thermoscientific

Thermo

ProSightPD

User Guide

Version 4.2 for Proteome Discoverer 3.0

XCALI-98509 Revision A ● May 2022

Thermo Fisher
SCIENTIFIC

Revision A, May 2022

Software version: ProSightPD 4.2 for Proteome Discoverer 3.0

For Research Use Only. Not for use in diagnostic procedures.

Contents

Preface

This guide describes how to use the Thermo ProSightPD™ 4.2 nodes in the Proteome Discoverer™ 3.0 application to perform top-down proteomics data analysis.

Contents

- Access documentation
- Software installation and activation
- Download the software
- Activate the software license
- Special notices
- Contact us

Access documentation

The ProSightPD application includes complete documentation.

❖ **To view user documentation from the Thermo Fisher website**

1. Go to Life Sciences Mass Spectrometry Software Download and Licensing Portal https://thermo.flexnetoperations.com.

2. Log in to your account.

3. In the navigation pane, click the **Product Search** link.

4. Type **ProSightPD 4.2.**

5. In the Product Search Results, select the entry for *ProSightPD 4.2 User Guide* in the description.

   The related software and documentation appear.

6. In the Results column, select the *ProSightPD 4.2 User Guide*.

7. On the Product Download page, select the link under File Name.

8. Click **Download Selected Files**.

About this guide

The *ProSightPD User Guide* describes features specific to top-down analysis. For information about the Proteome Discoverer application, on which the ProSightPD application relies, refer to the *Proteome Discoverer User Guide*.

Software installation and activation

ProSightPD 4.2 is a standalone installer that includes the Proteome Discoverer application and ProSightPD 4.2 nodes. After installing ProSightPD, activate your ProSightPD license in the License Manager tool in the Proteome Discoverer application.

If you have not received your license, contact ThermoMSLicensing@thermofisher.com. Include your sales order number in the email.

Alternatively, you can check your available licenses (and view their activation codes) on the Life Sciences Mass Spectrometry Software Download and Licensing Portal at https://thermo.flexnetoperations.com.

Download the software

If you have a license, but do not have ProSightPD nodes installed, complete the following:

❖ **To download the ProSightPD software**

1. Go to Life Sciences Mass Spectrometry Software Download and Licensing Portal: https://thermo.flexnetoperations.com.

2. Log in to your account.

3. In the navigation pane, click the **Product Search** link.

4. Type **ProSightPD 4.2**.

5. In the Product Search Results, select the entry for *ProSightPD 4.2*.

6. Under File Name (rightside column), click the file name link.

After you download and install the software, activate the license with the License Management interface.

If you have issues with licensing, send an email to ThermoMSLicensing@thermofisher.com. Include a short description of the problem and the ProSightPD license key.

To activate the license, you must have an activation code and product ID number from Thermo Fisher Scientific. Before you transfer a license to another computer, deactivate the license. Refer to the *Proteome Discoverer User Guide* for more information.

Activate the software license

To activate your license, follow these topics as necessary:

1. Enter the product ID and the activation code

2. Complete license activation on a computer that is connected to the internet

Enter the product ID and the activation code

To activate your license, you need to have the product ID (XCALI-*XXXXX*) and the activation code. You can obtain these from the email that Thermo MS Licensing sends you within one week of ordering the software. The email has the subject line "Your Order Is Ready."

❖ **To enter the licensing information**

1. Open the Proteome Discoverer application.

2. Choose **Administration > Manage Licenses**.

   The Administration page opens.

3. Click **Activate**.

   The License Activation dialog box opens to the Activation Code view.

Figure 1.   License Activation dialog box opened to the Activation Code view

4. If you have not received your activation code, do the following:

a.   Check your Junk Email folder.

If the email is not in your Junk Email folder, log in to your account at the following URL: https://thermo.flexnetoperations.com.

In the left navigation pane, under Software & Services, click **Order History**. Then, in the list of ordered products, click the order number.

b.   If you cannot find your account, send an email message to Licensing at ThermoMSLicensing@thermofisher.com.

Provide the following information in the body of the message:

• Software application: ProSightPD

• Sales order number or purchase order number

• End user name

• End user email

5. In the License Activation dialog box (Figure 1), enter the following:

    - Your company name.

    - Your full name.

    - Your contact email address.

    - The product ID for the ProSightPD application.

    - The activation code. You can type or paste the activation code.

    Table 1. Product IDs for the ProSightPD application

Material Order No.	Product ID	Description
OPTON-31090	XCALI-65179	SW, PROSIGHTPD 4.2
OPTON-31091	XCALI-65198	SW, PROSIGHTPD 4.1 TO PROSIGHTPD 4.2
OPTON-31092	XCALI-65199	SW,PROSIGHTPD BEFORE 4.1 TO PROSIGHTPD 4.2
OPTON-31093	XCALI-65200	SW, PROSIGHTPD 4.2 UPG FRM PROSIGHTPC

6. Continue to the following topic:

    - Complete license activation on a computer that is connected to the internet

Complete license activation on a computer that is connected to the internet

Follow these instructions if your processing computer has an Internet connection.

❖ **To complete license activation on an online computer**

1. If you have not already entered the licensing information, enter it in the Activation Code view of the License Activation dialog box.

2. Click **Online Activation** to process the activation code.

3. To close the dialog box, click **OK**.

This completes the online license activation process.

For additional information about deactivating a license or using an offline computer, refer to the *Proteome Discoverer User Guide*.

Special notices

Make sure you follow the special notices presented in this guide. Special notices appear in boxes; those concerning safety or possible system damage also have corresponding caution symbols.

IMPORTANT Highlights information necessary to prevent damage to software, loss of data, or invalid test results; or might contain information that is critical for optimal performance of the system.

Note Highlights information of general interest.

Tip Highlights helpful information that can make a task easier.

Contact us

Contact	Email	Telephone	QR Code
ProSightPD Support	prosightpdsupport@thermofisher.com		
Proteinaceous	info@proteinaceous.net		
U.S. Customer Service and Sales	us.customer-support.analyze@thermofisher.com	(U.S.) 1 (800) 532-4752	

Overview

This chapter introduces the ProSightPD application and provides a brief discussion of top-down proteomics.

Contents

- Introduction to ProSightPD
- Working with example data
- ProSightPD database search algorithms
- Label-free quantitation and sliding window deconvolution
- ProSightPD task overview
- New features in this version

Introduction to ProSightPD

The ProSightPD application is a set of nodes in the Proteome Discover platform. These nodes are designed to analyze top-down mass spectrometry data.

The ProSightPD nodes identify, quantify, and characterize proteoforms. The ProSightPD application is fully integrated into the Proteome Discoverer framework and utilizes a common user interface and results architecture. The primary difference between the ProSightPD nodes and the Proteome Discoverer nodes is the focus on proteoform identification and not peptide identification. Many of the notable differences such as database structure, tables in the results, and graphical result views stem from this fundamental practice.

The ProSightPD application supports several top-down workflows including:

- High-throughput LC/MS proteoform discovery
- Large proteoform discovery using proton transfer charge reduction (PTCR)
- Label-free quantitation (LFQ) of proteoforms
- Targeted proteoform searches

- Analysis of infused samples and manually collected data (not using a predefined MS method)

- Analysis of isotopically and non-isotopically resolved precursors

Introduction to top-down proteomics

The application works with the mass values inferred from mass spectral data from middle-down/bottom-up and top-down proteomics MS/MS experiments.

Figure 2 compares bottom-up and top-down methods.

Figure 2.   Comparing top-down and bottom-up proteomics
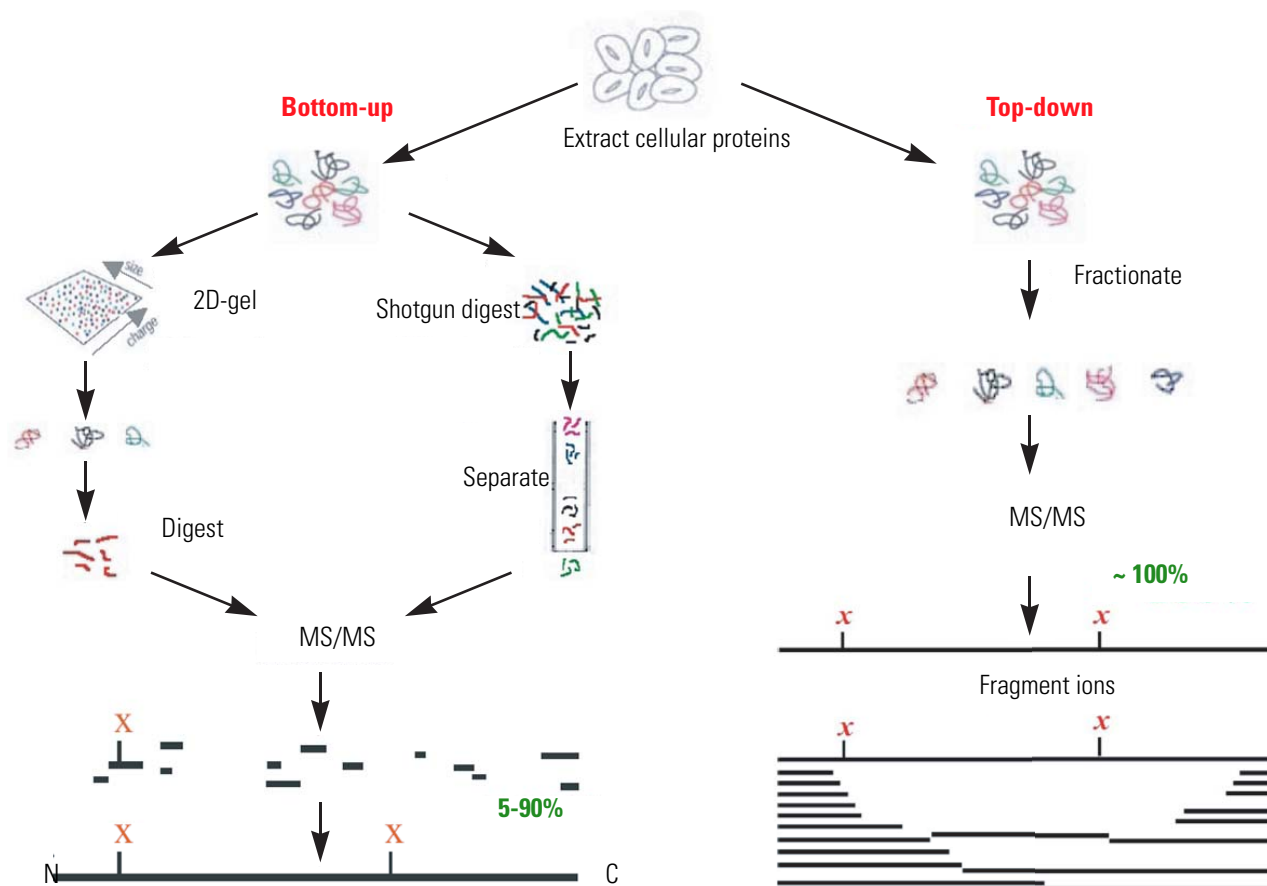
In top-down proteomics, proteins are introduced into a mass analyzer, where they are subjected to gas-phase fragmentation. The right side of Figure 2 illustrates this process.

Top-down proteomics is a technique for protein identification and characterization. Combining top-down proteomics searches with the shotgun annotation process provides a method for rapid and accurate protein definition.

Top-down proteomics can provide thorough sequence coverage of proteoforms, supporting characterization of complex combinations of PTMs.

Separation and ionization of intact proteins present many challenges beyond the scope of this guide. The bioinformatics of top-down proteomics, though, has some challenges that the application addresses.

• ProSightPD automates the determination of neutral masses from both precursor and fragmentation spectra. The resultant masses are then used for protein identification and characterization.

• Precursor masses may not originate from full-length, unmodified proteins. Therefore, multiple search strategies may be necessary to identify and characterize all proteoforms in a sample. ProSightPD provides several search modes to identify different proteoforms. Search modes such as the subsequence search can be used to find truncated forms while absolute mass searches can be used to identify highly modified proteoforms. More in-depth descriptions are provided in the ProSightPD database search algorithms section.

Working with example data

The ProSightPD application installs with a database and analysis workflow template for the Thermo Scientific™ Pierce™ Intact Protein Standard Mix. You can download an example of the Pierce Intact Protein Standard Mix data file at the Life Sciences Mass Spectrometry Software Download and Licensing Portal (https://thermo.flexnetoperations.com).

The Pierce Intact Protein Standard Mix contains six recombinant proteins. For additional information, refer to
https://www.thermofisher.com/order/catalog/product/A33527#/A33527.
This data file was collected on a Thermo Fisher Scientific Fusion Lumos Tribrid Mass Spectrometer with high resolution precursor and HCD fragment ions scans.

To become familiar with the ProSightPD data analysis process using example data, complete the following:

1. Download the Pierce_Intact_Protein_Standard_Top_Down_Example.raw data file at the Life Sciences Mass Spectrometry Software Download and Licensing Portal (https://thermo.flexnetoperations.com).

2. Open the Proteome Discoverer application.

3. On the Start page, click **New Study/Analysis**.

   The New Study and Analysis dialog box opens.

4. Type the Study Name.

5. For the Study Root Directory, click the browse icon and locate the path.

6. Click **OK**.

   The study opens.

7. Click **Open Analysis Template** and locate the PSPD HiHi TopDownStandard Analysis Template (C:\Users\Public\Documents\Thermo\Proteome Discoverer 3.0\Common Templates\ProSightPD 4.2\Analysis Templates).

8. Click **Open**.

9. Click the **Workflows** tab on the Study page.

10. In the Processing Step area of the Analysis pane, click **Edit**.

11. In the Workflow Tree, select the Annotated Proteoform Search node.

12. In the Parameters of the 'ProSightPD 4.2 Annotated Proteoform Search' area on the left side of the Workflow Tree, click the Input Database dropdown arrow and select the **PierceIntactProteinStandardMix** database file (it is the only database available for newly installed software).

13. Repeat step 12 for the second Annotated Proteoform search node and the Subsequence Search node and change the Input Database to the **PierceIntactProteinStandardMix** database file.

14. Click the **Input Files** tab on the Study page.

15. Click **Add Files**.

    The Add Files dialog box opens.

16. Locate the Pierce_Intact_Protein_Standard_Top_Down_Example.raw data file that you downloaded at the beginning of this process, then click **Open**.

    The file is now listed on the Input Files page.

17. Drag and drop the file into the Processing Step area of the Analysis pane on the right side.

    The Run button now activates and turns green.

18. Click **Run**.

    Once you click Run, the Administration page automatically opens where you can monitor the progress of your analysis. For additional information about viewing your results, see Chapter 4, "Working with results."

ProSightPD database search algorithms

The following topics describe the search algorithms specific to the ProSightPD application. Each search mode overcomes different issues of protein identification and characterization:

- Annotated proteoform search

- Subsequence search

- Delta M Mode searches

Annotated proteoform search

The annotated proteoform search is the standard search mode for top-down proteomics in ProSightPD.

For each precursor mass, a subset of proteoforms within the user-selected mass tolerance are selected from the database. Fragment ions from spectral fragmentation data that has been associated with the precursor mass are then matched against the theoretical fragment ions of the selected database proteoforms. The application determines the number of observed fragment ions matching the fragment tolerance and uses this value to score the identification (Figure 3).

The application calculates scores for each probable MS/MS to candidate proteoform sequence match (PrSM). An observed ion matches a theoretical ion when the two masses are within a user-defined fragment ion mass tolerance. The probability of the observed number of fragment ions matching by chance is then determined and reported as a P-Score.

Figure 3.   Annotated proteoforms searches

Intact precursor mass

Shotgun-annotated proteome database

Top-down fragments

Proteins within tolerance

Query proteome subset

Proteome subset

Identified/characterized protein
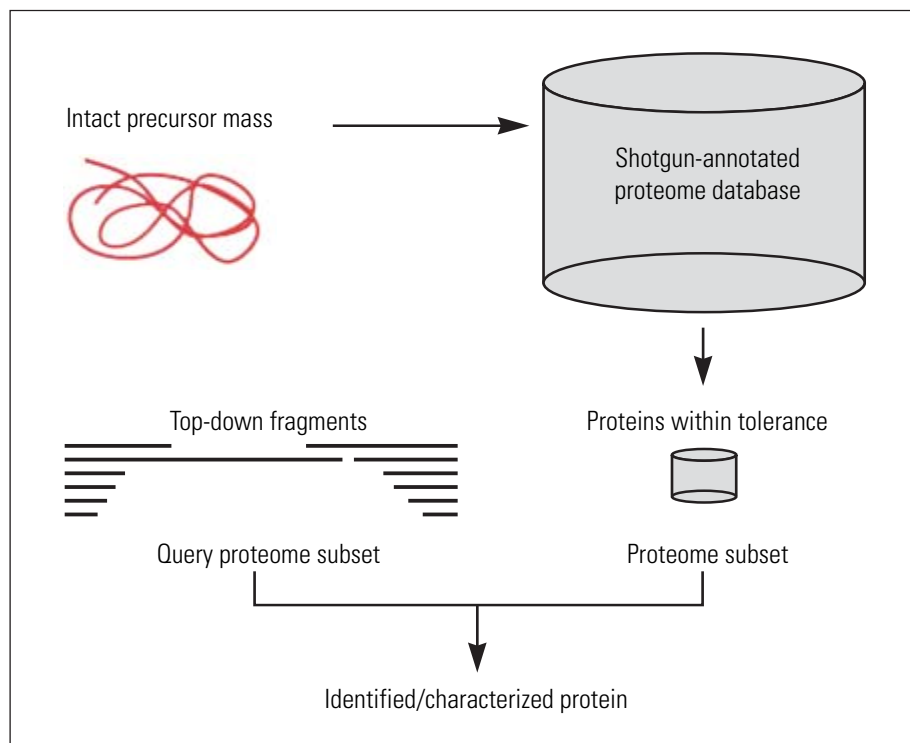
Although the application queries each proteoform with a theoretical precursor mass in the window, it displays only those proteoforms that meet the user-defined filtering of search results. These filters include the minimum number of matching fragments and false discovery rate (FDR) (Figure 4).

Figure 4. Annotated proteoforms search process
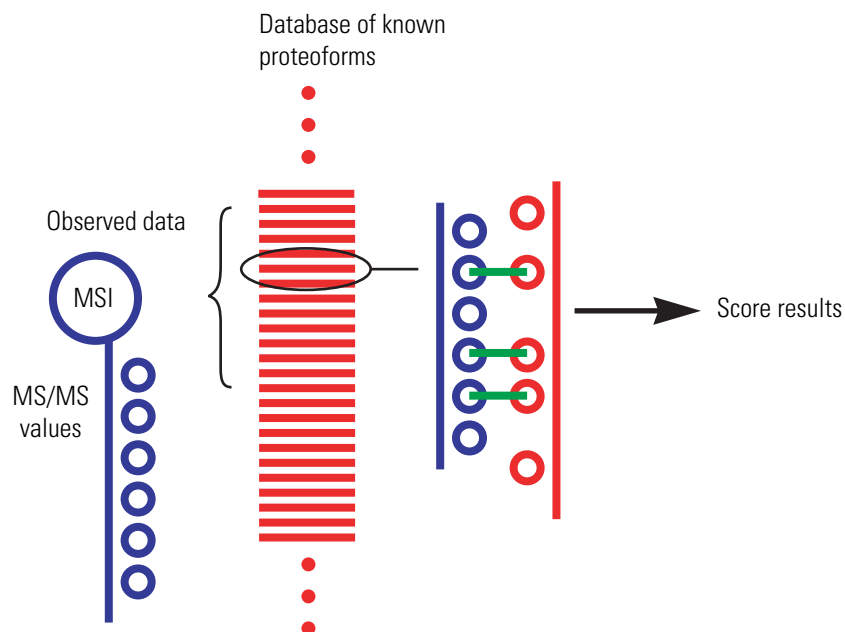
The following strategies may be helpful when running annotated proteoform searches:

- Use a large precursor mass window search when identifying an unknown proteoform. If there are several unknown modifications (modifications that are not in the search), a large window search often will still identify the protein. A 100 Da window is a good starting place but a 1000 Da window or larger may be necessary if a very large mass shift is present on the proteoform. Further proteoform interrogation to characterize the proteoform can then be performed once the identity of the protein is known.

  A large number of ions matching one terminal in a protein is evidence of a protein's identity. You can use the ProSight Lite and the TDValidator Lite applications for further characterization. For more information on these tools, see Use ProSight Lite and Use TDValidator Lite.

- Use Delta M (Δm) Mode to locate unknown modifications near either terminus. If the large precursor mass window annotated proteoform search fails to identify a proteoform, consider running another large precursor mass window annotated proteoform search with Delta M Mode enabled.

  Activating Delta M Mode increases the likelihood that the search will identify proteins with unknown modifications. However, this mode takes approximately two times longer than the corresponding annotated proteoform search. For more information, see Delta M Mode searches.

Subsequence search

A ProSightPD subsequence search matches MS/MS data against all subsequences of all proteoforms of proteins in a database. The subsequence search is similar to a bottom-up no enzyme search. The subsequence search produces many more possible candidates than the annotated proteoform search and will therefore take longer than an annotated proteoform search when using the same precursor window size. It looks at every possible subsequence of every base proteoform (unless mentioned otherwise) in the database and attempts to identify any subsequence that matches the observed intact ion mass within a tolerance. For each subsequence matching the intact ion mass, the subsequence search performs an absolute mass search and reports any subsequence that matches the observed intact ion mass and is able to generate the observed fragment ion pattern.

In a typical top-down experiment, not all of the polypeptides identified are intact proteins. A subsequence search identifies those proteins that are a product of biological degradation and cannot be logically predicted. This search compares the observed precursor mass to all possible entries of a particular database within a defined tolerance, for example, less than 10 ppm. The application theoretically fragments those entries that fall within the defined tolerance and compares the observed fragment ions.

> **Note** Use a subsequence search if an absolute mass analysis fails to identify a protein.

A subsequence search is a two-step process that is repeated for each base protein sequence in the proteome database:

1. Identify a candidate entry matching an observed precursor mass.

2. Calculate all possible theoretical fragment ions for the candidate entry, and then compare the theoretical fragment ion masses to the observed fragment ion masses.

Figure 5 shows the process involved in a subsequence search.

Figure 5. Subsequence searches
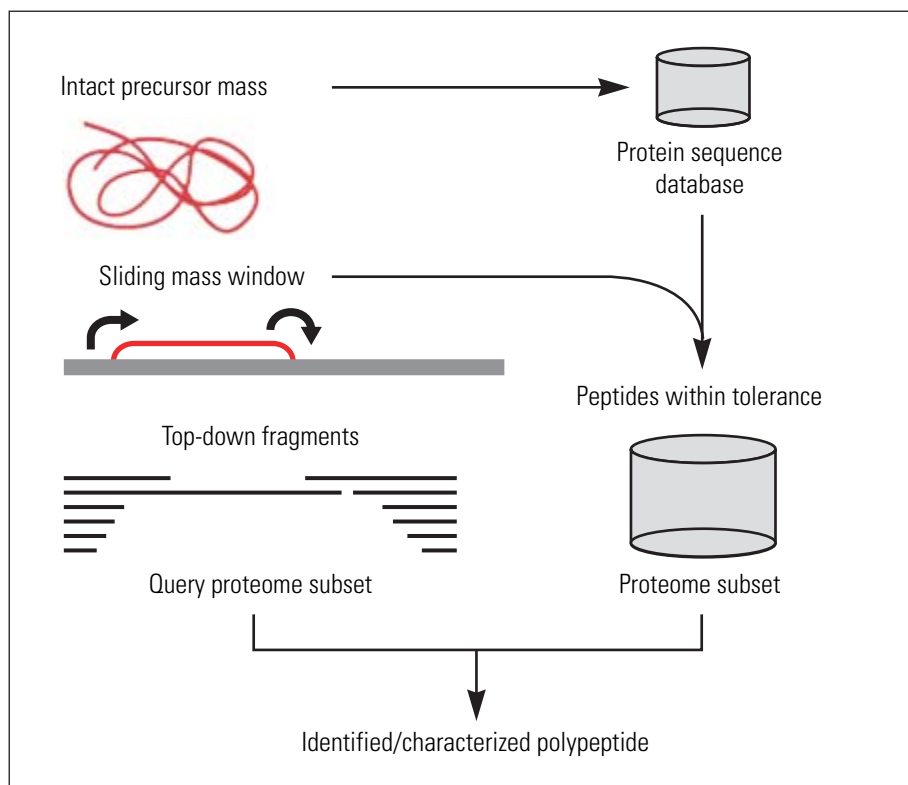
In a subsequence search, the precursor search tolerance is an estimate of measurement error on the observed precursor mass. The value is usually small compared to an absolute mass precursor search window.

Delta M Mode searches

Delta M (Δm) Mode is a feature for identifying proteoforms containing unknown mass shifts, for example, PTMs. The delta is the difference between the observed precursor mass and the theoretical precursor mass. When you perform a search in Delta M Mode, the ProSightPD software concurrently performs three queries per sequence to compare the following:

- The theoretical fragment ion masses of the protein sequence to the observed fragment ion list as usual

- The theoretical fragment ion masses derived from the sequence and the Delta M applied N terminal to the observed fragment ion mass list

- The theoretical fragment ion masses derived from the sequence and the Delta M applied C terminal to the observed fragment mass list

A Delta M Mode search takes approximately two times longer than the same search without Delta M Mode.

By carefully observing the pattern of fragments with and without Delta M (Δm) Mode, you can locate the source of the mass shift. For example, a particular result returns with the two smallest N-terminal fragments matching without the mass shift. If all other matching N-terminal fragments contain the mass shift, the unknown mass can be localized on an amino acid between the second and third N-terminal fragments (Figure 6).

When Delta M Mode is enabled, fragments found using the Delta M Mode display on the PrSMs page, along with two additional columns: Delta M Residue and Delta M Fragment(s), where the delta mass and the possible residue where the delta mass was localized. Delta masses also display in the ProSightPD fragment map with an orange square around the potential residue where the delta mass was localized.

Figure 6. Schematic of Δm mode
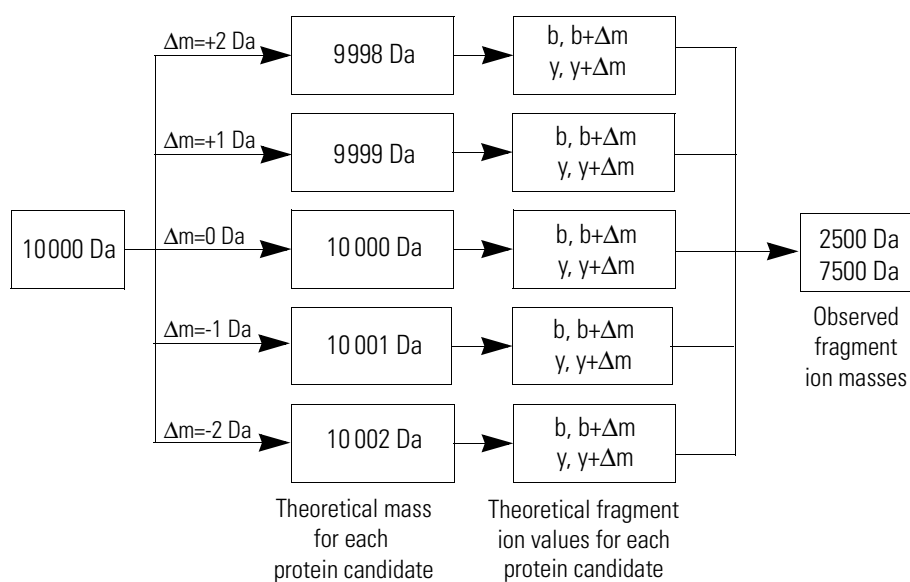
Label-free quantitation and sliding window deconvolution

The ProSightPD application performs label-free quantitation (LFQ) using the sliding window deconvolution algorithm.

In an LFQ experiment, the proteoforms are quantified across their elution profile. In ProSightPD, these quantitation results are called Feature Groups. The Feature Groups are determined by mass deconvolution in the Feature Detector nodes using Xtract for high resolution, isotopically-resolved data and kDecon for isotopically-unresolved data. A Feature Group is the decharged and deisotoped signal measured over time in an LC/MS experiment. Feature Groups are measured in the Feature detector nodes using the sliding window algorithm in conjunction with the Xtract (for high resolution, isotopically-resolved data) or kDecon (for low or medium resolution, isotopically unresolved data) algorithms for deconvolution.

The sliding window algorithm averages spectra over a succession of windows in retention time, deconvolves each average spectrum, and then merges similar masses from consecutive deconvolutions to form feature groups.

Figure 7 illustrates the concept of sliding windows.

> **Note** FWHM refers to full-width half-max.

Figure 7.   Sliding window overview
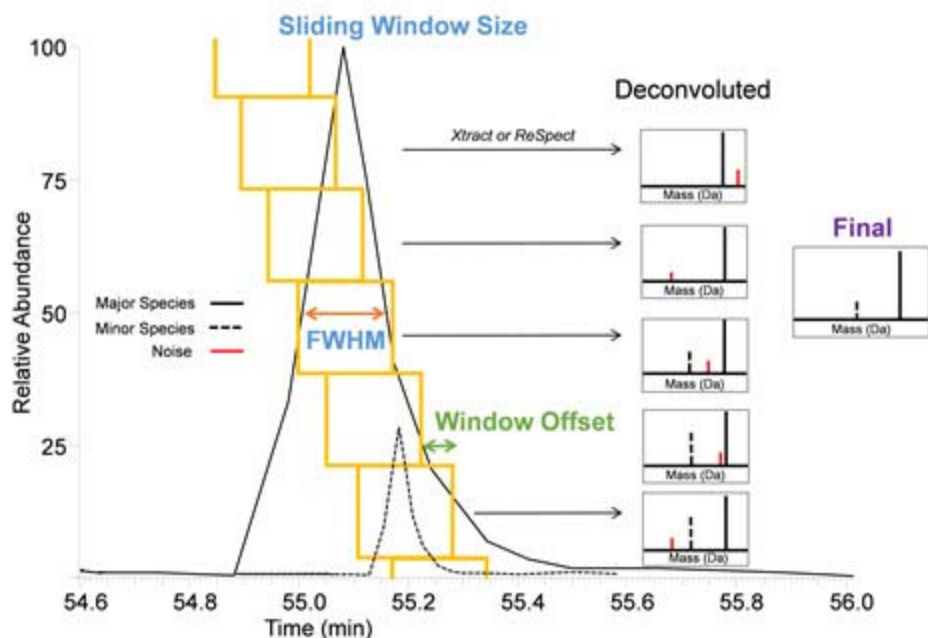
The sliding window algorithm benefits top-down LFQ in several ways:

- Reduces the number of false positives due to noisy data

- Improves sensitivity using signal averaging

- Identifies co-eluting species better

- Defines elution profiles for quantitation

The ProSightPD Hi Res. Feature Detector node and the ProSightPD Med. Res. Feature Detector node contain the parameters controlling the sliding window algorithm.

After the detector node determines the feature groups for data files, the node maps the feature groups to individual PrSMs and connects feature groups to quantitation traces. The node groups the features that are similar in mass and retention time between files. A group of feature groups across all files is a *consensus feature group*.

The Feature Mapper node maps the consensus feature groups to proteoforms based on a mass and retention time threshold.

The ProSightPD Quantifier node produces abundance ratios and statistics using the measured abundances and study factors you applied.

ProSightPD task overview

Follow the instructions in Figure 8 to specify preferences and perform the search.

Figure 8. The ProSightPD task flow

Start new study/analysis

Name the analysis.
Choose the location for results.
(Optional) Add the .raw files.
(Optional) Add the template workflow.
(Optional) Set the grouping and quantitation parameters.

↓

Create or select the database.

↓

Complete the workflow

Choose the workflow template:

* LFQ

* Single Protein Infusion

* High Throughput

* PTCR

Select databases and confirm parameters.

↓

(Optional) Customize the template.

↓

Run the analysis. (The application executes the processing workflow and the consensus workflow.)

↓

View the results.
(Optional) Export the results.

↓

(Optional) Save the analysis as a template.
(Optional) Create a Results View template.

New features in this version

The new and enhanced features in this release include:

New features

- **Proteoform specific database creation** - Isoforms can now be expanded into a collection of all possible proteoforms. These proteoforms can be further edited to create a fully customized database.

- **Improved searching to handle off by one deconvolution errors**- Precursors that are +/- 1 Da or 2 Da away from the deconvoluted masses are added to account for potential deconvolution artifacts.

- **ProForma annotated results** - Proteoform results now include a ProForma column.

- **Build TD database from exported PD FASTA** - You can import and edit databases directly in the Proteome Discoverer application using the FASTA tools utility and export in a top-down format, which is fully compatible with ProSightPD Database Manager.

- **Support for UniMod, RESID, and ProForma annotation in Single Proteoform Search** - You can now use popular modification databases to annotate single proteoform searches.

Enhanced features

- **Truncation Sites** - Results from Subsequence search indicate sites of truncation and the local amino acid motif at the truncation sites.

- **Custom mass modifications can be specified by molecular formula** - You can enter the molecular formula and the ProSight Annotator applies the correct mass.

- **Quan experiments can be restricted by time or scan ranges** - A new parameter in the Feature Detector nodes allows you to specify a scan or time range to quantify.

- **Max mods per proteoform at search time** - You can specify the number of PTMs per proteoform to avoid long search times or consider non-viable candidates.

- **Range features** - The range features dialog has been enhanced with its own user interface.

- **ProSight Annotator** - Enhancements have been implemented to support proteoform specific searching.

- **Delta M mode redundant hits removed** - Results which add and remove the same modification are ignored.

Creating and maintaining databases

This chapter describes working with top-down databases in the ProSightPD application.

Contents

Top-down databases in ProSightPD

ProSightPD Database Manager gives users flexibility when creating a database to search their top-down data. The Database Manager allows users to easily build databases from any UniProt formatted .fasta, or .xml file, or Proteome Discoverer formatted .mzIdentML file. Once these files are imported users can add or remove sequences, edit sequences and precisely annotate their database with proteoform level control using ProSight Annotator (previously Protein Annotator.)

You can download protein databases and full proteome databases from:

- UniProt—for a comprehensive source https://www.uniprot.org.

  Select an XML file.

- Proteinaceous— for selected, common proteome databases in XML and ProSight DataBase (.psdb) formats https://proteinaceous.net/databases.

The ProSightPD application creates and stores databases as .psdb files. The Database Manager can convert the XML and FASTA files to PSDB format.

Create a database in ProSightPD

Proteinaceous offers several ready-to-use databases for common organisms, which are in XML format that you can import through the database management tool in the ProSightPD application.

Download a ready-to-use database

> **Tip** For select proteins from a specific genome or proteins from different genomes, download .xml files from UniProt. For complete proteomes from commonly used organisms, download from the Proteinaceous website.

❖ **Download and configure an XML format database**

1. Go to one of the following websites:

   • https://www.uniprot.org/

   • https://www.proteinaceous.net/

2. Locate and download the XML file.

3. Save the file.

4. In the Proteome Discoverer application, click the ProSightPD Database Manager icon, , in the toolbar.

   The ProSightPD Database Manager dialog box opens.

   > **Tip** You can also access the Database Manager from the Help menu.

5. In the dialog box, do the following:

   a. Click **Create ProSightPD Database**.

   b. Click **Browse**, and locate the saved XML file.

   c. Select the file.

   The database file appears in the dialog box (Figure 9). The application displays the number of proteins found in the imported file.

Figure 9. ProSightPD Database Manager dialog box
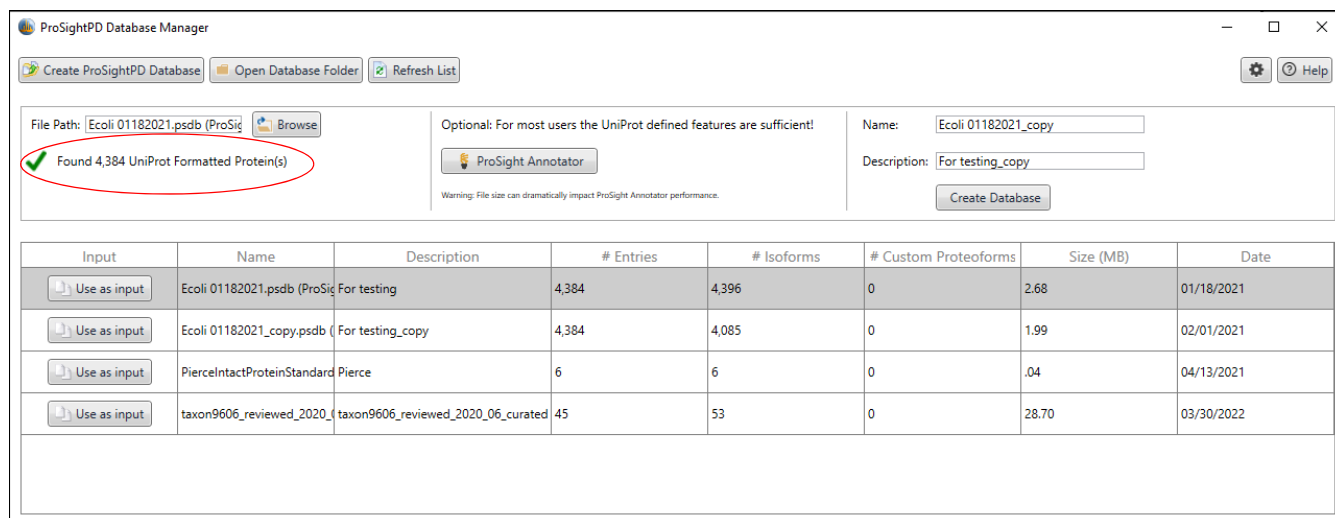
If the imported database is correctly formatted, a green check and the number of correctly formatted proteins appear. If the file is incorrectly formatted, a red X appears.

If the number of entries shown is fewer than you expected, the lower number might be due to the Max Isoform Mass setting being too low. For information on changing that setting, see Change the isoform mass threshold.

Other reasons that the number of isoforms might be unexpectedly low:

- A sequence contains an unrecognized amino acid (B,Y,Z, and so forth.) or a wildcard.

- The entry header is incorrectly formatted (that is, not UniProt format).

- Some of the entries were duplicated.

The Database Manager removes incorrectly formatted and duplicate entries.

❖ **(Optional) Edit isoforms in ProSight Annotator**

Figure 10. ProSight Annotator view

In the ProSight Annotator dialog box, you can do the following:

- Add or remove isoforms from a database.

- Add or remove modifications from an isoform.

1. To edit or annotate the database, click **ProSight Annotator** from the ProSightPD Database Manager.

2. To modify an isoform:

    a. Select the isoform.

    b. Select the residue to modify.

c. Select the desired modification in the Point Features pane. You can lock a feature by selecting the lock icon, 🔓 , next to the modification.

> **Note** After a modification is selected and locked, you cannot add additional modifications to the selected residue. The selected modification will be on all search candidates.

3. To add a disulfide bond or endogenous cleavage:

a. Click the plus icon, ⊞ , in the Range Features pane.

   The Add Range Features dialog box opens.

b. In the Add Range Feature dialog box, select the UniProt Feature Type and type the Modification Start Point and End Point.

c. Click the **Save and Exit** icon in the lower right corner.

4. To add a mutation or cSNP (single-nucleotide polymorphism), select the residue to mutate.

a. Select the **cSNPs** tab under Point Features.

b. Select the mutation.

c. (Optional) In the protein sequence pane, you can add one or more modifications in the Point Features pane.

> **Tip** To remove a modification, hold down CTRL and select the modification. To select multiple point features, hold down the CTRL key.

5. To add Custom Mass Modifications, select the residue to modify.

Figure 11. Custom Mass Modifications

a. Choose the **Custom** tab in the Point Features pane.

b. Type the Mono Mass in Daltons (Da) or the known Chemical Formula.

c. Click the green plus icon, ⊞, to add the modification to the modifications list and apply the modification to the residue. If there is more than one Custom Mass Modification listed, you can select multiple modifications to add to each residue.

> **Note** Unlike named PTMs, the addition of multiple Custom Mass Modifications will result in orange colored squares.

6. Click the **Save and Exit** icon in the lower right corner.

The downloaded database now appears in the list of searchable databases (Figure 12).

Figure 12. The ProSightPD Database Manager

ProSightPD Database Manager

Input	Name	▾	Description	# Entries	# Isoforms	Size (MB)	Date
Use as input	Yeast.psdb		yeast	6,049	6,077	5.78	06/12/2020
Use as input	Xenograft_IP.psdb		IP	990	1,362	1.38	10/12/2020
Use as input	UpdatedMouse08182020.ps		does have doub	13,027	17,544	18.25	08/18/2020
Use as input	TopDown Standard.psdb		TopDown Stand	5	5	.04	06/12/2020

❖ **(Optional) Expand Isoforms to Proteoforms**

The ProSight Annotator (formerly the Protein Annotator) now supports the ability to expand database isoforms into individual proteoforms. Once expanded, you can add, remove, and modify individual proteoforms for precise control of their search space.

In order to expand and curate proteoforms, complete the following steps:

1. Open an existing database file by clicking the **Browse** icon in ProSightPD Database Manager.

   Note Thermo Fisher Scientific recommends creating a database from a UniProt formatted .xml file as it contains additional annotated information not found in a .fasta file.

   Optionally, if you create a new database, you can expand the isoforms to proteoforms at that time.

2. Click **ProSight Annotator** (Figure 13).

   The ProSight Annotator dialog box opens.

   Figure 13. ProSight Annotator button

   Optional: For most users the UniProt defined features are sufficient!

   ProSight Annotator

   Warning: File size can dramatically impact ProSight Annotator performance.
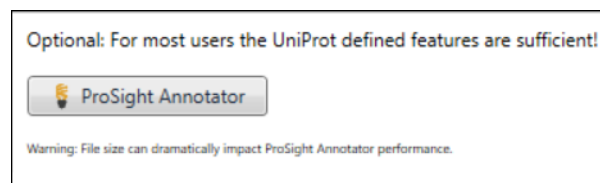
3. Before expanding, edit the isoform to include all proteoforms of interest.

4. Double-click the "Expand Isoform" icon, 🔀, in the Modify column (Figure 14).

   The Expand Isoform to Proteoforms dialog box opens (Figure 15). Any modifications added here are included once the isoform is expanded into proteoforms.

Figure 14. ProSight Annotator expand icon

Expand
Isoform
icon
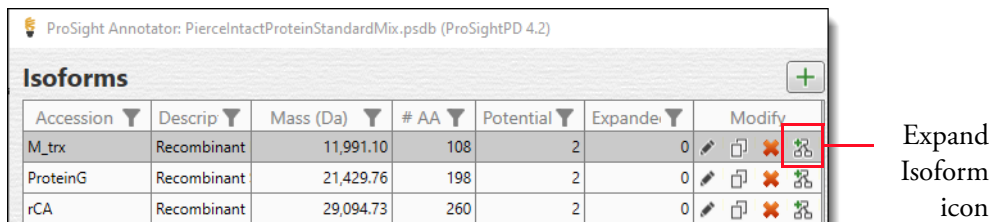
Figure 15. Expand Isoform to Proteoforms dialog box
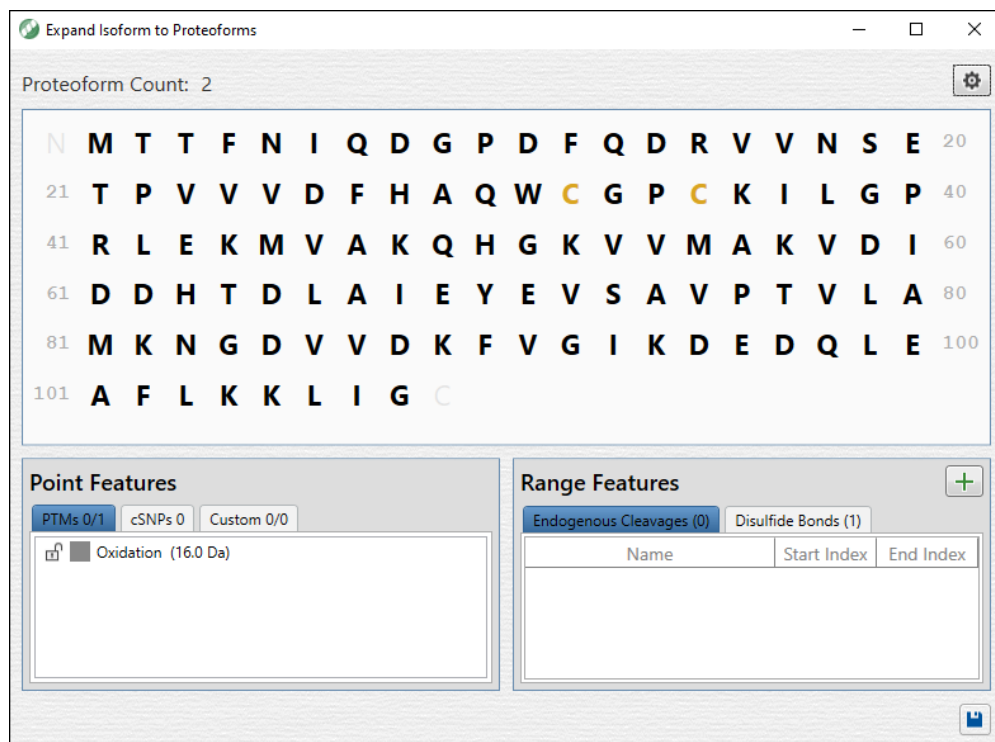
5. Proteoform expansion may be refined, if needed. Click the gear icon, [⚙], in the upper right corner of the Expand Isoform to Proteoform dialog box (Figure 15).

The Proteoform Expansion Parameters dialog box opens (Figure 16). Refer to Table 2 for information on expansion parameters

Figure 16. Proteoform Expansion Parameters dialog box
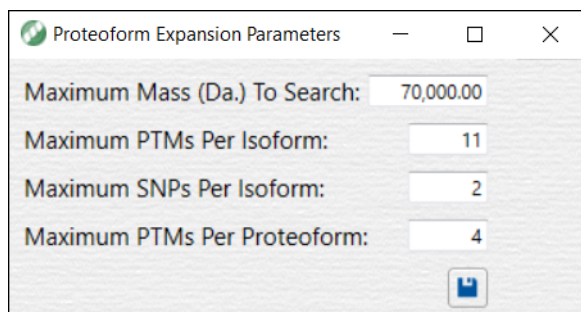
> **Note** Expanding a large number of proteoforms >1000 can be slow to complete and may result in reduced performance while editing the proteoform list.

Table 2. Parameters of the Proteoform Expansion Parameters dialog box

| Parameter	Description
Maximum Mass (Da.) To Search	Proteoforms larger than the default setting are not included in the database. The default is set to 70,000. Instrument performance should be taken into account when setting this parameter. It is unlikely to identify large numbers of proteoforms greater than 70 kDa.
Maximum PTMs Per Isoform	This parameter limits the number of PTMs expanded into proteoforms. Including a large number of PTMs increases the number of proteoforms exponentially resulting in long search times and large database files. This parameter should reflect experimental conditions. For example, in a targeted experiment with few sequences in the database, a larger number of PTMs per isoform can be considered while retaining good performance. However, for complex samples containing hundreds of sequences, a moderate number such as 11 ensures acceptable performance and search time. In the case where specific proteoforms must be included, Thermo Fisher Scientific recommends using the default expansion parameters and adding a specific proteoform after expansion.
Maximum SNPs Per Isoform	It is uncommon to find a sequence with many mutations, so the default number is set to 2. However, if a sequence of interest is known to have many mutations, adjust accordingly to reflect all possible proteoforms in the database.
Maximum PTMs Per Proteoform	This parameter limits the total number of possible modifications on a single proteoform. The default value is 4. However, there are certain cases such as histones where it is possible to encounter heavily modified proteoforms. As the number of max PTMs increase per proteoform, the size of the database increases as well. It is unlikely to identify a proteoform with more than 10 modifications on a single proteoform.

6. Once you set all modifications and Proteoform Expansion Parameters, click the **Save and Exit** icon in the lower right corner of the ProSight Annotator.

   A list of proteoforms displays. This list can be modified. Highlight the proteoform to delete, copy, or modify by adding Point Features or Range Features.

> **Note** If you want to keep a proteoform and add a modified form of the proteoform, you must copy that proteoform first.

7. After you have curated the database with proteoforms, click the **Save and Exit** icon in the lower right corner.

    The database now contains isoform and proteoform level information.

8. Click **Create Database** in the ProSightPD Database Manager just as you would when creating a database without proteoform level annotation.

9. In order to take advantage of a proteoform curated database, you must set the Candidate Source parameters to include proteoforms. See Table 3 for a list of parameters. After creating the database, select the database from the dropdown menu in the search node.

    > **IMPORTANT** There are new search node options that impact how your database is searched.

    The Candidate Source parameters determine how your database is searched.

Table 3. Candidate Source parameters

Parameter	Description
Isoforms	This parameter only searches isoforms.
Proteoforms	This parameter only searches proteoforms that you expanded.
Both-Exclude Isoforms with Proteoforms	This parameter searches both isoforms and proteoforms, but excludes isoforms that have expanded proteoforms. This search ignores proteoforms from expanded isoforms that are not explicitly added by the user (for example, for all expanded isoforms, only user-curated proteoforms are searched. Unexpanded isoforms are searched as usual).
Both-Include Isoforms with Proteoforms	This parameter searches both isoforms and proteoforms and includes searching for isoforms that have expanded proteoforms. This search includes proteoforms from expanded isoforms that are not explicitly added by the user. This search includes all user-expanded proteoforms and any proteoforms what would normally be searched using an isoform search. (**Note** Max PTM and Max SNP parameters dictate how large the isoform search space is.
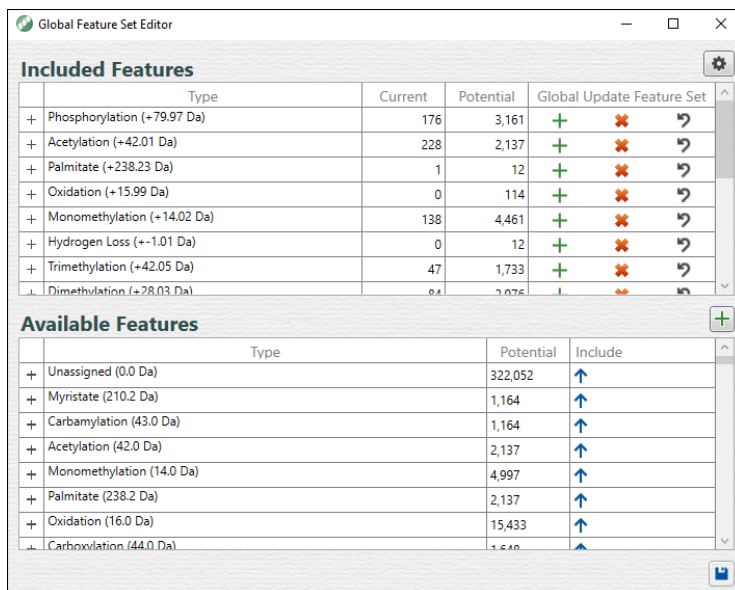
❖ **To add an available feature to the included feature set**

1. Click **Modify Available PTMs,** [Modify Available PTMs] .

   The Global Feature Set Editor opens (Figure 17).

   Figure 17. Global Feature Set Editor

2. If the modification is not available in Included Features, search for it under Available Features.

3. You can apply a modification on a group or an instance level:

   • To make the modification group available, in the Available Features list, click the up arrow for the row to move the complete modification group (that is, all the individual instances) to the Included Features list.

   • To make an instance available, expand the modification group. Select the up arrow on the instance row to move the instance to the Included Features list.

     You can add additional instances one at a time.

4. (Optional) To globally apply the modification: in the Included Features list, select the modification group or instance, and click the plus icon, [+] .

5. Click **Save**.

   The new modifications are available.

Create a ProSightPD database from an existing FASTA database in Proteome Discoverer

The Proteome Discoverer application has many database management utilities that you can utilize for editing FASTA databases prior to importing into the ProSightPD Database Manager, see the *Proteome Discoverer User Guide* for more information. However, these databases cannot be directly imported into the ProSightPD Database manager. The following steps describe the process of exporting a FASTA database from Proteome Discoverer for use in the ProSightPD Database Manager as a Top-down database.

❖ **To export a FASTA database from Proteome Discoverer**

1. Create a new study or use an existing study.

2. Click **New Analysis** to build a workflow consisting of only a Processing Step as shown in Figure 18.

Figure 18. Analysis with only a processing step
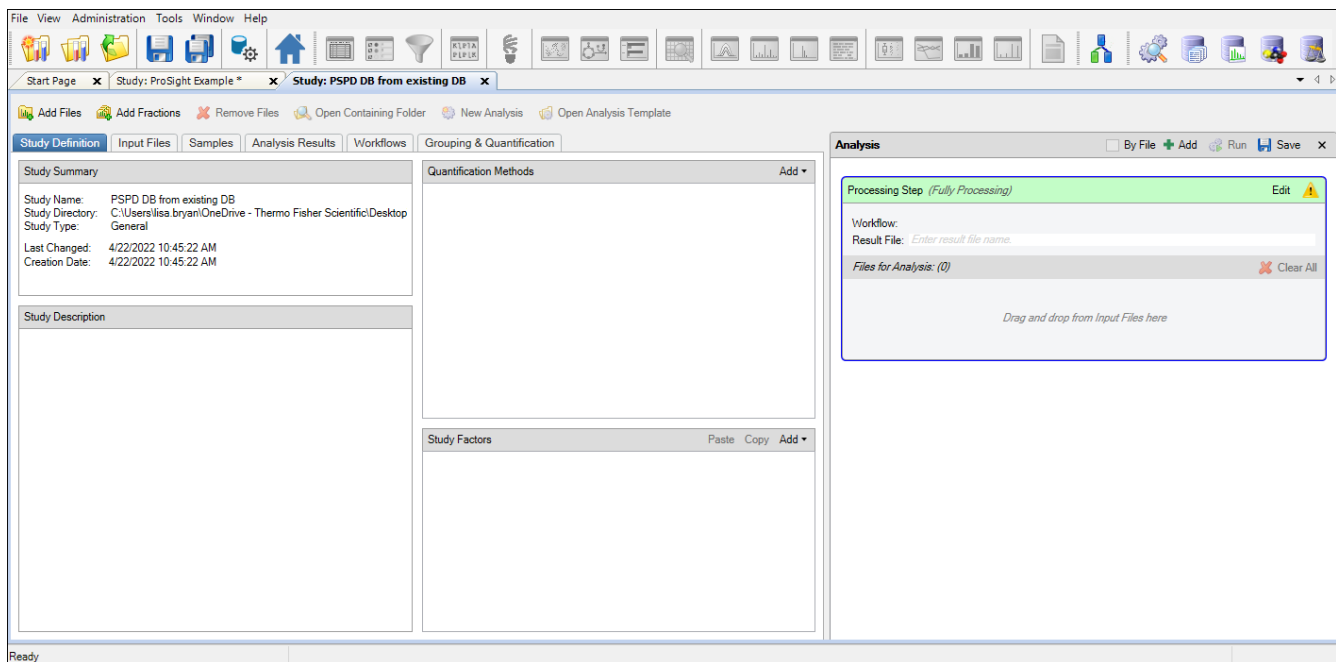
3. Click the **Workflows** tab.

   The Workflow Tree opens.

4. Drag the **ProSightPD FASTA Exporter** node to the Workflow Tree (Figure 19).

Figure 19. ProSightPD FASTA Exporter node in the Workflow Tree
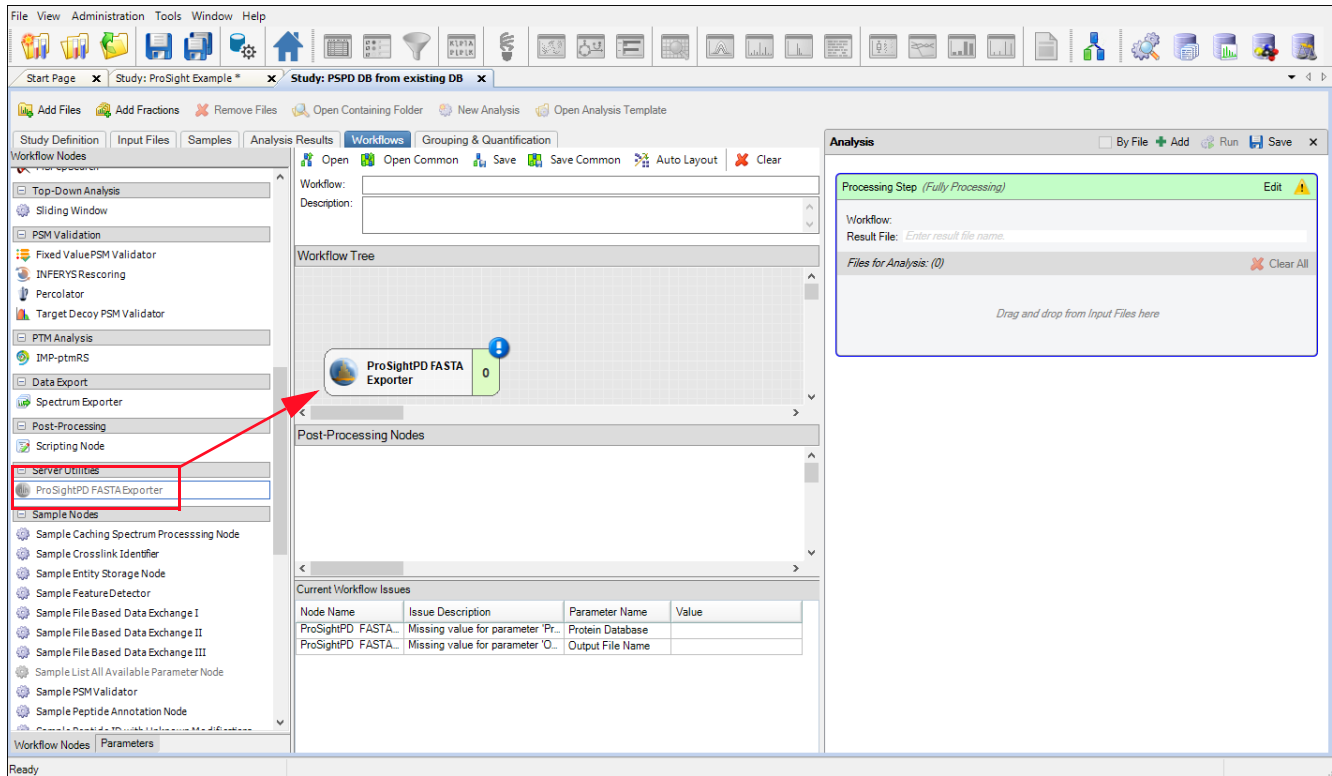
5. Click **Add Files**. Add any RAW data file.

> **Note** The data file will not be processed, however, a raw file must be present to run the workflow.

6. Click the Workflows tab and select the ProSightPD FASTA Exporter node.

7. In the Protein Database parameter, select the database to be exported.

8. In the Output File Name, type the name for the exported database.

9. Click **Run**.

> **Note** The FASTA database exports to the ProSightPD Database folder and must be imported through the ProSightPD Database Manager to create a correctly formatted Top-down database.
>
> This process is only applicable for exporting FASTA databases from Proteome Discoverer. To import any FASTA formatted file, please see this section.
>
> All FASTA file should be in Uniprot format (https://www.uniprot.org/help/fasta-headers). The ProSightPD FASTA Exporter node generates a FASTA file in Uniprot format.

Create a database from a FASTA format file

While XML files are strongly encouraged for database creation, you can also use FASTA files in ProSightPD. To export a FASTA file, see "Create a ProSightPD database from an existing FASTA database in Proteome Discoverer."

> **Note** Be aware that FASTA files do not contain any PTM information. As such, only N-terminal modifications defined in the search nodes or modifications manually added through ProSight Annotator are considered if FASTA files are used.

Make sure the FASTA file format matches the UniProt format, particularly in the format for the file description line. Otherwise, ProSightPD might not be able to process the file.

To enable the software to process the FASTA file, edit the description line to include:

- >sp or >tr
- unique accession for each entry
- pipe (|) before and after the unique accession and no spaces

This is the UniProt format:

```
>sp|P02144|MYG_HUMAN Myoglobin OS=Homo sapiens GN=MB PE=1 SV=2

MGLSDGEWQLVLNVWGKVEADIPGHGQEVLIRLFKGHPETLEKFDKFKHLKSEDEMKASE
DLKKHGATVLTALGGILKKKGHHEAEIKPLAQSHATKHKIPVKYLEFISECIIQVLQSKH
PGDFGADAQGAMNKALELFRKDMASNYKELGFQG
```

This is a non-conforming example:

```
>MYG_HUMAN

MGLSDGEWQLVLNVWGKVEADIPGHGQEVLIRLFKGHPETLEKFDKFKHLKSEDEMKASE
DLKKHGATVLTALGGILKKKGHHEAEIKPLAQSHATKHKIPVKYLEFISECIIQVLQSKH
PGDFGADAQGAMNKALELFRKDMASNYKELGFQG
```

This is an example of an acceptable user-generated FASTA file:

```
>sp|P02144|MYG_HUMAN

MGLSDGEWQLVLNVWGKVEADIPGHGQEVLIRLFKGHPETLEKFDKFKHLKSEDEMKASE
DLKKHGATVLTALGGILKKKGHHEAEIKPLAQSHATKHKIPVKYLEFISECIIQVLQSKH
PGDFGADAQGAMNKALELFRKDMASNYKELGFQG
```

Change the isoform mass threshold

The isoform mass threshold limits the size of isoforms included in the database. You can use the isoform mass threshold to improve search speed by limiting the database entries to only isoforms likely to be found in the sample.

For example, a typical top-down LC/MS experiment is unlikely to detect an intact form of Titin (over 1Mda).

To illustrate the effect of the isoform mass threshold on the number of entries being considered, applying a 70 000 Da threshold to the entire yeast proteome reduces the total number of isoform entries from 6050 to 4550. This 25% reduction speeds up database searching, particularly for subsequence searches.

❖ **To change the upper mass limit of isoforms**

1. Click the gear icon, ⚙, in the Database Manager.

   The Advanced Parameters dialog box opens (Figure 20).

2. For Max Isoform Mass, type a value.

   The default is 70 000 Da.

   Figure 20. Advanced Parameters dialog box
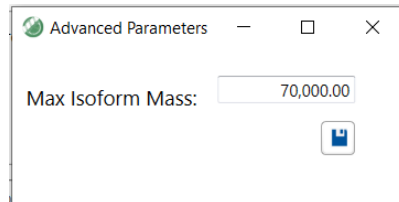
3. Click the **Save** icon.

4. Continue creating your database. See Create a database from a FASTA format file.

Building a top-down database from a bottom-up result file

Bottom-up and top-down databases can have complementary applications. One way to leverage both techniques is to combine the depth of protein IDs from bottom-up with the characterization capability of top-down. ProSightPD can utilize the protein ID results (with modification) from a bottom-up search to create a database for top-down. The steps below describe how to export bottom-up results and create a top-down database.

1. From the Proteome Discoverer Start Page, open the bottom-up result that you want to create a top-down database from.

2. Choose **File > Export > To mzIdentML** (Figure 21).

   The Export to mzIdentML dialog box opens.

Figure 21. File menu
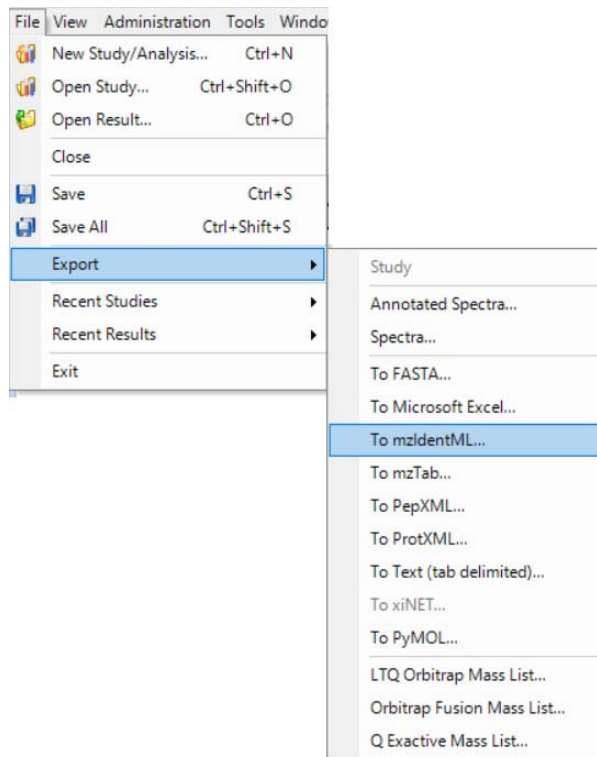
3. Click **Browse** and locate the Destination Folder (Figure 22).

Figure 22. Export to mzIdentML dialog box
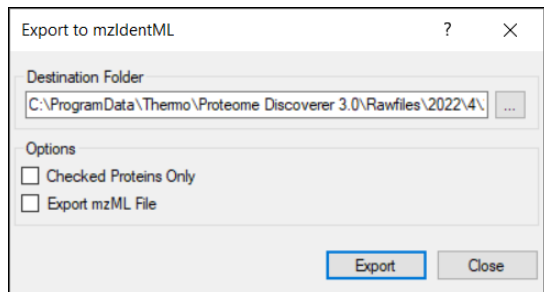
4. Click **Export**.

5. From the Proteome Discoverer application menu bar, click the ProSightPD Database Manager icon, ⚙ . See Create a database in ProSightPD for more information on creating databases in ProSightPD.

The ProSightPD Database Manager opens.

6. Click **Create ProSightPD Database**.

7. Click **Browse**.

The Open dialog box opens.

8. In the file type box at the bottom right corner, select **MzIdentML** (*.mzid file)
   ([Figure 23](#)).

Figure 23. Selecting MzIdentML in the Open dialog box
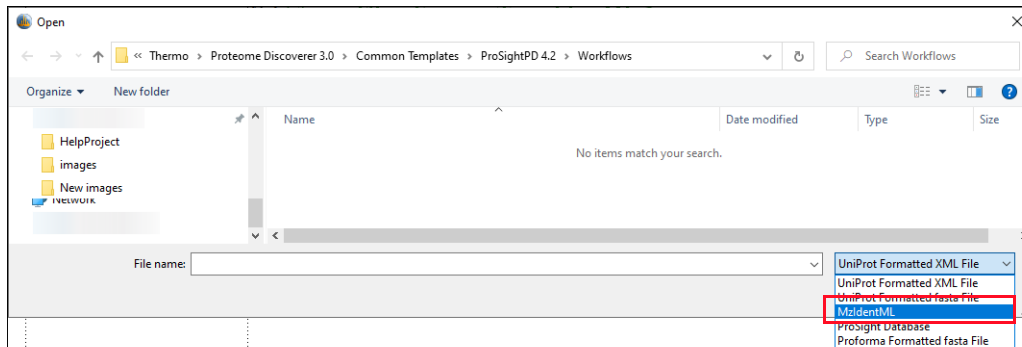
9. Select the exported results and click **Open**.

10. Create and save a top-down database using the ProSightPD Database Manager tool. See
    Create a database in ProSightPD for more information.

Working with studies

This chapter describes studies and analyses.

Contents

Studies

A study is a container for your data files and the analyses performed with those data files. The study is also where you structure and annotate your data with meta information, such as the biological condition (for example, disease versus control), analysis time points, biological replicate number, and so on.

The ProSightPD application analyzes data by running a processing workflow and a consensus workflow in the study.

Label-free quantitation

The ProSightPD application performs the label-free quantification (LFQ) of proteoforms using these nodes:

- ProSightPD Hi Res Feature Detector or the ProSightPD Med Res Feature Detector node in the processing workflow
- Feature Mapper node and ProSightPD Quantifier node in the consensus workflow

Figure 24 shows the workflows.

Figure 24. Processing workflow and consensus workflow for label-free quantitation
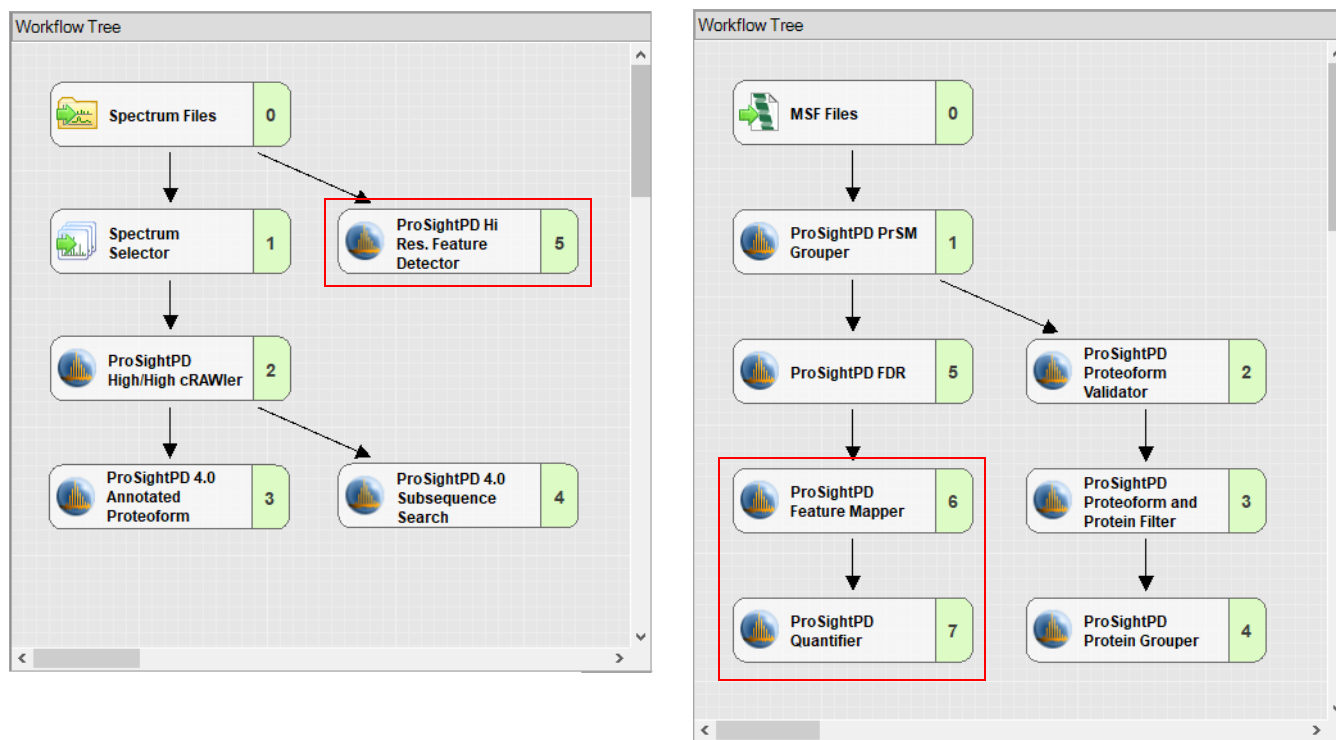
Create a study

The most convenient way to create a new study is to use the New Study and Analysis dialog box. You can add files to the study after creating it, but the import function is available only in the New Study and Analysis dialog box.

❖ **To create a study using the New Study and Analysis dialog box**

1. On the Start page, click **New Study/Analysis**.

   The New Study and Analysis dialog box opens.

   Figure 25. New Study and Analysis dialog box

2. Type the Study Name.

3. For the Study Root Directory, click **Browse** and locate the path.

Put the study directory on a local drive that has approximately the same free space as your raw data file size.

4. (Optional) For Import from File, click **Browse** and locate the file that contains the study definitions.

5. For the Processing Workflow, click **Browse** and locate the processing workflow template that matches the purpose of your study.

   To use an analysis template, which includes the processing workflow and the consensus workflow, leave the workflow fields empty. The analysis templates are not available until the study is created.

   For more information on workflows and templates, see Process workflows, consensus workflows, and analyses.

6. For the Consensus Workflow, click **Browse** and locate the consensus workflow template based on how you want to filter your results.

   For more information on workflows and templates, see Process workflows, consensus workflows, and analyses.

Process workflows, consensus workflows, and analyses

A workflow is composed of a series of nodes, and each node performs an individual step of the data analysis. Depending on the workflow's purpose, it usually consist of 5 to 10 nodes. Most nodes have their own parameters to set. The basic functions of the workflows are:

- In the processing workflow, the software searches the database and assigns PrSMs. If the analysis is quantitative, the processing workflow also performs feature detection.

- The consensus workflow filters the preliminary results (for example, by the FDR node) and groups them into proteoforms, isoforms, and proteins. The results display these groups. If there is quantitative analysis in the processing workflow, then the consensus workflow performs the required feature mapping and statistical analysis.

You can run multiple processing workflows separately and then combine them in the consensus workflow for unified reporting.

To simplify creating a workflow, the software includes a set of fit-for-purpose templates. Build your workflow from the provided templates, which have been tested and optimized for confident proteome analysis. Many of the node parameters already have their appropriate settings in the template.

Proteome Discoverer has three types of templates: processing workflow templates, consensus workflow templates, and analysis templates, which contain both the processing and consensus workflows. The templates are named for their intended use. For example, the Comprehensive Discovery template includes nodes that cover the largest search space.

Configure a workflow

The following sections describe how to add or change workflows and other related tasks:

- Add or change an analysis template

- Add or change a workflow after creating the study

- Replace a node in the workflow

- Perform Single Proteoform searches in the processing workflow

Add or change an analysis template

❖ **To add or change an analysis template**

1. Create a study in the dialog box without selecting either a processing or a consensus workflow.

    For instructions, see To create a study using the New Study and Analysis dialog box.

2. On the Study Definition page, click **Open Analysis Template**, 📁 Open Analysis Template .

3. In the File Explorer, open the analysis template:

    C:\Users\Public\Documents\Thermo\Proteome Discoverer *version*\Common Templates\ProSightPD *version*\Analysis Templates

4. Add files, study factors, and a database and adjust the parameters to suit your analysis.

Add or change a workflow after creating the study

> **Tip** You can overwrite the existing workflow with the new workflow, that is, you do not need to delete the existing workflow first.

❖ **To add or change a workflow after creating the study**

1. Do one of the following:

    - To use a workflow supplied in the application, click **Open Common**.

    - To use a workflow saved on your computer, click **Open**.

2. Locate the workflow, and click **Open**.

3. (Optional) Repeat to add or change another workflow.

4. Add files, study factors, and adjust the node parameters to suit your analysis.

Replace a node in the workflow

You might want to change a workflow node for many reasons, for example, you are getting unexpected results, too many results, or changing from a database to a single proteoform search.

❖ **To replace a node in the workflow**

1. Select the node that you want to replace.

   This example shows the ProSightPD High/High cRAWler node selected.
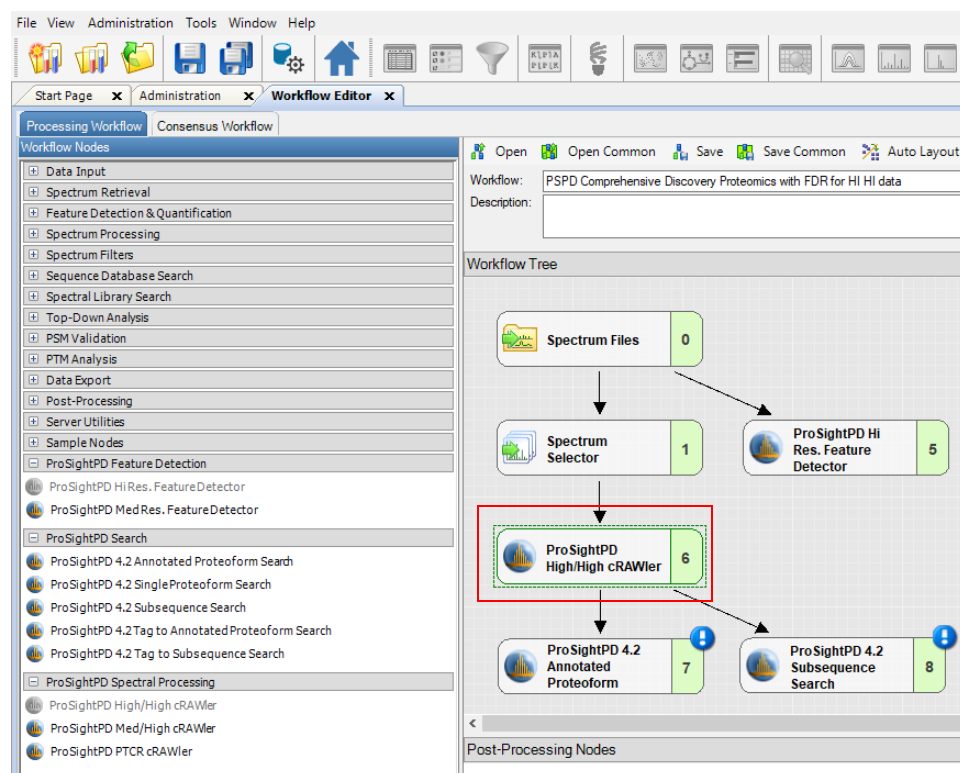
2. Press DELETE on the keyboard.

3. Drag the new node into the workflow pane, and connect the nodes.

Perform Annotated Proteoform searches in the processing workflow

The Annotated Proteoform Search node is designed to quickly analyze raw data. This node considers only proteoforms included in your database. This includes any known PTMs, truncations, or amino acid variants that are annotated in the database.

Annotated Proteoform searches can be applied in the following two ways:

- A narrow tolerance search

- A wide tolerance search

A narrow tolerance search is usually conducted with a 2.2 Da precursor mass tolerance to account for the common "off-by-one" data deconvolution artifacts. The narrow tolerance search completes quickly and the results do not require further analysis to improve confidence (a few exceptions might include: loss of hydrogen from S-S bonds and deamidation).

A wide tolerance search typically utilizes a 100-1000 Da precursor mass tolerance. This search returns many proteoforms whose precursor mass varies significantly from the theoretical precursor mass. The goal of this search is to find proteoforms that might have an unknown modification resulting in a larger precursor mass error, which could only be found using a wide tolerance. These results require additional work to determine the modification.

Thermo Fisher Scientific recommends opening individual results in the ProSight Lite application to further explore potential modifications. If the modification is identified, it can be added to the database. Then the search must be re-processed to find all proteoforms containing that modification.

To set advanced parameters for an Annotated Proteoforms search, see "Advanced parameters of the Annotated Proteoform search."

❖ **To add an Annotated Proteoform search in the processing workflow**

1. Open a workflow template that includes the ProSightPD Annotated Proteoform Search node.

   a. If the workflow does not include the Annotated Proteoform Search node, drag and drop the Annotated Proteoform Search node into the workflow.

   b. Connect the node by dragging the arrows from the upstream node to the Annotated Proteoform node.

2. Select the Annotated Proteoform node to view the parameters.

3. Set the Input Database by selecting from the dropdown menu.

   > **Note** To create a new database launch the Database Manager. To select a newly created database, you must close and reopen any open studies.

4. Set the Precursor Mass Tolerance.

   The following are recommended values for narrow or wide tolerance searches:

   - Narrow tolerance search is 2.2 Da

   - Wide-tolerance search 200 Da

     > **Note** For large modifications like signal peptides, consider setting your tolerance larger than the potential modification.

5. Set the Fragment Mass tolerance.

Thermo Fisher Scientific recommends 10 ppm. Use the narrowest tolerance possible for your data. As the tolerance widens, the search time and false positive rate increases.

6. Set the FDR calculation.

The default is True, which means that the application calculates the FDR based on the threshold set in the FDR node in the consensus workflow. To calculate the FDR, the FDR node must be present in the consensus workflow. If your sample is not complex or if you want to see all PrSMs, set this value to False to not calculate an FDR.

> **Note** Results are automatically displayed at the specified FDR cutoff for each result level (PrSM level, Proteoform level, Isoform level, and so on)

7. (Optional) Select a static modification to apply to a specific residue.

Currently, you can only apply one static modification in the search node. To set additional modifications, use the ProSight Annotator in the Database Manager.

Advanced parameters of the Annotated Proteoform search

The default parameters of the Annotated Proteoform search are designed to work in most situations. However, analyses can be customized to improve performance by optimizing the advanced parameters listed below. For a full list of advanced parameters, see "Annotated Proteoform Search node."

Table 4 is a list of advanced parameters and a brief description of how they impact the Annotated Proteoform search and results.

Table 4. Advanced parameters for the Annotated Proteoform

Parameter	Description
Maximum PTMs per Isoform	Limits the number of PTMs that are expanded and searched during search time. For example, if an isoform has 10 PTMs in the database and the maximum is set to 4, then only proteoforms with up to 4 PTMs are searched. This improves the search time. However, in cases of highly modified proteoforms, you should increase this value.
Maximum SNPs per Isoform	Limits the number of amino acid mutations of single nucleotide polymorphisms (SNPs) in the same way as the Maximum PTMs per isoform parameter.
Remove Disulfide Bonds	Thermo Fisher Scientific recommends setting this parameter to False. You can set this parameter to True if there are annotated disulfide bonds in the search database that you want to avoid searching, or when there is a known error in the disulfide bond annotation in the database. Errors in UniProt where disulfide bonds are placed on non-cysteine residues have been observed and can cause the search to crash.

Table 4. Advanced parameters for the Annotated Proteoform

Parameter	Description
Maximum Mass to Include PTMs	Limits the size of proteoforms that are considered with PTMs. It is unlikely to observe very large proteoforms, therefore it is unnecessary to consider all of the modified forms beyond the size expected to be observed in the experiment. Setting the limit to a reasonable size (100 kDa) allows the search to complete sooner.
N-Term Modifications to Include	N-terminal modifications are added at search time unless specifically added to the database. These modifications generally include acetylation or formylation, and the default is set to include N-terminal acetylation.
Delta M Mode	To identify unknown mass shifts, consider setting Delta M Mode to True. Searching with Delta M Mode will slow the search. Thermo Fisher Scientific recommends turning off FDR when utilizing Delta M Mode as the Delta M search affects the FDR threshold. Delta M Mode is suggested for data that has already been searched or in a target case.
Decoy Reps	Sets the number of times the decoy database is searched. The default setting is 1, which results in the fastest search. However, results can vary minimally run-to-run: 1-2 proteoform differences based on exactly where the FDR thresholds are set. Increasing the number of times the decoy database is searched to 3 results in slower searches but reduces the amount of variation between analyses.
Maximum PrSMs Per Precursor	Limits the number of spectral matches attributed to a single precursor. Any given precursor can have many associated spectral matches depending on search parameters and the data. In general, there are very few highly confident matches (based on P-Score or E-Value.) This parameter dictates the top N scoring PrSMs to return per precursor. The default value is 3, however not all 3 PrSMs can pass the FDR cutoff depending on their confidence. Due to the nature of the MS isolation window, the occurrence of off-by-one Dalton deconvolution errors and modifications such as deamination, which are ~1 Dalton, do not expect to see multiple high-scoring PrSMs for a single precursor (this is why the default setting is 3). If you expect multiple proteoforms to have the same precursor mass, (for example, positional isomers such as histones), you can increase this setting to 5 or 10.

Table 4. Advanced parameters for the Annotated Proteoform

Parameter	Description
Minimum Matched Fragments	Lets you discard spectra that deconvoluted with less than n fragment ions. The default value is 3 and you can increase it to 5 or 10 without major loss of IDs. For smaller proteoforms (< 5kDa), set the minimum to 5 or lower. By discarding sparse spectra, the search completes faster.
Include Labile Modification Mass Shift	Lets you apply a mass shift to all precursors to account for loss of a labile modification. To activate the setting, set Include Labile Modification Mass Shift to True and input a mass in the Labile Modification Mass Shift parameter.

Perform a Subsequence search in the processing workflow

The Subsequence search comprehensively searches subsequences of the isoforms in a specified database. The node considers any PTMs or other annotations found in the database. The Subsequence search is more time- and resource-intensive than the Annotated Proteofrom search. For any given sequence in the database, there could be hundreds of possible subsequences with additional complexity when considering PTMs and other modifications of each subsequence. Due to the increased search space, the Subsequence search utilizes more resources. Use the Subsequence search node to identify unexpected proteoforms in a sample and for discovering protein clipping products and endogenous truncations.

Thermo Fisher Scientific recommends first running the Subsequence search in the most restrictive manner possible and then incrementally increasing the search space as needed. A wide window Subsequence search (>20 ppm) can take days to complete and in cases of very wide precursor windows (>100 Da) may cause computer memory issues. To restrict the Subsequence search space, adjust the following parameters:

- For Precursor Mass Tolerance, set this to 15 ppm or fewer for complex databases.

- For small or targeted searches, tolerances of 2.2 Da can be useful to account for off-by-one errors.

❖ **To add a Subsequence search in the processing workflow**

1. Open a workflow template that includes the ProSightPD Subsequence Search node.

   a. If the workflow does not include the Subsequence Search node, drag and drop the Subsequence Search node into the workflow.

   b. Connect the node by dragging the arrows from the upstream node to the Subsequence Search node.

2. Select the Subsequence Search node to view the parameters.

3. Set the Input Database by selecting from the dropdown menu.

> **Tip** To create a new database, launch the Database Manager. To select a newly created database, you must close and reopen any open studies.

4. Set the Precursor Mass Tolerance.

   Thermo Fisher Scientific recommends 10 ppm.

   > **Note** Due to the large search space covered by the Subsequence Search, Thermo Fisher Scientific strongly suggests setting the precursor mass tolerance no greater than 25 ppm. For smaller data files or databases, you can use wider tolerances without causing the search time to increase.

5. Set the Fragment Mass Tolerance.

   Thermo Fisher Scientific recommends 10 ppm. Use the narrowest tolerance possible for your data. As the tolerance widens, the search time and false positive rate increases.

6. Set the FDR calculation.

   The default is True, which means that the FDRs are calculated based on the threshold set in the FDR node in the consensus workflow. The FDR node in the consensus workflow is required. If your sample is not complex or you want to see all PrSMs, set this value to False to not calculate the FDR.

   > **Note** Results are automatically displayed at the specified FDR cutoff for each result level (PrSM, Proteoform, Isoform, and so on).

7. (Optional) Select a static modification to apply to a specific residue.

   You can apply one only static modification in the search node. To set additional modification, use the ProSight Annotator tool in the Database Manager.

Advanced parameters of the Subsequence search

The default parameters of the Subsequence search are designed to work in most situations. However, most analyses can be improved/customized by optimizing the advanced parameters listed below. For a full list of advanced parameters, see "Subsequence Search node." Table 5 is a list of advanced parameters and a brief description of how they impact the Subsequence search and results

Table 5. Advanced parameters of the Subsequence search

Parameter	Description
Maximum PTMs per Isoform	Limits the number of PTMs that are expanded and searched during search time. For example, if an isoform has 10 PTMs in the database and the maximum is set to 4, then only proteoforms with up to 4 PTMs are searched. This improves the search time. However, in cases of highly modified proteoforms, it is advised that you increase this value.
Maximum SNPs per Isoform	Limits the number of amino acid mutations of single nucleotide polymorphisms (SNPs) in the same way as in the Maximum PTMs per Isoform parameter.
Maximum PTMs per Proteoform	Limits the number of PTMs that are allowed on each proteoform during search time. For example, if an isoform has 10 PTMs in the database, the maximum per isoform is set to 8, and the maximum per proteoform is set to 4, then only proteoforms (or subseqeunces) with up to 4 PTMs are searched. This improves the search time. However, in cases of highly modified proteoforms, it is advised that you increase this value.
	Tip The PSDB database contains isoforms and PTMs. All proteoforms to be searched are generated at search time according to the database and parameters selected in the search nodes. For the Subsequence search, each proteoform is expanded into its subsequences, so in the Subseqeunce search it is useful to limit the PTMs at the proteoform level.

Table 5. Advanced parameters of the Subsequence search

Parameter	Description
Maximum SNPs per Proteoform	Limits the number of amino acid mutations of single nucleotide polymorphisms (SNPs) in the same way as the Maximum PTMs per Proteoform parameter.
Remove Disulfide Bonds	Thermo Fisher Scientific recommends setting this parameter to False. You can set this parameter to True if there are annotated disulfide bonds in the search database that you want to avoid searching or when there is a known error in the disulfide bond annotation in the database. Errors in UniProt where disulfide bonds are placed on non-cysteine residues have been observed and can cause the search to crash.

Perform Single Proteoform searches in the processing workflow

The Single Proteoform search node allows you to search your data for a single user-defined proteoform without having to create a database. The user-defined proteoform is input in the ProForma format, which consists of the backbone residues in single letter amino acid abbreviation and modifications entered in brackets after the residue which is modified. The modifications can be entered in the following ways:

- a PSI-MOD number

- a RESID ID number

- a UNIMOD accession number

Each of the sources (PSI,RESID, and UNIMOD) are databases of known protein modifications. Modifications are linked to an ID or reference number. These numbers are used in the Single Proteoform search to represent the modification.

When entering a modification use the following syntax:

Table 6. List of compatible protein modification databases

Database	Example
PSI MOD	MOD:00057
	For example: …SGRG[MOD:00057]K…
UNIMOD	UNIMOD:1
	For example: …SGRG[UNIMOD:1]K…
RESID	RESID:AA0048
	For example: …SGRG[RESID:AA0048]K…

For a full list of modifications, visit the following websites:

- UNIMOD (https://unimod.org)

- RESID (https://proteininformationresource.org/cgi-bin/resid.pl)

- PSI-MOD (https://www.ebi.ac.uk/ols/ontologies/mod)

❖ **To perform single proteoform searches in the processing workflow**

1. Open a workflow template that includes the ProSightPD Single Proteoform Search node, or add the node after the cRAWler node.

2. In a web browser, navigate to https://proteininformationresource.org/cgi-bin/resid.pl.

3. Note the modification IDfrom the database of choice.

4. In the Proteome Discoverer application, click the **ProSightPD Single Proteoform Search** node, and then click **Advanced Parameters**.

5. In the Proforma Sequence, enter the protein sequence (Figure 26).

   The Proforma Sequence is the protein sequence plus the modification in the form of a Proforma database entry. Enter the modification number after the residue to be modified in brackets.

Figure 26. The Proforma Sequence entry for methylated lysine

Modification number in brackets
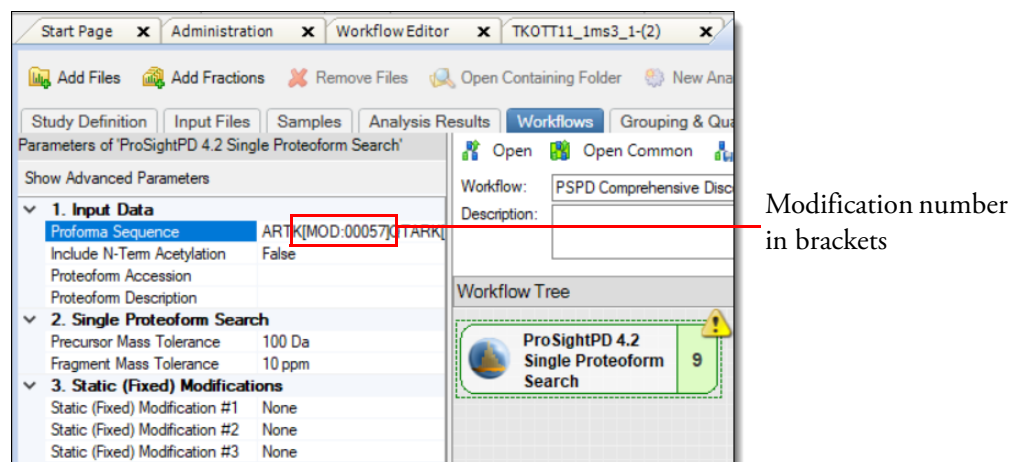
> **Note** Put the identifier in brackets after the amino acid containing the modification. For example, AA0057 N2-acetyl-L-lysine becomes AA0057, and an acetylation of a lysine residue in a protein sequence is expressed as [AA0057].

6. To add the modification, enter the modification number in brackets after the residue.

7. For manual precursor mass assignment, in the crawler node's Advanced Parameters, do the following:

   a. Set Use Manual Precursor to **Always** or **If Not Detected**.

b.   Type the expected Precursor Mass (Daltons).

8.   Configure the consensus workflow, or click **Run**.

For basic workflow information, refer to the *Proteome Discoverer User Guide*.

For more information about using the node, see Single Proteoform Search node.

Working with results

Contents

The results report

The ProSightPD application generates a results report that contains a list of the matching proteoforms identified by the search nodes that you selected (Figure 27). You can access the results report through the application's Start page or File menu. Through the Job Queue, you can access the results, including a failed workflow analysis. The application can save multiple search results within a single report.

The application saves the results report in .pdResult file format so you can also access the results through the Windows Explorer.

Figure 27. Results reports displayed in the application window

Field Chooser

List of open pages

Result Table Chooser

Results table

Status bar
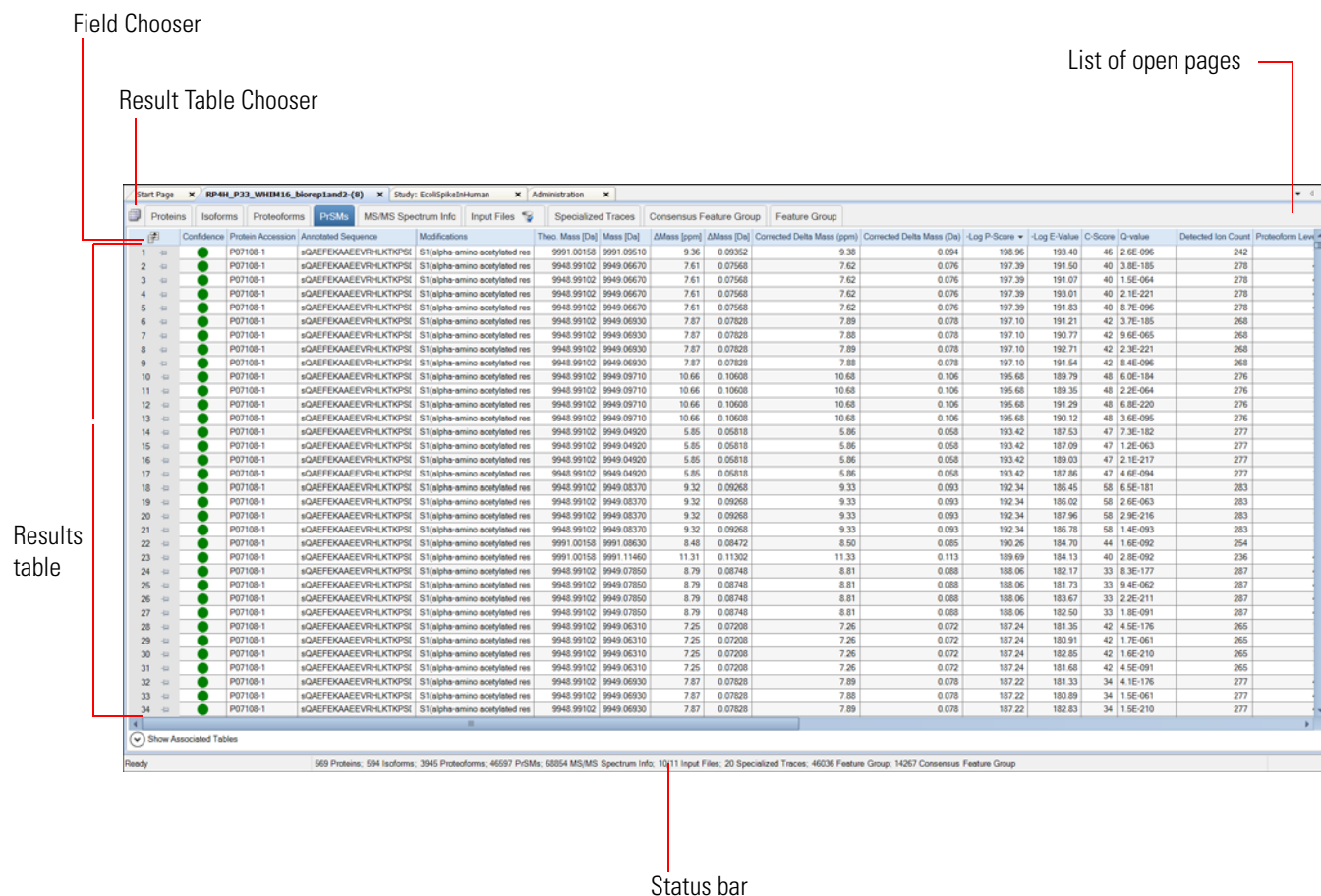
The results report includes the following pages:

- Proteins page
- Isoforms page
- PrSMs page
- MS/MS Spectrum Info page
- Input Files page
- Specialized Traces page

Proteins page

By default the results report (.pdResult file) opens to the Protein Groups page.

Figure 28. Commonly used columns on the Proteins page

	Checked	Accession	# of Isoforms	# of Isoforms with Characterized Proteoforms	# of Proteoforms	Proteoform Ch	# Characterized Proteoforms	# of PrSMs	Q-value
1	☐	P64463	1	1	2	●	1	9	5.2E-039
2	☐	P33219	1	1	1	●	1	18	2.7E-035
3	☐	P0AB14	1	1	1	●	1	2	6.1E-033
4	☐	P0ACE7	1	1	1	●	1	6	8.7E-027
5	☐	P0A800	1	1	1	●	1	6	4.7E-025
6	☐	P0AFH8	6	5	9	●	7	98	2.2E-024
7	☐	P0AE08	1	1	6	●	2	26	2.4E-022
8	☐	P0A763	2	2	2	●	2	6	3.7E-021
9	☐	P0AC62	1	1	1	●	1	8	1.0E-020
10	☐	P0ADU5	1	1	1	●	1	3	1.2E-018
11	☐	P0AF50	3	2	4	●	2	5	1.9E-009
12	☐	P45470	1	1	1	●	1	2	1.2E-008
13	☐	P76402	1	1	1	●	1	2	1.3E-007
14	☐	P0ADE6	1	1	1	●	1	1	1.5E-007
15	☐	P45578	1	1	1	●	1	1	1.1E-004

The Proteins page reports all of the proteins identified in the experiment that pass any user-defined threshold such as FDR. From this page, you can see the UniProt Accession and the number of proteoforms and isoforms of each protein identified in the experiment. Additionally, the number of PrSMs and characterized proteoforms (that is, proteoforms with C-Score greater than 40) are listed. If you enabled FDR, then the Q-Value for each protein is also displayed.

You can access additional information for each protein identified by selecting it and clicking the down arrow next to Show Associated Tables, ⊙ Show Associated Tables .

Table 7. Proteins page columns

Column	Description
Accession	UniProt™ accession identifiers used to cite UniProt entries.
# of Isoforms	The number of isoforms identified from the associated protein accession.
# of Isoforms with Characterized Proteoforms	The number of isoforms with proteoforms that have a C-Score greater than 40.
# Proteoforms	The number of identified proteoforms from the associated UniProt protein accession
Proteoform Characterization Confidence	The visual indication of C-Score confidence (default values are >40=green,<40->3=yellow, <3=red). You can define these thresholds in the ProSight Protein and Proteoform Filter node in the consensus workflow.
# Characterized Proteoforms	The number of proteoforms having a C-Score greater than 40.
Q-value	The likelihood of discovering the associated protein by chance given the experimental evidence.

Isoforms page

The Isoforms page displays all of the identified isoforms (Figure 29). The Isoforms page also includes the UniProt description for each isoform and the mass in kDa, pI, and, if enabled, the protein annotations for gene ontologies and various pathway analysis.
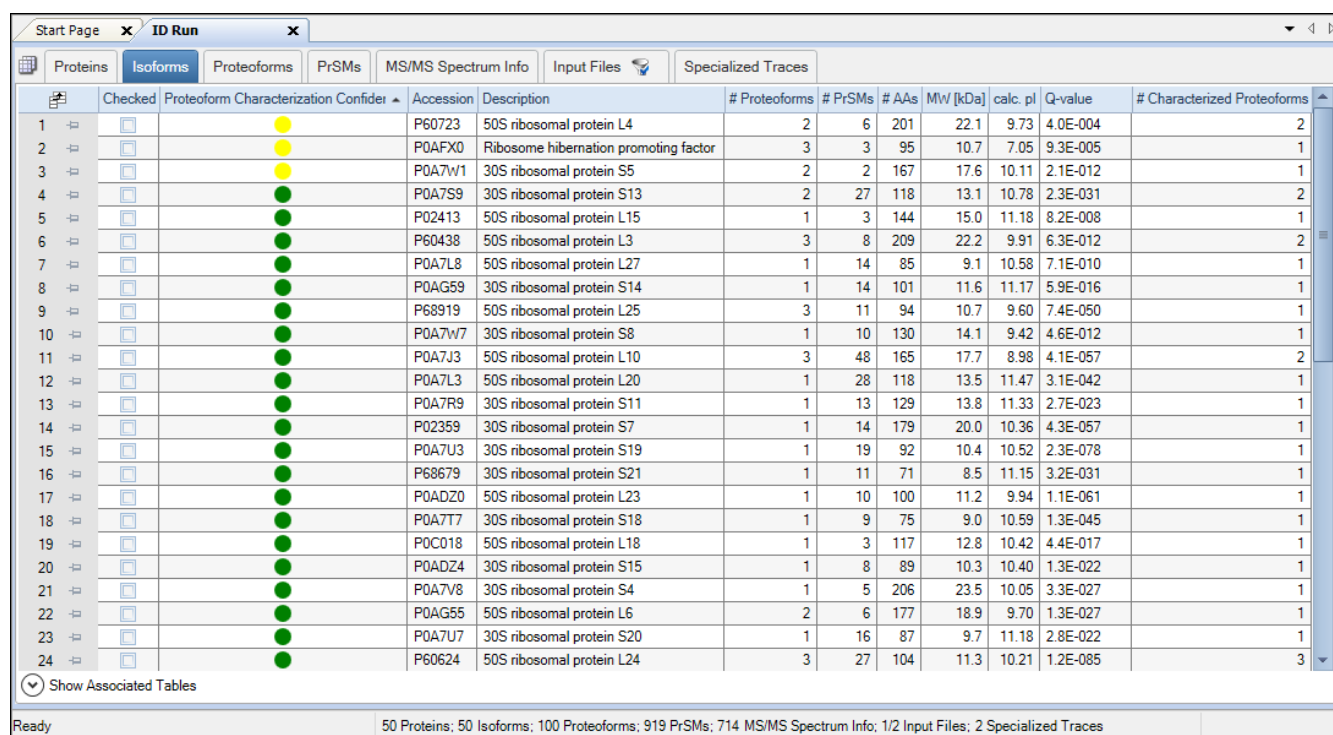
You can access additional information for each isoform identified by selecting the isoform and clicking the down arrow next to Show Associated Tables, [⌄ Show Associated Tables].

Figure 29. The Isoforms page

Table 8. Commonly used columns on the Isoforms page

Column	Description
Proteoform Characterization Confidence	Visual indication of C-Score confidence (>40=green, <40->3=yellow, <3=red). For information on defining the thresholds, see the PrSM Grouper node parameters.
Accession	UniProt accession identifiers used to cite UniProt entries.
Description	The protein description from UniProt.
# Proteoforms	The number of identified proteoforms from the associated UniProt protein accession.
# PrSMs	The number of PrSMs supporting this isoform assignment.
# Characterized Proteoforms	The number of proteoforms having a C-Score greater than 40.
Q-value	The chance of discovering the associated result given the experimental evidence.

Proteoforms page

The Proteoforms page displays all of the identified proteoforms.

Figure 30. The Proteoforms page

The Proteoforms page also includes the identified sequence for each proteoform and the number of supporting PrSMs, theoretical mass in kDa, search engine confidence, -Log P-Score, -Log E-Value, best PrSM C-score, and Sequence Coverage (as % residue cleavage.) If FDR is enabled, proteoform level Q-Values are also displayed.

For each result at the proteoform level, you can launch the ProSight Lite and TDValidator Lite applications. For additional information on these applications, see:

- Use ProSight Lite

- Use TDValidator Lite

You can access additional information for each identified proteoform by selecting the proteoform and clicking the down arrow next to Show Associated Tables, ⊙ Show Associated Tables .

Table 9. Commonly used columns on the Proteoforms page (Sheet 1 of 2)

Column	Description
Proteoform Characterization Confidence	Visual indication of C-Score confidence (>40=green, <40->3=yellow, <3=red). For information on defining the parameters, see PrSM Grouper node.
Sequence	Proteoform sequence.
# PrSMs	Number of proteoforms.

Table 9. Commonly used columns on the Proteoforms page (Sheet 2 of 2)

Column	Description
Theo Mass [Da]	Proteoform mass based on the matched sequence.
Confidence	Visual representation of the proteoform's PrSM confidence based on the -log E-value levels specified in the ProSightPD configuration.

Higher confidence = green, medium confidence = yellow, and low confidence = red |
-Log P-Score	Statistical confidence for proteoform assignment. Large positive numbers are more confident. Values less than 4 are considered poor.
-Log E-Value	Statistical confidence for proteoform assignment adjusted for the search space.
Best PrSM C-Score	Highest C-Score of PrSMs supporting the associated proteoform.
Average PrSM Detected Neutral Mass	Average observed mass from all PrSMs associated with a proteoform.
% Residue Cleavages	Percentage of observed fragments compared to the total possible cleavages. Also known as sequence coverage.
External Top Down Displays	Click to view the result in the ProSight Lite or the TDValidator Lite applications. For more information, see Use ProSight Lite and Use TDValidator Lite.
Proforma	The proforma sequence notation for the proteoform identification.

PrSMs page

The Proteoform Spectral Matches (PrSMs) page reports all of the PrSMs identified in the experiment.

The annotated sequence showing modifications, LogE,-LogP, are shown for each PrSM. To view additional information about related proteoforms, proteins, and isoforms, click the down arrow next to Show Associated Tables, [⌄ Show Associated Tables].

When Delta M Mode is enabled, two additional columns are available on the PrSMs tab: Delta M Residue and Delta M Fragment(s). The Delta M Residue column shows where the Delta M mass has been placed on the proteoform to produce the highest number of additional matching fragment ions. Each additional fragment ion found by Delta M is displayed in the Delta M Fragment(s) column. The delta masses also display in the ProSightPD fragment map with an orange square around the potential residue where the delta mass was localized.

Figure 31. PrSMs page

To further investigate each result, you can launch the ProSight Lite and the TDValidator Lite applications from the PrSMs table.

Table 10. Commonly used columns on the PrSMs page (Sheet 1 of 2)

Column	Description
Confidence	Visual representation of confidence (High Confidence = green, Medium=Yellow, Low= Red). You can define these thresholds.
Detected Ion Count	The number of fragment ions found in the fragment ion spectra.
Identifying Node	Search node used to identify this PrSM.
Annotated Sequence	The proteoforms backbone sequence and PTMs.
Modifications	List of modifications found on each proteoform.
# Proteins	The number of proteins each proteoform maps back to.
Protein Accessions	Parent protein accession.
Charge	Charge state isolated for fragmentation of the proteoform.
m/z [Da]	Observed mass.
Mass [Da]	Monoisotopic observed Mass.
Theo. Mass [Da]	Theoretical mass based on matched proteoform from the database.

Table 10. Commonly used columns on the PrSMs page (Sheet 2 of 2)

Column	Description
ΔMass [ppm]	Observed theoretical mass error in parts per million.
ΔMass [Da]	Observed-theoretical mass error in Daltons.
Matched Ions	Number of matched fragment ions supporting the PrSM.
Delta M Residue	Possible location of the delta mass supported by the corresponding Delta M fragments.
Delta M Fragment(s)	Fragments found using Delta M Mode.

MS/MS Spectrum Info page

The MS/MS Spectrum Info page lists all of the MS/MS spectra searched in the experiment and certain details such as scan number, injection time, # of PrSM (supported by this MS/MS scan), # of Precursors (associated with this MS/MS scan). You can use the MS/MS Spectrum Info table to interrogate which MS/MS spectra are associated with which proteoform or PrSM result.

Figure 32. MS/MS Spectrum Info page

❖ **Work with the MS/MS Spectrum Info page**

1. To display the associated precursor scans (precursor before and after the MS/MS scan), click the **MS/MS Spectrum Info** icon, ⊞, in the upper-left corner of the Field Chooser.

   The Field Chooser menu opens.
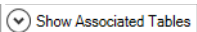
2. From the Field Chooser menu, select **Master Scan**.

3. To display Associated PrSMs, Proteoforms, and so on, select the down arrow next to Show Associated Tables, ⊙ Show Associated Tables .

Table 11. Commonly used columns on the MS/MS Spectrum Info page

Column	Description
File	The .raw data file name.
RT [min]	Retention time of the MS/MS scan (in minutes).
First Scan	First precursor scan associated with this MS/MS scan.
Activation Type	Activation type used to generate ions in the MS/MS scan.
NCE[%]	Normalized Collision Energy applied if HCD is used.
# PrSMs	Number of PrSMs that this MS/MS scan supports.
Isolation Interference [%]	Measure of co-isolated precursor interference.
Ion Inject Time [ms]	How long ions were collected prior to analysis. Shorter times indicate a more abundant signal.
m/z [Da]	Observed mass.
Mass [Da]	Monoisotopic observed mass.
Theo. Mass [Da]	Theoretical mass based on matched proteoform from database.
ΔMass [ppm]	Observed theoretical mass error in parts per million.
ΔMass [Da]	Observed theoretical mass error in Daltons.
Matched Ions	Number of matched fragment ions supporting the PrSM.
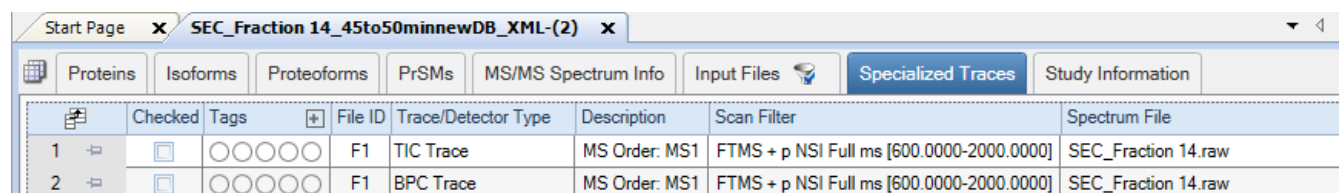
Input Files page

The Input Files page shows the raw data files searched in this experiment. Refer to the *Proteome Discoverer User Guide* for more information.

Specialized Traces page

The Specialized Traces page provides the total ion chromatogram (TIC) and base peak chromatogram (BPC) for each input file (Figure 33).

Figure 33. Specialized Traces page

Table 12 lists the columns of the Specialized Traces page.

Table 12. Columns of the Specialized Traces page

Column	Description
Tags	Not used with the ProSightPD workflows supplied with the software.
File ID	Number the application assigns when importing the raw data file.
Trace/Detector Type	Base Peak Chromatogram (BPC) or Total Ion Chromatogram (TIC).
Description	MS or MS/MS.
Scan Filter	The *m/z* range.
Spectrum File	The raw data file name.

Quantitative results in ProSightPD

The LFQ workflow generates additional results pages and columns:

- Feature Group page

- Consensus Feature Group page

Feature Group page

The Feature Group page reports a list of all features individually measured from all data files. Use the Feature Group page to see the mass and abundance of each detected feature as well as additional features (Figure 34).

Figure 34. Feature Group page columns

		Checked	File ID	Apex RT [min]	Area	Intensity	# LCMS Features	Scan Range	RT Range	Mass [Da]	Quan Info	# PrSMs
1		☐	F44	61.74	5.108e6	0.000e0	1	1595-1614	61.383-62.185	1996.14556	Shared	2
2		☐	F44	19.00	1.620e6	0.000e0	1	505-515	18.828-19.203	2011.19391	Shared	2
3		☐	F44	33.48	8.849e7	0.000e0	1	889-919	32.999-33.91	2012.22612	Shared	4
4		☐	F44	89.62	4.979e6	0.000e0	1	2231-2249	89.418-90.253	2014.12513	Shared	3
5		☐	F44	16.99	2.302e6	0.000e0	1	445-458	16.696-17.181	2029.14486		0
6		☐	F44	17.69	5.041e7	0.000e0	1	459-494	17.22-18.369	2038.08733	Shared	6
7		☐	F44	39.52	1.788e6	0.000e0	1	1060-1074	39.253-39.768	2047.00339	Shared	2
8		☐	F44	37.82	1.987e7	0.000e0	1	1010-1029	37.46-38.14	2055.21861	Shared	2
9		☐	F44	25.36	3.569e6	0.000e0	1	670-679	25.209-25.535	2057.17140	Shared	2
10		☐	F44	20.35	5.146e7	0.000e0	1	535-559	19.904-20.881	2070.18950		1
11		☐	F44	25.44	4.166e7	0.000e0	1	665-690	25.003-25.911	2075.12476	Shared	2
12		☐	F44	55.10	2.843e5	0.000e0	1	1440-1449	54.89-55.281	2077.20110		0
13		☐	F44	35.34	8.595e7	0.000e0	1	945-974	34.889-36.153	2083.26645	Shared	2
14		☐	F44	95.62	2.173e6	0.000e0	1	2361-2379	95.412-96.25	2085.16189	Shared	2
15		☐	F44	15.40	9.111e6	0.000e0	1	405-419	15.137-15.713	2091.99567		0
16		☐	F44	22.87	1.410e6	0.000e0	1	605-614	22.7-23.042	2099.20770		0
17		☐	F44	23.27	5.887e6	0.000e0	1	610-629	22.885-23.648	2105.10250		2

Table 13 lists the Feature Group page columns.

Table 13. Feature Group page columns

Column	Description
File ID	A short identifier of the input file in which the application detected the feature.
Apex RT [min]	The retention time (RT) at which the abundance is greatest for the feature.
Area	The measured peak area for a feature.
Intensity	Displays the peak height of the largest theoretical contained LCMS peak.
# LCMS Features	The number of LC/MS features found within your defined mass and retention time thresholds. This value is usually one. When many feature groups contain more than 2 or 3 LC/MS features, consider tightening the RT and Mass tolerance in the feature detector node. For information on feature detector nodes and their parameters, see Feature Detector nodes.
Scan Range	The scan numbers where the feature was detected.
RT Range	The retention time range when the feature was detected.
Mass (Da)	The detected mass of the feature.
Quan Info	The column indicates if the feature is shared or unique. • *Shared* indicates features mapping to multiple PrSMs • Empty column (indicating unique) indicates that the feature maps to 1 or 0 PrSMs
#PrSMs	The number of PrSMs.

Consensus Feature Group page

Consensus feature groups are a collection of similar (within mass and RT tolerance) feature groups, one from each file, representative of the feature across files. The Consensus Feature Groups page provides a list of all the consensus feature groups found.

Figure 35 shows some of the columns of the Consensus Feature Group page.

Figure 35. Consensus Feature Group page (detail)

Table 14 lists the columns on the Consensus Feature Group page.

Table 14. Consensus Feature Group page columns

Column	Description
Average Apex RT	The average apex retention time of all of the feature groups composing the consensus feature group.
Mass [Da]	The average mass of the feature groups composing the consensus feature groups.
# Files	The number of files where a feature was detected.
Abundance (Normalized)	The normalized abundance of each feature group in the consensus feature group.
Abundances (per File)	The Abundances column reports the measured abundance of each feature group in the consensus feature group.
Quan Info	*Shared* indicates that a consensus feature maps back to two or more proteoforms. This occurs when there are two species of similar masses with similar RT, which can occur when there are two isobaric proteoforms of the same isoform.

Quantitative results columns in the Proteoforms page

When you use a label-free quantitation workflow, quantitative results appear in the Proteoforms Table.

For information on the other columns in the Proteoform page, see Proteoforms page.

Figure 36. Quantitative information columns in the Proteoforms page

Table 15 lists the columns that contain quantitative information. For the other columns, see Table 9.

Table 15. Quantitative information columns in the Proteoforms page

Column	Description
Abundance Ratio	The ratio of the grouped abundances based on requested ratios from the grouping and analysis tab.
Abundance Ratios (log2)	The abundance ratio as a logarithm.
Abundance Ratio Adjusted P-Value	The p-values adjusted by using the Benjamini-Hochberg correction for the false discovery rate. For more information on p-values, refer to Calculating P-Values and Adjusted P-Values for Quantification Results in the *Proteome Discoverer User Guide*.
Abundances Grouped	Displays the abundance values of the sample groups. A grouped abundance value is calculated as the arithmetic mean of all the replicate abundance values within a sample group. This column appears when you group samples in the analysis setup and there is at least one sample group consisting of at least two samples. You can specify the sample grouping on the Grouping and Quantification page when you set up an analysis.
Abundances	The abundance values.

Visualizing top-down results

These topics describe tools for viewing top-down results:

- Use ProSight Lite
- Use TDViewer
- Use TDValidator Lite

Use ProSight Lite

To display the sequence maps for selected proteoforms in the results view, click the **Show ProSightPD Fragment Map** icon, . These sequence maps are annotated with modifications and fragments from experimental MS/MS data.

If the ProSight Lite application is not already installed, install it from the prosightlite.northwestern.edu website.

By clicking PSLite in the External Top-Down Displays column or from the ProSightPD Fragment Map, you can view the fragment map associated with the match in the ProSight Lite application (Figure 37).

Figure 37. Associated fragment map
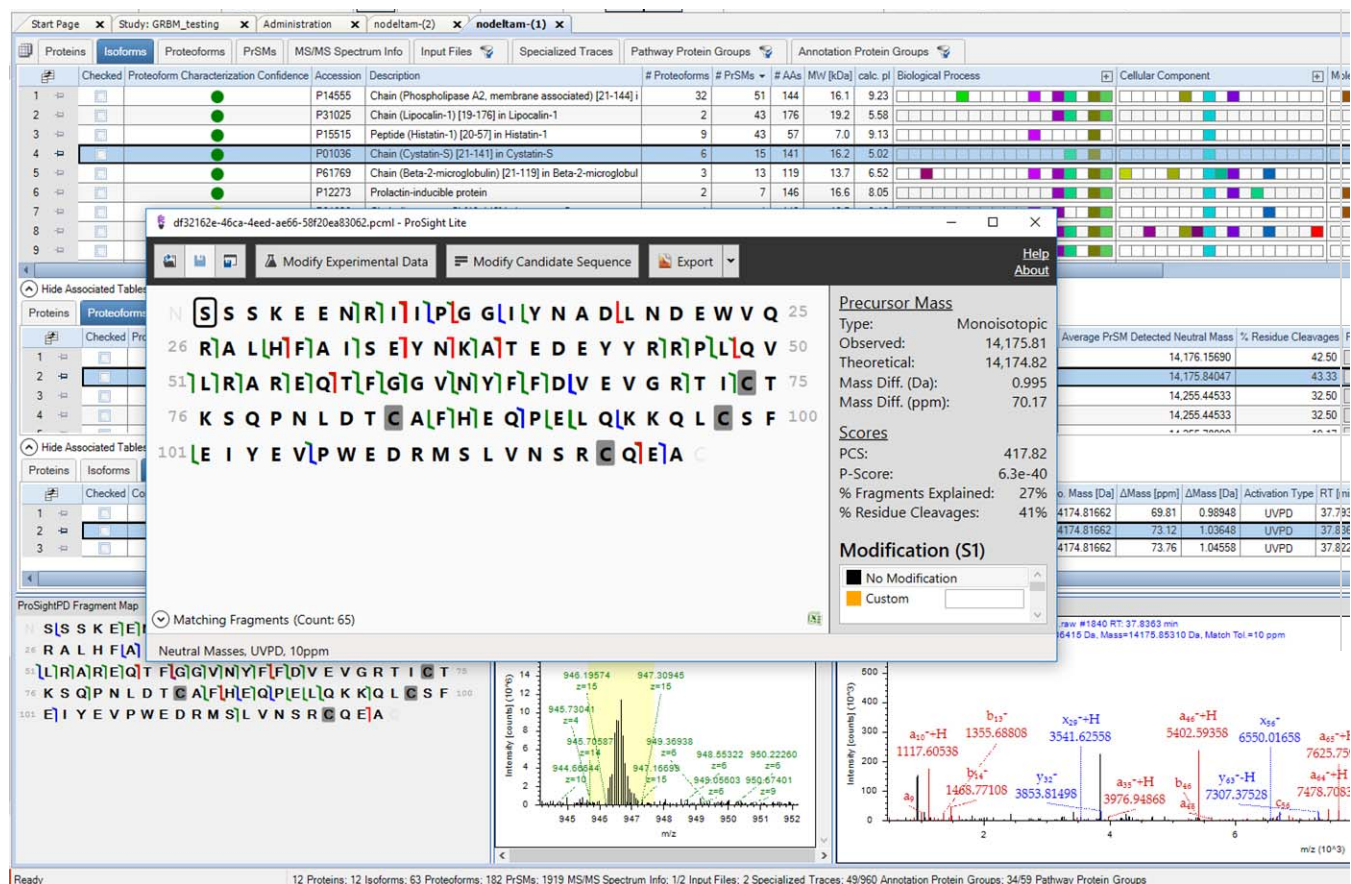
Use TDViewer

ProSightPD nodes are integrated with the TDViewer application. The TDViewer application lets you share Proteome Discoverer results outside of the application.

❖ **To install the TDViewer**

1. To download the TDViewer, navigate to http://psptm.kelleher.northwestern.edu/.

2. Follow the viewer's download and installation instructions.

❖ **To generate a compatible result file report**

Set the ProSightPD FDR node's Generate tdReport parameter to **True**. This node is available in the consensus workflow.

For more information, see FDR node.

Use TDValidator Lite

To launch the TDValidator Lite application, click **TDValidator Lite** in the Proteome Discoverer Results view.

The TDValidator Lite application provides a comprehensive overview of the matching fragment ions for a given proteoform and spectrum. The TDValidator Lite window opens with the associated precursor and fragmentation spectra loaded as well as the proteoform from the results table.

> **Note** If the raw files are not in the location specified in the study, TDValidator Lite will not work.

The TDValidator Lite application generates theoretical isotopic distributions based on the exact chemical formula of the proteoform and corresponding fragment ions. These isotopic distributions are then fit to the experimental data. Only ions relevant to the proteoform of interest are considered. As such, the matching fragment ions in TDValidator Lite can be slightly different than those found in ProSight searches from Xtract-detected fragment ions.

In addition to matching terminal ions of a proteoform, the TDValidator Lite application can be used to look at matching internal fragment ions. Because the number of possible internal fragment ions quickly scales for larger proteoforms, several additional features are included in TDValidator Lite to help users better interpret result output. An FDR feature is now included that generates randomized sequences of the same length as the proteoform. A distribution of matches for the randomized sequences can then be produced to understand how many fragment ions, terminal or internal, match on average to a random protein sequence. Another key part of the TDValidator Lite application is the fitter score. In general, scores greater than 0.7 can be considered fairly robust, but scores all the way down to 0.5 can still be acceptable. When looking at internal fragment ions, a higher score can be necessary to limit false positives, particularly for very dense spectral data or large proteoforms (>30 kDa). Lastly, ppm differences between experimental and theoretical ions can be used to limit the scope of the error differences that are allowed. For example, if terminal ions are generally clustering around 0 ppm mass difference but internal fragment ions are much more widely spaced, limiting the allowable mass differences can reduce the false positive internal fragment ions.

Filter results

Refer to the *Proteome Discoverer User Guide* for information on filtering your results.

Visualize results with charts and graphs

This section lists the visualization formats available for the ProSightPD application.

Histograms
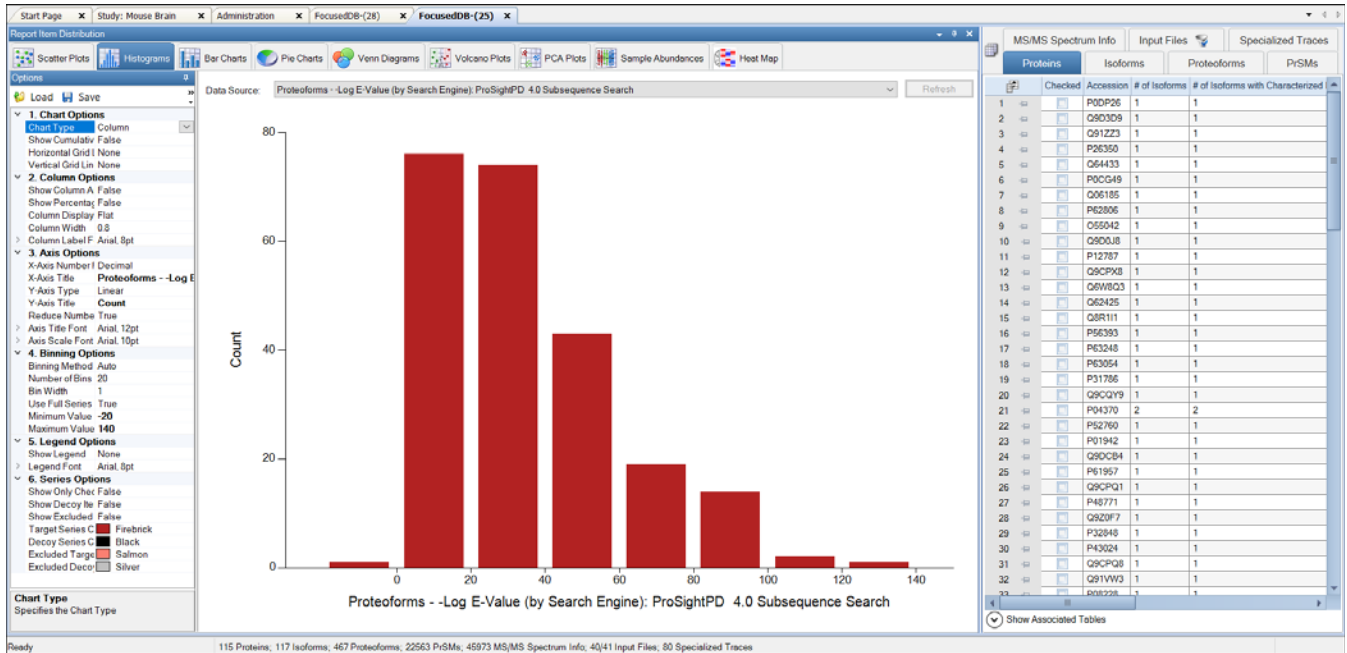
Abundances
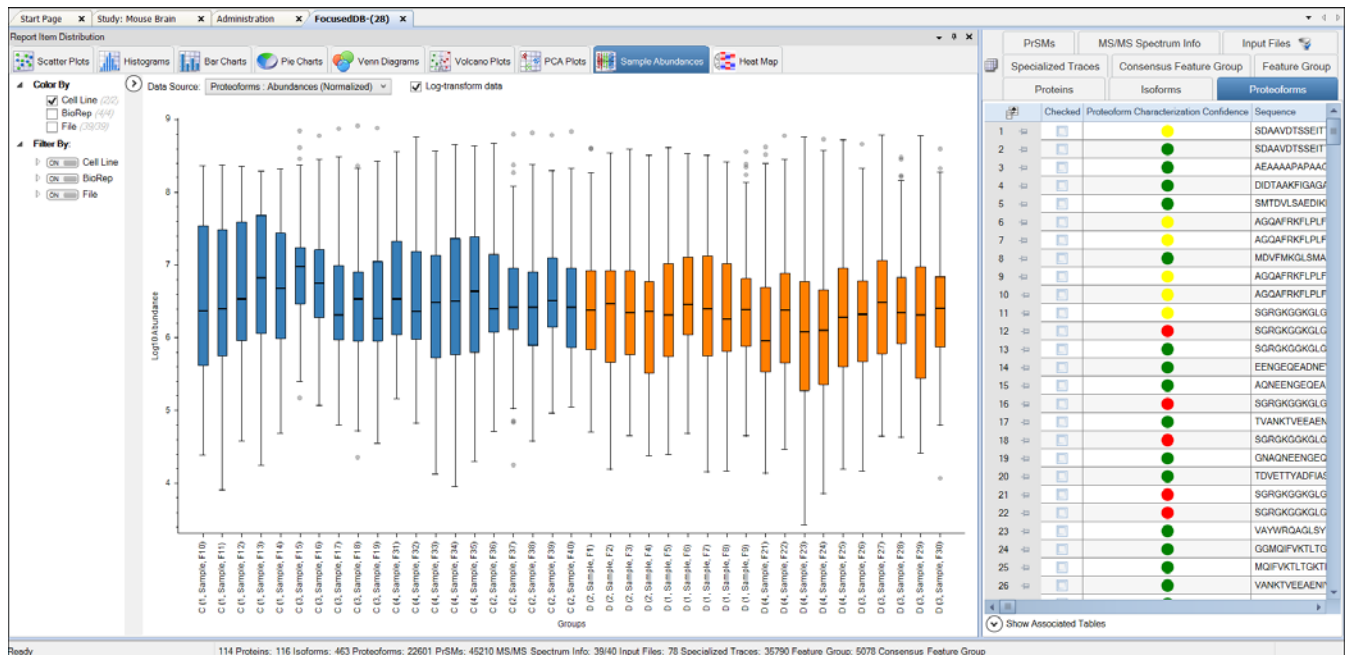
Volcano Plot

Scatter Plot
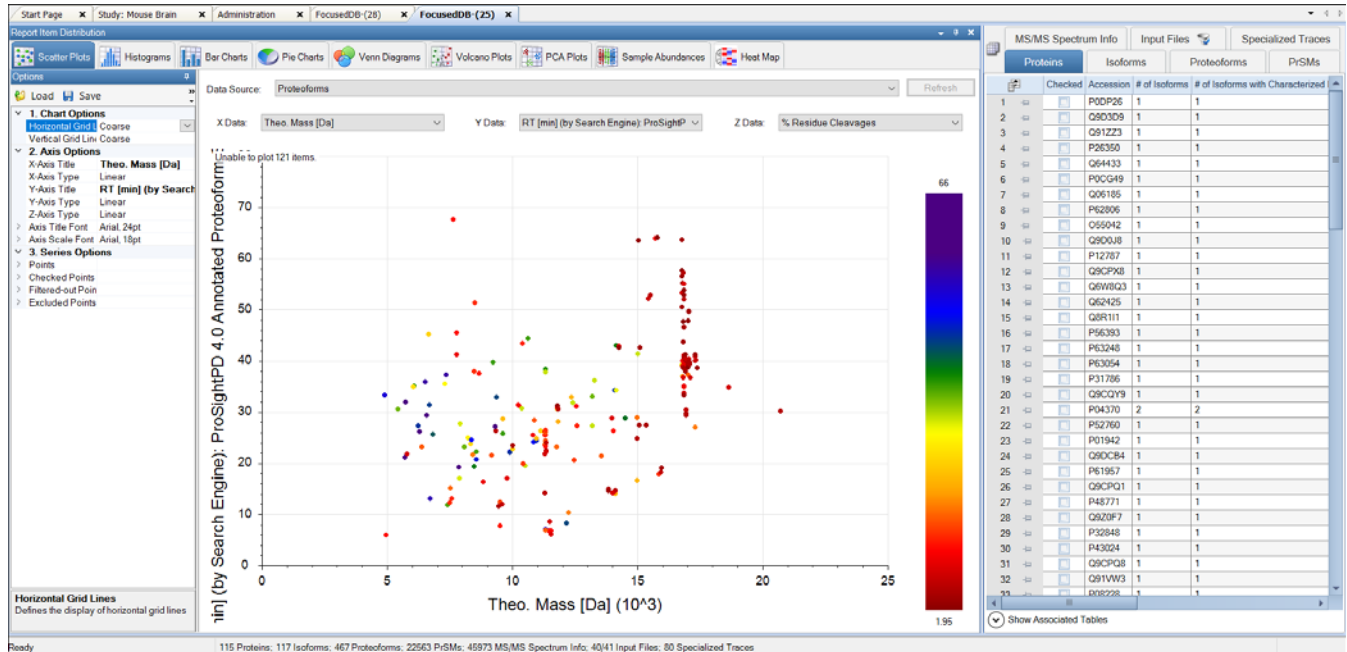
Chromatogram Traces view

The Chromatogram Traces view displays the following:

- The chromatographic profiles of individual features or peaks that the application uses in calculating label-free and precursor ion quantification results

- The exact peak and feature traces that are used to derive quantification values

- The feature traces that are summed up from their corresponding peaks

In addition, you can separately display the single peaks that are subsumed in feature traces to provide a more detailed view. Closely overlapping chromatograms (or traces) in the view indicate whether proteoforms eluted at the same time from the column and therefore are the same proteoform. They confirm the validity of the quantification results and the origin of the consensus features from the same proteoform. Conversely, peaks appearing far apart in the view indicate potential false positives.

The Chromatogram Traces view is available only when you perform label-free quantification or precursor ion quantification, but what it displays in the view depends on the result file page that is open.

Abundance is the area under the curve for each trace. In Figure 38, the Consensus Features page of the table associated with the selected proteoform group shows the abundances displayed in the Chromatogram Traces view.

Figure 38. Chromatogram Traces view
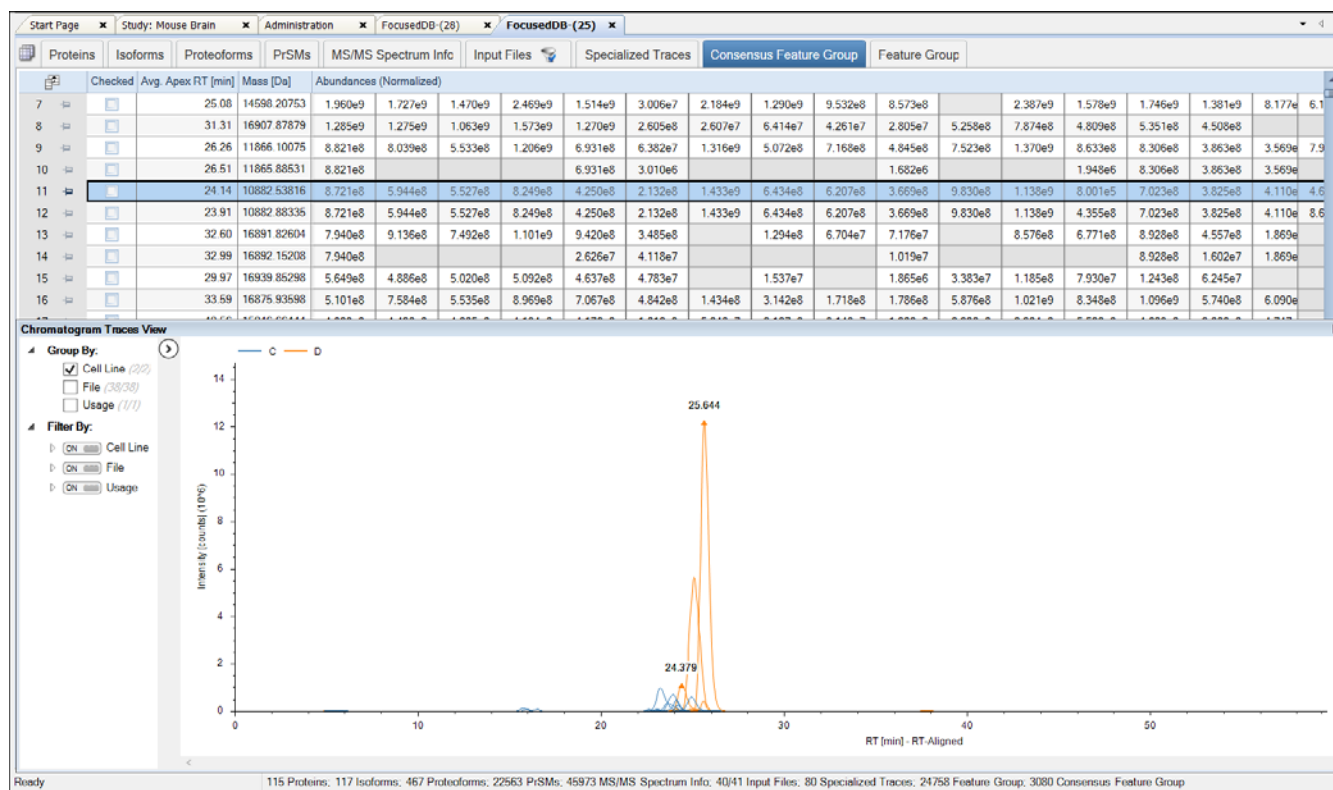
Workflow templates and analysis templates

This chapter provides information about the templates for top-down analysis.

Table 16 lists the analysis and workflow templates for study creation.

Table 16. Analysis Templates, Processing Workflow Templates, and Consensus Workflow Templates (Sheet 1 of 3)

Template Name	Description
Analysis Templates	Includes both processing workflow and consensus workflow.
PSPD 1 percent FDR LFQ for HI HI data.PDANALYSIS	Includes the discovery proteomics search nodes and the nodes required to perform label-free proteoform quantitation on high resolution data.
PSPD 1 percent FDR LFQ for Med HI data.PDANALYSIS	Includes the nodes for discovery proteomics searches and label-free proteoform quantitation on medium high resolution data.
PSPD FASTA Exporter.PDANALYSIS (analysis template)	Enables export of FASTA files from Proteome Discoverer for subsequent ProSight database creation in the ProSightPD Database Manager. Enables you to utilize the FASTA database utilities in PD and export FASTA files not formated in UniProt format.
PSPD PTCR tMS2 Discovery Proteomics with FDR.PDANALYSIS	Supports PTCR data collected as a pseudo MS1. In this method PTCR is used on a small window of the m/z spectrum. The resulting MS2 scan, which is the PTCR scan, is treated as the precursor containing scan. The subsequent MS3 scan is the fragmentation of the PTCR window used to make PrSMs. This workflow uses a special cRAWler node to analyze discovery PTCR data.
PSPD Hi Hi TopDownStandard.PDANALYSIS Analysis Template	Designed to be used with the provided Pierce TopDown Standard database and RAW file.
PSPD Med HI.PDANALYSIS Analysis Template	Built for data with isotopically unresolved precursors. Covers the largest proteoform space including all annotated forms (PTMs, SNPs, and so on) in your database (annotated proteoform search), unknown truncations (subsequence search) and unknown modifications through the wide mass (1000 Da) tolerance annotated proteoform search. This search is the most time consuming. Designed for traditional high throughput LC/MS based workflows. Results are filtered to 1% FDR.

Table 16. Analysis Templates, Processing Workflow Templates, and Consensus Workflow Templates (Sheet 2 of 3)

Template Name	Description
PSPD Discovery Proteomics with FDR.PDANALYSIS	Identifies all proteoforms found in the user's database, including known modifications. This search does not find unknown truncated proteoforms. This template is designed for quick analysis of traditional high throughput LC/MS based workflows. Results are filtered to 1% FDR.
PSPD Comprehensive Discovery Proteomics with FDR.PDANALYSIS	Covers the largest proteoform space including all annotated forms (PTMs, SNPs, and so on) in your database (annotated proteoform search), unknown truncations (subsequence search), and unknown modifications through the wide mass (1000 Da) tolerance annotated proteoform search. This search is the most time consuming. Designed for traditional high throughput LC/MS based workflows. Results are filtered to 1% FDR.
PSPD Med HI Analysis Template.PDANALYSIS	Intended for data with non-isotopically resolved precursors. Typically, this search template is used for lower resolution precursor scans or when analyzing larger or highly charged proteoforms. Results are filtered to 1% FDR.
PSPD Truncation Subsequence Search with FDR.PDANALYSIS	Searches for truncated sequences, including those with any known PTMs or modifications.
PSPD Single Proteoform.PDANALYSIS	Searches for a single user-defined proteoform, which does not require a database. This workflow is suitable for identifying one species in a complex LC/MS experiment or to confirm the identify of a single species from a mixture of infused proteoforms.
Processing Workflow Templates	
PSPD MS2 Precursor with MS3 Fragmentation HI HI.pdProcessingWF	Supports analysis of "complex-down" data. The initial fragmentation event generates monomers, which are targeted for MS3 fragmentation. The sequence information is derived from the MS3 data (PrSMs come from MS3) This template supports isotopically resolved species (monomer isotopes are resolved).
PSPD Comprehensive Discovery Proteomics with FDR for Unresolved Precursors.pdProcessingWF	ProSightPD processing workflow for unresolved precursor spectra. This workflow includes: Narrow Tolerance Annotated Proteoform Search, Wide Tolerance Annotated Proteoform Search, and Subsequence searches. These searches combined create a comprehensive search strategy.
PSPD PTCR Discovery Proteomics with FDR.pdProcessingWF	Identifies larger proteoforms from PTCR data.
PSPD Discovery Proteomics with FDR for HI HI data.pdProcessingWF	Performs searches for high-resolution data.

Table 16. Analysis Templates, Processing Workflow Templates, and Consensus Workflow Templates (Sheet 3 of 3)

Template Name	Description
PSPD Discovery Proteomics with FDR for Unresolved Med HI data.pdProcessingWF	Performs proteoform identification on data containing unresolved precursors. Must be used with FDR nodes in consensus workflow.
PSPD LFQ for HI HI data.pdProcessingWF	Performs searches for high resolution data with LFQ analysis. Must be used with LFQ nodes in consensus workflow for LFQ results.
PSPD LFQ for Med HI data.pdProcessingWF	Performs searches for medium high data with LFQ analysis. For label-free results, use with LFQ nodes in the consensus workflow.
PSPD Comprehensive Discovery Proteomics with FDR for HI HI data.pdProcessingWF	Performs both annotated and subsequence searches for high resolution data (resolved precursors). This parameter must be used with FDR nodes in consensus workflow.
PSPD Single Proteoform Search.pdProcessingWF	Performs single user-defined proteoform searches.
PSPD Truncation Search with FDR.pdProcessingWF	Performs subsequence search only for high resolution data. Must be used with FDR nodes in consensus workflow.
Consensus Workflow Templates	
PSPD 1 percent FDR Consensus.pdConsensusWF	Returns results filtered to 1% FDR.
PSPD 1 percent FDR LFQ for HI HI with PFR Annotation data.pdConsensusWF	Returns results filtered to 1%FDR includes PFR annotation.
PSPD 1 percent FDR LFQ for HI HI data.pdConsensusWF	Returns LFQ data with 1% FDR.
PSPD No FDR Consensus.pdConsensusWF	Returns results without FDR control.
PSPD 1 percent FDR LFQ for Unresolved Precursors data.pdConsensusWF	Returns Med/Hi LFQ data with 1% FDR.

ProSightPD nodes

This chapter provides information about ProSightPD nodes that are specific to top-down analysis.

Contents

Server Utilities node

The following topic describes the server utilities node.

FASTA Exporter node

Table 17. ProSightPD FASTA Exporter node parameters

Parameter	Description
Input Data	
Protein Database	Select the name of the database to export.
Output File Name	Type the name for the exported database.

Feature Detector nodes

The following topics describe the feature detection nodes:

- Med Res. Feature Detector

- Hi Res. Feature Detector

Med Res. Feature Detector

The Med Res. Feature Detector node uses the sliding window and the kDecon algorithms to perform spectral deconvolution and measure all of the deconvoluted features and their quantitation traces. The parameters used in this node are essential to the label-free quantitation workflows. Set the parameters according to the data being analyzed.

The Sliding Window Averaging Width RT parameter establishes the RT for the sliding window average. This value should match the full width at half max of a representative peak from the analyzed data. Thermo Fisher Scientific recommends measuring the peak width of an average feature in your data set. Do not choose the most abundant peak or one that is at noise level.

Table 18 lists the node parameters.

Table 18. (Sheet 1 of 4) ProSightPD Med Res. Feature Detector node

Parameter	Description
1. kDecon	
Precursor Maximum Mass	Specifies the maximum precursor mass to be considered by kDecon.
Precursor Minimum Mass	Specifies the minimum precursor mass to be considered by kDecon.
Number of Results	The maximum number of kDecon results to return for each averaged scan. Results are filtered by S/N prior to being returned.
Number of Iterations	kDecon can be run multiple times. After each iteration, charge state distributions from each mass are removed from the spectrum to limit false positives during subsequent iterations.
S/N Result Cutoff	The geometric mean signal-to-noise ratio S/N is calculated from the individual S/N of each charge state in the charge state distribution of a detected mass.
Peak Tolerance in PPM	Peak Tolerance PPM
2. Sliding Window Parameters	

Table 18. (Sheet 2 of 4) ProSightPD Med Res. Feature Detector node

Parameter	Description
Averaging Width RT	Specifies the retention time, or the width, of the sliding window, in minutes. Reducing this value improves time resolution but reduces execution speed and possibly sensitivity. Increasing this value increases execution speed but reduces time resolution and might increase sensitivity. You can achieve the best results when the window width is between one-quarter and twice the width of the characteristic peaks in the spectrum. For most usage, the optimum value might be half the width of the characteristic peaks. For example, if those peaks have a width of one minute, the optimum width would be 0.5 minutes.
Offset Type	Specifies the offset between successive sliding windows as a number of scans or as a percentage value. You can select from these two options: scan-based offset or percentage of averaging width RT offset.
Offset Scan(s)	This mode offsets each window from its predecessor by the user-specified number of scans. An offset of n means that each window begins n scans after the beginning of its predecessor.
Offset Percentage	This mode offsets each window from its predecessor by the user-specified percentage of the window width. An offset of 30% means that each window begins 30% after the beginning of its predecessor and overlaps the last 70%. An offset of 100% means that successive windows are adjacent with no overlap.
Merge Tolerance (ppm)	Determines how close two components in successive sliding windows must be in mass (ppm) for the application to identify them as a single component. A value that is too high might result in merging components that should remain separate. A value that is too low might result in false positives when components remain separate that should have merged.
Minimum Number of Charge States Observed	Minimum required number of observed charge states for data to be analyzed by KDecon. For top-down data, the recommended minimum is 3. Some very low-abundance species might only have 2 observable (greater than noise) peaks. However, setting this value to less than 3 will allow noise through and result in false positives.

Table 18. (Sheet 3 of 4) ProSightPD Med Res. Feature Detector node

Parameter	Description
Minimum Number of Sliding Window Detections	Specifies the minimum number of sliding window intervals that a component must appear in for the application to consider the component valid. A value that is too low might allow noise peaks to appear as false positives. A value that is too high might result in legitimate components being discarded. Set this parameter to a value large enough to exclude results that are implausibly narrow in retention time but small enough to include results of realistic duration. Values in the range of 3 through 8 generally give good results. A good approach is to use whichever is larger: 3 or the minimum number of windows that can fit into a peak.
Biggest Gap Type	Specifies the allowed separation between successive individual members of a merged component identified by the Sliding Window Deconvolution algorithm. You can select from these two options: Scans or Retention Time.
Biggest Gap Scans	The maximum allowed separation in scans between two successive individual members of a merged component identified by the Sliding Window Deconvolution algorithm. If the separation exceeds this value, the component will be divided into two components separated. As with the Merge Tolerance, too high a value can result in components that should have remained distinct being merged, while too low a value can result in false positives when components that should have been merged remain distinct. This parameter should be comparable too or slightly less than the expected separation in scans between distinct components with the same mass.
Biggest Gap Retention Time	The maximum allowed separation in retention time between two successive individual members of a merged component identified by the Sliding Window Deconvolution algorithm. If the separation exceeds this value, the component will be divided into two components separated by the observed gap in retention time. As with the Merge Tolerance, too high a value can result in components that should have remained distinct being merged, while too low a value can result in false positives when components that should have been merged remain distinct. This parameter should be comparable too or slightly less than the expected separation in retention time between distinct components with the same mass.
3. Feature Grouping	
Mass Tolerance	Mass tolerance used for calculating mass traces.
RT Threshold	Retention time limits used for calculating features.

Table 18. (Sheet 4 of 4) ProSightPD Med Res. Feature Detector node

Parameter	Description
4. Feature/Trace Connection	
Trace Mass Tolerance	Specifies the trace tolerance.
Number of Smoothing Points	Specifies the number of points to average. As the number increases, the data become smoother resulting in a loss of fine features. The recommended number of points is 3.
Time Range (min)	The allowable time difference between a trace and its linked feature. The recommended setting is 1 minute.
Trace Smoothing Type	Specifies the type of smoothing to perform: Gaussian, Moving Average, or None. The recommended setting is None. This should be used only in cases where unsmoothed data is not acceptable.
5. Multithreading Options	
CPU Usage	Amount of CPU to direct towards data processing.

Hi Res. Feature Detector

The Hi Res. Feature Detector node uses the sliding window and Xtract algorithms to perform spectral deconvolution and measure all of the deconvoluted features and their quantitation traces. The parameters used in this node are essential to the LFQ workflow. Set them according to the data you are analyzing.

One of the key parameters is the Sliding Window Averaging Width RT. This parameter establishes the RT for the sliding window average. Set this parameter to match the full width at half max of a representative peak from the analyzed data. It is recommended to measure the peak width of an average feature in your data set. Do not choose the most abundant peak or one that is at noise level.

The best parameter setting is based on your chromatography:

- If the window is too wide, then low intensity signals are averaged out of the data.

- If the window is too narrow, then you lose the advantages of signal averaging.

Table 19 lists the node parameters.

Table 19. ProSightPD Hi Res. Feature Detector parameters (Sheet 1 of 7)

Parameter	Description
1. Xtract	
S/N Threshold	Specifies the minimum signal-to-noise ratio for data that the node analyses.

Table 19. ProSightPD Hi Res. Feature Detector parameters (Sheet 2 of 7)

Parameter	Description
Lowest m/z	Specifies the portion of the input spectrum that the Xtract algorithm processes. Min: Specifies the lowest end of the input spectrum.
Highest m/z	Specifies the portion of the input spectrum that the Xtract algorithm processes. Max: Specifies the highest end of the input spectrum.
Min Precursor Mass	The minimum precursor mass that the Xtract algorithm considers.
Max Precursor Mass	The maximum precursor mass that the Xtract algorithm considers.
Lowest Charge	Sets the low end of the allowable range for the number of charge states that must appear for a component to be recognized. The Xtract algorithm rejects potential components with fewer than the minimum or greater than the maximum number of charge states.
Highest Charge	Sets the high end of the allowable range.
Minimum Number of Detected Charge States	Minimum required number of observed charge states for data to be analyzed by the Xtract algorithm. For top-down data, the recommended minimum is 3. Some very low-abundance species might have only 2 observable (greater than noise) peaks. However, setting this value to less than 3 lets noise through and results in false positives.
Relative Abundance Threshold	Specifies a threshold below which the node filters out data for data reporting. This option sets a relative threshold as a percentage of the most abundant component in the spectrum. The most abundant peak in the deconvolved spectrum has a relative abundance of 100 percent, and all other peaks are calculated relative to that one. For example, if the highest peak has an absolute abundance of 1000, the relative abundance is 1 percent, and no peaks below an absolute abundance of 10 appear in the deconvolved spectrum. Range: 0–100 Zero (0) displays all results; 100 displays only the most abundant component.
Resolution at m/z 400	Defines the resolution of the source spectrum at an *m/z* value of 400.

Table 19. ProSightPD Hi Res. Feature Detector parameters (Sheet 3 of 7)

Parameter	Description
Fit Factor	Measures the quality of the match between a measured isotope pattern and an averagine distribution of the same mass. Enter a value between 0 and 100% • 0% requires a low fit only. • 100% means that the measured isotope profile is identical to the theoretical averagine isotope distribution.
Remainder Threshold (%)	Specifies the height of the smaller overlapping isotopic cluster, as a percentage, with respect to the height of the most abundant isotopic cluster when the Xtract algorithm attempts to resolve overlapping isotopic clusters. For example, if one isotopic cluster in a spectrum has an abundance of 100, and you set the Remainder Threshold parameter to 30%, the Xtract algorithm ignores any overlapping clusters with an abundance less than 30.
Min. Intensity	Specifies a minimum intensity threshold to filter out possible background noise, including when you set the S/N Threshold parameter to zero.
Expected Intensity Err.	Specifies the permissible percentage of error allowed in calculating the ratio of the most abundant isotope to the next isotope higher in mass in the isotope series.
Consider Overlaps	When selected (default), indicates that the Xtract algorithm is more tolerant of errors when the spectrum intensity is significantly higher than expected for the theoretical isotopic cluster. This option can lead to increased false positives; select it only when you expect overlapping isotopic clusters in a data set.

Table 19. ProSightPD Hi Res. Feature Detector parameters (Sheet 4 of 7)

Parameter	Description
2. Sliding Window Parameters	
Averaging Width RT	Specifies the retention time, or the width, of the sliding window, in minutes. • Reducing this value improves time resolution but reduces execution speed and possibly sensitivity. • Increasing this value increases execution speed but reduces time resolution and might increase sensitivity. You can achieve the best results when the window width is between one-quarter and twice the width of the characteristic peaks in the spectrum. For most uses, the optimum value might be half the width of the characteristic peaks. For example, if those peaks have a width of one minute, the optimum width is 0.5 minutes.
Offset Type	Specifies the offset between successive sliding windows as a number of scans or as a percentage value. Select from: • Scan—Uses the Offset Scan setting • Percent—Uses the Offset Percentage setting
Offset Scan	Offsets each window from its predecessor by the user-specified number of scans. An offset of n means that each window begins n scans after the beginning of its predecessor.
Offset Percentage	Offsets each window from its predecessor by the user-specified percentage of the window width. • An offset of 30% causes each window to begin 30% after the beginning of its predecessor and overlap the last 70%. • An offset of 100% causes successive windows to be adjacent with no overlap.
Merge Tolerance	Determines how close two components in successive sliding windows must be in mass for the application to identify them as a single component. • A value that is too high might result in merging components that should remain separate. • A value that is too low might result in false positives when components remain separate that should have merged.

Table 19. ProSightPD Hi Res. Feature Detector parameters (Sheet 5 of 7)

Parameter	Description
Merge Mode	Determines how spectra are merged: • Batch—Merges spectra as a batch, which minimizes the effect of scan to scan variation (Recommended). • Incremental—Merges spectra incrementally, one-by-one, making the result sensitive to scan-by-scan variations for long eluting species (Legacy).
Minimum Number of Charge States Observed	Minimum required number of observed charge states for data to be analyzed by kDecon. For top-down data, the recommended minimum is 3. Some very low-abundance species might have only 2 observable (greater than noise) peaks. However, setting this value less than 3 lets noise through and result in false positives.
Minimum Number of Sliding Window Detections	Specifies the minimum number of sliding window intervals that a component must appear in for the application to consider the component valid. Set this parameter to a value large enough to exclude results that are implausibly narrow in retention time but small enough to include results of realistic duration. Values in the range of 3 through 8 generally give good results. A good approach is to use whichever is larger: 3 or the minimum number of windows that can fit into a peak. • A value that is too low might allow noise peaks to appear as false positives. • A value that is too high might result in legitimate components being discarded.
Biggest Gap Type	Specifies the maximum allowed separation in retention time between two successive individual members of a merged component identified by the sliding windows algorithm, in minutes. • Scans • Retention time

Table 19. ProSightPD Hi Res. Feature Detector parameters (Sheet 6 of 7)

Parameter	Description
Biggest Gap Scans	Specifies the greatest number of scans between subsequent feature detection for grouping similar features. - If the separation exceeds this value, the application divides the candidate component into two merged components separated by a gap in retention time. A value that is too high for this parameter might result in merging components that should remain separate. - A value that is too low might result in false positives when components remain separate that should have merged. Set this parameter comparable to or slightly less than the expected separation in retention time between distinct components with the same mass.
Biggest Gap Retention Time	Specifies the greatest amount of time between feature detection events for grouping similar features.
3. Feature /PrSM Connection	
Mass Tolerance	Delta mass (in Daltons) allowed for mapping PrSMs to features.
RT Threshold	Retention time threshold for mapping PrSMs to features: - Minimum value = 0 - Maximum value = (unchecked)
4. Feature/Trace Connection	
Trace Mass Tolerance	Specifies the trace tolerance. - Minimum value = 0.01 Da 0.1 ppm - Maximum value = (unchecked)
Number Of Smoothing Points	- Minimum value = 0 - Maximum value = (unchecked)
Time Range (min)	- Minimum value = 0 - Maximum value = (unchecked)
Trace Smoothing Type	Determines the method to smooth the traces: - Gauss - MovingMean (Default) - None

Table 19. ProSightPD Hi Res. Feature Detector parameters (Sheet 7 of 7)

Parameter	Description
5. Multithreading Options	
CPU Usage	Determines the amount of CPU to direct toward processing: • High—All available threads less 1 (Default) • Medium—50% of the available threads • Low—1 thread

Search nodes

This section describes the nodes that search the assigned database.

> **Note** The ProSightPD 4.2 application includes the legacy search nodes from ProSightPD 3.0 so that you can reproduce previous results in the ProSightPD 4.2 application. These legacy nodes require a database in the .pscw file format generated in ProSightPC. You cannot directly convert .pscw to .psdb.
> Use the source .xml file, which the .pscw file was generated from, to generate a .psdb-formatted database. To use a legacy search node, replace the 4.2 node with the equivalent legacy node and review all parameters to ensure agreement between your previous ProSightPD 3.0 workflows and your ProSightPD 4.2 workflows.

- Annotated Proteoform Search node

- Single Proteoform Search node

- Subsequence Search node

- Tag to Annotated Proteoform Search node

- Tag to Subsequence Search

Annotated Proteoform Search node

The Annotated Proteoform Search node performs a ProSight-annotated proteoform search.

Table 20 lists the parameters for the 4.2 version of the node. For information about the 3.0 version, refer to the user interface.

Table 20. ProSightPD 4.2 Annotated Proteoform Search node parameters (Sheet 1 of 4)

Parameter	Description
1. Input Data	
Input Database	Select the name of the database to search.
2. Annotated Proteoform (Absolute Mass) Search	
Precursor Mass Tolerance	Specifies the mass tolerance in Da or ppm used to find peak candidates. Range: 0.0–1 000 000 Da; 0.1–1000 ppm
Fragment Mass Tolerance	Specifies the mass tolerance in Da or ppm used for matching fragment peaks. Range: 0–2.0 Da; 1–50 ppm

Table 20. ProSightPD 4.2 Annotated Proteoform Search node parameters (Sheet 2 of 4)

Parameter	Description
Candidate Source	Specify particular proteoforms to consider for your search. • Isoforms: Only searches isoforms. • Proteoforms: Only searches proteoforms. • Both - Exclude Isoforms with Proteoforms: Searches both isoforms and proteoforms, but excludes isoforms with associated proteoforms. • Both - Include Isoforms with Proteoforms: Searches both isoforms and proteoforms, and includes isoforms with associated proteoforms.
Maximum PTMs per Isoform	Limits the number of PTMs to add to any given proteoform. If more than this number of PTMs are present on an isoform, the proteoforms created are limited to the ones with combinations of PTMs having this many PTMS or fewer.
Maximum SNPs per Isoform	Limits the number of SNPs to add to any given proteoform. If more than this number of SNPs are present on an isoform, the proteoforms created are limited to the ones with combinations of SNPs having this many SNPs or fewer.
Maximum PTMs per Proteoform	Limits the number of PTMs to add to any given proteoform. If more than this number of PTMs are present on an isoform, the proteoforms created are limited to the ones with combinations of PTMs having this many PTMS or fewer.
Remove Disulfide Bonds	Thermo Fisher Scientific recommends setting this parameter to False. You can set this parameter to True if there are annotated disulfide bonds in the search database that you want to avoid searching, or when there is a known error in the disulfide bond annotation in the database. Errors in UniProt where disulfide bonds are placed on non-cysteine residues have been observed and can cause the search to crash.
Maximum Mass to Include PTMs	Maximum proteoform mass to search with PTMs.
N-Term Modifications to Include	Selected modifications are included on all intact proteoforms n-termini.
Delta M Mode	Determines whether the application conducts the search in Delta M (Δm) Mode. For more information on this mode, see Delta M Mode searches.

Table 20. ProSightPD 4.2 Annotated Proteoform Search node parameters (Sheet 3 of 4)

Parameter	Description
Remove Redundant Delta M Hits	This parameter controls the results returned from a search that were performed with the Delta M mode set to True. When the Remove Redundant Delta M Hits is set to True, results that nullify a proteoform are removed. For example, if the parameter is set to True and there is an N-terminally acetylated result and a Delta M mode result with Delta M mass (-42) at the N-terminus, the Delta M mode hit will not be shown in the result because it removes a modification which was found through one of the database searches.
Delta M Mod Tolerance	Defines the tolerance when considering fragments with Delta M modifications.
Calculate FDR	Specifies whether the FDR is calculated. You might see fewer results passing the FDR threshold when you use Delta M Mode.
Decoy Reps	The number of times a decoy search is run. A higher number of repetitions produces more stable decoy results but increases search times.
Maximum PrSMs per Precursor	Maximum number of PrSMs matched per precursor limits the number of results to the highest scoring hits.
Minimum Matched Fragments	Return only hits with this many or more matched fragments.
3. Static (Fixed) Modifications	
Static (Fixed) Modification	Specifies a static (also known as fixed) modification from a list of known modifications. Static modifications apply universally to every instance of the specified residues.
4. Multithreading Options	
CPU Usage	The amount of CPU to direct to data processing: • High—The total number of CPU cores less 1 • Medium—Half the number of available cores • Low—1 core
5. Fragmentation Method	
Override Fragmentation Method	Manually set the fragmentation method. Auto: allows the software to determine the correct fragmentation type (Default).

Table 20. ProSightPD 4.2 Annotated Proteoform Search node parameters (Sheet 4 of 4)

Parameter	Description
6. UVPD Method	
Default UVPD Method	Set the number of ion types to consider for the score of UVPD fragment ion spectra: • 4—recommended for maximizing IDs • 9—recommended for maximizing characterization

Single Proteoform Search node

The Single Proteoform Search node performs proteoform searches against a single proteoform entry. This node also works for data without precursor scans (for example, direct infusion data). To run a search, the application requires a protein sequence and the appropriate modification ID. For more information, see Table 6. The sequence format is described in Table 21.

For an example of using this node in a workflow, see Perform Single Proteoform searches in the processing workflow.

Table 21 lists the parameters for the 4.2 version of the node.

Table 21. ProSightPD 4.2 Single Proteoform Search parameters (Sheet 1 of 3)

Parameter	Description
1. Input Data	
Proforma sequence	Input a Profroma sequence by inputting the single character abbreviation for each residue. You can add modifications to a residue by putting a modification ID in brackets before the residue to modify. See Table 6 for compatible databases. For example, a human Histone H3 proteoform with its N-terminal methionine removed modified by a K4 methylation and K36 trimethylation is input as: ARTK[AA0076]QTARKSTGGKAPRKQLATKAARKSAPATGGV K[AA0074]KPHRYRPGTVALREIRRYQKSTELLIRKLPFQRL VREIAQDFKTDLRFQSSAVMALQEACEAYLVGLFEDTNLC AIHAKRVTIMPKDIQLARRIRGERA
Include N-Term Acetylation	Applies N-term acetylation to the RESID sequence: • True • False
Proteoform Accession	Input an identifier that will appear in the pdResult file.

Table 21. ProSightPD 4.2 Single Proteoform Search parameters (Sheet 2 of 3)

Parameter	Description
Proteoform Description	Input a description that will appear in the pdResult file.
2. Single Proteoform Search	
Precursor Mass Tolerance	Specifies the mass tolerance in Da or ppm used to find peak candidates. Range: 0.0–1 000 000 Da; 0.1–1000 ppm
Fragment Mass Tolerance	Specifies the maximum mass difference for an observed fragment mass to be considered a match to a theoretical fragment mass: • Minimum value = 1 ppm • Maximum value = 2 Da |50 ppm
Delta M Mode	Determines whether the application conducts the search in Delta M (Δm) Mode. For more information on this mode, see Delta M Mode searches.
Static (Fixed) Modification	Specifies a static (also known as fixed) modification from a list of known modifications. Static modifications apply universally to every instance of the specified residues.
Maximum PrSMs per Precursor	Maximum number of PrSMs matched per precursor limits the number of results to the highest scoring hits.
Minimum Matched Fragments	Return only hits with this many or more matched fragments.
Include Labile Modification Mass Shift	Target mass will be shifted to account for loss of labile modification mass.
Labile Modification Mass Shift	Comma separated list of Target mass shifts.
3. Static (Fixed) Modifications	
Static (Fixed) Modification	Specifies a static (also known as fixed) modification from a list of known modifications. Static modifications apply universally to every instance of the specified residues.
4. Multithreading Options	
CPU Usage	The amount of CPU to direct to data processing: • High—The total number of CPU cores less 1 • Medium—Half the number of available cores • Low—1 core
5. Fragmentation Method	

Table 21. ProSightPD 4.2 Single Proteoform Search parameters (Sheet 3 of 3)

Parameter	Description
Override Fragmentation Method	Manually set the fragmentation method. Auto: allows the software to determine the correct fragmentation type (Default).
6. UVPD Method	
Default UVPD Method	Set the number of ion types to consider for the score of UVPD fragment ion spectra: • 4—recommended for maximizing IDs • 9—recommended for maximizing characterization

Subsequence Search node

The Subsequence Search node performs a subsequence search. Increasing the precursor tolerance results in longer run times.

To search for modified proteoforms, select the Include Modified Forms check box in the Predefined Search dialog box for biomarkers.

Table 22 lists the parameters for the 4.2 version of the node. For information about the 3.0 version, refer to the user interface.

Table 22. ProSightPD 4.2 Subsequence Search parameters (Sheet 1 of 3)

Parameter	Description
1. Input Data	
Input Database	Specifies the name of the database to search.
2. Subsequence (Biomarker) Search	
Subsequence Precursor Mass Tolerance	Specifies the maximum mass difference between the subsequence mass and the observed proteoform mass for a proteoform to be included in the search. Setting this to a large tolerance can potentially increase search time with large databases. Minimum value = 0.01 Da (10.1 ppm) Maximum value = 1 000 000 Da (1000 ppm)
Fragment Mass Tolerance	Specifies the mass tolerance in Da or ppm used for matching fragment peaks. Range: 0–2.0 Da; 1–50 ppm

Table 22. ProSightPD 4.2 Subsequence Search parameters (Sheet 2 of 3)

Parameter	Description
Maximum PTMs per Isoform	Limits the number PTMs to add to any given proteoform. If more than this number of PTMs are present on an isoform, the proteoforms created are limited to the ones with combinations of PTMs having this many PTMS or fewer.
Maximum SNPs per Isoform	Limits the number SNPs to add to any given proteoform. If more than this number of SNPs are present on an isoform, the proteoforms created are limited to the ones with combinations of SNPs having this many SNPs or fewer.
Remove Disulfide Bonds	Thermo Fisher Scientific recommends setting this parameter to False. You can set this parameter to True if there are annotated disulfide bonds in the search database that you want to avoid searching, or when there is a known error in the disulfide bond annotation in the database. Errors in UniProt where disulfide bonds are placed on non-cysteine residues have been observed and can cause the search to crash.
Maximum Mass to Include PTMs	Maximum proteoform mass to search with PTMs.
N-Term Modifications to Include	Selected modifications are included on all intact proteoforms n-termini.
Calculate FDR	Specifies whether the FDR is calculated. You might see fewer results passing the FDR threshold if you use Delta M Mode.
Decoy Reps	The number of times a decoy search is run. A higher number of repetitions produces more stable decoy results but increases search times.
Maximum PrSMs per Precursor	Maximum number of PrSMs matched per precursor limits the number of results to the highest scoring hits.
Minimum Matched Fragments	Return only hits with this many or more matched fragments.
Include Labile Modification Mass Shift	Target mass will be shifted to account for loss of labile modification mass.
Labile Modification Mass Shift	Comma separated list of Target mass shifts.

Table 22. ProSightPD 4.2 Subsequence Search parameters (Sheet 3 of 3)

Parameter	Description
Number of Off by 1	Specifies the number of off-by-one errors to accommodate. For example, if set to 1, the search considers three candidate precursor masses. • deconvoluted precursor mass • deconvoluted precursor mass plus 1.003354838 Da • deconvoluted precursor mass minus 1.003354838 Da Thermo Fisher Scientific recommends setting this parameter to 2.
3. Static (Fixed) Modifications	
Static (Fixed) Modification	Specifies a static (also known as fixed) modification from a list of known modifications. Static modifications apply universally to every instance of the specified residues.
4. Multithreading Options	
CPU Usage	The amount of CPU to direct to data processing: • High—The total number of CPU cores less 1 • Medium—Half the number of available cores • Low—1 core
5. Fragmentation Method	
Override Fragmentation Method	Manually set the fragmentation method. Auto: allows the software to determine the correct fragmentation type. (Default)
6. UVPD Method	
Default UVPD Method	Set the number of ion types to consider for the score of UVPD fragment ion spectra: • 4—recommended for maximizing IDs • 9—recommended for maximizing characterization

Tag to Annotated Proteoform Search node

The Tag to Annotated Proteoform Search node performs a sequence tag search first and then sends all the results from the sequence tag search to an annotated proteoforms search. The sequence tag search reduces the amount of spectra searched by the annotated proteoform search step and reduces the overall search time for large data.

Table 23 lists the parameters for the 4.2 version of the node. For information about the 3.0 version, refer to the user interface.

Table 23. ProSightPD 4.2 Tag to Annotate Proteoform Search parameters (Sheet 1 of 2)

Parameter	Description
1. Input Data	
Input Database	Specifies the name of the database to search.
2a. Sequence Tag	
Minimum Tag Score	Minimum probability that the probable tag is not found by chance.
Amino Acid Gap	Largest number allowed of sequential amino acids not having a fragment between them.
Max Tags	Greatest number of probable tags considered for each search per target.
Minimum Fragments Used	Minimum number of fragments required for each sequence tag.
Remove Water Losses	When set to True, water loss is not considered when searching sequence tags.
Consider Isobaric Dipeptide Gaps	When set to True, for example, both GA and Q are considered for gaps of 128.0 Da.
Compiler Tolerance	The permissible error, measured in ppm between two fragment ion masses that are still considered as matching an amino acid
2b. Annotated Proteoform (Absolute Mass) Search	
Fragment Mass Tolerance	Specifies the mass tolerance in Da or ppm used for matching fragment peaks. Range: 0–2.0 Da; 1–50 ppm
Maximum PTMs per Isoform	Limits the number PTMs to add to any given proteoform. If more than this number of PTMs are present on an isoform, the proteoforms created are limited to the ones with combinations of PTMs having this many PTMS or fewer.
Maximum SNPs per Isoform	Limits the number SNPs to add to any given proteoform. If more than this number of SNPs are present on an isoform, the proteoforms created are limited to the ones with combinations of SNPs having this many SNPs or fewer.
Remove Disulfide Bonds	Thermo Fisher Scientific recommends setting this parameter to False. You can set this parameter to True if there are annotated disulfide bonds in the search database that you want to avoid searching, or when there is a known error in the disulfide bond annotation in the database. Errors in UniProt where disulfide bonds are placed on non-cysteine residues have been observed and can cause the search to crash.

Table 23. ProSightPD 4.2 Tag to Annotate Proteoform Search parameters (Sheet 2 of 2)

Parameter	Description
N-Term Modifications to Include	Selected modifications are included on all intact proteoforms n-termini.
Delta M Mode	Determines whether the application conducts the search in Delta M (Δm) Mode. For more information on this mode, see Delta M Mode searches.
Calculate FDR	Specifies whether the FDR is calculated. You might see fewer results passing the FDR threshold if you use Delta M Mode.
Decoy Reps	The number of times a decoy search is run. A higher number of repetitions produces more stable decoy results but increases search times.
Maximum PrSMs per Precursor	Maximum number of PrSMs matched per precursor limits the number of results to the highest scoring hits.
Minimum Matched Fragments	Return only hits with this many or more matched fragments.
3. Static (Fixed) Modifications	
Static (Fixed) Modification	Specifies a static (also known as fixed) modification from a list of known modifications. Static modifications apply universally to every instance of the specified residues.
4. Multithreading Options	
CPU Usage	The amount of CPU to direct to data processing: • High—The total number of CPU cores less 1 • Medium—Half the number of available cores • Low—1 core
5. Fragmentation Method	
Override Fragmentation Method	Manually set the fragmentation method. Auto: allows the software to determine the correct fragmentation type. (Default)
6. UVPD Method	
Default UVPD Method	Set the number of ion types to consider for the score of UVPD fragment ion spectra: • 4—recommended for maximizing IDs • 9—recommended for maximizing characterization

Tag to Subsequence Search

The Tag to Subsequence Search node first performs a tag search and then performs a sequence tag search on the output of the tag search.

Table 24 lists the parameters for the 4.2 version of the node. For information about the 3.0 version, refer to the user interface.

Table 24. ProSightPD 4.2 Tag to Subsequence Search node parameters (Sheet 1 of 3)

Parameter	Description
1. Input Data	
Input Database	Select the search database.
2a. Subsequence Tag	
Minimum Tag Score	Minimum probability that the probable tag is not found by chance.
Amino Acid Gap	Largest number allowed of sequential amino acids not having a fragment between them.
Max Tags	Greatest number of probable tags considered for each search per target.
Minimum Fragments Used	Minimum number of fragments required for each sequence tag.
Remove Water Losses	If set to True, water loss is not considered when searching sequence tags.
Consider Isobaric Dipeptide Gaps	If set to True, for example, both GA and Q are considered for gaps of 128.0 Da.
Compiler Tolerance	The permissible error, measured in ppm between two fragment ion masses that are still considered as matching an amino acid.
2b. Subsequence (Biomarker) Search	
Subsequence Precursor Mass Tolerance	Specifies the tolerance when comparing subsequence mass to observed precursors. A large tolerance setting might increase search time with large databases. • Minimum value = 0.01 Da 10.1 ppm • Maximum value = 1 000 000 Da 1000 ppm
Fragment Mass Tolerance	Specifies the mass tolerance in Da or ppm used for matching fragment peaks. Range: 0–2.0 Da; 1–50 ppm

Table 24. ProSightPD 4.2 Tag to Subsequence Search node parameters (Sheet 2 of 3)

Parameter	Description
Maximum PTMs per Isoform	Limits the number PTMs to add to any given proteoform. If more than this number of PTMs are present on an isoform, the proteoforms created are limited to the ones with combinations of PTMs having this many PTMS or fewer.
Maximum SNPs per Isoform	Limits the number SNPs to add to any given proteoform. If more than this number of SNPs are present on an isoform, the proteoforms created are limited to the ones with combinations of SNPs having this many SNPs or fewer.
Maximum Mass to Include PTMs	Maximum proteoform mass to search with PTMs.
Remove Disulfide Bonds	Thermo Fisher Scientific recommends setting this parameter to False. You can set this parameter to True if there are annotated disulfide bonds in the search database that you want to avoid searching, or when there is a known error in the disulfide bond annotation in the database. Errors in UniProt where disulfide bonds are placed on non-cysteine residues have been observed and can cause the search to crash.
N-Term Modifications to Include	Selected modifications are included on all intact proteoforms n-termini.
Delta M Mode	Determines whether the application conducts the search in Delta M (Δm) Mode. For more information on this mode, see Delta M Mode searches.
Calculate FDR	Specifies whether the FDR is calculated. You might see fewer results passing the FDR threshold if you use Delta M Mode.
Decoy Reps	The number of times a decoy search is run. A higher number of repetitions produces more stable decoy results but increases search times.
Maximum PrSMs per Precursor	Maximum number of PrSMs matched per precursor limits the number of results to the highest scoring hits.
Minimum Matched Fragments	Return only hits with this many or more matched fragments.
3. Static (Fixed) Modifications	
Static (Fixed) Modification	Specifies a static (also known as fixed) modification from a list of known modifications. Static modifications apply universally to every instance of the specified residues.

Table 24. ProSightPD 4.2 Tag to Subsequence Search node parameters (Sheet 3 of 3)

Parameter	Description
4. Multithreading Options	
CPU Usage	The amount of CPU to direct to data processing: • High—The total number of CPU cores less 1 • Medium—Half the number of available cores • Low—1 core
5. Fragmentation Method	
Override Fragmentation Method	Manually set the fragmentation method. Auto: allows the software to determine the correct fragmentation type. (Default)
6. UVPD Method	
Default UVPD Method	Set the number of ion types to consider for the score of UVPD fragment ion spectra: • 4—recommended for maximizing IDs • 9—recommended for maximizing characterization

cRAWler nodes

This section describes the cRAWler nodes:

- PTCR cRAWler node

- High/High cRAWler node

- Med/High cRAWler node

PTCR cRAWler node

The PTCR cRAWler node processes PTCR precursor data with the APD algorithm and uses the Xtract algorithm to deconvolute fragmentation spectra.

Table 25 lists the node parameters.

Table 25. ProSightPD PTCR cRAWler node parameters (Sheet 1 of 4)

Parameter	Description
1. Chromatographic Parameters	
Max RT Difference [min]	The maximum time elapsed between MS/MS scans to be grouped together into a target.

Table 25. ProSightPD PTCR cRAWler node parameters (Sheet 2 of 4)

Parameter	Description
Precursor m/z Tolerance	The maximum difference in precursor isolation *m/z* allowed for grouping of subsequent MS/MS scans.
Grouper	Defines the grouping method: • Standard—Groups MS/MS scans within retention time and *m/z* tolerances and groups the corresponding MS scans • PTCR—Groups a single fragmentation scan with its corresponding PTCR precursor scan
2. Precursor APD Parameters	
Use Manual Precursor Mass	Selects from the following: • Never: Always uses APD detected masses • Always: Only uses the manually input Precursor Mass parameter for every target mass • If not detected: When no mass is detected by APD the manually input Precursor Mass parameter is used for the target
Precursor Mass (Da) Override	User-defined precursor mass in Daltons.
Use Instrument Determined Precursor Mass	Allows the instrument to determine the precursor mass.
Apply Precursor Spectral Shift	The decision whether or not to shift the spectrum by a ppm (part per million) shift.
Precursor Spectral Shift	The ppm amount to shift the spectrum. Minimum value: -100 Maximum value: 100
3. Fragmentation Xtract Parameters	
S/N Threshold	The minimum signal to noise threshold for a peak that the Xtract algorithm processes.
Lowest m/z	The lowest *m/z* considered by the Xtract algorithm.
Highest m/z	The highest *m/z* considered by the Xtract algorithm.
Lowest Charge	The lowest charge considered by the Xtract algorithm.
Highest Charge	The highest charge considered by the Xtract algorithm.

Table 25. ProSightPD PTCR cRAWler node parameters (Sheet 3 of 4)

Parameter	Description
Fit Factor	The minimum pattern accuracy for an isotopic distribution to be considered a valid Xtract mass.
Overlapping Remainder	Minimum relative remaining intensity of an isotope pattern overlapped by another pattern to be considered.
Resolution at m/z 400	Instrumental Resolution at m/z 400.
Apply Fragment Spectral Shift	The decision whether or not to shift the spectrum by a ppm (part per million) shift.
Fragment Spectral Shift (ppm)	The ppm amount to shift the spectrum. Minimum value: -100 Maximum value: 100
4. Multithreading Options	
CPU Usage	Determines the amount of CPU to direct toward processing: • High—All available threads less 1 (Default) • Medium—50% of the available threads • Low—1 thread
5. Data Reduction Parameters	
Enable Noise Reduction	Makes the following noise reduction parameters available.
Relative Multiple Precursor Threshold	Only consider additional precursors greater than the set value.
Absolute Minimum Fragment Intensity	Only considered fragments with intensity greater than the set value.
Window Size Below	Bin size below m/z 2000 for considering fragments.
Window Size Above	Bin size above m/z 2000 for considering fragments.
Fragments per window size above	Number of fragments per bin above m/z 2000 to consider.
Fragments per window size below	Number of fragments per bin below m/z 2000 to consider.
Minimum Fragment Count	Minimum number of fragments for a target to be searched.
Number of Precursors Per Experiment	Maximum number of precursors for a given experiment.

Table 25. ProSightPD PTCR cRAWler node parameters (Sheet 4 of 4)

Parameter	Description
6. Misc.	
Heavy Labeled	None: Standard
	C13: All carbons are C13.
	N15: All nitrogens are N15.
	C13 & N15: All carbons are C13 and all nitrogens are N15.

High/High cRAWler node

The High/High cRAWler node uses the Xtract algorithm to deconvolute the spectra from High/High data. The node then groups the deconvoluted spectra based on your parameter settings.

Table 26 lists the node parameters.

Table 26. ProSightPD High/High cRAWler node parameters (Sheet 1 of 3)

Parameter	Description
1. Chromatographic Parameters	
Max RT Difference [min]	The maximum time elapsed between MS/MS scans to be grouped together into a target.
Precursor m/z Tolerance	The maximum difference in precursor isolation *m/z* allowed for grouping of subsequent MS/MS scans.
Grouper	Determines the grouping method: • Standard—Groups MS/MS scans within retention time and *m/z* tolerances and groups the corresponding MS scans • PTCR—Groups a single fragmentation scan with its corresponding PTCR precursor scan
2. Precursor Xtract Parameters	
S/N Threshold	The minimum signal to noise threshold for a peak that the Xtract algorithm processes.
Lowest m/z	The lowest *m/z* considered by the Xtract algorithm.
Highest m/z	The highest *m/z* considered by the Xtract algorithm.
Lowest Charge	The lowest charge considered by the Xtract algorithm.
Highest Charge	Highest Charge considered by the Xtract algorithm.

Table 26. ProSightPD High/High cRAWler node parameters (Sheet 2 of 3)

Parameter	Description
Fit Factor	The minimum pattern accuracy for an isotopic distribution to be considered a valid Xtract mass.
Overlapping Remainder	Minimum relative remaining intensity of an isotope pattern overlapped by another pattern to be considered.
Resolution at m/z 400	Instrumental Resolution at *m/z* 400
Use Manual Precursor Mass	Selects from the following: • Never—Always uses the Xtract algorithm detected masses (Default) • Always—Only uses the manually input Precursor Mass parameter for every target mass • If not detected—When no mass is detected by the Xtract algorithm, the manually input Precursor Mass parameter is used for the target
Precursor Mass (Da) Override	User-defined precursor mass in Daltons.
Use Instrument Determined Precursor Mass	Sets precursor mass determined by the instrument.
3. Fragmentation Xtract Parameters	
S/N Threshold	The minimum signal-to-noise threshold for a peak that the Xtract algorithm processes.
Lowest m/z	The lowest *m/z* considered by the Xtract algorithm.
Highest m/z	The highest *m/z* considered by the Xtract algorithm.
Lowest Charge	The lowest charge considered by the Xtract algorithm.
Highest Charge	The highest charge considered by the Xtract algorithm.
Fit Factor	The minimum pattern accuracy for an isotopic distribution to be considered a valid Xtract mass.
Overlapping Remainder	Minimum relative remaining intensity of an isotope pattern overlapped by another pattern to be considered.
Resolution at m/z 400	Instrumental Resolution at *m/z* 400.

Table 26. ProSightPD High/High cRAWler node parameters (Sheet 3 of 3)

Parameter	Description
4. Multithreading Options	
CPU Usage	Determines the amount of CPU to direct toward processing: • High—All available threads less 1 (Default) • Medium—50% of the available threads • Low—1 thread
5. Data Reduction Parameters	
Enable Noise Reduction	Makes the following noise reduction parameters available.
Relative Multiple Precursor Threshold	Only consider additional precursors greater than the set value.
Absolute Minimum Fragment Intensity	Only considered fragments with intensity greater than the set value.
Window Size Below	Bin size below *m/z* 2000 for considering fragments
Window Size Above	Bin size above *m/z* 2000 for considering fragments.
Fragments per window size above	Number of fragments per bin above *m/z* 2000 to consider.
Fragments per window size below	Number of fragments per bin below *m/z* 2000 to consider.
Minimum Fragment Count	Minimum number of fragments for a target to be searched.
Number of Precursors Per Experiment	Maximum number of precursors for a given experiment.

Med/High cRAWler node

The Med/High cRAWler node provides mass assignments for precursor species in medium-high data with unresolved isotope distributions. The Med/High cRAWler node uses an iterative charge-state assignment approach to examine the most abundant *m/z* species in a spectrum and find real charge-state distributions. This approach is suitable for mass determination of larger mass precursor species.

Table 27 lists the node parameters.

Table 27. ProSightPD Med/High cRAWler Node parameters (Sheet 1 of 3)

Parameter	Description
1. Chromatographic Parameters	
Max. RT Difference[min]	The maximum time elapsed between MS/MS scans to be grouped together into a target.
Precursor m/z Tolerance	Specifies the maximum difference in precursor isolation *m/z* allowed for grouping of subsequent MS/MS scans.
Grouper	The grouping method: • Standard—Groups MS/MS scans within retention time and *m/z* tolerances and groups the corresponding MS scans. • PTCR—Groups a single fragmentation scan with its corresponding PTCR precursor scan.
2. Precursor kDecon Parameters	
Max Precursor Mass	Specifies the maximum precursor mass to consider.
Min Precursor Mass	Specifies the minimum precursor mass to consider.
Number Of Results	The maximum number of mass results to return for each averaged scan. The node filters the results by S/N before returning them.
Number of Iterations	The number of times to run the node. After each iteration, the node removes charge-state distributions from each mass contained in the spectrum to limit false positives during subsequent iterations.
S/N Result Cutoff	The geometric mean signal to noise ratio (S/N) is calculated from the individual S/N of each charge state in the charge state distribution of a detected mass.
Full Noise Reduction	Specifies whether to use a QuickHull algorithm to reduce noise. When set to False, the node removes peaks that are not above an exponential moving average.

Table 27. ProSightPD Med/High cRAWler Node parameters (Sheet 2 of 3)

Parameter	Description
Use Manual Precursor Mass	Selects from the following: • Never—Always uses algorithm detected masses (Default) • Always—Uses only the manually input Precursor Mass parameter for every target mass • If not detected—When the Xtract algorithm does not detect a mass, the manually input Precursor Mass parameter is used for the target
Precursor Mass (Da) Override	User-defined precursor mass in Daltons.
Use Instrument Determined Precursor Mass	Sets precursor mass determined by the instrument.
Charge State Intensity Type	The intensity returned for each mass is the sum of the detected charge states. If this parameter is set to Charge Normalized, the intensity of each charge state is divided by its charge.
Sum Centroid Intensities	If set to True, all intensity values that are part of the centroided peak are summed together. If set to False, only the most abundant *m/z* value is used.
3. Fragmentation Xtract Parameters	
S/N Threshold	The minimum S/N value for processing a peak.
Lowest m/z	Lowest *m/z* value to consider.
Highest m/z	Highest *m/z* value to consider.
Lowest Charge	The lowest charge considered by the Xtract algorithm.
Highest Charge	The highest charge considered by the Xtract algorithm.
Fit Factor	The minimum pattern accuracy for an isotopic distribution to be considered a valid Xtract mass.
Overlapping Remainder	Minimum relative remaining intensity of an isotope pattern overlapped by another pattern to be considered.
Resolution at m/z 400	Instrumental Resolution at *m/z* 400
4. Multithreading Options	
CPU Usage	Determines the amount of CPU to direct toward processing: • High—All available threads less 1 • Medium—50% of the available threads • Low—1 thread

Table 27. ProSightPD Med/High cRAWler Node parameters (Sheet 3 of 3)

Parameter	Description
5. Data Reduction Parameters	
Enable Noise Reduction	Makes the following noise reduction parameters available.
Relative Multiple Precursor Threshold	Consider only additional precursors greater than the set value.
Absolute Minimum Fragment Intensity	Consider only fragments with intensity greater than the set value.
Window Size Below	Bin size below *m/z* 2000 for considering fragments.
Window Size Above	Bin size above *m/z* 2000 for considering fragments.
Fragments per window size above	Number of fragments per bin above *m/z* 2000 to consider.
Fragments per window size below	Number of fragments per bin below *m/z* 2000 to consider.
Minimum Fragment Count	Minimum number of fragments for a target to be searched.
Number of Precursors Per Experiment	Maximum number of precursors for a given experiment.

ProSightPD Consensus nodes

This section describes all the top-down consensus nodes:

- PrSM Grouper node

- Protein Grouper node

- FDR node

- Proteoform Validator node

- Proteoform and Protein Filter node

- Feature Mapper node

- Quantifier node

- PFR Annotator node

PrSM Grouper node

The Proteoform Spectral Match Grouper node groups redundant PrSMs into proteoforms. In the consensus workflow, this node follows the MSF files node.

Table 28 lists the advanced node parameters. The node does not have regular parameters.

Table 28. ProSightPD PrSM Grouper advanced node parameters

Parameter	Description
1. C-Score Threshold	
Medium PrSM C-Score Threshold	PrSMs with C-Scores greater than or equal to this value, but less that the High PrSM C-Score Threshold are grouped together into medium confidence proteoforms.
	A proteoform with at least one PRSM of medium confidence is considered medium confidence.
	Default: 3
High PrSM C-Score Threshold	PrSMs with C-Scores greater than or equal to this value are grouped together into high confidences proteoforms.
	A proteoform with at least one PrSM of high confidence is considered high confidence.
	Default: 40
2. Grouping by Tolerance	
Should group by tolerance	If set to True, software groups proteoforms by mass in addition to sequence and modification.
	If set to False, software groups proteoforms only on their sequence and modifications. (Default)
Group Detected Mass Tolerance	Specifies the grouping mass tolerance.

Protein Grouper node

The Protein Grouper node groups proteoforms and protein isoforms into their parent proteins. The node includes results thresholding and grouping options.

The Protein Grouper node does not have parameters.

FDR node

You can use this node's parameter setting cutoff to include search results in the pdResult report.

The FDR calculations are implemented as described in the 2019 Mol Cell Proteomics article, "Accurate Estimation of Context–Dependent False Discovery Rates in Top–Down Proteomics," by LeDuc et. al. For more information, go to https://pubmed.ncbi.nlm.nih.gov/30647073/.

Table 29 lists the node parameters.

Table 29. ProSightPD FDR node parameters

Parameter	Description
1. FDR	
FDR cutoff	Enter a number that search results must exceed to be included in the pdResult report.
2. tdReport	
Generate tdReport	Generate a tdReport from the search results. The TD Viewer application can open the tdReport.

Proteoform Validator node

The Proteoform Validator node calculates proteoform confidences based on associated PrSMs.

The Proteoform Validator node does not have parameters.

Proteoform and Protein Filter node

The Proteoform and Protein Filter node filters proteoforms based on confidence.

Table 30 lists the node parameters.

Table 30. ProSightPD Proteoform and Protein Filter Node

Parameter	Description
1. Proteoform Filters	
Proteoform Confidence at Least	Proteoforms with a lower confidence than specified are excluded from the final result: • High (Default) • Medium • Low You can define confidence levels in the configuration panel of the Administration page.
Keep Lower Confidence PrSMs	If set to True, keep PrSMs with lower confidence than specified at the proteoform level. If set to False, do not keep them. (Default)

Feature Mapper node

The Feature Mapper node maps quantification consensus features onto proteoforms within the tolerances specified by the node.

Table 31 lists the node parameters.

Table 31. ProSightPD Feature Mapper node parameters

Parameter	Description
1. Consensus Feature	
Feature Presence in File Threshold	Species minimum percentage of files a species must be quantified to return a quantitation result. For example, if the threshold is set to 30%, then a proteoform must be detected in at least 30% of the files analyzed to report a quantitation value.
2. Consensus Features Group/Proteoform Connection	
Feature Group Proteoform Connection Time Tolerance (Minutes)	The maximum allowable separation in time for two feature groups to be grouped together.
Feature Group Proteoform Connection Mass Tolerance (Da or ppm)	The maximum allowable mass difference for two feature groups to be grouped together.

Quantifier node

The Quantifier node applies normalization and statistics to the resulting quantitative information. To use the volcano plots and heat maps, this node must be present in the consensus workflow.

Table 32 lists the node parameters.

Table 32. ProSightPD Quantifier node parameters

Parameter	Description
1. Settings	
Normalize	If set to True, normalization is applied to the proteoform level quantitation values.
Use Shared Results	If set to True, proteoforms can share quantification values.
	Due to minor differences in modification states (for example, positional isomers) frequently similar proteoforms coelute and therefore share elution profiles. If non-similar, non-coeluting species utilize shared quantitation values, Thermo Fisher Scientific recommends lowering the feature mapping tolerances or setting Used Shared Results to False.

PFR Annotator node

The PFR Annotator node assigns PFR Accessions to proteoforms that pass the FDR threshold in the ProSightPD FDR node. Only proteoforms associated with a protein entry and a Taxon Id will be annotated. If the search result is associated with multiple Taxon Ids, PFR Accessions will not be assigned.

The PFR Annotator node adds a column to the proteoform results table, which includes a proteoform reference number (PFR). The PFR lets you track proteoforms across different experiments.

The PFR Annotator node does not have parameters.

Interpreting results

This chapter provides information about scoring systems shown in the Sequence Gazer for ProSightPD results.

Contents

- P-Score
- C-Score
- Expectation Value (E-Value)

P-Score

A P-Score is the probability of obtaining at least as good a match between the observed fragment list and a sequence as by chance. It is a measure of confidence in the validity of a match. A low P-Score means that the probability of obtaining at least this many fragments matching a sequence is low. Thus, an association caused by random chance is unlikely.

The application calculates a P-Score as follows:

$$p(n) = \sum_{i=0}^{n-1} \frac{e^{-xf}(xf)^i}{i!}$$

where:

- n is the number of matching fragments.

- x is the probability of an observed fragment ion matching a random theoretical fragment ion by chance.

- f is the total number of fragment ions observed.

Matching more fragment ions leads to a lower, more confident P-Score. Alternatively, increasing the numbers of total detected fragment ions without a corresponding increase in matched fragment ions negatively affects the confidence of the result. To see the original score derivation, refer to:

https://www.nature.com/articles/nbt1001-952[1]

C-Score

The C-Score measures the level of characterization of a proteoform in relation to the others in the database. This score, described in LeDuc et al.[2], uses a Bayesian approach that assigns a likelihood to each candidate proteoform based on the observed MS data.

A C-Score of 3 indicates that there are two proteoforms in the database that equally explain the observed data. A score of 40 or higher is considered strong evidence of a unique characterization.

Expectation Value (E-Value)

The expectation value (E-Value) is the number of sequences in a database that are expected to have P-Scores equal to or better than what was observed by chance. Low E-Values represent better matches (less likely to be false positives) than high E-Values. Because the P-Score represents the probability of the n out of f fragments matching by chance, and if it is assumed that all sequences in the database are independent, the E-Value of a sequence-fragment set association is the P-Score multiplied by the number of sequences in the database.

If N is the number of proteoforms considered during a search, the E-Value currently reported by the software is:

$$e = N \times p(n)$$

---

[1] Meng, F.; Cargile, B.J.; Miller, L.M.; Forbes, A.J.; Johnson, J.R.; Kelleher, N.L. Informatics and multiplexing of intact protein identification in bacteria and the archaea. Nat. Biotechnol. 2001 Oct 1, 19(10):952–97.

[2] LeDuc, R.D.; Fellers, R.T.; Early, B.P.; Greer, J.B.; Thomas, P.M.; Kelleher, N.L. The C-score: a Bayesian framework to sharply improve proteoform scoring in high-throughput top-down proteomics. *J Proteome Res.* **2014** Jul 3, 13 (7):3231-40.