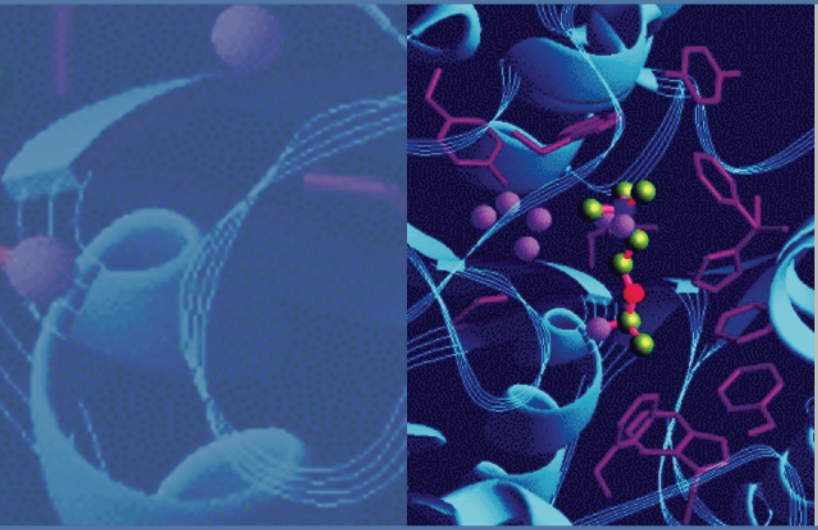


Xcalibur

Proteome Discoverer

User Guide

XCALI-97232 Revision A July 2008

Thermo
SCIENTIFIC

© 2008 Thermo Fisher Scientific Inc. All rights reserved.

Xcalibur is a registered trademarks of Thermo Fisher Scientific Inc. in the United States. ZCore, and Proteome Discoverer are trademarks of Thermo Fisher Scientific Inc. in the United States.

Adobe® and Acrobat® are registered trademarks of Adobe Systems Incorporated. Microsoft®, Excel®, Windows®, and ActiveX® are registered trademarks of Microsoft Corporation. SEQUEST® is a registered trademark of the University of Washington. Mascot is first defined in an Electrophoresis publication, 20(18) 3551-67 (1999).

Thermo Fisher Scientific Inc. provides this document to its customers with a product purchase to use in the product operation. This document is copyright protected and any reproduction of the whole or any part of this document is strictly prohibited, except with the written authorization of Thermo Fisher Scientific Inc.

The contents of this document are subject to change without notice. All technical information in this document is for reference purposes only. System configurations and specifications in this document supersede all previous information received by the purchaser.

Thermo Fisher Scientific Inc. makes no representations that this document is complete, accurate or error-free and assumes no responsibility and will not be liable for any errors, omissions, damage or loss that might result from any use of this document, even if the information in the document is followed properly.

This document is not part of any sales contract between Thermo Fisher Scientific Inc. and a purchaser. This document shall in no way govern or modify any Terms and Conditions of Sale, which Terms and Conditions of Sale shall govern all conflicting information between the two documents.

Release history: version 1.0

For Research Use Only. Not regulated for medical or veterinary diagnostic use by U.S. Federal Drug Administration or other competent authorities.

Contents

Preface

This guide describes how to use Proteome Discoverer™ 1.0 for peptide and protein mass spectrometry analyses.

Contents

Related Documentation

This guide includes information on procedures and parameters used in Proteome Discoverer. You can access this information by reading a PDF version of the *Xcalibur Proteome Discoverer User Guide* or searching the Help from within the Proteome Discoverer program using the Help menu. The User Guide (PDF) comes bundled with the application software.

❖ To open the help

- From the main Proteome Discoverer window, choose **Help > Proteome Discoverer Help**.
- If available for a specific window or view, click **Help** or press **F1** for information about setting parameters.

For more information, including upcoming application notes, visit www.thermo.com.

System Requirements

Proteome Discoverer requires a license. In addition, your system must meet these minimum requirements.

System	Requirements
Hardware	• 2 GHz processor with 2 GB RAM • CD-ROM drive • Video card and monitor capable of 1280x1024 resolution (XGA) • Screen resolution of 96 dpi (set in Windows Disaply Properties) • 75 GB or greater available on the C: drive • NTFS format
Software	• Microsoft™ Windows™ XP Professional with Service Pack 2 or Service Pack 3

Licenses

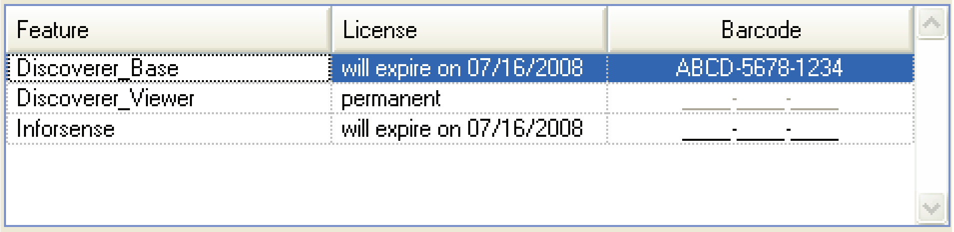
You can request a license for Proteome Discoverer as well as for your proteome computer lab, such as InforSense.

❖ To request your Proteome Discoverer license

1. Choose **Administration > Manage Licenses**.
2. In the barcode column Discoverer row, enter the number from the Proteome Discoverer jewel case.

> **Tip** The barcode is read-only for permanent licenses.

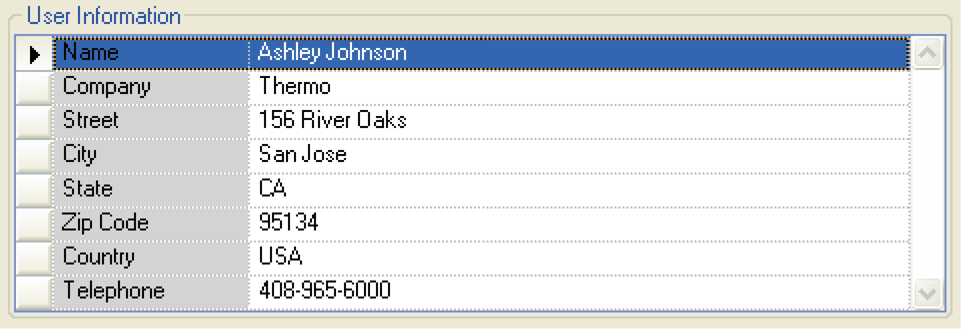
Figure 1. License Administration list of features



Feature	License	Barcode
Discoverer_Base	will expire on 07/16/2008	ABCD-5678-1234
Discoverer_Viewer	permanent	____-____-____
Inforsense	will expire on 07/16/2008	____-____-____

3. In the User Information pane, type your contact information.

Figure 2. License User Information pane



User Information

Name	Ashley Johnson
Company	Thermo
Street	156 River Oaks
City	San Jose
State	CA
Zip Code	95134
Country	USA
Telephone	408-965-6000

4. In the Host Information pane, click **Compose Email.**

Figure 3. License Host Information pane

1. Host Information

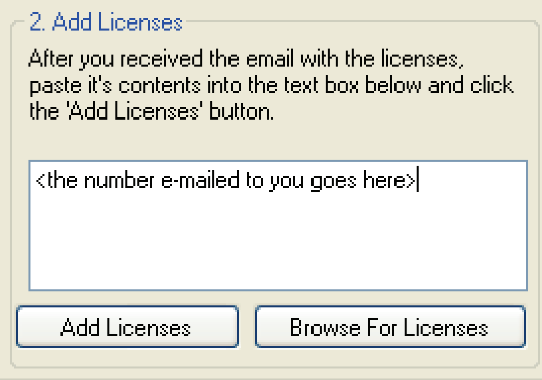
The Host Information is required to generate a license valid for your system. Send this information to licenses.ms@thermo.com.

Your default e-mail tool creates a new e-mail message, prepopulated with a message to Thermo Fisher Scientific, with a request for a Thermo Proteome Discoverer License.

5. When the license is sent back to you, copy and paste the license number into the Add Licenses box. See Figure 4.

6. Click **Add Licenses.**

Figure 4. Proteome Discoverer Add License pane



2. Add Licenses

After you received the email with the licenses, paste it's contents into the text box below and click the 'Add Licenses' button.

<the number e-mailed to you goes here>

❖ To enter other licenses, such as InforSense

1. Choose **Administration > Manage Licenses**.
2. In the Serial Number column `<application>` row, enter the `<application>` jewel case number.

Safety and Special Notices

Make sure you follow the precautionary statements presented in this guide. The safety and other special notices appear in boxes.

Safety and special notices include the following:

CAUTION Highlights hazards to humans, property, or the environment. Each CAUTION notice is accompanied by an appropriate CAUTION symbol.

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

There are several ways to contact Thermo Fisher Scientific for the information you need.

❖ To contact Technical Support

Phone	800-532-4752
Fax	561-688-8736
E-mail	us.techsupport.analyze@thermofisher.com
Knowledge base	www.thermokb.com

Find software updates and utilities to download at mssupport.thermo.com.

❖ To contact Customer Service for ordering information

Phone	800-532-4752
Fax	561-688-8731
E-mail	us.customer-support.analyze@thermofisher.com
Web site	www.thermo.com/ms

❖ To copy manuals from the Internet

Go to mssupport.thermo.com and click **Customer Manuals** in the left margin of the window.

❖ To suggest changes to documentation or to Help

- Fill out a reader survey online at www.thermo.com/lcms-techpubs.
- Send an e-mail message to the Technical Publications Editor at techpubs-lcms@thermofisher.com.

1

Getting to Know Proteome Discoverer

Proteome Discoverer™ 1.0 is a configurable software package for peptide and protein mass spectrometry analyses. It is a true end-to-end solution for workflow driven analysis. You can use this suite of applications to analyze spectral data from all Thermo Fisher Scientific and other mass spectrometers.

This chapter provides an overview of Proteome Discoverer functionality and discusses in detail its user interface.

Contents

Proteome Discoverer User Interface

In the Proteome Discoverer window, you can process data and view reports. You can customize the toolbar, and launch these tools easily from the main window.

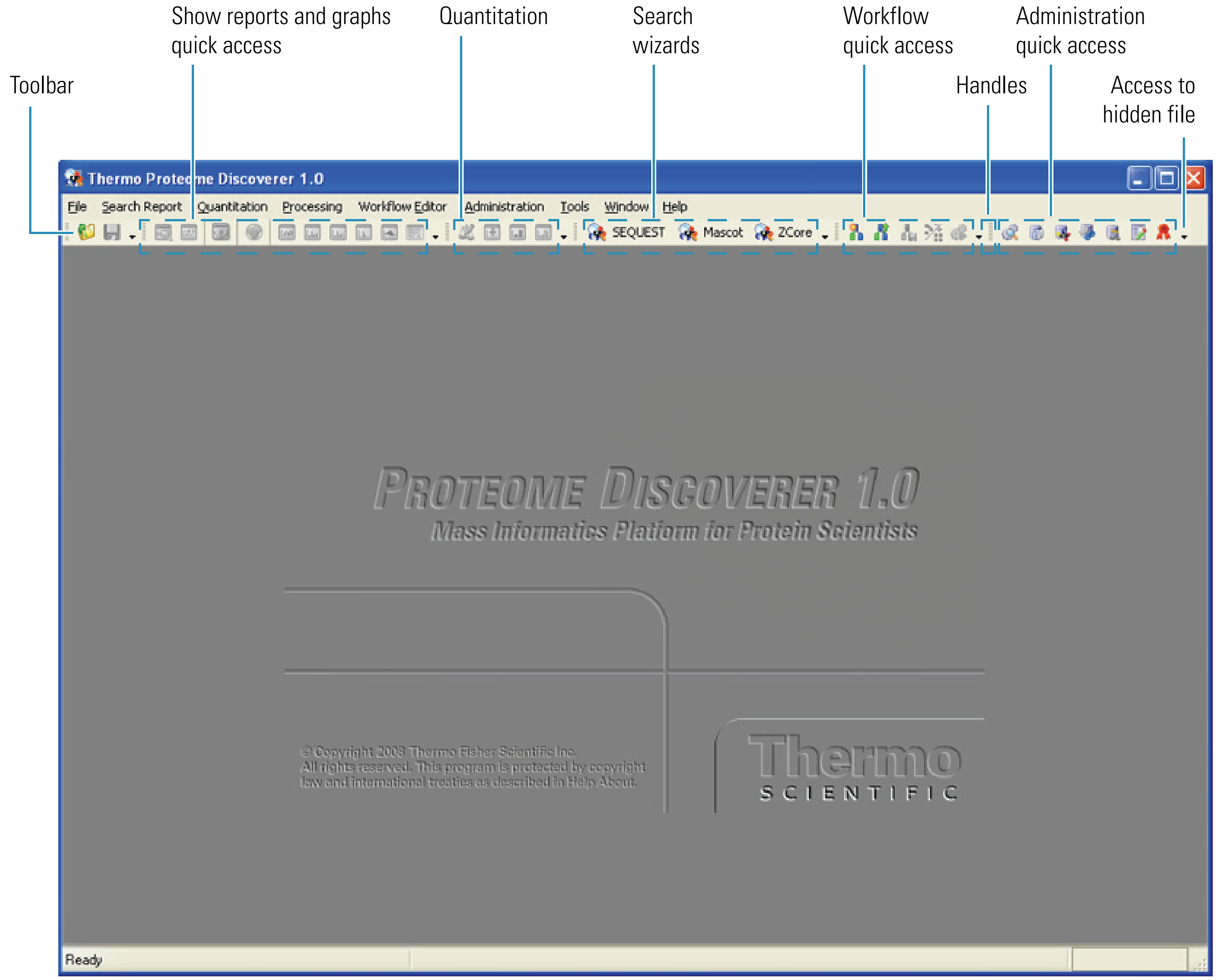
Main Window and Navigation

Figure 1, shows the toolbar options and button names. The buttons give you quick access to the following:

- Views
- Search wizards
- Administrative options and features

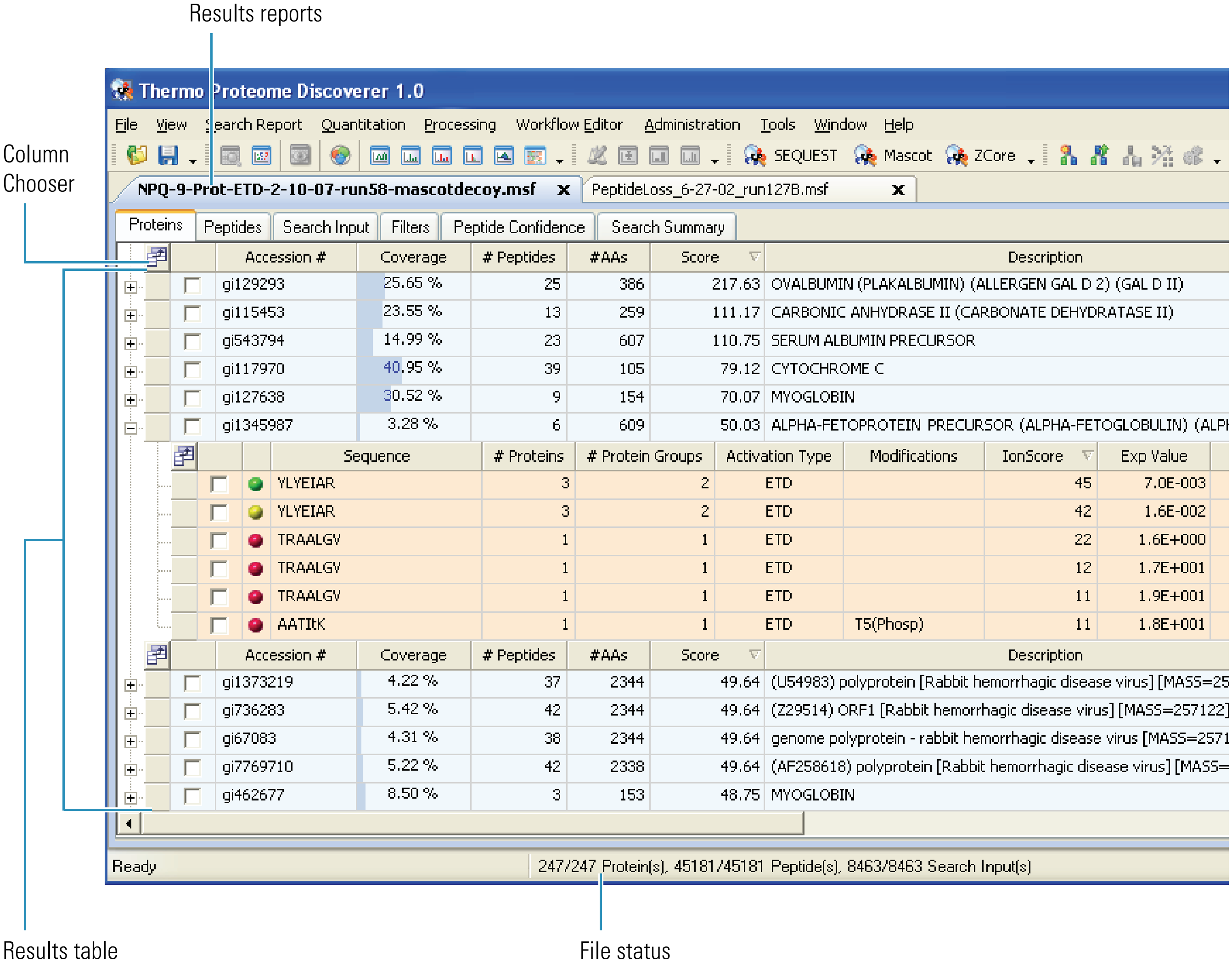
Use the toolbar handle to move a group of buttons to a different location on the toolbar. The Customize menu provides options to change the visual appearance of Proteome Discoverer.

Figure 1. Proteome Discoverer window



You can open two or more types of reports and the administration pages at the same time, as shown in Figure 2.

Figure 2. Results reports displayed in the Proteome Discoverer window



		Accession #	Coverage	# Peptides	#AAs	Score	Description
⊞	☐	gi129293	25.65 %	25	386	217.63	OVALBUMIN (PLAKALBUMIN) (ALLERGEN GAL D 2) (GAL D II)
⊞	☐	gi115453	23.55 %	13	259	111.17	CARBONIC ANHYDRASE II (CARBONATE DEHYDRATASE II)
⊞	☐	gi543794	14.99 %	23	607	110.75	SERUM ALBUMIN PRECURSOR
⊞	☐	gi117970	40.95 %	39	105	79.12	CYTOCHROME C
⊞	☐	gi127638	30.52 %	9	154	70.07	MYOGLOBIN
⊟	☐	gi1345987	3.28 %	6	609	50.03	ALPHA-FETOPROTEIN PRECURSOR (ALPHA-FETOGLOBULIN) (ALPH

		Sequence	# Proteins	# Protein Groups	Activation Type	Modifications	IonScore	Exp Value
☐	🟢	YLYEIAR	3	2	ETD		45	7.0E-003
☐	🟡	YLYEIAR	3	2	ETD		42	1.6E-002
☐	🔴	TRAALGV	1	1	ETD		22	1.6E+000
☐	🔴	TRAALGV	1	1	ETD		12	1.7E+001
☐	🔴	TRAALGV	1	1	ETD		11	1.9E+001
☐	🔴	AATItK	1	1	ETD	T5(Phosp)	11	1.8E+001

		Accession #	Coverage	# Peptides	#AAs	Score	Description
⊞	☐	gi1373219	4.22 %	37	2344	49.64	(U54983) polyprotein [Rabbit hemorrhagic disease virus] [MASS=25
⊞	☐	gi736283	5.42 %	42	2344	49.64	(Z29514) ORF1 [Rabbit hemorrhagic disease virus] [MASS=257122]
⊞	☐	gi67083	4.31 %	38	2344	49.64	genome polyprotein - rabbit hemorrhagic disease virus [MASS=2571
⊞	☐	gi7769710	5.22 %	42	2338	49.64	(AF258618) polyprotein [Rabbit hemorrhagic disease virus] [MASS=
⊞	☐	gi462677	8.50 %	3	153	48.75	MYOGLOBIN

Ready

247/247 Protein(s), 45181/45181 Peptide(s), 8463/8463 Search Input(s)

Menus

Proteome Discoverer has two broad types of toolbar menus to access tools:

- Display reports and views
- Configure database files and displays

The following are menu choices:

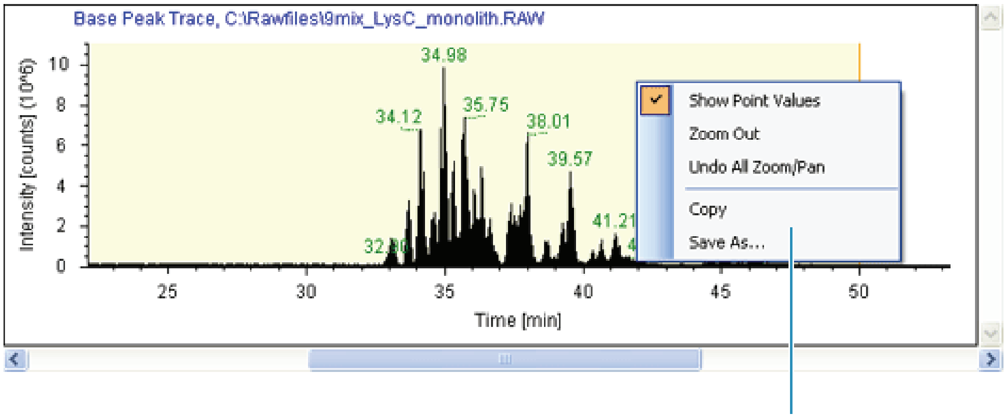
- Shortcut Menus
- File Menu
- Search Report Menu
- Search Report Menu
- Quantitation Menu
- Processing Menu
- Workflow Editor Menu
- Administration Menu
- Tools Menu
- Window Menu
- Help Menu

Shortcut Menus

Proteome Discoverer reports and views have shortcut menus with features you can use to help analyze data.

- Right-click anywhere in a view to display the shortcut menu. Figure 3 shows the View shortcut menu with commands to zoom into the chromatogram or copy it, or save the view to be included in another program.

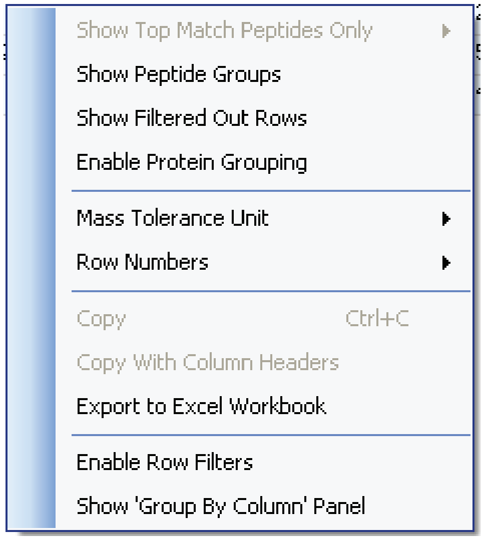
Figure 3. View shortcut menu



View shortcut menu

- Right-click anywhere in a report to display its shortcut menu as shown in Figure 4.

Figure 4. Report shortcut menu



File Menu

Use File menu commands (see Table 1) to open, close, or save your reports; connect or disconnect from a server; import a Bioworks .srf file; and close Proteome Discoverer.

Table 1. Description of File menu commands (Sheet 1 of 2)

Feature	Description	Menu access	Shortcut key access	Toolbar access
Open report	Use to upload a report.	**File > Open Report**	CTRL+O	📂
Save report	Use to save a modified report.	**File > Save Report**		💾
Close	Use to close a report without saving any modifications, such as applying filters. Proteome Discoverer remains open after the report is closed.	**File > Close**		
Import Search Results from SRF	Import search results that are in SRF format.	**File > Import Search Results from SRF**		
Connect to Server	(only if the Client_Server feature is licensed) Select a different server to use for your search process.	**File > Connect to Server**		

Table 1. Description of File menu commands (Sheet 2 of 2)

Feature	Description	Menu access	Shortcut key access	Toolbar access
Disconnect from Server	(only if the Client_Server feature is licensed) Select the server to disconnect from.	**File > Disconnect from Server**		
Exit	Close Proteome Discoverer and all opened views and reports without saving any unspecified modifications.	**File > Exit**		

Search Report Menu

Use the Search Report menu to access views and export commands.

Table 2. Description of the Search Report menu commands (Sheet 1 of 2)

Feature	Description	Menu access	Shortcut key access	Toolbar access
Show Details	When you select a protein row, the Protein Identification Details view appears. When you select a peptide row, the Peptide Identification Details view appears.	**Search Report > Show Details**	CTRL+D	
Show Distribution Chart	Assess different aspects of the search results including scores, delta masses, retention times, and such by plotting them in relationship to each other.	**Search Report > Show Distribution View**	CTRL+SHIFT+D	
Show proteins covered by this set of peptides	Explores what proteins are present and their associations through related peptides.	**Search Report > Show proteins covered by this set of peptides**	CTRL+SHIFT+H	
Show Chromatogram View	Shows the intensities of one or more masses as a function of time.	**Search Report > Show Chromatogram View**	CTRL+SHIFT+C	
Show Spectrum	Shows the MS/MS spectrum that was used for the peptide search.	**Search Report > Show Spectrum**	CTRL+SHIFT+S	

Table 2. Description of the Search Report menu commands (Sheet 2 of 2)

Feature	Description	Menu access	Shortcut key access	Toolbar access
Show Fragment Match Spectrum	Shows the MS/MS spectrum used for the search, annotated with the fragments that have been assigned within a predefined mass tolerance.	**Search Report > Show Fragment Match Spectrum**	CTRL+SHIFT+F	📊
Show Isotope Pattern	Shows a detailed perspective of the MS scan of the precursor isotope pattern of the selected precursor.	**Search Report > Show Isotope Pattern**	CTRL+SHIFT+P	📊
Show Sequence Comparison	Displays the complete sequences for different proteins associated with a single peptide.	**Search Report > Show Sequence Comparison**		▤
Show Extracted Ion Chromatogram	Shows the extracted mass chromatogram of the precursor mass of the selected peptide.	**Search Report > Show Extracted Ion Chromatogram**	CTRL+SHIFT+T	📈
Export Xcalibur Exclusion List	Exports the exclusion list, which is defined in this Search Report option.	**Search Report > Export Xcalibur Exclusion List**		
Export Spectra	Exports the spectra, which is defined in this Search Report option.	**Search Report > Export Spectra**		
ProtXML	Exports selected rows into prot format.	**Search Report > Export ProtXML**		
Layout	Provides a menu of options to save result layout.	**Search Report > Layout**		
Peptide Consensus View	A graphic view of the ion and peptide search results data.	**Search Report > Show Peptide Consensus View**		👁

Quantitation Menu

Use the Quantitation menu to access the quantitation methods.

Table 3. Description of the Quantitation menu commands

Feature	Description	Menu access	Shortcut key access	Toolbar access
Edit Quantitation Method	Edits the quantitation method of the current report.	**Quantitation > Edit Quantitation Method**		
Show peptide ratios	Shows the peptide ratios per protein view.	**Quantitation > Show Peptide Ratios**	CTRL+SHIFT+R	📊
Show Reporter Intensities	Shows the reporter intensities view.	**Quantitation > Show Reporter Intensities**	CTRL+SHIFT+N	📊
Show Quantitation Spectrum	Shows the spectrum used for quantitation.	**Quantitation > Show Quantitation Spectrum**	CTRL+SHIFT+Q	📊

Processing Menu

Use the wizards from the Processing menu to start your search process with predefined workflows. Process your .raw files and scans based on the parameters you set in the wizard. To establish your own search process, use the Workflow Editor Menu.

Workflow Editor Menu

Use the Workflow Editor to customize your search workflow, so you do not have to use the predefined search wizards.

Table 4. Description of the Workflow Editor menu commands (Sheet 1 of 2)

Feature	Description	Menu access	Shortcut key access	Toolbar access
New Workflow	Opens Workflow Editor page.	**Workflow Editor > New Workflow**		
Open From Template	Opens existing saved data analysis workflow.	**Workflow Editor > Open From Template**		
Save As Template	Saves a workflow.	**Workflow Editor > Save As Template**		

Table 4. Description of the Workflow Editor menu commands (Sheet 2 of 2)

Feature	Description	Menu access	Shortcut key access	Toolbar access
Auto Layout	Automatically adjusts and aligns the connecting arrows and nodes.	**Workflow Editor > Auto Layout**		
Start Workflow	Begins the data analysis search using your selected workflow.	**Workflow Editor > Start Workflow**		
Import Workflow from XML	Imports a data analysis workflow from data in XML format.	**Workflow Editor > Import Workflow from XML**		
Export Workflow to XML	Exports your workflow into an XML formatted file.	**Workflow Editor > Export Workflow to XML**		

Administration Menu

Use the Adminstation menu to manage processed data, methods, and job queues.

Table 5. Description of the Administration menu commands (Sheet 1 of 2)

Feature	Description	Menu access	Shortcut key access	Toolbar access
Show Job Queue	Displays the search queue and the status of current job searches.	**Administration > Show Job Queue**	CTRL+J	
Open .msf	Opens a recently completed, hightlighted report in the job queue.	**Administration > Show Job Queue > Open Report**		
Maintains FASTA Files	Adds, removes, and modifies FASTA files.	**Administration > Maintain FASTA Files**		
Maintain Chemical Modifications	Sets the chemical modifications to be used in the search process.	**Administration > Maintain Chemical Modifications**		
Maintain Cleavage Reagents	Adds, removes, and modifies the cleavage reagents.	**Administration > Maintain Cleavage Reagents**		
Maintain Quantitation Methods	Maintains and edits known quantitation methods.	**Administration > Maintain Quantitation Methods**		

Table 5. Description of the Administration menu commands (Sheet 2 of 2)

Feature	Description	Menu access	Shortcut key access	Toolbar access
Configuration	Configuration of parameters used in all searches.	**Administration > Configuration**		📝
Manage Licenses	View license status and add new licenses.	**Administration > Manage License**		🏅

Tools Menu

Table 6. Description of the Tools menu commands

Feature	Description	Menu access	Shortcut key access	Toolbar access
Open QualBrowser	Launches QualBrowser.	**Tools > Qual Browser**	CTRL+SHIFT+B	📊
Open InforSense	Launches InforSense.	**Tools > InforSense**	CTRL+SHIFT+I	📊
FASTA Database Utilities	Appends and adds to an existing FASTA file.	**Tools > FASTA Database Utilities**		

Window Menu

Use the Window menu to see all of the open windows available to view in your session of Proteome Discoverer.

Feature	Description	Menu access
Close All Windows	Closes all the open windows of your Proteome Discoverer session.	**Windows > Close All Windows**

Help Menu

Use the Help menu to access the Help and to determine the revision number of Proteome Discoverer.

Table 7. Description of the Help menu commands

Feature	Description	Menu access	Short key access	Toolbar access
About Proteome Discoverer	Displays the install version of Proteome Discoverer and its components. Shows the Thermo Fisher Scientific copyright notice.	**Help > About Proteome Discoverer**		
Proteome Discoverer Help	Opens Proteome Discoverer Help.	**Help > Proteome Discoverer Help**	F1	

Understanding Proteome Discoverer

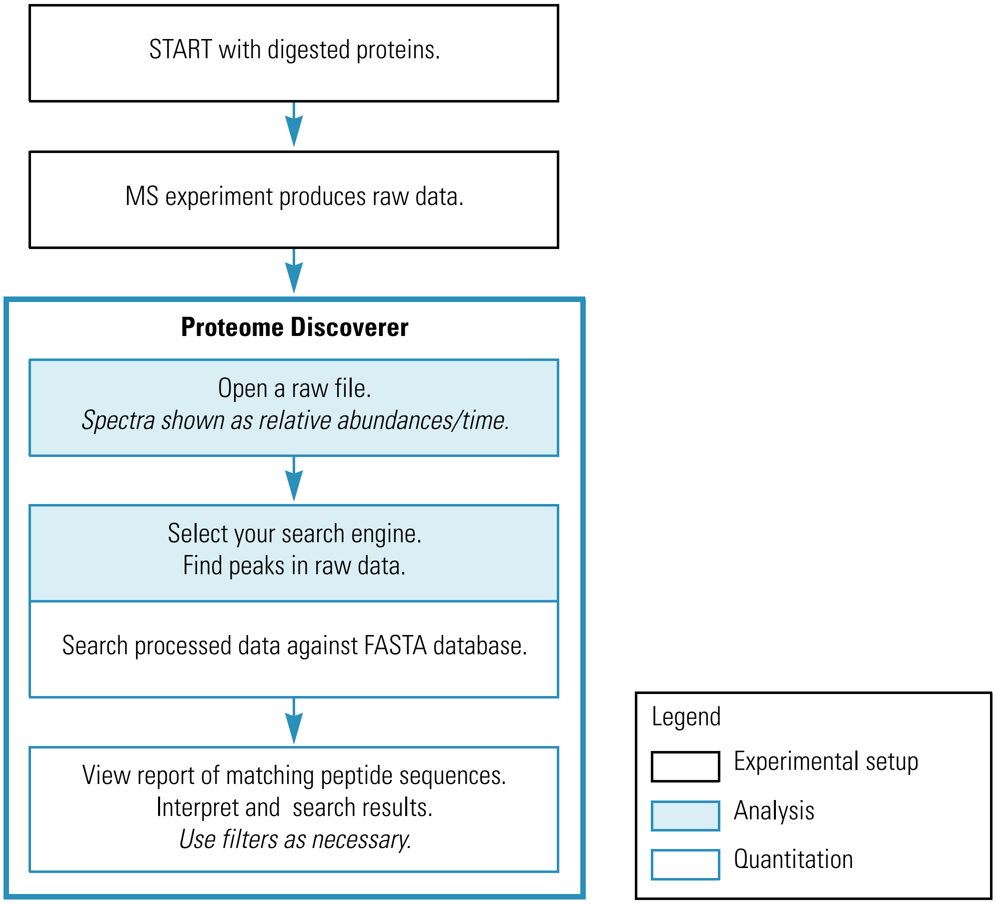
Use Proteome Discoverer to identify proteins from the mass spectra of digested fragments. The following concepts are highlights of Proteome Discoverer:

- Works with peak-finding search engines, such as SEQUEST™, Mascot™, and ZCore™, to process all data types collected from low- and high-mass accuracy MS instruments.
- Produces complementary data from a variety of dissociation methods and data-dependent stages of tandem mass spectrometry.
- Combines, filters, and annotates results from several search database engines and from multiple analysis iterations.

The peak-finding software searches the raw MS data and outputs a peak list and relative abundances. The peaks represent the peptides for a given mass and charge.

In the next step, the search engine correlates the uninterrupted tandem mass spectra of peptides with databases, such as FASTA. See "Databases and Sample Files" on page 18. Figure 5 outlines a standard workflow you can use with Proteome Discoverer.

Figure 5. Searching spectra flowchart



Proteome Discoverer Features

Proteome Discoverer includes the following features:

- Multiple search engines. See "Working with Search Wizards."
- Workflow editor for searching with multiple algorithms and merging results from multiple dissociation techniques. See "Using the Proteome Discoverer Workflow Editor."
- Database search results available from multiple raw files in a single protein/peptide report See "Understanding Reports and Views."
- Ability to export a peptide/protein report. See "Exporting Data to Other Programs."
- Integration with a local MASCOT installation.
- Ability to run tasks concurrently, such as database indexing and database searching.
- During database searching, the user interface remains responsive.

- Ability to import standard spectrum data formats, such as mzDATA, mzXml, and MGF.
- Ability to export standard spectrum data formats, such as mzDATA, dta, and MGF.
- Exports all or filtered results to the protXML format. See "Exporting Data to Other Programs."
- Merges filtered or unfiltered search results.
- Displays graphic comparisons of two or more protein sequences.
- Support for a FASTA database manager tool. See "Using FASTA Database Utilities."
- Connection to InforSense annotation workflows. See "InforSense Protein Annotation Discussion."
- Integration with GO™ annotation, through InforSense, for proteins in summary results. See "InforSense Protein Annotation Discussion."
- Support for TMT™ quantitation. See "Peptide Ratio Calculation."
- Support for iTRAQ™ quantitation. See "About iTRAQ Quantitation."
- Support for HMA ETD-Orbitrap data analysis.

Combining the traditional sensitivity of SEQUEST with the stringency of a probability-based scoring algorithm, Proteome Discoverer provides the flexibility to complete results and improve confidence in your peptide and protein matches.

Using Proteome Discoverer

With Proteome Discoverer, you can conduct data analysis searches of your spectrum using the search wizards or the Workflow Editor.

The Proteome Discoverer search wizards are predefined to quickly set your search parameters and get your results. There is one wizard for each of the supplied search engines: ZCore, SEQUEST, and Mascot.

For additional information on how to use the wizards, see "Working with Search Wizards" on page 29.

The Workflow Editor provides greater flexibility to create custom search results. Use the three-pane display to create a custom workflow. For additonal information, see "Using the Proteome Discoverer Workflow Editor" on page 49.

The next sections describe the following:

- Search Algorithms Overview
- Raw Data, Results, Reports, and Analysis
- Databases and Sample Files

Search Algorithms Overview

Proteome Discoverer has search algorithms, ZCore, SEQUEST and Mascot; each produces complementary data. ZCore and SEQUEST are peptide search engines distributed by Thermo Fisher Scientific. Mascot is a protein identification search engine created by Matrix Science.

Mascot uses mass spectrometry data to identify proteins from primary sequence databases. ZCore is specifically developed and optimized to evaluate both high-mass accuracy and low-mass accuracy ETD and ECD data. SEQUEST is capable of analyzing three types:

- Electron-transfer dissociation (ETD)
- Electron-capture dissociation (ECD)
- Collision-induced dissociation (CID)

ETD and CID ion fragmentation methods produce distinct ion fragment sets:

- ETD and ECD generate primarily C and Z ion fragments.
- CID generates primarily B and Y ion fragments.

Frequently, peptides identified by CID are not observed with ETD and vice versa so that combining results from 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 improve confidence in identifications.

Overview of Processing, Analyzing, and Interpreting Raw Data

Through user-specified settings in Proteome Discoverer, ZCore and SEQUEST algorithms can search, filter, and sort raw file data. In addition to creating reports from the analyzed data, Proteome Discoverer extracts relevant MS/MS spectra from the .raw file and determines the precursor charge state. Proteome Discoverer filters remove false positives and other non-relevant information with a variety of user-specified methods.

Note You can filter data according to false discovery rates that you define through the use of reverse FASTA databases.

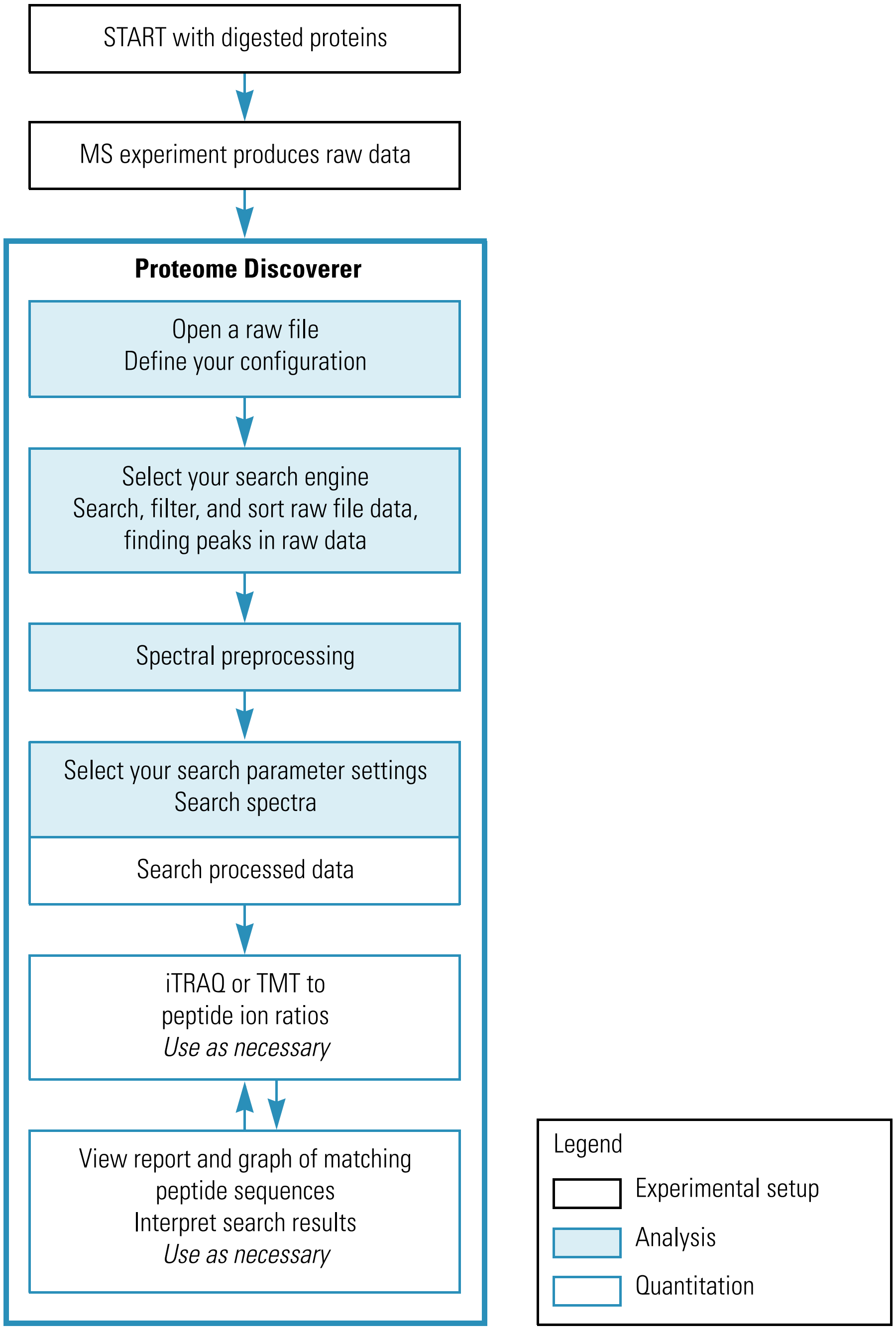
The following briefly outlines how to use Proteome Discoverer when processing, analyzing, and interpreting your MS data. See the flowchart in Figure 6 for more details.

1. Open a .raw file with Proteome Discoverer and define your configuration.
2. Select the activation and instrument types to initiate preprocessing;
   Proteome Discoverer automatically initates spectral preprocessing.
3. Select your search parameter settings and begin a search.
   Proteome Discoverer initiates a search against a FASTA database.
4. Review the generated reports and interpret the search results.

5. (Optional) Quantify the peptide ion ratios.
6. Review the data.

Figure 6 outlines a more detailed look at a standard workflow in Proteome Discoverer.

Figure 6. Using Proteome Discoverer to process, analyze, and interpret raw data



Mascot is a protein identification search engine created by Matrix Science.

Mascot Search Algorithm

Mascot uses mass spectrometry data to identify proteins from primary sequence databases. Mascot is unique in that it integrates all of the proven methods of searching. These different search methods can be categorized as follows:

- Peptide Mass Fingerprint: the experimental data are peptide mass values.
  A mass spectrum of the peptide mixture (an enzyme digested protein) provides a fingerprint of great specificity. So specific, that it is often possible to identify the protein from this information alone.
- Sequence Query: the peptide mass data is combined with amino acid sequence and composition information.
  When it is difficult determining a complete peptide sequence from an MS/MS spectrum, it can still be possible to find a series of peaks providing three or four residues of sequence data.
- MS/MS Ion Search: uninterpreted MS/MS data from one or more peptides.
  The MS/MS ions search accepts (mass and intensity pairs) peak lists. Only a single peptide may be searched, but this search mode is much more powerful when analysing an LC-MS/MS run containing data from multiple peptides. If you obtain matches to a number of peptides from a single protein you get a very high level of confidence for a correct result.

For more details on Mascot, visit www.matrixscience.com.

ZCore Search Algorithm

ZCore is specifically developed and optimized to evaluate both high-mass accuracy and nominal-mass accuracy ETD and ECD data. You can use ZCore to identify ETD and ECD MS/MS fragmentation spectra of precursor ions charge states between 2 and 7.

The ZCore search algorithm correlates experimental MS/MS spectra through comparisons to theoretical, in silico, peptide candidates derived from protein databases, even when sample sizes are limited and the signal-to-noise ratio of the spectra is low. ZCore evaluates the quality of the fragmentation spectrum and scores the spectra.

You can extract specific information from your results through the interactive data summary screens. With a click you can examine a fully annotated MS/MS spectrum or view the percent peptide coverage of an identified protein.

ZCore provides excellent search results on data acquired with Thermo Scientific ion trap mass spectrometers. Default search parameters are set in parts per million (ppm). Using accurate mass windows decreases the search time and increases the accuracy of the result, decreasing the false positive rate.

To distinguish a correct peptide match from an incorrect peptide match, the ZCore algorithm uses two calculations, a probability-based scoring of the fragment ion distribution and the total fraction of covered fragment ion intensity. The score is reported as a probability (using a -log10 value system) as well as an expectation value.

SEQUEST Search Algorithm

SEQUEST is specifically developed and optimized to evaluate both high-mass accuracy and low-mass accuracy ETD, ECD, and CID data. You can use SEQUEST in combination with automated LC/MS/MS and intelligent data acquisition tools to ensure the routine identification of low-abundance proteins in complex mixtures.

Proteome Discoverer extracts relevant MS/MS spectra from the .raw file and determines the precursor charge state and the quality of the fragmentation spectrum.

The SEQUEST search algorithm correlates experimental MS/MS spectra through comparisons to theoretical, in silico, peptide candidates derived from protein databases. The proprietary cross-correlation identification algorithm at the core of SEQUEST uses a sophisticated scoring system to help assess results. SEQUEST looks for characteristic spectral patterns and then critically evaluates the equivalence of experimental and theoretical MS/MS spectra. The identification algorithm extracts information and correctly identifies proteins even when protein samples sizes are limited and the signal-to-noise ratio of spectra is low.

You can extract specific information from your results through the interactive data summary screens. With a click you can examine a fully annotated MS/MS spectrum, or view the percent peptide coverage of an identified protein.

SEQUEST provides excellent search results on data acquired with Thermo Scientific ion trap mass spectrometers. Default search parameters are set in parts per million (ppm). Using accurate mass windows decreases the search time, increases the accuracy of the result, and decreases the false positive rate.

The Proteome Discoverer probability-based scoring system rates the relevance of the best matches found by the SEQUEST algorithm. With this probability-based scoring, Proteome Discoverer can independently rank the peptides and proteins, and increase the confidence in protein identification. Additionally, this scoring system minimizes the time needed for data interrogation or results review, increasing the overall throughput of the analysis.

You can also determine false discovery rates using reverse databases. Comparison of the results with forward and reversed databases provides an additional means of increasing confidence in protein identification.

Raw Data, Results, Reports, and Analysis

Proteome Discoverer has extensive options for exporting data and results into other file formats such as XML.

You can use reports to analyze and share your results. To create reports, use available file formats such as XML and Excel™. Define your report parameters through your Search Results view.

Databases and Sample Files

Proteome Discoverer includes the FASTA database (see page 62) and example .raw files. Use these files when exploring and learning how to use Proteome Discoverer.

Note If you have Bioworks installed, you can select the option to download the FASTA files.

The FASTA databases, supplied with the Proteome Discoverer program, are located in the Xcalibur database folder.

Proteome Discoverer Input and Output File Types

Proteome Discoverer accepts and creates the file types listed in Table 8.

Table 8. File types

File extension	File name definition	Description	File type
.dta	Mass spectral files produced during SEQUEST or ZCore analysis	Contains MS*n* data for single or grouped scans.	Output
.raw	Raw data	Raw data collected from an instrument.	Input
.msf	Mass spectrum search files	Contains relevant information from three separate sources (DTA file data, OUT file data, and setup information) - created by Proteome Discoverer.	Input and output
.out	Internal data files	Temporary output data file.	Output and input
.srf	Unified search files	Contains relevant information from three separate sources - not used by Proteome Discoverer.	Input only

Proteome Discoverer Workflow

In the Node Selection pane are five categories of different workflow choices. A typical workflow uses two or more options from these categories shown in Table 9. To start a new workflow, use a node from the Data Input category. See "Discoverer Workflow Nodes Details" on page 57 for detailed descriptions.

Table 9. Workflow nodes (Sheet 1 of 2)

Workflow category	Workflow node	Notes
Data Input:		
	Rawfile Selector	
	Spectrum Selector	
	SRF File Selector	.srf files are for input only.
Spectrum Processing:		
	ETD Spectrum Charger	
	Noise Peak Filter	
	Non-Fragment Filter	Removes peaks from an MS/MS spectrum that are not likely to be fragment peaks. Restricted to precursor ion peaks present in ETD and ECD spectra.
	Spectrum Grouper	Use group spectra from MS2 and MS3 scans together. Set **Allow MS Order Mismatch** to true; the default is false.
	Spectrum Normalizer	
	Xtract	Xtract only works with high-resolution data.
Spectrum Filters:		
	Scan Event Filter	
	Spectrum Properties Filter	
	Spectrum Score Filter	
Peptide/Protein ID:		
	Mascot	
	SEQUEST	

Table 9. Workflow nodes (Sheet 2 of 2)

Workflow category	Workflow node	Notes
	ZCore	ZCore can only process ETD spectra.
Quantitation:		
	Reporter Ions Quantitizer	
Data Export:		
	Spectrum Exporter	

For more information, see "Using the Proteome Discoverer Workflow Editor" on page 49.

Quantitation Overview

The quantitation section discusses iTRAQ and isobaric tandem mass tag (TMT) quantitation methods. The menu command, Quantitation > Edit Quantitation Method, is reserved only for reports analyzed by a Workflow Editor template that uses the Reporter Ions Quantitizer node.

CAUTION To access the quantitation menu option, you must first run a workflow template that uses the Reporter Ions Quantitizer node

See "How to Quantify Your Data Using Adminstration and Workflow" on page 67.

About iTRAQ Quantitation

iTRAQ™ is a protein quantitation technique that uses four or eight isobaric aminespecific tags. In single MS mode the differentially labeled versions of a peptide are indistinguishable. In tandem MS mode, where peptides are isolated and fragmented, each tag generates a unique reporter ion. You can derive protein quantitation by comparing the intensities of the four reporter ions in the MS/MS spectra.

Peptide Ratio Calculation

Protein ratios are the median of the peptides of the protein. If you want to recalculate, you must clear the Show Top Match Peptides only option so that all peptides are displayed. By default only unique peptides are considered in the calculation so that only peptides that have no other protein references are considered.

The menu command, Quantitation > Show Peptide Ratios, is reserved only for reports analyzed by a Workflow Editor template with the Reporter Ions Quantitizer node, Quant (see Figure 7).

Show Reporter Intensities

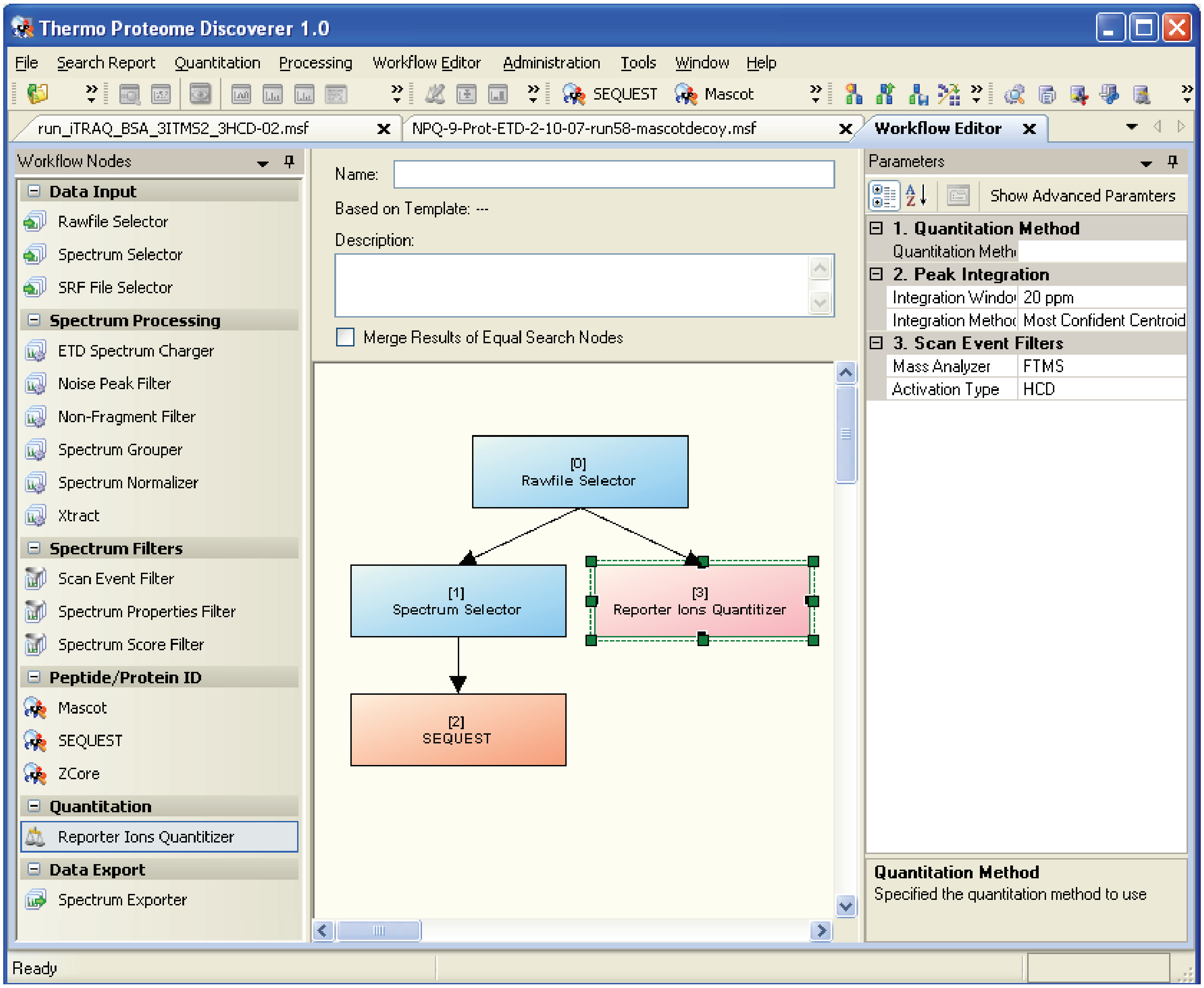
Protein ratios are the median of the peptides of the protein. If you want to recalculate, you must clear the Show Top Match Peptides only option so that all peptides are displayed. By default only unique peptides are considered in the calculation so that only peptides that have no other protein references are considered.

The menu command, Quantitation > Show Reporter Intensities, is reserved only for reports analyzed by a Workflow Editor template with the Reporter Ions Quantitizer node, Quant (see Figure 7).

Quantitation Spectrum

The menu command, Quantitation > Show Quantitation Spectrum, is reserved only for reports analyzed by a Workflow Editor template with the Reporter Ions Quantitizer node, Quant (see Figure 7).

Figure 7. Example of Workflow Editor for reporter ions experiment and quantitation options post search



☐ Merge Results of Equal Search Nodes

1. Quantitation Method
Quantitation Meth

2. Peak Integration
Integration Windo: 20 ppm
Integration Methoc: Most Confident Centroid

3. Scan Event Filters
Mass Analyzer: FTMS
Activation Type: HCD

Quantitation Method
Specified the quantitation method to use

Qual Browser Overview

With Qual Browser you can 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. When you choose the Tools > Qual Browser command, Proteome Discoverer passes the currently active raw file for Qual Browser operations. See "Using Qual Browser" on page 72 for more information.

InforSense Protein Annotation Discussion

InforSense Protein Annotation is part of the Proteome Discoverer toolset, which automatically determines the biological context of identified peptides.

Annotations are automatically retreived from a public database, such as the National Center for Biotechnology Information (NCBI) (www.ncbi.nlm.nih.gov/sites/entrez) or the Swiss Institute of Bioinformatics ExPASy proteomics server (www.expasy.com), for each protein identification in the results table. Use InforSense Protein Annotation to obtain predicted and known post-translational modifications, amino acid sequences, and biological functions.

Note When you are in InforSense Protein Annotation application, Proteome Discoverer exports your data into formatted XML files.

See Figure 8, Predefined Thermo Scientific InforSense workflows types, for the different predefined protein identfication workflows.

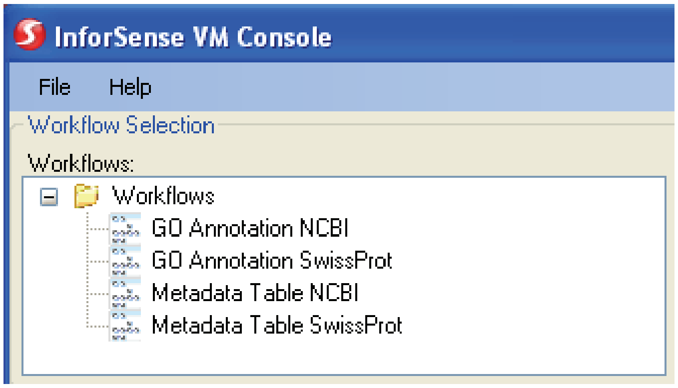
InforSense Workflow Types

Currently two types of information retrieval are available:

- Gene ontology (GO) annotations for your search results
- Metadata for your search results

Each of these has two different options, resulting in four predefined workflows for you to use on your protein MS*n* data.

Figure 8. Predefined Thermo Scientific InforSense workflows types



Introduction to GO Ontology

The Gene Ontology (GO) project 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 manner.

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 obtain more information at the GO Ontology site: www.geneontology.org/.

Introduction to Metadata Retrieval

Two Metadata workflows are available through Proteome Discoverer: Metadata Table NCBI (NCBI Web service) and Metadata Table SwissProt (ExPASy Web server). As noted earlier, each takes an XML output from Proteome Discoverer, extracts the protein accession numbers (translates them if necessary), and submits the corresponding accession numbers to the appropriate Web service. The workflow retrieves descriptive data for each accession submitted and then parses the data into different categories, including the sequence positions for known and potential post-translational modifications and metal binding sites.

Using Proteome Discoverer

Proteome Discoverer offers a variety of tools and features to assist you with spectrum analysis. This chapter discusses options that are most often used in Proteome Discoverer and describes methods of analysis.

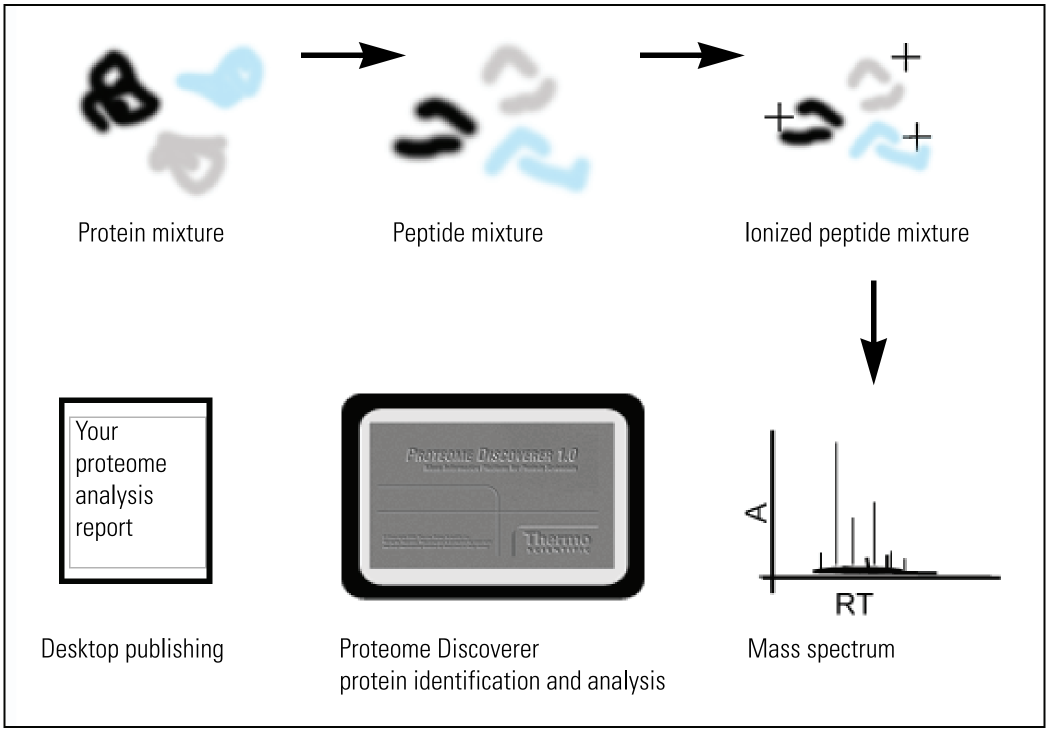
The quickest way to get started with Proteome Discoverer is to define your search parameters by using the search wizards.

Contents

How Proteome Discoverer Fits Into Your Lab Workflow

Figure 9 depicts a typical protein identification and characterization experimental workflow. Discoverer analyzes spectral data from all Thermo Scientific mass spectrometers.

Figure 9. Flow chart of typical protein identification



Features of Proteome Discoverer used in protein identification and analysis:

- Open .msf files: Multiple open options to evaluate your results.
- Job queue: Used to observe the spectrum search process.
- Workflow Editor: Used in an interactive way to build your search algorithm.
- Chemical modifications: Used to build and maintain the static and dynamic modifications data.
- Ability to quantify and annotate your results.

Overview of Menu Options

Proteome Discoverer features a diverse and multi-level user interface. You can choose your user paths, such as opening a report by choosing the menu command, File > Open Report (CTRL+O), or clicking the open reports icon ().

This section discusses both the options to graphically display spectra, tables lists, and also procedural methods to access the user interface features when working with raw data, results, reports, and analysis of reports.

The next sections describe the following:

- File Menu
- Working with Search Wizards
- Search Results, Reports, and Analysis

Start Proteome Discoverer

❖ To open Proteome Discoverer

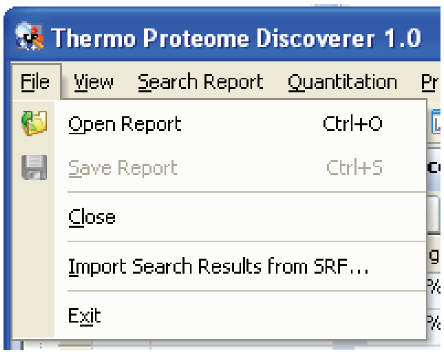
To open Proteome Discoverer from your desktop, choose **Start > Programs > Thermo Proteome Discoverer > Proteome Discoverer 1.0**.

Proteome Discoverer displays an empty startup window. From this window you can open search results, administer your session, or start a search.

File Menu

The File menu opens saved reports, saves and closes open reports, and imports search results in .srf format. Figure 10 shows details of these options.

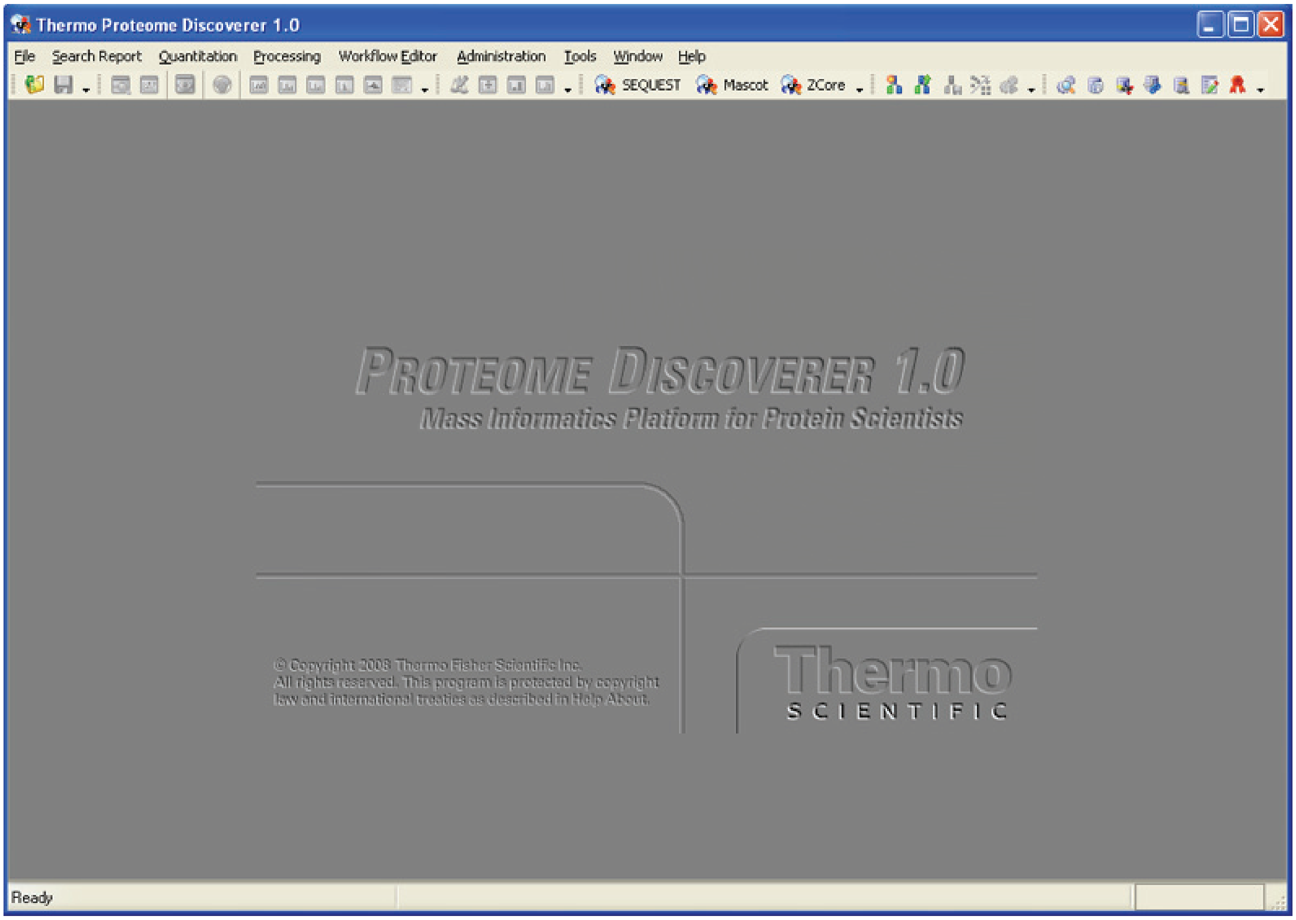
Figure 10. File menu options



❖ To open one or more reports

1. Choose **File > Open Report**.
2. Select a single file of interest or multiple files, using the CTRL key.

Figure 11. Proteome Discoverer initial view



© Copyright 2008 Thermo Fisher Scientific Inc.
All rights reserved. This program is protected by copyright law and international treaties as described in Help About.

❖ To exit and close all views and reports

1. Choose **File > Save Report** to save your changes.
2. Choose **Workflow > Save as Template** to save your workflow editor.
3. Click the data file **Filters** tab. Click Save... to save your filter settings.
4. In the Proteome Discoverer window, choose **File > Exit**.

❖ To change view size

You can turn on the feature to show large images on the tool bar.

1. Right-click anywhere on the tool bar. A shortcut menu appears.
2. Choose **Customize…**
3. Click the **Options** tab.
4. Select the desired check boxes.
5. Click **OK**.

❖ To turn on the animation option

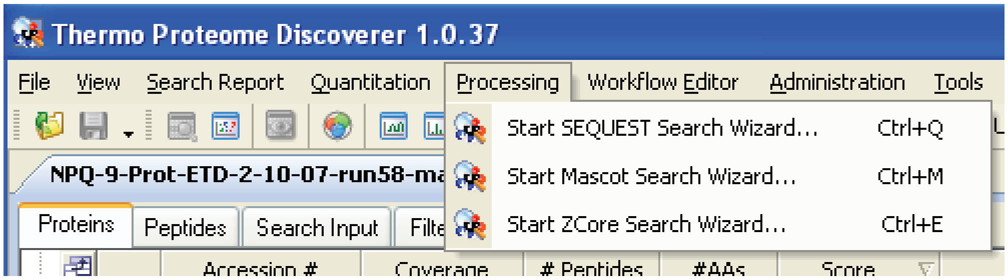
You can turn on the feature to show large images on the tool bar.

1. Right-click anywhere on the tool bar. A shortcut menu appears.
2. Choose **Customize…**
3. Click the **Options** tab.
4. Under Menu Animations, select the desired option.
5. Click **OK**.

Working with Search Wizards

You can access three wizard options from the processing menu: SEQUEST, Mascot, and ZCore. Figure 12 shows details of these options.

Figure 12. Three wizard options from the processing menu

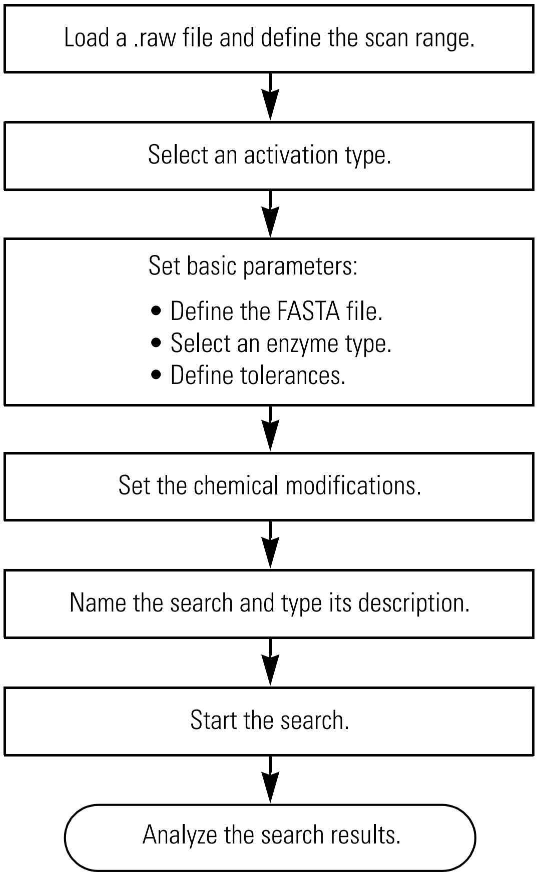


Use the Proteome Discoverer search wizards to set the search parameters, select a database and a search engine, and select the chemical modifications to be used to conduct your search. When using the search wizards, you must define four key settings:

- A .raw file
- An activation type
- A FASTA file
- A unique name for your search results

You can also set dynamic and static chemical modifications. Figure 13 defines the logical flow of the search engine wizard.

Figure 13. Proteome Discoverer search wizard process



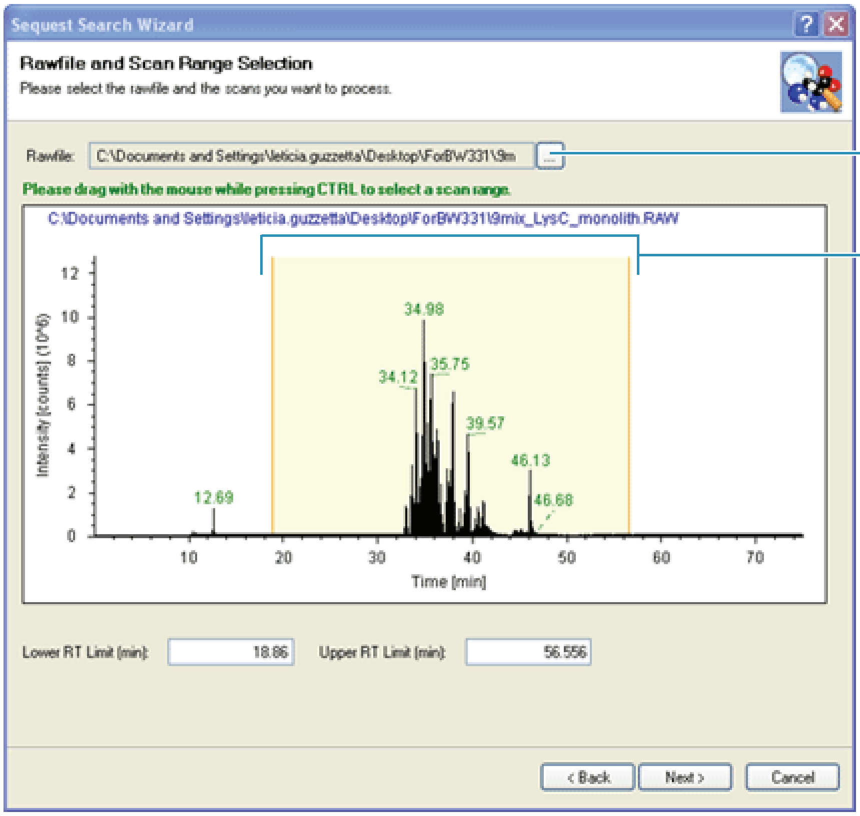
The following procedure describes how to use a search wizard, using SEQUEST as the example.

❖ To use the SEQUEST Search Wizard on a raw file

1. Choose **Processing > Start SEQUEST Search Wizard**.
   The wizard dialog box appears.

2. Click **Next**. The Rawfile and Scan Range Selection page appears. See Figure 14.

Figure 14. Rawfile and Scan Range Selection page



Sequest Search Wizard

Rawfile and Scan Range Selection

Please select the rawfile and the scans you want to process.

Rawfile: C:\Documents and Settings\leticia.guzzetta\Desktop\ForBW331\9m

Please drag with the mouse while pressing CTRL to select a scan range.

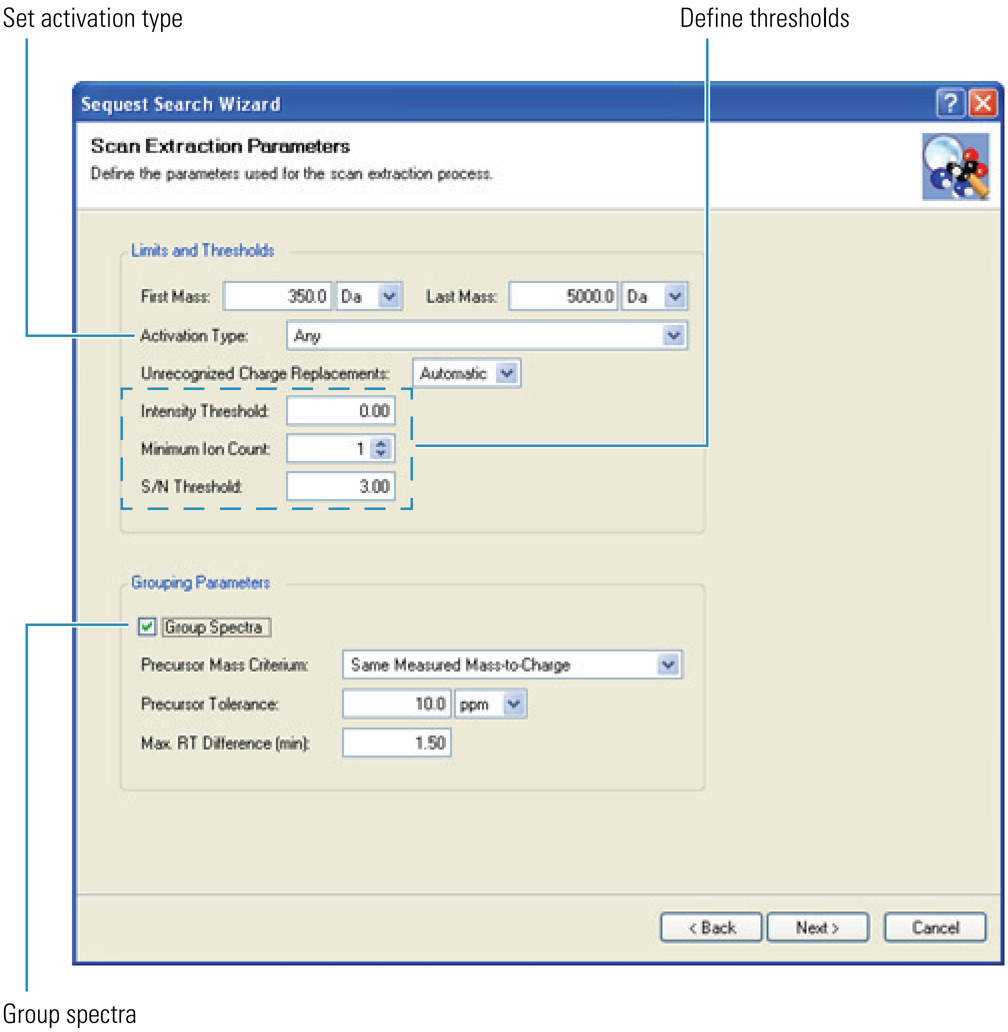
C:\Documents and Settings\leticia.guzzetta\Desktop\ForBW331\9mix_LysC_monolith.RAW

Lower RT Limit (min): 18.86

Upper RT Limit (min): 56.556

3. Set your basic search parameters:
   a. To find and select a .raw file, click the browse button (**...**).
   b. To select a scan range, press CTRL and drag your mouse over the desired range.
4. Click **Next** to move to the next page. The Scan Extraction Parameters page appears. See Figure 15.

Figure 15. Scan Extraction Parameters page



Scan Extraction Parameters

Define the parameters used for the scan extraction process.

Limits and Thresholds

First Mass: 350.0 Da
Last Mass: 5000.0 Da
Activation Type: Any
Unrecognized Charge Replacements: Automatic
Intensity Threshold: 0.00
Minimum Ion Count: 1
S/N Threshold: 3.00

Grouping Parameters

☑ Group Spectra
Precursor Mass Criterium: Same Measured Mass-to-Charge
Precursor Tolerance: 10.0 ppm
Max. RT Difference (min): 1.50

5. Set the scan extraction parameters:
   a. Enter the first and last mass limits.
   b. Set the mass units.
   c. Select an activation type.
   d. Type a value for the intensity threshold.
   e. Type a value for the minimum count.
   f. Type a value for the signal-to-noise threshold setting.
   g. (Optional) Determine your grouping parameters (seeTable 10).

Table 10. Four parameters that define the grouping criteria

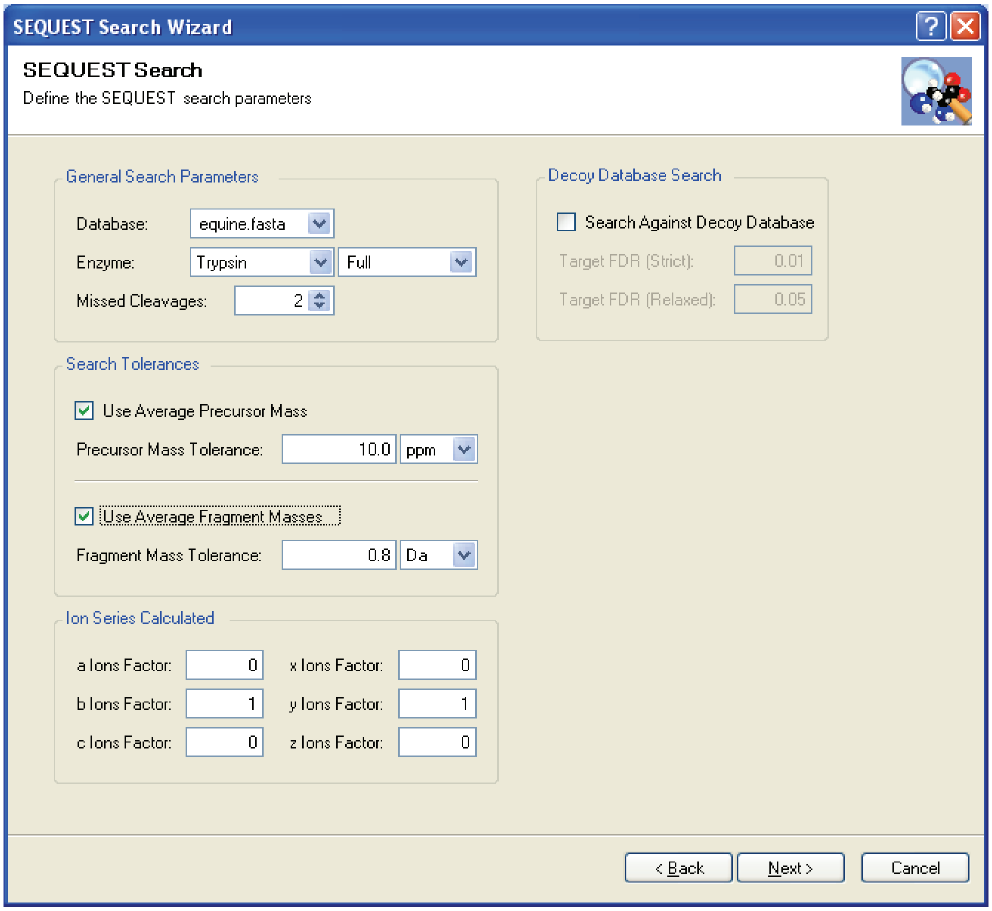
Search summary parameter	Corresponding Search Wizard parameter	Description
Allow Mass Analyzer Mismatch	Not available in the wizards.	(*Only for scans with the same precursor mas*s) When set to True, the fragment spectrum is sorted into the same group regardless of mass analyzer and activation type. Example: Both ITMS and FT-MS MS/MS scan are sorted into the same group as long as they have the same precursor mass. When set to 'False' the same two scans are added to separate groups.
Maximum Retention Time Difference	Max RT Difference	Sets the chromatographic window for the ions which are considered to be the same species, thereby adding to the same group. For example *m/z* 619 that elutes at 37.76 minutes is different from *m/z* 619 that elutes at 47.10 minutes

Table 10. Four parameters that define the grouping criteria

Search summary parameter	Corresponding Search Wizard parameter	Description
Precursor Mass Criterion	Precursor Mass Criterion	• Same Measured Mass-to-Charge: For the chromatographic peak at 37.76 minutes, only MS/MS fragment spectra that have *m/z* 619 as the precursor mass are grouped (similar for *m/z* 1236). • Same Singly Charged Mass: For the chromatographic peak at 37.76 minutes, MS/MS fragment spectra that have 619 or 1236 are grouped because both ions have the same singly charged mass. The charge associated with the combined peak list is the highest charge of the precursors in the group.
Precursor Mass Tolerance	Precursor Mass Tolerance	Sets the mass window where precursor ions are considered to be the same species, thereby adding to the same group.

6. Click **Next**. The SEQUEST Search parameters page appears.

Figure 16. SEQUEST Search parameters page



SEQUEST Search
Define the SEQUEST search parameters

General Search Parameters
Database: equine.fasta
Enzyme: Trypsin Full
Missed Cleavages: 2

Search Tolerances
☑ Use Average Precursor Mass
Precursor Mass Tolerance: 10.0 ppm
☑ Use Average Fragment Masses
Fragment Mass Tolerance: 0.8 Da

Ion Series Calculated
a Ions Factor: 0
b Ions Factor: 1
c Ions Factor: 0
x Ions Factor: 0
y Ions Factor: 1
z Ions Factor: 0

Decoy Database Search
☐ Search Against Decoy Database
Target FDR (Strict): 0.01
Target FDR (Relaxed): 0.05

7. Set the SEQUEST search parameters:
   a. Select a FASTA database.

   > **Note** The available .fasta files are registered and available through this Proteome Discoverer application. See "Using FASTA Database Administration" on page 144.

   b. Define your enzyme type.
   c. Define missed cleavages.
8. (Optional) Select the Decoy Database Search option and FDR parameters.

   > **Note** The decoy database option is neccesary if you want to see peptide confidence.

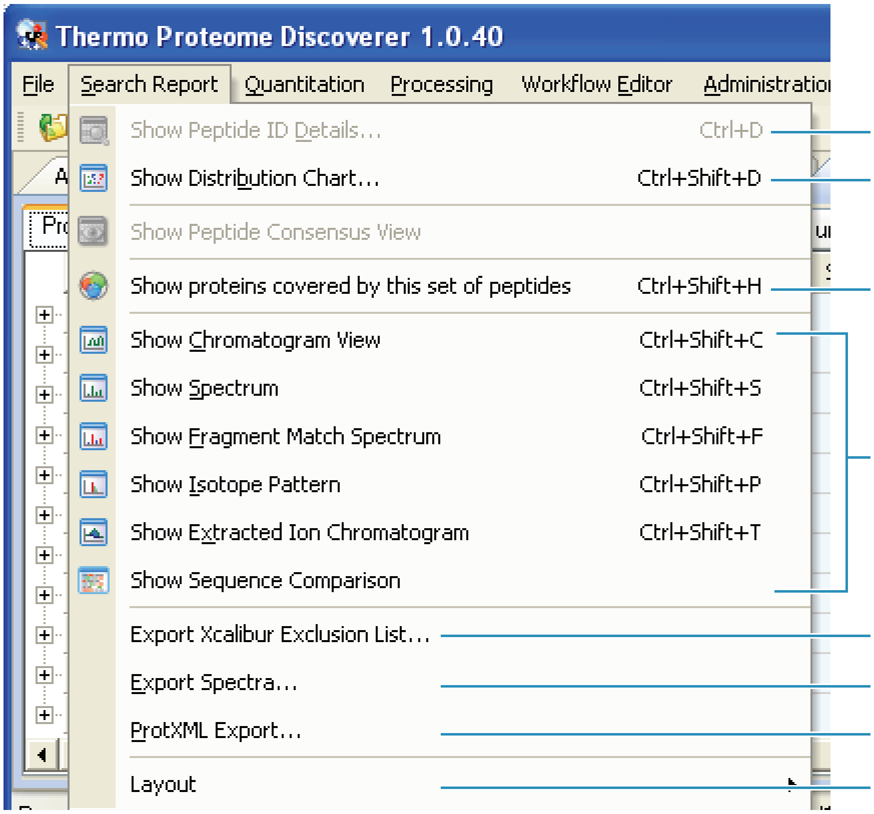
9. Set the precursor mass search tolerance parameters:
   a. (Optional) Select the Use Average Precursor Mass option.
   b. Define the precursor mass tolerance value.
   c. Choose the precursor mass units (Da, mmu, ppm).

10. Set the Ions Series, Calculated parameters as necessary for your experiment type.
11. Click Next. The Modifications page appears.
12. Analyze your search results:
    a. On the Job Queue page, click **Open Report** and select your report. Or choose **File > Open Report** to display your search results.
    b. Filter and sort your results. See "Working with Filters" on page 84.
    c. Use different views to aid in your analysis. See "Working with the Initial Results Report" on page 78.

Search Results, Reports, and Analysis

You can access search report options from the Search Report menu. Figure 17 shows details of these options.

Figure 17. Search Report menu options



Display pop-up of fragment ion detail.

Distribution plot of peptide data

Gives all proteins identified that contain the selected peptide.

Gives pop-up of indicated content.

Export user-defined mass exclusion list.

Export selected MSn spectral.

Export selected ProtXML data.

Gives pop-up for peptide data comparison.

Understanding Reports and Views

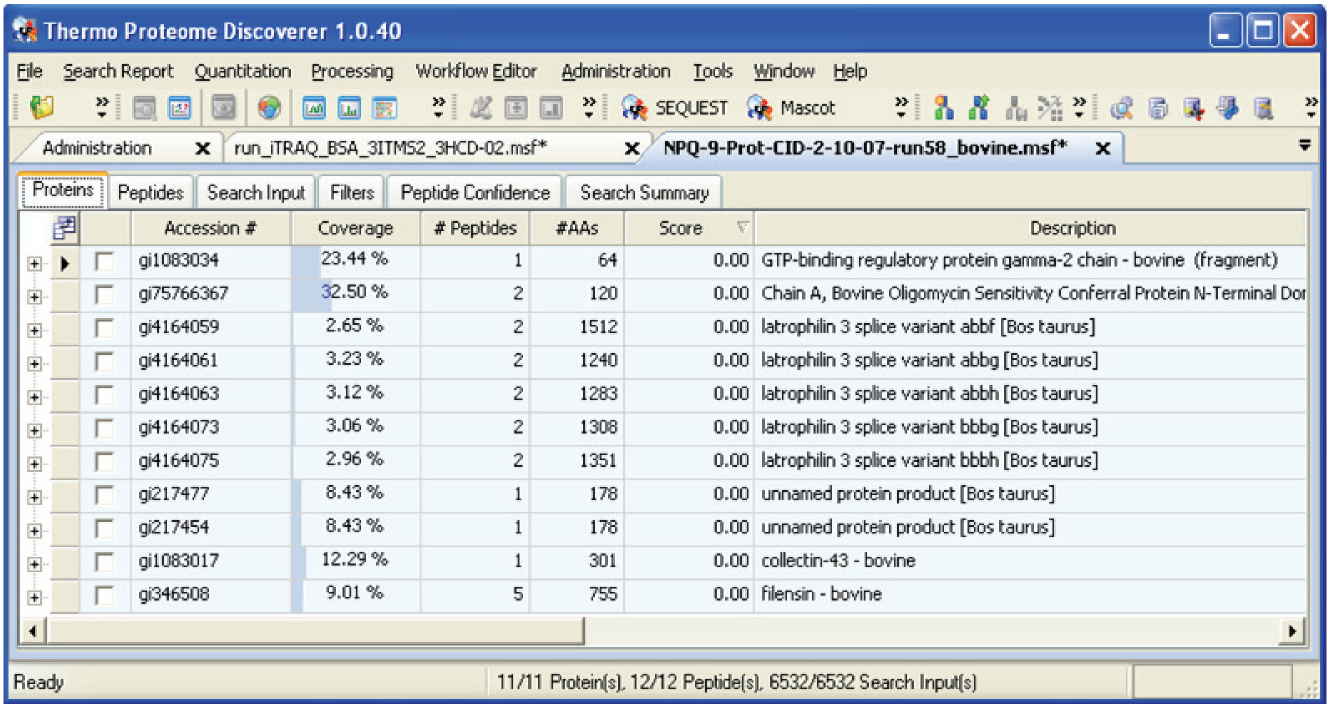
When you open a report, the Proteome Discoverer window displays your search results across multiple pages of descriptive information. Labeled tabs include Proteins, Peptides, Filters, Peptide Confidence, and a summary of the search parameters.

- Click a tab to access the corresponding page.

In the results report, you can navigate to other pages to access detailed information on proteins or peptides. You can also set different types of filtering to sort through your search results.

- Click a row or a cell to access graphical views of the corresponding protein or peptide.
- Click a protein row to access a chromatogram view of the protein or the protein coverage.

Figure 18. An example of the Results Report



Accession #	Coverage	# Peptides	#AAs	Score	Description
gi1083034	23.44 %	1	64	0.00	GTP-binding regulatory protein gamma-2 chain - bovine (fragment)
gi75766367	32.50 %	2	120	0.00	Chain A, Bovine Oligomycin Sensitivity Conferral Protein N-Terminal Dor
gi4164059	2.65 %	2	1512	0.00	latrophilin 3 splice variant abbf [Bos taurus]
gi4164061	3.23 %	2	1240	0.00	latrophilin 3 splice variant abbg [Bos taurus]
gi4164063	3.12 %	2	1283	0.00	latrophilin 3 splice variant abbh [Bos taurus]
gi4164073	3.06 %	2	1308	0.00	latrophilin 3 splice variant bbbg [Bos taurus]
gi4164075	2.96 %	2	1351	0.00	latrophilin 3 splice variant bbbh [Bos taurus]
gi217477	8.43 %	1	178	0.00	unnamed protein product [Bos taurus]
gi217454	8.43 %	1	178	0.00	unnamed protein product [Bos taurus]
gi1083017	12.29 %	1	301	0.00	collectin-43 - bovine
gi346508	9.01 %	5	755	0.00	filensin - bovine

Ready
11/11 Protein(s), 12/12 Peptide(s), 6532/6532 Search Input(s)

Search Summary Page

The Search Summary page is part of the results report. It provides a text version of the parameters that were set to perform a search. The summary page is divided into three categories that contain information about the .raw file, basic spectrum parameters, and search application-specific parameters. The categories represent the nodes used during the various search phases.

- The RawfileSelector section provides the name of the .raw file and the location of the .raw file used to conduct the analysis.
- The SpectrumSelector section contains the basic spectrum parameters. You defined each of these parameters in the Spectrum Extraction phase of the wizard when you prepared your search settings. The SpectrumSelector section contains the parameters used to select and to retrieve information for additional processing.
- The SpectrumGrouping section contains the parameters that are used to group and average the spectrum information.

IMPORTANT You defined each of these parameters in the Spectrum Extraction phase of the wizard when you were preparing your analysis setting. This section is only available if you select the Group spectra option in the wizard.

Protein Grid

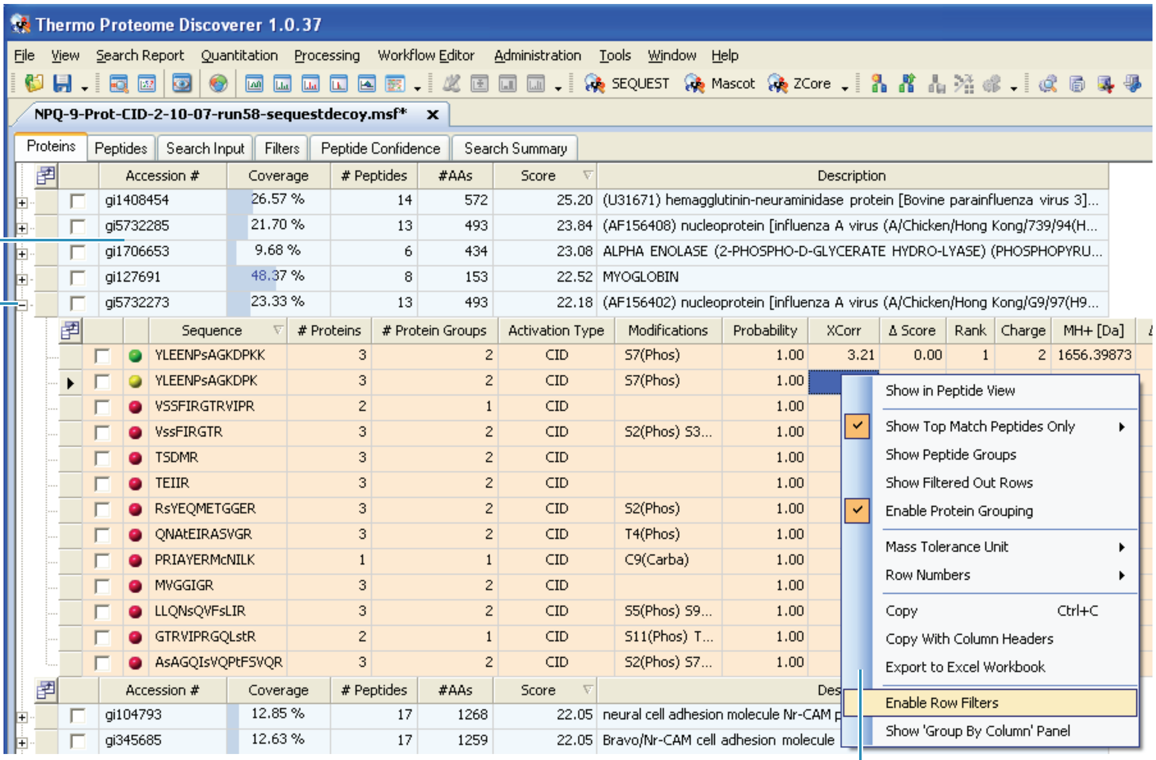
The Proteins page displays the proteins and the associated peptides identified during the search results. The Proteins page gives you detailed tabular information, a shortcut menu, and access to the peptide information. Table 11 on page 41 contains descriptions of the available columns in the protein table. To add columns to or remove columns from the table, use the shortcut menu. See "To view the protein grid" on page page 60.

The Proteins page is structured with different levels of access to the protein and peptide results:

- The blue-colored backgound shows rows of identified proteins with the associated statistics.
- The orange-colored background is a second-level table of peptides.
  Click **+** to open a protein row to see identified peptides in orange rows.
  - The columns in this second-level grid show detailed statistics for the peptide associated with the theoretical top-level protein.
  - Each row has a colored dot indicating a confidence level associated with the protein sequence at the top level; green for high confidence, yellow for modest confidence, and red for low confidence.

For a visual explanation of the Protein page, see Figures 19 and 20. For descriptions of the protein results, see Table 11 on page 41.

Figure 19. Protein view



	Accession #	Coverage	# Peptides	#AAs	Score	Description
☐	gi1408454	26.57 %	14	572	25.20	(U31671) hemagglutinin-neuraminidase protein [Bovine parainfluenza virus 3]...
☐	gi5732285	21.70 %	13	493	23.84	(AF156408) nucleoprotein [influenza A virus (A/Chicken/Hong Kong/739/94(H...
☐	gi1706653	9.68 %	6	434	23.08	ALPHA ENOLASE (2-PHOSPHO-D-GLYCERATE HYDRO-LYASE) (PHOSPHOPYRU...
☐	gi127691	48.37 %	8	153	22.52	MYOGLOBIN
☐	gi5732273	23.33 %	13	493	22.18	(AF156402) nucleoprotein [influenza A virus (A/Chicken/Hong Kong/G9/97(H9...

		Sequence	# Proteins	# Protein Groups	Activation Type	Modifications	Probability	XCorr	Δ Score	Rank	Charge	MH+ [Da]
☐	🟢	YLEENPsAGKDPKK	3	2	CID	S7(Phos)	1.00	3.21	0.00	1	2	1656.39873
☐	🟡	YLEENPsAGKDPK	3	2	CID	S7(Phos)	1.00					
☐	🔴	VSSFIRGTRVIPR	2	1	CID		1.00					
☐	🔴	VssFIRGTR	3	2	CID	S2(Phos) S3...	1.00					
☐	🔴	TSDMR	3	2	CID		1.00					
☐	🔴	TEIIR	3	2	CID		1.00					
☐	🔴	RsYEQMETGGER	3	2	CID	S2(Phos)	1.00					
☐	🔴	QNAtEIRASVGR	3	2	CID	T4(Phos)	1.00					
☐	🔴	PRIAYERMcNILK	1	1	CID	C9(Carba)	1.00					
☐	🔴	MVGGIGR	3	2	CID		1.00					
☐	🔴	LLQNsQVFsLIR	3	2	CID	S5(Phos) S9...	1.00					
☐	🔴	GTRVIPRGQLstR	2	1	CID	S11(Phos) T...	1.00					
☐	🔴	AsAGQIsVQPtFSVQR	3	2	CID	S2(Phos) S7...	1.00					

	Accession #	Coverage	# Peptides	#AAs	Score	Des
☐	gi104793	12.85 %	17	1268	22.05	neural cell adhesion molecule Nr-CAM p
☐	gi345685	12.63 %	17	1259	22.05	Bravo/Nr-CAM cell adhesion molecule

Double-click any cell to see coverage diagram.

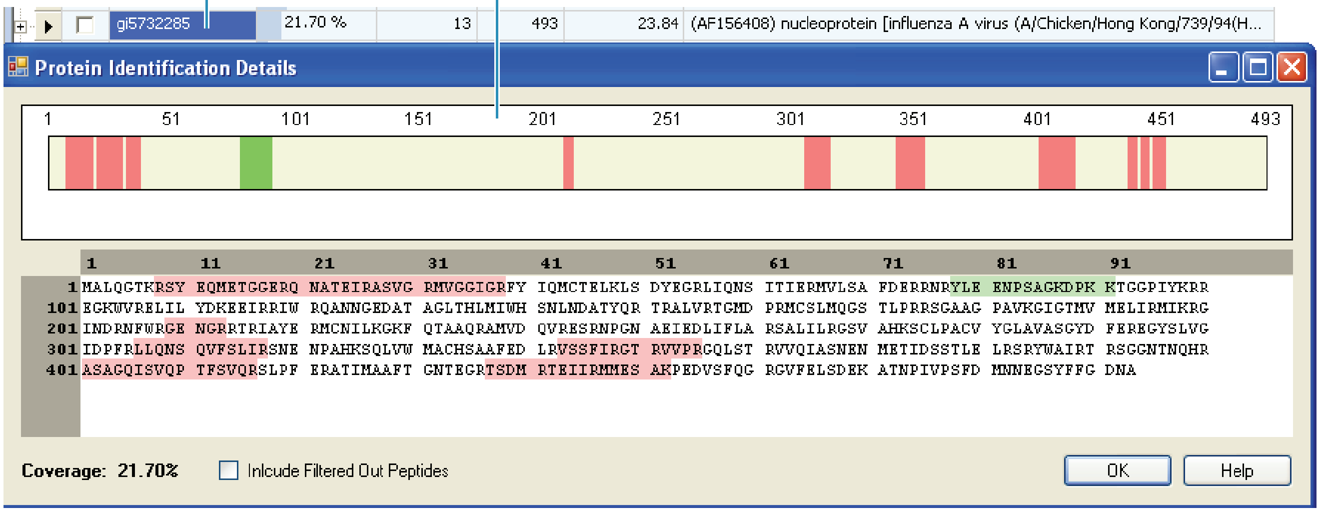
Click **+** to open the protein row.

Right-click for options.

Figure 20. Protein coverage diagram

Double-click a protein cell.

Resulting coverage diagram



	Accession #	Coverage	# Peptides	#AAs	Score	Description
☐	gi5732285	21.70 %	13	493	23.84	(AF156408) nucleoprotein [influenza A virus (A/Chicken/Hong Kong/739/94(H...

```
    1          11         21         31         41         51         61         71         81         91
  1 MALQGTKRSY EQMETGGERQ NATEIRASVG RMVGGIGRFY IQMCTELKLS DYEGRLIQNS ITIERMVLSA FDERRNRYLE ENPSAGKDPK KTGGPIYKRR
101 EGKWVRELIL YDKEEIRRIW RQANNGEDAT AGLTHLMIWH SNLNDATYQR TRALVRTGMD PRMCSLMQGS TLPRRSGAAG PAVKGIGTMV MELIRMIKRG
201 INDRNFWRGE NGRRTRIAYE RMCNILKGKF QTAAQRAMVD QVRESRNPGN AEIEDLIFLA RSALILRGSV AHKSCLPACV YGLAVASGYD FEREGYSLVG
301 IDPFRLLQNS QVFSLIRSNE NPAHKSQLVW MACHSAAFED LRVSSFIRGT RVVPRGQLST RVVQIASNEN METIDSSTLE LRSRYWAIRT RSGGNTNQHR
401 ASAGQISVQP TFSVQRSLPF ERATIMAAFT GNTEGRTSDM RTEIIRMMES AKPEDVSFQG RGVFELSDEK ATNPIVPSFD MNNEGSYFFG DNA
```

Coverage: 21.70% ☐ Inlcude Filtered Out Peptides

Table 11. Protein table

Feature	Description
Accession #	Displays the unique identifier used for the protein. Appears in table by default.
Coverage	Displays the percent coverage of the protein sequence covered by identified peptides. Appears in table by default.
# Peptides	Displays the total number of peptide matches found during the search. Appears in table by default.
#AAs	Shows the sequence length of the protein. Appears in table by default.
Score	Displays the total score of the protein. The score is the sum of the scores of the individual peptides.
Description	Provides the name of the protein excluding the Accession #. Appears in table by default.

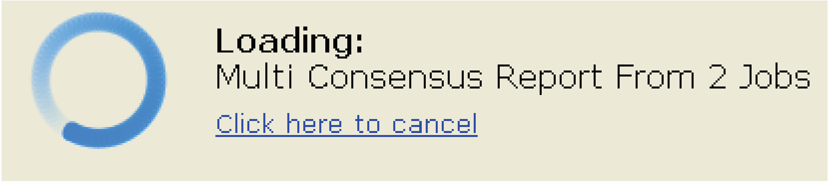
Creating a Report from Multiple Results

Use Proteome Discoverer to combine the results of multiple individual searches into one multiconsensus report. The combined report compiles information on the peptides and proteins identified in individual searches with the percent coverage combined into a unified results table, a multiconsensus report.

❖ To combine results into a report

1. From the Proteome Discoverer menu bar, choose **File > Open Report**. A Browse view appears.
2. Browse for the .msf files to upload. The files must all be in the same directory.
3. Hold down the CTRL key and select the files you need to analyze.
4. Click **OK**. A status message appears as the information is organized to display a combined report. The combined, multiconsensus report appears.

Figure 21. Multi Consensus opening message



Loading:
Multi Consensus Report From 2 Jobs
Click here to cancel

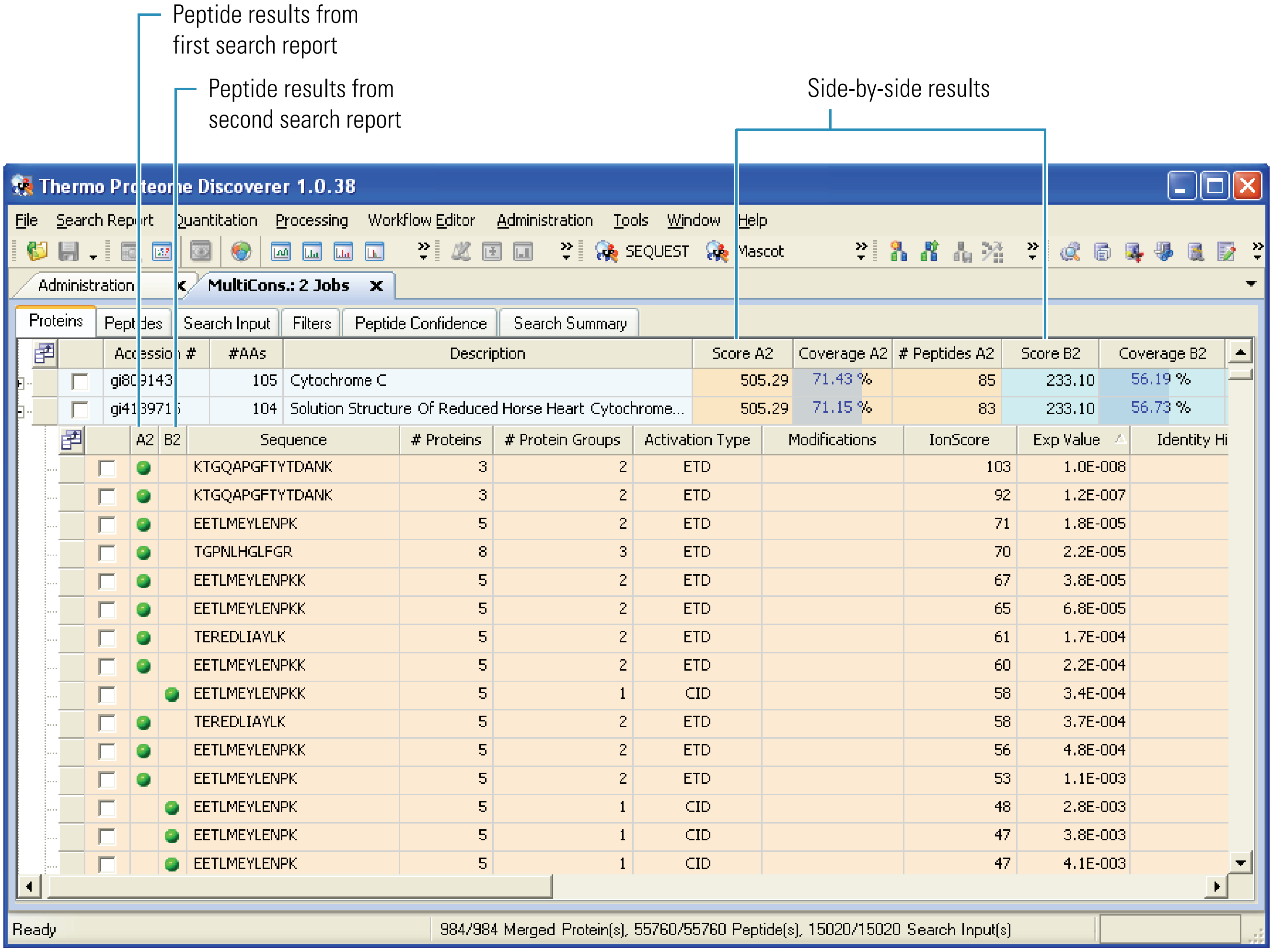
The combined results from the individual result files are numbered with alphanumeric labels. The columns and labels are defined as follows:

- A number following a letter, such as A3. The number identifies which search node in the workflow was used to create the result file.

- The column headers use labels to designate individual search results.
- Each of the column headers have tooltips that can help to match a search to the correct result column.

On the Proteins page, the individual search identification information, such as protein score, sequence coverage, and number of identified peptides, is displayed side-by-side for each protein (Figure 22). The proteins identified in the individual searches are compared by sequence to align with the results in the report. If the proteins referenced in the individual searches have different accession numbers, the accession number of the first search in the result set is chosen for display.

Figure 22. Multiconsensus report example



	Accession #	#AAs	Description	Score A2	Coverage A2	# Peptides A2	Score B2	Coverage B2
☐	gi809143	105	Cytochrome C	505.29	71.43 %	85	233.10	56.19 %
☐	gi4139715	104	Solution Structure Of Reduced Horse Heart Cytochrome...	505.29	71.15 %	83	233.10	56.73 %

	A2	B2	Sequence	# Proteins	# Protein Groups	Activation Type	Modifications	IonScore	Exp Value	Identity Hi
☐	●		KTGQAPGFTYTDANK	3	2	ETD		103	1.0E-008	
☐	●		KTGQAPGFTYTDANK	3	2	ETD		92	1.2E-007	
☐	●		EETLMEYLENPK	5	2	ETD		71	1.8E-005	
☐	●		TGPNLHGLFGR	8	3	ETD		70	2.2E-005	
☐	●		EETLMEYLENPKK	5	2	ETD		67	3.8E-005	
☐	●		EETLMEYLENPKK	5	2	ETD		65	6.8E-005	
☐	●		TEREDLIAYLK	5	2	ETD		61	1.7E-004	
☐	●		EETLMEYLENPKK	5	2	ETD		60	2.2E-004	
☐		●	EETLMEYLENPKK	5	1	CID		58	3.4E-004	
☐	●		TEREDLIAYLK	5	2	ETD		58	3.7E-004	
☐	●		EETLMEYLENPKK	5	2	ETD		56	4.8E-004	
☐	●		EETLMEYLENPK	5	2	ETD		53	1.1E-003	
☐		●	EETLMEYLENPK	5	1	CID		48	2.8E-003	
☐		●	EETLMEYLENPK	5	1	CID		47	3.8E-003	
☐		●	EETLMEYLENPK	5	1	CID		47	4.1E-003	

Ready	984/984 Merged Protein(s), 55760/55760 Peptide(s), 15020/15020 Search Input(s)

On the Peptides page, the results shown depend on the settings chosen through the shortcut menu of the results grid. The default setting shows the top hits per peptide and search engine (Figure 22). For the compiled report, the results are in titled and separate columns for the different searches.

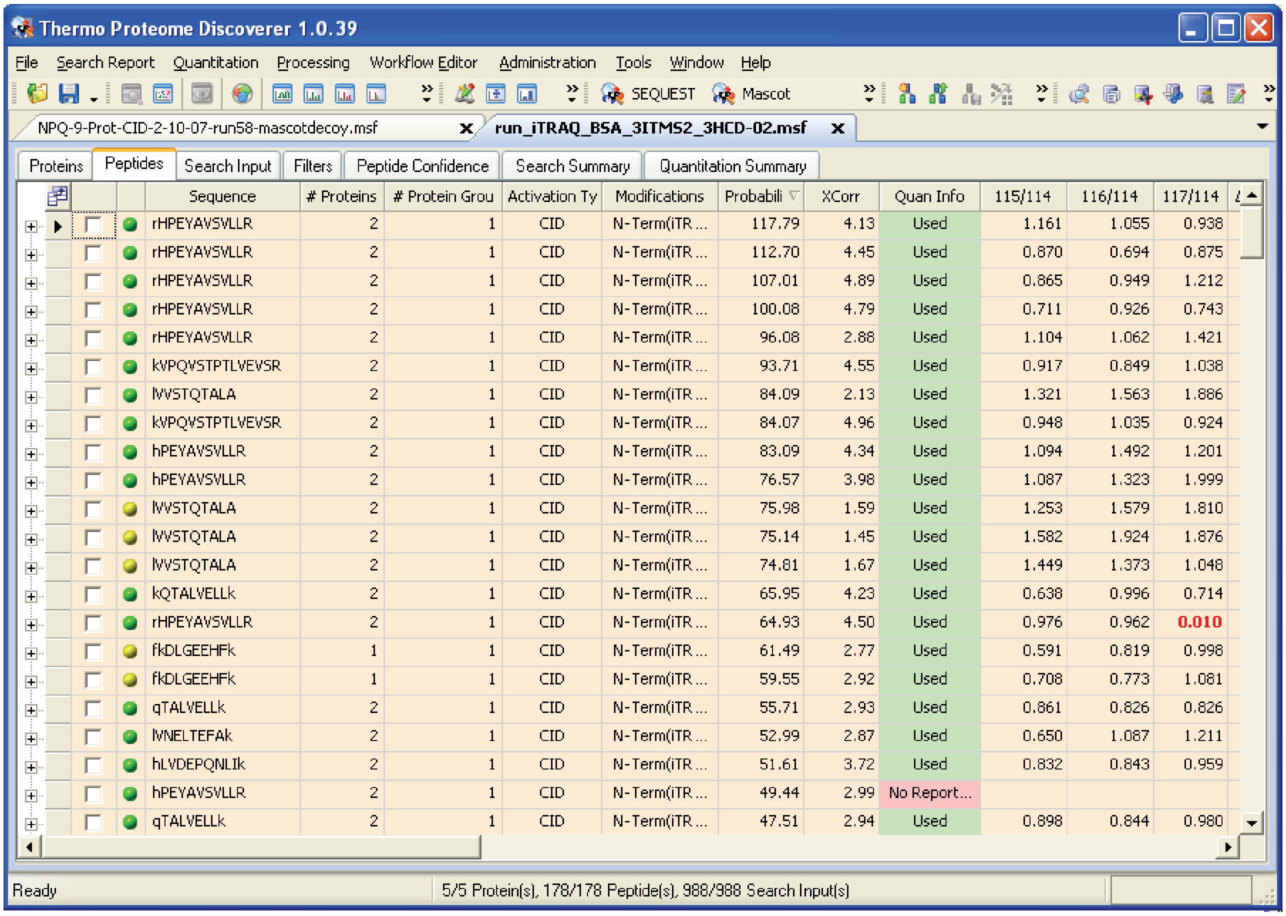
If the display options are set to show the peptide matches grouped by peptide sequence, then the identification information from the individual searches is displayed side-by-side, as shown in Figure 22. The grouped peptides do not represent actual matches found during the search but represent the unified information from all matches found for this particular peptide sequence. The peptide information can be accessed by examining the Proteins page for all peptides associated with a protein, and through the Peptides page for all peptides including those not associated with any protein.

As part of the Protein page, you can view the associated peptides. Click **+** to the right of the protein row to access the peptide rows. When a peptide is identified multiple times, only the top-match peptides are displayed in the table. Right-click to access the shortcut menu, and choose **Show Top Match Peptides Only** to display all the peptides. Table 12 describes the features of the peptide row.

Peptide Grid

The Peptides page displays the peptides and associated proteins that are identified during your search. The initial view is the list of peptides. Click **+** to the left of a peptide row to access the associated proteins. See "To view the peptide grid" on page 60.

Figure 23. Peptides view



		Sequence	# Proteins	# Protein Grou	Activation Ty	Modifications	Probabili	XCorr	Quan Info	115/114	116/114	117/114
☐	●	rHPEYAVSVLLR	2	1	CID	N-Term(iTR ...	117.79	4.13	Used	1.161	1.055	0.938
☐	●	rHPEYAVSVLLR	2	1	CID	N-Term(iTR ...	112.70	4.45	Used	0.870	0.694	0.875
☐	●	rHPEYAVSVLLR	2	1	CID	N-Term(iTR ...	107.01	4.89	Used	0.865	0.949	1.212
☐	●	rHPEYAVSVLLR	2	1	CID	N-Term(iTR ...	100.08	4.79	Used	0.711	0.926	0.743
☐	●	rHPEYAVSVLLR	2	1	CID	N-Term(iTR ...	96.08	2.88	Used	1.104	1.062	1.421
☐	●	kVPQVSTPTLVEVSR	2	1	CID	N-Term(iTR ...	93.71	4.55	Used	0.917	0.849	1.038
☐	●	lVVSTQTALA	2	1	CID	N-Term(iTR ...	84.09	2.13	Used	1.321	1.563	1.886
☐	●	kVPQVSTPTLVEVSR	2	1	CID	N-Term(iTR ...	84.07	4.96	Used	0.948	1.035	0.924
☐	●	hPEYAVSVLLR	2	1	CID	N-Term(iTR ...	83.09	4.34	Used	1.094	1.492	1.201
☐	●	hPEYAVSVLLR	2	1	CID	N-Term(iTR ...	76.57	3.98	Used	1.087	1.323	1.999
☐	●	lVVSTQTALA	2	1	CID	N-Term(iTR ...	75.98	1.59	Used	1.253	1.579	1.810
☐	●	lVVSTQTALA	2	1	CID	N-Term(iTR ...	75.14	1.45	Used	1.582	1.924	1.876
☐	●	lVVSTQTALA	2	1	CID	N-Term(iTR ...	74.81	1.67	Used	1.449	1.373	1.048
☐	●	kQTALVELLk	2	1	CID	N-Term(iTR ...	65.95	4.23	Used	0.638	0.996	0.714
☐	●	rHPEYAVSVLLR	2	1	CID	N-Term(iTR ...	64.93	4.50	Used	0.976	0.962	0.010
☐	●	fkDLGEEHFk	1	1	CID	N-Term(iTR ...	61.49	2.77	Used	0.591	0.819	0.998
☐	●	fkDLGEEHFk	1	1	CID	N-Term(iTR ...	59.55	2.92	Used	0.708	0.773	1.081
☐	●	qTALVELLk	2	1	CID	N-Term(iTR ...	55.71	2.93	Used	0.861	0.826	0.826
☐	●	lVNELTEFAk	2	1	CID	N-Term(iTR ...	52.99	2.87	Used	0.650	1.087	1.211
☐	●	hLVDEPQNLIk	2	1	CID	N-Term(iTR ...	51.61	3.72	Used	0.832	0.843	0.959
☐	●	hPEYAVSVLLR	2	1	CID	N-Term(iTR ...	49.44	2.99	No Report...			
☐	●	qTALVELLk	2	1	CID	N-Term(iTR ...	47.51	2.94	Used	0.898	0.844	0.980

Ready

5/5 Protein(s), 178/178 Peptide(s), 988/988 Search Input(s)

Table 12 describes the features of a peptide row.

Table 12. Peptide row (Sheet 1 of 2)

Feature	Description
Sequence	Displays the peptide sequence.
# Proteins	Number of proteins.
# Protein Groups	How many protein groups.
Activation Type	Displays the activation type of the spectrum where the peptide was identified.
Modifications	Displays the static and dynamic modifications identified in the peptide.
Probability	Displays the probability score for the peptide.
Score	Displays the score that the search algorithm calculated for the match.
XCorr	XCorr scores count the number of fragment ions that are common between X and Y and calculate the cross-correlation score for all candidate peptides queried from the database.
Quan Info	(Optional) Marking if quantitative analysis is available
115/114	(Optional) Related to Quan Info
116/114	(Optional) Related to Quan Info
117/114	(Optional) Related to Quan Info
Rank	Displays the rank ordering of the peptide.
Charge	Displays the charge state of the peptide.
MH+(Da)	Displays the weight differential of the peptides in Daltons.
ΔM (ppm)	Displays the difference between the theoretical mass of the peptide and the experimental mass of the precursor ion.
First Scan	Displays the first scan number where the peptide was initially identified.
Last Scan	Displays the last scan number in which this peptide match was identified.
RT (minutes)	Displays the retention time where the peptide was observed.

Table 12. Peptide row (Sheet 2 of 2)

Feature	**Description**
MSOrder	Indicates the order of the MS spectrum: • 1 is a one MS scan • 2 is an MS/MS scan (MS_2) • 3 is MS_3 The number of stages minus one of precursor ion *m/z* selection, followed by product ion detection such that MS_2 means one stage of precursor ion *m/z* selection is then followed by product ion detection
m/z	Mass-to-charge ratio
Intensity	Relative intensity
Ions Matched	Displays the number of ions found compared to the theoretical number of ions.
Annotation	Provides a location for you to insert and store notes about the search results, your analysis, and quality results.
Shortcut Menu	In the header, right-click to access the shortcut menu: • **Show Top Match Peptides Only** • **Show Peptide Groups** • **Show Filtered Out Rows** • **Enable Protein Grouping** • **Mass Tolerance Unit** • **Row Numbers Style** • **Copy, Ctrl+C** Copies the selected rows to the Clipboard. You can then paste the rows in a spreadsheet application such as Microsoft Excel. • **Copy With Column Headers** • **Export to Excel Workbook**. • **Enable Row Filters** • **Show 'Group By Column' Panel**

As part of the Peptides page, you can view the associated proteins. Click **+** to the left of the peptide row to access the protein rows. Right-click to access the shortcut menu, and choose **Show Top Match Peptides Only** to display all the peptides.

For a visual explanation of the Peptide page see Figure 23. For descriptions of the protein results, see Table 12 on page 44.

Peptide Confidence Page

Use the Peptide Confidence page to set the confidence levels for database searches (see "False Discovery Rates and Peptide Confidence Indicators" on page 97). In Proteome Discoverer, the filter settings are used to distribute the confidence indicators for the peptide matches (these are the green, yellow, and red dots attached to each peptide match). Whenever a decoy database search is performed during the database search and filter settings have been applied 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 confidence indicator, peptide matches that pass the filter associated with the relaxed FDR are assigned a yellow confidence indicator, and all other peptide matches receive a red indicator.

Note You can change the default confidence levels to alternative values within the Peptide Confidence page.

Related Topics

- To view the protein grid
- To view the peptide grid
- Peptide Confidence Page

Search Input Page

The Search Input page displays detailed information for the spectra (mass peak lists) that were created and processed during the search process. If you used the grouping parameters during your search wizard, the grouped spectra are listed in this report. You can also check to see how many peptides (# of identified peptides) were found and which proteins the peptide belongs to. From this report, you can also select spectra to create an Xcalibur exclusion/inclusion list.

Table 13. Search Input parameters (Sheet 1 of 2)

Parameters	Description
ID	Naming convention
# Identified Peptides	Displays the total number of peptide matches found during the search. Displayed in table by default.
Precursor MH+ [Da]	Displays the weight differential of the peptides in Daltons.
Precursor Charge	The charge for each peptide.
First Scan	Displays the first scan number where the peptide was initially identified.
Last Scan	Displays the last scan number where this peptide match was identified.
Master Scan(s)	Full scan in which this precursor has been selected.

Table 13. Search Input parameters (Sheet 2 of 2)

Parameters	Description
Scan(s)	Displays all scan numbers where this peptide match was identified.
RT (minutes)	Displays the retention time where the peptide was observed.
Activation Type	Displays the activation type of the spectrum where the peptide was identified.
Mass Analyzer	Displays the mass analyzer used to create the .raw file data. Proteome Discoverer recognizes the following mass analyzers: • ITMS (Ion Trap) • FTMS (Fourier Transform) • TOFMS (Time of Flight) • SQMS (Single Quad) • TQMS (Triple Quad) • SectorMS (Sector Field)

Filters Page

Use the Filters page to refine your search results and make your analysis quicker. By using filters, you can sort and filter your results by charge state, modifications, or even peptide probability. You can also create and apply more than one filter to your search results.

The protein and peptide filters have two sets of menu choices shown in Figure 24; the Filters page appears (see Figure 25).

Figure 24. Choices for protein and peptide filters

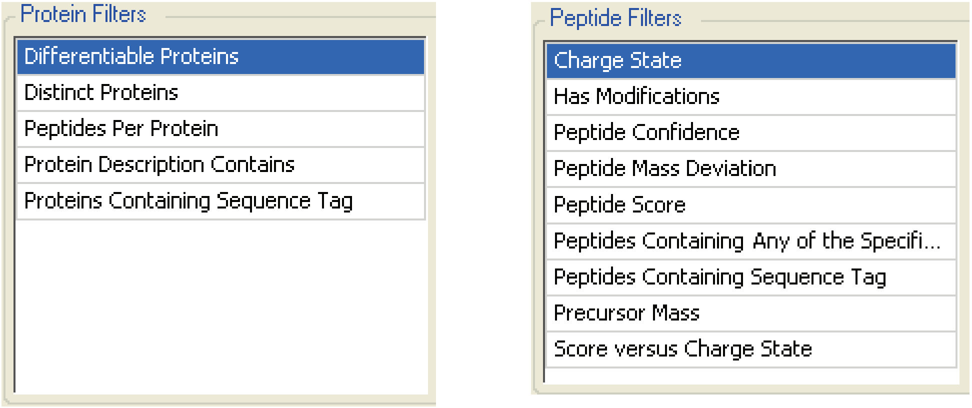
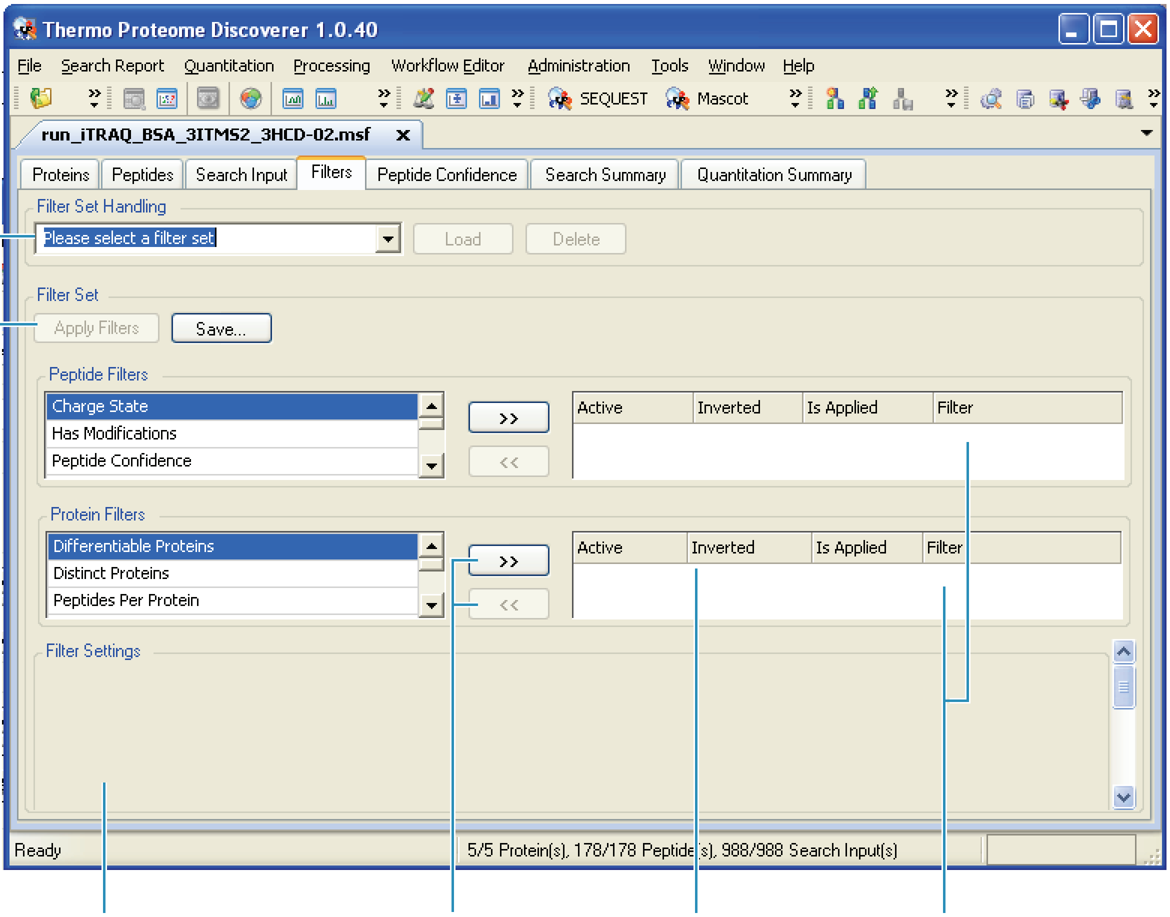


Figure 25. Filters page



Ready

5/5 Protein(s), 178/178 Peptide(s), 988/988 Search Input(s)

Saved filters list

Applies filters to search results

Filter settings status

Add and remove buttons

Invert the filter - good for disregarding a single property

Selected filter options and current status

Note For an inverted filter example, set the peptide confidence level to low confidence and select the inverted option so that only the modest- and high-confidence levels show in the report.

The different columns of the filter table provide specific information about the current status of the filter settings. See Figure 25 and Table 14.

Table 14. Filter options (Sheet 1 of 2)

Option	Description
Active	When the check box is selected, the filter is added to the filter set. You must also use Apply Filter to apply the filter setting to your current results. When the check box is not selected, the filter is temporarily removed from the filter set. By default, a newly added filter is selected and active.

Table 14. Filter options (Sheet 2 of 2)

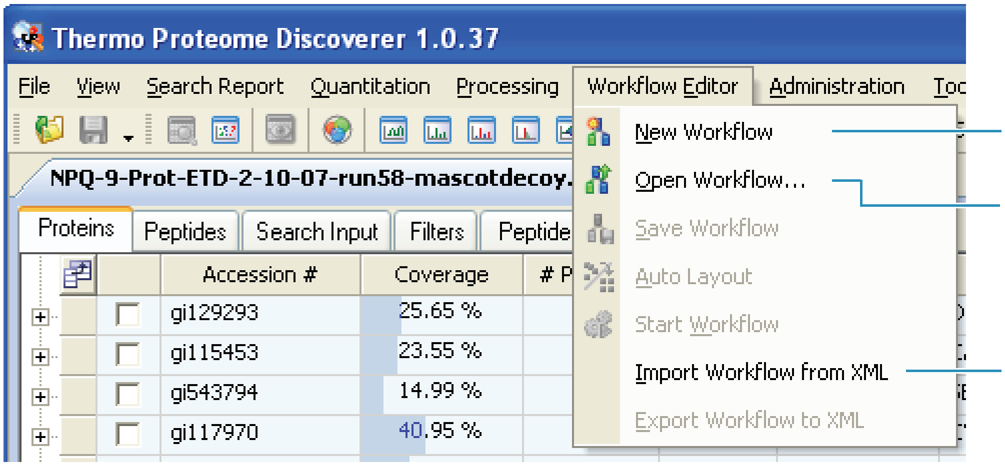
Option	**Description**
Inverted	When the check box is selected, the filter settings are inverted. All items that would normally be filtered out remain in the search results and all items that would normally remain in the result are hidden and not seen in the search results.
Is Applied	Displays if the filter has already been applied to the current search results. If the status is True, the filter was applied to the current search results.
Filter	Displays the name of the filter.
Filter Setting(s)	Provides a text summary of the status of the filter settings of parameters shown in the right pane. Modify the Filter Settings in the right pane.

Using the Proteome Discoverer Workflow Editor

- Workflow Workspace
- Creating and Saving a Search Workflow
- Discoverer Workflow Nodes Details

You can access the workflow editor options from the Search Report menu. Figure 26 shows details of these options.

Figure 26. Workflow Editor menu options



Accession #	Coverage
gi129293	25.65 %
gi115453	23.55 %
gi543794	14.99 %
gi117970	40.95 %

Design new user-defined data analysis.

Open user-defined data analysis workflow.

Import user-defined data analysis workflow.

The Proteome Discoverer Workflow Editor is a flexible and complex tool to create customized data processing workflows. Instead of using the standard wizards available through the Processing menu, you can develop a workflow specific to your needs. The workflow is the layout of processing nodes, which you then submit to process your data.

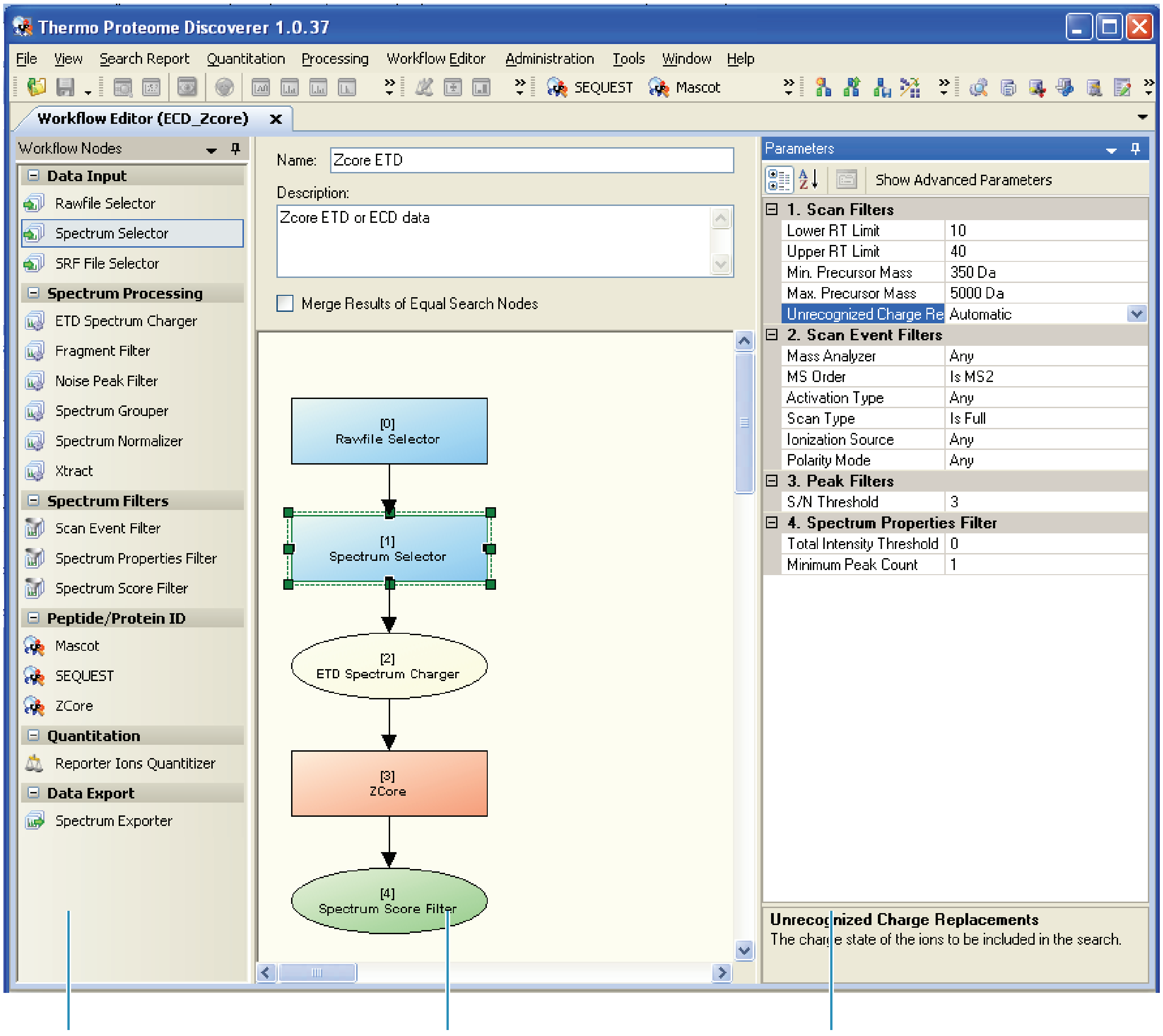
You can create a reusable processing workflow template by saving your design to load and use at another time. A unique workflow gives you the ability to set parameters that are normally static settings in the wizard or to use a function that would not normally be available such as Xtract, spectrum export, or cascading searches.

WARNING Prerequisites to using the Proteome Discoverer Workflow Editor are to first learn each "workflow node" functionality. If you do not understand the function (or interconnectivity) of these nodes, you can potentially build a sequence that creates bad results and makes no analytical sense.

Workflow Workspace

The unique three-pane layout of the Proteome Discoverer Workflow Editor provides node selections, a workspace, and parameters for each node. See Figure 27. The nodes are like building blocks that you can use to create a unique search sequence. Using the nodes, you can define your own search parameter tolerances and criteria.

Figure 27. Workflow Editor workspace definitions example



Name: Zcore ETD

Description: Zcore ETD or ECD data

☐ Merge Results of Equal Search Nodes

1. Scan Filters	
Lower RT Limit	10
Upper RT Limit	40
Min. Precursor Mass	350 Da
Max. Precursor Mass	5000 Da
Unrecognized Charge Re	Automatic
2. Scan Event Filters	
Mass Analyzer	Any
MS Order	Is MS2
Activation Type	Any
Scan Type	Is Full
Ionization Source	Any
Polarity Mode	Any
3. Peak Filters	
S/N Threshold	3
4. Spectrum Properties Filter	
Total Intensity Threshold	0
Minimum Peak Count	1

Unrecognized Charge Replacements
The charge state of the ions to be included in the search.

Node selection pane

Workspace pane

Parameter pane

Use the three-pane layout of the Proteome Discoverer Workflow Editor to do the following:

- Customize your workflow.
- Add nodes to your workflow.
- Set and define your search parameters (see Table 10).

❖ To learn the Workflow Editor nodes

1. Drag a node to the workspace.
2. Click the Workflow Node to activate its functions, displayed in the right pane.
3. In the right pane, examine the available options for that node.

   Note The same options are available in the wizards.

4. Delete the node by selecting the node and pressing DELETE.
5. Repeat steps 1 through 4 for each node.

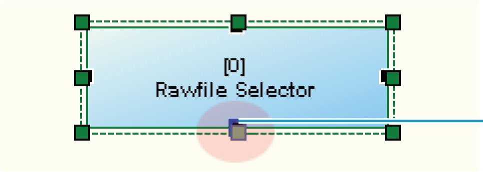
❖ To create a search sequence

1. Drag the **Rawfile Selector** node to the workspace.
2. Depending on your data needs, drag and drop ideal nodes to the workspace.
3. Organize the nodes to reflect a procedural order from top to bottom, so that the Rawfile selector remains on top.

   Joining the nodes together creates a step-wise path for Proteome Discoverer to follow as you feed data into the first node, the Rawfile Selector.

4. To join two nodes, click the node so that a blue handle is activated at the bottom-center of the node. See Figure 28.

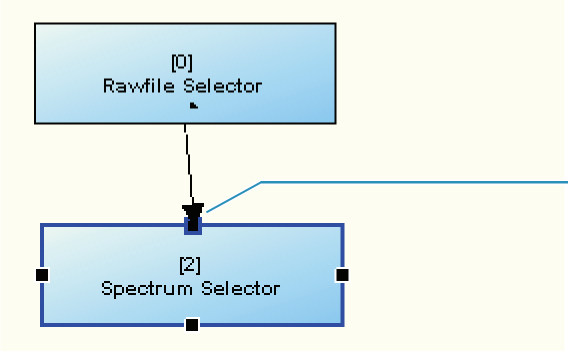
   Figure 28. Activated node example

   Blue handle

5. Drag the blue handle down to the top-center of the node below it (see Figure 29).

   Figure 29. Joining two nodes

   Join the top node to the bottom node.

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

6. Once all your chosen nodes are joined, align them. Choose **Workflow Editor > Auto Layout**.

7. For each node, do the following:
    a. Click the node to activate it.
    b. In the Show Advanced Parameters pane, review each line item for relevancy and accuracy. Choose and alter as fits the raw file base properties.

8. To save the workflow, choose **Workflow Editor > Save Workflow**.

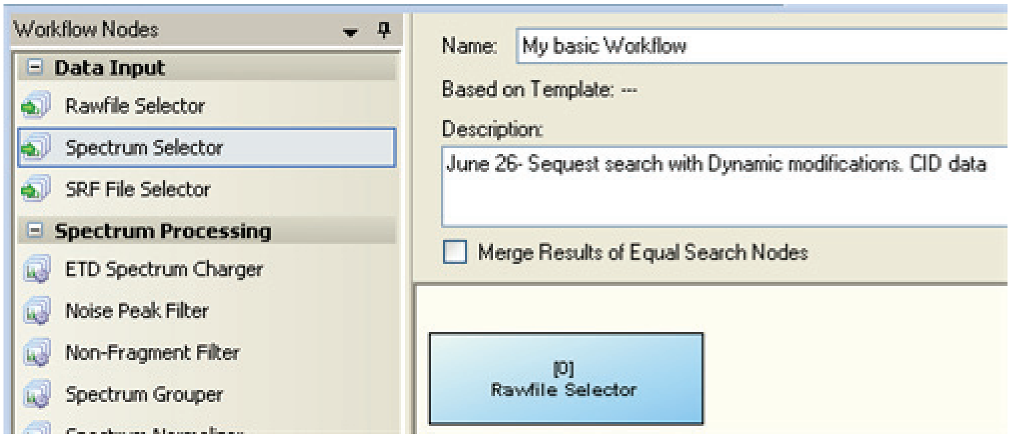
Creating and Saving a Search Workflow

❖ To create a search workflow

1. From the Proteome Discoverer toolbar, choose **Workflow Editor > New Workflow**. The Workflow Editor view opens.

2. Select and drag a **Data Input** node to the workspace, such as the one shown in Figure 30.

Figure 30. Data Input node Rawfile Selector on Workflow Editor pallette

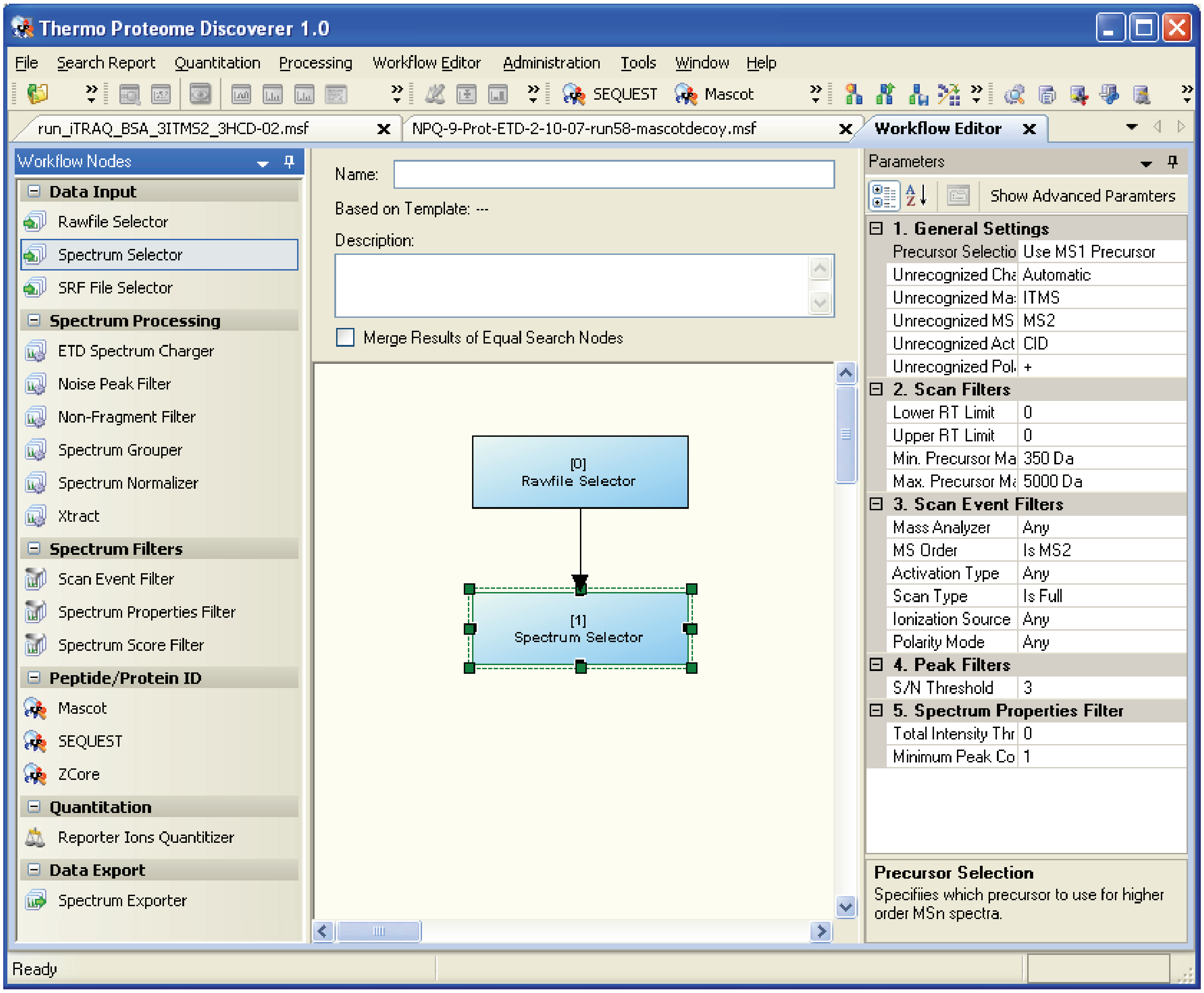


Name: My basic Workflow
Based on Template: ---
Description: June 26- Sequest search with Dynamic modifications. CID data
☐ Merge Results of Equal Search Nodes

3. Start creating your new workflow using a node from Data Input.
    a. From the Data Input category, drag the **Rawfile Selector** or **SRF File Selector** node to the workspace. This is your data input file.
    b. If you selected the Rawfile Selector node as your input, drag the **Spectrum Selector** to the workspace. Figure 31 shows the Rawfile Selector and Spectrum were selected and added to the workspace.

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

Figure 31. Adding Rawfile Selector and Spectrum Selector to a workflow



1. General Settings	
Precursor Selectio	Use MS1 Precursor
Unrecognized Cha	Automatic
Unrecognized Ma	ITMS
Unrecognized MS	MS2
Unrecognized Act	CID
Unrecognized Pol	+
2. Scan Filters	
Lower RT Limit	0
Upper RT Limit	0
Min. Precursor Ma	350 Da
Max. Precursor Ma	5000 Da
3. Scan Event Filters	
Mass Analyzer	Any
MS Order	Is MS2
Activation Type	Any
Scan Type	Is Full
Ionization Source	Any
Polarity Mode	Any
4. Peak Filters	
S/N Threshold	3
5. Spectrum Properties Filter	
Total Intensity Thr	0
Minimum Peak Co	1

Precursor Selection
Specifies which precursor to use for higher order MSn spectra.

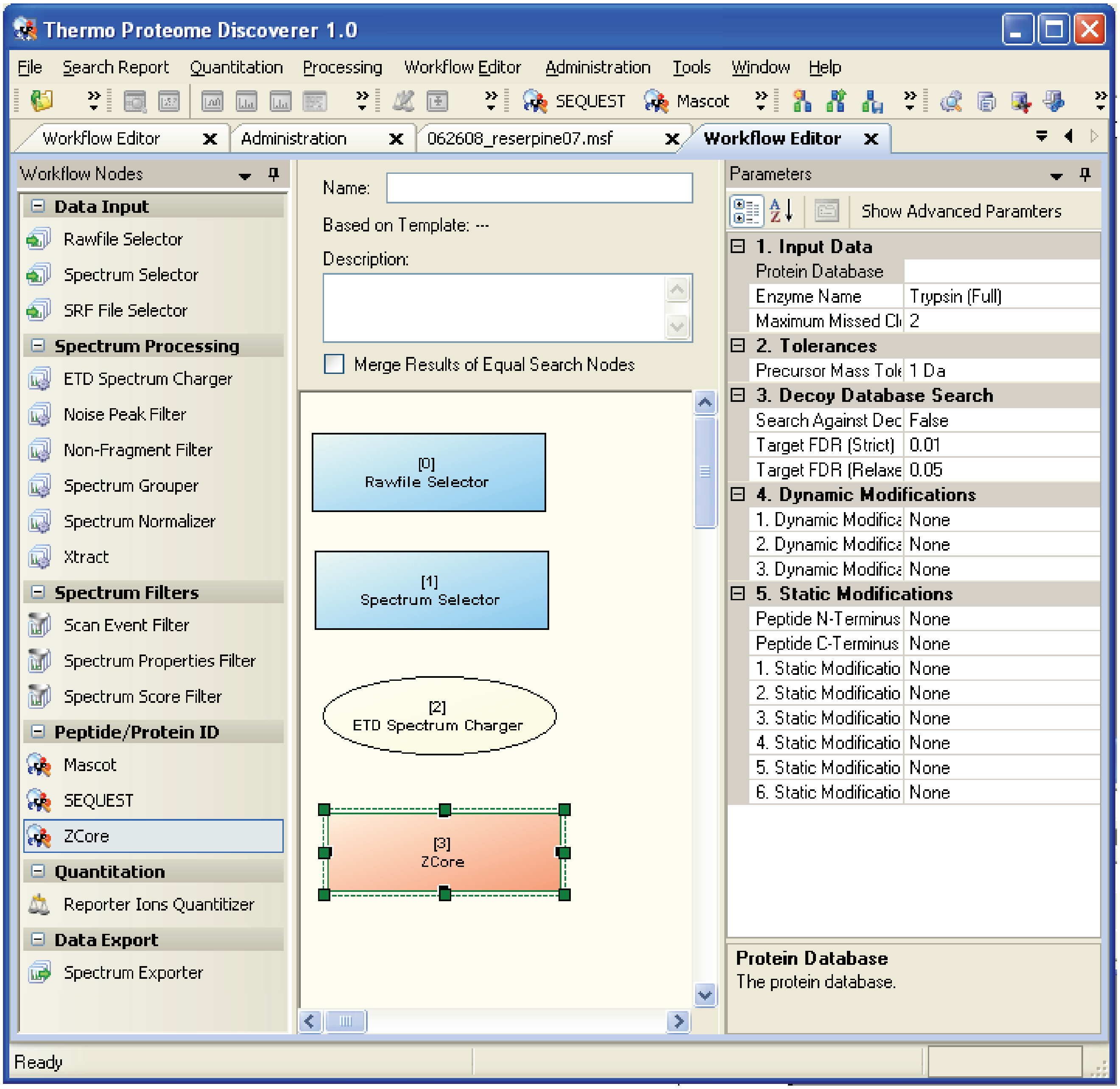
4. (Optional) Drag a node from the Spectrum Processing category to the workspace. You can use one or more nodes from the Spectrum Processing category for your workflow.
   - Use the ETD Spectrum Charger node to calculate precursor ion charge states for ETD spectra, used for all low-mass accuracy ETD data.
   - Use the Xtract node to deconvolve the precursors for all high-mass accuracy data regardless of the fragmentation type.
   - Use the Spectrum Grouper node to apply a grouping function to the data set.
   - Use the Noise Peak Filter node or the Spectrum Normalizer to define specialized filtering early in the search process.
5. (Optional) Drag a Spectrum Filters node or nodes to the workspace. See Figure 32.

Use Scan Event Filter for high-mass accuracy data such as Mascot analysis and SEQUEST analysis of mixed fragmentation mode type data (CID and ETD). It can filter on information such as fragmentation type and mass analyzer identity.

6. Drag a Peptide/Protein ID node to the workspace. Figure 32 shows that the ZCore search engine was selected.

   Both Mascot and SEQUEST can search ETD and CID data, whereas ZCore can only search ETD data.

Figure 32. Setting your parameters for your workflow



1. Input Data	
Protein Database	
Enzyme Name	Trypsin (Full)
Maximum Missed Cl	2
2. Tolerances	
Precursor Mass Tol	1 Da
3. Decoy Database Search	
Search Against Dec	False
Target FDR (Strict)	0.01
Target FDR (Relaxe	0.05
4. Dynamic Modifications	
1. Dynamic Modific	None
2. Dynamic Modific	None
3. Dynamic Modific	None
5. Static Modifications	
Peptide N-Terminus	None
Peptide C-Terminus	None
1. Static Modificatio	None
2. Static Modificatio	None
3. Static Modificatio	None
4. Static Modificatio	None
5. Static Modificatio	None
6. Static Modificatio	None

Protein Database
The protein database.

7. Select a node in the workspace. The available parameters for the node appear in the Parameters pane.
    - In the Parameters pane, define your parameters and input file. Complete this step for each node you selected.
8. Link the nodes to develop a workflow:
    a. Select the bottom edge of a node in the center.
    b. Drag your mouse to connect the two nodes. An arrow appears.
9. In the Name box, enter a unique identifier for the name of your workflow. This name is also the name of your .msf results file.
10. In the Description box, enter a description of your workflow.
11. To start your search, select **Workflow Editor > Start Workflow**. The Job Queue pane appears, showing the status of your search.
12. Use the Job Queue to check the status of your search. For details, see "Working with the Search Job Queue" on page 60.

Note To view your search results, see Chapter 3, "Interpreting Search Results."

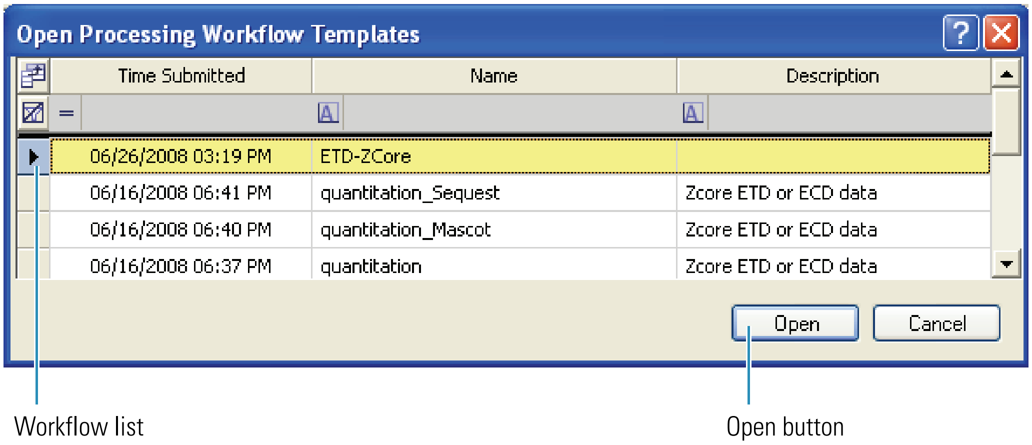
❖ To save a search workflow

1. After completing a new search workflow design, see "To create a search workflow" on page 53.
2. Click **Save As Template**. A dialog box appears.
3. Enter a name which describes the workflow contents.

❖ To open a saved workflow

1. Choose **Workflow > Open From Template**. The Open Processing WorkFlow Templates page appears.
2. Select a workflow from the list.
3. Click **Open**. The Workflow Editor window opens with the selected workflow displayed.

Figure 33. Open WorkFlow page

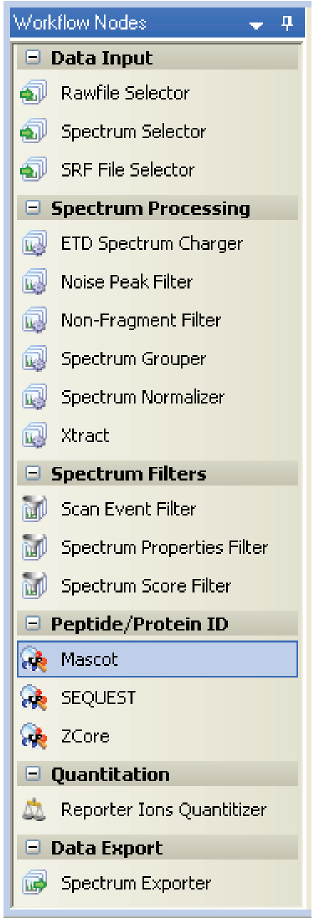

Open Processing Workflow Templates

Time Submitted	Name	Description
06/26/2008 03:19 PM	ETD-ZCore	
06/16/2008 06:41 PM	quantitation_Sequest	Zcore ETD or ECD data
06/16/2008 06:40 PM	quantitation_Mascot	Zcore ETD or ECD data
06/16/2008 06:37 PM	quantitation	Zcore ETD or ECD data

Discoverer Workflow Nodes Details

This section describes the six categories of workflow nodes. See "Proteome Discoverer Workflow" on page 19 for short definitions.

Figure 34. Workflow nodes



Data Input

A logical workflow contains a sequence of processing steps. Begin with raw spectra data, and then process the data with operators that you can choose from the workflow node selection pane. You can also set the workflow to export the results into other data formats.

You can begin your search flow with two data input nodes. Use either the Rawfile Selector or the SRF File Selector to specify your data input file.

IMPORTANT You must use Spectrum Selector with the Rawfile Selector.

Spectrum Processing

Once you have selected your input data, you can apply several processing functions to the data. Spectrum Processing provides six nodes to define your search parameters:

- ETD Spectrum Charger
- Noise Peak Filter
- Non-Fragment Filter
- Spectrum Grouper
- Spectrum Normalizer
- Xtract

Spectrum Filters

You can filter input data prior to a database search to remove lower quality spectral peak lists from your analysis. This can decrease search times and false positive identifications. The Spectrum Filter nodes provide three types of spectrum filters to use for your search. Use these pre-analysis filters to streamline your search results.

Peptide ID and Protein ID nodes

The Proteome Discoverer Workflow Editor provides an option to use any of the three search engines with the Peptide/Protein ID nodes: Mascot, SEQUEST, or ZCore.

Quantitation

The Proteome Discoverer Workflow Editor provides an option to quantitate your spectra with the Reporter Ions Quantitizer. For instructions to create a quantitation method, see the procedure, "To apply a quantitization node to a workflow" on page 69.

Data Export

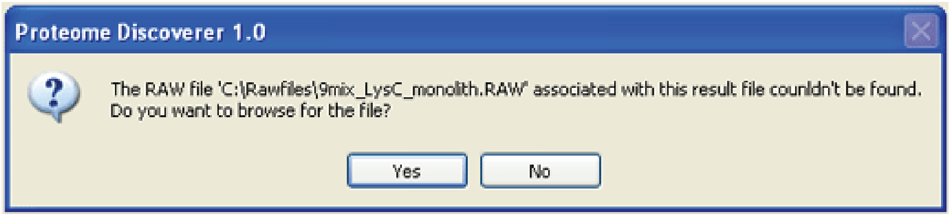
The Proteome Discoverer Workflow Editor provides an option to export your spectra with the Spectrum Exporter node.

Working with Reports

- Checking Your Job Queue Search Status
- Deleting Items from the Queue List

❖ To open or load a report (menu)

1. In the Discoverer window, choose **File > Open Report**.
2. Find the .msf file to upload and click **OK**. A status page appears as the information is organized to display the report.
3. (Optional) If you receive a .raw file query message, find the corresponding .raw file if it is available.
   a. Click **Yes** to find the .raw file.



The RAW file 'C:\Rawfiles\9mix_LysC_monolith.RAW' associated with this result file couldn't be found. Do you want to browse for the file?

   b. If it is found, click **Open** to accept the .raw file.

The search results are displayed in the Proteome Discoverer results window.

❖ To open an .srf file

1. Choose **File > Import Search Results from SRF**. A wizard appears.
2. Follow the instructions to open the appropriate file.

❖ To save a report

1. In the Proteome Discoverer window, click the appropriate report tab to save your modifications.
2. Choose **File > Save Report**. The filter settings are stored in the results file (.msf).

❖ To close a report without saving changes

1. In the Proteome Discoverer window, click the report tab you want to close.
2. Choose **File > Close**.
   –or–
   Right-click the report you want to close, and choose **Close** from the shortcut menu

   Your changes or filter settings are not automatically saved. You must save your changes to keep them for future use.

❖ To view the protein grid

In an open report, click the **Protein** tab. The protein view of your search report appears.

❖ To view the peptide grid

In an open report, click the **Peptide** tab. The peptide view of your search report appears.

Related Topics

- To open a Job Queue view
- To open the Qual Browser

Working with the Search Job Queue

Use the search queue to check the status of your search or remove search results from the Job Queue list. For detailed job queue information, see page 73.

- Checking Your Job Queue Search Status
- Deleting Items from the Queue List

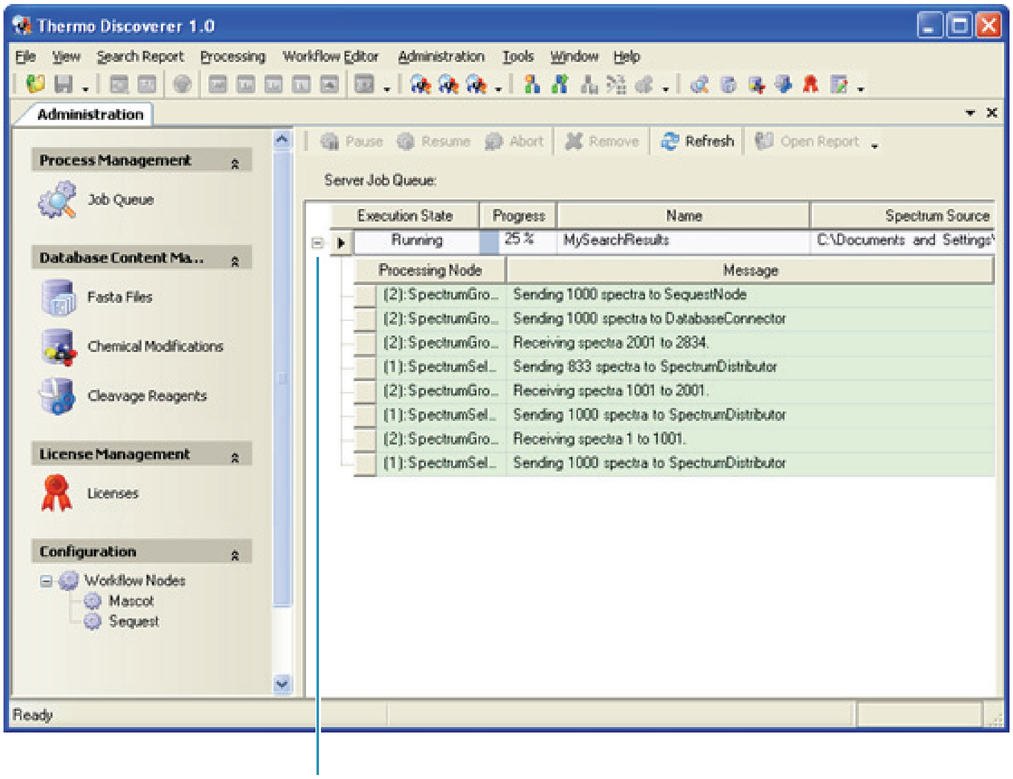
Checking Your Job Queue Search Status

You can check the status of your search and results, or you can review the result details from the Job Queue view.

❖ To check the search results status from the Job Queue list

1. Choose **Administration > Show Job Queue** to open the Job Queue view.

Figure 35. Job queue



Server Job Queue:

Execution State	Progress	Name	Spectrum Source
Running	25 %	MySearchResults	C:\Documents and Settings\

Processing Node	Message
(2):SpectrumGro...	Sending 1000 spectra to SequestNode
(2):SpectrumGro...	Sending 1000 spectra to DatabaseConnector
(2):SpectrumGro...	Receiving spectra 2001 to 2834.
(1):SpectrumSel...	Sending 833 spectra to SpectrumDistributor
(2):SpectrumGro...	Receiving spectra 1001 to 2001.
(1):SpectrumSel...	Sending 1000 spectra to SpectrumDistributor
(2):SpectrumGro...	Receiving spectra 1 to 1001.
(1):SpectrumSel...	Sending 1000 spectra to SpectrumDistributor

Click **+** to expand or **-** to collapse row details.

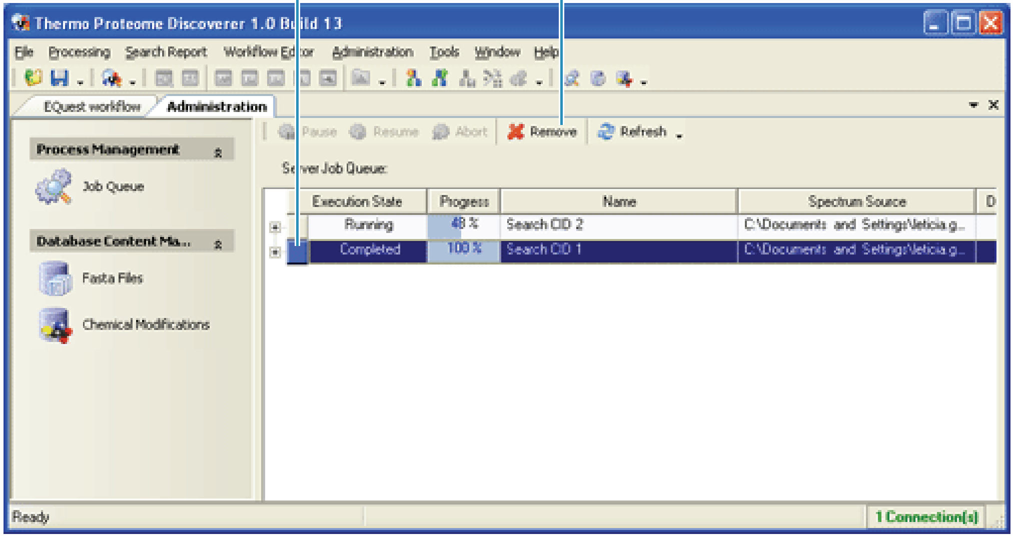
2. To expand and view the details of the search, click **+** to the left of the job.

Deleting Items from the Queue List

You can remove search results from the Job Queue list.

❖ To delete items from the Job Queue list

1. Choose **Administration > Show Job Queue** to open the Job Queue view.
2. Select the box to the left of the row of the job you want to delete. The job is selected.



Select job.

Click Remove to delete a row.

Execution State	Progress	Name	Spectrum Source	D
Running	48 %	Search CID 2	C:\Documents and Settings\leticia.g...	
Completed	100 %	Search CID 1	C:\Documents and Settings\leticia.g...	

3. Click **Remove** on the toolbar of the Job Queue view. A Delete Jobs message appears.
4. Click **OK.**

   The selected job is removed.

Using FASTA Database Utilities

- FASTA Files
- Adding Protein References to FASTA
- Performing a Decoy Database Search

With FASTA Database Utilities you can import your FASTA database files, review the properties of your FASTA file, determine if the database is in a readable format, or modify the way in which the protein titles in the database are parsed. When you select the Tools > FASTA Database Utilities command, Proteome Discoverer displays a view to perform these actions.

The FASTA Files view displays the processed .fasta file properties, such as the file name, file size, and number of proteins stored in the table. For each protein entry, the FASTA file structure is analyzed to determine if the .fasta file meets the requirements for Proteome Discoverer to use in a spectra search. The .fasta file is quickly processed and made available for use.

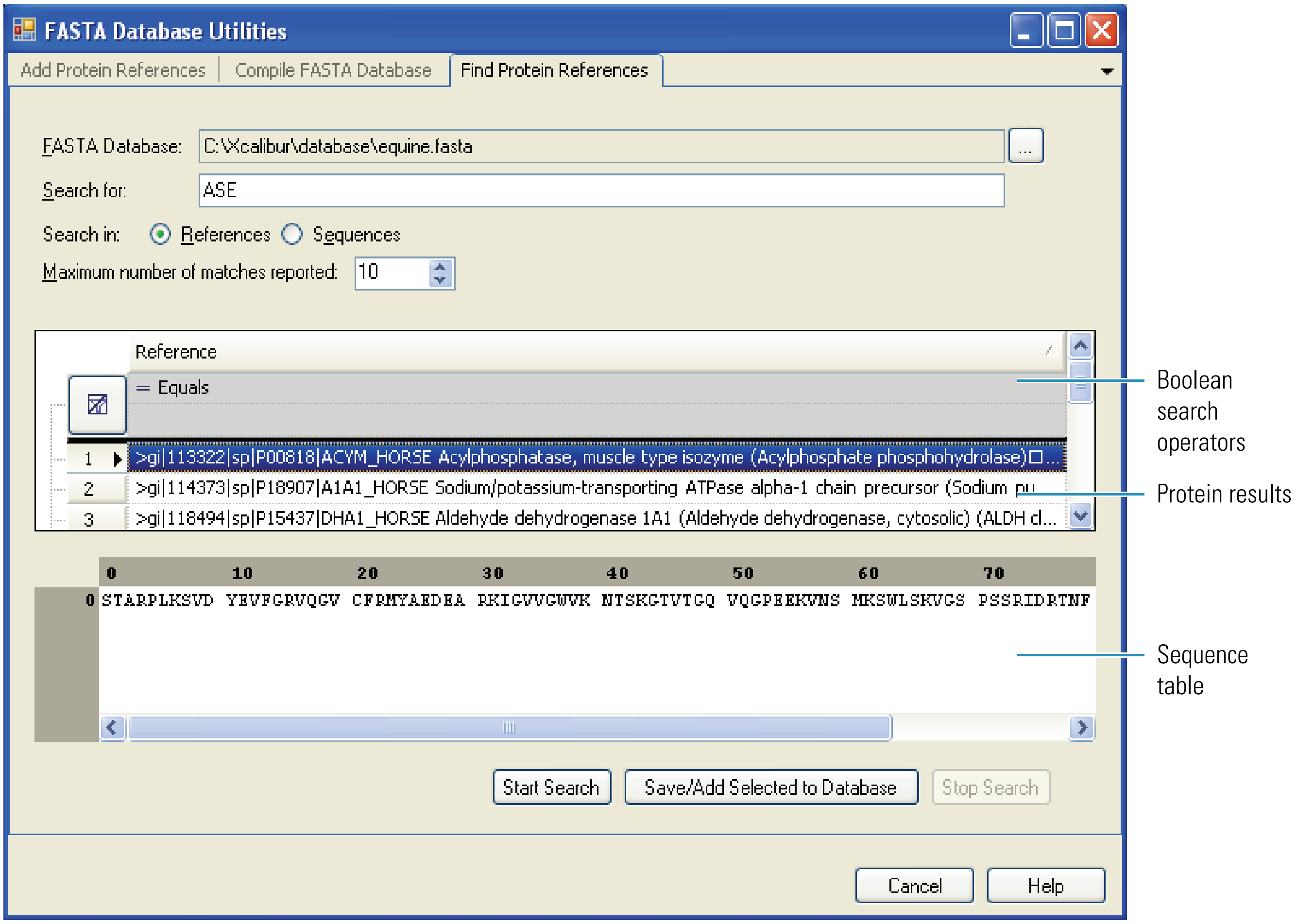
FASTA Files

This section describes how to add a protein reference to your FASTA database file.

❖ To find a protein reference

1. Choose **Tools > FASTA Database Utilities**.
2. Click **Find Protein Reference**. The Find Protein Reference page opens.

Figure 36. Find Protein References page in the FASTA Database Utilities window



FASTA Database: C:\Xcalibur\database\equine.fasta

Search for: ASE

Search in: ● References ○ Sequences

Maximum number of matches reported: 10

	Reference
1	>gi|113322|sp|P00818|ACYM_HORSE Acylphosphatase, muscle type isozyme (Acylphosphate phosphohydrolase)☐…
2	>gi|114373|sp|P18907|A1A1_HORSE Sodium/potassium-transporting ATPase alpha-1 chain precursor (Sodium pu
3	>gi|118494|sp|P15437|DHA1_HORSE Aldehyde dehydrogenase 1A1 (Aldehyde dehydrogenase, cytosolic) (ALDH cl…

	0	10	20	30	40	50	60	70
0	STARPLKSVD	YEVFGRVQGV	CFRMYAEDEA	RKIGVVGWVK	NTSKGTVTGQ	VQGPEEKVNS	MKSWLSKVGS	PSSRIDRTNF

Boolean search operators

Protein results

Sequence table

3. To find the .fasta file of interest, click the browse button (**...**).
4. Type the amino acid sequence of interest for a search string.
5. Set the Boolean search operators as needed.
6. Click **Start Search**. Results appear if the search parameters match the data.
7. (Optional) Select a protein result row. The sequence table below shows the theoretical amino acid results for the selected protein.

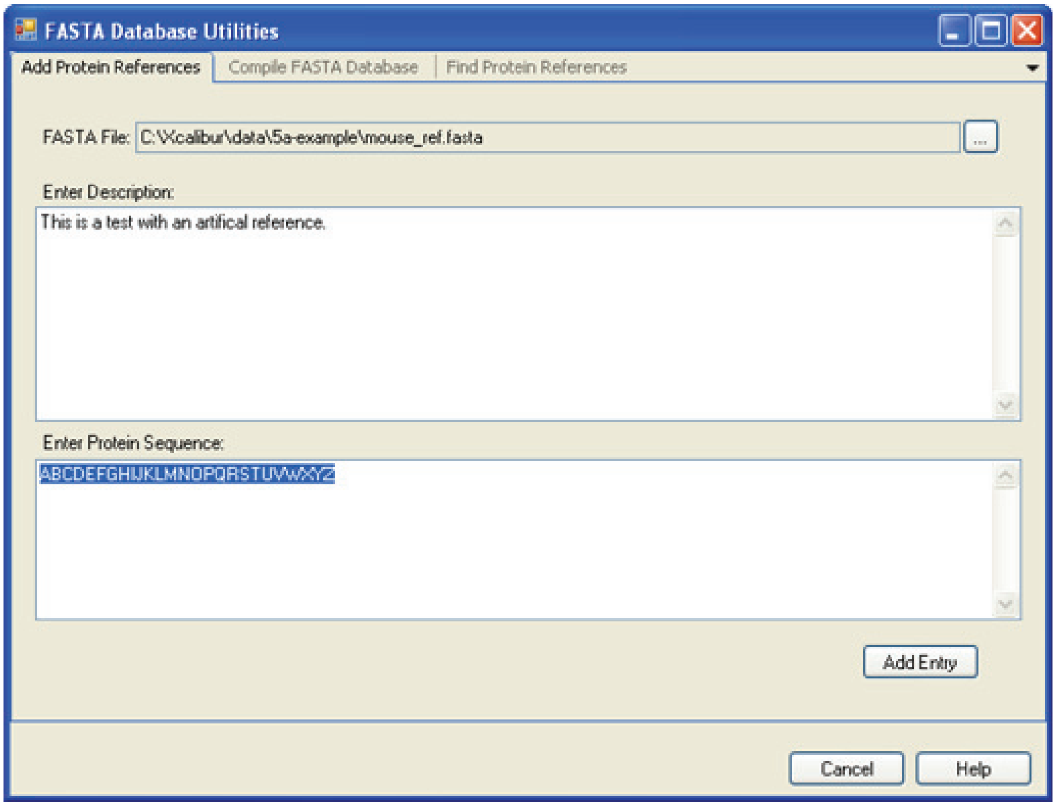
Adding Protein References to FASTA

Use the Add Protein References page to locate and modify an existing FASTA file. You can add a protein sequence or reference to a registered FASTA file.

❖ To open the Add Protein References page

1. Choose **Tools > FASTA Database Utilities**.
2. Click the **Add Protein References** tab. The Add Protein Reference page opens.
3. Click the browse button (**...**) to locate your file of interest.
4. In the Enter Description box, type the description of the .fasta file.

Figure 37. Add Protein References page in the FASTA Database Utilities window



FASTA File: C:\Xcalibur\data\5a-example\mouse_ref.fasta

Enter Description:
This is a test with an artifical reference.

Enter Protein Sequence:
ABCDEFGHIJKLMNOPQRSTUVWXYZ

Performing a Decoy Database Search

The false discovery rate (FDR) or false positive rate is a statistical value that estimates the number of false positive identifications among all identifications found by a peptide ID search. It is a measure of the certainty of the identification.

Although there is more than one way to determine FDRs, the following topics describe how to set and determine false discovery rates using Proteome Discoverer.

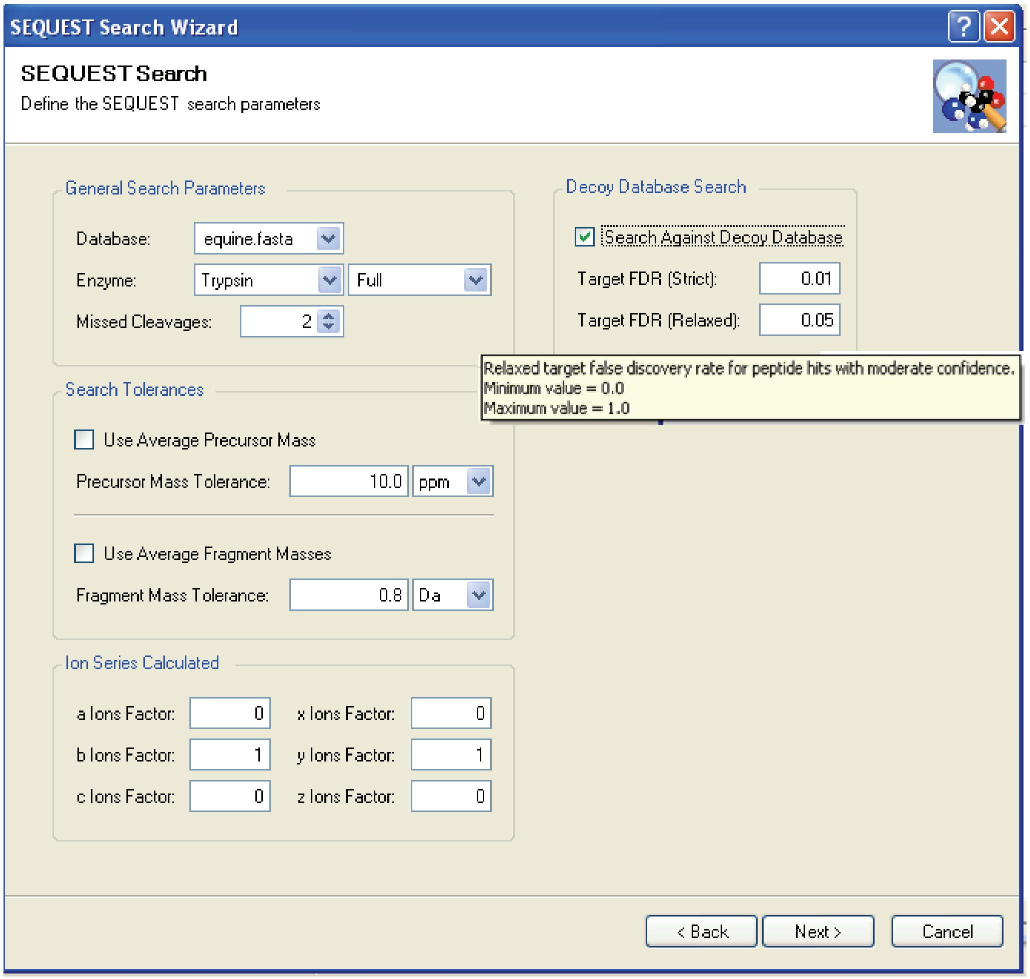
You can use FDRs to validate MS/MS searches of large data sets, but they are not effective on searches of a small number of spectra, or searches against a small number of protein sequences, because the number of matches will likely be too small to give a statistically meaningful estimate.

The following procedure describes how to use the search wizards to perform a decoy database search to achieve FDRs.

❖ To set a decoy database search

1. Start your search by using the search wizards. See the procedure, "To use the SEQUEST Search Wizard on a raw file" on page 30.
2. On the Search parameters page, select the **Search against decoy database** option. See Figure 38.
3. Set your target FDR for high and medium confidence.
4. Continue to define your remaining search criteria.

Figure 38. Decoy Database Search parameters



SEQUEST Search Wizard

SEQUEST Search
Define the SEQUEST search parameters

General Search Parameters
Database: equine.fasta
Enzyme: Trypsin Full
Missed Cleavages: 2

Search Tolerances
☐ Use Average Precursor Mass
Precursor Mass Tolerance: 10.0 ppm
☐ Use Average Fragment Masses
Fragment Mass Tolerance: 0.8 Da

Ion Series Calculated
a Ions Factor: 0
b Ions Factor: 1
c Ions Factor: 0
x Ions Factor: 0
y Ions Factor: 1
z Ions Factor: 0

Decoy Database Search
☑ Search Against Decoy Database
Target FDR (Strict): 0.01
Target FDR (Relaxed): 0.05

Relaxed target false discovery rate for peptide hits with moderate confidence.
Minimum value = 0.0
Maximum value = 1.0

Working with Chemical Modifications

The Chemical Modification view is used to build and maintain the static and dynamic modifications data that is available when you define your search settings.

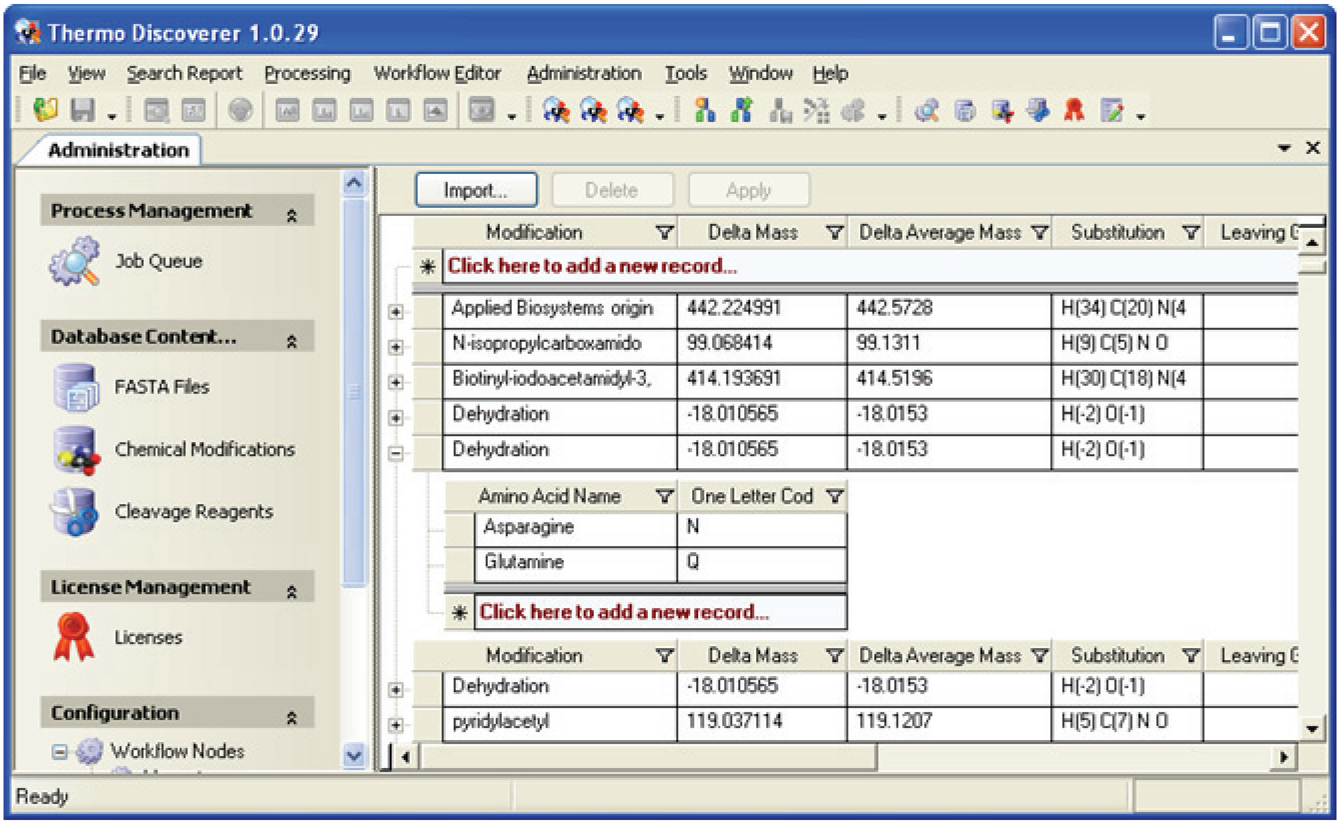
❖ To open the Chemical Modifications view

1. Choose **Administration > Chemical Modifications**. The Chemical Modifications view appears.

2. Explore the default types of modifications and their corresponding amino acids.

   The modifications table contains the modification's delta mass, amino acids, and substitution. By using the Chemical Modifications view, you can add amino acids to existing modifications and create new modifications.

Figure 39. Chemical Modifications Administration view



Modification	Delta Mass	Delta Average Mass	Substitution	Leaving G
✳ Click here to add a new record...				
Applied Biosystems origin	442.224991	442.5728	H(34) C(20) N(4	
N-isopropylcarboxamido	99.068414	99.1311	H(9) C(5) N O	
Biotinyl-iodoacetamidyl-3,	414.193691	414.5196	H(30) C(18) N(4	
Dehydration	-18.010565	-18.0153	H(-2) O(-1)	
Dehydration	-18.010565	-18.0153	H(-2) O(-1)	

Amino Acid Name	One Letter Cod
Asparagine	N
Glutamine	Q
✳ Click here to add a new record...	

Modification	Delta Mass	Delta Average Mass	Substitution	Leaving G
Dehydration	-18.010565	-18.0153	H(-2) O(-1)	
pyridylacetyl	119.037114	119.1207	H(5) C(7) N O	

How to Quantify Your Data Using Adminstration and Workflow

Proteome Discoverer includes a protein identification and quantitation package with SEQUEST and Zcore licenses. You can measure and report the relative quantitation of isotopically-labeled peptides.

Reporter Ion Based Quantitation in Proteome Discoverer

This section introduces reporter ion based quantitation (RIQ). Use the quantitation workflow and the data retrieval and storage process to quantify your results.

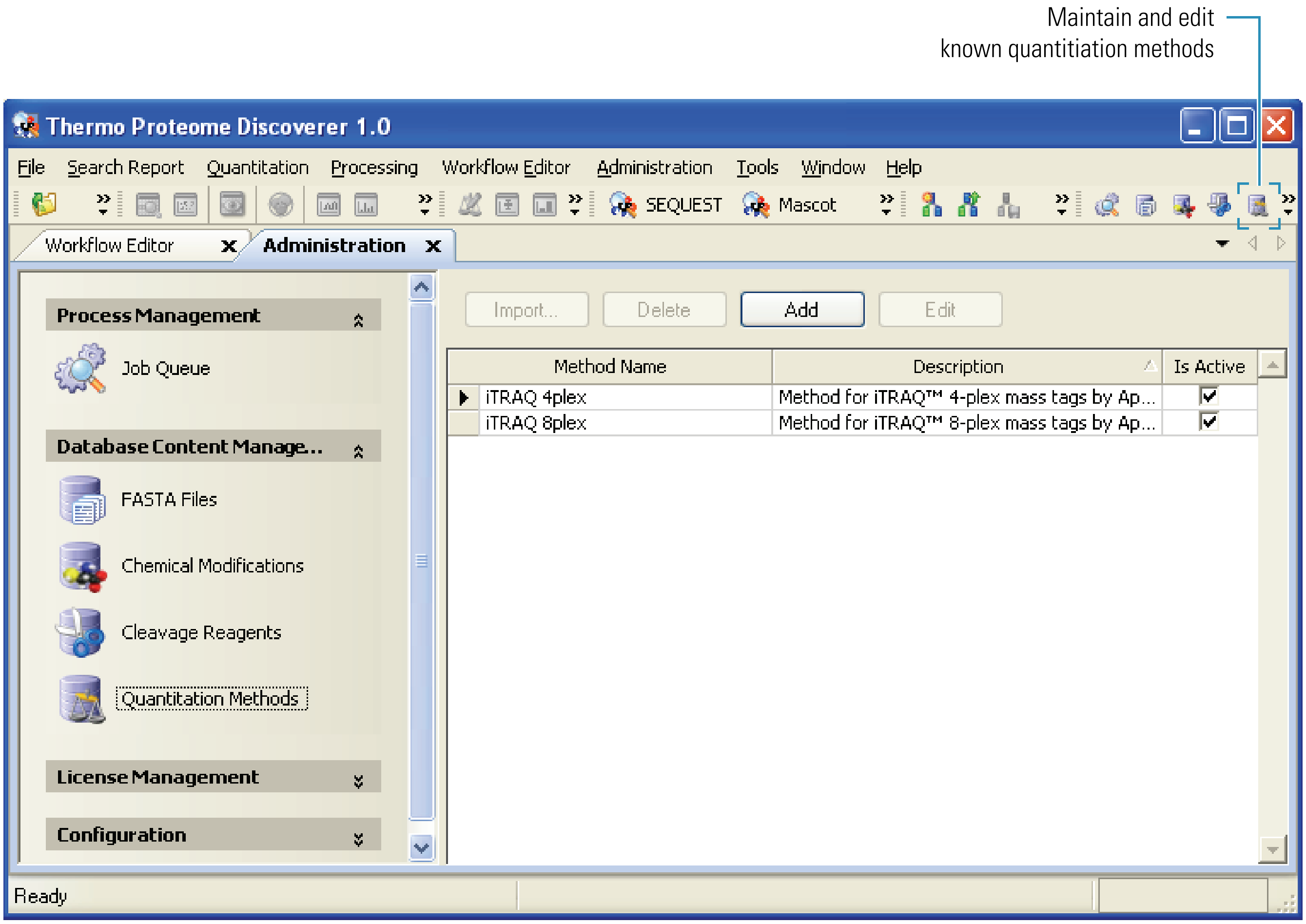
The following procedure discusses the reporting side, improvements in the calculation and statistics steps, and improvements for display and validation of quantitation results.

You can quantify all isobaric labeled samples. There are defaults available for iTRAQ-4plex and iTRAQ-8plex. You can also add new methods.

❖ To select a quantitation method

1. Choose **Administration > Quantitation Methods**. The reporter ion based quantitation methods are available through this option.
2. To activate the methods that are available when designing a Proteome Discoverer workflow, select the respective Method Name check box. See Figure 40.

Figure 40. Quantitation Methods view



Maintain and edit known quantitiation methods

Method Name	Description	Is Active
iTRAQ 4plex	Method for iTRAQ™ 4-plex mass tags by Ap...	☑
iTRAQ 8plex	Method for iTRAQ™ 8-plex mass tags by Ap...	☑

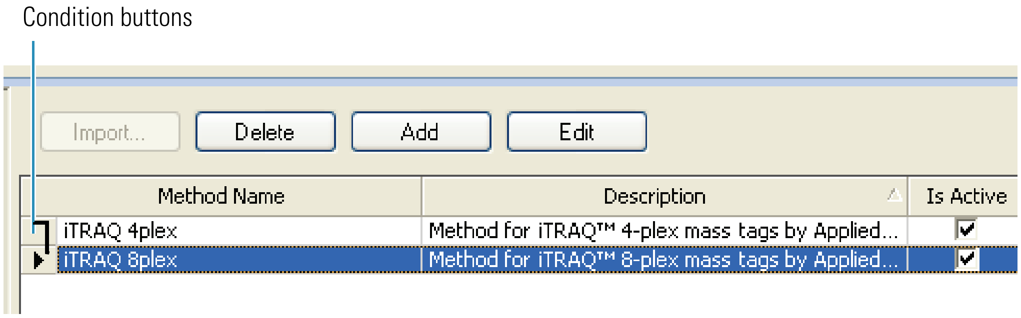
❖ To edit a method

1. Choose **Administration > Quantitation Methods**. The reporter ion based quantitation methods are available through this option.

2. Click ☐ of the a Method Name row. An arrow (▶) appears next to the selected name.

   The Edit-button is now active (see Figure 41).

Figure 41. Select method to edit

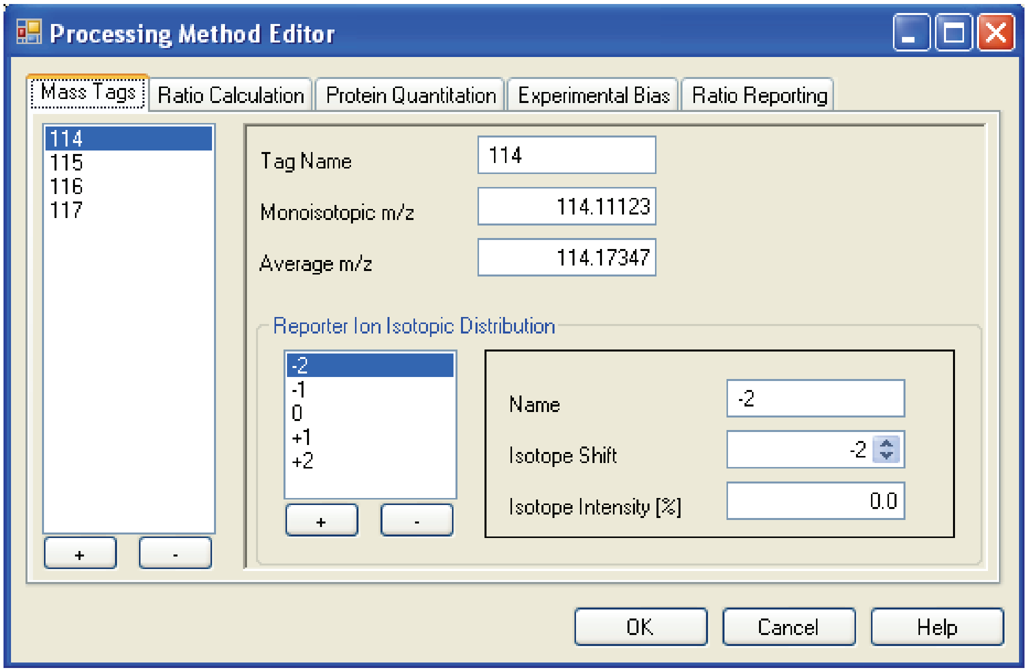


Condition buttons

Method Name	Description	Is Active
iTRAQ 4plex	Method for iTRAQ™ 4-plex mass tags by Applied...	☑
iTRAQ 8plex	Method for iTRAQ™ 8-plex mass tags by Applied...	☑

3. Click Edit
4. The Processing Method Editor appears.
5. Choose each of the following tabs to verify the options are set correctly for your raw data set. See Figure 42.
   - Mass Tags
   - Ration Calculation
   - Protein Quantitation
   - Experimental Bias
   - Ratio Reporting

Figure 42. Quantitation Processing Method Editor tab options



Tag Name: 114

Monoisotopic m/z: 114.11123

Average m/z: 114.17347

Name: -2

Isotope Shift: -2

Isotope Intensity [%]: 0.0

❖ To apply a quantitization node to a workflow

1. Open a workflow template. See “To create a search workflow” on page 53.
2. Choose **Quantitation >Reporter Ions Quantitizer**.
3. Drag the node to the workspace.
4. Connect the Rawfile Selector node to the Reporter Ions Quantitizer node.
5. Edit options for each node. See “To learn the Workflow Editor nodes” on page 52.
6. Select **Workflow Editor > Save as Template**.

Quantitation Summary

You can display the quantitation spectrum of a selected sequence with the quantitation summary view.

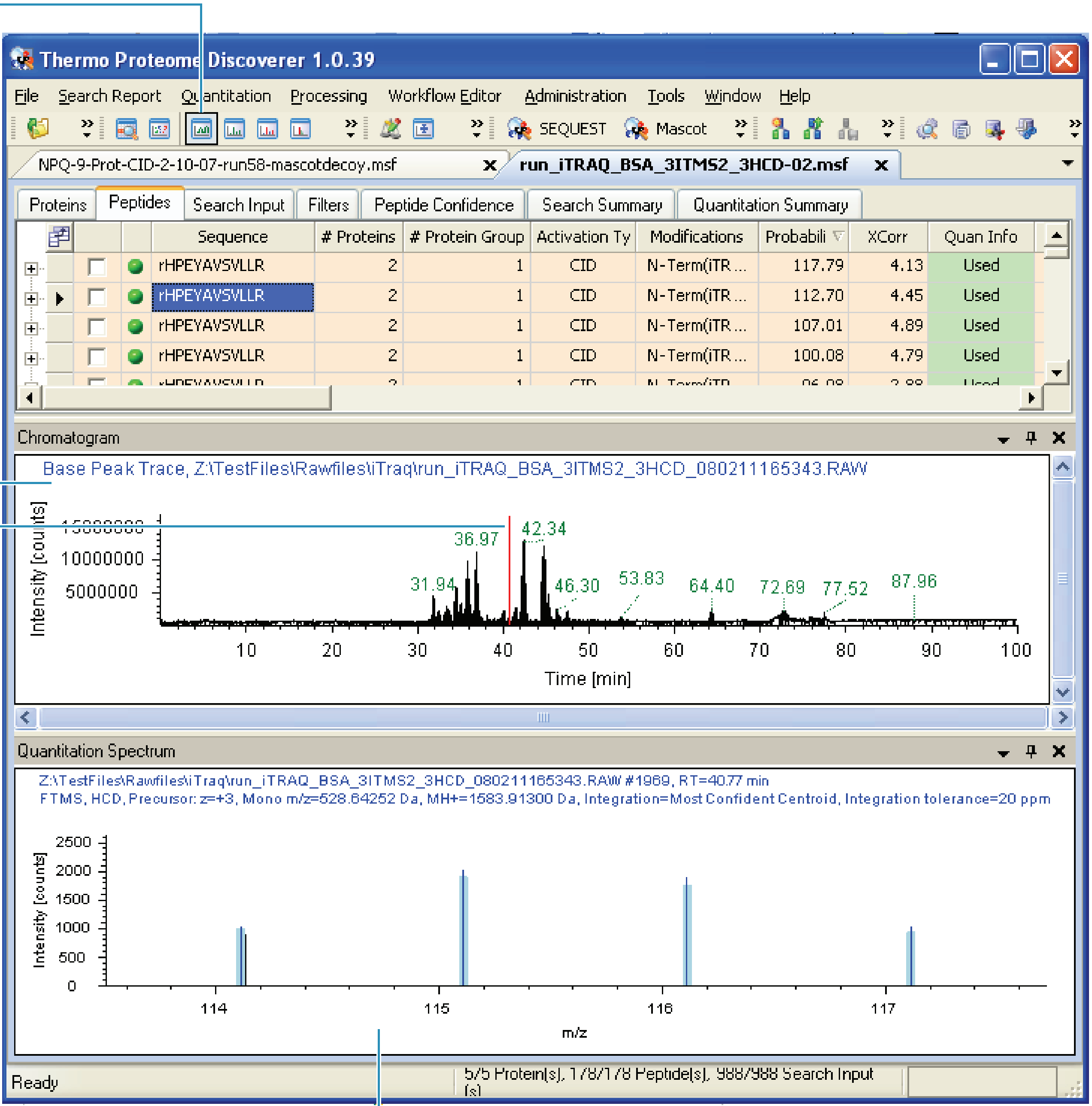
IMPORTANT To use the quantitation menu options, the raw data file must have been acted on by the Workflow Reporter Ion Qunatitizer node. See "Using the Proteome Discoverer Workflow Editor" on page 49.

❖ To show quantitation spectrum

1. Click the **Peptide** tab.
2. Select a sequence of interest.
3. Choose **Quantitation > Show Quantitation Summary**. A Quantitation Spectrum chart appears. See Figure 43.

Figure 43. Quantitation summary with a chromatogram view

A chromatogram view shows the intensities of masses as a function of time.



Sequence	# Proteins	# Protein Group	Activation Ty	Modifications	Probabili	XCorr	Quan Info
rHPEYAVSVLLR	2	1	CID	N-Term(iTR...	117.79	4.13	Used
rHPEYAVSVLLR	2	1	CID	N-Term(iTR...	112.70	4.45	Used
rHPEYAVSVLLR	2	1	CID	N-Term(iTR...	107.01	4.89	Used
rHPEYAVSVLLR	2	1	CID	N-Term(iTR...	100.08	4.79	Used

Base Peak Trace, Z:\TestFiles\Rawfiles\iTraq\run_iTRAQ_BSA_3ITMS2_3HCD_080211165343.RAW

Z:\TestFiles\Rawfiles\iTraq\run_iTRAQ_BSA_3ITMS2_3HCD_080211165343.RAW #1969, RT=40.77 min
FTMS, HCD, Precursor: z=+3, Mono m/z=528.64252 Da, MH+=1583.91300 Da, Integration=Most Confident Centroid, Integration tolerance=20 ppm

Ready

575 Protein(s), 178/178 Peptide(s), 988/988 Search Input (s)

The chromatogram shows the retention time when the selected peptide was eluted.

The quantitation spectrum of the selected peptide shows the intensities of the detected reporter ions.

Assess the Abundance of the Precursor

By using the graphically displayed intensity of the peptide, you are able correct for noise and move the baseline noise.

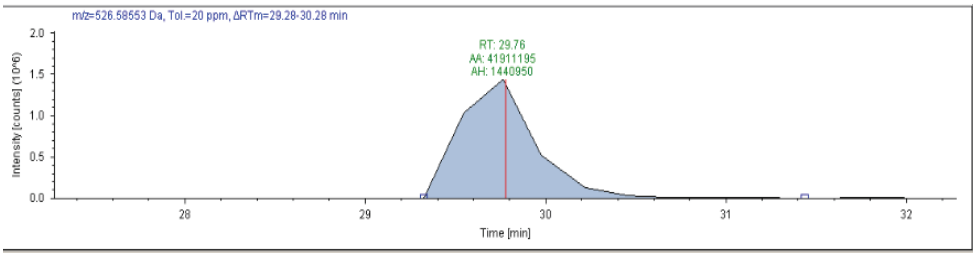
- The peak start and end points as well as the baseline are in blue.
- The peak area and the height values are automatically calculated.

You can use the view to assess the chromatographic peak shape of the associated precursor, and to reference the elution time of the identified peptide. The integrated area under the curve and height of the peak is displayed and can be used to assess the abundance of the precursor.

❖ To magnify a peak

Drag your mouse over the region of interest.

Figure 44. Extracted ion chromatograph



Using Qual Browser

Qual Browser automatically displays the elemental composition, theoretical mass, RDB, and delta values for your high-resolution data.

❖ To open the Qual Browser

1. Choose **Tools > Open Qual Browser** to open the Spectrum window.
2. Right-click and choose **Display Options** from the shortcut menu.
3. To automatically annotate your peaks, click the **Composition** tab and select the labels for display.

Job Queue

After you set up and start the search project, you can view a list of searches and their states in the Job Queue view. The Job Queue displays the search name, the .raw file name, the date the search was submitted, and the progress of the search and its status.

Using the Job Queue

The following procedure describes how to use the job queue options.

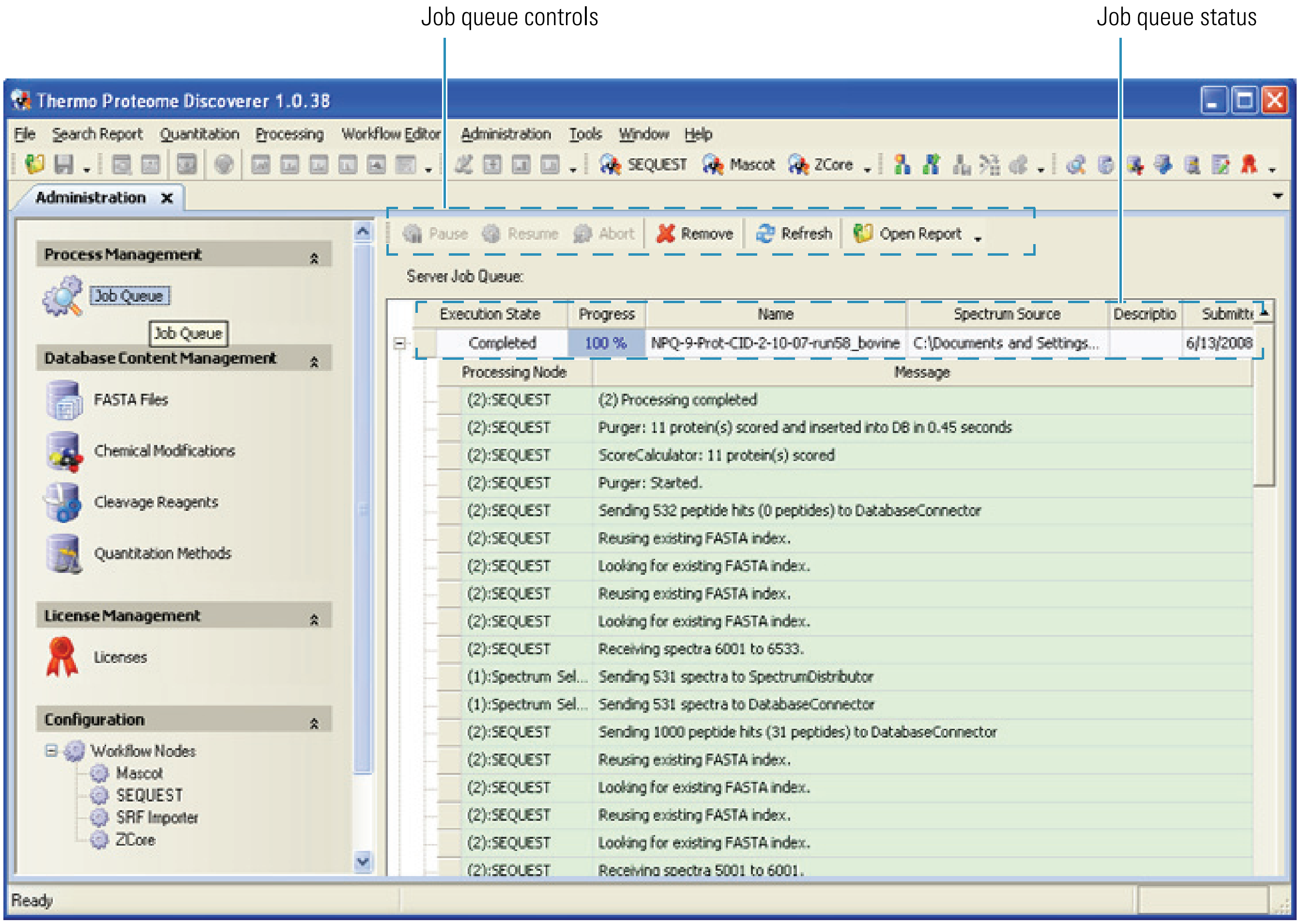
❖ To open a Job Queue view

Choose **Administration > Job Queue**.

From the Job Queue view, you can do the following:

- Refresh the job queue status.
- Check the status of a job.
- Pause or resume a job.
- Remove or cancel a job.
- Open a report in the Server Job Queue. See “Open a Completed Report From Job Queue” on page 74.

Figure 45. Job queue diagram of controls and status



Execution State	Progress	Name	Spectrum Source	Descriptio	Submitt
Completed	100 %	NPQ-9-Prot-CID-2-10-07-run58_bovine	C:\Documents and Settings...		6/13/2008

Processing Node	Message
(2):SEQUEST	(2) Processing completed
(2):SEQUEST	Purger: 11 protein(s) scored and inserted into DB in 0.45 seconds
(2):SEQUEST	ScoreCalculator: 11 protein(s) scored
(2):SEQUEST	Purger: Started.
(2):SEQUEST	Sending 532 peptide hits (0 peptides) to DatabaseConnector
(2):SEQUEST	Reusing existing FASTA index.
(2):SEQUEST	Looking for existing FASTA index.
(2):SEQUEST	Reusing existing FASTA index.
(2):SEQUEST	Looking for existing FASTA index.
(2):SEQUEST	Receiving spectra 6001 to 6533.
(1):Spectrum Sel...	Sending 531 spectra to SpectrumDistributor
(1):Spectrum Sel...	Sending 531 spectra to DatabaseConnector
(2):SEQUEST	Sending 1000 peptide hits (31 peptides) to DatabaseConnector
(2):SEQUEST	Reusing existing FASTA index.
(2):SEQUEST	Looking for existing FASTA index.
(2):SEQUEST	Reusing existing FASTA index.
(2):SEQUEST	Looking for existing FASTA index.
(2):SEQUEST	Receiving spectra 5001 to 6001.

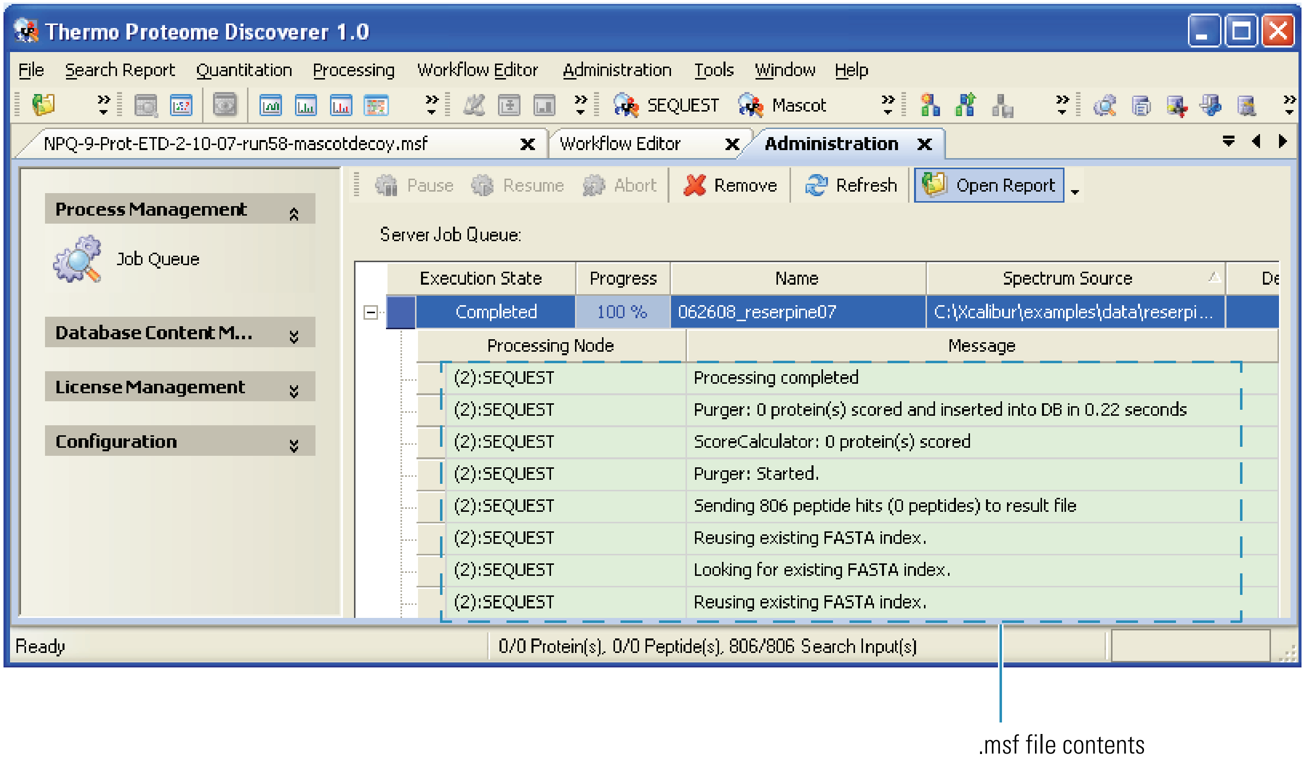
Open a Completed Report From Job Queue

The following procedure describes how to open a report you recently worked on from the Server Job Queue.

1. Choose **Administration > Show Job Queue.**
2. Click the Completed row in which you are interested.
3. **Click** Open Report. The Open dialog box appears (see Figure 46).
4. In the Open dialog box, select the file of interest. The example in Figure 46 uses **062608_reserpine07**.
5. Click **Open Report**. The .msf file appears in a new tab view.

Note The Job Queue view is still available by choosing the Job Queue tab name.

Figure 46. Job Queue Open Report option



Execution State	Progress	Name	Spectrum Source	De
Completed	100 %	062608_reserpine07	C:\Xcalibur\examples\data\reserpi...	

Processing Node	Message
(2):SEQUEST	Processing completed
(2):SEQUEST	Purger: 0 protein(s) scored and inserted into DB in 0.22 seconds
(2):SEQUEST	ScoreCalculator: 0 protein(s) scored
(2):SEQUEST	Purger: Started.
(2):SEQUEST	Sending 806 peptide hits (0 peptides) to result file
(2):SEQUEST	Reusing existing FASTA index.
(2):SEQUEST	Looking for existing FASTA index.
(2):SEQUEST	Reusing existing FASTA index.

Ready

0/0 Protein(s), 0/0 Peptide(s), 806/806 Search Input(s)

Related Topics

- To open or load a report (menu)
- To save a report
- To close a report without saving changes.

InforSense

InforSense Protein Annotation is part of the Proteome Discoverer tool set, which automatically retrieves descriptive information, allowing identified proteins to be placed in their biological context. You can use Proteome Discoverer InforSense as another way to look at your search results.

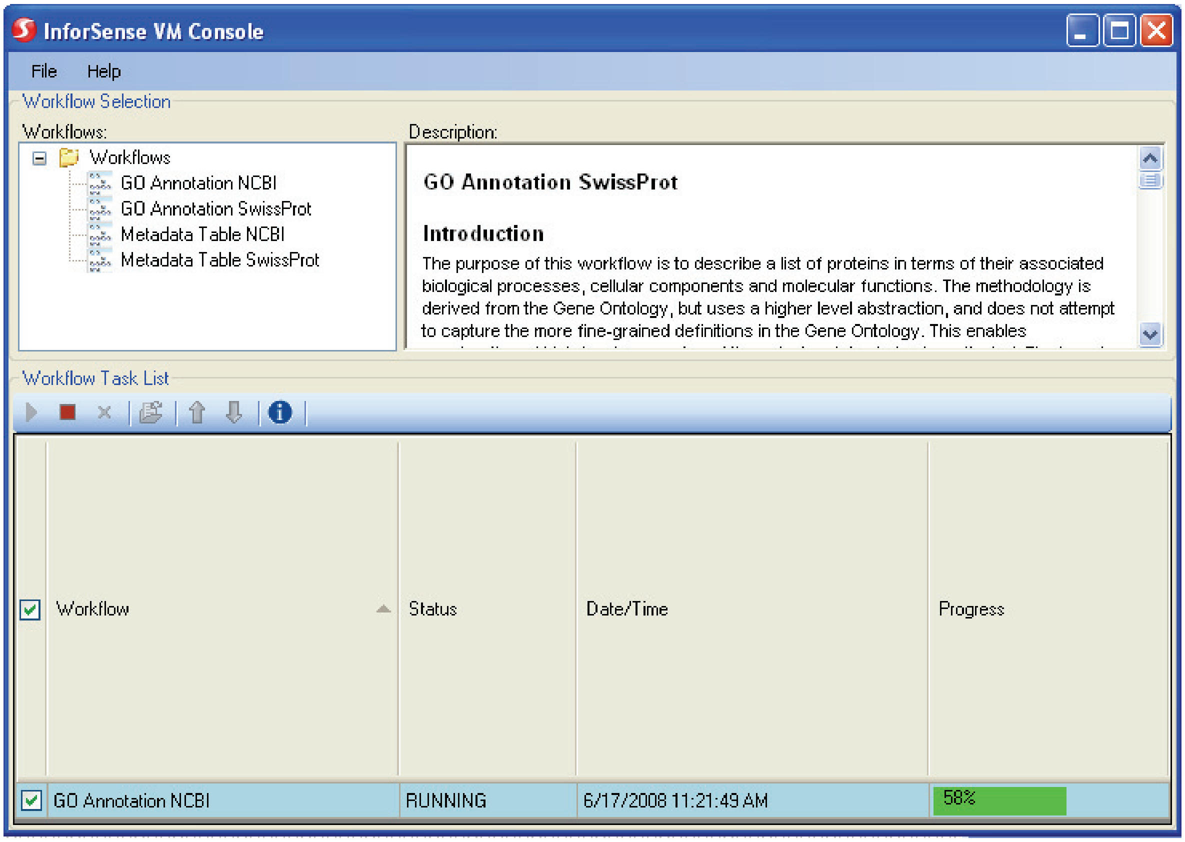
❖ To open InforSense

1. Open an .msf file.
2. Choose **Tools >InforSense**. A dialog box appears.
3. Select the criteria for your annotation project run.
4. Click **Invoke InforSense**. The InforSense VM Console appears.

5. In the Workflow pane, click **Workflow**. The Workflow begins running (see Figure 47).

   The bottom frame shows a completeness graph.

Figure 47. InforSense running outside of Proteome Discoverer



GO Annotation SwissProt

Introduction

The purpose of this workflow is to describe a list of proteins in terms of their associated biological processes, cellular components and molecular functions. The methodology is derived from the Gene Ontology, but uses a higher level abstraction, and does not attempt to capture the more fine-grained definitions in the Gene Ontology. This enables

☑	Workflow	Status	Date/Time	Progress
☑	GO Annotation NCBI	RUNNING	6/17/2008 11:21:49 AM	58%

IMPORTANT InforSense Protein Annotation comes with InforSense Workflow documentation available through the Help menu.

3

Interpreting Search Results

After Proteome Discoverer completes your search, it creates an .msf file. The .msf file contains data and results from your search. Open this .msf file so that you can access and interpret the search results. You can also display multiple search results within a single report.

The single or multiple results report displays a list of matching peptides and proteins identified by the search engine you specify. From your results report you can do the following:

- Sort and filter your data.
- Match likely proteins to your analyzed data.
- Access isotope, chromatogram, and other graphical views.
- Export your results, report, and views.
- Create reports to export into another file format for peptides and proteins identification or graphical display.

Contents

Working with the Initial Results Report

The results report is the main report that is initially displayed when you upload the search results (.msf file). From this report, click a protein or peptide row to navigate to other tabular and graphical views. The views provide detailed graphical information on your selected peptides. You can display more than one view to do a comparative analysis of your selected peptide or proteins. Use the shortcut menus to alter the details in your report or to copy information. To customize your report, you can move the columns to a different location in the report. You can also do a quick sort of your columns by clicking the column header.

Initial results report procedures:

- Organizing Rows and Columns
- Saving and Applying Results Report Layout Changes

Organizing Rows and Columns

In the initial results report, you can customize your table using these features:

- Grouping peptides in the protein or peptide tables
- Sorting columns by preference
- Adding row index numbering to help you sort

Adding, Removing, and Grouping Columns

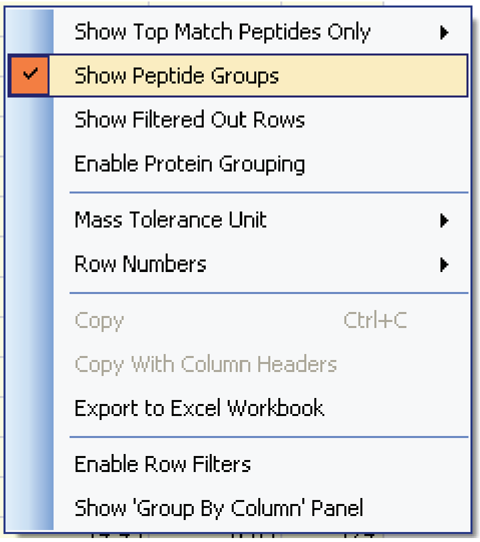
In the results report, you can group together peptide identifications with the same underlying sequence under the same collapsible heading. The individual identifications within the group might differ in detected modifications, activation type, and particular search engine result. Use this feature when viewing results from multiple searches in a single report. Entries within the peptide group should still display non-redundant (no duplicate) references.

Note If the Show Only Top Match Peptides option is active, Proteome Discoverer shows only the best matched peptides.

❖ To group by columns

1. In the Peptide or Protein table, right-click to display the shortcut menu (see Figure 48).
2. Choose **Show Peptides Group**. The table displays the peptides in a summary-type view, showing only the best matched peptides.

Figure 48. Shortcut menu



Sorting Columns

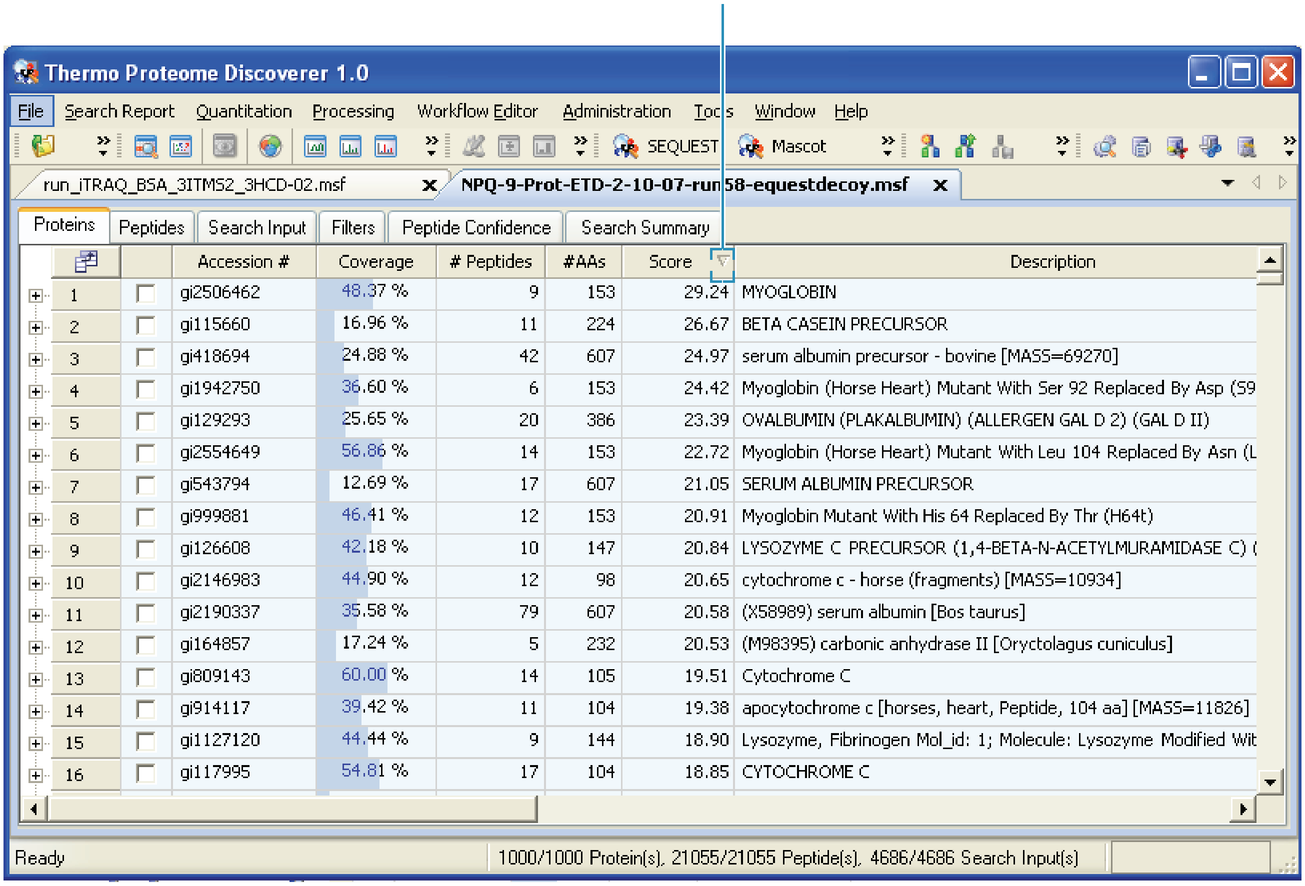
You can sort a column by ascending or descending order. The small triangle to the right of the column header shows if the column is in ascending or descending order.

❖ To sort a column

1. Open your search results. See "Understanding Reports and Views" on page 37.
2. Click the column header that you want to sort, as shown in Figure 49. You can also sort on multiple columns by holding down the CTRL key while clicking other column headers.

   The information in the column is sorted in descending order.

Figure 49. Column header sorting example


Column header sorted

		Accession #	Coverage	# Peptides	#AAs	Score	Description
1	☐	gi2506462	48.37 %	9	153	29.24	MYOGLOBIN
2	☐	gi115660	16.96 %	11	224	26.67	BETA CASEIN PRECURSOR
3	☐	gi418694	24.88 %	42	607	24.97	serum albumin precursor - bovine [MASS=69270]
4	☐	gi1942750	36.60 %	6	153	24.42	Myoglobin (Horse Heart) Mutant With Ser 92 Replaced By Asp (S9
5	☐	gi129293	25.65 %	20	386	23.39	OVALBUMIN (PLAKALBUMIN) (ALLERGEN GAL D 2) (GAL D II)
6	☐	gi2554649	56.86 %	14	153	22.72	Myoglobin (Horse Heart) Mutant With Leu 104 Replaced By Asn (L
7	☐	gi543794	12.69 %	17	607	21.05	SERUM ALBUMIN PRECURSOR
8	☐	gi999881	46.41 %	12	153	20.91	Myoglobin Mutant With His 64 Replaced By Thr (H64t)
9	☐	gi126608	42.18 %	10	147	20.84	LYSOZYME C PRECURSOR (1,4-BETA-N-ACETYLMURAMIDASE C) (
10	☐	gi2146983	44.90 %	12	98	20.65	cytochrome c - horse (fragments) [MASS=10934]
11	☐	gi2190337	35.58 %	79	607	20.58	(X58989) serum albumin [Bos taurus]
12	☐	gi164857	17.24 %	5	232	20.53	(M98395) carbonic anhydrase II [Oryctolagus cuniculus]
13	☐	gi809143	60.00 %	14	105	19.51	Cytochrome C
14	☐	gi914117	39.42 %	11	104	19.38	apocytochrome c [horses, heart, Peptide, 104 aa] [MASS=11826]
15	☐	gi1127120	44.44 %	9	144	18.90	Lysozyme, Fibrinogen Mol_id: 1; Molecule: Lysozyme Modified Wit
16	☐	gi117995	54.81 %	17	104	18.85	CYTOCHROME C

Ready 1000/1000 Protein(s), 21055/21055 Peptide(s), 4686/4686 Search Input(s)

3. Click the column header again to sort the information in ascending order.

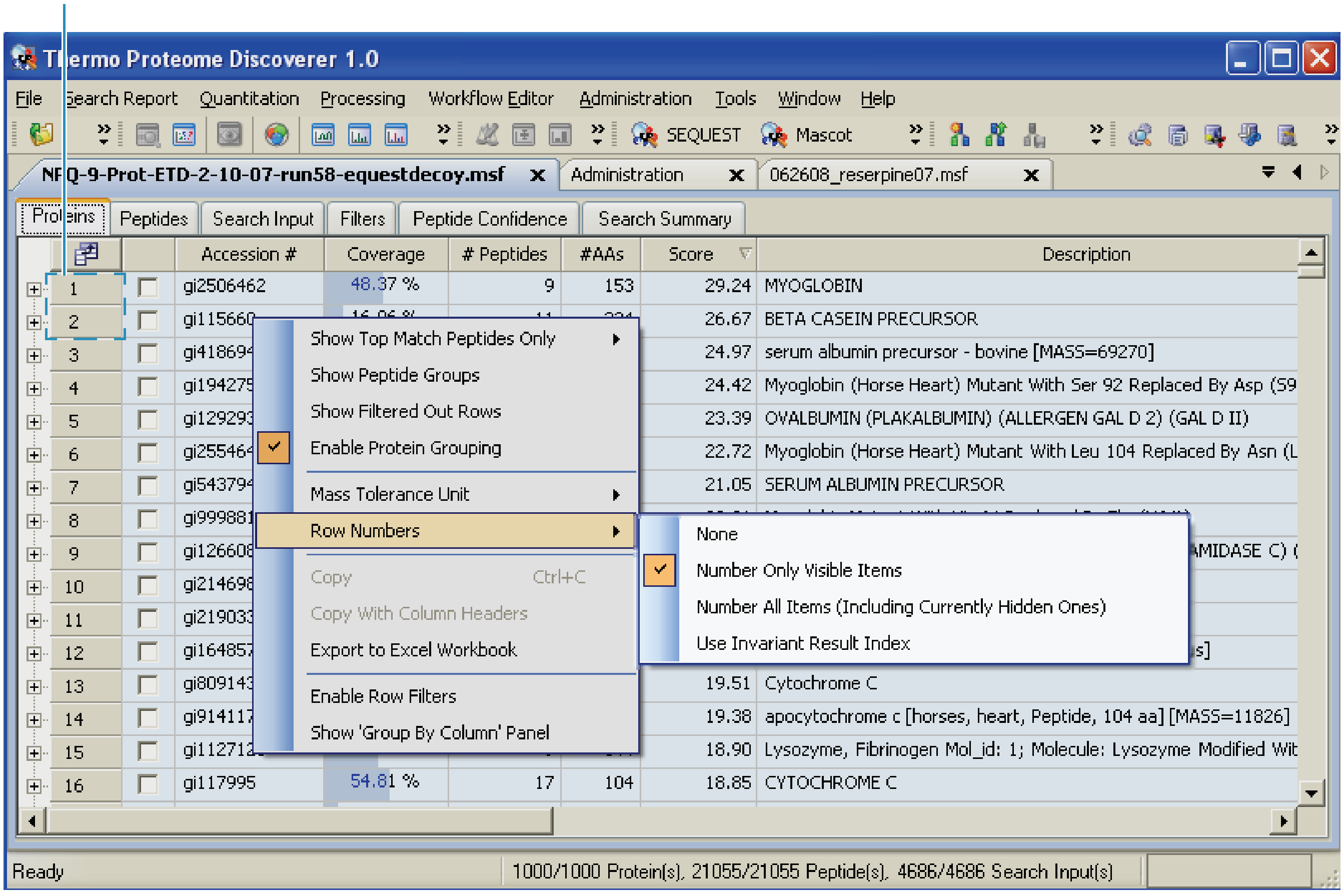
Labeling and Identifying Rows

❖ To add or remove row index numbers

1. Right-click the Peptides or Proteins table to display the shortcut menu.

2. Choose **Row Numbers**. A list of options is displayed.

Row number index



	Accession #	Coverage	# Peptides	#AAs	Score	Description
1	gi2506462	48.37 %	9	153	29.24	MYOGLOBIN
2	gi115660				26.67	BETA CASEIN PRECURSOR
3	gi418694				24.97	serum albumin precursor - bovine [MASS=69270]
4	gi194275				24.42	Myoglobin (Horse Heart) Mutant With Ser 92 Replaced By Asp (S9
5	gi129293				23.39	OVALBUMIN (PLAKALBUMIN) (ALLERGEN GAL D 2) (GAL D II)
6	gi255464				22.72	Myoglobin (Horse Heart) Mutant With Leu 104 Replaced By Asn (L
7	gi543794				21.05	SERUM ALBUMIN PRECURSOR
8	gi999881					
9	gi126608					AMIDASE C) (
10	gi214698					
11	gi219033					
12	gi164857					s]
13	gi809143				19.51	Cytochrome C
14	gi914117				19.38	apocytochrome c [horses, heart, Peptide, 104 aa] [MASS=11826]
15	gi112712				18.90	Lysozyme, Fibrinogen Mol_id: 1; Molecule: Lysozyme Modified Wit
16	gi117995	54.81 %	17	104	18.85	CYTOCHROME C

Ready | 1000/1000 Protein(s), 21055/21055 Peptide(s), 4686/4686 Search Input(s)

3. Select the row number style you want to use. Proteome Discoverer applies the row number option instantly.

4. (Optional) You can set the options as follows:

   - Number Only Visible Items—Numbers only visible rows from 1 to *N*.
   - Number All Items: (Including Currently Hidden Ones)—Numbers all rows from 1 to *N*.
   - Use Invariant Result Index—Numbers all rows by using an internal ID. (This option is the only numbering tag that remains on after sorting or filtering operations.)

Saving and Applying Results Report Layout Changes

Use the Layout menu to save layout changes, create a layout, and apply a layout to your results report. After you have used the Column Chooser options, you can save your changes to a default layout or a special layout for your results report.

❖ To define Column Chooser options

1. In the Protein or Peptide view, in the left corner, click . The Column Chooser appears.
2. Click the columns you want to view on the results page.

❖ To save a results report layout

1. Activate the modified results report with your preferred layout. See "Adding, Removing, and Grouping Columns" on page 78 and "Sorting Columns" on page 79.
2. Ensure the report rows are in the correct layout. See "Labeling and Identifying Rows" on page 81.
3. From the Proteome Discoverer menu, choose **Search Report > Layout > Save Layout As** to save the column and row changes. See Figure 50.

Figure 50. Save Layout As menu options

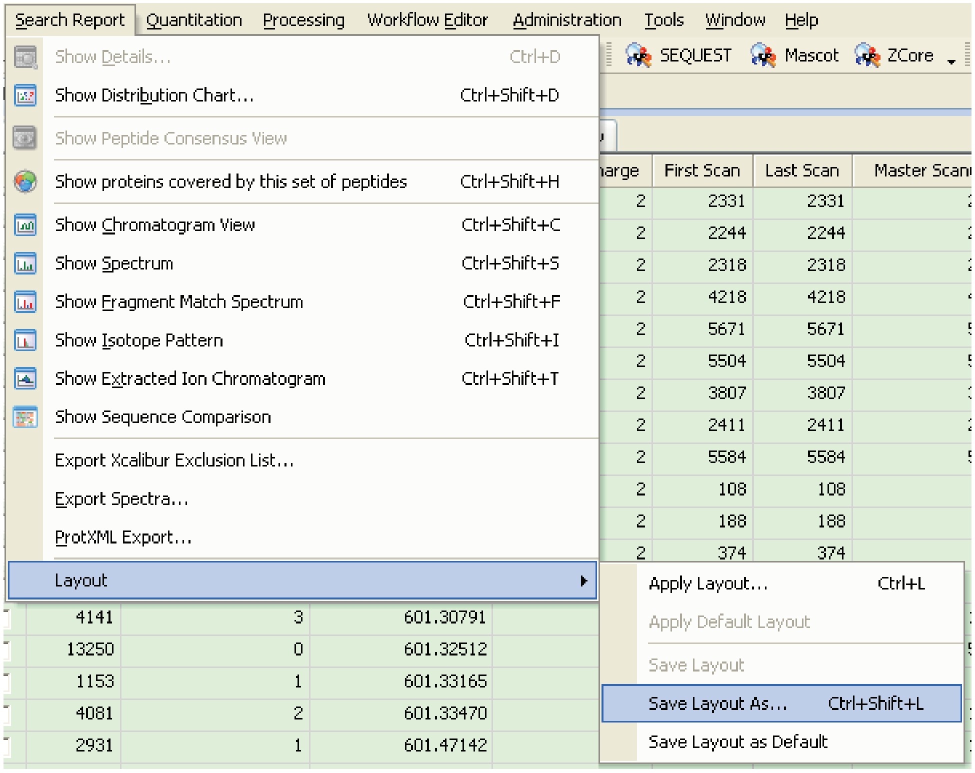
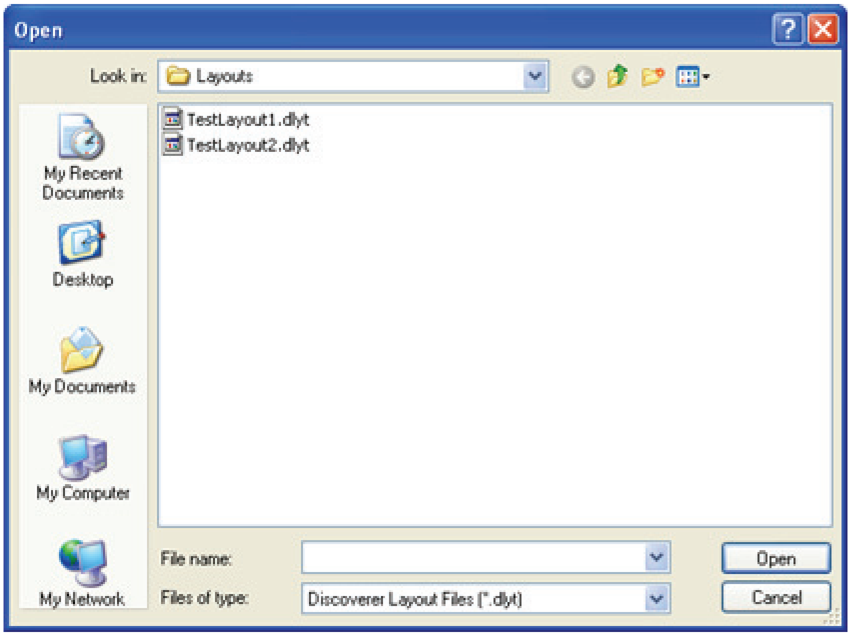


Figure 51. Saving a layout



Look in: Layouts

TestLayout1.dlyt
TestLayout2.dlyt

File name:

Files of type: Discoverer Layout Files (*.dlyt)

4. In the File name list, select or type the name of the layout.
5. Click **OK** to save the layout with the specified File name. The view closes and your layout properties are stored in a file in the layout folder.

❖ To create a default layout

1. Activate the modified results report with your preferred layout. See "Adding, Removing, and Grouping Columns" on page 78 and "Sorting Columns" on page 79.
2. From the Proteome Discoverer menu, choose **Search Report > Layout > Save Layout As Default** to save the column and row changes.

   You can now apply the row and column properties to any results report. To apply the default layout to any report, use the Apply Default Layout command.

❖ To apply the default layout to your report

1. Click the results report that you want to modify.
2. From the Proteome Discoverer menu, choose **Search Report > Layout > Apply Default Layout**.

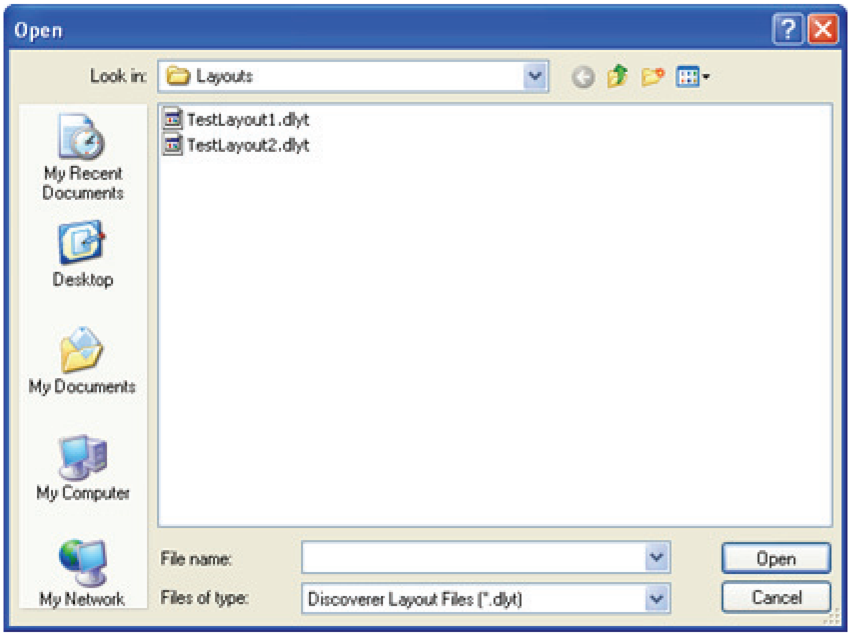
   The properties of the default layout are applied to the results report.

❖ To apply a layout to your report

1. Click the results report that you want to modify.
2. From the Proteome Discoverer menu, choose **Search Report > Layout > Apply Layout**. An Open dialog box appears as shown in Figure 52.

3. Select the layout you want to apply. The selected layout name appears in the File name box (see Figure 52).

Figure 52. Selecting a layout to apply



4. Click **Open** to apply the layout to your report. The properties of the selected layout are applied to the results report.

Working with Filters

Proteome Discoverer offers powerful capabilities for filtering your search results data by applying the available results filters on the Filters page. It also offers a quick method to filter your tabular search results. The quick filtering is called Row Filters and is accessed from the shortcut menu. These two different methods for filtering data, the results filters and the row filters, provide complementary options.

The filters are defined as follows:

- Results filters—Available on the Filters tab, these filters exclude peptides and proteins from the result set. As an effect, applying these results filters to filter out peptides changes the number of identified peptides and the percentage coverage values of the proteins. The numbers of filtered versus total number of peptides and proteins displayed in the status bar are also affected.
- Row filters—Display filters only. When displaying the filtered out rows, the affected lines for both filters are seen as grayed out rows. Excluding peptides by setting row filters does not change the number of identified peptides and the percentage coverage values of the proteins.

Use the filters feature to selectively hide and sort the visible results of the matched search results. Use the Filter page to separate the proteins and peptides based on the parameters selected from the Filters list. Use the Row Filters option with the Filters page feature to narrow your search results even further.

To filter and sort your results, you can apply any of these features:

- Apply a quick filter (see "Filtering Results" on page 85).
- Sort columns to organize the table (see "Filtering Results" on page 85).
- Add or remove filters (see "False Discovery Rates" on page 93).
- Activate more than one filter (see "False Discovery Rates" on page 93).
- Display the filtered row to do a visual check of the sorted results (see "Filtering Results" on page 85).

Filtering Results

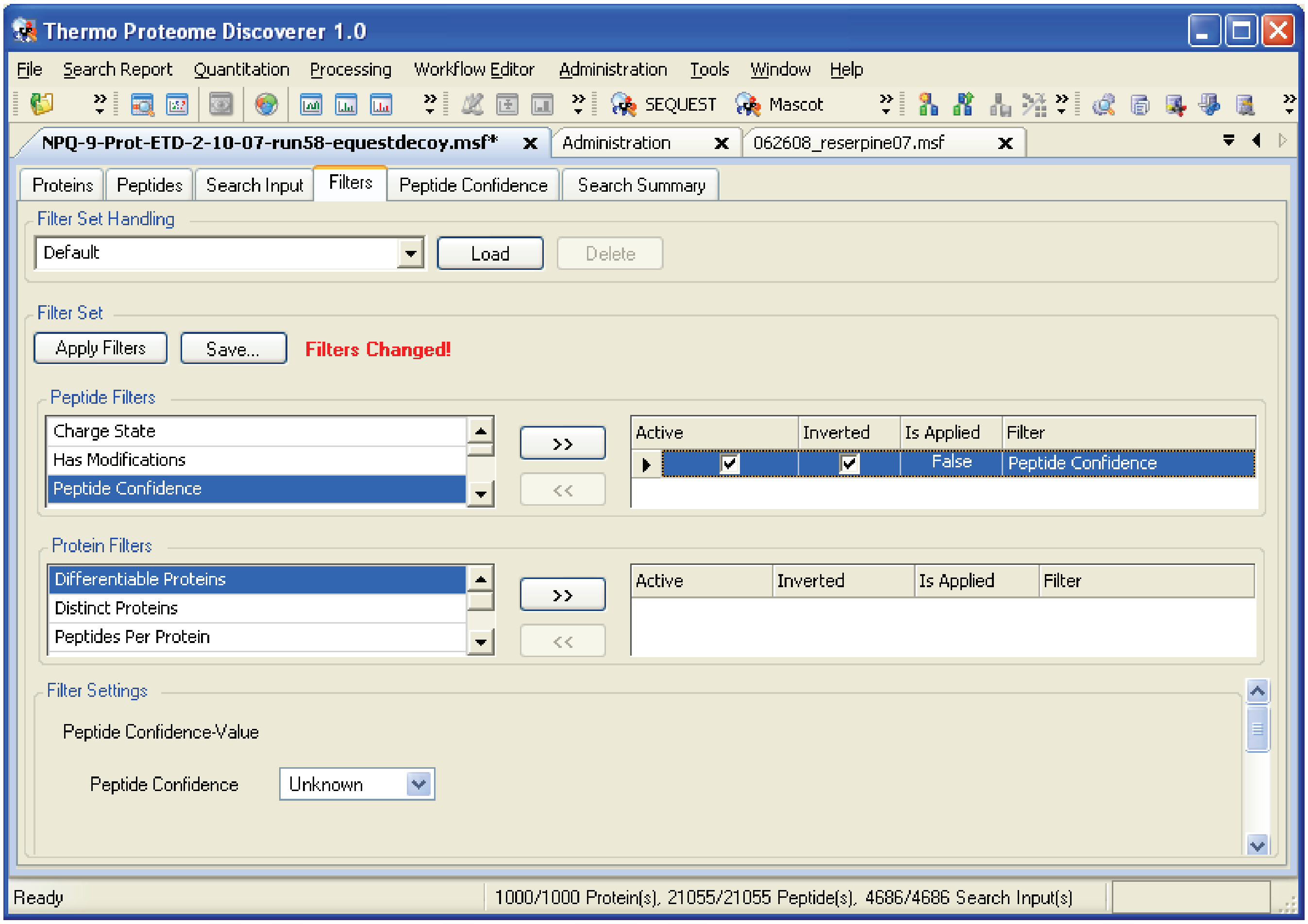
The following procedures describe how to filter your results using two filtering methods. If you save your report, you can save filters that you set using the Filters page with your results report. You cannot, however, save the filters you set using the Row Filters features with your results report. The Row Filters feature only works on the visible table.

❖ To filter your search results using the Filters page

1. Open your search results. See "Understanding Reports and Views" on page 37.
2. Click the tab of the search results you want to filter.
3. Click the **Filters** tab. The Filters page appears.
4. Select a filter type from either Filters list.
5. Click >>. The filter is included in the filter table. A message: Filters Changed! appears when a filter has been made active. (see Figure 53).
6. Select the check box to the left of the filter in the Active column.
7. (Optional) Select the check box to the left of the filter in the Inverted column.

> **Note** For an inverted filter example you set the peptide confidence level to low confidence and select the inverted option so that only the modest and high confidence levels show in the report.

Figure 53. Filters view



Filter Set Handling: Default

Filters Changed!

Peptide Filters:

- Charge State
- Has Modifications
- Peptide Confidence

Active	Inverted	Is Applied	Filter
☑	☑	False	Peptide Confidence

Protein Filters:

- Differentiable Proteins
- Distinct Proteins
- Peptides Per Protein

Active	Inverted	Is Applied	Filter

Filter Settings

Peptide Confidence-Value

Peptide Confidence: Unknown

Ready 1000/1000 Protein(s), 21055/21055 Peptide(s), 4686/4686 Search Input(s)

8. To update the results table, in the Filter Set area, click **Apply Filters**.

   The Is Applied column changes status from False to True as an indicator that the filter is applied to the active .msf file.

Note For an inverted filter example you set the peptide confidence level to low confidence and select the inverted option so that only the modest and high confidence levels show in the report.

Related Topics

- Removing and Deactivating Filters
- Quick Filtered Items

Removing and Deactivating Filters

Use the Filters page to set filter changes to the number of proteins and peptides visible in the Results grid. Add, remove, activate, or deactivate filters so that you can remove unneccessary information as you sort through the search results.

❖ To remove a filter

1. Open your search results. See "Understanding Reports and Views" on page 37.
2. Click the appropriate tab for your search results.
3. Click the **Filters** tab. The Filters page appears.
4. Click the filter on the filter table to highlight the row.
5. Click << . The filter is removed from the filter table. A message: Filters Changed! appears.
6. To update the results table, click **Apply Filters**.

❖ To deactivate a filter

1. Open your search results. See "Understanding Reports and Views" on page 37.
2. Click the appropriate tab for your search results.
3. Click the **Filters** tab. The Filters page appears.
4. Clear the check box to the left of the filter in the Active column.
5. To update the results table and disable the filter, click **Apply Filter**. The filter, however, is not removed from the Filters page.

Related Topics

- Filtering Results
- Quick Filtered Items

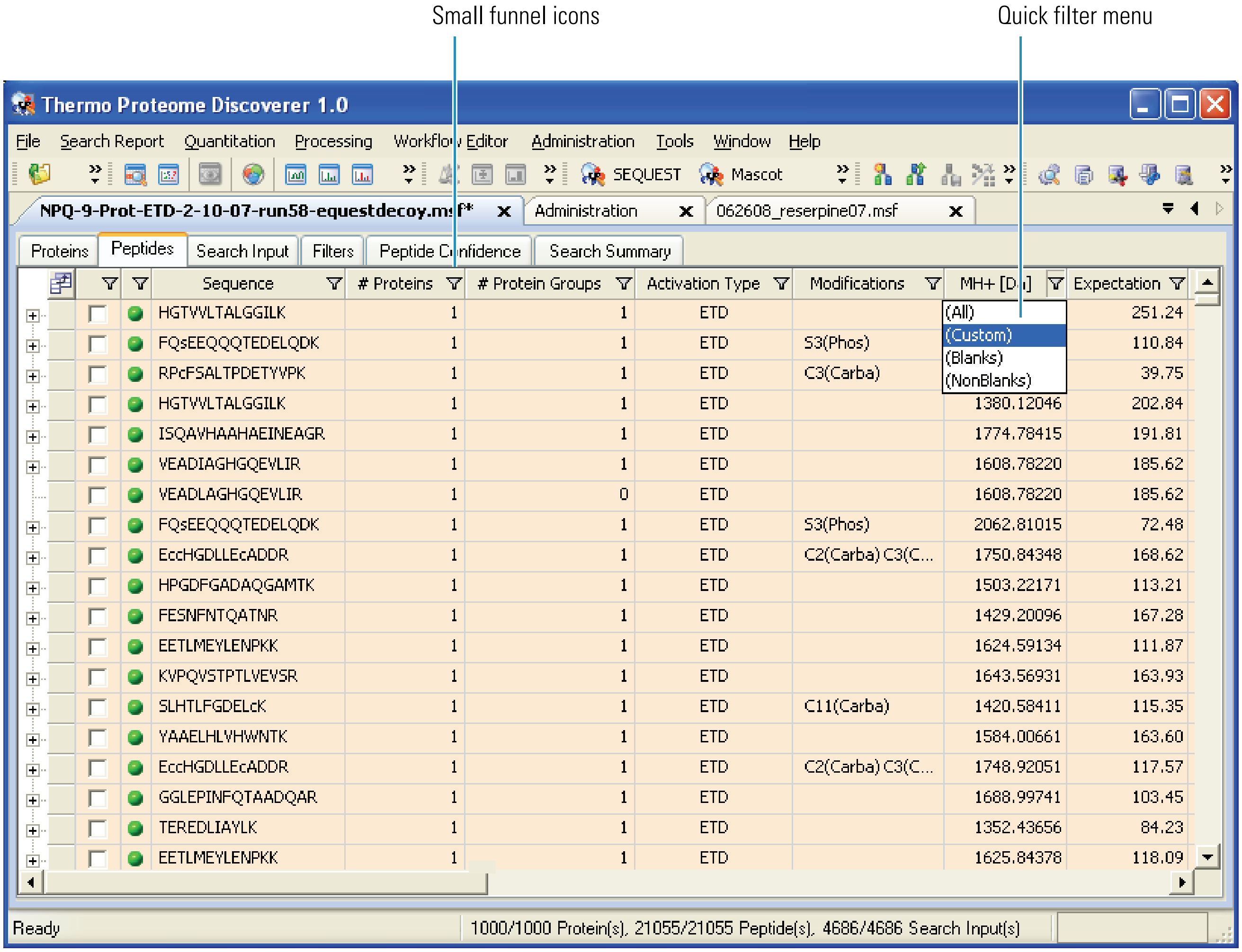
❖ To set a quick filter to your search results using row filters

1. Open your search results. See "Understanding Reports and Views" on page 37.

2. Select the Proteins or Peptides page. This example shows the Peptides page.

3. Right-click the table results to access the shortcut menu. Choose **Enable Row Filters**. Small funnel icons appear to the right of the column headers as seen in Figure 54.

Figure 54. Funnel filter menu



		Sequence	# Proteins	# Protein Groups	Activation Type	Modifications	MH+ [Da]	Expectation
☐	●	HGTVVLTALGGILK	1	1	ETD			251.24
☐	●	FQsEEQQQTEDELQDK	1	1	ETD	S3(Phos)		110.84
☐	●	RPcFSALTPDETYVPK	1	1	ETD	C3(Carba)		39.75
☐	●	HGTVVLTALGGILK	1	1	ETD		1380.12046	202.84
☐	●	ISQAVHAAHAEINEAGR	1	1	ETD		1774.78415	191.81
☐	●	VEADIAGHGQEVLIR	1	1	ETD		1608.78220	185.62
☐	●	VEADLAGHGQEVLIR	1	0	ETD		1608.78220	185.62
☐	●	FQsEEQQQTEDELQDK	1	1	ETD	S3(Phos)	2062.81015	72.48
☐	●	EccHGDLLEcADDR	1	1	ETD	C2(Carba) C3(C...	1750.84348	168.62
☐	●	HPGDFGADAQGAMTK	1	1	ETD		1503.22171	113.21
☐	●	FESNFNTQATNR	1	1	ETD		1429.20096	167.28
☐	●	EETLMEYLENPKK	1	1	ETD		1624.59134	111.87
☐	●	KVPQVSTPTLVEVSR	1	1	ETD		1643.56931	163.93
☐	●	SLHTLFGDELcK	1	1	ETD	C11(Carba)	1420.58411	115.35
☐	●	YAAELHLVHWNTK	1	1	ETD		1584.00661	163.60
☐	●	EccHGDLLEcADDR	1	1	ETD	C2(Carba) C3(C...	1748.92051	117.57
☐	●	GGLEPINFQTAADQAR	1	1	ETD		1688.99741	103.45
☐	●	TEREDLIAYLK	1	1	ETD		1352.43656	84.23
☐	●	EETLMEYLENPKK	1	1	ETD		1625.84378	118.09

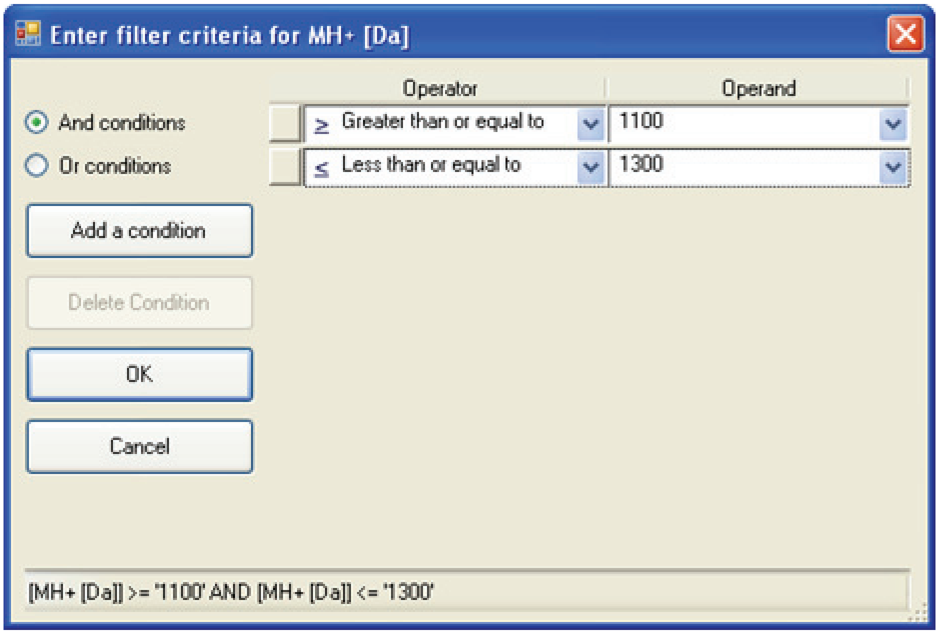
Ready

1000/1000 Protein(s), 21055/21055 Peptide(s), 4686/4686 Search Input(s)

4. Click a funnel icon. A menu list opens for you to set a row filter for that particular column. In this example, choose **Custom**. The dialog box to enter filter criteria for a particular view appears (see Figure 55).

Note The following steps provide specific details on how to set a precursor mass filter.

Figure 55. Enter filter criteria for MH+ [Da] example



● And conditions
○ Or conditions

Operator	Operand
≥ Greater than or equal to	1100
≤ Less than or equal to	1300

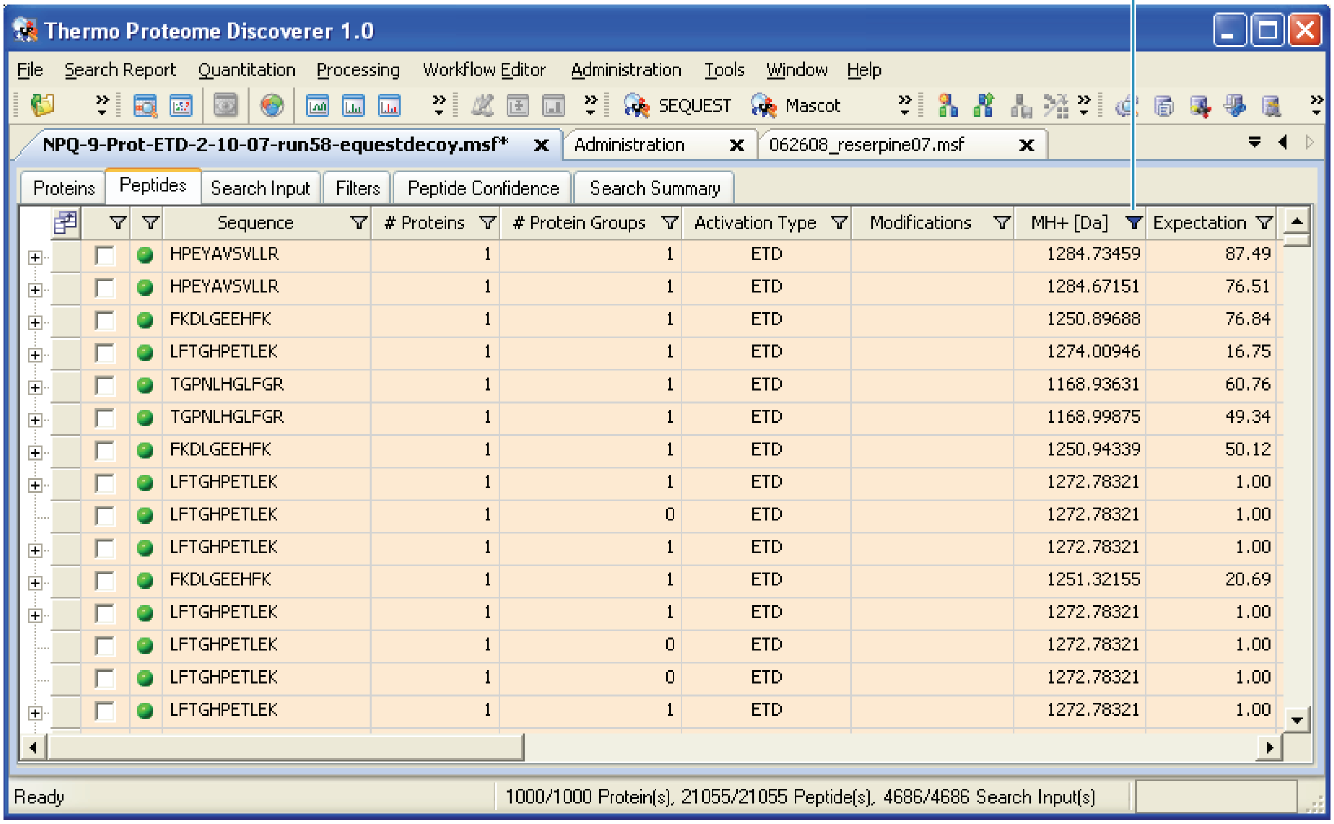
[MH+ [Da]] >= '1100' AND [MH+ [Da]] <= '1300'

5. From the Operator list, select **Greater than or equal to**.
6. In the Operand box, type **1100**.
7. To display another row in the Enter filter criteria dialog box, click **Add a condition**.
8. From the next Operator list, select **Less than or equal to**.
9. In the Operand box, type **1300**. The Enter filter criteria dialog box should look like Figure 55.
10. Click **OK** to accept the filter settings. For this example, only peptides that have a precursor MH+ mass between 1100 and 1300 are displayed.

Where this type of filter is active, the color of the column funnel icon is blue as shown in Figure 56.

Figure 56. Blue-colored funnel icon indicating an active column

Blue funnel icon



		Sequence	# Proteins	# Protein Groups	Activation Type	Modifications	MH+ [Da]	Expectation
☐	●	HPEYAVSVLLR	1	1	ETD		1284.73459	87.49
☐	●	HPEYAVSVLLR	1	1	ETD		1284.67151	76.51
☐	●	FKDLGEEHFK	1	1	ETD		1250.89688	76.84
☐	●	LFTGHPETLEK	1	1	ETD		1274.00946	16.75
☐	●	TGPNLHGLFGR	1	1	ETD		1168.93631	60.76
☐	●	TGPNLHGLFGR	1	1	ETD		1168.99875	49.34
☐	●	FKDLGEEHFK	1	1	ETD		1250.94339	50.12
☐	●	LFTGHPETLEK	1	1	ETD		1272.78321	1.00
☐	●	LFTGHPETLEK	1	0	ETD		1272.78321	1.00
☐	●	LFTGHPETLEK	1	1	ETD		1272.78321	1.00
☐	●	FKDLGEEHFK	1	1	ETD		1251.32155	20.69
☐	●	LFTGHPETLEK	1	1	ETD		1272.78321	1.00
☐	●	LFTGHPETLEK	1	0	ETD		1272.78321	1.00
☐	●	LFTGHPETLEK	1	0	ETD		1272.78321	1.00
☐	●	LFTGHPETLEK	1	1	ETD		1272.78321	1.00

Ready

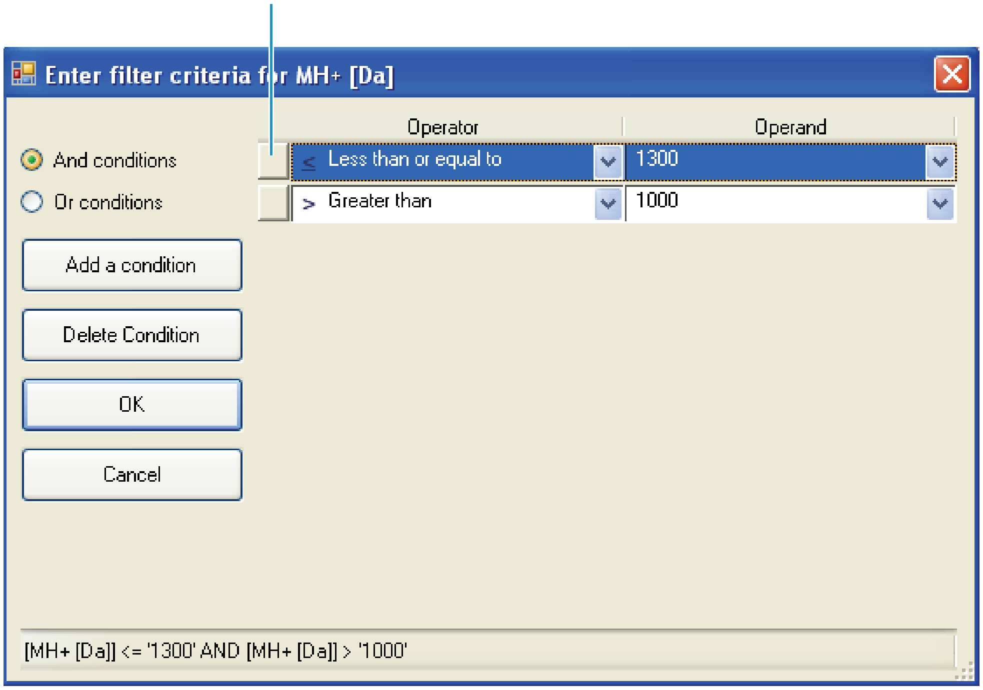
1000/1000 Protein(s), 21055/21055 Peptide(s), 4686/4686 Search Input(s)

❖ To remove a quick filter

1. Open your search results. See "Understanding Reports and Views" on page 37.
2. Select the Proteins or Peptides page with the quick filter set.
3. Click the blue funnel icon. A menu list opens for you to set a row filter for that particular column. In this example, choose **Custom**. The dialog box to enter filter criteria for a particular view appears.
4. In the Operand row of interest, click the condition button to the left of the Operator list. The condition is activated (see Figure 57).

Figure 57. Delete filter criteria for MH+ [Da] example

Condition button



Operator	Operand
≤ Less than or equal to	1300
> Greater than	1000

[MH+ [Da]] <= '1300' AND [MH+ [Da]] > '1000'

5. Click **Delete Condition**.
6. Click **OK**.

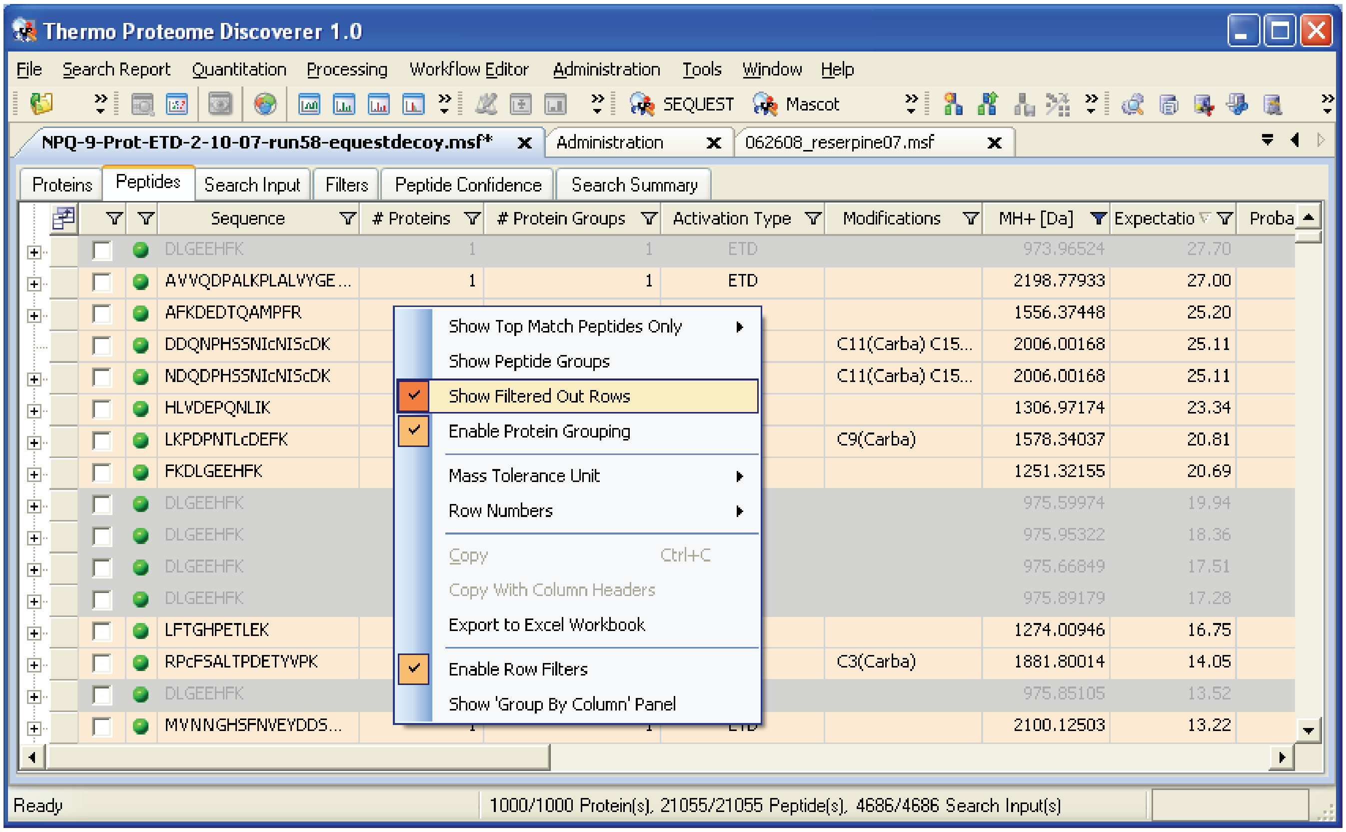
Quick Filtered Items

Quick filters hide rows in the tabular reports so that you can easily assess your results. By using the Show Filtered Out Rows command in the shortcut menu, you can see the hidden rows to do a comparative analysis. The hidden rows are grayed out when you choose Show Filtered Out Rows.

❖ To display quick filtered rows

1. Open your search results. See "Understanding Reports and Views" on page 37.
2. Select the Proteins or Peptides page. This example shows the Peptides page.
3. Click the **Peptides** tab. The Peptides page appears.
4. Right-click to display the shortcut menu as shown in Figure 58, and choose **Show Filtered Out Rows**. The grid changes to show filtered and unfiltered rows.

Figure 58. Showing quick filtered out rows



Sequence	# Proteins	# Protein Groups	Activation Type	Modifications	MH+ [Da]	Expectatio	Proba
DLGEEHFK	1	1	ETD		973.96524	27.70	
AVVQDPALKPLALVYGE...	1	1	ETD		2198.77933	27.00	
AFKDEDTQAMPFR					1556.37448	25.20	
DDQNPHSSNIcNIScDK				C11(Carba) C15...	2006.00168	25.11	
NDQDPHSSNIcNIScDK				C11(Carba) C15...	2006.00168	25.11	
HLVDEPQNLIK					1306.97174	23.34	
LKPDPNTLcDEFK				C9(Carba)	1578.34037	20.81	
FKDLGEEHFK					1251.32155	20.69	
DLGEEHFK					975.59974	19.94	
DLGEEHFK					975.95322	18.36	
DLGEEHFK					975.66849	17.51	
DLGEEHFK					975.89179	17.28	
LFTGHPETLEK					1274.00946	16.75	
RPcFSALTPDETYVPK				C3(Carba)	1881.80014	14.05	
DLGEEHFK					975.85105	13.52	
MVNNGHSFNVEYDDS...	1	1			2100.12503	13.22	

Ready | 1000/1000 Protein(s), 21055/21055 Peptide(s), 4686/4686 Search Input(s)

Related Topics

- Filtering Results
- Removing and Deactivating Filters

False Discovery Rates

The false discovery rate (or false positive rate) is a statistical value that estimates the number of false positive identifications among all identifications found by a peptide ID search. It is a measure of the certainty of the identification. This topic describes how to determine false discovery rates with Proteome Discoverer by using the decoy search feature.

A good decoy database should contain entries that look like real proteins, but do not contain genuine peptide sequences. The simplest approach to achieving such a decoy database is to reverse all protein sequences, which is the scheme currently used in Proteome Discoverer. 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. This is especially true with dynamic modifications, because it is possible to get 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.

There are two ways to perform the decoy database search:

- (more conservative approach) 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.
- 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 thinking of a case where two significant matches are found for a given spectrum. The first match is from the non-decoy database and the second one from the decoy database. As only the top matches are considered when calculating the false discovery rates, this is not considered as a false positive in the concatenated database approach, whereas it would count in the separate databases approach. This is why the latter case is considered the more conservative one, and this is the approach that is currently used in Proteome Discoverer.

How to Calculate False Discovery Rates

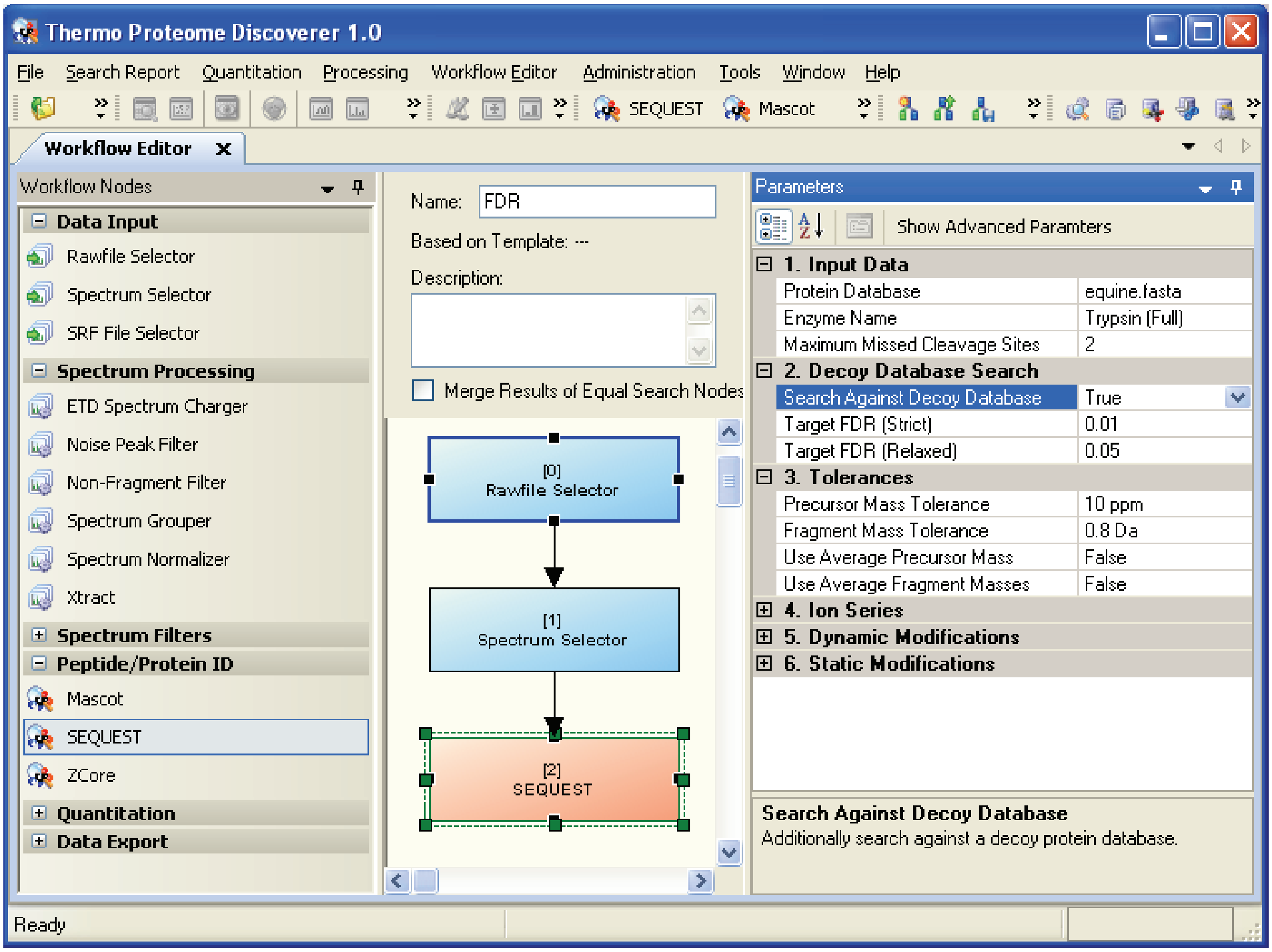
To calculate the false discovery rate, the matches are counted that pass a given set of filter thresholds from the decoy database and from the non-decoy database. Proteome Discoverer counts only the top match per spectrum, assuming that for any given spectrum only one peptide can be the correct hit.

What are Target False Discovery Rates (FDRs)?

If a false discovery rate target value is set for a decoy database search, Proteome Discoverer determines and applies filter thresholds to identified matches, such that the resulting false discovery rate is not higher than the set target value. The confidence indicators applied to each peptide match are distributed according to these calculated filter thresholds (see "False Discovery Rates and Peptide Confidence Indicators" on page 97).

You must specify two target values for a decoy database search: a strict target FDR and a more relaxed FDR. Figure 59 shows the decoy search setting with target false discovery rates 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.

Figure 59. Decoy Database Search in Workflow Editor



Name: FDR

Based on Template: ---

Description:

☐ Merge Results of Equal Search Nodes

1. Input Data	
Protein Database	equine.fasta
Enzyme Name	Trypsin (Full)
Maximum Missed Cleavage Sites	2
2. Decoy Database Search	
Search Against Decoy Database	True
Target FDR (Strict)	0.01
Target FDR (Relaxed)	0.05
3. Tolerances	
Precursor Mass Tolerance	10 ppm
Fragment Mass Tolerance	0.8 Da
Use Average Precursor Mass	False
Use Average Fragment Masses	False
4. Ion Series	
5. Dynamic Modifications	
6. Static Modifications	

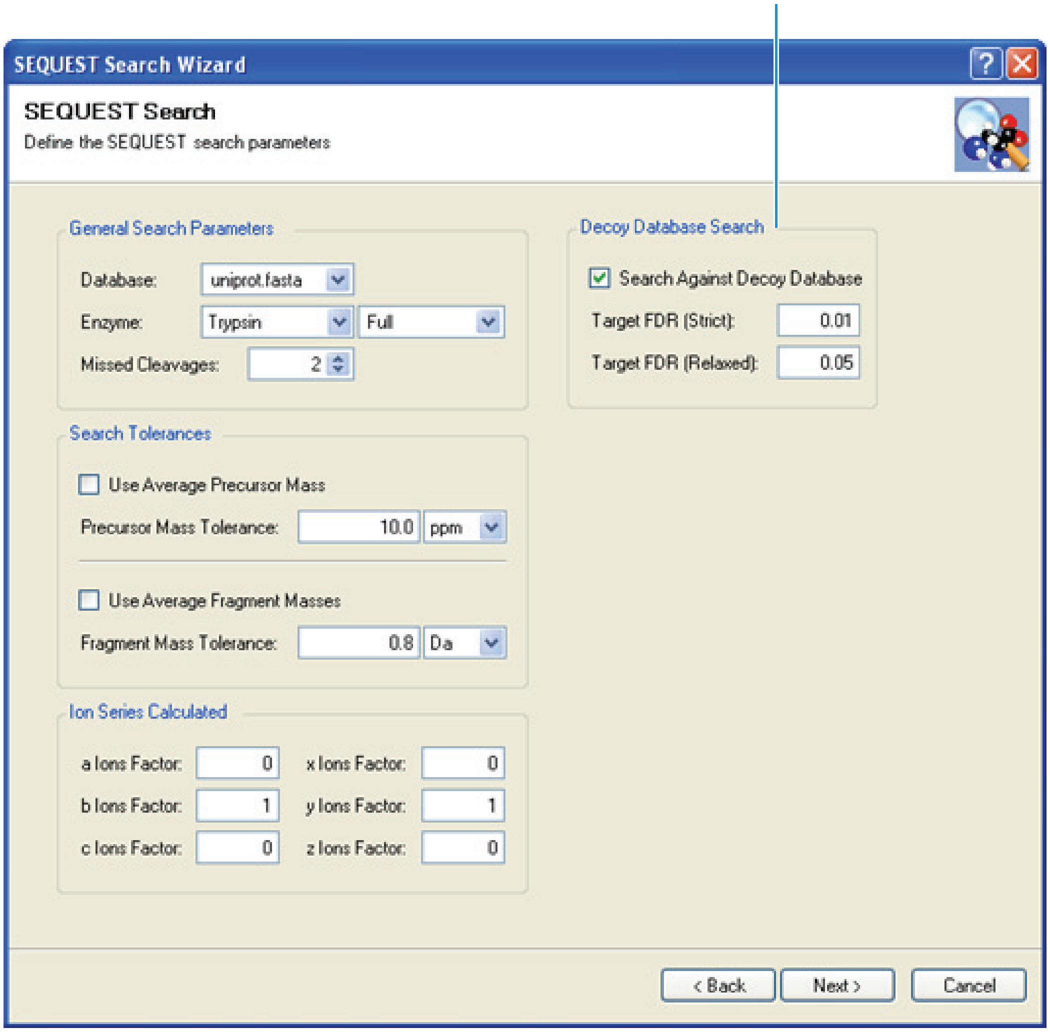
Search Against Decoy Database
Additionally search against a decoy protein database.

Determining False Discovery Rates

With Proteome Discoverer, you can determine the false discovery rate for every available search engine. If the search is set up through a search wizard, there is a single check box on the search parameter page (Figure 60) to enable automatic decoy database searching.

Figure 60. Enabling automatic decoy database searching through the search wizard

Setting up a decoy database search



SEQUEST Search Wizard

SEQUEST Search

Define the SEQUEST search parameters

General Search Parameters

Database: uniprot.fasta

Enzyme: Trypsin Full

Missed Cleavages: 2

Decoy Database Search

☑ Search Against Decoy Database

Target FDR (Strict): 0.01

Target FDR (Relaxed): 0.05

Search Tolerances

☐ Use Average Precursor Mass

Precursor Mass Tolerance: 10.0 ppm

☐ Use Average Fragment Masses

Fragment Mass Tolerance: 0.8 Da

Ion Series Calculated

a Ions Factor: 0

b Ions Factor: 1

c Ions Factor: 0

x Ions Factor: 0

y Ions Factor: 1

z Ions Factor: 0

When setting up the search with the Workflow Editor, you can find the options to do an automatic decoy database search in the search nodes under the Decoy Database Search pane (Figure 59).

After Proteome Discoverer completes the search and opens a result file (.msf), you can find the decoy database search result on the Peptide Confidence page (Figure 61). This page shows the two false discovery rates (relaxed and strict) with their corresponding filter settings listed above them.

Use the Peptide Confidence page to do the following:

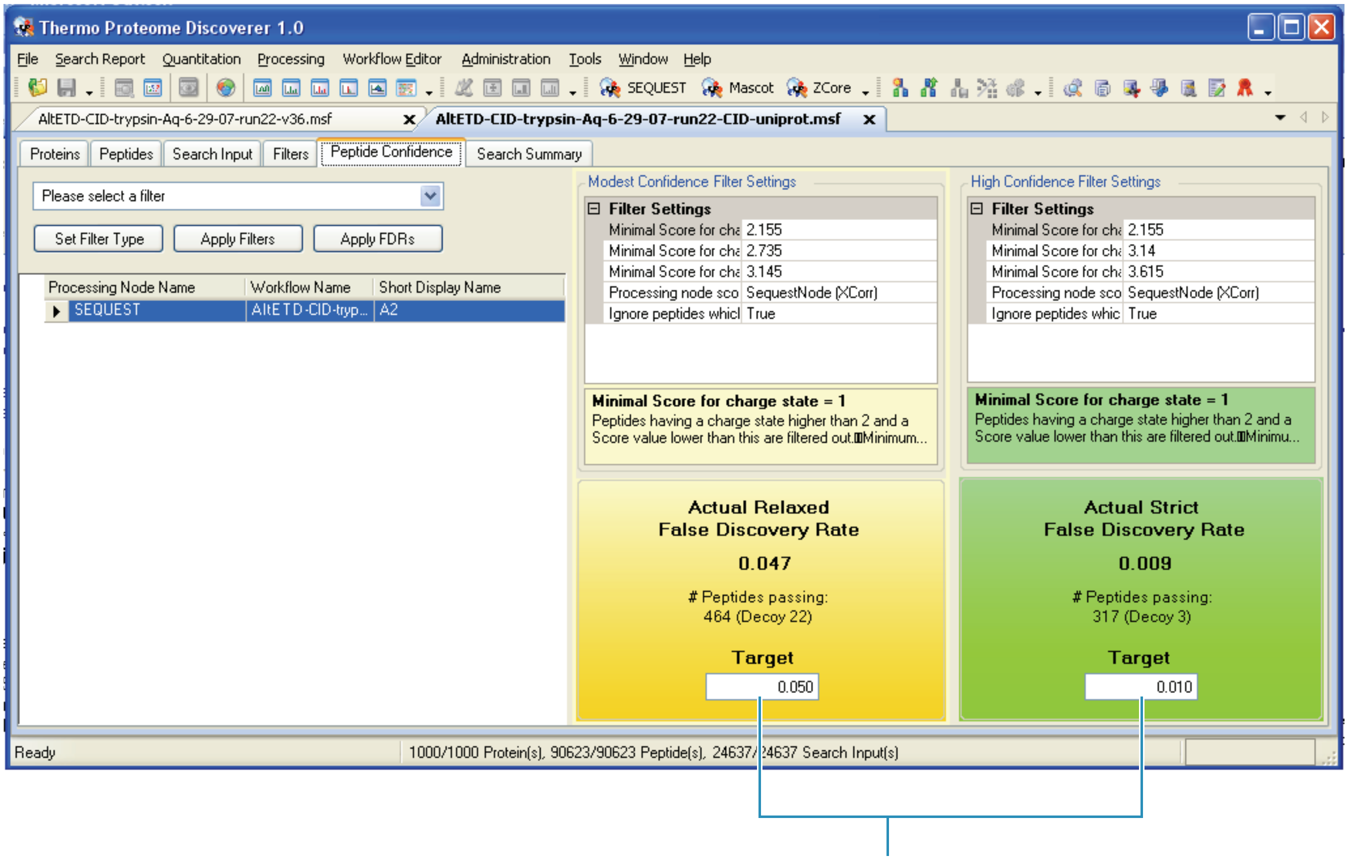
- Set new filters and recalculate new false discovery rates based on these new filter criteria.
- Set new target false discovery rates and then recalculate new filter settings that, when applied, lead to false discovery rates not worse than the specified ones.

❖ To recalculate the false discovery rates

1. Open a results page and click the **Peptide Confidence** tab.
2. Change the filter **Target** setting.
3. Recalculate the false discovery rates, and click **Apply Filters**.
4. Apply the new filter settings, and click **Apply FDRs**.

> **Note** You can change the default confidence levels to alternative values from the Peptide Confidence page.

Figure 61. Peptide Confidence view with the actual relaxed and strict false discovery rates given the current filter settings



Processing Node Name	Workflow Name	Short Display Name
SEQUEST	AltETD-CID-tryp...	A2

Modest Confidence Filter Settings

Filter Settings	
Minimal Score for cha	2.155
Minimal Score for cha	2.735
Minimal Score for cha	3.145
Processing node sco	SequestNode (XCorr)
Ignore peptides whicl	True

Minimal Score for charge state = 1
Peptides having a charge state higher than 2 and a Score value lower than this are filtered out.Minimum...

High Confidence Filter Settings

Filter Settings	
Minimal Score for cha	2.155
Minimal Score for cha	3.14
Minimal Score for cha	3.615
Processing node sco	SequestNode (XCorr)
Ignore peptides whic	True

Minimal Score for charge state = 1
Peptides having a charge state higher than 2 and a Score value lower than this are filtered out.Minimu...

Actual Relaxed
False Discovery Rate
0.047
Peptides passing:
464 (Decoy 22)
Target
0.050

Actual Strict
False Discovery Rate
0.009
Peptides passing:
317 (Decoy 3)
Target
0.010

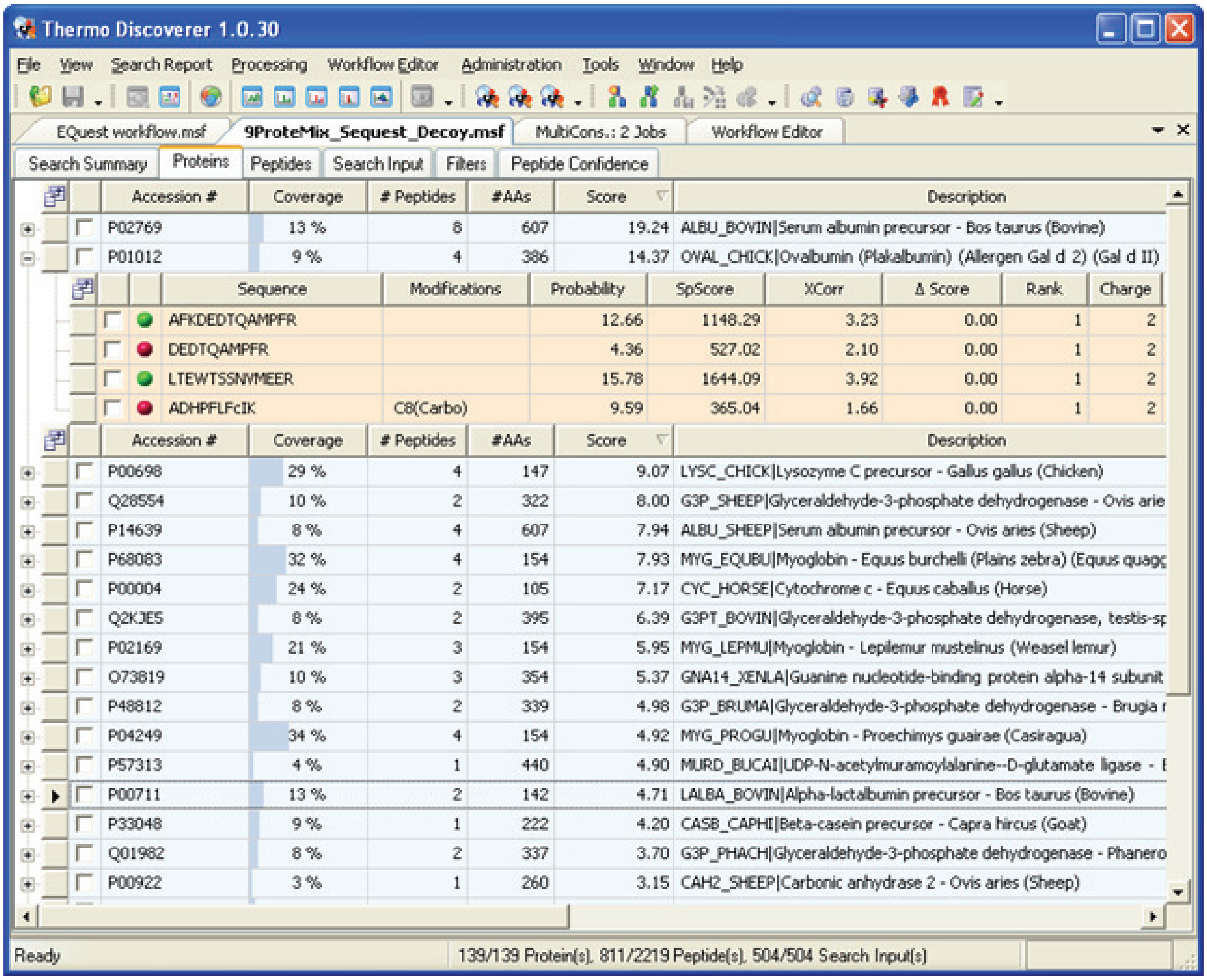
Ready
1000/1000 Protein(s), 90623/90623 Peptide(s), 24637/24637 Search Input(s)

Filter target setting

False Discovery Rates and Peptide Confidence Indicators

In Proteome Discoverer, the filter settings that determine false discovery rates are used to distribute the confidence indicators for the peptide matches: the green, yellow, and red dots attached to each peptide match (see Figure 62). Whenever a decoy database search is performed and filter settings are applied 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 indicator of high confidence, peptide matches that pass the filter associated with the relaxed FDR are assigned a yellow indicator of modest confidence, and all other peptide matches receive a red indicator of low confidence.

Figure 62. Decoy search results



	Accession #	Coverage	# Peptides	#AAs	Score	Description
☐	P02769	13 %	8	607	19.24	ALBU_BOVIN|Serum albumin precursor - Bos taurus (Bovine)
☐	P01012	9 %	4	386	14.37	OVAL_CHICK|Ovalbumin (Plakalbumin) (Allergen Gal d 2) (Gal d II)

		Sequence	Modifications	Probability	SpScore	XCorr	Δ Score	Rank	Charge
☐	🟢	AFKDEDTQAMPFR		12.66	1148.29	3.23	0.00	1	2
☐	🔴	DEDTQAMPFR		4.36	527.02	2.10	0.00	1	2
☐	🟢	LTEWTSSNVMEER		15.78	1644.09	3.92	0.00	1	2
☐	🔴	ADHPFLFcIK	C8(Carbo)	9.59	365.04	1.66	0.00	1	2

	Accession #	Coverage	# Peptides	#AAs	Score	Description
☐	P00698	29 %	4	147	9.07	LYSC_CHICK|Lysozyme C precursor - Gallus gallus (Chicken)
☐	Q28554	10 %	2	322	8.00	G3P_SHEEP|Glyceraldehyde-3-phosphate dehydrogenase - Ovis arie
☐	P14639	8 %	4	607	7.94	ALBU_SHEEP|Serum albumin precursor - Ovis aries (Sheep)
☐	P68083	32 %	4	154	7.93	MYG_EQUBU|Myoglobin - Equus burchelli (Plains zebra) (Equus quag
☐	P00004	24 %	2	105	7.17	CYC_HORSE|Cytochrome c - Equus caballus (Horse)
☐	Q2KJE5	8 %	2	395	6.39	G3PT_BOVIN|Glyceraldehyde-3-phosphate dehydrogenase, testis-sp
☐	P02169	21 %	3	154	5.95	MYG_LEPMU|Myoglobin - Lepilemur mustelinus (Weasel lemur)
☐	O73819	10 %	3	354	5.37	GNA14_XENLA|Guanine nucleotide-binding protein alpha-14 subunit
☐	P48812	8 %	2	339	4.98	G3P_BRUMA|Glyceraldehyde-3-phosphate dehydrogenase - Brugia
☐	P04249	34 %	4	154	4.92	MYG_PROGU|Myoglobin - Proechimys guairae (Casiragua)
☐	P57313	4 %	1	440	4.90	MURD_BUCAI|UDP-N-acetylmuramoylalanine--D-glutamate ligase -
☐	P00711	13 %	2	142	4.71	LALBA_BOVIN|Alpha-lactalbumin precursor - Bos taurus (Bovine)
☐	P33048	9 %	1	222	4.20	CASB_CAPHI|Beta-casein precursor - Capra hircus (Goat)
☐	Q01982	8 %	2	337	3.70	G3P_PHACH|Glyceraldehyde-3-phosphate dehydrogenase - Phanero
☐	P00922	3 %	1	260	3.15	CAH2_SHEEP|Carbonic anhydrase 2 - Ovis aries (Sheep)

Ready

139/139 Protein(s), 811/2219 Peptide(s), 504/504 Search Input(s)

Working with the Proteins Grid

Use the Proteins grid tab to display the list of identified proteins with the associated identified peptides. From this page you can examine the search results in terms of protein identification, as well as access more details about the peptide identifications and corresponding information from the search input.

You can combine the Chromatogram view with other views such as the extracted ion chromatogram (XIC). The following are examples of accessible views:

- Protein identification details view (protein sequence coverage map)
- Fragment match chart
- Peptide ID details view (sub-level view on the protein page)
- Chromatogram view
- XIC view (reconstructed ion chromatogram of precursor)
- Isotope view (full mass spectrum of precursor *m/z*)

For further discussion regarding the protein grid page, see the following:

- Researching Groups of Proteins
- Interpreting Your Results with the Chromatogram View
- Interpreting Your Results with the Protein Identification Details View

Researching Groups of Proteins

Although MS/MS-based proteomics studies are peptide-centric, you can also explore what proteins are present and their associations through related peptides. Deducing protein identities from a set of identified peptides becomes difficult due to sequence redundancy, such as the presence of proteins that have shared peptides. These redundant proteins are automatically grouped and are not initially displayed in the results report.

In the search results report, you can turn on or off protein grouping. However, protein grouping is an essential feature for the quantitation to help determine peptide uniqueness. By default, the Show Only Top Match Peptides option is set to off, so you can see all peptides considered for calculating protein ratios.

The proteins that are not distinct or differentiable are not displayed. The proteins within a group are ranked according to their protein score. If they have the same score, they are ranked by their sequence coverage. The top protein of a group becomes the master protein of that group. By default, only the master proteins are displayed on the main Proteins page.

❖ To display other proteins belonging to the same protein group

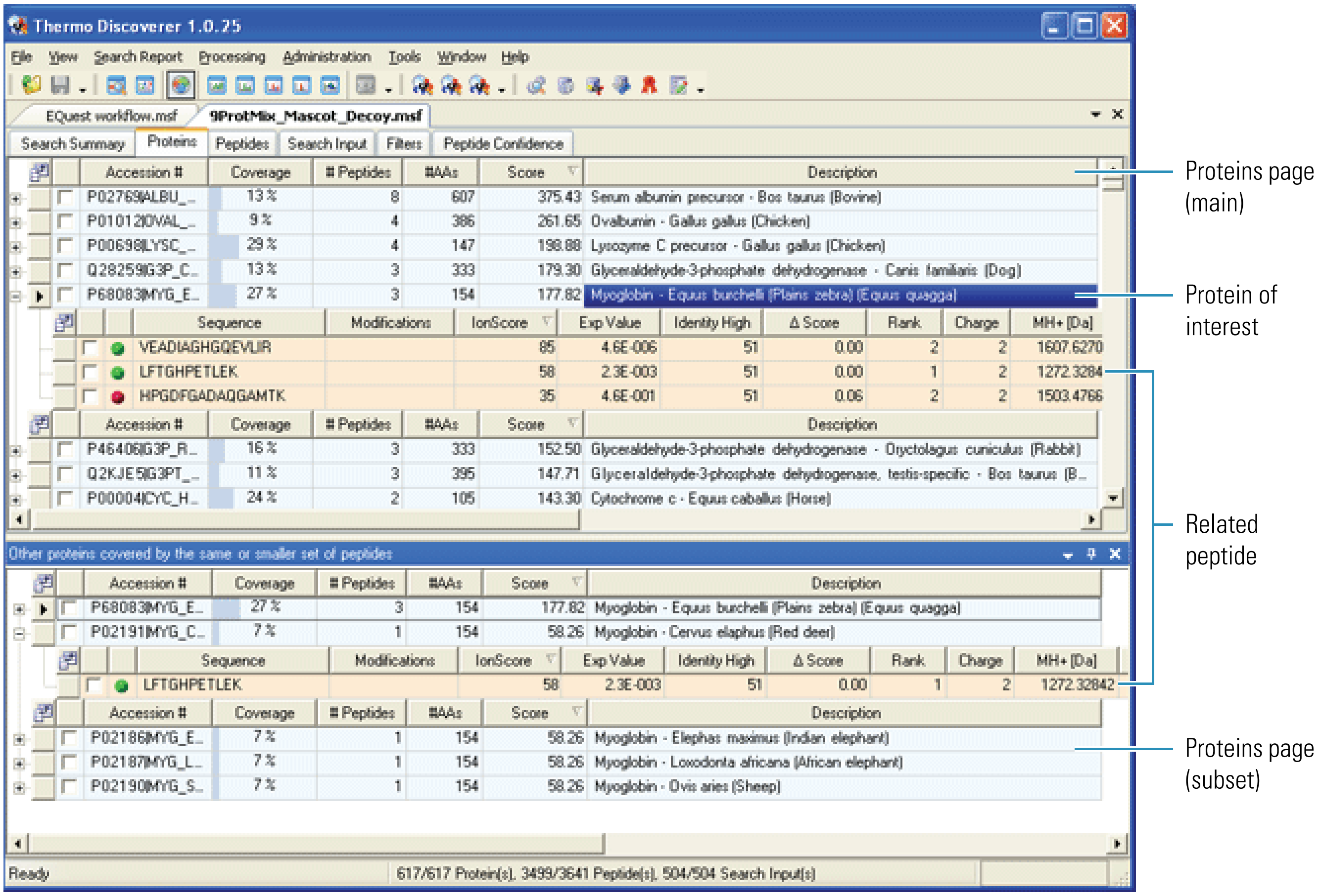
1. Open your search results. See "Understanding Reports and Views" on page 37.

2. On the Proteins page, click anywhere in a protein row.

3. From the Proteome Discoverer toolbar, choose **Search Report > Show proteins covered by this set of peptides**.

   Another Proteins table appears below the Proteins page.

Figure 63. Grouping of proteins



	Accession #	Coverage	# Peptides	#AAs	Score	Description
☐	P02769|ALBU_...	13 %	8	607	375.43	Serum albumin precursor - Bos taurus (Bovine)
☐	P01012|OVAL_...	9 %	4	386	261.65	Ovalbumin - Gallus gallus (Chicken)
☐	P00698|LYSC_...	29 %	4	147	198.88	Lysozyme C precursor - Gallus gallus (Chicken)
☐	Q28259|G3P_C...	13 %	3	333	179.30	Glyceraldehyde-3-phosphate dehydrogenase - Canis familiaris (Dog)
☐	P68083|MYG_E...	27 %	3	154	177.82	Myoglobin - Equus burchelli (Plains zebra) (Equus quagga)

		Sequence	Modifications	IonScore	Exp Value	Identity High	Δ Score	Rank	Charge	MH+ [Da]
☐	●	VEADIAGHGQEVLIR		85	4.6E-006	51	0.00	2	2	1607.6270
☐	●	LFTGHPETLEK		58	2.3E-003	51	0.00	1	2	1272.3284
☐	●	HPGDFGADAQGAMTK		35	4.6E-001	51	0.06	2	2	1503.4766

	Accession #	Coverage	# Peptides	#AAs	Score	Description
☐	P46406|G3P_R...	16 %	3	333	152.50	Glyceraldehyde-3-phosphate dehydrogenase - Oryctolagus cuniculus (Rabbit)
☐	Q2KJE5|G3PT_...	11 %	3	395	147.71	Glyceraldehyde-3-phosphate dehydrogenase, testis-specific - Bos taurus (B...
☐	P00004|CYC_H...	24 %	2	105	143.30	Cytochrome c - Equus caballus (Horse)

Other proteins covered by the same or smaller set of peptides

	Accession #	Coverage	# Peptides	#AAs	Score	Description
☐	P68083|MYG_E...	27 %	3	154	177.82	Myoglobin - Equus burchelli (Plains zebra) (Equus quagga)
☐	P02191|MYG_C...	7 %	1	154	58.26	Myoglobin - Cervus elaphus (Red deer)

		Sequence	Modifications	IonScore	Exp Value	Identity High	Δ Score	Rank	Charge	MH+ [Da]
☐	●	LFTGHPETLEK		58	2.3E-003	51	0.00	1	2	1272.32842

	Accession #	Coverage	# Peptides	#AAs	Score	Description
☐	P02186|MYG_E...	7 %	1	154	58.26	Myoglobin - Elephas maximus (Indian elephant)
☐	P02187|MYG_L...	7 %	1	154	58.26	Myoglobin - Loxodonta africana (African elephant)
☐	P02190|MYG_S...	7 %	1	154	58.26	Myoglobin - Ovis aries (Sheep)

Ready 617/617 Protein(s), 3499/3641 Peptide(s), 504/504 Search Input(s)

Proteins page (main)

Protein of interest

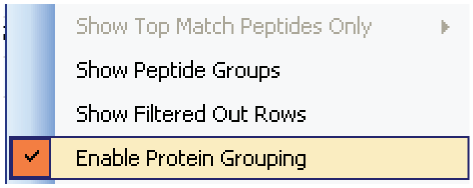
Related peptide

Proteins page (subset)

❖ To switch off protein grouping

1. In a protein grid cell or row, right-click to access the shortcut menu.
2. Clear the check mark for **Enable Protein Grouping**. See Figure 64. Your proteins immediately ungroup.

Figure 64. Protein grid shortcut menu options



Interpreting Your Results with the Chromatogram View

You can display the base peak chromatogram of the original .raw data file with the Chromatogram view. A review of the chromatogram can provide information on the specific peptides and peak shapes as well as the intensity of analytes.

❖ To display the Chromatogram view and peptide elution

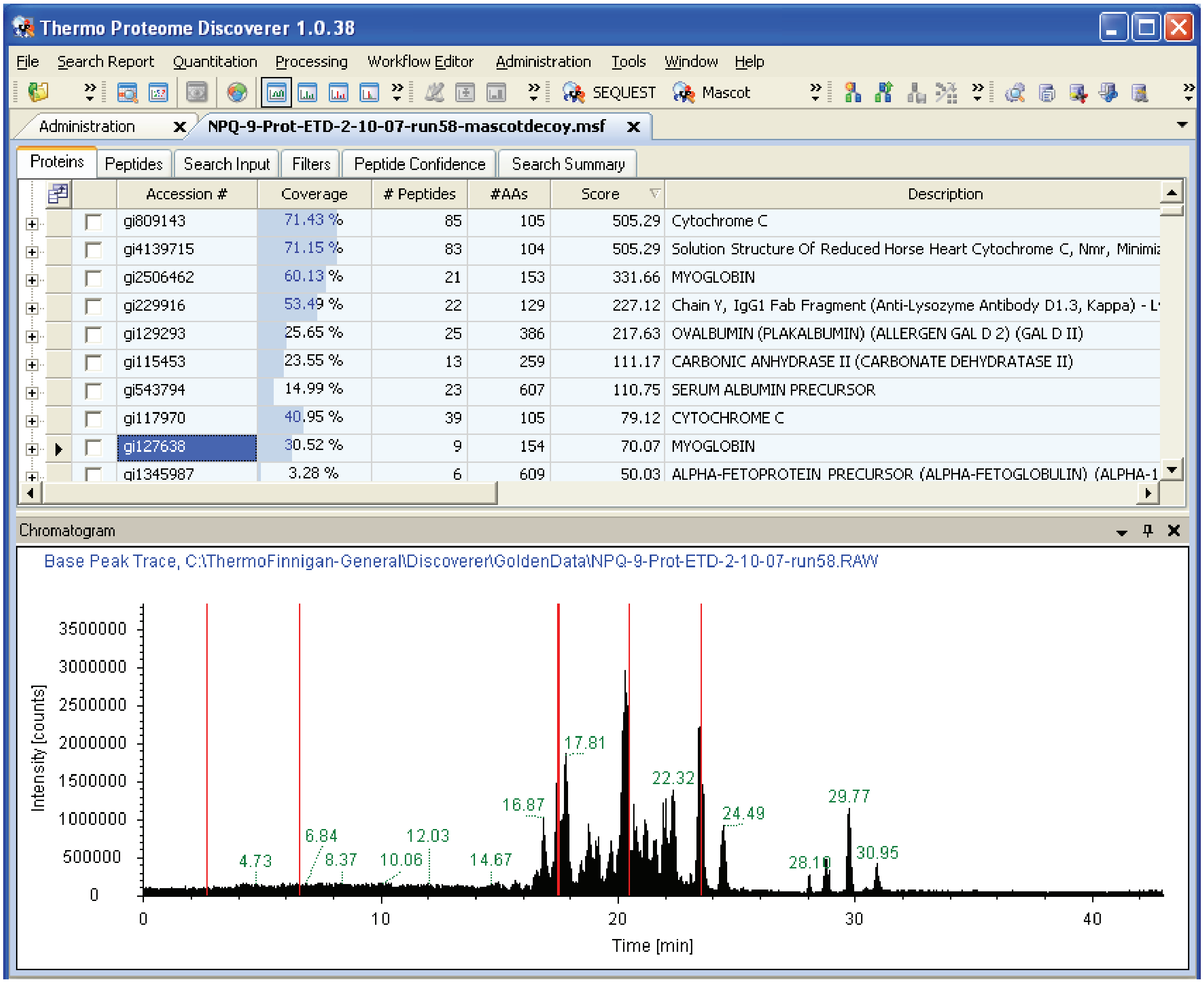
1. Open your search results. See "Understanding Reports and Views" on page 37.
2. From the Proteome Discoverer toolbar, choose **Search Report > Show Chromatogram View**. The Chromatogram view appears.
3. Select a row on the Proteins page. One or more red lines appear in the chromatogram. The red lines indicate the elution time of all identified peptides of the associated protein. See Figure 65.

   A yellow range can appear in the chromatogram view. This is the selected retention time range from which spectra are extracted and submitted for peptide identification.

❖ To interpret your results with the Chromatogram view

1. In the Chromatogram view, select one or more proteins, or one or more peptides.
2. Observe the red line on the yellow level of the grid. The red line indicates the elution time of each instance of this peptide.
3. Check to see if the amount and the profile are as expected.
   - Where is the peptide eluting in the chromatogram? Is it as expected?
   - Is the shape of the chromatogram as expected?

Figure 65. Example of a chromatogram view of proteins



Accession #	Coverage	# Peptides	#AAs	Score	Description
gi809143	71.43 %	85	105	505.29	Cytochrome C
gi4139715	71.15 %	83	104	505.29	Solution Structure Of Reduced Horse Heart Cytochrome C, Nmr, Minimi
gi2506462	60.13 %	21	153	331.66	MYOGLOBIN
gi229916	53.49 %	22	129	227.12	Chain Y, IgG1 Fab Fragment (Anti-Lysozyme Antibody D1.3, Kappa) - L
gi129293	25.65 %	25	386	217.63	OVALBUMIN (PLAKALBUMIN) (ALLERGEN GAL D 2) (GAL D II)
gi115453	23.55 %	13	259	111.17	CARBONIC ANHYDRASE II (CARBONATE DEHYDRATASE II)
gi543794	14.99 %	23	607	110.75	SERUM ALBUMIN PRECURSOR
gi117970	40.95 %	39	105	79.12	CYTOCHROME C
gi127638	30.52 %	9	154	70.07	MYOGLOBIN
gi1345987	3.28 %	6	609	50.03	ALPHA-FETOPROTEIN PRECURSOR (ALPHA-FETOGLOBULIN) (ALPHA-1

Base Peak Trace, C:\ThermoFinnigan-General\Discoverer\GoldenData\NPQ-9-Prot-ETD-2-10-07-run58.RAW

❖ To display the results of Chromatogram and Extracted Ion Chromatogram views

1. Open your search results. See "Understanding Reports and Views" on page 37.

2. From the Proteome Discoverer menu bar, choose **Search Report > Show Chromatogram View**. The Chromatogram view appears.

3. Click **+** to the right of the protein row. The peptides associated with the proteins appear.

4. Click a row header in the peptides row. A red line in the chromatogram appears. The red lines indicate the elution time of all identified peptides.

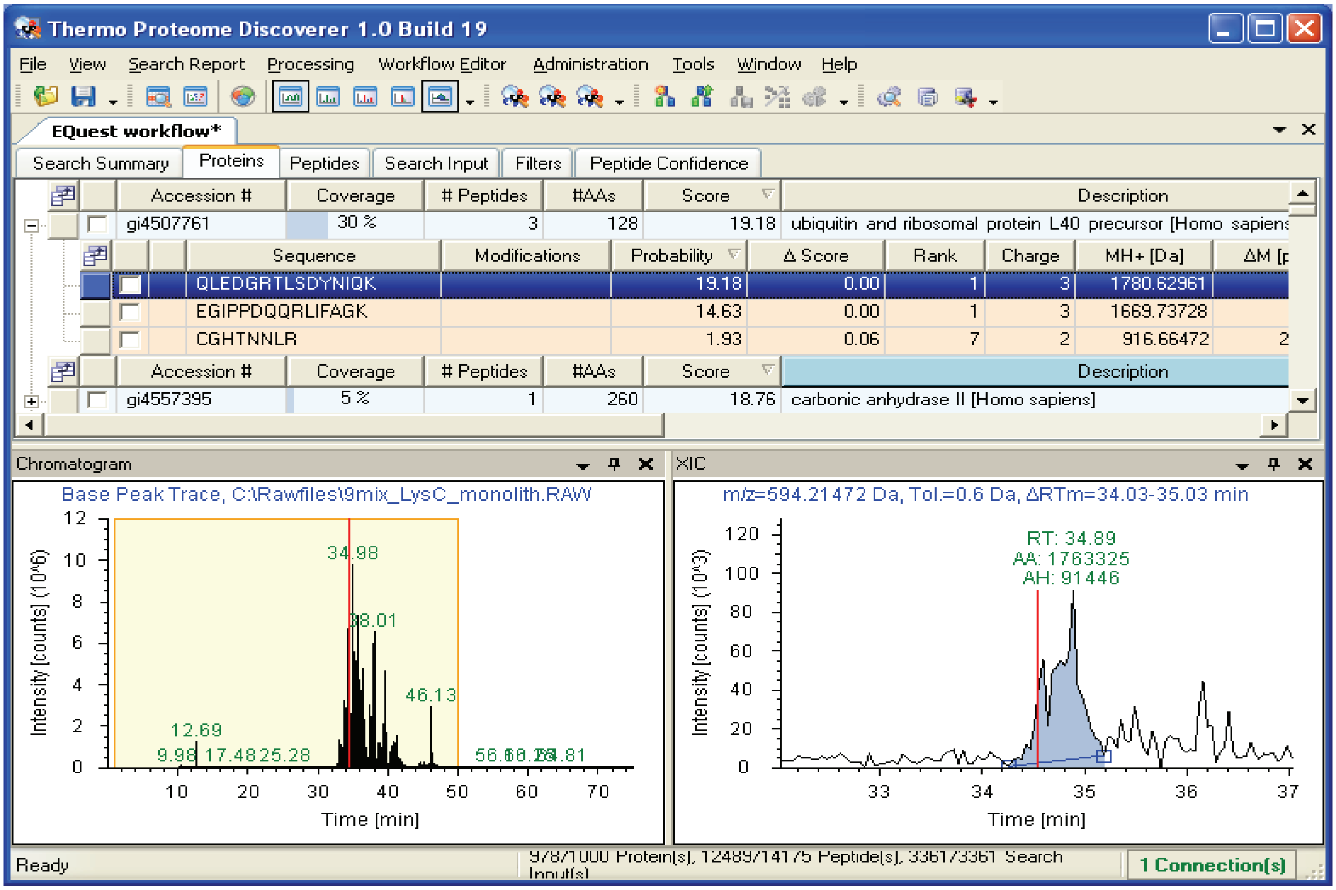
5. From the Proteome Discoverer window, choose **Search Report > Show Extracted Ion Chromatogram**. The Extracted Ion Chromatogram view corresponding to the *m/z* precursor of the selected peptide appears.

❖ To interpret your results with Chromatogram and Extracted Ion Chromatogram views

Select the row header of the peptide results.

- A red line on the Chromatogram view shows the elution position of this particular peptide.
- The view also displays the elution profile of this peptide in the Extracted Ion Chromatogram view.

Figure 66. Chromatogram view and Extracted Ion Chromatogram view of proteins



	Accession #	Coverage	# Peptides	#AAs	Score	Description
☐	gi4507761	30 %	3	128	19.18	ubiquitin and ribosomal protein L40 precursor [Homo sapiens

	Sequence	Modifications	Probability	Δ Score	Rank	Charge	MH+ [Da]	ΔM [p
☐	QLEDGRTLSDYNIQK		19.18	0.00	1	3	1780.62961	
☐	EGIPPDQQRLIFAGK		14.63	0.00	1	3	1669.73728	
☐	CGHTNNLR		1.93	0.06	7	2	916.66472	2

	Accession #	Coverage	# Peptides	#AAs	Score	Description
☐	gi4557395	5 %	1	260	18.76	carbonic anhydrase II [Homo sapiens]

Ready

978/1000 Protein(s), 12489/14175 Peptide(s), 3361/3361 Search Input(s)

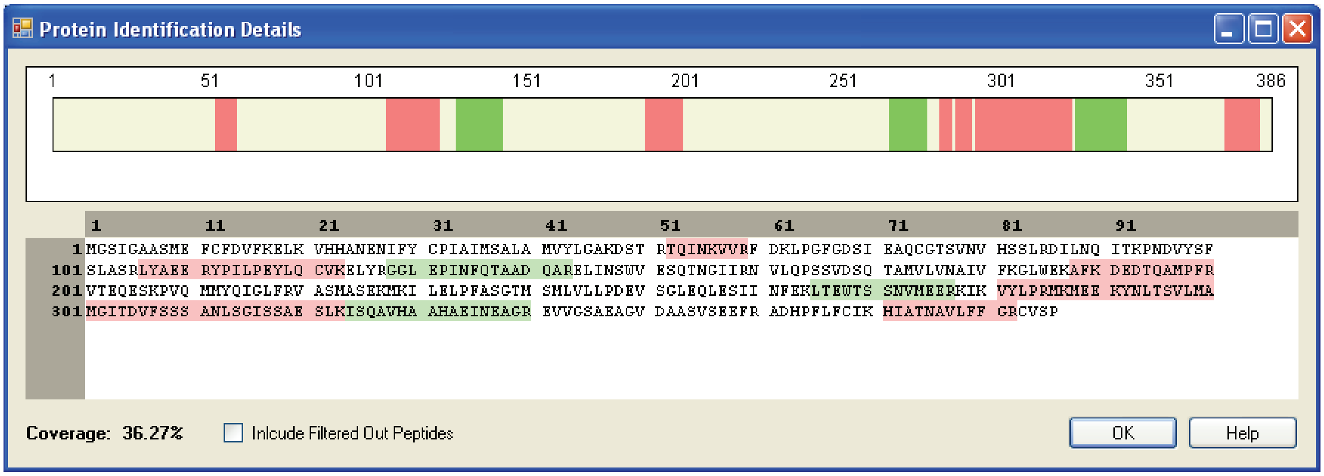
1 Connection(s)

Interpreting Your Results with the Protein Identification Details View

❖ To display Protein Identification Details

1. Open your search results. See “Understanding Reports and Views” on page 37.
2. Select a row header on the Proteins page.
3. From the Proteome Discoverer menu bar, choose **Search Report > Show Details**. The Protein Identification Details view appears. You can also access the Protein Identification page by highlighting a column and selecting Show Details.

Figure 67. Example of Protein Identification Details view



```
     1         11         21         31         41         51         61         71         81         91
  1 MGSIGAASME FCFDVFKELK VHHANENIFY CPIAIMSALA MVYLGAKDST RTQINKVVRF DKLPGFGDSI EAQCGTSVNV HSSLRDILNQ ITKPNDVYSF
101 SLASRLYAEE RYPILPEYLQ CVKELYRGGL EPINFQTAAD QARELINSWV ESQTNGIIRN VLQPSSVDSQ TAMVLVNAIV FKGLWEKAFK DEDTQAMPFR
201 VTEQESKPVQ MMYQIGLFRV ASMASEKMKI LELPFASGTM SMLVLLPDEV SGLEQLESII NFEKLTEWTS SNVMEERKIK VYLPRMKMEE KYNLTSVLMA
301 MGITDVFSSS ANLSGISSAE SLKISQAVHA AHAEINEAGR EVVGSAEAGV DAASVSEEFR ADHPFLFCIK HIATNAVLFF GRCVSP
```

Coverage: 36.27%

☐ Inlcude Filtered Out Peptides

The protein sequence coverage bar is colored to indicate the confidence of the peptide sequence identification. The peptides not highlighted were either not detected or not identified.

- Green—Peptides highlighted are of high confidence.
- Yellow—Peptides highlighted are of modest confidence.
- Red—Peptides highlighted are of low confidence.

4. In the protein sequence coverage bar, click a colored bar. The related sequence is highlighted in the same color.

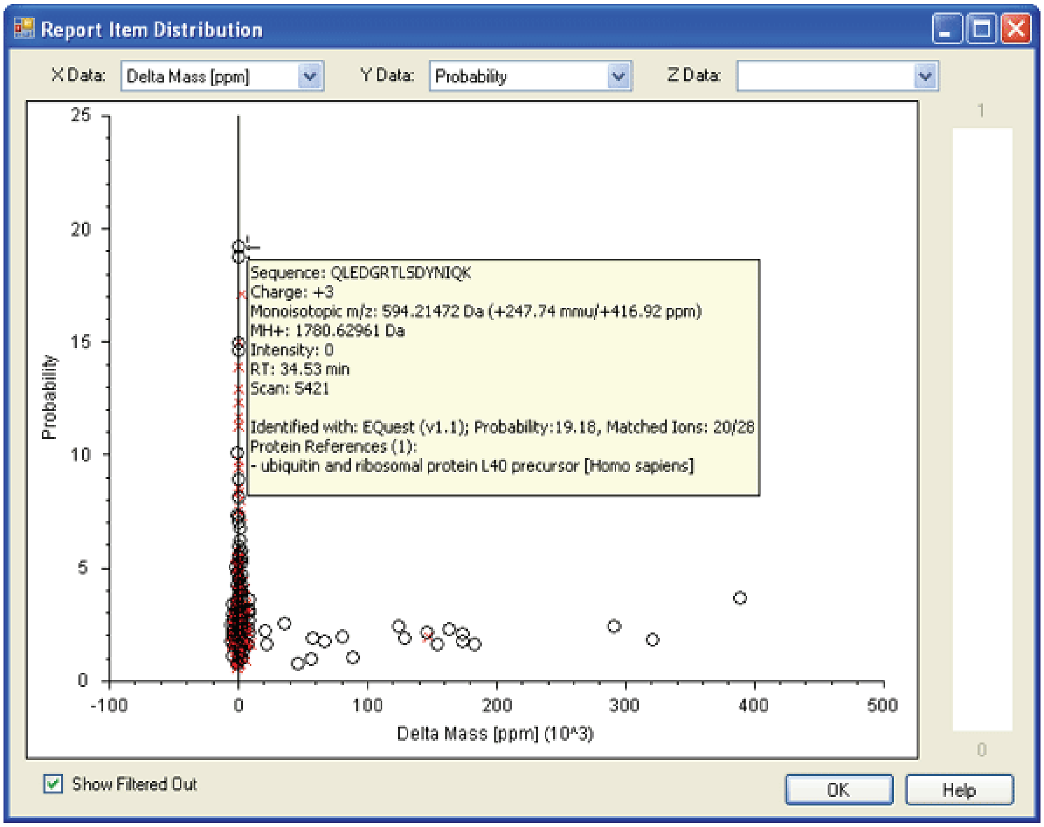
Report Item Distribution Chart

The Report Item Distribution chart provides flexibility to mix and match your search results. You can use the Report Item Distribution to assess different aspects of the search results including scores, delta masses, retention times, and so on by plotting them in relationship to each other. Using the chart, you can plot each property of the identified peptides against each other for comparative analysis.

The Report Item Distribution chart initially plots the score versus delta mass of the identified peptides. To display the extended tooltip information, hover your cursor over a spot in the chart that represents a peptide (as shown in Figure 68). The tooltip information describes the identified peptide, amino acid sequence, charge state, referenced proteins, and so on.

The default display shows the properties of all peptides currently visible. To display the filtered out peptides in the plot, select the **Show Filtered Out** check box at the bottom of the dialog box. The filtered out peptides are plotted as small red crosses (Figure 68 and Figure 69).

Figure 68. Report Item Distribution chart with hover text



X Data: Delta Mass [ppm]

Y Data: Probability

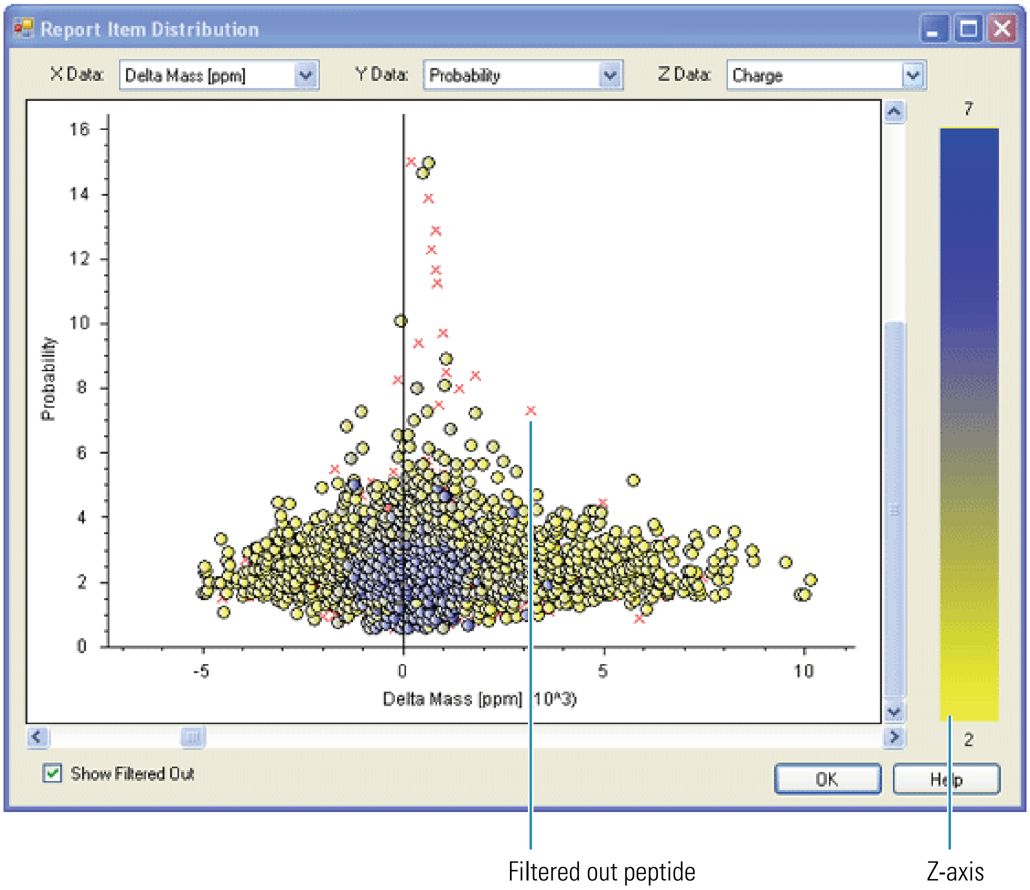
Z Data:

Sequence: QLEDGRTLSDYNIQK
Charge: +3
Monoisotopic m/z: 594.21472 Da (+247.74 mmu/+416.92 ppm)
MH+: 1780.62961 Da
Intensity: 0
RT: 34.53 min
Scan: 5421

Identified with: EQuest (v1.1); Probability:19.18, Matched Ions: 20/28
Protein References (1):
- ubiquitin and ribosomal protein L40 precursor [Homo sapiens]

☑ Show Filtered Out

Figure 69. Report Item Distribution chart with filtered out peptides


Report Item Distribution

X Data: Delta Mass [ppm]
Y Data: Probability
Z Data: Charge

☑ Show Filtered Out

The Report Item Distribution chart is interactive, and supports multi-level zooming and panning. Use the zoom options to look at the pattern in greater detail.

❖ To zoom in

Drag your cursor to the left and select the area you want to enlarge in size.

❖ To zoom out

Drag your cursor to the right and select the area you want to reduce in size.

❖ To use the right-click shortcut menu

Right-click anywhere in the Report Item Distribution chart. The shortcut menu appears.

Working with the Peptides Grid

Use the Peptides grid to accomplish most of your tabular and graphical analyses. From the Peptides grid, you can explore the tabular information of peptides. You can also access the various details views for the peptide matches. Select a row header on the first level (with red background) and select any toolbar icon to access a view such as Details or Spectrum.

On the Peptide page (if the Show Top Match Peptides Only option is active) a non-redundant list of all identified peptides is displayed. All displayed peptides are above the threshold set in any applied filters. Peptides below the threshold can be displayed in a grayed-out form if the Show Filtered Out Rows option is selected from the shortcut menu.

Note With quantitation reports the peptide non-redundant list is disabled.

To start a data validation from the Peptides grid, click the **Peptides** tab above the Results grid. Again, the same two levels of detail associated with a particular peptide sequence are available in the Results grid.

Interpreting Your Results with the Search Reports Views

❖ To view the spectrum

1. In an open report, click the **Peptides** tab. The peptide view of your search report appears.
2. Select a peptide row.
3. Choose **Search Report > Show Spectrum.** The Peptide Spectrum view appears.

❖ To view the fragment match spectrum

1. In an open report, click the **Peptide** tab. The peptide view of your search report appears.
2. Select a peptide row.
3. Choose **Search Report > Show Fragment Match Spectrum**. The Peptide Fragment Match view appears.

❖ To view the isotope pattern of the selected precursor

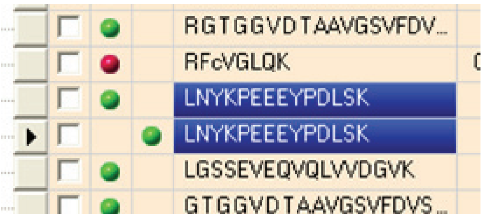
1. In an open report, click the **Peptides** tab. The peptide view of your search report appears.
2. Select a peptide row.
3. Choose **Search Report > Show Isotope Pattern**. The Isotope Pattern view appears.

❖ To view i the extracted ion chromatogram

1. In an open report, click the **Peptides** tab. The peptide view of your search report appears.
2. Select a peptide row.
3. Choose **Search Report > Show Extacted Ion Chromatogram**. The Extacted Ion Chromatogram view appears.

❖ To view the peptide consensus page

1. In an open report, click the **Peptides** tab. The peptide view of your search report appears.
2. Drag the cursor across boxes for identical peptides.

	☐	●		RGTGGVDTAAVGSVFDV...	
	☐	●		RFcVGLQK	(
	☐	●		LNYKPEEEYPDLSK	
▶	☐		●	LNYKPEEEYPDLSK	
	☐	●		LGSSEVEQVQLVVDGVK	
	☐	●		GTGGVDTAAVGSVFDVS...	

3. Click 👁. The peptide consensus view appears.

   In this view you can see the following:

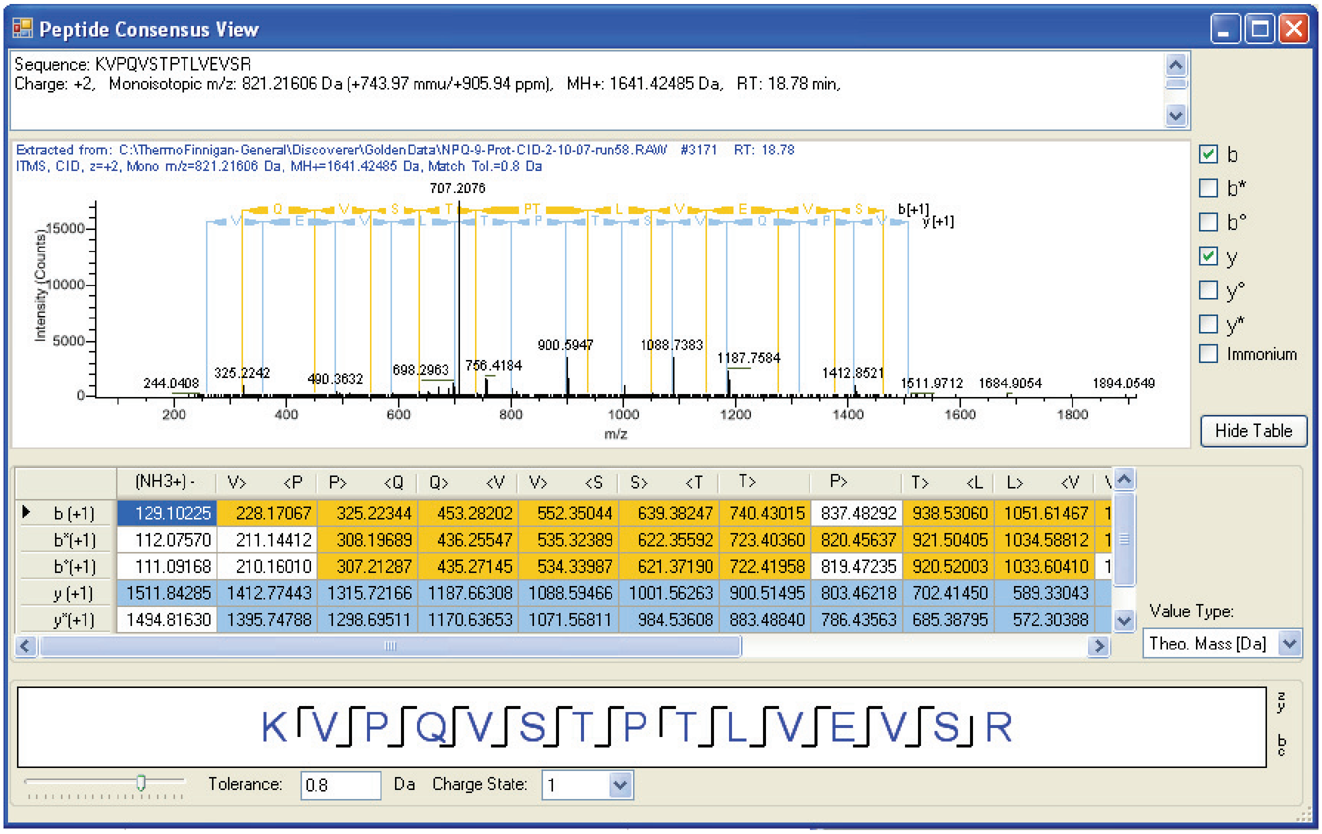
   - Spectra detail as a text header
   - Horizontal bars showing mass difference that confirms the presence of residues
   - Grid cell color indicating positive identification within tolerance and hue-indicated ion series
   - Fragment ion coverage showing which ion type confirms fragmentation
   - Theoretical values for ions formed by indicated bond breakage

4. (Optional) Drag your cursor over spectra region to zoom in. Right-click to reset to initial view.

❖ To view the peptide consensus page

1. In an open report, click the **Peptides** tab. The peptide view of your search report appears.
2. Select a peptide row.
3. Choose **Search Report > Show Peptide Consensus View**. The Peptide Consensus View appears. See Figure 70.

Figure 70. Peptide Consensus View



Sequence: KVPQVSTPTLVEVSR
Charge: +2, Monoisotopic m/z: 821.21606 Da (+743.97 mmu/+905.94 ppm), MH+: 1641.42485 Da, RT: 18.78 min,

Extracted from: C:\ThermoFinnigan-General\Discoverer\GoldenData\NPQ-9-Prot-CID-2-10-07-run58.RAW #3171 RT: 18.78
ITMS, CID, z=+2, Mono m/z=821.21606 Da, MH+=1641.42485 Da, Match Tol.=0.8 Da

☑ b
☐ b*
☐ b°
☑ y
☐ y°
☐ y*
☐ Immonium

	(NH3+) -	V> <P	P> <Q	Q> <V	V> <S	S> <T	T>	P>	T> <L	L> <V	\
b (+1)	129.10225	228.17067	325.22344	453.28202	552.35044	639.38247	740.43015	837.48292	938.53060	1051.61467	1
b*(+1)	112.07570	211.14412	308.19689	436.25547	535.32389	622.35592	723.40360	820.45637	921.50405	1034.58812	1
b°(+1)	111.09168	210.16010	307.21287	435.27145	534.33987	621.37190	722.41958	819.47235	920.52003	1033.60410	1
y (+1)	1511.84285	1412.77443	1315.72166	1187.66308	1088.59466	1001.56263	900.51495	803.46218	702.41450	589.33043	
y*(+1)	1494.81630	1395.74788	1298.69511	1170.63653	1071.56811	984.53608	883.48840	786.43563	685.38795	572.30388	

Value Type: Theo. Mass [Da]

K⌈V⌋P⌋Q⌋V⌋S⌋T⌋P⌈T⌋L⌋V⌋E⌋V⌋S⌋R

Tolerance: 0.8 Da Charge State: 1

Peptide Identification Details

Use the Peptides identification details page to show the analysed spectra of the selected peptide sequence in the peptide grid. You can refer to the following section on fragment ions while choosing your peptide identification details.

Fragment Ions

Fragment ions of peptides are produced by a collision-induced dissociation (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 N-CR bond and are used to determine amino acid sequence. 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. Fragment ions A, B, and C have a charge on the N-terminal side and fragment ions X, Y, and Z 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^{o}, B^{o}, and C^{o} are ions that have lost water (-18 Da). The subscript next to the letter indicates the number of residues in the fragment ion.[1]

Table 15 summarizes the fragment ions used in Proteome Discoverer:

Table 15. Fragment ions

Ions	Description
A	Charge on N-terminal side
B	Charge on N-terminal side
C	Charge on N-terminal side
Y	Charge on C-terminal side
Z	Charge on C-terminal side
B*	B ion that has lost ammonia (-17 Da)
Y*	Y ion that has lost ammonia (-17 Da)
B^{o}	B ion that has lost water (-18 Da)
Y^{o}	Y ion that has lost water (-18 Da)

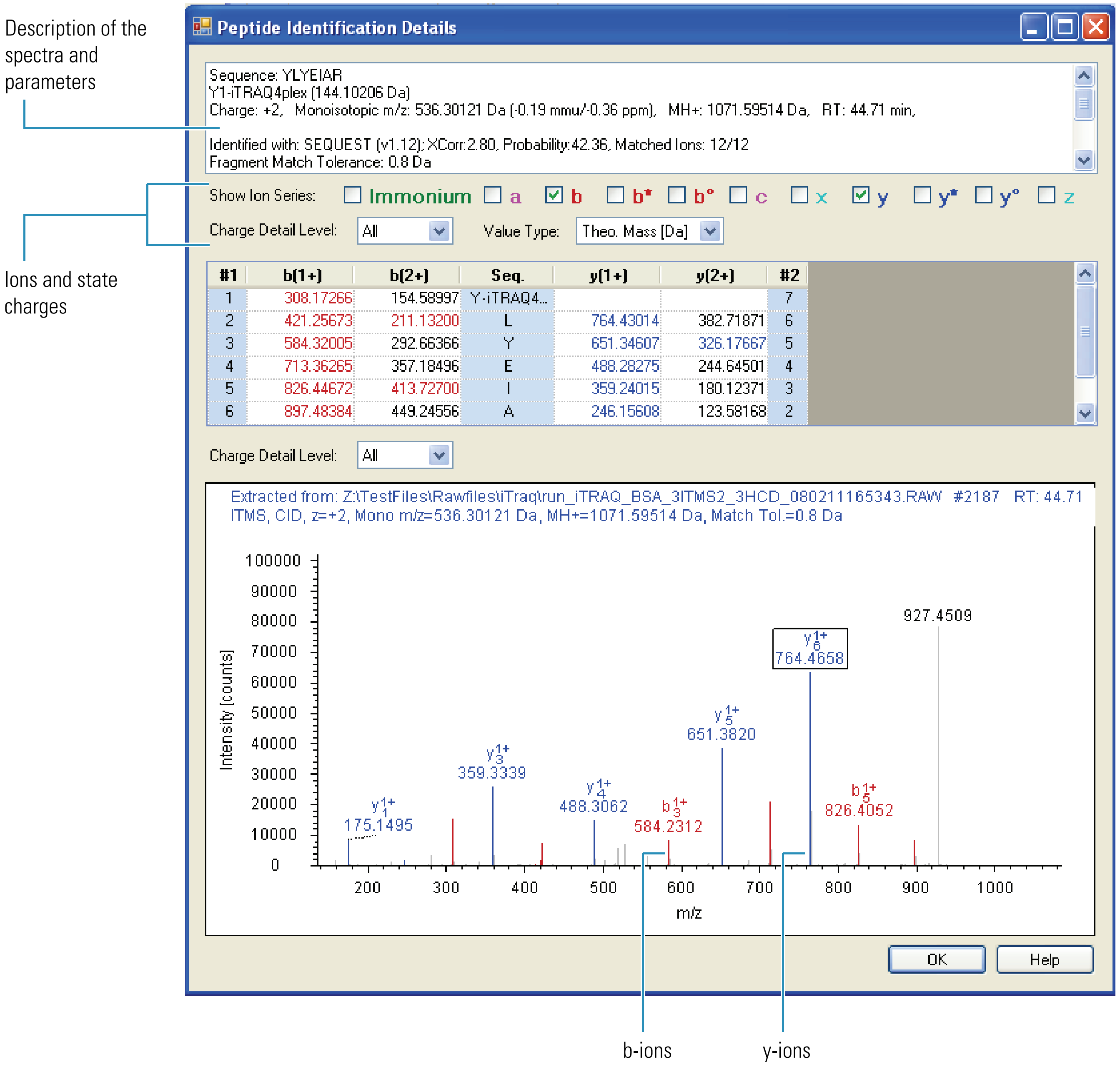
---

[1] For more information on fragment ions and nomenclature, see *Proposal for a Common Nomenclature for Sequence Ions in Mass Spectra of Peptides*; Roepstorff, P. and Fohlman, J.; Biomed Mass Spectrum, 11(11) 601 (1984).

❖ To view the peptide identification details page

1. In an open report, click the **Peptides** tab. The peptide view of your search report appears.
2. Select a peptide row and double-click a grid cell. The Peptide Identification Details page appears. See Figure 71 and Figure 72 for details.

Figure 71. Peptide identification details



Description of the spectra and parameters

Ions and state charges

b-ions

y-ions

Peptide Identification Details

Sequence: YLYEIAR
Y1-iTRAQ4plex (144.10206 Da)
Charge: +2, Monoisotopic m/z: 536.30121 Da (-0.19 mmu/-0.36 ppm), MH+: 1071.59514 Da, RT: 44.71 min,

Identified with: SEQUEST (v1.12); XCorr:2.80, Probability:42.36, Matched Ions: 12/12
Fragment Match Tolerance: 0.8 Da

Show Ion Series: ☐ Immonium ☐ a ☑ b ☐ b* ☐ b° ☐ c ☐ x ☑ y ☐ y* ☐ y° ☐ z

Charge Detail Level: All

Value Type: Theo. Mass [Da]

#1	b(1+)	b(2+)	Seq.	y(1+)	y(2+)	#2
1	308.17266	154.58997	Y-iTRAQ4...			7
2	421.25673	211.13200	L	764.43014	382.71871	6
3	584.32005	292.66366	Y	651.34607	326.17667	5
4	713.36265	357.18496	E	488.28275	244.64501	4
5	826.44672	413.72700	I	359.24015	180.12371	3
6	897.48384	449.24556	A	246.15608	123.58168	2

Charge Detail Level: All

Extracted from: Z:\TestFiles\Rawfiles\iTraq\run_iTRAQ_BSA_3ITMS2_3HCD_080211165343.RAW #2187 RT: 44.71
ITMS, CID, z=+2, Mono m/z=536.30121 Da, MH+=1071.59514 Da, Match Tol.=0.8 Da

Figure 72. Detailed view of the Peptide Identification Details page

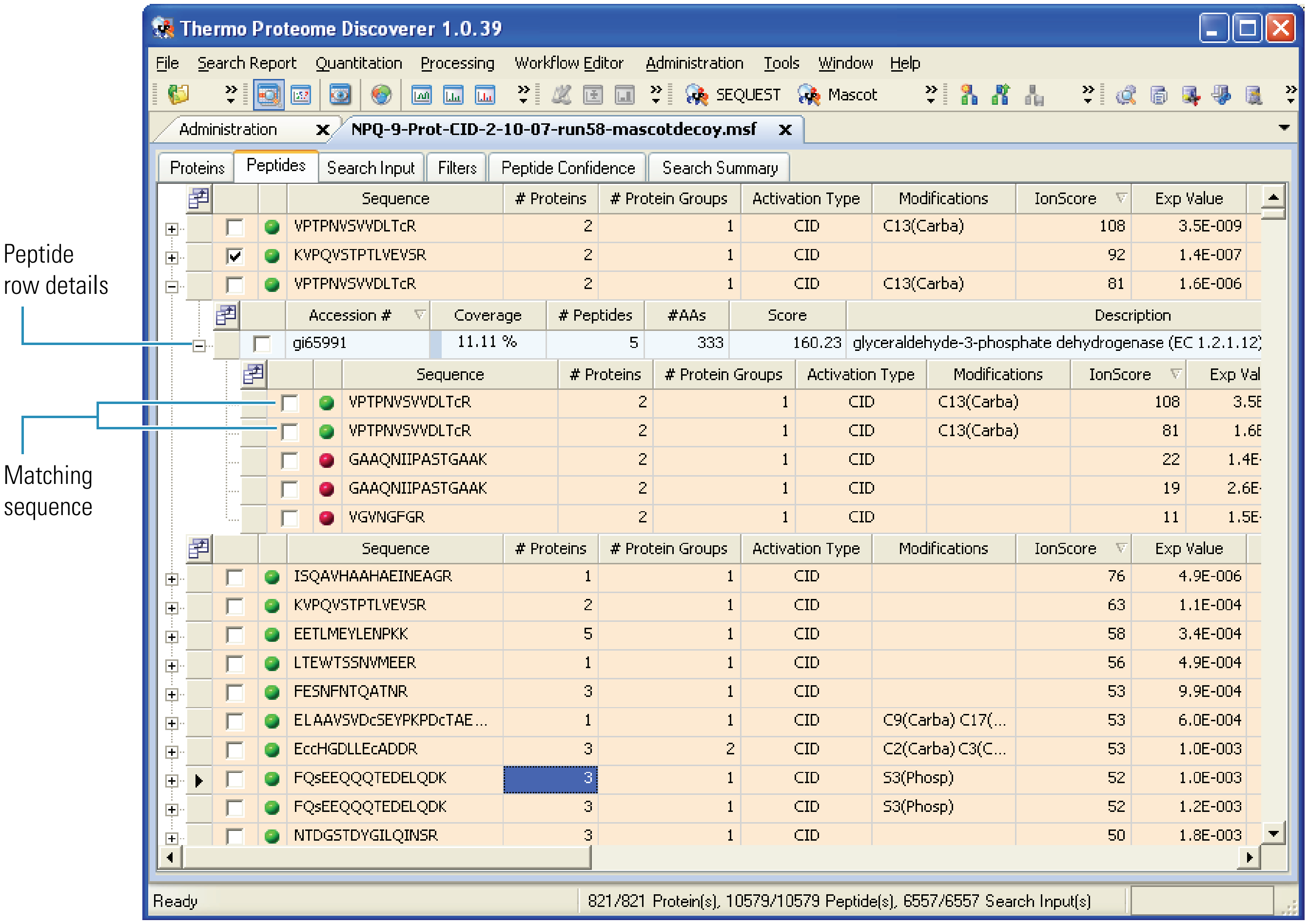
❖ To view the peptide grid, sequence view

1. In an open report, click the **Peptides** tab. The peptide view of your search report appears.
2. Select a peptide row.
3. In the peptide row, click **+**. Sequences matching that peptide appear in a grid under the peptide. See Figure 71.

   When drilling down into a peptide match, Proteome Discoverer lists all identified proteins that this particular peptide is contained in. When drilling down further into an identified protein, Proteome Discoverer lists all other peptides identified for this particular protein.

   If enabled in the Column Chooser of the Proteins grid, the second layer of the Peptides grid (that is, the identified proteins that this peptide is contained in) is also available in the Proteins grid as the third layer.

Figure 73. Peptide grid, sequence view



		Sequence	# Proteins	# Protein Groups	Activation Type	Modifications	IonScore	Exp Value
☐	🟢	VPTPNVSVVDLTcR	2	1	CID	C13(Carba)	108	3.5E-009
☑	🟢	KVPQVSTPTLVEVSR	2	1	CID		92	1.4E-007
☐	🟢	VPTPNVSVVDLTcR	2	1	CID	C13(Carba)	81	1.6E-006

	Accession #	Coverage	# Peptides	#AAs	Score	Description
☐	gi65991	11.11 %	5	333	160.23	glyceraldehyde-3-phosphate dehydrogenase (EC 1.2.1.12)

		Sequence	# Proteins	# Protein Groups	Activation Type	Modifications	IonScore	Exp Val
☐	🟢	VPTPNVSVVDLTcR	2	1	CID	C13(Carba)	108	3.5E
☐	🟢	VPTPNVSVVDLTcR	2	1	CID	C13(Carba)	81	1.6E
☐	🔴	GAAQNIIPASTGAAK	2	1	CID		22	1.4E-
☐	🔴	GAAQNIIPASTGAAK	2	1	CID		19	2.6E-
☐	🔴	VGVNGFGR	2	1	CID		11	1.5E-

		Sequence	# Proteins	# Protein Groups	Activation Type	Modifications	IonScore	Exp Value
☐	🟢	ISQAVHAAHAEINEAGR	1	1	CID		76	4.9E-006
☐	🟢	KVPQVSTPTLVEVSR	2	1	CID		63	1.1E-004
☐	🟢	EETLMEYLENPKK	5	1	CID		58	3.4E-004
☐	🟢	LTEWTSSNVMEER	1	1	CID		56	4.9E-004
☐	🟢	FESNFNTQATNR	3	1	CID		53	9.9E-004
☐	🟢	ELAAVSVDcSEYPKPDcTAE…	1	1	CID	C9(Carba) C17(…	53	6.0E-004
☐	🟢	EccHGDLLEcADDR	3	2	CID	C2(Carba) C3(C…	53	1.0E-003
☐	🟢	FQsEEQQQTEDELQDK	3	1	CID	S3(Phosp)	52	1.0E-003
☐	🟢	FQsEEQQQTEDELQDK	3	1	CID	S3(Phosp)	52	1.2E-003
☐	🟢	NTDGSTDYGILQINSR	3	1	CID		50	1.8E-003

Ready 821/821 Protein(s), 10579/10579 Peptide(s), 6557/6557 Search Input(s)

Working with the Search Input

Use the Search Input tab to display the grid for all individual peptide results.

- Interpreting the Isotope Pattern View
- Interpreting the Spectrum View
- Interpreting the Extracted Ion Chromatogram
- Interpreting the Fragment Match View

Interpreting the Isotope Pattern View

The Isotope Pattern view displays the isotope pattern of the precursor associated with the identified peptide (Figure 74). The components of this view are as follows:

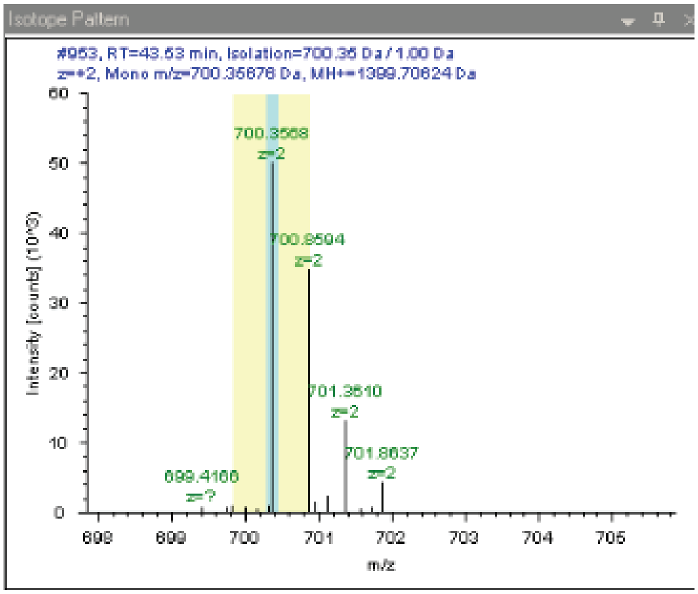
- The yellow region displays the isolation width of the instrument.
- A red line indicates the monoisotopic precursor mass-to-charge value determined by the instrument during acquisition. This is the isolation mass, which is displayed in the header of the isotope pattern view.
- A blue line marks the calculated monoisotopic precursor mass and represents the re-evaluated monoisotopic *m/z* value for the detected peptide.

This view can also be used to assess the abundance and intensity of the precursor or to reference the experimentally determined mass of the precursor and its isotopes, the isolation window, or other details.

❖ To display the Isotope Pattern view

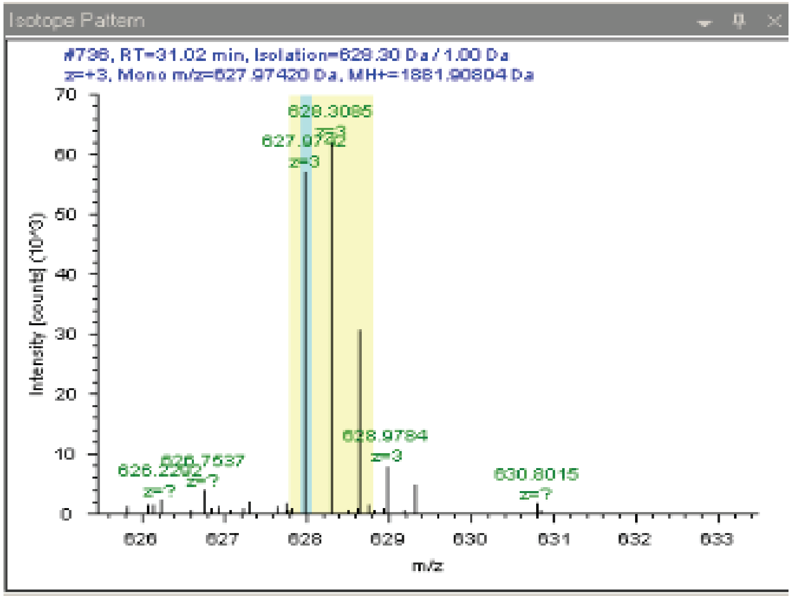
1. Open your search results. See “Understanding Reports and Views” on page 37.
2. From the Proteome Discoverer menu bar, choose **Search Report > Show Isotope Pattern**. The mass spectrum appears.
3. Select a row header on the Peptides page. A blue line and yellow bar in the Isotope Pattern appear. A red line can also appear if the monoisotopic mass was redetermined post-acquisition.

Figure 74. Isotope Pattern view



```
#953, RT=43.53 min, Isolation=700.36 Da / 1.00 Da
z=+2, Mono m/z=700.35678 Da, MH+=1399.70624 Da
```

4. Use the Isotope Pattern to check if the correct monoisotopic mass has been calculated, since the first isotope of a peptide is not always the most intense ion.



#736, RT=31.02 min, Isolation=628.30 Da / 1.00 Da
z=+3, Mono m/z=627.97420 Da, MH+=1881.90804 Da

Interpreting the Spectrum View

The Spectrum view displays the graphical spectrum of the submitted peak list used for the search. This spectrum view might differ from the original spectrum in the .raw file if spectral preprocessing (such as noise filter or special grouping) was applied to tandem mass spectra prior to searching associated with each peptide. Use this view to check the quality of the spectrum.

❖ To display a Spectrum view

1. Open your search results. See “Understanding Reports and Views” on page 37.
2. From the Proteome Discoverer menu bar, choose **Search Report > Show Spectrum**. The Spectrum view appears.
3. Select a row header on the Peptides page. The spectrum of the individual peptide appears.
4. Use the Spectrum view to check the peptide.

Interpreting the Extracted Ion Chromatogram

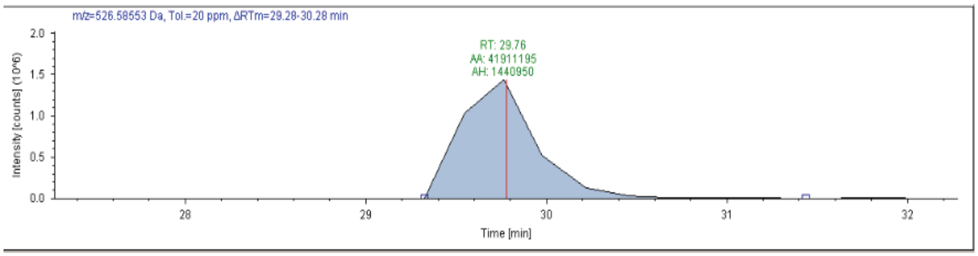
This view displays the extracted ion chromatogram of the precursor mass associated with each peptide. The extracted ion chromatogram is a plot of the intensity of an ion versus unit of time. The red line shows when the MS2 spectrum was taken.

❖ To display the Extracted Ion Chromatogram

1. Open your search results. See "Understanding Reports and Views" on page 37.
2. From the Proteome Discoverer menu bar, choose **Search Report > Show Extracted Ion Chromatogram**. The extracted mass chromatogram view appears. See Figure 75.
3. Select a row on the Peptides page.
4. Note the graphically displayed intensity of the peptide:
    - The peak start and end points as well as the baseline are in blue.
    - The peak area or the height value is automatically calculated.

   You can use the view to assess the chromatographic peak shape of the associated precursor, and to reference the elution time of the identified peptide. The integrated area under the curve and height of the peak is displayed and can be used to assess the abundance of the precursor.
5. Use the graph to determine answers to such questions as
    - Is the MS2 spectrum of poor signal-to-noise due to low abundance of the peptide?
    - Was the MS2 triggered too early (at the start of the peak) or too late (at the end of the peak)?
6. To magnify a particular peak, drag your cursor over the region of interest.

Figure 75. Extracted ion chromagraph



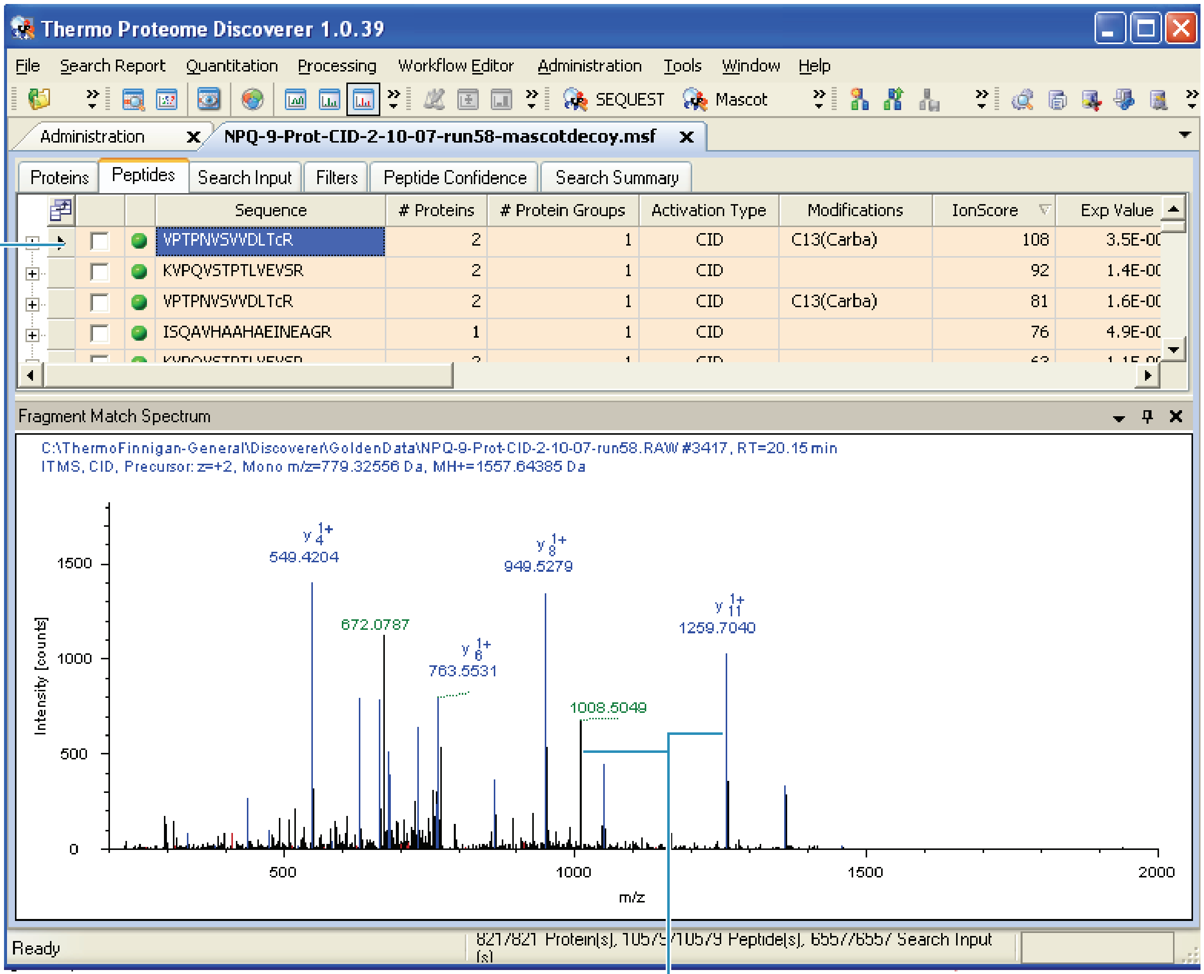
Interpreting the Fragment Match View

The Fragment Match view displays the annotated spectrum of the identified peptide. The matched fragments are colored according to their fragment type, such as blue for b-ions and red for y-ions.

❖ To display the Fragment Match view

1. Open your search results. See "Understanding Reports and Views" on page 37.
2. From the Proteome Discoverer menu bar, choose **Search Report > Show Fragment Match Spectrum**. The Fragment Match Spectrum view appears. See Figure 76. Select a row header on the Peptides page.
3. Confirm that all major fragments are assigned and colored coded.
4. Assess the quality of the match between the submitted spectrum peak list and the identified peptide.

Figure 76. Fragment mass view



		Sequence	# Proteins	# Protein Groups	Activation Type	Modifications	IonScore	Exp Value
☐	●	VPTPNVSVVDLTcR	2	1	CID	C13(Carba)	108	3.5E-0
☐	●	KVPQVSTPTLVEVSR	2	1	CID		92	1.4E-0
☐	●	VPTPNVSVVDLTcR	2	1	CID	C13(Carba)	81	1.6E-0
☐	●	ISQAVHAAHAEINEAGR	1	1	CID		76	4.9E-0

C:\ThermoFinnigan-General\Discoverer\GoldenData\NPQ-9-Prot-CID-2-10-07-run58.RAW #3417, RT=20.15 min
ITMS, CID, Precursor: z=+2, Mono m/z=779.32556 Da, MH+=1557.64385 Da

Ready
821/821 Protein(s), 10579/10579 Peptide(s), 6557/6557 Search Input (s)

Peptide row

Major fragments

Exporting Data to Other Programs

Proteome Discoverer offers extensive options for exporting data and results into other file formats, such as .xml. You can do the following:

- Export in common, open standard formats.
- Export spectral information in .mgf, mzData, .dta file formats.
- Export analysis results as ProtXML and as Excel files for detailed analysis of your search results.
- Export all grid data by copying and pasting into Excel files.
- Export all charts to the Clipboard to save in various image formats.

You can create a Peptide Report when the peptide information appears in the initial results report table. You can include the following types of information in the Peptide Report:

- The information displayed in the report table, including the peptides identified, XCorr, probability, and other scores.
- Peptide match information, which includes the sequences and scoring information for all the peptide matches identified for each scan.

Copying or Saving a View to an Image

This section describes how to copy your analysis results and save your results to another application.

❖ To copy a view

1. Open a view such as a Chromatogram view.
2. Right-click and choose **Copy** from the shortcut menu. The view is automatically stored as an image to the Clipboard. You can paste the image into another application.

❖ To save a view in another format

1. Open a view, such as a Chromatogram view.
2. Right-click and choose **Save as** from the shortcut menu.
3. Select the image type: **.emf**, **.png**, **.gif**, **.jpeg**, **.tiff**, or **.bmp**.
4. (Optional) Select the location to store the image.
5. In the File name box, type the name of the file.
6. Click **Save**. The image is saved in the format and location you selected.

Exporting Exclusion and Inclusion Mass Lists to Xcalibur

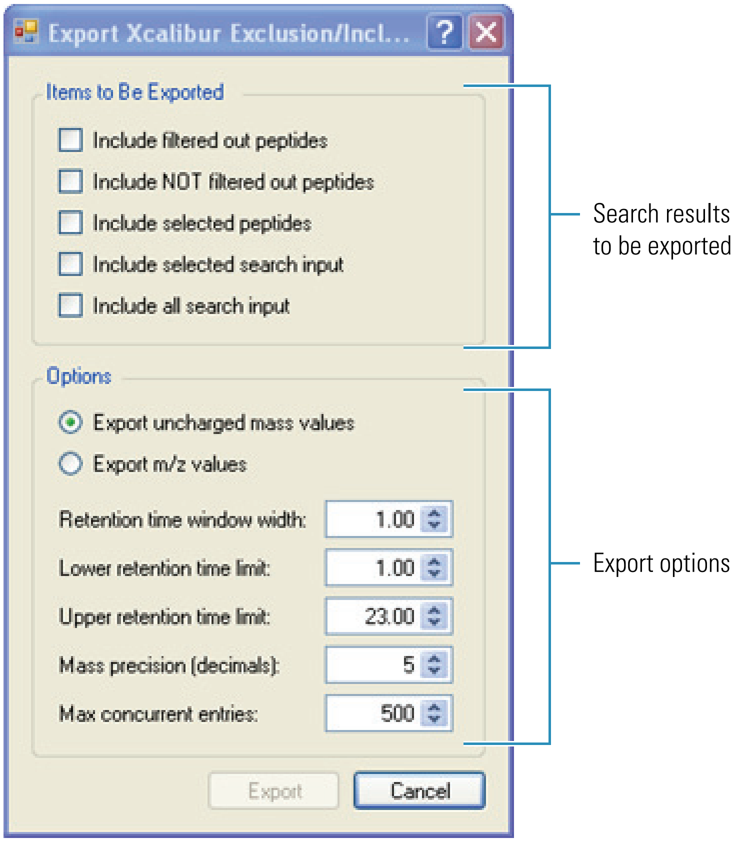
With Proteome Discoverer, you can export exclusion or inclusion mass lists based on your current search results. Exclusion and inclusion mass lists differ in their usage but have the same basic format. Use this feature to export to a format that can be used in Xcalibur. Exporting is a two-step process:

- Determine what portions of the search results to export.
- Define additional limits of the export.

❖ To export exclusion or inclusion mass lists

1. Open your search results. See "Understanding Reports and Views" on page 37.
2. From the Proteome Discoverer menu bar, choose **Search Report > Export Xcalibur Exclusion List**. The Export Xcalibur Exclusion/Inclusion Mass List dialog box appears as shown in Figure 77.

Figure 77. Export Xcalibur Exclusion dialog box


Export Xcalibur Exclusion/Incl...

Items to Be Exported
☐ Include filtered out peptides
☐ Include NOT filtered out peptides
☐ Include selected peptides
☐ Include selected search input
☐ Include all search input

Options
● Export uncharged mass values
○ Export m/z values

Retention time window width: 1.00
Lower retention time limit: 1.00
Upper retention time limit: 23.00
Mass precision (decimals): 5
Max concurrent entries: 500

Search results to be exported

Export options

3. In the Items to Be Exported area, specify which result items should be transferred to the exclusion mass list.

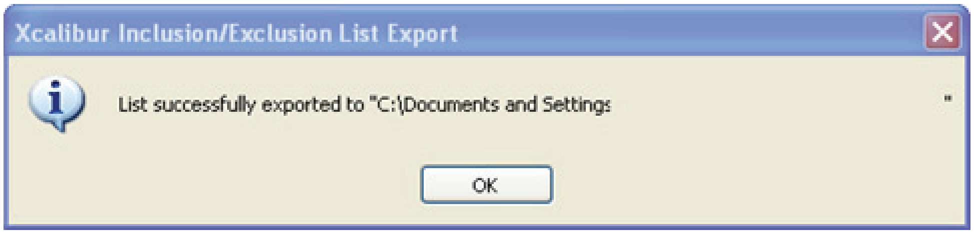
4. In the Options area, select either **Export uncharged mass values** or **Export m/z values**.

5. Choose the proper values for your data set for the time value options.

6. Click **Export**. The Save As dialog box opens.

7. Type the File name for the exported list. A success message appears (see Figure 78).

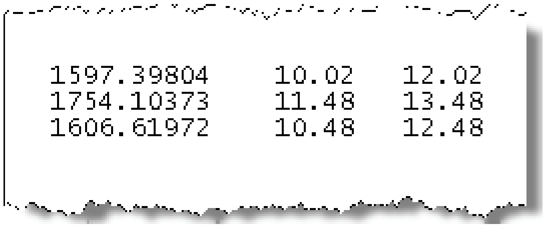
Figure 78. Export successful message



List successfully exported to "C:\Documents and Settings "

8. Use a standard text editor to view the resulting exclusion mass list (see Figure 79) from your hard drive.

Figure 79. Exported exclusion mass list example



```
1597.39804    10.02   12.02
1754.10373    11.48   13.48
1606.61972    10.48   12.48
```

Export Search Results to Excel

You can export protein and peptide identification results and also the search input and other grids, such as the fragment match matrix, directly into a spreadsheet application, such as Excel.

❖ To export search results to Excel (example for proteins and peptides)

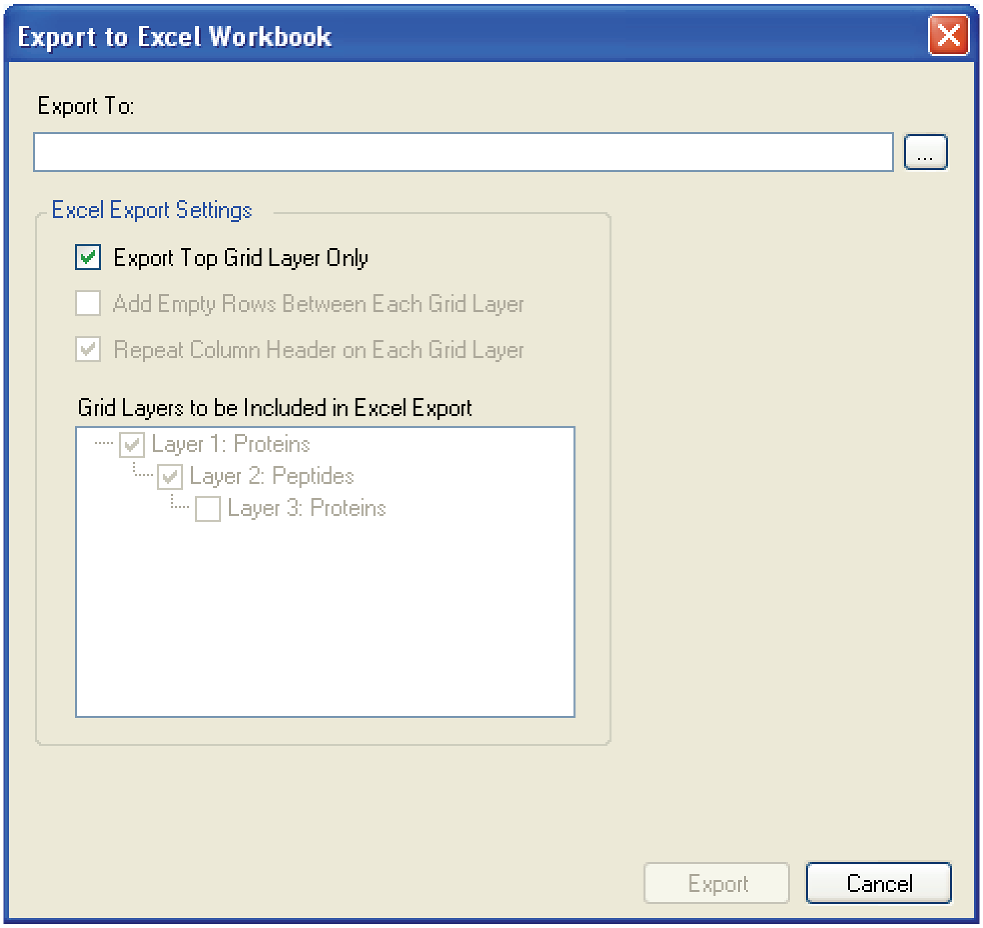
1. Open your search results. See "Understanding Reports and Views" on page 37.

2. Select the **Proteins** or **Peptides** tab.

3. On the Proteins or Peptides page, right-click anywhere in the table to display the shortcut menu.

4. Choose **Export to Excel Workbook**. The Export to Excel Workbook dialog box appears as shown in Figure 80.

Figure 80. Export to Excel Workbook dialog box



Export to Excel Workbook

Export To:

Excel Export Settings

☑ Export Top Grid Layer Only

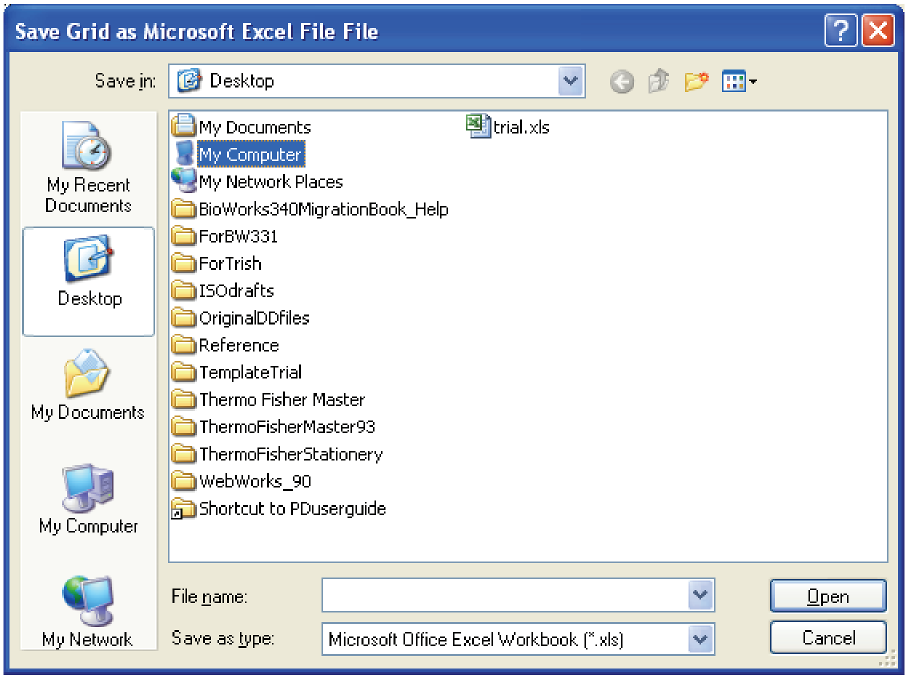
☐ Add Empty Rows Between Each Grid Layer

☑ Repeat Column Header on Each Grid Layer

Grid Layers to be Included in Excel Export

- ☑ Layer 1: Proteins
  - ☑ Layer 2: Peptides
    - ☐ Layer 3: Proteins

5. Click browse (**...**). The Save Grid as Microsoft Excel File dialog box appears.



Save in: Desktop

File name:

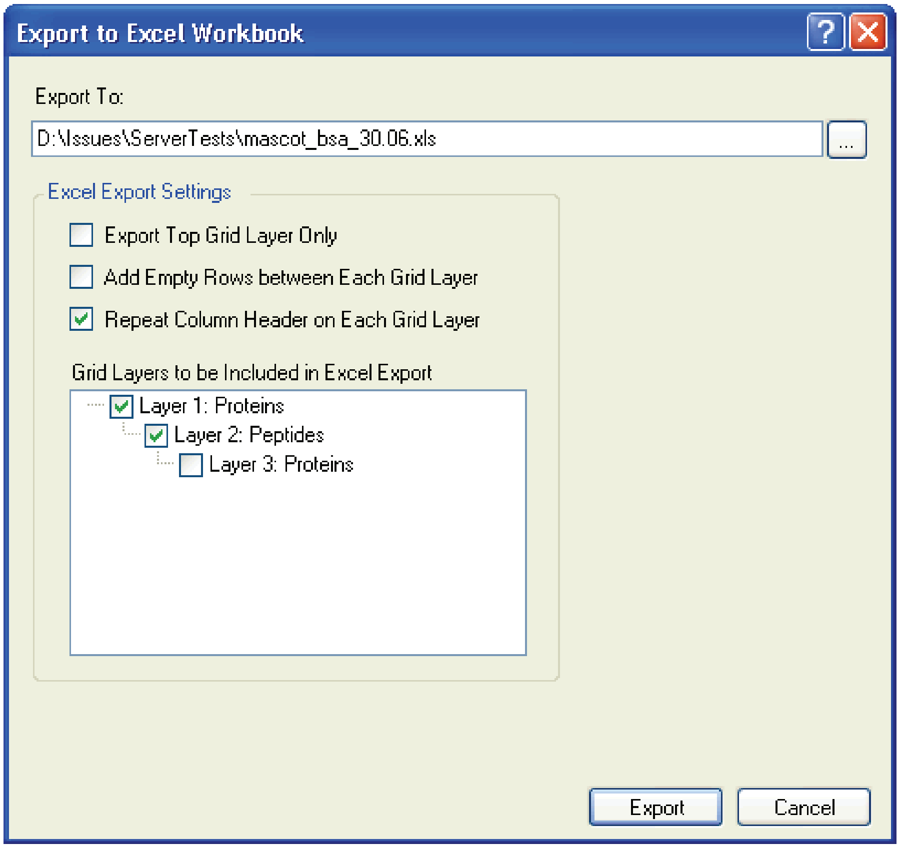
Save as type: Microsoft Office Excel Workbook (*.xls)

6. Browse and select a location to save the results file.

7. In the File name box, type the name of your results file.

8. Click **Open**. The Export to Excel Workbook dialog box appears.

9. To export only the information from the top layer of the current Results grid, select the **Export Top Grid Layer Only** check box in the Excel Export Settings area.
   –or–
   To export further grid layers and also the information from the top layer of the current Results grid, clear the **Export Top Grid Layer Only** check box, as shown in the next figure.



Export To:
D:\Issues\ServerTests\mascot_bsa_30.06.xls

Excel Export Settings

☐ Export Top Grid Layer Only
☐ Add Empty Rows between Each Grid Layer
☑ Repeat Column Header on Each Grid Layer

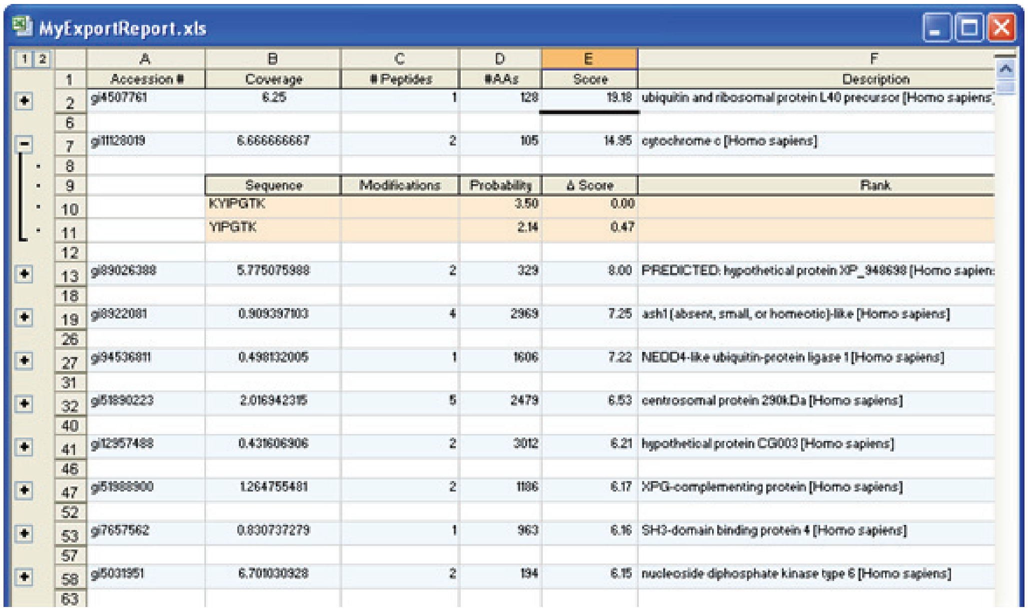
Grid Layers to be Included in Excel Export
- ☑ Layer 1: Proteins
  - ☑ Layer 2: Peptides
    - ☐ Layer 3: Proteins

10. To divide the layers from each other, select the **Add Empty Rows Between Each Grid Layer** check box.

11. Click **Export**. The status of the export appears. When the export is complete, you can open your exported file.

    See Figure 81 for an example of an exported report with two layers.

Figure 81. An exported report with two layers



MyExportReport.xls

	A	B	C	D	E	F
1	Accession #	Coverage	# Peptides	#AAs	Score	Description
2	gi4507761	6.25	1	128	19.18	ubiquitin and ribosomal protein L40 precursor [Homo sapiens
6						
7	gi11128019	6.666666667	2	105	14.95	cytochrome c [Homo sapiens]
8						
9		Sequence	Modifications	Probability	Δ Score	Rank
10		KYIPGTK		3.50	0.00	
11		YIPGTK		2.14	0.47	
12						
13	gi89026388	5.775075988	2	329	8.00	PREDICTED: hypothetical protein XP_948698 [Homo sapien
18						
19	gi8922081	0.909397103	4	2969	7.25	ash1 (absent, small, or homeotic)-like [Homo sapiens]
26						
27	gi94536811	0.498132005	1	1606	7.22	NEDD4-like ubiquitin-protein ligase 1 [Homo sapiens]
31						
32	gi51890223	2.016942315	5	2479	6.53	centrosomal protein 290kDa [Homo sapiens]
40						
41	gi12957488	0.431606906	2	3012	6.21	hypothetical protein CG003 [Homo sapiens]
46						
47	gi51988900	1.264755481	2	1186	6.17	XPG-complementing protein [Homo sapiens]
52						
53	gi7657562	0.830737279	1	963	6.16	SH3-domain binding protein 4 [Homo sapiens]
57						
58	gi5031951	6.701030928	2	194	6.15	nucleoside diphosphate kinase type 6 [Homo sapiens]
63						

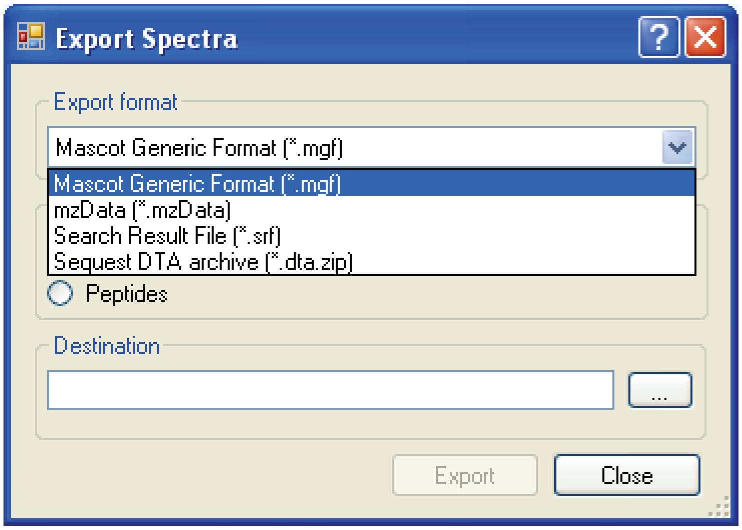
Exporting Spectra

Use the Export Spectra dialog box to select and save the search results or input them into another format.

❖ To export spectra

1. Open your search results. See "Understanding Reports and Views" on page 37.
2. From the Proteome Discoverer menu bar, choose **Search Report > Export Spectra**. The Export Spectra dialog box appears as shown in Figure 82.

Figure 82. Export Spectra dialog box



3. Click the browse button (**...**). The Save as dialog box appears.
4. In the File name box, type the name of the results file.
5. Click **Open**. The Export Spectra dialog box appears.
6. Click **Export**. The status of the export appears. When the export is complete, you can open your exported file.

Working with InforSense Discussion

Searching an LC-MS/MS raw file with a peptide-based search engine produces a list of identified proteins, along with peptide and statistical information. Determining the biological meaning of the results involves searching the literature in some manner to obtain pertinent information on each protein that is identified. InforSense workflows automate the process of searching the databases, which include the National Center for Biotechnology Information (NCBI, www.ncbi.nlm.nih.gov/sites/entrez) and the Swiss Institute of Bioinformatics ExPASy proteomics server (www.expasy.com).

WARNING Only SwissProt/Uniprot, NCBI, or IPI compatible protein databases yield search results compatible with InforSense Protein Annotation workflows.

Protein accessions denoted as International Protein Index (IPI), GenBank (GI), and SwissProt accession numbers or with TrEMBL names are all compatible with the workflows. Use of the ExPASy Web server requires that GI and IPI numbers be translated into the SwissProt/Uniprot format prior to submission; the NCBI workflows only needs to translate IPI accessions. The workflows facilitate the translations through the Protein Information Resource (PIR) Web service (http://pir.georgetown.edu/). TrEMBL names have undergone significant revisions in recent years and outdated names are not directly recognized by either of the Web services.

To ensure compatibility of search data obtained with older databases, both workflows submit any TrEMBL names they encounter to the automated ExPASy ID Tracker function (http://beta.uniprot.org/) to translate them into the latest representation prior to submission.

InforSense and the Internet

The InforSense Protein Annotation workflows are Internet intensive. Search speeds are a function of your Web service speed.

- Large result files run slower than small files.
- Higher network speeds result in faster search times.
- Loss of network connection, even intermittently (such as with a wireless connection), results in search failure.

Data Retrieval Content Levels

The richness of your data retrieval content depends on the content richness of the Web server. Information for each protein annotation searched might not be complete, due to the nature of the biological databases.

IMPORTANT For a more complete discussion, see the InforSense Protein Annotation Help written for Thermo Fisher Scientific Proteome Discoverer workflows.

Using InforSense Protein Annotation

Proteome Discoverer uses InforSense Protein Annotation to automatically determine the biological context of identified peptides.

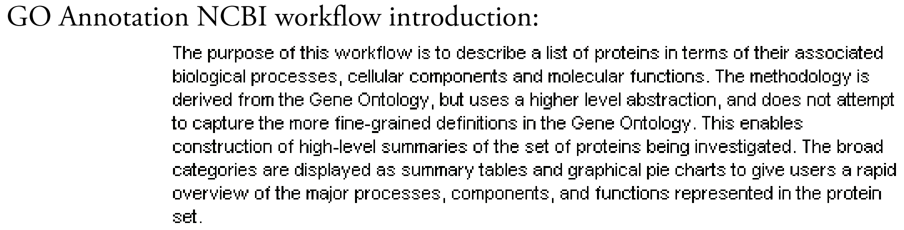
❖ To use InforSense

- Choose **Tools > InforSense**.

  InforSense opens a new window where you can choose four predefined workflow options:

  - GO Annotation NCBI
  - GO Annotation SwissProt
  - Metadata Table NCBI
  - Meta Table SwissProt

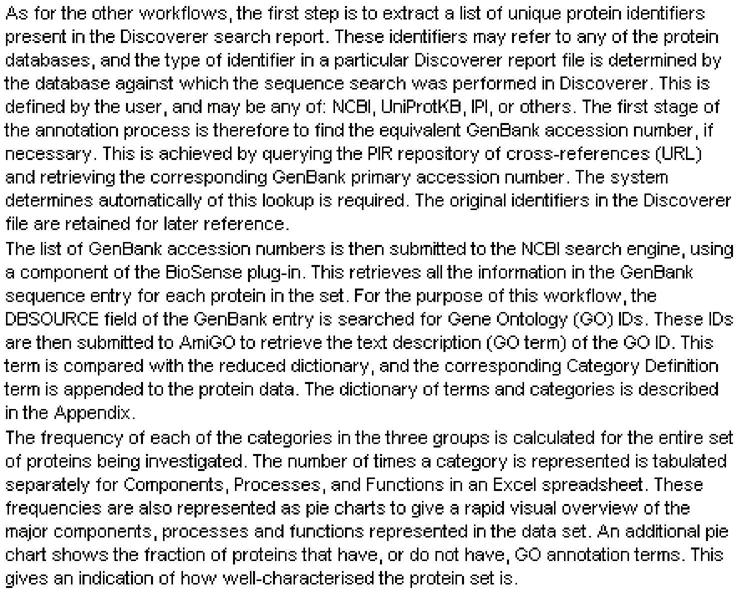
GO Annotation NCBI



GO Annotation NCBI workflow introduction:

The purpose of this workflow is to describe a list of proteins in terms of their associated biological processes, cellular components and molecular functions. The methodology is derived from the Gene Ontology, but uses a higher level abstraction, and does not attempt to capture the more fine-grained definitions in the Gene Ontology. This enables construction of high-level summaries of the set of proteins being investigated. The broad categories are displayed as summary tables and graphical pie charts to give users a rapid overview of the major processes, components, and functions represented in the protein set.

❖ Workflow GO Annotation NCBI description



As for the other workflows, the first step is to extract a list of unique protein identifiers present in the Discoverer search report. These identifiers may refer to any of the protein databases, and the type of identifier in a particular Discoverer report file is determined by the database against which the sequence search was performed in Discoverer. This is defined by the user, and may be any of: NCBI, UniProtKB, IPI, or others. The first stage of the annotation process is therefore to find the equivalent GenBank accession number, if necessary. This is achieved by querying the PIR repository of cross-references (URL) and retrieving the corresponding GenBank primary accession number. The system determines automatically of this lookup is required. The original identifiers in the Discoverer file are retained for later reference.

The list of GenBank accession numbers is then submitted to the NCBI search engine, using a component of the BioSense plug-in. This retrieves all the information in the GenBank sequence entry for each protein in the set. For the purpose of this workflow, the DBSOURCE field of the GenBank entry is searched for Gene Ontology (GO) IDs. These IDs are then submitted to AmiGO to retrieve the text description (GO term) of the GO ID. This term is compared with the reduced dictionary, and the corresponding Category Definition term is appended to the protein data. The dictionary of terms and categories is described in the Appendix.

The frequency of each of the categories in the three groups is calculated for the entire set of proteins being investigated. The number of times a category is represented is tabulated separately for Components, Processes, and Functions in an Excel spreadsheet. These frequencies are also represented as pie charts to give a rapid visual overview of the major components, processes and functions represented in the data set. An additional pie chart shows the fraction of proteins that have, or do not have, GO annotation terms. This gives an indication of how well-characterised the protein set is.

GO Annotation SwissProt

GO Annotation SwissProt workflow introduction:

The purpose of this workflow is to describe a list of proteins in terms of their associated biological processes, cellular components and molecular functions. The methodology is derived from the Gene Ontology, but uses a higher level abstraction, and does not attempt to capture the more fine-grained definitions in the Gene Ontology. This enables construction of high-level summaries of the set of proteins being investigated. The broad categories are displayed as summary tables and graphical pie charts to give users a rapid overview of the major processes, components, and functions represented in the protein set.

❖ Workflow GO Annotation SwissProt description

As for the other workflows, the first step is to extract a list of unique protein identifiers present in the Discoverer search report. These identifiers may refer to any of the protein databases, and the type of identifier in a particular Discoverer report file is determined by the database against which the sequence search was performed in Discoverer. This is defined by the user, and may be any of: NCBI, UniProtKB, IPI, or others. The first stage of the annotation process is therefore to find the equivalent UniProtKB accession number, if necessary. This is achieved by querying the PIR repository of cross-references (URL) and retrieving the corresponding UniProt primary accession number. The system determines automatically of this lookup is required. The original identifiers in the Discoverer file are retained for later reference.

The list of UniProt accession numbers is then submitted to the UniProt search engine (URL), using a component of the BioSense plug-in. This retrieves all the information in the UniProt sequence entry for each protein in the set. For the purpose of this workflow, the database cross-reference lines (DR lines) are searched for Gene Ontology (GO) terms. The text part of any GO terms found are then compared with the reduced dictionary, and the corresponding Category Definition term is appended to the protein data. The dictionary of terms and categories is described in the Appendix.

The frequency of each of the categories in the three groups is calculated for the entire set of proteins being investigated. The number of times a category is represented is tabulated separately for Components, Processes, and Functions in an Excel spreadsheet. These frequencies are also represented as pie charts to give a rapid visual overview of the major components, processes and functions represented in the data set. An additional pie chart shows the fraction of proteins that have, or do not have, GO annotation terms. This gives an indication of how well-characterised the protein set is.

Metadata Table NCBI

Metadata Table NCBI workflow introduction:

The purpose of this workflow is to automatically annotate a list of proteins using data retrieved for each protein from NCBI GenBank. Relevant fields are extracted from the GenBank entries, and reported in an Excel spreadsheet. In addition to reporting fields fields present in GenBank, a specialised lookup of terms in a Thermo-defined list of the major post-translational modifications is also extracted, and reported together with the sequence positions at which the modifications occur.

❖ Workflow Metadata Table NCBI description

The input to this workflow is a search report file from Discoverer in protXML format (ref, URL). A list of protein identifiers is extracted from the protXML file. These identifiers may refer to any of the protein databases, and the type of identifier in a particular Discoverer report file is determined by the database against which the sequence search was performed in Discoverer. This is defined by the user, and may be any of: NCBI, UniProtKB, IPI, or others. The first stage of the annotation process is therefore to find the equivalent GenBank accession number, if necessary. This is achieved by querying the PIR repository of cross-references and retrieving the corresponding GenBank primary accession number. The system determines automatically of this lookup is required. The original identifiers in the Discoverer file are retained for later reference.

The list of GenBank accession numbers is then submitted to GenBank using a component of the BioSense plug-in. This retrieves all the information in the GenBank sequence entry for each protein in the set.

Relevant fields from the returned sequence entries are then extracted and reported in the final Excel report. Two types of information are retrieved: 1. generic annotations extracted directly from the sequence entries, and 2. specific post-translational modifications (PTMs).

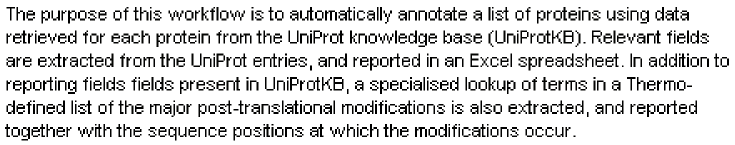
The generic annotations reported from a GenBank entry are:

1. accession number
2. GI number
3. protein description
4. comments
5. molecular weight (if reported)
6. PubMed cross-references
7. amino acid sequence
8. sequence length

The types of PTMs reported are listed in the PTM dictionary in the Appendix, and incorporated into the final report as follows. If any of the terms or keywords is found in the sequence entries retrieved from NCBI or UniProt, the corresponding category column is filled with the sequence locations of each occurrence in each protein.

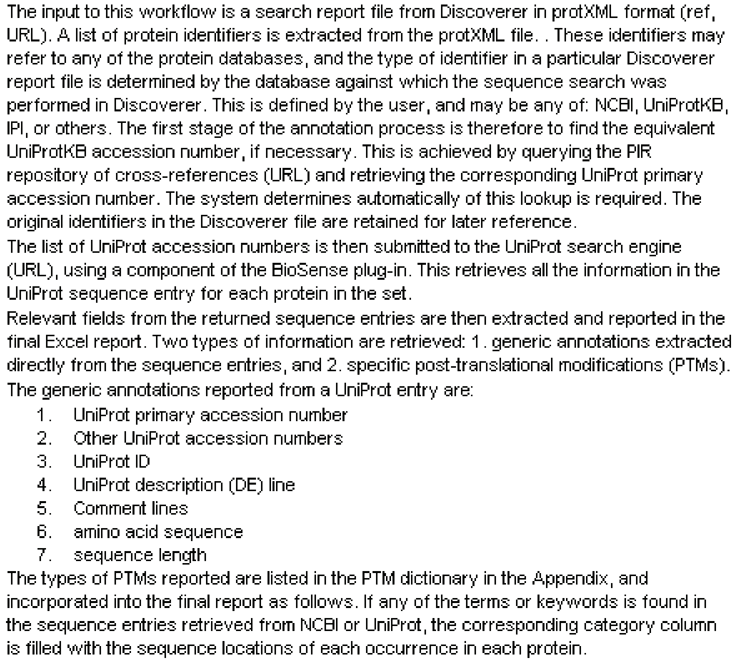
Meta Table SwissProt

The Meta Table SwissProt workflow introduction:



The purpose of this workflow is to automatically annotate a list of proteins using data retrieved for each protein from the UniProt knowledge base (UniProtKB). Relevant fields are extracted from the UniProt entries, and reported in an Excel spreadsheet. In addition to reporting fields fields present in UniProtKB, a specialised lookup of terms in a Thermo-defined list of the major post-translational modifications is also extracted, and reported together with the sequence positions at which the modifications occur.

❖ Workflow Metadata Table SwissProt description



The input to this workflow is a search report file from Discoverer in protXML format (ref, URL). A list of protein identifiers is extracted from the protXML file. . These identifiers may refer to any of the protein databases, and the type of identifier in a particular Discoverer report file is determined by the database against which the sequence search was performed in Discoverer. This is defined by the user, and may be any of: NCBI, UniProtKB, IPI, or others. The first stage of the annotation process is therefore to find the equivalent UniProtKB accession number, if necessary. This is achieved by querying the PIR repository of cross-references (URL) and retrieving the corresponding UniProt primary accession number. The system determines automatically of this lookup is required. The original identifiers in the Discoverer file are retained for later reference.

The list of UniProt accession numbers is then submitted to the UniProt search engine (URL), using a component of the BioSense plug-in. This retrieves all the information in the UniProt sequence entry for each protein in the set.

Relevant fields from the returned sequence entries are then extracted and reported in the final Excel report. Two types of information are retrieved: 1. generic annotations extracted directly from the sequence entries, and 2. specific post-translational modifications (PTMs).

The generic annotations reported from a UniProt entry are:

1. UniProt primary accession number
2. Other UniProt accession numbers
3. UniProt ID
4. UniProt description (DE) line
5. Comment lines
6. amino acid sequence
7. sequence length

The types of PTMs reported are listed in the PTM dictionary in the Appendix, and incorporated into the final report as follows. If any of the terms or keywords is found in the sequence entries retrieved from NCBI or UniProt, the corresponding category column is filled with the sequence locations of each occurrence in each protein.

4

Customizing Proteome Discoverer

Proteome Discoverer offers several ways to customize your search analysis experience:

- Customize the toolbar of the main window.
- Customize the quantitation methods available to use during a search.
- Customize the chemical modifications available to use during a search.
- Customize the cleavage reagents by modifying the reagents and their corresponding settings.
- Register a new .fasta file to use for your search.

Contents

Customizing the Toolbar

Proteome Discoverer comes with a number features and tools that are accessible with a click of a toolbar icon. The toolbar provides quick access to most of the commonly used Proteome Discoverer features. The following sections describe how you can change the display and the layout of the toolbar:

- Customizing the Toolbar Layout
- Customizing Toolbar Icons, Fonts, and Tooltips Display

Customizing the Toolbar Layout

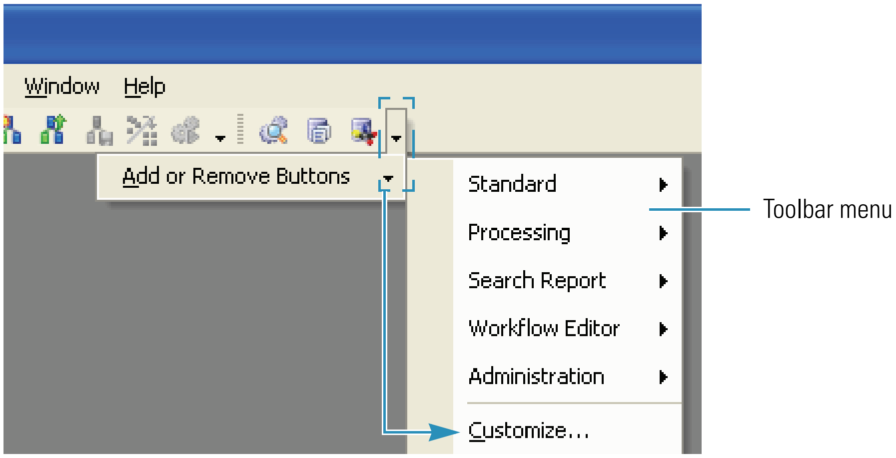
From the toolbar, you can add or remove access to other programs. These programs have the .exe extension. Toolbar buttons are available for most menu commands. You can also restore the toolbar to the default settings.

- Restoring Default Toolbar Settings
- Adding Shortcut Keys
- Removing and Repositioning Tools on the Toolbar

Restoring Default Toolbar Settings

❖ To restore your toolbar

1. On the toolbar, click ▼ to access the toolbar menu.

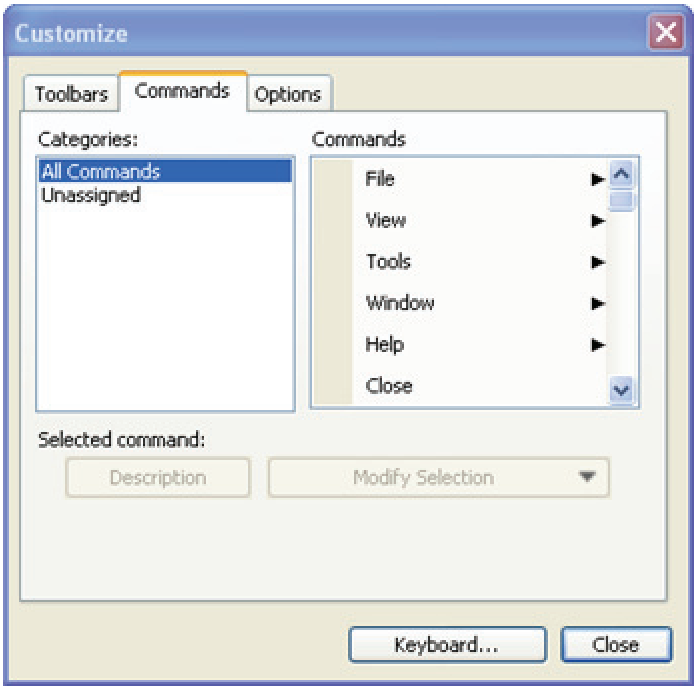


2. Choose **Customize**.
3. Select the menu you want to reset.
4. Click **Reset**. A message box appears to confirm your change.
5. Click **OK** to restore your menu selection.

Adding Shortcut Keys

❖ To add shortcut keys to a command

1. On the toolbar, click ▼ to access the toolbar menu.
2. Choose **Customize**. The Customize dialog box appears.
3. Click the **Commands** tab. The Commands page appears.

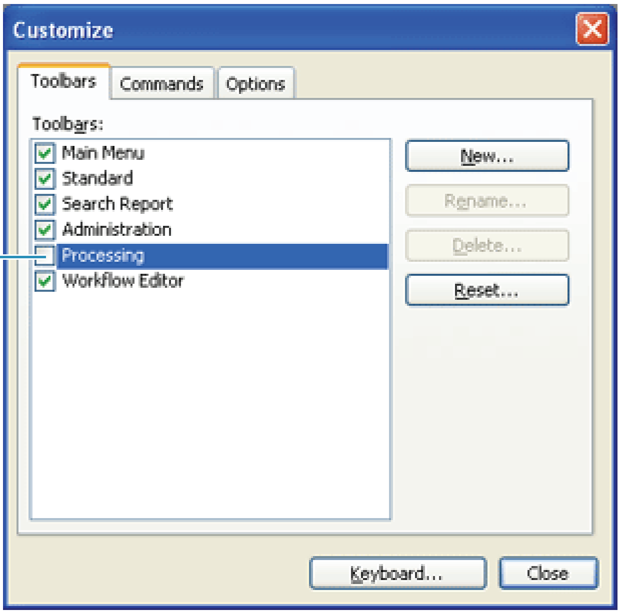


4. Click **Keyboard**. The Customize Keyboard dialog box opens.
5. In the Customize Keyboard dialog box, scroll through the Commands menu to find the command and its corresponding keyboard shortcut.
6. From the Specify a Shortcut list, select a shortcut.
7. Below the Specify a Shortcut list, check to see if the shortcut is unassigned.
8. If the shortcut is unassigned, click **Assign**. The shortcut is now assigned to your command choice.
9. In the Customize Keyboard dialog box, click **Close**. Your changes are saved.
10. In the Customize dialog box, click **Close**.

Removing and Repositioning Tools on the Toolbar

❖ To remove a set of tools from the toolbar

1. On the toolbar, click ▼ to access the toolbar menu.
2. Choose **Customize**. The Customize dialog box appears.
3. Clear the check box adjacent to the tool you want removed from the toolbar.



☑ Main Menu
☑ Standard
☑ Search Report
☑ Administration
☐ Processing
☑ Workflow Editor

Clear the check box to remove a tool from the toolbar.

The tool is visible in the menu, but the icon is removed from the toolbar.

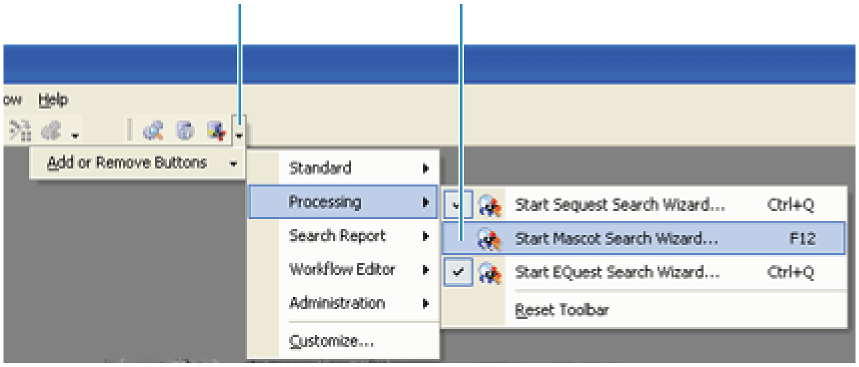
4. Click **Close** to close the Customize dialog box.

❖ To remove a single tool icon from the toolbar

1. On the toolbar, click ▼ to access the toolbar menu. The active toolbar commands have a checkmark next to them.

Access menu

Select tool to remove.



2. Clear the check mark adjacent to the tool you want removed from the toolbar. The tool is visible on the menu but the icon is removed from the toolbar.

❖ **To reposition toolbar buttons**

1. Select ⋮ immediately to the right of the toolbar section you want repositioned.

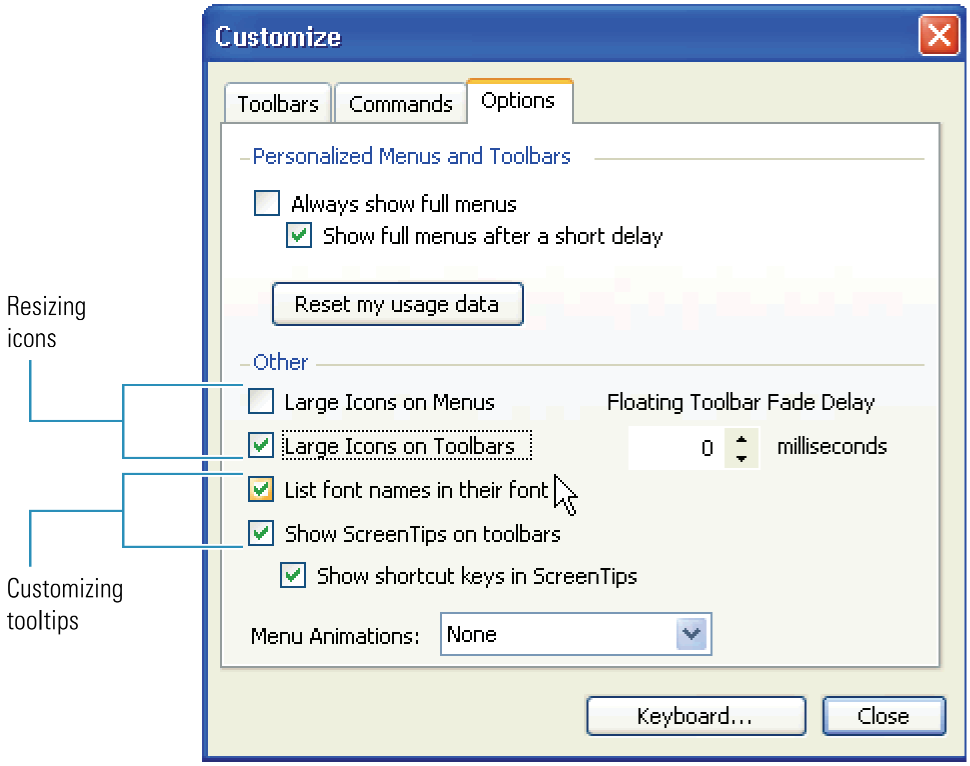
   A set of cross arrows (✥) appears when you select the toolbar section.

2. Drag the toolbar section to its new position.

Customizing Toolbar Icons, Fonts, and Tooltips Display

Use the Options page in the Customize dialog box to customize toolbar features. You can resize the toolbar icons and fonts, change your tooltips display to include corresponding icons and shortcut keys, or hide the tooltips display from view.

Figure 83. Customize dialog box



Personalized Menus and Toolbars

☐ Always show full menus
☑ Show full menus after a short delay

Other

☐ Large Icons on Menus
☑ Large Icons on Toolbars
☑ List font names in their font
☑ Show ScreenTips on toolbars
☑ Show shortcut keys in ScreenTips

Floating Toolbar Fade Delay: 0 milliseconds

Menu Animations: None

❖ To resize icons on toolbars and menus

1. On the toolbar, click ▼ to access the toolbar menu.
2. Choose **Customize**. The Customize dialog box appears.
3. Click **Options**.The Options page appears.
4. To increase the size of the menu and toolbar icons, select the **Large Icons on Menus** option and the **Large Icons on Toolbars** option.
5. To save your changes and close the Customize dialog box, click **Close**.

❖ To customize tooltips

1. On the toolbar, click ▼ to access the toolbar menu.
2. Choose **Customize**. The Customize dialog box appears.
3. Click **Options**.The Options page appears.
4. To display tooltips when you hover over the icons, select the **Show ScreenTips on toolbars** option.
5. (Optional) To display shortcut keys with the tooltips, select the **Show shortcut keys in ScreenTips** option. You must select Show ScreenTips on toolbars to use the Show shortcut keys in ScreenTips option.
6. To save your changes and close the Customize dialog box, click **Close**.

❖ To hide tooltips

1. On the toolbar, click ▼ to access the toolbar menu.
2. Choose **Customize**. The Customize dialog box appears.
3. To hide tooltips when you hover over the icons, clear the **Show ScreenTips on toolbars** option.
4. To save your changes and close the Customize dialog box, click **Close**.

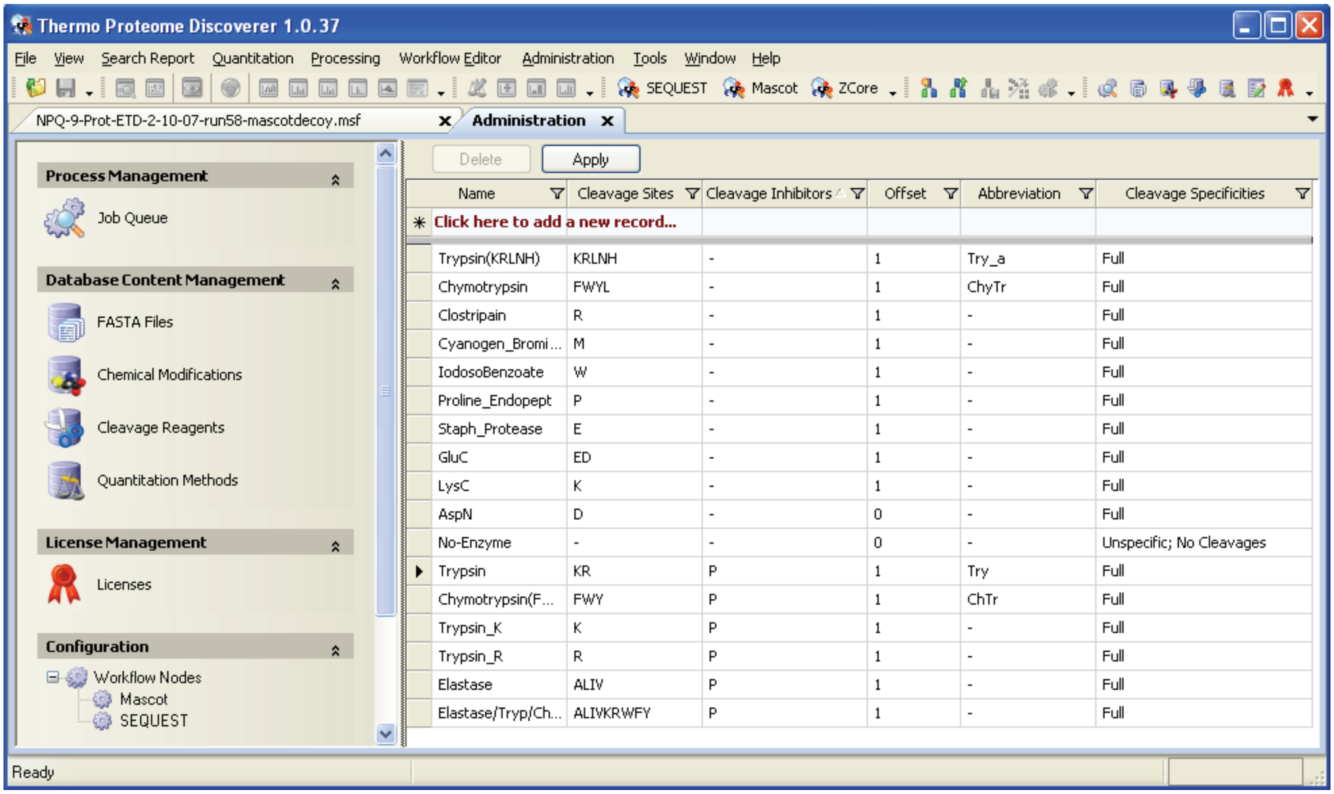
Customizing Cleavage Reagents

In the Cleavage Reagents view, you can add, remove, and modify the reagents and their corresponding settings. The reagents table contains the cleavage sites, cleavage inhibitors, abbreviations, and cleavage specificities.

❖ To open the Cleavage Reagents view

1. Choose **Administration > Cleavage Reagents**. The cleavage reagents view appears.

Figure 84. Cleavage reagents displayed in the Proteome Discoverer window



Name	Cleavage Sites	Cleavage Inhibitors	Offset	Abbreviation	Cleavage Specificities
Click here to add a new record...					
Trypsin(KRLNH)	KRLNH	-	1	Try_a	Full
Chymotrypsin	FWYL	-	1	ChyTr	Full
Clostripain	R	-	1	-	Full
Cyanogen_Bromi…	M	-	1	-	Full
IodosoBenzoate	W	-	1	-	Full
Proline_Endopept	P	-	1	-	Full
Staph_Protease	E	-	1	-	Full
GluC	ED	-	1	-	Full
LysC	K	-	1	-	Full
AspN	D	-	0	-	Full
No-Enzyme	-	-	0	-	Unspecific; No Cleavages
Trypsin	KR	P	1	Try	Full
Chymotrypsin(F…	FWY	P	1	ChTr	Full
Trypsin_K	K	P	1	-	Full
Trypsin_R	R	P	1	-	Full
Elastase	ALIV	P	1	-	Full
Elastase/Tryp/Ch…	ALIVKRWFY	P	1	-	Full

2. To modify the cleavage reagents list, click the Name column cell, **Click here to add**.

3. Type the information, such as **LysN**.

4. To save the modification, click **Apply**.

The following table lists the enzymes and reagents with cleavage properties that you can define when you add a new reagent.

Table 16. Enzymes and reagents with cleavage properties

Parameter	Description
Name	Specifies the name of the reagent used for the protein digestion.
Cleavage Sites	Specifies the position (amino acid) at which to cleave the sequence.
Cleavage Inhibitors	Specifies the amino acids that block cleavage when adjacent to the cleavage site.
Offset	Specifies whether the cleavage occurs before or after the amino acids listed in the Cleavage Sites column. • 0 - Cleavage occurs to the left of the amino acid. • 1 - Cleavage occurs to the right of the amino acid.
Abbreviation	Specifies the user-defined abbreviation.
Cleavage Specificities	You can select more than one cleavage specificity. The available cleavage specificities are as follows: • Full - Every cleavage must be at the specified cleavage site. • Semi - Only one end of the sequence needs to have the specified cleavage. • Semi (N-Term) - Only the N-terminal side of the sequence needs to have the specified cleavage. • Semi (C-Term) - Only the C-terminal side of the sequence needs to have the specified cleavage.

Customizing Chemical Modification Settings

With Proteome Discoverer, you can update the chemical modifications you use to conduct a peptide identification search. You can import a new list or the latest UNIMOD list. You can also modify the chemical modification list provided by adding amino acids to the modifications, by creating new modifications, or by activating or deactivating existing modifications.

Note A modification must be active to be usable during a search.

Use the chemical modification feature on the Administration page to customize the chemical modifications you use to do your search. Use the following list of tasks to update the chemical modifications:

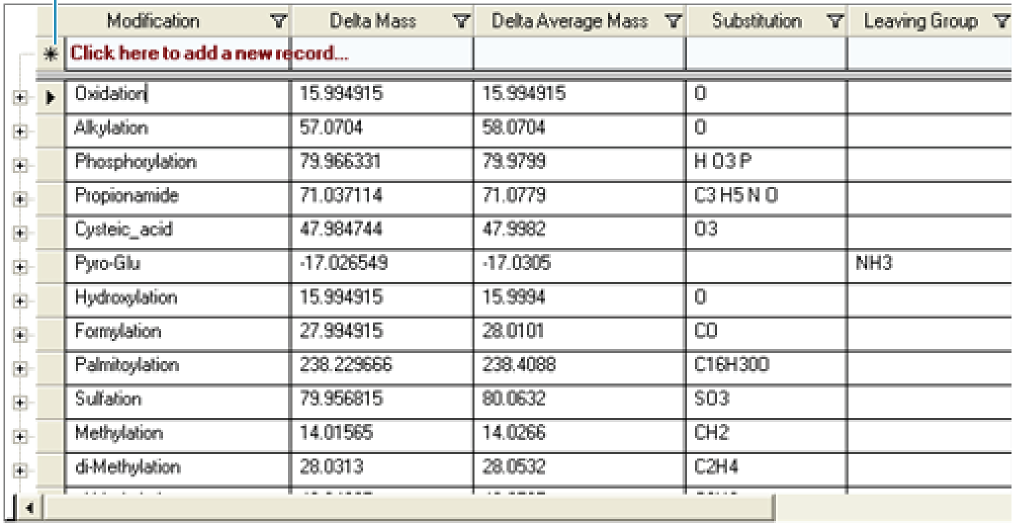
- Entering and Deleting Chemical Modifications
- Adding and Removing Amino Acids
- Importing Chemical Modifications

Entering and Deleting Chemical Modifications

❖ To add a new modification

1. Choose **Administration > Maintain Chemical Modifications**. The Chemical Modifications view appears.
2. Click the cell, **Click here to add a new record.**

   An empty row appears.



Click to add a row.

Modification	Delta Mass	Delta Average Mass	Substitution	Leaving Group
Click here to add a new record...				
Oxidation	15.994915	15.994915	O	
Alkylation	57.0704	58.0704	O	
Phosphorylation	79.966331	79.9799	H O3 P	
Propionamide	71.037114	71.0779	C3 H5 N O	
Cysteic_acid	47.984744	47.9982	O3	
Pyro-Glu	-17.026549	-17.0305		NH3
Hydroxylation	15.994915	15.9994	O	
Formylation	27.994915	28.0101	CO	
Palmitoylation	238.229666	238.4088	C16H30O	
Sulfation	79.956815	80.0632	SO3	
Methylation	14.01565	14.0266	CH2	
di-Methylation	28.0313	28.0532	C2H4	

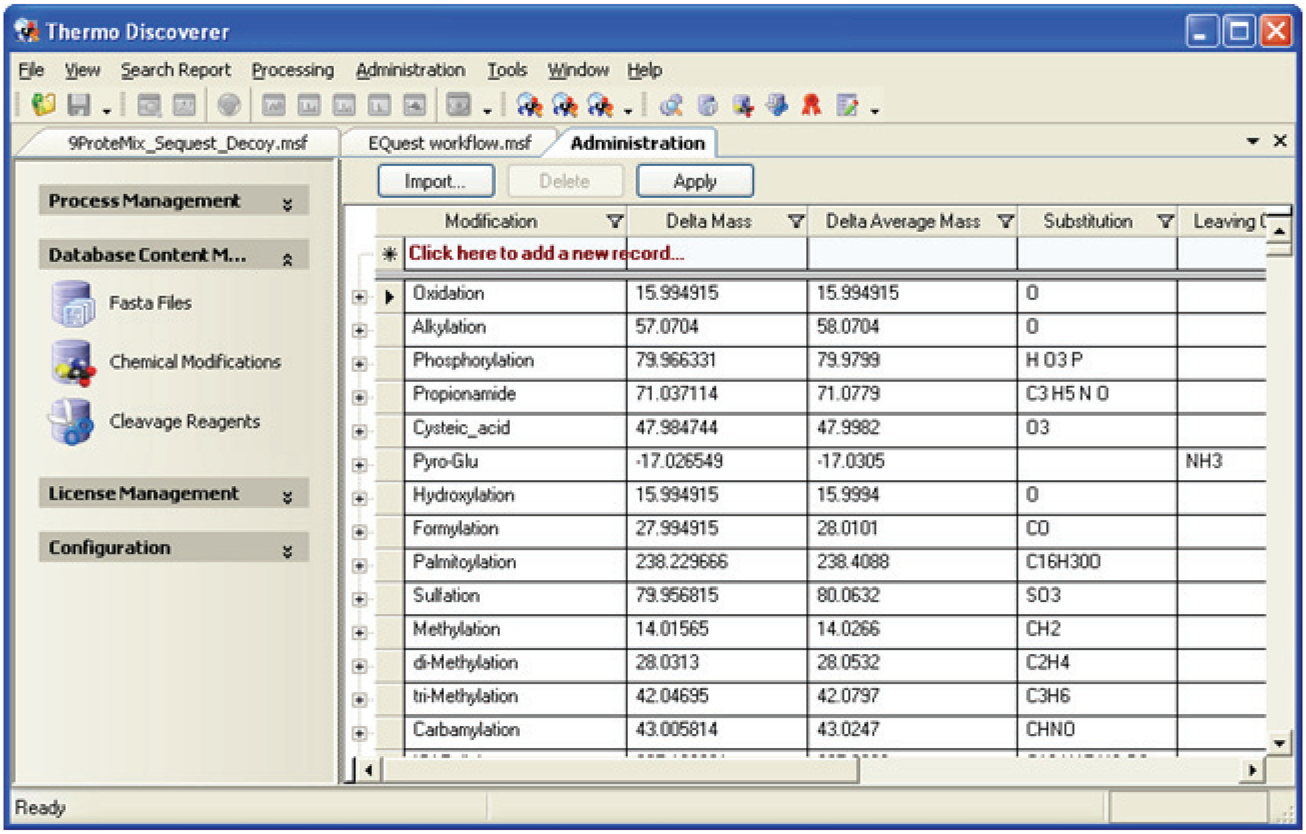
3. In the empty row, enter the name of the modification, the delta masses, the chemical substitution, the chemical group that is leaving, the position, and the abbreviations of the modifications.

   Note Both the substitution and leaving group are for display purposes only.

4. To accept the new modifications, click **Enter**.
5. Add an amino acid to the modifications. See "Adding and Removing Amino Acids" on page 141.

❖ To update an existing modification

1. Choose **Administration > Maintain Chemical Modifications**. The Chemical Modifications view appears.



	Modification	Delta Mass	Delta Average Mass	Substitution	Leaving
✳	Click here to add a new record...				
▶	Oxidation	15.994915	15.994915	O	
	Alkylation	57.0704	58.0704	O	
	Phosphorylation	79.966331	79.9799	H O3 P	
	Propionamide	71.037114	71.0779	C3 H5 N O	
	Cysteic_acid	47.984744	47.9982	O3	
	Pyro-Glu	-17.026549	-17.0305		NH3
	Hydroxylation	15.994915	15.9994	O	
	Formylation	27.994915	28.0101	CO	
	Palmitoylation	238.229666	238.4088	C16H30O	
	Sulfation	79.956815	80.0632	SO3	
	Methylation	14.01565	14.0266	CH2	
	di-Methylation	28.0313	28.0532	C2H4	
	tri-Methylation	42.04695	42.0797	C3H6	
	Carbamylation	43.005814	43.0247	CHNO	

2. In the Modification column, click the cell you want to update**.**
3. Type your changes for the delta masses, the substitution, the group it is leaving, the position, or the abbreviations of the modifications.
4. To accept the changes, click **Apply**.

❖ To delete a modification

1. Choose **Administration > Maintain Chemical Modifications**. The Chemical Modifications view appears.
2. Select the row of the modification you want to delete.
3. Click **Delete**. The row is removed from the chemical modifications table.

Related Topics

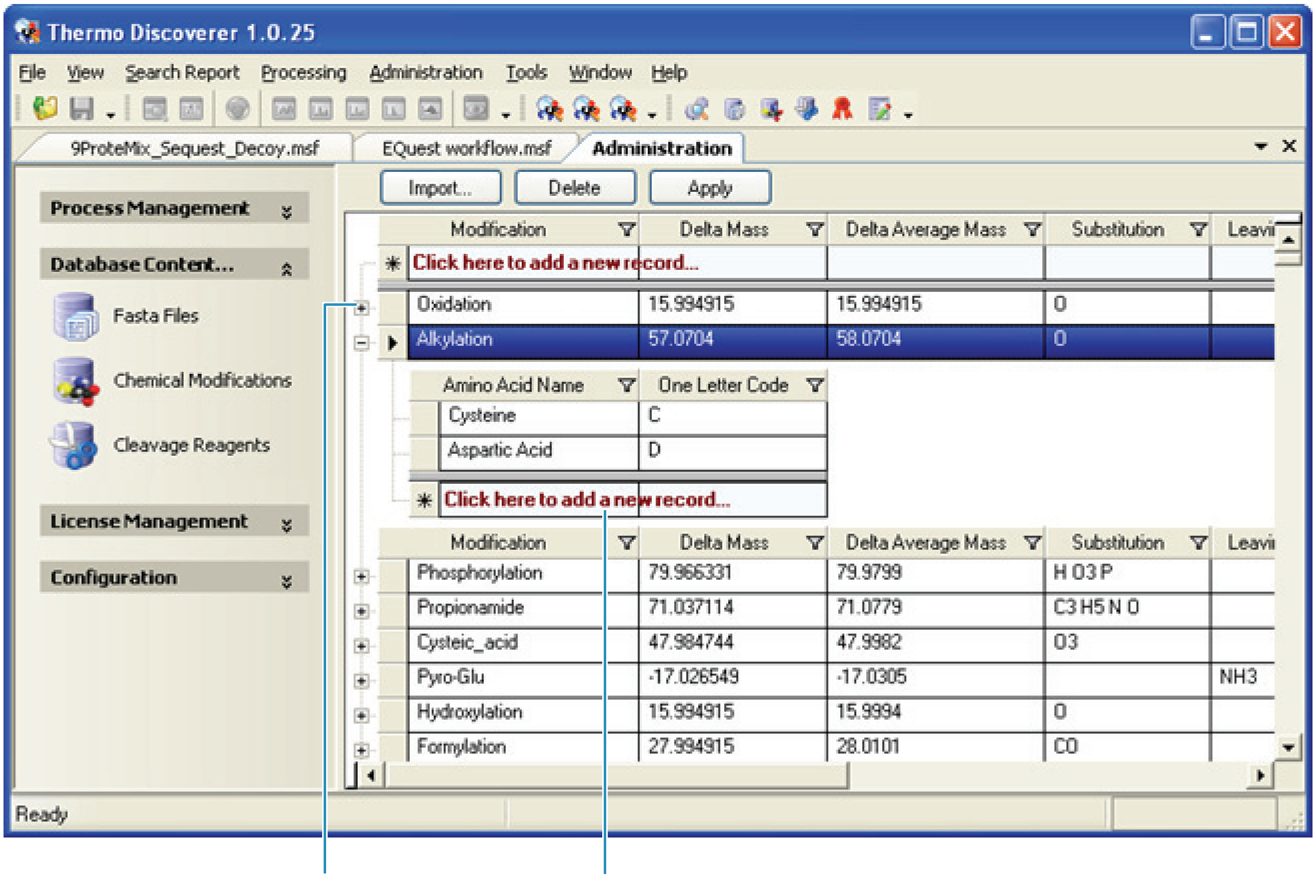
- Adding and Removing Amino Acids
- Importing Chemical Modifications

Adding and Removing Amino Acids

❖ To add an amino acid to a modification

1. Choose **Administration > Maintain Chemical Modifications**.

   The Chemical Modifications view appears.



Modification	Delta Mass	Delta Average Mass	Substitution	Leavi
Click here to add a new record...				
Oxidation	15.994915	15.994915	O	
Alkylation	57.0704	58.0704	O	

Amino Acid Name	One Letter Code
Cysteine	C
Aspartic Acid	D
Click here to add a new record...	

Modification	Delta Mass	Delta Average Mass	Substitution	Leavi
Phosphorylation	79.966331	79.9799	H O3 P	
Propionamide	71.037114	71.0779	C3 H5 N O	
Cysteic_acid	47.984744	47.9982	O3	
Pyro-Glu	-17.026549	-17.0305		NH3
Hydroxylation	15.994915	15.9994	O	
Formylation	27.994915	28.0101	CO	

Click to expand row.

Click to add a row.

2. Click **+** to the left of the modification row you want to update. The row expands and the associated amino acids appear.

3. Click the cell, **Click here to add a new record**. An empty row appears.

4. In the empty row, select the amino acid from the list. The amino acid and the one letter abbreviation appear.

5. To save the modifications, click **Apply**.

❖ To delete an amino acid from a modification

1. Choose **Administration > Maintain Chemical Modifications**. The Chemical Modifications view appears.
2. Click **+** to the left of the modification row 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 **Delete**. The row is removed from the chemical modifications table.

Related Topics

- Entering and Deleting Chemical Modifications
- Importing Chemical Modifications

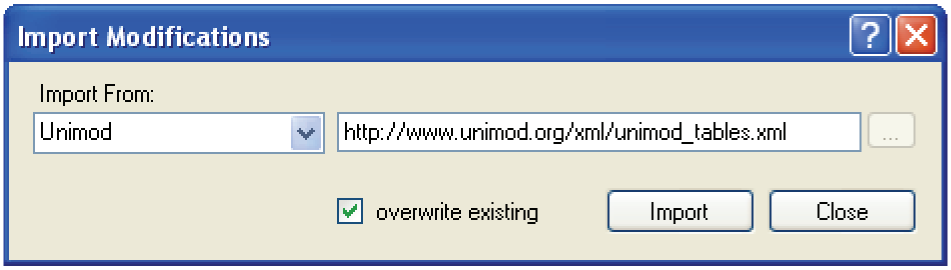
Importing Chemical Modifications

You can import chemical modifications from a local file or get an updated version from Unimod, a public domain database (www.unimod.org).

❖ To import chemical modifications from UNIMOD

1. Choose **Administration > Maintain Chemical Modifications**. The Chemical Modifications view appears.
2. Click **Import** and select **UNIMOD**.

   The UNIMOD URL appears.



Unimod

http://www.unimod.org/xml/unimod_tables.xml

☑ overwrite existing

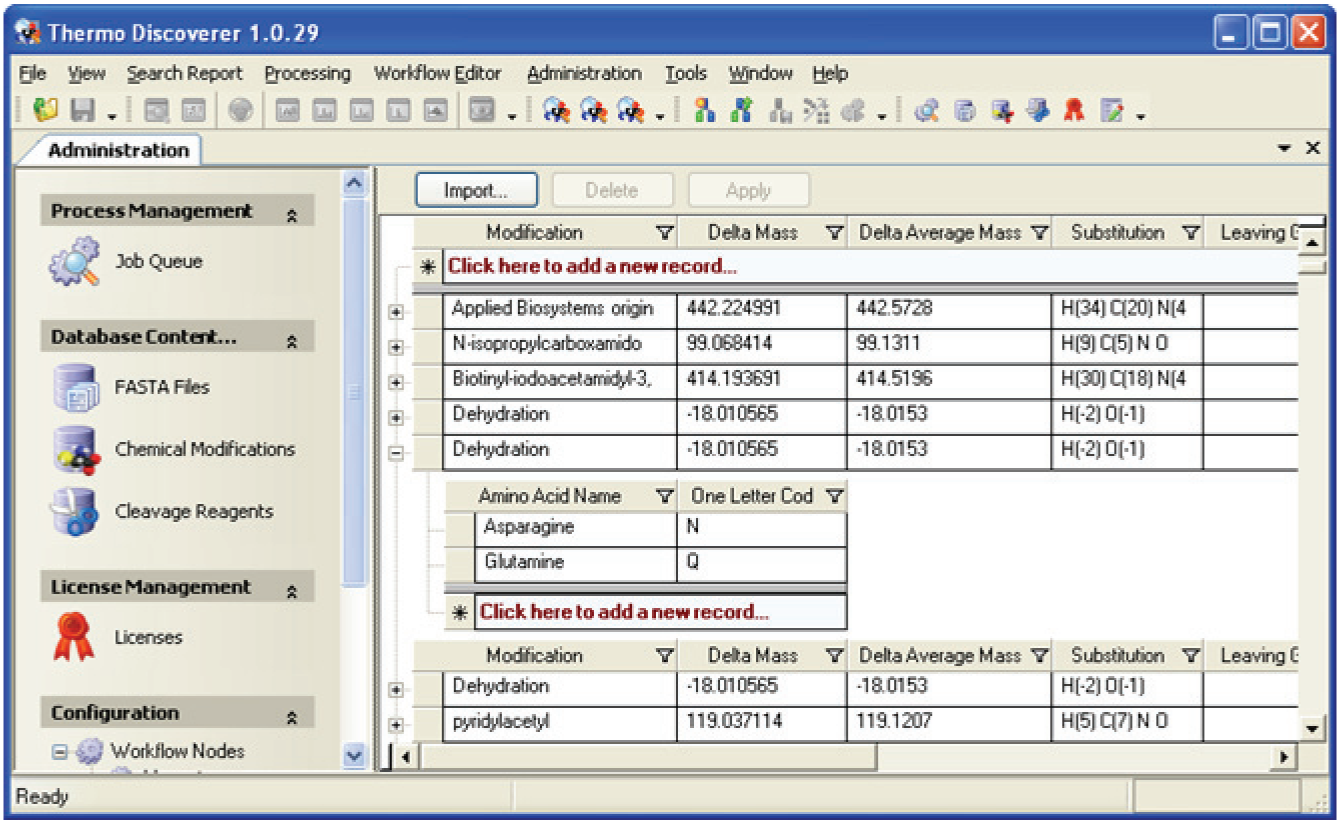
3. Click **Import**. A status message appears.
4. When the upload is complete, click **Close** .

❖ To import chemical modifications from a local file

1. Choose **Administration > Maintain Chemical Modifications**. The Chemical Modifications view appears.
2. Click **Import** and select **Local File**. Browse for your file.
3. Click **Import**. A status message appears.

4. When the upload is complete, click **Close**.

By using the Chemical Modifications view, you can add amino acids to existing modifications and create new modifications.



Modification	Delta Mass	Delta Average Mass	Substitution	Leaving G
Click here to add a new record...				
Applied Biosystems origin	442.224991	442.5728	H(34) C(20) N(4	
N-isopropylcarboxamido	99.068414	99.1311	H(9) C(5) N O	
Biotinyl-iodoacetamidyl-3,	414.193691	414.5196	H(30) C(18) N(4	
Dehydration	-18.010565	-18.0153	H(-2) O(-1)	
Dehydration	-18.010565	-18.0153	H(-2) O(-1)	

Amino Acid Name	One Letter Cod
Asparagine	N
Glutamine	Q
Click here to add a new record...	

Modification	Delta Mass	Delta Average Mass	Substitution	Leaving G
Dehydration	-18.010565	-18.0153	H(-2) O(-1)	
pyridylacetyl	119.037114	119.1207	H(5) C(7) N O	

Table 17. Chemical modifications parameters (Sheet 1 of 2)

Parameter	Description
Modification	Displays the type of modification.
Delta Mass	Displays the mass difference associated with the modification.
Delta Average Mass	Displays the difference between two adjacent average mass measurements.
Substitution	Displays the chemical group substitution that occurs with the modification.
Leaving group	Displays the chemical group that is missing.
Position	Displays the location where the modification might occur. The selections are Any, Any_C_Terminus, Any_N_Terminus, Protein_C_Terminus, and Protein_N_Terminus.
Abbreviation	Displays the abbreviation of the modification displayed in reports.

Table 17. Chemical modifications parameters (Sheet 2 of 2)

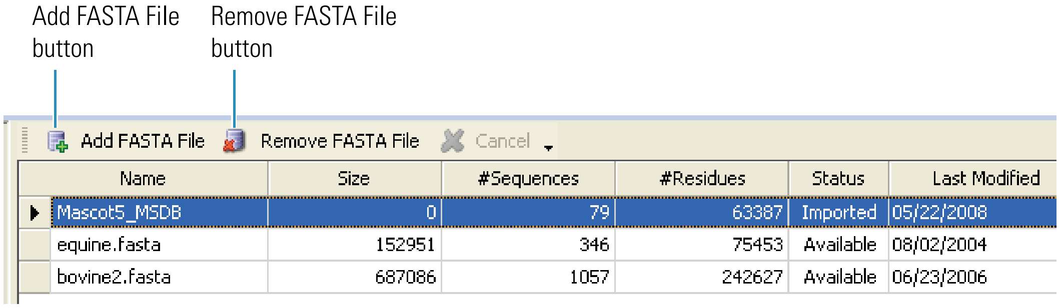
Parameter	Description
Amino Acid Name	Lists the amino acids that Proteome Discoverer can apply the modification to.
IsActive	Makes modifications available for search parameters.
UniMod Accession #	Go to http://www.unimod.org.
One letter code	Displays the single letter abbreviation of the amino acid.

Using FASTA Database Administration

When you use FASTA Database Utilities options, you can import FASTA files.

You can perform these actions in Proteome Discoverer and customize how your FASTA file is added to the modifications table.

Figure 85. FASTA file management



Name	Size	#Sequences	#Residues	Status	Last Modified
Mascot5_MSDB	0	79	63387	Imported	05/22/2008
equine.fasta	152951	346	75453	Available	08/02/2004
bovine2.fasta	687086	1057	242627	Available	06/23/2006

❖ To add a FASTA file

1. Choose **Administration > Maintain FASTA file**. The Administration page appears with the FASTA File view.
2. Click Add FASTA File.
3. Browse for and select the FASTA file that you want to process.
4. Click **OK**.

❖ To delete a FASTA file

1. Choose **Administration > Maintain FASTA file**. The Administration page appears with the FASTA File view.
2. Select ▶ at the beginning of a row to activate the row.
3. Click Remove FASTA File.

The amount of time it takes to process a FASTA file directly depends on the file size. When a FASTA file finishes processing, the status column displays the message, Available. The FASTA file is now available to be used for a peptide search with Proteome Discoverer.

Chemistry References

The following tables provide the mass values, the descriptions of enzyme properties, and the descriptions of fragment ions used throughout Proteome Discoverer:

- Amino Acid Mass Values
- Modification Values
- Enzyme Cleavage Properties
- Fragment Ions

Amino Acid Mass Values

Proteome Discoverer uses the amino acids symbols and mass values listed in Table 1 and Table 2.

Table 1. 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	57.02147	57.0517	C2H3NO
Alanine	A	Ala	71.03712	71.0787	C3H5NO
Serine	S	Ser	87.03203	87.078	C3H5NO2
Proline	P	Pro	97.05277	97.1168	C5H7NO
Valine	V	Val	99.06842	99.1328	C5H9NO
Threonine	T	Thr	101.04768	101.1051	C4H7NO2
Cysteine	C	Cys	103.00919	103.145	C3H5NOS
Isoleucine	I	Ile	113.08407	113.1598	C6H11NO
Leucine	L	Leu	113.08407	113.1598	C6H11NO
Asparagine	N	Asn	114.04293	114.1039	C4H6N2O2
Aspartic Acid	D	Asp	115.02695	115.0885	C4H5NO3
Glutamine	Q	Gln	128.05858	128.13091	C5H8N2O2
Lysine	K	Lys	128.09497	128.1745	C6H12N2O

Table 1. 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	129.0426	129.1156	$C5H7NO3$
Methionine	M	Met	131.0405	131.1994	$C5H9NOS$
Histidine	H	His	137.05891	137.1414	$C6H7N3O$
Phenylalanine	F	Phe	147.06842	147.1772	$C9H9NO$
Arginine	R	Arg	156.10112	156.188	$C6H12N4O$
Tyrosine	Y	Tyr	163.06332	163.17661	$C9H9NO2$
Tryptophan	W	Trp	186.07932	186.2141	$C11H10N2O$

Table 2. Special amino acids

Amino acid	One-letter code	Three-letter code	Monoisotopic mass	Average mass	Sum formula
Avrg. N/D	B	Bnd	114.53494	114.5962	$C4H5NO3$
Avrg. Q/E	Z	Zqe	128.55059	128.62326	$C5H7NO3$
Unknown acid (X)	X	Xxx	0	0	
Seleno cysteine (U)	U	Sec	150.0369	144.95959	$C3H5NOF$

Modification Values

Chemical modifications are listed in Table 3. See "Working with Chemical Modifications."

Table 3. Modification values (Sheet 1 of 2)

Modification	Monoisotopic mass	Average mass
Deamidation	0.98402	0.98480
Methylation	14.01565	14.02660
Hydroxylation	15.99492	15.99940
Oxidation	15.99492	15.99940
Formylation	27.99491	28.01010
Acetylation	42.01057	42.03670
Carboxyamidomethylation	57.02146	57.05130
Carboxymethylation	58.00548	58.03610
Propionamide	71.03711	71.07790
Sulfation	79.95682	80.06320

Table 3. Modification values (Sheet 2 of 2)

Modification	Monoisotopic mass	Average mass
Phosphorylation	79.96633	79.97990

Note Go to www.unimod.org/xml/unimod_tables.xml for more modification values.

Enzyme Cleavage Properties

Table 4 lists the enzymes and reagents with cleavage properties.

Table 4. Cleavage properties of enzymes and reagents

Enzymes/Reagents	Cleaves after	Cleaves before	Except when
Enzymes for digestion			
Trypsin K	K		P is after K
Trypsin R	R		P is after R
Trypsin (KR)	K or R		
Trypsin (KRLNH)	K, R, L, N, or H		
Trypsin (KR/P)	K or R		P is after K or R
Trypsin (KRLNH/P)	K, R, L, N, or H		P is after K, R, L, N, or H
Chymotrypsin	F, W, Y, or L		
Chymotrypsin (FWY)	F, W, or Y		P is after F, W, or Y
Clostripain	R		
Proline_Endopept	P		
Staph_protease	E		
GlucC	E or D		
LysC	K		
AspN		D	
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 A, L, I, V, K, R, W, F, or Y
Chemicals for degradation			
Cyanogen Bromide	M		
Iodobenzoate	W		

Fragment Ions

Fragment ions of peptides are produced by several different fragmentation techniques such as ECD, ETD, CID, higher-energy C-trap dissociation (HCD), and infrared multiphoton dissociation (IRMPD).

As an example, 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 N-CR 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. Fragment ions A, B, and C have a charge on the N-terminal side, and fragment ions X, Y, and Z 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^o, B^o, and C^o are ions that have lost water (-18 Da). The subscript next to the letter indicates the number of residues in the fragment ion.[1]

Table 5 summarizes the fragment ions used in Proteome Discoverer.

Table 5. Fragment ions

Ions	**Description**
A	Charge on N-terminal side
B	Charge on N-terminal side
C	Charge on N-terminal side
Y	Charge on C-terminal side
Z	Charge on C-terminal side
B*	B ion that has lost ammonia (-17 Da)
Y*	Y ion that has lost ammonia (-17 Da)
B^o	B ion that has lost water (-18 Da)
Y^o	Y ion that has lost water (-18 Da)

---

[1] For more information on fragment ions and nomenclature, see *Proposal for a Common Nomenclature for Sequence Ions in Mass Spectra of Peptides*; Roepstorff, P. and Fohlman, J.; Biomed Mass Spectrum, 11(11) 601 (1984).

B

FASTA Reference

This FASTA reference contains the mass values, the descriptions of enzyme properties, and the fragment ion descriptions used throughout Proteome Discoverer. It is an overview of the most important FASTA databases and the parsing rules Proteome Discoverer uses to obtain the accession# and the description.

NCBI

This non-redundant database is compiled by the NCBI (National Center for Biotechnology Information) as a protein database for Blast searches. It contains non-identical sequences from GenBank CDS translations, PDB, Swiss-Prot, PIR, and PRF.

http://www.ncbi.nih.gov/BLAST/blast_databases.html

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`

MSIPI

MSIPI is a database derived from IPI that contains additional information about cSNPs, N-terminus peptides, and known variants in a format suitable for mass spectrometry search engines. MSIPI is produced by the Max-Planck Institute for Biochemistry at Martinsried and the University of Southern Denmark, and distributed by the European Bioinformatics Institute (EBI).

ftp://ftp.ebi.ac.uk/pub/databases/IPI/msipi/current/

A typical MSIPI title line follows:

```
>MSIPI:IPI00000001.2| Gene_Symbol=STAU1 Isoform Long of
Double-stranded RNA-bin ding protein Staufen homolog 1 lng=577 #
CON[595,R,359,A] #
```

`FASTA ID:`

- `Accession#:  IPI00000001.2`
- `Description:  Isoform Long of Double-stranded RNA-bin ding protein Staufen homolog 1 lng=577 # CON[595,R,359,A] #`

IPI

IPI (International Protein Index) is compiled by the EBI to provide a top-level guide to the main databases that describe the human and mouse proteomes: SWISS-PROT, TrEMBL, NCBI RefSeq, and Ensembl.

http://www.ebi.ac.uk/IPI/

ftp://ftp.ebi.ac.uk/pub/databases/IPI/current/

A typical IPI title line follows:

```
>IPI:IPI00685094.1|SWISS-PROT:Q2KIJ2|ENSEMBL:ENSBTAP00000028878|REFSEQ:NP
_001073825;XP_593190 Tax_Id=9913 Gene_Symbol=MGC137286;LOC515210
Uncharacterized protein C1orf156 homolog
```

`FASTA ID:`

- `Accession#:  IPI00685094.1`
- `Description:  Uncharacterized protein C1orf156 homolog`

UniRef100

UniRef, also known as UniProt NREF, is a set of comprehensive protein databases curated by the Universal Protein Resource consortium. UniRef100 contains only non-identical sequences, where as UniRef90, and UniRef50 are non-redundant at a sequence similarity level of 90 percent and 50 percent, respectively.

http://www.ebi.ac.uk/uniref/

ftp://ftp.uniprot.org/pub/databases/uniprot/uniref/uniref100/uniref100.fasta.gz

A typical UniRef100 title line follows:

```
>UniRef100_Q4U9M9 Cluster: 104 kDa microneme-rhoptry antigen precursor;
n=1; Theileria annulata|Rep: 104 kDa microneme-rhoptry antigen precursor -
Theileria annulata
```

`FASTA ID:`

- `Accession#:   Q4U9M9`
- `Description:  Cluster: 104 kDa microneme-rhoptry antigen precursor; n=1; Theileria annulata|Rep: 104 kDa microneme-rhoptry antigen precursor - Theileria annulata`

SwissProt & TrEMBL

The SwissProt database is developed by the SWISS-PROT groups at SIB and EBI.

TrEMBL is a computer-annotated supplement of SWISS-PROT that contains all the translations of EMBL nucleotide sequence entries not yet integrated in SWISS-PROT.

http://www.expasy.org/sprot/

ftp://ftp.expasy.org/databases/uniprot/knowledgebase/uniprot_sprot.fasta.gz

ftp://ftp.ebi.ac.uk/pub/databases/uniprot/knowledgebase/uniprot_trembl.fasta.gz

A typical SwissProt title line follows:

```
>Q43495|108_SOLLC Protein 108 precursor - Solanum lycopersicum (Tomato)
(Lycopersicon esculentum)
```

`FASTA ID:        108_SOLLC`

- `Accession#:   Q43495`
- `Description:  Protein 108 precursor - Solanum lycopersicum (Tomato) (Lycopersicon esculentum)`

MSDB

The MSDB database is compiled by the Proteomics Group at Imperial College London using the following source databases: PIR, Trembl, GenBank, Swiss-Prot, and NRL3D.

http://csc-fserve.hh.med.ic.ac.uk/msdb.html

ftp://ftp.ncbi.nih.gov/repository/MSDB/

A typical MSDB title line follows:

```
>CBMS Ubiquinol-cytochrome-c reductase (EC 1.10.2.2) cytochrome b - mouse
mitochondrion
```

`FASTA ID:`

- `Accession#:   CBMS`
- `Description:  Ubiquinol-cytochrome-c reductase (EC 1.10.2.2) cytochrome b - mouse mitochondrion`

Custom Database Support

Proteome Discoverer also has two "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

This parsing rule is used, 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, would look like this one:

```
>tr|Q18FC3|Q18FC3_HALWD IS1341-type transposase - Haloquadratum walsbyi
(strain DSM 16790).
```

```
FASTA ID:        Q18FC3_HALWD
 • Accession#:   Q18FC3
 • Description:  IS1341-type transposase - Haloquadratum walsbyi (strain
   DSM 16790).
```

Custom Parsing Rule B

This parsing rule is used if the accession# and the description are separated by using a whitespace. A typical FASTA title line, which matches this parsing rule, would look like this one:

```
>HP0001 hypothetical protein {Helicobacter pylori 26695}
```

```
FASTA ID:
 • Accession#:   HP0001
 • Description:  hypothetical protein {Helicobacter pylori 26695}
```

Custom Parsing Rule C

This parsing rule is used if the FASTA title line only contains the accession#. A typical FASTA title line, which matches this parsing rule, would look like this one:

`>143B_HUMAN`

`FASTA ID:`

- `Accession#:   143B_HUMAN`
- `Description:  143B_HUMAN`

Index