

Top/Middle-Down Protein Sequencing

Biologics Explorer Software 4.0 Guidelines





Top/Middle-Down Protein Sequencing Guidelines



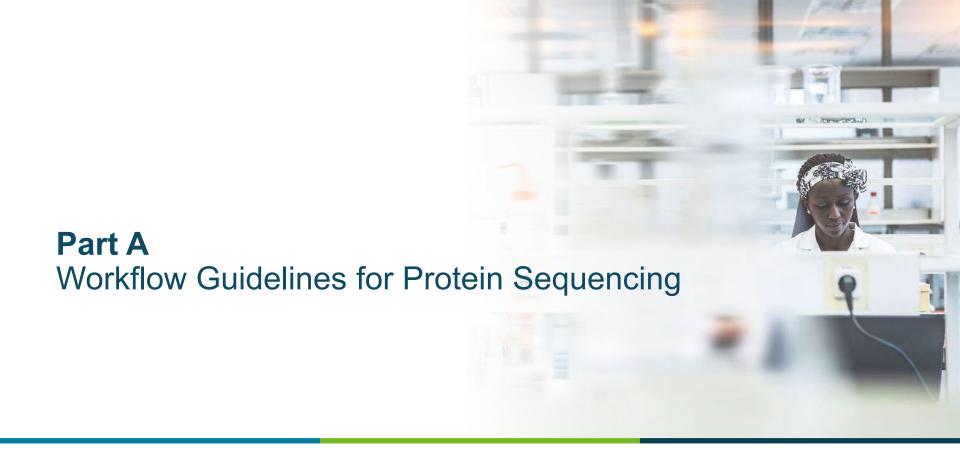
CONTENTS OF THIS GUIDE

Part A: Workflow Guidelines for Protein Sequencing

- 1. Overview of Applications and Workflows
- 2. Activity Nodes for Protein Sequencing:
 - Data Upload
 - Data Preparation
 - Peak Detection and Clustering
 - Fragment Mapping Search
 - Data Review
 - Export and Report Results

Part B: Workflow Guidelines to Review Snapshots

1. How to Use the Top_Middle-Down_ReviewSnapshots workflow



Overview of Applications for Top/Middle-Down Workflows



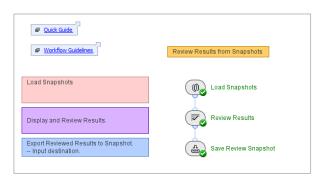
- These workflows are designed for top-down or middle-down protein sequencing with MS/MS data of:
 - Antibody fragments: for example, Fd', Fc/2 and Lc after IdeS digestion and reduction of disulfide bonds.
 - Antibody chains: for example, Hc and Lc after reduction of disulfide bonds.
 - Proteins with masses up to approximately 50 kDa.
- Analyze each antibody fragment or chain separately, and then use the Top_Middle-Down_ ReviewSnapshots workflow to show all of the results together.
- If the signal-to-noise ratio at high mass ranges is low, then 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.

4 of 33

Top/Middle-Down Protein Sequencing Workflows







Top_Middle-Down_ReviewSnapshots_Be4.0

Top_Middle-Down_ProteinSequencing _Be4.0

© 2024 DH Tech. Dev. Pte. Ltd.

Top/Middle-Down Protein Sequencing Workflow





Activity Nodes in the Protein Sequencing Workflow



- A. Load Raw Data
- B. RT Alignment
- C. Chromatogram View
- D. RT Range Restriction
- E. m/z Range Restriction
- 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
- L. Filters [Container]
 - i. Valid Feature Filter
 - ii. Highest Feature Filter
 - iii. Filter Annotated Peaks
- M. Define Experiment Name
- N. Review Results
- O. Save Results Snapshot

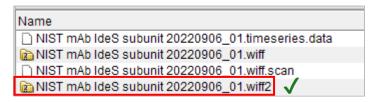
Load Raw Data: Data Files



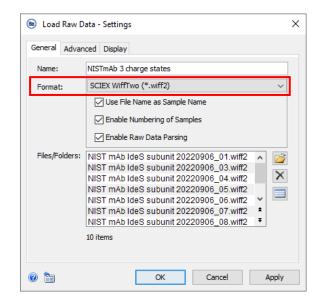
To upload raw data files, click the folder icon



- 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.



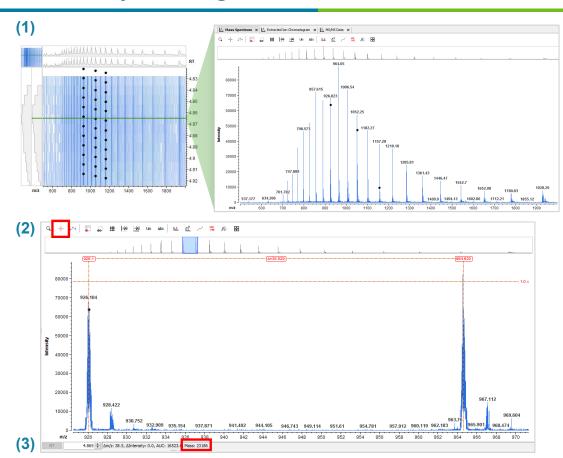
- To select samples in a wiff or wiff2 container file:
 - Double-click the wiff or wiff2 container to see the sample files.
 - Select the samples to upload from the list of sample files.



8 of 33 © 2024 DH Tech. Dev. Pte. Ltd.

Identify Charge States in Non-Deconvoluted Data





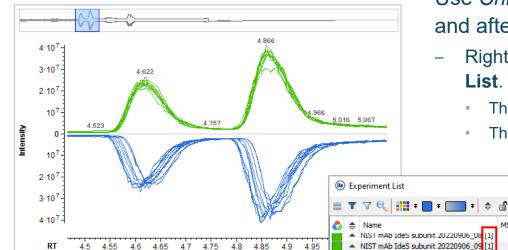
- To identify the charge state of a peak before deconvolution, for example, during optimization of the acquisition method and selection of the optimal charge states for MS/MS fragmentation:
 - Take a horizontal slice across the ion map to show the charge envelope in the Mass Spectrum window.
 - Use the 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.

9 of 33

RT Alignment







- Use *RT Alignment* to remove minor deviations in peak elution times between multiple technical replicates.
- 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.

MS Scans 1502

NIST mAb IdeS subunit 20220906 05 [2]

20 rows (1 selected)

1487

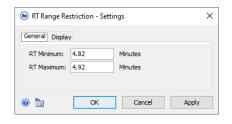
The samples with suffix [2] are before alignment.

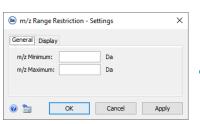
Restriction of RT and *m/z* Ranges

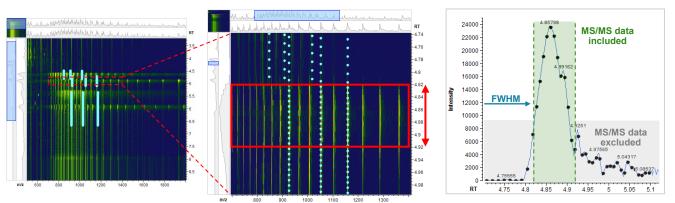




- To identify the RT ranges where there is meaningful data, open (double-click) Load Raw Data after the data is loaded.
- Identify a retention time (RT) range that includes MS/MS data from a single protein.
 - Select the RT range that contains MS/MS data acquired across the highest intensity region of the peak (above the full width at half maximum, FWHM) in most replicates.







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

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

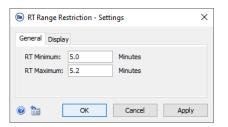


If protein sequencing data is acquired on only a single precursor, then the MS/MS data is

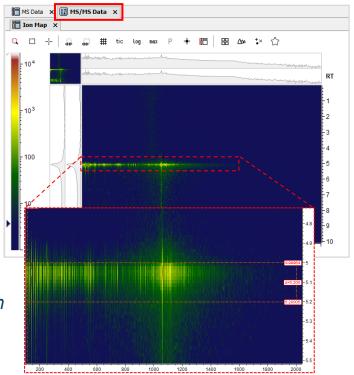
found in the MS/MS Data tab as an ion map.

 Select the RT range that includes the most intense region of the MS/MS.



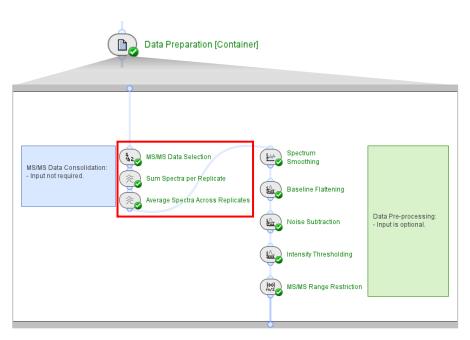


To include all MS/MS data, keep the m/z Range Restriction input fields empty or activate the Bypass icon.



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 further processing in the workflow.

Sum Spectra per Replicate

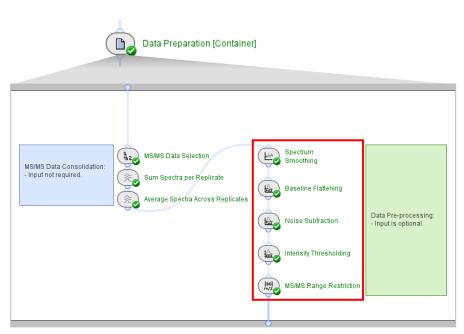
 Sums the MS/MS data from the selected RT range across all charge states for each replicate experiment.

Average Spectra Across Replicates

 Averages 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

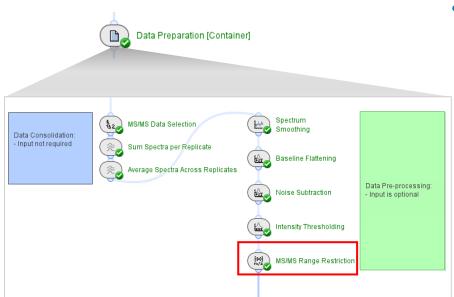
- Removes general noise and has an effect on all signals.
 - A decrease in the Quantile subtraction will increase sensitivity, but will also increase the number of false positive identifications significantly.
 - To increase sensitivity and thus increase bond coverage, it is recommended to decrease the thresholds in the *Filters* [Container].

Intensity Thresholding

Removes any signal below the defined threshold.

Data Preparation: MS/MS Range Restriction

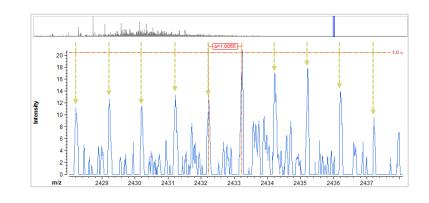




 To include all MS/MS data, keep the MS/MS Range Restriction input fields empty.

MS/MS Range Restriction

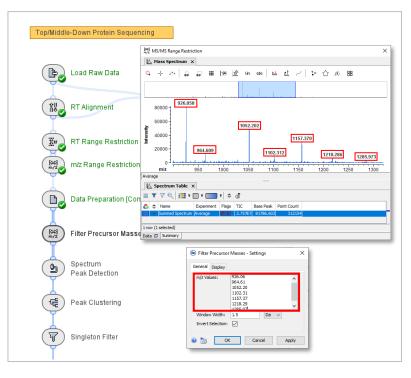
- If the MS/MS data is acquired over a large m/z range, then data above and below defined m/z values can be removed to simplify data interrogation.
 - For example, above 2000 m/z the inherent 1 Da 'picket fence' ladder peaks that are typical of this data type have higher intensities than at lower m/z. The higher intensity of these peaks decreases the robustness of the isotopic clustering of fragment ions.



Filter Precursor Masses

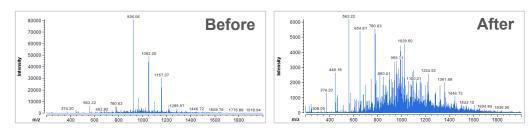






Filter Precursor Masses

- Removes data at defined m/z values.
 - To remove data related to the unfragmented precursors, type a list of masses that are applicable to the molecule of interest.
- Simplifies visualization of MS/MS clusters.



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

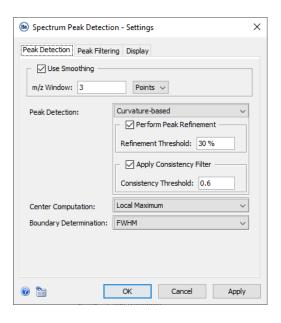
- 1. Open the results of MS/MS Range Restriction in the Data Preparation [Container].
- 2. Undock the MS/MS Range Restriction window.
- 3. Open Filter Precursor Masses from the workflow and specify values.

16 of 33

Spectrum Peak Detection







Use Smoothing

 Temporary smoothing is applied in the background for peak detection.

Peak Detection

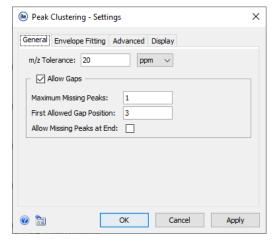
- Curvature-based peak detection is used to resolve shoulder peaks, which is recommended for analysis of data with complex and often overlapping isotopic cluster profiles.
 - Perform Peak Refinement: For detection of subtle shoulder peaks.
 - Refinement Threshold: To increase peak splitting with subtle 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.

17 of 33 © 2024 DH Tech. Dev. Pte. Ltd.

Peak Clustering: General

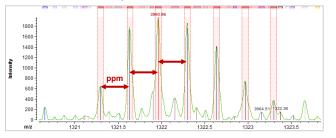






m/z Tolerance

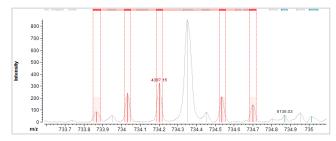
 Specify the maximum window of mass accuracy that each peak must have to be included in an isotopic cluster, related to either the detected or the recomputed monoisotopic mass.



• Note: This setting is not a reflection of the mass accuracy of the mass spectrometer.

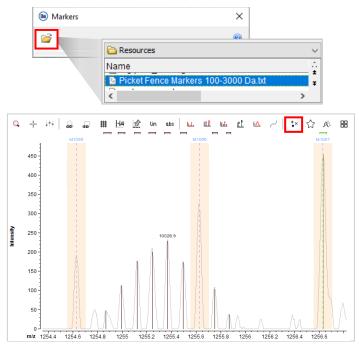
Allow Gaps

 Isotopic envelopes with missing or overlapping peaks can be clustered.



Peak Clustering: 'Picket Fence' Peaks





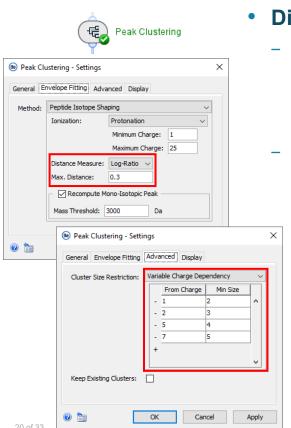
Note: Use the **Picket Fence Markers** file in the **Resources** folder to show the position of the 'picket fence' peaks to help with manual interrogation of isotopic clusters during data review. To load the Marker File included in the template workflow, use the **Show Markers** icon.

- In MS/MS data acquired from large peptides and proteins there is an inherent 'picket fence' laddersignal 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.

19 of 33 © 2024 DH Tech. Dev. Pte. Ltd.

Peak Clustering: Envelope Fitting and Advanced Tab





Distance Measure

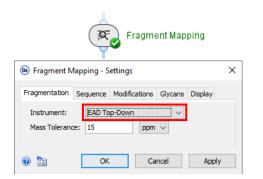
- Compares the measured isotopic cluster profile to the theoretical isotopic profile shape (based on the Averagine model).
 - **Log Ratio**: Treats all peaks in a cluster equally in comparisons of the experimental profile and the theoretical profile.
 - To decrease the required similarity between the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

Variable Charge Dependency

Specify the minimum number of peaks required in an isotopic cluster, based on the charge of that cluster.

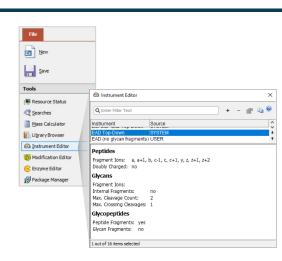
Fragment Mapping: Configure Settings (1)

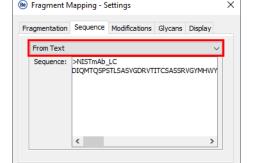




Fragmentation tab:

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





Cancel

Apply

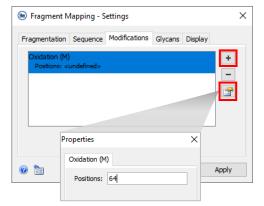
Sequence tab:

- Sequence: Select From Text and type the sequence, or select From Fasta File and add the applicable file.
- To analyze a mAb 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: Configure Settings (2)









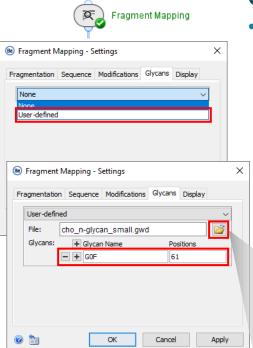
Modifications tab:

- All modifications are fixed.
- To see the list of available modifications, use 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.
 - Specify 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 will show 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.

22 of 33 © 2024 DH Tech. Dev. Pte. Ltd.

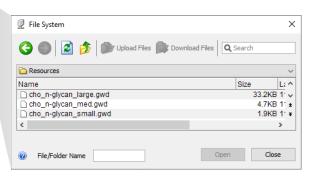
Fragment Mapping: Configure Settings (3)

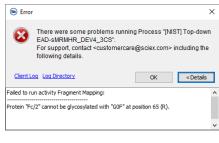




Glycans tab:

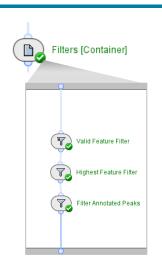
- Select **User-defined** if the sequence contains a glycosylation.
 - File: Browse to Resources in the File System, and then select a glycan library that is applicable to the molecule being investigated.
 - To review or change a glycan library: Browse to File > Tools > Library Browser.
 - Glycans: Type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





Filters: Remove Clusters





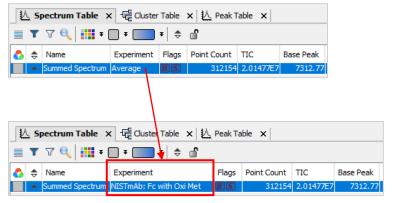
- Valid Feature Filter
 - To remove low-intensity clusters, specify the Validity Threshold.
 - If the most intense peak is below the defined threshold, then the cluster is removed.
- Highest Feature Filter
 - Specify the total number of clusters to keep.
- Filter Annotated Peaks
 - Removes any non-annotated peaks to simplify data visualization in Review Results.
 - To increase sensitivity, decrease the Valid Feature Filter threshold or increase the Keep Top number of clusters in Highest Feature Filter.
 An increase in sensitivity results in:
 - An increase in the number of detected clusters and, thus, in bond coverage.
 - An increased number of false positive identifications.
 - The increase in false positive identifications that results from 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





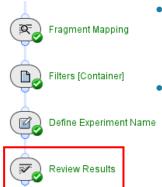


Note: To analyze results from different Protein Sequencing workflows that are saved as *Results Snapshots*, load them together in the Top_Middle-Down_ReviewSnapshots workflow.

- 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 specify 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 specified 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





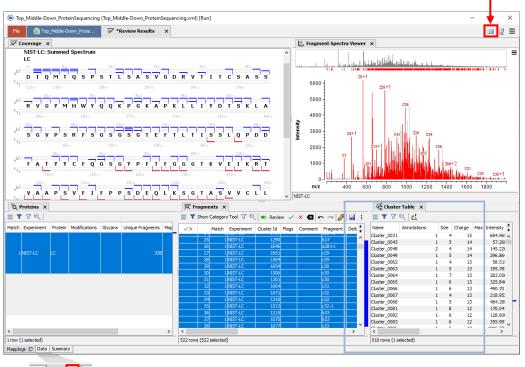
- Activate the **Review** mode, and then **Accept** or **Reject** annotations.
 - To add a comment, either type in the applicable row in the Comment column, or use the icon to add the same comment to multiple rows.
- To apply the changes, click the **Save** icon, and then select **Save and Reload**.



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 new 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 **Data** 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.



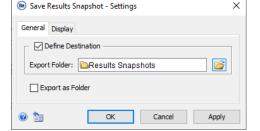
 The Save Result Snapshot creates an sbf file that contain the results of the Top_Middle-Down_ProteinSequencing workflow.

• To use a Save Snapshot activity node:

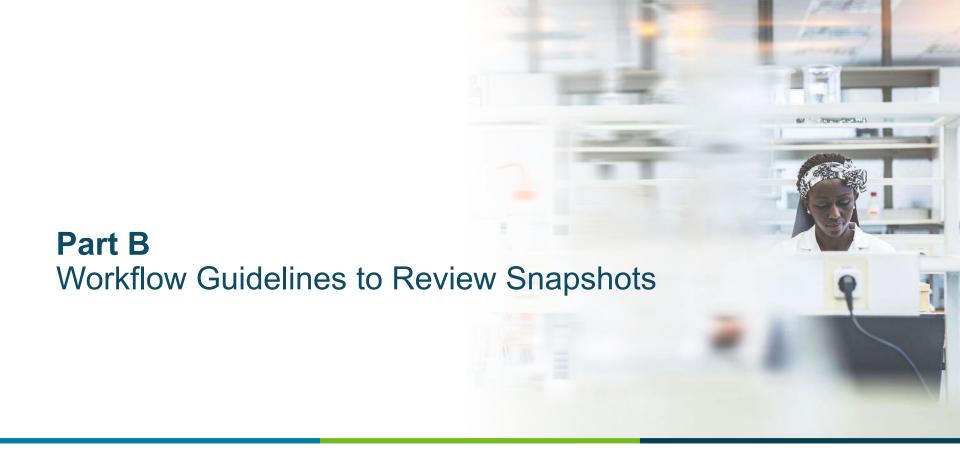


© 2024 DH Tech. Dev. Pte. Ltd.

- Deactivate the **Block** icon.
- Select or add the folders where the shf files will be saved

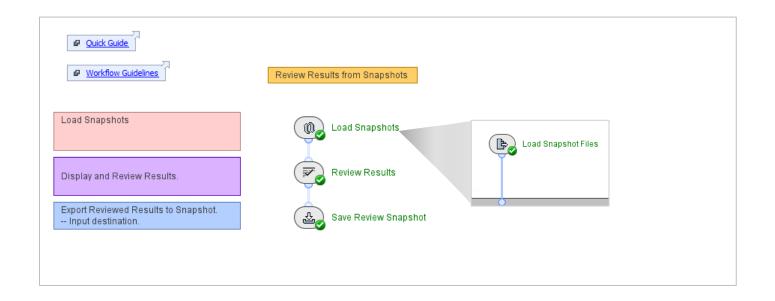


- To review saved data:
 - Select the sbf files to import into the Load Snapshots activity node in the Top_Middle-Down_ReviewSnapshots workflow.
 - Note: Select individual sbf files from the parent folder. Data will not load if the parent folder is selected.



Protein Sequencing Review Snapshots Workflow



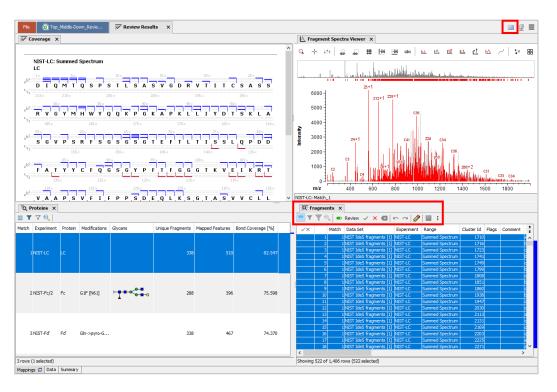


Top_Middle-Down_ReviewSnapshots_Be4.0

30 of 33

Review Snapshots Workflow: Overview





- This workflow can be used to open saved snapshot (sbf) files created from any Protein Sequencing workflow in Biologics Explorer software.
- Use this workflow to analyze previous results from Fragment Mapping searches.
 - For example, to analyze all mAb fragments from an IdeS digest together, or to compare the sequence coverage with different modification locations.
- The same saved layout can be applied here as in the Protein Sequencing workflow.
- Use Save Review Snapshot to save a final reviewed sbf file for each sample.



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.

Trademarks and/or registered trademarks mentioned herein, including associated logos, are the property of AB Sciex Pte. Ltd. or their respective owners, in the United States and/or certain other countries (see sciex.com/trademarks).

© 2024 DH Tech. Dev. Pte. Ltd. RUO-IDV-05-15407-B