Xcalibur

Proteome Discoverer

User Guide

XCALI-97232 Revision A July 2008
# Contents

**Preface** ................................................................. vii  
Related Documentation ............................................... vii  
System Requirements ................................................... viii  
Licenses .................................................................. viii  
Safety and Special Notices ................................................. x  
Contacting Us ........................................................... xi

## Chapter 1 Getting to Know Proteome Discoverer  ..................................................... 1  
Proteome Discoverer User Interface .............................................. 1  
  Main Window and Navigation ................................................. 2  
  Menus .................................................................. 4  
Understanding Proteome Discoverer ........................................... 11  
Proteome Discoverer Features ................................................ 12  
Using Proteome Discoverer .................................................... 13  
Search Algorithms Overview ................................................... 14  
  Overview of Processing, Analyzing, and Interpreting Raw Data .... 14  
  Mascot Search Algorithm .............................................. 16  
  ZCore Search Algorithm .............................................. 16  
  SEQUEST* Search Algorithm ............................................ 17  
Raw Data, Results, Reports, and Analysis .................................... 18  
  Databases and Sample Files .............................................. 18  
  Proteome Discoverer Input and Output File Types ................... 18  
Proteome Discoverer Workflow ................................................ 19  
Quantitation Overview .......................................................... 20  
  About iTRAQ Quantitation .............................................. 20  
  Peptide Ratio Calculation .............................................. 20  
  Show Reporter Intensities .............................................. 21  
  Quantitation Spectrum .................................................. 21  
Qual Browser Overview ........................................................ 22  
InforSense Protein Annotation Discussion .................................. 23  
  InforSense Workflow Types ............................................ 23  
  Introduction to GO Ontology ........................................... 24  
  Introduction to Metadata Retrieval .................................... 24
## Chapter 2  Using Proteome Discoverer  .................................. 25
How Proteome Discoverer Fits Into Your Lab Workflow .................. 26
Overview of Menu Options .................................................. 27
Start Proteome Discoverer .................................................. 27
File Menu ...................................................................... 27
Working with Search Wizards ................................................. 29
Search Results, Reports, and Analysis ......................................... 36
Understanding Reports and Views ............................................. 37
Search Summary Page ........................................................ 38
Protein Grid .................................................................... 39
Creating a Report from Multiple Results ..................................... 41
Peptide Grid ................................................................... 43
Peptide Confidence Page ....................................................... 46
Search Input Page ............................................................. 46
Filters Page .................................................................... 47
Using the Proteome Discoverer Workflow Editor ......................... 49
Workflow Workspace .......................................................... 51
Creating and Saving a Search Workflow ..................................... 53
Discoverer Workflow Nodes Details .......................................... 57
Working with Reports .......................................................... 59
Working with the Search Job Queue ......................................... 60
Checking Your Job Queue Search Status ................................... 60
Deleting Items from the Queue List .......................................... 61
Using FASTA Database Utilities ............................................. 62
FASTA Files .................................................................. 63
Adding Protein References to FASTA ....................................... 64
Performing a Decoy Database Search ....................................... 64
Working with Chemical Modifications ........................................ 66
How to Quantify Your Data Using Administration and Workflow ..... 67
  Reporter Ion Based Quantitation in Proteome Discoverer .......... 67
  Quantitation Summary ...................................................... 70
  Assess the Abundance of the Precursor ................................. 72
Using Qual Browser ............................................................ 72
Job Queue ...................................................................... 73
  Using the Job Queue ....................................................... 73
  Open a Completed Report From Job Queue ........................... 74
InforSense ..................................................................... 75

## Chapter 3  Interpreting Search Results  ................................... 77
Working with the Initial Results Report ...................................... 78
  Organizing Rows and Columns .......................................... 78
  Saving and Applying Results Report Layout Changes ............... 82
## Appendix A Chemistry References

- Amino Acid Mass Values .................................................. 147
- Modification Values ....................................................... 148
- Enzyme Cleavage Properties .............................................. 149
- Fragment Ions ............................................................... 150

## Appendix B FASTA Reference

- NCBI ................................................................. 151
- MSIP ................................................................. 152
- IPI ................................................................. 152
- UniRef100 .......................................................... 153
- SwissProt & TrEMBL ................................................ 153
- MSDB ................................................................. 154
- Custom Database Support ............................................. 154
  - Custom Parsing Rule A ........................................... 154
  - Custom Parsing Rule B .......................................... 154
  - Custom Parsing Rule C .......................................... 155

## Index ................................................................. 157
Preface

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

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.

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

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

<table>
<thead>
<tr>
<th>Feature</th>
<th>License</th>
<th>Barcode</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discoverer_Base</td>
<td>will expire on 07/16/2008</td>
<td>ABCC-5678-1234</td>
</tr>
<tr>
<td>Discoverer Viewer</td>
<td>permanent</td>
<td></td>
</tr>
<tr>
<td>InforSense</td>
<td>will expire on 07/16/2008</td>
<td></td>
</tr>
</tbody>
</table>
3. In the User Information pane, type your contact information.

**Figure 2.** License User Information pane

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
<td>Ashley Johnson</td>
</tr>
<tr>
<td>Company</td>
<td>Thermo</td>
</tr>
<tr>
<td>Street</td>
<td>156 River Oaks</td>
</tr>
<tr>
<td>City</td>
<td>San Jose</td>
</tr>
<tr>
<td>State</td>
<td>CA</td>
</tr>
<tr>
<td>Zip Code</td>
<td>95134</td>
</tr>
<tr>
<td>Country</td>
<td>USA</td>
</tr>
<tr>
<td>Telephone</td>
<td>408-965-6000</td>
</tr>
</tbody>
</table>

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

**Figure 3.** License Host Information pane

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.
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.
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
- Understanding Proteome Discoverer
- Proteome Discoverer Features
- Using Proteome Discoverer
- Search Algorithms Overview
- Raw Data, Results, Reports, and Analysis
- Proteome Discoverer Workflow
- Quantitation Overview
- Qual Browser Overview
- InforSense Protein Annotation Discussion

### 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

<table>
<thead>
<tr>
<th>Column Chooser</th>
<th>Results table</th>
<th>File status</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Proteins</strong></td>
<td><strong>Results table</strong></td>
<td><strong>File status</strong></td>
</tr>
<tr>
<td>Accession</td>
<td>Coverage</td>
<td># Peptides</td>
</tr>
<tr>
<td>g129293</td>
<td>25.65 %</td>
<td>25</td>
</tr>
<tr>
<td>g115453</td>
<td>23.55 %</td>
<td>13</td>
</tr>
<tr>
<td>g545894</td>
<td>14.99 %</td>
<td>23</td>
</tr>
<tr>
<td>g117970</td>
<td>40.98 %</td>
<td>39</td>
</tr>
<tr>
<td>g127630</td>
<td>30.02 %</td>
<td>9</td>
</tr>
<tr>
<td>g1345987</td>
<td>3.28 %</td>
<td>6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sequence</th>
<th># Proteins</th>
<th># Protein Groups</th>
<th>Activation Type</th>
<th>Modifications</th>
<th>IonScore</th>
<th>Exp Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>YLVEIR</td>
<td>3</td>
<td>2</td>
<td>ETD</td>
<td></td>
<td>45</td>
<td>7.0E-003</td>
</tr>
<tr>
<td>YLVEIR</td>
<td>3</td>
<td>2</td>
<td>ETD</td>
<td></td>
<td>42</td>
<td>1.6E-002</td>
</tr>
<tr>
<td>TRAILG</td>
<td>1</td>
<td>1</td>
<td>ETD</td>
<td></td>
<td>22</td>
<td>1.6E+000</td>
</tr>
<tr>
<td>TRAILG</td>
<td>1</td>
<td>1</td>
<td>ETD</td>
<td></td>
<td>12</td>
<td>1.7E+001</td>
</tr>
<tr>
<td>TRAILG</td>
<td>1</td>
<td>1</td>
<td>ETD</td>
<td></td>
<td>11</td>
<td>1.9E+001</td>
</tr>
<tr>
<td>AATL</td>
<td>1</td>
<td>1</td>
<td>ETD</td>
<td>T5(Phos)</td>
<td>11</td>
<td>1.0E+05</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Accession</th>
<th>Coverage</th>
<th># Peptides</th>
<th>#PAs</th>
<th>Score</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>g1273219</td>
<td>4.22 %</td>
<td>37</td>
<td>344</td>
<td>49.64</td>
<td>[U54883] polypeptide [Rabbit hemorrhagic disease virus] [MASS=24065.8]</td>
</tr>
<tr>
<td>g736283</td>
<td>5.42 %</td>
<td>42</td>
<td>321</td>
<td>49.64</td>
<td>[U20814] ORF1 [Rabbit hemorrhagic disease virus] [MASS=25712.2]</td>
</tr>
<tr>
<td>g67053</td>
<td>4.51 %</td>
<td>38</td>
<td>324</td>
<td>49.64</td>
<td>[Q5P898] polypeptide [Rabbit hemorrhagic disease virus] [MASS=257]</td>
</tr>
<tr>
<td>g7769710</td>
<td>5.22 %</td>
<td>42</td>
<td>333</td>
<td>49.64</td>
<td>[AP288415] polypeptide [Rabbit hemorrhagic disease virus] [MASS=257]</td>
</tr>
<tr>
<td>g425877</td>
<td>8.50 %</td>
<td>3</td>
<td>153</td>
<td>48.75</td>
<td>MYOGLOBIN</td>
</tr>
</tbody>
</table>
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
• Right-click anywhere in a report to display its shortcut menu as shown in Figure 4.

**Figure 4. Report shortcut menu**

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
<th>Menu access</th>
<th>Shortcut key access</th>
<th>Toolbar access</th>
</tr>
</thead>
<tbody>
<tr>
<td>Show Top Match Peptides Only</td>
<td></td>
<td>File &gt; Open Report</td>
<td>CTRL+O</td>
<td></td>
</tr>
<tr>
<td>Show Peptide Groups</td>
<td></td>
<td>File &gt; Open Report</td>
<td>CTRL+O</td>
<td></td>
</tr>
<tr>
<td>Show Filtered Out Rows</td>
<td></td>
<td>File &gt; Open Report</td>
<td>CTRL+O</td>
<td></td>
</tr>
<tr>
<td>Enable Protein Grouping</td>
<td></td>
<td>File &gt; Open Report</td>
<td>CTRL+O</td>
<td></td>
</tr>
<tr>
<td>Mass Tolerance Unit</td>
<td></td>
<td>File &gt; Open Report</td>
<td>CTRL+O</td>
<td></td>
</tr>
<tr>
<td>Row Numbers</td>
<td></td>
<td>File &gt; Open Report</td>
<td>CTRL+O</td>
<td></td>
</tr>
<tr>
<td>Copy</td>
<td></td>
<td>File &gt; Open Report</td>
<td>CTRL+O</td>
<td></td>
</tr>
<tr>
<td>Copy With Column Headers</td>
<td></td>
<td>File &gt; Open Report</td>
<td>CTRL+O</td>
<td></td>
</tr>
<tr>
<td>Export to Excel Workbook</td>
<td></td>
<td>File &gt; Open Report</td>
<td>CTRL+O</td>
<td></td>
</tr>
<tr>
<td>Enable Row Filters</td>
<td></td>
<td>File &gt; Open Report</td>
<td>CTRL+O</td>
<td></td>
</tr>
<tr>
<td>Show 'Group By Column' Panel</td>
<td></td>
<td>File &gt; Open Report</td>
<td>CTRL+O</td>
<td></td>
</tr>
</tbody>
</table>

**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)**

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
<th>Menu access</th>
<th>Shortcut key access</th>
<th>Toolbar access</th>
</tr>
</thead>
<tbody>
<tr>
<td>Open report</td>
<td>Use to upload a report.</td>
<td>File &gt; Open Report</td>
<td>CTRL+O</td>
<td></td>
</tr>
<tr>
<td>Save report</td>
<td>Use to save a modified report.</td>
<td>File &gt; Save Report</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Close</td>
<td>Use to close a report without saving any modifications, such as applying filters. Proteome Discoverer remains open after the report is closed.</td>
<td>File &gt; Close</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Import Search Results from SRF</td>
<td>Import search results that are in SRF format.</td>
<td>File &gt; Import Search Results from SRF</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Connect to Server</td>
<td>(only if the Client_Server feature is licensed) Select a different server to use for your search process.</td>
<td>File &gt; Connect to Server</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 1. Description of File menu commands (Sheet 2 of 2)

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
<th>Menu access</th>
<th>Shortcut key access</th>
<th>Toolbar access</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disconnect from Server</td>
<td>(only if the Client_Sever feature is licensed) Select the server to disconnect from.</td>
<td>File &gt; Disconnect from Server</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exit</td>
<td>Close Proteome Discoverer and all opened views and reports without saving any unspecified modifications.</td>
<td>File &gt; Exit</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**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)

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
<th>Menu access</th>
<th>Shortcut key access</th>
<th>Toolbar access</th>
</tr>
</thead>
<tbody>
<tr>
<td>Show Details</td>
<td>When you select a protein row, the Protein Identification Details view appears. When you select a peptide row, the Peptide Identification Details view appears.</td>
<td>Search Report &gt; Show Details</td>
<td>CTRL+D</td>
<td></td>
</tr>
<tr>
<td>Show Distribution Chart</td>
<td>Assess different aspects of the search results including scores, delta masses, retention times, and such by plotting them in relationship to each other.</td>
<td>Search Report &gt; Show Distribution View</td>
<td>CTRL+SHIFT+D</td>
<td></td>
</tr>
<tr>
<td>Show proteins covered by this set of peptides</td>
<td>Explores what proteins are present and their associations through related peptides.</td>
<td>Search Report &gt; Show proteins covered by this set of peptides</td>
<td>CTRL+SHIFT+H</td>
<td></td>
</tr>
<tr>
<td>Show Chromatogram View</td>
<td>Shows the intensities of one or more masses as a function of time.</td>
<td>Search Report &gt; Show Chromatogram View</td>
<td>CTRL+SHIFT+C</td>
<td></td>
</tr>
<tr>
<td>Show Spectrum</td>
<td>Shows the MS/MS spectrum that was used for the peptide search.</td>
<td>Search Report &gt; Show Spectrum</td>
<td>CTRL+SHIFT+S</td>
<td></td>
</tr>
</tbody>
</table>
### Table 2. Description of the Search Report menu commands (Sheet 2 of 2)

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

Use the Quantitation menu to access the quantitation methods.

Table 3. Description of the Quantitation menu commands

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
<th>Menu access</th>
<th>Shortcut key access</th>
<th>Toolbar access</th>
</tr>
</thead>
<tbody>
<tr>
<td>Edit Quantitation Method</td>
<td>Edits the quantitation method of the current report.</td>
<td>Quantitation &gt; Edit Quantitation Method</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Show peptide ratios</td>
<td>Shows the peptide ratios per protein view.</td>
<td>Quantitation &gt; Show Peptide Ratios</td>
<td>CTRL+SHIFT+R</td>
<td></td>
</tr>
<tr>
<td>Show Reporter Intensities</td>
<td>Shows the reporter intensities view.</td>
<td>Quantitation &gt; Show Reporter Intensities</td>
<td>CTRL+SHIFT+N</td>
<td></td>
</tr>
<tr>
<td>Show Quantitation Spectrum</td>
<td>Shows the spectrum used for quantitation.</td>
<td>Quantitation &gt; Show Quantitation Spectrum</td>
<td>CTRL+SHIFT+Q</td>
<td></td>
</tr>
</tbody>
</table>

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)

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
<th>Menu access</th>
<th>Shortcut key access</th>
<th>Toolbar access</th>
</tr>
</thead>
<tbody>
<tr>
<td>New Workflow</td>
<td>Opens Workflow Editor page.</td>
<td>Workflow Editor &gt; New Workflow</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Open From Template</td>
<td>Opens existing saved data analysis workflow.</td>
<td>Workflow Editor &gt; Open From Template</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Save As Template</td>
<td>Saves a workflow.</td>
<td>Workflow Editor &gt; Save As Template</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 4. Description of the Workflow Editor menu commands (Sheet 2 of 2)

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
<th>Menu access</th>
<th>Shortcut key access</th>
<th>Toolbar access</th>
</tr>
</thead>
<tbody>
<tr>
<td>Auto Layout</td>
<td>Automatically adjusts and aligns the connecting arrows and nodes.</td>
<td>Workflow Editor &gt; Auto Layout</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Start Workflow</td>
<td>Begins the data analysis search using your selected workflow.</td>
<td>Workflow Editor &gt; Start Workflow</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Import Workflow from XML</td>
<td>Imports a data analysis workflow from data in XML format.</td>
<td>Workflow Editor &gt; Import Workflow from XML</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Export Workflow to XML</td>
<td>Exports your workflow into an XML formatted file.</td>
<td>Workflow Editor &gt; Export Workflow to XML</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Administration Menu

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

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

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
<th>Menu access</th>
<th>Shortcut key access</th>
<th>Toolbar access</th>
</tr>
</thead>
<tbody>
<tr>
<td>Show Job Queue</td>
<td>Displays the search queue and the status of current job searches.</td>
<td>Administration &gt; Show Job Queue</td>
<td>CTRL+J</td>
<td></td>
</tr>
<tr>
<td>Open .msf</td>
<td>Opens a recently completed, highlighted report in the job queue.</td>
<td>Administration &gt; Show Job Queue &gt; Open Report</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maintains FASTA Files</td>
<td>Adds, removes, and modifies FASTA files.</td>
<td>Administration &gt; Maintain FASTA Files</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maintain Chemical Modifications</td>
<td>Sets the chemical modifications to be used in the search process.</td>
<td>Administration &gt; Maintain Chemical Modifications</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maintain Cleavage Reagents</td>
<td>Adds, removes, and modifies the cleavage reagents.</td>
<td>Administration &gt; Maintain Cleavage Reagents</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maintain Quantitation Methods</td>
<td>Maintains and edits known quantitation methods.</td>
<td>Administration &gt; Maintain Quantitation Methods</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 5. Description of the Administration menu commands (Sheet 2 of 2)

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
<th>Menu access</th>
<th>Shortcut key access</th>
<th>Toolbar access</th>
</tr>
</thead>
<tbody>
<tr>
<td>Configuration</td>
<td>Configuration of parameters used in all searches.</td>
<td>Administration &gt; Configuration</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Manage Licenses</td>
<td>View license status and add new licenses.</td>
<td>Administration &gt; Manage License</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Tools Menu

Table 6. Description of the Tools menu commands

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
<th>Menu access</th>
<th>Shortcut key access</th>
<th>Toolbar access</th>
</tr>
</thead>
<tbody>
<tr>
<td>Open QualBrowser</td>
<td>Launches QualBrowser.</td>
<td>Tools &gt; Qual Browser</td>
<td>CTRL+SHIFT+B</td>
<td></td>
</tr>
<tr>
<td>Open InforSense</td>
<td>Launches InforSense.</td>
<td>Tools &gt; InforSense</td>
<td>CTRL+SHIFT+I</td>
<td></td>
</tr>
<tr>
<td>FASTA Database</td>
<td>Appends and adds to an existing FASTA file.</td>
<td>Tools &gt; FASTA Database Utilities</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Window Menu

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

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
<th>Menu access</th>
</tr>
</thead>
<tbody>
<tr>
<td>Close All Windows</td>
<td>Closes all the open windows of your Proteome Discoverer session.</td>
<td>Windows &gt; Close All Windows</td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
<th>Menu access</th>
<th>Short key access</th>
<th>Toolbar access</th>
</tr>
</thead>
<tbody>
<tr>
<td>About Proteome Discoverer</td>
<td>Displays the install version of Proteome Discoverer and its components. Shows the Thermo Fisher Scientific copyright notice.</td>
<td>Help &gt; About Proteome Discoverer</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proteome Discoverer Help</td>
<td>Opens Proteome Discoverer Help.</td>
<td>Help &gt; Proteome Discoverer Help</td>
<td>F1</td>
<td></td>
</tr>
</tbody>
</table>

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.
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 additional 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 initiates 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

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

<table>
<thead>
<tr>
<th>Workflow category</th>
<th>Workflow node</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Input:</td>
<td>Rawfile Selector</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Spectrum Selector</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SRF File Selector</td>
<td>.srf files are for input only.</td>
</tr>
<tr>
<td>Spectrum Processing:</td>
<td>ETD Spectrum Charger</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Noise Peak Filter</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Non-Fragment Filter</td>
<td>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.</td>
</tr>
<tr>
<td></td>
<td>Spectrum Grouper</td>
<td>Use group spectra from MS2 and MS3 scans together. Set <strong>Allow MS Order Mismatch</strong> to true; the default is false.</td>
</tr>
<tr>
<td></td>
<td>Spectrum Normalizer</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Xtract</td>
<td>Xtract only works with high-resolution data.</td>
</tr>
<tr>
<td>Spectrum Filters:</td>
<td>Scan Event Filter</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Spectrum Properties Filter</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Spectrum Score Filter</td>
<td></td>
</tr>
<tr>
<td>Peptide/Protein ID:</td>
<td>Mascot</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SEQUEST</td>
<td></td>
</tr>
</tbody>
</table>
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 Administration and Workflow” on page 67.

About iTRAQ Quantitation

iTRAQ™ is a protein quantitation technique that uses four or eight isobaric amine-specific 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, (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, (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, (see Figure 7).
With Qual Browser you can view the entire ion chromatogram and browse individual precursor and MSn 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 retrieved 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 identification 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 MS² 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
- Overview of Menu Options
- Working with Search Wizards
- Search Results, Reports, and Analysis
- Using the Proteome Discoverer Workflow Editor
- Working with Reports
- Working with the Search Job Queue
- Using FASTA Database Utilities
- Working with Chemical Modifications
- How to Quantify Your Data Using Administration and Workflow
- Using Qual Browser
- Job Queue
- InforSense
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


  2. Select a single file of interest or multiple files, using the CTRL key.
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.
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.
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**.
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 (see Table 10).
Table 10. Four parameters that define the grouping criteria

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

Table 10. Four parameters that define the grouping criteria

<table>
<thead>
<tr>
<th>Search summary parameter</th>
<th>Corresponding Search Wizard parameter</th>
<th>Description</th>
</tr>
</thead>
</table>
| 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. |
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 neccessary 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.
**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 17.** Search Report menu options

<table>
<thead>
<tr>
<th>Thermo Proteome Discoverer 1.0.40</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>File</strong></td>
<td><strong>Search Report</strong></td>
</tr>
<tr>
<td><img src="image" alt="Show Peptide ID Details..." /></td>
<td>Display pop-up of fragment ion detail.</td>
</tr>
<tr>
<td><img src="image" alt="Show Distribution Chart..." /></td>
<td>Distribution plot of peptide data</td>
</tr>
<tr>
<td><img src="image" alt="Show Peptide Consensus View" /></td>
<td>Gives all proteins identified that contain the selected peptide.</td>
</tr>
<tr>
<td><img src="image" alt="Show proteins covered by this set of peptides" /></td>
<td>Gives pop-up of indicated content.</td>
</tr>
<tr>
<td><strong>Quantitation</strong></td>
<td><strong>Processing</strong></td>
</tr>
<tr>
<td><img src="image" alt="Show Chromatogram View" /></td>
<td>Export user-defined mass exclusion list.</td>
</tr>
<tr>
<td><img src="image" alt="Show Spectrum" /></td>
<td>Export selected MSn spectral.</td>
</tr>
<tr>
<td><img src="image" alt="Show Fragment Match Spectrum" /></td>
<td>Export selected ProtXML data.</td>
</tr>
<tr>
<td><img src="image" alt="Show Isotope Pattern" /></td>
<td>Gives pop-up for peptide data comparison.</td>
</tr>
<tr>
<td><img src="image" alt="Show Extracted Ion Chromatogram" /></td>
<td></td>
</tr>
<tr>
<td><img src="image" alt="Show Sequence Comparison" /></td>
<td></td>
</tr>
</tbody>
</table>
**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 60.

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

- The blue-colored background 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.
2 Using Proteome Discoverer
Search Results, Reports, and Analysis

Figure 19. Protein view

Double-click any cell to see coverage diagram.
Click + to open the protein row.

Figure 20. Protein coverage diagram

Double-click a protein cell. Resulting coverage diagram
Right-click for options.
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.

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.

### Table 11. Protein table

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

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.
Table 12 describes the features of a peptide row.

**Table 12. Peptide row (Sheet 1 of 2)**

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sequence</td>
<td>Displays the peptide sequence.</td>
</tr>
<tr>
<td># Proteins</td>
<td>Number of proteins.</td>
</tr>
<tr>
<td># Protein Groups</td>
<td>How many protein groups.</td>
</tr>
<tr>
<td>Activation Type</td>
<td>Displays the activation type of the spectrum where the peptide was identified.</td>
</tr>
<tr>
<td>Modifications</td>
<td>Displays the static and dynamic modifications identified in the peptide.</td>
</tr>
<tr>
<td>Probability</td>
<td>Displays the probability score for the peptide.</td>
</tr>
<tr>
<td>Score</td>
<td>Displays the score that the search algorithm calculated for the match.</td>
</tr>
<tr>
<td>XCorr</td>
<td>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.</td>
</tr>
<tr>
<td>Quan Info</td>
<td>(Optional) Marking if quantitative analysis is available</td>
</tr>
<tr>
<td>115/114</td>
<td>(Optional) Related to Quan Info</td>
</tr>
<tr>
<td>116/114</td>
<td>(Optional) Related to Quan Info</td>
</tr>
<tr>
<td>117/114</td>
<td>(Optional) Related to Quan Info</td>
</tr>
<tr>
<td>Rank</td>
<td>Displays the rank ordering of the peptide.</td>
</tr>
<tr>
<td>Charge</td>
<td>Displays the charge state of the peptide.</td>
</tr>
<tr>
<td>MH+(Da)</td>
<td>Displays the weight differential of the peptides in Daltons.</td>
</tr>
<tr>
<td>ΔM (ppm)</td>
<td>Displays the difference between the theoretical mass of the peptide and the experimental mass of the precursor ion.</td>
</tr>
<tr>
<td>First Scan</td>
<td>Displays the first scan number where the peptide was initially identified.</td>
</tr>
<tr>
<td>Last Scan</td>
<td>Displays the last scan number in which this peptide match was identified.</td>
</tr>
<tr>
<td>RT (minutes)</td>
<td>Displays the retention time where the peptide was observed.</td>
</tr>
</tbody>
</table>
Table 12. Peptide row (Sheet 2 of 2)

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSOrder</td>
<td>Indicates the order of the MS spectrum:</td>
</tr>
<tr>
<td></td>
<td>• 1 is a one MS scan</td>
</tr>
<tr>
<td></td>
<td>• 2 is an MS/MS scan (MS2)</td>
</tr>
<tr>
<td></td>
<td>• 3 is MS3</td>
</tr>
<tr>
<td></td>
<td>The number of stages minus one of precursor ion m/z selection, followed by product ion detection such that MS2 means one stage of precursor ion m/z selection is then followed by product ion detection</td>
</tr>
<tr>
<td>m/z</td>
<td>Mass-to-charge ratio</td>
</tr>
<tr>
<td>Intensity</td>
<td>Relative intensity</td>
</tr>
<tr>
<td>Ions Matched</td>
<td>Displays the number of ions found compared to the theoretical number of ions.</td>
</tr>
<tr>
<td>Annotation</td>
<td>Provides a location for you to insert and store notes about the search results, your analysis, and quality results.</td>
</tr>
<tr>
<td>Shortcut Menu</td>
<td>In the header, right-click to access the shortcut menu:</td>
</tr>
<tr>
<td></td>
<td>• Show Top Match Peptides Only</td>
</tr>
<tr>
<td></td>
<td>• Show Peptide Groups</td>
</tr>
<tr>
<td></td>
<td>• Show Filtered Out Rows</td>
</tr>
<tr>
<td></td>
<td>• Enable Protein Grouping</td>
</tr>
<tr>
<td></td>
<td>• Mass Tolerance Unit</td>
</tr>
<tr>
<td></td>
<td>• Row Numbers Style</td>
</tr>
<tr>
<td></td>
<td>• Copy, Ctrl+C</td>
</tr>
<tr>
<td></td>
<td>Copies the selected rows to the Clipboard. You can then paste the rows in a spreadsheet application such as Microsoft Excel.</td>
</tr>
<tr>
<td></td>
<td>• Copy With Column Headers</td>
</tr>
<tr>
<td></td>
<td>• Export to Excel Workbook.</td>
</tr>
<tr>
<td></td>
<td>• Enable Row Filters</td>
</tr>
<tr>
<td></td>
<td>• Show ‘Group By Column’ Panel</td>
</tr>
</tbody>
</table>

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)

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

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

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scan(s)</td>
<td>Displays all scan numbers where this peptide match was identified.</td>
</tr>
<tr>
<td>RT (minutes)</td>
<td>Displays the retention time where the peptide was observed.</td>
</tr>
<tr>
<td>Activation Type</td>
<td>Displays the activation type of the spectrum where the peptide was identified.</td>
</tr>
<tr>
<td>Mass Analyzer</td>
<td>Displays the mass analyzer used to create the .raw file data. Proteome Discoverer recognizes the following mass analyzers:</td>
</tr>
<tr>
<td></td>
<td>• ITMS (Ion Trap)</td>
</tr>
<tr>
<td></td>
<td>• FTMS (Fourier Transform)</td>
</tr>
<tr>
<td></td>
<td>• TOFMS (Time of Flight)</td>
</tr>
<tr>
<td></td>
<td>• SQMS (Single Quad)</td>
</tr>
<tr>
<td></td>
<td>• TQMS (Triple Quad)</td>
</tr>
<tr>
<td></td>
<td>• SectorMS (Sector Field)</td>
</tr>
</tbody>
</table>
Figure 25. Filters page

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)

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active</td>
<td>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.</td>
</tr>
</tbody>
</table>
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

**Table 14.** Filter options (Sheet 2 of 2)

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inverted</td>
<td>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.</td>
</tr>
<tr>
<td>Is Applied</td>
<td>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.</td>
</tr>
<tr>
<td>Filter</td>
<td>Displays the name of the filter.</td>
</tr>
<tr>
<td>Filter Setting(s)</td>
<td>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.</td>
</tr>
</tbody>
</table>

Table 14. Filter options (Sheet 2 of 2)
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

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

   ![Activated node example](image)

   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

   ![Joining two nodes](image)

   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.

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 node 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 MSn scan, such as choosing to pick the precursor from the parent scan.
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
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

![Figure 33](image)

**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

![Figure 34](image)
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.
   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.
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.
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

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.
2 Using Proteome Discoverer
Using FASTA Database Utilities

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

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
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
How to Quantify Your Data Using Administration and Workflow

- Reporter Ion Based Quantitation in Proteome Discoverer
- Assess the Abundance of the Precursor

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.
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
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
   - Ratio Calculation
   - Protein Quantitation
   - Experimental Bias
   - Ratio Reporting

Figure 42. Quantitation Processing Method Editor tab options

To apply a quantitation 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 Quantitzer 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.

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 to 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.
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](image). 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

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

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

**IMPORTANT** InforSense Protein Annotation comes with InforSense Workflow documentation available through the Help menu.
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
- Working with Filters
- False Discovery Rates
- Working with the Proteins Grid
- Working with the Peptides Grid
- Working with the Search Input
- Interpreting the Isotope Pattern View
- Interpreting the Spectrum View
- Interpreting the Extracted Ion Chromatogram
- Interpreting the Fragment Match View
- Exporting Data to Other Programs
- Working with InforSense Discussion
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

<table>
<thead>
<tr>
<th>Show Top Match Peptides Only</th>
</tr>
</thead>
<tbody>
<tr>
<td>Show Peptide Groups</td>
</tr>
<tr>
<td>Show Filtered Out Rows</td>
</tr>
<tr>
<td>Enable Protein Grouping</td>
</tr>
<tr>
<td>Mass Tolerance Unit</td>
</tr>
<tr>
<td>Row Numbers</td>
</tr>
<tr>
<td>Copy</td>
</tr>
<tr>
<td>Copy With Column Headers</td>
</tr>
<tr>
<td>Export to Excel Workbook</td>
</tr>
<tr>
<td>Enable Row Filters</td>
</tr>
<tr>
<td>Show 'Group By Column' Panel</td>
</tr>
</tbody>
</table>

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

![Figure 49. Column header sorting example](image-url)

**Thermo Proteome Discoverer 1.0**

<table>
<thead>
<tr>
<th>Accession #</th>
<th>Coverage</th>
<th># Peptides</th>
<th>Score</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>gi2505462</td>
<td>48.37 %</td>
<td>9</td>
<td>153</td>
<td>MYOGLOBIN</td>
</tr>
<tr>
<td>gi115860</td>
<td>16.96 %</td>
<td>11</td>
<td>224</td>
<td>BETA CASEIN PRECURSOR</td>
</tr>
<tr>
<td>gi410954</td>
<td>24.88 %</td>
<td>42</td>
<td>507</td>
<td>serum albumin precursor - bovine [MASS=69270]</td>
</tr>
<tr>
<td>gi1942750</td>
<td>38.68 %</td>
<td>5</td>
<td>153</td>
<td>Myoglobin (horse heart) Mutant with Ser 92 Replaced by Asp (92)</td>
</tr>
<tr>
<td>gi129253</td>
<td>25.65 %</td>
<td>20</td>
<td>306</td>
<td>OVALBUMIN (PLASALBUMIN) (ALLERGEN GAL D 2) (GAL D II)</td>
</tr>
<tr>
<td>gi2554649</td>
<td>52.06 %</td>
<td>14</td>
<td>153</td>
<td>Myoglobin (horse heart) Mutant with Leu 104 Replaced by Asn (L)</td>
</tr>
<tr>
<td>gi643704</td>
<td>12.69 %</td>
<td>17</td>
<td>807</td>
<td>SERUM ALBUMIN PRECURSOR</td>
</tr>
<tr>
<td>gi999001</td>
<td>40.41 %</td>
<td>12</td>
<td>153</td>
<td>Myoglobin Mutant with His 64 Replaced by Thr (H64T)</td>
</tr>
<tr>
<td>gi246868</td>
<td>43.18 %</td>
<td>19</td>
<td>197</td>
<td>LYSOZYME C PRECURSOR [1,β-BETA-N-ACETYLMURAMIDASE C C]</td>
</tr>
<tr>
<td>gi2414903</td>
<td>40.10 %</td>
<td>12</td>
<td>39</td>
<td>cytochrome c - horse (fragments) [MASS=10934]</td>
</tr>
<tr>
<td>gi2190337</td>
<td>35.58 %</td>
<td>79</td>
<td>507</td>
<td>Serum albumin [Bovine]</td>
</tr>
<tr>
<td>gi161457</td>
<td>17.24 %</td>
<td>5</td>
<td>252</td>
<td>lysozyme (horse, heart, Peptide, 104 Asx) [MASS=11826]</td>
</tr>
<tr>
<td>gi809416</td>
<td>60.99 %</td>
<td>105</td>
<td>105</td>
<td>Lysozyme C</td>
</tr>
<tr>
<td>gi314117</td>
<td>32.02 %</td>
<td>11</td>
<td>104</td>
<td>apocytochrome c (horses, heart, Peptide, 104 Asx) [MASS=11826]</td>
</tr>
<tr>
<td>gi1227120</td>
<td>41.44 %</td>
<td>9</td>
<td>144</td>
<td>Lysozyme, Fibrinogen Mol: 1; Molecule: Lysozyme Modified Wt</td>
</tr>
<tr>
<td>gi17705</td>
<td>54.81 %</td>
<td>17</td>
<td>104</td>
<td>CYTOCHROME C</td>
</tr>
</tbody>
</table>

**3 Interpreting Search Results**

**Working with the Initial Results Report**
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.

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
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.
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 unnecessary 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 \( \text{Filters Changed!} \). 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
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

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
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

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

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
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

![SEQUEST Search Wizard](Image)

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.
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
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
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.

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?
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

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.
3 Interpreting Search Results

Working with the Proteins Grid

Figure 67. Example of Protein Identification Details view

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).
3 Interpreting Search Results

Working with the Proteins Grid

Figure 68. Report Item Distribution chart with hover text
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.
3 Interpreting Search Results

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.

- **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.

- **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 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.

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.
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

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°, B°, and C° are ions that have lost water (-18 Da). The subscript next to the letter indicates the number of residues in the fragment ion.1

Table 15 summarizes the fragment ions used in Proteome Discoverer:

<table>
<thead>
<tr>
<th>Ions</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Charge on N-terminal side</td>
</tr>
<tr>
<td>B</td>
<td>Charge on N-terminal side</td>
</tr>
<tr>
<td>C</td>
<td>Charge on N-terminal side</td>
</tr>
<tr>
<td>Y</td>
<td>Charge on C-terminal side</td>
</tr>
<tr>
<td>Z</td>
<td>Charge on C-terminal side</td>
</tr>
<tr>
<td>B*</td>
<td>B ion that has lost ammonia (-17 Da)</td>
</tr>
<tr>
<td>Y*</td>
<td>Y ion that has lost ammonia (-17 Da)</td>
</tr>
<tr>
<td>B°</td>
<td>B ion that has lost water (-18 Da)</td>
</tr>
<tr>
<td>Y°</td>
<td>Y ion that has lost water (-18 Da)</td>
</tr>
</tbody>
</table>

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
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.
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
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.

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 chromatograph

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
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
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

   ![Export successful message](image)

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

   ![Exported exclusion mass list example](image)

**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

- [x] Layer 1: Proteins
- [x] Layer 2: Peptides
- [ ] Layer 3: Proteins

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

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.
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

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).

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 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 databases against which the sequence search was performed in Discoverer. This is defined by the user, and may be any of: NCBI, UniProtKB, 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 PR repository of cross-references (URL) and retrieving the corresponding GenBank primary accession number. The system determines automatically if this lookup is required. The original identifier 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-characterized 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 UniprotB 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 if 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 plugin. 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 present in GenBank, a specialized 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, IP 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 PR repository of cross-references and retrieving the corresponding GenBank primary accession number. The system determines automatically if 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 hierarchy 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.
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, text present in UniProtKB, a specialized 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 proXML format (ref. URL). A list of protein identifiers is extracted from the proXML 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, IP, or others. The first stage of the annotation process is to find the equivalent UniProt primary 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 if 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 plugin. 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 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 (DDB) 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 Excel report as above. If any of the terms or keywords is found in the sequence entries retrieved from NCBI or UniProt, the corresponding table column is filled with the sequence locations of each occurrence in each protein.
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.

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 \rightarrow 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.


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.

   ![Customize dialog box]

   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.

   ![Toolbar menu]

   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
To resize icons on toolbars and menus

1. On the toolbar, click \( \text{\textit{}} \) 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 \( \text{\textit{}} \) 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 \( \text{\textit{}} \) 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

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

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

### 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.

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.

   ![Table showing chemical modifications](image)

   **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.

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.

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.
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.

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

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modification</td>
<td>Displays the type of modification.</td>
</tr>
<tr>
<td>Delta Mass</td>
<td>Displays the mass difference associated with the modification.</td>
</tr>
<tr>
<td>Delta Average Mass</td>
<td>Displays the difference between two adjacent average mass measurements.</td>
</tr>
<tr>
<td>Substitution</td>
<td>Displays the chemical group substitution that occurs with the modification.</td>
</tr>
<tr>
<td>Leaving group</td>
<td>Displays the chemical group that is missing.</td>
</tr>
<tr>
<td>Position</td>
<td>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.</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Displays the abbreviation of the modification displayed in reports.</td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th>Name</th>
<th>Size</th>
<th>#Sequences</th>
<th>#Residues</th>
<th>Status</th>
<th>Last Modified</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mascot_IDE</td>
<td>0</td>
<td>70</td>
<td>63387</td>
<td>Imported</td>
<td>05/22/2008</td>
</tr>
<tr>
<td>equine.fasta</td>
<td>152951</td>
<td>246</td>
<td>75453</td>
<td>Available</td>
<td>06/02/2004</td>
</tr>
<tr>
<td>bovine2.fasta</td>
<td>667086</td>
<td>1057</td>
<td>242927</td>
<td>Available</td>
<td>06/23/2006</td>
</tr>
</tbody>
</table>

- **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](add fasta file.png).
  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 ![Select FASTA File](select fasta file.png) at the beginning of a row to activate the row.
  3. Click ![Remove FASTA File](remove fasta file.png).
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)

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>One-letter code</th>
<th>Three-letter code</th>
<th>Monoisotopic mass</th>
<th>Average mass</th>
<th>Sum formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glycine</td>
<td>G</td>
<td>Gly</td>
<td>57.02147</td>
<td>57.0517</td>
<td>C2H3NO</td>
</tr>
<tr>
<td>Alanine</td>
<td>A</td>
<td>Ala</td>
<td>71.03712</td>
<td>71.0787</td>
<td>C3H5NO</td>
</tr>
<tr>
<td>Serine</td>
<td>S</td>
<td>Ser</td>
<td>87.03203</td>
<td>87.078</td>
<td>C5H5NO2</td>
</tr>
<tr>
<td>Proline</td>
<td>P</td>
<td>Pro</td>
<td>97.05277</td>
<td>97.1168</td>
<td>C5H7NO</td>
</tr>
<tr>
<td>Valine</td>
<td>V</td>
<td>Val</td>
<td>99.06842</td>
<td>99.1328</td>
<td>C5H9NO</td>
</tr>
<tr>
<td>Threonine</td>
<td>T</td>
<td>Thr</td>
<td>101.04768</td>
<td>101.1051</td>
<td>C4H7NO2</td>
</tr>
<tr>
<td>Cysteine</td>
<td>C</td>
<td>Cys</td>
<td>103.00919</td>
<td>103.145</td>
<td>C3H5NOS</td>
</tr>
<tr>
<td>Isoleucine</td>
<td>I</td>
<td>Ile</td>
<td>113.08407</td>
<td>113.1598</td>
<td>C6H11NO</td>
</tr>
<tr>
<td>Leucine</td>
<td>L</td>
<td>Leu</td>
<td>113.08407</td>
<td>113.1598</td>
<td>C6H11NO</td>
</tr>
<tr>
<td>Asparagine</td>
<td>N</td>
<td>Asn</td>
<td>114.04293</td>
<td>114.1039</td>
<td>C4H6N2O2</td>
</tr>
<tr>
<td>Aspartic Acid</td>
<td>D</td>
<td>Asp</td>
<td>115.02695</td>
<td>115.0885</td>
<td>C4H5NO3</td>
</tr>
<tr>
<td>Glutamine</td>
<td>Q</td>
<td>Gln</td>
<td>128.05858</td>
<td>128.13091</td>
<td>C5H8N2O2</td>
</tr>
<tr>
<td>Lysine</td>
<td>K</td>
<td>Lys</td>
<td>128.09497</td>
<td>128.1745</td>
<td>C6H12N2O</td>
</tr>
</tbody>
</table>
### Table 1. Amino acid mass values (Sheet 2 of 2)

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>One-letter code</th>
<th>Three-letter code</th>
<th>Monoisotopic mass</th>
<th>Average mass</th>
<th>Sum formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glutamic Acid</td>
<td>E</td>
<td>Glu</td>
<td>129.0426</td>
<td>129.1156</td>
<td>C5H7NO3</td>
</tr>
<tr>
<td>Methionine</td>
<td>M</td>
<td>Met</td>
<td>131.0405</td>
<td>131.1994</td>
<td>C5H9NOS</td>
</tr>
<tr>
<td>Histidine</td>
<td>H</td>
<td>His</td>
<td>137.05891</td>
<td>137.1414</td>
<td>C6H7N3O</td>
</tr>
<tr>
<td>Phenylalanine</td>
<td>F</td>
<td>Phe</td>
<td>147.06842</td>
<td>147.1772</td>
<td>C9H9NO</td>
</tr>
<tr>
<td>Arginine</td>
<td>R</td>
<td>Arg</td>
<td>156.10112</td>
<td>156.188</td>
<td>C6H12N4O</td>
</tr>
<tr>
<td>Tyrosine</td>
<td>Y</td>
<td>Tyr</td>
<td>163.0632</td>
<td>163.1766</td>
<td>C9H9NO2</td>
</tr>
<tr>
<td>Tryptophan</td>
<td>W</td>
<td>Trp</td>
<td>186.07932</td>
<td>186.2141</td>
<td>C11H10N2O</td>
</tr>
</tbody>
</table>

### Table 2. Special amino acids

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

### Modification Values

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

### Table 3. Modification values (Sheet 1 of 2)

<table>
<thead>
<tr>
<th>Modification</th>
<th>Monoisotopic mass</th>
<th>Average mass</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deamidation</td>
<td>0.98402</td>
<td>0.98480</td>
</tr>
<tr>
<td>Methylation</td>
<td>14.01565</td>
<td>14.02660</td>
</tr>
<tr>
<td>Hydroxylation</td>
<td>15.99492</td>
<td>15.99940</td>
</tr>
<tr>
<td>Oxidation</td>
<td>15.99492</td>
<td>15.99940</td>
</tr>
<tr>
<td>Formylation</td>
<td>27.99491</td>
<td>28.01010</td>
</tr>
<tr>
<td>Acetylation</td>
<td>42.01057</td>
<td>42.03670</td>
</tr>
<tr>
<td>Carboxyamidomethylation</td>
<td>57.02146</td>
<td>57.05130</td>
</tr>
<tr>
<td>Carboxymethylation</td>
<td>58.00548</td>
<td>58.03610</td>
</tr>
<tr>
<td>Propionamide</td>
<td>71.03711</td>
<td>71.07790</td>
</tr>
<tr>
<td>Sulfation</td>
<td>79.95682</td>
<td>80.06320</td>
</tr>
</tbody>
</table>
Enzyme Cleavage Properties

Table 4 lists the enzymes and reagents with cleavage properties.

<table>
<thead>
<tr>
<th>Enzymes/Reagents</th>
<th>Cleaves after</th>
<th>Cleaves before</th>
<th>Except when</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Enzymes for digestion</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trypsin K</td>
<td>K</td>
<td>P is after K</td>
<td></td>
</tr>
<tr>
<td>Trypsin R</td>
<td>R</td>
<td>P is after R</td>
<td></td>
</tr>
<tr>
<td>Trypsin (KR)</td>
<td>K or R</td>
<td>P is after K or R</td>
<td></td>
</tr>
<tr>
<td>Trypsin (KRLNH)</td>
<td>K, R, L, N, or H</td>
<td>P is after K, R, L, N, or H</td>
<td></td>
</tr>
<tr>
<td>Trypsin (KR/P)</td>
<td>K or R</td>
<td>P is after K or R</td>
<td></td>
</tr>
<tr>
<td>Trypsin (KRLNH/P)</td>
<td>K, R, L, N, or H</td>
<td>P is after F, W, or Y</td>
<td></td>
</tr>
<tr>
<td>Chymotrypsin</td>
<td>F, W, Y, or L</td>
<td>P is after F, W, or Y</td>
<td></td>
</tr>
<tr>
<td>Chymotrypsin (FWY)</td>
<td>F, W, or Y</td>
<td>P is after F, W, or Y</td>
<td></td>
</tr>
<tr>
<td>Clostripain</td>
<td>R</td>
<td>P is after F, W, or Y</td>
<td></td>
</tr>
<tr>
<td>Proline_Endopept</td>
<td>P</td>
<td>P is after F, W, or Y</td>
<td></td>
</tr>
<tr>
<td>Staph_protease</td>
<td>E</td>
<td>P is after F, W, or Y</td>
<td></td>
</tr>
<tr>
<td>GlucC</td>
<td>E or D</td>
<td>P is after F, W, or Y</td>
<td></td>
</tr>
<tr>
<td>LysC</td>
<td>K</td>
<td>P is after F, W, or Y</td>
<td></td>
</tr>
<tr>
<td>AspN</td>
<td>D</td>
<td>P is after F, W, or Y</td>
<td></td>
</tr>
<tr>
<td>Elastase</td>
<td>A, L, I, or V</td>
<td>P is after A, L, I, or V</td>
<td></td>
</tr>
<tr>
<td><strong>Chemicals for degradation</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cyanogen Bromide</td>
<td>M</td>
<td>P is after F, W, or Y</td>
<td></td>
</tr>
<tr>
<td>Iodobenzoate</td>
<td>W</td>
<td>P is after F, W, or Y</td>
<td></td>
</tr>
</tbody>
</table>
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°, B°, and C° are ions that have lost water (-18 Da). The subscript next to the letter indicates the number of residues in the fragment ion.1

Table 5 summarizes the fragment ions used in Proteome Discoverer.

Table 5. Fragment ions

<table>
<thead>
<tr>
<th>Ions</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Charge on N-terminal side</td>
</tr>
<tr>
<td>B</td>
<td>Charge on N-terminal side</td>
</tr>
<tr>
<td>C</td>
<td>Charge on N-terminal side</td>
</tr>
<tr>
<td>Y</td>
<td>Charge on C-terminal side</td>
</tr>
<tr>
<td>Z</td>
<td>Charge on C-terminal side</td>
</tr>
<tr>
<td>B°</td>
<td>B ion that has lost water (-18 Da)</td>
</tr>
<tr>
<td>Y°</td>
<td>Y ion that has lost water (-18 Da)</td>
</tr>
</tbody>
</table>

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.


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:

>IP81:100685094.1|SWISS-PROT:Q2KIJI2|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, whereas 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.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


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
Symbols
.dta 18
.msfsf  introduction 18, 26
   load 59
   naming 56
.raw 18
.srf  introduction 18
   open 59
.xml 117

A
add a FASTA 144
add to a FASTA database file 62, 144
adding and removing amino acids 141
administration
   cleavage reagents, open 137
   maintain chemical modifications, add 141
   maintain chemical modifications, add new 139
   maintain chemical modifications, delete 142
   maintain chemical modifications, import 142
   maintain chemical modifications, import from UNIMOD 142
   maintain chemical modifications, update 140
   maintain chemical modifications, delete 140
   maintain FASTA file, add 144
   maintain FASTA file, delete 144
   quantitation methods 67, 68
amino acid
   corresponding 66
   mass values 147
   modification 141, 142
   modifications table 66
amino acid modifications
   add 141
   delete 142
   import from a local file 142
   import from local file 142
   import from UNIMOD 142
amino acid symbols 147
analysis, detailed 117
analyze raw MS data overview 14
animation option 29
assess the abundance and intensity of the precursor 113

C
chart
   chromatogram 100
   extracted ion chromatogram 101
   fragment match 115
   isotope 113
   peptide distribution 103
chemical modification, customizing 138
chemical modifications
   delete 140
   importing 142
   list 148
   new 139
   open 66
   quick access 9
   select 29
   update 140
Chemical Modifications view 143
chromatogram
   extracted ion chart 101
   yellow range, selected retention time range 100
chromatogram view
   red bars, elution time of all identified peptides 100
CID and ETD
   search algorithm discussion 14
CID, collision-induced dissociation 14
cleavage
   after 149
   before 149
   enzymes and reagent properties 149
   N-CR bond 109, 150
   reagents, customizing 137
   cleavage inhibitors, definition 138
   cleavage properties list 149
cleavage reagents
  quick access 9

cleavage sites
  cleavage reagents 137
  definition 138
  offset 138

close report 28
Column Chooser 82
complex mixtures 17
confidence levels, database searches 46
confidence settings, peptide 46, 96
configuration
  quick access 10
copying a view 117
create a search workflow 53
current job status, quick access 9
customize chemical modifications 143
customizing
  toolbar 131
  toolbar display 135
  toolbar layout 132
  tooltips 136
customizing chemical modification settings 138
customizing cleavage reagents 137

data files 18
database
  deleting a file 144
  renaming a file 144
database selection 29
deactivate a filter 87
decky database option
  peptide confidence 35
decky database search
  filter settings 46
  results 65
  using 65
  wizard setting 95
decky search feature 93
deleting
  modification 140
  toolbar button in Qual Browser 135
detailed analysis 117
display filtered rows 92

d E
ECD, electron-capture dissociation 14
edit quantitation
  overview 20
  peptide ratios 20
  quantitation spectrum 21
  show report intensities 21
enzyme manager, cleavage reagents 137
enzyme type, defining in SEQUEST Search Wizard 35
enzymes
  reagents, cleavage properties of 149
enzymes and reagents, list 138
ETD
electron-transfer dissociation 14
  with ECD, and CID data, SEQUEST 17
ETD and CID
  search algorithm discussion 14
ETD and ECD
  ZCore 16
ETD, electron-transfer dissociation 14
example
  column header sorting 79
  exported exclusion mass list 119
  filter settings 89
  filter your search results 85
  gene ontology (GO) three components 24
  inverted filter 48, 85, 86
  multiconsensus report 42
  proteins and peptides export 119
  retention time difference 33
  SEQUEST Search Wizard 30
  Workflow Editor 52
  Workflow Editor palette 51
  Workflow Report 38
  workflow workspace graphic 51
Excel, creating reports 18
experimental bias 69
export
  grids such as fragment match matrix 119
  protein and peptide identification results 119
  search input 119
exporting and reporting data, introduction 18
exporting data, how to 117
extracted example
  ion chromograph graph 72, 115
extracted ion chromatogram
display the results 101
view 107
XIC 98

F
false discovery rate
false positive rate 93
FDR 64
new target 96
SEQUEST 17
false discovery rates
determining 95
false positive identifications 64
false positive rate
false discovery rate 93
FDR 64
ZCore 16
FASTA Database Utilities 62, 144
FASTA file
add 144
deleting 144
renaming 144
FASTA files
quick access 9
FDR, false discovery rate 64
FDRs
not effective on small data sets 65
validate large data sets 65
file types 18
files, data 18
filter
Active 85
change settings 96
deactivate 87
hide rows 92
inverted 85
new settings 96
quick 92
remove 87
results 85
settings, distribute confidence indicators 46
filters
understanding descriptive information 37
filters tab 85
fragment ions 150
A, B, and C 109, 150
distribution 17
fragment match 115

G
getting a license viii
GO Annotation NCBI, using 126
GO Annotation SwissProt, using 127
graphical views, isotope and chromatogram 77
grouping
quantitation report, protein grouping is turned off 98
search results report, protein grouping 98

H
height value 72, 115
how to use
GO Annotation NCBI 126
GO Annotation SwissProt 127
Metadata Table NCBI 128
Metadata Table SwissProt 129

I
icon sizes 28
ImageQuest, getting a license viii
importing chemical modifications 142
InforSense
Protein Annotation 124
Virtual Machine (VM) 23, 75
workflow discussion 124
workflow documentation, complete set 76
workflows, compatibility caveat 124
InforSense Protein Annotation
compatible databases 75, 124
integrate data 23
join databases 23
license x
using 125
input files 18
instructions to create a quantitation method 69
integrate
automatically 115
data, InforSense Protein Annotation tasks 23
inverted filter example 48, 85, 86
invoke InforSense 75
ion trap mass spectrometers, SEQUEST 17
isotope pattern chart view 113
iTRAQ 4plex 67
iTRAQ 4plex method 68
iTRAQ 8plex 67
iTRAQ 8plex method 68
iTRAQ quantitation 20
J
job queue
controls and status 74
description 73
introduction 26
start workflow 56
join databases, InforSense Protein 23

L
layout menu 82
license
getting viii
getting, quick access 10
InforSense x
other x
to enter x
list of cleavage properties 149
low-abundance proteins 17

M
Mascot
high level of confidence with multiple peptides 16
peptide mass fingerprint 16
Mascot search algorithm 16
mass tags 69
mass values
amino acid 147
list 147
match likely proteins 77
measure and report 67
Meta Table, using SwissProt 129
Metadata Table
NCBI, using 128
missed cleavages, defining in SEQUEST Search Wizard 35
modifications
adding 139
adding amino acid 141
deleting 140
deleting amino acid 142
importing UNIMOD information from a local file 142
importing UNIMOD information from the web 142
setting 139
updating 140
values 148
modify a FASTA database file 62, 144
mouse, range definition 31
multiconsensus report
creating and viewing 41
opening 41

N
node options
workflow 57
nodes
workflow 51

O
open
one or more reports 27
open .srf 59
open a report
file menu option 59
job queue page option 74
open option 26
open Proteome Discoverer workflow 56
open report
quick access 9
output files 18
overview
InforSense Protein Annotation 23, 75
process and analyze raw MS data 14
Qual Browser 22
quantitation 20
search algorithms 14

P
parameters 29
peak
list 11
magnify 72, 115
peak area
automatically calculated 72, 115
peptide
access to the protein page 39
candidates 17
consensus page 107
consensus view 108, 110, 111
graphical view of distribution 103
identification details 107
individual results 112
intensity display 72, 115
isotope pattern of the selected precursor 106
quantitation 70
ratios 8
shortcut menu 45
show consensus view 108
show top matches only 45
view 60
view the fragment match spectrum 106
view the spectrum 106
peptide confidence
  decoy database option 35
  understanding descriptive information 37
Peptide Confidence page 95
peptide confidence settings, change the default 46, 96
peptide ID search 93
peptide identification view 110
peptide sequence match grid 111
peptides
  understanding descriptive information 37
percent coverag combined 41
PMF, Mascot 16
precursor
  detailed perspective of the MS scan 7
  experimentally determined mass 113
precursor view 113
probability-based scoring
  SEQUEST 17
  ZCore 17
process raw MS data overview 14
processing menu 29
protein
  add reference 64
  databases 17
  grouping 98
  view 60
protein identification
  details 98
  how to quantify 67
  increasing confidence with reverse database search 17
  overview 11
  SEQUEST 17
protein quantitation 69
protein sequence
  coverage 41
  protein table 39
proteins
  understanding descriptive information 37
quantitation 20
  basic feature 67
  iTRAQ 20
  menu option choices 8
  overview 20
  peptide ratios 20
  quantitation spectrum 21
  show report intensities 21
  spectrum 8
  TMT tagging 18
quantitation methods
  edit 68
  quick access 9
  select 67
quantitation overview 58
quantization node, in the workflow 69
queue list
  checking 60
  deleting search items 61
quick access
  configuration search programs 10
  current job status 9
  FASTA 10
  InforSense 10
  license status 10
  QualBrowser 10
  quantitation methods 9
  quantitation, edit method, edit method 8
  quantitation, reporter intensities 8
  quantitation, show peptide ratios 8
  quantitation, show spectrum 8
  Workflow Editor 8
quick filters
  display 92
  shortcut menu 92
  show filtered and unfiltered rows 92
R
  ratio calculation 69
  ratio reporting 69
raw data 11
  results, reports, and analysis 18
  reagents, enzymes 149
  recalculate new false discovery rate 96
  recalculate new filter settings 96
  red bars, chromatogram 100
  report
    close 59
    save 59
  Report Item Distribution graph 103
  reporter intensities 8
  reporter ion based quantitation 67
reporter ions quantitizer node
  about iTRAQ 20
  show peptide ratios 20
  show quantitation spectrum 21
  show reporter intensities, 21
reporting data, introduction 18
reports
  adding columns 79
  create 77
  export 77
  fragment 115
  open 27
  peptides 117
  removing columns 79
  save and close 28
results
  interpret the search overview 14
  reports and analysis 18
  validated 67
retention time 100, 103
row
  hide with quick filters 92
S
save a search workflow 56
save and close reports 28
save Proteome Discoverer workflow 56
scoring 17
search
  .msf (mass spectrum search files) 18
  .srf (unified search files) 18
  checking status of 60
  discussion of analysis 18
  file menu to import results 5
search algorithms
  Mascot 16
  ZCore 16
search engine 29
Search Input page 112
search parameters 29
search progress in percent
  description 73
  how to 60
search queue
  description 73
  jobs 9
search report menu
  access report options 36
  table of features 6
search wizards
  discussion 29
  process diagram 30
  to use on a raw file 30
select
  chemical modifications 29
  database 29
  search engine 29
  search parameters 29
SEQUEST
  ETD, ECD, and CID data 17
  false discovery rates 17
  ion trap mass spectrometers 17
  probability-based scoring 17
set the database 29
set the search parameters 29
shortcut keys, creating 133
shortcut menu
  peptide 45
  quick filters 92
  shortcut menus 4
  show quantitation spectrum 70
  size changes 28
  sort and filter 77
  start workflow 56
T
tabs
  filters 37
  peptide confidence 37
  peptides 37
  proteins 37
  summary 37
Terminal side
  C- 109, 150
  N- 109, 150
TMT
  quantitation 18
  tagging 18
toolbar
  customizing 131
  customizing layout 132
  display 135
tools
  FASTA Database Utilities 62, 64
  InforSense 75
  open Qual Browser 72
tooltips
  customizing 136
  hiding 136
total fraction 17
U
UNIMOD, importing data of 142
user interface change options 28

V
validation of quantitation results 67
view
peptide consensus 108, 110, 111
peptide consensus page 107
protein page 60
view size 28

W
wizards
four key settings 29
raw file 30
workflow
create a new Proteome Discoverer search 53
develop a specific template 50
Discoverer - align 53
Discoverer - blue box edge 52
Discoverer - open 56
Discoverer - rawfile 53
Discoverer - red box edge 52
InforSense - metadata table NCBI 128
InforSense - metadata table NCBI (how to use) 125
InforSense - metadata table SwissProt 129
InforSense - metadata table SwissProt (how to use) 125
InforSense - NCBI 126
InforSense - NCBI (how to use) 125
InforSense - SwissProt 127
InforSense - SwissProt (how to use) 125
InforSense discussion 124
Proteome Discoverer nodes 57
quantitization node 69
save a search in Proteome Discoverer 56
start in Proteome Discoverer 56
Workflow Editor
auto layout 53
cautitory note - learn workflow node functions 50
introduction to node selections 51
join two nodes 52
learn 52
menu 49
with respect to all Discoverer processes 19
working with InforSense Protein Annotation 125
workspace 51

X
XCorr
counts the number of fragment ions 44
report table 117
XIC
extracted ion chromatogram 98
XML, creating reports 18

Y
yellow range, chromatogram view 100

Z
ZCore
ETD and ECD data 16
false positive rate 16