Top/Middle-Down Protein Sequencing

Biologics Explorer Software 5.0 Guidelines

Powered by Genedata Expressionist®

Genedata solutions in silico

SCIEX The Power of Precision

Top/Middle-Down Protein Sequencing Guidelines

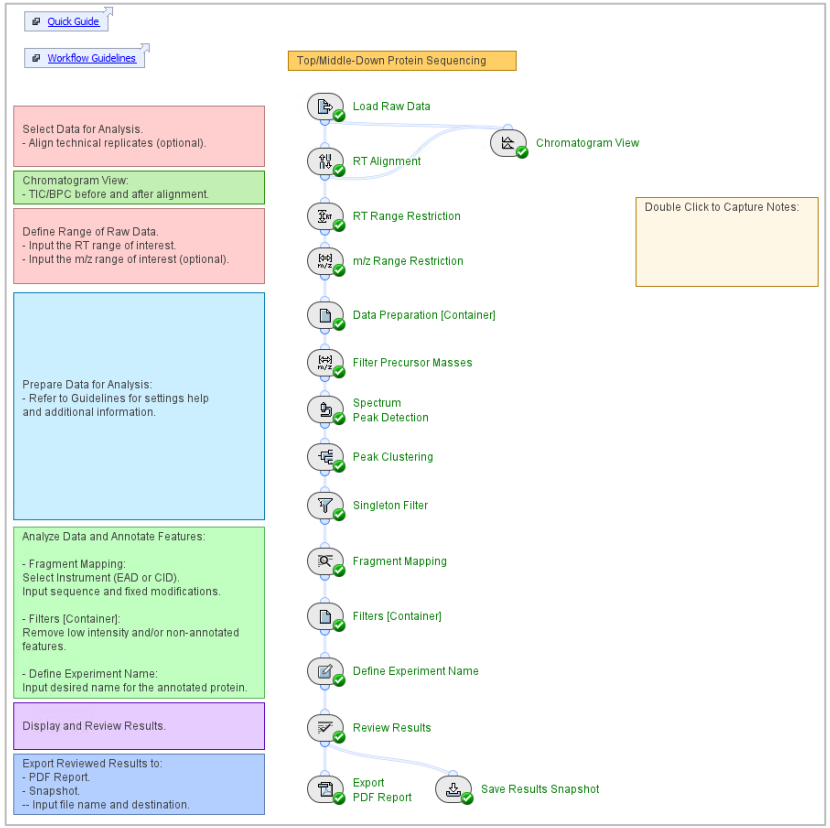
CONTENTS OF THIS GUIDE

Part A
General Guidelines for Protein Sequencing Workflows

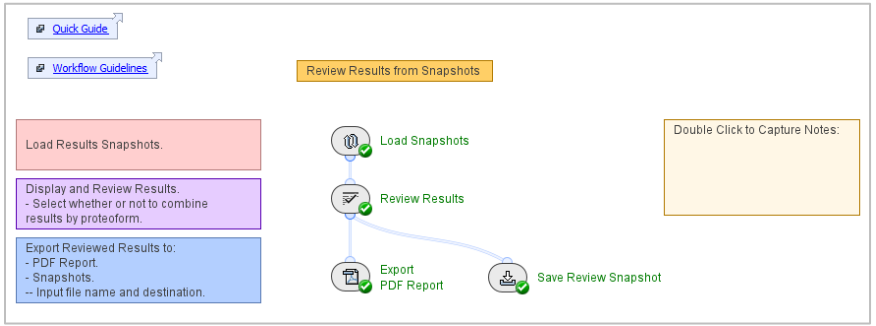
Overview of Applications for Top/Middle-Down Workflows

- These workflows are designed for top-down or middle-down protein sequencing with MS/MS data of:
  - Antibody subunits. For example, Fd', Fc/2 and LC after IdeS digestion and reduction of disulfide bonds.
  - Antibody chains. For example, HC and LC after reduction of disulfide bonds.
  - Proteins with masses up to approximately 50 kDa.
- Use the Top_Middle-Down_ProteinSequencing workflow to analyze each antibody subunit or chain separately
- Use the Top_Middle-Down_ ReviewSnapshots workflow to review all of the results together.
- Data quality for protein sequencing can be increased if multiple consistent technical replicates are analyzed at the same time in the Top_Middle-Down_ProteinSequencing workflow.
  - If the overall signal intensity is good, with a satisfactory signal-to-noise ratio, then a single sample can be sufficient for a successful analysis.

Top/Middle-Down Protein Sequencing Workflows



Top_Middle-Down_ProteinSequencing _Be5.0



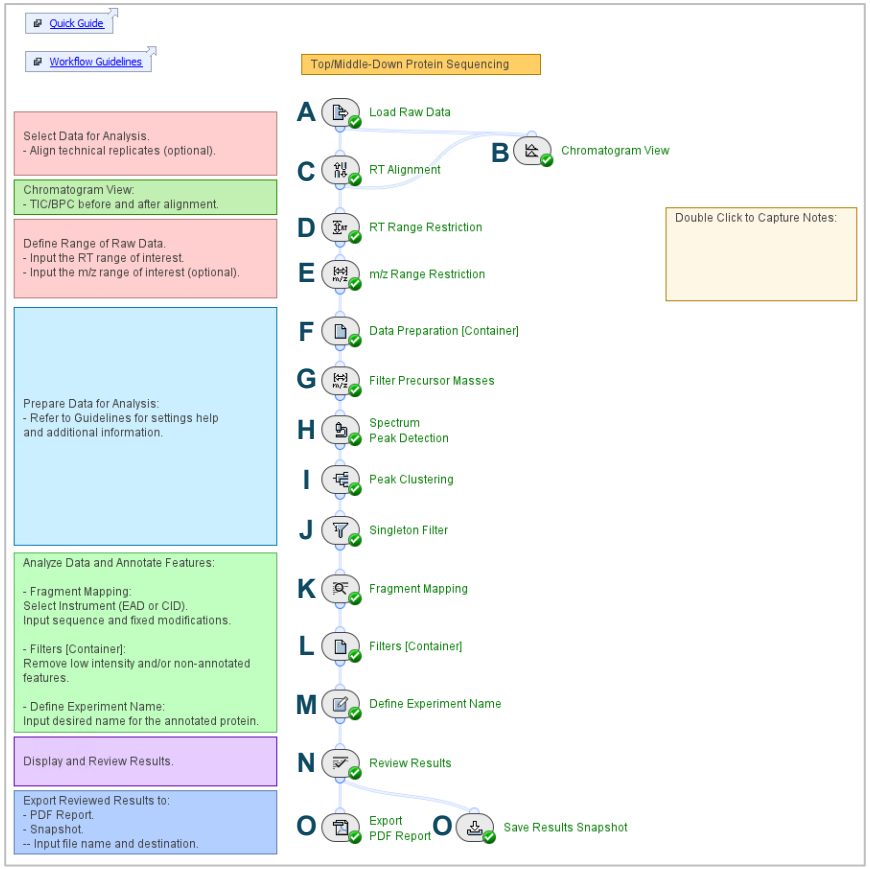
Top_Middle-Down_ReviewSnapshots_Be5.0

Part B
Guidelines for Protein Sequencing Workflows

1. Activity Nodes for Protein Sequencing

WORKFLOW SPECIFIC INFORMATION AND GUIDELINES

Top/Middle-Down Protein Sequencing Workflow



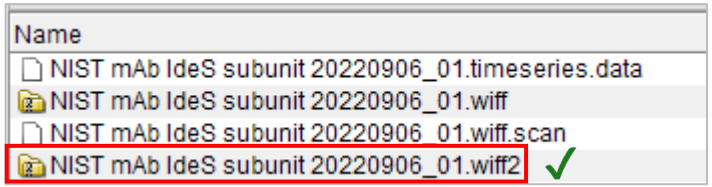
Activity Nodes in the Protein Sequencing Workflow

A. *Load Raw Data*
B. *Chromatogram View*
C. *RT Alignment*
D. *RT Range Restriction* → Input required*
E. *m/z Range Restriction* → Input required*
F. *Data Preparation [Container]*
    i. *MS/MS Data Selection*
    ii. *Sum Spectra per Replicate*
    iii. *Average Spectra Across Replicates*
    iv. *Spectrum Smoothing*
    v. *Baseline Flattening*
    vi. *Noise Subtraction*
    vii. *Intensity Thresholding*
    viii. *MS/MS Range Restriction*
G. *Filter Precursor Masses*
H. *Spectrum Peak Detection*
I. *Peak Clustering*
J. *Singleton Filter*
K. *Fragment Mapping* → Input required*
L. *Filters [Container]*
    i. *Valid Feature Filter*
    ii. *Highest Feature Filter*
    iii. *Filter Annotated Peaks*
M. *Define Experiment Name* → Input required*
N. *Review Results*
O. Report and Export Results

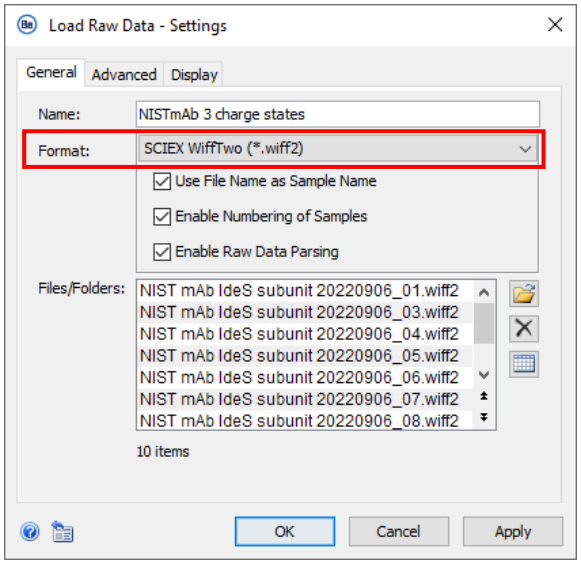
*Optimize settings for new samples

Load Raw Data: Data Files

- To upload raw data files, click the folder icon .
  - Select container files with the format wiff or wiff2.
    - If data was acquired with the ZenoTOF 7600 mass spectrometer, then select only the wiff2 format.
    - Do not select the auxiliary files with the same name.

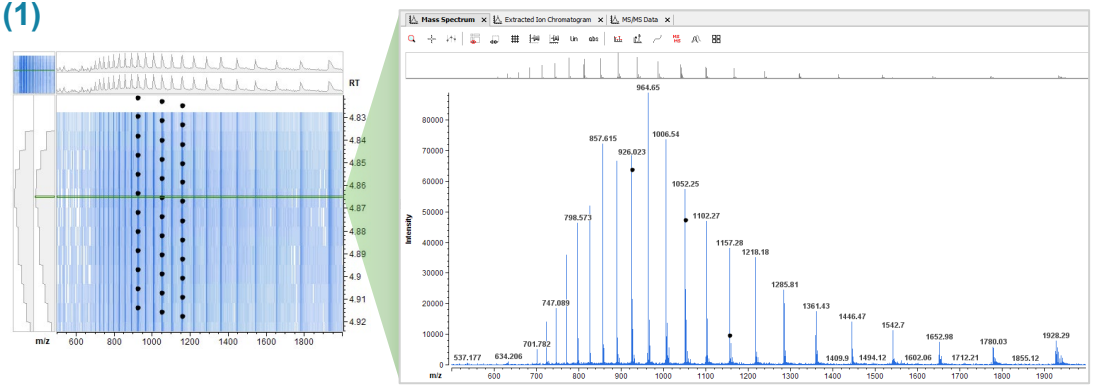


  - To select samples in a wiff or wiff2 container file:
    - Double-click the wiff or wiff2 container to see the sample files.
    - Select the samples to upload.



Identify Charge States in Non-Deconvoluted Data

(1)

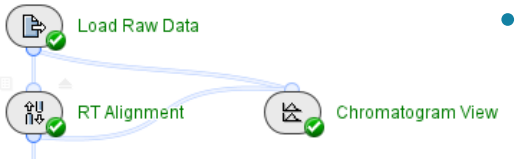


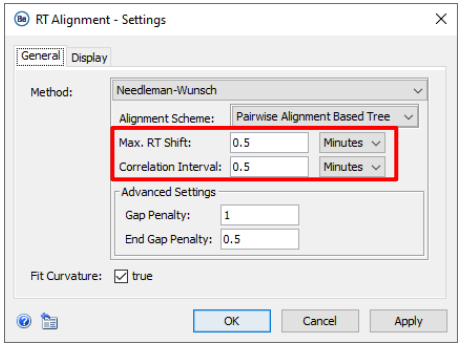
(2)

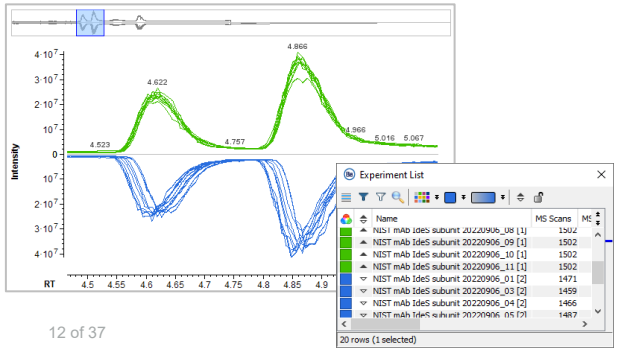
(3)

- To identify the charge state of a peak before deconvolution, for example, during optimization of the acquisition method and selection of the optimal charge states for MS/MS fragmentation:
  1. Take a horizontal slice across the ion map to show the charge envelope in the **Mass Spectrum** window.
  2. Use the **Measuring Tool** to measure the mass difference between two homologous peaks from consecutive charge states.
  3. Divide the given **Mass** value by the *m/z* of the peak of interest to calculate the charge state of that peak.

RT Alignment

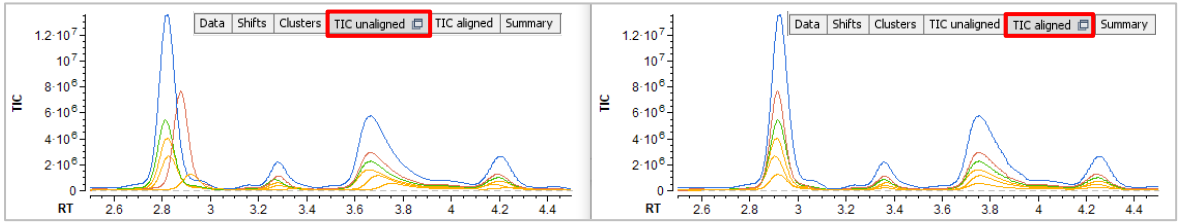






- Use *RT Alignment* to remove minor deviations in peak elution times between multiple technical replicates if required.
  - The **Pairwise Alignment Based Tree** uses hierarchical cluster analysis to align the most similar chromatograms until all samples are aligned.
  - The **Max. RT Shift** should be larger than the difference in RT between the multiple chromatograms.
  - The **Correlation Interval** should be at least half the width of a peak.

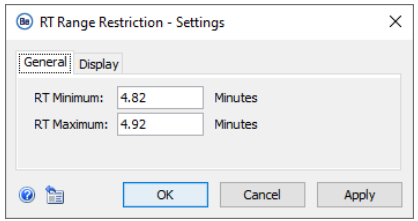
- The **TIC unaligned** and **TIC aligned** tabs show the data before and after *RT Alignment*.

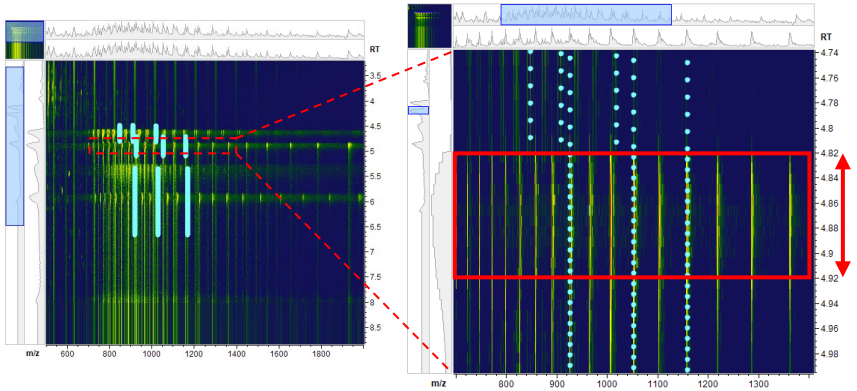


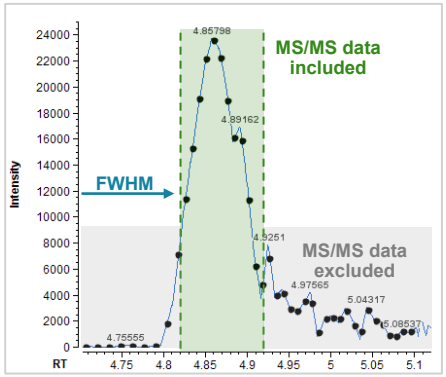
- Use *Chromatogram View* to compare samples before and after alignment.
  - Right-click the chromatogram to open the **Experiment List**.
    - The samples with suffix [1] are after alignment.
    - The samples with suffix [2] are before alignment.

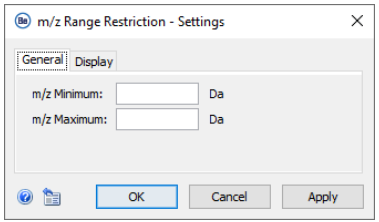
Restriction of RT and *m/z* Ranges

- To identify the RT ranges, open (double-click) either *Load Raw Data*, or *RT Alignment*, if this activity node was used.
- Identify a retention time (RT) range that includes MS/MS data from a single protein.
  - Select the RT range that contains MS/MS data acquired across the highest-intensity region of the peak (above the full width at half maximum, FWHM) in most replicates.







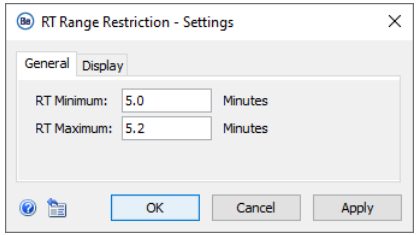


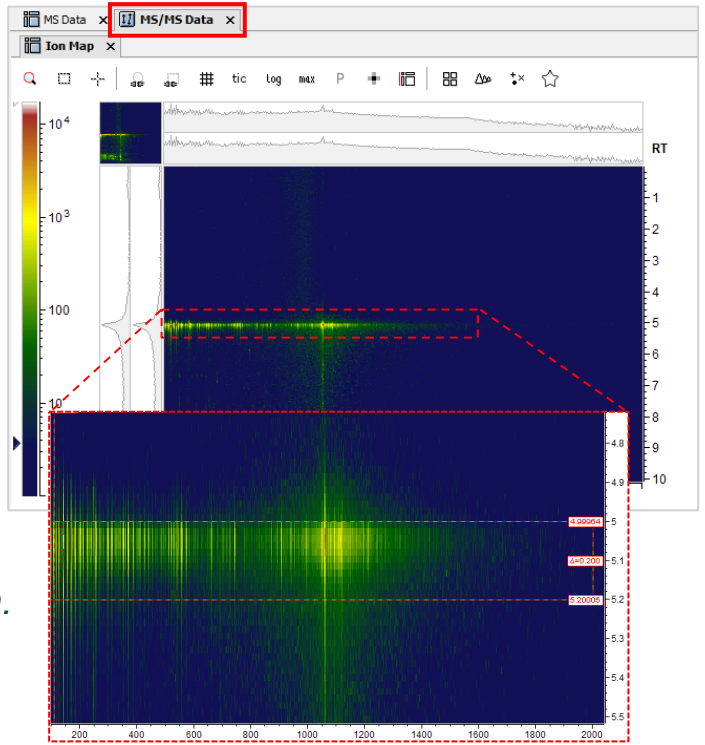
- Use *m/z Range Restriction* only if the specified RT range includes MS/MS data from other proteins or mAb fragments.

Note: If the fields are blank, or if the **Bypass** icon is activated, then the full RT or m/z ranges are used.

Restriction of RT and *m/z* Ranges: Special Case

- If protein sequencing data is acquired on only a single precursor, then the MS/MS data is found in the **MS/MS Data** tab as an ion map.
  - Select the RT range that includes the most intense region of the MS/MS data.

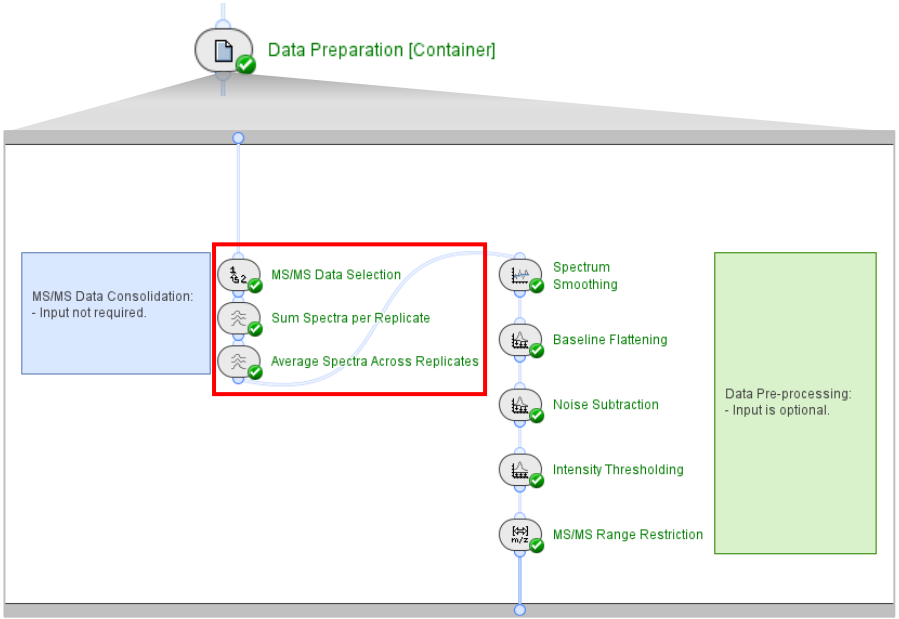




  - To keep all MS/MS data, do not use *m/z Range Restriction.*

Note: If the fields are blank, or if the **Bypass** icon is activated, then the full RT or m/z ranges are used.

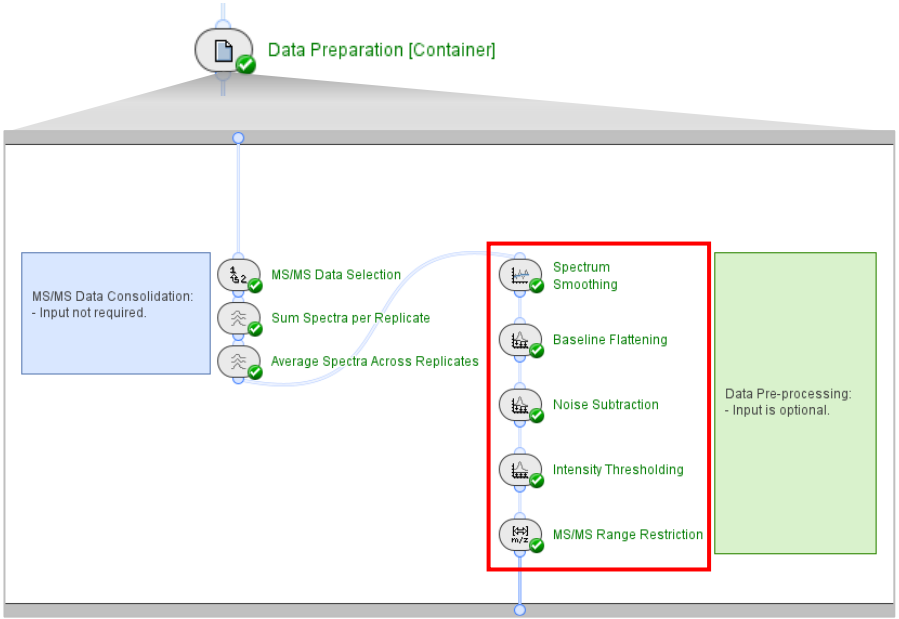
Data Preparation: MS/MS Data Consolidation



Note: The activity nodes used for consolidation of MS/MS data do not require any input or optimization, even if only a single replicate is analyzed.

- *MS/MS Data Selection*
  - Selects and extracts MS/MS data to create a new ion map for processing in the workflow.
- *Sum Spectra per Replicate*
  - Creates a single MS/MS spectrum from all of the MS/MS data in the selected RT range, across all of the charge states for each replicate experiment.
- *Average Spectra Across Replicates*
  - Calculates an average of the summed spectra from each replicate across all replicates, to create a single high-quality MS/MS spectrum for *Fragment Mapping*.

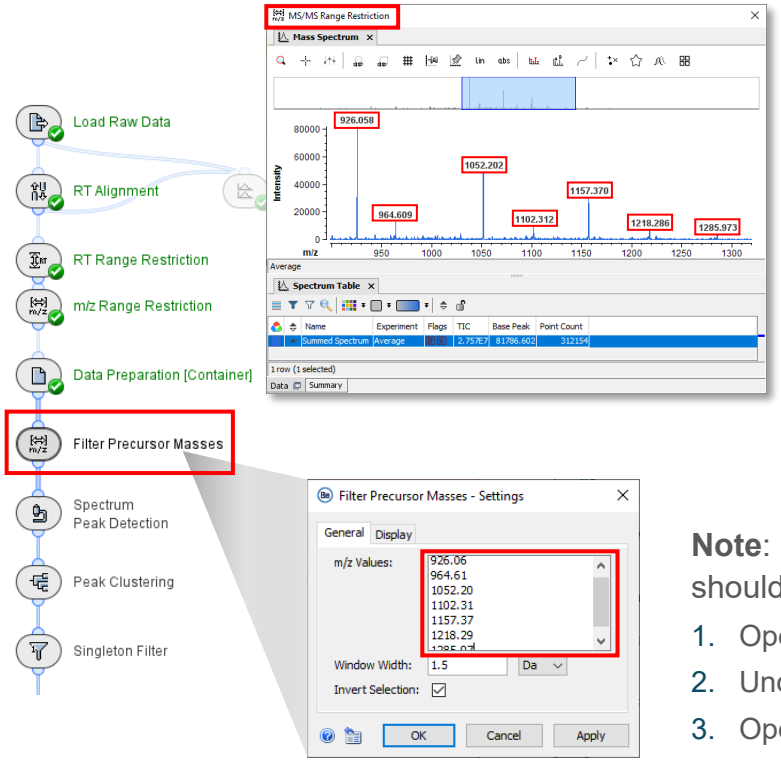
Data Preparation: MS/MS Data Pre-Processing



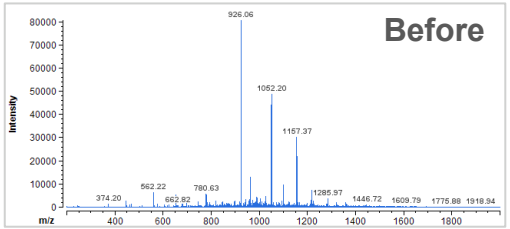
Note: The activity nodes related to pre-processing of MS/MS data require input only under unusual circumstances. The settings have been optimized across a range of datasets to give reliable results for protein sequencing applications.

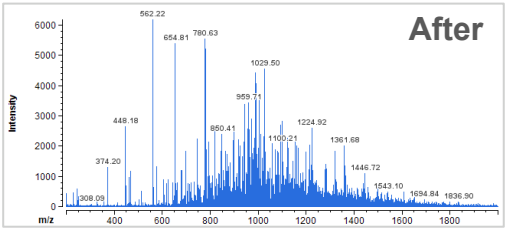
- *Spectrum Smoothing*
  - Removes small fluctuations in intensity to decrease the total number of detected peaks.
- *Baseline Flattening*
  - Removes low-intensity noise and decreases the baseline intensity.
- *Noise Subtraction*
  - Removes general noise and has an effect on all signals.
    - A decrease in the **Quantile** subtraction increases sensitivity, but it also increases the number of false positive identifications.
    - To increase sensitivity and thus increase bond coverage, it is recommended to decrease the thresholds in the *Filters [Container]*.
- *Intensity Thresholding*
  - Removes any signal below the specified threshold.
- *MS/MS Range Restriction*
  - If MS/MS data is acquired over a large *m/z* range, then data above and below specified *m/z* values can be removed to simplify data interrogation.
    - To keep all MS/MS data, do not use *MS/MS Range Restriction*.

Filter Precursor Masses



- *Filter Precursor Masses*
  - Removes data at specified *m/z* values.
    - To remove data related to the unfragmented precursors, enter a list of masses that are applicable to the molecule of interest.
  - Simplifies visualization of MS/MS clusters.

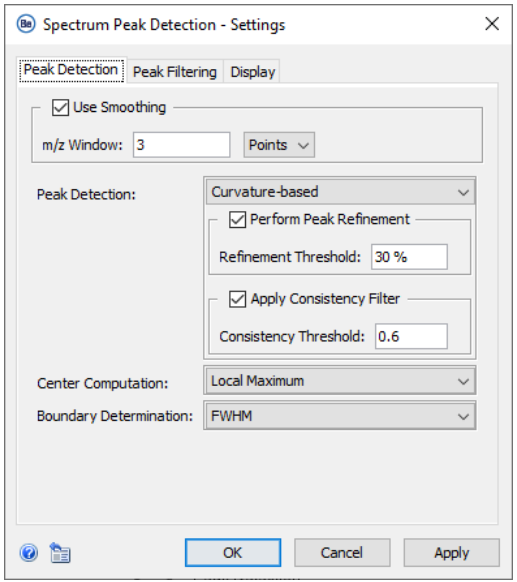




Note: Use the output of *MS/MS Range Restriction* to see the *m/z* values that should be removed:

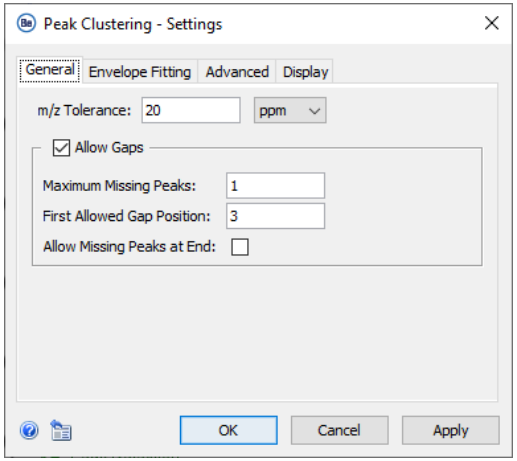
1. Open the results of *MS/MS Range Restriction* in the *Data Preparation [Container]*.
2. Undock the *MS/MS Range Restriction* window.
3. Open *Filter Precursor Masses* from the workflow and specify values.

Spectrum Peak Detection

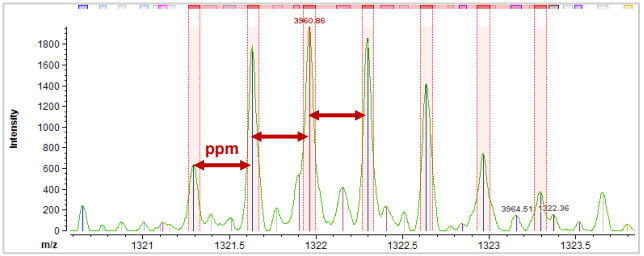


- **Use Smoothing**
  - Temporary smoothing is applied in the background for peak detection.
- **Peak Detection**
  - **Curvature-based** peak detection is used to resolve shoulder peaks. This type of peak detection is recommended for analysis of data with complex and often overlapping isotopic cluster profiles.
    - **Perform Peak Refinement**: For detection of subtle shoulder peaks.
      - **Refinement Threshold**: To increase peak splitting with subtle shoulder peaks, decrease the threshold percentage.
    - **Apply Consistency Filter**: Considers peak width and curvature.
      - **Consistency Threshold:** Values between 0.4 and 0.6 are recommended.
      - To decrease peak splitting, increase the threshold, up to a maximum of 2.

Peak Clustering: General

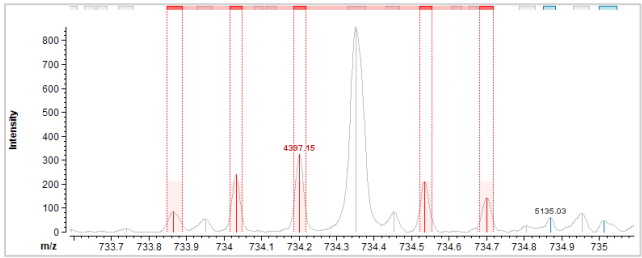


- ***m/z* Tolerance**
  - Enter the minimum acceptable mass accuracy that each peak in an isotopic envelope must have for them to be clustered together. An isotopic envelope is identified by either the detected or the recomputed monoisotopic mass.

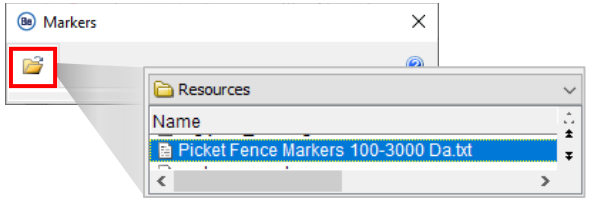


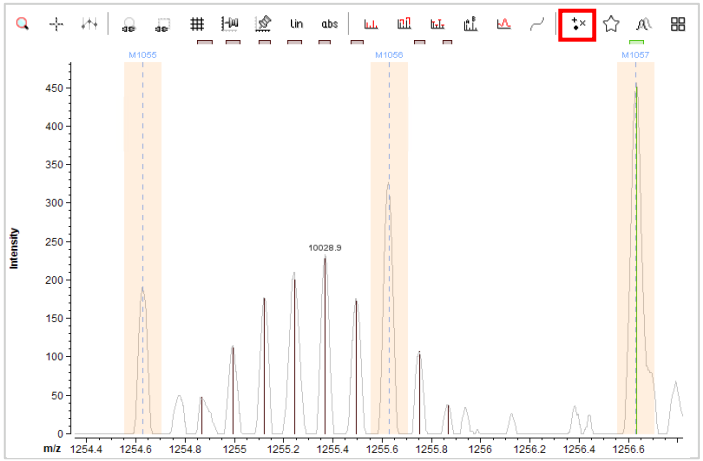
Note: This setting is not related to the mass accuracy of the mass spectrometer.

- **Allow Gaps**
  - Allows isotopic envelopes with missing or overlapping peaks to be clustered together.



Peak Clustering: 'Picket Fence' Peaks

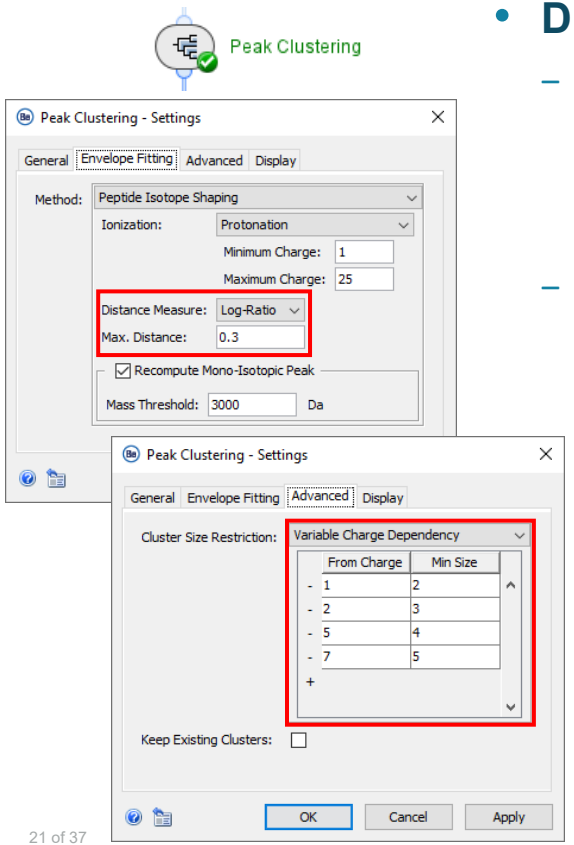




Note: Use the **Picket Fence Markers** file in the **Resources** folder to show the position of the 'picket fence' peaks to help with manual interrogation of isotopic clusters during data review. To load the Marker File included in the template workflow, use the **Show Markers** icon.

- In MS/MS data acquired from large peptides and proteins there is an inherent 'picket fence' ladder-signal of peaks separated by approximately 1 Da.
  - To decrease the effect of 'picket fence' peaks in the MS/MS data, decrease the energy transmitted to the molecule during the MS data acquisitions.
- Use **Allow Gaps** to compensate for 'picket fence' peaks in the MS/MS data.
  - Some 'picket fence' peaks will overlap with true peaks and cause a change to the isotopic cluster shape.
  - Overlapping 'picket fence' peaks prevent clusters from being correctly annotated, unless the **Allowed Gaps** setting is used to ignore changed peaks.

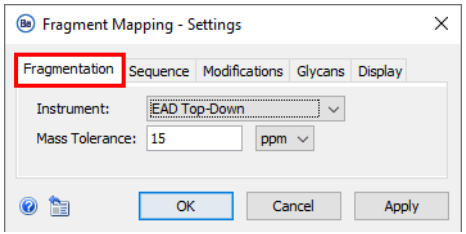
Peak Clustering: Envelope Fitting and Advanced Tab



- **Distance Measure**
  - Compares the measured isotopic cluster profile to the theoretical isotopic profile shape (based on the Averagine model).
    - **Log Ratio**: Gives all peaks in a cluster equal weight in comparisons between the experimental profile and the theoretical profile.
  - To decrease the required similarity between the measured and the theoretical isotopic clusters, increase the **Max. Distance**.
    - When **Max. Distance** is increased to improve detection of one cluster, there might be an effect on the isotopic clustering in another area of the data.
    - When the **Max. Distance** is increased, the number of false positive identifications also increases.

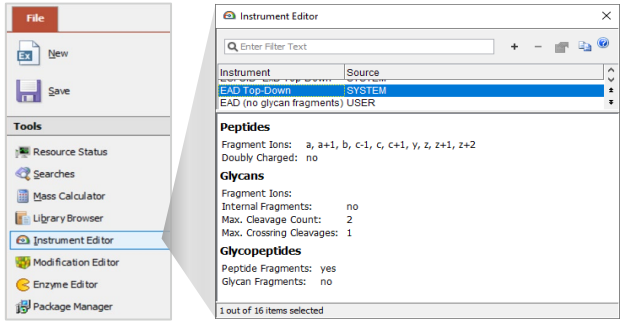
- **Variable Charge Dependency**
  - Enter the charge-dependent minimum number of peaks required for an isotopic cluster to be identified.

Fragment Mapping: Fragmentation



Fragmentation tab:

- **Instrument**: Select either **EAD Top-Down** or **CID Top-Down**.
  - To review or change the types of fragment ions used for identification:
    - Browse to **File** > **Tools** > **Instrument Editor**.

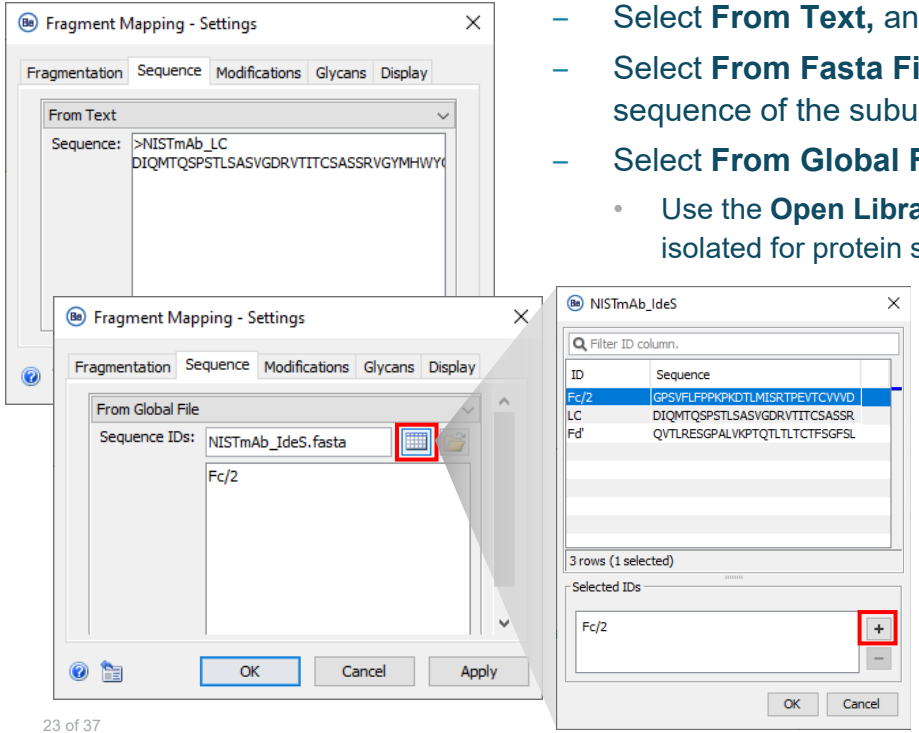


- **Mass Tolerance**: Enter the minimum acceptable mass accuracy that each isotopic cluster must have to match the experimental masses to theoretical mass for annotation of the MS/MS fragment data.

Fragment Mapping: Sequence

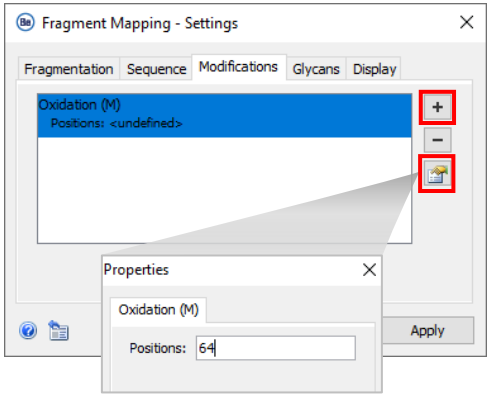
Sequence tab:

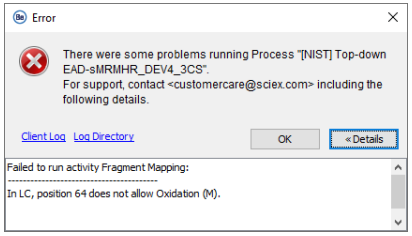
- **Sequence**:
  - Select **From Text,** and then type the sequence.
  - Select **From Fasta File**, and then add a FASTA file that contains the protein sequence of the subunit or chain that was isolated for protein sequencing.
  - Select **From Global File**, and then add the applicable FASTA file.
    - Use the **Open Library** icon to select the protein sequence of the subunit or chain that was isolated for protein sequencing.



- To analyze a mAb that was digested with IdeS:
  1. Run a separate *Fragment Mapping* search for each of Lc, Fd' and Fc/2.
  2. Use *Save Results Snapshot* to export the sbf file for each proteoform.
  3. Analyze the sbf files together with the Top_Middle-Down_ReviewSnapshots workflow.

Fragment Mapping: Modifications



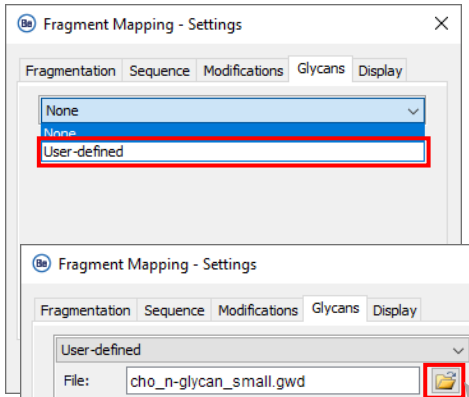


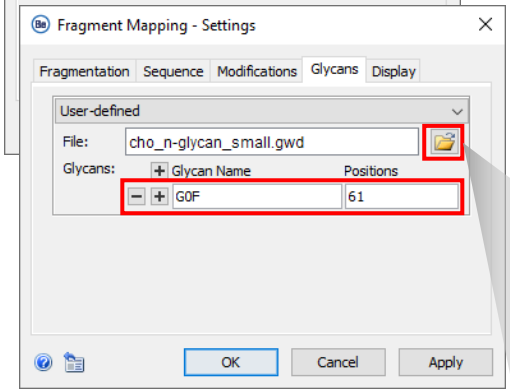
Modifications tab:

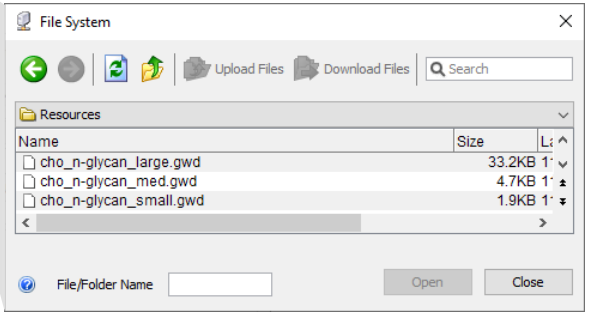
- All modifications are fixed.
- To see the list of available modifications, click the **+** icon to open the **Select Entries** dialog.
- To add information for the selected modification, click the **Configure selected entries** icon to open the **Properties** dialog.
  - Enter the expected amino acid position of the modification.
  - If the specified amino acid is not at that location in the protein sequence, then the activity node shows an error.

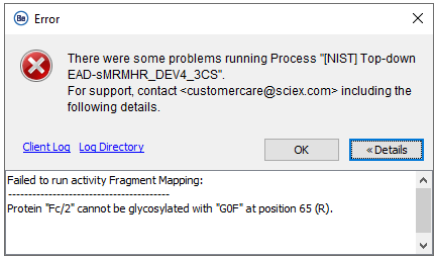
- To compare modifications specified at different amino acid positions:
  1. Run a separate *Fragment Mapping* search for each modification position.
  2. Use *Save Results Snapshot* to export the sbf file for each proteoform.
  3. Analyze the saved sbf files together with the Top_Middle-Down_Review Snapshots workflow.

Fragment Mapping: Glycans





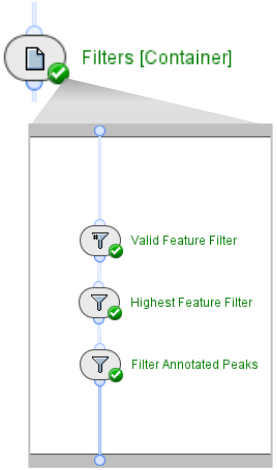




Glycans tab:

- Select **User-defined** if the sequence contains a glycosylation.
  - **File**: Browse to **Resources** in the **File System**, and then select a glycan library that is applicable to the molecule being investigated.
    - To review or change a glycan library: Browse to **File** > **Tools** > **Library Browser**.
  - **Glycans**: Enter the glycan name and the amino acid position in the sequence.
    - If multiple glycans are entered, then each is searched for as a separate fixed modification.
    - If the specified amino acid position cannot be glycosylated, then the activity node shows an error.

Filters



- Use the *Filter* activity nodes to decrease the number of peaks that are included in the results.
  - Keep the most relevant peaks, and remove those that originate from noise.
  - Activate the **Bypass** icons to keep all possible peaks.

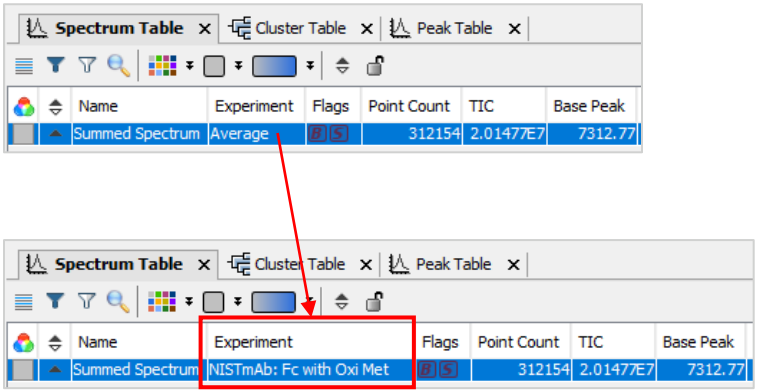
- *Valid Feature Filter*
  - To remove low-intensity isotopic clusters, enter an applicable **Validity Threshold**.
    - If the most intense peak is below the specified threshold, then the isotopic cluster is removed.
- *Highest Feature Filter*
  - Enter the maximum number of isotopic clusters to keep based on the selected Observable.

  Note: For information about Observables, refer to the Online Help (Observables - SCIEX BE 5.0).
- *Filter Annotated Peaks*
  - Removes any non-annotated peaks to simplify data visualization in *Review Results*.

- To increase sensitivity, decrease the *Valid Feature Filter* threshold or increase the number of clusters in *Highest Feature Filter*. An increase in sensitivity will:
  - Increase the number of detected clusters and, thus, bond coverage.
  - Increase the number of false positive identifications.
    - The increase in false positive identifications caused by a decrease in the *Valid Feature Filter* threshold is less than that caused by a decrease in the **Quantile** subtraction in the *Noise Subtraction* activity node.

Note: To optimize the *Filters* and change the sensitivity, compare the results for both the bond coverage and total number of detected clusters in the *Fragment Mapping* activity node with the results of the *Review Results* activity node.

Define Experiment Name



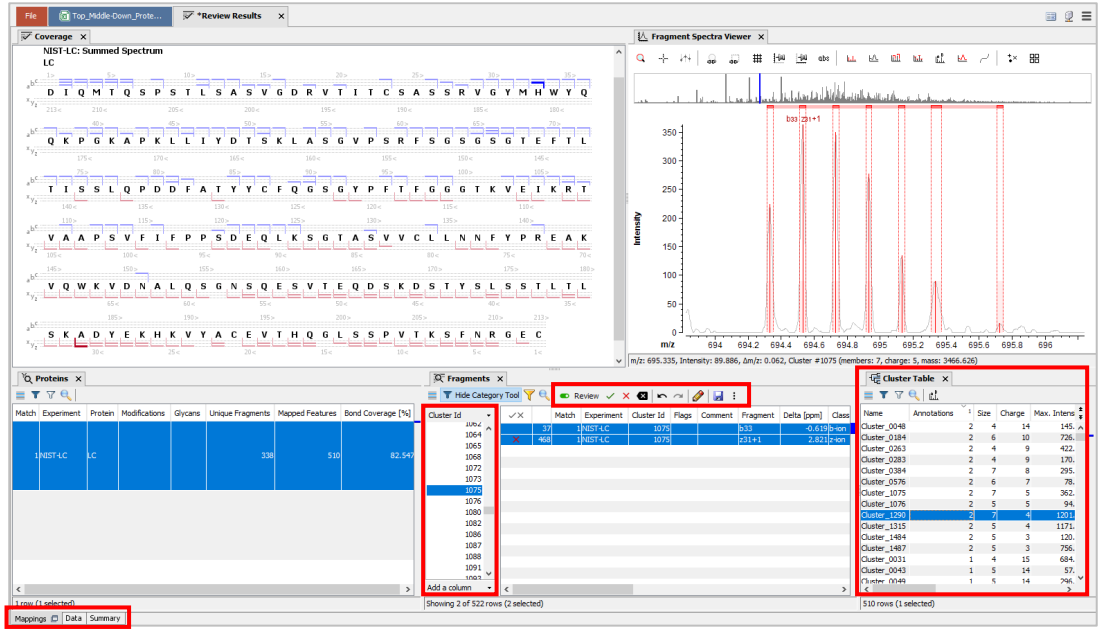
Note: To analyze results from different Protein Sequencing workflows that are saved as *Results Snapshots,* load them together in the Top_Middle-Down_ReviewSnapshots workflow.

- If multiple technical replicates are analyzed together, then the **Experiment** name is overwritten in the *Average Spectra Across Replicates* activity node.
- Use the *Define Experiment Name* activity node to enter a new **Experiment** name for the dataset.
- *Define Experiment Name* is also useful if the same data is analyzed with different search parameters in the *Fragment Mapping* activity node.
  - For example, for comparison of different results when a modification is searched for at different amino acid locations.
  - It is a good practice to use *Define Experiment Name* to include unique information about the specific sample set and the workflow settings used during data analysis.

Review Results: Review Protein Sequencing Results

- Results of *Fragment Mapping* can be reviewed in the Top_Middle-Down_ProteinSequencing workflow, or in the Top_Middle-Down_ReviewSnapshots workflow.

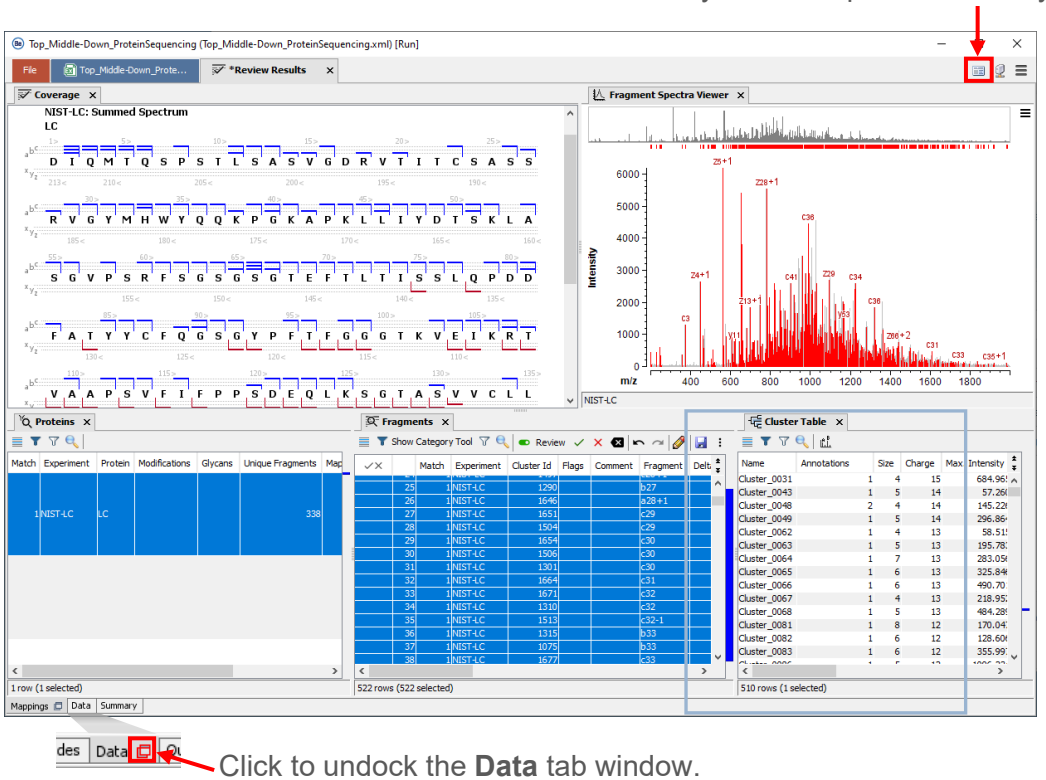
  Note: For information, refer to **B:** 2. Guidelines for the Review Snapshots Workflow.



- To review results:
  - Activate **Review** mode in the **Fragments** table.
  - To use the **Annotations** column in the **Cluster** table for data review, move the window from the **Data** tab to the **Mappings** tab.
    - Note: For more information, refer to the next page: *Review Results*: Create Custom Layouts.
  - **Accept** or **Reject** annotations as required.
  - To add a comment, either type in the applicable row in the **Comment** column, or use the icon to add the same comment to multiple rows.
  - To apply the changes, click the **Save** icon, and then select **Save and Reload**.

Review Results: Create Custom Layouts

Click to save the active layout or to open a saved layout.



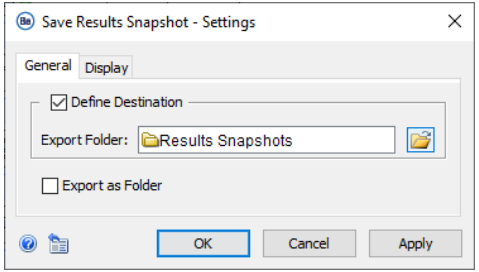
Click to undock the **Data** tab window.

- Each pane can be undocked and then docked at a different location.
- The location where the undocked pane will be docked is highlighted by a blue box.
- To move useful tables and visualizers to the **Mappings** tab:
  1. To undock the **Data** tab, click the icon.
  2. To undock any pane from the **Data** tab window, drag it to a new location on the **Mappings** tab.
- Favorite layouts can be saved and opened with the **Layout** icon.

Note: For more information, refer to the document: Biologics Explorer Software Quick Guide.

Export Intermediate Results for Further Analysis

- Snapshots are intermediate results that are saved permanently as sbf files.
  - The *Save Result Snapshot* creates an sbf file that contain the results of the Top_Middle-Down_ProteinSequencing workflow.



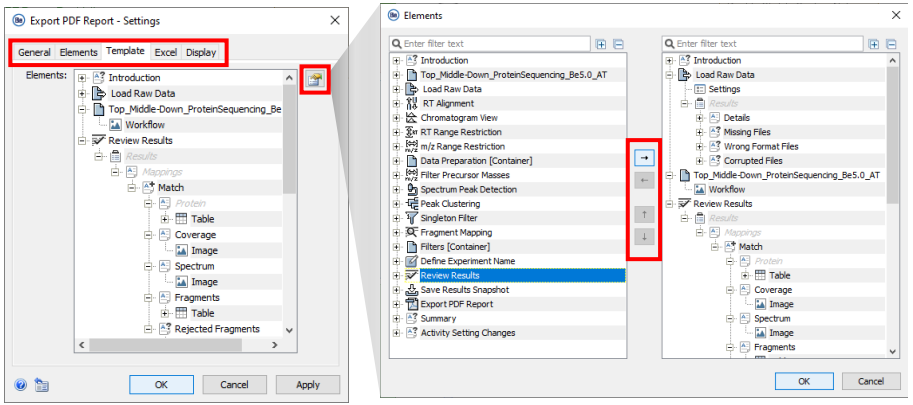
- To use a *Save Snapshot* activity node:
  - Deactivate the **Block** icon.
  - Select or add the folders where the sbf files will be saved.

- To review saved data:
  1. Select the sbf files to import into the *Load Snapshots* activity node in the Top_Middle-Down_ReviewSnapshots workflow.
     - Note: Select individual sbf files from the parent folder. Data will not load if the parent folder is selected.

Export PDF Report

- The output of *Export PDF Report* includes:
  - A PDF document.
  - An optional Excel file.
  - An embedded workflow file (xml) that includes all of the settings.
    - To open the xml file, drag the saved PDF Report into the workflow home page in the Biologics Explorer software.
    - Note: For more information, refer to the document: Biologics Explorer Software Quick Guide.

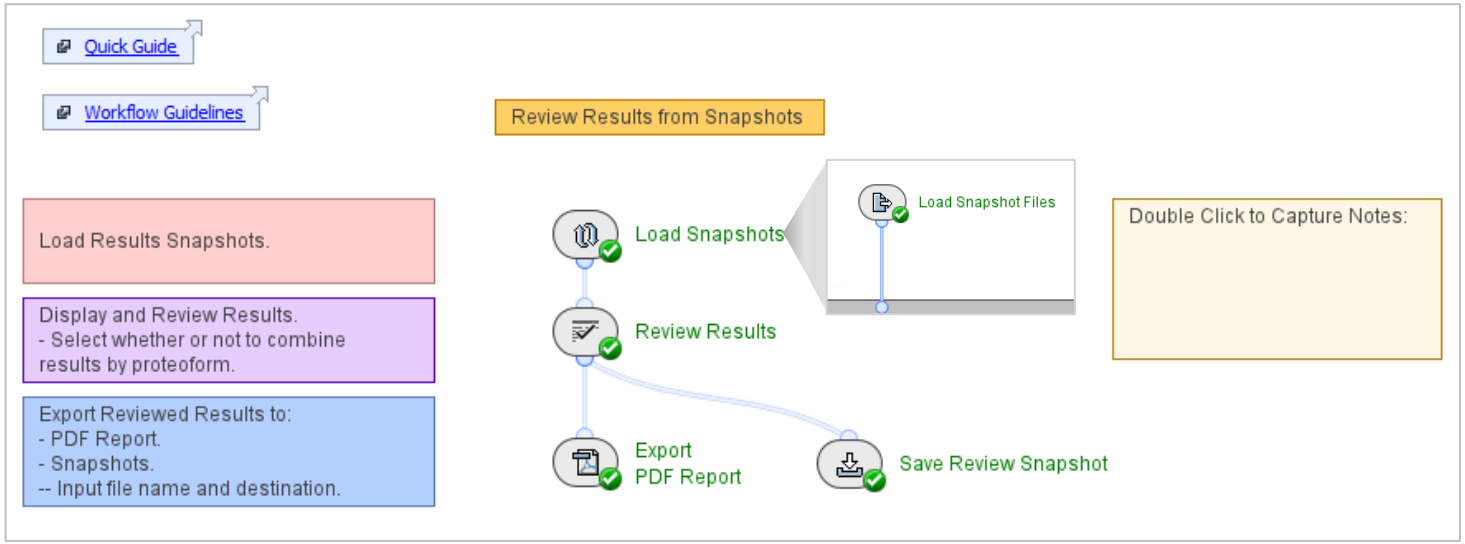
- **General** tab: Enter the name and saved location of the exported report.
- **Template** tab: Use the **Edit Selection** icon to select the **Elements** to be included in the report.
  - Select only columns of interest in reported tables. The layout of the tables is controlled by the number of columns.
- **Excel** tab: Use the **Edit Selection** icon to select the **Tables** to be included in the report.
  - All columns in a selected table are reported.



2. Guidelines for the Review Snapshots Workflow

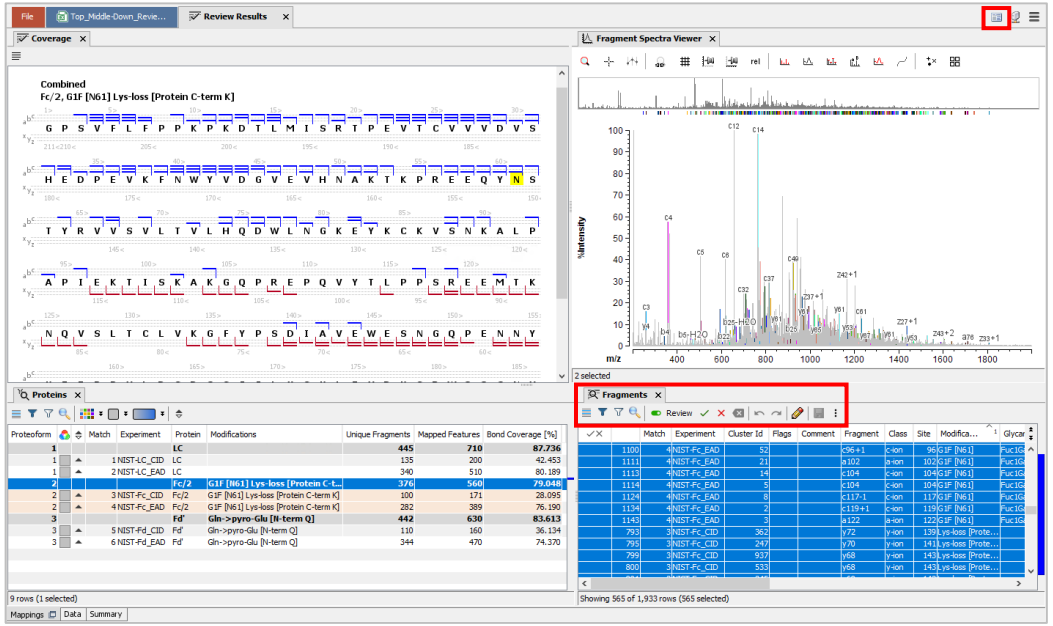
WORKFLOW SPECIFIC INFORMATION AND GUIDELINES

Protein Sequencing Review Snapshots Workflow



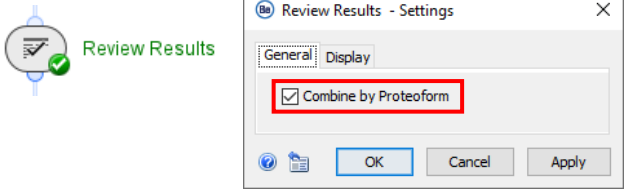
Top_Middle-Down_ReviewSnapshots_Be5.0

Review Snapshots Workflow: Overview



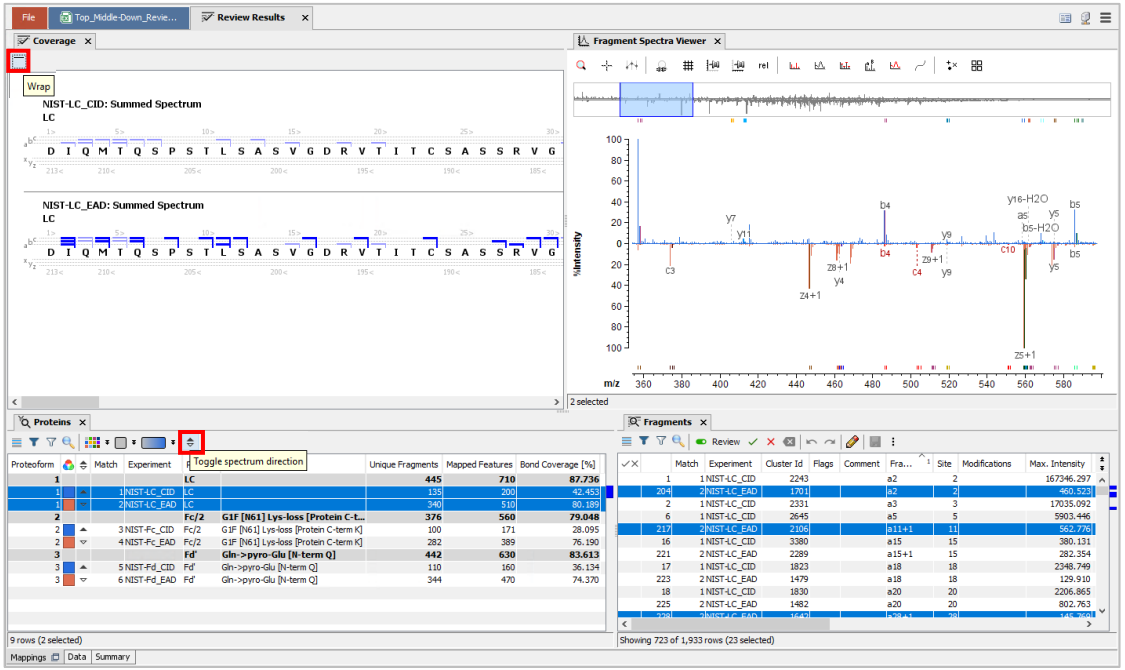
- This workflow can be used to open saved snapshot (sbf) files that are created with Protein Sequencing workflows.
- Use this workflow to review and compare results. For example:
  - To review results of all subunits from an IdeS digest.
  - To compare different sample acquisition parameters.
  - To create a combined sequence coverage from different fragmentation types.
- Use *Save Review Snapshot* to save a final reviewed sbf file for each sample.

Review Results



- To combine results from multiple data sources, for example data acquired with EAD and CID fragmentation or with different keV settings, select **Combine by Proteoform**.

- Use the **Wrap** icon to compare **Coverage** for different sequences
- Use the **Toggle spectrum direction** icon to create a mirror plot in the **Fragment Spectra Viewer**.
- To change **Labels, r**ight-click the **Fragment Spectra Viewer** and select **Settings**.
  - To increase the number of labeled peaks that are shown, select **Label Features: Offset**.
- Annotations that are selected in **Coverage**, **Fragment Spectra Viewer** and **Fragment Table** are synchronized.



SCIEX

The Power of Precision

For more information, visit the SCIEX website at sciex.com, or contact us in one of the following ways:

sciex.com/contact-us
sciex.com/request-support

Trademarks / Licensing

The SCIEX clinical diagnostic portfolio is For In Vitro Diagnostic Use. Rx Only. Product(s) not available in all countries. For information on availability, please contact your local sales representative or refer to https://sciex.com/diagnostics. All other products are For Research Use Only. Not for use in Diagnostic Procedures.