ProSightPC

Version 2.0

User Guide

XCALI-97244 Revision A   May 2009

DOCUMENTATION
SURVEY

Thermo
SCIENTIFIC

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

Contents

Preface

This guide describes how to use ProSightPC™ to identify and characterize proteins.

Contents

- Related Documentation
- Special Notices
- System Requirements
- Obtaining a License
- Contacting Us

To provide us with comments about this document, click the link below. Thank you in advance for your help.

DOCUMENTATION SURVEY

Related Documentation

ProSightPC includes Help and these manuals as PDF files:

- *ProSightPC User Guide*
- *ProSightPC Quick Start Guide*
- *ProSightPC Installation Guide*

❖ **To view ProSightPC manuals**

- *ProSightPC User Guide*: Go to **Start > Programs > ProSightPC > ProSightPC User Guide**.

- *ProSightPC Quick Start Guide*: Go to **Start > Programs > ProSightPC > Quick Start**.

- *ProSightPC Installation Guide*: Go to **Start > Programs > ProSightPC > Installation Guide**.

❖ **To open Help**

- From the ProSightPC window, choose **Help > ProSightPC Help**.

- If Help is available for a specific window or dialog box, click **Help** or press **F1** for information about setting parameters.

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

Special Notices

Make sure you follow the precautionary statements presented in this guide. Special notices appear in boxes.

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

> **Note** Highlights information of general interest.

> **Tip** Highlights helpful information that can make a task easier.

System Requirements

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

System	Requirements
PC	- 1 GHz processor (2-GHz dual core recommended)
	- 1 GB 666 MHz RAM (2 GB recommended)
	- CD-ROM drive
	- 32 MB graphics card, 64 MB or greater
	- 75 GB or greater available on the C: drive
	- Video card and monitor capable of 1280 × 1024 resolution (XGA)
	- NTFS format

System	Requirements
Instrument (supported or required)	• LTQ FT™ • LTQ Orbitrap™
Software	• Xcalibur™ 2.0.7 • MYSQL 5.0 (must be installed before ProSightPC installation) The following files are included in the ProSightPC installation CD: • Microsoft™ Windows™ XP Professional 32-bit with Service Pack 2 • Microsoft .NET Framework 2.0 • Microsoft Visual C++ 2008 redistributable package • XDK Extension 0803

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

Obtaining a License

Two licenses are available for ProSightPC:

• A 30-day day trial license that is installed automatically when you install ProSightPC

• A permanent license that you can obtain from Thermo Fisher Scientific

❖ **To obtain a license for ProSightPC**

When you install ProSightPC for the first time, a 30-day license is installed. After installation, ProSightPC displays the message box shown in Figure 1 to tell you in how many days the trial license will expire.

Figure 1. License Expiration Warning message box
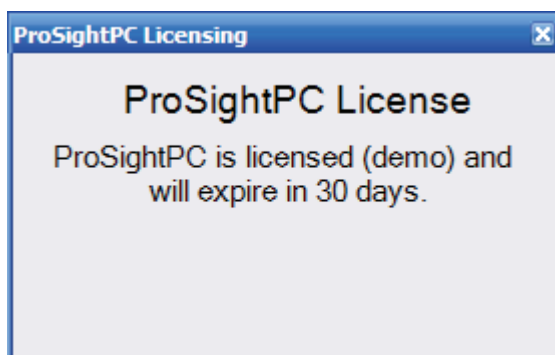
Obtaining a New Activation Code

❖ **To obtain a new activation code**

1. Open the **License Information** dialog box, shown in Figure 2, by using one of the following methods:

   • Run **Registration.exe** in the ProSightPC folder of your hard drive. For example, if you installed ProSightPC in C:\Program Files\, run C:\Program Files\ProSightPC\Registration.exe.

   –or–

   • Choose **Start > All Programs > ProSightPC**, or click the **ProSightPC** icon on your computer desktop to start ProSightPC and choose **Help > Manage License**.

Figure 2. License Information dialog box

2. Locate the bar code on the back of the ProSightPC CD jewel case, and type the serial number that appears below the bar code in the Product Key box.

3. Complete the information in the User Information area.

4. Send an e-mail message to Thermo Fisher Scientific with the license code:

   • Click **Copy to Clipboard** and paste the license code in the body of the e-mail message. Send the e-mail to license.ms@thermo.com.

   –or–

   • Click **Compose Email**, which directly creates an e-mail to send to Thermo Fisher Scientific.

When Thermo Fisher Scientific Customer Support sends you a new activation code, follow the instructions in "Installing the New Activation Code."

Installing the New Activation Code

After you receive your new activation code, start ProSightPC and open the License Information dialog box shown in Figure 2 to install it.

❖ **To install the activation code**

1. In the Add Activation Codes area of the License Information dialog box, paste the new activation code and click **Activate License**.

2. Click **OK** to accept the change in license.

3. Click **Close** to close the ProSightPC Browser License dialog box.

Contacting Us

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

❖ **To contact Technical Support**

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

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

❖ **To contact Customer Service for ordering information**

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

❖ **To copy manuals from the Internet**

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

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

- Fill out a reader survey online at www.thermo.com/lcms-techpubs.

- Send an e-mail message to the Technical Publications Editor at techpubs-lcms@thermofisher.com.

Introduction to ProSightPC

ProSightPC is a suite of tools designed to identify and characterize proteins and peptides from mass spectrometry data. This chapter introduces you to ProSightPC and to proteomics in general.

To install ProSightPC, see the *ProSightPC Installation Guide*.

Contents

- Features

- LC-MS/MS Workflow

- Inputs and Outputs

- Fragmentation Methods

- Ion Types

- Introduction to Proteomics

Features

ProSightPC is the only proteomics software suite that adequately supports high-mass-accuracy MS/MS experiments performed on LTQ FT and LTQ Orbitrap instruments. It operates on mass data from MS/MS experiments (or any MSn experiment) on intact and digested proteins.

As Figure 3 shows, ProSightPC first creates a new proteome database. Then it gathers intact protein sequences of a specific organism, along with information about known modifications, and loads them into a proteome warehouse (1). During loading, ProSightPC calculates all possible combinations of known modifications and applies them, along with single nucleotide polymorphisms (SNPs), splice variants, and sequence variants, to each protein sequence in a process called shotgun annotation (see "Shotgun Annotation" on page 12). Next, it imports the mass values inferred from mass spectral data from top-down and middle-down/bottom-up proteomics MS/MS experiments into a ProSightPC upload format (.puf) file (2). ProSightPC then searches the appropriate proteome databases for these mass values and compares them (3).

Figure 3. ProSightPC's protein and peptide identification process
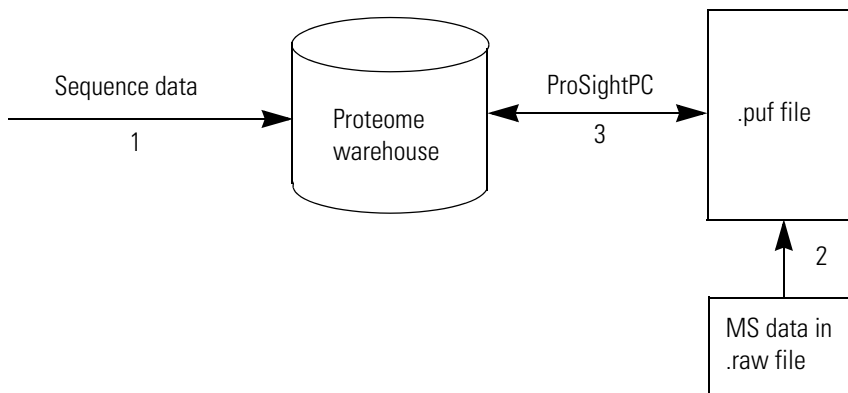
The basic unit of analysis in ProSightPC is the MS/MS experiment. An experiment is defined as one or more mass measurements of intact protein ions and the masses of one or more fragment ions that result from the disruption of those intact ions. Although many ProSightPC search modes accept multiple intact masses associated with a fragment ion mass list, performance improves when an MS/MS experiment consists of a single intact ion mass and a corresponding list of fragmentation masses. You must add complete MS/MS experiments in ProSightPC. You can add them to an existing .puf file or create a new .puf file for them. You can also edit existing MS/MS experiments. Each experiment is queried against the ProSightPC proteome warehouse in order to identify and characterize the proteins.

A search is a predefined query against the ProSightPC proteome warehouse. All experiments are associated with at least one search. By defining searches in the search logic (wizard) or during importations, you can use ProSightPC in a batch mode that facilitates high-throughput proteomic research.

The ProSightPC tool suite consists of the ProSightPC application and a small number of secondary applications to aid in managing the proteome database and experimental results.

The ProSightPC Application

ProSightPC can process a large number of searches to assist you in protein and peptide identification and characterization for high-resolution data. It can create automated iterative searches for batch processing, including search trees with decision points, to help create useful searches. It supports ultra-high-resolution MS/MS data, for example, top-down and middle-down/bottom-up LC-MS/MS data.

ProSightPC operates on a single .puf file that, when opened, is uploaded into memory and made available to a variety of search and data visualization tools. Additionally, ProSightPC includes several tools for importing LC-MS/MS and tandem MS data from Thermo Fisher Scientific .raw files, identifying and removing chemical noise peaks, and performing other utility functions. It can handle and store data in .raw format, ProSightPC upload format (.puf), or in a repository.

Proteome Warehouse

ProSightPC creates proteome warehouses, which are collections of databases that ProSightPC uses to identify and characterize protein data. It contains all the protein forms for a specific organism based on its sequenced genome. It stores many types of information, including known and predicted protein sequences, post-translational modifications (PTMs), alternate splice forms, and coding SNPs (cSNPs). The proteome warehouse contains both monoisotopic and average mass information and is organized to facilitate both protein identification and characterization. Each organism in the proteome warehouse receives its own database. You can create databases from Swiss-Prot or FASTA-formatted text files to create your own custom databases.

The databases in the warehouse are MySQL relational databases, which you can view by using other third-party applications. In addition, you can export them and move them between computers. ProSightPC searches these databases to try to find a match to the mass values inferred from mass spectral data from top-down and middle-down/bottom-up proteomics MS experiments. ProSightPC supports the creation of top-down and middle-down/bottom-up databases:

- Top-down (no sample proteolysis) databases are built around whole, intact protein sequences and everything that could potentially happen to them in a biological system.

- Middle-down/bottom-up (sample proteolysis) databases are built around peptide sequences that arose from proteolysis outside living organisms. Select this setting if anything in your sample preparation protocol involved trypsin or Lys-C or any other proteolysis agent.

For more information on top-down and middle-down/bottom-up databases, see "Top-Down Proteomics" on page 11 and "Middle-Down/Bottom-Up Proteomics" on page 10.

Chapter 8, "Using Proteome Databases," tells you how to create and manage databases in ProSightPC.

Search Types

ProSightPC supports the following types of searches. The section referenced for each search contains recommendations for running the search.

- Absolute mass searches, which are described in "Searching for Absolute Mass" on page 130

- Biomarker searches, which are described in "Searching for a Biomarker" on page 138

- Sequence tag searches, which are described in "Searching for a Sequence Tag" on page 145

- Single-protein searches, which are described in "Searching for Single Proteins" on page 177

- Gene-restricted absolute mass searches, which are described in "Searching for a Gene-Restricted Absolute Mass" on page 150

- Gene-restricted biomarker searches, which are described in "Searching for a Gene-Restricted Biomarker" on page 156

- MS^n hybrid searches, which are described in "Performing MS^n Hybrid Searches" on page 162

Iterative Searching

You can build an automatic, iterative, score-based search tree in ProSightPC. You select a predefined search, specify a condition, select an action, and select a category. All experiments pass through a first level of search logic, and the action taken next depends on the results of the search for each experiment. If the experiment results pass the condition that you set—for example, if at least one of the matching protein forms received an expectation value (e value) less than 1E-4—you can either load the experiment to the category selected or indicate that a second level of searching be performed. Figure 4 illustrates this methodology.

Figure 4. Iterative searching in ProSightPC
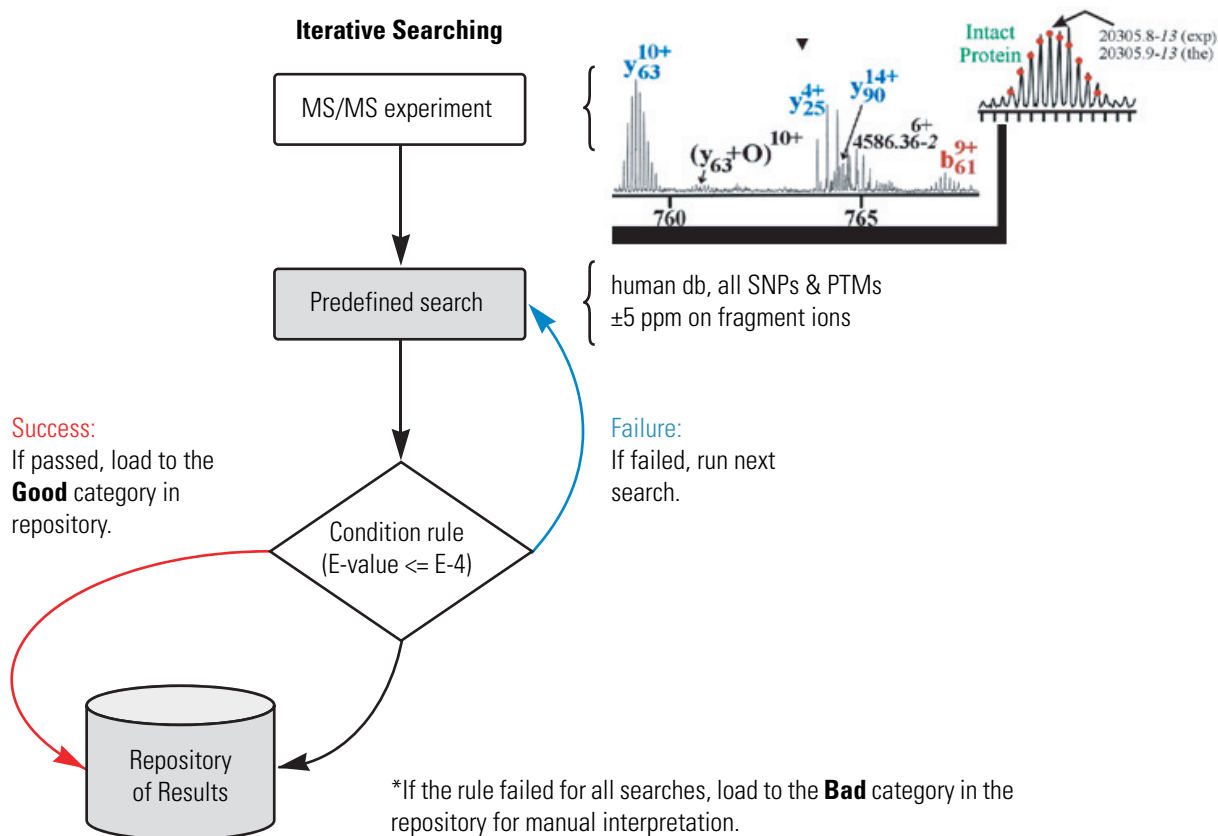
For a detailed explanation of ProSightPC's iterative search tree, see "Creating a Search Tree" on page 66.

ProSightPC's interface supports two levels of searching, but advanced users can define a search tree with unlimited levels by editing the .xml file that contains the search tree.

Database Manager

The Database Manager provides a point-and-click environment for managing the proteome warehouse and repositories. It imports and exports ProSightPC proteome warehouse (.pwf) files, enabling you to create your own proteome databases. The .pwf files are in a custom format that holds databases, patches, and repositories. For details on the functionality of the Database Manager, see Chapter 8, "Using Proteome Databases."

Data Manager

The Data Manager is part of ProSightPC's graphical user interface and provides a visual representation of all the information related to a single MS/MS experiment. Use it to view all information for a single experiment. You can use its context-sensitive controls to determine what information is displayed. For more information on the Data Manager, see Chapter 7, "Displaying Data in the Data Manager."

Sequence Gazer

The Sequence Gazer™ in ProSightPC's graphical user interface is an interactive environment for comparing MS/MS data to a known protein sequence. The Sequence Gazer characterizes previously identified proteins by selectively adding or removing PTMs or custom masses to amino acids in a protein sequence. Once you have made all your modifications to the amino acids, you can reevaluate the ion data. You use the Sequence Gazer to test hypotheses regarding which PTMs are present. You can also use it to fully characterize a protein. Chapter 6, "Searching for Single Proteins," explains how to perform single-protein searches by using the Sequence Gazer.

Experiment Manager

The Experiment Manager provides a simple interface for managing multiple MS/MS experiments in .puf files. For more information about the Experiment Manager, see Chapter 4, "Working with Experiments."

PTM Tier Editor

ProSightPC groups all PTMs into a multi-tier structure, enabling you to find and select PTMs quickly. Use the PTM Tier Editor to include or exclude PTMs and to view and change the tier assignment of PTMs. "Locating and Selecting PTMs with the PTM Tier Editor" on page 239 gives detailed information about the PTM Tier Editor.

Fragment Predictor

The Fragment Predictor takes a known protein sequence and returns all possible b, y, c, and Z$^\cdot$ fragment ion masses. You can use it to add PTMs or arbitrary custom masses to any amino acid in the protein sequence, and see the predicted fragment ion masses. The functionality of the Fragment Predictor is explained in "Viewing Fragments Ions with the Fragment Predictor" on page 242.

Noise Reducer

The Noise Reducer identifies and minimizes the effects of chemical noise in an MS/MS experiment. Once you have run the Noise Reducer, a new experiment identical to the source experiment is placed on the Data Grid. See "Reducing Chemical Noise with the Noise Reducer" on page 235 for more information on the Noise Reducer.

Font Converter

The Font Converter converts text into the ProSightPC fragment map font used to display N-terminal and C-terminal fragments. For more details on this feature, see "Converting Text to ProSightPC Font with the Font Converter" on page 245.

LC-MS/MS Workflow

Following are the general steps involved in using ProSightPC with LC-MS/MS data:

1. Load the proteome warehouse. ProSightPC uses shotgun annotation to apply sequence and PTM information to a proteome database in the proteome warehouse. This procedure is only performed once per proteome.

2. Import a .raw or .puf file into ProSightPC by using the High Throughput wizard.

3. If you imported a .raw file, ProSightPC performs an analysis to infer mass (AIM) to convert the data in the .raw data file to observed neutral mass values in a .puf file.

4. In the wizard, create a repository in which to store the search results.

5. In the wizard, define an iterative search tree.

6. Search for neutral mass data against the proteome warehouse. ProSightPC identifies and characterizes the observed proteins.

7. View the results in the user interface, generate a repository report, or both.

Figure 5 illustrates this flow.

Figure 5. ProSightPC wizard workflow
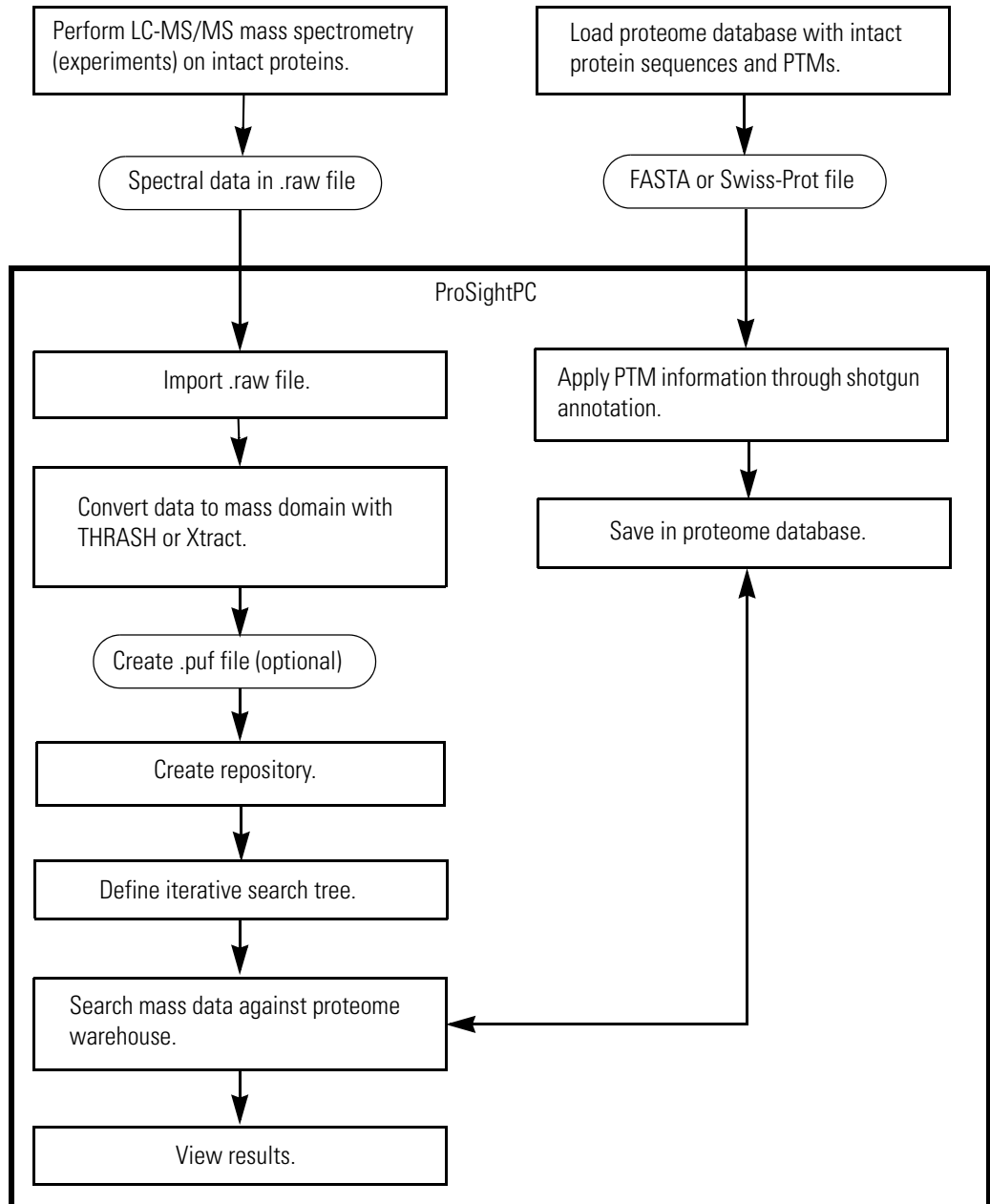
Inputs and Outputs

ProSightPC works with the following formats.

Inputs

ProSightPC works with three unique input file types:

- ProSightPC upload (.puf) files in XML format are used to store and transport ProSightPC results. Each .puf file can contain many MS/MS experiments, and each MS/MS experiment can contain searches. A single MS experiment can contain mass lists for both precursor and fragment ions extracted from the MS and MS/MS spectra.

  Only one .puf file can be open at a time. The active .puf file appears at the top of the Data Manager window. Each experiment is identified by a number that is unique in the .puf file.

- Proteome warehouse (.pwf) files are used to move databases and repositories from one computer to another. You can download prebuilt databases by choosing Databases > Download ProSightPC Databases in ProSightPC, or by going to ftp://prosightpc.scs.uiuc.edu. These databases include trypsin, Lys-C, and top-down databases for all the major model organisms. These databases are constructed and maintained quarterly in the Kelleher Laboratory and distributed by Thermo Fisher Scientific. You can also create your own .pwf files for exchanging proteome information.

- .raw files contain data that you must convert to the mass domain with the Xtract or THRASH algorithms by using the ProSightPC tools.

Outputs

As output, ProSightPC produces the following files:

- A .pwf file can contain any proteome databases and repositories that you have created and want to export to others.

- A .puf file contains experiments and searches.

Fragmentation Methods

ProSightPC supports the following fragmentation types:

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

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

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

- HCD: With the high-energy collision-induced dissociation (HCD) method of fragmentation, the projectile ion has laboratory-frame translation energy higher than 1 keV.

- ETD: With the electron transfer dissociation (ETD) method of fragmentation, singly charged reagent anions transfer an electron to multiply protonated peptides within an ion trap mass analyzer to induce fragmentation. ETD cleaves randomly along the peptide backbone while side chains and modifications such as phosphorylation are left intact. This method is used to fragment peptides and proteins.

Ion Types

ProSightPC supports both c/Z· and b/y ion types, which are shown in Figure 6.

Figure 6. c/Z· and b/y ion types
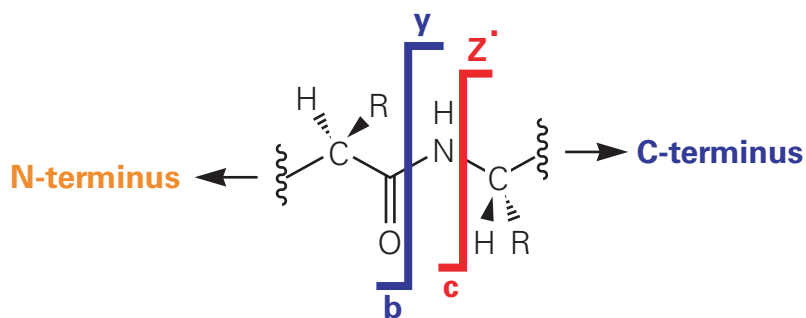
Introduction to Proteomics

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

Middle-Down/Bottom-Up Proteomics

Middle-down/bottom-up proteomics uses two methods to prepare the peptides for introduction into the mass analyzer, depending on the complexity of the protein mixture. For samples containing only a few proteins, you can separate the proteins by gel electrophoresis or chromatography. Enzymatic digestion breaks them down into smaller peptides with the aid of proteolytic agents such as trypsin or pepsin. For samples containing many different proteins, the proteins can be digested into peptides and then separated by chromatography and electrospray mass spectrometry (ESI-MS). These two methods are shown on the left in Figure 7. In either case, these peptides are then introduced to the mass analyzer.

Figure 7. Comparing top-down and bottom-up proteomics
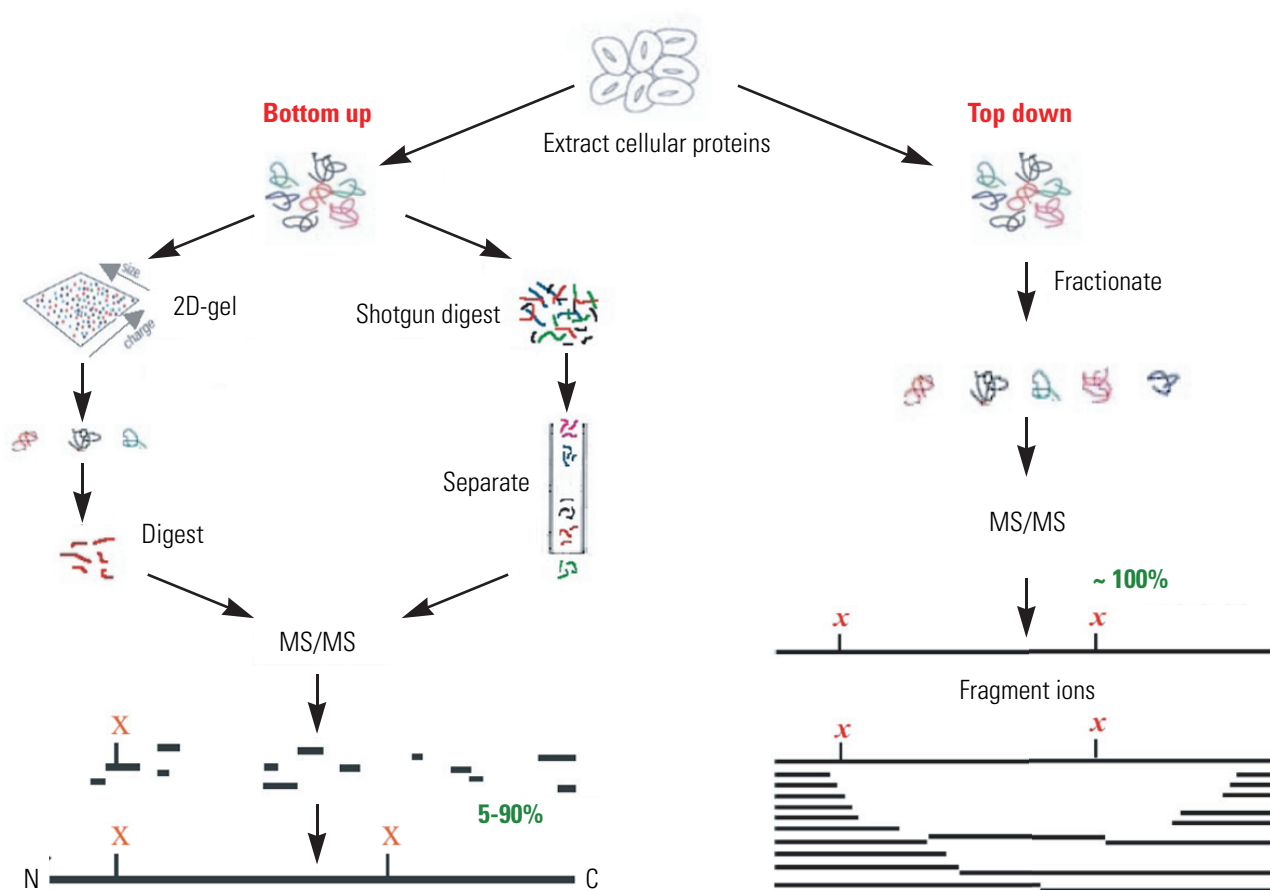
In top-down proteomics, electrospray ionization (ESI) or matrix-assisted laser desorption ionization (MALDI) ionize intact proteins. The proteins are then introduced into a mass analyzer, where they are subjected to gas-phase fragmentation. This methodology is shown on the right in Figure 7.

Top-Down Proteomics

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

Top-down proteomics is a rapidly emerging solution to the problem of protein identification and characterization. In contrast to other proteomic techniques, the unknown proteins in top-down proteomics are not digested into peptides before tandem mass spectrometry. This technique guarantees full sequence coverage on every protein, allowing easy characterization of complex combinations of PTMs.

Separation and ionization of intact proteins present many challenges beyond the scope of this manual. The bioinformatics of top-down proteomics, though, has 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 comparison with neutral masses. Once you collect MS and MS/MS spectra, you sum the relevant scans, and 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 search modes described in "Searching Databases" on page 125. Each search mode overcomes 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, which usually represents one charge state of the unknown protein but sometimes represents 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 parentheses before the amino acid containing the modification. For example, 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 an analysis process called shotgun annotation to take PTM events on a single protein and precalculate all possible combinations, regardless of whether the particular combination has ever been observed.

Shotgun annotation includes two components: constructing databases and searching databases.

Constructing Databases

Creating a shotgun-annotated database is based on the following principle. A given protein has *x* sites of modification—that is, sites where particular residues are observed or predicted to be modified in some way. A residue that can be modified is called a site. In a particular instance of a given protein, the modified sites are active and the unmodified sites are inactive. This instance is called a protein form.

Because you do not necessarily know which sites are simultaneously active in a living organism, within practical limitations, you want to precompute the masses and identities of all possible forms of a given protein, resulting in 2^n forms, where *n* is the number of sites, or modifications on the protein. For proteins with a limited number of sites, this growth rate is feasible; for proteins with a large number of sites, it is impractical to store all possible forms for highly modified proteins. The database creator uses an optimization algorithm to determine the most "important" sites. ProSightPC first determines if the protein is going to need more than 1000 records to fully describe it. If it does, ProSightPC prioritizes the protein forms, storing only those forms most likely to lead to protein identification. These forms are then shotgun-annotated. Shotgun annotation is therefore the process of generating potentially observable protein forms from the information known about a given protein.

For example, for a given protein that has only four phosphorylation sites and no other modifications, ProSightPC enters a record into the proteome database for the base sequence with no PTMs. It also enters the following:

- 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

ProSightPC processes all of these combinations, even if the phosphorylation events have only been observed separately.

If one of the multiphosphorylated forms occurs in nature and is observed in an MS/MS experiment, ProSightPC can readily identify it.

Searching Databases

You then search the resulting database. Any type of search eventually involves attempting to match observed masses to the masses that you precomputed and stored in your database (theoretical masses). Matches are not exact but are within a tolerance. The masses that you catch within the mass window—your observed mass plus or minus your tolerance—are then scored using various fragment-based scoring functions to determine your best match (see "Scores Box" on page 185 for detailed information about scoring methods).

The database searching component of shotgun annotation is illustrated in Figure 9. Every ball is a protein form that matches within a mass window.

Figure 8. Shotgun annotation search strategy

Legend:
- Correct Gene Family
- Correct Gene
- Observed Protein Form
- Correctly Predicted Protein Form
- Falsely Predicted Protein Form
- Protein Form
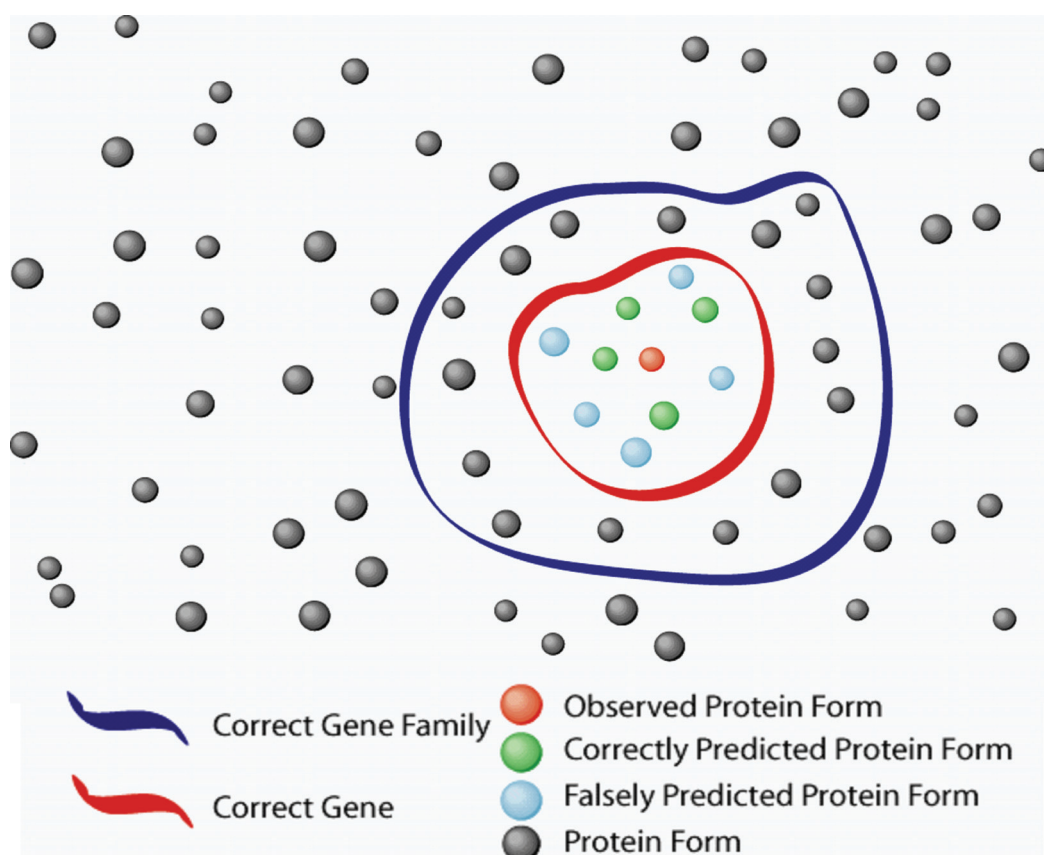
Most of these protein forms have negligible scores that you can ignore. They are represented by the gray balls outside any circle. The balls within the blue circle share a good number of fragments among the proteins encoded by a gene family, but their identification is at best partial. The balls in the concentric red circle have a better identification, because they match fragments that are unique to proteins encoded by a particular gene.

The blue balls are forms resulting from combinations of modifications that are abiological. They were created by the shotgun annotation algorithm, but they do not exist in living organisms. Usually you do not inherently know which protein forms in your database exist or do not exist in real life. The green balls are forms resulting from combinations of modifications that do, in fact, exist in living organisms—not the observed form, but real nonetheless. Finally, the red balls are the forms that you actually observed in your instrument.

Because you generate all possible protein forms in shotgun annotation, a large number are not going to exist in nature. You will also generate forms that do exist in nature. Shotgun annotation allows you to detect protein forms that you previously were not aware of or could not observe.

An example of a shotgun-annotated sequence is shown in Figure 9.

Figure 9. Shotgun-annotated sequence
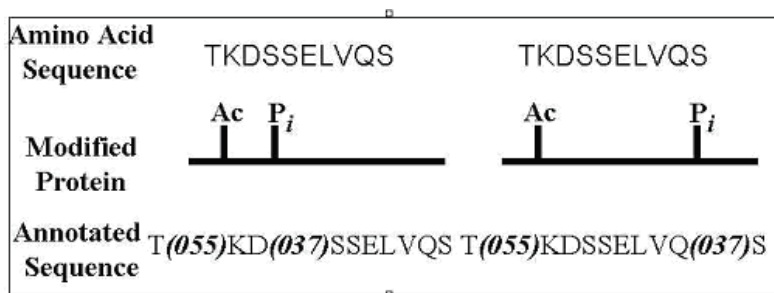
Search Modes and the Top-Down Funnel

In all but the most exceptional cases, top-down proteomics only receives partial fragment information in the MS/MS phase, so 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 10 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 identifying and partially characterizing the protein. In some cases, it is possible to only identify the protein.

Figure 10. Results of full characterization from initial data

Identified protein

Partially characterized protein

Identified and fully
characterized protein
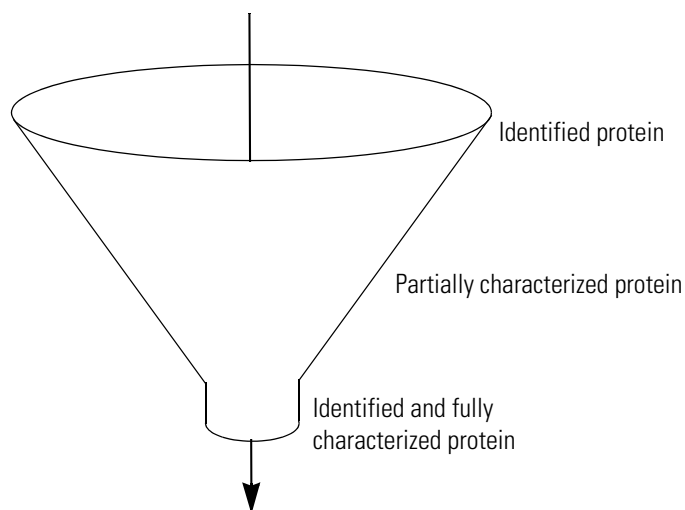
Note Partial characterization occurs in cases where it is possible to determine which PTM must be present on the protein, but the fragmentation data is not sufficient to determine on exactly which amino acid one or more of the PTMs must be. You can narrow the list of possibly modified amino acids to one or two residues in a short subsequence of the protein.

In some MS/MS experiments, you will have sufficient fragmentation data to fully characterize the proteins with the first search. If you shotgun-annotate the exact protein form observed into the proteome database and the MS spectra contains sufficient fragmentation information to uniquely identify this form, you can discover the correct answer by conducting an absolute mass search. This situation occurs frequently. In some well-annotated proteomes, unknown proteins are completely characterized on the first search, as shown in Figure 10.

Figure 11 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 is either a biomarker or a single protein mode search, though some search strategies conduct a gene-restricted search. These search modes are described in Chapter 5, "Searching Databases."

Figure 11. Multiple searches used for the identification and characterization of an unknown protein

Identified protein

Partially characterized protein

Identified and fully
characterized protein
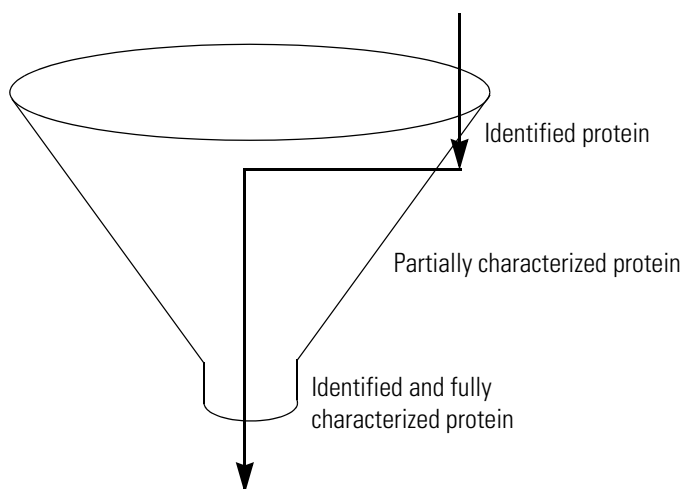
In Figure 12, the fragmentation data is insufficient to distinguish between two or more possible protein forms. In this case, full identification or partial characterization is the best possible result. When this occurs, re-run the MS/MS experiment to obtain better fragmentation data.

Figure 12. Results of identification, partial characterization, or both

Identified protein

Partially characterized protein

Identified and fully
characterized protein
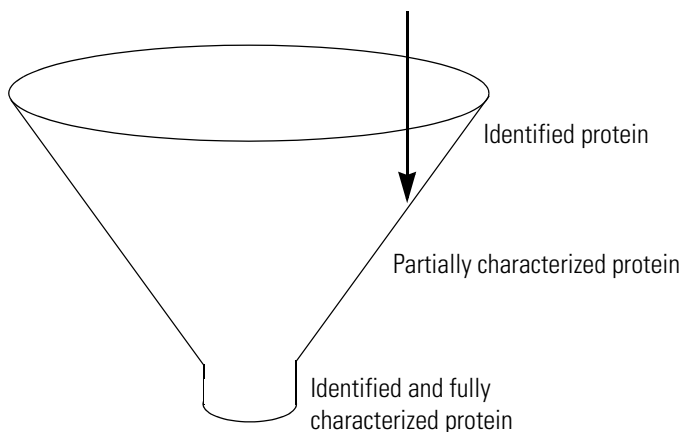
Using the ProSightPC Interface

ProSightPC has a unique approach to the organization of elements in the graphical user interface. This chapter describes the features of this interface.

Contents

- ProSightPC Interface
- Setting Default Options

ProSightPC Interface

Figure 13 shows some of the features of ProSightPC's graphical user interface.

Figure 13. ProSightPC window

Grid Display Preferences page   Menu bar   Toolbar   Data grid   Job queue   Data Manager   Tab Controller
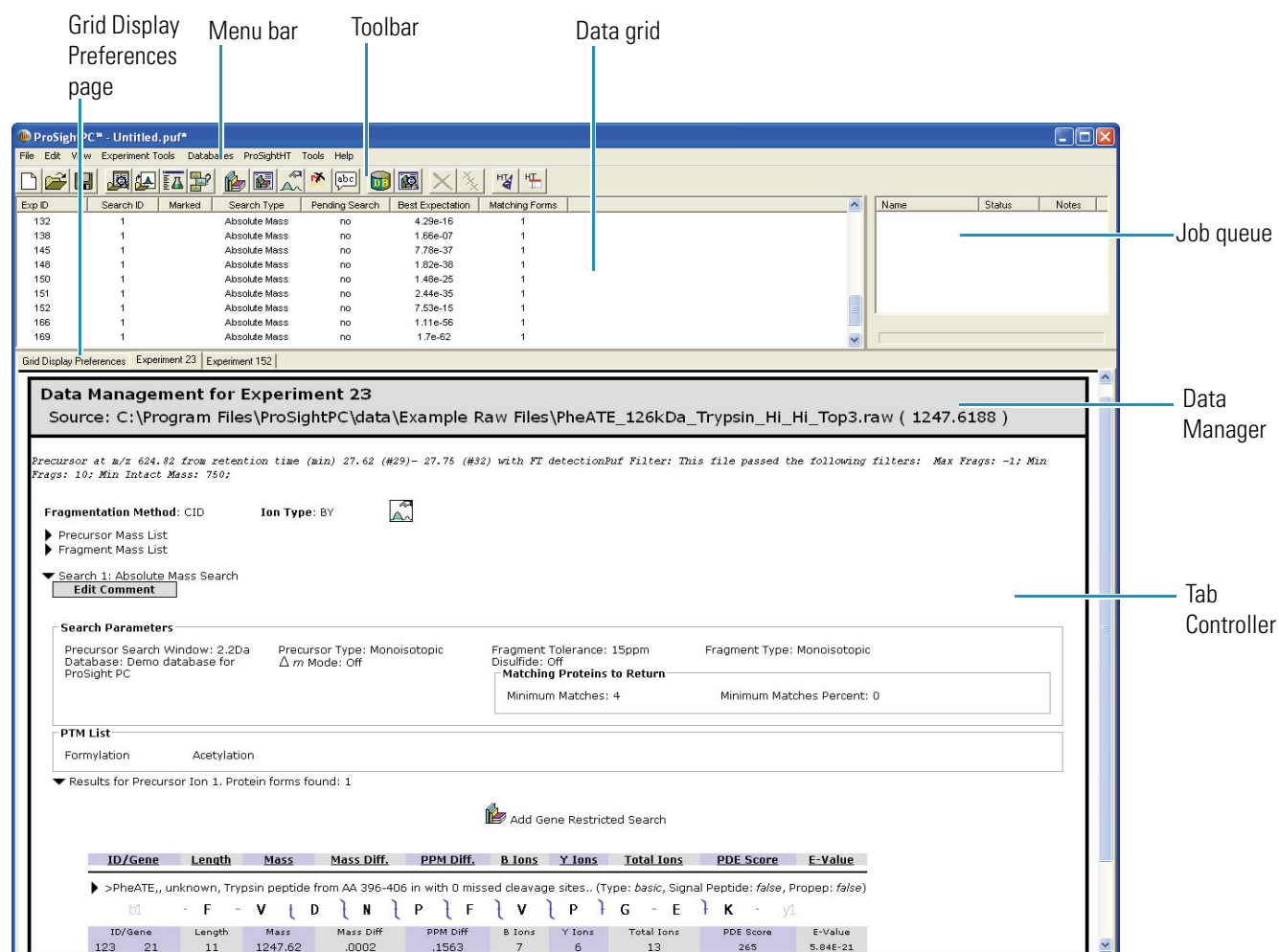
Menu Bar

The ProSightPC menu bar resides at the top of the ProSightPC graphical user interface. It contains the following menus. For detailed descriptions of each command on these menus, see Chapter 10, "ProSightPC Reference."

File	Use these commands to manipulate a .puf file, such as Open.
Edit	Use these commands to edit files, such as Copy and Paste.
View	Use these commands to display certain interface features, such as grid preferences, start screen, job queue, and toolbar.
Experiment Tools	Use these commands to perform operations on experiments, such as Add Search or Reduce Noise. This menu is only available when an experiment is open in the Data Manager.

Databases		Use these commands to handle proteome databases and repositories, import and export databases and repositories, create a custom database, and download pre-built databases.
ProSightHT		Use these commands to run the High Throughput wizard and edit and create repositories and search trees.
Tools		Use these commands to activate tools to process your data, such as Experiment Adder, PTM Tier Editor, and Individual Sequence Adder.
Help		Use these commands to view information about the current software release, manage licenses, and access the Help.

Toolbar

The ProSightPC toolbar, pictured in Figure 14, resides directly below the menu bar on the ProSightPC graphical user interface.

Figure 14. ProSightPC toolbar

Table 1 describes each of these icons.

Table 1. ProSightPC toolbar

Icon	Menu equivalent	Function
	File > New	Creates a new .puf file.
	File > Open	Opens an existing .puf file.
	File > Save	Saves a .puf file.
	File > Import .raw > Post Xtract	Builds a new experiment in the current .puf file by using Post XTRACT raw data.
	File > Import .raw > Profile	Builds a new experiment in the current .puf file by using high-resolution raw data obtained in profile mode.
	Tools > Experiment Adder	Builds a new experiment in the current .puf file by using manually input MS and MS/MS data.

Table 1. ProSightPC toolbar

Icon	Menu equivalent	Function
	View > Open Data Manager	Opens the last experiment using the open .puf file if there are no Experiment tabs open in the Tab Controller.
	Experiment Tools > Add Search	Adds a new search to the selected experiment. Available only when an experiment is open and shown in the Tab Controller.
	Experiment Tools > Append Predefined Search	Adds a predefined search to the selected experiment. This icon is available only when an experiment is open and shown in the Tab Controller.
	Experiment Tools > Edit Masses	Changes MS and MS/MS data in ProSightPC. This icon opens a new page showing the precursor and fragment masses of the current experiment. This icon is available only when an experiment is open and shown in the Tab Controller.
	Experiment Tools > Reduce Noise	Opens the Noise Reducer window. This icon is available only when an experiment is open and shown in the Tab Controller.
	Experiment Tools > Edit Comment	Enables you to edit the comment at the top of the current experiment.
	View Database Information	Opens the Database Manager, so that you can view information about the proteome databases in the proteome warehouse.
	Tools > Manage Predefined Searches	Enables you to modify the parameters of predefined searches.
	Abort Running Job	End the current search in the job queue. This icon is not available unless a job is running.
	Abort All Jobs	Ends all current and pending searches in the job queue. This icon is not available unless multiple jobs are running.
	ProSightHT > HighThroughput Wizard	Opens the High Throughput wizard so that you can start searching.
	ProSightHT > Repository Report	Opens the Repository Report dialog box so that you can generate a repository report.

❖ **To display the toolbar**

• Choose **View > Toolbar**.

Data Grid

The Data Grid displays summary information about each search in the open .puf file. It is organized into various columns, which you can change by using the Grid Display Preferences page, described in "Grid Display Preferences Page" on page 23.

Right-click a search in the Data Grid to access any of the following commands:

Parameter	Description
Refresh Grid	Redisplays the contents of the Data Grid.
Mark	Marks an experiment by placing the ProSightPC symbol to the left of the experiment and an asterisk (*) in the Marked column. This mark differentiates a particular experiment.
Add New Search	Opens the New Search in Experiment *X* dialog box so that you can add a new search to the experiment. For information on how to select options in this dialog box, see "Adding a Search" on page 126.
Append Predefined Search	Opens the Append Predefined Searches to Experiment *X* dialog box so that you can add a new predefined search to an experiment. For information on how to select options in this dialog box, see "Adding a Predefined Search to an Experiment" on page 169.
Edit Search	Opens the Edit Search in Experiment *X* dialog box for that type of search (this dialog box is the same as the New Search in Experiment *X* dialog box for that search type).
Edit Mass List	Opens a new page in the Tab Controller, showing the Precursor Mass List and the Fragment Mass List.
Run Search *x*	Runs a pending search or reruns a search after you have deleted the previous results of the search.
Remove Results	Removes the results of a specific search from an experiment in the Data Grid.
Remove Search *x*	Removes the specified search from the experiment.

Parameter	Description
Export Experiment to Repository	Replaces the specified experiment in the repository from which it was taken. This option is available only if the experiment was imported from the repository.
Remove Experiment *x*	Removes the specified experiment from the Data Grid.

Double-click a search in the Data Grid to open the relevant Data Manager. For more information about the Data Manager, see "Displaying Data in the Data Manager" on page 197.

"Working with Experiments" on page 117 explains how to import data into the Data Grid.

Job Queue

The job queue displays 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.

The job queue contains the following areas:

Area	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, the notes explain why the search failed.

Right-click the job queue pane to display the following commands:

Command	Description
Run	Performs a search.
Abort	Stops a search.
Clear Finished Jobs	Removes all jobs that have finished.

❖ **To display the job queue pane**

• Choose **View > Job Queue**.

❖ **To perform a search**

• If a search is not currently running, right-click and choose **Run** to add a search to the queue so that it runs next.

❖ **To cancel a search**

• If a search is running, right-click the search and choose **Abort** or click the **Abort Running Job** icon, ✕.

The search ends and the status changes to Failed.

• If you want to cancel several running searches, you can click the **Abort All Jobs** icon, ✕✕.

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 Data Manager in the Tab Controller.

• Right-click a page to hide the page.

• Right-click a page and choose **Close** to close the page, **Close All** to close all open pages, or **Close All But This** to close all open pages except the selected page.

• Right-click a page and choose **Refresh** to re-display the contents of the page.

For more information, see "Displaying Data in the Data Manager" on page 197.

Data Manager

The Data Manager appears in the ProSightPC interface when you double-click an experiment, choose View > Open Data Manager, or click the Open Data Manager icon, 🔲. "Displaying Data in the Data Manager" on page 197 describes the function of the Data Manager in detail.

Grid Display Preferences Page

Use the Grid Display Preferences page, shown in Figure 15, to select the type of information to display in the Data Grid. The Grid Display Preferences page automatically appears when you open a .puf file. Clicking View > Grid Preferences also displays this page. The Grid Display Preferences page consists of three areas: Show Columns, Quick Filters, and Custom Filters.

Figure 15. Grid Display Preferences page
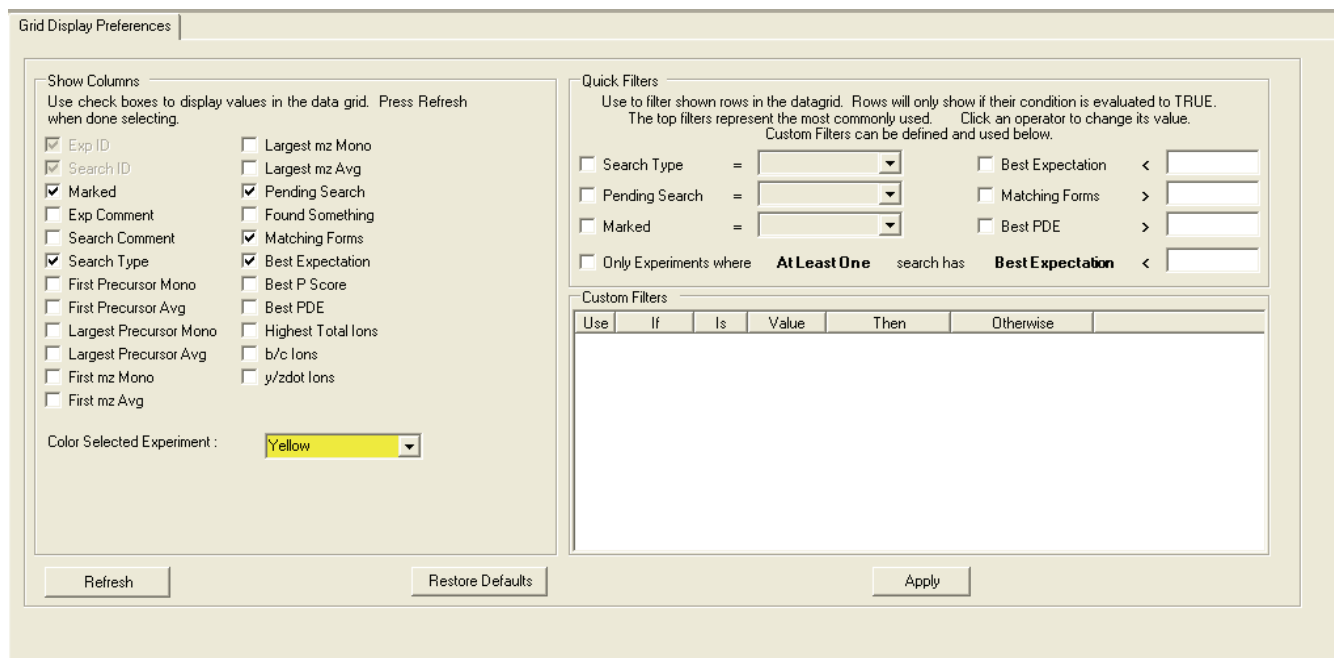
Show Columns Area

Use the Show Columns area to display or hide columns in the Data Grid. Each of the following options controls the appearance of a column in the Data Grid.

Parameter	Description
Exp ID	Displays a column showing the ProSightPC-assigned experiment number.
Search ID	Displays a column showing the ProSightPC-assigned search number.
Marked	Displays a column showing experiments marked by an asterisk. These experiments are also marked by a ProSightPC symbol to the left of the experiment.
Exp Comment	Displays a column showing a brief description of the experiment.
Search Comment	Displays a column showing a brief description of the search.
Search Type	Displays a column showing the type of search.
First Precursor Mono	Displays a column showing the monoisotopic mass of the first precursor ion.
First Precursor Avg	Displays a column showing the average mass of the first precursor ion.

Parameter	Description
Largest Precursor Mono	Displays a column showing the monoisotopic mass of the largest precursor ion.
Largest Precursor Avg	Displays a column showing the average mass of the largest precursor ion.
First mz Mono	Displays a column showing the monoisotopic mass-to-charge ratio (*m/z*) value of the first precursor ion for each experiment.
First mz Avg	Displays a column showing the average mass-to-charge ratio (*m/z*) value of the first precursor ion for each experiment.
Largest mz Mono	Displays a column showing the largest monoisotopic mass-to-charge ratio (*m/z*) value of all precursor entries for each experiment.
Largest mz Avg	Displays a column showing the largest average mass-to-charge ratio (*m/z*) value of all precursor entries for each experiment.
Pending Search	Displays a column indicating whether a search has been performed.
Found Something	Displays a column indicating whether a match in the database was found.
Matching Forms	Displays a column showing the number of matching forms.
Best Expectation	Displays a column with the best (lowest) expectation score of any match in the search results.
Best P Score	Displays a column with the best (lowest) p score of any match in the search results. For more information on the calculation of this score, see "p Score" on page 186.
Best PDE	Displays a column showing the best (highest) PDE (McLuckey) score of any match in the search results. For more information on the calculation of this score, see "PDE (McLuckey) Score" on page 188.
Highest Total Ions	Displays a column showing the highest total number of ions that matched the ions in the database.
b/c Ions	Displays a column showing the number of b and c fragment ions that matched in the database.

Parameter	Description
y/Z$^{\cdot}$ Ions	Displays a column showing the number of y and $^{Z\cdot}$ fragment ions that matched in the database.
Color Selected Experiment	Displays a column showing the color chosen for selected experiments.

❖ **To modify the defaults of the options in the Show Columns area**

• Click the Grid Display Preferences page or choose **Tools > Options > Grid Columns**.

❖ **To add a column to the Data Grid**

1. Select the check box for any columns that you want to display in the Data Grid.

2. Click **Refresh** to view your display.

   The columns that you selected appear in the Data Grid.

❖ **To remove a column from the Data Grid**

1. Clear the check box for any of the columns that you want to hide in the Data Grid.

2. Click **Refresh** to hide the columns.

3. Click **Restore Defaults** to reinstate the default settings. You can change the default settings with the Options dialog box. For information on this procedure, see "Setting Column Display Preferences" on page 34.

Quick Filters Area

You can use the Quick Filters area to quickly define conditions with which to filter the experiments and searches displayed in the Data Grid. Check one or more of the criteria to filter (hide) certain Data Grid rows. Click an operator to change its value.

Here are the quick filters that you can apply:

Filter	Description
Search Type	Filters the search by search type: • Absolute mass • Biomarker • GRAM (gene-restricted absolute mass) • GRBM (gene-restricted biomarker) • Sequence tag • Single protein
Pending Search	Filters the search by whether a search is pending or not: • Yes: A search is pending. • No: A search is not pending.
Marked	Filters the search by whether a search is marked or not: • Yes: A search is marked. • No: A search is not marked.
Best Expectation	Filters the search by expectation value (e value). For information about the expectation value, see "Expectation Value (e value)" on page 186.
Matching Forms	Filters the search by the number of matching forms.

Filter	Description
Best PDE	Filters the search by PDE (McLuckey) score. For information about this scoring method, see "PDE (McLuckey) Score" on page 188.
Only Experiments where *number* search has *option operator value*	Filters the search by experiments that meet the conditions set. *Number* can be • At Least One: Displays at least one search meeting the criteria. • All: Displays all searches meeting the criteria. • No: Does not display any of the searches meeting the criteria. *Option* can be • Best Expectation: Expectation value (e value). For information about the expectation value, see "Expectation Value (e value)" on page 186. • Best PDE: McLuckey score. For more information on the calculation of this score, see "PDE (McLuckey) Score" on page 188. • Best P Score: P score. For more information on the calculation of this score, see "p Score" on page 186. • Total Ions: Highest total number of ions that matched the ions in the database. • Matching Forms: Number of matching forms. *Operator* can be • = Equal to • NOT = Not equal to • < Less than • > Greater than • <= Less than or equal to • >= Greater than or equal to

❖ **To access the Quick Filters options**

• Click the **Grid Display Preferences** page.

❖ **To define quick filters for a search**

1. (Optional) In the Quick Filters area of the Grid Display Preferences page, select the Search Type, Pending Search, or Marked box, click any operators to set it, and select an option from the list to the right of the option.

2. (Optional) Select the Best Expectation, Total Fragments, or Best PDE box, click any operators to set it, and enter the appropriate value in the box to the right of the option.

3. (Optional) Select the **Only Experiments where** *number* **search has** *option operator value* box. Click *number*, *option*, and *operator* to display the choices available.

   The previous table in this section explains these choices.

4. When you have set all the filters that you want, click **Apply**.

❖ **To remove quick filters**

- Clear the box next to the filter that you want to remove.

Custom Filters Area

You can use the Custom Filters section of the Grid Display Preferences page to define completely custom conditions with which to filter the searches displayed in the Data Grid. Check one or more of the criteria to filter (hide) certain Data Grid rows. Click an operator to change its value.

❖ **To access the Custom Filters options**

- Click the **Grid Display Preferences** page.

❖ **To define custom filters for a search**

1. Right-click Custom Filters and choose **New** from the shortcut menu.

   The appearance of the Custom Filters area changes, as shown in Figure 16.

   Figure 16. Custom Filters area of the Grid Display Preferences page
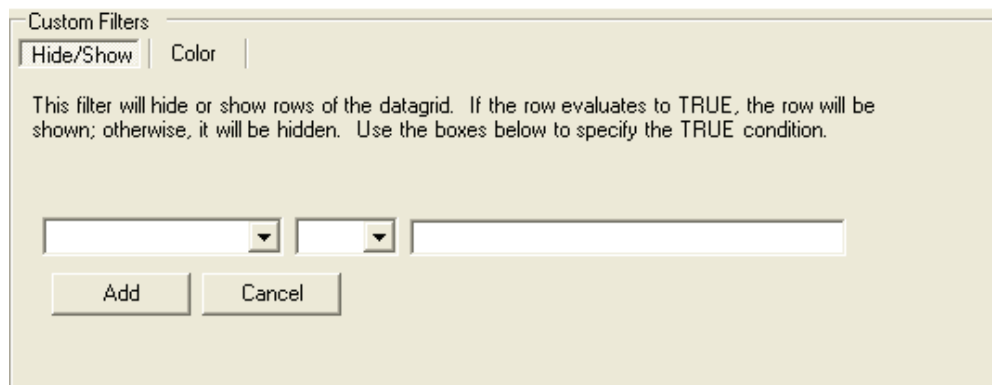
2. From the leftmost list, select a filter.

   For example, ExpID = 5 displays only the experiment whose identifier is 5.

   The parameters available in this list are the same as those given in "Show Columns Area" on page 24.

3. From the list in the center, select an operator.

The operators available in this list are the same as those given for the "Only Experiments where *number* search has *option operator value*" parameter in "Quick Filters Area" on page 26.

4. In the right box, specify an appropriate value. Here are some examples:

   largest precursor mono > 1000

   b/c ions > 7

   y/Z· ions < 20

5. To indicate that a search should be displayed in the Data Grid if it meets the specified condition, click **Show**. To indicate that a search that does not meet the specified condition should not be displayed, click **Hide**.

6. If you want to apply colors to specific columns on the basis of the condition that you just set, click **Color**.

   Two new boxes with drop-down lists now appear, as shown in Figure 17.

   Figure 17. Color selection lists
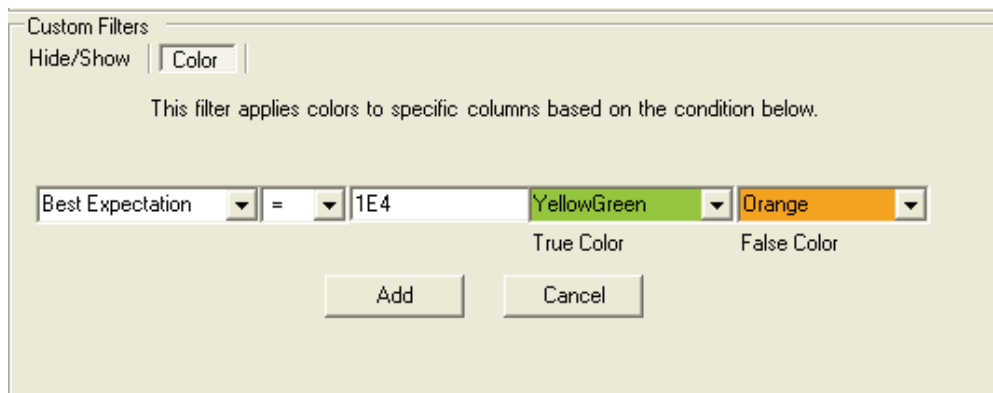
7. In the True Color list, select a color for columns containing data that meets your condition.

8. In the False Color list, select a color for columns containing data that does not meet your condition.

9. Click **Add**.

   The filter now appears with a small check box to the left, as shown in Figure 18.

Figure 18. Filter added to list of custom filters
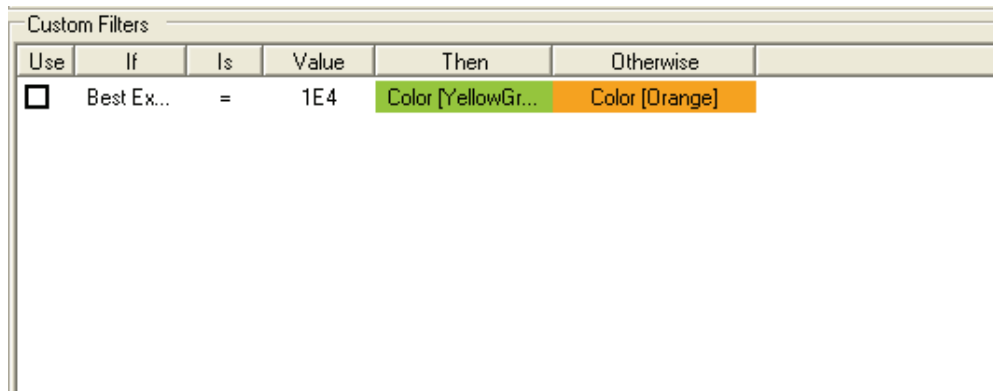

10. To apply the filter, select the check box next to the filter, and click **Apply**.

❖ **To remove custom filters**

1. Click a filter to highlight it.

2. Right-click and choose **Remove** from the shortcut menu.

Setting Default Options

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

You can set the following preferences in the Options dialog box:

- General: See "Setting General Preferences."

- Grid Columns: See "Setting Column Display Preferences" on page 34.

- THRASH: See "Setting THRASH Preferences" on page 37.

- Search Parameters: See "Setting Search Parameter Preferences" on page 39.

❖ **To access the Options dialog box**

- Choose **Tools > Options** to open the Options dialog box, shown in Figure 19.

Setting General Preferences

You can use the General Preferences page of the Options dialog box to set default values for options that are available throughout much of the ProSightPC interface.

❖ **To set general preferences**

1. From the left pane of the Options dialog box, click the **General** folder to open the General Preferences page, shown in Figure 19.

Figure 19. General Preferences page of the Options dialog box
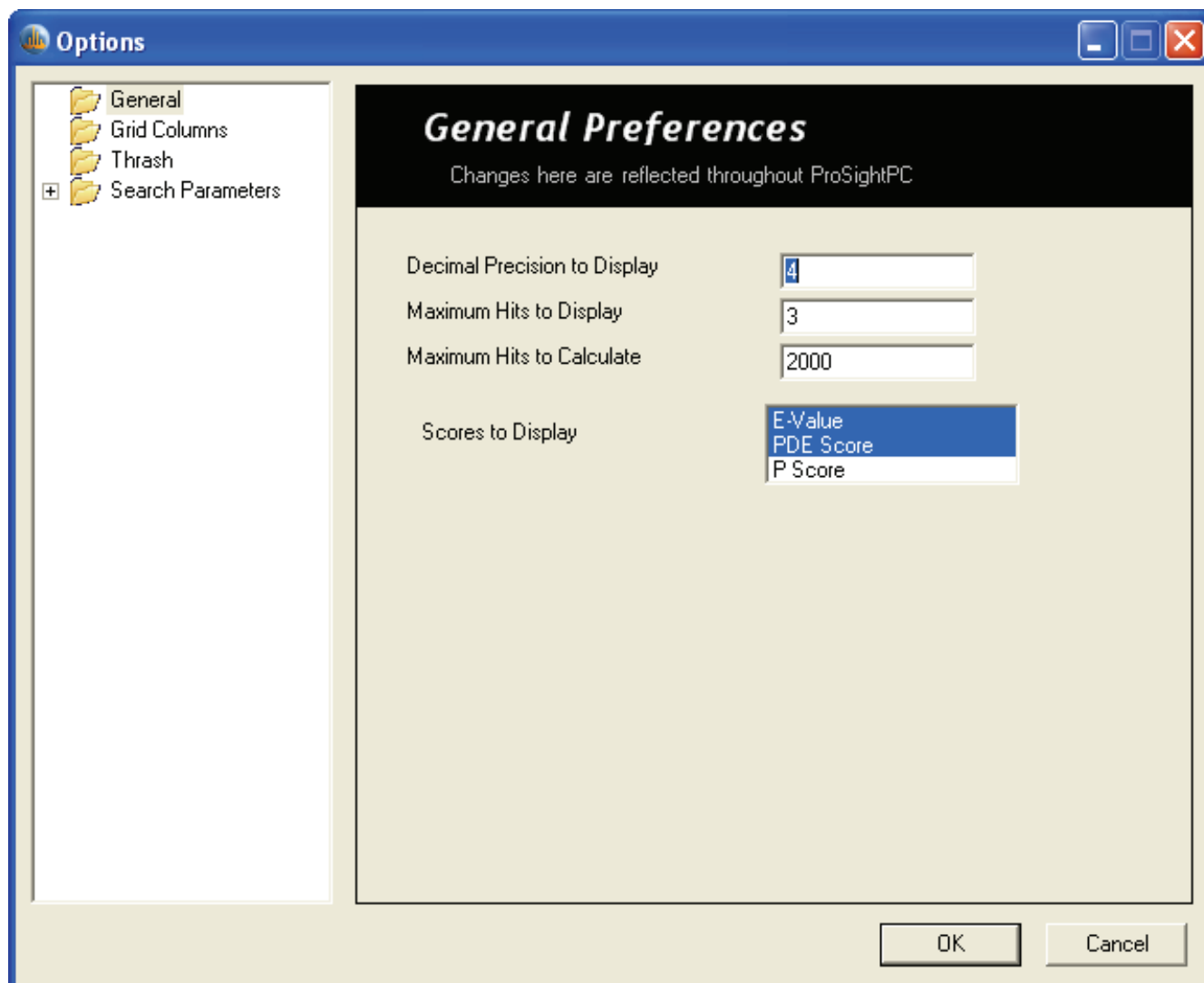
2. In the Decimal Precision to Display box, specify the number of decimal places to which most numbers are displayed.

3. In the Maximum Hits to Display box, specify the number of matching fragment tables to be displayed in the Data Manager.

   Only the best matches up to this number are displayed.

4. In the Maximum Hits to Calculate box, specify the maximum number of protein forms that a given search considers.

   When this number is exceeded, the search automatically stops, and ProSightPC issues a warning.

5. In the Scores to Display box, specify the types of scores to display in the statistics table in the Data Manager. You can choose from the following options:

E-Value	Displays the expectation value (e value), which is the number of sequences in a database that are expected to have p scores equal to or better than what was observed simply by chance. For more information on the e value, see "Expectation Value (e value)" on page 186.
PDE Score	Displays the PDE, or McLuckey, score, which is a way of scoring how well a set of observed fragment ions matches a protein. For more information on the PDE score, see "PDE (McLuckey) Score" on page 188.
P Score	Displays the p score, which is the probability of obtaining at least as good a match between the observed fragment list and a sequence as by chance. For more information on the p score, see "p Score" on page 186.

6. Click **OK**.

General Preferences Page Parameters

The General Preferences page of the Options dialog box contains the following parameters:

Parameter	Description
Decimal Precision to Display	Specifies the number of decimal places to which most numbers are displayed.
Maximum Hits to Display	Specifies the number of matching fragment tables that are displayed in the Data Manager.

Parameter	Description
Maximum Hits to Calculate	Specifies the maximum number of protein forms that a given search considers.
Scores to Display	Specifies the types of scores to display in the statistics table in the Data Manager. You can select from the following options:

    • E-Value: Displays the expectation value (e value), which is the number of sequences in a database that are expected to have p scores equal or better than what was observed simply by chance. See "Expectation Value (e value)" on page 186 for information on this score.

    • PDE Score: Displays the PDE, or McLuckey, score, which is a way of scoring how well a set of observed fragment ions matches a protein. See "PDE (McLuckey) Score" on page 188 for information on this score.

    • P Score: Displays the p score, which is the probability of obtaining at least as good a match between the observed fragment list and a sequence as by chance. See "p Score" on page 186 for information on this score.

For more information about setting general preferences, see "Setting General Preferences" on page 31.

Setting Column Display Preferences

You can use the Grid Columns page of the Options dialog box to set the defaults for the options in the Grid Display Preferences page.

❖ **To set default column display preferences**

1. From the left pane of the Options dialog box, click the **Grid Columns** folder.

    The Grid Columns page of the Options dialog box appears, as shown in Figure 20.

Figure 20. Grid Columns page of the Options dialog box
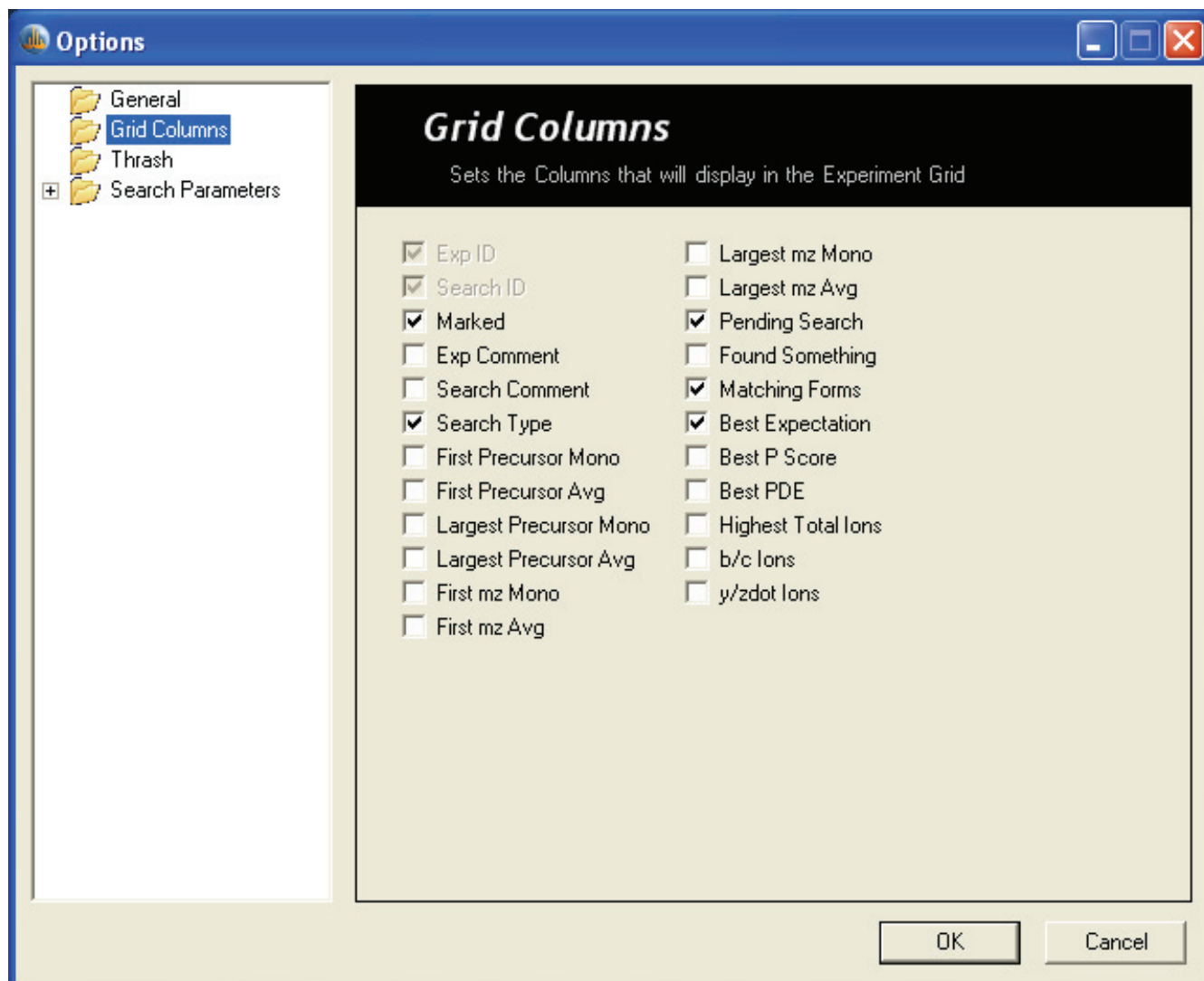
2. Select the check boxes of the column names to be displayed in the experiment grid. You can select from the column names shown in "Show Columns Area" on page 24.

3. Click **OK**.

Grid Columns Page Parameters

The Grid Columns page of the Options dialog box contains the following parameters:

Parameter	Description
Exp ID	Displays a column showing the ProSightPC-assigned experiment number.
Search ID	Displays a column showing the ProSightPC-assigned search number.

Parameter	Description
Marked	Displays a column showing experiments marked by an asterisk. These experiments are also marked by a ProSightPC symbol to the left of the experiment.
Exp Comment	Displays a column showing a brief description of the experiment.
Search Comment	Displays a column showing a brief description of the search.
Search Type	Displays a column showing the type of search.
First Precursor Mono	Displays a column showing the monoisotopic mass of the first precursor ion.
First Precursor Avg	Displays a column showing the average mass of the first precursor ion.
Largest Precursor Mono	Displays a column showing the monoisotopic mass of the largest precursor ion.
Largest Precursor Avg	Displays a column showing the average mass of the largest precursor ion.
First mz Mono	Displays a column showing the monoisotopic mass-to-charge ratio (m/z) value of the first precursor ion for each experiment.
First mz Avg	Displays a column showing the average mass-to-charge ratio (m/z) value of the first precursor ion for each experiment.
Largest mz Mono	Displays a column showing the largest monoisotopic mass-to-charge ratio (m/z) value of all precursor entries for each experiment.
Largest mz Avg	Displays a column showing the largest average mass-to-charge ratio (m/z) value of all precursor entries for each experiment.
Pending Search	Displays a column indicating whether a search has been run.
Found Something	Displays a column indicating whether a match in the database was found.
Matching Forms	Displays a column showing the number of matching forms.
Best Expectation	Displays a column with the best (lowest) expectation score of any match in the search results. See "Expectation Value (e value)" on page 186 for information on this score.

| Parameter	Description
Best P Score	Displays a column with the best (lowest) p score of any match in the search results. See "p Score" on page 186 for more information on the calculation of this score.
Best PDE	Displays a column showing the best (highest) PDE (McLuckey) score of any match in the search results. See "PDE (McLuckey) Score" on page 188 for information on the calculation of this score.
Highest Total Ions	Displays a column showing the highest total number of ions that matched the ions in the database.
b/c Ions	Displays a column showing the number of b and c fragment ions that matched in the database.
y/Z. Ions	Displays a column showing the number of y and Z· fragment ions that matched in the database.
Color Selected Experiment	Displays a column showing the color chosen for selected experiments.

Setting THRASH Preferences

You can use the Thrash Preferences page of the Options dialog box to set the default values for the THRASH algorithm, which is an analysis-to-infer-mass (AIM) operation that converts high-resolution mass spectral data from proteins or large peptides into neutral monoisotopic or average masses.

❖ **To set THRASH preferences**

1. From the left pane of the Options dialog box, click the **Thrash** folder.

   The Thrash Preferences page opens, as shown in Figure 21.

Figure 21. Thrash Preferences page of the Options dialog box
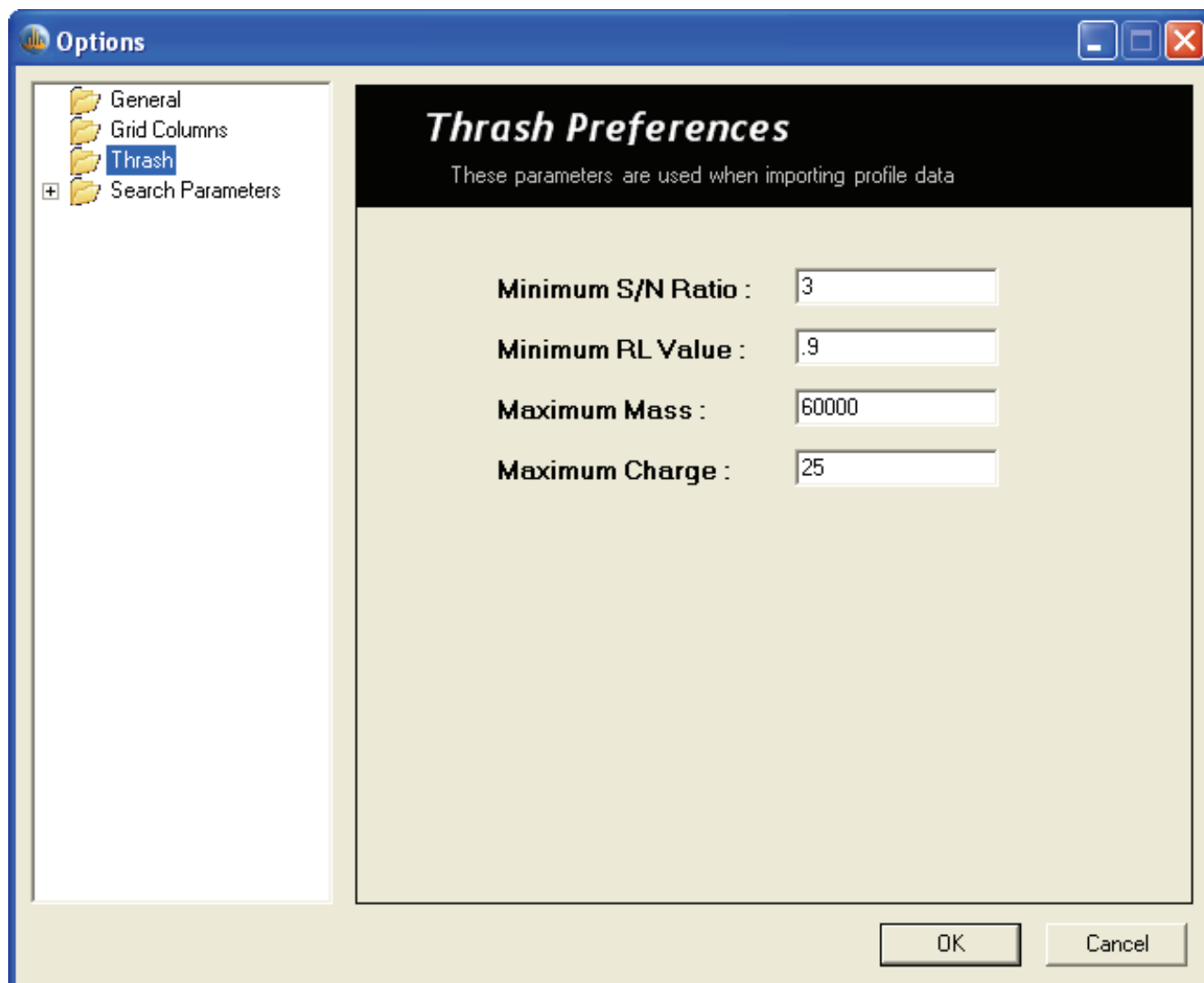
2. In the Minimum S/N Ratio box, enter the lowest signal-to-noise ratio that the THRASH algorithm will consider when trying to assign neutral mass to the data in your .raw files.

3. In the Minimum RL Value box, enter the minimum confidence level.

4. In the Maximum Mass box, enter the cutoff point for the THRASH algorithm when searching for masses.

5. In the Maximum Charge box, enter the maximum charge to be used by the THRASH algorithm.

6. Click **OK**.

Thrash Preferences Page Parameters

The Thrash Preferences page of the Options dialog box contains the following parameters:

Parameter	Description
Minimum S/N Ratio	Specifies the lowest signal-to-noise ratio that the THRASH algorithm considers when trying to assign neutral mass to the data in your .raw files.
Minimum RL Value	Specifies the minimum confidence level.
Maximum Mass	Specifies the cutoff point for the THRASH algorithm when searching for masses.
Maximum Charge	Specifies the maximum charge to be used by the THRASH algorithm.

Setting Search Parameter Preferences

You can use the Search Parameters pages of the Options dialog box to set defaults for each search type.

❖ **To set search parameter preferences**

1. From the left pane of the Options dialog box, click the **+** sign to the left of the **Search Parameters** folder.

2. Click one of the following options:

   • Absolute Mass: See "Setting Absolute Mass Search Preferences."

   • Biomarker: See "Setting Biomarker Search Preferences" on page 43.

   • Sequence Tag: See "Setting Sequence Tag Search Preferences" on page 47.

   • Single Protein: See "Setting Single-Protein Search Preferences" on page 51.
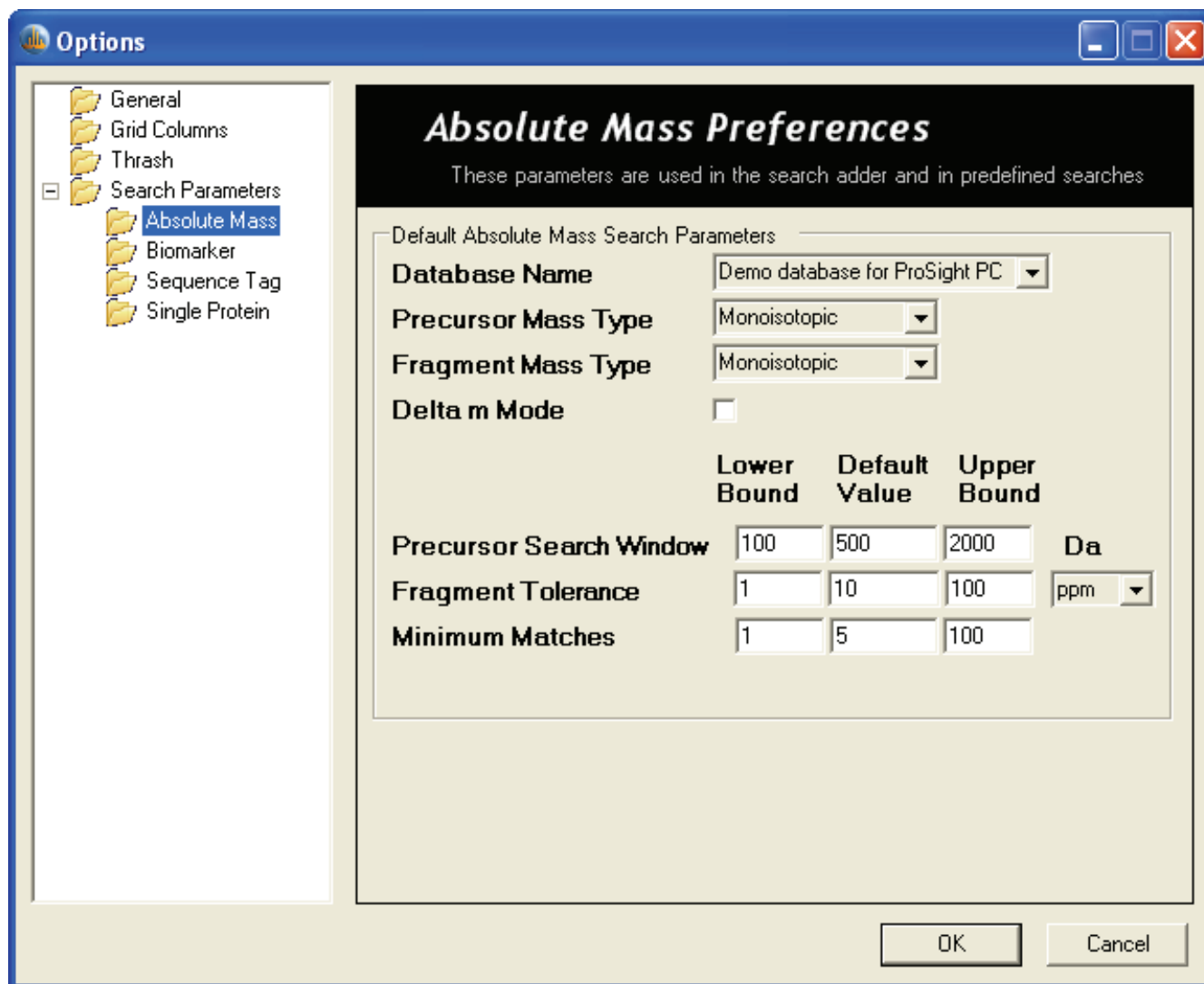
Setting Absolute Mass Search Preferences

Use the Absolute Mass Preferences page of the Options dialog box to set the default values used when you add new absolute mass searches. For information about absolute mass searches, see "Searching for Absolute Mass" on page 130.

❖ **To set absolute mass search preferences**

1. Click **Absolute Mass** in the Options dialog box.

   The Absolute Mass Preferences page of the dialog box opens, as shown in Figure 22.

Figure 22. Absolute Mass Preferences page of the Options dialog box

2. In the Database Name list, select the name of the database to search.

3. In the Precursor Mass Type list, select the type of precursor mass:

   • Monoisotopic: Specifies that the precursor mass is monoisotopic, which is the mass of the protein, peptide, or fragment ion, where all carbons are carbon-12.

   • Average: Specifies that the precursor mass is the mass of the most abundant isotope of the protein, peptide, or fragment ion.

4. In the Fragment Mass Type list, select the type of fragment mass:

    • Monoisotopic: Specifies that the fragment mass is monoisotopic, which is the mass of the protein, peptide, or fragment ion, where all carbons are carbon-12.

    • Average: Specifies that the fragment mass is the mass of the most abundant isotope of the protein, peptide, or fragment ion.

5. (Optional) Select the Delta m Mode box if you want to conduct the search in delta (Δm) mode, which is explained in "Performing a Search in Delta m Mode" on page 170.

6. In the Precursor Search Window boxes, specify the dimensions of the precursor search window of the observed intact ion mass in the selected units. For intact ion masses, the dimensions are always in daltons, but for fragments they can be in daltons or parts per million. Set the following parameters:

    • Lower Bound: Sets the minimum value for a precursor search window that does not trigger an "out of range" warning, which will be displayed as yellow background in the text box.

    • Default Value: Sets the default value for a precursor search window.

    • Upper Bound: Sets the maximum value for a precursor search window that does not trigger an "out of range" warning.

7. In the Fragment Tolerance boxes, specify the tolerance that determines whether the comparison of an observed fragment ion mass to a theoretical fragment ion mass is considered a match. Set the following parameters:

    • 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 a fragment tolerance.

    • Upper Bound: Sets the maximum value for a fragment tolerance that does not trigger an "out of range" warning.

    The tolerance can be absolute (set in daltons [Da]) or relative (set in parts per million [ppm]).

    A fragment tolerance is a mass value, either absolute or relative, within which your observed masses must match the theoretical fragment mass. For instance, if you set your tolerance to 0.005 Da (an absolute tolerance) and your theoretical fragment ion is at 1154.1126 Da, observed fragment ions of 1154.1090 Da (-0.0034 Da from theoretical) and 1154.1167 (+0.0041 Da from theoretical) fall within the tolerance, but 1154.2312 (+0.1222) does not, because the mass difference is greater than the tolerance that you set.

8. In the Minimum Matches boxes, specify the minimum number of matching ion fragments that you want the search algorithm to find in proteins. Set the following parameters.

- 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.

9. Click **OK**.

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

Absolute Mass Preferences Page Parameters

The Absolute Mass Preferences page of the Options dialog box contains the following parameters:

Parameter	Description
Database Name	Specifies the name of the database to search.
Precursor Mass Type	Specifies the type of precursor mass: • Monoisotopic: Specifies that the precursor mass is monoisotopic, which is the mass of the protein, peptide, or fragment ion, where all carbons are carbon-12. • Average: Specifies that the precursor mass is the mass of the most abundant isotope of the protein, peptide, or fragment ion.
Fragment Mass Type	Specifies the type of fragment mass: • Monoisotopic: Specifies that the fragment mass is monoisotopic, which is the mass of the protein, peptide, or fragment ion, where all carbons are carbon-12. • Average: Specifies that the fragment mass is the mass of the most abundant isotope of the protein, peptide, or fragment ion.
Delta m Mode	Determines whether ProSightPC conducts the search in delta (Δm) mode, which is explained in "Performing a Search in Delta m Mode" on page 170.

Parameter	Description
Precursor Search Window	Specifies the dimensions of the precursor search window of the observed intact ion mass in the selected units. For intact ion masses, the dimensions are always in daltons, but for fragments, they can be in daltons or parts per million. Set the following parameters: • Lower Bound: Sets the minimum value for a precursor search window that does not trigger an "out of range" warning. • Default Value: Sets the default value for a precursor search window. • Upper Bound: Sets the maximum value for a precursor search window that does not trigger an "out of range" warning.
Fragment Tolerance	Specifies the tolerance that determines whether the comparison of an observed fragment ion mass to a theoretical fragment ion mass is considered a match. Set the following parameters: • 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 a fragment tolerance. • Upper Bound: Sets the maximum value for a fragment tolerance that does not trigger an "out of range" warning. The tolerance can be absolute (set in daltons [Da]) or relative (set in parts per million [ppm]).
Minimum Matches	Specifies the minimum number of matching ion fragments that you want the search algorithm to find in proteins. Set the following parameters: • 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.

Setting Biomarker Search Preferences

When adding new biomarker searches, you can set the default values on the Biomarker Preferences page of the Options dialog box. For information on biomarker searches, see "Searching for a Biomarker" on page 138.

❖ **To set biomarker search preferences**

1. Click **Biomarker** in the Options dialog box.

   The Biomarker Preferences page of the Options dialog box opens, as shown in Figure 23.

   Figure 23.   Biomarker Preferences page of the Options dialog box

2. In the Database Name list, select the name of the database to search.

3. In the Precursor Mass Type list, select the type of precursor mass:

   • Monoisotopic: Specifies that the precursor mass is monoisotopic, which is the mass of the protein, peptide, or fragment ion, where all carbons are carbon-12.

   • Average: Specifies that the precursor mass is the mass of the most abundant isotope of the protein, peptide, or fragment ion.

4. In the Fragment Mass Type list, select the type of fragment mass:

   - Monoisotopic: Specifies that the fragment mass is monoisotopic, which is the mass of the protein, peptide, or fragment ion, where all carbons are carbon-12.

   - Average: Specifies that the fragment mass is the mass of the most abundant isotope of the protein, peptide, or fragment ion.

5. (Optional) Select the Delta m Mode box if you want to conduct the search in delta (Δm) mode, which is explained in "Performing a Search in Delta m Mode" on page 170.

6. In the Precursor Tolerance boxes, specify the tolerance that determines whether the comparison of an observed precursor ion mass to a theoretical precursor ion mass is considered a match. Set the following parameters:

   - Lower Bound: Sets the minimum value for a precursor search window that does not trigger an "out of range" warning.

   - Default Value: Sets the default value for a precursor search window.

   - Upper Bound: Sets the maximum value for a precursor search window that does not trigger an "out of range" warning.

7. In the Fragment Tolerance boxes, specify the tolerance that determines whether the comparison of an observed fragment ion mass to a theoretical fragment ion mass is considered a match. Set the following parameters:

   - 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 a fragment tolerance.

   - Upper Bound: Sets the maximum value for a fragment tolerance that does not trigger an "out of range" warning.

   The tolerance can be absolute (set in daltons [Da]), or relative (set in parts per million [ppm]).

   A fragment tolerance is a mass value, either absolute or relative, within which your observed masses must match the theoretical fragment mass. For instance, if you set your tolerance to 0.005 Da (an absolute tolerance) and your theoretical fragment ion is at 1154.1126 Da, observed fragment ions of 1154.1090 Da (-0.0034 Da from theoretical) and 1154.1167 (+0.0041 Da from theoretical) fall within the tolerance, but 1154.2312 (+0.1222) does not, because the mass difference is greater than the tolerance that you set.

8. In the Minimum Matches boxes, specify the minimum number of matching ion fragments that you want the search algorithm to find in proteins. Set the following parameters:

   - 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.

9. Click **OK**.

> **Note** Gene-restricted biomarker searches draw their default parameters from biomarker searches.

Biomarker Preferences Page Parameters

The Biomarker Preferences page of the Options dialog box contains the following parameters:

Parameter	Description
Database Name	Specifies the name of the database to search
Precursor Mass Type	Specifies the type of precursor mass: • Monoisotopic: Specifies that the precursor mass is monoisotopic, which is the mass of the protein, peptide, or fragment ion, where all carbons are carbon-12. • Average: Specifies that the precursor mass is the mass of the most abundant isotope of the protein, peptide, or fragment ion.
Fragment Mass Type	Specifies the type of fragment mass: • Monoisotopic: Specifies that the fragment mass is monoisotopic, which is the mass of the protein, peptide, or fragment ion, where all carbons are carbon-12. • Average: Specifies that the fragment mass is the mass of the most abundant isotope of the protein, peptide, or fragment ion.
Delta m Mode	Determines whether ProSightPC conducts the search in delta (Δm) mode, which is explained in "Performing a Search in Delta m Mode" on page 170.
Precursor Tolerance	Specifies the tolerance that determines whether the comparison of an observed precursor ion mass to a theoretical precursor ion mass is considered a match. Set the following parameters: • Lower Bound: Sets the minimum value for a precursor search window that does not trigger an "out of range" warning. • Default Value: Sets the default value for a precursor search window. • Upper Bound: Sets the maximum value for a precursor search window that does not trigger an "out of range" warning.

Parameter	Description
Fragment Tolerance	Specifies the tolerance that determines whether the comparison of an observed fragment ion mass to a theoretical fragment ion mass is considered a match. Set the following parameters: • 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 a fragment tolerance. • Upper Bound: Sets the maximum value for a fragment tolerance that does not trigger an "out of range" warning. The tolerance can be absolute (set in daltons [Da]), or relative (set in parts per million [ppm]).
Minimum Matches	Specifies the minimum number of matching ion fragments that you want the search algorithm to find in proteins. Set the following parameters: • 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.

Setting Sequence Tag Search Preferences

When you add new sequence tag searches, set the default values on the Sequence Tag Preferences page of the Options dialog box. For information on sequence tag searches, see "Searching for a Sequence Tag" on page 145.

❖ **To set sequence tag search preferences**

1. Click **Sequence Tag** in the Options dialog box.

   The Sequence Tag Preferences page of the Options dialog box opens, as shown in Figure 24.

Figure 24. Sequence Tag Preferences page in the Options dialog box
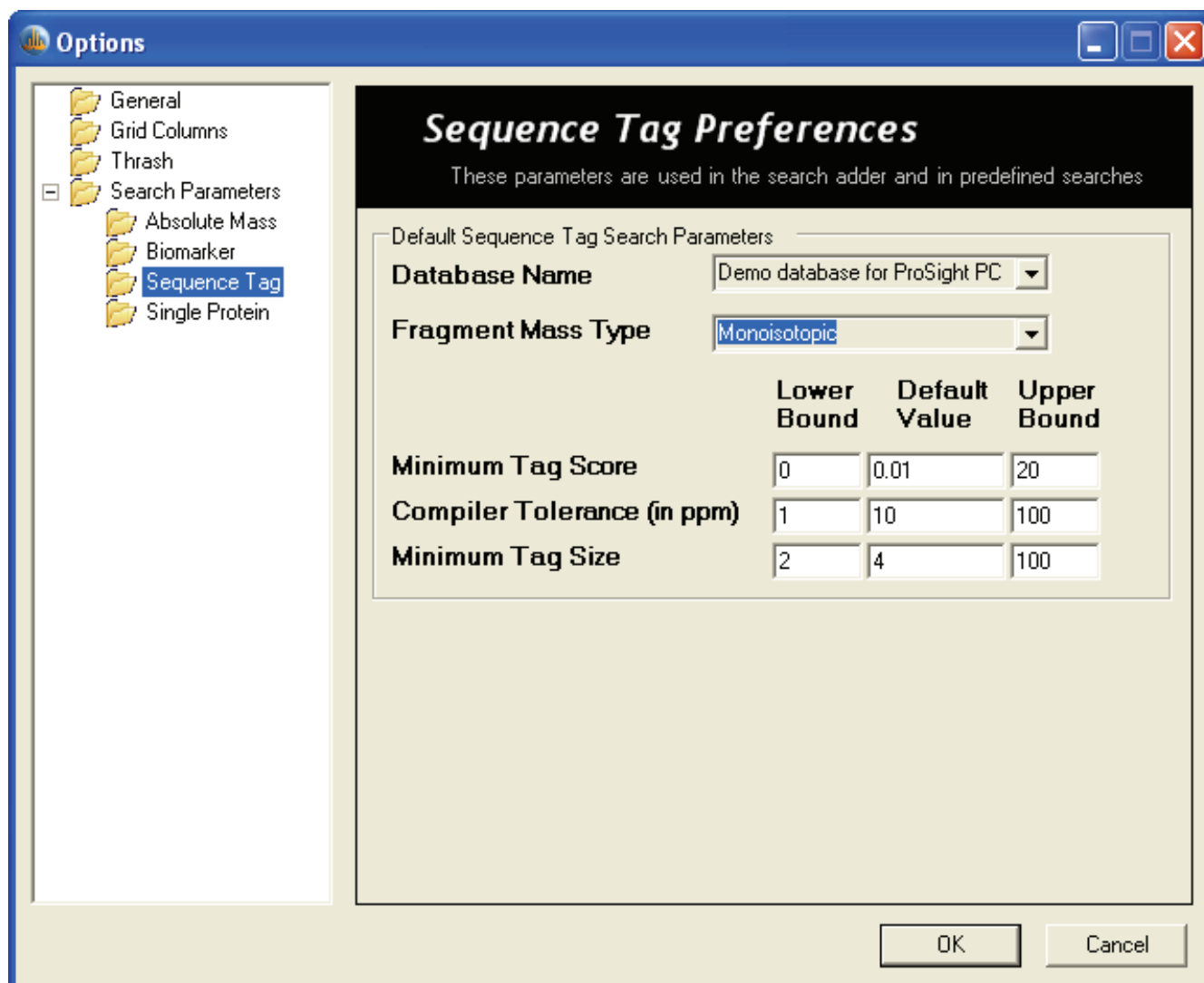
2. In the Database Name list, select the name of the database to search.

3. In the Fragment Mass Type list, select the type of fragment mass:

   - Monoisotopic: Specifies that the fragment mass is monoisotopic, which is the mass of the protein, peptide, or fragment ion, where all carbons are carbon-12.

   - Average: Specifies that the fragment mass is the mass of the most abundant isotope of the protein, peptide, or fragment ion.

4. In the Minimum Tag Score boxes, specify the minimum tag score for protein forms matched in a sequence tag search. ProSightPC scores any sequence found containing one or more of the sequence tags and reports any sequence scoring above this defined minimum tag score. Set the following parameters:

   • 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.

5. In the Compiler Tolerance (in ppm) boxes, enter the permissible error, measured in ppm, between two fragment ion masses that are still considered matching an amino acid. Set the following parameters:

   • 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.

6. In the Minimum Tag Size boxes, enter the lowest acceptable sequence tag score reported as a match. Set the following parameters:

   • 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.

7. Click **OK**.

Sequence Tag Preferences Page Parameters

The Sequence Tag Preferences page of the Options dialog box contains the following parameters:

Parameter	Description
Database Name	Specifies the name of the database to search.
Fragment Mass Type	Specifies the type of fragment mass: • Monoisotopic: Specifies that the fragment mass is monoisotopic, which is the mass of the protein, peptide, or fragment ion, where all carbons are carbon-12. • Average: Specifies that the fragment mass is the mass of the most abundant isotope of the protein, peptide, or fragment ion.
Minimum Tag Score	Specifies the minimum tag score for protein forms matched in a sequence tag search. ProSightPC scores any sequence found containing one or more of the sequence tags and reports any sequence scoring above this defined minimum tag score. Set the following parameters: • 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.

Parameter	Description
Compiler Tolerance (in ppm)	Specifies the permissible error, measured in ppm, between two fragment ion masses that are still considered matching an amino acid. Set the following parameters: • 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	Specifies the lowest acceptable sequence tag score reported as a match. Set the following parameters: • 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.

Setting Single-Protein Search Preferences

When you add new single protein searches, you can set the default values on the Single Protein Preferences page of the Options dialog box. For more information on single-protein searches, see "Searching for Single Proteins" on page 177.

❖ **To set single-protein search preferences**

1. Click **Single Protein** in the Options dialog box.

   The Single Protein Preferences page of the Options dialog box opens, as shown in Figure 25.

Figure 25. Single Protein Preferences page of the Options dialog box
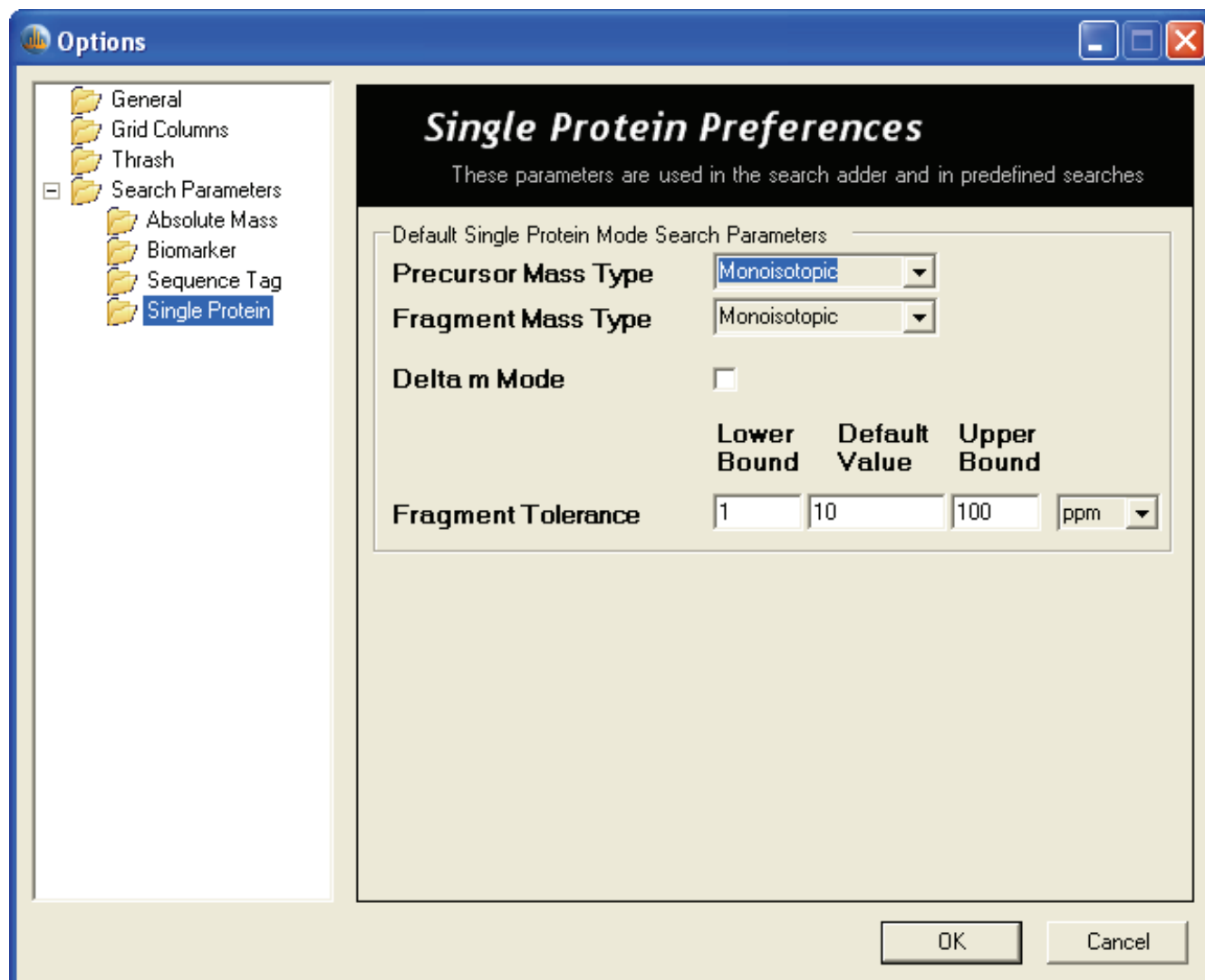
2. In the Precursor Mass Type box, specify the type of precursor mass:

   • Monoisotopic: Specifies that the precursor mass is monoisotopic, which is the mass of the protein, peptide, or fragment ion, where all carbons are carbon-12.

   • Average: Specifies that the precursor mass is the mass of the most abundant isotope of the protein, peptide, or fragment ion.

3. In the Fragment Mass Type box, specify the type of fragment mass:

   • Monoisotopic: Specifies that the fragment mass is monoisotopic, which is the mass of the protein, peptide, or fragment ion, where all carbons are carbon-12.

   • Average: Specifies that the fragment mass is the mass of the most abundant isotope of the protein, peptide, or fragment ion.

4. (Optional) Select the Delta m Mode box if you want to conduct the search in delta (Δ*m*) mode, which is explained in "Performing a Search in Delta m Mode" on page 170.

5. In the Fragment Tolerance boxes, specify the tolerance that determines whether the comparison of an observed fragment ion mass to a theoretical fragment ion mass is considered a match. Set the following parameters:

    *   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 a fragment tolerance.

    *   Upper Bound: Sets the maximum value for a fragment tolerance that does not trigger an "out of range" warning.

    The tolerance can be absolute (set in daltons [Da]), or relative (set in parts per million [ppm]).

    A fragment tolerance is a mass value, either absolute or relative, within which your observed masses must match the theoretical fragment mass. For instance, if you set your tolerance to 0.005 Da (an absolute tolerance) and your theoretical fragment ion is at 1154.1126 Da, observed fragment ions of 1154.1090 Da (-0.0034 Da from theoretical) and 1154.1167 (+0.0041 Da from theoretical) fall within the tolerance, but 1154.2312 (+0.1222) does not, because the mass difference is greater than the tolerance that you set.

6. Click **OK**.

Single Protein Preferences Page Parameters

The Single Protein Preferences page of the Options dialog box contains the following parameters:

Parameter	Description
Precursor Mass Type	Specifies the type of precursor mass:
	• Monoisotopic: Specifies that the precursor mass is monoisotopic, which is the mass of the protein, peptide, or fragment ion, where all carbons are carbon-12.
	• Average: Specifies that the precursor mass is the mass of the most abundant isotope of the protein, peptide, or fragment ion.
Fragment Mass Type	Specifies the type of fragment mass:
	• Monoisotopic: Specifies that the fragment mass is monoisotopic, which is the mass of the protein, peptide, or fragment ion, where all carbons are carbon-12.
	• Average: Specifies that the fragment mass is the mass of the most abundant isotope of the protein, peptide, or fragment ion.
Delta m Mode	Determines whether ProSightPC conducts the search in delta (Δm) mode, which is explained in "Performing a Search in Delta m Mode" on page 170.
Fragment Tolerance	Specifies the tolerance that determines whether the comparison of an observed fragment ion mass to a theoretical fragment ion mass is considered a match. Set the following parameters:
	• 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 a fragment tolerance.
	• Upper Bound: Sets the maximum value for a fragment tolerance that does not trigger an "out of range" warning.
	The tolerance can be absolute (set in daltons [Da]) or relative (set in parts per million [ppm]).

Getting Started

This chapter explains how to start using ProSightPC in three different ways according to the type of input data that you have:

- If you have an LC-MS/MS .raw file or a .puf file as input, you can use ProSightPC's High Throughput wizard.

- If you have a targeted .raw file as input, you can use the Post Xtract or THRASH algorithm to import it.

- If you want to enter data manually, you can use ProSightPC's Experiment Adder.

Contents

- Importing or Creating a Proteome Database

- Processing LC-MS/MS Data Files

- Importing Targeted .raw Files

- Entering Data Manually

- Importing Experiments

- Searching

Importing or Creating a Proteome Database

You have two options for importing or creating a proteome database: download databases from the ProSightPC FTP Web site or manually create a custom database.

❖ **To download a database from the ProSightPC FTP Web site**

1. Choose **Databases > Download ProSightPC Databases**.

2. Click the desired database to download it to your local computer.

❖ **To manually create a custom database**

1. Choose **Databases > Create a Custom Database** and click the **Create Database** icon, , on the Database Manager.

2. Follow the instructions in "Using Proteome Databases" on page 207.

Processing LC-MS/MS Data Files

If you want to load an LC-MS/MS .raw file or a .puf file as input, you can use ProSightPC's High Throughput wizard to process the data against the database that you downloaded or created. This section explains how to use this wizard, how to set custom processing options, and how to use its repositories.

You cannot use the High Throughput wizard to import a targeted .raw file as input or to enter data manually into ProSightPC. Instead, you must use the procedures given in "Importing Targeted .raw Files" on page 101 to import a targeted .raw file or the procedure given in "Entering Data Manually" on page 111 to enter data manually.

Using the High Throughput Wizard

You can quickly process LC-MS/MS data through ProSightPC's High Throughput wizard. If you want to load an LC-MS/MS .raw file or a .puf file, follow the procedures in this section. To view a demonstration of these procedures, see "Demonstrating the High Throughput Wizard" on page 78.

If you want to load a targeted .raw file, see "Importing Targeted .raw Files" on page 101.

Setting Processing Options

To set the processing options in the High Throughput wizard, follow this procedure.

❖ **To set the processing options in the High Throughput wizard**

1. Choose **ProSightHT > HighThroughput Wizard** or click the **HT Wizard** icon, .

   The Process a dataset page of the High Throughput wizard appears, as shown in Figure 26.

Figure 26. Process a dataset page of the High Throughput wizard
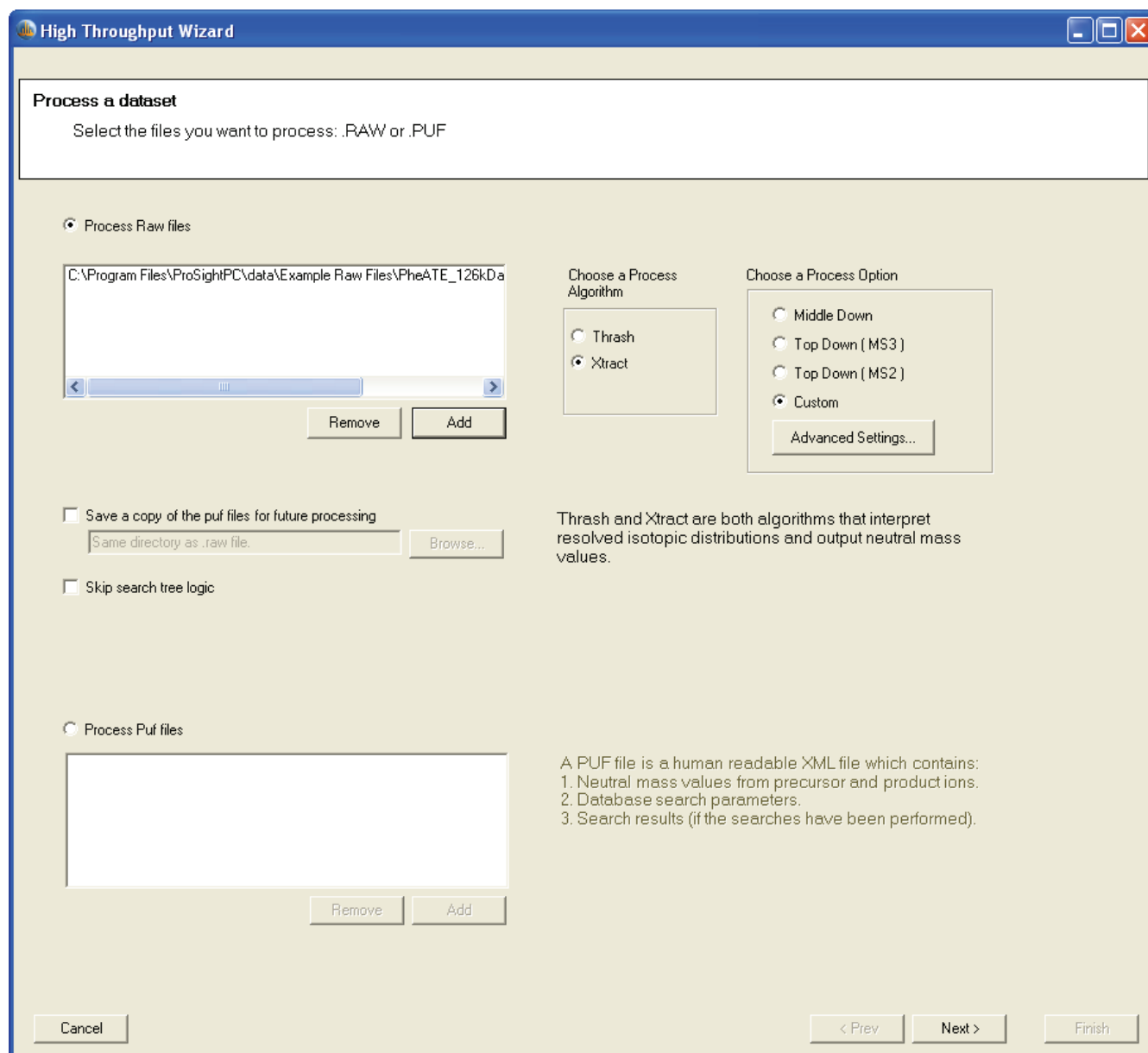
2. Select the **Process Raw files** or **Process Puf files** option, depending on the type of data that you want to import.

- Process Raw files: Converts LC-MS/MS .raw files to .puf files, using an extension of the THRASH or Xtract algorithm designed to analyze high-resolution profile LC-MS/MS data collected on Thermo Scientific Fourier Transform instruments.

   This option is the default. Example .raw files are available in the Example Raw Files folder for demonstration purposes.

- Process Puf files: Processes .puf files.

3. To add a .raw or .puf file, click **Add**, then browse to the file in the dialog box that opens, and select the file.

4. If you choose a .raw file, select one of the following methods in the Choose a Process Algorithm area for importing the data files:

   • Thrash: Uses the THRASH algorithm to process the input file.

   • Xtract: Uses the Xtract algorithm to process the input file. This option reduces analysis and search time and should give better results. This option is the default.

   As the dialog box itself notes, Xtract and THRASH are both algorithms that interpret resolved isotopic distributions and output neutral mass values. For more information on these algorithms, see "Importing Targeted .raw Files" on page 101.

5. If you choose a .raw file, select a processing option in the Choose a Process Option area for importing the data files:

   • Middle Down: Specifies the following default settings for the Xtract and THRASH processing algorithms.

   Xtract:

Precursor Minimum S/N	7.0
Precursor Maximum Charge	25
Precursor Minimum Fit	40
Precursor Remainder Threshold	20
Precursor Selection Criterion	Highest Intensity
Allow Multiple Precursors	Selected
Relative Precursor Threshold	10%
Precursor Add Remainder Afterwards	Unselected
Fragmentation Minimum S/N	3.0
Fragmentation Maximum Charge	25
Fragmentation Minimum Fit	10
Fragmentation Remainder Threshold	10
Minimum Fragmentation Base Peak Intensity	100
Fragmentation Add Remainder Afterwards	Selected

   THRASH:

Precursor Minimum S/N	7.0
Precursor Minimum RL	0.90
Precursor Maximum Charge	25

Precursor Maximum Mass (kDa)	25
Precursor Selection Criterion	Highest Intensity
Allow Multiple Precursors	Selected
Relative Precursor Threshold	10%
Fragmentation Minimum, S/N	3.0
Fragmentation Minimum RL	0.90
Fragmentation Maximum Charge	25
Fragmentation Maximum Mass (kDA)	25
Minimum Number of Fragmentation Scans	1
Minimum Fragmentation Base Peak Intensity	100

- Top Down (MS3): Specifies the following default settings for the Xtract and THRASH processing algorithms.

Xtract:

Precursor Minimum S/N	7.0
Precursor Maximum Charge =	30
Precursor Minimum Fit =	40
Precursor Remainder Threshold	20
Precursor Selection Criterion	Highest Intensity
Allow Multiple Precursors	Selected
Relative Precursor Threshold	10%
Precursor Add Remainder Afterwards	Unselected
Fragmentation Minimum S/N	3.0
Fragmentation Maximum Charge	30
Fragmentation Minimum Fit	10
Minimum Fragmentation Base Peak Intensity	100
Fragmentation Add Remainder Afterwards	Selected

THRASH:

Precursor Minimum S/N	7.0
Precursor Minimum RL	0.90
Precursor Maximum Charge	40
Precursor Maximum Mass (kDA)	35
Precursor Selection Criterion	Highest Intensity

Allow Multiple Precursors	Selected
Relative Precursor Threshold	10%
Fragmentation Minimum S/N	30
Fragmentation Minimum RL	0
Fragmentation Maximum Charge	25
Fragmentation Maximum Mass (kDA)	25
Minimum Number of Fragmentation Scans	1
Minimum Fragmentation Base Peak Intensity	100

- Top Down (MS2): Specifies the following default settings for the Xtract and THRASH processing algorithms.

Xtract

Precursor Minimum S/N	7.0
Precursor Maximum Charge	30
Precursor Minimum Fit	40
Precursor Remainder Threshold	20
Precursor Selection Criterion	Highest Intensity
Allow Multiple Precursors	Selected
Relative Precursor Threshold	10%
Precursor Add Remainder Afterwards	Unselected
Fragmentation Minimum S/N	3.0
Fragmentation Maximum Charge	30
Fragmentation Minimum Fit	10
Minimum Fragmentation Base Peak Intensity	100
Fragmentation Add Remainder Afterwards	Selected

THRASH

Precursor Minimum S/N	7.0
Precursor Minimum RL	0.90
Precursor Maximum Charge	40
Precursor Maximum Mass (kDA)	35
Precursor Selection Criterion	Highest Intensity
Allow Multiple Precursors	Selected
Relative Precursor Threshold	10%
Fragmentation Minimum S/N	30
Fragmentation Minimum RL	0
Fragmentation Maximum Charge	25
Fragmentation Maximum Mass (kDA)	25
Minimum Number of Fragmentation Scans	1
Minimum Fragmentation Base Peak Intensity	100

- Custom: Click **Advanced Settings** and use the Advanced Settings dialog box to specify your own settings. See "Using Custom Settings" on page 89 for instructions.

6. (Optional) If you selected the Process Raw files option, select the **Save a copy of the puf files for future processing** option to save a physical .puf file containing the results. Click **Browse** to browse to the directory where you want to save the .puf files.

   This option is useful for rapidly re-searching the data instead of processing the .raw file again. If you do not select this option, the results reside in a ProSightPC repository. You can always import them into ProSightPC and save them as a .puf file.

7. (Optional) If you choose not to search the data against a proteome database, select the **Skip search tree logic** option.

8. Click **Next**.

Process a Dataset Page Parameters

The Process a dataset page of the HighThroughput wizard contains the following parameters:

Parameter	Description
Process Raw files	Converts LC-MS/MS .raw files to .puf files using an extension of the THRASH or Xtract algorithm designed to analyze high-resolution profile LC-MS/MS data collected on Thermo Scientific Fourier Transform instruments.
Remove	Removes the selected .raw file displayed in the box.

Parameter	Description
Add	Opens a dialog box so that you can browse for a .raw file to process.
Choose a Process Algorithm	Specifies the method for converting mass spectral data to neutral mass values when importing the data files: • Thrash: Uses the THRASH algorithm to process the input file. • Xtract: Uses the Xtract algorithm to process the input file. This option reduces analysis and search time and should give better results. This option is the default. For more information on these algorithms, see "Importing Targeted .raw Files" on page 101.
Choose a Process Option	Specifies the settings for the Xtract and THRASH processing algorithms: • Middle Down: See step 5 of the Using the High Throughput Wizard section for this set of default settings. • Top Down (MS3): See step 5 of the Using the High Throughput Wizard section for this set of default settings. • Top Down (MS2): See step 5 of the Using the High Throughput Wizard section for this set of default settings. • Custom: Gives you the ability to specify your own settings by clicking on Advanced Settings and using the Advanced Settings dialog box. • Advanced Settings: Opens the Advanced Settings dialog box so that you can specify custom default settings for the Xtract and THRASH processing algorithms. See "Advanced Settings Dialog Box Parameters" on page 93 for information on the settings in this dialog box. The Advanced Settings option is only available when you select Custom.
Save a copy of the .puf files for future processing	Saves a physical .puf file containing the results of the .raw file conversion on your computer.
Browse	Enables you to browse to the directory in which to save the .puf file.
Skip search tree logic	Does not search the data against a proteome database.
Process Puf files	Processes .puf files.

Parameter	Description
Remove	Removes the selected .puf file displayed in the box.
Add	Opens a dialog box so that you can choose the .puf file to import.

Selecting or Creating a Repository

When you click Next in the Process a dataset page of the High Throughput wizard, the Running Highthroughput Logic page of the High Throughput wizard now appears, as shown in Figure 27, so that you can select or create a repository and define a search tree.

Figure 27. Running Highthroughput Logic page of the High Throughput wizard
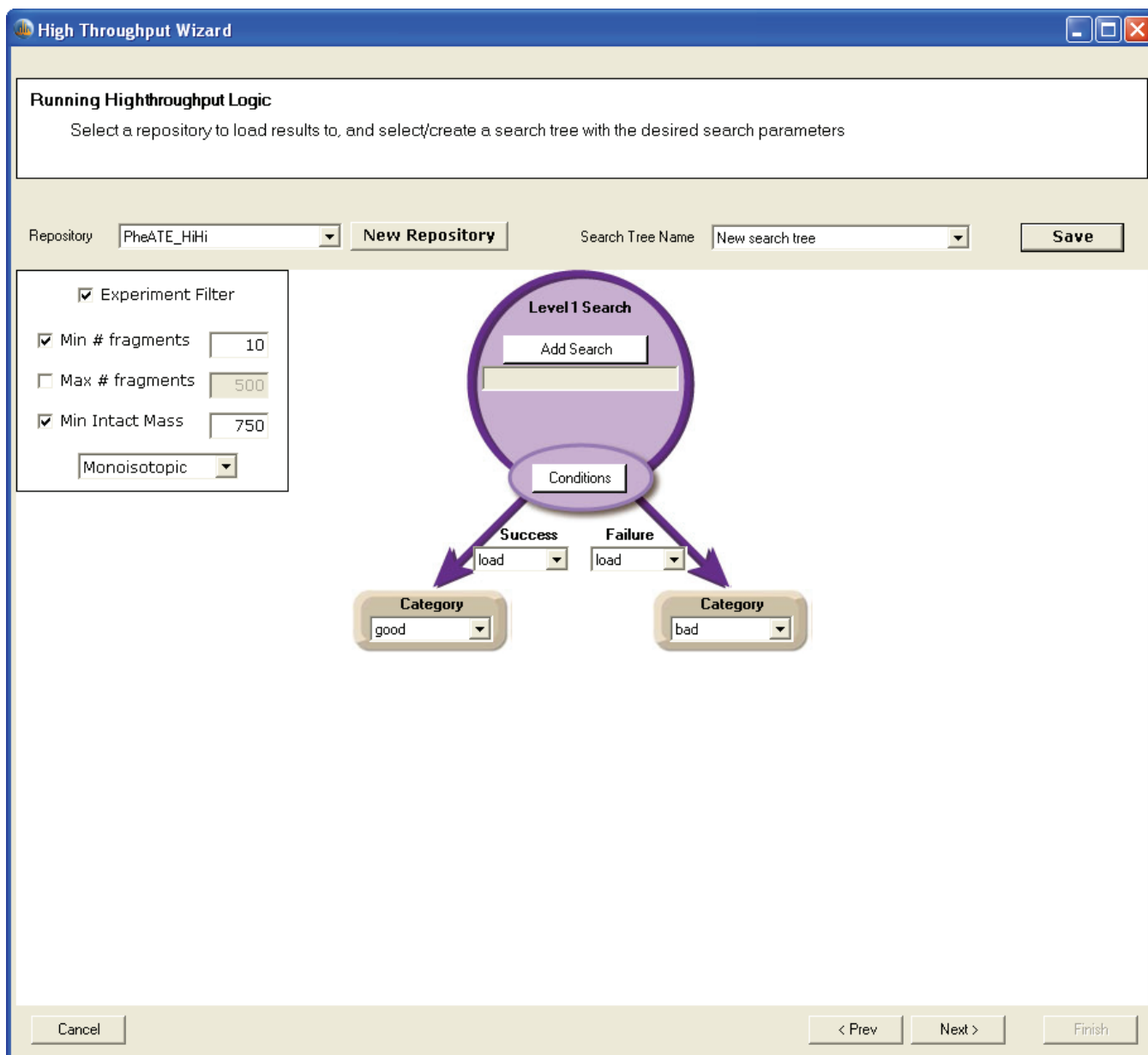
❖ **To select a repository**

1. From the Repository list on the Running Highthroughput Logic page, shown in Figure 27, select the name of the repository.

2. If you want to create a new repository, click **New Repository**, enter the new repository name in the New Repository dialog box, and click **OK**.

For more information on creating a repository, see "Creating a Repository" on page 95. For more information on editing a repository, see "Editing a Repository" on page 96.

Creating a Search Tree

ProSightPC automates searches through an iterative search tree to make the best use of your time. If you find a match during the first search, you do not have to run the second, longer search.

Each experiment created by ProSightPC enters the search tree at the top, as shown in Figure 27. ProSightPC conducts the first search according to the definitions in that box. The results of that search come back and are graded by the conditions set for that search node. Generally, ProSightPC uses the condition that the best expectation score returned by the search is less than 0.0001. If the expectation value is lower than 0.0001, ProSightPC loads the results into the "good" category, but if the results are greater than 0.0001, ProSightPC tries another search with "looser" search parameters, that is, a larger precursor search window, biomarker mode, or delta (Δm) mode. This usually means that the search will take longer. You now move down to the next node of the search tree, as shown in Figure 28. ProSightPC checks the results of that second search against the conditions that you set (again, for example, where the expectation value is less than 0.0001), and if the results meet the conditions, ProSightPC loads the results to the "good" category. If they do not meet the conditions, ProSightPC loads it to the "bad" category and you can manually try to run them again. "Good" and "bad" are arbitrary category names for searches that pass or fail the conditions set in the search tree, respectively. You can add results from searches, such as biomarker or delta (Δm) mode, to the results repository.

Figure 28. Second-level searching
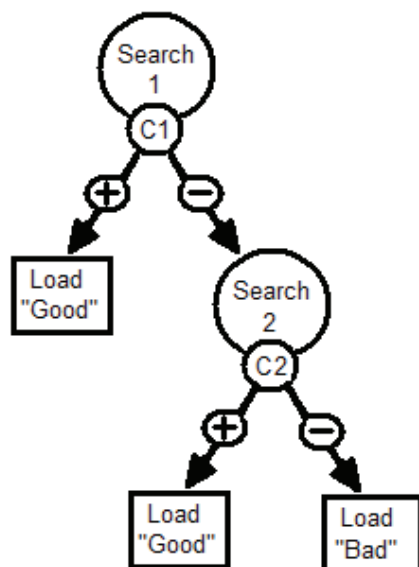
Creating or Editing a One-Level Search Tree

❖ **To create or edit a one-level search tree**

1. If you are creating a search tree, select **New search tree**, which is selected by default.

   If you are editing an existing search tree, select the name of the search tree from the Search Tree Name list on the Running Highthroughput Logic page.

2. Select the **Experiment Filter** check box to filter out experiments that will not yield matches.

   If you are looking for intact proteins, you might want to set a minimum precursor mass of 2000 Da to eliminate peptides from being searched.

   a. Select the **Min # fragments** check box (this option is selected by default), and in the box to the right of the option, enter the minimum number of fragments to search for.

      The default is 10, indicating that experiments that contain fewer than 10 fragments are ignored.

   b. (Optional) Select the **Max # fragments** check box, and in the box to the right of the option, enter the maximum number of fragments to search for.

      The default is 500, indicating that experiments that contain more than 500 fragments are ignored.

   c. Select the **Min Intact Mass check** box (this option is selected by default), and in the box to the right of the option, enter the minimum intact mass number.

      The default is 750 Da, indicating that experiments whose intact mass is less than 750 Da are ignored.

    d.   From the list beneath the Min Intact Mass option, specify the mass type:

       • Monoisotopic: Specifies that the mass is monoisotopic, which is the mass of the protein, peptide, or fragment ion, where all carbons are carbon-12.

       • Average: Specifies that the mass is the mass of the most abundant isotope of the protein, peptide, or fragment ion.

3. To define the first-level search, click **Add Search** in the Level 1 search area.

ProSightPC opens the SearchPickerHT dialog box, shown in Figure 29.

Figure 29. SearchPickerHT dialog box
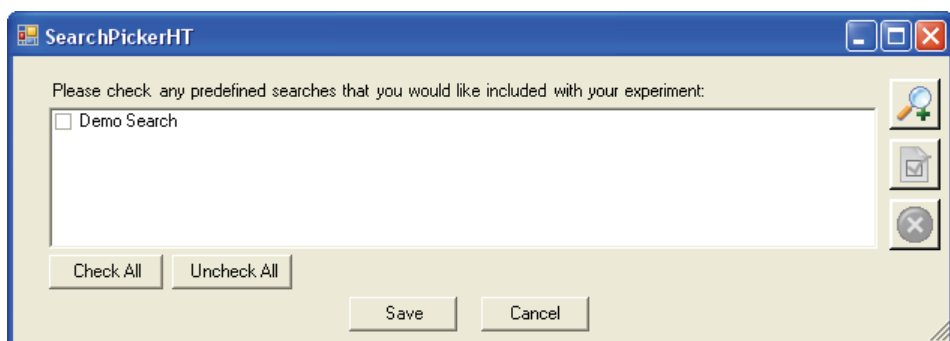
    a.   Select a predefined search to use by selecting the appropriate check box.

       A predefined search enables you to assign a name to a set of parameters that you can then add to any experiment. It reduces the repetition of identical searches on different sets of MS/MS data. For more information on predefined searches, see "Performing Predefined Searches" on page 163.

    b.   To add a search, click 🔍 on the dialog box.

       The New Predefined Search dialog box opens. Follow the instructions in "Creating a Predefined Search" on page 163 to create a new predefined search.

    c.   To edit a predefined search, select the name of the search and click ☑.

       The Edit Predefined Search dialog box opens. Follow the instructions in "Editing a Predefined Search" on page 166 to edit a predefined search.

    d.   To remove a predefined search, select the name of the search and click ❌.

    e.   Click **Save** in the SearchPickerHT dialog box.

4. To set the conditions for the first search, click **Conditions** in the purple circle in the Level 1 area.

The Condition dialog box appears, as shown in Figure 30. You can use operators and values to create conditions for the search.

Figure 30. Condition dialog box
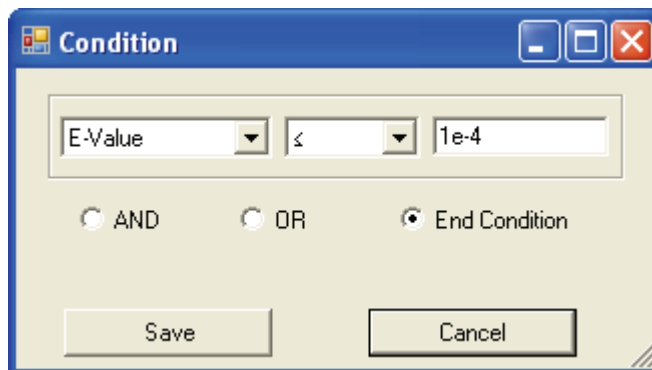
a. From the left list, select **Number of hits** or **E-Value**.

- Number of hits: Specifies the number of matches for an intact ion in the search.

- E-Value: Specifies the expectation value (e value) for the results of the search. If at least one search result received an e value of less than e-4, the search is loaded to the "good" category. For information on the expectation value, see "Expectation Value (e value)" on page 186. This setting is the default.

b. From the middle list, select the operator that indicates the relationship between the values in the left and right boxes.

- ≤: Indicates that the first value is less than or equal to the second value. This setting is the default.

- ≥: Indicates that the first value is greater than or equal to the second value.

c. From the right list, enter the desired value.

The default value is 1e-4. If you enter an illogical value, the background of the box becomes bright red.

d. Select one of the following operators:

- AND

- OR

- End Condition

If you only want to conduct a search with one condition, select the **End Condition** option.

If you want to add another condition, select **AND** or **OR**. When you select AND or OR, the Condition dialog box expands, as shown in Figure 31, so that you can add a second condition.

Figure 31. Expanded Condition dialog box
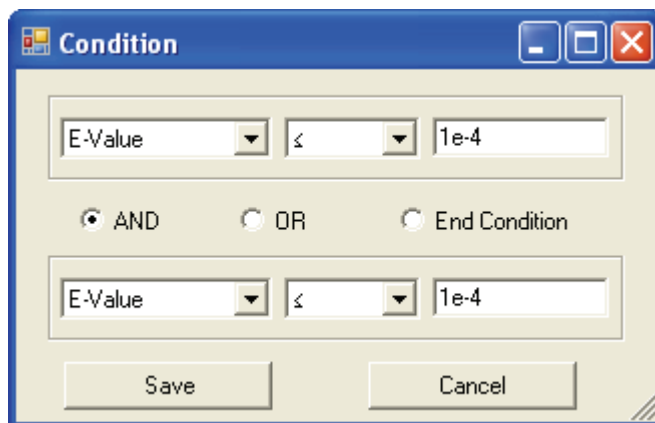
e. When you have set the conditions for all searches, click **Save** in the Condition
    dialog box.

5. From the Success list of the Level 1 search, select **load** or **run search**.

- Load: Loads the results to the selected category if the experiment passed the
  condition.

- Run search: Opens a second-level search tree if the experiment passed the condition,
  as shown in Figure 34. The experiment is re-searched with the second-level search
  and is later loaded to the categories that depend on the conditions set. For
  instructions on creating a two-level search, see "To create a two-level search tree" on
  page 76.

6. From the Failure list, select **load** or **run search**.

- Load: Loads the results to the selected category if the experiment failed the condition.

- Run search: Opens a second-level search tree if the experiment failed the condition, as
  shown in Figure 34. The experiment is re-searched with the second-level search and is
  later loaded to the categories that depend on the conditions set. For instructions on
  creating a two-level search, see "To create a two-level search tree" on page 76.

7. From the Category list under Success, select **good** to specify that searches that pass the
    conditions set in the search tree be stored in the repository in that specific category.

8. From the Category list under Failure, select **bad** to specify that searches that fail the
    conditions set in the search tree be stored in the repository in that specific category.

9. If you want to create a second-level search, see "To create a two-level search tree" on
    page 76.

10. Click **Save** in the upper right corner of the High Throughput wizard to save your search
    tree.

    The Save Search Tree dialog box appears, as shown in Figure 32.

Figure 32. Save Search Tree dialog box
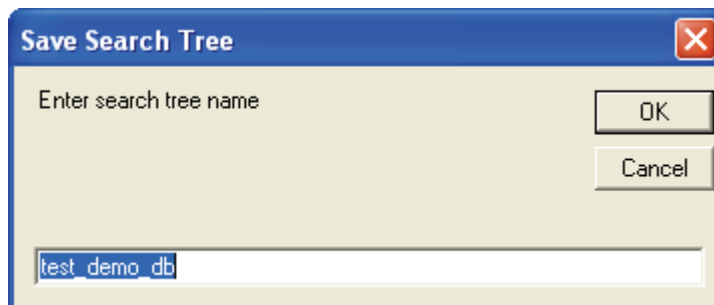
11. If you created a new search tree, type the name of the search tree and click **OK**. If you made changes to an existing search tree, a prompt box appears that asks if you want to replace the existing search tree. Click **Yes**.

12. Click **OK** in the message box.

13. Click **Next** in the Running Highthroughput Logic page.

You now see a summary of the parameters that you have set. Figure 33 gives an example.

Figure 33. Summary page of the High Throughput wizard
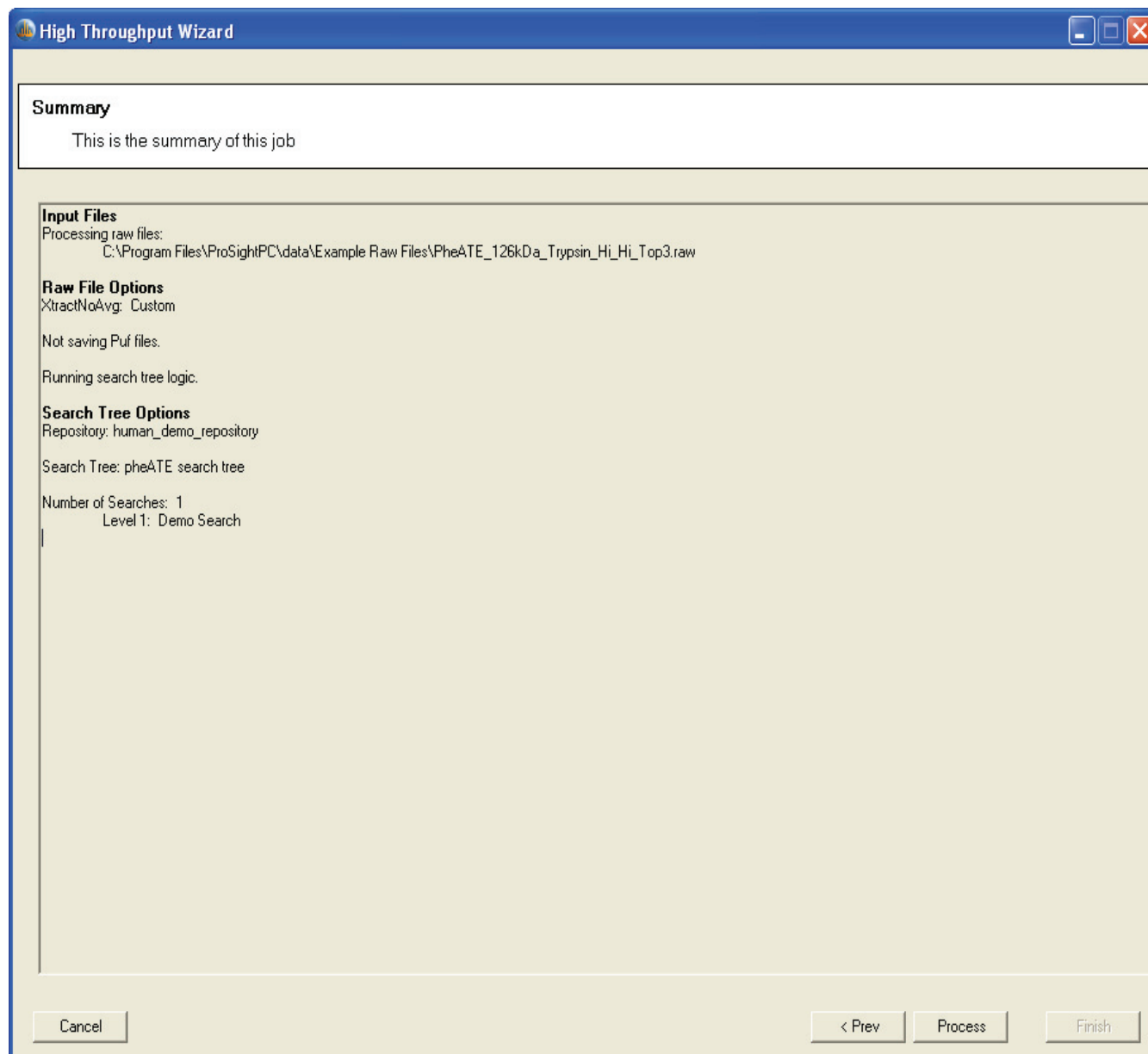
14. Click **Process**.

The searching now begins and can take a long time, depending on the length of the .raw file, the complexity of the database, and the parameters of the search. When ProSightPC finishes the search, you can view the repository report. For information on manipulating the data in this report, see "Generating a Repository Report" on page 79.

Running Highthroughput Logic Page Parameters

The Running Highthroughput Logic page of the High Throughput wizard, shown in
Figure 27 on page 65, contains the following parameters.

Parameter	Description
Repository	Specifies the name of the repository where the search results will be saved.
New Repository	Opens the New Repository dialog box, shown in Figure 42 on page 95, so that you can specify the name of the repository where the search results will be saved.
Search Tree Name	Specifies the name of the new or existing search tree.
Save	Opens the Save Search Tree dialog box, shown in Figure 32 on page 71, so that you can save a search tree under a new name.
Experiment Filter	Filters out experiments that will not yield matches.
Min # fragments	Specifies the minimum number of fragments to search for. The default is 500.
Max # fragments	Specifies the maximum number of fragments to search for. The default is 10.
Min Intact Mass	Specifies the minimum intact mass number. The default is 750 Da.
Mass type list	Specifies the mass type: • Monoisotopic: Specifies that the mass is monoisotopic, which is the mass of the protein, peptide, or fragment ion, where all carbons are carbon-12. • Average: Specifies that the mass is average, which is the mass of the most abundant isotope of the protein, peptide, or fragment ion.
Add search	Opens the SearchPickerHT dialog box, shown in Figure 29 on page 68, so that you can add a predefined search to your search.
Conditions	Opens the Condition dialog box, shown in Figure 30 on page 69, so you can sets the conditions for the search.

Parameter	Description
Success	Specifies the action to take on the results that passed the condition.
	• Load: Loads the results to the selected category if the experiment passed the condition.
	• Run search: Opens a second-level search tree if the experiment passed the condition, as shown in Figure 34 on page 77. The experiment will be re-searched with the second-level search and will later be loaded to the categories that depend on the conditions. See "To create a two-level search tree" on page 76 for instructions on creating a two-level search.
Category	Specifies that the searches that pass the conditions set in the search tree be stored in the repository in that specific category.
Failure	Specifies the action to take on the results that failed the condition.
	• Load: Loads the results to the selected category if the experiment failed the condition.
	• Run search: Opens a second-level search tree if the experiment failed the condition, as shown in Figure 34 on page 77. The experiment will be re-searched with the second-level search and will later be loaded to the categories that depend on the conditions. See "To create a two-level search tree" on page 76 for instructions on creating a two-level search.
Category	Specifies that the searches that fail the conditions set in the search tree be stored in the repository in that specific category.

SearchPickerHT Dialog Box Parameters

The SearchPickerHT dialog box, shown in Figure 29 on page 68, contains the following parameters:

Parameter	Description
Please check any predefined searches that you would like included with your experiment	Lists the available predefined searches.
Check All	Selects all the available predefined searches.
Uncheck All	Clears all the available predefined searches.
Save	Saves the changes that you have made to the predefined searches.

Parameter	Description
🔍	Adds a search.
☑	Edits a predefined search.
✖	Removes a predefined search.

Condition Dialog Box Parameters

The Condition dialog box, shown in Figure 30 on page 69, contains the following parameters:

Parameter	Description
Left list	Sets a condition that the results of the search must meet, either: • Number of hits: Specifies the number of matches for an intact ion in the search. • E-Value: Specifies the expectation value (e value) for the results of the search. If at least one search result received an e value of less than e-4, the search is loaded to the "good" category. See "Expectation Value (e value)" on page 186 for information on the expectation value. This setting is the default.
Middle list	Specifies the operator that indicates the relationship between the values in the left and right boxes. • $\leq$: Indicates that the first value is less than or equal to the second value. This setting is the default. • $\geq$: Indicates that the first value is greater than or equal to the second value.
Right list	Specifies the desired value.
AND	Expands the Condition dialog box so that you can add another condition. The search results must meet the first condition and the second.
OR	Expands the Condition dialog box so that you can add another condition. The search results must meet either the first condition or the second.
End condition	Indicates that the search has only one condition.
Save	Saves the condition or conditions for the search.

Save Search Tree Dialog Box Parameters

The Save Search tree dialog box, shown in Figure 32 on page 71, specifies a name for the search tree that you created.

Creating a Two-Level Search Tree

You might want to create a two-level search tree on the Success side. For example, you might have run a preliminary general search but want to search the modifications on the experiments that found matches.

You might also want to perform a two-level search on the Failure side. Suppose that you ran a first-level search and some of the experiments found no matches—that is, the experiments failed the conditions. To obtain good results, you might then want to run a different search, such as a broader search or a search of a different database, a search with different parameters, or a search in a different mode. But if some of the experiments found matches, you do not need to run another search.

❖ **To create a two-level search tree**

1. Follow the instructions in "Creating or Editing a One-Level Search Tree" on page 67, and select **Run search** in the Success or Failure list for the first-level search.

   A second-level search tree now opens, as shown in Figure 34.

Figure 34. Second-level search tree
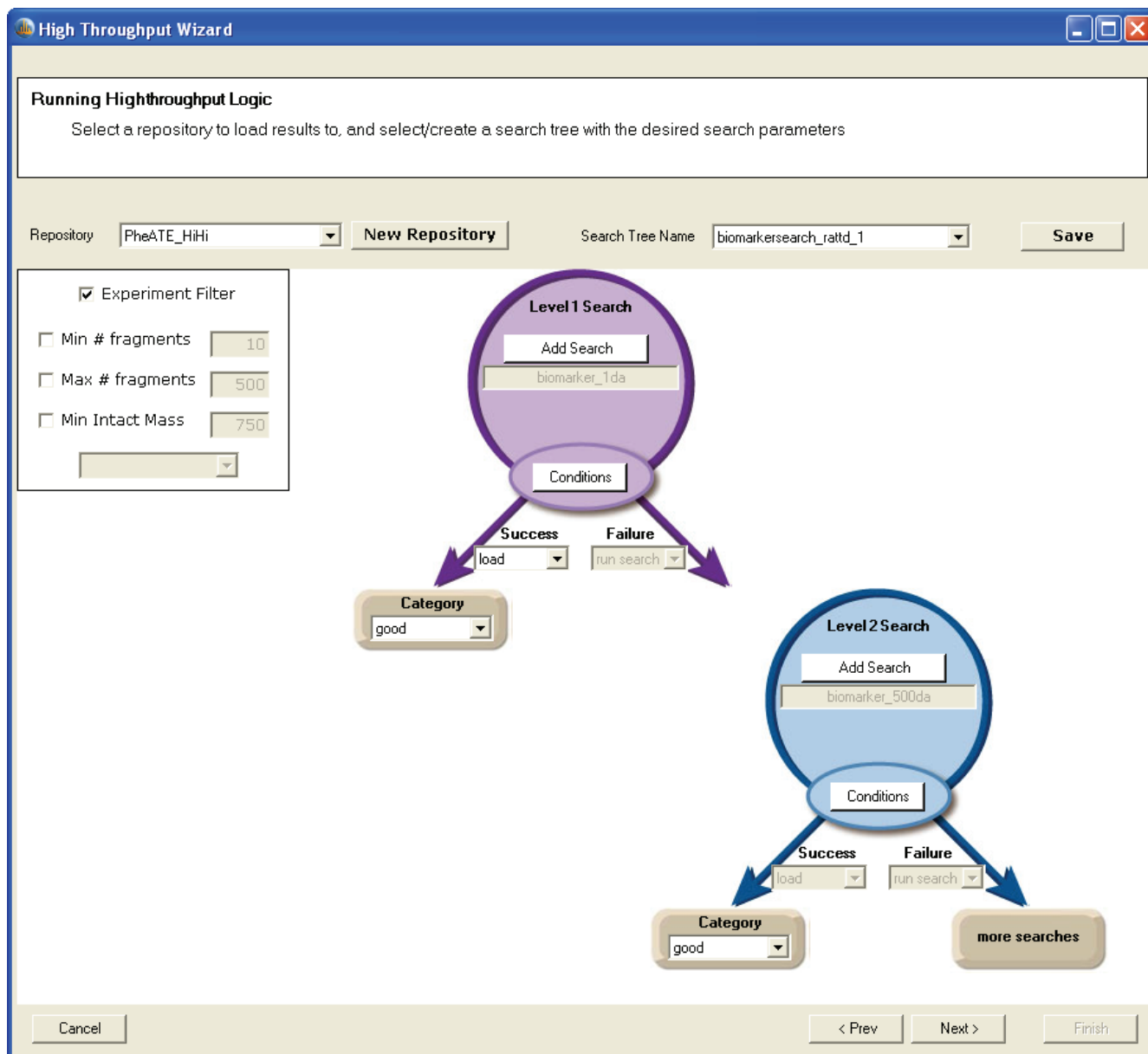
2. Starting with **Add Search**, perform the same steps as for the first-level search; however, the Success and Failure lists are not available in second-level searches, so you must skip this step. If you want to create a search tree of more than two levels, see the next section, "Creating a Search Tree with Three or More Levels" on page 77.

Creating a Search Tree with Three or More Levels

ProSightPC enables you to create two levels of searches in its user interface. However, you can create additional search levels by editing the .xml file that contains the search trees. After you create these levels in the .xml file and run the High Throughput wizard, these levels appear on the Summary page of the wizard, as shown in Figure 35.

Figure 35. Summary page for a five-level search

```
Input Files
Processing raw files:
        C:\Program Files\ProSightPC\data\Histone Raw Files\Histone_H4_actyl_25_scans_11299.409Da.raw

Raw File Options
XtractNoAvg:  Custom

Not saving Puf files.

Running search tree logic.

Search Tree Options
Repository: repository12

Search Tree: biomarkersearch_rattd_1

Number of Searches:  5
        Level 1:  biomarker_1da
        Level 2:  biomarker_500da
        Level 3:  absolute_mass_narrow
        Level 4:  absolute_mass_wide
        Level 5:  sequence_tag_search
```

If you want to edit the .xml file, contact Thermo Fisher Scientific's Marketing department.

Editing a Search Tree

❖ **To edit a search tree**

1. Choose **ProSightHT > Edit/Add Search Tree**.

   The Running Highthroughput Logic page appears in the High Throughput wizard, as shown in Figure 27 on page 65.

2. From the Search Tree Name list, select the search tree that you would like to edit.

3. Follow the instructions in "Creating or Editing a One-Level Search Tree" on page 67 and "Creating a Two-Level Search Tree" on page 76 to edit your search-tree settings.

4. Click **Save** on the Running Highthroughput Logic page.

Demonstrating the High Throughput Wizard

The following demonstration shows you how to use the High Throughput wizard.

Generating a Repository Report

You can generate a repository report that lists all the experiments that a repository contains. Use the repository report to focus on certain experiments that you want to investigate. You can control the display of many categories of information in this report. You can also set fixed and custom filters by which to refine the report data. Furthermore, you can import specified experiments into ProSightPC, manipulate them, and export the experiments back to the repository or to a Microsoft™ Excel™ spreadsheet.

The procedures in this section describe how to generate a repository report and manipulate its data. See "Demonstrating Repository Report Generation" on page 88 for a demonstration showing you how to generate a repository report.

❖ **To generate a repository report**

1. To generate a repository report, choose **ProSightHT > Repository Report** or **Tools > Reports > Repository Report** or click the **Repository Report** icon, ⊞ .

   The Repository Report dialog box opens, as shown in Figure 36.

   Figure 36. Repository Report dialog box
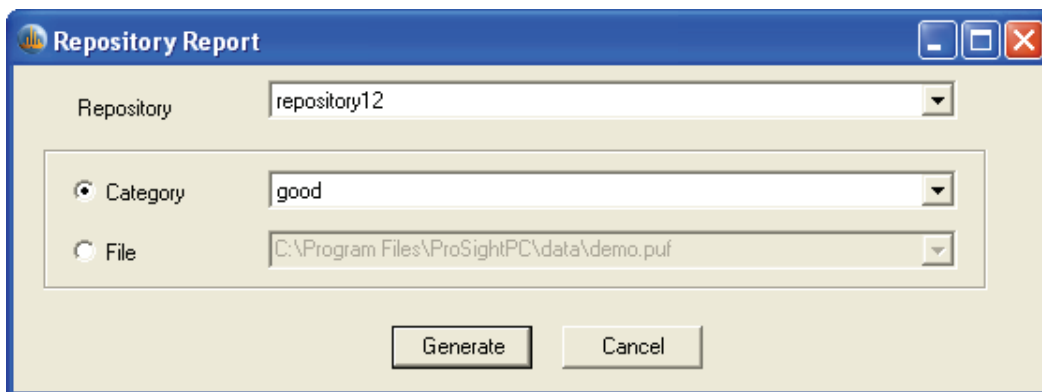
2. From the Repository list, select the name of the repository to generate a report for.

3. If you want to generate a report on all the experiments in a category (for example, all the "good" experiments), select the **Category** option and then select the name of the category from the list to the right of the option.

4. If you want to generate a report on all the experiments in a specific file, select the **File** option, and then select the name of the file from the list to the right of the option or type the name of the file in the box.

   Since each file represents the data that you obtained from an instrument in a specific run, selecting the File option is useful if you want to view the results of this run.

5. Click **Generate**.

The repository report automatically appears in the Tab Controller section of the interface, as shown in Figure 37. Each row in the Repository Report represents the best search result per intact ion in a search in the experiment.

Figure 37. Repository report

Changing the Display Columns in the Repository Report

The repository report includes the following columns:

Parameter	Description
Repository Name	Displays the name of the repository where an experiment belongs.
Category Name	Displays the name of the category where the experiment is saved.
Experiment Number	Displays the number of an experiment.
Search Type	Displays the type of search performed in an experiment: absolute mass, biomarker, sequence tag, single-protein, gene-restricted absolute mass (GRAM), or gene-restricted biomarker (GRBM).

Parameter	Description
Accession Number	Displays the accession number used by the major protein databases, such as UniProt or Swiss-Prot, to index a protein in a database.
E Value	Specifies the expectation value (e value) of the best search result in the search. See "Expectation Value (e value)" on page 186 for more information on the e value.
Sequence	Displays the protein sequence that forms the basis of an experiment.
Number of Matching Fragments	Displays the number of matching ion fragments in the protein identified.
PTMs	Displays the name of the PTM and the RESID number (the number of the amino acid that has the PTM) of the best search result.
Theoretical Mass	Displays the theoretical precursor mass of the protein identified.
Observed Mass Da	Displays the observed precursor mass of the precursor (experimental protein), in daltons.
Mass Diff Da	Displays the difference between the observed precursor mass and the theoretical precursor mass of the protein identified, in daltons.
Mass Diff ppm	Displays the difference between the observed precursor mass and the theoretical precursor mass of the protein identified, in parts per million.
Protein Description	Displays a description of the match—that is, the protein that was found in the search.
Source	Displays the path of the .raw or .puf file that an experiment was based on.
File Name	Displays the name of the .raw or .puf file that an experiment was based on.
Search Number	Displays the number of a search in an experiment. A report can contain multiple rows (searches) for an experiment, and for each search, it can have more rows if there were some intact ions.
Intact ID	Displays the number of an intact ion in an experiment.
Experiment Comment	Displays any comments about an experiment, such as the filters that it passed.

Parameter	Description
Number of Best Hits	Displays the number of matching search results with the best e value. For example, if an experiment had five matching search results, the best search result is the one that received the best (lowest) score. For more information on the e value, see "Expectation Value (e value)" on page 186.
P Score	Displays the p score of the best search result in a search. For more information on the p score, see "p Score" on page 186.
SeqTag Score	Displays the sequence tag score of the best search result in a search. You see a score only if the search type is sequence tag.

You can change the columns that are displayed in the repository report and the order of the columns.

❖ **To change the columns in the repository report**

• Select or clear the boxes in the **Columns To Display** area in the lower right corner of the Tab Controller.

❖ **To change the order of the columns in the repository report**

• Select the column header and move it to the desired location.

Repository Report Dialog Box Parameters

The Repository Report dialog box, shown in Figure 36 on page 80, contains the following parameters:

Parameter	Description
Repository	Specifies the name of the repository to generate a report for.
Category	Specifies the category of experiments in the repository to generate a report for.
File	Specifies the file of experiments in the repository to generate a report for.
Generate	Generates the repository report and displays it in the Tab Controller section of the graphical user interface, as shown in Figure 37.

Using the Repository Report to Import Experiments from a Repository

The repository report page displays the desired data from the repository, but the data is only for viewing. To manipulate the data, you must import the data into ProSightPC, perform any desired operations, such as adding or changing searches, and export it back to the repository to save the changes that you have made.

You can import experiments from a repository by using the following procedure or by using the procedure outlined in "Importing Experiments from a Repository" on page 97.

❖ **To use the repository report to import experiments from a repository**

1. In the repository report page, select the experiments that you want to import into ProSightPC.

   You can select experiments in the following ways:

   • Double-click each experiment.

   • Select the box to the extreme left of each experiment.

   • Use the SHIFT key to select consecutive experiments.

   • Use the CTRL key to select separate experiments.

   • Right-click an experiment and choose either **Select All** or **Check Selected Rows** from the shortcut menu. You can also choose **Unselect All** or **Uncheck Selected Rows** to clear rows.

2. Click **Import** in the Actions area, or right-click the selected experiments in the page and choose **Import** from the shortcut menu.

   If there are already experiments in the Data Grid, you are prompted to replace the current experiments in the Data Grid.

3. Click **Yes**, **No**, **Yes to all**, or **No to all**.

You can also import experiments from a repository by choosing **File > Import Data from Repository**.

Exporting Experiments to a Repository

Once you have imported the experiments shown on the repository report page and performed operations on them, you can export them back to the same repository.

You can export experiments from a repository by using the following procedure or by using the procedure outlined in "Exporting Experiments to a Repository" on page 99.

❖ **To export experiments to a repository**

1. In the repository report page, select the experiments that you want to export to the repository. You can select experiments in the following ways:

   • Select the box to the extreme left of each experiment.

   • Use the SHIFT key to select consecutive experiments.

   • Use the CTRL key to select separate experiments.

- Right-click an experiment and choose either **Select All** or **Check Selected Rows** from the shortcut menu. You can also choose **Unselect All** or **Uncheck Selected Rows** to clear rows.

2. Click the **Export to Repository** button in the Action area, or right-click the selected experiments on the page and choose **Export to Repository** from the shortcut menu. You can also right-click the experiment in the Data Grid, and choose **Export Experiment to Repository**.

   This option is available only if the experiment was imported from the repository.

3. When prompted if you want to replace the current experiments in the Data Grid, click **Yes**, **No**, **Yes to all**, or **No to all**.

You can also export experiments to a repository by choosing **File > Export Data to Repository**.

Exporting Experiments to an Excel Spreadsheet

You can export experiments to an Excel spreadsheet so that you can print, sort, manipulate, copy, and paste the data to other applications. The Excel spreadsheet opens with all the experiments that are selected in all the columns that are visible in the graphical user interface.

To export experiments to an Excel spreadsheet, you must have Excel installed.

❖ **To export experiments to an Excel spreadsheet**

1. In the repository report page, select the experiments that you want to export to the Excel spreadsheet. You can select experiments in the following ways:

   - Select the box to the extreme left of each experiment.

   - Use the SHIFT key to select consecutive experiments.

   - Use the CTRL key to select separate experiments.

   - Right-click an experiment and choose either **Select All** or **Check Selected Rows** from the shortcut menu. You can also choose **Unselect All** or **Uncheck Selected Rows** to clear rows.

2. Click **Export to Excel** in the Actions area, or right-click the selected experiments on the page and choose **Export to Excel** from the shortcut menu.

An Excel spreadsheet now opens, showing all the experiments that you selected.

Applying Filters to Repository Report Data

You can apply fixed filters or define custom filters by which to refine the type of data shown in the repository report. The fixed filters are the most common filters that users apply.

❖ **To apply fixed filters**

1. In the Fixed Filters section of the Actions area, shown in Figure 38, select one or more of the following filters:

   • Search Type: Displays all the experiments whose search type is the search type selected in the adjacent list: absolute mass, biomarker, sequence tag, single-protein, gene-restricted absolute mass (GRAM), or gene-restricted biomarker mass (GRBM).

   • E value (confident match): Displays all the experiments whose e value is less than the value that you entered in the box. The default value, 1E-4, is recommended for a confident match. For more information on the e value, see "Expectation Value (e value)" on page 186.

   • PTMs: Displays all the experiments with PTMs when you select Show.

   • Mass Difference: Displays all the experiments whose mass difference is less than the value that you entered in the box.

   • Category: Displays all the experiments whose category is the same as that selected in the adjacent list.

   • Unique Identifications: Displays all the experiments except for those with redundant accession numbers.

   Figure 38. Fixed Filters section
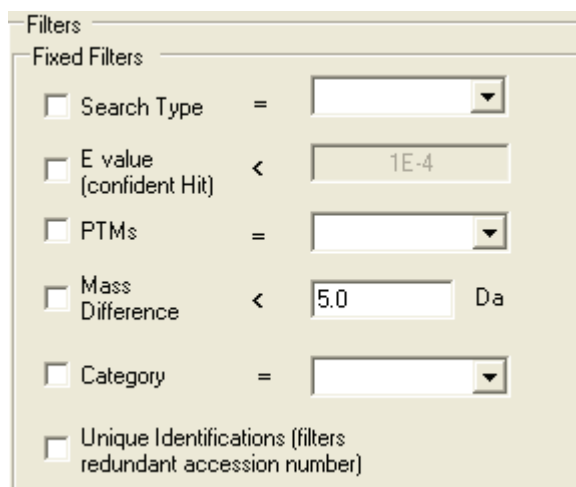
2. Click **Apply Filters**.

❖ **To apply existing custom filters**

1. In the Custom Filters section of the Actions area, select the **Show Custom Filters** check box.

   The Custom Filters section now expands, as shown in Figure 39.

Figure 39. Expanded Custom Filters section
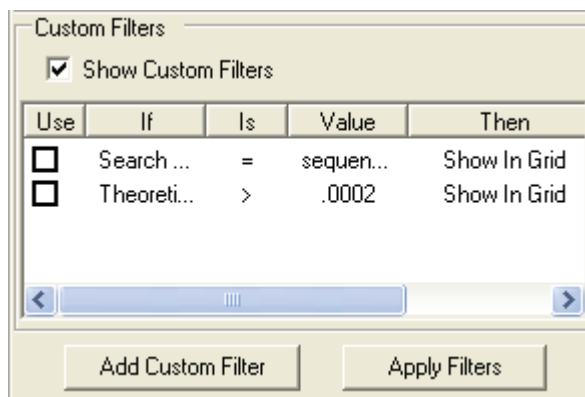
2. In the Use column, select the filter that you want to apply.

3. Click **Apply Filters**.

❖ **To add a new custom filter**

1. In the Custom Filters section of the Actions area, select the **Show Custom Filters** check box.

The Custom Filter section expands.

2. Click **Add Custom Filter**.

The Custom Filters section now resembles the section shown in Figure 40.

Figure 40. Expanded Custom Filters section
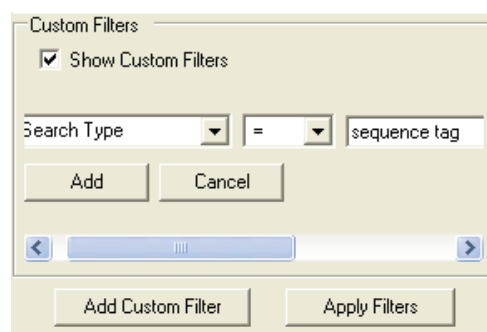
3. From the list on the left, select the parameter.

The parameters in this list are the same as the column names described in "Changing the Display Columns in the Repository Report" on page 81.

4. From the middle list, select an operator:

   • = Equal to

   • < Less than

   • > Greater than

- • <= Less than or equal to

- • >= Greater than or equal to

- • Not = Not equal to

5. In the box on the right, specify an appropriate value.

6. Click **Add**.

   The Custom Filters section now looks like the illustration shown in Figure 39.

7. In the Use column, select the filter that you want to apply.

8. Click **Apply Filters**.

Demonstrating Repository Report Generation

The following demonstration shows you how to generate a repository report, filter its data, and save the data to an Excel file.

Using ProSightPC's Advanced Wizard Functions

This section describes how to set custom processing options and how to use ProSightPC's repositories in the High Throughput wizard.

Using Custom Settings

If you do not want to use the predefined default settings for the Middle Down, Top Down (MS3), and Top Down (MS2) process options in the Choose a Process Option area in the High Throughput wizard, you can define your own custom settings for these options.

❖ **To set custom processing options**

1. On the Process a dataset page of the High Throughput wizard, shown on Figure 26 on page 59, click **Advanced Settings**.

   The Advanced Settings dialog box opens, as shown in Figure 41.

   Figure 41. Advanced Settings dialog box
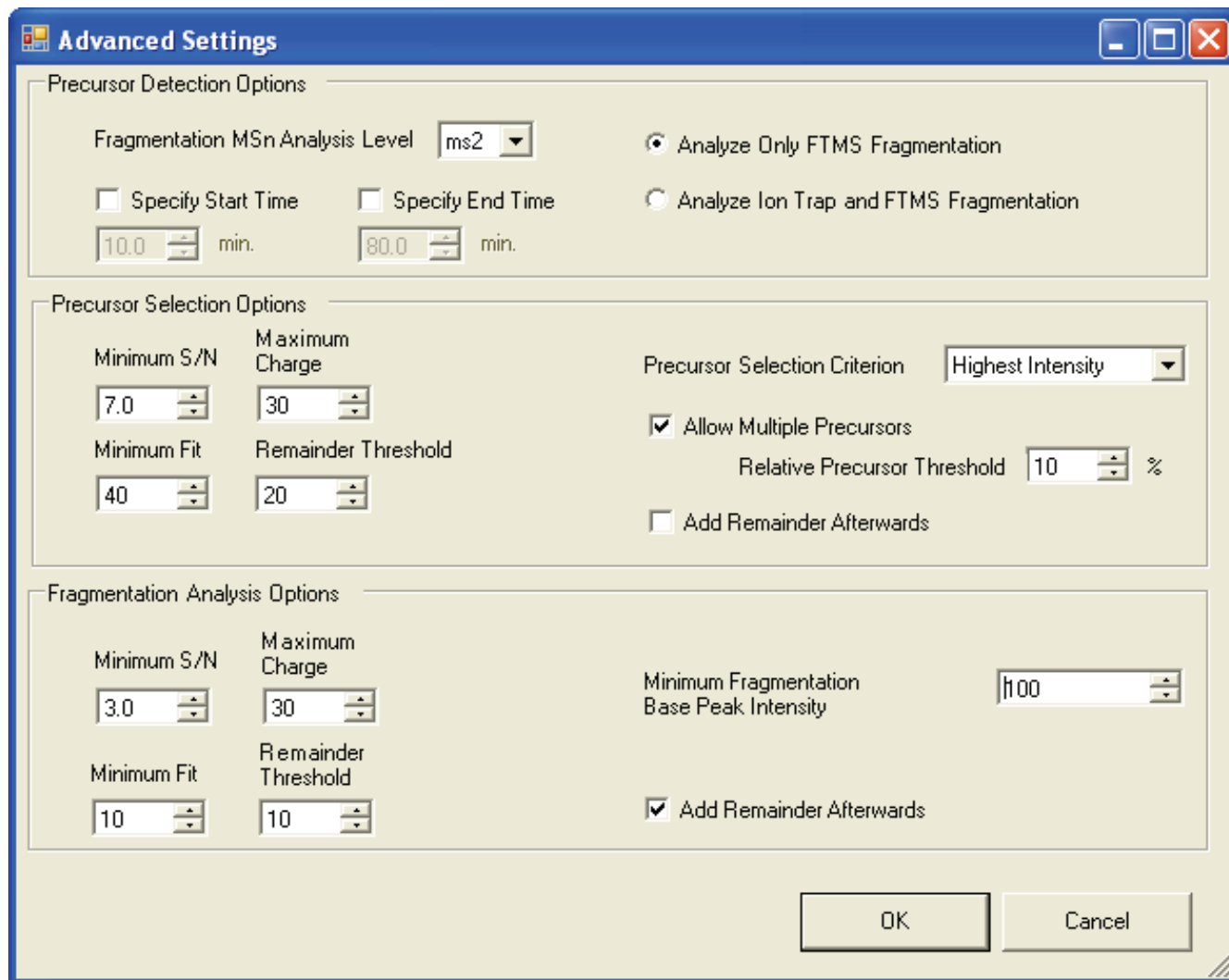
2. In the Precursor Detection Options section, specify the fragmentation scans in the .raw file where ProSightPC infers the precursor scan.

   a. From the Fragmentation MSn Analysis Level list, select the level of analysis that includes your fragmentation data:

   • ms2: For data-dependent LC-MS/MS experiments

   • ms3: For ion-trap marching experiments

b. To specify the start of the chromatographic time scale in which to analyze the data, select the **Specify Start Time** check box and enter the start time in the box underneath it.

This is the start of the first scan. The default is 10 minutes.

c. To specify the end of the chromatographic time scale in which to analyze the data, select the **Specify End Time** box and enter the end time in the box underneath it.

This is the end of the first scan. The default is 80 minutes.

If you do not specify a time range, every scan in the .raw file is analyzed.

d. If you want ProSightPC to process only fragmentation data from Thermo Scientific Fourier Transform instruments, select the **Analyze Only FTMS Fragmentation** option.

e. If you want ProSightPC to process fragmentation data from Thermo Scientific Fourier Transform and ion-trap instruments, select the **Analyze Ion Trap and FTMS Fragmentation** option.

3. In the Precursor Selection Options section, specify the THRASH parameters for analyzing precursor ions.

a. In the Minimum S/N box, enter the lowest signal-to-noise ratio that the THRASH algorithm will consider when trying to assign neutral mass to a charged (mass-to-charge ratio [*m/z*]) species.

The default is 7.

b. In the Maximum Charge box, enter the maximum charge to be used by the THRASH algorithm.

The default is 25.

c. In the Minimum Fit box, enter the minimum fit parameter used by the THRASH algorithm.

The default is 0.90.

d. In the Remainder Threshold box, enter the remainder of the fit that is left in the scan.

The Remainder Threshold option (as a percentage) determines whether a packet is further processed after an averagine pattern is subtracted. The allowed range is 0-100. The Remainder Threshold option is important if overlapping peaks are analyzed. If there is an overlapping pattern of two peptides and the first pattern has been identified, the first averagine pattern is subtracted. The remaining pattern is only processed if its peaks (the remainder) have an intensity that is greater than that specified by the Remainder Threshold option. Setting the Remainder Threshold option to 100 percent disables deconvolution of overlapping patterns. ProSightPC recognizes only the first, most intense pattern and ignores overlapping, less intense patterns. Setting Remainder Threshold to 10 percent allows the deconvolution of a peptide, even if it is overlapped by a peptide pattern with 10-fold intensity.

e. In the Precursor Selection Criterion list, select the type of precursor mass to use for searching:

- Highest Intensity: The precursor mass to use for searching is that of the most abundant ion in the precursor scan.

- Closest Average m/z: The precursor mass to use for searching is the closest to the mass-to-charge ratio (*m/z*) of the data-dependent scan.

The default is Highest Intensity.

f. If the data is intentionally multiplexed, do the following:

- Select the Allow Multiple Precursors check box so that fragmentation data can be multiplexed. If two different ions are fragmented at the same time in the mass spectrometer, you can use this setting to search both precursor ions against the same set of fragment ions.

- In the Relative Precursor Threshold box, enter a threshold cut-off percentage for the data. For example, suppose that the precursor scan contains three ions. Ion A is the major ion at 100 percent, ion B is lower at 15 percent, and ion C is very low at 3 percent. If you set the threshold at 10 percent, ProSightPC searches ions A and B with the fragmentation data but does not search ion C.

ProSightPC handles multiplexed scoring natively when you select the Allow Multiple Precursors option. Multiplexed scoring supports the detection of multiple precursors. When calculating the score for each precursor in one experiment (if you have multiple precursors), it optimizes the scoring as if there were only one precursor in the experiment. Sometimes when an experiment contains multiple precursors and some fragments match one precursor and other fragments match different precursors, a better and more accurate score results if the fragments that matched the others were removed, as if there had been one precursor.

g. Select the **Add Remainder Afterwards** check box if you want to add the remaining intensities to the output spectrum.

If a pattern is identified during the processing of the input file with the Xtract algorithm, the corresponding averagine pattern is subtracted from the input spectrum. The remaining intensities, or remainders, are then processed again with the Xtract algorithm so that Xtract can find an overlapping low-intensity pattern. If there is no overlapping second pattern but a small spike in the first pattern, the spike is not visible in the deconvoluted spectrum but will show up in the remainder spectrum, unless you used Add Remainder Afterwards. When you select Add Remainder Afterwards, the spike shows up in the deconvoluted spectrum, because unassigned remainders are added to the corresponding pattern.

4. In the Fragmentation Analysis Options section, specify the THRASH parameters for analyzing fragment ions.

a. In the Minimum S/N box, enter the lowest signal-to-noise ratio that the THRASH algorithm will consider when trying to assign neutral mass to data in your .raw files.

Set this option to 1 unless offline data is collected. The default is 3.

b. In the Maximum Charge box, enter the maximum charge to be used by the THRASH algorithm.

The default is 25.

c. In the Minimum Fit box, enter the minimum fit parameter used by the THRASH algorithm.

The default is 0.90.

d. In the Minimum Fragmentation Base Peak Intensity box, enter a value that will filter out noise and poor-quality data.

The default setting is 500, which corresponds to an NL value of 5e2. When a mass spectrometer is trying to fragment precursors, the data quality is often poor for some of them. If the fragmentation scan's base peak is below an intensity of 500, ProSightPC skips the scan and discards the precursor mass. This step eliminates the processing of bad MS2 experiments and prevents the analysis of noise, so it makes the application more efficient and increases the speed and performance of the searching.

If you are not certain what to select for this option, use the default setting.

e. If you want to add the remaining intensities to the output spectrum, select the **Add Remainder Afterwards** check box.

f. Click **OK**.

Advanced Settings Dialog Box Parameters

The Advanced Settings dialog box, shown in Figure 41 on page 90, contains the following parameters:

Parameter	Description
Fragmentation MS*n* Analysis Level	Specifies the level of analysis that includes your fragmentation data in the scan. You can select one of the following: • ms2: For data-dependent LC-MS/MS experiments • ms3: For ion-trap marching experiments
Specify Start Time	Specifies the start of the chromatographic time range in which to analyze the data. This is the start of the first scan.
Specify End Time	Specifies the end of the chromatographic time range in which to analyze the data. This is the end of the first scan.
Analyze Only FTMS Fragmentation	Indicates that ProSightPC is to process only high-resolution fragmentation data from Thermo Scientific Fourier Transform instruments.

Parameter	Description
Analyze Ion Trap and FTMS Fragmentation	Indicates that ProSightPC is to process both high- and low-resolution fragmentation data from Thermo Scientific Fourier Transform and ion-trap instruments.
Minimum S/N	Specifies the lowest signal-to-noise ratio that the THRASH algorithm will consider when trying to assign neutral mass to a charged (mass-to-charge ratio [m/z]) species while analyzing the precursor ions.
Maximum Charge	Specifies the maximum charge to be used by the THRASH algorithm while analyzing the precursor ions. The default is 25.
Minimum Fit	Specifies the minimum fit parameter used by the THRASH algorithm while analyzing the precursor ions. The default is 0.90.
Remainder Threshold	Specifies the remainder of the fit that is left in the scan during analysis of the precursor ions.
Precursor Selection Criterion	Specifies the type of precursor mass to use for searching: • Highest Intensity: The precursor mass to use for searching is that of the most abundant ion in the precursor scan. • Closest Average m/z: The precursor mass to use for searching is the closest to the mass-to-charge ratio (m/z) of the data-dependent scan. The default is Highest Intensity.
Allow Multiple Precursors	Allows fragmentation data to be multiplexed. If two different ions are fragmented at the same time in the mass spectrometer, you can search both precursor ions against the same set of fragment ions.
Relative Precursor Threshold	Specifies the threshold cut-off percentage for the data.
Add Remainder Afterwards	Adds the remaining intensities to the output spectrum during analysis of the precursor ions.
Minimum S/N	Specifies the lowest signal-to-noise ratio that the THRASH algorithm will consider when trying to assign neutral mass to a charged (mass-to-charge ratio [m/z]) species while analyzing the fragment ions. The default is 3.
Maximum Charge	Specifies the maximum charge to be used by the THRASH algorithm while analyzing the fragment ions. The default is 25.
Minimum Fit	Specifies the minimum fit parameter used by the THRASH algorithm while analyzing the fragment ions. The default is 0.90.
Remainder Threshold	Specifies the remainder of the fit that is left in the scan during analysis of the fragment ions.

Parameter	Description
Minimum Fragmentation Base Peak Intensity	Specifies a value that will filter out noise and poor-quality data during analysis of the fragment ions.
Add Remainder Afterwards	Adds the remaining intensities to the output spectrum during analysis of the fragment ions.

Using Repositories

ProSightPC places the output of its searches in repositories. A repository can store millions of matches. You can have an arbitrary number of repositories on any ProSightPC installation (one per project, for example). Each experiment is classified in a category.

Creating a Repository

You can create repositories for storing search results.

❖ **To create a repository**

1. Choose **ProSightHT > Edit/Add Repository** to open the Edit/Add Repositories dialog box, shown in Figure 43, and click **Add New Repository,** or click **New Repository** on the Running HighThroughput page.

   The New Repository dialog box shown in Figure 42 opens.

   Figure 42. New Repository dialog box
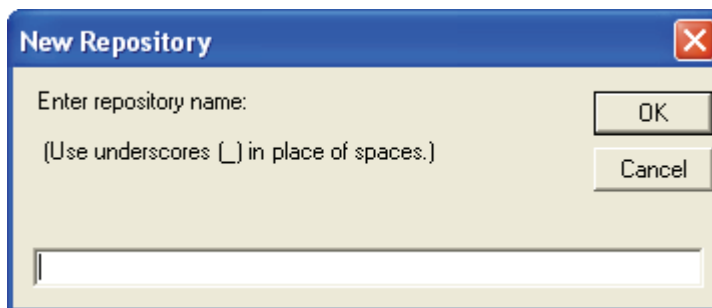
2. Type the name of the new repository. As indicated in the dialog box, do not use spaces or underscores in the repository name.

3. Click **OK**.

   By default, the name and the identifiers of the categories in the results now appear in the Categories box of the Edit/Add Repositories dialog box.

4. (Optional) In the Add Category box of the Edit/Add Repositories dialog box, type any new categories that are included in the results, and click **Add**.

5. Click **Save**.

New Repository Dialog Box Parameters

The New Repository dialog box, shown in Figure 42, specifies a name for the repository that you created.

Editing a Repository

You can edit existing repositories, as well as create new ones.

❖ **To edit a repository**

1. Choose **ProSightHT > Edit/Add Repository** to open the Edit/Add Repositories dialog box, shown in Figure 43.

   Figure 43. Edit/Add Repositories dialog box
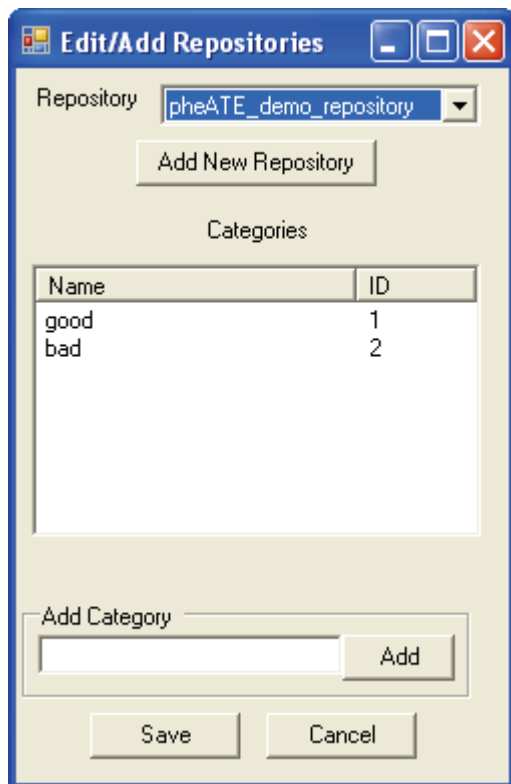
2. Select the repository that you would like to edit from the Repository list.

3. (Optional) In the Add Category box, type any new categories that are included in the results, and click **Add**.

4. Click **Save**.

You can also access the Edit/Add Repositories dialog box in the High Throughput wizard by going to the Running Highthroughput Logic page of the wizard, shown in Figure 27 on page 65, and selecting **Edit Repository** from the Repository list.

Edit/Add Repositories Dialog Box Parameters

The Edit/Add Repositories dialog box, shown in Figure 43, contains the following parameters:

Parameter	Description
Repository	Specifies the name of the repository to edit.
Add New Repository	Opens the New Repository dialog box, shown in Figure 42, so that you can add a new repository.
Categories: Name	Lists the names of the categories in the repository.
Categories: ID	Lists the identifiers of the categories in the repository.
Add Category	Specifies the name of the category to add to the repository.
Add	Adds the specified category to the repository.
Save	Saves the new or edited repository.

Importing Experiments from a Repository

You can import experiments from a repository into ProSightPC so that you can perform operations on them, such as adding or changing searches or using the Sequence Gazer.

❖ **To import experiments from a repository**

1. Choose **File > Import Data from Repository** to open the Import Data From Repository dialog box.

2. In the Repository list, select the name of the repository from which to import experiments into ProSightPC.

3. In the Category list, select the category of results to import.

   The categories available in the list include the default "good" and "bad" categories, as well as any that you defined.

4. In the File list, select the name of the .puf or .raw file whose data formed the basis of the experiments.

5. In the Experiments box, select the experiments to import. Click **Select All** to choose all the experiments in the repository.

   The Import Data from Repository dialog box should now resemble the example in Figure 44.

Figure 44. Import Data from Repository dialog box
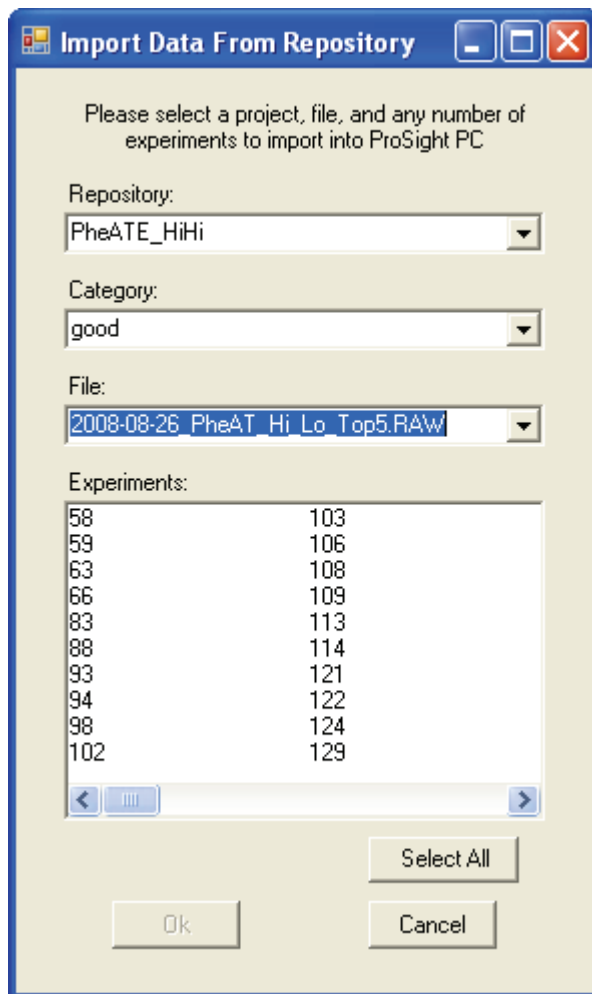
6. Click **OK**.

The experiments are now listed in the Data Grid.

For information on importing experiments into ProSightPC by using the repository report, see "Using the Repository Report to Import Experiments from a Repository" on page 83.

Import Data from Repository Dialog Box Parameters

The Import Data from Repository dialog box, shown in Figure 44, contains the following parameters:

Parameter	Description
Repository	Specifies the repository from which to import the experiments.
Category	Specifies the category of experiments in the repository to import. This list is not available unless you select the repository first. The categories available include the default "good" and "bad" categories, as well as any that you defined.
File	Specifies the .raw or .puf file containing the data on which the experiments were based. This list is not available unless you select the category first.
Experiments	Lists all the experiments in the selected file so that you can select the experiments to import.
Select All	Selects all the experiments in the selected file for importation.

Exporting Experiments to a Repository

You can export the experiments in the ProSightPC Data Grid into a repository. For example, suppose that you processed a .raw file and then viewed the repository report that was generated. You decide to further investigate a specific experiment that you find interesting. You would import it from the repository into ProSightPC and re-run the searches until you obtained good results. Then you save the results back to a repository, either overwriting the current experiment in the same repository or saving the results in a different repository.

❖ **To export experiments to a repository**

1. Choose **File > Export Data to Repository** to open the Export Data to Repository dialog box.

2. In the Experiments box, select the experiments that you want to export. Click **Select All** to choose all the experiments in the repository.

3. In the Repository list, select the name of the repository to which you would like to export the experiments.

4. From Category list, select the category in the repository to which the experiments will be exported.

   The categories available in the list include the default "good" and "bad" categories, as well as any that you defined.

5. (Optional) Select the **Set New File** option if you want to change the .puf or .raw file on whose data the experiments were based.

Otherwise, ProSightPC exports the experiments from the currently open .puf file. If the file is not a .puf file, ProSightPC names it *untitled*.puf.

In the File box that opens, select the name of the file or type the name of the file.

The Export Data to Repository dialog box should now resemble the example in Figure 45.

Figure 45. Export Data to Repository dialog box
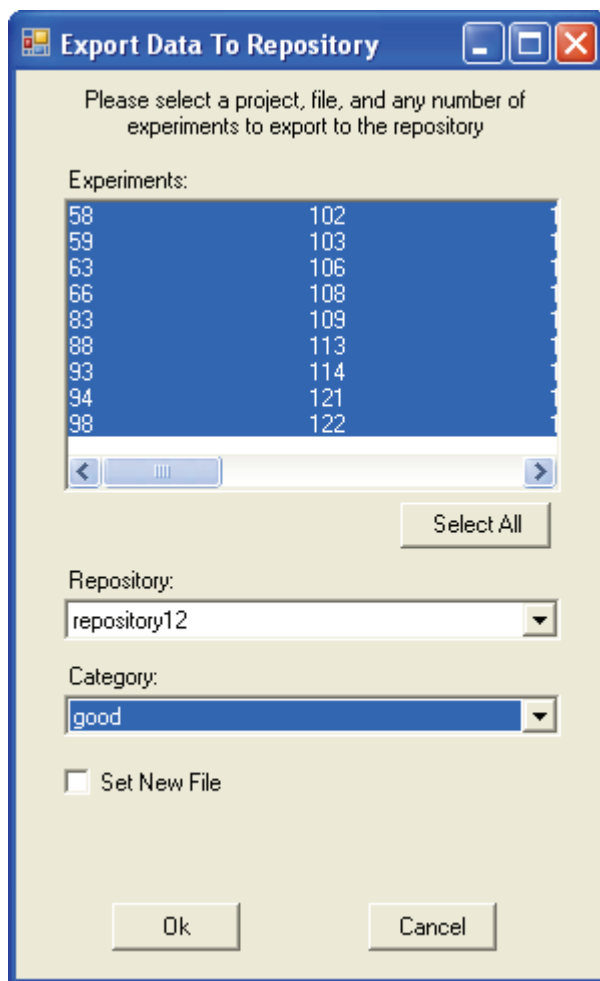
6. Click **OK**.

For information on exporting experiments to a repository by using the repository report, see "Exporting Experiments to a Repository" on page 99.

You can also right-click an experiment in the Data Grid, and choose **Export Experiment to Repository** from the shortcut menu to export it back to the same repository from which it was imported without having to specify the repository name, project name, and file name.

Export Data to Repository Dialog Box Parameters

The Export Data to Repository dialog box, shown in Figure 45, contains the following parameters:

Parameter	Description
Experiments	Lists all the experiments in the selected .raw or .puf file so that you can select the experiments that you want to export to a repository.
Select All	Selects all the experiments in the selected file for importation.
Repository	Specifies which repository to export the experiments to.
Category	Specifies the category in the repository to export the experiments to. The categories available in the list include the default "good" and "bad" categories, as well as any that you defined.
Set New File	Changes the .puf or .raw file on whose data the experiments were based. In the File box that opens, select the name of the file from the list or type the name of the file.

Importing Targeted .raw Files

If you want to import a targeted .raw file as input or if you want to enter data manually into ProSightPC, you cannot use the High Throughput wizard as you can with LC-MS/MS .raw files. You must use the procedures in this section to import targeted .raw files, or use the instructions in "Entering Data Manually" on page 111 if you want to enter data manually.

In order for ProSightPC to identify and characterize proteins, mass spectral data must be converted to neutral mass values. An analysis to infer mass (AIM) is an operation in which high-resolution mass spectral data from proteins or large peptides is converted into neutral monoisotopic or average masses.

> **IMPORTANT** ProSightPC works with neutral masses only.

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

ProSightPC supports three different AIMs:

- Post Xtract: Takes a small file generated by the Xtract algorithm within Qual Browser and uses it as the neutral mass data. This option is the default.

  To use this option, see "Importing a Targeted .raw File with the Post Xtract Option" on page 102.

---

[1]Horn, D. M.; Zubarev, R. A.; and McLafferty, F. W. Automated Reduction and Interpretation of High Resolution Electrospray Mass Spectra of Large Molecules. *J. Am. Soc. Mass Spectrom*. **2000**, *11*, 320–332.

- Profile: Uses the THRASH algorithm to process the input file. This algorithm takes raw mass-to-charge (*m/z*) data and finds the neutral mass values. At its most basic level, the THRASH algorithm infers monoisotopic or average masses from both precursor and data-dependent MS/MS scans and combines these mass lists into experiment sets (precursor mass and its corresponding fragments masses). These lists are then converted into a set of experiments in a ProSightPC upload format (.puf) file for searching with the ProSightPC suite of applications.

  To use this option, see "Importing a Targeted .raw File with the Profile Option" on page 107.

- Manual entry method: To use this option, see "Entering Data Manually" on page 111.

Table 2 summarizes the differences between the THRASH and Post Xtract methods of importing mass values.

Table 2. Comparison of Post Xtract and THRASH methods

AIM equivalency	Data type	Speed	Peaks A	S/N (approximate)
Post Xtract	Profile, centroid	Faster	Slightly more	3:1
THRASH	Profile	Slower	Slightly less	10:1

Importing a Targeted .raw File with the Post Xtract Option

The Post Xtract algorithm averages the data from all fragmentation scans and only analyzes the averaged fragmentation data once. This option reduces analysis and search time and should give better results. This option is the default.

To import a targeted .raw file with the Post Xtract option, follow this procedure. To view a demonstration of this procedure, see "Demonstrating Targeted .raw File Importation with Post Xtract" on page 104.

❖ **To import a targeted .raw file with the Post Xtract option**

1. Choose **File > Import .raw. > Post Xtract**, or click the **Import Xtract** icon, 🔍.

   The Build Experiment from Post Xtract RAW Data dialog box appears, as shown in Figure 46.

Figure 46. Build Experiment from Post Xtract RAW Data
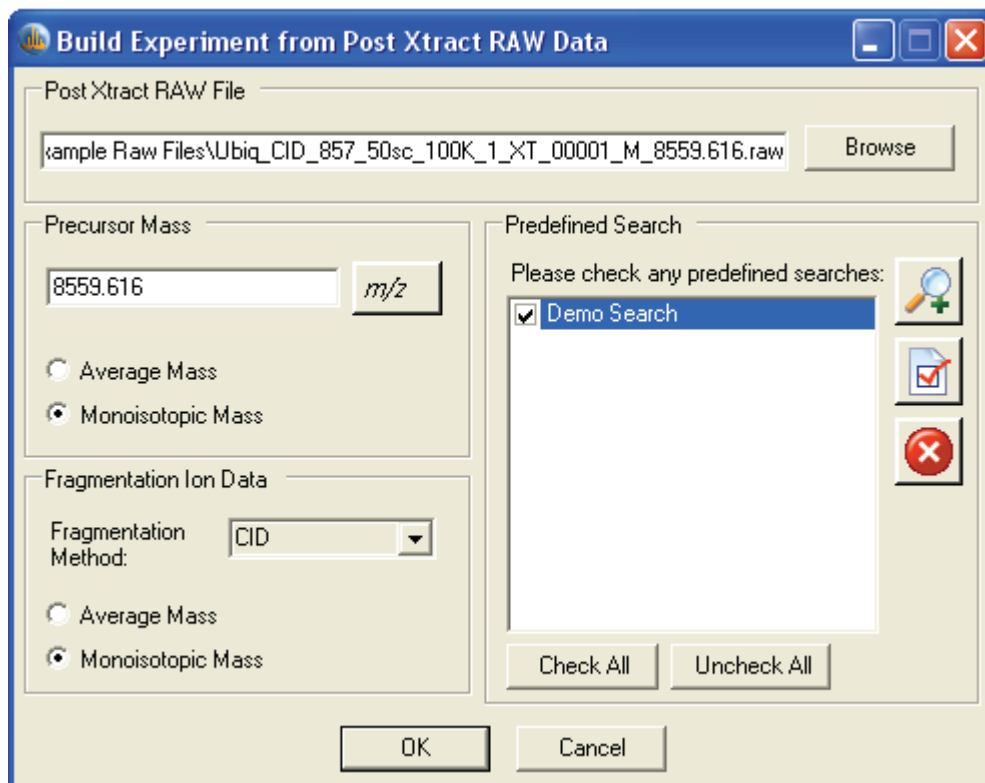
2. In the Post Xtract RAW File box, type the path and name of the .raw file that you want to import, or click **Browse** to browse for the file.

3. In the box in the Precursor Mass area, enter the mass of the precursor ion.

4. (Optional) Click *m/z* to have ProSightPC calculate the intact mass if you know only the mass-to-charge ratio and the charge.

The Intact Mass Calculator dialog box appears, as shown in Figure 47.

Figure 47. Intact Mass Calculator dialog box
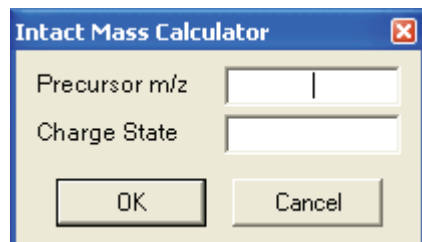
a. In the Precursor m/z box, enter the mass-to-charge ratio (*m/z*) value of the precursor ion.

b. In the Charge State box, enter the charge state, z, to assign to the mass-to-charge (*m/z*) data found in the data files.

c. Click **OK**.

5. In the Precursor Mass area, select the mass type of the precursor ions:

- Average Mass: Specifies that the precursor mass is monoisotopic, which is the mass of the protein, peptide, or fragment ion, where all carbons are carbon-12. The uncharged average mass data is scan 1 in the Post Xtract file.

- Monoisotopic Mass: Specifies that the precursor mass is the mass of the most abundant isotope of the protein, peptide, or fragment ion. The uncharged monoisotopic mass data is scan 2 in the Post Xtract file.

6. In the Fragmentation Method list, select one of the following fragmentation methods:

- CID

- ECD

- IRMPD

- HCD

- ETD

For information on these methods, see "Fragmentation Methods" on page 8.

7. In the Fragmentation Ion Data area, select the mass type of the fragment ions:

- Average Mass Data: Specifies that the fragment mass is the mass of the most abundant isotope of the protein, peptide, or fragment ion.

- Monoisotopic Mass Data: Specifies that the fragment mass is monoisotopic, which is the mass of the protein, peptide, or fragment ion, where all carbons are carbon-12.

8. (Optional) Select a predefined search in the Predefined Search area. Select **Check All** to select all of the searches or **Uncheck All** to clear all searches.

9. Click **OK**.

Demonstrating Targeted .raw File Importation with Post Xtract

The following demonstration shows you how to import a targeted .raw file with the Post Xtract option.

Build Experiment from Post Xtract RAW Data Dialog Box Parameters

The Build Experiment from Post Xtract RAW Data dialog box, shown in Figure 46 on page 103, contains the following parameters:

Parameter	Description
Post Xtract RAW File	Specifies the path and name of the .raw file that you want to import. You can also click Browse to find the file.
Precursor Mass	Specifies the mass of the precursor ion. • m/z: Calculates the intact mass if only the mass-to-charge ratio and the charge are known. It opens the Intact Mass Calculator dialog box, shown in Figure 47 on page 103. • Average Mass: Specifies that the precursor mass is monoisotopic, which is the mass of the protein, peptide, or fragment ion, where all carbons are carbon-12. The uncharged average mass data is scan 1 in the Post Xtract file. • Monoisotopic Mass: Specifies that the precursor mass is the mass of the most abundant isotope of the protein, peptide, or fragment ion. The uncharged monoisotopic mass data is scan 2 in the Post Xtract file.

Fragmentation Ion Data	• Fragmentation Method: Specifies one of the following fragmentation methods:
	
	    • CID
	
	    • ECD
	
	    • IRMPD
	
	    • HCD
	
	    • ETD
	
	For information on these methods, see "Fragmentation Methods" on page 8.
	
	• Average Mass: Specifies that the fragment mass is monoisotopic, which is the mass of the protein, peptide, or fragment ion, where all carbons are carbon-12.
	
	• Monoisotopic Mass: Specifies that the fragment mass is the mass of the most abundant isotope of the protein, peptide, or fragment ion.

Predefined Search	Displays the predefined searches set for an experiment.
Check All	Selects all predefined searches.
Uncheck All	Does not select any predefined searches.

Intact Mass Calculator Dialog Box Parameters

The Intact Mass Calculator dialog box, shown in Figure 47 on page 103, contains the following parameters:

Parameter	Description
Precursor m/z	Specifies the mass-to-charge ratio (*m/z*) value of the precursor ion.
Charge State	Specifies the charge state, z, to assign to the mass-to-charge (*m/z*) data found in the data files.

Importing a Targeted .raw File with the Profile Option

The Profile option applies the THRASH algorithm to the importation of mass values.

❖ **To import a targeted .raw file with the Profile Option**

1. Choose **File > Import .raw > Profile**, or click the **Import Profile** icon, .

The Build Experiment from Profile RAW Data dialog box opens, as shown in Figure 48.

Figure 48. Build Experiment from Profile RAW Data dialog box

2. To specify a .raw file to import, type the full path name in the box labeled "'RAW' file to be 'THRASHed,'" or click **Browse** to select the file name from the list.

3. In the Minimum Signal-to-Noise Ratio box, type the lowest signal-to-noise ratio that the Profile algorithm will consider when trying to assign neutral mass to data in your .raw files.

   Values less than 5:1 significantly slow down the analysis but can result in a greater number of identified isotopic envelopes.

4. In the Maximum Mass box, type the cutoff point for the THRASH algorithm when searching for masses.

5. In the First m/z box, type the lowest mass-to-charge ratio (*m/z*) value considered.

6. In the Minimum RL value box, type the minimum confidence level.

7. In the Maximum Charge box, type the maximum charge to be used by the THRASH algorithm.

8. In the Last m/z box, type the highest mass-to-charge ratio (*m/z*) value considered.

9. In the Summing Options area, type the first scan number scanned in the Start Scan Number box and the last scan number scanned in the End Scan Number box.

10. In the Type list in the Precursor Mass area, select the mass type:

    • Monoisotopic: Specifies that the precursor mass is monoisotopic, which is the mass of the protein, peptide, or fragment ion, where all carbons are carbon-12.

    • Average mass: Specifies that the precursor mass is the mass of the most abundant isotope of the protein, peptide, or fragment ion.

11. (Optional) Click **m/z** to have ProSightPC calculate the intact mass if only the mass-to-charge ratio and the charge are known.

    The Intact Mass Calculator dialog box appears, as shown in Figure 49.

    Figure 49. Intact Mass Calculator dialog box
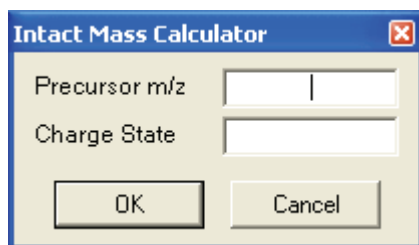
    a. In the Precursor m/z box, type the mass-to-charge ratio (*m/z*) value of the precursor ion.

    b. In the Charge State box, type the charge state, z, to assign to the mass-to-charge (*m/z*) data found in the data files.

    c. Click **OK**.

12. In the Fragmentation Method list, select one of the following fragmentation methods:

    • CID

    • ECD

    • IRMPD

    • HCD

    • ETD

    For information on these methods, see "Fragmentation Methods" on page 8.

13. (Optional) In the Predefined Search box, add a predefined search to the new MS/MS experiment by selecting the check box next to the search name.

14. Click **OK**.

You can change the THRASH default preferences by using the Options dialog box.

❖ **To change the THRASH default preferences**

- Choose **Tools > Options > Thrash**.

    The THRASH Preferences page of the Options dialog box opens.

For more information about changing THRASH defaults, see "Setting THRASH Preferences" on page 37.

Build Experiment from Profile RAW Data Dialog Box Parameters

The Build Experiment from Profile RAW Data dialog box, shown in Figure 48 on page 108, contains the following parameters:

Parameter	Description
"RAW" file to be "THRASHed"	Specifies the name of the .raw file to import.
Minimum Signal-to-Noise Ratio	Specifies the lowest signal-to-noise ratio that the THRASH algorithm will consider when trying to assign neutral mass to data in your .raw files.
Maximum Mass	Specifies the cutoff point for the THRASH algorithm when searching for masses.
First m/z	Specifies the lowest mass-to-charge ratio (*m/z*) value considered.
Minimum RL value	Specifies the minimum confidence level.
Maximum Charge	Specifies the maximum charge to be used by the THRASH algorithm.
Last m/z	Specifies the highest mass-to-charge ratio (*m/z*) value considered.
Start Scan Number	Specifies the first scan number scanned.
End Scan Number	Specifies the last scan number scanned.
Precursor Mass	Specifies the mass of the precursor ion.
Type	Specifies the type of precursor ion mass: • Monoisotopic: Specifies that the precursor mass is monoisotopic, which is the mass of the protein, peptide, or fragment ion, where all carbons are carbon-12. • Average: Specifies that the precursor mass is the mass of the most abundant isotope of the protein, peptide, or fragment ion.

Parameter	Description
m/z	Calculates the intact mass if only the mass-to-charge ratio and the charge are known. It opens the Intact Mass Calculator dialog box, shown in Figure 47.
Fragmentation Method	Specifies one of the following fragmentation methods: • CID • ECD • IRMPD • HCD • ETD For information on these methods, see "Fragmentation Methods" on page 8.
Predefined Search	Displays the predefined searches set for an experiment.
Check All	Selects all predefined searches.
Uncheck All	Does not select any predefined searches.

Entering Data Manually

As a third option, you can import MS/MS experiment data by manually entering the data from ProSightPC through the Tools menu. The data is then displayed in the Data Grid. You can also manually delete experiments.

❖ **To manually import MS/MS experiment data**

1. Choose **Tools > Experiment Adder**, or click the **Add Experiment** icon, 📇.

The Experiment Adder dialog box opens, as shown in Figure 50.

Figure 50. Experiment Adder dialog box
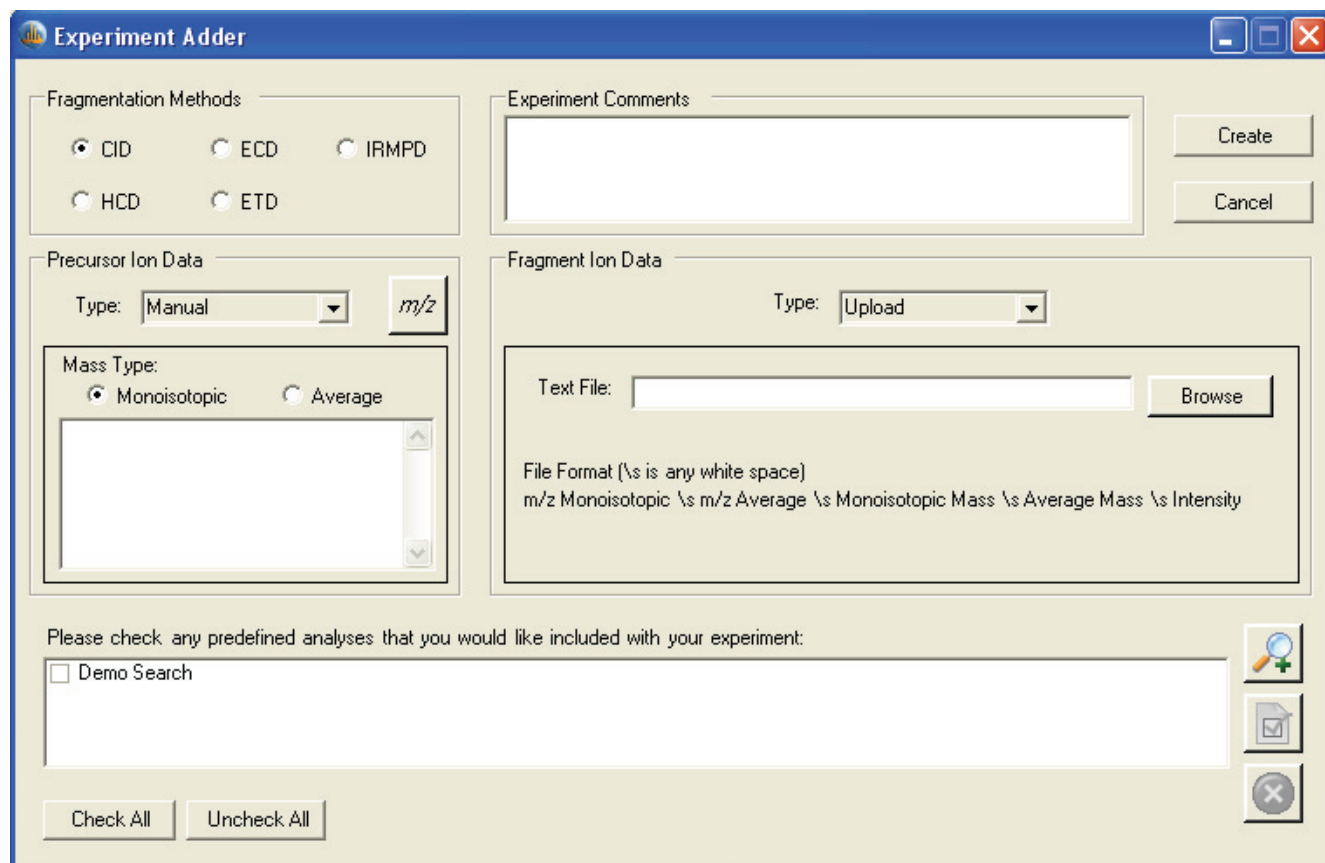
2. In the Fragmentation Methods area, select one of the following fragmentation methods:

   • CID

   • ECD

   • IRMPD

   • HCD

   • ETD

   For information on these methods, see "Fragmentation Methods" on page 8.

3. In the Precursor Ion Data area, select the method of inputting the precursor ion data.

   a. In the Type list, select **Manual** or **Upload**.

   b. If you select Manual in the Precursor Ion Data Type list, select the mass type of the precursor ion in the Mass Type box and enter the precursor mass in the box beneath it:

      • Monoisotopic: Specifies that the precursor mass is monoisotopic, which is the mass of the protein, peptide, or fragment ion, where all carbons are carbon-12.

- Average: Specifies that the precursor mass is the mass of the most abundant isotope of the protein, peptide, or fragment ion.

If you select Upload in the Precursor Ion Data Type list, enter the path and name of the ASCII text file or files containing the precursor ion data in the Text File box, or click **Browse** to browse for them.

These files must be properly formatted.

c. (Optional) Click *m/z* to have ProSightPC calculate the intact mass if only the mass-to-charge ratio and the charge are known.

The Intact Mass Calculator dialog box opens, as shown in Figure 47.

i. In the Precursor m/z box, type the mass-to-charge ratio (*m/z*) value of the precursor ion.

ii. In the Charge State box, type the charge state, z, to assign to the mass-to-charge (*m/z*) data that was found in the data files.

iii. Click **OK**.

4. (Optional) In the Experiment Comments box, enter any comments to help you remember or understand details about the experiment that you just added.

5. In the Fragment Ion Data area, select the method of inputting the fragment ion data.

a. In the Type list, select **Manual** or **Upload**.

b. If you select Manual in the Fragment Ion Data Type box, select the mass type of the fragment ion in the Mass Type area:

- Monoisotopic: Specifies that the fragment mass is monoisotopic, which is the mass of the protein, peptide, or fragment ion, where all carbons are carbon-12.

- Average: Specifies that the fragment mass is the mass of the most abundant isotope of the protein, peptide, or fragment ion.

If you select Upload in the Fragment Ion Data Type box, enter the path and name of the ASCII text file or files containing the precursor ion data in the Text File box, or click **Browse** to browse for them.

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

- Monoisotopic m/z: Specifies the monoisotopic mass-to-charge ratio (*m/z*) value corresponding to the fragment ion.

- Average m/z: Specifies the average mass-to-charge ratio (*m/z*) value corresponding to the fragment ion.

- Monoisotopic Mass: Displays the observed monoisotopic mass of the fragment ion, measured in Da.

- Average Mass: Displays the observed average mass of the fragment ion, measured in Da.

- Intensity: Specifies the abundance of the fragment ion.

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

6. (Optional) In the box listing predefined analyses, select any predefined searches to add.

7. Click **Create**.

ProSightPC now creates a new experiment from all the values entered for intact masses and fragment masses and adds it to the Data Grid. If the experiment already exists in the Data Grid, it receives the next available experiment number.

Experiment Adder Dialog Box Parameters

The Experiment Adder dialog box, shown in Figure 50 on page 112, contains the following parameters:

Parameter	Description
Fragmentation Methods	Specifies one of the following fragmentation methods: • CID • ECD • IRMPD • HCD • ETD For information on these methods, see "Fragmentation Methods" on page 8.
Precursor Ion Data: Type	Specifies the method of inputting the precursor ion data. You can select Manual or Upload from the Type list. • Manual: Inputs the precursor ion data. • Upload: Loads the precursor ion data from an ASCII text file or files.
Precursor Ion Data: m/z	Calculates the intact mass if only the mass-to-charge ratio and the charge are known. It opens the Intact Mass Calculator dialog box, shown in Figure 47 on page 103.

Parameter	Description
Precursor Ion Data: Mass Type	Specifies the mass type of the precursor ion, if you select Manual in the Type list. The mass type can be one of the following: • Monoisotopic: Specifies that the precursor mass is monoisotopic, which is the mass of the protein, peptide, or fragment ion, where all carbons are carbon-12. • Average: Specifies that the precursor mass is the mass of the most abundant isotope of the protein, peptide, or fragment ion.
Precursor Ion Data: Text File	Specifies the path and name of an ASCII text file if you select Upload in the Type list. Enter the path and name of the ASCII text file or files containing the fragment ion data, or click Browse to browse for them. These files must be properly formatted.
Experiment Comments	Displays any comments to help you remember or understand details about the experiment that you just added.
Create	Creates a new experiment from all the values entered for intact mass and fragment masses and adds it to the Data Grid.
Fragment Ion Data: Type	Specifies the method of inputting the fragment ion data. You can select Manual or Upload from the Type list. • Manual: Inputs the precursor ion data. • Upload: Loads the precursor ion data from an ASCII text file or files.
Fragment Ion Data: Mass Type	Specifies the mass type of the fragment ion if you select Manual in the Type list. The mass type can be one of the following: • Monoisotopic: Specifies that the fragment mass is monoisotopic, which is the mass of the protein, peptide, or fragment ion, where all carbons are carbon-12. • Average: Specifies that the fragment mass is the mass of the most abundant isotope of the protein, peptide, or fragment ion.
Fragment Ion Data: Text File	Specifies the path and name of an ASCII text file if you select Upload in the Type list. Enter the path and name of the ASCII text file or files containing the fragment ion data, or click Browse to browse for them. These files must be properly formatted.
Intensities	Specifies the abundance of the fragment ions.

Parameter	Description
Please check any predefined analyses that you would like included with your experiment	Selects any predefined searches to add to an experiment.
🔍	Opens the New Predefined Search dialog box so that you can create a new predefined search.
☑	Opens the Edit Predefined Search dialog box so that you can edit the parameters for the search.
✖	Removes the selected predefined search from the list of predefined searches to add to an experiment.
Check All	Selects all predefined searches listed to add to an experiment.
Uncheck All	Clears all predefined searches listed to add to an experiment.

Importing Experiments

Another way to import data into ProSightPC is to import experiments from a repository. See "Importing Experiments from a Repository" on page 97 and "Using the Repository Report to Import Experiments from a Repository" on page 83 for instructions on this procedure. For more information on handling experiments, see "Working with Experiments" on page 117.

Searching

After you import your data, you might want to search the proteome warehouse for matches. ProSightPC supports six different search modes. Each search mode represents a specific method used to query a proteome database within the proteome warehouse. You can either add a new search or add a predefined search.

To add a new search, see "Adding a Search" on page 126.

To add a predefined search, see "Performing Predefined Searches" on page 163.

Working with Experiments

This chapter explains how to work with the experiments in .puf files.

> **Contents**
>
> - Experiments in .puf Files
> - Creating a New .puf File
> - Opening an Existing .puf File
> - Adding and Removing Experiments in .puf Files
> - Managing Experiments in Multiple Files

Experiments in .puf Files

ProSightPC operates on a single .puf file that contains experiments. When you open the .puf file, ProSightPC uploads it into memory and makes the data visible in the Data Grid. You can use the Data Grid to manage the experiments in a single .puf file. To manage experiments in multiple files, use the Experiment Manager.

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

Creating a New .puf File

❖ **To create a new .puf file**

1. Choose **File > New**, or click the **New** icon, 📄.

   The experiments in any previously opened .puf file disappear from the Data Grid, and you are prompted to add an experiment or import data.

2. Add experiments or import data to the .puf file. See "Adding and Removing Experiments in .puf Files" on page 118.

Opening an Existing .puf File

Only one .puf file can be open at a time.

❖ **To open an existing .puf file**

1. Choose **File > Open**, or click the **Open** icon, .

   –or–

   Choose **File >** *filename*, where *filename* is the name of one of the four most recently opened .puf files listed at the bottom of the File menu.

2. If no other .puf file is open or if a .puf file is open but you have made no changes to it, in the Open a PUF File dialog box, browse to the .puf file that you want to open, select it, and click **Open**.

   –or–

   If a .puf file is already open and you have made changes to it, a prompt box appears.

   a. Click **Yes** to save the open .puf file. Click **No** to discard the changes.

   b. If you clicked Yes, specify the name of the file in the Save As dialog box, anad click **Save**.

   c. In the Open a PUF File dialog box that appears, browse to the file that you want to open or enter its path and name, and click **Open**.

The experiments from the opened .puf file now appear in the Data Grid.

Adding and Removing Experiments in .puf Files

There are several different ways to add experiments to the .puf file.

❖ **To add an experiment to the .puf file and to the Data Grid**

- Import .raw files with the Post Xtract option: See "Importing a Targeted .raw File with the Post Xtract Option" on page 102.

- Import .raw files with the Profile option: See "Importing a Targeted .raw File with the Profile Option" on page 107.

- Use the Experiment Adder: See "Entering Data Manually" on page 111.

- Import data from a repository: See "Importing Experiments from a Repository" on page 97 and "Using the Repository Report to Import Experiments from a Repository" on page 83.

❖ **To remove an experiment from the .puf file and the Data Grid**

1. In the Data Grid, right-click the experiment that you want to delete.

2. From the shortcut menu, choose **Remove Experiment *x***.

3. In the Confirm Delete confirmation box, click **Yes**.

   The experiment is only deleted from the Data Grid, but not from the .puf file, before you choose File > Save.

4. Choose **File > Save**, or click the **Save** icon, 💾.

Managing Experiments in Multiple Files

Use ProSightPC's Experiment Manager to manipulate experiments as objects, copy individual experiments between .puf files, or save them in their own .puf file.

Accessing the Experiment Manager

❖ **To access the Experiment Manager**

• Choose **Tools > Experiment Manager**.

   The Experiment Manager dialog box opens, as shown in Figure 51.

   Figure 51. Experiment Manager dialog box
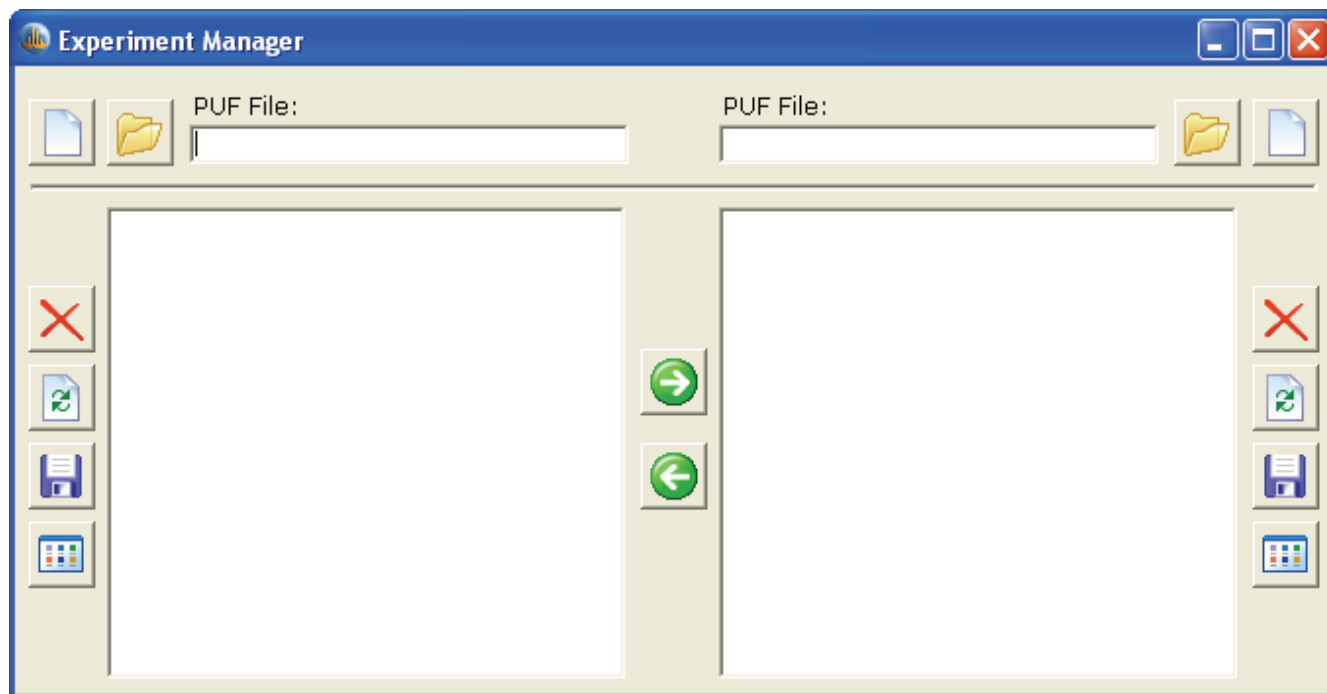
You can either open an existing .puf file or create a new .puf file.

Use the left side of the dialog box to perform operations on the source .puf file and the right side to perform operations on the destination .puf file.

Experiment Manager Dialog Box Parameters

The Experiment Manager dialog box, shown in Figure 51, contains the following parameters:

Parameter	Description
	Creates a new source .puf file (left side) or a new destination .puf file (right side).
	Opens an existing source .puf file (left side) or a destination .puf file (right side).
PUF File	Specifies the name of the source .puf file (left side) or the destination .puf file (right side).
	Removes the selected experiment from the source .puf file (left side) or the destination .puf file (right side).
	Reverts to the last version of the source .puf file saved (left side) or the destination .puf file saved (right side). Any experiments removed after the last time you saved reappear in the pane.
	Saves the source .puf file (left side) or the destination .puf file (right side).
	Opens a popup menu so that you can change how the experiments are displayed in the Experiment Manager dialog box.

- Details: Lists the experiments by number in a single column in the pane. A comment identifying each experiment appears in an adjoining column.

- List: Lists the experiments by number in multiple columns in the pane.

- Small Icons: Lists the experiments from left to right in the pane, using smaller icons than the Large Icons command does.

- Large Icons: Lists the experiments from left to right in the pane, using larger icons than the Small Icons command does.

This icon on the left displays the way experiments in the source .puf file are displayed, and this icon on the right displays the way experiments in the destination .puf file are displayed.

Parameter	Description
	Sends a copy of the selected experiments from the source .puf file to the destination .puf file.
	Sends a copy of the selected experiments from the destination .puf file to the source .puf file.

Creating a New .puf File

You can create a new .puf file by using the Experiment Manager.

❖ **To create a new .puf file**

1. At the top left in the Experiment Manager dialog box, click the **Create New PUF File** icon, , to create a new source .puf file. Click the same icon on the right to create a new destination .puf file.

2. In the dialog box that opens, select the desired directory and type the name of the new .puf file.

3. Click **Save**.

Opening an Existing .puf File

You can open an existing .puf file in the Experiment Manager.

❖ **To open an existing .puf file**

1. Choose **File > Open** or click the **Open Existing PUF File** icon, , on the left to open an existing source .puf file. Click the same icon on the right to open an existing destination .puf file.

2. From the dialog box that opens, select the .puf file from those listed.

3. Click **Open**.

The experiments in the .puf file are now displayed in the appropriate pane of the Experiment Manager, as shown in Figure 52.

Figure 52. Experiments listed in Experiment Manager dialog box
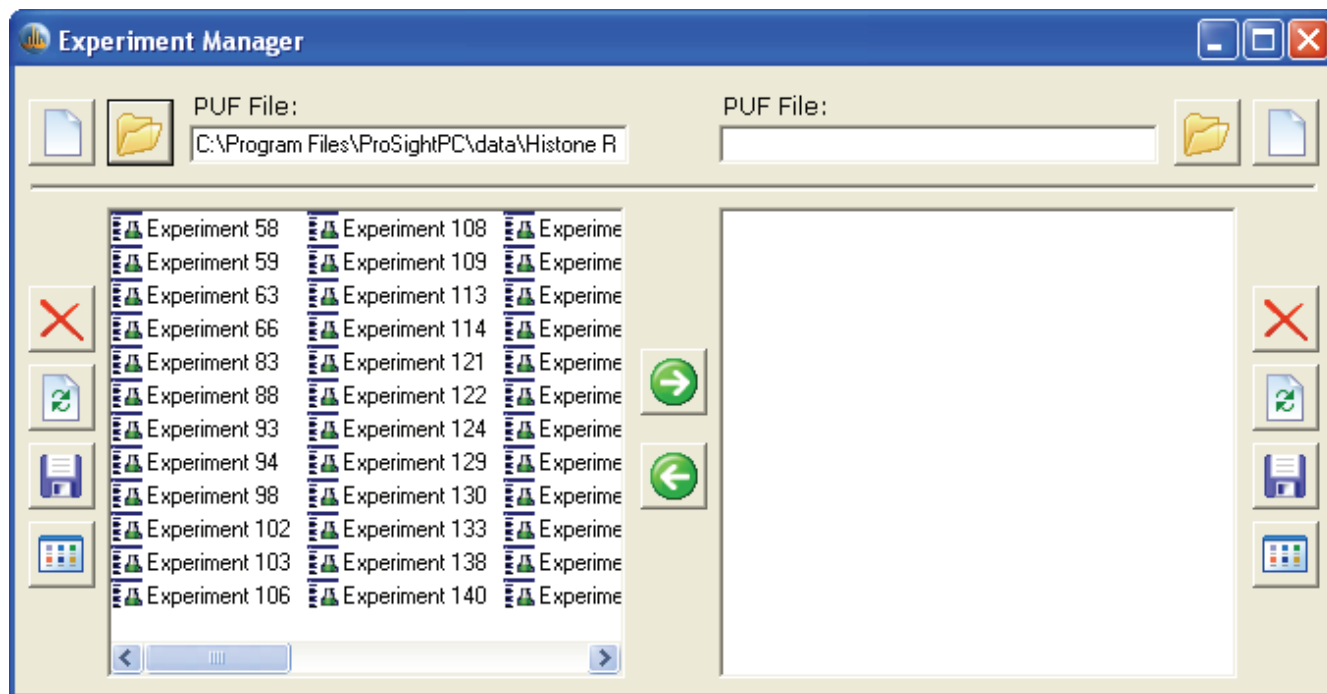
Saving a Changed .puf File

You might want to save a .puf file once you have added, deleted, or copied experiments. You can also revert to the last saved version of the .puf file.

❖ **To save a changed .puf file**

- Click the **Save This PUF** icon, 💾, on the left side of the dialog box to save the source .pdf file. Click the same icon on the right side to save the destination .puf file.

❖ **To revert to the last version of the .puf file saved**

- Click the **Revert to Last Saved** icon, 🔁, on the left side of the dialog box to revert to the last version of the source .puf file saved. Click the same icon on the right side to revert to the last version of the destination .puf file that you saved.

Any experiments removed after the last time you saved reappear in the pane.

Deleting Experiments from a .puf File

You can delete experiments from a .puf file by using the Experiment Manager.

❖ **To delete experiments from a .puf file in the Experiment Manager**

1. In the Experiment Manager dialog box, select the experiments that you want to delete.

2. Click the **Delete Selected Experiments** icon, ✕, on the left side of the dialog box to remove the experiments from the source .puf file. Click the same icon on the right side to remove the experiments from the destination .puf file.

Deleting a .puf File

You cannot delete a .puf file from ProSightPC. In Windows, you can delete it as you would a regular file by right-clicking and selecting **Delete** from the shortcut menu.

Copying Experiments from One .puf File to Another

You can copy experiments from a source .puf file to a destination .puf file using the two panes in the Experiment Manager dialog box.

❖ **To copy an experiment from one .puf file to another**

1. In the source pane of the dialog box, select an experiment to be copied.

2. Drag and drop the experiment from the source pane to the destination pane, or click the green arrow (➡ or ⬅) to send a copy of the experiment to the destination pane.

3. Click the **Save This PUF** icon, 💾, on the left side of the dialog box to save the source .puf file. Click the same icon on the right side to save the destination .puf file.

   If two or more experiments share experiment identification numbers when you are copying experiments to .puf files, a message informs you that the Experiment Manager will reassign the experiment number of the incoming experiment.

4. Click **OK** to confirm**.**

Changing the Experiment Display

You can change how the experiments in the .puf files are displayed in the Experiment Manager dialog box.

❖ **To change the experiment display**

1. Click the **Change View** icon, ▦, on the left side of the dialog box to change the display of the experiments in the left pane. Click the **Change View** icon, ▦, on the right side of the dialog box to change the display of the experiments in the right pane.

2. In the popup menu, select one of the following:

Details	Lists the experiments by number in a single column in the pane. A comment identifying each experiment appears in an adjoining column.
List	Lists the experiments by number in multiple columns in the pane.
Small Icons	Lists the experiments from left to right in the pane, using smaller icons than the Large Icons command does.
Large Icons	Lists the experiments from left to right in the pane, using larger icons than the Small Icons command does.

A heavy dot indicates the active command.

This transaction changes the display of the experiments in both the source and destination .puf files.

Searching Databases

This chapter describes ProSightPC's search modes and how to use them.

Contents

Search Types

ProSightPC has four basic types of searches. Each search mode represents a specific mechanism used to compare imported data to a proteome database in the proteome warehouse. The four search modes are the following:

- Absolute mass search
- Biomarker search

- Sequence tag search

- Single-protein search

In addition, you can use ProSightPC to perform the following types of advanced searches:

- Gene-restricted absolute mass search

- Gene-restricted biomarker search

- MS^n hybrid searches

You can compare MS/MS data to entries in a proteome database by adding a search to the MS/MS experiment containing the MS/MS data.

Adding a Search

You can add a single search to a specific MS/MS experiment when an experiment requires you to adjust parameters to identify the protein.

❖ **To add a search to an experiment**

1. Right-click any search in the experiment listed in the Data Grid and choose **Add New Search**. You can also choose **Experiment Tools > Add Search** or click the **Add Search** icon, .

   The New Search in Experiment *X* dialog box opens, as shown in Figure 53. The *X* in the dialog box title reflects the experiment number.

Figure 53. New Search in Experiment *X* dialog box

2. In the Search Type list, select one of these search types:

- Absolute mass search: See "Searching for Absolute Mass" on page 130.

- Biomarker search: See "Searching for a Biomarker" on page 138.

- Sequence tag search: See "Searching for a Sequence Tag" on page 145.

- Predefined search: See "Performing Predefined Searches" on page 163.

- Single-protein search: See Chapter 6, "Searching for Single Proteins."

New Search in Experiment *X* Dialog Box Parameters

The New Search in Experiment *X* dialog box, shown in Figure 53, contains the following parameters:

Parameter	Description
Search Type	Specifies the type of search to perform: • Absolute mass search: See "Searching for Absolute Mass" on page 130. • Biomarker search: See "Searching for a Biomarker" on page 138. • Sequence tag search: See "Searching for a Sequence Tag" on page 145. • Predefined search: See "Performing Predefined Searches" on page 163. • Single-protein search: See Chapter 6, "Searching for Single Proteins." • Gene-Restricted Absolute Mass Search: See "Searching for a Gene-Restricted Absolute Mass" on page 150. • Gene-Restricted Biomarker Mass Search: See "Searching for a Gene-Restricted Biomarker" on page 156.
Database Description	Describes the database that you want to search.
Precursor Mass Type	Specifies the type of precursor ion mass to use: • Monoisotopic: Specifies that the precursor mass is monoisotopic, which is the mass of the protein, peptide, or fragment ion, where all carbons are carbon-12. • Average: Specifies that the precursor mass is the mass of the most abundant isotope of the protein, peptide, or fragment ion.

Parameter	Description
Precursor Search Window	Specifies a range around the observed precursor mass, in daltons. ProSightPC queries all protein forms with a theoretical mass within this range.
Fragment Mass Type	Specifies the mass type of the fragment ions to use: • Monoisotopic: Specifies that the fragment mass is monoisotopic, which is the mass of the protein, peptide, or fragment ion, where all carbons are carbon-12. • Average: Specifies that the fragment mass is the mass of the most abundant isotope of the protein, peptide, or fragment ion.
Fragment Tolerance	Specifies the tolerance that determines whether the comparison of an observed fragment ion mass to a theoretical fragment ion mass is considered a match and indicates whether it is expressed as absolute, measured in Da, or relative, measured in ppm.
Δm Mode	Determines whether ProSightPC conducts the search in delta (Δm) mode. For details, see "Performing a Search in Delta m Mode" on page 170.
Disulfide	Indicates whether a protein's cysteines are oxidized.
Min # of Matching Fragments	Determines whether the search algorithm finds only proteins containing a minimum number of matching ion fragments. The box to the right specifies the minimum number of matching ion fragments.
Min % of Matching Fragments	Determines whether the search algorithm finds only proteins containing a minimum percentage of matching ion fragments. The box to the right specifies the percentage of matching ion fragments.
Min Score	Determines whether the search algorithm finds only proteins with an e value that matches the filter with the expectation value set in the left list, the operator in the middle list, and an appropriate value in the right box. • $\leq$: Indicates that the first value is less than or equal to the second value. This setting is the default. • $\geq$: Indicates that the first value is greater than or equal to the second value.
Max Proteins to Return	Specifies the maximum number of proteins to return in the search.
Fixed Modifications	Specifies the chemical modifications present on all instances of a given type of amino acid in the observed protein.

Parameter	Description
PTM Handling	Specifies the PTMs that you want queried.
Save	Saves the search information.

Removing a Search

❖ **To remove a search from an experiment**

- In the Data Grid, right-click the search number of an experiment and choose **Remove Search** *x*.

Removing Results and Rerunning a Search

You can remove the results of a specific search in an experiment and rerun the search.

❖ **To rerun a search**

1. Right-click the experiment and choose **Remove Results** from the shortcut menu.

2. In the Confirm Delete message box, click **Yes**.

   ProSightPC now sets the search as a pending search (the search parameters are not removed).

3. Right-click and choose **Run Search** *x*.

Searching for Absolute Mass

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 in the window of the observed precursor ion mass, the absolute mass search compares all theoretical fragments and masses to the observed fragment ion masses. ProSightPC determines the number of observed fragment ions matching the fragment tolerance and uses this value to score the identification. Figure 54 shows this methodology.

Figure 54. Absolute mass searches
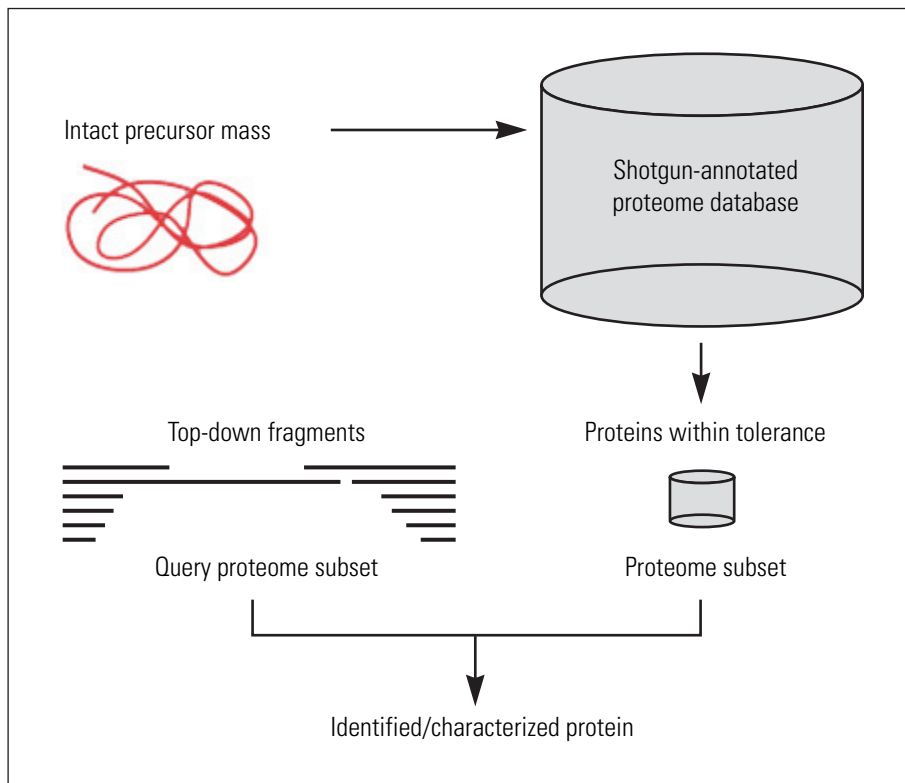
Although ProSightPC queries each protein form with a theoretical precursor mass in the window, only those protein forms meeting user-defined search result filtering (minimum number of matching fragments, minimum percentage of matching fragments, or minimum score) are displayed.

In summary, ProSightPC performs the following steps in a search for absolute mass:

1. In absolute mass searches, ProSightPC queries all protein forms in a user-defined window of the observed intact mass.

2. ProSightPC finds all proteins in the database with intact mass within the tolerance of the search window.

3. For each protein, it calculates all theoretical fragment ions.

4. It compares theoretical fragment ions with observed fragment ions.

5. It calculates its scores. An observed ion is said to match a possible ion if the two masses are within a user-defined tolerance. The probability of the observed number of fragment ions matching by chance is then determined and reported as a p score. For information on the calculation of the p score, see "p Score" on page 186.

This process is shown graphically in Figure 55.

Figure 55. Absolute mass search process
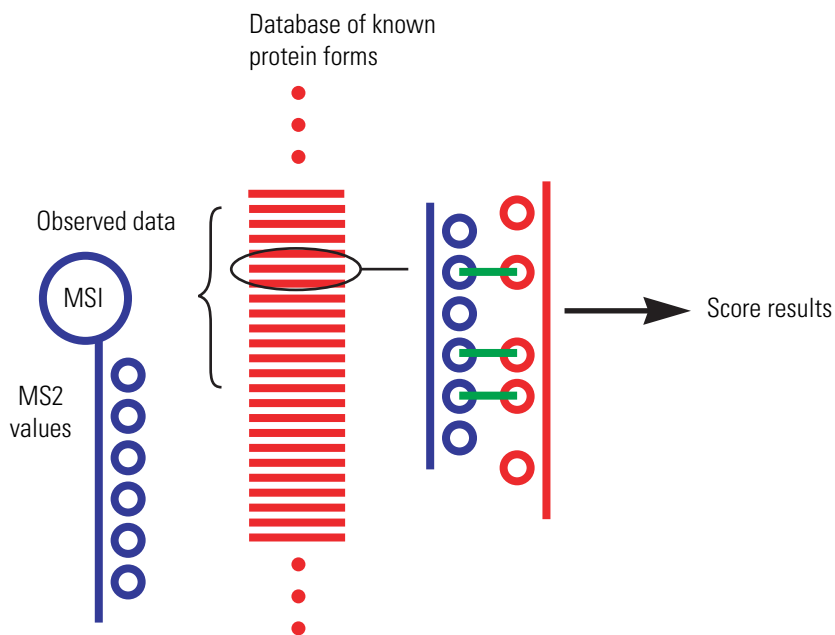
Use the following strategies when running absolute mass searches:

- Use a 1000-Da precursor search window search as the first search for an unknown protein. If there are few modifications on the unknown protein not in the proteome warehouse, a 1000-Da intact search will frequently identify, but not characterize, the protein.

  A large number of ions matching one terminal in a protein is evidence of a protein's identity. You can use the Sequence Gazer for further characterization.

- Use delta (Δm) mode (explained in "Performing a Search in Delta m Mode" on page 170) 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 delta (Δm) mode. Activating delta (Δm) mode increases the likelihood that the search will identify proteins with unknown modifications. However, this mode takes approximately three times longer than the corresponding absolute mass search.

To set preferences for absolute mass searches, see "Setting Absolute Mass Search Preferences" on page 39.

❖ **To search for absolute mass**

1. In the New Search for Experiment *X* dialog box, select **Absolute Mass** in the Search Type list.

   The appearance of the dialog box changes, as shown in Figure 56.

Figure 56. New Search in Experiment *X* dialog box for absolute mass

2. In the Database Description list, select a description of the database that you want to search.

3. From the Precursor Mass Type list, select the type of precursor ion mass to search for:

   • Monoisotopic: Specifies that the precursor mass is monoisotopic, which is the mass of the protein, peptide, or fragment ion, where all carbons are carbon-12.

   • Average: Specifies that the average mass is average, which is the mass of the most abundant isotope of the protein, peptide, or fragment ion.

4. In the Precursor Search Window box, type a range around the observed precursor mass, in daltons.

   The range is one number. For example, if you type 10, ProSightPC queries +10 and -10, meaning that the total range is 20. ProSightPC queries all protein forms with a theoretical mass within this range.

5. In the Fragment Mass Type list, select the type of fragment ion to search for:

   • Monoisotopic: Specifies that the fragment mass is monoisotopic, which is the mass of the protein, peptide, or fragment ion, where all carbons are carbon-12.

   • Average: Specifies that the fragment mass is the mass of the most abundant isotope of the protein, peptide, or fragment ion.

6. In the Fragment Tolerance box, specify the tolerance that determines whether the comparison of an observed fragment ion mass to a theoretical fragment ion mass is considered a match, and indicate whether it is expressed as absolute (measured in daltons) or relative (measured in parts per million).

   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.

7. Select the Δm **Mode** check box if you want to conduct the search in delta (Δm) mode. For more information on delta (Δm) mode, see "Performing a Search in Delta m Mode" on page 170.

8. Select the **Disulfide** check box if you know that the protein's cysteines are oxidized.

   ProSightPC looks for only one disulfide bond.

9. In the Hit Filtering section, set at least one of the following filters; otherwise, ProSightPC returns all protein forms that are searched, even proteins that have no matching fragments.

   a. Select the **Min # of Matching Fragments** check box if you want the search algorithm to find only proteins containing a minimum number of matching ion fragments (these protein forms are called hits). Then specify the minimum number of matching ion fragments in the box to the right.

b. Select the **Min % of Matching Fragments** check box if you want the search algorithm to find only proteins containing a minimum percentage of matching ion fragments. Specify the percentage of matching ion fragments in the box to the right.

c. Select the **Min Score** check box to determine whether the search algorithm finds only proteins with an e value that matches the filter with the expectation value set in the left list, the operator in the middle list, and an appropriate value in the right box. (See "Expectation Value (e value)" on page 186 for more information on the e value.)

- ≤: Indicates that the first value is less than or equal to the second value. This setting is the default.

- ≥: Indicates that the first value is greater than or equal to the second value.

d. From the Max Proteins to Return list, select the maximum number of proteins to return in the search.

With this option, you can truncate the results of a search because the data from all of the similar matching proteins do not need to be returned. You can load the results faster.

10. In the Fixed Modifications box, select no more than one fixed modification per amino acid type.

A fixed modification is a chemical modification that is present on all instances of a given type of amino acid in the observed protein.

11. In the PTM Handling box, select the PTMs that you want to search for.

The PTM Handling box displays PTMs arranged in one or more tiers based on the selected proteome database. ProSightPC only queries theoretical protein forms containing exclusively selected PTMs. Every form containing an unselected PTM is excluded from the interrogation.

12. Click **Save**.

The status of the new search for the experiment in the Pending Search column now changes to "Yes."

13. To execute the search from the Data Manager, right-click the pending search and then click **Run Search**.

New Search in Experiment *X* Dialog Box Parameters for Absolute Mass

The New Search in Experiment *X* dialog box for absolute mass, shown in Figure 56 on page 133, contains the following parameters:

Parameter	Description
Search Type	Specifies the type of search to perform:
	• Absolute mass search: See "Searching for Absolute Mass" on page 130.
	• Biomarker search: See "Searching for a Biomarker" on page 138.
	• Sequence tag search: See "Searching for a Sequence Tag" on page 145.
	• Predefined search: See "Performing Predefined Searches" on page 163.
	• Single-protein search: See Chapter 6, "Searching for Single Proteins."
	• Gene-Restricted Absolute Mass Search: See "Searching for a Gene-Restricted Absolute Mass" on page 150.
	• Gene-Restricted Biomarker Mass Search: See "Searching for a Gene-Restricted Biomarker" on page 156.
Database Description	Describes the database that you want to search.
Precursor Mass Type	Specifies the type of precursor ion mass to use:
	• Monoisotopic: Specifies that the precursor mass is monoisotopic, which is the mass of the protein, peptide, or fragment ion, where all carbons are carbon-12.
	• Average: Specifies that the precursor mass is the mass of the most abundant isotope of the protein, peptide, or fragment ion.
Precursor Search Window	Specifies a range around the observed precursor mass, in daltons. ProSightPC queries all protein forms with a theoretical mass within this range.

Parameter	Description
Fragment Mass Type	Specifies the mass type of the fragment ions to use: • Monoisotopic: Specifies that the fragment mass is monoisotopic, which is the mass of the protein, peptide, or fragment ion, where all carbons are carbon-12. • Average: Specifies that the fragment mass is the mass of the most abundant isotope of the protein, peptide, or fragment ion.
Fragment Tolerance	Specifies the tolerance that determines whether the comparison of an observed fragment ion mass to a theoretical fragment ion mass is considered a match and indicates whether it is expressed as absolute, measured in Da, or relative, measured in ppm.
Δm Mode	Determines whether ProSightPC conducts the search in delta (Δ*m*) mode, which is explained in "Performing a Search in Delta m Mode" on page 170.
Disulfide	Indicates whether a protein's cysteines are oxidized.
Min # of Matching Fragments	Determines whether the search algorithm finds only proteins containing a minimum number of matching ion fragments. The box to the right specifies the minimum number of matching ion fragments.
Min % of Matching Fragments	Determines whether the search algorithm finds only proteins containing a minimum percentage of matching ion fragments. The box to the right specifies the percentage of matching ion fragments.
Min Score	Determines whether the search algorithm finds only proteins with an e value that matches the filter with the expectation value set in the left list, the operator in the middle list, and an appropriate value in the right box. • ≤: Indicates that the first value is less than or equal to the second value. This setting is the default. • ≥: Indicates that the first value is greater than or equal to the second value.
Max Proteins to Return	Specifies the maximum number of proteins to return in the search.
Fixed Modifications	Specifies the chemical modifications present on all instances of a given type of amino acid in the observed protein.
PTM Handling	Specifies the PTMs that you want queried.
Save	Saves the search information.

Searching for a Biomarker

A biomarker search is the "brute force" search of an entire database. It looks at every possible subsequence of every base protein form (unless mentioned otherwise) in the database and attempts to identify any subsequence that has an observed intact ion mass. For each subsequence matching the intact ion mass, the biomarker search then performs an absolute mass search and reports any subsequence that matches the observed intact ion mass and is able to generate the observed fragment ion pattern.

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

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

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

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

Figure 57 shows the process involved in a biomarker search.

Figure 57. Biomarker searches
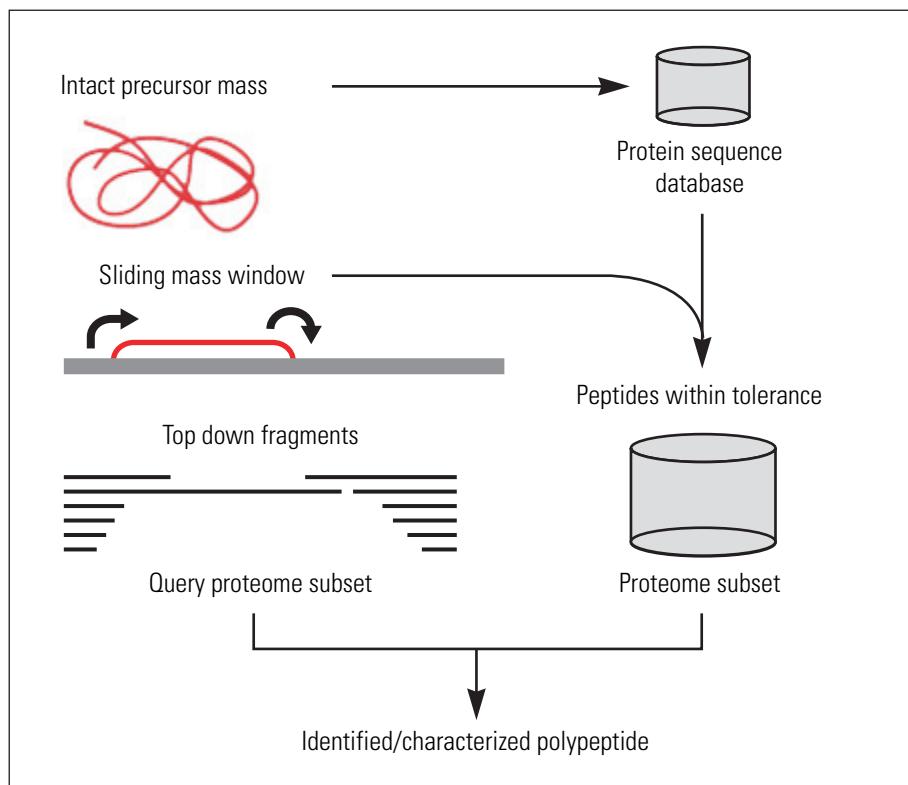
In a biomarker search, the precursor search tolerance is an estimate of measurement error on the observed precursor mass. The value is usually small compared to an absolute mass precursor search window.

Thermo Fisher Scientific recommends the following when you conduct biomarker searches:

- Use a biomarker search if an absolute mass analysis fails to identify a protein. The default biomarker search searches only for the basic forms with no known modifications. If you want to search for modified forms, you must select the Include Modified Forms check box in the New Search in Experiment *X* dialog box for biomarkers (see Figure 58). However, searching for modified forms increases the search run time.

- Increasing the precursor tolerance results in longer run times.

- Biomarker searches are well-suited for identifying biologically relevant proteolytic products.

- You can identify proteins or peptides containing disulfide bonds by setting the precursor search tolerance to 2.5 Da and running the search in delta (Δm) mode, or you can select the Disulfide check box in the New Search in Experiment *X* dialog box for biomarkers (see Figure 58).

To set preferences for biomarker searches, see "Setting Biomarker Search Preferences" on page 43.

❖ **To search for a biomarker**

1. In the New Search for Experiment *X* dialog box, select **Biomarker** from the Search Type list.

The appearance of the dialog box changes, as shown in Figure 58.

Figure 58. New Search in Experiment *X* dialog box for biomarkers

2. In the Database Description list, select a description of the database that you want to search.

3. In the Precursor Mass Type list, select the type of precursor ion mass to use:

   • Monoisotopic: Specifies that the precursor mass is monoisotopic, which is the mass of the protein, peptide, or fragment ion, where all carbons are carbon-12.

   • Average: Specifies that the precursor mass is the mass of the most abundant isotope of the protein, peptide, or fragment ion.

4. In the Precursor Tolerance box, enter the range value for tolerance when testing all protein forms for biomarker peptides. Indicate whether it is expressed as absolute (measured in Da) or relative (measured in ppm).

5. In the Fragment Mass Type list, select the mass type of the fragment ions to use:

   • Monoisotopic: Specifies that the fragment mass is monoisotopic, which is the mass of the protein, peptide, or fragment ion, where all carbons are carbon-12.

   • Average: Specifies that the fragment mass is the mass of the most abundant isotope of the protein, peptide, or fragment ion.

6. In the Fragment Tolerance box, specify the tolerance that determines whether the comparison of an observed fragment ion mass to a theoretical fragment ion mass is considered a match. Indicate whether it is expressed as absolute (measured in daltons) or relative (measured in parts per million).

   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.

7. Select the Δm **Mode** check box if you want to conduct the search in delta (Δm) mode. For details, see in "Performing a Search in Delta m Mode" on page 170.

8. Select the **Disulfide** check box if you know that the protein's cysteines are oxidized.

   ProSightPC looks for only one disulfide bond.

9. Select the **Include Modified Forms** check box if you want to include PTMs and polymorphisms when you perform a biomarker search.

   To detect biomarkers with modifications on them, select this option; however, analysis time increases as a result.

10. In the Hit Filtering section, set at least one of the following filters; otherwise, ProSightPC returns all protein forms that are searched, even proteins that have no matching fragments.

    a. Select the **Min # of Matching Fragments** check box if you want the search algorithm to find only protein forms containing a minimum number of matching ion fragments (these protein forms are called hits). Then specify the minimum number of matching ion fragments in the box to the right.

b. Select the **Min % of Matching Fragments** check box if you want the search algorithm to find only proteins containing a minimum percentage of matching ion fragments. Specify the percentage of matching ion fragments in the box to the right.

c. Select the **Min Score** check box to determine whether the search algorithm finds only proteins with an e value that matches the filter with the expectation value set in the left list, the operator in the middle list, and an appropriate value in the right box. (See "Expectation Value (e value)" on page 186 for more information on the e value.)

- ≤: Indicates that the first value is less than or equal to the second value. This setting is the default.

- ≥: Indicates that the first value is greater than or equal to the second value.

To return only "good" search results in your search, select this option.

d. In the Max Proteins to Return list, select the maximum number of proteins to return in the search.

With this option, you can truncate the results of a search because the data from all of the similar matching proteins does not need to be returned. You can load the results faster.

11. In the Fixed Modifications box, select 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.

12. Click **Save**.

The status of the new search for the experiment in the Pending Search column changes to "Yes."

13. To execute the search from the Data Manager, right-click on the pending search and then click **Run Search**.

New Search in Experiment *X* Dialog Box Parameters for Biomarkers

The New Search in Experiment *X* dialog box for biomarkers, shown in Figure 58 on page 141, contains the following parameters.

Parameter	Description
Search Type	Specifies the type of search to perform:
	• Absolute mass search: See "Searching for Absolute Mass" on page 130.
	• Biomarker search: See "Searching for a Biomarker" on page 138.
	• Sequence tag search: See "Searching for a Sequence Tag" on page 145.
	• Predefined search: See "Performing Predefined Searches" on page 163.
	• Single-protein search: See Chapter 6, "Searching for Single Proteins."
	• Gene-Restricted Absolute Mass Search: See "Searching for a Gene-Restricted Absolute Mass" on page 150.
	• Gene-Restricted Biomarker Mass Search: See "Searching for a Gene-Restricted Biomarker" on page 156.
Database Description	Describes the database that you want to search.
Precursor Mass Type	Specifies the type of precursor ion mass to use:
	• Monoisotopic: Specifies that the precursor mass is monoisotopic, which is the mass of the protein, peptide, or fragment ion, where all carbons are carbon-12.
	• Average: Specifies that the precursor mass is the mass of the most abundant isotope of the protein, peptide, or fragment ion.
Precursor Tolerance	Specifies the tolerance within which your sliding window must fall when you test all protein forms for biomarker peptides and indicates whether it is expressed as absolute (measured in Da) or relative (measured in ppm).
Fragment Mass Type	Specifies the mass type of the fragment ions to use:
	• Monoisotopic: Specifies that the fragment mass is monoisotopic, which is the mass of the protein, peptide, or fragment ion, where all carbons are carbon-12.
	• Average: Specifies that the fragment mass is the mass of the most abundant isotope of the protein, peptide, or fragment ion.

Parameter	Description
Fragment Tolerance	Specifies the tolerance that determines whether the comparison of an observed fragment ion mass to a theoretical fragment ion mass is considered a match and indicates whether it is expressed as absolute, measured in Da, or relative, measured in ppm.
Δm Mode	Determines whether ProSightPC conducts the search in delta (Δ*m*) mode, which is explained in "Performing a Search in Delta m Mode" on page 170.
Disulfide	Indicates whether a protein's cysteines are oxidized.
Include Modified Forms	Indicates whether to include PTMs and polymorphisms when you perform a biomarker search.
Min # of Matching Fragments	Determines whether the search algorithm finds only proteins containing a minimum number of matching ion fragments. The box to the right specifies the minimum number of matching ion fragments.
Min % of Matching Fragments	Determines whether the search algorithm finds only proteins containing a minimum percentage of matching ion fragments. The box to the right specifies the percentage of matching ion fragments.
Min Score	Determines whether the search algorithm finds only proteins with a p score that matches the filter with the expectation value set in the left list, the operator in the middle list, and an appropriate value in the right box. • ≤: Indicates that the first value is less than or equal to the second value. This setting is the default. • ≥: Indicates that the first value is greater than or equal to the second value.
Max Proteins to Return	Specifies the maximum number of proteins to return in the search.
Fixed Modifications	Specifies the chemical modifications present on all instances of a given type of amino acid in the observed protein.
Save	Saves the search information.

Searching for a Sequence Tag

Rich fragmentation data can generate mass differences that allow the inference of short lengths of amino-acid sequence. You can then use these sequence tags to identify a protein.

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

1. During compilation (also known as *de novo* sequencing), ProSightPC analyzes the fragment ion masses and 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 that you defined. The compiler tolerance is always given in parts per million (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.

2. ProSightPC queries the sequence tag list against every base sequence in the proteome database for the presence of any of the sequence tags. It scores any sequence found containing one or more of the sequence tags and reports any sequence scoring above the defined minimum tag score.

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

Thermo Fisher Scientific recommends the following when you conduct 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.

- You can enter the output of a sequence tag search in a series into a gene-restricted search to perform a hybrid search, which frequently identifies and characterizes a protein.

- A sequence tag search is frequently the first step in MS^n experiments.

- Manually enter unresolved amino acid pairs, such as isoleucine and leucine, as a pipe-separated list in square brackets with no spaces, for example, [I|L].

To set preferences for a sequence tag search, see "Setting Sequence Tag Search Preferences" on page 47.

❖ **To search for a sequence tag**

1. In the New Search for Experiment *X* dialog box, select **Sequence Tag** in the Search Type list.

   The appearance of the dialog box changes, as shown in Figure 59.

Figure 59. New Search in Experiment *X* dialog box for sequence tags
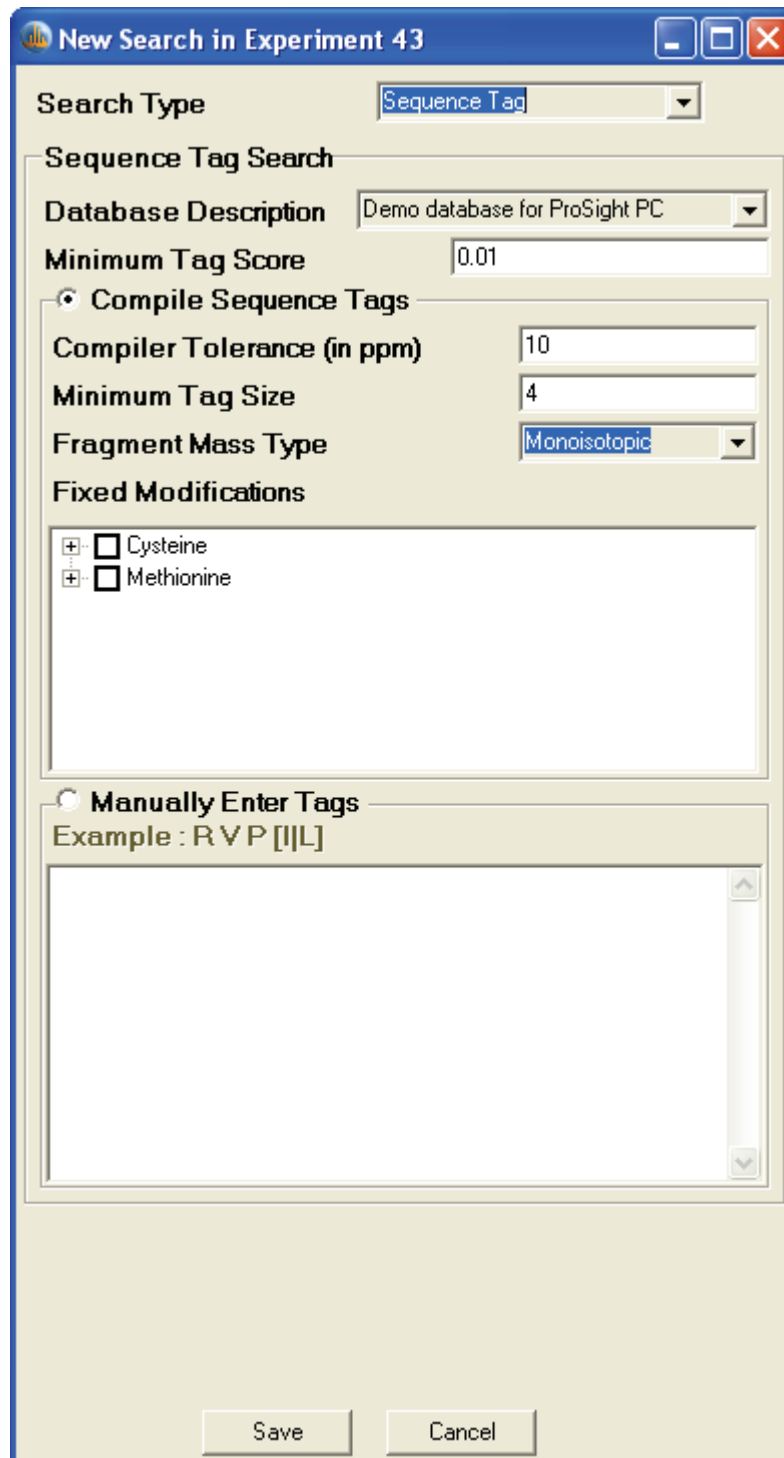
2. In the Database Description list, select a description of the database that you want to search.

3. In the Minimum Tag Score box, enter the lowest acceptable sequence tag score reported as a match.

4. Select either the **Compile Sequence Tags** or **Manually Enter Tags** option.

    •   Compile Sequence Tags: Determines the sequence tags and compiles them before searching them. If you select this option, complete step 5 through step 8 and step 10. This option is the default.

    •   Manually Enter Tags: Enables you to enter sequence tags that you have determined, possibly from manually analyzing a spectrum, and searches them. If you select this option, complete step 9 and step 10.

5. In the Compiler Tolerance box, enter the permissible error, measured in ppm, between two fragment ion masses that are still considered matching an amino acid.

6. In the Minimum Tag Size box, enter the lowest acceptable sequence tag score reported as a match.

7. In the Fragment Mass Type list, specify the type of ion mass fragment type to use:

    •   Monoisotopic: Specifies that the fragment mass is monoisotopic, which is the mass of the protein, peptide, or fragment ion, where all carbons are carbon-12.

    •   Average: Specifies that the fragment mass is the mass of the most abundant isotope of the protein, peptide, or fragment ion.

8. In the Fixed Modifications box, select 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.

9. If you selected the Manually Enter Tags option, enter into the box below it any sequence tags that you want to manually enter instead of compile.

    Each sequence tag consists of the single letter designation of the amino acid separated by a space.

    You can enter isobaric amino acids as a pipe-separated (|) list enclosed in square brackets, for example, [I|L].

    You can enter multiple sequence tags properly formatted, with one sequence tag per line in the space provided.

    The sequence tag search automatically searches for the entered sequence tag and its reverse.

10. Click **Save**.

The status of the new search for the experiment in the Pending Search column now changes to "Yes."

11. To execute the search from the Data Manager, right-click the pending search and then click **Run Search**.

New Search in Experiment *X* Dialog Box Parameters for Sequence Tags

The New Search in Experiment *X* dialog box for sequence tags, shown in Figure 59 on page 147, contains the following parameters.

Parameter	Description
Search Type	Specifies the type of search to perform: • Absolute mass search: See "Searching for Absolute Mass" on page 130. • Biomarker search: See "Searching for a Biomarker" on page 138. • Sequence tag search: See "Searching for a Sequence Tag" on page 145. • Predefined search: See "Performing Predefined Searches" on page 163. • Single-protein search: See Chapter 6, "Searching for Single Proteins." • Gene-Restricted Absolute Mass Search: See "Searching for a Gene-Restricted Absolute Mass" on page 150. • Gene-Restricted Biomarker Mass Search: See "Searching for a Gene-Restricted Biomarker" on page 156.
Database Description	Describes the database that you want to search.
Compile Sequence Tags	Determines the sequence tags and compiles them before searching them. This option is the default.
Compiler Tolerance (in ppm)	Specifies the permissible error, measured in ppm, between two fragment ion masses that are still considered matching an amino acid.
Minimum Tag Size	Specifies the lowest acceptable sequence tag score reported as a match.

Parameter	Description
Fragment Mass Type	Specifies the mass type of the fragment ions to use:
	• Monoisotopic: Specifies that the fragment mass is monoisotopic, which is the mass of the protein, peptide, or fragment ion, where all carbons are carbon-12.
	• Average: Specifies that the fragment mass is the mass of the most abundant isotope of the protein, peptide, or fragment ion.
Fixed Modifications	Specifies the chemical modifications present on all instances of a given type of amino acid in the observed protein.
Manually Enter Tags	Select to enter sequence tags that you have determined, possibly from manually analyzing a spectrum, and to search them.
Save	Saves the search information.

Searching for a Single Protein

For information on performing a single-protein search, see Chapter 6, "Searching for Single Proteins."

Performing Gene-Restricted Searches

Gene-restricted searches look at all protein forms of explicitly listed genes. They can only be made from the results, or match 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 consist of two functionally different, but closely related, types of searches:

• Gene-restricted absolute mass

• Gene-restricted biomarker mass

Searching for a Gene-Restricted Absolute Mass

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 minimum matches parameter are reported.

❖ **To perform a gene-restricted absolute mass search**
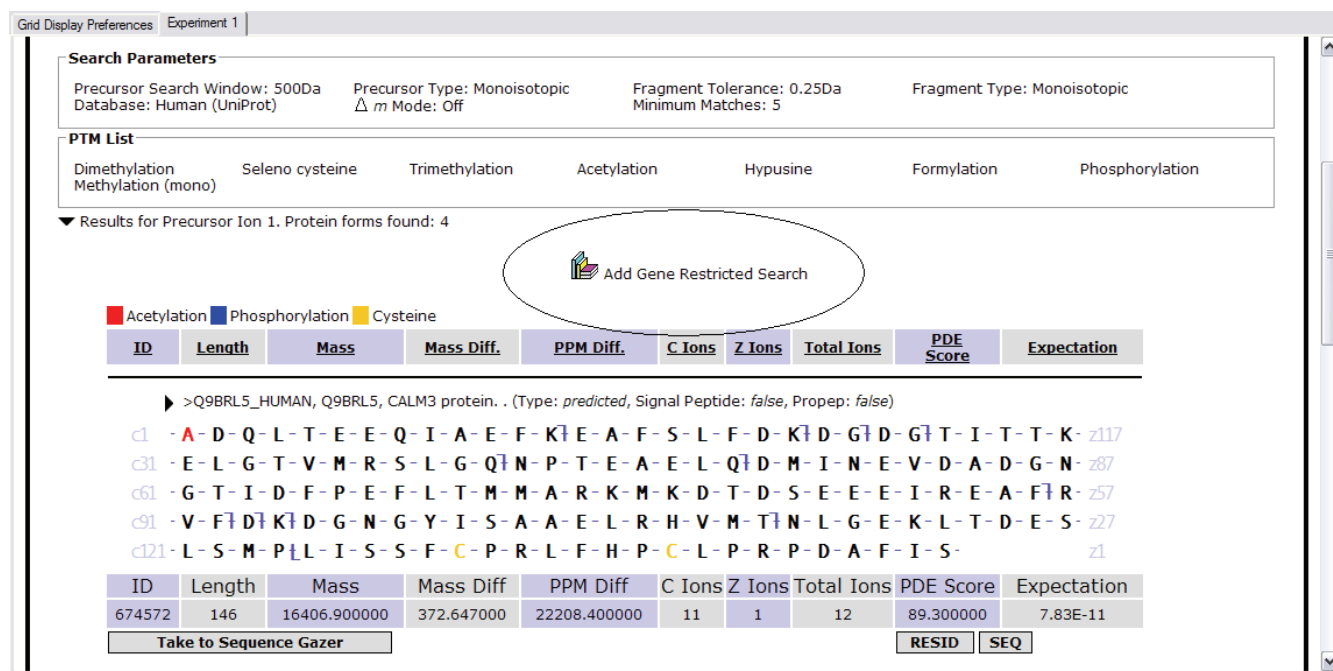
1. Perform any search.

2. Double-click an experiment in the Data Manager to view it.

3. Click the arrow next to **Search x**.

4. Click **Results for Precursor Ion 1** to view its results.

5. From the results list view, click the **Add Gene Restricted Search** icon, 📚, circled in Figure 60.

Figure 60. Performing a gene-restricted search

The New Gene Restricted Search in Experiment *X* dialog box opens, as shown in Figure 61.

Figure 61. New Gene Restricted Search in Experiment *X* dialog box for absolute mass

6. From the Search Type list, select **Gene-Restricted Absolute Mass**.

7. In the Database Description list, select the proteome database to which the entry or entries are being compared.

8. In the Precursor Mass Type list, specify the type of precursor ion mass to search for:

   • Monoisotopic: Specifies that the precursor mass is monoisotopic, which is the mass of the protein, peptide, or fragment ion, where all carbons are carbon-12.

   • Average: Specifies that the precursor mass is the mass of the most abundant isotope of the protein, peptide, or fragment ion.

9. In the Fragment Mass Type list, specify the type of fragment ion mass to search for:

   • Monoisotopic: Specifies that the fragment mass is monoisotopic, which is the mass of the protein, peptide, or fragment ion, where all carbons are carbon-12.

   • Average: Specifies that the fragment mass is the mass of the most abundant isotope of the protein, peptide, or fragment ion.

10. In the Fragment Tolerance box, specify the tolerance that determines whether the comparison of an observed fragment ion mass to a theoretical fragment ion mass is considered a match. From the adjacent list, select the units in which to express the fragment tolerance, either absolute (in daltons) or relative (in parts per million).

    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.

11. Select the Δm **Mode** check box if you want to conduct the search in delta (Δm) mode. For more information on delta (Δm) mode, see "Performing a Search in Delta m Mode" on page 170.

12. Select the **Disulfide** check box if you know that the protein's cysteines are oxidized.

    ProSightPC looks for only one disulfide bond.

13. In the Hit Filtering section, set at least one of the following filters; otherwise, ProSightPC returns all protein forms that are searched, even proteins that have no matching fragments.

    a. Select the **Min # of Matching Fragments** check box if you want the search algorithm to find only proteins containing a minimum number of matching ion fragments (these protein forms are called hits). Then specify the minimum number of matching ion fragments in the box to the right.

    b. Select the **Min % of Matching Fragments** check box if you want the search algorithm to find only proteins containing a minimum percentage of matching ion fragments. Specify the percentage of matching ion fragments in the box to the right.

    c. Select the **Min Score** check box to determine whether the search algorithm finds only proteins with a p score that matches the filter with the expectation value set in the left list, the operator in the middle list, and an appropriate value in the right box. (See "p Score" on page 186 for more information on the p score.)

       • ≤: Indicates that the first value is less than or equal to the second value. This setting is the default.

- ≥: Indicates that the first value is greater than or equal to the second value.

d. In the Max Proteins to Return list, select the maximum number of proteins to return in the search.

With this option, you can truncate the results of a search because the data from all of the similar matching proteins do not need to be returned. You can load the results faster.

14. In the Fixed Modifications box, select no more than one fixed modification per amino acid type.

15. In the PTM Handling box, select the PTMs that you want to search for.

The PTM Handling box displays PTMs arranged in one or more tiers based on the selected proteome database. ProSightPC only queries theoretical protein forms containing exclusively selected PTMs. Every form containing an unselected PTM is excluded from the interrogation.

16. Click **Save**.

The status of the new search for the experiment in the Pending Search column changes to "Yes."

17. To execute the search from the Data Manager, right-click the pending search and then click **Run Search**.

New Gene-Restricted Search in Experiment *X* Dialog Box Parameters for Absolute Mass

The New Gene Restricted Search in Experiment *X* dialog box for absolute mass, shown in Figure 61, contains the following parameters.

Parameter	Description
Search Type	Specifies the type of search to perform:
	• Absolute mass search: See "Searching for Absolute Mass" on page 130.
	• Biomarker search: See "Searching for a Biomarker" on page 138.
	• Sequence tag search: See "Searching for a Sequence Tag" on page 145.
	• Predefined search: See "Performing Predefined Searches" on page 163.
	• Single-protein search: See Chapter 6, "Searching for Single Proteins."
	• Gene-Restricted Absolute Mass Search: See "Searching for a Gene-Restricted Absolute Mass" on page 150.
	• Gene-Restricted Biomarker Mass Search: See "Searching for a Gene-Restricted Biomarker" on page 156.
Database Description	Describes the database that you want to search.
Precursor Mass Type	Specifies the type of precursor ion mass to use:
	• Monoisotopic: Specifies that the precursor mass is monoisotopic, which is the mass of the protein, peptide, or fragment ion, where all carbons are carbon-12.
	• Average: Specifies that the precursor mass is the mass of the most abundant isotope of the protein, peptide, or fragment ion.
Fragment Mass Type	Specifies the mass type of the fragment ions to use:
	• Monoisotopic: Specifies that the fragment mass is monoisotopic, which is the mass of the protein, peptide, or fragment ion, where all carbons are carbon-12.
	• Average: Specifies that the fragment mass is the mass of the most abundant isotope of the protein, peptide, or fragment ion.
Fragment Tolerance	Specifies the tolerance that determines whether the comparison of an observed fragment ion mass to a theoretical fragment ion mass is considered a match and indicates whether it is expressed as absolute, measured in Da, or relative, measured in ppm.

Parameter	Description
Δm Mode	Determines whether ProSightPC conducts the search in delta (Δ*m*) mode, which is explained in "Performing a Search in Delta m Mode" on page 170.
Disulfide	Indicates whether a protein's cysteines are oxidized.
Min # of Matching Fragments	Determines whether the search algorithm finds only proteins containing a minimum number of matching ion fragments. The box to the right specifies the minimum number of matching ion fragments.
Min % of Matching Fragments	Determines whether the search algorithm finds only proteins containing a minimum percentage of matching ion fragments. The box to the right specifies the percentage of matching ion fragments.
Min Score	Determines whether the search algorithm finds only proteins with a p score that matches the filter with the expectation value set in the left list, the operator in the middle list, and an appropriate value in the right box. • ≤: Indicates that the first value is less than or equal to the second value. This setting is the default. • ≥: Indicates that the first value is greater than or equal to the second value.
Max Proteins to Return	Specifies the maximum number of proteins to return in the search.
Fixed Modifications	Specifies the chemical modifications present on all instances of a given type of amino acid in the observed protein.
PTM Handling	Specifies the PTMs that you want queried.
Save	Saves the search information.

Searching for a Gene-Restricted Biomarker

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

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

❖ **To search for a gene-restricted biomarker**

1. Perform any search.

2. Double-click an experiment in the Data Manager to view it.

3. Click the arrow next to **Search x**.

4. Click **Results for Precursor Ion 1** to view its results.

5. From the results list view, click the **Add Gene Restricted Search** icon ![icon], circled in Figure 60.

   The New Gene Restricted Search in Experiment *X* dialog box opens, as shown in Figure 62.

Figure 62. New Gene Restricted Search in Experiment *X* dialog box for biomarkers

6. In the Search Type list, select **Gene-Restricted Biomarker**.

7. In the Database Description list, select the proteome database to which the entry or entries are being compared.

8. In the Precursor Mass Type list, specify the type of precursor ion mass to search for:

   • Monoisotopic: Specifies that the precursor mass is monoisotopic, which is the mass of the protein, peptide, or fragment ion, where all carbons are carbon-12.

   • Average: Specifies that the precursor mass is the mass of the most abundant isotope of the protein, peptide, or fragment ion.

9. In the Fragment Mass Type list, specify the type of fragment ion mass to search for:

   • Monoisotopic: Specifies that the fragment mass is monoisotopic, which is the mass of the protein, peptide, or fragment ion, where all carbons are carbon-12.

   • Average: Specifies that the fragment mass is the mass of the most abundant isotope of the protein, peptide, or fragment ion.

10. In the Fragment Tolerance box, specify the tolerance that determines whether the comparison of an observed fragment ion mass to a theoretical fragment ion mass is considered a match. From the adjacent list, select the units in which to express the fragment tolerance, either absolute (in daltons) or relative (in parts per million).

    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.

11. Select the Δm **Mode** check box if you want to conduct the search in delta (Δm) mode. For more information on delta (Δm) mode, see "Performing a Search in Delta m Mode" on page 170.

12. Select the **Disulfide** check box if you know that the protein's cysteines are oxidized.

    ProSightPC looks for only one disulfide bond.

13. Select the **Include Modified Forms** check box if you want to include PTMs and polymorphisms when you perform a biomarker search.

    To detect biomarkers with modifications on them, select this option; however, processor time increases as a result.

14. In the Hit Filtering section, set at least one of the following filters; otherwise, ProSightPC returns all protein forms that are searched, even proteins that have no matching fragments.

    a. Select the **Min # of Matching Fragments** check box if you want the search algorithm to find only proteins containing a minimum number of matching ion fragments (these protein forms are called hits). Then specify the minimum number of matching ion fragments in the box to the right.

     b. Select the **Min % of Matching Fragments** check box if you want the search algorithm to find only proteins containing a minimum percentage of matching ion fragments. Specify the percentage of matching ion fragments in the box to the right.

     c. Select the **Min Score** check box to determine whether the search algorithm finds only proteins with a p score that matches the filter with the expectation value set in the left list box, the operator in the middle list, and an appropriate value in the right box. (See "p Score" on page 186 for more information on the p score.)

         • ≤: Indicates that the first value is less than or equal to the second value. This setting is the default.

         • ≥: Indicates that the first value is greater than or equal to the second value.

     d. From the Max Proteins to Return list, select the maximum number of proteins to return in the search.

        With this option, you can truncate the results of a search because the data from all of the similar matching proteins does not need to be returned. You can load the results faster.

15. In the Fixed Modifications box, select no more than one fixed modification per amino acid type.

16. Click **Save**.

     The status of the new search for the experiment in the Pending Search column changes to "Yes."

17. To execute the search from the Data Manager, right-click the pending search and then click **Run Search**.

New Gene-Restricted Search in Experiment *X* Dialog Box Parameters for Biomarkers

The New Gene-Restricted Search in Experiment *X* dialog box for biomarkers, shown in Figure 62, contains the following parameters.

Parameter	Description
Search Type	Specifies the type of search to perform:
	• Absolute mass search: See "Searching for Absolute Mass" on page 130.
	• Biomarker search: See "Searching for a Biomarker" on page 138.
	• Sequence tag search: See "Searching for a Sequence Tag" on page 145.
	• Predefined search: See "Performing Predefined Searches" on page 163.
	• Single-protein search: See Chapter 6, "Searching for Single Proteins."
	• Gene-Restricted Absolute Mass Search: See "Searching for a Gene-Restricted Absolute Mass" on page 150.
	• Gene-Restricted Biomarker Mass Search: See "Searching for a Gene-Restricted Biomarker" on page 156.
Database Description	Describes the database that you want to search.
Precursor Mass Type	Specifies the type of precursor ion mass to use:
	• Monoisotopic: Specifies that the precursor mass is monoisotopic, which is the mass of the protein, peptide, or fragment ion, where all carbons are carbon-12.
	• Average: Specifies that the precursor mass is the mass of the most abundant isotope of the protein, peptide, or fragment ion.
Precursor Tolerance	Specifies the tolerance within which your sliding window must fall when you test all protein forms for biomarker peptides and indicates whether it is expressed as absolute (measured in Da) or relative (measured in ppm).
Fragment Mass Type	Specifies the mass type of the fragment ions to use:
	• Monoisotopic: Specifies that the fragment mass is monoisotopic, which is the mass of the protein, peptide, or fragment ion, where all carbons are carbon-12.
	• Average: Specifies that the fragment mass is the mass of the most abundant isotope of the protein, peptide, or fragment ion.

Parameter	Description
Fragment Tolerance	Specifies the tolerance that determines whether the comparison of an observed fragment ion mass to a theoretical fragment ion mass is considered a match and indicates whether it is expressed as absolute, measured in Da, or relative, measured in ppm.
Δm Mode	Determines whether ProSightPC conducts the search in delta (Δm) mode, which is explained in "Performing a Search in Delta m Mode" on page 170.
Disulfide	Indicates whether a protein's cysteines are oxidized.
Include Modified Forms	Indicates whether to include PTMs and polymorphisms when you perform a biomarker search.
Min # of Matching Fragments	Determines whether the search algorithm finds only proteins containing a minimum number of matching ion fragments. The box to the right specifies the minimum number of matching ion fragments.
Min % of Matching Fragments	Determines whether the search algorithm finds only proteins containing a minimum percentage of matching ion fragments. The box to the right specifies the percentage of matching ion fragments.
Min Score	Determines whether the search algorithm finds only proteins with a p score that matches the filter with the expectation value set in the left list, the operator in the middle list, and an appropriate value in the right box. • $\leq$: Indicates that the first value is less than or equal to the second value. This setting is the default. • $\geq$: Indicates that the first value is greater than or equal to the second value.
Max Proteins to Return	Specifies the maximum number of proteins to return in the search.
Fixed Modifications	Specifies the chemical modifications present on all instances of a given type of amino acid in the observed protein.
Save	Saves the search information.

Performing MS[n] Hybrid Searches

In some cases, you might need to use a sequence tag search to reduce the search space before performing an absolute mass search. MS[n] hybrid searches perform this function. A hybrid search first compiles a list of all possible sequence tags consistent with the observed fragment ions, and then ProSightPC uses these tags to identify all protein forms in the database that are consistent with the tags. The list of protein forms that match the sequence tags functions as input into an absolute mass search.

Performing Predefined Searches

Predefined searches are a strategy to simplify the repetition of identical searches on different sets of MS/MS data. They enable you to assign a name to a set of parameters that you can then add to any experiment.

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 add it to your experiments as needed.

The search parameters in predefined searches are persistent until you modify or delete them.

Creating a Predefined Search

❖ **To create a predefined search**

1. Choose **Tools > Manage Predefined Searches**, or click the **Manage Predefined Searches** icon, 🔍, to open the Predefined Search Manager dialog box, shown in Figure 63.

   Figure 63. Predefined Search Manager dialog box
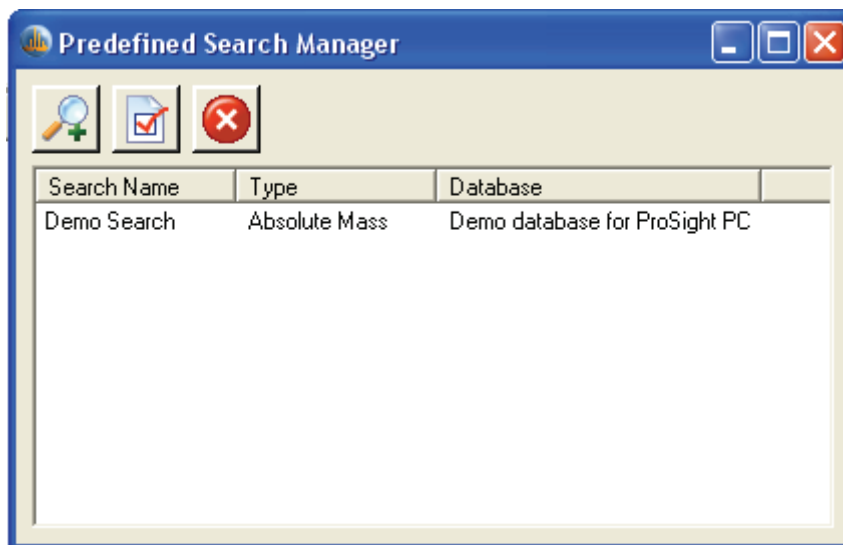
   All existing predefined searches appear in the Predefined Search Manager dialog box. In the example in Figure 63, the Demo Search default predefined search is available.

2. Click the **Create** icon, 🔍, on the Predefined Search Manager dialog box, or right-click the view area and choose **New** from the shortcut menu.

   The New Predefined Search dialog box opens, as shown in Figure 64.

Figure 64. New Predefined Search dialog box

3. In the Search Name box, type a name for the new predefined search.

4. From the Search Type list, select the search type, and follow the procedure for your selection:

   • Absolute mass search: See "Searching for Absolute Mass" on page 130.

   • Biomarker search: See "Searching for a Biomarker" on page 138.

   • Sequence tag search: See "Searching for a Sequence Tag" on page 145.

   • Single-protein search: See "Searching for Single Proteins" on page 177.

Predefined Search Manager Dialog Box Parameters

The Predefined Search Manager dialog box, shown in Figure 63, includes the following icons and parameters:

Parameter	Description
🔍	Opens the New Predefined Search dialog box so that you can create a new predefined search. See "Creating a Predefined Search" on page 163 for information on this dialog box.
☑	Opens the Edit Predefined Search dialog box so that you can edit the parameters for the search. See "Editing a Predefined Search" on page 166 for information on this dialog box.
❌	Removes the selected predefined search from the list of predefined searches to add to an experiment.
Search Name	Displays the name of the predefined search.
Type	Displays the type of predefined search. You can select absolute mass, biomarker, sequence tag, and single-protein searches.
Database	Displays the database on which the search is run.

New Predefined Search Dialog Box Parameters

The parameters in the New Predefined Search dialog box depend on the type of search that you select in the Search Type list:

• Absolute mass search: See "New Search in Experiment **X** Dialog Box Parameters for Absolute Mass" on page 136.

• Biomarker search: See "New Search in Experiment **X** Dialog Box Parameters for Biomarkers" on page 143.

• Sequence tag search: See "New Search in Experiment **X** Dialog Box Parameters for Sequence Tags" on page 149.

• Single-protein search: See Chapter 6, "Searching for Single Proteins."

Editing a Predefined Search

❖ **To edit a predefined search**

1. Select a predefined search name from the list in the Predefined Search Manager dialog box, shown in Figure 63 on page 163.

2. Click the **Edit** icon, [icon], on the Predefined Search Manager dialog box, or right-click the search name and choose **Edit** from the shortcut menu.

   The Edit Predefined Search dialog box opens, as shown in Figure 65.

Figure 65. Edit Predefined Search dialog box

3. Edit the parameters. See the following sections for more information.

- Absolute mass search: See "Searching for Absolute Mass" on page 130.

- Biomarker search: See "Searching for a Biomarker" on page 138.

- Sequence tag search: See "Searching for a Sequence Tag" on page 145.

- Predefined search: See "Performing Predefined Searches" on page 163

- Single-protein search: See Chapter 6, "Searching for Single Proteins."

- Gene-Restricted Absolute Mass Search: See "Searching for a Gene-Restricted Absolute Mass" on page 150.

- Gene-Restricted Biomarker Mass Search: See "Searching for a Gene-Restricted Biomarker" on page 156.

4. Click **Save**.

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

Edit Predefined Search Dialog Box Parameters

The parameters in the Edit Predefined Search dialog box depend on the type of search that you select in the Search Type list:

- Absolute mass search: See "New Search in Experiment **X** Dialog Box Parameters for Absolute Mass" on page 136.

- Biomarker search: See "New Search in Experiment **X** Dialog Box Parameters for Biomarkers" on page 143.

- Sequence tag search: See "New Search in Experiment **X** Dialog Box Parameters for Sequence Tags" on page 149.

- Single-protein search: See Chapter 6, "Searching for Single Proteins."

Removing a Predefined Search

❖ **To remove a predefined search**

1. Select a predefined search name from the list in the Predefined Search Manager dialog box (see Figure 63 on page 163).

2. Click the **Remove Selected Search** icon, ⊗, on the Predefined Search Manager dialog box, or right-click the search name and choose **Remove** from the shortcut menu.

3. Confirm the removal by clicking **Yes** or **No**.

You can return to the Predefined Search Manager without removing the selected search by clicking No.

Adding a Predefined Search to an Experiment

❖ **To add a predefined search to an experiment**

1. Right-click the experiment in the Data Grid and choose **Append Predefined Search**. You can also choose **Experiment Tools > Append Predefined Search** or click the **Add Predefined Search** icon, 📇.

   The Append Predefined Searches to Experiment *X* dialog box opens, as shown in Figure 66.

   Figure 66. Append Predefined Searches to Experiment *X* dialog box
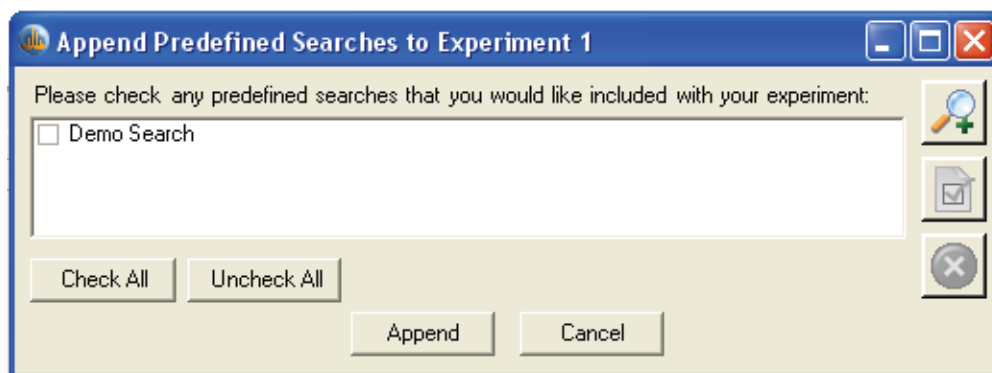
   The default predefined search is Demo Search.

2. 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 To process several predefined searches automatically, choose **Tools > Batch Run**.

Append Predefined Searches to Experiment *X* Dialog Box Parameters

The Append Predefined Searches to Experiment *X* dialog box, shown in Figure 66, contains the following parameters:

Parameter	Description
Please check any predefined searches that you would like included with your experiment	Lists the predefined searches available to add to an experiment.
Check All	Selects all the predefined searches listed to add to an experiment.
Uncheck All	Clears all the predefined searches.
🔍	Opens the New Predefined Search dialog box so that you can create a new predefined search. See "Creating a Predefined Search" on page 163 for information on this dialog box.

Parameter	Description
	Opens the Edit Predefined Search dialog box so that you can edit the parameters for the search. See "Editing a Predefined Search" on page 166 for information on this dialog box.
	Removes the selected predefined search from the list of predefined searches to add to an experiment.

Analyzing an MS/MS Experiment

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

1. Construct an MS/MS experiment with the MS/MS 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. Create an MS/MS experiment using the MS data as precursor ions and the MS/MS data as fragment ions.

3. Add a sequence tag search to the experiment from step 2 and manually enter the sequence tags from the first experiment. When you run this search, ProSightPC creates a gene list for all proteins containing sequences consistent with the MS^3 fragmentation data.

4. Add a gene-restricted absolute mass search to the second experiment. ProSightPC uses this search to identify and characterize the observed protein.

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

Performing a Search in Delta m Mode

Delta (Δm) mode is a technique for identifying protein forms containing unknown PTMs. The delta is the difference between the observed precursor mass and the theoretical precursor mass. When you perform a search in delta (Δm) mode, ProSightPC concurrently performs three queries per sequence to compare the following:

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

• The theoretical fragment ion masses derived from the sequence and the delta applied to the N terminal to the observed fragment ion mass list.

• The theoretical fragment ion masses derived from the sequence and the delta applied to the C terminal to the observed fragment mass list.

---

[1] Zabrouskov, V.; Senko, M. W.; Du, Y.; LeDuc, R. D.; Kelleher, N. L. New and Automated MS^n Approaches for Top-Down Identification of Modified Proteins. *J. Am. Soc. Mass Spectrom.* **2005**, *16* (12), 2027–2038.

A delta search takes approximately three times longer than the same search without delta mode.

By carefully observing the pattern of fragments with and without delta (Δm) mode, you can frequently locate the delta.

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

As shown in Figure 67, ProSightPC first checks the observed precursor mass against the theoretical precursor masses of every protein in the specified precursor mass window and calculates and stores the mass difference (delta). Next, it doubles the theoretical fragment ion list for each protein for each b/y or c/z ion. ProSightPC checks both the original fragment mass and the modified fragment mass (plus the delta) against the observed fragment ion mass list. As a result, ProSightPC returns any observed fragment ions having the same mass shift as the precursor protein as positive matches.

Figure 67. Schematic of Δm mode
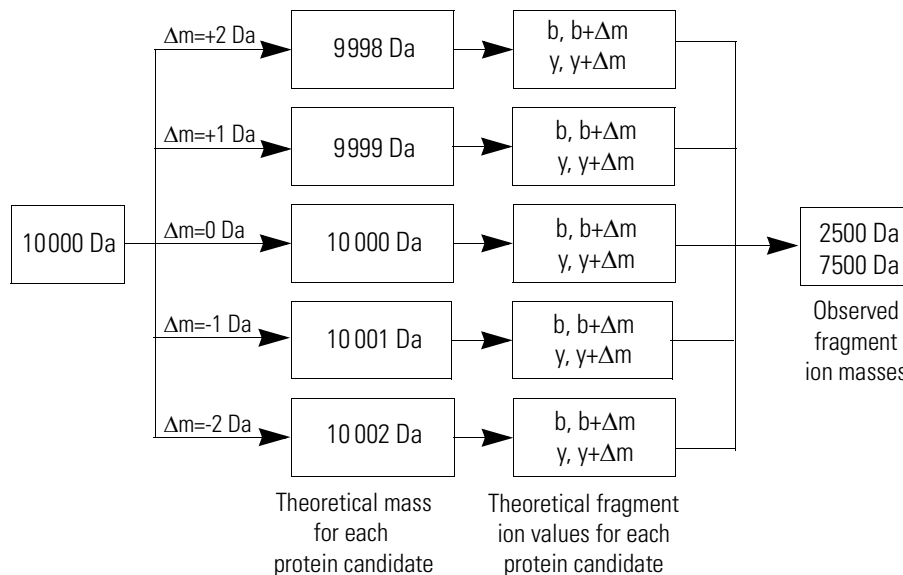
Performing Searches in Batch Mode

With batch processing, you can queue and run a large number of searches over any number of experiments in a single action. Use it when you have many pending searches in a .puf file and you would like to run all of them.

❖ **To perform searches in batch mode**

1. Select all the desired experiments, right-click, and choose **Add Search** from the shortcut menu.

2. Click **Save** in all the New Search in Experiment *X* dialog boxes.

3. Choose **Tools > Batch Run**.

Each pending search is queued and 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.

Using Search Reports

ProSightPC provides several batch-processing and reporting tools for managing large numbers of MS/MS experiments. They simplify working with several experiments in a single .puf file. This section describes how to use these tools to manage multiple experiments.

The following types of reports help you summarize your work:

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

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

- A best hit report displays the search result with the best score for each search that was run for each experiment in the Data Grid.

- A repository report lists all the experiments that a repository contains. See "Generating a Repository Report" on page 79 for information on this report.

❖ **To generate a status report**

1. Open the desired .puf file.

2. Choose **Tools > Reports > Status Report**.

A summary of all experiments and searches contained in the .puf file appears in a new window as a text document, as shown in Figure 68. This text document is organized by experiment number and is subdivided into the types of searches.

Figure 68. Status report
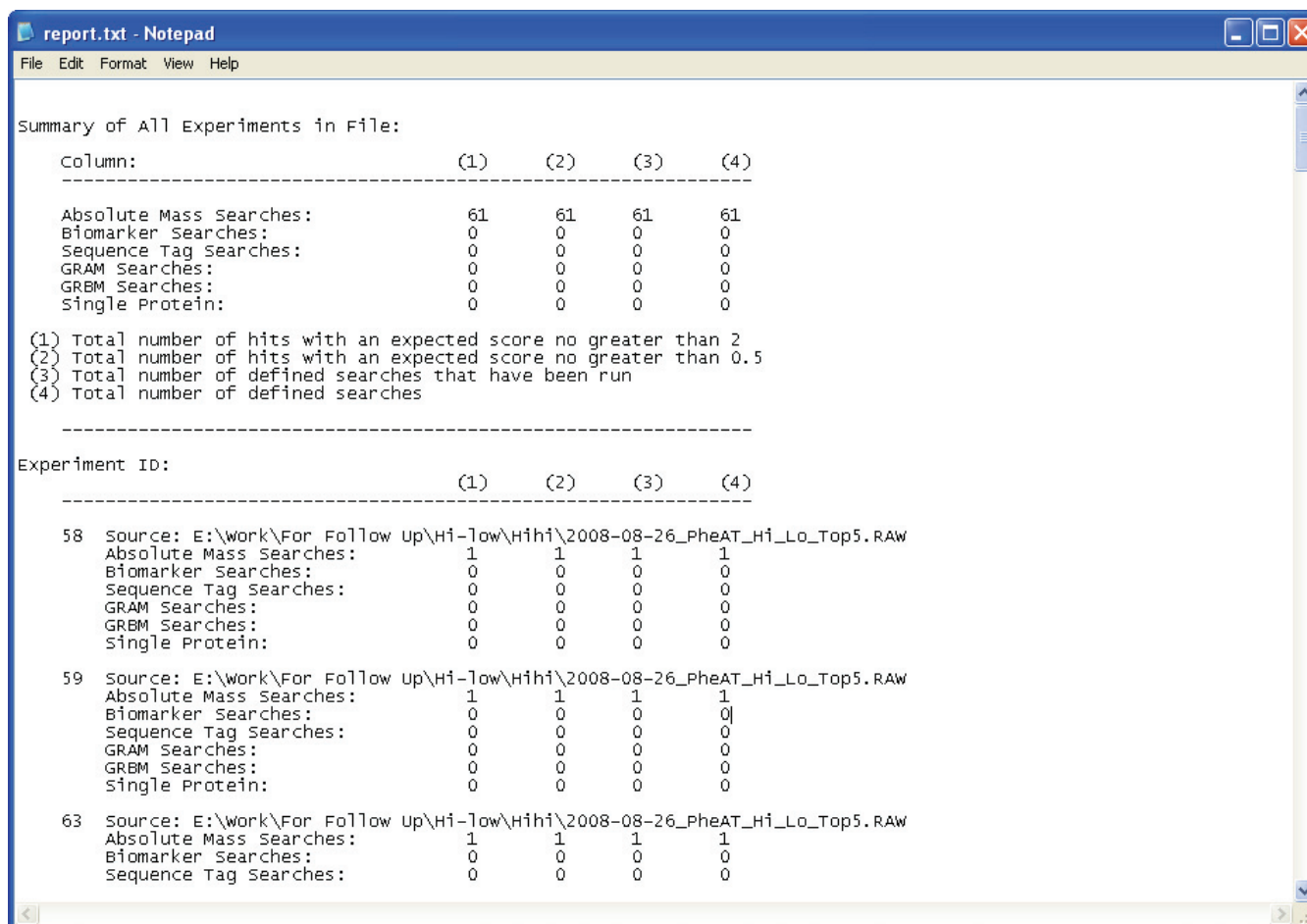
❖ **To generate a printable search report**

1. Select a search in the Data Grid.

2. Choose **Tools > Reports > Printable Report**.

The report appears in a Web browser window. Figure 69 shows an excerpt of this report.

Figure 69. Printable search report

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

❖ **To generate a best-hit report**

1. Choose **Tools > Reports > Best Hit Report**.

The report, shown in Figure 70, appears in a Web browser window.

Figure 70. Best-hit report

Data for Experiment 43

Custom created Experiment. To edit this comment, click Experiment Tools, Edit Comment Puf Filter: This file passed the following filters: Max Frags: -1; Min Frags: 10; Min Intact Mass: 750;
Search 1: Biomarker Search
 Sample BioMarker Search
 Results for Intact Ion 1. Protein forms found: 1

1-->39: RS28_HUMAN, P62857, 40S ribosomal protein S28.. (Type: *basic*, Signal Peptide: *false*, Propep: *false*)

ID	Length	Mass	Mass Diff	PPM Diff	C Ions	Z Ions	Total Ions
10	39	4337.2	-.0005	-.1071	24	16	40

Search 3: GRBM Search
 Sample Gene Restricted BioMarker (GRBM) Search. GRBM and GRAM Searches are created using results from other searches.
 Results for Intact Ion 1. Protein forms found: 2

1-->39: RS28_HUMAN, P62857, 40S ribosomal protein S28.. (Type: *basic*, Signal Peptide: *false*, Propep: *false*)

ID	Length	Mass	Mass Diff	PPM Diff	C Ions	Z Ions	Total Ions
10	39	4337.2	-.0005	-.1071	24	16	40

1-->39: RS28_HUMAN, P62857, 40S ribosomal protein S28.. (Type: *conflict*, Signal Peptide: *false*, Propep: *false*)

ID	Length	Mass	Mass Diff	PPM Diff	C Ions	Z Ions	Total Ions
5	39	4337.2	-.0005	-.1071	24	16	40

Searching for Single Proteins

This chapter describes how to use the Sequence Gazer to perform single-protein searches. You can fit any MS/MS experiment data to a single protein and hypothesize various permissible PTMs. As you test these different hypotheses, you can save the results. ProSightPC stores each saved result as a single-protein mode search result.

Contents

- Sequence Gazer
- Accessing the Sequence Gazer
- Navigating the Sequence Gazer
- Demonstrating the Sequence Gazer

Sequence Gazer

The Sequence Gazer is an interactive environment for comparing MS/MS data to a known protein sequence. The Sequence Gazer characterizes previously identified proteins by selectively adding or removing PTMs or custom masses to amino acids in a protein sequence. Once you have made all your modifications to the amino acids, you can reevaluate the ion data.

When you take a protein sequence to the Sequence Gazer, ProSightPC automatically scores the sequence on the basis of the initial search parameters. "Scores Box" on page 185 explains ProSightPC's scoring system.

You can change parameters and add or remove PTMs or fixed modifications. ProSightPC then rescores the modified sequence. Ideally, changes to the sequence followed by rescoring yield more matching fragments than before, narrowing the possible matching protein forms that explain the MS/MS data.

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

The Sequence Gazer is usually used for one of two purposes:

- MS/MS data might have been gathered on a known, pure protein containing one or more unknown PTMs. In this case, you build a single-protein mode search and add it to the MS/MS experiment. You use the Sequence Gazer to test hypotheses regarding which PTMs are present.

- The result of any other search mode might identify and partially characterize a protein, whereas the Sequence Gazer can fully characterize the protein.

Accessing the Sequence Gazer

You can access the 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

See "Demonstrating the Sequence Gazer" on page 195 for a demonstration showing how to use the Sequence Gazer to find modifications in fragment ions.

❖ **To perform a single-protein search**

1. Open the desired experiment in a Data Manager by double-clicking the experiment in the Data Grid.

2. Choose **Experiment Tools > Add Search**.

   The New Search in Experiment *X* dialog box opens.

3. In the Search Type list, select **Single Protein**.

   The New Search in Experiment *X* dialog box changes its appearance, as shown in Figure 71.

Figure 71. New Search in Experiment *X* dialog box for a single protein

4. In the Precursor Mass Type list, select one of the following:

   • Monoisotopic: Specifies that the mass is monoisotopic, which is the mass of the protein, peptide, or fragment ion, where all carbons are carbon-12.

   • Average Mass: Specifies that the mass is the mass of the most abundant isotope of the protein, peptide, or fragment ion.

5. In the Fragment Mass Type list, select one of the following:

   • Monoisotopic: Specifies that the mass is monoisotopic, which is the mass of the protein, peptide, or fragment ion, where all carbons are carbon-12.

   • Average Mass: Specifies that the mass is the mass of the most abundant isotope of the protein, peptide, or fragment ion.

6. In the Fragment Tolerance box, specify the tolerance that determines whether the comparison of an observed fragment ion mass to a theoretical fragment ion mass is considered a match, and indicate whether it is expressed as absolute (measured in Da) or relative (measured in ppm).

7. (Optional) Select the Δ***m* Mode** box to perform the search in delta (Δ*m*) mode. For more information on delta (Δ*m*) mode, see "Performing a Search in Delta m Mode" on page 170).

8. In the Fixed Modifications box, select 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.

9. In the Sequence box, either type the sequence or copy and paste a sequence from another source.

10. Click **Save.**

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

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

New Search in Experiment *X* Dialog Box Parameters for a Single Protein

The New Search in Experiment *X* dialog box for a single protein, shown in Figure 71 on page 179, contains the following parameters.

Parameter	Description
Search Type	Specifies the type of search to perform:
	• Absolute mass search: See "Searching for Absolute Mass" on page 130.
	• Biomarker search: See "Searching for a Biomarker" on page 138.
	• Sequence tag search: See "Searching for a Sequence Tag" on page 145.
	• Predefined search: See "Performing Predefined Searches" on page 163.
	• Single-protein search: See Chapter 6, "Searching for Single Proteins."
	• Gene-Restricted Absolute Mass Search: See "Searching for a Gene-Restricted Absolute Mass" on page 150.
	• Gene-Restricted Biomarker Mass Search: See "Searching for a Gene-Restricted Biomarker" on page 156.
Precursor Mass Type	Specifies the type of precursor ion mass to use:
	• Monoisotopic: Specifies that the precursor mass is monoisotopic, which is the mass of the protein, peptide, or fragment ion, where all carbons are carbon-12.
	• Average: Specifies that the precursor mass is the mass of the most abundant isotope of the protein, peptide, or fragment ion.
Fragment Mass Type	Specifies the mass type of the fragment ions to use:
	• Monoisotopic: Specifies that the fragment mass is monoisotopic, which is the mass of the protein, peptide, or fragment ion, where all carbons are carbon-12.
	• Average: Specifies that the fragment mass is the mass of the most abundant isotope of the protein, peptide, or fragment ion.

Parameter	Description
Fragment Tolerance	Specifies the tolerance that determines whether the comparison of an observed fragment ion mass to a theoretical fragment ion mass is considered a match and indicates whether it is expressed as absolute, measured in Da, or relative, measured in ppm.
Δm Mode	Determines whether ProSightPC conducts the search in delta (Δm) mode, which is explained in "Performing a Search in Delta m Mode" on page 170.
Fixed Modifications	Specifies the chemical modifications present on all instances of a given type of amino acid in the observed protein.
Sequence	Specifies the sequence. You can either type the sequence or use a sequence from another source.
Save	Saves the search information.

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

1. Select the desired search and click its corresponding side arrow in the Data Manager.

2. Locate the desired protein identification in the search results and click its corresponding side arrow.

3. Click **Take to Sequence Gazer** (circled in Figure 72).

Figure 72. Take To Sequence Gazer button

Navigating the Sequence Gazer

Fragment ion information in the Sequence Gazer interface is organized into the following regions, as shown in Figure 73:

- Search Parameter Display

- Scores Box

- Fragments Explained Box

- Mass Diagram

- Interactive Fragment Map

- Amino Acid Information Box

- Fixed Modifications Box

- Matching Fragments Table

- Non-Matching Fragments Table

Figure 73. Sequence Gazer window

Search Parameter Display

The search parameter display shows the data options and tolerances that you selected during the last round of scoring. User-defined selections appear in red. You can change these by clicking on a new selection. The new selection appears in red.

None of the changes made in the search parameter display are implemented until you click Rescore.

The search parameter display contains the following parameters.

Parameter	Description
Precursor Mass Type	Specifies the type of precursor ion mass to use: • Monoisotopic: Specifies that the precursor mass is monoisotopic, which is the mass of the protein, peptide, or fragment ion, where all carbons are carbon-12. • Average: Specifies that the precursor mass is the mass of the most abundant isotope of the protein, peptide, or fragment ion.
Fragment Tolerance	Displays the fragment tolerance and unit at the time that the search was last scored. The fragment tolerance is the tolerance that determines whether the comparison of an observed fragment ion mass to a theoretical fragment ion mass is considered a match. 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.
Δm Mode	Indicates whether delta (Δm) mode has been selected.
Mass Type	Specifies the type of ion mass fragment type to use: • Monoisotopic: Specifies that the mass is monoisotopic, which is the mass of the protein, peptide, or fragment ion, where all carbons are carbon-12. • Average: Specifies that the mass is the mass of the most abundant isotope of the protein, peptide, or fragment ion.

Scores Box

ProSightPC uses a number of different scoring systems to give you a greater degree of freedom when interpreting your results. The Scores box in the Sequence Gazer displays the following three scores:

• P score, as noted by Meng, et al.[1]

• Expectation value (e value), as noted by LeDuc, et al.[2]

• PDE (McLuckey), as noted by Reid, et al.[3]

---

[1] Meng, F.B.J. Cargile; Miller, L. H.; Forbes, A. J.; Johnson, J. R.; Kelleher, N. L. Informatics and multiplexing of intact protein identification in bacteria and the archaea. *Nat. Biotechno.* **2001**, *19*: 952–957.

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

[3] Reid, G. E.; Shang, H.; Hogan, J. M.; Lee, G. U.; McLuckey, S. A. Gas phrase concentration, purification, and identification of whole proteins from complex mixtures. *J. Am. Chem. Soc.* **2002**, *124*: 7353–7362.

These scores are described in the following sections.

p Score

A p score is the probability of obtaining at least as good a match between the observed fragment list and a sequence as by chance. It is a measure of confidence in the validity of a match. A low p score means that the probability of obtaining at least this many fragments matching a sequence is low, so it is unlikely that random chance is the cause of the association.

ProSightPC calculates a p score as follows:

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

where:

- n is the number of matching fragments.

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

- f is the total number of fragment ions observed.

Since the Poisson distribution allows n to go to infinity, $1 - n$ is calculated to determine the probability of getting at least this good of a result.

Expectation Value (e value)

The expectation value (e value) is the number of sequences in a database that are expected to have p scores equal to or better than what was observed simply by chance. Low e values represent better matches (less likely to be false positives) than high e values. Since the p score represents the probability of the n out of f fragments matching by chance, and if it is assumed that all sequences in the database are independent, the e value of a sequence-fragment set association is simply the association's p value times the number of sequences in the database.

If N is the number of protein forms considered during a search, the e value currently reported by ProSightPC is

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

Sample Calculation

Consider ubiquitin caroxyl-terminal hydrolase 12 EC 3.1.2.15. This protein has 355 amino acids and a theoretical intact mass of 41201 daltons. Consider a hypothetical MS/MS experiment that results in 32 fragment ions, of which n number matches this protein with a mass accuracy of plus or minus 2.5 daltons.

To calculate the p score of this assignment, apply the equation shown in "p Score" on page 186 to find the sum of the Poisson distribution for i=0 to n-1 (with lambda = $f*x$ or 32(2.5*4/111.1)), which is subtracted from 1. Figure 74 shows p(n) for all values of n between 0 and 32. As the number of matching fragments increases, it becomes highly unlikely that the fragment ion matching is due to chance.

Figure 74. Poisson value versus Poisson value or greater
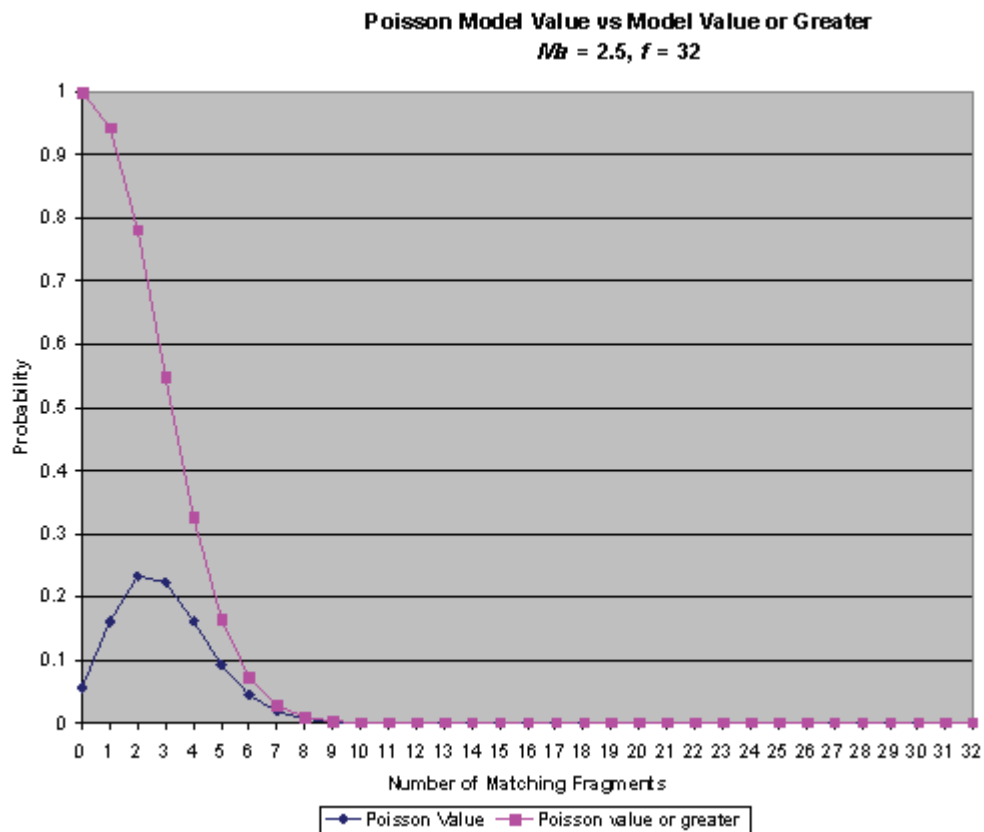
To find the p score for 12 matching fragments, sum the first 11 values as follows:

$$
\begin{aligned}
&0.056118597 \\
&+\ 0.161637722 \\
&+\ 0.232781598 \\
&+\ 0.223492684 \\
&+\ 0.160930825 \\
&+\ 0.092705426 \\
&+\ 0.044503055 \\
&+\ 0.018311659 \\
&+\ 0.006592857 \\
&+\ 0.002109925 \\
&+\ 0.000607719 \\
&+0.000159128 \\
&=\ 0.999951
\end{aligned}
$$

Then subtract 0.99951 from 1.000000:

1.000000
- 0.99951
= 4.9E-5

Therefore, the probability of 12 or more fragments matching by chance, out of a fragment ion list with 32 masses and a tolerance of plus or minus 2.5 daltons, is 4.9E-5.

PDE (McLuckey) Score

ProSightPC also reports a score calculated according to McLuckey. The McLuckey, or PDE, score is a way of scoring how well a set of observed fragment ions matches a protein. It takes into consideration which amino acids would have to have been cleaved in order to match the observed fragment ion data with the theoretical ion masses from the database. To find this score, use the following equation:

$$McLuckey\_score = 5nP\sum I_P + 5nD\sum I_D + 4nK\sum I_X + 2nE\sum I_E + nX\sum I_X$$

This equation includes the parameters:

nP   Specifies the number of product ions predicted with cleavage at an N terminal to a proline.

nD   Specifies the number of product ions predicted with cleavage at a C terminal to an aspartic acid.

nK   Specifies the number of product ions predicted with cleavage at a C terminal to a lysine.

nE   Specifies the number of product ions predicted with cleavage at a C terminal to a glutamic acid.

nX   Specifies the number of product ions predicted with cleavage at any other non-specific residue.

In   Specifies the sum of the intensities of the corresponding n values just given.

Sequence Tag Scores

ProSightPC uses a scoring system to rank the matches between a set of sequence tags and a sequence. The score for a single tag in a query that matches a sequence is calculated as follows:

$$score = \ln(\prod^n p_i)$$

where:

- ln is the length of the sequence.

- p_i is the probability of the ith amino acid occurring in a protein.

Since multiple tags can match the sequence, each tag is weighted by the number of independent possibilities for the tag to match the sequence. This is approximated in ProSightPC as follows:

$$score = \ln(\prod{}^n p_i)\left(\frac{n_i}{l}\right)$$

where:

- *l* is the overall length of the sequence.

- n_i is the length of the sequences in the tag.

The final score for a query is then the sum of all tag scores that matched.

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 includes three additional controls:

Parameter	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 that you have made and returns to the Data Manager.

Mass Diagram

The mass diagram displays the difference between the observed and theoretical mass, expressed in daltons and parts per million. It contains the following boxes or lists:

Parameter	Description
Observed	Contains a list that displays all the precursor masses detected by ProSightPC.
Theoretical	Displays the experimental precursor mass, including all user-input changes, as of the last score.
Difference	Displays the difference between the figure in the Observed list and the figure in the Theoretical box.

Interactive Fragment Map

The interactive fragment map, shown in Figure 75, is an interactive display of the protein sequence, along with any PTMs, and the matching fragment information.

Figure 75. Interactive fragment map

The theoretical protein sequence taken from the proteome warehouse is listed from left to right and from top to bottom. Depending on the ion type used in the experiment, blue bars with a serif at the top going to the left represent the termination of either b or c ions. A serif at the right bottom is the start of either a y or z ion.

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 that the matching PTM is currently assigned to that amino acid.

Note Click an amino acid to access 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. It displays the following information:

Parameter	Description
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.

Parameter	Description
Start PTM	Displays the PTM attached to the amino acid as of the last score.
PTM Choices	Adds "virtual PTMs" to each amino acid, which changes the sequences of the protein and therefore the score. • None: Removes the applied PTM from the amino acid. • Custom: Adds a custom mass shift (a PTM is a fixed mass shift). • PTM tiers: Reflects the priority of the PTMs. The PTMs in tier 1 are more common.

❖ **To add virtual PTMs to an amino acid**

1. Select an amino acid in the Sequence Gazer.

2. Select the desired PTM from the Tier *x* box.

   Observe that the amino acid changes color. Each amino acid has its own PTMs.

3. Click **Rescore**.

Fixed Modifications Box

The Fixed Modifications box lists each fixed modification supported by ProSightPC by amino acid.

You can select fixed modifications in the dialog box opened by choosing Tools > Add Search, or you can change them during rescoring.

To indicate that no fixed modifications are presently selected for that type of amino acid and will not be included in the next rescoring, select **None**.

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

Matching Fragments Table

The matching fragments table, shown in Figure 76, contains a summary of all fragment ions matching the protein.

Figure 76. Matching fragments table

The interactive fragment map and matching fragments table are linked for convenient data browsing. Click a fragment name in the table to select the terminal amino acid in the fragment map. You can also select the terminal amino acid of a fragment in the map to highlight the corresponding fragment name in the matching fragments table.

The matching fragments table displays the following columns:

Parameter	Description
ID	Displays a unique (in each MS/MS experiment) ProSightPC-assigned numerical identification of the matching fragment.
Name	Displays a name consisting of the ion type followed by the terminal amino acid number.
m/z *type*	Displays the mass-to-charge ratio (*m/z*) value corresponding to the fragment ion. The type is monoisotopic or average, depending on which you selected during the last rescoring.
Mass *type*	Displays the observed mass of the fragment ion, measured in Da. The type is monoisotopic or average, depending on which you selected during the last rescoring.

Parameter	Description
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	Displays the word `True` if the corresponding fragment is a match, considering the delta (Δm) mode.

Non-Matching Fragments Table

The non-matching fragments table, shown in Figure 77, lists every fragment that does not match the sequence.

Figure 77. Non-matching fragments table

b1 - H { M { Y } I } S } E { T } A } V } S } V } H } K - y1

▶ Show Matching Fragments (Total: 19 fragments)
▼ Show Non-Matching Fragments (Total: 59 fragments)

ID	m/z Monoisotopic	Mass Monoisotopic	Intensity
1	0.0000	413.1521	8127.41
2	0.0000	415.1480	431.09
5	0.0000	516.2519	1161.64
7	0.0000	550.3231	321.61
9	0.0000	576.3017	252.34
10	0.0000	594.3126	502.72
11	0.0000	612.3230	418.44
12	0.0000	613.2681	796.79
13	0.0000	621.3600	277.86
16	0.0000	651.3233	349.58
17	0.0000	722.4074	1520.2
18	0.0000	732.3266	200.95
21	0.0000	768.4018	430.13
22	0.0000	810.3869	1352.94
23	0.0000	814.3857	1667.18
24	0.0000	818.3804	185.72
25	0.0000	825.3465	264.93
26	0.0000	828.3977	1073.84
27	0.0000	833.3992	202.9
28	0.0000	843.3582	1656.58
29	0.0000	851.4497	1112.78
32	0.0000	876.4053	295.28
33	0.0000	894.4150	270.54
34	0.0000	896.3853	484.73
35	0.0000	904.4115	186.5
36	0.0000	905.4589	213.78
37	0.0000	913.4535	277.77
38	0.0000	914.3948	2378.14
39	0.0000	918.3890	239.61
40	0.0000	923.4701	1282.07
41	0.0000	931.4641	1000.97
43	0.0000	938.4808	3119.16

The non-matching fragments table displays the following columns:

Parameter	Description
ID	Displays a unique (in each MS/MS experiment) ProSightPC-assigned numerical identification of the fragment.
m/z (type)	Displays the mass-to-charge ratio (*m/z*) value corresponding to the fragment ion. The type is monoisotopic or average, depending on which you selected during the last rescoring.

Parameter	Description
Mass (type)	Displays the observed mass of the fragment ion, measured in Da. The type is monoisotopic or average, depending on which you selected during the last rescoring.
Intensity	Displays the abundance of the fragment ion.

Demonstrating the Sequence Gazer

Following is a demonstration that shows how to use the Sequence Gazer to refine a good result returned by an absolute mass search with a large mass error into an excellent result with no mass error, as evidenced by decreasing the score by several orders of magnitude.

Displaying Data in the Data Manager

This chapter describes the Data Manager in ProSightPC's graphical user interface.

Contents

- Data Manager
- Opening and Closing a Data Manager Window
- Adding or Editing an Experiment Comment
- Editing Mass Values
- Running a Pending Search
- Editing Search Comments
- Reading the Interactive Fragment Map
- Performing Gene-Restricted Searches

Data Manager

The Data Manager, shown in Figure 78, provides a visual representation of all the information related to a single MS/MS experiment. It appears when you double-click an experiment in the Data Grid. You can use it to view all information for a single experiment. The context-sensitive controls help you determine what information is displayed.

Figure 78. Data Manager
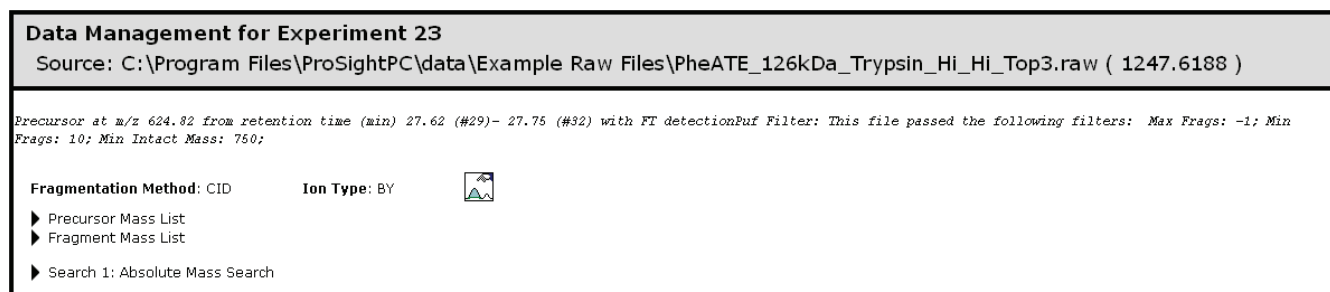
Data Management for Experiment 23
Source: C:\Program Files\ProSightPC\data\Example Raw Files\PheATE_126kDa_Trypsin_Hi_Hi_Top3.raw (1247.6188)

Precursor at m/z 624.82 from retention time (min) 27.62 (#29)– 27.75 (#32) with FT detectionPuf Filter: This file passed the following filters: Max Frags: -1; Min Frags: 10; Min Intact Mass: 750;

Fragmentation Method: CID   **Ion Type**: BY

▶ Precursor Mass List
▶ Fragment Mass List

▶ Search 1: Absolute Mass Search

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

A side arrow indicates hidden information related to a search. Click the side arrow to reveal the information. This step turns the side arrow icon into a down arrow, as shown in .

Figure 79. Information revealed by down arrow in the Data Manager

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

- Instrument data, which includes the mass values, fragmentation method, and ion type of the MS/MS experiment. If you have defined an experiment-level comment, it is displayed at the top of the Data Manager.

- Search data, which is a numerical list of searches arranged by search identification following the instrument data. gives an example. The search type and status are displayed. If the search is highlighted in blue, the search is pending and has yet to be run.

Figure 80. Search numbers

▶ Search 1: Single Protein Search
▶ Search 2: Absolute Mass Search
▶ Search 3: Absolute Mass Search
▶ Search 6: Sequence Tag Search
▶ Search 7: GRAM Search
▶ Search 8: GRAM Search

When an 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 and ETD) fragmentation is analyzed using c/z ions.

• Collision-induced dissociation (CID, HCD, and IRMPD) fragmentation is analyzed as b/y ions.

Opening and Closing a Data Manager Window

❖ **To open a Data Manager window**

• Double-click an experiment displayed in the Data Grid and choose **View > Open Data Manager**, or click the **Open Data Manager** icon, .

❖ **To close a Data Manager window**

• Choose **View > Close Data Tab**.

–or–

• Click the page containing the Data Manager and choose **View > Close Data Manager**.

The experiment page and all pages related to it, such as the Sequence Gazer, disappear from the screen.

Adding or Editing an Experiment Comment

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

❖ **To add or edit an experiment comment**

1. Click **Edit Comment** in the Data Manager, choose **Experiment Tools > Edit Comment**, or click the **Edit Comment** icon, abc.

2. Type the comment in the box that opens in the Data Manager.

      3. Click **Save** to save the comment.

Editing Mass Values

The Data Manager includes a facility for reviewing and editing mass values in an experiment.

If you would like to review the mass values, you can export them to an external application such as an Excel spreadsheet or a file.

You can also edit mass values by adding more precursor or fragment masses, deleting existing precursor or fragment masses, or changing values for precursor or fragment parameters.

In addition, you can change the fragmentation method.

❖ **To edit mass values**

1. Choose **Experiment Tools > Edit Masses**, click the Edit Masses icon, , in the Data Manager or the ProSightPC toolbar, or right-click an experiment and choose **Edit Mass List** from the shortcut menu.

   Each of these methods opens a new page in the Tab Controller, as Figure 81 shows.

Figure 81. Editing mass values

The top part of the Edit Masses *experiment_number* page contains the following parameters and icons:

Parameter	Description
💾	Saves the edits that you have made to the mass values.
❌	Does not save any of the edits that you have made to the mass values, closes the Edit Masses *experiment_number* page, and returns you to the Data Manager.
Fragmentation Method	Specifies the fragmentation method. For more information on fragmentation methods, see "Fragmentation Methods" on page 8.

The mass values are displayed in a series of columns in two areas, Precursor Mass List and Fragment Mass List, on the Edit Masses *experiment_number* page.

The Precursor Mass List area displays the following columns:

Parameter	Description
mz_monoisotopic	Specifies the monoisotopic mass-to-charge ratio (*m/z*) value of the precursor ion.
mz_average	Specifies a column showing the average mass-to-charge ratio (*m/z*) value of the precursor ion.
mass_monoisotopic	Specifies the monoisotopic mass of the precursor ion.
mass_average	Specifies the average mass of the precursor ion.
intensity	Specifies the abundance of the precursor ion.
id	Specifies a unique (in each MS/MS experiment) ProSightPC-assigned numerical identification of the precursor ion.

The Fragment Mass List area displays the following columns:

Parameter	Description
mz_monoisotopic	Specifies the monoisotopic mass-to-charge ratio (*m/z*) value of the fragment ion.
mz_average	Specifies the average mass-to-charge ration (*m/z*) value of the fragment ion.
mass_monoisotopic	Specifies the monoisotopic mass of the fragment ion.
mass_average	Specifies the average mass of the fragment ion.
intensity	Specifies the abundance of the fragment ion.
id	Specifies a unique (in each MS/MS experiment) ProSightPC-assigned numerical identification of the fragment ion.

2. Select any value from either of the two mass lists.

3. Click **Delete** or use the backspace key to remove the old value. Type a new value in the blank space.

4. Click ![save icon] to save the changes.

5. In the Save Masses Confirmation box, click **Yes**.

   ProSightPC returns you to the Data Manager.

❖ **To add a row to a mass list**

1. Click the last row of the Precursor Mass List and Fragment Mass List, which is marked with an asterisk (*).

   Zeros appear in all the columns of this row.

2. Replace the zeros with the desired values.

3. Click ![save icon] to save the changes.

4. In the Save Masses Confirmation box, click **Yes**.

   ProSightPC returns you to the Data Manager.

❖ **To remove a row from a mass list**

1. Click the margin to the left of the mass list to select an entire row.

2. Click **Delete** to remove the entire row from the mass list.

3. Click ![save icon] to save the changes.

4. In the Save Masses Confirmation box, click **Yes**.

   ProSightPC returns you to the Data Manager.

❖ **To copy mass values to an external application**

1. Copy the mass values:

   • To select contiguous rows, hold the SHIFT key down, click in the leftmost column of the first desired row, then click in the leftmost column in the last desired row.

   • To select disparate rows, hold the CTRL key down and click in the leftmost column of each desired row.

2. Press CTRL C (for copy) and select the external application to paste.

❖ **To change the fragmentation method**

• In the Fragmentation Method list, select the new fragmentation method.

   For more information on fragmentation methods, see "Fragmentation Methods" on page 8.

3. Click ![save icon] to save the changes.

4. In the Save Masses Confirmation box, click **Yes**.

   ProSightPC returns you to the Data Manager.

❖ **To return to the Data Manager without applying any changes to the mass values**

• Click ![red X icon].

ProSightPC does not save any of the edits that you have made to the mass values, closes the Edit Masses *experiment_number* page, and returns you to the Data Manager.

Running a Pending Search

❖ **To run a pending search**

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

2. Click **Run Search** to run the search.

Search parameters are discussed in "Searching Databases" on page 125.

A completed search generates a results list, as Figure 82 shows.

Figure 82. 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 in the result table is organized into three regions:

Parameter	Description
Description	Briefly describes the protein or peptide form.
Fragment Map	Graphically represents the protein form, showing the location of PTMs and matching fragment ions.
Statistics Table	Organizes information relating to the search.

The result statistics table is subdivided into the following display elements:

Parameter	Description
ID/Gene	Displays the internal identifier for the protein form and the gene identifier.
Length	Displays the number of amino acids in the protein form.
Mass	Displays the theoretical precursor mass of the protein form.
Mass Difference	Displays the observed mass minus the theoretical mass.
PPM Difference	Displays the mass difference in parts per million.
N-terminal Ions	Displays the total number of matching N-terminal ions.
C-terminal Ions	Displays the total number of matching C-terminal ions.
Total Ions	Displays the total number of matching ions.
PDE Score	Displays the PDE (McLuckey) score. For information on this score, see "PDE (McLuckey) Score" on page 188.
E-Value	Displays the expectation score. For information on this value, see "Expectation Value (e value)" on page 186.
P Score	Displays the p score. For information on this value, see "p Score" on page 186.

Each result has three context-sensitive buttons:

Take to Sequence Gazer	Creates a new single-protein search based on the result. See Chapter 6, "Searching for Single Proteins," for details.
RESID	Displays a RESID-annotated sequence.
SEQ	Displays the sequence.

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

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

Editing Search Comments

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

Reading the Interactive Fragment Map

The matching fragment table contains a summary of all fragment ions matching the protein. For information on the interactive fragment map, see "Interactive Fragment Map" on page 190.

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

Performing Gene-Restricted Searches

You can perform a gene-restricted search for any results list in the Data Manager. For information on this procedure, see "Performing Gene-Restricted Searches" on page 150.

Using Proteome Databases

This chapter describes the proteome warehouse and how to create, manipulate, and modify proteome databases.

Contents

Proteome Warehouse

ProSightPC searches require sequence information to identify and characterize proteins. This sequence information and modification information are stored in proteome databases in ProSightPC's proteome warehouse. The sequence and PTM information are combined using shotgun annotation, as explained in Chapter 1, "Introduction to ProSightPC."

Figure 83 shows all the known modifications, such as SNPs and sequence variants, that can be applied to a basic sequence.

Figure 83. Known modifications applied to a basic sequence

N-terminal processing
(in-vivo half life,
signal peptides)

Polymorphisms

Ac   Me   P$_i$ X

N   C

Known
modifications

Intron/exon
boundaries

Unexpected
modifications
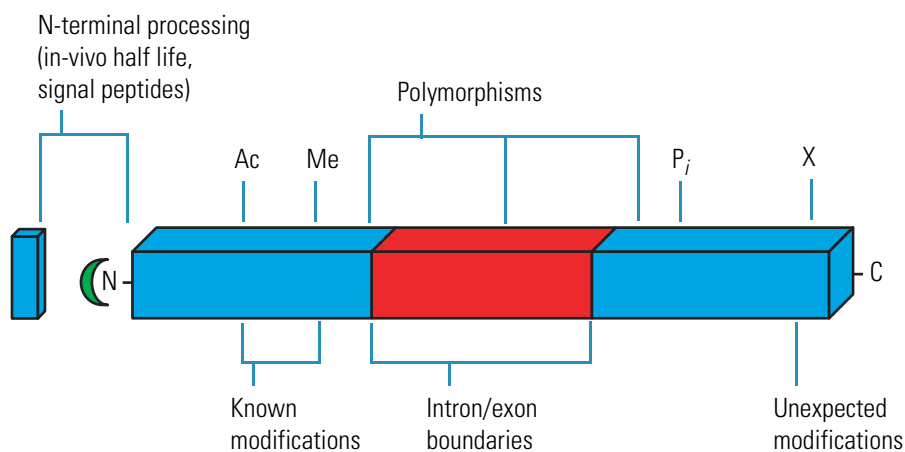
The proteome warehouse consists of a collection of proteome databases and a small amount of metadata. Proteome databases contain the shotgun annotation of all possible combinations of known modifications on each basic sequence in the proteome. A protein form refers to any given possible combination of modifications on a basic sequence.

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, you might designate E. coli (UniProt) as ecoli_uniprot.

Importing Data into the Proteome Warehouse

Use either of the following two methods to import data into the proteome warehouse:

- Load databases from .pwf files. See "Importing a Proteome Database or Repository" on page 210 for details.

- Create databases from Swiss-Prot or FASTA-formatted text files. See "Creating a Proteome Database" on page 215 for details.

Accessing the Database Manager

Use the Database Manager to handle all proteome warehouse and repository management and manipulation functions.

❖ **To access the Database Manager**

- Choose **Databases > Database Manager**, or click the **View Database Info** icon, .

  The Database Manager window opens, as shown in Figure 84.

Figure 84. Database Manager window
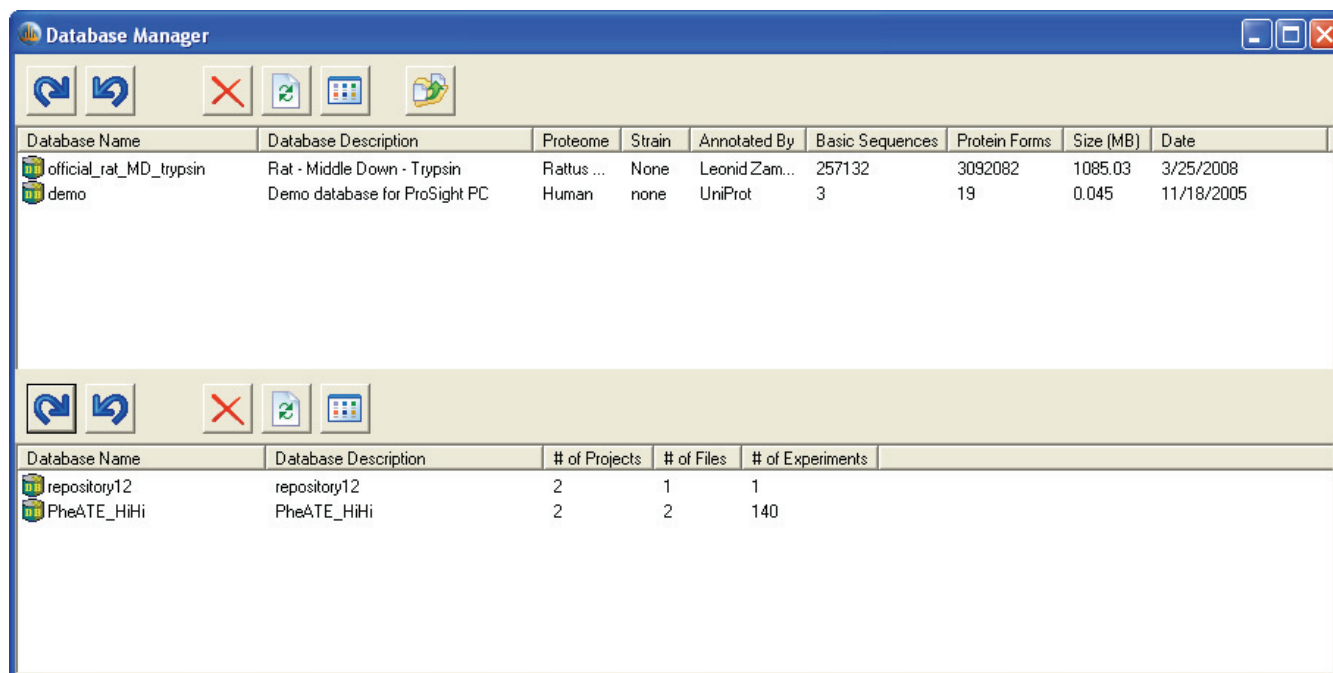
Database Manager Window Parameters

The Database Manager window contains the following parameters and toolbar icons.

Parameter	Description
Database Name	Displays the name of the proteome database. This name must be unique.
Database Description	Displays a brief description of the proteome database.
Proteome	Indicates the type of organism for the proteome database.
Strain	Lists the strain information for the proteome database.
Annotated By	Lists the source of the proteome data.
Basic Sequences	Lists the number of unmodified protein forms in the proteome database.
Protein Forms	Lists the total number of shotgun-annotated protein forms in the proteome database.
Size (MB)	Lists the physical size of the proteome database, in megabytes.
Date	Displays the date that the proteome database was created.
(top toolbar)	Imports a proteome database.

Parameter	Description
(top toolbar)	Exports a proteome database.
(top toolbar)	Removes a proteome database.
(top toolbar)	Refreshes the database data displayed in the top half of the Database Manager window.
(top toolbar)	Changes the view of the databases listed in the top half of the Database Manager window.
(top toolbar)	Activates the Create New Database wizard so that you can create a new proteome database.
(middle toolbar)	Imports a repository.
(middle toolbar)	Exports a repository.
(middle toolbar)	Removes a repository.
(middle toolbar)	Refreshes the repository data displayed in the bottom half of the Database Manager window.
(middle toolbar)	Changes the view of the repositories listed in the bottom half of the Database Manager window.

Importing a Proteome Database or Repository

Use the following procedures to load an existing proteome database or repository, respectively, into the ProSightPC proteome warehouse.

❖ **To import a proteome database**

1. Click the **Import Proteome Database** icon, , in the toolbar at the top of the Database Manager window.

   A dialog box appears so that you can enter or browse for the name of the .pwf.

2. Select the .pwf that you want, and click **Open**.

   The Import Database(s) dialog box appears, as shown in Figure 85.

Figure 85. Import Database(s) dialog box
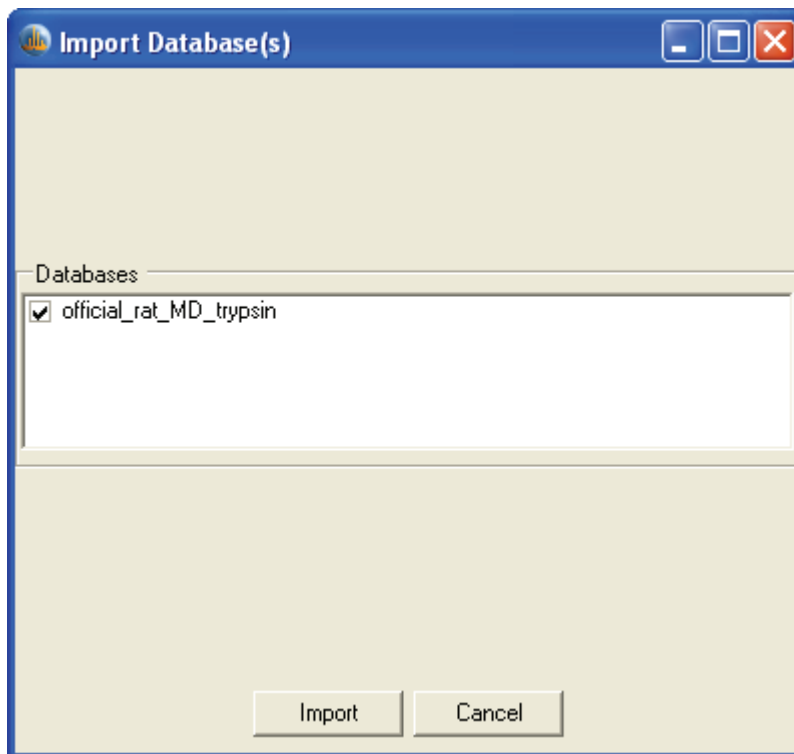
3. Select the check box next to the names of the database or databases that you want to import and click **Import**.

4. If a message box appears, indicating that the repository was imported successfully, click **OK**.

   If any internal name already exists in the proteome warehouse, this message box does not appear; instead, the Internal Name Conflict dialog box opens, as shown in Figure 86.

Figure 86. Internal Database Name Conflict dialog box
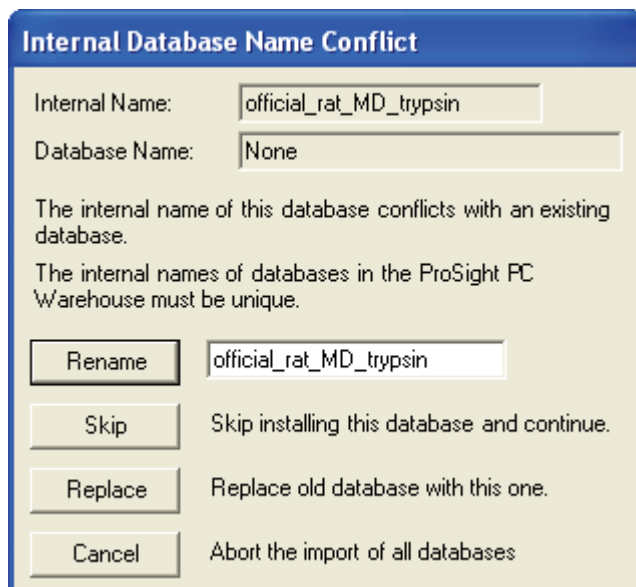
Use this dialog box to rename the database, skip the database, replace the database, or cancel changes.

> **Note** Importing databases as .pwf files is faster than recreating a proteome database with the Create Proteome Database option, a process detailed in "Creating a Proteome Database" on page 215.

❖ **To import a repository**

1. Click the **Import Repository** icon, , in the toolbar in the middle of the Database Manager window.

   A dialog box appears so that you can enter or browse for the name of the .pwf.

2. Browse to the repository .pwf, select it, and click **Open**.

3. In the Import Database(s) dialog box, shown in Figure 85, select the .pwf and click **Import**.

4. Click **OK** in the message box indicating that the repository was imported successfully.

Exporting a Proteome Database or Repository

You can also export one or more proteome databases or repositories from your proteome warehouse to a .pwf.

❖ **To export a proteome database**

1. Click the **Export Proteome Database** icon, , in the toolbar at the top of the Database Manager window.

The Export Database(s) dialog box opens, as shown in Figure 87.

Figure 87. Export Database(s) dialog box
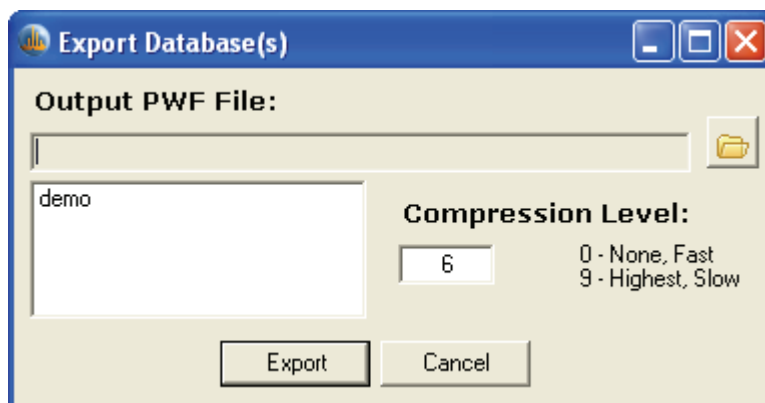
2. Select the database to export. To combine multiple databases into a single export file, hold the CTRL key down and select additional databases.

3. Click **Export**.

   A new dialog box opens.

4. Type the name of a destination .pwf, or browse to it.

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

6. Click **Export** to execute the procedure and create a new .pwf.

❖ **To export a repository**

1. Click the **Export Repository** icon, .

2. In the Export Database(s) dialog box, shown in Figure 87, select the database to export. To combine multiple databases into a single export file, hold down the CTRL key and select additional databases.

3. Click **Export**.

   A new dialog box opens.

4. Type a destination .pwf or browse to it.

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

6. Click **Export** to execute the procedure and create a new .pwf.

Removing a Proteome Database or Repository

You can remove unwanted proteome databases and repositories from the proteome warehouse by using the Database Manager.

❖ **To remove a proteome database**

1. Select the proteome database to remove in the Database Manager.

2. Click the **Remove Database** icon, ✕, in the toolbar at the top of the Database Manager window.

3. Click **Yes** to remove the proteome database from the proteome warehouse.

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

❖ **To remove a repository**

1. Select the repository to remove in the Database Manager.

2. Click the **Remove Repository** icon, ✕, in the toolbar in the middle of the Database Manager window.

3. Click **Yes** to remove the repository from the proteome warehouse.

> **CAUTION** Removing a repository from the proteome warehouse is a permanent change and cannot be undone except by reloading the data from the original source into the proteome warehouse.

Changing View

❖ **To change the display of databases in the Database Manager**

1. Click the **Change View** icon, ▦, on the Database Manager to change the display of databases.

2. You can choose one of the following display modes:

   • Details: Lists the names of the loaded databases and gives a brief description of each database, its proteome, its strain, by whom it was annotated, its basic sequences, its protein forms, its size, and the date that it was created.

   • List: Vertically lists the names of the loaded databases, starting in the upper left corner of the Database Manager.

   • Small Icons: Horizontally lists the names of the loaded databases, starting in the upper left corner of the Database Manager. Each name is preceded by a small icon.

- Large Icons: Horizontally lists the names of the loaded databases, starting in the upper left corner of the Database Manager. Each name is preceded by a large icon.

> **Note** You can drag files in the Database Manager.
>
> - .pwf files trigger the Import Proteome Database or Repository function.
>
> - FASTA and Swiss-Prot files trigger the Create Proteome Database function.

Creating a Proteome Database

ProSightPC supports the creation of top-down and middle-down/bottom-up databases.

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

- Swiss-Prot files, which store a large amount of known modification information. Swiss-Prot, a curated biological database of protein sequences, is a part of the UniProt database.

  All PTMs listed in RESID are available for shotgun annotation. However, only the PTM information in the source Swiss-Prot file can actually be processed into the proteome database.

- FASTA files, which contain no PTM information, so only predicted PTMs can be processed to their sequences. FASTA format represents either nucleic acid sequences or peptide sequences, where single-letter codes represent base pairs or amino acids. A sequence in FASTA format begins with a single-line description, followed by lines of sequence data.

  In FASTA format, sequences of nucleic acids or peptides begin with a lone-line description, followed by lines of sequence data. Single-letter codes represent base pairs or amino acids.

> **IMPORTANT** Creating a proteome database can take several hours or, in some cases, days to complete, depending on your hardware and the data being processed by the shotgun-annotation method.

❖ **To create a proteome database**

1. Choose **Databases > Create Custom Database** or click the **Create Database** icon, , on the Database Manager.

   The Welcome to the New Database Wizard page of the Create New Database wizard opens, as shown in Figure 88.

Figure 88. Welcome to the New Database Wizard page of the Create New Database wizard

2. Click **Next**.

   The Database Type page of the Create New Database wizard appears, as shown in Figure 89.

Figure 89. Database Type page of the Create New Database wizard
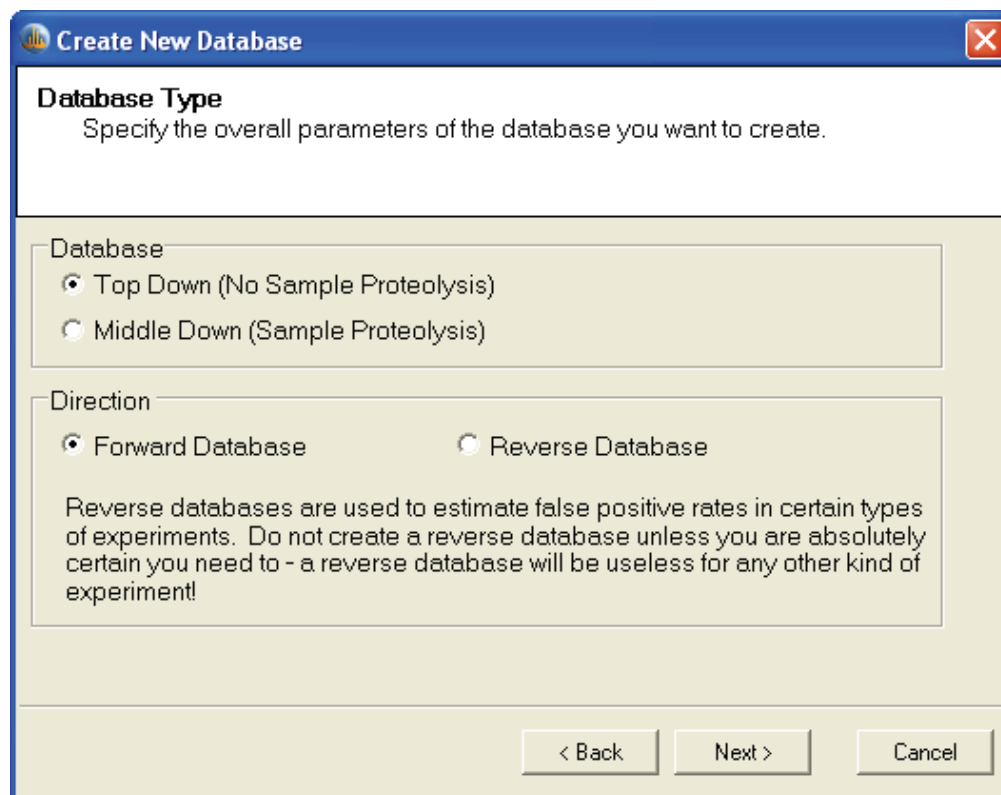
3. In the Database area, select the type of database to build:

   • Top Down (No Sample Proteolysis): Builds a database around whole, intact protein sequences and everything that could potentially happen to them in a biological system.

   • Middle Down (Sample Proteolysis): Builds a database around peptide sequences that arose from ex vivo proteolysis. If anything in your sample preparation protocol involves trypsin or Lys-C or any other proteolytic agent, select this setting.

4. In the Direction area, select the direction of the database to build. You can select either of the following options:

   • Forward Database: A forward database is a typical protein database.

   • Reverse Database: Creates a nonsense database consisting of correct masses but reverse sequences. A reverse database is a way to statistically validate a false positive/false negative confidence level. Reverse databases are sometimes used for estimating false positive rates in certain types of experiments. Typically, you would create a normal database first, and then use the same settings with reversal turned on to create a proper reverse database. Do not select this option unless it is absolutely necessary. You cannot use a reverse database in any other kind of experiment.

5. Click **Next**.

The Input File page of the Create New Database wizard appears, as shown in Figure 90.

Figure 90. Input File page of the Create New Database wizard
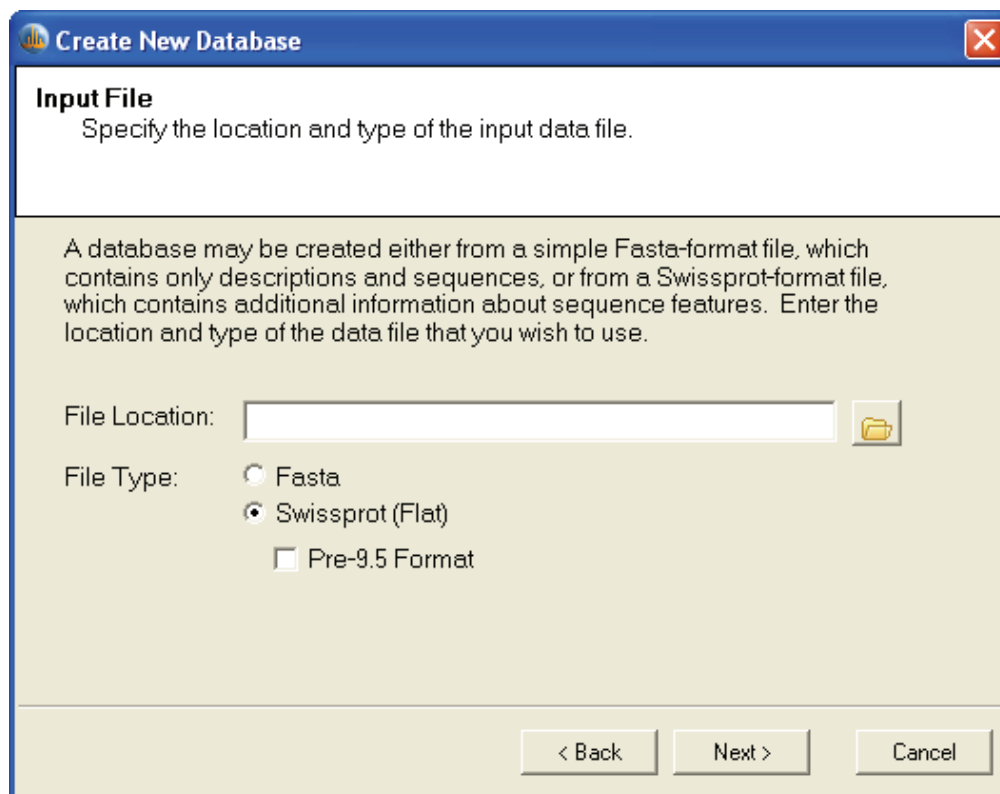
6. In the File Location box, enter the name and path of the file containing the sequence information, or browse for it by clicking the **Browse** button.

   ProSightPC generates the database from this data file. Before you load the file, open it in a text editor to ensure that it is free from errors. Most errors in loading result from bad input files. A good source for input files is the Web site of the UniProt consortium. If you are going to create your own input file, make sure that the encoding is correct. Notepad can sometimes mishandle the encoding of newline characters in the file. If you receive errors, try a proper text editor.

   The name of the data file must be unique.

7. For File Type, select the format of your input file:

   • FASTA, which is described at the beginning of "Creating a Proteome Database" on page 215

   • Swiss-Prot, which is described at the beginning of "Creating a Proteome Database" on page 215

Pre-9.5 Format: Indicates that the Swiss-Prot database is a UniProt database before version 9.5 when sequences were presented without the initial methionine. Residues are numbered starting from 1, so if there was an initial methionine that was cleaved off to produce the sequence, an INIT_MET tag was placed in the file with a location at residue 0. After version 9.5, initial methionines were included in the sequence, and if the initial methionine was cleaved off, the INIT_MET tag was added at position 1. When you select the Pre-9.5 Format option, ProSightPC adds a methionine to the input sequence to produce a form with an initial methionine. When you clear the Pre-9.5 Format option, ProSightPC removes a methionine from the input sequence to produce a form without an initial methionine.

8. Click **Next**.

The Initial Methionines page of the Create New Database wizard appears, as shown in Figure 91.

Figure 91. Initial Methionines page of the Create New Database wizard
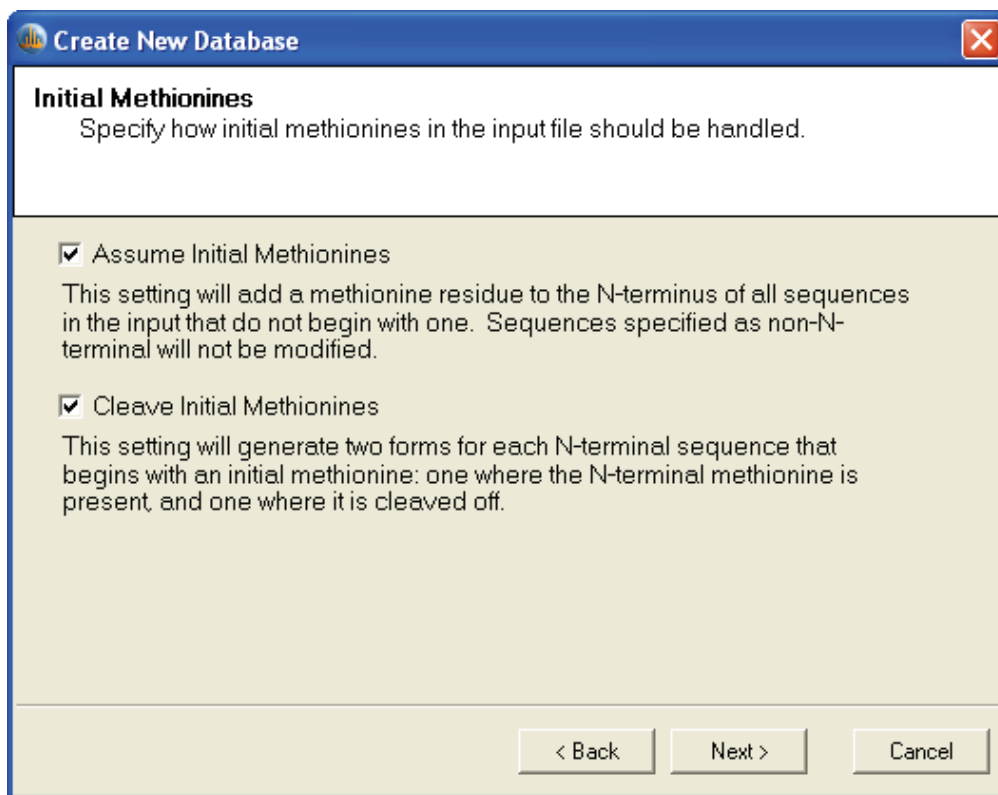
9. Select the method of handling initial methionines:

- Assume Initial Methionines (recommended): Adds an initial methionine to the N terminus of all sequences of the input file if you do not explicitly specify one in your input sequences. This means that when you are loading from a FASTA file, which has just a description and the raw sequence with no sequence markup, you must select this setting to obtain any initial methionine cleavage forms. If you are loading from a Swiss-Prot file, which has some known and predicted initial methionines, selecting this setting adds initial methionines to any input sequence that did not have them already and normalizes existing features, if necessary.

- Cleave Initial Methionines (recommended): Explicitly specifies the initial methionine cleavage that was either in the input file to begin with (Swiss-Prot files) or added earlier during database creation (FASTA and Swiss-Prot files). This method creates two forms for any N-terminal sequence with such a feature: an initial methionine present form and an initial-methionine absent form. This is a very common sequence modification. Disable it only if absolutely necessary.

10. Click **Next**.

The Feature Integration page of the Create New Database wizard appears, as shown in Figure 92.

Figure 92. Feature Integration page of the Create New Database wizard
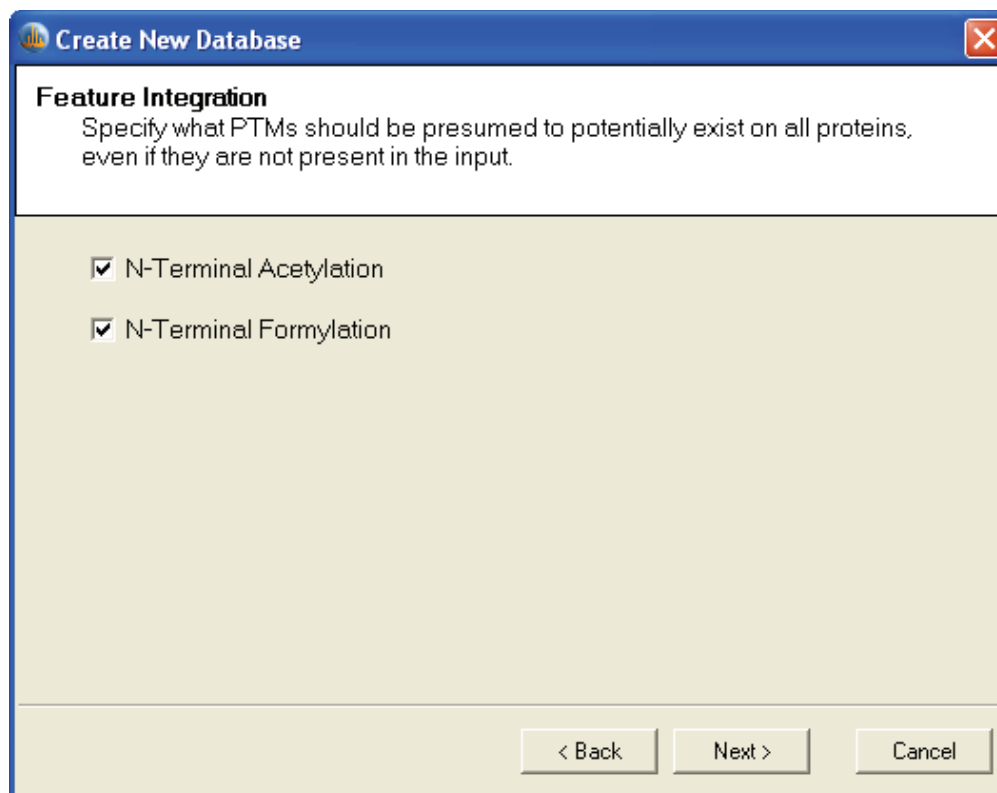
11. Specify the PTMs that should be presumed to exist on all proteins, even if the input does not contain them:

    • N-Terminal Acetylation: Adds N-terminal acetylation whenever it is possible, regardless of whether the input includes it. N-terminal acetylation is a very common PTM.

    • N-Terminal Formylation: Formylates N-terminal methionines. Select this option if you are building a prokaryotic database. Prokaryotes use N-formylmethionine for initiation.

    • Click **Next**.

    The Complexity page of the Create New Database wizard appears, as shown in Figure 93.

    Figure 93. Complexity page of the Create New Database wizard
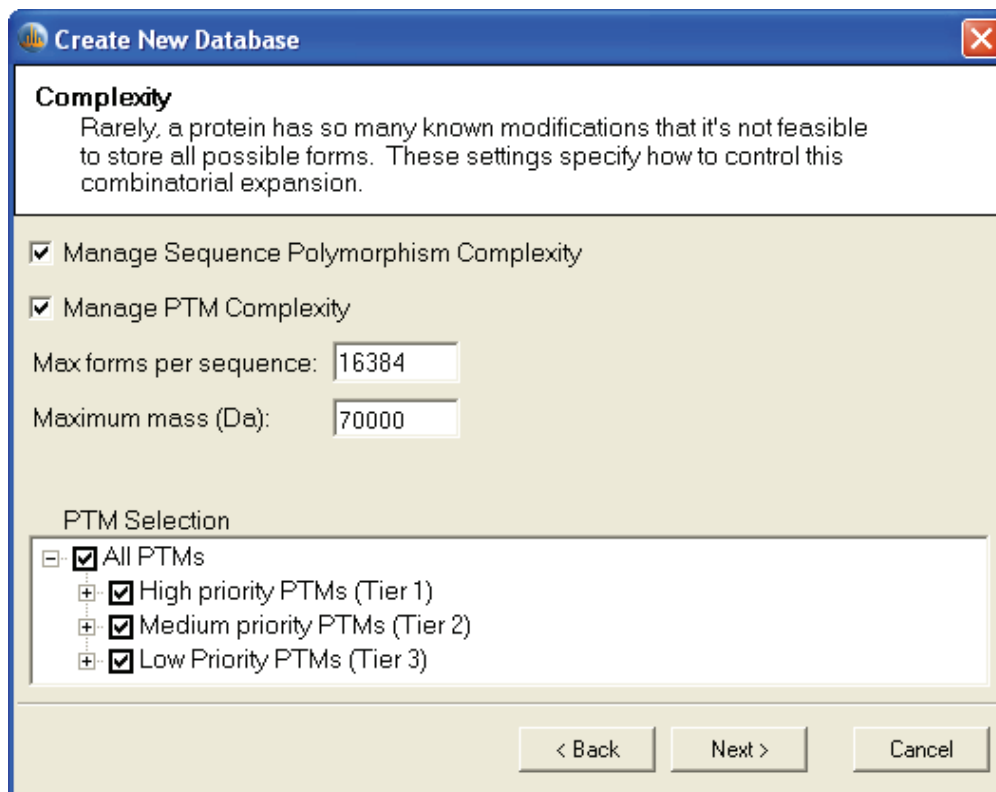
As indicated on the Complexity page, a protein may have so many known modifications that it is not feasible to store all possible forms. On this page, you can set options to specify how to control this combinatorial expansion. If you are uncertain about the values to set, use the default values.

12. If you are using an input file in Swiss-Prot/UniProt format and are creating a database of an organism with a large number of polymorphisms, select the **Manage Sequence Polymorphism Complexity** option.

Shotgun annotation scales as 2^n, where n is the total number of polymorphisms. This option limits the total number of polymorphisms that are included in the database to only the most relevant polymorphisms.

13. If you are creating a database of an organism with a large number of PTMs, select the **Manage PTM Complexity** option.

    This option limits the total number of PTMs that are included in the database to only the most relevant PTMs. It is only available if you are using an input file in Swiss-Prot/UniProt format.

14. In the Max forms per sequence box, enter the maximum number of forms per input sequence.

    This option sets the maximum number of database forms produced from a particular entry. If a protein has, for example, four PTMs, ProSightPC puts 2^4, or 16, forms into the database. The default value is 16384.

    Here is an example: A short peptide (SSS) has a phosphorylation feature on each residue. A PTM might or might not be present. The total number of forms that can be produced from this sequence is $2^3 = 8$: SSS, S(P)SS, SS(P)S, SSS(P), S(P)S(P)S, S(P)SS(P), SS(P)S(P), S(P)S(P)S(P). Forms are selected on the basis of which ones are most likely to be observed in the instrument: for the input sequence just given, if the restriction is set to $2^2 = 4$ forms, the forms selected are SSS, S(P)SS, SS(P), S(P)SS(P).

    Here is an example of an entry in the input with potential variation, such as polymorphisms and PTMs:

    MAAAVAAAPAAAA

    PTM: 3

    This protein *might have* a PTM at A3. A3 is a known site of modification.

    A form is in the database. It has no variation and is matched directly against the data.

    MA(PTM)AAVAAAPAAAA

    This protein *has* a PTM at A3.

    MAAAVAAAPAAAA

    This protein *does not have* a PTM at A3.

    This example also demonstrates shotgun annotation: from information about a known site that can be modified in an input entry, two database forms are produced: one where the site *is* modified and one where the site *is not* modified. If more known sites were known, database forms would be produced with all combinations.

15. In the Maximum mass (Da) box, enter the cutoff point for the THRASH algorithm when searching for masses.

The default for top-down databases is 70000 Da. This option is not seen in middle-down databases, where it is hardcoded to 50000 Da. If the mass of *just* the amino acids in your entry (PTM masses not considered) exceeds the cutoff, the optimizer does not determine which PTMs to pick; instead, it marks all PTMs as inactive.

This option sets the mass cutoff for complexity management; any entry exceeding the maximum mass will have variation (both polymorphisms and PTMs) discarded. Your instrument will probably not see anything beyond a certain size, and because bigger proteins typically have more PTMs, polymorphisms, or both, they will have a disproportionate impact on database size. This option can help resolve that problem.

16. In the PTM Selection area, specify which PTMs should be considered for inclusion in the database.

    If a PTM (or PTM category) is clear, those PTMs are not put into the database, whether or not they are present in the input data. This option is only available for Swiss-Prot-formatted input data, because standard FASTA format cannot encode information about PTMs.

17. Click **Next**.

    The Sequence Polymorphisms page of the Create New Database wizard appears, as shown in Figure 94. Use this page to specify how potentially observable protein forms should be generated from sequence polymorphisms.

Figure 94. Sequence Polymorphisms page of the Create New Database wizard
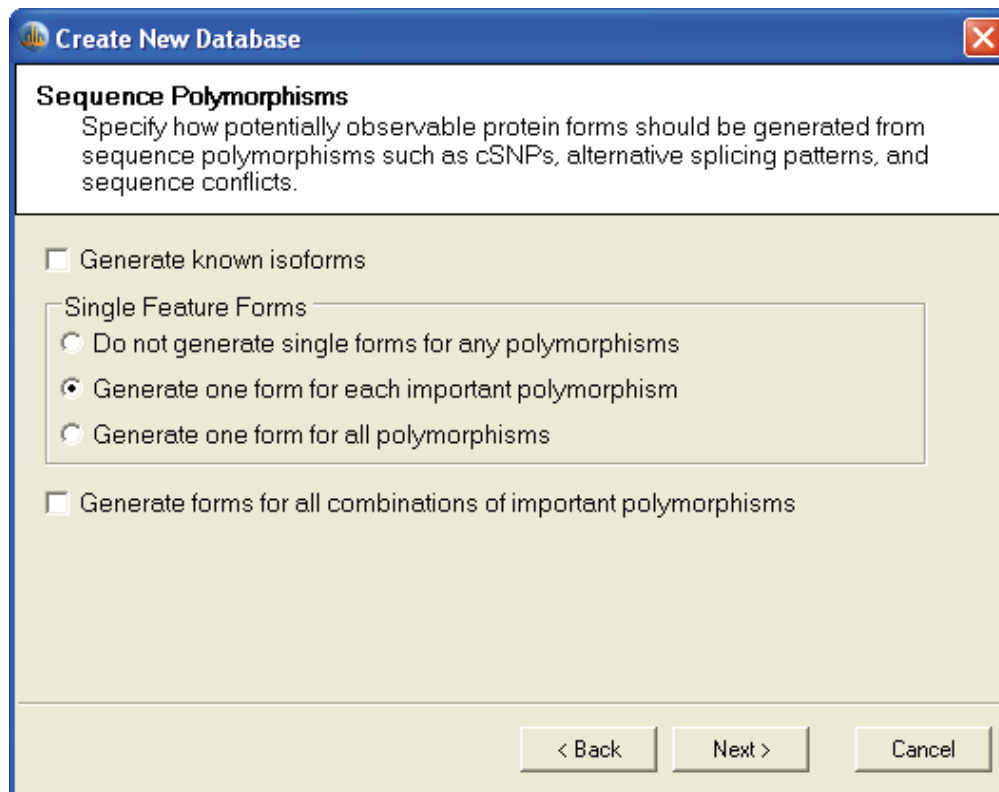
The options on this screen are available only when you are using Swiss-Prot-formatted data. FASTA format cannot encode information about sequence polymorphisms.

18. If you are loading eukaryotic sequences from a Swiss-Prot file, select the **Generate known isoforms** check box.

This option produces all known alternative splice (and alternative initiation and so forth) forms that are specified in the input file. It substantially increases database size and database creation time.

19. In the Single Feature Forms section, select an option for handling sequence polymorphisms:

- Do not generate single forms for any polymorphisms: Does not generate any new sequences for polymorphisms, such as SNPs and variants. Select this option unless you specifically know what type of polymorphism you want to generate a sequence for.

- Generate one form for each important polymorphism: Generates one new sequence for each "important" amino-acid-changing polymorphism where that polymorphism is applied (what is "important" is determined by the complexity optimizer). Select this option if you optimized cSNPs earlier on the Complexity page. For example, consider an input sequence of AAA with three AA-changing polymorphisms: an important polymorphism at the first position (A -> P), an unimportant polymorphism at the second position (A -> C), and an important polymorphism at the third position (A -> M). If this setting is selected, the following sequences are produced: AAA, PAA, AAM.

- Generate one form for all polymorphisms: Generates one form for all polymorphisms, regardless of the importance assigned to them by the optimizer. If you have not optimized cSNPs, the Generate one form for each important polymorphism option is identical to the Generate one form for all polymorphisms option. If you have not run the optimizer, all polymorphisms are equally important.

20. Only if you are performing targeted protein analysis, select the **Generate forms for all combinations of important polymorphisms** option.

This option performs full combinatorics on important polymorphisms. In the previous example, it would produce the following sequences: AAA, PAA, AAM, PAM. This increases the number of sequences exponentially, and each of the resulting sequences has another exponential expansion for PTM combinatorics. Use this option only in the rarest of cases.

21. Click **Next**.

The Endogenous Peptide Cleavage page of the Create New Database wizard appears, as shown in Figure 95. Use this page to specify whether known endogenous peptides should be generated from their precursors.

Figure 95. Endogenous Peptide Cleavage page of the Create New Database wizard
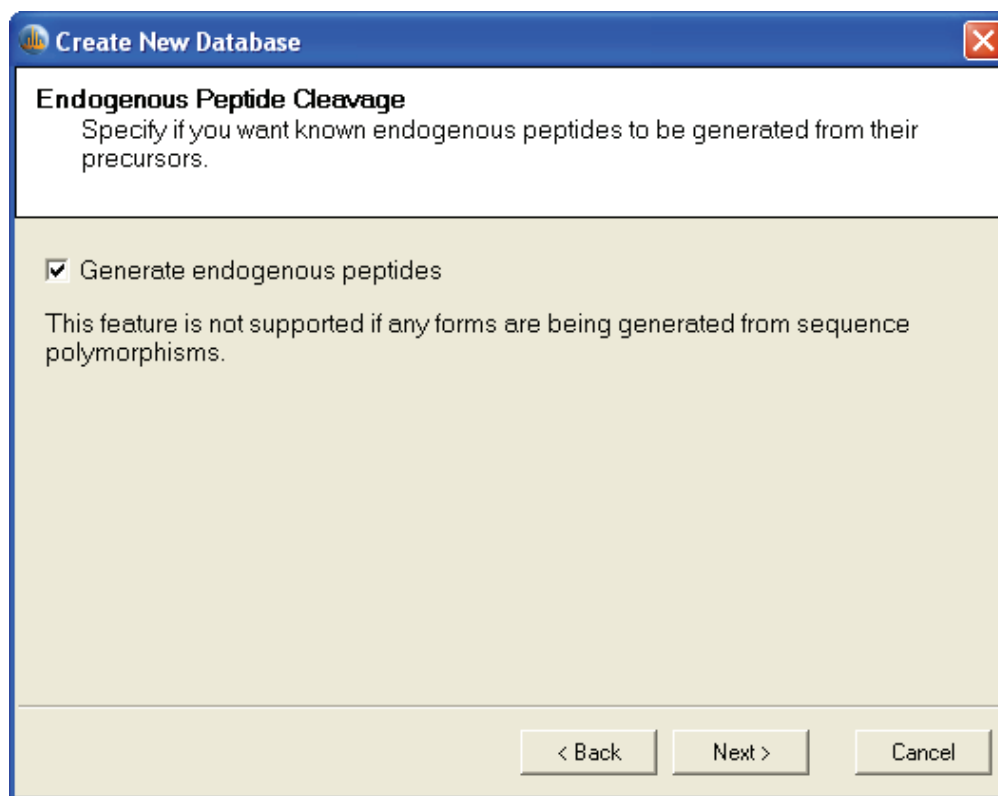
22. Select the **Generate endogenous peptides** check box.

    Endogenous peptide cleavage generates peptides known to be produced in vivo from protein entries. This type of cleavage is only available for Swiss-Prot-format input because it relies on the Swiss-Prot peptide identifiers, such as N-terminal signal peptides, transit peptides, and active chains, to tell it where to cleave.

23. Click **Next**.

    If you selected the Middle Down (Sample Proteolysis) option on the Database Type page, shown in Figure 89 on page 217, the Digestion page of the Create New Database wizard opens, as shown in Figure 96. Use it to specify the parameters for a sample proteolysis.

Figure 96. Digestion page of the Create New Database wizard

24. In the Method list, select the proteolytic method used to catalyze the breakdown of proteins into peptides.

25. In the Max missed cleavages box, type the maximum number of cleavage sites found in the generated peptides. No (0) missed cleavages indicates that there are no cleavage sites in the generated peptides. One (1) missed cleavage indicates that each peptide has one site in it, two (2) missed cleavages indicate that each peptide has two sites in it, and so on. The parameter in the Max missed cleavages box contains all values up to and including the set parameter. For example, if Max missed cleavages is set to 2, peptides with 0, 1, and 2 missed cleavages are generated.

Here is a longer example: If a peptide is AAAKAAAKAAA, and the digestion method is Lys-C, no missed cleavages result in the following peptides:

    AAAK

         AAAK

            AAA

If you select up to one missed cleavage, the peptides are the following:

AAAK

        AAAK

                AAA

AAAKAAA

      AAAKAAA

If you select two missed cleavages, the peptides are the following:

AAAK

        AAAK

                AAA

AAAKAAA

      AAAKAAA

AAAKAAAKAAA

26. In the Minimum peptide mass (Da) box, type the minimum mass that a peptide must have, in daltons, before it is allowed to be put into the database.

    No peptide less than the minimum peptide mass is put into the database; any theoretical peptide less than this mass is discarded and ignored. This parameter is useful because particularly small peptides sometimes cannot be identified but have a very strong impact on database size.

27. In the Maximum peptide mass (Da) box, type the maximum mass that a peptide can have, in daltons, before it is allowed to be put into the database.

    No peptide greater than this mass is put into the database.

28. Click **Next**.

    The Database Description page of the Create New Database wizard opens, as shown in Figure 97, so that you can enter identifying information about the database that you want to create.

Figure 97. Database Description page of the Create New Database wizard

a. In the Database Name box, type the name of the database that you want to create. Use only letters, numbers, and underscores.

b. In the Description box, type a brief description of the database.

c. In the Organism box, type the name of the organism for the proteome database that you are creating.

d. (Optional) In the Strain box, type the strain designation for the proteome database that you are creating.

e. In the Owner box, type either your name or the name of the data source.

f. In the Last Update box, type the date when the database was last updated or click the down arrow to display a calendar and select a different date.

29. Click **Finish**.

30. On the Ready to Load page, click **Go** to create the new database.

Create New Database Wizard Parameters

The pages of the Create New Database wizard contain the following parameters.

Database Type Page Parameters

The Database Type page of the Create New Database wizard, shown in Figure 89 on page 217, contains the following parameters:

Parameter	Description
Top Down (No Sample Proteolysis	Builds a database around whole, intact protein sequences and everything that could potentially happen to them in a biological system.
Middle Down (Sample Proteolysis)	Builds a database around peptide sequences that arose from ex vivo proteolysis.
Forward Database	Creates a database consisting of correct masses and forward sequences. A forward database is a typical protein database.
Reverse Database	Creates a database consisting of correct masses but reverse sequences.

Input File Page Parameters

The Input File page of the Create New Database wizard, shown in Figure 90 on page 218, contains the following parameters:

Parameter	Description
File Location	Specifies the name and path of the file containing the sequence information.
File Type	Specifies the format of the input file that the database will be generated from. You can select FASTA or Swiss-Prot, which are described at the beginning of "Creating a Proteome Database" on page 215.

Initial Methionines Page Parameters

The Initial Methionines page of the Create New Database wizard, shown in Figure 91 on page 219, contains the following parameters:

Parameter	Description
Assume Initial Methionines	Adds an initial methionine to the N terminus of all sequences of the input file if you do not explicitly specify one in your input sequences.
Cleave Initial Methionines	Creates two forms for each N-terminal sequence that begins with an initial methionine: an initial methionine present form and an initial methionine absent form.

Feature Integration Page Parameters

The Feature Integration page of the Create New Database wizard, shown in Figure 92 on page 220, contains the following parameters:

Parameter	Description
N-Terminal Acetylation	Adds N-terminal acetylation, which is a common PTM, whenever it is possible, regardless of whether the input includes it.
N-Terminal Formylation	Formylates N-terminal methionines.

Complexity Page Parameters

The Complexity page of the Create New Database wizard, shown in Figure 93 on page 221, contains the following parameters:

Parameter	Description
Manage Sequence Polymorphism Complexity	Limits the total number of polymorphisms that are included in the database to only the most relevant polymorphisms.
Manage PTM Complexity	Limits the total number of PTMs that are included in the database to only the most relevant PTMs.
Max forms per sequence	Specifies the maximum number of forms per input sequence.
Maximum mass (Da)	Specifies the cutoff point for the THRASH algorithm when searching for masses.
PTM Selection	Specifies which PTMs should be considered for inclusion in the database.

Sequence Polymorphisms Page Parameters

The Sequence Polymorphisms page of the Create New Database wizard, shown in Figure 94 on page 223, contains the following parameters:

Parameter	Description
Generate Known Isoforms	Generates all known alternative splice (and alternative initiation and so forth) forms that are specified in the input file.
Do not generate single forms for any polymorphisms	Does not generate any new sequences for polymorphisms.
Generate one form for each important polymorphism	Generates one new sequence for each "important" amino-acid-changing polymorphism where that polymorphism is applied. (What is "important" is determined by the complexity optimizer.)
Generate one form for all polymorphisms	Generates one form for all polymorphisms, regardless of the importance assigned to them by the complexity optimizer.
Generate forms for all combinations of important polymorphisms	Performs full combinatorics on important polymorphisms.

Endogenous Peptide Cleavage Page Parameters

The Endogenous Peptide Cleavage page of the Create New Database wizard, shown in Figure 95 on page 225, contains the following parameter:

Parameter	Description
Generate Endogenous Peptides	Generates peptides known to be produced in-vivo from protein entries. It is only available for Swiss-Prot-format input, because it relies on the Swiss-Prot peptide identifiers, such as N-terminal signal peptides, transit peptides, and active chains, to tell it where to cleave.

Digestion Page Parameters

The Digestion page of the Create New Database dialog box, shown in Figure 96 on page 226, contains the following parameters:

Parameter	Description
Method	Specifies the proteolytic method used to catalyze the breakdown of proteins into peptides.
Max missed cleavages	Specifies the maximum number of cleavage sites found in the generated peptides.

Parameter	Description
Minimum peptide mass (Da)	Specifies the minimum mass that a peptide must have, in daltons, before it is allowed to be put into the database. No peptide less than the minimum peptide mass is put into the database.
Maximum peptide mass (Da)	Specifies select the maximum mass that a peptide can have, in daltons, before it is allowed to be put into the database. No peptide greater than the maximum peptide mass is put into the database.

Database Description Page Parameters

The Database Description page of the Create New Database wizard, shown in Figure 97 on page 228, contains the following parameters:

Parameter	Description
Database Name	Specifies the name of the database that you want to create.
Description	Describes the database that you want to create.
Organism	Specifies the name of the organism for the proteome database that you want to create.
Strain	Specifies the strain designation for the proteome database that you want to create.
Owner	Specifies the name of the data source.
Last Update	Specifies the date when the database was last updated.

Linking to the UniProt Repository

UniProt is an international repository of organisms containing all the proteins and genes that are known for a specific organism. When you create a custom database, you must have a FASTA or flat text file that contains all the known proteins of interest. You can download those files from UniProt.

For example, suppose that you want to create a custom database for a fly. You would download a flat file from UniProt and use it in the Create New Database wizard.

❖ **To link to the UniProt repository**

- Choose **Databases > Link To Uniprot**.

  This command opens a Web browser with the appropriate UniProt address.

Using ProSightPC Tools

This chapter describes the utilities included in ProSightPC.

Contents

- Reducing Chemical Noise with the Noise Reducer
- Locating and Selecting PTMs with the PTM Tier Editor
- Viewing Fragments Ions with the Fragment Predictor
- Converting Text to ProSightPC Font with the Font Converter

Reducing Chemical Noise with the Noise Reducer

You can use ProSightPC's Noise Reducer to reduce chemical noise.

Fragmentation data frequently contains chemical noise peaks. These undesirable peaks consist 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. The Noise Reducer identifies and minimizes the effects of chemical noise in an MS/MS experiment.

Once you have run it, the Noise Reducer places a new experiment identical to the source experiment on the Data Grid. The new experiment lacks the undesirable fragment masses, and previous experiment results are removed because they might contain the deleted noise peaks.

Table 3 describes three sources of chemical noise.

Table 3. 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.

Table 3. Sources of chemical noise

Source	Description
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.

Reducing Noise

❖ **To reduce noise with the Noise Reducer**

1. Open the Data Manager of the experiment to be processed by double-clicking the experiment.

2. Choose **Experiment Tools > Reduce Noise** or click the **Reduce Noise** icon, 🏃 .

   The Noise Reducer window opens, as shown in Figure 98.

Figure 98. Noise Reducer window
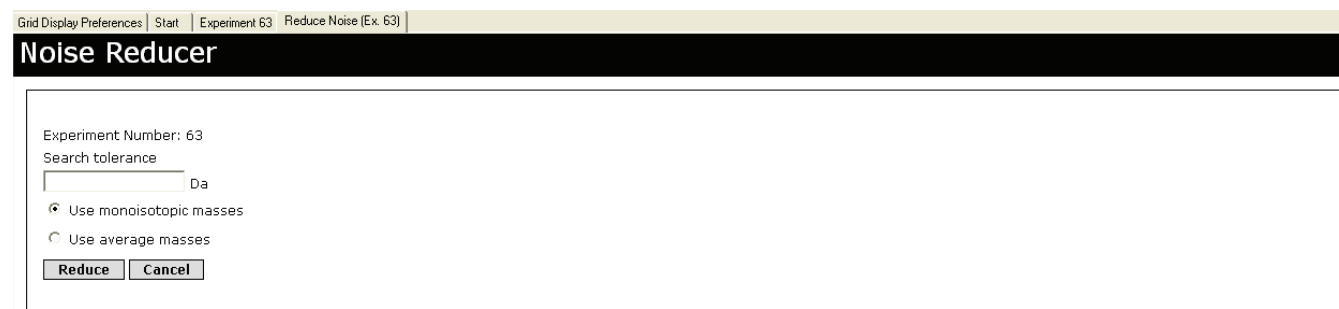
> **Note** The Noise Reducer applies to the MS/MS experiment in the active Data Manager. If no Data Manager is active, the Noise Reducer is unavailable.

3. In the Search tolerance box, type the allowable error, measured in daltons, between the expected and observed mass differences.

4. Indicate whether the Noise Reducer should consider monoisotopic or average mass types:

   • Monoisotopic: Specifies that the mass is monoisotopic, which is the mass of the protein, peptide, or fragment ion, where all carbons are carbon-12.

   • Average Mass: Specifies that the mass is the mass of the most abundant isotope of the protein, peptide, or fragment ion.

5. Click **Reduce** to activate the Noise Reducer.

ProSightPC opens the Reduce Noise page featuring the Noise Reducer output, shown in the example in Figure 99. The output consists of sets of fragment ion masses. The fragment ion masses are grouped together by the following.

- Differences of one dalton (Da)

- Loss of one NH_3

- Loss of one H_2O

Figure 99. Noise Reducer output
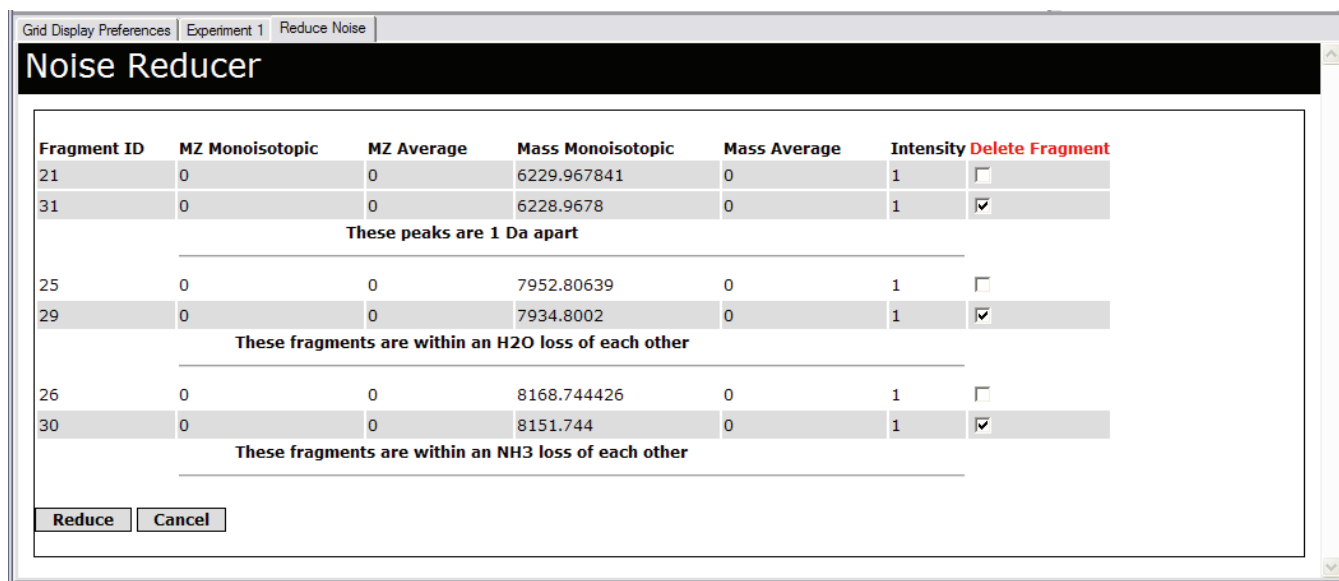
The Noise Reducer page contains the following columns:

Parameter	Description
Fragment ID	Specifies the fragment identifier assigned by ProSightPC.
MZ Monoisotopic	Shows the monoisotopic mass-to-charge ratio (*m/z*) value of the specific fragment.
MZ Average	Shows the average mass-to-charge ratio (*m/z*) value of the specific fragment.
Mass Monoisotopic	Shows the monoisotopic mass of the specific fragment.
Average Monoisotopic	Shows the monoisotopic average of the specific fragment.
Mass Average	Shows the mass of the most abundant isotope in the protein, peptide, or fragment ion.
Intensity	Shows the abundance of the fragment ion.
Delete Fragment	Indicates whether to delete a fragment.

Inspect the fragment data displayed in Figure 99 to determine which, if any, of the fragments are possibly not true b/y or c/z ions. Spurious fragments appear as shaded entries. They are generally undesirable in an experiment and should be deleted. For more information, see "Deleting Spurious Fragments" on page 238.

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

Noise Reducer Window Parameters

The Noise Reducer window contains the following parameters:

Parameter	Description
Search Tolerance	Specifies the allowable error, measured in Da, between the expected and observed mass differences.
Use monoisotopic masses	Specifies that the mass is monoisotopic, which is the mass of the protein, peptide, or fragment ion, where all carbons are carbon-12.
Use average masses	Specifies that the mass is the mass of the most abundant isotope of the protein, peptide, or fragment ion.
Reduce	Eliminates the 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.

Deleting Spurious Fragments

The Noise Reducer groups fragments that might be spurious.

❖ **To delete spurious fragments**

1. Select the check box next to each of the undesired peaks in each group.

2. Repeat step 1 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 in the new spurious fragment-free MS/MS experiment are pending.

Locating and Selecting PTMs with the PTM Tier Editor

You can use the PTM Tier Editor to view and to change the tier assignment of PTMs.

PTMs

ProSightPC groups all PTMs in a multi-tier structure, enabling you to find and select PTMs quickly. Assigning PTMs to tiers is intended to help you locate and select PTMs quickly and efficiently in your searches. All of the PTMs in the PTM Tier Editor come from the RESID database.

The PTM Tier Editor has two functions:

- To permit you to reassign PTMs in the tier system. ProSightPC automatically assigns many PTMs to Tier 1 and Tier 2. Tier 1 PTMs represent the most common PTMs, and rarer PTMs are assigned to Tier 2.

- To display which PTMs are currently included or excluded, ProSightPC comes with a preset list of included PTMs. Any PTM not listed is excluded. You select which PTMs are available to be included in or excluded from analyses conducted by ProSightPC.

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

> **Note**
>
> 1. You can enter tier assignments greater than 2.
>
> 2. The PTM Tier Editor does not append PTM information to databases. The PTM information must reside in the proteome database before ProSightPC analyzes MS data.

Accessing the PTM Tier Editor

❖ **To access the PTM Tier Editor**

- Choose **Tools > Tier Editor**.

  The PTM Tier Editor dialog box opens, as shown in Figure 100.

Figure 100. PTM Tier Editor dialog box
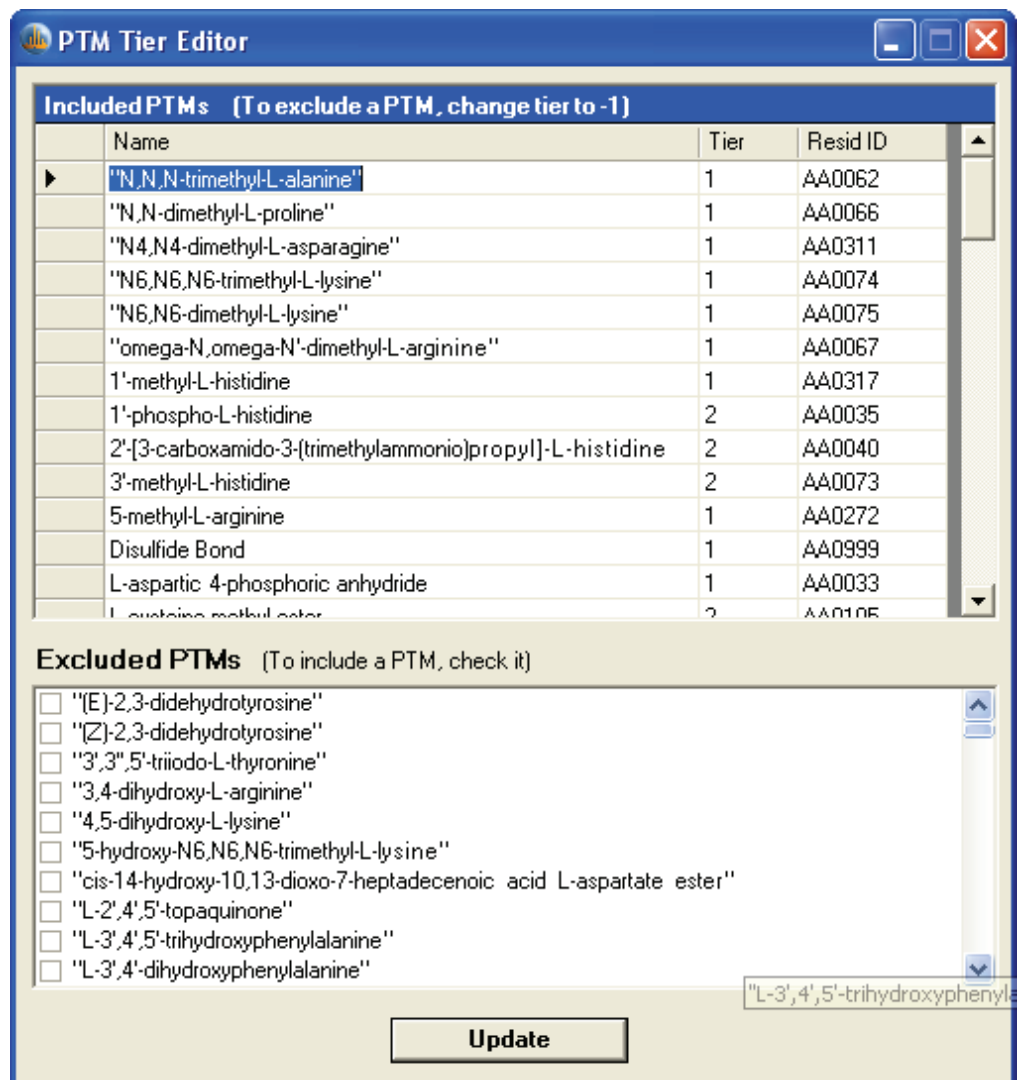
Sort the columns of Included PTMs in ascending or descending order by clicking the header.

PTM Tier Editor Dialog Box Parameters

The PTM Tier Editor dialog box contains the following parameters:

Parameter	Description
Included PTMs: Name	Specifies the RESID name of the included PTM.
Included PTMs: Tier	Specifies the current tier assignment of the included PTMs.
Included PTMs: Resid ID	Specifies the RESID identifier of the included PTMs.

Parameter	Description
Excluded PTMs	Lists all the presently excluded PTMs.
Update	Applies the changes that you made in the dialog box.

Including PTMs

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

❖ **To include a PTM in the database**

1. In the Excluded PTMs section of the PTM Tier Editor dialog box, select the check box to the left of the each PTM that you want to include.

    • To select more than one adjacent row, hold down the SHIFT key and click the first and last rows.

    • To select more than one row where the rows are not adjacent, hold down the CONTROL key and click the appropriate rows.

2. Click **Update** to make the changes.

    The PTMs appear in the Included PTMs list.

    > **Note** Reincluded PTMs are automatically designated as Tier 1.

Excluding PTMs

You can also exclude PTMs from the database.

❖ **To exclude a PTM from the database**

1. In the Included PTMs list of the PTM Tier Editor dialog box, click the number in the Tier column of the row of the PTM that you want to exclude.

2. Change the number in the Tier column to **-1**.

3. Click **Update**.

    The PTMs now appear in the Excluded PTMs list.

Moving PTMs Between Tiers

Use the Tier Editor to manually reassign a PTM to another tier. You can enter tier assignments greater than 2.

The Tier Editor does not append PTM information to databases. The PTM information must be present in the proteome database before ProSightPC analyzes the MS data. If the information for a given PTM is not in the proteome database, that PTM is not available for selection in database searches, even if the PTM is considered included by the tier editor.

❖ **To manually reassign a PTM to a tier in the Included PTMs list**

1. Click the number in the Tier column for the row of the PTM that you want to include.

2. Type a new positive integer in the Tier column.

3. Click **Update**.

> **Note** Once a tier is updated, any excluded PTMs reappear in the Excluded PTMs list.

Viewing Fragments Ions with the Fragment Predictor

You can use the Fragment Predictor to view all possible fragment ions for a specific sequence. You can also use it to add post-translational modifications (PTMs) or arbitrary custom masses to any amino acid in a known protein sequence and see the predicted b, y, c, and Z· fragment ion masses.

Before data collection, you might want to have a list of all theoretical fragment ion masses, particularly for modified protein sequences.

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.

❖ **To view the fragment ions for a sequence**

1. Choose **Tools > Fragment Predictor**.

   The Fragment Predictor window opens, as shown in Figure 101.

Figure 101. Fragment Predictor window
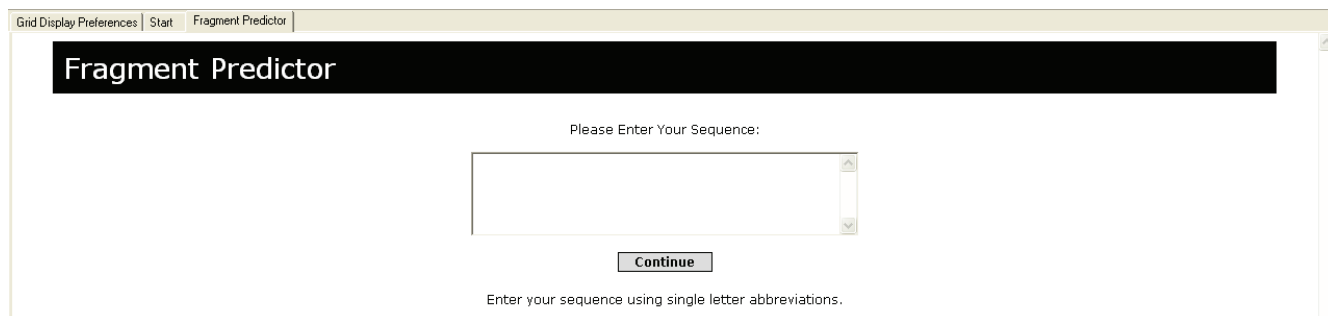
2. Enter a protein sequence in the protein sequence box, as shown in Figure 102.

Figure 102. Entering a sequence in the Fragment Predictor window
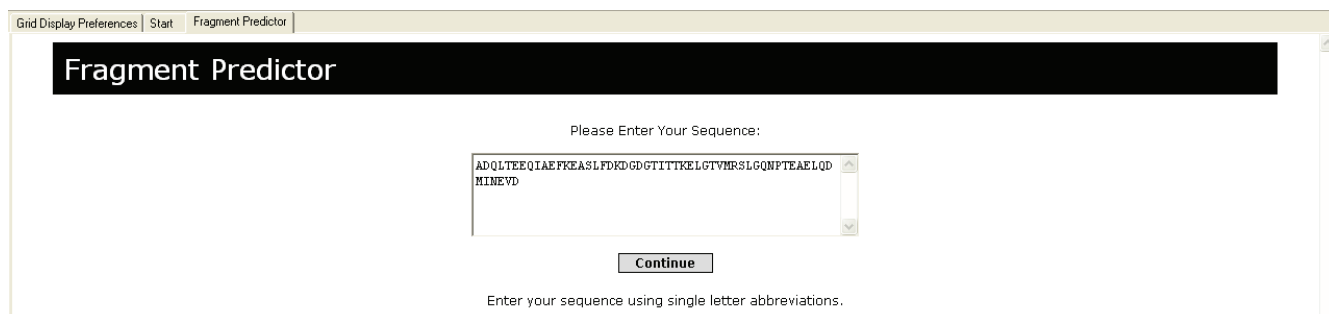
You can use two methods of entering a protein sequence in the Protein Sequence box:

- Manually enter a protein sequence.

- Copy and paste a protein sequence from another source.

> **Note** You can access the sequence from any successful search by clicking RESID or SEQ in the Data Manager. You can also acquire the protein sequence from external sources.

3. Click **Continue** after you enter the sequence in the area provided.

The Fragment Predictor displays a new window showing the protein sequences in an interactive sequence map, as shown in Figure 103.

Figure 103. Interactive sequence map in the Fragment Predictor window
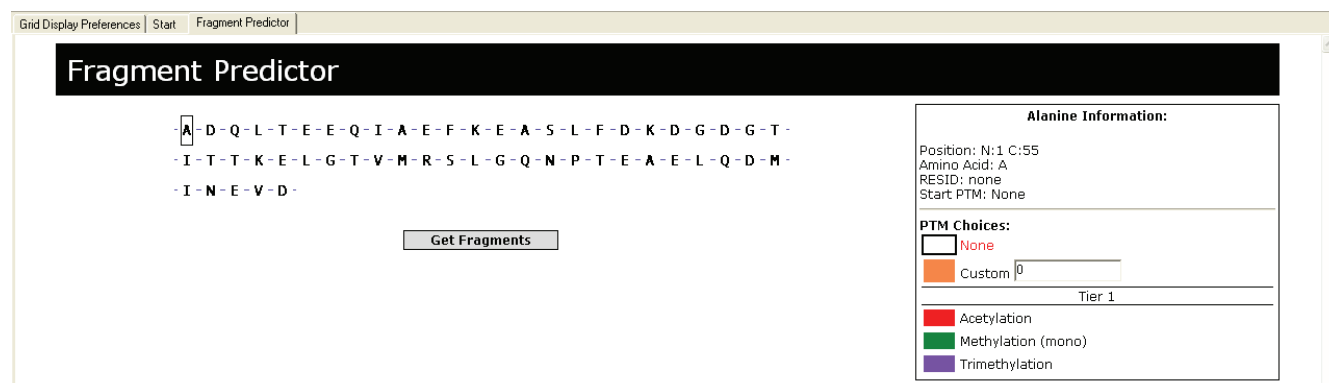
4. Click the sequence to select an amino acid.

A black box around an amino acid indicates that it is selected. For each selected amino acid, common PTMs appear in the Amino Acid information box to the right, as shown in Figure 104.

Figure 104. Amino Acid information box

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

You can customize the PTM tier assignment by using the PTM Tier Editor, described in "Locating and Selecting PTMs with the PTM Tier Editor" on page 239.

5. Click the name of the desired PTM.

The designated amino acid changes to match the color of the PTM selected.

If desired, you can enter a custom mass, in daltons, in the box provided.

6. Click **Get Fragments** when you have selected all the mass changes.

The Fragment Predictor displays the theoretical fragment masses in four columns in the results window, as shown in Figure 105. All theoretical fragment ion masses are arranged in ascending order and are classified as either b, y, c, or Z·.

Figure 105. Results window
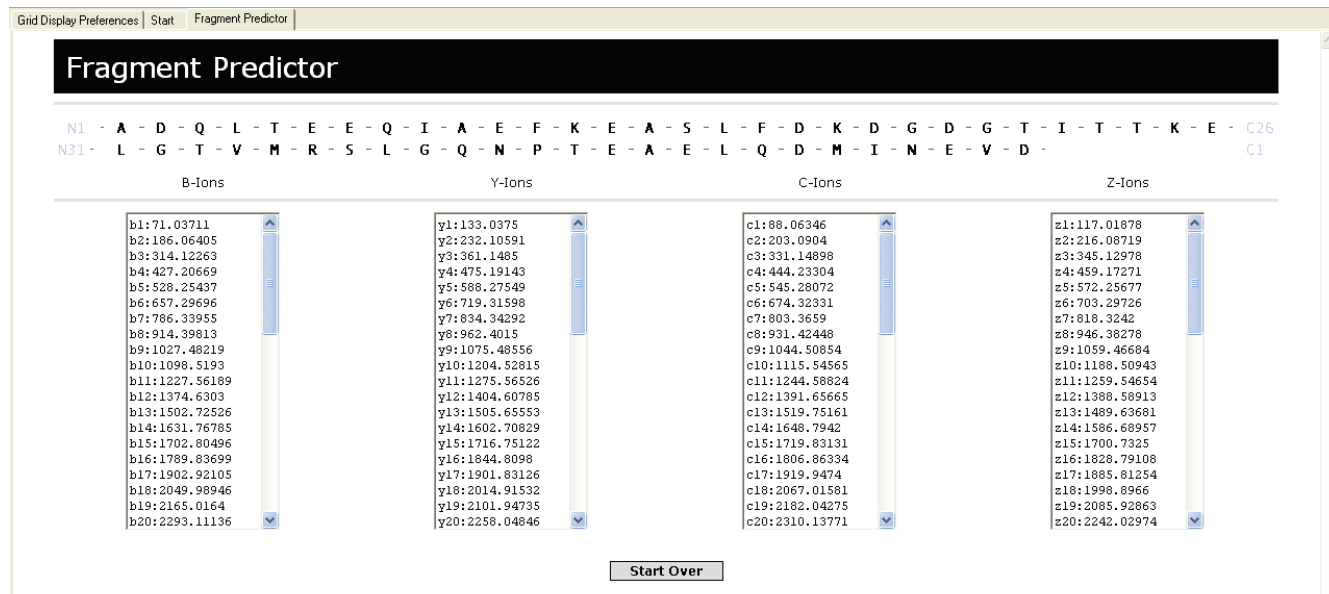
7. (Optional) Click **Start Over** to return to step 2.

Fragment Predictor Window Parameters

The Fragment Predictor window contains the following parameters:

Parameter	Description
Please Enter Your Sequence	Displays the protein sequence to which you want to add post-translational modifications (PTMs) or arbitrary custom masses.
Continue	Displays a new window showing the protein sequences in an interactive sequence map.

Converting Text to ProSightPC Font with the Font Converter

You can use the ProSightPC Font Converter to convert text into the ProSightPC fragment map font used to display N-terminal and C-terminal fragments. You can also use it to generate fragment maps to include in publications and presentations.

❖ **To convert text to ProSightPC fragment map font**

1. Choose **Tools > Font Converter**.

   The Font Converter dialog box opens, as shown in Figure 106.

Figure 106. Font Converter dialog box
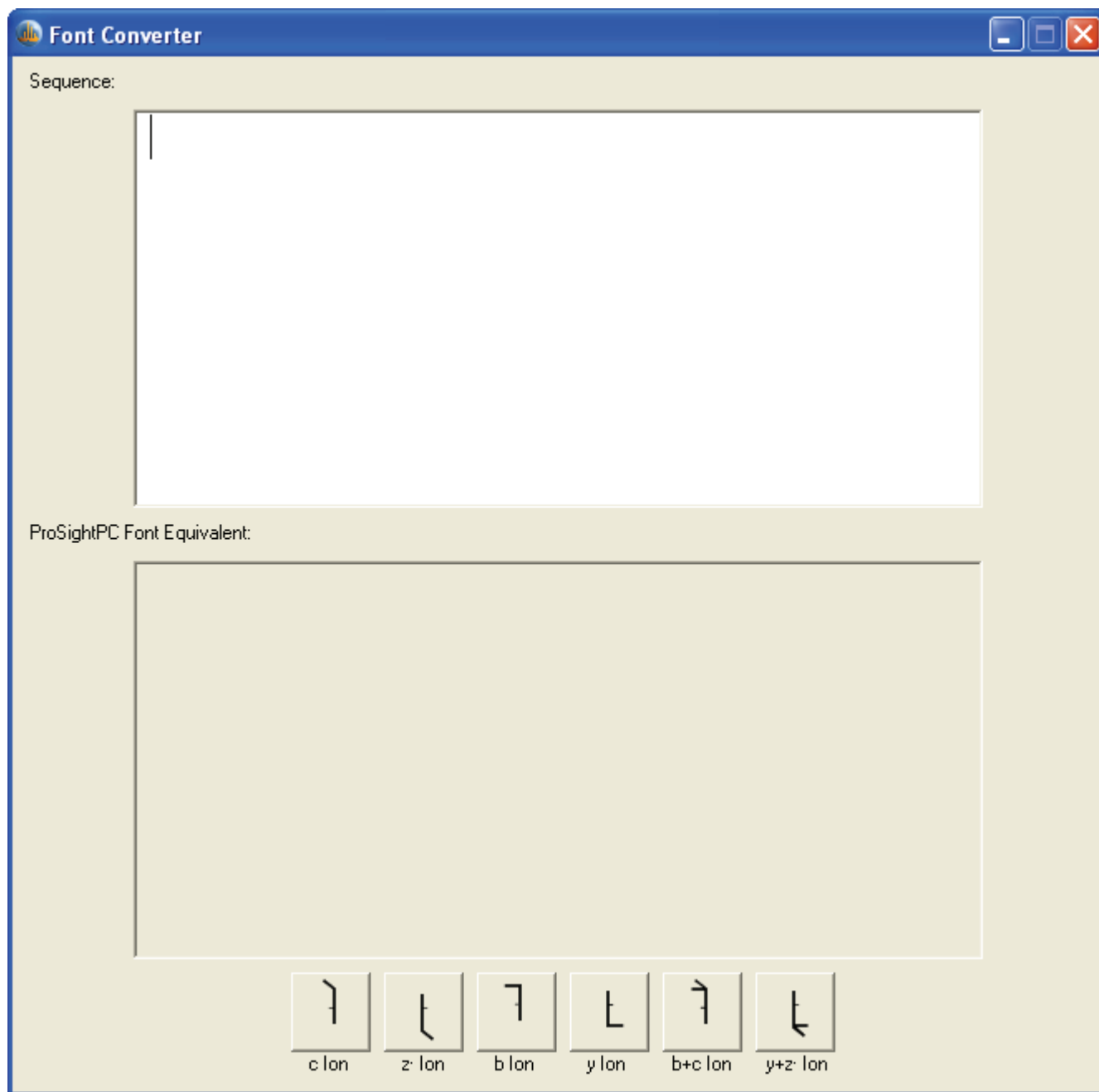
2. In the Sequence box, enter the amino acid sequence to be converted, as shown in Figure 107. You can either type the amino acid sequence in the box or paste it from another source.

Figure 107. Font Converter dialog box with amino acid sequence

Beneath the ProSightPC Font Equivalent box are six buttons that correspond to the N- and C-terminal fragment marks used for b/y and c/z fragment ions.

3. To add fragmentation tick marks, position the cursor between the two amino acid letters and click the appropriate box.

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

5. To transfer the converted font to another application, copy and paste text from the ProSightPC Font Equivalent to the other application.

You can resize the ProSightPC font after a paste operation.

Depending on 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 ProSightPC.

Font Converter Dialog Box Parameters

The Font Converter dialog box contains the following parameters:

Parameter	Description
Sequence	Specifies the amino acid sequence to be converted.
ProSightPC Font Equivalent	Displays the ProSightPC font equivalent of the sequence displayed in the Sequence box.
⌐	Adds the sign for a c ion.
⌊	Adds the sign for a Z· ion.
⌐	Adds the sign for a b ion.
⌊	Adds the sign for a y ion.
⌐	Adds the sign for a bc ion.
⌊	Adds the sign for a yZ· ion.

ProSightPC Reference

This chapter describes the commands on ProSightPC's menus. They are listed in the order in which they appear on the menus.

Contents

- File Menu
- Edit Menu
- View Menu
- Experiment Tools Menu
- Databases Menu
- ProSightHT Menu
- Tools Menu
- Help Menu
- Data Grid Shortcut Menu

File Menu

The File menu contains the following commands.

Command	Description
File > New	Opens a new .puf file.
File > Open	Opens an existing .puf file.
File > Close	Closes a .puf file.
File > Save	Saves a .puf file.
File > Save As	Saves a .puf file under another name.

Command	Description
File > Import Data from Repository	Imports experiments from a repository into ProSightPC so that you can perform operations on them, such as adding or changing searches or using the Sequence Gazer. It activates the Import Data from Repository dialog box, shown in Figure 44 on page 98.
File > Export Data from Repository	Exports the experiments in the ProSightPC Data Grid into a repository. It opens the Export Data to Repository dialog box, shown in Figure 45 on page 100. You must have experiments listed in the Data Grid before you can use the Export Data from Repository command.
File > Import .raw	Imports a targeted .raw file using one of the following analyses to infer mass (AIMs):
	• Post Xtract: Takes a small file generated by the Xtract algorithm within Qual Browser and uses it as the neutral mass data. It opens the Build Experiment from Post Xtract RAW Data dialog box, shown in Figure 46 on page 103. This option is the default.
	• Profile: Uses the THRASH algorithm to process the input file. This algorithm takes raw mass-to-charge (*m/z*) data and finds the neutral mass values. It opens the Build Experiment from Profile RAW Data dialog box, shown in Figure 48 on page 108.
Four most recently opened .puf files	Click the file name to open the file.
File > Exit	Closes ProSightPC.

Edit Menu

The Edit menu contains the following commands.

Command	Description
Edit > Copy	Copies text.
File > Paste	Pastes text.

View Menu

The View menu contains the following commands.

Command	Description
View > Grid Display Preferences	Determines whether the Grid Display Preferences page is displayed in the Tab Controller area. "Show Columns Area" on page 24 describes the function of each of the options on the Grid Display Preferences page.
View > Start	Determines whether the Start page is displayed in the Tab Controller area.
View > Job Queue	Determines whether the job queue pane is visible. See Figure 13 on page 18 for the location of the job queue and "Job Queue" on page 22 for a description of the job queue.
View > Toolbar	Determines whether the toolbar in the ProSightPC window is displayed. See "Toolbar" on page 19 for a description of the icons on the toolbar.
View > Close Data Tab	Closes the experiment page and all tabs related to it (for example, the Sequence Gazer) for the selected experiment.
View > Close All Data Tabs	Closes experiment pages and all pages related to them (for example, the Sequence Gazer) for all the experiments.
View > Close All Data Tabs But Selected	Closes the experiment pages and all pages related to them (for example, the Sequence Gazer) for all experiments except the one selected.

Command	Description
View > Open Data Manager	Opens the Data Manager in the Tab Controller for the selected experiment. If you do not select an experiment, it opens the Data Manager for the first experiment listed in the Data Grid. See "Displaying Data in the Data Manager" on page 197 for detailed information on the Data Manager.
View > Close Data Manager	Closes the Data Manager and all pages related to it (for example, the Sequence Gazer) for the selected experiment. This command is not available unless an experiment page is selected.

Experiment Tools Menu

The Experiment Tools menu is only available when an experiment is open in the Data Manager. It contains the following commands.

Command	Description
Experiment Tools > Add Search	Opens the New Search in Experiment *X* dialog box, shown in Figure 53 on page 127, so that you can add a new search to an experiment.
Experiment Tools > Add Predefined Search	Opens the Append Predefined Searches to Experiment *X* dialog box, shown in Figure 66 on page 169, so you can add a new predefined search to an experiment.
Experiment Tools > Edit Masses	Opens a new page in the Tab Controller, shown in Figure 81 on page 201, so you can review and edit mass values in an experiment.

Command	Description
Experiment Tools > Reduce Noise	Runs the Noise Reducer on the selected experiment. The Noise Reducer identifies and minimizes the effects of chemical noise in an MS/MS experiment. It activates the Noise Reducer window, shown in Figure 98 on page 236. The Data Manager must be active for the Reduce Noise command to be available.
Experiment Tools > Edit Comment	Opens a box in the Data Manager so that you can type a comment about an experiment or a search.

Databases Menu

The Databases menu contains the following commands.

Command	Description
Databases > Database Manager	Opens the Database Manager window, shown in Figure 84 on page 209. The Database Manager handles all proteome warehouse management and manipulation functions.
Databases > Create a Custom Database	Opens the Welcome to the New Database Wizard page of the Create New Database wizard, shown in Figure 88 on page 216, so you can manually create a custom database.
Databases > Download ProSightPC Databases	Downloads databases from the ProSightPC FTP Web site.
Databases > Link to UniProt Database	Connects you to the UniProt database, which is an international repository of organisms. It contains all the proteins and genes that are known for a specific organism.

ProSightHT Menu

The ProSightHT menu contains the following commands.

Command	Description
ProSightHT > HighThroughput Wizard	Opens the Process a dataset page of the High Throughput wizard, shown in Figure 26 on page 59, so you can import data from a .raw or .puf file, specify a repository in which to store the results of the search, and create a search tree.
ProSightHT > Edit/Add Repository	Opens the Edit/Add Repositories dialog box, shown in Figure 43 on page 96, so you can edit an existing repository or add a new one.
ProSightHT > Edit/Add Search Tree	Opens the Running Highthroughput Logic page of the High Throughput wizard, shown in Figure 27 on page 65. For information on the options on this page, see "Selecting or Creating a Repository" on page 65.
ProSightHT > Repository Report	Generates a repository report that lists all the experiments that a repository contains. Figure 37 on page 81 shows an example of this report. This command opens the Repository Report dialog box shown in Figure 36 on page 80.

Tools Menu

The Tools menu contains the following commands.

Command	Description
Tools > Experiment Adder	Imports experiment data into ProSightPC and displays it in the Data Grid. It opens the Experiment Adder dialog box, shown in Figure 50 on page 112.
Tools > Manage Predefined Searches	Assigns a name to a set of parameters that you can then add to any experiment. You can also use this command to set up defaults for frequently run searches. This command opens the Predefined Search Manager dialog box, shown in Figure 63 on page 163.

Command	Description
Tools > PTM Tier Editor	Opens the PTM Tier Editor dialog box, as shown in Figure 100 on page 240, so you can manually reassign a PTM to another tier.
Tools > Font Converter	Converts text into the ProSightPC fragment map font used to display N-terminal and C-terminal fragments. You can use the Font Converter to generate fragment maps for inclusion in publications and presentations. This command opens the Font Converter dialog box, shown in Figure 106 on page 246.
Tools > Fragment Predictor	Adds post-translational modifications (PTMs) or arbitrary custom masses to any amino acid in a known protein sequence and displays the predicted b, y, c, and Z$^\cdot$ fragment ion masses. It opens the Fragment Predictor window, shown in Figure 101 on page 242.
Tools > Experiment Manager	Opens the Experiment Manager dialog box, shown in Figure 51 on page 119, so you can manipulate experiments as objects, copy individual experiments between .puf files, or save them in their own .puf file.
Tools > Batch Run	Processes several predefined searches automatically. You can queue and run a large number of searches over any number of experiments in a single action.

Command	Description
Tools > Reports	Generates the following types of reports: • Status report: Gives a summary of every search in the open .puf file, including search type and best score. • Printable report: Contains all of the information related to one search, formatted for easy printing. • Best Hit report: Displays the match with the best score for each search that was run for each experiment in the Data Grid. • Repository report: Lists all the experiments that a repository contains.
Tools > Options	Opens the Options dialog box, which you can use to set default values for most of the interface elements in ProSightPC. See "Setting Default Options" on page 31 for instructions on setting default values.

Help Menu

The Help menu contains the following commands.

Command	Description
Help > Help	Opens the Help for ProSightPC.
Help > About	Opens a dialog box that displays the release version, the release date, and the trademark information.
Help > Manage License	Opens the License Information dialog box so that you can request a a new license activation code.

Data Grid Shortcut Menu

The Data Grid shortcut menu appears when you right-click an experiment in the Data Grid. It includes the following commands.

Command	Description
Refresh Grid	Redisplays the contents of the Data Grid.
Mark	Marks an experiment by placing the ProSightPC symbol to the left of the experiment and an asterisk (*) in the Marked column. This mark can differentiate a particular experiment.
Add New Search	Opens the New Search in Experiment *X* dialog box, so you can add a new search to the experiment. For information on how to select options in this dialog box, see "Adding a Search" on page 126.
Append Predefined Search	Opens the Append Predefined Searches to Experiment *X* dialog box, so you can add a new predefined search to an experiment. For information on how to select options in this dialog box, see "Adding a Predefined Search to an Experiment" on page 169.
Edit Search	Opens the Edit Search in Experiment *X* dialog box for that type of search (this dialog box is the same as the New Search in Experiment *X* dialog box for that search type).
Edit Mass List	Opens a new page in the Tab Controller, showing the Precursor Mass List and the Fragment Mass List.
Run Search *X*	Reruns a search after you have deleted the previous results of the search.
Remove Results	Removes the results of a specific search from an experiment in the Data Grid.
Remove Search *X*	Removes the specified search from the experiment.
Export Experiment to Repository	Replaces the specified experiment in the repository from which it was taken.
Remove Experiment *x*	Removes the specified experiment from the Data Grid.

Index

G

H