

PMI Byos®

User's Manual

June 2020
3.9

Protein Metrics Inc., Cupertino, California, USA

Contents

Introduction 1
Byos Default System Workflows 1
Data flow 2
Byos Menus 2
File 2
Help 3
Project Creation 4
Project Creation Actions 5
Project Creation tabs 6
Samples 6
Sequences 12
Processing nodes 16
Customizing System Workflows 16
Intact, Reduced, and ADC Workflows 16
Sequences and masses tab 17
Sample-protein input tab 25
Processing nodes 26
PTM, PTM (in-silico), HCP, S-S, SVA (C57) – Specific, SVA (C58) – Specific, HotSpot, System Suitability, Oxidative Footprinting and MAM Reference Characterization Workflows 29
Processing nodes 29
MAM Chromatogram Annotation (LC-MS), MAM Chromatogram Annotation (in-silico), MAM New Peak Detection, Comparison Chromatography, Comparison Chromatography (in-silico), Reference Chromatography, Reference Chromatography (in-silico) Workflows 50
Processing nodes 50
Detached Glycans (N-Linked, O-Linked, and IgG) Workflows 67
Glycan table 67
Glycan options 69
De novo Sequencing Workflow 74
Processing nodes 74
My Workflows 77
Saving a Customized Workflow or Portable Workflow 77
Save as Portable Workflow 77
Populate “My Workflows” 77
Create Project to Generate an Analysis and Report 78
Byos UNIFI API Integration 79
Customized Workflows 80

Introduction

Byos Default System Workflows

This manual will focus on the Byos® software program, which is based on the legacy products Byonic™, Byologic®, Byomap™, Intact Mass™, and Supernovo™. Byos is now the default application to open *.byrslt, *.blgc, *.ntms, and *.bmap project files. Byos currently includes several default System Workflows, as shown in Figure 1.

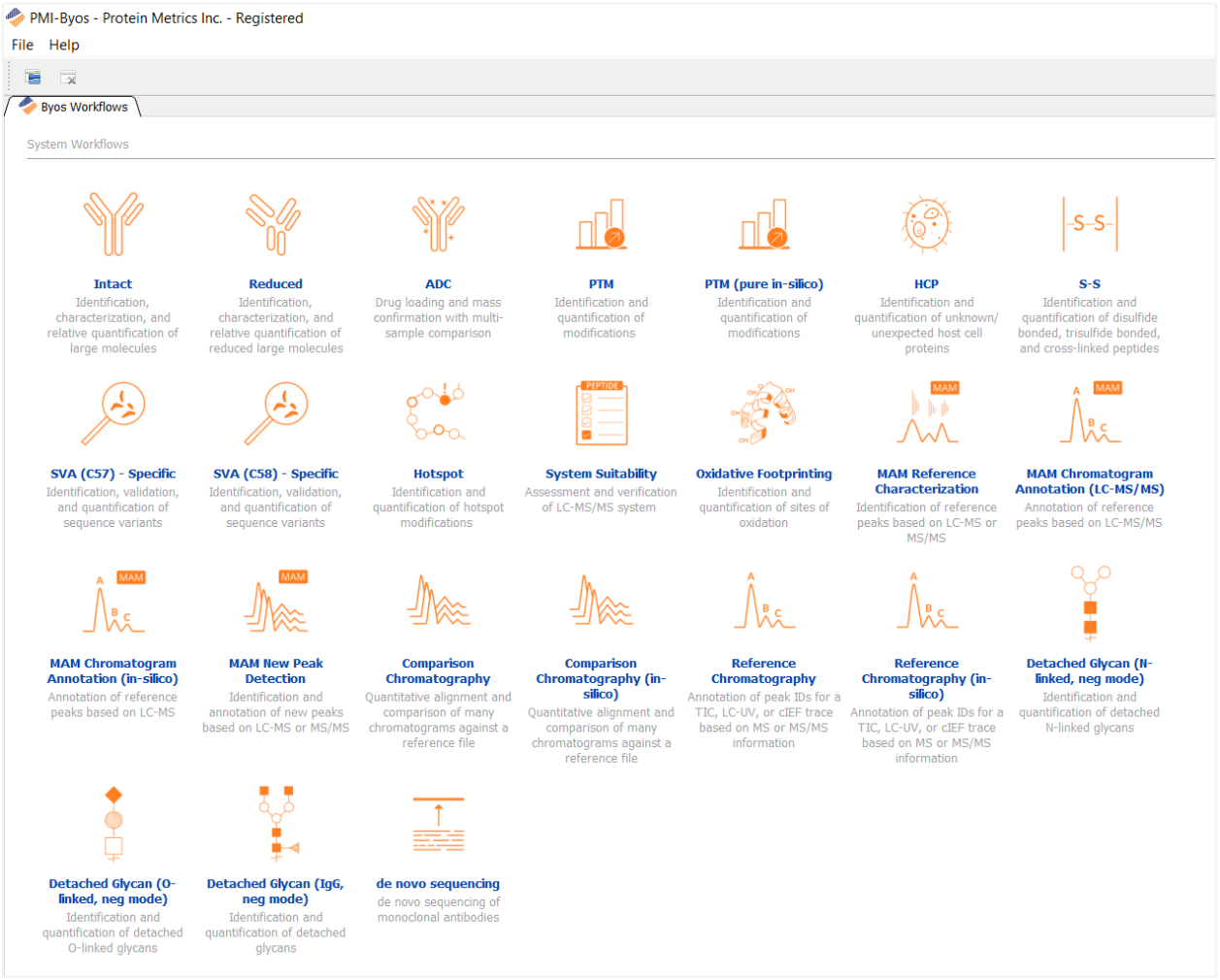


Figure 1: The Byos default System Workflows included with installation.

In this manual, the default System Workflows are separated into five sections:

1. **Intact Analysis** – Intact, Reduced, and ADC workflows
2. **Peptide Analysis** – PTM, PTM (in-silico), HCP, S-S, SVA (C57) – Specific, SVA (C58) – Specific, HotSpot, System Suitability, Oxidative Footprinting and MAM Reference Characterization workflows
3. **Chromatogram Analysis** – MAM Chromatogram Annotation (LC-MS), MAM Chromatogram Annotation (in-silico), MAM New Peak Detection, Comparison Chromatography, Comparison Chromatography (in-silico), Reference Chromatography, and Reference Chromatography (in-silico) workflows
4. **Deatched Glycan Analysis** – Detached Glycan (N-linked), Detached Glycan (O-linked), and Detached Glycan (IgG) workflows
5. **De novo Sequencing Analysis** – De novo Sequencing workflow

Data flow

The user may select any default workflow from the System Workflows section or customized workflow from the My Workflows section. Default parameters for each workflow, including a report template, have been set in the Sequence and masses and Processing nodes tabs.

To create a project, the user will click on a workflow icon, drag and drop their raw data file(s) and a FASTA file, and click **Create Project**. A completed analysis and optimized report will then be available. This data flow is represented in Figure 2 below.

A simplified diagram shows the data flow:

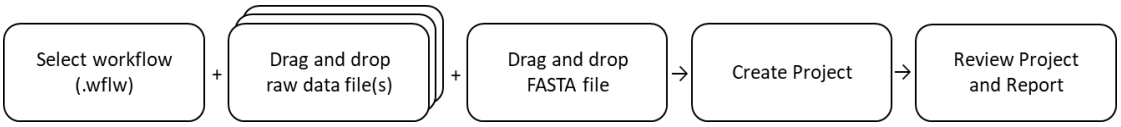


Figure 2: Simplified diagram of the Byos data flow.

Byos Menus

File

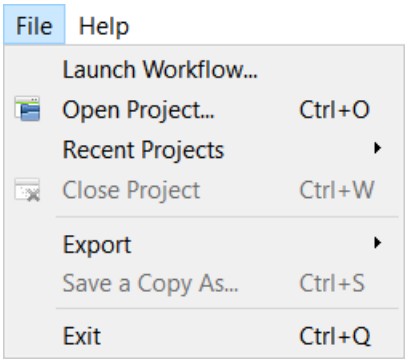


Figure 3: File menu options.

Menu	Sub Menu	Action and Notes
File		
	Launch Workflow	Launch a workflow to begin the creation of a project and its corresponding report. The user will be directed to the directory of default “System Workflows” available within the installer (at C:\Program Files\ProteinMetrics\PMI-Suite\Base\presets\Byos\workflows).
	Open Project	Open a Byos project that has been previously processed.
	Recent Projects	Select from recently created Byos projects.
	Close Project	Close the currently open Byos project.
	Export	Use this functionality to generate a Multi-Document Report. This feature allows the user to combine multiple Byos projects into one comprehensive report. Refer to the Reporting section for additional details.
	Save a Copy As	Save a copy of the currently open Byos project.
	Exit	Exit the Byos program.

Help

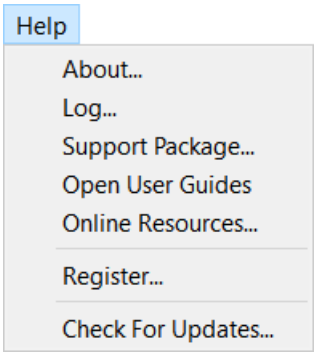


Figure 4: Help menu options.

Menu	Sub Menu	Actions and Notes
Help		
	About	Details on the version number, date, program path, copyright information, and patent details.
	Log	Displays the log file. The user is provided the options to “Open log folder”, “Copy All”, or “Clear”.
	Support Package	Collects all details necessary to assist a Support request. This file can be shared directly with support@proteinmetrics.com.
	Open User Guides	Directs the user to the Manuals, QSGs, Release Notes, Licensing Details, FAQs, etc. included with the installation package. These documents are for Byos along with the Protein Metrics suite of legacy products.
	Online Resources	Directs the user to the online resources available at https://www.proteinmetrics.com/secure-resources/. If Support is up to date, the password will be automatically entered and the user granted access to all application notes, videos, posters, etc.

	Register	Directs the user to the License Activation window. Please refer to the licensing instructions pdf provided with purchase.
	Check For Updates	Directs the user to upgrade if a newer version is available.

Project Creation

When a workflow is launched, the Project Creation window is activated. This is where the user directs Byos to create the project and corresponding report. The windows are shown below in Figure 5, Figure 6 and Figure 7, and options detailed in the following Project Creation Actions section.

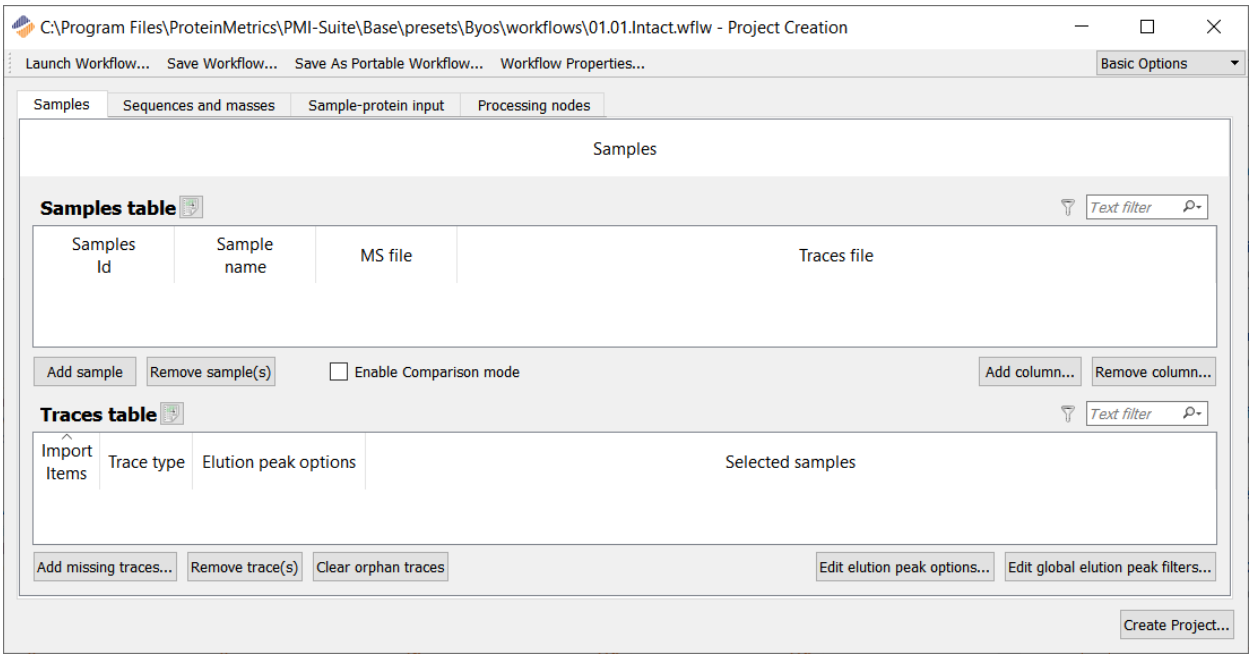


Figure 5: Intact Analysis project creation window.

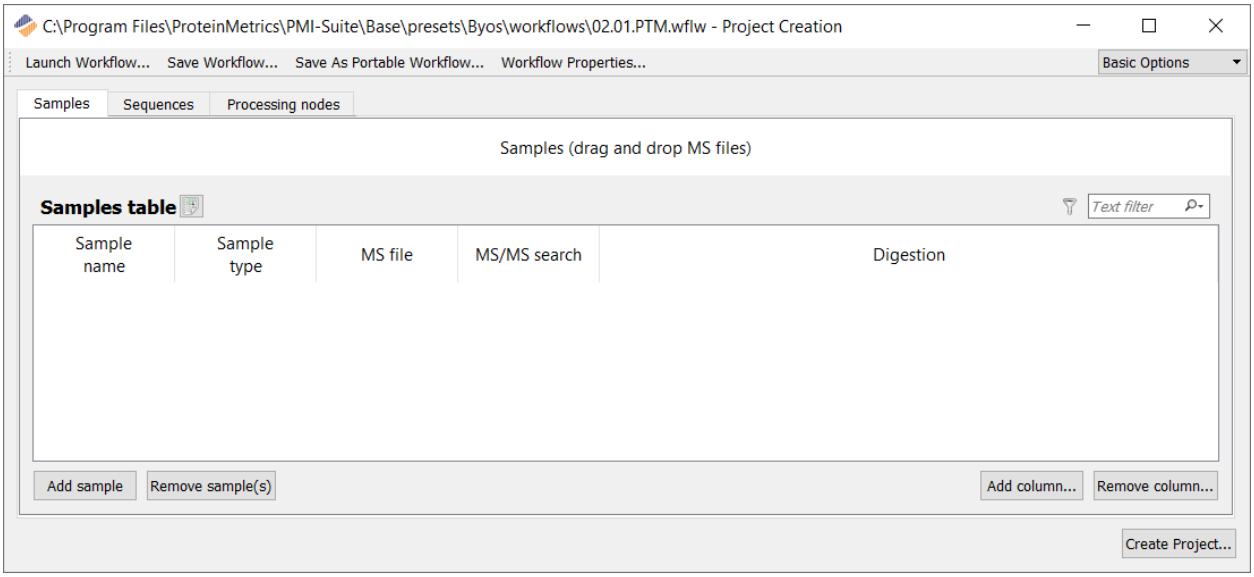


Figure 6: Peptide Analysis project creation window.

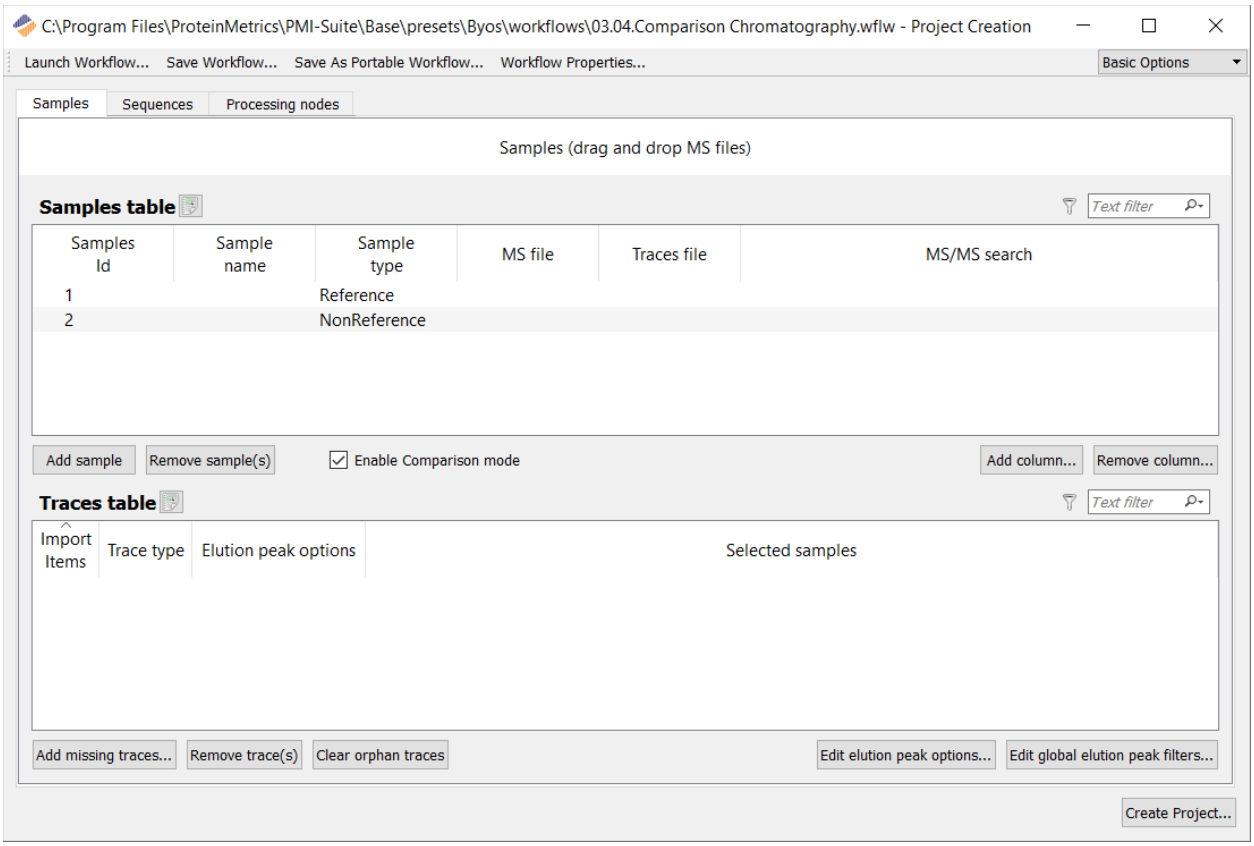


Figure 7: Chromatogram Analysis project creation window.

Project Creation Actions

Launch Workflow

This enables the user to open a different workflow from the one currently open.

Save Workflow

This enables the user to save the current workflow (*.wflw) file. This can be saved within a portable workflow (*.wflwp) folder or as a stand-alone .wflw file. It can be saved within the “My Workflows” folder, detailed in the section **Populate “My Workflows”**.

Save as Portable Workflow

This enables the user to move a workflow and all related components necessary for processing to another system. This process creates a folder that includes the workflow (.wflw), report template (.rptc), layout or filter (.ini), databases (.txt), etc. files. The user can then save changes to any *.wflwp component within this folder. It can be saved within the “My Workflows” folder, detailed in the **My Workflows** section.

Workflow Properties

This enables the user to edit the icon, name, and description of the workflow currently open. The icons are available to select from using a drop-down menu, as shown below:

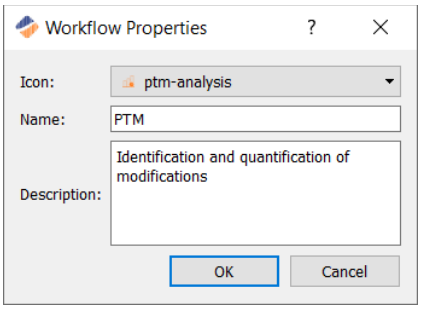


Figure 8: Workflow Properties – Icon, Name, and Description.

Please note each workflow has an icon, representative of the type of project being created. Byos also includes a “default” icon, if that is instead preferred to use. This is shown below:

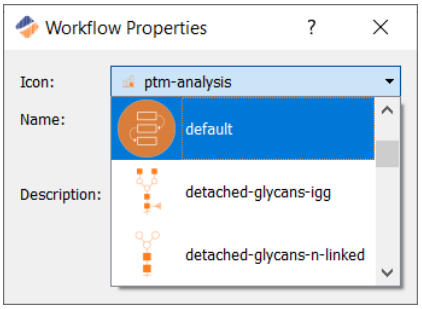


Figure 9: A default icon can also be assigned to the workflow.

Project Creation tabs

Samples

In the Samples tab, the user drags and drops mass spectrometry raw data files to be processed within the project. The process is the same for all workflows. The PTM workflow is shown below:

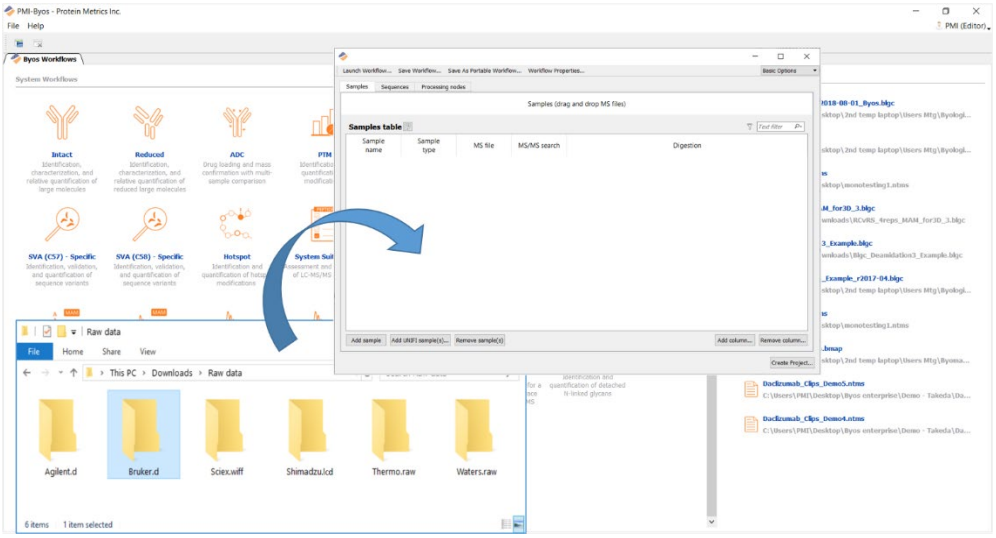


Figure 10: Drag and drop the raw data file(s) to be processed into the Samples tab.

The user can also use the Add row button to manually add a sample and then direct to the raw data file. **Remove sample(s)** will remove the highlighted line.

The user can use **Add column** to add an additional column to the default settings available. These include the options shown below:

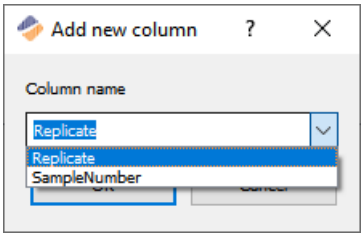


Figure 11: Add new column option.

The user can also create any custom names by typing in this dialog. The user can remove any column using the **Remove column** button.

For Intact Analysis and Chromatogram Analysis workflows, the MS file populates both the **Samples table** and the new **Traces table** at the bottom.

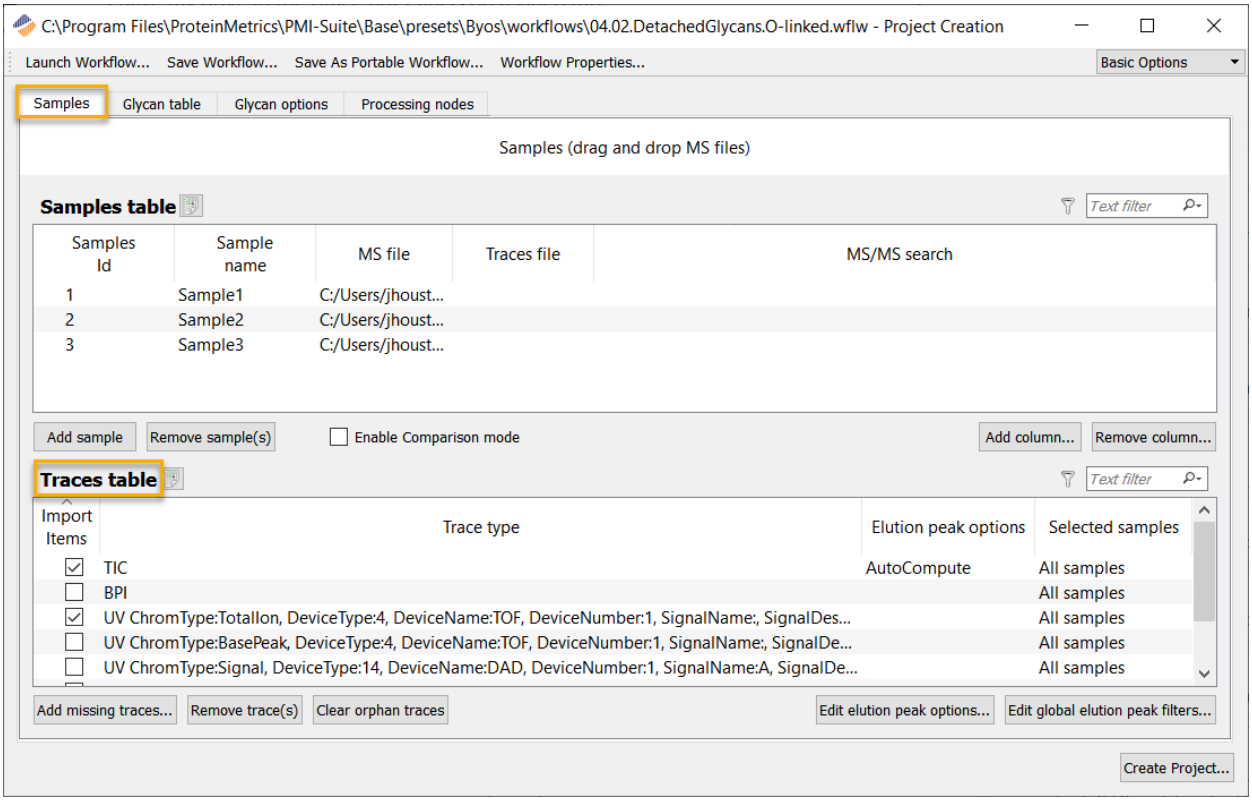


Figure 12: Samples tab with populated Traces table

Selected (not checked) traces can be removed by clicking **Remove trace(s)**. Removed traces can be restored by clicking **Add missing traces**, check the trace to restore and click **OK**. To associate specific traces to the samples, double-click the text in the Selected samples column, then click the ... button to open the Edit samples selection dialog:

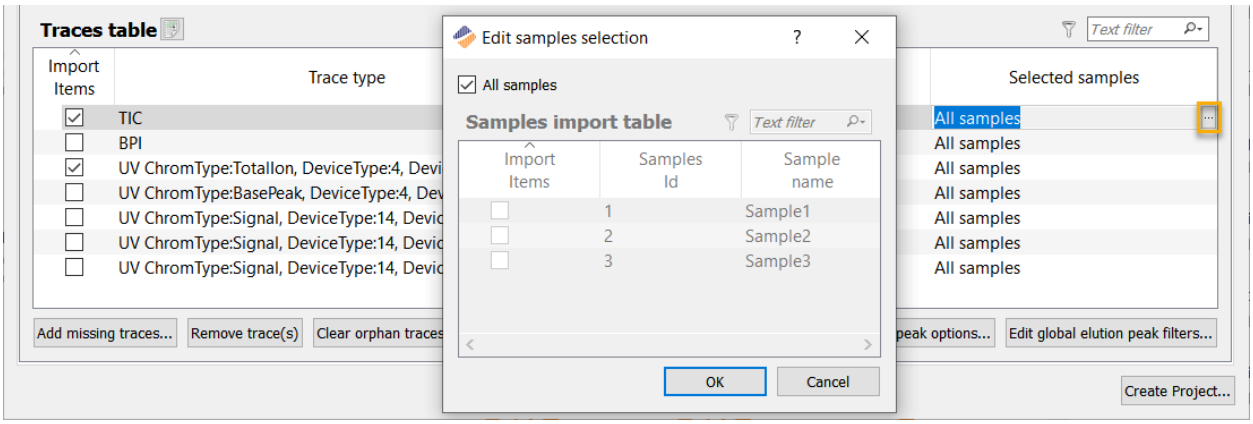


Figure 13: Edit samples selection dialog to associate traces with sample names

To associate the selected trace with individual samples, uncheck **All samples**, check those samples, and click **OK**. To import a trace from a *.csv file, drag the file into the **Traces files** column in the Samples table. Alternatively, double-click in the Samples table row under the Traces file column, click ⋯ and browse to the sample file. Click **Open**. The trace file is added to the sample name and the trace is added to the Traces table:

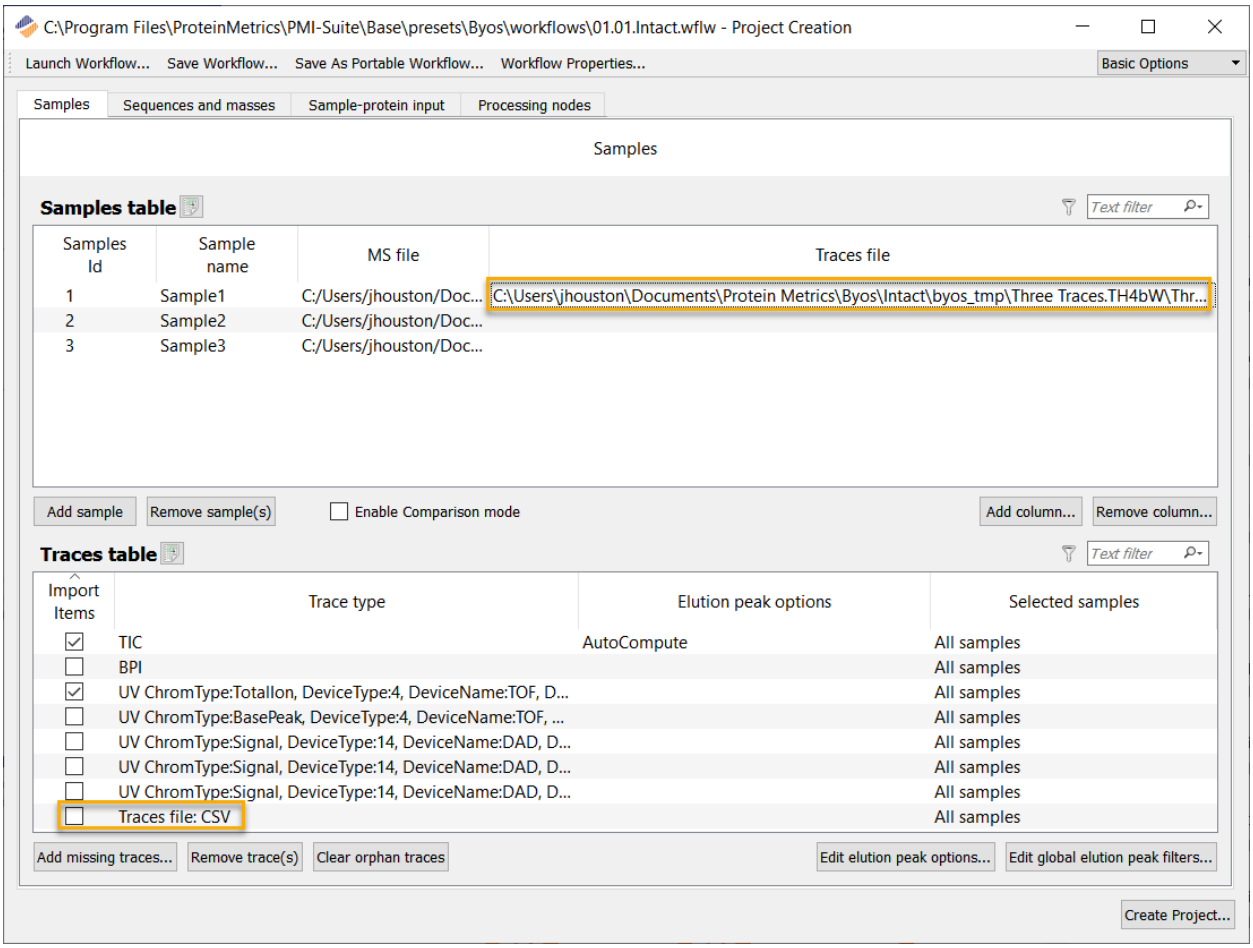


Figure 14: Traces imported from a *.csv file

To control how the elution peaks are processed, click **Edit elution peak options** at the bottom:

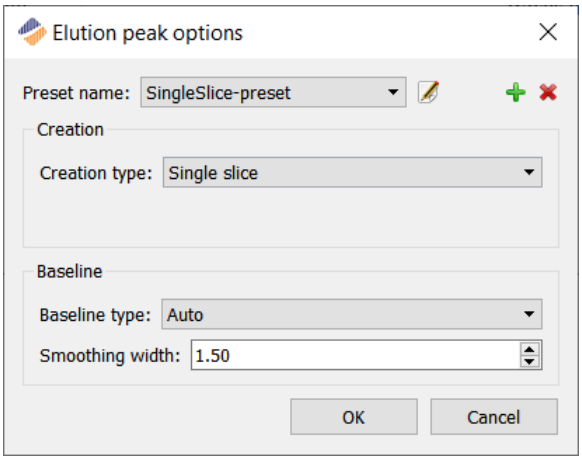


Figure 15: Elution peak options

At the bottom of the Elution peak options dialog are the Baseline parameters. The baseline type can be set to **Auto** or **Flat**. Baseline smoothing width is also set here.

The **Elution peak options** dialog contains five default peak processing presets:

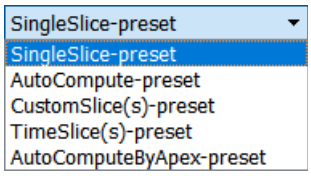


Figure 16: Default elution peak presets

Select a processing preset and edit the peak parameters. To create a custom preset, click the ✚ button, add a preset name, and click **OK**. Next, select a creation type and edit those peak parameters. The new preset is added to the list. Click the ✎ button to edit a selected preset name. Click the ✖ button to delete a selected preset.

The default and custom peak processing presets can use one of four **Creation types**:

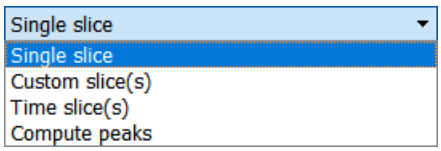


Figure 17: Elution peak Creation types

These Creation types use the following peak processing methods:

- **Single Slice** processes the entire chromatogram as single elution peak.
- **Custom slice(s)** divides elution peaks by user defined start and end times:

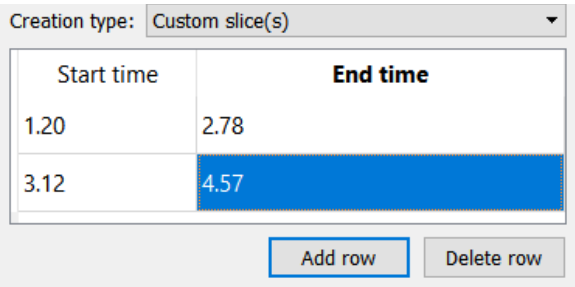


Figure 18: Elution peak Custom slice(s) option

Click **Add row** and enter start and end times for each slice. To remove a slice, select it and click **Delete row**.

- **Time slice(s)** divides elution peaks by regular time intervals defined by the user:

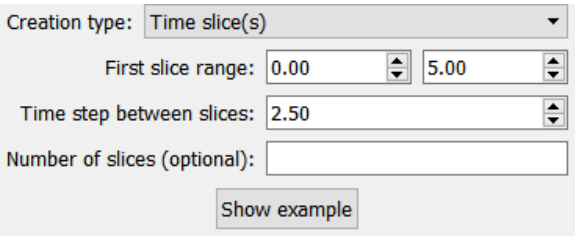


Figure 19: Elution peak Time slice(s) option

  - **First slice range** sets the start and end times for the first slice. This also sets the width for all time slices.
  - **Step** is the time delta that sets the start of subsequent time slices by adding the step value to the start of the preceding time slices. For example, the values above define the first slice as 0 - 5 min, the second slice as 2.5 – 5 min, etc., through the end of the chromatogram.
  - **Number of slices** sets the maximum number of defined slices. If blank, the last slice is determined by the end of the chromatogram time range.
  - **Show Example** displays an example time slice configuration with definitions and explanations:

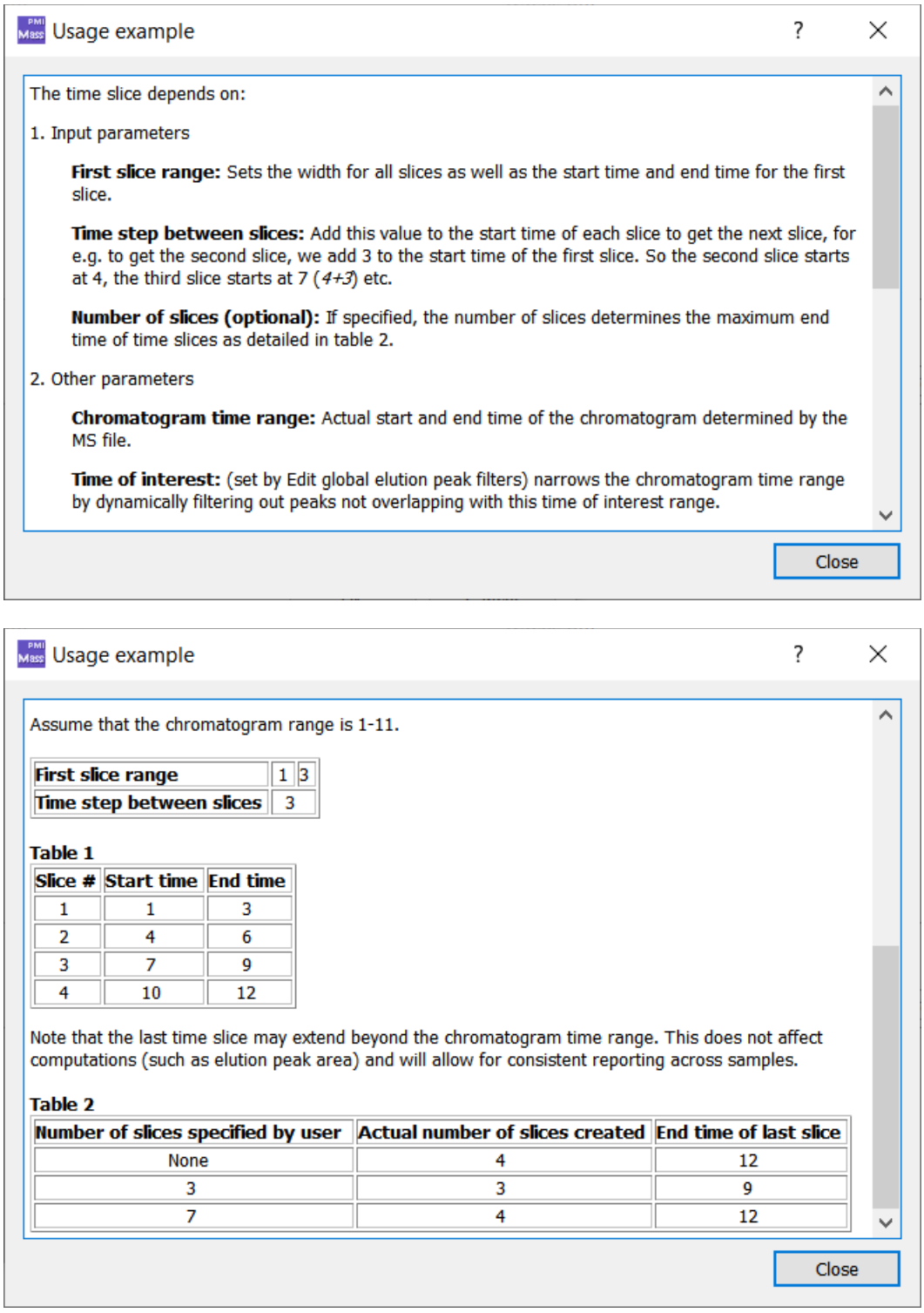


Figure 20: Time slice(s) Usage example

- **Compute peaks** automatically determines the elution peaks based on their minimum width, after smoothing:

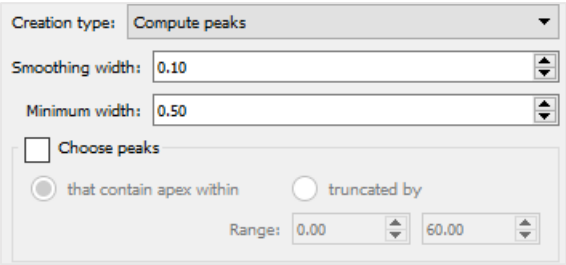


Figure 21: Elution peak Compute peaks option

- Choose peaks by apex, when checked, limits the detected peaks to the specified time range.
- Choose peaks truncated by, when checked, ensures all peaks within the specified time range will be chosen. Additionally, the first and last peak will be trimmed to fit within the specified time range.

Starting with Intact Analysis v. 3.8, individual trace filters are replaced with a single peak filter across all traces. To set peak property filters for all the traces, click **Edit Global elution peak filters**:

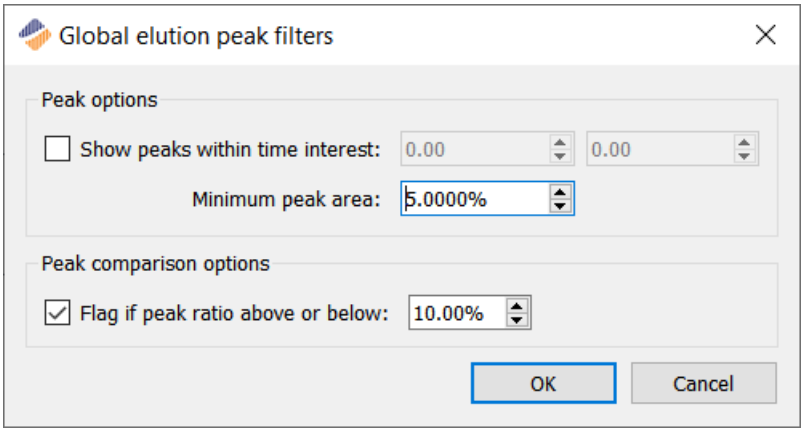


Figure 22: Global elution peak filters

Check **Show peaks within time interest** to manually set minimum and maximum times. **Minimum peak area** and a peak ratio threshold flag can be set in this dialog.

The new Traces table and elution peak settings are backwards-compatible. Traces associated with samples in projects from software versions before 3.8 will correctly populate the Traces table in the current version. Elution peak settings and presets from versions before 3.8 will correctly migrate into the current version. However, Intact and Chromatogram Analysis projects created in versions 3.8 and above will not correctly open in older versions.

Sequences

Sequences is where the user enters protein sequences to be investigated. The user can drag and drop the FASTA file, browse for the file, or add a row and direct Byos to the desired file. The user will then be prompted to select the **Protein candidates** (heavy chain, light chain, etc.) to be considered for processing. The PTM workflow is shown below:

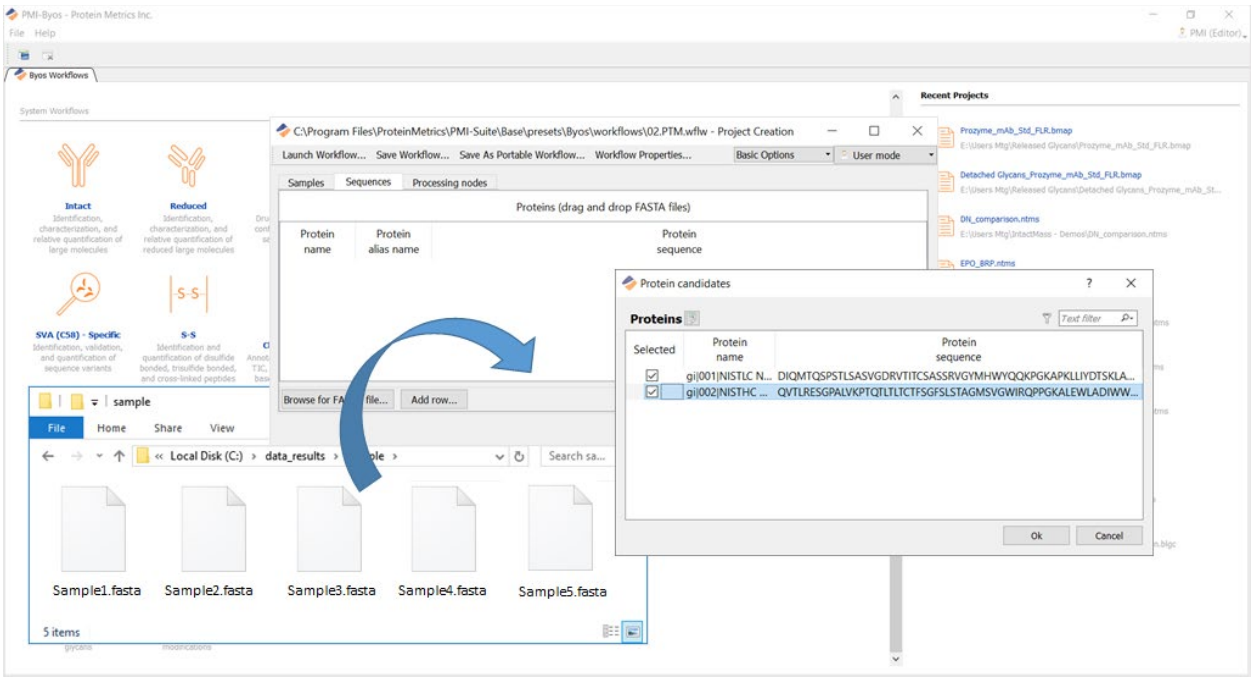


Figure 23: Drag and drop the FASTA file to be processed into the Sequences tab.

NOTE: In the HCP workflow, the user must instead point to a protein database. The user can either drag and drop as shown above or direct Byos to a database file. This is done by clicking in the space to activate the light blue "…" button. The user is then prompted to select a file, as shown:

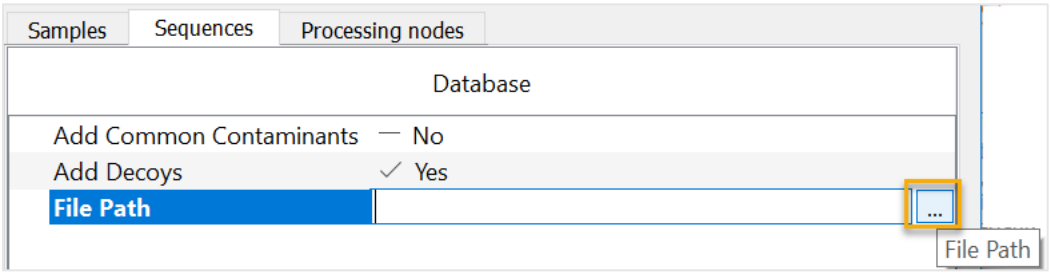


Figure 24: Click to activate the file path prompt.

NOTE: In the **System Suitability** workflow, the Sequences tab is populated by default to include the Pierce RT standard. This is shown below:

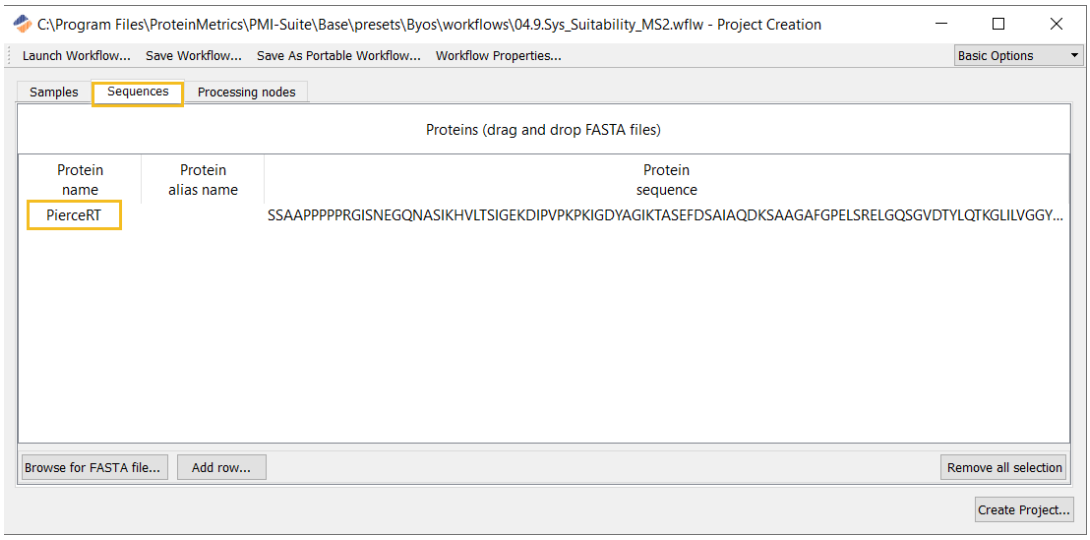


Figure 25: The Pierce RT standard is included in the Sequences tab of the System Suitability workflow.

The user can replace the Pierce RT standard with another protein if desired. The user can change this by clicking to highlight the line and click **Remove selected** (shown below). Another protein could be brought in using drag and drop (shown in Figure 10).

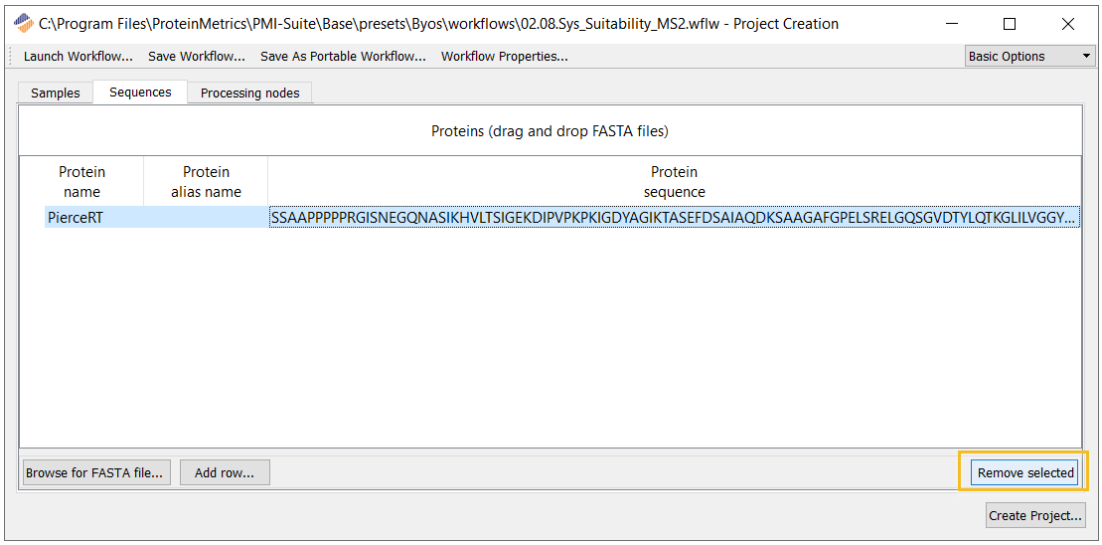


Figure 26: The user can change the default standard protein sequence.

NOTE: In the Intact, Reduced, and ADC workflows, this tab is named “Sequences and masses”. The user can add chains to the Chains table and based on these chains, add rows to the Sequences table. The user enters a protein sequence either by dragging and dropping a FASTA file (shown below), manually typing it in (shown in Figure 28), or selecting from a FASTA file (shown in Figure 29).

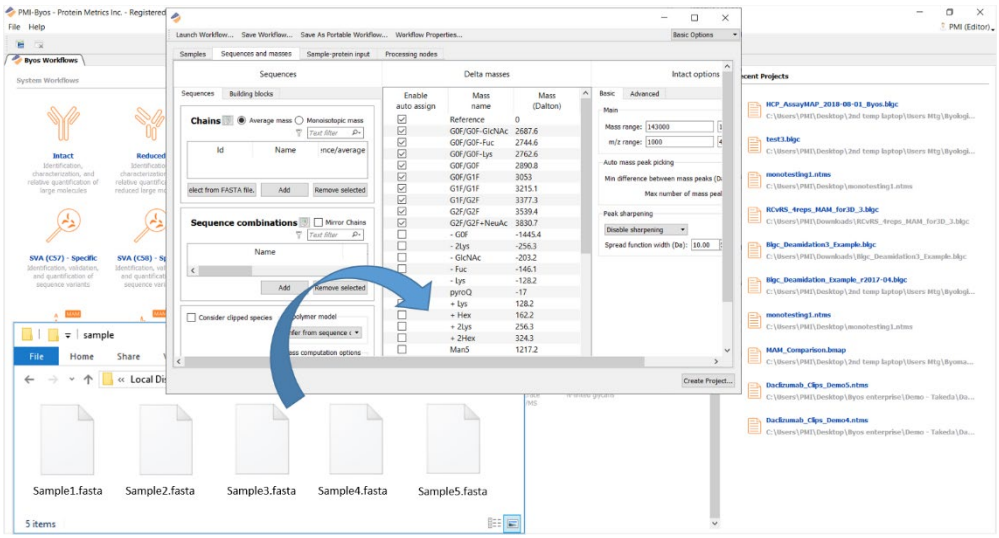


Figure 27: Drag and drop the FASTA file to be processed into the Protein and masses tab.

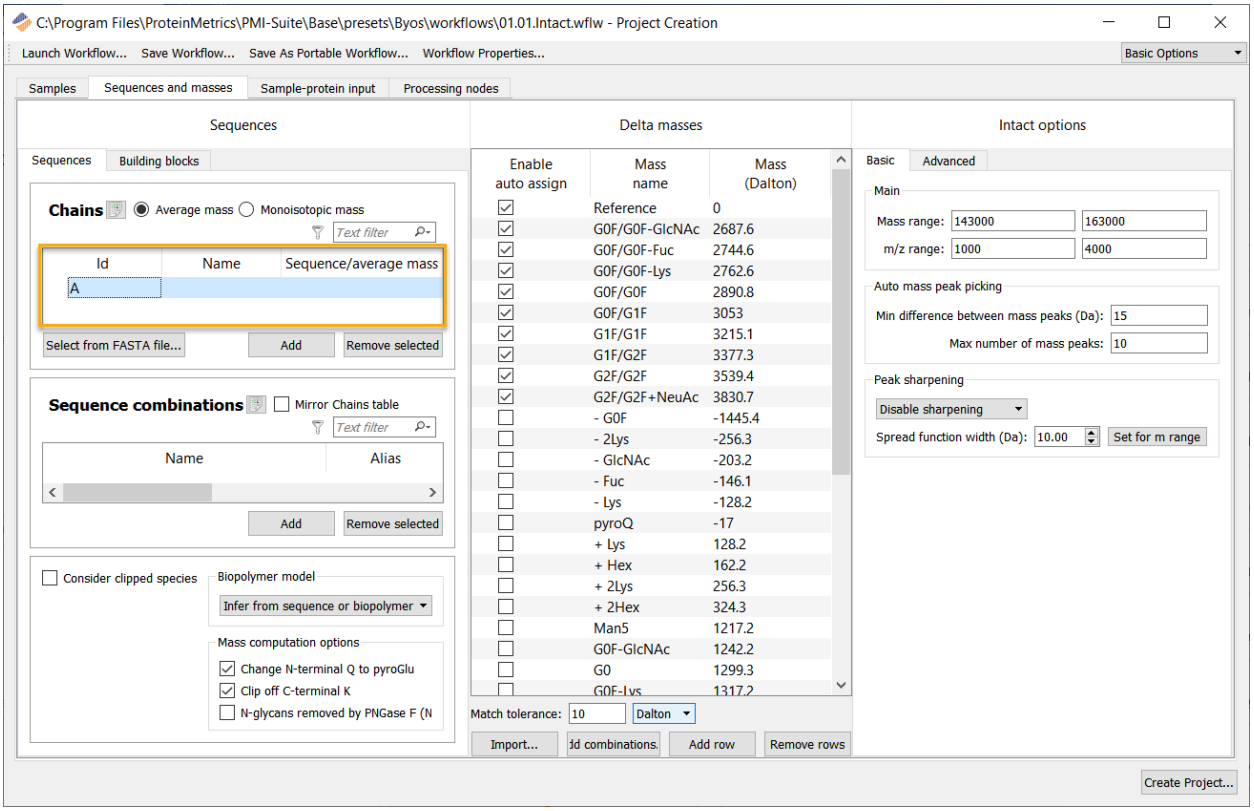


Figure 28: Click **Add** to add a row and manually enter the protein data (Id, Name, and Sequence/mass).

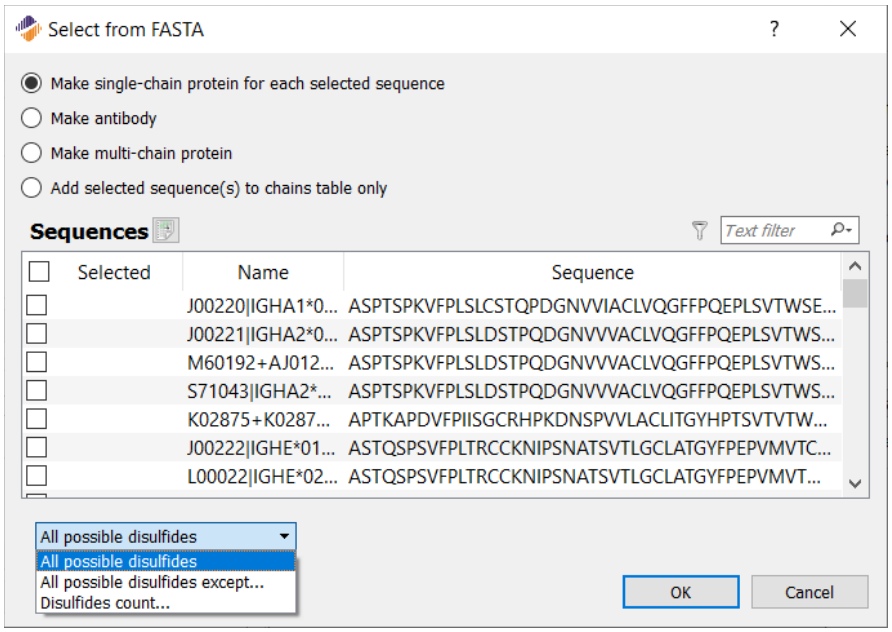


Figure 29: Select from FASTA. Select the appropriate protein and disulfide option.

Processing nodes

Processing nodes is where the user specifies the processing parameters to be applied during project creation and report generation. Each workflow is populated with default values, designed for the user to review once and save for future projects. This empowers every member of a team, scientist within a lab, or colleagues across a project to complete the exact same analysis and generate identical reports regardless of Byos user-skill level, scientific expertise or experience, or location. Please note the next section will only have to be completed once for each analysis type and the workflow used for perpetuity.

Customizing System Workflows

The user should review the parameters populated within the default set of System Workflows to generate customized My Workflows.

Intact, Reduced, and ADC Workflows

The following section is relevant to analyses based upon the Intact, Reduced, and ADC workflows. These workflows all have the Samples, Proteins and masses, Sample-protein input and Processing nodes tabs. The default parameters to review are included within the **Sequences and masses** and **Processing nodes** tabs, as shown below:

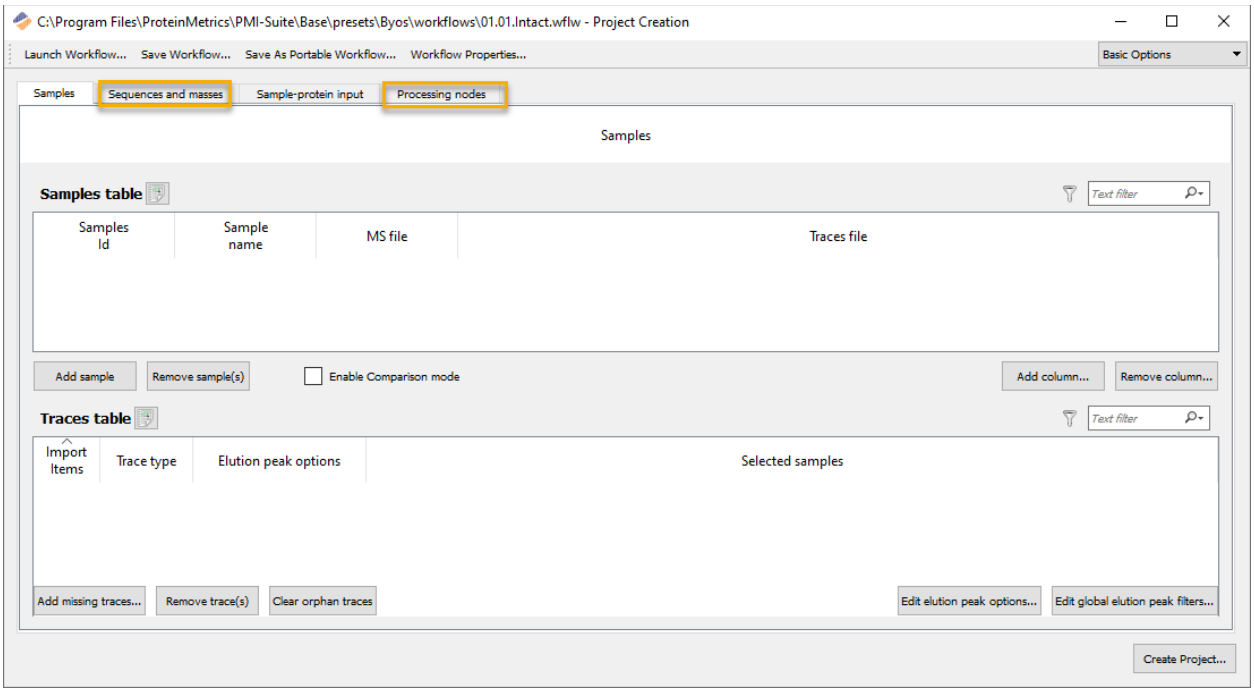


Figure 30: Sequences and masses and Processing nodes tabs

Sequences and masses tab

Sequences Sub-tab

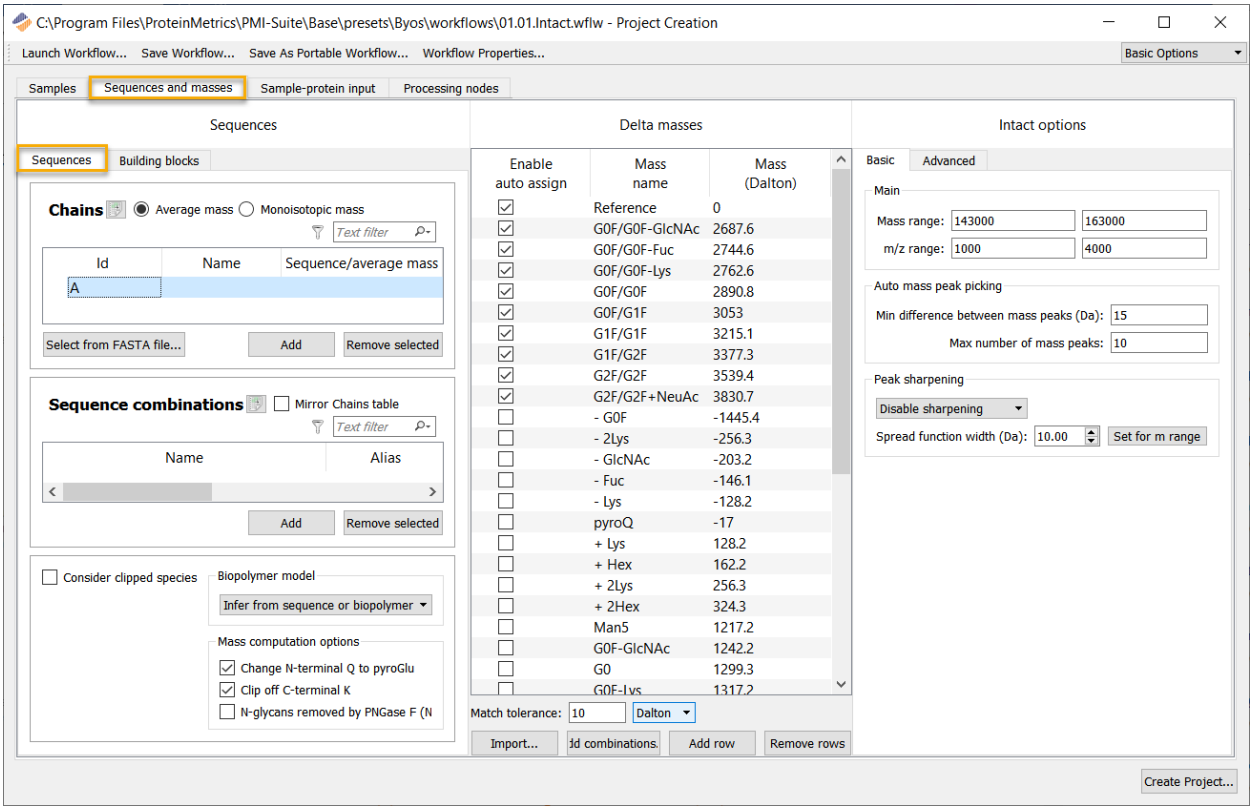


Figure 31: Sequences sub-tab.

The user can add chains to the Chains table and based on these chains add rows to the Proteins table. The user enters a protein sequence either by dragging and dropping a FASTA file (shown in Figure 27), manually typing it in (shown in Figure 28), or selecting from a FASTA file (shown in Figure 29).

Users can designate peptide chain masses without sequences as average mass or monoisotopic mass. In the Chains section, a radio button selects between the two types of masses and the mass header updates accordingly:

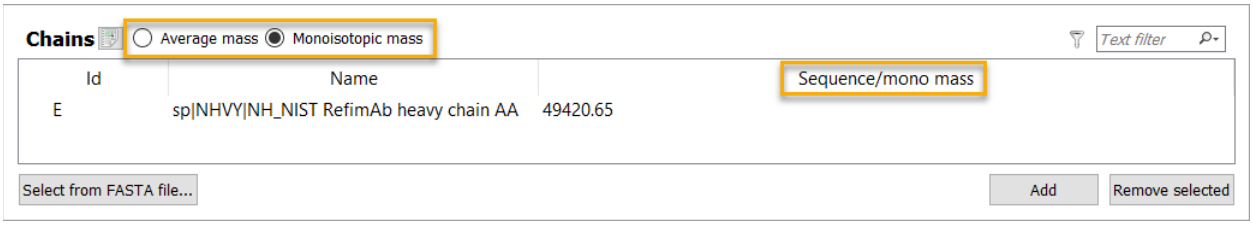


Figure 32: Average or monotopic mass assignment

The software automatically computes the average and monoisotopic mass for each selected FASTA entry, and these masses are then available for automatic peak assignment. The default computation assumes that Cys residues are disulfide bonded, but this default can be changed by the user. Protein sequences can be arbitrary if the user inputs the average or monoisotopic masses, overriding the computed masses. Average or monoisotopic masses give the software reference masses to assign mass peaks automatically based on mass deltas.

The **Mirror Chains table** check box populates the content from the Chains table into the Proteins table. This is shown below:

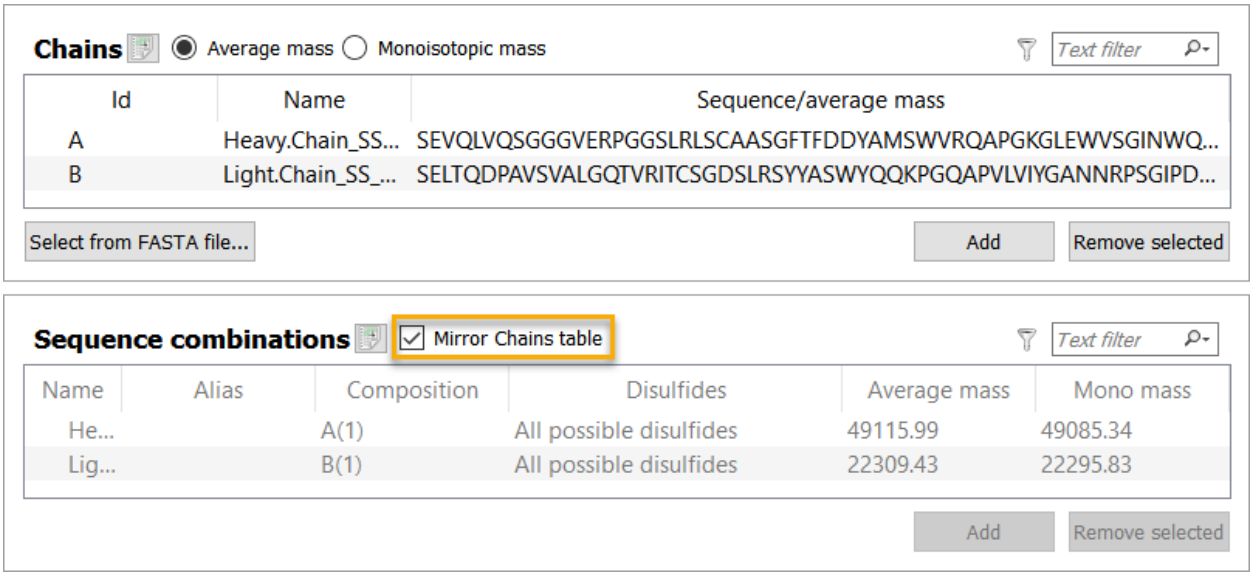


Figure 33: Mirror Chains table option.

For each row in the “Chains” table, a corresponding row is added to the “Proteins/protein complexes” table (with single-chain composition and all possible disulfides). The user can also set the number of disulfides within the Proteins/protein complexes table. The “Disulfides” window is activated by double clicking within the box within the table, as shown below:

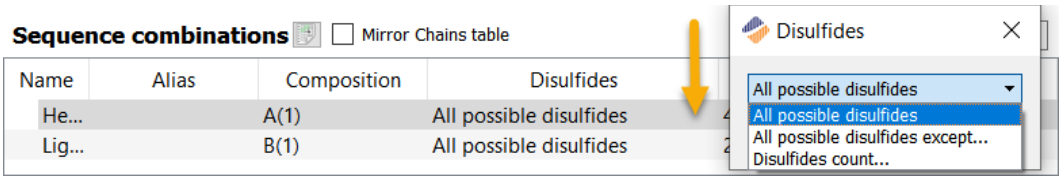


Figure 34: Disulfides.

To set a specific number, select “Disulfides count” and enter a value, as shown below:

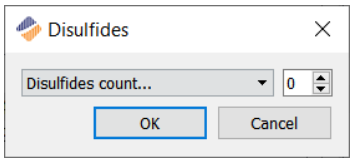


Figure 35: Disulfides count.

The user can also direct the software to consider clipping. As shown in Figure 36, select the box to activate – there are the options: 1. “Clip at specific sites” or 2. “Clip everywhere”. Click **Add** to enter specific residues, if that radio button is selected.

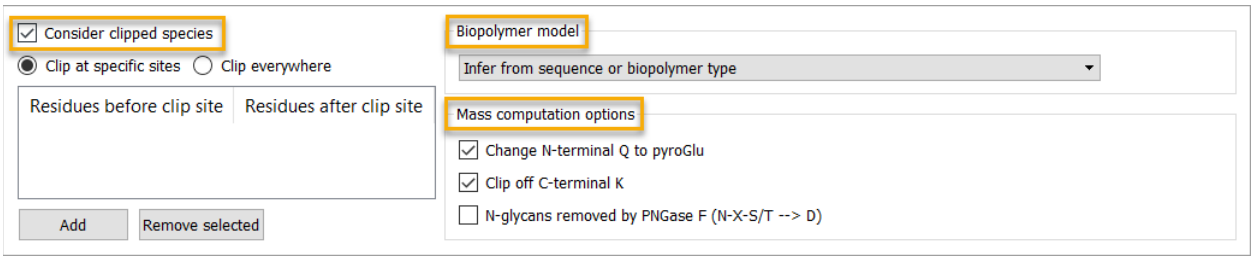


Figure 36: Consider clipped species, Biopolymer model and Mass computation options

Users can select the kind of biopolymer model used to convert average mass to monoisotopic mass and monoisotopic mass to average mass:

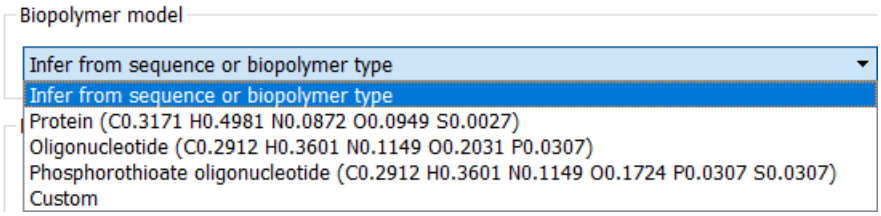


Figure 37: Biopolymer model choices

Biopolymer model options include:

- **Infer from sequence or biopolymer type** is the default selection, which uses the standard model that Intact Analysis has used to calculate mass.
- **Protein** (C0.3171 H0.4981 N0.0872 O0.0949 S0.0027) uses the displayed averagine formula to calculate mass.
- **Oligonucleotide** (C0.2912 H0.3601 N0.1149 O0.2031 P0.0307) uses the displayed formula to calculate mass.
- **Phosphorothioate oligonucleotide** (C0.2912 H0.3601 N0.1149 O0.1724 P0.0307 S0.0307) uses the displayed formula to calculate mass.

- **Custom** opens a cell to the right of the selection to add a custom formula to calculate mass.

Note that the formula values refer to atom counts (molar ratios), not mass ratios.

There are also a few Mass computation options shown in Figure 36. The user should select which to consider when calculating the masses.

Within the **Protein and masses** tab, there is also the **Building blocks** tab, as shown below:

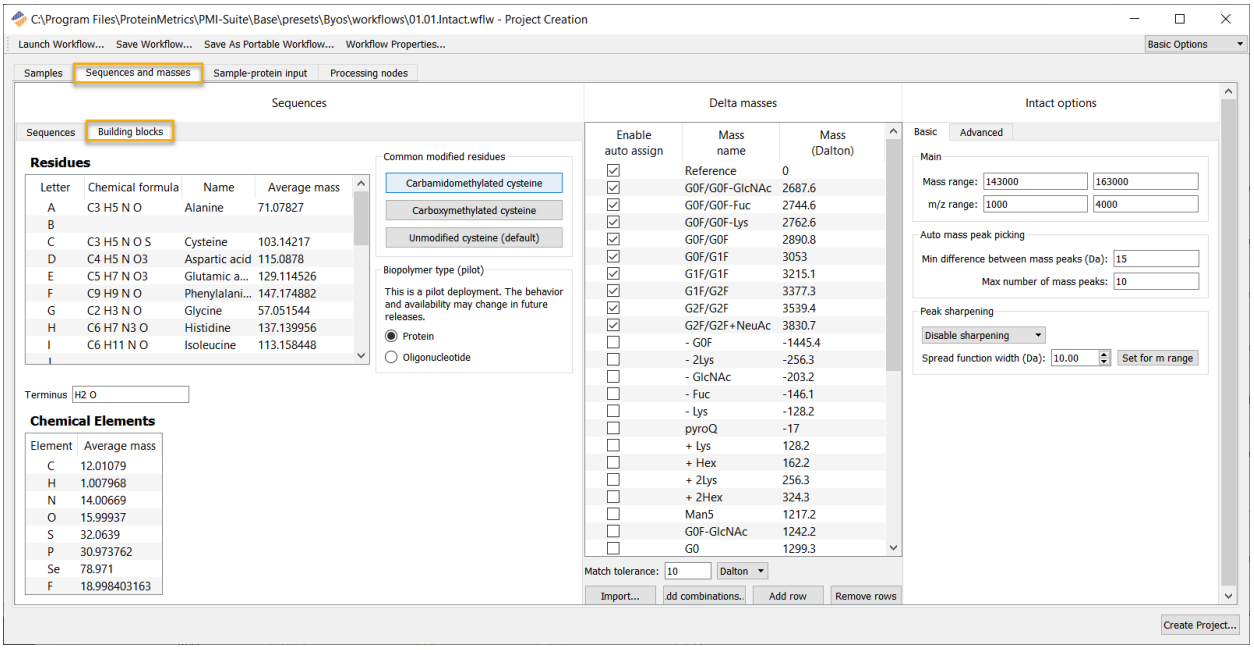


Figure 38: Building blocks sub-tab.

There are several parameters to consider in this tab. As shown in the three figures below, the cysteine modification can be clarified by selecting amongst the 3 options. The chemical formula will be adjusted accordingly.

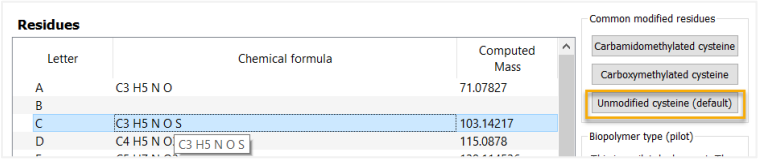


Figure 39: Unmodified cysteine (default).

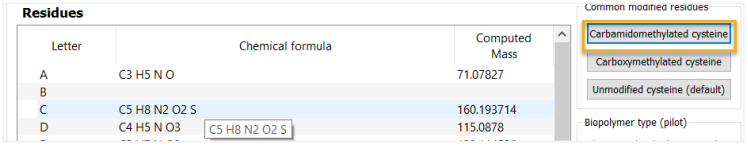


Figure 40: Carbamiodomethylated cysteine.

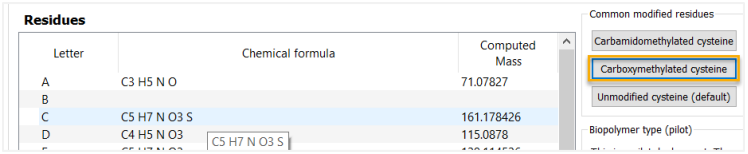


Figure 41: Carboxymethylated cysteine.

The user can also select the Biopolymer type (default Protein and Oligonucleotide; please note this is a pilot deployment). The content of the Residue table will be updated accordingly, as shown below:

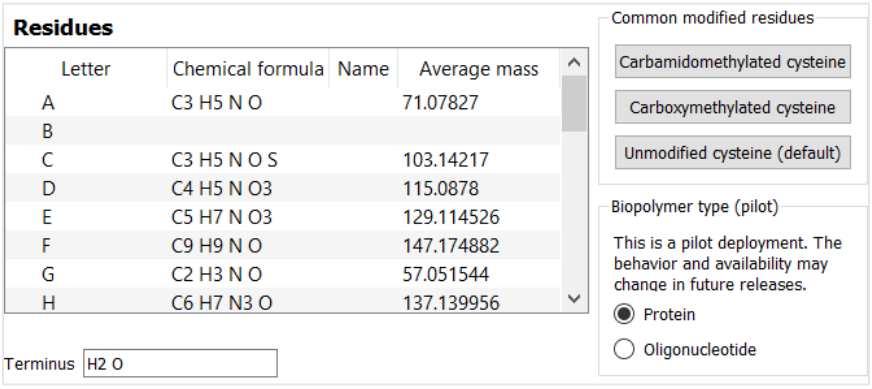


Figure 42: Biopolymer type: Protein

When Biopolymer type Oligonucleotide is selected, the content of the Residue table and the default Terminus value will be updated accordingly, as shown below:

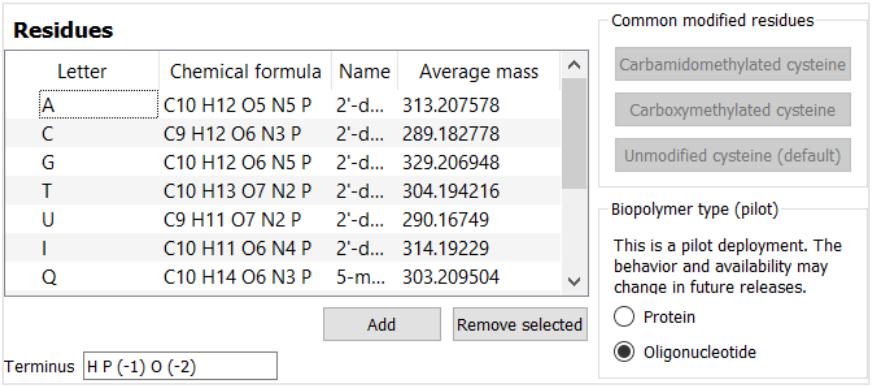


Figure 43: Biopolymer type: Oligonucleotide

The user can also specify the average mass of each chemical element, as shown below:

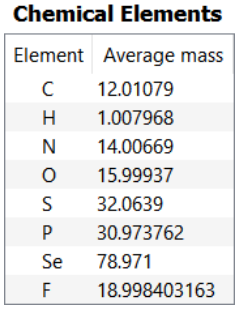


Figure 44: Chemical Elements

Delta masses

The Delta masses in the center of the Sequences and masses tab (shown in Figure 38) gives a table of likely mass differences between observed peaks and the reference mass (either input by the user or

computed from an amino acid sequence, as completed in the Proteins section). Byos uses this information to assign peaks based on mass differences from Reference peaks. Check each group to include it for peak assignments. The mass names and Dalton values can be edited by selecting and typing new values. The Delta mass table is also useful in automated mass assignments when processing many data files.

The user can also add delta masses not available by default by clicking **Add row** and entering a mass name and mass manually. This is shown below in 5.

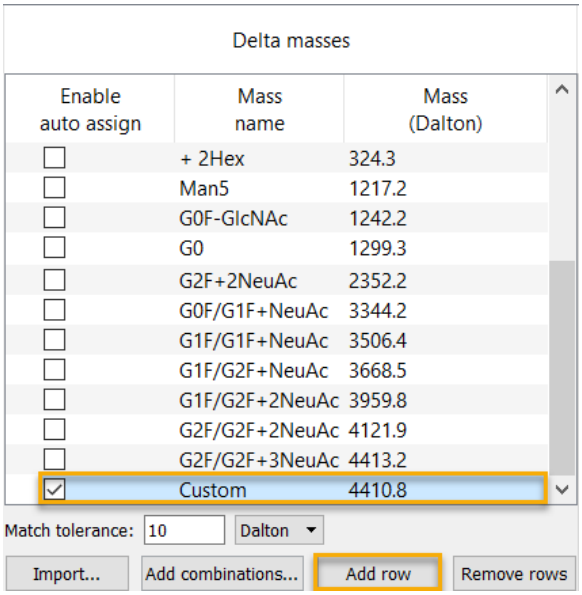


Figure 45: **Add row** to add a delta mass.

Any row can be deleted by highlighting and clicking the **Remove rows** button. Select multiple rows by using Ctrl-click or Shift-click. This is shown below in 6.

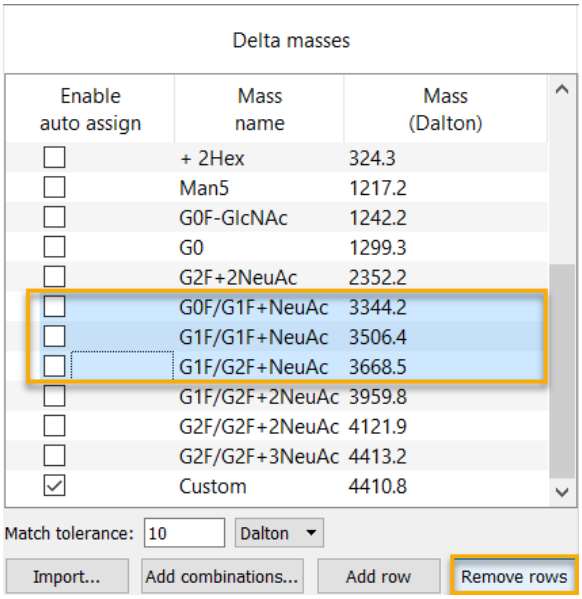


Figure 46: Remove rows.

The user can also import a custom list (.CSV) by clicking **Import** and selecting a saved delta mass table in CSV format. This is shown below:

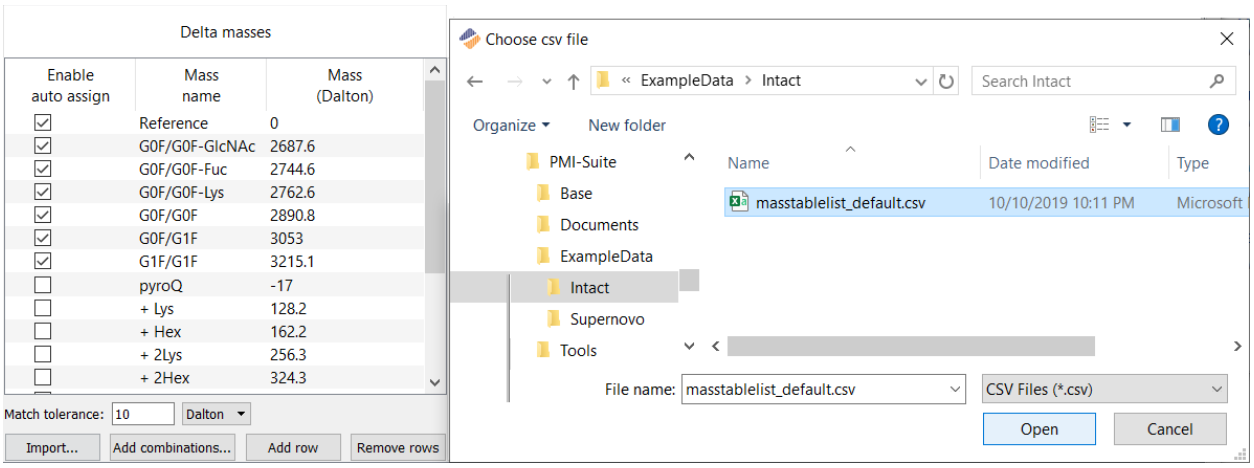


Figure 47: Import a custom mass deltas list.

Intact options

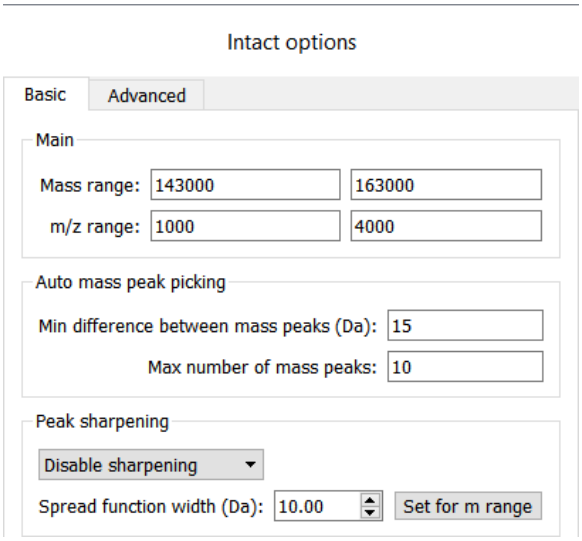


Figure 48: Intact options.

The Intact options section on the right side of the Protein and masses tab (shown in Figure 38) includes Basic and Advanced entries for Mass range, m/z range, mass peak picking, and peak sharpening parameters that are applied to all elution peaks.

The Basic tab contains the primary settings used in computing deconvolved masses. Mass range and m/z range set ranges for neutral masses and m/z, respectively. Mass range defines the range of neutral masses displayed in the Deconvolved mass spectrum. m/z range defines the segment of the MS1 spectrum used to compute neutral masses.

Within the Auto mass peak picking parameters, Min difference between mass peaks and Max number of mass peaks control peak picking. Min difference between mass peaks prevents the peak picker from picking multiple points on top of a ragged or isotope-resolved mass peak. Max number of mass peaks sets a limit on the number of picked peaks.

The Peak sharpening dropdown controls an optional “super-resolution step” that sharpens peaks beyond what is seen in the data. It deconvolves a “point spread function” to give super-resolved mass peaks and turn shoulders into separate peaks. The Spread function width sets the expected width of a peak when peak sharpening is enabled. A reasonable width is the square root of the peak mass in kilodaltons, for

example, 12 for a spectrum from 140,000 to 150,000 Da. A too-narrow width will not sharpen much, and a too-wide width might split peaks. Peak sharpening on mass spectra that span a range of more than about 20,000 Da is not recommended, because no single width will be optimal for the entire range. Peak sharpening should also be avoided on isotope-resolved mass spectra.

The **Advanced** options mainly concern resolution. The user will only need to use this tab if the aim is to produce isotopically resolved neutral mass spectra.

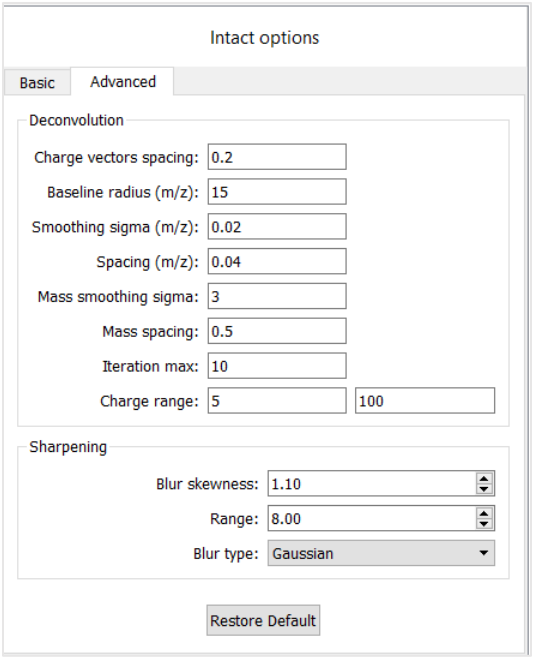


Figure 49: Advanced options.

Charge vectors give the charge assignment probabilities for each small interval of m/z points. Deconvolutions of most MS1 spectra will be almost exactly the same with any **Charge vectors spacing** from 0.2 to 1 m/z units (Thomsons), but a narrow spacing of 0.1 may give better results on isotope-resolved MS1 spectra with interleaved signals, and a wide spacing of 2 may give better results for native MS with broad m/z peaks.

Intact Analysis removes an m/z baseline before deconvolution; this step and removal of charge one or stop-list peaks are the only steps in the algorithm that do not conserve ions. **Baseline radius** controls the stiffness of the baseline. A baseline radius of 8 gives a flexible baseline that will cut into m/z peaks broader than 8 Thomsons; this will often give better visual separation of neutral mass peaks, but may distort peak areas. A baseline radius of 30 will give a stiffer baseline that cannot cut into m/z peaks narrower than 30 Thomsons. An even larger value, 100 or more, may be needed for native MS. The default value of 15 is a compromise.

Spacing (m/z) controls the spacing of sample points in the m/z spectrum. The raw MS1 data is represented as a continuous piecewise-linear function that can support any spacing of sample points, but m/z spacing finer than the finest spacing in the original data will slow the computation without adding resolution. Reasonable values for Spacing (m/z) are in the range 0.005 – 0.05 for QTOF instruments, which have almost the same resolution at all m/z; the low setting of 0.005 would be appropriate for Bruker maXis and the higher setting of 0.05 for older instruments with lower resolution. For Orbitrap, the resolution depends upon m/z; the setting of 0.005 shown above is for isotope-resolved 25 kDa masses (antibody subunits). For native MS on Exactive EMR with m/z's in the 5000 – 10,000 range, a spacing of 0.1 is fine enough.

Smoothing sigma is typically set to the same value as Spacing (m/z), but a larger value can be helpful for producing an appropriately smoothed neutral mass spectrum with less smoothing at lower mass and more smoothing at higher mass.

Mass spacing controls the spacing of points in the neutral mass spectrum. To preserve isotopic resolution, spacing should be set to 0.1 or even 0.05. If the MS1 spectrum does not have isotopic resolution, or isotopic resolution is not needed for analysis, mass spacing in the range 0.2 to 1 is best for target molecules below 200 kDa. Spacing of 10 Da or more is best for targets above 300 kDa. For mass spectra without isotopic resolution, **Mass smoothing sigma** in the range of 2 – 5 will smooth jittery peaks in the range 20 – 200 kDa; larger values will be needed for larger masses. For mass spectra with isotopic resolution 0.1 will work.

Iteration max set to 10 will work for most purposes. A larger value, for example 20 or 30, can be helpful for lower signal-to-noise spectra that take longer to converge.

Charge range is best set to a wide range, in which case the charge range will be implied by the mass and m/z ranges. The default range of 5 – 100 covers most applications, but 5 will need to be reduced for deconvolutions with mass range starting below 10 kDa, and 100 increased for targets that may have charges above 100.

Sharpening uses the Spread function width set in the Basic tab to deconvolve the data. **Blur skewness** controls the asymmetry of the point spread function; the default value of 1.1 means that the right tail has sigma (standard deviation) 10% bigger than the left tail. 1.2 gives even more tailing; 1.0 gives a symmetric point spread function. **Range** sets the length of the tails in standard deviations; a small value of 5 or 6 may work better in the case of Lorentzian point spread. **Blur type** has two choices: Gaussian (skinny tails) and Lorentzian (fat tails). Lorentzian should use a slightly smaller sigma than Gaussian.

Sample-protein input tab

The **Sample-protein input** tab allows for association of protein sequences with the samples that contain them:

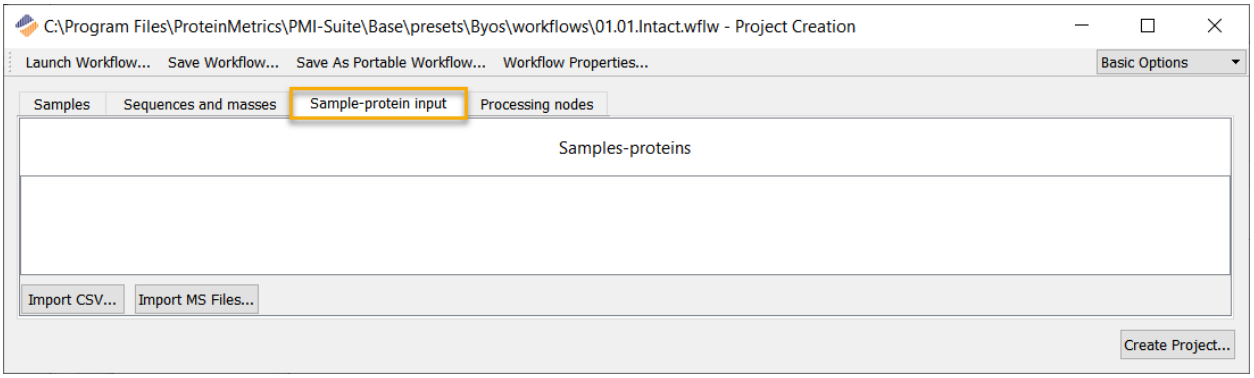


Figure 50: Sample-protein input tab

Sample-protein associations can be imported from *.csv files and from MS files. To create csv files from other projects, use **File > Export > Generate MS path template CSV**. This capability is useful for making a single Intact Analysis project with many different samples.

Processing nodes

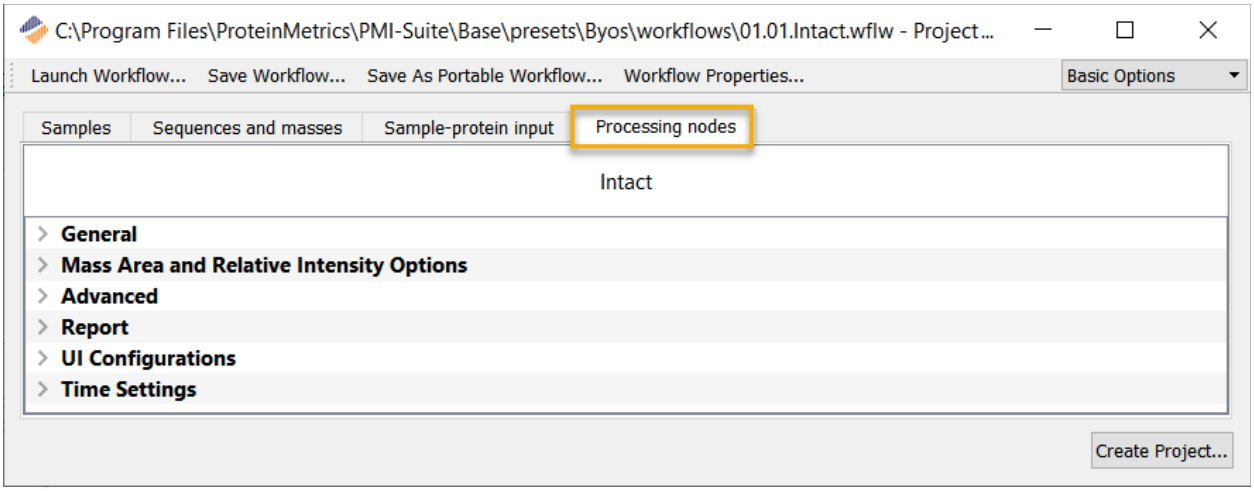


Figure 51: Intact Processing nodes tab.

- **General**

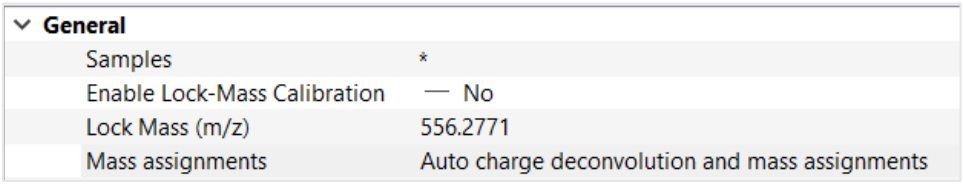


Figure 52: General parameters.

  - **Samples** - The "*" character applies all parameters to all samples dragged and dropped into the Samples tab.
  - **Enable Lock-Mass Calibration** can be set as yes or no.
  - **Lock Mass (m/z) sets** the calibrant m/z value. Several are available using the drop-down menu or the user can type in a numerical value. If empty, no calibration will be applied. The user can select:

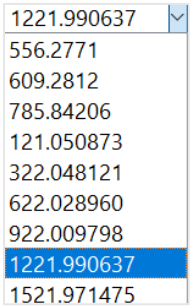


Figure 53: Lock mass drop-down values.

  - **Mass assignments**

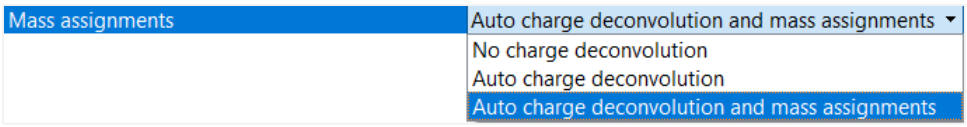


Figure 54: Mass assignments.

The Mass assignments parameter allows the user to turn on/off charge deconvolution and automatic mass assignment. The default selection is Auto charge deconvolution and mass assignments. For new projects, especially those with longer chromatography times, significant computation time may be saved by creating the project with the setting No charge deconvolution selected. The summed m/z spectra can then be viewed before deciding which elution peaks warrant deconvolution. Similarly, charge deconvolution without mass assignments will let the neutral mass spectra be viewed before mass peak assignment.

- **Mass Area and Relative Intensity Options**

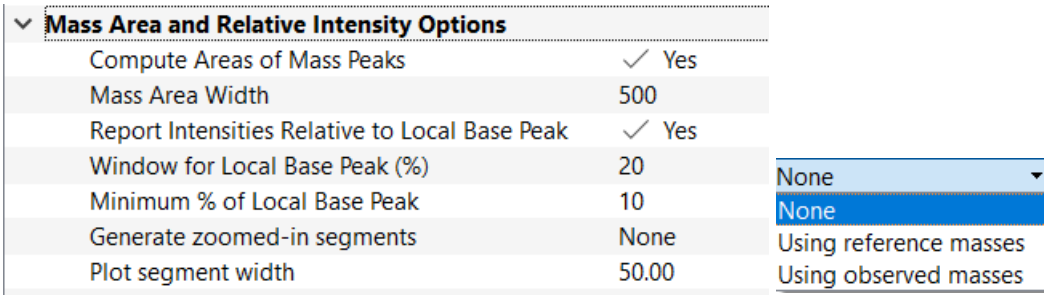


Figure 55: Mass Area and Relative Intensity Options

  - **Compute Areas of Mass Peaks**
  - **Mass Area Width**

    Compute peak area within a band around each mass, defined by the mass area width value
  - **Report Intensities Relative to Local Bas Peak**

    Report intensities based upon the two below parameters
  - **Window for Local Base Peak (%)**

    Set tolerance for local base peak (e.g. mass within +/- 20%),
  - **Minimum % of Local Bas Peak**

    Filters out masses below a certain % of local base peak
  - **Generate zoomed-in segments**

    The user has 3 options: None, using reference masses, Using observed masses (per highest local base peak).
  - **Plot segment width**

    Segments can be set to be automatically generated around reference/observed masses for deconvoluted mass spectrum and MS1 plots during project creation.

- **Advanced**



Figure 56: Advanced.

Several advanced commands can be applied to Intact Analysis processing by adding in the **Advanced configuration** text box. Please refer to the **Advanced Commands** section of the **Intact Analysis Manual**. The Ionization Mode has been removed from an advanced command to within the UI. The user can select between Positive and Negative using the drop-down menu.

- **Report**

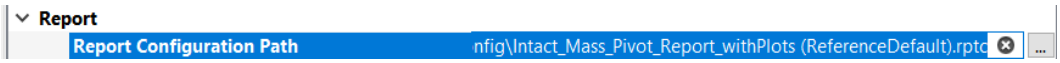


Figure 57: Report Configuration Path

Each Byos default workflow includes a report template created by our Customer Success team that is optimized for the specific type of analysis. If the user prefers a customized report template, they can direct Byos to this file using the light blue "…" button. They will be prompted to select a file.

- **UI Configurations**

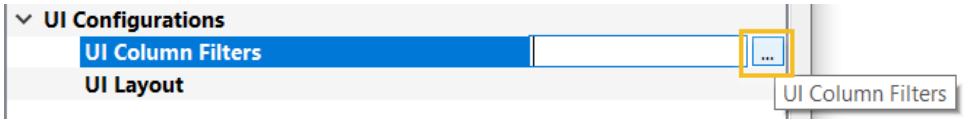


Figure 58: UI Configurations – UI Column Filters and UI Layout.

The user has the ability to import column filters as well as layout files. This is designed to standardize analyses across all users, labs, and sites. The user can direct Byos to the preferred file for each using the light blue "…" button. The user will be prompted to select a file.

- **Time Settings**



Figure 59: Time Settings

  - **Start time of interest** and **End time of interest** control the time limits within which the computation of the baseline and peaks are completed. The default values for both are set to 0.00, which means filtering will not be applied. The user can set them to apply the additional processing options.
  - **Alignment max time** sets the maximum alignment value between plots (for example, UV and TIC). This value will limit the allowed alignment time between the two signals. The default value is set to 1.00.

- **Peak Construction options** (older versions of Byos)

  The Peak Construction option parameters formerly found in Byos are now set directly in the Samples tab with the buttons **Edit elution peak options** and **Edit global elution peak filters**.

PTM, PTM (in-silico), HCP, S-S, SVA (C57) – Specific, SVA (C58) – Specific, HotSpot, System Suitability, Oxidative Footprinting and MAM Reference Characterization Workflows

The following section details parameters in the **Processing nodes** tab for the combination of MS/MS IDs and Quant. These parameters are therefore relevant to the HCP, HotSpot, PTM, PTM (in-silico), SVA (C57) – Specific, SVA (C58) – Specific, S-S, MAM Reference Characterization, System Suitability, and Oxidative Footprinting Workflows.

NOTE: Parameters specific to certain workflows (including SVA and System Suitability) are called out.

Processing nodes

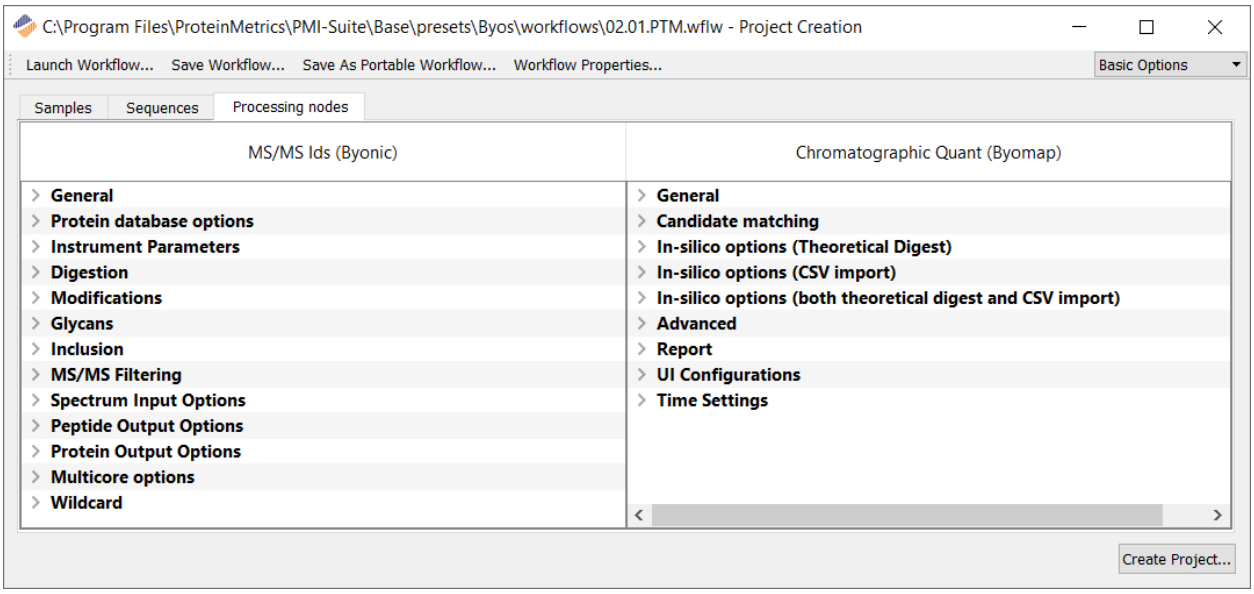


Figure 60: Peptide Analysis Processing nodes.

MS/MS Ids

NOTE: These parameters are not part of the PTM (in-silico) workflow. Please refer directly to the below Quant section.

- **General**

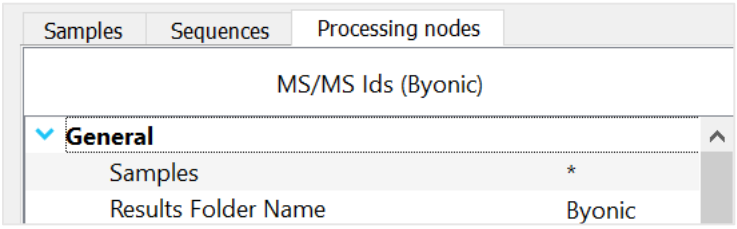


Figure 61: General.

  - **Samples** - The "*" will apply all parameters to all samples dragged and dropped into the Samples tab.

- **Results Folder Name** creates a folder of that name to save results, set to “Byonic” by default.

- **Protein database options**



Figure 62: Protein database options.

The protein database options **Add Common Contaminants** (default setting = No) and **Add Decoys** (default setting = Yes) were added to the Byonic Processing node in v. 3.9. The protein database should contain both targets and decoys (recognized by protein names beginning >Reverse or >Decoy) for false discovery rate (FDR) estimation. Byonic will automatically add decoys if the **Add decoys** box is checked and contaminant proteins (e.g., trypsin, bovine serum albumin, and human keratins) if the **Add common contaminants** box is checked. Typical folders to store input files are: C:\data_input\Mass_Spectra and C:\data_input\Protein_Databases.

- **Instrument Parameters**

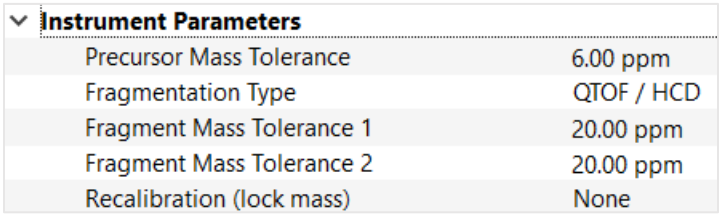


Figure 63: Instrument Parameters. Optimize for the experiment completed.

In Figure 63, the user set 6.0 ppm **Precursor Mass Tolerance**, 20.00 ppm **Fragment Mass Tolerance**, and QTOF/HCD as the **Fragmentation Type**. Both Dalton and ppm mass tolerances for precursors and fragments are supported, along with several fragmentation types. The Dalton tolerance applies to measured mass for precursors but measured m/z for fragments. The way scoring is completed changes at fragment tolerances of 0.1 Da or 100 ppm or less: high-resolution MS/MS is assumed, meaning resolution sufficient to distinguish charge states of fragment ions. For this reason, fragment tolerances larger than 0.1 Da should be used with low-resolution (ion trap) MS/MS analysis.

Internal models for most fragmentation types are included – CID low energy (ion trap), QTOF / HCD (beam-type CID), and ETD / ECD (electron transfer and electron capture dissociation), as well as a number of combinations of types. These internal models determine which fragment peak types will be scored and annotated. For example, prominent c- and z-ions and small y-ions are expected for ETD. Prominent oxonium ions are expected from glycopeptides with QTOF / HCD fragmentation, but small or missing oxonium ions from CID low energy.

- **Precursor Mass Tolerance**



Figure 64: Precursor Mass Tolerance.

The user can change this value by clicking within the text box and then clicking on the activated blue “…” square. The user can then modify the text value and mass accuracy, as required. .

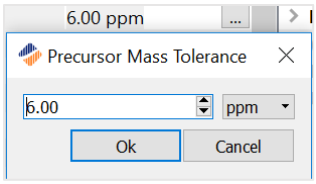


Figure 65: Modify the text value and mass accuracy.

- **Fragmentation Type**

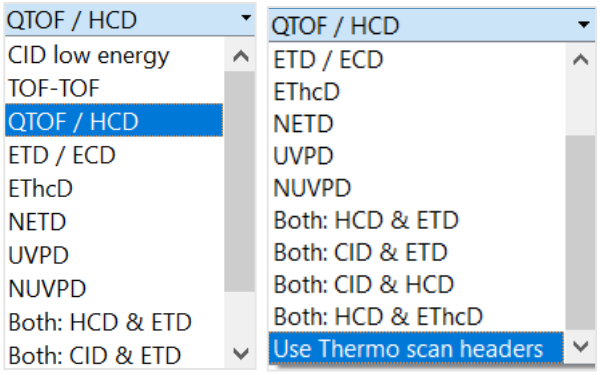


Figure 66: Fragmentation type.

The user can select from the available options using the drop-down menu. Additional options are visible using the scroll bar, including “Both:” for spectrum file(s) containing more than one fragmentation type and “Use Thermo scan headers” to read directly from Thermo raw data files.

- **Fragment Mass Tolerance 1**- The user can set the value used to acquire data in either ppm or Da. This is changed as shown in Figure 64.
- **Fragment Mass Tolerance 2** - If necessary, the user can set the value used to acquire data in either ppm or Da. This is only required for the “Both:” Fragmentation Types (selected as shown in Figure 66 above – the final entries). The user can set the value used to acquire data in either ppm or Da. This is changed as shown in Figure 64. No value will be applied if only a value for “Fragmentation Mass Tolerance 1” was entered.
- **Recalibration (lock mass)**

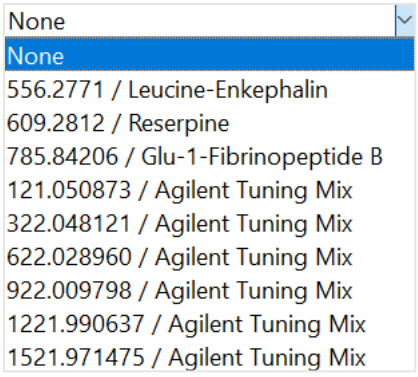


Figure 67: Lock mass calibration options.

The user can select from the available options using the drop-down menu.

- **Digestion**



Figure 68: Digestion options.

The Digestion settings allow the user to set the residues recognized by the digestion enzyme. In this example, the enzyme is trypsin, so the user entered RK for arginine and lysine for the Cleavage Site(s) and chose C-terminal for the Cleavage Side.

- **Cleavage Site(s)** - The user can change this value by entering text. If the user leaves the Cleavage Site(s) box empty, the only specific cleavage sites are protein termini.
- **Cleavage Side**

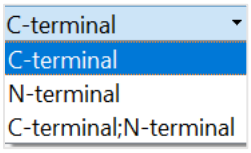


Figure 69: Cleavage Side options.

The user can change this by selecting another option from a drop-down menu. Click on the current selection to activate the drop-down menu to view the available options.

- **Digestion Specificity**

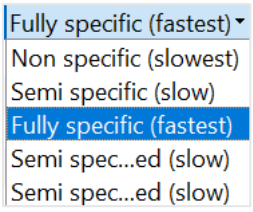


Figure 70: Digestion Specificity

In Figure 70, the user chose a “Fully specific (fastest)” search, meaning that both the N- and C-terminal cleavages must be C-terminal to R or K. Nonspecific cleavage at either or both endpoints is supported. A nonspecific search with RK in the Cleavage Site(s) box searches all peptides but favors tryptic peptides; the user must leave the Cleavage Site(s) box empty for a true no-enzyme search. Digestion Specificity can be changed by selecting another option from a drop-down menu. The user can click on the current selection to activate the drop-down menu to view the available options.

- **Missed Cleavages** - The user selected 2 Missed Cleavages, as shown in Figure 68. This limits the maximum number of internal Rs and Ks not followed by P to 2; leaving Missed Cleavages at its default value of -1, which means any number of internal Rs and Ks. Missed Cleavages can be changed by entering text.

- **Modifications**



Figure 71: Modifications options

Like most proteomics search engines, two types of modifications are supported: fixed and variable. A fixed modification is assumed to occur on all the residues of that type, but a variable modification is optional, so that each site for a variable modification is considered with and without the modification.

- **Modifications**

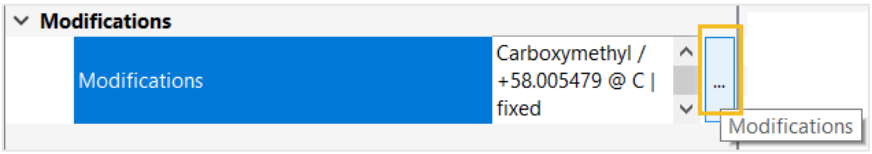


Figure 72: Click within the text box to activate the light blue "…"button

To view the list of modifications included with the default workflow, the user can click within the text box to activate the light blue "…"button. It will take a second or two for the window to open.

The user can then specify any number of modification rules via a pull-down menu containing all the modifications listed in www.unimod.org. For convenience, frequently used modifications are listed twice, at the top and again in the complete list. The three pull-down menus in each row select modification type, target residues, and fine control. There is a fourth pull-down, which lets the user delete, invert (as in (De)Carbamidomethyl), or add "attributes" to modifications. Attributes allow the user to define protein-specific modifications.

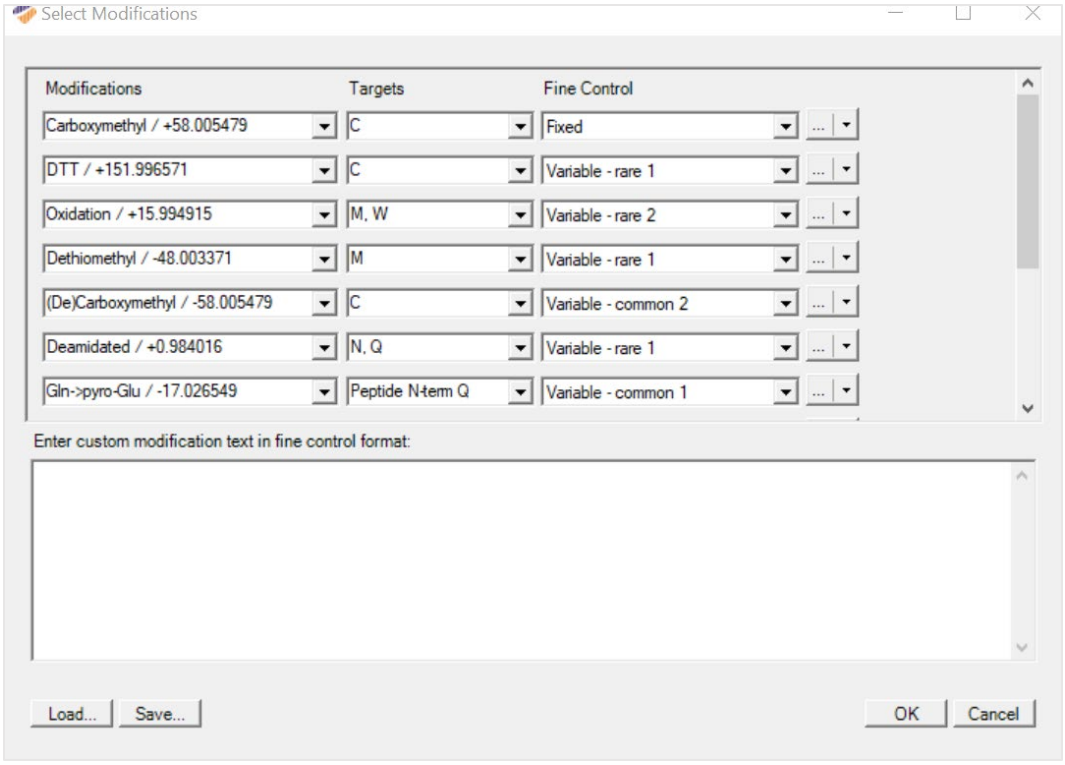


Figure 73: Select Modifications window

- **Total Common Max** and **Total Rare Max** - A unique feature not found in other search engines is offered: the user designates each variable modification as either “common” or “rare”, with the names suggesting their use. The user can define separate limits on the number of occurrences of each variable modification, so that “common 2” means at most two occurrences per peptides. Separate limits can also be set for the total number of common and rare modifications per peptide. A typical search allows a total of at most two common modifications and a total of at most one rare modification per peptide. To search for, say, three phosphoserines per peptide, the user can change Total common modification max to 3 or split phosphorylated serine between two rules: common2 and rare1. Depending upon the other modification rules, the latter approach may give a faster search. Please review the “Modification Fine Control” Application Note available at https://www.proteinmetrics.com/resources/.

  NOTE: The single most important factor in search time is **Total Common Max** (shown in Figure 71). Roughly speaking, the search time grows as CT where C is the number of common modifications enabled and T is Total Common Max.

  Conceptually, the search engine has one modification “slot” for each residue, along with slots for the peptide’s N- and C-termini. A variable modification such as +0.984016 @ N uses up the residue slot; a nonspecific terminal modification such as +57.021464 @ NTerm uses up the terminal slot; but residue-specific N-terminal modifications, such as -17.026549 @ NTerm Q, use up both the residue and the N-terminal slots.

  The big open box (shown in Figure 73) is a space for the user to type in custom modifications not listed in Unimod. The manual fine control format has the form:

  Modification_Name / Mass_Delta @ Targets | Fine_Control

  Modification_Name / is optional. The Targets field allows the 20 one-letter amino acid abbreviations, as well as four special locations: NTerm, CTerm, Protein NTerm, and

Protein CTerm. NTerm, CTerm, Protein NTerm, and Protein CTerm can also be used as modifiers of amino acid residues. Targets form a comma-separated list.

Here is an example of a real modification not (yet) in Unimod:

DehydroFormyl / +9.98435 @ NTerm S, NTerm T | rare1

A limited number of nonstandard amino acid residues can be supported by redefining one-letter amino acid abbreviations using fixed modifications. B, Z, U, O, J, and X are accepted within FASTA protein databases, with masses, respectively, of 114.042927 (same as N), 128.058578 (same as Q), 150.95363 (selenocysteine), 237.052645 (pyrrolysine), 100.0, and 110.05 (close to averagine). By placing, for example, a fixed modification of +13.04768 on J, the user can make J in a FASTA database have mass 113.04768, correct for hydroxyproline. However, the amino acid sequence is used to predict peak intensity, so this fixed modification on J will not give the same scores as a +15.9949 variable modification on P.

For comprehensive sequence variant searches, or other searches with large numbers of modifications, it is more convenient to paste in a list of modifications in the custom modification box than to add all the modifications via the drop-down menus. Sequence variant lists are available from Protein Metrics by contacting support@proteinmetrics.com.

- **Glycans**

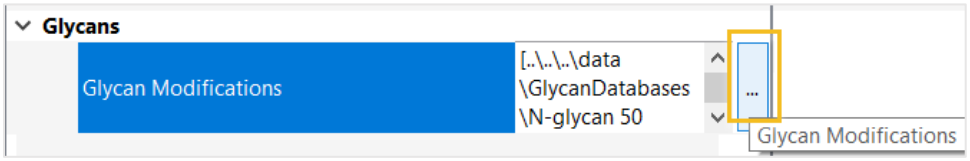


Figure 74: Click within the text box to activate the light blue "…" button

Three ways to define glycan modifications are offered: internal preset tables, external glycan databases, and user-defined glycans. The user clicks on the activated light blue "…" button to pop up a window labeled **Select Glycans**, as shown below.

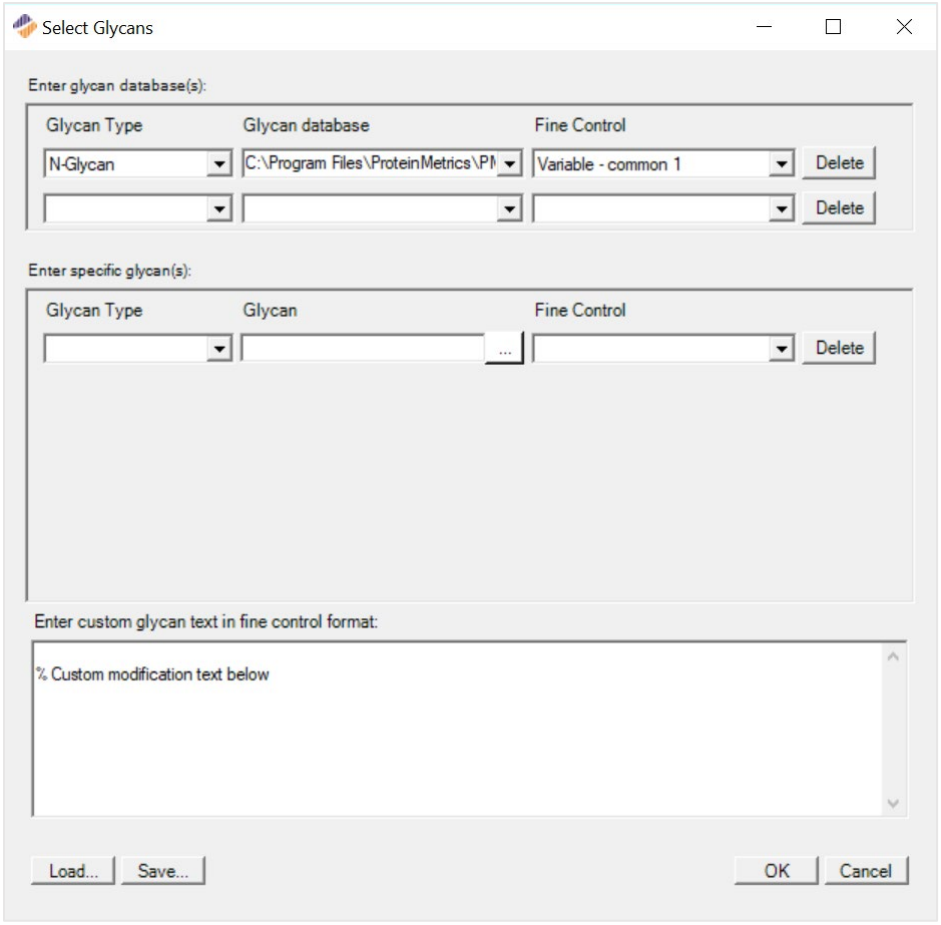


Figure 75: The Select Glycans window includes

- **Enter glycan databse(s)** allows the user to input a database of glycans. Clicking the drop-down menu and selecting **Browse** takes the user to the C:\Program Files\ProteinMetrics\PMI-Suite\Base\data\GlycanDatabases folder where the user can select from the databases currently included by default with installation. These databases are editable and the user can also add customized databases to this folder to select instead. We continually add to this set of glycan databses based on customer feedback – please reach out to support@proteinmetrics.com to request additional content.

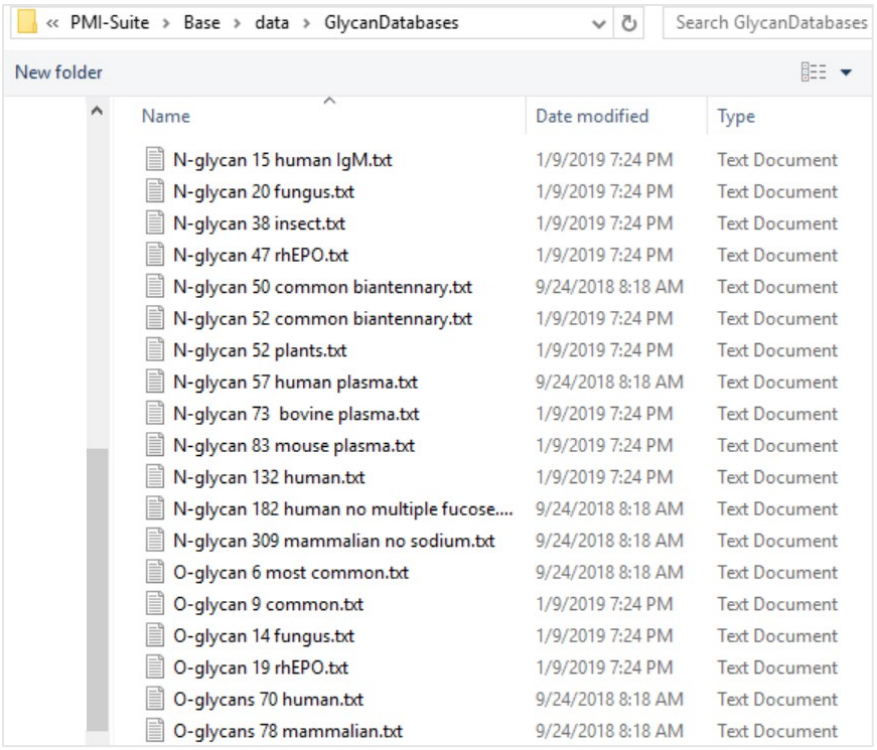


Figure 76: Glycan databases included with installation

The user can choose the Glycan type (N- or O-linked), browse for a text file of glycan compositions, and then set the Fine Control (rare1, common2, etc.).

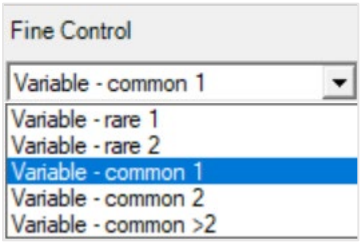


Figure 77: Modification Fine Control options

The text files include one glycan composition per line; for example, the following gives five of the most common human O-glycans. Spaces between monosaccharides are optional, and unused monosaccharides can be left out or included with zero (0) occurrences.

HexNAc(1) Hex(0)
HexNAc(1) Hex(1) Fuc(0) NeuAc(0)
HexNAc(1)Hex(1)Fuc(0)NeuAc(1)NeuGc(0)Na(0)
HexNAc(1)Hex(1)Fuc(0)NeuAc(2)NeuGc(0)Na(0)
HexNAc(1)Hex(1)Fuc(1)NeuAc(0)NeuGc(0)Na(0)

- **Enter specific glycan(s)** is located in the middle part of Select Glycans and allows the user to input glycans one at a time by specifying monosaccharide compositions.

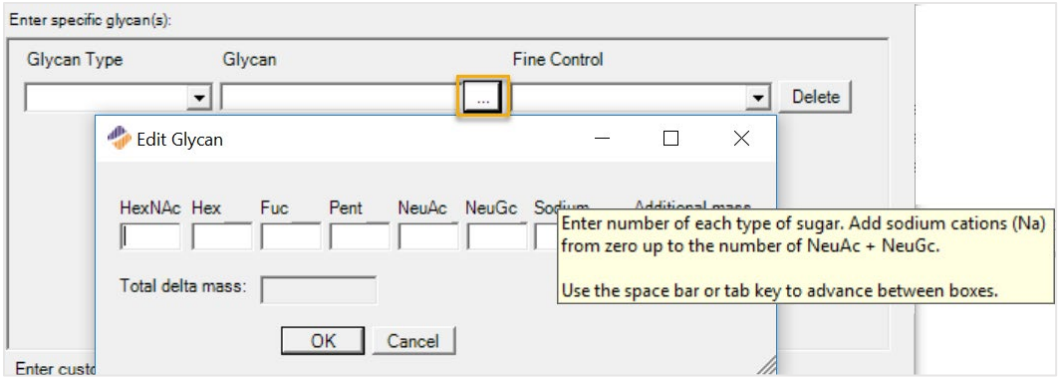


Figure 78: Enter specific glycans.

Here the user chooses Glycan type (N- or O-linked) from the drop-down menu, builds a custom glycan by clicking on the white "…" box and entering values in the **Edit Glycan** window, and then sets the Fine Control (rare1, common2, etc.). Six monosaccharide residues are allowed: HexNAc, Hexose, Fucose, Pentose (common in plants), NeuAc, and NeuGc (common in non-humans). There is also a box for Sodium because this is a common adduct on sialic acids. Other glycan masses and modifications such as sulfation and acetylation can be defined with the Additional mass box; this mass is added to the mass of the monosaccharides. The total delta mass will automatically populate. The user selects OK to close the window and include the value.

In the third section, the user can add additional glycans using the same format as within the glycan database text files.

For some helpful examples and best practices for conducting N-linked and O-linked glycan searches, see our Application Notes at https://www.proteinmetrics.com/resources/.

- **Inclusion** is used to import a *.csv file which defines m/z ratio ranges and/or elution time range segments

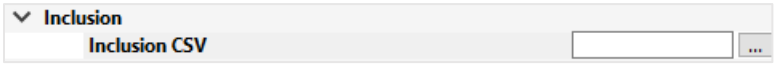


Figure 79: Inclusion option

- **MS/MS Filtering** creates MS/MS diagnostic peak filters.

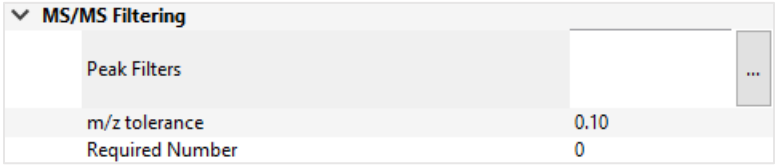


Figure 80: MS/MS diagnostic peak filtering

  - The **Peak Filters** … button in Byos opens the **Select Peaks** dialog:

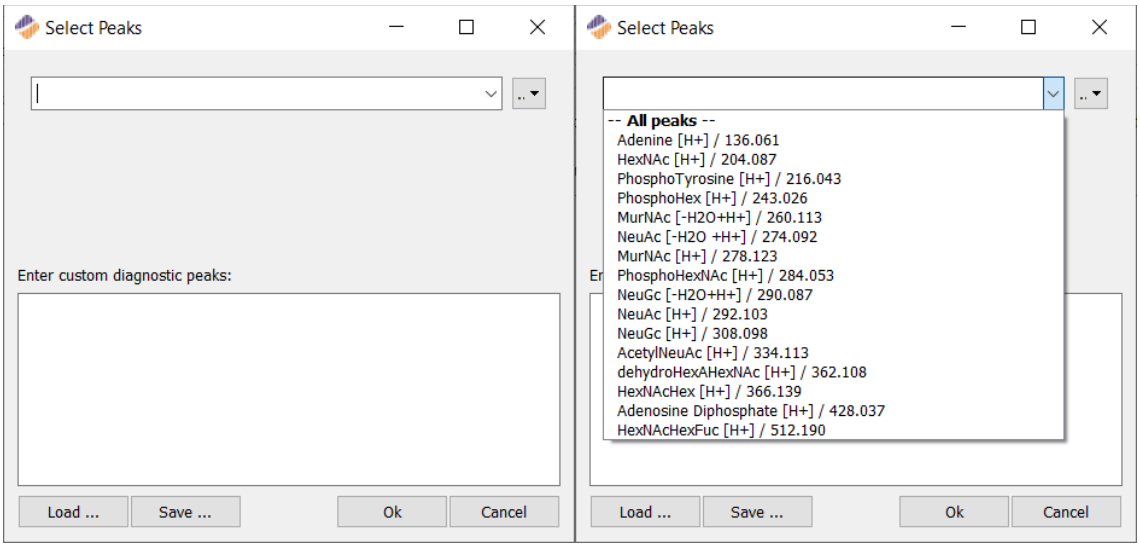


Figure 81: Select Peaks dialog for MS/MS Filtering, with dropdown

Click the drop-down arrow and select the diagnostic peak to filter. Alternatively, custom modifications can be entered manually in the box labeled **Enter custom diagnostic peaks**. The following is an example format of a diagnostic peak:

Kdn / 251.076

Set a maximum **m/z tolerance** value. **Required Number** sets a required count of diagnostic peaks to be applied from within the list. To set no minimum requirement, leave the value at zero.

- **S-S, Xlink** (NOTE: This set of parameters is unique to the S-S disulfide analysis workflow.)

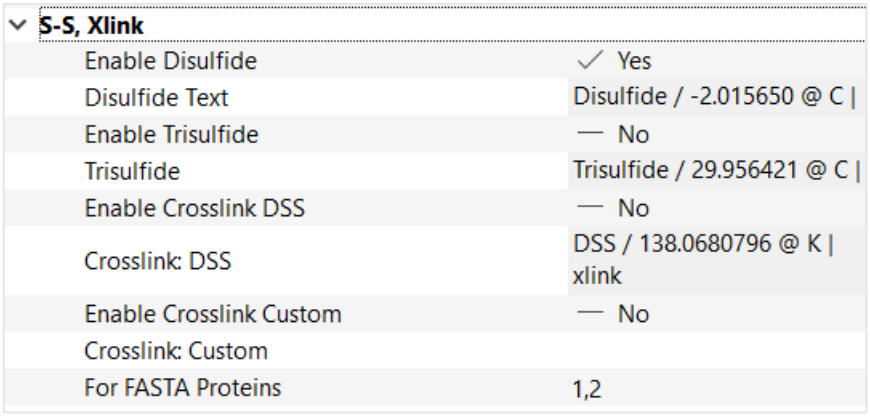


Figure 82: S-S and Xlink parameters.

This workflow allows the user to search for disulfide-bonded peptide pairs, trisulfide-bonded (also called persulfide-bonded) pairs, and more general cross-linking. This set of parameters also provides options to allow a user to search for expected and unexpected disulfide bonds. By clicking until “Yes” is visible, and designating numerically which protein sequences from the FASTA database to consider in the **For FASTA Proteins** parameter, an in-silico digestion will be completed based on the digestion parameters designated earlier, and consider every peptide that contains a cysteine and look to pair with other cysteine-containing peptides. The separation of numbers in the “For FASTA proteins” field below indicates how the potential pairings should be considered. For example, “1” searches for crosslinks in the first protein only. “4,5; 7” searches for

all potential crosslinks on the 4th, 5th, and 7th protein, #4 and #5 may crosslink to each other, but not with #7.

- The **Trisulfide** option allows a user to search for trisulfides within a single peptide and linking 2 peptides. Similarly, the Crosslink: DSS and Crosslink: Custom allow a user to search for crosslinks within a single peptide and linking 2 peptides.
- The user can specify the modification fine control used in a custom crosslink search in the “Crosslink: Custom” text box.

- **Spectrum Input Options**

Spectrum Input Options help Byos cope with imperfect inputs. For example, on many MS instruments, precursor ion charges are uncertain for some or all spectra.

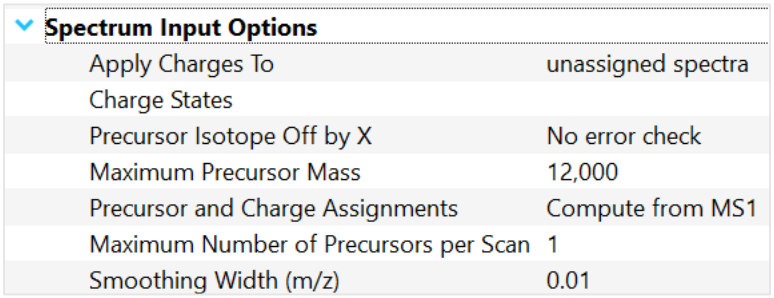


Figure 83: Spectrum Input Options.

- **Apply Charges To** - By default, the assigned charge will be used for all spectra with assigned charges and +1, +2, +3 will be used for all CID spectra and +2, +3, +4 for all ETD spectra without assigned charges. The Apply Charges To parameter allows the user to override this default setting by instead selecting unassigned spectra.
- **Charge States** - All comma-separated charges detailed will be applied to each spectrum (based on the values entered into the Charge States box).
- **Precursor Isotope Off By X**

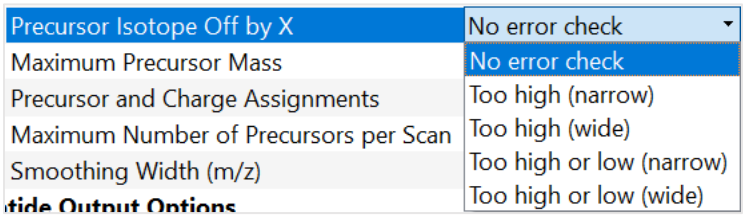


Figure 84: Precursor Isotope Off By X.

Similarly, on many instruments the nominal precursor mass may actually be the mass of a 13C isotope peak rather than of the base (all 12C monoisotopic) peak, so the true precursor mass will within 10 ppm of 2350.120 Da or within 10 ppm of 2351.123 Da. Precursor Isotope Off By X is a pulldown menu with several options.

No error check will use only the assigned precursor; **Too high (narrow**) will allow the assigned precursor to be up to 2 Da too high; **Too high (wide)** will allow the assigned precursor to be up to n Da too high for a precursor of mass at least 1000n Da; **Too high or low (narrow)** will allow the assigned precursor to be up to 2 Da too high or 2 Da too low; **Too high or low (wide)** will allow the assigned precursor to be up to n Da too high or 2 Da too low for a precursor of mass at least 1000n Da.

- **Maximum Precursor Mass** sets the Maximum Precursor Mass to be considered.

- **Precursor and Charge Assignments**

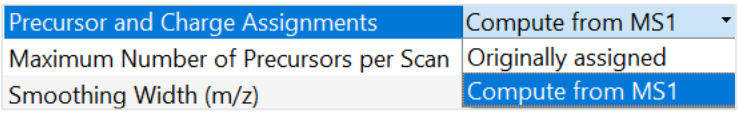


Figure 85: Precursor and Charge Assignments.

The precursor and charge assignments will be calculated directly from the MS1 data or the originally assigned values will be applied. The user can set this using a drop-down menu.

- **Maximum Number of Precursors per Scan** - Multiple precursors per scan can also be considered – it is recommended for the user to set this to 2 for complex samples and 5-10 if processing MSe or DIA data.
- **Smoothing Width** - The user can enter a sigma value in Thomsons for Gaussian smoothing and centroiding of Waters or Sciex data. Half-width at peak half maximum (~0.01 m/z) works well and is the default value already entered for the user.

- **Peptide Output Options**

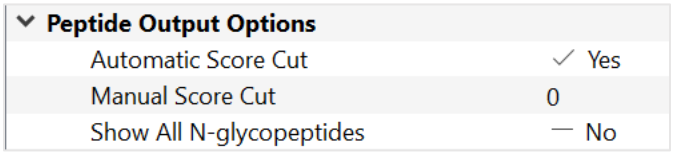


Figure 86: Peptide Output Options.

The Peptide Output Options parameters offer options for filtering the peptide-spectrum matches (PSMs) by score. By default, PSM filtering is deferred until after protein ranking, and then filters to control PSM FDR on the “true” proteins—those ranked above the top-ranking decoy protein. This method gains sensitivity while simultaneously reducing both protein and PSM FDRs. See:

Two-dimensional target decoy strategy for shotgun proteomics, Journal of Proteome Research 10 (12), 5296-5301, 2011.

- **Automatic Score Cut** - To filter PSMs before protein ranking, the user can click **Yes** to activate Automatic score cut and type in a minimum score.
- **Manual Score Cut** - The user can activate Automatic Score Cut and enter a minimum score. For example, a score threshold of 200 will remove weak matches and a threshold of 400 will remove all but the best matches. Filtering by score may be helpful in special cases, for example to eliminate from consideration all but the best wildcard PSMs.
- **Show All N-Glycopeptides** - The user also has the option to Show All N-Glycopeptides. This will show N-Glycopeptide matches regardless of score or FDR. This is recommended for simple samples only. This can be especially useful for low energy CID data.

- **Protein Output Options**



Figure 87: Protein Output Options.

- **Protein FDR**

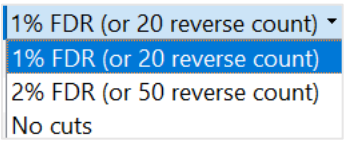


Figure 88: Protein FDR options

This gives the user control of the protein list cut-off. By default, the protein list is cut at 1% protein FDR or 20 decoy proteins, whichever comes last, but the user can ask for 2% protein FDR or a completely unfiltered (but still ranked) protein list, or No Cuts.

- **Create a Focused Database** - If the user clicks this to select Yes, the software is directed to output a new FASTA file (labeled focused and appearing in the output objs directory) containing only the proteins found in the search, along with suitable decoys (>Reverse) for unbiased FDR estimation. The focused database can then be used for subsequent wide searches, including more modifications and/or a wildcard. Of course, the user can also create focused databases outside of the software by editing existing FASTA files.
- **Export mzIdentML** - The user can also select to export this file as a mzIdentML file.

- **Multicore Options**

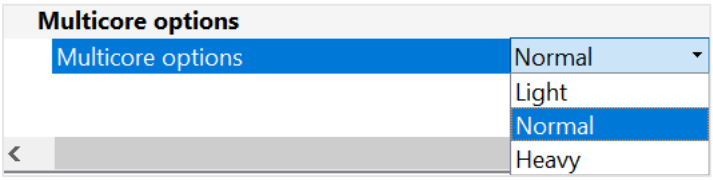


Figure 89: Multicore processing options.

The user can control the number of computer cores of the CPU used through the drop-menu. The Light searches uses one core, Normal search uses all available cores minus two, and Heavy search uses all available cores.

- **Wildcard**



Figure 90: Wildcard parameters.

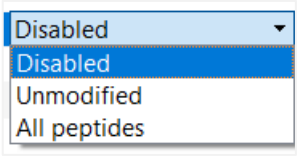


Figure 91: Wildcard Search drop-down menu options.

Wildcard lets the user turn on wildcard searches, set the range for the wildcard mass, and restrict the wildcard to certain residues if desired. The options Disabled, Unmodified, and All peptides are available through a drop-down menu.

The **Restrict to residues** box uses the common 20 single-letter residue abbreviations, and (lower case) n denotes peptide N-terminus and (lower case) c denotes peptide C-terminus. A wildcard, even one with a mass range of only 50 or 60 Da, greatly increases the size of the search. It is best used with a focused database (see the Advanced tab section below) and used either alone or with only a few other modifications enabled. Most wildcard mass shifts will be recognizable by an expert; hence, a wildcard can be used to discover which known modifications should be enabled in a subsequent search. For more details about the wildcard search, see the application note "Byonic™: Wildcard Search™" at https://www.proteinmetrics.com/resources/#application-notes.

By specifying most modifications as rare, it is quite feasible to search for 10 – 20 modification types at once with Byonic. Even larger searches are possible with focused protein databases, for example with therapeutic proteins. Such a focused database easily allows efficient mutation searches with 200+ possible substitutions, or oxidative footprinting searches with 50+ types of oxidations. Glycans and wildcards can easily enlarge the search space by 2 to 3 orders of magnitude, so these options should be used with care, and in conjunction with only the most common variable modifications (such as oxidized methionine or pyro-Glu N-terminus). **NOTE**: The single most important factor in search time is Total common max. Roughly speaking, the search time grows as C^T where C is the number of common modifications enabled and T is Total common max.

The **Appendix** of the Byonic Manual provides examples of frequently found modifications and appropriate syntax for including those modifications in a Byonic search.

Quant

NOTE: As mentioned in the above MS/MS Ids section, these parameters apply alone to the PTM (in-silico) Workflow.

- **General**



Figure 92: Samples to include in the project.

  - **Samples** - The "*" will apply all parameters to all samples dragged and dropped into the Samples tab.

- **In-silico options**

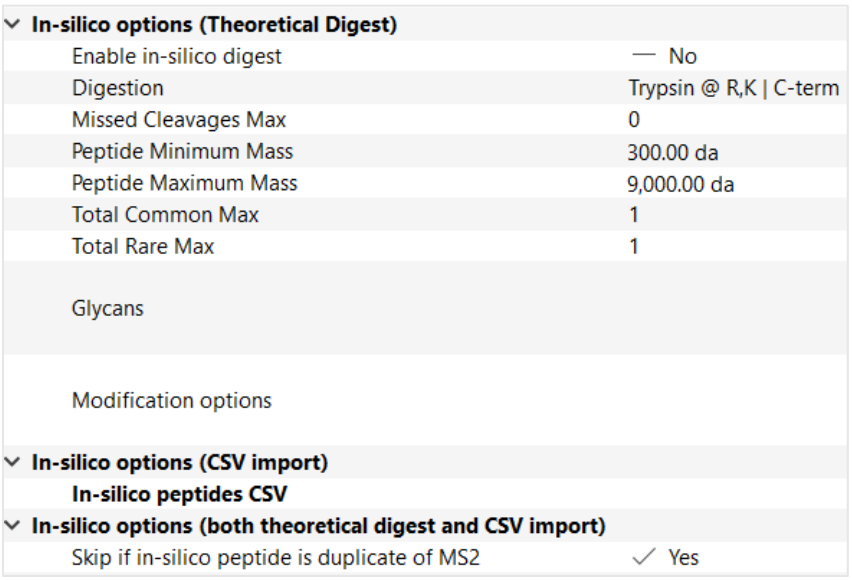


Figure 93: In-silico options processing parameters.

Parameters are separated into three sections to clarify which pertain to a **Theoretical Digest**, **CSV import** and **both theoretical digest and CSV import.**

- **Enable In-Silico** - The user has the option to click **Yes** to enable this parameter to generate an in-silico list of peptides. The below additional parameters must then be set.

  NOTE: The option to add a list of in-silico peptides from a CSV file is offered through In-Silico Peptides CSV.

- **Digestion**

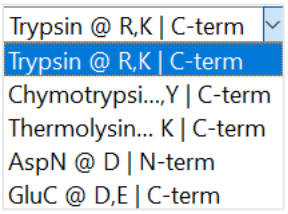


Figure 94: Digestion options.

The user can view several options through the drop-down menu. These are detailed as one letter residue codes adjacent to cleavage points. The user may enter a customized digestion by using the following syntax: Name @ Amino acid letters | C-term or N-term.

- **Missed Cleavages Max**

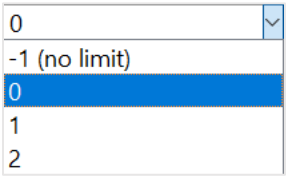
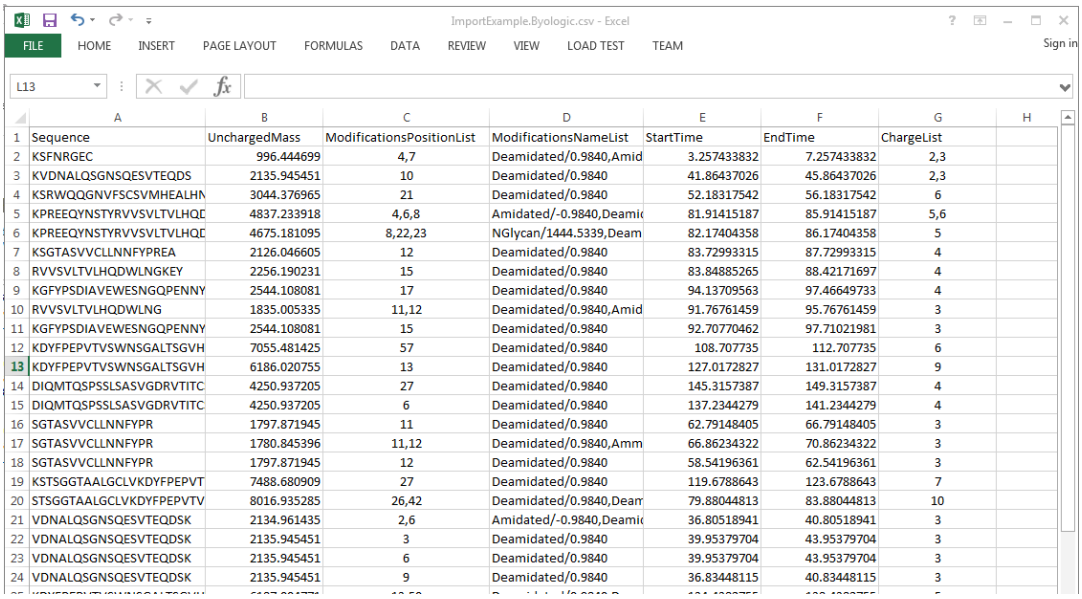


Figure 95: Missed Cleavages Max options.

The user can set the maximum number of missed cleavages per peptide by selecting a value using the drop-down menu. A value of -1 allows any number of cleavages.

- **Peptide Minumum Mass** - The user can set the minimum value of the peptide mass range.
- **Peptide Maximum Mass** - The user can set the maximum value of the peptide mass range.
- **Total Common Max** - The user can set the maximum number of common modifications per peptide. The search size grows by an order of magnitude with each increase in Total Common Max.
- **Total Rare Max** - The user can set the maximum number of rare modifications per peptide. The search size grows by an order of magnitude with each increase in Total Rare Max.
- **Glycans** - The user can specify glycan databsesor custom glycans to include in the in-silico generation. This is done the same as detailed in the MS/MS Ids section.
- **Modifications Options** - The user can specify modifications to include in the in-silico generation. This is done the same as detailed in the MS/MS Ids section
- **In-Silico Peptides CSV** - This option allows the user to to add a list of in-silico pepitdes from an imported CSV file, thus there is no need to run an MS2 search if there are known modifications with known masses and retention times. The format for the CSV is shown below.



	A	B	C	D	E	F	G	H
1	Sequence	UnchargedMass	ModificationsPositionList	ModificationsNameList	StartTime	EndTime	ChargeList	
2	KSFNRGEC	996.444699	4,7	Deamidated/0.9840,Amid	3.257433832	7.257433832	2,3	
3	KVDNALQSGNSQESVTEQDS	2135.945451	10	Deamidated/0.9840	41.86437026	45.86437026	2,3	
4	KSRWQQGNVFSCSVMHEALHN	3044.376965	21	Deamidated/0.9840	52.18317542	56.18317542	6	
5	KPREEQYNSTYRVVSVLTVLHQD	4837.233918	4,6,8	Amidated/-0.9840,Deamic	81.91415187	85.91415187	5,6	
6	KPREEQYNSTYRVVSVLTVLHQD	4675.181095	8,22,23	NGlycan/1444.5339,Deam	82.17404358	86.17404358	5	
7	KSGTASVVCLLNNFYPREA	2126.046605	12	Deamidated/0.9840	83.72993315	87.72993315	4	
8	RVVSVLTVLHQDWLNGKEY	2256.190231	15	Deamidated/0.9840	83.84885265	88.42171697	4	
9	KGFYPSDIAVEWESNGQPENNY	2544.108081	17	Deamidated/0.9840	94.13709563	97.46649733	4	
10	RVVSVLTVLHQDWLNG	1835.005335	11,12	Deamidated/0.9840,Amid	91.76761459	95.76761459	3	
11	KGFYPSDIAVEWESNGQPENNY	2544.108081	15	Deamidated/0.9840	92.70770462	97.71021981	3	
12	KDYFPEPVTVSWNSGALTSGVH	7055.481425	57	Deamidated/0.9840	108.707735	112.707735	6	
13	KDYFPEPVTVSWNSGALTSGVH	6186.020755	13	Deamidated/0.9840	127.0172827	131.0172827	9	
14	DIQMTQSPSSLSASVGDRVTITC	4250.937205	27	Deamidated/0.9840	145.3157387	149.3157387	4	
15	DIQMTQSPSSLSASVGDRVTITC	4250.937205	6	Deamidated/0.9840	137.2344279	141.2344279	4	
16	SGTASVVCLLNNFYPR	1797.871945	11	Deamidated/0.9840	62.79148405	66.79148405	3	
17	SGTASVVCLLNNFYPR	1780.845396	11,12	Deamidated/0.9840,Amm	66.86234322	70.86234322	3	
18	SGTASVVCLLNNFYPR	1797.871945	12	Deamidated/0.9840	58.54196361	62.54196361	3	
19	KSTSGGTAALGCLVKDYFPEPVT	7488.680909	27	Deamidated/0.9840	119.6788643	123.6788643	7	
20	STSGGTAALGCLVKDYFPEPVTV	8016.935285	26,42	Deamidated/0.9840,Deam	79.88044813	83.88044813	10	
21	VDNALQSGNSQESVTEQDSK	2134.961435	2,6	Amidated/-0.9840,Deamic	36.80518941	40.80518941	3	
22	VDNALQSGNSQESVTEQDSK	2135.945451	3	Deamidated/0.9840	39.95379704	43.95379704	3	
23	VDNALQSGNSQESVTEQDSK	2135.945451	6	Deamidated/0.9840	39.95379704	43.95379704	3	
24	VDNALQSGNSQESVTEQDSK	2135.945451	9	Deamidated/0.9840	36.83448115	40.83448115	3	

Figure 96: Format for an In-Silico Peptides CSV.

- **Skip if in-silico peptide is duplocate of MS2** - The user can click to select **Yes**. This will skip any generated in-silico peptide if one already exists from the MS2 data. It is recommended to make this selction when generating in-silico peptides.

- **MS extract options**

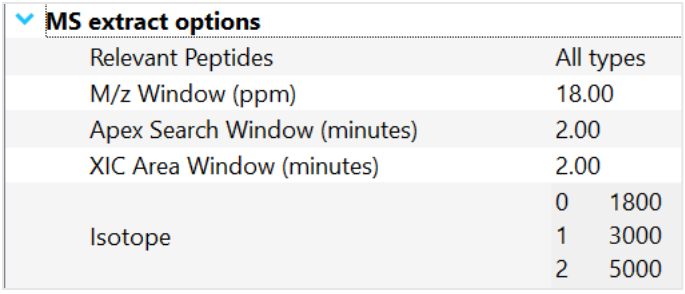


Figure 97: MS extract options.

There are various aspects to the XIC of a peptide which can help in distinguishing a true from false identification, or whether the peptide hit is relevant based on intensity.

- **Relevant Peptides**

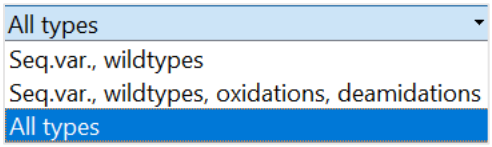


Figure 98: Relevant Peptide options.

The user can select what types of peptides to consider relevant using the drop-down menu. Only peptides of interest will be considered for extracting additional information, such as XIC, isotope profile, and MS2 fragments. This pulldown menu option may affect the file size significantly. For SVA, the user may only need to choose Sequence variants and wildtypes. For peptide mapping, the user should choose All types.

- **M/z Window (ppm)** - The user can define the maximum m/z error tolerance to be applied to the search.
- **Apex Search Window (minutes)** - The user can automatically define the search window for the peak apex.
- **XIC Area Window (minutes)** - The user can automatically set the time window for integration. The defined time limits will be visible as two vertical lines for each XIC in the project. These lines can be dragged by the user's mouse to adjust the integration time if needed.
- **Isotope**

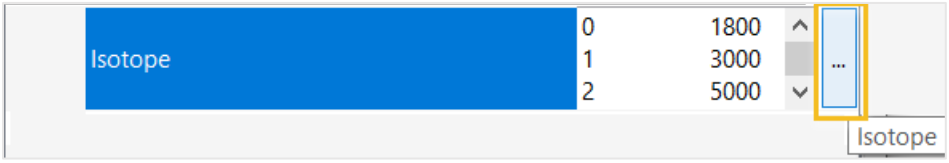


Figure 99: Isotope settings.

The user can set the mass ranges for the offset from the monoisotopic mass. Click within the box to activate the light blue "…" button and view the default values. These values can be modified by typing in values.

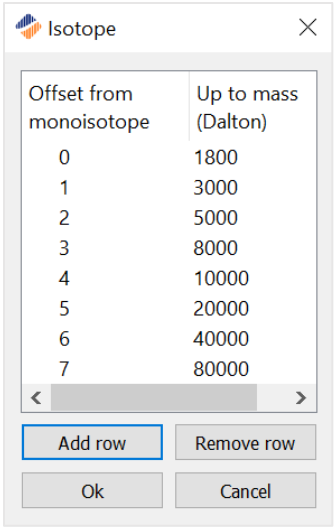


Figure 100: Isotope settings

- **Advanced**

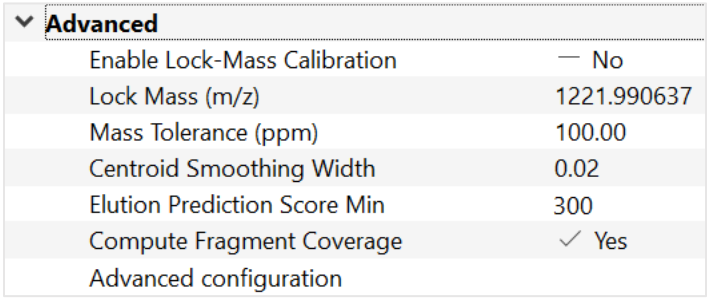


Figure 101: Advanced options

  - **Enable Lock-Mass Calibration** - The user can click to select between **No** and **Yes**.
  - **Lock Mass (m/z)**

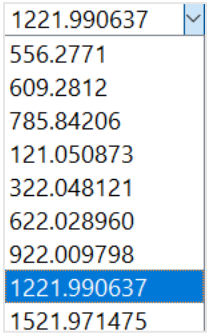


Figure 102: Lock mass m/z value options

The user can enter the calibrant m/z value. Several are also available using the drop-down menu or the user can type in a numerical value. If empty, no calibration will be applied.

- **Mass Tolerance (ppm)** - The user can enter the calibrant m/z mass tolerance value in ppm. The user can type in a numerical value or use the up and down arrow keys to increase or decrease the value. If empty, no calibration will be applied.
- **Centroid Smoothing Width** - The user can enter a sigma value in Thomsons for Gaussian smoothing and centroiding of Waters or Sciex data. Half-width at peak half maximum (~0.01 m/z) works well and is the default value already entered for the user.
- **Elution Prediction Score Min** - The user can set a value for this filter to find wildtype peptides with scores above this value to adjust the elution time.
- **Compute Fragment Coverage** - The user can click between **Yes** and **No** to compute fragmentation coverage at project creation.
- **Advanced Configuration** - The user can enter text commands to complete advanced procesing. These are detailed in the Release Notes included with each quarterly release. Please reach out to support@proteinmetrics.com for additional details.

- **Report**

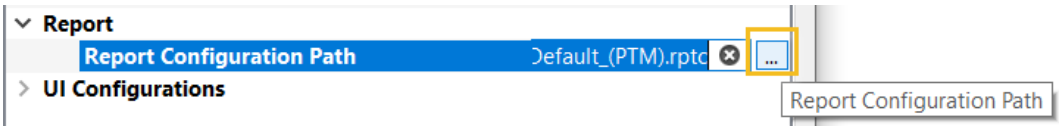


Figure 103: Report Configuration Path

Each Byos default workflow includes a report template created by our Customer Success team that is optimized for the specific type of analysis. If the user prefers a customized report template, they can direct Byos to this file using the light blue "…" button. They will be prompted to select a file.

- **UI Configurations**

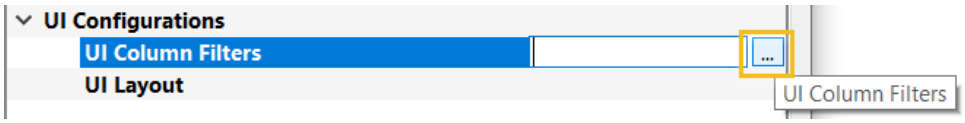


Figure 104: UI Configurations – UI Column Filters and UI Layout

The user has the ability to import column filters as well as layout files. This is designed to standardize analyses across all users, labs, and sites. The user can direct Byos to the preferred file for each using the light blue "…" button. The user will be prompted to select a file.

- **Feature Finder**

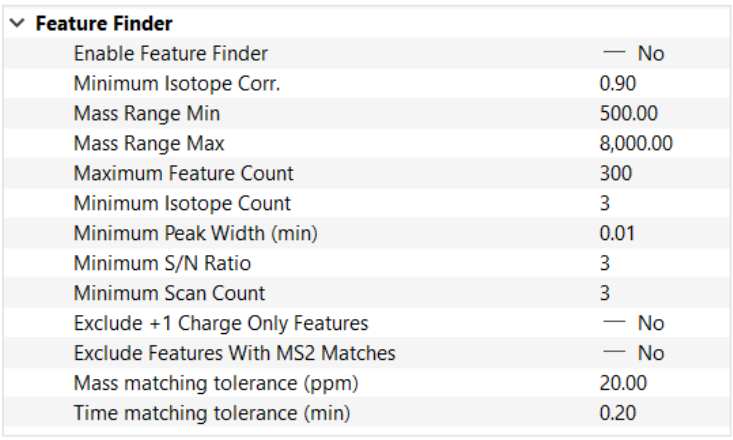


Figure 105: Feature Finder

Feature Finder is an algorithm that allows the user to scan the MS1 domain for all existent peptides in a sample. In a typical data dependent acquisition assay, peptides are identified only if an MS2 scan is triggered from a precursor ion in an MS1 scan. Therefore, if a peptide ion does not trigger an MS2 scan, or its score excludes it from identification, it will not be detected. Feature Finder ensures that a peak will not go undetected by identifying all possible isotopic distributions in a sample that could potentially be a peptide.

Feature Finder also allows the user to match these features to the current identifications and to add a specified number of the unidentified features as unknowns to the project. Doing so will generate XICs for the unknowns and allow reporting on features such as retention time, mass, intensity, and fold change across samples.

- **Enable Feature Finder** - The user can click to select between **No** and **Yes** to enable Feature Finder.
- **Minimum Isotope Corr.** - The user can set the minimum cutoff to match the found isotopic distribution to the theoretical distribution using Pearson's correlation. If a feature has an isotopic distribution that is vastly different from its theoretical, this can be used to exclude it.
- **Mass Range Min** - The user can set the minimum mass for a feature to be considered.
- **Mass Range Max** - The user can set the maximum mass for a feature to be considered.
- **Maximum Features Count** - The user can set the maximum feature count to include the most intense features.
- **Minimum Isotope Count** - The user can set the minimum number of isotope peaks required for a feature to be considered. This applies only if its charge is greater than 2.
- **Minimum Peak Width (min)** - The user can set the minimum peak width in minutes for a feature to be considered.
- **Minimum S/N Ratio** - The user can set the minimum S/N ratio for a feature to be considered.
- **Minimum Scan Count** - The user can set the minimum scans required across the peak for a feature to be considered.
- **Exclude +1 Charge Only Features** - The user can click to select between **Yes** and **No** to exclude singly charged features.
- **Exclude Features With MS2 Matches** - The user can click to select between **No** and **Yes** to exclude features with MS2 matches. Exclude will remove the feature if it coincides

within the time and mass tolerances set – it will not be included in the top X features to avoid duplicates. It is recommended to check this box for projects containing MS2 data and Byonic searches.

- **Mass matching tolerance (ppm)** and **Time matching tolerance (min)** - These parameters are used to match and remove features that have already been identified by MS2 (when **Exclude Features With MS2 Matches** is checked/set to Yes) so that they do not contribute to the maximum feature count. They help to match the retention time and mass of an unknown feature to a MS2 feature so that it may be excluded from the Peptides table in Peptide Analysis.

MAM Chromatogram Annotation (LC-MS), MAM Chromatogram Annotation (in-silico), MAM New Peak Detection, Comparison Chromatography, Comparison Chromatography (in-silico), Reference Chromatography, Reference Chromatography (in-silico) Workflows

The following section details parameters for the combination of MS/MS IDs and Chromatographic Quant. They are therefore relevant to the Reference Chromatography, Characterization Chromatography, and MAM Chromatogram Annotation (LC-MS) Workflows.

Processing nodes

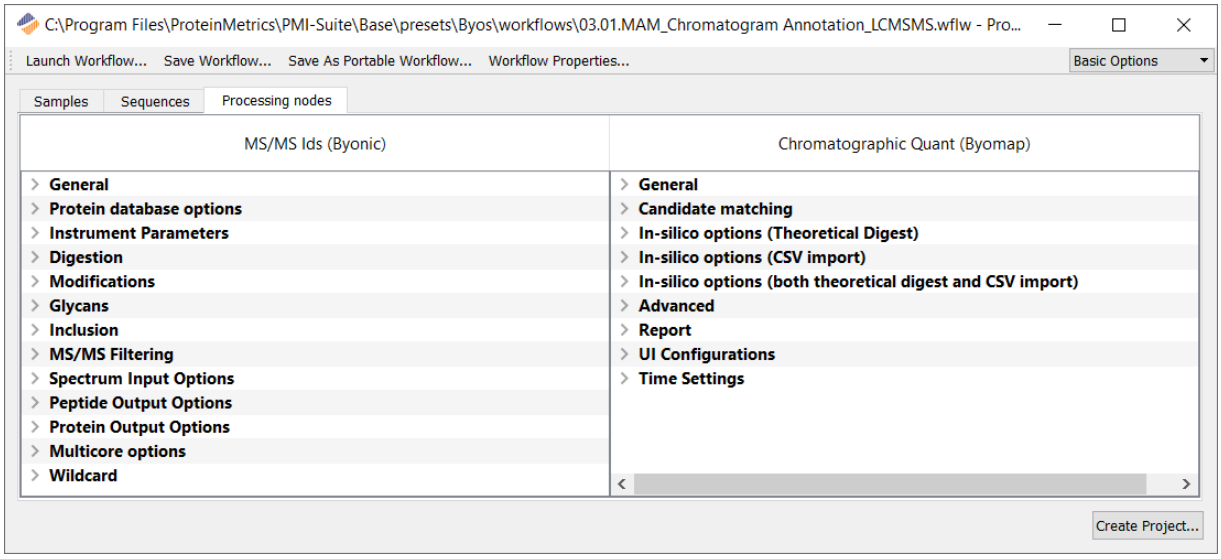


Figure 106: Processing nodes parameters – MAM Chromatogram Annotation.

MS/MS Ids

NOTE: These parameters are not part of the Reference Chromatography (in-silico), Characterization Chromatography (in-silico), MAM Chromatogram Annotation (in-silico), and MAM New Peak Detection Workflows. Please refer directly to the below Chromatographic Quant section.

- **General**

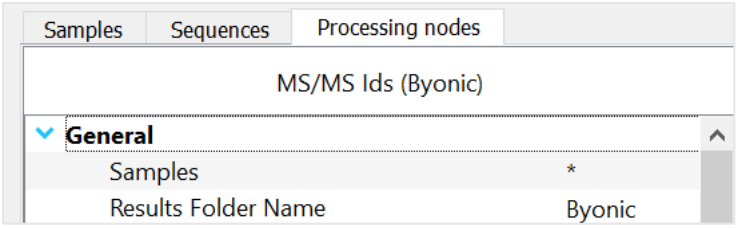


Figure 107: General

- **Samples** - The "*" will apply all parameters to all samples dragged and dropped into the Samples tab.
- **Results Folder Name** - This is automatically set to "Byonic" and will be applied to the additional processing (as defined in the workflow).

- **Protein database options**



Figure 108: Protein database options

The protein database options **Add Common Contaminants** (default setting = No) and **Add Decoys** (default setting = Yes) were added to the Byonic Processing node in v. 3.9. The protein database should contain both targets and decoys (recognized by protein names beginning >Reverse or >Decoy) for false discovery rate (FDR) estimation. Byonic will automatically add decoys if the **Add decoys** box is checked and contaminant proteins (e.g., trypsin, bovine serum albumin, and human keratins) if the **Add common contaminants** box is checked. Typical folders to store input files are: C:\data_input\Mass_Spectra and C:\data_input\Protein_Databases.

- **Instrument Parameters**

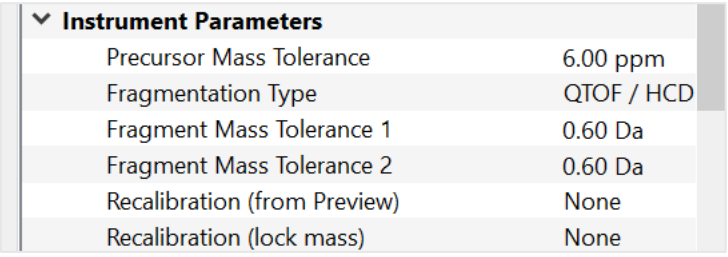


Figure 109: Instrument Parameters. Optimize for the experiment completed.

In Figure 109, the user set 6.0 ppm **Precursor Mass Tolerance**, 0.60 Da **Fragment Mass Tolerance**, and QTOF/HCD as the **Fragmentation Type**. Both Dalton and ppm mass tolerances for precursors and fragments are supported, along with several fragmentation types. The Dalton tolerance applies to measured mass for precursors but measured m/z for fragments. The way scoring is completed changes at fragment tolerances of 0.1 Da or 100 ppm or less: high-resolution MS/MS is assumed, meaning resolution sufficient to distinguish charge states of fragment ions. For this reason, fragment tolerances larger than 0.1 Da should be used with low-resolution (ion trap) MS/MS analysis.

Internal models for most fragmentation types are included – CID low energy (ion trap), QTOF / HCD (beam-type CID), and ETD / ECD (electron transfer and electron capture dissociation), as

well as a number of combinations of types. These internal models determine which fragment peak types will be scored and annotated. For example, prominent c- and z-ions and small y-ions are expected for ETD. Prominent oxonium ions are expected from glycopeptides with QTOF / HCD fragmentation, but small or missing oxonium ions from CID low energy.

- **Precursor Mass Tolerance**



Figure 110: Precursor Mass Tolerance

The user can change this value by clicking within the text box and then clicking on the activated blue "…" square. The user can then modify the text value and mass accuracy, as required. This is shown below in Figure 111.

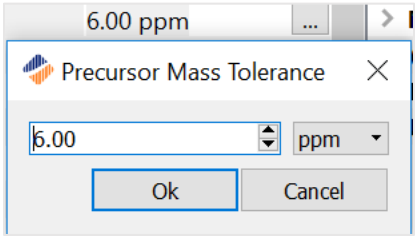


Figure 111: Modify the text value and mass accuracy

- **Fragmentation Type**

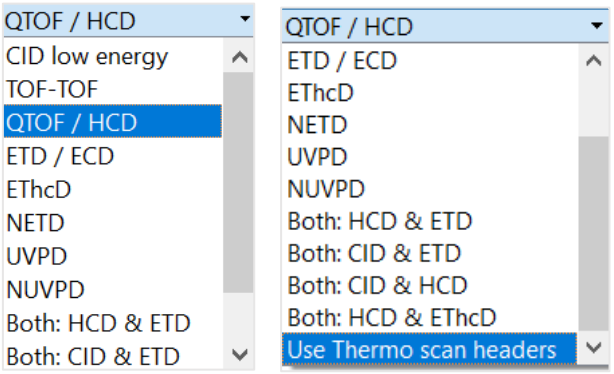


Figure 112: Fragmentation type.

The user can select from the available options using the drop-down menu. Additional options are visible using the scroll bar, including "Both:" for spectrum file(s) containing more than one fragmentation type.

- **Fragment Mass Tolerance 1** - The user can set the value used to acquire data in either ppm or Da.
- **Fragment Mass Tolerance 2** - If necessary, the user can set the value used to acquire data in either ppm or Da. This is only required for the "Both:" Fragmentation Types. No value will be applied if only a value for "Fragmentation Mass Tolerance 1" was entered.
- **Recalibration (lock mass)**

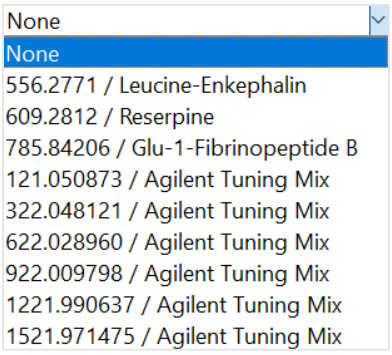


Figure 113: Lock mass calibration options

The user can select from the available options using the drop-down menu.

- **Digestion**



Figure 114: Digestion options

The Digestion settings allow the user to set the residues recognized by the digestion enzyme. In this example, the enzyme is trypsin, so the user entered RK for arginine and lysine for the Cleavage Site(s) and chose C-terminal for the Cleavage Side.

- **Cleavage Site(s)** - The user can change this value by entering text. If the user leaves the Cleavage Site(s) box empty, the only specific cleavage sites are protein termini.

- **Cleavage Side**

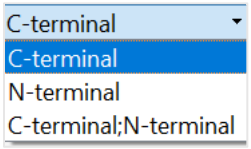


Figure 115: Cleavage Side options

The user can change this by selecting another option from a drop-down menu. Click on the current selection to activate the available options.

- **Digestion Specificity**

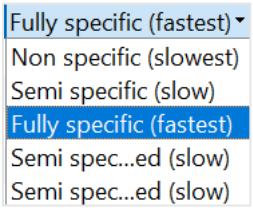


Figure 116: Digestion Specificity

Here the user chose “Fully specific (fastest)” search, meaning that both the N- and C-terminal cleavages must be C-terminal to R or K. Nonspecific cleavage at either or both endpoints is supported. A nonspecific search with RK in the Cleavage Site(s) box searches all peptides but favors tryptic peptides; the user must leave the Cleavage Site(s) box empty for a true no-enzyme search. Digestion Specificity can be changed by selecting another option from a drop-down menu. The user can click on the current selection to activate the drop-down menu to view the available options.

- **Missed Cleavages** - The user selected 2 Missed Cleavages, as shown in Figure 114. This limits the maximum number of internal Rs and Ks not followed by P to 2; leaving Missed Cleavages at its default value of -1, which means any number of internal Rs and Ks. Missed Cleavages can be changed by entering text.

- **Modifications**



Figure 117: Modifications options

Like most proteomics search engines, two types of modifications are supported: fixed and variable. A fixed modification is assumed to occur on all the residues of that type, but a variable modification is optional, so that each site for a variable modification is considered with and without the modification.

- **Modifications**

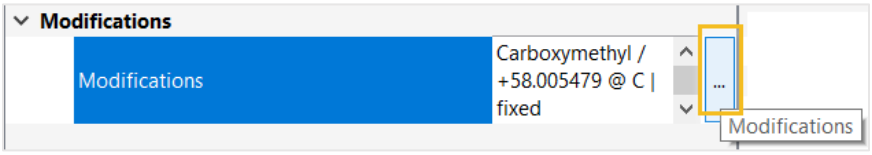


Figure 118: Click within the text box to activate the light blue “…” button

To view the list of modifications included with the default workflow, the user can click within the text box to activate the light blue “…” button. It will take a second or two for the window to open.

The user can then specify any number of modification rules via a pull-down menu containing all the modifications listed in www.unimod.org. For convenience, frequently used modifications are listed twice, at the top and again in the complete list. The three pull-down menus in each row select modification type, target residues, and fine control. There is a fourth pull-down, which lets the user delete, invert (as in (De)Carbamidomethyl), or add “attributes” to modifications. Attributes allow the user to define protein-specific modifications.

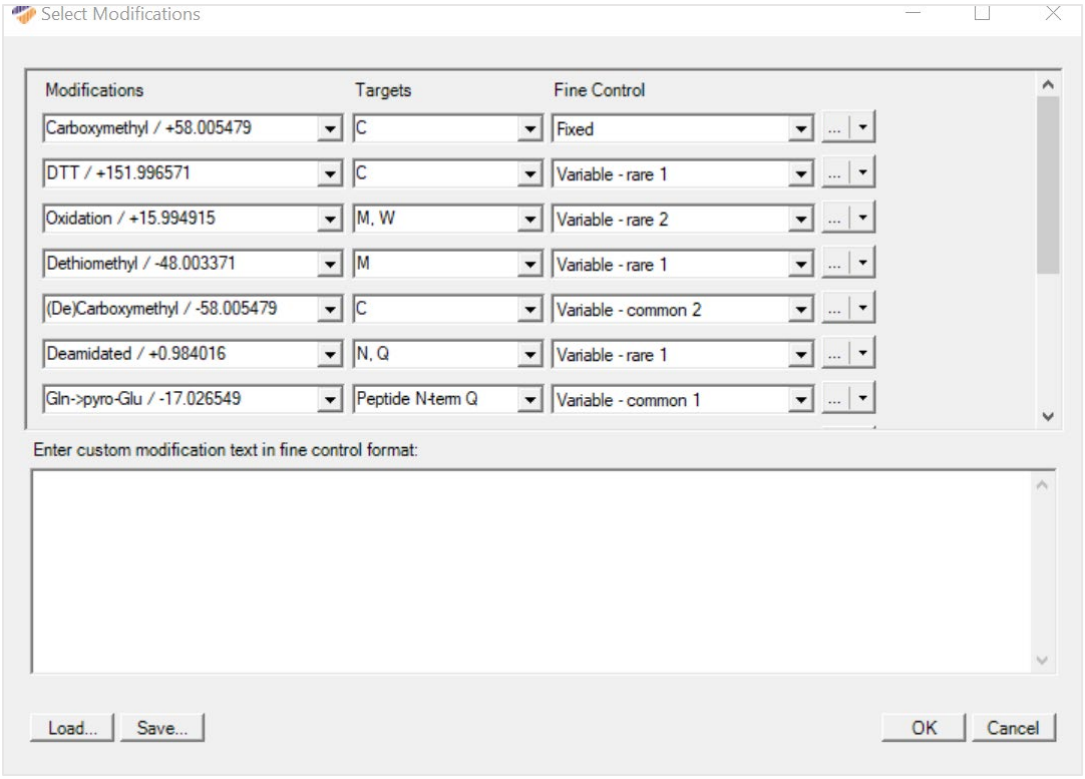


Figure 119: Select Modifications window

- **Total Common Max and Total Rare Max** - A unique feature not found in other search engines is offered: the user designates each variable modification as either “common” or “rare”, with the names suggesting their use. Separate limits on the number of occurrences of each variable modification can be set, so that “common 2” means at most two occurrences per peptides. Separate limits can also be set on the total number of common and rare modifications per peptide. A typical search allows a total of at most two common modifications and a total of at most one rare modification per peptide. To search for, say, three phosphoserines per peptide, the user can change Total common modification max to 3 or split phosphorylated serine between two rules: common2 and rare1. Depending upon the other modification rules, the latter approach may give a faster search. Please review the “Modification Fine Control” Application Note available at https://www.proteinmetrics.com/resources/.

  NOTE: The single most important factor in search time is **Total Common Max**. Roughly speaking, the search time grows as CT where C is the number of common modifications enabled and T is Total Common Max.

  Conceptually, there is one modification “slot” for each residue, along with slots for the peptide’s N- and C-termini. A variable modification such as +0.984016 @ N uses up the residue slot; a nonspecific terminal modification such as +57.021464 @ NTerm uses up the terminal slot; but residue-specific N-terminal modifications, such as -17.026549 @ NTerm Q, use up both the residue and the N-terminal slots.

  The big open box is a space for the user to type in custom modifications not listed in Unimod. The manual fine control format has the form:
  Modification_Name / Mass_Delta @ Targets | Fine_Control

  Modification_Name / is optional. The Targets field allows the 20 one-letter amino acid abbreviations, as well as four special locations: NTerm, CTerm, Protein NTerm, and

Protein CTerm. NTerm, CTerm, Protein NTerm, and Protein CTerm can also be used as modifiers of amino acid residues. Targets form a comma-separated list.

Here is an example of a real modification not (yet) in Unimod:

DehydroFormyl / +9.98435 @ NTerm S, NTerm T | rare1

A limited number of nonstandard amino acid residues is supported by redefining one-letter amino acid abbreviations using fixed modifications. B, Z, U, O, J, and X are accepted in FASTA protein databases, with masses, respectively, of 114.042927 (same as N), 128.058578 (same as Q), 150.95363 (selenocysteine), 237.052645 (pyrrolysine), 100.0, and 110.05 (close to averagine). By placing, for example, a fixed modification of +13.04768 on J, the user can make J in a FASTA database have mass 113.04768, correct for hydroxyproline. However, amino acid sequence is used to predict peak intensity, so this fixed modification on J will not give the same scores as a +15.9949 variable modification on P.

For comprehensive sequence variant searches, or other searches with large numbers of modifications, it is more convenient to paste in a list of modifications in the custom modification box than to add all the modifications via the drop-down menus. Sequence variant lists are available from Protein Metrics by contacting support@proteinmetrics.com.

- **Glycans**

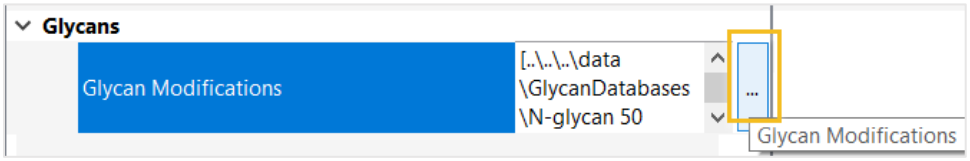


Figure 120: Click within the text box to activate the light blue "…" button

Three ways to define glycan modifications are offered: internal preset tables, external glycan databases, and user-defined glycans. Clicking on the activated light blue "…" button pops up a window labeled **Select Glycans** as shown below.

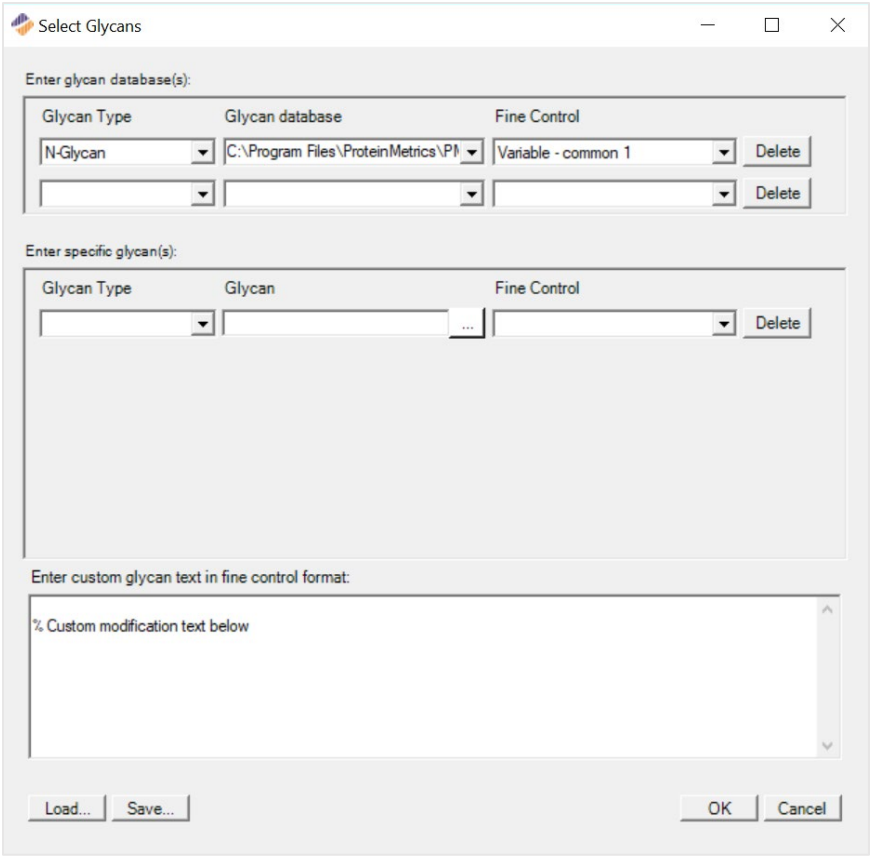


Figure 121: The Select Glycans window

Enter glycan databse(s) allows the user to input a database of glycans. Clicking on the drop-down menu and selecting **Browse** takes the user to the C:\Program Files\ProteinMetrics\PMI-Suite\Base\data\GlycanDatabases folder where the user can select from the databases currently included by default with installation. These databases are editable and the user can also add customized databases to this folder to select instead. We continually add to this set of glycan databses based on customer feedback – please reach out to support@proteinmetrics.com to request additional content.

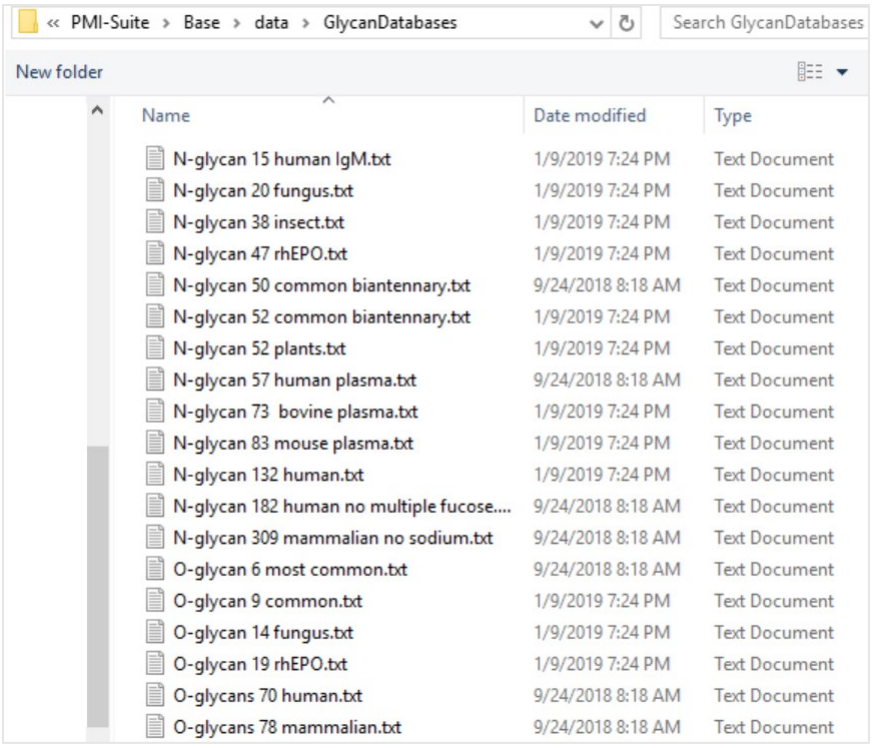


Figure 122: Glycan databases included with installation

The user can choose Glycan type (N- or O-linked), browse for a text file of glycan compositions, and then set the Fine Control (rare1, common2, etc.).

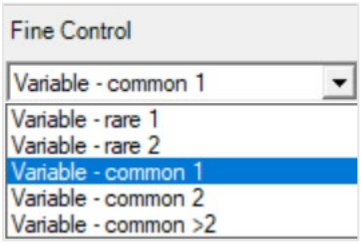


Figure 123: Modification Fine Control options

The text files include one glycan composition per line; for example, the following gives five of the most common human O-glycans. Spaces between monosaccharides are optional, and unused monosaccharides can be left out or included with zero (0) occurrences.

HexNAc(1)
HexNAc(1) Hex(1)
HexNAc(1)Hex(1) NeuAc(1)
HexNAc(1)Hex(1) NeuAc(2)
HexNAc(1)Hex(1)Fuc(1)

Enter specific glycan(s) is located in the middle part of Select Glycans and allows the user to input glycans one at a time by specifying monosaccharide compositions.

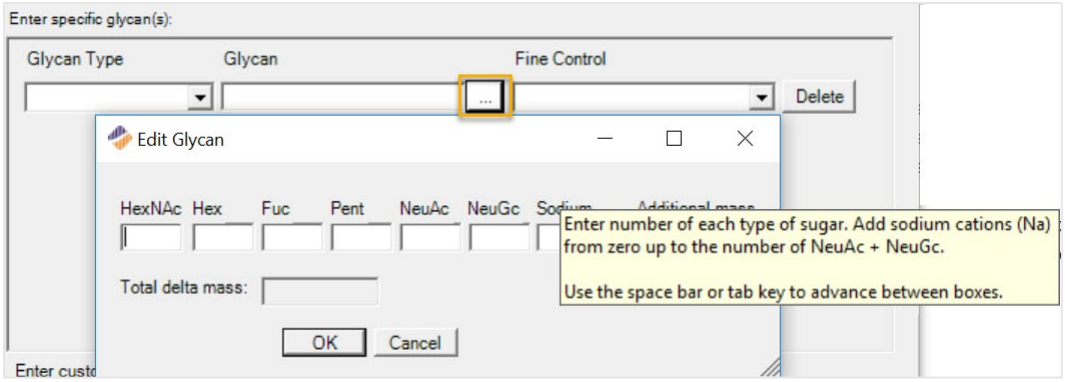


Figure 124: Enter specific glycans

Here the user chooses Glycan type (N- or O-linked) from the drop-down menu, builds a custom glycan by clicking on the white "…" box and entering values in the **Edit Glycan** window, and then sets the Fine Control (rare1, common2, etc.). Six monosaccharide residues are allowed: HexNAc, Hexose, Fucose, Pentose (common in plants), NeuAc, and NeuGc (common in non-humans). A box for Sodium is also included because this is a common adduct on sialic acids. Other glycan masses and modifications such as sulfation and acetylation can be defined with the Additional mass box; this mass is added to the mass of the monosaccharides. The total delta mass will automatically populate. The user selects OK to close the window and include the value.

In the third section, the user can add additional glycans using the same format as within the glycan database text files.

For some helpful examples and best practices for conducting N-linked and O-linked glycan searches, see our Application Notes at https://www.proteinmetrics.com/resources/.

- **Spectrum Input Options**

  These parameters help Byos cope with imperfect inputs. For example, on many MS instruments, precursor ion charges are uncertain for some or all spectra.

Spectrum Input Options	
Apply Charges To	unassigned spectra
Charge States	
Precursor Isotope Off by X	No error check
Maximum Precursor Mass	12,000
Precursor and Charge Assignments	Compute from MS1
Maximum Number of Precursors per Scan	1
Smoothing Width (m/z)	0.01

Figure 125: Spectrum Input Options

  - **Apply Charges To** - By default, the assigned charge will be used for all spectra with assigned charges and +1, +2, +3 will be used for all CID spectra and +2, +3, +4 for all ETD spectra without assigned charges. The Apply Charges To parameter allows the user to override this default setting by instead selecting unassigned spectra.
  - **Charge States** - All comma-separated charges detailed to each spectrum will be applied (based on the values the user entered into the Charge States box).
  - **Precursor Isotope Off By X**

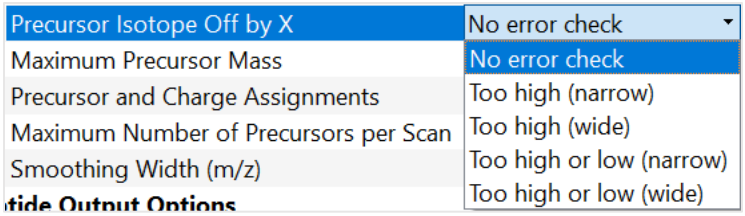


Figure 126: Precursor Isotope Off By X

Similarly, on many instruments the nominal precursor mass may actually be the mass of a 13C isotope peak rather than of the base (all 12C monoisotopic) peak, so the true precursor mass will within 10 ppm of 2350.120 Da or within 10 ppm of 2351.123 Da. Precursor Isotope Off By X is a pulldown menu with several options.

No error check will use only the assigned precursor; **Too high (narrow)** will allow the assigned precursor to be up to 2 Da too high; **Too high (wide)** will allow the assigned precursor to be up to n Da too high for a precursor of mass at least 1000n Da; **Too high or low (narrow)** will allow the assigned precursor to be up to 2 Da too high or 2 Da too low; **Too high or low (wide)** will allow the assigned precursor to be up to n Da too high or 2 Da too low for a precursor of mass at least 1000n Da.

- **Maximum Precursor Mass** - The user can set the Maximum Precursor Mass to be considered.
- **Precursor and Charge Assignments**

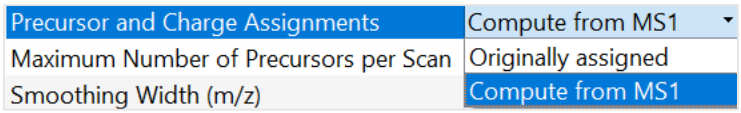


Figure 127: Precursor and Charge Assignments

The precursor and charge assignments can also be calculated directly from the MS1 data or use the originally assigned values. The user can set this using a drop-down menu.

- **Maximum Number of Precursors per Scan** - Multiple precursors per scan is also considered – it is recommended for the user to set this to 2 for complex samples and 5-10 if processing MSe or DIA data.
- **Smoothing Width** - The user can enter a sigma value in Thomsons for Gaussian smoothing and centroiding of Waters or Sciex data. Half-width at peak half maximum (~0.01 m/z) works well and is the default value already entered for the user.

- **Peptide Output Options**

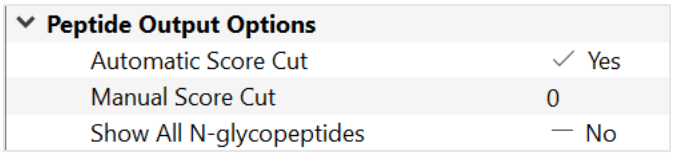


Figure 128: Peptide Output Options

The Peptide Output Options parameters offer options for filtering the peptide-spectrum matches (PSMs) by score. By default, PSM filtering is deferred until after protein ranking, and then filters to control PSM FDR on the “true” proteins—those ranked above the top-ranking decoy protein. This method gains sensitivity while simultaneously reducing both protein and PSM FDRs. (Two-dimensional target decoy strategy for shotgun proteomics, Journal of Proteome Research 10 (12), 5296-5301, 2011.)

- **Automatic Score Cut** - To filter PSMs before protein ranking, the user can click **Yes** to activate Automatic score cut and type in a minimum score.
- **Manual Score Cut** - The user can activate Automatic Score Cut and enter a minimum score. For example, a score threshold of 200 will remove weak matches and a threshold of 400 will remove all but the best matches. Filtering by score may be helpful in special cases, for example to eliminate from consideration all but the best wildcard PSMs.
- **Show All N-Glycopeptides** - The user also has the option to Show All N-Glycopeptides. This will show N-Glycopeptide matches regardless of score or FDR. This is recommended for simple samples only. This can be especially useful for low energy CID data.

- **Protein Output Options**



Figure 129: Protein Output Options

- **Protein FDR**

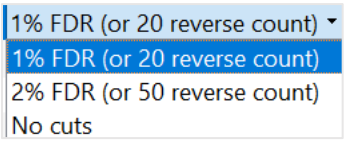


Figure 130: Protein FDR options

This gives the user control of the protein list cut-off. By default, the protein list is cut at 1% protein FDR or 20 decoy proteins, whichever comes last, but the user can ask for 2% protein FDR or a completely unfiltered (but still ranked) protein list, or No Cuts.

- **Create a Focused Database** - If the user clicks this to select **Yes**, the software is directed to to output a new FASTA file (labeled focused and appearing in the output objs directory) containing only the proteins found in the search, along with suitable decoys (>Reverse) for unbiased FDR estimation. The focused database can then be used for subsequent wide searches, including more modifications and/or a wildcard. Of course, the user can also create focused databases outside of the software by editing existing FASTA files.
- **Export mzIdentML** - The user can also select to export this file as a mzIdentML file.

- **Multicore Options**

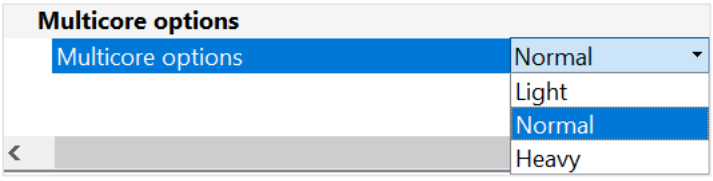


Figure 131: Multicore processing options

The user can control the number of computer cores of the CPU used through the drop-menu. The Light searches uses one core, Normal search uses all available cores minus two, and Heavy search uses all available cores.

- **Wildcard**



Figure 132: Wildcard parameters.

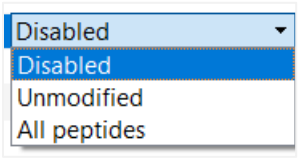


Figure 133: Wildcard Search drop-down options.

- Wildcard lets the user turn on wildcard searches, set the range for the wildcard mass, and restrict the wildcard to certain residues if desired. The options Disabled, Unmodified, and All peptides are available through a drop-down menu.

  The **Restrict to residues** box uses the common 20 single-letter residue abbreviations, and (lower case) n denotes peptide N-terminus and (lower case) c denotes peptide C-terminus. A wildcard, even one with a mass range of only 50 or 60 Da, greatly increases the size of the search. It is best used with a focused database (see the Advanced tab section below) and used either alone or with only a few other modifications enabled. Most wildcard mass shifts will be recognizable by an expert; hence, a wildcard can be used to discover which known modifications should be enabled in a subsequent search. For more details about the wildcard search, see the application note “Byonic™: Wildcard Search™” at https://www.proteinmetrics.com/resources/#application-notes.

- By specifying most modifications as rare, it is quite feasible to search for 10 – 20 modification types at once with Byonic. Even larger searches are possible with focused protein databases, for example with therapeutic proteins. Such a focused database easily allows efficient mutation searches with 200+ possible substitutions, or oxidative footprinting searches with 50+ types of oxidations. Glycans and wildcards can easily enlarge the search space by 2 to 3 orders of magnitude, so these options should be used with care, and in conjunction with only the most common variable modifications (such as oxidized methionine or pyro-Glu N-terminus). **NOTE**: The single most important factor in search time is Total common max. Roughly speaking, the search time grows as C^T where C is the number of common modifications enabled and T is Total common max.

- The **Appendix** of the Byonic Manual provides examples of frequently found modifications and appropriate syntax for including those modifications in a Byonic search.

Chromatographic Quant

NOTE: As mentioned in the above MS/MS Ids section, these parameters apply alone to the Reference Chromatography (in-silico), Characterization Chromatography (in-silico), MAM Chromatogram Annotation (in-silico), and MAM New Peak Detection Workflows.

- **General**

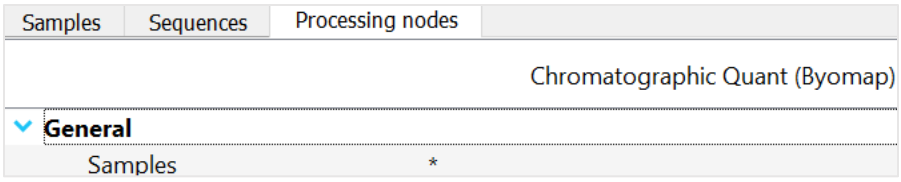


Figure 134: Samples to include in the project

- **Samples** - The "*" will apply all parameters to all samples dragged and dropped into the Samples tab.

- **Candidate Matching**

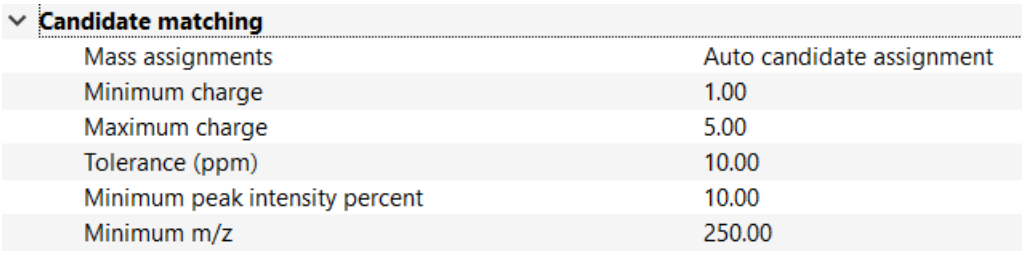


Figure 135: Candidate matching

These parameters allow the user to set mass assignment, charge range and other candidate matching filters before project creation.

- **In-silico options**

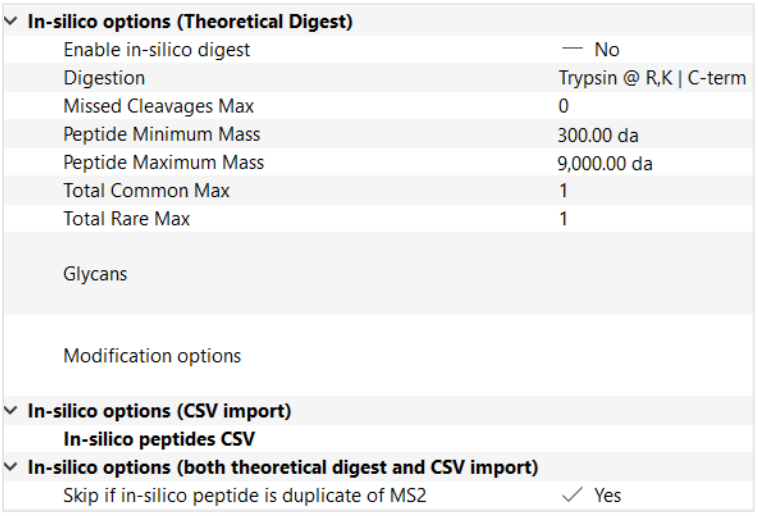


Figure 136: In-silico options

Parameters are separated into three sections to clarify which pertain to a **Theoretical Digest**, **CSV import** and **both theoretical digest and CSV import.**

- **Enable In-Silico** - The user has the option to click **Yes** to enable this parameter to generate an in-silico list of peptides. The below additional paremeters must then be set. NOTE: The option to add a list of in-silico peptides from a CSV file is offered through In-Silico Peptides CSV.
- **Digestion**

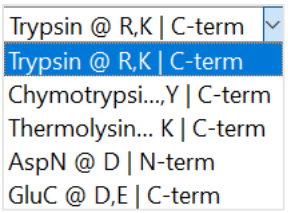


Figure 137: Digestion options

The user can view several options through the drop-down menu. These are detailed as one letter residue codes adjacent to cleavage points. The user may enter a customized digestion by typing using the following syntax: Name @ Amino acid letters | C-term or N-term.

- **Missed Cleavages Max**

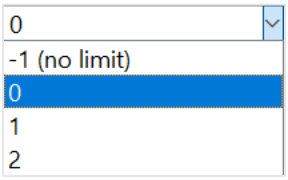
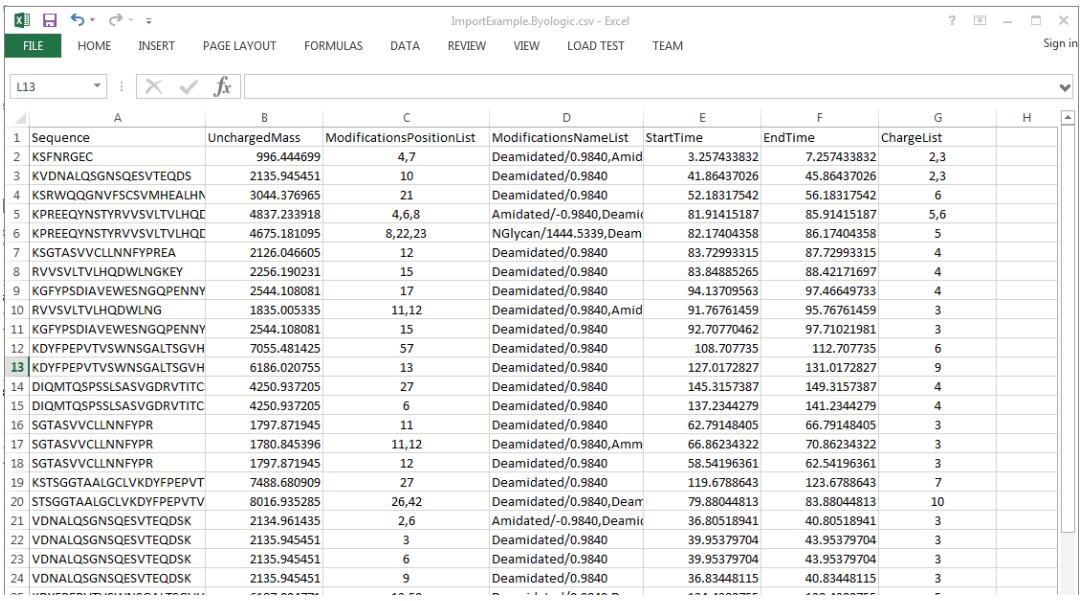


Figure 138: Missed Cleavages Max options

The user can set the maximum number of missed cleavages per peptide by selecting a value using the drop-down menu. A value of -1 allows any number of cleavages.

- **Peptide Minumum Mass** - The user can set the minimum value of the peptide mass range.
- **Peptide Maximum Mass** - The user can set the maximum value of the peptide mass range.
- **Total Common Max** - The user can set the maximum number of common modifications per peptide. The search size grows by an order of magnitude with each increase in Total Common Max.
- **Total Rare Max** - The user can set the maximum number of rare modifications per peptide. The search size grows by an order of magnitude with each increase in Total Rare Max.
- **Glycans** - The user can specify glycan databsesor custom glycans to include in the in-silico generation. This is done the same as detailed in the MS/MS Ids section.
- **Modifications Options** - The user can specify modifications to include in the in-silico generation. This is done the same as detailed in the MS/MS Ids section.
- **In-Silico Peptides CSV** - This option allows the user to to add a list of in-silico pepitdes from an imported CSV file, thus there is no need to run an MS2 search if there are known modifications with known masses and retention times. The format for the CSV is shown below.



	A	B	C	D	E	F	G	H
1	Sequence	UnchargedMass	ModificationsPositionList	ModificationsNameList	StartTime	EndTime	ChargeList	
2	KSFNRGEC	996.444699	4,7	Deamidated/0.9840,Amid	3.257433832	7.257433832	2,3	
3	KVDNALQSGNSQESVTEQDS	2135.945451	10	Deamidated/0.9840	41.86437026	45.86437026	2,3	
4	KSRWQQGNVFSCSVMHEALHN	3044.376965	21	Deamidated/0.9840	52.18317542	56.18317542	6	
5	KPREEQYNSTYRVVSVLTVLHQD	4837.233918	4,6,8	Amidated/-0.9840,Deamic	81.91415187	85.91415187	5,6	
6	KPREEQYNSTYRVVSVLTVLHQD	4675.181095	8,22,23	NGlycan/1444.5339,Deam	82.17404358	86.17404358	5	
7	KSGTASVVCLLNNFYPREA	2126.046605	12	Deamidated/0.9840	83.72993315	87.72993315	4	
8	RVVSVLTVLHQDWLNGKEY	2256.190231	15	Deamidated/0.9840	83.84885265	88.42171697	4	
9	KGFYPSDIAVEWESNGQPENNY	2544.108081	17	Deamidated/0.9840	94.13709563	97.46649733	4	
10	RVVSVLTVLHQDWLNG	1835.005335	11,12	Deamidated/0.9840,Amid	91.76761459	95.76761459	3	
11	KGFYPSDIAVEWESNGQPENNY	2544.108081	15	Deamidated/0.9840	92.70770462	97.71021981	3	
12	KDYFPEPVTVSWNSGALTSGVH	7055.481425	57	Deamidated/0.9840	108.707735	112.707735	6	
13	KDYFPEPVTVSWNSGALTSGVH	6186.020755	13	Deamidated/0.9840	127.0172827	131.0172827	9	
14	DIQMTQSPSSLSASVGDRVTITC	4250.937205	27	Deamidated/0.9840	145.3157387	149.3157387	4	
15	DIQMTQSPSSLSASVGDRVTITC	4250.937205	6	Deamidated/0.9840	137.2344279	141.2344279	4	
16	SGTASVVCLLNNFYPR	1797.871945	11	Deamidated/0.9840	62.79148405	66.79148405	3	
17	SGTASVVCLLNNFYPR	1780.845396	11,12	Deamidated/0.9840,Amm	66.86234322	70.86234322	3	
18	SGTASVVCLLNNFYPR	1797.871945	12	Deamidated/0.9840	58.54196361	62.54196361	3	
19	KSTSGGTAALGCLVKDYFPEPVT	7488.680909	27	Deamidated/0.9840	119.6788643	123.6788643	7	
20	STSGGTAALGCLVKDYFPEPVTV	8016.935285	26,42	Deamidated/0.9840,Deam	79.88044813	83.88044813	10	
21	VDNALQSGNSQESVTEQDSK	2134.961435	2,6	Amidated/-0.9840,Deamic	36.80518941	40.80518941	3	
22	VDNALQSGNSQESVTEQDSK	2135.945451	3	Deamidated/0.9840	39.95379704	43.95379704	3	
23	VDNALQSGNSQESVTEQDSK	2135.945451	6	Deamidated/0.9840	39.95379704	43.95379704	3	
24	VDNALQSGNSQESVTEQDSK	2135.945451	9	Deamidated/0.9840	36.83448115	40.83448115	3	

Figure 139: Format for an In-Silico Peptides CSV

- **Skip if in-silico peptide is duplocate of MS2** - The user can click to select **Yes**. This will skip any generated in-silico peptide if one already exists from the MS2 data. It is recommended to make this selction when generating in-silico peptides.

- **Advanced**

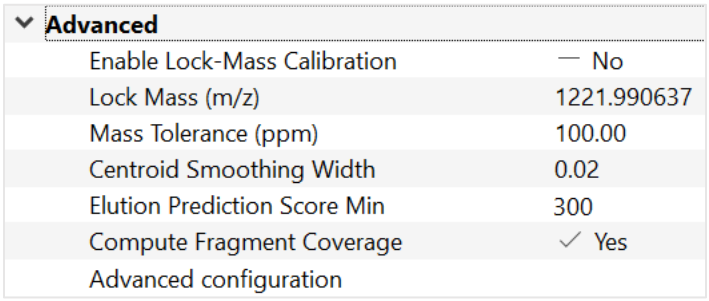


Figure 140: Advanced parameters

- **Enable Lock-Mass Calibration** - The user can click to select between **No** and **Yes**.
- **Lock Mass (m/z)**

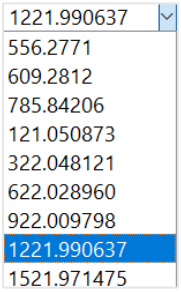


Figure 141: Lock mass m/z value options

The user can enter the calibrant m/z value. Several are also available using the drop-down menu or the user can type in a numerical value. If empty, no calibration will be applied.

- **Mass Tolerance (ppm)** - The user can enter the calibrant m/z mass tolerance value in ppm. The user can type in a numerical value or use the up and down arrow keys to increase or decrease the value. If empty, no calibration will be applied.
- **Centroid Smoothing Width** - The user can enter a sigma value in Thomsons for Gaussian smoothing and centroiding of Waters or Sciex data. Half-width at peak half maximum (~0.01 m/z) works well and is the default value already entered for the user.
- **Elution Prediction Score Min** - This filter will find wildtype peptides with scores above this value to adjust the elution time.
- **Compute Fragment Coverage** - The user can click between **Yes** and **No** to compute fragmentation coverage at project creation.
- **Advanced Configuration** - The user can enter text commands to complete advanced processing. These are detailed in the Release Notes included with each quarterly release. Please reach out to support@proteinmetrics.com for additional details.

- **Report**

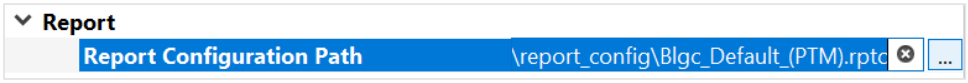


Figure 142: Report Configuration Path

Each Byos default workflow includes a report template created by our Customer Success team that is optimized for the specific type of analysis. If the user prefers a customized report template, they can direct Byos to this file using the light blue "…" button. They will be prompted to select a file.

- **UI Configurations**

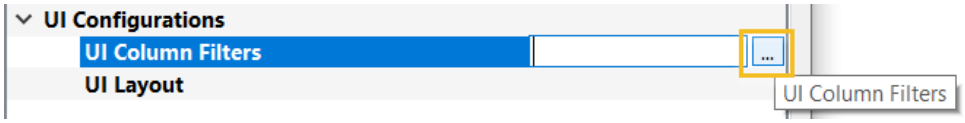


Figure 143: UI Configurations – UI Column Filters and UI Layout.

The user has the ability to import column filters as well as layout files. This is designed to standardize analyses across all users, labs, and sites. The user can direct Byos to the preferred file for each using the light blue "…" button. The user will be prompted to select a file.

- **Time Settings**



Figure 144: Time Settings

- **Start time of interest** and **End time of interest** control the time limits within which the computation of the baseline and peaks are completed.
- **Alignment max time** sets the maximum alignment value between plots (for example, UV and TIC). This value will limit the allowed alignment time between the two signals.

The default values for both are set to 0.00, which means filtering will not be applied. The user can set them to apply the additional processing options. The default value for "Alignment max time" is set to 1.00.

- **Peak Construction options** (older versions of Byos)

  The Peak Construction option parameters formerly found in Byos are now set directly in the Samples tab with the buttons **Edit elution peak options** and **Edit global elution peak filters**.

Detached Glycans (N-Linked, O-Linked, and IgG) Workflows

The following section details parameters in the Glycan table, Glycan options, and Processing nodes tabs for these three workflows.

Glycan table

Each default workflow is populated with a table of Glycans specific to the workflow – N-linked, O-linked, or IgG. The "Glycan alias name" and "Glycan composition" are detailed for each entry. Please note the start and end times are left blank by default.

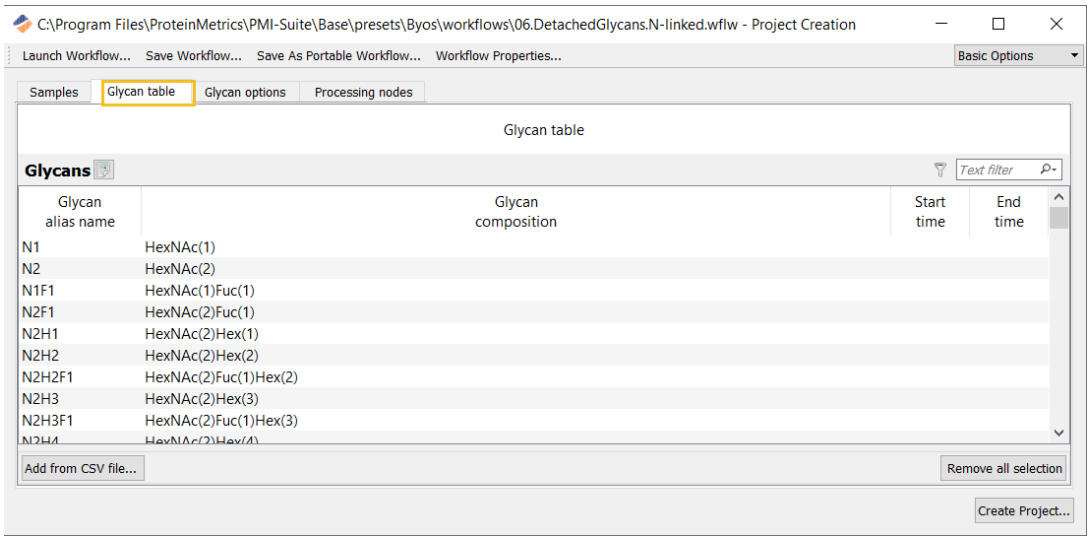


Figure 145: Glycan table tab in the Detached Glycan workflows

The user has many options while working with these databases – the current content is editable. The user may also opt to add to the current content or even delete the current content and replace it with a customized database. This is done by highlighting the content and clicking on **Remove all selection**, as shown below:

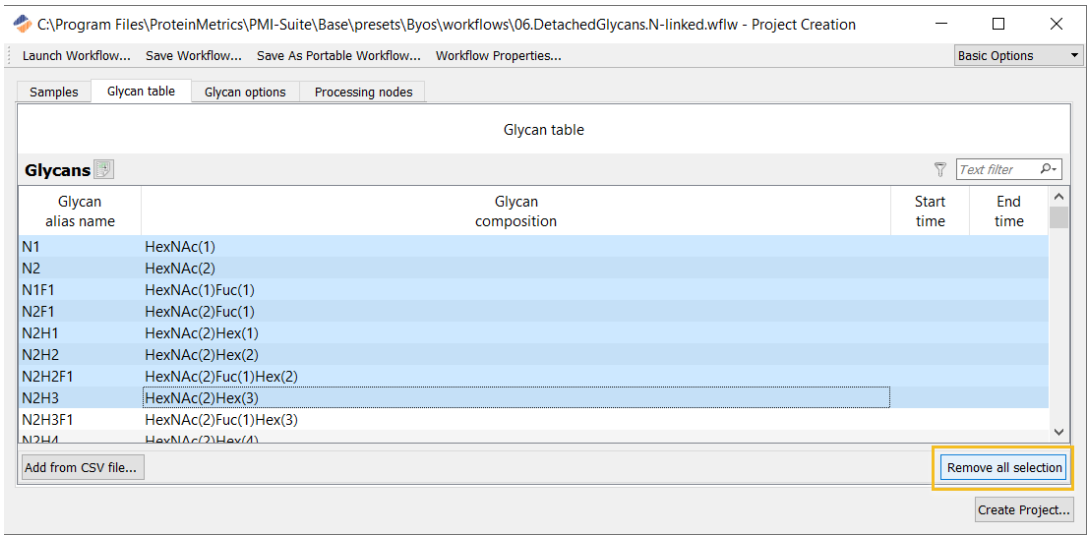


Figure 146: Remove all selection

The user would then click **Add from CSV file** and direct Byos to the desired database, as shown below:

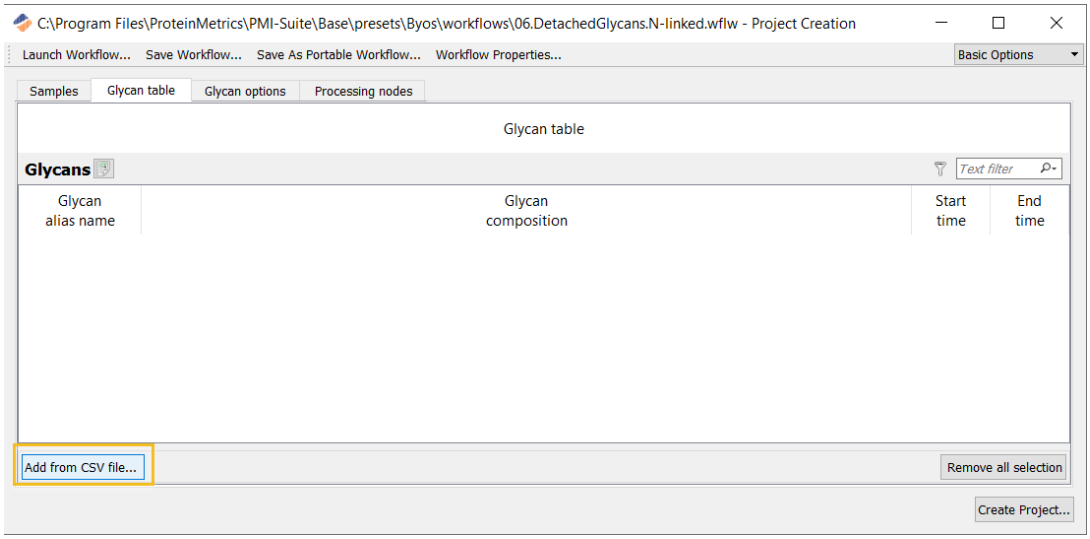


Figure 147: Glycan table, Add from CSV file to import a custom Glycan database

Glycan options

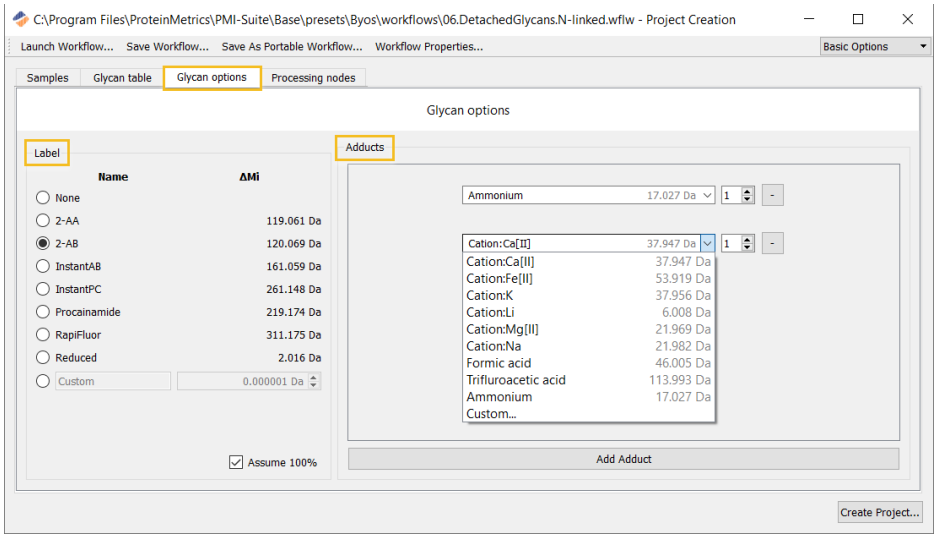


Figure 148: Glycan options tab – identify the label used and set Adducts

The user has the option to define the label used during the experiment. The delta mass of each tag is defined. The user can also define 100% labeling.

The user can also include adducts to consider in the mass calculation in this same tab. Click **Add Adduct** will make the displayed drop-down menu available. The user can click the button again to add another adduct.

Processing nodes

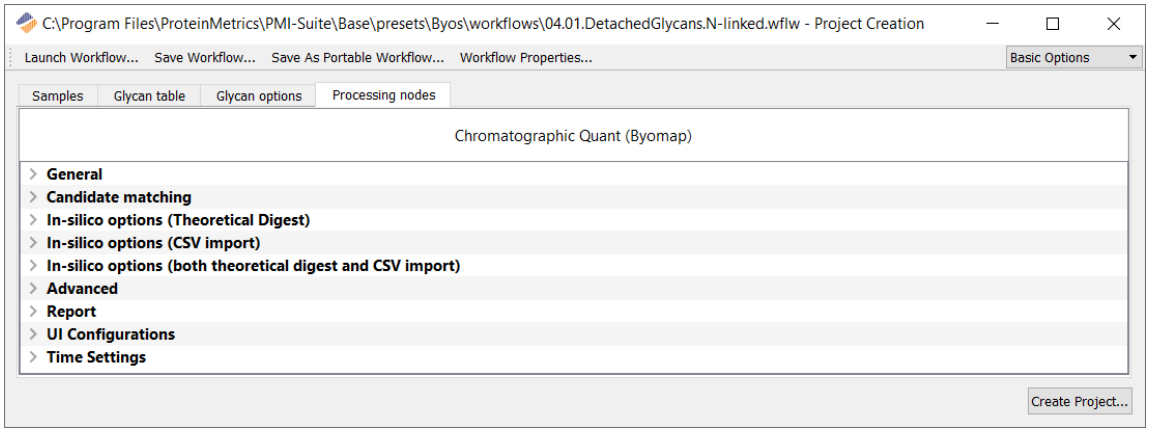


Figure 149: Processing nodes

Chromatographic Quant

- **General**

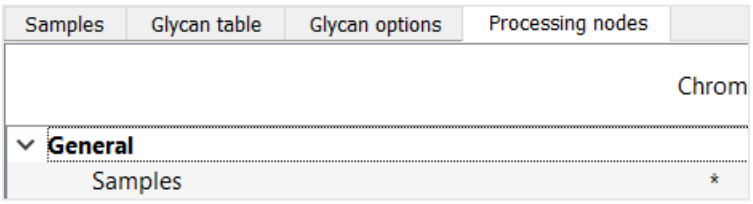
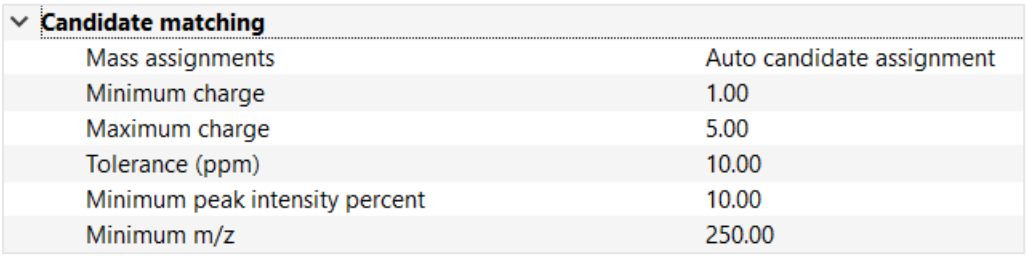


Figure 150: Samples to include in the project

- **Samples** - The "*" will apply all parameters to all samples dragged and dropped into the Samples tab.

- **Candidate matching**



These parameters allow the user to set mass assignment, charge range and other candidate matching filters before project creation.

- **In-silico options**

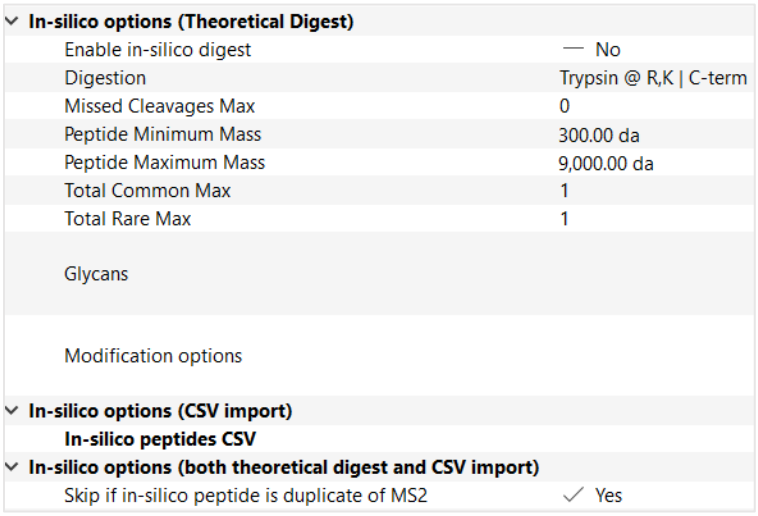


Figure 151: In-silico options processing parameters.

Parameters are separated into three sections to clarify which pertain to a **Theoretical Digest**, **CSV import** and **both theoretical digest and CSV import.**

- **Enable In-Silico** - The user has the option to click **Yes** to enable this parameter to generate an in-silico list of peptides. The below additional parameters must then be set.

  NOTE: The option to add a list of in-silico peptides from a CSV file is offered through In-Silico Peptides CSV.

- **Digestion**

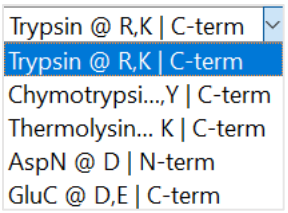


Figure 152: Digestion options

The user can view several options through the drop-down menu. These are detailed as one letter residue codes adjacent to cleavage points. The user may enter a customized digestion by typing using the following syntax: Name @ Amino acid letters | C-term or N-term.

- **Missed Cleavages Max**

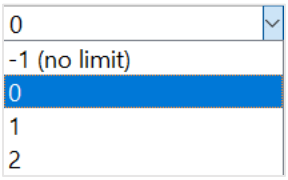
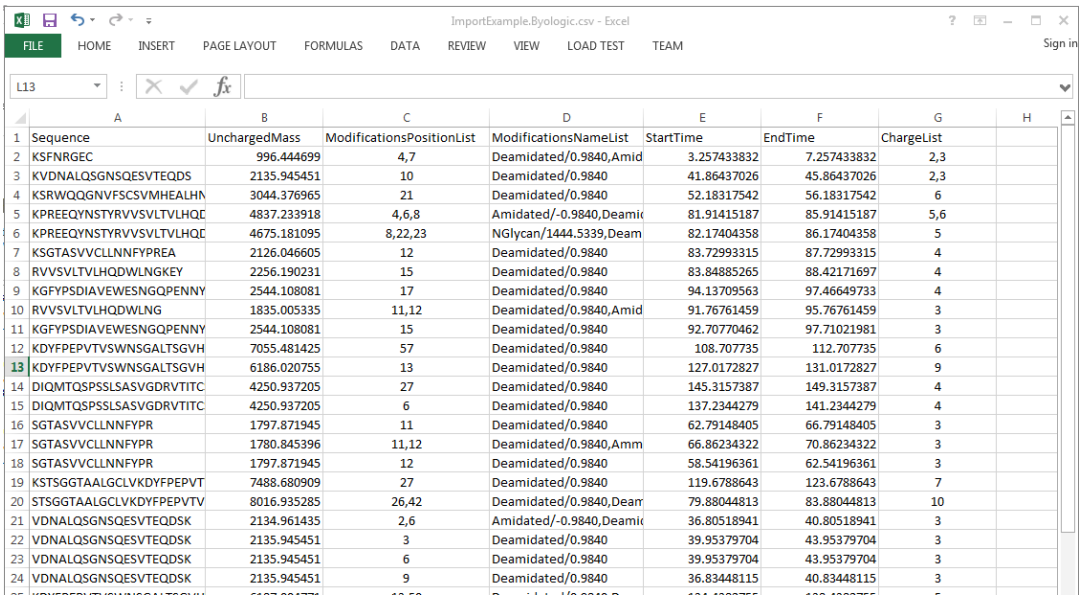


Figure 153: Missed Cleavages Max options

The user can set the maximum number of missed cleavages per peptide by selecting a value using the drop-down menu. A value of -1 allows any number of cleavages.

- **Peptide Minimum Mass** - The user can set the minimum value of the peptide mass range.
- **Peptide Maximum Mass** - The user can set the maximum value of the peptide mass range.
- **Total Common Max** - The user can set the maximum number of common modifications per peptide. The search size grows by an order of magnitude with each increase in Total Common Max.
- **Total Rare Max** - The user can set the maximum number of rare modifications per peptide. The search size grows by an order of magnitude with each increase in Total Rare Max.
- **Glycans** - The user can specify glycan databases or custom glycans to include in the in-silico generation. This is done the same as detailed in the MS/MS Ids section.
- **Modifications Options** - The user can specify modifications to include in the in-silico generation. This is done the same as detailed in the MS/MS Ids section.
- **In-Silico Peptides CSV** - This option allows the user to add a list of in-silico peptides from an imported CSV file, thus there is no need to run an MS2 search if there are known modifications with known masses and retention times. The format for the CSV is shown below.



ImportExample.Byologic.csv - Excel

FILE HOME INSERT PAGE LAYOUT FORMULAS DATA REVIEW VIEW LOAD TEST TEAM Sign in

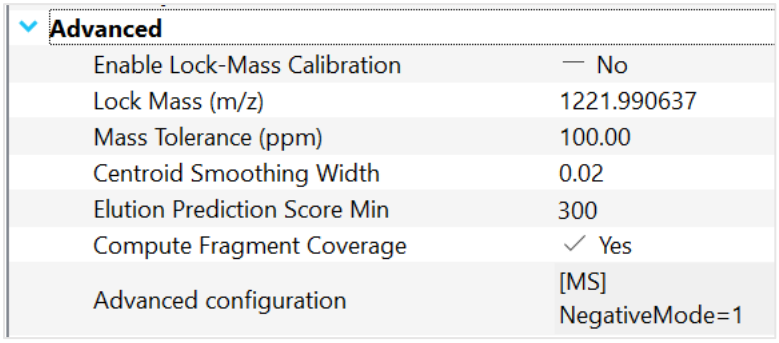
L13

	A	B	C	D	E	F	G	H
1	Sequence	UnchargedMass	ModificationsPositionList	ModificationsNameList	StartTime	EndTime	ChargeList	
2	KSFNRGEC	996.444699	4,7	Deamidated/0.9840,Amid	3.257433832	7.257433832	2,3	
3	KVDNALQSGNSQESVTEQDS	2135.945451	10	Deamidated/0.9840	41.86437026	45.86437026	2,3	
4	KSRWQQGNVFSCSVMHEALHN	3044.376965	21	Deamidated/0.9840	52.18317542	56.18317542	6	
5	KPREEQYNSTYRVVSVLTVLHQD	4837.233918	4,6,8	Amidated/-0.9840,Deamic	81.91415187	85.91415187	5,6	
6	KPREEQYNSTYRVVSVLTVLHQD	4675.181095	8,22,23	NGlycan/1444.5339,Deam	82.17404358	86.17404358	5	
7	KSGTASVVCLLNNFYPREA	2126.046605	12	Deamidated/0.9840	83.72993315	87.72993315	4	
8	RVVSVLTVLHQDWLNGKEY	2256.190231	15	Deamidated/0.9840	83.84885265	88.42171697	4	
9	KGFYPSDIAVEWESNGQPENNY	2544.108081	17	Deamidated/0.9840	94.13709563	97.46649733	4	
10	RVVSVLTVLHQDWLNG	1835.005335	11,12	Deamidated/0.9840,Amid	91.76761459	95.76761459	3	
11	KGFYPSDIAVEWESNGQPENNY	2544.108081	15	Deamidated/0.9840	92.70770462	97.71021981	3	
12	KDYFPEPVTVSWNSGALTSGVH	7055.481425	57	Deamidated/0.9840	108.707735	112.707735	6	
13	KDYFPEPVTVSWNSGALTSGVH	6186.020755	13	Deamidated/0.9840	127.0172827	131.0172827	9	
14	DIQMTQSPSSLSASVGDRVTITC	4250.937205	27	Deamidated/0.9840	145.3157387	149.3157387	4	
15	DIQMTQSPSSLSASVGDRVTITC	4250.937205	6	Deamidated/0.9840	137.2344279	141.2344279	4	
16	SGTASVVCLLNNFYPR	1797.871945	11	Deamidated/0.9840	62.79148405	66.79148405	3	
17	SGTASVVCLLNNFYPR	1780.845396	11,12	Deamidated/0.9840,Amm	66.86234322	70.86234322	3	
18	SGTASVVCLLNNFYPR	1797.871945	12	Deamidated/0.9840	58.54196361	62.54196361	3	
19	KSTSGGTAALGCLVKDYFPEPVT	7488.680909	27	Deamidated/0.9840	119.6788643	123.6788643	7	
20	STSGGTAALGCLVKDYFPEPVTV	8016.935285	26,42	Deamidated/0.9840,Deam	79.88044813	83.88044813	10	
21	VDNALQSGNSQESVTEQDSK	2134.961435	2,6	Amidated/-0.9840,Deamic	36.80518941	40.80518941	3	
22	VDNALQSGNSQESVTEQDSK	2135.945451	3	Deamidated/0.9840	39.95379704	43.95379704	3	
23	VDNALQSGNSQESVTEQDSK	2135.945451	6	Deamidated/0.9840	39.95379704	43.95379704	3	
24	VDNALQSGNSQESVTEQDSK	2135.945451	9	Deamidated/0.9840	36.83448115	40.83448115	3	

Figure 154: Format for an In-Silico Peptides CSV

- **Skip if in-silico peptide is duplicate of MS2** - The user can click to select **Yes**. This will skip any generated in-silico peptide if one already exists from the MS2 data. It is recommended to make this selection when generating in-silico peptides.

- **Advanced**



Advanced	
Enable Lock-Mass Calibration	— No
Lock Mass (m/z)	1221.990637
Mass Tolerance (ppm)	100.00
Centroid Smoothing Width	0.02
Elution Prediction Score Min	300
Compute Fragment Coverage	✓ Yes
Advanced configuration	[MS] NegativeMode=1

Figure 155: Advanced options

- **Enable Lock-Mass Calibration** - The user can click to select between **No** and **Yes**.
- **Lock Mass (m/z)**

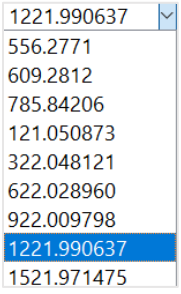


Figure 156: Lock mass m/z value options

The user can enter the calibrant m/z value. Several are also available using the drop-down menu or the user can type in a numerical value. If empty, no calibration will be applied.

- **Mass Tolerance (ppm)** - The user can enter the calibrant m/z mass tolerance value in ppm. The user can type in a numerical value or use the up and down arrow keys to increase or decrease the value. If empty, no calibration will be applied.
- **Centroid Smoothing Width** - The user can enter a sigma value in Thomsons for Gaussian smoothing and centroiding of Waters or Sciex data. Half-width at peak half maximum (~0.01 m/z) works well and is the default value already entered for the user.
- **Elution Prediction Score Min** - The user can set a value for this filter to find wildtype peptides with scores above this value to adjust the elution time.
- **Compute Fragment Coverage** - The user can click between **Yes** and **No** to compute fragmentation coverage at project creation.
- **Advanced Configuration** - **The command to process the data in negative mode is included by default**. The user can enter text commands to complete advanced processing. These are detailed in the Release Notes included with each quarterly release. Please reach out to support@proteinmetrics.com for additional details.

- **Report**

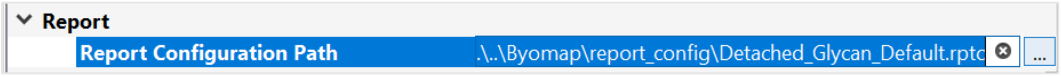


Figure 157: Report Configuration Path

Each Byos default workflow includes a report template created by our Customer Success team that is optimized for the specific type of analysis. If the user prefers a customized report template, they can direct Byos to this file using the light blue "…" button. They will be prompted to select a file.

- **UI Configurations**

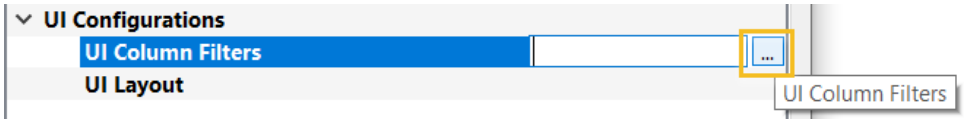


Figure 158: UI Configurations – UI Column Filters and UI Layout.

The user has the ability to import column filters as well as layout files. This is designed to standardize analyses across all users, labs, and sites. The user can direct Byos to the preferred file for each using the light blue "…" button. The user will be prompted to select a file.

- **Time Settings**



Figure 159: Time Settings

- **Start time of interest** and **End time of interest** control the time limits within which the computation of the baseline and peaks are completed.

- **Alignment max time** sets the maximum alignment value between plots (for example, UV and TIC). This value will limit the allowed alignment time between the two signals.

  The default values for both are set to 0.00, which means filtering will not be applied. The user can set them to apply the additional processing options. The default value for "Alignment max time" is set to 1.00.

- **Peak Construction options** (older versions of Byos)
  The Peak Construction option parameters formerly found in Byos are now set directly in the Samples tab with the buttons **Edit elution peak options** and **Edit global elution peak filters**.

De novo Sequencing Workflow

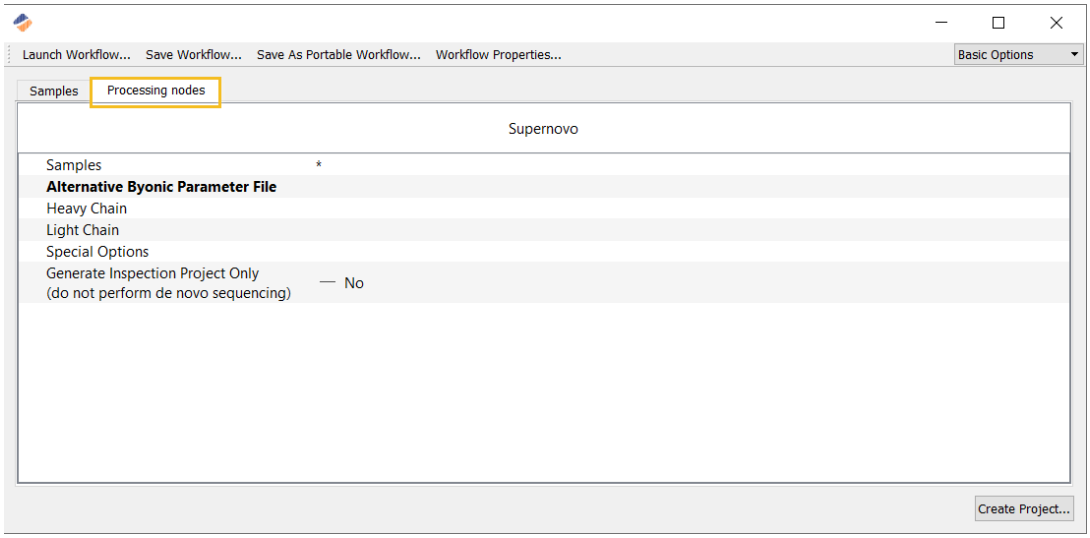


Figure 160: Processing nodes

Processing nodes

- **Samples** - The "*" will apply all parameters to all samples dragged and dropped into the Samples tab.
- **Alternative Byonic Parameter File** [OPTIONAL]

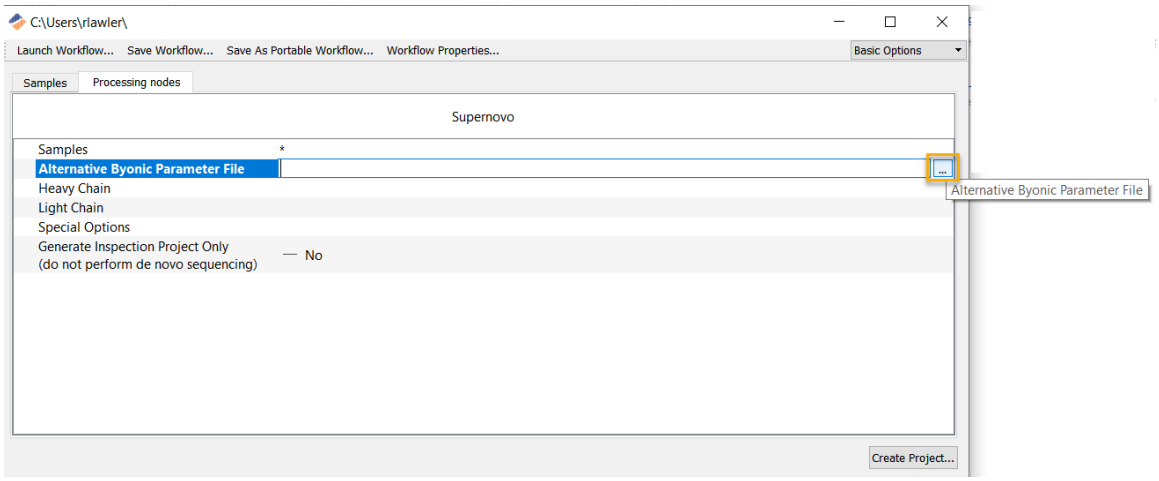


Figure 161: Alternative Byonic Parameter File parameter

Specify an alternative Byonic parameter file (*.byparms) by clicking within the box to activate the light blue "…"button, as detailed in Figure 159. If left blank, Byos uses a default Byonic parameter file, which specifies 10 ppm precursor mass tolerance, 20 ppm fragment mass tolerance, and carbamidomethyl (delta mass +57) for the cysteine alkylation.

- **Heavy Chain and Light Chain** [OPTIONAL]

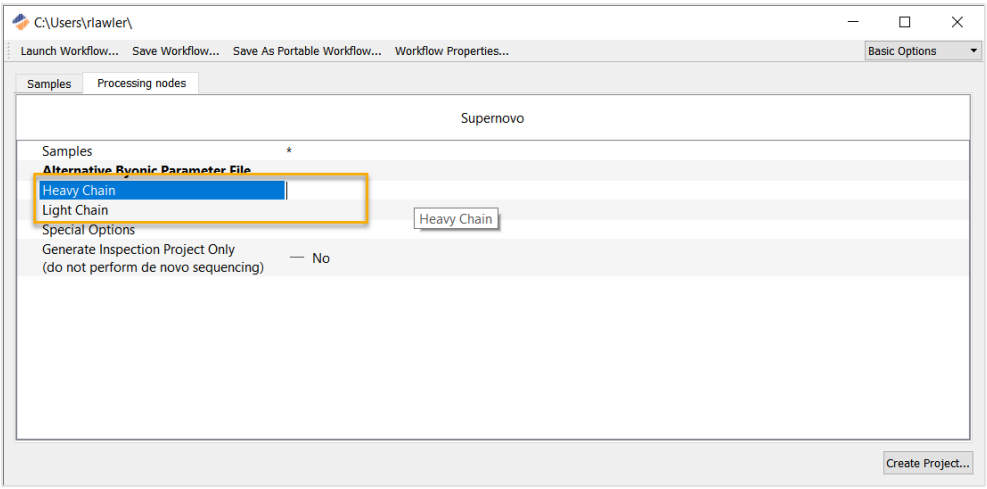


Figure 162: The sequences for the Heavy Chain and Light Chain can be pasted in as text

If a starting sequence template is to be used, paste the Heavy Chain and Light Chain sequences. If left blank, Byos will automatically determine an antibody scaffold sequence as the starting point for de novo analysis. This works best for antibodies from human, mouse, rat, or related species.

  - **Special Options** [OPTIONAL] - Leave this blank generally. Please reach out to support@proteinmetrics.com for additional details.
  - **General Inspection Project Only** - the user can toggle between Yes/No. Check this box to only generate metrics / visualizations for a particular specified sequence. Byos will not perform de novo sequencing for this inspection project. Please note this type of project runs faster.

- Click **Create Project** to start analysis and report generation. Please note Byos will save projects based upon this type of processing as a .blgc file.

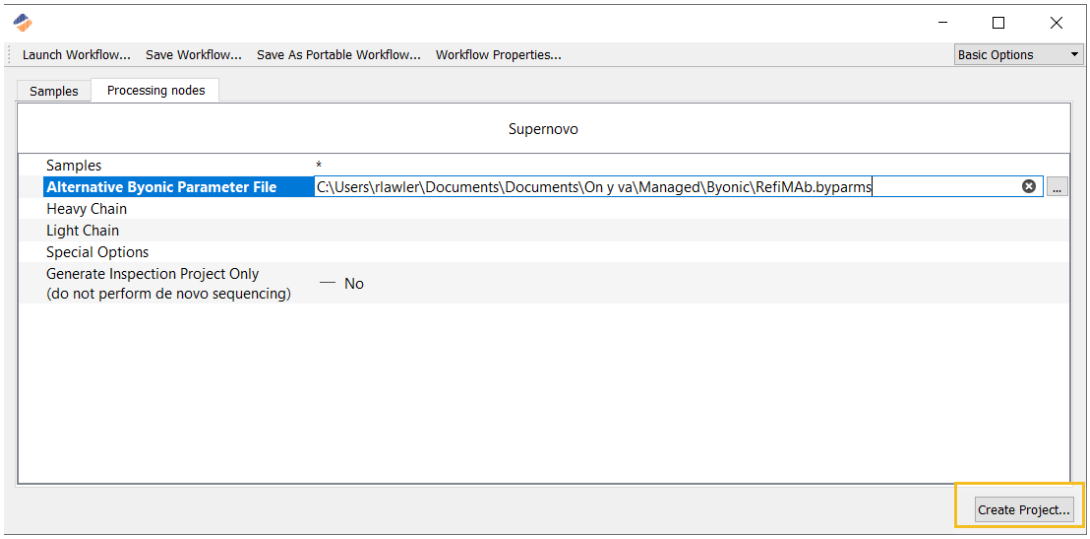


Figure 163: Click Create Project

- Name the project and save in a convenient location.

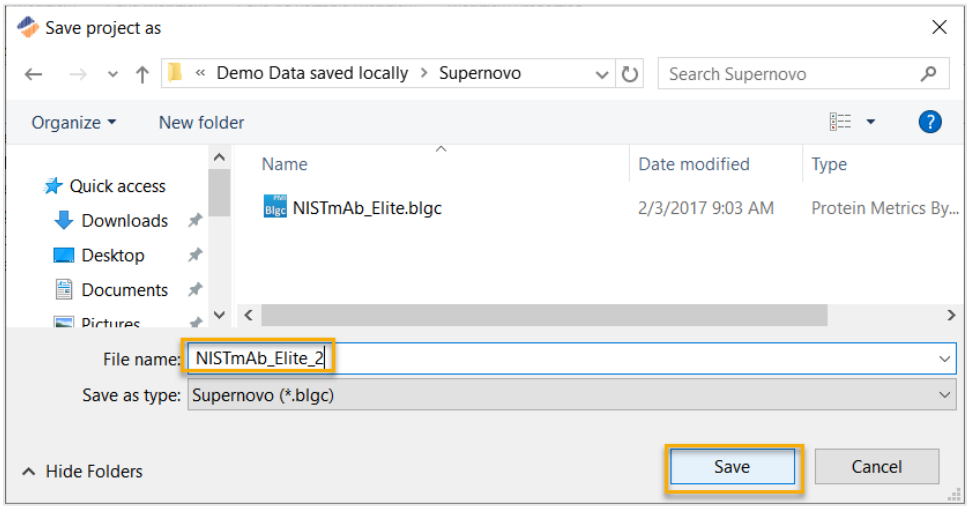


Figure 164: Name the project and save in a convenient location

Processing will commence. Once complete, the project and associated report will appear as additional tabs after the “Byos Workflows” tab, as shown in Figure 163.

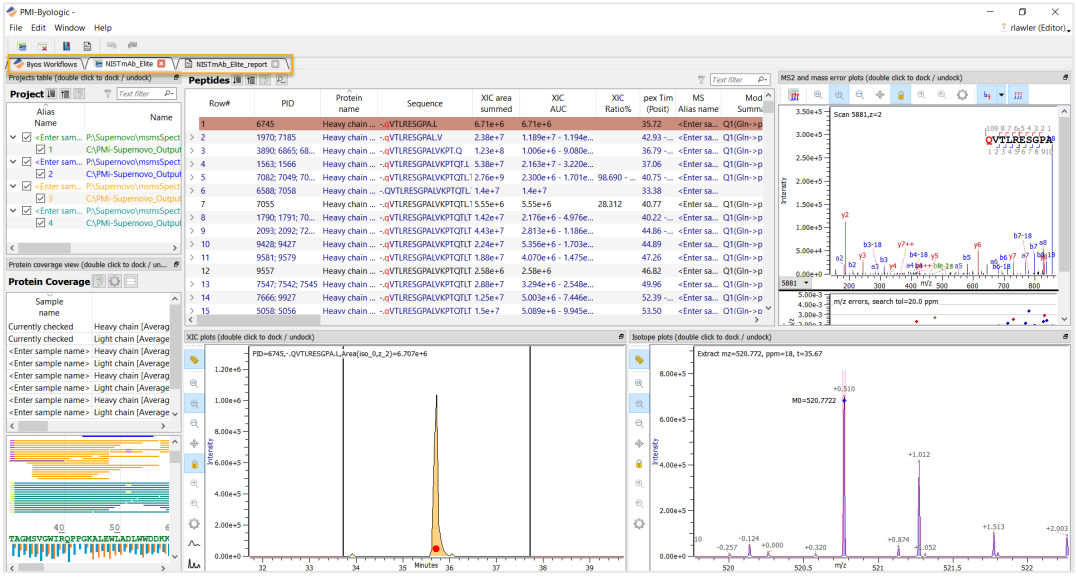


Figure 165: Byos opening tab, completed project, and associated report in separate tabs

Please note the analysis can take hours. Early in the analysis, a constant region depth of coverage is calculated. The number of distinct peptide sequences covering each amino acid residue in the constant regions of the heavy chain and the light chain is counted, and the average is calculated to give the average depth of coverage. The depth of coverage provides an estimate of how likely the full analysis will find the complete sequence accurately and provides feedback to the user on the sequence depth of their data set.

- Greater than 20 – good chance of success.
- Between 10 and 20 – moderate chance of success.

- Less than 10 – poor chance of success; recommend collecting more data and/or revising data collection procedure.

My Workflows

Saving a Customized Workflow or Portable Workflow

Once the user has reviewed the default parameters and made all desired changes as detailed in the "Customizing System Workflows to Create My Workflows" section, he or she is ready to save this workflow as one of the "My Workflows".

Save as Portable Workflow

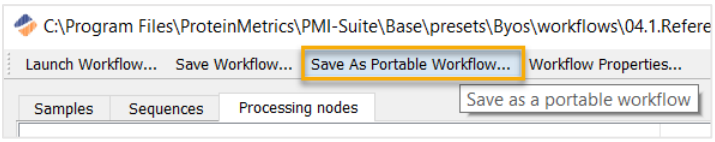


Figure 166: Save as Portable Workflow

The user can click **Save As Portable Workflow** to save the customized workflow file.

The user will be prompted to define the folder name. The user must save the portable workflow into a preferred folder location.

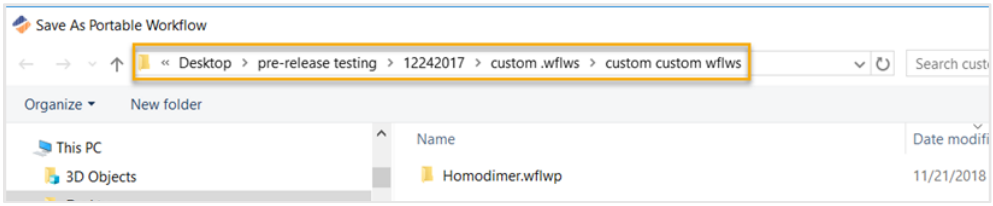


Figure 167: Select the folder location

NOTE: The portable workflow folder contains the layout, report template, and workflow comprising the Portable Workflow, as shown below. The customized content is contained within a folder.

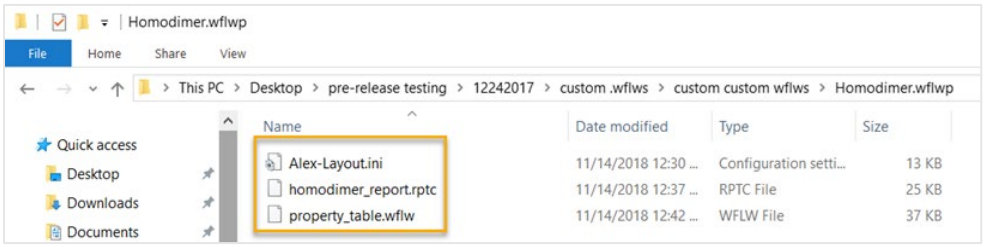


Figure 168: Content of portable workflow folder

Populate "My Workflows"

The user is now ready to populate the "My Workflows" section (called out as 1). The user will do this by directing Byos to the folder where their customized portable workflow file(s) are saved. This is done by clicking on the "…" square and selecting workflow folder to define the folder directory.

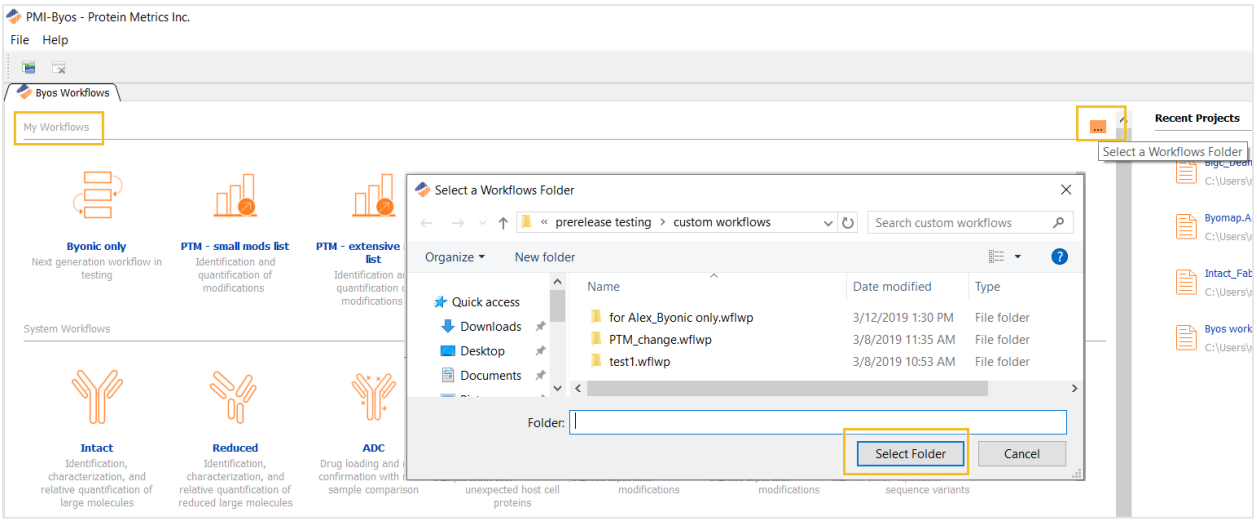


Figure 169: Select the My Workflows folder location

Create Project to Generate an Analysis and Report

Now that the user has saved customized portable workflows into their My Workflows section, they are ready to launch a Byos workflow to generate customized projects and reports. Please note these same steps could be followed using one of the default System Workflows as well. My Workflows simply provides the user with the option to customize.

The below example is for an Intact project; please note the steps to follow are exactly the same for every workflow.

1. Click **Create Project** to start processing and to generate the report.

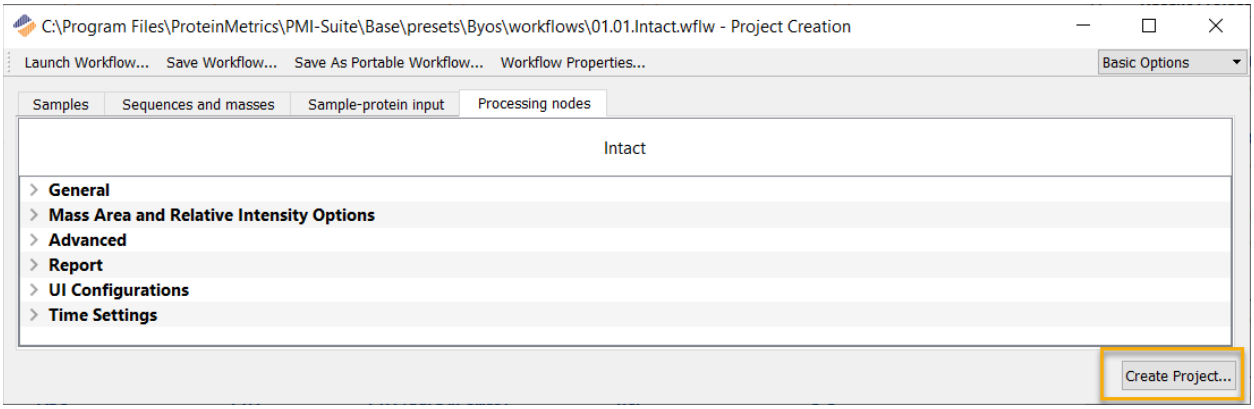


Figure 170: Click Create Project

2. The user will be prompted to name and save the project file. Please note Byos will save projects based upon the type of processing as a .ntms, .blgc file, or .bmap file.

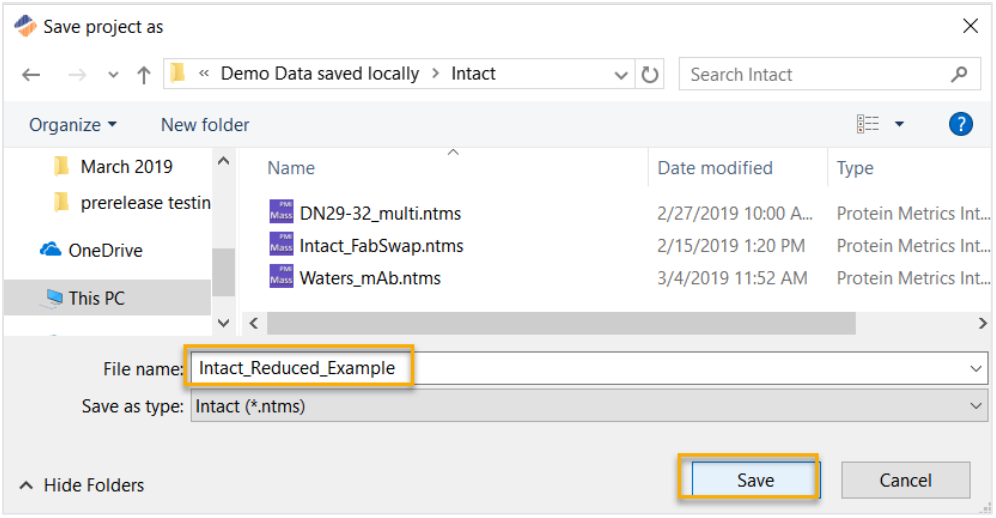


Figure 171: Name the project and save in a convenient location

Processing will commence. Once complete, the project and associated report will appear as additional tabs after the “Byos Workflows” tab, as in the example below:

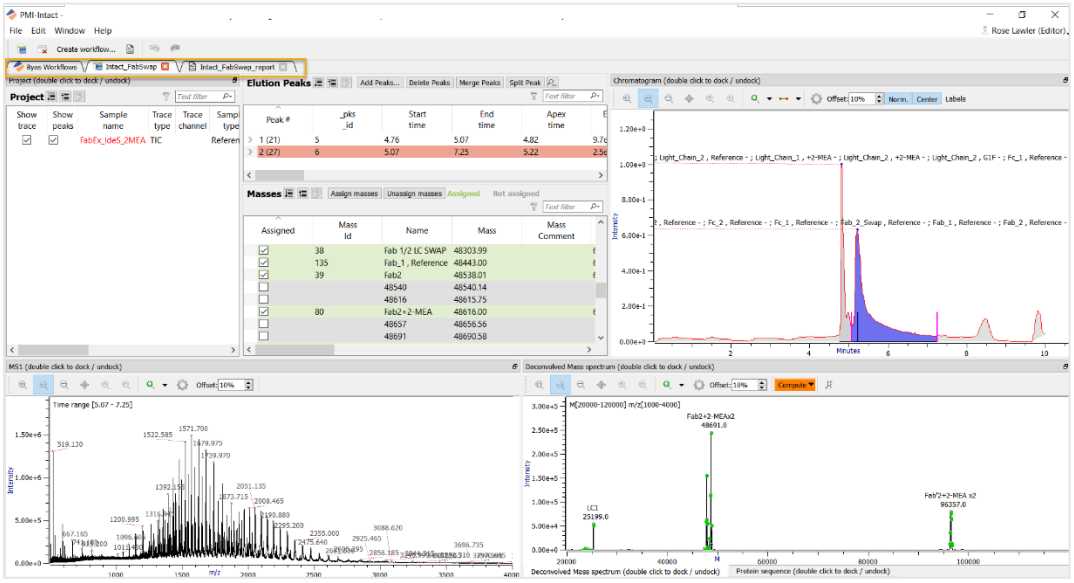


Figure 172: Byos opening tab, completed project, and associated report are in separate tabs.

Byos UNIFI API Integration

Sample results stored in the Waters UNIFI Application Programming Interface (API) can be loaded directly into Byos. The integration requires installation of the Waters UNIFI API and configuration of both UNIFI and Byos. The **Add UNIFI sample(s)** button will then appear in the **Samples** tab in Byos:

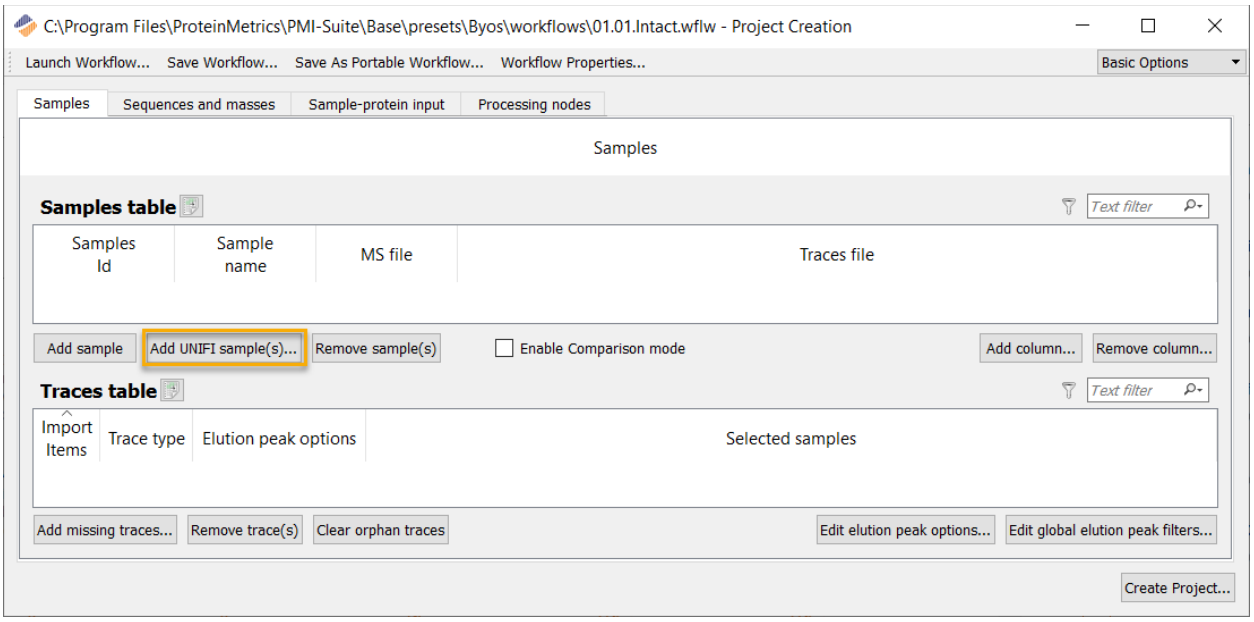


Figure 173: Byos Samples tab configured for UNIFI sample addition

This opens the **Select UNIFI sample** dialog where MS data stored in UNIFI can be browsed and selected for download to Byos:

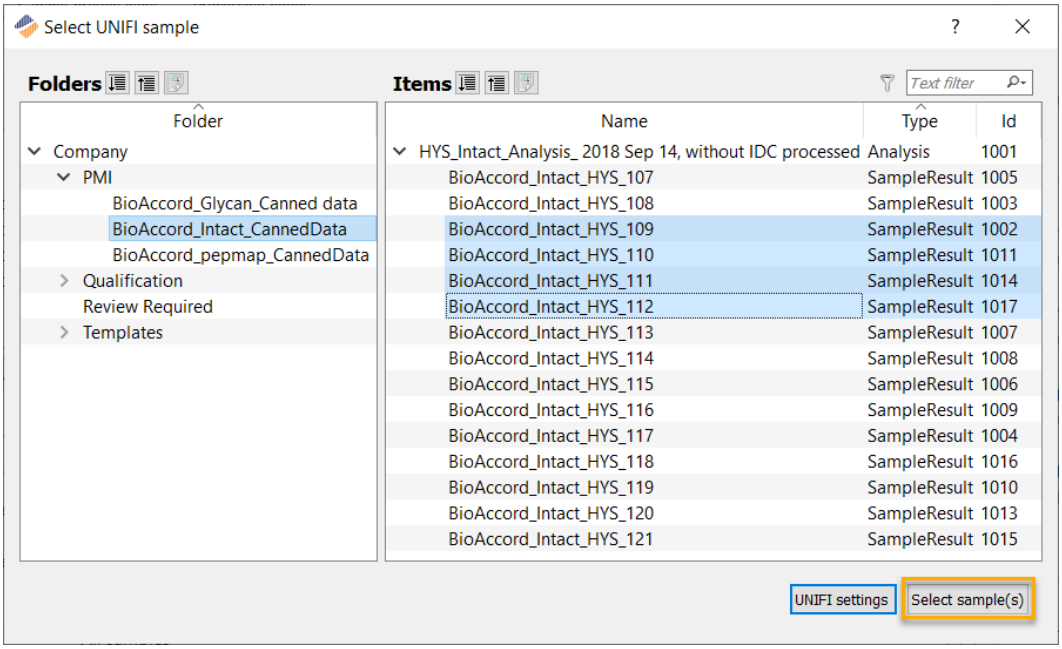


Figure 174: Select UNIFI sample dialog opened from within Byos

To learn how to access UNIFI from Byos, please refer to **PMI UNIFI Integration User Guide.docx**. Please start with the **PMI Byos UNIFI Integration Prerequisites Guide.docx**.

Customized Workflows

Contact support@proteinmetrics.com for information on a variety of existing and new workflow templates.

Up to date release and product information are listed on our website. Please visit:
https://www.proteinmetrics.com.