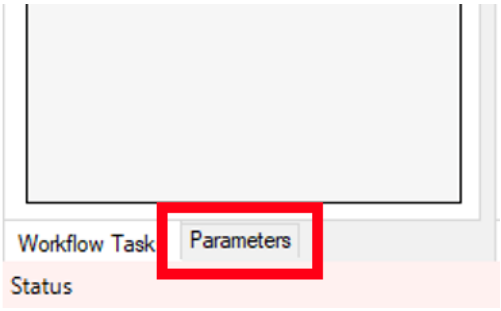


Figure 63: Parameters Tab Location

Within the tab, the Common Parameters section includes the Sequence Data and Sequence Table Parameters (see figure 64).

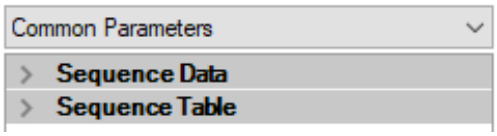


Figure 64: Common Parameters Categories

11.1 Sequence Data Parameters

The Sequence Data Parameters can be seen in the figure below

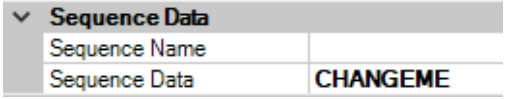


Figure 65: Sequence Data Parameters

- *Sequence Name:* Users can name their sequence.
- *Sequence Data:* Users can manually type in their sequence.

11.2 Sequence Table Parameters

The different Sequence Table Parameters can be seen in the figure below

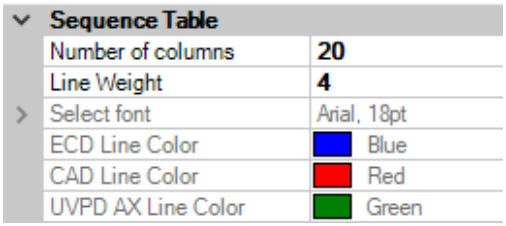


Figure 66: Sequence Table Parameters

- *Number of Columns:* Users can indicate the number of columns of amino acids in the sequence table.
- *Line Weight:* Users can indicate the weight of the fragmentation lines in the sequence table.
- *Select Font:* This shows the font and size of the sequence table. This parameter is unchangeable.
- *ECD Line Color:* This shows the color of ECD fragmentation lines in the sequence table. This parameter is unchangeable.
- *CAD Line Color:* This shows the color of CAD fragmentation lines in the sequence table. This parameter is unchangeable.
- *UVPD AX Line Color:* This shows the color of UVPD fragmentation lines in the sequence table. This parameter is unchangeable.

12 Configuration and Log Files

The MASH Native 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.

12.1 Configuration Files

When you first start MASH Native, 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 Native from running, because it is possible to produce an invalid configuration file if you manually edit it.)

Currently MASH Native uses the following configuration files:

- MASH_Explorer_FilePaths.txt
- MASH_Explorer_Modifications.xml
- MASH_Explorer_NormalDeconSearch.txt
- MASH_Explorer_Privacy.txt

- `MASH_Explorer_QuickDecon.txt`
- `MASH_Explorer_ShowDialogs.txt`

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

12.2 Log File

MASH Native will create a log file called `MASH_Explorer_LogBook.txt` when it runs. Each time you run MASH Native, 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 Native log file contains a record of the actions you perform each time you run the application.

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

```
--------------------------------
2021-11-19 12:09:31:  MASH Native Version 1.0.0.21105
```

This tells you when MASH Native 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 Native – 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 Native. 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 Native - 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.