AutoHX (HX-DIA) Workflow

Updated 2024-02-20

Installation:

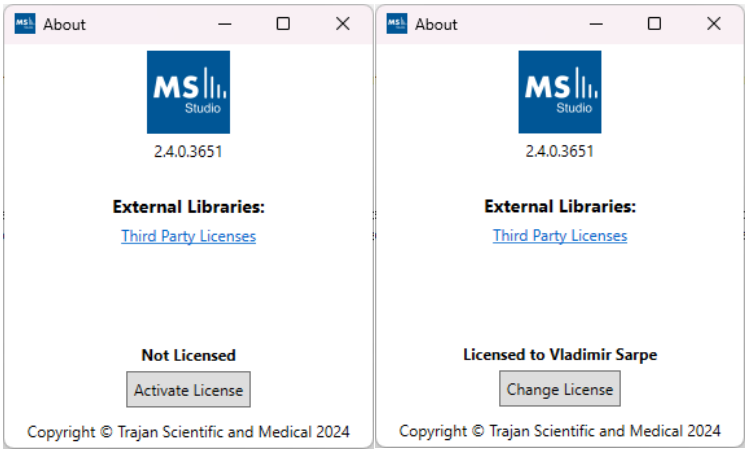
1. Install .NET Framework 3.5 SP1. For detailed instructions, visit this link.
2. [*Optional*] Install the latest version of Java 64-bit. This is required for some external tools (MS-GF+ and Percolator) used in HX-PIPE.
3. Run **setup.exe**. If you run the .msi directly, it will skip the check for additional system requirements. See below for more info.
4. During installation, you may be prompted to also install multiple "Visual C++ Redistributable 20XX" versions (2008, 2010, 2012, 2013, 2015+), if they are not already on your system. These pre-requisites are needed to make vendor software and other dependencies work. If you wish to install them manually, visit this link.

System requirements:

- Windows 10+ x64 with .NET 4.7.2 or higher.
- Multi-core CPU.
- 16GB+ RAM.
- 100GB+ of free disk space.

Enabling AutoHX Processing (Activate License):

A license is required to enable the full functionality behind AutoHX (HX-DIA). Please contact vsarpe@trajanscimed.com and/or kbachus@trajanscimed.com to obtain a license. Once you have a license, you can enable it from the "Main Menu -> Help -> About" and click the "Activate License" button.

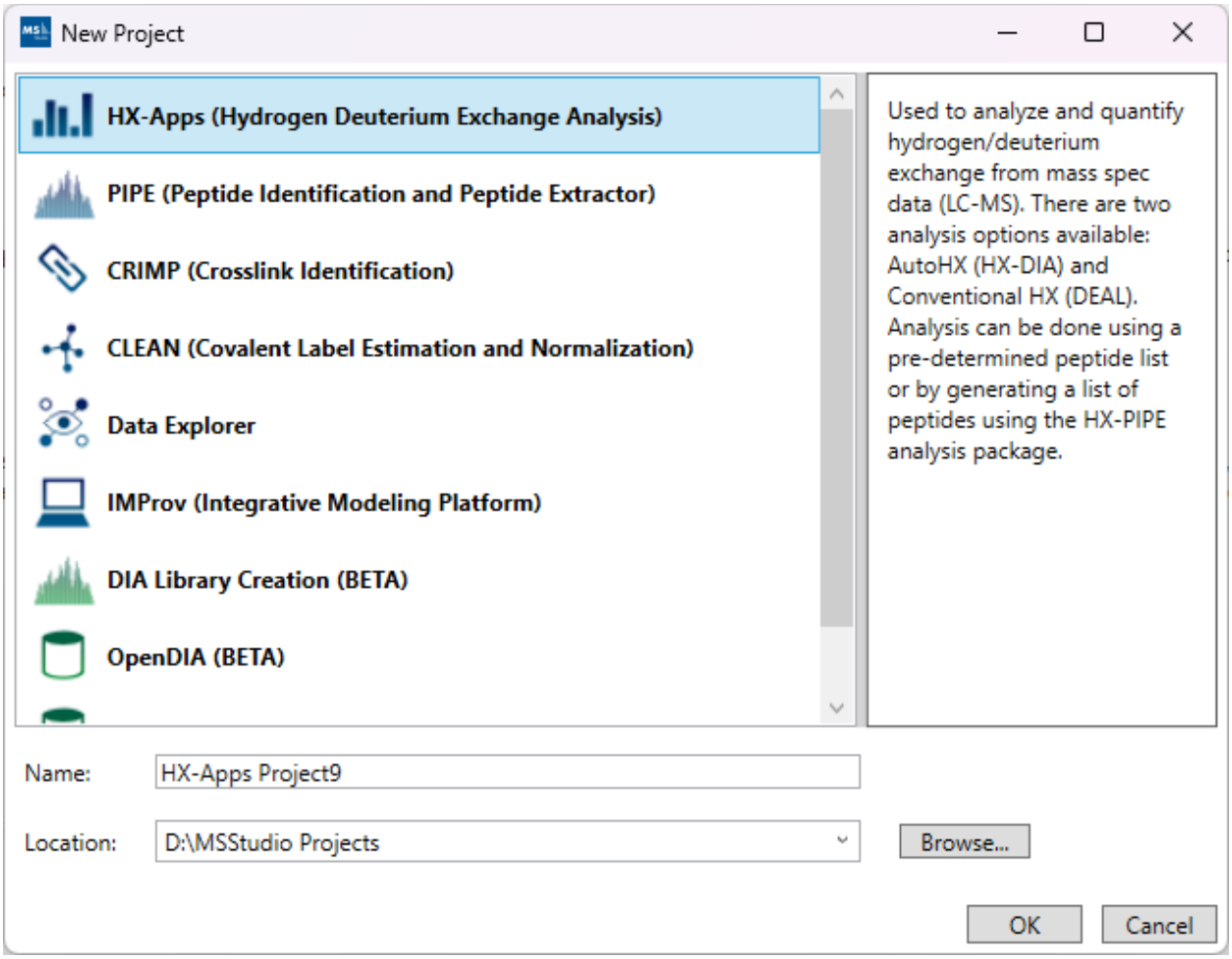


When prompted, enter the license text and click OK. If the license activates successfully, you will see a pop-up confirmation and the About page will say "Licensed to *<Your Name>*". Following the one-time activation, the full AutoHX option will be available in the Processing Routines area of the Processing window.

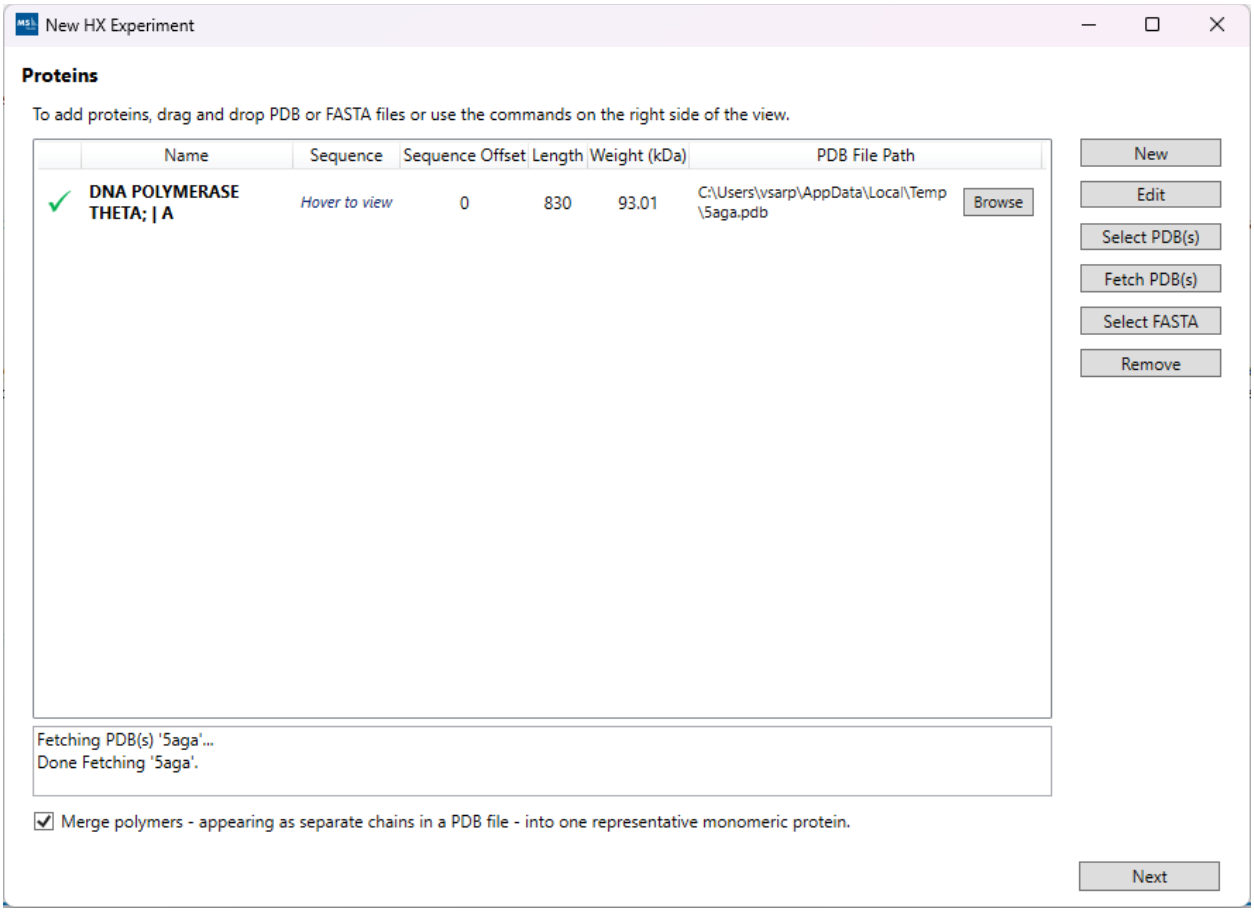
Create Project

Presently, AutoHX is a sub-module under the standard HX- Apps (formerly "HX-DEAL") workflow. Therefore, project creation for AutoHX in Mass Spec Studio follows the same overall wizard-based flow as regular HX-Apps. The major difference is that DIA files are required to use the new HX-DIA processing routine and DIA bins need to be supplied for all the runs. Please follow these steps to get started:

1. **"New Project":** Create a new HX-Apps project by selecting "New Project" and selecting "HX-Apps". The default location for new projects will be inside your home directory. You can select a different location by clicking "Browse" and selecting a new directory.

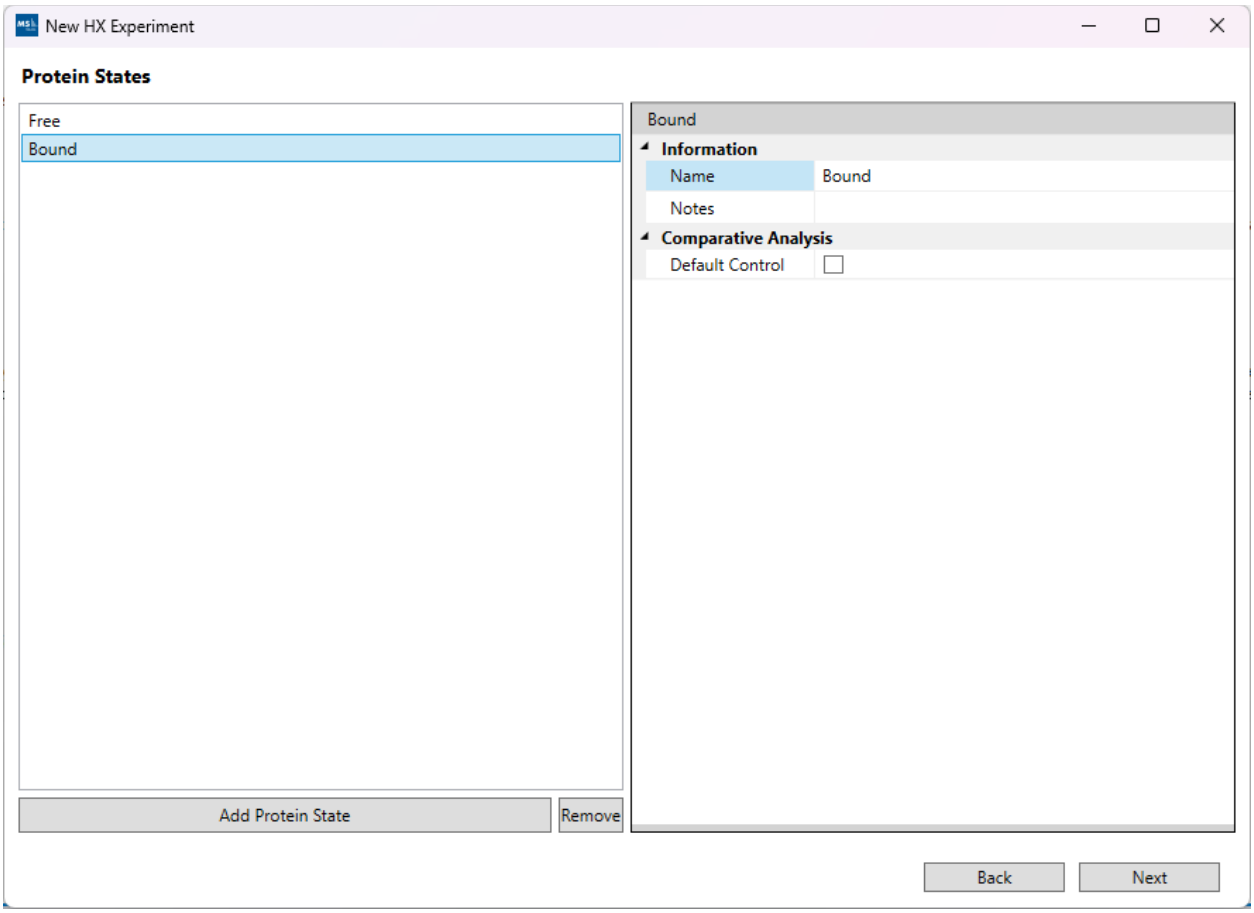


2. **“Proteins”:** If you have more than a single protein or wish to use sequence visualizations, you can add your protein sequences either manually (“New”), via a FASTA file (“Select FASTA”), via a PDB file (“Select PDB(s)”), or via a 4-letter PDB code (“Fetch PDB(s)”, example: “1jff”). Each protein will be listed as a separate row. If you have both PDB and FASTA, you can add the FASTA file first and then use the “Browse” button to link the PDB file. If you add both the PDB file and the FASTA file separately, you may end up with duplicate proteins in the table. **Important:** For the sequence visualizations to work, the names of the proteins must match those found in the Peptide .csv file (explained later).
The “Merge polymers” option can be used to remove duplicate sequences from multimeric protein complexes which may appear as separate chains inside a PDB file. Example PDB-code: “1fu1”.

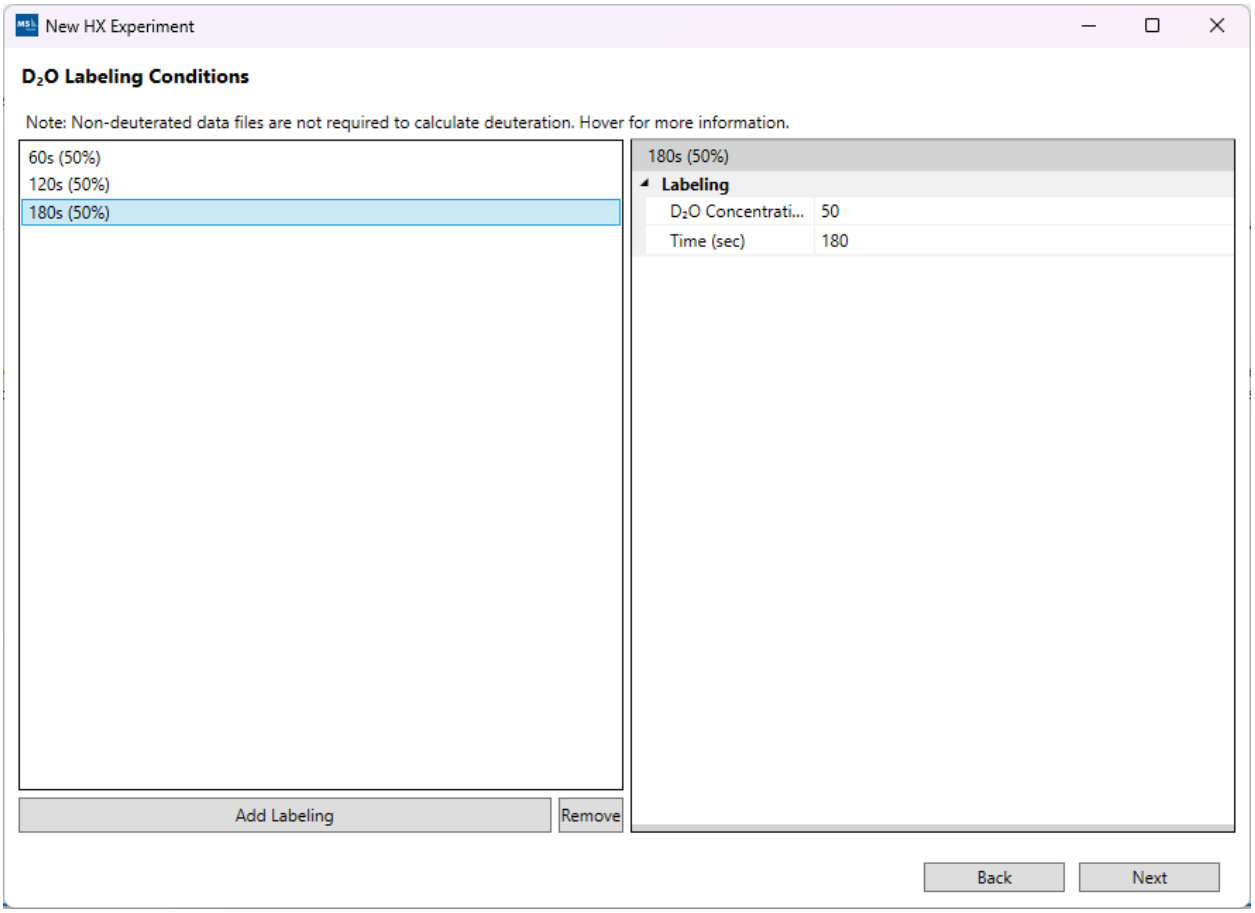


3. **“Protein States”:** For simple analysis, at least 1 protein state is required. For the additional comparative analysis tools and visualizations to become available, you must supply at least 2 proteins states. You can set the “Default Control” checkbox ON for the state you wish to be your control state – this will make it so all views that have a comparative option are pre-set with the default control. If none of the states are “Default Control”, then the most views will show one state in isolation (non-comparative).

Important: *For the current version, if you wish to use create a multi-target project (>2 states), you should turn OFF the “Fragment Model Refinement ->**Generate Optimal Fragment Model**” advanced parameter in the processing dialog. We will reiterate this recommendation with more details later in the **Processing** section of this document.*



4. **"D2O Labeling Conditions":** At least 1 labeling condition is required (both time and %D_2O). For kinetics visualizations to become available, you must supply at least 2 timepoints.



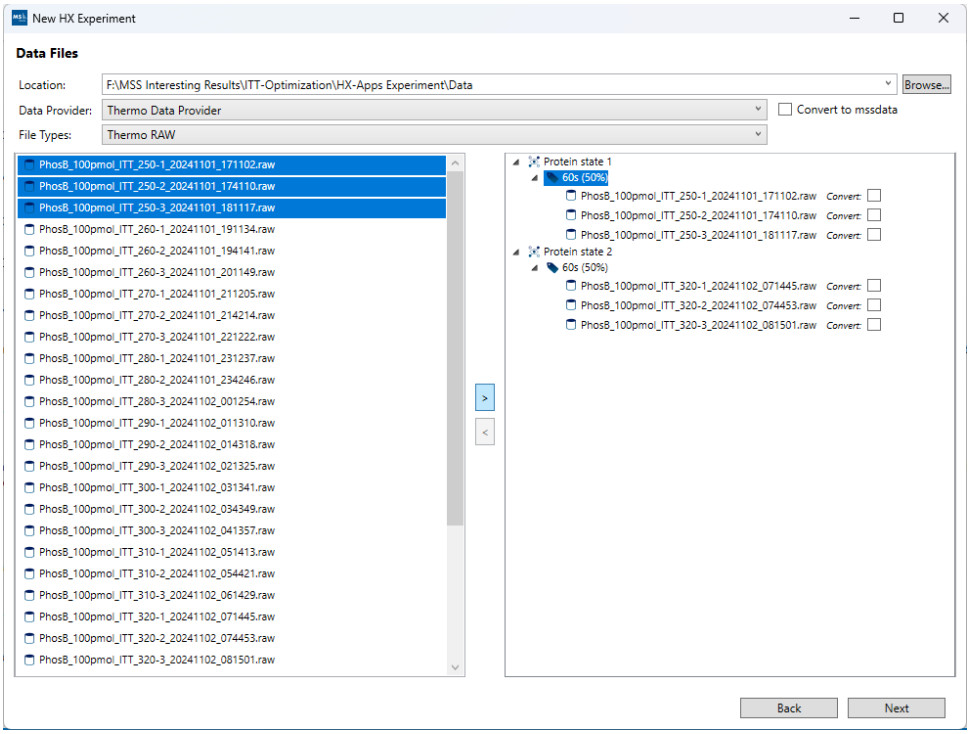
5. **“Data Files”:** Follow these steps to add your data:

a. Click “Browse” and select the root folder which contains your raw data files. In cases where the data files themselves are directories (example: Waters, Bruker), please make sure the containing folder is selected not the top-level .raw or .d directory itself. Note: Not all files have to be located under the same root folder from the start. After you add some files, you can still “Browse” to a different root directory and select additional files from the new location.

b. For Thermo data, select the Thermo Data Provider, which does not require conversion to Mass Spec Studio data format. For other raw vendor files (Waters, Bruker, Sciex, Agilent), we recommend using ProteoWizard Data Provider. If you already have Mass Spec Studio pre-converted files (.mssdata, .mssmeta), select “Mass Spec Studio Data Provider”.

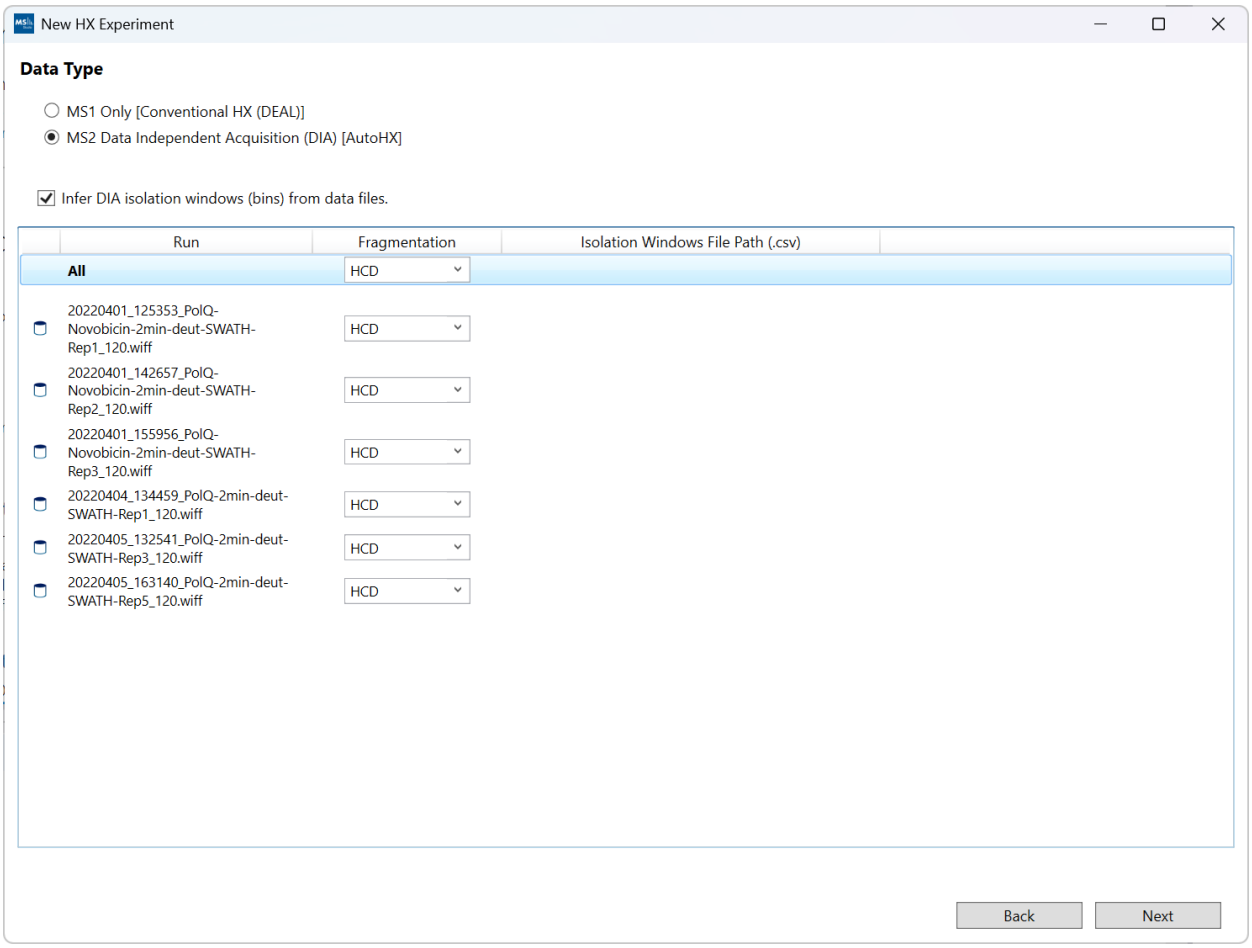
c. If you are not using Thermo data, the “Convert to mssdata” checkbox will the ON by default. When enabled, raw vendor files will be converted to the .mssdata format which enables super-fast searching at the cost of additional disk space for the .mssdata files. You can still proceed without converting the files, but the processing will be very slow. The “Noise filter” value is a multiplier of the minimum signal in each spectrum. A noise filter of “2” will remove any intensities smaller than 2 * minimum. A value of “0” will not remove any low-intensity data. Again, converting files is not required nor recommended for Thermo data.

d. Select your replicates (shift+click or ctrl+click for multi-select), select the appropriate Protein State/Labeling node and click the “>” button. If you make mistakes, you can remove runs using the “<” button.



6. **"Data Type":** For DIA files, select the "MS2 Data Independent Acquisition (DIA) [AutoHX]" option. This will present a table with a pre-selected fragmentation mode and a blank Isolation Windows File Path column. AutoHX only supports HCD and CID fragmentation enabling full hydrogen scrambling. To proceed, you will have to decide on how to define the DIA isolation windows. There are two options:

Option 1: [Default] Enable the "Infer DIA isolation windows (bins) from data files" checkbox. This will make it so AutoHX will attempt to automatically extract the DIA windows from each data file.



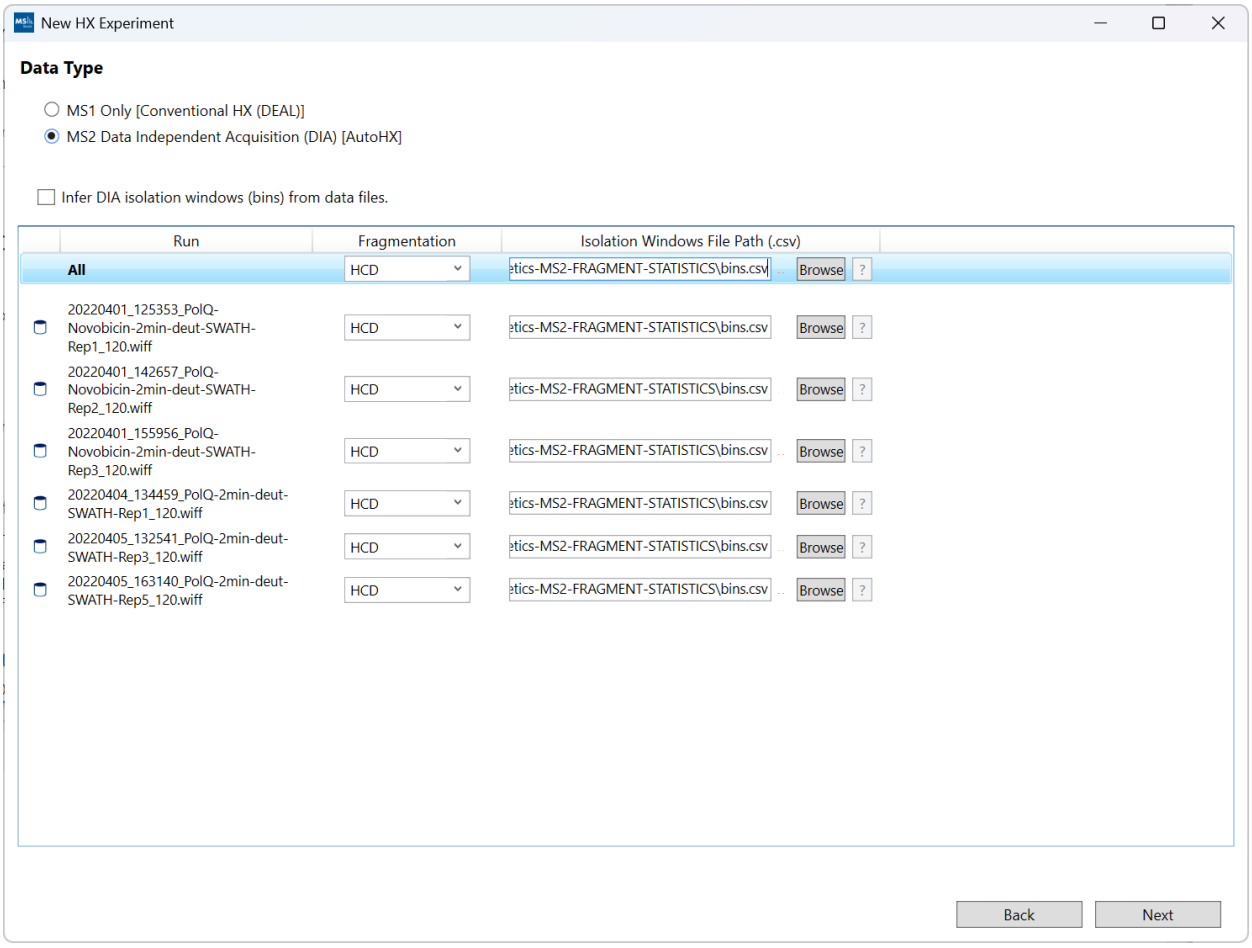
Option 2: Disable the "Infer DIA isolation windows (bins) from data files" checkbox and supply your own DIA isolation windows in .csv format.

The isolation windows should be in ".csv" format with "Start" and "End" columns which represent the boundaries of each DIA isolation window. If all runs have the same isolation windows, the file can be supplied once under the first row representing "All" runs. For example:

Start	End
100	150
145	195

190	240
…	…

Note: Direct .csv export of the DIA windows form Xcalibur works as well in the single-column “m/z Ranges” format where each row is a range: “500-520”.

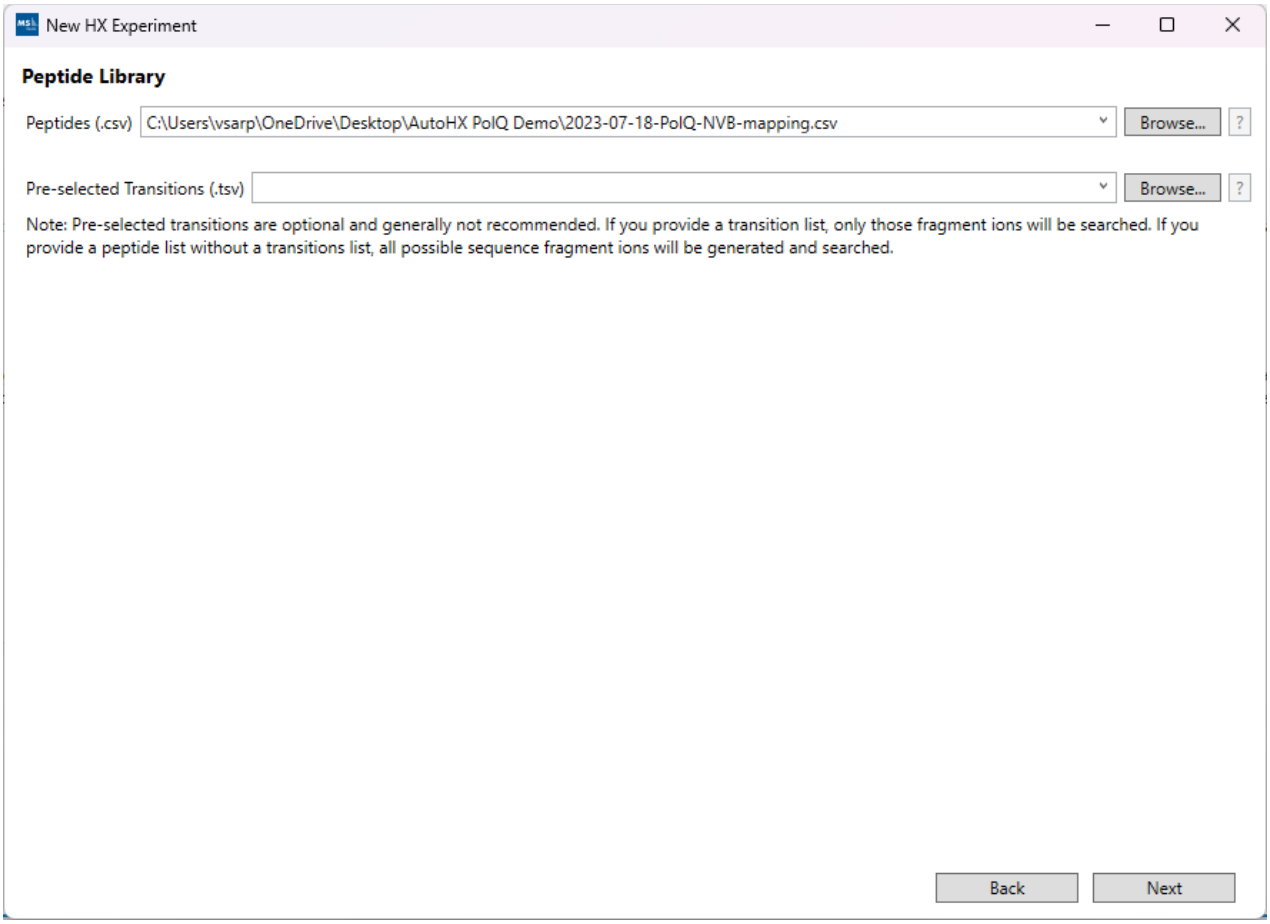


7. **“Peptide Library”:** Click “Browse” and select a peptide list. Note: To avoid duplicates, it’s recommended that your peptide identifications are grouped by sequence and charge such that each peptide (sequence/charge pair) has only one "RT". If you start with multiple separate peptide-spectrum matches (PSMs) for the same peptide, you can group them by sequence/charge and average their RTs to get the final peptide "RT". If you expect high discrepancies between the mapping runs and the DIA runs, you can define a larger “RT Variance” column in your peptides file. HX-DIA does not need a pre-selected transitions file because it will check all possible fragments for each peptide. However, if you wish to constrain the deuteration calculations to a pre-selected list of fragments, please supply them in the pre-selected transitions file.

Important: The name of your protein assigned to each peptide in the .csv must match one of the proteins in the initial protein selection window (where you select a FASTA and/or a PDB). The start/stop values of your peptide list must also match up with the FASTA protein sequence.

A sample peptide list: https://s3.us-west-2.amazonaws.com/www.msstudio.ca/assets/tutorials/hx-deal/sample_peptides.csv

[Optional - Not Recommended] A sample transitions list: https://s3.us-west-2.amazonaws.com/www.msstudio.ca/assets/tutorials/hx-deal/sample_transitions.tsv



Processing

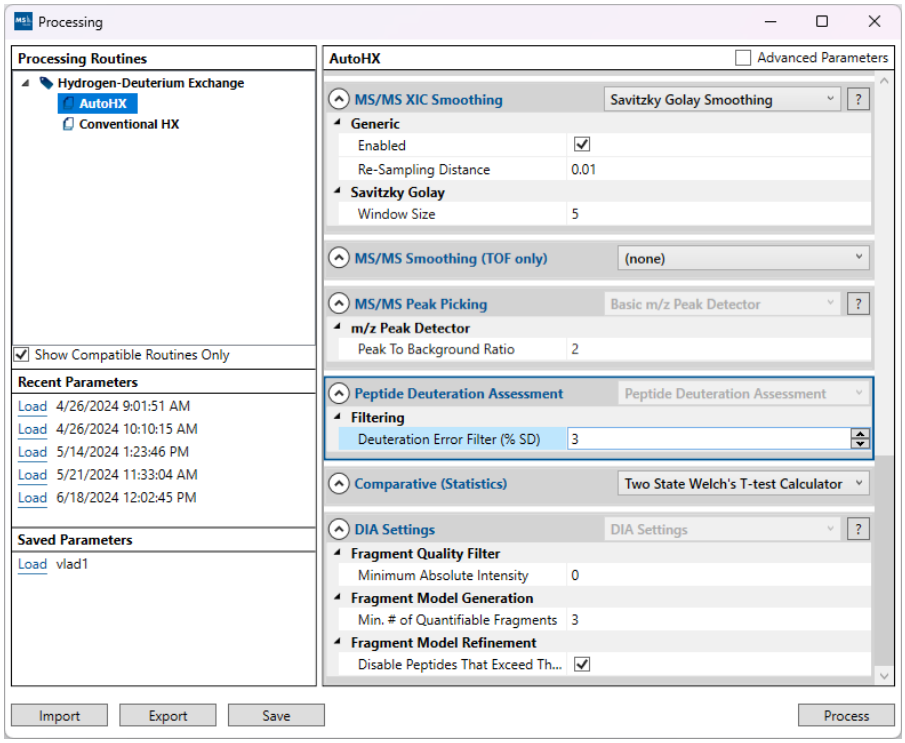
Important: *A license is required to enable the full capacity of AutoHX – if you do not see the "AutoHX Peptide HX-DIA: processing routine, please contact us to obtain a license.*

To get started with processing, you can open the processing window from the Process menu and select the "AutoHX - Peptide HX-DIA" routine. The most important parameters to set correct are the mass accuracy and the peptide elution time parameters. The rest of the parameters are already set to the default values we found most useful during our internal testing. For advanced users, the advanced parameters to fine-tune the search can be enabled via the top-right "Advanced Parameters" checkbox. If you wish to know more about the parameters, you can click on the help tooltip buttons ("?") next to each processing step/subsection.

The most impactful parameters for analysis are the DIA Settings -> "Min # of Quantifiable Fragments" (default 3) and the Peptide Deuteration Assessment -> "Deuteration Error Filter %" (measured in % corrected D, default 3%). The default of 3 minimum quantifiable fragments is a very relaxed setting which can result in some outlier peptide measurements. If your instrument is producing rich fragmentation in the DIA windows, it is beneficial to increase the minimum number of quantifiable fragments to achieve better consensus and thus strong automatic outlier removal. Our recommendation is to increase the minimum number of quantifiable fragments until you start seeing a sizable drop-off in sequence coverage. The optimum is maintaining a desired coverage percentage while the minimum number of fragments is as high as possible. A good range that works for our runs on the Eclipse is 5-8 minimum quantifiable fragments.

Note that in recent version you can now tune the "Min # of Quantifiable Fragments" parameter after the result is already generated directly from the bottom **Summary Peptides Table**. The main views will automatically update to reflect the peptides which pass the new fragment threshold.

Important: *For the current version, if you wish are processing a multi-target project (>2 states), you should turn OFF the "Fragment Model Refinement -> **Generate Optimal Fragment Model**" advanced parameter in the **Processing** window. This parameter is not shown by default and can be visible after checking the **Advanced Parameters** checkbox at the top-right of the main **Processing Dialog** window. The optimization process selects specific fragments for comparative analysis for each binary comparison (target1 vs control). When there are more than 2 target states present, this mechanism can produce some unexpected error bars on the Woods plot. This issue is a work in progress and will be addressed in a future non-beta version of the software. For now, disabling the optimization step means that the MS2 deuteration comes from ALL the quantifiable fragments for any given peptide.*



Once everything is set correctly, click "Process". The total processing time can span from minutes to hours, depending on the number of runs and the number of peptides. One can expect 2-5 minutes per file, depending on the size of the data and the number of peptides.

Results

After processing is finished, the result will be saved to disk inside the "Results" folder of the project and will be auto-selected in the left-hand-side project tree. AutoHX calculates the "Rescued" (MS2) deuteration in addition to the standard MS1-based deuteration using the MS2 DIA fragments. The "Final" deuteration for each peptide is based on a decision whether to use the MS1 or MS2 deuteration, depending on the quality of the MS1 and MS2 data. For the most part, MS2 deuteration is used for the "Final" deuteration because of its measurement redundancy -- multiple fragments representing the same peptide deuteration during full scrambling, even if the precursor is overlapped in MS1. All of the deuteration values for each peptide are displayed in the right-hand-side Properties tab (per replicate).

The standard MS-1 visualization from HX-Apps have been updated to show the final AutoHX results. Most notably, the Woods plot now displays certainty bars for each peptide which are determined by the degree of consensus on the peptide deuteration after sampling different sets of MS2 fragments.

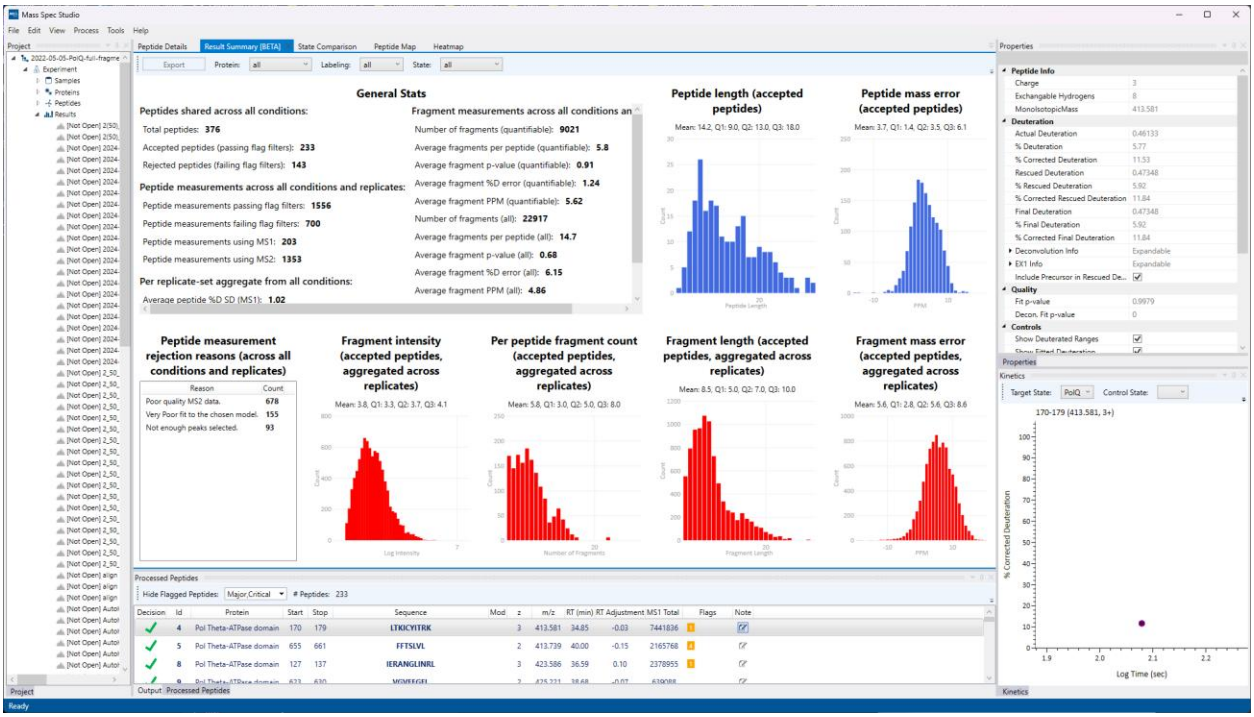
The Peptide Map view will map all the peptide deuteration values onto the protein sequence. You can select your protein, state and timepoint to filter the data. The view is fully customizable all the way from colors, D value ranges, size of peptide bars, etc.

In Recent versions, we added a "Localized Uptake Solver" with some options available to attempt to solve the deuterium placement down to individual residue wherever available. This is similar to the heatmap view in HDExaminer. It uses a method close to the one described in this paper

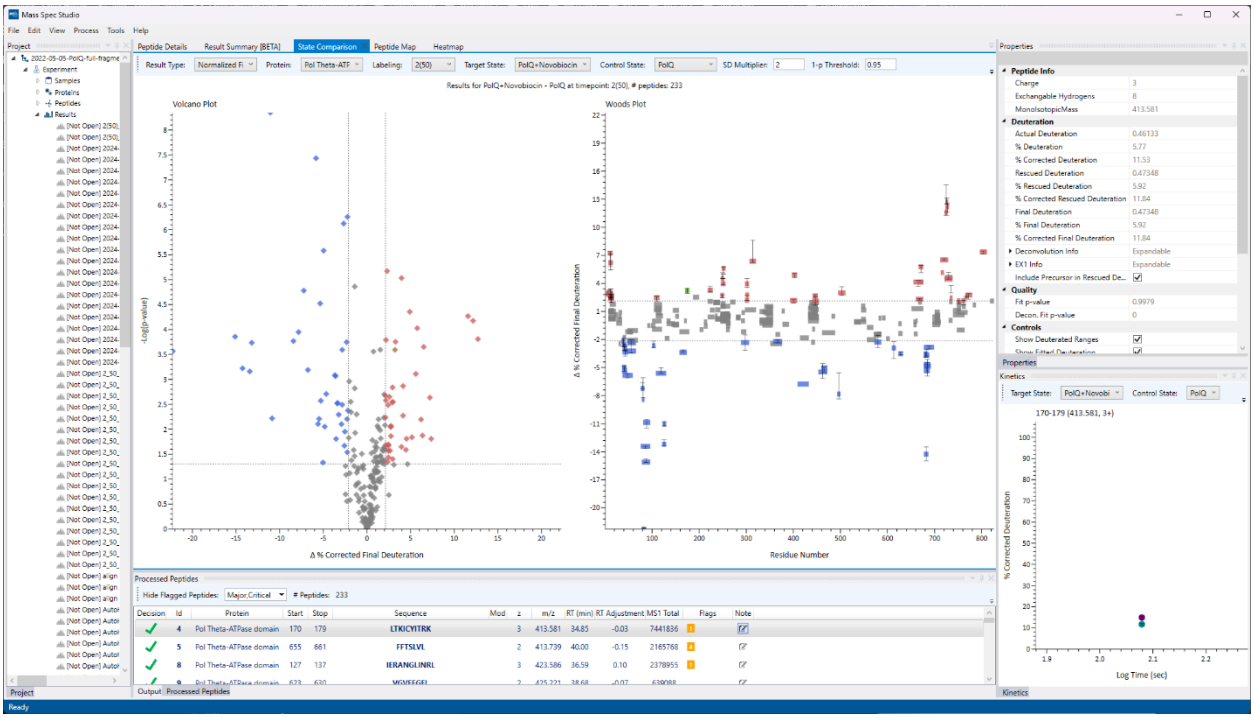
http://dx.doi.org/10.1002/cem.2876. Both the “Babic & Smith” and the “NNLS” options will use the matrix algebra to assess solvability, they just use slightly different methods to solve for the localized deuteration. The former “Babic & Smith” uses matrix algebra only via the Moore-Penrose pseudoinverse, the latter (NNLS) is a regression which constricts the problem boundaries to not be able to produce negative uptake values for each segment. Additionally, it also has an option to use the “Keppel & Weis” weighted averaging method described here: https://doi.org/10.1007/s13361-014-1033-6. The confidence of the solved localized uptake is displayed as a thin bar above each segment that typically ranges from green to red based on the residual error after solving. Areas that appear with a black certainty bar may contain multiple solutions, so they are considered not solvable by the matrix-based solutions.

Hint: If you are using a smaller monitor, some of the toolbar controls may be hidden. They can be found by clicking the small black arrow on the right-hand side of the top toolbars.

The “Result Summary” view provides summaries of the metrics used by AutoHX to determine the final accepted peptide list.

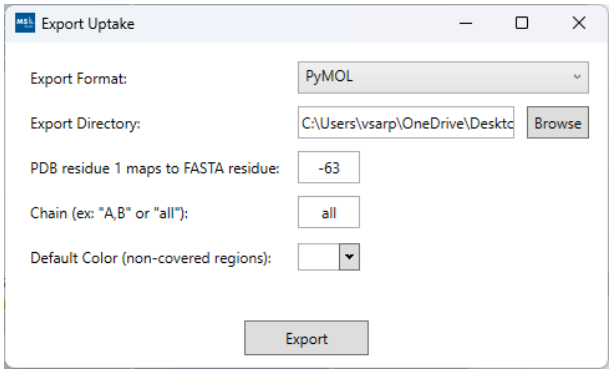


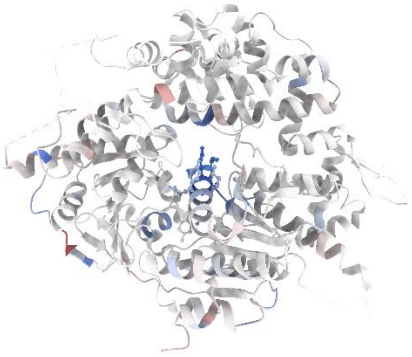
The State Comparison tab remains unchanged and will display the Volcano plot as well as the Woods plot for the selected Target/Control state pair. The errors bars represent the variance in the deuteration measurements coming from different sub-sets of fragments (randomly sampled).



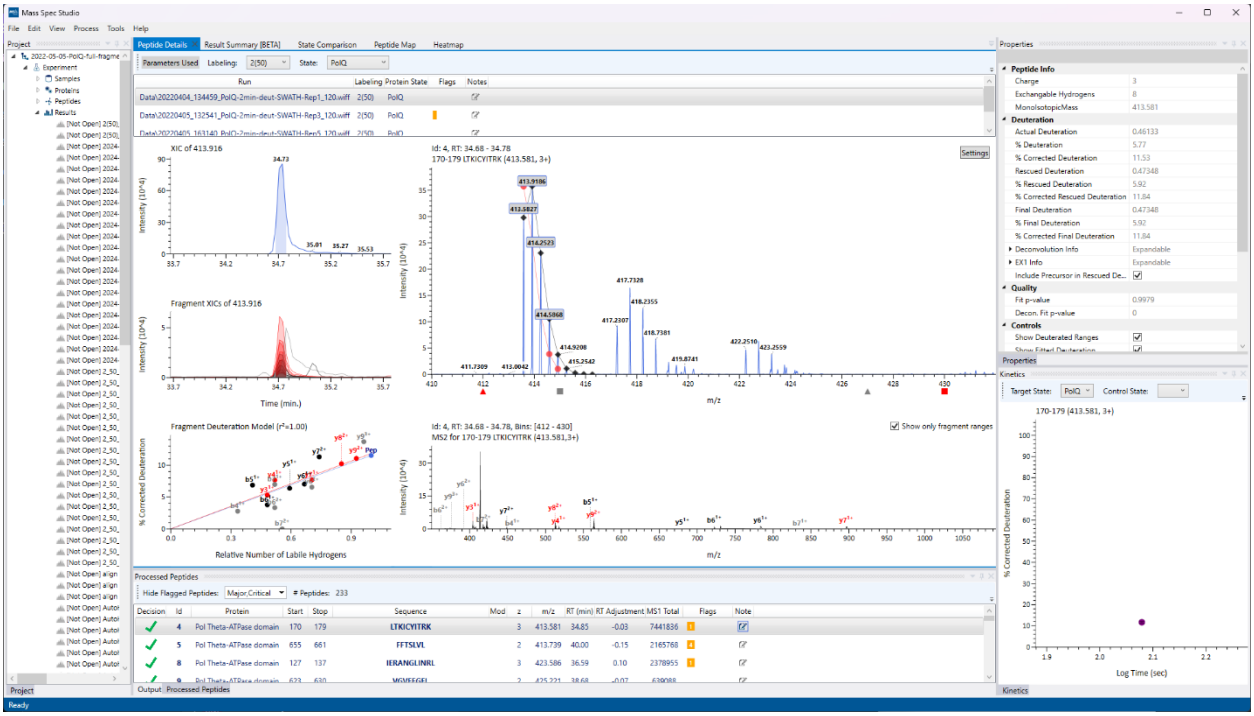
In the Peptide Map, the “Export” button next to the “Localized Uptake Solver” can be used to export the localized deuteration values to a PyMOL, ChimeraX, Chimera, HDXViewer or a plain text file. In most cases, two files will be generated at your selected export location: 1) the deuteration values, labelled as “uptake” in the file name and 2) the certainty bars, labelled as “confidence” in the file name.

Note that if your FASTA file and PDB file have different residues, you should apply the offset correction via the “PDB residue 1 maps to FASTA residue __” control prior to export.

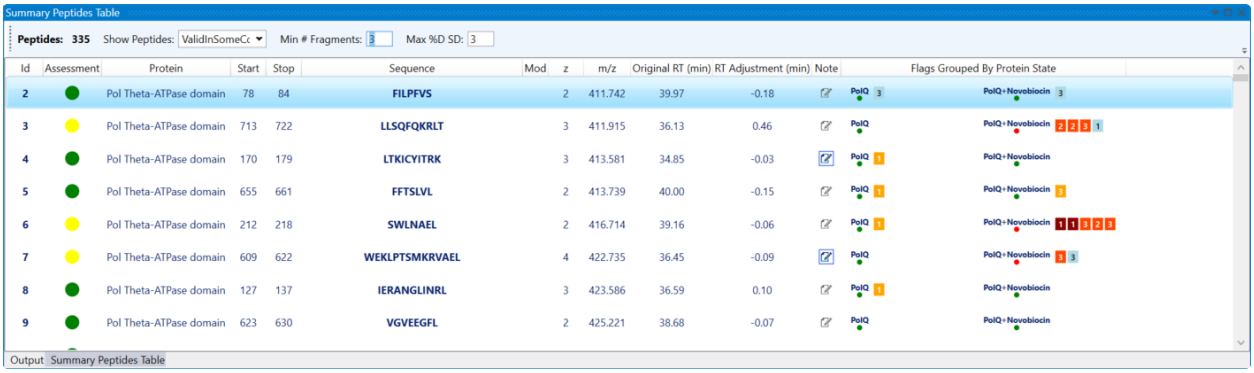




The “Peptide Details” view (formerly known as Manual Validation) will not automatically appear after processing. With AutoHX, we generally discourage cherry-picking of individual peptides whenever possible. However, if you do have outliers that you would like to inspect, simply click on a peptide in the Peptides table at the bottom or in any other view and the “Peptide Details” will pop up.

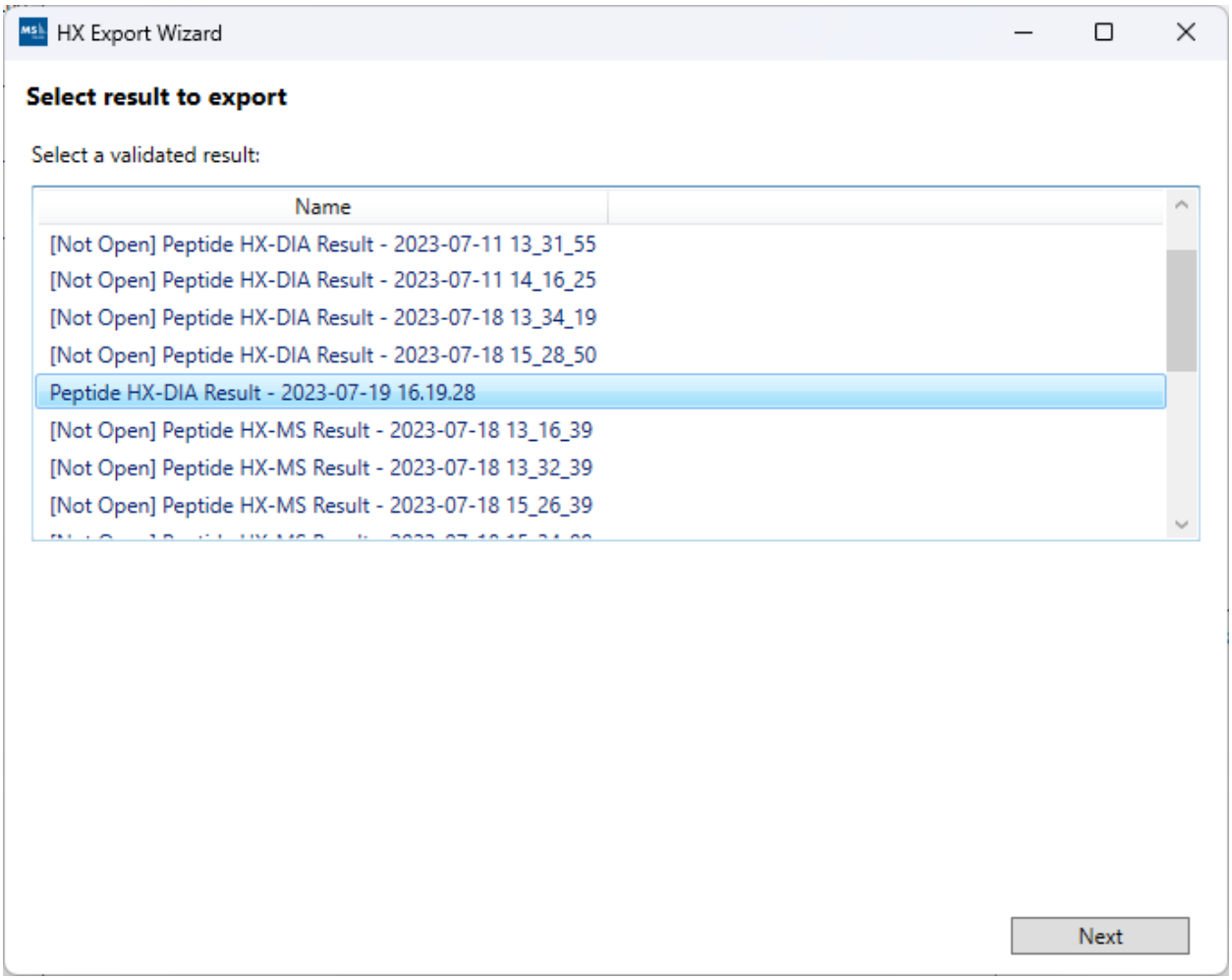


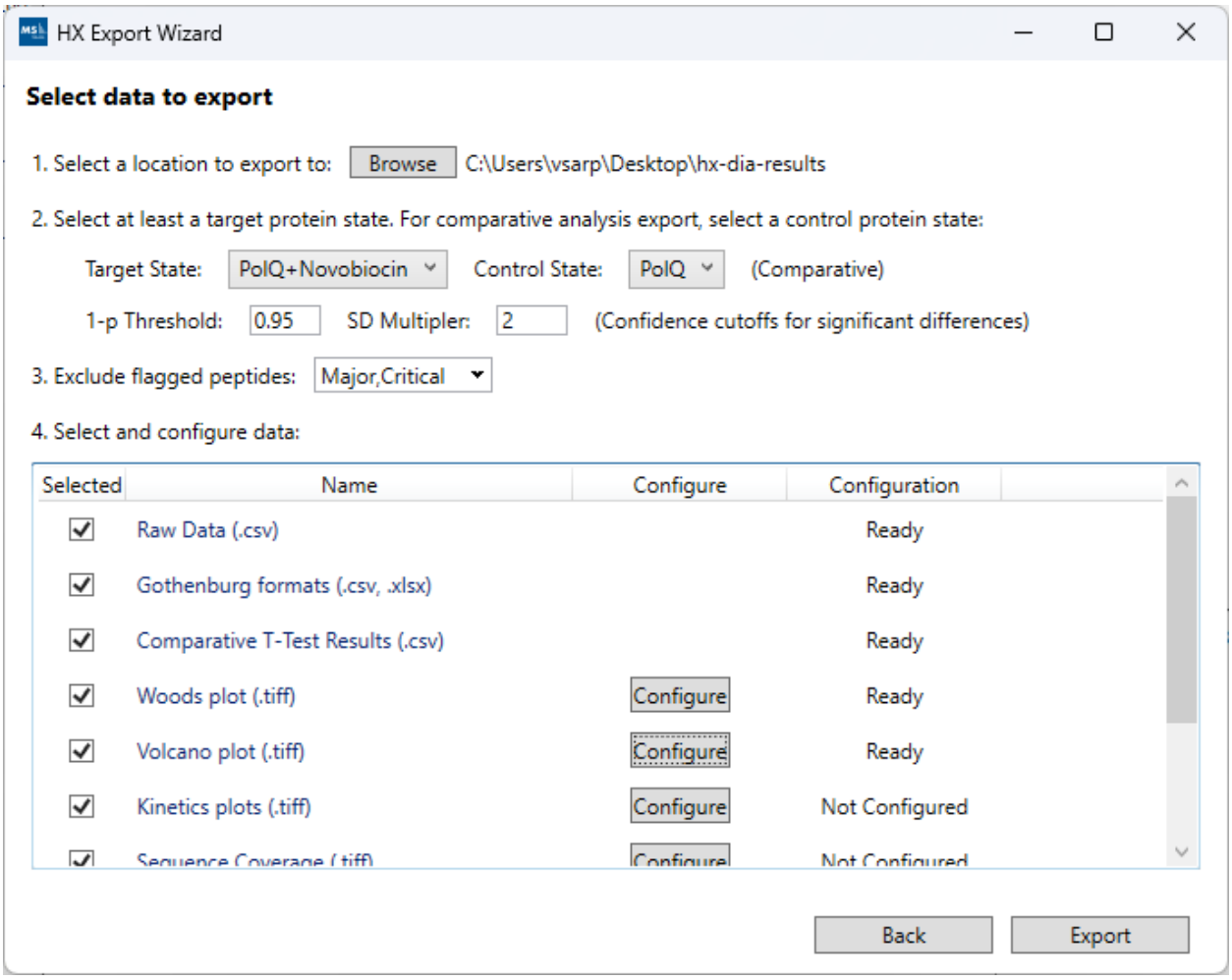
The Processed Peptides view at the bottom of the screen will display all the accepted peptides by default. If you wish to view rejected peptides, you can select the “Not Valid” option in the “Show Peptides” dropdown menu. Additionally, you can also adjust the minimum # of quantifiable fragments and the Max %D SD (maximum %D standard deviation) filters post-processing. We recommend increasing the minimum fragments and the Max %D SD until you notice a significant drop in the protein sequence coverage and peptide redundancy. This will ensure that the highest quality and most reliable measurements remain while some potential outliers with poor MS2 and large variability are removed.



Export

To export the results, click on File -> Export and follow the export wizard. This follows the standard HX-DEAL export wizard which allows you to select and configure everything from peptide-level .csv exports to snapshots of the commonly used visualizations (Sequence Coverage, Peptide Map, Kinetics, Woods Plot, Volcano Plot). Note that if you wish to export the visualization, they will need to be configured once before the export. The “Configuration” column will display all the selected exports that require manual configuration before the final export bundle is created. Configuration usually involves setting a template for how the visual should look in the final export bundle (size, DPI, fonts, etc.).




HX Export Wizard
Select data to export
1. Select a location to export to:
Browse
C:\Users\vsarp\Desktop\hx-dia-results
2. Select at least a target protein state. For comparative analysis export, select a control protein state:
Target State:
PolQ+Novobiocin
Control State:
PolQ
(Comparative)
1-p Threshold:
0.95
SD Multipler:
2
(Confidence cutoffs for significant differences)
3. Exclude flagged peptides:
Major,Critical
4. Select and configure data:
Selected
Name
Configure
Configuration
Raw Data (.csv)
Ready
Gothenburg formats (.csv, .xlsx)
Ready
Comparative T-Test Results (.csv)
Ready
Woods plot (.tiff)
Configure
Ready
Volcano plot (.tiff)
Configure
Ready
Kinetics plots (.tiff)
Configure
Not Configured
Back
Export