thermoscientific

Thermo

Proteome Discoverer

User Guide

Software Version 2.2

XCALI-97808   Revision A   •   June 2017

Thermo Fisher
S C I E N T I F I C

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

Contents

Contents

Contents

Preface

This guide describes how to use the Thermo Proteome Discoverer™ application for peptide and protein mass spectrometry analyses.

Contents

❖ **To suggest changes to the documentation or to the Help**

- Accessing Documentation
- System Requirements
- Special Notices
- Contacting Us

Complete a brief survey about this document by clicking the button below. Thank you in advance for your help.

Accessing Documentation

The Proteome Discoverer application includes complete documentation. For system requirements, refer to the release notes on the software DVD.

❖ **To view the product manual**

From the Microsoft™ Windows™ taskbar, do the following:

- For a Thermo Scientific™ application, choose **Start > All Programs >** …. Or, from the application, choose **Help > Manuals** and so on.

- For a Thermo Scientific mass spectrometer, choose **Start > All Programs >** ….

- For an LC instrument controlled by a Thermo Scientific application, choose **Start > All Programs > Thermo Instruments > Manuals** and so on.

❖ **To view user documentation on the Thermo Fisher Scientific website**

1. Go to www.thermoscientific.com.

2. Point to **Services & Support** and click **Manuals** on the left.

3. In the Refine Your Search box, search by the product name.

4. From the results list, click the title to open the document in your web browser, save it, or print it.

   To return to the document list, click the browser **Back** button.

❖ **To view the Proteome Discoverer Help**

- From the application window, choose **Help > Proteome Discoverer Help**, or press the F1 key.

- In applications that have a Communicator bar, click a field or parameter to display definitions, required actions, ranges, defaults, and warnings.

System Requirements

The Proteome Discoverer application requires a license. In addition, ensure that the system meets the minimum requirements listed in Table 1.

Table 1. System requirements

System	Minimum requirements
Computer	• 2 GHz processor • (Recommended) Two Intel™ Xeon™ 6-core processors, 2.4 GHz • 2 GB RAM • Video card and monitor capable of 1280 × 1024 resolution (XGA) • Screen resolution of 96 dpi • 1 TB available on drive C • NTFS format

Software	• Adobe™ Reader™ 10
	
	• Microsoft Windows 7 Professional with Service Pack 1
	
	   – 64-bit operating system for the Proteome Discoverer application
	
	   – 32- or 64-bit operating system for the Proteome Discoverer Daemon utility
	
	• Microsoft .NET Framework 4.5.1

Mascot^{SM} Server	• Mascot Server 2.1
	
	   – Mascot servers running version 2.2 are usable, but retrieving the result files (protein sequences) from the servers can be a lengthy process because you can only retrieve the protein sequences one at a time.
	
	   – Mascot servers running version 2.1 should have all available updates, patches, or both from Matrix Science already installed. Ensure that you have installed a patch that enables MIME format for the result files; otherwise, the Proteome Discoverer application cannot receive the search results from the Mascot server.
	
	• Mascot Server 2.2: Proteome Discoverer 2.2 does not support error-tolerant searches.
	
	• Mascot Server 2.3: Proteome Discoverer 2.2 does not support error-tolerant searches and searches against multiple-sequence databases.
	
	• Mascot Server 2.4: Proteome Discoverer 2.2 does not support error-tolerant searches.

> **Note** Ensure that port 28199 is not blocked by firewalls.

> **Note** Before installing the Proteome Discoverer application, ensure that the Windows operating system has the latest Microsoft .NET Framework and Windows updates installed.

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.

Contacting Us

Contact	Email	Telephone	QR Code[a]
U.S. Technical Support	us.techsupport.analyze@thermofisher.com	(U.S.) 1 (800) 532-4752	
U.S. Customer Service and Sales	us.customer-support.analyze@thermofisher.com	(U.S.) 1 (800) 532-4752	
Global support	❖ **To find global contact information or customize your request** 1. Go to thermofisher.com. 2. Click **Contact Us**, select the country, and then select the type of support you need. 3. At the prompt, type the product name. 4. Use the phone number or complete the online form. ❖ **To find product support, knowledge bases, and resources** Go to thermofisher.com/us/en/home/technical-resources. ❖ **To find product information** Go to thermofisher.com/us/en/home/brands/thermo-scientific.		

Note To provide feedback for this document, go to surveymonkey.com/s/PQM6P62 or send an email message to Technical Publications (techpubs-lcms@thermofisher.com).

[a] You can use your smartphone to scan a QR Code, which opens your email application or browser.

Introduction

This chapter introduces you to the Proteome Discoverer application and describes its features and functionality.

Contents

Features

The Proteome Discoverer application is a client-server application that uses workflows to process and report mass spectrometry data. It compares the raw data taken from mass spectrometry or spectral libraries to the information from a selected FASTA database and identifies proteins from the mass spectra of digested fragments. You can use this application to analyze spectral data from all Thermo Scientific and other mass spectrometers. Specifically, the application does the following:

- Works with peak-finding search engines such as Sequest™ HT and Mascot to process all data types collected from low- and high-mass-accuracy mass spectrometry (MS) instruments. The peak-finding algorithm searches the raw mass spectrometry data and generates a peak list and relative abundances. The peaks represent the fragments of peptides for a given mass and charge.

- Produces complementary data from a variety of dissociation methods and data-dependent stages of tandem mass spectrometry.

- Combines, filters, and annotates results from several database search engines and from multiple analysis iterations. The search engines correlate the uninterrupted tandem mass spectra of peptides with databases, such as FASTA. (See "Using FASTA Databases" on page 151.)

Proteome Discoverer Application

The Proteome Discoverer application consists of the main client, called Proteome Discoverer; the Magellan server; the Proteome Discoverer Daemon utility; and the different client applications that you can connect to the Proteome Discoverer application.

Proteome Discoverer Client

In general, you interact with the main window of the Proteome Discoverer client. You use it for creating and submitting the workflows to be processed; monitoring job progress; configuring the server; administering repository data, such as FASTA files, spectral libraries, and modifications; and displaying the results.

You create workflows in the application's Workflow Editor, create processing and consensus analysis sequences within a study, and then submit them for execution to the Magellan server running a local or remote process.

Magellan Server

The Magellan server is a command-line application that performs all workflow-based data processing. It can run on the same computer or on a remote machine. It is responsible for registering and managing processing nodes that you use within the data processing workflows, processing workflow jobs, storing data, and general data maintenance. The server is extendable with new processing nodes. The Magellan server publishes client gateways by .NET remoting, and multiple clients can connect to these gateways to communicate with the server. The Magellan server maintains a repository of the registered processing nodes, their configurations, and metadata about the available nodes. It also maintains repositories of registered data such as FASTA files, spectral libraries, chemical modifications, and so forth.

The Magellan server interprets user-specified workflows and converts them to jobs that can be processed. It controls the execution of data processing jobs and provides real-time status updates to connected clients.

Proteome Discoverer Daemon Utility

You use the Proteome Discoverer Daemon utility to automatically start workflow jobs from a remote system. You can use this utility to start batches of files to process. It can perform multiple searches on multiple raw data files at any given time. You can use it to perform searches on multiple raw data files taken from multiple samples or replicates from the same sample. Through its command line, you can automate the submission of workflow jobs to the Magellan server. You can also use the Proteome Discoverer Daemon utility to submit jobs manually. For more information on the Proteome Discoverer Daemon utility, see "Using the Proteome Discoverer Daemon Utility" on page 123.

The Proteome Discoverer application includes the following additional features.

Study Management

The Proteome Discoverer application uses a study management system to organize your experiment data and to automate processing. The study management system comprises studies and analyses. A study contains your input files, the information about the samples contained in these files, information about the treatment of the samples, the workflows used to process the input files, and the results of the processing. An analysis contains the processing and consensus workflows that you assemble in the Workflow Editor to process your data. Each analysis is associated with a study.

Search Engines

The Proteome Discoverer application supports the Sequest HT and Mascot search engines; each produces complementary data. These search engines are available as nodes in the Workflow Editor.

Sequest HT Search Engine

The Sequest HT search engine is distributed by Thermo Fisher Scientific. It can analyze different data types:

- Electron-transfer dissociation (ETD)
- Electron transfer dissociation with HCD or CID activation (EThcD)
- Electron-capture dissociation (ECD)
- Collision-induced dissociation (CID)
- High-energy collision-induced dissociation (HCD)
- Pulsed Q collision-induced dissociation (PQD)

ETD and ECD generate primarily c and z fragment ions with preferences for precursor ion charge states of +3 or higher. CID and HCD generate primarily b and y fragment ions with preferences for precursor ion charge states of +3 or lower. PQD and HCD do not exhibit a low-mass cutoff and are good for reporter-ion experiments.

Frequently, peptides identified by CID, PQD, or HCD are not observed with ETD or ECD, and vice versa, so that combining results from, for example, CID and ETD can enhance sequence coverage. Many times CID and ETD identify the same peptides, often with different precursor ion charge states. Combining ETD and CID results improves confidence in identifications.

The Sequest HT search engine calculates XCorr scores for peptide matches and provides the peptide matches having the best XCorr score for each spectrum. It calculates a preliminary SpScore score and uses it to filter peptide candidates. It calculates XCorr values for peptide spectrum matches (PSMs) only if they pass the SpScore filter. The Sequest HT node calculates the XCorr value for every peptide candidate.

Mascot Search Engine

The Mascot search engine, created by Matrix Science, uses mass spectrometry data to identify proteins from primary sequence databases. For more details on Mascot, visit www.matrixscience.com.

Workflow Editor

For ease of use and greater flexibility, the Proteome Discoverer application offers a dual-workflow search capability to search for matching proteins and peptides.

You can use the application's Workflow Editor to create customized searches and customized results reports. You create two workflows: a customized processing workflow that generates the primary search results from the input data files and a customized consensus workflow to display the results. The Workflow Editor can accept single or multiple input raw data files for its processing workflow and single or multiple MSF files for its consensus workflow. It can search with multiple algorithms and merges results from multiple fragmentation methods. For detailed information about the Workflow Editor, see "Workflow Editor" on page 34.

Quantification

You can perform reporter ion quantification (for example, iTRAQ™ and Tandem Mass Tag™ [TMT]), x precursor ion quantification (for example, SILAC), and label-free quantification with the Proteome Discoverer application. For details, see "Performing Reporter Ion Quantification" on page 347, "Performing Label-Free Quantification" on page 467, and "Performing Precursor Ion Quantification" on page 411.

Reporter ion quantification is an isobarically labeled quantification method that uses external reagents, or tags, to chemically label proteins and peptides to detect differences in abundances. TMT quantification offers default 2plex, 6plex, and 10plex quantification methods, and iTRAQ offers 4plex and 8plex quantification methods. You can use these methods to create your own quantification templates. iTRAQ and TMT use the Reporter Ions Quantifier node in the quantification workflow.

Precursor ion quantification is an isotopically labeled quantification method that uses in-vivo metabolic labeling to detect differences in the abundance of proteins in multiple samples. It uses the Minora Feature Detector node in the processing workflow, and the Precursor Ions Quantifier node and the Feature Mapper node in the consensus workflow.

Label-free quantification is the quantification of MS peptide signals in multidimensional LC/MS2 data without an isotopic labeling reagent. It performs untargeted feature detection; that is, it does not require an identification to quantify a peptide or protein. It requires only a single PSM to quantify across all samples when you use retention time alignment and gap filling. Label-free quantification uses the Minora Feature Detector node in the processing workflow, and the Precursor Ions Quantifier node and the Feature Mapper in the consensus workflow.

ProteinCenter Access

The Proteome Discoverer application gives you access to annotation information from ProteinCenter, which is a web-based application that you can use to download biologically enriched annotation information for a single protein. You can download information such as the following:

- Molecular functions, cellular components, and biological processes from the Gene Ontology (GO) database

- Classification information for protein families from the Protein Family (Pfam) database from the Wellcome Trust Sanger Institute (WTSI)

- Biochemical pathways from the KEGG™ PATHWAY and the Wiki Pathways databases

- Gene identifications from the Entrez Gene database maintained by the National Center for Biotechnology Information (NCBI)

- Genomic annotations of genetically sequenced organisms from the Ensembl genome database, which is a joint project of the European Bioinformatics Institute (EBI) and the WTSI

- Post-translational modification information from the UniProt™ database

The data in ProteinCenter is updated biweekly. You can use this information to annotate the proteins in your results report (.pdResult) file if you have an account with ProteinCenter. For information, see "Obtaining Protein Annotation Information" on page 259.

Graphical Views

The application includes a number of graphical tools to help you analyze your data.

- Data tables in the results report are organized by proteins, protein groups, PSMs, peptide groups, and MS/MS spectrum information. Depending on the nodes that you use in the workflows, you can optionally display a result statistics table and tables on decoy proteins, decoy protein groups, decoy PSMs, and decoy peptide groups. In addition, you can display tables on annotated modifications, found modifications, and unknown modifications.

- Several graphical views contain detailed information about the selected peptides and proteins. You can display more than one view to perform a comparative analysis of your selected peptide or proteins. Refer to the Help.

- Data distribution maps display the areas and ratios calculated by the quantification nodes that show the distribution of values across the available files, samples, and sample groups, or across ratios and ratio groups.

FASTA Databases and Indexes

You can download FASTA databases that you can use when searching for and analyzing data in the Proteome Discoverer application and its indexes. See "Using FASTA Databases" on page 151.

Data Filtering and Validation

A number of protein and peptide filtering, validation, and grouping options are available to help you sort and filter your data. For information on Proteome Discoverer's filtering capabilities, see "Filtering Data" on page 217. For information on validating your data, see "Validating Results" on page 247. For information on grouping, see "Grouping Peptides and Proteins" on page 253.

Export Formats

You can export the spectra from protein and peptide reports in standard spectrum data formats, such as MZDATA, DTA, MZML, and MGF. You can also export search results to XML and tab-delimited TXT files. In addition, you can export annotated spectra for selected peptides into a ZIP file that includes an HTML page with peptide information and links to spectrum images. The Help describes how to export your data to these and other formats.

Qual Browser Application

The Proteome Discoverer application can start the Qual Browser application if Xcalibur is installed on the same computer. You can use Qual Browser to view the entire ion chromatogram and browse individual precursor and MS^n data. You can filter the results in a variety of ways, for example, to produce a selected ion chromatogram. Qual Browser automatically displays the elemental composition, theoretical mass, delta values, and ring and double-bond (RDB) equivalents for your high-resolution data. For more information, see "Using the Qual Browser Application" on page 208.

Fragmentation Methods

The Proteome Discoverer application supports the following fragmentation types:

- CID—Uses the collision-induced dissociation method of fragmentation, where molecular ions are accelerated to high kinetic energy in the vacuum of a mass spectrometer and then allowed to collide with neutral gas molecules such as helium, nitrogen, or argon. The collision breaks the bonds and fragments the molecular ions into smaller pieces.

- ECD—Uses the electron capture dissociation method of fragmentation, where multiply protonated molecules are introduced to low-energy free electrons. Capture of the electrons releases electric potential energy and reduces the charge state of the ions by producing odd-electron ions, which easily fragment.

- HCD—Uses the high-energy collision-induced dissociation method of fragmentation, where the projectile ion has laboratory-frame translational energy higher than 1 keV. HCD produces a highly abundant series of reporter ions for TMT and iTRAQ quantification.

- ETD—Uses the electron transfer dissociation method of fragmentation, where singly charged reagent anions transfer an electron to multiply protonated peptides within an ion trap mass analyzer to induce fragmentation. ETD cleaves along the peptide backbone while side chains and modifications, such as phosphorylation, are left intact. This method is used to fragment peptides and proteins. ETD activation includes the ETciD activation type.

- EThcD—Uses the electron transfer dissociation method of fragmentation but produces additional b and y ions.

- IRMPD—With the infrared multi-photon dissociation method of fragmentation, an infrared laser is directed at the ions in the vacuum of the mass spectrometer. The target ions absorb multiple infrared photons until they reach more energetic states and begin to break bonds, resulting in fragmentation.

- PQD—Uses the pulsed Q collision-induced dissociation method of fragmentation, where precursor ions are activated at a high value, a time delay occurs to allow the precursor to fragment, and then a rapid pulse is applied at a low value where all fragment ions are trapped. The product ions can then be scanned out of the ion trap and detected. PQD fragmentation produces precise, reproducible fragmentation and has been used for iTRAQ peptide quantification on the LTQ™ mass spectrometer using both electrospray and MALDI source ionization.

Peptides and Fragment Ions

The types of fragment ions observed in an MS/MS spectrum depend on several factors, such as the primary sequence, the energy source, and the charge state.

Fragment ions of peptides are produced by a CID process in which a peptide ion is fragmented in a collision cell. Low-energy CID spectra are generated by MS/MS and ESI, and are sequence-specific. The fragment ion spectra contain peaks of the fragment ions formed by cleavage of the peptide bond and are used to determine the amino acid sequence. A fragment must have at least one charge for it to be detected. If this charge is retained on the N terminal fragment, the ion is classed as a, b, or c. If the charge is retained on the C terminal fragment, the ion type is x, y, or z. A subscript indicates the number of residues in the fragment.

In addition to the proton carrying the charge, c ions and y ions abstract an additional proton from the precursor peptide, as shown in Figure 1.

Figure 1. Structures of six singly charged sequence ions
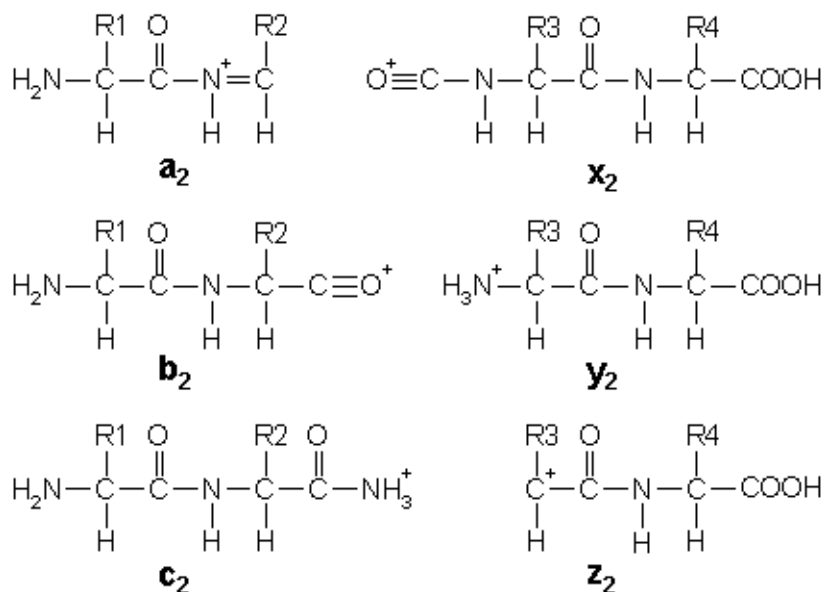
MudPIT Experiments

Multidimensional Protein Identification Technology (MudPIT) experiments investigate complex proteomes by applying multidimensional chromatography to the samples before acquisition in the mass spectrometer. Typically, this process results in several dozen or even a few hundred fractions that are separately analyzed by LC/MS, resulting in one raw data file per sample fraction. Analyzing gel slices or performing in-depth follow-up acquisitions also results in multiple fractions. Because all these fractions belong to the same sample, the Proteome Discoverer application can process all raw data files from these fractions as one contiguous input file and generate a single result file. For detailed information about processing MudPIT samples, see "Using the Proteome Discoverer Daemon Utility" on page 123.

Workflow

Processing, analyzing, and interpreting mass spectrometry data involves several steps in the standard Proteome Discoverer workflow. For the entire workflow, see Figure 2.

Figure 2. The Proteome Discoverer workflow

Experiments produce raw data.

If you intend to use Sequest HT, download a FASTA database.

Proteome Discoverer application

Create a study and an analysis. Identify the RAW data file or files that contain data in the sample.

In the Workflow Editor, create a processing and a consensus workflow.

Select your search parameter settings.

Search the database. The application initiates a FASTA database search.

NOTE You can filter data according to false discovery rates that you define through the use of decoy databases that you specify.

Sort and filter the search report, generate graphs and views, and interpret the search results.

(Optional) Review the quantification results and change any parameters.

(Optional) Reanalyze the quantification results.

(Optional) Export the search results and data to other applications.

Inputs and Outputs

The Proteome Discoverer application can accept several different file formats as input and can export data in several formats.

FASTA Databases

The Proteome Discoverer application includes FASTA databases, including several example FASTA databases and example RAW data files. Use these files when exploring and learning how to use the application. For a description of the different types of FASTA databases and their purpose, see "Using FASTA Databases" on page 151.

Inputs

You can use the application with data from the following mass spectrometers:

- Ion traps, such as the LTQ Velos™

- Orbitraps, such as the Q Exactive™

- Hybrid mass spectrometers, which combine ion traps and Fourier transform ion cyclotron resonance (FT-ICR) or Orbitrap™ analyzers, such as the LTQ Orbitrap Velos or the Orbitrap Fusion Lumos™.

- Quadrupole time-of-flight (QTOF) mass spectrometers

- Triple quadrupole mass spectrometers

The application accepts the following file types as input to the processing workflow:

- Xcalibur raw data files contain raw data collected from a mass spectrometer.

- Mascot Generic Format (MGF) files are mass spectral files produced during Mascot analysis. They contain a list of precursor ions, their fragments, and the masses of the fragments.

- MZDATA files are common data format files developed by the Human Proteome Organization (HUPO) for proteomics mass spectrometry data. These files are in version 1.05 format. They are exported with XML indentation enabled so that the different XML tags are broken into multiple lines instead of merged into one line.

- MZXML files are standard 2.*x* mass spectrometer data format files developed at the Seattle Proteome Center at the Institute for Systems Biology (ISB) that contain a list of precursor ions, their fragments, and the masses of each fragment.

- MZML files are a combination of .mzData and .mzXML formats developed by the Human Proteome Organization Standard Initiative (HUPO-PSI) and the ISB. The Proteome Discoverer application supports version 1.1.0 of the MZML format.

The application accepts MSF files as input to the consensus workflow.

Outputs

The Proteome Discoverer application creates MSF files as output to its processing workflow. The application creates .pdResult files as standard output to its consensus workflow. You can also export spectra in the following formats:

- DTA Archive (DTA) files are files containing MS^n data for single or grouped scans.

- Mascot Generic Format (MGF) files are mass spectral files produced during Mascot analysis. They contain a list of precursor ions, their fragments, and the masses of the fragments.

- MZDATA files are common data format files developed by the HUPO-PSI for proteomics mass spectrometry data. These files are in version 1.05 format. They are exported with XML indentation enabled so that the different XML tags are broken into multiple lines instead of merged into one line.

- MZML files are a combination of MZDATA and MZXML formats developed by the HUPO-PSI and the ISB. The Proteome Discoverer application supports version 1.1.0 of the MZML format.

You can export peptide and protein search results in the following formats:

- ProtXML files contain protein identifications from MS/MS-derived peptide sequence data.

- PepXML files contain peptides that are included in the results of searches performed by the Sequest HT and Mascot search engines. They are in PepXML format version 1.14, which is an open data format developed by SPC/Institute for Systems Biology for storing, exchanging, and processing peptide sequence assignments from MS/MS scans.

- Tab-delimited TXT files are in a simple text format that stores tabular data and is widely used to exchange data between different computer programs.

- mzIdentML (MZID) files contain protein identification results for archiving and sharing proteomics information. You can use this format to upload to a Proteomics Identifications (PRIDE) database (www.ebi.ac.uk/pride/archive).

- mzTab files contain protein identification and quantification results in mzTab format, which is a format for reporting proteomics and metabolomics results.

New Features in This Release

The 2.2 release of the Proteome Discoverer application introduces the following new features.

Calculation of P-values for Samples with Replicates

P-values provide a measure of confidence in quantitative proteomic results. For more information on p-values, see "Calculating P-Values and Adjusted P-Values for Quantification Results" on page 512.

Label-Free Quantification

Label-free quantification is the quantification of MS peptide signals in multidimensional LC/MS2 data without an isotopic labeling reagent. First, the algorithm detects LC/MS peaks in the individual raw data files and maps them to identified PSMs, if they are available. Next, it creates features from peaks that are not associated with any feature. It takes all LC/MS peaks within a small retention-time range and checks for peaks that build isotope patterns. It also uses feature mapping for retention-time alignment and feature linking across files.

ProSightPD Application

The ProSightPD™ application consists of a set of nodes within the Proteome Discoverer application for analysis of top-town proteomics data. The ProSightPC™ software must be installed for you to access these nodes. ProSightPC and ProSightPD are products of Proteinaceous, Inc., exclusively licensed to Thermo Fisher Scientific. For more information, visit proteinaceous.net. For a permanent license, you must purchase ProSightPD software separately. 60-day trial versions of both the Base and the High Mass versions of ProSightPD are available from the Proteome Discoverer Third-Party Node Installer.

XlinkX Algorithm

The XlinkX algorithm identifies chemically crosslinked peptides. The Proteome Discoverer application supports different types of crosslinkers, mainly MS-cleavable crosslinkers that are fragmented by the CID/HCD fragmentation method and non-cleavable crosslinkers that do not break into signature fragments in the mass spectrum.

For a full description of the XlinkX algorithm and the associated MS-based methods, refer to *Liu, Fan, et al. 2017. Optimized fragmentation schemes and data analysis strategies for proteome-wide cross-link identification. Nat. Commun. (8) 15473*. You can access this paper from the following link: https://www.nature.com/articles/ncomms15473.

The XlinkX algorithm is implemented in new processing and consensus workflow nodes. For information on these nodes, refer to the *XlinkX QuickStart Guide*.

The XlinkX algorithm requires a separate license, but the Third-Party Installer software contains a 60-day demonstration version.

Export Formats

You can now export data from the .pdResult file to files in the following formats:

- Export of protein identification results to mzIdentML format

- Export of protein and quantification results to mzTab format

- Export of proteins containing crosslinked peptides to xiNET, a web-based crosslink viewer

Processing Workflow Nodes

This release adds the following new nodes to the processing workflow:

- Spectrum Files RC node—A starting node in the processing workflow. For each spectrum file, it calculates a constant mass shift, in ppm, by performing a quick Sequest HT search using the Fixed Value PSM Validator node to identify highly confident PSMs. It calculates the PSM delta masses between the theoretical and experimental mass, in ppm, and then calculates the mass shift curve as the median of the PSM delta masses. It can calculate a mass shift only when it can find at least 50 PSMs of high confidence.

- Minora Feature Detector node—Detects chromatographic peaks and features according to the specified quantification approach and writes the results to the LCMS peak table. It also links the LC/MS peaks to the PSMs from a given data file. For precursor ion processing, it links the different isotopic forms at this step.

Consensus Workflow Nodes

This release adds the following new nodes to the consensus workflow:

- Feature Mapper node—Performs retention-time alignment and feature linking across data sets.

- Precursor Ions Quantifier node—Performs precursor ion and label-free quantification.

- Reporter Ions Quantifier node—Performs reporter ion quantification, which uses external reagents, or tags, to enzymatically or chemically label proteins and peptides.

Additional Pages in the .pdResult Report

The .pdResult report includes the following new pages:

- Input Files page—Lists all the files used as input to the workflow that produced the .pdResult report.

- Specialized Traces page—Lists the specialized traces—total ion current chromatograms (TICs) and base peak chromatograms (BPCs)—that the application found in the processing workflow.

- Consensus Features page—Lists the consensus features for the analysis and shows the calculated abundances for those features across all the data files.

- LCMS Features page—Lists the LC/MS features detected across all the raw data files.

Additional Result Views

This release adds several new views and charts to help you visualize the data.

Views for Quantification Data

The following charts and views are available only for results that contain quantification data:

- Trend chart—Shows the abundance of proteins or peptide groups across all sample groups. It displays the data as a trend line for each protein or peptide group. For detailed information, see "Displaying the Trend Chart" on page 565.

- Volcano plot—A type of scatter plot used to quickly identify protein or peptide changes in large data sets composed of sample pairs. The left upper quadrant (highlighted in green) displays those proteins or peptide groups that are significantly down-regulated. The upper right quadrant (highlighted in red) displays those proteins or peptide groups that are significantly up-regulated.

  The volcano plot is one of the distribution charts available when you choose View > Distribution charts. For detailed information, see "Displaying the Report Item Distribution Chart as a Volcano Plot" on page 610.

- Principal Component Analysis (PCA) chart—Identifies the major components in a set of data. Principal components are groups of features in the data that contribute the most to distinguish one sample from another. The application automatically calculates these principal components.

  The PCA plot is one of the distribution charts available when you choose View > Distribution charts. For detailed information, see "Displaying the Report Item Distribution Chart as a Principal Component Analysis Plot" on page 618.

- Sample Abundances chart—Displays as a box-and-whisker plot the peak abundance values for all items currently displayed on the Proteins page or the Peptide Groups page.

  The Sample Abundances chart is one of the distribution charts available when you choose View > Distribution charts. For detailed information, see "Displaying the Report Item Distribution Chart as a Sample Abundances Chart" on page 569.

- Mass Recalibration view—Indicates how well the algorithm performed that recalculates precursor mass independently per file. It calculates a correction curve by using support vector regression. This view can also provide insight into potential issues with the acquisition or chromatography itself. It displays the input data (called landmarks in this context) used for the regression and the calculated correction curve. For detailed information, see "Displaying the Mass Recalibration View" on page 629.

- File Alignment view—Indicates how well the retention-time alignment algorithm performed. This algorithm calculates retention time across files. It calculates a correction curve by using support vector regression. This view can also provide insight into potential issues with the acquisition or chromatography itself. For detailed information, see "Displaying the File Alignment View" on page 632.

- Chromatogram Traces view—Displays the chromatographic profiles of individual features or peaks that the application uses in calculating label-free and precursor ion quantification results.

Layout Selector

You can now arrange a custom layout on a page of a .pdResult file, save it, and apply it to other .pdResult files. You can also restore the original layout of a .pdResult file—that is, the layout that was applied by the consensus workflow that originally produced the results. The application provides a Layout Selector area and icon so that you can arrange and save a layout.

Biological Replicate Study Factors

This release adds biological replicate study factors to help you specify how to calculate p-values. You can set up these replicates as nested designs or non-nested designs. For information on this procedure, see "Calculating P-Values for Replicate Data by Using Biological Replicate Study Factors" on page 513.

Missing Value Imputation

The Proteome Discoverer application can now impute missing abundance values. In the Precursor Ions Quantifier node or the Reporter Ions Quantifier node in the consensus workflow, set the Imputation Mode parameter to one of two settings to indicate how the application should treat missing values. For information on missing value imputation, see "Imputing Missing Values" on page 520.

Annotation Databases

This release adds four annotation databases to ProteinCenter.

KEGG PATHWAYS Database from ProteinCenter

The Kyoto Encyclopedia of Genes and Genomes (KEGG) PATHWAY database contains biochemical pathway maps for many species.

Wiki Pathways Database from ProteinCenter

The Wiki Pathways database collects user-provided biochemical pathway maps and links the pathways to the proteins.

Getting Started

This chapter describes how to set up a search in the Proteome Discoverer application.

Contents

To start performing a search immediately, see "Performing a Search" on page 37.

Opening the Proteome Discoverer Application

Open the Proteome Discoverer application by using the Start menu or by clicking the desktop icon.

❖ **To open the Proteome Discoverer application**

From the Start menu, choose **Programs > Thermo Proteome Discoverer 2.1** or click the **Proteome Discoverer** icon, , on your desktop.

The Proteome Discoverer Start Page opens (Figure 3).

Figure 3.　Proteome Discoverer Start Page
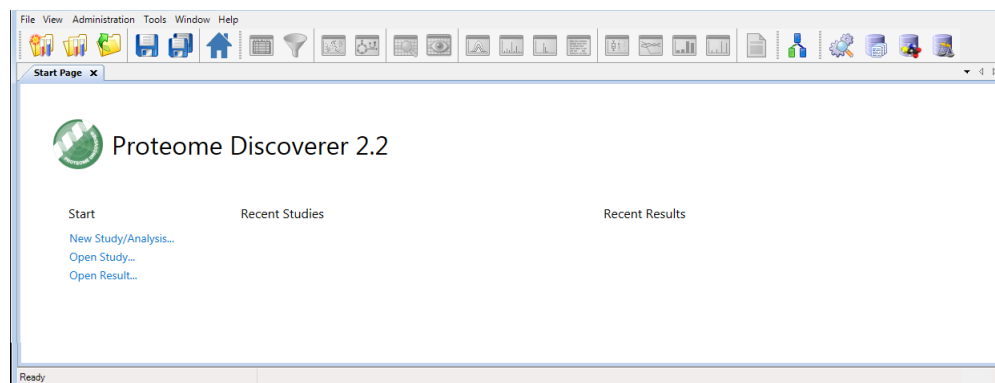
For information on the features of the Proteome Discoverer interface and how to customize them, refer to the Help.

To open a .pdResult file, refer to the Help.

To use the Start Page, see .

Closing the Proteome Discoverer Application

Save your changes before you exit the Proteome Discoverer application, because it does not prompt you.

To close a results report file before you close the application, refer to the Help.

❖ **To close the Proteome Discoverer application**

1. (Optional) Save any changes that you made to the report, study, or workflow.

2. Choose **File > Exit**, or click the **X** in the upper right corner of the main Proteome Discoverer window.

Finding Help in the Proteome Discoverer Application

You can find information about Proteome Discoverer parameters and procedures by pressing the F1 key in the application. You can also find help by using the Help menu on the menu bar.

Using the Start Page

The Start Page provides links to the most important commands used in working with the Proteome Discoverer application. It features three areas:

- Start, which contains the commands required to create a new study or analysis, open studies, and open results

- Recent Studies, which lists the 20 most recently opened studies

- Recent Results, which lists the 20 most recently opened .pdResult files

The links on the Start Page correspond to File menu commands.

- Opening the Start Page

- Opening Result Files on the Start Page

- Retaining Study Names or Result File Names on the Start Page

- Deleting a Study Name or a Result File Name from the Start Page

- Clearing the Start Page

- Closing the Start Page

Opening the Start Page

❖ **To open the Start Page**

- Start the Proteome Discoverer application (see Opening the Proteome Discoverer Application).

–or–

- Choose **View > Start Page**, or click the **Start Page** icon, 🏠 .

Opening Result Files on the Start Page

For instructions on opening .pdResult files on the Start Page, refer to the Help.

Retaining Study Names or Result File Names on the Start Page

When the number of study or file names exceeds 20 on the Start Page, the application begins removing individual files, oldest first, to keep the number at 20. However, you might want to continually work with some studies or result files over a period of time. You can prevent their names from disappearing from the lists of recent items by "pinning" them.

❖ **To retain study names or .pdResult file names on the Start Page**

1. On the Start Page, hold the mouse pointer over the name of the study or .pdResult file that you want to retain.

2. Click the pin symbol that appears to the left of the entry.

   Figure 4 shows the Start Page with pinned entries. A pinned entry still changes its position in the list, depending on its age, but it does not disappear from the list. You can remove an item from its "pinned" status by clicking the pin.

Figure 4. Pinned items on the Start Page
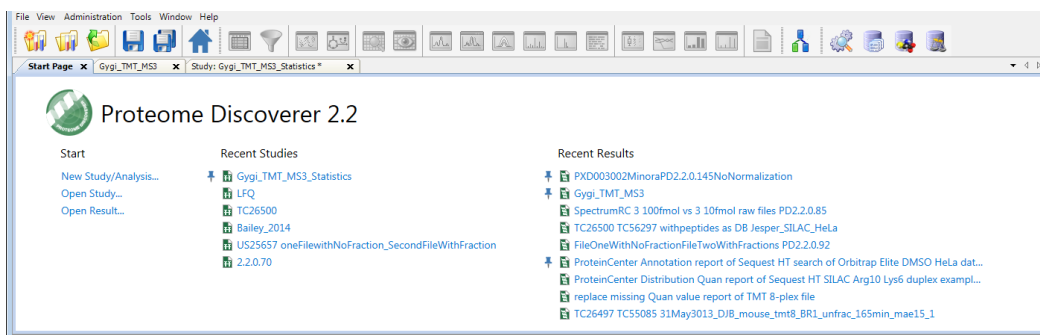
Deleting a Study Name or a Result File Name from the Start Page

You can delete the name of an individual study name or a .pdResult file name from the Start Page.

❖ **To delete an individual study or results file name from the Start Page**

1. Right-click the name of the .pdResult file that you want to delete.

2. Choose **Remove from List**.

–or–

1. In Windows Explorer, delete the study or .pdResult file.

2. On the Start Page, click the name of the study or .pdResult file that you deleted.

   An error message box opens.

3. Click **OK**.

Clearing the Start Page

You can delete all the study names or file names from the Start Page at once.

❖ **To delete all study names or results file names from the Start Page**

1. Choose **File > Recent Studies > Clear Recent Study List**, or **File > Recent Results > Clear Recent Result List**.

   The Clear Recent Study List dialog box opens for studies (Figure 5), or the Clear Recent Result List dialog box opens for result files.

   Figure 5. Clear Recent Study List
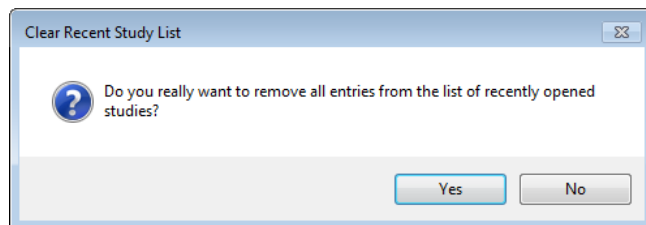
2. Click **OK**.

   All the study or results file names disappear from the Start Page and the File menu.

Closing the Start Page

❖ **To close the Start Page**

Click the **X** on the Start Page tab.

Configuring Search Engine Parameters

Before you begin your search, you can configure certain search parameters for the Sequest HT and Mascot search engines.

❖ **To configure search parameters**

1. Choose **Administration > Configuration**.

   The Administration page changes to the Configuration view. By default, this view is set to the Display Settings view (see Figure 6).

Figure 6. Default configuration view of the Administration page
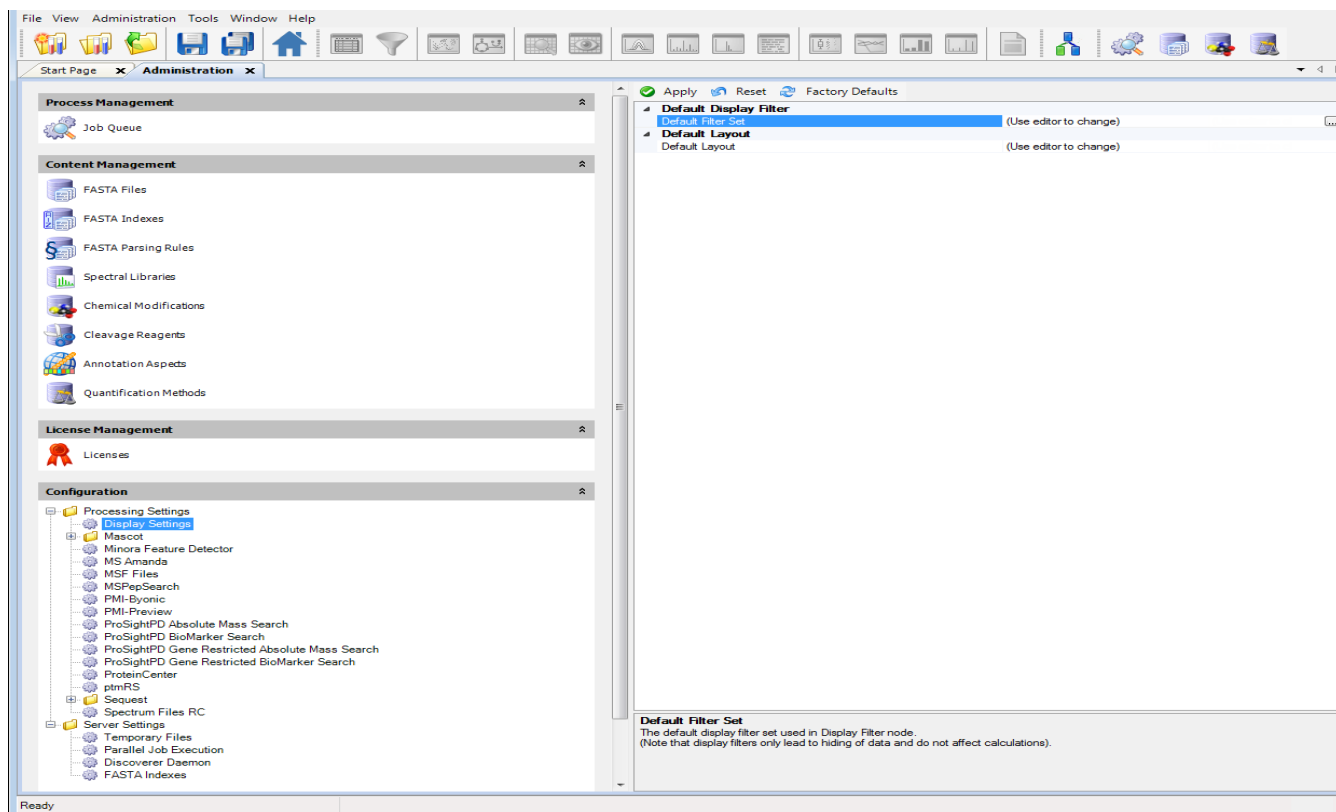
2. Follow these procedures:

- Configuring the Sequest HT Search Engine

- Configuring the Mascot Search Engine

Configuring the Sequest HT Search Engine

Before using the Sequest HT search engine, set the parameters that configure the SequestHT node and the Sequest protein-scoring calculation.

❖ **To configure the Sequest HT search engine**

1. On the Administration page, expand the **Sequest** folder under Processing Settings in the Configuration area.

2. Click **Sequest HT Node**.

All of the SequestHT node configuration parameters appear (see Figure 7).

Figure 7. Sequest HT node configuration parameters

For information about these parameters, refer to the Help.

3. In the Automatic box, specify whether you want the Proteome Discoverer application to automatically estimate the workload level.

   The default is True, which means that the application automatically estimates the workload level.

4. (Optional) If you set the Automatic parameter to False, do the following:

   a. In the Number of Spectra Processed at Once box, specify the maximum number of spectra that the Sequest HT search engine can process at once.

      The minimum value is 1000, and there is no maximum. The default is 3000.

      The larger the value, the more memory is required.

   b. In the Number of Parallel Tasks box, specify the number of search tasks that Sequest HT can perform at the same time.

      The minimum value is 0, and there is no maximum. The default is 0.

      If you set this parameter to 0, this search engine performs as many parallel tasks as the number of available CPUs can handle.

5. If you are using the Sequest HT search engine to search low-resolution data, set the XCorr confidence thresholds under the XCorr Confidence Thresholds (low-resolution data) parameter.

   The application uses these values only when you include the Fixed Value PSM Validator node in a workflow. The Target Decoy PSM Validator node and the Percolator node assign false discovery rates differently.

   The default values appear in Figure 7.

6. If you are using the Sequest HT search engine to search high-resolution data, set the XCorr confidence thresholds under the XCorr Confidence Thresholds (high-resolution data) parameter.

   The application considers these values only when the workflow includes the Fixed Value PSM Validator node.

   The default values appear in Figure 7 on page 23.

7. To apply any changes to the settings of the node configuration parameters, click
   ✔ Apply .
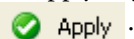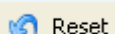
   The following message box opens (Figure 8):

   Figure 8. Administration message box
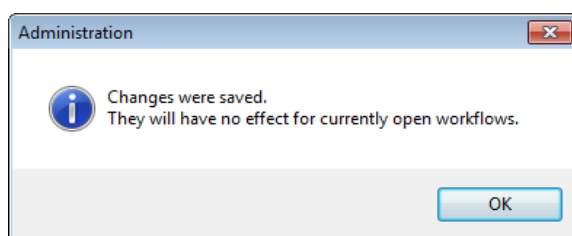
8. Click **OK**.

   > **Note** Click **↺ Reset** to return to the default values.

❖ **To configure the Sequest HT protein-scoring calculation**

1. On the Administration page, expand the Sequest folder under Processing Settings.

2. Click **Sequest Protein Score**.

   The Protein Scoring Options configuration parameter opens (Figure 9).

   Figure 9. Sequest Protein Scoring Options configuration parameter
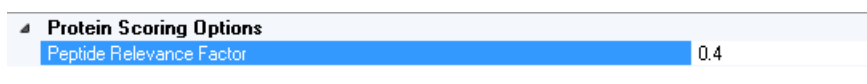
   For information on this parameter, refer to the Help.

3. In the Peptide Relevance Factor box, specify a factor to apply to the protein score.

   The Proteome Discoverer application calculates the protein score for a protein as follows:

   protein_score = (sum_of_all_cross-correlation_factors_of_0.8_or_above) + (peptide_charge × peptide_relevance_factor)

   The minimum value is 0.0, and the maximum value is 0.8. The default is 0.4.

You can review the setting of the Sequest HT protein-scoring configuration parameters by choosing View > Result Summaries, clicking the Configuration tab, clicking the Consensus WF tab, and viewing the Configuration for Protein Scorer section (see Figure 10) on the Consensus WF subpage. For information on the Result Summaries, refer to the Help.

Figure 10. Protein-scoring configuration parameter settings in the Protein Scorer section of the Search Summary
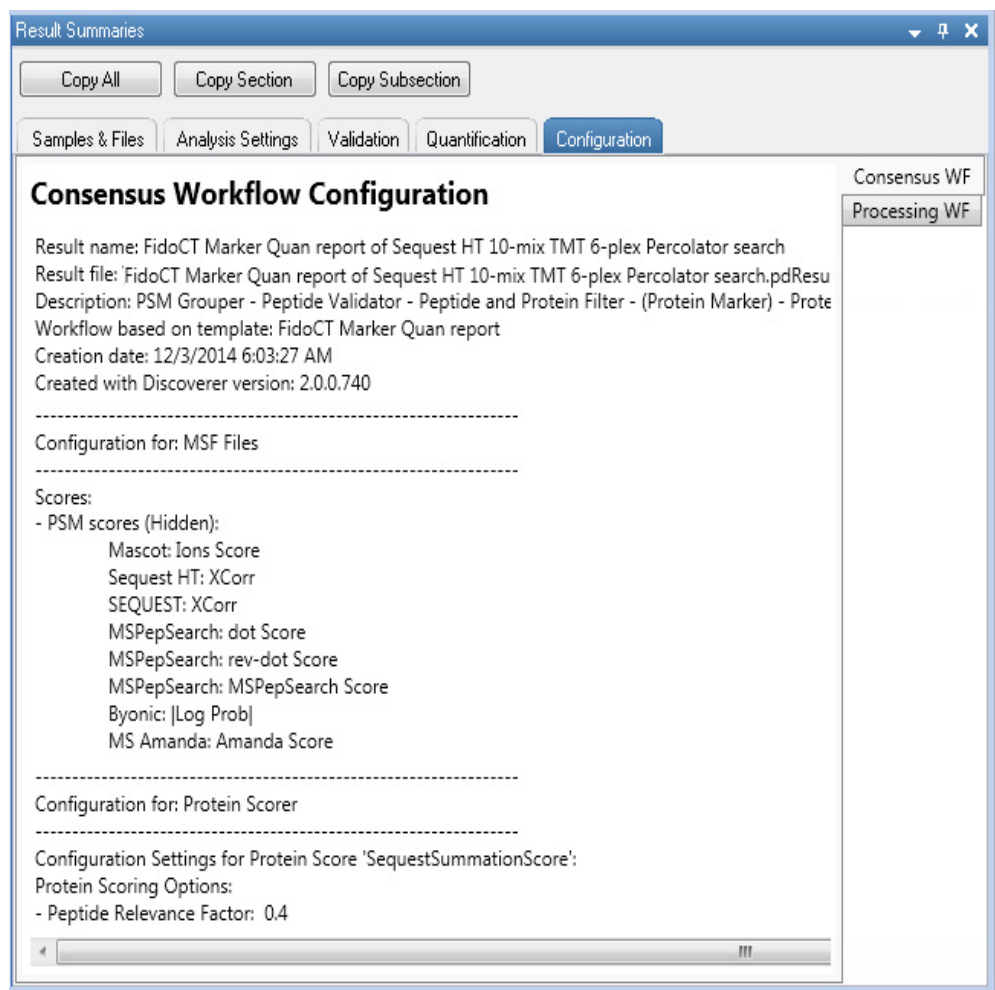
4. To apply any changes to the settings of the Sequest protein-scoring configuration parameters, click [✓ Apply].

5. When the message box opens (Figure 8 on page 24), click **OK**.

> **Note** Click [↺ Reset] to return to the default values.

Configuring the Mascot Search Engine

Before using the Mascot search engine, you must direct the Proteome Discoverer application to the location of the Mascot server and configure the parameters that control access to the server. If your Mascot search fails, the troubleshooting guidelines can help you check for server problems.

- Directing the Proteome Discoverer Application to the Mascot Server Location

- Configuring Mascot Parameters

- Troubleshooting Failed Mascot Searches

Directing the Proteome Discoverer Application to the Mascot Server Location

Open a web browser and try to access the Mascot server through its URL. If you cannot, the Mascot server might not be running, or the URL might not be correct. In this case, contact your system administrator to assist you.

To test the connection between the Proteome Discoverer application and the Mascot server, follow this procedure.

❖ **To test the connection to the Mascot server**

1. Start the Mascot Communication Tester by choosing **Tools > Mascot Communication Tester**.

   The Mascot Communication Tester window opens (see the top of Figure 11). The Proteome Discoverer application automatically populates the fields of the window, including the URL in the Mascot Server Url box.

2. If the URL does not automatically appear or you want to enter it manually, enter the URL of the Mascot server in the Mascot Server Url box.

   This is the same URL that you entered in the Mascot Server URL box in the Configuration view of the Proteome Discoverer application.

   If the Web server that hosts the Mascot server enables any user authentication, provide a user name and a password for this authentication in the corresponding boxes in the Mascot Communication Tester window.

3. Click **Test Communication**.

   An automated communication test starts (see Figure 11).

Figure 11. Test running in the Mascot Communication Tester window
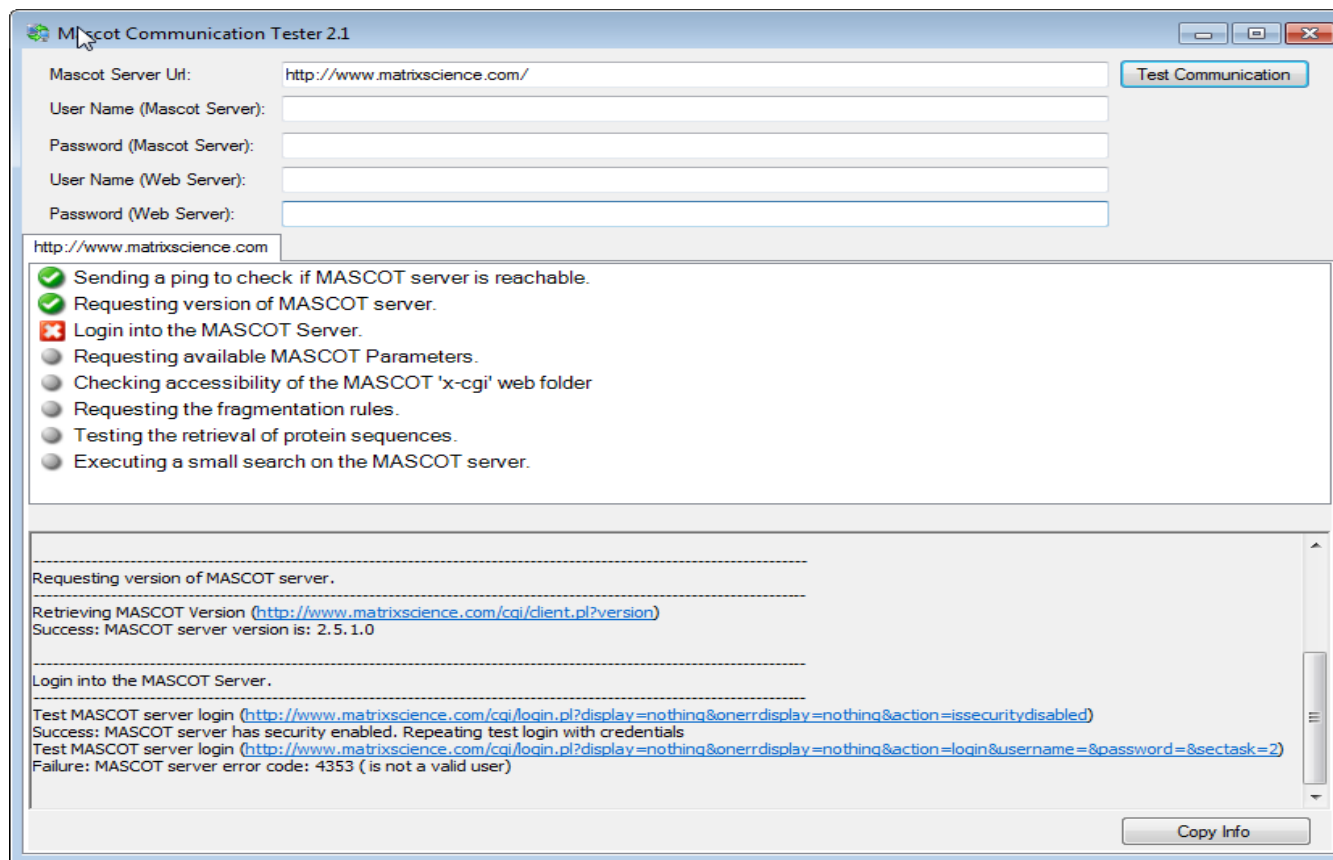
The upper pane displays the progress of the communication test, and the lower pane provides detailed information. When errors occur, the lower pane displays the particular test that failed, which helps you determine why the connection failed. You can click the blue highlighted links shown in the lower pane to test whether the communication works from within a web browser.

If this information does not resolve the problem, use the Copy Info button to copy the log information into an email and send it to Thermo Fisher Scientific Customer Service (see "Contacting Us" on page xxi).

Configuring Mascot Parameters

Before using the Mascot search engine, set the parameters that configure the Mascot server, the Mascot node, and the Mascot protein-scoring calculation.

❖ **To configure the Mascot server**
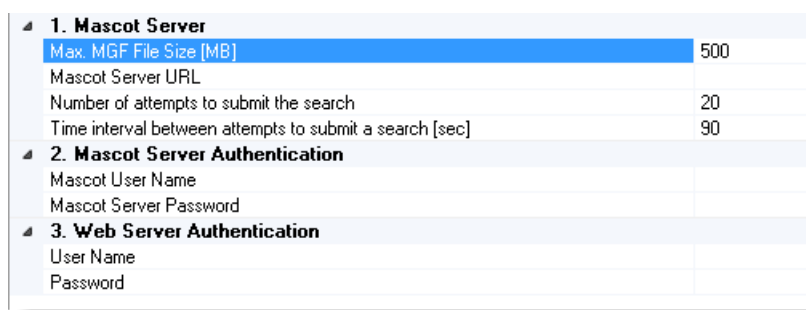
1. Choose **Administration > Configuration**.

2. On the Administration page, expand the Mascot folder under Processing Settings in the Configuration section.

3. Click **Mascot Server** to open the parameters (see Figure 12).

   The Proteome Discoverer application generates an MGF file that contains the search settings and all mass spectral information. It submits this file to the Mascot server through a web server, which might have a file size limitation. A search that generates large amounts of data—for example, a search with multiple raw data files—could create an MGF file that exceeds this limitation. The Max. MGF File Size [MB] parameter avoids this limitation by performing several separate Mascot searches and merging the results.

4. To split the MGF file and avoid any potential file-size limitations on the web server, enter the maximum size that the MGF file can be in the Max. MGF File Size [MB] box, in megabytes (see Figure 12).

   Figure 12. Maximum MGF file size for Mascot

   Make sure the size is less than the file size permitted by the web server.

   The minimum file size is 20, and there is no maximum. The default file size is 500 megabytes.

   For information on these parameters, refer to the Help.

5. In the Number of Attempts to Submit the Search box, specify the number of times that the application tries to submit the search when the Mascot server is busy.

   The minimum value is 0, and there is no maximum value. The default is 20.

6. In the Time Interval Between Attempts to Submit a Search [sec] box, specify the interval of time, in seconds, that elapses between attempts to submit a search when the Mascot server is busy.

   The minimum value is 20, and there is no maximum value. The default is 90 seconds.

7. If you are accessing a Mascot server through your own network and security for that server is enforced, enter your user name and password in the boxes under Mascot Server Authentication.

8. If you are accessing a Mascot server through the web and security for that server is enforced, enter your user name and password in the boxes under Web Server Authentication.

9. If you changed any settings of the Mascot server configuration parameters, click ✓ Apply .
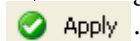
10. Click **OK** in the administration message box.

> **Note** Click ![Reset] to return to the default values.

❖ **To configure the Mascot node**

1. On the Administration page, expand the Mascot folder under Processing Settings in the Configuration section.

2. Click **Mascot Node**.

The configuration parameters for the Mascot node appear (Figure 13).

Figure 13. Configuration parameters for the Mascot node

For information on these parameters, refer to the Help.

3. Set the Default Confidence Thresholds parameters as follows:

   • Significance High: Calculates the thresholds for high-confidence peptides. The application automatically sets this value to the calculated relaxed significance when it performs a decoy search. The minimum value is 0.0, and the maximum value is 1.0. The default is 0.01.

   • Significance Middle: Calculates the thresholds for medium-confidence peptides. The application automatically sets this value to the calculated relaxed significance when it performs a decoy search. The minimum value is 0.0, and the maximum value is 1.0. The default is 0.05.

4. To apply any changes to the settings of the Mascot node configuration parameters, click ![Apply] .

5. Click **OK** in the Administration message box.

> **Note** Click ![Reset] to return to the default values.

❖ **To configure the Mascot protein-scoring calculation**

1. On the Administration page, expand the Mascot folder under Processing Settings.

2. Click **Mascot Protein Score**.

The configuration parameters for Protein Scoring Options appear (Figure 14).

Figure 14. Protein Scoring Options configuration parameters

For information on these parameters, refer to the Help.
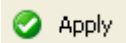
You can review these parameter settings in the Protein Scorer section of the Consensus Workflow page of the Result Summaries. (Figure 10 on page 25 shows this section for the equivalent Sequest HT parameters.) For information on the Result Summaries, refer to the Help.
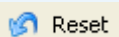
3. Specify whether to use MudPIT scoring or normal scoring.

   MudPIT scoring, an option for Mascot searches, is used for generating the protein score from data sets with large numbers of peptide matches, as in a complex proteome digest. Mascot stores 10 possible matches for each spectrum. The protein score is normally a sum of all possible matches. However, this might result in a high "protein" score that comprises numerous low-scoring erroneous matches. The Use MudPIT Scoring parameter prevents low-scoring peptide matches from contributing to a falsely reported high-scoring protein.

   • (Default) Automatic: Automatically determines whether to use MudPIT scoring.

   • True: Always uses MudPIT scoring.

   • False: Never uses MudPIT scoring.

   For information about MudPIT scoring, consult the Mascot Help.

4. If you changed any settings of the Mascot protein-scoring configuration parameters, click ✅ Apply .

5. Click **OK** in the administration message box.

> **Note** Click ↺ Reset to return to the default values.

Troubleshooting Failed Mascot Searches

If all your searches using the Mascot search engine fail, follow these instructions to locate the problem.

❖ **To troubleshoot failed Mascot searches**

1. Verify that the Mascot server is running and accessible from the computer that is running the Proteome Discoverer application. For details, see "Directing the Proteome Discoverer Application to the Mascot Server Location" on page 26.

2. If your Mascot server is running, verify that it is operating properly by submitting a simple search from the Mascot web interface. Do one of the following:

   • If the search from the Mascot web interface is successful, go to step 3.

   • If the search fails, contact your system administrator. There might be a problem with the Mascot server itself.

3. If your Mascot server is operating properly and you can access it from the Proteome Discoverer application, try to perform a very simple search using the Mascot node. Do one of the following as applicable:

   • If simple searching fails, there might be a general problem in the interaction between the Proteome Discoverer application and the Mascot server. In this case, file an error report.

   • If you can perform simple Mascot searches, investigate your failing searches more closely:

     Does the search finish successfully on the Mascot server according to the Mascot search log?

     Do the process messages sent to the job queue during the search indicate the problem?

4. If these suggestions still do not resolve the search problem, file an error report.

Configuring Parallel Job Execution Parameters

To avoid overloading your machine, you can specify how many processing and consensus workflows to execute at the same time. (For definitions of processing and consensus workflows, see "Workflow Editor" on page 34.) You must restart the application after you set the appropriate configuration parameters.

> **IMPORTANT** Use caution when setting these parameters. Workflows share resources other than processors, such as hard disks or network connections, so running many jobs in parallel could cause Proteome Discoverer, other applications, or the operating system to become unresponsive.

❖ **To specify the number of workflows to execute in parallel**

1. Choose **Administration > Configuration**.

2. In the Configuration area in the lower left corner of the Proteome Discoverer window, click the + symbol to the left of Server Settings.

3. Click **Parallel Job Execution**.

4. In the Max. Number of Processing Workflows in Parallel Execution box, specify the number of processing workflows to execute at the same time.

   Default: 1

   Range: 1 to half the number of dual-processor cores on your machine.

   For example, if your machine has a dual-core processor with hyperthreading, it has four logical cores. The maximum value for the parallel execution of processing workflows would therefore be half of four, or two.

5. In the Max. Number of Consensus Workflows in Parallel Execution box, specify the number of consensus workflows to execute at the same time.

   Default: 1

   Range: 1 to half the number of dual-processor cores on your machine

   For example, if your machine has a dual-core processor with hyperthreading, it has four logical cores. The maximum value for the parallel execution of consensus workflows would therefore be half of four, or two.

6. Click ✓ Apply .

   The Settings Applied confirmation box opens (Figure 15).

   Figure 15. Settings Applied confirmation box
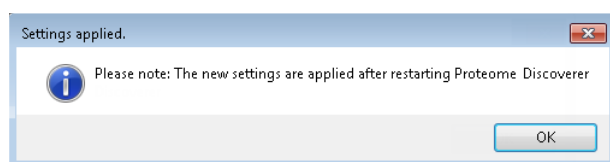
7. Save any appropriate files, and restart the Proteome Discoverer application.

Configuring Temporary Files Parameters

The Proteome Discoverer application generates a number of temporary files until it creates the final result file, especially when experiments involve label-free and precursor ion quantification. If you are concerned that the input raw data files and the temporary result files generated by your experiment might be too large for the default directory on your hard drive, you can select another drive and folder to store the temporary files in.

By default, the application stores temporary files in the C:\ProgramData\Thermo\Proteome Discoverer 2.2\Scratch directory.

❖ **To select a temporary folder**

1. Choose **Administration > Configuration**.

2. Under Server Settings, click **Temporary Files**.

3. In the New Directory box (Figure 16), type the drive, path, and name of the folder where you want to store the temporary files, or browse to the location of the folder and click **OK**.

   The new folder must be named Scratch. If it is not, the application creates a subfolder named Scratch.

   The new temporary folder location must be accessible, and you must have write permission. To avoid slowing down the data processing, ensure that the read/write performance is high. Because storage on USB drives or network shares might cause slow performance, use local drives for storage instead.

Figure 16. Temporary Files pane
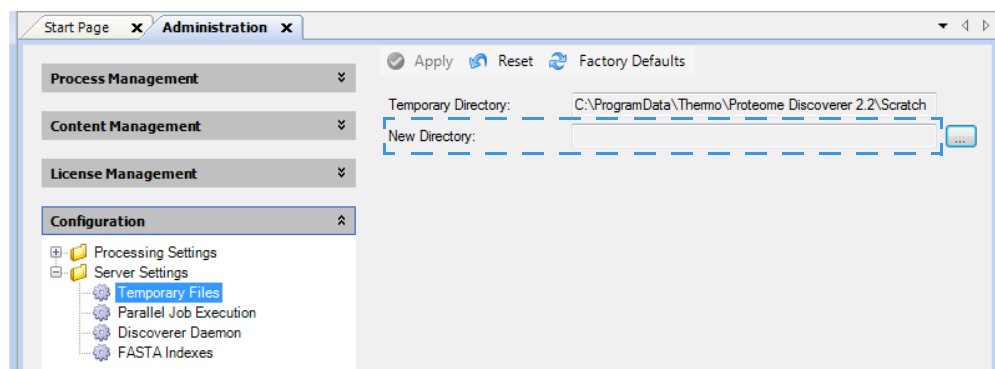
4. Restart the application.

   If the application encounters a problem during the initialization of the temporary folder, it displays an error message in the Temporary Files pane (see Figure 17). When you use an invalid Scratch folder as the current directory, all job processing fails.

Figure 17. Error message in the Temporary Files pane
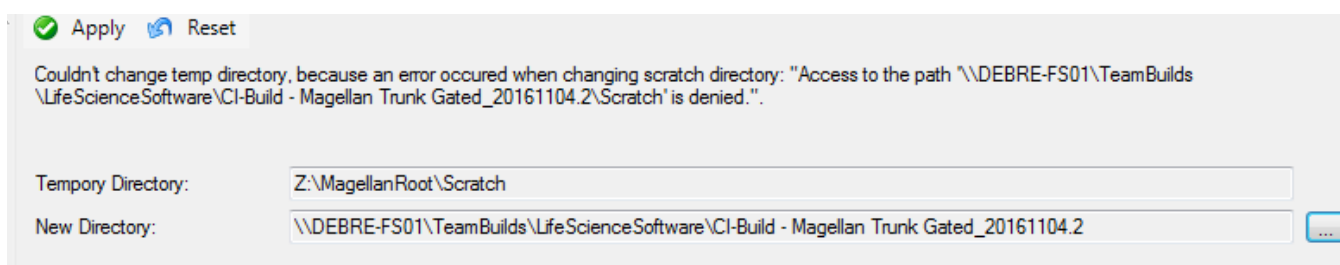
Downloading the FASTA Files

You must add a FASTA file to the Proteome Discoverer application before you can conduct a search with the Sequest HT.

If you are using the Mascot search engine, you must ask your administrator to download a FASTA file or files to the Mascot server.

You can download FASTA files from two sources:

• You can download a controlled protein database directly from ProteinCenter and add it to a FASTA file. These controlled databases offer you access to the latest protein information. The ProteinCenter database service provides extensive information about proteins, peptides, and genes. In addition, it gives you the ability to select proteins of a specified taxonomy to download and use as a FASTA file.

• You can add a FASTA file that you download from other sources onto your hard drive and then register it.

For complete instructions, see "Downloading FASTA Files to Proteome Discoverer" on page 153.

Studies in the Proteome Discoverer Application

The first step in conducting a search in the Proteome Discoverer application is to create a study. A study is primarily a container that you use to structure and annotate your data with meta information, such as how you treated the samples, how you performed the chromatography, or how you set up the data acquisition. A study also contains the input files, the information about the samples contained in those files, and their treatment information. It also keeps track of your processed results.

You can use a study to organize all the quality control files that you run regularly; to track all input and result files when you test different sample preparations, chromatography, or acquisition settings; or to organize true biological experiments with technical and biological replicates, different treatments, time points, and so forth.

To create a study, see "Using Studies" on page 38.

Analyses in the Proteome Discoverer Application

An analysis in the Proteome Discoverer application is the actual data processing in the context of a study. It consists of the different data processing and reporting steps to perform, their associated workflows, their parameter settings, their input files, and other information related to how your data is being processed. A study can contain multiple analyses.

An analysis consists of multiple data-processing steps in the form of a workflow. The processing step and the consensus step are described in the next topic, "Workflow Editor."

To create an analysis, see "Using Analyses" on page 64.

After you have set up an analysis, you can save it as a .pdAnalysis file, which you can then use as a template to set up new analyses.

Workflow Editor

The Workflow Editor is both a stand-alone window in the Proteome Discoverer application and a window included in a study. It is a flexible, complex tool that you can use to create customized searches and a customized results report. Use it to create one workflow to perform the search and another workflow to produce the results. Each workflow consists of a string of nodes that you can choose from a group of nodes unique to that workflow. You can further customize each workflow by setting parameters for each node.

In the Workflow Editor, you create the following two workflows:

- A processing workflow, which processes data in one or more RAW data files, generates the primary search results (for example, the results of a sequence database search by an engine such as Sequest HT) or extracting the raw quantification values (for example, the extracted reporter peak intensities where samples use isobaric tags). The processing workflow accepts spectrum container files such as RAW and MZML files as input and produces MSF files as output.

- A consensus workflow, which collects and assembles the data from one or more processing workflow results. The consensus workflow accepts an MSF file or files as input and produces a .pdResult file as output. The application can create different .pdResult files from the same MSF file and can combine different MSF files into one .pdResult file called a multiconsensus report.

You can create a reusable processing and consensus workflow template by saving your design to load and use at another time.

As a stand-alone window, the Workflow Editor is intended for use when you just want to edit or create a workflow template for later reuse.

Figure 18 shows the stand-alone Workflow Editor.

Figure 18. Workflow Editor

Parameters pane

Workflow Tree pane

Workflow Nodes pane
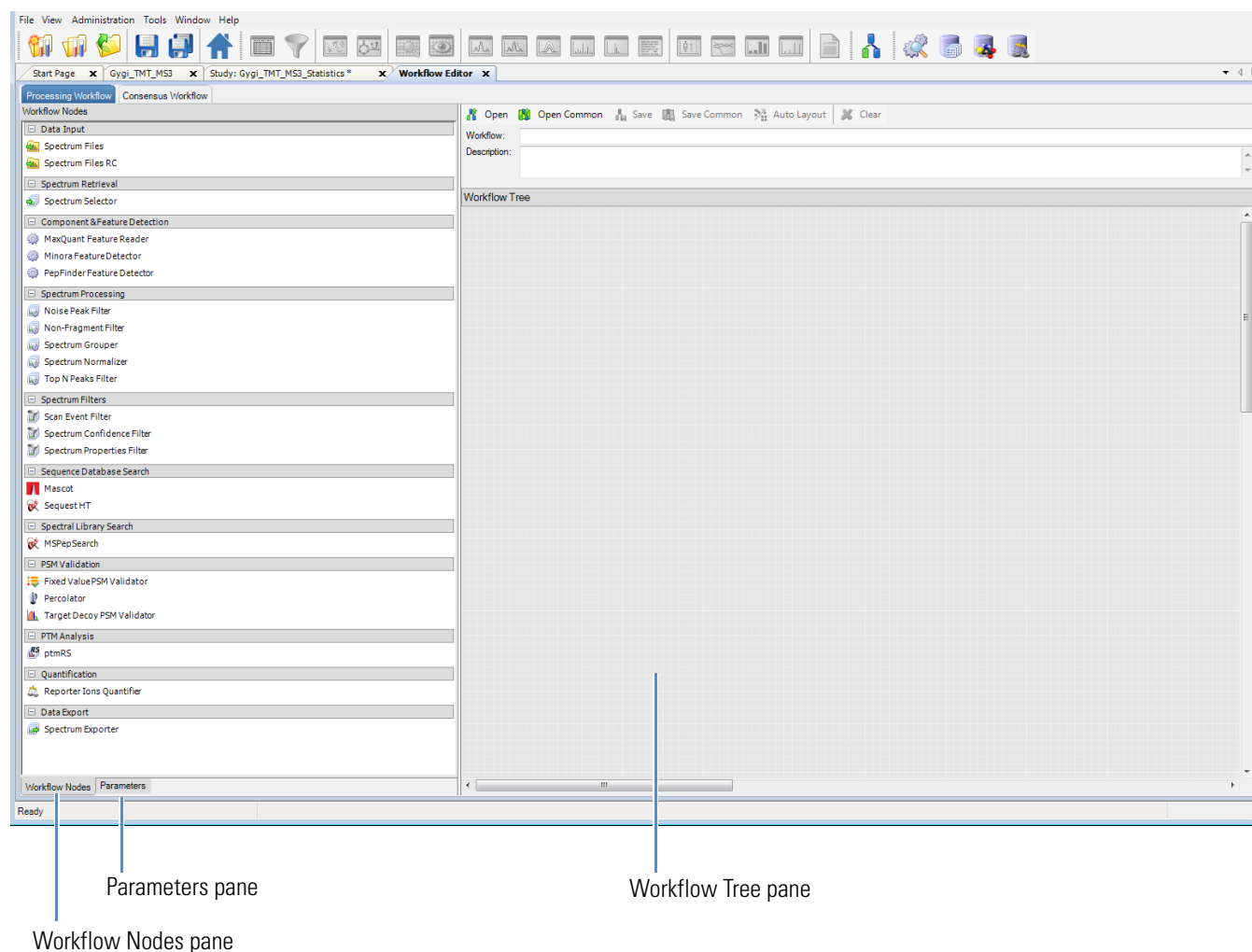
The Workflow Editor consists of the following features:

- Workflow Nodes pane—Contains the nodes available for a processing workflow on one page, Processing Workflow, and the nodes available for a consensus workflow on another page, Consensus Workflow.

- Parameters pane—Displays the parameters available for the selected node.

- Workflow Tree pane—Where you connect nodes together to create a processing or consensus workflow.

- Post-Processing Nodes pane—Where you place the Post-Processing nodes, which process the results generated by the entire workflow. This pane is only visible when you create a consensus workflow.

For instructions on using the Workflow Editor to create a workflow, see "Using the Workflow Editor" on page 97.

Performing a Search

Follow these topics to perform a search.

- Before Performing a Search

- Using Studies

- Using Analyses

- Specifying Quantification Ratios from Selected Sample Groups

- Performing the Search

- Working with the Search Results

- Using the Workflow Editor

- Using Workflow Templates

- Creating Specific Types of Workflows

- Searching Multiple Sequence Databases with Mascot

- Creating a Multiconsensus Report

Before Performing a Search

Follow these steps before using the Workflow Editor either as a stand-alone tool or in the context of a study to create a workflow:

- Configure the search engine that you will use. See "Configuring the Sequest HT Search Engine" on page 22 or "Configuring the Mascot Search Engine" on page 26.

- If you are using the Sequest HT search engine, download a FASTA file or files if necessary. See "Downloading FASTA Files to Proteome Discoverer" on page 153.

  If you are using the Mascot search engine, you must ask your administrator to download a FASTA file or files to the Mascot server.

- Make spectrum source files available as RAW, MGF, MZDATA, MZXML, or MZML files.

Using Studies

This topic explains how to create a study and how to use the Proteome Discoverer application to perform a database search. For illustrative purposes, it uses an example study called Bailey_2014, which uses a publicly available data set from the Chorus Project (https://chorusproject.org)[1]. This project resides under the Elution Order Algorithm project and includes data about the following.

- Two sets of biological replicates

  Four mice (replicates) were sacrificed and dissected. Individual organs of interest from them were homogenized, and the proteins were extracted from them and labeled. Then the differentially labeled organ-specific proteomes were mixed together if they came from the same mouse. Each pooled mouse sample was then run twice, using different acquisition method parameters.

  For information on replicates, see "Technical and Biological Replicates" on page 643.

- The TMT 8plex quantification method

- Label switching

- One biological factor: different tissues

- One technical factor: different acquisition methods

Figure 19 shows how different tissue samples are distributed over four biological mouse replicates.

Figure 19. Distribution of different tissue samples over four biological mouse replicates

---
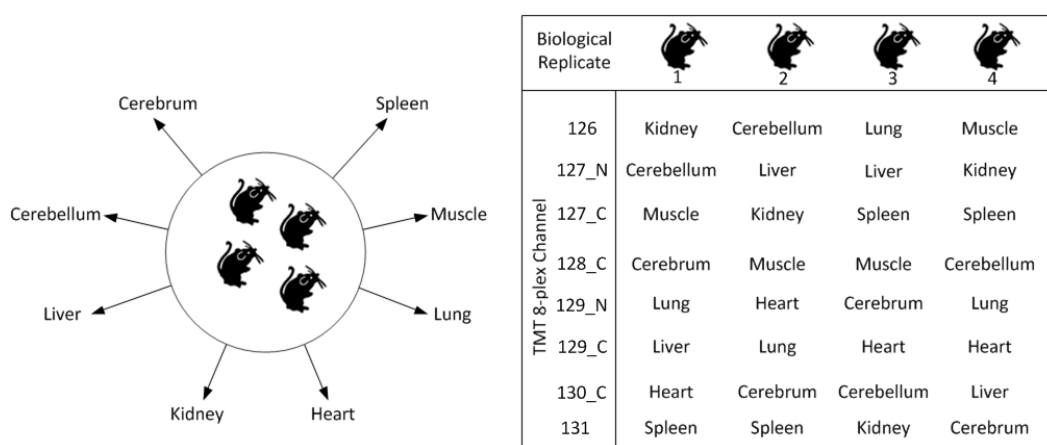
[1] Bailey, D.J.; McDevitt, M.T.; Westphall, M.S.; Pagliarini, D. J; Coon, J. J. Intelligent data acquisition blends targeted and discovery methods. *Journal of Proteome Research*, **2014**, 13 (4): 2152–2161.

See these topics:

- Adding a Quantification Method to the Study

- Opening an Existing Study

- Creating a Study

- Adding a Description

- Adding Input Files

- Importing MSF and .pdResult Files

- Specifying the Quantification Method for Multiple Input Files

- Setting the Factor Values for the Samples

- Setting Values for Multiple Samples at the Same Time on the Input Files Page

- Setting Values for Multiple Samples at the Same Time on the Samples Page

- Saving a Study

- Copying a Study to Another Computer

- Study Window Parameters

Adding a Quantification Method to the Study

If you are performing reporter ion quantification or precursor ion quantification, you must add a quantification method to the study. See the appropriate procedure:

- For reporter ion quantification, see "Creating a Quantification Method for Reporter Ion Quantification" on page 352.

- For precursor ion quantification, see "Creating a Quantification Method for Precursor Ion Quantification" on page 417.

Label-free quantification does not require you to associate a quantification method with the raw data file or files.

Setting the Study Options

You can set options that apply to all studies and that do not need to be changed for each study.

1. Choose **Tools > Options**.

2. In the Options dialog box, click **Study Options**.

3. (Optional) Set the following options:

    - Auto Save When Starting Analysis—Determines whether the application automatically saves the contents of a study when you click Run .

- Enable Study Variables Without Multiple Values—Determines whether you can select all study variables or only those with multiple values in the Grouping & Quantification page. Study variables with just a single value are displayed in italics.

- Show Warning When Running a Quan Analysis Without Specified Ratios—Determines whether the application displays a warning if you start an analysis without creating any ratios in the Grouping & Quantification page.

4. Click **OK**.

Opening an Existing Study

Now you open an existing study or create a study. (See "Creating a Study" on page 40.)

❖ **To open an existing study**

- In the Recent Studies area of the Start Page, click the study name.

- In the Recent Studies area of the Start Page, right-click the study name and choose **Open**.

–or–

- Choose **File > Open Study**, or click the **Open Existing Study** icon, 📁 , browse to the study folder, select the .pdStudy file, and then open it.

–or–

1. In the Start area of the Start Page, click **Open Study**.

2. In the Open Study dialog box, browse to and select the name of the study, and click **Open**.

Creating a Study

Note If you want to use a custom quantification method, create it before you create or open a study.

- For reporter ion quantification, see "Creating a Quantification Method for Reporter Ion Quantification" on page 352.

- For precursor ion quantification, see "Creating a Quantification Method for Precursor Ion Quantification" on page 417.

Label-free quantification does not require you to associate a quantification method with the raw data file or files.

❖ **To create a study**

- On the Start Page, click **New Study/Analysis**.

- Choose **File > New Study/Analysis**.

–or–

- Click the **Create New Study/Analysis** icon, .

  The New Study and Analysis dialog box opens (see Figure 20).

  Figure 20. New Study and Analysis dialog box
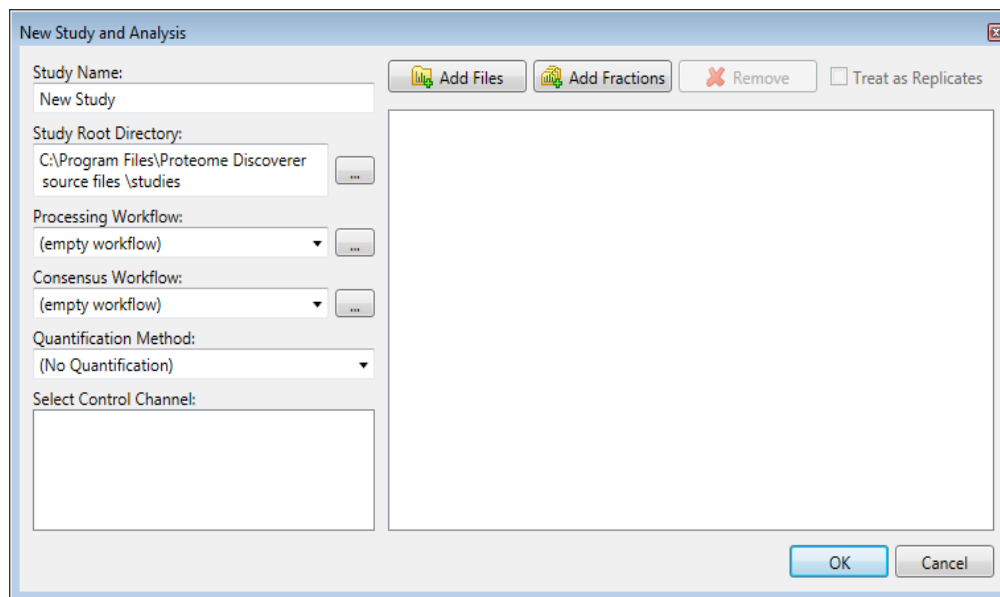
  In this example, you only specify the name of the study and a root directory to save the study in.

3. In the Study Name box, specify the mandatory study name.

   The application generates a default study name by searching for the common part of the file names when you add multiple files at once and using this common part as the suggestion for the name of the new study.

4. In the Study Root Directory box, specify the folder where you will store the study folder. Click the Browse button (...), and in the Select Folder dialog box, specify the folder and click **Select Folder**.

5. Click **OK**.

   The application creates a new study folder in the folder that you specified as the root directory and opens a new page with the study name (Study: Bailey_2014 in the example); see Figure 21. It appends a .pdStudy extension to the study file name.

Figure 21. New Study: *Study_name* page
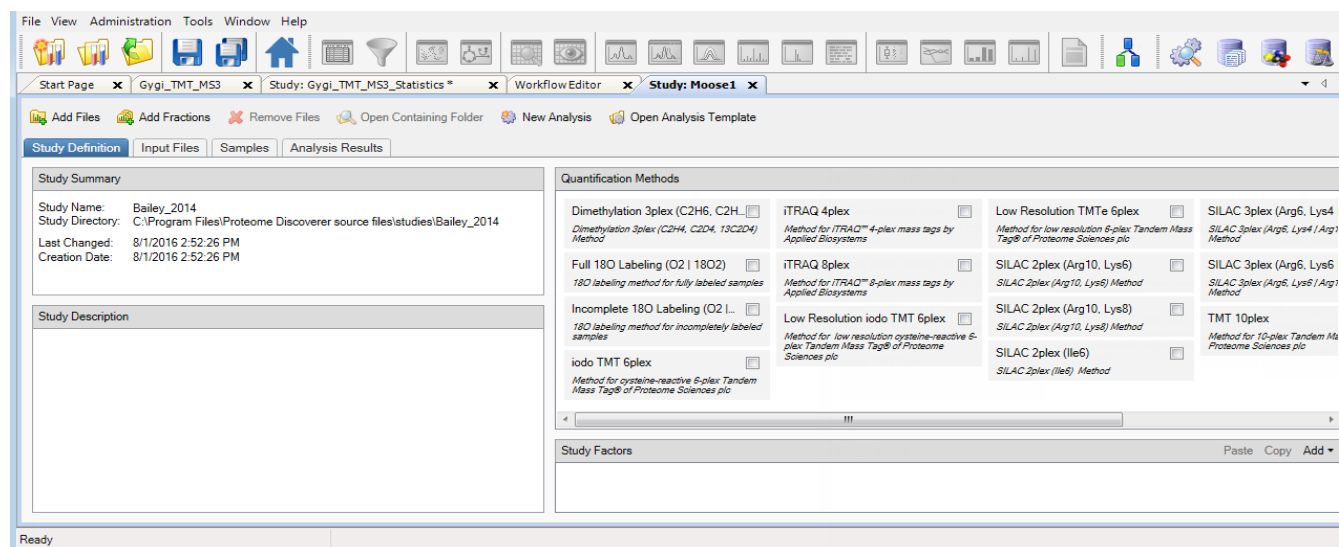
On the Study Definition page, you add a description of your study, select the quantification method or methods to use in the study, and set up the new factors to use to describe and distinguish your samples.

Adding a Description

You can optionally add a description of the study by typing it in the Study Description area of the Study Definition page.

Adding Input Files

Add the input files from the example data set to your study. (You can add input files to the study at any point.) For the types of input files supported, see "Inputs" on page 10. You can add individual input files, multiple unrelated input files, or fractions.

Adding a Single Input File or Multiple Unrelated Input Files

❖ **To add a single input file or multiple unrelated input files**

1. Click .

2. In the Add Files dialog box, browse to the location of the input files, and select them.

> **Note** You can also add existing MSF files to a study. The application adds the MSF file as a new result to the Analysis Results page and adds the raw data files that were processed to create the MSF files to the Input Files page.

3. Click **Open**.

—or—

- Drag the input file or files from Windows Explorer and drop them onto the Input Files page.

  The input files appear on the Input Files page (see Figure 22). Each file on the page receives a unique identifier: F1, F2, …, F*n*. The Proteome Discoverer application adds each file as a single study file.

Figure 22. Input files on the Input Files page
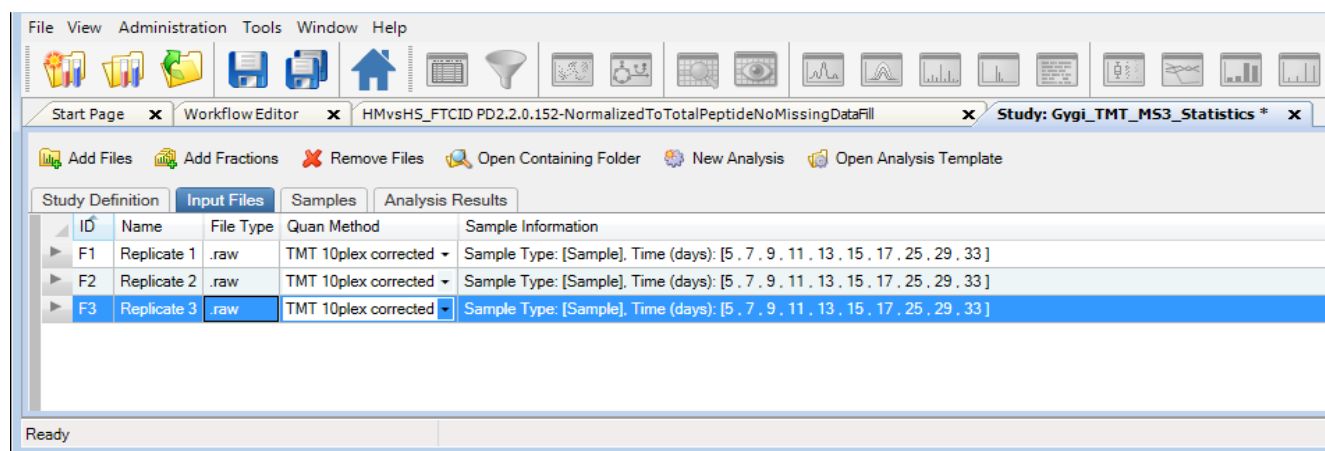
Adding Fractions

Fractions are raw data files derived from a sample that was separated into multiple vials, with each vial subjected to different treatment. For example, a sample could be divided into eight vials, with a different pH level applied to each vial.

One type of fraction is a MudPIT file. MudPIT experiments investigate complex proteomes by applying multidimensional chromatography to the samples before acquisition in the mass spectrometer. Typically, this process results in several dozen or even a few hundred fractions that are separately analyzed by LC/MS, resulting in one raw data file per sample fraction. Analyzing gel slices or performing in-depth follow-up acquisitions also results in multiple fractions. Because all these fractions belong to the same sample, the Proteome Discoverer application can process all raw data files from these fractions as one contiguous input file, and it generates only one result file.

You can search MudPIT fractions in the application by using the Add Fractions button, Add Fractions , or in the Proteome Discoverer Daemon utility by using the MudPIT button. However, Thermo Fisher Scientific recommends that you use the Proteome Discoverer application to process local batch or MudPIT samples, and use the Proteome Discoverer Daemon utility only when you need remote access and automatic processing from the command line—for example, when you use the Xcalibur data system to send data to a remote Proteome Discoverer computer (host) for acquisition. For information on processing MudPIT and batch samples with the Proteome Discoverer Daemon utility, see "Using the Proteome Discoverer Daemon Utility" on page 123.

❖ **To add fractions**

1. Click ![Add Fractions] .

2. In the Add Fractions dialog box, browse to the location of the input files, and select them.

3. Click **Open**.

   The Proteome Discoverer application adds all selected files as a single study file.

Creating Subsets of Fractions

You can optionally use only parts of a file set in a fractionated sample by re-adding them as a subset on the Input Files page.

❖ **To create subsets of fractions**

1. Add the original input files as follows:

   a. On the Input Files page, click ![Add Fractions] .

   b. In the Add Fractions dialog box, browse to the directory where the fraction files are located, select the fraction files, and click **Open**.

      The fraction files appear as a file set (F29) on the Input Files page (see Figure 23).

Figure 23. Fraction files added to the Input Files page as file set 29
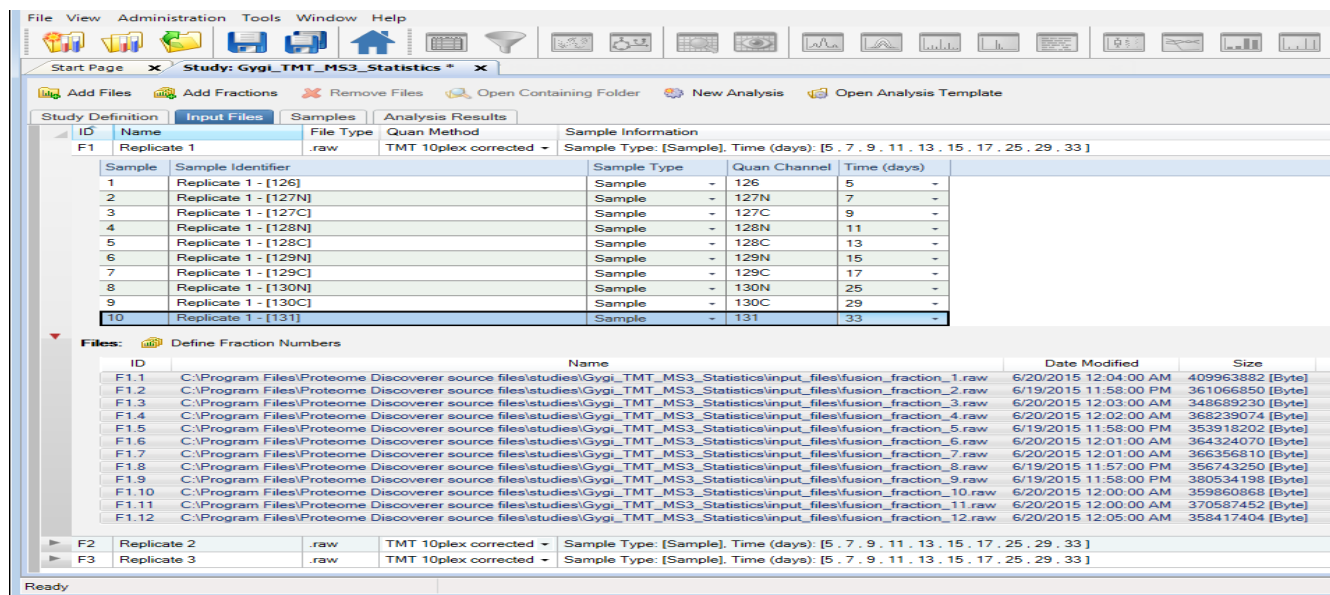
2. Re-add some of the fraction files as follows:

   a. Click ![Add Fractions] .

   b. In the Add Fractions dialog box, browse to the same directory of the original fraction files, select the fraction files that you want to re-add, and click **Open**.

The Proteome Discoverer application adds the selected fraction files to the original file set as a subset. Figure 24 shows three of the original fraction files added to the file set as a subset of F29 called F30.

Figure 24. Three fraction files re-added as a subset of F29
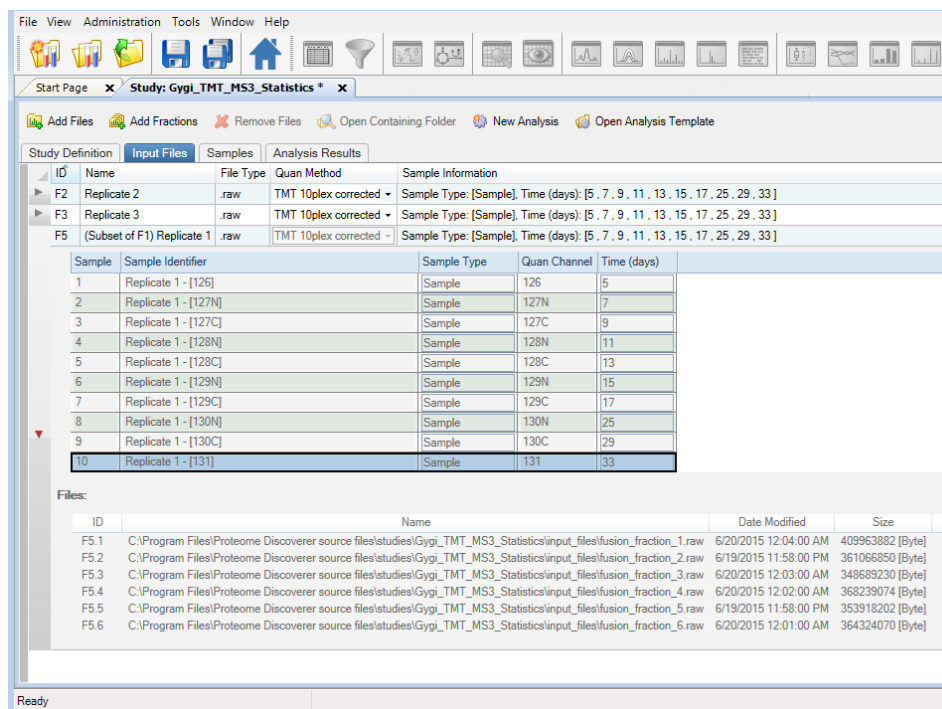
You can create any number of subsets.

When adding fractions to a study, the application first checks to see whether the study already contains any new files whose location changed within the study. If it finds these kinds of files, it updates the file locations but does not create new sample files. For example, suppose that the study contains a file called *drive*:\xxx\file1. You add file1 through file5. The application updates the path of the file1 file but does not create a new sample file.

If all selected files are new—that is, if the study does not already contain the files—the application creates a new sample file set that contains all the selected files. For example, suppose that you want to add file1 through file5, and all these files are new. The application creates a new sample file containing files file1 through file5.

If the study contains at least one new file, the application checks to see how many existing study files contain new files that are not subsets. For example, suppose that the existing sample file consists of all new files. You add file1 through file5. The study contains a sample file set called F1 consisting of file1 through file5 and no other files. In this case, the application does nothing.

As another example, suppose that an existing sample file set called F1 contains all new files but also some others, for instance, file1 through file8. In this case, the application creates a new subset containing file1 through file5 and names it [Subset of F1] F2… .

As a another example, suppose that an existing file set called F1 contains only new files, such as file1 through file3. The application creates a new file set, F2, which consists of file1 through file5 and is actually a superset of F1. But it adds F2 as a subset to the original F1 file set. The F2 subset added to F1 is called [Subset of F2] F1… .

As a final example, suppose that an existing file set called F1 contains only new files, such as file1 through file3. You want the application to create a new file set, F2, which consists of file1 through file3 and file6, as a superset of F1. But because file4 and file5 are missing, the application issues an error message and takes no action.

The application also does not support multiple file sets containing new files. When it encounters this situation, it issues an error message and takes no action.

Assigning the Order of Fractions

You can specify the order of the fraction numbers for the files in an input file set. Because files from different fractions can be very different, this step is required for label-free quantification. In addition, you must perform retention-time alignment and feature mapping only between files having the same fraction number.

The Input Files page of a study displays the fraction numbers. It consists of a table that lists the identification number (ID) of a file, the path and name of the file, modification date, and size of each file in the file set. The identification number in the ID column follows this format:

F*fileID*.*fraction*

where:

- *FileID* is the file number. For example, in the file identification number F21.3, 21 is the file number.

- *Fraction* is the fraction number. For example, in the file identification number F21.3, 3 is the fraction number.

❖ **To assign fraction numbers**

1. Add the file set to the study. See "To add fractions" on page 44.

2. Click the **Input Files** page of the study if that page is not already visible.

3. Click **Define Fraction Numbers** on the Input Files page (see Figure 25).

   This button is available only when the input file set contains multiple files and when the file set is not a subset of another file set.

Figure 25. Input Files page displaying fractions
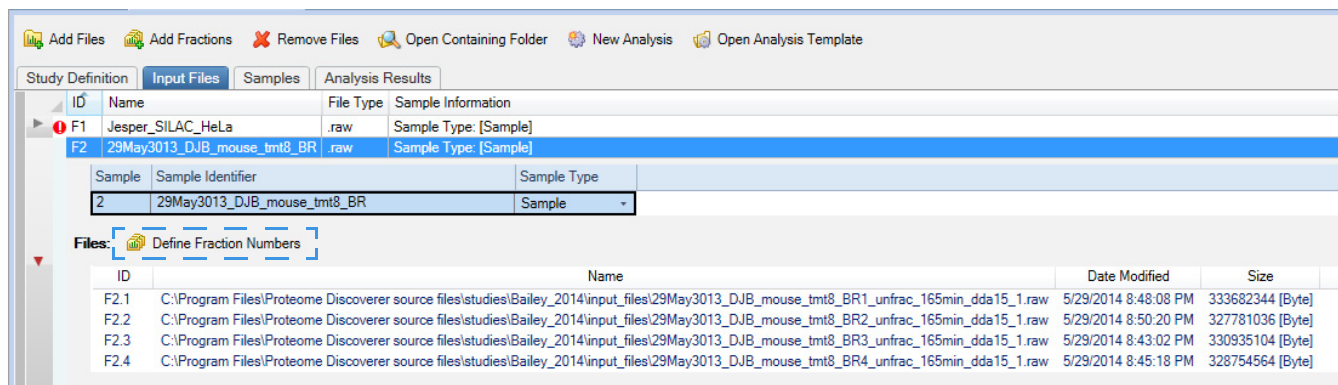
The Specify File Fractions dialog box opens (Figure 26), displaying all the files in a file set. The files in this example are not in numerical order.

Figure 26. Specify File Fractions dialog box displaying files not in numerical order
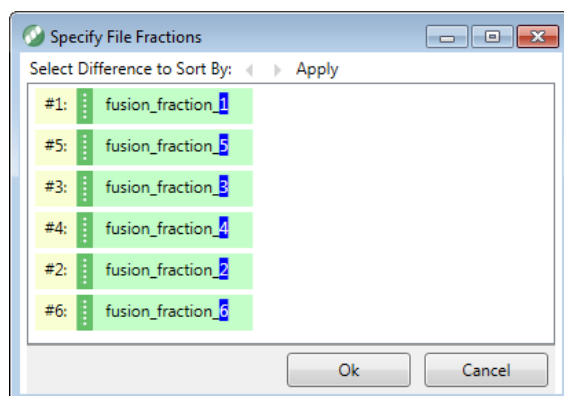
4. Place the files in order by doing one of the following:

   • Click **Apply** to sort the fractions automatically by the selected file part.

   The application sorts file names numerically when all the selected file parts start with a numeric value; otherwise, it sorts them alphabetically.

   • Use the arrow keys to the left of Apply, ◄ ►, to highlight the part of the file name to sort by.

   When there are multiple different file parts, you can choose a file part by clicking the arrows; the selected file part is highlighted in blue, while other parts are highlighted in gray.

   –or–

   • Manually change the fraction numbers by dragging the file names. Use the variable placement handle (▦) to drag each file.

   The files are now in numeric order (Figure 27).

Figure 27. Specify File Fractions dialog box displaying files in numerical order
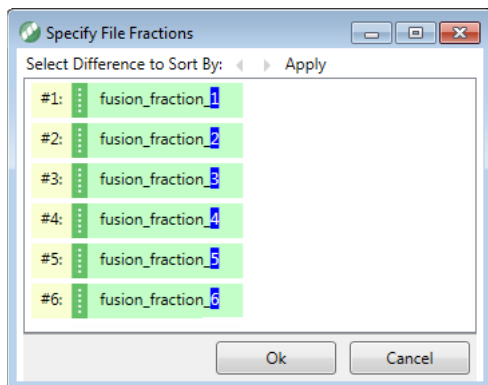
5. Click **OK** to closes the dialog box and store the new fraction numbers, or click **Cancel** to restore the previous fraction numbers.

Importing MSF and .pdResult Files

You can add existing MSF and .pdResult files to the study.

When you add an MSF file, the application adds the input files of the MSF file as fractions. If it cannot, you cannot add the MSF file to the study.

When you add a .pdResult file, the application adds the input files from each MSF file of the .pdResult file as one fraction. For example, suppose that a .pdResult file was created from two MSF files, msf1 and msf2. Msf1 was created from file1 through file4, and msf2 was created from file5 through file8. The application adds file1 through file4 as one file set and file5 through file8 as a second file set. If the application cannot do this, you cannot add the .pdResult file to the study.

Specifying the Quantification Method for Multiple Input Files

In this step, you specify the quantification method that was used for each of the files. In the example data set, all samples were labeled with the TMT 8plex custom quantification method.

❖ **To set the quantification method for each of the input files**

1. Click the **Input Files** tab if it is not already selected.

2. In each sample row, click the Quan Method column and select the quantification method (for this example, **TMT 10plex**) from the list. See Figure 28.

Figure 28. Setting the quantification method for each raw data file

Setting the Factor Values for the Samples

When you select a quantification method for a file, the Proteome Discoverer application generates a sample placeholder for each quantification channel.

The term "sample" refers to a distinct measurement of an analyte. At a minimum, a RAW data file contains at least a single sample, or in the case of sample multiplexing, it uses isobaric or metabolic labeling for multiple samples. If you measure the analyte from the same vial in two acquisitions—that is, if you measure two technical replicates—the application considers them two different samples. If you split the analyte, label each with a different isobaric tag, mix them, and acquire them into a single RAW data file, the application considers the two differently labeled parts as two different samples.

Each sample is associated with one of the following sample types: (Currently only quantification uses sample types.)

- Sample: A specimen from a larger biological entity.

- Control: A sample typically used as a reference sample in a quantification experiment.

- Blank: A sample consisting only of solvent and no sample mixture.

- Standard: A sample consisting of a standard quality-control peptide mixture.

You can create ad hoc relative quantitative ratios between any study factors or sample types, for example, blank/standard, control/sample, blank/control, and standard/blank. You can leave all of the channels as samples or make all of the channels any combination of sample, control, blank, or standard.

The Sample Type column on the Samples page of the study displays the sample type of each sample. The default sample type is Sample. If a file has samples for different quantification channels, mark one of the samples Control. In this example, the channel that was used to label the mouse kidney tissues is used as the control for generating the quantification ratios.

You can also use the control channels in scaling. For more information, see "Normalizing Peptide Groups and Protein Abundances" on page 535.

Each sample is associated with a quantification channel shown in the Quan Channel column and with values for each of the factors that you specified for your study. Previously, you specified a factor for the acquisition method used and a factor for the tissue that was extracted and labeled. You now set the correct factor values for each of the samples in the study.

Each sample has an automatically generated sample name composed of the raw data file name and the appended name of the quantification channel. You can change this name, but the name must be unique among all samples in the study.

❖ **To view the samples**

1. On the Input Files page, click the gray arrow to the left of a sample to display its constituent file entries.

   A hierarchical view opens, showing the samples contained in a raw data file. For each of the raw data files in the example, there are eight samples for the eight quantification channels of the TMT 8plex method. See Figure 29.

Figure 29. Samples in a raw data file
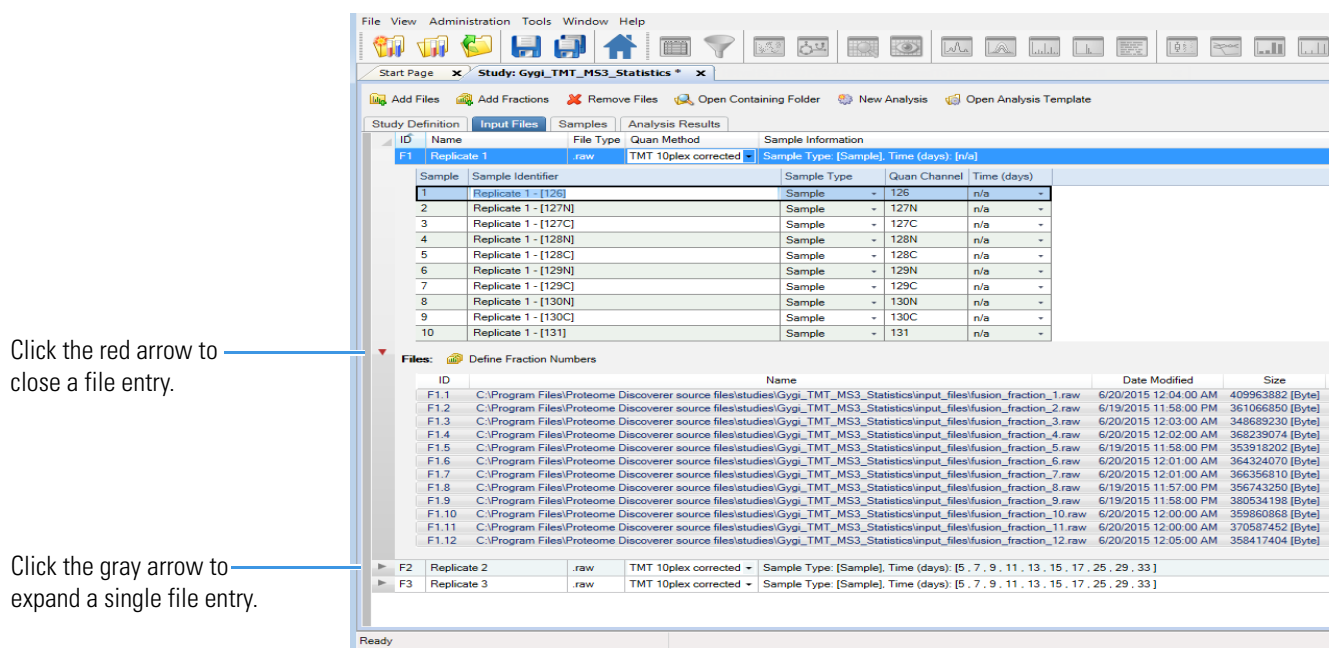
Click the red arrow to close a file entry.

Click the gray arrow to expand a single file entry.

❖ **To set the factor values for the samples**

1. Click the **Input Files** tab if it is not already selected.

2. Click the gray arrow next to the first sample to expand the information about the sample.

3. For the first factor, set the value for each sample in each raw data file by selecting the down arrow in the factor column and then selecting the value from the list.

4. For the second factor (in the example, Tissue), set the value for each sample in each raw data file by selecting the down arrow in the factor column and then selecting the value from the list.

Figure 30 shows this process for the Tissue factor.

Figure 30. Setting the factor values for the Tissue factor
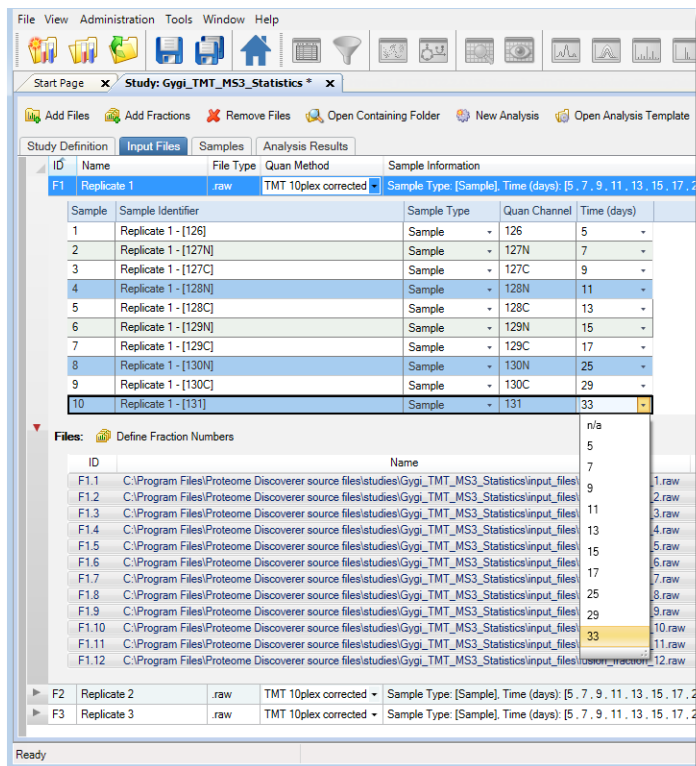
5. Set the same values in the Acquisition and the Tissue columns for the rest of the samples.

After you finish setting the factor values for each sample, the Input Files page resembles Figure 31. (In the example data set, you must set 128 factor values for eight files with eight samples each with two factors each.)

> **Note** For instructions on changing the values for factors or other study variables for multiple samples at once, see "Setting Values for Multiple Samples at the Same Time on the Input Files Page" on page 53.

Figure 31. Completed sample factors
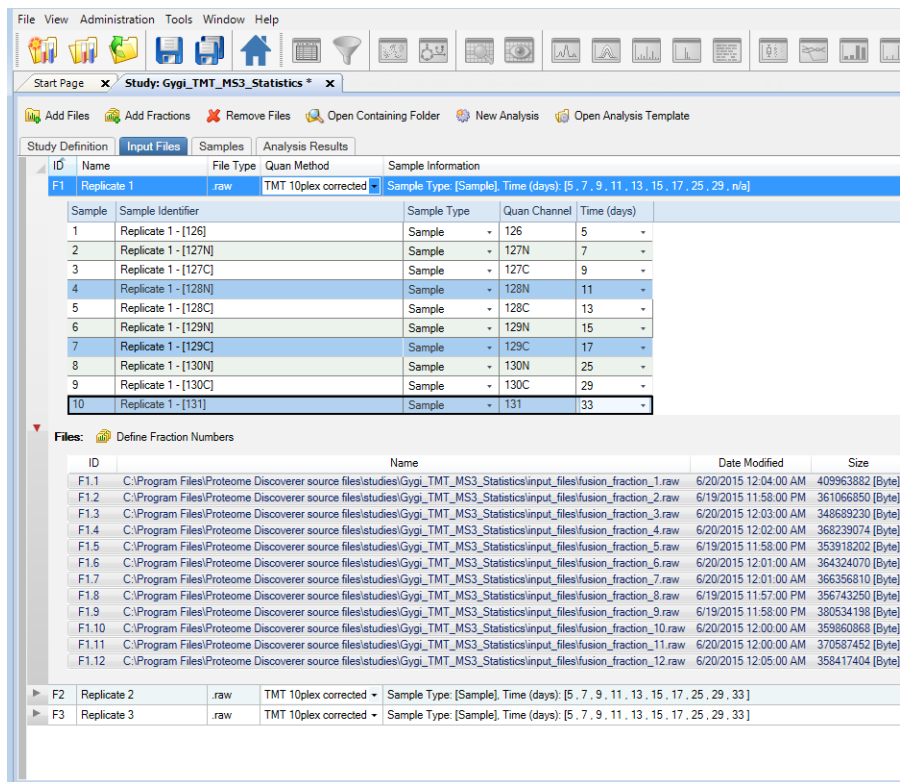
6. Click the red down arrow next to each sample to compress the information display.

   The Sample Information column summarizes the information about the samples contained in a file (see Figure 32).

Figure 32. Sample Information Column
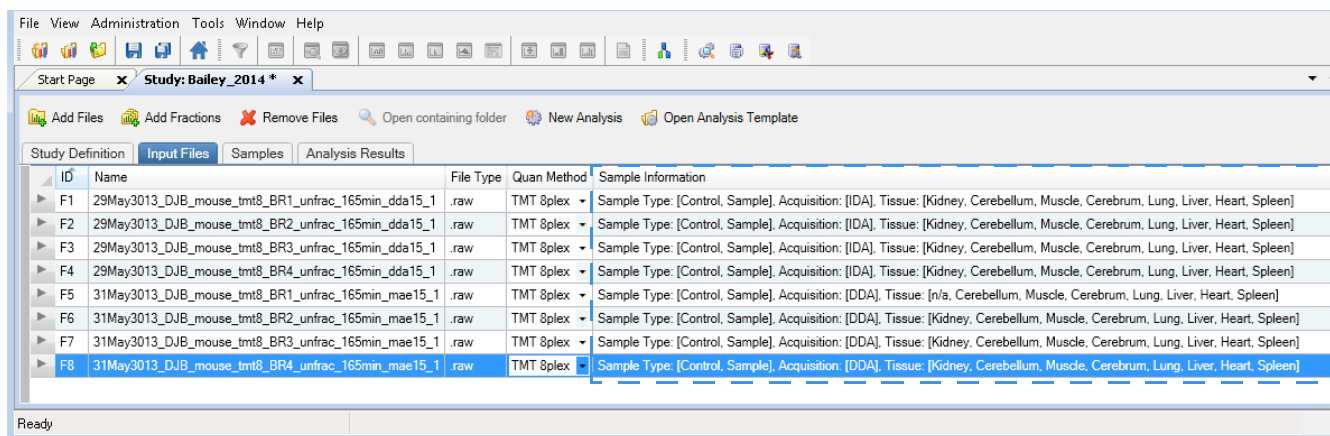
7. (Optional) Click the **Samples** tab.

   Figure 33 shows the Samples page. It displays the same sample information as the Input files page.

Figure 33. Samples page

Setting Values for Multiple Samples at the Same Time on the Input Files Page

Highly multiplexed data are results obtained from processing several samples from one raw data file mixed and analyzed together in one LC/MS analysis, where isotopic and isobaric labels were used in quantification to distinguish the contribution of the individual samples. If you have highly multiplexed data, several files and samples with many study variables to set (sample type, quantification channel, study factors), or both, you can set values for study variables for multiple samples at once. You can use either the mouse or the keyboard to set these values on the Input Files page.

You can also use multicell editing on the Samples page to enter factor values for several samples. For example, if you have acquired data from several different tissues and the tissue is encoded in the file and sample names, you can use the filter option on the Samples page to filter all samples that contain "liver" and then set the Tissue factor to "liver" for all these samples.

❖ **To set values for multiple sample cells at the same time by using the mouse**

1. Click the **Input Files** tab.

2. To select multiple samples for one study variable, select the first cell and drag the cursor to select the remaining cells.

Figure 34 shows multiple cells being selected in the Acquisition column.

Figure 34. Selecting multiple cells at one time
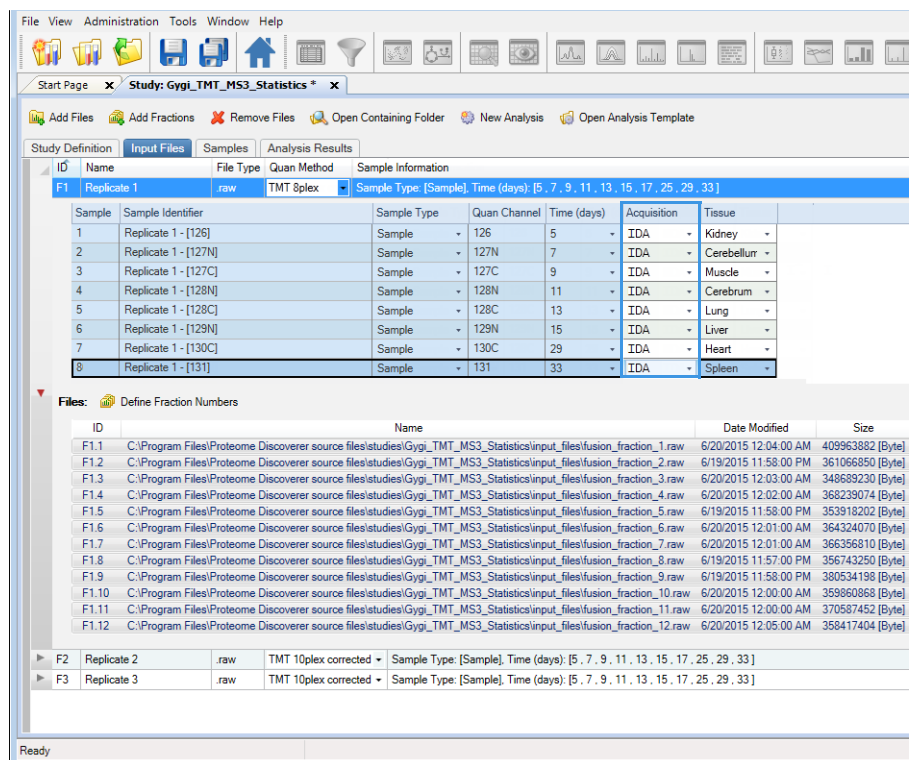
3. Press the F2 key to enter multicell editing mode.

4. Select the new value from the list in the last cell that you selected.

5. Press the RETURN or ENTER key, or click elsewhere in the application.

In the example, the values in the cells of the Acquisition column change from IDA to DDA, as shown in Figure 35.

Figure 35. Multiple values changed at one time
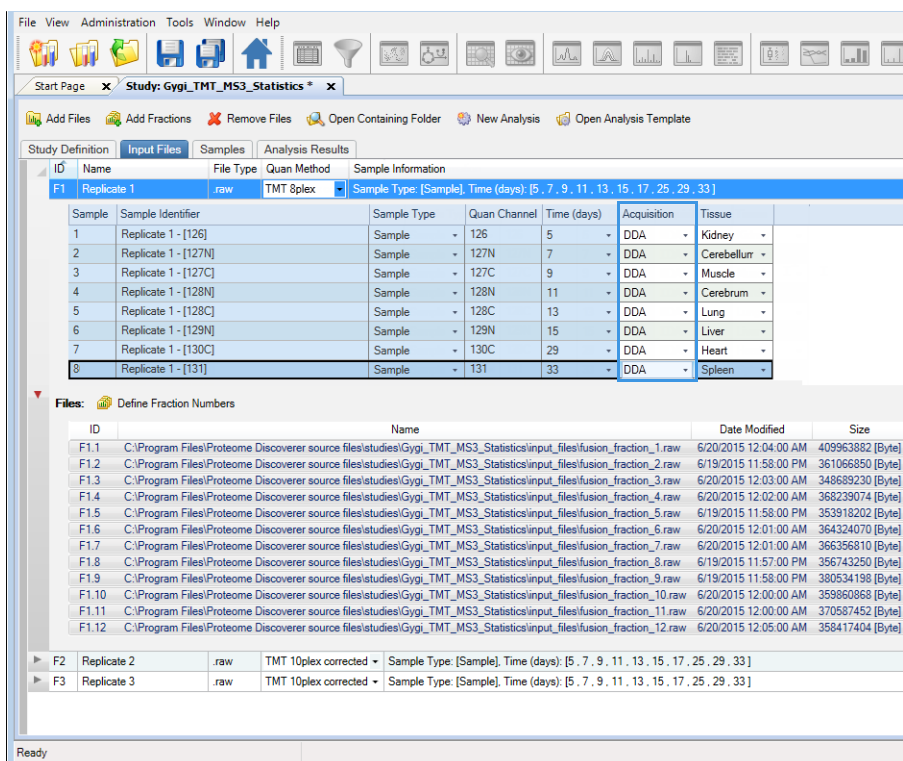
❖ **To set values for multiple sample cells at the same time by using the keyboard**

1. Click the **Input Files** tab.

2. Click the first cell to edit.

3. Press the ESC key twice until you are no longer in cell editing mode.

4. Holding down the SHIFT key, use the up and down arrow keys on your keyboard to select the cells that you want to edit.

5. Press the F2 key to enter multicell editing mode.

6. In the last cell selected, use the up and down arrow keys to select a new cell value from the list.

7. Press the RETURN or ENTER key to apply your change to all selected cells.

Setting Values for Multiple Samples at the Same Time on the Samples Page

Use the following procedure to change values for multiple samples on the Samples page.

❖ **To set values for multiple samples at the same time on the Samples page**

1. Click the **Samples** tab.

2. Click the **Contains** icon, ▣ , if necessary, in the Acquisition column.

3. In the box next to the icon, type the name of the filter.

   For example, suppose you want to change the acquisition method for a number of samples from IDA to DDA. If you want to display only samples that contain "DDA," type **dda** in the filter box. Figure 36 shows the results.

   Figure 36. Multicell editing of samples using filters
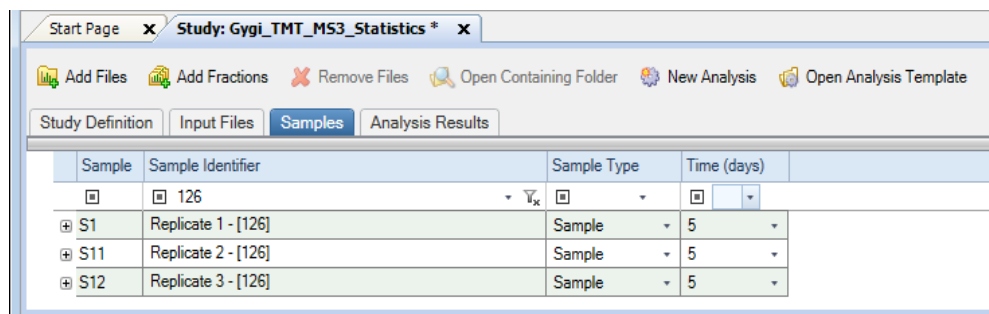
4. Select the first cell to change, and then drag the cursor to select the remaining cells that you want to change.

5. Press the F2 key to enter multicell editing mode.

6. Select the new value from the list in the last cell that you selected.

7. Press the RETURN or ENTER key, or click elsewhere in the application.

To return the samples to their unfiltered state, select the **Clear All Filters** icon, ⧩.

Saving a Study

You can save a study manually or automatically.

You can save a study manually at any time. A change in a study that requires you to save it is indicated with an asterisk (*) in the tab after the study name.

> **Note** Studies and analyses in the Proteome Discoverer application are separate and must be saved separately. Saving a study does not save an analysis, and saving an analysis does not save a study.

You can set an option to have the application save studies automatically when you click [Run]. This option also saves the results generated in the study. It saves the analysis containing the workflow, but when you close the study and reopen it, you must access the analysis by clicking the Analysis Results tab, and then doing one of the following:

- Click [Reprocess ▾] , and choose **All Analysis Steps** (to open both the processing and consensus workflow) or **Last Consensus Step** (to open just the consensus workflow).

–or–

- Right-click the name of the result file and choose **Show Details**.

❖ **To save a study manually**

Choose **File > Save**.

The Proteome Discoverer application saves the study in the *study_name*.pdStudy file in the study directory.

❖ **To save a study automatically**

1. At some point before you click   Run , choose **Tools > Options**.

2. In the Options dialog box, select **Study Options** in the left pane.

3. Select the **Auto Save When Starting Analysis** check box.

4. Click **OK**.

The Auto Save When Starting Analysis option does not take effect for any open studies. The application automatically saves only newly opened studies.

❖ **To save all open studies**

Choose **File > Save All**.

The Proteome Discoverer application saves all open studies.

Copying a Study to Another Computer

If your computer has run out of storage space or no longer works, you might want to move your study data to a new computer to continue working with it. You can copy a study folder with all of its data to another computer and continue to use it on another installation of the Proteome Discoverer application.

Your computer system must meet the following prerequisites to effect this transfer:

• All file references in a study must not point to fixed locations on a specific computer. You must be able to update the storage paths to a new location.

The study management system automatically tries to locate files in a study folder or any of its subfolders. You can resolve all missing file issues in one of the following ways:

– Copy all missing input files or missing result files into the study folder or a subfolder. For example, within the study folder you can create a folder called Input Files and copy all raw data files into the Input Files folder.

– Update the file location by re-adding the files to the study. If the study can no longer find a *drive*:\data\abc.raw file because you have moved the file to *drive*:\data\project xy\abc.raw, you can add the abc.raw file again from the new location. The study management system stores the new location for the file. It also retains all quantification methods, factors, and other study variables set for the file.

• The quantification methods stored in a study do not keep their connection to the initially connected processing server from which they were received.

When you go to the Study Definition page of an open study, you find all quantification methods currently available on the connected processing server listed in the Quantification Methods box in the upper right corner. By clicking the check box next to a listed quantification method, you can add this method to the study. When you save a study, the study management system stores the added quantification methods in the study file.

The quantification methods stored in the study are kept independent of the processing server. However, once you add a quantification method to a study, the method is independent of the connected processing server, and any changes you make to the quantification methods do not appear in the quantification methods already added to your study. You must be able to display and edit the quantification methods within the study itself. If you hold the mouse pointer over one of the check boxes representing a quantification method, two small icons for editing a method and exporting a method appear (see Figure 37).

Figure 37. Icons for editing and exporting a quantification method

Save icon
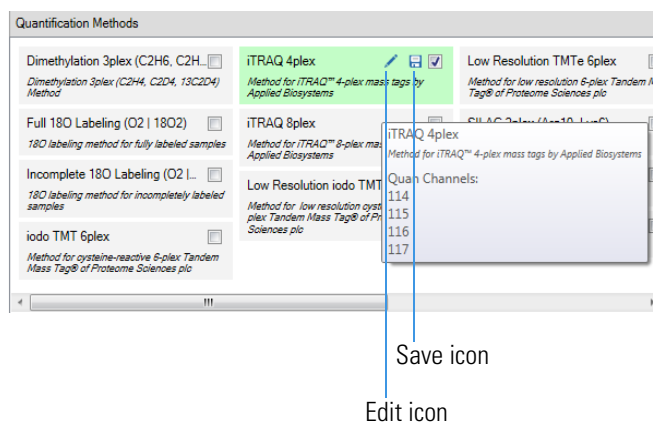
Edit icon

Clicking the Edit icon opens the Quantification Method Editor.

> **Note** If a quantification method is already set for one or more input files, you cannot change the number and names of the quantification channels. In this case, you can only change the modifications set for the method or the channels. Click the **Export** icon to save the quantification method to a METHOD file that you can then import into another Proteome Discoverer instance.

The quantification methods are now automatically set for the quantification nodes contained in the workflow to avoid any accidental mismatches between the quantification methods specified for an input file and the quantification methods set in a workflow. The analysis validation detects any inconsistencies.

You can also use the Proteome Discoverer Daemon utility to select a quantification method. For information on this procedure, see "Running the Daemon Utility from the Window" on page 124.

Study Window Parameters

Table 2 describes the parameters available on the pages of a study.
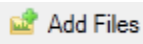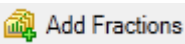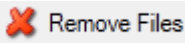
Table 2. Study page parameters (Sheet 1 of 6)
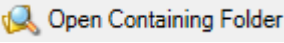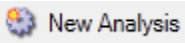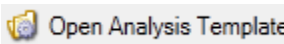
Parameter	Description
Command buttons	
Add Files	Opens the Add Files dialog box so that you can select the files to submit to the processing workflow. You can add RAW, MSF, MGF, MZXML, MZDATA, MZML, and .pdResult files.
	This button is available only when the Input Files page is selected.
Add Fractions	Opens the Add Fraction dialog box so that you can select a series of input spectrum files in any format that are fractions. The formats can be RAW, MGF, MZXML, MZDATA, or MZML.
	This button is available only when input files are selected on the Input Files page.
Remove Files	Removes the selected input files from the Input Files page.
	This button is available only when the Input Files page is selected.
Open Containing Folder	Opens the folder or folders containing the .pdResult file and the MSF file corresponding to a selected analysis result on the Analysis Result page. This button opens the folder or folders in Windows Explorer.
New Analysis	Opens a new Analysis window.
Open Analysis Template	Opens the Open Analysis Template dialog box so that you can select a custom .pdAnalysis file to use as a template.
Study Definition page	Displays basic study information, the quantification method, and study factors.
Study Summary	Displays information about the study.
Study Name	Displays the name of the study that you entered in the Study Name box of the Create New Study/Analysis dialog box.
Study Directory	Displays the directory where the study is located. You entered this location in the Study Root Directory box of the Create New Study/Analysis dialog box.
Last Changed	Displays the date and time that you saved the most recent change.
Creation Date	Displays the time and date that you clicked OK in the New Study and Analysis dialog box to create the study.
Study Description	Displays your description of the study.
Quantification Methods	Displays the quantification methods available for the selected input files in the study.

Table 2. Study page parameters (Sheet 2 of 6)

Parameter	Description
Study Factors	Displays information about the factors that distinguish two or more samples. Examples of factors are drug treatment; time of drug application; differences in tissue, organism, patient, and so forth; differences in sample preparation, chromatography settings, acquisition parameters; or differences in the isobaric or metabolic labels used.
Paste (Study Factors area)	Pastes the copied study factor box into the Study Factors area and assigns each copy the study factor name and an incrementing number: *study_factor1*, *study_factor2*, … *study_factor_n*. This command is only available after you select the Copy command.
Copy (Study Factors area)	Copies the selected study factor box.
Add (Study Factors area)	Specifies the type of study factor to add to the study. It can be one of the following: • Categorical Factor—Adds a new study factor box for a factor that belongs to a non-numerical class. See Figure 262 on page 364 for reporter ion quantification, Figure 353 on page 478 for label-free quantification, or Figure 318 on page 432 for precursor ion quantification. • Numeric Factor—Adds a new study factor box for a factor that is numerical in nature. See Figure 256 on page 361 for reporter ion quantification, Figure 347 on page 475 for label-free quantification, or Figure 312 on page 429 for precursor ion quantification.
Input Files page	Lists the input files submitted to the Proteome Discoverer application for the analysis.
ID	Displays the identification number that the Proteome Discoverer application assigns to the raw data file when you add the file to the study. File names begin with an F, and sample names begin with an S.
Name	Displays the name of the raw data file.
File Type	Displays the type of the input file, for example, a raw data file.
Quan Method	Displays the type of quantification performed on the samples. A quantification method specifies the available quantification channels. The Proteome Discoverer application currently supports precursor ion quantification methods using stable isotopes and MS/MS reporter ion quantification methods using isobaric tags. You can specify a quantification method for each of the input files.
Sample Information	Displays information about the samples in the input files, such as the sample type, the acquisition method, and the tissue type.
Samples page	Displays information about the samples contained in the input files.
Sample	Specifies the number of the sample.
Sample Identifier	Specifies the name of the sample.

Table 2. Study page parameters (Sheet 3 of 6)

Parameter	Description
Sample Type	Displays the sample type associated with a sample for quantification purposes. The sample type can be one of the following: • Sample—A specimen from a larger biological entity. • Control—A sample typically used as a reference sample in a quantification experiment. • Blank—A sample consisting only of solvent and no sample mixture. • Standard—A sample consisting of a standard quality-control peptide mixture.
Study_variables	Display the factors that distinguish two or more samples, for example, drug treatment; time of drug application; differences in tissue, organism, or patient; differences in sample preparation, chromatography settings, or acquisition settings; or differences in isobaric or metabolic labels.

Samples page shortcut menu	Displays the following commands when you right-click the Samples page:
	
	• Copy with Headers—Copies the selected rows and the header captions to the Clipboard so that you can paste them into a third-party file application file such as Excel™.
	
	• Copy—Copies the selected rows to the Clipboard so that you can paste them into a third-party file application such as Excel.
	
	• Clear Selection—Clears all selected items.
	
	• Row Selection Mode—Copies cell ranges on the Samples page to the Clipboard so that you can paste them into a third-party application file such as Excel.
	
	• Sample Type To—Sets a sample to one of the following:
	
	  – Sample: A specimen from a larger biological entity.
	
	  – Control: A sample typically used as a reference sample in a quantification experiment.
	
	• Set *Study_factor* to—Sets multiple samples to be the same study factor.

Analysis Results page	Displays the results of the search.
Open Result	Opens the .pdResult file containing the search results for the selected file.

Table 2. Study page parameters (Sheet 4 of 6)

Parameter	Description
🔄 Reprocess ▾	Repeats the analysis if you change the input files or the workflow used to process the RAW data files. You can change the workflows before rerunning the processing or consensus workflows. • All Analysis Steps—Re-performs all the analysis steps. • Last Consensus Step—Re-performs just the last step in the consensus workflow. • Use Results to Create New (Multi) Consensus—Uses a new consensus workflow to reprocess MSF results associated with one or more .pdResult files. • Use as Analysis Template—Creates a new analysis using the same sequence steps as that in the analysis result selected in the grid. However, it removes the input files and clears the result file name, if they were automatically generated from the input files.
Search	Specifies the string to search for in the search results listed on the Analysis Results page. Specify the string in the Search box. By default, the application searches only in the File Name and Description columns. If you select the column name from the Search For menu to the right of the Search box, the application searches for the string only in that column.
Search for	Specifies the name of the column to search for the string specified in the Search box.
Execution State	Displays the status of the search producing the results. These status states are the same as those displayed by the job queue. For information on these states, refer to the Help.
Creation Date	Displays the time and date that the Proteome Discoverer application began the analysis.
File Name	Displays the name of the input file being processed or completed. The application uses the name of the first sample as a default if you do not specify an analysis name.
File Type	Displays the file name extension of the analysis results file, either .msf or .pdResult.
Description	Displays your brief description of the analysis. This description has no relation to any other descriptions already introduced on any other level of the study.

Table 2. Study page parameters (Sheet 5 of 6)

Parameter	Description
Analysis Results page shortcut menu	Displays the following commands when you right-click the Analysis Results page: • Open Result—Performs the same function as ⟨icon⟩ Open Result on the Analysis Results page; that is, it opens a .pdResult file. • Open Containing Folder—Opens the folder containing the files associated with a result. • Show Details—Opens an Analysis Sequence Details window, which displays the Workflow Editor and the Analysis window containing the processing and consensus workflows used to generate the selected results file. • Reprocess—Performs the same function as ⟨icon⟩ Reprocess ▾ on the Analysis Results page; that is, it reprocess MSF results associated with one or more .pdResult files after you change the input files or the workflow used to process the raw data files.
Workflow page	For information on the parameters of the Workflow page, see Table 4 on page 100.

Table 2. Study page parameters (Sheet 6 of 6)

Parameter	Description
Grouping & Quantification page	Displays information about how the application groups quantification ratios and samples.
Study Variables	Specifies the categorical study factors to group samples and quantification ratios by. This area contains check boxes for all the study factors that you selected in the Study Factors Area of the Study Definition page.
	• Files—Groups samples and quantification ratios by files.
	• Quan Channels—Groups samples and quantification ratios by quantification channels.
	• *Study_factors*—Groups samples and quantification ratios by user-defined study factors.
	• Sample Types—Groups factors that include Sample, Control, Standard, and Blank. Variables displayed in italics contain only a single value.
Manual Ratio Generation	Contains menus where you choose the numerator and denominator for each quantification ratio.
	• Numerator: Specifies the numerator of the quantification ratio.
	• Denominator: Specifies the denominator of the quantification ratio.
Bulk Ratio Generation	Displays the study factor values to use as the denominators for semiautomatically generated quantification ratios.
Generated Sample Groups	Displays samples grouped by the values set for the selected study factors. The application ignores any other differences among the samples that might be present.
Generated Ratios	Displays the quantification ratios generated.

Using Analyses

When you have finished setting up a study, you can create an analysis. See these topics:

- Creating an Analysis

- Opening an Existing Analysis

- Adding Input Files to an Analysis

- Using Multiple Processing Steps in an Analysis

- Adding or Deleting a Processing Step

- Creating the Workflows

- Saving an Analysis
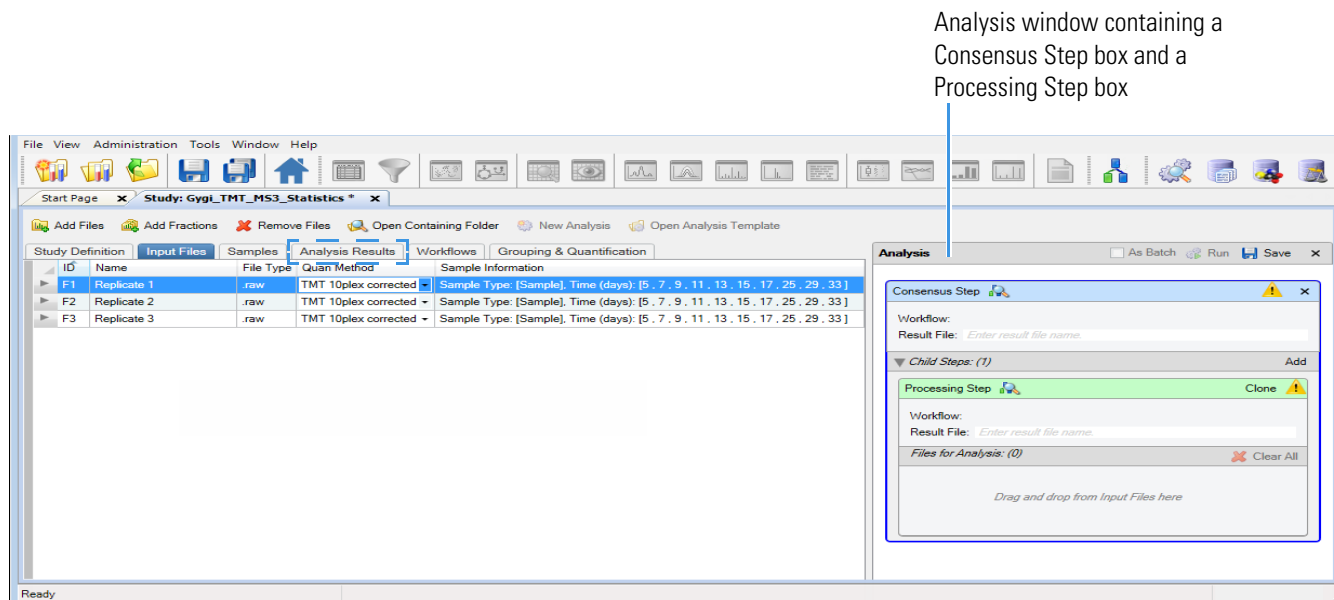
- Analysis Window Parameters

Creating an Analysis

❖ **To create an analysis**

On the Study: *Study_name* page, click [⚙ New Analysis] .

An Analysis window opens on the right side of the Study: *Study_name* page (see Figure 38).

Figure 38. Analysis window

Analysis window containing a Consensus Step box and a Processing Step box

An Analysis window contains the following items:

- Consensus Step box—Represents the consensus workflow step of the data processing.

- Processing Step box—Represents the processing workflow step of the data processing

- Child Steps bar—Contains an Add button that you can use to add another Processing Step box. Multiple Processing Step boxes are useful when you want to process the same data in different ways—for example, by using different nodes or different node settings.

In addition, two new tabs appear on the Study: *Study_name* page: the Workflows tab and the Grouping & Quantification tab, as shown in Figure 38.

If you switch to an existing study that includes an Analysis window containing an existing consensus workflow, processing workflow, or both, and you want to open a new analysis, you must close the existing Analysis window and open a new Analysis window. To close the window, click **X** in the upper right corner.

Opening an Existing Analysis

You can open an existing analysis through a template or through a results file.

❖ **To open an existing analysis through a template**

1. In an open study, click [icon] Open Analysis Template .

2. In the Open Analysis Template dialog box, browse to and select the .pdAnalysis file that you want to open.

3. Click **Open**.

❖ **To open an existing analysis through a results file**

1. In an open study, click the **Analysis Results** tab in the study.

2. Select the appropriate .pdResult file.

3. Click [icon] Reprocess ▾ .

4. To open an Analysis window containing both the processing and consensus workflows, choose **All Analysis Steps** in the Reprocess list.
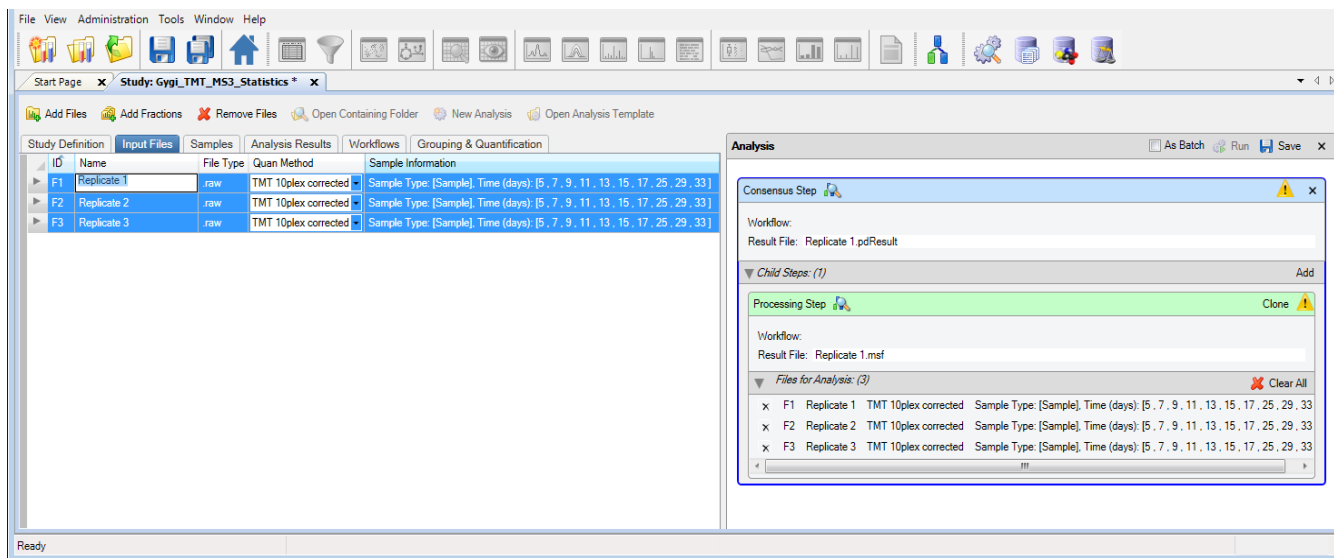
   –or–

   To open an Analysis window to execute just the consensus workflow, choose **Last Consensus Step** in the Reprocess list.

Adding Input Files to an Analysis

❖ **To add the input files to an analysis**

Select and drag the files from the Input Files page to the Input Files box of the Processing Step box in the Analysis window.

The input files are listed in the Input Files area of the Processing Step box (see Figure 39).

Figure 39. Input files in the Input Files area of the Processing Step box of the Analysis window

> **Note** You can remove a file from the Input Files area of the Processing Step box by clicking the X to the left of the file name.

Using Multiple Processing Steps in an Analysis

An analysis can contain more than one processing step. You can use additional processing steps when you want to process one set of input files differently from another set of input files. For example, you might want to run a Percolator validation separately on a set of files, such as the replicates of the control group and the replicates of the treatment group.

❖ **To add multiple processing steps**

In an open analysis window, click **Add** on the Child Steps bar.

A new processing step without a workflow appears.

❖ **To duplicate an existing processing step, including its workflow**

In an open analysis window, click **Clone** on the processing step bar.

Another window with a child step opens. You can now define a different processing workflow for the same file or a different file.

Adding or Deleting a Processing Step

You might want to perform only the processing step in a workflow. For example, you only need the processing step if you want to export peak lists as an MGF file with the Spectrum Exporter node. You can remove the Consensus step and re-add it later.

❖ **To perform only the processing step in an analysis**

In the title bar of the Consensus Step box, click the **X** in the upper right corner to remove the consensus step from the analysis.

The Consensus Step box closes, and the button, 🔵 Add , appears in the Analysis title bar.

❖ **To re-add a deleted consensus step in an analysis**

In the Analysis title bar, click 🔵 Add .

The Consensus Step box reopens, and the Add button disappears.

Creating the Workflows

The next step in performing a search is to create the workflows to use for the processing and consensus steps. This topic gives a brief overview of the steps required to create processing and consensus workflows. For information on the Workflow Editor, see "Workflow Editor" on page 34.

Creating a Processing Workflow

This procedure is the same whether you use the Workflow Editor as a stand-alone tool or in the context of an analysis.

For a description of all the nodes available in the consensus workflow, refer to the Help.

❖ **To create a processing workflow**

1. To use the Sequest HT search engine or the Mascot search engine in the workflow, configure it first by following the instructions in "Configuring the Sequest HT Search Engine" on page 22 or "Configuring the Mascot Search Engine" on page 26.

2. Click the **Workflows** tab, choose **View > Workflow Editor**, or click the **Workflow Editor** icon, 🔷 .

   Note The Workflows tab does not appear until you add or open an analysis.

   The Workflow Editor opens, as shown in Figure 18 on page 36.

3. Click the **Show Workflow** icon, 🔍, on the title bar of the Processing Step box to indicate that you want to create a processing workflow.

   The Workflow Nodes pane lists the nodes available for use in the processing workflow.

4. Click the **Workflow Nodes** tab in the Workflow Nodes pane.

5. Click the **Processing Workflow** tab on the Workflow Nodes page.

Default processing workflows are available for several instruments. You can use these default workflows as is or modify them to suit your needs. For information on accessing them, see "Using Common Workflow Templates" on page 102. You must check the node parameters in the workflow to make sure they are appropriate for your data and add modifications to the search for the quantification to be successful.

The example shown in Figure 48 on page 77 includes the following nodes:

- Spectrum Files node

- Spectrum Selector node

- Sequest HT node

- Percolator node

- Reporter Ions Quantifier node

For information about these nodes, refer to the Help.

6. From the Data Input area of the Workflow Nodes pane, drag the **Spectrum Files** or the **Spectrum Files RC** node to the Workflow Tree pane.

Either the Spectrum Files node or the Spectrum Files RC node is required in the workflow. For the location of these nodes in the Workflow Nodes pane, see Figure 40. For more information on them, refer to the Help or the Help, respectively.

Figure 40. Selected data input node in the Processing Workflow window
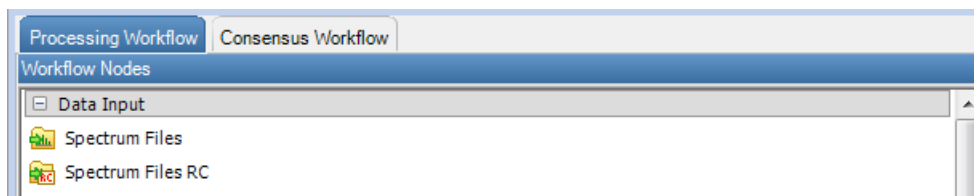
7. Drag the **Spectrum Selector** node to the workspace and place it beneath the Spectrum Files node.

Selecting the Spectrum Selector node in the Workflow Tree pane displays its parameters in the right pane.

> **Note** You can set the Spectrum Selector node to select which precursor mass to use for a given MS^n scan, such as choosing the precursor from the precursor scan.

8. Drag a search engine node (**Mascot** or **Sequest HT**) to the workspace and place it below the Spectrum Selector node.

9. Drag a PSM Validation node (**Fixed Value PSM Validator**, **Percolator**, or **Target Decoy PSM Validator**) to the workspace and place it below the search engine node.

10. Depending on your data needs, drag any other appropriate nodes from the Workflow Nodes pane to the Workflow Tree pane.

For a description of the nodes that you can select, refer to the Help.

You can also add third-party nodes that are in your installation but are not documented in this manual. For further information on those nodes, consult the relevant third-party documentation.

The numbers that appear on each workflow node indicate the order in which the application processes the nodes.

You cannot drag workflow nodes into the Workflow Tree pane that cannot logically be added. For example, you cannot add the Reporter Ions Quantifier node and the Minora Feature Detector node to the workspace together because they are mutually exclusive.

11. Organize the nodes to reflect a procedural order from top to bottom so that the Spectrum Files node remains on top as the root node.

    Delete a node by selecting the node in the Workflow Tree pane and pressing DELETE or by right-clicking the node and choosing **Cut** (or **CTRL+X**).

    You can use the Cut command and the Paste (or CTRL+P) command in the shortcut menu to move a node to another place in the workspace or use the Copy (or CTRL+C) and Paste commands to duplicate a node in the workspace.

    You can paste copied or cut nodes into other workflows.

12. Connect the nodes if they do not automatically connect:

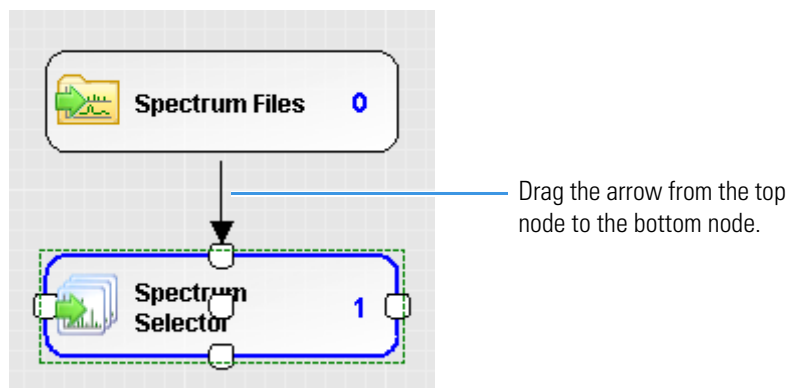   a. Click the top node to enable a blue handle at the bottom center of the node (see Figure 41).

    Figure 41. Activated node
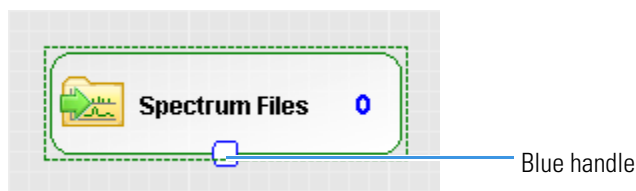
    Blue handle

    Joining the nodes together creates a sequence of steps for the application to follow.

b. Drag the blue handle down to the node below it (Figure 42).

Figure 42. Joining two nodes

Drag the arrow from the top
node to the bottom node.

IMPORTANT If the next node appears with a red edge at this point, you cannot
connect it to the previous node.

If the Workflow Editor prevents you from connecting two nodes, the workflow is
erroneous.

c. Link all the nodes to develop a workflow.

In this example, the only connections that you must make are from the Spectrum Selector
node to the Sequest HT node to the Percolator node.

13. After you join all your chosen nodes, align them by clicking ![Auto Layout] or
    right-clicking a node and choosing **Auto Layout**.

14. (Optional) To renumber the nodes, right-click and choose **Auto Number**.

Figure 43 shows a basic processing workflow.

Figure 43. Basic processing workflow
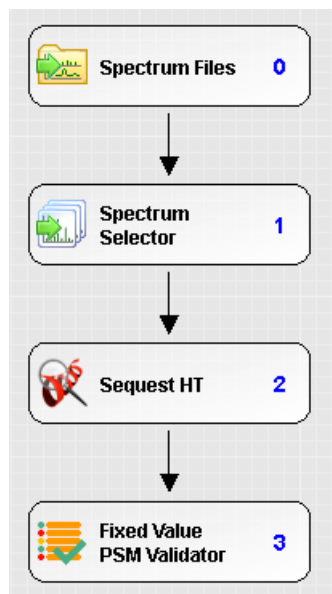
15. Set the parameters for each node in the workflow:

a. Click the node.

b. (Optional) Click **Show Advanced Parameters** so that you can view all the node's parameters.

The available parameters for the node appear in the Parameters pane, as shown in Figure 44 for the Spectrum Selector node.

**Figure 44.** Parameters for the Spectrum Selector node in the Parameters pane

16. For this example, set the parameters of the Sequest HT node as follows:

> **Note** With the Mascot and Sequest HT nodes, you can specify multiple FASTA protein databases to search at one time. To search multiple databases with Mascot server 2.3 or later, see "Searching Multiple Sequence Databases with Mascot" on page 116.

- From the dropdown list of the Sequest HT node's Protein Database parameter, select the check boxes next to the names of the appropriate FASTA files or file, for example, **SwissProt**, as shown in Figure 45. There is no limit to the number of FASTA files that you can select.

Figure 45. Selecting multiple FASTA databases for the Sequest node
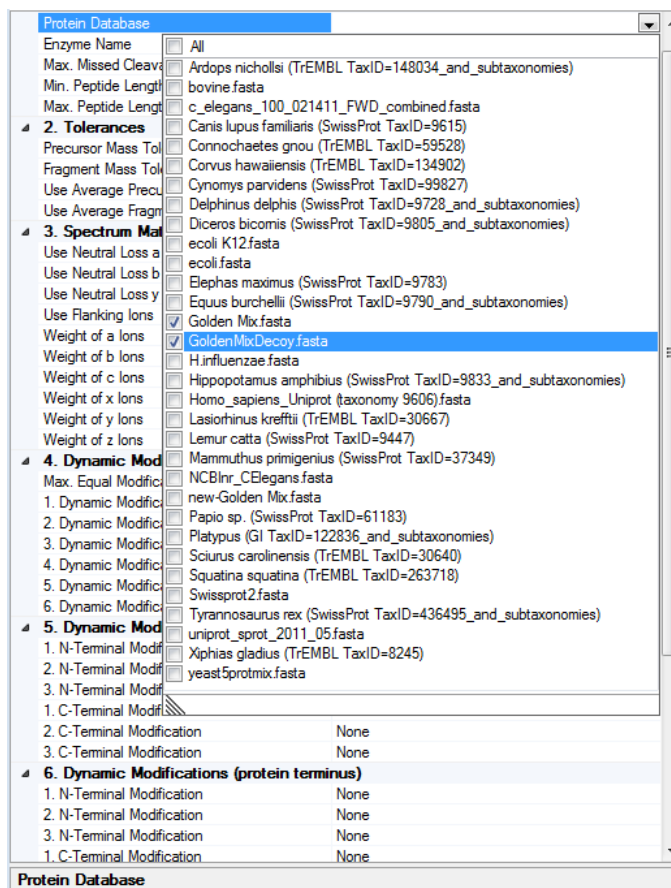
- Set the N-Terminal Modification parameter under Dynamic Modifications (Peptide Terminus), for example, **TMT 6plex/+229.163 Da**.

- Set a Dynamic Modification parameter, for example, **TMT6plex /+229.163 Da (K)**.

- Set a Static Modification parameter, for example, **Carbamidomethyl/+57.021 Da (C)**.

Figure 46 shows the parameter settings for the Sequest HT node.

Figure 46. Example of parameters in the Sequest HT node

When you click some parameters, two lists open, as in the example in Figure 47.

Figure 47. Settings and filters

You can use the list on the left to apply a filter to the setting that you select in the list on the right. The list on the left consists of three options:

- Is: Applies the setting selected in the list on the right. In Figure 47, "Is" means that the workflow processes data from the CID activation type.

- Is Not: Applies all settings in the list on the right except the selected setting. In Figure 47, "Is Not" means that the workflow processes data from all activation types except CID.

- (Default) Any: Applies all settings available for the parameter in the list on the right. In Figure 47, "Any" means that the workflow processes data from any activation type available in the list on the right.

You can filter input data before searching the database to remove lower-quality spectral peak lists from your analysis. This step might help to decrease search times and false positive identifications. The Spectrum Filters area of the Workflow Nodes pane for the processing workflow provides three types of spectrum filters to use for your search. Use these pre-analysis filters to streamline your search results. For information about these nodes, refer to the Help.

Use the Scan Event Filter node for high-mass-accuracy data, such as Mascot analysis and Sequest HT analysis of mixed fragmentation-mode-type data (CID and ETD). It can filter information according to fragmentation type, mass analyzer identity, and other parameters. For information about the Scan Event Filter node, refer to the Help.

17. In the Workflow box above the Workflow Tree pane, type a name for the processing workflow.

18. (Optional) In the Description box, type a brief description of the processing workflow.

    Providing a workflow description is highly recommended.

    Figure 48 shows an example of a completed processing workflow.

Figure 48. Processing workflow example
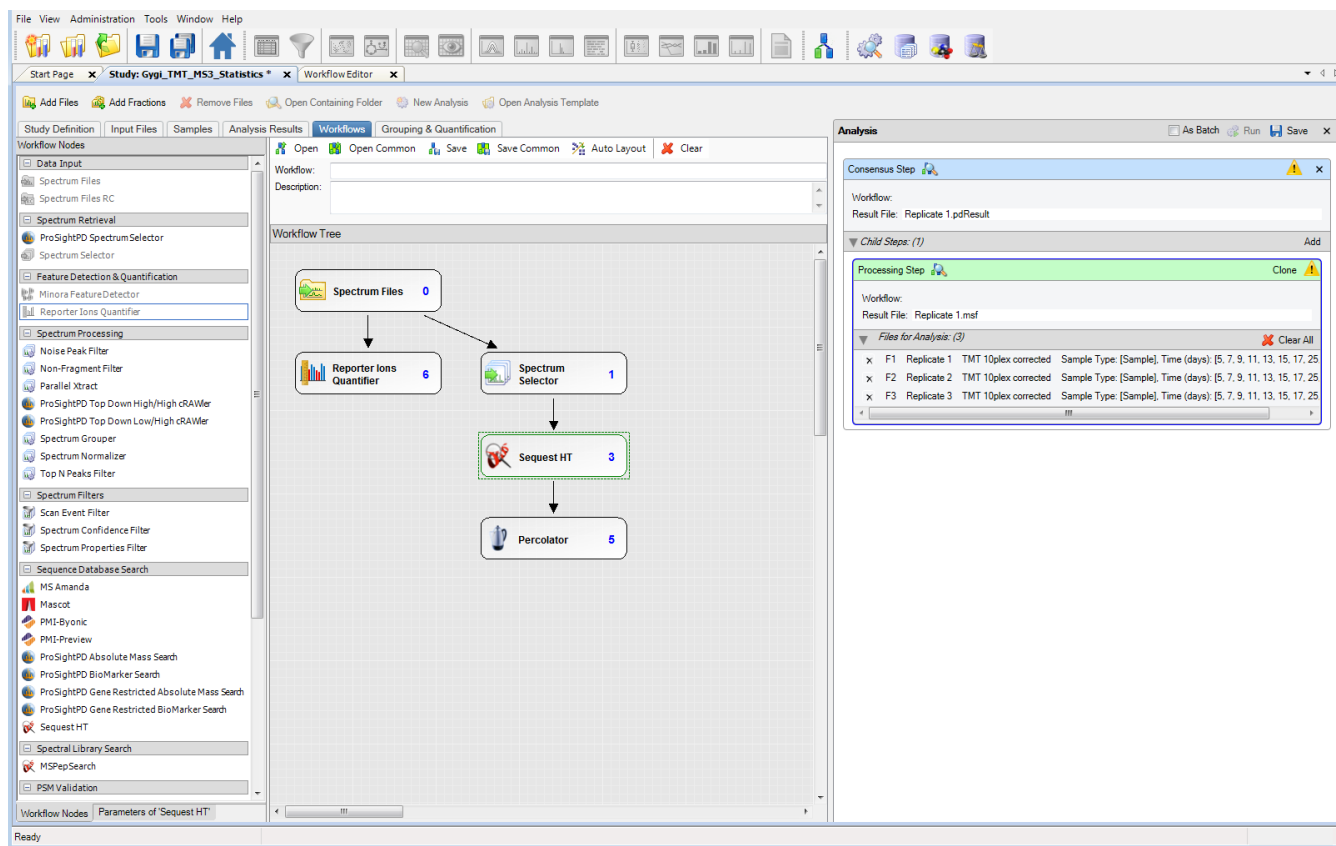
19. (Optional) Save the processing workflow as a template in a .pdProcessingWF file:

    a. In the Workflow Editor, click ![Save icon] **Save** .

    b. In the Save Workflow dialog box, do the following:

        i. Select the file to save the workflow in, or type a file name in the File Name box. You can save the workflow in the study folder or in the Common Templates folder (click ![Save Common icon] **Save Common** , in this case), or in a separate folder of workflows.

        ii. Click **Save**.

    The application saves the workflow in the *file_name*.pdProcessingWF file.

> **Note** An exclamation mark inside a yellow triangle in the upper right corner of the Processing Step box (⚠) usually indicates that a node in the workflow is obsolete, a parameter is missing, a required node is missing, or the output file name is invalid or missing. Point to the triangle to display further details about what is missing.

For detailed template information, see "Using Workflow Templates" on page 101.

The application returns you to the Processing Workflow window.

20. Do one of the following:

   • Create a consensus workflow (see "Creating a Consensus Workflow" on page 78).

   –or–

   • To run just a processing workflow, click   **Run**   in the Analysis window.

   The job queue opens, showing the status of your search (see Figure 49).

   Figure 49. Status of the processing workflow in the job queue
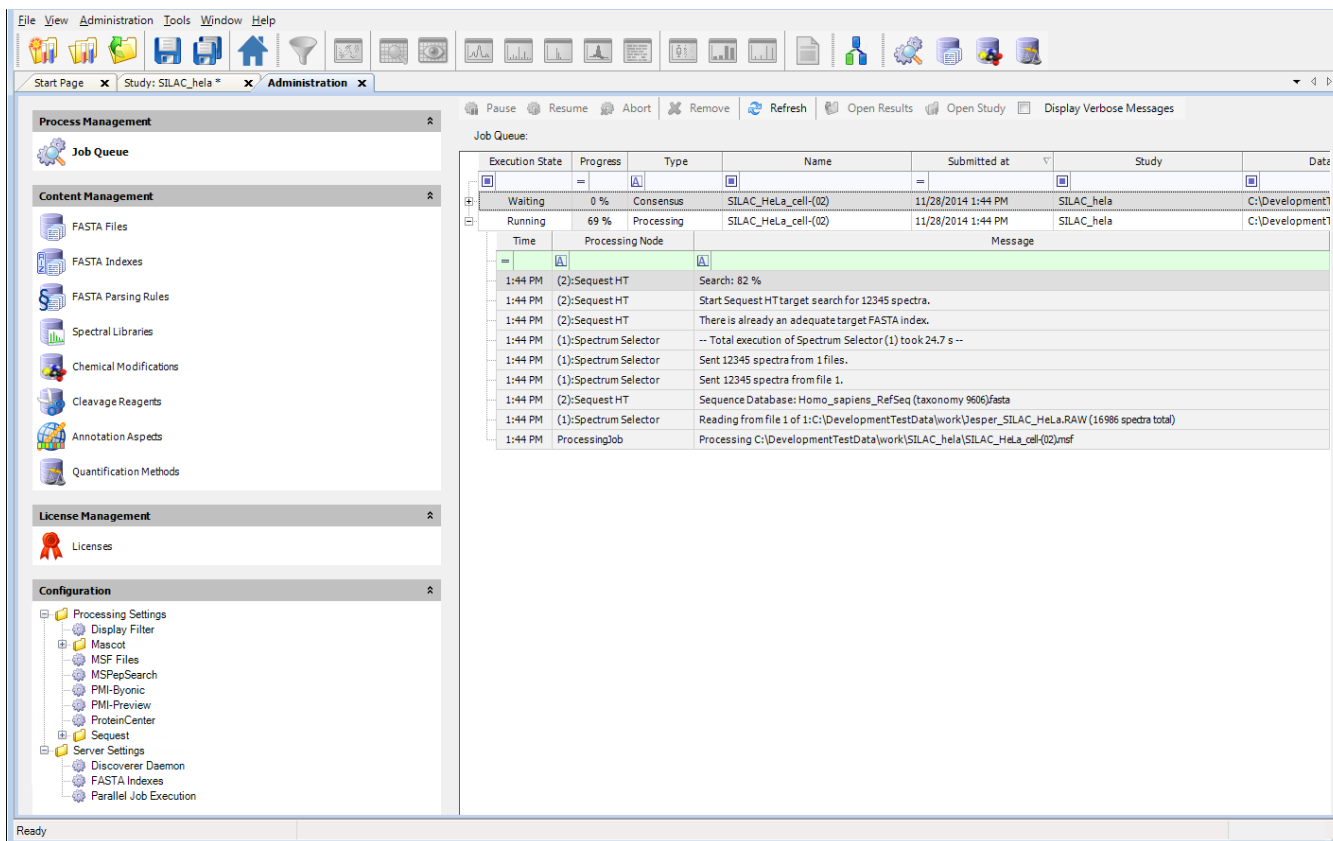
Use the job queue to check the status of your search.

For information about the job queue, refer to the Help.

The application creates an MSF file containing the results. The Processing Files pane lists the name of the output MSF file, which is the same as the name of the input file but with an .msf extension.

Creating a Consensus Workflow

You can create a new consensus workflow that generates a results report for a new processing workflow. Or, you can create a stand-alone consensus workflow if you have a preexisting MSF file for which you want to generate a results report.

For a description of all the nodes available in the consensus workflow, refer to the Help.

❖ **To create the consensus workflow**

1. Click the **Workflows** tab, choose **View > Workflow Editor**, or click the **Workflow Editor** icon, 🔧 .

   > **Note** The Workflows tab does not appear until you add or open an analysis.

   The Workflow Editor opens (Figure 18 on page 36).

2. Click the **Show Workflow** icon, 🔍, in the title bar of the Consensus Step box.

   The Workflow Nodes pane lists the nodes available for use in the consensus workflow.

3. Click the **Consensus Workflow** tab in the Workflow Nodes pane.

   Default consensus workflows are available for several instruments. You can use these default workflows right away or modify them to suit your needs. To access them, see "Using Common Workflow Templates" on page 102. You must check the node parameters in the workflow to make sure they are appropriate for your data and add modifications to the search for the quantification to be successful.

   The example workflow in Figure 52 includes the following nodes in the Workflow Tree pane:

   - MSF Files node

   - PSM Grouper node

   - Peptide Validator node

   - Peptide and Protein Filter node

   - Protein Scorer node

   - Protein Grouping node

   - Precursor Ions Quantifier node or Reporter Ions Quantifier node

   The example workflow in Figure 52 also includes the following nodes in the Post-Processing Nodes area:
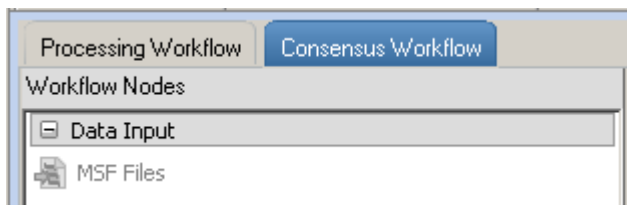
   - Result Statistics node

   - Data Distributions node

   For information about these nodes, refer to the Help.

4. From the Data Input area of the Workflow Nodes pane, drag the **MSF Files** node to the Workflow Tree pane.

   The MSF Files node is required. Figure 50 shows the location of this node.

Figure 50. Selected data input node in the Consensus Workflow window

5. Drag the following nodes to the workspace and place them beneath the MSF Files node.

> **IMPORTANT** Although these nodes are not required, they generate meaningful results. Reports generated from a workflow consisting of only the MSF Files node contain only the Proteins, PSMs, and MS/MS Spectrum Info pages. The PSMs and proteins have no confidences, and the proteins have no scores.

- PSM Grouper node: For information, refer to the Help.

- Peptide Validator node: For information, refer to the Help.

- Peptide and Protein Filter node: For information, refer to the Help.

- Protein Scorer node: For information, refer to the Help.

- Protein Grouping node: For information refer to the Help.

6. If you performed quantification in the Processing workflow, drag the **Precursor Ions Quantifier** node (for precursor ion quantification or label-free quantification) or the **Reporter Ions Quantifier** node (for reporter ion quantification) to the Workflow Tree pane.

   This node is required if you performed quantification. For more information, refer to the Help.

7. Depending on your data needs, drag any other appropriate nodes from the Workflow Nodes pane to the Workflow Tree pane.

   For a description of the nodes that you can select, refer to the Help.

   The numbers that appear on each workflow node indicate the order in which the application processes the nodes.

   You can also add third-party nodes that are in your installation but are not documented in this manual. For further information on those nodes, consult the relevant third-party documentation.

   You cannot drag workflow nodes into the Workflow Tree pane that cannot logically be added.

8. Organize the nodes to reflect a procedural order from top to bottom so that the MSF Files node remains on top as the root node.

   Delete a node by selecting the node in the Workflow Tree pane and pressing DELETE, or by right-clicking the node and choosing **Cut** (or **CTRL+X**).

You can use the Cut command and the Paste (or CTRL+P) command in the shortcut menu to move a node to another place in the workspace, or use the Copy (or CTRL+C) and Paste commands to duplicate a node in the workspace.

You can paste copied or cut nodes into other workflows.

The application automatically connects the nodes in the consensus flow. If two adjacent nodes do not connect, there is an error in the flow, and you must rearrange the nodes until they connect.
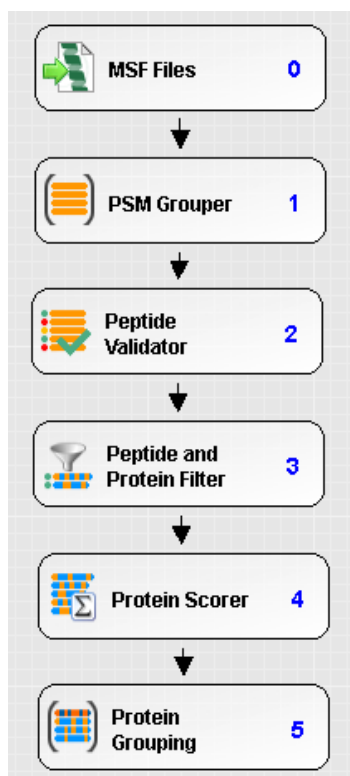
9. Align the nodes by clicking 🖾 Auto Layout or right-clicking a node and choosing **Auto Layout**.

10. (Optional) To renumber the nodes, right-click and choose **Auto Number**.

11. Set the appropriate parameters for each node in the Workflow Tree pane.

Figure 51 shows a basic consensus workflow that does not involve quantification.

**Figure 51.** Basic consensus workflow

12. Drag any appropriate nodes from the Post-Processing category in the Workflow Nodes pane to the Post-Processing Nodes pane.

You cannot place nodes from the Post-Processing category in the Workflow Tree pane. Similarly, you cannot place nodes from any other category in the Post-Processing Nodes pane.

Figure 52 shows an example of a completed consensus workflow.

Figure 52. Consensus workflow example
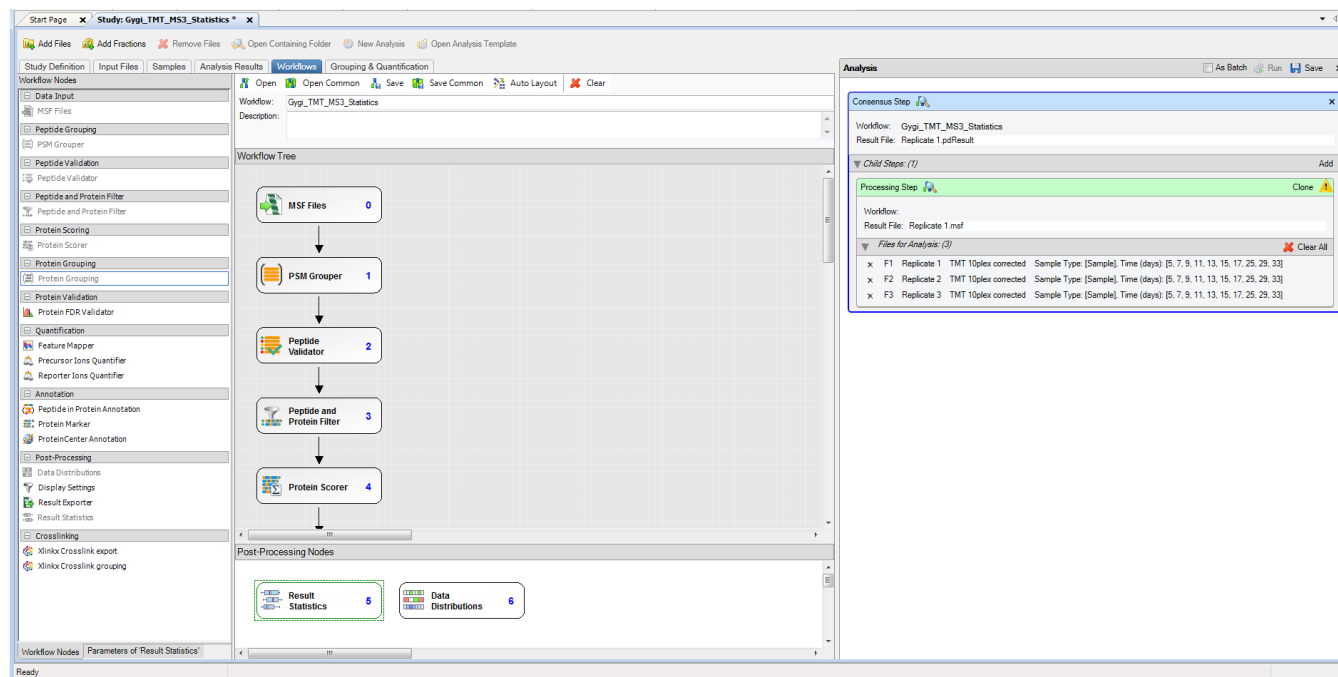
13. (Optional) In the Name box, type a name for the workflow.

14. (Optional) In the Description box, type a brief description of the workflow.

15. (Optional) Save the consensus workflow as a template in a .pdConsensusWF file as follows:

   a. In the Workflow Editor, click 📇 Save .

   b. In the Save Workflow dialog box, do the following:

      i. Select the file to save the workflow in, or type a file name in the File Name box. You can save the workflow in the study folder or in the Common Templates folder (click 📇 Save Common , in this case), or in a separate folder of workflows.

      ii. In the Save As Type box, select **Consensus Workflow File (*.pdConsensusWF)**.

      iii. Click **Save**.

   The application saves the workflow in a *file_name*.pdConsensusWF file.

   > **Note** An exclamation mark inside a yellow triangle in the upper right corner of the Consensus Step box (⚠) usually indicates that a node in the workflow is obsolete, a parameter is missing, a required node is missing, or the output file name is invalid or missing. Point to the triangle to display details about what is missing.
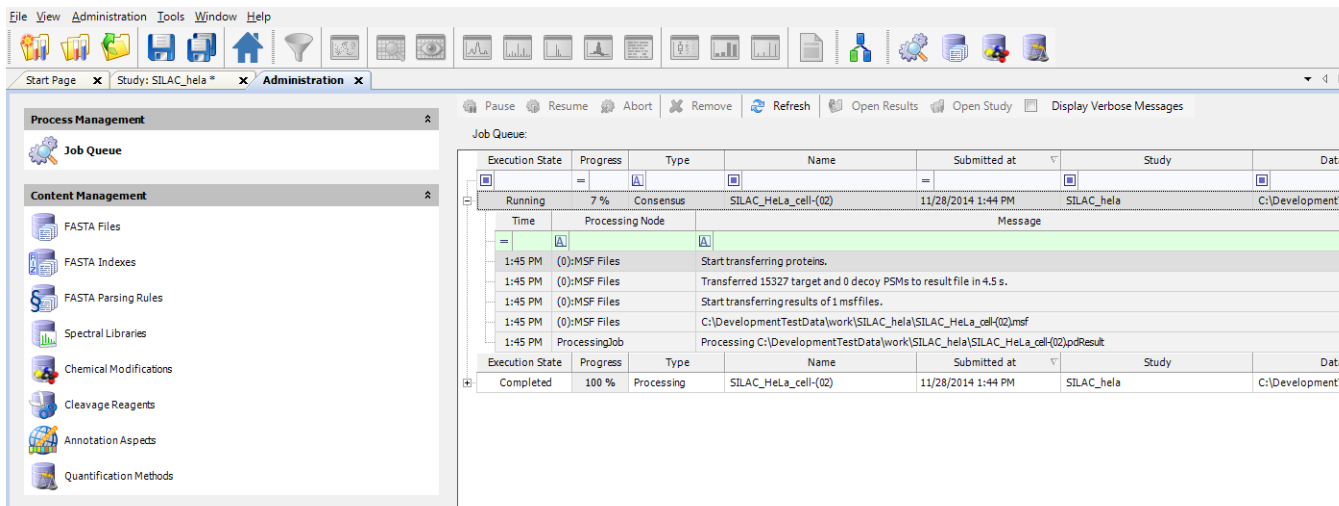
   For detailed template information, see "Using Workflow Templates" on page 101.

   The application returns you to Consensus Workflow window.

16. In the Analysis window, click [Run] .

The job queue opens, displaying the status of the consensus flow (Figure 53).

Figure 53. Status of the consensus workflow in the job queue

Use the job queue to check the status of the consensus workflow.

For information about the job queue, refer to the Help.

The Proteome Discoverer application generates a .pdResult file. For information on opening and using the .pdResult file, refer to the Help.

Saving an Analysis

To use an analysis as a template for later reuse, you can save it as a .pdAnalysis template file.

> **Note** Studies and analyses in the Proteome Discoverer application are separate, so you must save them separately. Saving a study does not save an analysis, and saving an analysis does not save a study.

❖ **To save an analysis as a template for later reuse**

1. In the Analysis window, click [Save] .

2. In the Save Analysis Template dialog box, browse to the location where you want to store the template.

3. In the File Name box, browse to the study folder, and type or select the template file name.

4. In the Save as Type box, select **Analysis Templates (*pdAnalysis)**.

5. Click **Save**.

    The application saves the analysis in a file with a .pdAnalysis extension.

    The .pdAnalysis template file saves the processing and consensus workflows. It saves neither the input files nor the study variables that were selected to group the samples and quantification ratios.

Viewing the Workflow and the Analysis After the Search

To view the workflow and the analysis that produced a results file after you perform a search, see "Viewing the Workflow and the Analysis from the Results" on page 90.

Analysis Window Parameters

Table 3 describes the parameters available in the Analysis window.
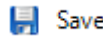
Table 3. Analysis window parameters (Sheet 1 of 2)

Parameter	Description
As Batch	Determines whether the application performs an analysis in batch mode. For information, see "Performing a Search in Batch Mode" on page 88.
Run	Performs an analysis.
Save	Opens the Save Analysis Template dialog box so that you can save the analysis in a .pdAnalysis template file.
Consensus Step box	Displays the name of the consensus workflow, the name of the consensus workflow results file, and the processing workflow that generated the MSF file that was submitted to the consensus workflow as an input file.
(lock icon)	Indicates that you cannot change the consensus workflow.
(magnifier icon)	Opens the Workflow Tree window containing the consensus workflow.
(warning icon)	Indicates that the consensus workflow contains an error.
Workflow	Displays the name of the consensus workflow.
Result File	Displays the name of the file containing the results of the consensus workflow.
Child Steps	Displays the processing workflow.
Add	Adds an empty Processing Step box.

Table 3. Analysis window parameters (Sheet 2 of 2)

Parameter	Description
Processing Step box	Displays the name of the processing workflow, the name of the processing workflow results file, and the input files that were submitted to the processing workflow as input files.
🔒	Indicates that you cannot change the processing workflow.
🔍	Opens the Workflow Tree window containing the processing workflow.
⚠	Indicates that the processing workflow contains an error.
Reprocess	Opens a locked processing workflow so that you can make changes to it.
Clone	Opens a child step window so that you can create a different processing workflow for the same file or a different file.
Workflow	Displays the name of the processing workflow.
Result File	Displays the name of the file containing the processing workflow results.
Input Files	Lists the input raw data files that were submitted to the processing workflow.

Specifying Quantification Ratios from Selected Sample Groups

If you intend to perform quantification, you must specify quantification ratios for the appropriate sample groups. For information on this procedure, see the following topics:

- Reporter ion quantification: "Specifying Quantification Ratios from Selected Sample Groups for Reporter Ion Quantification" on page 393.

- Precursor ion quantification: "Specifying Quantification Ratios from Selected Sample Groups for Precursor Ion Quantification" on page 456.

- Label-free quantification: "Specifying Quantification Ratios from Selected Sample Groups for Label-Free Quantification" on page 498.

Performing the Search

You can perform a search in group mode or batch mode.

## Performing a Search in Group Mode

❖ **To perform a search in group mode**

In the Analysis window, click [Run] .

The application validates the analysis setup before it starts processing and issues error or warning messages in the Analysis Validation Issues dialog box if it finds errors (see Figure 54). For example, it might issue an error message if not all the input files have the same quantification method. Or, it might issue a warning message if you added several input files with quantification but did not set any of the study variables to group your samples and quantification ratios.

You can ignore warnings, which are marked by an exclamation mark inside a yellow triangle. Because warnings are only hints that the analysis might not be set up correctly, you can click Ignore in the Analysis Validation Issues dialog box. Figure 54 shows examples of warning messages.

Figure 54. Warnings in the Analysis Validation Issues dialog box
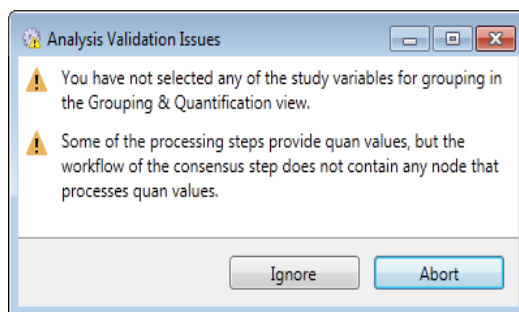
You must resolve any validation errors that are marked with a red exclamation mark (Figure 55).

**Figure 55.** Error in the Analysis Validation Issues dialog box
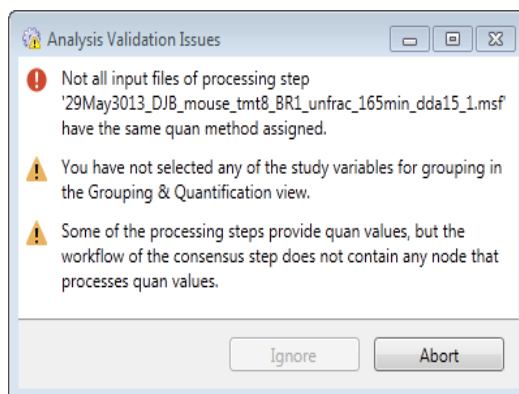
The validation cannot detect every potential problem but can check for these specific problems or inconsistencies:

- The set workflows have major errors because of missing mandatory node types such as validator nodes, missing connections, or missing parameters. In most cases, you catch these errors earlier because you cannot click the Run button.

- When you use quantification nodes or have assigned quantification methods to your input files, the validation checks for the following:

  – All quantification nodes used in the workflow of a particular processing step have the same kind of quantification method; for example, you cannot have mixed precursor and reporter ion quantification nodes in a workflow.

  – Not all input files of a particular processing step have the same quantification method assigned; for example, you have mixed precursor and reporter ion quantification files in a processing step.

  – The kind of quantification node used in the workflow of a particular processing step is not compatible with the quantification method set for the input files of this step; for example, you try to process reporter quantification files with a precursor quantification node or vice versa.

  – All input files have a quantification method set, but no quantification node is used in the workflow of the processing step.

  – A quantification node is used in the workflow of a processing step, but none of the input files has a quantification method set.

  – You have not selected any of the study variables for grouping on the Grouping & Quantification page, which is only shown if you process more than one file. If you use only one file, you usually do not want to group. However, you could create biological replicates using different channels of a quantification method such as iTRAQ in an experiment on a single raw data file in which you want to group by tissue, matrix or experimental condition.

  – Some of the processing steps provide quantification values, but the workflow of the consensus step does not contain any node that processes quantification values.

  – The quantification method used for the input files of a particular processing step is not valid.

After the application validates the analysis, it begins processing it. The job queue opens so that you can monitor the progress of the job (see Figure 56).

Figure 56. Search progress in the job queue
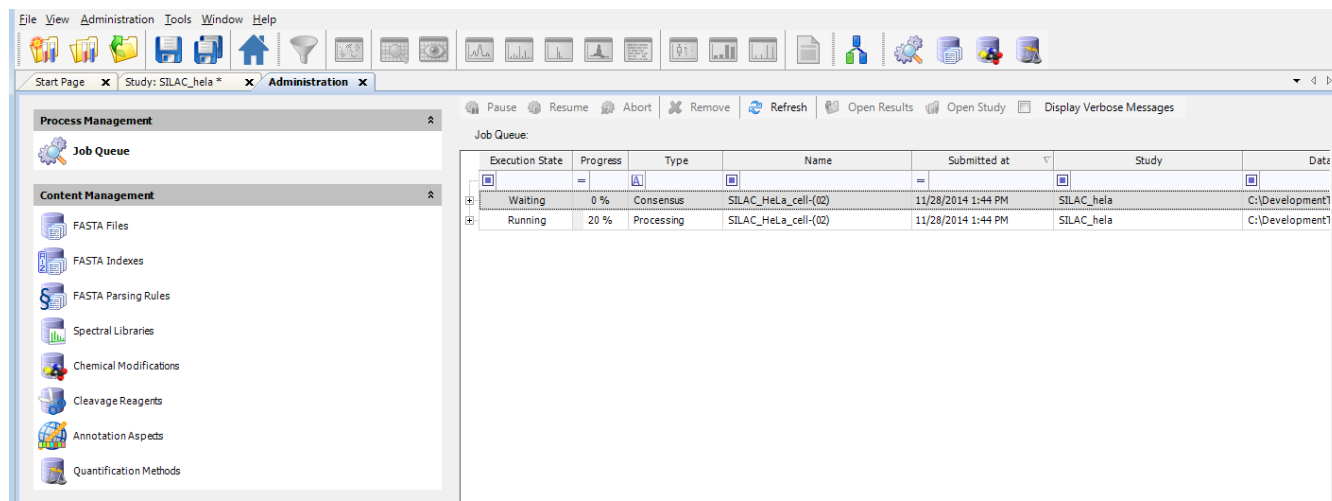
The job is done when "Completed" appears in the Execution State column for the processing and consensus workflows.

The application uses the name of the first raw data file or sample as the default name of the results file as a whole.

Performing a Search in Batch Mode

You might want to process each file in a set of files with the same processing workflow and the same consensus workflow. Processing a set of files in this way is called batch mode. Batch mode is only available if there is more than one input file and if the analysis has just one processing step—that is, if there is just one Processing box in the Analysis window.

❖ **To perform a search in batch mode**

1. Set up or open a study (see "Using Studies" on page 38), and add the input files (see "Adding Input Files" on page 42).

2. Set up or open an analysis (see "Using Analyses" on page 64).

3. Drag the input files to the Processing Step box.

4. In the title bar of the Analysis window, select the **As Batch** check box (see Figure 57).

Figure 57. As Batch check box selected
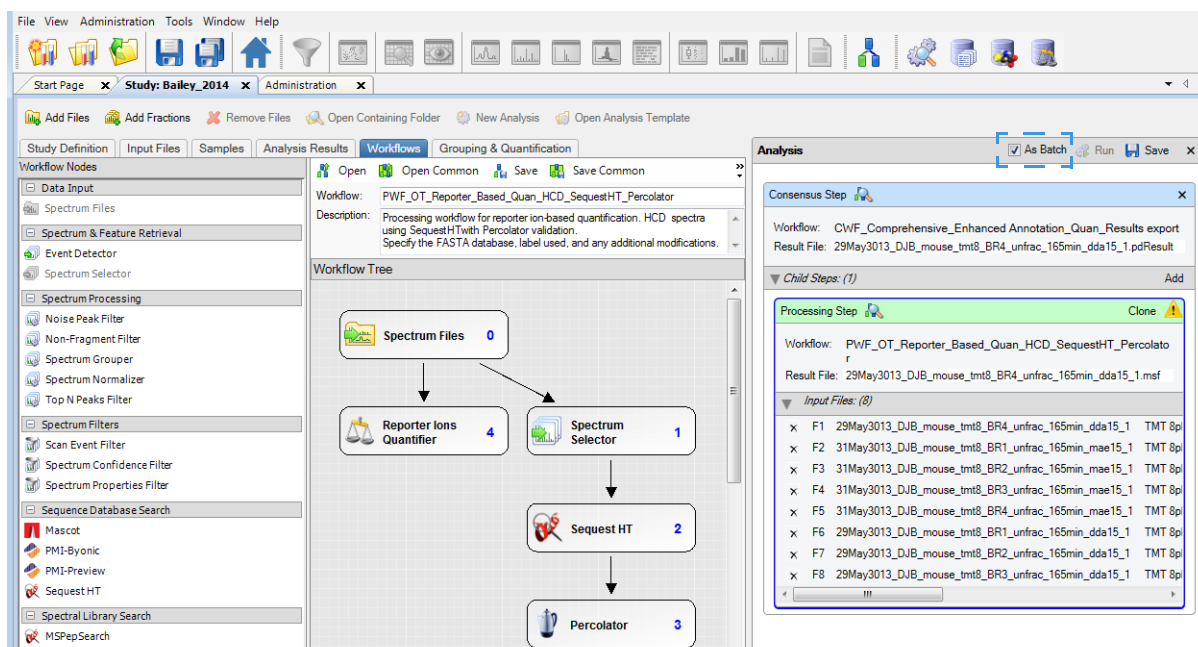
5. Click ![Run] .

The application processes all of the input files separately and generates one .pdResult file for each of the input files. The result files have the same name as the single input file.

Working with the Search Results

The following topics explain how to open, delete, view, convert, and reprocess search results.

- Opening the Search Results
- Deleting Search Results
- Viewing the Workflow and the Analysis from the Results
- Converting Results
- Reprocessing an Existing Analysis
- Using an Analysis as a Template

Opening the Search Results

For information about opening the search results, refer to the Help. You can also open the search results from the Analysis Results page of a study. In addition, you can open an MSF file and the folder containing the files associated with a result from the Analysis Results page.

❖ **To open a .pdResult file from the Analysis Results page**

Right-click the appropriate result on the Analysis Results page, and choose **Open Result**, or click 🗐 Open Result on the Study: *Study_name* page.

❖ **To open the folder containing the files associated with a result**

Right-click the appropriate result on the Analysis Results page, and choose **Open Containing Folder**.

Deleting Search Results

You can delete search results from a study.

❖ **To delete a study's search results**

1. In an open study, click the **Analysis Results** tab if it is not already selected.

2. Select the row of results that you want to delete, and click ✖ Remove Files or press the DELETE key.

3. In the Remove Analysis Result dialog box, click **Yes**.

## Viewing the Workflow and the Analysis from the Results

After you perform a search, you can see the workflow and the analysis that produced the search results.

❖ **To access the workflow and the analysis from the results**

1. In an open study, click the **Analysis Results** tab.

2. Right-click name of the results file and choose **Show Details**.

   The Analysis Sequence Details window opens (see Figure 58). It displays the Workflow Editor and the Analysis window that contains the processing and consensus workflows used to generate the selected results file.

Figure 58. Analysis Sequence Details window
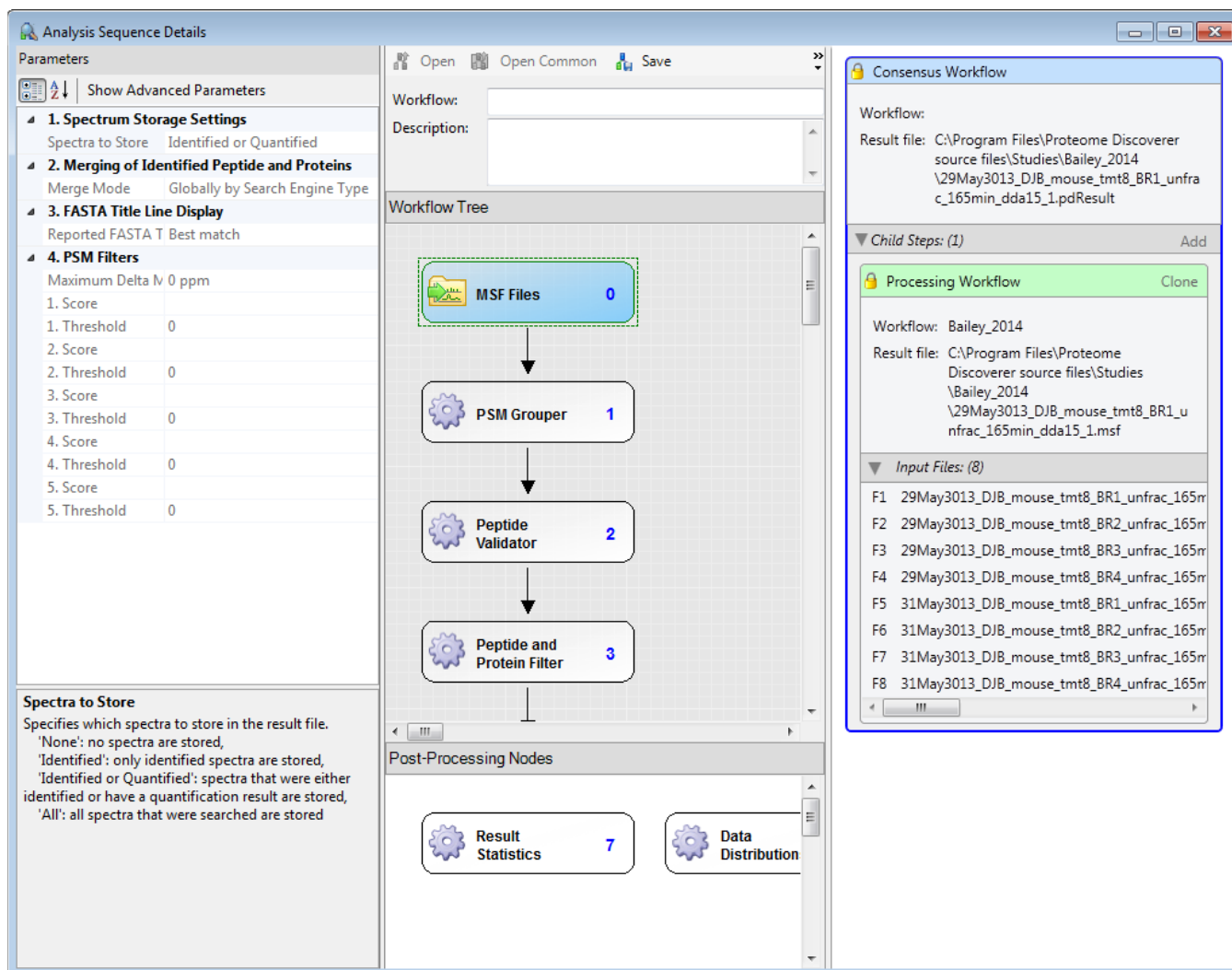
If the nodes in the workflow in this window display one or more of the warning symbols described in "Correcting Template Errors" on page 106, such as the yellow triangles shown in Figure 59, click the Analysis Result tab and choose **Reprocess > All Analysis Steps** or **Last Consensus Step**, as appropriate.

Figure 59. Analysis Sequence Details window with obsolete symbols in the workflow
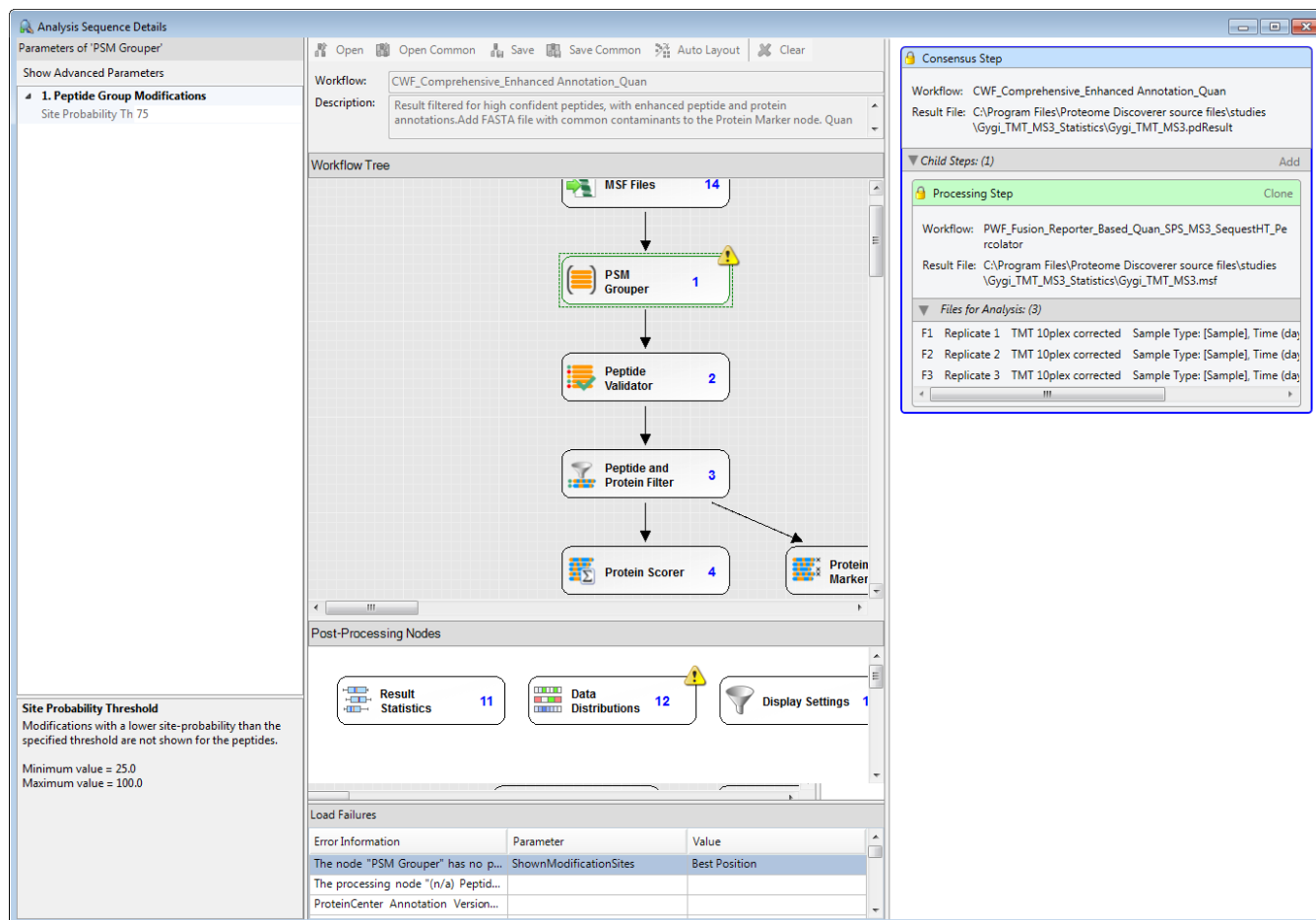
Searching Analysis Results

You can search for a specified string in the search results on the Analysis Results page of a study. Filtering the analysis results in studies enables you to reprocess previous workflows more quickly.

❖ **To search for a specified string**

1. In an open study, click the **Analysis Results** tab.

2. In the Search box (Figure 60), type the string that you want to search for.

    By default, the application searches for the string in the File Name and Description columns.

3. (Optional) To search in a single column, select the name of the column from the Search For menu to the right of the Search box.

Figure 60. Search box and Search For menu on the Analysis Results page
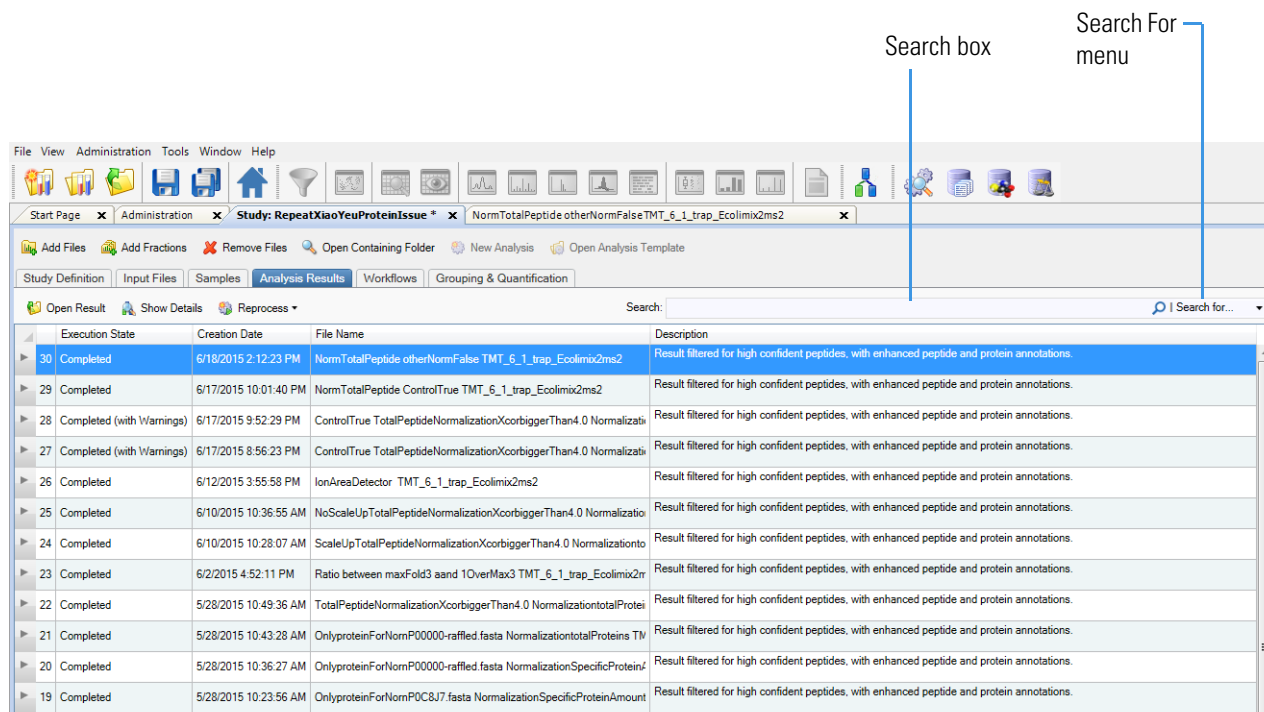
Converting Results

To obtain a different results file, you can convert MSF files to .pdResult files or use .pdResult files. You have two ways to perform this conversion: by adding result files to a new study or by adding result files to an existing study.

You can also process these files as a batch and generate a .pdResult file for each input file or one .pdResult file for multiple input files.

Adding Result Files to a New Study

You can add existing MSF or .pdResult files to a new study.

❖ **To add MSF or .pdResult files to a new study**

1. Create a study (see "Creating a Study" on page 40).

2. In the New Study and Analysis dialog box, click ![Add Files].

3. In the Add Input File dialog box, select the MSF files, .pdResult files, or both that you want to import.

4. Click **Open**.

5. In the New Study and Analysis dialog box, click **OK**.

   An Analysis window opens that contains a processing step for each file added to the study.

6. Create a consensus workflow. For instructions, see "Incorporating an Existing Workflow into a Study and an Analysis" on page 99.

7. (Optional) To generate a .pdResult file for each input file, click the **As Batch** check box in the Analysis window title bar.
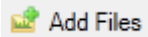
   When you select the As Batch check box, the application creates a .pdResult file for each input file. When you do not select it, it creates one .pdResult file for all input files.

8. Click Run .

Adding Result Files to an Existing Study
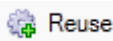
You can add MSF or .pdResult files to an existing study.

❖ **To add result files to an open study**

1. In an open study, click Add Files .

2. In the Add Files dialog box, browse to the location of the MSF or .pdResult files that you want to add to the existing results in the study, select them, and click **Open**.

   Figure 61 shows an added MSF file on the Analysis Results page with a .pdResult file.

   Figure 61. Existing results file and the added MSF file on the Analysis Results page
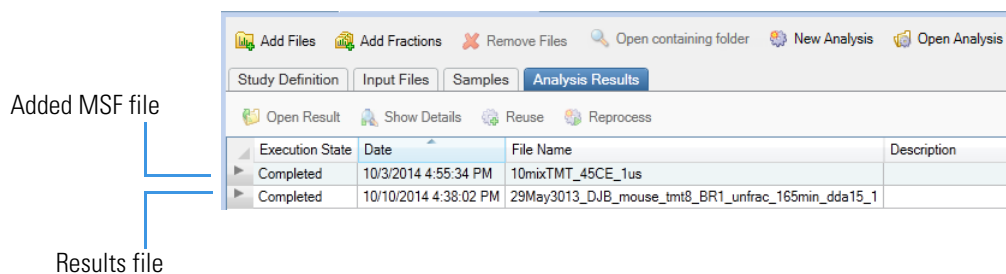
1. Click the **Analysis Results** tab.

2. On the Analysis Results page, select the existing result files and the added result files.

3. Click Reuse .

   An Analysis window opens that contains a processing step for each result file in the study.

4. Create a consensus workflow. For instructions, see "Incorporating an Existing Workflow into a Study and an Analysis" on page 99.

5. (Optional) To generate a .pdResult file for each input file, click the **As Batch** check box in the Analysis window title bar.

   When you select the As Batch check box, the application creates a .pdResult file for each input file. When you do not select it, it creates one .pdResult file for all input files.

6. Click Run .

Reprocessing an Existing Analysis

You can change the input files or the workflow used to process the raw data files and then reprocess the existing results. You can decide to reprocess all analysis steps or only the last consensus step. You can also use a new consensus workflow to reprocess MSF results associated with one or more .pdResult files.

Using the Reprocess command with the Reuse command, you can select multiple analysis results at once to reuse for a new consensus from existing analysis results.

❖ **To reprocess an existing search by using different parameters**

1. In an open study, click the **Analysis Results** tab if it is not already selected.

2. Select the search that you want to reprocess.

3. Click Reprocess ▾ :

   • To rerun all the analysis steps, choose **All Analysis Steps**.

   –or–

   • To rerun just the last step in the consensus workflow, choose **Last Consensus Step**.

4. Change the appropriate workflows, settings, parameters, input files, and so forth.

5. In the Analysis window, click Run .

   The job queue opens, displaying the status of your search (see Figure 53 on page 83). Use the job queue to check the status of your search as the search progresses. For information about the job queue, refer to the Help.

❖ **To reprocess existing MSF results by using a new consensus workflow**

1. Click the **Analysis Results** tab.

2. Select the first .pdResult file associated with the MSF file that you want to reprocess with a new consensus workflow.

3. Click Reprocess ▾ and choose **Use Results to Create New (Multi) Consensus**.

   The application adds the analysis associated with the selected .pdResult file, as shown in Figure 62.

Figure 62. Consensus and processing steps associated with a .pdResult file
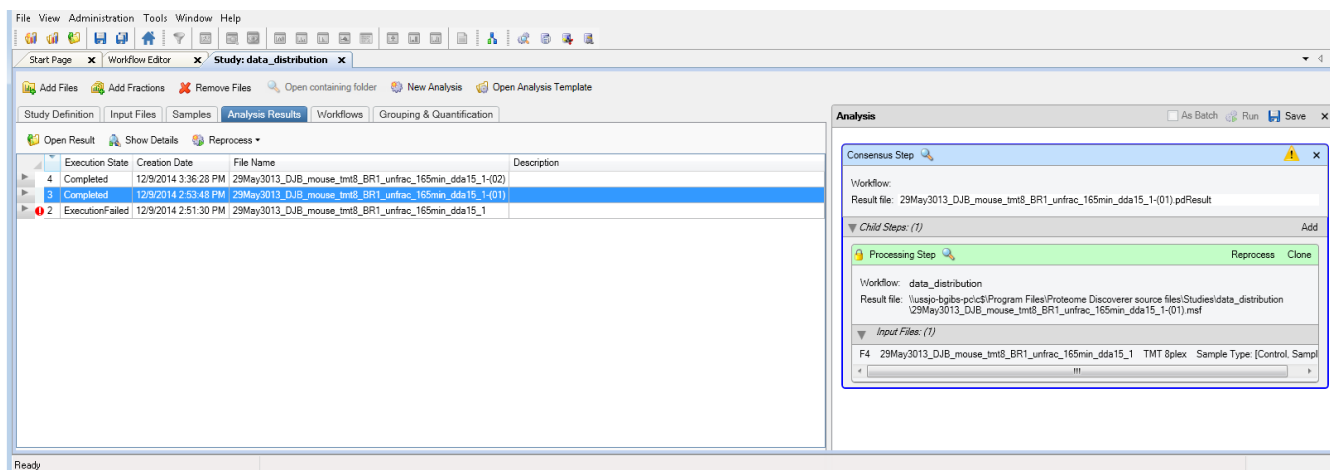
The lock symbol, 🔒, indicates that you cannot change the processing workflow. (If you want to change the processing workflow click **Reprocess** on the Processing Step title bar.)

4. (Optional) Select any other .pdResult files associated with additional MSF files that you want to reprocess with a new consensus workflow, click 🔄 Reprocess ▾ , and choose **Use Results to Create New (Multi) Consensus**.

5. Click the **Workflows** tab, and then click the **Consensus** tab if it is not already open.

6. Create a new consensus workflow. For instructions, see "Incorporating an Existing Workflow into a Study and an Analysis" on page 99.

   –or–

   Open an existing workflow by doing one of the following:

   • In the Workflow Editor within the analysis, click 🔧 Open , browse to the location where you stored the consensus workflow, select the .pdConsensusWF file, and click **Open**.

   –or–

   • If you stored the consensus workflow in the Common Templates directory (see "Using Custom Workflow Templates" on page 106), click 📋 Open Common , select the .pdConsensusWF file, and click **Open**.

   The template opens in the Workflow Tree pane of the Workflow Editor.

7. Click 🔄 Run .

Using an Analysis as a Template

From a selected analysis result, you can create a new analysis template that uses the same sequence steps as the selected result.

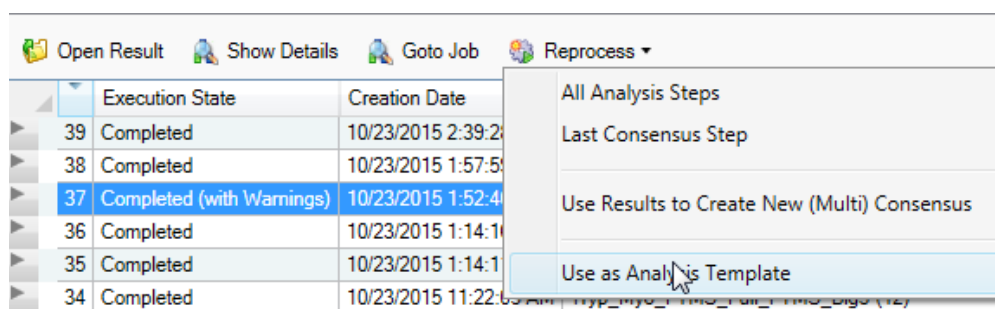❖ **To create an analysis template from selected analysis results**

1. Click the **Analysis Results** tab.

2. Select the appropriate analysis result on the page.

3. Click 🐵 Reprocess ▼ and choose **Use as Analysis Template** (Figure 63).

Figure 63. Use as Analysis Template command

This procedure removes the input files from the analysis and clears the result file names, if they were automatically generated from the input files.

Using the Workflow Editor

For information about creating processing and consensus workflows in the Workflow Editor, see these topics.

- Opening the Stand-alone Workflow Editor

- Closing the Workflow Editor

- Creating a Processing Workflow

- Creating a Consensus Workflow

- Incorporating an Existing Workflow into a Study and an Analysis

- Incorporating an Existing Workflow into a Study and an Analysis

- Workflow Editor Parameters

Opening the Stand-alone Workflow Editor

You can open the Workflow Editor in the context of an analysis, or you can open the stand-alone Workflow Editor outside of an analysis.

❖ **To open the Workflow Editor from inside an analysis**

1. In a study, create an analysis. See "Creating an Analysis" on page 65.

2. Click the **Workflows** tab.

–or–

To create a consensus workflow, click the **Show Workflow** icon, 🔍, on the title bar of the Consensus Step box. To create a processing workflow, click the same icon on the title bar of the Processing Step box.

The Workflow Editor opens in the middle of the Study: *Study_name* page.

❖ **To open the stand-alone Workflow Editor**

Choose **View > Workflow Editor**, or click the **Workflow Editor** icon, 🔣 .

The Workflow Editor opens on a separate page called Workflow Editor.

Closing the Workflow Editor

To close the Workflow Editor in the context of an analysis, you must close the analysis itself.

❖ **To close the Workflow Editor inside an analysis**

Click the **X** on the Analysis title bar.

❖ **To close the standalone Workflow Editor**

Click the **X** on the Workflow Editor tab.

Creating a Processing Workflow

See "Creating a Processing Workflow" on page 68.

Creating a Consensus Workflow

See "Creating a Consensus Workflow" on page 78.

Incorporating an Existing Workflow into a Study and an Analysis

You might want to create a workflow outside of the context of a study or an analysis and incorporate it into a study and an analysis later.

❖ **To incorporate an existing workflow into a study and an analysis**

1. Choose **View > Workflow Editor**, or click the **Workflow Editor** icon, .

   The stand-alone Workflow Editor opens on a separate page.

2. Create the processing workflow (see "Creating a Processing Workflow" on page 68).

3. Click Save , browse to the location where you want to save the .pdProcessingWF file, and click **Save**.

4. Create the consensus workflow. For instructions, see "Incorporating an Existing Workflow into a Study and an Analysis" on page 99.

5. Click Save , browse to the location where you want to save the .pdConsensusWF file, and click **Save**.

6. Create or open a study and an analysis:

   • To create a study, see "Creating a Study" on page 40.

   • To open an existing study, see "Opening an Existing Study" on page 40.

   • To create an analysis, see "Creating an Analysis" on page 65.

   • To open an existing analysis, see "Opening an Existing Analysis" on page 66.

7. Click the **Workflows** tab to open the Workflow Editor within the analysis.

8. In the Workflow Editor within the analysis, click Open, browse to the location where you stored the processing workflow, select the .pdProcessingWF file, and click **Open**.

   –or–

   If you stored the processing workflow in the Common Templates directory (see "Using Custom Workflow Templates" on page 106), click Open Common , select the .pdProcessingWF file, and click **Open**.

   The template opens in the Workflow Tree pane of the Workflow Editor.

9. In the Workflow Editor within the analysis, click Open , browse to the location where you stored the consensus workflow, select the .pdConsensusWF file, and click **Open**.

   –or–

   If you stored the consensus workflow in the Common Templates directory (see "Using Custom Workflow Templates" on page 106), click Open Common , select the .pdConsensusWF file, and click **Open**.

   The template opens in the Workflow Tree pane of the Workflow Editor.

Workflow Editor Parameters

Table 4 describes the parameters in the Workflow Editor.

Table 4. Workflow Editor parameters (Sheet 1 of 2)

Parameter	Description
Workflow Nodes pane	Lists all the nodes available for constructing a processing and a consensus workflow and the parameters for each node. This pane includes the Processing Workflow page and the Consensus Workflow page. For a detailed description of each node shown in these pages, refer to the Help.
Processing Workflow page	Lists all the nodes available for constructing a processing workflow.
Consensus Workflow page	Lists all the nodes available for constructing a consensus workflow.
Parameters page	Lists all the parameters that you can set for a selected node.
Show/Hide Advanced Parameters	Determines whether more advanced parameters are displayed or hidden in the Parameters pane.
Open	Opens the Open Workflow dialog box so that you can select an existing workflow.
Open Common	Opens the Open Workflow dialog box in the Common Templates area so that you can select a processing or consensus workflow template.
Save	Opens the Save Workflow dialog box so that you can save a processing or consensus workflow. By default, the dialog box opens in the study directory. The type of workflow that it saves depends on whether you selected the Processing Workflow tab or the Consensus Workflow tab in the Workflow Nodes pane.
Save Common	Opens the Save Workflow dialog box in the Common Templates area so that you can save a processing or consensus workflow template to the Common Templates directory.
Auto Layout	Automatically adjusts and aligns the connecting arrows and nodes in the Workflow Tree pane in the Workflow Editor. For information about using this command, see "Creating a Processing Workflow" on page 68 and "Incorporating an Existing Workflow into a Study and an Analysis" on page 99.

Table 4. Workflow Editor parameters (Sheet 2 of 2)

Parameter	Description
✖ Clear	Removes the contents of the Workflow Tree pane.
Workflow	Displays the name of the current workflow in the Workflow Tree pane.
Description	Displays a brief description of the current workflow in the Workflow Tree pane.
Workflow Tree pane	Provides a workspace for placing all nodes except the post-processing nodes into a workflow layout.
Post-Processing Nodes pane	Provides a workspace for placing the post-processing nodes in a consensus workflow.

Using Workflow Templates

A workflow template is an XML file that contains information about a processing workflow or a consensus workflow. For processing workflows, this file has a .pdProcessingWF suffix. For consensus workflows, it has a .pdConsensusWF suffix.

The Proteome Discoverer application offers standard workflow templates, or you can create your own.

See these topics:

- Using Common Workflow Templates
- Using Custom Workflow Templates
- Correcting Template Errors
- Saving a Workflow as a Common Template
- Saving a Workflow as a Custom Template
- Deleting a Workflow Template
- Opening a Workflow from a .pdResult File

Using Common Workflow Templates

When you install the Proteome Discoverer application, it creates a Common Templates folder containing subfolders of factory default templates for processing workflows, consensus workflows, analyses, and filter sets. The processing workflows are divided into subfolders of templates appropriate for Q Exactive, LTQ Orbitrap, and Orbitrap Fusion output. You and other users of the system can also use this folder to store common workflow templates that you can then share.

The application installs the Common Templates folder in the following location:

C:\Users\Public\Public Documents\Thermo\Proteome Discoverer *release_number*\Common Templates\

You can use either of the following two methods to access the Common Templates folder and copy your preferred workflow templates to and from it.

❖ **To open the Common Templates folder through the Documents library in Windows Explorer**

1. Open Windows Explorer.

2. Choose **Libraries > Documents > Public Documents > Thermo > Proteome Discoverer 2.1 > Common Templates** (see Figure 64).

Figure 64. Location of the Common Templates folder through the Documents library
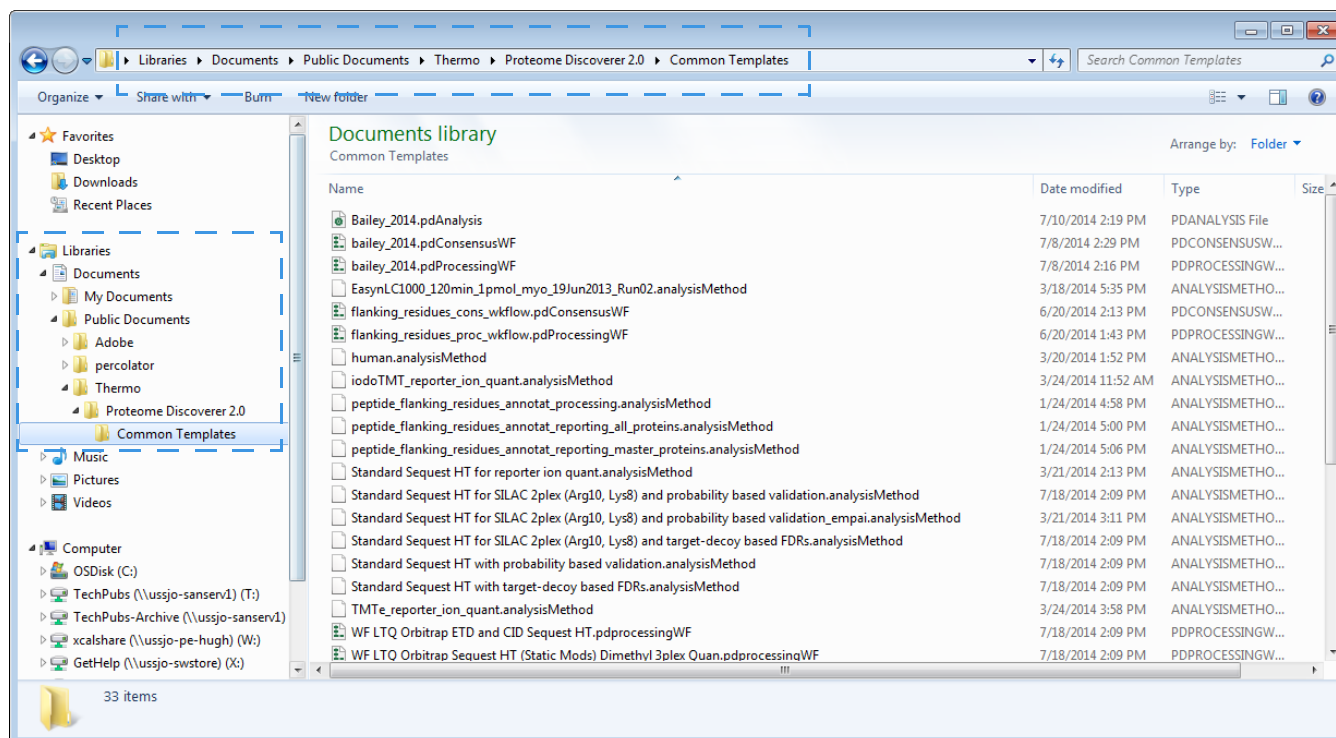
❖ **To open the Common Templates folder directly in Windows Explorer**

1. Open Windows Explorer.

2. Choose **Drive C: > Users > Public > Public Documents > Thermo > Proteome Discoverer 2.1 > Common Templates** (see Figure 65).

Figure 65. Location of the Common Templates folder through the local disk directory
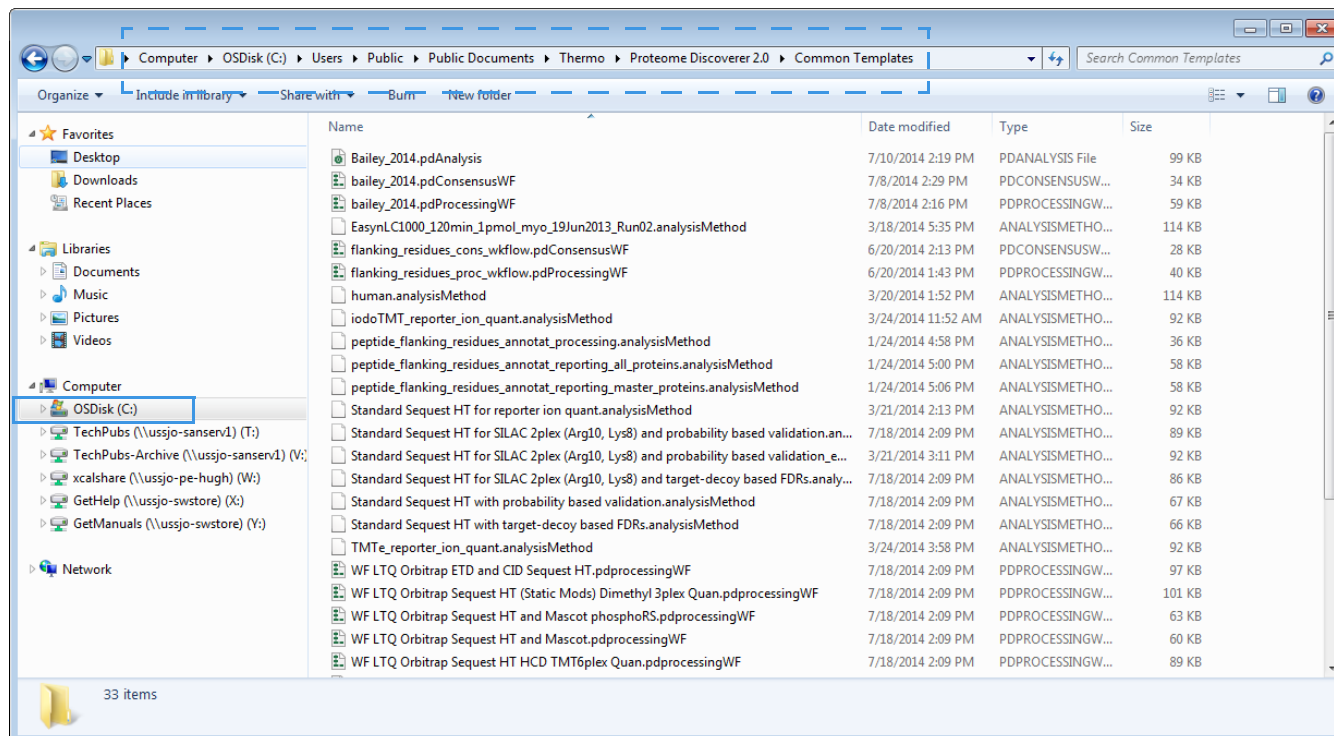
> **Note** The C:\Users\Public\Documents\ folder is a special folder in Windows. The absolute path to this folder, as well as the folder names shown in Windows Explorer, might differ on different systems, depending on your Windows version. Programmatically, always access these special folders by their symbolic names. To find out their exact location on your system, you can refer to the PUBLIC environment variable that is set to the following:
>
> PUBLIC = C:\Users\Public
>
> or to the following registry entry:
>
> HKEY_LOCAL_MACHINE\SOFTWARE\Microsoft\Windows\CurrentVersion\Explorer\Shell Folders\Common Documents
>
> The Documents library is often a concatenation of two different folders. To see them, do the following:
>
> 1. Open Windows Explorer.
>
> 2. Choose **Libraries > Documents**.
>
> 3. Click **2 locations**.
>
> The Documents Library Locations dialog box (Figure 66), opens to show the actual folders that are joined together.

Figure 66. Document Library Locations dialog box
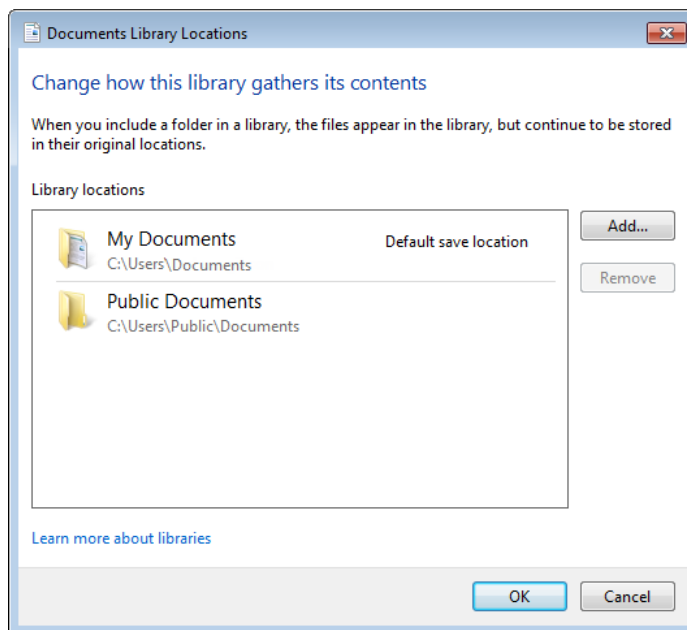
Follow these instructions to open a template in the Common Templates folder.

❖ **To open a common workflow template**

1. Create an analysis (see "Creating an Analysis" on page 65), or open an existing analysis (see "Opening an Existing Analysis" on page 66).

2. Click the **Workflows** tab.

3. Click the Consensus Step bar or the Processing Step bar in the Analysis window.

4. In the Workflow Editor window, click 📇 Open Common .

   If you clicked the Processing Step bar, the Open Workflow dialog box displays .pdProcessingWF files in the Common Templates folder. If you clicked the Consensus Step bar, the dialog box displays .pdConsensusWF files in the Common Templates folder.

5. In the Open Workflow dialog box, select a .pdProcessingWF file for a processing workflow or a .pdConsensusWF file for a consensus workflow from the Common Templates folder.

6. Click **Open**.

The workflow opens in the Workflow Tree area. If you need to correct any template errors, see "Correcting Template Errors" on page 106.

❖ **To save a template as a common workflow template**

1. Complete the processing or consensus workflow in the Workflow Tree pane of the Workflow Editor page.

2. In the Workflow box above the Workflow Tree pane, type a name for the workflow.

3. (Optional) In the Description box, type a brief description of the processing workflow.

4. Click 📇 Save Common .

5. In the Save Workflow dialog box, do the following:

   a. In the File Name box, type a name for the workflow.

   b. Click **Save**.

The application saves the processing workflow in a *file_name*.pdProcessingWF file and the consensus workflow in the *file_name*.pdConsensusWF file.

Using Custom Workflow Templates

Custom workflow templates are templates that you created yourself and stored. If you saved a workflow as part of an analysis, you can access the workflow by opening the analysis template.

❖ **To open a custom workflow template for a new analysis**

1. Create an analysis. For instructions, see "Using Analyses" on page 64.

2. Click ⬚ Open Analysis Template .

3. In the Open Analysis Template dialog box, select the .pdAnalysis file that you want to use as a template.

4. Click **Open**.

❖ **To open a custom workflow template in an existing analysis**

1. Open an existing analysis.

2. Click the **Workflows** tab.

3. Click the Consensus Step or the Processing Step title bar in the Analysis window.

4. In the **Workflow Editor** window, click ⬚ Open .

5. In the Open Workflow dialog box, browse to the .pdProcessingWF or .pdConsensusWF, file that the custom template is stored in, and click **Open**.

The workflow opens in the Workflow Tree area. To correct any template errors, see "Correcting Template Errors."

Correcting Template Errors

When you open a template file, either standard or custom, the Proteome Discoverer application validates parameter settings and displays the selected workflow in the Workflow Tree area of the Workflow Editor. The name of the workflow appears in the Workflow box, and a description of the workflow appears in the Description box. The application uses warning symbols to indicate outdated nodes and to display error information in the Workflow Failures pane.

If the application cannot load the selected file, it displays a message box with information about the issue. It cannot load files that cannot be updated, are read-only or invalid, or were created with a newer version of the Proteome Discoverer application.

When you open an existing workflow template, some of the nodes in the Workflow Tree pane might show a yellow triangle warning symbol (Figure 67). This symbol indicates that the version of the node that was used when the template was created has been superseded by a later version in the current Proteome Discoverer application. When you save or re-save the workflow, the application automatically updates the node to the latest version.

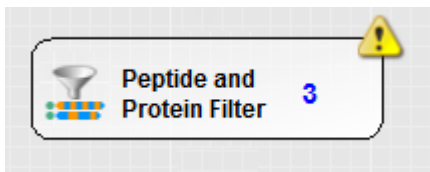**Figure 67.** Warning symbol of an outdated node version

Figure 68 shows an exclamation mark inside a blue warning circle, which indicates that one or more of the parameter settings for the node are incorrect or outdated. Click on the node and reset the parameters in the Parameters pane.

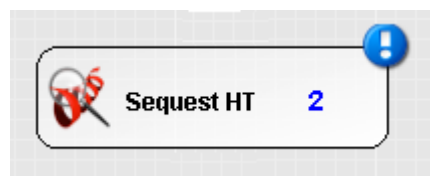**Figure 68.** Warning symbol of incorrect or outdated parameter settings

Figure 69 shows a blocked gear (or settings) symbol, which indicates that the flow contains a node that is no longer available for the given workflow because it is outdated, not installed, installed but not licensed, or has incorrect parameter settings.

Figure 69. Warning symbol of a node no longer available for the given workflow
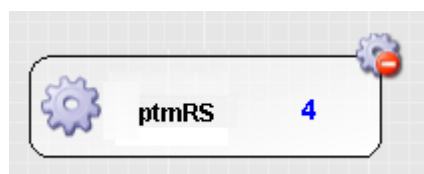
When you open a workflow that contains an unavailable node, a Load Failures pane opens beneath the Workflow Nodes pane (see Figure 70).

If this warning symbol is attached to a node, the application cannot update it because it is not licensed or no longer licensed, it has been replaced by a different node, or the node was moved to the consensus workflow. In these cases, it does not permit you to save the workflow.

Figure 70. Load Failures pane

Load Failures pane

The Load Failures pane has three columns:

- Error Information: Displays information about the problem that the application encountered in the workflow.

- Parameter: Displays the name of the node parameter that has an erroneous setting.

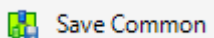- Value: Displays the erroneous setting of the node parameter.

When a warning symbol is attached to a node, the application automatically updates the node with the correct version and saves the previous parameter values in the updated node. It does not include node parameters that are no longer available, and it adds any new parameters set to their default values.

If the Parameter and Value columns indicate a problem with the parameter settings, enter the correct parameter settings in the Parameters pane.

Saving a Workflow as a Common Template

If you want to share a changed standard template or a new custom workflow with others, you can save it as a standard template in the Common Templates folder.

❖ **To save a workflow as a common template**

1. To save a common workflow under a new file name or to save a new custom workflow, type a new name in the Workflow box of the Workflow Editor and a new description in the Description box.

2. Click ![Save Common icon] Save Common .

   By default, the Save Workflow dialog box opens in the Common Templates directory. For the location of this folder, see "Using Common Workflow Templates" on page 102.

3. In the Save Workflow dialog box, do the following:

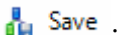   a. In the File Name box, type the file name.

      If you altered a standard workflow template, the name of this file must be different from the name of the original template.

   b. Click **Save**.

Saving a Workflow as a Custom Template

You can save a new workflow as a custom template for your own use.

❖ **To save a workflow as a custom template**

1. Type a new name in the Workflow box of the Workflow Editor and a new description in the Description box.

2. Click ![Save icon] Save .

   By default, the Save Workflow dialog box opens in the last directory used.

3. In the Save Workflow dialog box, do the following:

   a. If you do not want to save the workflow template in the study directory, browse to the directory where you want to save the custom template.

   b. In the File Name box, type the file name.

   c. Click **Save**.

Deleting a Workflow Template

You can delete a workflow template.

❖ **To delete an existing workflow template**

1. In Windows Explorer, navigate to the folder containing the template.

2. Select the template to delete, and either press the DELETE key or right-click and choose **Delete**.

3. In the Delete File dialog box, click **Yes**.

Opening a Workflow from a .pdResult File

From a .pdResult file, you can open the workflow that generated the file.

❖ **To open the workflow that generated a .pdResult file**

Do one of the following:

a. If you recently generated a .pdResult file, choose **Administration > Show Job Queue** to open the job queue.

b. Select the appropriate search.

c. Click Open Study .

–or–

a. In the Workflow Editor window of an open analysis, click Open .

b. In the File Type box of the Open Workflow dialog box, select **Proteome Discoverer Result File (*.pdResult)**.

c. Browse to the location of the appropriate .pdResult file, select the file, and click **Open**.

You can also obtain information about the workflow that generated a .pdResult file by using the file's Search Summary. For information on the Search Summary, refer to the Help.

Creating Specific Types of Workflows

Follow these procedures to create specific types of workflows.

- Creating Quantification Workflows
- Creating Protein Annotation Workflows
- Creating PTM Analysis Workflows
- Creating Parallel Workflows

- Creating HCD/EThcD Workflows

- Creating Spectra Exportation Workflows

- Creating a Marked Contaminants Workflow

Creating Quantification Workflows

To perform quantification, you must run quantification processing and consensus workflows. A quantification processing workflow includes one of the quantification nodes found in the Feature Detection & Quantification section of the Workflow Nodes pane of the Processing Workflow window. A quantification consensus workflow includes the Reporter Ions Quantifier or the Precursor Ions Quantifier node, found in the Quantification section of the Workflow Nodes pane of the Consensus Workflow window.

Table 5 lists these nodes and where to find information about creating a quantification workflow for each one.

Table 5. Quantification nodes in the processing and consensus workflows

Quantification node	Use	For more information
Processing workflow		
Minora Feature Detector	For precursor ion quantification (for example, SILAC) and label-free quantification	See "Creating a Processing Workflow for Precursor Ion Quantification" on page 449 or "Creating a Processing Workflow for Label-Free Quantification" on page 492.
Reporter Ions Quantifier	For reporter ion quantification (for example, iTRAQ and TMT)	See "Creating a Processing Workflow for Reporter Ion Quantification" on page 386.
Consensus workflow		
Precursor Ions Quantifier	For precursor ion and label-free quantification	See "Creating a Consensus Workflow for Precursor Ion Quantification" on page 452 and "Creating a Consensus Workflow for Label-Free Quantification" on page 495.
Reporter Ions Quantifier	For reporter ion quantification	See "Creating a Consensus Workflow for Reporter Ion Quantification" on page 390.
Feature Mapper	For precursor ion and label-free quantification	See "Creating a Consensus Workflow for Precursor Ion Quantification" on page 452 and "Creating a Consensus Workflow for Label-Free Quantification" on page 495.

In the processing workflow, you must attach the Minora Feature Detector node directly to the Spectrum Files node.

Creating Protein Annotation Workflows

To create a workflow that uses the ProteinCenter Annotation node to retrieve annotation information, see "Creating a Protein Annotation Workflow" on page 265. This node loads information from the GO Slim, Pfam, Entrez Gene, Ensembl Genome, and UniProt databases from ProteinCenter and installs it in the Proteome Discoverer results files.
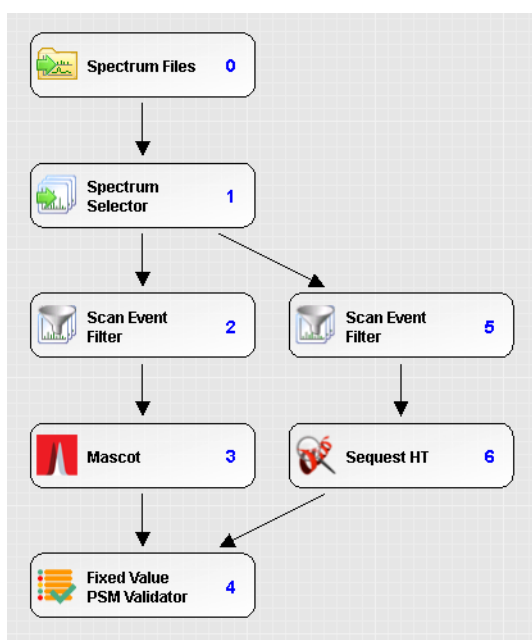
Creating PTM Analysis Workflows

To create a workflow that enables you to focus on studying the biologically relevant post-translational modifications of proteins, you can create a PTM analysis workflow to display phosphorylation modifications in the .pdResult results. For instructions, see "Creating a PTM Analysis Workflow with the ptmRS Node" on page 320.

Creating Parallel Workflows

Parallel workflows are workflows that search the same raw data file and the same part of the spectrum but specify different criteria, different search nodes for the search, or both. The processing part of these workflows resembles the example in Figure 71. You can use parallel workflows to conduct two or more searches that use two or more search engines on the same raw data and, at the same time, compare the results of these two searches. For example, you might want to search both CID and ETD data from the same raw data file to increase the chances of finding a match. CID data contains b and y ions, and ETD data contains b, c, and z ions, so the two types of data are complementary. You can also use a parallel workflow for quantification.

Figure 71. Parallel processing workflow

The following instructions show you how to create the basic parallel workflow shown in Figure 71.

❖ **To create a parallel processing workflow**

1. Create or open a study and an analysis:

    • To create a study, see "Creating a Study" on page 40.

    • To open an existing study, see "Opening an Existing Study" on page 40.

    • To create an analysis, see "Creating an Analysis" on page 65.

    • To open an existing analysis, see "Opening an Existing Analysis" on page 66.

2. Follow the general instructions in "Creating a Processing Workflow" on page 68.

3. Drag the **Spectrum Files** node to the Processing Workflow Tree pane, and specify the name and path of the raw data file in the Parameters pane.

4. Drag the **Spectrum Selector** node to the Processing Workflow Tree pane and place it directly under the Spectrum Files node. Set the parameters.

5. Drag two **Scan Event Filter** nodes to the Processing Workflow Tree pane and place them side by side beneath the Spectrum Selector node. In the Parameters pane, set the Activation Type parameter to **CID** for one node and to **ETD** for the other node.

6. Drag the a search engine node, such as **Sequest HT**, to the Processing Workflow Tree pane and place it beneath the Scan Event Filter node set to the **CID** activation type.

7. Drag another search engine node, such as **Mascot**, to the Processing Workflow Tree pane and place it beneath the Scan Event Filter node set to the **ETD** activation type.

8. Drag a PSM Validator node, such as **Fixed Value PSM Validator**, beneath one of the search engine nodes.

9. Connect the nodes as shown in Figure 71 on page 112.

10. Create a consensus workflow, following the instructions in "Incorporating an Existing Workflow into a Study and an Analysis" on page 99.

11. Save the analysis (see "Saving an Analysis" on page 83).

12. Save the study (see "Saving a Study" on page 56).

13. Click ⚙ Run in the Analysis window.

    The job queue opens (see Figure 49 on page 78), displaying the status of your search. Use the job queue to check the status of your search as the search progresses. For job queue information, refer to the Help.

Creating HCD/EThcD Workflows

To improve the quality of ETD fragmentation, mass spectrometers, such as the Orbitrap Fusion, can supplement ETD with HCD or CID activation. These supplemental activations are called EThcD and ETciD activations in recent instrument firmware.

- ETciD activation, formerly known as "supplemental activation," produces the same fragment ions as ETD-only activation, so changing the search parameters is unnecessary. ETciD scans or ETD scans with "sa" in their names are simply recognized as ETD scans.

- EThcD activation generates additional b and y ions, so you must change the search parameters. These parameters are different from those for ETD- or HCD-only activation.

If the acquisition method used HCD and EThcD activations simultaneously, the raw data file contains spectra from both. In these cases, the search workflow splits the scan by activation type and searches HCD and EThcD spectra separately with the appropriate settings.

A typical processing workflow for raw data files containing both HCD scans and EThcD scans that are acquired in parallel uses Scan Event Filter nodes to split the scan by activation type. The Proteome Discoverer application then searches the HCD scans with b and y ions and the EThcD scans with c and z ions and a certain number of b and y ions.

Figure 72 shows a typical processing workflow for searching a raw data file with both HCD and EThcD spectra. In this workflow, the Activation Type parameter of one Scan Event Filter node is set to HCD, and the same parameter for the other Scan Filter node is set to EThcD.

Figure 72. Typical search workflow for the simultaneous search of EThcD and HCD spectra
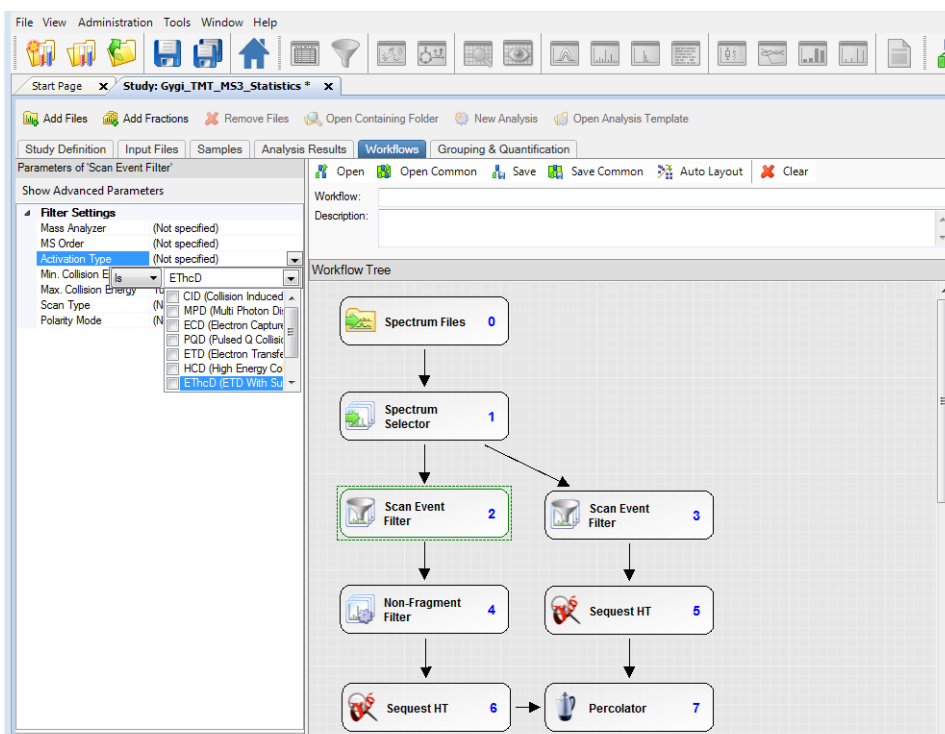
Figure 73 shows the results of the workflow presented in Figure 72.

Figure 73. Search results from the simultaneous search of HCD and EThcD spectra in the same raw data file
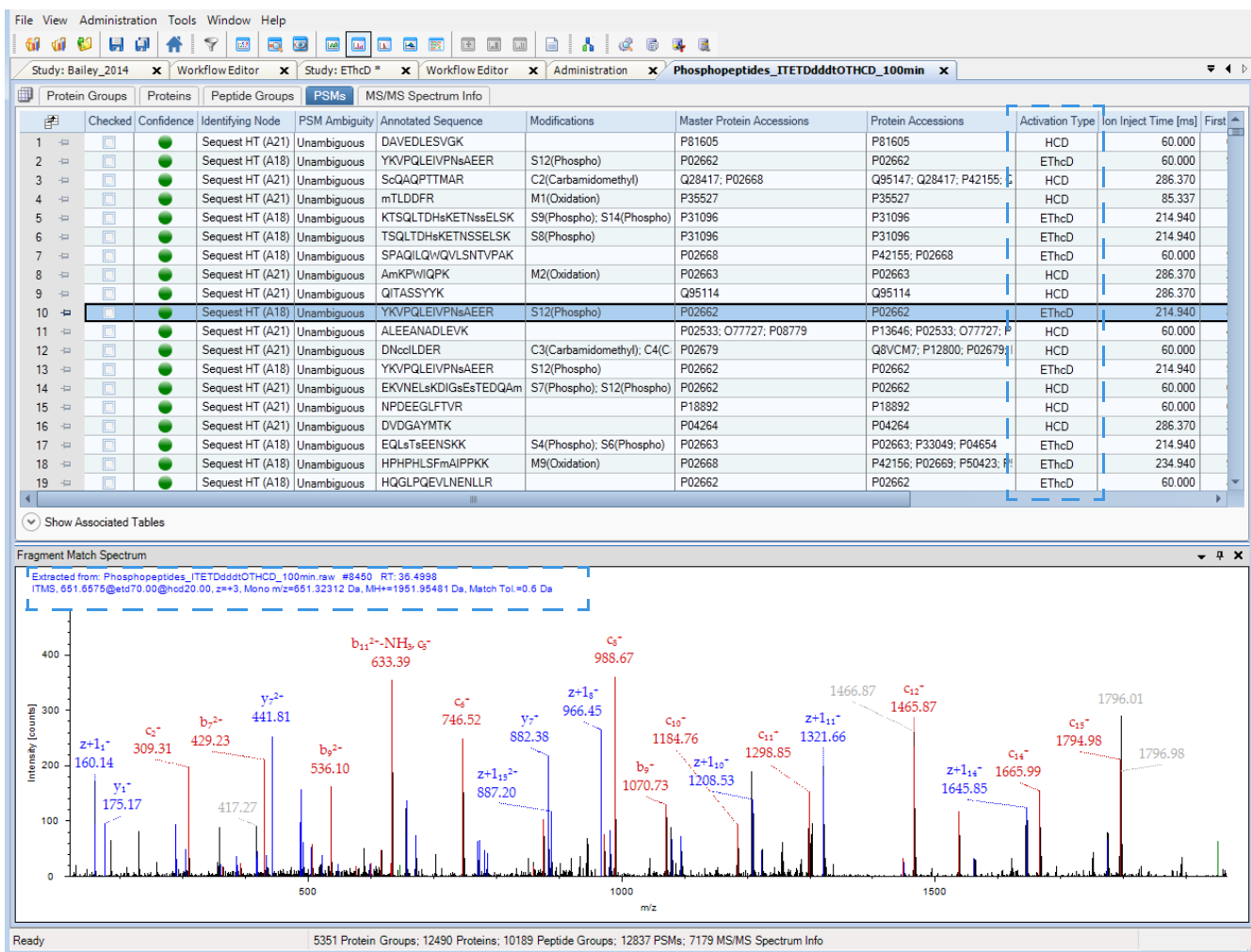
Creating Spectra Exportation Workflows

You can create a workflow that exports data to DTA, MGF, MZDATA, and MZML files. For information on creating this workflow, refer to the Help.

Creating a Marked Contaminants Workflow

You can mark as contaminants all proteins in the results file that are listed in a designated FASTA file or files. To create this workflow, see "Identifying Contaminants During Searches" on page 190.

Searching Multiple Sequence Databases with Mascot

If you connect the Proteome Discoverer application to Mascot server 2.3 or later, you can search more than one sequence database in the same search.

❖ **To search multiple sequence databases with Mascot**

1. Connect to Mascot server 2.3 or later. For instructions, see "Configuring the Mascot Search Engine" on page 26.

2. Create or open a study and an analysis:

   - To create a study, see "Creating a Study" on page 40.

   - To open an existing study, see "Opening an Existing Study" on page 40.

   - To create an analysis, see "Creating an Analysis" on page 65.

   - To open an existing analysis, see "Opening an Existing Analysis" on page 66.

3. Create a processing workflow (see "Creating a Processing Workflow" on page 68).

4. For the search engine node, drag the **Mascot** node to the Workflow Tree pane in the Processing Workflow window.

5. To set the parameters for the Mascot node, select two or more databases in the Protein Database box (see Figure 74).

Figure 74. Selecting multiple sequence databases for the Mascot node
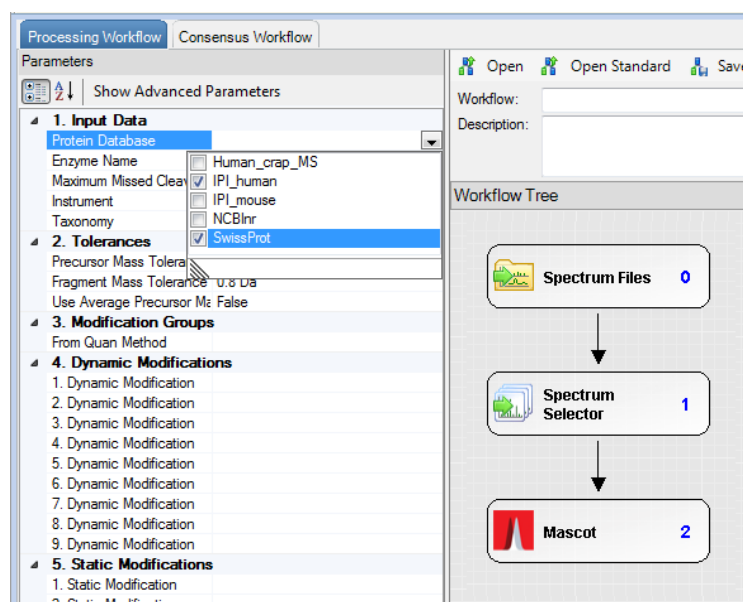
6. Continue with the workflow in "Creating a Processing Workflow" on page 68.

When you run a search against multiple sequence databases, the results include proteins from both or all databases (see Figure 75).

Figure 75. Result of performing a Mascot search against multiple sequence databases

Creating a Multiconsensus Report

A multiconsensus report is a .pdResult file generated from a workflow that includes multiple search engines or multiple input files. For the input files, you can use any of the input files listed in "Inputs and Outputs" on page 10 or MSF files.

The .pdResult report displays the combined results in columns that are labeled by search engine type, for example, Score Mascot and Score Sequest HT.

• On the Proteins page, the #PSMs, # Peptides, Coverage, and Score columns display information side by side for each protein. When the proteins referenced in the individual searches have different accession numbers, the application displays the accession number of the first search in the result set.

• On the PSMs page, the Search ID column displays the order in which the search was submitted to the job queue. A multiconsensus report resulting from a workflow that includes both Sequest HT and Mascot contains XCorr (Sequest HT) or Ions Score (Mascot) columns. These columns score the number of fragment ions that are common to two different peptides with the same precursor mass and calculate the cross-correlation score for all candidate peptides queried from the database. The default setting shows the top matches per peptide and search engine.

The identification information from the individual search nodes appear side by side on the PSMs and Peptide Groups pages. The grouped peptides do not represent actual matches found during the search but rather the unified information from all matches found for this particular peptide sequence.

You can access the peptide information by examining both the Proteins page for all peptides associated with a protein and the PSMs page for all peptides, including those not associated with any protein. You can view the associated peptides on the Proteins page. Select a protein row and click **Show Associated Tables** at the bottom of the page.

To filter multiconsensus reports for specific results, see "Finding Common or Unique Proteins in Multiple Searches" on page 242 and "Applying Filters Specific to Different Searches in Multiconsensus Reports" on page 246.

❖ **To create a multiconsensus report from a search using multiple input files**

See "Adding Input Files" on page 42.

❖ **To create a multiconsensus report by using multiple search engines**

1. Create or open a study and an analysis:

   • To create a study, see "Creating a Study" on page 40.

   • To open an existing study, see "Opening an Existing Study" on page 40.

   • To create an analysis, see "Creating an Analysis" on page 65.

   • To open an existing analysis, see "Opening an Existing Analysis" on page 66.

2. Add the appropriate input file or files.

3. Create a processing workflow (see "Creating a Processing Workflow" on page 68).

   In the workflow, include at least two of the search engine nodes listed under Sequence Database Search in the Workflow Nodes pane (Figure 76).

   Figure 76. Workflow with multiple search engines
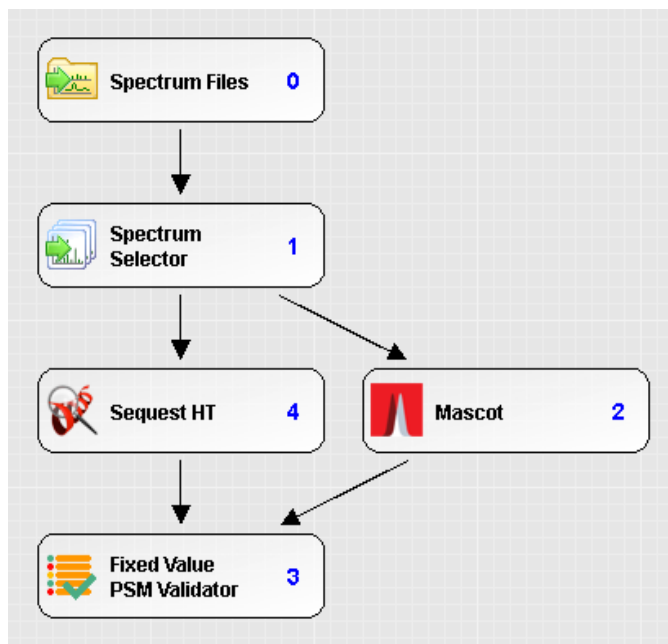
4. Create a consensus workflow by following the instructions in "Incorporating an Existing Workflow into a Study and an Analysis" on page 99.

5. For multiconsensus reports that include multiple occurrences of the same type of search nodes, set the Merge Mode parameter of the MSF Files node (Figure 77) to determine how the application merges the data. Use one of the following settings:

- Globally by Search Engine Type: See "Globally by Search Engine Type Setting" on page 120.

- Per File and Search Engine Type: See "Per File and Search Engine Type Setting" on page 121.

- Do Not Merge: See "Do Not Merge Setting" on page 122.

For more information on the MSF Files node, refer to the Help.

Figure 77. Setting the Merge Mode parameter of the MSF Files node
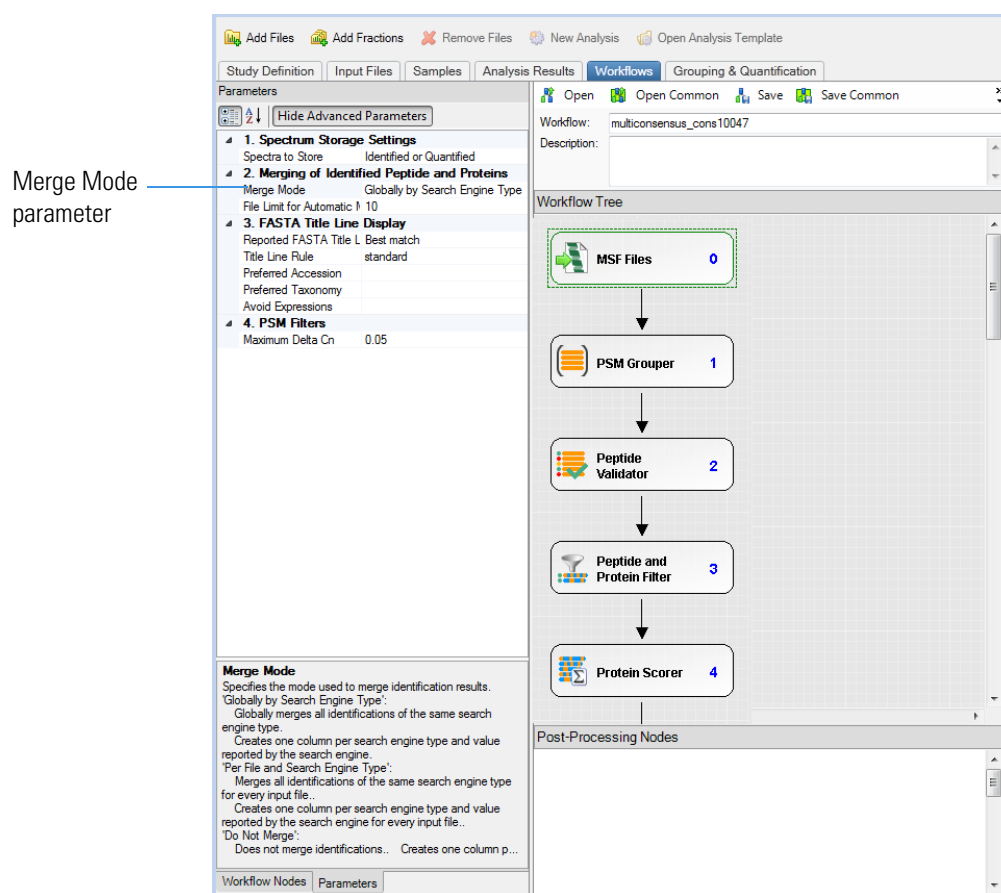
6. Click [Run] in the Analysis window.

The job queue opens (Figure 53 on page 83), displaying the status of your search. Use the job queue to check the status of your search as the search progresses. For information about the job queue, refer to the Help.

Merging Results by Search Engine Node

When you generate a .pdResult file from a workflow containing multiple search engines, you can merge the search results of each search engine type (such as Sequest HT or Mascot) to prevent too many result columns from appearing in the report. In addition, you might want to divide the results by search engine—for example, if someone wants to review the differences between instrument methods used in the different files or compare search engine settings.

You might want to merge search results in the following cases:

- The application creates a report from many MSF files (for example, 50), and each contains at least one search node.

- The application creates a report from only a few MSF files (for example, three), but each workflow contains 10 search nodes.

In both cases, you might prefer to reduce the number of data columns to obtain a clear overview of the results.

When you create a results report generated by a workflow that contains MSF files generated by several search nodes, you can use the following settings for the Merge Mode parameter of the MSF Files node to display the search-engine-specific data on the Protein Groups and the Proteins pages:

- Globally by Search Engine Type
- Per File and Search Engine Type setting
- Do Not Merge

The following topics illustrate the Merge Node parameter settings. They are based on the processing of two files with a processing workflow that contains one Mascot node and two Sequest HT nodes that use different sequence databases.

For more information on the MSF Files node, refer to the Help.

Globally by Search Engine Type Setting

The default setting of the Merge Mode parameter of the MSF Files node, Globally by Search Engine Type, merges the results of all search nodes of the same type. It displays the best PSM values on the Peptide Groups page and shows the combined protein score that was calculated from all PSMs identified by the search engine type.

Figure 78 displays the merged results of all Sequest HT and Mascot nodes and one protein score by node type on the Peptide Groups page. The results from the two different search engine types appear together in two different column groups. The top illustration shows them grouped on the Peptide Groups page, and the bottom illustration shows them grouped on the Proteins page.

Figure 78. Results grouped by search engine type on the Peptide Groups page and the Proteins page

Per File and Search Engine Type Setting

The Per File and Search Engine Type setting of the Merge Mode parameter of the MSF Files node merges the results of all search engine nodes of the same type for each input file.

In Figure 79, the results report displays one group of search-engine-specific values per file and the search engine type. There are two groups of Sequest HT columns and two groups of Mascot columns. The top illustration shows the data grouped on the Peptide Groups page, and the bottom illustration shows it grouped on the Proteins page.

Figure 79. Results grouped by search engine type for each input file on the Peptide Groups page and the Proteins page

Do Not Merge Setting

The Do Not Merge setting of the Merge Mode parameter of the MSF Files node does not merge nodes. It displays the search-engine-specific values for each search engine. The File Limit for Automatic Merge parameter of the MSF Files Node sets an upper limit on the number of files that the report can display for unmerged data.

Figure 80 shows four groups of Sequest HT columns and two groups of Mascot columns. The top illustration shows the data grouped on the Peptide Groups page, and the bottom illustration shows it grouped on the Proteins page.

Figure 80. Results in one group of search-engine-specific values per search node on the Peptide Groups page and the Proteins page

Using the Proteome Discoverer Daemon Utility

This chapter describes the Proteome Discoverer Daemon utility, which you can use to perform a search either locally or remotely by using its own application window, the command line, or a parameter file that contains command-line commands. You can use it to start workflows, monitor job execution on the configured server, perform batch processing, and process Multidimensional Protein Identification Technology (MudPIT) samples. The Daemon utility can perform multiple searches on multiple raw data files taken from multiple samples or from one sample.

Contents

- Starting the Daemon Utility in a Window

- Selecting the Server

- Running the Daemon Utility from the Window

- Monitoring Job Execution in the Daemon Utility

- Logging in to a Remote Server

- Running the Daemon Utility from the Xcalibur Data System

- Running the Daemon Utility from the Command Line

For information about MudPIT and creating a MudPIT workflow, see Adding Input Files.

Starting the Daemon Utility in a Window

You can start the Proteome Discoverer Daemon utility on the command line or in a window. To run it on the command line, see Running the Daemon Utility from the Command Line.

❖ **To start the utility in a window**

1. Start the utility in Windows by choosing **Start > All Programs > Thermo Proteome Discoverer** *release_number* **> Proteome Discoverer Daemon** *release_number* or by clicking the **Daemon** icon, , on your desktop.

2. After the utility window opens, connect to a computer that is running the Proteome Discoverer application.

Selecting the Server

The Daemon utility can connect to a remote computer running the Magellan server so that you can perform searches on multiple raw data files from multiple samples or a single sample on a remote computer. It can also connect to a local server.

❖ **To specify the server to connect to**

1. Click the **Configuration** tab in the Daemon utility window.

2. From the Host list, select the name of the server that you want to use, or type the server name.

   You must connect the utility to a computer running the Magellan server. Your local host is the default server, that is, the computer that you are working on. To connect remotely to the Magellan server, see "Logging in to a Remote Server" on page 128.

3. In the User box, type the user's login name on the server.

   The Configuration page now resembles Figure 81.

   Figure 81. Configuration page of the Proteome Discoverer Daemon utility
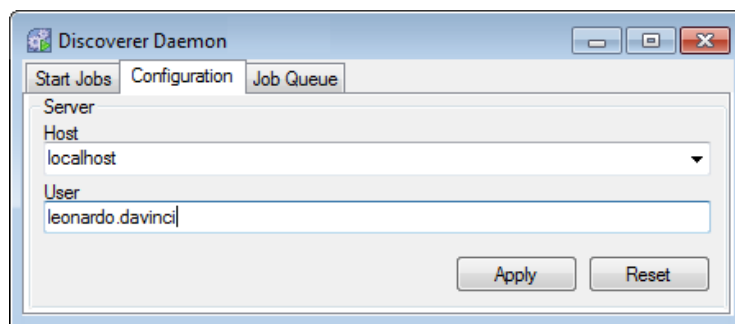
4. Click **Apply** to activate the new settings.

5. To return to the previous settings, click **Reset**.

Running the Daemon Utility from the Window

You can use the Daemon utility to run processing and consensus workflows on a remote server. These workflows are saved in templates that you created in the Workflow Editor. You must set the values of the required parameters for the workflow templates before you run the utility. The specified values—for example, FASTA databases—must be available on the server where they will be processed. You cannot change the workflows by using the Proteome Discoverer Daemon interface.

You can choose quantification methods from predefined quantification methods on the server or from a method file that you load with the Browse button.

❖ **To start a workflow for batch processing or MudPIT processing**

1. Click the **Start Jobs** tab.

   The Start Jobs page opens (Figure 82).

   Figure 82. Start Jobs page of the Daemon utility
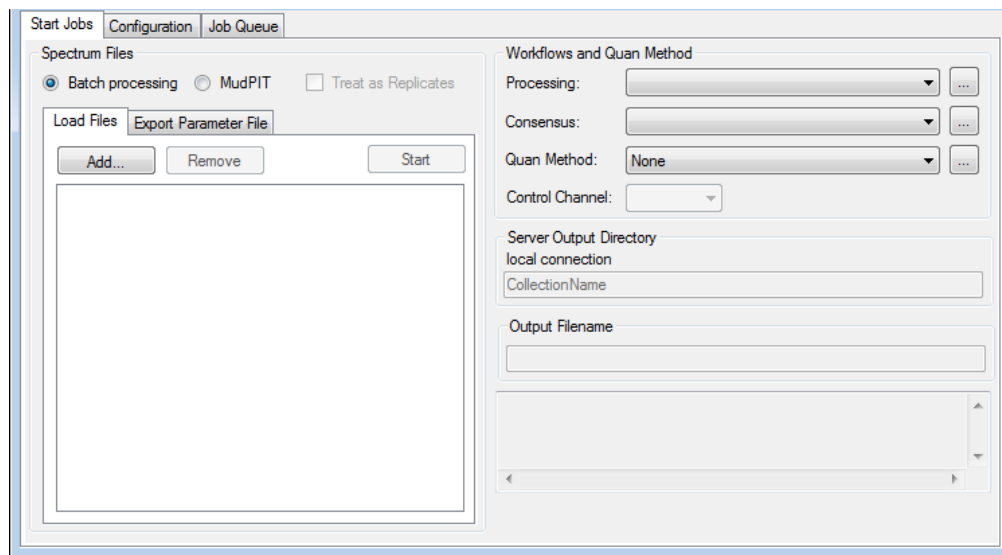
2. Click the **Load Files** tab if it is not already selected.

3. Click **Add**.

4. In the Open dialog box, locate the folder that contains the raw data files, select the spectrum (RAW) files that you want to load, and click **Open**.

   The selected spectrum files appear on the Load Files page.

   To remove a file from the Load Files page, select the file and click **Remove**.

5. To specify the type of processing, select the **Batch Processing** or **MudPIT** option.

   - Batch processing (the default): Executes the workflow once for each spectrum file.

   - MudPIT: Feeds all spectrum files into one workflow.

   When you select the MudPIT option, the Treat as Replicates check box and the Output Filename box become available.

6. (Optional) For the MudPIT option, select the **Treat as Replicates** check box to group the samples by quantification channel and to average the abundance values for each channel across files.

   This option is only available when you select the MudPIT option.

   If you do not select the Treat as Replicates option, the application groups the samples by quantification channel and by file. It creates the sample groups and ratios only within files, not across files.

7. Select a processing template from template files located on the Daemon server.

   • In the Workflows and Quan Method area of the Start Jobs page, do one of the following:

   Click the **Browse** button (...), select the processing template from the Open dialog box, and click **Open**.

   –or–

   Select a processing workflow template from the Processing menu.

   The menu box lists only the templates that you selected in the current run. When you restart the application, the menu box does not list any templates. When you connect the application to another server, the application verifies that the templates are still valid and can be run by the connected server.

   The template files for processing workflows have a .pdProcessingWF extension.

8. Select a consensus template.

   • In the Workflows and Quan Method area of the Start Jobs page, do one of the following:

   Click the **Browse** button (...), select a consensus workflow template from the Open dialog box, and click **Open**.

   –or–

   Select a consensus workflow template from the Consensus menu.

   The menu box lists only the templates that you selected in the current run. When you restart the application, the menu box does not list any templates. When you connect the application to another server, the application verifies that the templates are still valid and can be run by the connected server.

   The template files for consensus workflows have a .pdConsensusWF extension.

9. To obtain quantification results, select a quantification method from the Quan Method menu in the Workflows and Quan Method area of the Start Jobs page.

10. (Optional) To specify the control channel to use in creating ratios when you process quantification samples, select the control channel from the Control Channel list.

11. If you selected the MudPIT option, in the Output Filename box, type the name of the output file where you want to store the results of the search.

    The Start Jobs page should now resemble Figure 83 for batch processing or Figure 84 for MudPIT processing.

12. Click **Start** to execute the job.

Figure 83.   Start Jobs page of the Daemon utility for batch processing
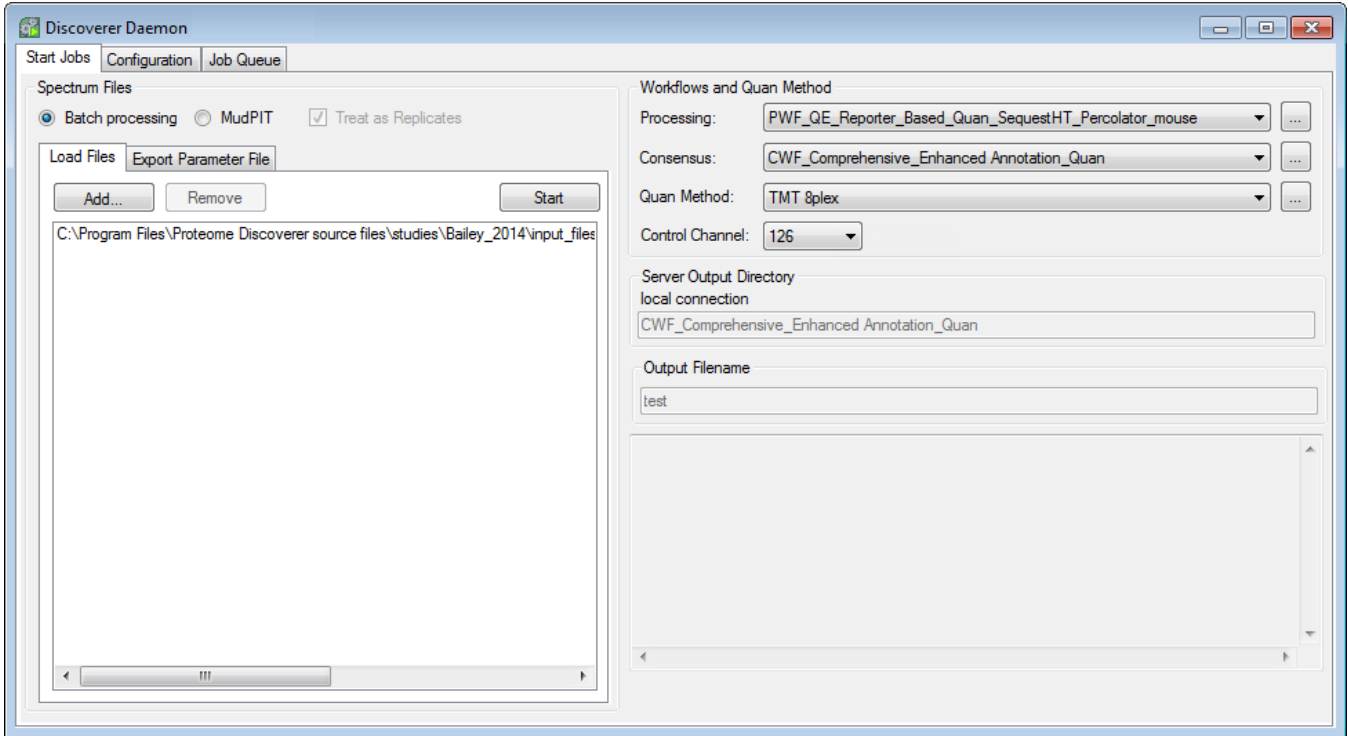
Figure 84.   Start Jobs page of the Daemon utility for MudPIT processing
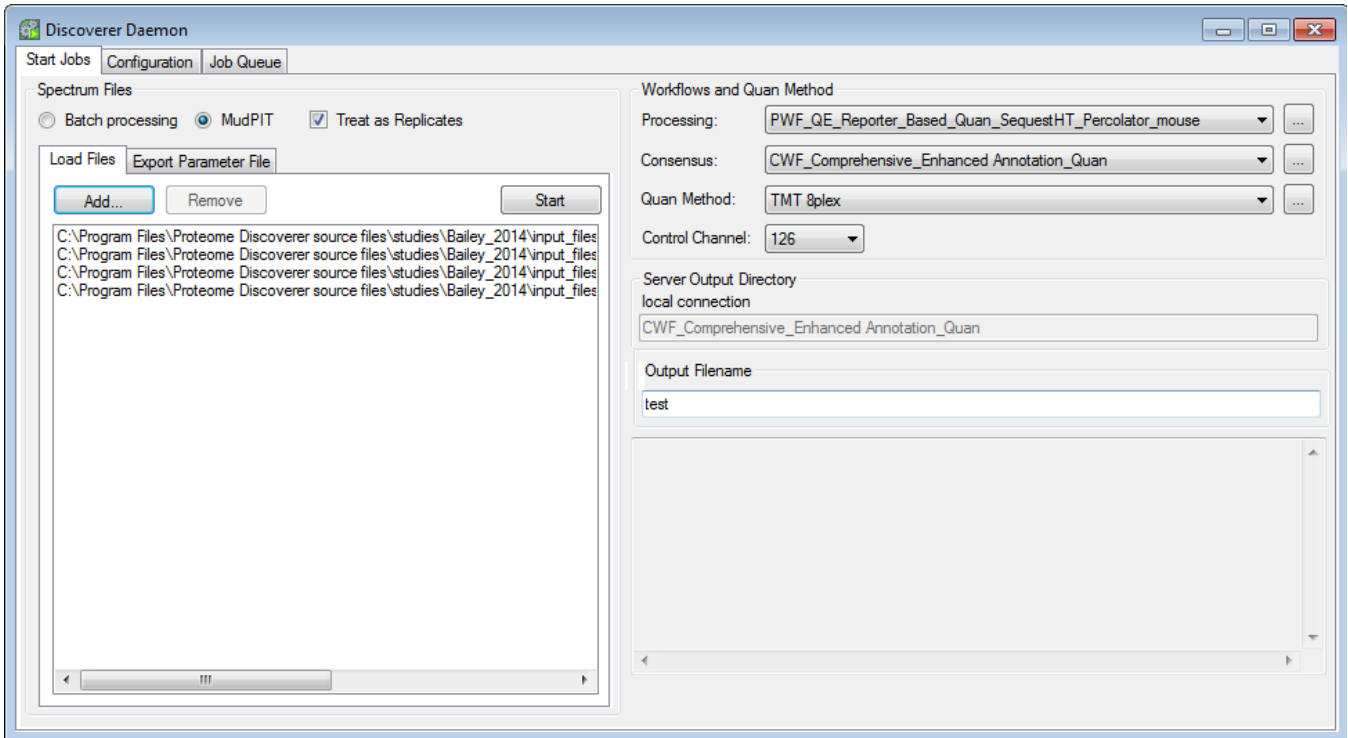
Monitoring Job Execution in the Daemon Utility

You can use the Job Queue page in the Daemon utility window to monitor the execution of the jobs that you submit. It performs the same function as the job queue on the Proteome Discoverer interface. For information about the features of the job queue on the Proteome Discoverer interface, refer to the Help.

A progress bar displays the progress of the overall batch processing. This progress bar is visible only if you have started batch jobs.

❖ **To monitor the job execution**

Click the **Job Queue** tab of the Daemon utility window.

Figure 85 shows the completed job for batch processing.

Figure 85. Job Queue page of the Daemon utility for batch processing
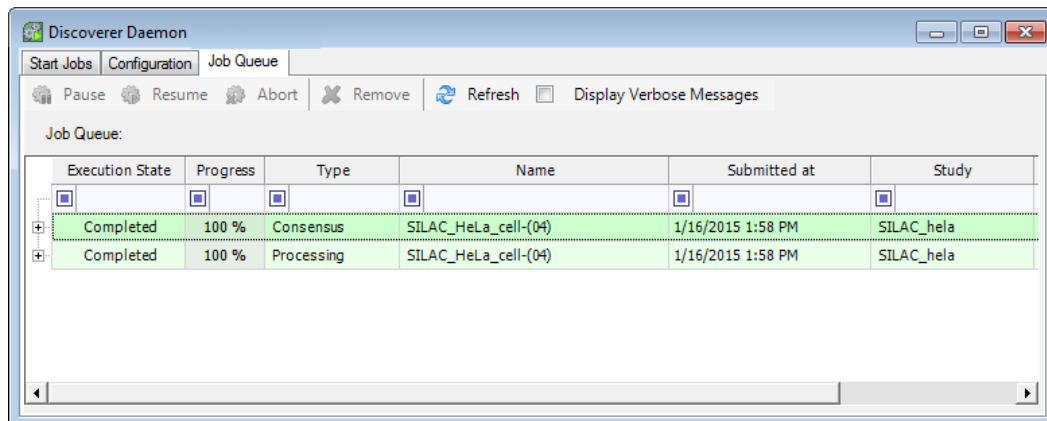
Logging in to a Remote Server

The searches started by the Proteome Discoverer application consume memory and can potentially cause the data-acquiring computer to crash and lose the sample in the mass spectrometer. To avoid this outcome, Thermo Fisher Scientific recommends that you connect the Proteome Discoverer Daemon utility to a remote computer running the Magellan server before data acquisition.

❖ **To log in to a remote server**

1. Start the Proteome Discoverer application on the remote machine.

2. To store the output files in a location other than the default folder, do the following:

   a. Choose **Administration > Configuration > Server Settings > Discoverer Daemon**.

      The PublicFiles folder is the default folder displayed in the Current File Directory box (Figure 86).

   b. In the New Directory box, browse to the storage location for the output files.

c. Click ✅ Apply .

The Daemon utility issues a message informing you that the Proteome Discoverer application will apply the change the next time that you start it.
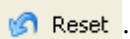
To return to the default directory, click 🔄 Reset .

Figure 86. Proteome Discoverer Daemon area of the Configuration view
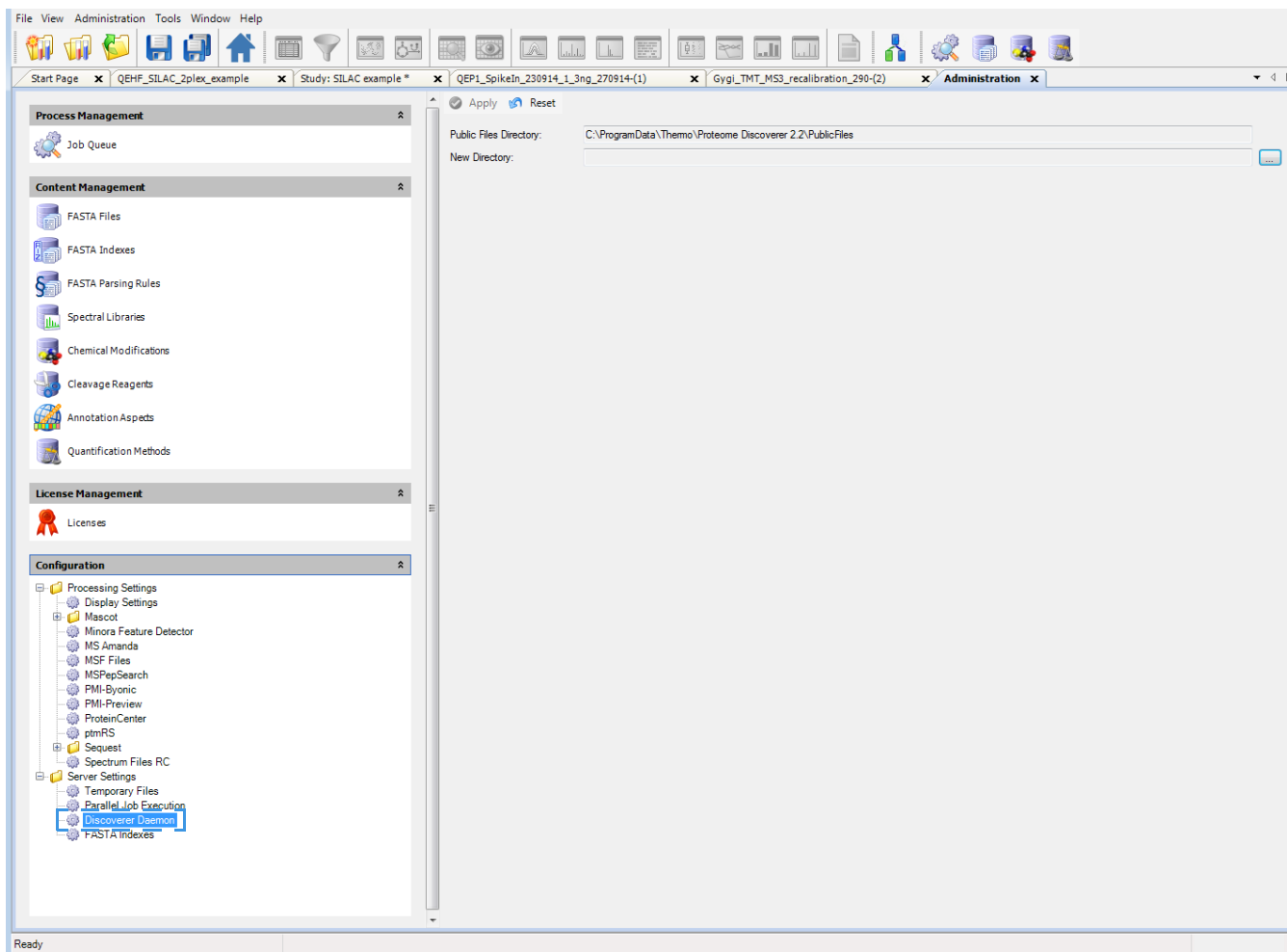
3. Start the Daemon utility on the local machine.

A message box informs you that the utility cannot connect to the server.

4. Click **OK** in the message box.

The utility opens with the Configuration page selected.

5. In the Host box, type the name of the remote computer.

6. In the User box, type the login name of the remote server.

7. Click **Apply**.

Running the Daemon Utility from the Xcalibur Data System

You can use the parameter file created in the Daemon utility to call the application from the Xcalibur data system. You can start the utility by adding a parameter file to the processing method specified in the Xcalibur injection sequence.

To run the utility from the Xcalibur data system, follow these topics:

- Before You Start

- Running the Daemon Utility from a Parameter File

- Creating a Processing Method That Calls the Daemon Utility

- Batch Processing with a Processing Method That Calls the Daemon Utility

- Batch Processing with Multiple Processing Methods

- Processing MudPIT Samples by Using a Processing Method

Before You Start

Before you run the Daemon utility from the Xcalibur data system, follow these procedures to set up the interface between the utility and the Xcalibur data system.

❖ **To prepare to run the Daemon utility in Xcalibur**

1. Before you start the Daemon utility, install the Proteome Discoverer application on a remote computer to decouple data processing from data acquisition.

   Thermo Fisher Scientific strongly recommends that you perform data analysis and data recording on two different computers to avoid disturbing the data acquisition by resource-consuming data processing.

2. Install the utility on the same computer where the Xcalibur data system is running.

3. Start the Proteome Discoverer application.

4. In the Proteome Discoverer application, prepare the workflow to be used by the Daemon utility (Figure 87). Save this workflow.

Figure 87. Basic workflow used for the samples
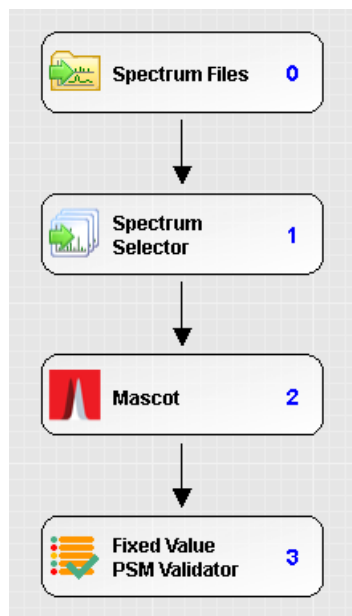
After you install the utility, the Proteome Discoverer application places the directory where it saves the raw data files and stores the results in the following folder:

Drive:\ProgramData\Thermo\Proteome Discoverer *<release_number>*\PublicFiles

This directory might be invisible to you because the *drive*:\Documents and ProgramData directory is hidden. To display hidden directories, choose **Tools > Folder Options > View > Hidden files and folders > Show hidden files and folders** in Windows Explorer.

5. (Optional) To change this directory for easier data access, open the Proteome Discoverer application, choose **Administration > Configuration**, click **Discoverer Daemon** beneath Server Settings in the Configuration area in the left pane, and change the directory in the New Directory box (see Figure 88).

The settings are applied after you restart the Proteome Discoverer application.

Figure 88. Change in the destination directory where results from the Daemon utility are stored
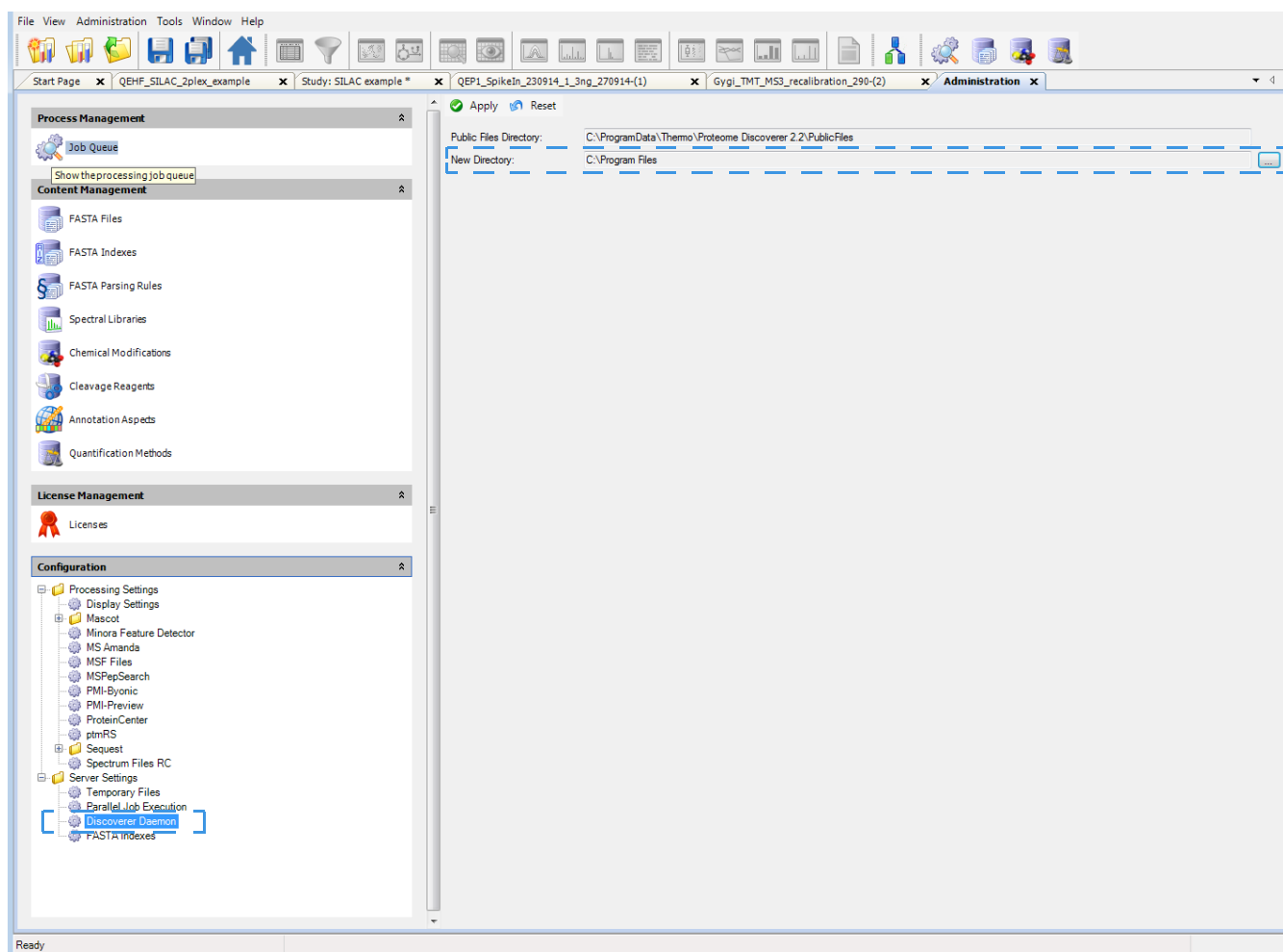
Running the Daemon Utility from a Parameter File

In the Daemon utility, you can create a parameter file that you can use to call the application from the Xcalibur data system. The application automatically translates the options that you set on the utility interface and in the workflow used for the search into text commands in the parameter file.

❖ **To create a parameter file that calls the Daemon utility**

1. Set up the search according to the instructions in "Running the Daemon Utility from the Window" on page 124. You do not need to have files loaded to create a parameter file.

2. Click the **Export Parameter File** tab (Figure 89) on the Start Jobs page.

Figure 89. Export Parameter File page
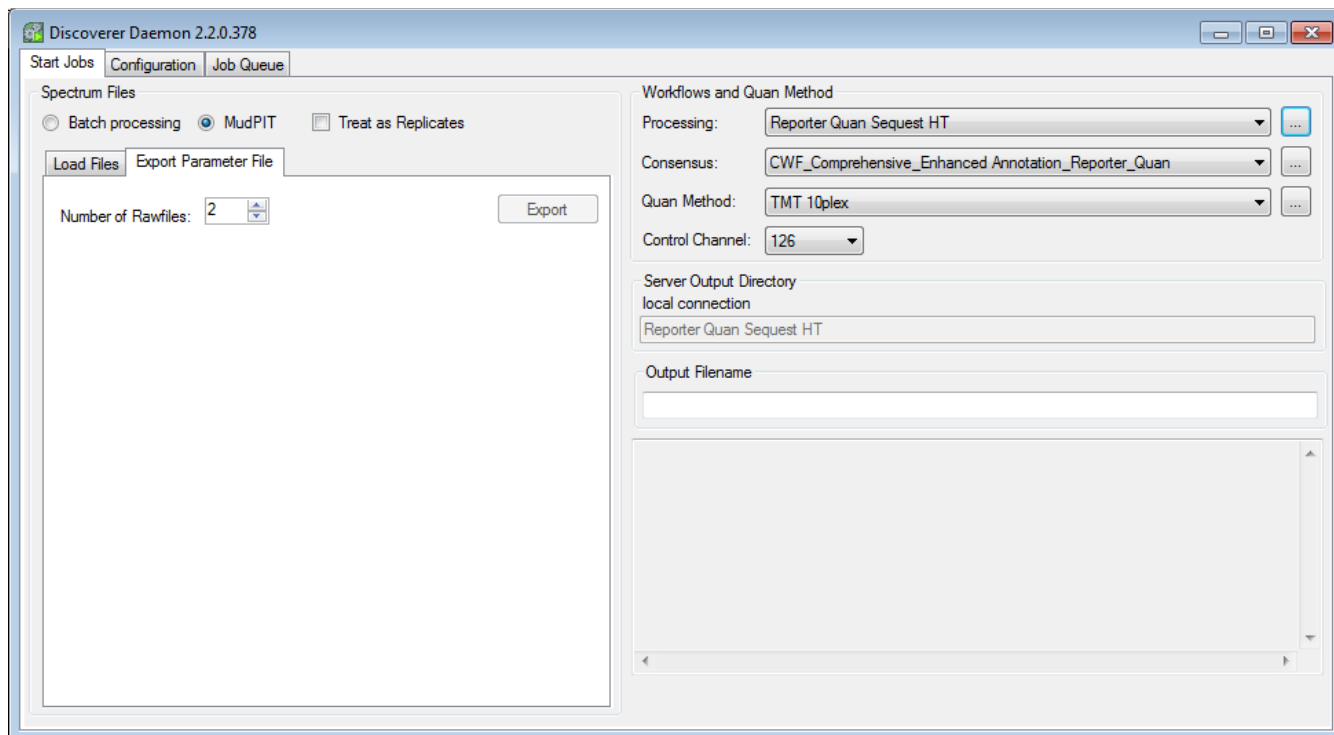
3. In the Number of Rawfiles box for a MudPIT search, select the number of files that will appear in the Xcalibur Sequence Setup dialog box.

This option is not available when you select the Batch Processing option.

4. Click **Export**.

The Save a Parameter File dialog box opens.

5. Specify the path and name of the parameter file, and click **Save**.

The Proteome Discoverer application writes the parameter file in XML format to the specified directory.

To call the utility through the parameter file, see "Running the Daemon Utility from the Xcalibur Data System" on page 130.

Creating a Processing Method That Calls the Daemon Utility

The following procedure describes how to create a processing method that calls the Daemon utility. It assumes that you have already created an appropriate processing method for your raw data files. Processing methods have a .pmd file extension.

❖ **To add a processing method that calls Proteome Discoverer Daemon to a processing method**

1. Choose **Start > All Programs > Thermo Xcalibur > Xcalibur** to start the Xcalibur data system.

2. In the Xcalibur Roadmap view, choose **GoTo > Processing Setup**.

   The Processing Setup window opens.

3. Open the processing method that you want to modify as follows:

   a. Choose **File > Open**.

   b. Browse to the location of the processing method file and select the file.

   c. Click **Open**.

4. Open the Programs view of the Processing Setup window as follows:

   a. Choose **View > View Bar**.

   The view bar appears on the left side of the dialog box.

   b. On the view bar, click the **Programs** icon, 🧿 .

   The Programs view of the Processing Setup window opens (Figure 90).

   Figure 90. Programs view with an empty table

	Enable	Std	QC	Unk	Other	Action	Program or Macro Name	Sync	Parameters
*	☐	Yes	Yes	Yes	Yes	Run Program		Yes	

5. If the Programs view has an empty table, right-click the table and choose **Insert Row**.

   A new row appears above the placeholder row (Figure 91). An asterisk to the left side of a table row defines it as a placeholder row.

Figure 91. Programs view with an unedited table row

6. In the added table row, specify the name and location of the parameter file as follows.

a. In the Enable column, select the check box.

b. In the Action list column, select **Run Program**.

c. Right-click the **Program or Macro Name** column and choose **Browse** (Figure 92).

Figure 92. Programs view with the shortcut menu displayed

d. Browse to the following executable, and click **Open**:

C:\Program Files\Thermo\Proteome Discoverer Daemon 2.1\System\Release\
System\Release\DiscovererDaemon.exe

If the following warning appears, "The file 'DiscovererDaemon' does not exist on this computer," click **OK**.

e. In the Parameters column, type the location of the parameter file containing the commands that will execute the utility:

-**p** *path_to_parameter_file\parameter_filename* **%R**

> **IMPORTANT** If the name of the parameter file contains a space, you must enclose the name in quotation marks, as in this example:
>
> -p "C:\Xcalibur\methods\batch processing.param" %R

7. In the Std, QC, Unk, Other, and Sync columns, accept the default settings or modify them according to your requirements. For information about setting the sample types to be sent, see "To specify the sample types to be sent to Proteome Discoverer Daemon."

To send all sample types to the Daemon utility, make sure that all of the sample type columns are set to **Yes** (see Figure 93).

Figure 93. Program table with a call to the Daemon utility

-p "C:\Daemon\data\daemon.param" %R

8. Click **OK** to save the changes to the processing method.

9. Choose **File > Save**.

❖ **To specify the sample types to be sent to Proteome Discoverer Daemon**

1. If the processing method that you want to modify is not open, open it and make sure that the parameter file and location are specified as described in To add a processing method that calls Proteome Discoverer Daemon to a processing method.

2. In the Std, QC, Unk, and Other columns, do the following:

   • To send a sample to the Daemon utility, make sure that "Yes" appears in the column for its sample type.

   • To avoid processing a sample with the utility, clear the column for its sample type.

   > **Tip** Use the Other column for the Blank sample type. For example, if you do not want to send blank samples to the utility for further processing, clear the Other column.

3. Save the processing method.

Batch Processing with a Processing Method That Calls the Daemon Utility

To inject samples and to acquire and process data files with the Xcalibur data system, you must create one or more instrument methods, one or more processing methods, and a sequence that defines the sample injection set.

For information about creating an instrument method for your LC/MS system, refer to the Help for the LC devices and the Help for the mass spectrometer. For information about creating processing methods and sequences, refer to the Xcalibur Help.

> **Tip** For a typical LC/MS experiment, an autosampler automates the sample injection process, and the position nomenclature depends on the autosampler tray type.
>
> For information about specifying the autosampler tray type and the position nomenclature for the specified tray type, refer to the Help for the autosampler.
>
> For some autosamplers, you can change the tray type from the Sequence Setup view by choosing Change > Tray Name and then selecting a different tray type.

To start the utility from the Xcalibur data system version 2.10 or later, you must add a processing method that calls the application to the sequence.

❖ **To set up and run an injection sequence with a processing method that starts Proteome Discoverer Daemon**

1. From the Home Page window of the Xcalibur data system, do one of the following:

   • Click the **Sequence View** icon, [icon], in the toolbar.

   –or–

   • Click the **Sequence Setup** icon, [icon], in the Roadmap view.

   The Sequence Setup view opens with an empty sequence table. Refer to the Xcalibur – Sequence Setup view Help for information about filling out the sequence table.

2. In the Proc Meth column, select a processing method with a parameter file that calls the utility as follows:

   • Type the file location and name of the processing method.

   –or–

   • Double-click the column to open the Select Processing Method dialog box, where you can browse to and select the processing method.

   You can now start the sequence without first saving it, or you can save the sequence for later use.

3. In the sequence table, select the row or rows that you want to run.

4. Choose **Actions > Run Sequence** or click the **Run Sequence** icon, [icon].

   If you have changed the instrument configuration in Foundation platform after the previous sequence run, the Change Instruments In Use dialog box opens. Otherwise, the Run Sequence dialog box opens (Figure 94).

   For an LC/MS system, the autosampler (or device with an autosampler) is specified as the start instrument. When the autosampler makes an injection, it triggers the mass spectrometer to begin data acquisition.

Figure 94. Run Sequence dialog box
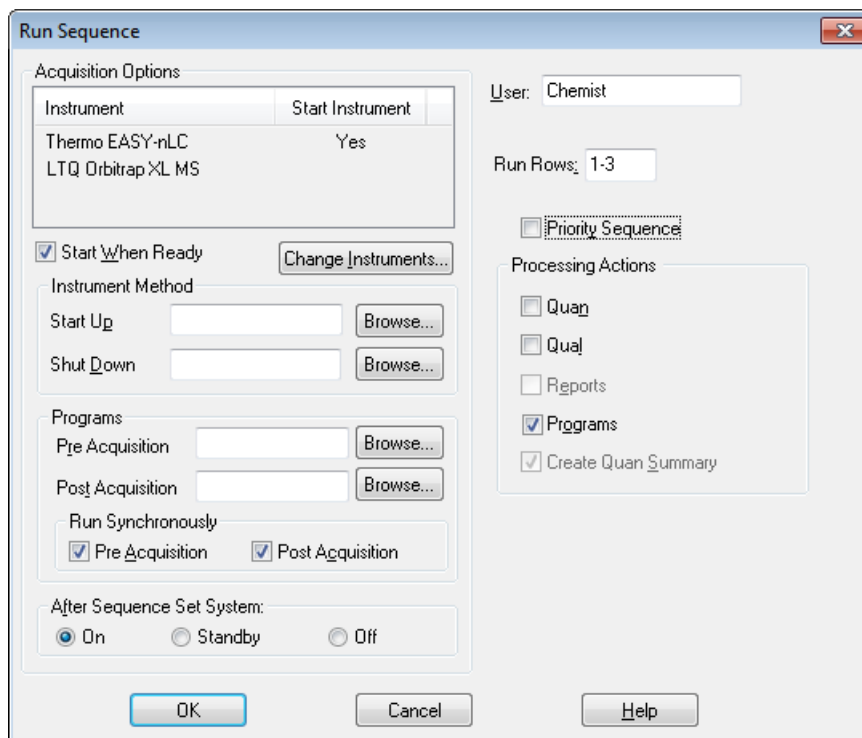
5. Select the **Programs** check box if it is not already selected.

    You must select this check box to start the utility.

6. Click **OK**.

    If you did not save the sequence, the File Summary Information dialog box opens.

7. Save the sequence as follows:

    a. In the File Summary Information box, click **OK**.

    b. In the File Name box, type a unique name for the sequence.

    c. In the Save In list, select the appropriate folder location for the sequence.

    d. Click **Save**.

    The Xcalibur data system adds the sequence to the acquisition queue.

For each sequence row, after the data system acquires a raw data file, it sends the processing method and the raw data file to the Proteome Discoverer application, which stores the raw data file and the .pdResult file in the server output directory specified in the Server Output Directory box of the Export Parameter File page of the Start Jobs page. All the search results of the batch processing are stored in the same directory. If the same directory name is used for the results of another batch process, the application appends the date and an incremental index number to the folder name.

Batch Processing with Multiple Processing Methods

In some cases, you might need to use more than one processing method in the sequence. For example, the sequest.pmd method runs the Daemon utility with a parameter file containing a simple Sequest HT workflow, and the export.pmd method runs the utility with an export workflow.

❖ **To use more than one processing method in a sequence**

1. In the Sequence Setup view, choose **File > New**.

   The New Sequence Template dialog box opens.

2. Enter the appropriate values in each of the boxes.

3. In the Bracket Type area, select the **None** option (Figure 95).

   You can now change the processing methods individually for each sample.

   Figure 95. New Sequence Template with the selection of None for the bracket type

Figure 96 shows a sequence using two different processing methods.

Figure 96. Sequence with two different processing methods

	Sample Type	File Name	Sample ID	Path	Inst Meth	Proc Meth	Position	Inj Vol
1	Unknown	BSA1	1	C:\Projects\Daemon	C:\Xcalibur\methods\Daemon	C:\Xcalibur\methods\batch	A1	10.00
2	Unknown	BSA2	2	C:\Projects\Daemon	C:\Xcalibur\methods\Daemon	C:\Xcalibur\methods\batch_export	A2	10.00
3	Unknown	BSA3	3	C:\Projects\Daemon	C:\Xcalibur\methods\Daemon	C:\Xcalibur\methods\batch	A3	10.00
*								0.00

4. Click **OK**.

In this example, the Xcalibur data system starts two different workflows (performing a Sequest HT search and exporting a raw data file) for the recorded raw data files in the Proteome Discoverer application (see Figure 97).

Figure 97. Two workflows in the job queue started by two different processing methods

Processing MudPIT Samples by Using a Processing Method

You can process MudPIT samples by using a processing method.

❖ **To process MudPIT samples**

1. Start the Daemon utility and export a parameter file for MudPIT processing. For information about exporting a parameter file, see "Running the Daemon Utility from a Parameter File" on page 132.

   Figure 98 shows how to configure the Export Parameter File page in the utility to export a parameter file. In this example, the parameter file is saved in C:\Xcalibur\methods.

   Figure 98. Selecting MudPIT processing in the Start Jobs page
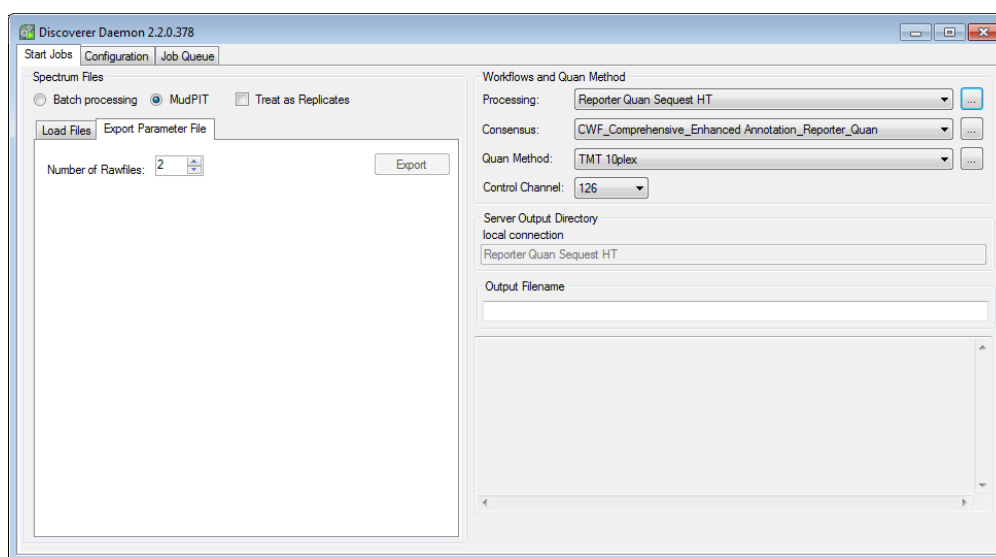
   This example features two MudPIT samples. Each one consists of two raw data files (for a total of four raw data files).

2. Define a processing method (see "Creating a Processing Method That Calls the Daemon Utility" on page 134) using the parameter file exported in step 1, and select the method as the processing method in the Proc Meth column (Figure 99).

Figure 99. Sequence used for MudPIT processing

	Sample Type	File Name	Sample ID	Path	Inst Meth	Proc Meth	Position	Inj Vol	Level
1	Unknown	BSA1	1	C:\Projects\Daemon	C:\Xcalibur\methods\Daemon	C:\Xcalibur\methods\mudPIT	1	10.00	
2	Unknown	BSA2	2	C:\Projects\Daemon	C:\Xcalibur\methods\Daemon	C:\Xcalibur\methods\mudPIT	2	10.00	
3	Unknown	BSA3	3	C:\Projects\Daemon	C:\Xcalibur\methods\Daemon	C:\Xcalibur\methods\mudPIT	3	10.00	
4	Unknown	BSA4	3	C:\Projects\Daemon	C:\Xcalibur\methods\Daemon	C:\Xcalibur\methods\mudPIT	3	10.00	
*								0.00	

3. Start processing the MudPIT samples in the Run Sequence dialog box (Figure 100).

Figure 100. Starting the processing of the MudPIT samples
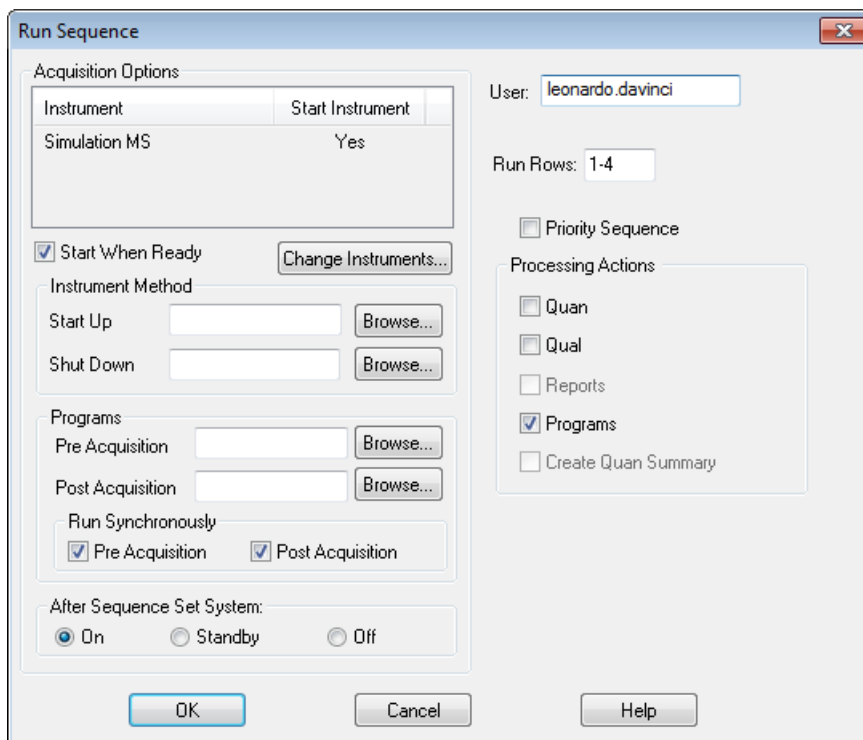
The utility processes the two samples as MudPIT (Figure 101).

Figure 101. Two MudPIT samples processed in the Daemon utility
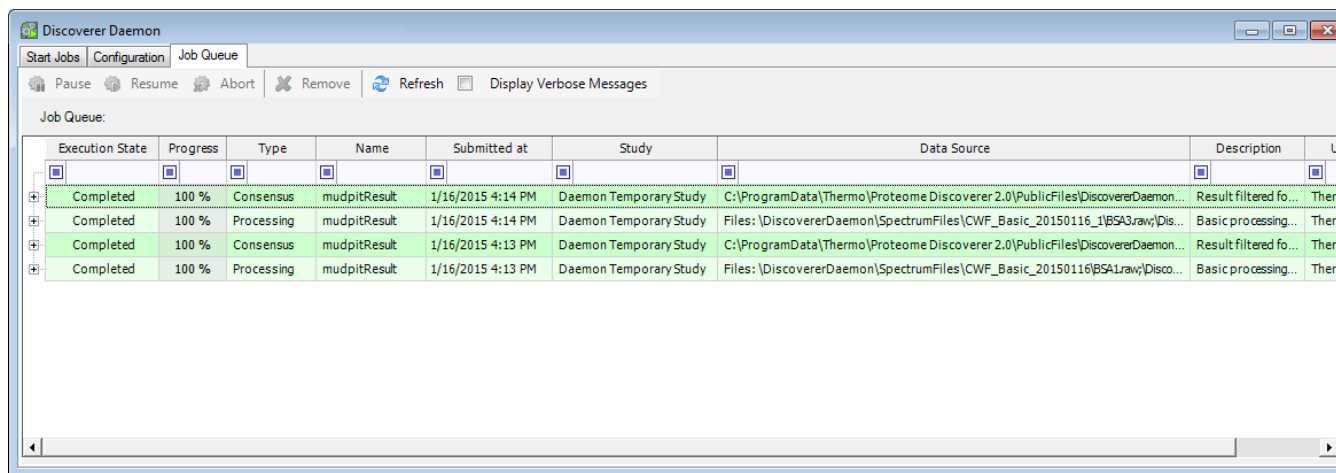
The Proteome Discoverer application saves the data in the two MudPIT samples in two directories, each one containing the raw data files of one MudPIT sample (in this example, two raw data files) (see Figure 102).

Figure 102. Two directories where the raw data files of each MudPIT group are saved
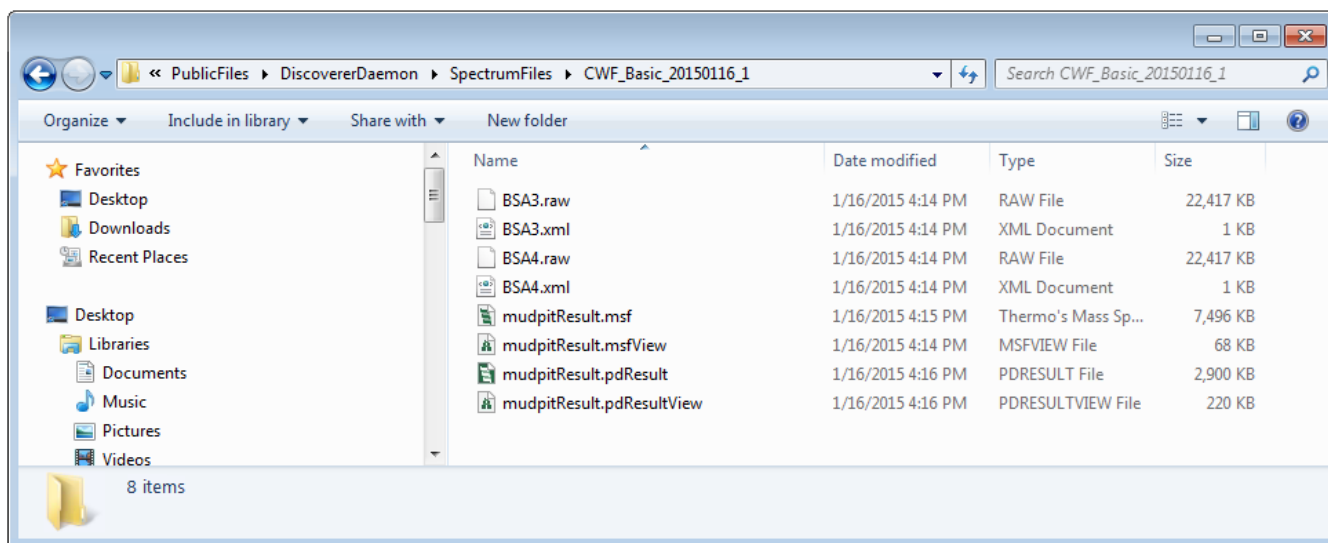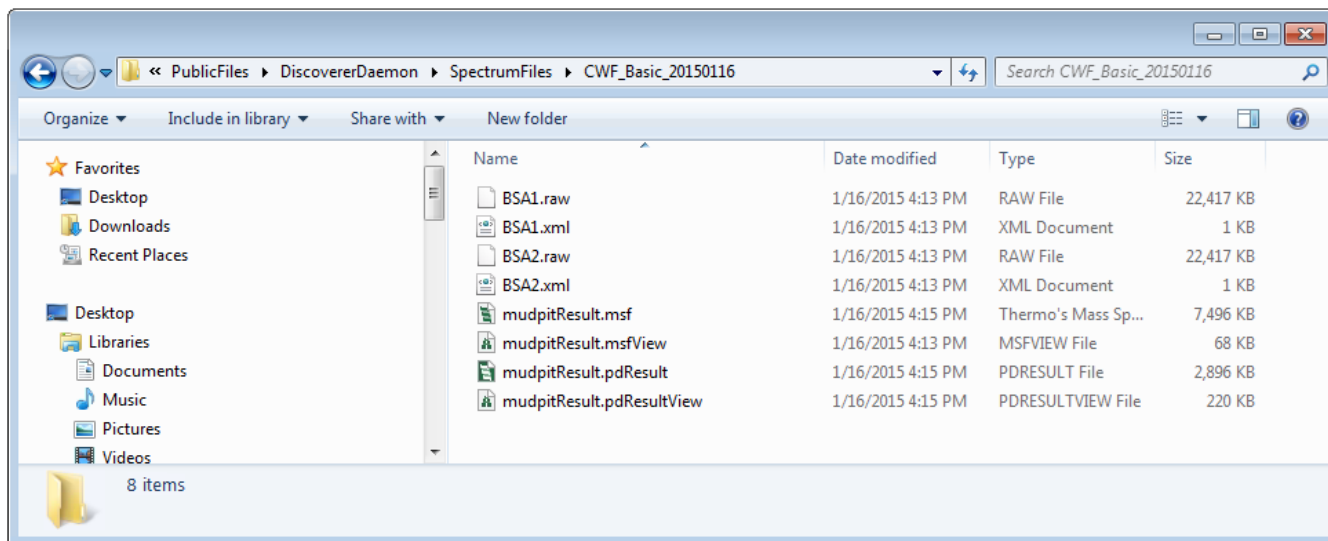
Running the Daemon Utility from the Command Line

You can run the Daemon utility from the command line or from the interface window.

❖ **To run the utility on the command line**

1. Open a command shell and use the cd command to move to **Program Files > Thermo > Proteome Discoverer Daemon 2.1 > System > Release**.

   When you run the Daemon utility from the command line, you must start it in the Proteome Discoverer Daemon 2.1 folder, not in the Proteome Discoverer folder.

2. Type **DiscovererDaemon** and any of the following options on the command line:

   DiscovererDaemon

   [-e *folderName FileCount workflows quanMethodFile ParameterAssignment*]

   [-c *folderName*]

   [-a *folderName SpectrumFile*]

   [-l *serverName userName*]

   [-r *outputFilename*]

   [-p *parameterFile rawFile*]

   [-f *folderName*]

For more information, see these topics:

- Syntax

- Examples

Syntax

The Daemon utility command-line syntax includes the following parameters:

- [-e *folderName filecount workflows quanMethodFile parameterAssignment*]

  Executes the workflow on the server using these specified parameters:

  – *folderName*: Specifies the location where the raw data files are stored. You can give it any name, for example, RawFiles or Fractions.

  – *FileCount*: Specifies the number of spectrum files that must be included before the workflow is executed. Intended to be used with MudPIT experiments and acquisition on several machines. If the workflow should be executed regardless of the number of files contained in the file collection, use ANY instead of a number.

  – *workflows*: Specifies a list of the template workflow files to run. You must have created these files in the Proteome Discoverer application by choosing Workflow Editor > *processing_workflow|consensus_workflow* > Save > *filename*.

  You must list the workflows in the order that they are to be executed. Separate the workflow names by semicolons. For example, to run a processing workflow, C:\test.pdProcessingWF, and a consensus workflow, test.pdConsensusWF, use this syntax:

  test.pdProcessingWF;test.pdConsensusWF

  – *quanMethodFile*: Specifies the name of the file that contains the quantification method.

  – *ParameterAssignment*: Specifies the name and value of a parameter in the format of *parameter=value*. Some examples follow.

This example sets the FASTA database for any node to equine.fasta:

`FastaDatabase=equine.fasta`

The next example sets the FASTA database for all Mascot nodes to equine.fasta:

`Mascot.FastaDatabase=equine.fasta`

The last example sets the FASTA database for Mascot nodes having 4 as the processing node number to equine.fasta. It is equivalent to [4].FastaDatabase=equine.fasta because the processing node numbers are unique.

`Mascot[4].FastaDatabase=equine.fasta`

- `[-c folderName]`

  – Remote server: Creates a user-named folder in the PublicFiles folder on the server where you store output files. The PublicFiles folder is the default folder in the Current File Directory box in the view displayed in the Proteome Discoverer application when you select Administration > Configuration > Server Settings > Discoverer Daemon. The -c option automatically appends the name to the date and, if the directory already exists, an incremental index number.

    You can create a folder only in the directory configured in the view opened by the Administration > Configuration > Server Settings > Discoverer Daemon command on the remote server. If you attempt to create a folder in a location other than the PublicFiles folder in the Current File Directory box, a message informs you that the Proteome Discoverer application will apply the change the next time that you start it.

    This option performs the same function as the -f *foldername* option, except that you can use the name of the folder more than once. When you use the name more than once, the utility appends the date and an incremental index number to the name.

  – Local server: Does nothing.

- `[-a folderName spectrumFile]`

  – Remote server: Uploads the spectrum file to the location that you specified on the configured server. *spectrumFile* is the name of the spectrum file.

  – Local server: Does nothing.

- `[-l serverName userName]`: Connects the Daemon utility to the specified local or remote host machine.

  – *serverName*: Specifies the name of the local or remote host.

  – *userName*: Specifies the name to log in.

- `[-r outputFilename]`: Specifies the name of the output file. You must use this option with the `-e` option, as in this example:

  `DiscovererDaemon.exe -e sfcid any mascot3.xml -r silac1noMT_AS4DE.msf`

- `[-p parameterFile rawFile]`: Processes the specified raw data file with all the parameters given in the parameter file, including the connection to the server.

- *parameterFile*: Specifies the name of the parameter file. In addition to the parameters, the parameter file also contains the workflows to run.

- *rawFile*: Specifies the name of the raw data file.

In the following example of the -p syntax, the utility processes the 9mix_LysC_monolith.raw file with the parameters given in the parameter file called C:\Xcalibur\methods\batchprocessing.param.

```
DiscovererDaemon.exe -p C:\Xcalibur\methods\batchprocessing.param
9mix_LysC_monolith.raw
```

- [-f *folderName*]: On a remote server, creates a user-named folder in the PublicFiles folder of the server where the local version of the raw data file and the result files are stored. If the directory already exists, the utility issues an error message, and the process returns with exit code -1 (standard exit code 0).

  If you attempt to create a file other than in the PublicFiles folder in the Current File Directory box, a message informs you that the Proteome Discoverer application will apply the change the next time that you start it.

  Performs the same function as the -c *foldername* option, except that you cannot use the name of the folder more than once.

  Local server: Does nothing.

Examples

The following are examples of the Daemon utility command-line syntax.

Example 1

This example constructs the spectrum file collection called Rawfiles, adds the TrypMyo.raw file to the collection, and executes the SequestEquine workflow using the raw data file in the Rawfiles directory:

```
DiscovererDaemon.exe -c Rawfiles -a Rawfiles
C:\Rawfiles\TrypMyo.raw -e Rawfiles ANY C:\Workflows\SequestEquine.xml
```

Example 2

In the following example, the Daemon utility evaluates several fractions in a single workflow:

```
DiscovererDaemon.exe -f Fractions

DiscovererDaemon.exe -a Fractions C:\rawfiles\fraction1.raw

DiscovererDaemon.exe -a Fractions C:\rawfiles\fraction2.raw

DiscovererDaemon.exe -a Fractions C:\rawfiles\fractionN.raw

DiscovererDaemon.exe -e Fractions ANY
pathToFile.pdProcessingWF;pathToOtherFile.pdConsensusWF
```

Example 3

The next example demonstrates that you can start several workflows with one invocation of the utility.

```
DiscovererDaemon.exe

-f RawFile

-a RawFile C:\Rawfiles\TrypMyo.raw

-e RawFile ANY pathToFile.pdProcessingWF;pathToOtherFile.pdConsensusWF

-a RawFile C:\Rawfiles\BSADigest.raw

-e RawFile ANY pathToFile.pdProcessingWF;pathToOtherFile.pdConsensusWF
```

Example 4

The following example runs the Daemon utility on a remote host called protlab2, uploads the iTRA_BSA_3ITMS2_3HCD.raw spectrum file to the server, and executes the workflows in C:\Workflows:

```
DiscovererDaemon.exe -l protlab2 leo_davinci -c sfcid -a sfcid
iTRA_BSA_3ITMS2_3HCD.raw -e sfcid any C:\Workflows\MascotEcoli\
pathToFile.pdProcessingWF;pathToOtherFile.pdConsensusWF
```

Example 5

In this example, the following sequence of commands submits multiple raw data files for processing on a remote server. The 020110303 notation indicates the date that the Daemon command was issued.

```
DiscovererDaemon.exe -c AllTrypMyo

DiscovererDaemon.exe -a AllTrypMyo_020110303
C:\DaemonTest\mudpit4\Tryp_Myo.raw

DiscovererDaemon.exe -a AllTrypMyo_020110303
C:\DaemonTest\mudpit4\Tryp_Myo_1.raw

DiscovererDaemon.exe -a AllTrypMyo_020110303
C:\DaemonTest\mudpit4\Tryp_Myo_2.raw

DiscovererDaemon.exe -e AllTrypMyo_020110303 3
C:\DaemonTest\mudpit4\pathToFile.pdProcessingWF;
pathToOtherFile.pdConsensusWF
```

Example 6

In this example, the next sequence of commands submits multiple raw data files for processing on a local server:

```
DiscovererDaemon.exe -a AllTrypMyo C:\DaemonTest\mudpit4\Tryp_Myo.raw
```

```
DiscovererDaemon.exe -a AllTrypMyo
C:\DaemonTest\mudpit4\Tryp_Myo_1.raw
```

```
DiscovererDaemon.exe -a AllTrypMyo
C:\DaemonTest\mudpit4\Tryp_Myo_2.raw
```

```
DiscovererDaemon.exe -e AllTrypMyo 3
C:\DaemonTest\mudpit4\pathToFile.pdProcessingWF;
pathToOtherFile.pdConsensusWF
```

Example 7

If you want to perform quantification, the commands resemble this example:

```
DiscovererDaemon.exe -c iTRAQInput
```

```
DiscovererDaemon.exe -a iTRAQInput C:\testData\iTraq\
myo_8plex_hcd_pqd.raw
```

```
DiscovererDaemon.exe -e iTRAQInput 1
C:\Users\Public\Documents\Thermo\Proteome_Discoverer_2.1\
Common_Templates\iTRAQ_8plex_Sequest_HT_equine.pdProcessingWF;
C:\Users\Public\Documents\Thermo\Proteome_Discoverer_2.1\
Common_Templates\WF_standard_automatic_validation_with_Quan
.pdConsensusWF;
C:\Users\Public\Documents\Thermo\Proteome_Discoverer_2.1\Common
Templates\iTRAQ_8plex.method
```

The Daemon utility appends a time stamp to each file when it processes the file on a remote server.

Searching for Data

This chapter describes the features that you can use when searching for and analyzing data in the Proteome Discoverer application.

Contents

- Using FASTA Databases
- Identifying Contaminants During Searches
- Displaying Species Names for Proteins and Peptide Groups
- Searching Spectrum Libraries
- Defining Chemical Modifications
- Using the Qual Browser Application
- Customizing Cleavage Reagents

Using FASTA Databases

You can use the FASTA database utilities to add, delete, and find protein references and sequences. You can also extract information from an existing FASTA file, place it into a new FASTA file, and compile it for availability in the Proteome Discoverer application.

For more information about FASTA databases, see "FASTA Reference" on page 635.

- Identifying Contaminants During Searches
- Downloading FASTA Files to Proteome Discoverer
- Deleting FASTA Files
- Adding Protein Sequences and References to a FASTA Database File
- Finding Protein Sequences and References
- Compiling a FASTA Database
- Excluding Individual Protein References and Sequences from a FASTA Database

- FASTA Database Utilities Dialog Box Parameters

- Managing FASTA Indexes

- Adding or Modifying FASTA Parsing Rules

- Identifying Contaminants During Searches

Displaying FASTA Files

You can list all the FASTA files that you have downloaded from other sources onto your hard drive and that you have registered.

❖ **To list the available FASTA files**

Choose **Administration > Maintain FASTA Files**, or click the **Maintain FASTA Files** icon, , either in the toolbar or on the Administration page.

The FASTA Files view shown in Figure 103 opens. It lists all the FASTA files that you have downloaded from other sources and registered. It displays the processed FASTA file properties, such as the file name, file size, and the number of proteins stored. The Proteome Discoverer application analyzes each protein entry to determine if the FASTA file meets the application requirements for use in a spectra search. It processes the FASTA file and makes it available for use.

Figure 103. FASTA Files view

Add button
Download button
Update button
Remove button
Cancel button
Export button
Refresh button
Check for Updates button
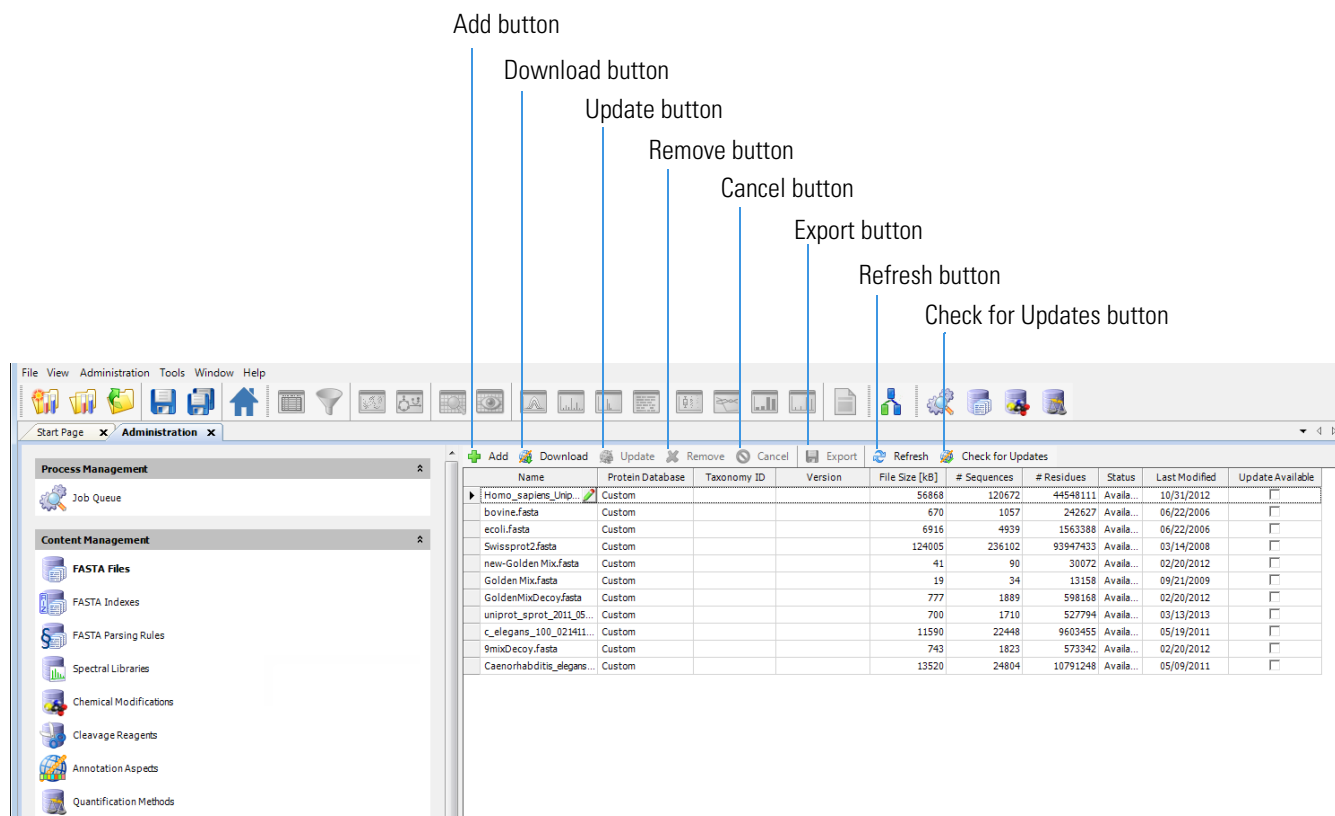
Downloading FASTA Files to Proteome Discoverer

You must add a FASTA file to the Proteome Discoverer application before you can conduct a search with the Sequest HT node. You must add a FASTA file to the server that is running Mascot before you can conduct a search with the Mascot search engine. You can download FASTA files from two sources:

- You can download a controlled protein database directly from ProteinCenter and add it to a FASTA file. These controlled databases offer you access to the latest protein information. The ProteinCenter database service provides extensive information about proteins, peptides, and genes. In addition, it gives you the ability to select proteins of a specified taxonomy to download and use as a FASTA file.

- You can add a FASTA file that you have downloaded from other sources onto your hard drive.

❖ **To download a FASTA file from ProteinCenter**

1. If you want to download a FASTA file to a location other than the default URL shown when you choose Administration > Configuration > ProteinCenter > ProteinCenter URL, see "Configuring the ProteinCenter Web Server" on page 264.

2. Choose **Administration > Maintain FASTA Files**, or click the **Maintain FASTA Files** icon, 📠 .

    The Administration page opens with the FASTA Files view (see Figure 103).

3. Click ![Check for Updates] Check for Updates .

    The Proteome Discoverer application updates the available databases in the job queue.

4. Click **FASTA Files** under Content Management in the Configuration view to return to the FASTA Files view.

    The Download button on the FASTA Files view becomes available.

5. Click ![Download] Download .

    The Download from ProteinCenter dialog box opens (see Figure 104).

    Figure 104. Download from ProteinCenter dialog box
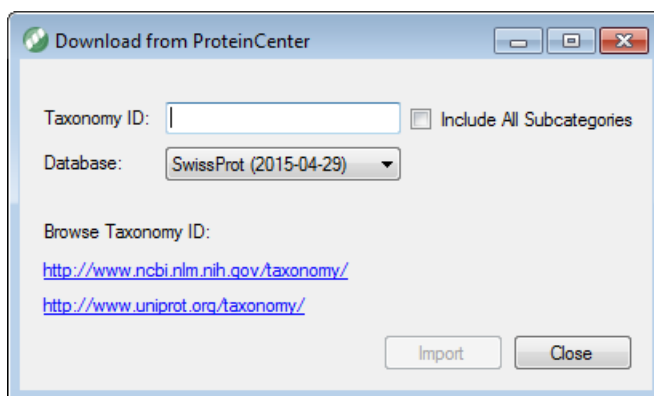
6. In the Taxonomy ID box, do the following:

    a. Type the taxonomy identification number of the appropriate organism-specific sequence database in ProteinCenter.

        The taxonomy identification number is a unique number identifying a biological species, a special subspecies, or a bacterial strain. You can find the organism's taxonomy identification number on the UniProt™ website (http://www.uniprot.org/taxonomy) or at the NCBI.

        i.   In the box to the right of the **Taxonomy** menu on the UniProt website, type the name of the organism that you are interested in, for example, **Baker's yeast**.

        ii. Click the **Search** icon, 🔍 .

        iii. (Optional) Under the Taxon heading, click the name of the subspecies that you are interested in, for example, **Saccharomyces cerevisiae**.

        The taxonomy identification number appears on the Taxon Identifier line, as shown in Figure 105.

Figure 105. Taxonomy identification on the UniProt website

Taxonomy identifier number
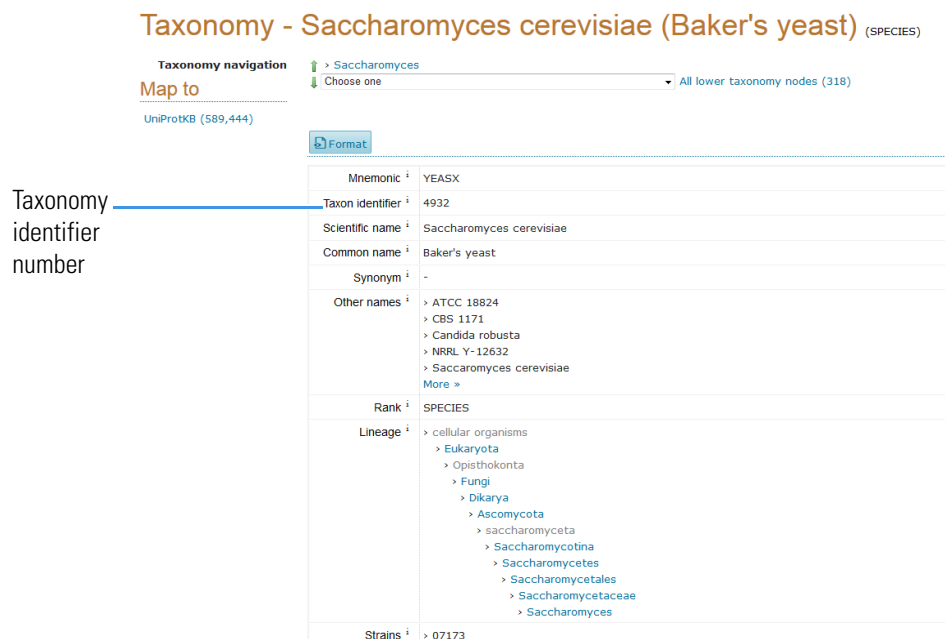
b. (Optional) To include data for a subspecies or subcategory of the selected species in the downloaded database, select the **Include All Subcategories** check box in the Download from ProteinCenter dialog box. For more information on this option, see the Help.

c. From the list in the Database box of the Download from ProteinCenter dialog box, select the name of the original source database to download the proteins from.

The default is SwissProt.

Figure 106 shows the completed Download from ProteinCenter dialog box.

Figure 106. Completed Download from ProteinCenter dialog box
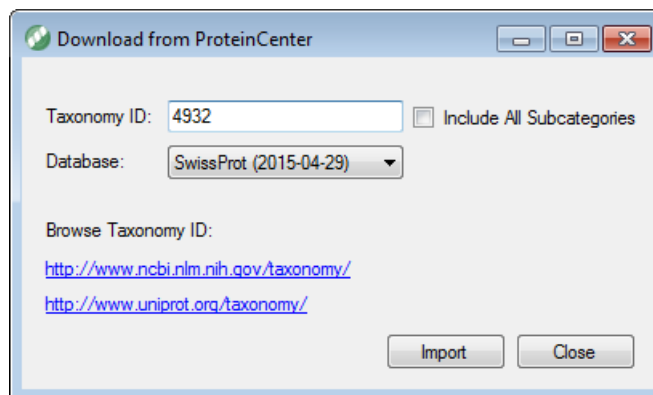
d. Click **Import**.

The application now displays the download as a job running in the job queue.

7. When the job queue displays "Completed" in the Execution State column, click **FASTA Files** under Content Management in the Configuration view to return to the FASTA Files view.

   The downloaded database might take several minutes to appear and opens in the FASTA Files view.

8. If you do not see the downloaded database after a few minutes, click [Refresh icon] Refresh.

   Figure 107 shows the Saccharomyces cerevisiae (4932) species database downloaded from the SwissProt database.

Figure 107. Database downloaded from ProteinCenter

Database downloaded
from ProteinCenter

If you select the SwissProt database, all proteins of the specified organism that are present in the SwissProt database are downloaded. The FASTA file is automatically given the scientific name of the organism and the database source, with the taxonomy identification in parentheses, for example, Saccharomyces cerevisiae (SwissProt TaxID =4932). It also gives the date the database was updated on the server, for example, (SwissProt - 2015-04-29).

If no databases appear in the list, follow the procedure in "To update a FASTA file from ProteinCenter."

❖ **To download a FASTA file from sources other than ProteinCenter**

1. Choose **Administration > Maintain FASTA Files**, or click the **Maintain FASTA Files** icon, [icon].

   The Administration page opens with the FASTA files view (Figure 103).

2. Click [Add icon] Add.

3. In the Open dialog box, browse to and select the FASTA file that you want to register, and then click **Open**.

   The FASTA file that you selected appears as a job in the job queue. To cancel the addition of this file, click [Abort icon] Abort.

   When you see "Completed" in the Execution State column, the database has finished downloading.

4. To add another FASTA file, wait until the Execution State column indicates that the addition of the FASTA file is completed, click **FASTA Files** in the left pane of the Administration page under Content Management, and then click **Add**.

The amount of time that it takes to import a FASTA file depends on the file size. When the application finishes importing a FASTA file, it displays "Available" in the Status column. The FASTA file is now available to use for a protein or peptide search with the Proteome Discoverer application.

❖ **To update a FASTA file from ProteinCenter**

> **Note** You can only use this procedure to update FASTA databases that you originally downloaded directly from ProteinCenter.

1. Choose **Administration > Maintain FASTA Files**, or click the **Maintain FASTA Files** icon, 📇 .

   The Administration page opens with the FASTA files view (Figure 103).

2. Select the database that you want to check for the availability of updated information. Click ▶ at the beginning of a database row to select the row.

3. Click 🔄 Check for Updates .

   The application now displays the search for updates as a job running in the job queue.

4. When the job queue displays Completed in the Execution State column, close the Proteome Discoverer application and reopen it.

5. Choose **Administration > Maintain FASTA Files**, or click the **Maintain FASTA Files** icon, 📇 .

   If an update is available for the selected database, a check mark appears in the Update Available column of the FASTA Files view (see Figure 108).

Figure 108. Check marks indicating update availability
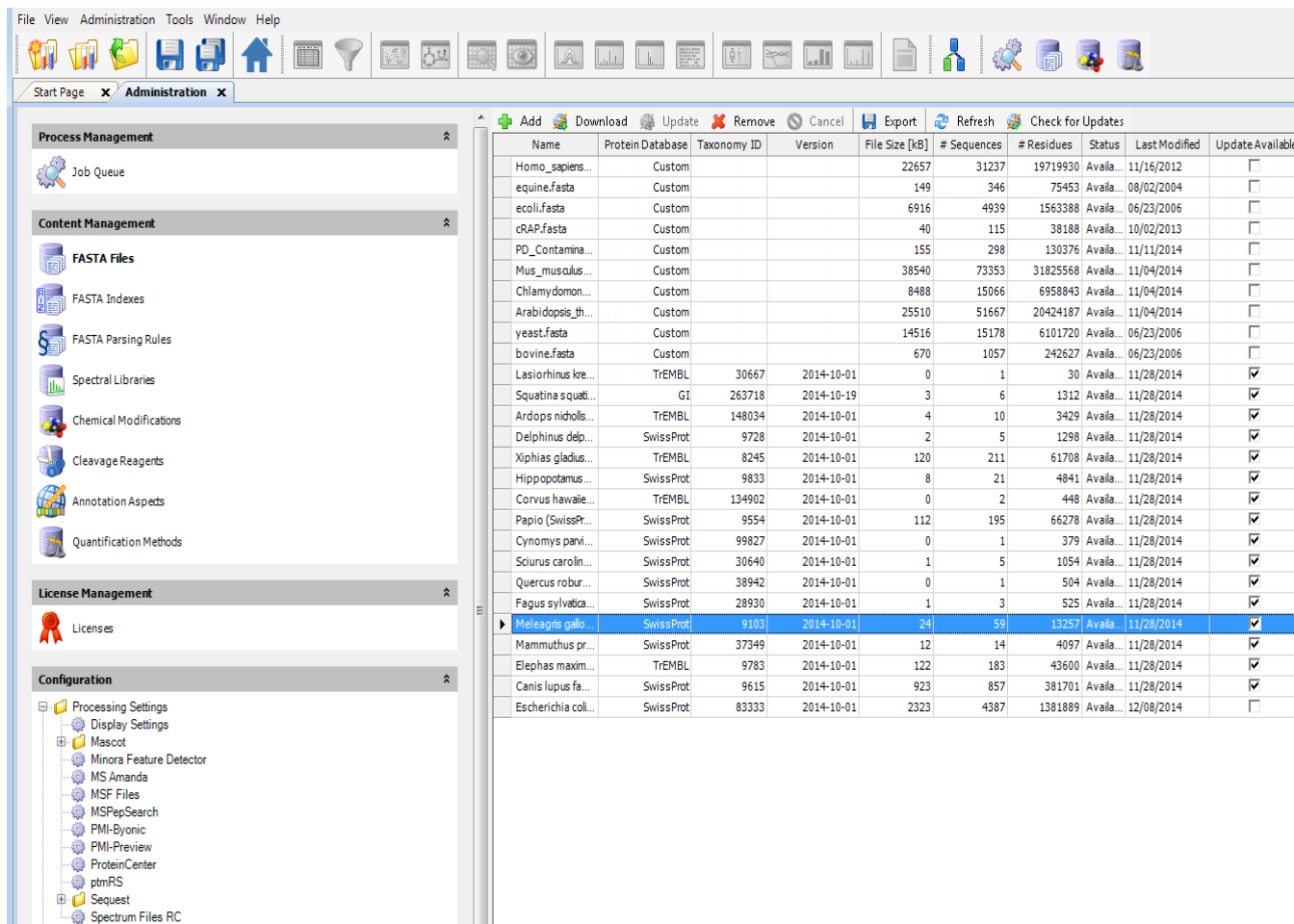
6. Select the database with the check mark in the Update Available column. Click ▶ at the beginning of a row to select the row.

   The Update button becomes available.

7. Click 🔄 Update .

   The application now displays the update as a job running in the job queue.

8. (Optional) When the job queue displays Completed in the Execution State column, click **FASTA Files** under Content Management to return to the FASTA Files view.

   The data on the ProteinCenter server is updated every two weeks. When the ProteinCenter server contains new data, the downloaded FASTA files might change.

   An update on the ProteinCenter server does not mean that new proteins for the FASTA files are automatically available. When you can access the new proteins depends on the timing of submissions to a specific database and that database's release cycle. For example, the UniProt databases are updated every month, so a new ProteinCenter version after two weeks does not necessarily contain new proteins.

Exporting FASTA Files

You can export FASTA files from the view when you choose Administration > Maintain FASTA Files; these include FASTA files added to the Proteome Discoverer application and files downloaded from ProteinCenter.

The exported FASTA file might be different from the original FASTA file. The application removes invalid and duplicate protein sequences during importation and cannot export them, although it retains protein title lines. The exported FASTA file lists the duplicate proteins with their titles lines in consecutive order.

❖ **To export a FASTA file**

1. Choose **Administration > Maintain FASTA Files**.

2. Select the FASTA file that you want to export.

3. Click 💾 Export .

4. In the Save As dialog box, browse to the directory where you want to save the FASTA file, and click **Save**.

   The job queue opens and displays a status of "Completed" in the Execution State column when the application finishes exporting the file.

Deleting FASTA Files

You can delete a FASTA file from the application.

❖ **To delete a FASTA file**

1. Choose **Administration > Maintain FASTA Files**, or click the **Maintain FASTA Files** icon, 🗄 .

   The Administration page opens with the FASTA files view (Figure 103).

2. Click ▶ at the beginning of a row to select it for deletion.

3. Click ✖ Remove .

4. In the Remove FASTA Databases dialog box, click **OK**.

The FASTA file that you selected appears as a job in the job queue. After you start the deletion of the file, you cannot cancel the deletion. You can remove the completed job from the job queue by clicking ✖ Remove , and then clicking OK in the Delete Jobs dialog box.

Adding Protein Sequences and References to a FASTA Database File

You can add protein sequences and protein references to a registered FASTA database file. The protein sequence refers to the sequence of amino acids that constitute the protein, and the protein reference refers to the name or reference of the protein.

❖ **To add a protein sequence and reference to a FASTA database file**

1. Choose **Tools > FASTA Database Utilities**.

2. In the FASTA Database Utilities dialog box, click the **Add Protein References** tab.

   The Add Protein References page of the dialog box opens.

3. Click the **Browse** button (...) next to the FASTA File box.

4. In the Save/Add to FASTA File dialog box, select the FASTA database that you want to add the protein sequence and reference to, and click **Save**.

5. In the Enter Description box of the FASTA Database Utilities dialog box, type a description of the protein sequence that you are adding.

6. In the Enter Protein Sequence box, type the protein sequence that you want to add to the FASTA database.

   Ensure that the Add Protein References page resembles Figure 109.

   Figure 109. Add Protein References page of the FASTA Database Utilities dialog box
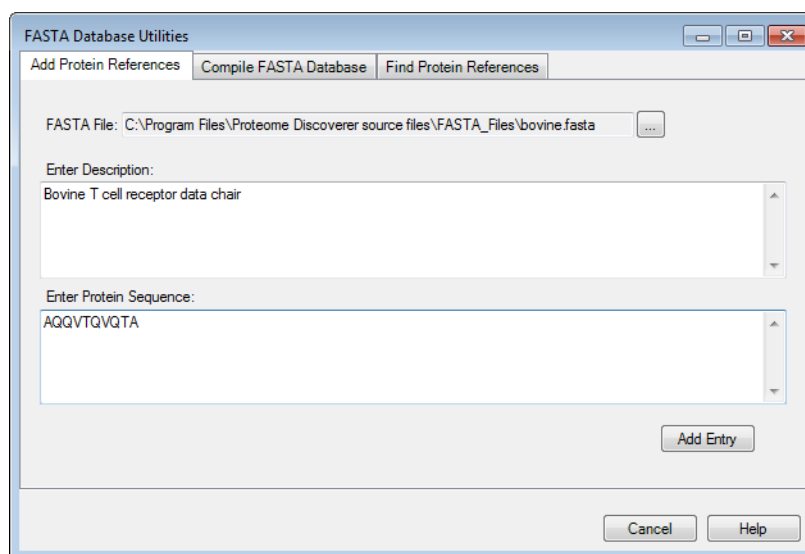
7. Click **Add Entry** to add the protein sequence.

For information on the Add Protein References page parameters, refer to the Help.

Finding Protein Sequences and References

You can find a protein sequence or reference in an existing FASTA database file.

- To find a protein sequence or reference
- To filter a protein reference search
- To refine a filtered protein reference search
- To delete conditions in filtered protein reference searches

❖ **To find a protein sequence or reference**

1. Choose **Tools > FASTA Database Utilities**.

2. In the FASTA Database Utilities dialog box, click the **Find Protein References** tab.

   The Find Protein References page opens (see Figure 110).

   Figure 110. Find Protein References page of the FASTA Database Utilities dialog box
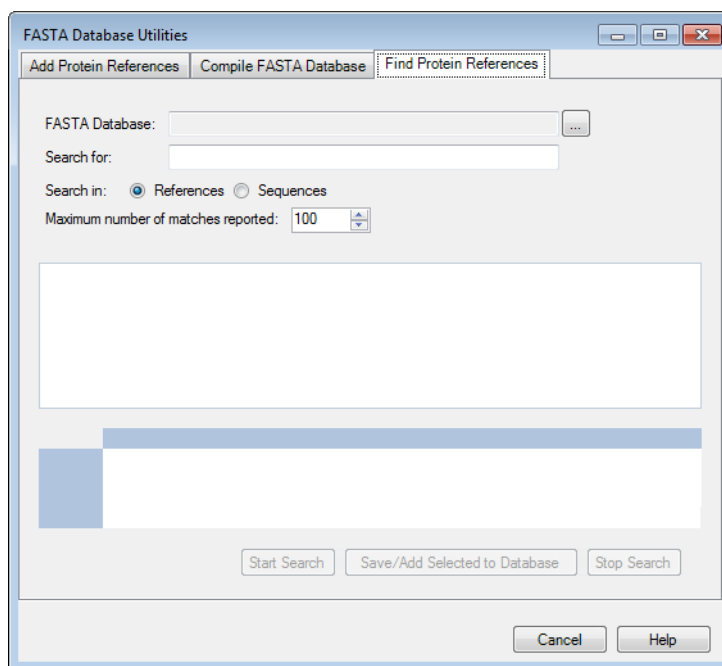
3. Click the **Browse** button (...) next to the FASTA Database box to locate the FASTA file of interest.

4. In the Please Select a FASTA Database dialog box, select the FASTA file, and click **Open**.

5. In the Search For box of the Find Protein References page, type an amino acid sequence or a protein reference search string.

6. In the Search In area, select the **References** option (the default) to search for a reference or the **Sequences** option to search for a sequence.

- References: Searches for the search string in the protein references.

- Sequences: Searches for the specified amino acid sequence within the protein sequences.

You can further refine the results by using filters either before or after you run the search. For instructions, see "To filter a protein reference search" on page 163.

7. In the Maximum Number of Matches Reported box, select the maximum number of references or sequences to report.

8. Click **Start Search**.

Results appear if the search parameters match the data (see Figure 111). Click a protein row to see the amino acid sequences that constitute that protein.

9. To suspend the search, click **Stop Search**.

Figure 111. Find Protein References page of the FASTA Database Utilities dialog box
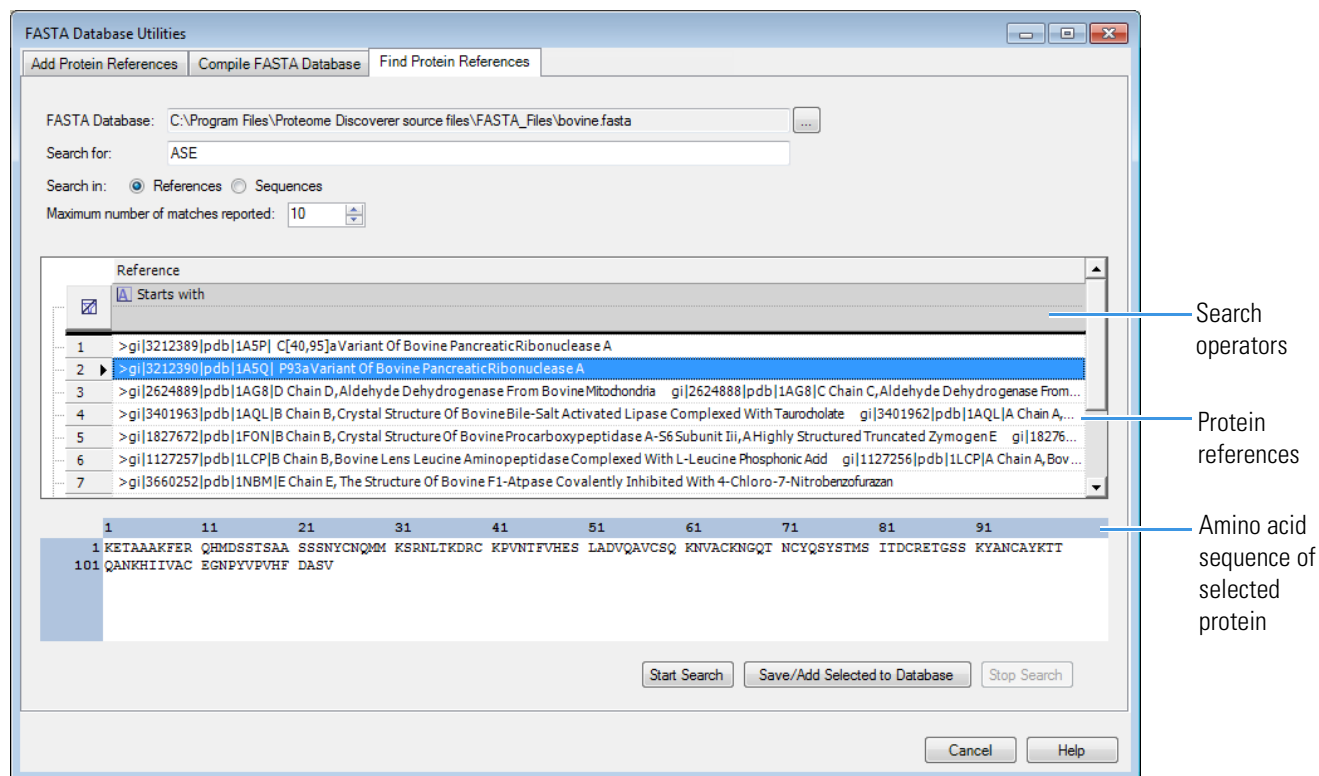
10. (Optional) To save a protein result row in another FASTA database, select the protein row, click **Save/Add Selected to Database**, select the database in the Save/Add to FASTA File dialog box, and click **Save**.

For information on the Find Protein References page parameters, refer to the Help.

❖ **To filter a protein reference search**

1. On the Find Protein References page of the FASTA Database Utilities dialog box, click the line below "Reference" in the middle of the page to access a list of operators that you can use to filter the references. (The default operator is "Starts with.") For a list of all operators, refer to the Help.

2. In the line below the operator that you selected, type the search string or condition that you want the operator to apply to.

The example in Figure 112 filters out those protein references that contain "fragment."

Figure 112. Filtering out protein references containing "fragment"
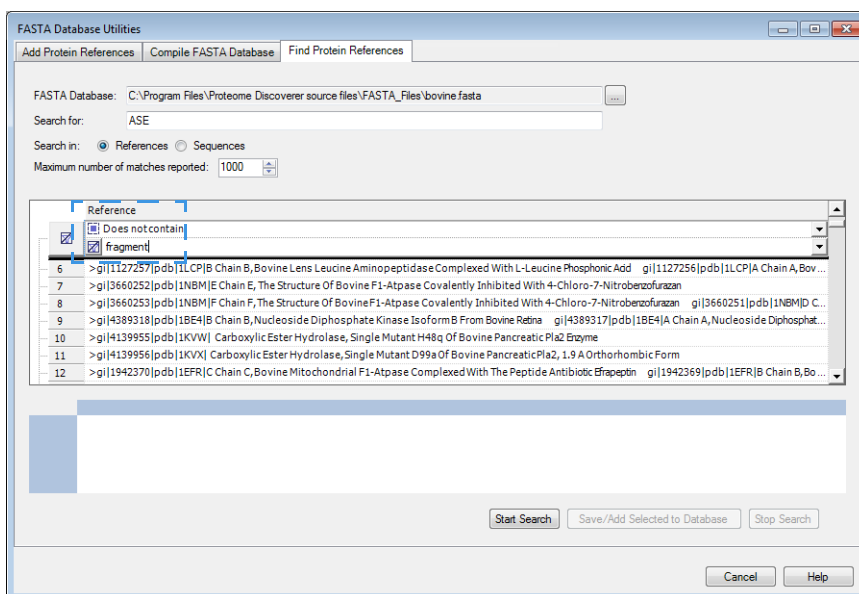
❖ **To refine a filtered protein reference search**

1. Select the **Custom** option from the list in the line below the operator.

To make the Custom option available, click the down arrow in the line below the operator (Figure 113) and select **Custom**.

Figure 113. Selecting the Custom option

Click this down arrow.
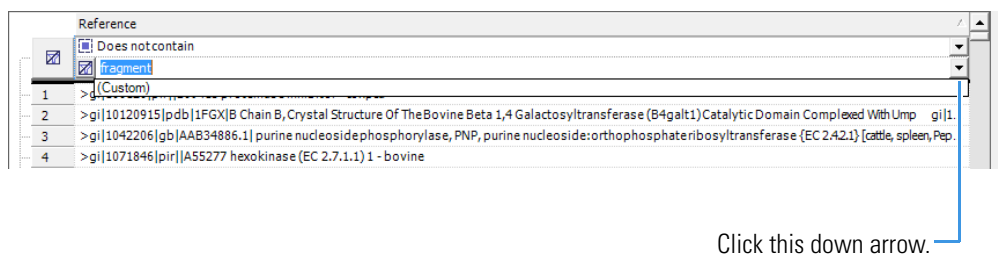
The Custom option opens the Custom Filter dialog box (Figure 114), so you can add multiple conditions.

Figure 114. Custom Filter dialog box
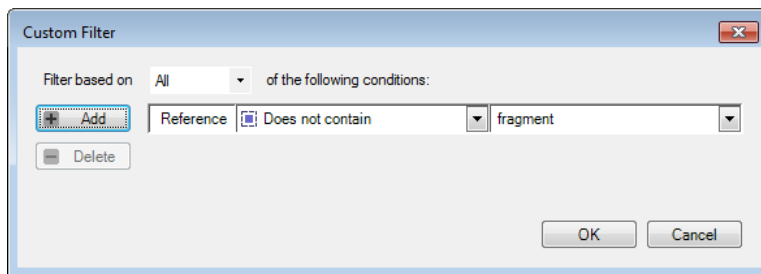
2. Click **Add**.

    A new line appears in the Operator (left) and Operand (right) lists.

3. Select an operator from the Operator list.

4. Type an operand on the line in the Operand column.

5. In the Filter Based On list, do one of the following:

    • Select the **All** option to indicate whether the search algorithm should search for protein references that meet both conditions.

    • Select the **Any** option to indicate whether the search algorithm should search for protein references that meet only one of the conditions.

    Figure 115 gives an example of a search for protein references that meet both of the conditions.

Figure 115. Custom filter with a second condition
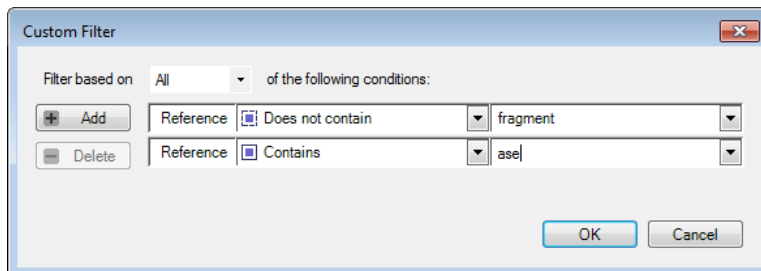
6. Click **OK**.

❖ **To delete conditions in filtered protein reference searches**

• To delete a condition in the Custom Filter dialog box, select the check box to the left of the appropriate condition in the Operator column, and click **Delete**.

• To delete the condition in the Reference area on the Find Proteins References page, click the **Clear Reference Filter Criteria** icon, , in the line below the operator.

• To delete all conditions in both the Custom Filter dialog box and the Reference area on the Find Proteins References page, click the **Clear All Filter Criteria** icon, , to the left of the filters.

Compiling a FASTA Database

You can extract information from an existing FASTA file and place it into a new FASTA file, replace an existing FASTA file, or append it to an existing FASTA file. Then you must compile the new or changed FASTA file to make it available in the Proteome Discoverer application.

❖ **To compile a FASTA database**

1. Choose **Tools > FASTA Database Utilities**.

2. In the FASTA Database Utilities dialog box, click the **Compile FASTA Database** tab.

   The Compile FASTA Database page opens.

3. In the Original box, browse to the FASTA file that you are taking the information from, or type its path and name.

4. In the Please Select a FASTA Database dialog box, click **Open**.

5. In the Target box, browse to the FASTA file that you are placing the extracted information into, or type its path and name.

6. In the Save/Add to FASTA File dialog box, select the file, verify that the file extension is .fasta, and click **Save**.

7. In the Target Database Options area, select one of the following options to indicate what you want to do with the extracted information:

   • (Default) **Create/Replace**: Creates a new FASTA file for storing the information or overwriting an existing FASTA file.

   • **Append**: Adds the extracted information to an existing FASTA file.

8. In the Search In area, specify whether the application should search for the search string in the protein references or sequences.

   • **References**: Searches for the search string in the protein references.

   • **Sequences**: Searches for the specified amino acid sequence within the protein sequences.

9. To disregard the case of the information to be extracted, select the **Ignore Case of Reference Strings** check box.

10. Specify the information to be extracted:

    a. Click ✛ above the Step 1: String(s) to Include box.

       A line where you can specify the first set of conditions appears in the box.

b. Click the first line in the Select Operator column, and select the operator to apply to the information to be extracted. You can select from the following:

- **Starts With**: Extracts information that begins with this string.

- **Does Not Start With**: Extracts information that does not begin with this string.

- **Ends With**: Extracts information that ends with this string.

- **Does Not End With**: Extracts information that does not end with this string.

- **Contains**: Extracts information that includes this string.

- **Does Not Contain**: Extracts information that does not includes this string.

c. Click the first line in the Condition column, and type the condition that the information must meet in order to be extracted.

d. Repeat step a through step c to add more sets of conditions for the information to be extracted.

e. To delete a set of conditions, in the Active column select the line that you want to delete and click ![X].

Ensure that the Compile FASTA Database page resembles Figure 116.

Figure 116. Compile FASTA Database page of the FASTA Database Utilities dialog box
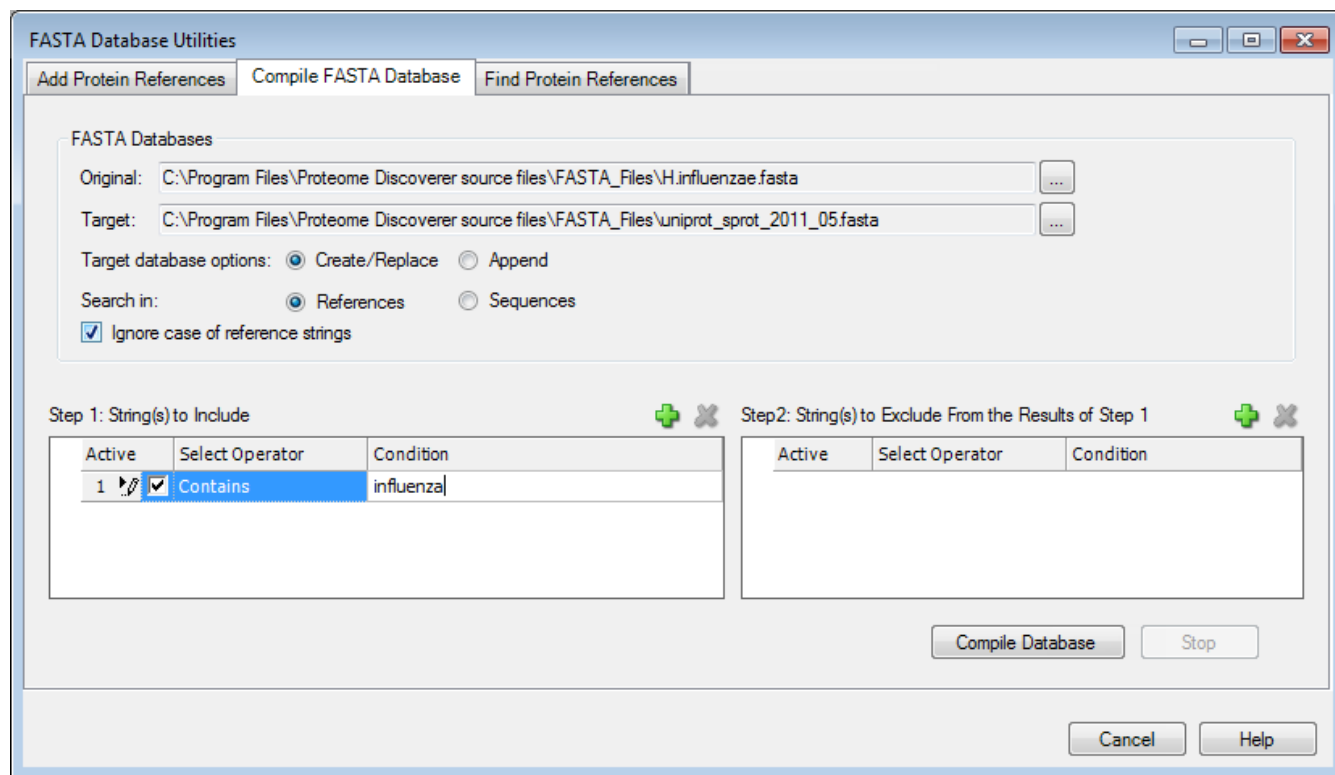
11. Click **Compile Database**.

    Click **Stop** to halt the compilation.

12. After the compilation, click **Start Search** on the Find Protein References page to view the results of the extraction. For an example, see Figure 117.

    You do not have to enter information into the Search For box.

Figure 117. Results of a search
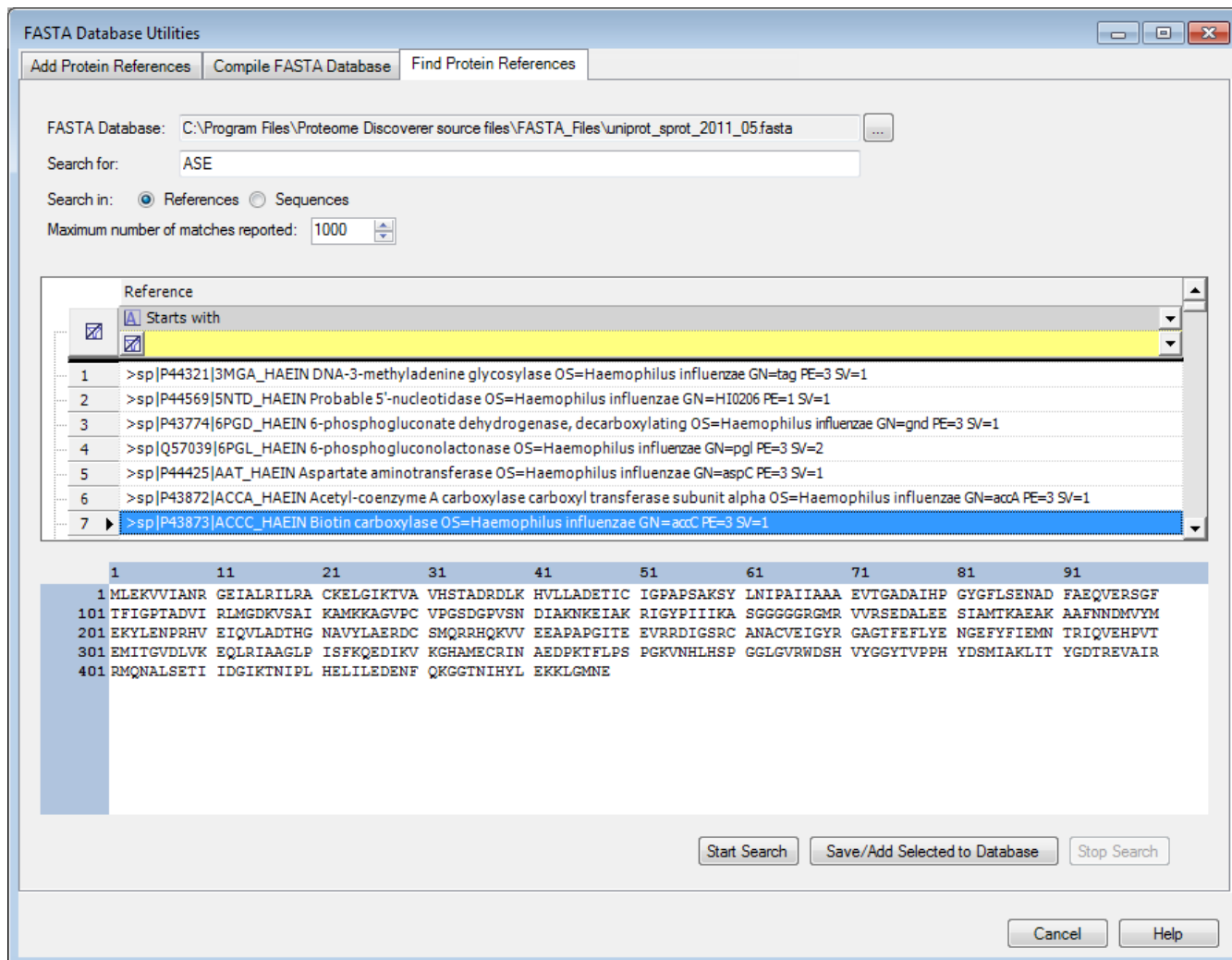
13. (Optional) To specify any information that you want to exclude from the extracted results, follow these steps:

    a. Click ![plus icon] above the Step 2: String(s) to Exclude From the Results of Step 1 box on the Compile FASTA Database page.

        A line where you can specify the first set of conditions now appears in the box.

b. Click the first line in the Select Operator column, and select the operator to apply to the information from the list. You can choose from the following:

- **Starts With**: Excludes information that begins with this string.

- **Does Not Start With**: Excludes information that does not begin with this string.

- **Ends With**: Excludes information that ends with this string.

- **Does Not End With**: Excludes information that does not end with this string.

- **Contains**: Excludes information that includes this string.

- **Does Not Contain**: Excludes information that does not include this string.

c. Click the first line in the Condition column, and type the condition that the information must meet in order to be excluded.

d. Repeat step a through step c to add more sets of conditions for the information that you want to exclude.

e. To delete a set of conditions, in the Active column select the line that you want to delete and click ![X icon].

14. Click **Compile Database**.

15. Click **Start Search** on the Find Protein References page to view the results of the extraction (see Figure 117).

You do not have to enter information into the Search For box.

For information on the Compile FASTA Database page parameters, refer to the Help.

Excluding Individual Protein References and Sequences from a FASTA Database

You can exclude individual entries from a FASTA file.

❖ **To exclude individual protein references and sequences from a FASTA file**

1. Choose **Tools > FASTA Database Utilities**.

2. In the FASTA Database Utilities dialog box, click the **Compile FASTA Database** tab.

3. In the Original box, browse to the FASTA database that contains the protein that you want to remove, or type its path and name. In the Please Select a FASTA Database dialog box, click **Open**.

4. In the Target box, browse to the output FASTA file or type its path and name. In the Save/Add to FASTA File dialog box, select the file, verify that the file extension is .fasta, and click **Save**.

5. Select the **Ignore Case of References Strings** check box.

6. Click ![plus icon] above the Step 1: String(s) to Include box.

A line to specify the first set of conditions now appears in the box.

7. Click the first line in the Select Operator column, and select **Contains** if it is not already selected. Leave the first line in the Condition column blank.

8. Click ➕ above the Step 2: String(s) to Exclude From the Results of Step 1 box.

   A line to specify the first set of conditions now appears in the box.

9. Click the first line in the Select Operator column, and select **Contains**.

10. In the first line of the Condition column, type the protein reference or sequence that you want to remove.

11. Click **Compile Database**.

   The compiling process creates the target FASTA file that excludes protein entries that match the condition.

FASTA Database Utilities Dialog Box Parameters

For parameter information on the pages of the FASTA Database Utilities dialog box, see these topics.

- Add Protein References Page

- Compile FASTA Database Page

- Find Protein References Page

Add Protein References Page

Table 6 describes the parameters on the Add Protein References page of the FASTA Database Utilities dialog box.

Table 6. Add Protein References page parameters

Parameters	Description
FASTA File	Specifies the name of the FASTA file to modify.
Enter Description	Describes the protein sequence that you are appending to the selected FASTA file. Include the name of the protein and the source tissue.
Enter Protein Sequence	Specifies the amino acid sequence of the protein that you are appending to the selected FASTA file.
Add Entry	Appends and saves the protein sequence and reference to the currently selected FASTA file.

Compile FASTA Database Page

Table 7 describes the parameters on the Compile FASTA Database page of the FASTA Database Utilities dialog box.

Table 7. Compile FASTA Database page parameters (Sheet 1 of 2)

Parameters	Description
Original	Specifies the path and name of the source FASTA database that you are taking the information from. It must be an existing FASTA file.
Target	Specifies the path and name of the FASTA file that you are placing or adding the extracted information to.
Target Database Options	Determines whether the extracted information is placed in a new FASTA file or is appended to an existing FASTA file. • Create/Replace (Default): Creates a new FASTA file where you will store the extracted information. • Append: Adds the extracted information to an existing FASTA file.
Search In	Determines whether to search for the search string in the protein references or sequences. • References: Searches for the search string in the protein references. • Sequences: Searches for the specified amino acid sequence within the protein sequences.
Ignore Case of Reference Strings	Disregards the case of the text in the Step 1: String(s) to Include box and the Step 2: String(s) to Exclude From the Results of Step 1 box.
Step 1: String(s) to Include:	Specifies the information to be extracted from the original file.
Active	Specifies the number of the condition set and allows you to select it.
Select Operator	Specifies the action to apply to the condition selected in the Condition column: • Starts With: Extracts information that begins with this string. • Does Not Start With: Extracts references that do not begin with this string. • Ends With: Extracts information that ends with this string. • Does Not End With: Extracts references that do not end with this string. • Contains: Extracts information that includes this string. • Does Not Contain: Extracts references that do not contain this string.
Condition	Specifies the condition that the information must meet in order to be extracted.
➕	Adds a line to the Step 1: String(s) to Include box.
✖	Deletes a line from the Step 1: String(s) to Include box.

Table 7. Compile FASTA Database page parameters (Sheet 2 of 2)

Parameters	Description
Step 2: String(s) to Exclude From the Results of Step 1	Specifies the information to be excluded from the extracted results.
Active	Specifies the number of the condition set and allows you to select it.
Select Operator	Specifies the action to apply to the condition selected in the Condition column: • Starts With: Excludes information that begins with this string. • Does Not Start With: Excludes information that does not begins with this string. • Ends With: Excludes information that ends with this string. • Does Not End With: Excludes information that does not end with this string. • Contains: Excludes information that includes this string. • Does Not Contain: Excludes information that does not include this string.
Condition	Specifies the condition that the information must meet in order to be excluded.
	Adds a line to the Step 2: String(s) to Exclude From the Results of Step 1 box.
	Deletes a line from the Step 2: String(s) to Exclude From the Results of Step 1 box.
Compile Database	Compiles and saves the changes to the FASTA file specified in the Target box.
Stop	Suspends the compilation.
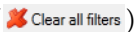
Find Protein References Page

Table 8 describes the parameters on the Find Protein References page of the FASTA Database Utilities dialog box.

Table 8. Find Protein References page parameters (Sheet 1 of 2)

Parameters	Description
FASTA Database	Specifies the name of the FASTA file that you want to search for sequences or references in.
Search For	Specifies the amino acid sequence to search for.

Table 8. Find Protein References page parameters (Sheet 2 of 2)

Parameters	Description
Search In	Determines whether to search for the search string in the protein references or sequences. • References: Searches for the search string in the protein references. • Sequences: Searches for the specified amino acid sequence within the protein sequences.
Maximum Number of Matches Reported	Specifies the maximum number of matching references or sequences to report.
Start Search	Begins the search for the reference or sequence.
Save/Add Selected to Database	Adds a selected protein result row to a specified database.
Stop Search	Suspends the active search.
Reference: List of Operators	Specifies the operator for the condition: • Equals • Does Not Equal • Less Than • Less Than or Equal To • Greater Than • Greater Than or Equal To • Like • Matches Regular Expression • Starts With • Contains • Ends With • Does Not Start With • Does Not Contain • Does Not End With • Does Not Match • Not Like
Reference: List of Condition Settings	Custom: Opens the Enter Filter Criteria for Reference dialog box so that you can define additional conditions.
Clear Reference Filter icon (⊞)	Deletes the condition under "Reference" on the Find Protein References page.
Clear All Filters icon (✖ Clear all filters)	Deletes all conditions in the Enter Filter Criteria for Reference dialog box and the condition under "Reference" on the Find Protein References page.

Managing FASTA Indexes

A FASTA index is a type of lookup table containing masses, theoretical peptide sequences, and associated proteins, which minimizes search time. The index lists all possible amino acid sequences that can be produced when an enzyme digests a protein or peptide. The peptide fragments are listed by molecular weight. The index stores information about every nominal mass, every peptide that has that mass, every protein that contains this peptide, and the location of its protein description in the FASTA file.

Rather than read all protein sequences from the FASTA file, digest them in silico with the specified enzyme, calculate the mass of each peptide, and compare it to the given precursor mass, the Proteome Discoverer application looks for the specific mass in the FASTA index and uses it to find the peptides that have this mass and the associated proteins that contain the peptides.

For full enzymatic searches, the application automatically creates FASTA indexes as they are needed. It does not automatically create FASTA indexes during semi-enzymatic or no-enzyme searches because these searches usually consume a large amount of space on a computer's hard disk. However, you can manually create FASTA indexes for these types of searches.

- Specifying the Location and Number of FASTA Indexes Stored

- Displaying the FASTA Indexes View

- Specifying the Columns to Display

- Manually Creating a FASTA Index

- Controlling Automatic FASTA Index Removal

- Deleting a FASTA Index

- Changing Number and Location of Stored FASTA Indexes

- Removing FASTA Indexes When a FASTA File Is Deleted

Specifying the Location and Number of FASTA Indexes Stored

If you do not want to store the FASTA indexes in the default directory (see Figure 118), you can specify an alternate directory in the FASTA Indexes configuration view. You can also change the maximum number of stored FASTA indexes.

❖ **To specify the location and number of the stored FASTA indexes**

1. Choose **Administration > Server Settings > FASTA Indexes**.

   The configuration view opens (see Figure 118).

Figure 118. FASTA Indexes configuration view
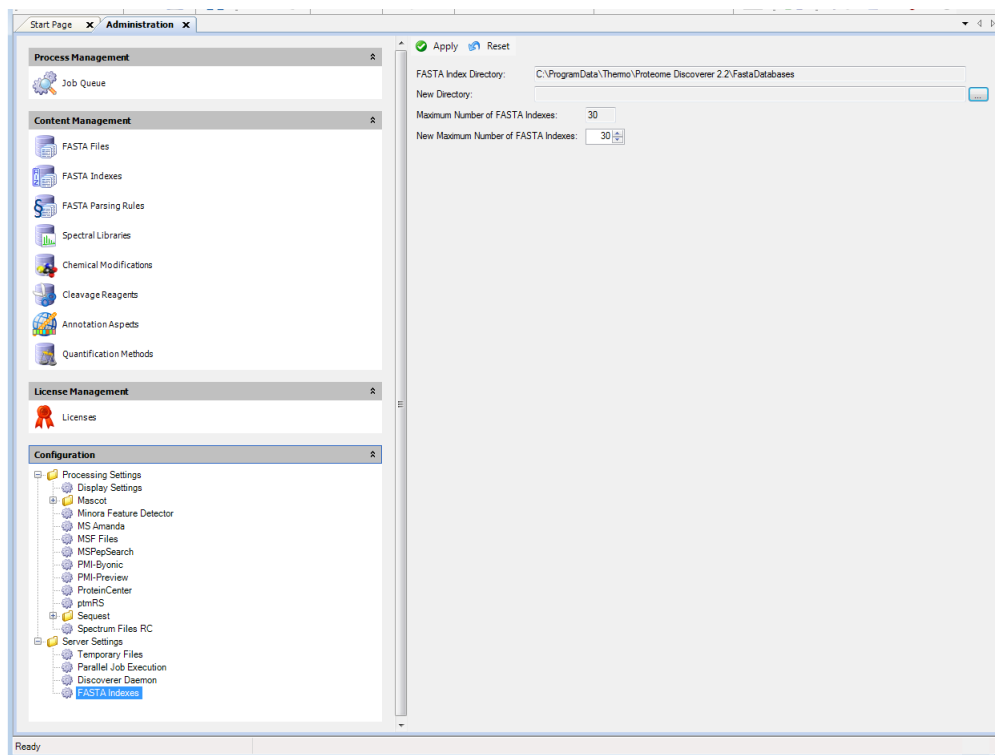
2. In the New Directory box, browse to the location of the folder to store the FASTA indexes in.

3. In the New Maximum Number of FASTA Indexes box, select the maximum number of FASTA indexes to store.

   If you generate more FASTA indexes than the number to store in the New Maximum Number of FASTA Indexes box, the application discards the difference from the oldest FASTA indexes the next time that you restart the application.
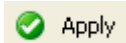
4. If you changed any settings, click **Apply**.

5. Click **OK** in the message box that opens (Figure 119).

Figure 119. Administration message box
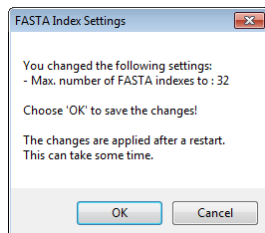
> **Note** Click **Reset** to return to the default values.

6. Restart your machine.

Displaying the FASTA Indexes View

You can access FASTA indexes through the FASTA Indexes view.

❖ **To display the FASTA Indexes view**

1. Choose **Administration > Maintain FASTA Indexes**.

   The FASTA Indexes view opens (Figure 120).

   Figure 120. FASTA Indexes view
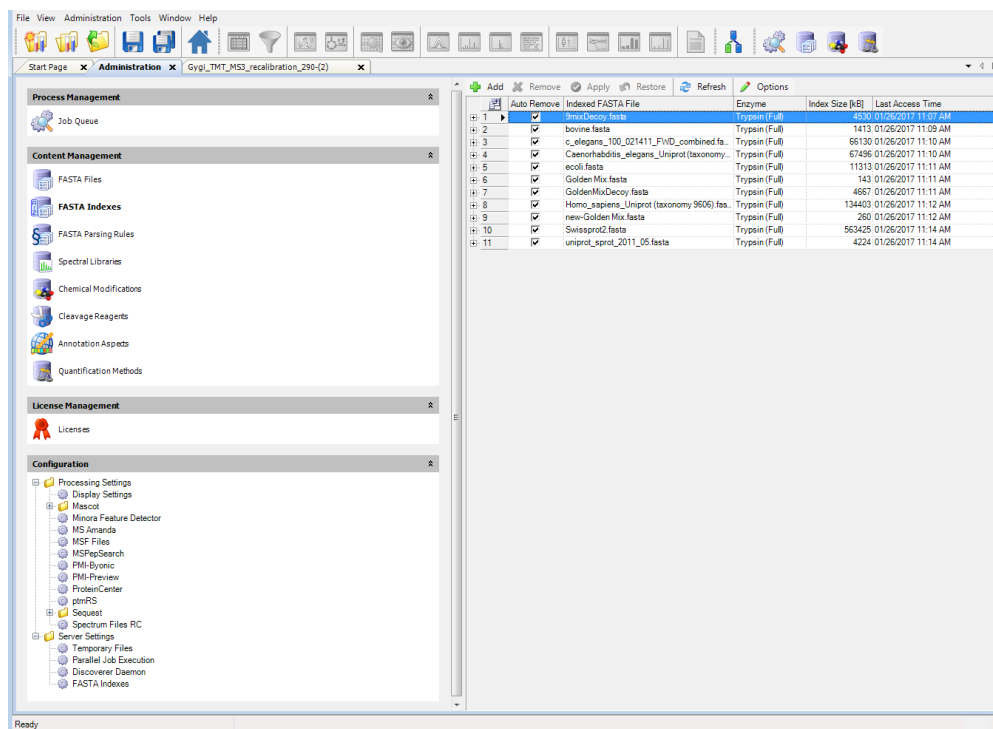
2. Click the plus (+) sign to the left of a database name to display the settings for that database, as shown for the swissprot2.fasta database in Figure 121.

Figure 121. Database settings in the FASTA Indexes view

Specifying the Columns to Display

Use the Field Chooser to specify the columns that you want to display.

❖ **To set the columns that you want to display**

1. Click the **Field Chooser** icon, ▣.

2. In the Field Chooser dialog box (Figure 122), select the check boxes corresponding to the columns that you want to display in the FASTA Indexes view.

Figure 122. Field Chooser dialog box in the FASTA Indexes view
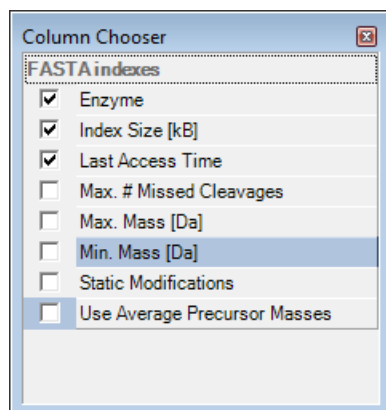
The application instantly makes the selected columns visible and the cleared columns invisible. For a description of these columns, see the Help.

Manually Creating a FASTA Index

The application automatically creates FASTA indexes for a full enzymatic digestion during a Sequest HT search, if an adequate FASTA index does not already exist. You can manually create a FASTA index for a semi-enzymatic or non-specific digestion (see "Controlling Automatic FASTA Index Removal" on page 180).
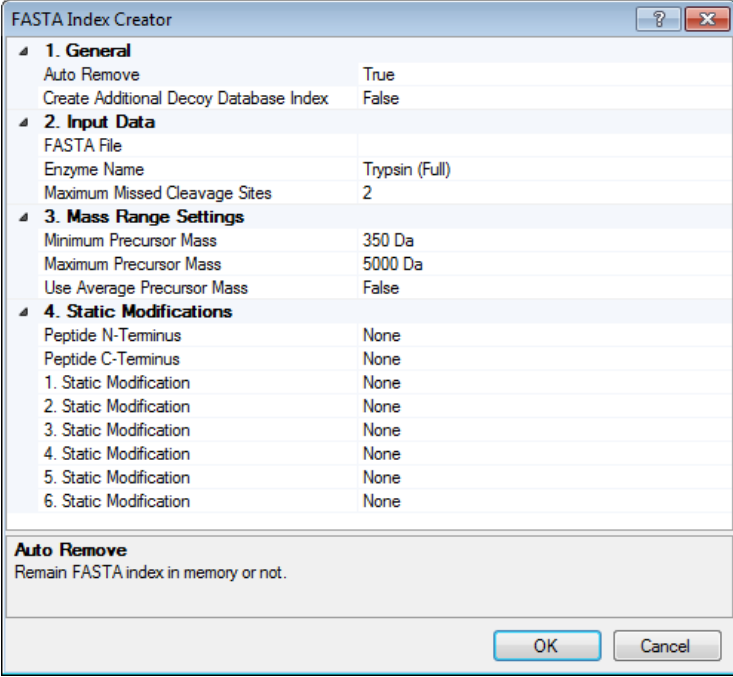
You can only create a specific FASTA index once.

❖ **To manually create a FASTA index**

1. Choose **Administration > Maintain FASTA Indexes**.

2. Click 🗟 Add .

   The FASTA Index Creator dialog box opens (Figure 123).

   Figure 123. FASTA Index Creator dialog box

3. In the General section, specify whether the available FASTA indexes will be removed from memory after the number of indexes reaches the specified maximum.

   • True (Default): Automatically removes the FASTA indexes from memory.

   • False: Keeps the FASTA indexes in memory.

   For information about how the application removes FASTA indexes after the maximum has been reached, see "Controlling Automatic FASTA Index Removal" on page 180. For instructions on specifying the maximum number of indexes, see "Changing Number and Location of Stored FASTA Indexes" on page 182.

4. In the Input Data section, specify the basic information that the application needs to create the index:

   • **FASTA File**: Select the FASTA database to be indexed from the list.

   • **Enzyme Name**: Select the enzyme used in the digestion from the list on the left (the enzymes on this list are set in the Cleavage Reagents window) and the type of digestion from the list on the right:

     – Full: Specifies a full enzymatic digestion.

     – Semi: Specifies semi-enzymatic digestion.

     – Unspecific: Specifies a non-specific digestion.

     – No Cleavages: Specifies that no cleavages occur.

   • **Maximum Missed Cleavage Sites**: Specifies the maximum number of internal cleavage sites per peptide fragment that is acceptable for an enzyme to miss when cleaving peptides during digestion. Normally the digestion time is too short to enable the enzyme to cleave the peptide at all positions, so you must specify the number of missed positions in one resulting peptide fragment where the enzyme could cleave but did not.

     Minimum: 0; maximum: 12; default: 2

5. In the Mass Range Settings section, set the limits of the mass range of the singly charged precursor ion to be processed:

   • **Minimum Precursor Mass**: Specifies the minimum mass of the precursor ion.

     Minimum: 0.0 Da; maximum: 10 000.0 Da; default: 350 Da

   • **Maximum Precursor Mass**: Specifies the maximum mass of the precursor ion.

     Minimum: 0.0 Da; maximum: 10 000.0 Da; default: 5000 Da

   • **Use Average Precursor Mass**: Determines whether the average mass is used to match the precursor ion.

     – True: Uses the average mass to match the precursor ion.

     – False (Default): Uses the monoisotopic mass to match the precursor ion, which is the mass of the most abundant isotope of the protein, peptide, or fragment ion.

6. In the Static Modifications area, specify the static modifications that occur on the amino acid:

   • **Peptide N-Terminus**: Select the static modification that occurs on the N terminus of the peptide.

   • **Peptide C-Terminus**: Select the static modification that occurs on the C terminus of the peptide.

- **Static Modification**: Select the static modification that occurs on the amino acid side chain.

7. Click **OK**.

   The application starts creating the FASTA index, and the job queue appears (Figure 124).

Figure 124. Creation of the FASTA index displayed in the job queue
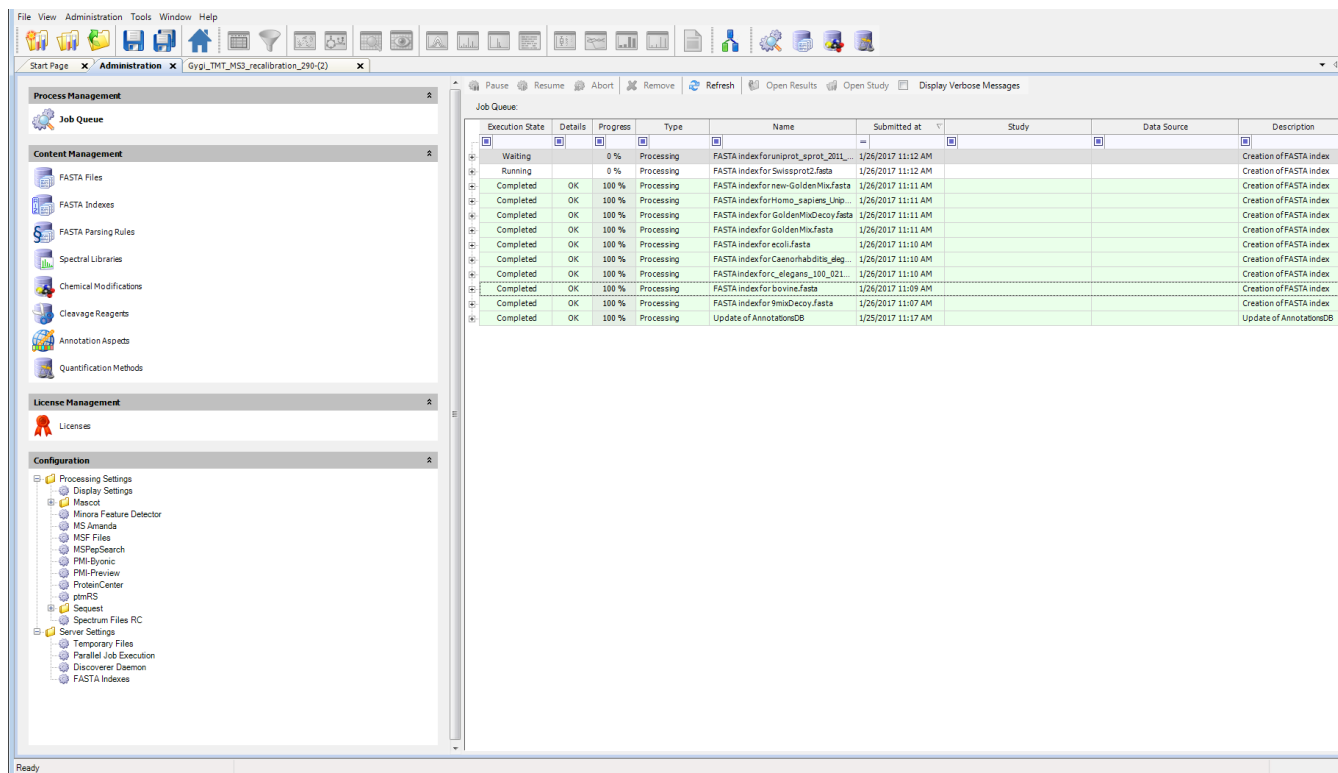
8. When the job finishes, choose **Administration > Maintain FASTA Indexes** to display the FASTA Indexes view.

9. Click [Refresh].

   The new FASTA index appears in the FASTA Indexes view on the Administration page (see Figure 125). The application creates an index for the specified FASTA file and the decoy version of the FASTA file.

Figure 125. FASTA Indexes view
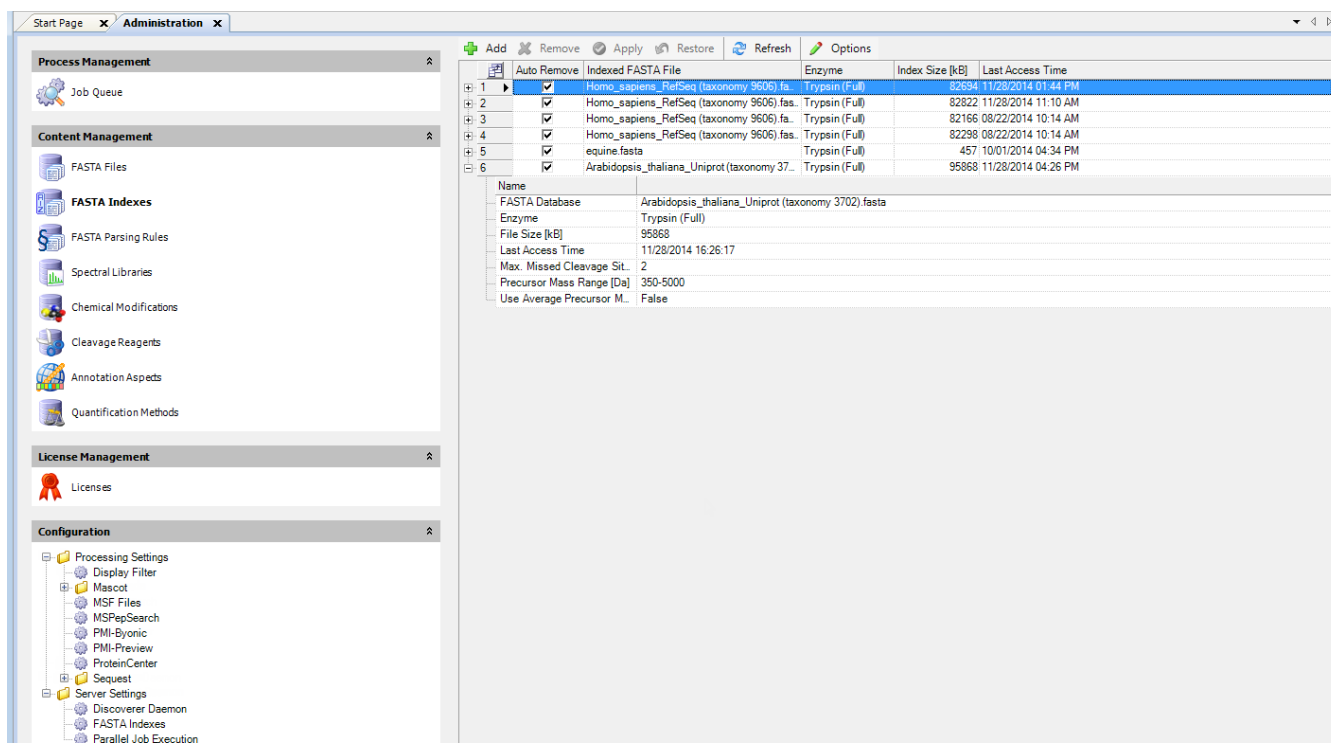
Controlling Automatic FASTA Index Removal

After the number of FASTA indexes reaches the specified maximum, the application automatically removes from memory the number of FASTA indexes over the maximum. It first removes the oldest indexes (that is, the ones with the earliest access time). However, you can mark specific FASTA indexes so that they will not be removed from memory, even after the maximum is reached.

❖ **To deactivate automatic FASTA index removal**

1. In the FASTA Indexes view on the Administration page, clear the **Auto Remove** check box.

   The Apply button now becomes available.

2. Click ⬛ Apply .

3. In the Remove FASTA indexes confirmation box, click **OK**.

❖ **To activate automatic FASTA index removal**

1. Select the **Auto Remove** check box.

2. Click ⬛ Apply .

3. In the Remove FASTA indexes confirmation box, click **OK**.

Deleting a FASTA Index

You can only delete FASTA indexes that have the Auto Remove check box selected.

❖ **To delete a FASTA index**

1. Be sure that the Auto Remove check box is selected for the index that you want to delete.

2. Select the index to delete by clicking the box next to the plus (+) sign.

   The box changes to an arrow (▶).

3. Click the arrow (▶).

4. Click ✖ Remove .

5. In the Remove FASTA indexes confirmation box, click **OK**.

The name of the deleted index is removed from the FASTA Indexes table and appears in a separate table called Deleted FASTA Indexes (see Figure 126). It no longer appears in the FASTA Indexes table. However, because the FASTA index might be used in some calculations, its removal from the application only takes place the next time that the server starts.
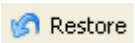
Figure 126. Deleted FASTA Indexes table

❖ **To restore a deleted FASTA index**

1. In the Deleted FASTA Indexes table, select the deleted index by clicking the arrow (▶).

2. Click 🔄 Restore .

3. In the Restore FASTA indexes confirmation box, click **OK**.

   The restored index appears in the FASTA Indexes table and disappears from the Deleted FASTA Indexes table.

Changing Number and Location of Stored FASTA Indexes

You can specify a new directory for storing the FASTA indexes and change the maximum number of FASTA indexes stored. The application counts all FASTA indexes, even the indexes that cannot be automatically removed with the Auto Remove option.

❖ **To change the number and location of stored FASTA indexes**

1. Click 🖊 Options .

   The FASTA Indexes Options dialog box opens (Figure 127).

   Figure 127. FASTA Indexes Options dialog box
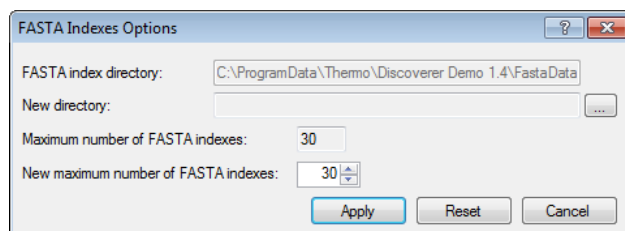
   > **Note** Another way to access these options is to choose Administration > Configuration and click FASTA Indexes in the Server Settings area.

   The FASTA Indexes Options dialog box contains two read-only parameters:

   • The FASTA Index Directory box displays the name of the current directory where the FASTA indexes are saved.

   • The Maximum Number of FASTA Indexes box displays the current maximum number of FASTA indexes allowed.

2. In the New Directory box, browse to the directory where you want to store the FASTA indexes.

   You can change the directory only if the server runs on the local machine.

3. In the New Maximum Number of FASTA Indexes box, type the new maximum number of FASTA indexes allowed.

4. Click **OK**.

5. In the FASTA index settings confirmation box, click **OK**.

After you confirm the changes, the application saves them, but the changes are only executed the next time that the server starts. You can undo the changes made since the last time that the server started and before the next time that the server starts, even though you clicked OK in the FASTA Indexes Options dialog box and closed it. For example, when you change the location of the directory in the FASTA Indexes Options dialog box, click OK, and close the dialog box, the server moves all FASTA indexes to the new target directory when the server restarts. But if you reinvoke the dialog box and click Reset before restarting the server, the changes that you made previously are deleted, and the directory reverts to its previous location.

❖ **To reset the changes made in a previous FASTA index session**

1. Click ✎ Options .

   The FASTA Indexes Options dialog box opens.

2. Click **Reset**.

Removing FASTA Indexes When a FASTA File Is Deleted

When you or the Proteome Discoverer application deletes a FASTA file, it removes the FASTA indexes belonging to the deleted FASTA file the next time that the server starts.

Adding or Modifying FASTA Parsing Rules

The protein descriptions and accessions shown on the Proteins page of the .pdResult file are taken from the title lines of the added FASTA file. The rules for extracting these values are defined as regular expressions (see http://en.wikipedia.org/wiki/Regular_expression). If you want to use unsupported FASTA files that might contain descriptions or accessions that are difficult to read—for example, FASTA files from the Saccharomyces Genome Database (SGD) or the Arabidopsis Information Resource (TAIR) web pages—you might need to add new parsing rules to the system or modify existing parsing rules.

If you want to add or modify a FASTA parsing rule, download a FASTA file from an appropriate website, add or change the parsing rule, test the parsing rule, correct the parsing rule until it meets your needs, and then apply it.

❖ **To add or modify FASTA parsing rules**

Choose **Administration > Maintain FASTA Parsing Rules**.

The FASTA Parsing Rules view opens (see Figure 128).

Figure 128. FASTA Parsing Rules view
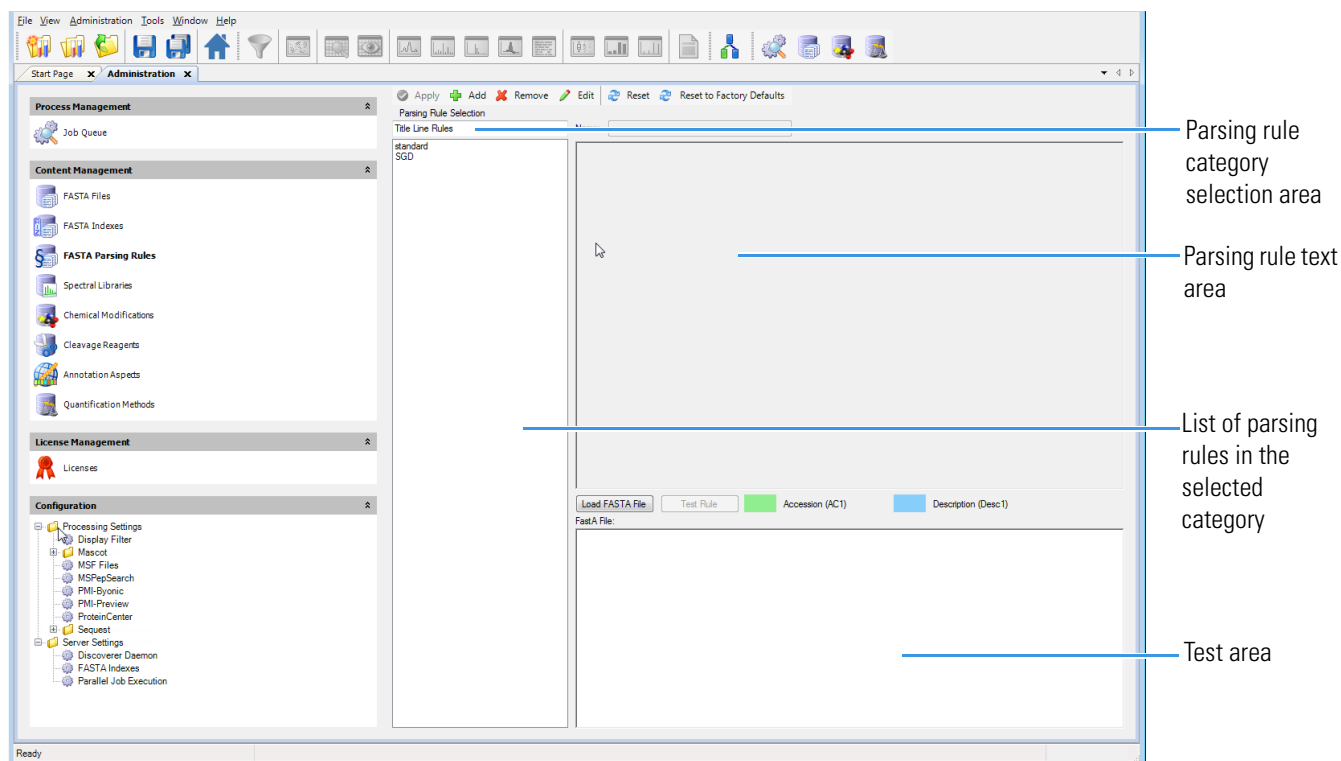
The FASTA Parsing Rules view includes these features:

- Parsing rule category selection area: Displays the four categories (Title Line Rules, Accession Rules, Taxonomy Rules, and Avoid Expression Rules) into which the application groups parsing rules.

- Parsing rule text area: Displays the regular expression of the parsing rule. Each line corresponds to a single regular expression. These expressions are tested as alternatives ("or" connected). Figure 129 shows regular expressions for SwissProt accessions in the parsing rules text area.

- List of parsing rules in the selected category: Displays all available parsing rules in the selected category. This list corresponds to the available values of the appropriate parameter of the MSF Files node. If you select a single entry, the FASTA Parsing Rules view displays the parsing rule in the parsing rule area on the right.

- Test area: Loads the title lines of a sample FASTA file to test the matching of the expression.

Figure 129. Regular expressions for SwissProt accessions in the parsing rule text area
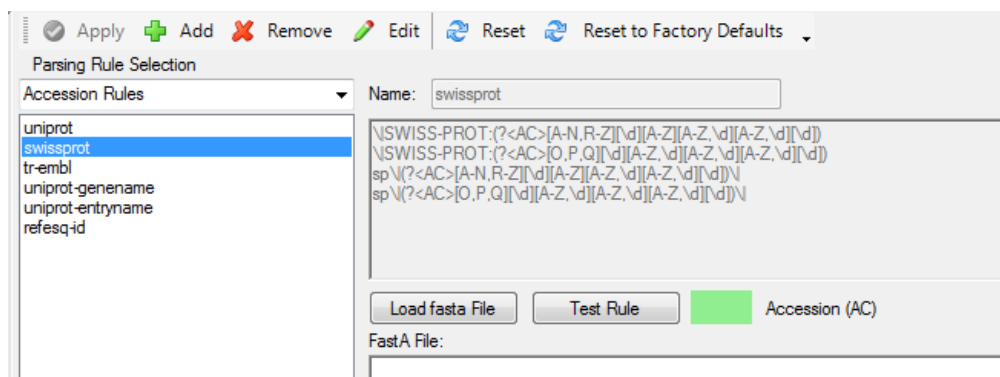
❖ **To add a regular expression**

1. Click ➕ Add .

2. Enter the new parsing rule as a regular expression in the parsing rule text area, and change the preset rule name, if needed.

   You can find a guide explaining the syntax of regular expressions on the Microsoft Developer Network (MSDN) home page (Regular Expression Language - Quick Reference page) at http://msdn.microsoft.com/en-us/library/az24scfc.aspx. The title line rules and the accession rules use named capture groups to capture the accession and description of the protein.

   • Title line rules (Title Line Rules area) must specify a named capture group, AC1, for the protein accession and a group, Desc1, for the protein description.

   • Accession rules (Accession Rules area) require a capture group, AC, for the accession.

   • Taxonomy rules (Taxonomy Rules area) require a matched taxonomy string.

   • Avoid expression rules (Avoid Expression Rules area) require a matched string.

❖ **To change or rename a regular expression**

Select the parsing rule in the list, and click 🖊 Edit .

You can now edit the rule in the parsing rule text area.

❖ **To test a parsing rule**

1. Click **Load FASTA File** to load the FASTA file for which the rule is designed.

   The first and last five title lines of the file are displayed in the text box of the test area (see Figure 130).

Figure 130. Loaded FASTA file from UniProt with first and last five title lines in the test area
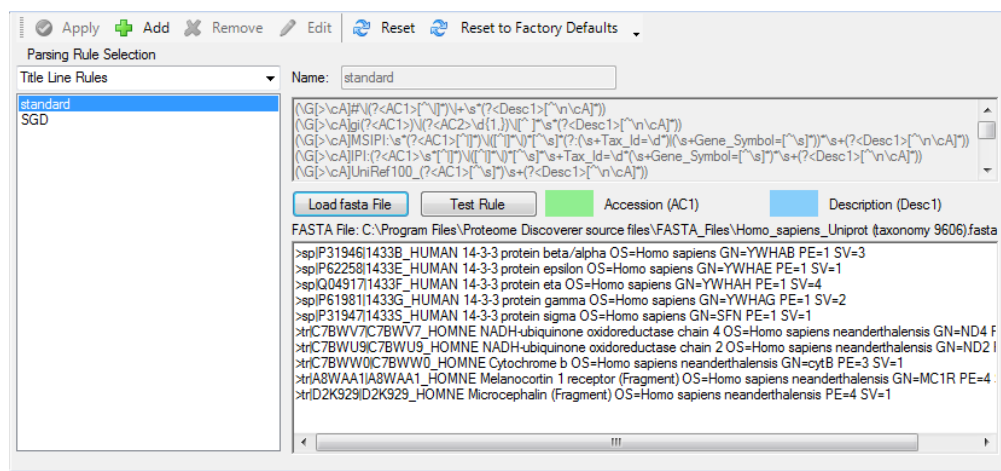
2. Click **Test Rule**.

The application applies the newly defined parsing rule to the small sample set of FASTA title lines read from the file. It highlights the matched values with a colored background, as shown in the following examples. If the application cannot match anything in the FASTA title line, it displays [No Match].

Test of the standard title line rule:

Test of the SwissProt accession rule:

Test of the human taxonomy rule:

Test of the refseq-id accession rule with no match:

4. Correct the parsing rule until it meets your needs.

❖ **To apply the changes**

To apply the changes, click .

After you apply the changes, the changed or newly defined parsing rules are available in the MSF Files node. For more information on the MSF Files node, refer to the Help.

Using the MSF Files Node to Change Title Line Parsing

You can use the MSF Files node to determine how the application parses the FASTA title lines of found proteins. When the application parses these title lines, it applies a set of predefined parsing rules to extract the accession and description of the protein. If it finds a protein in more than one results file from which it generates a report, and if the FASTA files of the two reports differ, the application displays the first available description and accession on the Proteins page. You can view all other accessions and descriptions when you move the mouse over the cells in the corresponding columns (see Figure 131).

Figure 131. Available accessions and descriptions

Accession

Description

The MSF Files node has advanced parameters that you can set to change the title line parsing. Changing these settings affects how the application displays protein accession and description information in the report. You can use the node's Title Line Rule parameter to define an alternative parsing rule to use to parse the accession and description. If the specified rule does not match anything in the FASTA title line, the Proteome Discoverer application tries the standard parsing rules to match the protein accession and description.

In cases where multiple accessions and descriptions are available for a given protein (see Figure 131 on page 187), you can determine the order of the displayed accessions and descriptions by using three other node parameters. The application applies them in this order:

> **Note**   The application applies the following rules only if multiple equally scored proteins are returned for the same set of PSMs, which rarely happens. Otherwise, it marks the highest-scoring protein reference as the master protein and displays it as such on the Protein Groups page. Even then, the protein marked as master is typically the longest of the group of equally scored protein references (the others are alternate master protein candidates).

1. Preferred Accession—Select a parsing rule to extract the preferred protein accession from the FASTA entry. If the application finds a preferred accession, it displays it instead of the primary accession.

   If you select a rule for the Preferred Accession parameter that matches one of the accessions, the application moves this accession and description to the first position in the list of available accessions and descriptions.

2. Preferred Taxonomy—Select a parsing rule to extract the preferred taxonomy from the FASTA entry. If the application finds a preferred taxonomy, it displays the accession and description of this entry, except when an entry containing a preferred accession is better than an entry containing preferred taxonomy and no preferred accession.

   If you performed the search without a preferred taxonomy and the application identifies proteins with the same sequence from different species, you can select a rule with the Preferred Taxonomy parameter to display the accession and description from the right species.

   Here is an example showing the precedence of the Preferred Taxonomy parameter. Suppose that you have one protein with more than one accession and description. Both descriptions contain some common keywords, for example:

   Description 1: *xxxxxxxxxx* human abc

   Description 2: *yyyyyyyyyy* human abc

   You set the Preferred Taxonomy parameter to a rule that includes "human." Then you set the Avoid Expressions parameter to a rule that includes "human."

   In this case, the Preferred Taxonomy parameter has a higher precedence than the Avoid Expression parameter.

3. Avoid Expressions—Select the terms that the application should avoid when parsing the protein description. If more than one description is available, the application prefers the description containing none of the specified terms.

   Some of the publicly available protein databases, such as UniProt or NCBI, collect experimental verified and curated proteins as well as unverified proteins, for example, the output from bioinformatic algorithms predicting potential proteins in a sequenced genome. In many cases, these unverified proteins contain words like "predicted" or

"hypothetical" in the description. The Avoid Expressions parameter matches these words in the description. Where there are two different descriptions from two different databases used in a complex search, the application displays the description that does not contain a word that you selected with the Avoid Expressions parameter.

The Proteome Discoverer application only applies the Avoid Expressions parameter if there are different FASTA title lines for the same protein that conform to the other title line rules described in this topic.

Here is an example showing how the Avoid Expressions parameter works. Suppose that you have one protein with more than one accession and description. Both descriptions contain some common keywords, for example:

Description 1: xxxxxxxxxx human abc

Description 2: yyyyyyyyyy human abc

You set the Avoid Expressions parameter to a rule that includes "abc."

In this case, the application takes the first accession in the list.

The application saves the parsing rules for the FASTA title lines in the FastaTitleParsingRules.xml file, which is stored in the C:\ProgramData\Thermo\Proteome Discoverer 2.1\MagellanDBs folder or equivalent. This file contains the parsing rules, along with name and meta information. A parsing rule is a list of regular expressions. If the application uses the parsing rule, it applies all regular expressions in the list in the order of the list, starting with the first. It uses the first rule that matches the title line to read out the accession and, if declared, the description.

The file contains a section for each of the four parsing rule parameters of the MSF Files node. The basic *.xml document with the four sections without rules looks like this:

```
<?xml version="1.0" encoding="utf-16"?>
<FastaTitlelineRules
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:xsd="http://www.w3.org/2001/XMLSchema">
  <TitlelineRules>
    ... enter rules here ...
  </TitlelineRules>
  <AccessionRules>
    ... enter rules here ...
  </AccessionRules>
  <TaxonomyRules>
    ... enter rules here ...
  </TaxonomyRules>
  <AvoidExpressionRules>
    ... enter rules here ...
  </AvoidExpressionRules>
</FastaTitlelineRules>
```

The parsing rules with their lists of regular expressions are defined as follows as in this example showing an accession rule to match SwissProt accessions:

```
<ParsingRule name="swissprot" isVisible ="true" changable="true">

<RuleParts>

<RulePart>\|SWISS-PROT:(?&lt;AC&gt;[A-N,R-Z][\d][A-Z][A-Z,\d][A-Z,\d][\d])</RulePart>

<RulePart>\|SWISS-PROT:(?&lt;AC&gt;[O,P,Q][\d][A-Z,\d][A-Z,\d][A-Z,\d][\d])</RulePart>

<RulePart>sp\|(?&lt;AC&gt;[A-N,R-Z][\d][A-Z][A-Z,\d][A-Z,\d][\d])\|</RulePart>

<RulePart>sp\|(?&lt;AC&gt;[O,P,Q][\d][A-Z,\d][A-Z,\d][A-Z,\d][\d])\|</RulePart>

</RuleParts>

</ParsingRule>
```

The first line of the rule defines the name that is displayed in the node parameter. It specifies whether the rule is visible in the list for the parameter. For new rules, always set `isVisible` and `changeable` to `true`; otherwise, it is impossible to apply the rule.

The body of the rule is a list of statements containing one or more elements named `RulePart`. All regular expressions are "or" connected in the final parsing rule. For accession rules, a named capture group, `(?<AC>)`, must match the accession. The Proteome Discoverer application evaluates it to extract the accession for display. You must use `<` instead of < and `>` instead of > because the < and > are not allowed in the XML entry. The title line rules contain two named capture groups, `?<AC1>` and `?<Desc1>`, for accession and description, respectively.

To add or modify FASTA parsing rules, see "Adding or Modifying FASTA Parsing Rules" on page 183.

Identifying Contaminants During Searches

In many cases, a proteomics sample contains contaminant proteins from sample preparations that are identified in the search. These contaminants have no impact on the scientific interpretation of the data, so you want to remove them from the displayed results file. You might also want to remove proteins from the sample preparation (for example, antibodies used in immunoprecipitation or proteins from affinity purification) and proteins that are known carryovers from the previous experiment (for example, peptides that still adhere to the HPLC column after the washing procedure).

Suppose that a core facility receives a sample from a microbiological laboratory and a request to search the data against the proteins of all known bacteria. If the preceding experiment involved a proteome of E. coli, it is likely that the HPLC column retains some E. coli peptides and that proteins are detected on the basis of these peptides. It is helpful if you can specify the E. coli FASTA file as an additional contaminant database and have the proteins highlighted in an additional column.

You can use a consensus workflow that includes the Protein Marker node to mark as contaminants all proteins in the results file that are listed in a designated FASTA file or files as contaminants. You can specify an original FASTA file and up to three additional FASTA files containing contaminants. You can also specify the name of the column in the results report that displays an X symbol marking the proteins.

Before you begin creating a workflow, add the FASTA file or files containing the proteins marked as contaminants to the Proteome Discoverer application. For instructions, see "Downloading FASTA Files to Proteome Discoverer" on page 153.

❖ **To create a marked contaminants processing workflow**

1. Create or open a study and an analysis:

   • To create a study, see "Creating a Study" on page 40.

   • To open an existing study, see "Opening an Existing Study" on page 40.

   • To create an analysis, see "Creating an Analysis" on page 65.

   • To open an existing analysis, see "Opening an Existing Analysis" on page 66.

2. Follow the general instructions for creating a processing workflow with the Workflow Editor. See "Creating a Processing Workflow" on page 68.

   The basic processing workflow for identifying contaminants is the same as the basic processing workflow shown in Figure 43 on page 72.

❖ **To create a marked contaminants consensus workflow**

1. Create a basic consensus workflow as described in "Incorporating an Existing Workflow into a Study and an Analysis" on page 99, but attach the Protein Grouping node to the ProteinScorer node, and attach the Protein Marker node directly to the Peptide and Protein Filter node (Figure 132).

Figure 132. Basic consensus workflow for marking proteins as contaminants
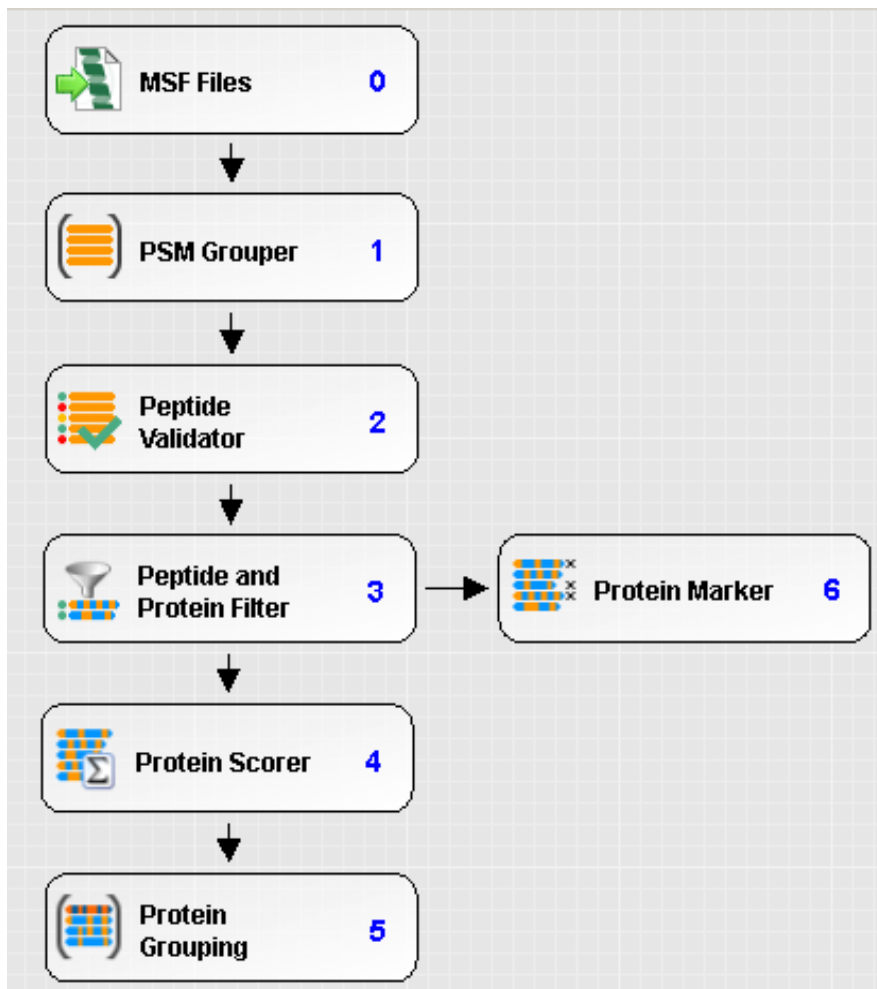
2. Set the parameters for the nodes, including the Protein Marker node. For more information, see the Help.

3. Run the workflow.

The results report contains a Contaminant column on the Proteins, Peptide Groups, and PSMs pages that displays an X symbol next to the proteins marked as contaminants in the searched FASTA file or files. Figure 133 shows the Contaminant column on the Proteins page.

Figure 133. Contaminant column in results file

Displaying Species Names for Proteins and Peptide Groups

In some cases, a proteomics sample might contain more than one species, either intentionally or accidentally, as a result of sample preparation. These different species can affect the scientific interpretation of the data, so you might want to filter them in the result report.

You can display the species names of the proteins and the peptide groups that they belong to as distribution maps or as text strings separated by semicolons in columns on the Proteins and Peptide Groups pages of the .pdResult file. The application extracts these names from the FASTA title lines of all proteins found in the sequence database search, removes duplicates, and assigns a unique color to each species name if you choose to display the names as a distribution map. Figure 134 shows the protein and peptide group names displayed in the Species Map and the Species columns on the Proteins page of the .pdResult report.

Figure 134. Species names as distribution maps and as text columns

The application can display up to 30 species names in a distribution map. If it extracts more names, it issues a warning and displays the additional names as text strings separated by semicolons.

Use the As Species Map and As Species Names parameters of the ProteinMarker node to display species names in the result file. For a description of these parameters, see the Help. Use the node in the same workflow as that given in "Identifying Contaminants During Searches" on page 190. It connects to the Peptide and Protein Filter node.

The annotation is restricted to FASTA database formats of UniProt (Swiss-Prot, TrEMBL), NCBI (RefSeq), and ProteinCenter downloads.

Searching Spectrum Libraries

Spectrum library search is a different search approach from the sequence database search commonly used in shotgun proteomics. The main difference between a database search and a spectrum library search is in the origin of the spectra that the measured spectra from your experiments are compared to. Sequence database searches use theoretical spectra generated from peptide sequences, but spectrum libraries are libraries of measured (consensus) spectra from actual previous experiments.

Using a library of already well-identified peptides avoids identifying already known peptides over and over again by a time-consuming database search. Restricting the library to previously identified peptides also drastically reduces the search space and therefore the search time. In addition, comparisons that use consensus spectra consider the measured peak intensities, increasing the selectivity and making the identification more accurate.

You can use the MSPepSearch node to search large spectrum libraries downloaded from the NIST or the PeptideAtlas home page.

- Displaying Spectrum Libraries

- Adding a Spectrum Library

- Deleting a Spectrum Library

- Searching Spectrum Libraries with the MSPepSearch Node

- Visually Verifying Spectrum Library Matches

Displaying Spectrum Libraries

You can display a list of all the spectrum libraries that you registered in the Proteome Discoverer application.

❖ **To list the available spectrum libraries**

Choose **Administration > Maintain Spectrum Libraries**.

The Spectrum Libraries view opens (see Figure 135). It lists all the spectrum libraries that you downloaded from NIST or the Peptide Atlas home page and that you registered. It displays the processed spectrum library properties, such as the file name, file size, the number of proteins stored, and the library type, which determines the search node to use. The application processes the spectrum library and makes it available for use.

Figure 135. Spectrum Libraries view

Add button

Remove button

Refresh button
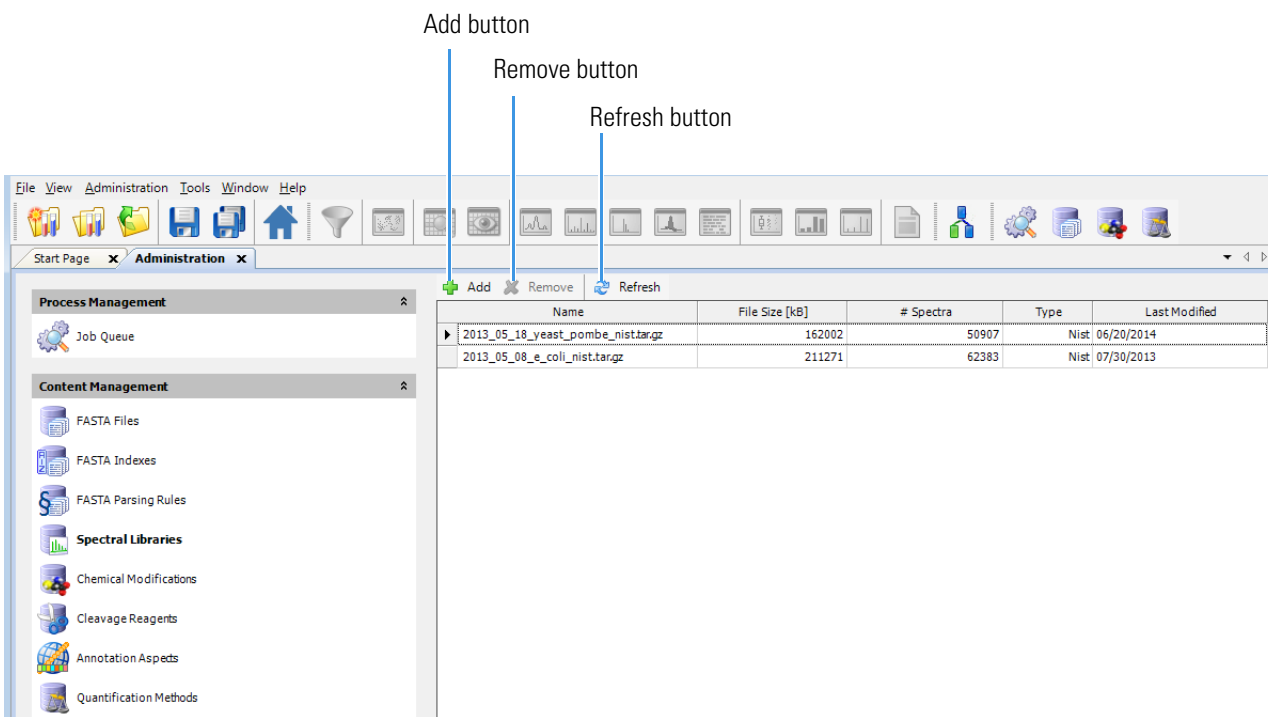
Adding a Spectrum Library

You must add a spectrum library to the Proteome Discoverer application before you can conduct a search with the MSPepSearch node. In the registration process, the application automatically recognizes the type of the spectral library. The type determines the search node that you can use the library with. Adding the spectrum libraries is similar to the procedure for adding FASTA files.

❖ **To add a spectrum library for searching with the MSPepSearch node**

1. Download the appropriate spectrum libraries from the National Institute of Standards and Technology (NIST) at http://peptide.nist.gov or from Peptide Atlas at http://www.peptideatlas.org/speclib.

   The Proteome Discoverer application recognizes the following file formats for searching spectrum libraries with the MSPepSearch node:

   • *.zip/*.gz files from NIST or PeptideAtlas. You can find these files in the *_nist.tar.gz file on the library download site at NIST or the *_nist.zip file on the PeptideAtlas home page. The file must contain a complete spectrum library in MSPepSearch. If files are missing, the Proteome Discoverer application does not add the library.

2. In the application, choose **Administration > Maintain Spectrum Libraries**.

3. Click 🞤 Add .

4. In the Select a Spectrum Library dialog box, do the following:

   a. In the list box in the lower right of the Select a Spectrum Library dialog box, select **All Spectrum Library Files** (*.gz, *.msp, *.zip) or Zip archives (*.gz, *.zip).

   b. Browse to the location of the spectrum library where you downloaded and unpacked the *_nist.tar.gz file.

   c. Select the *filename*.**gz** file.

   d. Click **Open**.

When you add a spectrum library file, the application takes the following steps:

- Extracts the archive file.

- Extracts spectra for visualization.

During library creation, the job queue in the Administration view displays each step (see Figure 136).

Figure 136. Adding a spectrum library for searching with the MSPepSearch node
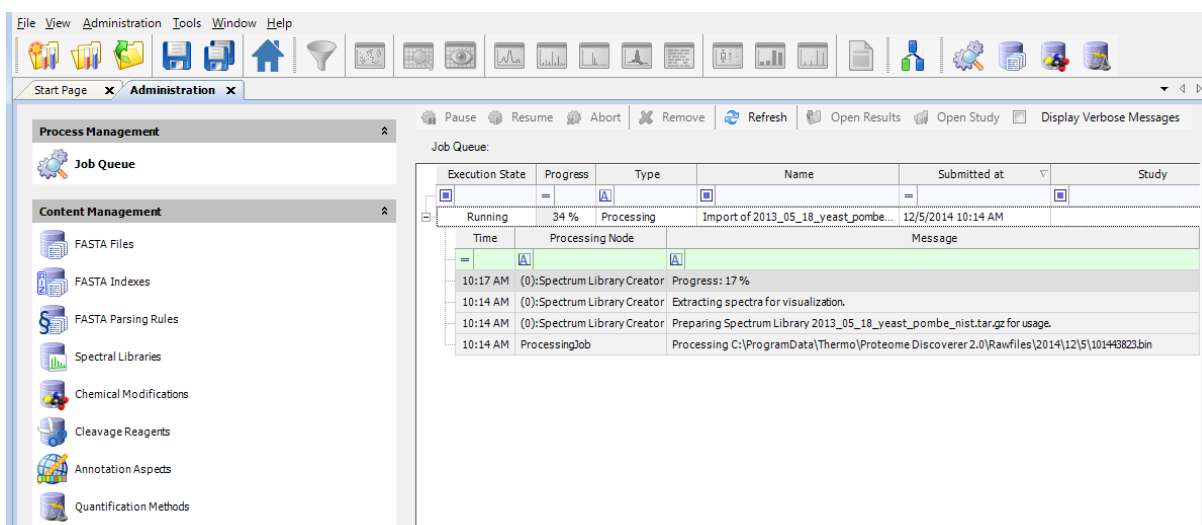
When the application finishes adding the spectrum library, the spectrum library file appears in the Spectrum Libraries view (see Figure 135 on page 195).

Now you are ready to search the spectrum library. To search with the MSPepSearch node, see "Searching Spectrum Libraries with the MSPepSearch Node." For more information on the MSPepSearch node, refer to the Help.
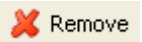
Deleting a Spectrum Library
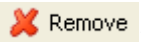
You can delete a spectrum library from the application.

❖ **To delete a spectrum library**

1. Choose **Administration > Maintain Spectrum Libraries**.

   The Administration page opens with the Spectrum Libraries view.

2. Click ▶ at the beginning of a row to select the row.

3. Click ✖ Remove .

4. In the Remove Spectrum Libraries Databases dialog box, click **OK**.

The Spectrum Libraries file that you selected appears as a job in the job queue. After you start the deletion of the file, you cannot cancel the deletion. You can remove the completed job from the job queue by clicking ✖ Remove , and then clicking OK in the Delete Jobs dialog box.

Searching Spectrum Libraries with the MSPepSearch Node

You can use the MSPepSearch node as an alternative to a search node such as Sequest HT.

The spectrum library search reports the three scores shown in Table 9. Dot score and reversed dot score are secondary scores, and their values are not shown by default.

Table 9. Scores generated by the MSPepSearch node

Score	Description
MSPepSearch	Is the main score of MSPepSearch.
Dot score	Is the score from a cross-correlation computed between two spectra.
Reverse dot score	Is the reversed spectral dot product.

❖ **To create a processing workflow for searching spectrum libraries with the MSPepSearch node**

Follow the instructions in "Creating a Processing Workflow" on page 68, but substitute the MSPepSearch node for the search engine node.

Figure 137 shows the basic processing workflow for searching spectrum libraries with the MSPepSearch node.

Figure 137. Basic processing workflow using the MSPepSearch node to search spectrum libraries
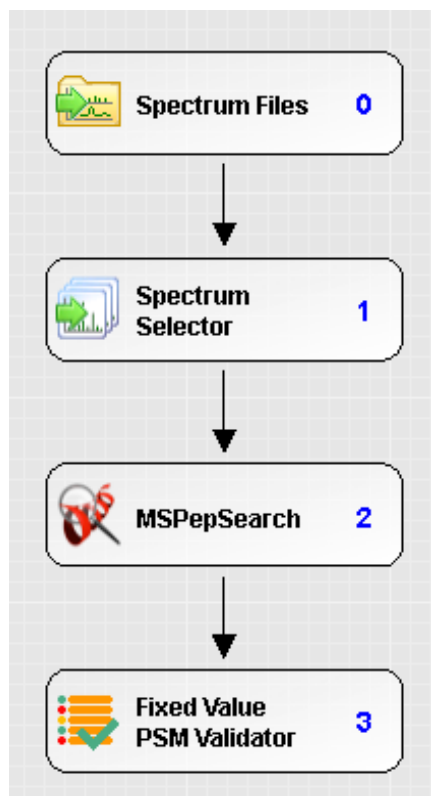
The Fixed Value PSM Validator node is the only possible peptide validator for the MSPepSearch node. It is impossible to perform a decoy search because there is no proper decoy spectrum library.

For a description of the parameters available in the MSPepSearch node, refer to the Help.

❖ **To create a consensus workflow for searching a spectrum libraries with the MSPepSearch node**

Follow the instructions in "Incorporating an Existing Workflow into a Study and an Analysis" on page 99.

Visually Verifying Spectrum Library Matches

You can visually verify matches between measured spectra from your experiment and the reference spectra in the spectrum library for peptides identified with the MSPepSearch node. In the Peptide Spectrum Match Identification Details view, you can display a mirror plot of the matching peptides (see Figure 138). You can use the reference spectrum with the fragment match settings (refer to the Help).

Figure 138. Mirror plot in the Peptide Spectrum Match Identification Details view
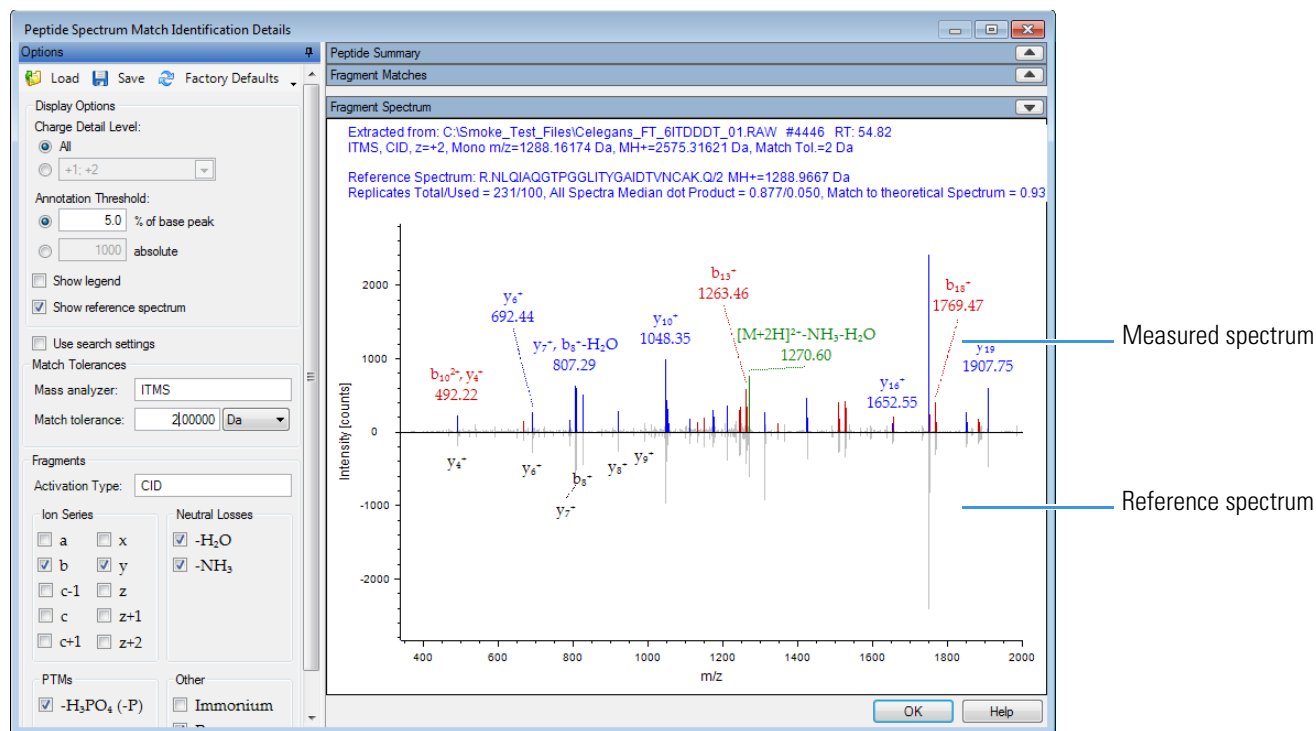
The application displays the reference spectrum using intensities multiplied by –1 in the same plot as the measured spectrum. In the reference spectrum, it also labels peaks of the a, b, c, ion series and the x, y, and z ion series, as well as the peaks from the precursor peptide. It does not display labels for all fragments with a mass difference, isotope peaks, and "?" peaks in the spectrum library.

❖ **To generate a mirror plot**

1. Open the .pdResult file for the results of the spectrum library search performed with the MSPepSearch node.

2. Follow the instructions for generating a Peptide Spectrum Match Identification Details view in the Help.

Defining Chemical Modifications

You can update the chemical modifications that you use to conduct a peptide identification search. The available modifications are defined in the Chemical Modifications view on the Administration page that opens when you choose Administration > Maintain Chemical Modifications. Use this view to customize the chemical modifications that you use to do your search. You can import a new list or the latest UNIMOD list. You can also modify the chemical modification list by adding amino acids to the modifications, creating new modifications, or activating or deactivating existing modifications.

> **Note** A modification must be active to be usable during a search.

- Dynamic Modifications
- Static Modifications
- Opening the Chemical Modifications View
- Adding Chemical Modifications
- Adding Amino Acids
- Deleting Amino Acids
- Importing Chemical Modifications

The Proteome Discoverer application offers two types of modifications, dynamic and static.

Dynamic Modifications

Dynamic modifications, also known as variable amino acid modifications, are modifications that might or might not be present. They are mainly used for determining post-translational modifications (PTMs). For example, some phosphorylated peptide serines are modified, and some are not modified.

Static Modifications

Static modifications apply the same specific mass to all occurrences of that named amino acid, as in an exhaustive chemical modification.

A static modification might result from derivatization or isotopic labeling of an amino acid. For example, carbamidomethylated cysteine has a delta mass of 57.021464, which is added to each cysteine residue appearing in a protein.

In static searches, the application assumes that every amino acid residue will be modified in that way. Constant mass is changed. The search engines perform static modification searches by adding the specified constant value to the mass of the specified amino acid.

Opening the Chemical Modifications View

The Chemical Modifications view is an advanced feature of the Proteome Discoverer application. You use it to build and maintain the static and dynamic modifications data that is available when you define your search settings.
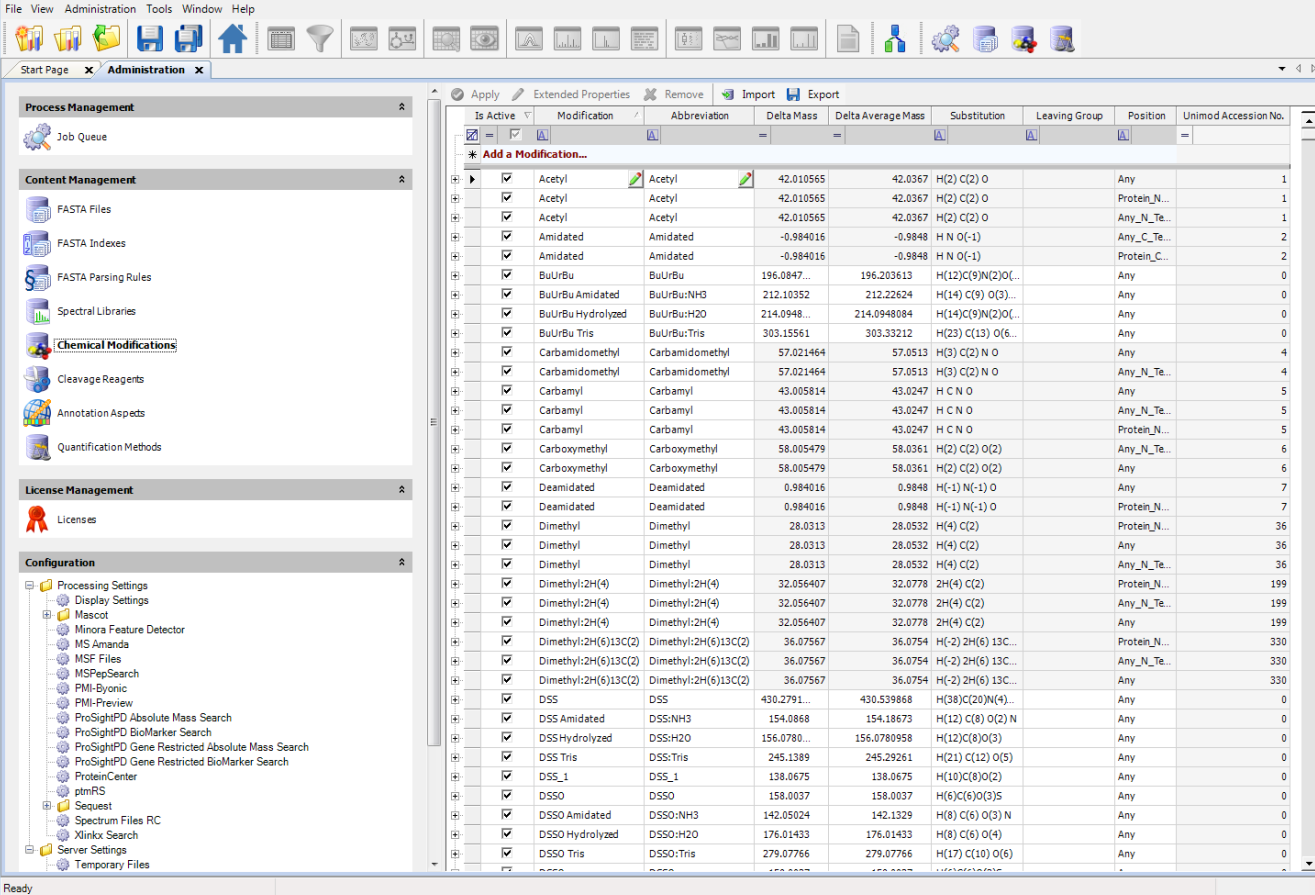
In the Chemical Modifications view, you can explore the default types of modifications and their corresponding amino acids. It contains the modification's delta mass, amino acids, and substitutions. By using the Chemical Modifications view, you can add amino acids to existing modifications and create new modifications.

❖ **To open the Chemical Modifications view**

1. Choose **Administration > Maintain Chemical Modifications**, or click the **Maintain Chemical Modifications** icon, 🔬, either in the toolbar or on the Administration page.

   The Chemical Modifications view opens on the Administration page (Figure 139). The amino acids listed are those where the modifications can appear.

Figure 139. Chemical Modifications view

2. Click the plus (+) sign to the left of each modification row to see the amino acids that the modification is found on, the letter abbreviation of this amino acid, and the modification type or category. As an example, Figure 140 shows information for the Acetyl modification. Table 10 lists the available modification categories.

Figure 140. Modification information for Acetyl displayed in columns

Table 10. Available modification categories (Sheet 1 of 2)

Classification	Description
Post-translational	Protein modification after translation (*in vivo*)
Co-translational	Amino acid modified in translation (for example, myristyl glycine)
Pre-translational	Amino acid modified before integration into a protein (for example, formyl methionine)
Chemical derivative	Chemically induced modification (for example, during sample preparation)
Artifact	Modification made during sample preparation
N-linked glycosylation	Glycosylation (in vivo)

Table 10. Available modification categories (Sheet 2 of 2)

Classification	Description
O-linked glycosylation	Glycosylation (in vivo)
Other glycosylation	Glycosylation (in vivo)
Synthetic peptide protection group	Protection group used in chemical peptide synthesis (for example, trityl (triphenylmethyl))
Isotopic label	Label for quantification
Non-standard residue	Amino acid derivative like selenomethionine
Multiple	More than one classification possible
AA substitution	Amino acid replaced by another amino acid (mutation)
Other	Modification not fitting into another category

The application automatically imports the classifications from unimod.org, the protein modifications online database for mass spectrometry applications. You can also manually define your own classifications.

Adding Chemical Modifications

You can create new chemical modifications and add them to the Chemical Modifications view. For example, you might have a new or experimental label that you want to add to the list of chemical modifications.

❖ **To add a new chemical modification**

1. Choose **Administration > Maintain Chemical Modifications**, or click the **Maintain Chemical Modifications** icon, , either in the toolbar or on the Administration page.
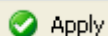
   The Chemical Modifications view opens (Figure 139).

2. Click the **Add a Modification** heading.

   An empty row appears in yellow (Figure 141).

Figure 141. Adding a row in the Chemical Modifications view

Is Active		Modification	Abbreviation	Delta Mass	Delta Average Mass	Substitution	Leaving Group	Position	Unimod Accession No.
☑	= ☑	A	A	=	=	A	A	A	=
✳	☑	✏	✏	1 ✏	1 ✏	✏	✏	Any	0
⊞	☑	Acetyl	Acetyl	42.010565	42.0367	H(2) C(2) O		Any	1
⊞	☑	Acetyl	Acetyl	42.010565	42.0367	H(2) C(2) O		Protein_N...	1
⊞	☑	Acetyl	Acetyl	42.010565	42.0367	H(2) C(2) O		Any_N_Te...	1
⊞	☑	Amidated	Amidated	-0.984016	-0.9848	H N O(-1)		Protein_C...	2
⊞	☑	Amidated	Amidated	-0.984016	-0.9848	H N O(-1)		Any_C_Te...	2
⊞	☑	Carbamidomethyl	Carbamidomethyl	57.021464	57.0513	H(3) C(2) N O		Any	4
⊞	☑	Carbamidomethyl	Carbamidomethyl	57.021464	57.0513	H(3) C(2) N O		Any_N_Te...	4
⊞	☑	Carbamyl	Carbamyl	43.005814	43.0247	H C N O		Any	5
⊞	☑	Carbamyl	Carbamyl	43.005814	43.0247	H C N O		Any_N_Te...	5

3. In the empty row, enter the name of the modification, the delta masses, the chemical substitution, the chemical group that is leaving, the position, and the abbreviations of the modifications.

4. If you select Any in the Position column, a message box opens to inform you that you must specify which amino acids (target amino acids) will possibly have the modification. For more information, see "Adding Amino Acids."

5. To accept the new modifications, click ✅ Apply .

6. Add an amino acid to the modifications. See "Adding Amino Acids."

❖ **To update an existing chemical modification**

1. Choose **Administration > Maintain Chemical Modifications**.

   The Chemical Modifications view opens (Figure 139 on page 201).

2. In the Modification column, click the cell that you want to update**.**

3. Type your changes for the delta masses, the substitution, the group that it is leaving, the position, or the abbreviations of the modifications.

   For chemical modifications that you add yourself, you can edit any column except the Unimod Accession No. column. The Unimod Accession No. column identifies these modifications by a zero. For chemical modifications that you import from UNIMOD, you can edit only the Modification and Abbreviation columns. UNIMOD chemical modifications are identified by a number greater than zero in the Unimod Accession No. column.

   Columns that you can edit activate an edit button when you click them. Columns that you cannot edit display a gray background.

4. To accept the changes, click ✅ Apply .

Adding Amino Acids

You can add amino acids to a modification that has been set up for any position.

❖ **To add an amino acid to a modification**

1. Choose **Administration > Maintain Chemical Modifications**.

   The Chemical Modifications view opens (Figure 139 on page 201).

2. Click the plus (**+**) sign to the left of the modification row that you want to update.

   The row must display Any in the Position column.

   The list of classifications appears (Figure 140 on page 202).

3. Click the **Add a Modification** line (Figure 142 on page 205), below the list of amino acids.

Figure 142. Adding an amino acid to a modification

An empty row appears.

4. In the empty row, select the amino acid from the list in the Amino Acid Name column.

The amino acid and the one-letter abbreviation appear.

5. From the list in the Classification column, select the type of modification.

6. To save the modifications, click Apply .

When you reimport data from unimod.org, the application retains the modification that you added. However, if you want to change the classification of an amino acid, you must do so before reimporting the Unimod data. After you import the Unimod data, the only way to change the classification is to delete the amino acid and re-add it with another classification.

Deleting Amino Acids

You can also delete amino acids from chemical modifications.

❖ **To delete an amino acid from a chemical modification**

1. Choose **Administration > Maintain Chemical Modifications**.

The Chemical Modifications view opens (see Figure 139 on page 201).

2. Click the plus (**+**) sign to the left of the modification row that you want to delete.

The row expands and the associated amino acids appear.

3. Select the amino acid row that you want to delete.

4. Click Remove .

5. In the Delete Row dialog box, click **Yes**.

Deleting Chemical Modifications

You can remove chemical modifications from the Chemical Modifications view.

❖ **To delete a modification**

1. Choose **Administration > Maintain Chemical Modifications**.

   The Chemical Modifications view opens (Figure 139 on page 201).

2. Select the row of the modification that you want to delete.

3. Click ![X] Remove .

4. In the Delete Row dialog box, click **Yes**.

Importing Chemical Modifications

You can import chemical modifications from a local file or obtain an updated version from unimod.org, a public domain database.

When you install the Proteome Discoverer application, it automatically imports accessions from unimod.org as chemical modifications.

❖ **To import chemical modifications from a local file**

1. Choose **Administration > Maintain Chemical Modifications**.

   The Chemical Modifications view opens (Figure 139 on page 201).

2. Click ![] Import .

3. In the Import From List of the Import Modifications dialog box, select **Local File**.

4. In the adjacent box, click the **Browse** button (…) to browse to your file, or type the name and path of the file in the box.

5. To overwrite an existing upload, select the **Overwrite Existing** check box.

6. Click **Import**.

   A status message opens.

7. When the upload is complete, click **Close**.

❖ **To import chemical modifications from unimod.org**

1. Choose **Administration > Maintain Chemical Modifications**.

   The Chemical Modifications view opens (Figure 139 on page 201).

2. Click ![] Import .

   The Import Modifications dialog box opens (Figure 143).

Figure 143. Import Modifications dialog box
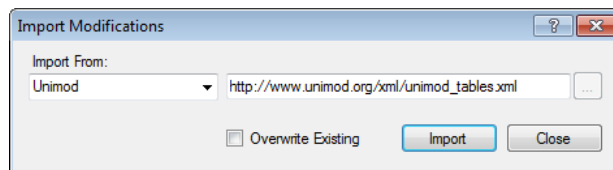
3. In the Import From list, select **Unimod**.

   The UNIMOD URL appears in the adjacent box.

4. To overwrite an existing upload, select the **Overwrite Existing** check box.

5. Click **Import**.

   A status message opens.

6. When the upload is complete, click **Close**.

For chemical modifications imported from unimod.org, you can only edit the Is Active, Modification, and Abbreviation columns. You do not have access to the Delta Mass, Delta Average Mass, Substitution, Leaving Group, Position, and UNIMOD Accession No. columns. Chemical modifications imported from unimod.org have a number greater than zero in the Unimod Access No. column.

If you select the Overwrite Existing check box, the application does the following when it imports chemical modifications from unimod.org:

- Updates the columns that are inaccessible to you.

- Updates the names and the abbreviations of the modifications.

- Adds any new amino acids found in unimod.org.

- Adds any amino acids that you removed if they are defined in unimod.org.

- Removes any amino acids that you added if they are defined in unimod.org.

If you do not select the Overwrite Existing check box, the application performs the same tasks as it does during installation:

- Updates the columns that are inaccessible to you.

- Leaves the modification name and abbreviation unchanged.

- Adds any new amino acids found in unimod.org.

- Adds any amino acids that you removed if they are defined in unimod.org.

- Leaves unchanged any amino acids that you added.

Using the Qual Browser Application

The Proteome Discoverer application provides a link to the Qual Browser application from selected PSMs if the application was already installed on the computer. With Qual Browser, you can examine spectra and chromatograms in detail, view the entire ion chromatogram, and browse individual precursor and MS^n data. You can filter the results in a variety of ways, for example, to produce a selected ion chromatogram. The Qual Browser application automatically displays the elemental composition, theoretical mass, delta values, and ring and double-bond (RDB) equivalents for your high-resolution data. (RDB equivalents measure the number of unsaturated bonds in a compound and limit the calculated formulas to only those that make sense chemically.)

You must have the Xcalibur data system installed to use Qual Browser. For information about using this application, refer to the *Thermo Xcalibur Qual Browser User Guide*.

You must also have a search results file open and a specific peptide or search input row selected before the Qual Browser application becomes available. If you are viewing the Administration page, Qual Browser does not open a raw data file.

❖ **To open the Qual Browser application**

1. Open a .pdResult file.

2. Choose **Tools > Open QualBrowser**, choose **Tools > Open QualBrowser**, or press CTRL+SHIFT+B to open the Spectrum window.

> **Note** You must have a search results (.pdResult) file open and selected before the Open QualBrowser command becomes available on the Tools menu. In addition, the Open QualBrowser command is available only when you select at least a single peptide or a search input item first. You cannot use QualBrowser if the original raw data file or files are missing. The .pdResult file and the raw data file must reside in the same directory.

The Qual Browser application opens (see Figure 144).

Figure 144. Qual Browser application window
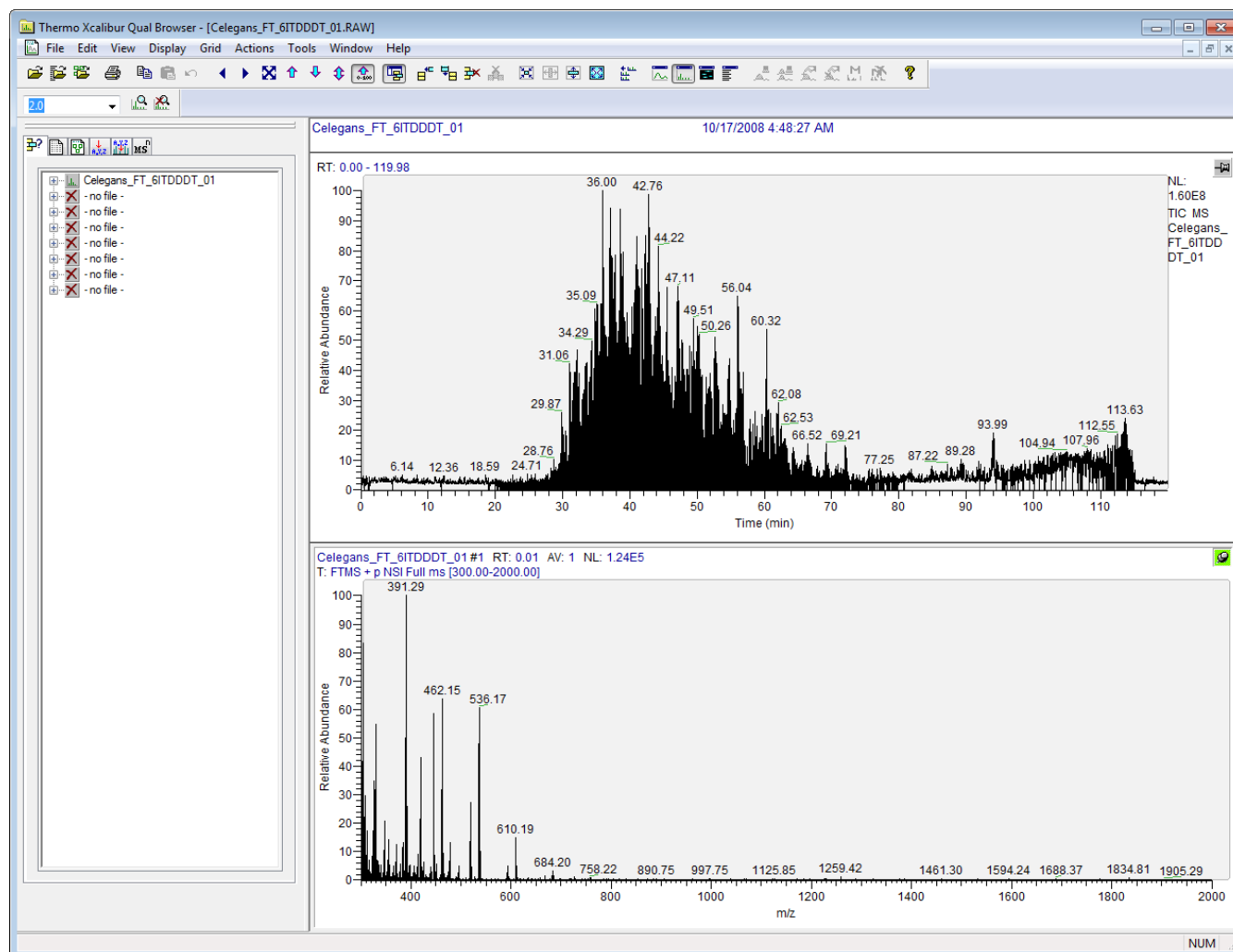
3. Right-click in the application window and choose **Display Options** from the shortcut menu.

4. To automatically annotate your peaks with the elemental composition, theoretical mass, RDB equivalent, or mass delta, click the **Composition** tab and select the labels for display.

Customizing Cleavage Reagents

In the Cleavage Reagents view, you can explore the default types of reagents and their corresponding settings. You can also add, remove, and modify the reagents and their corresponding settings. The Cleavage Reagents view contains the cleavage sites, cleavage inhibitors, abbreviations, and cleavage specificities.

For information about enzyme cleavage, see "Enzyme Cleavage Properties" on page 640.

- Opening the Cleavage Reagents View

- Adding a Cleavage Reagent

- Deleting a Cleavage Reagent

- Modifying a Cleavage Reagent

- Filtering Cleavage Reagent Data

Opening the Cleavage Reagents View

❖ **To display the Cleavage Reagents view**

Choose **Administration > Maintain Cleavage Reagents**.

The Cleavage Reagents view opens (see Figure 145).

Figure 145. Cleavage Reagents view
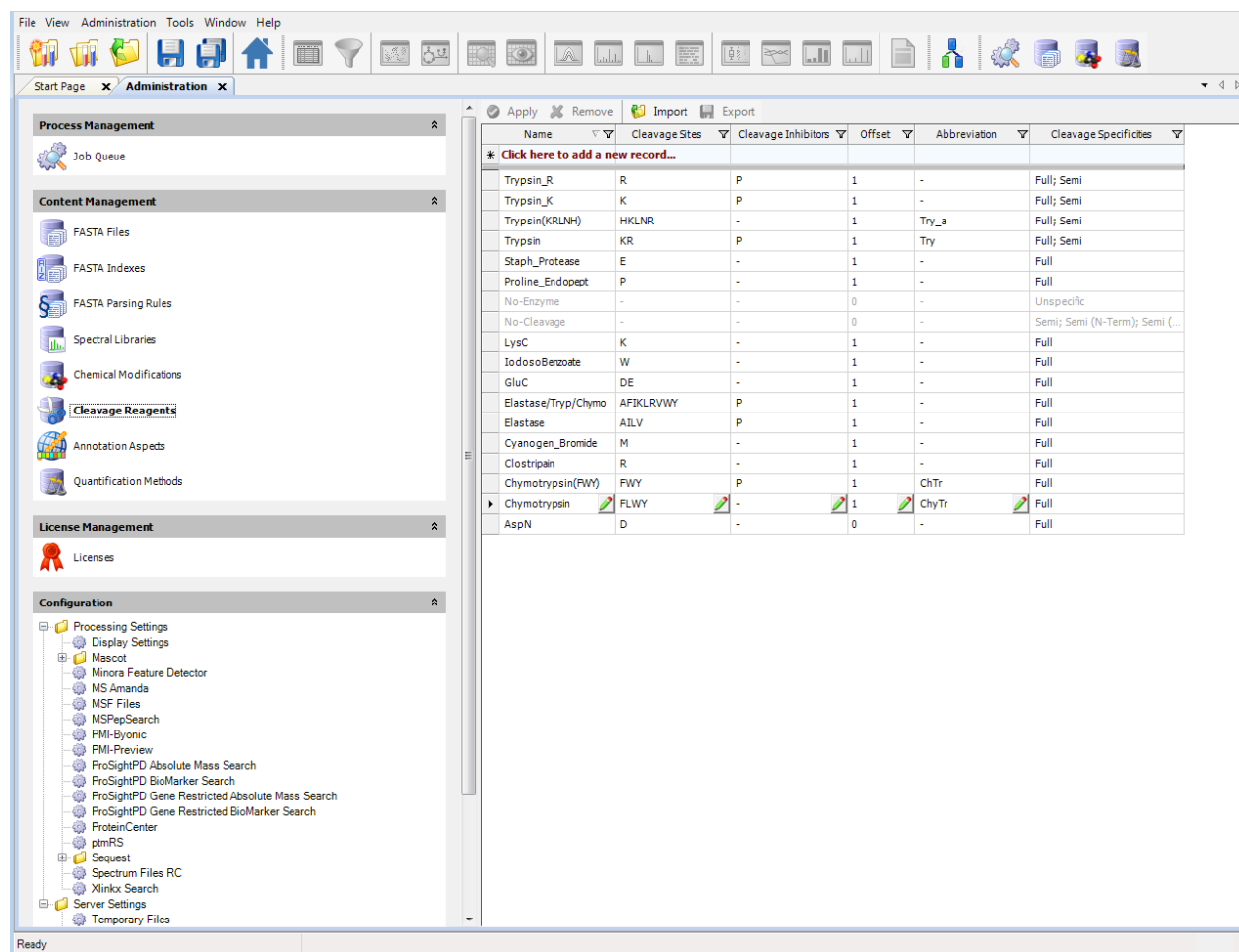
Cleavage Enzyme Specificities

The application supports the following cleavage specificities for enzymes:

- Full: Specifies a full enzymatic digestion.

- Semi: Specifies a semi-enzymatic digestion. The cleavage is specific at one terminal end.

- Semi (N-Term): Specifies that the cleavage be specific at the amino terminal end and unspecific at the carboxyl terminal end.

- Semi (C-Term): Specifies that the cleavage be specific at the carboxyl terminal end and unspecific at the amino terminal end.

- Enzymatic Unspecific: Specifies an unspecific digestion with an extra weight for enzymatically digested peptides.

The list of cleavage reagents includes the No-Enzyme and No-Cleavage reagents:

- The No-Enzyme reagent digests a sequence at every amino acid. For example, if the FASTA file contains the ACDEFGHIK sequence, No-Enzyme cleaves it into peptides A, AC, ACD, ..., C, CD, CDE, ..., D, DE, DEF.

  The application supports the following specificity for the No-Enzyme reagent:

  - Unspecific: Specifies a non-specific digestion.

- The No-Cleavage reagent does not cleave a sequence. For example, if the FASTA file contains the ACDEFGHIK sequence, the full sequence is the only possible sequence.

  The application supports the following cleavage specificities for the No-Cleavage reagent:

  - No-Cleavage: Specifies that no cleavages occur, so intact proteins result.

- Semi, Semi-N, Semi-C: Specifies terminal signal peptides that are cleaved off in vivo.

Cleavage Reagents View Parameters

Table 11 describes the parameters in the Cleavage Reagents view.

Table 11. Cleavage Reagents view parameters (Sheet 1 of 3)

Parameter	Description
Remove	Deletes the highlighted cleavage reagent.
Apply	Saves any changes made to the cleavage reagents.
Name	Specifies the name of the reagent used for the protein digestion.
	• All: Returns the filtered search results to the results that were first loaded.
	• Custom: Opens the Custom Filter dialog box (Figure 146 on page 215).
	• Blanks: Filters out rows that have data-filled cells in the column whose funnel icon you clicked.
	• NonBlanks: Filters out rows that have empty cells in the column whose funnel icon you clicked.
	• List of enzymes: Lists the specific names of all available enzymes that act as reagents.

Table 11. Cleavage Reagents view parameters (Sheet 2 of 3)
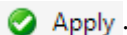
Parameter	Description
Cleavage Sites	Specifies the position (amino acid) at which to cleave the sequence.
	• All: Returns the filtered search results to the results that were first loaded.
	• Custom: Opens the Custom Filter dialog box (Figure 146 on page 215).
	• Blanks: Filters out rows that have data-filled cells in the column whose funnel icon you clicked.
	• NonBlanks: Filters out rows that have empty cells in the column whose funnel icon you clicked.
	• List of amino acids: Lists the names of the amino acids at which the enzyme cleaves the peptide.
Cleavage Inhibitors	Specifies the amino acids that block cleavage when adjacent to the cleavage site.
	• All: Returns the filtered search results to the results that were first loaded.
	• Custom: Opens the Custom Filter dialog box (Figure 146 on page 215).
	• Blanks: Filters out rows that have data-filled cells in the column whose Filter icon you clicked.
	• NonBlanks: Filters out rows that have empty cells in the column whose Filter icon you clicked.
	• -: No amino acids inhibit cleavage.
	• P: Adjacent proline amino acids inhibit cleavage.
Offset	Specifies at which side of the given amino acids the protease should cleave:
	• All: Returns the filtered search results to the results that were first loaded.
	• Custom: Opens the Custom Filter dialog box (Figure 146 on page 215).
	• Blanks: Filters out rows that have data-filled cells in the column whose Filter icon you clicked.
	• NonBlanks: Filters out rows that have empty cells in the column whose Filter icon you clicked.
	• 0: The protease cleaves at the N-terminal side of the specified cleavage sites.
	• 1: The protease cleaves at the C-terminal side of the specified cleavage sites.

Table 11. Cleavage Reagents view parameters (Sheet 3 of 3)

Parameter	Description
Abbreviation	Specifies the user-defined abbreviation.
	• All: Returns the filtered search results to the results that were first loaded.
	• Custom: Opens the Custom Filter dialog box (Figure 146 on page 215).
	• Blanks: Filters out rows that have data-filled cells in the column whose Filter icon you clicked.
	• NonBlanks: Filters out rows that have empty cells in the column whose Filter icon you clicked.
	• List of abbreviated enzymes: Lists the existing user-defined abbreviations.
Cleavage Specificities	Selects the cleavage specificity for enzymes. You can select more than one cleavage specificity. The available cleavage specificities are as follows:
	• Full: Specifies a full enzymatic digestion. Sequences are not cleaved.
	• Semi: Specifies a semi-enzymatic digestion. The cleavage is specific at one terminal end.
	• Semi (N-Term): Specifies that the cleavage be specific at the amino terminal end and unspecific at the carboxyl terminal end.
	• Semi (C-Term): Specifies that the cleavage be specific at the carboxyl terminal end and unspecific at the amino terminal end.
	• Enzymatic Unspecific: Specifies an unspecific digestion with an extra weight for enzymatically digested peptides.
Click here to add a new record	Opens a row in the reagents table so that you can enter the name and information about a new cleavage reagent.

Adding a Cleavage Reagent

❖ **To add a new cleavage reagent**

1. Click the Name column cell and click **Click Here To Add a New Record**.

2. Modify the default values in the row of that new reagent.

3. Click ✓ Apply .
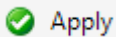
Deleting a Cleavage Reagent

❖ **To delete a cleavage reagent**

1. Click the box in the * column next to the row that you want to delete.

2. Click **Delete** and then click **Yes** in the confirmation box.

Modifying a Cleavage Reagent

❖ **To modify a cleavage reagent**

1. Click in the column for the reagent you want to modify, select the current contents, and enter the new information.

2. Click ✅ Apply .

Filtering Cleavage Reagent Data

❖ **To filter cleavage reagent data**

1. Click the **Filter** icon, 🔽 , next to the header of the column.

2. Select one of the following:

   • All: Returns the filtered search results to the results that were first loaded.

   • Custom: Opens the Custom Filter dialog box (Figure 146).

   Figure 146. Custom Filter dialog box
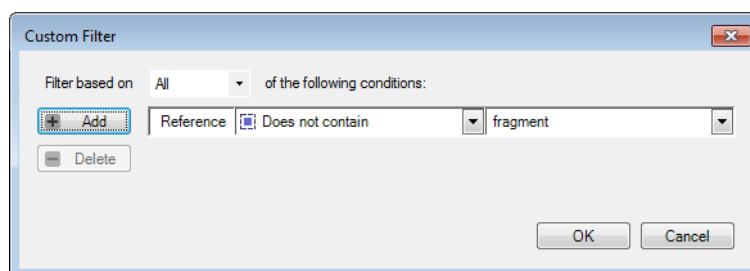
   For information about using this type of dialog box, see "Filtering with Display Filters" on page 225.

   • Blanks: Filters out rows that have data-filled cells in the column whose Filter icon you clicked.

   • NonBlanks: Filters out rows that have empty cells in the column whose Filter icon you clicked.

Filtering Data

A single or multiconsensus .pdResult file displays a list of PSMs and proteins identified by the search engine that you specify. This chapter explains how to sort and filter the data in your Proteome Discoverer results report.

Contents

- Filtering Results
- Filtering PSMs
- Filtering Proteins
- Filtering Phosphorylation Site Probabilities
- Filtering with Display Filters

Filtering Results

You can filter the results of an analysis during processing or afterward in the opened results report.

- In the processing workflow, you can use the Maximum Delta Cn parameter of the PSM validation nodes (Percolator, Target Decoy PSM Validator, Fixed Value PSM Validator) to reduce the number of PSMs stored and to affect the PSM FDR estimation. For information on these nodes, refer to the Help.

- In the consensus workflow, you can use the filtering parameters of the MSF Files node, the Peptide and Protein Filter node, and the Protein Grouping node to determine the set of protein groups, proteins, peptide groups, and PSMs displayed in a .pdresult file. For information on these nodes, refer to the Help.

- In the .pdResult file, you can apply an extensive set of editable filters, called display filters, to any of the open pages to simplify and streamline the data to review. You can use display filters to show or hide data with particular properties. Applying display filters does not affect calculations performed by workflows, such as confidence levels, coverage, or rank.

Filtering PSMs

Search engines often provide multiple possible matching peptides as explanations for the same spectrum. Most of the time you can clearly distinguish the top-scoring match from the other PSMs, but sometimes, especially in the presence of dynamic modifications, the best-scoring matches of the same spectrum have very similar scores. In this case, you can filter the results to select the best-scoring PSMs and the matches that have very similar scores by using the Maximum Delta Cn parameter or the Maximum Rank parameter of the following nodes:

- Processing workflow:

  – Filtering PSMs with the Percolator Node

  – Filtering PSMs with the Fixed Value PSM Validator Node

  – Filtering PSMs with the Target Decoy PSM Validator Node

- Consensus workflow:

  – Filtering PSMs with the MSF Files Node

The Δ Cn value displays the normalized score difference between the currently selected PSM and the highest-scoring PSM for that spectrum:

$$\Delta Cn(\text{rank i}) = \frac{score_{\text{rank 1}} - score_{\text{rank i}}}{score_{\text{rank 1}}}$$

The Maximum Delta Cn parameter filters out all PSMs with a Δ Cn score larger than the specified value.

On the PSMs page of the .pdResult file, the Δ Cn column displays the Δ Cn values. For example, Figure 147 shows how the score of a peptide ranked 2 compares to other high-confidence peptides from the same spectrum.

Figure 147. Δ Cn scores for multiple high-confidence peptides from the same spectrum

Filtering PSMs with the Percolator Node

Figure 148 shows the parameters available for the Percolator node (for information about these parameters, refer to the Help). Setting its Maximum Delta Cn parameter to 0 ensures that the application consider only rank 1 PSM matches for further processing with Percolator. When you set this parameter to a slightly higher value, the application might consider PSMs with a rank higher than 1.

A slightly higher value can be a good option in case of phosphopeptide analysis, where the first two or more matches often refer to the same peptide sequence and differ solely in the potential location of the phosphorylation site. Their scores are usually very close, if not identical, so they generate very low Delta Cn values. The default setting of the Maximum Delta Cn parameter, 0.05, is intended to accommodate such cases. You can specify a maximum value of 0.1 for the Maximum Delta parameter. Allowing PSMs with even higher Delta Cn values negatively affects Percolator's performance. In addition, it hugely increases the volume of data to be processed.

Figure 148. Percolator node parameters
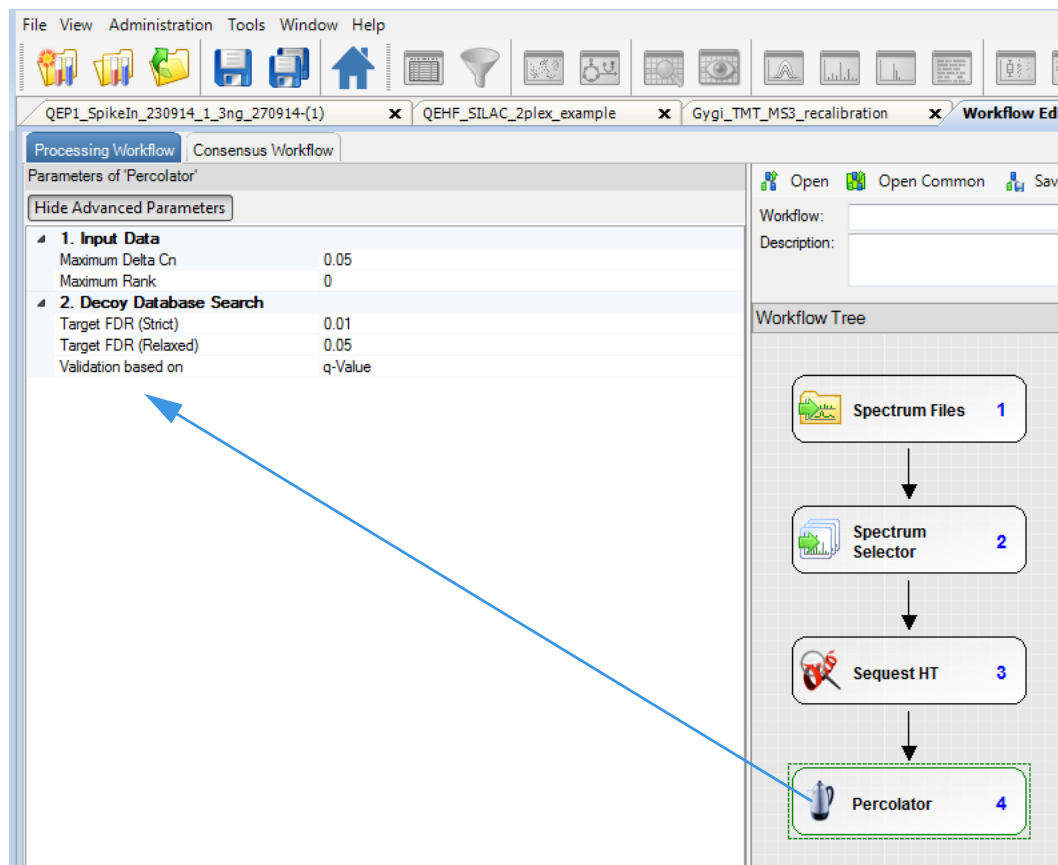
Filtering PSMs with the Fixed Value PSM Validator Node

For information on filtering with the Fixed Value PSM Validator node in the processing workflow, refer to the Help.

Filtering PSMs with the Target Decoy PSM Validator Node

For information on filtering with the Target Decoy PSM Validator node in the processing workflow, refer to the Help.

Filtering PSMs with the MSF Files Node

You can also use four parameters of the MSF Files node in the consensus workflow to filter PSMs: the Maximum Delta Cn, the Maximum Rank, the Maximum Delta Mass, and the Score/Threshold parameters of the individual search node.

❖ **To filter PSMs with the Maximum Delta Cn parameter**

1. Create a consensus workflow that contains the MSF Files node.

   For information about creating a basic consensus workflow, see "Incorporating an Existing Workflow into a Study and an Analysis" on page 99.

2. Set any parameters for the MSF Files node.

   For information on these parameters, refer to the Help.

3. Click **Show Advanced Parameters** in the Parameters pane of the Workflow Editor, and specify a value for the Maximum Delta Cn parameter to set a threshold that determines which PSMs the application transfers to the report file.

   Figure 149 shows this parameter.

   The threshold can be within a range of 0 to 0.1. The default is 0.05.

   The application transfers only those PSMs with a Δ Cn value greater than the specified threshold to the report file.

Figure 149. MFS Files node parameters
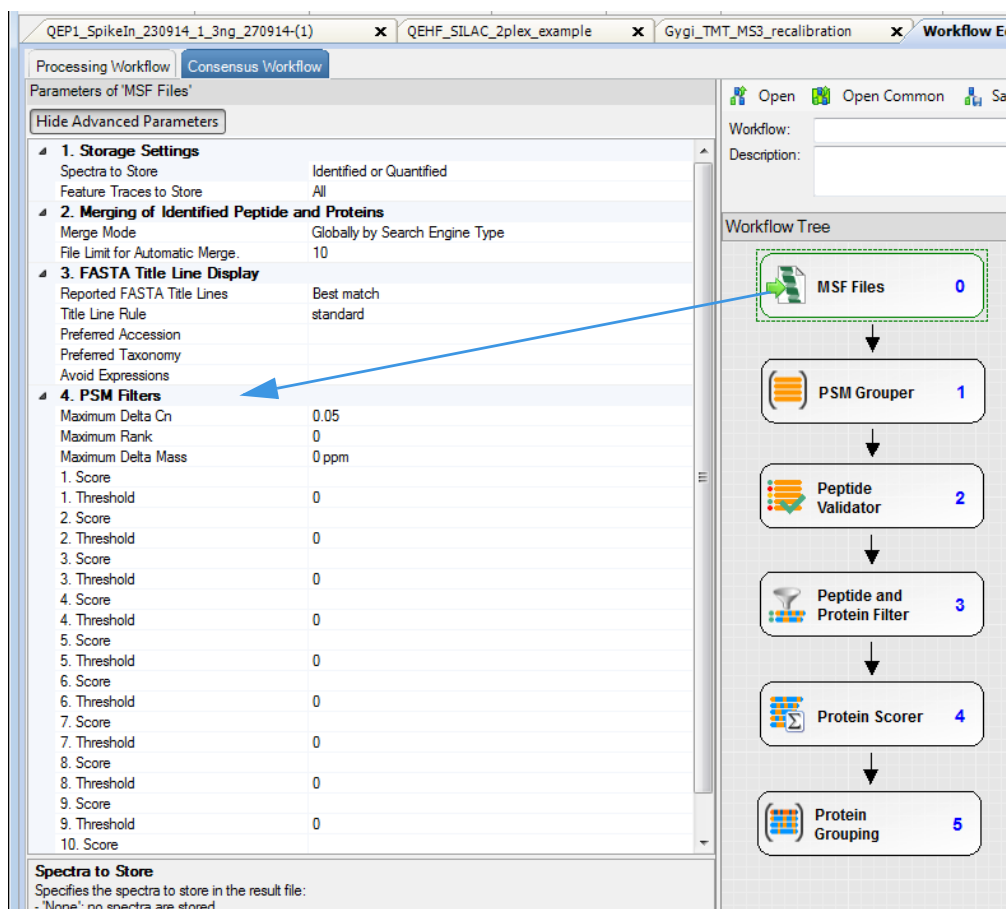
❖ **To filter PSMs with the Maximum Rank parameter**

1. Create a consensus workflow that contains the MSF Files node.

    For information, see Incorporating an Existing Workflow into a Study and an Analysis.

2. Set any parameters for the MSF Files node.

    For information on these parameters, refer to the Help.

3. Click **Show Advanced Parameters** in the Parameters pane of the Workflow Editor.

4. Specify a value for the Maximum Rank parameter to set a threshold that determines which PSMs the application transfers to the report file.

    Figure 149 shows this parameter.

    You can set the threshold to any positive value. The default is 0, which means that the Maximum Rank filter is not applied.

    The application transfers to the report file only those PSMs with a rank better than or equal to the specified threshold.

❖ **To filter PSMs with the Maximum Delta Mass parameter**

1. Create a consensus workflow that contains the MSF Files node.

    For information about creating a basic consensus workflow, see Incorporating an Existing Workflow into a Study and an Analysis.

2. Set any parameters for the MSF Files node.

    For information on these parameters, refer to the Help.

3. Specify a value for the Maximum Delta Mass parameter to indicate whether the application should apply a delta mass filter and exclude from the final result peptide matches with a larger mass difference between the theoretical and the found peptide *m/z*.

    Use a value within the range of 0.0–5.0 Da or 0.0–5000 ppm. The default is 0 ppm. If you set this parameter to zero, the application does not apply a delta mass filter.

❖ **To filter PSMs with the Score and Threshold parameters**

1. Create a consensus workflow that contains the MSF Files node.

    For information about creating a basic consensus workflow, see Incorporating an Existing Workflow into a Study and an Analysis.

2. Set any parameters for the MSF Files node.

    For information about these parameters, refer to the Help.

3. Set the filter parameters for the node:

a. In the 1. Score box under PSM Filters, select the name of the score to filter by from the list, as shown in Figure 150.

Figure 150. Selecting the score to filter by
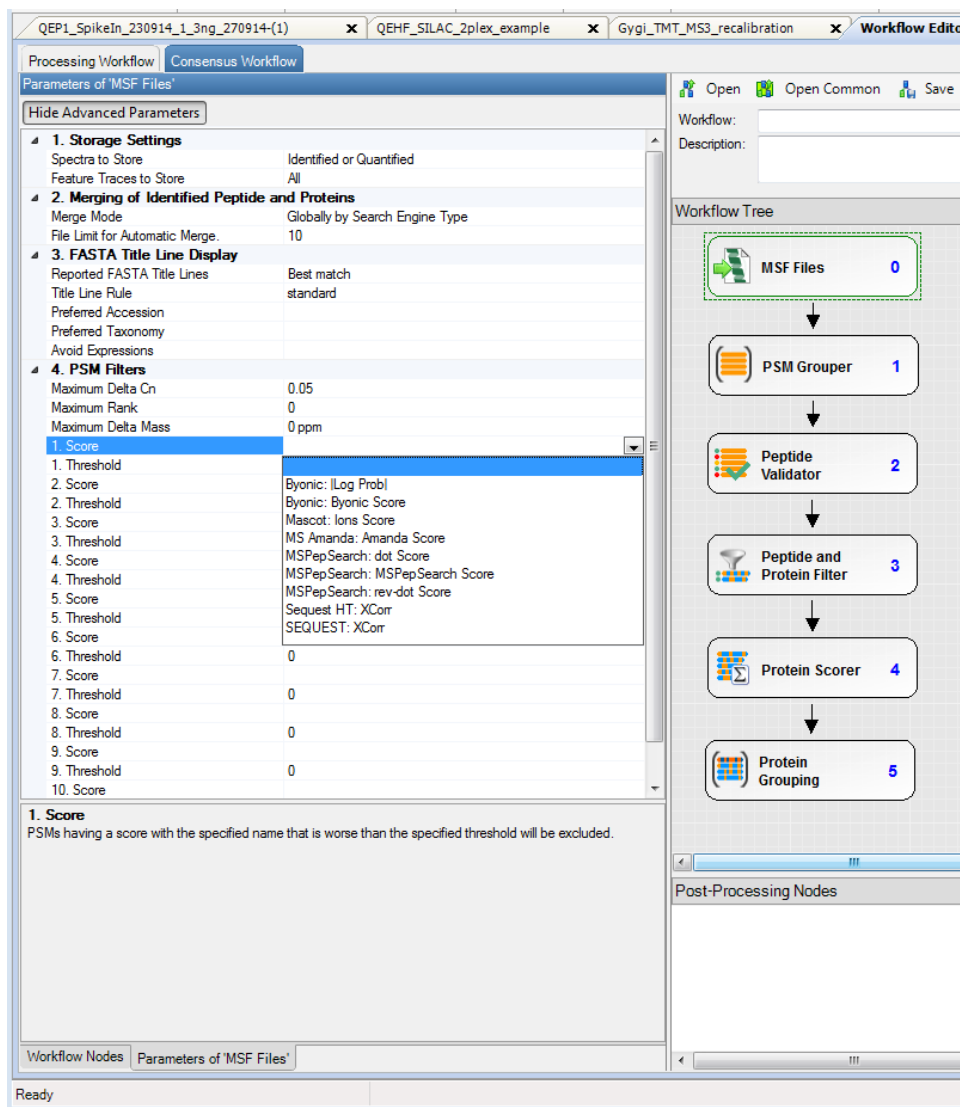
If you do not see the name of the filter that you want to filter by, see To specify the list of score names to filter by for instructions on adding score names to the list.

b. In the 1. Threshold box, type the score to filter by.

c. Repeat the previous two steps for any additional filters.

You can specify up to 10 PSM scores and thresholds to filter by.

For each score type, the application excludes scores lower than the specified threshold.

❖ **To specify the list of score names to filter by**

1. Choose **Administration > Configuration > MSF Files**.

2. Click the **Browse** button (...) next to the PSM Scores parameter.

   The Edit Parameter Text for PSM Scores dialog box opens, as shown in Figure 151.

   Figure 151. Edit Parameter Text for PSM Scores dialog box
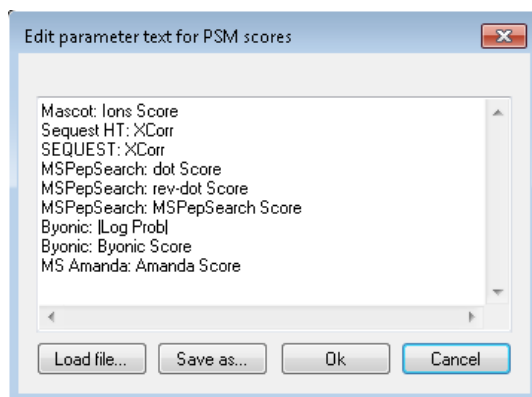
3. Beneath the last line, type the name of the score type that you want to add in the format *search_node_name:score_name*.

   You can find the search node name in the Workflow Editor or in the Result Summaries when you open the .pdResult file. The column on the PSMs page that shows the score also displays the search node name.

   The specified string for the score type name must be in the format *search_node_name:score_name*. If a string does not match this format, it is not available in the score parameters of the filter node. If the search node name and score names are not correct, the filter node cannot filter the PSMs on the basis of that score. If the node cannot find a specified score, it issues a warning message.

4. Click **OK**.

5. Close the study and reopen it.

   The added score type is now available in the list in the Edit Parameter Text dialog box.

Filtering PSMs with the Peptide and Protein Filter Node

You can also filter PSMs by using the Peptide and Protein Filter node in the consensus workflow. This node filters by score for the type of score specified. For information on this node, refer to the Help.

Filtering Proteins

You can filter proteins with the Peptide and Protein Filter node. For information on this node, refer to the Help.

When you place the Display Settings node in the consensus workflow, the Proteins page displays only master proteins by default. If you want to filter the proteins on the Proteins page with a filter set other than the default, follow the procedure in To apply stored filters to a .pdResult file.

Filtering Phosphorylation Site Probabilities

Filtering phosphorylation site probabilities involves both display filters and the ptmRS node in the processing workflow. For information on filtering phosphorylation modification data, refer to the Help.

Filtering with Display Filters

You can use the View > Display Filter command or the Display Filter icon, ▽ , to display only those items that meet a particular condition that you define. Refining your search results in this way reduces the complexity of the displayed results, allowing for a quicker and more productive analysis. You can sort and filter your results by a number of criteria, such as charge state, modifications, or even peptide probability. You can also create and apply multiple filters to your search results. You can save such filter sets to disk and load them in into the same or other report files.

- Filtering the Data on a Results Report Page

- Example 1

- Example 2

- Example 3

- Example 4

- Filtering with Filter Sets

- Using Display Filters to Filter Numerical Values by a Specified Precision Value

- Finding Common or Unique Proteins in Multiple Searches

- Applying Filters Specific to Different Searches in Multiconsensus Reports

Filtering the Data on a Results Report Page

You can use the application's display filters to refine the results in the .pdResult file.

❖ **To filter the data on a .pdResult data page**

1. Open a .pdResult file, and click on the tab of the page that you want to view.

2. Choose **View > Display Filter**, or click the **Display Filter icon**, ![filter icon] .

   The dockable Display Filter pane opens at the top of the results file, as shown in .

Figure 152. Display Filters pane
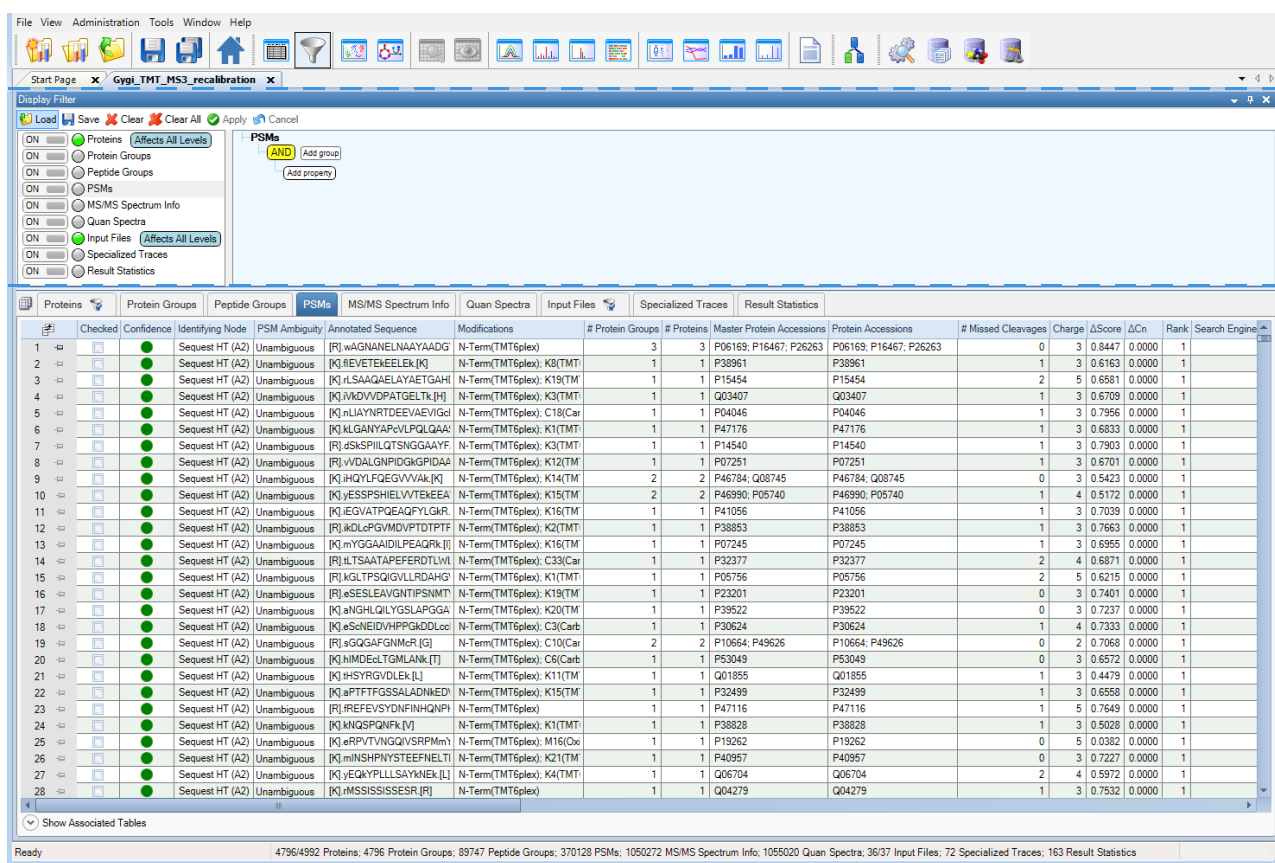
The left pane of the filter area lists the names of the filter groups. These names correspond to the names of the pages in the results report, such as Protein Groups, Proteins, PSMs, Decoy Peptide Groups, Quan Spectra, and so forth.

For convenience, the application automatically highlights a filter group when you navigate though different pages in the results report.

3. Select the filter category, for example, **PSMs**.

4. Set the first condition:

   a. Click the Add Property box in the right pane, which has a blue background, and select the property that you want from the menu.

   An operator box and a values box appear to the right of the element, as shown in Figure 153. The Apply button is now highlighted.

   Figure 153. Elements of the Display Filter pane

   Temporary enable/disable filter buttons

   Remove box

   Values box

   Operator box

   Property box
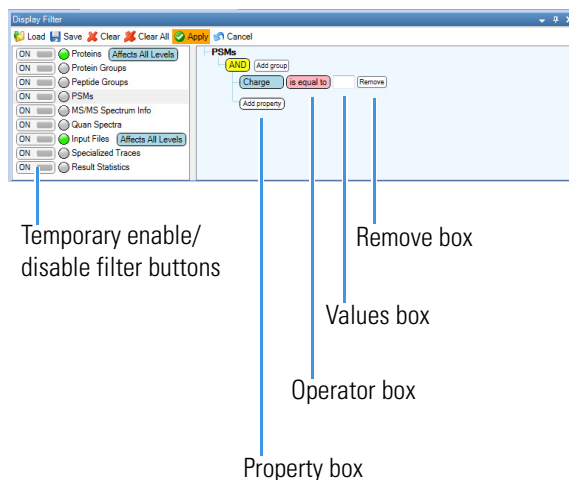
   b. Click the operator box, and choose an operator for the condition from the menu.

   For numerical values, the default is Is Equal To. For textual values, the default is Contains.

   c. In the values box, type a value.

   You can use simple wildcards to filter for specific patterns in strings. Choose the Contains Pattern operator and type a question mark (?) in the values box for any single character or an asterisk (*) for a sequence of arbitrary characters of any length.

5. (Optional) Set the additional conditions, if applicable, by clicking in the Add Property box and repeating the previous steps.

   Note Multiple conditions are arranged as siblings of an And box, which means that both conditions must apply to display a particular item. Change this box to Or if only one of the conditions applies.

6. (Optional) To add another filter, click **Add Property**.

   Another level of filtering appears above the original level, and a new Add Property box appears under the original, as shown in Figure 154.

Figure 154. Display Filter pane with two filters
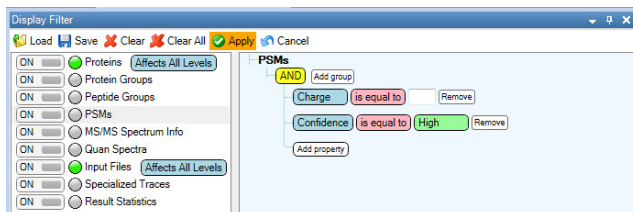
7. From the menu to the right of the selected filter group, choose the level that you want the filter to apply to, as shown in Figure 155:

- **Affects First Level Only**: Applies the filter to the items on a page, but not to the items on any associated pages.

- **Affects All Levels**: Applies the filter to both the items on a page and to the items on any associated pages at all levels.

Figure 155. Selecting the level of items to apply the filter to
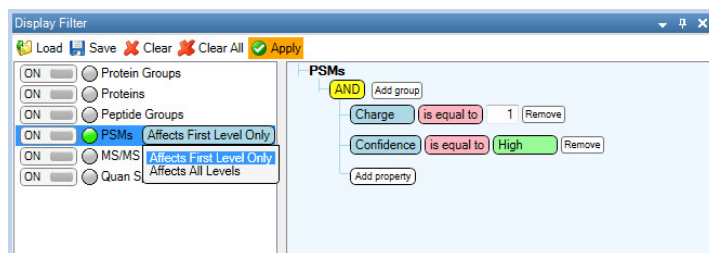
8. To apply the filters in the filter set to the results report, click ✅ Apply .

For examples showing how to build increasingly complex filters, see "Example 1" on page 232, "Example 2" on page 234, and "Example 3" on page 235.

❖ **To temporarily disable filters (method 1)**

1. In the left pane of the Display Filter pane, click ON next to the filter category that you want to disable temporarily.

   The icon changes to an Off icon.

2. Click ✅ Apply .

   Using the example in Figure 155, Figure 156 shows the results when the PSM filter is temporarily disabled. The filters for Protein Groups, Proteins, Peptide Groups, MS/MS Spectrum Info, and Quan Spectra are still enabled. The disabled filter is denoted by the gold circle.

3. To re-enable the filter categories, click the **On/Off** icon so that On appears and click the **Apply** icon.

Figure 156. Temporarily disabled Protein Groups and PSMs filters
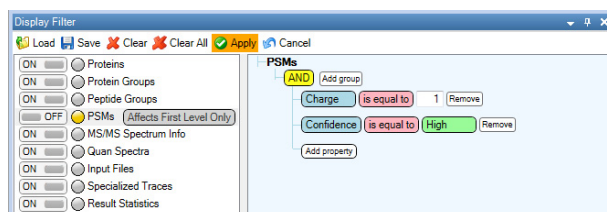
❖ **To temporarily disable filters (method 2)**

1. On the tab of a page displaying the or the filter icon (see Figure 157), click the icon.

Figure 157. Tab with filter icon
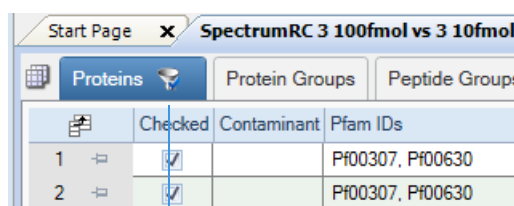
Filter icon on page tab

The application temporarily disables the active filter for the page, as indicated by the disabled filter icon on the tab, and the application displays the filtered-out rows.

2. (Optional) Click the icon to redisplay the filtered data.

Clicking on a filter icon is a faster way of temporarily disabling active filters than using the method described in "To temporarily disable filters (method 1)" on page 228. The ability to easily switch back and forth between filtered and unfiltered views of the data is useful during data review.

Temporarily disabling the row filters and displaying all rows affects only the contents of the page itself. It has no effect on the defined row filters themselves or on other views, charts, or pages. It has no effect on the contents of any exported CSV and Excel files.

Table 12 explains the different filter icons that appear on the page tabs.

Table 12. Filter icons

Icon	Meaning
-	No filter is defined for this page.
	A filter is defined, and all rows are visible, including the filtered-out rows.
	A filter is defined, but at least one row is currently filtered out.
	A filter is defined but is currently disabled for this table, and all rows are visible.

❖ **To clear conditions and filters**

- To remove a condition, click **Remove** next to it.

- To clear the filter on the currently selected page, click ✖ Clear, and then click ✔ Apply .

  For example, if the PSMs page is currently selected, the Clear icon clears only the filter on the PSMs page.

- To clear the filters on all pages of the open results report, click ✖ Clear All , and then click ✔ Apply .

❖ **To save a filter set**

1. Click 💾 Save to save the filter set in a file on the disk.

2. In the Export Filter dialog box, browse to the file to save the filter set in, or type the name in the File Name box.

3. Click **Save**.

   The file has a .filterset suffix.

When you save the .pdResult file containing the saved filter set and reopen it, the Filter icon, 🔻, appears on the tabs of the pages that have filters set so that you know immediately that the results on that page are filtered. In Figure 158, the results on the PSMs page are filtered.

Figure 158. .pdResult file containing filtered results for the PSMs page

PSMs tab indicating that the
page contains filtered results
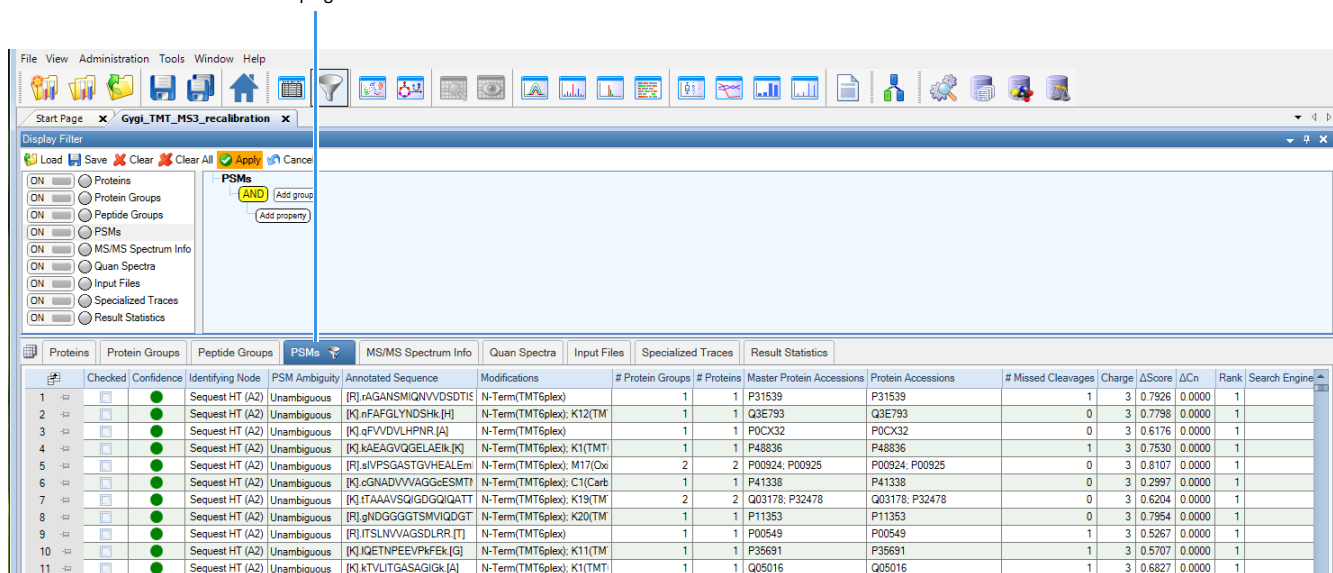
❖ **To load a filter set**

1. Choose **View > Display Filter**.

   The Display Filter pane opens, as shown in Figure 152 on page 226.

2. Click ⊡ 🗀 Load... .

3. In the Load Filter Set dialog box, browse to the filter set (*name*.filterset) file that you want to load, or type the name of the file in the File Name box.

4. Click **Open**.

   The selected filter set now appears in the Display Filter pane.

   When you load filter sets that do not apply to the type of pages that the report contains, the application still preserves the filters but makes them unavailable. When you load filter sets with conditions that refer to unavailable properties (columns) in the report, application shows the condition as unavailable with a line through it.

   For examples of inapplicable but preserved filters, see "Example 4" on page 238.

5. To remove conditions that are temporarily disabled, click **Remove** next to the condition. To remove all such unused conditions from all filters, click **Remove Unused Items**.

❖ **To view the filter settings in the Result Summaries**

1. Open a .pdResult file.

2. Choose **View > Summaries**.

3. In the Result Summaries pane, click the **Validation** tab.

4. Click the **Filters** tab in the upper right corner of the Validation page.

   The filters that you set in the Display Filter pane are reflected in the Applied Display Filters area of the Filters and Counts page of the Validation page, as shown in Figure 159.

Figure 159. Filters shown in the Display Filter pane and in the Result Summaries pane
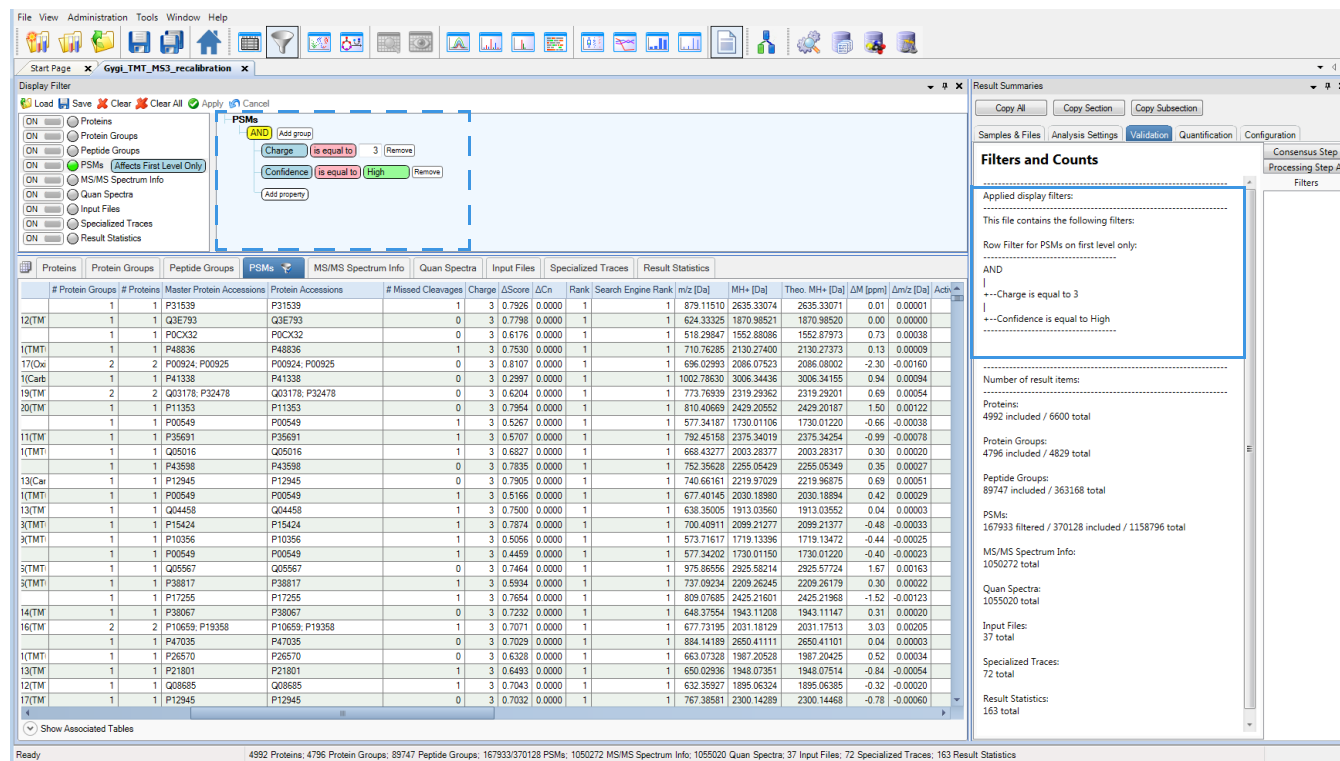
Following are some examples showing how to use display filters.

Example 1

Suppose that you would like to view only PSMs that are associated with a protein that contains the number "Q03771" in their corresponding protein accessions.

1. Open a .pdResult file.

2. Click the **Display Filter** icon, ⧉ , on the main toolbar.

3. In the left pane of the Display Filter pane, select **PSMs**.

4. In the right pane of the window, click in the Add Property box and select **Protein Accessions**.

5. Click in the operator box, and choose **Contains** if it is not already selected.

6. In the values box to the right of the operator box, type **Q03771**.

   The Display Filter pane now resembles Figure 160.

Figure 160. Filtering protein accessions
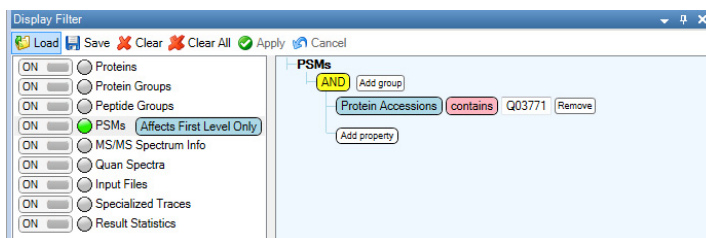
7. Click Apply .

Figure 161 shows only entries that contain Q03771 in their protein accessions.

Figure 161. Entries on the PSMs page that contain Q03771 in their protein accessions

Protein accessions containing Q03771

Example 2

The following example extends the filter defined in Example 1 by adding a second condition that requires a top search rank.

1. In the Display Filter pane shown in Figure 160 on page 233, click in the Add Property box, and select **Search Engine Rank**.

2. Leave the default **Is Equal To** setting in the operator box.

3. In the values box to the right of the operator box, type **1**.

   The Display Filter pane now resembles Figure 162.

   Figure 162. Search Engine Rank added as a second filter
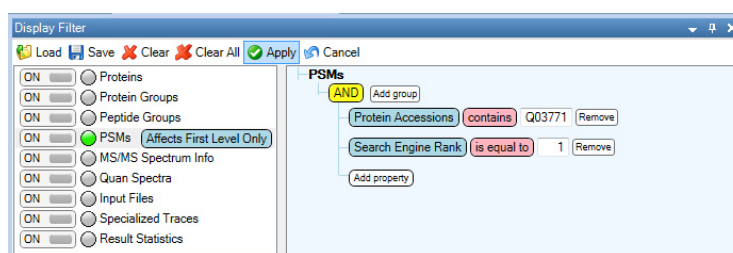
4. Click Apply .

   Figure 163 shows the results of filtering PSMs by protein accessions that contain the word "Q03771" and a search engine rank of "1."

Figure 163. Entries on the PSMs page that contain Q03771 in their protein accessions and 1 in their search engine rank

Search engine rank of 1

Both conditions must be met for a data row to pass the filter because they are connected by the AND Boolean operator. If you want only one condition to apply, click in the Boolean operator box, which displays AND in Figure 163, and select **OR**.

Example 3

Suppose that you want to build a more complex filter from the examples shown in Example 1 and Example 2. The new filter corresponds to "protein accession contains Q03771" and "search rank is equal to 1" or "protein accession contains P14540."

1. In the Display Filter pane shown in Figure 162, click **Add Group**.

   The Display Filter pane displays a new group, as shown in Figure 164.

Figure 164. Adding a group

Added group
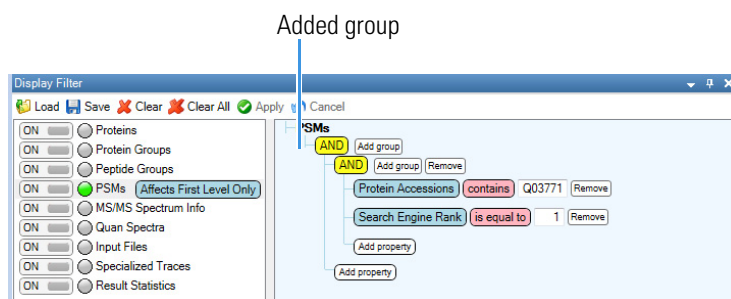
2. In the Boolean operator box of the new condition, select **OR**, as shown in Figure 165.

Figure 165. Selecting the OR Boolean operator
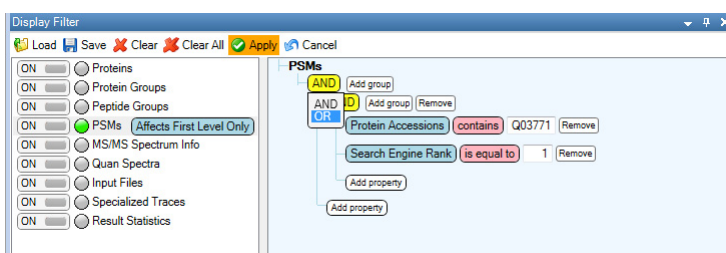
3. Click in the Add Property box at the lowest level, as shown in Figure 166, and select **Protein Accessions.**

   You must add the new condition to the outer level corresponding to OR rather than to the inner level corresponding to AND.

4. In the operator box, select **Contains**.

5. In the values box to the right of the operator box, type **P14540**.

   The Display Filter pane resembles Figure 166.

Figure 166. OR condition added
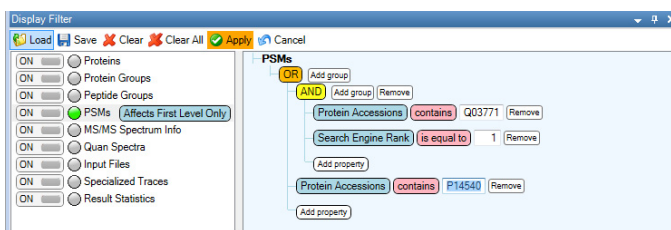
6. Click .

   Figure 167 shows the results of the filtering in the Proteins column. In this case, all proteins were filtered out except proteins P14540 and Q03771.

Figure 167. Entries on the PSMs page that contain Q03771 in their protein accessions and 1 in their search engine rank or contain P14540in their protein accessions

Protein Accessions column

Example 4

You can save and load filter sets to and from the disk, even though they might not entirely apply to another report. As an example, assume that you create a filter set that only displays protein groups with more than four associated proteins and proteins that are connected to at least ten identified peptides. Figure 168 shows these filters.

Figure 168. Specifying protein group filters and protein filters
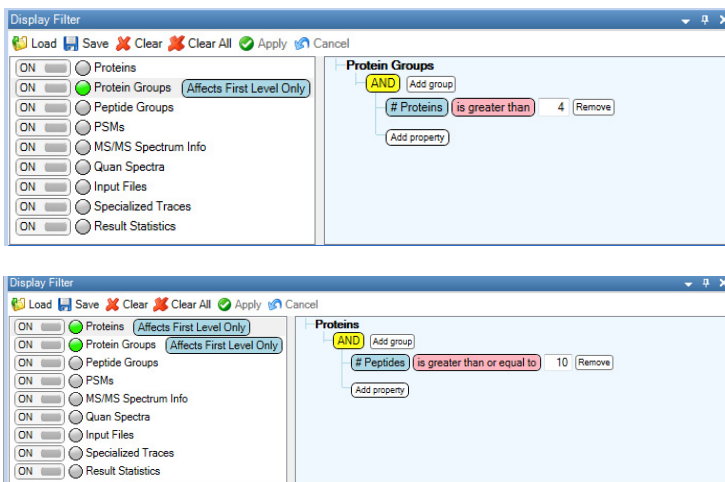
When you load this filter set into a different results report that, for example, does not contain the Protein Groups page, the protein groups filter is still preserved but is unavailable, as shown in Figure 169.

Figure 169. Preserved but unavailable protein groups filter
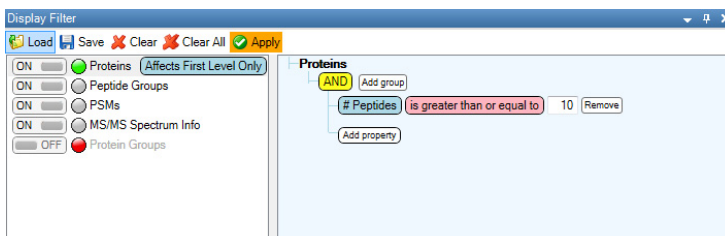
When you select the unavailable filter set, the message shown in Figure 170 appears.

Figure 170. Selecting the preserved but unavailable protein groups filter
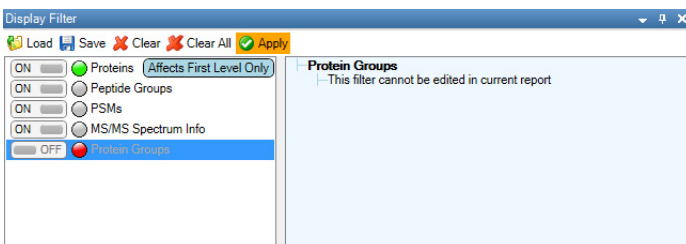
To remove all filters that refer to unavailable pages, click ✖ Clear .

Now suppose that you define the filter for the heavy-light ratio of a quantification report that only shows ratios above 2.5. Figure 171 shows this filter.

Figure 171. Specifying a quantification filter
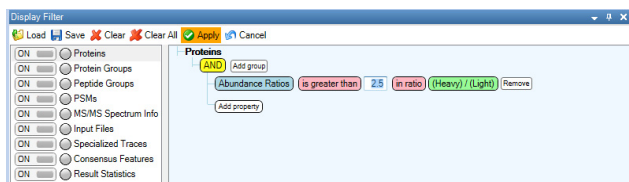
When you load this filter into a report that does not contain quantification information and therefore has no Heavy/Light column, this condition is unavailable, so it appears with a line drawn through it, as shown in Figure 172.

Figure 172. Loading a quantification filter into a report containing no quantification data
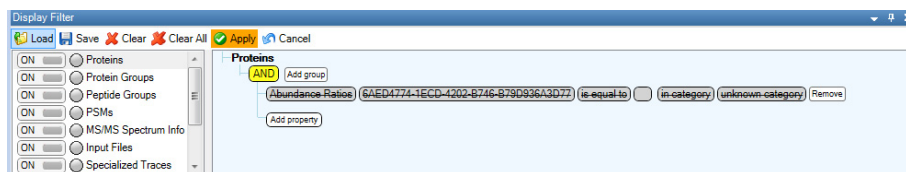
The application still preserves such conditions in the internal representation of the filter but ignores them for the current report. You can remove a temporarily disabled condition by clicking the Remove button next to the condition. To remove all such unused conditions from all filters, click ✖ Clear .

Filtering with Filter Sets

You can save a set of your favorite display filters as the default filters of a .pdResult file. You set the display filters and save them to a .filterSet file, and then apply them in a newly generated .pdResult file by using the Display Settings node in the consensus workflow.

Saving a Filter Set in a .pdResult File

❖ **To save a filter set in a .pdResult file**

1. Open the .pdResult file and set the filters in the Display Filter pane, following the instructions in "Filtering with Display Filters" on page 225.

2. Save the filter set as described in "To save a filter set" on page 230.

3. (Optional) Choose **File > Save**.

   This command saves the current layout and filter set in the .pdResult file and in the .filterSet file. It also saves the filter set in the following directory in the Common Templates area:

C:\Users\Public\Public Documents\Thermo\Proteome Discoverer 2.1\
Common Templates\FilterSets

You can apply this filter set as a template when you generate other .pdResult files by using
the Display Settings node.

Applying Stored Filters to a .pdResult File

Use the following procedure to apply stored filters to a .pdResult file or to temporarily
override the display of only master proteins on the Proteins page of the .pdResult file.

❖ **To apply stored filters to a .pdResult file**

1. In an open study, create a consensus workflow. For instructions, see "Incorporating an
   Existing Workflow into a Study and an Analysis" on page 99.

2. Drag the Display Settings node to the Post-Processing Nodes pane.

3. Click the **Display Settings** node.

4. Click the **Browse** button (...) next to the Filter Set parameter.

5. The Parameter Text for Filter Set (Read Only) dialog box opens, as shown in Figure 173.

   > **Note** The default filter set in this dialog box displays only master proteins by default
   > on the Proteins page.

   Figure 173. Parameter Text for Filter Set (Read Only) dialog box
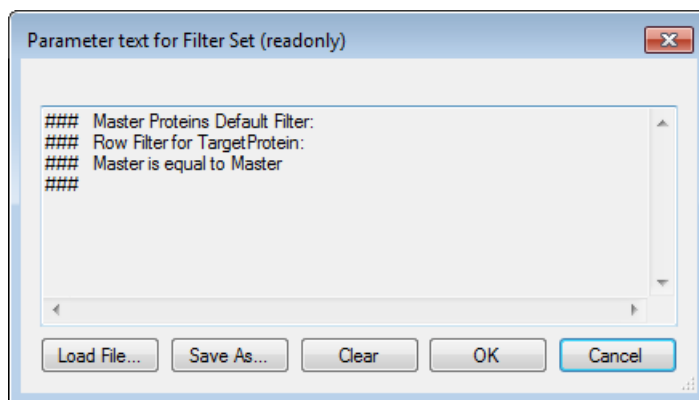
6. Click **Load File**.

7. In the Open dialog box, browse to the .filterSet file that you want to load, and click
   Open.

8. In the Parameter Text for Filter Set dialog box, click **OK**.

   The new filter set replaces the default filter set.

9. Run the consensus workflow.

10. (Optional) Open the .pdResult file that contains the filter set.

Changing the Default Filter for the Proteins Page

By default, the Proteins page displays only master proteins. If you want the Proteins page to display a different set of proteins by default, follow these steps.

❖ **To change the default filter set for the Proteins page**

1. Choose **Administration > Configuration**.

2. Click the **Browse (...)** button next to the Default Filter Set parameter.

3. In the Parameter Text for Filter Set (Read Only) dialog box, shown in Figure 173, click **Load File**.

4. In the Open dialog box, browse to the new default .filterSet file that you want to load, and click **Open**.

5. In the Parameter Text for Filter Set dialog box, click **OK**.

   The new default filter set replaces the factory default filter set.

Using Display Filters to Filter Numerical Values by a Specified Precision Value

In reporting, the application displays most numerical values with a significantly lower precision than it uses to store the values internally. The display filters can filter with high precision, but you might not want to choose such precision during filtering. In addition, the precision might be data-dependent; for example, filtering the delta mass requires greater precision than is required in filtering probabilities.

If an operator tests for the equality of two values, you can control the numerical precision with which the application filters data that are represented as floating-point values. The filter automatically extends any value that you enter to at least two digits for the fractional part of the value.

To use a higher precision, enter a value with more digits. The example in Figure 174 uses a precision of five digits (3.22528) in the filter. The filter for such values automatically rounds the internal data before filtering according to the precision of the given value.

Figure 174. Automatic rounding when filtering high-precision data
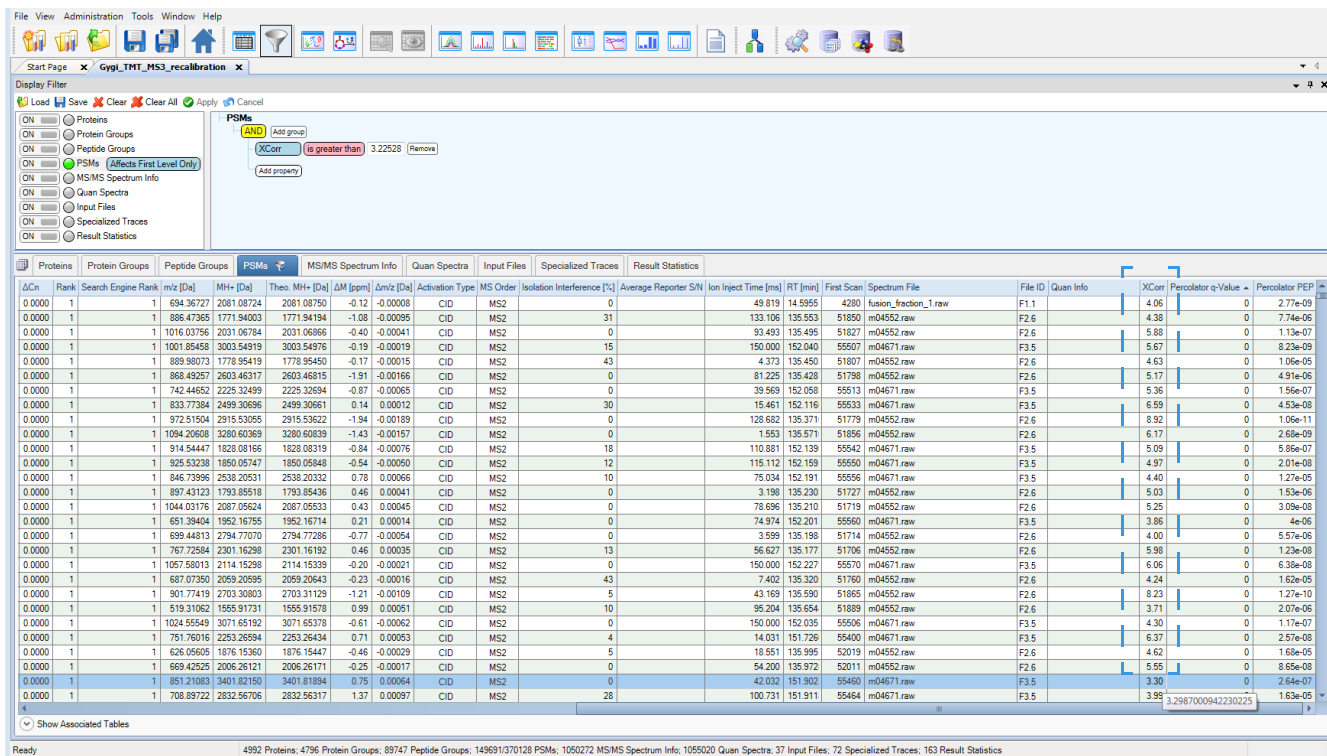
Filtering numerical data inherently involves a precision problem. Numbers categorized as "double" or "float" follow the IEEE 754 format, which specifies that they should always be compared to a tolerance interval, which is effectively equivalent to rounding. This comparison is especially important in the Proteome Discoverer application, which stores data internally with full precision but rounds them for user convenience in the displayed pages. For example, although the data shown in the XCorr column in Figure 174 is only displayed with two digits of precision (see value 3.30), the internal precision of the value is higher (3.2987000942230225), as shown in the ToolTip.

Finding Common or Unique Proteins in Multiple Searches

A multiconsensus report can help you find common or unique proteins between two or more searches. You can use the display filters of the Proteins, PSMs, or MS/MS Spectrum Info page to automate the task of creating a list of all proteins unique to one search or search node or a list of all proteins common to all searches or search nodes.

❖ **To find proteins common to two or more searches**
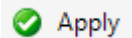
1. Choose **View > Display Filter**, or click the **Display Filter** icon, ▼ .

2. Select **Proteins** in the left pane of the Display Filters pane.

3. In the right pane, click in the Add Property box and select **Coverage** *first_search_engine*, for example, **Coverage Mascot**.

4. In the operator box, select **Is Greater Than**.

5. In the values box, select **0**.

6. Click in the Add Property box again, and select **Coverage *second_search_engine***, for example, **Coverage Sequest HT**.

7. In the operator box, select **Is Greater Than**.

8. In the values box, select **0**.

   The filter in the Display Filters pane now resembles Figure 175.

   Figure 175. Filtering to select common proteins
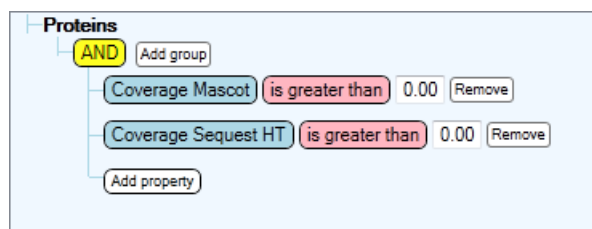
9. For each additional search engine, repeat step 3 through step 5 earlier in this procedure.

10. Click Apply .

    The Coverage *Search_engine* columns in the results report display the proteins found by both searches. The Coverage Mascot and Coverage Sequest HT columns display these proteins in Figure 176.

Figure 176. Multiconsensus report showing proteins common to Mascot and Sequest HT search in results report
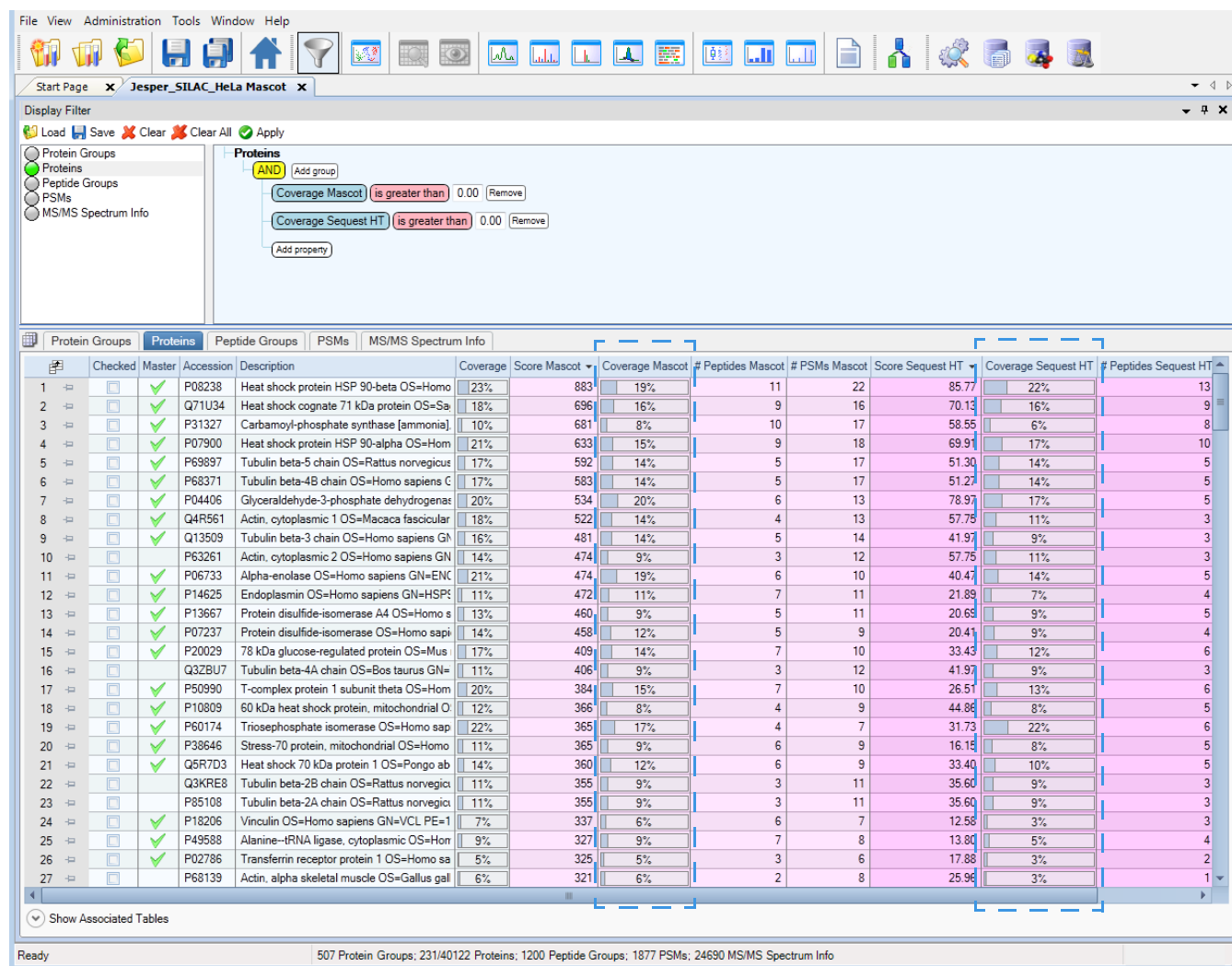
❖ **To find proteins unique to each search**

1. Choose **View > Display Filter**, or click the **Display Filter** icon, 🔻 .

2. Select **Proteins** in the left pane of the Display Filter pane.

3. In the right pane, click in the Add Property box, and select **# PSMs** *first_search_engine*, for example, **# PSMs Mascot**.

4. In the operator box, select **Is Not Set**.

   The filter in the Display Filter pane now resembles Figure 177.

Figure 177. Filtering to select unique proteins
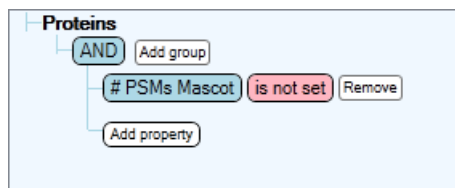
5. Click ✅ Apply .

Zeros in the Coverage *Search_engine* column in the results report indicate the proteins found only by the selected search. Figure 178 displays the proteins unique to the search with Mascot.

Figure 178. Multiconsensus report showing proteins unique to the Mascot search

You can repeat the procedure using # PSMs *Search_engine* to find the proteins unique to the *Search_engine* search.

In general, if one search or search node uniquely identifies a protein, the number of identified peptides for this protein is non-zero for this search or search node but zero in all other searches or search nodes. On the other hand, if a protein is common to two or more searches or search nodes, the number of identified peptides is non-zero for all considered searches or search nodes.

Applying Filters Specific to Different Searches in Multiconsensus Reports

You cannot apply filters specific to different searches in multiconsensus reports, but you can apply filters that are specific to the different search engines used (Sequest HT and Mascot). The available score-based filters automatically change the list of scores that you can select for filtering, depending on the search engines used for the loaded reports.

For a Sequest HT search report, you can apply the XCorr and Charge display filters, as shown in Figure 179.

Figure 179. XCorr and Charge filters applied to Sequest HT multiconsensus reports
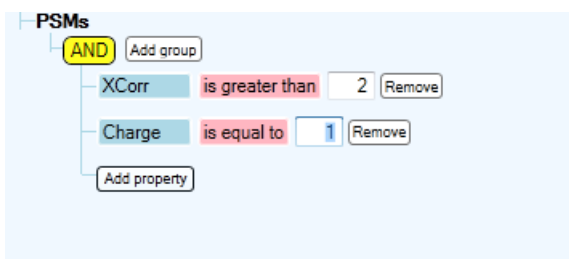
For a Mascot search report, you can apply the Ions Score display filter, as shown in Figure 180.

Figure 180. Ions Score filter applied to Mascot multiconsensus reports
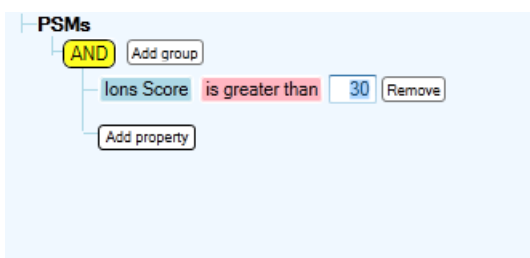
Validating Results

This chapter explains how to validate your results by calculating false discovery rates.

Contents

The false discovery rate (FDR), or the false positive rate, is a statistical value that estimates the number of false positive identifications among all identifications found by a peptide identification search. It is a measure of the certainty of the identification. You can use the Proteome Discoverer decoy database search feature to determine FDRs.

You can use FDRs to validate MS/MS searches of large data sets, but they are not effective on searches of a small number of spectra or searches against a small number of protein sequences, because the number of matches will likely be too small to give a statistically meaningful estimate.

A decoy database strategy gives a probability value to identifiers and the percentage of false discoveries that you can expect. A one percent false discovery rate is a typical target for searches.

A good decoy database should contain entries that resemble real proteins but do not contain genuine peptide sequences. The simplest approach to obtaining such a decoy database is to reverse all protein sequences, which is the scheme that the Proteome Discoverer application uses by default. However, you can request that the search be performed against the concatenated database, when required. It is a suitable approach for enzymatic MS/MS searches.

> **CAUTION** Reversing the database is not suitable for peptide mass fingerprinting or no-enzyme MS/MS searches, especially for dynamic modifications. You might see mass shifts at each end of a peptide sequence that transform a genuine y-series match into a false b-series match or vice versa.

You can perform the decoy database search in two ways:

- Perform two separate searches, one against the non-decoy database and one against the decoy database. Then count the number of matches from both searches to determine the false discovery rates. This approach is the more conservative approach.

- Create a concatenated database from the non-decoy and the decoy database and then perform the search against this concatenated database.

The difference between the two approaches becomes clear when you find two significant matches for a given spectrum. The first match is from the non-decoy database, and the second one is from the decoy database. Because the Proteome Discoverer application application considers only the top matches when calculating the FDRs, finding two significant matches for a given spectrum is not considered a false positive in the concatenated database approach, but it counts in the separate databases approach. The latter case is more conservative and is the default approach that the application currently uses. You can select the validation methodology by setting the Concatenated FDR Calculation parameter of the Peptide Validator node.

To calculate the FDR, the application counts the matches that pass a given set of filter thresholds from the decoy database and from the non-decoy database. It counts only the top match per spectrum, assuming that for any given spectrum only one peptide can be the correct match.

For detailed information on how the application calculates the FDRs for PSMs, peptide groups, proteins, and protein groups, refer to the Help.

Target FDRs

When you set an FDR target value for a decoy database search, the application determines and applies filter thresholds to identified matches so that the resulting FDR is not higher than the set target value. The application distributes the confidence indicators applied to each peptide match according to these calculated filter thresholds.
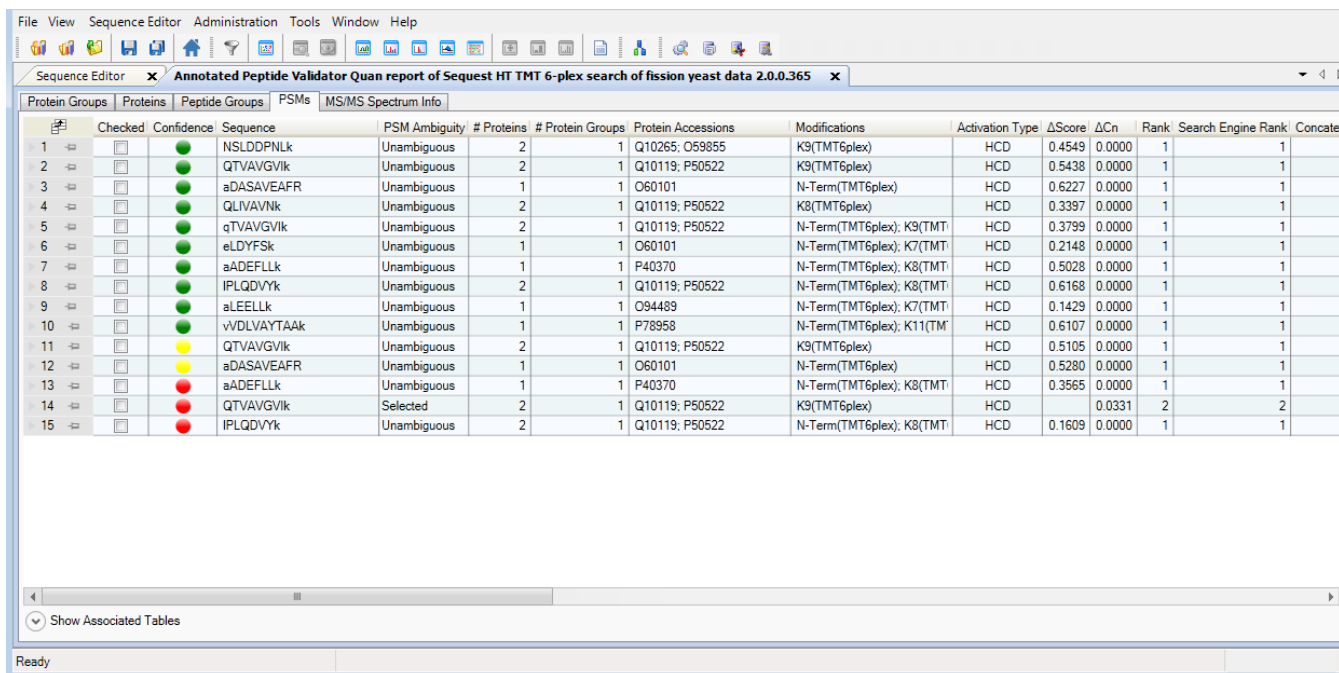
You must specify two target values for a decoy database search: a strict target FDR and a more relaxed FDR. Figure 182 on page 250 shows the decoy search setting with target FDRs of one percent and five percent, respectively. After completing the search, the system automatically determines two sets of filter settings so that the resulting separate FDRs do not exceed their corresponding target value.

Peptide Confidence Indicators

The application uses the filter settings that determine FDRs to distribute the confidence indicators for the peptide matches; these are the green, yellow, and red circles attached to each peptide match. Whenever you perform a decoy database search during the database search and apply filter settings to achieve the specified target FDRs, the same filters are used to

distribute the confidence indicators. Peptide matches that pass the filter associated with the strict FDR are assigned a green high-confidence indicator, peptide matches that pass the filter associated with the relaxed FDR but do not pass the filter associated with the strict FDR are assigned a yellow medium-confidence indicator, and all other peptide matches receive a red low-confidence indicator. Figure 181 gives an example of these confidence indicators.

Figure 181. Decoy database search results

Setting Up FDRs

You can set up FDRs with any of the following nodes in the workflow.

To validate the identifications from different search engines, the Proteome Discoverer application application offers the Percolator node and the Target Decoy PSM Validator node in the processing workflow. Percolator is a superior validation algorithm that uses a machine learning approach, but it requires a sufficient number of target and decoy matches that are not always available. In these cases, you can use the Target Decoy PSM Validator node. This node triggers a target and decoy search and calculates score thresholds to achieve the specified target false discovery rate (FDR). The derived score thresholds for the strict and relaxed FDR separate the identified PSMs into high-, medium-, and low-confidence identifications.

Figure 182 illustrates the basic processing workflow to use for setting up a decoy database search. For more information on creating a processing workflow, see "Creating a Processing Workflow" on page 68.

> **Note** In this workflow, you can use the Percolator node in place of the Target Decoy PSM Validator node.
>
> To work properly, Percolator needs a sufficient number of PSMs from the target and the decoy search. If the search identified fewer than 200 target or decoy PSMs, or if fewer than 20 percent decoy PSMs are available compared to the number of target matches, Percolator rejects them for processing and displays an appropriate message in the Proteome Discoverer job queue or in the Search Summary of an open report.

Figure 182. Basic processing workflow for setting up a decoy database search
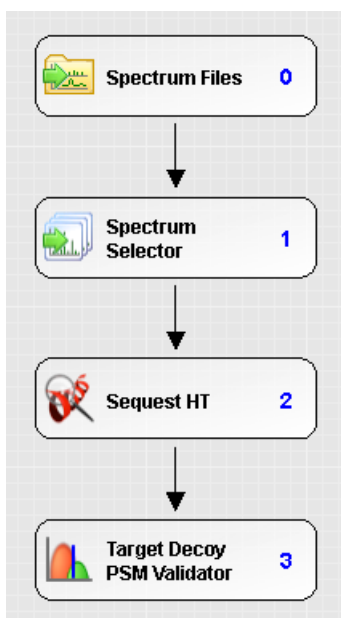
Figure 183 illustrates the basic consensus workflow to use for setting up a decoy database search. For more information on creating a consensus workflow, see "Incorporating an Existing Workflow into a Study and an Analysis" on page 99.

Figure 183. Basic consensus workflow for setting up a decoy database search
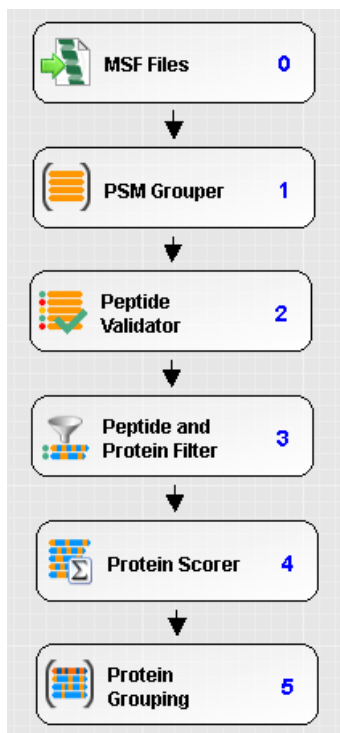
Calculating FDRs for PSMs, Peptide Groups, Proteins, and Protein Groups

This topic directs you to sources of detailed information on calculating FDRs for PSMs, peptide groups, proteins, and protein groups.

Calculating False Discovery Rates for PSMs

Use the Percolator node to calculate the FDRs for PSMs. For information on this node, refer to the Help and "Filtering PSMs with the Percolator Node" on page 219.

Calculating False Discovery Rates for Peptide Groups

Use the quality algorithm in the Percolator node to calculate the FDRs for peptide groups. For information on this node, refer to the Help.

Calculating False Discovery Rates for Proteins

Use the Protein FDR Validator node to calculate the FDRs for proteins. For detailed information on this node, refer to

Grouping Peptides and Proteins

This chapter describes how the Proteome Discoverer application groups proteins and peptides.

Contents

Grouping Proteins

The application groups proteins through the Protein Grouping node in the consensus workflow. This topic explains the algorithm that it uses and the workflow that you should use to group proteins.

Master Proteins

For information on master proteins, see the Help.

Protein Grouping Algorithm

The application uses a protein grouping inference process to group proteins when you use the Protein Grouping node in the consensus workflow. Figure 184 shows the steps involved in this process.

The application initially collects all peptide spectrum matches (PSMs) that were not filtered out by the Peptide and Protein filter node.

Figure 184. Protein grouping inference process in the Proteome Discoverer application
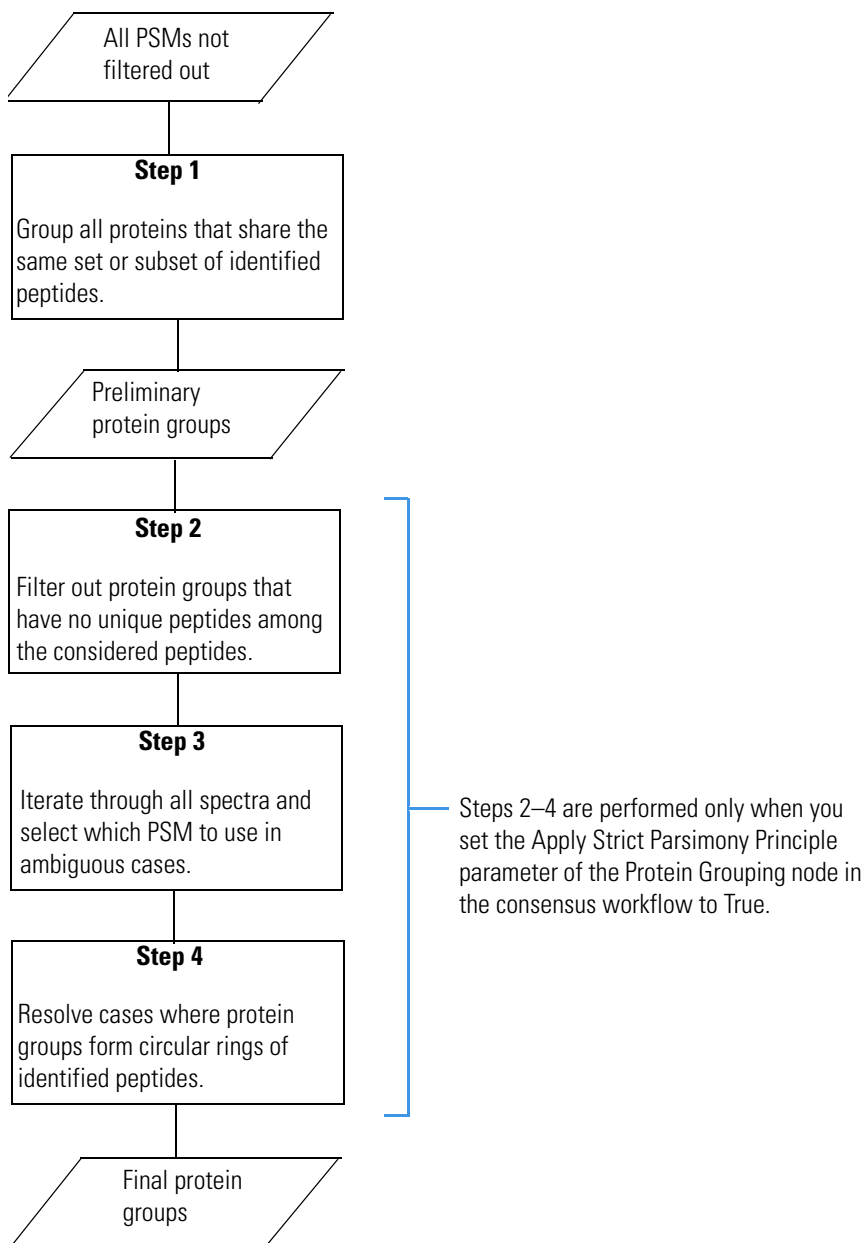
Step 1: The application creates preliminary protein groups from the PSMs collected. It combines all proteins into one protein group that contains the same subset of peptides.

The application takes the next steps in the protein grouping process if you set the Apply Strict Parsimony Principle parameter of the Protein Grouping node in the consensus workflow to True.

Step 2: The application removes all protein groups that have no unique peptides among the peptides that it considers for the protein grouping process. If a protein group does not contain at least one unique peptide, the application also includes all of the peptides included by other protein groups, so there is no supporting evidence for the existence of this protein group. At this point, the application explicitly retains all protein groups that form circular rings of overlapping shared peptides. For example, suppose a circular ring comprises these protein groups:

- ABCD (identified by peptides a, b, c, and d)

- CDEF (identified by peptides c, d, e, and f)

- EFAB (identified by peptides e, f, a, and b)

To explain all identified peptides, only two of the three protein groups are needed, but at this point it is not clear which to take and which to reject. The application postpones the resolution of this issue until step 5.

Step 3: The application first collects all spectra with more than one peptide match to consider for the protein grouping process. It then resolves these ambiguous cases and selects one of the PSMs to use for the protein grouping process while rejecting the remaining peptide matches of a spectrum. In cases where the application considers more than one PSM for a spectrum, it resolves this ambiguity by selecting the PSM that is connected to the "best" protein group and rejecting the other PSMs. The "best" protein group is the group with the highest number of unambiguous and unique peptides.

Step 4: The application resolves the cases where protein groups form circular rings of overlapping identified peptides. This step is the last step of the protein group inference process, resulting in the final list of protein groups that the application reports in the Proteins page of the .pdResult file.

The PSM Ambiguity column on the PSMs and MS/MS Spectrum Info pages can help you understand the process of selecting PSMs for the protein group. This column is available for every PSM, every search input entry (representing the searched spectra), and every peptide group. For the search input entries and the peptide groups, this column displays the best PSM ambiguity from all connected PSMs. Refer to the Help for a description of the categories of ambiguity in this column.

In the example shown in Figure 185, the application identifies eight different PSMs for search input 20. Even though only seven PSMs ranked 1 through 7 are of high confidence, all eight PSMs meet the specified protein grouping criteria because, on the basis of user-specified criteria in the Peptide and Protein Filter node, all PSMs for the top-scored proteins were retained. However, because the search input cannot be unambiguously assigned to a single protein, the PSM ambiguity is set to Ambiguous.

Figure 185. PSMs shown for search input

Protein Grouping Workflow

To group proteins, use the Protein Grouping node in the consensus workflow. Figure 186 shows the basic workflow to use to group proteins with the Protein Grouping node.

Figure 186. Basic workflow for grouping proteins
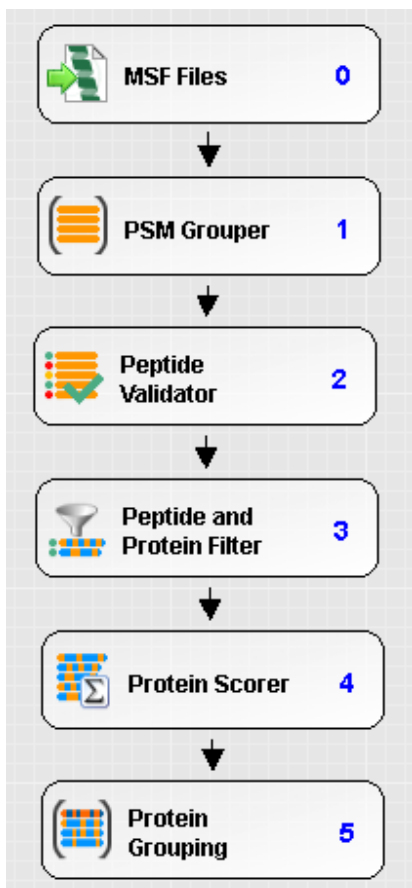
For information on the parameters of the Protein Grouping node, refer to the Help.

Number of Unique Peptides Column on the Proteins Page

The application counts the number of peptides that are only contained in a protein group and displays the value in the # Unique Peptides column on the Proteins page. The Proteome Discoverer application counts only the peptides that display a status of Selected or Unambiguous in the PSM Ambiguity column, because assessing the uniqueness of peptides that were not used to form protein groups has no relevance.

PSMs Identified by Multiple Workflow Nodes

In search results where the application identifies PSMs by multiple search nodes within a single workflow, the protein grouping algorithm selects one of the PSMs identified for the same spectrum for building the protein groups.

In search results where the application identifies PSMs by multiple search nodes from multiple workflows (multi-consensus report), it treats PSMs and spectra from the different workflows as being separate, even if it searched the same raw data files and therefore the same spectra. In this case, it is difficult to determine whether the application searched the exact same spectra, because they might have changed in the different workflows.

Grouping PSMs

You can group PSMs into peptide groups by using the PSM Grouper node. For information on this node, refer to the Help.

Obtaining Protein Annotation Information

This chapter explains how the Proteome Discoverer application retrieves annotation information from ProteinCenter. Annotation information is metadata for investigated proteins from available public databases and includes widely used annotations like the GO (Gene Ontology) terms for each protein, Pfam (Protein Families) classification of proteins, KEGG PATHWAY and Wiki Pathways database pathways, Entrez Gene identifiers, Ensembl gene identifications from the Ensembl reference genomes of genetically sequenced organisms, and UniProt information about post-translational modifications (PTMs).

Contents

ProteinCenter

ProteinCenter is a web-based application that you can use to download biologically enriched annotation information for a single protein. You can download the following types of information:

- Molecular functions, cellular components, and biological processes from the GO database

- Classification information for protein families from the Pfam database

- Biochemical pathway maps for many species from the KEGG PATHWAY database

- User-provided biochemical pathway maps and links between the pathways to the proteins from the Wiki Pathways database

- Gene identifications from the Entrez Gene database

- Genomic annotations of genetically sequenced organisms from the Ensembl genome database

- Post-translational modification information from the UniProt database

- Official symbol value from the Entrez Gene database

The data in ProteinCenter is updated biweekly.

You can also download FASTA databases from ProteinCenter (see "Downloading FASTA Files to Proteome Discoverer" on page 153).

The application gives you access to ProteinCenter in two ways:

- You can create a consensus workflow that includes the ProteinCenter Annotation node, which retrieves the types of annotation information just given from ProteinCenter and stores it in the Proteome Discoverer results files. The application displays this information in columns on the Proteins page of the .pdResult file. For information on setting up a protein annotation workflow to achieve these results, see "Downloading Files from ProteinCenter" on page 263 and "Creating a Protein Annotation Workflow" on page 265.

- The ProteinCard page available for each protein displays the annotation data available in ProteinCenter and displays it on a page of the Protein Identification Details view (see "Accessing the ProteinCard Page" on page 297). You can display this information for the following proteins:

  – Proteins on the Proteins page of the .pdResult file

  – Proteins associated with identified peptides

  You can access the ProteinCard page for each protein by double-clicking its row in the .pdResult report; or clicking its row, choosing View > Protein Details, and then clicking the ProteinCard tab of the Protein Identification Details view. The ProteinCard page itself is divided into separate pages representing different aspects of that protein: General, Keys, Features, Molecular Functions, Cellular Components, Biological Processes,

Diseases, and External Links. You can display a ProteinCard page for every identified protein whose accession is tracked in ProteinCenter. For information on the ProteinCard page, see "Accessing the ProteinCard Page" on page 297 and "ProteinCard Page Parameters" on page 298.

You can download organism-specific sequence databases from ProteinCenter. For instructions, see "Downloading FASTA Files to Proteome Discoverer" on page 153.

Gene Ontology (GO) Database Annotation

The Gene Ontology (GO) database is a collaborative effort, incorporating community input from database and genome annotation groups to address the need for consistent descriptions of gene products in different databases. The GO project has developed three structured, controlled vocabularies (ontologies) that describe gene products in a species-independent way.

biological processes

cellular components

molecular functions

Each gene ontology is divided into categories and subcategories called GO terms, which define the protein in more specific terms. For example, chloroplast, a term in the cellular component ontology, is subdivided as follows:

chloroplast

[p] *chloroplast* envelope

[p] *chloroplast* membrane

[i] *chloroplast* inner membrane

[i] *chloroplast* outer membrane

You can find more information on the GO Ontology website at www.geneontology.org/.

GO Slim annotations are a collection of annotations assembled by ProteinCenter from various sources such as the European Bioinformatics Institute (EBI) and the GO consortium. These annotations are displayed in the Molecular Function, Cellular Component, and Biological Process columns of the results report (.pdResult) file. In addition, you can define your own categories of GO Slim annotations.

Pfam Database Annotation

In addition to GO and GO Slim annotations, you can also use ProteinCenter annotations from the Pfam database at the Wellcome Trust Sanger Institute (WTSI) (http://pfam.xfam.org/). These are annotations of protein families, which are proteins with similar sequences and similar biological functions. A special sequence comparison algorithm called the Hidden Markov Model groups proteins into the families by comparing the sequences. Each family has its own identification number that starts with Pf … . The Proteins page of the .pdResult file displays this number in the Pfam IDs column. You can use the Pfam identification number to search in the Pfam database to obtain more details about the protein family.

The Pfam annotation system is an alternative to GO annotations. You might want to use the Pfam system to filter your proteins when you want the results to be traceable, scored, and uniformly grouped. You might also consider its computationally based data more reliable. However, it might be easier to use the hierarchy and grouping of the GO system to help you interpret results.

Table 13 compares the features of the GO and Pfam databases.

Table 13. Comparison of GO and Pfam features

GO features	Pfam features
Proteins grouped in biologically meaningful categories	Proteins grouped by similarity
Deep hierarchical order of terms	Few hierarchies
Data input by experts with different confidence levels and differing opinions	Computational data input with no human influence or expert knowledge

KEGG PATHWAY and Wiki Pathways Database Annotations

The Kyoto Encyclopedia of Genes and Genomes (KEGG) PATHWAY database contains biochemical pathway maps for many species. The Wiki Pathways database collects user-provided biochemical pathway maps and links the pathways to the proteins. The ProteinCenter Annotation node retrieves the pathway identifications and pathway names from ProteinCenter and displays them in separate columns of the .pdResult file when it annotates a result.

Entrez Gene Database Annotation

The database maintained by the National Center for Biotechnology Information (NCBI) includes the Entrez Gene database, which contains gene-specific information. Each gene stored in the Entrez Gene database has a unique identification. All proteins derived from the same gene have the same gene identification. When you use the Proteome Discoverer

application to retrieve the Entrez Gene identifications from ProteinCenter, the Proteins page of the results report displays these identifications in the Entrez Gene ID column. You can use this information to group or cluster together the proteins that are biologically meaningful. Because not all genes are stored in the Entrez Gene database, some proteins do not have a valid gene identification. In this case, the application displays a blank or empty cell in the Entrez Gene ID column on the Proteins page of the results file.

Ensembl Genome Database Annotation

The Ensembl genome database is a joint project of the European Bioinformatics Institute (EBI) and the Wellcome Trust Sanger Institute (WTSI). It stores and automatically annotates the reference genomes of the genetically sequenced organisms. Because this database focuses on completed gene-sequencing projects, it contains fewer species than the large protein databases like the UniProt and NCBI databases, which also collect data from partially sequenced organisms.

UniProt Database Annotation

From ProteinCenter, you can retrieve information on known PTMs from the UniProt database and compare it with information on found PTMs. For details on this feature, see "Viewing PTM Information on the Protein Identification Details View" on page 338.

Official Gene Symbol

The official Gene Symbol is published by HUGO Gene Nomenclature Committee (HGNC) for the gene that encodes the protein of interest. When the gene is from an organism well-documented by a major proteomics organization like the Human Genome Organization (HUGO), the gene symbol is considered official. Proteome Discoverer displays the official gene symbol whenever it is available. It also displays gene symbols that are available but that have not been recognized as official. If no symbol is available, it displays the primary gene key that is available for a given protein in the ProteinCenter database.

Downloading Files from ProteinCenter

You can download FASTA files, download protein annotations during report processing, and access the ProteinCard page from ProteinCenter. To download protein annotations from ProteinCenter, include the ProteinCenter Annotation node in the consensus workflow.

Configuring the ProteinCenter Web Server

You do not need to configure the Proteome Discoverer application to download FASTA files, to download protein annotations, or to access ProteinCard. The application automatically provides the URL during installation. However, if you need to use a server other than the default server, follow this procedure.

❖ **To configure the ProteinCenter Web server**

1. Choose **Administration > Configuration**.

2. In the Configuration pane of the Administration page, click **ProteinCenter**.

   The ProteinCenter Server area appears on the page, as shown in Figure 187.

   Figure 187. ProteinCenter Server area of the Administration page

1. ProteinCenter Server	
ProteinCenter URL	http://webservice.proteincenter.proxeon.com/ProXweb/
Number of attempts to submit the annotation request	3
Time interval between attempts to submit the annotation request [sec]	90
Timeout of the annotation request [min]	15

3. In the ProteinCenter URL box, type the URL of the ProteinCenter Web server to download the data to if you do not want to use the default shown in Figure 187.

   The Proteome Discoverer application connects to this server when you download files from ProteinCenter.

4. On the Number of Attempts to Submit the Annotation Request line, specify the number of times that the application tries to submit the search when the ProteinCenter server is busy.

   Default: 3

5. On the Time Interval Between Attempts to Submit the Annotation Request [sec] line, specify the interval of time, in seconds, that elapses between attempts to submit a search when the ProteinCenter server is busy.

   Default: 90

6. On the Timeout of the Annotation Request [min] line, specify the amount of time, in minutes, that elapses before the application ceases to try to access the ProteinCenter server when the server is busy.

   Default: 15

7. Click **OK**.

Downloading FASTA Files from ProteinCenter

To download FASTA files from ProteinCenter, see "Downloading FASTA Files to Proteome Discoverer" on page 153.

Downloading Annotations Associated with Identified Proteins

To download protein annotations from ProteinCenter, see "Creating a Protein Annotation Workflow."

Accessing the ProteinCard Page of the Report Item Distribution Chart

To access the ProteinCard page of the Report Item Distribution Chart, refer to the Help.

Exporting Files Downloaded from ProteinCenter

To export FASTA files that you downloaded from ProteinCenter, see "Exporting FASTA Files" on page 159.

Creating a Protein Annotation Workflow

You can retrieve annotations of all identified proteins from ProteinCenter by using the ProteinCenter Annotation node in the consensus workflow. This node can retrieve the types of annotations listed in "ProteinCenter" on page 260 and places it in the following columns in the .pdResult file:

- Gene Ontology (GO) annotations, which are displayed in the GO Accessions column of the results report (.pdResult) file. These also include GO Slim annotations, which are displayed in the Molecular Function, Cellular Component, and Biological Process columns of the .pdResult file.

  In addition, you can define your own categories of GO Slim annotations. In the Proteome Discoverer application, the categories containing user-defined GO Slim terms are called annotation aspects. For information on defining new annotation aspects, see "Defining Annotation Aspects" on page 282.

- Protein family (Pfam) annotations, which are displayed in the Pfam IDs column of the .pdResult file.

- KEGG PATHWAY and Wiki Pathways annotations, which are displayed in the following columns:

  – For KEGG PATHWAYS annotations, the node displays the KEGG Pathways, KEGG Pathway Accessions, and Gene Symbol columns of the .pdResult file.

  – For Wiki Pathways annotations, the node displays the WikiPathways, WikiPathway Accessions, and Gene Symbol columns of the .pdResult file.

- Gene identifications from the Entrez Gene database, which are displayed in the Entrez Gene ID column of the .pdResult file.

- Gene identifications from the reference genomes of the genetically sequenced organisms in the Ensembl genome database, which are displayed in the Ensemble Gene ID and Chromosome columns of the Proteins page of the .pdResult file.

- UniProt PTM modifications documented in the UniProt database, which are displayed on the Protein Identification Details view in the Proteins page of the .pdResult file.

The data in ProteinCenter is updated biweekly.

❖ **To create a processing protein annotation workflow**

1. Follow the general instructions for creating a processing workflow with the Workflow Editor. See "Creating a Processing Workflow" on page 68.

    The basic processing workflow for protein annotation is the same as the basic processing workflow shown in Figure 43 on page 72.

2. Include the following nodes in the processing workflow:

    - The Spectrum Files node

    - The Spectrum Selector node

    - A search engine node (Mascot or Sequest HT)

    - The Fixed Value PSM Validator node, Percolator node, or Target Decoy PSM Validator node

3. Add any other appropriate nodes.

4. Connect the nodes together.

5. Set the parameters for each node.

6. (Optional) Save the workflow:

    a. In the Name box above the Workflow Tree pane, type a name for the processing workflow.

    b. (Optional) In the Description box above the Workflow Tree pane, type a brief description of the processing workflow.

    c. In the Workflow Editor, click ⬛ Save .

    d. In the Save Workflow dialog box, do the following:

        i. Browse to the file to save the template in, or type a file name in the File Name box.

        ii. In the Save As Type box, select **Processing Workflow File (*.pdProcessingWF)**.

        iii. Click **Save**.

    The application saves the workflow in the *file_name*.pdProcessingWF file.

❖ **To create a consensus protein annotation workflow**

1. To use the ProteinCenter Annotation node in the consensus workflow, configure this node by following the instructions in "Downloading Files from ProteinCenter" on page 263.

2. Create a consensus workflow by following the general instructions in "Incorporating an Existing Workflow into a Study and an Analysis" on page 99.

3. Include the following nodes in the consensus workflow:

   • MSF Files node

   • PSM Grouper node

   • Peptide Validator node

   • Peptide and Protein Filter node

   • ProteinScorer node

   • ProteinCenter Annotation node

   • Protein Grouping node

   Connect the ProteinCenter Annotation node to the ProteinScorer node if the application does not automatically connect it.

   Figure 188 shows the basic protein annotation consensus workflow.

Figure 188. Basic protein annotation consensus workflow

4. (Optional) Add any other appropriate nodes.

5. Set the parameters for each node.

6. (Optional) Save the workflow:

   a. In the Name box above the Workflow Tree pane, type a name for the consensus workflow.

   b. (Optional) In the Description box above the Workflow Tree pane, type a brief description of the consensus workflow.

   c. In the Workflow Editor, click 🔧 Save .

   d. In the Save Workflow dialog box, do the following:

      i. Browse to the file to save the template in, or type a file name in the File Name box.

      ii. In the Save As Type box, select **Consensus Workflow File (*.pdConsensusWF)**.

      iii. Click **Save**.

   The application saves the workflow in the *file_name*.pdConsensusWF file.

7. Save the analysis. See "Saving an Analysis" on page 83.

8. Save the study. See "Saving a Study" on page 56.

9. Select the input files.

10. Click 🔄 Run in the upper right corner of the Analysis window.

   The job queue appears, as shown in Figure 53 on page 83, displaying the status of your search.Use the job queue to check the status of your search as the search progresses.

Displaying the Annotated Protein Results

When it finishes processing all search nodes, the Proteome Discoverer application retrieves the types of annotations listed in "ProteinCenter" on page 260 from ProteinCenter. You can display both predefined and user-defined annotated protein results in the results report (.pdResult file).

> **Note** The application cannot retrieve annotations from searches conducted in the UniRef FASTA database because of the prefix appended to the accession number.

- Displaying Predefined GO Protein Annotation Results

- Displaying Annotation Aspects

- Displaying GO Accessions

- Displaying Protein Family (Pfam) Annotations

- Displaying KEGG PATHWAY and Wiki Pathways Database Identifications

- Displaying Entrez Gene Database Identifications

- Displaying Ensembl Genome Database Annotations

- Displaying UniProt Annotations

Displaying Predefined GO Protein Annotation Results

Follow these procedures to display predefined GO protein categories, called annotation aspects, in the results report.

To display user-defined annotation aspects, see "Displaying Annotation Aspects" on page 282.

❖ **To display predefined GO protein annotation aspects in the results report**

1. Generate the .pdResult file by following the instructions in "Creating a Protein Annotation Workflow" on page 265.

2. If the Molecular Function, Cellular Component, and Biological Processes columns are not visible on the Proteins page (they are visible by default), select them in the Field Chooser dialog box.

   The application displays the results on the Proteins page of the .pdResult file as colored boxes similar to those shown in ProteinCenter. Figure 189 gives an example.

Figure 189. GO Slim category boxes for the protein groups shown in the results of an annotation search
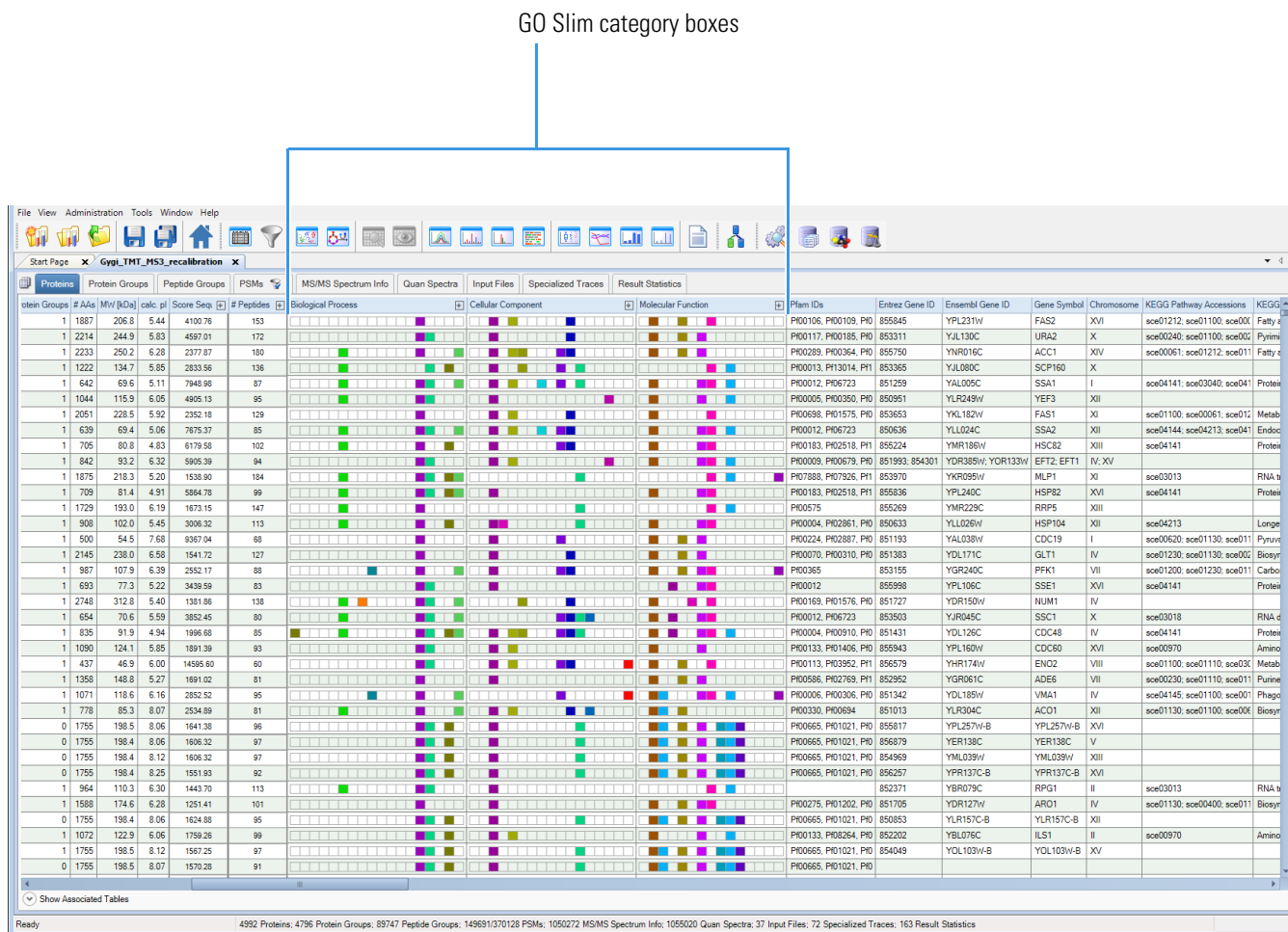
The application represents each aspect of the annotation (biological processes, cellular components, and molecular functions) in a separate column. Each box represents a GO Slim category, which is a selected subset of the Gene Ontology annotations. If the protein annotation is included in one of these subsets, the corresponding box is highlighted by a color specific to this GO Slim category. Figure 190 provides the column names and shows the meaning of the GO Slim category colors.

Figure 190. GO Slim category colors

Molecular Function
- antioxidant activity
- catalytic activity
- DNA binding
- enzyme regulator activity
- metal ion binding
- motor activity
- nucleotide binding
- protein binding
- receptor activity
- RNA binding
- signal transducer activity
- structural molecule activity
- transcription regulator activity
- translation regulator activity
- transporter activity

Cellular Component
- cell surface
- chromosome
- cytoplasm
- cytoskeleton
- cytosol
- endoplasmic reticulum
- endosome
- extracellular
- golgi
- membrane
- mitochondrion
- nucleus
- organelle lumen
- proteasome
- ribosome
- spliceosome
- vacuole

Biological Process
- cell communication
- cell death
- cell differentiation
- cell division
- cell growth
- cell organization and biogenesis
- cell proliferation
- cellular component movement
- cellular homeostasis
- coagulation
- conjugation
- defense response
- development
- metabolic process
- regulation of biological process
- reproduction
- response to stimulus
- transport

When you move the mouse over the GO Slim category boxes, a ToolTip showing the category name appears, as shown in the Molecular Function column in Figure 191.

Figure 191. ToolTip identifying the annotation category
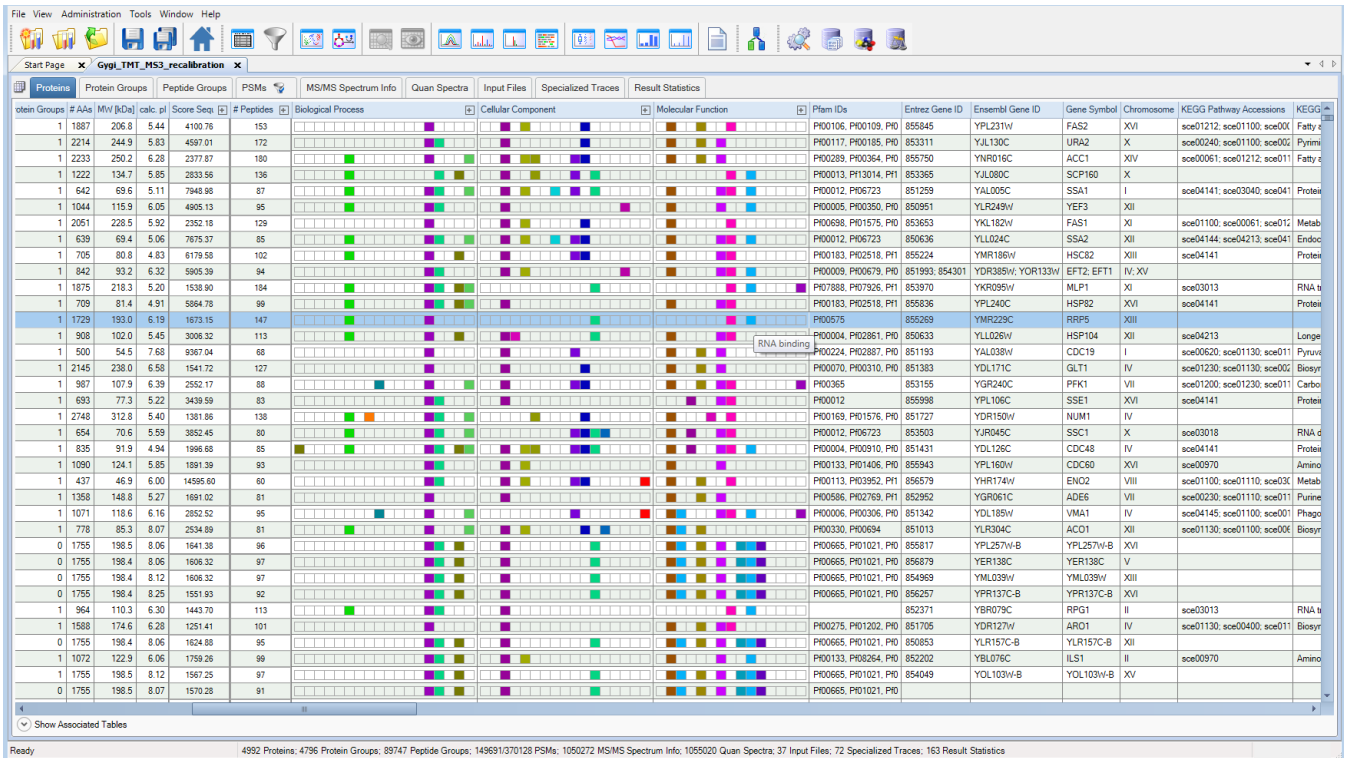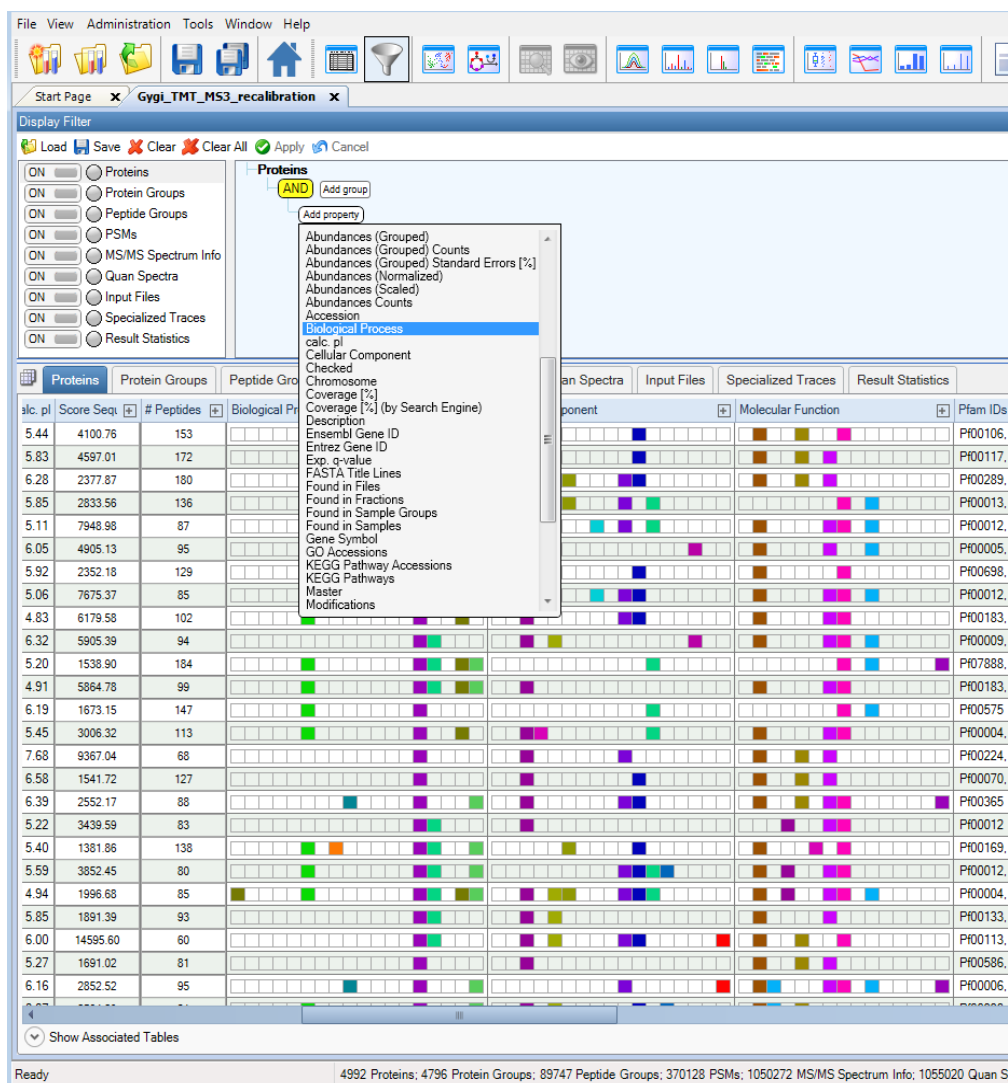
❖ **To filter the identified proteins by GO Slim categories**

1. In the .pdResult file, choose **View > Display Filter** or click the **Display Filter** icon, 🔻 .

2. In the Display Filters pane, select **Proteins** in the left column.

3. Click **Add Property**, and choose a property that represents an annotation aspect, such as **Molecular Function**, **Cellular Component**, or **Biological Process**.

   The example in Figure 192 shows the selection of Biological Processes.

Figure 192. Filtering annotation aspects

The application creates an expression with an "is true" operator.

4. In the blank target value box in the dialog box, select a predefined Go Slim value for the annotation, as shown in Figure 193.

Figure 193. Selecting a predefined value for the annotation

5. Click Apply .

Only those proteins that feature the selected aspect are now visible in the .pdResult file, as shown in Figure 194.

Figure 194. Proteins filtered by Go Slim category
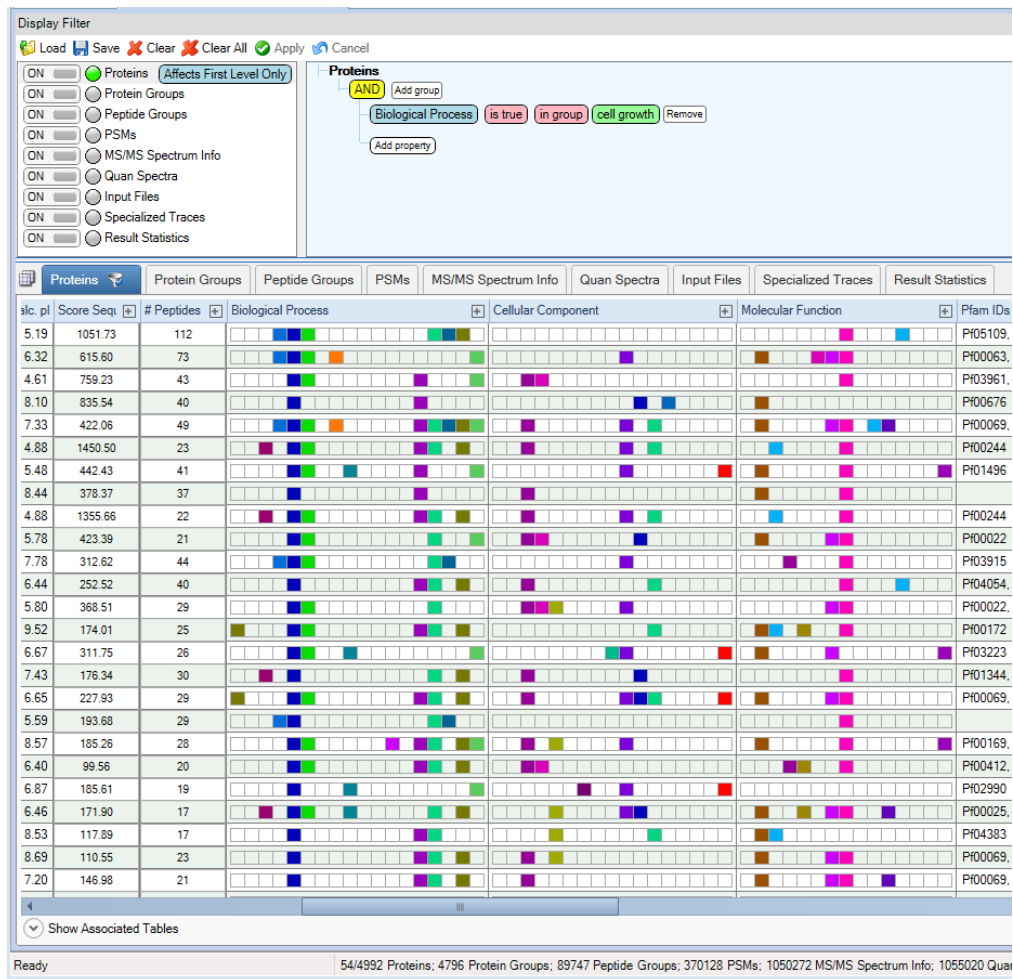
You can combine several aspect filters by using AND or OR as with any other numerical or textual data type.

Displaying GO Accessions

Gene ontology terms are related in hierarchical graphs. The GO term annotated to a special protein is always part of a complex directed graph. All ancestor elements—that is, the elements between the annotated GO term and one of the three top-level terms (Molecular Function, Cellular Component, and Biological Process)—are additional less-specific descriptions of the annotated value. For example, the "iron ion binding (GO:0005506)" term contains in its graph the "metal ion binding (GO:0046872)" value, which is less specific. All GO terms contained in the graph of the annotated GO term of the protein are represented in the GO Accessions column on the Proteins page.

❖ **To display GO accessions**

1. Open the generated .pdResult file.

2. Click the **Proteins** tab.

3. In the Field Chooser dialog box of the proteins page, select the **GO Accessions** column.

   The application displays the protein's GO terms contained in the graph of the annotated GO term on the Proteins page of the .pdResult report in the GO Accessions column, as shown in Figure 195.

Figure 195. Go Accessions column on the Proteins page of the .pdResult file

The square parentheses around a GO term indicate that the term is a child term of the GO terms without parentheses on the same line. For example, on the first line in Figure 196, [GO:0000287] is the child term of the higher-level terms GO:0046872, GO:0043169, GO:0043167, GO: 0005488, and GO: 0003674. The latter five terms represent a higher level in the biological context of the protein.

Figure 196. Child and higher-level GO terms

4. Move the cursor over the GO Accessions column.

The application displays the annotated GO term and all ancestor terms associated with a protein, as shown in Figure 197. It shows the term annotated to the protein in brackets, followed by their ancestor terms. Each annotated GO term starts on a new line. If you want all proteins to have a higher-level annotation that is not provided by the Molecular Function, Cellular Component, and Biological Process annotation columns, you can filter for the GO term in this column.

Figure 197. The complete list of GO terms associated with a protein

Displaying Protein Family (Pfam) Annotations

As noted in "Pfam Database Annotation" on page 262, you can retrieve Pfam annotations from the Pfam database as an alternative to GO annotations.

❖ **To display Protein Family (Pfam) annotations**

1. Open the results report (.pdResult) file by following the instructions in the Help.

2. Click the **Proteins** tab.

3. If the Pfam IDs column is not visible on the Proteins page, select it in the Field Chooser dialog box.

   For information on the Field Chooser dialog box, refer to the Help.

   Figure 198 shows the Pfam IDs column on the Proteins page.

Figure 198. Pfam ID annotations for the protein groups shown in the results of an annotation search

Displaying KEGG PATHWAY and Wiki Pathways Database Identifications

"KEGG PATHWAY and Wiki Pathways Database Annotations" on page 262 describes the identifications that the ProteinCenter Annotation node retrieves from the KEGG PATHWAY and Wiki Pathways databases.

❖ **To display KEGG PATHWAY and Wiki Pathways database identifications**

1. Open the results report (.pdResult) file by following the instructions in the Help.

2. Click the **Proteins** tab.

3. In the Field Chooser dialog box of the Proteins page, select the following columns:

   • Gene Symbol: Displays the official symbol of the gene.

   • KEGG Pathway Accessions: Displays the accessions from the KEGG PATHWAY database.

   • KEGG Pathways: Displays the descriptions from the KEGG PATHWAY database.

   • WikiPathways Accessions: Displays the accessions from the Wiki Pathways database.

   • WikiPathways: Displays the descriptions from the Wiki Pathways database.

For information on the Field Chooser dialog box, refer to the Help.

The application displays the gene identifications in these column of the Proteins page of the .pdResult) file, as shown in Figure 199.

Figure 199. Gene Symbol, KEGG PATHWAY, and Wiki Pathways annotations on the Proteins page

To display KEGG PATHWAY and Wiki Pathways maps, refer to the Help.

Displaying Entrez Gene Database Identifications

Entrez Gene database identifications are unique identifications assigned to all genes stored in the Entrez Gene database, NCBI's database of gene-specific information. The Proteome Discoverer application displays these identifications in the Entrez Gene ID column on the Proteins page, as shown in Figure 200 on page 280. All proteins derived from the same gene have the same gene identification. You can use this information to group or cluster biologically meaningful proteins together. Because not all genes are stored in the Entrez Gene database, some proteins do not have a valid gene identification. In this case, the column is empty. For more information on the Entrez Gene identifications, see "Entrez Gene Database Annotation" on page 262.

❖ **To display Entrez Gene database identifications**

1. Open the results report (.pdResult) file by following the instructions in the Help.

2. Click the **Proteins** tab.

3. In the Field Chooser dialog box of the Proteins page, select the **Entrez Gene ID** column.

    For information on the Field Chooser dialog box, refer to the Help.

    The application displays the gene identifications in the Entrez Gene ID column of the Proteins page of the .pdResult) file, as shown in Figure 200. Because not all genes are stored in the Entrez Gene database, some proteins do not have a valid gene identification. In this case, the column displays a value of 0.

Figure 200. Entrez Gene ID column on the Proteins page

Displaying Ensembl Genome Database Annotations

The Ensemble genome database stores and automatically annotates the reference genomes of the genetically sequenced organisms. For fully sequenced species, the Ensembl gene identification and chromosome information is available if you retrieve protein annotations from ProteinCenter. If this information is available, it appears in the Ensembl Gene ID and Chromosome columns on the Proteins page of the .pdResult file, as shown in Figure 201. As with the Entrez Gene identifications, you can use this information to group or cluster proteins by these values.

❖ **To display Ensembl genome database annotations**

1. Open the results report (.pdResult) file by following the instructions in the Help.

2. Click the **Proteins** tab.

3. In the Field Chooser dialog box of the proteins page, select the **Ensembl Gene ID** and the **Chromosome** column.

   For information on the Field Chooser dialog box, refer to the Help.

   The application displays the genome and chromosome identifications in the Ensembl Gene ID and Chromosome columns, respectively, of the Proteins page of the .pdResult file, as shown in Figure 201.

Figure 201. Columns for Ensembl Genome database identifications and chromosome number

Displaying UniProt Annotations

For information on displaying UniProt PTM annotations, refer to the Help.

Displaying Annotation Aspects

Although 50 GO Slim terms are available in the Molecular Function, Cellular Component, and Biological Processes categories, these annotation categories, called annotation aspects, might not be specific enough for your needs. You might want to define your own aspects to display in the results report. You can define these aspects in the Administration view.

To display predefined aspects such as Molecular Function, see "Displaying Predefined GO Protein Annotation Results" on page 269.

- Defining Annotation Aspects
- Using User-Defined Annotation Aspects
- Editing User-Defined Annotation Aspects
- Removing User-Defined Annotation Aspects
- Exporting User-Defined Annotation Aspects
- Importing User-Defined Annotation Aspects
- Defining New Annotation Aspects for the First Time After Installation
- Annotation Aspect Editor Parameters
- Annotation Aspects View Parameters

Defining Annotation Aspects

The Annotation Aspects view, shown in Figure 202, displays the currently available annotation aspects. When you click + to the left of an aspect, the application displays the GO Slim terms, annotation groups, or pathways associated with that aspect. It also displays the color code used in the results report for each group and the group name used in the ToolTip that appears when you move the mouse over the GO Slim category boxes, as shown in Figure 191 on page 271. Figure 202 shows these groups for the Biological Processes category.

You can create user-defined annotation aspects based on pathway annotations. In the Annotation Aspect Editor, you create an aspect where each of the colored boxes can represent one or more pathways. As with the GO and Pfam aspects, you cannot use more than one annotation database to define the single categories of an annotation aspect.

Figure 202. Annotation Aspects view
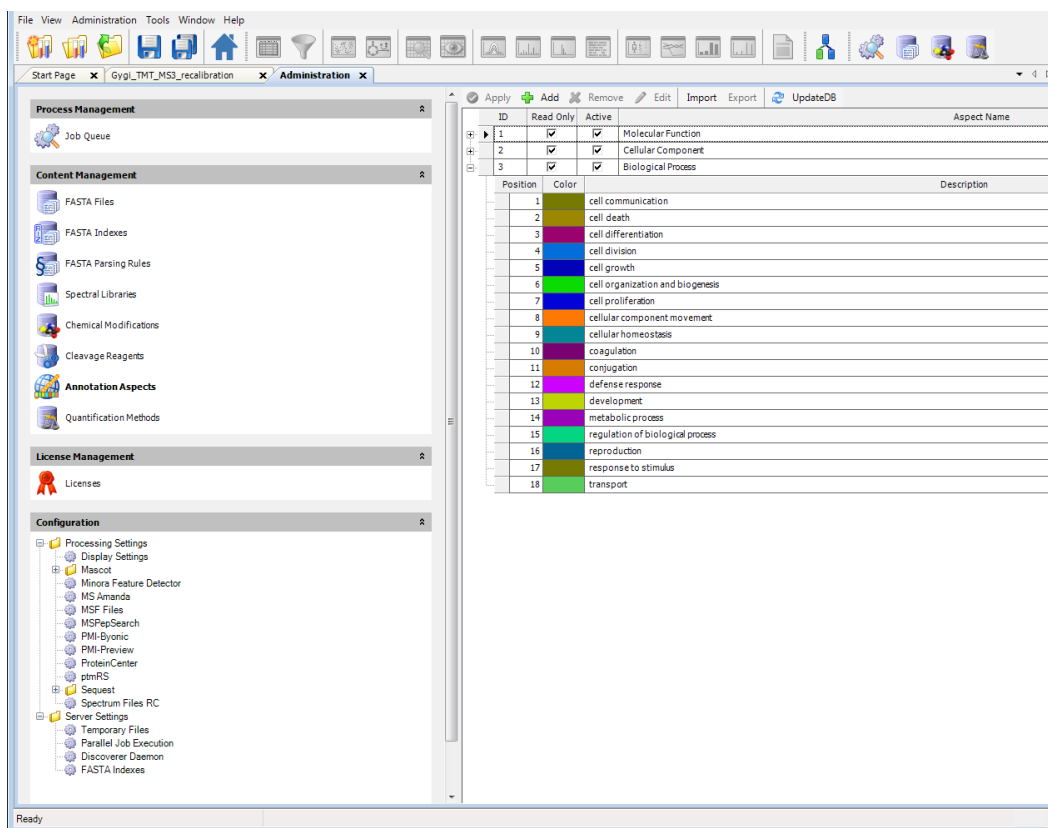
❖ **To define an annotation aspect**

1. Choose **Administration > Maintain Annotation Aspects**.

2. In the Annotation Aspect view on the Administration page, click 🞢 Add in the toolbar.

   The Annotation Aspect Editor opens and loads the hierarchical definitions of the GO terms, Pfam terms, KEGG PATHWAY pathways, or Wiki Pathways pathways in the Available Annotations area in the lower right part of the Editor. Figure 203 shows GO terms loaded into the Editor.

Figure 203. Annotation Aspect Editor
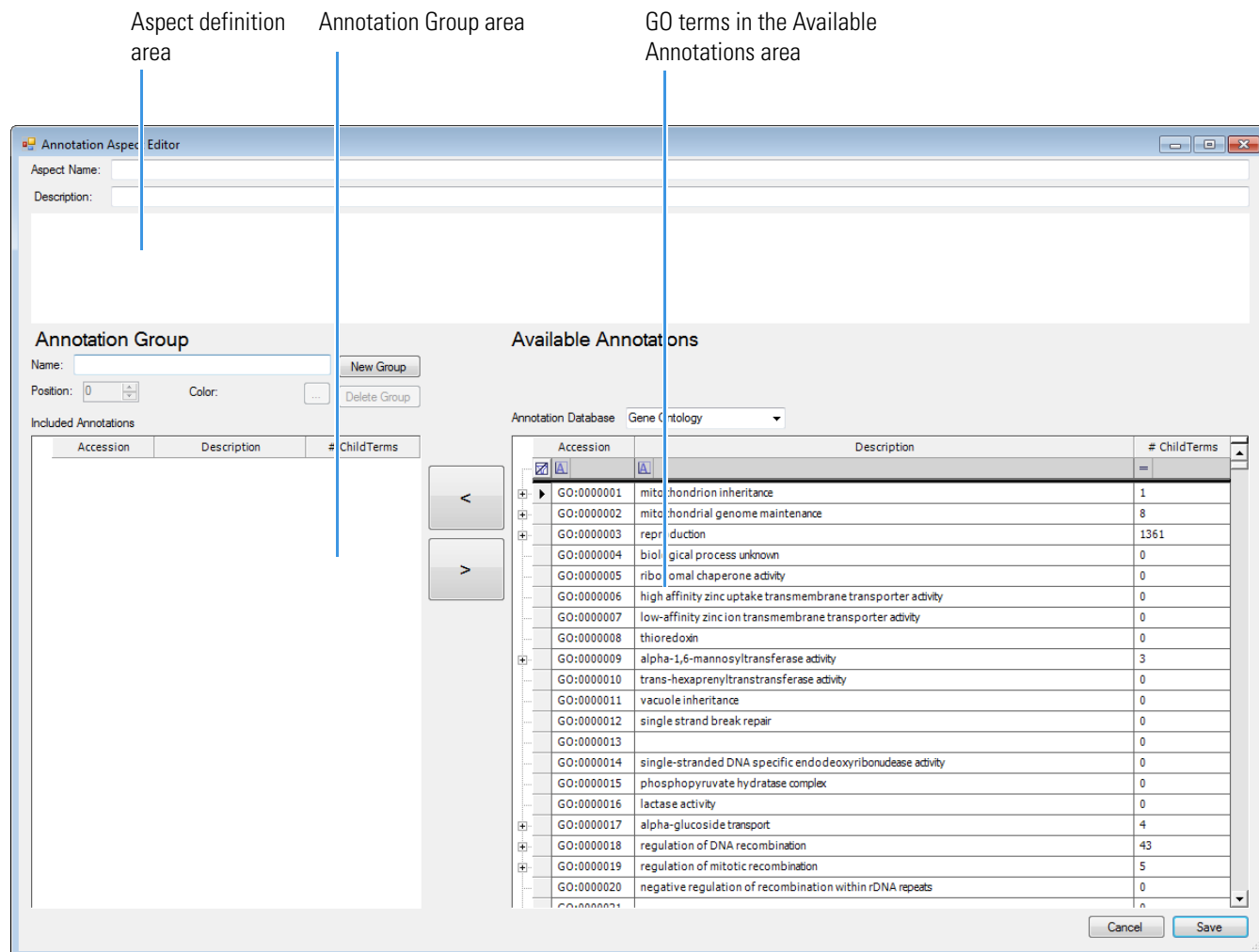
The Annotation Aspects Editor consists of three parts:

- On the top is the aspect definition area, which includes the name and description of the annotation aspect and colored boxes representing the annotation groups of the aspect.

- On the lower left is the Annotation Group area, which contains the definitions of a single annotation group.

- On the lower right is the Available Annotations area, which lists the available annotations.

3. In the Aspect Name box in the aspect definition area, enter the aspect name, for example, **Metal_Ion_Binding**.

You can use letters, numbers, and underscores (_) in the aspect name. Spaces are not allowed.

4. In the Description box, type a brief definition of the new aspect, for example, "Chemical element bound to the protein."

5. Click **New Group** in the Annotation Group area of the Editor.

   A red box appears in the aspect definition area of the Editor beneath the Description box.

6. In the Name box in the Annotation Group area, enter the name of the group, for example, "Magnesium" (see Figure 204).

7. Specify the fill color of the group rectangle used if a protein belongs to the new group (see Figure 204):

   a. Click [...] next to Color to open the Color dialog box.

   b. Click the color that you want, or define a custom color.

   c. Click **OK**.

8. (Optional) If you specify more than one group, change the position of a rectangle in the aspect definition area.

   Starting from the left, the rectangles in the aspect definition area are assigned a sequential position number in the Position box. To change the position of a rectangle in the aspect definition area, select the rectangle and enter the new position number into the Position box. You can also use the up and down arrows in the Position box to set the new position number. For example, if the rectangle corresponding to potassium is the fourth from the left, and you want it to be the seventh rectangle from the left, enter 7 in the Position box.

   The Available Annotations area lists all GO annotations with the accession number and a description. When you click + to the left of a GO annotation, the Editor displays all child terms, if any, for the annotation.

9. In the Annotation Database box, select the database of annotation to use, either **Gene Ontology, Protein Families (Pfam)**, **KEGG PATHWAY**, or **Wiki Pathways**.

   For information on the Gene Ontology (GO) annotation, see "Gene Ontology (GO) Database Annotation" on page 261. For information on Pfam annotation, see "Pfam Database Annotation" on page 262. For information on KEGG PATHWAY and Wiki Pathways annotation, see "KEGG PATHWAY and Wiki Pathways Database Annotations" on page 262.

10. Scroll down the list of GO, Pfam, KEGG PATHWAY, or Wiki Pathways terms to find the appropriate term, in this case, "metal ion binding." For faster access, use the column filters in the gray row below the column headers. Type the name you are looking for in the filter cell. The entries in the table are automatically restricted to values containing the given name.

    Figure 204 shows a partial list of the available annotations and the child terms for metal ion binding.

Figure 204. List of child terms for metal ion binding
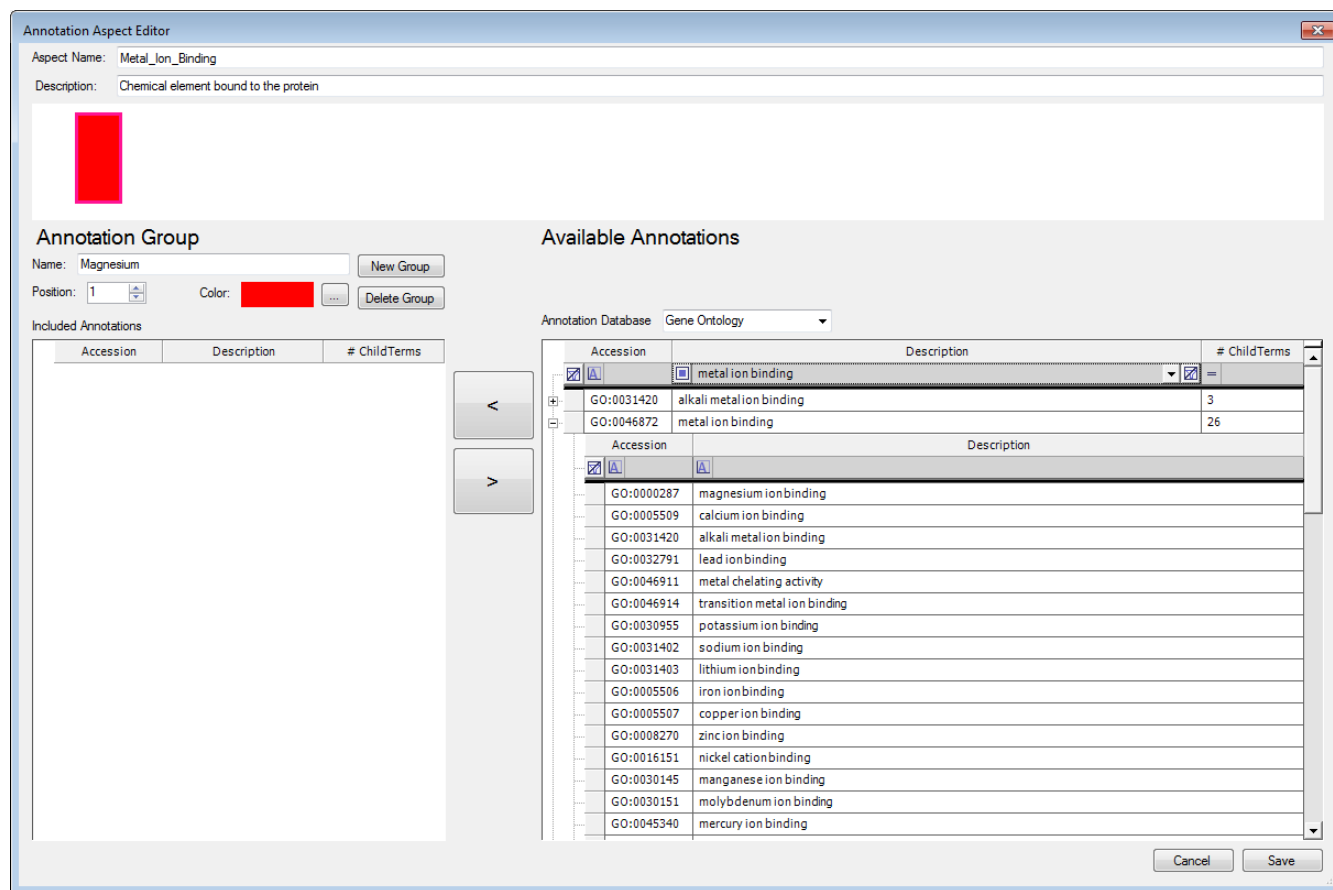
11. To associate GO terms with the group, select the terms to include in the list of available annotations, and click the left arrow, ⬚, to move them to the Included Annotations pane.

All GO terms are listed with an accession number and a description. The proteins in the results report (.pdResult file) reflect all GO or Pfam terms, including the appended child terms, that you move to the group definition. For example, when you define a group with 17 different GO or Pfam terms, the results report displays a colored rectangle if the corresponding protein is annotated with one of these 17 terms. Otherwise, the rectangle remains empty.

Figure 205 shows the complete definition of an aspect containing one group, which is highlighted for all proteins that are annotated with "GO:0000287- magnesium ion binding." With this annotation aspect, you can filter the results report (.pdResult file) for all proteins containing magnesium.

Figure 205. Annotation aspect with completed definition for one group
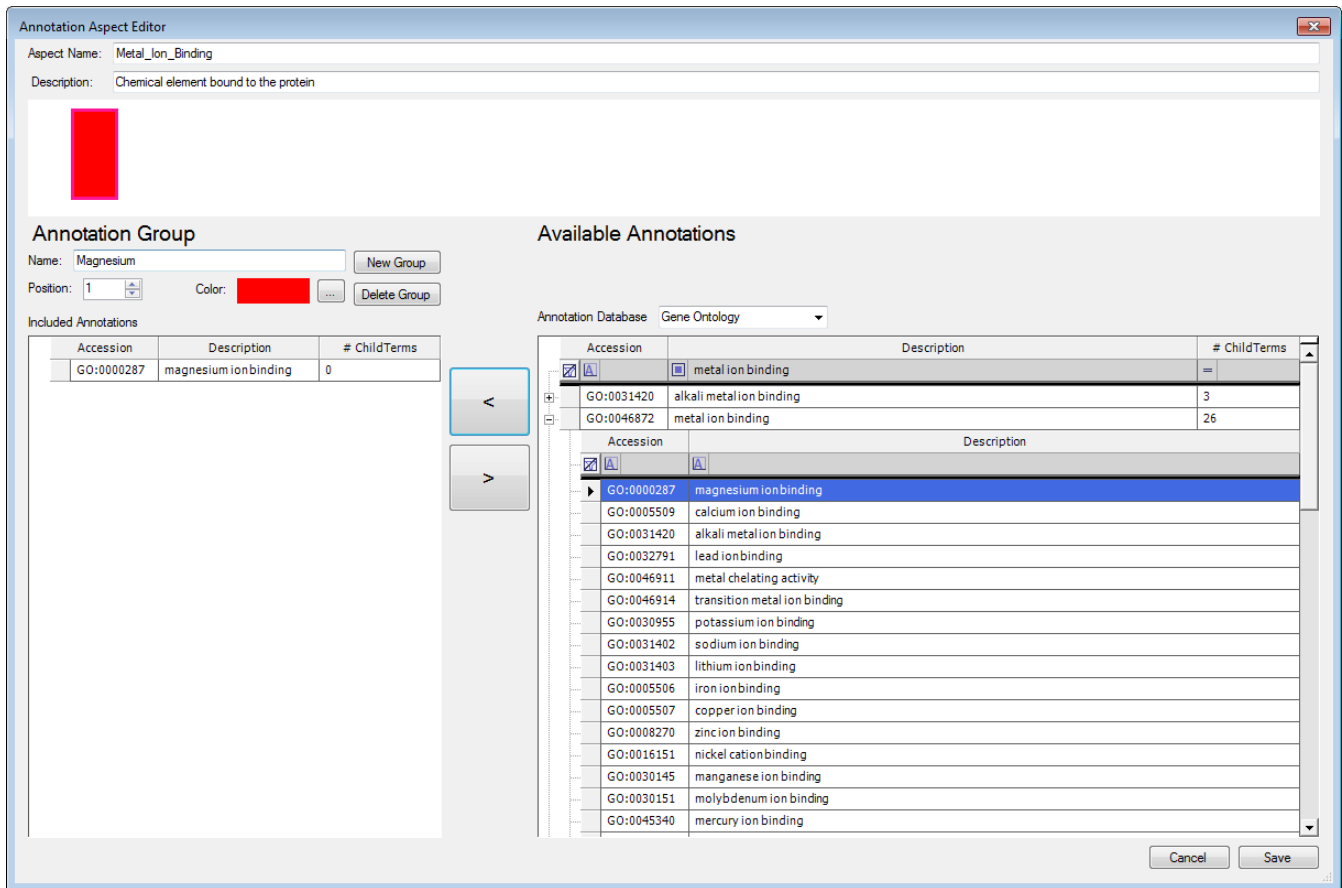
Figure 206 shows the metal ion binding aspect with more group definitions. You can define a maximum of 25 groups per aspect.

Figure 206. Annotation aspect with eleven groups

Figure 207 shows KEGG PATHWAY annotations.

Figure 207. Annotation Aspect Editor using KEGG PATHWAY annotations to define the colored annotation groups
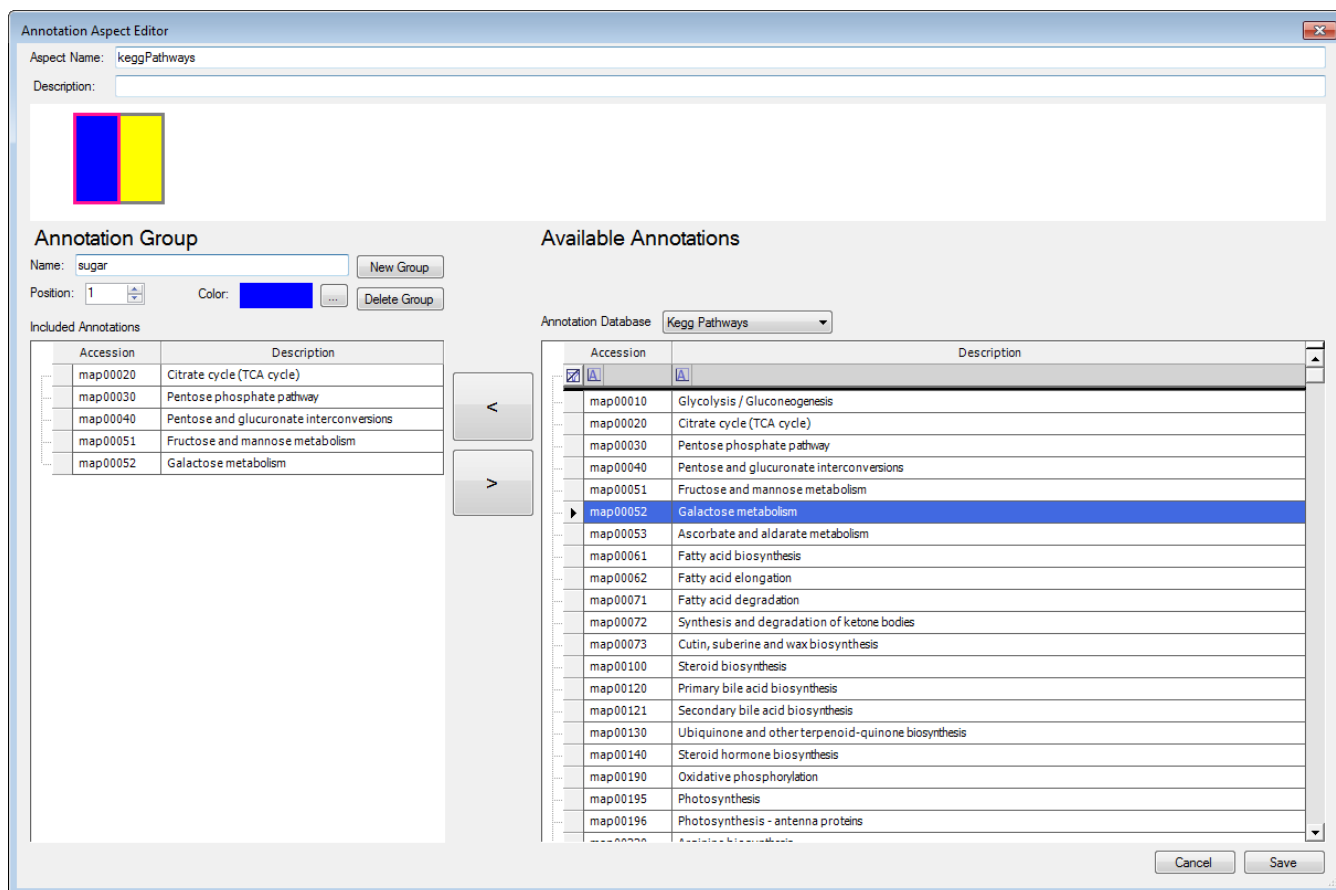
To remove a group, select the colored box corresponding to the group and click **Delete Group** in the Annotation Group area.

Figure 208 shows the annotation aspects for the KEGG PATHWAY database pathways in the KEGG Pathways column on the Proteins page. This column appears on the Proteins page when you include the ProteinCenter Annotation node in the consensus workflow.

Figure 208. Annotation Aspect for KEGG PATHWAY pathways shown in the .pdResult file

❖ **To save user-defined aspect annotations**

Click **Save** in the lower right corner of the Annotation Aspect Editor.

Before the application saves an aspect, it verifies that the aspect name and the group names are defined and valid. It also verifies that you set the color and the list of group definitions. If it finds missing mandatory values, it opens a message box listing the missing values. You must set all mandatory values before the application saves the annotation aspect. After the application saves the annotation aspect, it lists the aspect in the Administration view, as shown in Figure 209. Click the plus sign, **+**, to the left of the row to display the single groups of the aspect, as shown in Figure 209.

The Read Only column indicates annotation aspects that are defined in the application as read-only. You cannot change these annotation aspects and cannot delete them. The Edit and Delete buttons in the toolbar above the list are therefore disabled. However, you can change and delete any user-defined annotation aspects.

Figure 209. User-defined aspect in the Annotation Aspects view
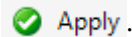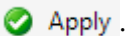
Using User-Defined Annotation Aspects

By default, a user-defined aspect is active in the Annotation Aspects view and therefore usable in the ProteinCenter Annotation node.

❖ **To deactivate a user-defined annotation aspect**

1. In the table in the Annotation Aspects view, click the check box in the Active column for the appropriate annotation aspect.

2. Click ✅ Apply .

3. To make an annotation aspect active, select its check box in the Active column, and click ✅ Apply .

You can select user-defined aspects in the ProteinCenter Annotation node in a consensus workflow. As shown in Figure 210, the ProteinCenter Annotation node has up to six annotation aspects available for the Annotation Aspects parameter that you can apply to the annotation. When you click on each numbered aspect, a dropdown list appears from which you can choose any predefined and user-defined aspect. The list contains all aspect marked "Active" in the Administration view.

Figure 210. User-defined annotation aspect in the Parameters pane of the Workflow Editor
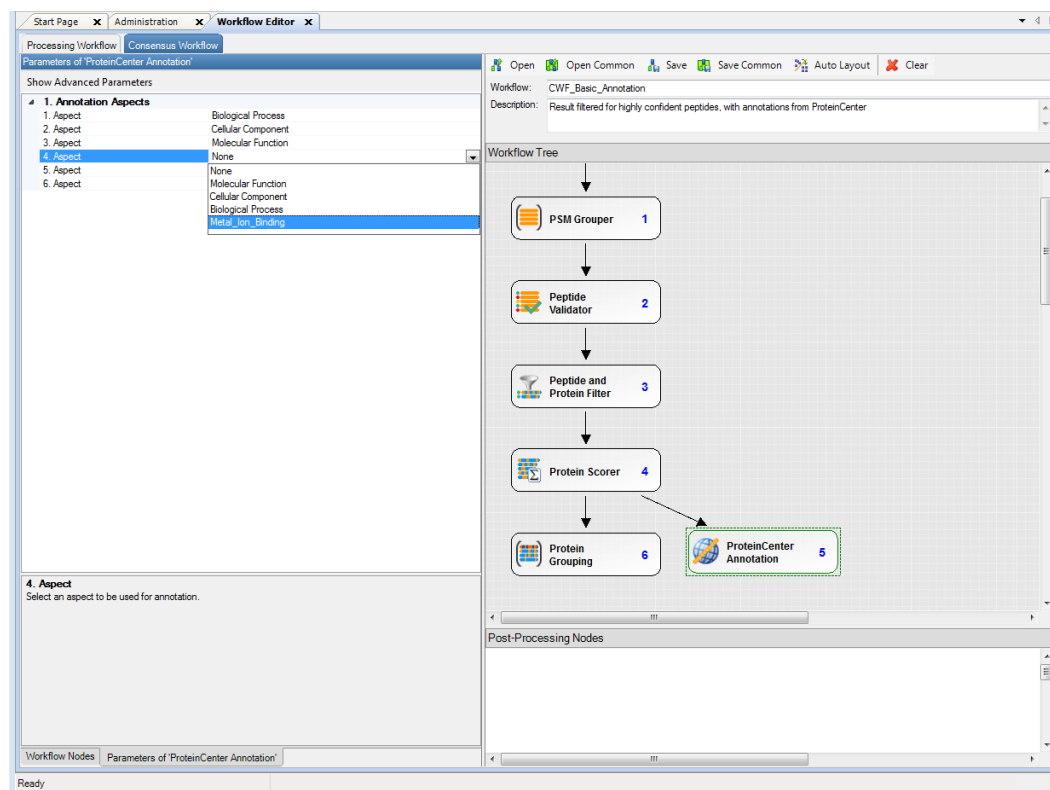
The application uses the new annotation aspects in its consensus workflow and displays the results of the annotation in the .pdResult file, as shown in Figure 211.

Figure 211. User-defined annotation aspects in the results report

Editing User-Defined Annotation Aspects

You can change an existing user-defined annotation aspect.

❖ **To edit user-defined annotation aspects**

1. In the table in the Annotation Aspects view on the Administration page, double-click the annotation aspect that you want to edit.

   –or–

   Do the following:

   a. Select the row containing the user-defined annotation aspect that you want to edit. Click the row in the column to the left of the ID column to select the aspect.

   b. Click ✎ Edit in the toolbar.

   The Annotation Aspects Editor opens.

   When you double-click a predefined aspect like Biological Process, the Annotation Aspect Editor opens, but you cannot change any of the settings. You can only edit the settings for user-defined annotation aspects.

2. Make the changes that you want, and click **Save** in the lower right-hand corner of the Annotation Aspects Editor.

The Annotation Aspect Editor closes, leaving the Annotation Aspects view displayed on the Administration page.

Removing User-Defined Annotation Aspects

You can remove a user-defined annotation aspect.

❖ **To remove user-defined annotation aspects**

1. In the table on the Annotation Aspects view on the Administration page, select the row containing the user-defined annotation aspect that you want to remove. Click the row in the column to the left of the ID column to select the aspect.

2. Click ✖ Remove on the toolbar.

The annotation aspect no longer appears in the table in the Annotation Aspects view.

Exporting User-Defined Annotation Aspects

You can export user-defined annotation aspects to an XML file. You might want to use this XML file to share user-defined annotation aspects between different installations of the Proteome Discoverer application, or use it as a backup if you install the application on a new computer.

❖ **To export a user-defined annotation aspect**

1. In the table on the Annotation Aspects view on the Administration page, select the row containing the user-defined annotation aspect that you want to export. Click the row in the column to the left of the ID column to select the aspect.

2. Click Export in the toolbar.

3. In the Export Annotation Aspect dialog box, enter the name of the XML file where you want to save the annotation aspects.

4. Click **Save**.

Importing User-Defined Annotation Aspects

You can import user-defined annotation aspects from an XML file.

❖ **To import a user-defined annotation aspect**

1. In the table in the Annotation Aspects view of the Administration page, select the row containing the user-defined annotation aspect that you want to import. Click the row in the column to the left of the ID column to select the aspect.

2. Click Import in the toolbar.

3. In the Import Annotation Aspect dialog box, select the XML file containing the annotation aspects that you want to import.

4. Click **Open**.

The imported annotation aspect now appears in the table in the Annotation Aspects view.

Defining New Annotation Aspects for the First Time After Installation

After you install the Proteome Discoverer application, only an initial version of the annotation database is available, and it does not contain any annotation terms. You must have an updated annotation database available before you can define new aspects. The application updates the annotation database during the first use of the ProteinCenter Annotation node in a consensus workflow. If you define new aspects before using the ProteinCenter Annotation node for the first time or after a long period of no usage in a consensus workflow (in this case, the local annotation database would be outdated), manually update the annotation database by following this procedure.

❖ **To update the annotation aspect database for the first time after installation**

Click 🔄 UpdateDB in the Annotation Aspects view.

The view changes to the job queue view, as shown in Figure 212.

Figure 212. Job queue showing update of annotation aspect database
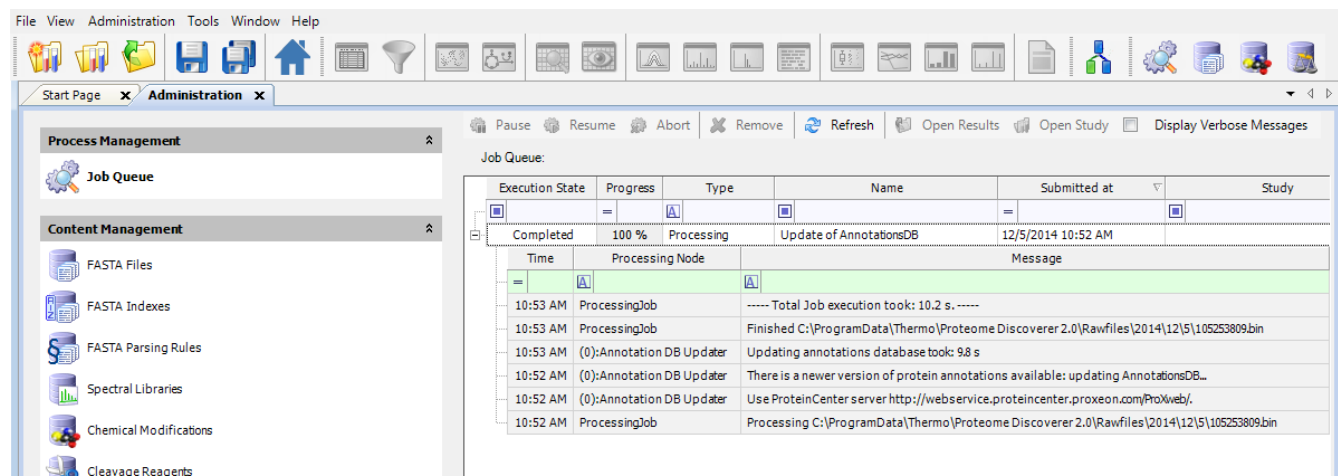
Annotation Aspect Editor Parameters

Table 14 describes the parameters available in the Annotation Aspect Editor.
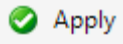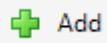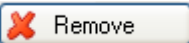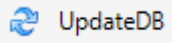
Table 14. Annotation Aspect Editor parameters

Parameter	Description
Aspect Name	Specifies the name of the annotation aspect.
Description	Briefly describes the annotation aspect.
Aspect definition area	Displays colored rectangles corresponding to the GO or Pfam terms (called groups) assigned to the annotation aspect.
Annotation Group	Contains parameters that define a group.
Name	Specifies the name of the group.
New Group	Places a rectangle in the aspect definition area to represent the new group being defined.
Position	Specifies the position of a rectangle representing a group in the aspect definition area.
Color	Specifies the color that the group displays in the results report (.pdResult file).
Delete Group	Deletes a group.
Included Annotations	Lists the annotations assigned to a group.
Accession	Displays the accession number of the annotation.
Description	Briefly describes the annotation.
# Child Terms	Displays the number of child terms that the annotation has.
<	Assigns a GO or Pfam term to an annotation group.
>	Removes a GO or Pfam term from an annotation group.
Available Annotations	Lists the annotations in the selected database.
Annotation Database	Selects the database of annotations to display.
Accession	Displays the accession number of the annotation.
Description	Briefly describes the annotation.
# Child Terms	Displays the number of child terms that the annotation has.
Cancel	Discards any changes to the annotation aspect and closes the Annotation Aspect Editor.
Saves	Saves the annotation aspect.

Annotation Aspects View Parameters

Table 15 describes the parameters available in the Annotation Aspects view on the Administration page.

Table 15. Annotation Aspects view parameters

Parameter	Description
Apply	Implements any changes that you make to the ID, Read Only, or Active columns.
Add	Opens the Annotation Aspects Editor so that you can define a new annotation aspect.
Remove	Removes an annotation aspect from the table on the Annotation Aspects view.
Edit	Opens the Annotation Aspects Editor so that you can change the settings of an existing annotation aspect.
Import	Imports an XML file containing annotation aspects and displays it in the table on the Annotation Aspects view.
Export	Places the selected annotation aspects in an XML file.
UpdateDB	Updates an annotation database before you define a new annotation aspect for the first time after installation of the Proteome Discoverer application.
ID	Displays the sequential identification number of the annotation aspect.
Read Only	Indicates whether an annotation aspect is read-only. If it is, you cannot change it or delete it. The Edit and the Delete buttons in the toolbar above the list are therefore disabled. However, you can change and delete any user-defined annotation aspects.
Active	Determines whether the annotation aspect is displayed in the ProteinCenter Annotation node.
Aspect Name	Displays the name of the annotation aspects.
Position	Specifies the position of the colored rectangle representing a group assigned to the annotation aspect.
Color	Specifies the color that a group of an annotation aspect displays in the results report (.pdResult file).
Description	Briefly describes the group of an annotation aspect.

Accessing the ProteinCard Page

You can access the data in ProteinCenter through the ProteinCard page of the Protein Identification Details view for each protein. In the ProteinCard page, a protein is considered a specific amino acid sequence in a given species.

❖ **To open the ProteinCard page**

1. If you want to download the data in ProteinCenter to a location other than the default URL shown when you choose Administration > Configuration > ProteinCenter > ProteinCenter URL, see "Configuring the ProteinCenter Web Server" on page 264.

2. Double-click a grid cell on the Proteins page of the .pdResult file.

   –or–

   Select a cell and choose **View > Protein Details**.

   –or–

   Click the **Protein Details** icon, .

3. In the Protein Identification Details view, click the **ProteinCard** tab.

   You might experience a short delay as the Proteome Discoverer application accesses the URL.

   After loading data from the ProteinCenter server, the Proteome Discoverer application displays the data in the ProteinCard page. By default, the ProteinCard page shows the General page, shown in Figure 214 on page 299.

4. Click on the tab of the page containing the information that you are seeking:

   • General Page

   • Keys Page

   • Features Page

   • Molecular Functions Page

   • Cellular Components Page

   • Biological Processes Page

   • Diseases Page

   • External Links Page

5. Click **OK** to close the Protein Identification Details dialog box.

If the entire protein is not found in ProteinCenter but a protein with the same sequence exists, the ProteinCard page displays a warning that the displayed information is from a protein with different accession, as shown in Figure 213. If there is more than one protein with the same sequence but from different organisms, an additional list box appears so that you can select the correct species.

Figure 213. Warning displayed for protein with different accession
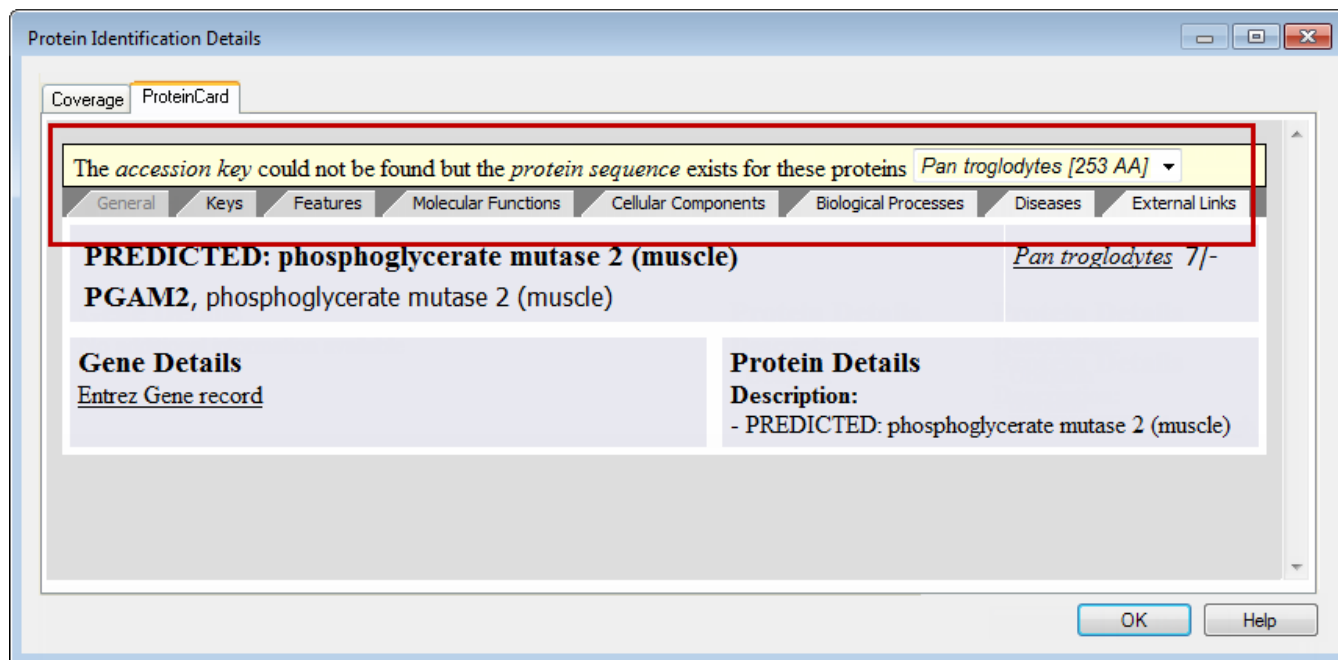
ProteinCard Page Parameters

The ProteinCard page of the Protein Identification Details dialog box contains the following pages.

- General Page

- Keys Page

- Features Page

- Molecular Functions Page

- Cellular Components Page

- Biological Processes Page

- Pathways Page

- Diseases Page

- External Links Page

General Page

The General page of the ProteinCard page, shown in Figure 214, displays information about the protein: its name, its description, its function, the keywords that produce it in a database search, and the gene that ultimately directs the protein's synthesis through RNA.

Figure 214. General page of the ProteinCard page
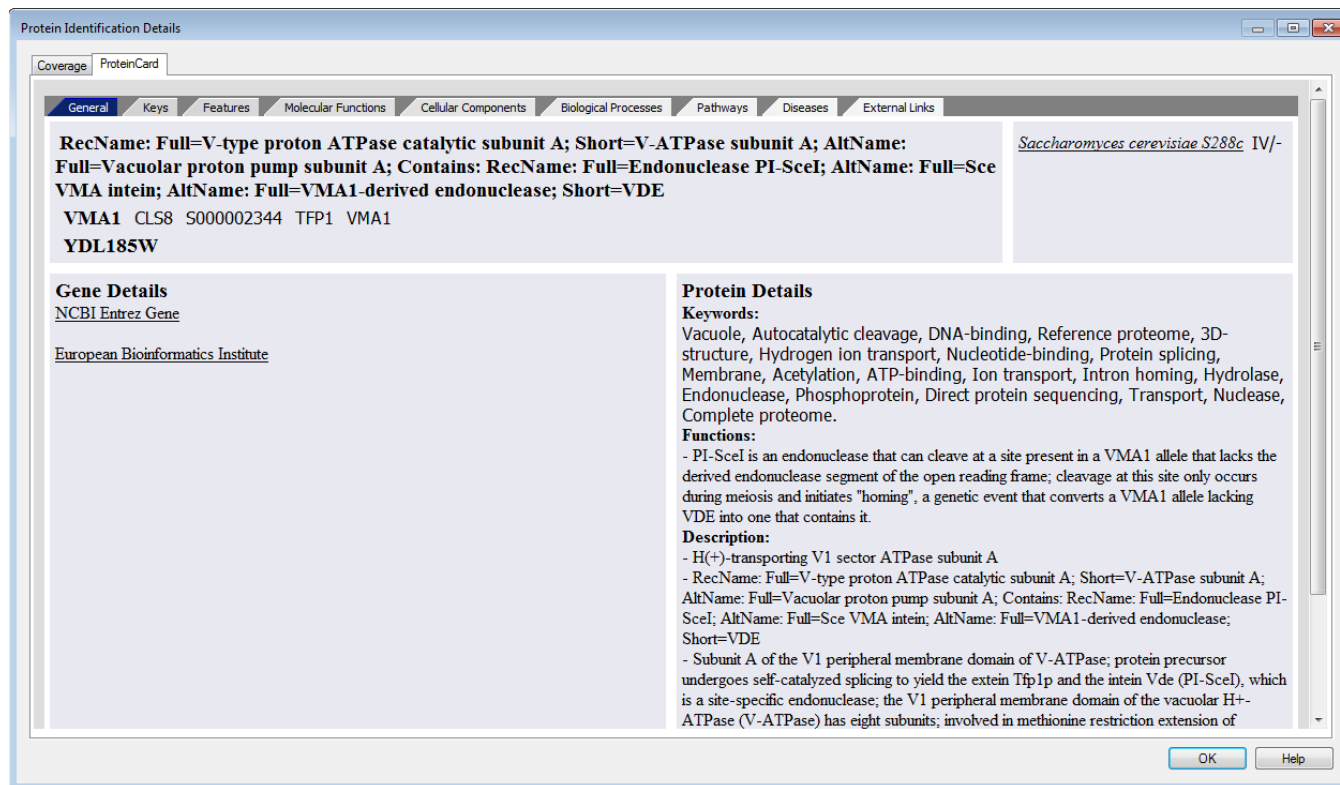
Table 16 lists the parameters on the General page of the ProteinCard page.

Table 16. Parameters on the General page of the ProteinCard page (Sheet 1 of 2)

Parameter	Description
Top area	Displays the protein name in bold font on the first line. The second line in bold font is the official symbol of the gene that ultimately directs the synthesis of the protein through RNA, and the text following it is the alternative name or names of the gene.
Top right area	Displays the name of the species that contains the gene that ultimately directs the synthesis of this protein through RNA, the number of the chromosome that the gene resides on, and the location of the chromosome that the gene resides on. The name of the species links to the National Center for Biotechnology Information (NCBI) taxonomy browser.

Table 16. Parameters on the General page of the ProteinCard page (Sheet 2 of 2)

Parameter	Description
Gene Details area	Displays information about the gene that directs the synthesis of the protein. If no information about the gene is available, the application provides a link to the Entrez Gene website.
Protein Details area	Lists the keywords that produce this protein in a database search, the functions of the protein, and a description of the protein.

Keys Page

The Keys page of the ProteinCard page, shown in Figure 215, lists all the accession keys for a given protein.

Figure 215. Keys page of the ProteinCard page
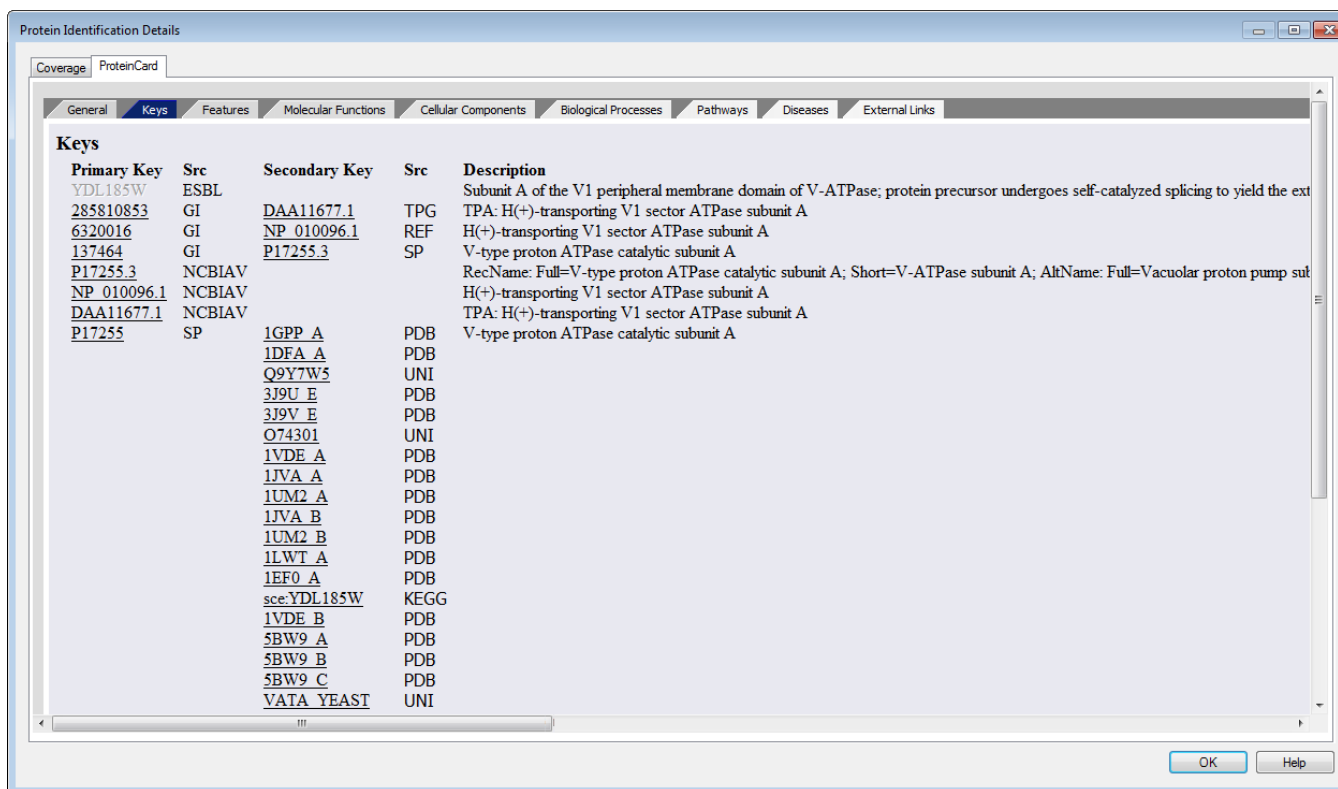
Table 17 lists the parameters on the Keys page of the ProteinCard page.

Table 17. Parameters on the Keys page of the ProteinCard page

Parameter	Description
Primary Key	Lists the accession key of the database that the sequence was imported from. It is linked to the original database records in the source database, such as Ensembl, SGD, NRDB, IPI or UniProt. This parameter emphasizes the preferred type of accession.
Src	Specifies the abbreviation of the primary source database.
Secondary Key	Lists the secondary accession key, which is either an alternative key used in the source database or the key of the original database.
Src	Specifies the abbreviation of the secondary source database.
Description	Displays the original description for the original database entry.

An exclamation mark flags outdated protein keys, and the keys are linked to the outdating history in their respective source database.

Features Page

The Features page of the ProteinCard page, shown in Figure 216, includes a selection of sequence features from UniProt, from various conserved domain predictions, and from the computational enrichment undertaken by ProteinCenter. (Computational enrichment refers to information that has no experimental evidence but was found by using a computer prediction program.) The features are sorted according to their start positions in the protein sequence.

Figure 216. Features page of the ProteinCard page

Table 18 lists the parameters on the Features page of the ProteinCard page.

Table 18. Parameters on the Features page of the ProteinCard page (Sheet 1 of 2)

Parameter	Description
Source	Specifies the name of the database that provided information about the feature: • InterPro • Tmap (computational enrichment) • PrediSi (computational enrichment) • Pfam (computational enrichment) • UniProt
Category	Displays the type of information that UniProt, InterPro, and Tmap include for each row. For example, UniProt might include "CARBOHYD" as one of its types of information, and InterPro might include "SSF57184" as one of its types of information.
From	Specifies the start position of the amino acid.

Table 18. Parameters on the Features page of the ProteinCard page (Sheet 2 of 2)

Parameter	Description
To	Specifies the end position of the amino acid.
Acc	Specifies the accession identifier for the domain linked to InterPro or Pfam.
Description	Describes the feature.

Molecular Functions Page

The Molecular Functions page of the ProteinCard page, shown in Figure 217, summarizes information about the function of the protein. It consolidates GO data and Enzyme Category (EC) information. The EC designation indicates whether a protein has been categorized with a certain enzyme function.

Figure 217. Molecular Functions page of the ProteinCard page

Table 19 lists the parameters on the Molecular Functions page of the ProteinCard page.

Table 19. Parameters on the Molecular Functions page of the ProteinCard page

Parameter	Description
GO Id	Lists the GO code for each of the protein's molecular functions. Each code is linked to the QuickGO browser of the European Bioinformatics Institute (EBI), which hosts several databases and services.
Evidence Codes	Lists the evidence codes for each of the protein's molecular functions for GO annotation. Evidence codes describe how the GO information was proven—for example, by computer prediction or by experiment.
PMIDs	Lists the molecular function codes in the PubMed database, which is maintained by the U.S. National Library of Medicine (NLM) and the National Institutes of Health (NIH). Each code is linked to the PubMed browser.
Go Slim	Specifies the basic GO Slim category for the GO term. GO Slim categories are reduced versions of the GO ontologies containing a subset of the terms in the entire GO database. They give a broad overview of the ontology content without the detail of the specific fine-grained terms. Table 22 on page 310 provides the Go Slim categories for molecular functions.
Name	Describes the molecular function for a GO term. This description is created by the GO consortium.

Enzymes with an EC number for IUBMB Enzyme Nomenclature are displayed with links to detailed information at the International Union of Biochemistry and Molecular Biology.

Cellular Components Page

The Cellular Components page of the ProteinCard page, shown in Figure 218, summarizes information about where the protein carries out its function in the cell.

Figure 218. Cellular Components page of the ProteinCard page
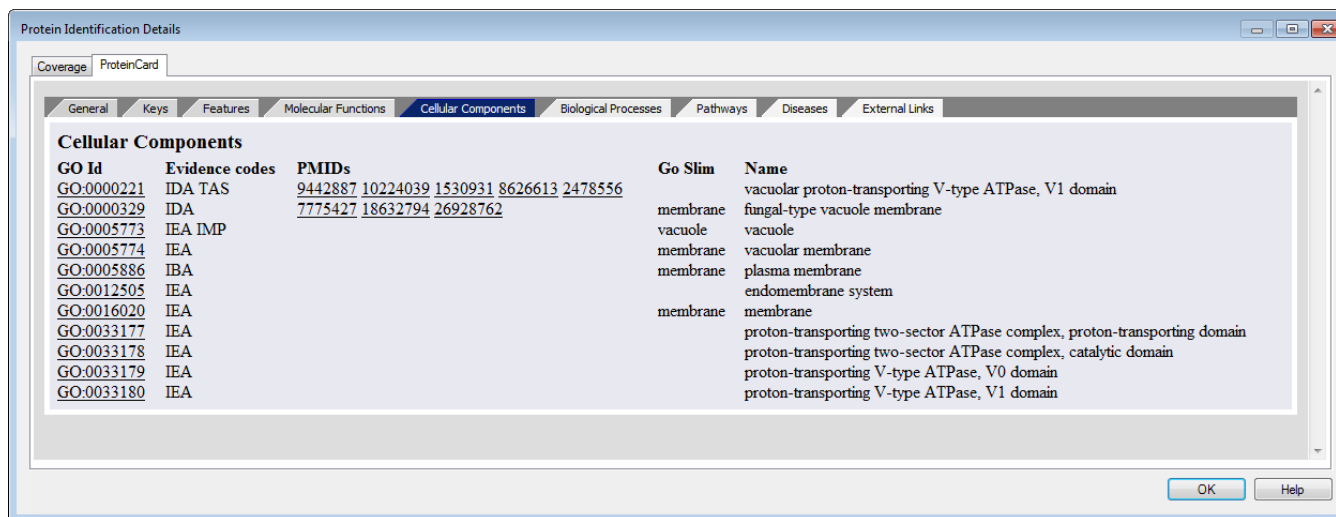
Table 20 lists the parameters on the Cellular Components page of the ProteinCard page.

Table 20. Parameters on the Cellular Components page of the ProteinCard page (Sheet 1 of 2)

Parameter	Description
GO Id	Lists the GO code for each of the protein's molecular functions. Each code is linked to the QuickGO browser of the EBI, which hosts a number of databases and services.
Evidence Codes	Lists the evidence codes for each of the protein's cellular components for GO annotation. Evidence codes describe how the GO information was proven—for example, by computer prediction or by experiment.
PMIDs	Lists the cellular component codes in the PubMed database, which is maintained by the NLM and the NIH. Each code is linked to the PubMed browser.
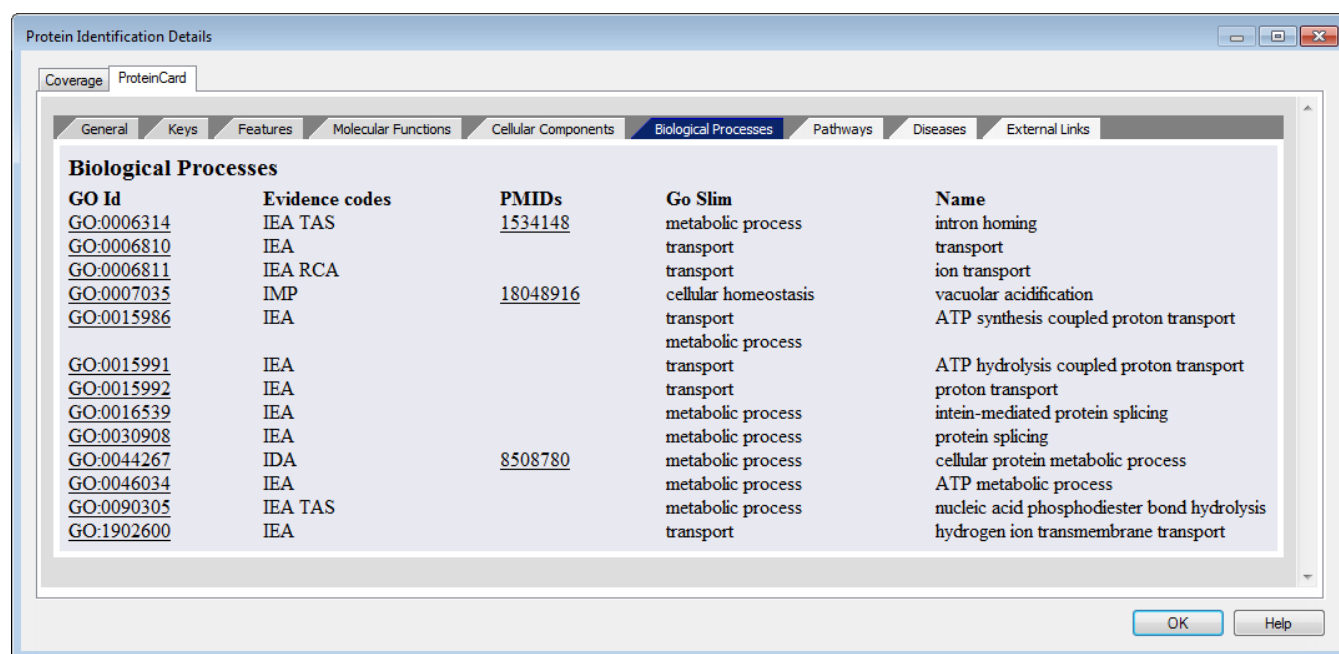
Table 20. Parameters on the Cellular Components page of the ProteinCard page (Sheet 2 of 2)

Parameter	Description
Go Slim	Specifies the basic GO Slim category for the GO term. GO Slim categories are reduced versions of the GO ontologies containing a subset of the terms in the entire GO database. They give a broad overview of the ontology content without the detail of the specific fine-grained terms. Table 23 on page 311 provides the Go Slim categories for cellular components.
Name	Describes the cellular component for a GO term. This description is created by the GO consortium.

Enzymes with an EC number for IUBMB Enzyme Nomenclature are displayed with links to detailed information at the International Union of Biochemistry and Molecular Biology.

Biological Processes Page

The Biological Processes page of the ProteinCard page, shown in Figure 219, summarizes information about the biological processes that the protein belongs to.

Figure 219. Biological Processes page of the ProteinCard page

Table 21 lists the parameters on the Biological Processes page of the ProteinCard page.

Table 21. Parameters on the Biological Processes page of the ProteinCard page

Parameter	Description
GO Id	Lists the GO code for each of the protein's molecular functions. Each code is linked to the QuickGO browser of the EBI, which hosts a number of databases and services.
Evidence Codes	Lists the evidence codes for each of the protein's biological processes for GO annotation. Evidence codes describe how the GO information was proven—for example, by computer prediction or by experiment.
PMIDs	Lists the biological process codes in the PubMed database, which is maintained by the NLM and the NIH. Each code is linked to the PubMed browser.
Go Slim	Specifies the basic GO Slim category for the GO term. GO Slim categories are reduced versions of the GO ontologies containing a subset of the terms in the entire GO database. They give a broad overview of the ontology content without the detail of the specific fine-grained terms. Table 24 provides the Go Slim categories for biological components.
Name	Describes the biological process for a GO term. This description is created by the GO consortium.

Enzymes with an EC number for IUBMB Enzyme Nomenclature are displayed with links to detailed information at the International Union of Biochemistry and Molecular Biology.

Pathways Page

The Pathways page of the ProteinCard page, shown in Figure 220, lists the pathways that the protein belongs to and the links to them.

Figure 220. Pathways page of the ProteinCard page
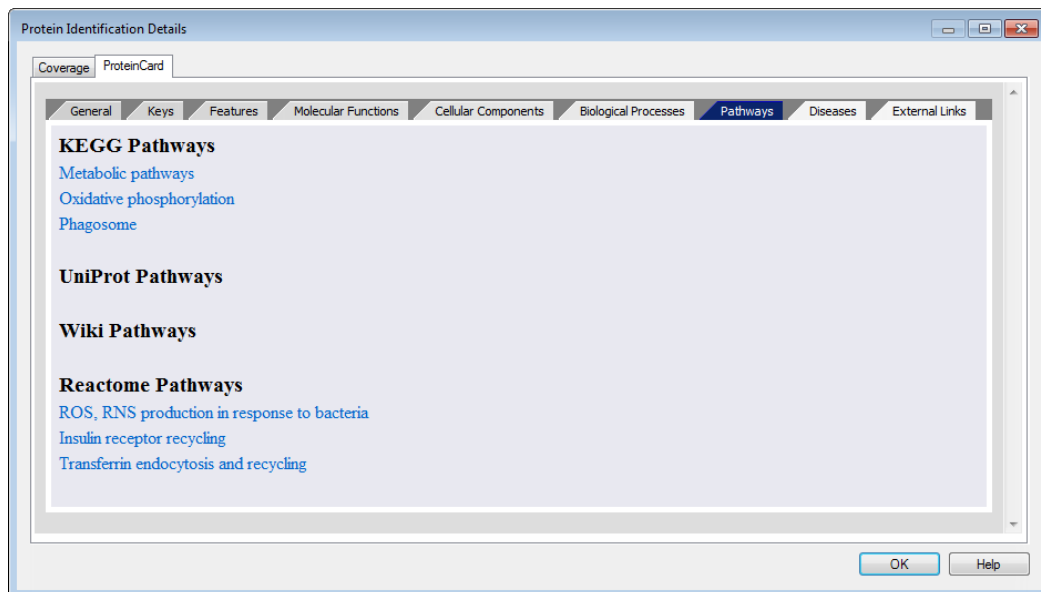
Diseases Page

The Diseases page of the ProteinCard page, shown in Figure 221, lists the diseases that the selected protein is associated with.

Figure 221. Diseases page of the ProteinCard page
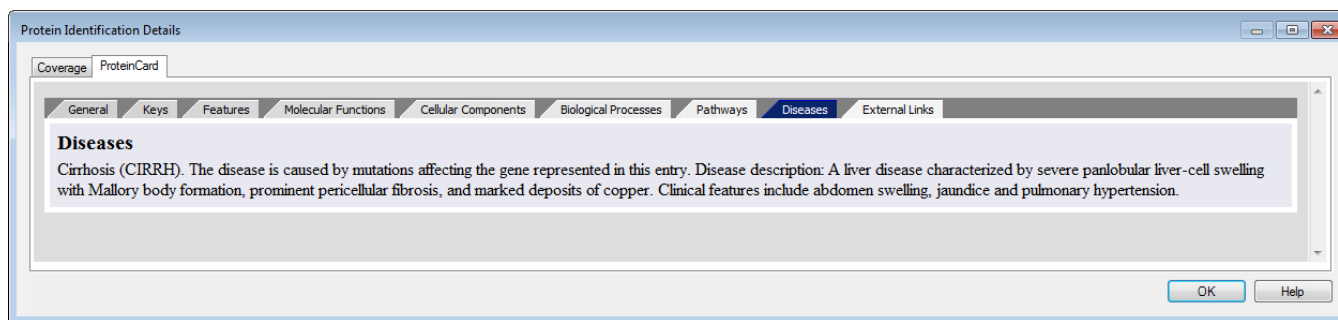
External Links Page

The External Links page of the ProteinCard page, shown in Figure 222, lists the web links to resources containing information about the protein.

Figure 222. External Links page of the ProteinCard page
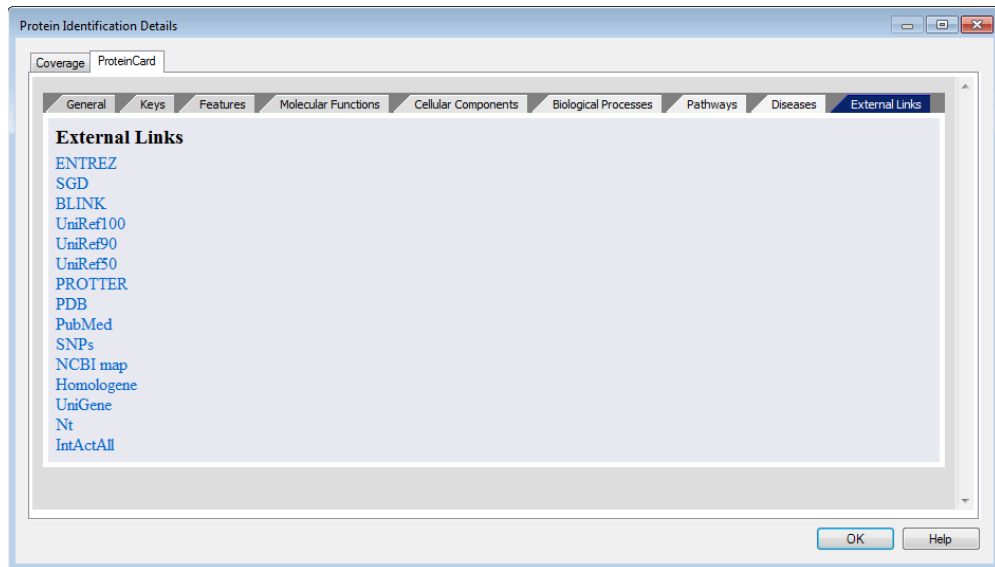
Click the appropriate link to open the browser for the database. The external links contain links to resources containing information about the respective protein.

GO Slim Categories

This topic defines the GO Slim terms for molecular functions, cellular components, and biological processes.

GO Slim Categories for Molecular Functions

Table 22 describes the GO Slim categories for molecular functions.

Table 22. GO Slim categories for molecular functions (Sheet 1 of 2)

GO Slim molecular function	Description
Antioxidant activity	Inhibition of the reactions brought about by dioxygen (O2) or peroxides. Usually the antioxidant is effective because it can be more easily oxidized than the substance protected. The term is often applied to components that can trap free radicals, breaking the chain reaction that normally leads to extensive biological damage.
Catalytic activity	Catalysis of a biochemical reaction at physiological temperatures. In biologically catalyzed reactions, the reactants are known as substrates, and the catalysts are naturally occurring macromolecular substances known as enzymes. Enzymes possess specific binding sites for substrates and are usually composed wholly or largely of protein.
DNA binding	Selective interaction with DNA (deoxyribonucleic acid).
Enzyme regulator activity	Modulation of an enzyme.
Metal ion binding	Selective interaction with any metal ion.
Motor activity	Catalysis of movement along a polymeric molecule such as a microfilament or microtubule, coupled to the hydrolysis of a nucleoside triphosphate.
Nucleotide binding	Selective interaction with a nucleotide, which is any compound consisting of a nucleoside that is esterified with (ortho)phosphate or an oligophosphate at any hydroxyl group on the ribose or deoxyribose moiety.
Protein binding	Selective interaction with any protein or protein complex (a complex of two or more proteins that might include other nonprotein molecules).
Receptor activity	The mediation by protein or gene products of a signal from the extracellular environment to a intracellular messenger.
RNA binding	Selective interaction with an RNA molecule or a portion of it.
Signal transducer activity	Mediation of the transfer of a signal from the outside to the inside of a cell by means other than the introduction of the signal molecule itself into the cell.

Table 22. GO Slim categories for molecular functions (Sheet 2 of 2)

GO Slim molecular function	Description
Structural molecule activity	The action of a molecule that contributes to the structural integrity of a complex or assembly within or outside a cell.
Transcription regulator activity	Activity that plays a role in regulating transcription; it might bind a promoter or enhancer DNA sequence or interact with a DNA-binding transcription factor.
Translation regulator activity	The initiation, activation, perpetuation, repression, or termination of polypeptide synthesis at the ribosome.
Transporter activity	Activity that enables the directed movement of substances (such as macromolecules, small molecules, ions) into, out of, within, or between cells.

GO Slim Categories for Cellular Components

Table 23 describes the GO Slim categories for cellular components.

Table 23. GO Slim categories for cellular components (Sheet 1 of 4)

GO Slim cellular component	Description
Cell surface	Proteins that are attached to the external part of the cell wall, cell membrane, or both.
Chromosome	A structure composed of a very long molecule of DNA and associated proteins (for example, histones) that carry hereditary information.
Cytoplasm	All of the contents of a cell excluding the plasma membrane and nucleus but including other subcellular structures.
Cytoskeleton	Any of the various filamentous elements that form the internal framework of cells and that typically remain after treatment of the cells with mild detergent to remove membrane constituents and soluble components of the cytoplasm. The term embraces intermediate filaments, microfilaments, microtubules, the microtrabecular lattice, and other structures characterized by a polymeric filamentous nature and long-range order within the cell. The various elements of the cytoskeleton not only serve in the maintenance of cellular shape but also have roles in other cellular functions, including cellular movement, cell division, endocytosis, and movement of organelles.
Cytosol	That part of the cytoplasm that does not contain membranous or particulate subcellular components.

Table 23. GO Slim categories for cellular components (Sheet 2 of 4)

GO Slim cellular component	Description
Endosome	A membrane-bound organelle that carries materials newly ingested by endocytosis. It passes many of the materials to lysosomes for degradation.
Endoplasmatic reticulum	The irregular network of unit membranes, visible only by electron microscopy, that occurs in the cytoplasm of many eukaryotic cells. The membranes form a complex meshwork of tubular channels, which are often expanded into slit-like cavities called cisternae. The endoplasmatic reticulum takes two forms, rough (or granular), with ribosomes adhering to the outer surface, and smooth, with no ribosomes attached.
Extracellular	The space external to the outermost structure of a cell. For cells without external protective or external encapsulating structures, this term refers to the space outside of the plasma membrane. It only applies to proteins that are not attached to the cell surface. It covers the host cell environment outside an intracellular parasite.
Golgi	A compound membranous cytoplasmic organelle of eukaryotic cells consisting of flattened, ribosome-free vesicles arranged in a more or less regular stack. The Golgi apparatus differs from the endoplasmic reticulum in often having slightly thicker membranes, appearing in sections as a characteristic shallow semicircle so that the convex side (cis or entry face) abuts the endoplasmic reticulum, secretory vesicles emerging from the concave side (trans or exit face). In vertebrate cells, there is usually one such organelle, but in invertebrates and plants, where they are known usually as dictyosomes, there might be several scattered in the cytoplasm. The Golgi apparatus processes proteins produced on the ribosomes of the rough endoplasmic reticulum. Such processing includes modification of the core oligosaccharides of glycoproteins and the sorting and packaging of proteins for transport to a variety of cellular locations.

Table 23. GO Slim categories for cellular components (Sheet 3 of 4)

GO Slim cellular component	Description
Membrane	Double layer of lipid molecules that encloses all cells, and, in eukaryotic cells, many organelles. The membrane can be a single or double lipid bilayer. It also includes associated proteins.
	Note This term is not restricted to the plasma membrane but applies to all types of membranes present in the cell, that is, nuclear membranes and mitochondrial membranes.
Mitochondrion	A semiautonomous, self-replicating organelle that occurs in varying numbers, shapes, and sizes in the cytoplasm of virtually all eukaryotic cells. It is notably the site of tissue respiration.
Nucleus	A membrane-bounded organelle of eukaryotic cells where chromosomes are housed and replicated. In most cells, the nucleus contains all of the cell's chromosomes except the organellar chromosomes and is the site of RNA synthesis and processing. In some species or in specialized cell types, RNA metabolism or DNA replication might be absent.
Spliceosome	A ribonucleoprotein complex containing RNA and small nuclear ribonucleoproteins (snRNPs), which is assembled during the splicing of messenger RNA primary transcript to excise an intron.
Protein complex	Any protein group composed of two or more subunits, which might or might not be identical. Protein complexes might have other associated non-protein prosthetic groups, such as nucleic acids, metal ions, or carbohydrate groups.
Ribosome	An intracellular organelle, about 200 angstroms in diameter, consisting of RNA and protein. It is the site of protein biosynthesis resulting from translation of messenger RNA (mRNA).

Table 23. GO Slim categories for cellular components (Sheet 4 of 4)

GO Slim cellular component	Description
Vacuole	A closed structure found only in eukaryotic cells, completely surrounded by unit membrane and containing liquid material. Cells contain one or several vacuoles that might have different functions from each other. Vacuoles have a diverse array of functions. They can act as a storage organelle for nutrients or waste products, as a degradative compartment, as a cost-effective way of increasing cell size, and as a homeostatic regulator controlling both the turgor pressure and the pH of the cytosol.
Organelle lumen	The volume enclosed by the membranes of a particular organelle, for example, endoplasmic reticulum lumen or the space between the two lipid bilayers of a double membrane surrounding an organelle (for example, nuclear membrane lumen).

GO Slim Categories for Biological Processes

Table 24 describes the GO Slim categories for biological processes.

Table 24. GO Slim categories for biological processes (Sheet 1 of 3)

Go Slim biological process	Description
Cell communication	Any process that mediates interactions between a cell and its surroundings. Cell communication encompasses interactions such as signaling or attachment between one cell and another cell, between a cell and an extracellular matrix, or between a cell and any other aspect of its environment.
Cell death	The specific activation or halting of processes within a cell so that its vital functions markedly cease, rather than simply deteriorate gradually over time, which culminates in cell death.
Cell differentiation	The process where relatively unspecialized cells—for example, embryonic or regenerative cells—acquire specialized structural features, functional features, or both that characterize the cells, tissues, or organs of the mature organism or some other relatively stable phase of the organism's life history. Differentiation includes the processes involved in commitment of a cell to a specific fate.

Table 24. GO Slim categories for biological processes (Sheet 2 of 3)

Go Slim biological process	Description
Cell division	The processes resulting in the physical partitioning and separation of a cell into daughter cells.
Cell growth	The process by which a cell irreversibly increases in size over time by accretion and biosynthetic production of matter similar to that already present.
Cell homeostasis	The processes involved in the maintenance of an internal equilibrium at the level of the cell.
Cell motility	Any process involved in the controlled movement of a cell.
Cell organization and biogenesis	A process that is carried out at the cellular level and that results in the formation, arrangement of constituent parts, or disassembly of a cellular component. The process includes the plasma membrane and any external encapsulating structures, such as the cell wall and cell envelope.
Cell proliferation	The multiplication or reproduction of cells, resulting in the rapid expansion of a cell population.
Coagulation	The process by which a fluid solution, or part of it, changes into a solid or semisolid mass.
Conjugation	The union or introduction of genetic information from compatible mating types that results in a genetically different individual. Conjugation requires direct cellular contact between the organisms.
Defense response	Reactions triggered in response to the presence of a foreign body or the occurrence of an injury, which result in restriction of damage to the organism attacked or prevention and recovery from the infection caused by the attack.
Development	The biological process whose specific outcome is the progression of an organism over time from an initial condition (for example, a zygote or a young adult) to a later condition (for example, a multicellular animal or an aged adult).
Metabolic process	Processes that cause many of the chemical changes in living organisms, including anabolism and catabolism. Metabolic processes typically transform small molecules but also include macromolecular processes such as DNA repair and replication, and protein synthesis and degradation.

Table 24. GO Slim categories for biological processes (Sheet 3 of 3)

Go Slim biological process	Description
Regulation of biological process	Any process that modulates the frequency, rate, or extent of a biological process. Biological processes are regulated by many means, for example, control of gene expression, protein modification, or interaction with a protein or substrate molecule.
Reproduction	The production by an organism of new individuals that contain some portion of their genetic material inherited from that organism.
Response to stimulus	A change in state or activity of a cell or an organism (in terms of movement, secretion, enzyme production, gene expression, and so forth) as a result of a stimulus.
Transport	The directed movement of substances (such as macromolecules, small molecules, ions) into, out of, within, or between cells.

Searching for Post-Translational Modifications

This chapter describes how to search for peptides containing amino acids with post-translation modifications, such as phosphorylation or glycosylation, and how to interpret the results.

Contents

Using the ptmRS Node

The ptm*RS* node in the processing workflow provides a confidence measure for the localization of phosphorylation and any other modifications in peptide sequences. You can use it with all common fragmentation techniques, such as CID, ETD, and HCD, and peptide identifications from all available database search engines. This node uses parallelized calculations with multiple processor cores, when available, to improve performance. It also determines the optimal number of peaks to consider for localization of PTM sites for each *m/z* window individually, which increases the sensitivity of site localization for CID data. Depending on the applied fragmentation technique, the ptm*RS* node uses different fragment ion types for scoring to provide the highest sensitivity. For CID data, it scores only singly and doubly charged b and y ions. For analysis of HCD spectra, it also considers neutral loss ions. In contrast, when localizing PTM sites in ETD spectra, the node exclusively considers singly charged c, z, and y+H ions.

The input to ptm*RS* is a list of filtered MS/MS fragment spectra with corresponding peptide identifications.

How the ptm*RS* Node Calculates Sequence and Site Probabilities

The calculation of both ptm*RS* sequence probabilities and ptm*RS* site probabilities is based on the assumption that if the sequence of amino acids assigned to the respective MS/MS spectrum is correct and the MS/MS spectrum acquired did not result from fragmentation of two co-eluting peptides with similar *m/z* ratios, one of the putative isoforms must be the true assignment. Therefore, the sum of all sequence probabilities must equal 100 percent. When one or both of these assumptions is not correct, the calculated probabilities might not correlate with the true probabilities for correct modification localization. Figure 223 shows the basic internal workflow of the ptm*RS* node for phosphorylation.

Figure 223. Internal ptm*RS* workflow example for phosphorylation
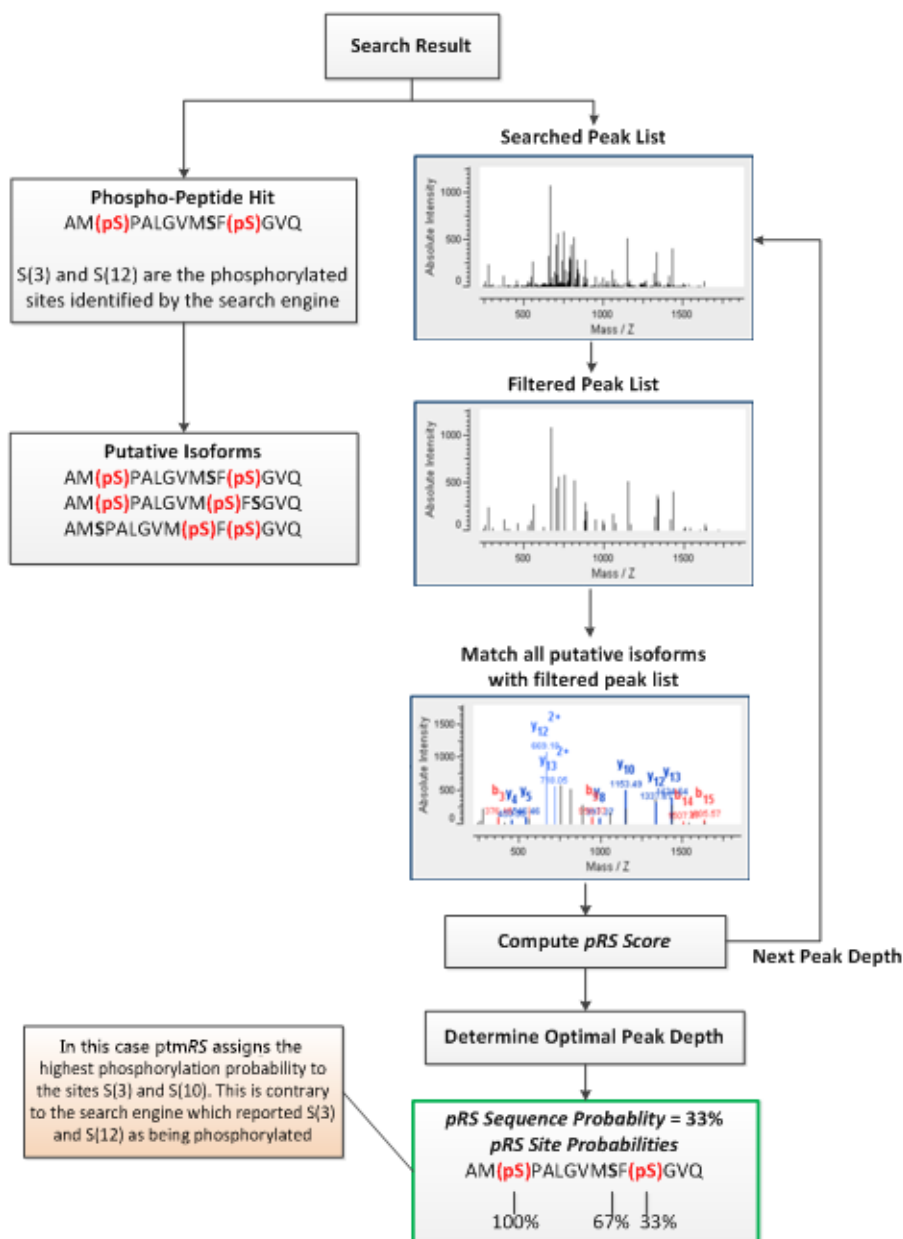
The ptm*RS* node performs the following steps:

1. On the basis of all potential modification sites in the peptide sequence, ptm*RS* calculates all putative phosphorylation isoforms.

2. ptm*RS* divides the fragment spectrum into 100-Th windows and extracts the i most intense peaks for each window, where i is the peak depth used for the filtered fragment spectrum.

3. For each phosphorylation isoform, S, and each peak depth, i, ptm*RS* calculates the list of theoretical fragment ions and matches them to the extracted peaks of the experimental spectrum.

   The probability, p, of matching a fragment peak purely by chance is calculated as follows:

   $$p = \frac{N_{Peaks} \times d}{w}$$

   where:

   - N_{Peaks} is the number of extracted peaks.

   - d is the mass tolerance for matching peaks to the theoretical fragment ions.

   - w is the extracted mass range.

4. From the number of theoretical fragment ions, n, the number of matched fragment peaks, kS, and the probability, p, ptm*RS* calculates the probability, $P_{S,i}$ of matching kS or more peaks purely by chance as the cumulative binomial probability of matching kS or more peaks in n attempts[2,3]:

   $$P_{S,i}(X \geq k_s) = \sum_{k=k_S}^{n} \binom{n}{k} p^k \times (1-p)^{(n-k)}$$

5. ptm*RS* reports the binomial peptide score of a specific isoform at peak depth, i, as follows:

   $$pRS_{S,i} = 10 \times \log_{10}(P_{S,i}(X \geq k_S))$$

6. ptm*RS* determines the optimal peak depth $i_{optimal}$ as the peak depth with the largest difference between the best-scoring phosphorylation isoform and the second-best-scoring isoform[2].

7. From the binomial peptide score at the optimal peak depth, ptm*RS* calculates an isoform confidence sequence probability and modification site probabilities. It calculates the ptm*RS* sequence probability as follows[3]:

   $$ptmRS\ sequence\ probability = \frac{P^{-1}_{S,i_{optimal}}}{\sum_S P^{-1}_{S,i_{optimal}}}$$

---

[2] Beausoleil S.A., et al. *Nat. Biotech.* **2006**, *Volume 24*: 1285-1292.

[3] Olsen J.V., et al. *Cell.* **2006**, *127*: 635-648.

Table 25 gives an example showing how ptm*RS* calculates sequence probabilities.

Table 25. Calculating ptm*RS* sequence probabilities

Putative isoform	pRS score	1/P value	pRS sequence probability
AM(pS)PALGVM(pS)FSGVQ	121	1.26×10^{12}	$1.26 \times 10^{12}/1.89 \times 10^{12=0.67}$
AM(pS)PALGVMSF(pS)GVQ	118	6.31×10^{12}	0.33
AMSPALGVM(pS)F(pS)GVQ	59	7.94×10^{12}	0.00
		$\Sigma = 1.89 \times 10^{12}$	$\Sigma = 1.00$

8. The ptm*RS* node calculates the ptm*RS* site probability, for example, for a particular phosphorylation site by summing up the ptm*RS* sequence probabilities of those isoforms where the respective site is phosphorylated[4]. Table 26 gives an example showing how ptm*RS* calculates site probabilities.

Table 26. Calculating ptm*RS* site probabilities

pRS sequence probability	Putative isoform						
0.67	AM	(pS)	PALGVM	(pS)	F	S	GVQ
0.33	AM	(pS)	PALGVM	S	F	(pS)	GVQ
0.00	AM	S	PALGVM	(pS)	F	(pS)	GVQ
ptm*RS* site probabilities		1.00		0.67		0.33	

Creating a PTM Analysis Workflow with the ptm*RS* Node

To focus on studying the biologically relevant post-translational modifications of proteins, you can create a PTM analysis workflow. The processing workflow must include the ptm*RS* node (refer to the Help), which generates all PSM-specific modification site data. The Peptide in Protein Annotation node (refer to the Help) in the consensus workflow provides additional information about the modifications on the protein level. For in-depth PTM analysis, use both nodes together.

---

[4] Olsen, J., et al. *Cell.* **2006**, *127*: 635–648.

The ptm*RS* node calculates modification site probabilities and makes them available in the Protein Identification Details view when you choose View > Protein Details on the Proteins page of the .pdResult file. This view color-codes the found modification above the amino acid sequences to indicate the probability of the modification being found on those portions of the amino acid. The PTM Site Probabilities area to the left of the sequence table displays a legend explaining the color-coding. For more information on this view, see "Viewing PTM Information on the Protein Identification Details View" on page 338.

For information on the parameters of the ptm*RS* node, refer to the Help.

❖ **To create a basic PTM analysis processing workflow**

1. Create or open a study and an analysis:

    • To create a study, see "Creating a Study" on page 40.

    • To open an existing study, see "Opening an Existing Study" on page 40.

    • To create an analysis, see "Creating an Analysis" on page 65.

    • To open an existing analysis, see "Opening an Existing Analysis" on page 66.

2. Create a basic processing workflow, as described in "Creating a Processing Workflow" on page 68, by using the following nodes:

    • Spectrum Files

    • Spectrum Selector

    • Search engine node, such as Sequest HT or Mascot

    • PSM Validation node, such as Fixed Value PSM Validator, Percolator, or Target Decoy PSM Validator

3. Attach a ptm*RS* node to the PSM validator node (in this case the Target Decoy PSM Validator node), as shown in Figure 224.

Figure 224. Basic PTM analysis processing workflow
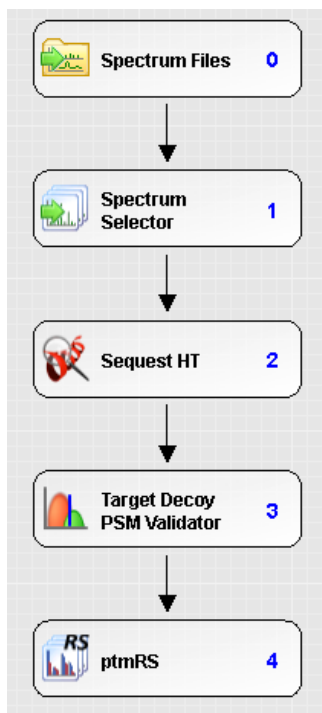
4. (Optional) Save the workflow:

   a. In the Name box above the Workflow Tree pane, type a name for the processing workflow.

   b. (Optional) In the Description box above the Workflow Tree pane, type a brief description of the processing workflow.

   c. In the Workflow Editor, click Save or Save Common .

   d. In the Save Workflow dialog box, do the following:

      i. Browse to the file to save the template in, or type a file name in the File Name box.

      ii. In the Save As Type box, select **Processing Workflow File (*.pdProcessingWF)**.

      iii. Click **Save**.

   The application saves the workflow in the *file_name*.pdProcessingWF file.

❖ **To create a basic PTM analysis consensus workflow**

1. Follow the general instructions for creating a consensus workflow in the Workflow Editor. See "Incorporating an Existing Workflow into a Study and an Analysis" on page 99.

2. Include the following nodes in the consensus workflow:

   • MSF Files node

   • PSM Grouper node

- Peptide Validator node

- Peptide and Protein Filter node

- Protein Scorer node

- Protein Grouping node

- Peptide in Protein Annotation node

Figure 225 shows the nodes required to create a basic PTM analysis consensus workflow.

Figure 225. Basic PTM analysis consensus workflow
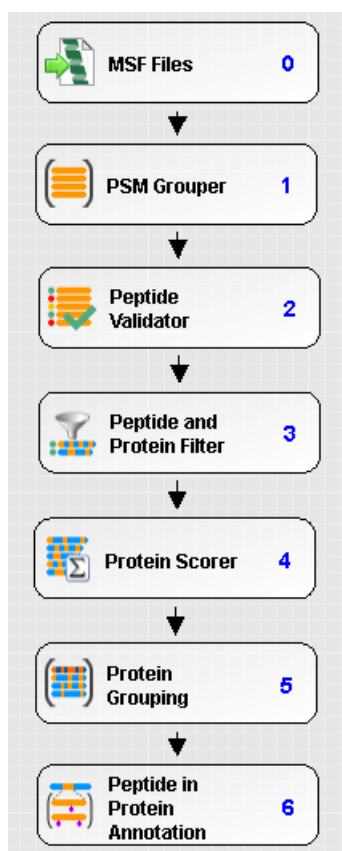
3. (Optional) Add any other appropriate nodes.

4. Connect the nodes together. Connect the Peptide in Protein Annotation node to the Protein Grouping node, if it not already connected.

5. Set the parameters for each node.

6. (Optional) Save the workflow:

   a. In the Name box above the Workflow Tree pane, type a name for the processing workflow.

   b. (Optional) In the Description box above the Workflow Tree pane, type a brief description of the processing workflow.

    c.   In the Workflow Editor, click ![Save icon] Save or ![Save Common icon] Save Common .

    d.   In the Save Workflow dialog box, do the following:

        i.   Browse to the file to save the template in, or type a file name in the File Name box.

        ii.   In the Save As Type box, select **Consensus Workflow File (*.pdConsensusWF)**.

        iii.   Click **Save**.

    The application saves the workflow in the *file_name*.pdConsensusWF file.

7.   Save the study. See "Saving a Study" on page 56.

8.   Click ![Run icon] Run in the upper right corner of the Analysis window.

    The job queue appears, as shown in Figure 53 on page 83, displaying the status of your search as the search progresses.

Columns in the Results Report

The PSMs page of the .pdResult report displays the primary results of the ptm*RS* processing. The node adds the following new columns to the report. These columns report the probabilities for single modifications, score, and isoform confidences.

- ptmRS Best Site Probabilities column: Displays the most likely positions of the modifications and their site score for each PSM. For each modification site, this value is an estimate of the probability (0–100%) that the site is truly modified. Any ptm*RS* site probabilities above 75% indicate that a site is truly modified. For peptide groups, this column shows the best site probabilities of the first PSM. This column is visible by default.

- ptmRS *Modification* Site Probabilities: Displays the modification site probabilities for all possible modification positions for a single modification. In the column title, *Modification* is replaced by the name of the actual modification, for example, Oxidation Site Probabilities.

- ptmRS Binomial Peptide Score: Reflects the quality of the match between a modification position isoform and the respective tandem spectrum. This peptide score is based on the cumulative binomial probability that the observed match is a random event. The value of the binomial peptide score heavily depends on the data scored, but usually scores above 50 indicate a good match. You can find details on the calculation of this score elsewhere.[5]

- ptmRS Isoform Confidence Probability: Displays the estimate of the probability (0–100%) that an individual isoform is correct. This estimate assumes that the search engine correctly identified the corresponding peptide in terms of amino acid sequence, as well as the number and identity of the modifications.[4]

---

[5] Taus, T. et al., *J. Proteome Res*, **2011**, 5354–5362

• ptmRS Isoform Group Report: Displays the confidence estimate of an individual isoform. This estimate assumes that the search engine correctly identified the corresponding peptide in terms of amino acid sequence, as well as the number and identity of the modifications.

Figure 226 shows the columns that the ptm*RS* node adds to the PSMs page of the .pdResult report.

Figure 226. Columns added to the PSMs page by the ptm*RS* node

Figure 227 shows another example of the results of ptm*RS* processing on the PSMs page. The top of Figure 227 shows the ptmRS Best Site Probabilities column. The bottom of the figure shows all the ptmRS *Modification* Site Probabilities columns for oxidation, phosphorylation, and acetylation, along with the ptmRS Binomial Peptide Score column, the ptmRS Isoform Confidence Probability column, the ptmRS Isoform Group report column, and the ptmRS Best Site Probabilities column.

Figure 227. ptm*RS* modification site probability score columns on the PSMs page

ptmRS *Modification* Site
Probabilities columns

On the Peptide Groups page, the PSM Grouper node places the modification site scores in the Modifications column, as shown in Figure 228. The application uses the modification site scores to select the best modification sites.

Figure 228. ptm*RS* modification site probability scores in the Modifications column on the Peptide Groups page

On the Proteins page, the Peptide in Protein Annotation node creates modification site scores in the Modifications column, as shown in Figure 229.

Figure 229. ptm*RS* modification site probability scores in the Modifications column on the Proteins page

In the Protein Identification Details view, the application displays the modification site probabilities calculated by the ptmRS node as color-coded modification abbreviations above the sequence, as shown in Figure 230.

Figure 230. Modification site probabilities shown as color-coded modification abbreviations above the sequence in the Protein Identification Details view
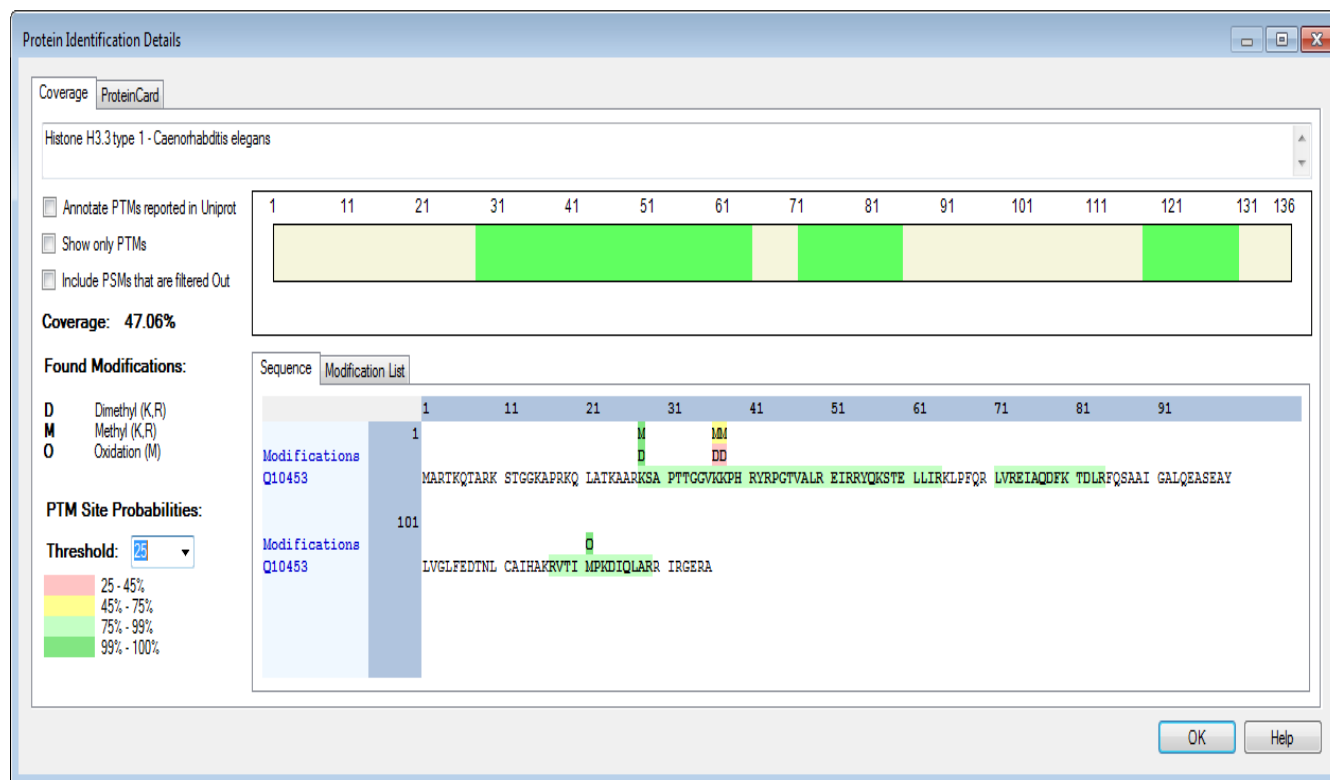
Methylation

Methylation is a PTM that can occur more than once on certain amino acids. For example, lysine might be modified by one, two, or three methyl groups. Search engines like Sequest HT or Mascot use three different modifications (methyl, dimethyl, and trimethyl) to denote these three modifications.

When there is more than one possible modification target, the methyl groups might be distributed differently over the sequence. For example, a peptide with a dimethyl modification on lysine might actually carry two single methyl groups on two lysine residues next to each other (both peptides have the same precursor mass in the MS1 scan). The ptmRS node calculates site probabilities for these different modification isoforms. After the ptmRS modification site scoring, a peptide with a dimethyl modification identified by the search engine might have two single methyl modifications. The ptmRS scores are the preferred criteria for the modification display for peptide groups and proteins.

Using the Peptide in Protein Annotation and the PSM Grouper Nodes

Proteomics results contain data that is determined by the position of an identified peptide in a protein sequence, such as the position of the identified peptide in a protein itself, derived data like the flanking residues, and the positions of modifications in the protein sequence.

When you search for proteins with a given modification, you must have the modification information (that is, type and position) directly available on the Proteins page of the .pdResult file. When your research focuses on PTMs, you might work mainly with the Peptide Groups page. In this case, you might want to know the position of the modification in the proteins containing the peptide. You might also need to identify the flanking residues when the digestion of the protein in the sample preparation breaks potential modification sites into two pieces that you can no longer identify.

Consider an example involving flanking residues. Enzymes add biologically relevant PTMs to the protein inside the cell. These enzymes bind to the target protein only on a defined short sequence called the motif. Suppose that you are looking for potential glycosylation sites with the NxS/T sequence motif. The glycosylation occurs only in the local sequence Asn-*Xaa*-Ser/Thr/Cys, where *Xaa* is a random amino acid except proline. *Xaa* might be arginine or lysine and therefore a cleavage site of trypsin. When the glycosylated peptide contains two asparagine residues, one in the middle of the sequence and one at the position before the last amino acid, you must use the flanking residues of the peptide to decide whether the asparagine residue at the end of the sequence is a valid glycosylation site. If so, the flanking residue must be either serine, threonine, or cysteine.

You can use the Peptide in Protein Annotation node in the consensus workflow to add columns to the Proteins and Peptide Groups pages of the results report that display the positions of modifications in the protein sequence. For detailed information about the parameters of these nodes, refer to the Help.

Peptide Positions in Proteins

The Protein Positions for Peptides parameter of the Peptide in Protein Annotation node in the consensus workflow determines whether the Proteome Discoverer application displays the peptide or PSM position for all proteins or only for the master proteins. When you set this parameter to For All Proteins, the application adds a Position in Proteins column to the Peptide Groups page that displays the position of the peptides or PSMs for all proteins that it finds the peptide or PSM in. It displays these positions in the following form:

protein accession [*start - end*]

When you set the parameter to Only for Master Proteins, the application adds a Positions in Master Proteins column to the Peptide Groups page that displays the position of the peptides or PSMs only for the master protein of each protein group. Figure 231 shows this column.

Figure 231. Positions in Master Proteins column on the Peptide Groups page

On the Peptide Groups page, there is often more than one protein for each peptide. You can click Show Associated Tables to display tables that show distinct peptide-protein relations, so the position of the peptide in the protein is unambiguous and displayed without protein accession. When a peptide occurs more than once in a protein, the tables list both ranges.

Flanking Residues

The flanking residues of a peptide are the amino acids that border a cleavage site. The Annotate Flanking Residues of the Peptide parameter of the Peptide in Protein Annotation node determines whether the Proteome Discoverer application annotates the amino acids before and after the peptide in the protein sequence. When you set this parameter to True, the application replaces the Sequence column on the Peptide Groups page by an Annotated Sequence column that displays the one-letter code representing the flanking residues of a peptide in a protein. Figure 232 shows the Annotated Sequence column on the Peptide Groups page.

Figure 232. Annotated Sequence column on the Peptide Groups page

#	Checked	Confidence	Annotated Sequence	Modifications	Modifications in Proteins	# Protein Groups	# Proteins	# PSMs	Master Protein Accessions	Positions in Master Protein
1		●	[K].LGEMWNNTAADDKQPYEK.[K]			1	5	1	P10103	P10103 [129-146]
2		●	[K].HPDSSVNFAEFSK.[K]			1	4	1	P17741	P17741 [31-43]
3		●	[R].KTVTAMDVVYALKR.[Q]			1	19	4	Q6LAF1; P84043	Q6LAF1 [80-93]; P84043
4		●	[R].KTVTAMDVVYALKR.[Q]		P02310 1×Oxidation [M85(100)]; Q41811 1×Oxidatic	1	19	6	Q6LAF1; P84043	Q6LAF1 [80-93]; P84043
5		●	[K].KNPEVPVNFAEFSKK.[C]			1	4	2	P40618	P40618 [30-44]
6		●	[K].ADRDESSPYAAMLAAQDVAQR.[C]		Q9CAX6 1×Oxidation [M74(100)]; P19950 1×Oxidati	1	6	1	P13471	P13471 [64-84]
7		●	[R].SKEEAEALYHSKYEELQVTVGR.[H]			1	1	1	P35908	P35908 [369-390]
8		●	[R].KTVTSMDVVYALKR.[Q]			1	1	2	P27996	P27996 [80-93]
9		●	[R].LLELDPPPKDVEDR.[G]			1	1	1	P19661	P19661 [57-70]
10		●	[R].KEEDPKTASQSLLVNLR.[A]			1	2	1	Q60611	Q60611 [411-427]
11		●	[R].SAGKYDVYLINPQGK.[A]			1	4	1	Q00566	Q00566 [116-130]
12		●	[K].VQASLAANTFTITGHAETK.[Q]			1	2	1	P20290	P20290 [123-141]
13		●	[K].SKAEAESLYQSKYEELQITAGR.[H]			1	1	1	P04264	P04264 [365-386]
14		●	[K].NQPGKYSQLVVETIR.[R]			1	1	2	Q92522	Q92522 [43-57]
15		●	[K].FADEHVPGSPFTVK.[I]			1	2	1	O75369	O75369 [2075-2088]
16		●	[R].IKDDPEDDGYFAPPKEDIKPLK.[R]	1×Methyl [K2(100)]	Q04750 1×Methyl [K119(100)]	0	1	1	Q04750	Q04750 [118-139]
17		●	[R].LILPGELAKHAVSEGTK.[A]			2	88	6	P07794; P10854; Q96A08; P0C1H3; Q8CGP0	P07794 [99-115]; P10854
18		●	[R].AQTLPTSVVTITSESSPGKR.[E]			1	2	1	Q01082	Q01082 [2326-2345]
19		●	[R].TPEVDDEALEKFDK.[A]			1	2	1	P02755	P02755 [143-156]
20		●	[R].KESYSIYVYKVLK.[QR]			1	14	1	P0C1H3; Q8CGP0	P0C1H3 [35-47]; Q8CGP
21		●	[R].LVREIAQDFKTDLR.[F]			2	32	1	P68429; Q10453; P59169; P68432	P68429 [71-84]; Q10453
22		●	[R].LAREIAQDFKTDLR.[F]	1×Dimethyl [R3(100)]	Q5RCC9 1×Dimethyl [R73(100)]	1	1	1	Q5RCC9	Q5RCC9 [71-84]
23		●	[K].TYFPHFDLSHGSGQIK.[A]			0	1	1	P07414	P07414 [41-56]
24		●	[K].TYFPHFDLSHGSGQVK.[GA]	1×Methyl [K16(100)]	P01954 1×Methyl [K56(100)]; P01936 1×Methyl [K5(	0	2	1	P01936	P01936 [41-56]
25		●	[K].TYFPHFDLSHGSAQVK.[GAD]			1	70	1	P01966	P01966 [42-57]
26		●	[R].SLETENSALQLQVTEREEVR.[G]			1	3	1	P14733	P14733 [53-72]
27		●	[KR].TVTAMDVVYALKRQGR.[T]		P02310 1×Oxidation [M85(100)]; Q41811 1×Oxidatic	1	20	1	Q6LAF1; P84043	Q6LAF1 [81-96]; P84043
28		●	[K].AQNDLIWNVKDELKK.[A]			1	1	1	P18493	P18493 [243-257]
29		●	[K].ITFRPDSADGMLLYNGQK.[R]		P98160 1×Oxidation [M3700(100)]; Q05793 1×Oxida	1	2	1	P98160	P98160 [3690-3707]
30		●	[K].LAAAILGGVDQIHIKPGAK.[V]			1	3	1	P22509	P22509 [150-168]
31		●	[R].VDVLPVNLPGEHGQR.[L]			1	3	2	P04937; P07589	P04937 [938-952]; P0758
32		●	[K].LASVPAGGAVAVSAAPGSAAPAAGSAI			1	4	1	P05387	P05387 [62-94]
33		●	[R].IKDEPEDDGYFAPPKEDIKPLK.[R]			1	1	1	Q9WUL0	Q9WUL0 [118-139]
34		●	[R].GENLVSMTVEGPPPKDTGIAR.[V]		P27048 1×Oxidation [M80(100)]; P17136 1×Oxidatic	1	10	1	P14678	P14678 [74-94]
35		●	[K].VKEPSVQEATSTSDILK.[V]			1	2	1	Q5R5Q2	Q5R5Q2 [230-246]
36		●	[K].KLGEMWNNTAADDKQPYEK.[K]		Q9UGV6 1×Oxidation [M132(100)]; P63158 1×Oxida	1	5	2	P10103	P10103 [128-146]
37		●	[K].KHPDASVNFSFESKK.[C]				4	4	P10103	P10103 [30-44]

Show Associated Tables

Ready — 139 Protein Groups; 1093 Proteins; 360 Peptide Groups; 707 PSMs; 2893 MS/MS Spectrum Info

You can click Show Associated Tables to display pages that show the positions of the peptide in the proteins and the flanking residues. You can use these pages to obtain detailed information about the flanking residues of the peptide in different proteins. To specify the number of flanking residues to display in the associated pages, use the Number Flanking Residues in Connection Tables parameter of the Peptide in Protein Annotation node. You can display up to five flanking amino acids in the associated pages.

The application displays the flanking residues in brackets before and after the peptide sequence and separates them by a dot in the sequence, as shown in the first sequence in Figure 233. When the application finds a peptide more than once in a protein or when it finds a peptide in more than one protein, it might display more than one flanking residue. In these cases, it displays all possible residues within a pair of brackets, which are commonly used to display a sequence consensus.

Figure 233 shows different flanking residues for two proteins in brackets. The sequences compared in the bottom of the view show the residues, which are identical, for the two proteins, in green. The flanking residues are different on both sides of the peptide, so there are gaps (since different residues are not aligned) or alternative shading to indicate the difference between the two sequences.

Figure 233. Peptide group consensus sequence with different flanking residues from two proteins

Consensus sequence showing flanking residues in brackets

Two proteins

Full protein sequences of the different flanking residues shown at top in brackets

Modifications of Proteins on the PSMs Page

The Proteome Discoverer application displays the positions of the identified modifications in a protein the same way that it displays the positions of a peptide in a protein. The way it displays peptide modifications is described in the Help.

You can display modification positions for all proteins or only for master proteins.

When you select the Peptide in Protein Annotation node in the consensus workflow and set its Protein Modifications Reported parameter to For All Proteins, the application adds a Modification in Proteins column to the Peptides Group page that displays the positions of the modifications in all proteins.

When you set the node's Protein Modifications Reported parameter to For Master Proteins, the application adds a Modification in Master Proteins column to the Peptides Group page that displays the modifications in only the master proteins.

The application always creates two different modification columns. One modification column contains only the best possibilities for the modifications. The other contains all possible positions. The Modification Sites Shown parameter of the PSM Grouper node determines whether the application shows one or both of these columns. The column not shown is still available, and you can display it by using the Field Chooser (refer to the Help). For detailed information on the parameters of the PSM Grouper node, refer to the Help and the decision tree for the modification display (Figure 236 on page 335) and the accompanying examples.

You can also use the PSM Grouper node's Site Probability Threshold parameter to determine whether the application should display only sites with a minimal site probability score.

Modifications of Proteins on the Proteins Page

The Modification Sites Reported parameter of the Peptide in Protein Annotation node determines which modifications the application annotates on the Proteins page of the .pdResult file. These are the options:

All Combined—The application annotates all modifications and displays the annotations in a single column called Modifications.

Only Specific—The application annotates only the modifications specified by the modification parameters that reside below the Modification Sites Reported parameter in the same section and displays two columns for each specified modification as follows:

- The first column, named by the modification name, shows the positions of all identified modifications of this kind in the protein.

- The second column, also named by the modification name plus an additional count, contains the total count of this kind of modification in each protein.

All and Specific—The application annotates all identified modifications of the protein and displays the annotations in a single column called Modifications. It also annotates the specified modifications and displays the positions and count in single columns.

You can restrict the concatenated list of modifications in the Modification column on the Proteins page. Use the Site Probability Threshold parameter of the PSM Grouper node and the Report Only PTMs parameter of the Peptide in Protein Annotation node to show only modification sites above a given probability threshold or modifications that are PTMs.

You can also use the Peptide in Protein Annotation node's Minimum PSM Confidence parameter to restrict the modifications shown to those with a minimal confidence level of the PSMs that identified the modifications.

Depending on the protein, the list of modifications can become long. When you are interested in only one type of modification, use the Modification parameter (under Modifications in Protein) to display a set of five (maximum) selected modifications in extra columns. For each of these specified modifications, the application creates two new columns. One column contains the modifications and all sites where the modification is found, and the second column shows how many times this modification occurs in the protein.

Using phosphorylation as an example, Figure 234 shows in the Modifications column all modifications and the sites where the modifications are found. It also shows the number of modifications in the Phospho Count column and the separate modifications in the Phospho Positions column.

Figure 234. Phosphorylation modifications on the Proteins page

Modifications of Peptides on the Peptide Groups Page

When you use the PSM Grouper node in the consensus workflow, the Proteome Discoverer application adds columns of information about modifications to the Peptide Groups page of the .pdResult file. For information on the Peptide Groups page, refer to the Help.

The content of the Modifications column shown in the Peptide Groups page is a merged list of all the modifications found in the PSMs belonging to the peptide group. Figure 235 shows the format of each kind of modification.

Figure 235. Modification format

2xPhospho[S4(98);T7(97.6)]

Site probability, when available

Target amino acid and position in the peptide

Type of modification as shown in the Chemical Modifications view

Number of modifications of this type in the peptide
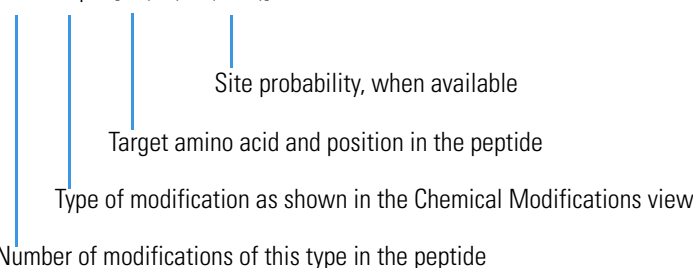
The PSM Grouper node creates two different modification columns. One column contains all possible modification sites, and the other one contains only the best possible site for each modification. When a modification has no unique "best" position, the application displays the modification with the target amino acid but without specifying the best position. In this case, you must review the PSMs of the peptide group by clicking Show Associated Tables on the results pages. Depending on the setting of the Modification Sites Shown parameter of the

node, the application displays one or both of these columns. If one of the columns is invisible, you can display it by using the Field Chooser in the upper left corner of the result page. You can restrict the display of modifications in the Peptide Groups page through the parameter settings of the PSM Grouper node. For information on these parameters, refer to the Help.

Figure 236 shows the decisions that determine how the application displays a modification on the Peptide Groups page.

Figure 236. Decision tree determining the display of modifications on the Peptide Groups page
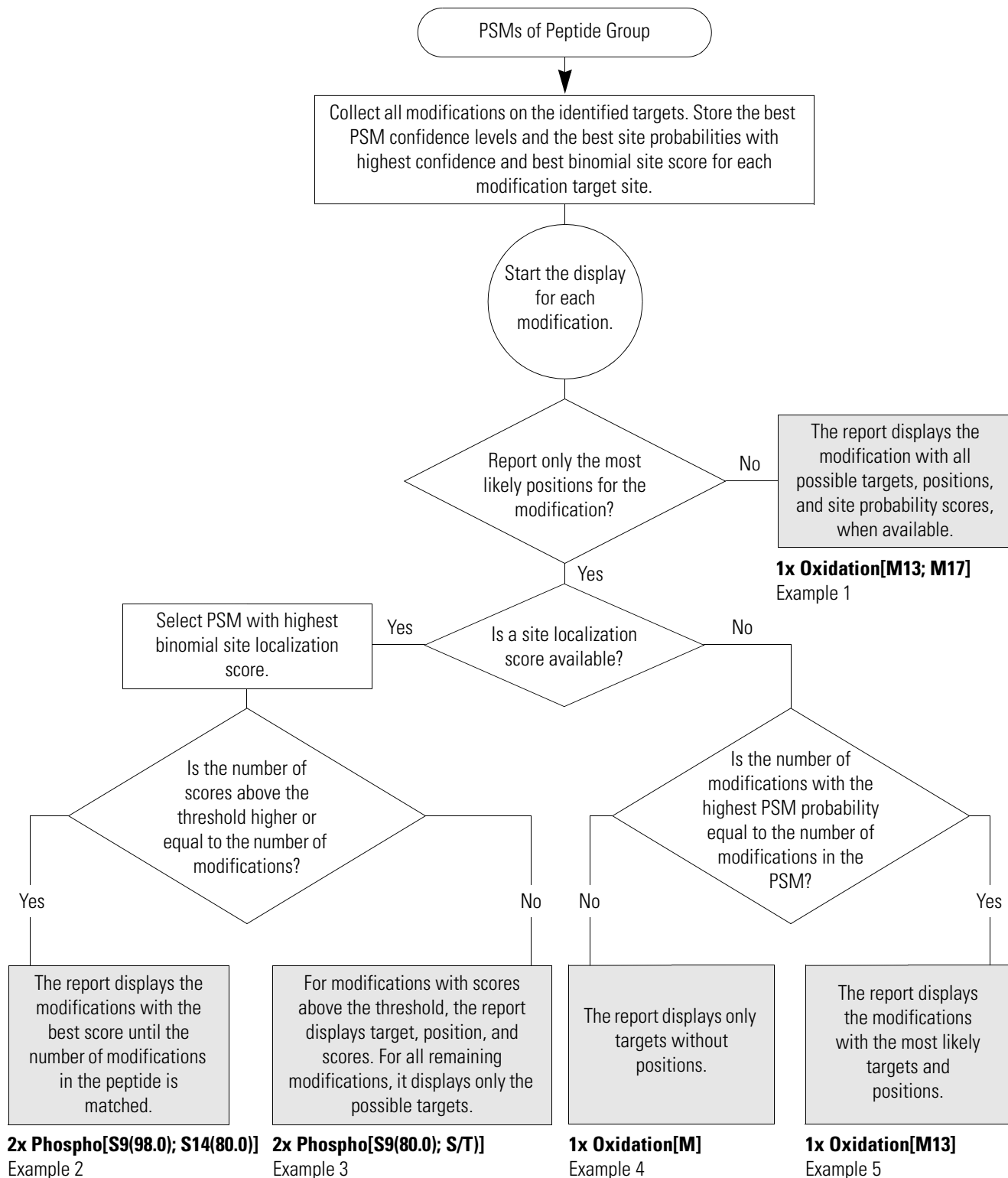
The following examples illustrate how the application displays different modifications.

Example 1

In this example, there is one modification with different positions and different confidences.

PSM number	PSM confidence level	Modified PSM sequence
PSM 1	🟢	APEPTIDEWITHM(Ox)ANYMETHIQNIN
PSM 2	🟢	APEPTIDEWITHM(Ox)ANYMETHIQNIN
PSM 3	🟡	APEPTIDEWITHMANYM(Ox)ETHIQNIN

The application displays the modification as 1x Oxidation[M13; M17] in the Modifications (All Possible Sites) column of the Peptide Groups page and 1x Oxidation[M13] in the Modifications column, which displays the best modification sites.

Example 2

In this example, the site probabilities are available, and enough scores are above the threshold.

The peptide has two phosphorylations (from the precursor mass).

PSM number	Modified PSM sequence
PSM 1	APEPTIDES(Phospho 80.0)WITHS(Phospho 80.0)ERANDT(Phospho 40.0)HR Binomial Score: 112
PSM 2	APEPTIDES(Phospho 98.0)WITHS(Phospho 76.0)ERANDT(Phospho 36.0)HR Binomial Score: 42

The application displays the modifications as 2x Phospho[S9(80.0) and S14(80.0)] in the Modifications column and 2xPhospho [S9(80.0), S14(80.0), T20(40)] in the Modifications (All Possible Sites) column.

Example 3

In this example, the site probabilities are available, but not enough scores are above the threshold.

A peptide has two phophorylations (from the precursor mass). The Site Probability Threshold parameter of the PSM Grouper node is set to 75.0. See the next table.

PSM number	Modified PSM sequence
PSM 1	APEPTIDES(Phospho 80.0)WITHS(Phospho 60.0)ERANDT(Phospho 60.0) HR Binomial Score: 112
PSM 2	APEPTIDES(Phospho 98.0)WITHS(Phospho 66.0)ERANDT(Phospho 35.0) HR Binomial Score: 42

The application displays the modifications as 2x Phospho[S9(80.0); S/T)] in the Modifications column.

One phosphorylation has an 80% probability of being located on serine 9. The other phosphorylation might be located on any serine or threonine in the sequence. When no threonine is present, the string becomes phospho[S9(98.0); S)]. The Modifications (All Possible Sites) column contains 2xPhospho[S9(80);S14(60);T20(60)].

Example 4

In this example, the position of the modifications with the highest PSM confidence is not unambiguous.

PSM number	PSM confidence level	Modified PSM sequence
PSM 1	🟢	APEPTIDEWITHM(Ox)ANYMETHIQNIN
PSM 2	🟢	APEPTIDEWITHMANYM(Ox)ETHIQNIN
PSM 3	🟡	APEPTIDEWITHMANYM(Ox)ETHIQNIN

The sites with highest PSM probability are M13 and M17.

The mass of the precursor ion indicates that the peptide contains one modification. In the list of PSMs are two high-confidence (green) PSMs where the modification is in different positions. The application cannot determine which position of the modification is the correct one, so it reports no position.

The application displays the modifications as Oxidation[M] in the Modifications column and Oxidation[M13; M17] in the Modifications (All Possible Sites) column.

Example 5

In this example, the position of modifications with the highest PSM confidence is unambiguous.

PSM number	PSM confidence level	Modified PSM sequence
PSM 1	🟢	APEPTIDEWITHM(Ox)ANYMETHIQNIN
PSM 2	🟢	APEPTIDEWITHM(Ox)ANYMETHIQNIN
PSM 3	🟡	APEPTIDEWITHMANYM(Ox)ETHIQNIN

The site with the highest PSM probability is M13.

The mass of the precursor ion indicates that the peptide contains one modification. All high-confidence PSMs report the modification in the same position. In this case, there is a reasonable certainty that the modification is on the position identified in the high-confidence matches, so the application reports it.

The application displays the modifications as 1x Oxidation[M13] in the Modifications column and 1x Oxidation [M13; M17] in the Modifications (All Possible Sites) column.

Viewing PTM Information on the Protein Identification Details View

To view information on the Proteins page about the modifications found in the search, use the Protein Identification Details view. The following topics explain how to use this view to investigate PTMs. For basic information about the Protein Identification Details view, refer to the Help.

❖ **To display a protein sequence**

1. Open the .pdResult file.

2. Click the Proteins tab.

3. Double-click a grid cell on the Proteins page.

   –or–

   Select a grid cell and choose **View > Show Protein Details**.

   –or–

   Click the **Show Protein Details** icon, .

   The Protein Identification Details view opens, as shown in Figure 237.

Figure 237. Protein Identification Details dialog box

The Protein Identification Details view features two pages, Coverage and ProteinCard.

- For detailed information on the Coverage page, refer to the Help.

- For detailed information on the ProteinCard page, see "Accessing the ProteinCard Page" on page 297.

❖ **To display a protein's PTMs**

1. In the Protein Identification Details view, click the **Coverage** tab if it is not already selected.

2. Click the **Modification List** tab.

   The Modification List page displays all identified modifications in the sequence view. An example of the Modification List page is shown in the Help.

   For information on the columns in this page, refer to the Help.

3. In the Number of Amino Acids Before and After Modification box, specify the length of the sequence motif before and after the site of the PTM.

❖ **To show only PTMs**

1. In the Protein Identification Details view, click the **Coverage** tab if it is not already selected.

2. Select the **Show Only PTMs** check box.

This option displays only the PTMs on the Sequence and Modification List pages, so it helps you focus on biologically relevant modifications. It excludes all modifications classified as artifacts, chemical derivatives, isotopic labels, and synthetic peptide protection groups from the display.

❖ **To display filtered-out PSMs**

1. In the Protein Identification Details view, click the **Coverage** tab if it is not already selected.

2. Select the **Include PSMs That Are Filtered Out** check box.

   This option filters out PSMs that have been found but filtered out by the display filters.

❖ **To control the display of the probability of a PTM occurring on a site**

1. In the Protein Identification Details view, click the **Coverage** tab if it is not already selected.

   When the site probability scores are available (that is, when you included the ptm*RS* node in the processing workflow), the application color-codes the scores, and you can select the threshold.

2. From the Threshold list in the PTM Site Probabilities area, select the category of probability for a modification to occur on a sequence site.

   For example, when you choose 25 from the Threshold list, the Protein Identification Details view displays modifications that have a low (25–100%) probability of occurrence on a site, as shown in Figure 238.

   > **Note** The application does not display modifications with less than a 25% chance of occurring on a site.

Figure 238. Protein Identification Details view showing modifications with a low probability of occurring on a site
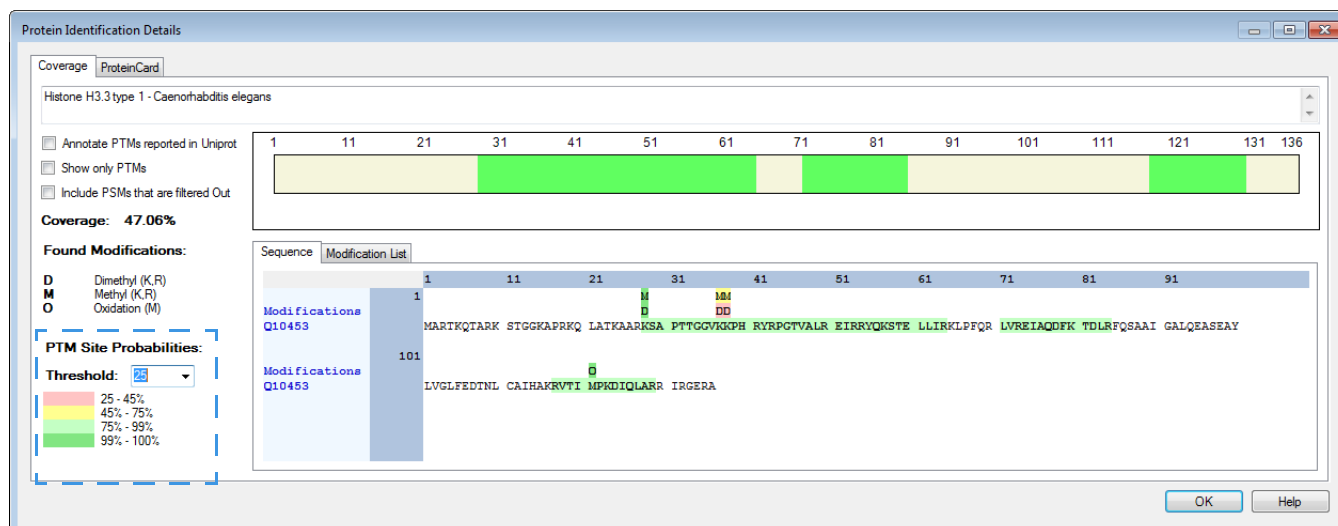
As another example, to display only those modifications with a very high probability (99–100%) of a modification occurring on a site, select 99 from the Threshold list, as shown in Figure 239.

Figure 239. Protein Identification Details view showing modifications with a high probability of occurring on a site
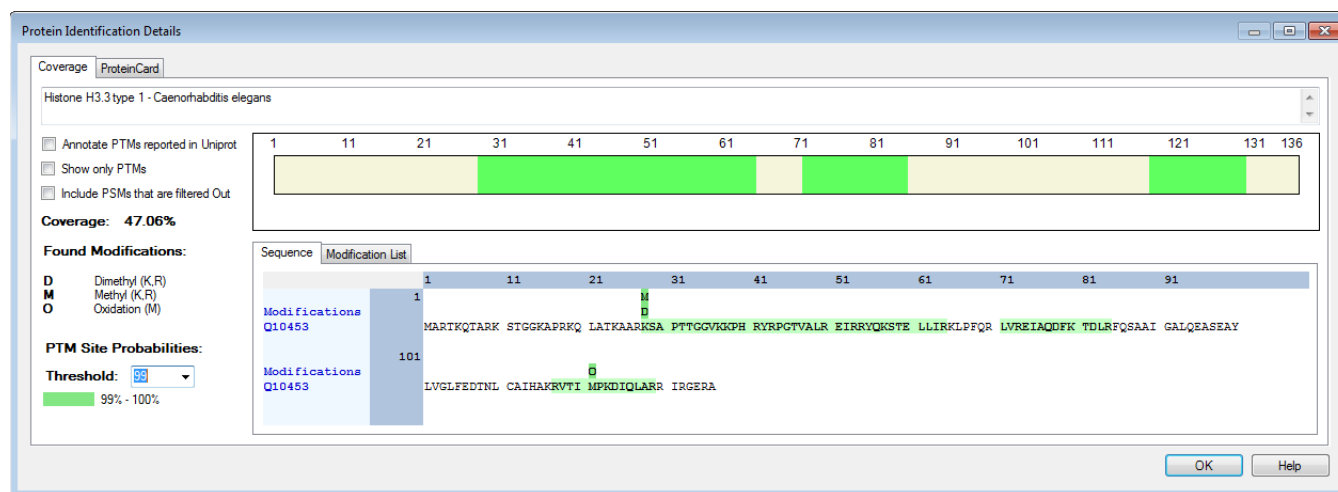
❖ **To display UniProt annotations of the protein**

1. Set up and run a protein annotation workflow that includes the ProteinCenter Annotation node (see "Creating a Protein Annotation Workflow" on page 265).

   This node retrieves UniProt database information, including PTMs, from ProteinCenter and saves it in Proteome Discoverer results files.

2. Open the resulting .pdResult file, select the protein of interest, and open the Protein Identification Details view.

3. Select the **Annotate PTMs Reported in Uniprot** check box.

   You can find out if the protein has modification information by clicking the ProteinCard tab and then clicking on the Features tab. All lines containing the MOD_RES or CARBOHYD keyword should be displayed as PTMs in the sequence.

   When annotation information is available, the modification appears as a single-letter abbreviation in black font in a small gray box beneath the amino acids of the protein in the sequence table. Information about the PTMs found in the UniProt database appear under the UniProt PTMs category in the area to the left of the sequence table. In addition, the likelihood of a PTM occurring on an amino acid appears under the PTM Site Probabilities heading in this same area. The Help illustrates this information.

❖ **To copy the colored bar on the Sequence page**

Right-click the bar and choose **Copy**.

❖ **To save the colored bar on the Sequence page**

1. Right-click the bar and choose **Save As**.

2. In the Save As dialog box, browse to the location where you want to save the file, and type the file name in the File Name box.

3. Click **Save**.

❖   **To export the Modification List page to Excel**

1. Right-click the Modification List table and choose **Export to Excel Workbook**.

2. In the Save Modification List as Microsoft Excel File dialog box, browse to the location where you want to save the Excel file, or specify the path and file name in the File Name box.

3. Click **Save**.

Filtering Phosphorylation Site Probabilities

You can filter phosphorylation site probability data by using display filters. For phosphorylation probabilities produced by using the ptm*RS* node in the processing workflow, you can use the Exceed operator in the Display Filters dialog box to display the following:

*   Any peptides that contain a phosphorylation of an arbitrary amino acid

*   All peptides that contain a phosphorylation of an arbitrary amino acid where each probability exceeds a given probability

*   Any peptides that contain a phosphorylation of an amino acid that exceeds a given probability value

You can combine these conditions with other filter conditions by using AND and OR.

For general information about filtering data, see "Filtering with Display Filters" on page 225.

For information about the PTM*RS* node, see "Using the ptmRS Node" on page 317 and the Help.

❖   **To filter phosphorylation site probabilities**

1. In a .pdResult containing phosphorylation modification data (see Figure 240), choose **View > Display Filter**, or click the **Display Filter** icon, 🔽 .

Figure 240. .pdResult file containing phosphorylation modification data in the ptmRS Best Site Probabilities column

2. In the pane on the left, select **Peptide Groups**.

3. In the Add Property box, select **ptmRS PhosphoSite Probabilities**.

4. In the operator box, select **Exceed** if it is not already selected.

   For Amino Acid appears in the values box.

5. Type the appropriate value in the values box to the right of the For Amino Acid box.

6. Click ✅ **Apply** .

   Figure 241 shows an empty filter that you can further configure in the three different ways shown in "Example 1," "Example 2," and "Example 3."

Figure 241. Empty Display Filters pane that addresses ptm*RS* probabilities
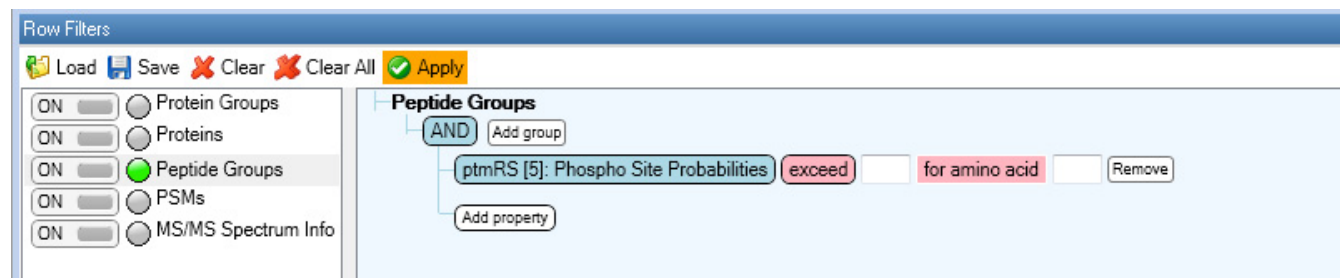
To reverse an addition or change immediately after you make it, click **Remove** next to it. For more information on using display filters, see "Filtering with Display Filters" on page 225.

Example 1

To show only those peptides that contain a phosphorylation of an arbitrary amino acid with positive probability, type **0** in the values box to the right of the operator box and **Any** in the values box to the right of the For Amino Acid box, as shown in Figure 242. Click .

Figure 242. Peptides that contain a phosphorylation of an arbitrary amino acid with positive probability

	Checked	Confidence	Sequence	Modifications	# Protein Groups	# Proteins	# PSMs	Master Protein Accessions	PhosphoRS: Best Site Probabilities
1	☐	●	LCDFGSASHVADNDITPYLVSR	1×Phospho [Y18(100)]	1	3	10	IPI00013721.3	Y18(Phospho): 100
2	☐	●	LRSYEDMIGEEVPSDQYYWAPLAQHER	1×Phospho [Y4(99.6)]	1	29	2	IPI00182540.7	Y4(Phospho): 99.56
3	☐	●	YDEEGGGEEDQDYDLSQLQQPDTVEPDAIKPV	1×Phospho [S16(99)]	1	2	2	IPI00290085.2	S16(Phospho): 99.03
4	☐	●	KISENSYSLDDLEIGPGQLSSSTFDSEK	1×Phospho [S3(98.8)]	1	7	4	IPI00883896.1	S3(Phospho): 98.8
5	☐	●	AYSQEEITQGFEETGDTLYAPYSTHFQLQNQPP	1×Phospho [S3(99.2)]	1	1	2	IPI00022624.1	S3(Phospho): 99.23
6	☐	●	RYSDLTTLVAFPSSSVYPTK	1×Phospho [S3(99.7)]	1	2	4	IPI00013214.2	S3(Phospho): 99.47
7	☐	●	REPGYTPPGAGNQNPPGMYPVTGPK	1×Phospho [T6(99.6)]	1	3	6	IPI00023704.1	T6(Phospho): 99.51
8	☐	●	REPGGWGAGASAPVEDDSDAETYGEENDEGG	1×Phospho [S/T/Y]	1	1	3	IPI00006904.1	S18(Phospho): 49.92; T22(Phospho): 49.92
9	☐	●	SYDVPPPPMEPDHPFYSNISK	1×Phospho [S/Y]	1	3	10	IPI00549725.6	S1(Phospho): 50; Y2(Phospho): 50
10	☐	●	HSSYPAGTEDDEGMGEEPSPFR	1×Phospho [S19(100)]	1	2	3	IPI00024291.2	S2(Phospho): 98.72
11	☐	●	SSSSGSDDYAYTQALLLHQR	1×Phospho [S/Y/T]	1	2	5	IPI00166840.4	S3(Phospho): 33.07; S6(Phospho): 33.07; Y9(Phospho): 33.07
12	☐	●	TRSYDNLTTACDNTVPLASR	1×Phospho [T/S/Y]	1	4	3	IPI00185018.7	T1(Phospho): 33.33; S3(Phospho): 33.33; Y4(Phospho): 33.33
13	☐	●	AFSSRSYTSGPGSR	2×Phospho [S3(100); S6(97.4)]	1	8	1	IPI01021414.1	S3(Phospho): 99.99; S4(Phospho): 33.34; S6(Phospho): 33.33; Y7(Phosp
14	☐	●	ADVQLFMDDDSYSHHSGLEYADPEK	1×Phospho [S11(99.5)]	1	2	8	IPI00019870.1	S11(Phospho): 99.52
15	☐	●	ADRDESSPYAAMLAAQDVAQR	1×Phospho [S7(99.1)]	1	3	3	IPI00026271.5	S7(Phospho): 99.11
16	☐	●	EGMNPSYDEYADSDEDQHDAYLER	2×Phospho [S13(100); S/Y]	1	1	2	IPI00005154.1	S6(Phospho): 50; Y7(Phospho): 50; S13(Phospho): 100
17	☐	●	RGLLYDSDEEDEERPAR	1×Phospho [S7(99.4)]	1	2	1	IPI00184330.5	S7(Phospho): 99.43
18	☐	●	YTLENKEEGSLSDTEADAVSGQLPDPTTNPSAC	1×Phospho [Y/T/S]	1	1	4	IPI00170935.1	Y1(Phospho): 19.95; T2(Phospho): 19.95; S10(Phospho): 19.95; S12(Ph
19	☐	●	GGPGSAVSPYPTFNPSSDVAALHK	1×Phospho [S/Y/T]	1	3	3	IPI00218918.5	S8(Phospho): 50; Y10(Phospho): 50
20	☐	●	TIDLGAAAHYTGDKASPDQNASTHTPQSSVK	1×Phospho [S16(100)]	1	3	3	IPI00975549.1	T11(Phospho): 49.69; S16(Phospho): 49.69
21	☐	●	AYSFCGTVEYMAPEVVNR	1×Phospho [S3(99.7)]	2	11	9	IPI00477982.1; IPI00020898.1	S3(Phospho): 99.67
22	☐	●	YCTETSGVHGDSPYGSGTMDTHSLESK	1×Phospho [Y14(99.2)]	1	2	3	IPI00018370.4	Y14(Phospho): 99.19
23	☐	●	QQAAYYAQTSPQGMPQHPPAPQGQ	1×Phospho [Y6(97.2)]	2	3	14	IPI00641948.4; IPI00983652.2	Y6(Phospho): 97.16

Example 2
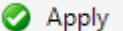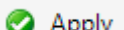
To show only those peptides that contain a phosphorylation of an arbitrary amino acid where each probability exceeds a given probability (90% in this example), type **90** in the box to the right of the operator box and **Any** in the values box to the right of the For Amino Acid box, as shown in Figure 243. Click ☑ Apply .

Figure 243. Peptides that contain a phosphorylation of an arbitrary amino acid where each probability exceeds a given probability

		Checked	Confidence	Sequence	Modifications	# Protein Groups	# Proteins	# PSMs	Master Protein Accessions	PhosphoRS: Best Site Probabilities
1		☐	●	LCDFGSASHVADNDITPYLVSR	1×Phospho [Y18(100)]	1	3	10	IPI00013721.3	Y18(Phospho): 100
2		☐	●	LRSYEDMIGEEVPSDQYYWAPLAQHER	1×Phospho [Y4(99.6)]	1	29	2	IPI00182540.7	Y4(Phospho): 99.56
3		☐	●	SHYADVDPENQNFLLESNLGK	1×Phospho [S17(100)]	1	2	2	IPI00410034.2	Y3(Phospho): 99.41
4		☐	●	IADPEHDHTGFLTEYVATR	1×Phospho [Y15(99.2)]	1	8	1	IPI00018195.3	Y15(Phospho): 99.17
5		☐	●	QISYNYSDLDQSNVTEETPEGEEHHPVADTENK	1×Phospho [Y6(98.2)]	1	2	11	IPI00410693.4	Y6(Phospho): 98.19
6		☐	●	IGEGTYGVVYK	1×Phospho [Y10(100)]	2	11	3	IPI01025117.1; IPI00026689.4	Y6(Phospho): 99.64
7		☐	●	YCTETSGVHGDSPYGSGTMDTHSLESK	1×Phospho [Y14(99.2)]	1	2	3	IPI00018370.4	Y14(Phospho): 99.19
8		☐	●	AELSYRGPVSGTEPEPVYSMEAADYR	1×Phospho [Y5(99.4)]	1	6	2	IPI01013939.1	Y5(Phospho): 99.39
9		☐	●	QQAAYYAQTSPQGMPQHPPAPQGQ	1×Phospho [Y6(97.2)]	2	3	14	IPI00641948.4; IPI00983652.2	Y6(Phospho): 97.16
10		☐	●	YMIGVTYGGDDIPLSPYR	1×Phospho [Y17(99.6)]	1	9	1	IPI00900293.1	Y17(Phospho): 99.59
11		☐	●	EELTPGAPQHQLPPVPGSPEPYPGQQAPGPEF	1×Phospho [T4(99.9)]	1	1	4	IPI00012199.1	Y22(Phospho): 98.36
12		☐	●	LIEDNEYTAR	1×Phospho [Y7(99.7)]	1	14	1	IPI00013981.4	Y7(Phospho): 99.66
13		☐	●	YTPTQQGNMQVLVTYGGDPIPK	1×Phospho [Y1(99.2)]	1	9	2	IPI00900293.1	Y1(Phospho): 99.2
14		☐	●	YSPSQNSPIHHIPSRR	2×Phospho [Y1(98.9); S7(100)]	1	10	3	IPI00006079.1	Y1(Phospho): 98.9; S7(Phospho): 100
15		☐	●	SESVVYADIR	1×Phospho [Y6(100)]	1	4	1	IPI00022558.2	Y6(Phospho): 100
16		☐	●	VADPDHDHTGFLTEYVATR	1×Phospho [Y15(100)]	1	3	2	IPI00003479.3	Y15(Phospho): 99.98
17		☐	●	VADPDHDHTGFLTEYVATR	2×Phospho [T13(100); Y15(100)]	1	3	3	IPI00003479.3	T13(Phospho): 100; Y15(Phospho): 99.99
18		☐	●	AAYGDLSSEEEEENEPESLGVVYK	2×Phospho [Y3(99); S/Y]	1	2	1	IPI00007343.1	Y3(Phospho): 99.01; S7(Phospho): 50.48; S8(Phospho): 50.48
19		☐	●	VIEDNEYTAR	1×Phospho [Y7(99.7)]	1	8	1	IPI00940936.1	Y7(Phospho): 99.68
20		☐	●	IGEGTYGVVYK	2×Phospho [T5(100); Y6(100)]	2	11	2	IPI01025117.1; IPI00026689.4	T5(Phospho): 100; Y6(Phospho): 100
21		☐	●	KPYRIESDEEEDFENVGK	1×Phospho [Y3(99.3)]	1	3	2	IPI00290652.4	Y3(Phospho): 99.33
22		☐	●	IFQGSFGGPTLYENPHYQSPNMHR	1×Phospho [Y12(97.3)]	1	1	1	IPI00181728.1	Y12(Phospho): 97.29
23		☐	●	GFYKGTGPALMAYVAENSVLFMCYGFCQQFVF	1×Phospho [Y13(96.3)]	1	1	1	IPI00011190.3	Y13(Phospho): 96.32

Example 3

To show only those peptides that contain a phosphorylation of a specific amino acid (serine) that exceeds a given probability (90% in this example), type **90** in the box to the right of the operator box and **S** in the values box to the right of the For Amino Acid box, as shown in Figure 244. Click ☑ Apply .

Figure 244. Peptides that contain a phosphorylation of an amino acid that exceeds a probability of 90%

Performing Reporter Ion Quantification

This chapter describes how to perform reporter ion quantification.

Contents

- Reporter Ion Quantification
- Performing Reporter Ion Quantification
- Calculating and Validating Raw Quantification Values in Reporter Ion Quantification
- Correcting Reporter Ion Quantification Results for Isotopic Impurities
- Filtering Quantification Data with Signal-to-Noise Values

Reporter Ion Quantification

Isobarically labeled reporter ion quantification methods use reagents, or tags, to enzymatically or chemically label proteins and peptides. Reporter ion quantification uses tags that have the same mass, but produce fragments with different mass in the MS/MS spectrum that can be used for quantification. (A reporter ion is a fragment ion produced from the MS/MS of a peptide with an attached tag.)

The application supports reporter ion quantification for Tandem Mass Tag (TMT), Isobaric Tag for Relative and Absolute Quantification (iTRAQ), and any user-defined tags. The MS/MS scan performs identification and quantification with both TMT and iTRAQ.

> **Note** User-defined tags require an additional license key provided by Thermo Fisher Scientific.

For iTRAQ, 4plex and 8plex default methods are available. For TMT, 2plex, 6plex, 10plex, and 11plex default methods are available. You can also add new methods by either deriving them from an existing method or creating new ones from the beginning.

The discussion of reporter ion quantification includes the following topics:

- TMT Quantification
- iTRAQ Quantification

TMT Quantification

TMT quantification is a reproducible, highly accurate quantification method that provides both comparative and absolute MS/MS-based quantification of proteins and peptides in biological samples. TMT tagging produces data to calculate the relative abundances of proteins. You can evaluate differential protein expression in two to eleven samples in a single experiment.

Each sample is labeled with chemically identical tags before mixing the samples, which generates a single isotope cluster for each peptide, irrespective of which tag it has been given. Between the normalizer and reporter is a cleavable linker that fragments during MS/MS. The subsequent mass reporter ion is measured by the mass spectrometer.

Only MS/MS fragmentation can differentiate the tagged peptides. The reporter ion, measured by the mass spectrometer, generates a different low-mass peak for each sample. As a result, the peak height for each reporter denotes the relative amount of peptide originating from each of the labeled samples.

Proteome Discoverer includes the following default methods for TMT quantification:

- TMT 2plex

- iodo TMT 6plex

- Low-resolution iodo TMTe 6plex

- TMTe 6plex is the standard method for TMTe 6plex using Orbitrap MS detection. The different TMT channels do not interfere so correction factors are not required.

- Low-resolution TMTe 6plex is the method that you must use when you acquire MS/MS data in an ion trap or other instrument with low resolution. In this method, different channel impurities affect the quantitation (mass differences between 15N and 13C are not resolved) so correction factors are suggested to generate the most accurate quantification values.

- TMT 10plex

- TMT 11plex

> **Note** The TMT 6plex quantification method, created for an earlier version of 6-plex labels, is no longer listed. The TMTe 6plex methods replace it. When you require an analysis of the earlier 6-plex method, you must derive a new method with the appropriate reporter ions from the 10plex method, deactivating all the "_N" reporter ion tags in the method.

You can use these default methods to create your own quantification templates. For information on adding quantification methods, see "Setting Up the Quantification Method" on page 543.

Table 27 lists the masses of the reporter ions of the tags available in the different TMT kits. The masses for the original TMT reagents, which are no longer available, are included for reference.

Table 27. Monoisotopic masses of the reporter ions after CID or HCD fragmentation of the tags in the different TMT kits

TMT 2plex		TMTe 6plex (Current)		TMT 10plex		iodo TMT 6plex		TMT 11plex	
Tag	**Mass**	**Tag**	**Mass**	**Tag**	**Mass**	**Tag**	**Mass**	**Tag**	**Mass**
126	126.12773	126	126.12773	126	126.127726	126	126.12773	126	126.127726
127	127.13108	127	127.124760	127_N	127.124761	127	127.124760	127_N	127.124761
				127_C	127.131081			127_C	127.131081
		128	128.13444	128_N	128.128116	128	128.13444	128_N	128.128116
				128_C	128.134436			128_C	128.134436
		129	129.13147	129_N	129.13147	129	129.13147	129_N	129.131471
				129_C	129.13779			129_C	129.13779
		130	130.14115	130_N	130.134825	130	130.14115	130_N	130.134825
				130_C	130.14115			130_C	130.14115
		131	131.13818	131	131.13818	131	131.13818	131_N	131.13818
								131_C	131.14499

The iodo TMT 6plex is a cysteine reactive TMT reagent.

The TMT 10plex and TMT 11plex leverage the high resolution of recent mass spectrometers to routinely differentiate the ^{13}C isotopes from the ^{15}N isotopes[6, 7]. For the 127, 128, 129, 130, and 131 tags, the TMT 11plex contains two reagents: the ^{13}C reagent and the ^{15}N reagent. For the monoisotopic masses of the different reporter ions after CID or HCD fragmentation, see Table 27.

Figure 245 shows the position of the ^{13}C and ^{15}N atoms in the different reagents. In this illustration, the stars indicate the positions of the ^{13}C and the ^{15}N substitutions, the red lines indicate the position of the ETD fragmentation sites, and the blue lines indicate the position of the CID fragmentation sites for the 10plex method.

For more information on the TMT 11plex method, go to the Thermo Fisher Scientific website:
https://www.thermofisher.com/us/en/home/life-science/protein-biology/protein-mass-spectrometry-analysis/protein-quantitation-mass-spectrometry/tandem-mass-tag-systems.html

---

[6] McAlister, G. C.; Huttlin, E. L.; Haas, W.; Ting, L.; Jedrychowski, M. P.; Rogers, J. C.; Kuhn, K.; Pike; I.; Grothe, R. A.; Blethrow, J. D.; Blethrow, G. S. P. Increasing the Multiplexing Capacity of TMTs Using Reporter Ion Isotopologues with Isobaric Masses. *Anal. Chem.* **2012,** *84,* 7469–7478.

[7] Werner, T.; Becher, I.; Sweetman, G.; Doce, C.; Savitski, M. M.; Savitski, B. M. High-Resolution Enabled TMT 8-plexing. *Anal. Chem.* **2012,** *84,* 7188–7194.

Figure 245. Structures of the TMT reagents contained in the TMT 10plex quantification method
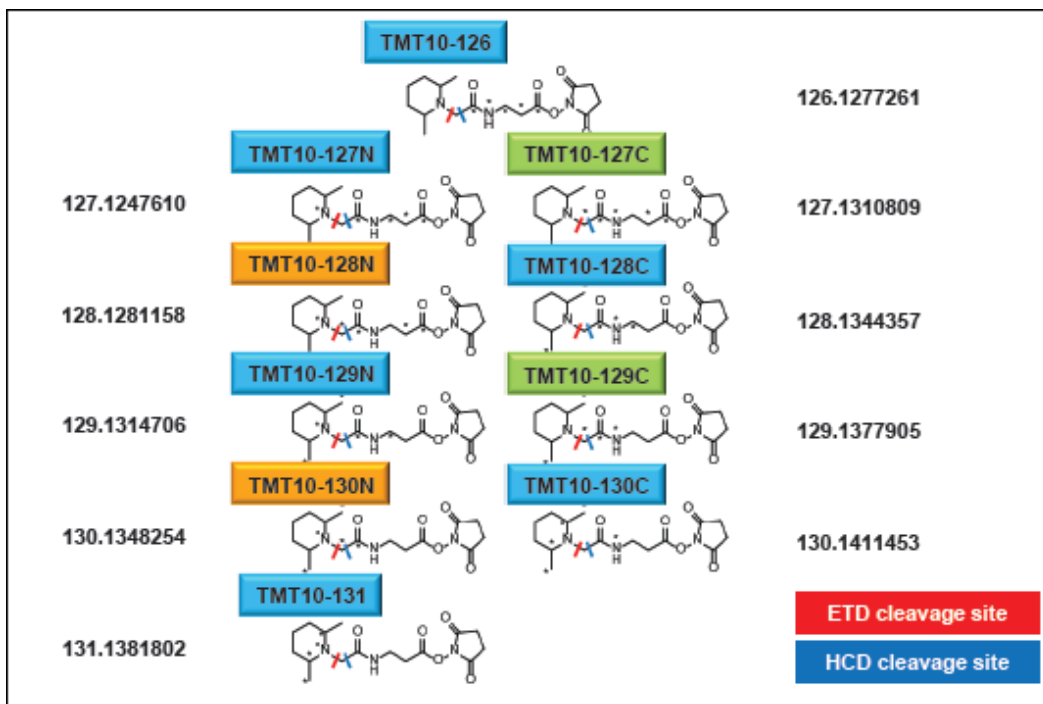
Recent research concludes that, when you avoid applying any correction for isotopic impurities, the quantification results improve for the TMTe 6plex, TMT 10plex, TMT 11plex, and iodo TMT 6plex kits, so the default methods for these kits turn off the purity correction. For the low resolution TMTe 6plex, the TMT 10plex, TMT 11plex, and iodo TMT 6plex kits, the default methods do not include isotope purity correction. However, the addition of correction factors can significantly improve the accuracy for these ratios and they are recommended. Consult the certificate of analysis for the TMT reagent for the study factors. To learn how to enter study factors for a given method, see "Creating a Quantification Method for Reporter Ion Quantification" on page 352.

iTRAQ Quantification

iTRAQ is a protein quantification technique that uses isobaric amine-specific, stable isotope reagents to label all peptides in up to eight different samples simultaneously. The labeled peptides from each sample are combined and analyzed with liquid chromatography tandem mass spectrometry (LC/MS/MS). The same peptide from each sample appears as a single peak in the MS spectrum. In single MS mode, the differentially labeled versions of a peptide are indistinguishable. In tandem MS mode, which isolates and fragments peptides, each tag generates a unique reporter ion. Protein quantification compares the peak intensity of the reporter ions in the MS/MS spectra to assess the relative abundance of the peptides and, therefore, the proteins that they are derived from.

iTRAQ includes two default methods available from AB Sciex that you can use to label all peptides:

- iTRAQ 4plex

- iTRAQ 8plex

The application includes default quantification methods for processing data from iTRAQ 4plex- and iTRAQ 8plex-labeled samples. You can use these methods to create your own workflow templates. For information on adding quantification methods, see "Changing a Quantification Method" on page 554.

iTRAQ quantification works exactly the same as TMT quantification, except that TMT quantification offers 2plex, 6plex, 10plex, and 11plex quantification methods, while iTRAQ offers 4plex and 8plex quantification methods.

Performing Reporter Ion Quantification

This topic describes the default procedure to use to perform reporter ion quantification and then describes the processing and consensus workflows in detail.

- Allocating Sufficient Storage Space

- Downloading FASTA Files

- Checking for the Quantification Method

- Creating a Quantification Method for Reporter Ion Quantification

- Opening an Existing Study

- Creating a Study

- Adding a Description

- Selecting a Quantification Method for the Study

- Adding the Study Factors

- Adding Input Files

- Setting the Factor Values for the Samples

- Saving a Study

- Creating an Analysis

- Opening an Existing Analysis

- Adding Input Files to an Analysis

- Using Multiple Processing Steps in an Analysis

- Adding or Deleting a Processing Step

- Creating a Processing Workflow for Reporter Ion Quantification

- Creating a Consensus Workflow for Reporter Ion Quantification

- Saving an Analysis

- Specifying Quantification Ratios from Selected Sample Groups for Reporter Ion Quantification

- Working with the Search Results

- Specifying Quantification Ratios from Selected Sample Groups for Reporter Ion Quantification

- Performing TMT Quantification on HCD and CID Scans

Allocating Sufficient Storage Space

Be sure that you allocate enough space to store the input raw data files and the temporary result files that the application creates until it creates the final result file. See "Configuring Temporary Files Parameters" on page 32.

Downloading FASTA Files

The first step in performing reporter ion quantification is to download an appropriate FASTA file. For information on downloading a FASTA file, see "Downloading FASTA Files to Proteome Discoverer" on page 153.

Checking for the Quantification Method

Before you open or create a study, check to see if the application includes the quantification method that you want to use.

❖ **To view the quantification methods available**

Choose **Administration > Maintain Quantification Methods**.

Figure 246 on page 353 shows the Quantification Methods view, which lists all the quantification methods.

Creating a Quantification Method for Reporter Ion Quantification

When you do not see an appropriate quantification method, you can create one. A typical way of creating a new quantification method would be to apply the TMT10plex correction factors using the values from the certificate of analysis shipped with the TMT reagents. This step is not necessary when you use the TMTe6plex reagents because there is no isotope interference.

❖ **To create a quantification method for reporter ion quantification**

1. Choose **Administration > Maintain Quantification Methods**, or click the **Maintain Quantification Methods** icon, 📇 .

The Quantification Methods view opens, as shown in Figure 246. It lists all of the available methods for precursor ion, and reporter ion quantification.

Figure 246. Quantification Methods view
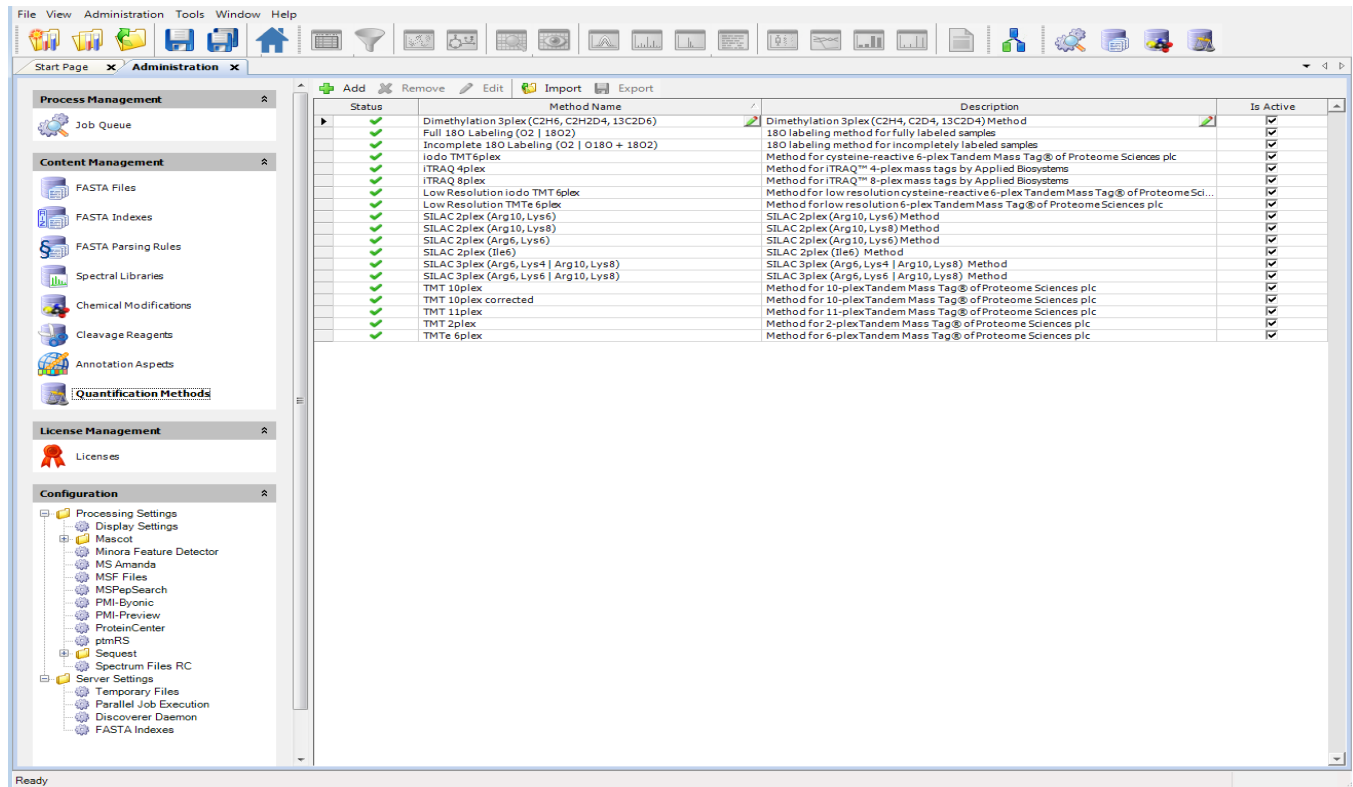
The Status column indicates whether the quantification method is valid for use in quantification:

• A green check mark means that the quantification method is valid and can be used for quantification.

• An exclamation point in a yellow triangle means that the quantification method is not valid. Double-click this mark to view a message that describes the error and provides information on how to fix it.

Figure 247 gives examples of these symbols in the Status column.

Figure 247. Method validity symbols in the Quantification Methods view
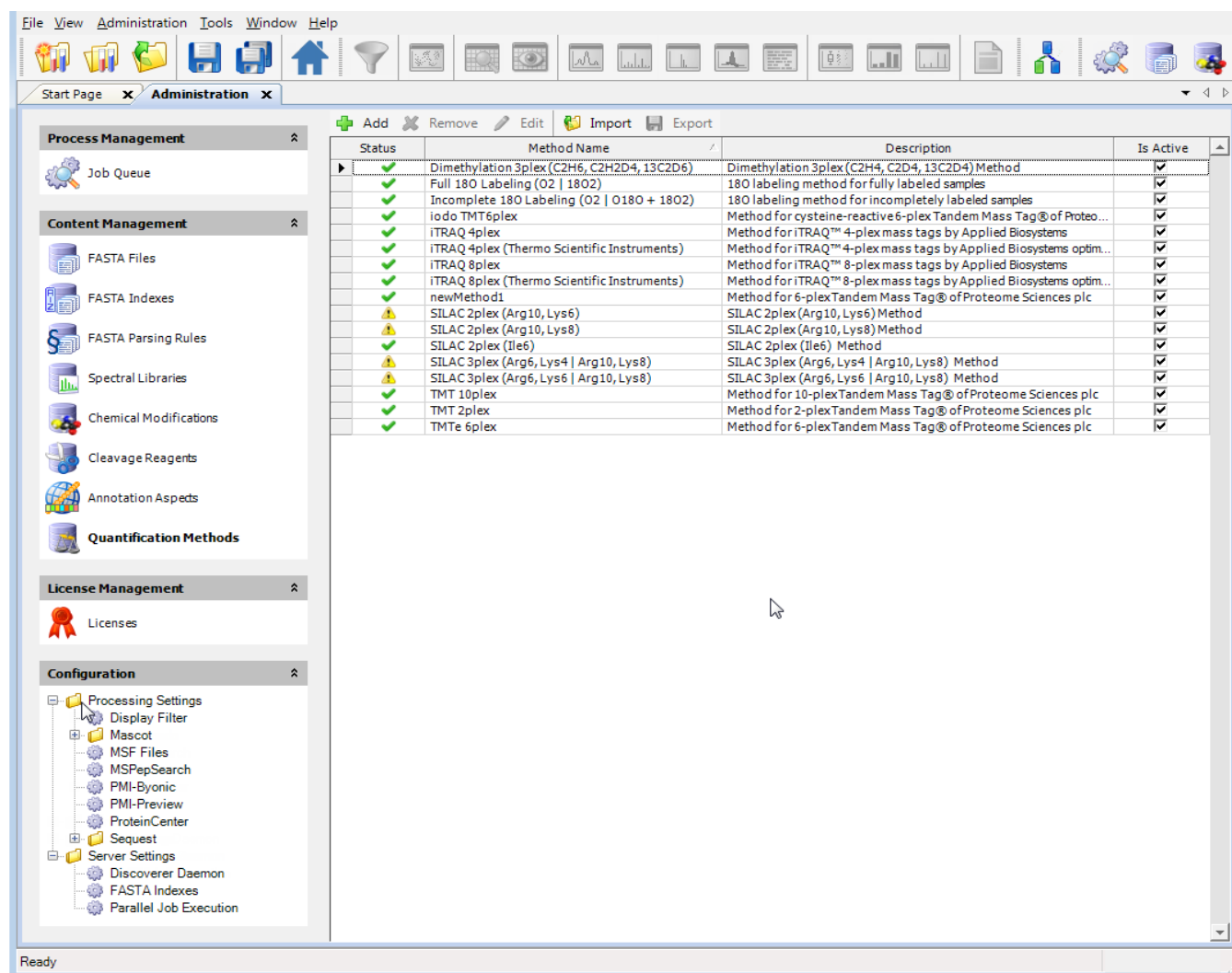
2. Double-click the row for the appropriate method in the Method Name or Description column, or click the column to the left of the Status column for the method and click ➕ Add .
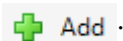
The Create New Quantification Method dialog box opens, as shown in Figure 248.

Figure 248. Create New Quantification Method dialog box
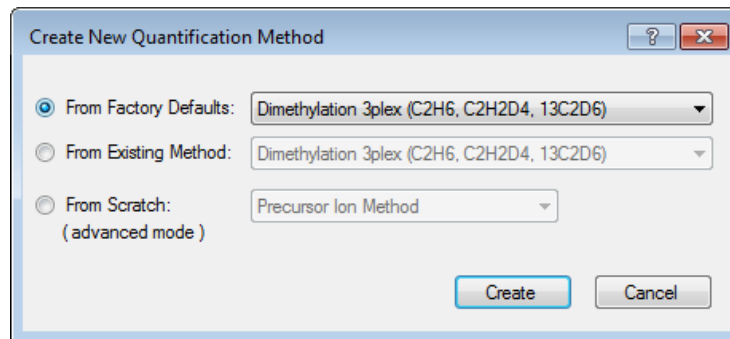
You can select one of the following methods of creating a quantification method:

- (Default) From Factory Defaults: Creates a new method using the same settings from one of the listed default settings that come with a new application installation.

- From Existing Method: Uses the same settings as those of the existing quantification method that you select from the list. The list of methods is the same as that given in the Method Name column of the Quantification Methods view, shown in Figure 246 on page 353.

- From Scratch: Uses one of the following templates so that you can build a new processing method from the beginning:

  – Reporter Ion Quan Method: Provides a template for reporter ion quantification. This method requires an additional license.

  – Precursor Ion Quan Method: Provides a template for precursor ion quantification.

3. Select the **From Existing Method** option, and from the adjacent list, select a method similar to the one that you intend to create, as shown in Figure 249.

   The example shows the **TMT 10plex** method.

4. Click **Create**.

   The Quantification Method Editor opens, as shown in Figure 249.

Figure 249. Quantification Method Editor

5. (Optional) In the Active column of the Quantification Method Editor, clear the check boxes for channels that are not used for the given experiment.

6. (Optional) Enter the isotopic impurities from the certificate of analysis for TMT reagent labels used for experiment.

Figure 250. Quantification Editor with impurities entered

7. Click **OK**.

The Save Quantification Method dialog box opens.

Figure 251. Save Quantification Method dialog box
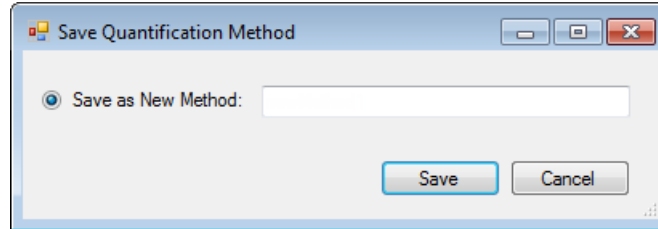
8. In the Save Quantification Method dialog box, type the name of the quantification method that you want to create, for example, **TMT 10plex corrected**.

9. Click **Save**.

The application adds the TMT 10plex corrected method to the Quantification Methods view, as shown in Figure 252.

Figure 252. New custom TMT 10plex corrected method in the Quantification Methods view
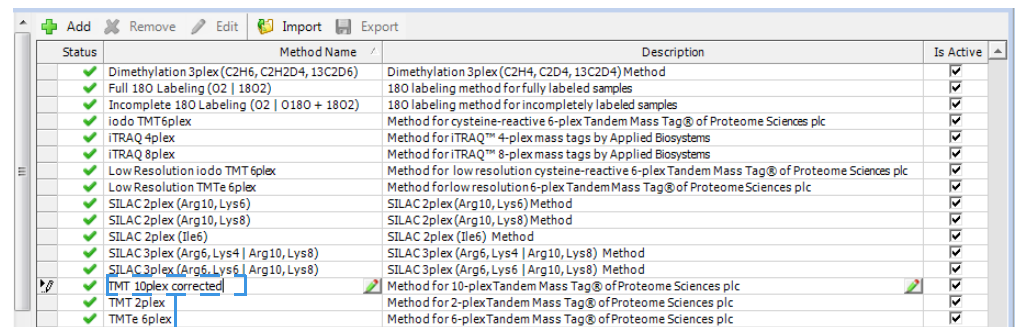
New custom TMT 10plex corrected
quantification method

To create a new quantification method to correct for isotopic impurities in TMT 10plex kits, see "Excluding PSMs with High Levels of Coisolation" on page 528.

Opening an Existing Study

Now you open an existing study or create a study. To create a study, see "Creating a Study" on page 358.

❖ **To open an existing study**

In the Recent Studies area of the Start Page, click the study name.

–or–

In the Recent Studies area of the Start Page, right-click the study name and choose **Open**.

–or–

Choose **File > Open Study** or click the **Open Existing Study** icon, 📊 , and browse to the study folder and then select and open the a .pdStudy file.

–or–

1. In the Start area of the Start Page, click **Open Study**.

2. In the Open Study dialog box, browse to and select the name of a study and then click **Open**.

Creating a Study

> **Note** When you want to use a custom quantification method, create it before you create or open a study. For instructions, see "Selecting a Quantification Method for the Study" on page 360.

❖ **To create a study**

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

–or–

Choose **File > New Study/Analysis**.

–or–

Click the **Create New Study/Analysis** icon, 📊 .

The New Study and Analysis dialog box opens, as shown in Figure 253.

Figure 253. New Study and Analysis dialog box

In this example, you only specify the name of the study and a root directory to save the study in.

2. In the Study Name box, specify the mandatory study name.

   The application generates a default study name by searching for the common part of the file names when you add multiple files at once and using this common part as the suggestion for the name of the new study.

3. In the Study Root Directory box, specify the folder where you will store the study folder. Click the **Browse** button (...), and in the Select Folder dialog box, specify the folder and click **Select Folder**.

4. Click **OK**.

   The application creates a new study folder in the folder that you specified as the root directory and opens a new page with the study name (Study: Gygi_TMT_MS3_Statistics in the example), as shown in Figure 254. It appends a .pdStudy extension to the study file name.

Figure 254. New Study: *Study_name* page

On the Study Definition page, you add a description of your study, select the quantification method or methods to use in the study, and set up the new factors to use to describe and distinguish your samples.

Adding a Description

You can optionally add a description of the study by typing it in the Study Description area of the Study Definition page.

Selecting a Quantification Method for the Study

A quantification method contains the specification of the available quantification channels. You can specify a quantification method for each of the input files. You can either select an existing quantification method, or create a quantification method and then select it.

The example used in this topic uses a custom TMT 8plex method, which you must create.

❖ **To select the quantification method to use in the study**

1. Click the **Study Definition** tab in the study, if it is not already selected.

   The Quantification Methods area of the Study Definition page lists all the quantification methods that are currently available.

2. Select the check box corresponding to the quantification method or methods that you want to use.

   In this example, the samples are labeled with TMT 10plex corrected, so you would select the TMT 10plex corrected check box, as shown in Figure 255.

   When the Quantification Methods pane does not include the TMT 10plex corrected method, choose **File > Save All**, and close and reopen the study.

Figure 255. Quantification method selected
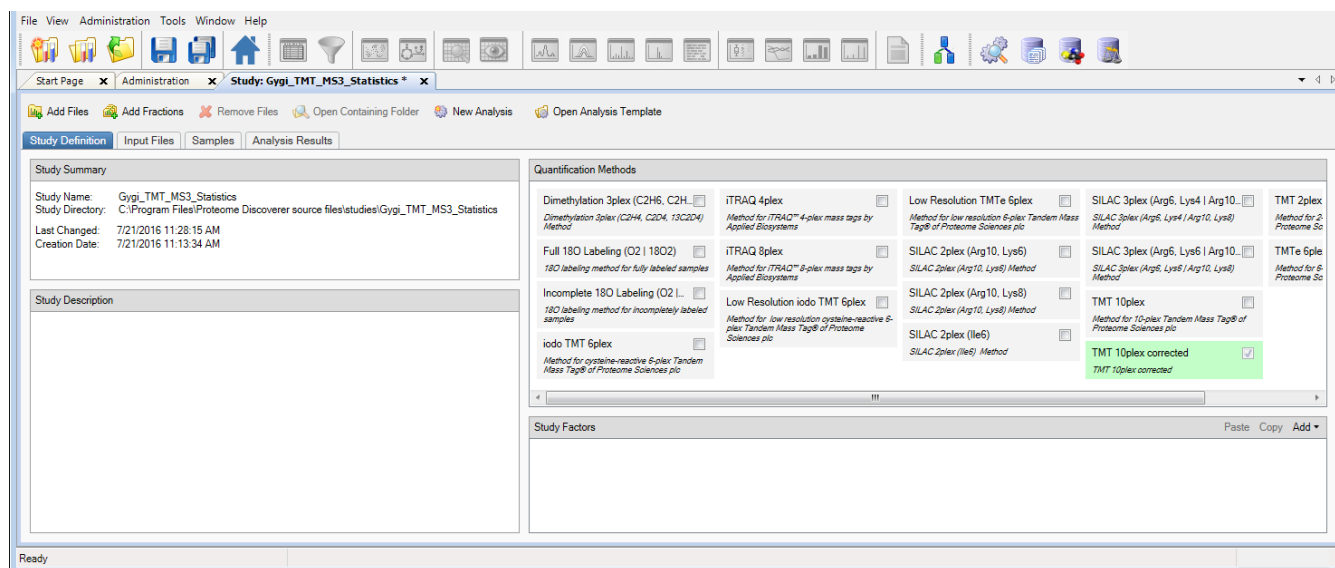
3. After you add the input files, specify the quantification method for each input file (see "Adding Input Files" on page 367). For instructions, see "Specifying the Quantification Method for Multiple Input Files" on page 373.

Adding the Study Factors

In this step, you add the study factors that you want to use for your samples.

A factor is a single biological or technical parameter that you control, for example, genotype, diet, environmental stimulus, age, column length, spray voltage, or collision energy.

In theory, you can track every parameter as a factor in your study, but normally you only track the parameters that actually differentiate samples from each other. For example, the gradient that you use for the chromatography is a parameter in your experiment. However, if you use the same gradient for all your samples, you would not add this as an explicit factor to your study.

❖ **To add numerical study factors**

1. Add the first numerical factor:

   a. In the Study Factors area of the Study Definition page, choose **Add > Numerical Factor**.

   The numerical factor dialog box shown in Figure 256 opens.

   Figure 256. Numerical factor dialog box
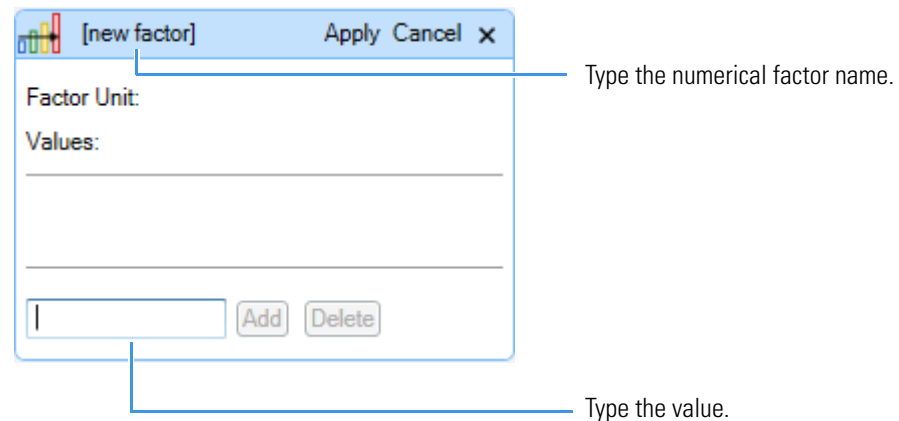
   Type the numerical factor name.

   Type the value.

   [new factor] is highlighted.

   b. Type a name over [new factor] for the new numerical factor, for example, **Time (days)**. See Figure 257.

   > **Note** If the full numerical box becomes compressed, click **Edit** to restore it to its original size.

   c. In the box to the left of the Add and Delete buttons, type the name of the first value, which is 5, and click **Add**. See Figure 257. Continue by adding values 7, 9, 11, 13, 15, 17, 25, 29, and 33.

   Figure 257 shows the completed numerical factor dialog box.

Figure 257. Completed numerical factor dialog box
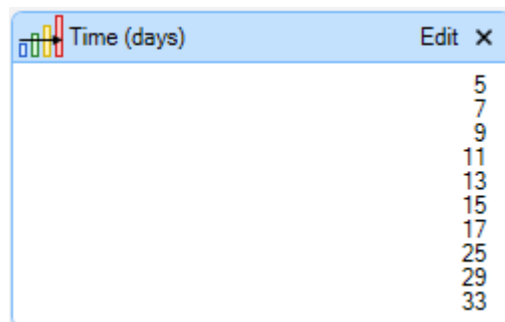
d. Click **Apply** in the numerical factor dialog box.

Confirm that the Study Definition page resembles Figure 258.

Figure 258. Numerical factor added to the study
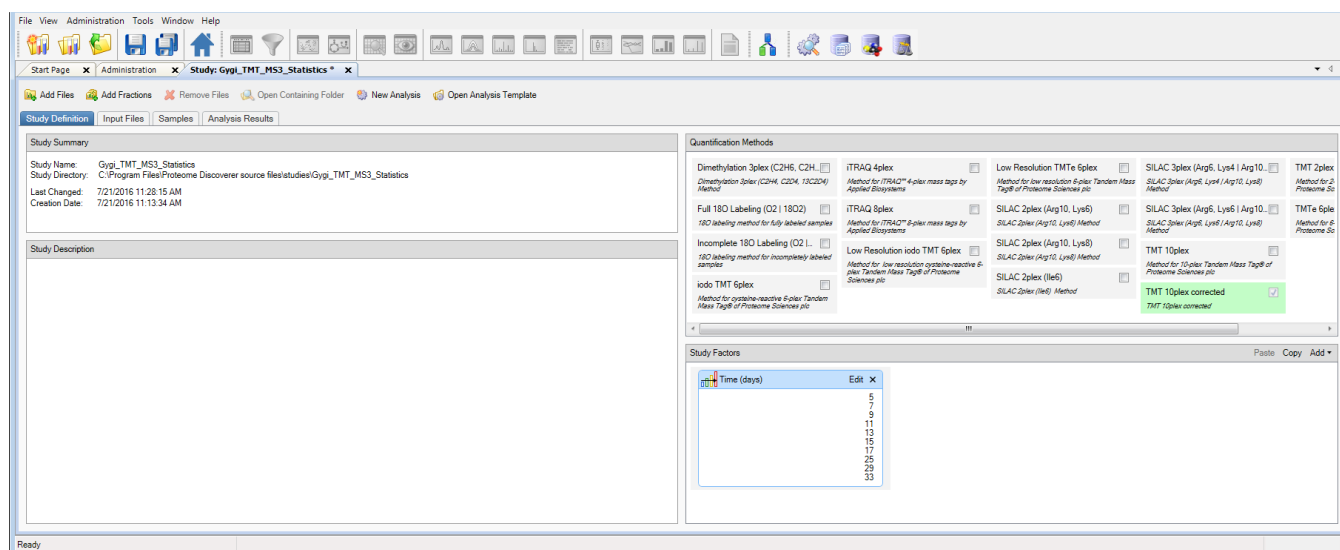
2. Add a second numerical study factor, if applicable.

However, the example in this section does not use a second numerical study factor.

❖ **To add biological replicate study factors**

1. Add the first biological replicate factor:

a. In the Study Factors area of the Study Definition page, choose **Add > Biological Replicate Factor**.

The biological replicate factor dialog box shown in Figure 259 opens.

Figure 259. Biological replicate factor dialog box
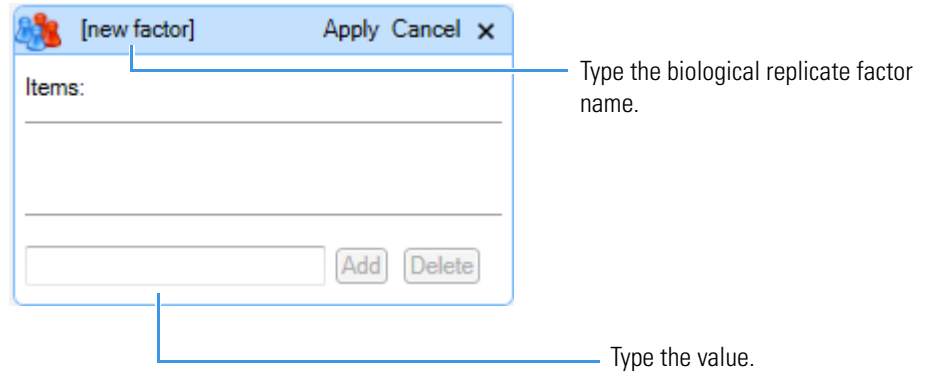
Type the biological replicate factor name.

Type the value.

[new factor] is highlighted.

b. Type a name over [new factor] for the new numerical factor, for example, **Patient**. See Figure 260.

> **Note** If the full numerical box becomes compressed, click **Edit** to restore it to its original size.

c. In the box to the left of the Add and Delete buttons, type the name of the first value, for example, Patient1, and click **Add**. See Figure 260. Continue by adding values Patient2, Patient3, Patien4, and so forth.

Figure 260 shows the completed biological factor dialog box.

Figure 260. Completed biological replicate factor dialog box
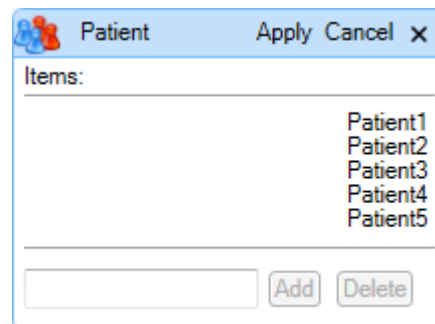
d. Click **Apply** in the numerical factor dialog box.

Confirm that the Study Definition page resembles Figure 261.

Figure 261. Biological replicate factor added to the study
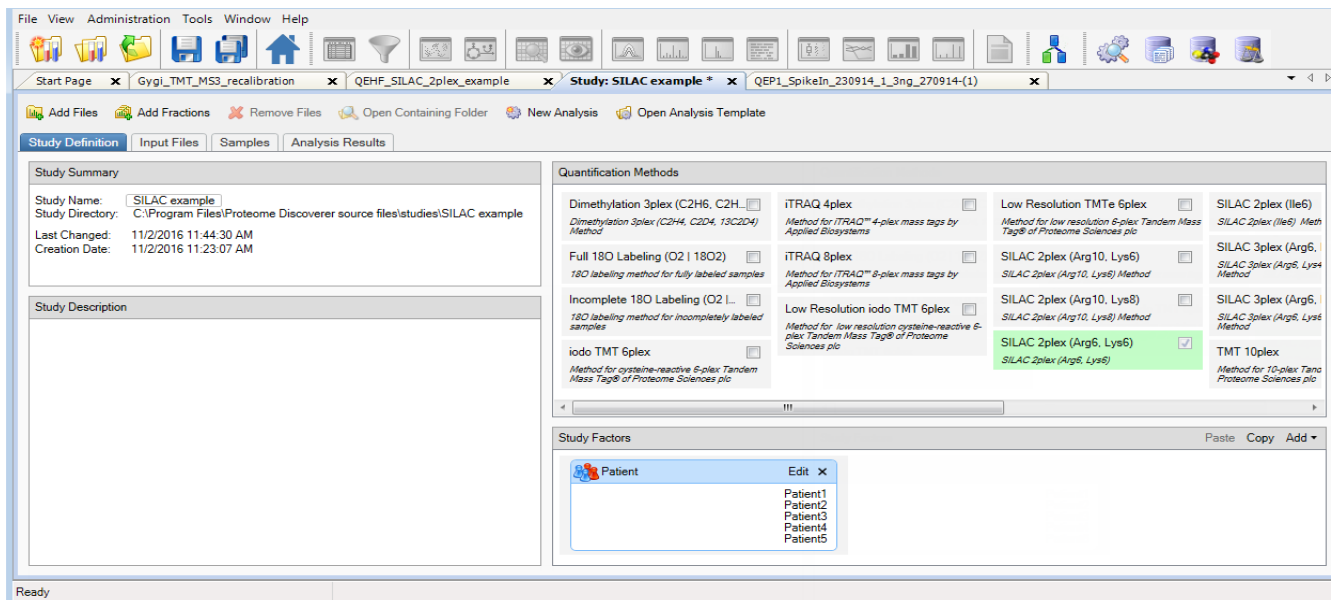
2. Add a second biological replicate study factor, when applicable.

❖ **To add categorical study factors**

1. Add the first categorical (non-numeric) factor (Acquisition in the example):

   a. In the Study Factors area of the Study Definition page, choose **Add > Categorical Factor**.

   The categorical factor dialog box shown in Figure 262 opens.

   Figure 262. Categorical factor dialog box
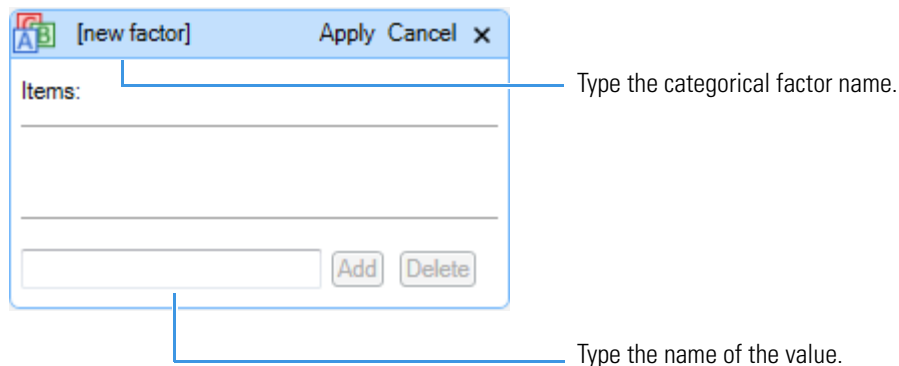
   Type the categorical factor name.

   Type the name of the value.

   [new factor] is highlighted.

   b. Type a name over [new factor] for the new categorical factor, for example, **Acquisition**. See Figure 263.

> **Note** If the full categorical box becomes compressed, click **Edit** to restore it to its original size.

c. In the box to the left of the Add and Delete buttons, type the name of the first value and click **Add**. See Figure 263.

In this example, the value is the acquisition method, and the first acquisition method is intelligent data acquisition (IDA).

d. In the same box, type the name of the second acquisition method and click **Add**.

In this example, the second acquisition method is data-dependent acquisition (DDA). See Figure 263.

Figure 263 shows the completed categorical factor dialog box.

Figure 263. Completed categorical factor dialog box
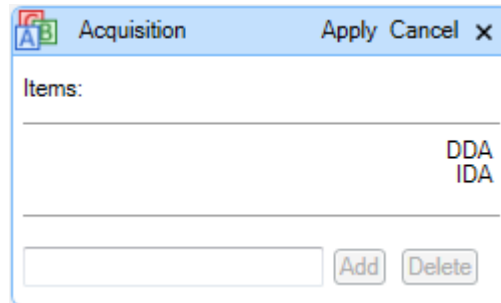
e. Click **Apply** in the categorical factor dialog box.

Confirm that the Study Definition page resembles Figure 264.

Figure 264. First categorical factor added to the study
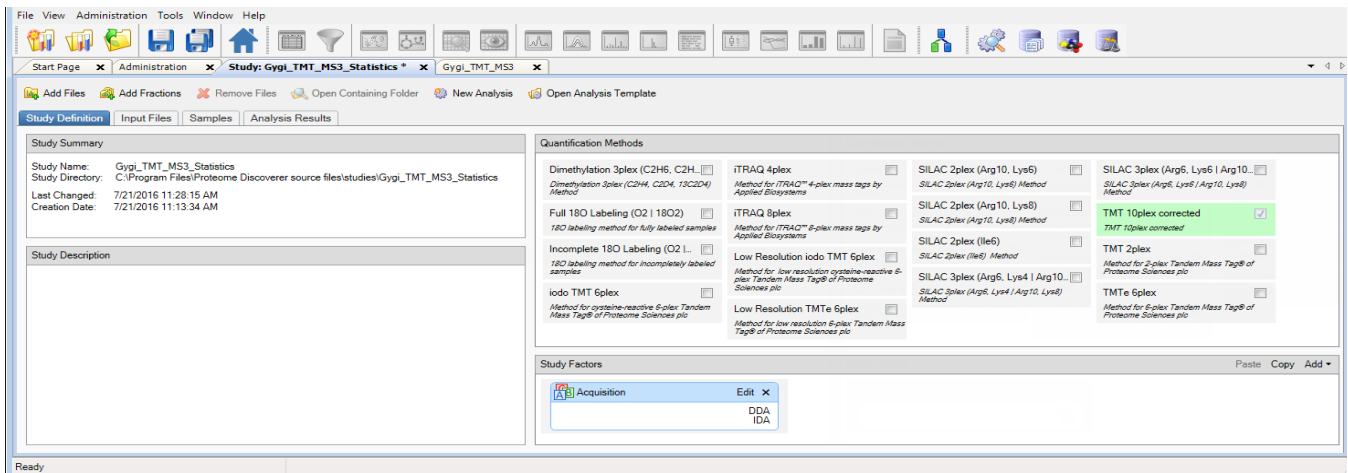
2. Add the second categorical factor (Tissue in the example):

a. In the Study Factors area of the Study Definition page, choose **Add > Categorical Factor**.

The dialog box shown in Figure 262 on page 364 opens.

[new factor] is highlighted.

b. Type a name over [new factor] for the new factor (see Figure 263 on page 365), for example, Tissue.

c. In the box to the left of the Add and Delete buttons (see Figure 263 on page 365), type the name of the first type of tissue and click **Add**.

In this example, the first type of tissue is Cerebellum.

d. In the same box, type the name of any additional types of tissue and click **Add** after each one.

The example adds the following types of tissue to the study:

- Cerebellum

- Cerebrum

- Heart

- Kidney

- Liver

- Lung

- Muscle

- Spleen

e. Click **Apply** in the Tissue dialog box.

Confirm that the Study Definition page resembles Figure 265.

Figure 265. Second categorical factor added to the study

Adding Input Files

Add the input files from the example data set to your study. (You can add input files to the study at any point.) For the types of input files supported, see "Inputs" on page 10. You can add individual input files, multiple unrelated input files, or fractions.

Adding a Single Input File or Multiple Unrelated Input Files

❖ **To add a single input file or multiple unrelated input files**

1. Click Add Files .

2. In the Add Files dialog box, browse to the location of the input files, and select them.

   > **Note** You can also add existing MSF files to a study. The application adds the MSF file as a new result to the Analysis Results page and adds the raw data files that were processed to create the MSF files to the Input Files page.

3. Click **Open**.

–or–

Drag the input file or files from Windows Explorer and drop them onto the Input Files page.

The input files appear on the Input Files page, as shown in Figure 266. Each file on the page receives a unique identifier: F1, F2, …, F*n*. The Proteome Discoverer adds each file as a single study file.

Figure 266. Input files on the Input Files page
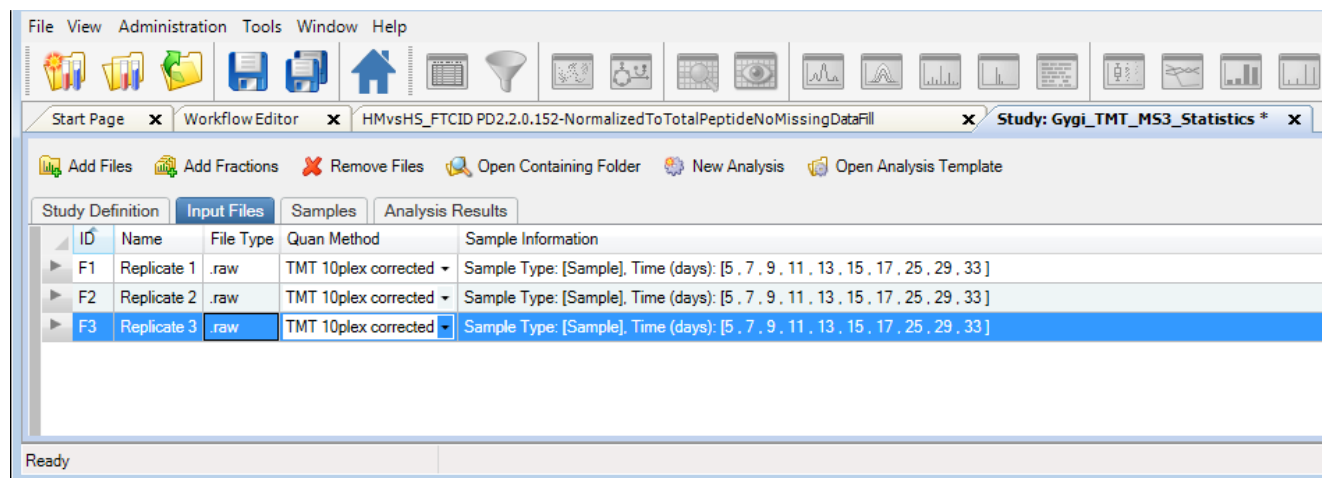
Adding Fractions

Fractions are datasets where a sample was prefractionated prior to LC/MS/MS analysis. In this case, the results associated with the sample are found in multiple raw data files, each of which is a fraction of the original sample.

MudPIT is an example of a popular method that produces fractions. MudPIT experiments investigate complex proteomes by applying multidimensional chromatography to the samples before acquisition in the mass spectrometer. Typically, this process results in several dozen or even a few hundred fractions that are separately analyzed by LC/MS, resulting in one raw data file per sample fraction. Analyzing gel slices or performing in-depth follow-up acquisitions also results in multiple fractions. Because all these fractions belong to the same sample, the Proteome Discoverer can process all raw data files from these fractions as one contiguous input file, generating a single result file.

You can search MudPIT fractions in the Proteome Discoverer by using ![Add Fractions] or in the Proteome Discoverer Daemon utility by using the MudPIT button. However, Thermo Fisher Scientific recommends that you use the Proteome Discoverer to process local batch or MudPIT samples, and use the Proteome Discoverer Daemon utility only when you need remote access and automatic processing from the command line—for example, when you use the Xcalibur data system to send data to a remote Proteome Discoverer computer (host) for acquisition. For information on processing MudPIT and batch samples with the Proteome Discoverer Daemon utility, see "Using the Proteome Discoverer Daemon Utility" on page 123.

❖ **To add fractions**

1. Click ![Add Fractions] .

2. In the Add Fractions dialog box, browse to the location of the input files, and select them.

3. Click **Open**.

The Proteome Discoverer adds all selected files as a single study file.
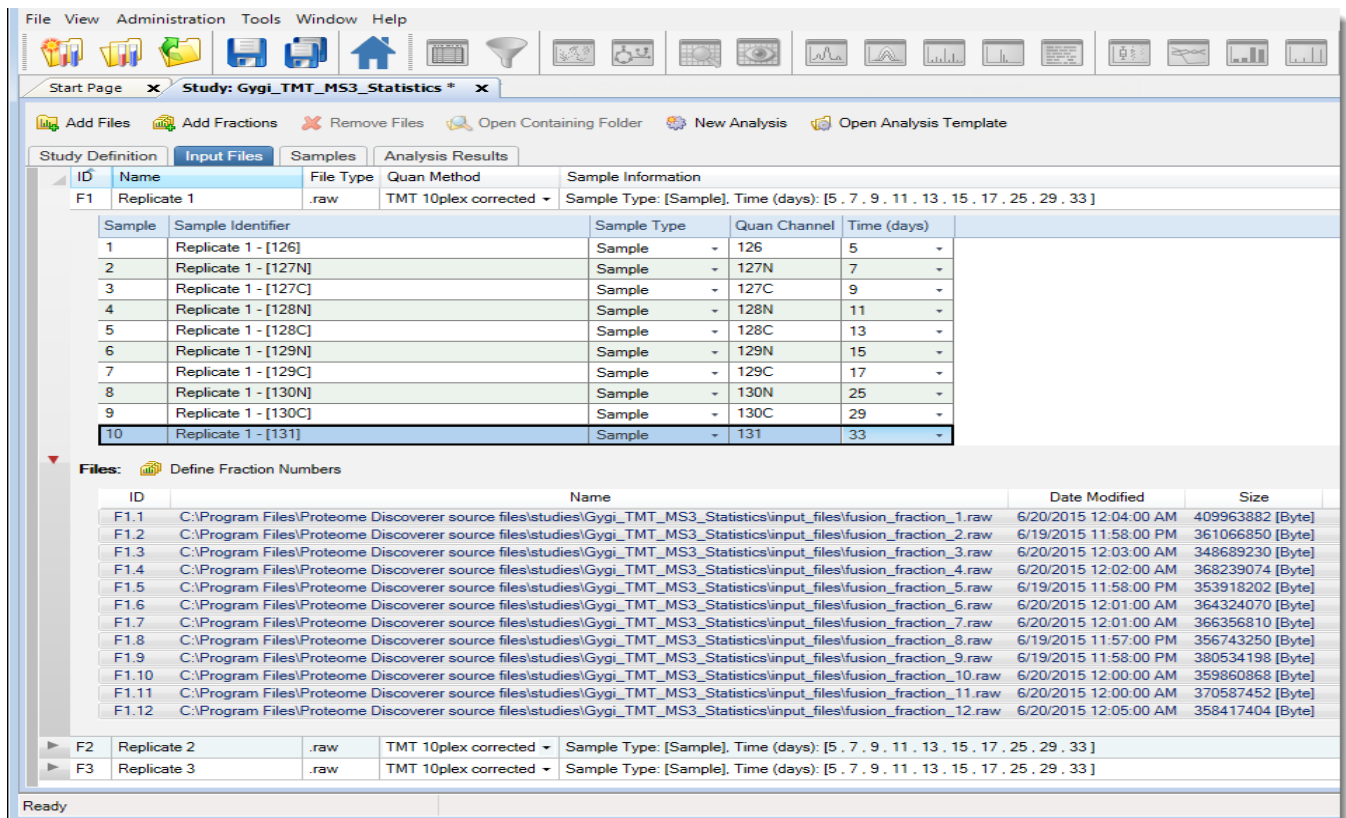
Creating Subsets of Fractions

You can optionally use only parts of a file set in a fractionated sample by re-adding them as a subset on the Input Files page.

❖ **To create subsets of fractions**

1. Add the original input files as follows:

   a. On the Input Files page, click ⊞ Add Fractions .

   b. In the Add Fractions dialog box, browse to the directory where the fraction files are located, select the fraction files, and click **Open**.

   The fraction files appear as file sets called F2, F3, and F4 on the Input Files page, as shown in Figure 267.

Figure 267. Fraction files added to the Input Files page as file sets F2, F3, and F4

2. Re-add some of the fraction files as follows:

   a. Click ⊞ Add Fractions .

b. In the Add Fractions dialog box, browse to the same directory where the original fraction files are located, select the fraction files that you want to re-add, and click **Open**.

The Proteome Discoverer adds the selected fraction files to the original file set as a subset. Figure 268 shows three of the original fraction files added to the file set as a subset of F3 called F5.

Figure 268. Three fraction files re-added as a subset of F3

You can create any number of subsets.

When adding fractions to a study, the application first checks to see whether the study already contains any new files whose location changed within the study. When it finds these kinds of files, it updates the file locations but does not create new sample files. For example, suppose that the study contains a file called *drive*:\xxx\file1. You add file1 through file5. The application updates the path of the file1 file but does not create a new sample file.

When all selected files are new—that is, if the study does not already contain the files—the application creates a new sample file set that contains all the selected files. For example, suppose that you want to add file1 through file5, and all these files are new. The application creates a new sample file containing files file1 through file5.

When the study contains at least one new file, the application checks to see how many existing study files contain new files that are not subsets. For example, suppose that the existing sample file consists of all new files. You add file1 through file5. The study contains a sample file set called F1 consisting of file1 through file5 and no other files. In this case, the application does nothing.

As another example, suppose that an existing sample file set called F1 contains all new files but also some others, for instance, file1 through file8. In this case, the application creates a new subset containing file1 through file5 and names it [Subset of F1] F2… .

As a another example, suppose that an existing file set called F1 contains only new files, such as file1 through file3. The application creates a new file set, F2, which consists of file1 through file5 and is actually a superset of F1. But it adds F2 as a subset to the original F1 file set. The F2 subset added to F1 is called [Subset of F2] F1… .

As a final example, suppose that an existing file set called F1 contains only new files, such as file1 through file3. You want the application to create a new file set, F2, which consists of file1 through file3 and file6, as a superset of F1. But because file4 and file5 are missing, the application issues an error message and takes no action.

The application also does not support multiple file sets containing new files. When it encounters this situation, it issues an error message and takes no action.

Assigning the Order of Fractions

You can specify the order of the fraction numbers for the files in an input file set. Because files from different fractions can be very different, this step is required for label-free quantification. Keeping the same order of fractions in each file set ensures that retention time alignment and feature mapping are performed between corresponding fractions.

The Input Files page of a study displays the fraction numbers. It consists of a table that lists the identification number (ID) of a file, the path and name of the file, modification date, and size of each file in the file set. The identification number in the ID column follows this format:

F*fileID*.*fraction*

where:

- *FileID* is the file number. For example, in the file identification number F21.3, 21 is the file number.

- *Fraction* is the fraction number. For example, in the file identification number F21.3, 3 is the fraction number.

❖ **To assign fraction numbers**

1. Add the file set to the study. See "To add fractions" on page 368.

2. Click the **Input Files** page of the study if that page is not already visible.

3. Click the **Define Fraction Numbers** icon on the Input Files page, shown in Figure 269.

This icon is available only when the input file set contains multiple files and when the file set is not a subset of another file set.

Figure 269. Input Files page displaying fractions
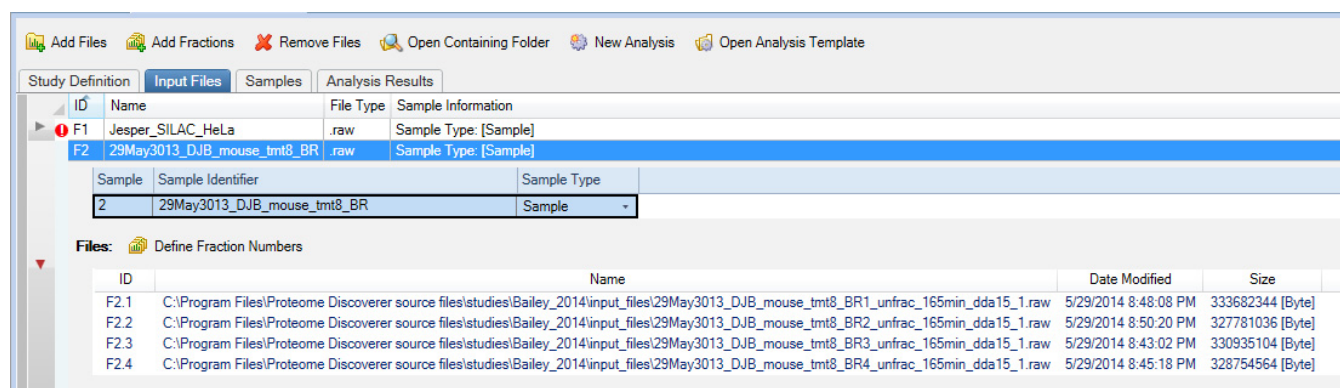
The Specify File Fractions dialog box opens, as shown in Figure 270, displaying all the files in a file set. The files in this example are not in numerical order.

Figure 270. Specify File Fractions dialog box displaying files not in numerical order
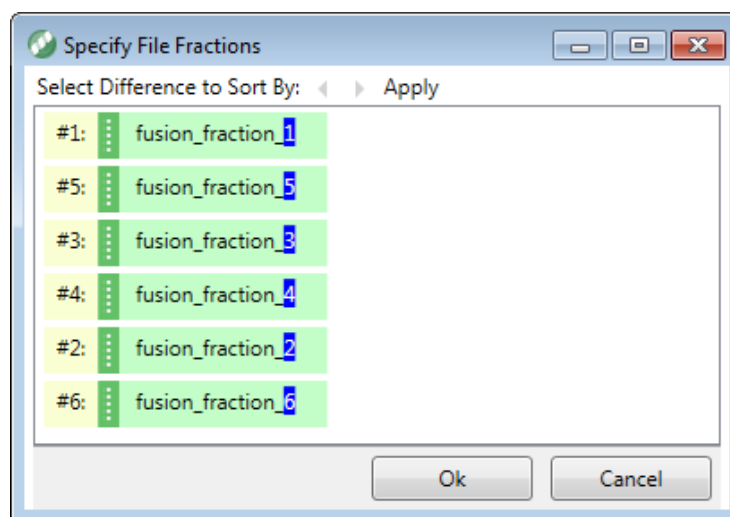
4. Place the files in order by doing one of the following:

Click **Apply** to sort the fractions automatically by the selected file part.

The application sorts file names numerically when all the selected file parts start with a numeric value; otherwise, it sorts them alphabetically.

–or–

Use the arrow keys to the left of Apply, ◄ ►, to highlight the part of the file name to sort by.

When there are multiple different file parts, you can choose a file part by clicking the arrows; the selected file part is highlighted in blue, while other parts are highlighted in gray.

–or–

Manually change the fraction numbers by dragging and dropping the file names. Use the variable placement handle ( ) to drag each file.

Figure 271. Specify File Fractions dialog box displaying files in numerical order
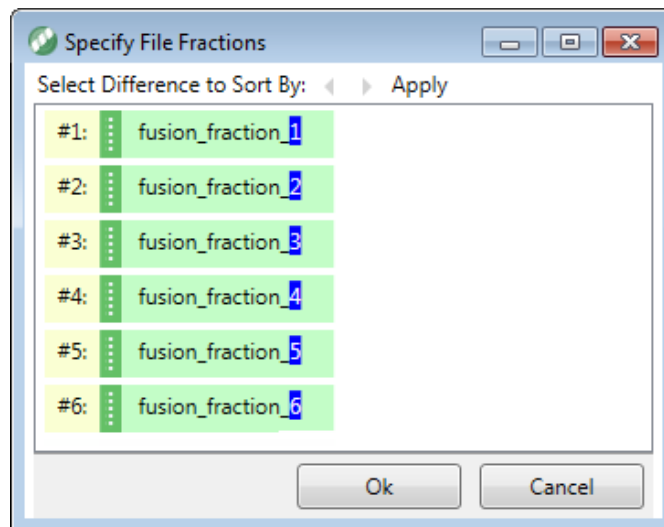
5. Click **OK**.

   Clicking the OK button closes the dialog box and stores the new fraction numbers.

   Clicking the Cancel button restores the previous fraction numbers.

Specifying the Quantification Method for Multiple Input Files

In this step, you specify the quantification method that was used for each of the files. In the example data set, all samples were labeled with the TMT 10plex corrected custom quantification method.

You do not need to assign the quantification method to the input files for label-free quantification.

❖ **To set the quantification method for each of the input files**

1. Click the **Input Files** tab, if it is not already selected.

2. In each sample row, click the Quan Method column and select the quantification method (for this example, **TMT 10plex corrected**) from the list. See Figure 272. Only items in list are those that are selected in the Study Definition page.

Figure 272. Setting the quantification method for each raw data file
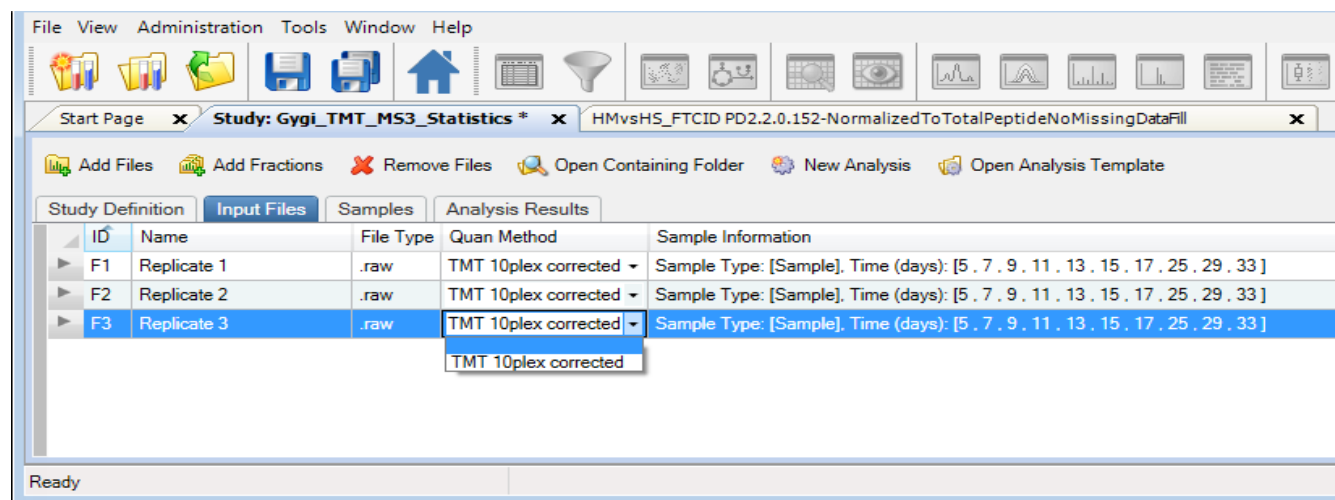
Setting the Factor Values for the Samples

When you select a quantification method for a file, the Proteome Discoverer generates a sample placeholder for each quantification channel.

The term "sample" refers to a distinct measurement of an analyte. At a minimum, a raw data file contains at least a single sample, or in the case of sample multiplexing, it uses isobaric or metabolic labeling for multiple samples. When you measure the analyte from the same vial in two acquisitions—that is, if you measure two technical replicates—the application considers them two different samples. When you split the analyte, label each with a different isobaric tag, mix them, and acquire them into a single raw data file, the application considers the two differently labeled parts as two different samples.

Each sample is associated with a sample type. Currently only quantification uses sample types. The sample type can be one of the following:

- Sample: A specimen from a larger biological entity

- Control: A sample typically used as a reference sample in a quantification experiment

- Blank: A sample consisting only of solvent and no sample mixture

- Standard: A sample consisting of a standard quality-control peptide mixture

You can create ad hoc relative quantitative ratios between any study factors or sample types, for example, blank/standard, control/sample, blank/control, and standard/blank. You can leave all of the channels as samples or make all of the channels any combination of sample, control, blank, or standard.

The Sample Type column on the Samples page of the study displays the type of each sample. The default type is Sample.

You can also use the control channels in scaling. For more information, see "Normalizing Peptide Groups and Protein Abundances" on page 535.

Each sample is associated with a quantification channel shown in the Quan Channel column and with values for each of the factors that you specified for your study. Previously, you specified a factor for time. You now set the correct factor values for each of the samples in the study.

Each sample has an automatically generated sample name composed of the raw data file name and the appended name of the quantification channel. You can change this name, but the name must be unique among all samples in the study.

❖ **To view the samples**

On the Input Files page, click the gray arrow to the left of a sample to display its constituent file entries, as shown in Figure 273.

A hierarchical view opens, showing the samples contained in a raw data file. For each of the raw data files in the example, there are eight samples for the eight quantification channels of the TMT 10plex corrected method. See Figure 273.

Figure 273. Samples in a raw data file

Click the red arrow to close a file entry.

Click the gray arrow to expand a single file entry.
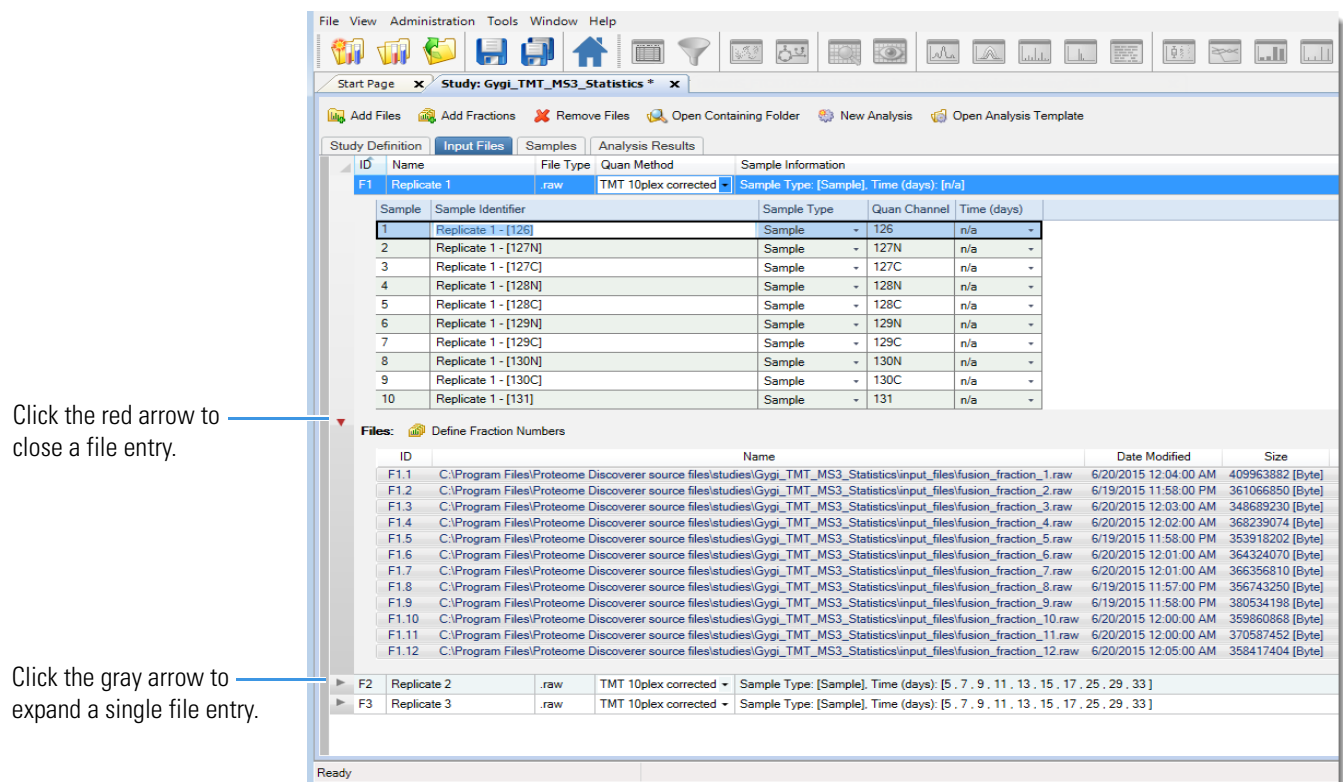
For reporter ion quantification, you can set the factor values for the samples on the Input Files page or the Samples page. The following procedure shows you how to set them on the Input Files page.

❖ **To set the factor values for the samples**

1. Click the **Input Files** tab, if it is not already selected.

2. Click the gray arrow next to the first file to expand the information about the sample.

3. For the factor (in the example, Time), set the value for each sample in each raw data file by selecting the down arrow in the factor column and then selecting the value from the list.

   Figure 274 shows this process for the Time factor.

   Figure 274. Setting the factor values for the Time factor

4. Set the same values in the Time column for the rest of the samples.

   After you finish setting the factor values for each sample, the Input Files page resembles Figure 275.

> **Note** For instructions on changing the values for factors or other study variables for multiple samples at once, see "Setting Values for Multiple Samples at the Same Time on the Input Files Page" on page 379.

Figure 275. Completed sample factors

5. (Optional) Click the **Samples** tab.

Figure 276 shows the Samples page. It displays the same sample information as the Input files page.

Figure 276. Samples page

Setting Values for Multiple Samples at the Same Time on the Input Files Page

Highly multiplexed data are results obtained from processing several samples from one raw data file mixed and analyzed together in one LC/MS analysis, where isotopic and isobaric labels were used in quantification to distinguish the contribution of the individual samples. When you have highly multiplexed data, several files and samples with many study variables to set (sample type, quantification channel, study factors), or both, you can set values for study variables for multiple samples at once. You can use either the mouse or the keyboard to set these values on the Input Files page.

You can also use multi-cell editing on the Samples page to enter factor values for several samples. For example, when you have acquired data from several different tissues and the tissue is encoded in the file and sample names, you can use the filter option on the Samples page to filter all samples that contain "liver" and then set the Tissue factor to "liver" for all these samples.

❖ **To set values for multiple sample cells at the same time by using the mouse**
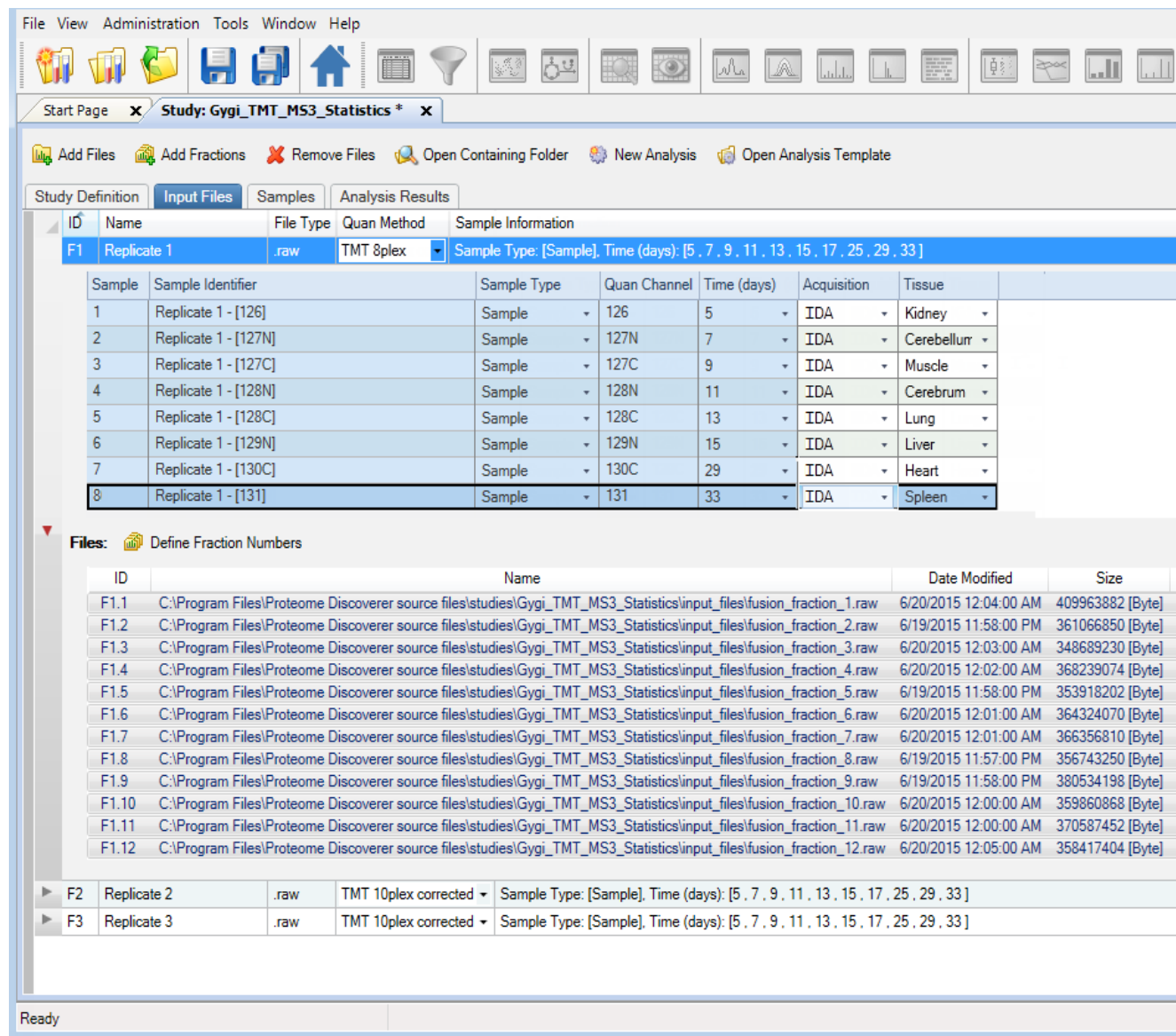
1. Click the **Input Files** tab.

2. To select multiple samples for one study variable, select the first cell and drag the cursor to select the remaining cells.

   Figure 277 shows multiple cells being selected in the Time column.

Figure 277. Selecting multiple cells at one time

3. Right-click and choose the column that you want to apply the changes to and choose the value from the submenu.

    For example, suppose you wanted to set all samples in the Time (days) column to 11. After you select all cells, right-click and choose **Set (Time (days) to > 11**, as shown in Figure 278.

Figure 278. Selecting a new value to apply to all cells in multi-cell editing mode
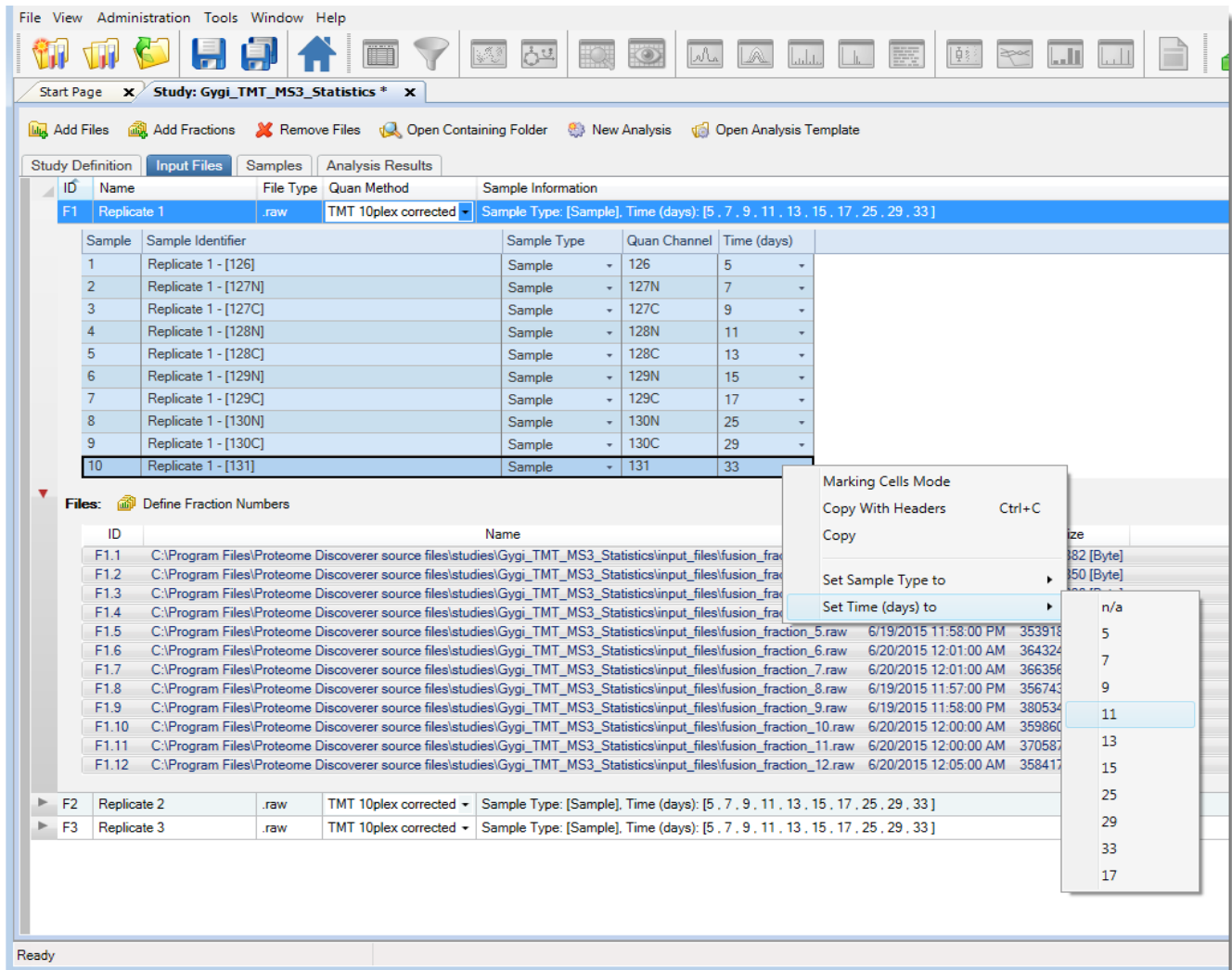
Setting Filters for Multiple Samples at the Same Time on the Samples Page

Use the following procedure to change filters for multiple samples on the Samples page.

❖ **To set filters for multiple samples at the same time on the Samples page**

1. Click the **Samples** tab.

2. Click the **Contains** icon, ▣ , when necessary, in the appropriate column of the Samples page.

3. In the box next to the icon, type the name of the filter.

   For example, suppose you want to change the sample identifier in the Sample Identifier column to 126. Figure 279 shows the results.

Figure 279. Multicell editing of samples using filters
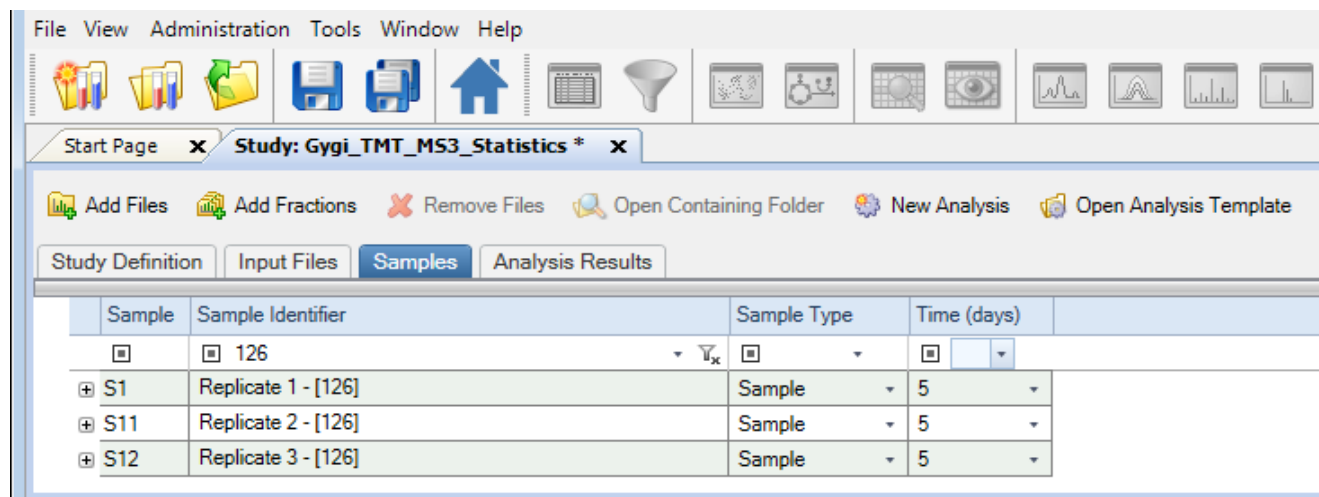
4. SHIFT-click or CTRL-click to highlight the rows to be modified.

5. Right-click the highlighted cells and select the study factor to apply.

To return the samples to their unfiltered state, select the **Clear All Filters** icon, [icon].

Saving a Study

You can save a study manually or automatically.

You can save a study manually at any time. A change in a study that requires you to save it is indicated with an asterisk (*) in the tab after the study name.

> **Note** Studies and analyses in the Proteome Discoverer are separate and must be saved separately. Saving a study does not save an analysis, and saving an analysis does not save a study.

You can set an option to have the application save studies automatically when you click [Run] . This option also saves the results generated in the study. It saves the analysis containing the workflow, but when you close the study and reopen it, you must access the analysis by clicking the Analysis Results tab, and then doing one of the following:

❖ **To save a study manually**

Choose **File > Save**.

The Proteome Discoverer saves the study in the *study_name*.pdStudy file in the study directory.

❖ **To save a study automatically**

1. Before you click , select **Tools > Options**.

2. In the Options dialog box, select **Study Options** in the left pane.

3. Select the **Auto Save When Starting Analysis** check box.

4. Click **OK**.

   The Auto Save When Starting Analysis option does not take effect for any open studies. The application automatically saves only newly opened studies.
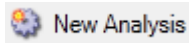
❖ **To save all open studies**

   Choose **File > Save All**.

   The Proteome Discoverer saves all studies open in the application.

Creating an Analysis

❖ **To create an analysis**

   On the Study: *Study_name* page, click .

   An Analysis window opens on the right side of the Study: *Study_name* page, as shown in Figure 280.

   Figure 280. Analysis window
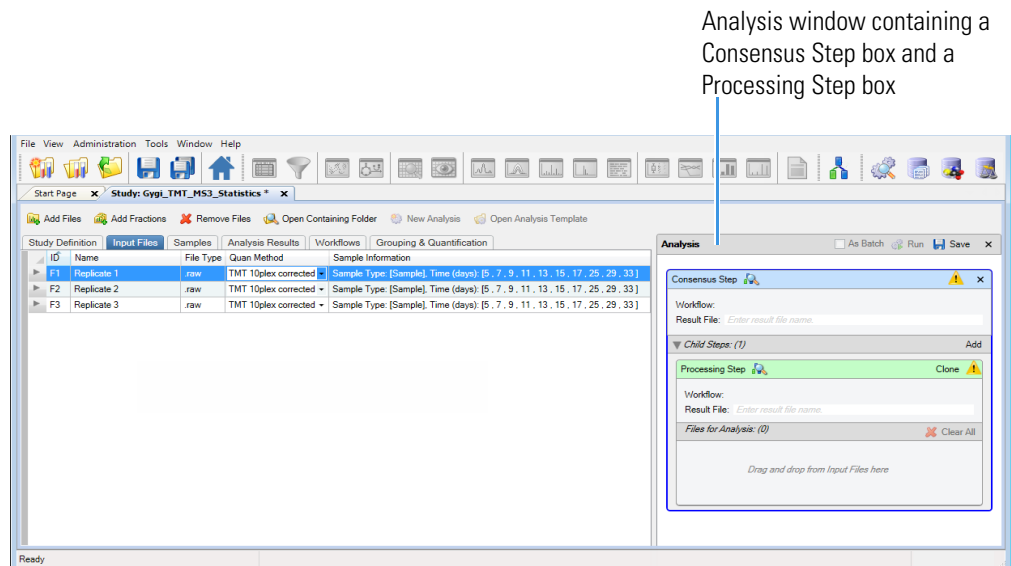
Analysis window containing a Consensus Step box and a Processing Step box

An Analysis window contains the following items:

- A Consensus Step box, which represents the consensus workflow step of the data processing

- A Processing Step box, which represents the processing workflow step of the data processing

- A Child Steps bar, which contains an Add button that you can use to add another Processing Step box. Multiple Processing Step boxes are useful when you want to process the same data in different ways—for example, by using different nodes or different node settings.

In addition, two new tabs appear on the Study: *Study_name* page: the Workflows tab and the Grouping & Quantification tab, as shown in Figure 280.

If you switch to an existing study that includes an Analysis window containing an existing consensus workflow, processing workflow, or both, and you want to open a new analysis, you must close the existing Analysis window and open a new Analysis window. To close the window, click **X** in the upper right corner.

Opening an Existing Analysis

You can open an existing analysis through a template or through a results file.

❖ **To open an existing analysis through a template**

1. In an open study, click 🔘 Open Analysis Template .

2. In the Open Analysis Template dialog box, browse to and select the .pdAnalysis file that you want to open.

3. Click **Open**.

❖ **To open an existing analysis through a results file**

1. In an open study, click the **Analysis Results** tab in the study.

2. Select the appropriate .pdResult file.

3. Click 🔄 Reprocess ▾ .

4. To open an Analysis window containing both the processing and consensus workflows, choose **All Analysis Steps** in the Reprocess list.

   –or–

   To open an Analysis window in order to execute just the consensus workflow, choose **Last Consensus Step** in the Reprocess list.

Adding Input Files to an Analysis

❖ **To add the input files to an analysis**

Select and drag the files from the Input Files page to the Input Files box of the Processing Step box in the analysis window.

The input files are listed in the Input Files area of the Processing Step box, as shown in Figure 281.

Figure 281. Input files in the Input Files area of the Processing Step box of the Analysis window

> **Note** You can remove a file from the Input Files area of the Processing Step box by clicking the X to the left of the file name.

Using Multiple Processing Steps in an Analysis

An analysis can contain more than one processing step. You can use additional processing steps when you want to process one set of input files differently from another set of input files. For example, you might want to run a Percolator validation separately on a set of files, such as the replicates of the control group and the replicates of the treatment group.

❖ **To add multiple processing steps**

In an open analysis window, click **Add** on the Child Steps bar.

A new processing step without a workflow appears.

❖ **To duplicate an existing processing step, including its workflow**

In an open analysis window, click **Clone** on the processing step bar.

Another window with a child step appears. You can now define a different processing workflow for the same file or a different file.

Adding or Deleting a Processing Step

You might want to perform only the processing step in a workflow. For example, you only need the processing step when you want to export peak lists as an MGF file with the Spectrum Exporter node. You can remove the Consensus step and re-add it later.

❖ **To perform only the processing step in an analysis**

In the title bar of the Consensus Step box, click the **X** in the upper right corner to remove the consensus step from the analysis.

The Consensus Step box closes, and the Add icon appears in the Analysis title bar.

❖ **To re-add a deleted consensus step in an analysis**

In the Analysis title bar, click ➕ Add .

The Consensus Step box reopens, and the Add icon disappears.

Creating a Processing Workflow for Reporter Ion Quantification

To create a reporter ion quantification method, you must set up a processing workflow that includes the Reporter Ions Quantification node.

Default processing workflows are available for several instruments. You can use these default workflows as is or modify them to suit your needs. To access them, click [Open Common] on the Workflows page, or see "Using Common Workflow Templates" on page 102. When you use a default workflow, you do not need to perform step 1 through step 7 in the following procedure because these steps are already included in the workflow. However, you must check the node parameters in the workflow to make sure they are appropriate for your data and add modifications to the search for the quantification to be successful.

The processing workflow must be selected in the Workflow Tree window for the Open Common command to display the processing workflows available.

The default processing workflows for reporter ion quantification are the following:

- ProcessingWF_Fusion\PWF_Fusion_Reporter_Based_Quan_SPS_SequestHT_ Percolator.pdProcessingWF

- ProcessingWF_Fusion\PWF_Fusion_Reporter_Based_Quan_MS2_SequestHT_ Percolator.pdProcessingWF

- ProcessingWF_QExactive\PWF_QE_Reporter_Based_Quan_SequestHT_ Percolator.pdProcessingWF

- ProcessingWF_LTQ_Orbitrap\PWF_OT_Reporter_Based_Quan_HCD_SequestHT_ Percolator.pdProcessingWF

Setting up the workflow for TMT and iTRAQ quantification is basically the same.

❖ **To create a processing workflow for reporter ion quantification**

> **Note** This procedure uses a TMT SPS MS3 10plex example.

1. In the Workflow Editor, follow the general instructions for creating a processing workflow. See "Creating a Processing Workflow" on page 68.

   The processing workflow must include the following nodes as a minimum:

   - Spectrum Files node

   - Spectrum Selector node

   - A search engine node (Mascot or Sequest HT)

   - Reporter Ions Quantifier node

   - PSM Validator node, Target Decoy PSM Validator node, or Percolator node

2. In the Processing Workflow window, drag the **Spectrum Files** node from the Data Input area of the Workflow Nodes pane to the Workflow Tree pane.

3. Drag the **Spectrum Selector** node to the workspace.

4. Connect the Spectrum Selector node to the Spectrum Files node.

5. Drag the **Reporter Ions Quantifier** node to the Workflow Tree pane, and attach it directly to the Spectrum Files node.

   The Reporter Ions Quantifier node performs quantification for isobarically labeled amino acids.

   > **Note** You cannot use the Reporter Ions Quantifier node in a workflow that includes the Minora Feature Detector node.

6. Drag the appropriate search engine node—for example, **Sequest HT**—to the Workflow Tree pane, and attach it to the Spectrum Selector node.

7. Drag the **Fixed Value PSM Validator** node, **Target Decoy PSM Validator** node, or the **Percolator** node to the Workflow Tree pane and attach it to the search engine node.

   Figure 282 illustrates the basic processing workflow for reporter ion quantification.

Figure 282. Basic processing workflow for reporter ion quantification
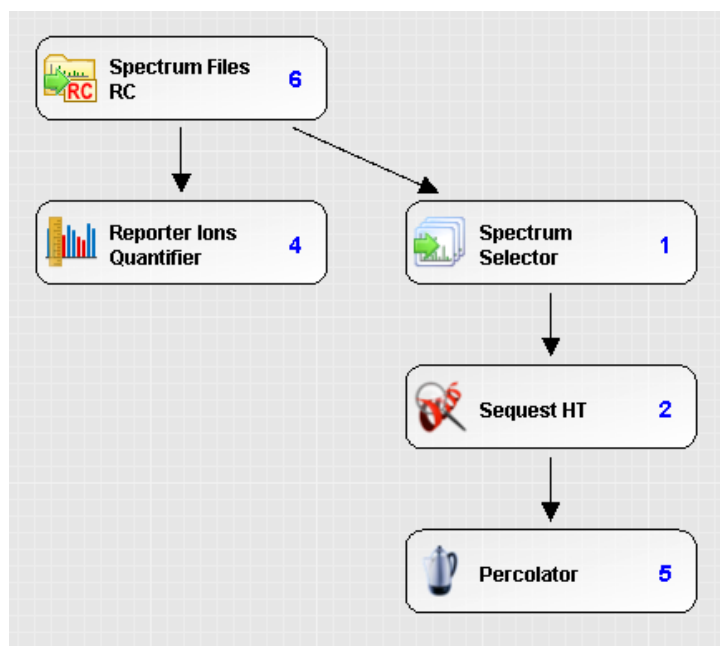
8. (Optional) Add any other nodes that you want and connect them to the workflow.

9. In the Parameters pane of the Workflow Editor, click **Show Advanced Parameters**.

10. Click the search engine node—for example, **Sequest HT**—and set the parameters for it in the Parameters pane:

    a. In the Protein Database box, select the appropriate FASTA database.

    b. In the Dynamic Modifications area, select the dynamic modifications from these options:

    > **Note** When you are using the Mascot search engine node, you can use the Quan Modifications parameter rather than the Dynamic Modifications parameters to specify the modifications to search for. For instructions on using this method, see "Checking the Quantification Method" on page 554.

    -   TMT 2plex:
        -   TMT 2plex for lysine and N-terminal (Use these as static or dynamic modifications.)

    -   TMTe 6plex or TMT 6plex:
        -   TMT 6plex for lysine and N-terminal (Use these as static or dynamic modifications.)

    -   TMT 10plex or TMT 11plex: the same modifications as for TMT 6plex

    -   iodo TMT 6plex: iodo TMT 6plex for cysteine (Use these as static or dynamic modifications.)

For example, for TMTe 6plex quantification, you would select a static or dynamic modification of **TMT6plex / +229.163 Da (K)**. If you do not find this label, you can enable it by following the instructions in "Defining Chemical Modifications" on page 200.

    c. In the Static Modifications area, select the static modifications. For example, for TMTe 6plex quantification, you would select **TMT6plex / +229.163 Da** in the Peptide N-Terminus box.

    d. Set any other parameters that you prefer. For information on the available search engine parameters, refer to the Help.

11. Click the **Reporter Ions Quantifier** node and set the parameters for it in the Parameters pane.

For information on the parameters that you can set for the Reporter Ions Quantifier node, refer to the Help.

12. Set the parameters for all other nodes in the Parameters pane.

For information about the parameters that you can set for each node, refer to the Help.

13. (Optional) Save the processing workflow:

    a. In the Name box above the Workflow Tree pane, type a name for the processing workflow.

    b. (Optional) In the Description box above the Workflow Tree pane, type a brief description of the processing workflow.

    c. In the Workflow Editor, click   Save   or   Save Common   .

    d. In the Save Workflow dialog box, do the following:

        i. Browse to the file to save the template in, or type a file name in the File Name box.

        ii. In the Save As Type box, select **Processing Workflow File (*.pdProcessingWF)**.

        iii. Click **Save**.

The application saves the workflow in the *file_name*.pdProcessingWF file.

Creating a Consensus Workflow for Reporter Ion Quantification

To use a reporter ion quantification method, you must use a consensus workflow that includes the Reporter Ions Quantifier node.

Default consensus workflows are available. You can use these default workflows as is or modify them to suit your needs. To access them, click [Open Common] on the Workflows page, or see "Using Common Workflow Templates" on page 102. When you use a default workflow, you do not need to perform step 1 through step 3 in the following procedure because these steps are already included in the workflow. However, you must check the node parameters in the workflow to make sure they are appropriate for your data and add modifications to the search for the quantification to be successful.

The consensus workflow must be selected in the Workflow Tree window for the Open Common command to display the processing workflows available.

The default consensus workflow for reporter ion quantification is ConsensusWF\CWF_Comprehensive_Enhanced Annotation_Reporter_Quan.pdConsensus WF.

Setting up the workflow for TMT and iTRAQ quantification is basically the same.

❖ **To create a consensus workflow for reporter ion quantification**

1. Follow the general instructions for creating a consensus workflow. See "Incorporating an Existing Workflow into a Study and an Analysis" on page 99.

   The consensus workflow must include the following nodes as a minimum:

   - MSF Files node
   - PSM Grouper node
   - Peptide Validator node
   - Peptide and Protein Filter node
   - Protein Scorer node
   - Protein Grouping node
   - Reporter Ions Quantifier node

2. In the Processing Workflow window, drag the **MSF Files** node to the workspace.

3. Drag the **Reporter Ions Quantifier** node to the Workflow Tree pane, and connect it to the MSF Files node.

4. Drag the **PSM Grouper** node to the workspace, and connect it to the MSF Files node.

5. Drag the **Peptide Validator** node to the workspace, and connect it to the PSM Grouper node.

6. Drag the **Peptide and Protein Filter** node to the workspace, and connect it to the Peptide Validator node.

7. Drag the **Protein Scorer** node to the workspace, and connect it to the Peptide and Protein Filter node.

8. Drag the **Protein Grouping** node to the workspace, and connect it to the Protein Scorer node.

9. (Optional) Drag the **Data Distributions** node to the Post-Processing Nodes pane.

   This node adds heat maps, which aid in data review.

   Figure 283 illustrates the basic consensus workflow for reporter ion quantification.

   Figure 283. Basic consensus workflow for reporter ion quantification
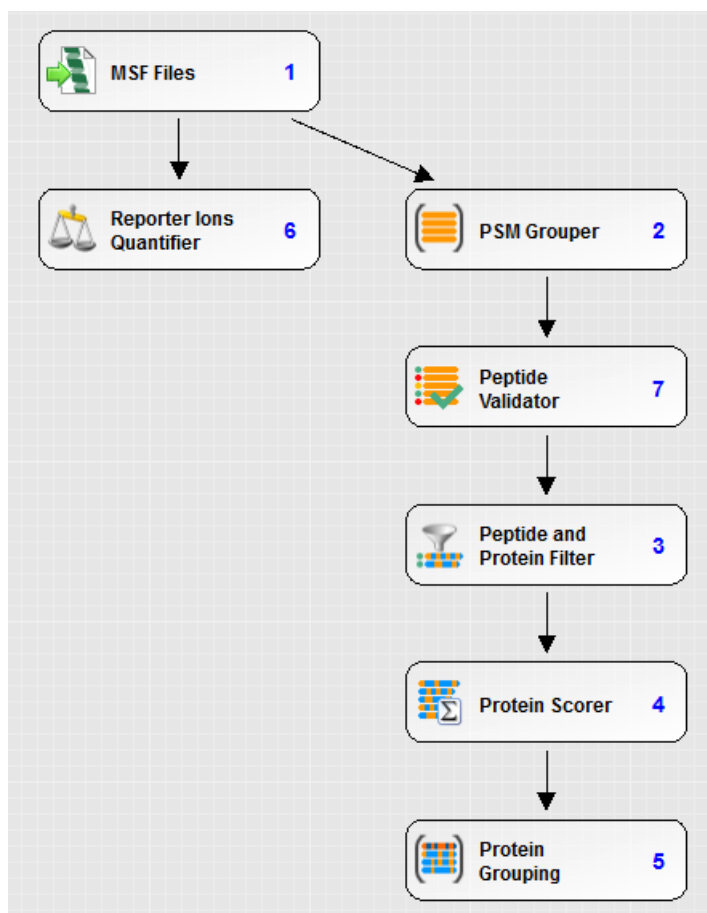
10. (Optional) Add any other appropriate nodes and connect all the nodes together.

11. Set the parameters for each node.

    For information about all the parameters that you set for each node, refer to the Help.

12. For the Reporter Ions Quantifier node, use the following settings:

    • Imputation Mode: See "Imputing Missing Values" on page 520.

    • Normalization and Scaling: See "Normalizing Peptide Groups and Protein Abundances" on page 535.

13. (Optional) Save the consensus workflow:

    a. In the Name box above the Workflow Tree pane, type a name for the consensus workflow.

    b. (Optional) In the Description box above the Workflow Tree pane, type a brief description of the consensus workflow.

    c. In the Workflow Editor, click 🔩 Save .

    d. In the Save Workflow dialog box, do the following:

        i. Browse to the file to save the template in, or type a file name in the File Name box.

        ii. In the Save As Type box, select **ConsensusWorkflow File (*.pdConsensusWF)**.

        iii. Click **Save**.

    The application saves the workflow in a *file_name*.pdConsensusWF file.

    > **Note** A yellow triangle containing an exclamation mark in the upper right corner of the Consensus Step box (⚠) usually indicates that a node in the workflow is obsolete, a parameter is missing, a required node is missing, or the output file name is invalid or missing. Hold the mouse pointer over the triangle to display details about what is missing.

Saving an Analysis

To use an analysis as a template for later reuse, you can save it as a .pdAnalysis template file.

> **Note** Studies and analyses in the Proteome Discoverer are separate, so you must save them separately. Saving a study does not save an analysis, and saving an analysis does not save a study.

❖ **To save an analysis as a template for later reuse**

1. In the upper right corner of the Analysis window, click 💾 Save .

2. In the Save Analysis Template dialog box, browse to the location where you want to store the template.

3. In the File Name box, browse to the study folder, and type or select the template file name.

4. In the Save as Type box, select **Analysis Templates (*pdAnalysis)**.

5. Click **Save**.

The application saves the analysis in a file with a .pdAnalysis extension.

The .pdAnalysis template file saves the processing and consensus workflows. It saves neither the input files nor the study variables that were selected to group the samples and quantification ratios.

Specifying Quantification Ratios from Selected Sample Groups for Reporter Ion Quantification

After you set up the workflow to use for the analysis, you can specify the ratios to report for the quantification and how to group your samples in relation to the specified factor values.

You can generate arbitrary ratios from selected sample group abundances without being restricted to predefined sample and control specifications. Previous versions of the Proteome Discoverer application automatically generated quantification ratios that relied on the specification of sample or control sample types. This feature is appropriate for some experiments but too restrictive for experiments in which the same sample should occur in the denominator for some ratios and in the numerator for others.

For example, in a typical mutant-versus-wild-type experiment, you usually would monitor ratios like this to study the effect of the treatment on a mutant:

$$\frac{wt\_treat}{wt}, \frac{m\_treat}{m}, \frac{m\_treat}{wt\_treat}$$

where:

- *wt_treat* is a treated wild type.

- *wt* is a wild type.

- *m_treat* is a treated mutant.

- *m* is a mutant.

In some experiments there is no clear control available and all pair-wise comparisons of available samples are of interest.

On the Grouping & Quantification page of a study, you can generate custom quantification ratios that do not restrict you to predefined sample and control specifications.

In the Grouping & Quantification view, you first select the study variables that were used to group your samples and select the numerators and denominators of your ratios. Then, you manually specify the quantification ratios to generate or semiautomatically create all possible quantification ratios against an ad-hoc selection of denominator values.

❖ **To open the Grouping & Quantification page**

On the Study: *Study_name* page, click the **Grouping & Quantification** tab.

> **Note** The Grouping & Quantification tab does not appear until you add input files that are associated with a single quantification method.

Figure 284 shows the Grouping & Quantification page of the study.

Figure 284. Grouping and Quantification page

The page contains the following areas:

- Study Variables area: At the top left is a list of the selectable study factors, or study variables, that you specified in the Study Factors area of the Study Definition page. A study factor is anything that captures the difference between two samples, for example, drug treatment, the time of drug application, or differences in tissue, organisms, or patients. It can be differences in sample preparation, chromatography settings, or acquisition parameters; or differences in the isobaric or metabolic labels used. Study factors can vary from sample to sample. In the example used throughout this topic, the study factors are the quantification channels associated with each sample and the factors for the acquisition method and tissue. Selecting the study factors to use for grouping provides an effective means of sorting the samples and quantification ratios into replicate and treatment groups.

  – Files: Groups samples and quantification ratios by files.

  – Quan Channels: Groups samples and quantification ratios by quantification channels.

  – *Study_factors*: Groups samples and quantification ratios by user-defined study factors.

– Sample Types: Groups factors that include Sample, Control, Standard, and Blank. Variables displayed in italics contain only a single value.

Study factors are user-defined, but the File, Quan Channel, and Sample Type variables always appear on the Grouping & Quantification page for all studies.

- Manual Ratio Generation area: Contains menus where you can select the numerator and denominator for each quantification ratio.

- Bulk Ratio Generation area: Displays the study factor values to use as the denominators for semiautomatically generated quantification ratios.

- Generated Sample Groups area: Displays samples grouped by the values set for the selected study factors. The application ignores any other differences among the samples that might be present.

Selecting these variables indicates that you want to group your samples and quantification ratios according to the specified study variables and ignore any other difference between the samples that might be present. For example, if you group by tissue, you put into the same sample group all samples that share the same tissue factor. You also group all quantification ratios together that have the same tissue in the numerator or denominator. However, you ignore any differences in the acquisition method used.

The grouping of samples affects how identifications are displayed in the distribution maps (see "Sample Information Used to Display Identifications and Quantifications Among Files and Samples" on page 578) and the calculation of areas from precursor ion quantification. For the latter, the application reports separate area values for each sample or sample group.

When you group ratios, the application calculates an averaged ratio for a group of ratios that are similar with respect to the selected study variable in the numerator and denominator.

> **Note** The application first groups samples to average abundance values from replicate samples and then calculates ratios from the grouped abundances. The ratios are freely defined in the Grouping and Quantification view, without being restricted to predefined control samples.

If the application does not use all available sample group values in at least one ratio, it displays a warning in the Generated Sample Groups area, like that shown in Figure 285, and marks the unused sample groups.

Figure 285. Warning in the Generated Sample Groups area
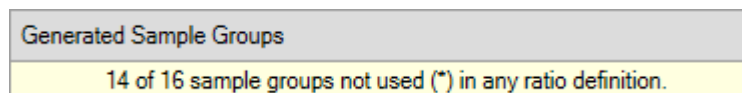
Generated Sample Groups
14 of 16 sample groups not used (*) in any ratio definition.

- Generated Ratios area: Displays the ratios generated.

❖ **To generate custom quantification ratios semiautomatically**

1. In the Study Variables area, select the check box of the study factors, or variables, that you want to use to group your samples and from which you want to draw the numerators and denominators of the ratios. For the example used throughout this topic, select the **(Time (days)** check box to indicate that the samples and quantification ratios are grouped by time interval.

   After you select the study factors, the Generated Sample Groups area displays the generated sample groups. When performing the quantification, the application calculates abundance values for each sample and averages the abundance values of all samples in a sample group.

   This step creates sample groups using the selectable study factors available for the files to be analyzed. The application places all samples that have the same combination of study factor values in the same group.

   The order of the study factors is relevant for the semiautomatic generation of ratios.

2. (Optional) To change the placement of a study factor in the list of study factors, do the following:

   a. In the Study Variables area, select the check box for a study factor.

   A placement handle in the form of a green rectangle appears to the left of the selected check box, as shown in Figure 286.

   Figure 286. Placement handle in the Study Variables area

   b. Hold the cursor over the placement handle.

   White up and down arrows now appear on the handle.

   c. Drag the cursor up or down to move the variable to its new place in the list of variables, or click the up or down arrows to move the study factor.

3. (Optional) To sort study factors in the Bulk Ratio Generation area and sample groups in the Generated Sample Groups area, click one of the following to the right of each study factor in the Study Variable area.

   • To sort these items in descending order, click the **Sort Descending** icon, ⯆.

   • To sort these items in ascending order, click the **Sort Ascending icon**, ⯅.

   • To leave these items unsorted, click the **No Sorting** icon, ☰.

4. In the Bulk Ratio Generation area, select the check box for the time point to use in the denominator of the ratio. For the example, select **Time (days): 5**.

The Bulk Ratio Generation area displays a list of the denominator values for this type of study factor. If you select only one study factor, it displays a list of the available denominator values for this factor, as shown in Figure 287.

Figure 287. Denominator values available for a single study factor

If you select multiple study factors, the Bulk Ratio Generation area displays the denominator values available for each factor.

5. (Optional) To select the same study factor for all the denominators, do the following:

a. Hold the cursor over a denominator value.

An icon containing four check boxes in a square appears on the left side of that item, as shown in Figure 288.

b. Click the icon.

The application selects the same study factor for all denominators (see Figure 288).

Figure 288. Icon for multiple selection of denominators in the Bulk Ratio Generation area
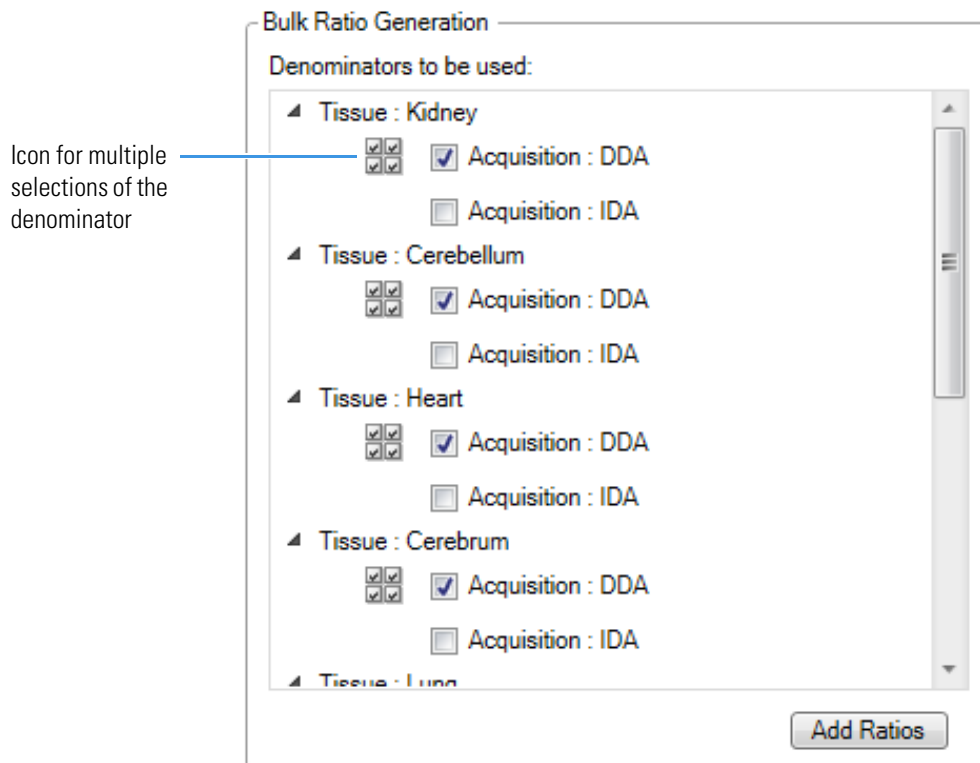
Icon for multiple
selections of the
denominator

6. Click **Add Ratios**.

The application generates all possible ratios against the selected denominator values and
adds them to the Generated Ratios area. Figure 289 shows the generated quantification
ratios and ratio groups in the Generate Ratios area after the selection of Time (days) study
variable to group by and 5 as the denominator to use.

Figure 289. Quantification ratios and ratio groups generated by grouping the samples and ratios by time point
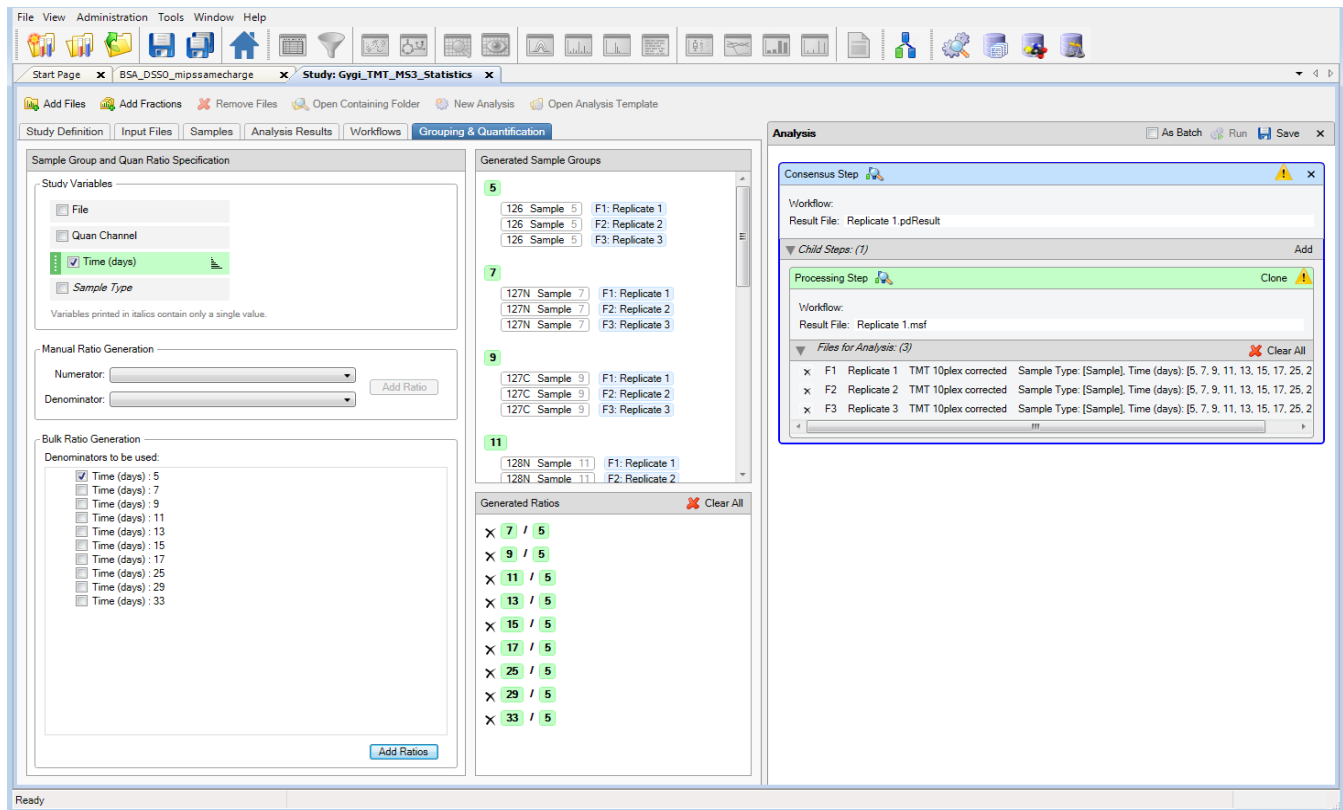
❖ **To generate custom quantification ratios manually**

1. On the Study: *Study_name* page, click the **Grouping & Quantification** tab.

   > **Note** The Grouping & Quantification tab does not appear until you add or open an analysis and specify Input Files for the analysis.

2. In the Study Variables area, select the check box of the study factor or factors that you want to use to group your samples and from which you want to draw the numerators and denominators of the ratios. For the example used throughout this chapter, select the **Time (days)** check box to indicate that the samples and quantification ratios are grouped by time interval.

3. In the Manual Ratio Generation area, select the numerator value from the Numerator list, for example, **7**.

4. Select the denominator value from the Denominator list, for example **33**.

5. Click **Add Ratio**.

   The generated ratio appears in the Generated Ratios area, as shown in Figure 290.

Figure 290. Ratio manually generated for one study factor
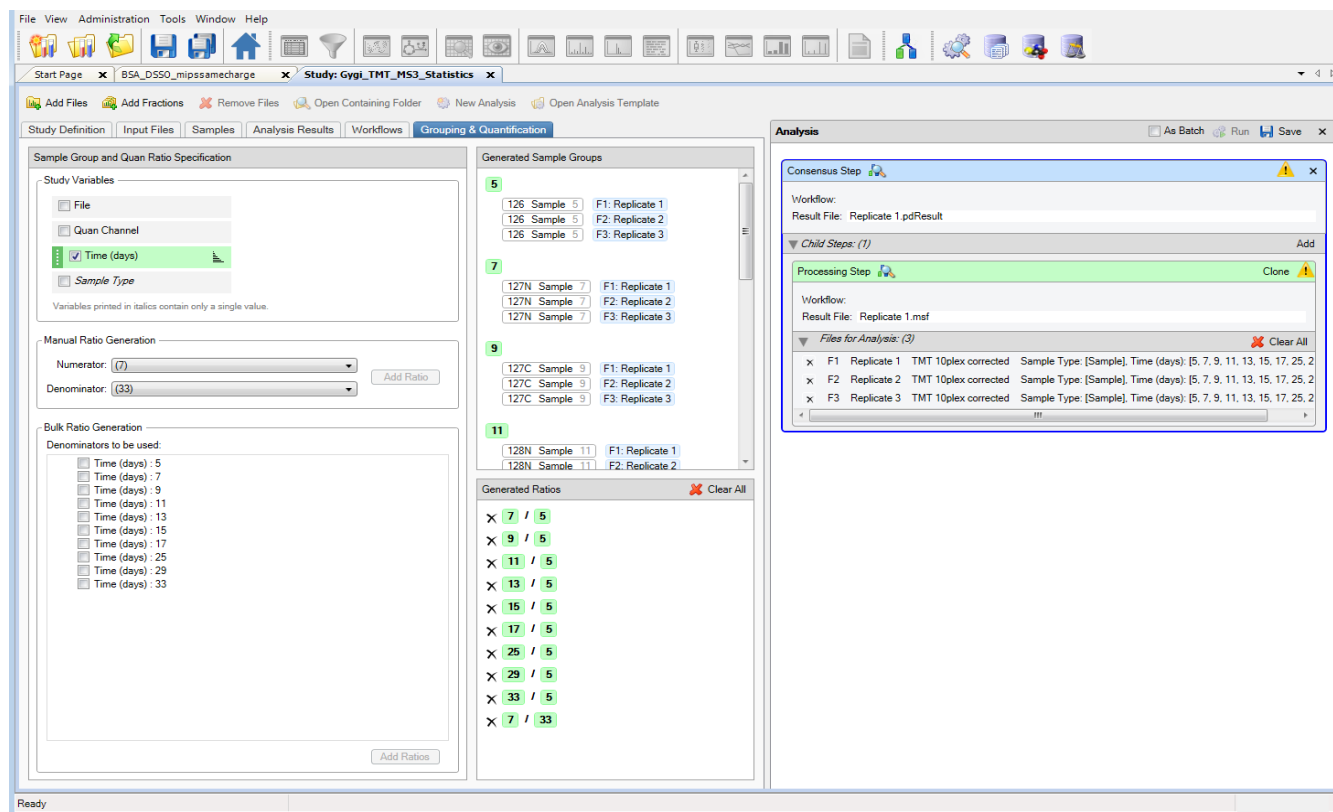
❖ **To generate custom quantification ratios based on channels**

1. In the Study Variables area, select the **Quan Channel** check box.

2. To generate quantification ratios semiautomatically, follow the procedure "To generate custom quantification ratios semiautomatically" on page 396.

   –or–

   To generate quantification ratios manually, follow the procedure "To generate custom quantification ratios manually" on page 399.

   Figure 291 shows the ratios generated in the Generated Ratios area.

Figure 291. Ratios generated by quantification channel

❖ **To save the settings on the Grouping & Quantification page**

> **Note** The application does not save the settings on the Grouping & Quantification page with a study or with an analysis. Instead, it associates the settings with search results, so you must load them from data sets that have already been processed within the study or recreate them from the beginning.

1. Click the **Analysis Results** tab of the study.

2. Select the result on the Analysis Results page, and choose either **Reprocess > All Analysis Steps** or **Reprocess > Last Consensus Step**.

    If you choose Use Results to Make New (Multi) Consensus, the Grouping and Quantification page no longer displays ratios.

Performing the Quantification

❖ **To perform the quantification**

Click ⟨ Run ⟩ .

The job queue appears, as shown in Figure 53 on page 83, displaying the status of your search. Use the job queue to check the status of your search as the search progresses. For information about the job queue, refer to the Help.

The application creates one or more MSF and .pdResult files, depending on whether the data files to be analyzed are batch-processed. It might reuse MSF files to produce a new .pdResult file if a different report for the given data is required.

The .pdResult file has the same name as the input file but with a .pdResult extension. The Analysis Results page displays the name of the .pdResult file.

Working with the Search Results

To open, delete, convert, and reprocess search results and to view the workflow and the analysis after the search, see "Working with the Search Results" on page 89.

Performing TMT Quantification on HCD and CID Scans

If a raw data file contains both CID scans for identification and HCD scans for quantification, you can use the following workflow to both quantify the HCD scans and identify peptides in the CID scans, the HCD scans, or both.

❖ **To perform TMT quantification on HCD and CID scans**

1. Drag the **Reporter Ions Quantifier** node to the Workflow Tree pane and connect it to the workflow.

2. Set the Activation Type parameter for the Reporter Ions Quantifier node to **HCD**.

3. Set the Activation Type parameter for the Spectrum Selector node to **Any**, **Is CID**, **HCD**, or **Is CID**, depending on your method setup and identification strategy.

4. Set all other parameters—modifications, tolerances, FASTA files, and so forth.

5. Click ⟨ Run ⟩ .

Calculating and Validating Raw Quantification Values in Reporter Ion Quantification

In reporter ion quantification, a Quan Spectra page displays all spectra that were used for quantification. A quantification spectrum shows the abundances of the different mass or reporter tags detected in the spectrum. The Reporter Ion Quantifier node creates and fills this table, but it does not fill in abundance values. Instead, it stores the detected abundances of the different mass and reporter tags in another hidden table. The node also connects quantification spectra with PSMs that were identified for the same precursor ion.

The Reporter Ions Quantifier node determines abundance values and the average Reporter S/N value for the quantification spectra. Then it validates the quantification spectra. The Quan Info and Quan Info Details columns on the Quan Spectra page show whether a quantification spectrum can be used for quantification (the Quan Info Details column is hidden by default and shows why a quantification spectrum is rejected by method):

- No Quan Values: Indicates that no tag was found.

- Missing Values: Indicates that a tag value is missing, the Reject Quan Results with Missing Channels parameter of the Reporter Ions Quantifier node is set to true, and no imputation is used.

- Rejected By Method: Indicates that the precursor interference is above the user-set threshold or that the average reporter ion signal noise value is too low.

- Filtered by Isolation Interference: Indicates that the spectrum isolation interference is above a user-defined threshold.

- Filtered by Average S/N: Indicates that the average signal-to-noise value is below a user-defined threshold.

- Shared: Indicates that different peptide sequences are associated with a quantification spectrum.

Correcting Reporter Ion Quantification Results for Isotopic Impurities

This topic explains how to set up a quantification method for isotopic impurities.

When you buy a TMT kit from Pierce, you receive a product data sheet called the Certificate of Analysis (COA), which summarizes the essential information about the kit and the lot that it was taken from. The COA includes a table with the exact masses of the different labels and their isotopic impurities. For reporter ion methods, you can add quantification channel data like monoisotopic mass, isotopic impurities, and associated isotopic variant relations as they are stated in the COA to the corresponding quantification method. With existing reporter ion quantification methods, you can only change the values of the isotopic impurities. The impurity values of the default methods are zero, so you must first enter the correct values from the COA to make the purity corrections work.

You can create a new method for correcting reporter ion quantification results through any of the following means:

- Add a new method based on a factory default method as a template and fill in the reporter ion correction factor table.

- Use an existing method as a template.

- Create a new method from the beginning. This method requires an additional license.

In this new method, you can use all of the available labels in a kit or only a subset.

When you create a new quantification method from the factory default template or an existing method, you cannot change the reporter ion masses, change the mass tags, or edit the relations among the impurities. For instructions on creating a new method from the first two templates, see "Creating a Quantification Method for Reporter Ion Quantification" on page 352. To create a new method from the beginning, see the next topic.

❖ **To create a quantification method for correcting reporter ion quantification results**

> **Note** The following procedure uses a TMT 10plex kit as an example.

1. Choose **Administration > Maintain Quantification Methods**, or click the **Maintain Quantification Methods** icon, .

   The Quantification Methods view opens, as shown in Figure 246 on page 353. It lists all of the available methods for both precursor ion and reporter ion quantification.

2. Click ✚ Add .

   The Create New Quantification Method dialog box now opens, as shown in Figure 292.

   Figure 292. Create New Quantification Method dialog box
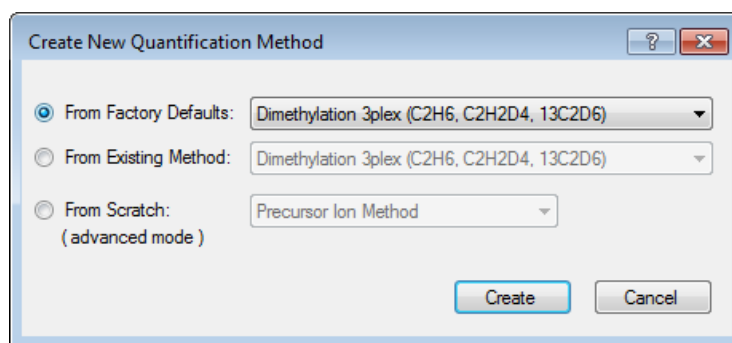
3. Select the **From Factory Defaults** option, and then select the TMT 10plex method as shown in the following figure.

Figure 293. Create New Quantification Method dialog box for creating TMT 10plex method
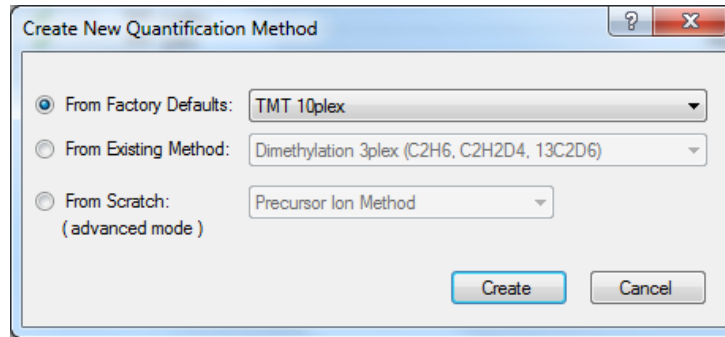
4. Click **Create**.

The Quantification Method Editor dialog box opens, as shown in Figure 294. The Quantification Method Editor dialog box for reporter ion quantification is different from the Quantification Method Editor dialog box for precursor ion quantification.

The Quantification Method Editor dialog box for reporter ion quantification displays channels with correction factors set to zero (see Figure 294).

Figure 294. Quantification Method Editor dialog box showing correction factors set to zero
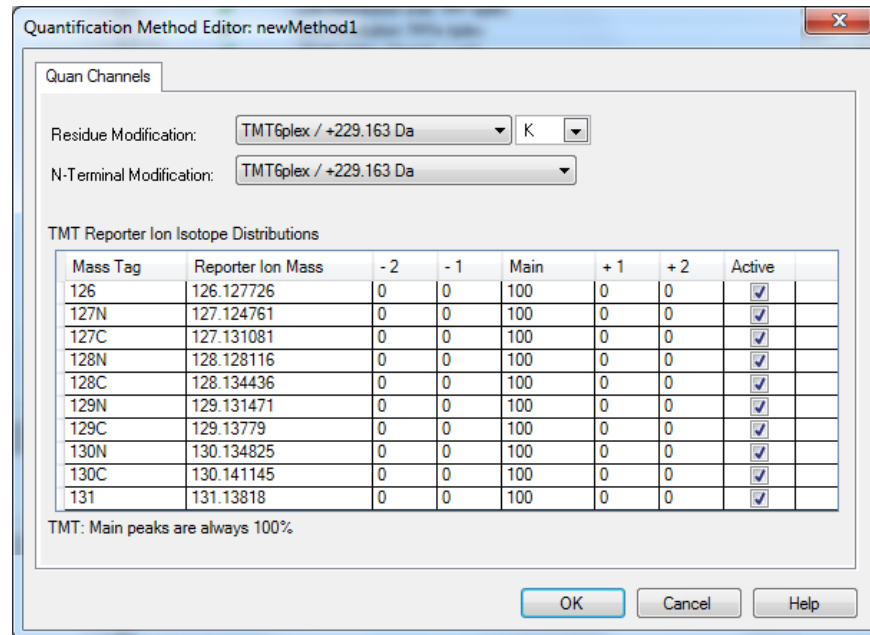
5. Type the data from the TMT 10plex kit into the appropriate columns of the Quantification Method Editor dialog box, as shown in Figure 295.

Figure 295. Quantification Method Editor dialog box with values entered from a TMT 10plex kit

The application displays channels that are not active with a gray background, as shown in Figure 296.

Figure 296. Quantification Method Editor dialog box showing inactive channels

6. Click **OK**.

   The Save Quantification Method dialog box opens, as shown in Figure 297.

   Figure 297. Save Quantification Method dialog box
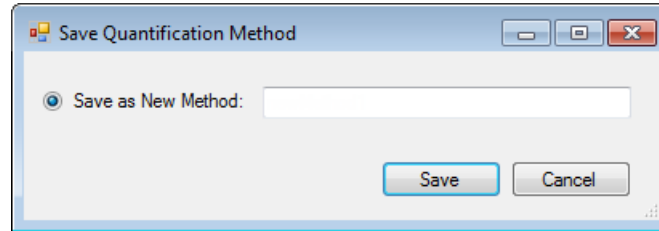
7. Type the name of the new method in the Save as New Method box, and click **Save**.

   The name of the new method now appears in the Method Name column in the Quantification Methods view.

   When you edit factory-provided methods, the application automatically sets the mass tag relations and stores them in the method definition. You cannot edit them when you create a new method from the beginning. If you want to edit a mass tag, you must create a new method.

   When you double-click the name of the new method in the Quantification Methods view, the Quantification Method Editor opens, but now a usage status column replaces the Active column. It displays text indicating whether each channel is used or not used, as shown in Figure 298.

   Figure 298. Quantification Method Editor showing usage status column
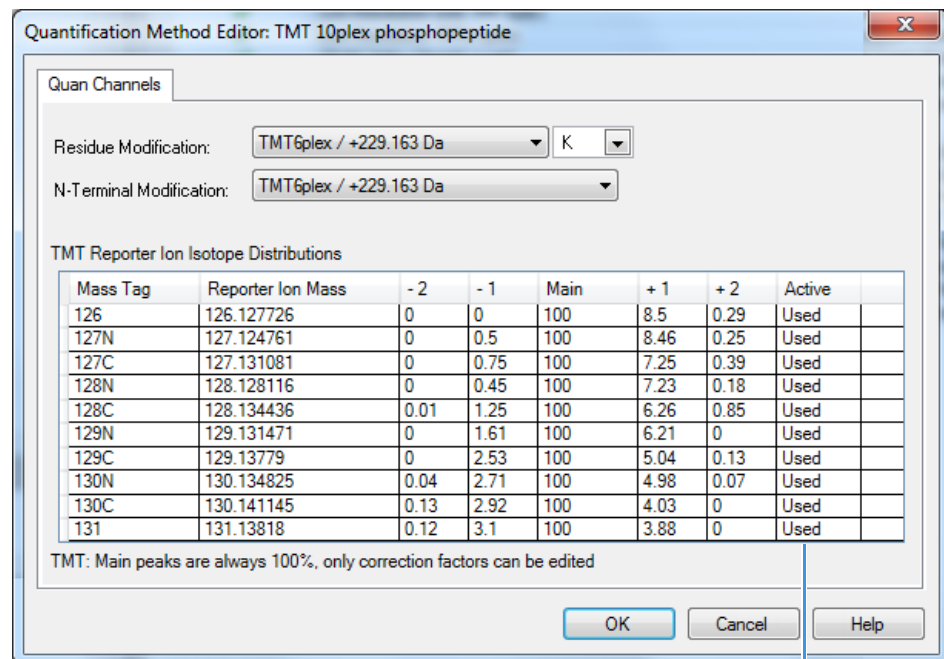
   Usage status column

Filtering Quantification Data with Signal-to-Noise Values

For Orbitrap data, you can filter reporter ion quantification spectra that have too much variability in the intensities of the reporter ions by using signal-to-noise values instead of intensities.

Using Signal-to-Noise Values as Quantification Channel Values

Because of ion statistical effects, measurements of intensities are more variable and therefore less accurate when the mass peaks arise from a smaller population of ions. When you perform precision quantification, you might want to filter out reporter peaks or spectra that have a variability that is too high because of ion statistical effects. However, with an Orbitrap analyzer, the intensity of a signal is not proportional to the number of ions corresponding to the signal. Instead, the number of ions is proportional to the S/N value of a peak (*number_of_ions* = 6 × S/N for D30 Orbitraps, and *number_of_ions* = 4 × S/N for D20 Orbitraps).

To filter reporter ion quantification input data that is detected from a number of ions that is too small, you use the signal-to-noise (S/N) values of the reporter ion peaks instead of their intensities to provide such a measurement when ions are acquired in the Orbitrap. You can use the settings of the Reporter Abundance Based On parameter of the Reporter Ions Quantifier node to select signal-to-noise intensity values. For information on this parameter, see the Help.

Filtering Quantification Data with Average Reporter Ion Signal-to-Noise Values

For Orbitrap data, you can obtain more reliable TMT quantification results by filtering out quantification spectra with too few ions for the reporter ion signals. The Reporter Ions Quantifier node includes a parameter, Average Reporter S/N Threshold, that you can use to specify an average reporter S/N threshold value that determines which PSMs the application excludes from quantification. The application filters quantification values by excluding PSMs that have an average reporter ion S/N value smaller than this threshold value.

In addition, the Average Reporter S/N threshold parameter adds the Average Reporter S/N column, which displays average reporter ion S/N values on the PSMs and Quan Spectra pages. The application calculates the values as the sum of S/N values found, divided by the number of defined tags.

In the following example, the application calculates the average S/N values for a .pdResult file created from an MSF file containing TMT 6plex data. The quantification method has six channels, and the application calculates the average S/N value by sum(S/N) divided by 6. Table 28 shows the results.

Table 28. Average S/N values for a single MSF file containing TMT 6plex data

Peptide	126	127	128	129	130	131	Average S/N
1	10	10	12	8	7	3	50/6=8.333
2	10		8		2		20/6=3.333
3							0/6=0

In the next example, the application calculates the average S/N values for a .pdResult file created from two MSF files containing TMT 6plex data. The quantification method for both files has six channels, and the application calculates the average S/N by sum(S/N) divided by 6. Table 29 shows the results.

Table 29. Average S/N values for two combined MSF files containing TMT 6plex data

Peptide	126	127	128	129	130	131	Average S/N
A1	10	10	12	8	7	3	50/6=8.333
A2	10		8		2		20/6=3.333
B1	20	21	17	22		10	80/6=13.333
B2							0/6=0

When you combine MSF files containing reporter ion quantification data with MSF files lacking reporter ion quantification data, the PSMs without reporter ion quantification data display no values in the Average Reporter S/N column on the PSMs and Quan Spectra pages.

Performing Precursor Ion Quantification

This chapter describes how to perform precursor ion quantification.

Contents

Precursor Ion Quantification

In precursor ion quantification, also called isotopically labeled quantification, protein abundance is determined from the relative MS signal intensities of an isotopically labeled sample and an unlabeled control sample. (However, this type of quantification does not necessarily have to use labeled and unlabeled samples. Some laboratories use heavy and medium labels to eliminate the bias of unlabeled peptides.)

Stable-isotope labeling by amino acid in cell culture (SILAC™) is a proteomics identification and quantification technique that uses in-vivo metabolic labeling to detect differences in the abundance of proteins in multiple samples. It is a type of isotopically labeled quantification, which uses stable (non-radioactive) heavy isotopes as labels. You can also introduce the stable isotopes by chemical labeling at the protein or peptide level with the isotopomeric tags, for example, dimethyl labeling.

This discussion of precursor ion quantification includes the following topics:

- SILAC 2plex Methods

- SILAC 3plex Methods

- Dimethylation 3plex Method

- 18O Labeling Method

Identification

Proteome Discoverer quantifies all features in the sample regardless of whether or not it was first identified. The matching of quantification results to peptide identifications is first performed in the Processing workflow in the Minora Feature Detector node for Precursor Ion Quantification. For information on the Minora algorithm, see "Minora Feature Detection" on page 468.

Chromatographic Peaks Used

Instead of using the area of all isotope pattern peaks as the quantification value for a chromatographic feature, precursor ion quantification uses the largest peak by default to obtain more accurate results. By default, it uses the heights of the peaks at the apex of the chromatographic profile instead of the integrated peak area, but this can be changed in the Precursor Ions Quantifier node in the Consensus workflow.

Default Quantification Methods

The following default quantification methods are available for precursor ion (isotopically labeled) quantification:

- SILAC 2plex (Arg10, Lys6): Uses label 13C(6)15N(4) for arginine and label 13C(6) for lysine.

- SILAC 2plex (Arg10, Lys8): Uses label 13C(6)15N(4) for arginine and 13C(6)15N(2) for lysine.

- SILAC 2plex (Ile6): Uses label 13C(6) on isoleucine.

- SILAC 3plex (Arg6, Lys4|Arg10, Lys8): Uses heavy label 13C(6)15N(4) for arginine and 13C(6)15N(2) for lysine. Uses medium label 13C(6) for arginine and 2H(4) for lysine.

- SILAC 3plex (Arg6, Lys6|Arg10, Lys8): Uses heavy label 13C(6)15N(4) for arginine and 13C(6)15N(2) for lysine. Uses medium label 13C(6) for arginine and 13C(6) for lysine.

- Dimethylation 3plex: Chemically adds isotopically labeled dimethyl groups to the N-terminus and to the ε-amino group of lysine.

- Full ^{18}O labeling: Introduces 4 Da mass tags through the enzyme-catalyzed exchange reaction of C-terminal oxygen atoms with ^{18}O. This method is for fully labeled samples.

- Incomplete ^{18}O labeling: Introduces 2 or 4 Da mass tags through the enzyme-catalyzed exchange reaction of C-terminal oxygen atoms with ^{18}O. This method is for incompletely labeled samples.

Chemical Elements Supported in Precursor Ion Quantification

You can use any chemical element and each of its available isotopes for precursor ion quantification. However, you must adhere to certain syntactical restrictions when you enter chemical compositions. The notation from unimod.org, which sets the nominal mass prefix in parentheses—for example, (13)C4 (15)N2—only applies to the elemental isotopes ^{2}H, ^{13}C, ^{15}N, and ^{18}O.

SILAC 2plex Methods

In a typical SILAC quantification experiment, two cell populations grow in media that are deficient in lysine and arginine. One population grows in a medium containing normal ("light") amino acids, such as lysine ($^{12}C_6$ $^{14}N_2$). The other population grows in a medium containing amino acids where stable heavy isotopes, such as lysine 6 ($^{13}C_6$ $^{14}N_2$) or lysine 8 ($^{13}C_6$ $^{15}N_2$), have been substituted for normal atoms. SILAC quantification usually uses "heavy" arginine and lysine, because these are the cleavage sites for the generally used trypsin protease. Both populations incorporate these amino acids into proteins through natural cellular protein synthesis. The cells growing in the medium with the heavy isotopes incorporate these isotopes into all of their proteins.

After altering the proteome in one sample through chemical treatment or genetic manipulation, you combine equal amounts of protein from both cell populations and then digest with trypsin before MS analysis. Because peptides labeled with "heavy" and "light" amino acids are chemically identical, they co-elute during reverse-phase chromatographic separation. This means they are detected simultaneously during MS analysis. To determine the average change in protein abundance in the treated sample, you use the relative peak intensities of multiple isotopically distinct peptides from each protein, as shown in Figure 299.

Figure 299. Schematic workflow for SILAC-based peptide and protein quantification
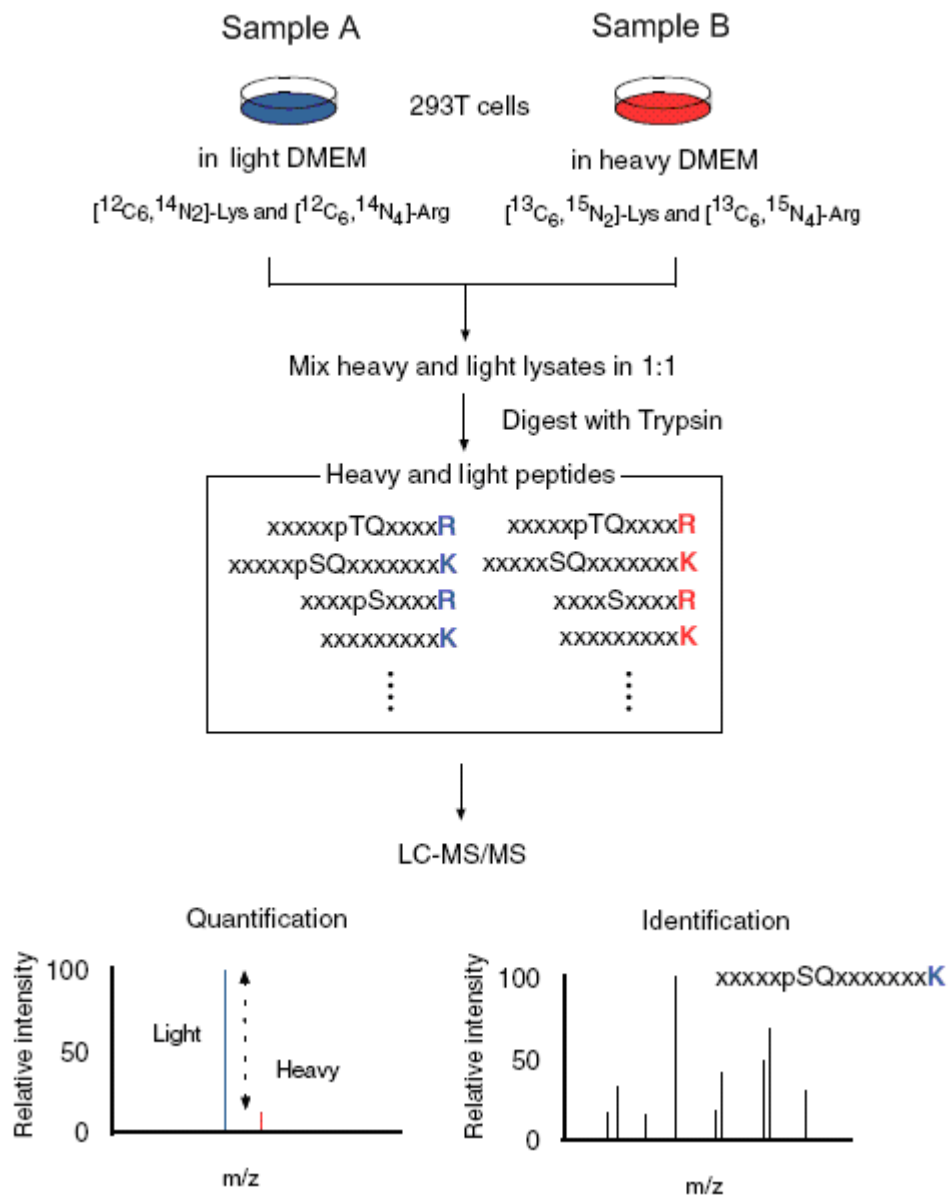
SILAC can differentiate peptides in single MS mode without requiring you to perform tandem mass spectrometry. However, SILAC cannot identify peptides, so you must use tandem mass spectrometry for that purpose.

You can choose from several SILAC 2plex methods, for example, (Arg10, Lys6) and (Arg10, Lys8), to compare two samples.

SILAC 3plex Methods

SILAC 3plex methods are similar to SILAC 2plex methods except, in addition to a "heavy" sample (containing, for example, Arg10 and Lys8), they also use a "medium" sample (containing, for example, Arg6 and Lys4). Protein abundance is determined from the relative MS signal intensities of the heavy sample, medium sample, and a control sample containing "light" (^{12}C and ^{14}N) arginine and lysine.

Dimethylation 3plex Method

The application also includes the dimethylation 3plex method, which is not metabolomic labeling in cell culture but a form of peptide chemical labeling. This method uses formaldehyde and sodium cyanoborohydride to add dimethyl groups $(CH_3)_2$ to the N-terminus and to the ε-amino group of lysine. By choosing the isotopomers of formaldehyde and sodium cyanoborohydride, you can create light, medium, and heavy labels:

- For the light label, the (natural-isotope) dimethyl group is $^{12}C_2{}^1H_6$.

- For the medium label, the dimethyl group is $^{12}C_2{}^2H_4{}^1H_2$, which is 4 Da more massive.

- For the heavy label, the dimethyl group is $^{13}C_2{}^2H_6$, which is an additional 4 Da more massive.

You can use the dimethylation 3plex method to compare up to three samples.

^{18}O Labeling Method

The ^{18}O labeling method introduces 2 or 4 Da mass labels through the enzyme-catalyzed exchange reaction of C-terminal oxygen atoms with ^{18}O.

- Full ^{18}O labeling: Introduces 4 Da mass tags through the enzyme-catalyzed exchange reaction of C-terminal oxygen atoms with ^{18}O. This method is for fully labeled samples.

- Incomplete ^{8}O labeling: Introduces 2 or 4 Da mass tags through the enzyme-catalyzed exchange reaction of C-terminal oxygen atoms with ^{18}O. This method is for incompletely labeled samples.

Performing Precursor Ion Quantification

This topic describes the default procedures to use to perform precursor ion quantification and then describes the processing and consensus workflows in detail.

This topic describes the general procedure to use to perform reporter ion quantification and then describes the processing and consensus workflows in detail.

- Allocating Sufficient Storage Space

- Downloading FASTA Files

- Checking for the Quantification Method

- Creating a Quantification Method for Precursor Ion Quantification

- Opening an Existing Study

- Creating a Study

- Adding a Description

- Selecting a Quantification Method for the Study

- Adding the Study Factors

- Adding Input Files

- Setting the Study Factor Values for the Samples

- Saving a Study

- Creating an Analysis

- Opening an Existing Analysis

- Adding Input Files to an Analysis

- Using Multiple Processing Steps in an Analysis

- Adding or Deleting a Processing Step

- Creating a Processing Workflow for Precursor Ion Quantification

- Creating a Consensus Workflow for Precursor Ion Quantification

- Saving an Analysis

- Performing the Quantification

- Working with the Search Results

- Specifying Quantification Ratios from Selected Sample Groups for Precursor Ion Quantification

Allocating Sufficient Storage Space

Be sure that you allocate enough space to store the input raw data files and the temporary result files that the application creates until it creates the final result file. See "Configuring Temporary Files Parameters" on page 32.

Downloading FASTA Files

The first step in performing reporter ion quantification is to download an appropriate FASTA file. For information on downloading a FASTA file, see "Downloading FASTA Files to Proteome Discoverer" on page 153.

Checking for the Quantification Method

Before you open or create a study, check to see if the application includes the quantification method that you want to use.

❖ **To view the quantification methods available**

Choose **Administration > Maintain Quantification Methods**.

Figure 403 on page 546 shows the Quantification Methods view, which lists all the quantification methods.

Creating a Quantification Method for Precursor Ion Quantification

If you do not see an appropriate quantification method on the Quantification Methods view, you can create one.

1. Choose **Administration > Maintain Quantification Methods**, or click the **Maintain Quantification Methods** icon, .

   The Quantification Methods view opens, as shown in Figure 300. It lists all of the available methods for precursor ion, label-free, and reporter ion quantification.

Figure 300. Quantification Methods view
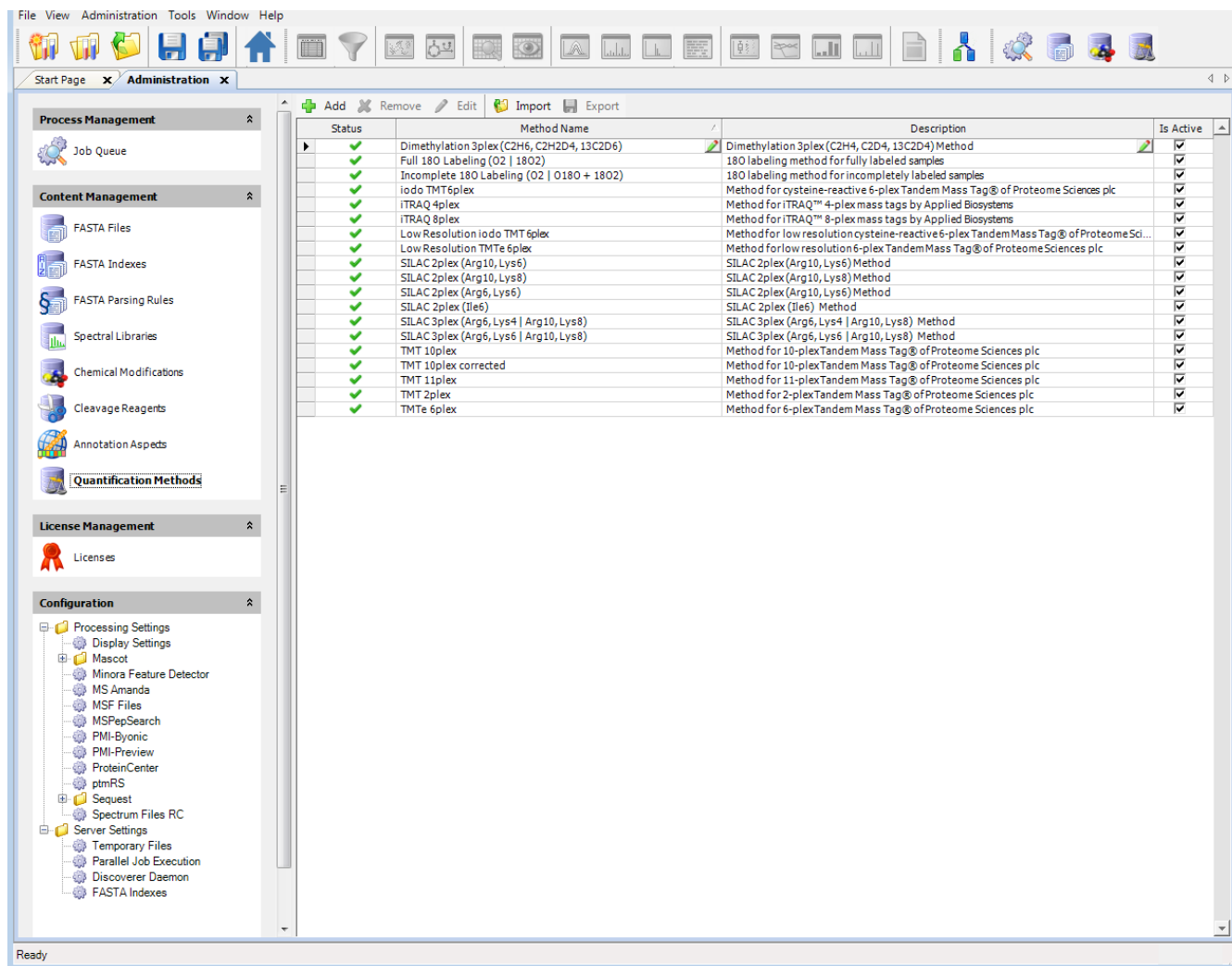
The Status column indicates whether the quantification method is valid for use in quantification:

• A green check mark means that the quantification method is valid and can be used for quantification.

• An exclamation point in a yellow triangle means that the quantification method is not valid. Double-click this mark to view a message that describes the error and provides information on how to fix it.

Figure 301 gives examples of these symbols in the Status column.

Figure 301. Method validity symbols in the Quantification Methods view
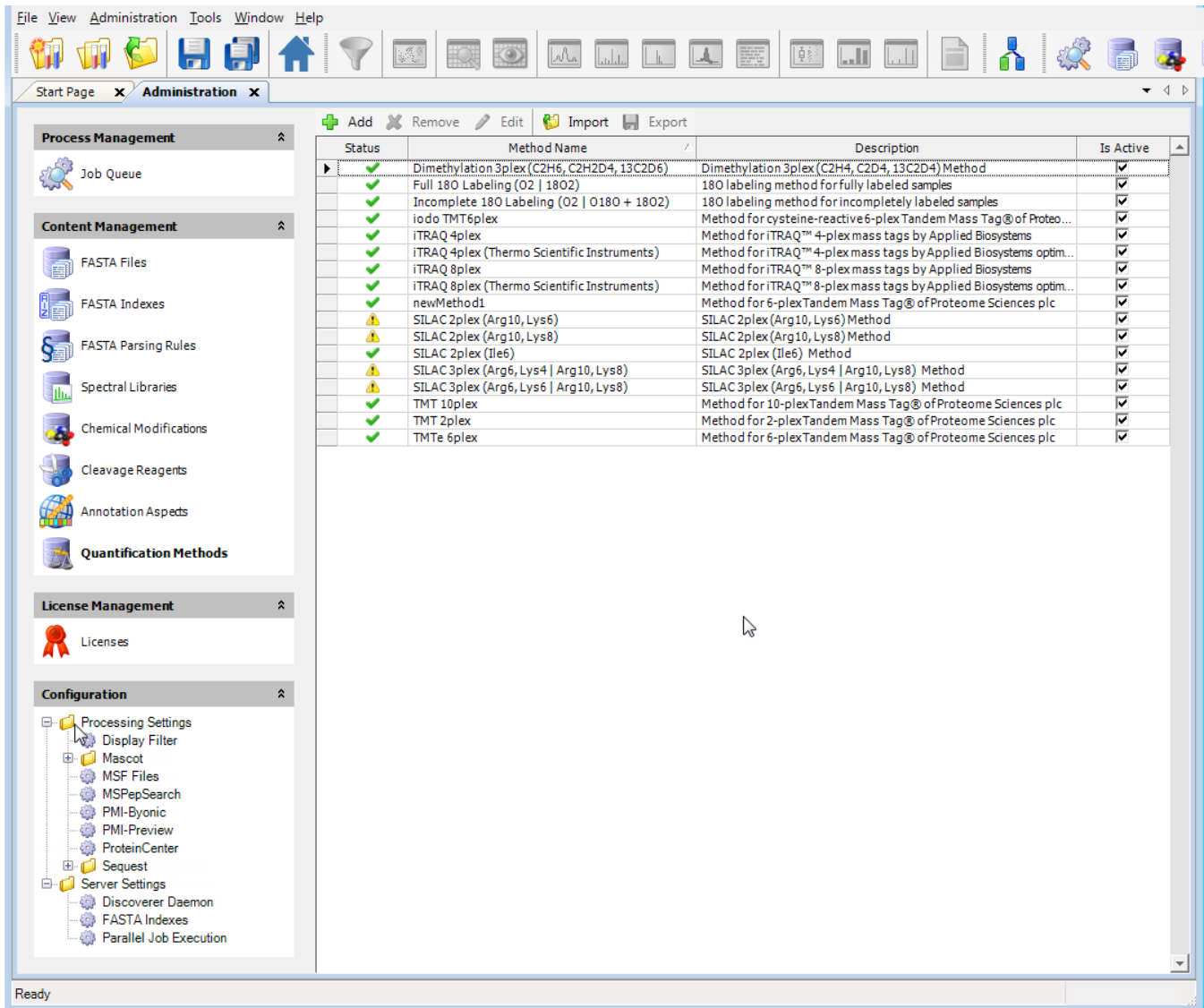
2. Double-click the row for the appropriate method in the Method Name or Description column, or click the column to the left of the Status column for the method and click
   ➕ Add .
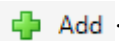
   The Create New Quantification Method dialog box opens, as shown in Figure 302.

Figure 302. Create New Quantification Method dialog box
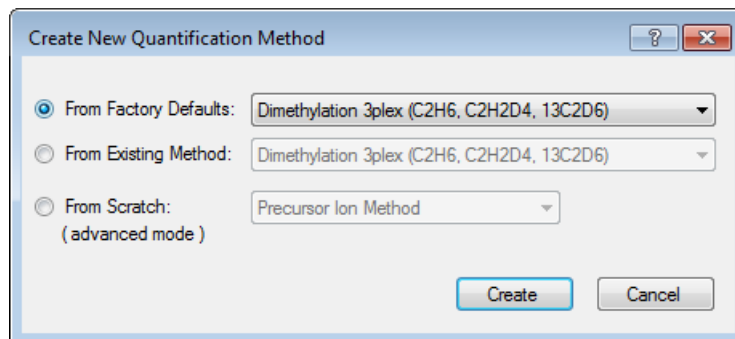
You can select one of the following methods of creating a quantification method:

- (Default) From Factory Defaults: Creates a new method using the same settings from one of the default settings that come with a new application installation.

- From Existing Method: Uses the same settings as those of the existing quantification method that you select from the list. The list of methods is the same as that given in the Method Name column of the Quantification Methods view, shown in Figure 300 on page 418.

- From Scratch: Uses one of the following templates so that you can build a new processing method from the beginning:

  - Reporter Ion Quan Method: Provides a template for reporter ion quantification. This method requires an additional license.

  - Precursor Ion Quan Method: Provides a template for precursor ion quantification.

3. Select the **From Factory Defaults** option, and from the adjacent list, select a method similar to the one that you intend to create, as shown in Figure 303.

The example shows the **SILAC 2plex (Arg10, Lys6)** method.

Figure 303. Completed Create New Quantification Method dialog box
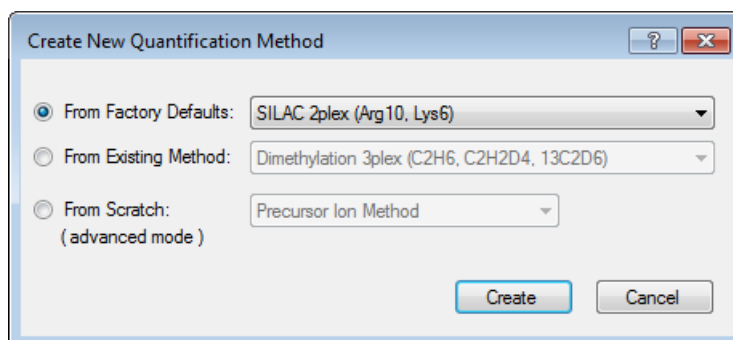
4. Click **Create**.

The Quantification Method Editor opens, as shown in Figure 304.

Figure 304. Initial Quantification Method Editor
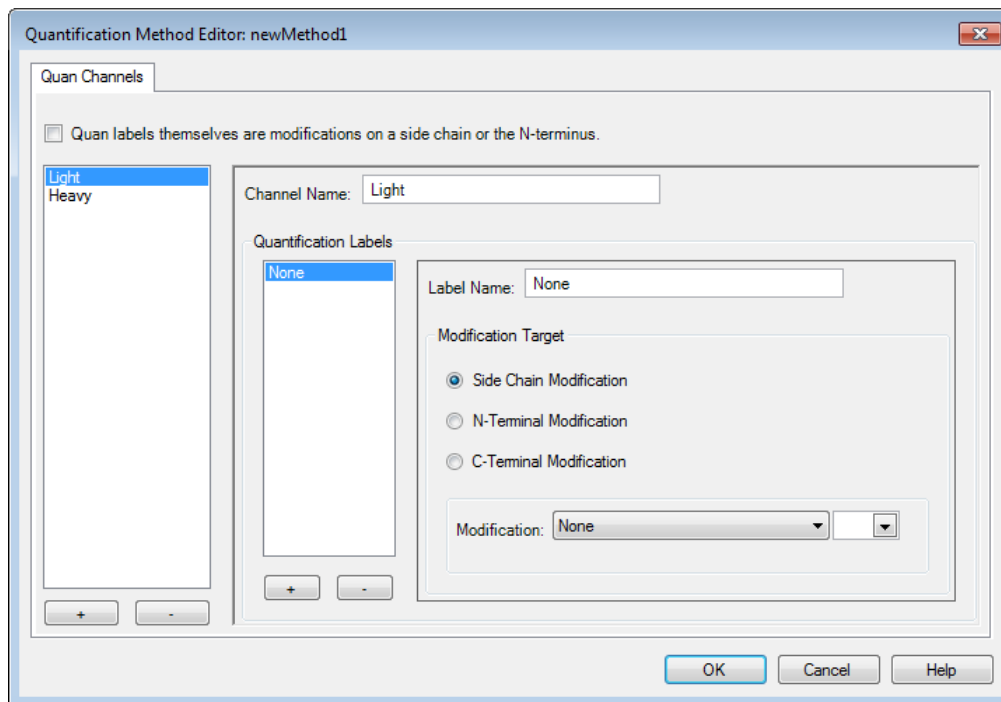
5. If you want the application to ignore modifications on the N terminus, clear the **Quan Labels Themselves Are Modifications on a Side Chain or the N-Terminus** check box.

    By default, this check box is cleared. Select this check box for dimethylation and clear it for SILAC modifications.

6. (Optional) In the box at the far left of the dialog box, select the type of isotopes used as labels:

    *   Light: Refers to amino acid labels that use normal isotopes.

    *   Medium (3plex methods only): Refers to amino acid labels that use less massive isotopes, for example, Arg6 and Lys4.

    *   Heavy: Refers to amino acid labels that use heavy isotopes, for example, Arg10 and Lys8.

    The following example uses the **Heavy** label.

    "Heavy" now appears in the Channel Name box, and "Arg10" and "Lys6" appear in the Quantification Labels box, as shown in Figure 305.

7. In the Quantification Labels box, click the name of the channel that you want to change, in this case, **Arg10**.

    "Arg10" appears in the Label Name box.

Figure 305. Arg10 shown in the Quantification Method Editor
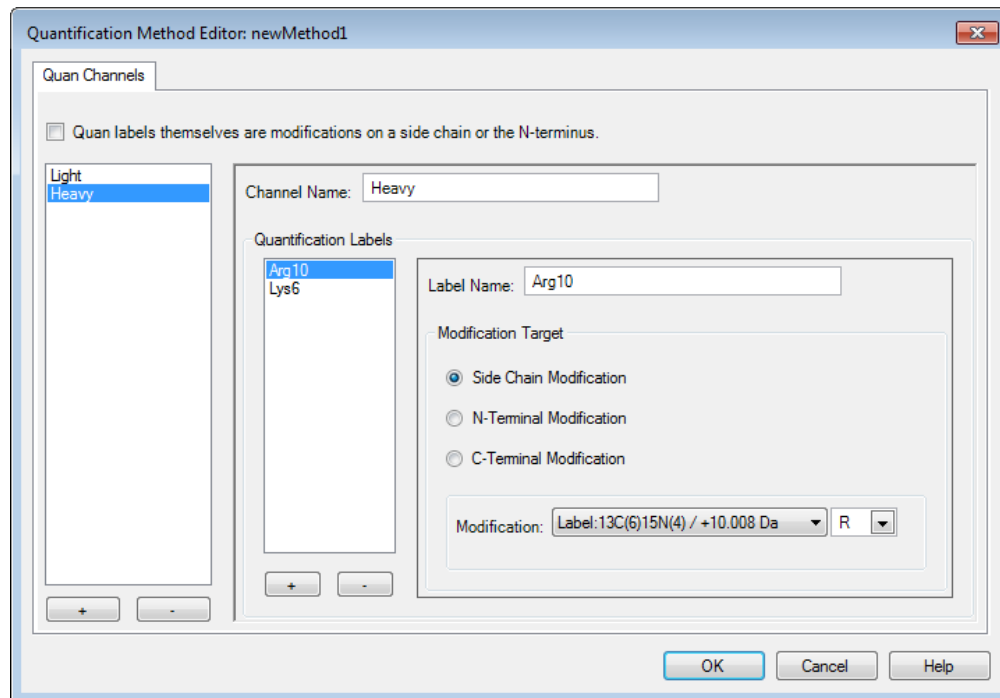
8. In the Label Name box, type the name of the new label—in this example, **Arg6**.

9. From the Modification list, choose the modification label: **13C(6)/+6.020 Da** in this example.

10. From the list to the right of the Modification box, choose the modification type, in this case, **R**.

Figure 306 shows the completed Quantification Method Editor.

Figure 306. Completed Quantification Method Editor
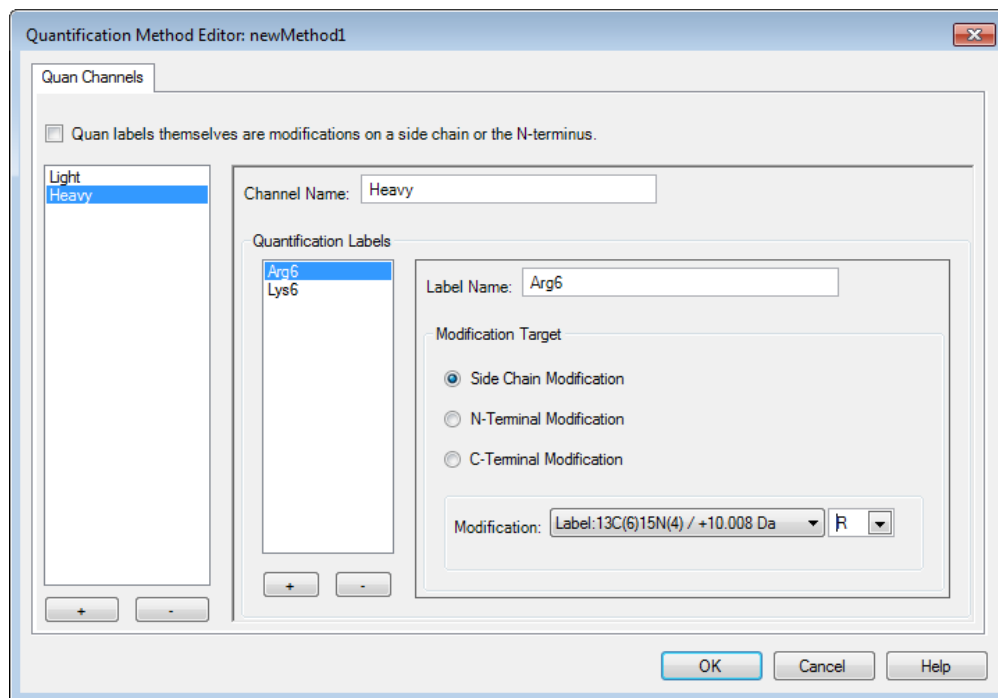
11. Click **OK**.

    The Save Quantification Method dialog box opens.

12. In the Save as New Method box, type the name of the quantification method that you
    want to create, for example, **SILAC 2plex (Arg6, Lys6)**. See Figure 307.

    Figure 307. Save Quantification Method dialog box
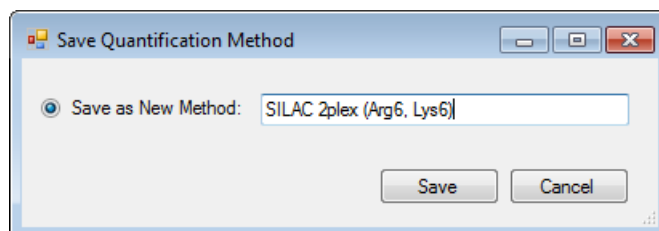
13. Click **Save**.

    The application adds the SILAC 2plex (Arg6, Lys6) method to the Quantification
    Methods view, as shown in Figure 308.

Figure 308. New custom SILAC 2plex method in the Quantification Methods view

New custom SILAC 2plex
quantification method

❖ **To add a new label rather than modifying an existing label**

If you want to add a new channel rather than modifying an existing channel template, follow these instructions when the Quantification Method Editor opens.

1. Click + beneath the list of labels in the box on the far left.

   The default name of New *number* now appears in the list of quantification channels in the box at the far left of the dialog box and in the Channel Name box.

   (To remove a quantification channel, select the quantification channel in the list of quantification channels and click – beneath the list.)

2. To specify a name for the new quantification channel, backspace over the default name in the Channel Name box and type the new name.

   The new name now appears in the quantification channel (far left) box.

3. To specify a quantification label to assign to the new quantification channel, click + beneath the Quantification Labels box.

   A default quantification label of New *number* now appears in the Quantification Labels box and the Label Name box.

   (To remove an existing quantification label, select the label in the Quantification Labels box and click – beneath the box.)

   Backspace over the label name in the Label Name box and type the new name.

4. In the Modification box, select the modification for the amino acid label.

5. In the adjacent box to the right of the Modification box, select the abbreviation of the amino acid selected in the Quantification Labels box that the modification should occur on.

6. Click **OK**.

Opening an Existing Study

Now you open an existing study or create a study. To create a study, see "Creating a Study" on page 425.

❖ **To open an existing study**

In the Recent Studies area of the Start Page, click the study name.

–or–

In the Recent Studies area of the Start Page, right-click the study name and choose **Open**.

–or–

Choose **File > Open Study** or click the **Open Existing Study** icon, , and browse to the study folder and then select and open a .pdStudy file.

–or–

1. In the Start area of the Start Page, click **Open Study**.

2. In the Open Study dialog box, browse to and select the name of the study, and click **Open**.

Creating a Study

> **Note** If you want to use a custom quantification method, create it before you create or open a study. For instructions, see "Selecting a Quantification Method for the Study" on page 427.

❖ **To create a study**

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

–or–

Choose **File > New Study/Analysis**.

–or–

Click the **Create New Study/Analysis** icon, .

The New Study and Analysis dialog box opens, as shown in Figure 309.

Figure 309. New Study and Analysis dialog box
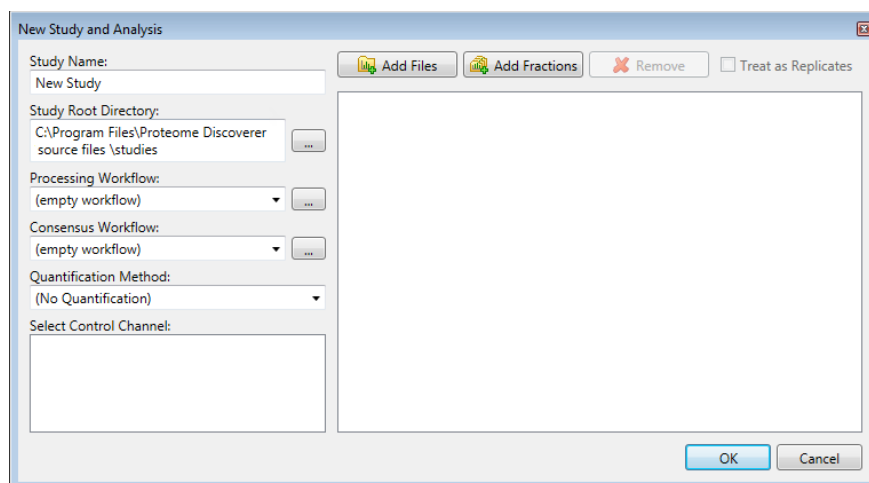
In this example, you only specify the name of the study and a root directory to save the study in.

2. In the Study Name box, specify the mandatory study name.

The application generates a default study name by searching for the common part of the file names when you add multiple files at once and using this common part as the suggestion for the name of the new study.

3. In the Study Root Directory box, specify the folder where you will store the study folder. Click the **Browse** button (...), and in the Select Folder dialog box, specify the folder and click **Select Folder**.

> **Note** When the data files, quan method, and workflows are already known, you can enter these into the New Study and Analysis dialog before clicking OK. This automatically adds the data files to the Input Files section. The quantification method is automatically applied to the data files, and a new Analysis with the processing and consensus workflows appears in the study. When the analysis requires study factors, you must enter these as described in the Adding the Study Factors section. Subsequently, you must associate these study factors with the different quan channels using either the Input Files or Samples tab in the study.

4. Click **OK**.

The application creates a new study folder in the folder that you specified as the root directory and opens a new page with the study name (Study: SILAC example in this case), as shown in Figure 310. It appends a .pdStudy extension to the study file name.

Figure 310. New Study: *Study_name* page
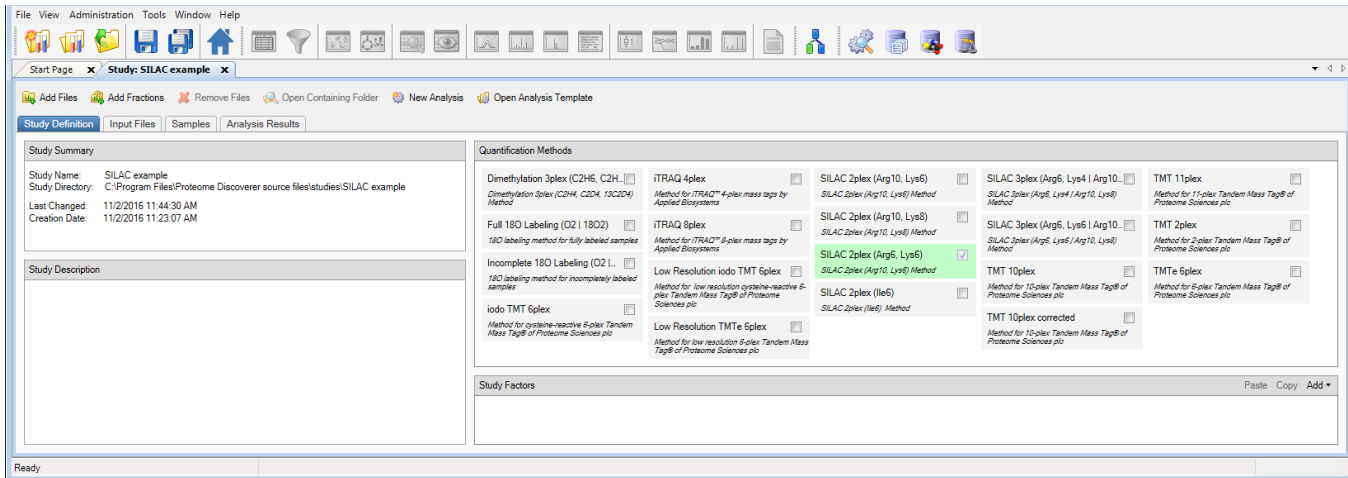
On the Study Definition page, add a description of your study, select the quantification method or methods to use in the study, and set up the new factors to use to describe and distinguish your samples.

Adding a Description

You can optionally add a description of the study by typing it in the Study Description area of the Study Definition page.

Selecting a Quantification Method for the Study

A quantification method contains the specification of the available quantification channels. You can specify a quantification method for each of the input files. You can either select an existing quantification method, or create a quantification method and then select it.

The example in this topic uses a custom SILAC 2plex method, but the following steps are applicable for any precursor ion quantification method.

❖ **To select the quantification method to use in the study**

1. Click the **Study Definition** tab in the study, if it is not already selected.

   The Quantification Methods area of the Study Definition page lists all the quantification methods that are currently available.

2. Select the check box corresponding to the quantification method or methods that you want to use.

   In this example, the samples are labeled with SILAC 2plex (Arg6, Lys6), so you would select the **SILAC 2plex (Arg6, Lys6)** check box, as shown in Figure 311.

   If the Quantification Methods pane does not include the SILAC 2plex (Arg6, Lys6) method, choose **File > Save All**, and close and reopen the study.

Figure 311. Quantification method selected

3. After you add the input files, specify the quantification method for each input file (see "Adding Input Files" on page 435). For instructions, see "Specifying the Quantification Method for Multiple Input Files" on page 436.

Adding the Study Factors

In this step, you add the study factors that you want to use for your samples.

A factor is a single biological or technical parameter that you control, for example, genotype, diet, environmental stimulus, age, column length, spray voltage, or collision energy.

In theory, you can track every parameter as a factor in your study, but normally you only track the parameters that actually differentiate samples from each other. For example, the gradient that you use for the chromatography is a parameter in your experiment. However, if you use the same gradient for all your samples, you would not add this as an explicit factor to your study.

❖ **To add numerical study factors**

1. Add the first numerical factor:

   a. In the Study Factors area of the Study Definition page, choose **Add > Numerical Factor**.

   The numerical factor dialog box shown in Figure 312 opens.

Figure 312. Numerical factor dialog box
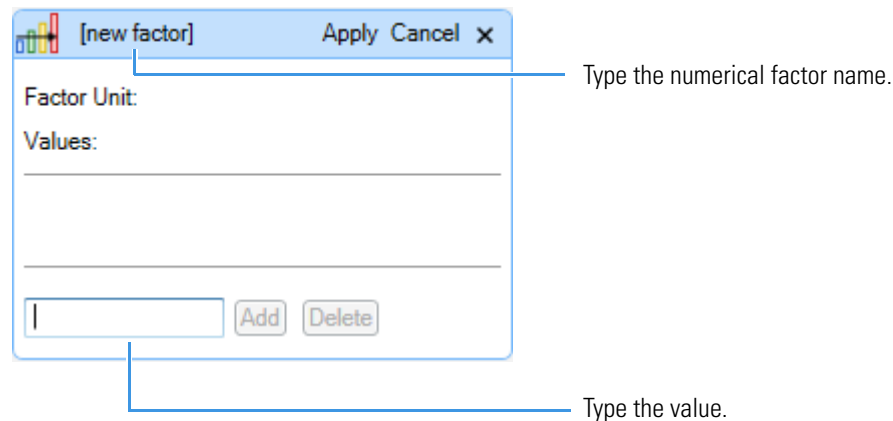
Type the numerical factor name.

Type the value.

[new factor] is highlighted.

b. Type a name over [new factor] for the new numerical factor, for example, **Time (days)**. See Figure 313.

> **Note** If the full numerical box becomes compressed, click **Edit** to restore it to its original size.

c. In the box to the left of the Add and Delete buttons, type the name of the first value, which is 5, and click **Add**. See Figure 313. Continue by adding values 7, 9, 11, 13, 15, 17, 25, 29, and 33.

Figure 313 shows the completed numerical factor dialog box.

Figure 313. Completed numerical factor dialog box
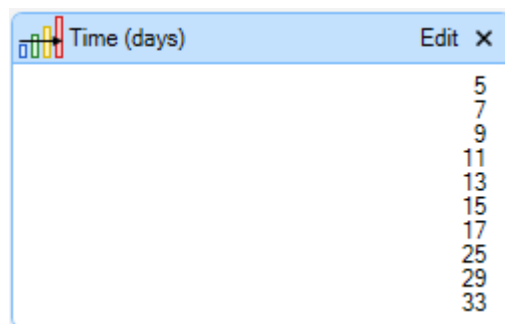
d. Click **Apply** in the numerical factor dialog box.

Confirm that the Study Definition page resembles Figure 314.

Figure 314. Numerical factor added to the study
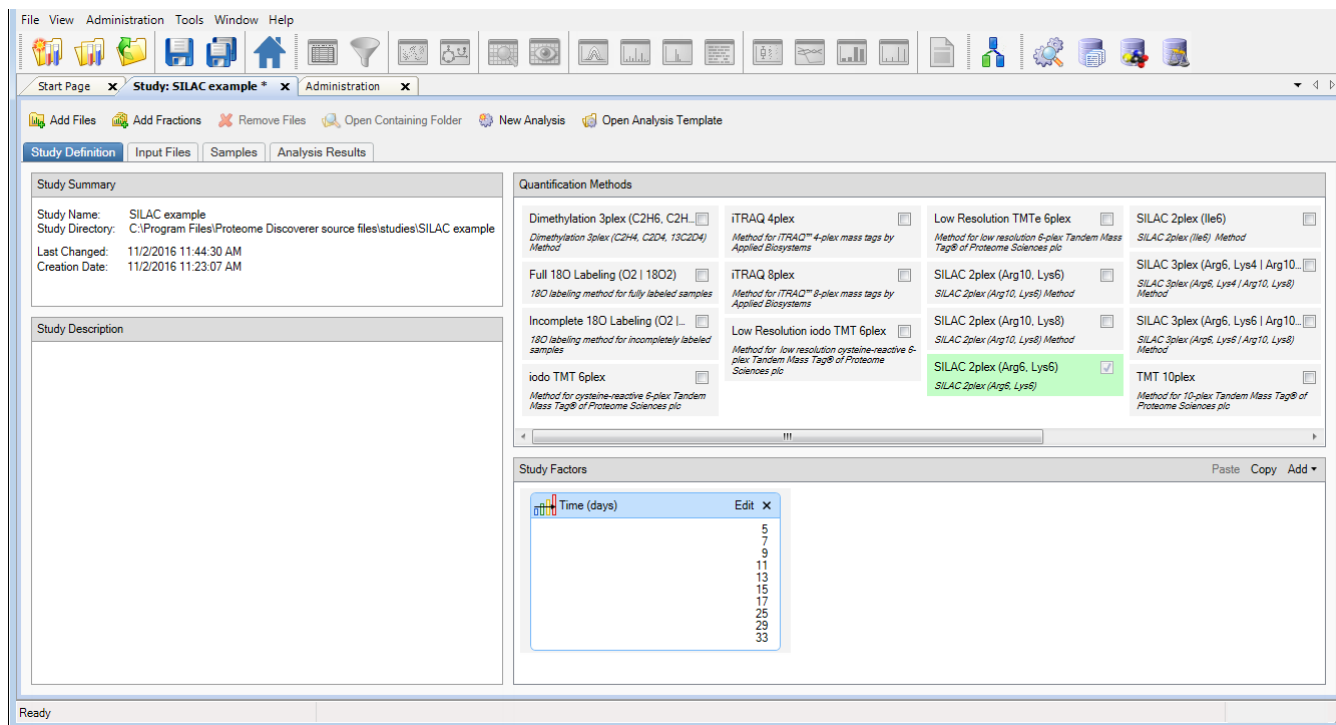
❖ **To add biological replicate study factors**

> **Note** For a description of the biological replicate study factor, see "Calculating P-Values for Replicate Data by Using Biological Replicate Study Factors" on page 513.

1. Add the first biological replicate factor:

   a. In the Study Factors area of the Study Definition page, choose **Add > Biological Replicate Factor**.

   The biological replicate factor dialog box shown in Figure 315 opens.

   Figure 315. Biological replicate factor dialog box
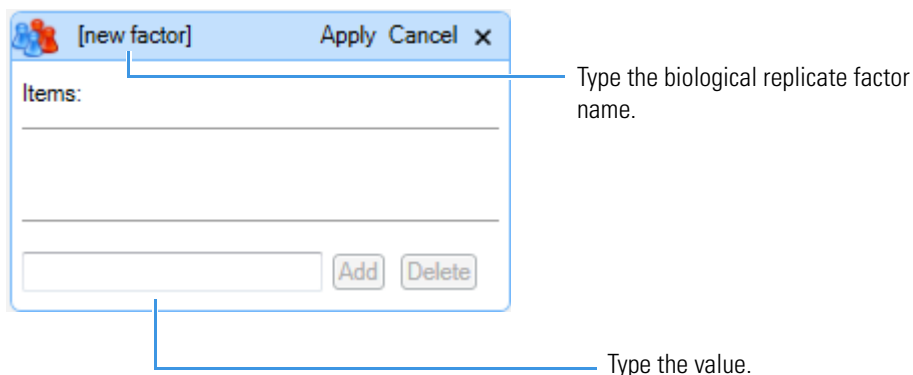
   Type the biological replicate factor name.

   Type the value.

   [new factor] is highlighted.

b. Type a name over [new factor] for the new numerical factor, for example, **Patient**. See Figure 316.

> **Note** If the full numerical box becomes compressed, click **Edit** to restore it to its original size.

c. In the box to the left of the Add and Delete buttons, type the name of the first value, for example, Patient1, and click **Add**. See Figure 316. Continue by adding values Patient2, Patient3, Patient4, and so forth.

Figure 316 shows the completed biological factor dialog box.

Figure 316. Completed biological replicate factor dialog box
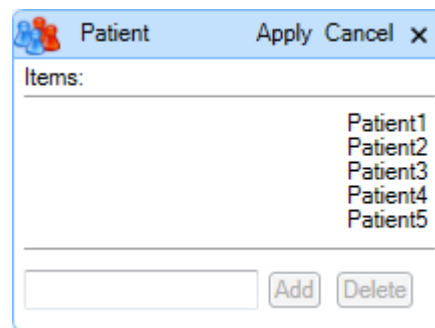
d. Click **Apply** in the numerical factor dialog box.

Confirm that the Study Definition page resembles Figure 317.

Figure 317. Biological replicate factor added to the study
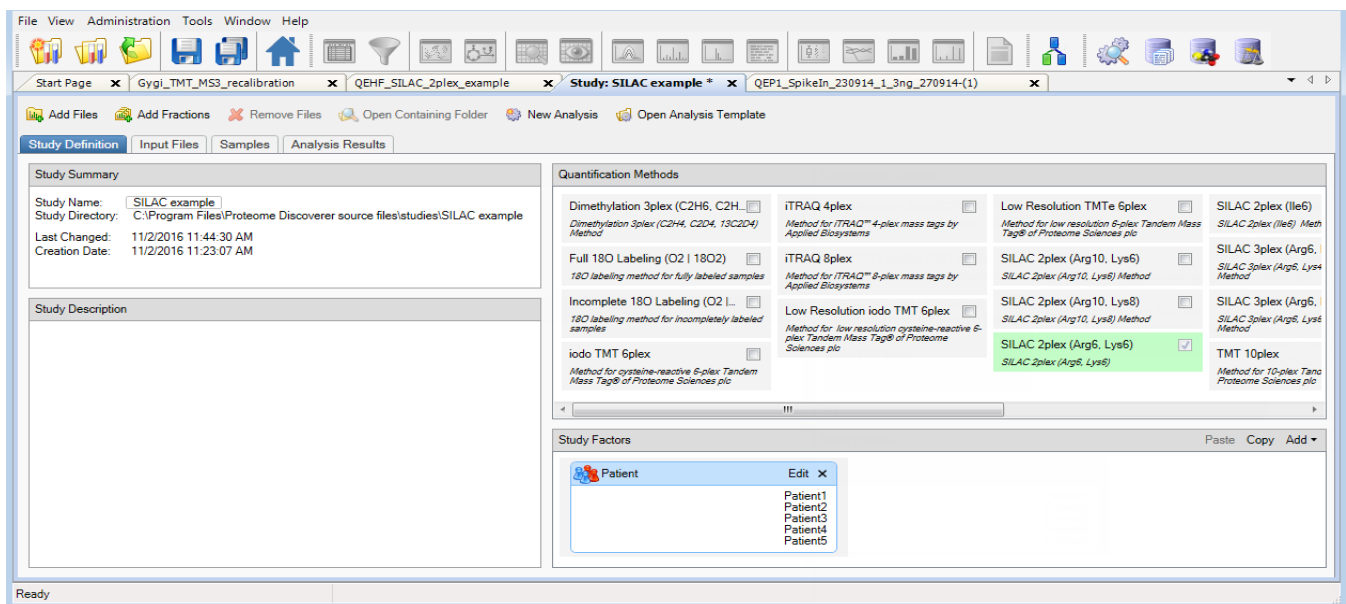
2. Add a second biological replicate study factor, if applicable.

❖ **To add categorical study factors**

1. Add the first categorical (non-numeric) factor (Acquisition in the example):

   a. In the Study Factors area of the Study Definition page, choose **Add > Categorical Factor**.

   The categorical factor dialog box shown in Figure 318 opens.

   Figure 318. Categorical factor dialog box
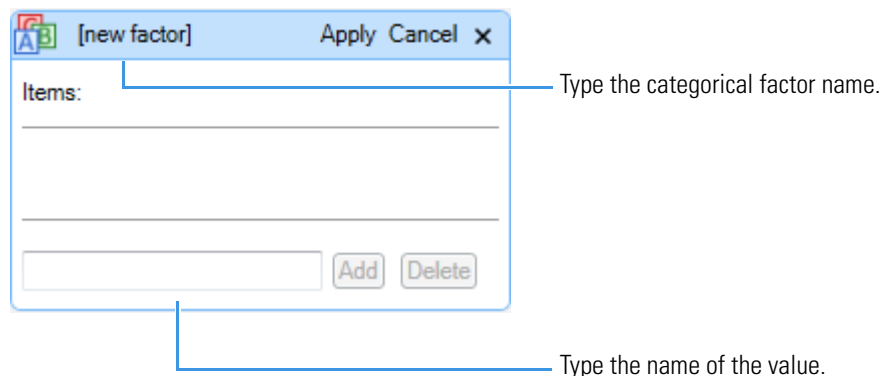
   Type the categorical factor name.

   Type the name of the value.

   [new factor] is highlighted.

   b. Type a name over [new factor] for the new categorical factor, for example, **Acquisition**. See Figure 319.

   > **Note** If the full categorical box becomes compressed, click **Edit** to restore it to its original size.

   c. In the box to the left of the Add and Delete buttons, type the name of the first value and click **Add**. See Figure 319.

   In this example, the value is the acquisition method, and the first acquisition method is intelligent data acquisition (IDA).

   d. In the same box, type the name of the second acquisition method and click **Add**.

   In this example, the second acquisition method is data-dependent acquisition (DDA). See Figure 319.

   Figure 319 shows the completed categorical factor dialog box.

Figure 319. Completed categorical factor dialog box
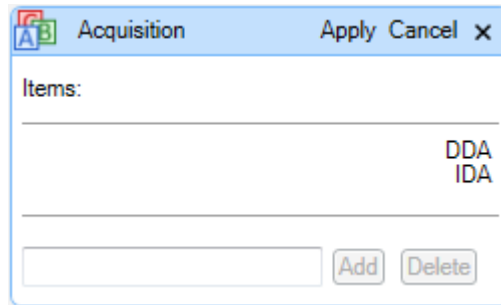
e. Click **Apply** in the categorical factor dialog box.

    Confirm that the Study Definition page resembles Figure 320.

Figure 320. First categorical factor added to the study
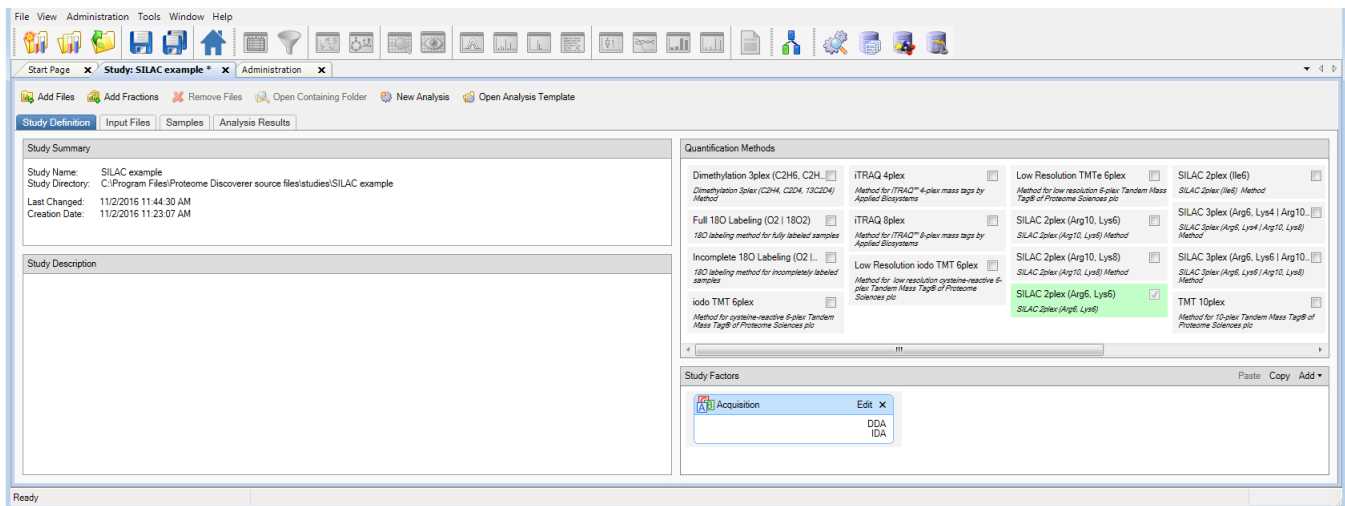
2. Add the second categorical factor (Tissue in the example):

    a. In the Study Factors area of the Study Definition page, choose **Add > Categorical Factor**.

        The dialog box shown in Figure 318 on page 432 opens.

        [new factor] is highlighted.

    b. Type a name over [new factor] for the new factor (see Figure 319 on page 433), for example, Tissue.

    c. In the box to the left of the Add and Delete buttons (see Figure 319 on page 433), type the name of the first type of tissue and click **Add**.

        In this example, the first type of tissue is Cerebellum.

d. In the same box, type the name of any additional types of tissue and click **Add** after each one.

The example adds the following types of tissue to the study:

- Cerebellum

- Cerebrum

- Heart

- Kidney

- Liver

- Lung

- Muscle

- Spleen

e. Click **Apply** in the Tissue dialog box.

Confirm that the Study Definition page resembles Figure 321.

Figure 321. Second categorical factor added to the study

Adding Input Files

Add the input files from the example data set to your study. (You can add input files to the study at any point.) For the types of input files supported, see "Inputs" on page 10. You can add individual input files, multiple unrelated input files, or fractions.

Adding a Single Input File or Multiple Unrelated Input Files

❖ **To add a single input file or multiple unrelated input files**

1. Click .

2. In the Add Files dialog box, browse to the location of the input files, and select them.

   Note You can also add existing MSF files to a study. The application adds the MSF file as a new result to the Analysis Results page and adds the raw data files that were processed to create the MSF files to the Input Files page.

3. Click **Open**.

   –or–

   Drag the input file or files from Windows Explorer and drop them onto the Input Files page.

   The input files appear on the Input Files page, as shown in Figure 322. Each file on the page receives a unique identifier: F1, F2, …, Fn. The Proteome Discoverer adds each file as a single study file.

   There is only one input file for the SILAC example study.

Figure 322. Input files on the Input Files page
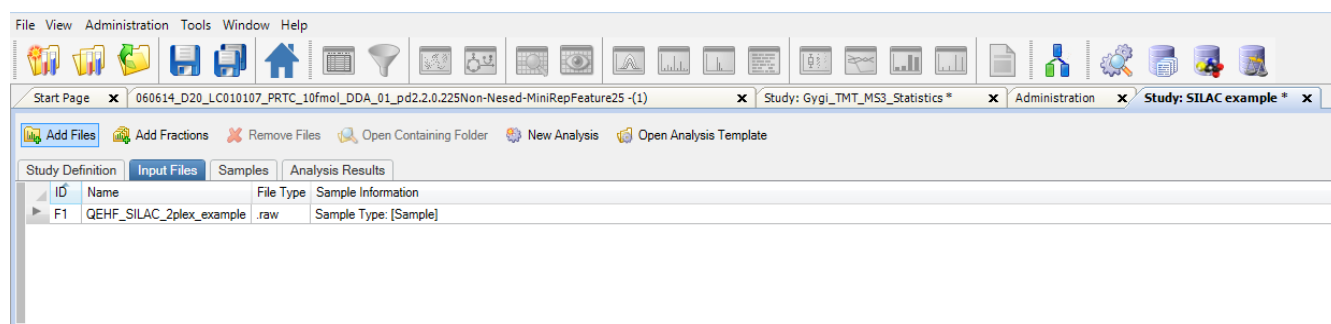
Adding Fractions

The SILAC example study does not use fractions. For information on adding fractions, see "Adding Fractions" on page 368.

Specifying the Quantification Method for Multiple Input Files

To specify the quantification method that was used for each of the files, see "Specifying the Quantification Method for Multiple Input Files" on page 373.

Setting the Study Factor Values for the Samples

When you select a quantification method for a file, the Proteome Discoverer generates a sample placeholder for each quantification channel.

The term "sample" refers to a distinct measurement of an analyte. At a minimum, a RAW data file contains at least a single sample, or in the case of sample multiplexing, it uses isobaric or metabolic labeling for multiple samples. If you measure the analyte from the same vial in two acquisitions—that is, if you measure two technical replicates—the application considers them two different samples. If you split the analyte, label each with a different isobaric tag, mix them, and acquire them into a single RAW data file, the application considers the two differently labeled parts as two different samples.

Each sample is associated with a sample type. Currently only quantification uses sample types. The sample type can be one of the following:

- Sample: A specimen from a larger biological entity

- Control: A sample typically used as a reference sample in a quantification experiment

- Blank: A sample consisting only of solvent and no sample mixture

- Standard: A sample consisting of a standard quality-control peptide mixture

You can create ad hoc relative quantitative ratios between any study factors or sample types, for example, blank/standard, control/sample, blank/control, and standard/blank. You can leave all of the channels as samples or make all of the channels any combination of sample, control, blank, or standard.

The Sample Type column on the Samples page of the study displays the sample type of each sample. The default sample type is Sample. If a file has samples for different quantification channels, mark one of the samples Control. In this example, the channel that was used to label the mouse kidney tissues is used as the control for generating the quantification ratios.

You can also use the control channels in scaling. For more information, see "Normalizing Peptide Groups and Protein Abundances" on page 535.

Each sample is associated with a quantification channel shown in the Quan Channel column and with values for each of the factors that you specified for your study. Previously, you specified a factor for the acquisition method used and a factor for the tissue that was extracted and labeled. You now set the correct factor values for each of the samples in the study.

Each sample has an automatically generated sample name composed of the raw data file name and the appended name of the quantification channel. You can change this name, but the name must be unique among all samples in the study.

❖ **To view the samples**

1. On the Input Files page, click the gray arrow to the left of a sample to display its constituent file entries, as shown in Figure 323.

   A hierarchical view opens, showing the samples contained in a raw data file. For each of the raw data files in the example, there are two samples for the two quantification channels of the SILAC 2plex. See Figure 323.

Figure 323. Samples in a raw data file
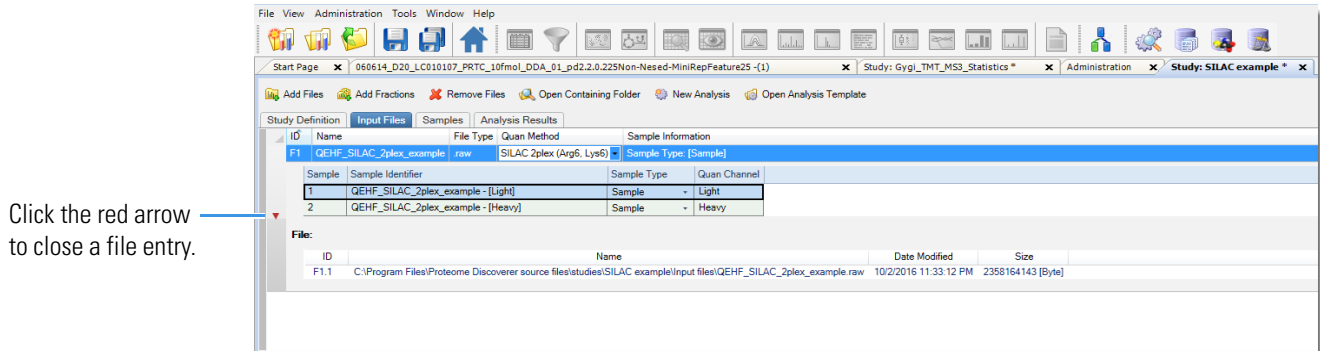
Click the red arrow to close a file entry.

❖ **To set the factor values for the samples**

1. Click the **Input Files** tab, if it is not already selected.

2. Click the gray arrow next to the first file to expand the information about the sample.

3. For the first factor (in the example, Acquisition), set the value for each sample in each raw data file by selecting the down arrow in the factor column and then selecting the value from the list.

4. For the second factor (in the example, Tissue), set the value for each sample in each raw data file by selecting the down arrow in the factor column and then selecting the value from the list.

5. Set the same values in the Acquisition and the Tissue columns for the rest of the samples.

   After you finish setting the factor values for each sample, the Input Files page resembles Figure 324. (In the example data set, you must set 128 factor values for ten files with ten samples each with two factors each.)

   Note For instructions on changing the values for factors or other study variables for multiple samples at once, see "Setting Values for Multiple Samples at the Same Time on the Input Files Page" on page 440.

Figure 324. Completed Input Files page

6. Click the red down arrow next to each sample to compress the information displayed.

The Sample Information column summarizes the information about the samples contained in a file, as shown in Figure 325.

Figure 325. Sample Information Column
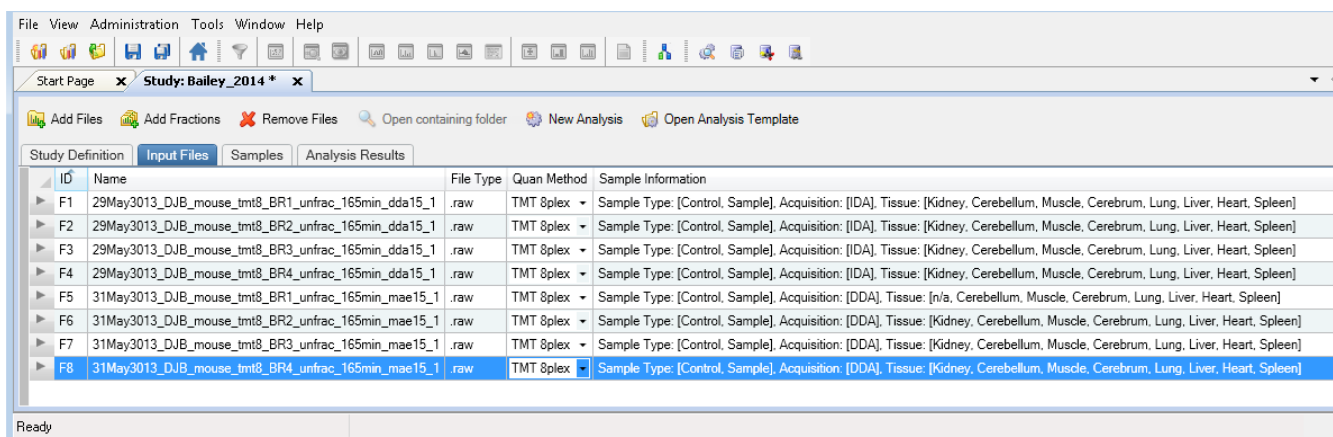
7. (Optional) Click the **Samples** tab.

Figure 326 shows the Samples page. It displays the same sample information as the Input files page.

Figure 326. Samples page

Setting Values for Multiple Samples at the Same Time on the Input Files Page

Highly multiplexed data are results obtained from processing several samples from one raw data file mixed and analyzed together in one LC/MS analysis, where isotopic and isobaric labels were used in quantification to distinguish the contribution of the individual samples. If you have highly multiplexed data, several files and samples with many study variables to set (sample type, quantification channel, study factors), or both, you can set values for study variables for multiple samples at once. You can use either the mouse or the keyboard to set these values on the Input Files page.

You can also use multicell editing on the Samples page to enter factor values for several samples. For example, if you have acquired data from several different tissues and the tissue is encoded in the file and sample names, you can use the filter option on the Samples page to filter all samples that contain "liver" and then set the Tissue factor to "liver" for all these samples.

❖ **To set values for multiple sample cells at the same time by using the mouse**

1. Click the **Input Files** tab.

2. To select multiple samples for one study variable, select the first cell and drag the cursor to select the remaining cells, or SHIFT-click multiple consecutive rows.

    Figure 327 shows multiple cells being selected in the Acquisition column.

Figure 327. Selecting multiple cells at one time
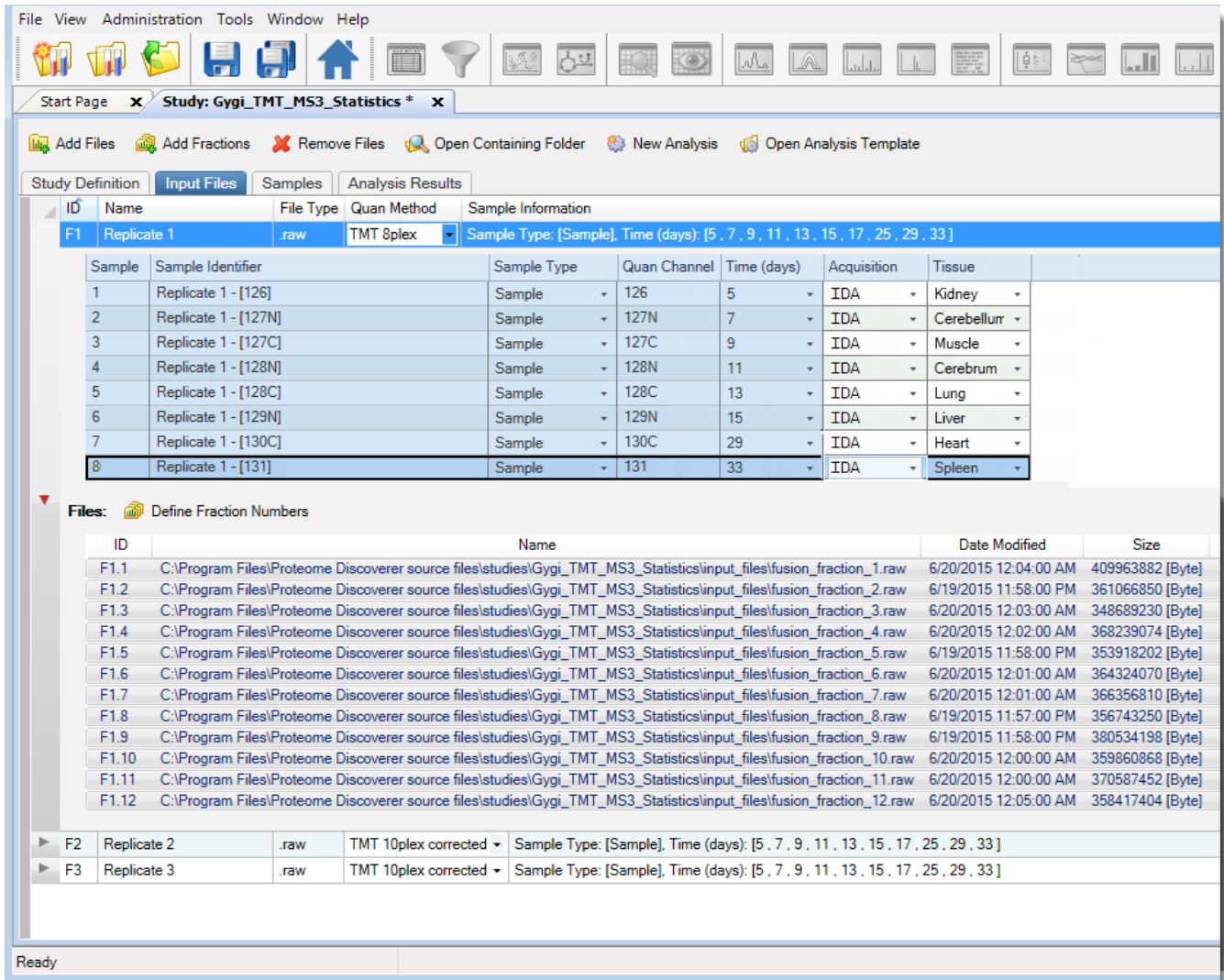
3. Right-click and select the study factor to be applied to the highlighted rows.

Figure 328. Selecting a new value to apply to all cells in multicell editing mode

In the example, the values in the cells of the Acquisition column change from IDA to DDA, as shown in Figure 329.

Figure 329. Multiple values changed at one time
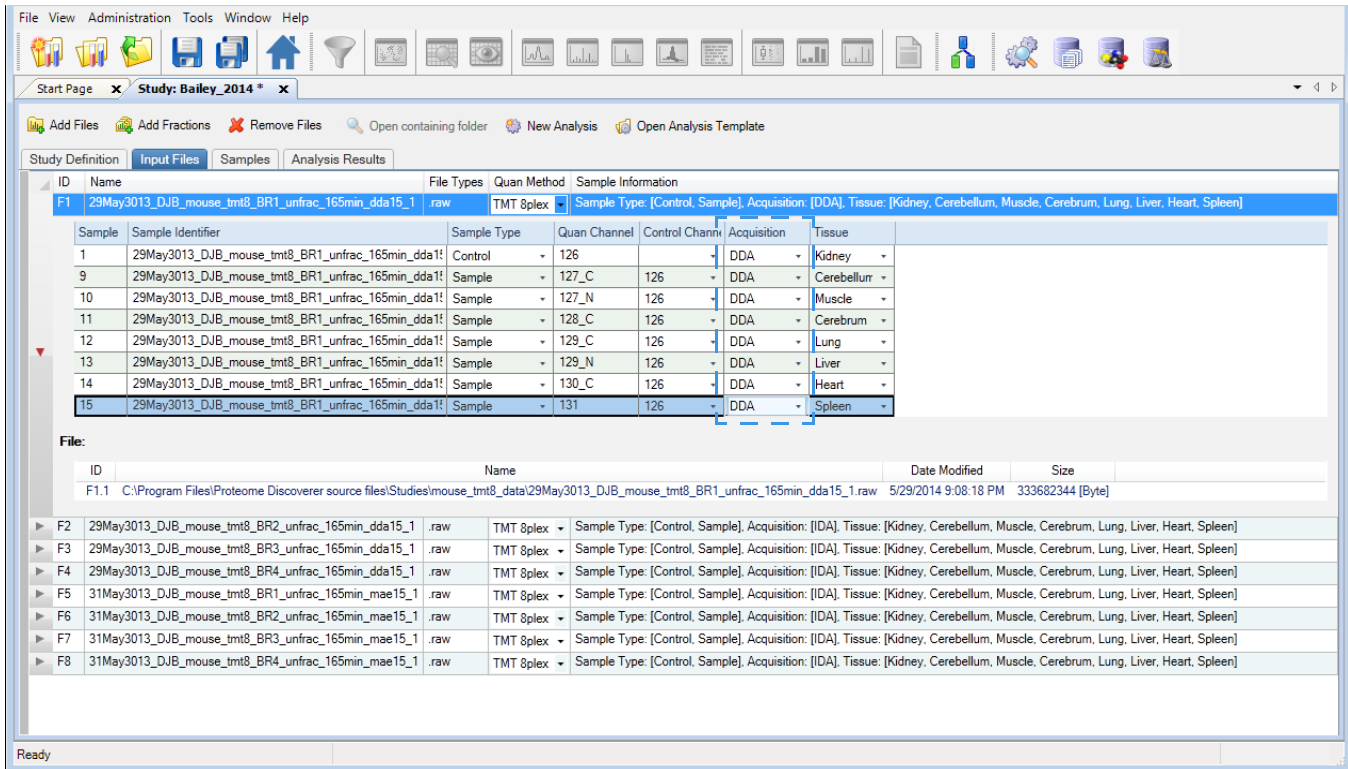
Setting Values for Multiple Samples at the Same Time on the Samples Page

Use the following procedure to change values for multiple samples on the Samples page.

❖ **To set values for multiple samples at the same time on the Samples page**

1. Click the **Samples** tab.

2. Click the **Contains** icon, [▣], if necessary, in the Acquisition column.

3. In the box next to the icon, type the name of the filter.

   For example, suppose you want to change the acquisition method for a number of samples from IDA to DDA. If you want to display only samples that contain "DDA," you would type **dda** in the filter box. Figure 330 shows the results.

Figure 330. Multicell editing of samples using filters

4. Select the first cell to change, and then drag the cursor to select the remaining cells that you want to change.

5. Right-click the highlighted rows and choose the study factor to apply to all the rows.

To return the samples to their unfiltered state, select the **Clear All Filters** icon, ![icon].

Saving a Study

You can save a study manually or automatically.

You can save a study manually at any time. A change in a study that requires you to save it is indicated with an asterisk (*) in the tab after the study name.

> **Note** Studies and analyses in the Proteome Discoverer are separate and must be saved separately. Saving a study does not save an analysis, and saving an analysis does not save a study.

You can set an option to have the application save studies automatically when you click
 Run . This option also saves the results generated in the study. It saves the analysis containing the workflow, but when you close the study and reopen it, you must access the analysis by clicking the Analysis Results tab, and then doing one of the following:

Click Reprocess ▼ , and choose **All Analysis Steps** (to open both the processing and consensus workflow) or **Last Consensus Step** (to open just the consensus workflow).

–or–

Right-click the name of the result file and choose **Show Details**.

❖ **To save a study manually**

Choose **File > Save**.

The Proteome Discoverer saves the study in the *study_name*.pdStudy file in the study directory.

❖ **To save a study automatically**

1. Before you click Run , select **Tools > Options**.

2. In the Options dialog box, select **Study Options** in the left pane.

3. Select the **Auto Save When Starting Analysis** check box.

4. Click **OK**.

   The Auto Save When Starting Analysis option does not take effect for any open studies. The application automatically saves only newly opened studies.

❖ **To save all open studies**

Choose **File > Save All**.

The Proteome Discoverer saves all studies open in the application.

Creating an Analysis

❖ **To create an analysis**

On the Study: *Study_name* page, click ⚙ New Analysis .

An Analysis window opens on the right side of the Study: *Study_name* page, as shown in Figure 331.

Figure 331. Analysis window

Analysis window containing a
Consensus Step box and a
Processing Step box

An Analysis window contains the following items:

- A Consensus Step box, which represents the consensus workflow step of the data processing

- A Processing Step box, which represents the processing workflow step of the data processing

- A Child Steps bar, which contains an Add button that you can use to add another Processing Step box. Multiple Processing Step boxes are useful when you want to process the same data in different ways—for example, by using different nodes or different node settings.

In addition, two new tabs appear on the Study: *Study_name* page: the Workflows tab and the Grouping & Quantification tab, as shown in Figure 331.

If you switch to an existing study that includes an Analysis window containing an existing consensus workflow, processing workflow, or both, and you want to open a new analysis, you must close the existing Analysis window and open a new Analysis window. To close the window, click **X** in the upper right corner.

Opening an Existing Analysis

You can open an existing analysis through a template or through a results file.

❖ **To open an existing analysis through a template**

1. In an open study, click Open Analysis Template .

2. In the Open Analysis Template dialog box, browse to and select the .pdAnalysis file that you want to open.

3. Click **Open**.

❖ **To open an existing analysis through a results file**

1. In an open study, click the **Analysis Results** tab in the study.

2. Select the appropriate .pdResult file.

3. Click Reprocess ▾ .

4. To open an Analysis window containing both the processing and consensus workflows, choose **All Analysis Steps** in the Reprocess list.

   –or–

   To open an Analysis window in order to execute just the consensus workflow, choose **Last Consensus Step** in the Reprocess list.

Adding Input Files to an Analysis

❖ **To add the input files to an analysis**

Select and drag the files from the Input Files page to the Input Files box of the Processing Step box in the analysis window.

The input files are listed in the Input Files area of the Processing Step box, as shown in Figure 332.

Figure 332. Input files in the Input Files area of the Processing Step box of the Analysis window

> **Note** You can remove a file from the Input Files area of the Processing Step box by clicking the X to the left of the file name.

Using Multiple Processing Steps in an Analysis

An analysis can contain more than one processing step. You can use additional processing steps when you want to process one set of input files differently from another set of input files. For example, you might want to run a Percolator validation separately on a set of files, such as the replicates of the control group and the replicates of the treatment group.

❖ **To add multiple processing steps**

In an open analysis window, click **Add** on the Child Steps bar.

A new processing step without a workflow appears.

❖ **To duplicate an existing processing step, including its workflow**

In an open analysis window, click **Clone** on the processing step bar.

Another window with a child step appears. You can now define a different processing workflow for the same file or a different file.

Adding or Deleting a Processing Step

You might want to perform only the processing step in a workflow. For example, you only need the processing step if you want to export peak lists as an MGF file with the Spectrum Exporter node. You can remove the Consensus step and re-add it later.

❖ **To perform only the processing step in an analysis**

In the title bar of the Consensus Step box, click the **X** in the upper right corner to remove the consensus step from the analysis.

The Consensus Step box closes, and the Add icon appears in the Analysis title bar.

❖ **To re-add a deleted consensus step in an analysis**

In the Analysis title bar, click 🟢 Add .

The Consensus Step box reopens, and the Add icon disappears.

Creating a Processing Workflow for Precursor Ion Quantification

To create a precursor ion quantification method, you must set up a processing workflow that includes the Minora Feature Detector node.

Default processing workflows are available for several instruments. You can use these default workflows as is or modify them to suit your needs. To access them, click 🔷 Open Common on the Workflows page, or see "Using Common Workflow Templates" on page 102. When you use a default workflow, you do not need to perform step 1 through step 7 in the following procedure because these steps are already included in the workflow. However, you must check the node parameters in the workflow to make sure they are appropriate for your data and add modifications to the search for the quantification to be successful.

The default processing workflows for precursor ion quantification are the following:

- ProcessingWF_Fusion\PWF_Fusion_Precursor_Quan_and_LFQ_IT HCD_SequestHT_Percolator.pdProcessingWF

- ProcessingWF_LTQ_Orbitrap\PWF_OT_Precursor_Quan_and_LFQ_CID_SequestHT_ Percolator.pdProcessingWF

- ProcessingWF_QExactive\PWF_QE_Precursor_Quan_and_LFQ_SequestHT_ Percolator.pdProcessingWF

❖ **To create a processing workflow for precursor ion quantification**

> **Note** This procedure uses a SILAC 2plex example.

1. In the Workflow Editor, follow the general instructions for creating a processing workflow. See "Creating a Processing Workflow" on page 68.

   The processing workflow must include the following nodes as a minimum:

   - Spectrum Files node

   - Spectrum Selector node

   - A search engine node (for example, Mascot or Sequest HT)

- Minora Feature Detector node

- Fixed Value PSM Validator node, Target Decoy PSM Validator node, or Percolator node

2. In the Processing Workflow window, drag the **Spectrum Files RC** node to the workspace.

3. Drag the **Spectrum Selector** node to the workspace.

4. Connect the Spectrum Files node to the Spectrum Selector node.

5. Drag the **Minora Feature Detector** node to the Workflow Tree pane and attach it directly to the Spectrum Files node.

   The Minora Feature Detector node performs quantification for isotopically labeled amino acids.

6. Drag the appropriate search engine node—for example, **Sequest HT**—to the Workflow Tree pane and attach it to the Spectrum Selector node.

7. Drag the **Fixed Value PSM Validator**, **Target Decoy PSM Validator**, or the **Percolator** node to the Workflow Tree pane and attach it to the search engine node.

   Figure 333 illustrates the basic processing workflow for precursor ion quantification.

   Figure 333. Basic processing workflow for precursor ion quantification
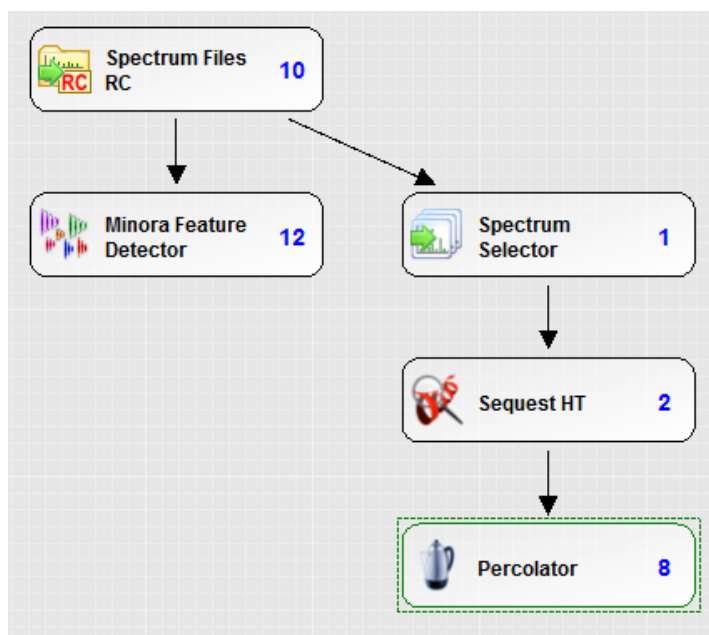
8. (Optional) Add any other appropriate nodes and connect them to the workflow.

9. In the Parameters pane of the Workflow Editor, click **Show Advanced Parameters**.

10. Click the **Spectrum Files RC** node and set the parameters for it in the parameters pane:

    a. In the Protein Database box, select the FASTA database.

b. In the Static Modifications area, select the static modifications. For example, select **Carbamidomethyl/+57.021 Da (C)** in the Static Modification box.

c. Set any other parameters, including precursor and product ion mass tolerance.

For information on the available search engine parameters, refer to the Help.

10. Click the **Spectrum Selector** node, and set the parameters for it in the Parameters pane:

a. Change the setting in the Max. Precursor Mass box to an appropriate setting. For example, for SILAC 2plex (Arg6, Lys6) quantification, set this option to **6500**.

b. Change the setting in the S/N Threshold box to an appropriate setting. For example, for SILAC 2plex (Arg6, Lys6) quantification, use the default of **1.5**.

11. Click the search engine node—for example, **Sequest HT**—and set its parameters in the Parameters pane:

a. In the Protein Database box, select the **FASTA** database.

b. In the Dynamic Modifications area, select the dynamic modifications.

> **Note** If you are using the Mascot search engine node, you can use the Quan Modifications parameter rather than the Dynamic Modifications parameters to specify the modifications to search for. For instructions on using this method, see "Checking the Quantification Method" on page 554.

For example, for SILAC 2plex (Arg6, Lys6) quantification, you might select the following dynamic modification:

- **13C(6)/ +6.020 Da (K, R)**

If you do not find these labels, you can enable them by following the instructions in "Defining Chemical Modifications" on page 200.

c. In the Static Modifications area, select the static modifications. For example, for SILAC 2plex (Arg6, Lys6) quantification, select **Carbamidomethyl/ +57.021 Da (C)** in the Static Modification box.

d. Set any other parameters that you prefer. For information on the search engine parameters available, refer to the Help.

12. Click the **Fixed Value PSM Validator** node, the **Percolator** node, or the **Target Decoy PSM Validator** node, and set its Maximum Delta Cn parameter.

13. Set the parameters for all other nodes in the Parameters pane.

For information about all the parameters that you can set for each node, refer to the Help.

14. (Optional) Save the workflow as follows:

a. In the Name box above the Workflow Tree pane, type a name for the processing workflow.

b. (Optional) In the Description box above the Workflow Tree pane, type a brief description of the processing workflow.

c. In the Workflow Editor, click ⊞ Save .

d. In the Save Workflow dialog box, do the following:

  i. Browse to the file to save the template in, or type a file name in the File Name box.

  ii. In the Save As Type box, select **Processing Workflow File (*.pdProcessingWF)**.

  iii. Click **Save**.

The application saves the workflow with a .pdProcessingWF file name extension.

Creating a Consensus Workflow for Precursor Ion Quantification

To use a precursor ion quantification method, you must set up a consensus workflow that includes the Feature Mapper and the Precursor Ions Quantifier node.

Default consensus workflows are available. You can use these default workflows as is or modify them to suit your needs. To access them, click 🗂 Open Common on the Workflows page, or see "Using Common Workflow Templates" on page 102. When you use a default workflow, you do not need to perform step 1 through step 9 in the following procedure because these steps are already included in the workflow. However, you must check the node parameters in the workflow to make sure they are appropriate for your data and add modifications to the search for the quantification to be successful.

The consensus workflow must be selected in the Workflow Tree window for the Open Common command to display the processing workflows available.

The default consensus workflow for precursor ion quantification is ConsensusWF\CWF_Comprehensive_Enhanced Annotation_Quan.

❖ **To create a consensus workflow for precursor ion quantification**

1. Follow the general instructions for creating a consensus workflow with the Workflow Editor. See "Incorporating an Existing Workflow into a Study and an Analysis" on page 99.

The consensus workflow must include the following nodes as a minimum:

  • MSF Files node

  • PSM Grouper node

  • Peptide Validator node

  • Peptide and Protein Filter node

  • Protein Scorer node

- Protein Grouping node

- Feature Mapper node

- Precursor Ions Quantifier node

2. In the Consensus Workflow window, drag the **MSF Files** node to the workspace.

3. Drag the **PSM Grouper** node to the workspace, and connect it to the MSF Files node.

4. Drag the **Peptide Validator** node to the workspace, and connect it to the PSM Grouper node.

5. Drag the **Peptide And Protein Filter** node to the workspace, and connect it to the Peptide Validator node.

6. Drag the **Protein Scorer** node to the workspace, and connect it to the Peptide And Protein Filter node.

7. Drag the **Protein Grouping** node to the workspace, and connect it to the Protein Scorer node.

8. Drag the **Feature Mapper** node to the workspace, and connect it to the MSF Files node.

9. Drag the **Precursor Ions Quantifier** node to the workspace, and connect it to the Feature Mapper node.

10. (Optional) Drag the **Data Distributions** node to the Post-Processing Nodes pane.

   This node adds heat maps, which aid in data review.

   Figure 334 illustrates the basic consensus workflow for precursor ion quantification.

Figure 334. Basic consensus workflow for precursor ion quantification

11. (Optional) Add any other appropriate nodes and connect all the nodes together.

12. Set the parameters for each node.

   For information about all the parameters that you set for each node, refer to the Help.

13. (Optional) Save the workflow:

   a. In the Name box above the Workflow Tree pane, type a name for the consensus workflow.

   b. (Optional) In the Description box above the Workflow Tree pane, type a brief description of the consensus workflow.

   c. In the Workflow Editor, click   Save .

   d. In the Save Workflow dialog box, do the following:

      i.   Browse to the file to save the template in, or type a file name in the File Name box.

      ii. In the Save As Type box, select **ConsensusWorkflow File (*.pdConsensusWF)**.

      iii. Click **Save**.

   The application saves the workflow in the *file_name*.pdConsensusWF file.

14. Save the analysis. See "Saving an Analysis" on page 83.

15. Save the study. See "Saving a Study" on page 56.

16. Click 🔛 Run , in the upper right corner of the Analysis window.

   The job queue appears, as shown in Figure 53 on page 83, displaying the status of your search. Use the job queue to check the status of your search as the search progresses. For information about the job queue, refer to the Help.

Saving an Analysis

To use an analysis as a template for later reuse, you can save it as a .pdAnalysis template file.

> **Note** Studies and analyses in the Proteome Discoverer are separate, so you must save them separately. Saving a study does not save an analysis, and saving an analysis does not save a study.

❖ **To save an analysis as a template for later reuse**

1. In the upper right corner of the Analysis window, click 💾 Save .

2. In the Save Analysis Template dialog box, browse to the location where you want to store the template.

3. In the File Name box, browse to the study folder, and type or select the template file name.

4. In the Save as Type box, select **Analysis Templates (*.pdAnalysis)**.

5. Click **Save**.

   The application saves the analysis in a file with a .pdAnalysis extension.

   The .pdAnalysis template file saves the processing and consensus workflows. It saves neither the input files nor the study variables that were selected to group the samples and quantification ratios.

Specifying Quantification Ratios from Selected Sample Groups for Precursor Ion Quantification

After you set up the workflow to use for the analysis, you can specify the ratios to report for the quantification and how to group your samples in relation to the specified factor values.

You can generate arbitrary ratios from selected sample group abundances without being restricted to predefined sample and control specifications. Previous versions of the Proteome Discoverer application automatically generated quantification ratios that relied on the specification of sample or control sample types. This feature is appropriate for some experiments but too restrictive for experiments in which the same sample should occur in the denominator for some ratios and in the numerator for others.

For example, in a typical mutant-versus-wild-type experiment, you usually would monitor ratios like this to study the effect of the treatment on a mutant:

$$\frac{wt\_treat}{wt}, \frac{m\_treat}{m}, \frac{m\_treat}{wt\_treat}$$

where:

- *wt_treat* is a treated wild type.

- *wt* is a wild type.

- *m_treat* is a treated mutant.

- *m* is a mutant.

Also, in some experiments there is no clear control available, and all pair-wise comparisons of available samples are of interest.

On the Grouping & Quantification page of a study, you can generate custom quantification ratios that do not restrict you to predefined sample and control specifications.

In the Grouping & Quantification view, you first select the study variables that were used to group your samples and select the numerators and denominators of your ratios. Then, you manually specify the quantification ratios to generate or semiautomatically create all possible quantification ratios against an ad-hoc selection of denominator values.

❖ **To open the Grouping & Quantification page**

On the Study: *Study_name* page, click the **Grouping & Quantification** tab.

> **Note** The Grouping & Quantification tab does not appear until you add input files that are associated with a single quantification method.

Figure 335 shows the Grouping & Quantification page of the study.

Figure 335. Grouping and Quantification page
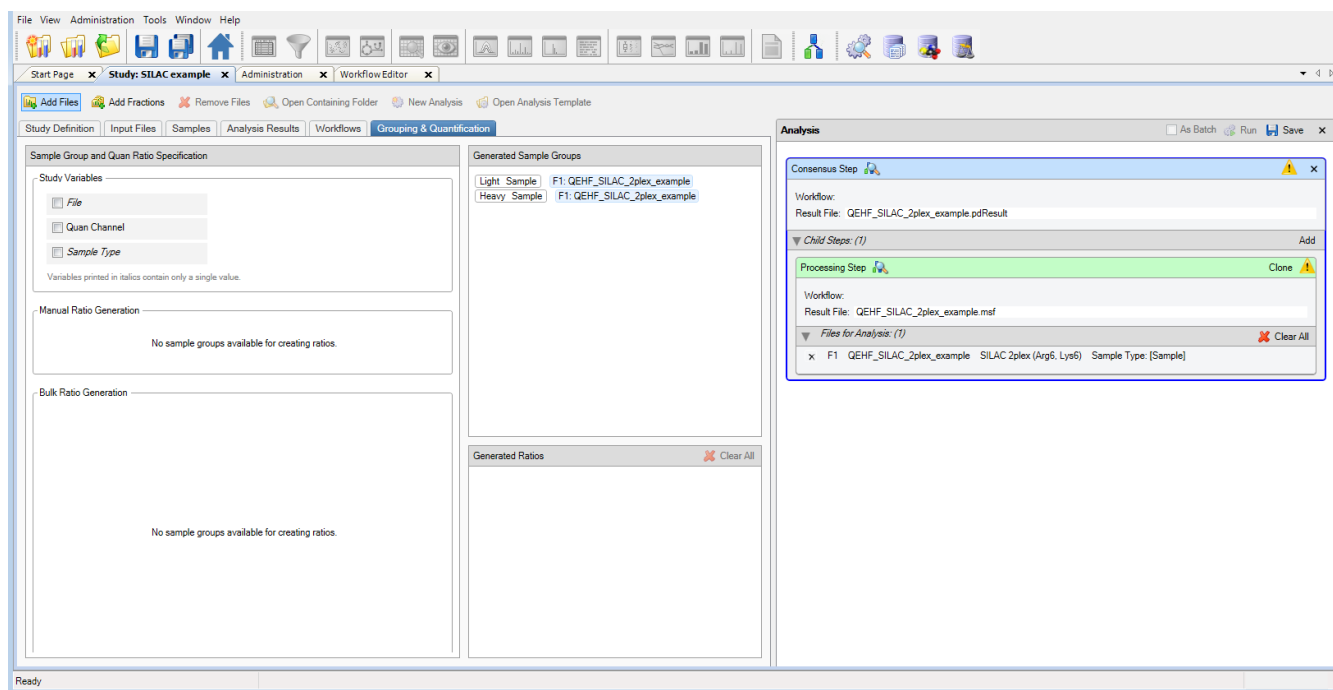
Use this study factor if you want to create ratios from the different forms of the precursor label.

The page contains the following areas:

- Study Variables area: At the top left is a list of the selectable study factors, or study variables, that you specified in the Study Factors area of the Study Definition page. A study factor is anything that captures the difference between two samples, for example, drug treatment, the time of drug application, or differences in tissue, organisms, or patients. It can be differences in sample preparation, chromatography settings, or acquisition parameters; or differences in the isobaric or metabolic labels used. Study factors can vary from sample to sample. In the example used throughout this topic, the study factors are the quantification channels associated with each sample and the factors for the acquisition method and tissue. Selecting the study factors to use for grouping provides an effective means of sorting the samples and quantification ratios into replicate and treatment groups.

    - Files: Groups samples and quantification ratios by files.

    - Quan Channels: Groups samples and quantification ratios by quantification channels.

    - *Study_factors*: Groups samples and quantification ratios by user-defined study factors.

    - Sample Types: Groups factors that include Sample, Control, Standard, and Blank. Variables displayed in italics contain only a single value.

Study factors are user-defined, but the File, Quan Channel, and Sample Type variables always appear on the Grouping & Quantification page for all studies.

- Manual Ratio Generation area: Contains menus where you can select the numerator and denominator for each quantification ratio.

- Bulk Ratio Generation area: Displays the study factor values to use as the denominators for semiautomatically generated quantification ratios.

- Generated Sample Groups area: Displays samples grouped by the values set for the selected study factors. The application ignores any other differences among the samples that might be present.

Selecting these variables indicates that you want to group your samples and quantification ratios according to the specified study variables and ignore any other difference between the samples that might be present. For example, if you group by tissue, you put into the same sample group all samples that share the same tissue factor. You also group all quantification ratios together that have the same tissue in the numerator or denominator. However, you ignore any differences in the acquisition method used.

The grouping of samples affects how identifications are displayed in the distribution maps (see "Sample Information Used to Display Identifications and Quantifications Among Files and Samples" on page 578) and the calculation of areas from precursor ion quantification. For the latter, the application reports separate area values for each sample or sample group.

When you group ratios, the application calculates an averaged ratio for a group of ratios that are similar with respect to the selected study variable in the numerator and denominator.

> **Note** The application first groups samples to average abundance values from replicate samples and then calculates ratios from the grouped abundances. The ratios are freely defined in the Grouping and Quantification view, without being restricted to predefined control samples.

If the application does not use all available sample group values in at least one ratio, it displays a warning in the Generated Sample Groups area, like that shown in Figure 336, and marks the unused sample groups.

Figure 336. Warning in the Generated Sample Groups area
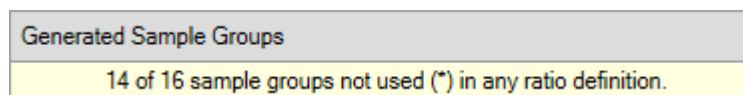
Generated Sample Groups

14 of 16 sample groups not used (*) in any ratio definition.

- Generated Ratios area: Displays the ratios generated.

❖ **To generate custom quantification ratios semiautomatically**

1. In the Study Variables area, select the check box of the study factors, or variables, that you want to use to group your samples and from which you want to draw the numerators and denominators of the ratios. For the example used throughout this topic, select the **Tissue**

check box to indicate that the samples and quantification ratios are grouped by the values set for the tissue factor.

After you select the study factors, the Generated Sample Groups area displays the generated sample groups. When performing the quantification, the application calculates abundance values for each sample and averages the abundance values of all samples in a sample group.

This step creates sample groups using the selectable study factors available for the files to be analyzed. The application places all samples that have the same combination of study factor values in the same group. In the example shown in Figure 340 on page 461, the application groups the samples by quantification channel.

The order of study factors is relevant for the semiautomatic generation of ratios.

2. (Optional) To change the placement of a study factor in the list of study factors, do the following:

   a. In the Study Variables area, select the check box for a study factor.

   A placement handle in the form of a green rectangle appears to the left of the selected check box, as shown in Figure 337.

   Figure 337. Placement handle in the Study Variables area

   b. Hold the cursor over the placement handle.

   White up and down arrows now appear on the handle.

   c. Drag the cursor up or down to move the variable to its new place in the list of variables, or click the up or down arrows to move the study factor.

3. (Optional) To sort study factors in the Bulk Ratio Generation area and sample groups in the Generated Sample Groups area, click one of the following to the right of each study factor in the Study Variable area.

   • To sort these items in descending order, click the **Sort Descending** icon, .

   • To sort these items in ascending order, click the **Sort Ascending icon**, .

   • To leave these items unsorted, click the **No Sorting** icon, .

4. In the Bulk Ratio Generation area, select the check box for the type of tissue to use in the denominator of the ratio. For the example, select **Quan Channel: Light**.

The Bulk Ratio Generation area displays a list of the denominator values for this type of study factor. If you select only one study factor, it displays a list of the available denominator values for this factor, as shown in Figure 338.

Figure 338. Denominator values available for a single study factor
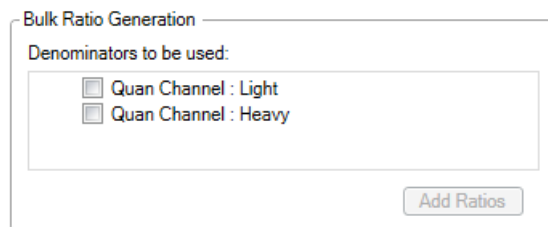
If you select multiple study factors, the Bulk Ratio Generation area displays the denominator values available for each factor.

5. (Optional) To select the same study factor for all the denominators, do the following:

   a. Hold the cursor over a denominator value.

      An icon containing four check boxes in a square appears on the left side of that item, as shown in Figure 339.

   b. Click the icon.

   The application selects the same study factor for all denominators (see Figure 339).

Figure 339. Icon for multiple selection of denominators in the Bulk Ratio Generation area
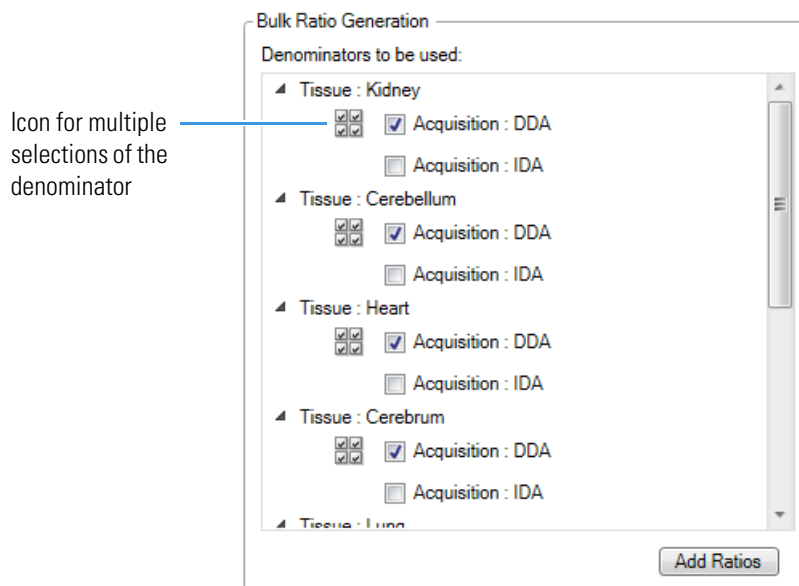
6. Click **Add Ratios**.

The application generates all possible ratios against the selected denominator values and adds them to the Generated Ratios area. Figure 340 shows the generated quantification ratios and ratio groups in the Generate Ratios area after the selection of Tissue as the study variable to group by and Kidney as the denominator to use.

Figure 340. Quantification ratios and ratio groups generated by grouping the samples and ratios by quantification channel
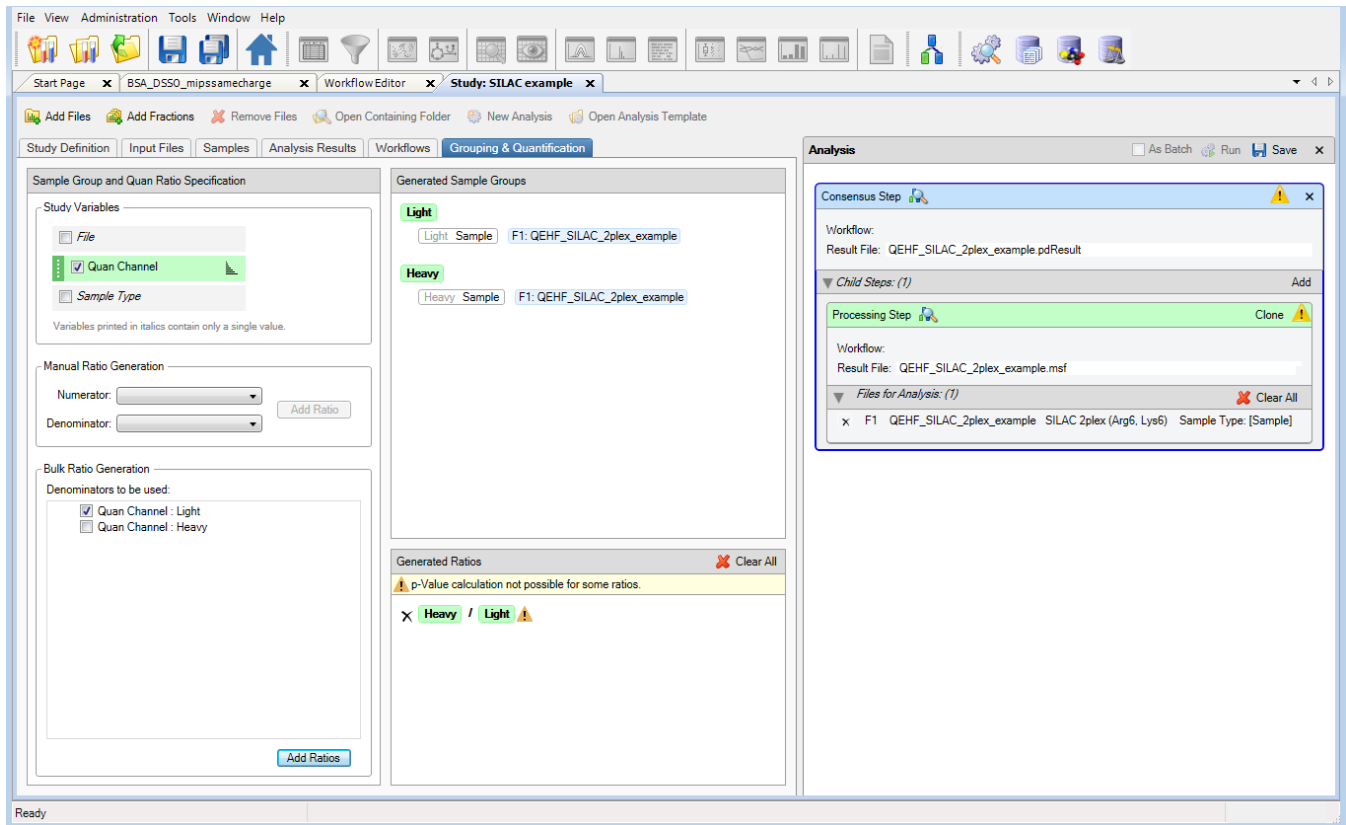
❖   **To generate custom quantification ratios manually**

1. On the Study: *Study_name* page, click the **Grouping & Quantification** tab.

   > **Note** The Grouping & Quantification tab does not appear until you add or open an analysis and specify Input Files for the analysis.

2. In the Study Variables area, select the check box of the study factor or factors that you want to use to group your samples and from which you want to draw the numerators and denominators of the ratios. For the example used throughout this chapter, select the **Tissue** check box to indicate that the samples and quantification ratios are grouped by the values set for the tissue factor.

3. In the Manual Ratio Generation area, select the numerator value from the Numerator list.

4. Select the denominator value from the Denominator list.

5. Click **Add Ratio**.

The generated ratio appears in the Generated Ratios area, as shown in Figure 341.

Figure 341. Ratio manually generated for one study factor

❖ **To save the settings on the Grouping & Quantification page**

> **Note** The application does not save the settings on the Grouping & Quantification page with a study or with an analysis. Instead, it associates the settings with search results, so you must load them from data sets that have already been processed within the study or recreate them from the beginning.

1. Click the **Analysis Results** tab of the study.

2. Select the result on the Analysis Results page, and choose either **Reprocess > All Analysis Steps** or **Reprocess > Last Consensus Step**.

   If you choose Use Results to Make New (Multi) Consensus, the Grouping and Quantification page no longer displays ratios.

Performing the Quantification

❖ **To perform the quantification**

Click ⚙ Run .

The job queue appears, as shown in Figure 49 on page 78 for the processing workflow and in Figure 53 on page 83 for the consensus workflow, displaying the status of your search. Use the job queue to check the status of your search as the search progresses. For information about the job queue, refer to the Help.

The application creates one or more MSF and .pdResult files, depending on whether the data files to be analyzed are batch-processed. It might reuse MSF files to produce a new .pdResult file if a different report for the given data is required.

The .pdResult file has the same name as the input file but with a .pdResult extension. The Analysis Results page displays the name of the .pdResult file.

Working with the Search Results

To open, delete, convert, and reprocess search results and to view the workflow and the analysis after the search, see "Working with the Search Results" on page 89.

Calculating and Validating Raw Quantification Values in Precursor Ion Quantification

The Precursor Ions Quantifier node first validates the LC/MS features. It determines whether they contain valid values or consist only of a single peak. In addition, it marks them as shared when they are connected to multiple PSMs. The node creates a Quan Info column for the LCMS Features page that displays the following values:

- No Quan Values: Indicates that the feature area is 0.

- Rejected by Method: Indicates that the peptide was not used in the quantification.

- Single Peak: Indicates that the feature consists of only a single LC/MS peak and that single peak features should not be used.

- Missing Replicates: Indicates that not enough replicate files have detected features. For details, see "Missing Replicate Features" on page 464.

- Shared: Indicates that different PSMs are associated with a feature.

When there are different labeling states for a peptide, the node creates a Quan Channel
column for peptide groups and PSMs, where the channel name of its labeling state is shown.
The node also creates a Quan Result ID column for the Peptide Groups page, showing an
integer number. Peptide groups belonging to the same quantification result with different
labeling states have the same quantification result identifications. The node also ensures that
all consensus features of a quantification result are connected to all the different peptide
groups of this quantification result.

In the next step, the Precursor Ions Quantifier node validates consensus features and calculates
the quantification values for the different files. It checks which LC/MS features can be used
and excludes invalid features. It creates two columns, Quan Info and Quan Info Details, for
the Consensus Features page, showing the following values. (The Quan Info Details column is
hidden and only shows information about why a quantification result is labeled Rejected by
Method.)

- No Quan Values: Indicates that no valid features are associated with a consensus feature.

- Has Missing Values: Indicates that not all expected labeling states for a consensus feature
  group are available, the Reject Quan Results with Missing Channels parameter of the
  Precursor Ions Quantifier node is set to True, and no imputation is used. This value
  appears only when labeling was used.

- Rejected by Method (only when labeling was used): Indicates that the peptide was not
  used in the quantification.

- Unknown Quan Channel: Indicates that no channel could be assigned.

- Unknown Quan Result: Indicates that no quantification result identification could be
  assigned.

- Shared: Indicates that multiple peptide groups are associated with a consensus feature.

Note The node checks for missing values and unknown quantification channels after you
specify quantification usage of consensus features for peptide groups and assign
quantification channels and quantification results to the consensus features.

A consensus feature shows abundances for a retention-time and *m/z* value over the different
spectrum files used in the processing step. For calculating abundances, the Precursor
Quantifier node sums areas of associated LC/MS features for the different files.

Missing Replicate Features

The Min. # Replicate Features [%] parameter of the Precursor Ions Quantifier node defines
the minimum percentage of replicate files that a feature must be detected in to be used. Its
default value is 0. This parameter affects consensus features in label-free quantification. It is
not supported for labeled precursor quantification, because it can be hard to identify the
replicate files.

When you set this parameter to a value greater than zero, the node determines which files belong to one replicate. Which files that belong to a replicate group depends on your study setup. If you use a biological replicate factor and multiple files are grouped into one biological replicate group, the files in one biological replicate group belong to one replicate group. If you do not have a biological replicate factor, the files of a sample group belong to the one replicate group.

When the Precursor Ions Quantifier node calculates the abundances of a consensus feature, it determines whether quantification values were found for enough replicate files. If there are not enough files, it excludes the LC/MS features of the replicate group from quantification. It displays Rejected by Method in the Quan Info column, Missing Replicates in the Quan Info Details column, and Not Used in the Quan Usage column.

The filter checks the LC/MS features by replicate group. If one replicate group does not have not enough quantification values, the node can still use the LC/MS features from other replicate groups when there are enough features.

Identifying Isotope Patterns in Precursor Ion Quantification

The quantification spectra on the pages of the .pdResult file show the isotope pattern used for quantifying the peptides. The algorithm used in precursor ion quantification finds isotope patterns by identifying target components—that is, known elemental compositions from event lists. It identifies the peptides and searches in the event lists for the isotope patterns of these identified peptides. After peptide identification, the algorithm follows the steps shown in Figure 342 to identify the isotope patterns.

Figure 342. Identifying isotope patterns
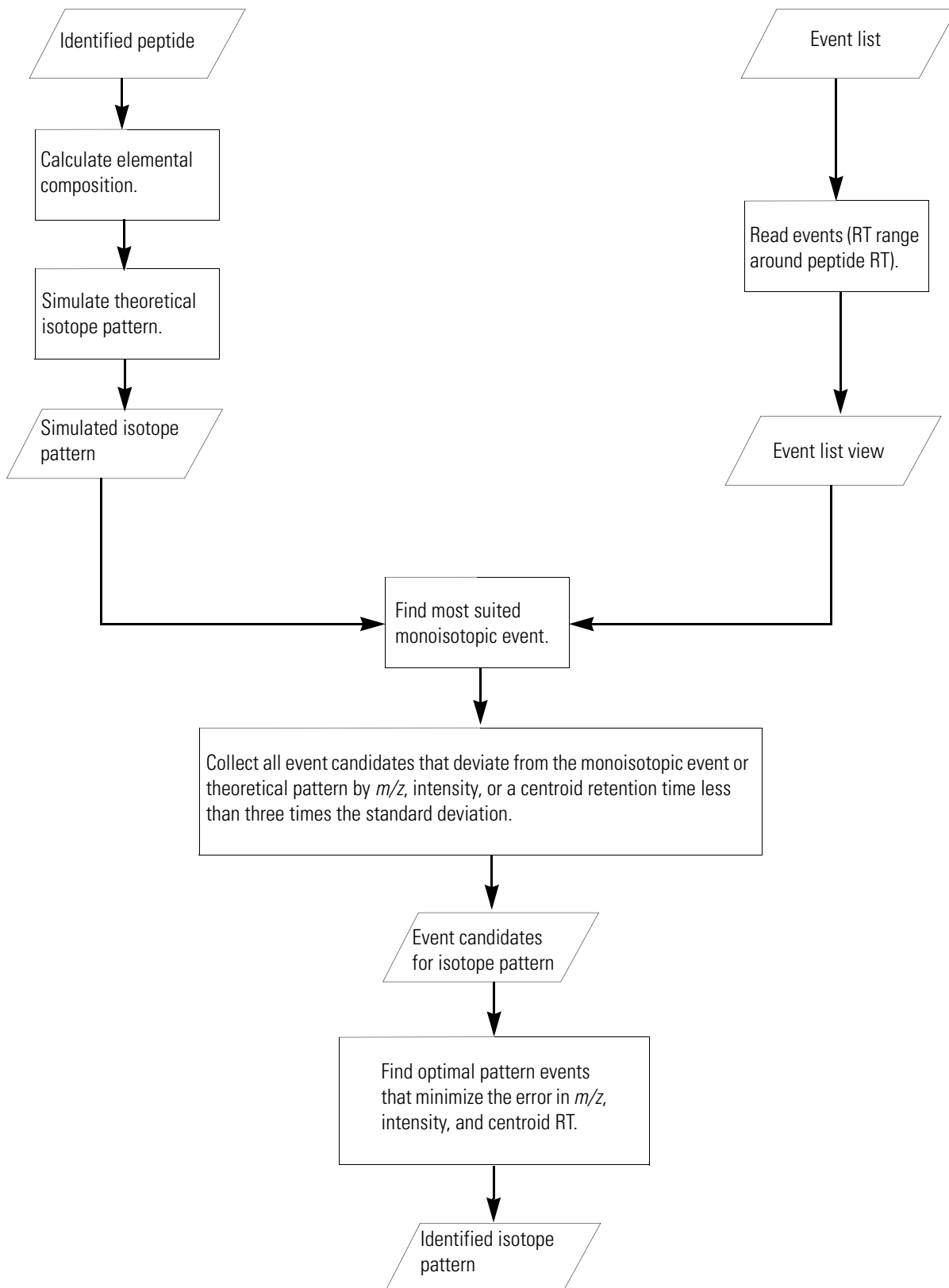
Performing Label-Free Quantification

This chapter describes how to perform label-free quantification.

Contents

Label-Free Quantification

Label-free quantification is the quantification of MS peptide signals in multidimensional LC/MS/MS data without an isotopic labeling reagent. It uses the Minora algorithm to perform untargeted feature detection; that is, it does not require an identification to quantify a peptide or protein. In theory, it quantifies the most peptides and arguably provides the widest dynamic range for untargeted workflows. It requires only a single PSM to quantify across all samples when you use retention time alignment and gap filling.

However, any variability in sample preparation, liquid chromatography, or mass spectrometry more strongly affects label-free quantification than it does isotopically labeled techniques where differently labeled peptides coelute. In addition, it provides the lowest throughput because of a lack of multiplexing. Finally, peptides might have different levels of ion suppression across the various samples, leading to a decrease in quantification accuracy.

The processing workflow for label-free quantification must include the Minora Feature Detector node, and the consensus workflow must include the Feature Mapper and Precursor Ions Quantifier nodes.

This discussion of label-free quantification includes the following topics:
- Minora Feature Detection
- Retention-Time Alignment and Feature Mapping
- Fractionated Samples
- Mass Recalibration
- Calculating Label-Free Quantification Values

Minora Feature Detection

Label-free quantification uses the Minora Feature Detection node in the processing workflow to detect features in individual raw data files and to map them to PSMs, if they are available. A feature is signal for a single molecule and charge state across the elution profile.

As a first step, the Minora Feature Detector node detects LC/MS peaks in the raw data files and then maps them to identified PSMs. It calculates the theoretical isotope pattern of a PSM, finds the LC/MS peaks mapping to the isotope pattern, and creates a feature from the fitting peaks.

After the Minora Feature Detector node maps all identified PSMs to features, it creates features from peaks that are not associated with any feature. To do this, it takes all LC/MS peaks within a small retention-time range and checks for peaks that build isotope patterns. Mass deviations and intensities of the peaks must fit a theoretical pattern of an averagine peptide.

For finding untargeted features, the node splits the complete retention-time range of a file into slices having the equal retention-time range width, although the time slices slightly overlap. The overlap is twice the value of the Pattern RT Tolerance parameter setting. The node takes all peaks within a time slice and searches for other peaks that build an isotope pattern for different charges. It searches the most intense peaks first. If it finds pattern peaks, it creates a feature from them and labels the peaks "used" to ensure that each peak is used only in a single pattern.

The Minora Feature Detector node creates LC/MS features for all identified patterns and stores them in a feature table where targeted features are already stored.

Retention-Time Alignment and Feature Mapping

> **Note** Retention-time alignment is also called chromatographic alignment.

The Feature Mapper node in the consensus workflow performs retention-time alignment and feature linking across data sets. It first chromatographically aligns the input files in a sample set. Specifically, it uses the results from the Minora algorithm to find the most abundant features and then aligns them to have the same retention time across all sample files. It uses the file in which most of these features are found as the reference for the rest of the alignment. For each file, the node tries to find a matching feature for every feature from the reference file. For each of these feature matches, it records the retention time difference in the two files. Finally, it fits a regression curve into the (RT, ΔRT) points. The node uses this curve to correct the retention times of the file relative to the chosen reference file.

For multidimensional data, the Feature Mapper node only aligns files having the same fraction number. It uses the grouping information in the Grouping & Quantification page in the study to align them. Before you start an analysis that uses the Feature Mapper node in the consensus workflow for label-free quantification, you can use this page to define sample groups of files that should be grouped together. Usually, you group files having equal properties, such as replicate files. See "Grouping Similar Files" on page 507.

The Feature Mapper node subsequently links peptides and features and maps features across multiple files. It converts the LC/MS features from each individual raw data file into consensus features by matching retention time, monoisotopic *m/z*, and charge. It compares isotope patterns, retention time, and mass to determine if the peptide is the same. The node also tries to fill the gaps for files where it did not find a feature in the processing step. Only one data file needs an identified PSM for the Feature Mapper node to quantify a peptide across all files.

Whenever a feature is missing in one file but found in others—for example, in the case of features measured near the detection limit—the node's feature-mapping algorithm identifies a feature gap and tries to fill it with an unused neighboring single peak. Each single peak used for gap-filling that is above the threshold specified by the node's Single Peak Score Threshold parameter is included as a single-peak feature with the same charge state as all other features of this consensus feature.

Figure 343 shows the feature mapping on the left and its resulting consensus feature on the right, where:

- F stands for a feature.

- P stands for a single peak above the threshold.

- P_T stands for a single peak below the threshold.

Figure 343. Gap filling
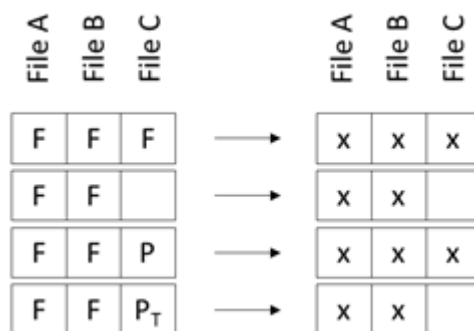
A consensus feature can consist of only single-peak features if there are no missing values or there is at least one PSM assigned to it. In a final step, the node assigns all features that cannot be identified in the processing step to consensus features. This step uses the parts per million and retention time tolerances set in the Feature Mapper node for grouping. The features must also have the same charge state.

Figure 344 shows the feature mapping on the left and its resulting consensus feature on the right, where P stands for a single peak above the threshold.

Figure 344. Gap filling for single-peak features
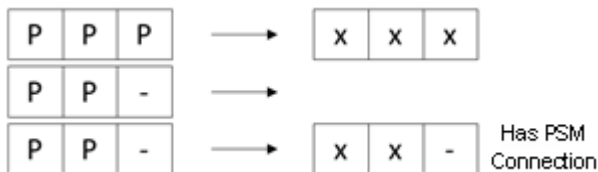
Fractionated Samples

The Proteome Discoverer application only performs retention-time alignment and feature mapping between the corresponding fractions in a fractionated sample. For example, it might align fraction 1 across all samples, where all fraction 1's contain the same peptide. Therefore, you must order fractions when adding them to a study. For information on ordering fractions, see "Assigning the Order of Fractions" on page 46.

Mass Recalibration

Spectrum recalibration improves the feature mapping for label-free quantification. In this type of quantification, you must use the Spectrum Files RC node instead of the Spectrum Files node as the starting node in the processing workflow. All the subsequent nodes in the workflow, such as Sequest HT, receive recalibrated spectra. The Minora Feature Detector, Feature Mapper, and Precursor Ions Quantifier nodes also use recalibrated LCMS peaks, LCMS features, and quantification spectra.

For information on the Mass Recalibration view, see "Displaying the Mass Recalibration View" on page 629. For information on the Mass Recalibrations page in the .pdResult file, see "" on page ccxxxix.

Calculating Label-Free Quantification Values

The Precursor Ions Quantifier node in the consensus workflow validates the quantification values and calculates abundances for peptide groups and proteins.

As a first step, it calculates and validates the raw quantification values using the Minora Feature Detector node in the Processing workflow. A quantification value is the intensity, area detected for a given quantification channel depending on the setting in the Precursor Ions Quantifier in the Consensus method. These label-free quantification values are subsequently mapped across runs to become consensus features using the Feature Mapper node in the Consensus workflow after a retention time alignment step has been performed.

Performing Label-Free Quantification

This topic describes the default procedure to use to perform label-free quantification and then describes the processing and consensus workflows in detail.

- Allocating Sufficient Storage Space
- Downloading FASTA Files
- Checking for the Quantification Method
- Opening an Existing Study
- Creating a Study
- Adding a Description
- Adding the Study Factors
- Adding Input Files
- Setting the Factor Values for the Samples
- Saving a Study
- Creating an Analysis
- Opening an Existing Analysis
- Adding Input Files to an Analysis
- Using Multiple Processing Steps in an Analysis
- Adding or Deleting a Processing Step
- Creating a Processing Workflow for Label-Free Quantification
- Creating a Consensus Workflow for Label-Free Quantification
- Saving an Analysis
- Performing the Quantification
- Working with the Search Results
- Specifying Quantification Ratios from Selected Sample Groups for Label-Free Quantification
- Grouping Similar Files

Allocating Sufficient Storage Space

Be sure that you allocate enough space to store the input raw data files and the temporary result files that the application creates until it creates the final result file. See "Configuring Temporary Files Parameters" on page 32.

Downloading FASTA Files

The first step in performing reporter ion quantification is to download an appropriate FASTA file. For information on downloading a FASTA file, see "Downloading FASTA Files to Proteome Discoverer" on page 153.

Checking for the Quantification Method

For label-free quantification, you should not select a specific quantification method.

Opening an Existing Study

Now you open an existing study or create a study. To create a study, see "Creating a Study" on page 472.

❖ **To open an existing study**

In the Recent Studies area of the Start Page, click the study name.

–or–

In the Recent Studies area of the Start Page, right-click the study name and choose **Open**.

–or–

Choose **File > Open Study** or click the **Open Existing Study** icon, , and browse to the study folder and then select and open a .pdStudy file.

–or–

1. In the Start area of the Start Page, click **Open Study**.

2. In the Open Study dialog box, browse to and select the name of the study, and click **Open**.

Creating a Study

❖ **To create a study**

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

–or–

Choose **File > New Study/Analysis**.

–or–

Click the **Create New Study/Analysis** icon, .

The New Study and Analysis dialog box opens, as shown in Figure 345.

Figure 345. New Study and Analysis dialog box

In this example, you only specify the name of the study and a root directory to save the study in.

2. In the Study Name box, specify the mandatory study name.

   The application generates a default study name by searching for the common part of the file names when you add multiple files at once and using this common part as the suggestion for the name of the new study.

3. In the Study Root Directory box, specify the folder where you will store the study folder. Click the **Browse** button (...), and in the Select Folder dialog box, specify the folder and click **Select Folder**.

4. Click **OK**.

   The application creates a new study folder in the folder that you specified as the root directory and opens a new page with the study name (Study: LFQ Example in the example), as shown in Figure 346. It appends a .pdStudy extension to the study file name.

Figure 346. New Study: *Study_name* page
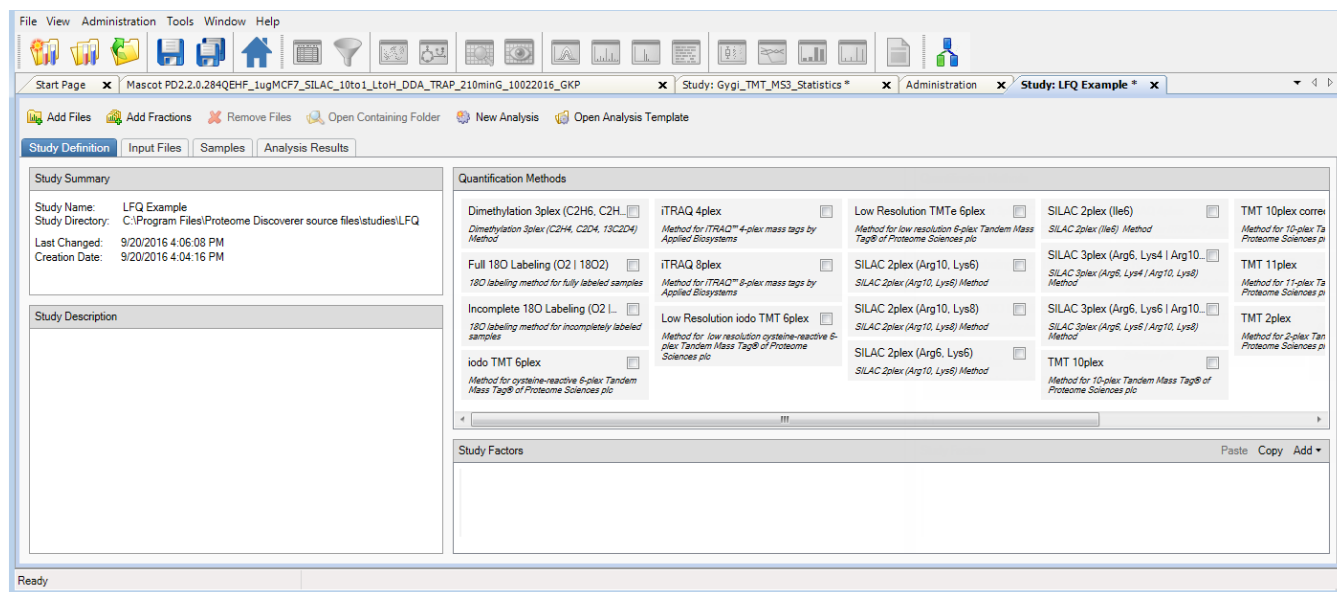
On the Study Definition page, you add a description of your study and set up the new factors to use to describe and distinguish your samples.

Adding a Description

You can optionally add a description of the study by typing it in the Study Description area of the Study Definition page.

Adding the Study Factors

In this step, you add the study factors that you want to use for your samples.

A factor is a single biological or technical parameter that you control, for example, genotype, diet, environmental stimulus, age, column length, spray voltage, or collision energy.

In theory, you can track every parameter as a factor in your study, but normally you only track the parameters that actually differentiate samples from each other. For example, the gradient that you use for the chromatography is a parameter in your experiment. However, if you use the same gradient for all your samples, you would not add this as an explicit factor to your study.

❖ **To add numerical study factors**

1. Add the first numerical factor:

   a. In the Study Factors area of the Study Definition page, choose **Add > Numerical Factor**.

   The numerical factor dialog box shown in Figure 347 opens.

Figure 347. Numerical factor dialog box
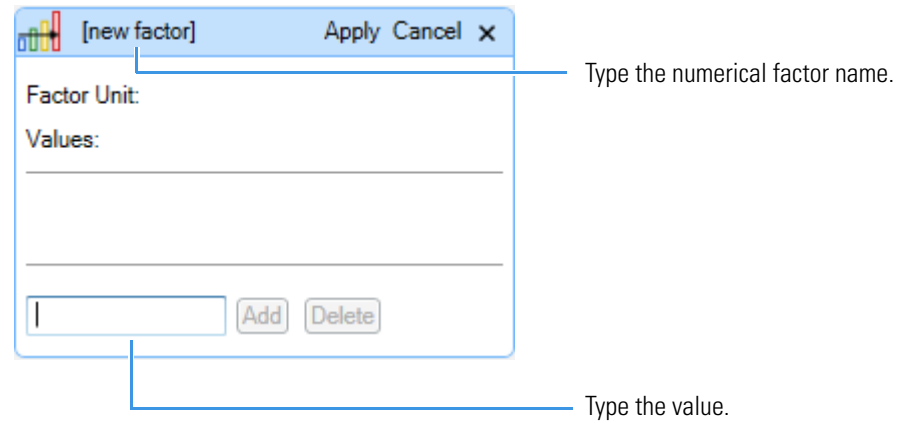
Type the numerical factor name.

Type the value.

[new factor] is highlighted.

b. Type a name over [new factor] for the new numerical factor, for example, **Time (days)**. See Figure 348.

> **Note** If the full numerical box becomes compressed, click **Edit** to restore it to its original size.

c. In the box to the left of the Add and Delete buttons, type the name of the first value, which is 5, and click **Add**. See Figure 348. Continue by adding values 7, 9, 11, 13, 15, 17, 25, 29, and 33.

Figure 348 shows the completed numerical factor dialog box.

Figure 348. Completed numerical factor dialog box
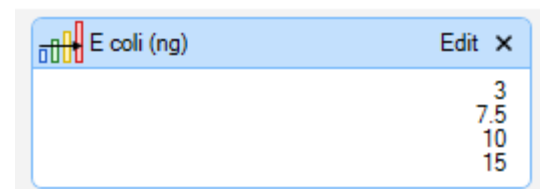
d. Click **Apply** in the numerical factor dialog box.

Confirm that the Study Definition page resembles Figure 349.

Figure 349. Numerical factor added to the study
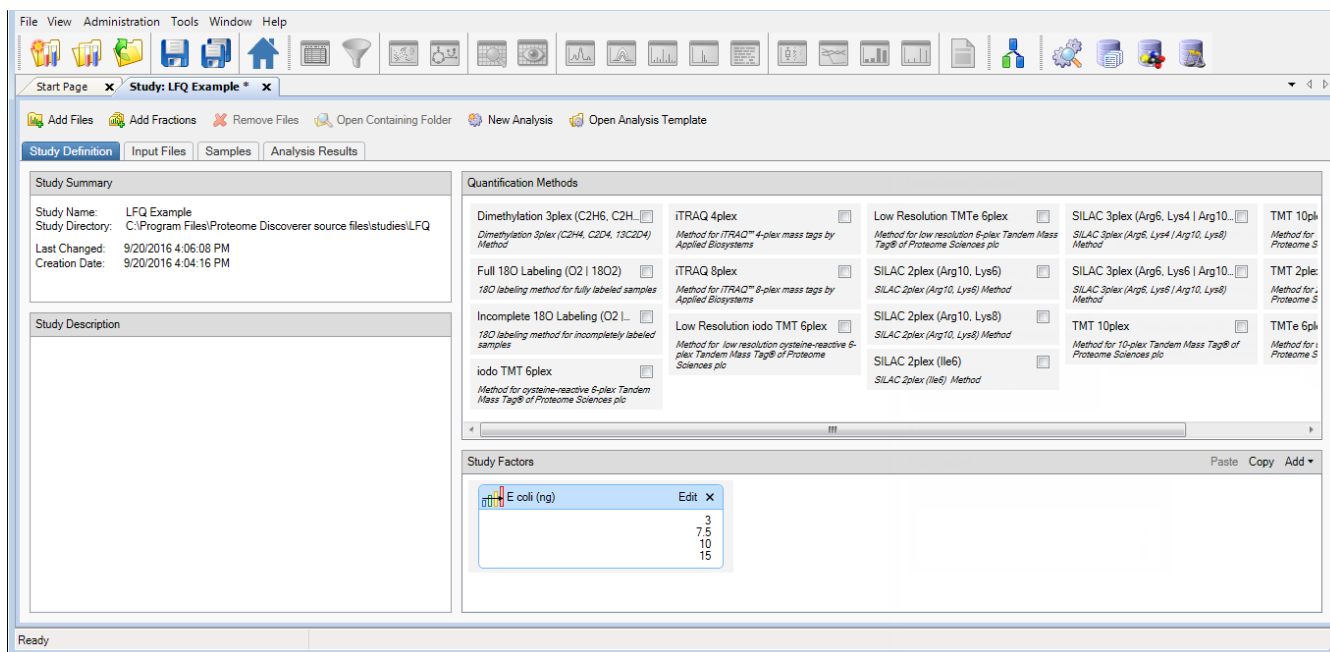
❖ **To add biological replicate study factors**

1. Add the first biological replicate factor:

   a. In the Study Factors area of the Study Definition page, choose **Add > Biological Replicate Factor**.

   The biological replicate factor dialog box shown in Figure 350 opens.

   Figure 350. Biological replicate factor dialog box
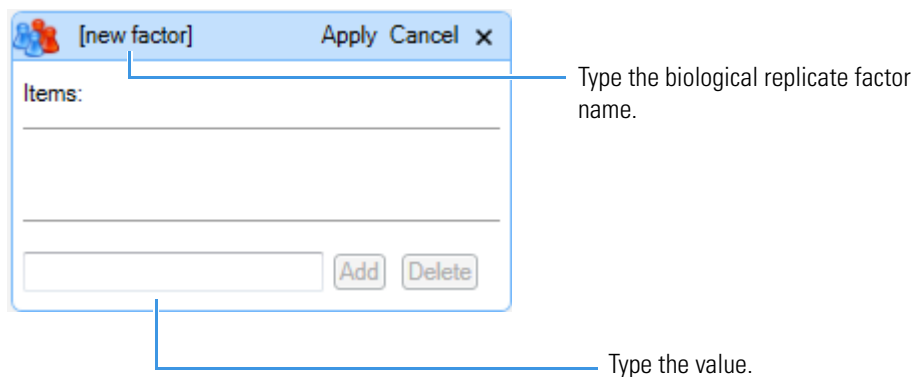
   [new factor] is highlighted.

   b. Type a name over [new factor] for the new numerical factor, for example, **Patient**. See Figure 351.

   > **Note** If the full numerical box becomes compressed, click **Edit** to restore it to its original size.

c. In the box to the left of the Add and Delete buttons, type the name of the first value, for example, Patient1, and click **Add**. See Figure 351. Continue by adding values Patient2, Patient3, Patient4, and so forth.

Figure 351 shows the completed biological factor dialog box.

Figure 351. Completed biological replicate factor dialog box
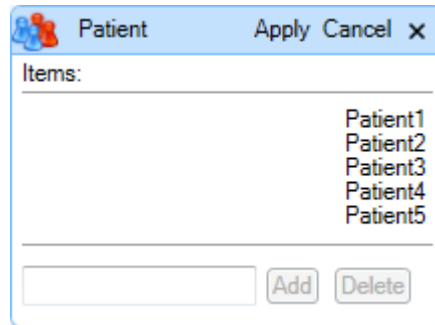
d. Click **Apply** in the numerical factor dialog box.

Confirm that the Study Definition page resembles Figure 352.

Figure 352. Biological replicate factor added to the study
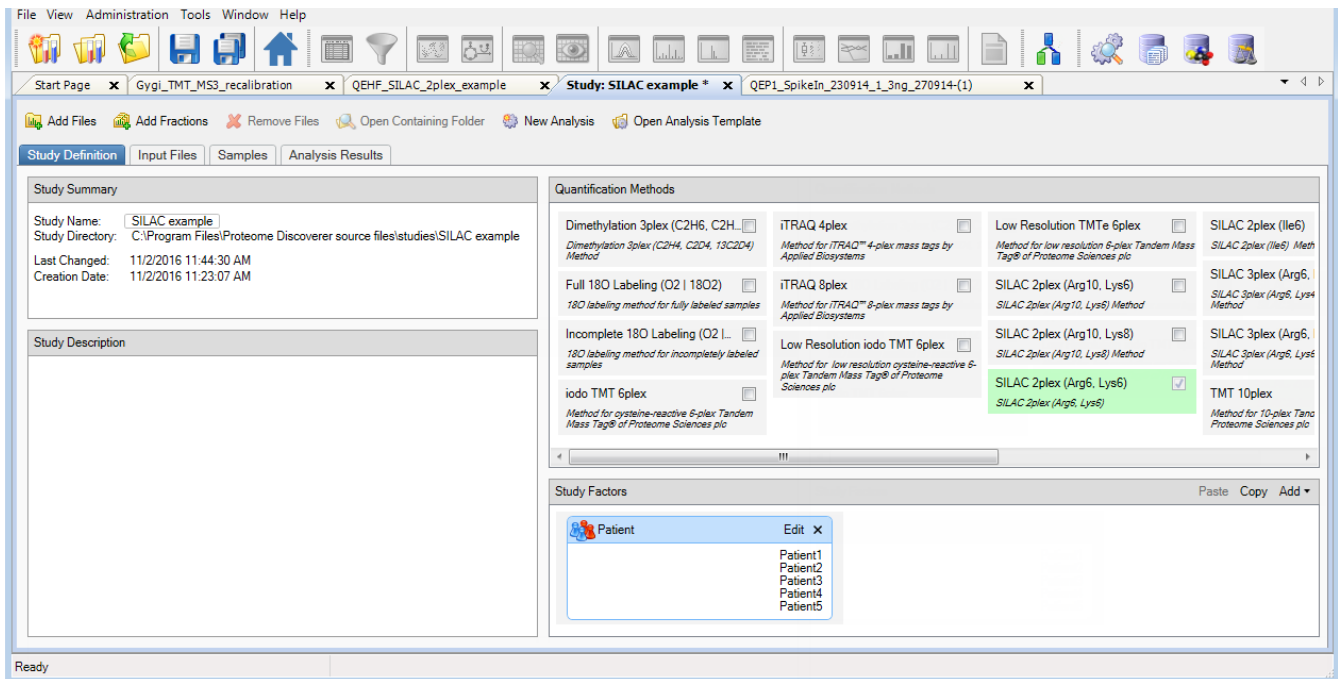
2. Add a second biological replicate study factor, if applicable.

❖ **To add categorical study factors**

1. Add the first categorical (non-numeric) factor (Acquisition in the example):

a. In the Study Factors area of the Study Definition page, choose **Add > Categorical Factor**.

The categorical factor dialog box shown in Figure 353 opens.

Figure 353. Categorical factor dialog box
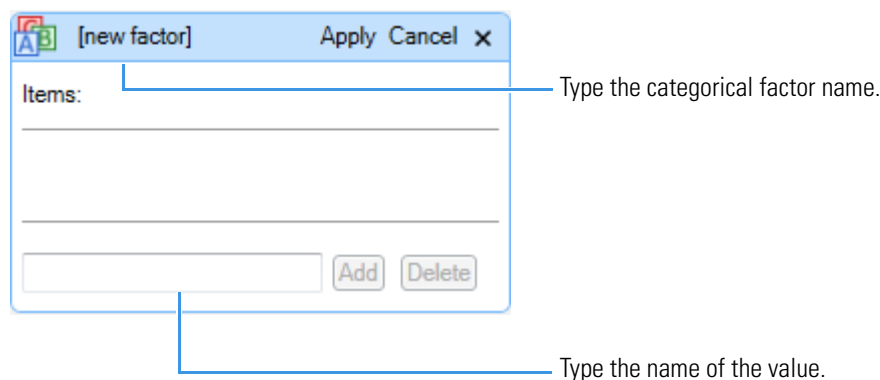
[new factor] is highlighted.

b. Type a name over [new factor] for the new categorical factor, for example,
Acquisition. See Figure 354.

> **Note** If the full categorical box becomes compressed, click **Edit** to restore it to its
> original size.

c. In the box to the left of the Add and Delete buttons, type the name of the first value
and click **Add**. See Figure 354.

In this example, the value is the acquisition method, and the first acquisition method
is intelligent data acquisition (IDA).

d. In the same box, type the name of the second acquisition method and click **Add**.

In this example, the second acquisition method is data-dependent acquisition
(DDA). See Figure 354.

Figure 354 shows the completed categorical factor dialog box.

Figure 354. Completed categorical factor dialog box
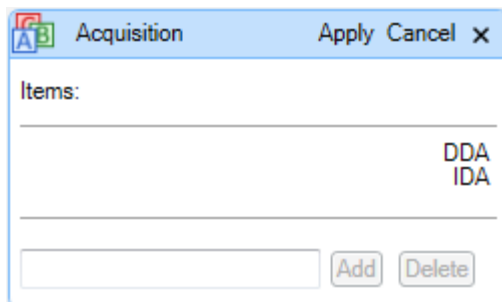
e. Click **Apply** in the categorical factor dialog box.

Confirm that the Study Definition page resembles Figure 355.

Figure 355. First categorical factor added to the study
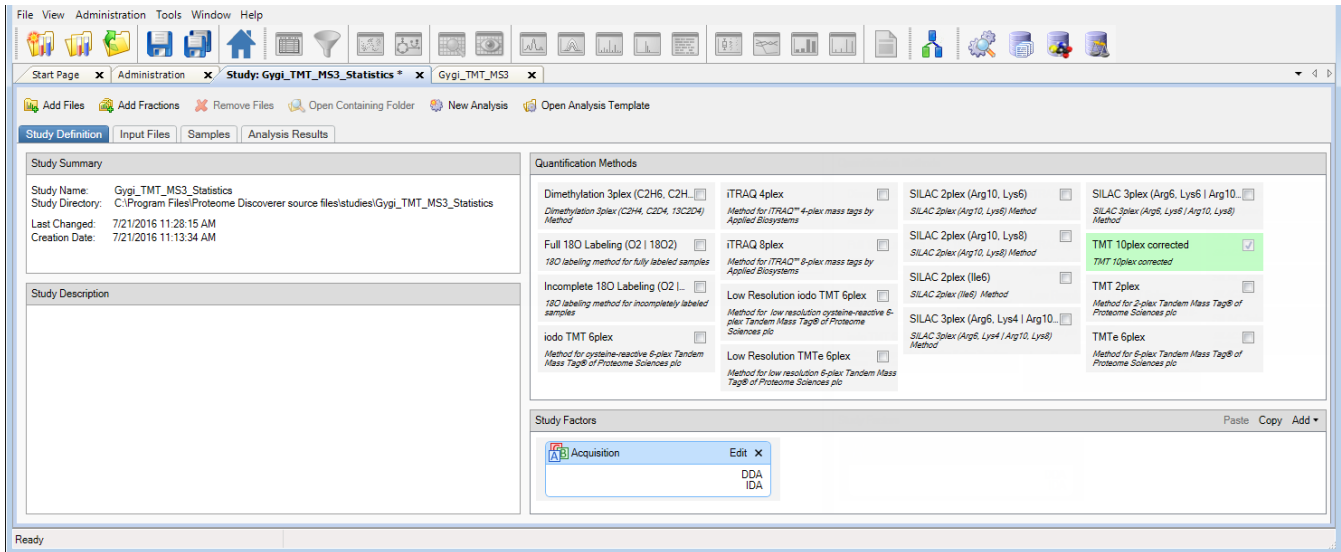
2. Add the second categorical factor (Tissue in the example):

   a. In the Study Factors area of the Study Definition page, choose **Add > Categorical Factor**.

   The dialog box shown in Figure 353 on page 478 opens.

   [new factor] is highlighted.

   b. Type a name over [new factor] for the new factor (see Figure 354 on page 478), for example, Tissue.

   c. In the box to the left of the Add and Delete buttons (see Figure 354 on page 478), type the name of the first type of tissue and click **Add**.

   In this example, the first type of tissue is Cerebellum.

   d. In the same box, type the name of any additional types of tissue and click **Add** after each one.

   The example adds the following types of tissue to the study:

   • Cerebellum

   • Cerebrum

   • Heart

   • Kidney

   • Liver

   • Lung

> • Muscle
>
> • Spleen

e. Click **Apply** in the Tissue dialog box.

> Confirm that the Study Definition page resembles Figure 356.

Figure 356. Second categorical factor added to the study

Adding Input Files

Add the input files from the example data set to your study. (You can add input files to the study at any point.) For the types of input files supported, see "Inputs" on page 10. You can add individual input files, multiple unrelated input files, or fractions.

Adding a Single Input File or Multiple Unrelated Input Files

❖ **To add a single input file or multiple unrelated input files**

1. Click ![Add Files] .

2. In the Add Files dialog box, browse to the location of the input files, and select them.

> **Note** You can also add existing MSF files to a study. The application adds the MSF file as a new result to the Analysis Results page and adds the raw data files that were processed to create the MSF files to the Input Files page.

3. Click **Open**.

–or–

Drag the input file or files from Windows Explorer and drop them onto the Input Files page.

The input files appear on the Input Files page, as shown in Figure 357. Each file on the page receives a unique identifier: F1, F2, …, F*n*. The Proteome Discoverer adds each file as a single study file.

Figure 357. Input files on the Input Files page
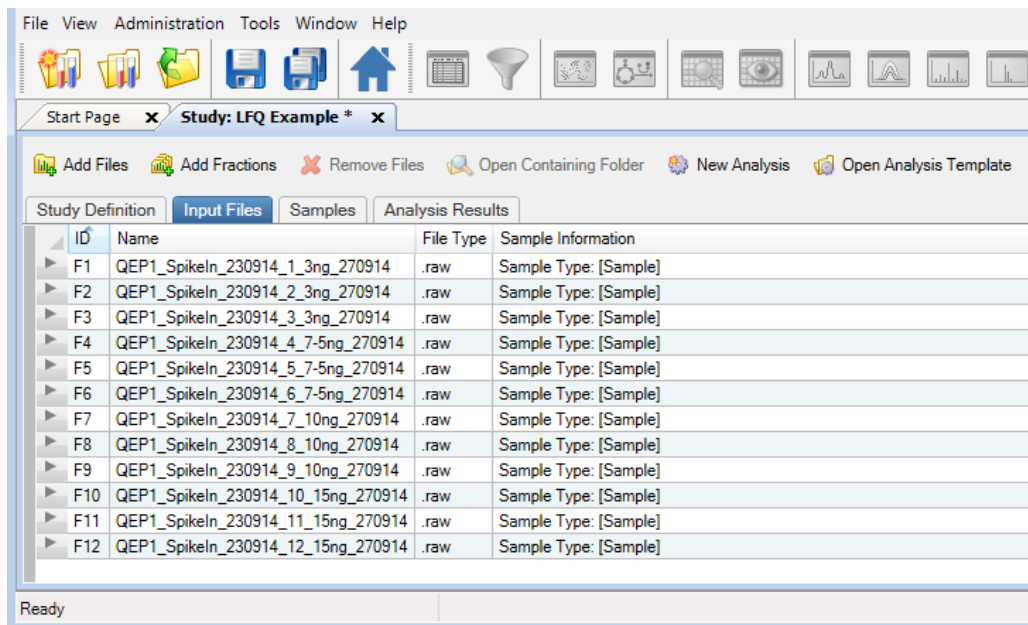
Adding Fractions

To add fractions, follow the instructions in "Adding Fractions" on page 481.

Creating Subsets of Fractions

To create subsets of fractions, see "Creating Subsets of Fractions" on page 369.

Assigning the Order of Fractions

To assign the order of fractions, see "Assigning the Order of Fractions" on page 371.

Specifying the Quantification Method for Multiple Input Files

Specifying the quantification method is not necessary for label-free quantification.

Setting the Factor Values for the Samples

The term "sample" refers to a distinct measurement of an analyte. At a minimum, a RAW data file contains at least a single sample, or in the case of sample multiplexing, it uses isobaric or metabolic labeling for multiple samples. If you measure the analyte from the same vial in two acquisitions—that is, if you measure two technical replicates—the application considers them two different samples. If you split the analyte, label each with a different isobaric tag, mix them, and acquire them into a single RAW data file, the application considers the two differently labeled parts as two different samples.

Each sample is associated with a sample type. Currently only quantification uses sample types. The sample type can be one of the following:

- Sample: A specimen from a larger biological entity

- Control: A sample typically used as a reference sample in a quantification experiment

You can create ad hoc relative quantitative ratios between any study factors or sample types.

The Sample Type column on the Samples page of the study displays the sample type of each sample. The default sample type is Sample. If a file has samples for different quantification channels, mark one of the samples Control. In this example, the channel that was used to label the mouse kidney tissues is used as the control for generating the quantification ratios.

You can also use the control channels in scaling. For more information, see "Normalizing Peptide Groups and Protein Abundances" on page 535.

Each sample is associated with a quantification channel shown in the Quan Channel column and with values for each of the factors that you specified for your study. Previously, you specified a factor for the acquisition method used and a factor for the tissue that was extracted. You now set the correct factor values for each of the samples in the study.

Each sample has an automatically generated sample name composed of the raw data file name and the appended name of the quantification channel. You can change this name, but the name must be unique among all samples in the study.

❖ **To view the samples**

1. On the Input Files page, click the gray arrow to the left of a sample to display its constituent file entries, as shown in Figure 358.

   A hierarchical view opens, showing the samples contained in a raw data file.

Figure 358. Samples in a raw data file

Click the red arrow to close a file entry.

Click the gray arrow to expand a single file entry.
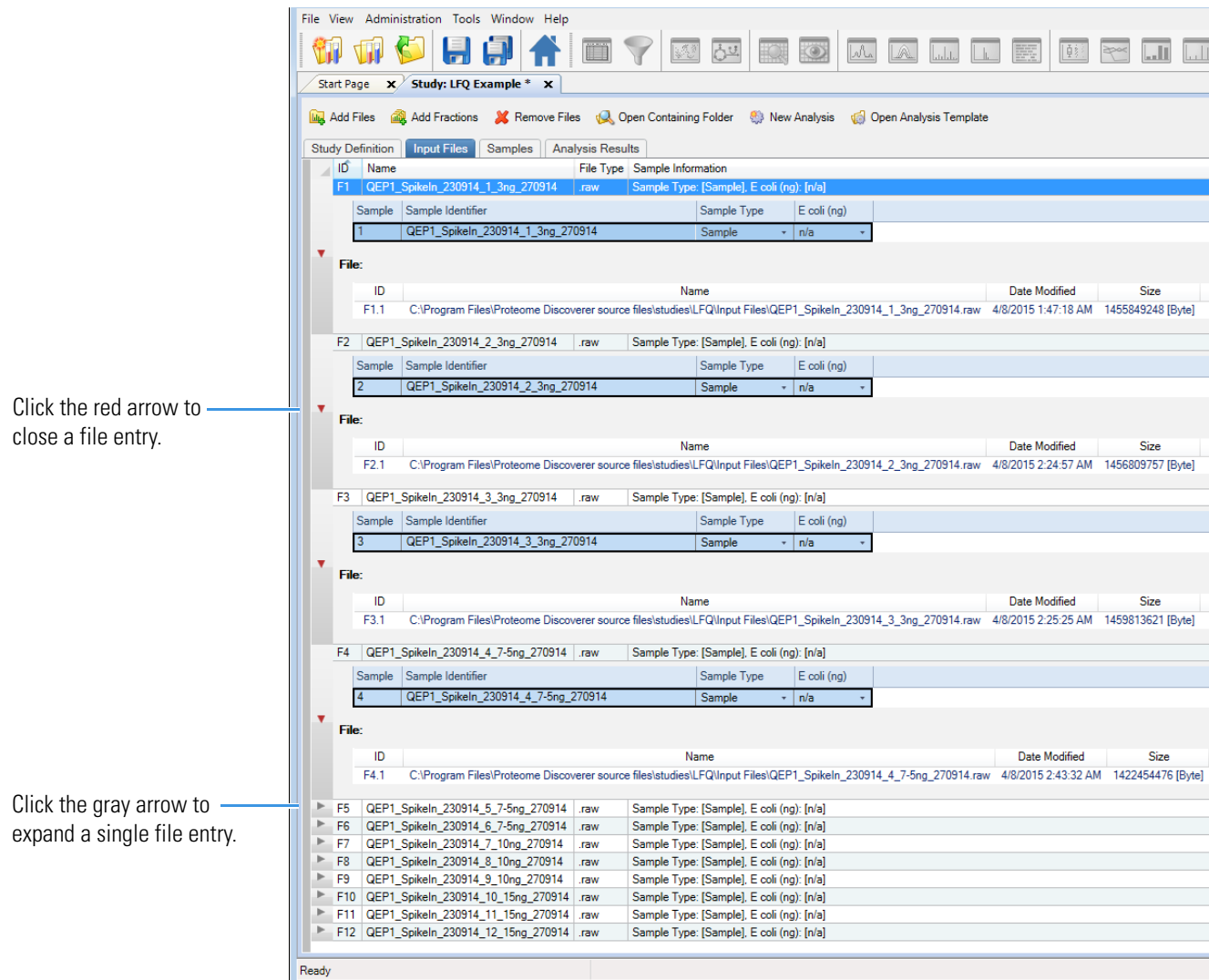
For label-free quantification, which involves raw data files containing only one sample, it is faster to set the factor values for the samples on the Samples page rather than on the Input Files page.

❖ **To set the factor values for the samples**

1. Click the **Samples** tab, if it is not already selected.

2. Click the gray arrow next to the first sample on the Samples page to expand the information about the sample.

3. For the factor (in the example, E coli (ng)), set the value for each sample in each raw data file by selecting the down arrow in the factor column and then selecting the value from the list, as shown in Figure 359.

Figure 359. Setting the factor values for the Ecoli (ng) factor on the Samples page
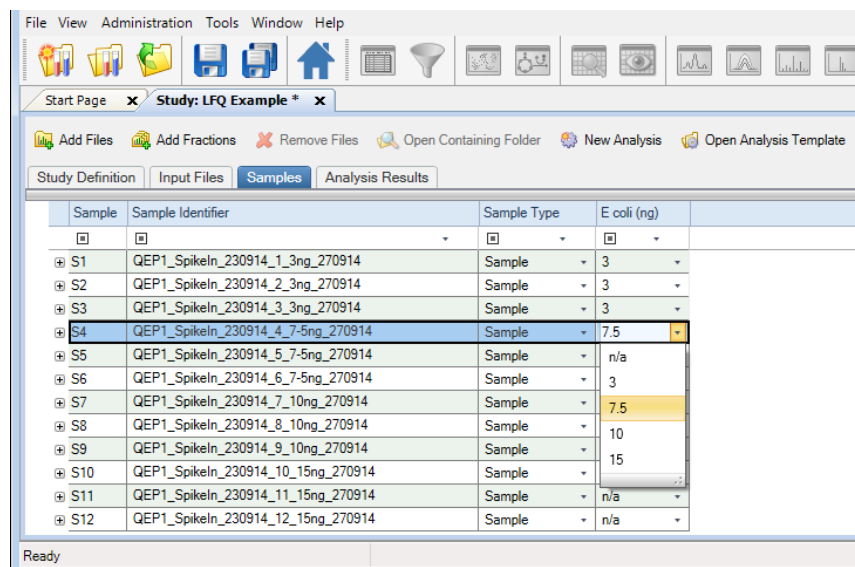
Figure 360 shows all the factors set on the Samples page. It displays the same sample information as the Input files page.

Figure 360. Samples page
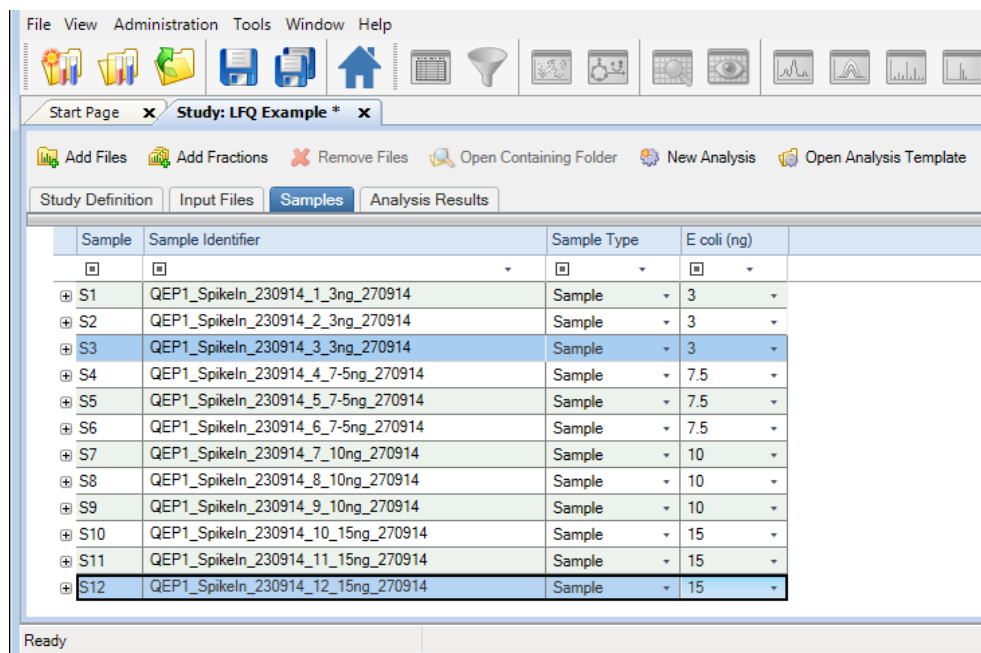
Setting Values for Multiple Samples at the Same Time on the Samples Page

If you have several files and samples with many study variables to set (sample type, quantification channel, study factors), you can set values for study variables for multiple samples at once. You can use either the mouse or the keyboard to set these values on the Samples page.

You can also use multicell editing on the Samples page to enter factor values for several samples. For example, if you have acquired data from several different tissues and the tissue is encoded in the file and sample names, you can use the filter option on the Samples page to filter all samples that contain "liver" and then set the Tissue factor to "liver" for all these samples.

❖ **To set values for multiple sample cells at the same time by using the mouse**

1. Click the **Samples** tab.

2. To select multiple samples for one study variable, position the cursor to the immediate right of the cell in the factor column (in the example, the E coli (ng) column) and drag the cursor to select the remaining cells.

–or–

Holding down the SHIFT key, use the up and down arrow keys on your keyboard to select the cells that you want to edit.

Figure 361 shows multiple cells being selected in the E coli (ng) column.

Figure 361. Selecting multiple cells at one time

3. To the immediate right of the last cell that you want to select in the factor column, right-click and choose **Set *factor_name* to > *factor_value***, as shown in Figure 362.

Figure 362. Selecting a new value to apply to multiple cells
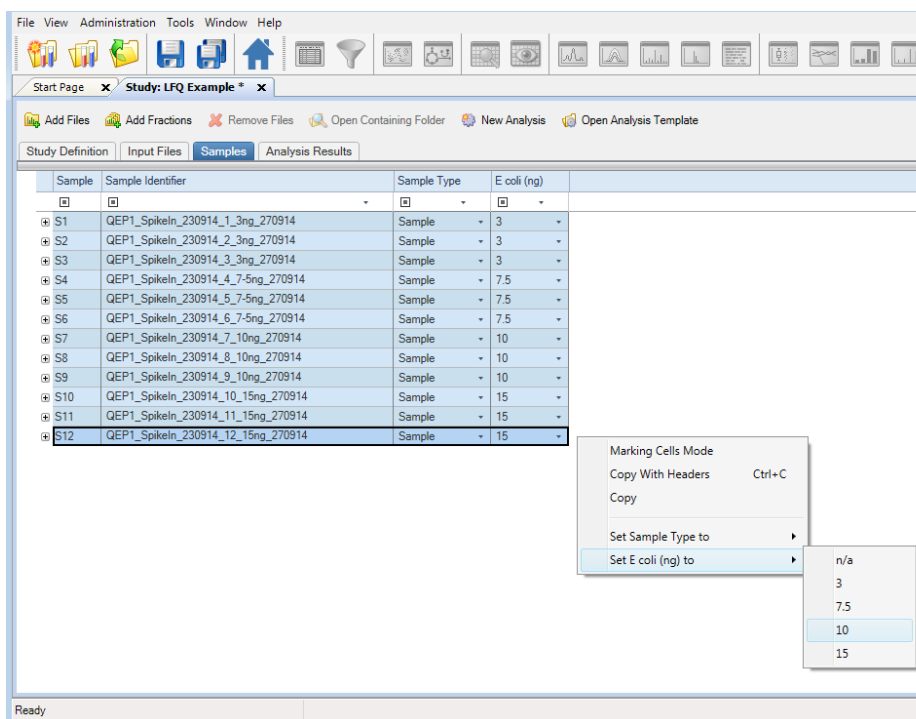
In the example shown in Figure 363, the values in the selected cells of the E coli (ng) column change from various values to 10.

Figure 363. Multiple values changed at one time
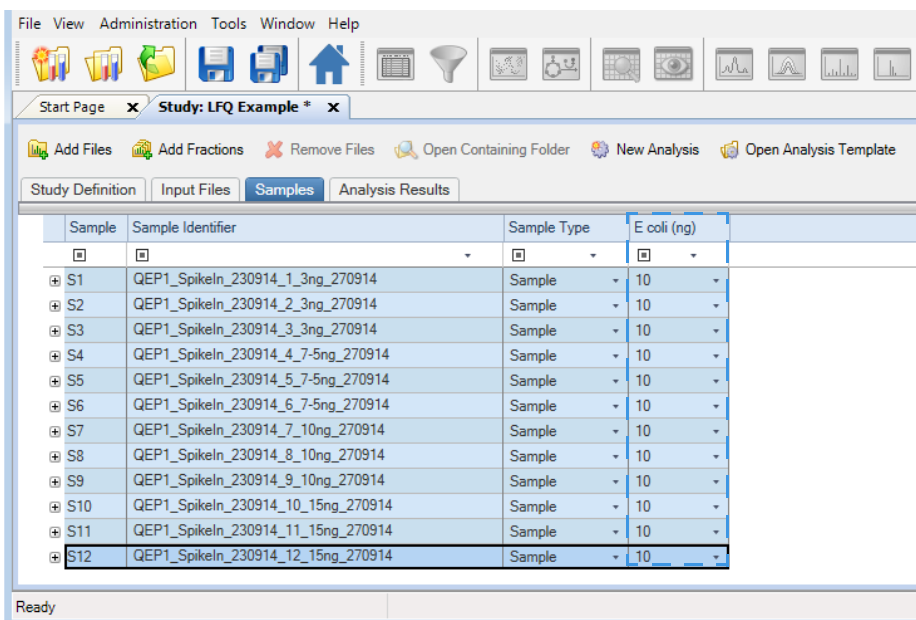
Setting Filters for Multiple Samples at the Same Time on the Samples Page

Use the following procedure to change a filter for multiple samples on the Samples page.

❖ **To set a filter for multiple samples at the same time on the Samples page**

1. Click the **Samples** tab.

2. Click the **Contains** icon, ▣ , if necessary, in the appropriate column of the Samples page, and choose the type of filter from the menu.

> **Note** In columns displaying categorical or biological replicate factors, you cannot use the Equals, Not Equals, Less Than, Less Than or Equal To, Greater Than, and Greater Than or Equal To filters. In columns displaying numerical factors, you cannot use the Contains, Does Not Contain, Like, Not Like, Match, Does Not Match, Starts With, Does Not Start With, Ends With, and Does Not End With filters.

3. In the box next to the icon, type the value of the filter.

   For example, suppose you want to filter out sample factors to show only those with values of 10. From the menu, you would choose "contains" and type 10 in the adjacent box. Figure 364 shows the results.

Figure 364. Filter in the factor column

To return the samples to their unfiltered state, select the **Clear All Filters** icon, ⊽x .

Saving a Study

You can save a study manually or automatically.

You can save a study manually at any time. A change in a study that requires you to save it is indicated with an asterisk (*) in the tab after the study name.

> **Note** Studies and analyses in the Proteome Discoverer are separate and must be saved separately. Saving a study does not save an analysis, and saving an analysis does not save a study.

You can set an option to have the application save studies automatically when you click ⚙ Run . This option also saves the results generated in the study. It saves the analysis containing the workflow, but when you close the study and reopen it, you must access the analysis by clicking the Analysis Results tab, and then doing one of the following:

Click ⚙ Reprocess ▾ , and choose **All Analysis Steps** (to open both the processing and consensus workflow) or **Last Consensus Step** (to open just the consensus workflow).

–or–

Right-click the name of the result file and choose **Show Details**.

❖ **To save a study manually**

Choose **File > Save**.

The Proteome Discoverer saves the study in the *study_name*.pdStudy file in the study directory.

❖ **To save a study automatically**

1. Before you click ⚙ Run , select **Tools > Options**.

2. In the Options dialog box, select **Study Options** in the left pane.

3. Select the **Auto Save When Starting Analysis** check box.

4. Click **OK**.

The Auto Save When Starting Analysis option does not take effect for any open studies. The application automatically saves only newly opened studies.

❖ **To save all open studies**

Choose **File > Save All**.

The Proteome Discoverer saves all studies open in the application.

Creating an Analysis

❖ **To create an analysis**

On the Study: *Study_name* page, click 🔧 New Analysis .

An Analysis window opens on the right side of the Study: *Study_name* page, as shown in Figure 365.

Figure 365. Analysis window
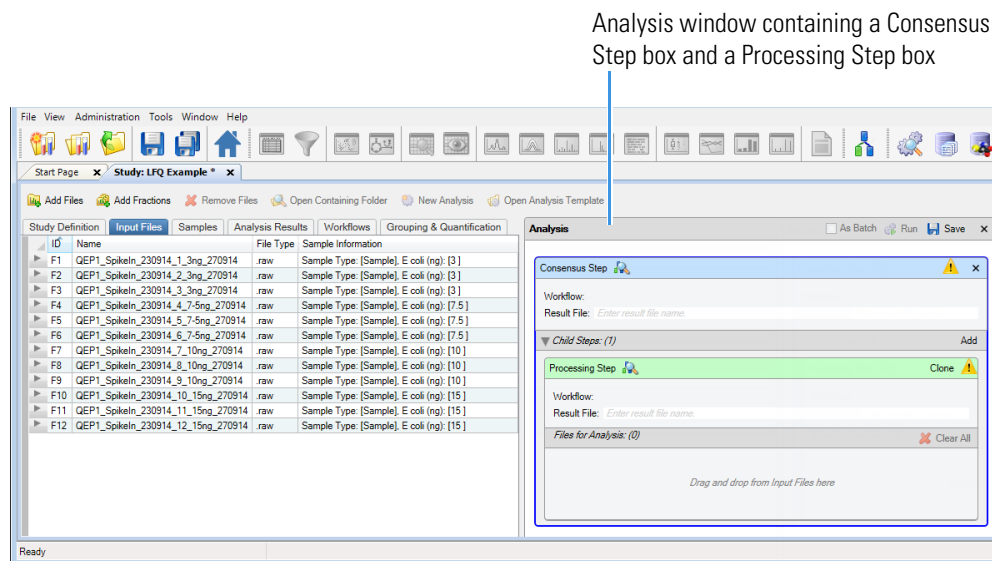
Analysis window containing a Consensus Step box and a Processing Step box

An Analysis window contains the following items:

- A Consensus Step box, which represents the consensus workflow step of the data processing

- A Processing Step box, which represents the processing workflow step of the data processing

- A Child Steps bar, which contains an Add button that you can use to add another Processing Step box. Multiple Processing Step boxes are useful when you want to process the same data in different ways—for example, by using different nodes or different node settings.

In addition, two new tabs appear on the Study: *Study_name* page: the Workflows tab and the Grouping & Quantification tab, as shown in Figure 365.

If you switch to an existing study that includes an Analysis window containing an existing consensus workflow, processing workflow, or both, and you want to open a new analysis, you must close the existing Analysis window and open a new Analysis window. To close the window, click **X** in the upper right corner.

Opening an Existing Analysis

You can open an existing analysis through a template or through a results file.

❖ **To open an existing analysis through a template**

1. In an open study, click Open Analysis Template .

2. In the Open Analysis Template dialog box, browse to and select the .pdAnalysis file that you want to open.

3. Click **Open**.

❖ **To open an existing analysis through a results file**

1. In an open study, click the **Analysis Results** tab in the study.

2. Select the appropriate .pdResult file.

3. Click Reprocess ▾ .

4. To open an Analysis window containing both the processing and consensus workflows, choose **All Analysis Steps** in the Reprocess list.

   –or–

   To open an Analysis window in order to execute just the consensus workflow, choose **Last Consensus Step** in the Reprocess list.

Adding Input Files to an Analysis

❖ **To add the input files to an analysis**

Select and drag the files from the Input Files page to the Input Files box of the Processing Step box in the analysis window.

The input files are listed in the Input Files area of the Processing Step box, as shown in Figure 366.

Figure 366. Input Files area of the Processing Step box of the Analysis window

Note You can remove a file from the Input Files area of the Processing Step box by clicking the X to the left of the file name.

Using Multiple Processing Steps in an Analysis

An analysis can contain more than one processing step. You can use additional processing steps when you want to process one set of input files differently from another set of input files. For example, you might want to run a Percolator validation separately on a set of files, such as the replicates of the control group and the replicates of the treatment group.

❖ **To add multiple processing steps**

In an open analysis window, click **Add** on the Child Steps bar.

A new processing step without a workflow appears.

❖ **To duplicate an existing processing step, including its workflow**

In an open analysis window, click **Clone** on the processing step bar.

Another window with a child step appears. You can now define a different processing workflow for the same file or a different file.

Adding or Deleting a Processing Step

You might want to perform only the processing step in a workflow. For example, you only need the processing step if you want to export peak lists as an MGF file with the Spectrum Exporter node. You can remove the Consensus step and re-add it later.

❖ **To perform only the processing step in an analysis**

In the title bar of the Consensus Step box, click the **X** in the upper right corner to remove the consensus step from the analysis.

The Consensus Step box closes, and the Add icon appears in the Analysis title bar.

❖ **To re-add a deleted consensus step in an analysis**

In the Analysis title bar, click ➕ Add .

The Consensus Step box reopens, and the Add icon disappears.

Creating a Processing Workflow for Label-Free Quantification

To create a label-free quantification analysis, you must set up a processing workflow that includes the Minora Feature Detector node.

Default processing workflows are available for several instruments. You can use these default workflows as is or modify them to suit your needs. To access them, see "Using Common Workflow Templates" on page 102. When you use a default workflow, you do not need to perform step 1 through step 6 in the following procedure because these steps are already included in the workflow. However, you must check the node parameters in the workflow to make sure they are appropriate for your data and add modifications to the search for the quantification to be successful.

The default processing workflows for label-free quantification are the following:

- ProcessingWF_QExactive\PWF_QE_Precursor_Quan_and_LFQ_SequestHT_ Percolator.pdProcessingWF

- ProcessingWF_LTQ_Orbitrap\PWF_OT_Precursor_Quan_and_LFQ_CID_SequestHT_ Percolator.pdProcessingWF

- ProcessingWF_Fusion\PWF_Fusion_Precursor_Quan_and_LFQ_IT_HCD_SequestHT_ Percolator.pdProcessingWF

❖ **To create a processing workflow for label-free quantification**

1. In the Workflow Editor, follow the general instructions for creating a processing workflow. See "Creating a Processing Workflow" on page 68.

   The processing workflow must include the following nodes as a minimum:

   • Spectrum Files RC node

   • Spectrum Selector node

   • Minora Feature Detector node

   • A search engine node (for example, Mascot or Sequest HT)

   • Fixed Value PSM Validator node, Target Decoy PSM Validator node, or Percolator node

2. In the Processing Workflow window, drag the **Spectrum Files RC** node to the workspace.

   For information about this node, refer to the Help.

3. Drag the **Spectrum Selector** node to the workspace.

4. Connect the Spectrum Files RC node to the Spectrum Selector node.

5. Drag the appropriate search engine node—for example, **Sequest HT**—to the Workflow Tree pane and attach it to the Spectrum Selector node.

6. Drag the **Fixed Value PSM Validator**, **Target Decoy PSM Validator**, or the **Percolator** node to the Workflow Tree pane and attach it to the search engine node.

   Figure 367 illustrates the basic processing workflow for label-free quantification.

   Figure 367. Basic processing workflow for label-free quantification
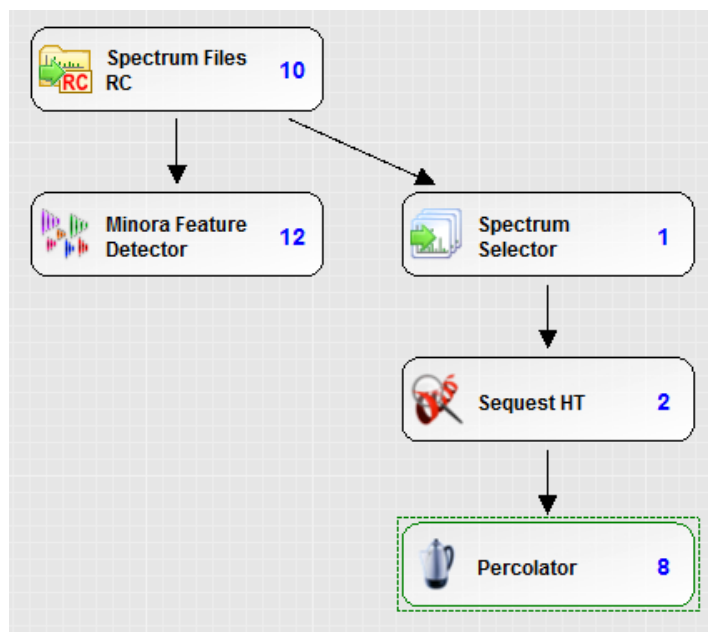
7. (Optional) Add any other appropriate nodes and connect them to the workflow.

8. In the Parameters pane of the Workflow Editor, click **Show Advanced Parameters**.

9. Click the **Spectrum Selector** node, and set the parameters for it in the Parameters pane:

   a. Change the setting in the Max. Precursor Mass box to an appropriate setting. For example, set this option to **6500**.

   b. Change the setting in the S/N Threshold box to an appropriate setting. For example, use the default of **1.5**.

   For optional parameters that you can set for the Spectrum Selector node, refer to the Help.

10. Click the search engine node—for example, **Sequest HT**—and set its parameters in the Parameters pane:

    a. In the Protein Database box, select the **FASTA** database.

    b. In the Dynamic Modifications area, select the dynamic modifications.

    c. In the Static Modifications area, select the static modifications. For example, select **Carbamidomethyl/ +57.021 Da (C)** in the Static Modification box.

    d. Set any other parameters that you prefer. For information on the search engine parameters available, refer to the Help.

11. Click the **Fixed Value PSM Validator** node, the **Percolator** node, or the **Target Decoy PSM Validator** node, and set its Maximum Delta Cn parameter.

    For information on this parameter, refer to the Help.

12. Set the parameters for all other nodes in the Parameters pane.

    For information about all the parameters that you can set for each node, refer to the Help.

13. (Optional) Save the workflow as follows:

    a. In the Name box above the Workflow Tree pane, type a name for the processing workflow.

    b. (Optional) In the Description box above the Workflow Tree pane, type a brief description of the processing workflow.

    c. In the Workflow Editor, click ⊟ Save .

    d. In the Save Workflow dialog box, do the following:

       i. Browse to the file to save the template in, or type a file name in the File Name box.

       ii. In the Save As Type box, select **Processing Workflow File (*.pdProcessingWF)**.

       iii. Click **Save**.

    The application saves the workflow with a .pdProcessingWF file name extension.

Creating a Consensus Workflow for Label-Free Quantification

To create a label-free quantification analysis, you must set up a consensus workflow that includes the Feature Mapper and the Precursor Ions Quantifier node.

Default consensus workflows are available. You can use these default workflows as is or modify them to suit your needs. To access them, see "Using Common Workflow Templates" on page 102. When you use a default workflow, you do not need to perform step 1 through step 9 in the following procedure because these steps are already included in the workflow. However, you must check the node parameters in the workflow to make sure they are appropriate for your data. Be sure the normalization and scaling parameters in the Precursor Ions Quantifier node are set appropriately.

❖ **To create a consensus workflow for label-free quantification**

1. Follow the general instructions for creating a consensus workflow with the Workflow Editor. See "Incorporating an Existing Workflow into a Study and an Analysis" on page 99.

   The consensus workflow must include the following nodes as a minimum:

   - MSF Files node

   - PSM Grouper node

   - Peptide Validator node

   - Peptide And Protein Filter node

   - Protein Scorer node

   - Protein Grouping node

   - Feature Mapper node

   - Precursor Ions Quantifier node

2. In the Consensus Workflow window, drag the **MSF Files** node to the workspace.

3. Drag the **PSM Grouper** node to the workspace, and connect it to the MSF Files node.

4. Drag the **Peptide Validator** node to the workspace, and connect it to the PSM Grouper node.

5. Drag the **Peptide And Protein Filter** node to the workspace, and connect it to the Peptide Validator node.

6. Drag the **Protein Scorer** node to the workspace, and connect it to the Peptide And Protein Filter node.

7. Drag the **Protein Grouping** node to the workspace, and connect it to the Protein Scorer node.

8. Drag the **Feature Mapper** node to the workspace, and connect it to the MSF Files node.

9. Drag the **Precursor Ions Quantifier** node to the workspace, and connect it to the Feature Mapper node.

10. (Optional) Drag the **Data Distributions** node to the Post-Processing Nodes pane.

   This node adds heat maps, which aid in data review.

   For information on all these nodes, refer to the Help.

   Figure 368 illustrates the basic consensus workflow for precursor ion quantification.

   Figure 368. Basic consensus workflow for label-free quantification

11. (Optional) Add any other appropriate nodes and connect all the nodes together.

12. Set the parameters for each node.

   For information about all the parameters that you can set for each node, refer to the Help.

13. (Optional) Save the workflow:

   a. In the Name box above the Workflow Tree pane, type a name for the consensus workflow.

   b. (Optional) In the Description box above the Workflow Tree pane, type a brief description of the consensus workflow.

    c.   In the Workflow Editor, click   🖫 Save .

    d.   In the Save Workflow dialog box, do the following:

        i.   Browse to the file to save the template in, or type a file name in the File Name box.

        ii.   In the Save As Type box, select **ConsensusWorkflow File (*.pdConsensusWF)**.

        iii.   Click **Save**.

    The application saves the workflow in the *file_name*.pdConsensusWF file.

14.   Save the analysis. See "Saving an Analysis" on page 83.

15.   Save the study. See "Saving a Study" on page 56.

16.   Click   🔁 Run   in the upper right corner of the Analysis window.

    The job queue appears, as shown in Figure 53 on page 83, displaying the status of your search. Use the job queue to check the status of your search as the search progresses. For information about the job queue, refer to the Help.

Saving an Analysis

To use an analysis as a template for later reuse, you can save it as a .pdAnalysis template file.

> **Note** Studies and analyses in the Proteome Discoverer are separate, so you must save them separately. Saving a study does not save an analysis, and saving an analysis does not save a study.

❖ **To save an analysis as a template for later reuse**

1.   In the upper right corner of the Analysis window, click   💾 Save .

2.   In the Save Analysis Template dialog box, browse to the location where you want to store the template.

3.   In the File Name box, browse to the study folder, and type or select the template file name.

4.   In the Save as Type box, select **Analysis Templates (*pdAnalysis)**.

5.   Click **Save**.

    The application saves the analysis in a file with a .pdAnalysis extension.

    The .pdAnalysis template file saves the processing and consensus workflows. It saves neither the input files nor the study variables that were selected to group the samples and quantification ratios.

Specifying Quantification Ratios from Selected Sample Groups for Label-Free Quantification

After you set up the workflow to use for the analysis, you can specify the ratios to report for the quantification and how to group your samples in relation to the specified factor values.

You can generate arbitrary ratios from selected sample group abundances without being restricted to predefined sample and control specifications. Previous versions of the Proteome Discoverer application automatically generated quantification ratios that relied on the specification of sample or control sample types. This feature is appropriate for some experiments but too restrictive for experiments in which the same sample should occur in the denominator for some ratios and in the numerator for others.

For example, in a typical mutant-versus-wild-type experiment, you usually would monitor ratios like this to study the effect of the treatment on a mutant:

$$\frac{wt\_treat}{wt}, \frac{m\_treat}{m}, \frac{m\_treat}{wt\_treat}$$

where:

- *wt_treat* is a treated wild type.

- *wt* is a wild type.

- *m_treat* is a treated mutant.

- *m* is a mutant.

Also, in some experiments there is no clear control available, and all pair-wise comparisons of available samples are of interest.

On the Grouping & Quantification page of a study, you can generate custom quantification ratios that do not restrict you to predefined sample and control specifications.

In the Grouping & Quantification view, you first select the study variables that were used to group your samples and select the numerators and denominators of your ratios. Then, you manually specify the quantification ratios to generate or semiautomatically create all possible quantification ratios against an ad-hoc selection of denominator values.

❖ **To open the Grouping & Quantification page**

On the Study: *Study_name* page, click the **Grouping & Quantification** tab.

> **Note** The Grouping & Quantification tab does not appear until you add input files to the analysis.

Figure 369 shows the Grouping & Quantification page of the study.

Figure 369. Grouping and Quantification page
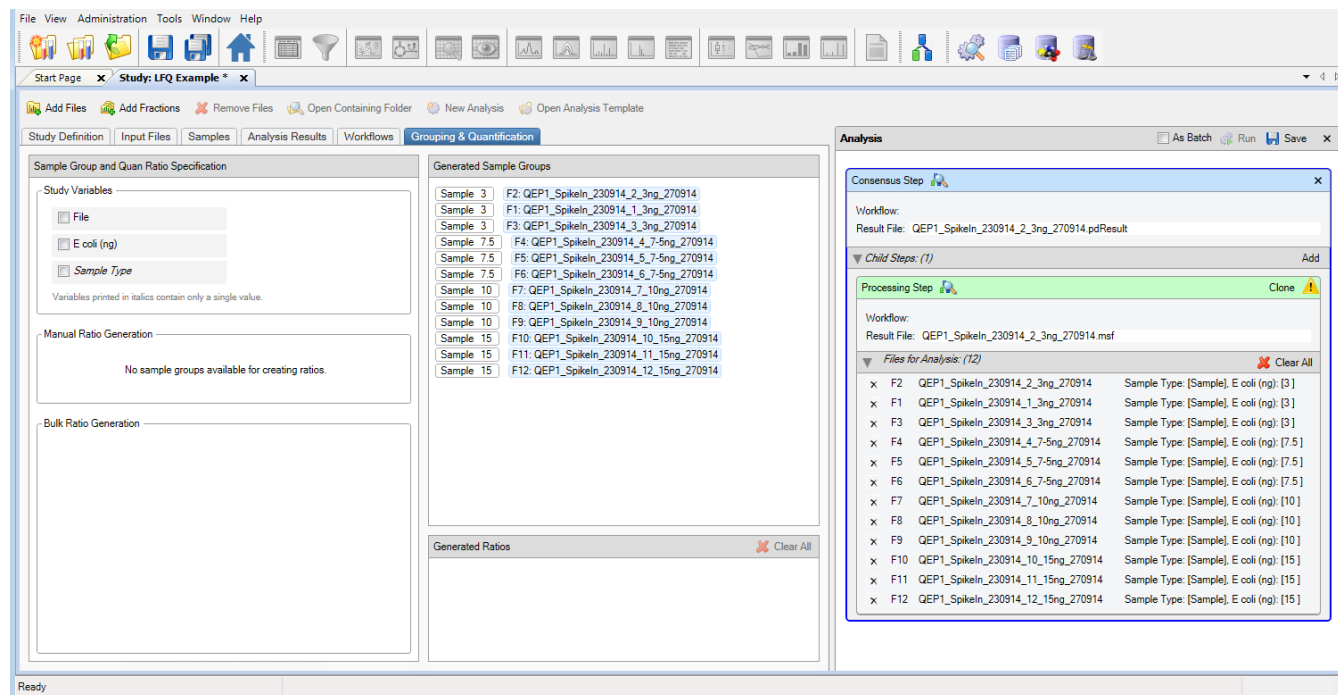
The page contains the following areas:

- Study Variables area: At the top left is a list of the selectable study factors, or study variables, that you specified in the Study Factors area of the Study Definition page. A study factor is anything that captures the difference between two or more samples, for example, drug treatment, the time of drug application, or differences in tissue, organisms, or patients. It can be differences in sample preparation, chromatography settings, or acquisition parameters; or differences in the isobaric or metabolic labels used. Study factors can vary from sample to sample. In the example used throughout this topic, the study factors are the concentration of *E. coli* proteins, in nanograms. Selecting the study factors to use for grouping provides an effective means of sorting the samples and quantification ratios into replicate and treatment groups.

  - File: Groups samples and quantification ratios by files.

  - *Study_factors*: Groups samples and quantification ratios by user-defined study factors.

  - Sample Types: Groups factors that include Sample, Control, Standard, and Blank. Variables displayed in italics contain only a single value.

  Study factors are user-defined, but the File and Sample Type variables always appear on the Grouping & Quantification page for all studies.

- Manual Ratio Generation area: Contains menus where you can select the numerator and denominator for each quantification ratio.

- Bulk Ratio Generation area: Displays the study factor values to use as the denominators for semiautomatically generated quantification ratios.

- Generated Sample Groups area: Displays samples grouped by the values set for the selected study factors. The application ignores any other differences among the samples that might be present.

  Selecting these variables indicates that you want to group your samples and quantification ratios according to the specified study variables and ignore any other difference between the samples that might be present. For example, if you group by the concentration of *E. coli* proteins, you put into the same sample group all replicates with the same concentration. You also group all quantification ratios together that have the same concentration in the numerator or denominator.

  The grouping of samples affects how identifications are displayed in the distribution maps (see "Sample Information Used to Display Identifications and Quantifications Among Files and Samples" on page 578) and the calculation of areas from precursor ion quantification. For the latter, the application reports separate area values for each sample or sample group.

  When you group ratios, the application calculates an averaged ratio for a group of ratios that are similar with respect to the selected study variable in the numerator and denominator.

  > **Note** The application first groups samples to average abundance values from replicate samples and then calculates ratios from the grouped abundances. The ratios are freely defined in the Grouping and Quantification view, without being restricted to predefined control samples.

  If the application does not use all available sample group values in at least one ratio, it displays a warning in the Generated Sample Groups area, like that shown in Figure 370, and marks the unused sample groups.

  Figure 370. Warning in the Generated Sample Groups area
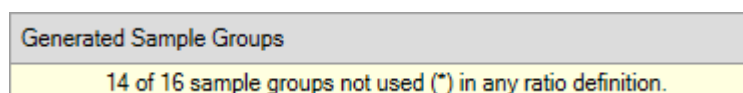
  Generated Sample Groups
  14 of 16 sample groups not used (*) in any ratio definition.

- Generated Ratios area: Displays the ratios generated.

❖ **To generate custom quantification ratios semiautomatically**

1. In the Study Variables area, select the check box of the study factors, or variables, that you want to use to group your samples and from which you want to draw the numerators and denominators of the ratios. For the example used throughout this topic, select the **E. coli (ng)** check box to indicate that the samples and quantification ratios are grouped by the values set for the *E. coli* (ng) factor.

   After you select the study factors, the Generated Sample Groups area displays the generated sample groups. When performing the quantification, the application calculates abundance values for each sample and averages the abundance values of all samples in a sample group.

This step creates sample groups using the selectable study factors available for the files to be analyzed. The application places all samples that have the same combination of study factor values in the same group. In the example shown in Figure 375 on page 504, the application groups the samples by the concentration of *E. coli* (ng) proteins.

The order of study factors is relevant for the semiautomatic generation of ratios.

2. (Optional) To change the placement of a study factor in the list of study factors, do the following:

    a. In the Study Variables area, select the check box for a study factor.

        A placement handle in the form of a green rectangle appears to the left of the selected check box, as shown in Figure 371.

        Figure 371. Placement handle in the Study Variables area

    b. Hold the cursor over the placement handle.

        White up and down arrows now appear on the handle.

    c. Drag the cursor up or down to move the variable to its new place in the list of variables, or click the up or down arrows to move the study factor.

3. (Optional) To sort study factors in the Bulk Ratio Generation area and sample groups in the Generated Sample Groups area, click one of the following to the right of each study factor in the Study Variable area.

    *   To sort these items in descending order, click the **Sort Descending** icon, ![icon].
    *   To sort these items in ascending order, click the **Sort Ascending icon**, ![icon].
    *   To leave these items unsorted, click the **No Sorting** icon, ![icon].

4. In the Bulk Ratio Generation area, select the check box for the concentration of *E. coli* protein to use in the denominator of the ratio. For the example, select **E. coli (ng): 10.**

    The Bulk Ratio Generation area displays a list of the denominator values for this type of study factor. If you select only one study factor, it displays a list of the available denominator values for this factor, as shown in Figure 372.

Figure 372. Denominator values available for a single study factor
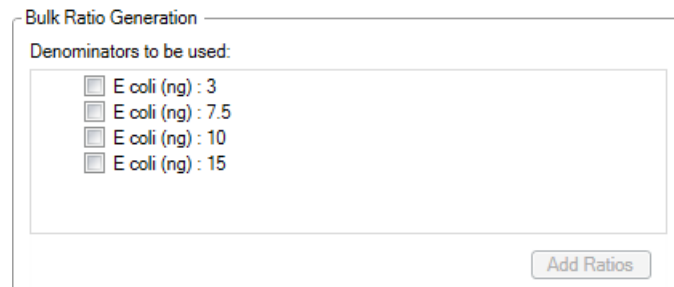
If you select multiple study factors, the Bulk Ratio Generation area displays the denominator values available for each factor.

5. (Optional when there are multiple study factors) To select the same study factor for all the denominators, do the following:

   a. Hold the cursor over a denominator value.

      An icon containing four check boxes in a square appears on the left side of that item, as shown in Figure 373.

   b. Click the icon.

   The application selects the same study factor for all denominators.

Figure 373. Icon for multiple selection of denominators in the Bulk Ratio Generation area
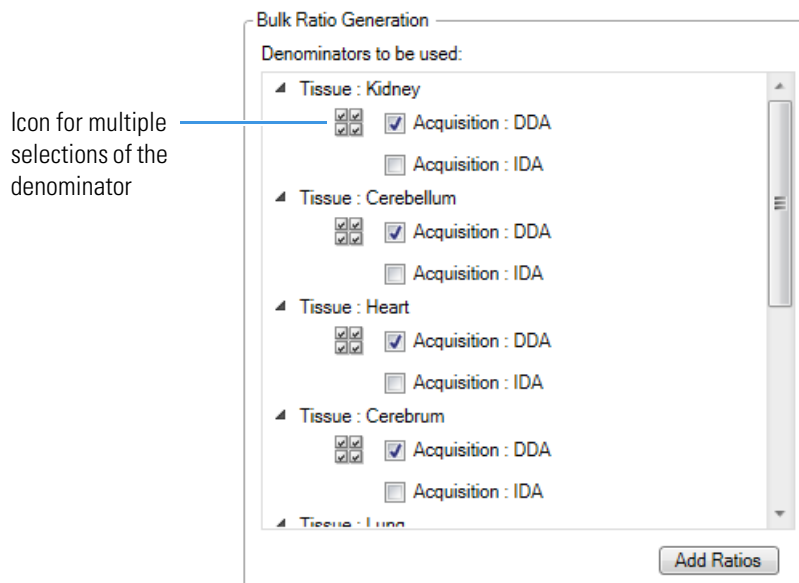
6. Click **Add Ratios**.

   The application generates all possible ratios against the selected denominator values and adds them to the Generated Ratios area. Figure 374 shows the generated quantification ratios and ratio groups in the Generate Ratios area after the selection of *E. coli* (ng) as the study variable to group by and 10 as the denominator to use.

Figure 374. Quantification ratios and ratio groups generated by grouping the samples and ratios by concentration of *E. Coli* (ng) proteins
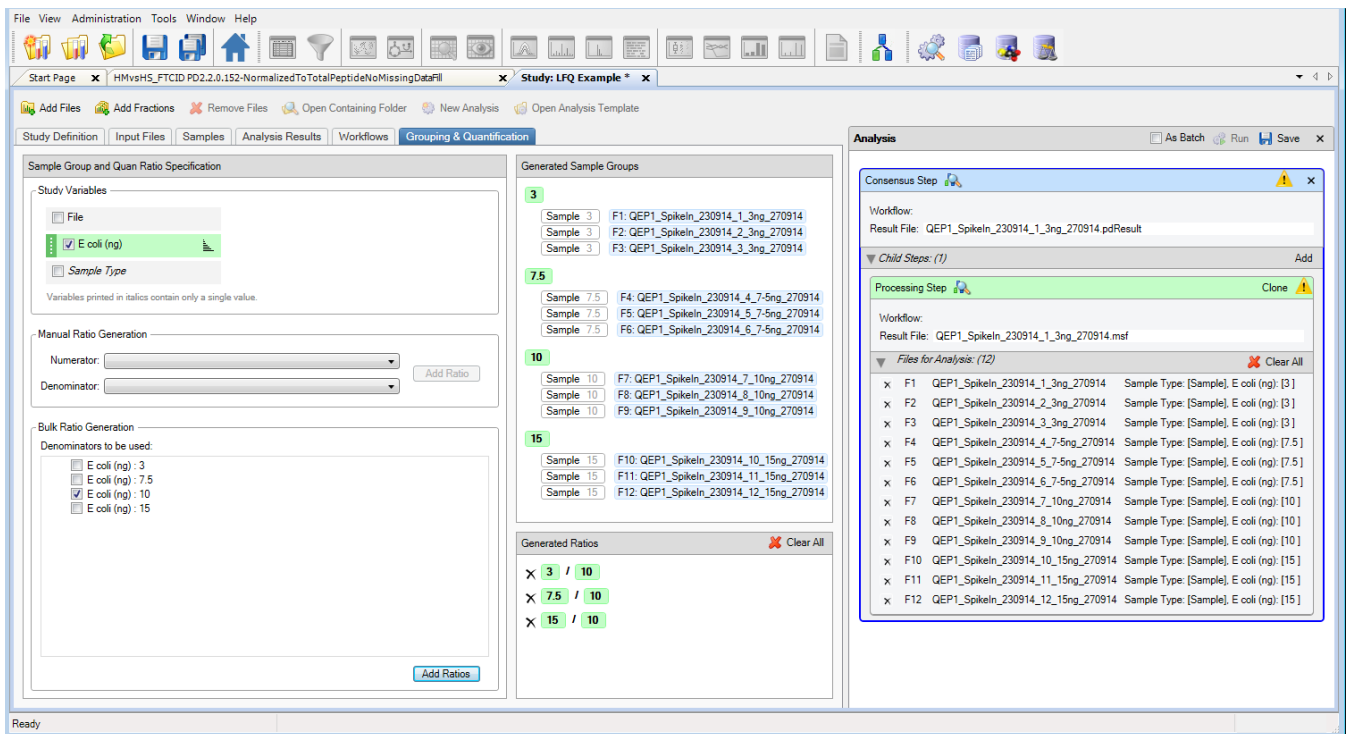
7. (Optional) Select another variable or variables. In the example, select the **File** check box in the Study Variables area.

    When you add a second variable, the information in the Generated Ratios area becomes invalid and appears in a gold color.

8. In the Generate Ratios area, click ❌ Clear All to delete the previous ratios.

9. In the Bulk Ratio Generation area, select the check box for the type of tissue to use in the denominator. For the example, select **E coli (ng): 3.**

10. Click **Add Ratios**.

    Figure 375 shows the generated quantification ratios and ratio groups on the Generated Ratio Groups page after selecting *E. coli* (ng) and File as the study variables to group by.

Figure 375. Quantification ratios and ratio groups generated by grouping the samples and ratios by E. coli (ng) and File
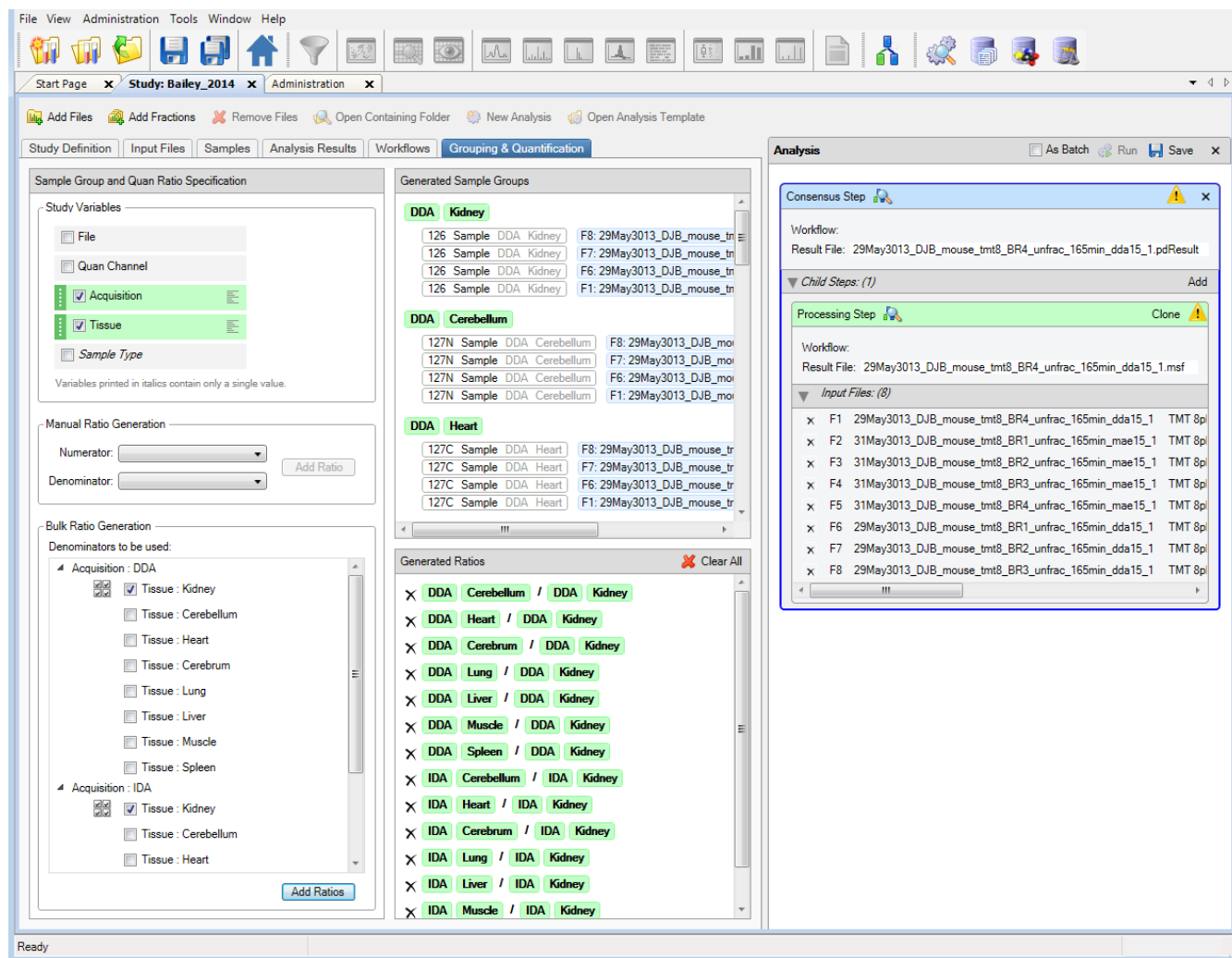
❖ **To generate custom quantification ratios manually**

1. On the Study: *Study_name* page, click the **Grouping & Quantification** tab.

> **Note** The Grouping & Quantification tab does not appear until you add or open an analysis and specify Input Files for the analysis.

2. In the Study Variables area, select the check box of the study factor or factors that you want to use to group your samples and from which you want to draw the numerators and denominators of the ratios. For the example used throughout this chapter, select the **E. coli (ng)** check box to indicate that the samples and quantification ratios are grouped by the values set for the *E. coli (ng)* factor.

3. In the Manual Ratio Generation area, select the numerator value from the Numerator list.

4. Select the denominator value from the Denominator list.

5. Click **Add Ratio**.

The generated ratio appears in the Generated Ratios area, as shown in Figure 376.

Figure 376. Ratio manually generated for one study factor
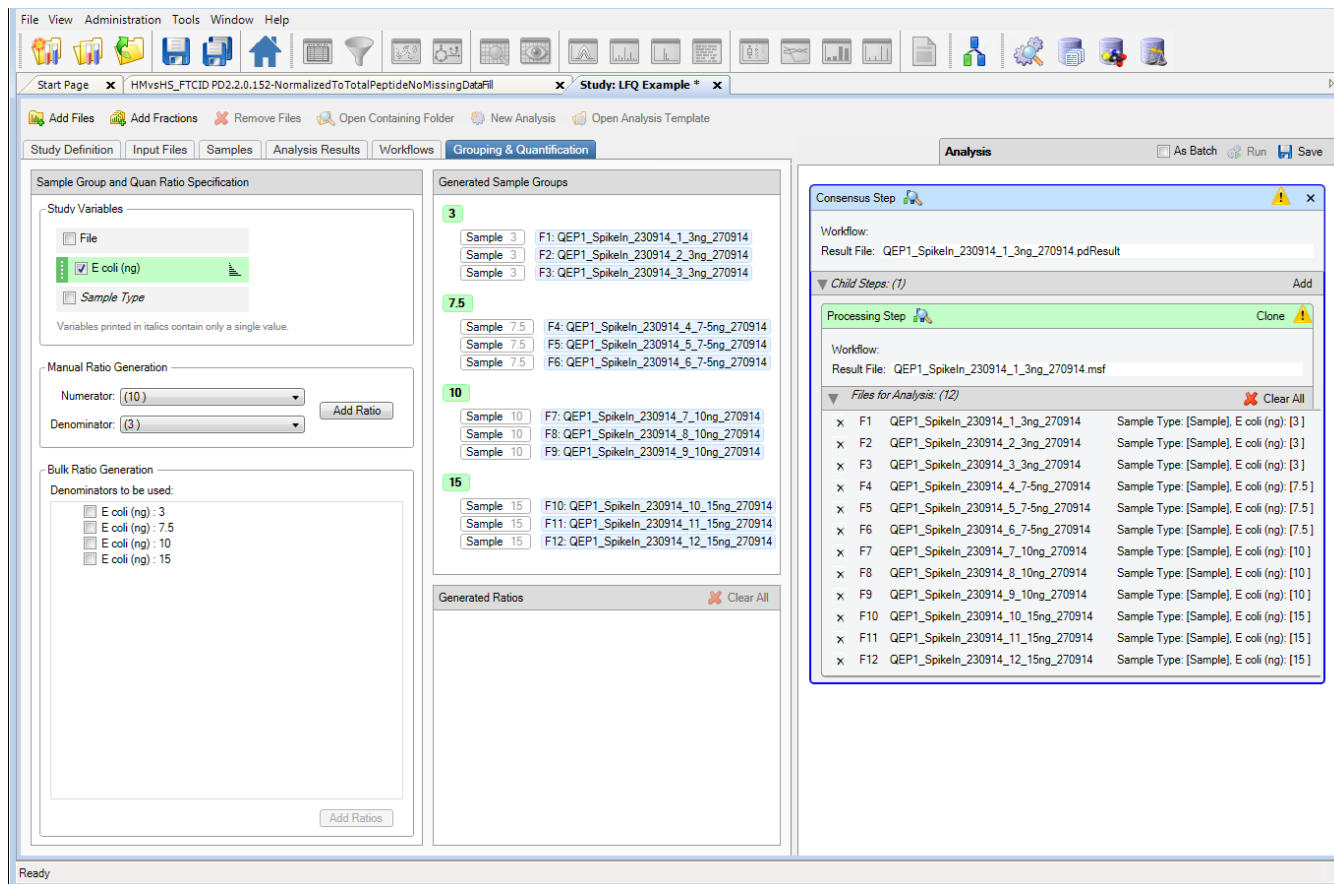
6. (Optional) Select another study factor or factors in the Study Variables area, for example, **File**.

When you add a second factor, the information in the Generated Ratios area becomes invalid and appears in a gold color.

7. In the Generated Ratios area, click ✖ Clear All to delete the previous ratios.

8. In the Manual Ratio Generation area, select the numerator value from the Numerator list, for example, **(10)**.

9. Select the denominator value from the Denominator list, for example, **(3)**.

10. Click **Add Ratio**.

Figure 377 shows the ratios generated in the Generated Ratios area.

Figure 377. Ratios manually generated for two study factors

❖ **To save the settings on the Grouping & Quantification page**

> **Note** The application does not save the settings on the Grouping & Quantification page with a study or with an analysis. Instead, it associates the settings with search results, so you must load them from data sets that have already been processed within the study or recreate them from the beginning.

1. Click the **Analysis Results** tab of the study.

2. Select the result on the Analysis Results page, and choose either **Reprocess > All Analysis Steps** or **Reprocess > Last Consensus Step**.

   If you choose Use Results to Make New (Multi) Consensus, the Grouping and Quantification page no longer displays ratios.

Performing the Quantification

❖ **To perform the quantification**

Click   Run   .

The job queue appears, as shown in Figure 49 on page 78 for the processing workflow and in Figure 53 on page 83 for the consensus workflow, displaying the status of your search. Use the job queue to check the status of your search as the search progresses. For information about the job queue, refer to the Help.

The application creates one or more MSF and .pdResult files, depending on whether the data files to be analyzed are batch-processed. It might reuse MSF files to produce a new .pdResult file if a different report for the given data is required.

The .pdResult file has the same name as the input file but with a .pdResult extension. The Analysis Results page displays the name of the .pdResult file.

Working with the Search Results

To open, delete, convert, and reprocess search results and to view the workflow and the analysis after the search, see "Working with the Search Results" on page 89.

Grouping Similar Files

Before you start an analysis, you can use the Grouping & Quantification page in the study to define sample groups of files that the Feature Mapper node should group, or align, together. You usually group files having equal properties, such as replicate files.

First, the aligner node uses the grouping information to align files within a sample group because the files are more similar than the files from different sample groups. Next, the node aligns the reference files of the just-aligned sample groups to each other. Then the node applies the calculated alignments to all files of the sample group.

For the example shown in this chapter, these samples are grouped in the manner shown in Figure 378.

Figure 378.

For these data, the application first separately aligns each of the sample groups based on the *E. coli* (ng) study factor. It first independently aligns the three files found for the sample group *E. Coli* (ng): 3. Then it independently aligns the three files for *E. coli* (ng): 7.5, *E. coli* (ng): 10, and *E. coli* (ng): 15. Subsequently, it aligns the selected reference files for each sample group to each other. The selection of the reference file depends on the alignment node. The Feature Mapper node takes the file with the most features.

Calculating and Validating Raw Quantification Values in Label-Free Quantification

The Precursor Ions Quantifier node in the Consensus Workflow first validates the LC/MS features. It determines whether they contain valid values or consist only of a single peak. In addition, it marks them as shared when they are connected to multiple PSMs. The node creates a Quan Info column for the LCMS Features page that displays the following values:

- Rejected by Method: Indicates that the peptide was not used in the quantification.

The node also creates a Quan Result ID column for the Peptide Groups page, showing an integer number. Peptide groups belonging to the same quantification result with different states have the same quantification result identifications. The node also ensures that all consensus features of a quantification result are connected to all the different peptide groups of this quantification result.

In the next step, the Precursor Ions Quantifier node validates consensus features and calculates the quantification values for the different files. It checks which LC/MS features can be used and excludes invalid features. It creates two columns, Quan Info and Quan Info Details, for the Consensus Features page, showing the following values. (The Quan Info Details column is hidden and only shows information about why a quantification result is labeled Rejected by Method.)

- No Quan Values: Indicates that no valid features are associated with a consensus feature.

- Unknown Quan Result: Indicates that no quantification result identification could be assigned.

- Shared: Indicates that multiple peptide groups are associated with a consensus feature.

A consensus feature shows abundances for a retention-time and *m/z* value over the different spectrum files used in the processing step. For calculating abundances, the Precursor Quantifier node sums areas of associated LC/MS features for the different files.

Missing Replicate Features

The Min. # Replicate Features [%] parameter of the Precursor Ions Quantifier node defines the minimum percentage of replicate files that a feature must be detected in to be used. Its default value is 0. This parameter affects consensus features in label-free quantification. It is not supported for labeled precursor quantification, because it can be hard to identify the replicate files.

When you set this parameter to a value greater than zero, the node determines which files belong to one replicate. Which files that belong to a replicate group depends on your study setup. If you use a biological replicate factor and multiple files are grouped into one biological replicate group, the files in one biological replicate group belong to one replicate group. If you do not have a biological replicate factor, the files of a sample group belong to the one replicate group.

When the Precursor Ions Quantifier node calculates the abundances of a consensus feature, it determines whether quantification values were found for enough replicate files. If there are not enough files, it excludes the LC/MS features of the replicate group from quantification. It displays Rejected by Method in the Quan Info column, Missing Replicates in the Quan Info Details column, and Not Used in the Quan Usage column.

The filter checks the LC/MS features by replicate group. If one replicate group does not have not enough quantification values, the node can still use the LC/MS features from other replicate groups when there are enough features.

Quantification Methodologies

This chapter explains what you can do to ensure that your quantification results are optimal and can be easily interpreted. It also describes aspects of quantification that are common to reporter ion, label-free, and precursor ion quantification.

Contents

- Generating Quantification Ratios
- Calculating P-Values and Adjusted P-Values for Quantification Results
- Imputing Missing Values
- Treating Missing Quantification Channels for Quantification Methods
- Calculating Quantification Results
- Using a Quantification Method
- Searching for Quantification Modifications with Mascot
- Troubleshooting Quantification

Generating Quantification Ratios

The following topics use examples that are appropriate for the three types of quantification, and you use the same procedure for each type to specify the quantification ratios.

- For reporter ion quantification, see "Specifying Quantification Ratios from Selected Sample Groups for Reporter Ion Quantification" on page 393.
- For label-free quantification, see "Specifying Quantification Ratios from Selected Sample Groups for Label-Free Quantification" on page 498.
- For precursor ion quantification, see "Specifying Quantification Ratios from Selected Sample Groups for Precursor Ion Quantification" on page 456.

Calculating P-Values and Adjusted P-Values for Quantification Results

The previous chapters describe how to set up reporter ion quantification, precursor ion quantification, and label-free quantification workflows. After the analyses are complete, a final report presents the list of peptide groups and proteins with scaled abundances and selected ratios. The Proteome Discoverer application includes a feature for assessing the significance of differential expression by providing p-values and adjusted p-values for those ratios selected on the Grouping and Quantification page prior to the analysis.

The p-value displayed for these selected ratios for a given peptide group or protein is a measurement of how likely the abundance is unchanged between samples. A larger p-value indicates that the abundances are the same; a lower p-value indicates that the abundances are more likely to be significantly different. A p-value threshold of 0.05 or 0.01 is typically applied as a threshold to determine whether a given peptide group or protein has significantly different abundances. The probability of a false-positive detection for one or more of the ratios increases with an increasing number of measurements. This error rate is managed by adjusting the p-value to lower significance based on the number of measurements using a correction such as the Benjamini-Hochberg method.

There are four different methods to calculate p-values for selected ratios:

- ANOVA (Background-Based): To use this method, select it in the Precursor Ions Quantifier or Reporter Ions Quantifier nodes in the consensus workflow. This method uses the background population of ratios for all peptides and proteins to determine whether any given a single peptide or protein is significantly changing relative to that background. This method does not require that a given sample group contains replicates, but it does require that most of the protein abundances are unchanged between samples. If this is not the case or if there are fewer than 500 protein IDs in the sample, you must use the ANOVA (Background-Based) method. To set up replicates samples, see "Calculating P-Values for Replicate Data by Using Biological Replicate Study Factors" on page 513. The ANOVA (Individual Proteins) method also calculates an adjusted p-value using the Benjamini-Hochberg method.

- ANOVA (Individual Proteins): This method performs an ANOVA for all conditions (or a t-test when there are only two conditions) and requires that there are replicates associated with the study factors used to create the ratio. This method also calculates an adjusted p-value using the Benjamini-Hochberg method.

- Biological Replicates study factor using a Non-Nested Design: When the biological replicates are set up using a non-nested (or unpaired) design, the application uses either the ANOVA (Background-Based) or the ANOVA (Individual Proteins) method to calculate the p-values and adjusted p-values. To set up biological replicate study factors for a non-nested design, see "Calculating P-Values for Replicate Data by Using Biological Replicate Study Factors" on page 513.

- Biological replicates study factors using a Nested Design: For a nested or paired design, the biological replicates are set up such that the measured effect is observed for the same individual or sample under multiple conditions. The application uses either the Paired (Background-Based) or the ANOVA (Individual Proteins) method to calculate the p-values and adjusted p-values. To set up a nested design experiment using biological replicate study factors, see "Calculating P-Values for Replicate Data Without Using Biological Replicate Study Factors" on page 519.

❖ **To determine whether a study contains replicates**

1. Click the **Analysis Results** tab in the study.

2. Right-click the result that you are interested in, and choose **Reprocess > All Analysis Steps**.

3. Click the **Grouping & Quantification** tab.

4. On the Grouping & Quantification page, view the Generated Sample Groups area.

   See Figure 381 on page 516 or Figure 384 on page 519 for examples of studies with replicate samples. There must be more than one file or quantification channel under each study factor or combination of study factors in the Generated Sample Groups pane.

For more information on the Grouping & Quantification page, see "Specifying Quantification Ratios from Selected Sample Groups for Reporter Ion Quantification" on page 393.

Calculating P-Values for Replicate Data by Using Biological Replicate Study Factors

Use the biological replicate study factor for experiments where there are both technical and biological replicates in the same experiment. You can also use the biological replicate study factor to create a nested design study. Biological replicate study factors help you specify how to calculate p-values. You can set up these replicates as nested designs or non-nested designs.

Nested Designs

Nested design involves measurements taken repeatedly from the same biological source across multiple time points—for example, drawing blood from the same patients for several consecutive days. Using nested design is potentially a more sensitive way to spot trends.

❖ **To create a nested design**

1. On the Study Definition page, create the study factors to assign to the quantification channels (see Figure 379).

   - For reporter ion quantification, see "Adding the Study Factors" on page 361.

   - For label-free quantification, see "Adding the Study Factors" on page 474.

   - For precursor ion quantification, see "Adding the Study Factors" on page 428.

The example in Figure 379 assigns study factors 1, 2, and 3 to the quantification channels.

Figure 379. Study factors assigned to the quantification channels
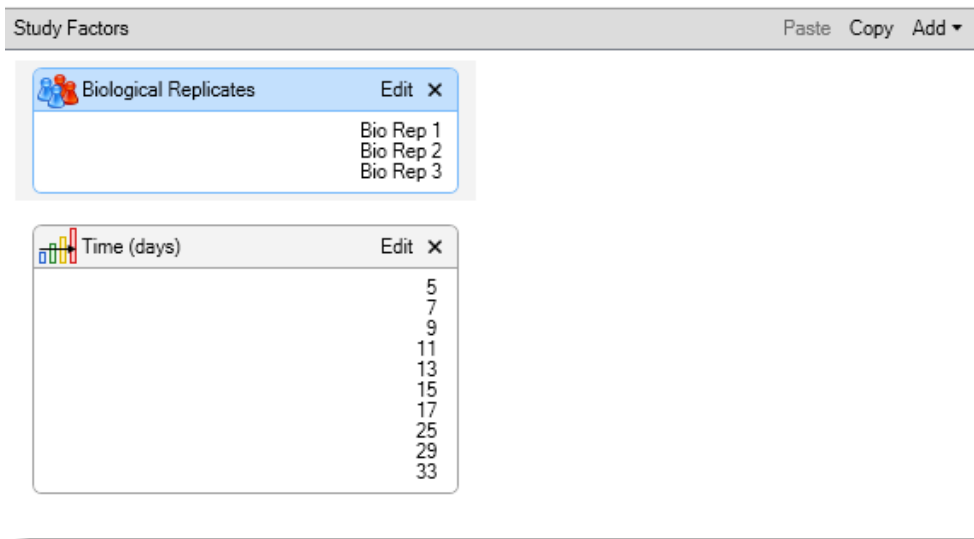
2. On the Input Files page, assign the study factors to the quantification channels:

   a. Click the gray arrow to the left of the first replicate row.

   b. Select all the quantification channels used for a given replicate.

   c. Right-click the Replicate column, and choose **Set Replicate To >** ***study_factor_number*** (see Figure 380).

   The example in Figure 380 assigns study factor Bio Rep 2 to the replicate 2 dataset.

Figure 380. Assignment of study factors to quantification channels in a nested design
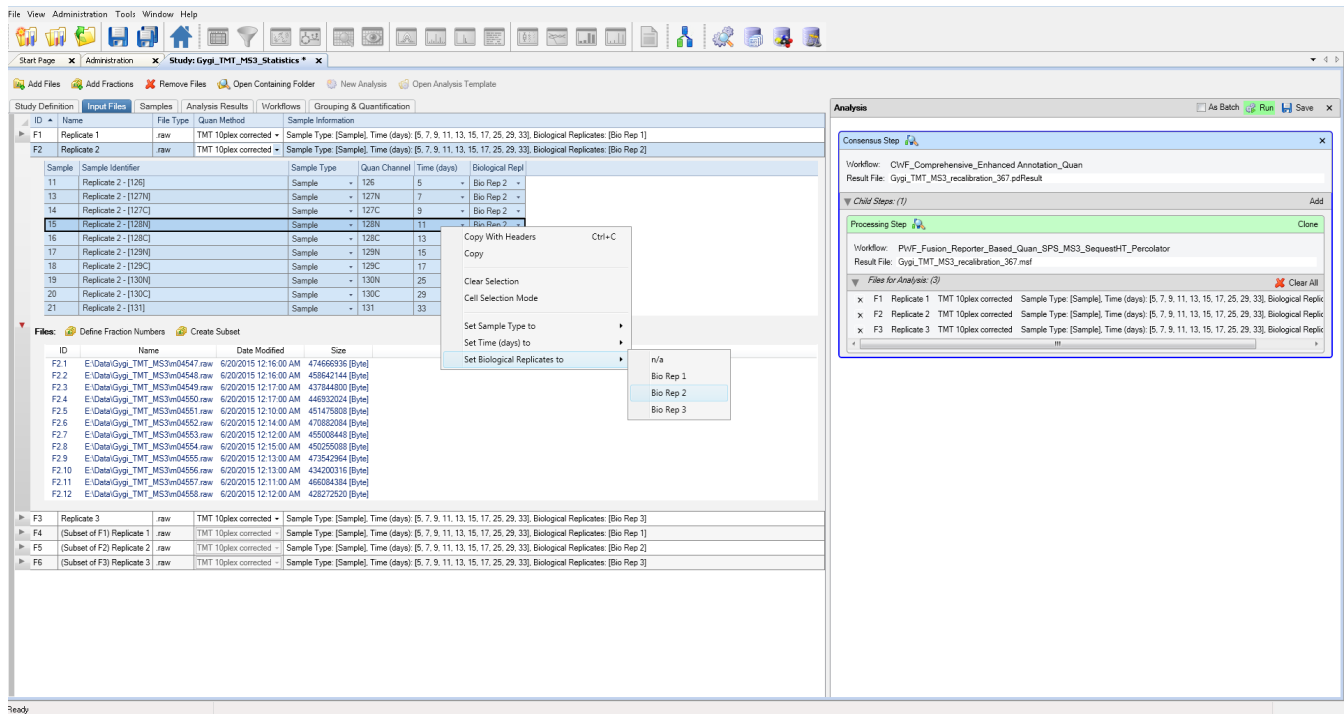
3. Click the **Grouping & Quantification** tab.

The Generate Sample Groups pane in Figure 381 shows three biological replicates that are being remeasured multiple times. The Generated Ratios pane of the Grouping & Quantification page confirms that the design is nested.

Figure 381. Generated Sample Groups page showing replicates in a nested design

Non-Nested Designs

Non-nested design involves measurements taken from different biological sources or taken at different times—for example, comparing healthy patients to diseased patients. They are not taken from the same sample or samples repeatedly over a course of time.

❖ **To create a non-nested design**

1. On the Study Definition page, create the study factors to assign to the quantification channels (see Figure 382).

   • For reporter ion quantification, see "Adding the Study Factors" on page 361.

   • For label-free quantification, see "Adding the Study Factors" on page 474.

   • For precursor ion quantification, see "Adding the Study Factors" on page 428.

The example in Figure 382 assigns study factors 1 through 30 to the quantification channels.

Figure 382. Study factors assigned to the quantification channels
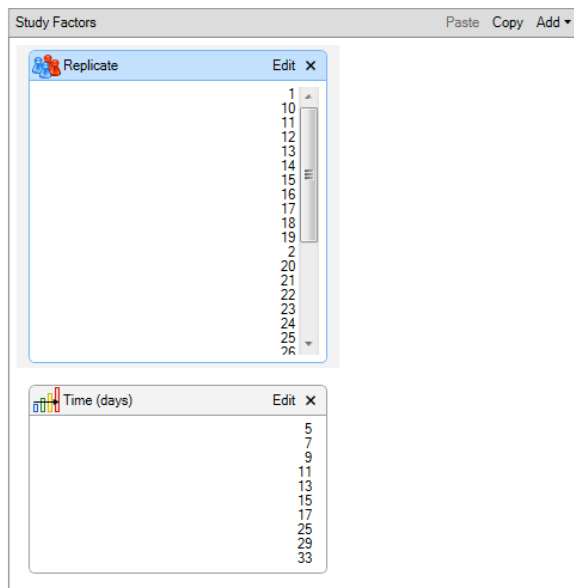
2. On the Input Files page, assign the study factors to the quantification channels.

    a. Click the gray arrow to the left of the first replicate row.

    b. For each quantification channel, select a unique replicate number in the Replicate column.

    The example in Figure 383 assigns replicate factor 10 to quantification channel 131.

Figure 383. Assignment of study factors to quantification channels in non-nested design

3. Click the **Grouping & Quantification** tab.

As shown in Figure 384, the Generated Sample Groups pane channels (that are grouped by one or more selected study factors) and the Generated Ratios pane of the Grouping & Quantification page confirm that the design is non-nested.

Figure 384. Generated Sample Groups pane showing replicates in a non-nested design
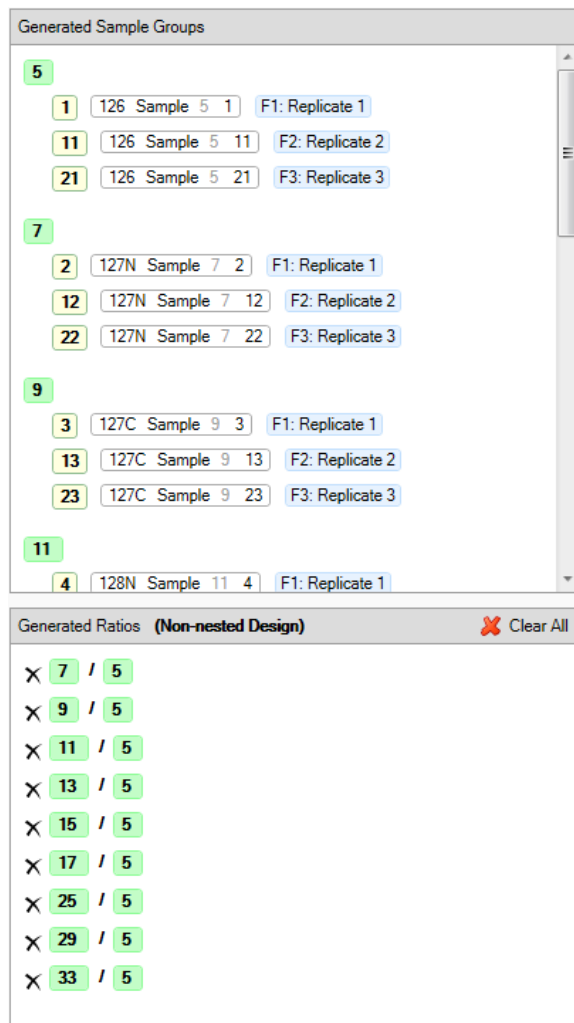
Calculating P-Values for Replicate Data Without Using Biological Replicate Study Factors

If a study factor includes multiple quantification channels, the application calculates the p-value for ratios that use those study factors the same way that it calculates the p-value for a non-nested biological replicate design (see "Non-Nested Designs" on page 516). The advantage of this type of calculation is that it does not require you to create study factors.

When you do not use study factors, the resulting Generated Sample Groups pane and the Generated Ratios pane on the Grouping & Quantification page resemble the panes in Figure 385.

Figure 385. Generated Sample groups pane and Generated Ratios pane when using study factors
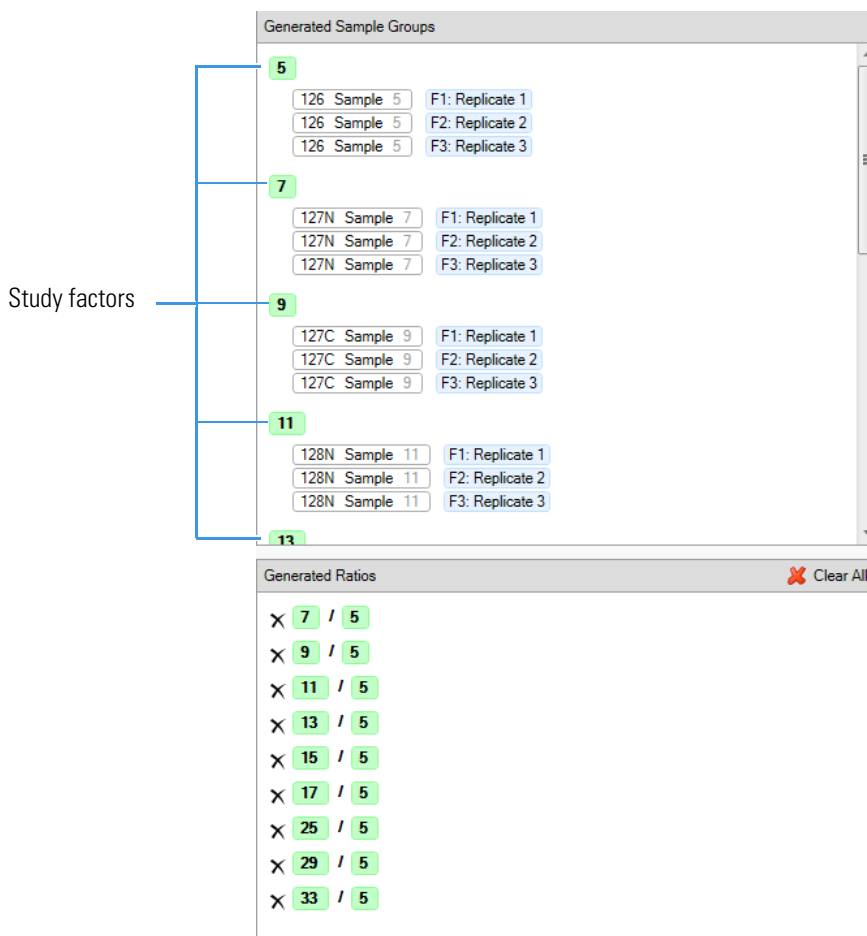
Ensuring That P-Values Are Calculated

When you use the ANOVA (Individual Proteins) p-value calculation method, to ensure that the application calculates p-values for ratios that include those study factors, you must group quantification channels underneath a study factor; see the Generated Sample Groups pane in Figure 385. In this example, the green boxes show study factors 5, 7, 9, and so forth.

Imputing Missing Values

To analyze the statistical significance of changes observed in abundances of proteins and peptides, you usually apply statistical tests, like t-tests or ANOVA, to your matrix of abundance values. This matrix traditionally shows the peptides or proteins as columns and the samples as rows. Most statistical tests have issues when values are missing in the data, so you usually try to remove these gaps in the data before further statistical analysis by either removing variables with missing values for some observations or trying to fill the gaps with some sensible values. Filling the gaps with reasonable values is called *missing value imputation*.

You can use a number of strategies to impute missing values. Imputed values are artificial and are never optimal, but the following recommendations apply:

- That they are not all the same constant value but exhibit some kind of randomness

- That they do not lower the variance of the detected values

- That, at best, they do not create significance that would not exist without them being imputed

To impute missing values, set the Imputation Mode parameter in the Precursor Ions Quantifier node or the Reporter Ions Quantifier node in the consensus workflow to one of the following settings. These nodes specify how to treat missing values.

- (Default) None: Does not impute any missing values.

- Replicate-Based Resampling: Uses the detected values from replicate measurements. For each sample group, this method first uses a linear model to assess the dependence of the standard deviation of the replicate measurements of a protein or peptide on the median abundance of the protein or peptide. Then for each protein or peptide with missing values, it uses the median of the detected values and draws random values from a normal distribution around this median with a standard deviation derived in the first step. When it does not detect any values in a sample group for a protein or peptide, it imputes low abundance values by using the already available method random sampling from the distribution of the lower fifth percentile of abundance values.

- Low Abundance Resampling: Replaces missing values with random values sampled between the minimum and the lower 5 percent of all detected values.

The application applies the missing value imputation before the optional scaling of abundance values and after the optional normalization of abundance values. If the application performs normalization, the raw abundances still contain gaps but it fills the gaps on the normalized abundances level according to the chosen imputation method.

Imputation has little to no effect when you use the pairwise ratio-based method. Use this method only when you select the summed abundance ratio method in the Precursor Ions Quantifier or the Reporter Ions Quantifier nodes in the consensus workflow. When it imputes missing values, it creates a new Abundance Origin column on the Proteins and Peptide Groups page with information about where an abundance value came from.

- Det: The value is derived from a detected signal.

- Imp: The value was imputed.

Treating Missing Quantification Channels for Quantification Methods

If quantification channels are missing, you can use the parameter settings of the Reporter Ions Quantifier node or the Precursor Ions Quantifier node in the consensus workflow to determine how the application handles this problem. Generally, imputation handles missing quantification channels, but there are options in the Reporter Ions Quantifier node and the Precursor Ions Quantifier node that you can use to include or ignore PSMs, peptide groups, or proteins that do not measure all expected quantification channels.

You can do one of the following:

- Accept spectra with missing quantification channels, report the minimum or maximum allowed fold change, and use the quantification results (for example, x/0 = 100.000, 0/x = 0.010). The default consensus quantification workflows use his option.

- Reject the quantification results.

- Replace the missing reporter with imputed abundance values.

The following topics discuss these options in more detail.

Accepting Spectra with Missing Quantification Channels

If one or more of the quantification channels are missing, the calculated ratios using these channels are either zero or infinity, depending on which channel is the numerator and which is the denominator. Even if all channels are present, the calculated ratios might be very high or very low. You can use the Maximum Allowed Fold Change parameter of the Reporter Ions Quantifier node or Precursor Ions Quantifier node to replace such extremely high or extremely low ratios with the maximum or minimum allowed fold change, respectively—that is, extremely high ratios (greater than 100) are replaced by 100, and extremely low ratios (less than 0.01) are replaced by 0.01. If you set the Maximum Allowed Fold Change parameter to 1000, the application replaces any ratios higher than 1000 by 1000. It replaces any ratios less than 0.001 by 0.001. By default, the Maximum Allowed Fold Change parameter is set to 100. You can find this parameter in the General Quantification Settings category in the Parameters area (see Figure 386). You can expect an inherent dynamic range to be valid or detectable with the given instrumentation and method.

To accept spectra with missing quantification channels, set the Reject Quan Results with Missing Channels parameter to False in the Reporter Ions Quantifier node or Precursor Ions Quantifier node (Figure 386).

Figure 386. Default parameters settings for the Reporter Ions Quantifier node

Rejecting the Quantification Results

When you reject any quantification results from spectra with missing quantification channels, the application does not use the corresponding Quan Spectra to calculate the protein and peptide quantification values. The Peptide Quan Usage column of the .pdResult report displays "Not Used," and the Quan Info column displays "No Quan Values."

You can use the Reject Quan Results with Missing Channels parameter of the Reporter Ions Quantifier node or the Precursor Ions Quantifier node to reject the quantification measurement if any of the quantification channels are zero. When you set this parameter to True, the application does not use a quan spectrum or consensus feature for calculating peptide group and protein quantification values if one or more of the quantification channels have a detected abundance of zero. If you set it to False, which is the default, the application calculates and displays the quantification values for those channels that it found and retains the missing channels.

Replacing the Missing Quantification Channels

The third option is to replace the missing quantification channels. Use the Imputation Mode parameter. For information on this parameter, see "Imputing Missing Values" on page 520.

Calculating Quantification Results

The next topics give an overview of the methodology that the Proteome Discoverer application uses to perform quantification and then examine each step in detail.

Quantification Methodology

The application uses the following general methodology to perform quantification:

1. It calculates the raw quantification values and associates them with the PSMs identified by Sequest HT, Mascot, or other search engines. The quantification values are peak areas or heights for precursor ion quantification, and intensities or S/N values of the mass or reporter tags in reporter ion quantification.

2. It sums the quantification channel values linked to the PSMs to the peptide groups and proteins. PSMs that do not meet the criteria set by the method, such as minimum S/N threshold or isolation purity, are not included in this sum. Otherwise, this methodology aggregates all contributing signals of the protein, irrespective of the charge and modification state of the peptides. The result is quantification channel abundance values for all channels from all files for the peptides and proteins.

3. The application optionally normalizes the abundance values according to the selected normalization method in the consensus workflow.

4. It optionally scales the normalized abundance values for each protein and peptide so that the average abundance is 100. For instance, if you select a TMT 6plex quantification method, the application scales the calculated abundance values of each protein so that their sum is 600 (100 × 6). If the abundances are equal, each one is 100.

5. It groups the optionally normalized and scaled abundance values according to the sample grouping that you specified when you set up the analysis. Grouping in this case means effectively taking the median of the abundance values of all replicates.

6. As the last step, the application calculates the quantification ratios from the grouped abundance values according to the ratios selected when you set up the analysis.

The Precursor Ions Quantifier node or the Reporter Ions Quantifier node in the consensus workflow calculates quantification abundances and ratios from the raw quantification values used in the processing workflows. It calculates quantification abundances from the data processed by the Minora Feature Detector node or the Reporter Ions Quantifier node.

The Precursor Ions Quantifier node and the Reporter Ions Quantifier node perform the following steps to calculate the quantification ratios:

- Calculating PSM Abundances

- Classifying Quan Results

- Calculating Peptide Group Abundances

- Classifying Peptide Groups

- Calculating Protein Abundances

- Normalizing Peptide Groups and Protein Abundances

- Using Sample Information to Calculate and Display Quantification Results

- Calculating Group Abundances

- Calculating Peptide Group and Protein Ratios

- Calculating Abundance CVs and Ratio Variability

- Calculating Protein Group Ratios

Calculating PSM Abundances

A quantification value is the intensity, S/N, or area detected for a given quantification channel. For reporter ion quantification, a quantification channel is one of the mass or reporter tags, and for precursor ion quantification, it is one of the different possible labeling states of a PSM corresponding to the different heavy amino acids used in the cell cultures. "Intensity" refers to both the intensity of the reporter peaks in reporter ion quantification and to the apex intensity or areas detected in precursor ion quantification. The S/N corresponds to the intensity of the peak divided by the noise values calculated for each peak during acquisition.

You can set the following parameters to control this step:

- Use Single-Peak Quan Channels: Determines whether the application uses quantification channels identified with only a single peak for the precursor ion pattern for precursor quantification.

- Average Reporter S/N Threshold: Specifies the minimum average reporter signal-to-noise threshold value to use. The application ignores PSMs with quantification results below this threshold.

- Precursor Abundance Based On: For label-free or precursor ion quantification, you can choose the abundance associated with a PSM to be either the intensity at the apex of the chromatographic profile of the selected feature or the total peak area.

- Reporter Abundance Based On: For reporter ion quantification, the PSM abundance can be based on either S/N or peak intensity. The Automatic setting bases the abundance on S/N for quantification spectra acquired using Orbitrap MS and bases the abundance on intensity for quantification spectra acquired suing ion trap MS.

Table 30 lists some of the different circumstances that can arise in calculating quantification ratios for PSMs from the selected quantification values. The parameters in the table belong to the Reporter Ions Quantifier node.

Table 30 does not include cases resulting from PSM uniqueness and protein grouping. It focuses on cases where some or all of the quantification channels are zero. In these cases, the application detects nothing on a channel because the spectrum does not contain one of the reporter peaks, the heavy or light isotope pattern is missing, a quantification value falls below a specified minimum threshold, or the calculated ratios are very high or very low.

Table 30 also lists the different possible cases exemplified by arbitrary values. The values in the tables have [counts] as units if the cases are presented for reporter ion quantification. For precursor ion quantification, 114 and 115 are replaced by Light and Heavy and the quantification values have [counts × seconds] as units.

Table 30. Ratios calculated from grouped abundances

Case	Reject Quan Results with Missing Quan Channels parameter	Detected quantification values			Displayed or used quantification values		
		114	115	116	114	115	116
All quan. values detected	Irrelevant	100	50	300	100	50	300
Quan. value missing for a quan. channel	No	100	0	300	100	0	300
	Yes	100	0	300	0	0	0
	Irrelevant	100	0	300	100	33	300
Quan. value missing for all quan. channels	Irrelevant	0	0	0	0	0	0

Using Reporter Ion Isotopic Distribution Values to Correct for Impurities

iTRAQ and TMT kits consist of labels that contain different numbers of ^{13}C atoms, ^{15}N atoms, or both. For simplicity, assume that a 4plex kit yields peaks at 114, 115, 116, and 117 *m/z*, which correspond to $^{13}C1$, $^{13}C2$, $^{13}C3$, and $^{13}C4$, respectively. Because the label substances are not 100 percent isotopically pure, each label contains a certain number of other atoms. For example, the 116 label would not consist only of label molecules having three ^{13}C atoms but might also contain label molecules with only one or two ^{13}C atoms or even four or five ^{13}C atoms. As a result, these impurities lead to an observed peak at 116 *m/z*, which is smaller than might be expected if the tag were 100 percent isotopically pure, and

additional peaks at positions –2, –1, +1, +2 Da apart from 116 *m/z*. The intensities of the latter peaks are proportional to the amount of the described isotopic impurities. When the 116 label and the 114, 115, and 117 labels are used, these latter three labels contribute to the peak at 116 *m/z* because of their isotopic impurities.

The intensity of the peak at 116 *m/z* effectively includes the following contributions:

(*observed intensity 116*) = (*true intensity 116*) – (*intensity loss because of 116 impurities*) + (*intensity gain because of other label impurities*)

To obtain the true intensity value of the 116 label—that is, the amount of the substance initially labeled with the 116 tag—you must correct the experimentally observed peak for the impurity of the labels.

For a 4plex sample, four formulas use the above equation for each of the labels, and the proper correction would consider both contributions in the formula by solving the system of coupled linear equations:

(*intensity_of_loss_because_of_116_impurity* and
intensity_of_gain_because_of_other_label_impurities)

For this correction, you must enter the isotopic distribution of each of the labels used in the quantification method, as described in "Excluding PSMs with High Levels of Coisolation." The values are part of each of the iTRAQ or TMT label kits that were used.

You can also deconvolve the overlapping labels using other methods. Compatible with the Mascot search engine, the application uses a first-order approximation to the solution. The error made is small when the intensities of all possible contributing labels are of similar height, and it becomes larger if the intensity differences become larger. You can find more information on this topic, at www.matrixscience.com/help/quant_config_help.html. In the Sections menu, choose **Component > Correction > Type AB Certificate**.

In TMT kits, the values of impurities in reporter ion labels are normalized to the highest peak. In the application, they are normalized to 100 percent, which the application considers to be the sum of all impurities.

For TMTe 6plex kits, using correction factors is optional, because the resolution in Q-Exactive or Orbitrap Elite instruments is enough to separate ^{13}C and ^{15}N isotopes.

TMT 10plex kits use no correction factors by default. You can add correction factors to these kits by following the instructions in "Excluding PSMs with High Levels of Coisolation."

In the kits, correction factors are the same for the same batch of isotopes and are usually valid for three to five years. See "Checking the Quantification Method" on page 554.

Excluding PSMs with High Levels of Coisolation

To create a fragment spectrum, you select a precursor mass for isolation, isolate and fragment the ions within a mass window that you define, and record the product ion masses created.

Ideally, you would isolate and fragment only the precursor ions of a single selected component. However, in practice you isolate the precursor ions within a user-specified window—typically 1 or 2 daltons around the isolation mass. Coeluting components with a mass falling into this isolation window are also isolated and fragmented. This process is called coisolation. The coisolating components are likely to be peptides whose fragments are observed in the created fragment spectra. The coisolation can prevent the identification of the selected peptide or lower the identification confidence.

Coisolation is also an issue for reporter ion quantification. In this type of quantification, the peptides from different samples—for example, different treatment states—are modified with isobaric labels. The isobaric labels fragment during precursor ion fragmentation and create reporter tags that appear in the low-mass region of the fragment spectra. The intensity ratio of the observed fragment tags is used for relative quantification of the peptides from the different sample charges.

The coisolating peptides also create reporter tags with the same masses as those from the selected peptide. The intensities of the reporter ions are therefore the sum of the intensities of the reporter ions for all coisolated peptides rather than the target peptide. As a a result, the intensities are perturbed and are not accurate representations of the true abundance of the selected peptide. Furthermore, the perturbed ratios of the selected peptides that are greatly affected by coisolation can also adversely affect the ratios that the application calculates for the proteins that include these peptides.

Determining the extent to which the real reporter tag ratios of the selected peptides are perturbed depends on the level of coisolation and the isolation characteristics of the instrument. The application calculates and displays the percentage of interference within the precursor isolation window. This percentage is the relative amount of ion current within the isolation window that is not attributed to the precursor itself:

$$\%\_isolation\_interference \;=\; 100 \times \left[1 - \left(\frac{precursor\_intensity\_in\_isolation\_window}{total\_intensity\_in\_isolation\_window}\right)\right]$$

The application displays the calculated interference value in the % Isolation Interference column on the PSM and MS/MS Spectrum Info pages. For reporter ion quantification, a high isolation interference value could indicate that a calculated peptide ratio is skewed by the presence of coisolated peptide species.

> **Note** The application only calculates the % Isolation Interference value if the precursor scans are high-resolution, high-mass-accuracy scans.

You can use the Co-Isolation Threshold parameter of the Peptide and Protein Quantification parameter to specify a threshold of between 0 and 100 percent for the allowed coisolation interference. The default value is 100 percent, which means that no PSM is excluded. This parameter is only used for reporter ion quantification.

Classifying Quan Results

Seven different scenarios can occur when you derive protein quantification ratios from peptide quantification ratios. These cases show how the validity of using a given quantification result for the quantification of a certain protein depends on whether this particular quantification result is unique or shared among other peptides.

The peptide quantification ratios are taken from the associated quantification results. The term *quantification result* in this topic refers to MS/MS reporter intensities taken from the same scan as the identification (for example, ID-CID) or from a separate quantification scan (for example, Quan-HCD). The term also refers to intensities derived from the precursor scans in precursor ion quantification. A quantification result here is a general quantity associated with one or more peptides that are, in turn, associated with one or more proteins.

Case 1: Quantification Result Associated with One Spectrum, One Peptide, and One Protein

Case 1 (see Figure 387) is the simplest case. The quantification result is associated with one identification spectrum—whether the quantification results come from the same identification spectrum, from a different quantification spectrum, or from the precursor ion—and one peptide that is contained in one protein. The quantification result is unique for this protein. The application can mark peptide A "Unique" in the Quan Info column of the PSMs page if the quantification result meets other criteria.

Figure 387. Case 1: Quantification result associated with one identification spectrum, one peptide, and one protein
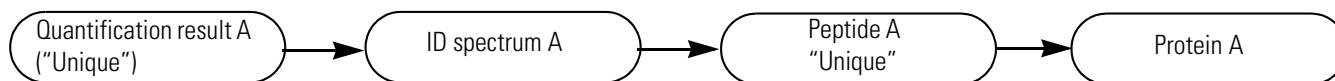
Case 2: Two Quantification Results Associated with Two Spectra, One Peptide, and One Protein

Case 2 (see Figure 388) is a variant of case 1. Each of two different quantification results is associated with a different identification spectrum. Both identification spectra identify peptide A, which is a peptide with the same sequence. Peptide A is only contained in one protein. Each of the two different quantification results is unique for just one protein. The peptides are redundantly identified and quantified, and you can use both for the quantification of protein A.

Figure 388. Case 2: Two different quantification results associated with two identification spectra, one peptide, and one protein
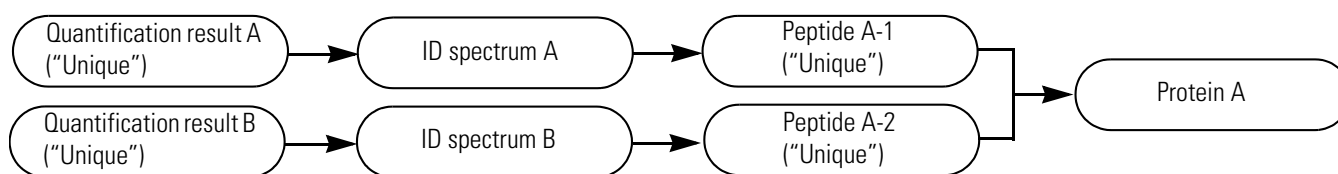
Case 3: Quantification Result Associated with Two Spectra, Two Peptides, and One Protein

Case 3 (Figure 389) is similar to case 2 but varies from it in a slight but important way. In case 3, the two identification spectra are associated with the same quantification result rather than with two different quantification results. For example, you might obtain these results if you trigger the same precursor two times for MS/MS. It does not matter whether peptide A and peptide B are the same peptides (redundantly identified) or different peptides that are accidentally contained in the same protein. It also does not matter whether they are identified by the same search engine or by two different search engines, for example, a CID spectrum and an ETD spectrum. The quantification result is still unique for just one protein. However, you cannot use the quantification ratio of both peptides A and B to calculate the quantification ratio of protein A, because it is the same quantification result, and you do not want to use the same quantification result multiple times for the same protein. In this case, the application marks peptide A—the peptide with the better identification—as "Unique" and the other peptide as "Redundant" for quantification (rather than redundant for identification).

Figure 389. Case 3: Quantification result associated with two identification spectra, two peptides, and one protein
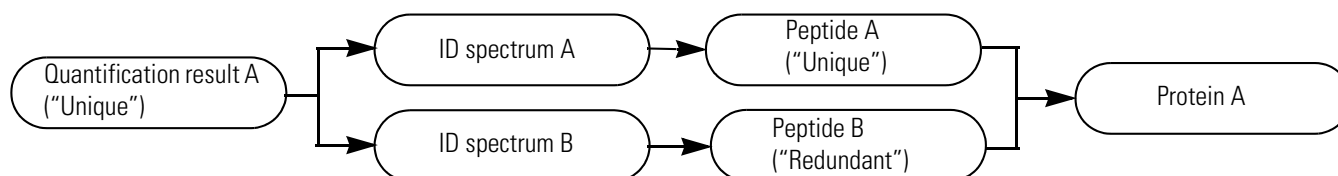
Case 4: Quantification Result Associated with One Spectrum, Two Peptides, and One Protein

In case 4 (Figure 390) the two peptides can be identified by the same search engine and have different ranks, or they can be identified by different search engines and both have rank 1. It does not matter whether peptide A and B have the same sequence with different PTM states or different sequences. The quantification result is unique for protein A. You can use it to calculate the protein ratio, but you must only use it once. The application marks the "better" peptide as "Unique" and the other as "Redundant" for quantification.

Figure 390. Case 4: Quantification result associated with one identification spectrum, two peptides, and one protein
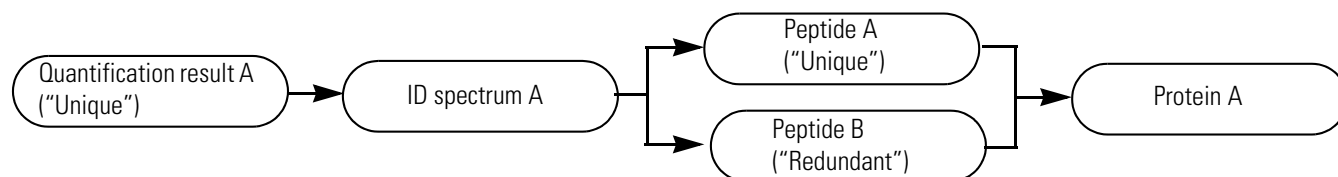
Case 5: Quantification Result Associated with One Spectrum, One Peptide, Two Proteins

In case 5 (Figure 391) the quantification result is associated with one identification spectrum and one peptide, but this peptide is contained in more than one protein. The quantification result is potentially shared between these proteins, and you do not know how to share it. If the quantification method specifies using only unique peptides for protein quantification, you would not use peptide A in this case. If the quantification method specifies using all peptides for protein quantification, the quantification result of peptide A would be divided equally between both proteins.

Figure 391. Case 5: Quantification result associated with one identification spectrum, one peptide, and two proteins
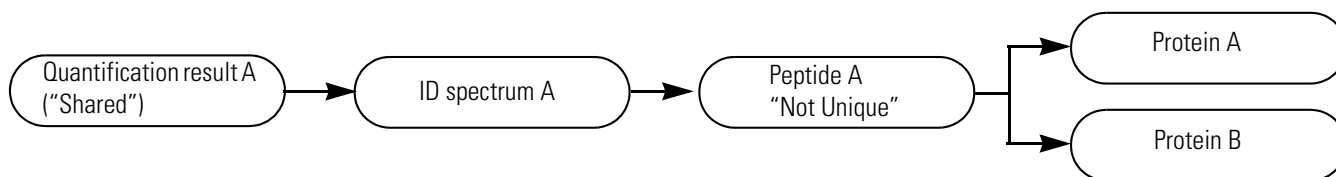
Case 6: Quantification Result Associated with One Spectrum, Two Peptides, and Two Proteins

In case 6 (Figure 392) the quantification result is associated with one identification spectrum from which two different peptides are identified either by the same search engine as different ranks or by different search engines. The two different peptides are contained in two different proteins. The two different peptides are both unique to just one protein. Nevertheless, the associated quantification result is the same, and you do not want to use it to calculate the protein ratios if you specified in the quantification method to use only unique peptides. Only if you specify using all peptides can you use them for protein quantification. This case illustrates the discrepancy between the uniqueness of peptides and the uniqueness of the quantification results.

Figure 392. Case 6: Quantification result associated with one identification spectrum and two peptides unique to one protein
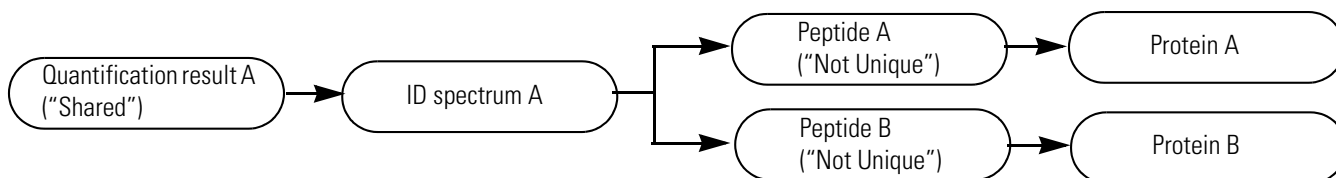
Case 7: Quantification Result Associated with Two Spectra, Two Peptides, and Two Proteins

Case 7 (Figure 393) is a variant of case 6. Either the same search engine or different search engines identify different identification spectra, for example, CID and ETD spectra. As in case 6, the peptides are unique, but the quantification result is not. The result depends on whether you specified in the quantification method to use only unique peptides or all peptides.

Figure 393. Case 7: Quantification result associated with two identification spectra and two different peptides unique to one protein
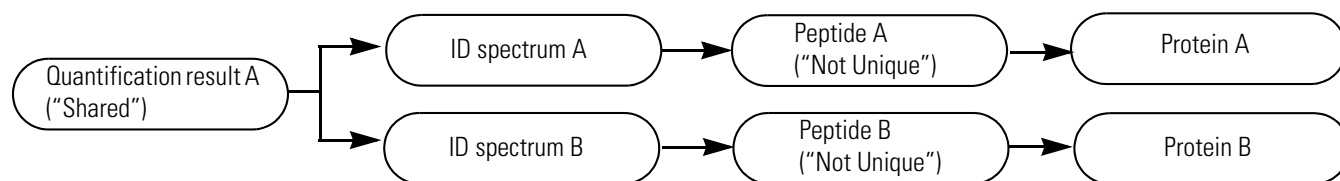
Classification Flow Chart

The application calculates the quantification values for the protein as the sum of the intensity or S/N values for all PSMs belonging to the proteins that are marked as being usable. Whether the application considers a PSM usable is determined by the parameters of the Reporter Ions Quantifier node and the Precursor Ions Quantifier node, including those labeled Peptides to Use and Consider Protein Groups for Peptide Uniqueness. The Use Only Unique Peptides parameter includes the quantification peptides that do not occur in other proteins. The Proteins Groups for Peptide Uniqueness parameter defines peptide uniqueness on the basis of protein groups rather than individual proteins.

When it determines peptide uniqueness for classification in the PSM Ambiguity column on the PSMs page, the application only considers the PSMs that it considered when creating the protein groups, if you select the Use Only Unique Peptides parameter. For example, it does not use for quantification a PSM of low confidence that it did not use to create the protein groups.

The flowchart in Figure 394 shows how the application classifies PSMs for protein quantification. It displays this classification in the Quan Info column of the results report.

Figure 394. Classifying peptides for protein quantification

Depending on the settings of the parameters of the Reporter Ions Quantifier node or the Precursor Ions Quantifier node, the application excludes the following peptides:
- No Quan Labels: No reporter label.
- Indistinguishable Channels: Not all defined channels can be distinguished.
- Inconsistently Labeled: Labels are from different channels.
- Excluded by Method: Quan channels are missing, single-peak channels are missing, ratios exceed limits, and so forth.

The application does not consider high-scoring peptides if they have no protein links.

The application marks peptides that are not unique as Redundant.

This check depends on the setting of the Consider Protein Groups for Peptide Uniqueness parameter of the Reporter Ions Quantifier node and the Precursor Ions Quantifier node. It classifies the peptides as Unique if they match the proteins within the same protein group.
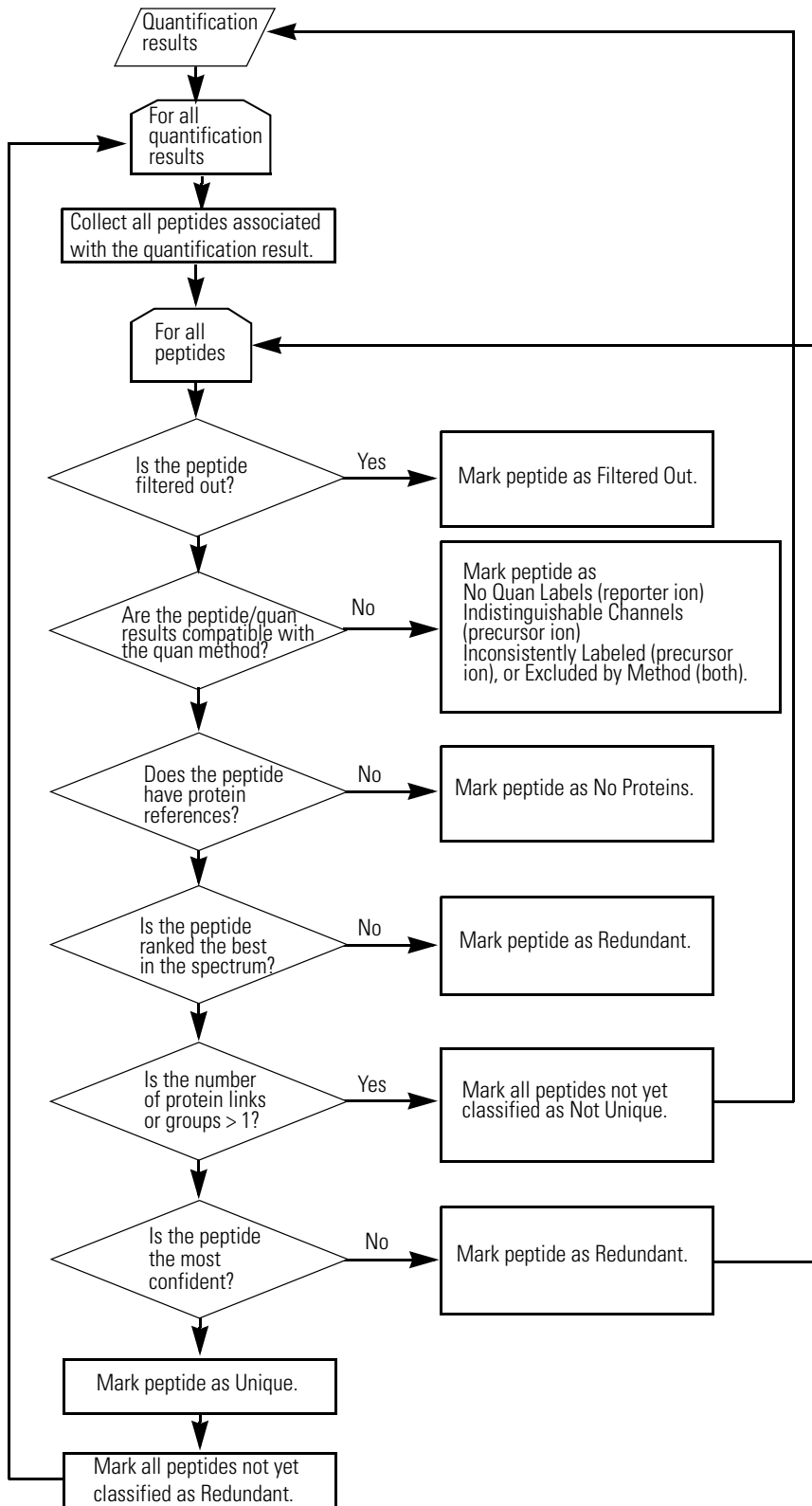
Calculating Peptide Group Abundances

After classifying Quan Spectra, the Reporter Ions Quantifier node and the Precursor Ions Quantifier node calculate peptide group abundances for the different samples from their associated Quan Spectra. It calculates the abundance for a peptide group as a simple summation of its associated and used Quan Spectra abundances.

When the node calculates peptide group abundances, it creates the following columns:

- Peptide Quan Usage column on the PSMs page, which shows the PSMs that are used for calculating peptide group abundances. This column displays all PSMs, whether or not they are unique.

- Abundances column on the Peptide Groups page, which displays the peptide group abundances.

- Abundance Counts column on the Peptide Groups page, which displays the number of PSM abundances used for calculating the peptide group abundances.

Classifying Peptide Groups

The Reporter Ions Quantifier node and the Precursor Ions Quantifier node classify which peptide groups are used for protein quantification.

Use the following parameters to control the classification of peptide groups:

- The Consider Protein Groups for Peptide Uniqueness parameter determines which peptide groups are unique. If you set it to True, the Reporter Ions Quantifier node and the Precursor Ions Quantifier node consider a peptide unique if it is included in only one protein group. If you set it to False, the node considers a peptide unique if it is included in only one protein.

- The Peptides to Use parameter determines which peptides to use.

  – Unique: Uses only unique peptide groups.

  – Unique + Razor: Uses both unique peptide groups and peptide groups containing razor peptides for the best associated master protein. Razor peptides are shared among multiple protein groups or proteins. When you set the Consider Protein Groups for Peptide Uniqueness parameter to True, the node uses peptide groups that are not unique for all proteins belonging to the protein group of the best master protein. The best master protein is the protein with the largest value in the # Protein Unique Peptides column on the Proteins page and with the smallest value in the Coverage column (the longest protein).

  – All: Uses all peptide groups, whether they are unique or not.

When it classifies peptide groups, the Reporter Ions Quantifier node and the Precursor Ions Quantifier node create the following columns in the .pdResult file:

- Quan Info column on the Peptide Groups page, which displays No Quan Values for peptide groups that have no abundance value. It displays Not Unique or Unique for all other peptide groups.

- # Razor Peptides column on the Proteins page, which displays the number of razor peptides for proteins when you use razor peptides for quantification.

Calculating Protein Abundances

After the Reporter Ions Quantifier node and the Precursor Ions Quantifier node classify peptide groups, they calculate protein abundances for the samples from the associated peptide group abundances. They calculate a protein abundance as the simple summation of its associated and used peptide group abundances.

The nodes create the Abundance Counts column on the Proteins page, which shows the number of peptide group abundances used for calculating abundances.

Normalizing Peptide Groups and Protein Abundances

After the Reporter Ions Quantifier node and the Precursor Ions Quantifier node calculate the protein abundances for the samples, they normalize the peptide groups and protein abundances, impute missing values, and scale them. In general, the nodes apply the normalization of the total abundance values for each channel across all files, equalizing the total abundance between different runs.

You can use the following parameters of the Reporter Ions Quantifier node and the Precursor Ions Quantifier node to normalize peptide groups and protein abundances.

Normalization Mode Parameter

Use the settings of the nodes' Normalization Mode parameter to specify how they perform normalization:

- (Default) None: Does not perform normalization.

- Total Peptide Amount: Sums the peptide group abundances for each sample and determines the maximum sum for all files. The normalization factor is the factor of the sum of the sample and the maximum sum in all files.

- Specific Protein Amount: Calculates the normalization factor from the abundances of selected proteins in the specified FASTA file. You can specify a FASTA file (use the Proteins for Normalization parameter to specify the name of the FASTA file) that can contain one or multiple proteins. The Reporter Ions Quantifier node and the Precursor Ions Quantifier node use all proteins in the FASTA file that are contained in the result file and that have any protein abundance. This setting calculates the maximum sum for all files. The normalization factor is the factor of the sum of a sample and the maximum sum in all files.

- RT-Dependent Normalization: This method corrects for datasets where different normalization factors would be required for different retention time ranges within the files.

If you set the Normalization Mode parameter to Specific Protein Amount but do not select a FASTA file, or if the .pdResult file contains no proteins that appear in the FASTA file, the nodes cannot perform normalization.

When normalization is based on specific proteins, the Reporter Ions Quantifier node and the Precursor Ions Quantifier node sum the abundance values of all proteins in the .pdResult file that reside in the selected FASTA file. They compare the protein sequences in the FASTA file to the sequences in the .pdResult file rather than to the accession strings. You can therefore use the same FASTA file when the title line is slightly different.

After calculating the normalization factors, the Reporter Ions Quantifier node and the Precursor Ions Quantifier node normalize quan spectra (for reporter ion quantification) or consensus features (for label-free and precursor ion quantification) by dividing abundances with the normalization factor over all samples.

Scaling Mode Parameter

After normalization, the Reporter Ions Quantifier node and the Precursor Ions Quantifier node scale the optionally normalized and imputed abundances if you set its Scaling Mode parameter to On All Average or On Controls Average. By default, the nodes do not scale abundances.

- On All Average: After aggregating all the abundance or normalized abundance values per sample, the nodes scale the abundance values of each sample so that the average of all samples is 100.

- On Controls Average: After aggregating all the abundance or normalized abundance values per sample, the nodes scale the abundance values of each sample so that the average of all control samples is 100 and then scale all other samples up or down relative to 100. When you use multiplexed files, the nodes process the samples from each file separately.

The application displays normalized abundances in the Abundances (Normalized) column and scaled abundances in the Abundances (Scaled) column. The Abundances (Normalized) column appears on the Proteins and Peptide Groups pages of the .pdResult file when you set the Normalization Mode parameter of the Reporter Ions Quantifier node or the Precursor Ions Quantifier node to Total Peptide Amount or Specific Protein Amount. The Abundances

(Scaled) column appears on the same pages when you set the Scaling Mode parameter of the Reporter Ions Quantifier node or the Precursor Ions Quantifier node to On All Average or On Controls Average. These columns do not appear if you do not apply normalization and scaling.

Excluded Peptide Modification Parameter

You can use from one to three of the Excluded Peptide Modification parameters of the Reporter Ions Quantifier node and the Precursor Ions Quantifier node to determine the peptides to exclude from the rollup from the peptide group quantification results to the protein level.

Figure 395 shows how to select a modification and the abbreviation of the amino acid to determine what peptides to exclude from the normalization calculation.

Figure 395. Excluded Peptide Modification parameters of the Reporter Ions Quantifier or the Precursor Ions Quantifier node
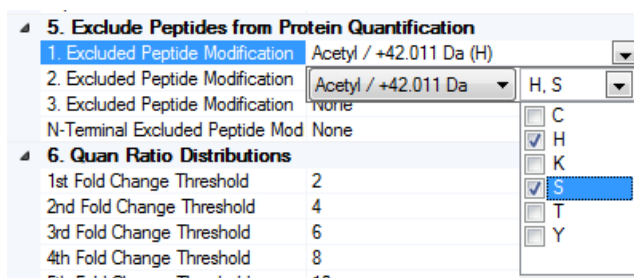
Using Sample Information to Calculate and Display Quantification Results

The application displays the Abundance (Scaled) column and the Abundance (Grouped) column in blue, white, and red tones (see Figure 396). For information about how the application calculates the values in these columns, see "Calculating Group Abundances" on page 539.

Figure 396. Colors used to display the Abundance (Scaled) and Abundance (Grouped) columns

Fold-Change Color Threshold	Scaled Abundance Color Threshold	Color
0.10	18.2	
0.13	22.2	
0.17	28.6	
0.25	40.0	
0.50	66.7	
0.00	100.0	
2.00	133.3	
4.00	160.0	
6.00	171.4	
8.00	177.8	
10.00	181.8	

The application uses the settings of the 1st Fold Change Threshold, 2nd Fold Change Threshold, 3rd Fold Change Threshold, 4th Fold Change Threshold, and 5th Fold Change Threshold parameters of the Reporter Ions Quantifier node and the Precursor Ions Quantifier node to color the scaled abundance values and the ratios calculated from the abundance values shown in the Abundance (Scaled) column and the Abundance (Grouped) columns. For the ratios, the application directly uses the fold changes as thresholds for applying colors from dark blue (downregulation) to white (no regulation) to dark red (upregulation). For the scale abundance values, the application colors the values according to their deviation from 100. It calculates the thresholds of the scaled and scaled-and-grouped abundances as follows:

$$abundance\ coloring\ threshold\ =\ 100\frac{fold\ change\ threshold-1}{fold\ change\ threshold+1}$$

T is the given fold change, and 100.0 is used to scale the average of all channels to 100.0 (per file or across files).

Figure 396 shows the thresholds of the scaled abundance values for the default fold changes.

Figure 397 shows the fold-change coloring scheme for the grouped abundance and scaled abundance values.

Figure 397. Abundances (Grouped) and Abundances (Scaled) columns on the Proteins page
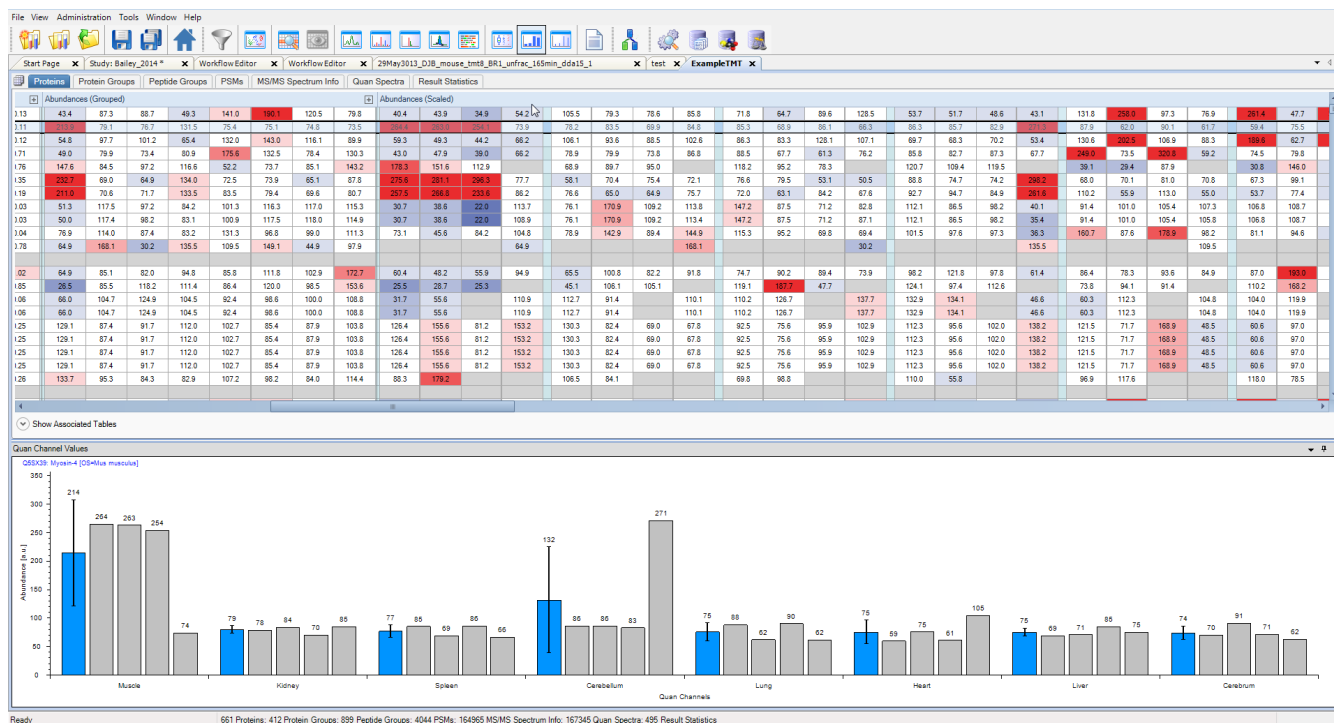
Calculating Group Abundances

After normalization, the Reporter Ions Quantifier node and the Precursor Ions Quantifier node calculate the grouped abundances for peptide groups and proteins. They use the following preprocessed abundances as a basis, when they are available:

- Ungrouped raw abundances, where no normalization and no scaling are applied

- Ungrouped normalized abundances

- Ungrouped scaled abundances (optionally also normalized)

The nodes calculate these as the average of the preprocessed sample abundances that belong to the same sample group. They calculate group abundances only when there is at least one sample group consisting of multiple samples.

Calculating Peptide Group and Protein Ratios

See these topics.

- Calculating Ratios Using Summed Abundance Approach

- Calculating Ratios Using the Pairwise Ratio Approach

Calculating Ratios Using Summed Abundance Approach

This method is used when the Ratio Calculation Parameter in the Precursor Ions Quantifier and the Reporter Ions Quantifier is set to Summed Abundance Based. When the study design is non-nested, the protein ratio is calculated by dividing the summed abundance values described in Calculating Group Abundances for the two samples chosen for that ratio. The peptide group ratio is also calculated by dividing the summed abundance values for that specific peptide.

For precursor ion quantification, the peptide group ratios are calculated slightly differently because the various isotopic forms consist of different consensus features. In this case, the peptide group abundances and the ratio are shown for only one of the isotopic forms used for the ratio when more than one of the isotopic forms has a confidently identified PSM. For an example, see Figure 398. In this example, the highlighted row shows that the heavy-labeled form of the peptide includes a ratio column and grouped abundance values for both the light and heavy forms of the peptide, while the light form does not show a ratio nor abundance values. For protein quantification, the ratios are calculated as a sum of the individual peptide group abundances as just described.

Figure 398. Peptide group abundances and ratios for a SILAC 2plex example

For a nested study design with biological replicates, the application calculates the peptide group and protein ratios differently. In this case, ratios for the summed abundances for each biological replicate are calculated separately, and the geometric median of the resulting ratios is used as the peptide group and protein ratios. Figure 399 shows an example calculation of protein ratios for non-nested and nested designs.

Figure 399. Calculating ratios in nested and non-nested design studies using the summed abundance-based approach

a) Calculating ratios in nested-design studies

Abundances of biological replicates

Untreated			Treated		
Mouse 1	Mouse 2	Mouse 3	Mouse 1	Mouse 2	Mouse 3
97.5	106	104	164	147	170

Ratios (treated / untreated) within individuals

Mouse 1	Mouse 2	Mouse 3
1.68	1.37	1.63

Median ratios

Treated / Untreated
1.63

b) Calculating ratios in non-nested design studies

Abundances of sample groups

Untreated	Treated
102	163

Calculate the ratios of the treatments

Treated / Untreated
1.60

Thermo Fisher Scientific recommends that you use the summed abundance approach for reporter ion quantification, such as TMT where there are few missing values. For precursor ion quantification and label-free quantification, use the pairwise ratio approach described in Calculating Ratios Using the Pairwise Ratio Approach.

The Maximum Allowed Fold Change parameter defines the maximum fold change that the node calculates. All ratios are between 1/fold change and fold change. When the numerator or the denominator is 0, the ratio is the fold change when the denominator is 0 or 1/fold change when the numerator is 0.

Calculating Ratios Using the Pairwise Ratio Approach

Use this method when the Ratio Calculation parameter in the Precursor Ion Quantifier and the Reporter Ions Quantifier nodes are set to Pairwise Ratio Based. For non-nested study designs, the application calculates the peptide group ratios as the geometric median of all combinations of ratios from all the replicates for the selected study factors. The protein ratio is subsequently calculated as the geometric median of the peptide group ratios. Figure 400 shows an example calculation of peptide group and protein ratios for a non-nested design.

Figure 400. Calculating ratios in non-nested design studies using the pairwise-based approach

	Treatment T1			Treatment T2		Pairwise based ratios T1/T2						T1/T2
	A1	A2	A3	A4	A5	A1/A4	A1/A5	A2/A4	A2/A5	A3/A4	A3/A5	
Peptide 1	100	110	105	90	80	1,111111	1,25	1,222222	1,375	1,1666667	1,3125	1,23611111
Peptide 2	50		40	53	50	0,943396	1			0,754717	0,8	0,87169811

	T1/T2
Protein	1,138889

For precursor ion quantification, the pairwise ratios for the peptide groups for the different isotopic forms of a given peptide are combined and displayed on only one of the isotopic forms for that peptide.

For a nested study design, the replicates are calculated as the geometric median of all the pairwise ratios for the technical replicates for the given biological replicate. The peptide sample group ratios are then calculated as the geometric median of the biological replicate ratios just as is done for the summed abundance method above. The protein ratio is then calculated as the geometric median of the peptide group ratios. Figure 401 shows an example calculation of peptide group and protein ratios for a nested study that contains both biological and technical replicates.

Figure 401. Calculating ratios in nested design studies using the pairwise ratio approach

	Treatment T1				Treatment T2				Pairwise based ratios — T1/T2 (M1)				T1/T2 (M2)				Ratios for Bio. Reps. T1/T2		
Bio. Rep.	M1		M2		M1		M2										T1/T2 M1	T1/T2 M2	
Tech. Rep	M 1.1.1	M 1.1.2	M 1.2.1	M 1.2.2	M 2.1.1	M 2.1.2	M 2.2.1	M 2.2.2	1.1.1/2.1.1	1.1.2/2.1.1	1.1.2/2.1.1	1.1.2/2.1.2	1.2.1/2.2.1	1.2.1/2.2.2	1.2.2/2.2.1	1.2.2/2.2.2	T1/T2 M1	T1/T2 M2	T1/T2
Peptide 1	100	110	105	90	80	82	78	76	1,25	1,2195122	1,375	1,34146341	1,3461538	1,38157895	1,15384615	1,1842105	1,295732	1,265182	1,280457
Peptide 2	50		40	53		52	45		0,9615385				0,8888889		1,17777778		0,961538	1,033333	0,997436

	Ratios for Bio. Reps. T1/T2		
	T1/T2 M1	T1/T2 M2	
Protein	1,25	1,180994	1,215497

The Maximum Allowed Fold Change parameter defines the maximum fold change that the node calculates. All ratios are between 1/fold change and fold change. When the numerator or the denominator is 0, the ratio is the fold change when the denominator is 0 or 1/fold change when the numerator is 0.

Calculating Abundance CVs and Ratio Variability

The Reporter Ions Quantifier node and the Precursor Ions Quantifier node calculate the coefficients of variation of the grouped abundances from the abundances of the individual replicates within each group, as follows:

grouped abundances CV = $100 \times std.dev$ / *median(abundance(replicate_1), …, abundance(replicate_n))*

Other CVs are calculated accordingly.

The variability of the pairwise ratios is calculated from the all pairwise ratios of all peptides of a protein, as follows:

variability pairwise ratio = 100×1.483 *MAD(ratio_1, …, ratio_m)* / *(median_fold_change)*

where

- *MAD* is the median absolute deviation

- *median_fold_change* is the median_ratio when this is larger or equal to one; otherwise, *median_fold_change* is 1/median_ratio

In all cases where medians are taken, we use the geometric median; that is, for an even number of elements, we take the geometric_average of the two elements in the middle.

Calculating Protein Group Ratios

The Reporter Ions Quantifier node and the Precursor Ions Quantifier node do not display quantification information for Protein Groups.

Using a Quantification Method

The following topics explain how to set up, add, check, change, remove, import, and export a quantification method.

Setting Up the Quantification Method

Setting up the quantification method is similar for both precursor ion quantification and reporter ion quantification. Both methods use values called quantification (quan) channels as the basis for the ratio reporting.

For reporter ion quantification, a quantification channel is one of several masses, states, or tags (depending on which quantification method you use) for which you measure a quantification value. The application calculates the reported quantification ratios from the quantification values of the different quantification channels. For example, for iTRAQ 4plex, the different reporter tags (114, 115, 116, 117) are the four quantification channels of the iTRAQ 4plex method. The application calculates the ratios from the detected quantification values of the four quantification channels.

For precursor ion quantification, a quantification channel is one of the different possible labeling states of a peptide corresponding to the different heavy amino acids used in the cell cultures. For example, the SILAC 2plex methods are normally used with two quantification channels named "light" and "heavy." The light quantification channel uses the natural isotopes of lysine ($^{12}C_6{}^{14}N_2$) and arginine ($^{12}C_6{}^{14}N_4$). In the heavy quantification channel, arginine 10 ($^{13}C_6{}^{15}N_4$) replaces all arginines, and either lysine 6 ($^{13}C_6{}^{14}N_2$) or lysine 8 ($^{13}C_6{}^{15}N_2$) replaces all lysines.

❖ **To set up the quantification method**

On the Study Definition page of the study, select the check box of the appropriate quantification method or methods in the Quantification Methods area (see Figure 402).

The example shows the iodo TMT 6plex method selected.

If the quantification method that you want to use is not in the list, you must create it. For instructions, see "Specifying the Quantification Channels" on page 544.

Figure 402. Quantification Methods area of the Study Definition page
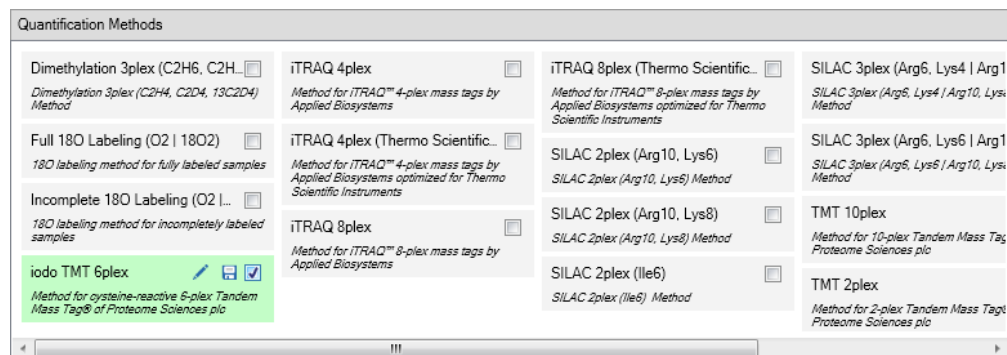
If you use an input file in one processing step and another input file in another processing step, you can select a different quantification method for each file.

> **Note** After you add a quantification method to a study, it is completely separated from the original method on the server. Edits that you make to the quantification method in the study do not propagate to the method on the server and vice versa, although they seem to be the same at first glance. You must be able to copy a study in the study folder to another computer that might contain different quantification methods or none at all. For the study to work on the new computer, the application saves the quantification method with the study rather than on the server.

Specifying the Quantification Channels

The first step in setting up the quantification is to specify the quantification channels to use. The process of specifying label modifications is similar for precursor ion quantification and reporter ion quantification, but it also has some differences:

- For precursor ion quantification, you specify the label modifications for each quantification channel. For reporter ion quantification, you set the label modification that produces all reporter ion quantification channels.

- For precursor ion quantification, specifying the label modifications for quantification channels other than the unlabeled channel is mandatory. For reporter ion quantification, specifying the label modifications is optional because the information about the modification of the peptides is not necessary for processing the data. It is only used to verify the peptides when the application loads the reports.

When you specify at least one of the label modifications in the quantification method, the application verifies that each identified peptide has at least one of the specified modifications during processing.

- When the application identifies a peptide with none of the specified label modifications, this peptide cannot be the source of reporter peaks in the MS/MS spectra. As a result, the application marks the peptide "No Quan Labels" in the .pdResult file. It does not use these peptides when it calculates the protein quantification values.

• When the application finds a peptide that does not have an iTRAQ or TMT label as a modification, even though reporter ions were present, it leaves the *Ratio* columns in the .pdResult file blank.

When you install the application, the default methods for TMT and iTRAQ do not include the correct label modification. When you set up a workflow, ensure that these modifications are selected as either fixed or variable. The application does not automatically update already existing reporter methods; you must manually specify the label modifications.

When you open old .pdResult files that contain reporter quantification data, the label modifications of the quantification method of the .pdResult file appear as None on the Quan Channels page of the Quantification Method Editor dialog box. You can manually specify the label modification, which then triggers the validation of the peptides, and save the change in the quantification method in the .pdResult file.

When you do not set the label modifications on the Quan Channels page, the application does not perform the validation.

The process of specifying quantification channels for precursor ion quantification is slightly different from the process of specifying label modifications for reporter ion quantification.

❖ **To specify quantification channels for precursor ion quantification**

1. Choose **Administration > Maintain Quantification Methods**, or click the **Maintain Quantification Methods** icon, 🗄 .

   The Quantification Methods view opens (see Figure 403). It lists all of the available methods for both precursor ion and reporter ion quantification.

Figure 403. Quantification Methods view
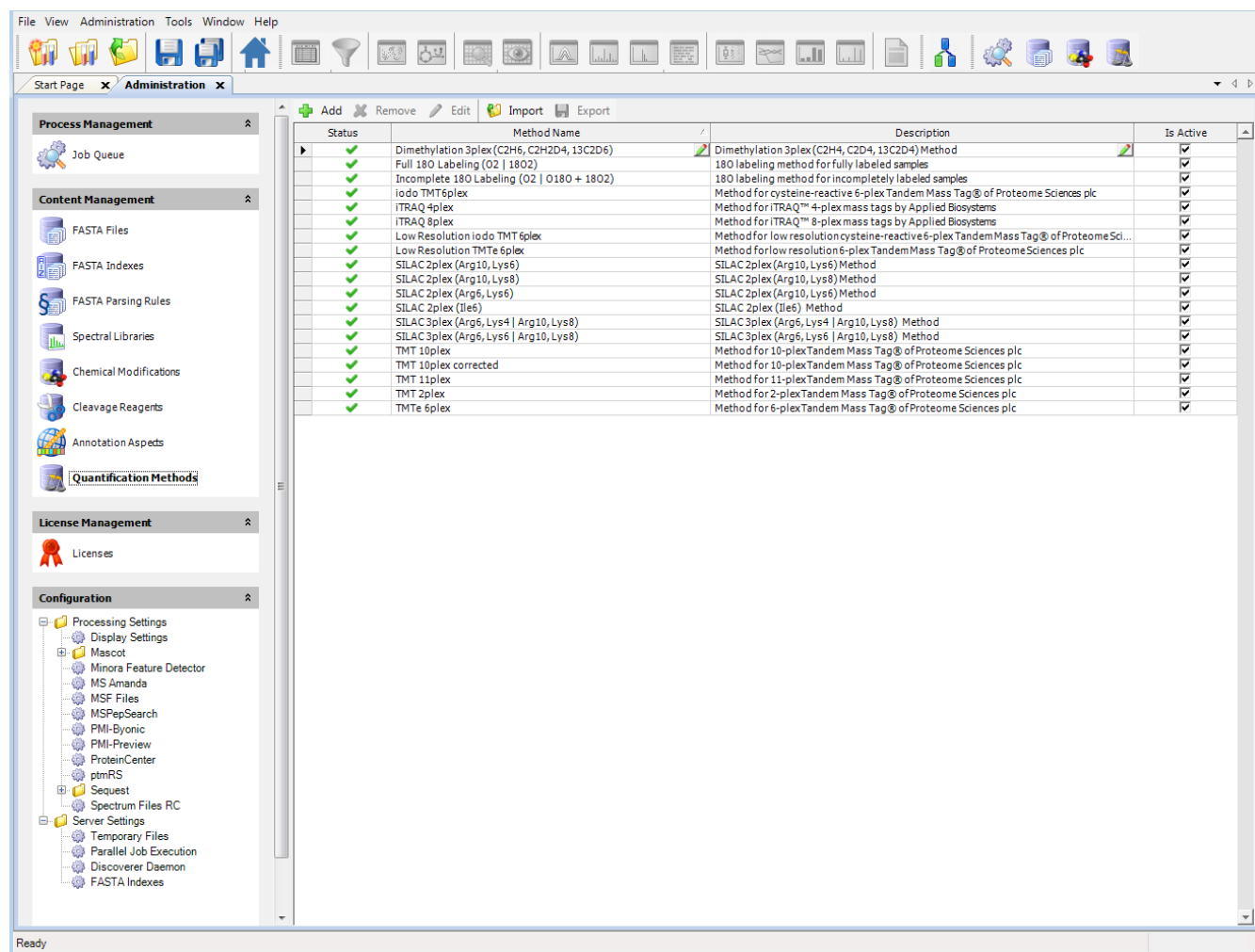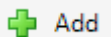
The Status column indicates whether the quantification method is valid for use in quantification:

- Green check mark—The quantification method is valid and can be used for quantification.

- Exclamation point in a yellow triangle—The quantification method is not valid. Double-click this mark to view a message that describes the error and how to fix it.

Figure 404 gives examples of these symbols in the Status column.

Figure 404. Method validity symbols in the Quantification Methods view
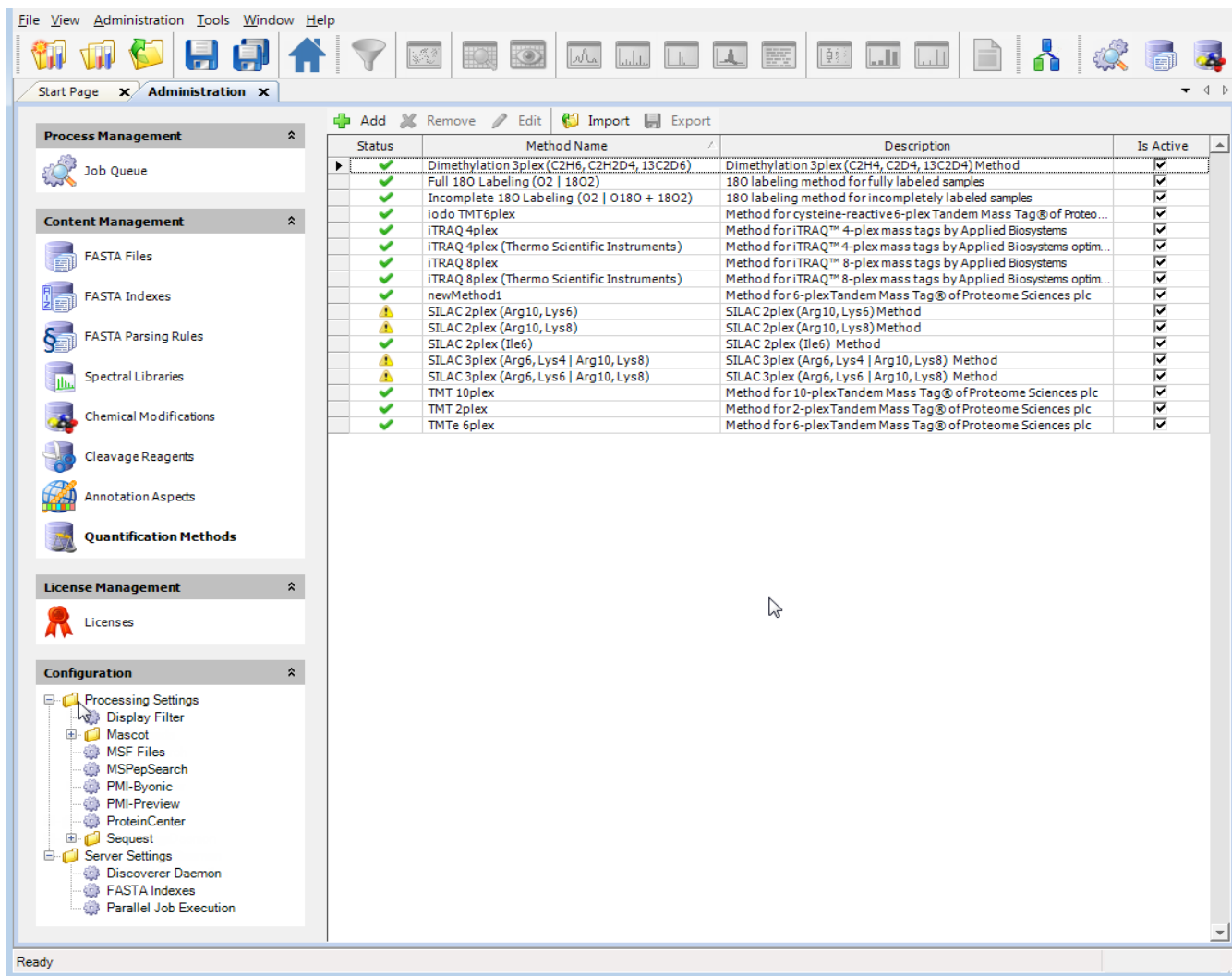
2. Double-click the row for the appropriate method in the Method Name or Description column, or click the column to the left of the Status column for the method and click ➕ Add .

The Create New Quantification Method dialog box opens (see Figure 405).

Figure 405. Create New Quantification Method dialog box

3. Select one of the following methods of creating a quantification method:

- (Default) **From Factory Defaults**: Creates a new method using the same settings from one of the default settings that come with a new application installation.

- **From Existing Method**: Uses the same settings as those of the existing quantification method that you select from the list. The list of methods is the same as the list given at the beginning of "Setting Up the Quantification Method" on page 543.

- **From Scratch**: Uses one of the following templates so that you can build a new processing method from the beginning:

  – **Reporter Ion Quan Method**: Provides a template for reporter ion quantification. This method requires an additional license.

  – **Precursor Ion Quan Method**: Provides a template for precursor ion quantification.

4. Click **Create**.

The Quantification Method Editor dialog box opens (see Figure 406).

Figure 406. Quantification Method Editor dialog box for precursor ion quantification
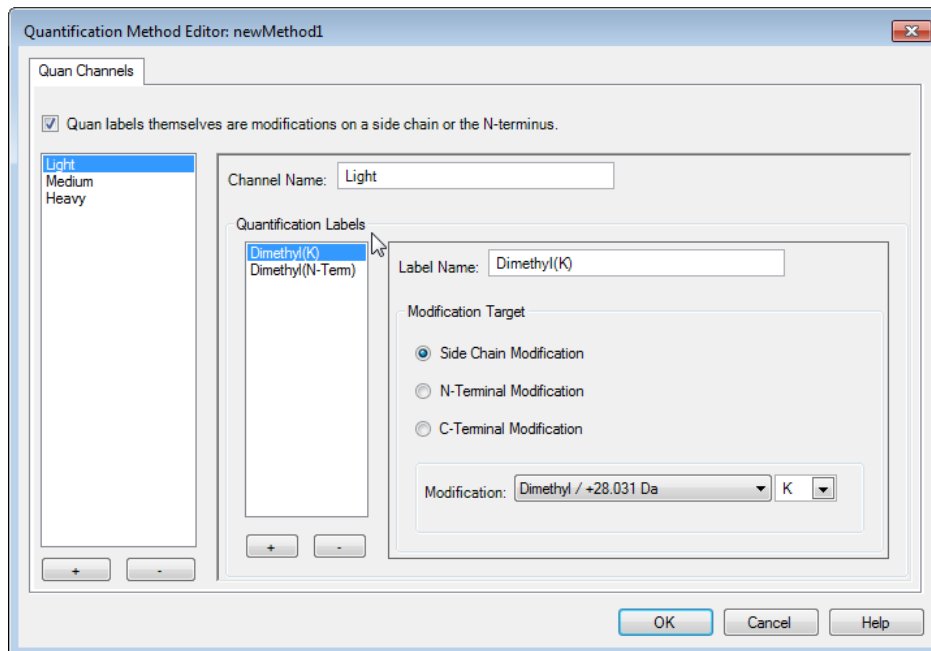
5. If you want the application to ignore modifications on the N terminus, clear the **Quan Labels Themselves Are Modifications on a Side Chain or the N-Terminus** check box.

   By default, this check box is selected. Select this check box for dimethylation and clear it for SILAC modifications.

6. In the box at the left, select the name of the labeling method to use:

   • **Heavy**: Refers to amino acid labels that use heavy isotopes, for example, Arg10 and Lys8.

   • **Medium (3plex methods only)**: Refers to amino acid labels that use less massive isotopes, for example, Arg6 and Lys4.

   • **Light**: Refers to amino acid labels that use normal isotopes.

7. To add a quantification channel, click **+** below the list of quantification channels on the left.

   The default name of New *number* appears in the list and in the Channel Name box (see Figure 407).

Figure 407. New quantification channel on the Quan Channels page for precursor ion quantification
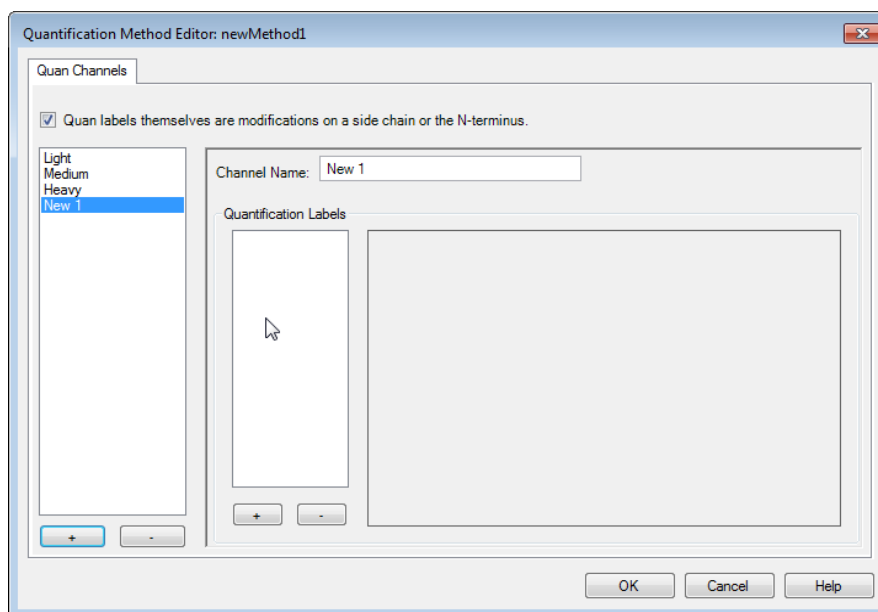
To remove a quantification channel, select the quantification channel from the list on the left and click **–** below the list.

8. To specify a name for the new quantification channel, backspace over the default name in the Channel Name box and type the new name.

   The new name appears in the quantification channel list.

9. To specify a quantification label to assign to a quantification channel, click **+** below the Quantification Labels box.

   A default quantification label of New *number* appears in the Quantification Labels box and the Label Name box.

   To remove an existing quantification label, select the label in the Quantification Labels box and click **–** below it.

10. To change the default quantification channel name, backspace over the name in the Label Name box and type the new name. The example in Figure 408 uses Arg6, Lys6.

11. In the Modification Target area, select the location of the label on the peptide:

    • **Side Chain Modification**: The label occurs on a side chain.

    • **N-Terminal Modification**: The label occurs on the N terminus.

    • **C-Terminal Modification**: The label occurs on the C terminus.

12. From the Modification list, select the modification for the amino acid label.

    The example in Figure 408 shows Label:13C(6) / +6.020 Da.

13. From the list next to the Modification list, select the abbreviation of the amino acid selected in the Quantification Labels box that the modification should occur on.

In this example, K is selected.

The completed Quan Channels page resembles Figure 408.

Figure 408. Completed Quan Channels page
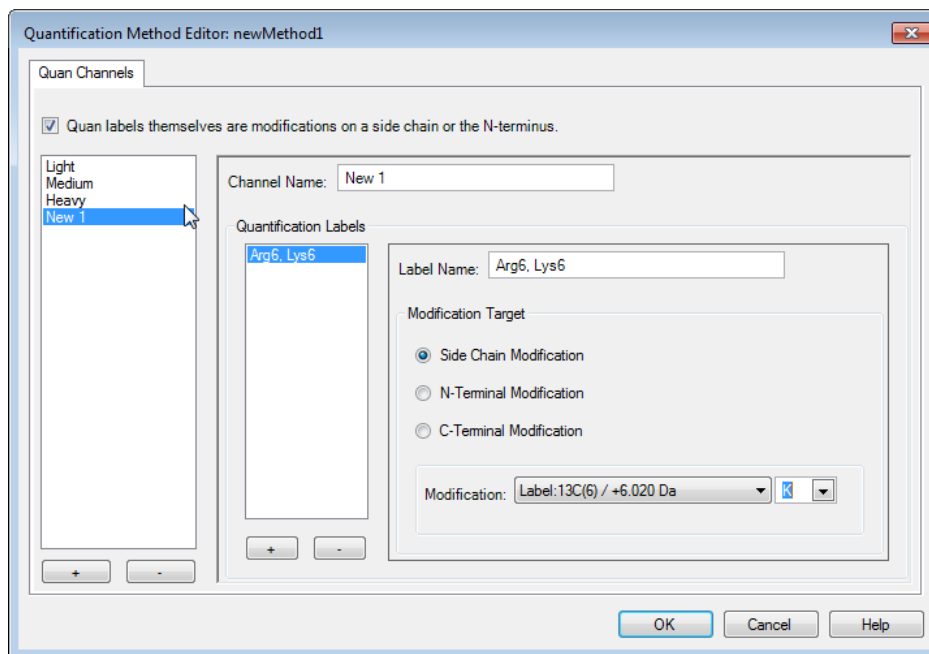
14. When you finish setting the parameters in the Quantification Method Editor, click **OK**.

The Save Quantification Method dialog box opens (see Figure 409).

Figure 409. Save Quantification Method dialog box
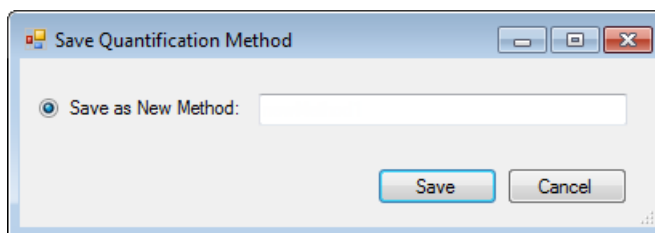
15. In the Save as New Method box, type the name of the new method.

16. Click **Save**.

The new quantification method appears in the Quantification Methods view.

❖ **To specify quantification channels for reporter ion quantification**

1. Choose **Administration > Maintain Quantification Methods**, or click the **Maintain Quantification Methods** icon, .

   The Quantification Methods view opens (see Figure 403 on page 546). It lists all of the available methods for both precursor ion and reporter ion quantification. For information about the Status column in this view, see "To specify quantification channels for precursor ion quantification" on page 545.

2. Double-click the row for the appropriate method in the Method Name or Description column, or click the column to the left of the Status column for the method and click Add .

   The Create New Quantification Method dialog box opens (see Figure 405 on page 548).

3. Select one of the following methods of creating a quantification method:

   • (Default) **From Factory Defaults**: Creates a new method using the same settings from one of the default settings that come with a new application installation.

   • **From Existing Method**: Uses the same settings as those of the existing quantification method that you select from the list. The list of methods is the same as the list given at the beginning of "Setting Up the Quantification Method" on page 543.

   • **From Scratch**: Uses one of the following templates so that you can build a new processing method from the beginning:

      – **Reporter Ion Quan Method**: Provides a template for reporter ion quantification. This method requires an additional license.

      – **Precursor Ion Quan Method**: Provides a template for precursor ion quantification.

4. Click **Create**.

   The Quantification Method Editor dialog box opens (see Figure 410).

Figure 410. Quantification Method Editor dialog box for reporter ion quantification
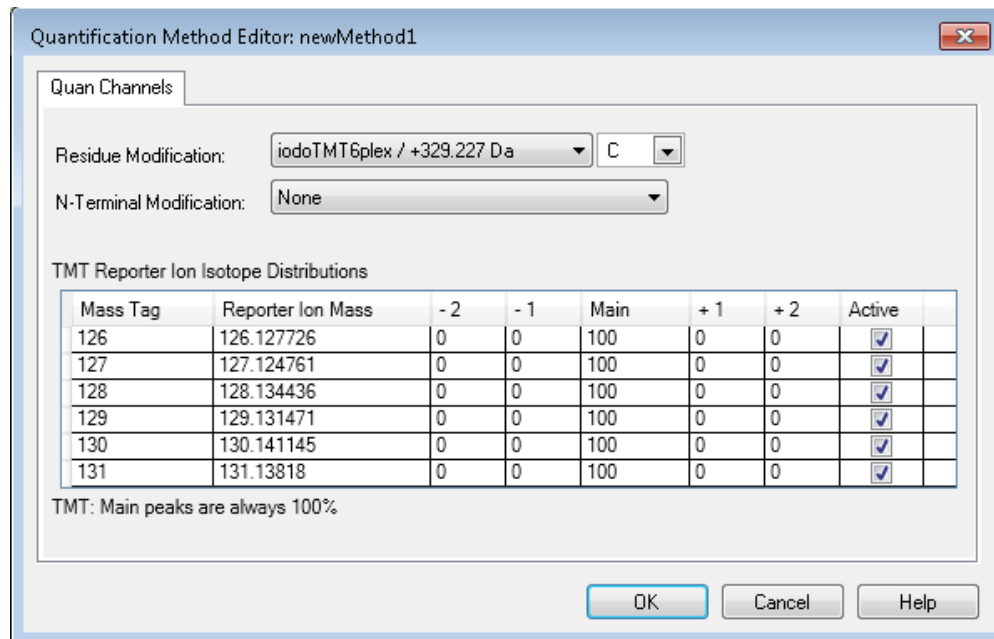
5. From the Residue Modification list, select the label modification to be found on the target amino acid residue. From the adjacent list, select the appropriate letter to indicate that the modification should occur on the indicated residue and will have an increased mass.

6. From the N-Terminal Modification list, select the label modification to be found on the N terminus of each peptide.

7. In the TMT Reporter Ion Isotope Distributions or iTRAQ Software Correction Factors area, place the isotope intensities for specific isotope shifts in the following columns:

   a. In the –2 column, type the value that is 2 Da lower than that of the reporter ion.

   b. In the –1 column, type the value that is 1 Da lower than that of the reporter ion.

   c. In the +1 column, type the value that is 1 Da higher than that of the reporter ion.

   d. In the +2 column, type the value that is 2 Da higher than that of the reporter ion.

   e. Select the check box in the Active column if you want to use that reporter ion in the method.

   f. If only some, but not all, of the tags are used—for example, all but 129—clear the active check box for that reporter ion mass.

   g. Click **OK**.

   The Save Quantification Method dialog box opens.

8. In the Save as New Method box, type the name of the new method.

9. Click **Save**.

   The new quantification method appears in the Quantification Methods view.

   To create a new quantification method to correct for isotopic impurities in TMT 10plex kits, see "Excluding PSMs with High Levels of Coisolation" on page 528.

Checking the Quantification Method

The application checks the parameters that you have set for the quantification method. For reporter ion quantification, it verifies that the method has at least two channels. For precursor ion quantification, it checks for the following:

- At least one quantification channel.

- At least one label for each quantification channel. The label can be None, which the default SILAC methods use.

- Unique label names in a channel.

- The modification of each label applied to at least one amino acid, unless you chose None for a modification.

- Each amino acid labeled only once in a channel (labels must have a defined elemental composition).

- Each label mass used only once (label masses vary by at least 1 Da).

You cannot apply changes to a quantification method unless the method meets all these criteria.

Changing a Quantification Method

You can change a quantification method before you generate the results report. You can also change a quantification method after you generate a results report, but you can make only limited changes.

Changing a Quantification Method Before Generating a Results Report

Follow these steps to change a quantification method before generating a results report.

❖ **To change an existing quantification method before generating a results report**

1. Open the Quantification Method Editor dialog box by doing one of the following:

   In the Quantification Methods area of the Study Definition page of a study, select the check box for the appropriate quantification method, and click the **Edit Quantification Method** icon, , for the selected quantification method.

   —or—

a. Choose **Administration > Maintain Quantification Methods**, or click the **Maintain Quantification Methods** icon, 🗄.

The Quantification Methods view opens (see Figure 403 on page 546). It lists all of the available methods for both precursor ion and reporter ion quantification.

b. Double-click the row for the appropriate method in the Method Name or Description column, or click the column to the left of the Status column for the method and click 🖉 Edit .

The Quantification Method Editor dialog box opens for precursor ion quantification (see Figure 406 on page 549) or for reporter ion quantification (see Figure 410 on page 553).

2. In the Quantification Method Editor dialog box, make any appropriate modifications.

For information on setting the parameters in this dialog box, see "Specifying the Quantification Channels" on page 544.

3. Click **OK**.

The application checks the parameters that you have changed to be sure that they conform to the guidelines in "Checking the Quantification Method" on page 554. It does not apply the changes to a quantification method unless the method meets all these criteria.

Changing a Quantification Method After Generating a Results Report

After you perform quantification, you can change the quantification method and reprocess the data.

❖ **To change a quantification method after generating a results report**

1. On the Input Files page of the study, choose a new quantification method in the Quan Method box.

   –or–

   Open the Quantification Method Editor by doing the following.

   a. In the Quantification Methods area of the Study Definition page, click the **Edit Quantification Method** icon, 🖉, for the selected quantification method.

      –or–

      Choose **Administration > Maintain Quantification Methods**, or click the **Maintain Quantification Methods** icon, 🗄, and then double-click one of the methods in the Method Name column of the Quantification Methods view.

   b. In the Quantification Method Editor dialog box, add, or change the modifications for the preexisting channels.

      However, you cannot define mass tags or labels as when you set up the initial quantification method, because they have already been measured.

2. Click the **Analysis Results** tab.

3. Click 🔄 Reprocess ˑ.

4. Choose **All Analysis Steps**.

5. In the Analysis window, click 🔄 Run .

Removing a Quantification Method

You can delete a quantification method if it is no longer useful, or make a quantification method temporarily unavailable to new workflows.

❖ **To remove a quantification method**

1. Choose **Administration > Maintain Quantification Methods**, or click the **Maintain Quantification Methods** icon, 🗄️ .

   The Quantification Methods view opens (see Figure 403 on page 546). It lists all of the available methods for both precursor ion and reporter ion quantification.

2. Select the row of the method that you want to remove.

   The Remove button becomes available.

3. Click ✖ Remove .

4. In the Delete Methods dialog box, click **OK**.

❖ **To deactivate a quantification method**

1. Choose **Administration > Maintain Quantification Methods**, or click the **Maintain Quantification Methods** icon, 🗄️ .

   The Quantification Methods view opens (see Figure 403 on page 546). It lists all of the available methods for both precursor ion and reporter ion quantification.

2. Clear the check box in the Is Active column on the line containing the quantification method that you want to make inactive.

   To make the quantification method active again, select the same check box.

Importing a Quantification Method

You can import a new quantification method from another computer.

❖ **To import a quantification method**

1. Choose **Administration > Maintain Quantification Methods**, or click the **Maintain Quantification Methods** icon, 🗄️ , either in the toolbar or on the Administration page.

   The Quantification Methods view opens (see Figure 403 on page 546). It lists all of the available methods for both precursor ion and reporter ion quantification.

2. Click Import .

3. In the Import Quan Method dialog box, select the METHOD file containing the method that you want to import, and click **Open**.

   • If the new method is valid, the Quantification Method Editor dialog box opens, showing the new method.

   • If the new method is not valid, a message box opens that describes the error.

4. For a valid method, click **OK** in the Quantification Method Editor dialog box.

5. Change the name of the imported quantification method by changing it in the Method Name column of the Quantification Methods view.

Exporting a Quantification Method

You can save a quantification method to use on another computer.

❖ **To export a quantification method**

1. Choose **Administration > Maintain Quantification Methods**, or click the **Maintain Quantification Method**s icon, , either in the toolbar or on the Administration page.

   The Quantification Methods view opens (see Figure 403 on page 546). It lists all of the available methods for both precursor ion and reporter ion quantification.

2. Select the method that you want to export in the Quantification Methods view by clicking in the column to the left of the Status column.

3. Click Export .

4. In the Export Quan Method dialog box, browse to the folder where you want to store the METHOD file containing the quantification method, and click **Save**.

   The application saves the method in a file with the name of the method and a .method suffix, for example, TMT 2plex.method.

Restoring Quantification Method Template Defaults

If you have altered one of the quantification method templates listed at the beginning of "Setting Up the Quantification Method" on page 543, you can restore the original template in effect when the application was newly installed.

❖ **To restore the original template**

1. Choose **Administration > Maintain Quantification Methods**, or click the **Maintain Quantification Methods** icon, .

   The Quantification Methods view opens (see Figure 403 on page 546). It lists all of the available methods for both precursor ion and reporter ion quantification.

2. To open the Quantification Method Editor dialog box, click ➕ Add in the Quantification Methods view.

   The Create Quantification Method dialog box opens (see Figure 405 on page 548).

3. Select the appropriate template from the Create from Factory Defaults list.

4. Set up the quantification method according to the instructions in "Setting Up the Quantification Method" on page 543.

Searching for Quantification Modifications with Mascot

When you use the Mascot node as the search engine in a quantification workflow, you can set the dynamic and static modifications as parameters. For samples with isotopic labels and several PTMs, you might need to specify several dynamic modifications usable within a single search, but the current number that you can specify is limited to nine.

To avoid this limitation, you can configure quantification methods on the Mascot server. In a quantification method, modifications are organized into groups classified as fixed, variable, or exclusive. You can define modification groups as variable or exclusive at the component level, where they usually characterize the component. You can also define them at the method level, but only as fixed or variable. Defining modifications at the method level is convenient for modifications that are important to the method and saves having to choose them in the Workflow Editor. Exclusive groups are effectively a choice of fixed modifications, so the restrictions that apply to fixed modifications also apply to them.

With the Mascot node, you can use the modification groups specified as part of a quantification method on the Mascot server. You can use the node's From Quan Method parameter in the Parameters pane to select the dynamic modifications to search for, rather than manually specifying each modification with a Dynamic Modifications parameter.

In the editor in the Mascot server window, you can specify that these groups be variable, fixed, or exclusive. You can also define them directly for the method in report ion quantification or for each component in precursor ion quantification.

❖ **To specify the quantification modifications to search for**

1. Choose **Administration > Configuration > Mascot**, and configure Mascot by following the instructions in "Configuring the Mascot Search Engine" on page 26. In the Mascot Server URL box, be sure that you enter the URL of the Mascot server to be used for Mascot searches.

2. Set up a processing workflow that includes, at a minimum, the nodes shown in these workflows:

   • For reporter ion quantification, see the workflow in Figure 282 on page 388.

   • For label-free quantification, see the workflow in Figure 367 on page 493.

   • For precursor ion quantification, see the workflow Figure 333 on page 450.

Follow the instructions to set up the applicable workflow.

3. For the search engine node, drag the **Mascot** node to the workflow, and connect it to the other nodes.

4. Set the parameters of the Mascot node.

5. Select the dynamic modifications to search for from the following:

   • Select a dynamic modification from the list in each Dynamic Modification parameter.

   You can select up to nine modifications.

   –or–

   • Click the **From Quan Method** parameter in the Parameters pane under Modification Groups, and from the list (see Figure 411 for an example), select the modifications that you want to search for.

   You can select more than nine modifications.

   > **Note** Do not use the modifications that you specify as part of the modification groups in the selected quantification method as additional dynamic or static modifications.

Figure 411. From Quan Method list

6. (Optional) If you want to group these modifications, go to the editor in the Mascot server window and choose **Configuration Editor > Quantitation**.

   After you group the modifications, you can define them as fixed, variable, or exclusive. You can also define them directly for the method in reporter ion quantification or for each component in precursor ion quantification. Refer to the Mascot documentation for information on grouping modifications and defining the groups.

For the final search results, whether you explicitly specify a modification as dynamic or static or indirectly specify a modification from the chosen quantification method does not matter. As an exception, when you select an exclusive modification group, Mascot modifies all or none of the affected residues of a peptide sequence. Therefore, peptide matches with inconsistent labeling no longer occur.

> **Note** Using a Mascot quantification method to retrieve the modification groups to use does not affect how the application performs the quantification. The application itself exclusively performs the quantification. You must specify in the application's methods any quantification labels used for the quantification.

Troubleshooting Quantification

The following procedures can help you obtain optimal results when performing quantification.

❖ **To troubleshoot reporter ion quantification**

If you obtain unexpected quantification results, verify that all node settings in your processing workflow are reasonable.

– Make sure that the Integration Tolerance parameter of the Reporter Ions Quantifier node fits the data that you are processing. The default is 20 ppm, which is too low if you are processing PQD data from an ion trap.

– Make sure that the settings of the Mass Analyzer, MS Order, and Activation Type parameters of the Reporter Ions Quantifier node are correct for the data that you are processing. Figure 412 shows the typical settings to use if you want to quantify HCD scans from the Orbitrap mass spectrometer.

Figure 412. Typical settings for quantifying iTRAQ or iodo TMT 6plex tags from HCD scans
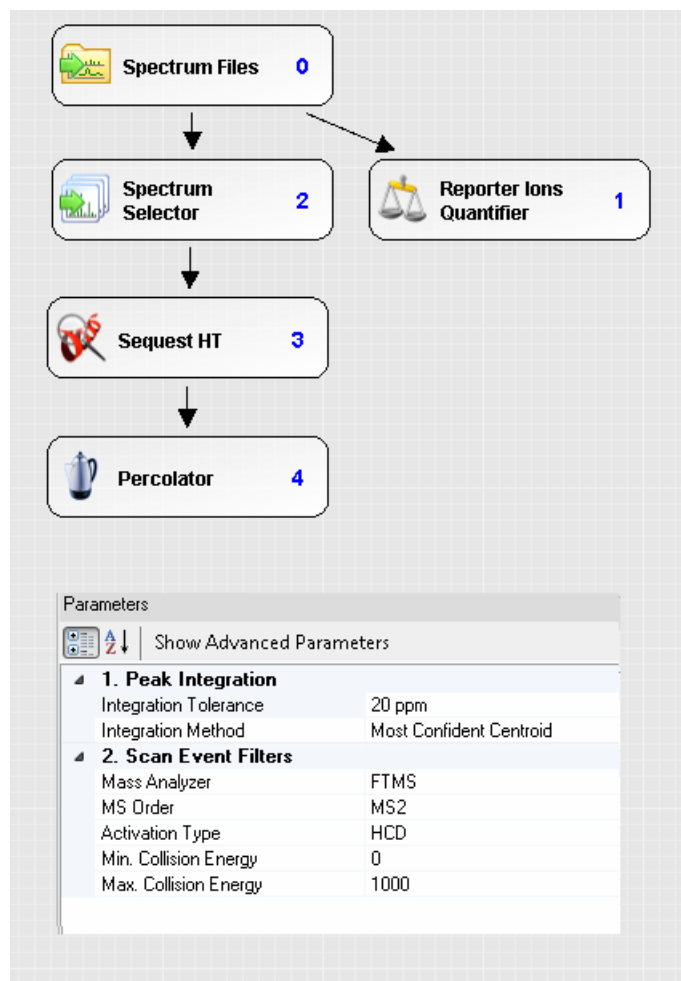
To quantify PQD scans from an ion trap, use the typical settings shown in
Figure 413.

Figure 413. Typical settings for quantifying ITRAQ or TMT tags from ion trap PQD scans
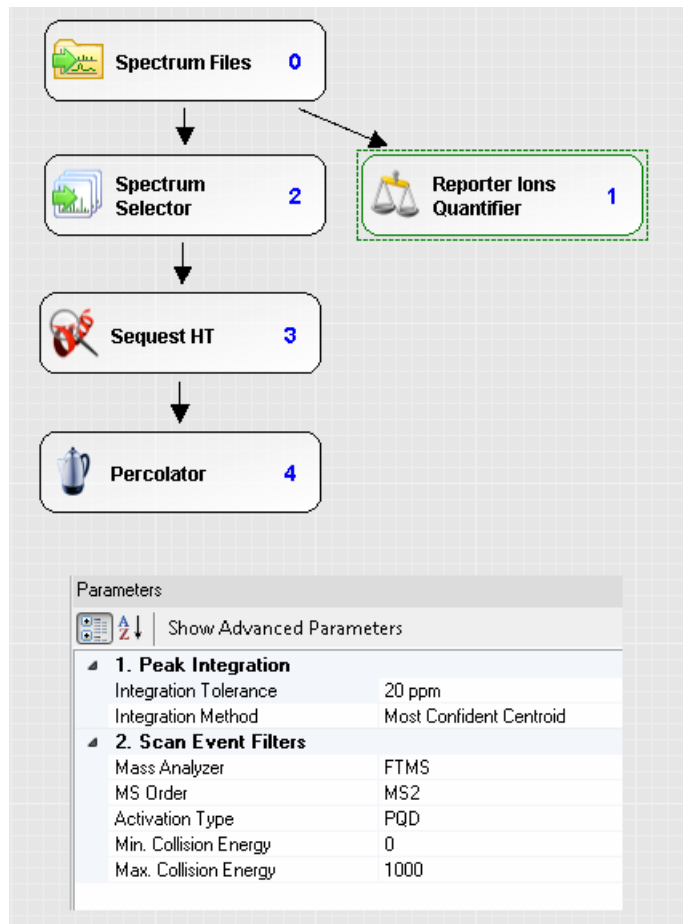
Make sure that you use the correct set of static and dynamic modifications for the
search engine. For example, if you are searching iodo TMT 6plex data with Sequest
HT, check that your settings resemble those in Figure 414.

Figure 414. Modifications required for searching iodo TMT 6plex samples

Figure 414. Modifications required for searching iodo TMT 6plex samples

Parameters

Show Advanced Parameters

4. Dynamic Modifications
Max. Equal Modifications Per Pept	3
1. Dynamic Modification	Phospho / +79.966 Da (S, T)
2. Dynamic Modification	None
3. Dynamic Modification	None
4. Dynamic Modification	None
5. Dynamic Modification	None
6. Dynamic Modification	None

5. Dynamic Modifications (peptide terminus)
1. N-Terminal Modification	None
2. N-Terminal Modification	None
3. N-Terminal Modification	None
1. C-Terminal Modification	None
2. C-Terminal Modification	None
3. C-Terminal Modification	None

6. Dynamic Modifications (protein terminus)
1. N-Terminal Modification	None
2. N-Terminal Modification	None
3. N-Terminal Modification	None
1. C-Terminal Modification	None
2. C-Terminal Modification	None
3. C-Terminal Modification	None

7. Static Modifications
Peptide N-Terminus	TMT6plex / +229.163 Da (Any N-Terminus)
Peptide C-Terminus	None
1. Static Modification	Carbamidomethyl / +57.021 Da (C)
2. Static Modification	TMT6plex / +229.163 Da (K)
3. Static Modification	None
4. Static Modification	None
5. Static Modification	None

❖ **To troubleshoot precursor ion quantification**

If you obtain unexpected precursor ion quantification results, verify that all settings of your processing workflow are reasonable.

- Check the dynamic modification parameters in the Mascot or Sequest HT search engines. Ensure that they match your isotope labeling sample.

- Check the node parameters that you set before performing the quantification to see if they are appropriate for your sample.

  See "Performing Precursor Ion Quantification" on page 411 for more information.

- Verify that your isotopic labeling is one of the following options in the protein ID/search node (either SequestHT or Mascot):

  – SILAC 2plex (Arg10, Lys6): Uses arginine 10 and lysine 6.

  – SILAC 2plex (Arg10, Lys8): Uses arginine 10 and lysine 8.

  – SILAC 2plex (Ile6): Uses isoleucine 6.

  – SILAC 3plex (Arg6, Lys4|Arg10, Lys8): Uses arginine 10 and lysine 8 for "heavy" labels and arginine 6 and lysine 4 for "medium" labels.

  – SILAC 3plex (Arg6, Lys6|Arg10, Lys8): Uses arginine 10 and lysine 8 for "heavy" labels and arginine 6 and lysine 6 for "medium" labels.

  – Dimethylation 3plex: Chemically adds isotopically labeled dimethyl groups to the N-terminus and to the ε-amino group of lysine.

  – ^{18}O labeling: Introduces 2 or 4 Da mass tags through the enzyme-catalyzed exchange reaction of C-terminal oxygen atoms with ^{18}O.

  Note Low-mass accuracy MS 1 full-scan data cannot be used for precursor ion quantification.

- Check your tolerance window. If you get too many results, decrease the size of the window. For too few results, increase the size of the window.

- Make sure you chose the right database.

- Check the species listed to make sure the samples came from that species.

- Verify that the activation type used is correct.

- Verify that the instrument type in the Mascot search engine is correct.

- Use only the ETD Spectrum Charger node for low-mass resolution ETD data.

Interpreting Quantitative Results

You can view the results of quantification by using the following charts, data maps, and pages of the .pdResult file.

Contents

Displaying the Trend Chart

The Trend chart shows the abundance of proteins or peptide groups across all sample groups. It displays the data as a trend line for each protein or peptide group. You define the sample groups by selecting one or more check boxes under Group By in the collapsible pane to the left of the chart.

The .pdResult file must contain multiple quantification values if you want to display a trend chart.

❖ **To display a Trend chart**

1. Open the .pdResult file of interest.

2. Click the **Proteins** or **Peptide Groups** tab.

3. Click the protein or peptide group of interest.

4. Choose **View > Trend Chart**, or click the **Trend Chart View** icon, [icon].

   The Trend Chart opens as a docked window at the bottom of the result pages (see Figure 415). This example shows five groups of mitochondrial proteins increasing in relative abundance over time.

Figure 415. Trend chart
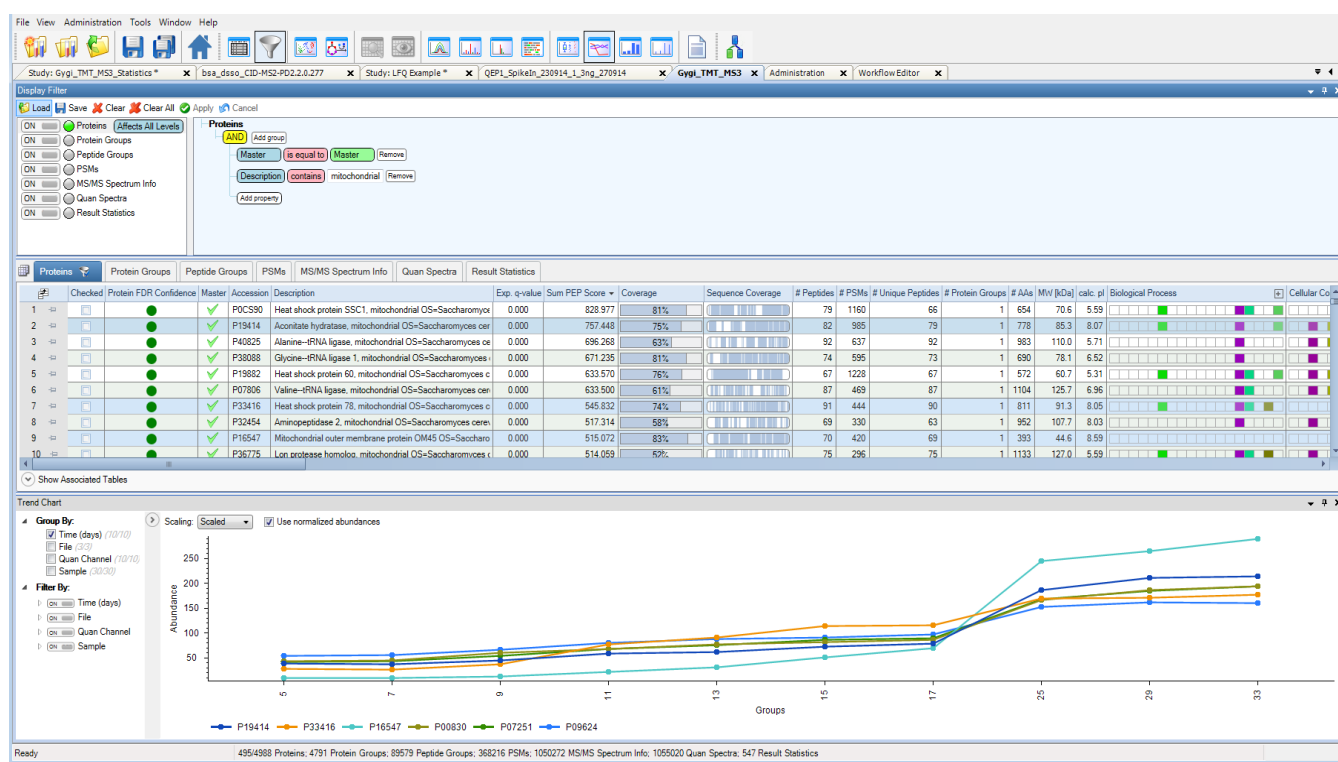
If the active result table does not contain a quantification results, the following text appears in the graph area:

   No Results Available to Plot.

To plot the peak areas for more proteins or peptide groups, press the CTRL key and select one or more rows in the results table, taking care to avoid clicking an editable table cell.

5. (Optional) In the Group By area at the far left, select the appropriate check boxes to specify how to create sample groups from the data:

   • ***Study_factors***: Creates sample groups according to the study factors that you defined in the study.

- **File**: Creates sample groups according to file.

- **Quan Channel**: Creates sample groups according to quantification channel.

- **Sample**: Creates sample groups according to the sample type study variable.

6. (Optional) In the Filter By area at the far left, specify how to filter the data. Set the type of filter to **On**, click the arrow to the left of the On/Off filter box, and select the appropriate study factors, input files, quantification channels, or samples.

  - *Study_factor*: Filters by study factor, for example, Time (days).

  - File: Filters by input file.

  - Quan Channel: Filters by quantification channel.

  - Sample filter: Filters by sample.

7. (Optional) Select the **Use Normalized Abundances** check box to display the normalized abundances.

  The check box is available if the Reporter Ions Quantifier node or the Precursor Ions Quantifier node normalized the quantification data because you set the Normalization Mode parameter to Total Peptide Amount or Specific Protein Amount.

  If you do not select the Use Normalized Abundances check box, the application displays the absolute abundance values at each study factor point.

8. (Optional) From the Scaling menu, choose a value by which to scale the abundance values represented by the trend line.

  This value resets the height of the trend line and changes the y axis. Especially when you select a number of proteins to view in the Trend chart, rescaling makes it easier to compare the trend lines and therefore to interpret the results.

  - **Unscaled**: Uses absolute abundance values and displays the raw abundance values at each study factor point.

  - **Scaled**: Displays scaled abundance values.

  - **Scaled to x**: Scales the points in the trend line to a selected study factor.

  If you selected Show Legend from the shortcut menu (right-click the Trend chart), the legend displays the name and color of each selected protein or peptide group.

Trend Chart Parameters

Table 31 describes the parameters on the Trend chart.

Table 31. Trend Chart parameters (Sheet 1 of 2)

Parameter	Description
Group By	Creates sample groups according to the following variables: • *Study_factor*: Creates sample groups according to the study factors that you defined in the study. • File: Creates sample groups according to file. • Quan Channel: Creates sample groups according to quantification channel. • Sample: Creates sample groups according to the sample type study variable.
Filter By	Filters the data according to the following filter types: • *Study_factor*: Filters by study factor, for example, Time (days). • File: Filters by input file. • Quan Channel: Filters by quantification channel. • Sample filter: Filters by sample.
Scaling	Determines whether and how the application scales the abundance values in the trend line. • Unscaled: Uses absolute abundance values and displays the raw abundance values at each study factor point. • Scaled: Displays scaled abundance values. • Scaled to *x*: Scales the points in the trend line to a selected study factor.

Table 31. Trend Chart parameters (Sheet 2 of 2)

Parameter	Description
Use Normalized Abundances	Determines whether the chart displays normalized abundances. The check box is available if the Reporter Ions Quantifier node or the Precursor Ions Quantifier node normalized the quantification data because you set the Normalization Mode parameter to Total Peptide Amount or Specific Protein Amount. If you do not select the Use Normalized Abundances check box, the application displays the absolute abundance values at each study factor point. • Selected: Displays normalized abundances. • Clear: Displays the absolute abundance values at each study factor point.
Groups (*x*-axis)	Displays the groups selected in the study—that is, the available study factors when the run starts.
Abundances (*y*-axis)	Displays the abundance of proteins or peptide groups.

Displaying the Report Item Distribution Chart as a Sample Abundances Chart

The Sample Abundances chart displays as a box-and-whisker plot the abundance of the peak intensity values for all items currently displayed on the Proteins page or the Peptide Groups page.

A box-and-whisker plot displays the data for a variable as a rectangular box with a set of lines at each end (see Figure 416). The line through the rectangle represents the median value in the data set. The lower portion of the rectangle represents the data points that fall within the second quartile and the upper portion represents the data points that fall within the third quartile. The circles that fall outside the fence lines are outliers.

Figure 416. Box-and-whisker plot of the data distribution
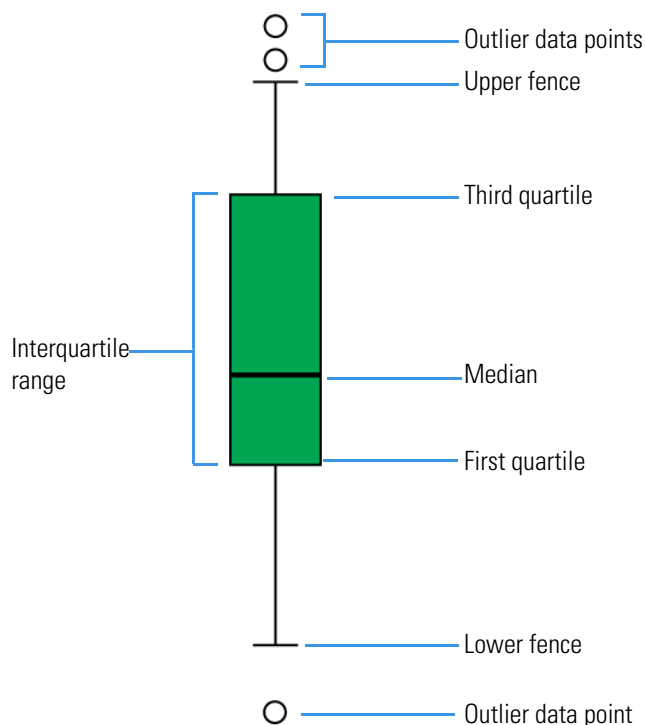
Fences and whiskers are usually slightly different in a box-and-whisker plot. Fences are never drawn and are normally calculated as follows:

Interquartile range (IQR) = Quartile 3 (Q3) – Quartile 1 (Q1)

Upper fence = min(Q3 + IQR × 1.5, maximum value)

Lower fence = max(Q1 – IQR × 1.5, minimum value)

> **Note** To calculate the quartiles, the application uses a method that is similar to the type 6 method in the R statistical computing software.

Whiskers are drawn up to the last point between the quartile and the fence—that is, to the maximum value between Q3 and the upper fence or the minimum value between Q1 and the lower fence (see Figure 417). They would only extend to the fence values if there were observations equal to the fence values; otherwise, the whiskers extend to the most extreme observations that lie within the fences.

Figure 417. Whiskers
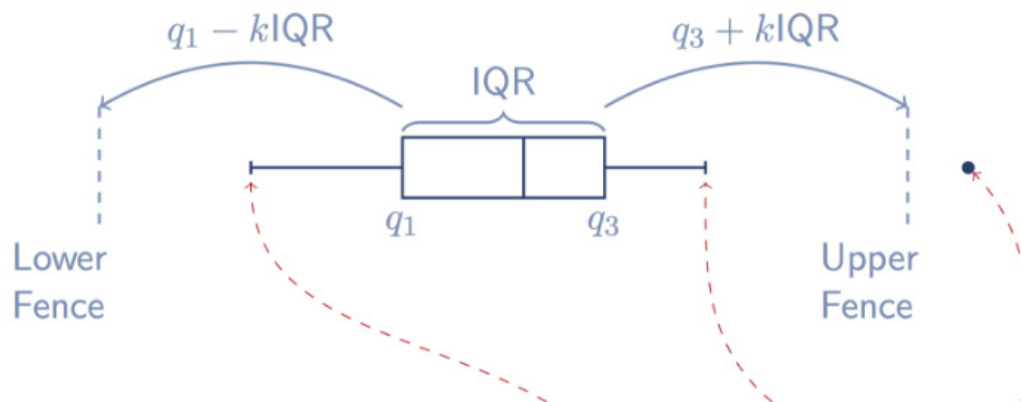
- The "whiskers" extend to the smallest and largest observations that are not outliers.
- Observations that are smaller than the lower fence or larger than the upper fence are identified as dots

By default, the Sample Abundances chart displays the data for all the samples and duplicates the grouping on the Grouping & Quantification page of the analysis. Each group appears in a different color. The legend shows the colors of the sample groups.

Use the Group By check boxes in the collapsible pane to change the grouping. Use the Filter By check boxes to remove samples from the plot.

❖ **To display the Report Item Distribution chart as a Sample Abundances chart**

1. Open the .pdResult file of interest.

2. Click the **Proteins** or **Peptide Groups** tab.

3. Choose **View > Distribution Charts** to open the Report Item Distribution chart.

4. Click the **Sample Abundances** tab.

5. (Optional) In the Group By area at the far left, select one or more of the following check boxes to create sample groups according to the study factors for quantification:

    - *Study_factors*: Lists each study factor that you defined in the study.

    - **Sample**: Selects the sample mixture of all samples. The Sample check box is based on the Sample Type default study variable on the Grouping & Quantification page. It appears whether or not you select Sample Type on the Grouping & Quantification page.

6. (Optional) In the Filter By area at the far left, set one or more of the following variables to filter the data:

    a. Set the *Study_factor* filter to **On**.

b. Set the Sample filter to **On**.

c. Click the arrow to the left of the On/Off filter box for *Study_factor* or Sample.

d. Select the appropriate check boxes beneath *Study_factor*, Sample, or both.

7. From the Data Source list, select the result category to plot the data from, either **Proteins** or **Peptide Groups**.

8. (Optional) To display the data in linear scale, clear the **Log-Transform Data** check box.

    If you do not select this check box, the application displays the data in linear scale.

9. (Optional) To display the normalized chromatographic peak areas, select the **Use Normalized Areas** check box.

    The check box is available if the Reporter Ions Quantifier node or the Precursor Ions Quantifier node normalized the quantification data but did not scale it.

    If you do not select this check box, the application displays the chromatographic peak areas without normalization.

    Figure 418 gives an example of a Sample Abundances chart. It shows the abundance of the peptides within the proteins in the samples.

Figure 418. Sample Abundances page of the Report Item Distribution chart
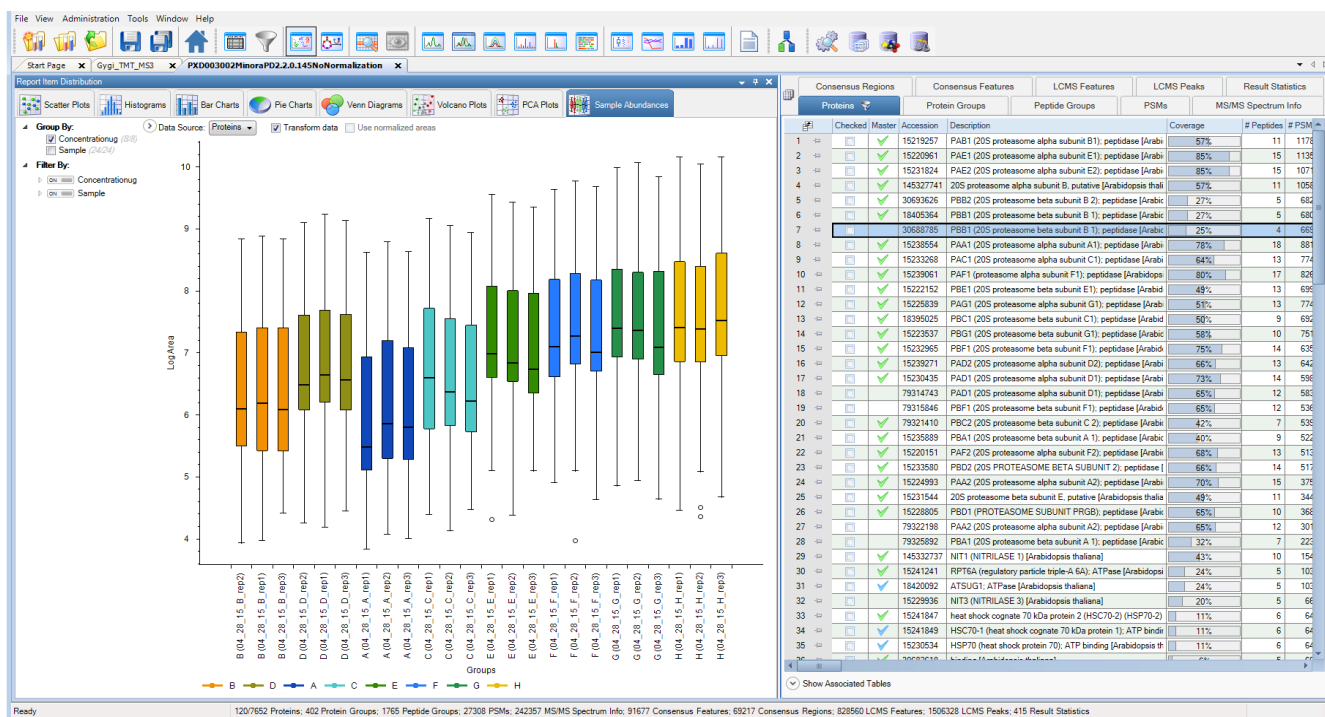
❖ **To display the legend**

Right-click the plot and choose **Show Legend**.

❖ **To display ToolTips**

Right-click the view and choose **Show Position Tooltips**.

The ToolTips display the information in Table 32.

❖ **To zoom in**

Drag your cursor to the right and select the area to enlarge in size.

❖ **To zoom out**

• Drag your cursor to the left and select the area to reduce in size.

–or–

• Right-click the view and choose **Zoom Out** from the shortcut menu.

❖ **To return to the default view**

Right-click the chart and choose **Undo All Zoom/Pan**.

❖ **To copy or save the data as an image**

Right-click the plot and choose **Copy** to copy an image to the Clipboard.

–or–

Right-click the plot and choose **Save As** to save the data to an image file.

❖ **To copy or save the data as editable text**

Right-click the plot and choose **Copy Information to Clipboard** to copy the text to the Clipboard.

You can paste this text to Notepad, an Excel spreadsheet, and so on. The data appears in columnar format.

–or–

Right-click the plot and choose **Save Information As** to save the data to a tab-delimited CSV file.

The data appears as a single column. After you transform the tab-delimited text to columns in a spreadsheet application, the spreadsheet contains two sets of columns. The first set consists of these columns from left to right: Groups, Name, Minimum Value, Maximum Value, Std. Deviation, Mean, Median, Q1 Value, Q2 Value, and Q3 Value. The second set lists the outlier data points and consists of these columns from left to right: Groups, Name, and Outlier.

❖ **To view the table row for an outlier data point**

In the plot, double-click the data point.

❖ **To select the check box for an outlier data point**

In the plot, right-click the data point and choose **Check Point**.

In the result table, the check box is selected for the corresponding compound or expected compound.

❖ **To clear the check box for an outlier data point**

Right-click the data point in the volcano plot, and choose **Uncheck Point**.

❖ **To select the check boxes for all of the visible data points**

Right-click the plot and choose **Check All Visible Points**.

❖ **To clear all visible data points**

Right-click the plot and choose **Uncheck All Visible Points**.

Parameters in the Sample Abundances View of the Report Item Distribution Chart

Table 32 lists the parameters in the Sample Abundances view of the Report Item Distribution chart.

Table 32. Sample Abundances view parameters (Sheet 1 of 3)

Parameter	Description
Group By	Creates sample groups according to the study factors for quantification:
	• *Study_factors*: Lists each study factor that you defined in the study.
	• Sample: Selects the sample mixture of all samples. The Sample check box is based on the Sample Type default study variable on the Grouping & Quantification page. It appears whether or not you select Sample Type on the Grouping & Quantification page.
Filter By	Filters the data according to the following variables:
	• *Study_factors*: Filters by the study factor that you defined in the study.
	• Sample: Filters by sample (see the description of Sample for the Group By parameter).
Data Source	Specifies the result category to plot the data from:
	• Proteins: Plots from protein data.
	• Peptide Groups: Plots from peptide group data.

Table 32. Sample Abundances view parameters (Sheet 2 of 3)

Parameter	Description
Transform Data	Determines whether the data appears in a linear scale or the $\log_{10}$ scale.
	Selecting this check box transforms the area counts to the $\log_{10}$ scale.
Use Normalized Areas	Displays normalized chromatographic peak areas.
	The check box is available if the Reporter Ions Quantifier node or the Precursor Ions Quantifier node normalized the quantification data but did not scale it.
Graph	
Groups (*x* axis)	Displays the name of the sample groups.
LogArea/Area (10^9) (*y* axis)	Displays the area in a linear scale or in a $\log_{10}$ scale, depending on the setting of the Log-Transform Data parameter. • LogArea: Displays the peak area in $\log_{10}$ scale. • Area (10^9): Displays the peak area in linear scale.
Rectangles	The rectangles represent the second and third quartiles for the data set. When you choose the Show Position ToolTips command, placing the cursor over the rectangle displays the following information: • Name: The name of the raw data file that the sample group belongs to. • Group: The name of the sample group. • Max: The maximum value of the chromatographic peak area within the sample group, including the outliers. • 3rd Q: The third quartile of the chromatographic peak area within the sample group. • Median: The median of the chromatographic peak area within the sample group. • Mean: The mean of the chromatographic peak area within the group. • 1st Q: The first quartile the chromatographic peak area within the sample group. • Min: The minimum value of the chromatographic peak area within the sample group, including the outliers. • StdDev: The standard deviation of the chromatographic peak area within the sample group.

Table 32. Sample Abundances view parameters (Sheet 3 of 3)

Parameter	Description
Circles	Represent outlier data points.
	When you choose the Show Position ToolTips command, pointing to a data point displays the following information:
	• *Accession_number*: The accession number.
	• MW *number* [kDA]=*number*: The calculated molecular weight of the protein.
	• Calc pI=*number*: The theoretically calculated isoelectric point for the protein—that is, the pH at which a particular molecule carries no net electrical charge.
	• # Peptides=*number*: The number of distinct peptide sequences in the protein.
	• # PSMs=*number:* The number of identified peptide spectrum matches identified from all included searches, including those redundantly identified.
	• #AAs=*number*: The length of the protein sequence.
Legend	Displays the group colors.
	By default, the legend is hidden. Choose Show Legend from the shortcut menu to display it.

Displaying the Quantification Page of the Result Summaries

For information on the Quantification page of the Result Summaries pane, refer to the Help.

Displaying Data Distribution Maps for Quantification

In workflows that include quantification nodes, you can use the Reporter Ions Quantifier node or the Precursor Ions Quantifier node in the consensus workflow to display ratio, standard error, abundances, and count columns as data distribution maps in .pdResult reports. These maps show the distribution of values across sample groups and ratios. They can help you visually validate the results of the quantification.

You can also use the Data Distributions node in the consensus workflow to display the Found in Samples, Found in Sample Groups, Found in Files, and Found in Fractions columns in .pdResult reports as data distribution maps. These maps show the distribution of values across the available files, samples, and sample groups.

The application combines all values for each generated group of quantification values (ratios, standard errors, abundances, and counts) into one data distribution map. You can see all grouped ratios in relation to the ratios of the individual replicates and the associated variabilities.

For detailed information about the Reporter Ions Quantifier node, refer to the Help. For detailed information about the Precursor Ions Quantifier node, refer to the Help.

Sample Information Used to Calculate and Display Quantification Results

When the application processes the data, it first extracts the raw quantification values during the processing step. For reporter ion quantification, these values are the intensities of the reporter peaks for each MSn spectrum. For precursor ion quantification, these values are the integrated peak areas or height. In the consensus workflow, the Reporter Ions Quantifier node or the Precursor Ions Quantifier, respectively, refers to the information generated during the analysis setup to determine which ratios to calculate for each file and which of these ratios to group by calculating an average ratio from them.

The ratio changes are currently color-coded. The area values are also color-coded, with the color ranging from yellow to red with increasing area.

Figure 419 shows the data distribution maps of the found ratio changes for the specified sample groups.

Figure 419. Distribution maps with color-coded ratio change for the specified sample groups

You can use the *n*th Fold Change Threshold parameters of the Reporter Ions Quantifier node or the Precursor Ions Quantifier node to adjust the thresholds of the color highlighting in the Abundance and Abundance Ratios (log2) column of the Proteins page. For more information on these parameters and the color highlighting, refer to the Help.

The display filters provide a means to filter data on the basis of the data distribution maps (for more information on display filters, see "Filtering with Display Filters" on page 225). The example in Figure 420 filters for the proteins that have a fold change of at least four in at least five of the ratio groups present.

Figure 420. Display filters applied to the distribution map columns
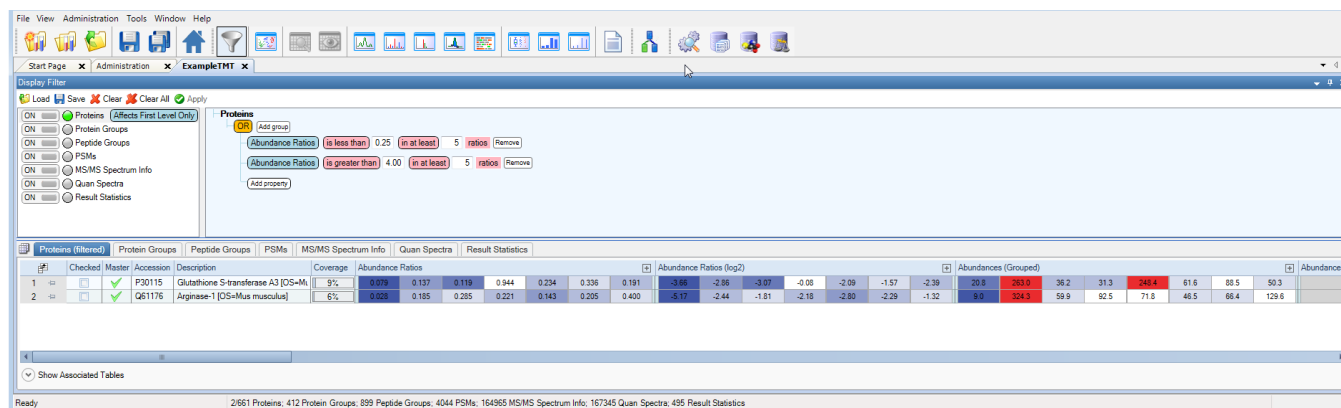
Sample Information Used to Display Identifications and Quantifications Among Files and Samples

The application uses the information about the grouping of samples in the data distribution maps to show which protein group, protein, or peptide was identified in which file, sample, and sample group (see Figure 421). The color-coding refers to the confidence level of the identification—that is, the confidence of the best PSM found for the particular file, sample, or sample group.

Figure 421. Data distribution maps with color-coded identification confidence levels for proteins in sample groups, samples, and files

In precursor ion identification, there is frequently no confidence information from an identified MS*n* spectrum, but the application found a quantification value for a particular sample. For example, it identified only the light peptide in a SILAC experiment, but it also found an integrated extracted ion current (XIC) chromatogram for the heavy peptide. The application displays such cases in blue on the distribution map.

Generating Data Distribution Maps

❖ **To generate a distribution map**

1. Create or open a study and an analysis as follows:

   - See "Creating a Study" on page 40.

   - See "Opening an Existing Study" on page 40.

   - See "Creating an Analysis" on page 65.

   - See "Opening an Existing Analysis" on page 66.

2. Create a quantification processing workflow. See "Creating a Consensus Workflow for Reporter Ion Quantification" on page 390, "Creating a Processing Workflow for Label-Free Quantification" on page 492, or "Creating a Processing Workflow for Precursor Ion Quantification" on page 449.

3. Create a quantification consensus workflow. See "Creating a Consensus Workflow for Reporter Ion Quantification" on page 390, "Creating a Consensus Workflow for Label-Free Quantification" on page 495, or "Creating a Consensus Workflow for Precursor Ion Quantification" on page 452.

4. Set the appropriate parameters of the Reporter Ions Quantifier node or the Precursor Ions Quantifier node to display the appropriate columns in the output .pdResult file.

    For information about these parameters, refer to the Help.

5. (Optional) To have the .pdResult report display the Found in Samples, Found in Sample Groups, Found in Files, and Found in Samples columns, drag the **Data Distributions** node to the to the Post-Processing area of the Workflow Editor.

6. (Optional) Save the workflow:

    a. In the Name box above the Workflow Tree pane, type a name for the consensus workflow.

    b. (Optional) In the Description box, type a brief description of the consensus workflow.

    c. In the Workflow Editor, click 🔧 Save or 🔧 Save Common .

    d. In the Save Workflow dialog box, do the following:

        i.   Browse to the file to save the template in, or type a new name for the file.

        ii. In the Save As Type list, select **Consensus Workflow File (*.pdConsensusWF)**.

        iii. Click **Save**.

    The application saves the workflow in the *file_name*.pdConsensusWF file.

7. Save the analysis. See "Saving an Analysis" on page 83.

8. Save the study. See "Saving a Study" on page 56.

9. Click 🔄 Run at the top right of the Analysis window.

    The job queue appears (see Figure 53 on page 83), displaying the status of your search as it progresses. For job queue information, refer to the Help.

Displaying Data Distribution Maps

❖ **To display a distribution map**

1. Open the .pdResult file and use the Field Chooser to make the appropriate columns visible in the .pdResult file.

2. Click the **Protein Groups**, **Proteins**, or **Peptide Groups** tab, as appropriate.

3. In the Field Chooser, select the check boxes of the appropriate data distribution columns. Examples of these columns include the following:

   - Found in Samples

   - Found in Files

   - Found in Sample Groups

   - Found in Fractions

   These columns display the best confidence of the PSMs of the protein that the application identified in the files.

   The colors of the boxes in the Found in Samples, Found in Sample Groups, Found in Files, and Found in Fractions columns represent the following:

   – Green: High confidence

   – Yellow: Medium confidence

   – Red: Low confidence

   – Blue: Found but unidentified PSM. Only the results from precursor ion quantification searches contain blue boxes.

   Each column pertaining to data distribution features an Expand icon, ⊞ , to the right of the column name (see Figure 422).

Figure 422. Data Distribution columns marked by Expand icons
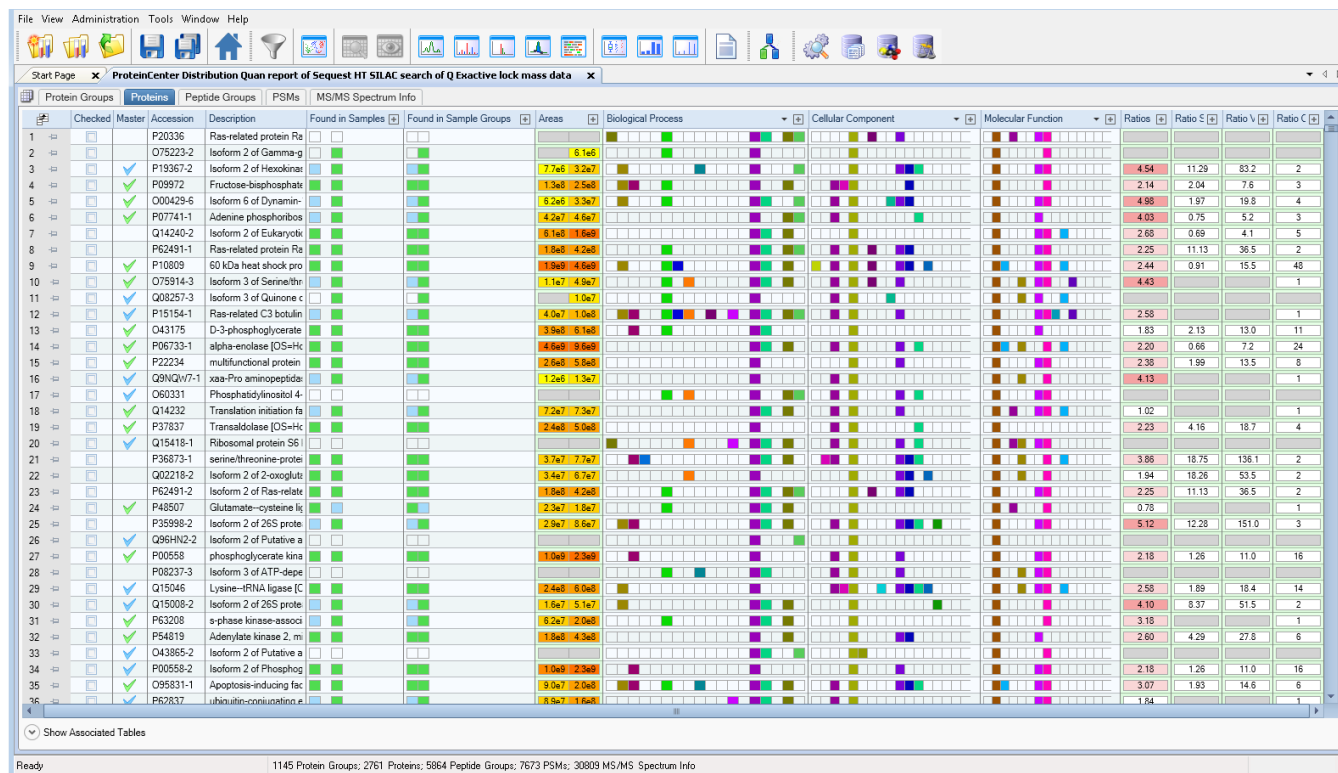
4. (Optional) In the appropriate columns pertaining to data distribution, click the **Expand** icon, ⊞ , to open a legend that displays the identification number of the input file (shown in the ID column on the Input Files page) and the quantification label.

Figure 423 shows the legends for the Found in Samples, Found in Sample Groups, Areas, Biological Process, and Ratios columns of the .pdResult report.

Figure 423. Legends in data distribution columns

❖ **To sort the data in distribution map columns**

Refer to the Help.

Displaying the Quantification Channel Values Chart

When you select a PSM or a quantification spectrum for reporter ion quantification, the Quantification Channel Values chart displays the absolute intensity of the quantification values detected for the available quantification channels. When you select a protein or a peptide group, the chart displays the abundance of the item.

When you select a PSM or a quantification spectrum for precursor ion quantification, the Quantification Channel Values chart displays the area of the quantification values detected for the available quantification channels. When you select a protein or a peptide group, the chart displays the abundance of the item.

The chart shows the normalized abundances, if available, or the raw abundances. It can also display grouped abundances if they are available. Grouped abundances, which are the abundances of the sample groups defined on the Grouping and Quantification page of the study, appear as blue bars. Ungrouped abundances, which are the abundances of the samples defined on the Grouping and Quantification page of the study, appear as pastel-colored bars.

Ungrouped abundances are ordered. When the results contain grouped and ungrouped abundances, you can use the shortcut menu commands in the chart to choose the data to display. The Show Legend command labels the ungrouped bars with the sample identifier. It also displays a legend for the colored bars.

The processing workflow must include a quantification node for this chart to appear in the results report.

❖ **To display the Quantification Channel Values chart**

1. Click the appropriate tab (**Proteins**, **Peptide Groups**, **PSMs**, or **Quan Spectra**) in the .pdResult file.

2. Select the row of the protein, peptide group, PSM, or quantification spectrum that interests you.

3. Choose **View > Quan Channel Values**, or click the **Quan Channel Values** icon, .

To view the results, see the following topics:

- Displaying Quantification Channel Values for Reporter Ion Quantification

- Displaying Quantification Channel Values for Precursor Ion Quantification

- Displaying Quantification Channel Values Chart for Label-Free Quantification

Displaying Quantification Channel Values for Reporter Ion Quantification

The Quantification Channel Values chart for reporter ion quantification displays the absolute intensity of the reporter ions detected for the available quantification channels when you select a PSM or a quantification spectrum. When you select a protein or a peptide group, the chart displays the abundance of the item.

Displaying the Quantification Channel Values Chart for PSMs, MS/MS Spectrum Info Input, and Quantification Spectra

Reporter ions, also called *reporters*, are produced from the MS/MS of the labels affixed to peptide samples in reporter ion quantification. You can use the quantification value intensity to calculate the relative ratio of a peptide. You might also want to view the absolute quantification value intensity to verify that the peptide ratio calculation is correct.

❖ **To display the Quantification Channel Values chart for PSMs, MS/MS Spectrum Info input, and quantification spectra for reporter ion quantification**

1. Click the **PSMs, MS/MS Spectrum Info,** or **Quan Spectra** tab in the .pdResult file.

2. Select the row of the PSM or quantification spectrum that interests you.

3. Choose **View > Quan Channel Values**, or click the **Quan Channel Values** icon, .

The TMT 10plex quantification method has ten reporter ions. Suppose that they are used to label six biological samples: 126, 127N, 127C, 128N, 128C, 129N, 129C, 130N, 130C, and 131. Figure 424 shows the Quantification Channel Values chart created by the Show Quan Channel Values command for these samples. It shows the relative intensities of the samples labeled with the 126, 127N, 127C, 128N, 128C, 129N, 129C, 130N, 130C, and 131 reporter ions. Clearly, the sample labeled 128C is the sample with the greatest reporter ion intensity.

Figure 424. Quantification Channel Values chart showing the reporter ion raw S/N values for a selected PSM
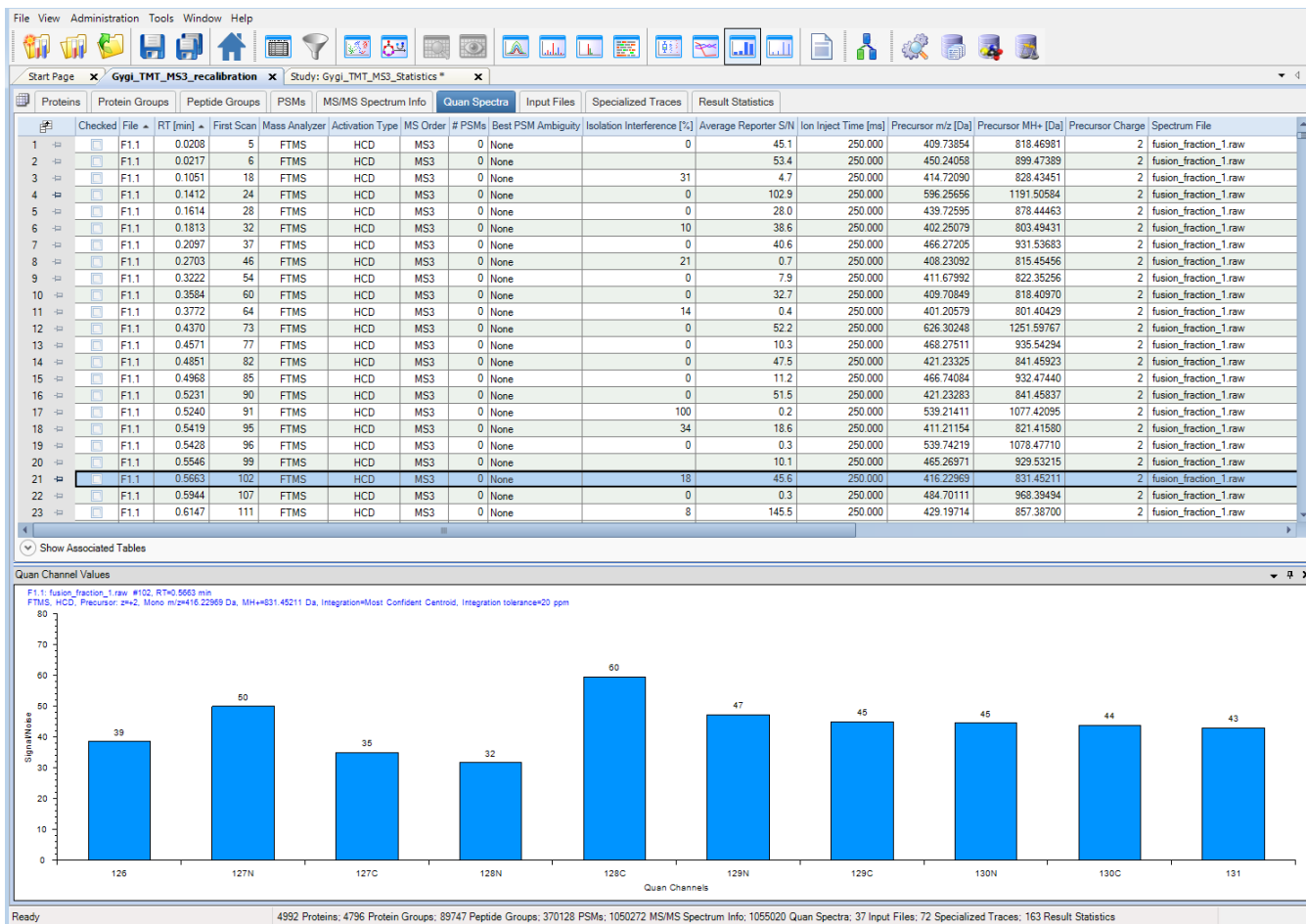
The *x* axis of the chart shows the names of the quantification channels, and the *y* axis shows the intensity of the reporter ions, in counts.

A legend is not available in the Quantification Channel Values chart when you display PSMs and quantification spectra.

Displaying the Quantification Channel Values Chart for Proteins and Peptide Groups

Figure 425 shows the grouped and ungrouped abundances of a selected protein. The *x* axis of the chart shows the quantification channels, and the *y* axis shows the abundance of the selected protein for the quantification channel, defined by arbitrary units. In this example, the quantification channels are labeled as study factors rather than as reporter ion masses. Error bars denoting the standard error are displayed above the blue bar that shows the average abundance of the ungrouped abundances. An optional legend at the bottom displays the grouped and ungrouped abundances.

❖ **To display the Quantification Channel Values chart for proteins and peptide groups for reporter ion quantification**

1. Click the **Proteins** or **Peptide Groups** tab in the .pdResult file.

2. Select the row of the protein or peptide group that interests you.

3. Choose **View > Quan Channel Values**, or click the **Quan Channel Values** icon, [icon].

Figure 425. Quantification Channel Values chart showing the abundance of a protein in reporter ion quantification
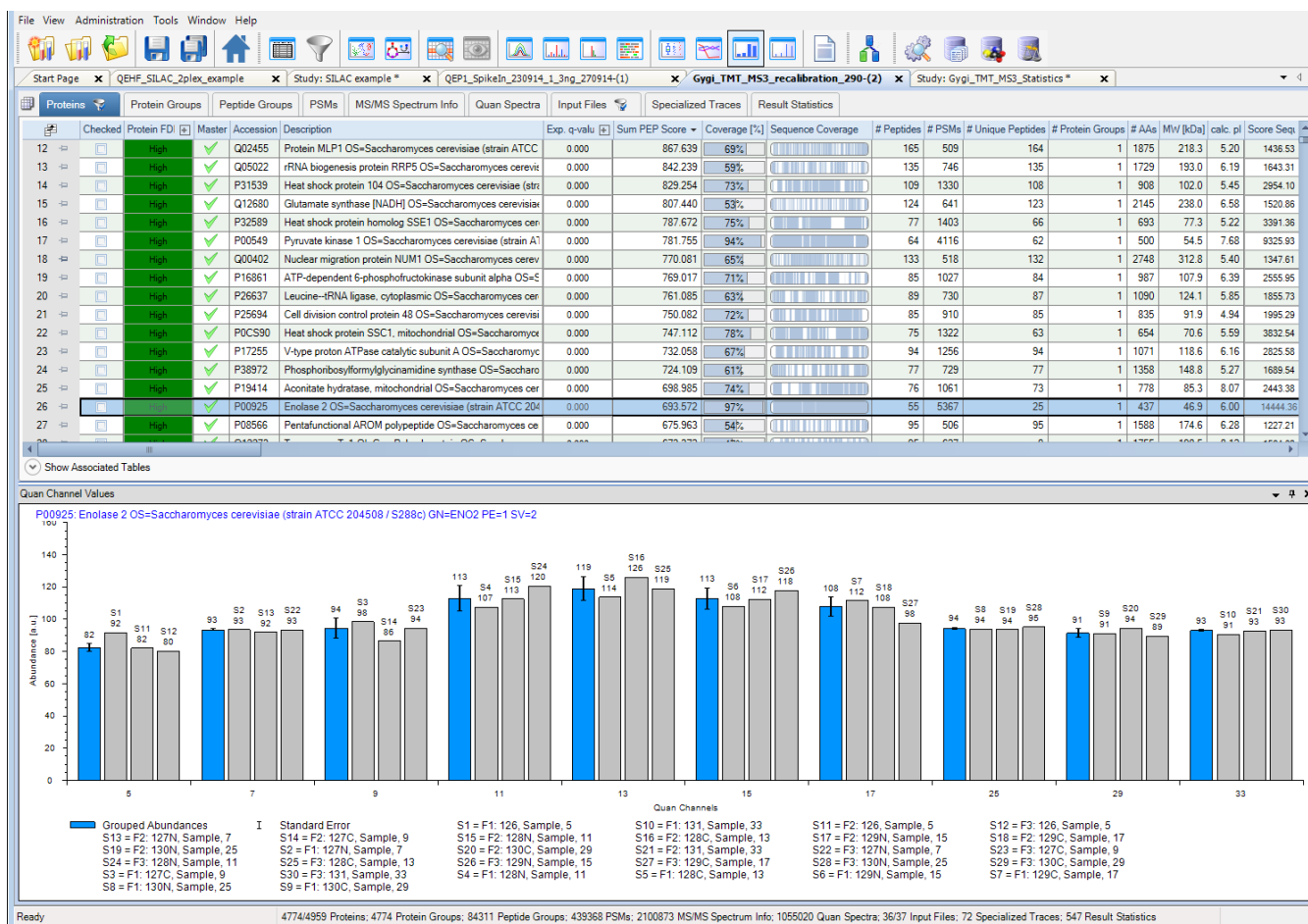
The *x* axis of the chart shows the names of the quantification channels, and the *y* axis shows the scaled abundance or summed S/N values, depending on whether you chose scaling in the Reporter Ions Quantifier node.

❖ **To display the legend**

Right-click the Quan Channel Values pane and choose **Show Legend**.

❖ **To display grouped abundances**

Right-click and choose **Show Grouped Abundances**.

❖ **To display ungrouped abundances**

Right-click and choose **Show Ungrouped Abundances**.

Displaying Quantification Channel Values for Precursor Ion Quantification

The Quantification Channel Values chart for precursor ion quantification displays the area of the isotopes detected for the available quantification channels when you select a PSM or a quantification spectrum. When you select a protein or a peptide group, the chart displays the abundance of that item.

Displaying the Quantification Channel Values Chart for PSMs and Quantification Spectra

Heavy isotopes are incorporated into proteins in precursor ion quantification. You can use the quantification value area to calculate the relative ratio of a peptide. You might also want to view the quantification value area to verify that the peptide ratio calculation is correct.

❖ **To display the Quantification Channel Values chart for PSMs for precursor ion quantification**

1. Click the **PSMs** tab in the .pdResult file.

2. Select the row of the PSM or quantification spectrum that interests you.

3. Choose **View > Quan Channel Values**, or click the **Quan Channel Values** icon, .

The SILAC 2plex (Arg6, Lys6) quantification method in SILAC has a mixture of samples labeled with Light and Heavy isotopic labels. Figure 426 shows the chart created by the Show Quan Channel Values command for these samples. It shows the relative peak area for the samples with the Light and Heavy isotope labels. The sample labeled Light is the sample with the greatest area.

Figure 426. Quantification Channel Values chart showing abundances on the PSMs page for precursor ion quantification
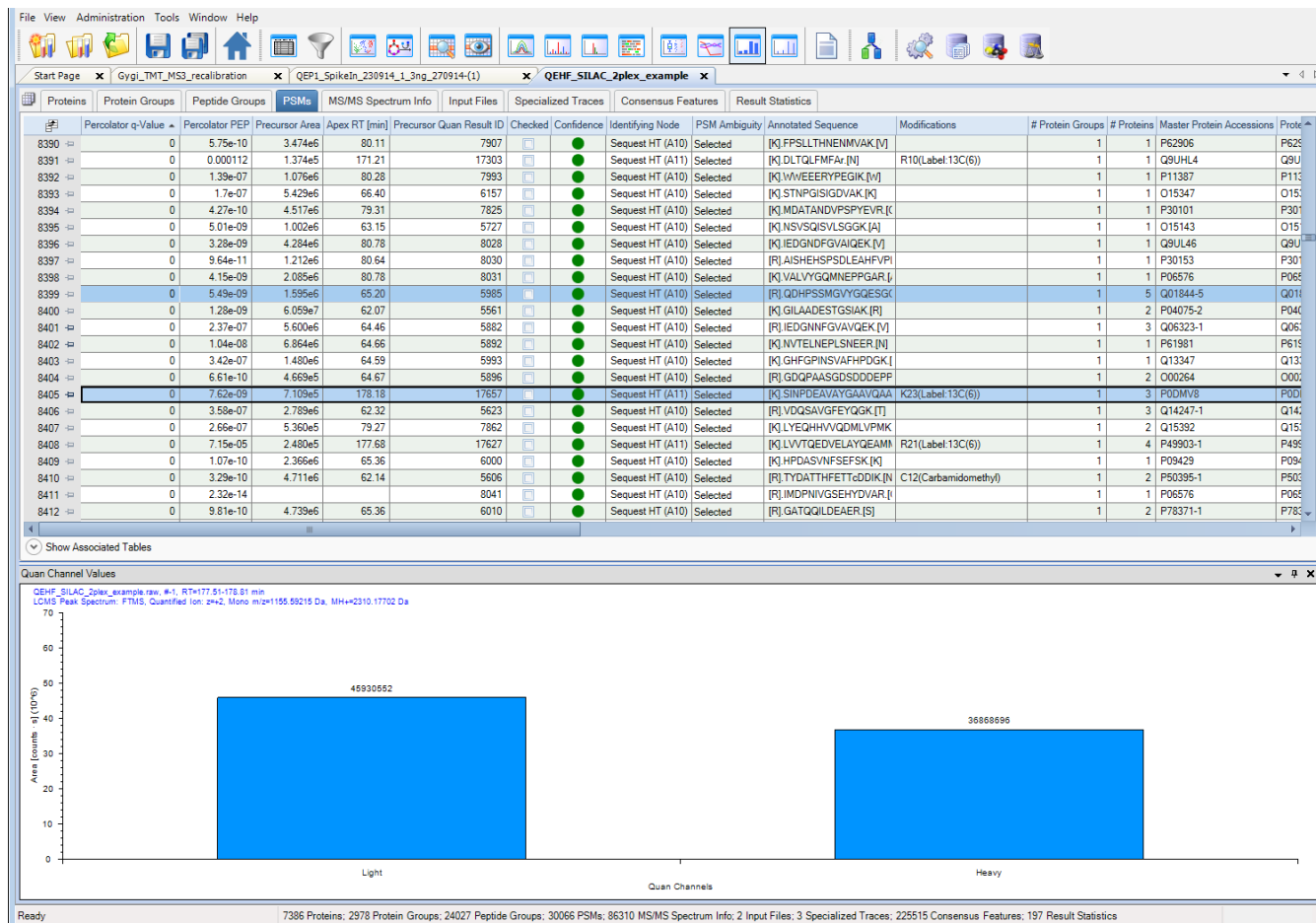
The *x* axis of the chart shows the quantification channels, and the *y* axis shows the detected area for the given quantification channel, defined by counts × seconds.

A legend is not available in the Quantification Channel Values chart when you display PSMs and quantification spectra. You also cannot display grouped and ungrouped abundances.

Displaying the Quantification Channel Values Chart for Proteins and Peptide Groups

Figure 427 shows the grouped and ungrouped abundances of a selected protein. The *x* axis of the chart shows the quantification channels, and the *y* axis shows the abundance of the selected protein for the quantification channel, defined by arbitrary units. In this example, the quantification channel names are replaced by the names of the study factors. An optional legend at the bottom displays the grouped and ungrouped abundances.

❖ **To display the Quantification Channel Values chart for proteins and peptide groups for precursor ion quantification**

1. Click the **Proteins** or **Peptide Groups** tab in the .pdResult file.

2. Select the row of the protein or peptide group that interests you.

3. Choose **View > Quan Channel Values**, or click the **Quan Channel Values** icon,

Figure 427 shows the grouped and ungrouped abundances on the Proteins page of the .pdResult file for precursor ion quantification.

Figure 427. Quantification Channel Values chart showing grouped and ungrouped abundances on Proteins page for precursor ion quantification
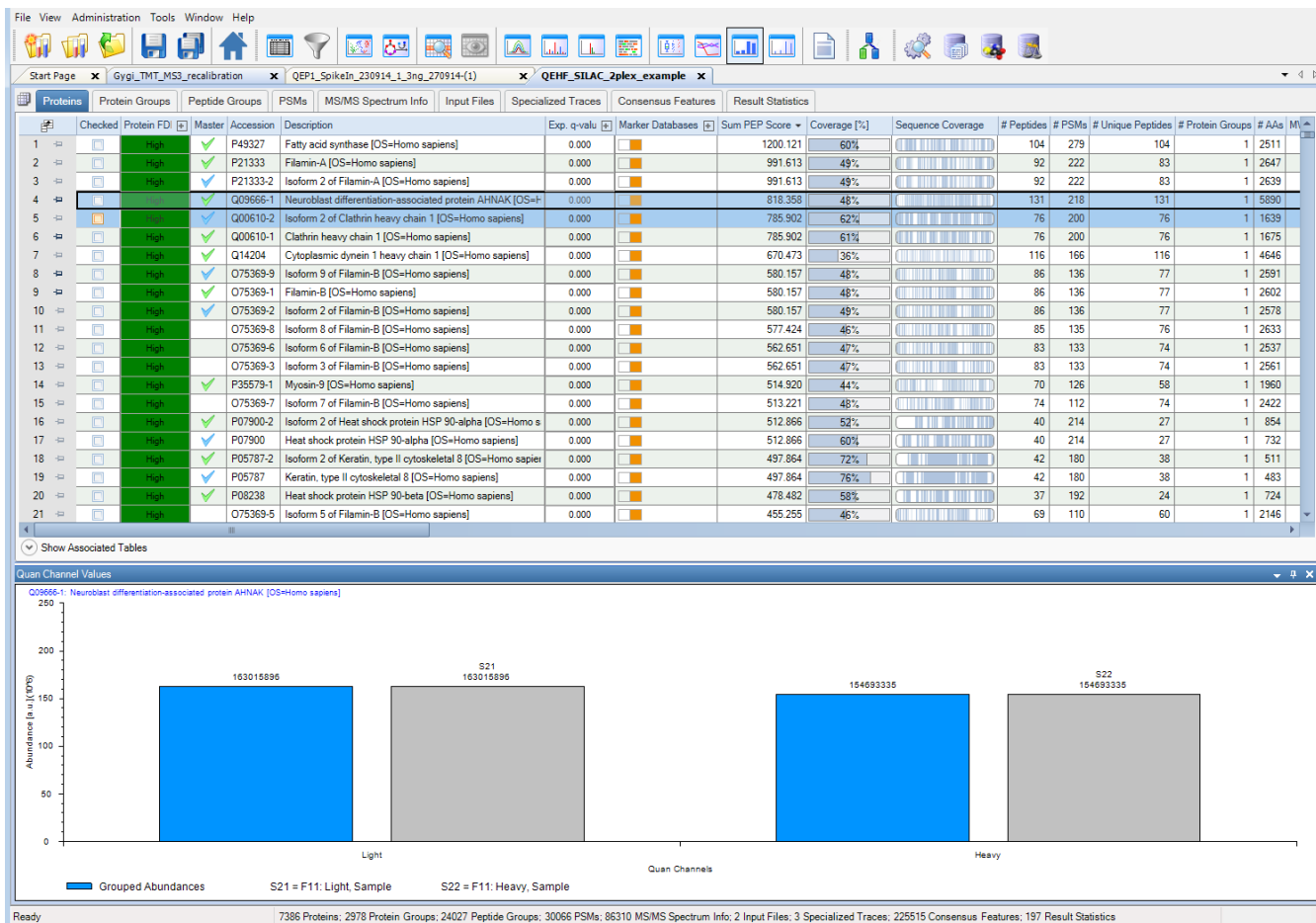
The *x* axis of the chart shows the quantification channels, and the *y* axis shows scaled abundance or summed S/N values, depending on whether you chose scaling in the Precursor Ions Quantifier node.

❖ **To display the legend**

Right-click the Quan Channel Values pane and choose **Show Legend**.

❖ **To display grouped abundances**

Right-click and choose **Show Grouped Abundances**.

❖ **To display ungrouped abundances**

Right-click and choose **Show Ungrouped Abundances**.

Displaying Quantification Channel Values Chart for Label-Free Quantification

❖ **To display the Quantification Channel Values chart**

1. Click the **Proteins** or **Peptide Groups** tab in the .pdResult file.

2. Select the row of the protein or peptide group that interests you.

3. Choose **View > Quan Channel Values**, or click the **Quan Channel Values** icon, .

   Figure 428 shows the grouped and ungrouped abundances on the Proteins page of the .pdResult file for label-free quantification.

Figure 428. Quantification Channel values chart showing grouped and ungrouped abundances for label-free quantification
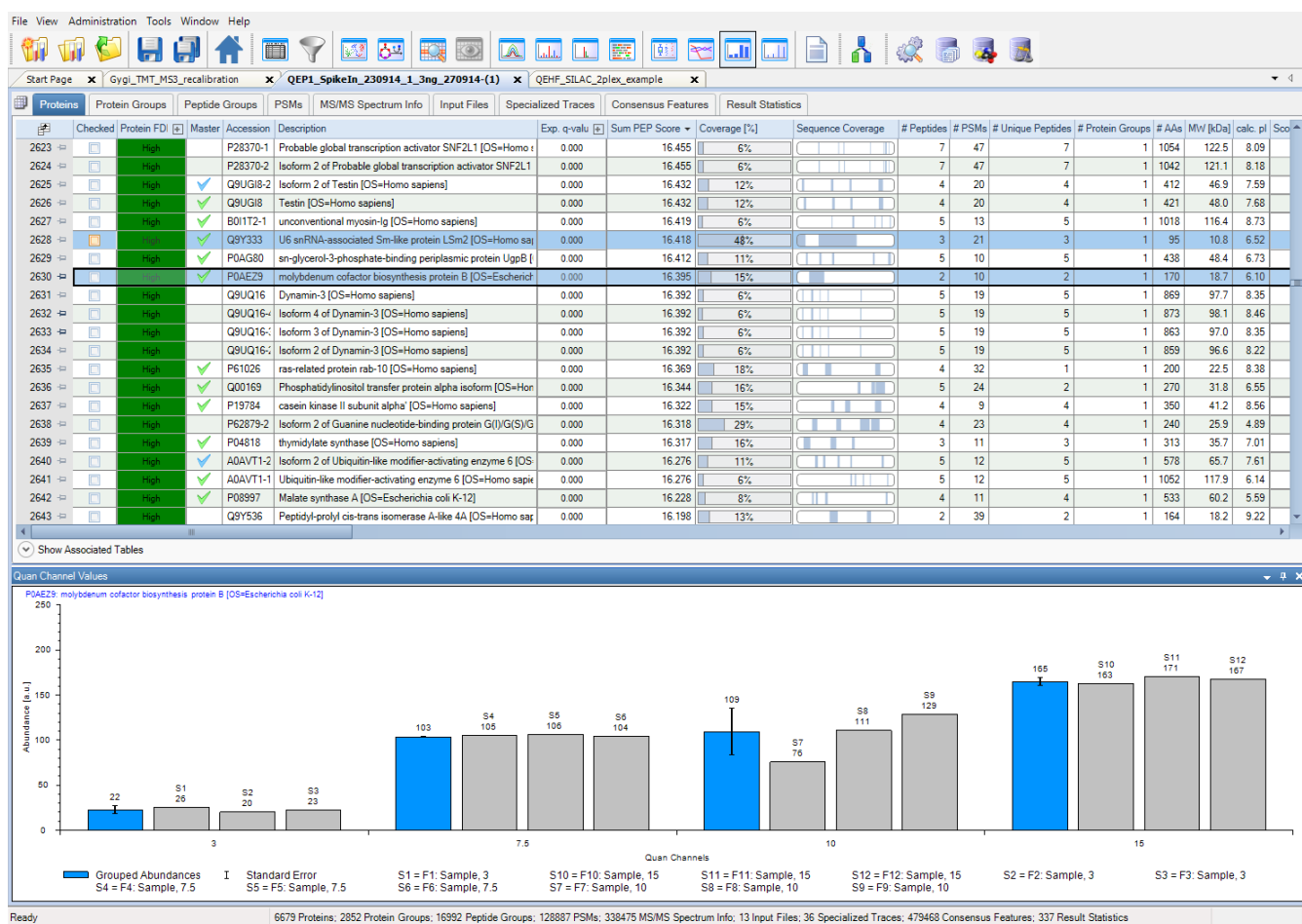
The *x* axis of the chart shows the quantification channels, and the *y* axis shows the scaled abundance or summed S/N values, depending on whether you chose scaling in the Precursor Ions Quantifier node.

❖ **To display the legend**

Right-click the Quan Channel Values pane and choose **Show Legend**.

❖ **To display grouped abundances**

Right-click and choose **Show Grouped Abundances**.

❖ **To display ungrouped abundances**

Right-click and choose **Show Ungrouped Abundances**.

Quantification Channel Values Chart Parameters

Table 33 describes the features of the Quantification Channel Values chart.

Table 33. Quantification Channel Values chart parameters

Parameter	Description
Header	Displays the accession number and description for proteins.
	Displays the identified peptide in the header for peptide groups.
	Displays the following for PSMs:
	• Name and location of the raw data file
	• Numbers of any grouped spectra
	• Retention time of the precursor ion
	• Mass spectrometer used
	• Activation type used
	• Mass-to-charge ratio (*m/z*) of the "light" ion (precursor ion quantification) or the mass-to-charge ratio of the precursor ion (reporter ion quantification), in daltons
	• MH+ (the protonated monoisotopic mass), in daltons
	• Integration method (reporter ion quantification)
	• Integration tolerance (reporter ion quantification)
Quan Channels	Specifies the names of the quantification channels used. If you have defined study factors, the quantification channel names are replaced by those study factors.
Area [counts × seconds] (precursor ion quantification)	Specifies the area of the extracted ion chromatogram of each channel. The relative peak intensities of multiple isotopically distinct peptides from each protein are then used to determine the average change in protein abundance in the treated sample.
Intensity [counts] (reporter ion quantification)	Specifies the intensity of the reporter ions, in counts.

Quantification Channel Values Chart Shortcut Menu

Table 34. Quantification Channel Values shortcut menu commands

Command	Description
Show Position Tooltips	No functionality.
Zoom Out	No functionality.
Undo All Zoom/Pan	No functionality.
Copy	Copies the data in the Quantification Channel Values chart.
Save As	Opens the Save As dialog box so that you can save the Quantification Channel Values chart in an Enhanced Metafile (EMF) file.
Show Legend	Displays a legend at the bottom of the Quantification Channel Values chart that can display grouped or ungrouped abundances or both when you select a protein or a peptide group.
Show Grouped Abundances	Displays only grouped abundances when you select a protein or a peptide group.
Show Ungrouped Abundances	Displays only ungrouped abundances when you select a protein or a peptide group.

Displaying the Quantification Spectrum Chart

You can generate a chart showing the spectrum used for quantification. This chart is available for every peptide with an associated quantification result. You must conduct a search with a workflow that includes a quantification node for this page to appear in the results report.

❖ **To display the Quantification Spectrum chart**

1. Click the appropriate tab in the .pdResult file.

2. Select the row that interests you.

3. Choose **View > Quantification Spectrum**, or click the **Quantification Spectrum** icon, .

To see the results, see the following topics:

- Displaying the Quantification Spectrum Chart for Reporter Ion Quantification

- Displaying the Quantification Spectrum Chart for Precursor Ion Quantification

- Displaying the Quantification Spectrum Chart for Label-Free Quantification

Displaying the Quantification Spectrum Chart for Reporter Ion Quantification

For reporter ion quantification, the Quantification Spectrum chart displays the intensity of the reporter ions, in counts. It shows a spectrum for each peptide, except for those peptides labeled with "No Quan Values" in the Quan Info column of the PSMs page.

❖ **To display the Quantification Spectrum chart for reporter ion quantification**

1. Click the **PSMs**, **Quan Spectra**, or **MS/MS Spectrum Info** tab in the .pdResult file.

2. Select the row of the quantification spectrum that interests you.

3. Choose **View > Quantification Spectrum**, or click the **Quantification Spectrum** icon, ![icon].

   Figure 429 shows an example of a quantification spectrum from a TMT 10plex sample.

Figure 429. Quantification Spectrum chart for a TMT 10plex showing the intensity of reporter ions
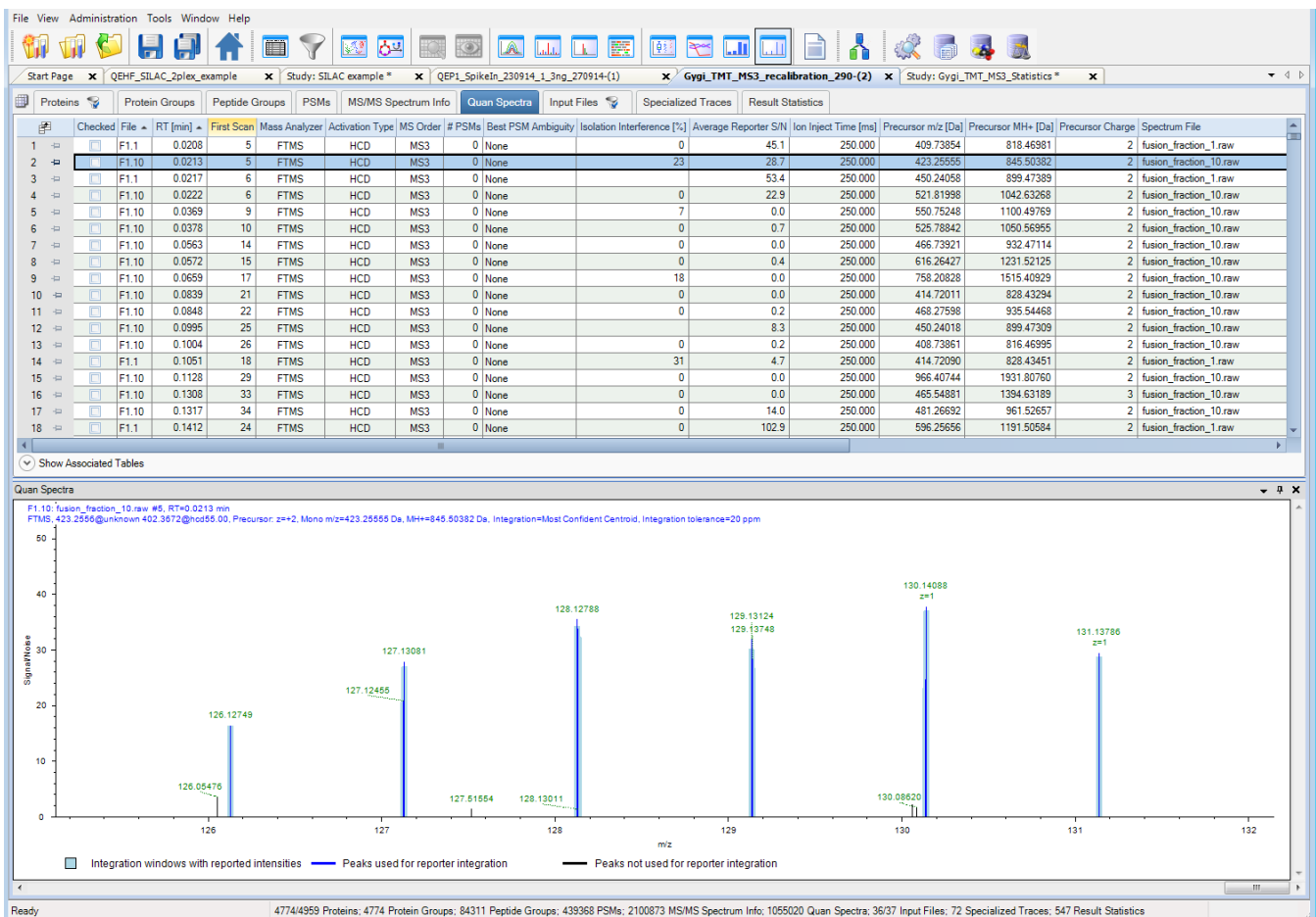
The width of the bars shown in the Quantification Spectrum chart is determined by the setting of the Integration Tolerance parameter in the Reporter Ions Quantifier node in the processing workflow.

The Quantification Spectrum chart for reporter ion quantification includes the following features:

- The light blue bars represent the integration windows for the reporter tags. The boxes are centered on the masses of the reporter tags, as specified in the quantification method. The width of boxes is the integration window used for extracting the reporter tags. It is ±20 ppm, as specified by the settings of the parameters in the Reporter Ions Quantifier node in the processing workflow (you can look up all these values on the Quantification Summary page). The height of the line in the box represents the actual tag intensity used for calculating the peptide ratios. The height of the bars represent the corrected tag intensity. The height depends on the setting of the Integration Method parameter specified in the Reporter Ions Quantifier node in the processing workflow. It is always the value that results from correction for isotopic impurities, as specified in the TMT Reporter Ion Isotope Distributions area of the Quan Channels page of the Quantification Method Editor dialog box (see Figure 249 on page 356).

- The bright blue lines represent the peaks used for reporter integration. To calculate the actual intensity of a particular tag, the application chooses the blue fragment peaks from the spectrum and considers only peaks in the integration window.

- The pink bars denote the theoretical quantification windows that indicate where the peaks should be.

  Figure 430. Theoretical integration windows marked by pink bars in the Quantification Spectrum chart
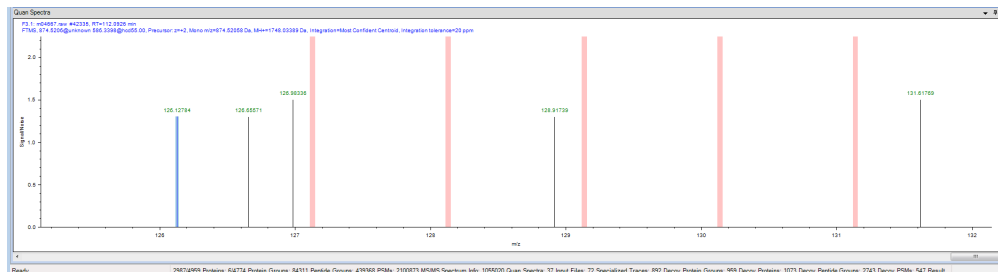
- The black fragment peaks represent peaks that are present in the spectrum but that are not chosen for calculating the tag intensities. They might not be chosen because the peaks lie outside of any integration window, or because the setting of the Integration Method parameter specified in the Reporter Ions Quantifier node determined that only one peak per integration window should be chosen from any integration window. A different peak was picked for this integration window according to the criterion specified by the Integration Method setting.

Displaying the Quantification Spectrum Chart for Precursor Ion Quantification

For precursor ion quantification, the Quantification Spectrum chart displays a quantification spectrum for each peptide. It also displays the different abundances of the identified Light, Medium, and Heavy isotopic peak patterns used to quantify a peptide. The abundances are measured by calculating the area of the extracted ion chromatogram of each isotope of a pattern. The chart highlights the corresponding isotope pattern peaks and labels them with the quantification channel names, as shown in Figure 431. It also includes any peaks that are not part of an isotope pattern.

❖ **To display the Quantification Spectrum chart for precursor ion quantification**

1. Click the **PSMs** tab in the .pdResult file.

2. Select the row of the PSM that interests you.

3. Choose **View > Quantification Spectrum**, or click the **Quantification Spectrum** icon, .

Figure 431 gives an example of the Quantification Spectrum chart for precursor ion quantification.

Figure 431. Quantification Spectrum chart for precursor ion quantification
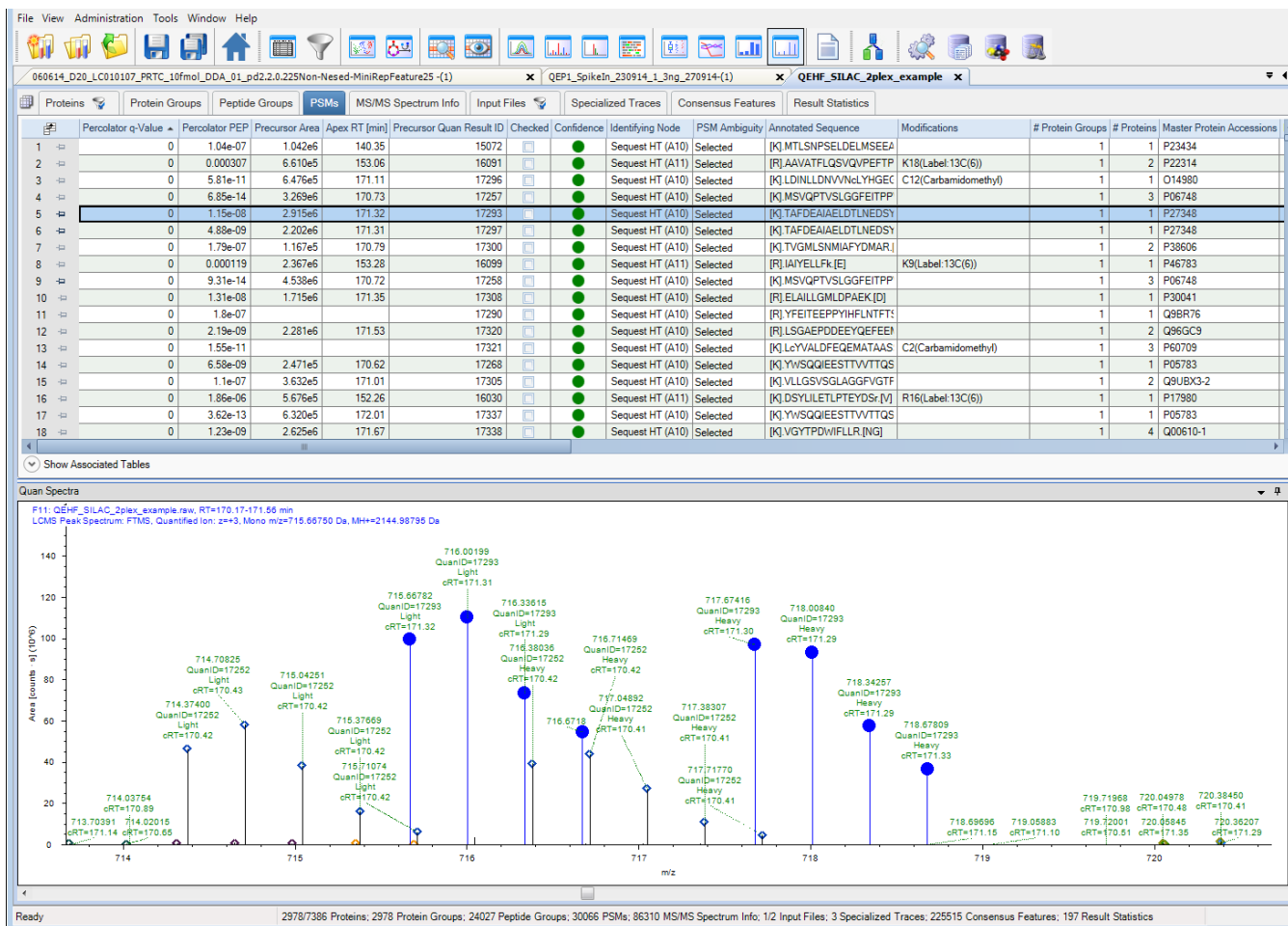
The *x* axis of the chart displays the mass-to-charge ratio of the isotopes, and the *y* axis displays the area of the extracted ion chromatogram for the isotopes. Filled blue circles mark the isotope pattern peaks for the different isotopic forms of the identified peptide. Only the most abundant peak is used for quantification. Other isotope distributions that coelute are labeled with diamonds in different colors, with peaks from the same isotope clusters having the same color of diamond.

The Quantification Spectrum chart can also indicate whether an expected quantification pattern peak is absent. Regions in pink indicate where a quantification pattern peak was expected but is absent (see Figure 432). This ion pattern peak is not used in calculating the quantification values for the different quantification channels.

Figure 432. Expected but absent peak in the Quantification Spectrum chart
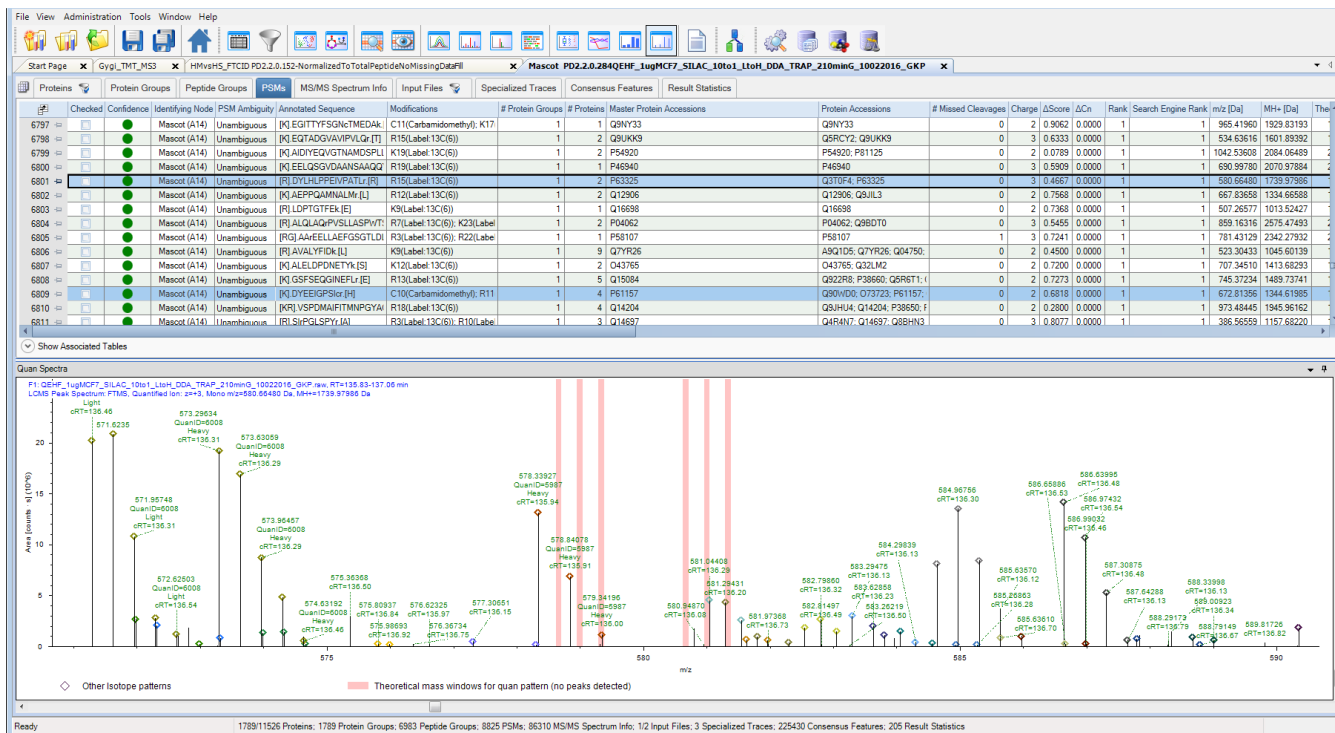
Regions in green, shown in Figure 433, indicate where a quantification pattern peak was expected but is unsuitable. Pattern peaks might be unsuitable because of the wrong centroid retention time, out-of-delta mass range, wrong intensity, or a peak that has been used by another isotopic pattern. This ion pattern peak is not used in calculating the quantification values for the different quantification channels.

Figure 433. Expected but unsuitable peaks in the Quantification Spectrum chart
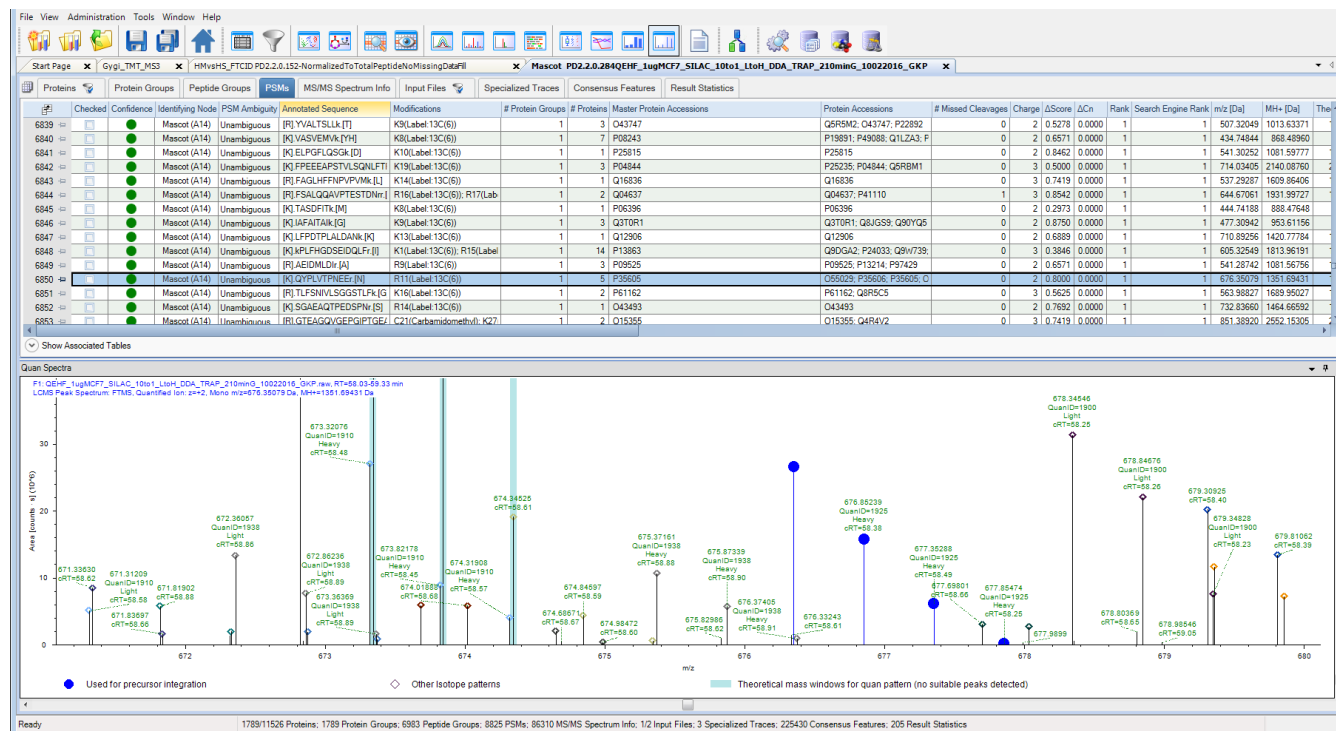
Table 35 shows what the various colors mean on the Quantification Spectrum charts in Figure 432 and Figure 433.

Table 35. The meaning of colors in the Quantification Spectrum chart

Color	Meaning
Filled blue circle	Indicates the isotope pattern peaks that are used in calculating the quantification values for the different quantification channels.
Yellow box	Indicates that the pattern includes peaks from indistinguishable channels. This ion pattern peak is not used in calculating the quantification values for the different quantification channels.
Pink bar	Indicates that a quantification pattern peak is expected but is missing. This ion pattern peak is not used in calculating the quantification values for the different quantification channels.
Green bar	Indicates that a quantification pattern peak is present but is unsuitable because of errors in peptide labeling or because of the wrong centroid retention time, the out-of-delta mass range, the wrong intensity, or a peak that has been used by another isotopic pattern. This ion pattern peak is not used in calculating the quantification values for the different quantification channels.

Displaying the Quantification Spectrum Chart for Label-Free Quantification

For label-free quantification, the Quantification Spectrum chart displays the precursor isotope distribution for a selected PSM, LCMS feature, or MS/MS spectrum input. The abundances are measured by calculating either the height or the area of the extracted ion chromatogram of the most abundant isotope in the pattern. When you select a PSM or LCMS peak, the peaks of the selected isotope cluster are colored in blue with a filled blue circle above it. Other isotope distributions that coelute display diamonds in different colors above them, with peaks from the same isotope clusters having the same color of diamond.

❖ **To display the Quantification Spectrum chart for label-free quantification**

1. Click the **PSMs**, **MS/MS Spectrum Info**, or **LCMS** tab in the .pdResult file.

2. Select the row of the PSM that interests you.

3. Choose **View > Quantification Spectrum**, or click the **Quantification Spectrum** icon, .

Figure 434 gives an example of the Quantification Spectrum chart for label-free quantification.

Figure 434. Quantification Spectrum chart for label-free quantification
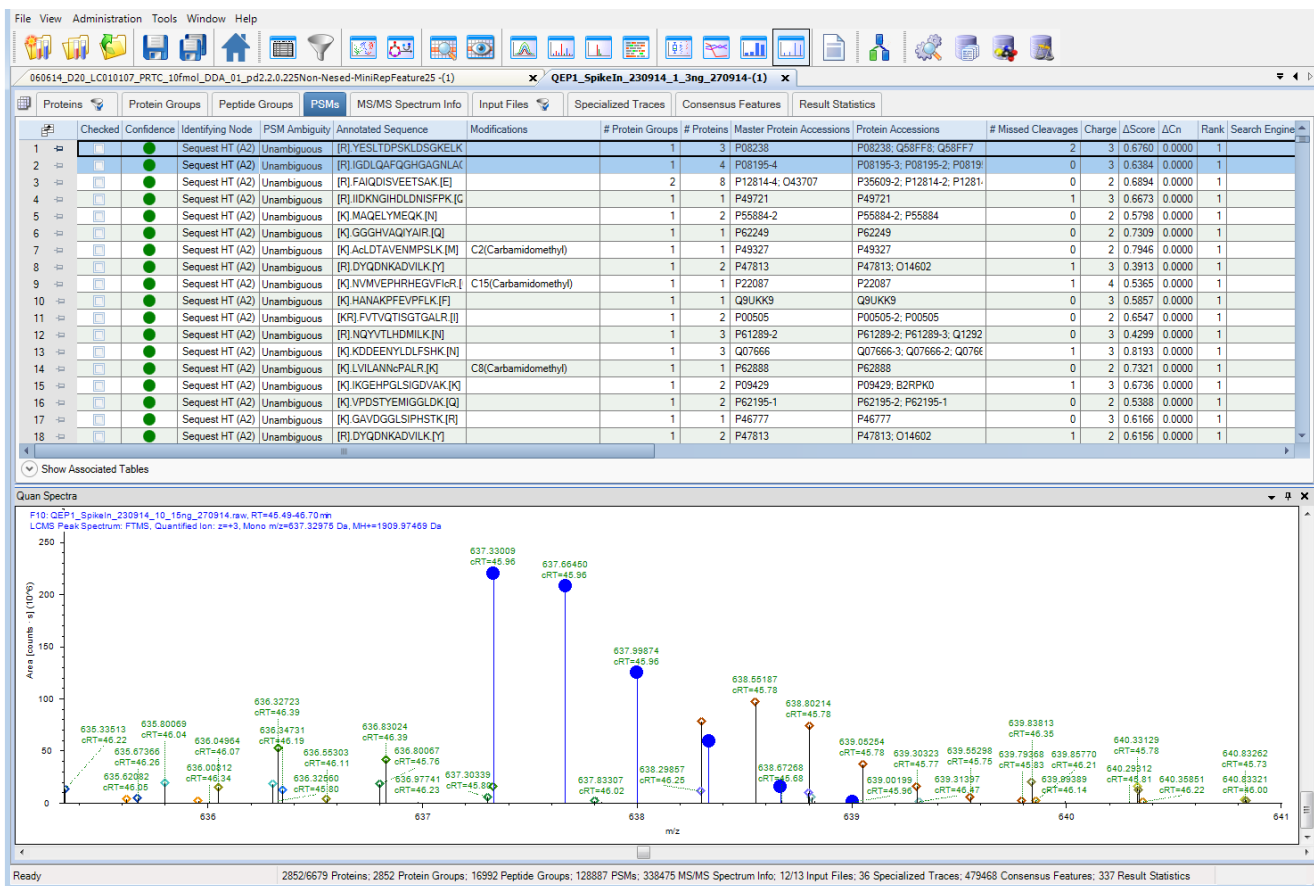
The *x* axis of the chart displays the mass-to-charge ratio of the isotopes, and the *y* axis displays the area of the extracted ion chromatogram for the isotopes.

The colors and symbols on the chart are the same as those for precursor ion quantification. See Table 35.

Quantification Spectrum Chart Parameters

Table 36 describes the features of the Quantification Spectrum chart.

Table 36. Quantification Spectrum chart parameters

Parameter	Description
Header	Displays the following: • Name and location of the raw data file • Scan numbers of any grouped spectra • Retention time of the precursor ion • Type of mass spectrometer used • Activation type (reporter ion quantification only) • Mass-to-charge ratio of the precursor ion, in daltons (reporter ion quantification) or the mass-to-charge ratio of the light isotope of the peptide (precursor ion quantification). • Monoisotopic mass of the peptide of interest • MH+ (the protonated monoisotopic mass), in daltons • Integration method (reporter ion quantification only) • Integration tolerance (reporter ion quantification only)
m/z	Displays the mass-to-charge ratio of the reporter ion tags (for reporter ion quantification) or the isotopes (for precursor ion quantification).
Intensity [counts]/ Area [counts × seconds]	Displays the intensity of the reporter ions, in counts, (for reporter ion quantification) or the extracted chromatographic area of the isotopes (for precursor ion quantification).

Displaying the Quan Spectra Page

The Quan Spectra page of the .pdResult report lists all the spectra that were used in reporter ion quantification. It displays quantification values for identified spectra and spectra that were quantified but not identified. The application generates the Quan Spectra page only if you include a Reporter Ions Quantification node in your processing workflow. When the consensus workflow includes the Reporter Ions Quantifier node, the Quan Spectra page contains columns listing the raw quantification values and the average reporter signal-to-noise values if they are available.

Figure 435. Quan Spectra page

When you select a row on the Quan Spectra page, you can view data in the Quantification Channels view and the Quantification Spectrum view to see the details of the spectra used for quantification.

❖ **To display the Quan Spectra page**

1. Perform reporter ion quantification.

2. Open the .pdResult report and click the **Quan Spectra** tab.

Quan Spectra Page Parameters

The columns of the Quan Spectra page are basically the same as those on the MS/MS Spectrum Info page (see the Help). However, they also include reporter ion abundances of the intensity of the fragmented tag in a sample to the intensity of the fragmented tag in the control sample for all spectra, regardless of whether they have been identified.

Table 37 describes the columns on the Quan Spectra page.

Table 37. Quan Spectra page columns (Sheet 1 of 4)

Parameters	Description
# PSMs	Displays the total number of PSMs attached to the quantification spectrum.
Abundances	Displays the abundances of the quantification channels.
Activation Type	Displays the fragmentation method used to produce the product spectrum.
Average Reporter S/N	Displays the average reporter S/N values, which the application calculates as the sum of S/N values found, divided by the number of defined tags.

Table 37. Quan Spectra page columns (Sheet 2 of 4)

Parameters	Description
Best PSM Ambiguity	Displays the "best" ambiguities of the PSMs attached to a spectrum. This column can display the following categories: • Unambiguous: Indicates that this PSM is the only match that the application considered for this spectrum. There is no ambiguity that it needs to resolve. • Selected: Indicates that the application selected this PSM from a set of two or more matches that it considered for the protein group inference process. • Rejected: Indicates that the application rejected this PSM from a set of two or more matches that it considered for the protein group inference process. • Ambiguous: Indicates that the application considered two or more peptide matches to the same spectrum for the protein group inference process. If you set the Apply Strict Parsimony Principle parameter of the Protein Grouping (Enabled) node to True, the application resolves this ambiguity to Selected or Rejected. • Unconsidered: Indicates that the application did not consider this PSM for the protein group inference process because it did not meet the requirements specified by the settings of the Protein Grouping node. For more information on the protein group inference process that the application uses to assign the grouping status of PSMs, see Grouping Proteins.
Checked	Indicates whether the item is selected.
Creator Node No	Displays the number of the node in the workflow that created the spectrum.
File ID	Displays a short identifier of the input file that the spectrum was taken from.
First Scan	Displays the number of the first scan in grouped scans used to produce the composite search input scan, if grouping is performed.

Table 37. Quan Spectra page columns (Sheet 3 of 4)

Parameters	Description
Ion Inject Time [ms]	Specifies the time in milliseconds used to accumulate ions in the mass spectrometer to reach their target value before they are scanned out. Low ion injection times are desirable because they indicate high transmission and fast cycle times. Ion injection times that slowly increase might indicate lower transmission, the need for multiplier gain calibration, or both. The MS/MS Spectrum page displays the Ion Inject Time [ms] column by default for every searched spectrum and every PSM for newly processed data.
Isolation Interference [%]	Displays the percentage of interference by coisolation within the precursor isolation window. The application calculates the isolation interference only for precursors from MS1 spectra with high resolution. For more information on coisolation, see Excluding PSMs with High Levels of Coisolation.
Last Scan	Displays the number of the last scan in grouped scans used to produce the composite search input scan, if grouping is performed. The last scan might not be the same as the first scan if the workflow enabled grouped scans.
Mass Analyzer	Displays the mass analyzer used to create the raw data file. The application recognizes the following mass analyzers: • ITMS (Ion Trap) • FTMS (Fourier Transform) • TOFMS (Time of Flight) • SQMS (Single Quad) • TQMS (Triple Quad) • SectorMS (Sector Field)
Master Scan(s)	Displays the master scans of the scans that the application used to derive the actual scan, usually one MS1 scan.
MS Order	Indicates the order of the MS spectrum: • MS1 is one MS scan. • MS2 is an MS/MS scan. • MS3 is an MS^3 scan.
Original Mass	Displays the original mass of the precursor.
Original Precursor Charge	Displays the charge state of the precursor as stored in the scan header of the raw data file. This value is different from the actual charge state if you use charge deconvolution, for example, by using Xtract.
Precursor Charge	Displays the charge state of the precursor ion.

Table 37. Quan Spectra page columns (Sheet 4 of 4)

Parameters	Description
Precursor Intensity	Displays the intensity of the precursor ion.
Precursor m/z [Da]	Displays the mass-to-charge ratio of the precursor ion, in daltons.
Precursor MH+ [Da]	Displays the singly charged mass of the precursor ions, in daltons.
Quan Info	Displays the status of the validity of this quantification result for the quantification. For reporter ion quantification, it can be one of the following values: • Rejected by Method: The quantification spectrum was not used in the quantification. • No Quan Values: No tag was found. • Missing Values: A tag value was missing, the Reject Quan Results with Missing Channel parameter of the Reporter Ions Quantification node was set to True, and no imputation was used. • Shared: Different peptide sequences are associated with a quantification spectrum.
Quan Info Details	Displays the reason for rejection when the status in the Quan Info column is Rejected by Method. • Filtered by Isolation Interference: The spectrum isolation interference is above a user-defined threshold. • Filtered by Average S/N: The average signal-to-noise value is below a user-defined threshold.
RT [min]	Displays the retention time when the application observed the peptide, in minutes.
Scan(s)	Displays the scans from which the actual scan was derived. Usually it is a single scan, but a grouping node might group several scans together.
SearchID	Displays the order in which the search was submitted to the job queue.
Spectrum File	Displays the name of the file containing the spectra.

Displaying Quantification Ratio Distribution Charts

For both precursor ion and reporter ion quantification, the application calculates protein and peptide group ratios from the protein and peptide group abundances. The ratios are not calculated from PSM ratios.

You can display a chart of the peptide group ratios, which shows the distribution of peptide group ratios for the selected protein and displays the ratios of the peptide groups associated with the selected protein as a log2-fold change. For this chart to appear in the results report, you must conduct a search with a workflow that includes a quantification node.

❖ **To display quantification ratio distribution charts**

1. On the Proteins page of the .pdResult file that contains quantification results, click the row of the protein that you are interested in.

2. Choose **View > Quan Ratio Distributions**, or click the **Quan Ratios Distribution** icon, .

The Quan Ratio Distributions chart opens for reporter ion quantification (see Figure 436), or for precursor ion quantification (see Figure 437 on page 608).

Figure 436. Quan Ratio Distributions chart for reporter ion quantification
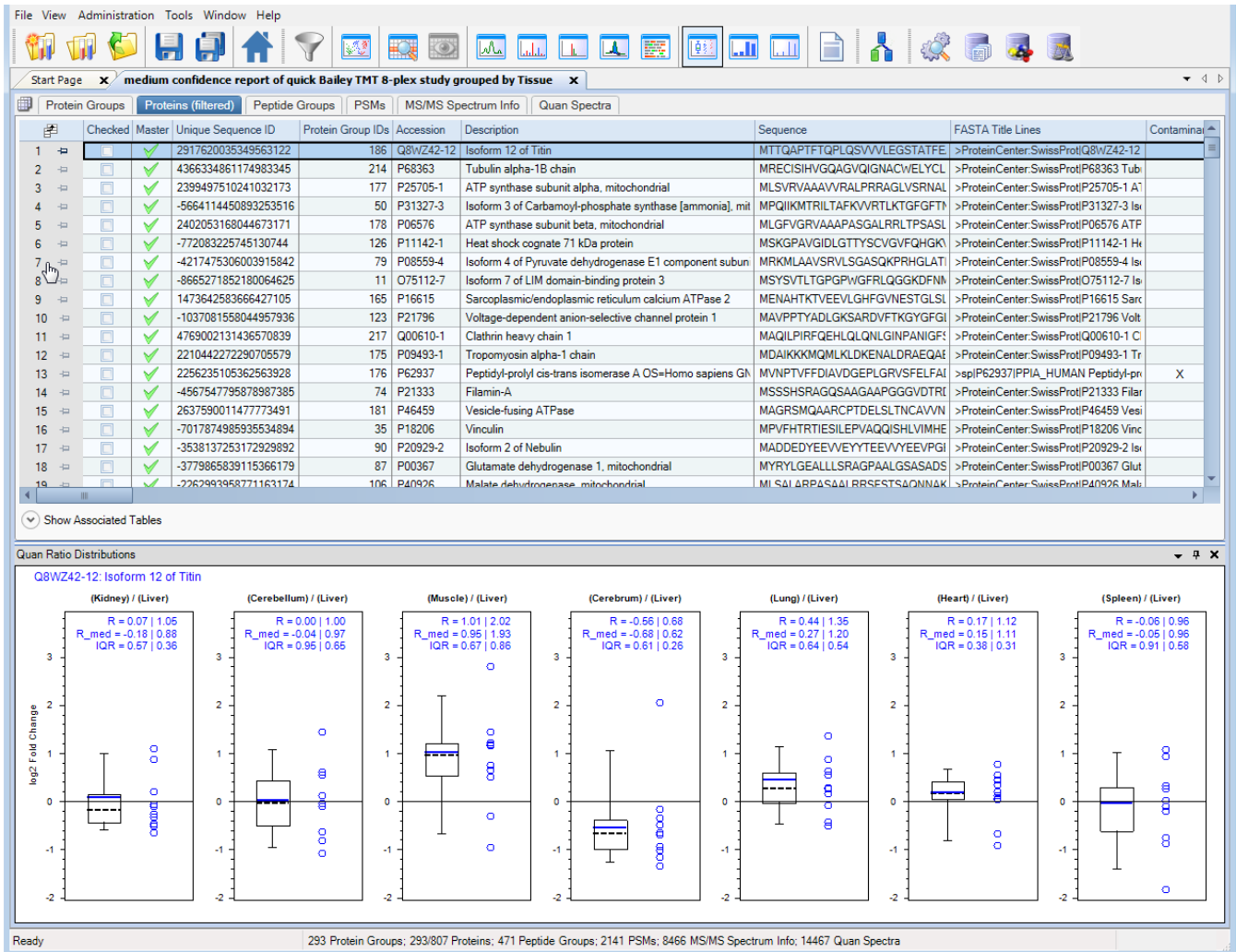
Figure 437. Quan Ratio Distributions Chart for precursor ion quantification
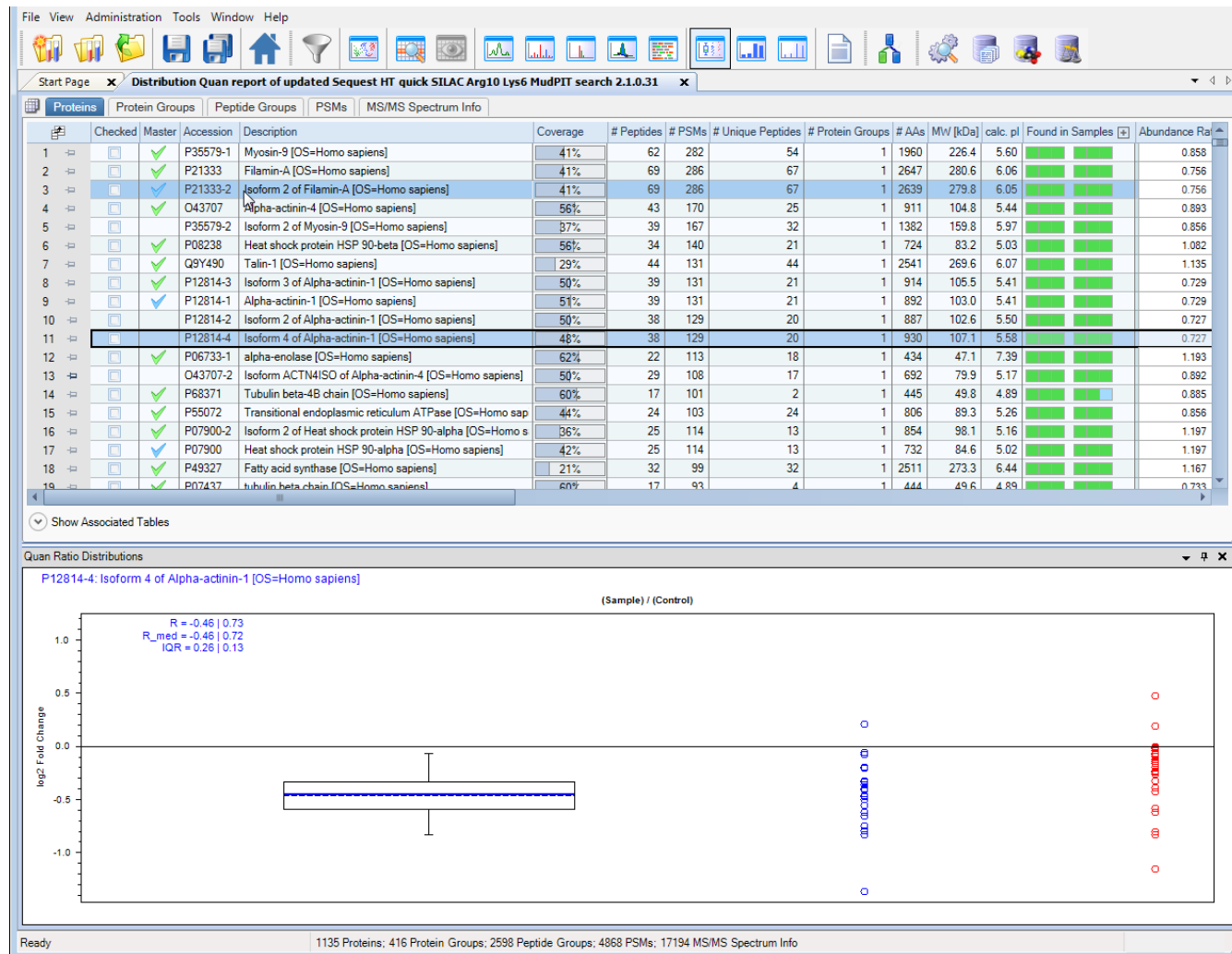
The Quantification Ratio Distributions chart shows the distribution and spread of the ratios of all peptides belonging to a particular protein. Figure 438 shows an example for the albumin protein.

Figure 438. Quantification Ratio Distributions chart
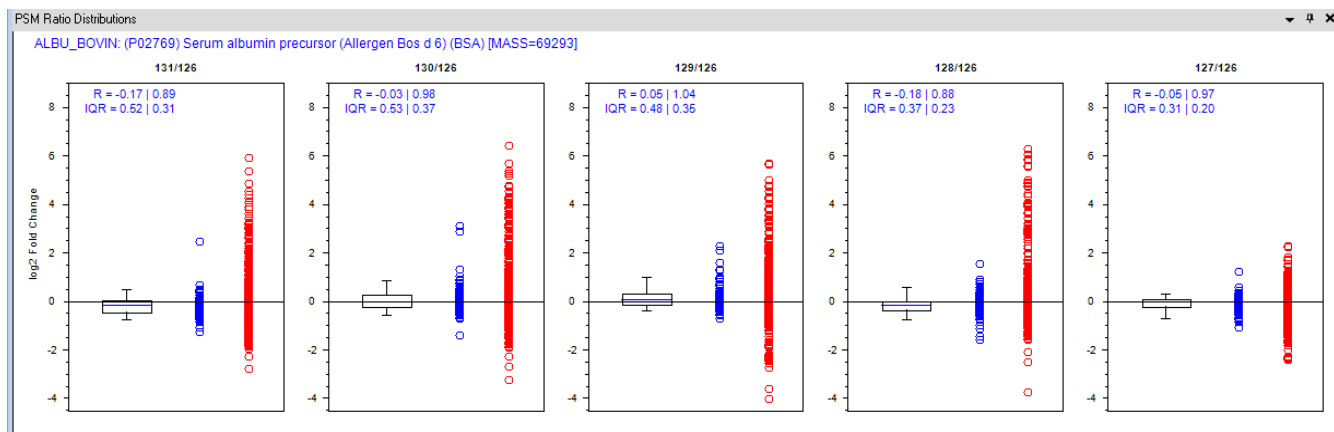
The chart shows the distribution of peptide group ratios for each of the ratios reported, as defined in the quantification method for this search. Each of the ratio distribution charts displays the peptide group ratios as the binary logarithm. The logarithmic form is common for such displays, because it provides a reasonable display, even when there is a large spread of the displayed values. In binary logarithmic form, a value of 1 means a two-fold increase, a value of 2 means a four-fold increase, a value of 3 means an eight-fold increase, and so forth.

Each of the separate distribution charts displays the peptide group ratios in three sections. The chart legend explains the meaning of these sections. You can access the chart legend by right-clicking the chart and choosing Show Legend.

The Quan Ratio Distribution charts contain the three sections illustrated in Figure 439:

- The first section displays the distribution of the ratios of all peptides considered for calculating the abundance of this protein as a box-and-whisker plot. A box-and-whisker plot is a convenient way of graphically depicting groups of numerical data through a five-number summary: 5 percent lower bound, lower quartile, median, upper quartile, 95 percent upper bound. The range between the lower and upper quartile (this is the range of the box) is also known as the inter-quartile range (IQR) and, like the standard deviation for normally distributed data, is a measure of the spread of the data.

  – The box represents the peptide group ratios between the 25th and the 75th percentiles.

  – The error bars represent the peptide group ratios below the 5th and the 95th percentiles.

  – The blue lines inside the horizontal bar represent the median of the distribution.

- The second section (blue circles) displays the distribution of the ratios of all peptide groups considered in calculating the protein abundances.

- The third section (red circles) displays the distribution of the ratios of all peptide groups that were not considered in calculating the protein abundance (for example, the peptide group is not unique to this protein or this protein group) according to the rules defined in the quantification method.

Figure 439. Quan Ratio Distributions chart for reporter ion quantification
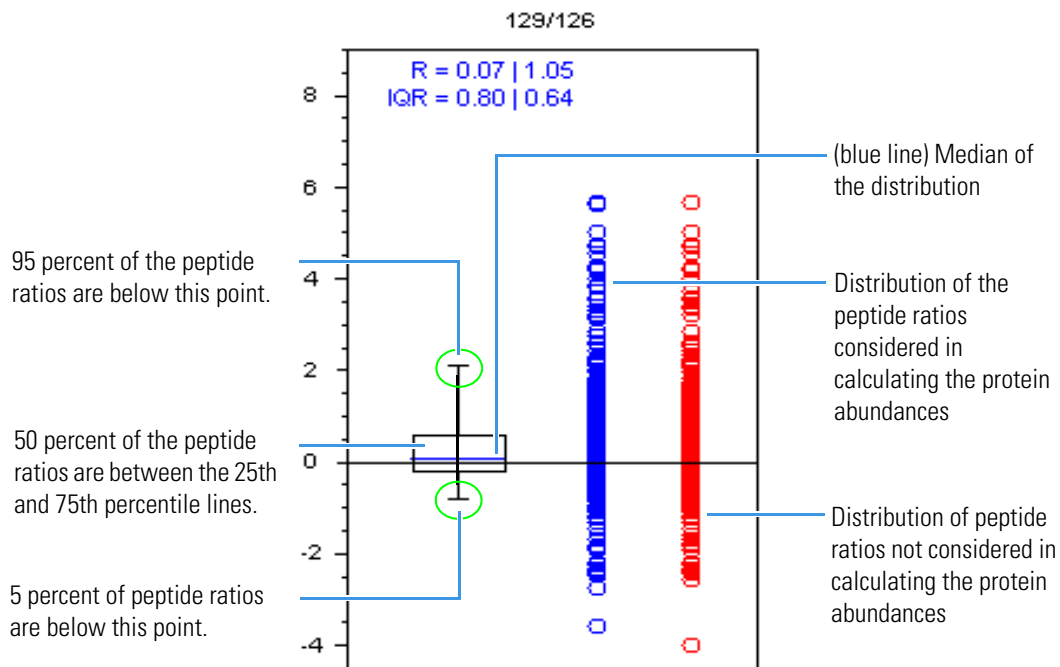
In addition, each chart displays the median ratio (R) and the inter-quartile range (IQR) in linear and logarithmic format. The header of the chart identifies the protein that the PSM belongs to. Right-click the chart and choose **Show Legend** for the identity of other notations on the chart.

Displaying the Report Item Distribution Chart as a Volcano Plot

A volcano plot is a type of scatter plot used to quickly identify protein or peptide changes in large data sets composed of sample pairs. The left upper quadrant (highlighted in green) displays those proteins or peptide groups that are significantly downregulated. The upper right quadrant (highlighted in red) displays those proteins or peptide groups that are significantly upregulated. See Figure 440 on page 612 for an example. You can choose where the green and red highlighted areas are displayed by using the P-Value and Log_2Fold Change sliders. Use these areas to choose proteins or peptides that might be biologically relevant to your study.

The x axis displays the abundance ratio in log_2, and the y axis displays p-values in log_{10}.

For information on p-values, see Calculating P-Values and Adjusted P-Values for Quantification Results.

The .pdResult report must contain replicate quantification data in order for you to display this plot or to run a new analysis.

❖ **To display the Report Item Distribution chart as a volcano plot**

1. Open the .pdResult file of interest.

2. Click the **Proteins** or **Peptide Groups** tab.

3. Click the protein or peptide group of interest.

4. Choose **View > Distribution Charts**, or click the **Distribution Charts** icon, , in the toolbar.

5. Click the **Volcano Plots** tab.

6. From the Data Source list, select the type of data to plot, either **Proteins** or **Peptide Groups**.

7. From the Comparison list, select the items to be compared.

8. Drag the **P-Value** slider to set the range of p-values (shaded areas) along the y axis, or type the appropriate number in the box to the right of the slider and press the RETURN or ENTER key.

9. Drag the **Log2 Fold Change** slider to set the range of abundance values (shaded areas) along the x axis, or type the appropriate number in the box to the right of the slider and press the RETURN or ENTER key.

Figure 440 gives an example of a volcano plot that compares two tissues, kidney and liver.

The data points are color-coded as defined by the legend.

Figure 440. Volcano plot
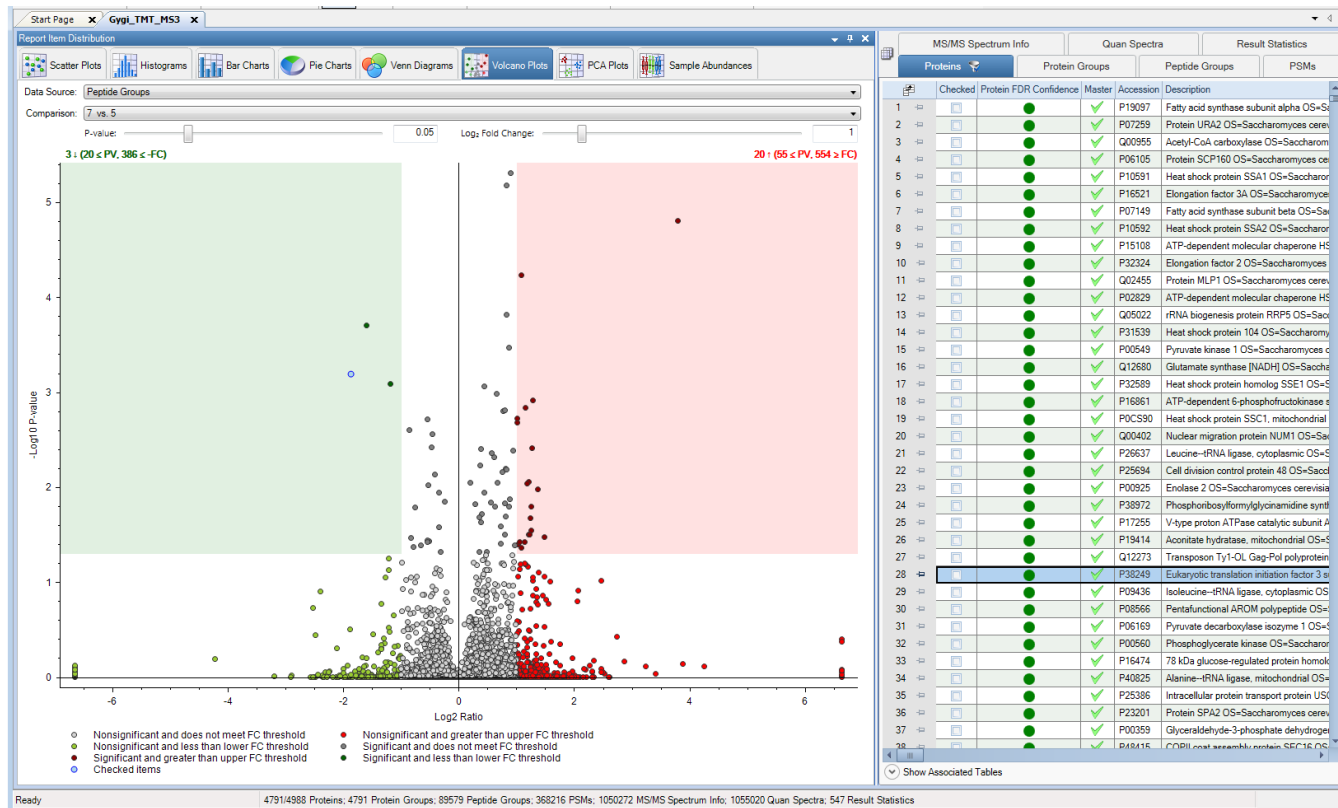
If the .pdResult file contains p-values and ratios, the volcano plot opens with the differential analysis from the data processing. The ratios in the Comparison list match the ratios on the Ratios and Ratio Groups page of the Quantification page in the Result Summaries view (see Figure 441). The initial p value setting is 0.05 (–log10 0.05 = 1.3), and the initial Log_2 Fold Change setting is 1 (a ratio of 2 to 1). Depending on the setting of the Log_{10} Transform parameter in the Reporter Ions Quantifier node or the Precursor Ions Quantifier node, the y-axis scale spans the p-value range (0–1) or the –log10 p-value range (–$\log_{10}$ 0 = infinity, –$\log_{10}$ 1 = 0).

Figure 441. Identical protein ratios in the Comparison list and on the Ratios and Ratio Groups summary page
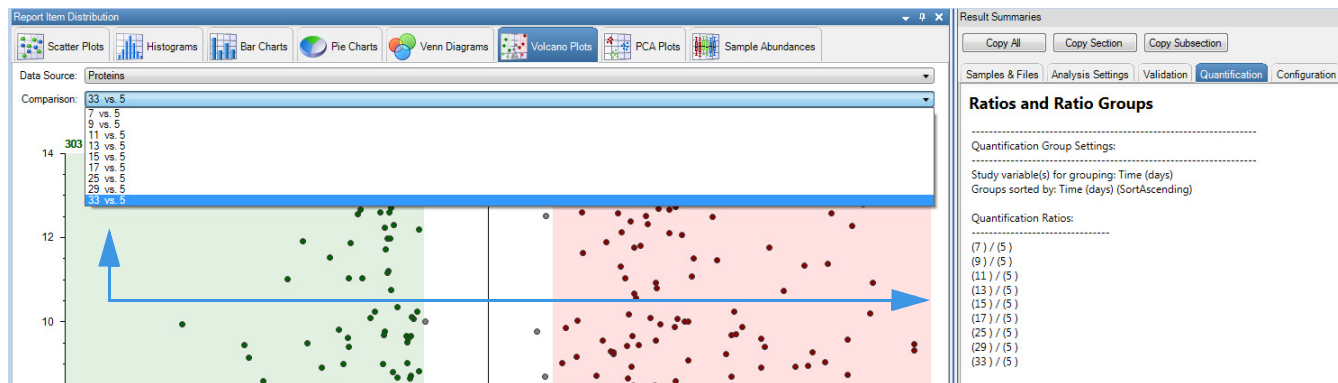
The volcano plot highlights selected items in blue. You can select and clear items of interest. The corresponding items are then selected or cleared in the .pdResult file page.

❖ **To display the legend**

Right-click the plot and choose **Show Legend**.

Figure 440 gives an example of a legend below the plot.

❖ **To change the analysis displayed in the volcano plot**

Select a different ratio from the Comparison list.

This step changes the data points in the plot.

❖ **To change the p-value setting**

Drag the **P-Value** slider to the left or right, or type the p-value in the box to the right of the slider and press the RETURN or ENTER key.

This step changes the *y*-axis range of the shaded areas.

❖ **To change the log$_2$ fold change setting**

Drag the **Log$_2$ Fold Change** slider to the left or right, or type the fold change value in the box to the right of the slider and press the RETURN or ENTER key.

This step changes the *x*-axis range of the shaded areas.

❖ **To select a single data point in the volcano plot**

Right-click the data point in the volcano plot, and choose **Check Point**.

The selected point turns blue in the plot, and in the Checked column of the .pdResult report, the application selects the check box of the corresponding protein or peptide group.

❖ **To clear a single data point in the volcano plot**

Right-click the data point in the volcano plot, and choose **Uncheck Point**.

The selected point resumes its original color in the plot, and in the Checked column of the .pdResult report, the application clears the check box of the corresponding protein or peptide group.

❖ **To select all visible data points**

Right-click the volcano plot, and choose **Check All Visible Points**.

All visible data points turn blue in the plot, and the legend disappears. In the Checked column of the .pdResult report, the application selects the check boxes of the corresponding proteins or peptide groups.

❖ **To clear all visible data points**

Right-click the volcano plot, and choose **Uncheck All Visible Points**.

All visible data points resume their original colors in the plot, and the legend reappears. In the Checked column of the .pdResult report, the application clears the check boxes of the corresponding proteins or peptide groups.

❖ **To display ToolTips**

Right-click the view and choose **Show Position Tooltips**.

The ToolTips display the location of the cursor by time and peak height.

❖ **To enlarge the view**

Drag your cursor to the right and select the area to enlarge in size.

❖ **To to reduce the view**

• Drag your cursor to the left and select the area to reduce in size.

–or–

• Right-click the view and choose **Zoom Out**.

❖ **To return to the default view**

Right-click the chart and choose **Undo All Zoom/Pan**.

❖ **To copy the data points in a volcano plot in an image format**

1. Right-click the chart and choose **Copy > Image**.

2. Paste the contents of the chart into another document.

   The contents appear as an image of the scatter plot in the document.

❖ **To copy the data points in a volcano plot as text**

1. Right-click the chart and choose **Copy > Points**.

2. Paste the contents of the chart into another document.

   The contents appear as two textual columns in the document, Log_2 Ratio and $-Log_{10}$ P-value, which correspond to each point's location coordinates in the plot.

❖ **To copy the Proteins or Peptide Groups page as text in CSV format**

1. Right-click the chart and choose **Copy > Point Details**.

2. Paste the contents of the chart into another document.

   All items on the selected page (Proteins or Peptide Groups) appear in a CSV-format text file.

❖ **To export the data points from the volcano plot as an image**

1. Right-click the volcano plot and choose **Export > Image As**.

2. In the Save As dialog box, type the name of the file to save the image in.

3. Click **Save**.

   The application saves the volcano plot as an Enhanced Metafile (EMF) file.

❖ **To export the data points from the volcano plot as text**

1. Right-click the volcano plot and choose **Export > Points As**.

2. In the Save As dialog box, type the name of the file to save the image in.

3. Click **Save**.

   The application saves the volcano plot as a text file that contains $\log_2$ ratios and Log_{10} p-values.

❖ **To export the Proteins or Peptide Groups page as text in CSV format**

1. Right-click the volcano plot and choose **Export > Point Details As**.

2. In the Save As dialog box, type the name of the file to save the image in.

3. Click **Save**.

   All items on the selected page (Proteins or Peptide Groups) appear in a CSV-format text file.

Parameters in the Volcano Plots View of the Report Item Distribution Chart

Table 38 lists the parameters in the Volcano Plots view of the Report Item Distribution chart.

Table 38. Parameters in the Volcano Plots view of the Report Item Distribution chart (Sheet 1 of 2)

Parameter	Description
Data Source	Specifies the type of data to plot from: • Proteins: From protein data. • Peptide Groups: From peptide group data.
Comparison	Specifies the items to be compared.

Table 38. Parameters in the Volcano Plots view of the Report Item Distribution chart (Sheet 2 of 2)

Parameter	Description
P-value	Sets the p-value, which is the likelihood that abundances of a particular peptide across different samples are not different. For information on p-values, see Calculating P-Values and Adjusted P-Values for Quantification Results.
	Drag the slider to change the y-axis range of the shaded areas, or type the appropriate number in the box to the right of the slider and press the RETURN or ENTER key.
Log_2 Fold Change	Sets the Log_2 fold change, which is the ratio of the difference between the last value and the first value over the original value. (If the initial value is A and final value is Z, the fold change is (Z - A)/A or Z/A - 1.)The fold change is expressed in log_2 format.
	Drag the slider to change the x-axis range of the shaded areas, or type the appropriate number in the box to the right of the slider and press the RETURN or ENTER key.
Log_{10} P-value (y axis)	Displays p-values in log_{10}.
Log_2 Ratio (x axis)	Displays the abundance ratio in log_2.
Legend	Displays the meaning of the colors of groups of proteins or peptide groups.
	By default, the legend is hidden. Choose Show Legend from the shortcut menu to display the legend.

Shortcut menu commands in the Volcano Plots View of the Report Item Distribution Chart

Table 39 describes the commands in the shortcut menu of the Volcano Plots view.

Table 39. Commands in the shortcut menu of the Volcano Plots view (Sheet 1 of 2)

Command or Option	Description
Show Position Tooltips	Displays the following information:
	• *number, number*: Log$_2$ ratio
	• P*number*: P value
	• MW *number* [kDa]: Molecular weight of the protein, in kilodaltons
	• calc. PI=*number*: Isoelectric point of the protein
	• #Peptides=*number*: Number of unique peptides
	• #PSMs=*number*: Number of peptide spectrum matches
	• #AAs=*number*: Number of amino acids in the protein
Zoom Out	Reduces the view.
Undo All Zoom/Pan	Returns the volcano plot to the default view.
Copy	Copies the data in the plot view.
Image	Copies the data in the plot view as an image.
Points	Copies the data in the plot view as text.
Point Details	Copies all items on the selected page (Proteins or Peptide Groups) into a CSV-format text file.
Export	Exports the data in the plot view to a file.
Image As	Exports the data in the plot view as an image in an Enhanced Metafile (EMF) file.
Points As	Exports the data in the plot view as text in a TXT file.
Point Details As	Exports all items on the selected page (Proteins or Peptide Groups) into a CSV-format text file.
Show Legend	Displays the meaning of the colors of groups of proteins or peptide groups.
Check Point	Selects a data point in the plot view, colors the point blue, and selects the check box of the corresponding protein or peptide group in the Checked column of the .pdResult report.

Table 39. Commands in the shortcut menu of the Volcano Plots view (Sheet 2 of 2)

Command or Option	Description
Uncheck Point	Clears a data point in the plot view and returns the point to its original color in the plot. In the Checked column of the .pdResult report, the application clears the check box of the corresponding protein or peptide group.
Check All Visible Points	Selects all data points in the plot view, colors the points blue, and selects the check boxes of the corresponding proteins or peptide groups in the Checked column of the .pdResult report.
Uncheck All Visible Points	Clears all data points in the plot view, returns the points to their original colors, and re-displays the legend. In the Checked column of the .pdResult report, the application clears the check boxes of the corresponding proteins or peptide groups.

Displaying the Report Item Distribution Chart as a Principal Component Analysis Plot

A principal component analysis (PCA) plot identifies the major components in a set of data. Principal components are groups of features in the data that help the most to distinguish one sample from another. The Proteome Discoverer application automatically calculates these principal components.

A principal component analysis displays the correlation between multivariate data in a set of observations. It transforms a set of observations of possibly correlated variables into an artificial set of independent linear combinations of the original variables known as principal components (PC1, PC2, PC3, and so on). PC1 has the most variation, and the highest principal component has the least variation.

The principal component analysis plot has three pages:

- The Scores Plot page—Shows the correlation among the observations. For details, see Scores Plot.

- The Loadings Plot page—Shows the relations among the variables for a given pair of principal components. For details, see Loadings Plot.

- The Variances Plot page—Shows the proportion and the cumulative proportion of the variance contributed by each principal component. For details, see Variances Plot.

The .pdResult report must contain multiple quantification values for you to display this plot or run a new differential analysis.

❖ **To display the Report Item Distribution chart as a PCA plot**

1. Open the .pdResult file of interest.

2. Click the **Proteins** or **Peptide Groups** tab.

3. Choose **View > Distribution Charts**, or click the **Distribution Charts** icon, [icon], in the toolbar.

4. Click the **PCA Plots** tab.

5. (Optional for the scores plot only) In the Group By area at the far left, select one or more of the following check boxes to create sample groups according to the study factors for quantification:

    • *Study_factors*: Lists each study factor that you defined in the study.

    • **Sample**: Selects the sample mixture of all samples. The Sample check box is based on the Sample Type default study variable on the Grouping & Quantification page. It appears whether or not you select Sample Type on the Grouping & Quantification page.

6. (Optional for the scores plot and the loadings plot only) In the Filter By area at the far left, set one or more of the following variables to filter the data:

    a. Set the *Study_factor* filter to **On**.

    b. Set the Sample filter to **On**.

    c. Click the arrow to the left of the On/Off filter box for *Study_factor* or Sample.

    d. Select the appropriate check boxes below *Study_factor*, Sample, or both.

7. From the Data Source list, select the result category to plot the data from, either **Proteins** or **Peptide Groups**.

8. In the X Data list, select an appropriate principal component.

9. In the Y Data list, select an appropriate principal component.

10. (Optional) Select the **Center and Scale** check box when you want to mean-center the values in a covariance matrix.

    If you clear this check box, the chart uses the original values in the covariance matrix.

11. (Optional) Select the **Use Normalized Abundances** check box when you want the chart to display normalized abundance values.

    This check box, selected by default, is only available when you normalize the data with the normalization parameters of the Reporter Ions Quantifier node or the Precursor Ions Quantifier node in the consensus workflow. If you clear this check box, the chart uses abundance values that are not normalized.

12. Click the tab of the type of plot that you want to view: **Scores Plot**, **Loadings Plot**, or **Variances Plot**.

Figure 442, Figure 443, Figure 444, and Figure 445 give examples of these three types of PCA plots. They show three different biological replicates of yeast growing on glucose and the effect of glucose depletion on them over time.

Scores Plot

The plot on the Scores Plot page indicates the similarity between samples.

Circles represent samples (which are linked to data files) as data points in this plot. Samples that lie physically near each other are similar. Both the *x* axis and *y* axis display principal components.

You can check the percentage values for the principal components. The labels on the *x* axis and *y* axis include the proportion of variance that the principal components add to the total variance as a percentage.

Point to a data point to view its coordinates, sample number, and replicate number.

Figure 442 and Figure 443 show examples of the PCA plot on the Scores Plot page.

In Figure 442, the Sample check box is selected, and the Time (days) check box is cleared. The three samples separate along the *x* axis and are differentiated by principal component 1. This grouping indicates that the samples in replicates 2 and 3 are more closely related than the samples in replicate 1.

Figure 442. PCA plot on the Scores Plot page with the Sample check box selected
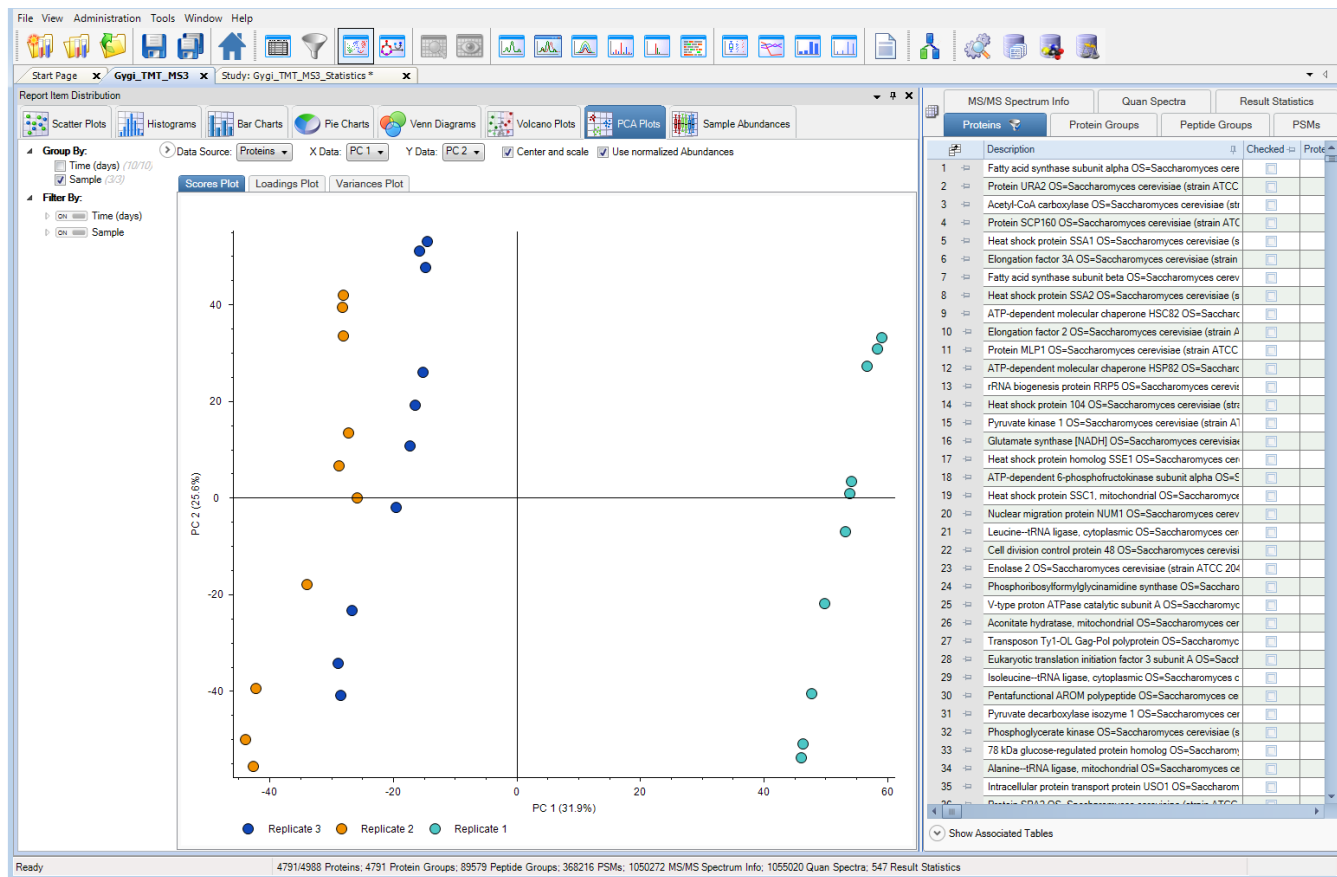
In Figure 443, the Time (days) check box is selected, and the Sample check box is cleared. The longitudinal pattern of the circle colors is the same in all three samples. The three samples are differentiated along the *y* axis, and the plot shows a common effect over time by principal component 2; that is, the effect of the decreasing amount of glucose is the same on all three samples. This plot implies that the same features are changing over the course of time.

Figure 443. PCA plot on the Scores Plot page with the Time (days) check box selected
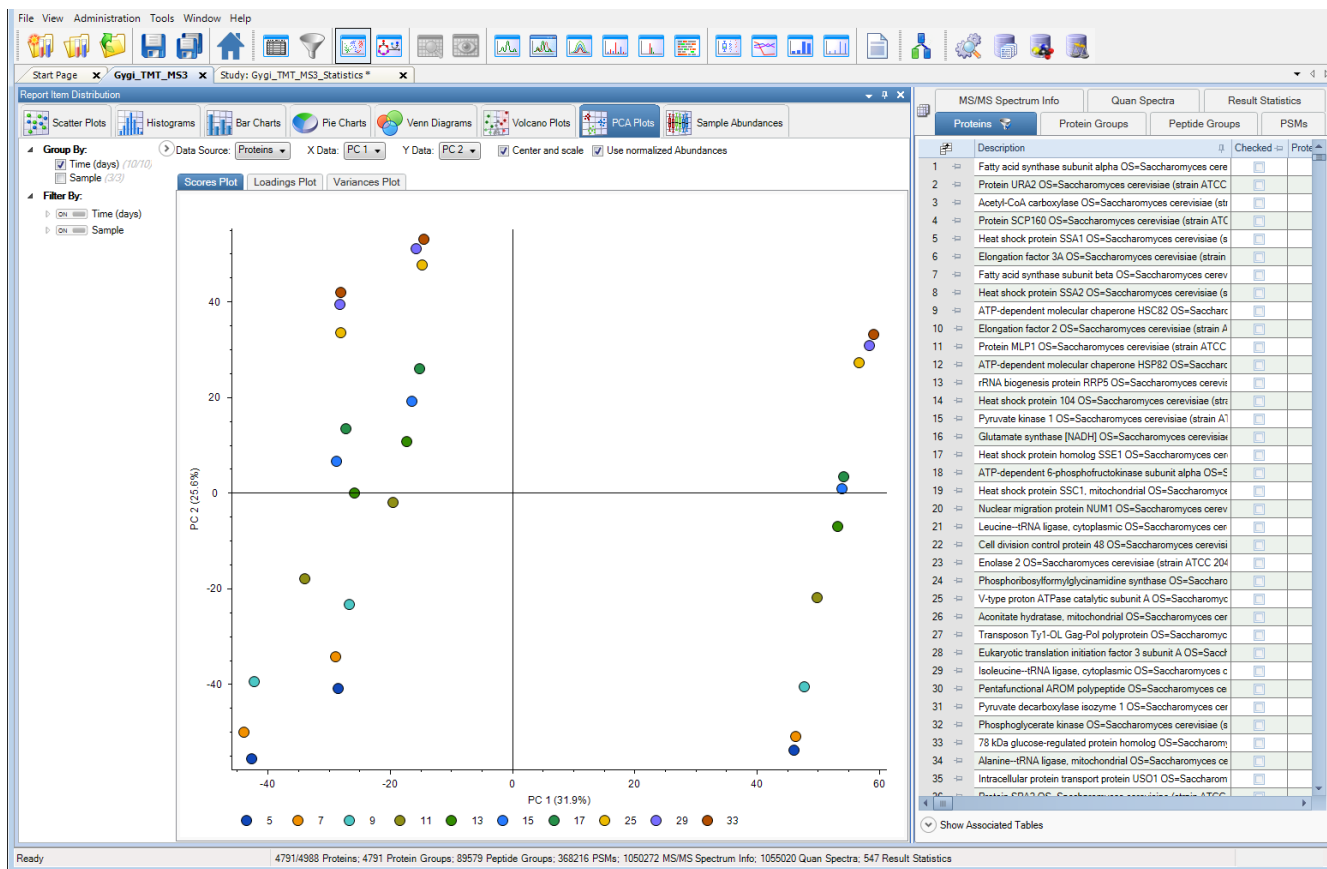
Loadings Plot

Use the plot on the Loadings Plot page to determine how each protein contributes to a principal component and to select individual proteins on the plot, especially those that have a comparable profile, to understand where proteins are changing similarly across all the samples.

Each blue circle, or data point, represents an individual protein measured across all samples. Data points that lie physically near each other are similar.

Both the *x* axis and *y* axis display principal components.

Data points that are on the opposite sides of the origin have an inverse correlation to each other. Data points that are higher on an axis or that are the farthest away from the origin (0) represent proteins that contribute more to the principal component that helps the most to distinguish one sample from another.

Point to a data point to view the following information:

- *number, number*: The coordinates of the data point on the plot.

- *Accession_number*: The accession number.

- MW *number* [kDA]=*number*: The calculated molecular weight of the protein.

- Calc pI=*number*: The theoretically calculated isoelectric point for the protein, which is the pH at which a particular molecule carries no net electrical charge.

- # Peptides=*number*: The number of distinct peptide sequences in the protein.

- # PSMs=*number:* The number of identified peptide spectrum matches identified from all included searches, including those redundantly identified.

- #AAs=*number*: The length of the protein sequence.

Figure 444 gives an example of a PCA plot on the Loadings Plot page. The more tightly clustered areas show similar proteins. The right and the left areas show proteins that explain the difference between biological replicates. The top and the bottom areas show the change between different time points.

Figure 444. PCA plot on the Loadings Plot page
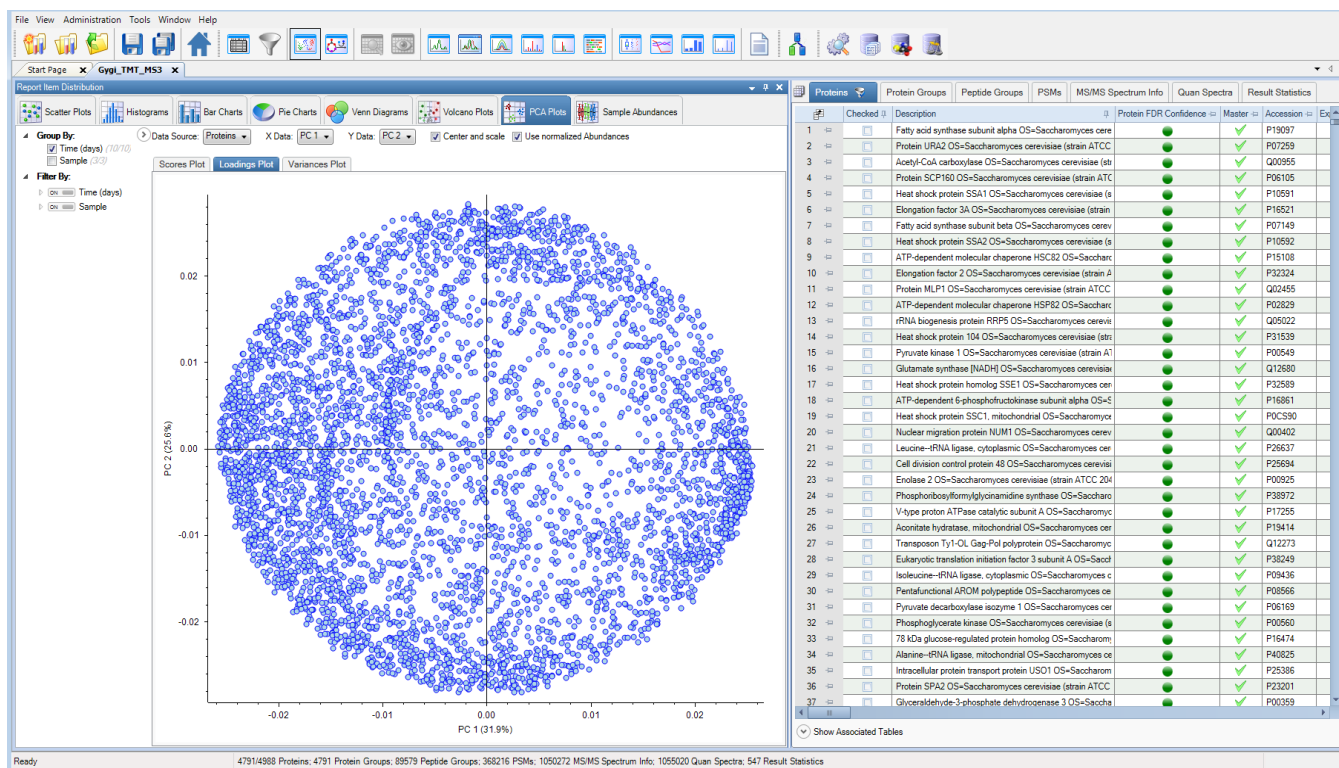
Variances Plot

Use the plot on the Variances Plot page to determine what proportion and cumulative proportion of the variance between samples is contributed by each principal component.

In this plot, squares represent the proportion of the variance as data points. Circles represent the cumulative proportion of the variance as data points. The *x* axis is the number of principal components, and the *y* axis is the proportion and the cumulative proportion of the variance contributed by each principal component.

Point to a data point square to view the number of the principal component and the sum of the proportion of variance of that principal component and that of its preceding principal components. Point to a data point circle to view the number of the principal component and the proportion of variance of that principal component.

In general, as the proportion of variance increases for the first two or three principal components, the dissimilarity between the sample groups increases.

Figure 445 gives an example of a PCA plot on the Variances Plot page. The first three principal components summed together equal 82 percent of the total variation between samples. In general, this plot shows that the principal component analysis is able to distinguish between the samples.

Figure 445. PCA plot on the Variances Plot page
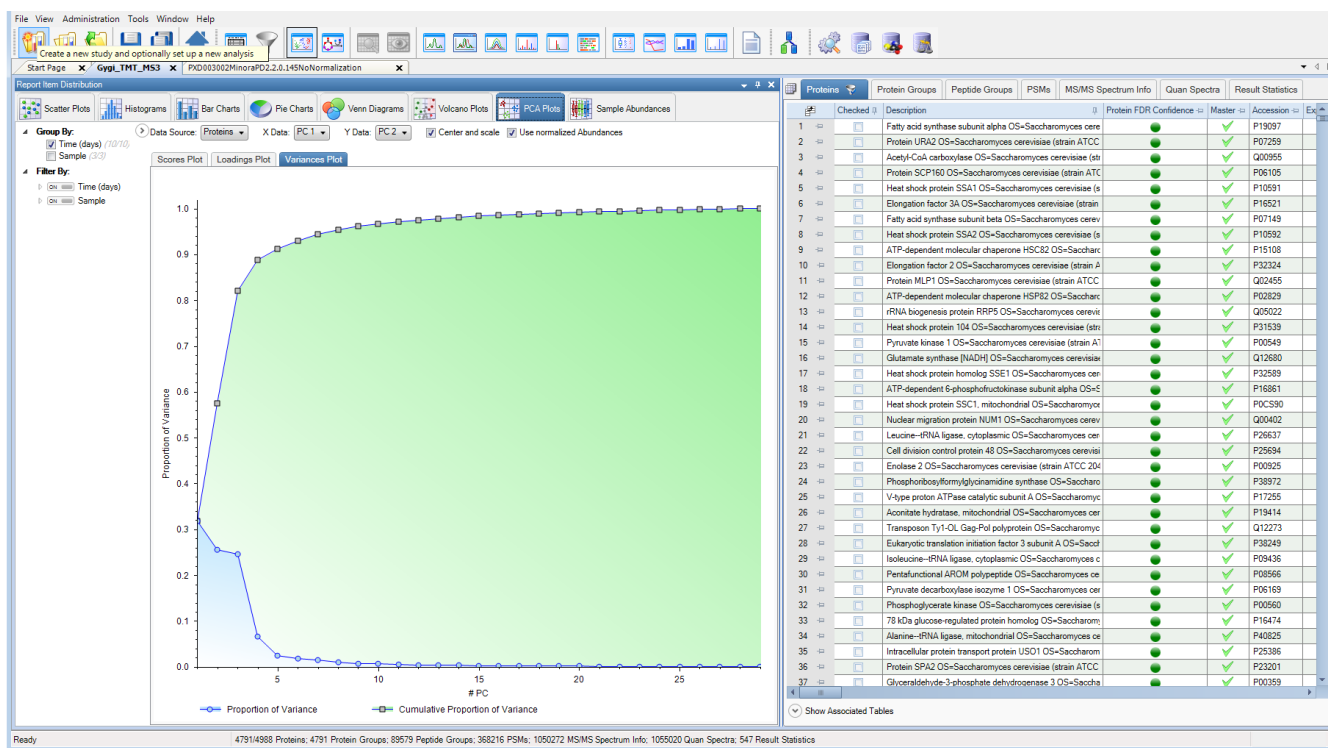
Parameters in the PCA Plots View of the Report Item Distribution Chart

Table 40 lists the parameters in the PCA Plots view of the Report Item Distribution chart.

Table 40. Parameters in the PCA Plots view of the Report Item Distribution chart (Sheet 1 of 3)

Parameter	Description
Group By	Creates sample groups according to the study factors for quantification: • *Study_factors*: Lists each study factor that you defined in the study. • Sample: Selects the sample mixture of all samples. The Sample check box is based on the Sample Type default study variable on the Grouping & Quantification page. It appears whether or not you select Sample Type on the Grouping & Quantification page. The Group By parameter only applies to the plots on the Scores Plot page.
Filter By	Filters the data according to the following variables: • *Study_factors*: Filters by the study factor that you defined in the study. • Sample: Filters by sample (see the description of Sample for the Group By parameter). The Filter By parameter only applies to the plots on the Scores Plot page and the Loadings Plot page.
Data Source	Specifies the result category to plot the data from: • Proteins: From protein data. • Peptide Groups: For peptide group data.
X Data	Specifies a principal component.
Y Data	Specifies a principal component.
Center and Scale	Mean-centers the values in a covariance matrix. If you clear this check box, the chart uses the original values in the covariance matrix.
Use Normalized Abundances	Displays normalized abundance values in the chart. The Use Normalized Abundances check box, selected by default, is only available when you normalize the data with the normalization parameters of the Reporter Ions Quantifier node or the Precursor Ions Quantifier node in the consensus workflow. If you clear this check box, the chart uses abundance values that are not normalized.

Table 40. Parameters in the PCA Plots view of the Report Item Distribution chart (Sheet 2 of 3)

Parameter	Description
Legend	Displays the group colors.
	Default: Hidden (Choose **Show Legend** from the shortcut menu to display the legend.)
Scores Plot page	
Circles	Represent samples as data points.
	When you choose the Show Position ToolTips command, pointing to a data point displays the following information:
	• *number*: The sample number.
	• Sample: The data point is a sample.
	• Replicate *number*: The number of the replicate.
	• *number, number*: The coordinates of the data point on the plot.
x-axis	Displays the principal component that you selected in the X Data box.
y-axis	Displays the principal component that you selected in the Y Data box.
Loadings Plot page	
Circles	Represent individual proteins as data points.
	When you choose the Show Position ToolTips command, pointing to a data point circle displays the following information:
	• *number, number*: The coordinates of the data point on the plot.
	• *Accession_number*: The accession number.
	• MW *number* [kDA]=*number*: The calculated molecular weight of the protein, in kilodaltons.
	• Calc pI=*number*: The theoretically calculated isoelectric point for the protein, which is the pH at which a particular molecule carries no net electrical charge.
	• # Peptides=*number*: The number of distinct peptide sequences in the protein.
	• # PSMs=*number:* The number of identified peptide spectrum matches identified from all included searches, including those redundantly identified.
	• #AAs=*number*: The length of the protein sequence.

Table 40. Parameters in the PCA Plots view of the Report Item Distribution chart (Sheet 3 of 3)

Parameter	Description
x-axis	Displays the values that you selected in the X Data box.
y-axis	Displays the values that you selected in the Y Data box.
Variances Plot page	
Circles	Displays the proportion of variance.
Squares	Displays the cumulative proportion of variance.
# PC (*x*-axis)	Displays the number of principal components.
Proportion of Variance (*y*-axis)	Displays the proportion and the cumulative proportion of the variance contributed by each principal component.

Shortcut Menu Commands in the PCA Plots View of the Report Item Distribution Chart

Table 39 describes the commands in the shortcut menu of the PCA Plots view.

Table 41. Commands in the shortcut menu of the PCA Plots view

Command or Option	Description
Show Position Tooltips	Displays the following information:
	• *number*: Group number
	• *channel_number*: Channel number
	• Replicate *number* - [*channel_number*]:
	• MW *number* [kDa]: Molecular weight of the protein, in kilodaltons
	• *Sample_type*: Indicates the sample type, which can be one of the following:
	– Sample: A specimen from a larger biological entity
	– Control: A sample typically used as a reference sample in a quantification experiment
	– Blank: A sample consisting only of solvent and no sample mixture
	– Standard: A sample consisting of a standard quality-control peptide mixture
	• F*number*: Replicate *number*:
	• *number*, *number*:
Zoom Out	Reduces the view.
Undo All Zoom/Pan	Returns the PCA plot to the default view.
Copy	Copies the data in the plot view.
Image	Copies the data in the plot view as an image.
Points	Copies the data in the plot view as text.
Export	Exports the data in the plot view to a file.
Image As	Exports the data in the plot view as an image in an Enhanced Metafile (EMF) file.
Points As	Exports the data in the plot view as text in a TXT file.
Show Legend	Displays the meaning of the colors of groups of proteins or peptide groups.

Displaying the Mass Recalibration View

The Proteome Discoverer application includes an algorithm for mass recalibration. This algorithm recalculates precursor mass independently per file. It calculates a correction curve by using support vector regression. The Mass Recalibration view indicates how well the regression performed. It can also provide insight into potential issues with the acquisition or chromatography itself. It displays the input data (called landmarks in this context) used for the regression and the calculated correction curve.

For mass recalibration, perform a fast Sequest HT search with a fairly wide precursor mass tolerance (2 ppm by default) to identify enough confident identifications to deduce the mass error from. You must use the Spectrum Files RC node instead of the Spectrum Files node in the processing workflow to generate a .pdResults file that can display the Mass Recalibration view.

For more information, see Mass Recalibration. For information on the Mass Recalibrations page of the .pdResult file, see .

❖ **To display the Mass Recalibration view**

1. In the .pdResult file, click the **Input Files** or the **Mass Recalibrations** tab (use the Result Table Chooser to make these tabs visible if necessary).

   The Input Files page displays the input data for the regression and the calculated correction curve.

   The Mass Recalibrations page indicates how well the recalibration performed for all the input files.

2. Select an input file on the Input Files page or the Mass Recalibrations page.

3. Choose **View > Mass Recalibration View**.

   The Mass Recalibration view opens (see Figure 446).

Figure 446. Mass Recalibration view
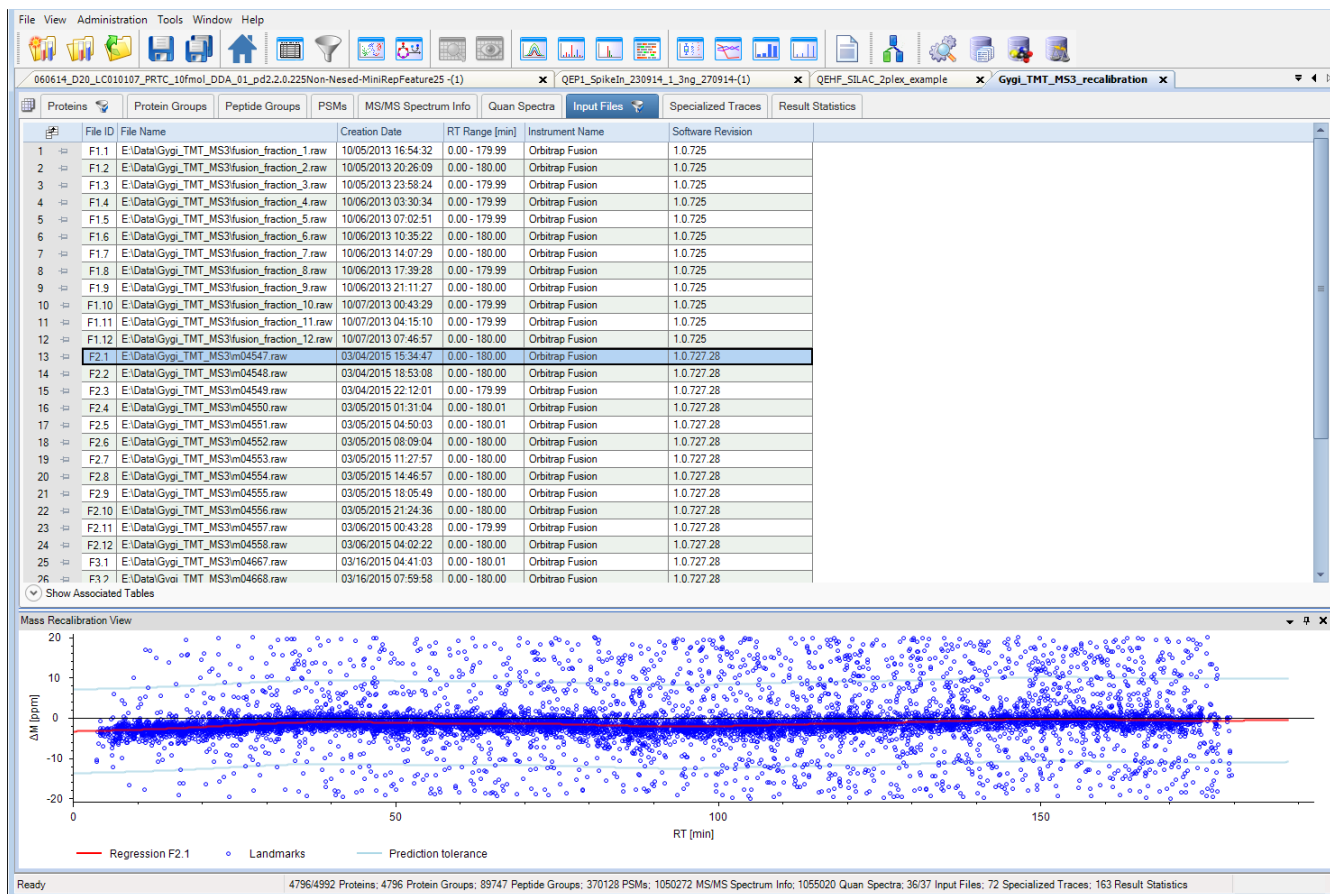
The Mass Recalibration view displays the distribution of detected mass errors over the retention time range. It also shows you how the application fit the correction curve to the data.

- The *x*-axis value used for the regression is the mass error between the measured mass for a given peptide and the theoretical mass based on the identified sequence, in parts per million.

- The *y*-axis value is the deviation of the experimental mass of the peptide match from the theoretical mass of the peptide.

Parameters in the Mass Recalibration View

Table 42 lists the parameters in the Mass Recalibration view.

Table 42. Mass Recalibration view parameters

Parameter	Description
ΔM [ppm]	Displays the mass error between the measured mass for a given peptide and the theoretical mass based on the identified sequence, in parts per million.
RT [min]	Displays the deviation of the experimental mass of the peptide match from the theoretical mass of the peptide.

Shortcut Menu Commands in the Mass Recalibration View

The shortcut menu in the Mass Recalibration view provides options to copy or export the plot as an image or the underlying data.

Table 43 describes the commands in the shortcut menu of the Mass Recalibration view.

Table 43. Shortcut menu commands in the Mass Recalibration view

Command or Option	Description
Show Position Tooltips	Displays the coordinates of the location of the mass error and the retention time of the landmark peptide.
Zoom Out	Reduces the view.
Undo All Zoom/Pan	Returns the Mass Recalibration view to the default view.
Copy	Copies the data in the Mass Recalibration view.
Image	Copies the data in the Mass Recalibration view as an image.
Points	Copies the data in the Mass Recalibration view as text.
Export	Exports the data in the Mass Recalibration view to a file.
Image As	Exports the data in the Mass Recalibration view as an image in an Enhanced Metafile (EMF) file.
Points As	Exports the data in the Mass Recalibration view as text in a TXT file.

Displaying the File Alignment View

The application includes an algorithm for retention time alignment. This algorithm calculates retention time across files. Like the mass recalibration algorithm, it calculates a correction curve by using support vector regression. The File Alignment view indicates how well the regression performed. It can also provide insight into potential issues with the acquisition or chromatography itself. It displays the input data (called *landmarks* in this context) used for the regression and the calculated correction curve.

Retention time alignment is based on finding common landmarks between two files, where one file is the reference file. As input for the regression, the application uses similar landmarks found in both files.

The .pdResult file must contain quantification results for the File Alignment view to be available.

❖ **To display the File Alignment view**

1. In the .pdResult file, click the **Input Files** tab (use the Result Table Chooser to make it visible if necessary).

   The Input Files page displays the input data for the regression and the calculated correction curve.

   Each file is aligned to one reference file.

2. Select an input file.

3. Choose **View > File Alignment View**.

   The File Alignment view opens (see Figure 447).

Figure 447. File Alignment view
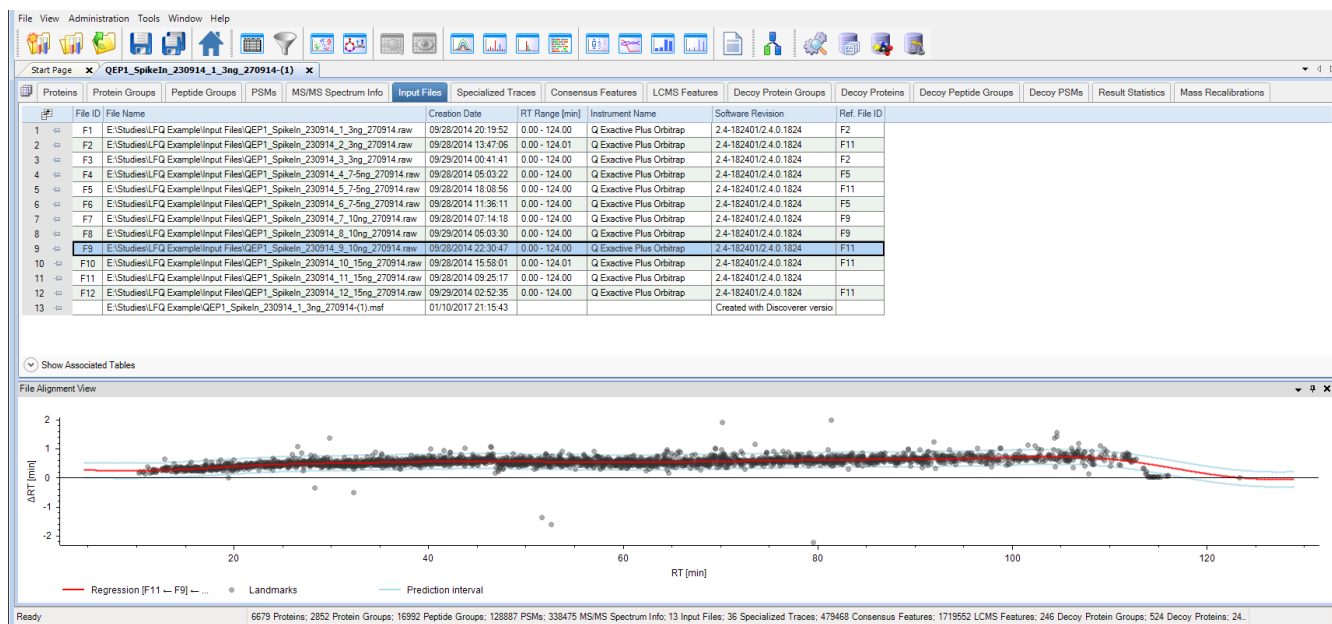
The File Alignment view displays how the selected file is aligned to its reference file. The reference file is shown in the Ref. File ID column of Input Files page and in the legend of the correction plots.

- The *x*-axis value used for the regression is the retention time of the identified peptides in the reference file.

- The *y*-axis value is the retention time difference between the landmarks.

Parameters in the File Alignment View

Table 44 lists the parameters in the File Alignment view.

Table 44. File Alignment view parameters

Parameter	Description
RT [min]	Displays the retention time of the identified peptides in the reference file.
ΔRT [min]	Displays the retention time difference between the landmarks.

Shortcut Menu Commands in the File Alignment View

The shortcut menu in the File Alignment view provides options to copy or export the plot as an image or the underlying data.

Table 45 describes the commands in the shortcut menu of the File Alignment view.

Table 45. Shortcut menu commands in the File Alignment view

Command or Option	Description
Show Position Tooltips	Displays the retention time of the selected file and the delta retention time of the same landmark in the reference file.
Zoom Out	Reduces the view.
Undo All Zoom/Pan	Returns the Retention Tine Alignment view to the default view.
Copy	Copies the data in the File Alignment view.
Image	Copies the data in the File Alignment view as an image.
Points	Copies the data in the File Alignment view as text.
Export	Exports the data in the File Alignment view to a file.
Image As	Exports the data in the File Alignment view as an image in an Enhanced Metafile (EMF) file.
Points As	Exports the data in the File Alignment view as text in a TXT file.

FASTA Reference

This appendix lists the most important FASTA databases and parsing rules that the Proteome Discoverer application uses to obtain protein sequences, accession numbers, and descriptions.

- FASTA Databases

- Custom Database Support

FASTA Databases

These are the most important FASTA databases that the application uses.

- NCBI

- UniRef100

- SwissProt and TrEMBL

Follow the links to each database if you would like to download the database and save it to your local machine. Some databases are more time-consuming to load than others.

NCBI

NCBI is a nonredundant database compiled by the NCBI (National Center for Biotechnology Information) as a protein database for Blast searches. It contains nonidentical sequences from GenBank CDS translations, Protein Data Bank (PDB), SwissProt, Protein Information Resource (PIR), and Protein Research Foundation (PRF).

http://blast.ncbi.nlm.nih.gov/Blast.cgi

ftp://ftp.ncbi.nih.gov/blast/db/FASTA/nr.gz

A typical NCBI title line follows:

> `>gi|70561|pir||MYHO myoglobin - horse_i|418678|pir||MYHOZ myoglobin - common zebra (tentative sequence) [MASS=16950]`

> FASTA ID:
> - `Accession#:gi70561`
> - `Description:myoglobin - horse_i`

UniRef100

UniRef, also known as UniProt NREF, is a set of comprehensive protein databases curated by the Universal Protein Resource consortium. UniRef100 contains only nonidentical sequences, and UniRef90, and UniRef50 are nonredundant at a sequence similarity level of 90 percent and 50 percent, respectively.

ftp://ftp.uniprot.org/pub/databases/uniprot/current_release/uniref/uniref100/

A typical UniRef100 title line follows:

> `>UniRef100_4U9M9 Cluster: 104 kDa microneme-rhoptry antigen precursor; n=1; Theileria annulata|Rep: 104 kDa microneme-rhoptry antigen precursor - Theileria annulata`

> FASTA ID:
> - `Accession#:4U9M9`
> - `Description:Cluster: 104 kDa microneme-rhoptry antigen precursor; n=1; Theileria annulata|Rep: 104 kDa microneme-rhoptry antigen precursor - Theileria annulata`

SwissProt and TrEMBL

The SwissProt database is developed by the SwissProt groups at the Swiss Institute of Bioinformatics (SIB) and the European Bioinformatics Institute (EBI).

TrEMBL is a computer-annotated supplement of SwissProt that contains all the translations of EMBL nucleotide sequence entries not yet integrated into SwissProt.

http://www.expasy.org/sprot/

ftp://ftp.ebi.ac.uk/pub/databases/uniprot/knowledgebase/uniprot_trembl.fasta.gz

A typical SwissProt title line follows:

> `>43495|108_SOLLC Protein 108 precursor - Solanum lycopersicum (Tomato) (Lycopersicon esculentum)`

> FASTA ID:108_SOLLC
> - `Accession#:43495`
> - `Description:Protein 108 precursor - Solanum lycopersicum (Tomato) (Lycopersicon esculentum)`

Custom Database Support

The Proteome Discoverer application has three "general" parsing rules to support custom sequence database formats. The generic parsing rules are applied only if no other parsing rule matches the given FASTA title line.

- Custom Parsing Rule A

- Custom Parsing Rule B

- Custom Parsing Rule C

Custom Parsing Rule A

The application uses custom parsing rule A if the FASTA ID, the accession#, and the description are separated by a pipe (|) symbol. A typical FASTA title line, which matches this parsing rule, looks like this:

```
>tr|18FC3|18FC3_HALWD IS1341-type transposase - Halouadratum walsbyi
(strain DSM 16790).
```

```
FASTA ID:18FC3_HALWD
```

- `Accession#:18FC3`

- `Description:IS1341-type transposase - Halouadratum walsbyi (strain DSM 16790).`

Custom Parsing Rule B

The application uses custom parsing rule B if the accession# and the description are separated by a space. A typical FASTA title line, which matches this parsing rule, looks like this:

```
>HP0001 hypothetical protein {Helicobacter pylori 26695}
```

```
FASTA ID:
```

- `Accession#:HP0001`

- `Description:hypothetical protein {Helicobacter pylori 26695}`

Custom Parsing Rule C

The application uses custom parsing rule C if the FASTA title line only contains the accession#. A typical FASTA title line, which matches this parsing rule, looks like this:

```
>143B_HUMAN
```

```
FASTA ID:
```

- `Accession#:143B_HUMAN`

- `Description:143B_HUMAN`

Chemistry References

The tables in this appendix list amino acid symbols and mass values, enzyme cleavage properties, and the fragment ions used in the Proteome Discoverer application.

- Amino Acid Mass Values

- Enzyme Cleavage Properties

- Fragment Ions

Amino Acid Mass Values

The application uses the amino acid symbols and mass values listed in Table 46 and Table 47.

Table 46. Amino acid mass values (Sheet 1 of 2)

Amino acid	One-letter code	Three-letter code	Monoisotopic mass	Average mass	Sum formula
Glycine	G	Gly	57.02146	57.05177	C_2H_3NO
Alanine	A	Ala	71.03711	71.07855	C_3H_5NO
Serine	S	Ser	87.03203	87.07796	$C_3H_5NO_2$
Proline	P	Pro	97.05276	97.11623	C_5H_7NO
Valine	V	Val	99.06841	99.13211	C_5H_9NO
Threonine	T	Thr	101.04768	101.10474	$C_4H_7NO_2$
Cysteine	C	Cys	103.00918	103.14464	C_3H_5NOS
Isoleucine	I	Ile	113.08406	113.15890	$C_6H_{11}NO$
Generic Leu/lle	J	Jli	113.08406	113.15890	$C_6H_{11}NO$
Leucine	L	Leu	113.08406	113.15890	$C_6H_{11}NO$
Asparagine	N	Asn	114.04293	114.10354	$C_4H_6N_2O_2$
Aspartic Acid	D	Asp	115.02694	115.08826	$C_4H_5NO_3$
Glutamine	Q	Gln	128.05858	128.13032	$C_5H_8N_2O_2$
Lysine	K	Lys	128.09496	128.17358	$C_6H_{12}N_2O$

Table 46. Amino acid mass values (Sheet 2 of 2)

Amino acid	One-letter code	Three-letter code	Monoisotopic mass	Average mass	Sum formula
Glutamic Acid	E	Glu	129.04259	129.11504	$C_5H_7NO_3$
Methionine	M	Met	131.04048	131.19820	C_5H_9NOS
Histidine	H	His	137.05891	137.14062	$C_6H_7N_3O$
Phenylalanine	F	Phe	147.06841	147.17571	C_9H_9NO
Arginine	R	Arg	156.10111	156.18707	$C_6H_{12}N_4O$
Tyrosine	Y	Tyr	163.06333	163.17512	$C_9H_9NO_2$
Tryptophan	W	Trp	186.07931	186.21220	$C_{11}H_{10}N_2O$

Table 47. Special amino acids

Amino acid	One-letter code	Three-letter code	Monoisotopic mass	Average mass	Sum formula
Avrg. N/D	B	Bnd	114.53494	114.59590	
Avrg. Q/E	Z	Zqe	128.55059	128.62326	
Unknown acid (X)	X	Xxx	0	0	N/A
Pyrrolysine	O	Pyl	237.14773	237.29874	$C_{12}H_{19}N_3O_2$
Seleno Cysteine	U	Sec	150.95364	150.0369	C_3H_5NOSe

Enzyme Cleavage Properties

Table 48 lists the enzymes and reagents with cleavage properties.

Table 48. Cleavage properties of enzymes and reagents (Sheet 1 of 2)

Enzymes/Reagents	Cleaves after	Cleaves before	Except when
Enzymes for digestion			
AspN		D	
Chymotrypsin	F, W, Y, or L		
Chymotrypsin (FWY)	F, W, or Y		P is after F, W, or Y
Clostripain	R		
Elastase	A, L, I, or V		P is after A, L, I, or V
Elastase/Tryp/Chymo	A, L, I, V, K, R, W, F, or Y		P is after Al, L, I, V, K, R, W, F, or Y

Table 48. Cleavage properties of enzymes and reagents (Sheet 2 of 2)

Enzymes/Reagents	Cleaves after	Cleaves before	Except when
GluC	E or D		
LysC	K		
No-Cleavage			
No-Enzyme			
Proline_Endopept	P		
Staph_protease	E		
Trypsin	K or R		P is after K or R
Trypsin (KRLNH)	K, R, L, N, or H		
Trypsin_K	K		P is after K
Trypsin_R	R		P is after R
Chemicals for degradation			
Cyanogen bromide	M		
Iodobenzoate	W		

Fragment Ions

Several different fragmentation techniques produce fragment ions of peptides, such as ECD, ETD, CID, EThcD, higher-energy C-trap dissociation (HCD), and infrared multi-photon dissociation (IRMPD).

As an example, MS/MS and ESI generate low-energy CID spectra, which are sequence-specific. The fragment ion spectra contain peaks of the fragment ions formed by the cleavage of the peptide bond and are used to determine amino acid sequences. A fragment must have at least one charge for it to be detected.

The fragment ions produced are identified according to where they are fragmented in the peptide. A, b, and c fragment ions have a charge on the N-terminal side, and x, y, and z fragment ions have a charge on the C-terminal side. Fragment ions a*, b*, and y* are ions that have lost ammonia (–17 Da), and fragment ions a°, b°, and c° are ions that have lost water (–18 Da). The subscript next to the letter indicates the number of residues in the fragment ion.[8]

Table 49 summarizes the fragment ions used in the application.

---

[8] For more information about fragment ions and nomenclature, see Roepstorff, P.; Fohlman, J. Proposal for a Common Nomenclature for Sequence Ions in Mass Spectra of Peptides. *Biomed. Mass Spectrum.* **1984**, *11* (11) 601.

Table 49. Fragment ions

Ions	Description
a	A ion with charge on the N-terminal side
b	B ion with charge on the N-terminal side
c	C ion with charge on the N-terminal side
y	Y ion with charge on the C-terminal side
z	Z ion with charge on the C-terminal side
a*	A ion that has lost ammonia (–17 Da)
b*	B ion that has lost ammonia (–17 Da)
y*	Y ion that has lost ammonia (–17 Da)
a°	A ion that has lost water (–18 Da)
b°	B ion that has lost water (–18 Da)
c°	C ion that has lost water (–18 Da)

Technical and Biological Replicates

This appendix explains what replicates are and how to specify them in your experiments.

- Definitions

- Specifying Replicates

Definitions

Replicates are measurements of samples with the same conditions for the study variables that differentiate them within your analysis.

- Biological replicates are derived from different biological samples. They are repetitions of an experimental condition that allow you to estimate the variability associated with the phenomenon.

- Technical replicates are derived from the same biological sample. They are repeated measurements of the same item that allow you to calculate measurement errors. The application deals with biological and technical replicates differently in data analysis.

Specifying Replicates

You can specify biological replicates in a study. This information is required for calculating abundances and p-values with the Precursor Ions Quantifier node or the Reporter Ions Quantifier node.

Experiments can be divided into two groups, non-nested and nested.

Non-Nested Experiments

In non-nested experiments, biological replicates are independent. Figure 448 shows a study that investigates six different mice: three are untreated and three are treated. Each mouse has three different measurements. In this example, each mouse is a biological replicate, and the different measurements are technical replicates.

Figure 448. Non-nested biological and technical replicates
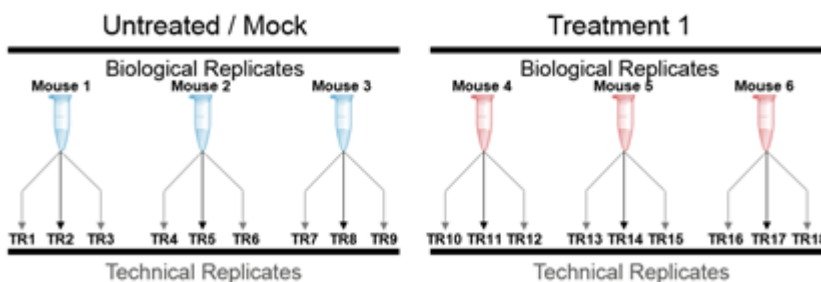
Non-nested longitudinal experiments are an extension of non-nested experiments. In these types of experiments, experimental groups are ordered, and you can test for temporal trends in the data. Figure 449 shows non-nested longitudinal biological and technical replicates.

Figure 449. Non-nested longitudinal biological and technical replicates

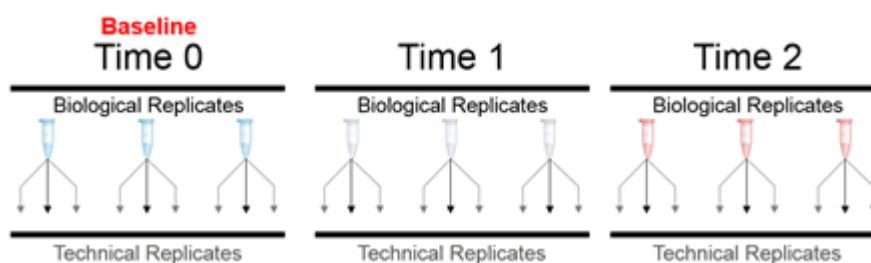In non-nested experiments, the application calculates abundances (see Figure 450). After summing up sample abundances and normalization, it calculates abundances for the replicates as the median of the technical replicates, and then it calculates the sample group abundances as the median of the biological replicate abundances. It calculates ratios as the ratio of sample group abundances.

Figure 450. Calculation of abundances in non-nested experiments

Summed abundances of all technical replicates **separately:**

	Untreated						Treated					
Mouse 1		Mouse 2		Mouse 3		Mouse 4		Mouse 5		Mouse 6	
TR1	TR2	TR1	TR2	TR1	TR2	TR1	TR2	TR1	TR2	TR1	TR2
98	97	105	107	101	103	158	168	150	144	172	168

For each bio rep calculate the median of the tech reps

Untreated		
Mouse 1	Mouse 2	Mouse 3
97.5	106	102

Treated		
Mouse 4	Mouse 5	Mouse 6
163	147	170

For each treatment calculate the median of the bio reps

| Untreated |
|---|
| 102 |

Treated
163

Calculate the ratios of the treatments

Treated / Untreated
1.60

Nested Experiments

In nested experiments, biological replicates are dependent; they might or might not be ordered. The different sample groups might represent the same individual under different conditions. In this case, you can have three different individuals, such as mouse 1, 2, and 3, under two different conditions, Untreated and Treatment 1 (see Figure 451).

Figure 451. Calculation of abundances in nested experiments
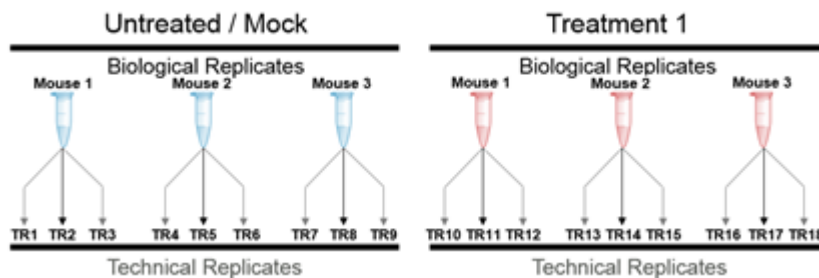
In nested experiments, the application calculates abundances for biological replicates and sample groups as it does for non-nested designs, but it calculates the ratios differently. First, it calculates the ratios for biological replicates as a ratio of their abundances, and then it calculates sample group ratios as the median of the replicate ratios.

Figure 452. Calculation of abundances in nested experiments

Summed abundances of all technical replicates separately:

Untreated						Treated					
Mouse 1		Mouse 2		Mouse 3		Mouse 1		Mouse 2		Mouse 3	
TR1	TR2	TR1	TR2	TR1	TR2	TR1	TR2	TR1	TR2	TR1	TR2
98	97	105	107	105	103	158	168	150	144	172	168

Median within technical replicates

Untreated			Treated		
Mouse 1	Mouse 2	Mouse 3	Mouse 1	Mouse 2	Mouse 3
97.5	106	104	164	147	170

Ratios (treated / untreated) within individuals

| Mouse 1	Mouse 2	Mouse 3
1.68	1.37	1.63

Median ratios

Treated / Untreated
1.63

Biological Replicates in the Proteome Discoverer Application

To specify biological replicates, you must add a study factor called the Biological Replicate Factor (see Figure 453). You can add only one biological replicate factor; if one already exists, the application makes the Biological Replicate Factor menu item unavailable.

Figure 453. Biological Replicate Factor study factor
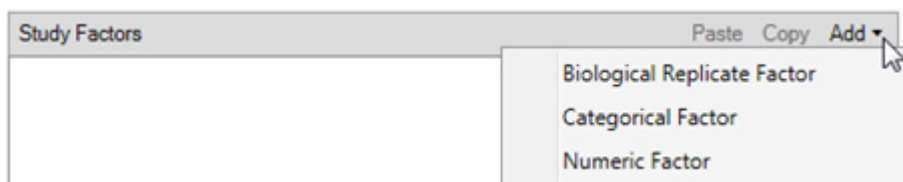
The application displays a biological replicate factor with a specific Replicate icon (see Figure 454) so that you can easily see which factor is a replicate.

Figure 454. Replicate icon
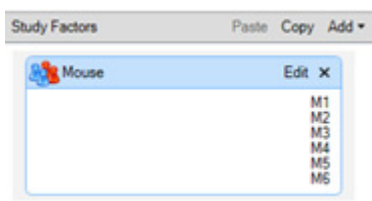
When the study contains a replicate factor, the Grouping and Quantification page in the study shows which biological replicates are contained in the generated sample groups. Within each sample group, the application groups samples belonging to the same biological replicate and highlights them with a yellow box showing the replicate name. Figure 455 shows samples grouped by the time course time point. There are several sample groups, each a different time point in the time course of this experiment. The replicates are grouped together in yellow boxes containing the replicate name appended to the front of each quan channel.

Figure 455. Samples grouped by treatment
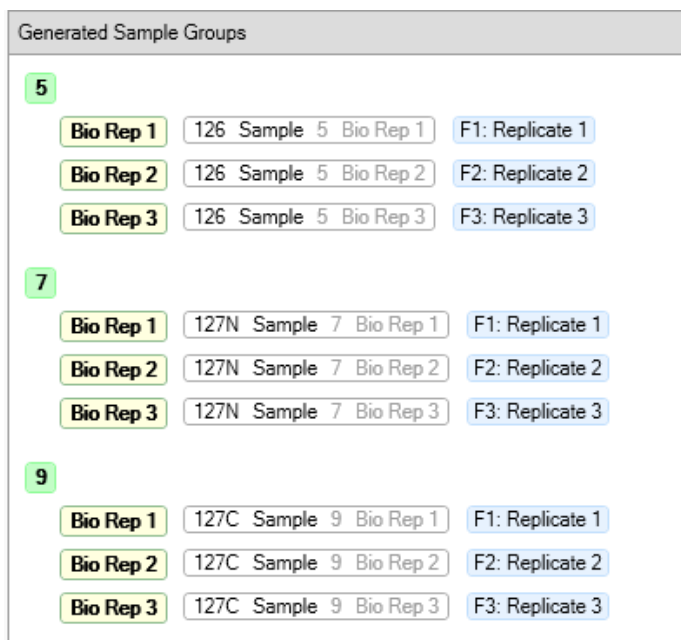
The Generated Ratios pane on the Grouping and Quantification page shows which ratios you created.

In nested experiments, where each sample group contains the same biological replicates, the pane also shows that sample group ratios are created using replicate ratios.

When there are biological replicates in the study, the Precursor Ions Quantifier node or the Reporter Ions Quantifier node creates columns showing abundances of biological replicates for proteins and peptide groups. It calculates abundances and ratios as described earlier. It calculates abundances for biological replicates using sample abundances, which you can normalize and scale, as appropriate.

The node can calculate p-values for non-nested experiments. If the ANOVA (Individual Proteins) p-value method is selected, it can calculate p-values for nested experiments only when each sample group includes two biological replicates. For the ANOVA (Background-Based) method, p-values can be calculated even when there is only a single biological replicate.