ProSightPC®

User Guide

XCALI-97172 July 2006

Analyze • Detect • Measure • Control™

Thermo ELECTRON CORPORATION

ProSightPC, Sequence Gazer, and PTM Warehouse are registered trademarks of The Board of Trustees of the University of Illinois.

Celeron and Pentium are registered trademarks of Intel Corporation in the United States and/or other countries.

Windows is a registered trademark of Microsoft Corporation in the United States and/or other countries.

RESID is a registered trademark of John Garavelli.

The procedure THRASH is based on routines in *Numerical Recipes: The Art of Scientific Computing*, Published by Cambridge University Press, and are used by permission.

All other trademarks are the property of Thermo Electron Corporation and its subsidiaries.

Printing History: Revision A released in July 2006

Software Version: 1.0

Contents

Preface

About This Guide

Welcome to ProSightPC® 1.0. The ProSightPC suite provides you with an extensive array of tools for the identification and characterization of proteins. Additional tools enable you to create, edit, and save the results of your analyses.

This guide describes the ProSightPC suite, additional ProSightPC applications, as well as the methods for identifying and characterizing proteins. Along with a thorough review of all functions and options, this guide provides helpful advice and tips on how to use ProSightPC to optimize day-to-day use of the software suite.

Safety and Special Notices

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 avoid damage to software, loss of data, invalid test results, or information critical for optimal performance of the system.

Note Highlights information of general interest.

Tip Helpful information that can make a task easier.

Contacting Us

There are several ways to contact Thermo Electron.

Assistance

For new product updates, technical support, and ordering information, contact us in one of the following ways:

Visit Us on the Web

www.thermo.com/finnigan

Contact Technical Support

Phone: 1-800-685-9535
Fax: 1-561-688-8736
techsupport.finnigan@thermo.com

Contact Customer Service

In the US and Canada for ordering information:
Phone: 1-800-532-4752
Fax: 1-561-688-8731

International contacts for ordering information:
Visit www.thermo.com/finnigan for the current listing,

Changes to the Manual

To suggest changes to this guide, use either of the following methods:

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

Chapter 1 Top-Down Proteomics

Top-down proteomics is a powerful technique for protein identification and characterization.

This chapter provides an overview of top-down proteomics. It contains the following sections:

Introduction to Proteomics

Top-down proteomics is a rapidly emerging solution to the problems of protein identification and characterization. Unlike other proteomic techniques, in top-down proteomics the unknown proteins are not digested into peptides prior to tandem mass spectrometry. The primary advantage of this technique is that full sequence coverage on every protein is guaranteed. This enables complex combinations of post-translational modifications (PTMs) to be easily characterized.

Separation and ionization of intact proteins present many challenges beyond the scope of this manual. The bioinformatics of top-down proteomics, though, does have some challenges that ProSightPC addresses. First, because the precursor ions are large, they are almost always multiply charged. This complicates spectral comparison techniques used in certain other proteomic strategies. Fortunately, this issue can be avoided by doing all comparisons with neutral masses. Once MS and MS/MS spectra are collected, sum the relevant scans, then run an automated analysis to infer mass using the resulting mass values for protein identification and characterization. Second, because the precursor mass can represent either a highly modified protein or an internal fragment of the intact protein, no single strategy of comparing the observed mass values to a proteome database is guaranteed to identify the protein. For this reason, ProSightPC provides the following search modes:

- Absolute mass
- Biomarker
- Sequence tag
- Gene restricted

Each search mode overcomes the different issues of protein identification and characterization.

The fundamental unit of analysis in top-down proteomics is the MS/MS experiment. In this experiment, intact precursor protein molecules are ionized and subjected to mass spectrometry. A single peak (usually representing one charge state of the unknown protein, but sometimes representing a small number of isobaric proteins) is isolated and subjected to fragmentation. The accurate mass measurement of the resulting MS/MS fragment ions provides the second vital piece of information. This MS and MS/MS mass data is then compared to prior information about protein sequences and known or predicted PTMs in order to identify and characterize the unknown protein

Note ProSightPC includes the RESID® database. All post-translational modifications are identified to the system by a truncated form of their RESID identification number. ProSightPC removes leading zeros and the letters AA from the start of a RESID identifier. This identifier is placed in parenthesis before the amino acid containing the modification. Thus, AA0049 N-acetyl-L-methionine becomes 49, and an acetylation of a methionine residue in a protein sequence is expressed as (49)M.

Shotgun Annotation

ProSightPC relies on shotgun annotation of PTM information.

This process takes PTM events on a single protein and precalculates all possible combinations, regardless of whether the particular combination has ever been observed.

If, for example, a given protein is known to have only four phosphorylation sites and no other modifications, during shotgun annotation a record is entered into the proteome database for the base sequence with no PTMs. ProSightPC also enters

- Four records for the four protein forms, each containing one modification
- Six records for the possible combination of two phosphorylations
- Four records for the three triphosphorylated forms
- One record for the form with all four possible phosphorylations

All of these combinations are processed, even if the phosphorylation events have only been observed separately.

The advantage of this technique is if one of the multiphosphorylated forms occurs in nature and is observed in an MS/MS experiment, it will be readily identified.

Unfortunately, this also means the number of records needed to describe a protein in a shotgun annotated database grows as 2 to the power of *n*, where *n* is the number of modifications on the protein. This growth rate means it is impractical to store all possible forms for highly modified proteins. ProSightPC overcomes this limitation by first determining if the protein is going to need more than 1,000 records to fully describe it. If it does, then ProSightPC prioritizes the protein forms, storing only those forms most likely to lead to protein identification.

Search Modes and the Top-Down Funnel

In all but the most exceptional cases, top-down proteomics only receives partial fragment information in the MS/MS phase, meaning there is no guarantee that all the information necessary to fully characterize an unknown protein will be observed. This limitation leads to what is known as the top-down funnel.

Figure 1 shows a schematic representation of the top-down proteomics funnel. The top of the funnel represents the space of all possible observed combinations of MS and MS/MS data. A certain area at the top contains those combinations allowing the identification of the unknown protein and full characterization of any PTM present. Additional combinations allow for the identification and a partial characterization of the protein. In some cases, it is possible to only identify the protein.

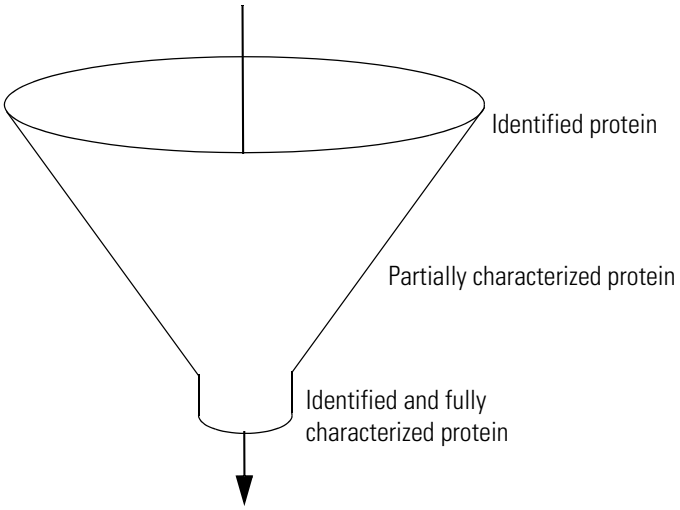

Figure 1. Results of full characterization from initial data

Note Partial characterization occurs in cases where it is possible to determine which PTM must be present on the protein, but the fragmentation data is not sufficient to make the determination of exactly which amino acid one or more of the PTMs must be on. It is frequently possible to narrow the list of possibly modified amino acids down to one or two residues within a short subsequence of the protein.

In some MS/MS experiments you have sufficient fragmentation data to fully characterize the proteins with the first search. If the exact protein form observed was shotgun annotated into the proteome database and sufficient fragmentation information was present in the MS spectra to uniquely identify this form, then the correct answer is discovered by conducting an

absolute mass search. This situation occurs frequently. In some well-annotated proteomes, unknown protein are completely characterized on the first search. Figure 1 illustrates this situation.

Figure 2 illustrates another common situation when the initial search only identifies, and perhaps partially characterizes, the unknown protein. In this case, conducting a second search fully characterizes the protein. Often the second search will be either a biomarker or single protein mode search, though in some search strategies a gene-restricted search will be conducted. These search modes are described in Chapters 6 and 7.

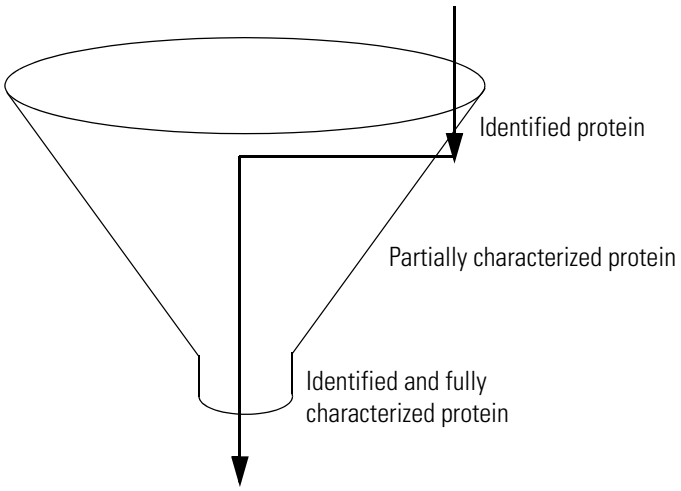

Figure 2. Multiple searches can be used for the identification and characterization of the unknown protein

Figure 3 illustrates when fragmentation data is insufficient to distinguish between two or more possible protein forms. In this case, full identification or partial characterization is the best possible result. When this occurs, re-running the MS/MS experiment can yield greater fragmentation data.

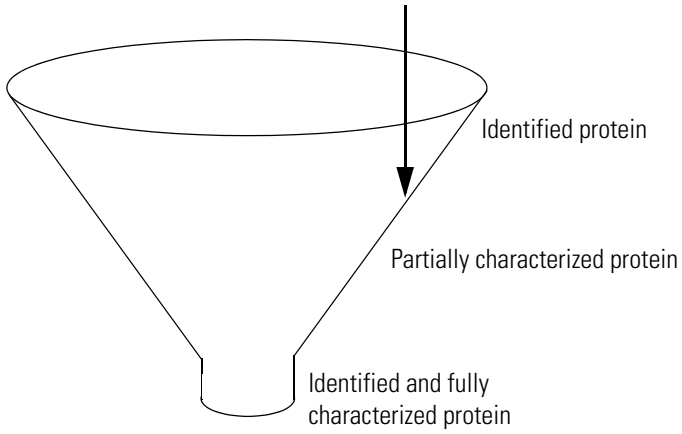

Figure 3. Results of identification and/or partial characterization

Chapter 2 Installing the Software

This chapter describes how to install ProSightPC. It contains the following sections:

System Requirements

Thermo Electron, Inc. recommends the following hardware and software requirements.

Hardware Requirements

Table 1 describes the minimum and recommended hardware requirements.

Table 1. Minimum and recommended hardware requirements

Minimum Requirements	Recommended Requirements
Intel Pentium® III or Celeron® 500 mHz processor	Pentium IV or Celeron 1 GHz processor
500 MB RAM 333 mHz	1 GB 666 MHz RAM
32 MB graphics card	64 MB or greater
24x CD-ROM	
Color monitor	
Keyboard	
Mouse	

Software Requirements

The following software is required:

- Minimum: Microsoft® Windows® 2000 (Recommended: Windows® XP)
- MYSQL (must be installed prior to ProSightPC installation)
- Additional proteome databases (for installation into the PTM Warehouse®)

Note ProSightPC operates on the Windows XP environment and is not guaranteed to function on any other platform.

Installing ProSightPC

To install ProSightPC

1. Insert the ProSightPC CD-ROM into your CD drive. The installation wizard automatically starts, as Figure 4 shows.

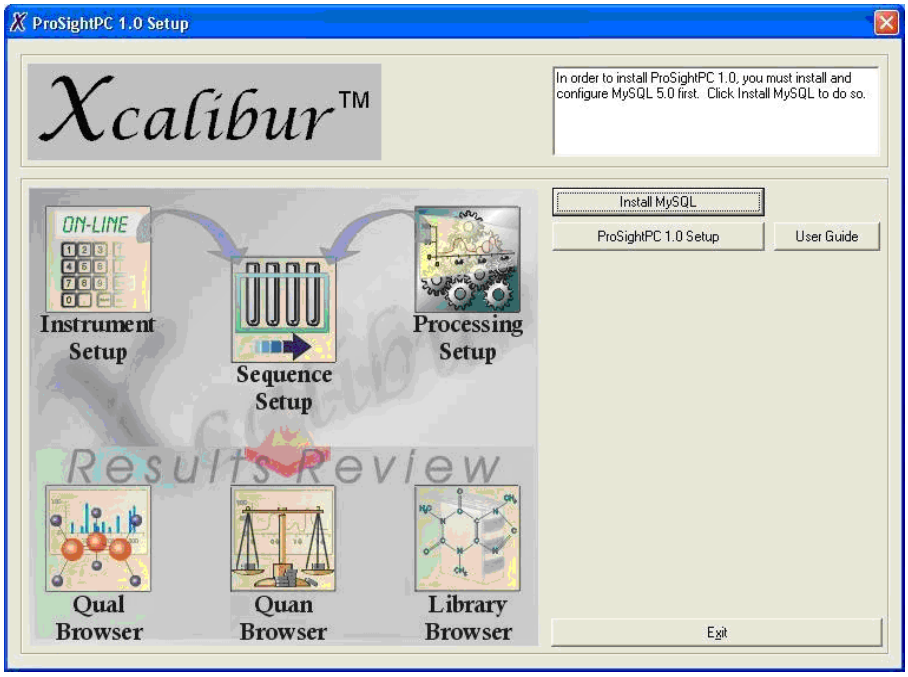

Figure 4. ProSightPC Setup Wizard

If the installation wizard does not start automatically, select **Start > Run.** Browse to the CD-ROM and select **Xinstall**.

2. Click **Install MySQL** to launch the MySQL installer. The installation wizard automatically starts, as Figure 5 shows.

Note ProSightPC requires that you install MySQL before installing ProSight PC.

Figure 5. MySQL Setup Wizard screen

3. Click **Next**. The MySQL License Agreement dialog box appears, as Figure 6 shows.

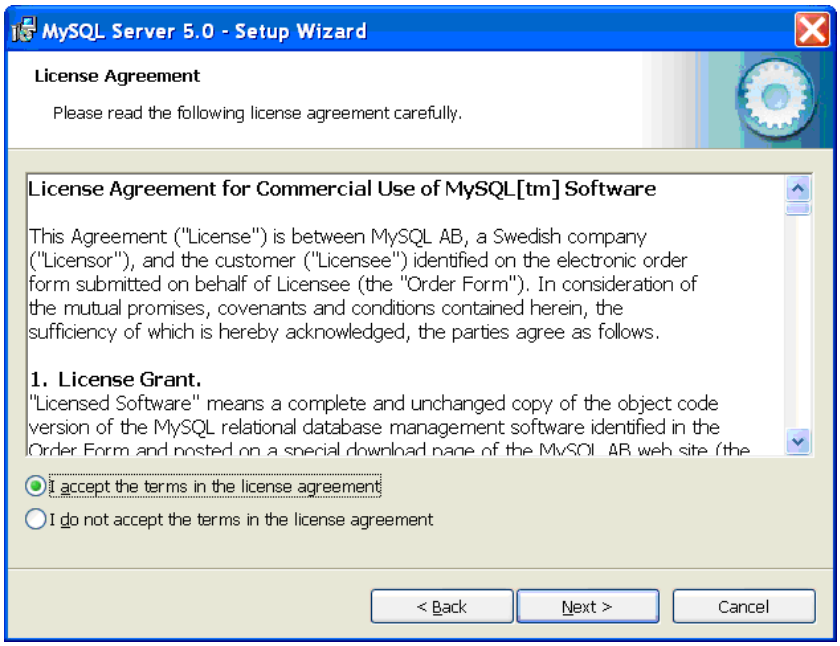

Figure 6. MySQL License Agreement screen

4. Click **I accept** and click **Next** after you read the agreement.

5. Follow the installation instructions.

6. Choose any one of the options shown in Figure 7 and click **Next**. It is not necessary to sign up for an account to continue the installation process.

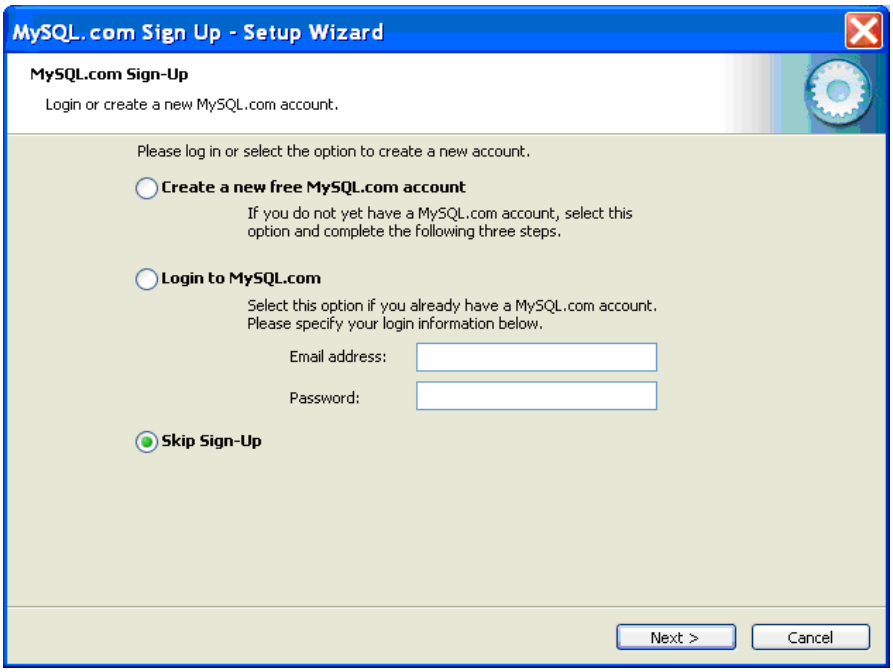

Figure 7. MySQL Account Sign Up screen

7. Select **Configure the MySQL Server now** and click **Finish**, as Figure 8 shows.

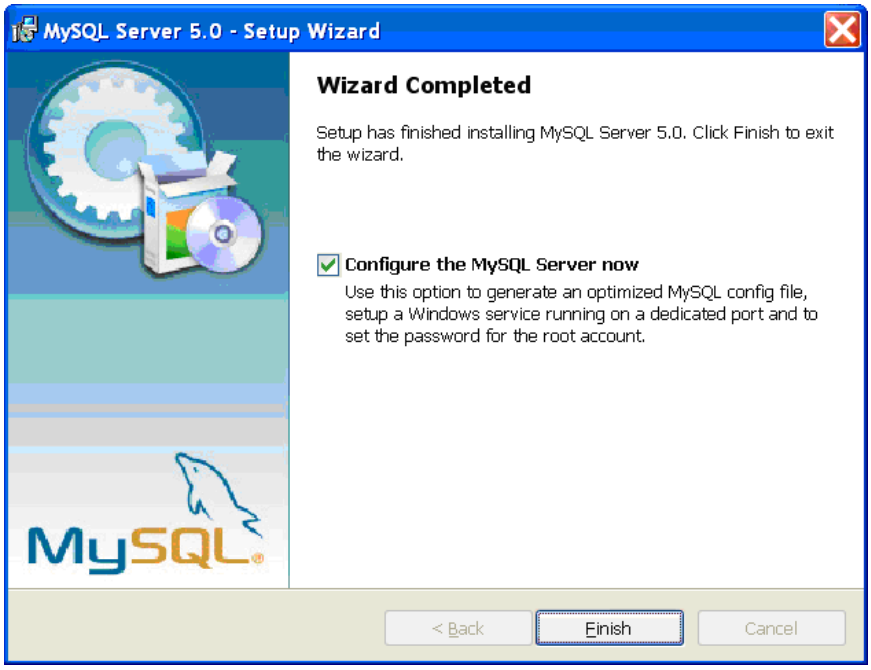

Figure 8. MySQL Server Configuration screen

8. Select **Standard Configuration** and click **Next**, as Figure 9 shows.

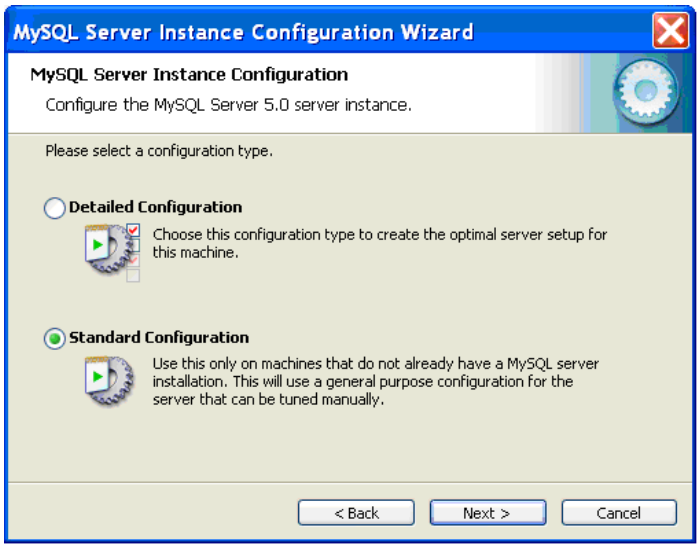

Figure 9. MySQL Server Instance Configuration Type dialog box

9. Select **Install as Windows Service** and click **Next**, as Figure 10 shows.

Figure 10. MySQL Server Instance Windows options dialog box

10. Deselect **Modify Security Settings** and click **Next**, as Figure 11 shows.

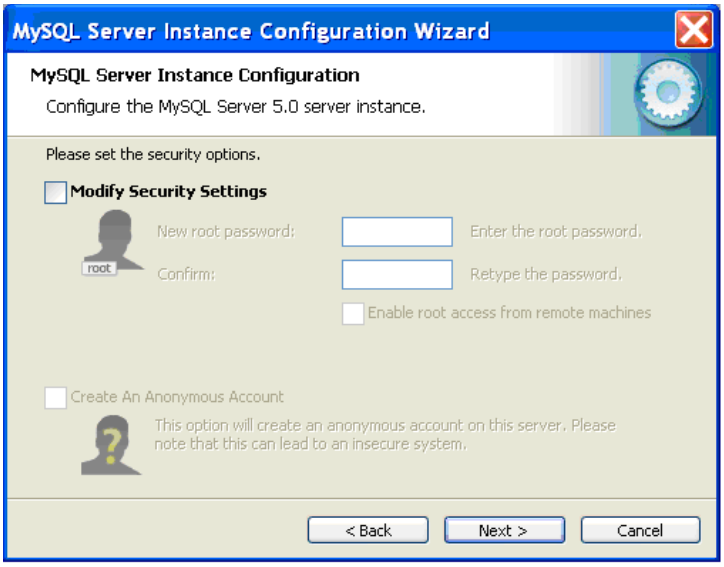

Figure 11. MySQL Server Instance Security options dialog box

Note If you deselect **Modify Security Settings**, no password is created for the root account. If you create a password, you must remember this password in order to complete the ProSightPC installation, as Figure 17 shows.

11. Click **Execute** to begin the MySQL Server configuration, as Figure 12 shows.

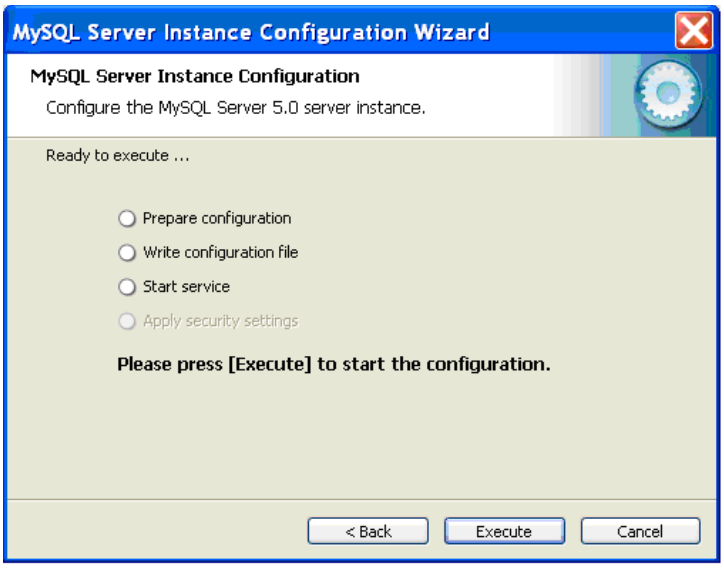

Figure 12. MySQL Server Instance Execute options dialog box

12. Click **ProSightPC 1.0 Setup** and follow the instructions to install ProSightPC, as Figure 13 shows.

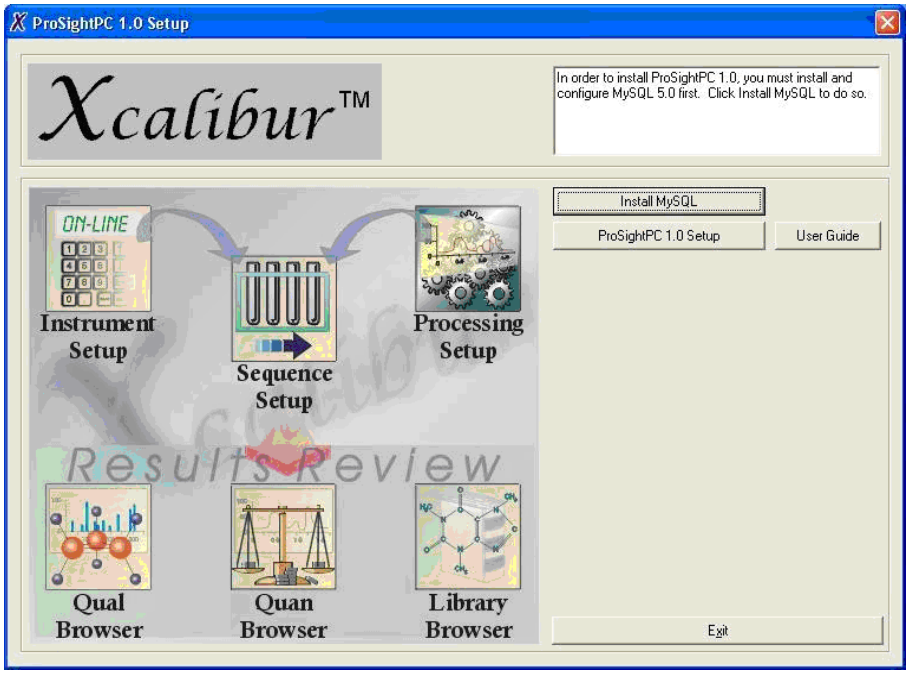

Figure 13. ProSightPC Setup window

The License Agreement dialog box appears, as Figure 14 shows.

Figure 14. ProSightPC License Agreement screen

13. Click **I agree** and click **Next** after you read the agreement.

    The Select Installation Folder dialog box appears, as Figure 15 shows. The default installation folder is C:\Program Files\ProSightPC\.

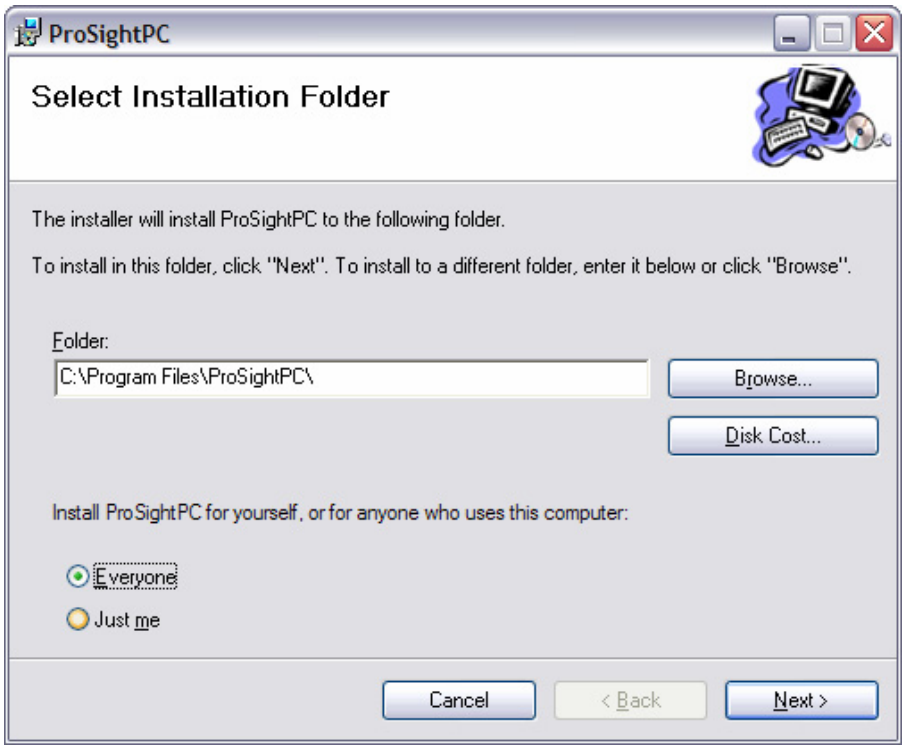

Figure 15. Select Installation Folder screen

14. Click **Browse** to change the default installation folder. Choose a location from the list that appears.

15. Select **Everyone** or **Just me** and click **Next** to define access to ProSightPC.

    The Add Shortcuts dialog box appears, as Figure 16 shows. Thermo recommends that, at a minimum, you select the Start Menu shortcut.

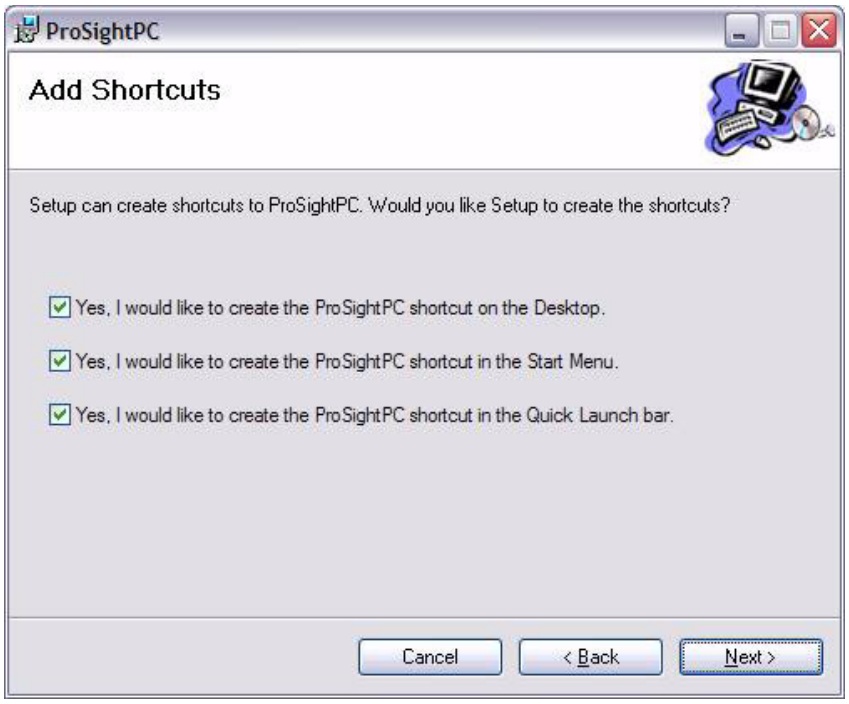

Figure 16. Add Shortcuts screen

16. Click **Next** to begin the installation. The installation wizard prompts you for your MySQL administrator user name and password, as Figure 17 shows. Use the password you created in step 10. If you did not create a password, leave the box empty. Click **OK** to continue the installation.

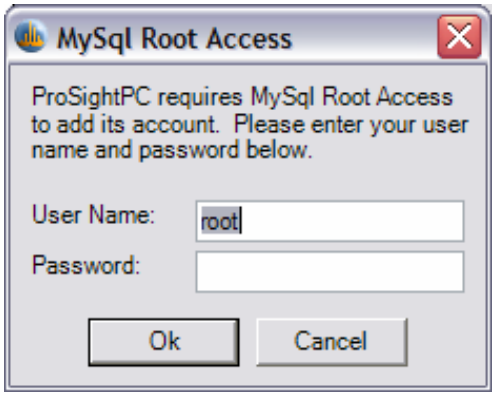

Figure 17. MySQL Root Access dialog box

The wizard begins to install the software, as Figure 18 shows.

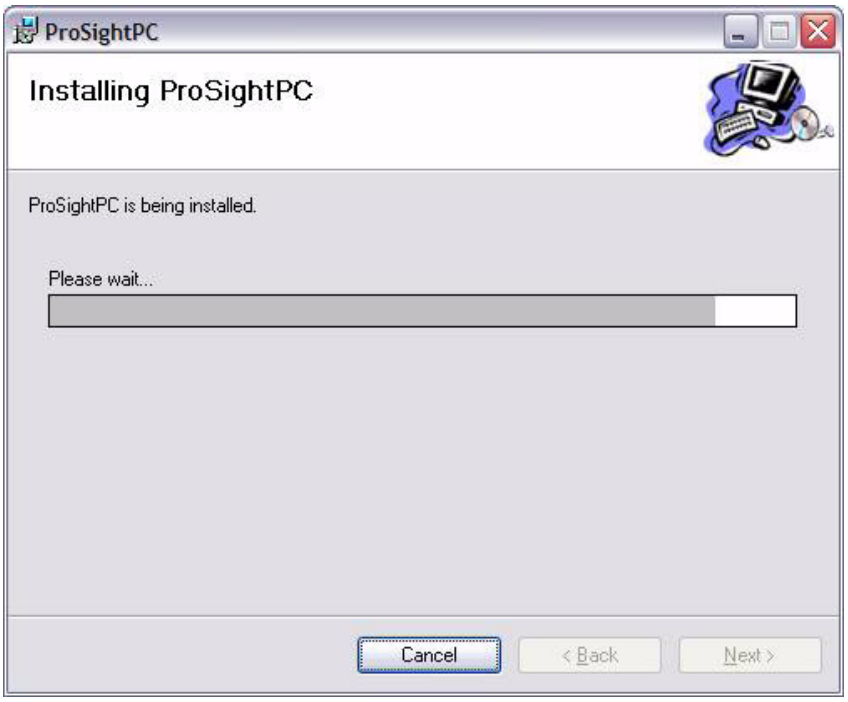

Figure 18. Installing ProSightPC status screen

17. Click **Close** to end the installation process, as Figure 19 shows.

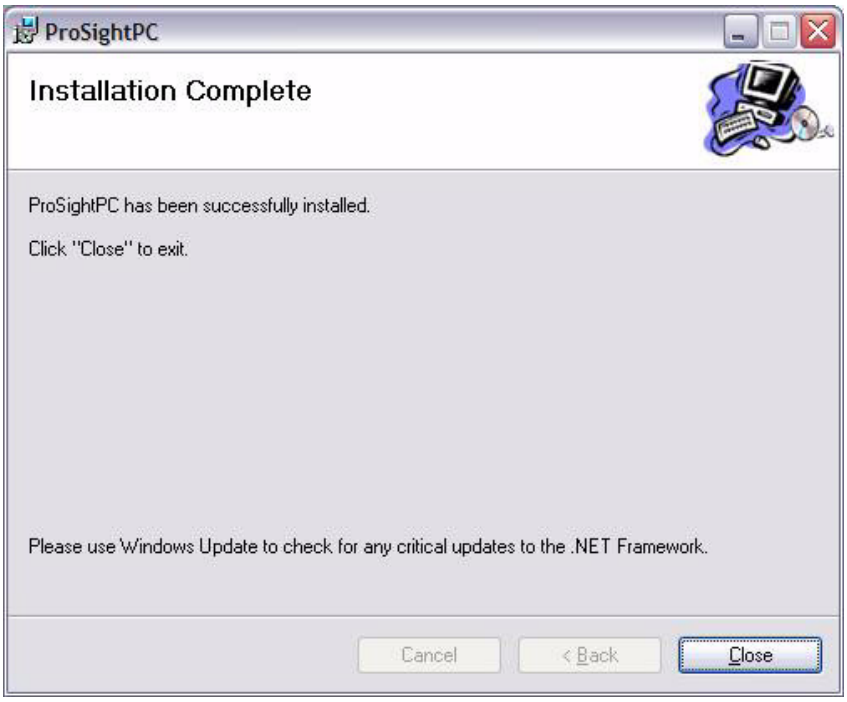

Figure 19. Installation Complete screen

To install additional proteome databases

1. Select **Start > All Programs > ProSightPC > Database Manager** to open the Database Manager.

   The Database Manager window appears, as Figure 20 shows.

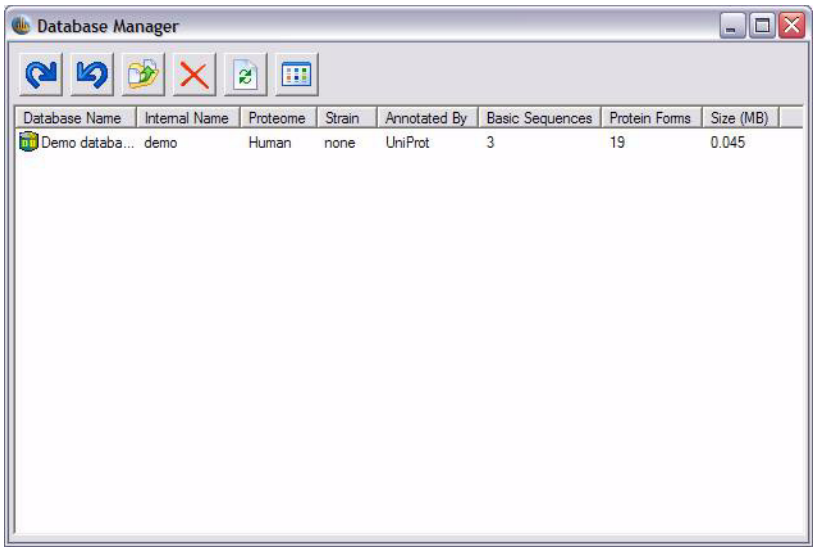

Figure 20. Database Manager window

2. Click **Import Database**. Select the proteome database to install.

   Additional proteome databases are saved as ProSightPC Warehouse Files or PWF files. These files are provided in the ProSightPC Warehouse Files folder on the ProSightPC CD-ROM. For more information, see Chapter 9, "Proteome Databases."

3. Click **Import** to install the additional database.

   If a database already exists with the same internal name as the database you are importing, you can choose to skip the conflicting database, rename the database you want to import, or overwrite the existing database.

4. Confirm that the database information is correct in the Database Manager.

Chapter 3 Understanding ProSightPC

ProSightPC is composed of several components. Understanding the organization of these components makes the application easier to use.

This chapter provides an overview of the organization of ProSightPC. It contains the following sections:

ProSightPC Overview

ProSightPC is designed to solve the problem of identification and characterization of proteins from mass spectrometry data. ProSightPC functions in the following manner. Prior protein sequences along with PTM information are gathered and loaded into a proteome database. During loading, the PTM information is shotgun annotated into the protein sequence information. The shotgun-annotation process calculates all possible combinations or known PTMs onto each protein sequence. After the annotation process, a new proteome database is created. Next, the mass values inferred from mass spectral data from top-down proteomics MS experiments are imported into a ProSightPC Upload File (PUF). These mass values are compared to the appropriate proteome databases.

Figure 21 illustrates an overview of the flow of information within ProSightPC. In step 1, sequence and PTM information are shotgun annotated into a proteome database within the PTM Warehouse. This is only done once per proteome. Step 2 involves an analysis to infer mass to acquire the observed neutral mass values from the RAW data file into a PUF file. Step 3 depicts how ProSightPC identifies and characterizes the observed proteins.

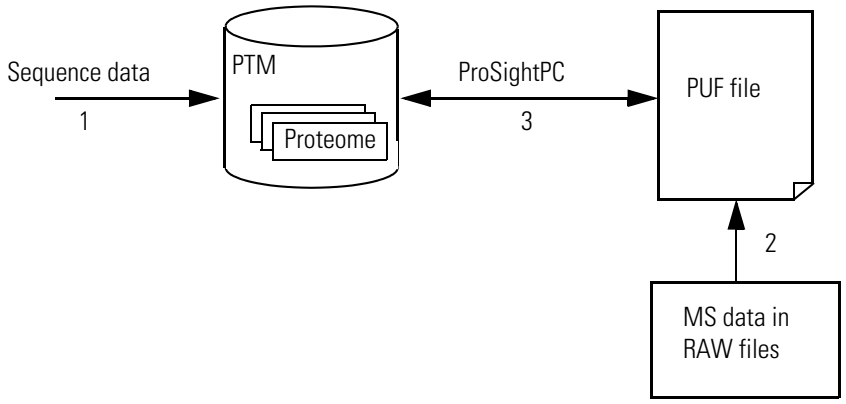

Figure 21. ProSightPC information flow

The ProSightPC Suite

ProSightPC is a suite of tools designed to help identify and characterize proteins. It consists of the ProSightPC application, one instance of the PTM Warehouse, and a small number of secondary applications to aid in proteome database and experimental results management.

The ProSightPC Application

ProSightPC operates on a single PUF file which, when opened, is uploaded into memory and made available to a variety of search and data visualization tools. Additionally, ProSightPC has several tools for importing tandem MS data from Thermo Electron RAW files, identifying and removing chemical noise peaks, and other utility functions.

The PTM Warehouse

The PTM Warehouse is a relational database containing the information ProSightPC needs to function. The primary component of the PTM Warehouse is a series of schemas, one for each proteome database containing PTM and protein sequence information.

Secondary Applications

ProSightPC comes with two secondary applications: the Database Manager and the Experiment Manager. The Database Manager provides a point-and-click environment for managing the PTM Warehouse. The Database Manager imports and exports PWF files as well enabling you to create your own proteome databases. The Experiment Manager provides a simple interface for moving MS/MS experiments between different PUF files.

File Types

ProSightPC has two unique file types. The first file type is the ProSightPC Warehouse File, identified by the .pwf file extension. These files are a transport format for moving proteome databases between PTM Warehouses. Thermo Electron supplies the standard ProSightPC proteome databases in this file format. You can also create your own PWF files for exchanging proteome information. The second file type is the ProSightPC upload format, identified by the .puf file extension. The PUF file is a unique XML file format used to store and transport ProSightPC results. Each PUF file can consist of one or more MS experiments. A single MS experiment can contain mass lists for both precursor and fragment ions extracted from the MS and MS/MS spectra. Each PUF file includes this information, related search parameters, and any search results associated with this data.

Fixed Modifications

A fixed modification is a chemical alteration of an amino acid residue that is present on all amino acids of a given type (for example, alkylation of cysteine residues).

Chapter 4 Navigating the Program

ProSightPC has a unique approach to the organization of elements within the graphical user interface.

This chapter describes how to navigate the ProSightPC program. It contains the following sections:

The Organization of the Interface

Figure 22 shows the three window segments of ProSightPC:

- Data Grid
- Tab Controller
- Job Queue

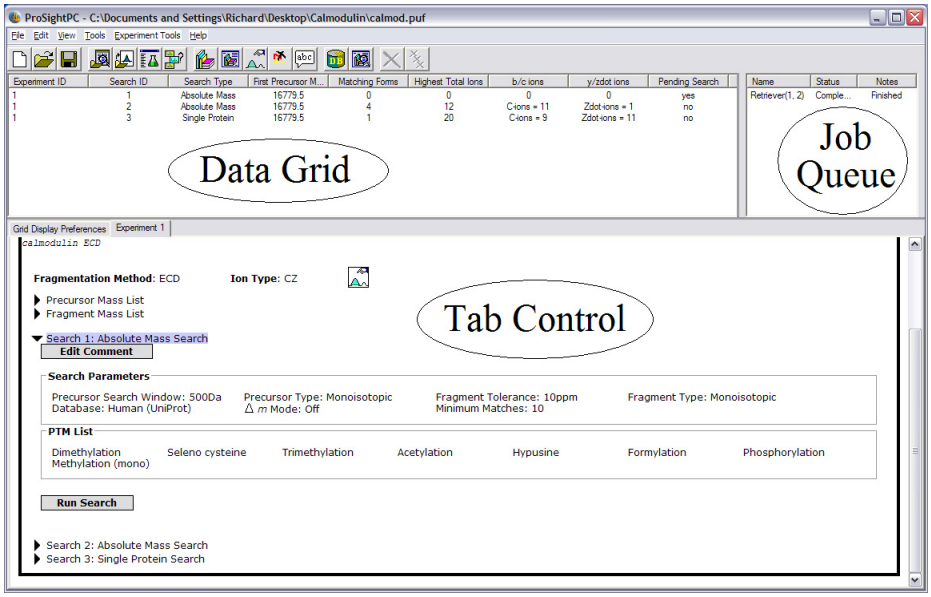

Figure 22. ProSightPC window

Data Grid

The Data Grid displays summary information about each search contained within the open PUF file. It is organized into various columns, which you can change by using the Grid Display Preferences Tab Controller, discussed in detail in the following sections.

Tip

1. Right-click on a search within the Data Grid to apply several actions.
2. Double-click on a search within the Data Grid to open the relevant Data Manager. For more information, see Chapter 8, "Understanding the Data Manager."

Tab Controller

Many of the more complex interface elements of ProSightPC appear in the Tab Controller. Double-click an experiment in the Data Grid to display the experiment in the Tab Controller. For more information, see Chapter 8, "Understanding the Data Manager."

Tip Right-click on a tab to close or hide the tab.

Job Queue

Use the job queue to display the status of any previously run or currently running searches in the ProSightPC session. While a search is running, a status bar displays the progress of that search.

Table 2 describes the organization of the job queue.

Table 2. Job Queue

Item	Description
Name	Lists the MS/MS search identifiers associated with the job.
Status	Displays the status of the current search: Pending indicates the search has yet to be run Running indicates the search is currently running Completed indicates the search has been successfully run Failed indicates the search ended abnormally
Notes	Displays additional information about searches. For example, if the search failed, this includes notes explaining why the search failed.

Tip Right-click on a job within the job queue to display the following options:

1. If the job is not currently running, right-click and select run to queue the experiment to run next.
2. If the job is running, right-click the experiment and select **Abort**. The search ends and the status changes to Failed.

The Menu Bar

The various features within ProSightPC are contained in the menu bar. Table 3 describes how the menu bar is organized.

Table 3. Menu Bar

Item	Description
File	Includes functions related to PUF file manipulation, including importing data from the RAW files.
Edit	Includes Cut, Copy, and Paste.
View	Includes user-selected displays: grid preferences, start screen, job queue, and toolbar
Tools	Includes Experiment Adder, Manage Predefined Searches, PTM Tier Editor, View Database Information, Individual Sequence Adder, Font Converter, Fragment Predictor, Batch Run, Reports, and Options.
Experiment Tools	Includes Add Search, Append Predefined Search, Edit Masses, Reduce Noise, and Edit Comment. This menu is only available when an experiment is open in the Data Manager.
Help	Includes About.

The Toolbar

Table 4 describes how the toolbar is organized.

Table 4. The Toolbar

Item	Description
New, Open, Save	Use to create, load or save PUF files
Import XTRACT	Use for building a new experiment within the current PUF file using post-XTRACT RAW data.
Import Profile	Use for building a new experiment within the current PUF file using high-resolution RAW data obtained in profile mode.
Add Experiment	Use for building a new experiment within the current PUF file using manually input MS and MS/MS data.
Open Data Manager	Use to open the last experiment using the in the open PUF file if there are no Experiment tabs open in the Tab Controller.
Add Search	Use to append a new search to the selected experiment. Available only when an experiment is open and shown in the Tab Controller.
Append Predefined Search	Use to append a predefined search to the selected experiment. Available only when an experiment is open and shown in the Tab Controller.
Edit Masses	Use to change MS and MS/MS data within ProSightPC. Opens a new tab showing the precursor and fragment masses of the current experiment.
Reduce Noise	Use to display the noise reducer options.
Edit Comment	Use to edit the comment at the top of the current experiment.
View Database Information	Use to view information about the proteome databases in the PTM Warehouse.
Manage Predefined Searches	Use to modify the parameters of predefined searches.
Abort Running Job	Use to end the current search in the job queue.
Abort All Jobs	Use to end all current and pending searches in the job queue.

The Grid Preferences Tab

The Grid Display Preferences tab enables you to control what information is displayed within the Data Grid. The Grid Display Preferences tab automatically appears when a PUF file is opened.

Figure 23 shows the Grid Preferences tab.

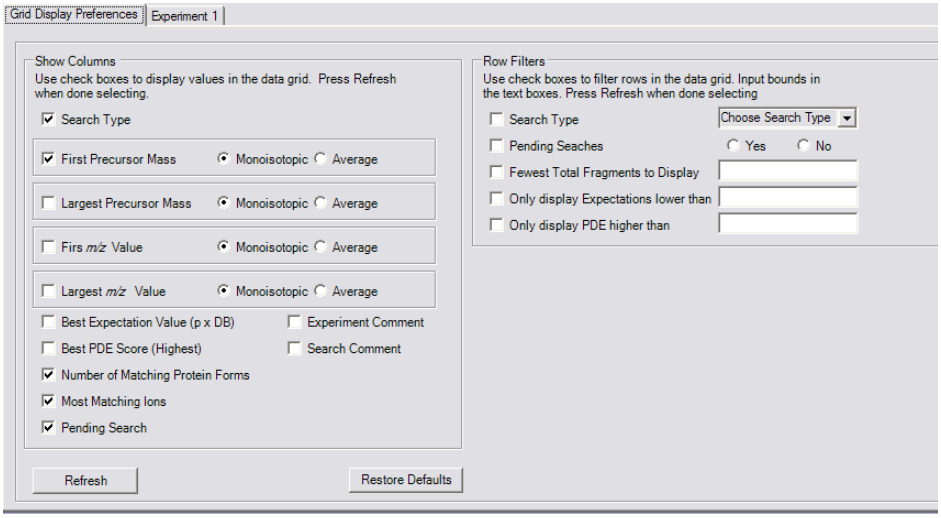

Figure 23. Grid Preferences tab

Tip Clicking **View > Grid Preferences** also displays this tab.

The Grid Display Preferences tab consists of two controls:

- Show Columns
- Row Filters

Show Columns

The Show Columns option defines which columns are visible in the Data Grid.

To add a column to the Data Grid

1. Place a check beside any columns you want to display within the Data Grid.
2. Click **Refresh** to implement your selections.

   The columns you selected appear in the Data Grid.

To remove a column from the Data Grid

1. Click to remove a check from beside any columns you want removed from the Data Grid.

2. Click **Refresh** to implement your selections.

Note You can also access the Show Columns options available in the Data Grid by selecting **Tools > Options**.

Each option in Table 5 controls the appearance of a column in the Data Grid.

Table 5. Show Columns and Descriptions

Column	Description
Search Type	Displays a column showing the search type.
First Precursor Mass	Displays a column showing the first precursor mass in the mass list.
Largest Precursor Mass	Displays a column of the precursor mass with the highest intensity value.
First *m/z* Value	Displays a column with the *m/z* value of the first precursor entry.
Largest *m/z* Value	Displays a column with the largest *m/z* value of all precursor entries.
Best Expectation Value (p x DB)	Displays a column with the lowest expectation score of any hit in the search results.
Best PDE Score (Highest)	Displays a column with the highest PDE score of any hit in the search results.
Number of Matching Protein Forms	Displays a column with the total number of protein forms returned by the search.
Most Matching Ions	Displays a column with the largest number of matching fragments of any hit.
Pending Search	Displays a column indicating whether a search has been run.

Row Filters

The Row Filters option controls which searches are displayed within the Data Grid. By checking one or more of the criteria, you filter (or omit from view), certain Data Grid rows.

Table 6 describes row filters.

Table 6. Row Filters

Row Filter	Description
Search Type	Displays only those searches of the type selected.
Pending Searches	Displays pending or completed searches.
Fewest Total Fragments to Display	Displays those searches that have one or more hits with at least the indicated number of matching fragments.
Only Display Expectations Lower Than	Displays only those searches that have one or more hits with an expectation score lower than the indicated value.
Only Display PDE Higher Than	Displays only those searches that have one or more hits with a PDE score higher than the indicated value.

Note Click **Restore Defaults** to reinstate the default settings. The default settings can be changed with the Options dialog box.

The Options dialog box

Use the Options dialog box to set default values for most of the interface elements within ProSightPC.

To access the Options dialog box

- Choose **Tools > Options**. The dialog box shown in Figure 24 appears.

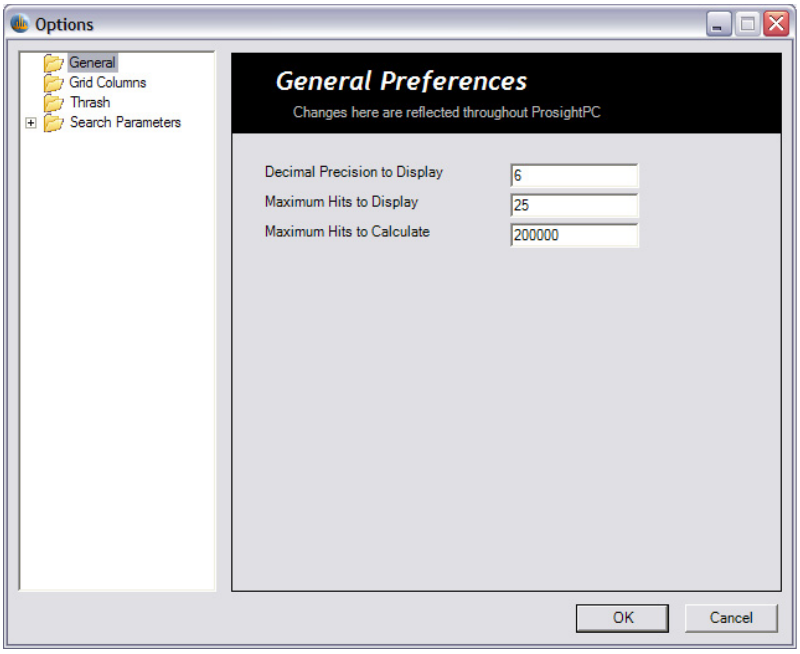

Figure 24. Options dialog box

Set the following preferences from the Options dialog box:

- General Preferences
- Grid Columns
- Thrash Preferences
- Search Parameters

General Preferences

The General Preferences window has three options that are available throughout much of the ProSightPC interface. These preferences are described in Table 7.

Table 7. General Preferences

Preference	Description
Decimal Precision to Display	Defines the number of decimal places to which most numbers are displayed.
Maximum Hits to Display	Defines the number of matching fragment tables that will be displayed within the Data Manager. Only the best matches up to this number will be displayed.
Maximum Hits to Calculate	Defines the maximum number of protein forms that a given search will consider. When exceeded, the search aborts and a warning is issued.

Grid Columns

The Grid Columns dialog box has the options shown in Figure 25. Use these options to set defaults in the Grid Preferences tab.

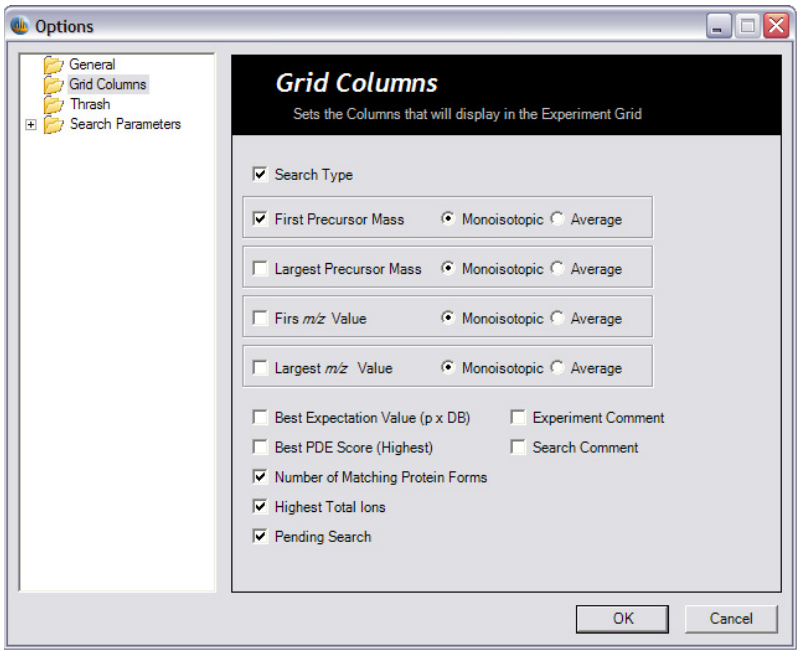

Figure 25. Grid Columns dialog box

THRASH Preferences

Use the THRASH Preferences dialog box shown in Figure 26 to set the following default values related to the THRASH algorithm:

- Minimum S/N Ratio
- Minimum RL Value

- Maximum Mass
- Maximum Charge

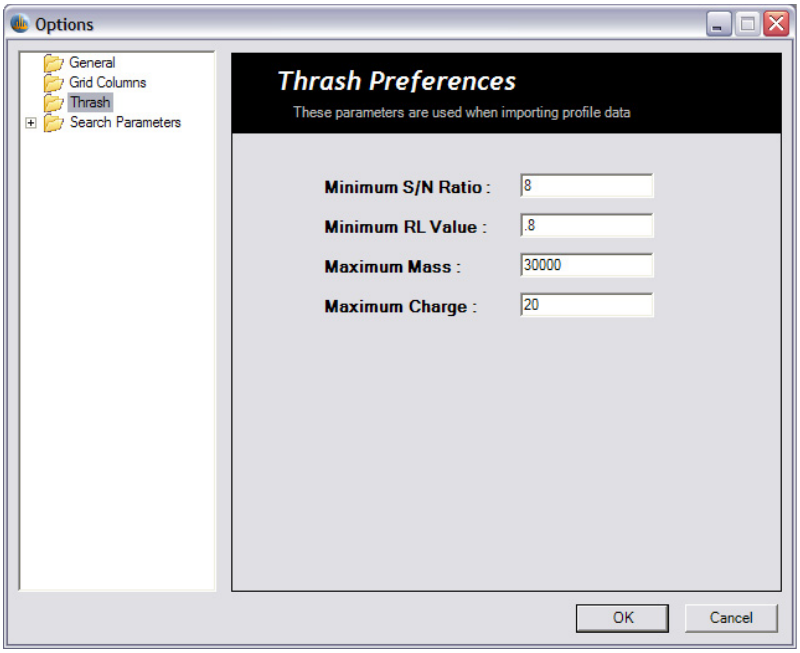

Figure 26. Thrash Preferences dialog box

Search Parameters

Click on Search Parameters to display the following search modes:

- Absolute Mass
- Biomarker
- Sequence Tag
- Single Protein

Absolute Mass Preferences

Use the Absolute Mass Preferences tab to set the default value used when adding new absolute mass searches. Click on the Absolute Mass Preferences tab and the dialog box shown in Figure 27 appears.

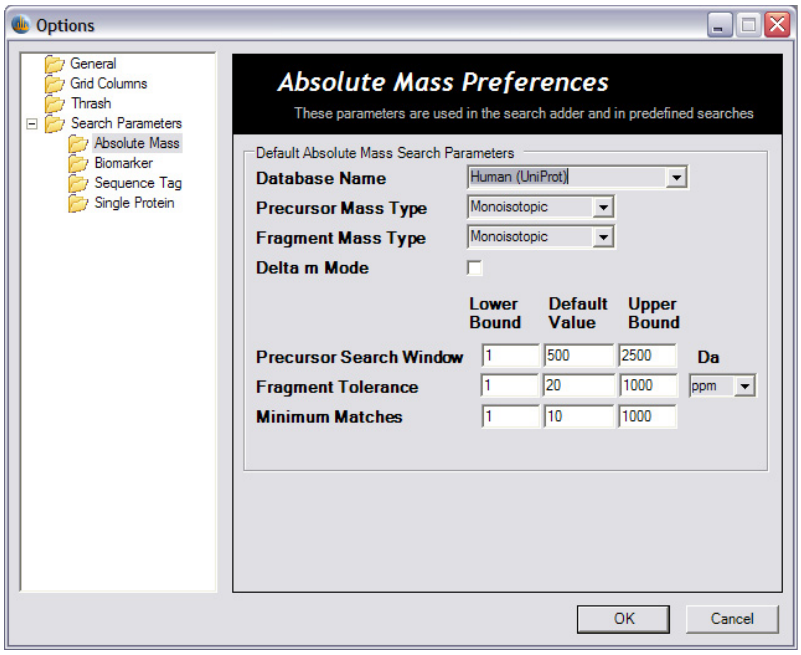

Figure 27. Absolute Mass Preferences dialog box

Set the following parameters from this interface:

- Database Name
- Precursor Mass Type
- Fragment Mass Type
- Δ*m* Mode

Table 8 lists those elements that must be assigned a numerical value.

Table 8. Absolute Mass Parameters

Preferences	Description
Precursor Search Window	Lower bound sets the minimum value for an precursor search window that does not trigger an "out of range" warning. Default value sets the default value for an precursor search window. Upper bound sets the maximum value for an precursor search window that does not trigger an "out of range" warning.

Table 8. Absolute Mass Parameters, continued

Preferences	Description
Fragment Tolerance	Lower bound sets the minimum value for an fragment tolerance that does not trigger an "out of range" warning. Default value sets the default value for an fragment tolerance. Upper bound sets the maximum value for an fragment tolerance that does not trigger an "out of range" warning. You can set values as either Da or ppm.
Minimum Matches	Lower bound sets the minimum value for minimum matches that does not trigger an "out of range" warning. Default value sets the default value for an minimum matches. Upper bound sets the maximum value for an minimum matches that does not trigger an "out of range" warning.

Note Gene-restricted absolute mass searches draw their parameters from absolute mass searches.

Biomarker Preferences

Use the Biomarker Preferences window to set the default values used when adding new biomarker searches. Click the Biomarker Preferences tab and the dialog box shown in Figure 28 appears.

Set the following options from the Biomarker Preferences dialog box:

- Database Name
- Precursor Mass Type
- Fragment Mass Type
- Δm Mode

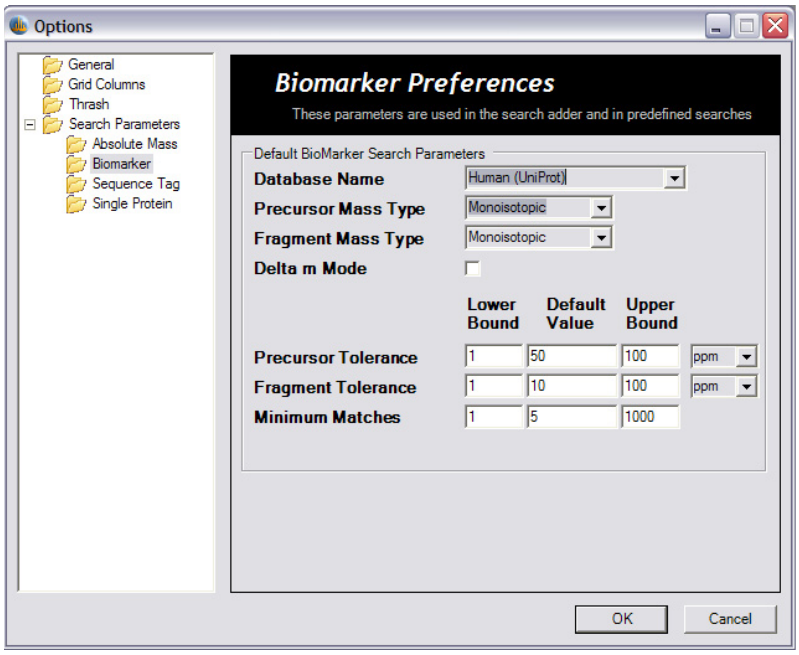

Figure 28. Biomarker Preferences dialog box

Table 9 describes the preferences in the Biomarker tab that must have numerical input.

Table 9. Biomarker Preferences

Preference	Description
Precursor Tolerance	Lower bound sets the minimum value for an Precursor Tolerance that does not trigger an "out of range" warning. Default value sets the default value for an Precursor Tolerance. Upper bound sets the maximum value for an Precursor Tolerance that does not trigger an "out of range" warning. Displayed in Da or ppm.
Fragment Tolerance	Lower bound sets the minimum value for a Fragment Tolerance that does not trigger an "out of range" warning. Default value sets the default value for an Fragment Tolerance. Upper bound sets the maximum value for an Fragment Tolerance that does not trigger an "out of range" warning. Displayed in Da or ppm.
Minimum Matches	Lower bound sets the minimum value for Minimum Matches that does not trigger an "out of range" warning. Default value sets the default value for Minimum Matches. Upper bound sets the maximum value for Minimum Matches that does not trigger an "out of range" warning.

Note Gene-Restricted Biomarker searches draw their default parameters from Biomarker searches.

Sequence Tag Preferences

Use the Sequence Tag Preferences window to set the default value used when adding new Sequence Tag searches. Click the Sequence Tag Preferences tab and the dialog box shown in Figure 29 appears.

You can set the following options from the Sequence Tag Preferences dialog box:

- Database Name
- Fragment Mass Type

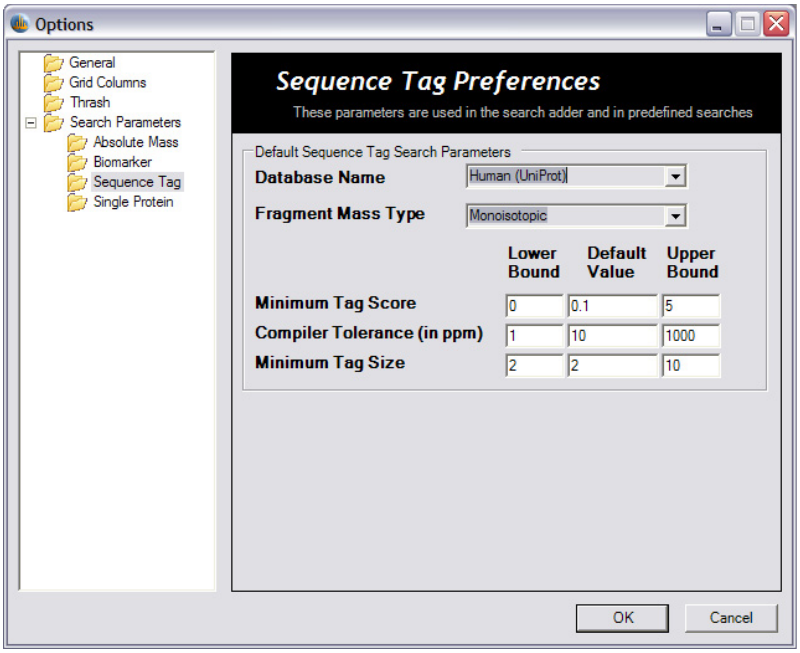

Figure 29. Sequence Tag Preferences dialog box

Table 10 describes the preferences in the Biomarker tab that must have numerical input.

Table 10. Sequence Tag Preferences

Preference	Description
Minimum Tag Score	Lower bound sets the minimum value for a Minimum Tag Score that does not trigger an "out of range" warning. Default value sets the default value for a Minimum Tag Score. Upper bound sets the maximum value for a Minimum Tag Score that does not trigger an "out of range" warning.
Compiler Tolerance	Lower bound sets the minimum value for a Compiler Tolerance that does not trigger an "out of range" warning. Default value sets the default value for a Compiler Tolerance. Upper bound sets the maximum value for a Compiler Tolerance that does not trigger an "out of range" warning.
Minimum Tag Size	Lower bound sets the minimum value for a Minimum Tag Size that does not trigger an "out of range" warning. Default value sets the default value for a Minimum Tag Size. Upper bound sets the maximum value for a Minimum Tag Size that does not trigger an "out of range" warning.

Single Protein Preferences

Use the Single Protein Preferences window to set the default value used when adding new Single Protein searches. Click the Single Protein tab and the dialog box shown in Figure 30 appears.

Set the following options from the Single Protein Preferences dialog box:

- Fragment Mass Type
- Δm Mode

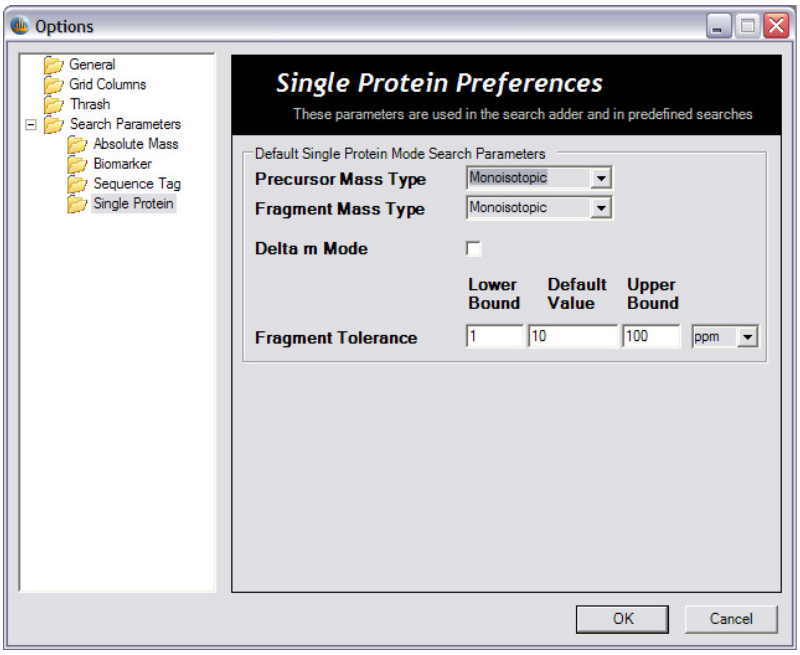

Figure 30. Single Protein Preferences dialog box

Table 11 describes the preferences in the Biomarker tab that must have numerical input.

Table 11. Single Protein Preferences

Preference	Description
Fragment Tolerance	Lower bound sets the minimum value for fragment tolerance that does not trigger an "out of range" warning. Default value sets the default value for an fragment tolerance. Upper bound sets the maximum value for an fragment tolerance that does not trigger an "out of range" warning. Displayed in Da or ppm

Chapter 5 Importing MS/MS Data

In order for ProSightPC to identify and characterize proteins, mass spectral data must be converted to neutral mass values. ProSightPC supports different ways of entering MS and MS/MS data into the experiment.

This chapter describes how to import MS/MS data into ProSightPC. It contains the following sections:

Analysis to Infer Mass

An analysis to infer mass (AIM) is an operation in which high resolution mass spectra data from proteins or large peptides is converted into neutral monoisotopic or average masses.

ProSightPC and XCalibur provide two methods for AIM.

- THRASH (in ProSightPC)
- XTRACT (part of the Qual Browser tools)

IMPORTANT ProSightPC works with neutral masses only.

Selecting an AIM

The relative advantages of different AIMs are beyond the scope of this manual. For more information, see the XTRACT Manual or Horn et al. (2000).[1]

[1]David M. Horn, Roman A Zubarev, and Fred W. McLafferty, (2000) Automated Reduction and Interpretation of High Resolution Electrospray Mass Spectra of Large Molecules. Journal of the American Society of Mass Spectrometry 11, 320-332.

Entering Mass Values

There are three possible methods for importing MS/MS data into a ProSightPC MS/MS experiment:

- THRASH Method
- XTRACT Method
- Manual Entry Method

Table 12 contains a summary of the differences between the THRASH and XTRACT methods of importing mass values.

Table 12. Comparison of XTRACT and THRASH methods

AIM Equivalency	Data Type	Speed	Peaks A	S/N (approximate)
XTRACT	Profile, centroid	Faster	Slightly more	3:1
THRASH	Profile	Slower	Slightly less	10:1

THRASH Method

The THRASH method uses a modified version of the algorithm described by Horn et al (2000) to automatically infer the observed mass values from the spectra in a RAW file. For this analysis to work, the source RAW file must contain data gathered in profile mode and all summed scans must be in a numerically continuous block.

Note This analysis is sometimes referred to as THRASHing.

THRASHing a RAW File

THRASHing a RAW file is accomplished from within the ProSightPC interface.

To THRASH a RAW file

1. Choose **File > Import RAW > Profile > Build Experiment** from the RAW Data dialog box shown in Figure 31.
2. Enter the RAW file to be THRASHed. Type the full path name or click **Browse** to select the file name from the list.
3. Type the desired parameters and select one or more predefined searches to append to the new experiment. Table 13 lists the THRASH parameters.
4. Click **OK**.

A progress bar indicates the progress of the THRASH process.

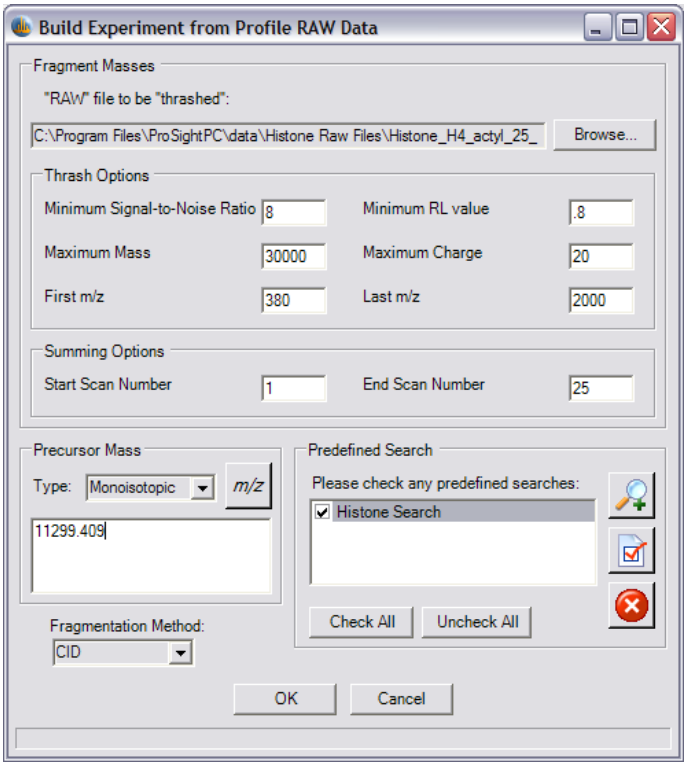

Figure 31. Build Experiment from Profile RAW Data dialog box

The THRASH options are described in Table 13.

Table 13. THRASH Options

Option	Description
Minimum Signal-to-Noise Ratio	As per Hort et al (2000). Values less than 5:1 significantly slow down the analysis, but can result in a greater number of identified isotopic envelopes.
Minimum RL Value	As per Hort et al (2000).
Maximum Mass	Cutoff point for THRASH when searching for masses.
Maximum Charge	Highest charge state THRASH will consider.
Start *m/z*	Lowest *m/z* value considered.
End *m/z*	Highest *m/z* value considered.
Start Scan Number	First scan number scanned.
End Scan Number	Last scan number scanned.
Precursor Type	Precursor mass type is either monoisotopic or average mass.

Table 13. THRASH Options, continued

Option	Description
Precursor Mass	Type the precursor mass, or click *m/z* to use the Precursor Mass Calculator to calculate the neutral mass from an *m/z* value and charge state. If needed, several values can be entered, in Da.
Fragmentation Method	Method used: either CID, ECD, or IRMPD.
Predefined Search	(Optional) Predefined searches can be added to the new MS/MS experiment.

THRASH Defaults

The THRASH default preferences can be changed by using the ProSightPC Tools dialog box.

To change the THRASH default preferences

- Choose **Tools > Options > THRASH**. A new window appears.

For more information about changing THRASH defaults, see Figure 26 on page 37.

XTRACT Method

Use the XTRACT method (one of the Qual Browser tools) to infer neutral mass values. First, use XTRACT to create a RAW file containing the neutral monoisotopic mass values, then import MS/MS experiments from this RAW file.

Importing an XTRACTed RAW file

To XTRACT a single scan from a RAW file

1. Open a RAW file in the Qual Browser.
2. Right-click on the MS pane. A pop-up menu appears.
3. Choose **Export > XTRACT RAW file**, choose the parameters, and click **Xtract**.

   The resulting XTRACTed file appears in an additional window.

To save a monoisotopic mass post-XTRACTed file

- Select the appropriate mass filter (in Ranges) and choose **Export > Write to RAW file**.

This new RAW file can be imported into ProSightPC.

To XTRACT a multiple scan from a RAW file

1. Open a RAW file in the Qual Browser.
2. Right-click on the MS pane. A pop-up menu appears.
3. Right-click on the LC pane, choose **Export > XTRACT RAW file**, choose the parameters, and click **Xtract**.

   The XTRACTed scans appear in an additional window. The scans can be saved individually and imported into ProSightPC.

Manual Entry Method

A third option for importing MS/MS experiment data is to manually enter the data.

Manually Entering Data

Enter MS/MS experiment data from within ProSightPC through the Tools menu.

To manually enter the data

- Choose **Tools > Experiment Adder**.

  The Experiment Adder window appears, as Figure 32 shows.

- Select from the options and parameters listed in Table 14.
- Click **Create**.

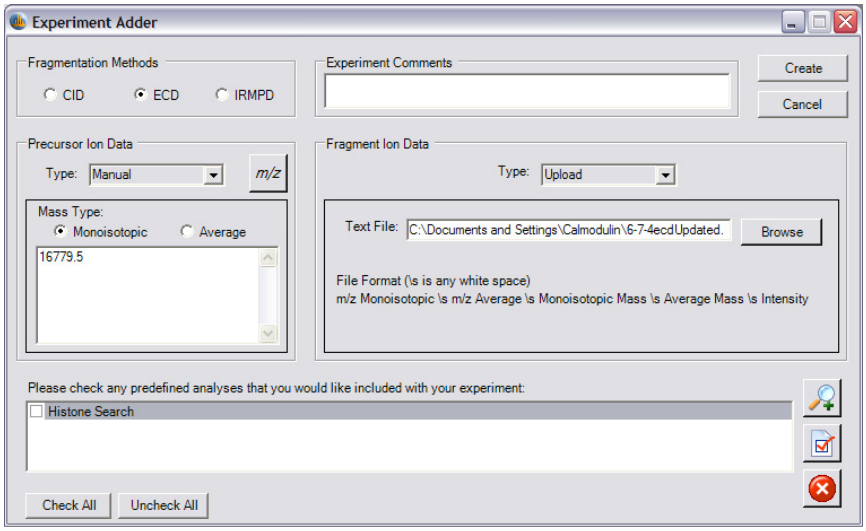

Figure 32. Experiment Adder dialog box

Manual entry parameters or options are described in Table 14.

Table 14. Manual Data Entry Parameters and Options

Option or Parameter	Description
Fragmentation Method	Select either CID, ECD, or IRMPD.
Precursor Ion Data	Select Manual or Upload. When Manual is selected, type one or more monoisotopic or average masses or use the Precursor Mass Calculator. When inputting fragment ion data using the Upload option, the ASCII text files must be properly formatted.
Fragment Ion Data	Select Manual or Upload. When Manual is selected, type one or more monoisotopic or average masses. When inputting fragment ion data using the Upload option, the ASCII text files must be properly formatted
Predefined Search	(Optional) Select any predefined searches.

Note

1. When inputting fragment ion density data manually, you can leave Intensity Values empty. In this case, the default intensity of 1 is assigned to each fragment ion.

2. ASCII text files must be formatted with five white-space-separated columns of numbers. Each row represents a separate ion mass. The columns must be arranged as:

   - Monoisotopic *m/z*
   - Average *m/z*
   - Monoisotopic Mass
   - Average Mass
   - Intensity

Chapter 6 Search Modes

ProSightPC has four search modes. Each search mode represents a specific mechanism used to query a proteome database within the PTM Warehouse.

This chapter describes how to use each search mode. It contains the following sections:

For more information about single protein searches, see Chapter 7.

About Search Modes

Each search mode represents a specific method used to query a proteome database within the PTM Warehouse.

ProSightPC has four basic search modes:

- Absolute mass
- Biomarker
- Sequence tag
- Single protein search
  For more information, see Chapter 7, "Sequence Gazer and Single Protein Searches."

Use ProSightPC to perform the following types of advanced searches:

- Gene-restricted absolute mass
- Gene-restricted biomarker
- MS^n hybrid searches

To query MS/MS data against a a proteome database, add a search to the MS/MS experiment containing the MS/MS data.

Adding A Search

Add a single search to a specific MS/MS experiment when an experiment requires adjusting parameters to identify the protein.

Add a search to an experiment by using the Experiment tools.

To add a search to an experiment

1. Open the Data Manager of the experiment.
2. Choose **Experiment Tools > Add Search**. The New Search in Experiment "X" dialog box appears, as Figure 33 shows. The X of the dialog box reflects the proper experiment number.

   Alternatively, right-click any search in the experiment listed on the Data Grid and select **Add Search**.
3. Select a search type, database name, and parameters.
4. Click **Save**.

You can also execute the search from the Data Manager by clicking **Run Search.**

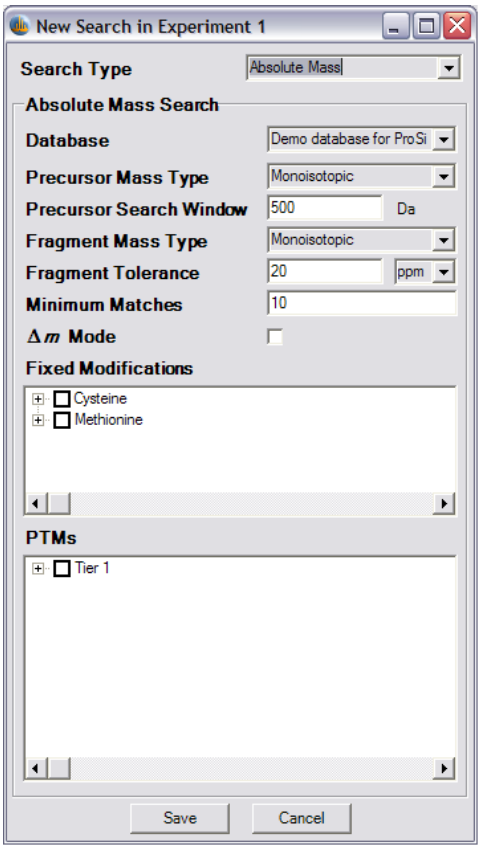

Figure 33. Adding a New Search dialog box

Absolute Mass Search

The absolute mass search is the defining search mode for top-down proteomics. Absolute mass searches use the precursor mass to restrict the portion of the proteome database to query.

For each protein form with a theoretical precursor mass within a window of the observed precursor ions mass, the absolute mass search compares all theoretical fragments and masses to the observed fragment ion masses. The number of observed fragment ions matching to within the fragment tolerance is determined and used to score the identification.

Although each protein form with a theoretical precursor mass within the window is interrogated, only those protein forms meeting user-defined criteria for number of fragment ions are displayed (see Table 5 on page 33 and Table 6 on page 34).

Parameters

Absolute mass searches include the parameters listed in Table 15.

Table 15. Absolute Mass Search Parameters

Parameter	Description
Database	Proteome database against which the search is being run.
Precursor Type	Defines whether monoisotopic or average precursor ions mass is used.
Precursor Search Window	Defines a range around the observed precursor mass. All protein forms with a theoretical mass within this range will be interrogated.
Fragment Type	Defines whether the search uses monoisotopic or average mass of the fragment ions.
Fragment Tolerance	Defines whether fragment tolerance is expressed as absolute, measured in Da; or relative, measured in ppm. An observed fragment ion matches a theoretical fragment ion if the observed fragment ion mass is within plus or minus the fragment tolerance of the theoretical fragment ion mass.
Minimum Matches	Defines the minimum number of matching fragment ions that must be present before a protein form will be reported. This parameter acts as a filter of returned protein forms.
Δm Mode	Defines whether the search is conducted in Δm Mode.
Fixed Modifications	Enables the selection of no more than one fixed modification per amino acid type.
PTMs	Displays PTMs arranged in one or more tiers based upon the selected proteome database. Only theoretical protein forms containing exclusively selected PTMs will be interrogated. Every form containing a nonselected PTM will be excluded from the interrogation.

Advice on Use

Use the following strategies when running absolute mass searches.

Precursor Search Window Size

A 1000 Da Precursor Search Window search is often used as the first search for an unknown protein. If there are few modifications on the unknown protein not in the PTM Warehouse, a 1000 Da intact search will frequently identify, but not characterize, the protein.

If a protein has a large number of ions matching one terminal, that is evidence of a protein's identity and further characterization in the Sequence Gazer® is usually warranted.

Δm Mode

Use the Δm mode to locate unknown modifications near either terminus.

If the 1000 Da absolute mass search fails to identify a protein, consider running another absolute mass search with a 1000 Da precursor search window in Δm mode. By activating the Δm mode, the search is more likely to identify proteins with unknown modifications but will take approximately three times longer then the corresponding absolute mass search.

Sequence Tag Search

The sequence tag search as a two-step process to identify, but not characterize, proteins. The two steps are compilation and search.

Compilation

Compilation (also known as *De Novo* sequencing) orders the mass list from largest to smallest, looking for sets or ladders of mass differences exactly equal to the mass of an amino acid, within the compiler tolerance you defined. The compiler tolerance is always given in ppm.

Compilation returns only those sequence tags equal to or longer than the defined minimum tag size.

Note Multiple independent sequence tags are frequently found with ECD data.

Search

The sequence tag list is queried against every base sequence within the proteome database for the presence of any of the sequence tags. Any sequence found containing one or more of the sequence tags will be scored. Any sequence scoring above the defined minimum tag score will be reported.

The sequence tag search automatically searches both the forward and reverse direction of every sequence tag.

Parameters

Sequence tag searches require the parameters defined in Table 16.

Table 16. Sequence Tag Search Parameters

Parameter	Description
Database	The proteome database against which the search is being run.
Minimum Tag Score	The lowest acceptable sequence tag score reported as a match.
Compiler Tolerance	The permissible error, measured in ppm, between two fragment ion masses that will still be considered matching an amino acid.
Minimum Tag Size	The shortest length of a sequence tag ladder to be reported by the sequence tag compiler. Always an integer value equal to or greater than 2.
Fragment Type	Declares whether the sequence tag compiler uses monoisotopic or average mass.
Fixed Modifications	A chemical modification present on all instances of a given type of amino acid in the observed protein. Enables the selection of no more than one fixed modification per amino acid type.

Table 16. Sequence Tag Search Parameters, continued

Parameter	Description
Manually Enter Tags	Instead of compiling sequence tags, you can manually enter sequence tags. Each sequence tag consists of the single letter designation of the amino acid separated by a space. Isobaric amino acids can be entered as a Pipe '|' separated list enclosed in square brackets, for example: [I|L]. Multiple sequence tags can be entered, properly fomatted, with one Sequence Tag per line in the space provided. The sequence tag search will automatically search for the entered sequence tag and its reverse.

Advice on Use

The following suggestions are recommended when conducting sequence tag searches.

- If absolute mass or biomarker searches fail to identify the protein in the presence of rich fragmentation data, a sequence tag search can frequently identify, but not characterize, the protein.
- The output of a sequence tag search can be entered in series into a gene-restricted search to perform a hybrid search. This frequently identifies and characterizes a protein.
- A sequence tag search is frequently the first step in MS^n experiments.
- Unresolved amino acid pairs, such as isolucine and lucine, should be manually entered as a pipe-separated list in square brackets; for example, [I|L] with no spaces.

Biomarker Search

In typical top-down experimental preparation, not all polypeptides are intact proteins. Use a Biomarker Search to identify those proteins that are a product of biological degradation and cannot be logically predicted. A Biomarker Search compares the observed precursor mass to all possible subsequences of a particular database within a defined tolerance (for example, less than 10ppm). Those subsequences that fall within the defined tolerance are theoretically fragmented and the observed fragment ions are compared.

A biomarker search is a two-step process which is repeated for each base protein sequence in the proteome database. The steps are:

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

2. Calculate all possible theoretical fragment ions for the candidate subsequence then compare the theoretical fragment ion masses against the observed fragment ion masses.

In a biomarker search, the precursor search tolerance is an estimation of measurement error on the observed precursor mass. The value is usually small compared to an absolute mass precursor search window.

Parameters

Biomarker searches require the parameters described in Table 17.

Table 17. Biomarker Search Parameters

Parameter	Description
Database	Defines the proteome database the search is being run against.
Precursor Type	Defines whether monoisotopic or average precursor ions will be used.
Precursor Search Tolerance	Defines precursor tolerance as either absolute, measured in Da; or relative, measured in ppm. A protein subsequence matches the observed precursor mass if the protein subsequences theoretical mass is within plus or minus the Precursor Search Tolerance of the observed precursor mass.
Fragment Type	Defines whether the search uses monoisotopic or average mass of the fragment ions.
Fragment Tolerance	Defines fragment tolerance as either absolute, measured in Da; or relative, measured in ppm. An observed fragment ion matches a theoretical fragment ion if the observed fragment ion mass is within plus or minus the fragment tolerance of the theoretical fragment ion mass.

Table 17. Biomarker Search Parameters, continued

Parameter	Description
Minimum Matches	Defines the minimum number of matching fragment ions that must be present before a protein form will be reported. This parameter acts as a filter of returned protein forms.
Δm Mode	Defines whether the search is conducted in Δm mode.
Fixed Modifications	Enables the selection of no more than one fixed modification per amino acid type. A fixed modification is a chemical modification present on all instances of a given type of amino acid in the observed protein.

Advice on Use

Thermo recommends the following while conducting biomarker searches:

- If an absolute mass analysis fails to identify a protein, use a biomarker search.
- Increasing the precursor tolerance results in longer run times.
- Biomarker searches are well suited for identifying biologically relevant proteolytic products.
- Proteins or peptides containing non-reduced disulfide bonds can be identified by setting the precursor search tolerance to 2.5 Da and running the search in Δm mode.

Predefined Searches

Predefined searches represent a strategy to simplify repeating identical searches on different sets of MS/MS data. This is accomplished by assigning a name to a set of parameters, which can then be appended to any experiment.

Note The search parameters within predefined searches are persistent until modified or deleted.

Tip Use a predefined search to set up defaults for frequently run searches. Instead of using Add Search to manually append the same search to multiple experiments, create a predefined search and append it to your experiments as needed.

Predefined Search Manager

Use the Predefined Search Manager to create, edit, or delete predefined searches.

Accessing Predefined Search Manager

Access the Predefined Search Manager through the Tools menu of ProSightPC.

To access the Predefined Search Manager

- Choose **Tools > Manage Predefined Searches** to open the Predefined Search Manager dialog box, as Figure 34 shows.

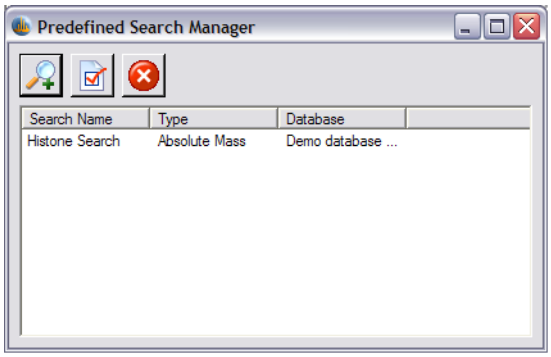

Figure 34. Predefined Search Manager dialog box

Navigating Predefined Search Manager

All predefined searches already created appear in the Predefined Search Manager grid.

Table 18 describes the options in the Predefined Search Manager dialog box.

Table 18. Predefined Search Manager

Item	Description
Search Name	Displays the name attributed to the Predefined Search.
Type	Displays the Search Mode attributed to the Predefined Search.
Database	Displays the Database the Search is run against.

The Predefined Search Manager has three primary functions:

- Create a New Predefined Search
- Edit a Predefined Search
- Remove a Predefined Search

Create a New Predefined Search

To create a new predefined search

1. Select **Create**. The New Predefined Search dialog box appears, as Figure 35 shows.
2. Type a **Name** for the New Predefined Search in the Search Name field.
3. Select **Search Type** from the pull-down menu.
4. Select parameters, based on search type, from the pull-down menu.
5. Click **Save**.

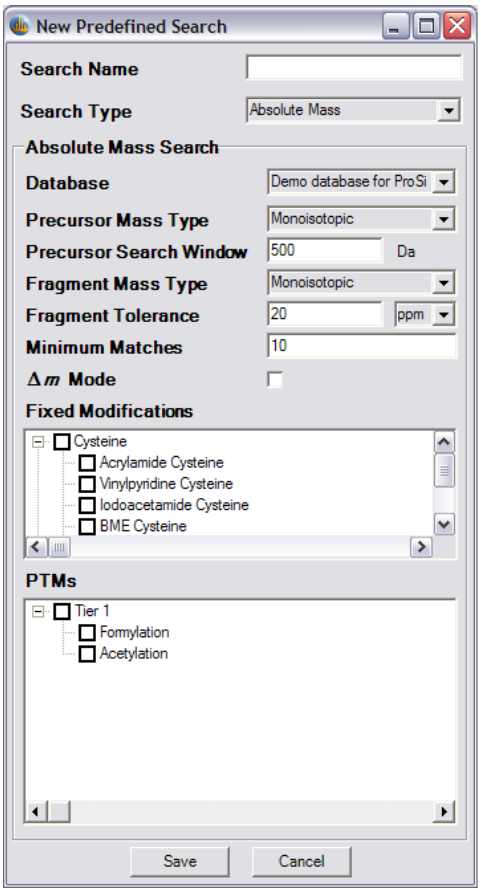

Figure 35. New Predefined Search dialog box

Edit a Predefined Search

To edit a predefined search

1. Select a predefined search name from the list in the predefined search manager Data Grid.
2. Click **Edit.** The Edit Predefined Search dialog box appears, as Figure 36 shows.
3. Edit the parameters as desired.
4. Click **Save**.

Note Altering search parameters has no effect on searches already added to MS/MS experiments.

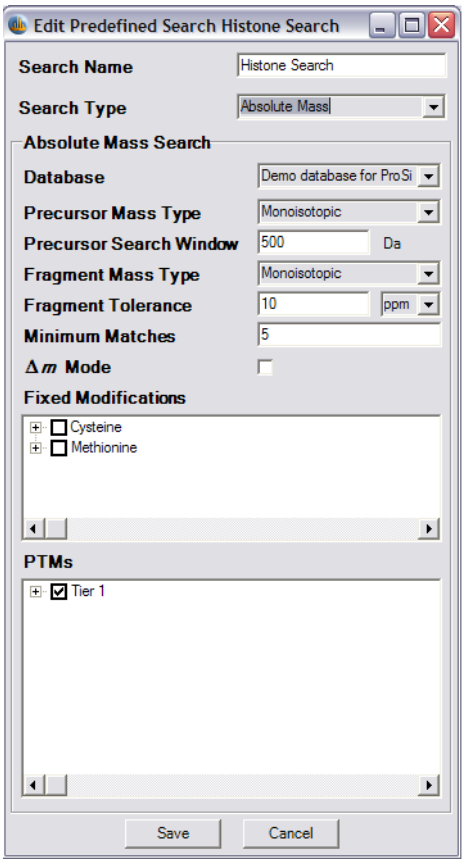

Figure 36. Edit Predefined Search dialog box

Remove Predefined Search

To remove a predefined search

1. Select a predefined search name from the list in the predefined search manager Data Grid.
2. Click **Remove**.
3. Confirm Remove by clicking **Yes** or **No**.

Note Click No to return to the Predefined Search Manager without removing the selected Predefined Search.

Add A Predefined Search

To add a predefined search to an experiment

1. Open the Data Manager of the experiment.

2. Choose **Experiment Tools > Append Predefined Search**. The Predefined Search dialog box appears, as Figure 37 shows.

   Alternatively, right-click any search in the experiment listed on the Data Grid and select **Add Search**.

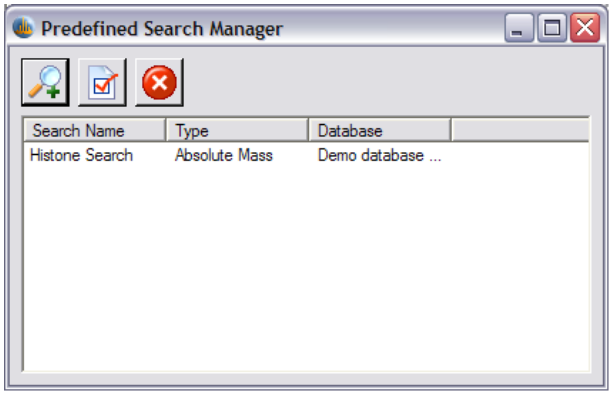

Figure 37. Predefined Search Manager dialog box

3. Select a predefined search to append to the experiment and click **Append**.

   You can also execute the search from the Data Manager by clicking **Run Search**.

Tip Choose **Tools > Batch Run** to process several predefined searches automatically.

Advanced Searches

Gene-restricted searches look at all protein forms of explicitly listed genes. They can only be made from the results, or hit list, of a previously completed absolute mass, biomarker, or sequence tag search. Gene-restricted searches are most often performed with a sequence tag search to form a hybrid search.

ProSightPC automatically generates a gene ID list from the results of a previous search.

Gene-restricted searches consists of two functionally different, but closely related, types of searches:

- Gene-Restricted Absolute Mass Search
- Gene-Restricted Biomarker Mass Search

Gene-Restricted Absolute Mass Search

Use a gene-restricted absolute mass (GRAM) search to perform an absolute mass search on every protein form of each gene in the gene list, regardless of theoretical precursor mass. Only those protein forms meeting the defined minimum matches parameter will be reported.

Parameters

Gene-restricted absolute mass searches require the parameters described in Table 19.

Table 19. Gene-Restricted Absolute Mass Search Parameters

Parameter	Description
Database	Set automatically to the proteome database of the parent search.
Fragment Type	Declares whether the search uses monoisotopic or average mass of the fragment ions.
Fragment Tolerance	Defines whether fragment tolerance is expressed as absolute, measured in Da; or relative, measured in ppm. An observed fragment ion matches a theoretical fragment ion if the observed fragment ion mass is within plus or minus the fragment tolerance of the theoretical fragment ion mass.
Minimum Matches	Defines the minimum number of matching fragment ions that must be present before a protein form will be reported. Acts as a filter for which protein forms are returned.
Δm Mode	Defines whether the search is conducted in Δm mode.
Fixed Modifications	A chemical modification present on all instances of a given type of amino acid in the observed protein. Enables the selection of no more than one fixed modification per amino acid type.

Table 19. Gene-Restricted Absolute Mass Search Parameters, continued

Parameter	Description
PTMs	Displays PTMs arranged in one or more tiers based upon the proteome database. Only theoretical protein forms containing zero or more of the selected PTMs will be interrogated. Every form containing a nonselected PTM will be excluded.

Gene-Restricted Biomarker Mass Search

Use a gene-restricted biomarker search (GRBM) to perform a biomarker search on all protein forms of the genes listed in the gene ID list.

Because the query is limited to the gene IDs listed, a gene-restricted biomarker mass search runs substantially quicker than a simple biomarker search.

Parameters

Gene-restricted biomarker mass searches require the parameters described in Table 20.

Table 20. Gene-restricted Biomarker Mass Search Parameters

Parameter	Description
Database	Set to the proteome database of the parent search automatically.
Precursor Type	Defines whether monoisotopic or average precursor ions are used.
Precursor Search Tolerance	Defines whether precursor tolerance is expressed as absolute, measured in Da; or relative, measured in ppm. A protein subsequence matches the observed precursor mass if the protein subsequence's theoretical mass is within plus-or-minus the precursor search tolerance of the observed precursor mass.
Fragment Type	Defines whether the search uses monoisotopic or average mass of the fragment ions.
Fragment Tolerance	Defines whether fragment tolerance is expressed as either absolute, measured in Da; or relative, measured in ppm. An observed fragment ion matches a theoretical fragment ion if the observed fragment ion mass is within plus-or-minus the fragment tolerance of the theoretical fragment ion mass.
Minimum Matches	Defines the minimum number of matching fragment ions that must be present before a protein form will be reported. Acts as a filter for which protein forms are returned.
Δm Mode	Defines whether the search is conducted in Δm mode.
Fixed Modifications	A chemical modification present on all instances of a given type of amino acid in the observed protein. Enables the selection of no more than one fixed modification per amino acid type.

MS/MS Experiments

ProSightPC is built on the concept of the MS/MS experiment. Analyzing an MS^{3+} experiment requires the following steps:

1. An MS/MS experiment is constructed with the MS^2 data as precursor masses and the MS^3 data as fragment ion masses. This MS/MS experiment is used to run a sequence tag search.

2. Next an MS/MS experiment is created using the MS data as precursor ions and the MS^2 data as fragment ions.

3. A sequence tag search is added to the experiment from step 2, and the sequence tags from the first experiment are manually entered. When this search is run, a gene list will be created for all proteins containing sequences consistent with the MS^3 fragmentation data.

4. A gene-restricted absolute mass search is then added to the second experiment and used to identify and characterize the observed protein.

For additional information on MS^n of top-down proteomics data, see Zabrouskov (2005).[1]

Δ*m* Mode Demystified

The Δ*m* mode is a technique for identifying protein forms containing unknown PTMs.

The Δ*m* is the difference between the observed precursor mass and the theoretical precursor mass. When a search is run in Δ*m* mode, three interrogations per sequence are performed concurrently:

- The theoretical fragment ion masses of the protein sequence are compared to the observed fragment ion list as usual.
- The theoretical fragment ion masses derived from the sequence and the Δ*m* applied to the N-terminal are compared to the observed fragment ion mass list.
- The theoretical fragment ion masses derived from the sequence and the Δ*m* applied to the C-terminal are compared to the observed fragment mass list.

Note A Δ*m* mode search takes approximately three times longer than the same search without Δ*m* mode.

[1]Zabrouskov, V., M. W. Senko, Y. Du, R. D. LeDuc and N. L. Kelleher (2005) New and Automated MSn Approaches for Top-Down Identification of Modified Proteins. Journal of the Americal Society of Mass Spectrometry 16:12, 2027-2038.

Interpreting the Output

By carefully observing the pattern of fragments with and without Δ*m*, you can frequently locate the Δ*m*.

For example, if a particular result returned with the two smallest N-terminal fragments matching without the Δ*m,* but all other matching N-terminal fragments contain the Δ*m*, the unknown PTM must be on an amino acid between the second and third N-terminal fragment.

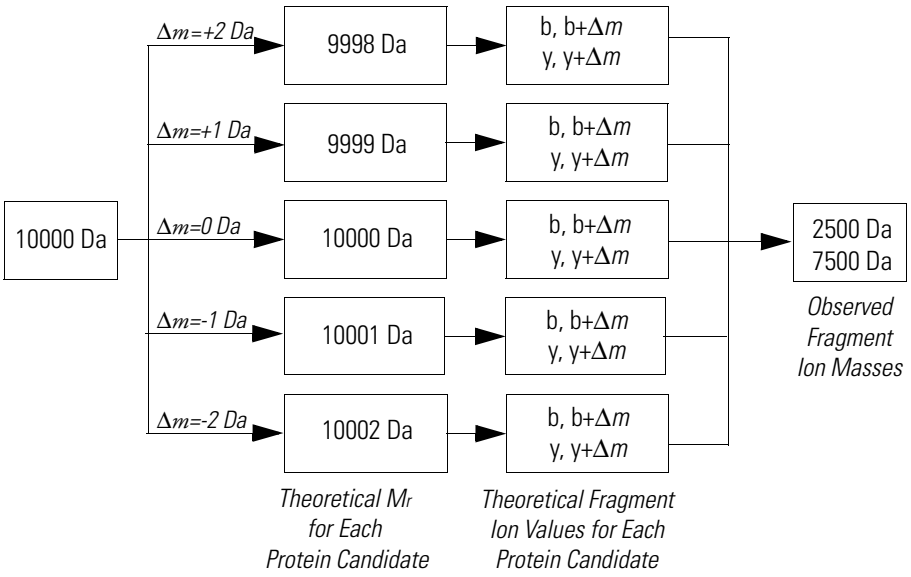

Figure 38. Schematic of Δ*m* Mode

In Figure 38, the observed precursor mass is first checked against the theoretical precursor masses of every protein within the specified precursor mass window and the mass difference (Δ*m)* is calculated and stored. Next, the theoretical fragment ion list for each protein is doubled; for each b/y or c/z**.** ion. Both the original fragment mass and the modified fragment mass (plus the Δm) are checked against the observed fragment ion mass list. The result is that any observed fragment ions having the same mass shift as the precursor protein will be returned as positive matches.

Chapter 7 Sequence Gazer and Single Protein Searches

The Sequence Gazer is an interactive environment for interrogating MS/MS data.

This chapter describes how to use the Sequence Gazer. It contains the following sections:

About Sequence Gazer

The Sequence Gazer is an interactive environment for interrogating MS/MS data against a known protein sequence. Use the Sequence Gazer to characterize identified proteins by selectively adding or removing PTMs or custom masses to amino acids within a protein sequence. Once you have made all your modifications to the amino acids, the ion data can be rescored.

The Sequence Gazer is usually used for one of two purposes. First, MS/MS data might have been gathered on a known, and pure, protein containing one or more unknown PTMs. In this case, a single protein mode search is built and added to the MS/MS experiment and the Sequence Gazer is used to test hypotheses of which PTMs are present.

Alternatively, the result of any other search mode might identify, and possibly partially characterize a protein, but full characterization does not occur. In this case, the Sequence Gazer can be used to fully characterize the protein.

Accessing Sequence Gazer

Access Sequence Gazer through one of two strategies:

- By running a single protein search
- By clicking **Take to Sequence Gazer** from any protein identification of a completed search

To run a single protein search

1. Open the desired experiment in a Data Manager.
2. Choose **Experiment Tools > Add Search**. The New Search dialog box appears, as Figure 39 shows.
3. Select **Search Type > Single Protein**.
4. Select Precursor Mass Type as monoisotopic or average mass.
5. Select Fragment Type as monoisotopic or average mass.
6. Type the Fragment Tolerance, in Da or ppm.
7. Select Δm mode, if desired.
8. Select Fixed Modifications, if desired.
9. Enter the sequence in the field by either typing it or using a sequence from another source.
10. Click **Save.**

> **Note** After step 10, a new search is added to the Data Manager. It appears in blue highlighted text to indicate the new search is pending and has not yet been run.

11. From the Data Manager, click **Run Search**.

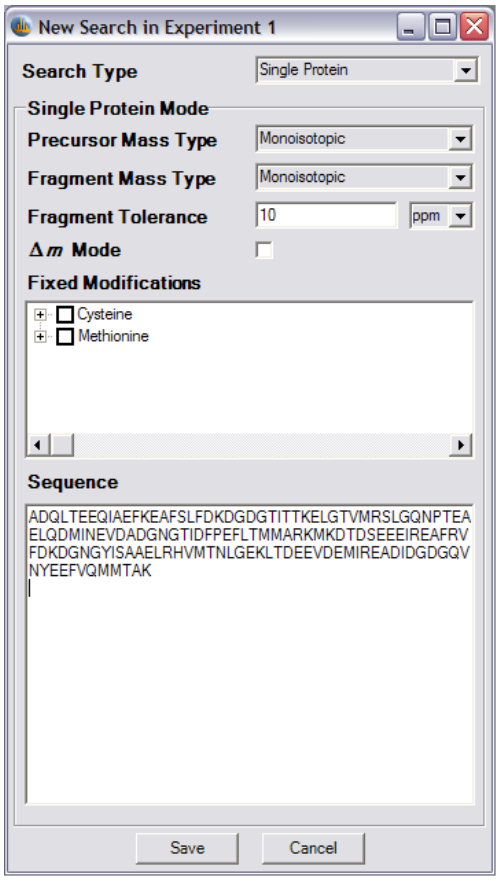

Figure 39. Add Single Protein Search dialog box

To access Sequence Gazer from any protein identified in a completed search

1. Select the desired search and click its corresponding side arrow.
2. Locate the desired protein identification within the search results and click its corresponding side arrow.
3. Click **Take to Sequence Gazer** (shown with the oval in Figure 40).

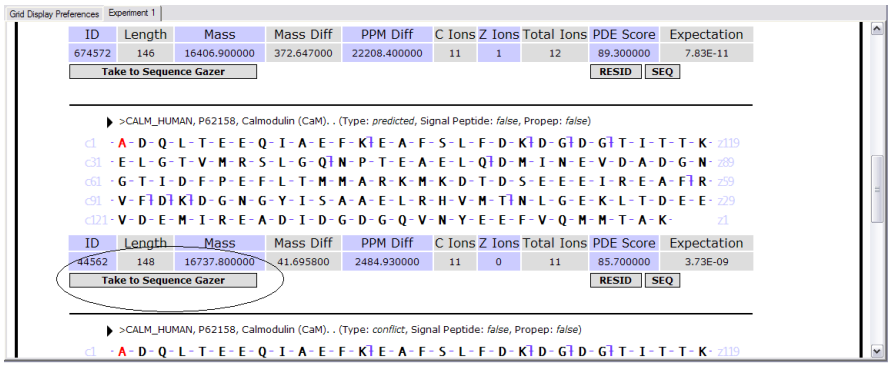

Figure 40. Take To Sequence Gazer Button

Navigating Sequence Gazer

Fragment ion information within the Sequence Gazer interface is organized into the following regions:

- Search Parameter Display
- Scores Box
- Fragments Explained Box
- Mass Diagram
- Interactive Fragment Map
- Amino Acid Information Box
- Fixed Modifications Box
- Matching Fragment Table
- Nonmatching Fragments Table

These regions are illustrated in Figure 41.

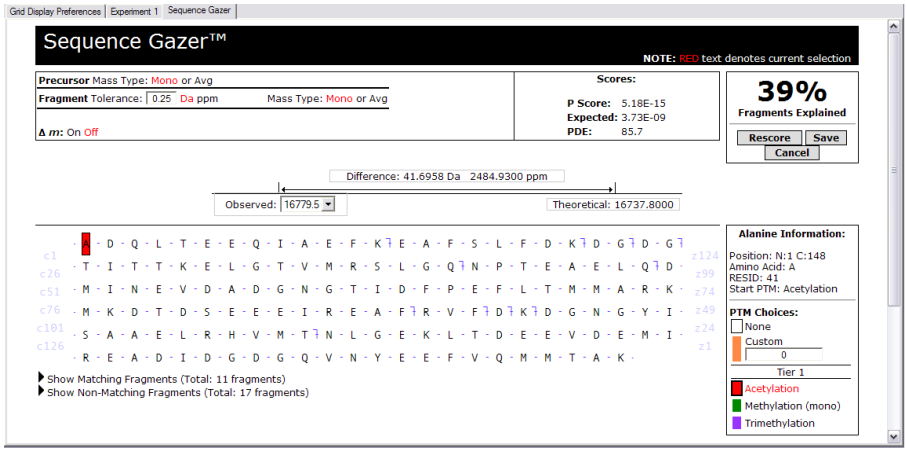

Figure 41. Sequence Gazer dialog box

Search Parameter Display

Use the search parameter display to display the data options and tolerances you selected during the last round of scoring, as Table 21 shows.

Table 21. Search Parameters

Display	Function
Fragment Tolerance	Displays the Fragment Tolerance at the time the search was last scored.
Precursor Mass Type	Displays Mass Type as monoisotopic or average mass
Fragment Mass Type	Displays Mass Type as monoisotopic or average mass
Δm Mode	Displays whether Δm Mode has been selected.

Note

1. User-defined selections appear in red. These can be changed by clicking on a new selection. The new selection appears in red.
2. None of the changes made in the Search Parameter Display are implemented until you click **Rescore**.

Scores Box

Three scores are computed. These scores are explained in Table 22.

Table 22. Scores

Score	Source
P_Score	Lists the P_Score as noted by Meng, et al.*
Expectation	Lists the Expected Score as noted by LeDuc, et al.**
PDE	Lists the PDE Score as noted by Reid, et al.***

*Meng, F. B. J. Cargile, L. H. Miller, A. J. Forbes, J. R. Johnson and N. L. Kelleher. (2001) Informatics and Multiplexing of intact protein identification in bacteria and the archaea. Nature Biotechnology 19: 952-957.

**LeDuc, R. D., G. K. Taylor, Y. B. Kim, T. E. Januszyk, L. H. Bynum, J. V. Sola, J. S. Garavelli and N. L. Kelleher. (2004) ProSIght PTM: an integrated environment for protein identification and characterization by top-down mass spectrometry. Nucleic Acids Research 32: W340-W345.

***Reid, G. E., H. Shang, J. M. Hogan, G. U. Lee and S. A. McLuckey. (2002) Gas-phase concentration, purification, and identification of whole proteins from complex mixtures. Journal of the American Chemical Society 124: 7353-7362.

Fragments Explained Box

The fragments explained box displays a percentage representing the number of matching fragments divided by the total number of fragments.

The fragments explained box has three additional controls described in Table 23.

Table 23. Fragments Explained Box

Control	Description
Rescore	Recalculates all scores and matching fragment information.
Save	Adds a new completed single protein mode search to the experiment.
Cancel	Discards the changes the user has made and returns to the Data Manager.

Mass Diagram

The mass diagram displays the difference between the observed and theoretical mass, expressed in Da and ppm. The functions of these two elements are described in Table 24.

Table 24. Mass Diagram

Item	Function
Observed Precursor Mass	Displays the experimental precursor mass.
Theoretical Precursor Mass	Displays the theoretical Precursor Mass, including all user-input changes, as of the last Score.

Tip The drop-down menu displays the observed precursor mass used in scoring. When toggled, you can switch between the precursor masses.

Interactive Fragment Map

The interactive fragment map is an interactive display of the protein sequence, along with any PTMs, and the matching fragment information.

A black box around an amino acid indicates the amino acid selected. Choosing a PTM from the amino acid box affixes that PTM to the selected amino acid. A colored background behind an amino acid indicates the matching PTM is currently assigned to that amino acid.

Note Click on an amino acid to access to all available PTMs contained in the RESID database that can be applied to that amino acid.

Amino Acid Information Box

The Amino Acid Information box refers to the selected amino acid in the interactive fragment map.

The functions of the amino acid information are described in Table 25.

Table 25. Amino Acid Information

Item	Function
Position	Displays the relation of the selected amino acid to the N and C terminals.
Amino Acid	Displays the IUPAC single letter designation of the selected amino acid.
RESID	Displays the RESID designation of the selected PTM.
Start PTM	Displays the PTM attached to the amino acid as of the last score.
PTM Tiers	Displays PTMs in user-defined tiers that can be added to this amino acid.

Fixed Modifications Box

Each fixed modification supported by ProSightPC is listed by amino acid.

Tip You can select fixed modifications in the search adder, or they can be changed during rescoring.

Select **None** to indicate that no fixed modifications are presently selected for that type of amino acid and will not be included at the next rescore.

Each amino acid can have no more than one fixed modification.

Matching Fragment Table

The matching fragment table contains a summary of all fragment ions matching the protein.

The interactive fragment map and matching fragment table are linked for convenient data browsing. Click a fragment name in the table to select the terminal amino acid within the fragment map. Alternatively, select the terminal amino acid of a fragment in the graphical map to highlight the corresponding fragment name in the fragment table.

Information relating to the matching fragments is described in Table 26.

Table 26. Matching Fragment Table

Display	Function
ID	Displays a unique (within each MS/MS experiment) ProSightPC-assigned numerical identification of the fragment.
Name	Displays a name formed by noting the ion type followed by the terminal amino acid number.

Table 26. Matching Fragment Table, continued

Display	Function
Mass (type)	Displays the observed mass of the fragment ion, measured in Da. The type will be monoisotopic or average depending on which was selected during the last rescoring.
Theoretical Mass	Displays the mass of the corresponding theoretical fragment ion.
Error (Da)	Displays the difference between the fragment mass and the theoretical fragment mass, measured in Da.
Error (ppm)	Displays the difference between the fragment mass and the theoretical fragment mass, measured in ppm.
Δm Mode	Displays the word `True` if the corresponding fragment is a match considering the Δm.

Nonmatching Fragments Table

The nonmatching fragments table lists every fragment not matching the sequence.

The information within the nonmatching fragments table is described in Table 27.

Table 27. Nonmatching Fragments Table

Display	Function
Mass (type)	Displays the observed mass of the fragment ion, measured in Da.
m/z (type)	Displays the *m/z* value corresponding to the fragment ion.
Intensity	Displays the intensity of the fragment ion.

Using Sequence Gazer

Use Sequence Gazer to characterize a previously identified protein.

When a protein sequence is taken to Sequence Gazer, ProSightPC automatically scores the sequence based on the initial search parameters.

You can change parameters and add or remove PTMs or fixed modifications. The modified sequence is then rescored. Ideally, changes to the sequence followed by rescoring net more matching fragments than before, thus narrowing the possible matching protein forms explaining the MS/MS data.

By rescoring, ProSightPC compares the new protein sequence configuration with all changes in place to the fragment ion data to determine the new number of fragments explained, along with all corresponding scores.

Chapter 8 Understanding the Data Manager

The Data Manager provides a visual representation of all the information related to a single MS/MS experiment.

This chapter describes how the Data Manager displays information. It contains the following sections:

About Data Manager

Use the Data Manager to view all information regarding a single experiment. Use the context-sensitive controls to determine what information is displayed. Figure 42 shows the Data Manager.

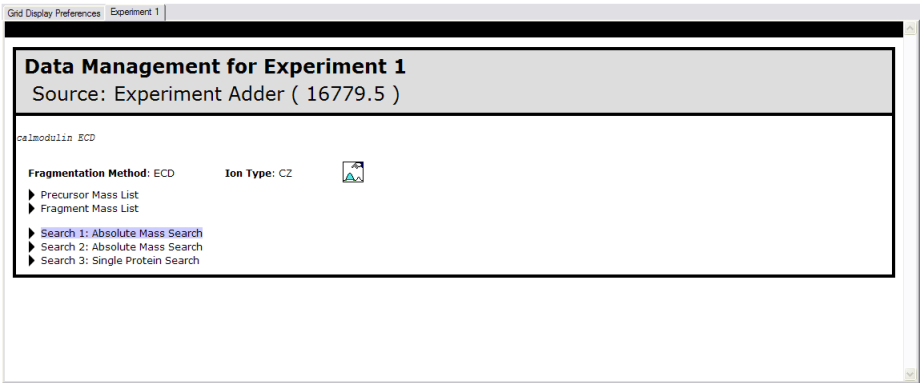

Figure 42. Data Manager

Display Controls

Click the display control arrows to reveal or conceal various displays The arrows indicate the status of the displayed information.

Side Arrows

A side arrow indicates that information relating to a search is concealed. Click the side arrow, as shown in Figure 43, to reveal the information. This turns the side arrow icon into a down arrow.

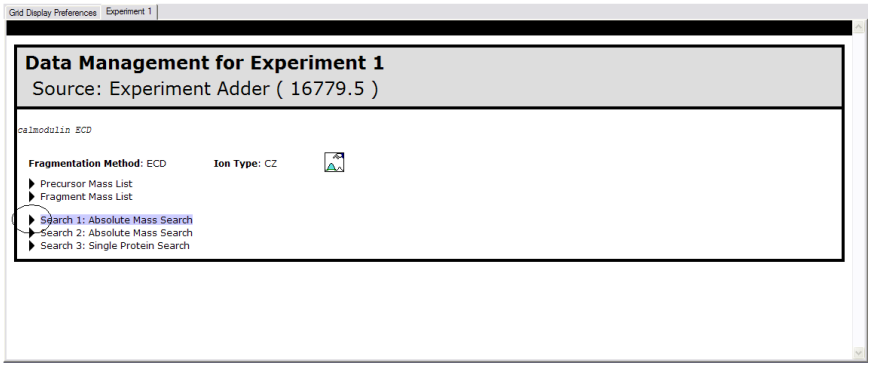

Figure 43. Data Manager Side Arrow

Down Arrows

Down arrows indicate that information relating to a search is presently displayed. Click the down arrow to conceal the information. The arrow changes to a side arrow.

Data Manager Organization

Information within the Data Manager is displayed in two functional groups:

- Instrument Data
- Search Data

Instrument Data

Instrument data includes the mass values, fragmentation method, and ion type of the MS/MS experiment.

Note If you have defined an experiment-level comment, it is displayed at the top of the Data Manager.

Adding or Editing an Experiment Comment

Add or edit an experiment comment by using the **Experiment Tools > Edit Comment** menu.

To add or edit an experiment comment

1. Choose **Experiment Tools > Edit Comment**.
2. Type the comment in the field provided.
3. Click **Save** to save the comment.

Tip Use an experiment comment to record information relating to all the searches, which liquid chromatography fraction the data came from, and information regarding the mass spectra used to create the mass list.

Editing Mass Values

You can edit mass values within an experiment.

Use one of the following two methods to edit mass values:

- Click **Edit Masses** within the Data Manager.
- Choose **Experiment Tools > Edit Masses**.

Either method opens a new tab in the Data Manager, as Figure 44 shows.

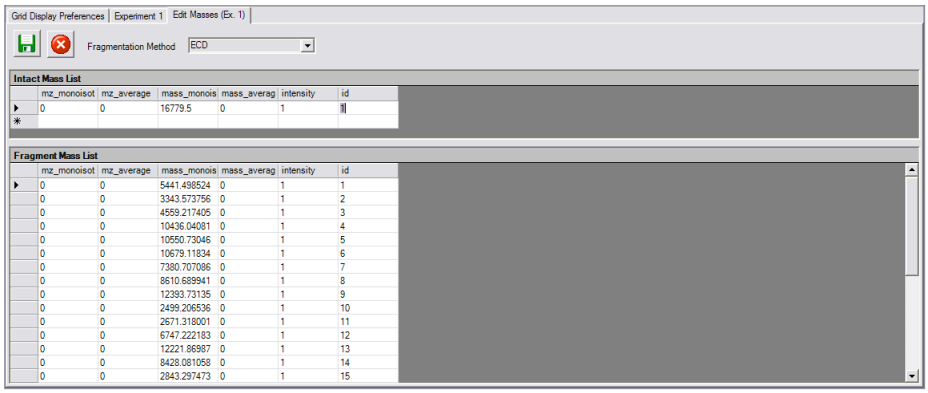

Figure 44. Edit Mass Values

The mass values display in a series of columns within two fields:

- Intact Mass List
- Fragment List

To edit mass values

1. Select any value from either of the two mass lists.
2. Click **Delete** or use the backspace key to remove the old value. Type a new value in the blank space.
3. Click **Save Changes** to save the changes. ProSightPC returns you to the Data Manager.

Tip

1. The last row of the Data Grid is marked with a "*." You can add a new value, if you prefer.
2. Click in the margin to the left of the mass list to select an entire row. Click **Delete** to remove the entire row from the mass list

Fragmentation and Ion Types

When a MS/MS experiment is generated, the fragmentation method used to generate the MS/MS data is declared. From this input, ProSightPC determines which of the following ion types to use during searches:

- Electron capture dissociation (ECD) fragmentation is analyzed using *c/z.* ions.

- Collisionally-induced dissociation (CID) fragmentation is analyzed as *b/y* ions.

Search Data

A numerical list of searches arranged by search ID follows the instrument data. The search type and status are displayed.

If the search is highlighted in blue, the search is Pending and has yet to be run.

Pending Search

Click the side arrow of a pending search to reveal the search parameters and a **Run Search** button. Click the **Run Search** button to run the search.

Note Search Parameters are discussed in **Chapter 6: Search Modes**

Edit Comment

Click the **Edit Comment** button to enter and save information specific to the search.

Results

A completed search generates a results list, as Figure 45 shows.

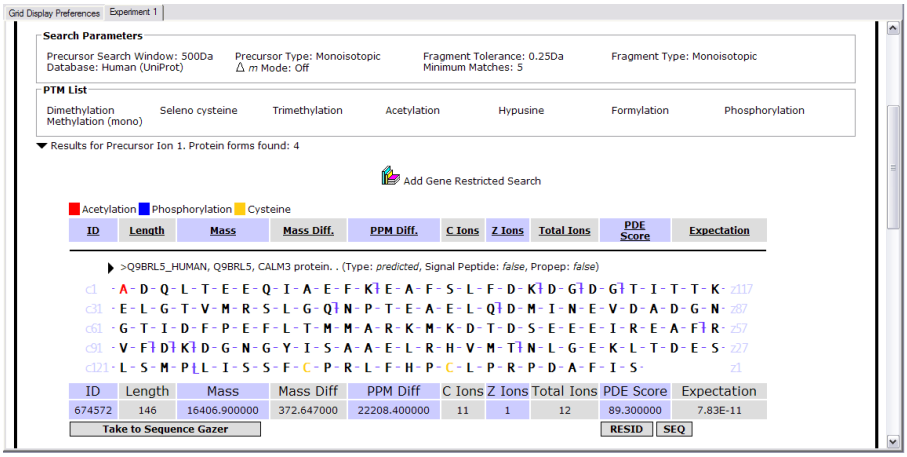

Figure 45. Results List

IMPORTANT Absolute mass and biomarker searches return one result list for each precursor ion.

Each result list displays the number of protein forms found. Click the side arrow next to the results list to display the results contained in the result table. Each result table contains complete information about each matching protein form. Information within the result table is organized into the three regions described in Table 28

Table 28. Result Table Regions

Region	Description
Description	This is a short description of the protein form.
Fragment Map	This is a graphical representation of the protein form, showing the location of PTMs and matching fragment ions.
Statistics Table	This organizes information relating to the search. See Table 29.

The statistics table is subdivided into the display elements described in Table 29.

Table 29. Result Statistics Table

Statistic	Description
ID	The internal identifier for the protein form.
Length	The number of amino acids in the protein form.
Mass	The theoretical precursor mass of the protein form.
Mass Difference	The observed mass minus the theoretical mass.
PPM Difference	The mass difference in terms of ppm.
N-terminal Ions	The total number of matching N-Terminal ions.
C-terminal Ions	The total number of matching C-Terminal Ions.
Total Ions	The total number of matching ions.
PDE Score	The PDE score.
Expectation	The expectation score.

Each result has three context-sensitive buttons, described in Table 30.

Table 30. Context-Sensitive Buttons

Button	Description
Take to Sequence Gazer	Creates a new Single Protein Mode search based on the result. See Chapter 7 for details.
RESID	Displays a RESID-annotated sequence.
SEQ	Displays the sequence.

Note

1. Click the text within the header column to sort the results list in ascending or descending order. Click again to reverse the order.

2. ProSightPC automatically generates a color-coded legend. An amino acid bearing a PTM will be color coded according to this legend. Cysteines are always colored yellow.

Reading the Fragment Map

The matching fragment table contains a summary of all fragment ions matching the protein.

Absolute mass, biomarker, single protein, gene-restricted absolute mass, and gene-restricted biomarker mass searches all return similar results.

Table 31 describes the functions displayed in the matching fragments table.

Table 31. Matching Fragment Table

Display	Function
ID	A unique (within each MS/MS experiment) ProSightPC-assigned numerical identification of the matching fragment.
Name	The ion type followed by the distance from the terminus.
Mass (type)	The observed mass of the fragment ion, measured in Da. The type is either monoisotopic or average depending on search parameters.
Theoretical Mass	The mass of the corresponding theoretical fragment ion.
Error (Da)	The difference between the fragment mass and the theoretical fragment mass, measured in Da.
Error (ppm)	The difference between the fragment mass and the theoretical fragment mass, measured in ppm.
Δm Mode	The word `True` if the corresponding fragment is a match considering the mass shift, Δm.

Building Gene-Restricted Searches from a Results List

Build a gene-restricted search for any results list within the Data Manager.

To access the new gene-restricted search dialog box

1. Click the **Add Gene Restricted Search** icon, shown in the oval in Figure 46.

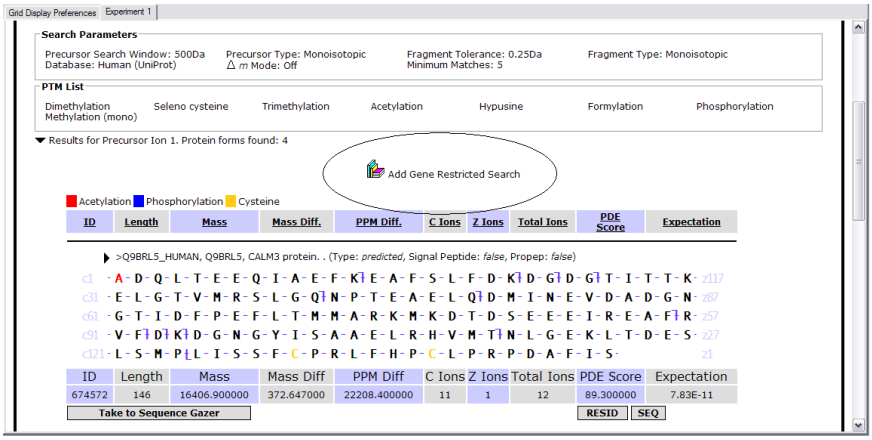

Figure 46. Add Gene Restricted Search

2. The New Gene Restricted Search dialog box appears, as Figure 47 shows.

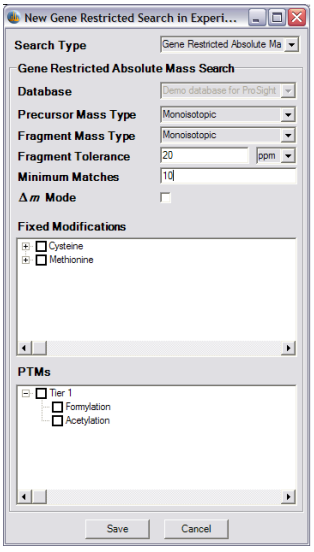

Figure 47. New Gene-Restricted Search dialog box

Building a New Gene Restricted Search

Build a new gene-restricted biomarker or gene-restricted absolute mass search after the New Gene Restricted Search dialog box appears.

To build a new gene-restricted search

1. Select the Search Type.
2. Type the appropriate gene-restricted parameters. These parameters are described in Chapter 6, "Search Modes."
3. Click **Save** to append the new search to the experiment and return to the Data Manager. The status of the new search is pending.
4. Click **Run Search** to run the search.

Chapter 9 Proteome Databases

Proteome databases are a critical element of ProSightPC. The various searches require sequence information in order to identify and characterize proteins. This sequence information is stored in proteome databases within the PTM Warehouse.

This chapter describes the PTM Warehouse and how to create, manipulate, and modify proteome databases. It contains the following sections:

The PTM Warehouse

The PTM Warehouse stores protein sequence and PTM information in proteome databases. The sequence and PTM information are combined using shotgun annotation, as explained in Chapter 1, "Top-Down Proteomics."

About Proteome Databases

The PTM Warehouse consists of a collection of proteome databases and a small amount of metadata. Proteome databases contain the shotgun annotation of all possible combinations of PTMs on each basic sequence in the proteome. Any given possible combination of PTMs on a basic sequence is referred to as a protein form.

Internal Database Names

Each proteome is stored as a proteome database and is uniquely identified by an internal name consisting of one or more letters A-Z without spaces or punctuation. For example, E. coli (UniProt) might be designated as ecoli_uniprot.

Importing Data into the PTM Warehouse

Use the following two methods for importing data into the PTM Warehouse. Each method is covered in more detail in the Database Manager section of this chapter.

- Load databases from PWF Files
- Create Databases from Swiss-Prot or FASTA-formatted text files.

The Database Manager

Use the database manager to handle all PTM Warehouse management and manipulation functions. Access the database manager from the ProSightPC Applications folder located on your desktop.

To access the Database Manager

1. Double-click **ProSightPC Applications**.
2. Double-click **Database Manager**. A new window appears, as Figure 48 shows.

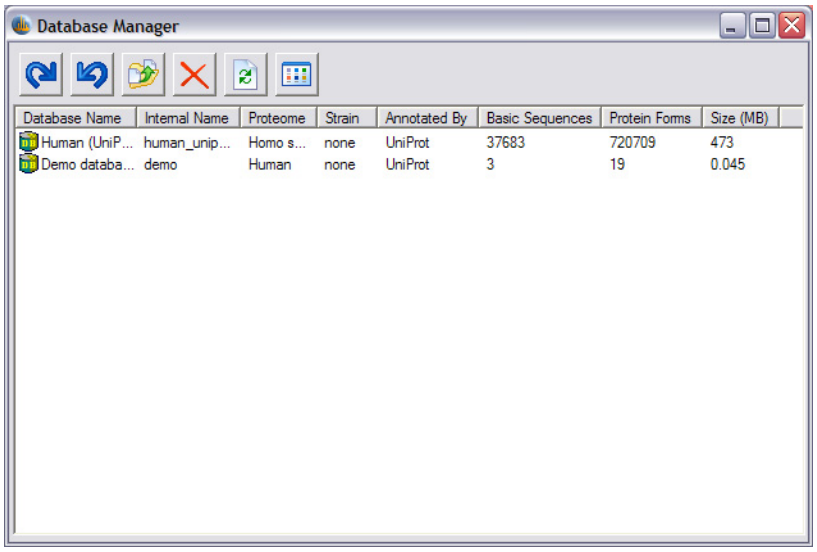

Figure 48. Database Manager

Database Manager Organization

The database manager graphical user interface (GUI) consists of the following two elements:

- The Data Grid
- The Toolbar

The Data Grid

The data grid contains one row for each proteome database within the PTM Warehouse.

Table 32 describes the organization of the Data Grid.

Table 32. Data Grid Columns

Component	Description
Database Name	Describes the proteome database.
Internal Name	Displays the internal name.
Proteome	Describes the proteome from which the protein sequence data is derived.
Strain	Displays optional information about the strain of the organism from which the proteome was derived.
Annotated By	Describes the source of the proteome data.
Basic Sequence	Lists the number of basic sequences in the database.
Protein Forms	Lists the number of protein forms in the database.
Size (Mb)	Lists the physical size of the database in MB.

The Toolbar

Use the toolbar to do the following:

- Import a Proteome Database
- Export a Proteome Database
- Create a Proteome Database
- Remove a Proteome Database
- Change View

Import a Proteome Database

Use the Import Proteome Database tool in the ProSightPC Application folder to load an existing proteome database into your PTM Warehouse.

When opened, the Import Databases dialog box shown in Figure 49 displays a list of all patches and proteome databases in the PWF file. All patches and proteome databases, when present, are preselected. You must manually deselect them to prevent them from being loaded.

To import a proteome database

1. Click **Import Proteome Database**.
2. Select the desired PWF file.
3. Select one or more proteome databases and patches to be loaded.
4. Click **Import**.

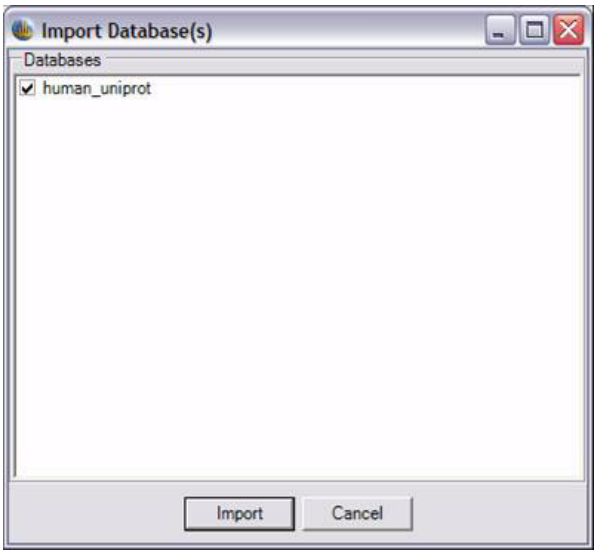

Figure 49. Import Database(s) dialog box

If any internal name already exists within the PTM Warehouse, the Internal Name Conflict dialog box appears, as Figure 50 shows.

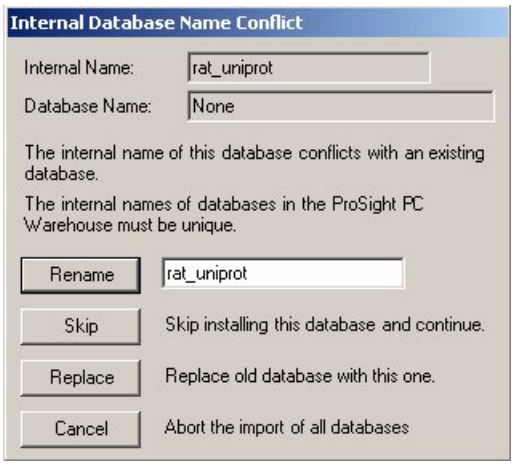

Figure 50. Internal Name Conflict dialog box

The Internal Name Conflict dialog box alerts you if the entered internal name already exists within the PTM Warehouse. The appearance of this dialog box prompts you for one of the following actions:

- Rename
- Skip
- Replace
- Cancel

Note Importing databases as PWF files is faster than recreating a proteome database with the Create Proteome Database option, a process detailed in the following sections.

Export a Proteome Database

Click the Export Proteome Database icon to export proteome databases from your PTM Warehouse to a PWF file. The Export Database(s) dialog box appears, as Figure 51 shows.

Use the Export Databases tool, located in the Database Manager, to export one or more databases.

To export a proteome database

1. Select the database to export. To combine multiple databases into a single export file, press the CTRL key and select additional databases.

2. Click **Export ProSight Database**. A new dialog box appears, as Figure 51 shows.

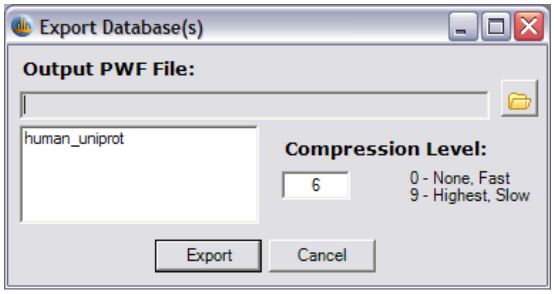

Figure 51. Export Database(s) dialog box

3. Type a destination PWF file, either manually or with the file picker provided.

4. Select a compression level by typing an integer ranging from 9(slow export but small file size) to 0 (fast and large file size). This controls the final PWF file size.

5. Click **Export** to execute the procedure and create a new PWF file.

Create a Proteome Database

Create your own shotgun-annotated proteome databases. These databases are restricted to one of the following two input file formats:

- Swiss-Prot
- FASTA

Accessing Create Proteome Database

Use the Create Proteome Database tool in the database manager to create a proteome database.

To access create proteome database

- Click **Create Proteome Database**. The following dialog box appears, as Figure 52 shows.

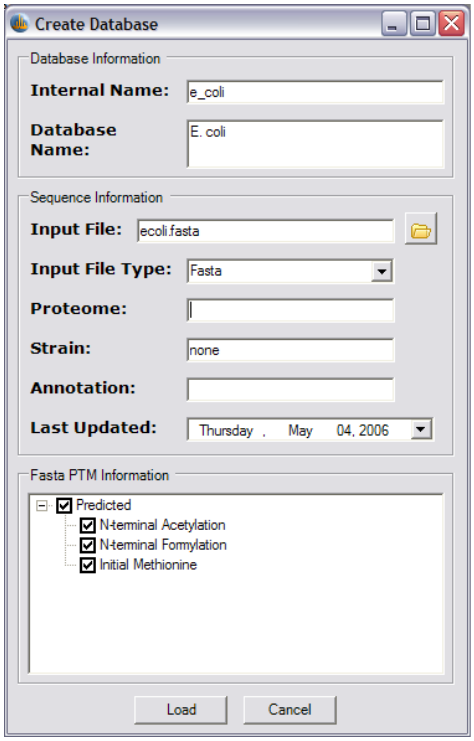

Figure 52. Create Proteome Database dialog box

Organization of Create Proteome Database

Information within the Create Proteome Database dialog box appears in the following three regions:

- Database Information
- Sequence Information

- PTM Information

Database Information

Database information is divided into two fields, as Table 33 describes:

Table 33. Database Information

Item	Description
Internal Name	Type the internal name of the new proteome database. This name must be unique.
Database Name	Type the name for the new proteome database. This is the name that displays in the ProSightPC interface.

Sequence Information

Sequence information is divided into the following fields, as Table 34 describes.

Table 34. Sequence Information

Item	Description
Input File	Type or use the tool to select the path to the input files.
Input File Type	Select Swiss-Prot or FASTA.
Proteome	Enter a description of the proteome.
Strain	(Optional) Enter strain information for the proteome database.
Annotation	Type a short description of the source of the data.
Last Updated	Type the date when the database was last updated.

PTM Information

This frame filters which PTM information is shotgun annotated into the new proteome database. The selectable content of the PTM information frame is determined by file type (Swiss-Prot or FASTA).

FASTA

FASTA files contain no PTM information, thus only predicted PTMs can be shotgun annotated to their sequences.

Swiss-Prot

Swiss-Prot files store a large amount of PTM information. For this reason, all PTMs listed in RESID are available for shotgun annotation. However, only the PTM information in the source Swiss-Prot file can actually be shotgun annotated into the proteome database.

PTMs are arranged in PTM tiers. These tiers are described in Chapter 11, "PTM Tier Editor."

Creating a Proteome Database

Create proteome databases from within the Database Manager.

To create a proteome database from either a FASTA or Swiss-Prot file

1. Click **Create Proteome Database**.
2. Type the database name in the field provided.
3. Type the internal name in the field provided.
4. Type the path to the file containing the sequence information.
5. Select the input file type from the drop down menu.
6. Type the proteome name in the field provided.
7. (Optional) Type the Strain name in the field provided.
8. Type the annotation information in the field provided.
9. Click the drop-down menu to display a calendar and select a different date. The current date defaults to the Last Modified field.
10. Select the PTMs you want to use from the list.
11. Click **Load** to create the new database.

IMPORTANT This process can take several hours or, in some cases, days to complete, depending on your hardware and the data being shotgun annotated.

Remove a Proteome Database

Eliminate unwanted proteome databases from the PTM Warehouse by selecting the proteome databases in the Data Grid and using the **Remove Database** function.

To remove a proteome database

1. Select the unwanted proteome database in the data grid.
2. Click **Remove Proteome Database**.
3. Click **Yes** to remove the proteome database from the PTM Warehouse.

CAUTION Removing a proteome database from the PTM Warehouse is a permanent change and cannot be undone except by reloading the data from the original source into the PTM Warehouse.

Change View

Click the **Change View** button to change the display of databases in the database manager. Choose whether the databases are displayed in detail, list, and small or large icons.

Note

You can drag-and-drop files within the database manager.

- PWF files trigger the Import Proteome Database function.
- FASTA and Swiss-Prot files trigger the Create Proteome Database function.

ProSightPC Proteome Database Tools

ProSightPC provides two additional tools for managing the PTM Warehouse:

- Individual Sequence Adder
- View Database Information

Individual Sequence Adder

Use the Individual Sequence Adder to add all shotgun-annotated forms of a single protein to an existing proteome database. Access the Individual Sequence Adder through the Tools menu.

To access the Individual Sequence Adder

- Choose **Tools > Individual Sequence Adder**. A new dialog box appears, as Figure 53 shows.

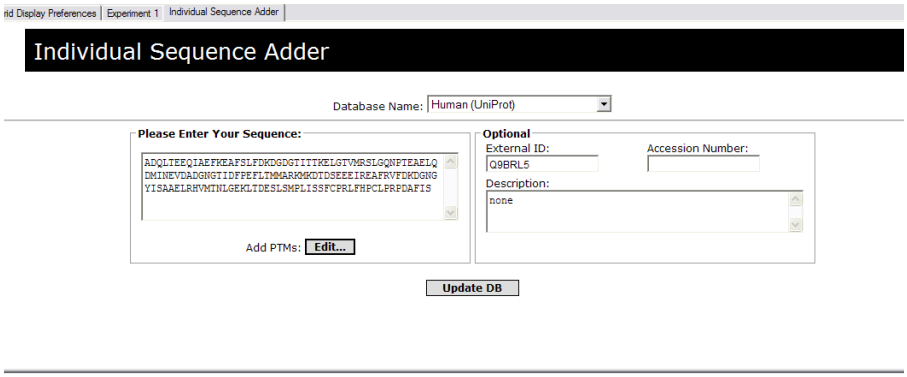

Figure 53. Individual Sequence Adder dialog box

To use the Individual Sequence Adder

1. Type the sequence, without spaces or extra characters, in the field provided.
2. Click **Edit** to add PTMs. A new dialog box appears that displays the sequence, as Figure 54 shows.
3. First click an amino acid to select a PTM. A box appears around the amino acid. Any PTMs known to occur on the selected amino acid appear in the Amino Acid Information Box to the right.

4. Click the name of the PTM. A colored box matching the PTM appears around the amino acid.

5. Select an amino acid and click **None** to remove the PTM from the amino acid.

6. Click **Accept** to integrate the PTM changes to the Individual Sequence Adder.

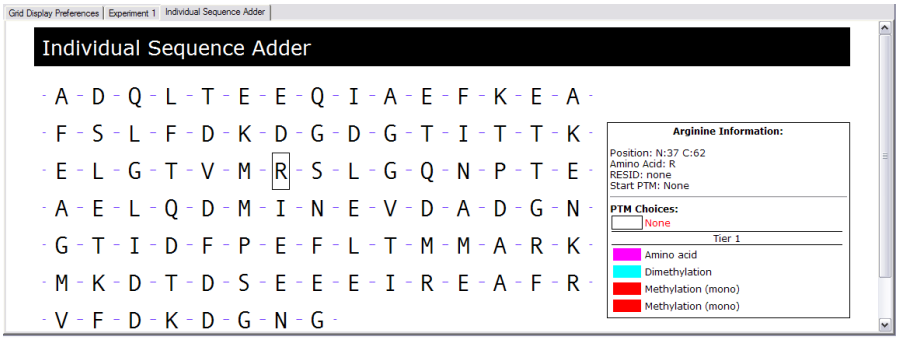

Figure 54. Individual Sequence Adder PTM Selection

Entering Optional Data

Add additional information into the Individual Sequence Adder before or after PTM selection This information is optional data and is described in Table 35.

Table 35. Optional Data

Data	Description
External ID	Type any sequence identifier.
Accession Number	Type the Accession Number, if available.
Description	Type a description of the protein, up to 255 characters.

Note

1. All of the data elements in Table 35 display if the sequence is found during a search.

2. ProSightPC functions without any of these three optional selections but including at least one of these data elements aids in identification.

View Database Information

View Database Information displays a summary of all proteome databases currently loaded into the PTM Warehouse.

To view Database Information

- Choose **Tools > View Database Information**. A new window appears, as Figure 55 shows.

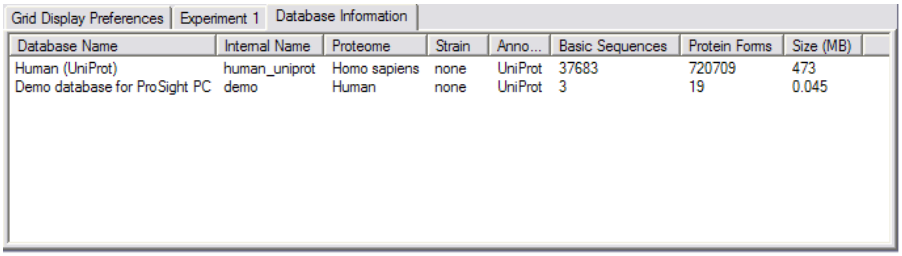

Figure 55. View Database Information dialog box

Table 36 describes the organization of the Database Information tab.

Table 36. Database Information

Item	Description
Internal Name	Displays the internal name of the proteome database. This name must be unique.
Proteome	Lists the description of the proteome.
Strain	Lists the strain information for the proteome database.
Annotated By	Lists the source of the proteome data.
Basic Sequence	Lists the number of unmodified protein forms within the proteome database.
Protein Forms	Lists the total number of shotgun annotated protein forms within the Proteome Database.
Size	Lists the physical size of the proteome database in MB.

Chapter 10 Noise Reducer

Fragmentation data frequently contains chemical noise peaks. Use the Noise Reducer to identify and remove these fragment masses.

This chapter describes how to use the Noise Reducer. It contains the following sections:

About Noise Reducer

Fragmentation data frequently contains chemical noise peaks. These undesirable peaks consists of fragment mass values representing observed ions that do not correspond to *b/y* or *c/z***.** fragments of the precursor ion or are data processing noise. These values arise from errors in peak picking algorithms or as a by-product of the fragmentation mechanism. Use the Noise Reducer to identify and remove some of these undesirable masses from an MS/MS experiment.

Once Noise Reducer has run, a new experiment identical to the source experiment is placed on the Data Grid. The new experiment lacks the undesirable fragment masses and previous experiment results are removed as they might contain the deleted noise peaks.

Table 37 describes three sources of chemical noise.

Table 37. Sources of Chemical Noise

Source	Description
NH_3 loss	Describes two or more peaks with masses differing by 17.01 Da plus or minus the user supplied tolerances. The lighter peak would be considered noise from the heavier.
H_2O loss	Describes two or more peaks with masses differing by 18.01 Da plus or minus the user supplied tolerances. The lighter peak would be considered noise from the heavier.
Isotope	Describes two or more peaks from different charge states, with masses differing exactly by 1.0 Da.

Accessing Noise Reducer

Access the Noise Reducer through the Experiment Tools menu.

To access Noise Reducer

1. Open the Data Manager of the experiment to be processed.
2. Choose **Experiment Tools > Reduce Noise**. The Noise Reducer window appears, as Figure 56 shows.

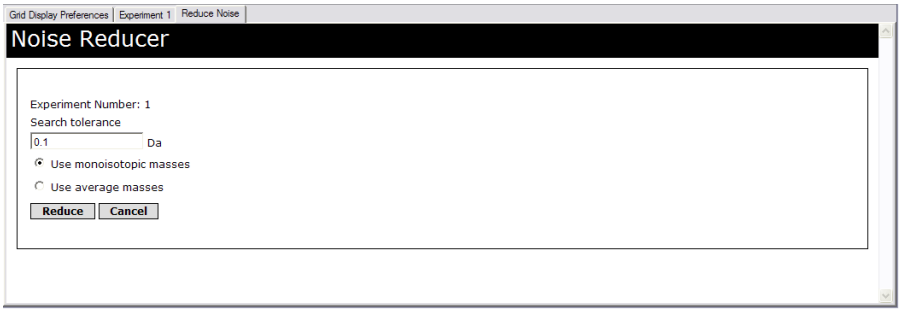

Figure 56. Noise Reducer

Note The Noise Reducer applies to the MS/MS experiment in the active Data Manager. If no Data Manager is active, the Noise Reducer is unavailable.

Using Noise Reducer

Use the Noise Reducer to identify and minimize the effects of chemical noise within an experiment.

The Noise Reducer requires two user-defined pieces of information, as Table 38 describes.

Table 38. Noise Reducer Inputs

Element	Description
Search Tolerance	Search tolerance is an allowable error, measured in Da, between the expected or observed mass differences.
Mass Type	Mass type indicates whether the noise reducer should consider monoisotopic or average mass types.

To run Noise Reducer

1. Type the desired search tolerance, measured in Da, in the field provided.
2. Select Monoisotopic or Average Mass Type.
3. Click **Reduce** to activate the Noise Reducer.

Interpreting the Result

The output of Noise Reducer consists of sets of fragment ion masses. The fragment ion masses are grouped together by

- Differences of one Dalton (Da)
- Loss of one NH_3
- Loss of one H_2O

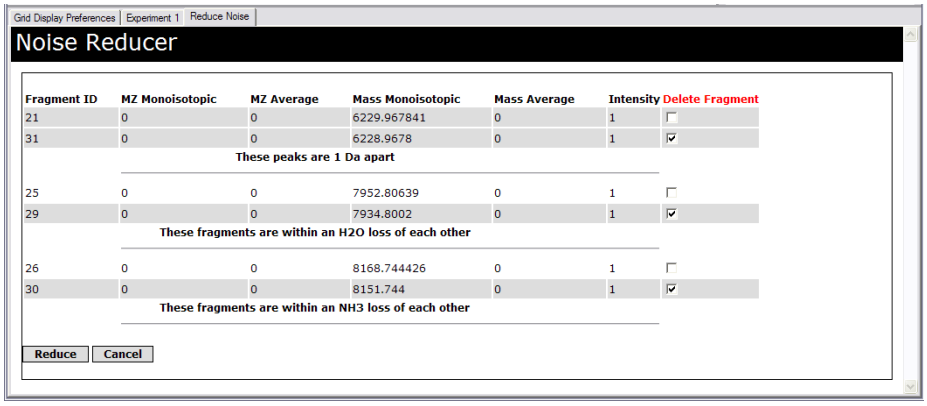

Figure 57. Noise Reducer Output dialog box

Inspect the fragment data displayed in Figure 57 to determine which, if any, of the fragments are possibly not true *b/y* or *c/z.* ions. Spurious fragments appear as shaded entries. Spurious fragments are generally undesirable within an experiment and should be deleted. For more information, see Deleting Spurious Fragments.

Note If all fragment ions in a set are shaded, the Noise Reducer is unable to determine which is most likely to be spurious.

Deleting Spurious Fragments

The Noise Reducer groups fragments that might be spurious.

To delete spurious fragments

1. Check the box next to each of the undesired peaks within each group.
2. Repeat for each group in the list.
3. Click **Reduce** to remove the spurious fragment masses.

   A duplicate experiment is added. The spurious fragments are omitted from the Mass List of the new MS/MS experiment.

   Note All searches within the new spurious fragment-free MS/MS experiment will be pending.

Chapter 11 PTM Tier Editor

ProSightPC groups all PTMs in a multitier structure, allowing you to find and select PTMs quickly.

This chapter describes how to use the Tier Editor to view and to change the tier assignment of PTMs. It contains the following sections:

About PTM Tier Editor

Assigning PTMs into tiers is intended to help you locate and select PTMs quickly and efficiently in your searches. All of the PTMs in the Tier Editor come from the RESID database.

The PTM Tier Editor has two functions:

- To permit you to reassign PTMs within the tier system
- To display which PTMs are currently included or excluded

  You select which PTMs will be available to be included in or excluded from analyses conducted by ProSightPC.

ProSightPC automatically assigns many PTMs to Tier 1 and Tier 2. Tier 1 PTMs represent the most common PTMs while rarer PTMs are assigned to Tier 2.

Note

1. Tier assignments greater than 2 can be entered.
2. The Tier Editor does not append PTM information into databases. The PTM information must be present in the proteome database prior to analyzing MS data.

Accessing PTM Tier Editor

Access the Tier Editor through the **Tools** menu.

To access Tier Editor

- Choose **Tools > Tier Editor**. The PTM Tier Editor dialog box appears, as Figure 58 shows.

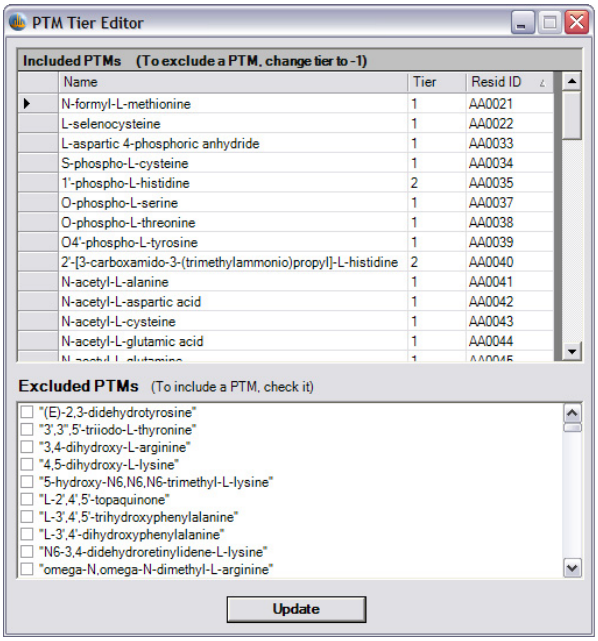

Figure 58. PTM Tier Editor dialog box

The PTM Tier Editor has two main display areas:

- Included PTMs
- Excluded PTMs

Including and Excluding PTMs

Use the Tier Editor to include or exclude PTMs. Included PTMs are available to the Sequence Gazer, the Database Loader, and to all search modes.

ProSightPC comes with a preset list of included PTMs. Any PTM not listed is excluded.

Included PTMs

Within the Included PTMs area, information relating to all presently included PTMs is arranged in three columns, as Table 39 shows.

Table 39. Included PTM Information

PTM Information	Description
Name	States the RESID name of the included PTM.
Tier	States the current tier assignment of the included PTM.
RESID ID	States the RESID identifier of the included PTM.

Tip Sort the columns of Included PTMs in ascending or descending order by clicking on the header.

Excluded PTMs

All presently excluded PTMs are listed in the Excluded PTMs field. Use the PTM Tier Editor to reclassify an excluded PTM as included.

To reclassify an Excluded PTM

1. Check the box to the left of any desired PTMs.
2. Click **Update** to make the changes.
3. The reclassified PTMs reappear within the Included PTM field, above.

Note Reclassified PTMs are automatically designated as Tier 1.

Moving PTMs Between Tiers

Use the Tier Editor to manually reassign a PTM to another tier.

To manually reassign a PTM to a tier within the Included PTMs field

1. Click the Tier Number to the right of the PTM to be reassigned.
2. Type a new positive integer in the Tier Number column.
3. Click **Update** to commit the change.

Note

1. Tier assignments greater than 2 can be entered.
2. The Tier Editor does not append PTM information into databases. The PTM information must be present in the proteome database prior to analyzing MS data. If the information for a given PTM is not in the proteome database, that PTM will not be available for selection in database searches, even if the PTM is considered included by the Tier Editor.

Excluding Included PTMs

Remove PTMs from the Included PTMs list by changing the Tier Number value.

To remove a PTM from the Included PTM list

1. Select the Tier Number to the right of the PTM you want to remove.
2. Click **Delete** to remove the current value.
3. Type -1 as the new Tier Number.
4. Click **Update** to commit the change.

Note Once **Update** executes, any excluded PTMs reappear in the Excluded PTMs field.

Chapter 12 Manage Data

ProSightPC provides several batch processing and reporting tools for managing large numbers of MS/MS experiments. Use these tools to simplify working with several experiments within a single PUF file.

This chapter describes how to manage multiple experiments. It contains the following sections:

Complex Data Analysis Tasks

Use the following tools to manage multiple experiments:

- Batch Processing
- Reports

Batch Processing

Batch processing enables a large number of searches, over any number of experiments, to be queued and run in a single action. Use Batch Processing when you have many pending searches in a PUF file and you would like to run all of them.

Batch Processing is available from with the **Tools** menu.

To access batch processing

1. Choose **Tools > Batch Run**.
2. Each pending search will be queued and will run in turn.

Tip To save time, use predefined searches as you import data and run all your predefined searches as a single batch job.

Reports

You can summarize your work with two types of reports.

A status report gives a summary of every search in the open PUF file, including search type and best score.

Status reports are available from the ProSightPC Tools menu.

To access status reports

1. **Open** the desired PUF file.
2. Choose **Tools > Reports > Status Report**.

A summary of all experiments and searches contained within the PUF file displays in a new window as a text document, as Figure 59 shows. This text document is organized by experiment number and is subdivided into the types of searches.

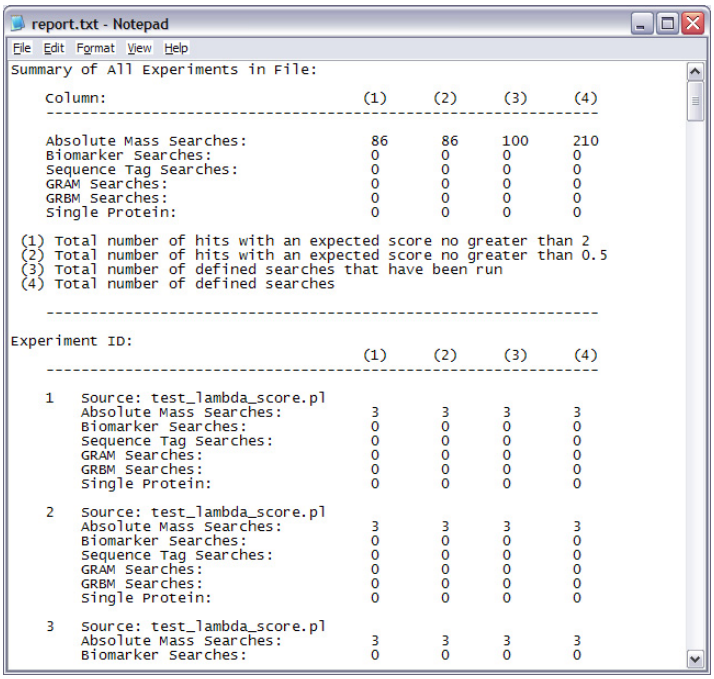

Figure 59. Status Report

Printable Search Report

A printable search report contains all of the information relate to one search, formatted for easy printing.

Accessing Printable Search Report

The Printable Search Report is accessible from the **Tools** menu.

To access the printable search report

1. Select an search in the Data Grid.
2. Choose **Tools > Reports > Printable Report**.

A new window appears, as Figure 60 shows.

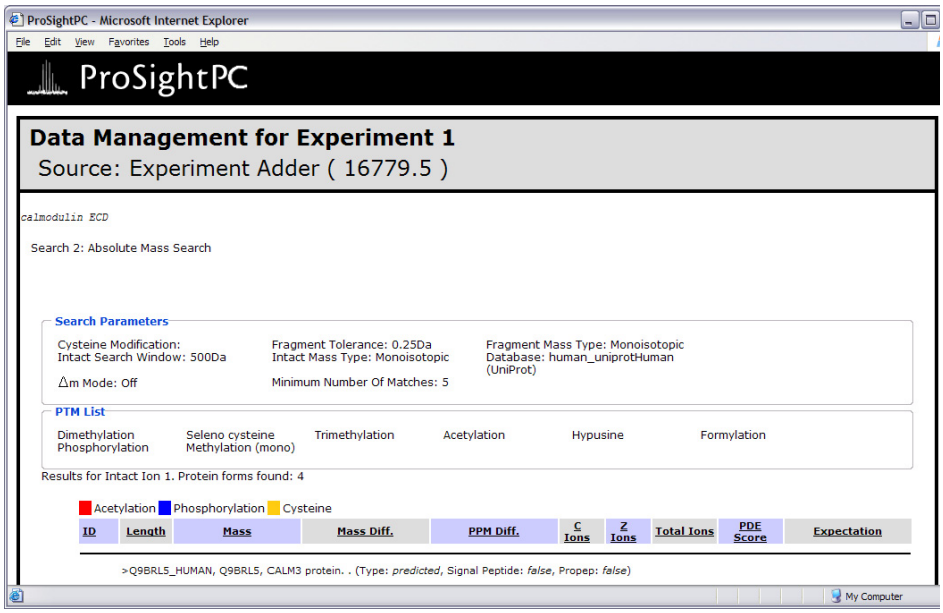

Figure 60. Printable Search Report

The printable report presents all relevant data for a search in a printable form similar to the Data Manager, but only contains information from the selected search.

Experiment Manager

Use the Experiment Manager to manipulate experiments as objects, move individual experiments between PUF files or save them in their own PUF file.

Note An experiment is defined as one or more precursor masses, one or more fragment masses, and all related searches.

Accessing the Experiment Manager

The Experiment Manager is accessible in the ProSightPC applications folder on the desktop.

To access the Experiment Manager

1. Double-click **ProSightPC Applications**.
2. Double-click **Experiment Manager** to start the Experiment Manager. A new window appears, as Figure 61 shows.

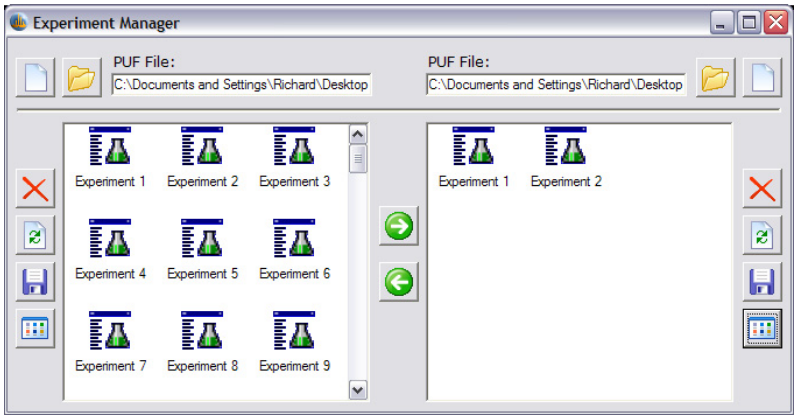

Figure 61. Experiment Manager dialog box

The top portion of the Experiment Manager contains two text boxes. Each represents a single PUF file.

You can either open an existing PUF file or create a new, but empty, file.

Creating a New PUF File

To create a new PUF file

1. Click **Create New PUF File**. A new window appears.
2. Select the desired directory and type the name of the new PUF file.

3. Click **Save** to save the PUF file.

Opening an Existing PUF File

To open an existing PUF file

1. Click **Open Existing PUF File**. A new window appears.
2. Select the PUF file from those listed.
3. Click **Open** to open the PUF file.

Note When a PUF file is opened, information about all experiments in the file is populated in the area below its name.

Moving Experiments Between PUF Files

Experiments can be moved between two open PUF files.

To move an experiment between PUF files

1. Select an experiment to be copied.
2. Drag and drop the experiment from the source field and onto the destination field. You can also click the green arrow to send a copy of the experiment to the destination field.
3. Click **Save**.

IMPORTANT When copying experiments to PUF files, if two or more experiments share experiment identification numbers, the Experiment Manager reassigns the experiment number of the incoming experiment.

Single PUF Files

The Experiment Manager also provides four features affecting a single PUF file, as Table 40 describes.

Table 40. Tools Affecting a Single PUF File

Modification	Description
Delete	Select one or more experiments and press Delete to remove the experiments from the PUF file.
Revert	Restores to the last Save, any experiments removed after the Save reappear in the field.
Save	Saves your work to a destination PUF file.
Change View	Changes the display of the experiments within the Source and Destination PUF file. A heavy dot indicates the selection is active.

Chapter 13 Fragment Predictor

The Fragment Predictor takes a known protein sequence and returns all possible *b*, *y*, *c*, and *z*• fragment ion masses. Use the Fragment Predictor to add post-translational modifications, or arbitrary custom masses, to any amino acid in the protein sequence, and see the predicted fragment ion masses.

This chapter describes how to use the Fragment Predictor. It contains the following sections:

Researchers might want to possess a list of all theoretical fragment ion masses particularly for modified protein sequences, prior to data collection.

Note The Fragment Predictor does not directly handle cross-linked proteins, such as trans-peptide bonds or cyclized species like disulfides. Compute these by using the Enter Custom Mass function.

Accessing Fragment Predictor

Access the Fragment Predictor through the **Tools** menu

To access the Fragment Predictor

- Choose **Tools > Fragment Predictor**. A window appears, as Figure 62 shows.

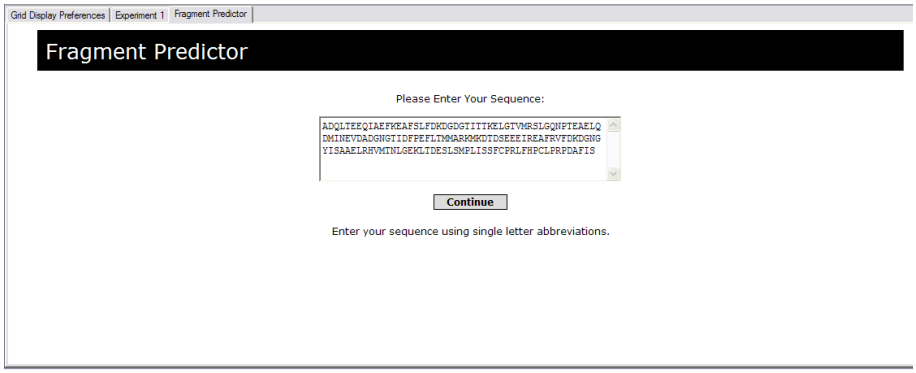

Figure 62. Fragment Predictor dialog box

Entering A Sequence

The Fragment Predictor accepts two methods for entering a protein sequence in the Protein Sequence Field.

- Manually enter a protein sequence.
- **Copy** and **Paste** a protein sequence from another source.

  The Fragment Predictor window displays one sequence text box, as Figure 63 shows.

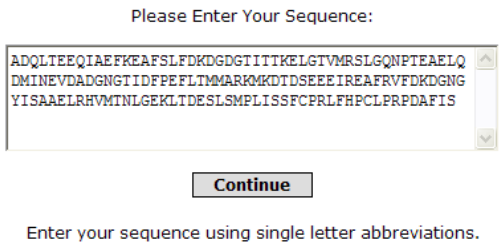

Figure 63. Protein Sequence dialog box

Note The sequence from any successful search can be accessed by clicking **RESID** or **SEQ** within the Data Manager. You can also acquire the protein sequence from external sources.

- Click **Continue** after typing the sequence in the field provided. A new window is displayed, as Figure 64 shows.

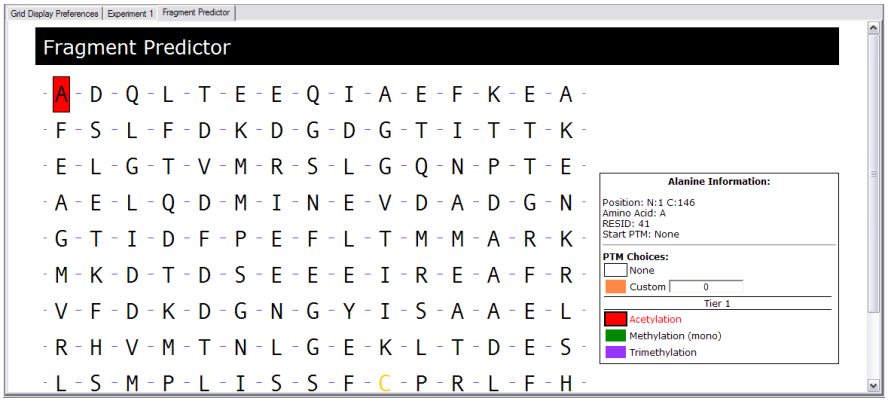

Figure 64. Fragment Predictor

The Fragment Predictor displays the protein sequences in an interactive sequence map. Click to select an amino acid. A black box around an amino acid indicates it is selected. For each selected amino acid, common PTMs display in the Amino Acid Information box to the right, as Figure 65 shows.

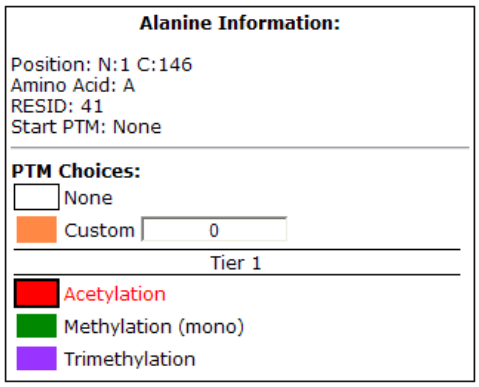

Figure 65. Amino Acid Box

PTMs are arranged in tiers. The PTM listed in red text is the current selection for the amino acid.

Tip Customize the PTM Tier assignment by using the PTM Tier Editor, covered in Chapter 11, "PTM Tier Editor."

To add a PTM to an amino acid

1. Select the amino acid.
2. Click the name of the desired PTM.

   The designated amino acid changes to match the color of the PTM selected.
3. If desired, a custom mass in Da can be entered in the field provided.
4. Click **Get Fragments** when all mass changes have been selected.

   The Fragment Predictor tool displays the theoretical fragment masses in the same tab of the Tab controller.

Interpreting the Output

Four columns of data display in the Results Window, as Figure 66 shows. All theoretical fragment ion masses are arranged in ascending order and are classified as either *b*, *y*, *c*, or *z*.

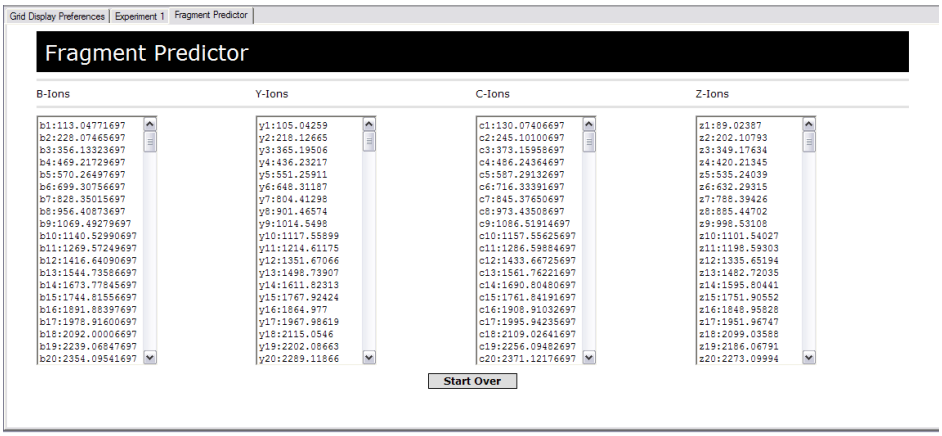

Figure 66. Results Window

Click **Start Over** to return to Step 1 of Adding a Sequence.

Chapter 14 Font Converter

The Font Converter converts text into the ProSightPC fragment map font. Use the Font Converter to generate fragment maps for inclusion in publications and presentations.

This chapter describes how to use the Font Converter. It contains the following sections:

About Font Converter

The Font Converter converts typed text into the ProSightPC font used to display N-Terminal and C-Terminal fragments. Use the Font Converter to generate fragment maps for inclusion in other media, such as publications and electronic presentations.

You can find the Font Converter in the **Tools** menu.

Accessing Font Converter

Access the Font Converter through the **Tools** menu of the ProSightPC Data window.

To access the Font Converter

- Choose **Tools > Font Converter**.

The Font Converter dialog box displays information in two fields:

- Sequence Field
- ProSightPC Font Output

Underneath these fields are six buttons that correspond to the N- and C-terminal fragment marks used for *b/y* and *c/z***.** fragment ions.

Using Font Converter to Convert Fonts

Use the Font Converter to convert typed text into the ProSightPC font.

1. Choose **Tools > Font Converter**. The Font Converter dialog box appears, as Figure 67 shows.

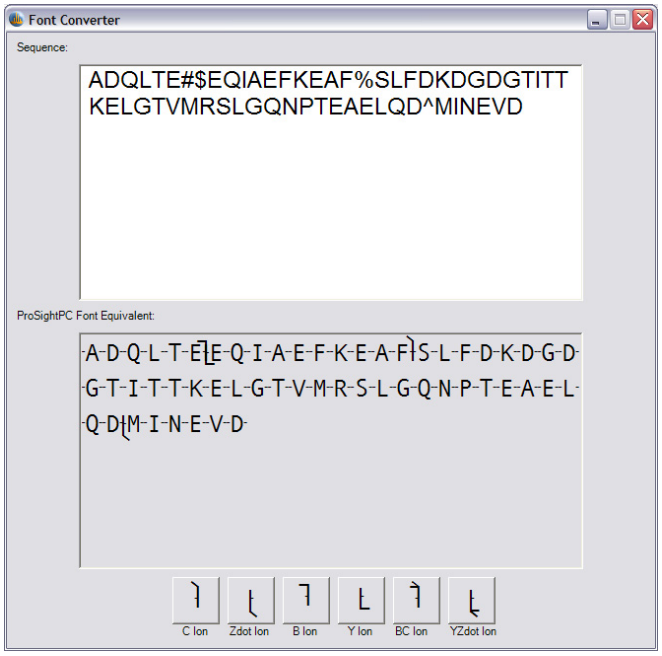

Figure 67. Font Converter dialog box

2. Enter the amino acid sequence to be converted into the Sequence field.

> **Note** The amino acid sequence can either be typed in the field, or you can **Paste** from another source.

3. To add fragmentation tick marks, position the cursor between the two amino acid letters and click the appropriate box. There are six boxes which correspond to b-ions, c-ions, z-ions, both b-ions and c-ions, and both y-ions and z-ions.

4. To display a complementary pair, click the appropriate N-terminal fragment and then click the appropriate C-terminal.

5. To transfer the converted font to another application, copy and paste text from the ProSightPC font-equivalent field to the other application.

You can resize the ProSight font after a paste operation.

Note Depending upon your system configuration, the font information might not transfer during a paste operation and might be displayed in another font. Correct this by selecting the incorrectly displayed output and manually changing the font to ProSight.

Index

G

H

I

J

L

M

N

O

P

R

S

T

V

Y

Z