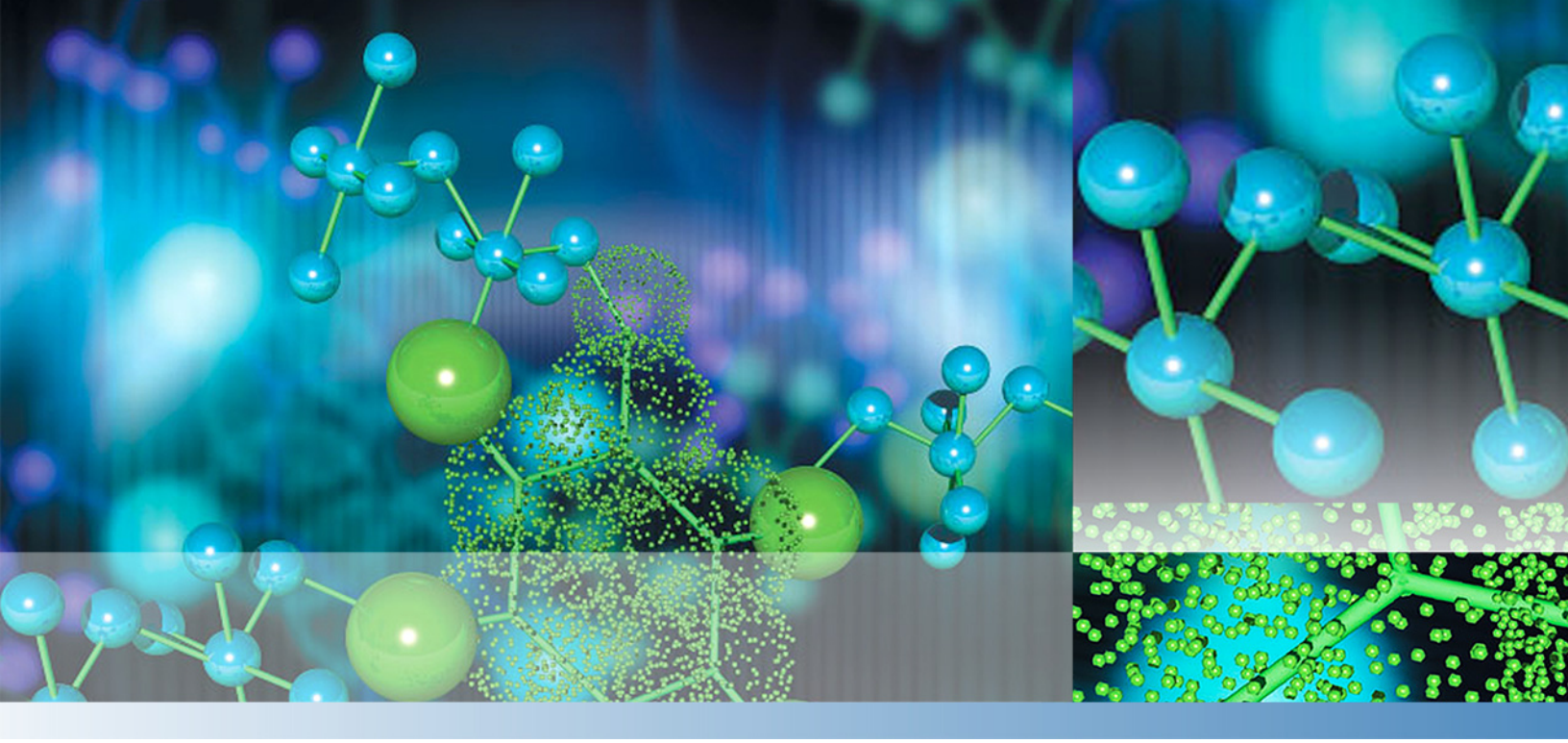


Thermo

ProSightPC

User Guide

Software Version 4.0

XCALI-97801 Revision A September 2016

DOCUMENTATION SURVEY

thermoscientific

ProSightPC, PTM Warehouse, and Sequence Gazer are registered trademarks of Proteinaceous, Inc. in the United States.

Q Exactive, Orbitrap Elite, and Proteome Discoverer are trademarks, and LTQ, LTQ FT, Orbitrap, Tribrid, and Thermo Scientific are registered trademarks of Thermo Fisher Scientific Inc. in the United States.

The following are registered trademarks in the United States and possibly other countries: UniProt is a registered trademark of European Molecular Biology Laboratory Incorporated Association. SQLite is a registered trademark of Hipp, Wyrick & Company, Inc. iTRAQ is a registered trademark of AB SCIEX.

Microsoft and Excel are registered trademarks of Microsoft Corporation in the United States and other countries.

RESID is a registered service mark of John S. Garavelli (individual) in the United States.

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

TMT is a registered trademark of Proteome Sciences plc in the United Kingdom.

All other trademarks are the property of Thermo Fisher Scientific Inc. and its subsidiaries.

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

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

Release history: Revision A September 2016

Software version: ProSightPC 4.0

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

Contents

Preface

This guide describes how to use the Thermo ProSightPC™ 4.0 application to identify and characterize proteins.

Contents

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

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

Accessing Documentation

The ProSightPC application includes complete documentation. For system requirements, refer to the release notes on the flash drive.

❖ **To view the product manuals**

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

- For a Thermo Scientific™ application, choose **Start > All Programs >** …. Or, from the application, choose **Help > Manuals** and so on.
- For a Thermo Scientific mass spectrometer, choose **Start > All Programs >** ….
- For an LC instrument controlled by a Thermo Scientific application, choose **Start > All Programs > Thermo Instruments > Manuals** and so on.

❖ To view user documentation from the Thermo Fisher Scientific website

1. Go to thermofisher.com.
2. Point to **Services & Support** and click **Manuals** on the left.
3. In the Refine Your Search box, search by the product name.
4. From the results list, click the title to open the document in your web browser, save it, or print it.

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

❖ To view ProSightPC Help

From the application window, choose **Help > ProSightPC Help** or press the F1 key.

Installing ProSightPC

Follow these instructions to install the ProSightPC software and to use the accompanying example data files.

❖ To install the ProSightPC software

1. Open the PSPC 4.0 folder on the distributed flash drive that contains a setup.exe file.
2. In the setup folder, double-click **setup.exe**.

   The InstallShield Wizard opens.
3. On the first page of the wizard, click **Install** to install the prerequisites.
4. On the Welcome to the InstallShield Wizard for ProSightPC page, click **Next**.
5. On the License Agreement page, select **I accept the terms of the license agreement**, and click **Next**.

   The Thermo License Activation dialog box appears.
6. Enter your activation code in the Activation Code boxes and any contact information in the Contact Information boxes, and click **Activate**.

   Obtain the activation code from the identification card included in the box that contains the flash drive.
7. On the Choose Destination Location page, click **Next**.
8. On the Select Features page, click **Next**.
9. On the Ready to Install the Program page, click **Install**.
10. In the Installation Qualification dialog box, click **Yes** or **No** to view the Installation Qualification report.

11. On the Setup Status page, click **Finish**.

❖ **To use the example data files**

Download the example data files from the distributed flash drive.

If you did not obtain this flash drive, you can download these files from the following site:

http://proteinaceous.net/prosightpc40-demonstration-data/

Special Notices

Make sure you follow the special notices presented in this guide. Special notices appear in boxes; those concerning safety or possible system damage also have corresponding caution symbols.

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

Note Highlights information of general interest.

Tip Highlights helpful information that can make a task easier.

Contacting Us

There are several ways to contact Thermo Fisher Scientific for the information you need. You can use your smartphone to scan a QR code, which opens your email application or browser.

Contact us	Customer Service and Sales	Technical Support
	(U.S.) 1 (800) 532-4752	(U.S.) 1 (800) 532-4752
	(U.S.) 1 (561) 688-8731	(U.S.) 1 (561) 688-8736
	us.customer-support.analyze@thermofisher.com	us.techsupport.analyze@thermofisher.com

Contact us	Customer Service and Sales	Technical Support

❖ **To find global contact information or customize your request**

1. Go to thermofisher.com.
2. Click **Contact Us** and then select the type of support you need.
3. At the prompt, type the product name.
4. Use the phone number or complete the online form.

❖ **To find product support, knowledge bases, and resources**

Go to thermofisher.com/us/en/home/technical-resources.

❖ **To find product information**

Go to thermofisher.com/us/en/home/brands/thermo-scientific.

Note To provide feedback for this document:

- Send an email message to Technical Publications (techpubs-lcms@thermofisher.com).
- Complete a survey at surveymonkey.com/s/PQM6P62.

1

Introduction to the ProSightPC Application

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

To install the ProSightPC software, see "Installing ProSightPC" on page x.

Contents

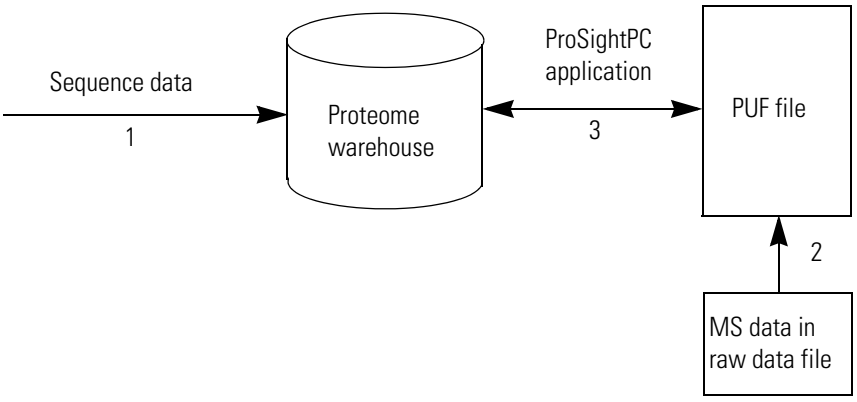
Features

The ProSightPC application is the only proteomics software suite that effectively supports high-mass-accuracy MS/MS experiments performed on Orbitrap™- and LTQ™ FT™-based mass spectrometers, including the Q Exactive™ and Fusion™ Tribrid™. It operates on mass data from MS/MS experiments (or any MS^n experiment) on intact and digested proteins. For accurate MS/MS data, it produces highly confident identifications and also automatically detects and annotates post-translational modifications in database files that are in the UniProtKB XML and flat file formats. The application can identify more than one peptide or protein in a spectrum and includes a biomarker search mode to determine if a protein has been truncated.

The ProSightPC application complements the Proteome Discoverer™ application and is best used with it to find new or unexpected modifications. To identify these unexpected modifications, you can use the ProSightPC delta-m (Δm) mode with its ability to search databases in UniProtKB XML and flat files. Although you can use either tool to search bottom-up and top-down experiments, the ProSightPC application is uniquely suited to top-down experiments, and the Proteome Discoverer application is better suited to bottom-up experiments.

As Figure 1 shows, the ProSightPC application first creates a new proteome database called a PTM Warehouse™ or a proteome warehouse. 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, the ProSightPC application calculates all possible combinations of known modifications and applies them, along with single nucleotide polymorphisms (SNPs) and sequence variants, to each protein sequence in a process called shotgun annotation (see "Shotgun Annotation" on page 13). 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). The ProSightPC application then searches the appropriate proteome databases for these mass values and compares them (3).

Figure 1. The ProSightPC protein and peptide identification process



The basic unit of analysis in the ProSightPC application 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 the ProSightPC application. 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. The application queries each experiment against the ProSightPC proteome warehouse 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 the ProSightPC application 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

The ProSightPC application 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.

The ProSightPC application 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, the ProSightPC application includes several tools for importing LC/MS/MS and tandem MS data from Thermo Fisher Scientific raw data files, identifying and removing chemical noise peaks and performing other utility functions. It can handle and store data in raw data format, in ProSightPC upload format (PUF), or in a repository.

Proteome Warehouse

The ProSightPC application creates proteome warehouses, which are collections of databases that it uses to identify and characterize protein data. It contains all the proteoforms 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). A proteome warehouse contains both monoisotopic and average mass information and is organized to facilitate both protein identification and characterization. Each organism in a proteome warehouse receives its own database. You can create custom databases from UniProtKB XML and flat or FASTA-formatted text files.

The databases in a warehouse are SQLite relational databases, which you can view by using other third-party applications. In addition, you can move them between computers. The ProSightPC application 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. The ProSightPC application 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 See "Top-Down Proteomics" on page 12. and Middle-Down/Bottom-Up Proteomics.

"Using Proteome Databases" on page 215 tells you how to create and manage databases in the ProSightPC application.

Search Types

The ProSightPC application supports the types of searches shown in Table 1. The topic referenced for each search contains recommendations for running the search.

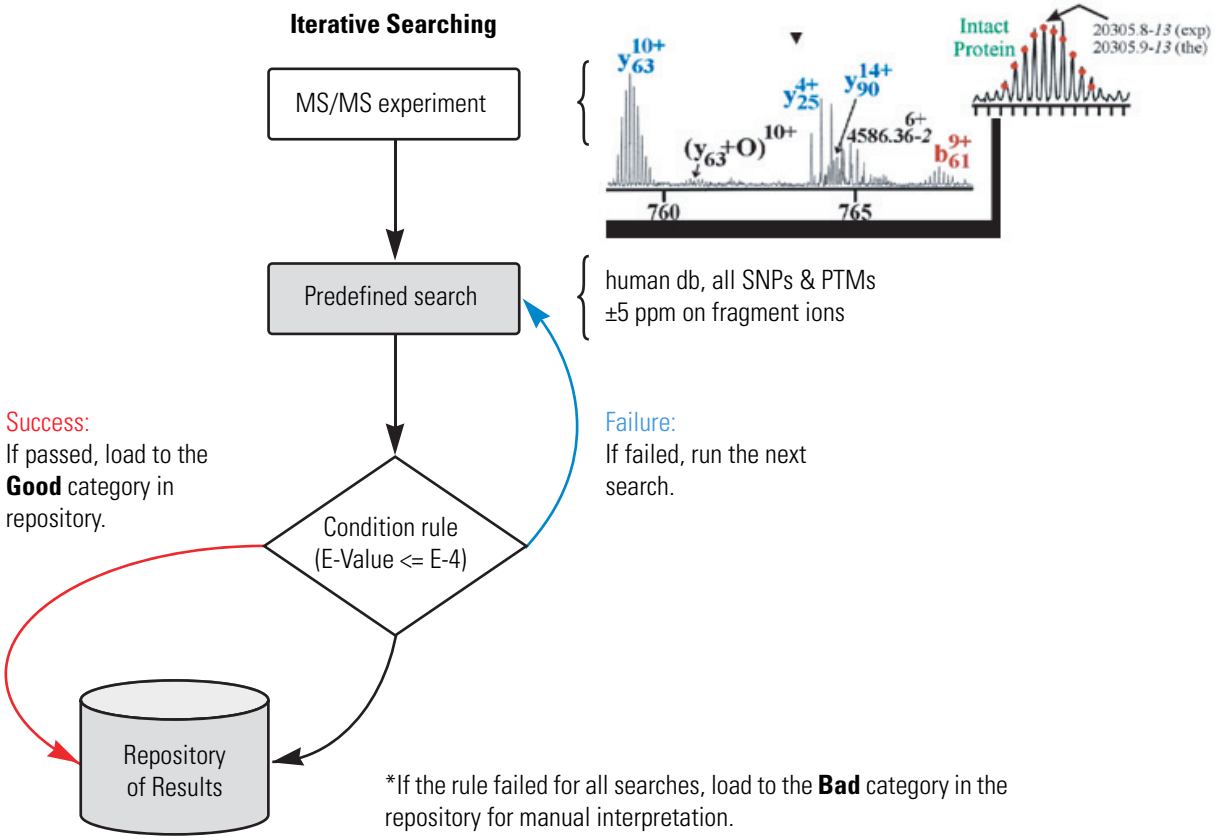
Table 1. Types of searches available in the ProSightPC application

Type of search	Location of information
Absolute mass searches	"Searching for Absolute Mass" on page 116
Biomarker searches	"Searching for Biomarkers" on page 128
Sequence tag searches	"Searching for Sequence Tags" on page 139
Single-protein searches	"Searching for Single Proteins" on page 189
Gene-restricted absolute mass searches	"Searching for Gene-Restricted Absolute Masses" on page 153
Gene-restricted biomarker searches	"Searching for Gene-Restricted Biomarkers" on page 159
MS^n hybrid searches	"Performing MS^n Hybrid Searches" on page 165

Iterative Searching

You can build an automatic, iterative, score-based search tree in the ProSightPC application. 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 proteoforms 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 2 illustrates this methodology.

Figure 2. Iterative searching in the ProSightPC application



For a detailed explanation of the ProSightPC iterative search tree, see "Creating a Search Tree" on page 37.

The ProSightPC application supports two levels of searching, but if you are an advanced user, you 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 (PSCW) files and repository (PSPH) files, enabling you to create your own proteome databases and repositories. The PSCW and PSPH files are in a custom format that holds databases, patches, and repositories. For details on the functionality of the Database Manager, see "Using Proteome Databases" on page 215.

Data Manager

The Data Manager is part of the ProSightPC 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 "Displaying Data in the Data Manager" on page 207.

Sequence Gazer

The Sequence Gazer™ in the ProSightPC 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. "Searching for Single Proteins" on page 189 explains how to perform single-protein searches by using the Sequence Gazer.

ProSight Lite

ProSight Lite is a separate piece of free software that you can use to view a single proteoform result and produce publication-ready fragment maps. For information and a link to install this software, see "Accessing ProSight Lite" on page 192.

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 "Working with Experiments" on page 95.

PTM Tier Editor

The ProSightPC application 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 233 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· 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. For information on the functionality of the Fragment Predictor, see "Viewing Fragments Ions with the Fragment Predictor" on page 236.

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 239.

LC/MS/MS Workflow

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

1. (Optional) Set the default values for the ProSightPC options. See "Setting Default Options" on page 21.
2. Load the proteome warehouse. The ProSightPC application 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. For information on this procedure, see "Importing or Creating a Proteome Database" on page 24.
3. Customize the chemical modifications that you use to conduct a search. See "Editing Modifications" on page 25.
4. Specify the type of search to perform and set the parameters for that search. See Searching Databases, specifically "Creating a Predefined Search" on page 105.
5. Use the High Throughput Wizard to do the following:
   a. Import a raw data or PUF file into the ProSightPC application.
   b. If you import a raw data file, select the Xtract or THRASH algorithm to interpret resolved isotopic distributions and output neutral mass values in a PUF file.
   c. Define an iterative search tree.
   d. Create a repository in which to store the search results.
   e. Perform the search.

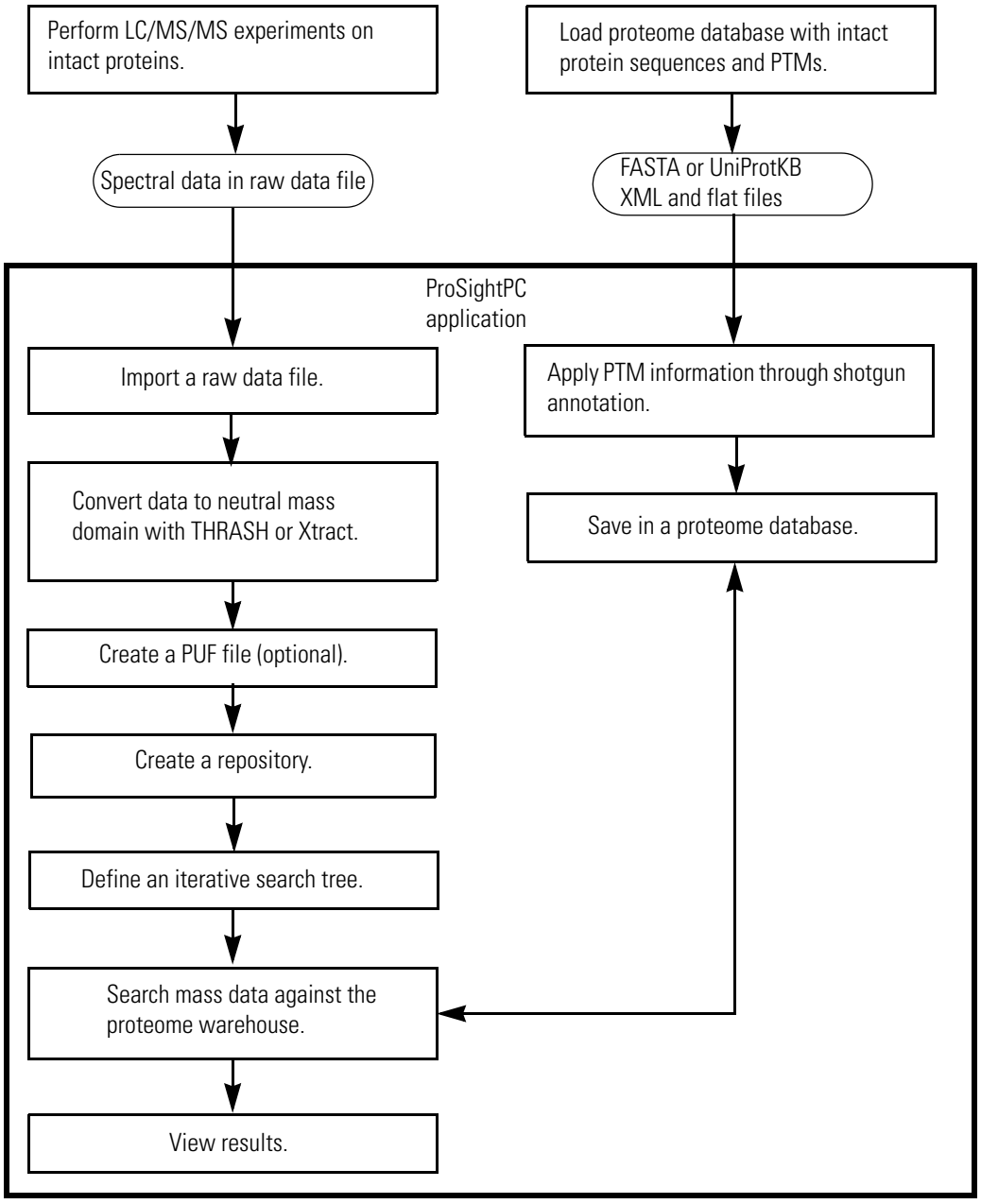
      ProSightPC searches for neutral mass data against the proteome warehouse. It identifies and characterizes the observed proteins.

   For information on using the High Throughput Wizard, see "Using the High Throughput Wizard to Process LC/MS/MS Data Files" on page 26.

6. View the results in the user interface, view the reports, or generate a repository report. For information on viewing the results of the search, see "Viewing Search Results" on page 167.

Figure 3 illustrates this flow.

Figure 3. ProSightPC wizard workflow



Inputs and Outputs

The ProSightPC application works with the following formats.

Inputs

The ProSightPC application 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 (PSCW) 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 the ProSightPC application, or by going to http://proteinaceous.net/database-warehouse. This warehouse includes top-down and bottom-up databases for all major model organisms. These databases are constructed and distributed quarterly by Proteinaceous, Inc. You can also download the UniProtKB XML file to create your own PSCW file.

- Raw data 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, the ProSightPC application produces the following files:

- PSCW file: Can contain any proteome databases and repositories that you have created and want to export to others.

- PUF file: Contains experiments and searches.

Fragmentation Methods

The ProSightPC application supports the following fragmentation types:

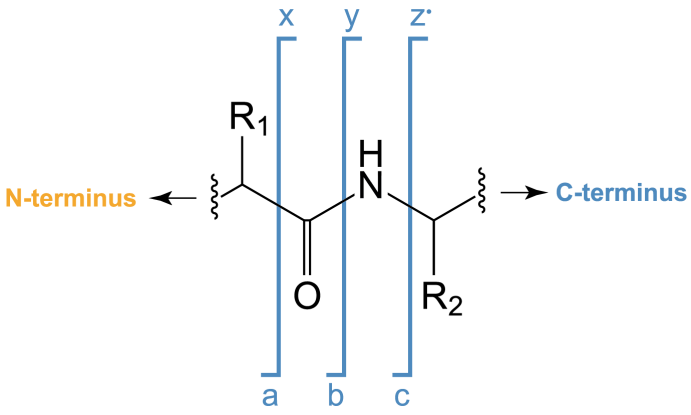
- CID: With the collision-induced dissociation (CID) method of fragmentation, molecular ions are accelerated to high kinetic energy in the vacuum of a mass spectrometer and then allowed to collide with neutral gas molecules such as helium, nitrogen, or argon. The collision breaks the bonds and fragments 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.
- UVPD: With the ultraviolet photodissociation (UVPD) method of fragmentation, the proteins are activated by using ultraviolet photons.
- EThcD: Uses the electron transfer higher-energy collision dissociation (EThcD) method of fragmentation but produces additional b and y ions.

Ion Types

The ProSightPC application supports both c/z· and b/y ion types, which are shown in Figure 4. It also supports a/x ion types with UVPD support.

Figure 4. a/x, b/y, and c/z· ion types



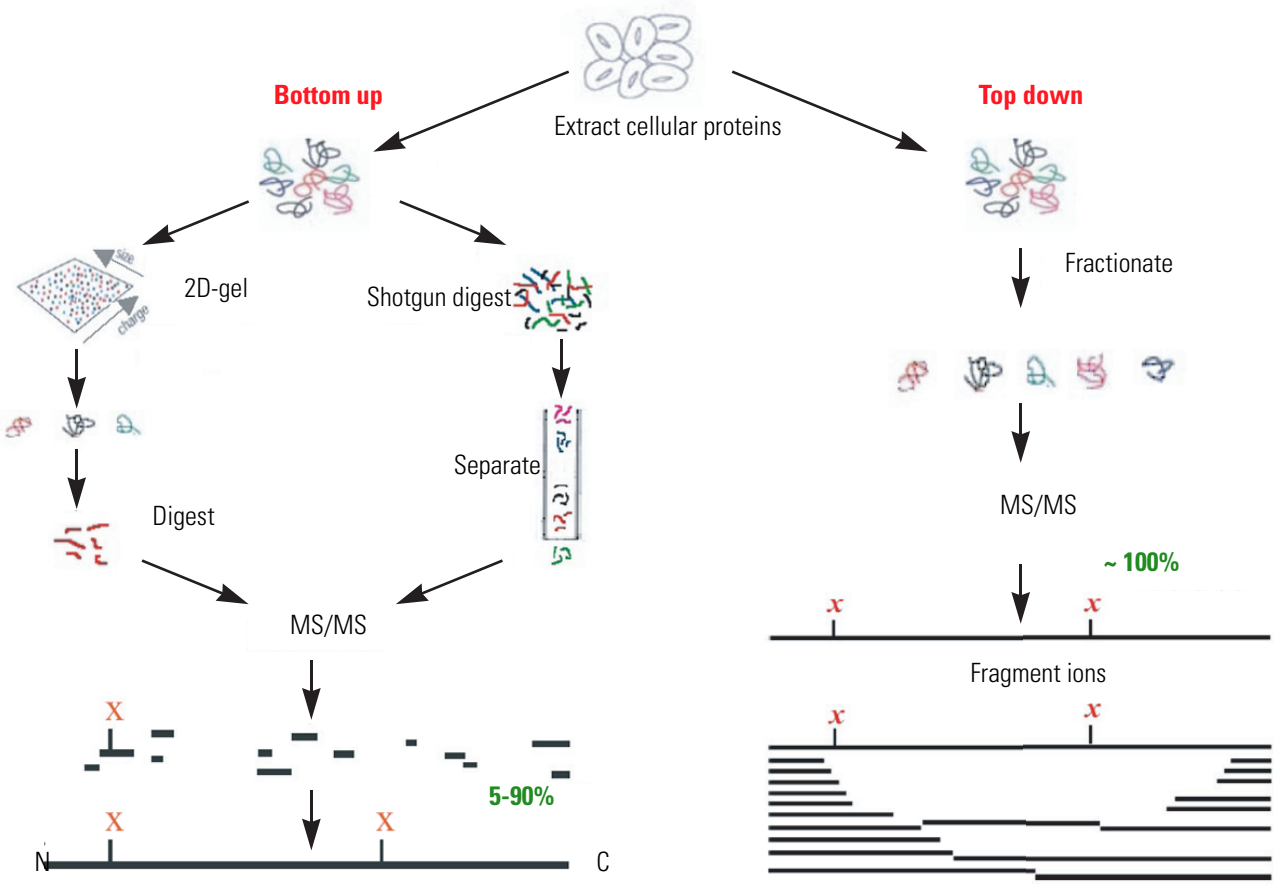
Introduction to Proteomics

The ProSightPC application 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 samples 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 Lys-C. For complex samples containing many different proteins, the proteins can be digested into peptides and then separated by several orthogonal methods before electrospray mass spectrometry (ESI-MS). The left side of Figure 5 illustrates these two methods. In either case, these peptides are then introduced to the mass analyzer.

Figure 5. 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. The right side of Figure 5 illustrates this methodology.

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 a method for rapid and accurate protein definition.

Top-down proteomics is an 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 the ProSightPC application 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 is avoided by comparison with neutral masses. Once you collect MS and MS/MS spectra, you sum the relevant scans and then run through 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, the ProSightPC application provides the search modes described in "Searching Databases" on page 103. 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 The ProSightPC application includes the RESID℠ database (http://pir.georgetown.edu/resid/). The system identifies all post-translational modifications as a truncated form of their RESID identification number. The ProSightPC application 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

The ProSightPC application 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 proteoform.

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 proteoforms of a given protein, resulting in 2^n proteoforms, where *n* is the number of sites of 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 proteoforms for highly modified proteins. To address this issue, the ProSightPC application first determines if the protein is going to need more than 1000 records to fully describe it. If it does, the application prioritizes the proteoforms, storing only those proteoforms most likely to lead to protein identification. These proteoforms are then shotgun-annotated. Shotgun annotation is therefore the process of generating potentially observable proteoforms from the information known about a given protein.

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

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

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

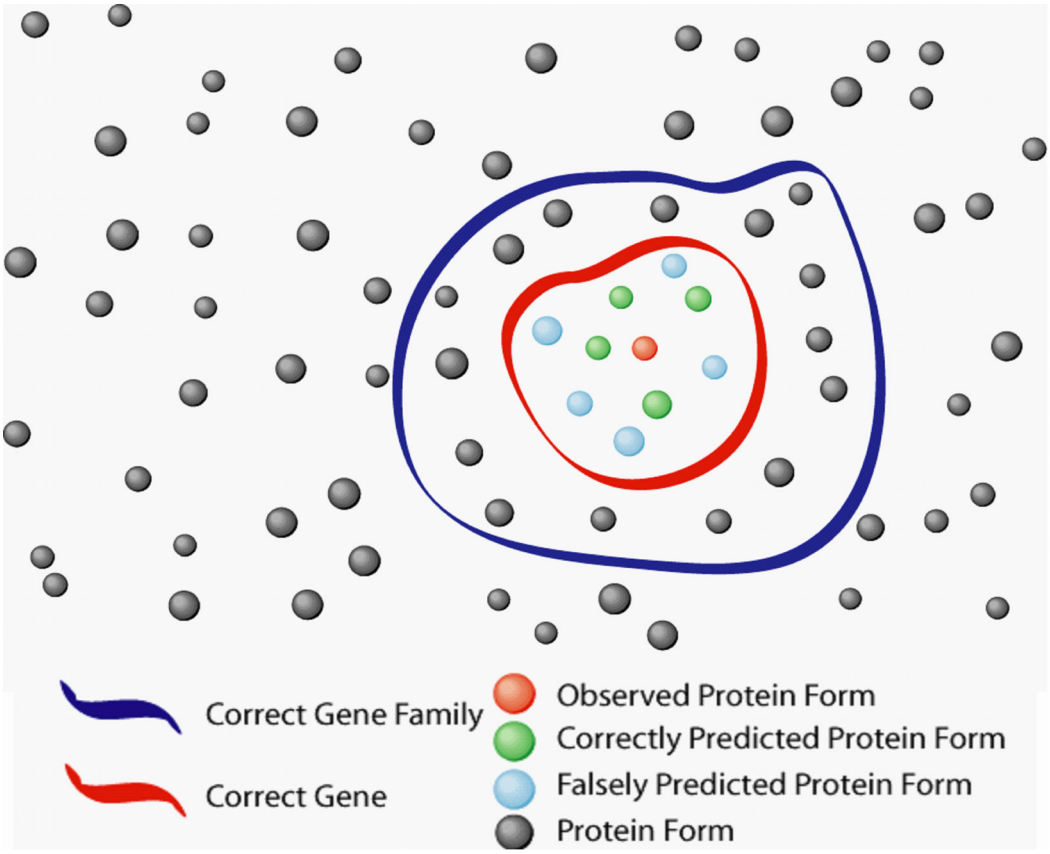
If one of the multiphosphorylated proteoforms occurs in nature and is observed in an MS/MS experiment, the ProSightPC application can readily identify it.

Searching Databases

You then search the resulting database. All searches require matching observed masses to the masses stored in your database (theoretical masses). Matches are not exact but are within a tolerance. The matches are then scored using various fragment-based scoring functions to determine the best match (see "Scores Box" on page 195 for detailed information about scoring methods).

Figure 6 illustrates the database searching component of shotgun annotation. Every ball is a proteoform that matches within a mass window.

Figure 6. Shotgun annotation search strategy



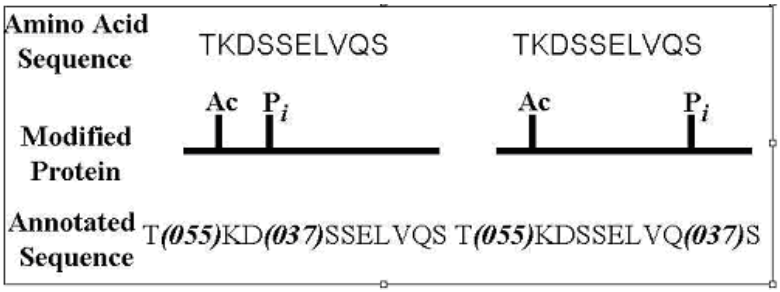
Most potentially matching proteoforms have negligible scores that you can ignore. They are represented by the gray balls outside any circle. The balls within the blue circle share many fragments among the proteins encoded by a gene family, but each identification is at best partial. The balls in the concentric red circle represent better identifications, because they match fragments that are unique to proteins encoded by a particular gene.

The blue balls are proteoforms resulting from combinations of modifications that might be abiological. The shotgun annotation algorithm created them, but they do not exist in living organisms. Usually you do not inherently know which proteoforms in your database exist or do not exist in real life. The green balls are proteoforms resulting from combinations of modifications that do, in fact, exist in living organisms—not the observed proteoform, but real nonetheless. Finally, the red balls represent the proteoforms that you actually observed in the mass spectrometer.

Because you generate all potential proteoforms in shotgun annotation, a large number are not going to exist in nature. Using shotgun annotation, you can detect proteoforms that you previously were not aware of or could not observe.

Figure 7 shows an example of a shotgun-annotated sequence.

Figure 7. Shotgun-annotated sequence

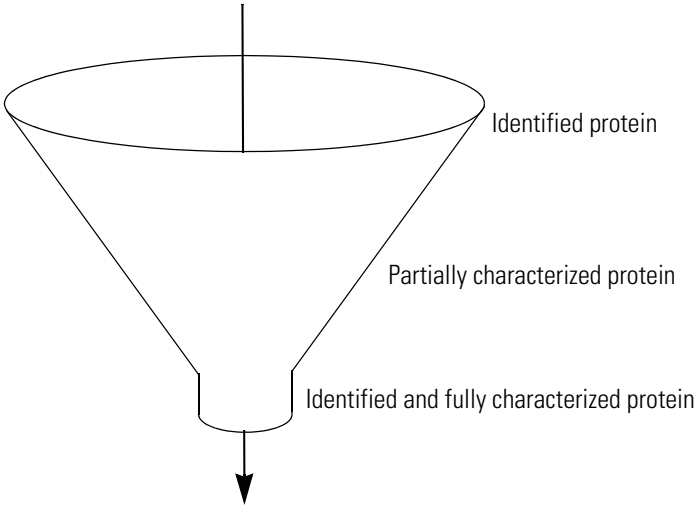


Search Modes and the Top-Down Funnel

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

Figure 8 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 that identify the unknown protein and fully characterize any PTMs present. Additional combinations allow for identifying and partially characterizing the protein. In some cases, only identifying the protein is possible.

Figure 8. Results of full characterization from initial data

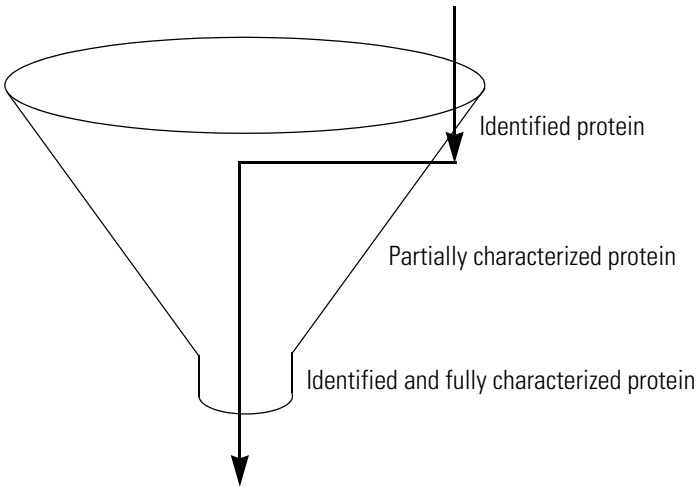


Note Partial characterization occurs in cases where it is possible to determine which PTM must be present on the protein, but the fragmentation data is not sufficient to determine exactly which amino acid one or more of the PTMs must be on. You can narrow the list of possibly modified amino acids to 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 proteoform observed into the proteome database and the MS spectra contains sufficient fragmentation information to uniquely identify this proteoform, 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 8.

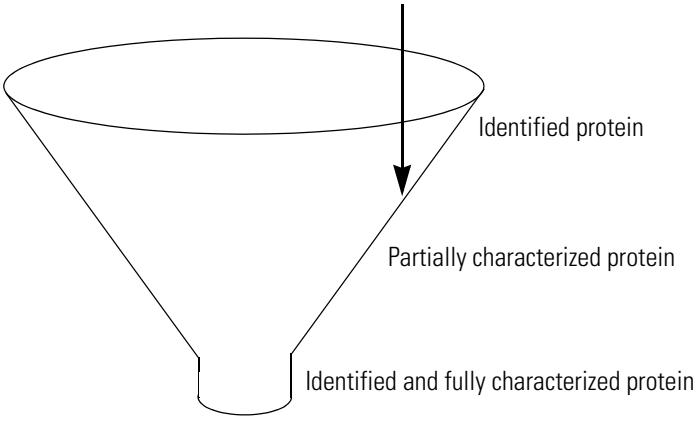
Figure 9 illustrates another common situation when the initial search only identifies and perhaps partially characterizes the unknown protein. In this case, conducting a second search fully characterizes the protein. Often the second search is either a biomarker or a single protein mode search, though some search strategies conduct a gene-restricted search. For descriptions of these search modes, see "Searching Databases" on page 103.

Figure 9. Multiple searches used for the identification and characterization of an unknown protein



In Figure 10, the fragmentation data is insufficient to distinguish between two or more possible proteoforms. In this case, full identification or partial characterization is the best possible result. When this occurs, rerun the MS/MS experiment to obtain better fragmentation data.

Figure 10. Results of identification, partial characterization, or both



2

Getting Started

This chapter explains how to set up and conduct the search of the databases that the ProSightPC application performs to identify proteins and peptides.

Contents

For detailed information on the ProSightPC features noted in this chapter, see ProSightPC Application Window.

Starting the ProSightPC Application

Open the ProSightPC application by choosing a Start menu command or clicking a desktop icon.

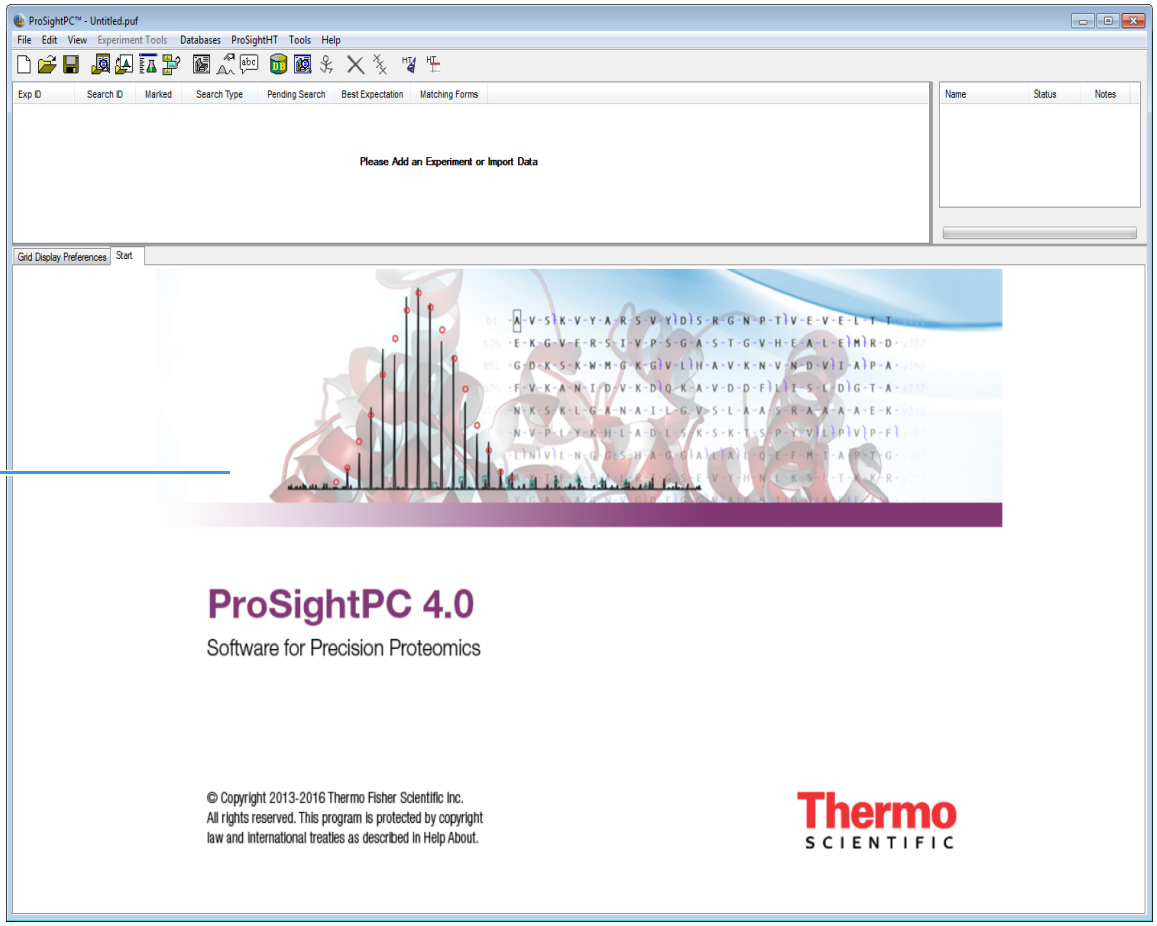
If you have never installed a ProSight PC product, you must install the correct font when you first start the application.

❖ To start the ProSightPC application

From the Start menu, choose **All Programs > Proteinaceous Inc > ProSight > ProSightPC**, or click the **ProSightPC** icon,  , on your desktop.

The ProSightPC main window displays the Start page, as shown in Figure 11.

Figure 11. ProSightPC main window



Start page

Note You can access the Start page any time by choosing **View > Start**.

Figure 107 on page 256 highlights the main features of the ProSightPC window. For information on these features and how to customize them, see "ProSightPC Application Window" on page 255.

❖ To install the ProSightPC font

1. Choose **Help> Install Font**.
2. In the dialog box that opens, click **Install**.

This procedure only needs to be done once.

Closing the ProSightPC Application

IMPORTANT Save your changes before you exit the ProSightPC application, because it does not prompt you.

❖ To exit the ProSightPC application

Choose **File > Exit**.

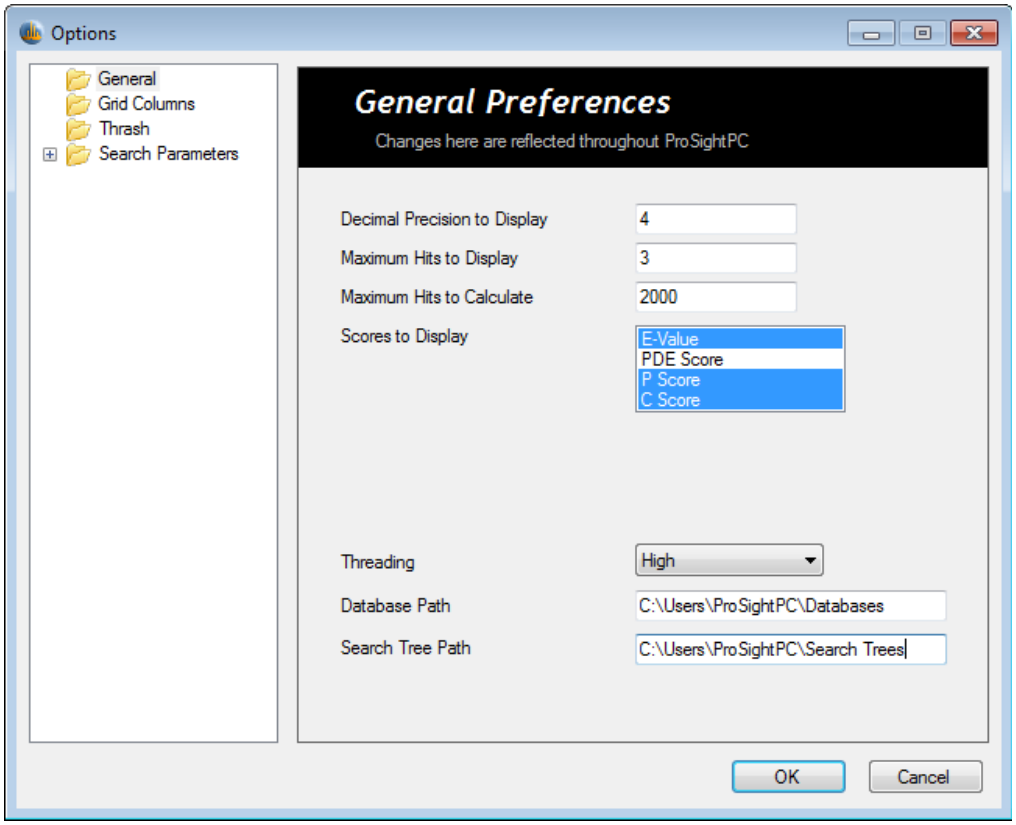
Setting Default Options

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 window.

❖ To set general preferences

1. Choose **Tools > Options** to open the Options dialog box.
2. From the left pane of the Options dialog box, click the **General** folder to open the General Preferences page, shown in Figure 12.

Figure 12. General Preferences page of the Options dialog box



3. In the Decimal Precision to Display box, specify the number of decimal places for displaying most numbers.
4. 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.
5. In the Maximum Hits to Calculate box, specify the maximum number of proteoforms that a given search considers.

   When this number is exceeded, the search automatically stops, and the ProSightPC application issues a warning.
6. 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
   - PDE Score
   - P Score
   - C Score
7. Click **OK**.

General Preferences Page Parameters

Table 2 lists the parameters on the General Preferences page of the Options dialog box.

Table 2. General Preferences page parameters (Sheet 1 of 2)

Parameter	Description
Decimal Precision to Display	Specifies the number of decimal places to display most numbers in.
Maximum Hits to Display	Specifies the number of matching fragment tables that are displayed in the Data Manager.
Maximum Hits to Calculate	Specifies the maximum number of proteoforms that a given search considers.

Table 2. General Preferences page parameters (Sheet 2 of 2)

Parameter	Description
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. For more information on this score, see "Expectation Value (e value)" on page 196. • 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 this score, see "PDE (McLuckey) Score" on page 198. • 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 this score, see "P Score" on page 196. • C Score: Displays the C score, which is a measure of the confidence in the characterization of the proteoform. For more information on this score, see "C Score" on page 199.
Threading	Specifies the number of processors to use when the High Throughput Wizard performs parallel operations or searches, or deconvolves fragments and precursors. • High: Uses n–1 processors. • Medium: Uses n/2 processors. • Low: Uses one processor. N is the number of processors.
Database Path	Displays the default location of the folder where the application saves the database (PSCW) and repository (PSPH) files.
Search Tree Path	Displays the default location of the folder where the application saves the search tree (XML) files.

For more information about setting general preferences, see "Setting Default Options" on page 21.

Importing or Creating a Proteome Database

The first step in using the ProSightPC application is to download or create a proteome database that you can search to identify the peptides and proteins in your mass spectrometry data. You can add proteome databases with multiple formats:

- FASTA databases
- UniProtKB XML and flat file databases containing PTM and sequence-variant information
- PSCW databases that are included on the ProSightPC-distributed flash drive for various organisms

You have two options for importing or creating a proteome database: download databases from the ProSightPC website to your local computer or manually create a custom database.

❖ To download a proteome database from the ProSightPC website

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

   This step takes you to proteinaceous.net with access to top-down and bottom-up databases.

2. Click the date of the database that you are interested in, for example, **July 2016**.
3. Click **Archaebacteria**, **Eukaryotes**, **Prokaryotes**, or **Custom**, as appropriate.

   You must know the taxonomy of the species of the database that you want to download.

4. Click the name of the appropriate species database.

   You can choose from one of the following database types:

   - TD Complex PSCW: Includes N-terminal acetylation and initial methionine cleavage. This database contains 12 to 15 modifications per entry.
   - TD Simple PSCW: Includes N-terminal acetylation and initial methionine cleavage. This database contains up to three modifications per entry.
   - Bottom Up PSCW: Includes trypsin digestion using two missed cleavages.
   - UniProt XML File: Recreates a PSCW database by using the Database Manager.

   The site automatically downloads the database once you select it. The downloaded database (PSCW or XML file) appears in the Downloads folder in the following directory:

   C:\Users*your_name_folder*\Downloads

5. Import the database into the ProSightPC application by following the instructions in "Importing a Proteome Database or Repository" on page 216.

❖ **To manually create a custom database**

1. Choose **Databases > Create a Custom Database**, or click 

in the Database Manager.
2. Follow the instructions in "Creating a Proteome Database" on page 217.

Editing Modifications

You can customize the chemical modifications that you use to conduct a search. You can use the Fixed Modification Editor to add fixed modifications, which apply the same specific mass to all occurrences of the named amino acid.

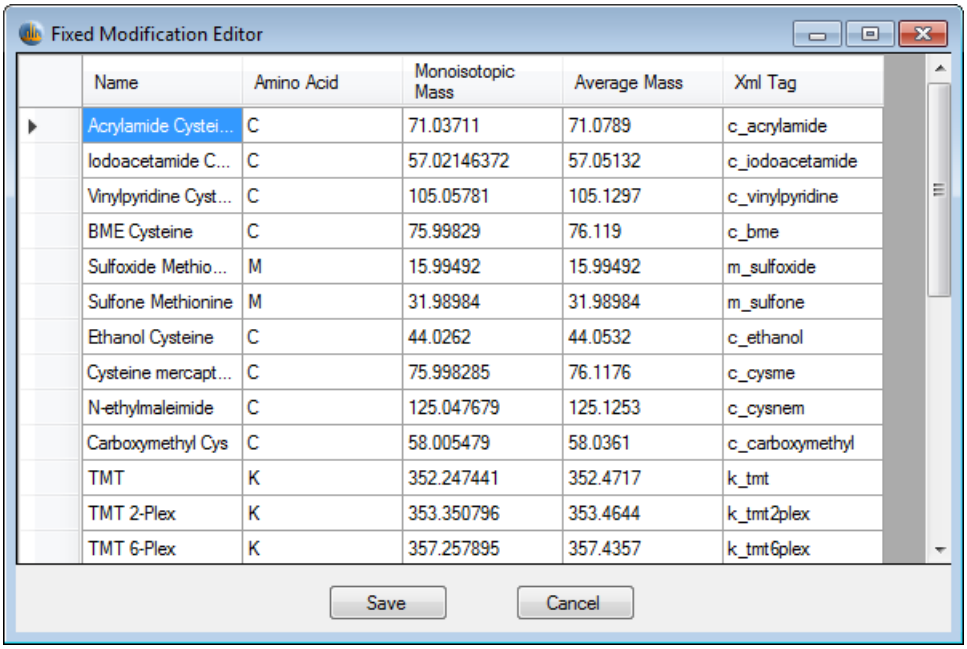
Both types of modifications are used more frequently in bottom-up searches.

Specify these modifications before you process LC/MS/MS data through the ProSightPC High Throughput Wizard.

❖ **To edit fixed modifications**

1. Choose **Tools > Fixed Modification Editor** to open the Fixed Modification Editor, shown in Figure 13.

Figure 13. Fixed Modification Editor



2. In the dialog box, do the following:
   a. Scroll down to the last row, which is marked by an asterisk (*).
   b. In the Name box, type the name of the modification.
   c. In the Amino Acid box, type the symbol of the amino acid being modified.

d. In the Monoisotopic Mass box, type the monoisotopic mass of the chemical formula of the modification.

e. In the Average Mass box, type the average mass of the chemical formula of the modification.

f. In the XML Tag box, type a single word without spaces to indicate the modification.

3. Click **Save**.

   The window closes. The modification appears when you create searches.

Fixed Modification Editor Parameters

Table 3 lists the parameters in the Fixed Modification Editor.

Table 3. Fixed Modification Editor dialog box parameters

Parameter	Description
Name	Specifies the name of the modification.
Amino Acid	Specifies the symbol of the amino acid being modified.
Monoisotopic Mass	Specifies the monoisotopic mass of the chemical formula of the modification.
Average Mass	Specifies the average mass of the chemical formula of the modification.
XML Tag	Specifies the XML tag used in the search. The tag must be one word without spaces.

Setting the Parameters for the Search

The next step is to specify the type of search to perform and to set the parameters for that search. To do this, create a predefined search by following the instructions in "Creating a Predefined Search" on page 105.

Using the High Throughput Wizard to Process LC/MS/MS Data Files

If you want to load an LC/MS/MS raw data file or a PUF file as input, you can use the ProSightPC High Throughput Wizard to search the database that you downloaded or created for matches to your mass spectrometry data. This topic explains how to use this wizard and how to set custom processing options.

You cannot use the High Throughput Wizard to import a targeted raw data file as input or to enter data manually into the ProSightPC application. Instead, you must use the procedure given in "Importing Targeted Raw Data Files" on page 75 to import a targeted raw data file or the procedure given in "Entering Data Manually" on page 88 to enter data manually.

To load an LC/MS/MS raw data file or a PUF file, follow these procedures:

- Setting Processing Options
- Selecting or Creating a Repository
- Selecting an Existing Search Tree
- Creating a Search Tree
- Editing or Adding a Search Tree
- Deleting a Search Tree
- Viewing the Summary
- Processing the Data
- Demonstrating the High Throughput Wizard
- Using Custom Settings in the High Throughput Wizard

To view a demonstration of these procedures, see "Demonstrating the High Throughput Wizard" on page 51.

Setting Processing Options

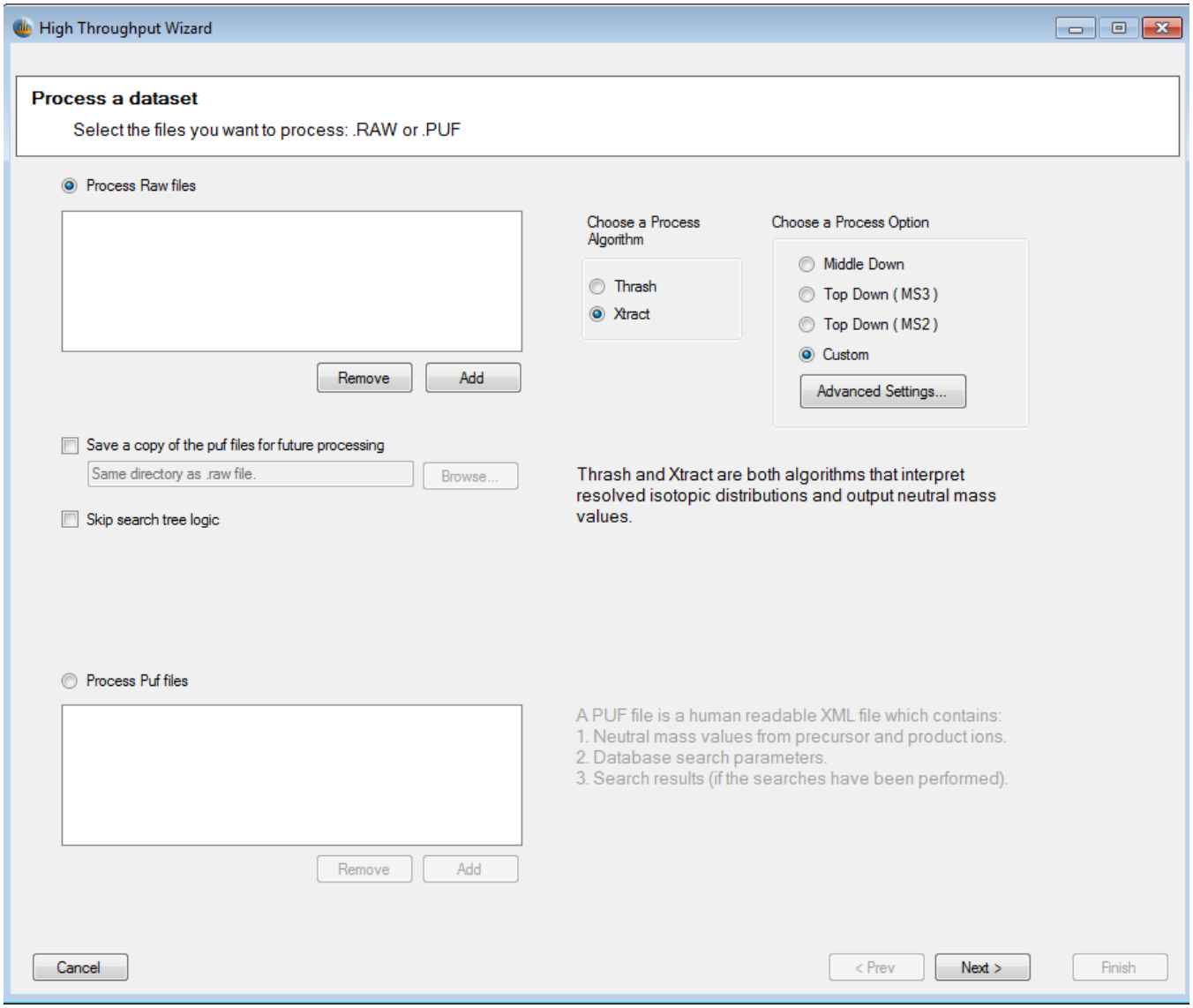
To set the processing options in the High Throughput Wizard, follow this procedure.

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

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

   The Process a Dataset page of the High Throughput Wizard appears, as shown in Figure 14.

Figure 14. Initial 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.

   - (Default) Process Raw Files: Converts LC/MS/MS raw data 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, such as the Orbitrap Elite™.

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

   - Process Puf Files: Processes PUF files.

3. To add a raw data or a 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 data 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.
   - (Default) Xtract: Uses the Xtract algorithm to process the input file. This option reduces analysis and search time and should give better results.

   Both Xtract and THRASH are algorithms that interpret resolved isotopic distributions and output neutral mass values. For more information on these algorithms, see "Importing Targeted Raw Data Files" on page 75.
5. If you choose a raw data 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	Cleared
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	Cleared
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	0.9
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	Cleared
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	0.9
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 in the High Throughput Wizard" on page 55 for instructions.

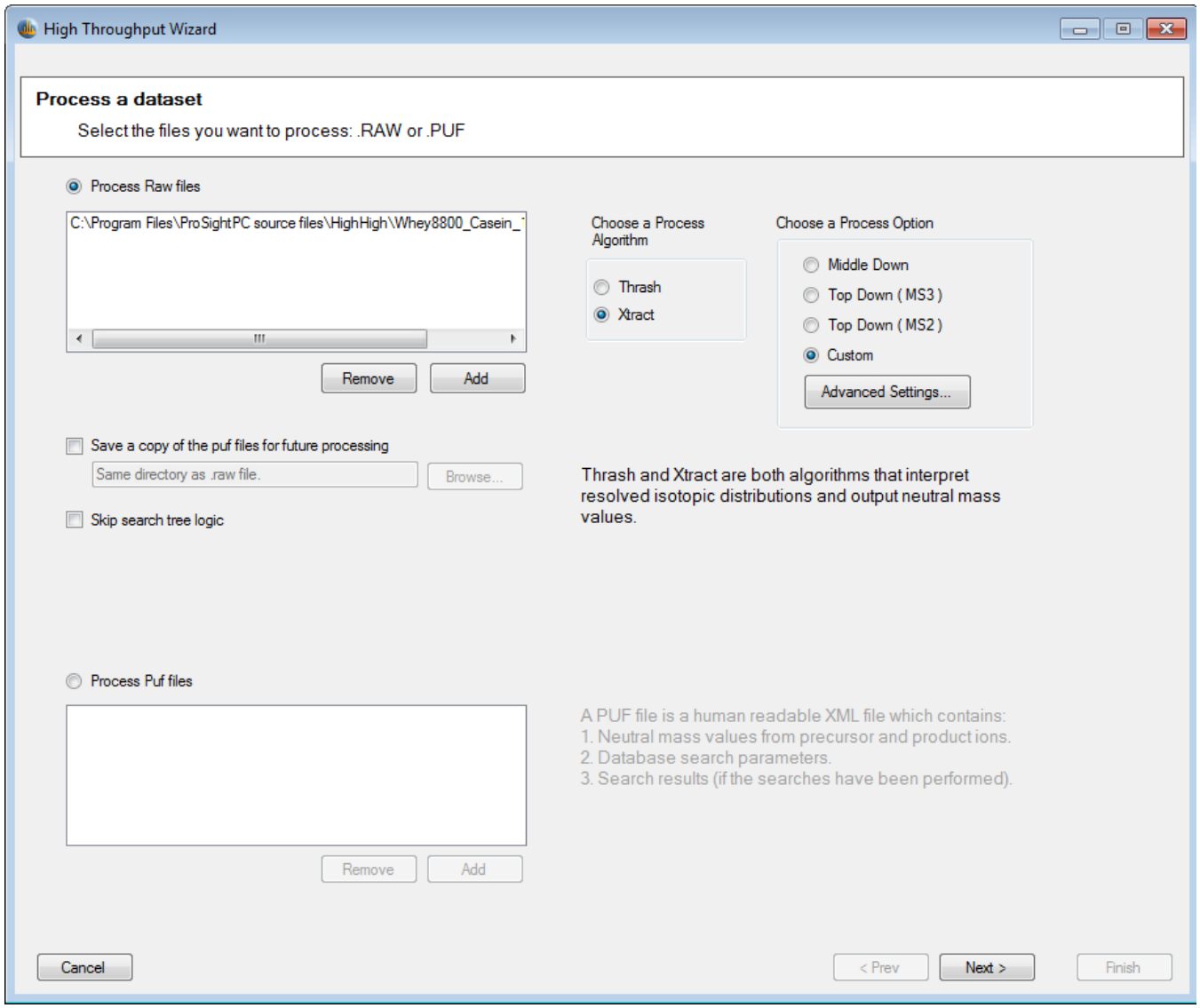
6. (Optional) If you select the Process Raw files option, select the **Save a Copy of the Puf Files for Future Processing** option to save a physical PUF file containing the results. Click **Browse** to browse to the directory where you want to save the PUF files.

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

   The Process a Dataset page now resembles Figure 15.

Figure 15. Completed Process a Dataset page of the High Throughput Wizard



8. Click **Next**.

Process a Dataset Page Parameters

Table 4 lists the parameters in the Process a Dataset page of the High Throughput Wizard.

Table 4. Process a Dataset page parameters (Sheet 1 of 2)

Parameter	Description
Process Raw Files	Converts LC/MS/MS raw data 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, such as the Orbitrap Elite.
Remove	Removes the selected raw data file displayed in the box.
Add	Opens a dialog box so that you can browse for a raw data file to process.
Choose a Process Algorithm	Specifies the method for converting mass spectral data to neutral mass values when importing the data files: • (Default) 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. For more information on these algorithms, see "Importing Targeted Raw Data Files" on page 75.

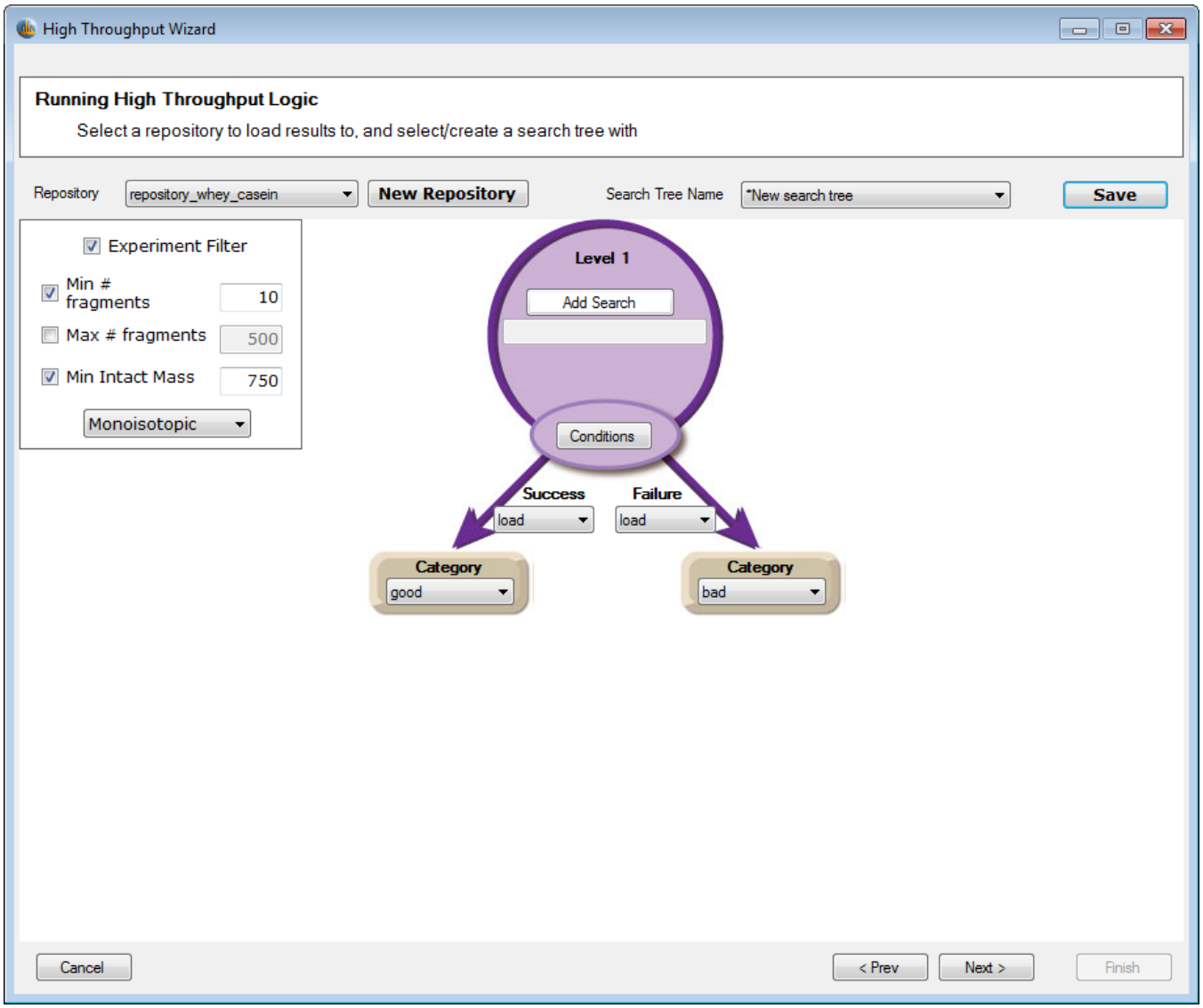
Table 4. Process a Dataset page parameters (Sheet 2 of 2)

Parameter	Description
Choose a Process Option	Specifies the settings for the Xtract and THRASH processing algorithms: • Middle Down: See step 5 of "Setting Processing Options" on page 27 for this set of default settings. • Top Down (MS3): See step 5 of "Setting Processing Options" on page 27 for this set of default settings. • Top Down (MS2): See step 5 of "Setting Processing Options" on page 27 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 61 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 data 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.
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 High Throughput Logic page of the High Throughput Wizard appears, as shown in Figure 16, so that you can select or create a repository and define a search tree.

Figure 16. Initial Running High Throughput Logic page of the High Throughput Wizard



For information about the parameters on this page of the dialog box, see Table 5.

❖ To select a repository

From the Repository list on the Running High Throughput Logic page, shown in Figure 16, select the name of the repository.

❖ To create a repository

1. Click **New Repository**.
2. Enter the new repository name in the New Repository dialog box, shown in Figure 31.

3. Click **OK**.

   The Edit/Add Repositories dialog box appears, as shown in Figure 32.

4. If you do not want to edit the repository, click **Save**. If you want to edit the repository, follow the instructions in "Editing a Repository" on page 69, and click **Save** in the Edit/Add Repositories dialog box.

   The name of the repository appears in the Repository box of the Running High Throughput Logic page of the High Throughput Wizard.

For more information on creating a repository, see "Creating a Repository" on page 69.

Selecting an Existing Search Tree

❖ **To select an existing search tree**

From the Search Tree Name list on the Running High Throughput Logic page of the High Throughput Wizard dialog box, shown in Figure 16, select the name of the search tree.

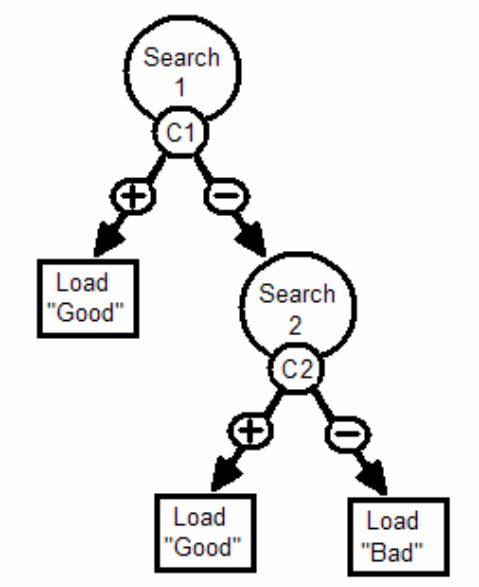
Creating a Search Tree

The ProSightPC application 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 the ProSightPC application enters the search tree at the top, as shown in Figure 16. The ProSightPC application 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, the application 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, the ProSightPC application loads the results into the "good" category, but if the results are greater than 0.0001, the application tries another search with "looser" search parameters—that is, a larger precursor search window, biomarker mode, or delta-m (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 17. The ProSightPC application 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, the application loads the results to the "good" category. If they do not meet the conditions, the application loads them 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-m (Δm) mode searches, to the results repository.

Figure 17. 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** from the Search Tree Name list on the Running High Throughput Logic page of the High Throughput Wizard.

   This option 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 High Throughput 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.

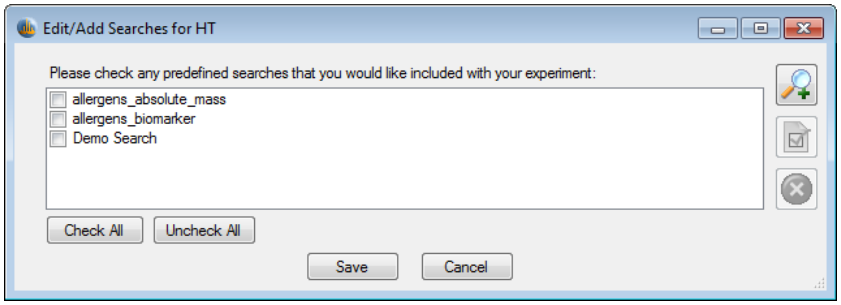
d. From the list beneath the Min Intact Mass option, specify the mass type:

- 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.

The ProSightPC application opens the Edit/Add Searches for HT dialog box, shown in Figure 18.

Figure 18. Edit/Add Searches for HT 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 104.

b. To add a search, click  in the dialog box.

The New Predefined Search dialog box opens. Follow the instructions in "Creating a Predefined Search" on page 105 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 109 to edit a predefined search.

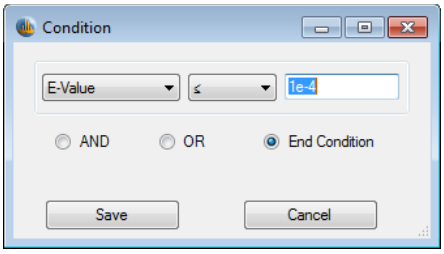
d. To remove a predefined search, select the name of the search and click  .

e. Click **Save** in the Edit/Add Searches for HT dialog box.

4. To set the conditions for the first search, click **Conditions** in the purple circle in the Level 1 area of the Running High Throughput Logic page.

The Condition dialog box appears, as shown in Figure 19. You can use operators and values to create conditions for the search.

Figure 19. Condition dialog box



a. From the list on the left, select **Number of Hits**, **P Score**, **E-Value**, or **C-Score**.

- Number of Hits: Specifies the number of matches for an intact ion in the search.
- P Score: Specifies 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 196.
- 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 1e-4, the search is loaded to the "good" category. For more information on the expectation value, see "Expectation Value (e value)" on page 196. This setting is the default.
- C-Score: Specifies the C score, which is a measure of the confidence in the characterization of the proteoform. For more information on the C score, see "C Score" on page 199.

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 appropriate value.

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

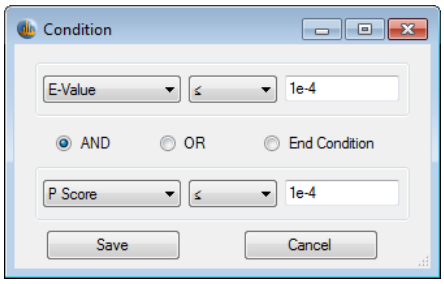
d. Select one of the following operators:

- 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 20, so that you can add a second condition.

Figure 20. Expanded Condition dialog box



e. When you have set the conditions for all searches, click **Save** to return to the Running High Throughput Logic page.

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 25. 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 "Creating a Two-Level Search Tree" on page 45.

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 25. 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 "Creating a Two-Level Search Tree" on page 45.

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.

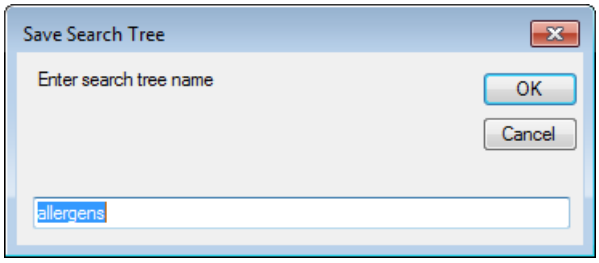
9. (Optional) To create a second-level search, see "Creating a Two-Level Search Tree" on page 45.

10. (Optional) To create a search with three or more levels, see "Creating a Search Tree with Three or More Levels" on page 47.

11. 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 21.

Figure 21. Save Search Tree dialog box



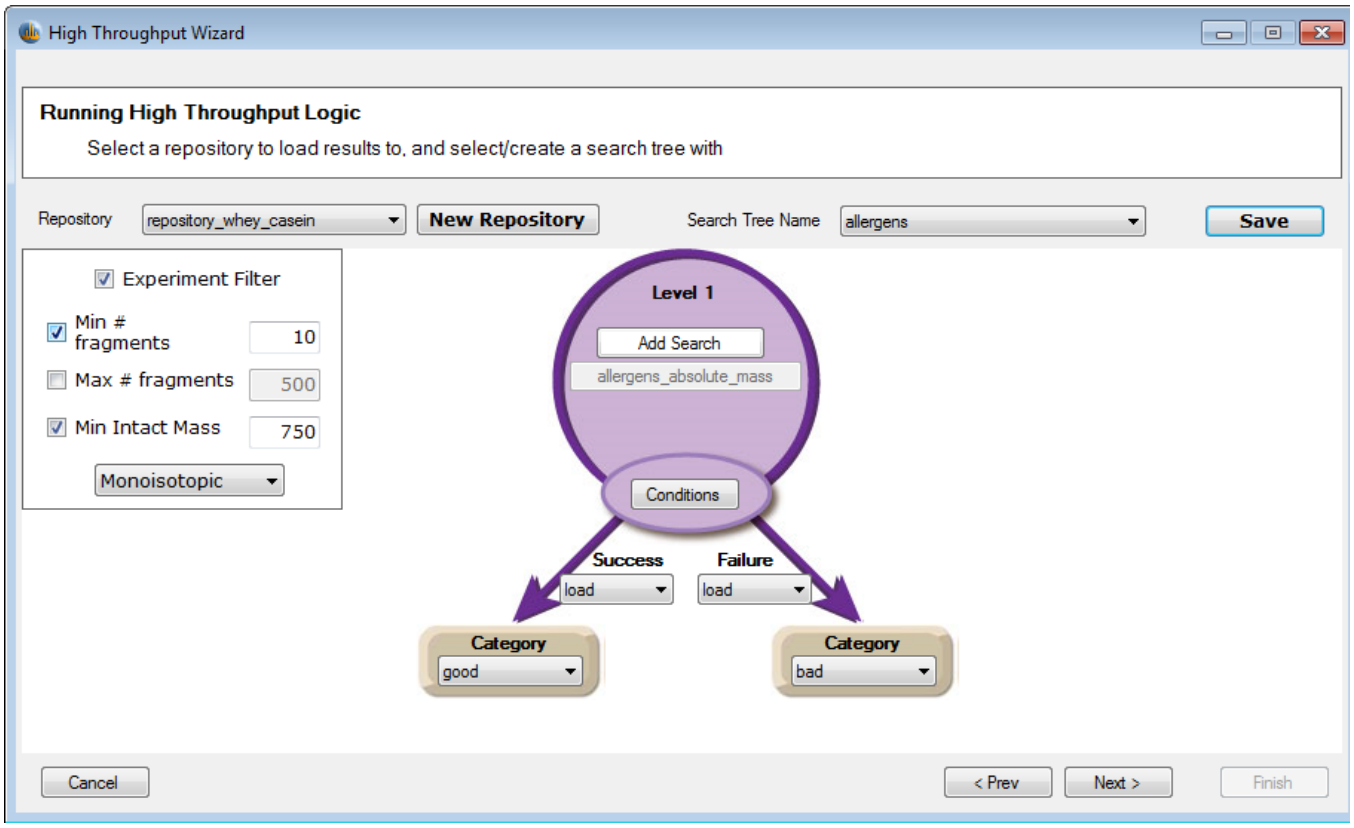
12. If you created a new search tree, type the name of the search tree and click **OK**.

    If you made changes to an existing search tree, a prompt box appears to confirm that you want to replace the existing search tree. Click **Yes**.

13. Click **OK** in the message box that appears.

    The completed Running High Throughput Logic page now resembles Figure 22.

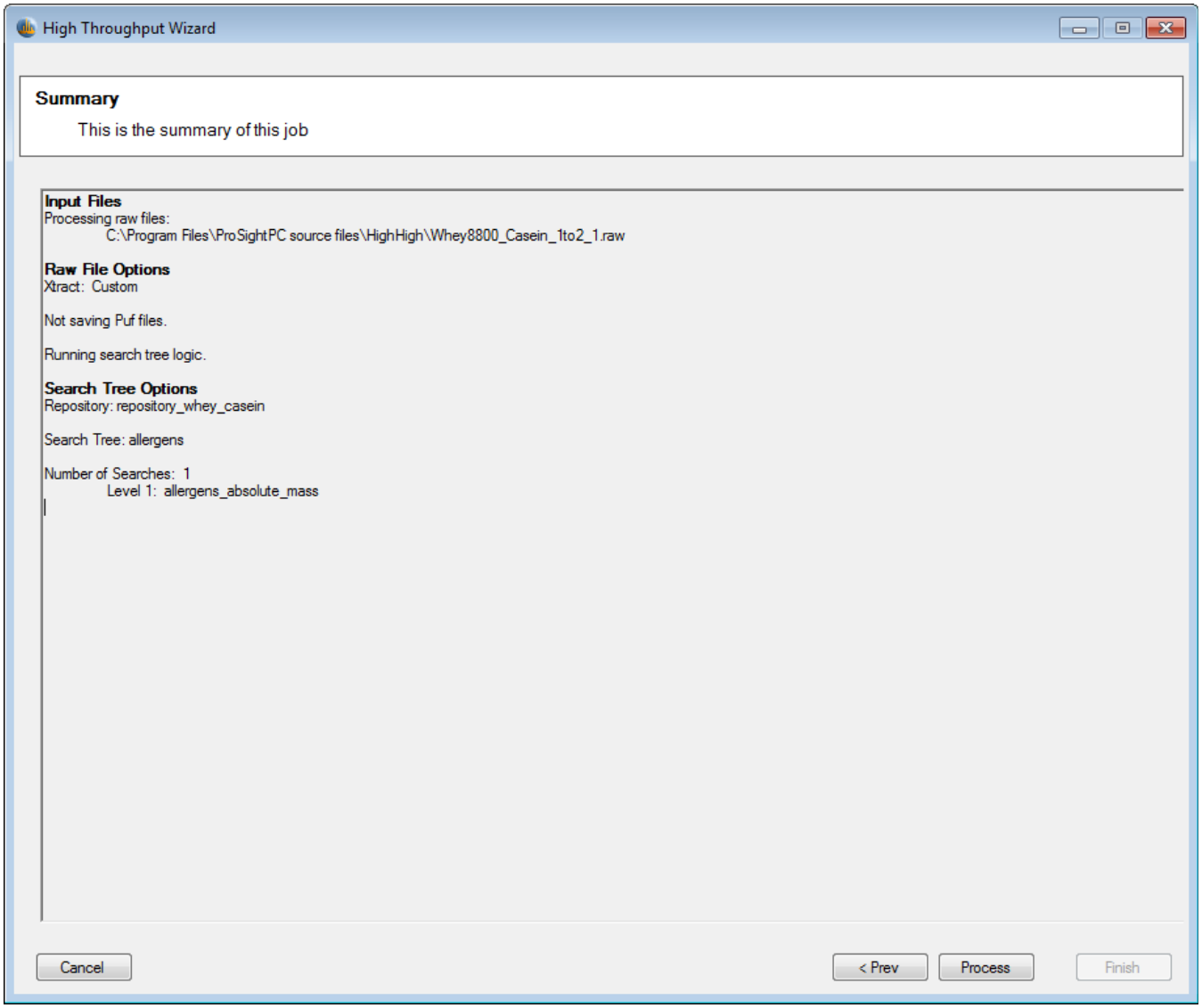
Figure 22. Completed Running High Throughput Logic page of the High Throughput Wizard



14. Click **Next** on the Running High Throughput Logic page.

    A summary of the parameters that you have set appears, as shown in Figure 23.

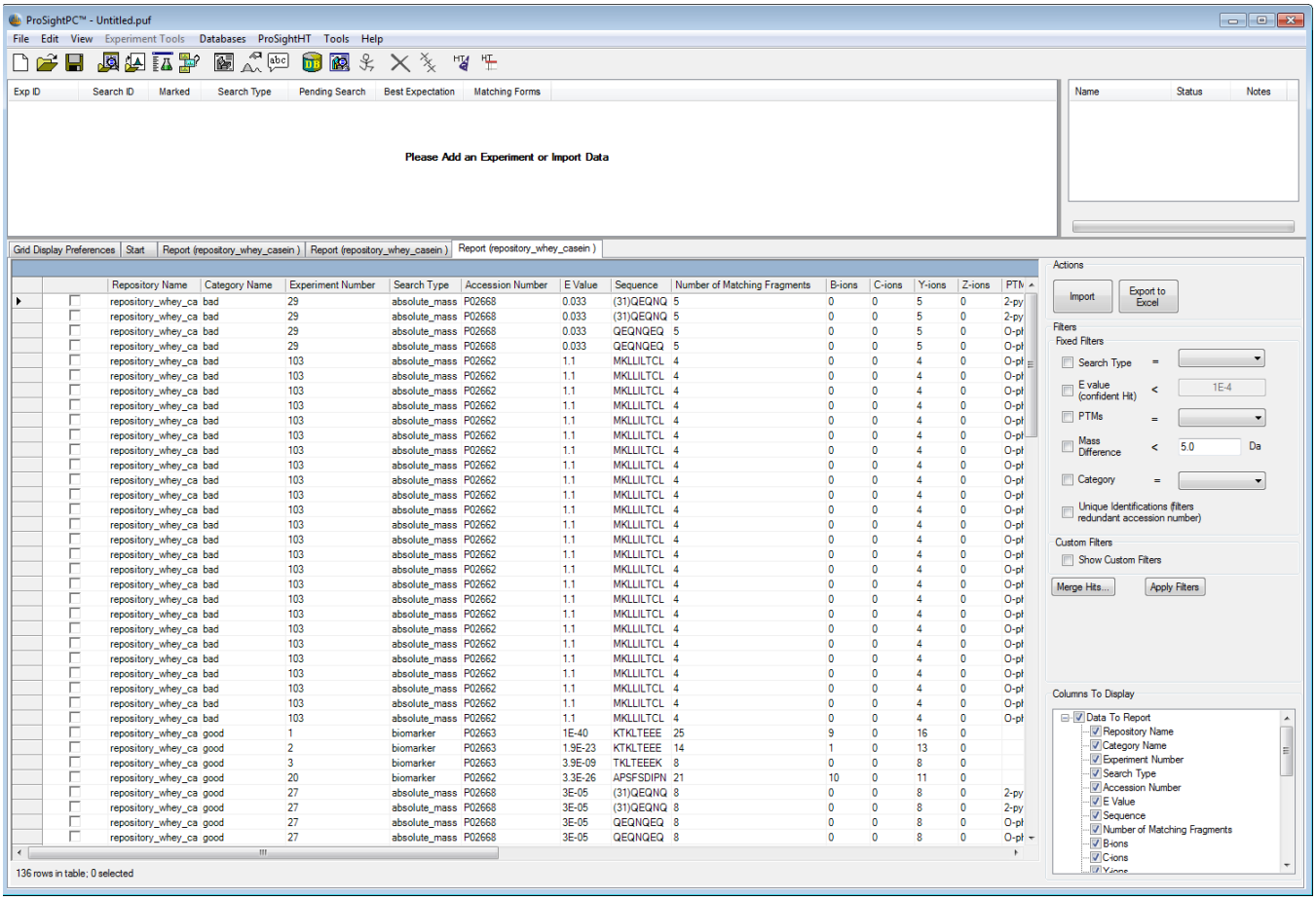
Figure 23. Summary page of the High Throughput Wizard



15. Click **Process**.

The searching begins and can take a several minutes, depending on the length of the raw data file, the complexity of the database, and the parameters of the search. The application deconvolves all the MS1 and MS/MS scans in the raw data file. When the ProSightPC application finishes the search, it opens the repository report (see Figure 24).

Figure 24. Initial repository report



For information on manipulating the data in this report, see "Viewing the Results in a Repository Report" on page 174.

If you selected a PUF file as input, the application also generates a PUF file in the directory containing the raw data file. The PUF file contains a list of deconvolved spectra but does not contain protein and peptide identification information. You must manually add the identification information to the PUF file.

❖ To add peptide and protein identifications to the PUF file

1. Follow the instructions in "Adding Predefined Searches to an Experiment" on page 108.
2. Choose **Tools > Batch Run**, or click the **Batch Run** icon, .
3. Choose **File > Save**, or click the **Save** ( ) icon.

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.

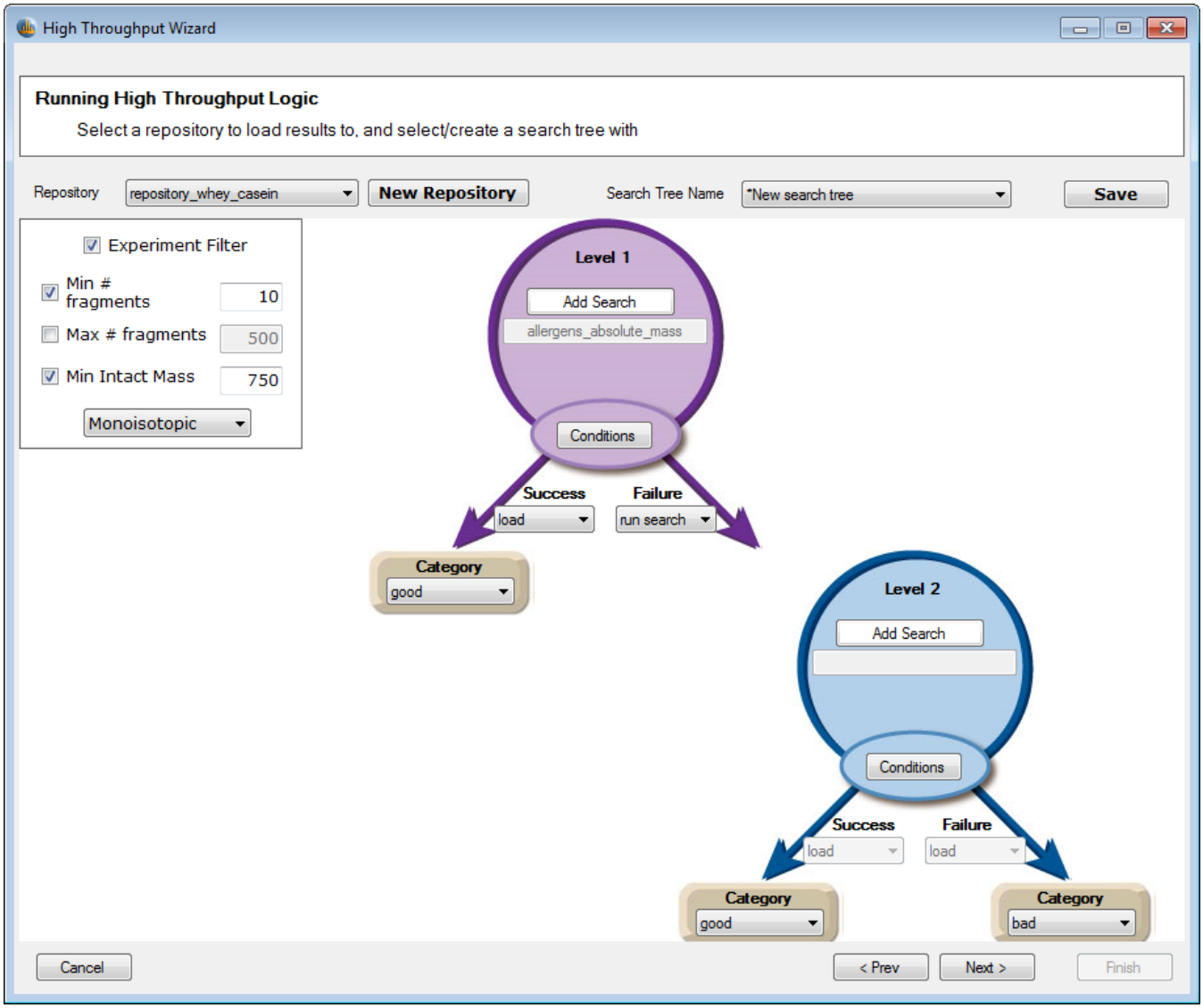
You might also want to perform a two-level search on the Failure side. Suppose that you ran a first-level search and some of the experiments found no matches—that is, the experiments failed the conditions. To obtain good results, you might then want to run a different search, such as a broader search 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 38, and select **Run Search** in the Success or Failure list for the first-level search.

   A second-level search tree opens, as shown in Figure 25.

Figure 25. Second-level search tree

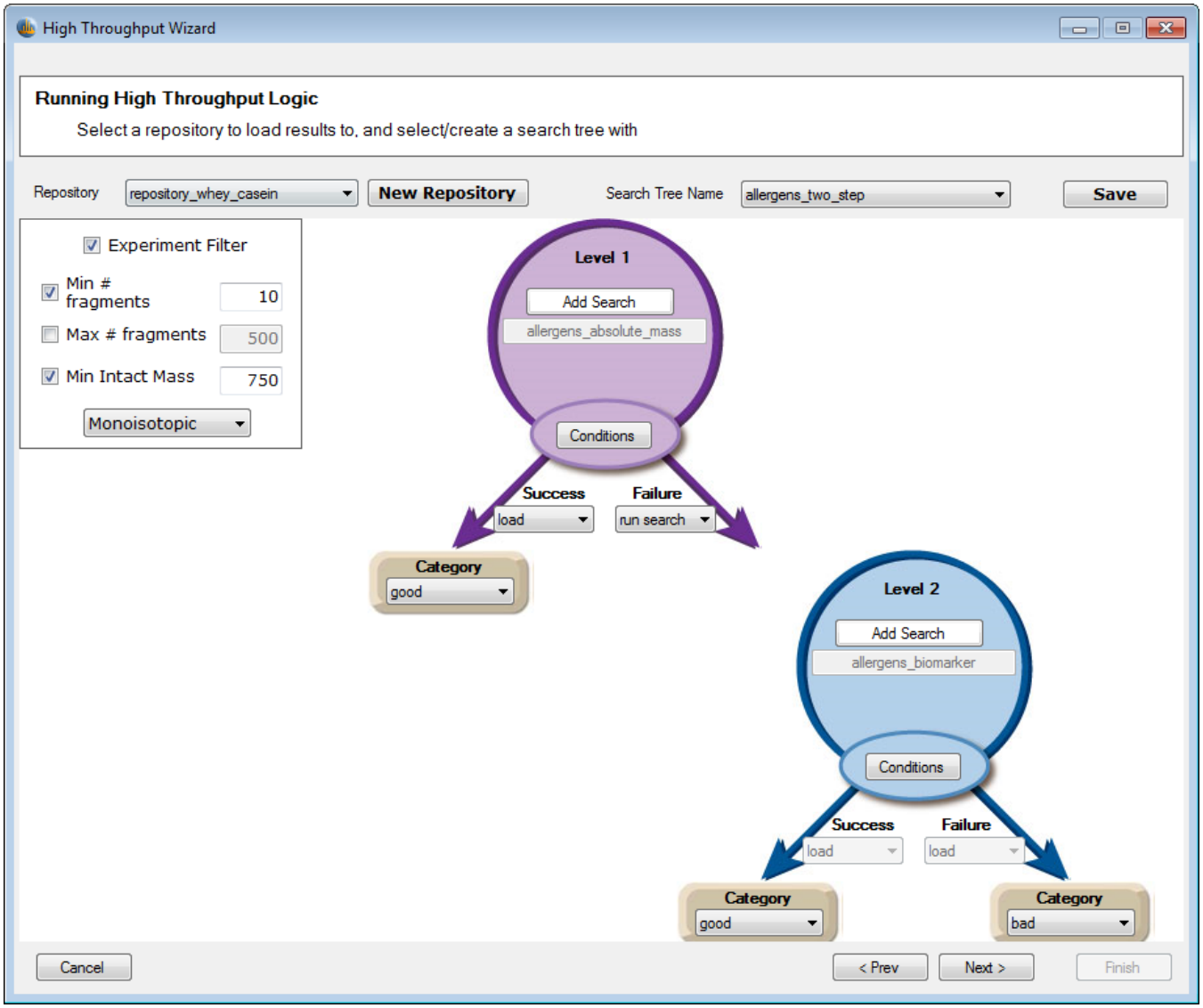


Note To delete the second search level after you have added it, select **Load** in the same Success or Failure list.

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 topic, "Creating a Search Tree with Three or More Levels" on page 47.

   Figure 26 shows a completed second-level search.

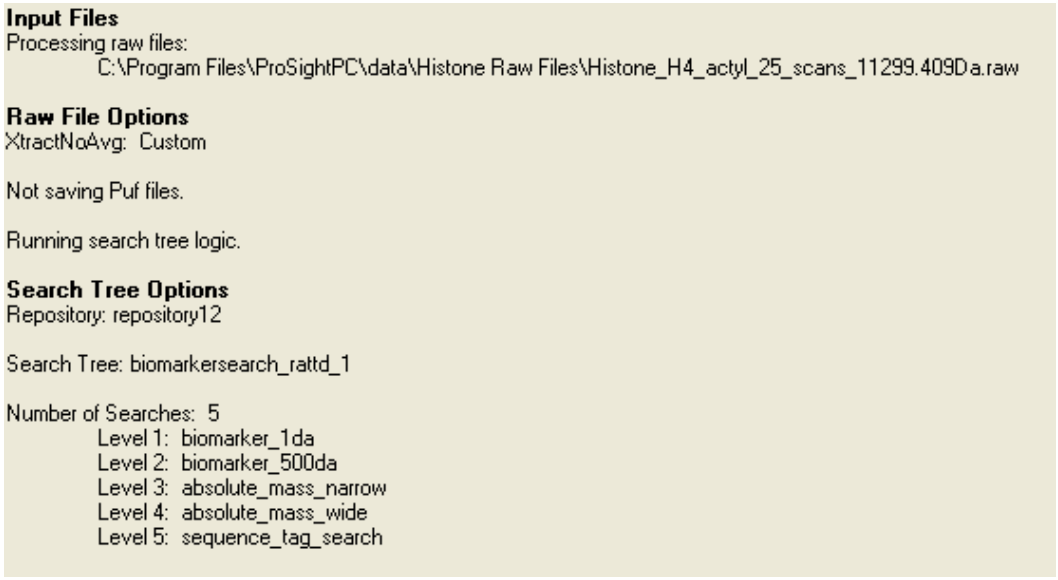
Figure 26. Completed second-level search tree



Creating a Search Tree with Three or More Levels

You can use the ProSightPC application 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 27.

Figure 27. Summary page for a five-level search



For more information on how to edit this XML file to add more than two levels, contact Proteinaceous at info@proteinaceous.net for assistance.

Editing or Adding a Search Tree

Once you set up a search tree, you can go back and edit it or add another level to it.

If you selected the Skip Search Tree Logic check box on the Process a Dataset page of the High Throughput Wizard, you can also add a search tree.

❖ To edit a search tree

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

   The Running High Throughput Logic page appears in the High Throughput Wizard, as shown in Figure 16.

2. If you are editing an existing search tree, from the Search Tree Name list, select the search tree that you would like to edit or add a level to.

   If you are adding a search tree, follow the next steps.

3. Follow the instructions in "Creating or Editing a One-Level Search Tree" on page 38 and "Creating a Two-Level Search Tree" on page 45 to edit your search-tree settings or to add a search tree.

4. Click **Save** on the Running High Throughput Logic page.

Deleting a Search Tree

You must delete a search tree from the Search Trees folder in the default folder of your search tree path.

❖ To delete a search tree

1. Close the ProSightPC application.
2. Navigate to the following folder under the ProSightPC search tree path, unless you changed it in the Search Tree Path box of the General Preferences dialog box activated by the Tools > Options > General command:

   C:\Users*user_name*\My Documents\ProSightPC\Search Trees
3. Delete the XML file that shares the same name as your search tree.
4. Reopen the ProSightPC application.

Viewing the Summary

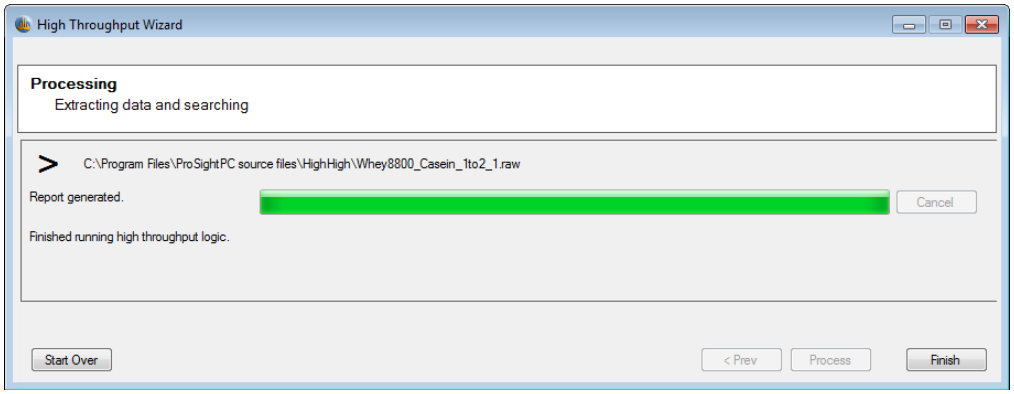
Review your parameter settings on the Summary page of the High Throughput wizard.

Processing the Data

❖ To process the data

1. To begin the ProSightPC High Throughput processing, click **Process** at the bottom.
2. When the High Throughput wizard finishes processing the data, as shown in Figure 28, click **Finish**.

Figure 28. Processing page of the High Throughput Wizard



Once the search is complete (which might take several minutes, depending on the size of the raw data file, the complexity of the database, and the search parameters), you can view the repository report. You can use filter, import, and export options to view the report. For each raw data file, a table displaying all the matches found appears in the ProSightPC application. Each row represents the best match per MS/MS experiment. See "Viewing the Results in a Repository Report" on page 174.

Demonstrating the High Throughput Wizard

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

Running High Throughput Logic Page Parameters

Table 5 lists the parameters on the Running High Throughput Logic page of the High Throughput Wizard, shown in Figure 16.

Table 5. Running High Throughput Logic page parameters (Sheet 1 of 3)

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 31, 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 21, so that you can save a search tree under a new name.
Experiment Filter	Filters out experiments that will not yield matches.

Table 5. Running High Throughput Logic page parameters (Sheet 2 of 3)

Parameter	Description
Min # Fragments	Specifies the minimum number of fragments to search for. Default: 10
Max # Fragments	Specifies the maximum number of fragments to search for. Default: 500
Min Intact Mass	Specifies the minimum intact mass. Default: 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 Edit/Add Searches for HT dialog box, shown in Figure 18, so you can add a predefined search to your search.
Conditions	Opens the Condition dialog box, shown in Figure 19, so you can sets the conditions for the search.
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 25. 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 "Creating a Two-Level Search Tree" on page 45 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.

Table 5. Running High Throughput Logic page parameters (Sheet 3 of 3)

Parameter	Description
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 25. 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 "Creating a Two-Level Search Tree" on page 45 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.

Edit/Add Searches for HT Dialog Box Parameters

Table 6 lists the parameters in the Edit/Add Searches for HT dialog box, shown in Figure 18.

Table 6. Edit/Add Searches for HT dialog box parameters

Parameter	Description
Please Check Any Predefined Searches That You Would Like Included with Your Experiment	Lists the available predefined searches.
Demo Search	Searches the demonstration database included in the installation of the ProSightPC software.
Check All	Selects all the available predefined searches.
Uncheck All	Clears all the available predefined searches.
Save	Saves the changes that you made to the predefined searches.
	Adds a predefined search.
	Edits a predefined search.
	Removes a predefined search.

Condition Dialog Box Parameters

Table 7 lists the parameters in the Condition dialog box, shown in Figure 19.

Table 7. Condition dialog box parameters

Parameter	Description
Left list	Sets one of two conditions that the results of the search must meet: • Number of Hits: Specifies the number of matches for an intact ion in the search. • P Score: Specifies 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 196. • (Default) 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 196 for information on the expectation value. • C-Score: Specifies the C score, which is a measure of the confidence in the characterization of the proteoform. For more information on the C score, see "C Score" on page 199.
Middle list	Specifies the operator that indicates the relationship between the values in the left and right boxes. • (Default) ≤: Indicates that the first value is less than or equal to the second value. • ≥: Indicates that the first value is greater than or equal to the second value.
Right list	Specifies the applicable 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 or that this is the last 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 21, has one parameter for specifying the name of the search tree that you created.

Using Custom Settings in the High Throughput Wizard

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 in Figure 14, click **Advanced Settings.**

   If you chose the THRASH algorithm in the Choose a Process Algorithm area, the version of the Advanced Settings dialog box shown in Figure 29 opens. If you chose the Xtract algorithm, the version of the Advanced Settings dialog box shown in Figure 30 opens.

Figure 29. Advanced Settings dialog box for THRASH

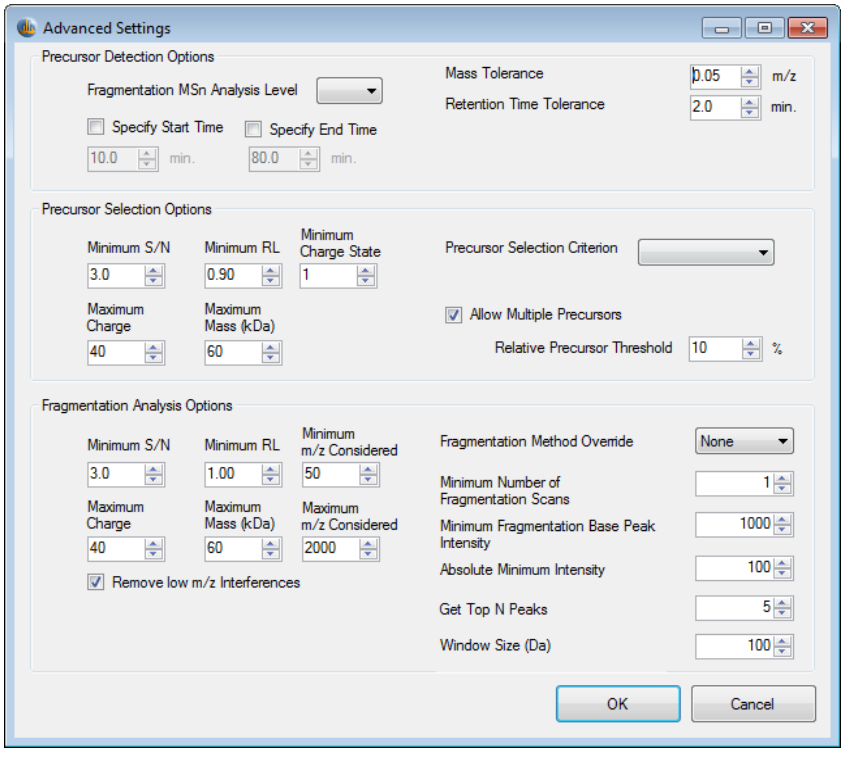
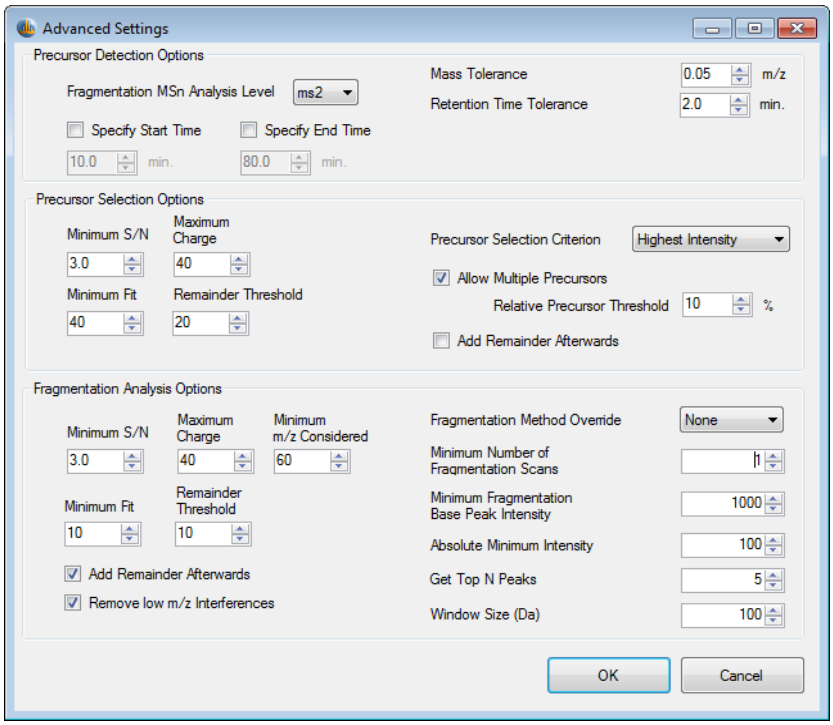


Figure 30. Advanced Settings dialog box for Xtract



2. In the Precursor Detection Options area, specify the level of analysis that includes your fragmentation scans in the raw data file where the ProSightPC application infers the precursor scan.

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

      - (Default) ms2: For data-dependent LC/MS/MS experiments
      - ms3: For ion-trap marching experiments

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

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

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

      This value 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 data file is analyzed.

   d. In the Mass Tolerance check box, specify an *m/z* tolerance that determines which scan filters are summed together.

      If the mass (and retention time) is within the tolerance, the ProSightPC application combines the scan filters.

The minimum value is 0.01, and the maximum value is 1.0 *m/z*. The default is 0.05 *m/z*.

e. In the Retention Time Tolerance box, specify a retention time tolerance, in minutes, that determines which scan filters are summed together.

You can specify a value of at least 0.1 minutes, and there is no maximum value. The default is 2.0 minutes.

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

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

The minimum value is 1, and there is no maximum value. The default is 3.0.

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

The minimum value is 1, and there is no maximum value. The default is 40.

c. (Xtract only) In the Minimum Fit box, enter the minimum fit parameter used by the Xtract algorithm.

The minimum value is 0, and the maximum value is 100. The default is 40.

d. (THRASH only) In the Minimum RL box, enter the minimum confidence level.

The minimum value is 0, and the maximum value is 1.0. The default is 0.90.

e. (THRASH only) In the Maximum Mass (kDa) box, enter the highest mass to be considered for the precursor.

The minimum value is 1, and there is no maximum value. The default is 60.

f. (THRASH only) In the Minimum Charge State box, select the smallest charge state to be considered for the precursor.

The minimum value is 1, and there is no maximum. The default is 1.

g. (Xtract only) 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. This 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. The ProSightPC application 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 or protein, even if it is overlapped by a pattern with 10-fold greater intensity.

The minimum value is 0, and the maximum value is 100. The default is 20.

h. 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.

i. 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. This parameter is selected by default.
- In the Relative Precursor Threshold box, specify a threshold for selecting the precursor intensities when there are multiple precursors within the window. The ProSightPC application selects only precursors with intensities within the top *x* percent of the top precursor. 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, the ProSightPC application searches ions A and B with the fragmentation data but does not search ion C. The default is 10 percent.

The ProSightPC application 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.

j. (Xtract only) 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 deconvolved 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 deconvolved spectrum, because unassigned remainders are added to the corresponding pattern.

4. In the Fragmentation Analysis Options area, specify the parameters for analyzing fragment ions.
   a. In the Minimum S/N box, enter the lowest signal-to-noise ratio that the algorithm considers when trying to assign neutral mass to data in your raw data files.

      The minimum value is 1, and there is no maximum value. The default is 3.0.

   b. (THRASH only) In the Minimum RL box, enter the minimum confidence level.

      The minimum value is 0, and the maximum value is 1.0. The default is 1.0.

   c. In the Minimum m/z Considered box, select the lowest *m/z* to be considered for fragments.

      The minimum value is 1, and there is no maximum value. The default is 50 for THRASH and 60 for Xtract.

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

      The minimum value is 1, and there is no maximum value. The default is 40.

   e. (Xtract only) In the Minimum Fit box, enter the minimum fit parameter used by the algorithm.

      The minimum value is 0, and the maximum value is 100. The default is 10.

   f. (THRASH only) In the Maximum Mass (kDa) box, enter the highest mass to be considered for the fragments.

      The minimum value is 1, and there is no maximum value. The default is 60.

   g. (THRASH only) In the Maximum m/z Considered box, select the lowest *m/z* to be considered for fragments.

      The minimum value is 1, and there is no maximum value. The default is 2000.

   h. (Xtract only) In the Remainder Threshold box, enter the remainder of the fit that is left in the scan during analysis of the precursor ions.

      The minimum value is 0, and the maximum value is 100. The default is 10.

   i. (Xtract only) If you want to add the remaining intensities to the output spectrum, select the **Add Remainder Afterwards** check box. This parameter is selected by default.

   j. If you want the ProSightPC application to remove fragments arising from immonium ions and reagent ions from TMT™ and iTRAQ™ quantifications, select the **Remove Low m/z Interferences** check box. This parameter is selected by default.

k. In the Fragmentation Method Override list, specify the fragmentation method used if you do not want to use the fragmentation method given in the scan header.

   You can select from the following fragmentation methods:

   - (Default) None: The High Throughput Wizard uses the fragmentation method given in the scan header.
   - CID
   - ECD
   - ETD
   - HCD
   - IRMPD
   - UVPD
   - EThcD

   For information on these methods, see "Fragmentation Methods" on page 9.

l. In the Minimum Number of Fragmentation Scans box, enter a value to filter out low-quality spectra.

   The minimum value is 1, and there is no maximum value. The default is 1. In some cases, you might want to only consider precursors that have been fragmented twice or more.

m. In the Minimum Fragmentation Base Peak Intensity box, enter a value that will filter out noise and poor-quality data during analysis of the fragment ions.

   You can enter a minimum value of 1, and there is no maximum. The default is 1000. A value of 500 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, the ProSightPC application skips the scan and discards the precursor mass. This step eliminates the processing of bad MS/MS 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.

n. In the Absolute Minimum Intensity box, enter the minimum intensity to be accepted for fragmentation peaks.

   The ProSightPC application excludes any deisotoped peaks below this threshold, so it removes low-intensity fragment ions that might be spurious.

   The minimum value is 1, and there is no maximum value. The default is 100.

o. In the Get Top N Peaks box, specify the number of the most intense peaks per window size that the ProSightPC application considers.

This parameter works with the Window Size parameter to filter the deisotoped and decharged data. The default settings mean that the ProSightPC application considers only the most intense 5 peaks in a 100-Da window. Therefore, this setting removes low-intensity fragment ions that might be spurious.

The minimum value is 1, and there is no maximum value. The default is 5.

p. In the Window Size box, specify the size of the window containing the number of the most intense peaks that the ProSightPC application considers.

This parameter works with the Get Top N Peaks parameter to filter the deisotoped and decharged data.

The minimum value is 1, and there is no maximum value. The default is 100.

q. Click **OK**.

Advanced Settings Dialog Box Parameters

Table 8 lists the parameters in the Advanced Settings dialog box, which is shown in Figure 29 for the THRASH algorithm and in Figure 30 for the Xtract algorithm. This table includes the parameters for both the THRASH and the Xtract algorithms.

Table 8. Advanced Settings dialog box parameters (Sheet 1 of 8)

Parameter	Description
Precursor Detection Options	
Fragmentation MS*n* Analysis Level	Specifies the level of analysis that includes your fragmentation data in the scan in the raw data file where the ProSightPC application infers the precursor scan. You can select one of the following: • (Default for Xtract) 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 value is the start of the first scan. Default: 10 minutes
Specify End Time	Specifies the end of the chromatographic time range in which to analyze the data. This value is the end of the first scan. Default: 80 minutes

Table 8. Advanced Settings dialog box parameters (Sheet 2 of 8)

Parameter	Description
Mass Tolerance	Specifies a tolerance that determines which scan filters are summed together. If the mass (and retention time) is within the tolerance, the ProSightPC application combines the scan filters. Range: 0.01–1.0 *m/z* Default: 0.05 *m/z*
Retention Time Tolerance	Determines which scan filters are summed together. If the retention time (and mass) is within the tolerance, the ProSightPC application combines the scan filters. Range: 0.1–no maximum Default: 2.0 minutes
Precursor Selection Options	
Minimum S/N	Specifies the lowest signal-to-noise ratio that the THRASH algorithm considers when trying to assign neutral mass to a charged (mass-to-charge ratio [*m/z*]) species while analyzing the precursor ions. Range: 1–no maximum Default: 3.0
Minimum RL (THRASH only)	Specifies the minimum confidence level. Range: 0–1.0 Default: 0.90
Minimum Charge State (THRASH only)	Specifies the smallest charge state to be considered for the precursor. Range: 0–no maximum Default: 1
Maximum Charge	Specifies the maximum charge used by the algorithm while analyzing the precursor ions. Range: 0–no maximum Default: 40

Table 8. Advanced Settings dialog box parameters (Sheet 3 of 8)

Parameter	Description
Maximum Mass (kDa) (THRASH only)	Specifies the highest mass to consider for the precursor. Range: 0–no maximum Default: 60
Minimum Fit (Xtract only)	Specifies the minimum fit parameter used by the algorithm while analyzing the precursor ions. Range: 0–100 Default: 10
Remainder Threshold (Xtract only)	Specifies the remainder of the fit that is left in the scan during analysis of the precursor ions. The Remainder Threshold option (as a percentage) determines whether a packet is further processed after an averagine pattern is subtracted. This 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. The ProSightPC application 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. Range: 0–100 Default: 20
Precursor Selection Criterion	Specifies the type of precursor mass to use for searching: • Highest Intensity: Uses the precursor mass of the most abundant ion in the precursor scan for searching. • Closest Average m/z: Uses the precursor mass that is closest to the mass-to-charge ratio (*m/z*) of the data-dependent scan for searching.

Table 8. Advanced Settings dialog box parameters (Sheet 4 of 8)

Parameter	Description
Allow Multiple Precursors	Determines whether the ProSightPC application multiplexes fragmentation data—that is, whether it scores multiple precursors in a single experiment. 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. • (Default) Selected: Multiplexes fragmentation data. • Unselected: Creates a new experiment for each precursor detected.
Relative Precursor Threshold	Specifies the threshold for selecting the precursor intensities when there are multiple precursors within the window. The ProSightPC application selects only precursors with intensities within the top *x* percent of the top precursor. Range: 1–100 Default: 10
Add Remainder Afterwards (Xtract only)	Determines whether the ProSightPC application adds the remaining intensities to the output spectrum during analysis of the precursor ions. 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 deconvolved 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 deconvolved spectrum, because unassigned remainders are added to the corresponding pattern. • Selected: Adds the remaining intensities to the output spectrum during analysis of the precursor ions. • (Default) Cleared: Does not add the remaining intensities to the output spectrum during analysis of the precursor ions.

Table 8. Advanced Settings dialog box parameters (Sheet 5 of 8)

Parameter	Description
Fragmentation Analysis Options	
Minimum S/N	Specifies the lowest signal-to-noise ratio that the algorithm considers when trying to assign neutral mass to a charged (mass-to-charge ratio [*m/z*]) species while analyzing the precursor ions. Range: 1–no maximum Default: 3.0
Minimum RL (THRASH only)	Specifies the minimum confidence level. Range: 0–1.0 Default: 0.9
Maximum Charge	Specifies the maximum charge used by the algorithm while analyzing the precursor ions. Range: 1–no maximum Default: 40
Minimum m/z Considered	Specifies the low end of the *m/z* range that the ProSightPC application analyzes. Range: 1–no maximum Default: 50 for THRASH and 60 for Xtract
Maximum Mass (kDa) (THRASH only)	Specifies the highest mass to consider for the precursor. Range: 1–no maximum Default: 60
Maximum m/z Considered	Specifies the high end of the *m/z* range that the ProSightPC application analyzes. Range:1–no maximum Default: 2000
Minimum Fit (Xtract only)	Specifies the minimum fit parameter used by the algorithm while analyzing the precursor ions. Range: 0–100 Default: 10

Table 8. Advanced Settings dialog box parameters (Sheet 6 of 8)

Parameter	Description
Remainder Threshold (Xtract only)	Specifies the remainder of the fit that is left in the scan during analysis of the precursor ions. Range: 0–100 Default: 10
Add Remainder Afterwards (Xtract only)	Determines whether the ProSightPC application adds the remaining intensities to the output spectrum during analysis of the precursor ions. • (Default) Selected: Adds the remaining intensities to the output spectrum during analysis of the precursor ions. • Cleared: Does not add the remaining intensities to the output spectrum during analysis of the precursor ions.
Remove Low m/z Interferences	Determines whether the ProSightPC application removes fragments arising from immonium ions and reagent ions from TMT and iTRAQ quantifications. • (Default) Selected: Removes fragments arising from immonium ions and reagent ions from TMT and iTRAQ quantifications. • Cleared: Does not remove fragments arising from immonium ions and reagent ions from TMT and iTRAQ quantifications.

Table 8. Advanced Settings dialog box parameters (Sheet 7 of 8)

Parameter	Description
Fragmentation Method Override	Specifies the fragmentation method used if you do not want to use the fragmentation method given in the scan header. You can select from the following fragmentation methods: • (Default) None: The High Throughput Wizard uses the fragmentation method given in the scan header. • CID • ECD • ETD • HCD • IRMPD • UVPD • EThcD For information on these methods, see "Fragmentation Methods" on page 9.
Minimum Number of Fragmentation Scans	Specifies a value to filter out low-quality spectra. Range: 1–no maximum Default: 1 In some cases, you might want to consider only precursors that have been fragmented twice or more.
Minimum Fragmentation Base Peak Intensity	Specifies a value that will filter out noise and poor-quality data during analysis of the fragment ions. Range: 1–no maximum Default: 1000
Absolute Minimum Intensity	Specifies the minimum intensity that the ProSightPC application accepts for fragmentation peaks. The application excludes deisotoped peaks below this value, so it removes low-intensity fragment ions that might be spurious. Range: 1–no maximum Default: 100

Table 8. Advanced Settings dialog box parameters (Sheet 8 of 8)

Parameter	Description
Get Top N Peaks	Specifies the number of the most intense peaks per window size that the ProSightPC application considers. This parameter works with the Window Size parameter to filter the deisotoped or decharged data. The default settings mean that the application considers only the most intense 5 peaks in a 100-Da window. Therefore, this setting therefore removes low-intensity fragment ions that might be spurious. Range: 1–no maximum Default: 5
Window Size (Da)	Specifies the size of the window containing the number of the most intense peaks that the ProSightPC application considers, in daltons. This parameter works with the Get Top *N* Peaks parameter to filter the deisotoped or decharged data. Range: 1–no maximum Default: 100

Using Repositories

The ProSightPC application 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.

For information on creating and viewing repository reports, see "Viewing the Results in a Repository Report" on page 174.

- Creating a Repository
- Editing a Repository
- Deleting a Repository
- Importing Experiments from a Repository
- Exporting Experiments to a Repository

Creating a Repository

You can create repositories for storing search results.

❖ To create a repository

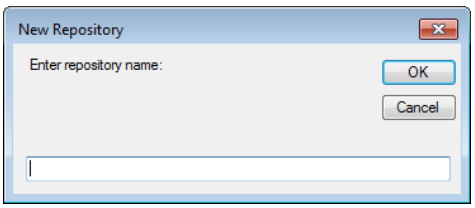
1. Choose **ProSightHT > Edit/Add Repository** to open the Edit/Add Repositories dialog box, shown in Figure 32, and click **Add New Repository**.

   –or–

2. On the Running High Throughput Logic page of the High Throughput Wizard, click **New Repository**.

   The New Repository dialog box shown in Figure 31 opens.

   Figure 31. New Repository dialog box

3. Type the name of the new repository. As indicated in the dialog box, do not use spaces in the repository name; use underscores in place of spaces.

4. Click **OK**.

   By default, the name and the identifiers of the categories in the results appear in the Categories box of the Edit/Add Repositories dialog box, shown in Figure 32.

5. (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**.

6. Click **Save**.

New Repository Dialog Box Parameters

The New Repository dialog box, shown in Figure 31, specifies a name for the repository that you created.

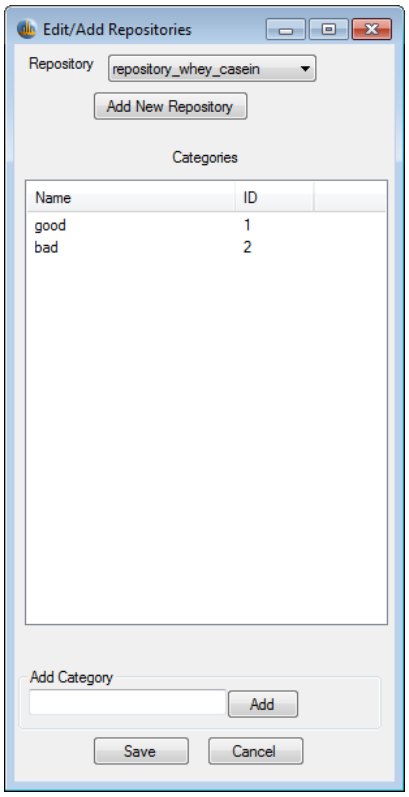
Editing a Repository

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 32.

Figure 32. 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 High Throughput Logic page of the Wizard, shown in Figure 16, and selecting Edit Repository from the Repository list.

Edit/Add Repositories Dialog Box Parameters

Table 9 lists the parameters in the Edit/Add Repositories dialog box, shown in Figure 32.

Table 9. Edit/Add Repositories dialog box (Sheet 1 of 2)

Parameter	Description
Repository	Specifies the name of the repository to edit.
Add New Repository	Opens the New Repository dialog box, shown in Figure 31, 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.

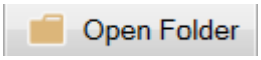
Table 9. Edit/Add Repositories dialog box (Sheet 2 of 2)

Parameter	Description
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.

Deleting a Repository

Use the Database Manager to delete a repository.

❖ To remove a repository

1. Choose **Databases > Database Manager** to open the Database Manager window.
2. Click 

in the toolbar at the top of the window.
3. In the system folder that opens, right-click the repository file (*database_name*.psph) to remove and choose **Delete**.
4. In the Delete File dialog box, click **Yes**.

IMPORTANT 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.

Importing Experiments from a Repository

You can import experiments from a repository into the ProSightPC application 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

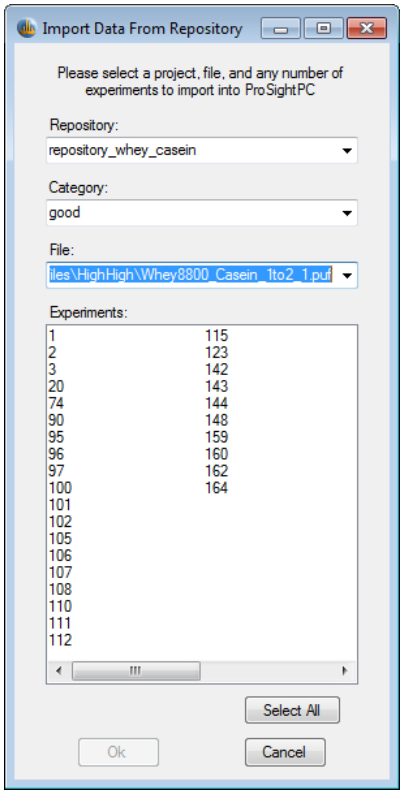
1. Choose **File > Import Data from Repository** to open the Import Data From Repository dialog box (see Figure 33).
2. In the Repository list, select the name of the repository from which to import experiments into the ProSightPC application.
3. In the Category list, select the category of results to 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 data 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.

   Verify that the Import Data from Repository dialog box resembles the example in Figure 33.

   Figure 33. Import Data from Repository dialog box



6. Click **OK**.

The experiments are now listed in the data grid.

For information on importing experiments into the ProSightPC application by using the repository report, see "Using the Repository Report To Import Experiments from a Repository into the PUF File" on page 181.

Import Data from Repository Dialog Box Parameters

Table 10 lists the parameters in the Import Data from Repository dialog box, shown in Figure 33.

Table 10. Import Data from Repository dialog box 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 data 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 data 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 the ProSightPC application 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 area of the dialog 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 where you would like to export the experiments.
4. From Category list, select the category in the repository where 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.

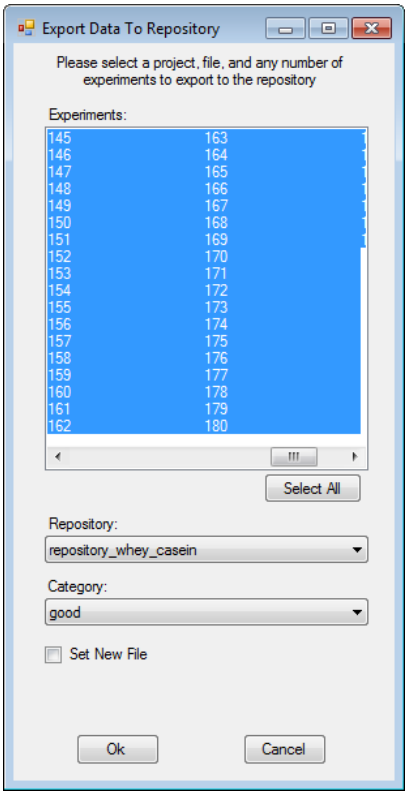
5. (Optional) Select the **Set New File** option if you want to change the PUF or raw data file on whose data the experiments were based.

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

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

   Verify that the Export Data to Repository dialog box resembles the example in Figure 34.

   Figure 34. 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 73.

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 specifying the repository name, project name, and file name.

Export Data to Repository Dialog Box Parameters

Table 11 lists the parameters in the Export Data to Repository dialog box, shown in Figure 34.

Table 11. Export Data to Repository dialog box

Parameter	Description
Experiments	Lists all the experiments in the selected raw data 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 data 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 Data Files

If you want to import a targeted raw data file as input or if you want to enter data manually into the ProSightPC application, you cannot use the High Throughput Wizard as you can with LC/MS/MS raw data files. You must use the procedures in this topic to import targeted raw data files, or use the instructions in "Entering Data Manually" on page 88 if you want to enter data manually.

In order for the ProSightPC application 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 The ProSightPC application 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]

The ProSightPC application supports three different targeted data import methods:

- Post Xtract: Takes the small file generated by the Xtract algorithm within Qual Browser and uses it as the neutral mass data. This algorithm has a better mass accuracy than THRASH but is a little slower. This option is the default.

[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 Spectrum*. **2000**, *11*, 320–332.

To use this option, see "Importing a Targeted Raw Data File with the Post Xtract Option" on page 76.

- 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. The THRASH algorithm is a little faster than the Post Xtract algorithm.

  To use this option, see "Importing a Targeted Raw Data File with the Profile Option" on page 81.

- Manual entry method: To use this option, see "Entering Data Manually" on page 88.

Table 12 summarizes the differences between the THRASH and Post Xtract methods of importing mass values.

Table 12. 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 Data 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.

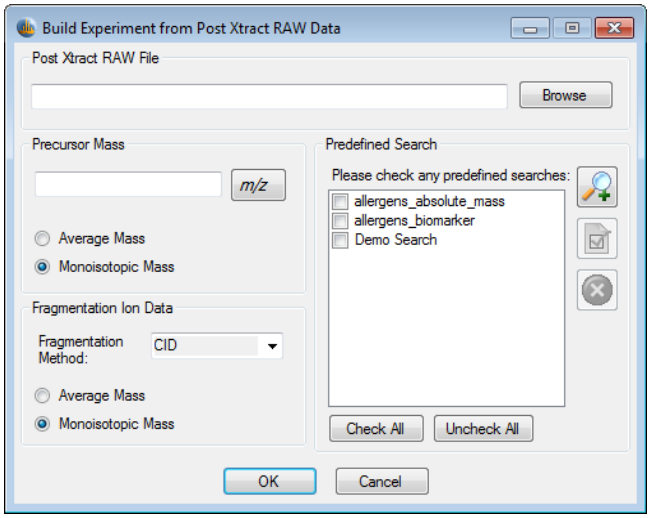
To import a targeted raw data file with the Post Xtract option, follow this procedure. To view a demonstration of this procedure, see "Demonstrating Targeted Raw File Importation with Post Xtract" on page 81.

❖ To import a targeted raw data 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 35.

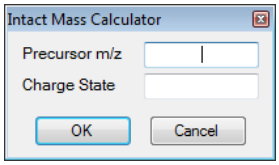
Figure 35. Build Experiment from Post Xtract RAW Data dialog box



2. In the Post Xtract RAW File box, type the path and name of the raw data 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 the ProSightPC application 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 36.

   Figure 36. 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:

   - Monoisotopic 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.
   - Average 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
   - ETD
   - HCD
   - IRMPD
   - UVPD
   - EThcD

   For information on these methods, see "Fragmentation Methods" on page 9.
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.**

Build Experiment from Post Xtract RAW Data Dialog Box Parameters

Table 13 lists the parameters in the Build Experiment from Post Xtract RAW Data dialog box, shown in Figure 35.

Table 13. Build Experiment from Post Xtract Raw Data dialog box parameters (Sheet 1 of 2)

Parameter	Description
Post Xtract RAW File	Specifies the path and name of the raw data file that you want to import. You can also click Browse to find the file.
Precursor Mass	
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 36.
Monoisotopic 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.

Table 13. Build Experiment from Post Xtract Raw Data dialog box parameters (Sheet 2 of 2)

Parameter	Description
Average 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: • (Default) CID • ECD • ETD • HCD • IRMPD • UVPD • EThcD For information on these methods, see "Fragmentation Methods" on page 9.
Monoisotopic 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.
Average Mass	Specifies that the fragment mass is the mass of the most abundant isotope of the protein, peptide, or fragment ion.
Predefined Search	
Please Check Any Predefined Searches	Displays the predefined searches set for an experiment.
Check All	Selects all predefined searches.
Uncheck All	Does not select any predefined searches.
Create New Search icon, 	Opens the New Predefined Search dialog box so that you can create a predefined search. For more information on this dialog box, see "Creating a Predefined Search" on page 105.
Edit Selected Search icon, 	Opens the Edit Predefined Search dialog box so that you can edit the parameters for the search.
Remove Selected Search icon, 	Removes the selected predefined search from the list of predefined searches to add to an experiment.

Intact Mass Calculator Dialog Box Parameters

Table 14 lists the parameters in the Intact Mass Calculator dialog box, shown in Figure 36 on page 77.

Table 14. Intact Mass Calculator dialog box 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.

Demonstrating Targeted Raw File Importation with Post Xtract

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

Importing a Targeted Raw Data File with the Profile Option

The Profile option applies the THRASH algorithm to the importation of mass values. The THRASH algorithm is an AIM operation that converts high-resolution mass spectral data from proteins or large peptides into neutral monoisotopic or average masses.

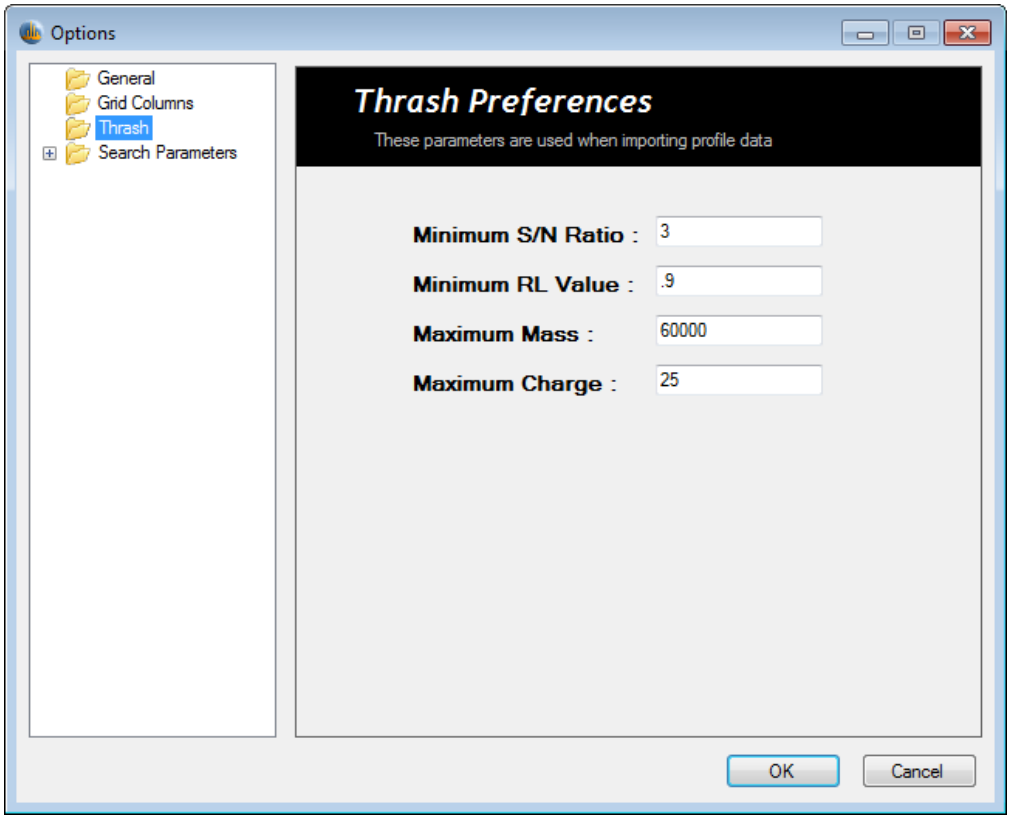
Setting THRASH Preferences

Before you import a targeted raw data file, you might want to set the default values for the THRASH algorithm. Use the Thrash Preferences page of the Options dialog box.

❖ **To set THRASH preferences**

1. Choose **Tools > Options** to open the Options dialog box.
2. From the left pane of the Options dialog box, click the **Thrash** folder to open the Thrash Preferences page, shown in Figure 37.

Figure 37. Thrash Preferences page of the Options dialog box



3. 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 data files.

   Range: >0–no maximum

   Default: 3

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

   Range: >0–1

   Default: 0.9

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

   Range: >0–no maximum

   Default: 60 000

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

   Range: >0–no maximum

   Default: 25

7. Click **OK**.

Thrash Preferences Page Parameters

Table 15 lists the parameters on the Thrash Preferences page of the Options dialog box.

Table 15. Thrash Preferences page 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 data files. Range: >0–no maximum Default: 3
Minimum RL Value	Specifies the minimum confidence level. Range: >0–1 Default: 0.9
Maximum Mass	Specifies the cutoff point for the THRASH algorithm when searching for masses. Range: >0–no maximum Default: 60 000
Maximum Charge	Specifies the maximum charge to be used by the THRASH algorithm. Range: >0–no maximum Default: 25

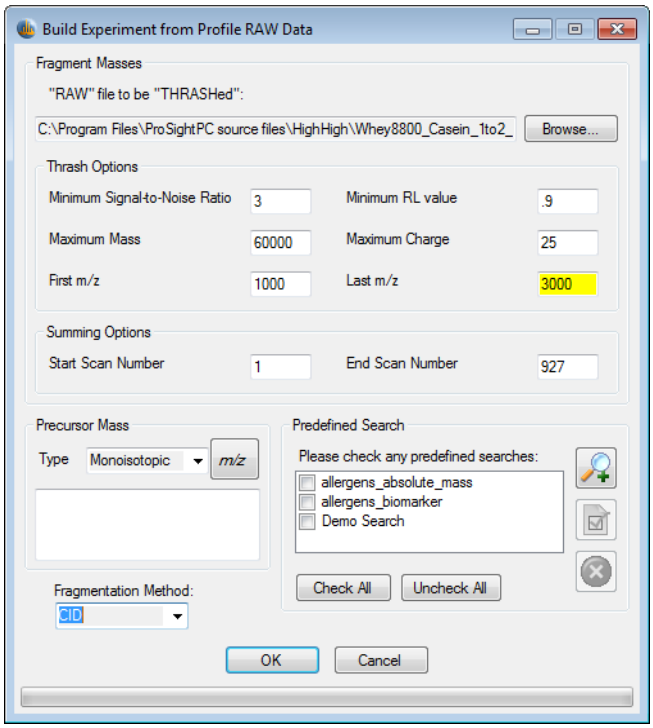
Importing the Targeted Raw Data File

❖ **To import a targeted raw data 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 38.

Figure 38. Build Experiment from Profile RAW Data dialog box



2. To specify a raw data file to import, type the full path name in the box labeled raw data 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 data files.

   Range: 0–no maximum

   Default: 3

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

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

   Range: 1–no maximum

   Default: 60 000

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

   Range: 1–no maximum

   Default: Taken from the raw data file

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

   Range: 0–1

   Default: 0.9

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

   Range: 1–no maximum

   Default: 25

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

   Range: 1–no maximum

   Default: Taken from the raw data file

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.

   Defaults: Taken from the raw data file

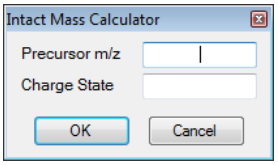
10. In the Type list in the Precursor Mass area, select the mass type:

    - 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 the ProSightPC application 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 39.

    Figure 39. 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:
    - (Default) CID
    - HCD
    - ECD
    - ETD
    - IRMPD
    - UVPD
    - EThcD

    For information on these methods, see "Fragmentation Methods" on page 9.
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.

Build Experiment from Profile RAW Data Dialog Box Parameters

Table 16 lists the parameters in the Build Experiment from Profile RAW Data dialog box, shown in Figure 38.

Table 16. Build Experiment from Profile RAW Data dialog box parameters (Sheet 1 of 3)

Parameter	Description
RAW File to Be THRASHed	Specifies the name of the raw data file to import.
THRASH Options	
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 data files. Range: 1–no maximum Default: 3
Maximum Mass	Specifies the cutoff point for the THRASH algorithm when it searches for masses. Range: 1–no maximum Default: 60 000

Table 16. Build Experiment from Profile RAW Data dialog box parameters (Sheet 2 of 3)

Parameter	Description
First m/z	Specifies the lowest mass-to-charge ratio (*m/z*) value considered. Range: 1–no maximum Default: Taken from the raw data file
Minimum RL Value	Specifies the minimum confidence level. Range: 0–1 Default: .9
Maximum Charge	Specifies the maximum charge to be used by the THRASH algorithm. Range: 1–no maximum Default: 25
Last m/z	Specifies the highest mass-to-charge ratio (*m/z*) value considered. Range: 1–no maximum Default: Taken from the raw data file
Summing Options	
Start Scan Number	Specifies the first scan number scanned. Default: Taken from the raw data file
End Scan Number	Specifies the last scan number scanned. Default: Taken from the raw data file
Precursor Mass	
Type	Specifies the type of precursor ion mass: • (Default) 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.
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 36.

Table 16. Build Experiment from Profile RAW Data dialog box parameters (Sheet 3 of 3)

Parameter	Description
Fragmentation Method	Specifies one of the following fragmentation methods: • (Default) CID • HCD • ECD • ETD • IRMPD • UVPD • EThcD For information on these methods, see "Fragmentation Methods" on page 9.
Predefined Search	
Please Check Any Predefined Searches	Displays the predefined searches set for an experiment.
Check All	Selects all predefined searches.
Uncheck All	Does not select any predefined searches.
Create New Search icon, 	Opens the New Predefined Search dialog box so that you can create a predefined search. For more information on this dialog box, see "Creating a Predefined Search" on page 105.
Edit Selected Search icon, 	Opens the Edit Predefined Search dialog box so that you can edit the parameters for the search.
Remove Selected Search icon, 	Removes the selected predefined search from the list of predefined searches to add to an experiment.

Entering Data Manually

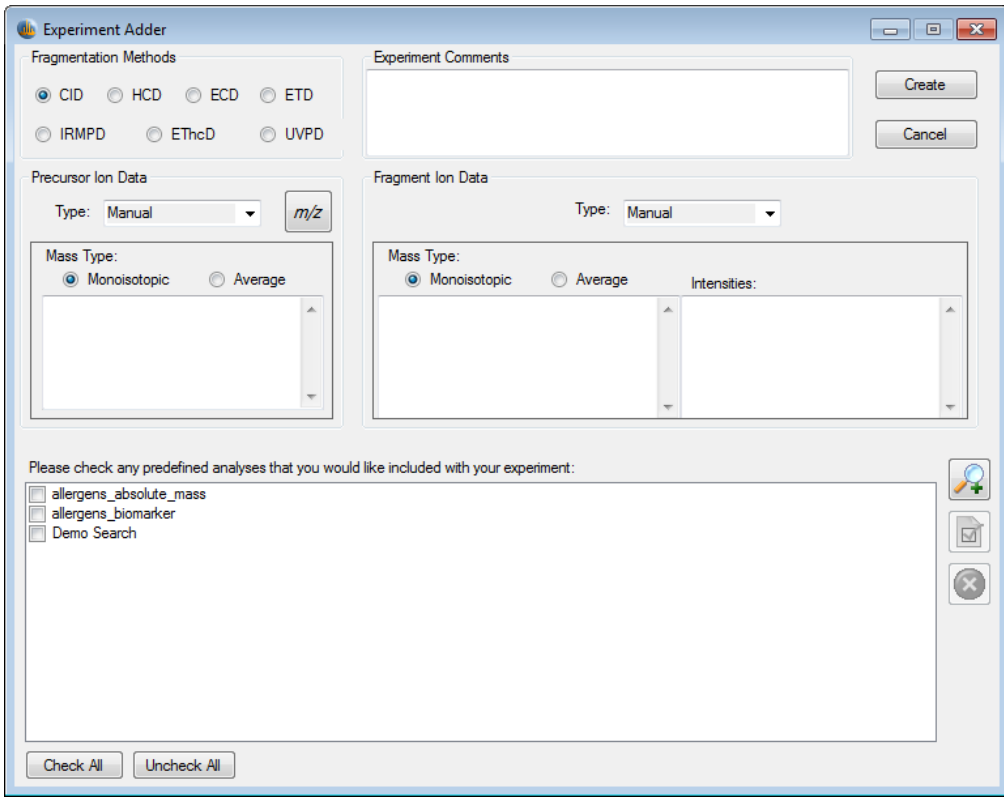
As a third option, you can import MS/MS experiment data by manually entering the data from the ProSightPC application through the Tools menu. The data is then displayed in the data grid. You can also manually delete experiments.

❖ 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 40.

Figure 40. Experiment Adder dialog box



2. In the Fragmentation Methods area, select one of the following fragmentation methods:
   - (Default) CID
   - HCD
   - ECD
   - ETD
   - IRMPD
   - EThcD
   - UVPD

   For information on these methods, see Fragmentation Methods.
3. In the Precursor Ion Data area, select the method of inputting the precursor ion data.
   a. In the Type list, specify the method of inputting the precursor ion data.
      - (Default) Manual: Inputs the precursor ion data.
      - Upload: Loads the precursor ion data from an ASCII text file or files.

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:

- (Default) 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 the ProSightPC application 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 36.

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, specify the method of inputting the fragment ion data.

- (Default) Manual: Inputs the precursor ion data.
- Upload: Loads the precursor ion data from an ASCII text file or files.

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:

- (Default) 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.
- Intensities: Specifies the intensity of the fragment mass.

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 intensity of the fragment mass.

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 Please Check Any Predefined Analyses That You Would Like Included with Your Experiment box, select any predefined searches to add. Click **Check All** to add all listed predefined searches. Click **Uncheck All** to clear all listed predefined searches.
7. Click **Create**.

The ProSightPC application 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

Table 17 lists the parameters in the Experiment Adder dialog box, shown in Figure 40.

Table 17. Experiment Adder dialog box parameters (Sheet 1 of 2)

Parameter	Description
Fragmentation Methods	Specifies one of the following fragmentation methods: • CID • HCD • ECD • ETD • IRMPD • EThcD • UVPD For information on these methods, see "Fragmentation Methods" on page 9.
Precursor Ion Data	
Type	Specifies the method of inputting the precursor ion data. You can select Manual or Upload from the Type list. • (Default) Manual: Inputs the precursor ion data. • Upload: Loads the precursor ion data from an ASCII text file or files.
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 36.
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: • (Default) 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.
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.

Table 17. Experiment Adder dialog box parameters (Sheet 2 of 2)

Parameter	Description
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. • (Default) Manual: Inputs the precursor ion data. • Upload: Loads the precursor ion data from an ASCII text file or files.
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: • (Default) 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. • Intensities: Specifies the intensity of the fragment mass.
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.
Please Check Any Predefined Analyses That You Would Like Included with Your Experiment	Selects any predefined searches to add to an experiment.
Create New Search icon, 	Opens the New Predefined Search dialog box so that you can create a new predefined search. For more information on this dialog box, see "Creating a Predefined Search" on page 105.
Edit Selected Search icon, 	Opens the Edit Predefined Search dialog box so that you can edit the parameters for the search.
Remove Selected Search icon, 	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 the ProSightPC application is to import experiments from a repository. See "Importing Experiments from a Repository" on page 71 and "Using the Repository Report To Import Experiments from a Repository into the PUF File" on page 181 for instructions on this procedure. For more information on handling experiments, see "Working with Experiments" on page 95.

Searching the Proteome Warehouse for Matches

After you import your data, you might want to search the proteome warehouse for matches. The ProSightPC application supports six different search modes. Each search mode represents a specific method used to query a proteome database within the proteome warehouse. You can add a predefined search by following the instructions in "Searching Databases" on page 103.

3

Working with Experiments

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

Contents

Experiments in PUF Files

The ProSightPC application operates on a single PUF file that contains experiments. When you open the PUF file, the application loads 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

For high-throughput experiments, the High Throughput Wizard creates PUF files. For targeted experiments, you can create a PUF file by creating an empty file and importing data into it or by using the Experiment Manager.

❖ To create a PUF file by importing data into an empty 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.

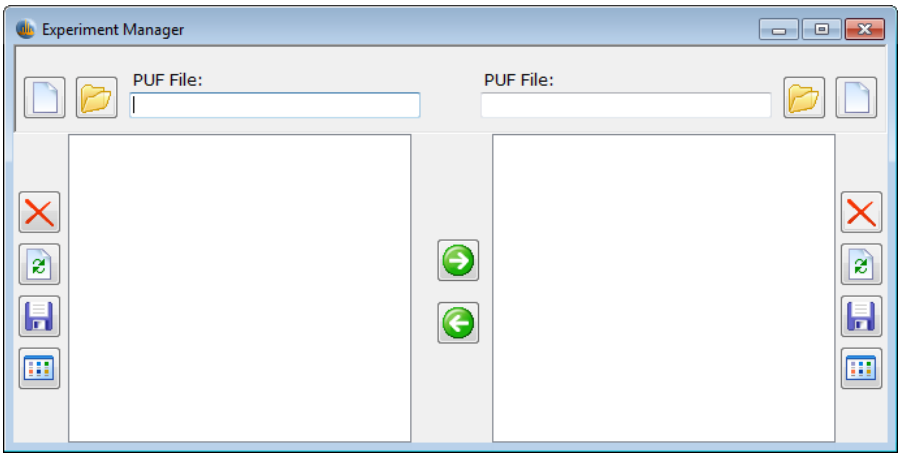
2. Add experiments or import data to the PUF file. See "Adding Experiments to PUF Files" on page 98.

❖ To create a PUF file by using the Experiment Manager

1. Choose **Tools > Experiment Manager**.

   The Experiment Manager opens, as shown in Figure 41. Usually you use this dialog box to copy experiments from one PUF file to another (see "Copying Experiments from One PUF File to Another" on page 98).

   Figure 41. Experiment Manager

   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.

2. At the top left in the Experiment Manager, 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.

3. In the dialog box that opens, select the applicable directory and type the name of the new PUF file.

4. Click **Save**.

Opening an Existing PUF File

You can open an existing PUF file by using a command on the File menu or by using the Experiment Manager.

Only one PUF file can be open at a time.

❖ To open an existing PUF file from the File menu

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, and 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 appear in the data grid.

❖ To open an existing PUF file from the Experiment Manager

1. Choose **Tools > Experiment Manager**.

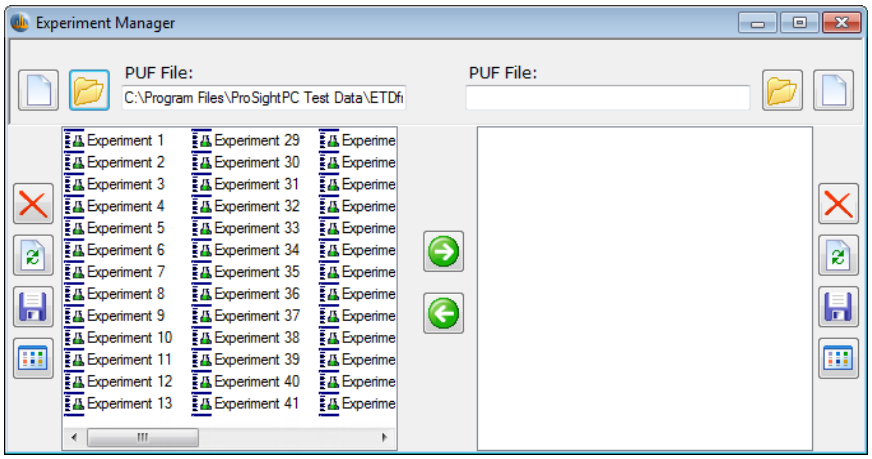
2. In the Experiment Manager, shown in Figure 41, click the **Open Existing PUF File** icon,  , on the left to open an existing source PUF file. Click the same icon on the right to open an existing destination PUF file.

3. Select the PUF file from those listed.

4. Click **Open**.

   The experiments in the PUF file are now displayed in the appropriate pane of the Experiment Manager, as shown in Figure 42.

Figure 42. Experiments listed in Experiment Manager



Adding Experiments to 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 data files with the Post Xtract option: See "Importing a Targeted Raw Data File with the Post Xtract Option" on page 76.
- Import raw data files with the Profile option: See "Importing a Targeted Raw Data File with the Profile Option" on page 81.
- Use the Experiment Adder: See "Entering Data Manually" on page 88.
- Import data from a repository: See "Importing Experiments from a Repository" on page 71 and "Using the Repository Report To Import Experiments from a Repository into the PUF File" on page 181.

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.

❖ **To copy an experiment from one PUF file to another**

1. Choose **Tools > Experiment Manager**.
2. In the source (left) pane of the Experiment Manager, shown in Figure 41 on page 96, select an experiment to be copied.
3. Drag the experiment from the source (left) pane to the destination (right) pane, or click the green arrow ( ) to send a copy of the experiment to the destination (right) pane.

4. Click the **Save This PUF** icon,  , on the source (left) side of the dialog box to save the source PUF file. Click the same icon on the destination (right) side to save the destination PUF file.

   If two or more experiments share the same experiment identification number 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.

5. Click **OK** to confirm**.**

You can also copy experiments from the destination (right) pane to the source (left) pane.

Removing Experiments from PUF Files

You can remove experiments from a PUF file by using a shortcut menu or by using the Experiment Manager.

❖ **To remove an experiment from the PUF file and the data grid by using the shortcut menu**

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,  .

❖ **To delete experiments from a PUF file by using the Experiment Manager**

1. Choose **Tools > Experiment Manager**.
2. In the Experiment Manager, shown in Figure 41 on page 96, select the experiments that you want to delete.
3. 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.

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**

1. Choose **Tools > Experiment Manager**.
2. In the Experiment Manager, shown in Figure 41 on page 96, 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**

1. Choose **Tools > Experiment Manager**.
2. In the Experiment Manager, shown in Figure 41 on page 96, 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.

Changing the Experiment Display

You can change how the experiments in the PUF files are displayed in the Experiment Manager.

❖ **To change the experiment display**

1. Choose **Tools > Experiment Manager**.
2. In the Experiment Manager, shown in Figure 41 on page 96, 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.
3. 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.

Deleting PUF Files

You cannot delete a PUF file from the ProSightPC application. In Windows, you can delete it as you would a regular file by right-clicking on the file name and choosing Delete from the shortcut menu.

Experiment Manager Parameters

Table 18 lists the parameters in the Experiment Manager, shown in Figure 41 on page 96.

Table 18. Experiment Manager parameters (Sheet 1 of 2)

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).

Table 18. Experiment Manager parameters (Sheet 2 of 2)

Parameter	Description
	Opens a popup menu so that you can change how the experiments are displayed in the Experiment Manager. • 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.
	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.

4

Searching Databases

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

Contents

Search Types

The ProSightPC application 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 the ProSightPC application 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.

Performing Searches

When you perform searches with the ProSightPC application, start with a more specific search rather than a more general search. The database search takes considerably longer with increasing search complexity, so identifying as many spectra as possible with a quick, simple search saves a lot of search time overall.

There is also another reason for starting with a more specific search. The e value for a result is inversely proportional to the size of the searched database. For the exact same spectrum, a search against a smaller database produces a better identification than a search against a larger one, assuming that the correct protein identification resides in both databases.

- Performing Predefined Searches
- Performing Searches in Delta-m Mode
- Performing Searches in Batch Mode

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. 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
- Adding Predefined Searches to an Experiment
- Editing a Predefined Search
- Running a Predefined Search
- Canceling a Predefined Search

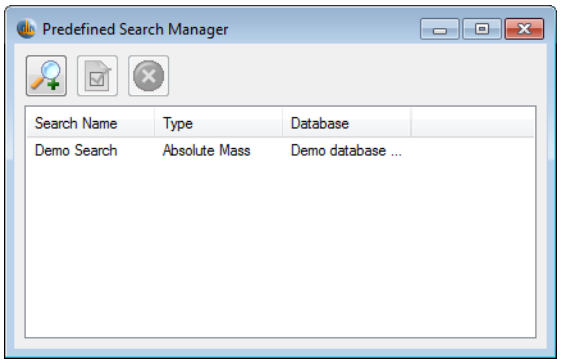
- Removing a Predefined Search
- Removing Search Results from a Search
- Removing an Experiment from the Data Grid

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 43.

   Figure 43. Predefined Search Manager dialog box

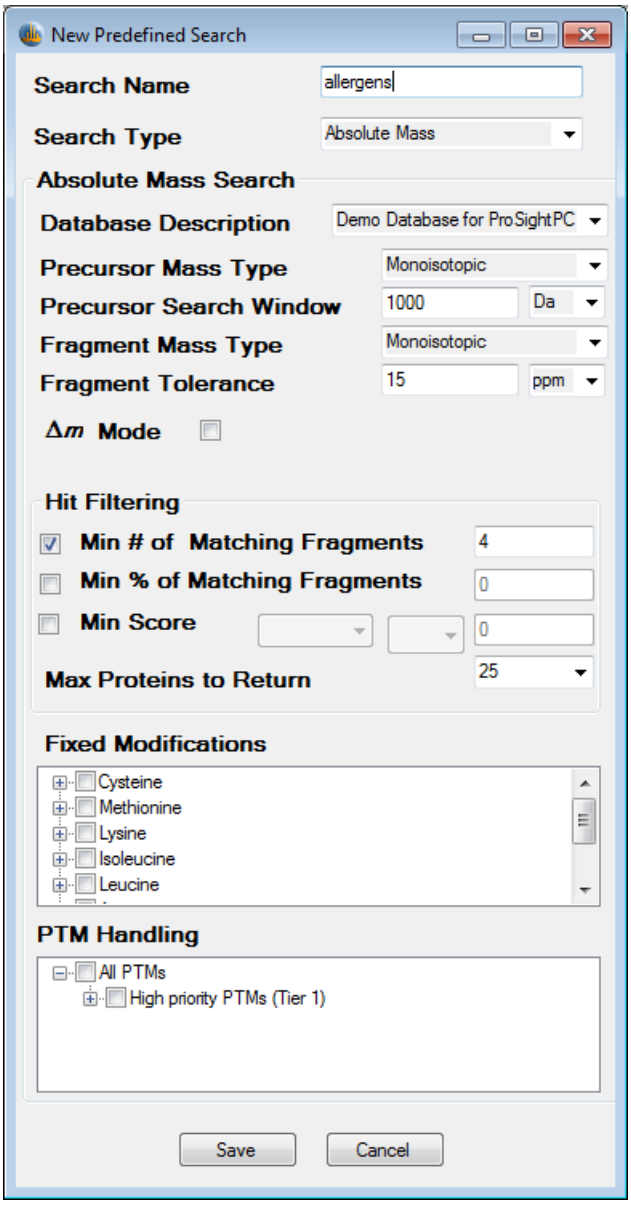
   

   All existing predefined searches appear in the Predefined Search Manager dialog box. In the example in Figure 43, a search of the demonstration database included in the installation of the ProSightPC software is available.

2. Click the **Create New Search** icon,  , in 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 44.

Figure 44. New Predefined Search dialog box



3. In the Search Name box, type a name for the new predefined search.
4. In the Search Type list, select the search type, and follow the procedure for your selection:
   - Absolute Mass: See "Searching for Absolute Mass" on page 122.
   - BioMarker: See "Searching for Biomarkers" on page 133.
   - Sequence Tag: See "Searching for Sequence Tags" on page 143.
   - Single Protein: See "Searching for Single Proteins" on page 149.

- Gene-Restricted Absolute Mass: See "Searching for Gene-Restricted Absolute Masses" on page 153.
- Gene-Restricted BioMarker: See "Searching for Gene-Restricted Biomarkers" on page 159.

Predefined Search Manager Dialog Box Parameters

Table 19 lists the parameters in the Predefined Search Manager dialog box, shown in Figure 43 on page 105.

Table 19. Predefined Search Manager dialog box parameters

Parameter	Description
Create New Search icon, 	Opens the New Predefined Search dialog box so that you can create a predefined search. For more information on this dialog box, see "Creating a Predefined Search" on page 105.
Edit Selected Search icon, 	Opens the Edit Predefined Search dialog box so that you can edit the parameters for the search. For more information on this dialog box, see "Editing a Predefined Search" on page 109.
Remove Selected Search icon, 	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, single-protein, gene-restricted absolute mass, and gene-restricted biomarker 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, shown in Figure 44 on page 106, depend on the type of search that you select in the Search Type list:

- Absolute mass search: See "New Predefined Search Dialog Box Parameters for Absolute Mass" on page 126.
- Biomarker search: See "New Predefined Search Dialog Box Parameters for Biomarkers" on page 137.
- Sequence tag search: See "New Predefined Search Dialog Box Parameters for Sequence Tags" on page 145.
- Single-protein search: See "Searching for Single Proteins" on page 189.

- Gene-restricted absolute mass search: See "Searching for Gene-Restricted Absolute Masses" on page 153.
- Gene-restricted biomarker search: See "Searching for Gene-Restricted Biomarkers" on page 159.

Adding Predefined Searches to an Experiment

You can append a single predefined search or multiple predefined searches to an experiment.

❖ To add a single predefined search to experiments

In the data grid, right-click the appropriate experiment and choose **Append Predefined Search > *search_name***.

To add a single predefined search to all the experiments in the PUF file, click the data grid and press CTRL+A.

The experiment that the search has been appended to appears in the data grid with the same experiment number in the Exp ID column and a different number in the Search ID column.

–or–

Follow the next procedure, To add multiple predefined searches to experiments, and select the appropriate search.

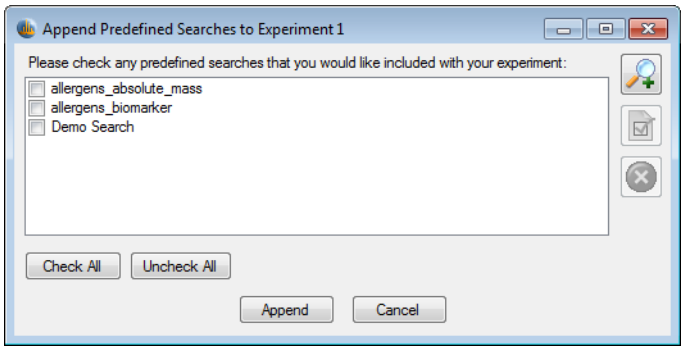
❖ To add multiple predefined searches to experiments

1. Right-click the experiment in the data grid and choose **Append Predefined Searches**. You can also choose **Experiment Tools > Append Predefined Search**.

   To add a multiple predefined searches to all the experiments in the PUF file, click the data grid and press CTRL+A.

   The Append Predefined Searches to Experiment *X* dialog box opens, as shown in Figure 45.

   Figure 45. Append Predefined Searches to Experiment *X* dialog box

   The default predefined search is Demo Search, which searches the demonstration database included in the installation of the ProSightPC software.

2. Select the predefined searches to append to the experiment and click **Append**. To select all of the searches listed, click **Check All**.

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

   Tip To process several predefined searches automatically, see "Performing Searches in Batch Mode" on page 115.

Append Predefined Searches to Experiment *X* Dialog Box Parameters

Table 20 lists the icons and parameters in the Append Predefined Searches to Experiment *X* dialog box, shown in Figure 45 on page 108.

Table 20. Append Predefined Searches to Experiment *X* dialog box 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. • Demo Search: Searches the demonstration database included in the installation of the ProSightPC software.
Check All	Selects all the predefined searches listed to add to an experiment.
Uncheck All	Clears all the predefined searches.
Create New Search icon, 	Opens the New Predefined Search dialog box so that you can create a predefined search. For more information on this dialog box, see "Creating a Predefined Search" on page 105.
Edit Selected Search icon, 	Opens the Edit Predefined Search dialog box so that you can edit the parameters for the search. For more information on this dialog box, see "Editing a Predefined Search" on page 109.
Remove Selected Search icon, 	Removes the selected predefined search from the list of predefined searches to add to an experiment.

Editing a Predefined Search

❖ To edit a predefined search

1. Open the Edit Predefined Search dialog box by doing one of the following:
   - Follow this procedure:
     i. Choose **Tools > Manage Predefined Searches**.
     ii. In the Predefined Search Manager dialog box, shown in Figure 43 on page 105, click the name of the appropriate predefined search.

iii. Click the **Edit** icon,  , in the Predefined Search Manager dialog box, or right-click the search name and choose **Edit** from the shortcut menu.

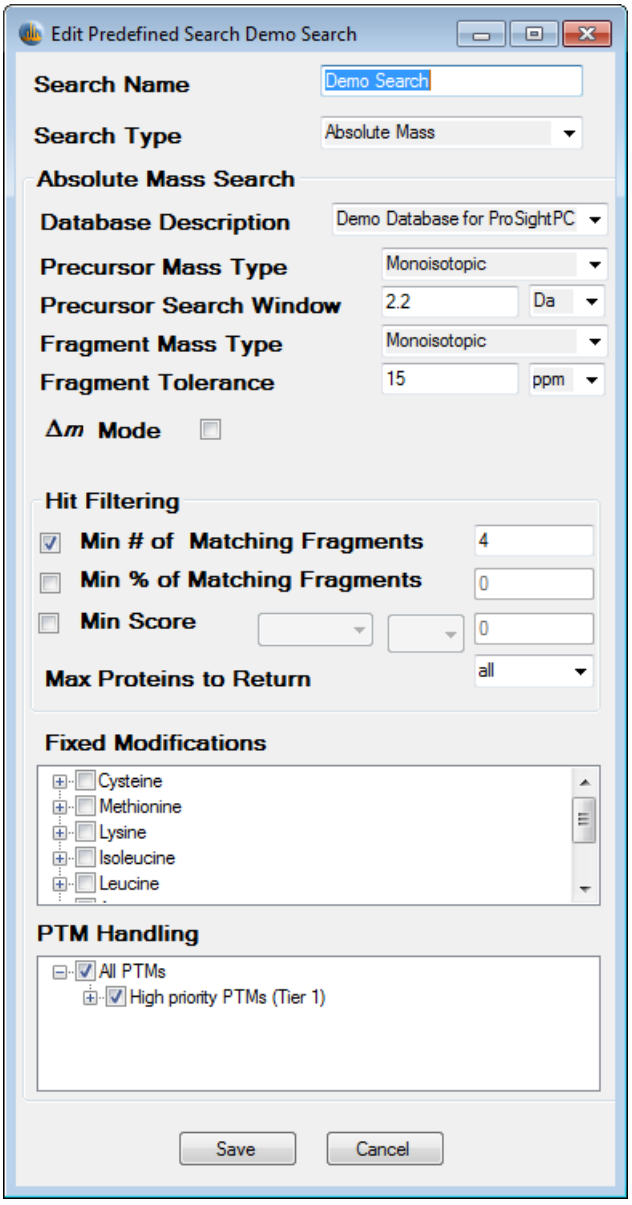
The Edit Predefined Search dialog box opens, as shown in Figure 46.

–or–

- Right-click an experiment in the data grid and choose **Edit Search *x***.

A dialog box similar to the one shown in Figure 46 appears, although it has a slightly different header. Instead of Edit Predefined Search *search_name*, the header reads Edit Predefined Search *x*.

Figure 46. Edit Predefined Search dialog box



2. Edit the parameters in the dialog box. See the following topics for more information.
    - Absolute mass search: See "Searching for Absolute Mass" on page 116.
    - Biomarker search: See "Searching for Biomarkers" on page 128.
    - Sequence tag search: See "Searching for Sequence Tags" on page 139.
    - Single-protein search: See "Searching for Single Proteins" on page 189.
    - Gene-restricted absolute mass search: See "Searching for Gene-Restricted Absolute Masses" on page 153.
    - Gene-restricted biomarker mass search: See "Searching for Gene-Restricted Biomarkers" on page 159.
3. 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: See "New Predefined Search Dialog Box Parameters for Absolute Mass" on page 126.
- BioMarker: See "New Predefined Search Dialog Box Parameters for Biomarkers" on page 137.
- Sequence Tag: See "New Predefined Search Dialog Box Parameters for Sequence Tags" on page 145.
- Single Protein: See "Searching for Single Proteins" on page 149.
- Gene-Restricted Absolute Mass: See "Searching for Gene-Restricted Absolute Masses" on page 153.
- Gene-Restricted BioMarker: See "Searching for Gene-Restricted Biomarkers" on page 159.

Running a Predefined Search

You can run a predefined search by using either of the following procedures.

❖ To run a predefined search with the Run Search command

1. Select the applicable experiments in the data grid.
    - To help you sort entries in the data grid, you can click the title row of the column to sort entries from lowest to highest value or highest to lowest.

- To select contiguous experiment names, click the name of the first experiment, hold down the SHIFT key, and click the last experiment name that you want to select.
- To select noncontiguous experiment names, click the name of the first experiment, hold down the CTRL key, and click each separate experiment name.

2. Right-click and choose **Append Predefined Searches**.
3. In the Append Predefined Searches to Experiment *X* dialog box, select the predefined searches that you would like to include with your experiment, and click **Append**.
4. Reselect the experiments that you selected in step 1.
5. Right-click and choose **Run Search *x***, where *x* is the number that appears in the Search ID column.

   The Run Search *x* command is only available when the Pending Search command displays "yes" for the appropriate search.

❖ **To run a predefined search with the Batch Run command**

1. Perform step 1 through step 3 in the previous procedure.
2. Choose **Tools > Batch Run** or click the **Batch Run** icon,  .

   The ProSightPC application queues and runs each pending search in turn. You can use the Batch Run command or icon to run a single search.

Canceling a Predefined Search

You can cancel a predefined search that has started running.

❖ **To cancel a search in the job queue**

- If a search is running, right-click the search in the job queue, and choose **Abort** or click the **Abort Running Job** icon,  , in the toolbar.

  The search ends and the status changes to Failed.
- If you want to cancel all running searches, you can click the **Abort All Jobs** icon,  .

Removing a Predefined Search

You can remove a predefined search by using a shortcut command in the data grid or by using the Predefined Search Manager.

❖ **To remove a predefined search from an experiment by using a data grid shortcut command**

In the data grid, right-click the search number of an experiment and choose **Remove Search *x***, where *x* is the name of the search that you want to remove.

❖ **To remove a predefined search from an experiment by using the Predefined Search Manager**

1. Choose **Tools > Manage Predefined Searches**.
2. In the Predefined Search Manager dialog box, select a predefined search name from the list (see Figure 43 on page 105).
3. Click the **Remove Selected Search** icon,  , in the Predefined Search Manager dialog box, or right-click the search name and choose **Remove** from the shortcut menu.
4. Confirm the removal by clicking **Yes** or **No**.

   You can return to the Predefined Search Manager without removing the selected search by clicking No.

Removing Search Results from a Search

You might want to remove search results from a search that has already been run if you want to rerun the search with different parameters.

❖ **To remove search results from a search**

1. Select the experiment in the data grid.
2. Right-click and choose **Remove Results**.

   This command is only available when search results are present—that is, when the Pending Search column displays “no” for the appropriate search.
3. In the confirmation box, click **Yes**.

Removing an Experiment from the Data Grid

❖ **To remove an experiment**

1. Right-click an experiment in the data grid, and choose **Remove Experiment *x***.
2. In the Confirm Delete dialog box, click **Yes**.

   The ProSightPC application deletes the experiment from the data grid.

Performing Searches in Delta-m Mode

Delta-m (Δm) mode is a technique for identifying proteoforms 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-m (Δm) mode, the ProSightPC application 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-m applied to the N terminal to the observed fragment ion mass list
- The theoretical fragment ion masses derived from the sequence and the delta-m applied to the C terminal to the observed fragment mass list

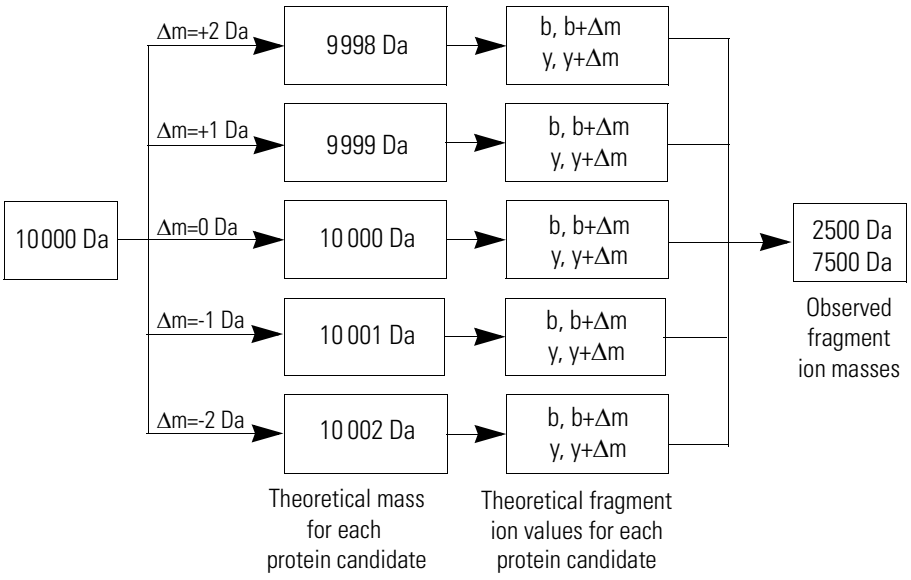
A delta-m search takes approximately two times longer than the same search without delta-m (Δm) mode.

By carefully observing the pattern of fragments with and without delta-m (Δm) mode, you can frequently locate the delta-m.

For example, if a particular result returns with the two smallest N-terminal fragments matching 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 47, the ProSightPC application 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-m). Next, it doubles the theoretical fragment ion list for each protein for each b/y or c/z ion. The ProSightPC application checks both the original fragment mass and the modified fragment mass (plus the delta) against the observed fragment ion mass list. As a result, the ProSightPC application returns any observed fragment ions having the same mass shift as the precursor protein as positive matches.

Figure 47. Schematic of Δm mode



Performing Searches in Batch Mode

The ProSightPC application offers two ways of performing multiple searches at the same time.

- With batch processing, you can queue and run a large number of searches over any number of experiments in a single action. The ProSightPC application runs any search in the grid with pending searches. Use batch processing when you have many pending searches in a PUF file and you would like to run all of them.
- The Run Searches command runs any searches that are selected (highlighted) in the data grid. If a search is not selected, the ProSightPC application does not run it.

❖ To perform searches in batch mode

1. Select the applicable experiments in the data grid.
   - To help you sort entries in the data grid, you can click the title row of the column to sort entries from lowest to highest value or highest to lowest.
   - To select contiguous experiment names, click the name of the first experiment, hold down the SHIFT key, and click the last experiment name that you want to select.
   - To select noncontiguous experiment names, click the name of the first experiment, hold down the CTRL key, and click each separate experiment name.
2. Right-click and choose **Append Predefined Searches** from the shortcut menu.
3. In the Append Predefined Searches to Experiment *X* dialog box, select the predefined searches that you would like to include with your experiment, and click **Append**.
4. Choose **Tools > Batch Run**, or click the **Batch Run** icon,  .

   The ProSightPC application queues and runs each pending search in turn.

You can also use the Batch Run command or icon to quickly run a single search.

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

❖ To run multiple searches

1. Select the applicable pending searches in the data grid.
   - To select contiguous experiment names, click the name of the first experiment, hold down the SHIFT key, and click the last experiment name that you want to select.
   - To select noncontiguous experiment names, click the name of the first experiment, hold down the CTRL key, and click each separate experiment name.
2. Right-click and choose **Append Predefined Searches**.

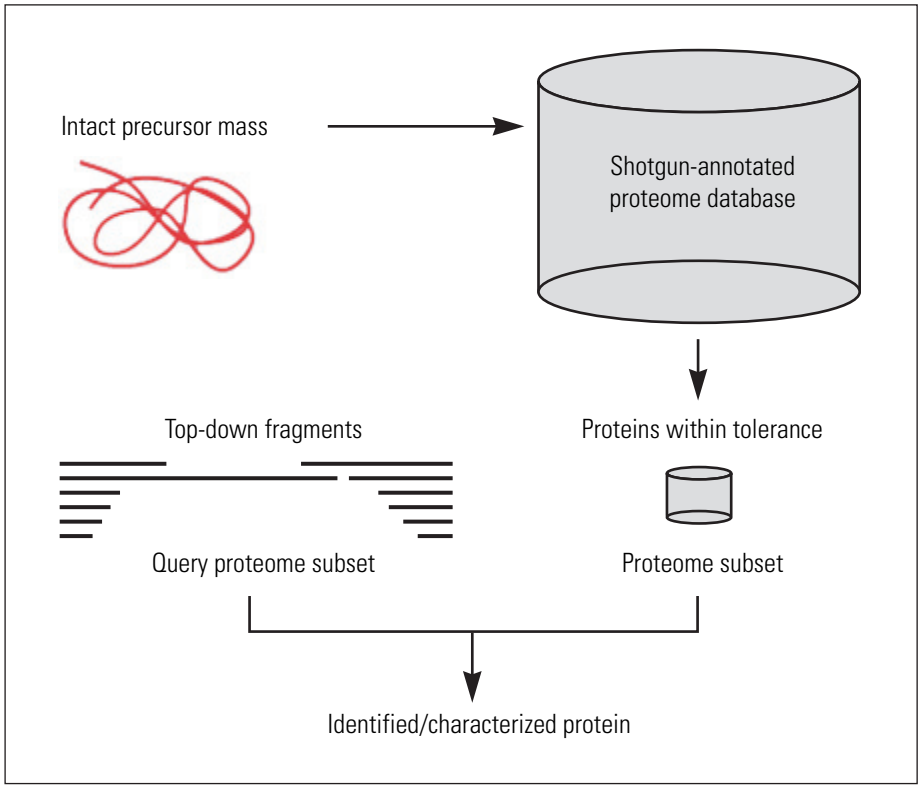
3. In the Append Predefined Searches to Experiment *X* dialog box, select the predefined searches that you would like to include with your experiment, and click **Append**.
4. In the data grid, reselect the resulting pending searches (that is, the searches with "yes" in the Pending Search column).
5. Right-click and choose **Run Searches**.

Searching for Absolute Mass

The absolute mass search matches MS/MS data against all intact proteoforms of proteins in a database. It is the defining search mode for top-down proteomics. Absolute mass searches use the precursor mass to generate a subset of the proteome database to query.

For each proteoform with a theoretical precursor mass within the window of the observed precursor ion mass, plus or minus the defined search tolerance, the absolute mass search compares all theoretical fragments and masses to observed fragment ion masses. The ProSightPC application determines the number of observed fragment ions matching the fragment tolerance and uses this value to score the identification. Figure 48 shows this methodology.

Figure 48. Absolute mass searches



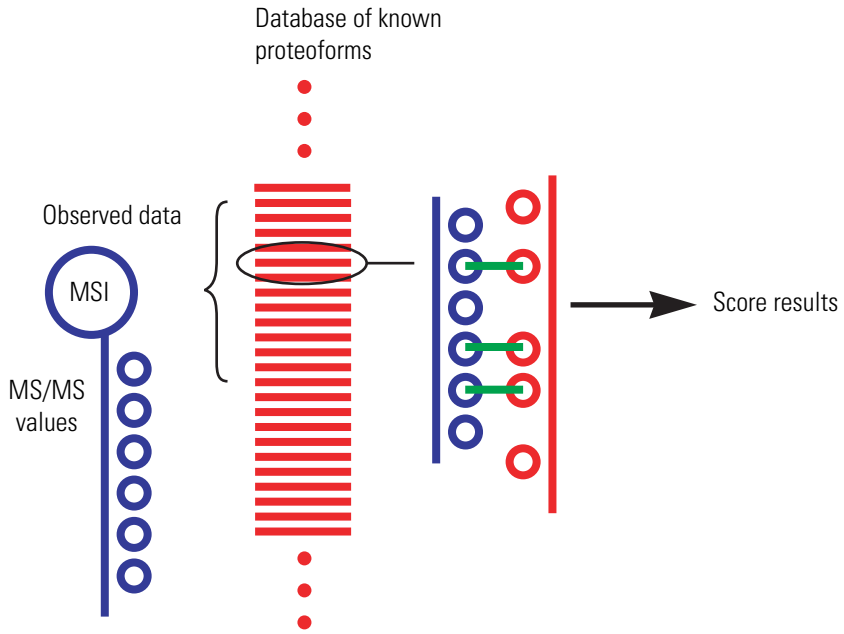
Although the ProSightPC application queries each proteoform with a theoretical precursor mass in the window, it displays only those proteoforms that meet the user-defined filtering of search results. These filters include the minimum number of matching fragments, minimum percentage of matching fragments, or minimum score.

In summary, the ProSightPC application performs the following steps in a search for absolute mass:

1. In absolute mass searches, it queries all proteoforms in a user-defined window of the observed intact mass.
2. The ProSightPC application 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 matches a theoretical 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 196.

Figure 49 shows this process graphically.

Figure 49. 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-m (Δm) mode (explained in "Performing Searches in Delta-m Mode" on page 113) 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-m (Δm) mode. Activating delta-m (Δm) mode increases the likelihood that the search will identify proteins with unknown modifications. However, this mode takes approximately two times longer than the corresponding absolute mass search.

Follow these sets of instructions to specify preferences and perform the search:

- Setting Absolute Mass Search Preferences
- Searching for Absolute Mass

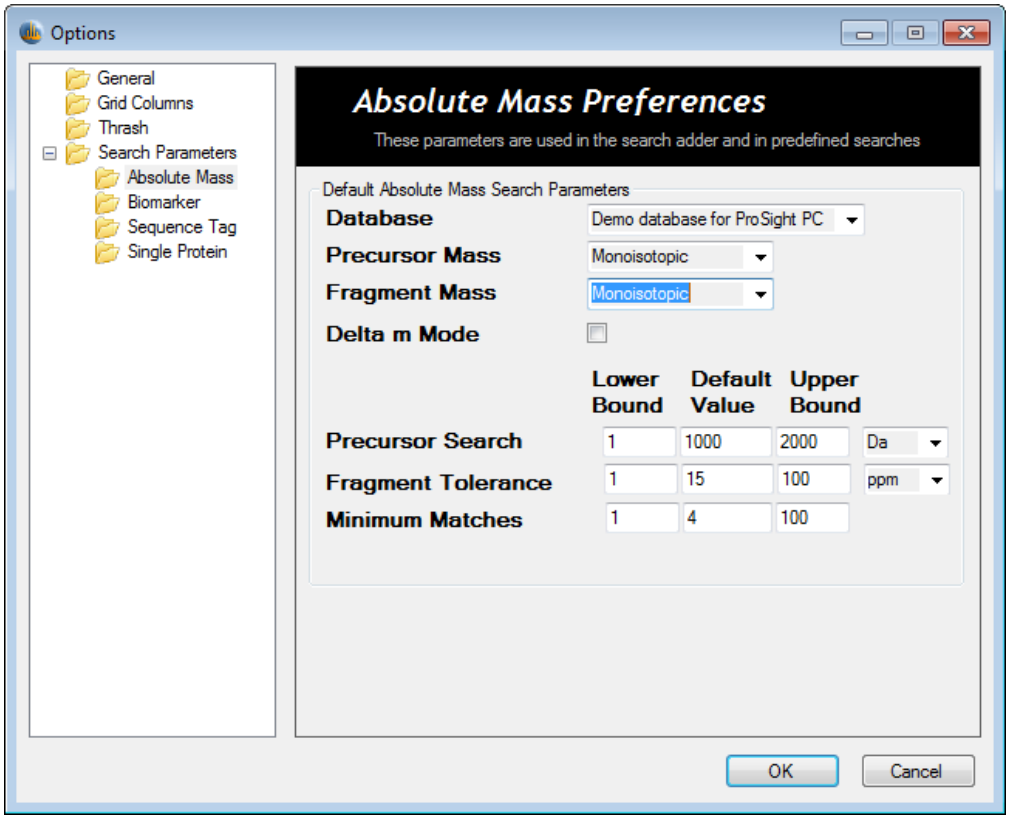
Setting Absolute Mass Search Preferences

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 116.

❖ To set absolute mass search preferences

1. Choose **Tools > Options** to open the Options dialog box.
2. From the left pane of the Options dialog box, click the **Search Parameters** folder and then click the **Absolute Mass** folder to open the Absolute Mass Preferences page, shown in Figure 50.

Figure 50. Absolute Mass Preferences page of the Options dialog box



3. In the Database list, select the name of the database to search.
4. In the Precursor Mass 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.
5. In the Fragment Mass 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.
6. (Optional) Select the **Delta m Mode** check box if you want to conduct the search in delta-m (Δm) mode.

   For more information on this mode, see "Performing Searches in Delta-m Mode" on page 113.
7. In the Precursor Search 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: Sets the minimum value for a precursor search window that does not trigger an "out of range" warning, which is displayed as yellow background in the text box.
- Default: Sets the default value for a precursor search window.
- Upper: Sets the maximum value for a precursor search window that does not trigger an "out of range" warning.

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

   - Lower: Sets the minimum value for a fragment tolerance that does not trigger an "out of range" warning.
   - Default: Sets the default value for a fragment tolerance.
   - Upper: 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.

9. 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: Sets the minimum value for minimum matches that will not trigger an "out of range" warning.
   - Default: Sets the default value for minimum matches.
   - Upper: Sets the maximum value for minimum matches that will not trigger an "out of range" warning.

10. Click **OK**.

Absolute Mass Preferences Page Parameters

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

Table 21 lists the parameters on the Absolute Mass Preferences page of the Options dialog box, shown in Figure 50 on page 119.

Table 21. Absolute Mass Preferences page parameters (Sheet 1 of 2)

Parameter	Description
Database	Specifies the name of the database to search.
Precursor Mass	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	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 the ProSightPC application conducts the search in delta-m (Δm) mode. For more information on this mode, see "Performing Searches in Delta-m Mode" on page 113.
Precursor Search	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 will 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 will not trigger an "out of range" warning.

Table 21. Absolute Mass Preferences page parameters (Sheet 2 of 2)

Parameter	Description
Fragment Tolerance	Specifies the tolerance that determines whether comparing an observed fragment ion mass to a theoretical fragment ion mass is considered a match. Set the following parameters: • Lower: Sets the minimum value for a fragment tolerance that will not trigger an "out of range" warning. • Default: Sets the default value for a fragment tolerance. • Upper: Sets the maximum value for a fragment tolerance that will 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: Sets the minimum value for minimum matches that will not trigger an "out of range" warning. • Default: Sets the default value for minimum matches. • Upper: Sets the maximum value for minimum matches that will not trigger an "out of range" warning.

Searching for Absolute Mass

❖ To search for absolute mass

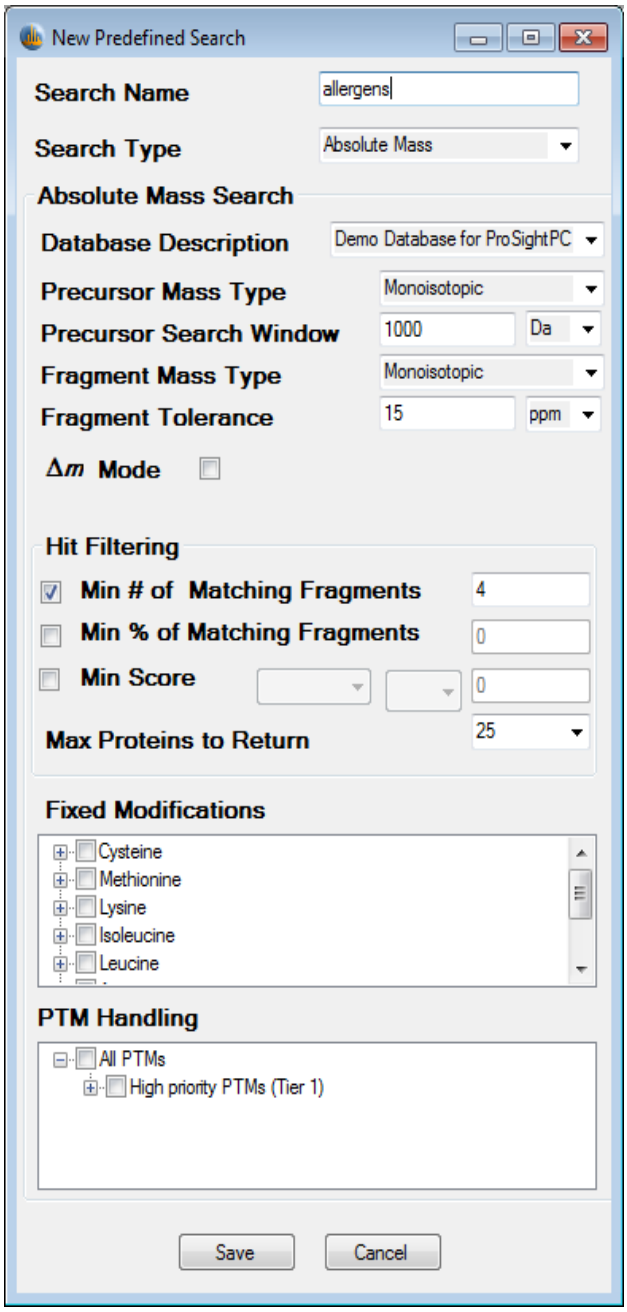
1. Start a search by following the instructions in "Creating a Predefined Search" on page 105.

   The New Predefined Search dialog box opens, as shown in Figure 44 on page 106.

2. In the Search Name box, type the name of the search.
3. In the Search Type list, select **Absolute Mass** if it is not already selected.

   Because absolute mass is the default type of search, the New Predefined Search dialog box does not change (see Figure 51).

Figure 51. New Predefined Search dialog box for absolute mass



4. In the Database Description list, select a description of the database that you want to search.

5. In 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.

6. In the Precursor Search Window box, specify the tolerance that determines whether comparing an observed precursor mass to a theoretical precursor mass is considered a match, and indicate whether it is expressed as absolute (measured in daltons) or relative (measured in parts per million).

   The window is one number. For example, if you type 10 and select Da, the ProSightPC application queries +10 Da and –10 Da around the observed precursor for a total range of 20 Da. The ProSightPC application queries all proteoforms with a theoretical mass within this range.

7. 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.

8. In the Fragment Tolerance box, specify the tolerance that determines whether comparing 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.

9. Select the **Δ*m* Mode** check box if you want to conduct the search in delta-m (Δ*m*) mode. For more information on delta-m (Δ*m*) mode, see "Performing Searches in Delta-m Mode" on page 113.

10. In the Hit Filtering section, set at least one of the following filters; otherwise, the ProSightPC application returns all proteoforms 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 proteoforms 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 196 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.

11. 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.

12. 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.

13. 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. The ProSightPC application only queries theoretical proteoforms containing exclusively selected PTMs. It excludes every proteoform containing an unselected PTM from the interrogation.

14. Click **Save**.

    The new search appears in the data grid, with "yes" appearing in the Pending Search column.

15. To execute the search from the data grid, right-click the pending search and then choose **Run Search *number***.

To view the results of the search, refer to "Viewing Search Results" on page 167.

New Predefined Search Dialog Box Parameters for Absolute Mass

Table 22 lists the parameters in the New Predefined Search dialog box for absolute mass, shown in Figure 51 on page 123.

Table 22. New Predefined Search dialog box parameters for absolute mass (Sheet 1 of 3)

Parameter	Description
Search Name	Specifies the name of the search.
Search Type	Specifies the type of search to perform: • Absolute Mass: Performs an absolute mass search, which matches MS/MS data against all intact proteoforms of proteins in a database. For detailed information, see "Searching for Absolute Mass" on page 116. • BioMarker: Performs a biomarker search, which matches MS/MS data against all subsequences of all proteoforms of proteins in a database. For detailed information, see "Searching for Biomarkers" on page 128. • Sequence Tag: Performs a sequence tag search, which searches for a series of consecutive fragment ions from which you might infer a partial protein sequence. You can then use these sequence tags to identify the protein when you search the sequence database. For detailed information, see "Searching for Sequence Tags" on page 139. • Single Protein: Performs a search for a single protein. For more information, see "Searching for Single Proteins" on page 149. • Gene-Restricted Absolute Mass: Performs a gene-restricted absolute mass (GRAM) search, which is an absolute mass search for every proteoform of each gene in the gene list, regardless of the theoretical precursor mass. For detailed information, see "Searching for Gene-Restricted Absolute Masses" on page 153. • Gene-Restricted BioMarker: Performs a gene-restricted biomarker (GRBM) search, which is a biomarker search for all proteoforms of the genes listed in the gene identification list. For detailed information, see "Searching for Gene-Restricted Biomarkers" on page 159.
Database Description	Describes the database that you want to search.

Table 22. New Predefined Search dialog box parameters for absolute mass (Sheet 2 of 3)

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.
Precursor Search Window	Specifies a range around the observed precursor mass, in daltons. The ProSightPC application queries all proteoforms 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 comparing 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 the ProSightPC application conducts the search in delta-m (Δ*m*) mode. "Performing Searches in Delta-m Mode" on page 113 explains this mode.
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.

Table 22. New Predefined Search dialog box parameters for absolute mass (Sheet 3 of 3)

Parameter	Description
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. • (Default) ≤: Indicates that the first value is less than or equal to the second value. • ≥: 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 Biomarkers

A biomarker search matches MS/MS data against all subsequences of all proteoforms of proteins in a database. It is similar to a bottom-up no enzyme search. A biomarker search is a "brute force" search of an entire database and can take a long time. It looks at every possible subsequence of every base proteoform (unless mentioned otherwise) in the database and attempts to identify any subsequence that matches the observed intact ion mass within a tolerance. 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.

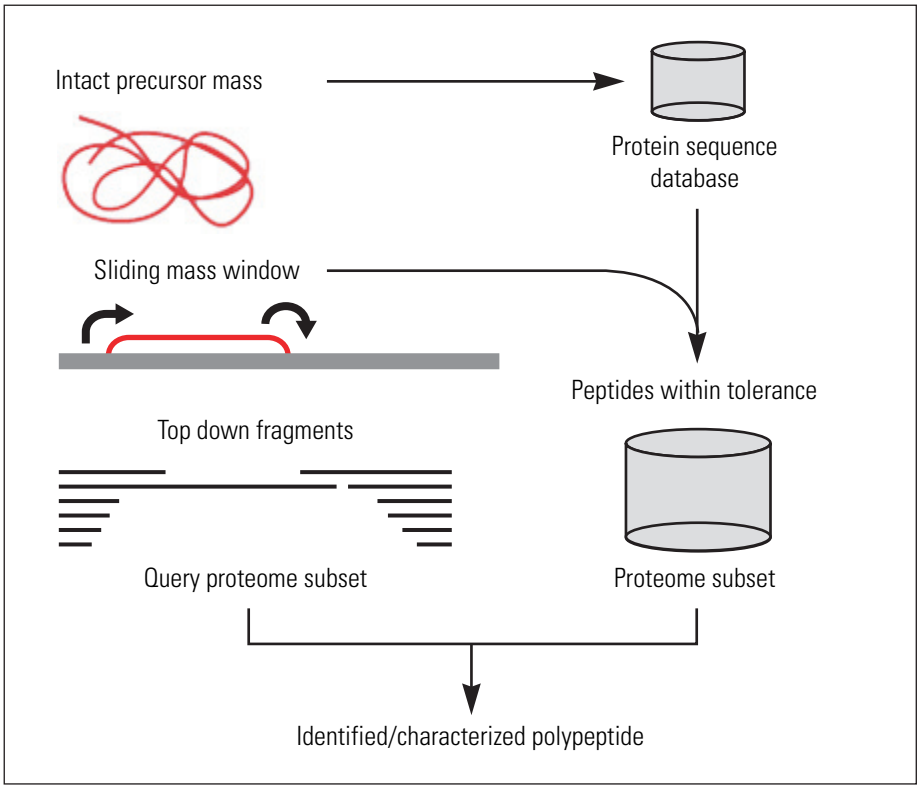
In a typical top-down experiment, not all polypeptides identified are intact proteins. A biomarker search identifies those proteins that are a product of biological degradation and cannot be logically predicted. It compares the observed precursor mass to all possible entries of a particular database within a defined tolerance, for example, less than 10 ppm. The ProSightPC application 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 52 shows the process involved in a biomarker search.

Figure 52. 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 proteoforms with no known modifications. If you want to search for modified proteoforms, you must select the Include Modified Forms check box in the New Predefined Search dialog box for biomarkers (see Figure 54 on page 134). However, searching for modified proteoforms 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-m (Δm) mode.

Follow these sets of instructions to specify preferences and perform the search:

- Setting Biomarker Search Preferences

- Searching for Biomarkers

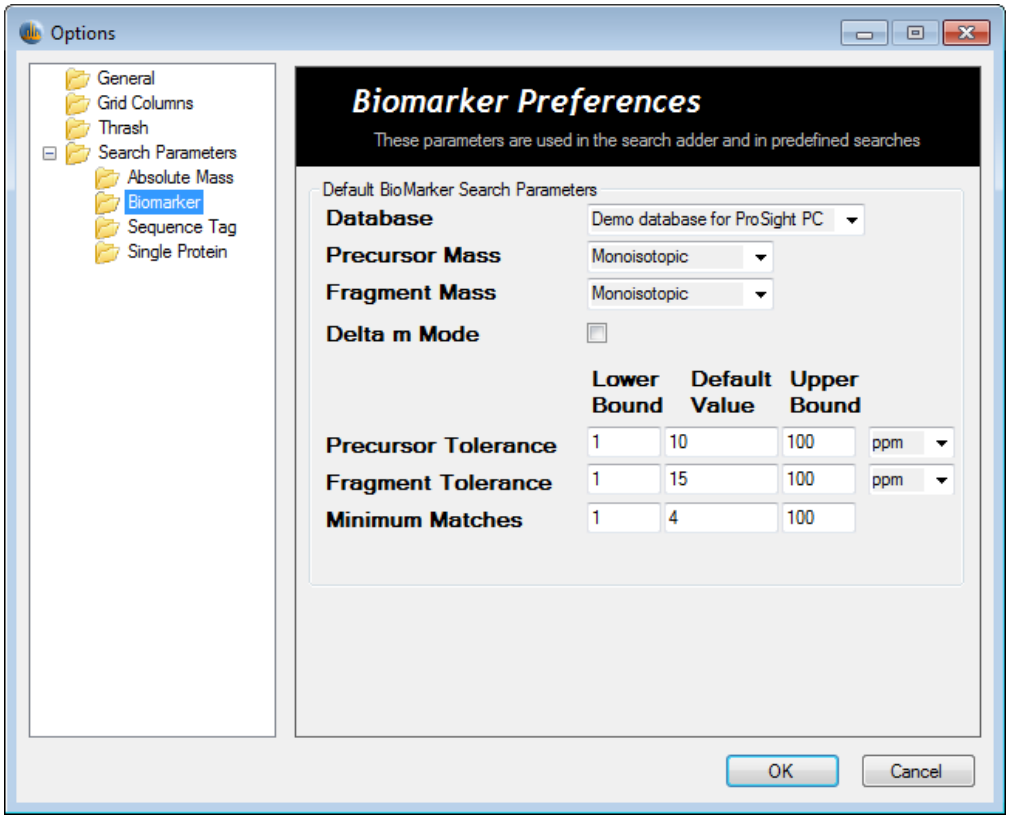
Setting Biomarker Search Preferences

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 Biomarkers" on page 128.

❖ To set biomarker search preferences

1. Choose **Tools > Options** to open the Options dialog box.
2. From the left pane of the Options dialog box, click the **Search Parameters** folder and then click the **Biomarker** folder to open the Biomarker Preferences page, shown in Figure 53.

Figure 53. Biomarker Preferences page of the Options dialog box



3. In the Database list, select the name of the database to search.
4. In the Precursor Mass 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.

5. In the Fragment Mass 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.

6. (Optional) Select the **Delta m Mode** check box if you want to conduct the search in delta-m (Δm) mode.

   For more information on this mode, see "Performing Searches in Delta-m Mode" on page 113.

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

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

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

   - Lower: Sets the minimum value for a fragment tolerance that does not trigger an "out of range" warning.
   - Default Sets the default value for a fragment tolerance.
   - Upper: 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.

9. 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: Sets the minimum value for minimum matches that does not trigger an "out of range" warning.
   - Default: Sets the default value for minimum matches.
   - Upper: Sets the maximum value for minimum matches that does not trigger an "out of range" warning.
10. Click **OK**.

Biomarker Preferences Page Parameters

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

Table 23 lists the parameters on the Biomarker Preferences page of the Options dialog box, shown in Figure 53 on page 130.

Table 23. Biomarker Preferences page parameters (Sheet 1 of 2)

Parameter	Description
Database	Specifies the name of the database to search
Precursor Mass	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	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 the ProSightPC application conducts the search in delta-m (Δm) mode. For more information on this mode, see "Performing Searches in Delta-m Mode" on page 113.

Table 23. Biomarker Preferences page parameters (Sheet 2 of 2)

Parameter	Description
Precursor Tolerance	Specifies the tolerance that determines whether comparing 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.
Fragment Tolerance	Specifies the tolerance that determines whether comparing 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 a minimum match that does not trigger an "out of range" warning. • Default Value: Sets the default value for a minimum match. • Upper Bound: Sets the maximum value for a minimum match that does not trigger an "out of range" warning.

Searching for Biomarkers

❖ To search for a biomarker

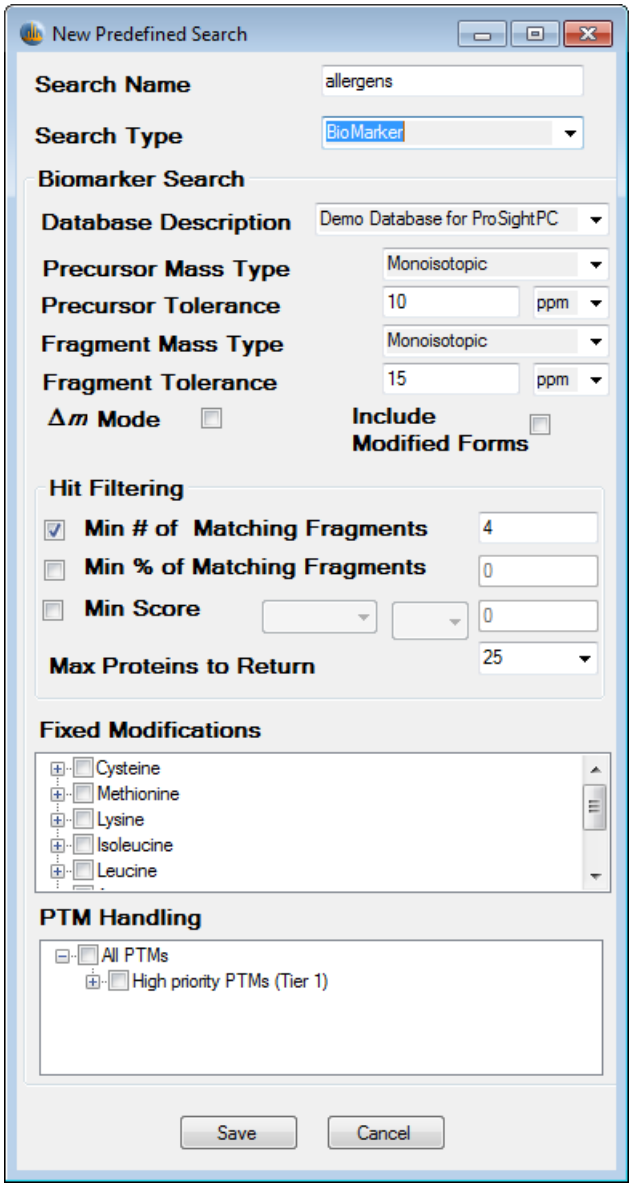
1. Start a search by following the instructions in "Creating a Predefined Search" on page 105.

   The New Predefined Search dialog box opens, as shown in Figure 44 on page 106.

2. In the Search Name box, type the name of the search.
3. In the Search Type list, select **BioMarker**.

   The New Predefined Search dialog box changes to the version shown in Figure 54.

   Figure 54. New Predefined Search dialog box for biomarkers



4. In the Database Description list, select a description of the database that you want to search.
5. 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.

6. In the Precursor Tolerance box, enter the range value for tolerance when testing all proteoforms for biomarker peptides. Indicate whether it is expressed as absolute (measured in Da) or relative (measured in ppm).
7. 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.
8. In the Fragment Tolerance box, specify the tolerance that determines whether comparing 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.
9. Select the **Δ*m* Mode** check box if you want to conduct the search in delta-m (Δ*m*) mode. For details, see in "Performing Searches in Delta-m Mode" on page 113.
10. 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.
11. In the Hit Filtering section, set at least one of the following filters; otherwise, the ProSightPC application returns all proteoforms 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 proteoforms containing a minimum number of matching ion fragments (these proteoforms 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 196 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.

12. 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.

13. 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.

14. 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. The ProSightPC application only queries theoretical proteoforms containing exclusively selected PTMs. It excludes every proteoform containing an unselected PTM from the interrogation.

15. Click **Save**.

    The new search appears in the data grid, with "yes" appearing in the Pending Search column.

16. To execute the search from the data grid, right-click on the pending search and then choose **Run Search *number***.

To view the results of the search, refer to "Viewing Search Results" on page 167.

New Predefined Search Dialog Box Parameters for Biomarkers

Table 24 lists the parameters in the New Predefined Search dialog box for biomarkers, shown in Figure 54 on page 134.

Table 24. New Predefined Search dialog box parameters for biomarkers (Sheet 1 of 2)

Parameter	Description
Search Name	Specifies the name of the search.
Search Type	Specifies the type of search to perform: • Absolute Mass: See "Searching for Absolute Mass" on page 116. • BioMarker: See "Searching for Biomarkers" on page 128. • Sequence Tag: See "Searching for Sequence Tags" on page 139. • Single Protein: See "Searching for Single Proteins" on page 149. • Gene-Restricted Absolute Mass: See "Searching for Gene-Restricted Absolute Masses" on page 153. • Gene-Restricted BioMarker: See "Searching for Gene-Restricted Biomarkers" on page 159.
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 proteoforms for biomarker peptides and indicates whether it is expressed as absolute (measured in Da) or relative (measured in ppm).

Table 24. New Predefined Search dialog box parameters for biomarkers (Sheet 2 of 2)

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 comparing 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 the ProSightPC application conducts the search in delta-m (Δm) mode. This mode is explained in "Performing Searches in Delta-m Mode" on page 113.
Include Modified Forms	Determines 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. • (Default) ≤: Indicates that the first value is less than or equal to the second value. • ≥: 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 Sequence Tags

Tandem mass spectrometry experiments are known to create series of consecutive fragment ions from which you might infer a partial protein sequence. You can then use these sequence tags to identify the protein when you search the sequence database.

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), the ProSightPC application 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 a single amino acid or select amino acid pairs within the compiler tolerance that you defined. The application always gives the compiler tolerance 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 and ETD data.

2. The ProSightPC application 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 score is based on the negative log of the probability of the sequences existing in nature.

   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]**.

Follow these sets of instructions to specify preferences and perform the search:

- Setting Sequence Tag Search Preferences
- Searching for Sequence Tags

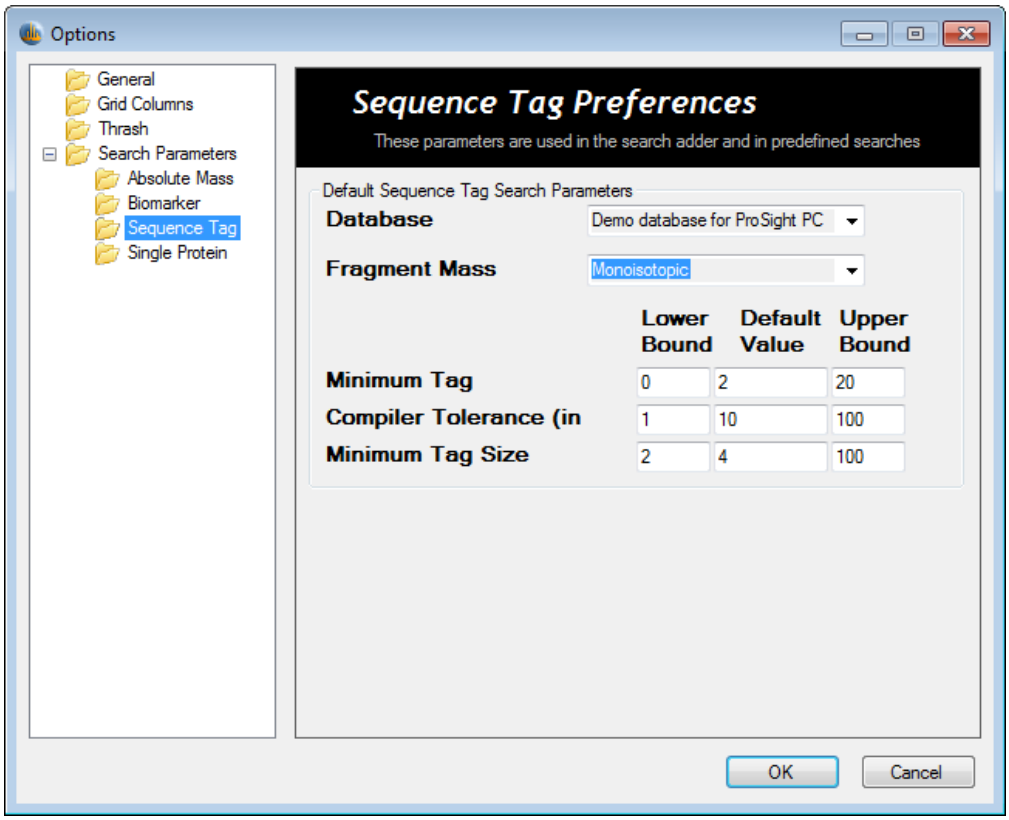
Setting Sequence Tag Search Preferences

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 Sequence Tags" on page 139.

❖ To set sequence tag search preferences

1. Choose **Tools > Options** to open the Options dialog box.
2. From the left pane of the Options dialog box, click the **Search Parameters** folder and then click the **Sequence Tag** folder to open the Sequence Tag Preferences page, shown in Figure 55.

Figure 55. Sequence Tag Preferences page in the Options dialog box



3. In the Database list, select the name of the database to search.
4. In the Fragment Mass 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. In the Minimum Tag boxes, specify the minimum tag score for proteoforms that are matched in a sequence tag search.

   The ProSightPC application 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: Sets the minimum value for a minimum tag score that does not trigger an "out of range" warning.
   - Default: Sets the default value for a minimum tag score.
   - Upper: Sets the maximum value for a minimum tag score that does not trigger an "out of range" warning.

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

   - Lower: Sets the minimum value for a compiler tolerance that does not trigger an "out of range" warning.
   - Default: Sets the default value for a compiler tolerance.
   - Upper: Sets the maximum value for a compiler tolerance that does not trigger an "out of range" warning.

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

   - Lower: Sets the minimum value for a minimum tag size that does not trigger an "out of range" warning.
   - Default: Sets the default value for a minimum tag size.
   - Upper: Sets the maximum value for a minimum tag size that does not trigger an "out of range" warning.

8. Click **OK**.

Sequence Tag Preferences Page Parameters

Table 25 lists the parameters on the Sequence Tag Preferences page of the Options dialog box, shown in Figure 55 on page 140.

Table 25. Sequence Tag Preferences page parameters

Parameter	Description
Database	Specifies the name of the database to search.
Fragment Mass	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	Specifies the minimum tag score for proteoforms that are matched in a sequence tag search. The ProSightPC application 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.
Compiler Tolerance (in ppm)	Specifies the permissible error, measured in ppm, between two fragment ion masses that are still considered as 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.

Searching for Sequence Tags

❖ To search for a sequence tag

1. Start a search by following the instructions in "Creating a Predefined Search" on page 105.

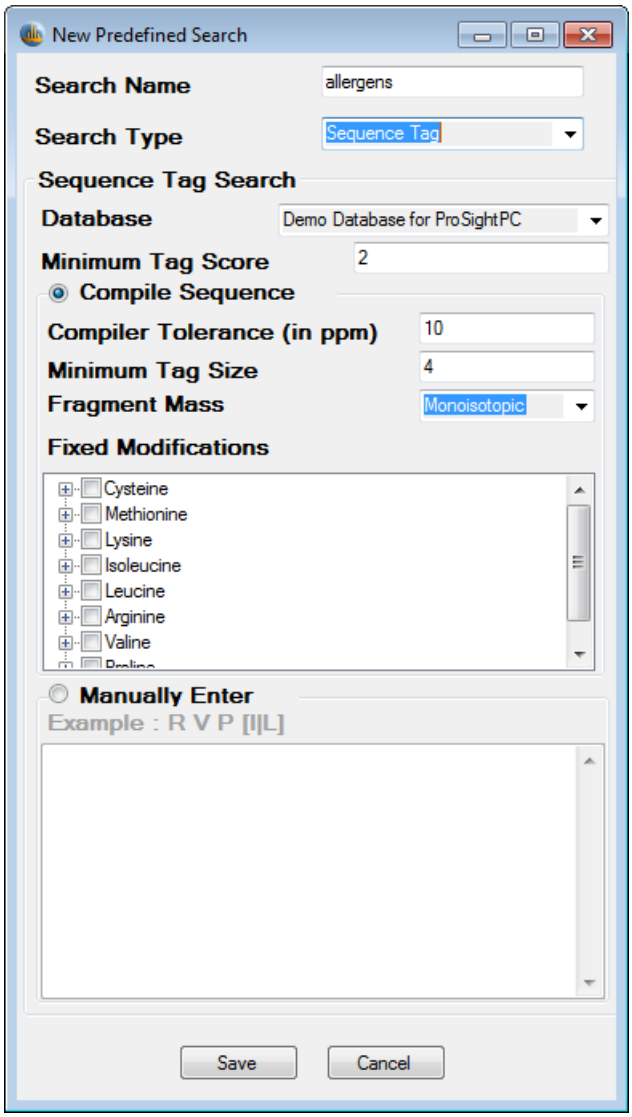
   The New Predefined Search dialog box opens, as shown in Figure 44 on page 106.

2. In the Search Name box, type the name of the search.

3. In the Search Type list select **Sequence Tag**.

   The New Predefined Search dialog box changes to the version shown in Figure 56.

Figure 56. New Predefined Search dialog box for sequence tags



4. In the Database list, select a description of the database that you want to search.
5. In the Minimum Tag Score box, enter the lowest acceptable sequence tag score reported as a match.
6. Select either the **Compile Sequence** or the **Manually Enter** option.
   - Compile Sequence: Determines the sequence tags and compiles them before searching them. If you select this option, complete step 7 through step 10 and step 12. This option is the default.
   - Manually Enter: 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 11 and step 12.
7. 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.
8. In the Minimum Tag Size box, enter the lowest acceptable sequence tag score reported as a match.
9. In the Fragment Mass 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.
10. 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.
11. If you selected the Manually Enter 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.
12. Click **Save**.

    The new search appears in the data grid, with “yes” appearing in the Pending Search column.

13. To execute the search from the data grid, right-click the pending search and then choose **Run Search *number***.

To view the results of the search, see "Viewing Search Results" on page 167.

New Predefined Search Dialog Box Parameters for Sequence Tags

Table 26 lists the parameters in the New Predefined Search dialog box for sequence tags, shown in Figure 56 on page 143.

Table 26. New Predefined Search dialog box parameters for sequence tags (Sheet 1 of 2)

Parameter	Description
Search Name	Specifies the name of the search.
Search Type	Specifies the type of search to perform: • Absolute Mass: See "Searching for Absolute Mass" on page 116. • Biomarker: See "Searching for Biomarkers" on page 128. • Sequence Tag: See "Searching for Sequence Tags" on page 139. • Single Protein: See "Searching for Single Proteins" on page 149. • Gene-Restricted Absolute Mass: See "Searching for Gene-Restricted Absolute Masses" on page 153. • Gene-Restricted BioMarker: See "Searching for Gene-Restricted Biomarkers" on page 159.
Database	Describes the database that you want to search.
Minimum Tag Score	Specifies the lowest acceptable sequence tag score reported as a match.
Compile Sequence Tags	Determines the sequence tags and compiles them before searching them. You can select either the default Compile Sequence Tags option or the Manually Enter option.
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.

Table 26. New Predefined Search dialog box parameters for sequence tags (Sheet 2 of 2)

Parameter	Description
Fragment Mass	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	Select to enter sequence tags that you have determined—possibly from manually analyzing a spectrum—and to search them. You can select either the default Compile Sequence Tags option or the Manually Enter option.
Save	Saves the search information.

Searching for Single Proteins

Single-protein searches match MS/MS data against a single user-defined amino acid sequence.

Follow these sets of instructions to specify preferences and perform the search:

- Setting Single-Protein Search Preferences
- Searching for Single Proteins

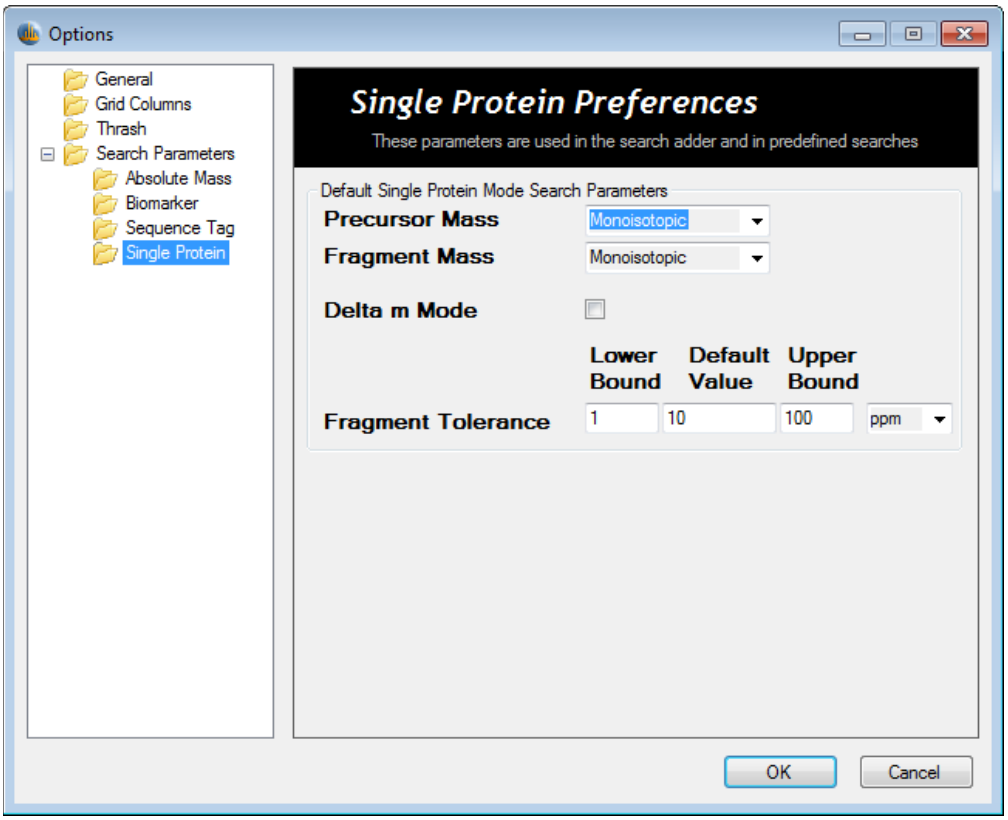
Setting Single-Protein Search Preferences

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 189.

❖ To set single-protein search preferences

1. Choose **Tools > Options** to open the Options dialog box.
2. From the left pane of the Options dialog box, click the **Search Parameters** folder and then click the **Single Protein** folder to open the Single Protein Preferences page, shown in Figure 57.

Figure 57. Single Protein Preferences page of the Options dialog box



3. In the Precursor Mass 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.

4. In the Fragment Mass 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.

5. (Optional) Select the **Delta m Mode** check box if you want to conduct the search in delta-m (Δm) mode.

   For more information on this mode, see See “Performing Searches in Delta-m Mode” on page 113..

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

   - Lower: Sets the minimum value for a fragment tolerance that does not trigger an "out of range" warning.
   - Default: Sets the default value for a fragment tolerance.
   - Upper: 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.

7. Click **OK**.

Single Protein Preferences Page Parameters

Table 27 lists the parameters on the Single Protein Preferences page of the Options dialog box, shown in Figure 57.

Table 27. Single Protein Preferences page parameters (Sheet 1 of 2)

Parameter	Description
Precursor Mass	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	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.

Table 27. Single Protein Preferences page parameters (Sheet 2 of 2)

Parameter	Description
Delta m Mode	Determines whether the ProSightPC application conducts the search in delta-m (Δm) mode. For more information on this mode, see "Performing Searches in Delta-m Mode" on page 113.
Fragment Tolerance	Specifies the tolerance that determines whether comparing 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]).

Searching for Single Proteins

❖ To perform a single-protein search

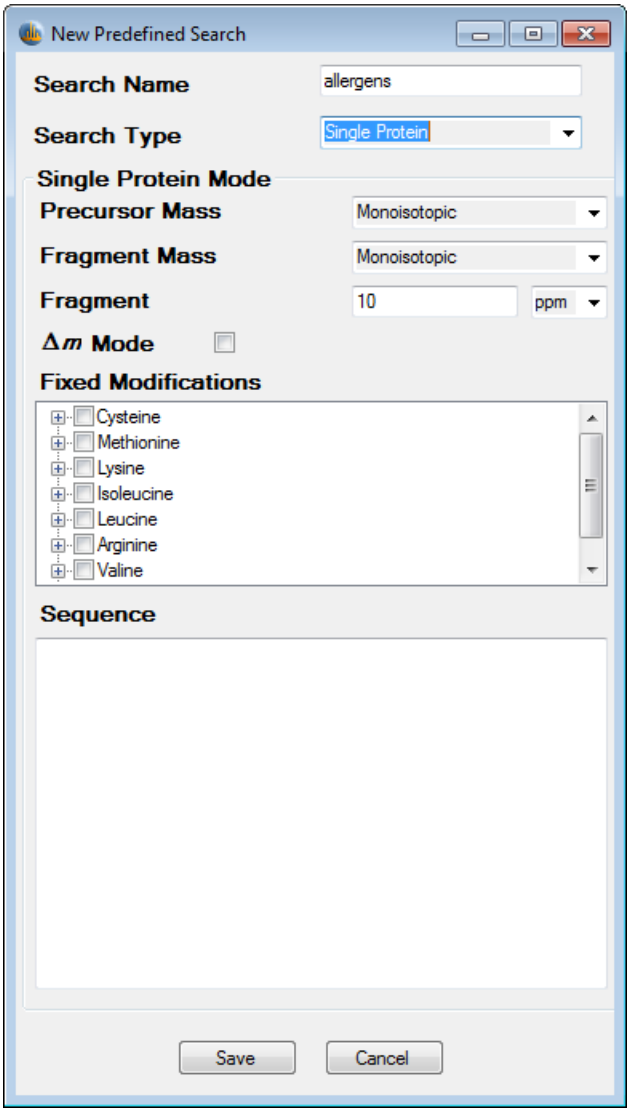
1. Start a search by following the instructions in "Creating a Predefined Search" on page 105.

   The New Predefined Search dialog box opens, as shown in Figure 44 on page 106.

2. In the Search Name box, type the name of the search.
3. In the Search Type list, select **Single Protein**.

   The New Predefined Search dialog box changes to the version shown in Figure 58.

Figure 58. New Predefined Search dialog box for single proteins



4. In the Precursor Mass 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 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 box, specify the tolerance that determines whether comparing 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-m (Δ*m*) mode. For more information on delta-m (Δ*m*) mode, see "Performing Searches in Delta-m Mode" on page 113).
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.**

    The search appears in the Append Predefined Searches to Experiment *X* dialog box.
11. Click **Append**.

> **Note** After step 10, the ProSightPC application 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.

12. In the Data Manager, click **Run Search**.

To view the results of the search, see "Viewing Search Results" on page 167 and "Searching for Single Proteins" on page 189.

New Predefined Search Dialog Box Parameters for a Single Protein

Table 28 lists the parameters in the New Predefined Search dialog box for a single protein, shown in Figure 58 on page 150.

Table 28. New Predefined Search dialog box parameters for a single protein (Sheet 1 of 2)

Parameter	Description
Search Name	Specifies the name of the search.
Search Type	Specifies the type of search to perform: • Absolute Mass: See "Searching for Absolute Mass" on page 116. • BioMarker: See "Searching for Biomarkers" on page 128. • Sequence Tag: See "Searching for Sequence Tags" on page 139. • Single Protein: See "Searching for Single Proteins" on page 149. • Gene-Restricted Absolute Mass: See "Searching for Gene-Restricted Absolute Masses" on page 153. • Gene-Restricted BioMarker: See "Searching for Gene-Restricted Biomarkers" on page 159.
Precursor Mass	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	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.

Table 28. New Predefined Search dialog box parameters for a single protein (Sheet 2 of 2)

Parameter	Description
Fragment	Specifies the tolerance that determines whether comparing an observed fragment ion mass to a theoretical fragment ion mass is considered a match. This parameter also indicates whether it is expressed as absolute, measured in Da, or relative, measured in ppm.
Δm Mode	Determines whether the ProSightPC application conducts the search in delta-m (Δ*m*) mode. "Performing Searches in Delta-m Mode" on page 113 explains this mode.
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.

Performing Gene-Restricted Searches

Gene-restricted searches look at all proteoforms 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.

The ProSightPC application 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 Gene-Restricted Absolute Masses

Use a gene-restricted absolute mass (GRAM) search to perform an absolute mass search on every proteoform of each gene in the gene list, regardless of theoretical precursor mass. The ProSightPC application reports only those proteoforms that meet the minimum matches parameter.

❖ To set gene-restricted absolute mass search preferences

See "To set absolute mass search preferences" on page 118.

❖ To perform a gene-restricted absolute mass search

1. Perform one of the following procedures:

   a. Start a search by following the instructions in "Creating a Predefined Search" on page 105.

      The New Predefined Search dialog box opens, as shown in Figure 44 on page 106.

   b. In the Search Name box, type the name of the search.

   c. In the Search Type list, select **Gene-Restricted Absolute Mass**.

   –or –

   a. Perform any search.

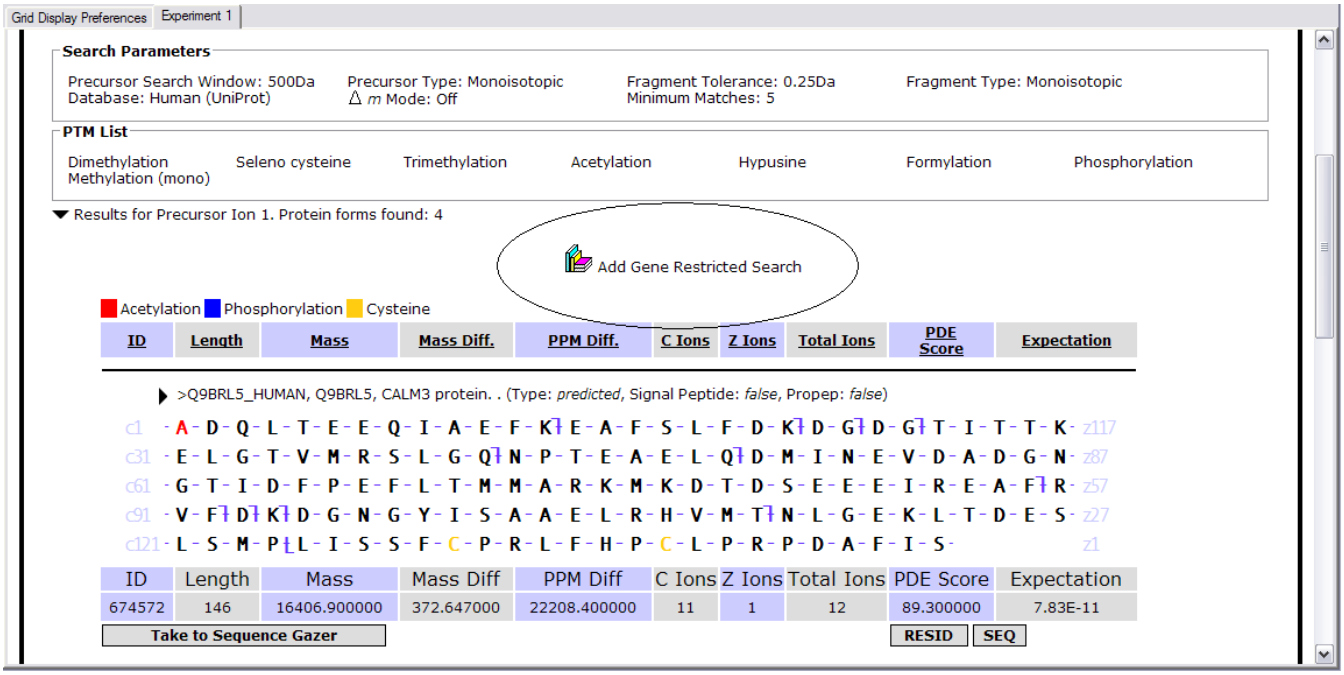
   b. Double-click an experiment in the Data Manager to view it.

   c. Click the display results arrow next to Search *x*.

   d. Expand the view for Results for Precursor Ion 1.

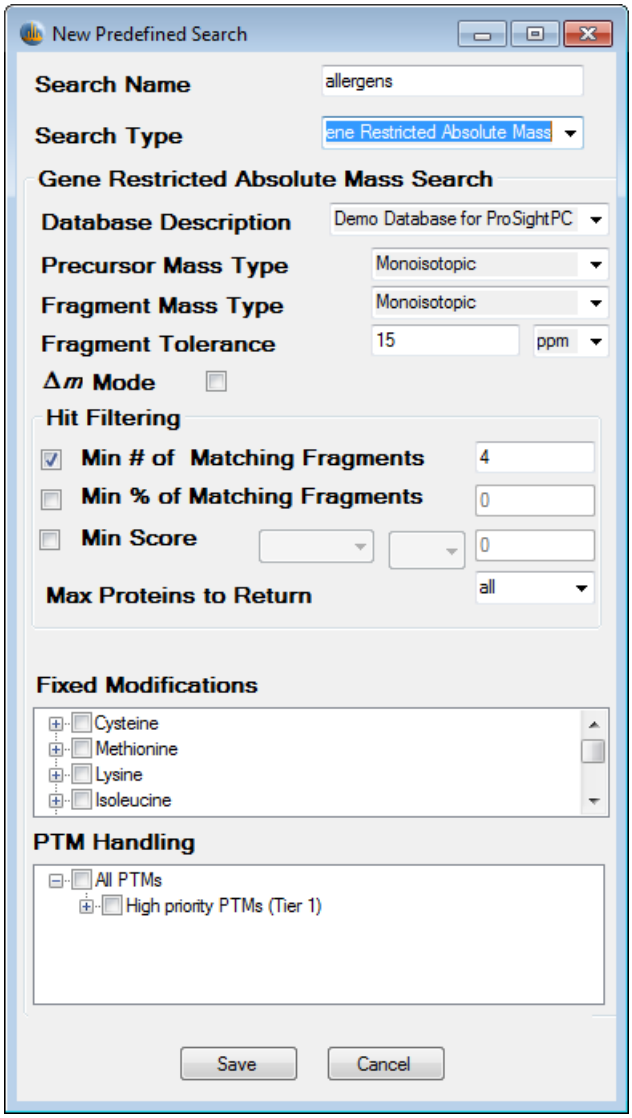
   e. From the results list view, click the **Add Gene Restricted Search** icon,  , circled in Figure 59.

Figure 59. Performing a gene-restricted search



The New Predefined Search dialog box for gene-restricted absolute mass opens, as shown in Figure 60.

Figure 60. New Predefined Search dialog box for gene-restricted absolute mass



f. In the Search Type list, select **Gene-Restricted Absolute Mass**.

2. In the Database Description list, select the proteome database to compare the entry or entries to.

3. 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.

4. 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.

5. In the Fragment Tolerance box, specify the tolerance that determines whether comparing 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.

6. Select the **Δ*m* Mode** check box if you want to conduct the search in delta-m (Δ*m*) mode. For more information on delta-m (Δ*m*) mode, see "Performing Searches in Delta-m Mode" on page 113.

7. In the Hit Filtering section, set at least one of the following filters; otherwise, the ProSightPC application returns all proteoforms 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 proteoforms 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 196 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.

8. 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.

9. In the Fixed Modifications box, select no more than one fixed modification per amino acid type.

10. 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. The ProSightPC application only queries theoretical proteoforms containing exclusively selected PTMs. It excludes every proteoform containing an unselected PTM from the interrogation.

11. Click **Save**.

    The new search appears in the data grid, with "yes" appearing in the Pending Search column.

12. To execute the search from the data grid, right-click the pending search and then choose **Run Search *number***.

New Predefined Search Dialog Box Parameters for Gene-Restricted Absolute Mass

Table 29 lists the parameters in the New Predefined Search dialog box for gene-restricted absolute mass, shown in Figure 60 on page 155.

Table 29. New Predefined Search dialog box parameters for gene-restricted absolute mass (Sheet 1 of 3)

Parameter	Description
Search Name	Specifies the name of the search.
Search Type	Specifies the type of search to perform: • Absolute Mass: See "Searching for Absolute Mass" on page 116. • BioMarker: See "Searching for Biomarkers" on page 128. • Sequence Tag: See "Searching for Sequence Tags" on page 139. • Single Protein: See "Searching for Single Proteins" on page 149. • Gene-Restricted Absolute Mass: See "Searching for Gene-Restricted Absolute Masses" on page 153. • Gene-Restricted BioMarker: See "Searching for Gene-Restricted Biomarkers" on page 159.
Database Description	Describes the database that you want to search.

Table 29. New Predefined Search dialog box parameters for gene-restricted absolute mass (Sheet 2 of 3)

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 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 comparing 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 the ProSightPC application conducts the search in delta-m (Δm) mode. "Performing Searches in Delta-m Mode" on page 113 explains this mode.
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.

Table 29. New Predefined Search dialog box parameters for gene-restricted absolute mass (Sheet 3 of 3)

Parameter	Description
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. • (Default) ≤: Indicates that the first value is less than or equal to the second value. • ≥: 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 Gene-Restricted Biomarkers

Use a gene-restricted biomarker (GRBM) search to perform a biomarker search on all proteoforms 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 set gene-restricted biomarker search preferences

Follow the instructions "To set biomarker search preferences" on page 130.

❖ To search for a gene-restricted biomarker

1. Perform one of the following procedures:
   a. Start a search by following the instructions in "Creating a Predefined Search" on page 105.

      The New Predefined Search dialog box opens, as shown in Figure 44 on page 106.

   b. In the Search Name box, type the name of the search.
   c. In the Search Type list, select **Gene-Restricted Biomarker**.

   –or –

   a. Perform any search.
   b. Double-click an experiment in the Data Manager to view it.

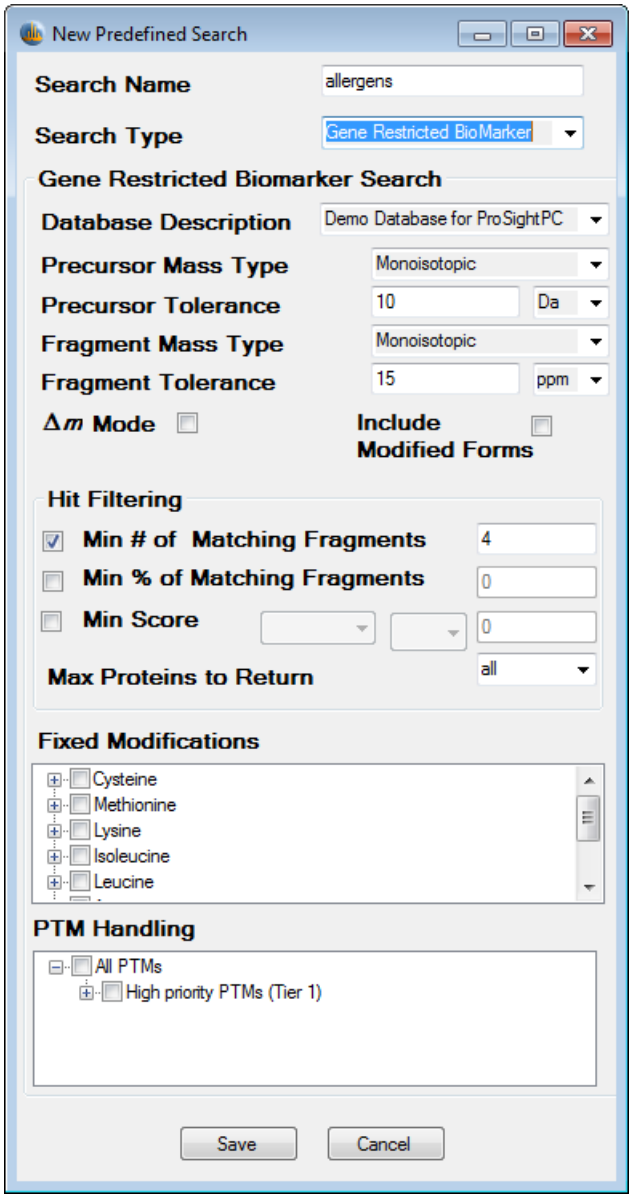
c. Click the display results arrow next to Search *x*.

d. Expand the view for Results for Precursor Ion 1.

e. From the results list view, click the **Add Gene Restricted Search** icon  , circled in Figure 59 on page 154.

The New Predefined Search dialog box for gene-restricted biomarkers opens, as shown in Figure 61.

Figure 61. New Predefined Search dialog box for gene-restricted biomarkers



f. In the Search Type list, select **Gene-Restricted BioMarker**.

2. In the Database Description list, select the proteome database to compare the entry or entries to.
3. 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.
4. In the Precursor Tolerance box, enter the range value for tolerance when testing all proteoforms 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, 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.
6. In the Fragment Tolerance box, specify the tolerance that determines whether comparing 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.
7. Select the **Δ*m* Mode** check box if you want to conduct the search in delta-m (Δ*m*) mode. For more information on delta-m (Δ*m*) mode, see "Performing Searches in Delta-m Mode" on page 113.
8. 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.
9. In the Hit Filtering section, set at least one of the following filters; otherwise, the ProSightPC application returns all proteoforms 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 proteoforms 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 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.

10. 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.

11. In the Fixed Modifications box, select no more than one fixed modification per amino acid type.

12. 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. The ProSightPC application only queries theoretical proteoforms containing exclusively selected PTMs. It excludes every proteoform containing an unselected PTM from the interrogation.

13. Click **Save**.

    The new search appears in the data grid, with "yes" appearing in the Pending Search column.

14. To execute the search from the data grid, right-click the pending search and then choose **Run Search *number***.

New Predefined Search Dialog Box Parameters for Gene-Restricted Biomarkers

Table 30 lists the parameters in the New Predefined Search dialog box for gene-restricted biomarkers, shown in Figure 61 on page 160.

Table 30. New Predefined Search dialog box for gene-restricted biomarkers (Sheet 1 of 2)

Parameter	Description
Search Name	Specifies the name of the search.
Search Type	Specifies the type of search to perform: • Absolute mass: See "Searching for Absolute Mass" on page 116. • BioMarker: See "Searching for Biomarkers" on page 128. • Sequence Tag: See "Searching for Sequence Tags" on page 139. • Single Protein: See "Searching for Single Proteins" on page 149. • Gene-Restricted Absolute Mass: See "Searching for Gene-Restricted Absolute Masses" on page 153. • Gene-Restricted BioMarker: See "Searching for Gene-Restricted Biomarkers" on page 159.
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 proteoforms for biomarker peptides and indicates whether it is expressed as absolute (measured in Da) or relative (measured in ppm).

Table 30. New Predefined Search dialog box for gene-restricted biomarkers (Sheet 2 of 2)

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 comparing 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 the ProSightPC application conducts the search in delta-m (Δm) mode. Performing Searches in Delta-m Mode explains this mode.
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. • (Default) $\leq$: Indicates that the first value is less than or equal to the second value. • $\geq$: 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.
PTM Handling	Specifies the PTMs that you want queried.
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 the ProSightPC application uses these tags to identify all proteoforms in the database that are consistent with the tags. The list of proteoforms that match the sequence tags functions as input into an absolute mass search.

Analyzing MS/MS Experiments

The ProSightPC application 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, the ProSightPC application 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. The ProSightPC application uses this search to identify and characterize the observed protein.

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

[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.

5

Viewing Search Results

You can view the results of a ProSightPC search in the data grid, Data Manager, search report, or repository report.

Contents

Viewing the Results in the Data Manager

Using the Data Manager is the fastest way to see the results of your search.

❖ To display results in the Data Manager

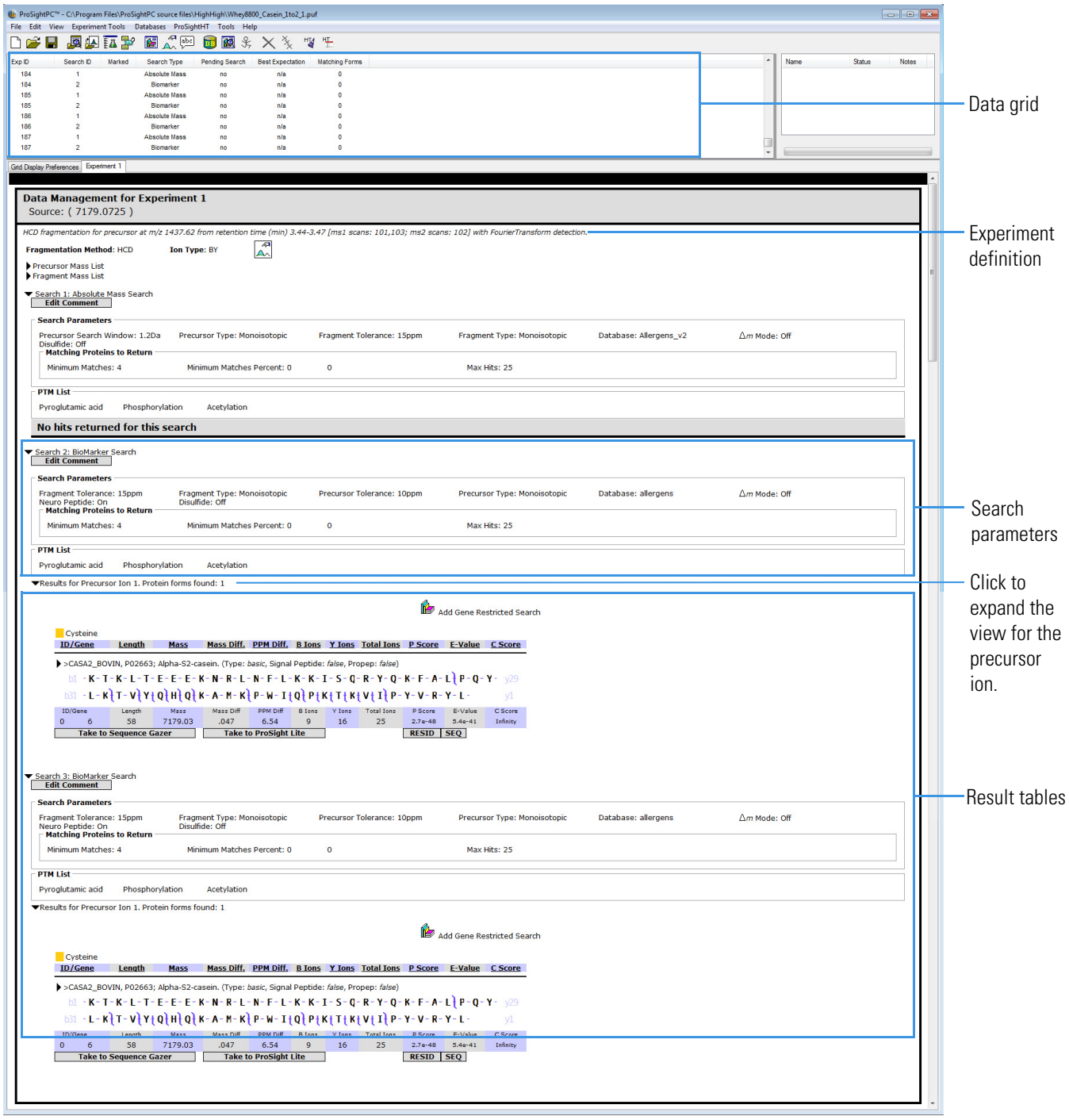
1. When the job queue indicates that a search has finished running, double-click the corresponding experiment identification (Exp ID) line in the data grid to open the Data Manager for the experiment.

   The ProSightPC application automatically highlights this line.

2. In the Data Manager, expand the view for the precursor ion that you are interested in.

   Search results open that are similar to those shown in Figure 62.

Figure 62. Typical search results in the Data Manager



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

Each result list displays the number of protein isoforms found. Expand the results list to display the information contained in the result table. Each result table contains complete information about each matching protein isoform. Information in the result table is organized into the three regions shown in Figure 63. Table 31 describes these regions.

Figure 63. Search result table elements

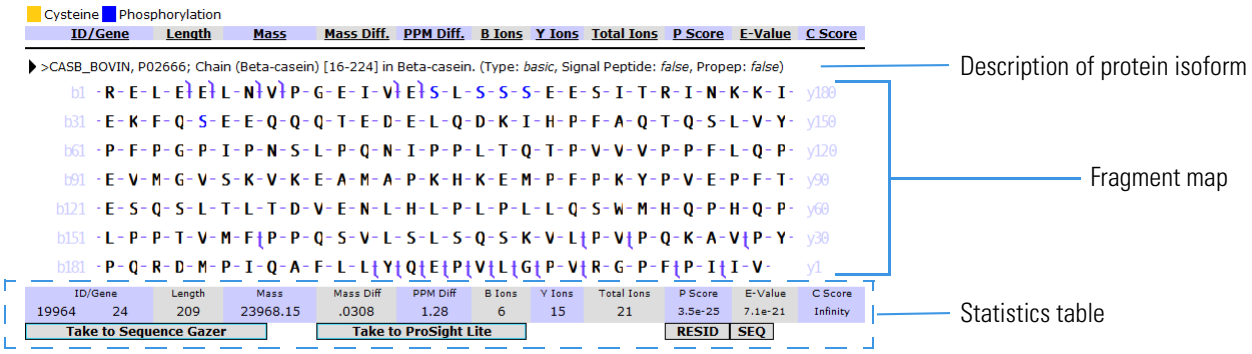


Table 31. Search result table elements

Parameter	Description
Description	Briefly describes the protein or peptide isoform.
Fragment map	Graphically represents the protein isoform, showing the location of PTMs and matching fragment ions.
Statistics table	Organizes information relating to the search.

The statistics table is subdivided into the display elements shown in Table 32.

Table 32. Result statistics table elements (Sheet 1 of 2)

Parameter	Description
ID/Gene	Displays the internal identifier for the proteoform and the gene identifier.
Length	Displays the number of amino acids in the proteoform.
Mass	Displays the theoretical precursor mass of the proteoform.
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.
Other ions	Displays the ions that match the candidate sequence.
Total Ions	Displays the total number of matching ions.
E-Value	Displays the expectation score (e value). For information on this value, see "Expectation Value (e value)" on page 196.

Table 32. Result statistics table elements (Sheet 2 of 2)

Parameter	Description
P Score	Displays the P score. For information on this value, see "P Score" on page 196.
C Score	Displays the C score, which is a measure of the confidence in the characterization of the proteoform. For more information on the C score, see "C Score" on page 199.

Each result has the three context-sensitive buttons as described in Table 33.

Table 33. Result buttons

Button	Description
Take to Sequence Gazer	Creates a new single-protein search based on the result. See "Searching for Single Proteins" on page 189 for details.
Take to ProSight Lite	Opens the ProSight Lite window, shown in Figure 79 on page 193. For information on ProSight Lite, see "Accessing ProSight Lite" on page 192.
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.

The ProSightPC application automatically generates a color-coded legend. An amino acid bearing a PTM is color-coded according to this legend. Cysteines are always colored yellow.

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

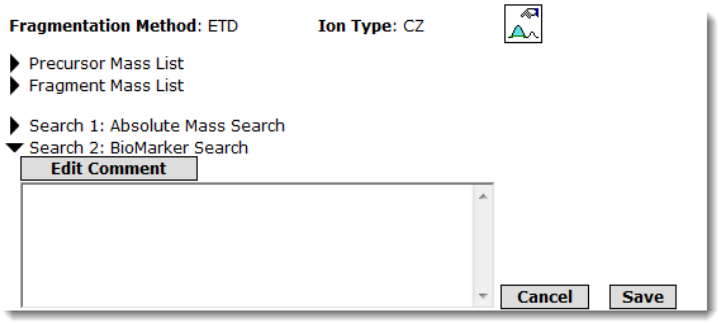
You can perform a gene-restricted search for any results list in the Data Manager. For information on this procedure, see "Performing Gene-Restricted Searches" on page 153.

❖ **To enter and save information specific to the search**

Click **Edit Comment** (see Figure 62 on page 168).

A box opens so that you can type your comments, as shown in Figure 64.

Figure 64. Edit comment box



Viewing the Results in a Search Report

The ProSightPC application 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 topic describes how to use these tools to manage multiple experiments.

The following types of reports help you summarize your work:

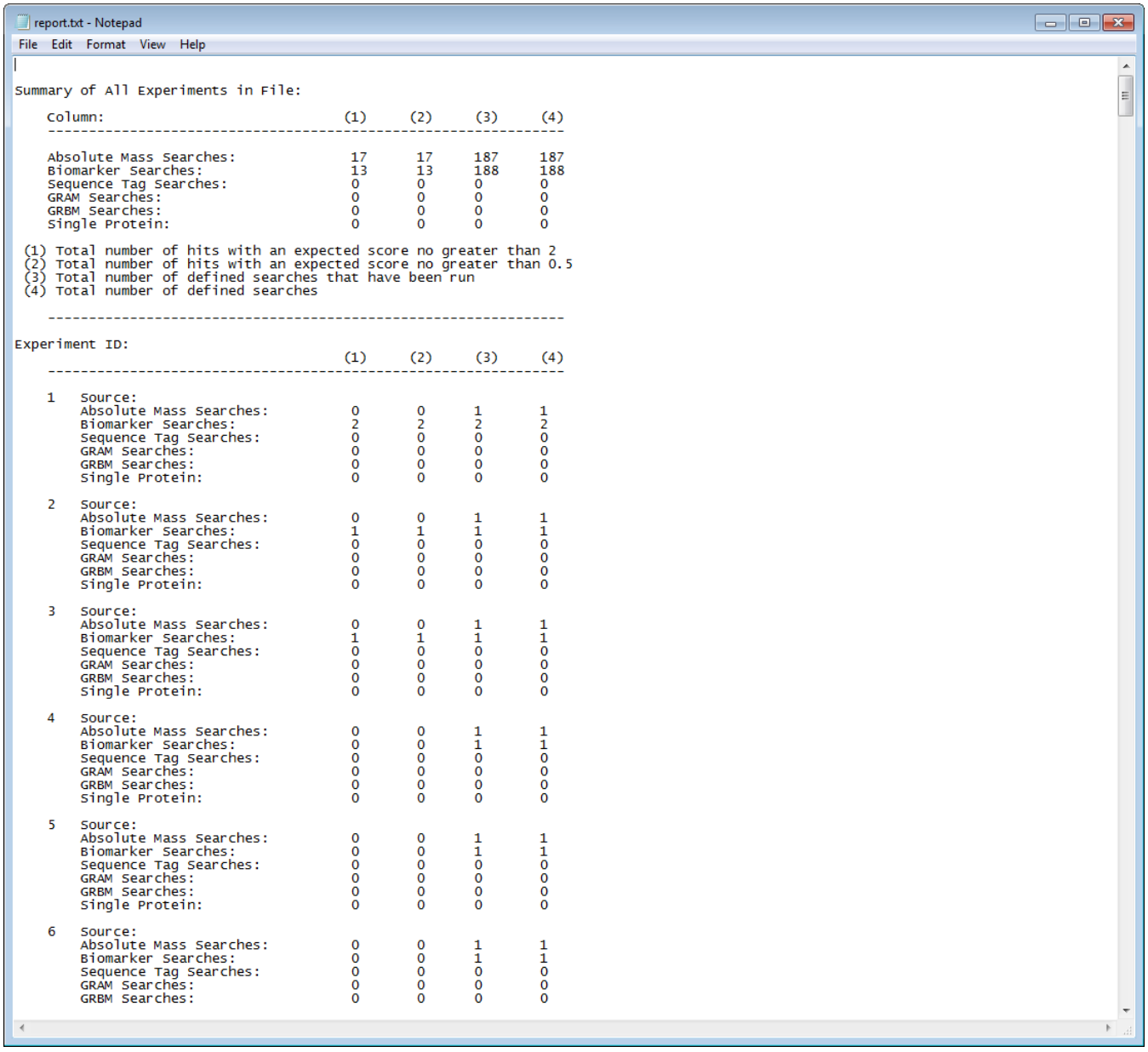
- A status report gives a summary of every search in the open PUF file, including search type and best score.
- A printable search report contains all of the information related to one search, formatted for easy printing.
- A best hit report displays the search result with the best score for each search that was run for each experiment in the data grid.
- A repository report lists all the experiments that a repository contains. For information on this report, see "Viewing the Results in a Repository Report" on page 174.

❖ To generate a status report

1. Open the applicable 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 65. This text document is organized by experiment number and is subdivided into the types of searches.

Figure 65. Status report



report.txt - Notepad

File Edit Format View Help

```
Summary of All Experiments in File:

    Column:                              (1)     (2)     (3)     (4)
    ---------------------------------------------------------------

    Absolute Mass Searches:               17      17     187     187
    Biomarker Searches:                   13      13     188     188
    Sequence Tag Searches:                0       0      0       0
    GRAM Searches:                        0       0      0       0
    GRBM Searches:                        0       0      0       0
    Single Protein:                       0       0      0       0

 (1) Total number of hits with an expected score no greater than 2
 (2) Total number of hits with an expected score no greater than 0.5
 (3) Total number of defined searches that have been run
 (4) Total number of defined searches

    ---------------------------------------------------------------

Experiment ID:
                                     (1)     (2)     (3)     (4)
    ---------------------------------------------------------------

    1   Source:
        Absolute Mass Searches:       0       0       1       1
        Biomarker Searches:           2       2       2       2
        Sequence Tag Searches:        0       0       0       0
        GRAM Searches:                0       0       0       0
        GRBM Searches:                0       0       0       0
        Single Protein:               0       0       0       0

    2   Source:
        Absolute Mass Searches:       0       0       1       1
        Biomarker Searches:           1       1       1       1
        Sequence Tag Searches:        0       0       0       0
        GRAM Searches:                0       0       0       0
        GRBM Searches:                0       0       0       0
        Single Protein:               0       0       0       0

    3   Source:
        Absolute Mass Searches:       0       0       1       1
        Biomarker Searches:           1       1       1       1
        Sequence Tag Searches:        0       0       0       0
        GRAM Searches:                0       0       0       0
        GRBM Searches:                0       0       0       0
        Single Protein:               0       0       0       0

    4   Source:
        Absolute Mass Searches:       0       0       1       1
        Biomarker Searches:           0       0       1       1
        Sequence Tag Searches:        0       0       0       0
        GRAM Searches:                0       0       0       0
        GRBM Searches:                0       0       0       0
        Single Protein:               0       0       0       0

    5   Source:
        Absolute Mass Searches:       0       0       1       1
        Biomarker Searches:           0       0       1       1
        Sequence Tag Searches:        0       0       0       0
        GRAM Searches:                0       0       0       0
        GRBM Searches:                0       0       0       0
        Single Protein:               0       0       0       0

    6   Source:
        Absolute Mass Searches:       0       0       1       1
        Biomarker Searches:           0       0       1       1
        Sequence Tag Searches:        0       0       0       0
        GRAM Searches:                0       0       0       0
        GRBM Searches:                0       0       0       0
```

❖ 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 66 shows an excerpt of this report.

Figure 66. Printable search report

ProSightPC

Data Management for Experiment 1

Source: (7179.0725)

HCD fragmentation for precursor at m/z 1437.62 from retention time (min) 3.44-3.47 [ms1 scans: 101,103; ms2 scans: 102] with FourierTransform detection.

Search 3: Biomarker Search

Search Parameters

Cysteine Modification:
Δm Mode: Off
Fragment Tolerance: 15ppm
Minimum Number Of Matches:
Fragment Mass Type: Monoisotopic
Intact Tolerance: 10ppm
Intact Mass Type: Monoisotopic
Database: allergensallergens

Results for Intact Ion 1. Protein forms found: 1

ID	Length	Mass	Mass Diff.	PPM Diff.	B Ions	Y Ions	Total Ions	Expectation	P Score

>CASA2_BOVIN, P02663; Alpha-S2-casein. (Type: *basic*, Signal Peptide: *false*, Propep: *false*)

ID	Length	Mass	Mass Diff	PPM Diff	B Ions	Y Ions	Total Ions	Expectation	P Score
0	58	7179.03	.047	6.54	9	16	25	5.4e-41	2.7e-48

Fragment Details

Ion	Fragment ID	Observed Mass (Da)	Theoretical Mass (Da)	Mass Error (Da)	Mass Error (PPM)	Delta M
B27	11	3334.8692	3334.8877199999997	-.0185	-5.5534	--
B32	26	3964.2202	3964.2414099999996	-.0212	-5.3503	--
B34	14	4164.3547	4164.3575	-.0028	-.6724	--
B36	19	4455.452	4455.47941	-.0274	-6.152	--
B37	13	4592.497	4592.53832	-.0413	-8.9972	--
B38	23	4720.5708	4720.5969	-.0261	-5.529	--
B42	24	5178.8476	5178.8644200000008	-.0168	-3.2478	--
B46	9	5703.1691	5703.1391300000005	.03	5.255	--
B52	33	6369.5252	6369.58196	-.0568	-8.9111	--
Y6	3	809.4417	809.443565	-.0019	-2.3041	--
Y7	7	922.5259	922.52762500000006	-.0017	-1.8699	--
Y8	10	1021.5949	1021.596035	-.0011	-1.111	--
Y9	17	1149.6911	1149.690995	.0001	.0913	--
Y10	8	1250.735	1250.7386749999998	-.0037	-2.9383	--
Y11	30	1378.8182	1378.8336349999997	-.0154	-11.1942	--
Y12	2	1475.885	1475.8863949999998	-.0014	-.9452	--
Y13	16	1603.95	1603.9449749999999	.005	3.1329	--
Y16	6	2000.1536	2000.161105	-.0075	-3.7522	--
Y20	21	2458.4303	2458.428625	.0017	.6813	--
Y21	18	2586.4654	2586.487205	-.0218	-8.4304	--
Y22	12	2723.5423	2723.546115	-.0038	-1.4007	--
Y23	27	2851.5846	2851.604695	-.0201	-7.0469	--
Y24	5	3014.6649	3014.668025	-.0031	-1.0366	--
Y26	15	3214.7696	3214.784115	-.0145	-4.5151	--
Y31	28	3844.1122	3844.137805	-.0256	-6.6608	--

Fragmentation Method: HCD **Ion Type**: BY

Intact Mass List

ID	MZ Monoisotopic	MZ Average	Mass Monoisotopic	Mass Average	Intensity
1	1436.8218		7179.0725		119235.84

Fragment Mass List

ID	MZ Monoisotopic	MZ Average	Mass Monoisotopic	Mass Average	Intensity
1	1436.8064		7178.9956		29806.24
2	1476.8929		1475.885		14077.49
3	810.449		809.4417		10804.960000000001
4			7162.9637		2296.14
5	1508.3397		3014.6649		3750.95
6	2001.163		2000.1536		6920.65
7	923.5332		922.5259		5017.9400000000005
8	1251.7423		1250.735		4273.52
9	1426.7996		5703.1691		1795.45
10	1022.6022		1021.5949		3835.86
11	1668.4419		3334.8692		2248.87
12	1362.7784		2723.5423		2821.59
13	1531.8396		4592.497		867.72
14			4164.3547		846.12
15	1608.3921		3214.7696		1892.58
16	1604.9573		1603.95		2858.17
17	1150.6984		1149.6911		2183.16
18	1294.24		2586.4654		1494.97
19	1486.158		4455.452		1561.3
20	1001.5888		2001.1629		1003.96
21	1230.2224		2458.4303		1580.75
22	739.4504		1476.8862		924.48
23			4720.5708		824.46
24			5178.8476		1143.14
25			7135.9745		335.87
26			3964.2202		663.12
27	1426.7996		2851.5846		940.34
28			3844.1122		835.84
29	1112.9663		3335.8771		514.58
30	1379.8254		1378.8182		1146.96
31			5574.0429		324.89
32	1252.7488		1251.7415		1168.06
33			6369.5252		456.5

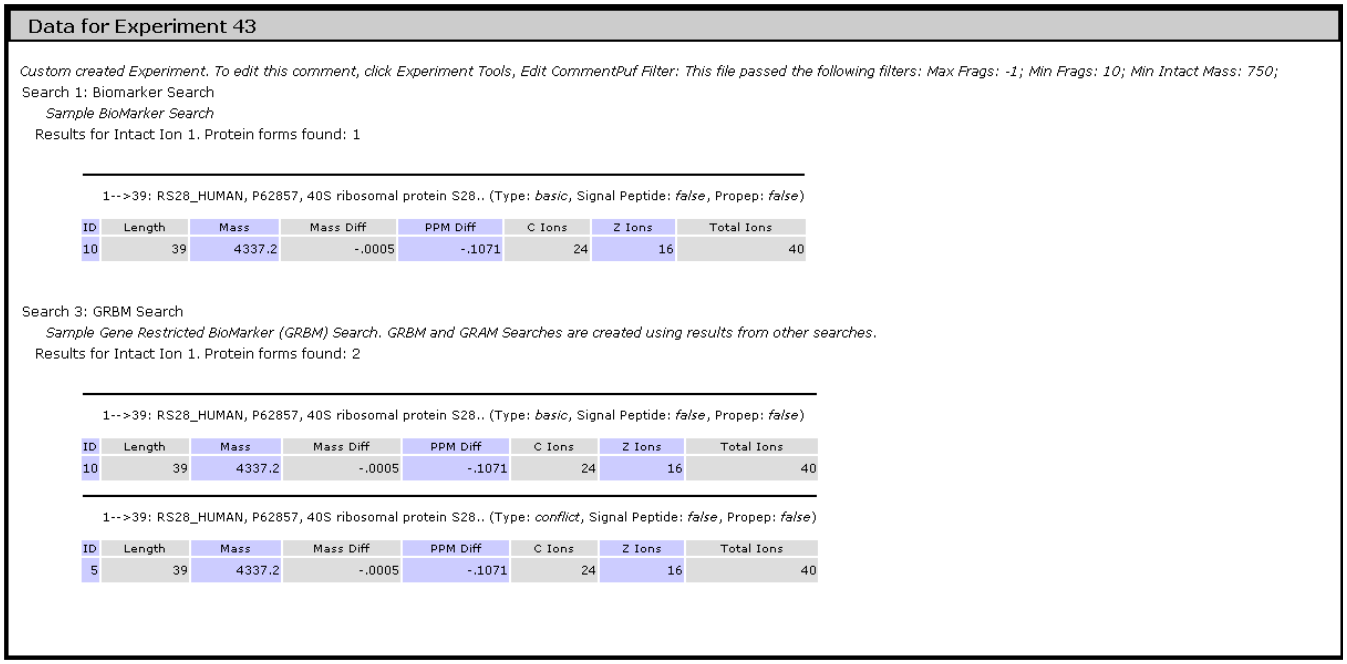
The report presents all relevant data for a search in a printable form similar to that of the Data Manager but only contains information from the selected search.

❖ **To generate a best-hit report**

Choose **Tools > Reports > Best Hit Report**.

The report, shown in Figure 67, appears in a Web browser window.

Figure 67. Best-hit report



Data for Experiment 43

Custom created Experiment. To edit this comment, click Experiment Tools, Edit CommentPuf 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

Viewing the Results in 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 the ProSightPC application, manipulate them, and export the experiments back to the repository or to a Microsoft Excel™ spreadsheet.

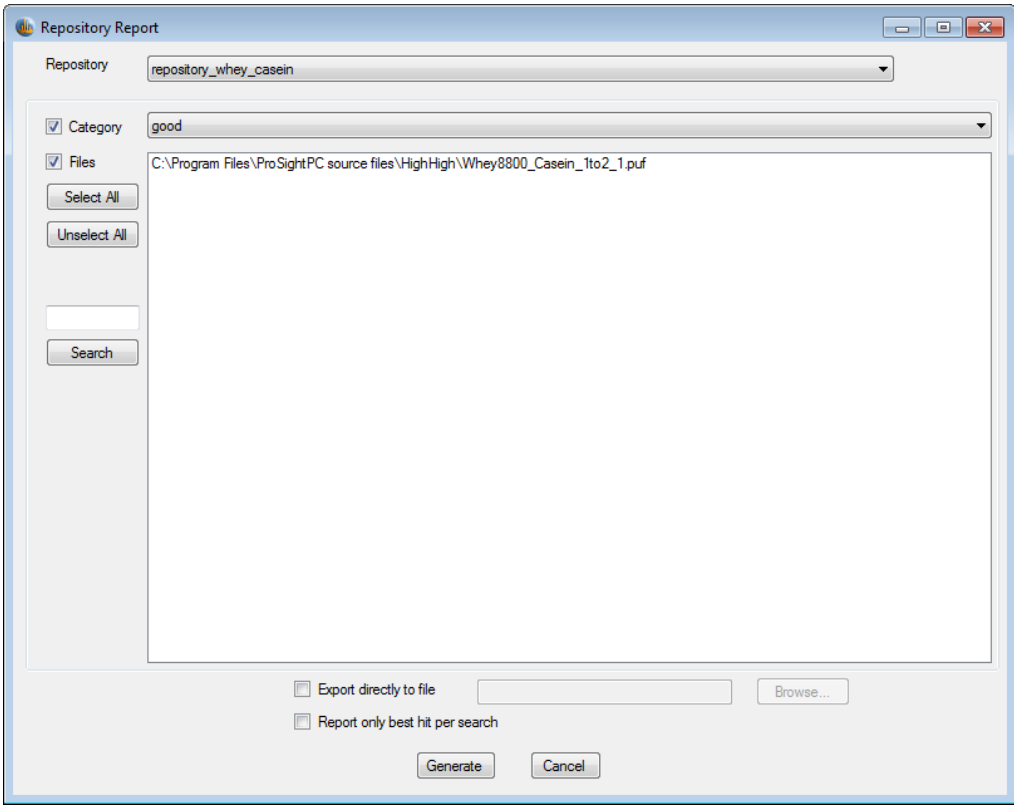
The ProSightPC High Throughput Wizard automatically generates a repository report when it finishes its processing. However, you can also manually generate a repository report. The procedures in this topic describe how to generate a repository report and manipulate its data. See "Demonstrating Repository Report Generation" on page 187 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 68.

   Figure 68. Repository Report dialog box



2. From the Repository list, select the name of the repository to generate a report for.
3. Select the **Category** option and then select the name of the category from the list to the right of the option.

   You assign a search category on the Running High Throughput Logic page of the High Throughput Wizard. For more information on selecting a search category, see "Creating a Search Tree" on page 37.

4. (Optional) To generate a report on the experiments in a specific file or files, select the **File** option, and then select the name of the file or files from the list to the right of the option or type the name of the file in the box.

   When you use the High Throughput Wizard or the Export to Repository command or button, the ProSightPC application automatically adds the names of the PUF files used to the Repository Report dialog box.

   Because 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.

If you do not select a specific file, the ProSightPC application generates a report on the experiments in all files.

- To select all the listed files, click **Select All**.
- To clear all the listed files, click **Unselect All**.
- To clear a particular file, click it.
- To select files that contain a certain text string, type the text string in the box above the Search button.

5. If you want to export the repository report data to a text file instead of to the repository report, select the **Export Directly to File** check box. Figure 70 gives a partial example of a report in a Microsoft Notepad file.
6. If you want to generate a report on only the best result per precursor ion, select the **Report Only Best Hit Per Search** check box.

   The ProSightPC application attempts to discriminate between very close e values or P scores by examining the intact mass differences and choosing the one with the smallest mass difference.
7. Click **Generate**.

The repository report automatically opens on the Report (*Report_name*) page, as shown in Figure 69. Each row in the repository report represents the best search result per intact ion in a search in the experiment.

Table 34 describes the columns displayed in the repository report. You can change the columns that are displayed in the repository report and the order of the columns.

Figure 69. Repository report on the Report (*Report_name*) page

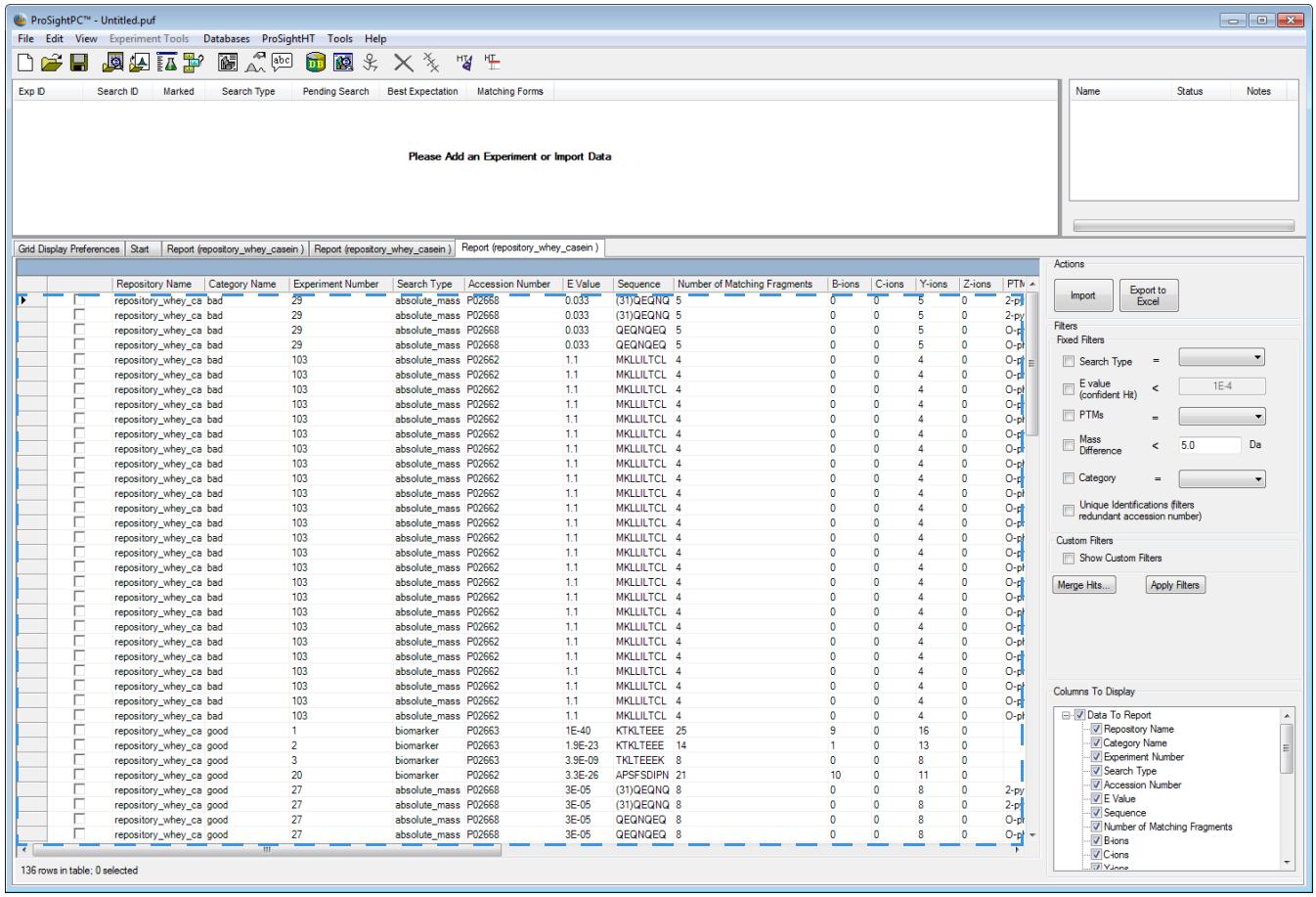
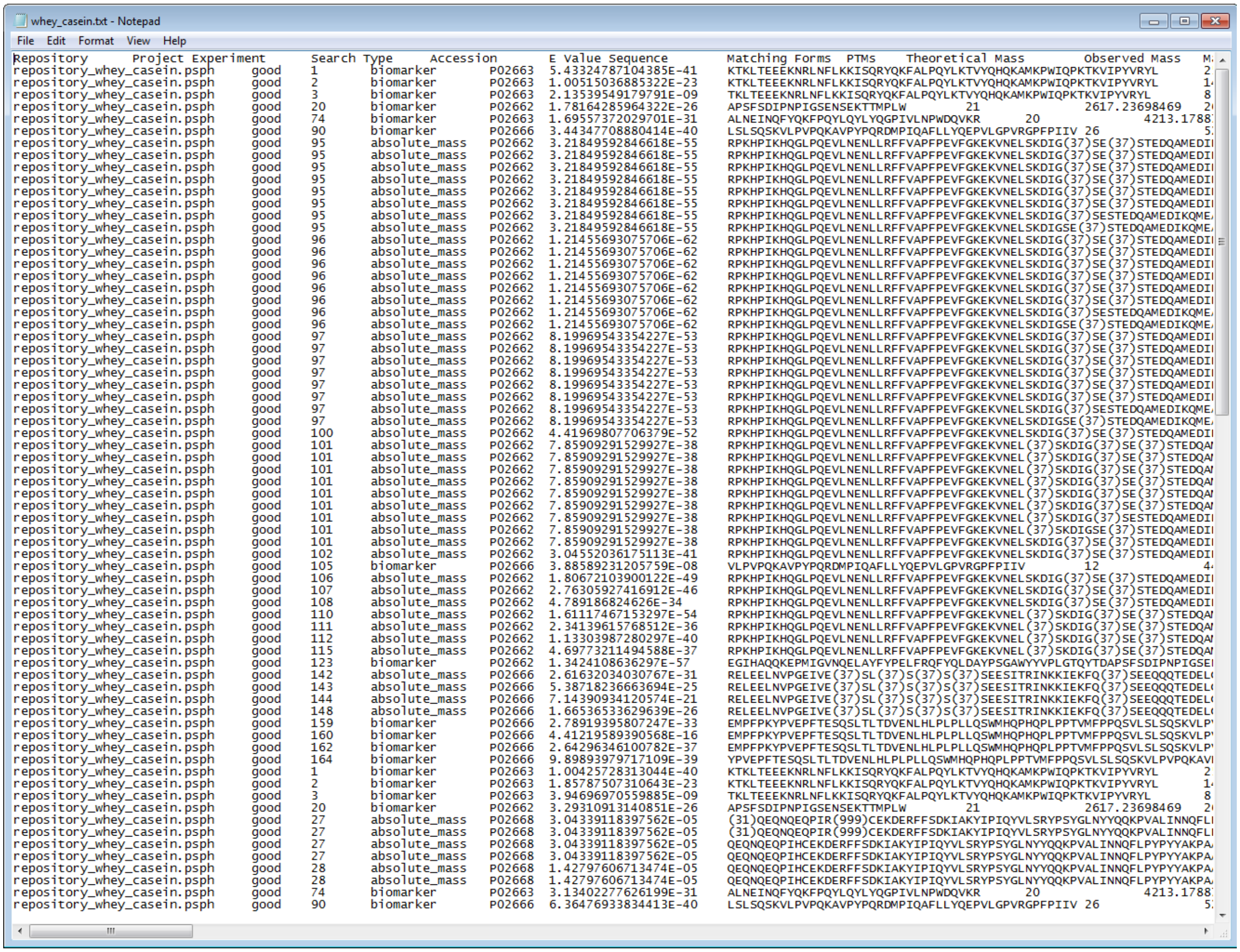


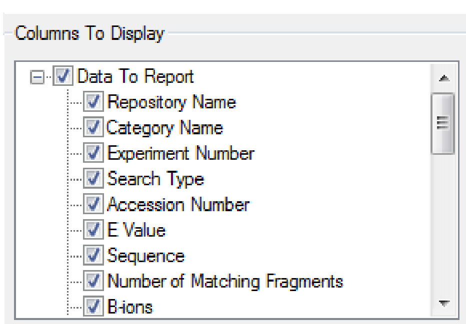
Figure 70. Repository report exported directly to a text file



❖ To change the columns in the repository report

Select or clear the check boxes in the Columns to Display area in the lower right corner of the Report (*Report_name*) page (see Figure 71).

Figure 71. Columns to Display area



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

Select the column header and move it to the applicable location.

Display Columns in the Repository Report

Table 34 lists the parameters in the repository report.

Table 34. Repository report columns (Sheet 1 of 2)

Parameter	Description
Check box	Selects or clears an experiment for export to an Excel spreadsheet.
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).
Accession Number	Displays the accession number used by the major protein databases, such as UniProt™, to index a protein in a database.
E Value	Specifies the expectation value (e value) of the best search result in the search. For more information on the e value, see "Expectation Value (e value)" on page 196.
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.

Table 34. Repository report columns (Sheet 2 of 2)

Parameter	Description
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 data or PUF file that an experiment was based on.
File Name	Displays the name of the raw data 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.
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 196.
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 196.
C Score	Displays the C score, which is a measure of the confidence in the characterization of the proteoform. For more information on the C score, see "C Score" on page 199.
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.

Repository Report Dialog Box Parameters

Table 35 lists the parameters in the Repository Report dialog box, shown in Figure 68 on page 175.

Table 35. Repository Report dialog box 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.
Files	Specifies the files to include in the report.
Select All	Selects all the listed files. This button is not available unless you select Files.
Unselect All	Clears all the listed files. This button is not available unless you select Files.
Text box (originally blank)	Specifies the text that is used to search for files in the list. If files in this list match the search text, they are selected for inclusion in the report.
Search	Searches for the text specified in the text box.
Unselect	Clears a particular search.
Export Directly to File	Exports a file without displaying the report.
Report Only Best Hit Per Search	Attempts to "break ties" in e values or p scores by examining the intact mass differences and choosing the one with the smallest mass difference.
Generate	Generates the repository report and displays it on the Report (*Report_name*) page, as shown in Figure 69 on page 177.

Using the Repository Report To Import Experiments from a Repository into the PUF File

The repository report page displays the applicable data from the repository, but the data is only for viewing. To manipulate the data, you must import the data into the PUF file, perform any appropriate 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 71.

❖ 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 the ProSightPC application.

   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 the data grid already contains experiments, 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 an Excel Spreadsheet

You can export experiments to an Excel spreadsheet file 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 check box to the extreme left of each experiment row.
   - 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.
3. In the Save As dialog box, type a file name in the File Name box, select **Excel File (*.xlsx)** in the Save as Type box, and click **Save**.
4. In the directory where you stored the .xlsx file, open the Excel spreadsheet by right-clicking the file name and choosing **Open**, or double-clicking the file name.

   The 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.

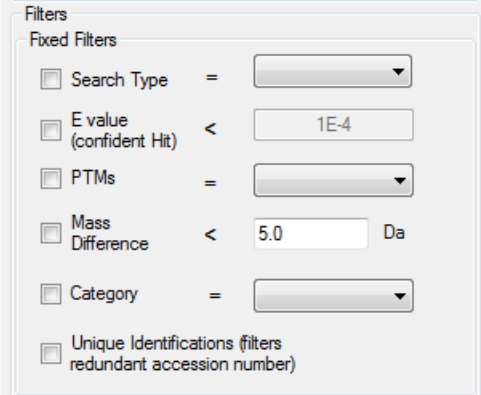
You can also set a tolerance that causes the ProSightPC application to merge matches that are very similar but differ by a small amount. This merging reduces the size of the data.

❖ To apply fixed filters

1. In the Fixed Filters section of the Actions area, shown in Figure 72, 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 196.
   - PTMs: Displays all the experiments with PTMs when you select Yes.
   - 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 the hit with the best e value for a set of hits with redundant accession numbers.

Figure 72. Fixed Filters section



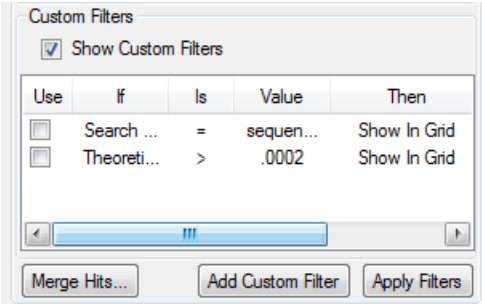
2. Click **Apply Filters**.

❖ To apply existing custom filters

1. In the Custom Filters section of the Actions area (see Figure 69 on page 177), select the **Show Custom Filters** check box.

   The Custom Filters section expands.

   Figure 73. 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

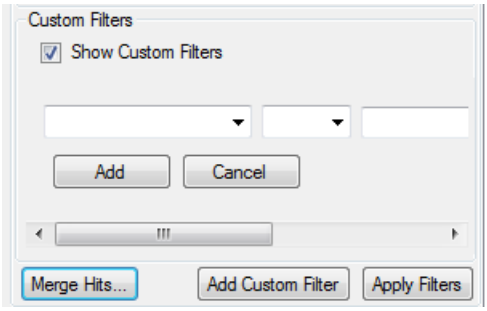
1. In the Custom Filters section of the Actions area (see Figure 69 on page 177), select the **Show Custom Filters** check box.

   The Custom Filter section expands, as shown in Figure 73.

2. Click **Add Custom Filter**.

   The Custom Filters section resembles Figure 74.

Figure 74. 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 "Display Columns in the Repository Report" on page 179.

4. From the middle list, select an operator:

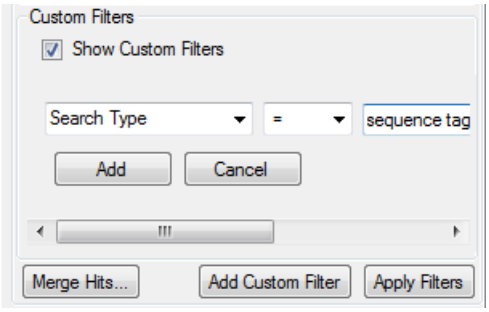
   - = Equal to
   - < Less than
   - > Greater than
   - <= Less than or equal to
   - >= Greater than or equal to
   - Not = Not equal to

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

6. Click **Add**.

   The Custom Filters section resembles the illustration shown in Figure 75.

   Figure 75. Completed Custom Filters section



7. To add another filter, click **Add Custom Filter** again or right-click and choose **New**.

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

9. Click **Apply Filters**.

❖ To remove a custom filter

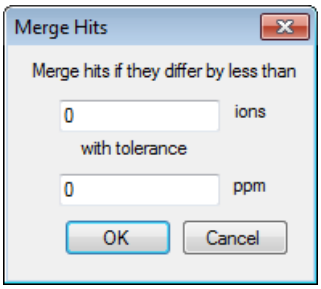
- In the Custom Filters area (see Figure 69 on page 177), select the filters that you want to delete, right-click, and choose **Remove**.

❖ To merge a set of matches with similar values

1. In the Custom Filters area (see Figure 69 on page 177), click **Merge Hits**.

   The Merge Hits dialog box appears, as shown in Figure 76.

   Figure 76. Merge Hits dialog box

2. In the box to the left of Ions, type the number of fragment ions by which the two matches can differ.
3. In the box to the left of Ppm, type a tolerance that the mass of the fragment ions must fall within for the sets of matches to be merged together.
4. Click **OK**.

To reduce redundancy, the ProSightPC application merges together a set of matches if the difference between the matches is fewer than the number of fragment ions specified and with an intact mass tolerance within the number of ppm specified.

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.

6

Searching for Single Proteins

This chapter describes how to use the Sequence Gazer to perform single-protein searches, add modifications, and display sequences and fragment maps. 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. The ProSightPC application stores each saved result as a single-protein mode search result.

For information on how to perform a basic single-protein search, see "Searching for Single Proteins" on page 146.

Contents

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, the ProSightPC application automatically scores the sequence on the basis of the initial search parameters. "Scores Box" on page 195 explains the scoring system in ProSightPC.

You can change parameters and add or remove PTMs or fixed modifications. The ProSightPC application then rescores the modified sequence. Ideally, changes to the sequence followed by rescoring yield more matching fragments than before, narrowing the possible matching proteoforms that explain the MS/MS data.

By rescoring, the ProSightPC application 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.

Using the Sequence Gazer

Follow these procedures to display and interpret sequences:

- Accessing the Sequence Gazer
- Navigating the Sequence Gazer
- Demonstrating the Sequence Gazer

Accessing the Sequence Gazer

You can access the Sequence Gazer through one of two strategies:

- By performing 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 206 for a demonstration showing how to use the Sequence Gazer to find modifications in fragment ions.

Searching for Single Proteins and Accessing the Sequence Gazer

Running a single protein search opens the Sequence Gazer, which you use to run the search. You can either run a Sequence Gazer search by running a single-protein search from the predefined search list or by taking an existing hit to the Sequence Gazer.

❖ **To perform a single-protein search and access the Sequence Gazer**

Follow the procedure in "Searching for Single Proteins" on page 146.

The Sequence Gazer window opens, as shown in Figure 77.

Figure 77. Sequence Gazer

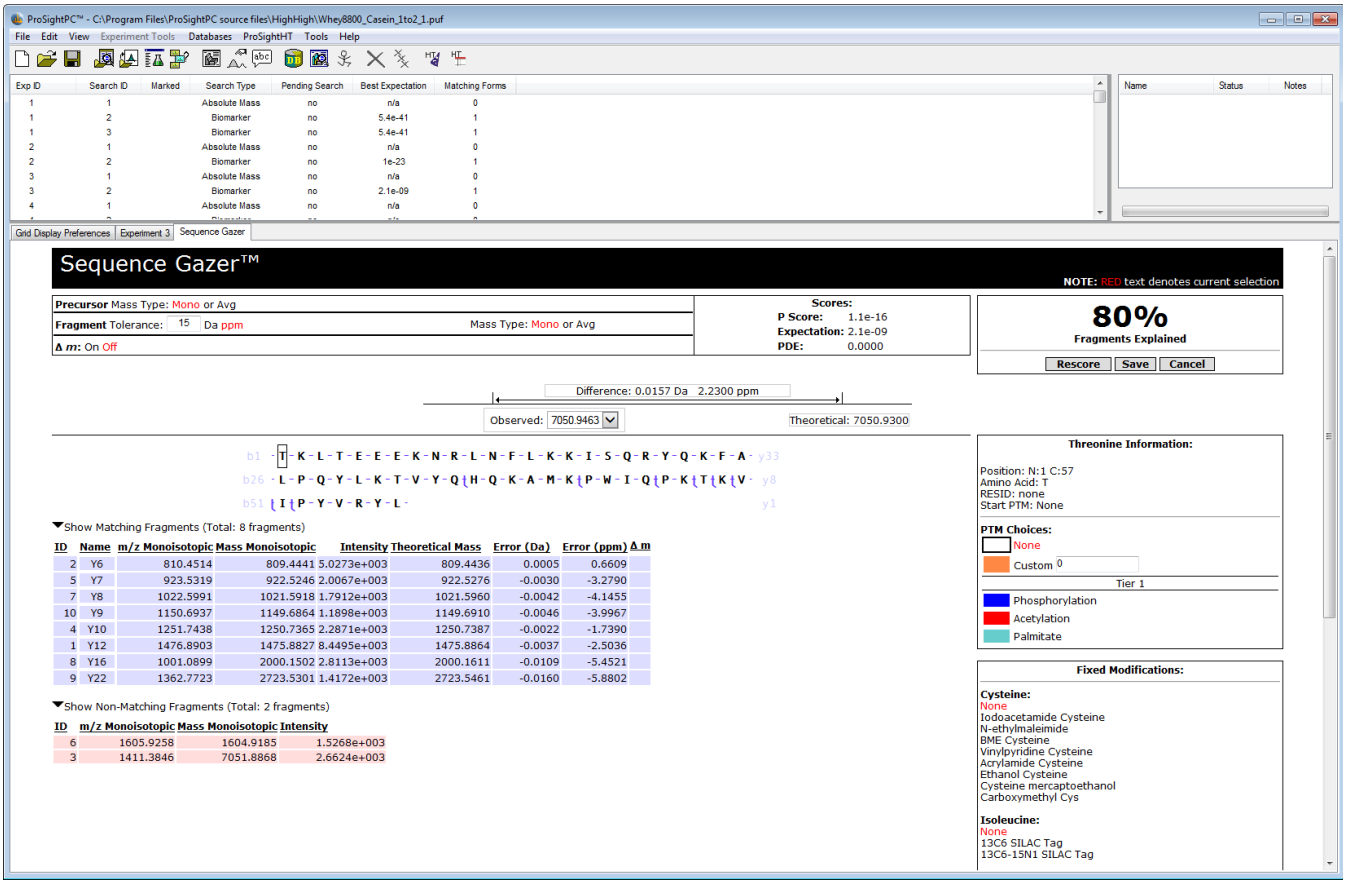


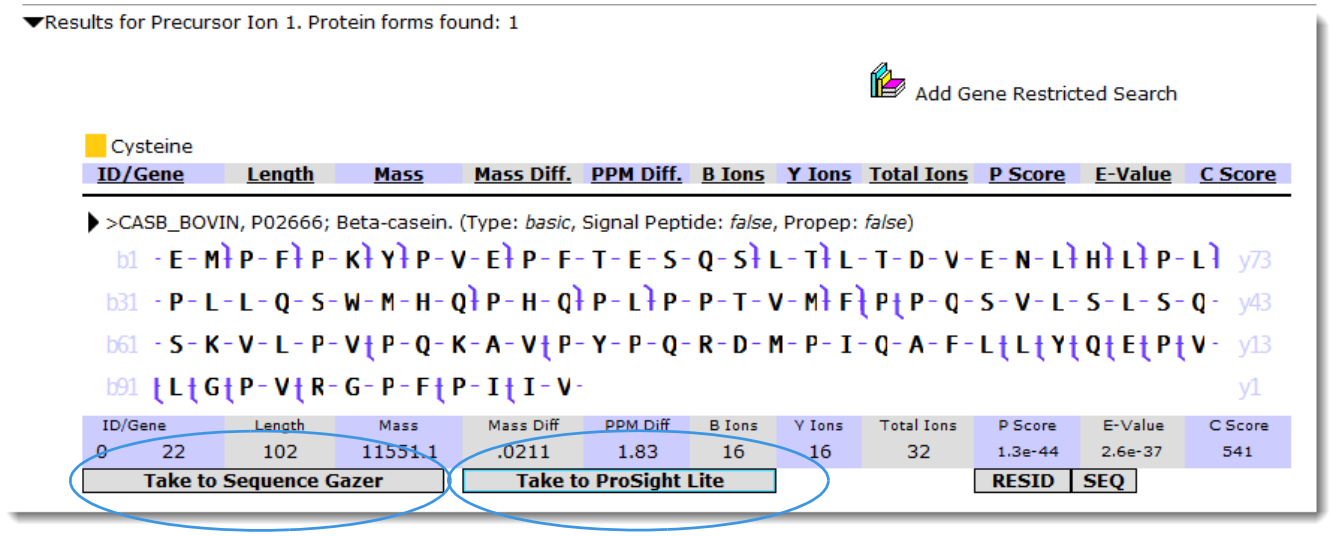
Figure 80 displays the features of the Sequence Gazer window, and "Navigating the Sequence Gazer" on page 193 explains them.

Identifying a Protein and Accessing the Sequence Gazer

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

1. Select the search and expand the view in the Data Manager.
2. Locate the applicable protein identification in the search results and expand its view.
3. Click **Take to Sequence Gazer** (see Figure 78).

Figure 78. Take to Sequence Gazer and Take to ProSight Lite buttons



Accessing ProSight Lite

ProSight Lite is a free and simplified version of the ProSightPC application that is mostly used for single protein analysis by infusion. You can use it to view the marked-up spectrum results and, as with the Sequence Gazer feature, edit the sites of modifications. You can also use it to produce publication-ready fragment maps.

Click the following link to access a video containing more information about ProSight Lite and to install the application:

http://prosightlite.northwestern.edu

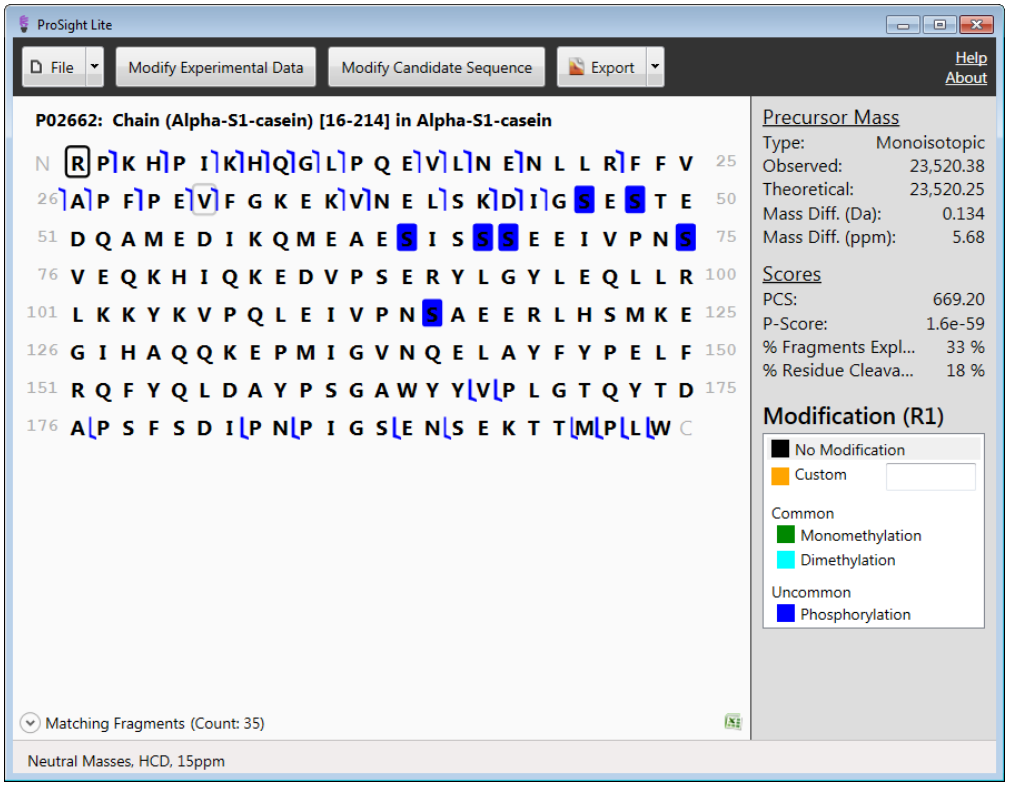
❖ To access ProSight Lite

1. In the Data Manager, select the appropriate search and expand the results view.
2. Locate the appropriate protein identification in the search results and expand its view.
3. Click **Take to ProSight Lite** (see Figure 78).

   You must have the ProSight Lite application installed before you click Take to ProSight Lite.

   The ProSight Lite window opens, as shown in Figure 79.

Figure 79. ProSight Lite window

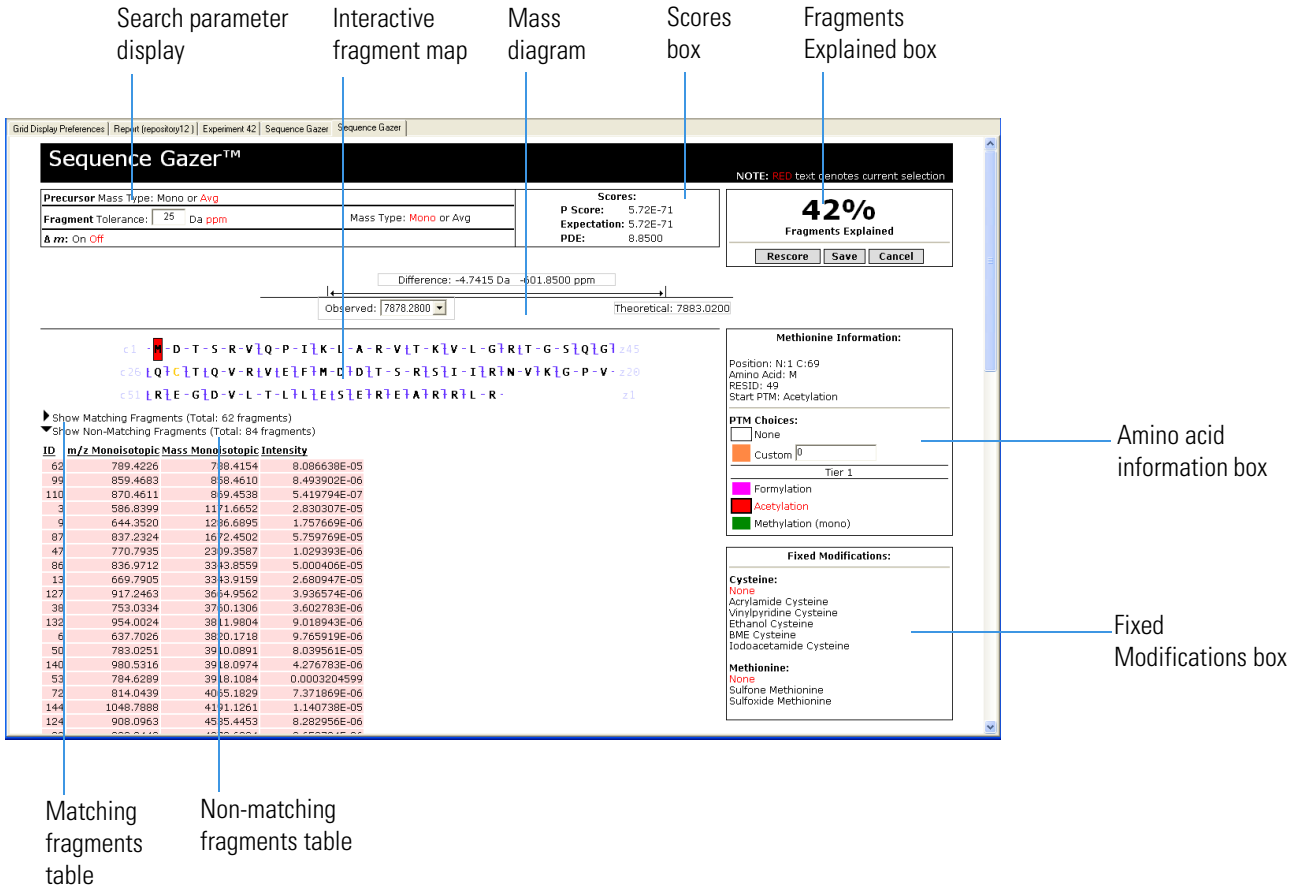


Navigating the Sequence Gazer

Fragment ion information in the Sequence Gazer interface is organized into the following regions, as shown in Figure 80:

- 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 80. 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.

You must click Rescore to implement the changes made in the search parameter display.

Table 36 lists the parameters in the search parameter display.

Table 36. Search parameter display 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 comparing 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-m (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

The ProSightPC application 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.

- C score, as noted by LeDuc, et al.[4]

These scores are described in the following topics.

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.

The ProSightPC application 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 proteoforms considered during a search, the e value currently reported by the ProSightPC application is

$$e = N \times p(n)$$

[3] Reid, G. E.; Shang, H.; Hogan, J. M.; Lee, G. U.; McLuckey, S. A. Gas phase concentration, purification, and identification of whole proteins from complex mixtures. *J. Am. Chem. Soc.* **2002**, *124*: 7353–7362.

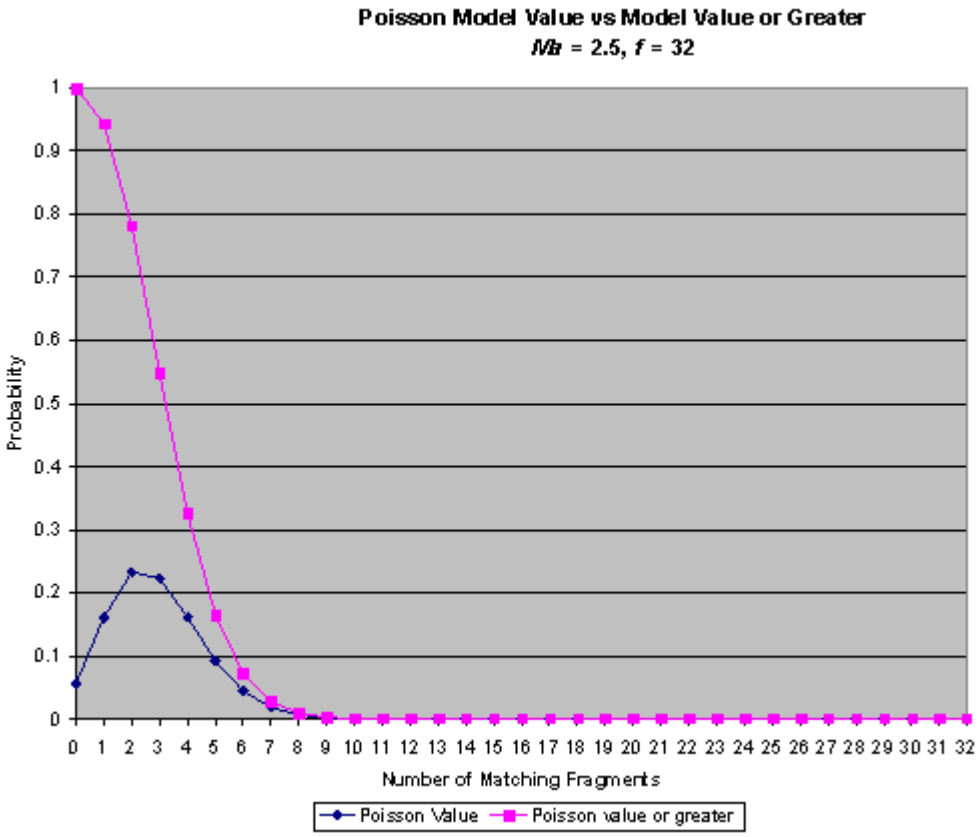
[4] LeDuc, R.D.; Fellers, R.T.; Early, B.P.; Greer, J.B.; Thomas, P.M.; Kelleher, N.L. The C-score: a Bayesian framework to sharply improve proteoform scoring in high-throughput top-down proteomics. *J Proteome Res.* **2014** Jul 3, 13 (7):3231-40.

Sample Calculation

Consider ubiquitin carboxyl-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 196 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 81 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 81. Poisson value versus Poisson value or greater



To find the P score for 12 matching fragments, sum the first 11 values as follows:

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

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

The ProSightPC application 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.

C Score

The C score measures the level of characterization of a proteoform in relation to the others in the PTM Warehouse. This score, described in LeDuc et al.[5], uses a Bayesian approach that assigns a likelihood to each candidate proteoform based on the observed MS data.

A C score of 3 indicates that there are two proteoforms in the database that equally explain the observed data. A score of 40 or higher is considered strong evidence of a unique characterization.

Sequence Tag Scores

The ProSightPC application 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 *i*th 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 the ProSightPC application 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.

[5] LeDuc, R.D.; Fellers, R.T.; Early, B.P.; Greer, J.B.; Thomas, P.M.; Kelleher, N.L. The C-score: a Bayesian framework to sharply improve proteoform scoring in high-throughput top-down proteomics. *J Proteome Res.* **2014** Jul 3, 13 (7):3231-40.

Fragments Explained Box

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

Table 37 lists the three additional controls in the The Fragments Explained box.

Table 37. Fragments Explained box parameters

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 boxes or lists described in Table 38.

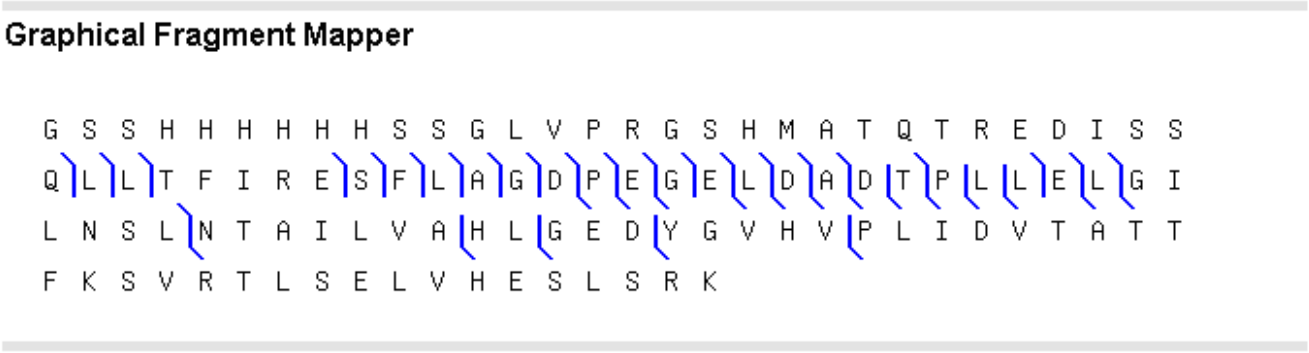
Table 38. Mass diagram parameters

Parameter	Description
Observed	Contains a list that displays all the precursor masses detected by the ProSightPC application.
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 82, is an interactive display of the protein sequence, along with any PTMs, and the matching fragment information.

Figure 82. 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.

Tip 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

Table 39 lists the parameters in the Amino Acid Information box, which refers to the selected amino acid in the interactive fragment map.

Table 39. Amino Acid Information box parameters (Sheet 1 of 2)

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.

Table 39. Amino Acid Information box parameters (Sheet 2 of 2)

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 appropriate 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 the ProSightPC application by amino acid.

You can select fixed modifications in the dialog box by choosing Tools > Fixed Modification Editor (see "Editing Modifications" on page 25), or you can change them during rescoring.

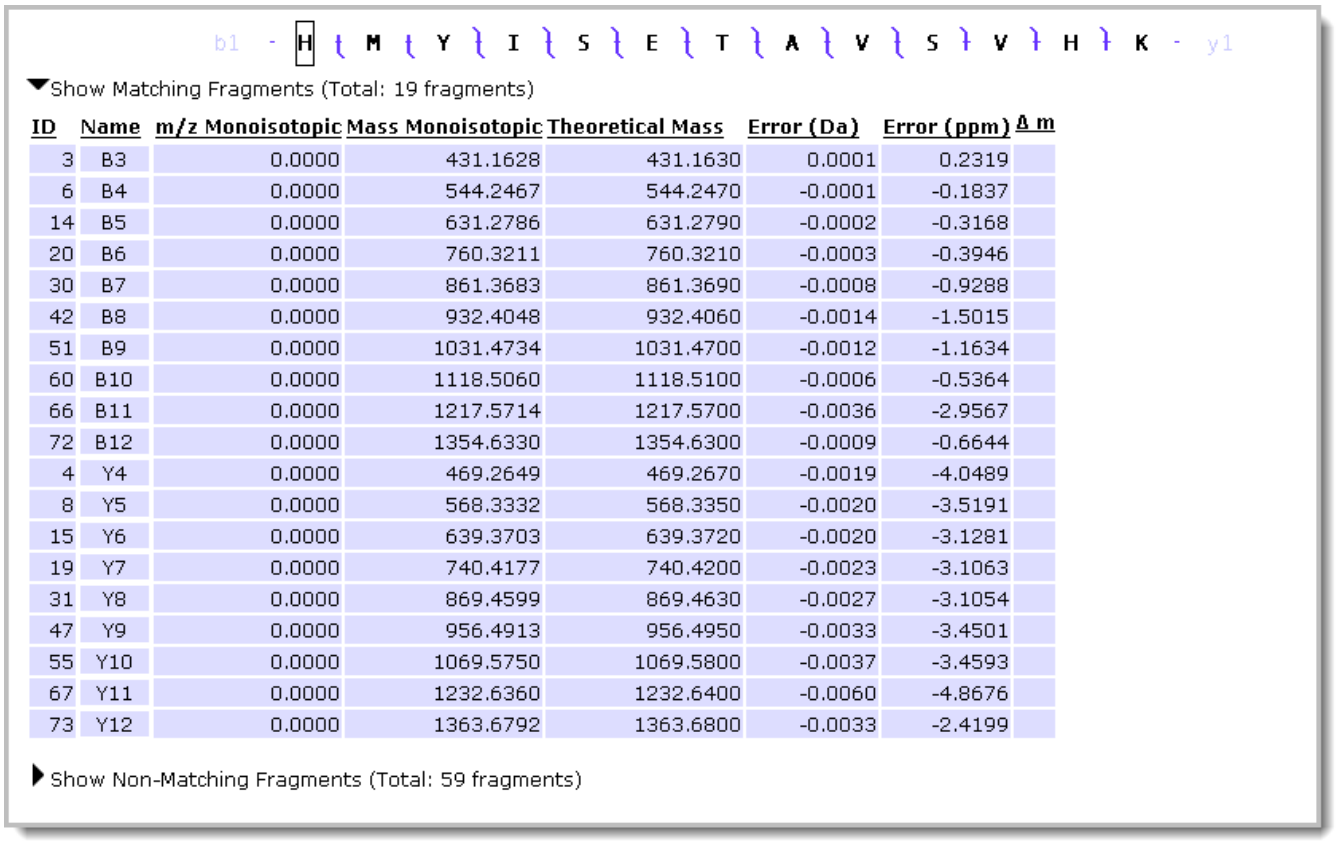
To indicate that no fixed modifications are presently selected for that type of amino acid and will not be included in the next rescoring, select **None**.

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

Matching Fragments Table

The matching fragments table, shown in Figure 83, contains a summary of all fragment ions matching the protein.

Figure 83. Matching fragments table



ID	Name	m/z Monoisotopic	Mass Monoisotopic	Theoretical Mass	Error (Da)	Error (ppm)	Δ m
3	B3	0.0000	431.1628	431.1630	0.0001	0.2319	
6	B4	0.0000	544.2467	544.2470	-0.0001	-0.1837	
14	B5	0.0000	631.2786	631.2790	-0.0002	-0.3168	
20	B6	0.0000	760.3211	760.3210	-0.0003	-0.3946	
30	B7	0.0000	861.3683	861.3690	-0.0008	-0.9288	
42	B8	0.0000	932.4048	932.4060	-0.0014	-1.5015	
51	B9	0.0000	1031.4734	1031.4700	-0.0012	-1.1634	
60	B10	0.0000	1118.5060	1118.5100	-0.0006	-0.5364	
66	B11	0.0000	1217.5714	1217.5700	-0.0036	-2.9567	
72	B12	0.0000	1354.6330	1354.6300	-0.0009	-0.6644	
4	Y4	0.0000	469.2649	469.2670	-0.0019	-4.0489	
8	Y5	0.0000	568.3332	568.3350	-0.0020	-3.5191	
15	Y6	0.0000	639.3703	639.3720	-0.0020	-3.1281	
19	Y7	0.0000	740.4177	740.4200	-0.0023	-3.1063	
31	Y8	0.0000	869.4599	869.4630	-0.0027	-3.1054	
47	Y9	0.0000	956.4913	956.4950	-0.0033	-3.4501	
55	Y10	0.0000	1069.5750	1069.5800	-0.0037	-3.4593	
67	Y11	0.0000	1232.6360	1232.6400	-0.0060	-4.8676	
73	Y12	0.0000	1363.6792	1363.6800	-0.0033	-2.4199	

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.

Table 40 describes the columns in the matching fragments table.

Table 40. Matching fragments table columns (Sheet 1 of 2)

Column	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.

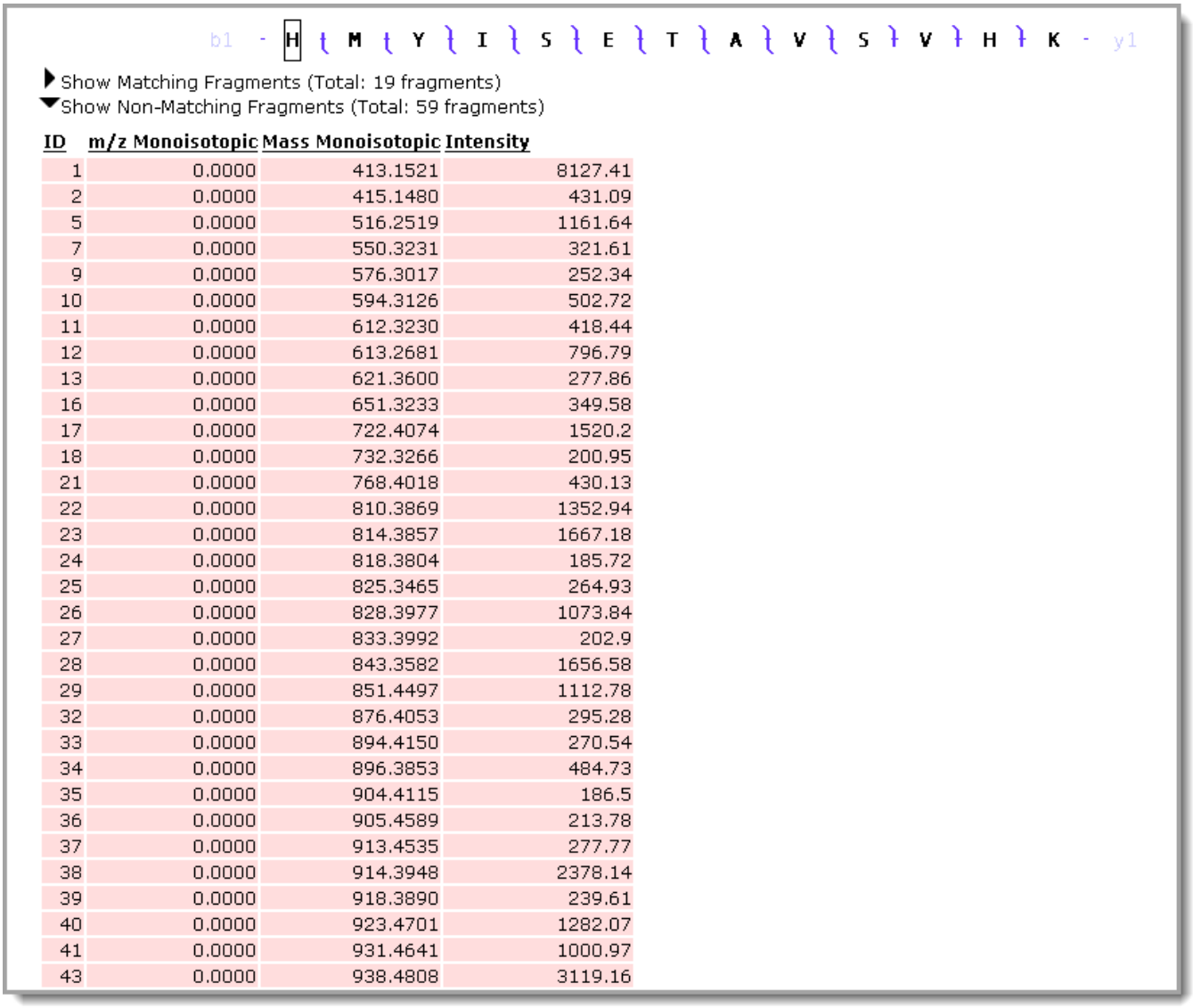
Table 40. Matching fragments table columns (Sheet 2 of 2)

Column	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.
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 84, lists every fragment that does not match the sequence.

Figure 84. Non-matching fragments table



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 columns shown in Table 41.

Table 41. Non-matching fragments table parameters (Sheet 1 of 2)

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.

Table 41. Non-matching fragments table parameters (Sheet 2 of 2)

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

The following demonstration shows how to use the Sequence Gazer to turn a good result from 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.

7

Displaying Data in the Data Manager

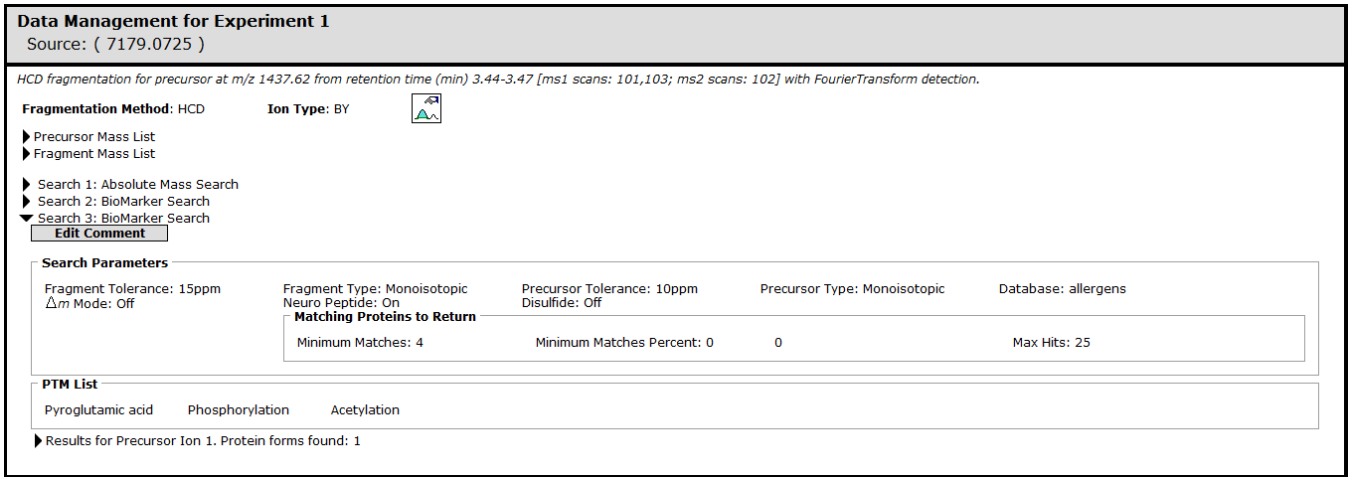
This chapter describes the Data Manager in the ProSightPC graphical user interface.

Contents

Data Manager

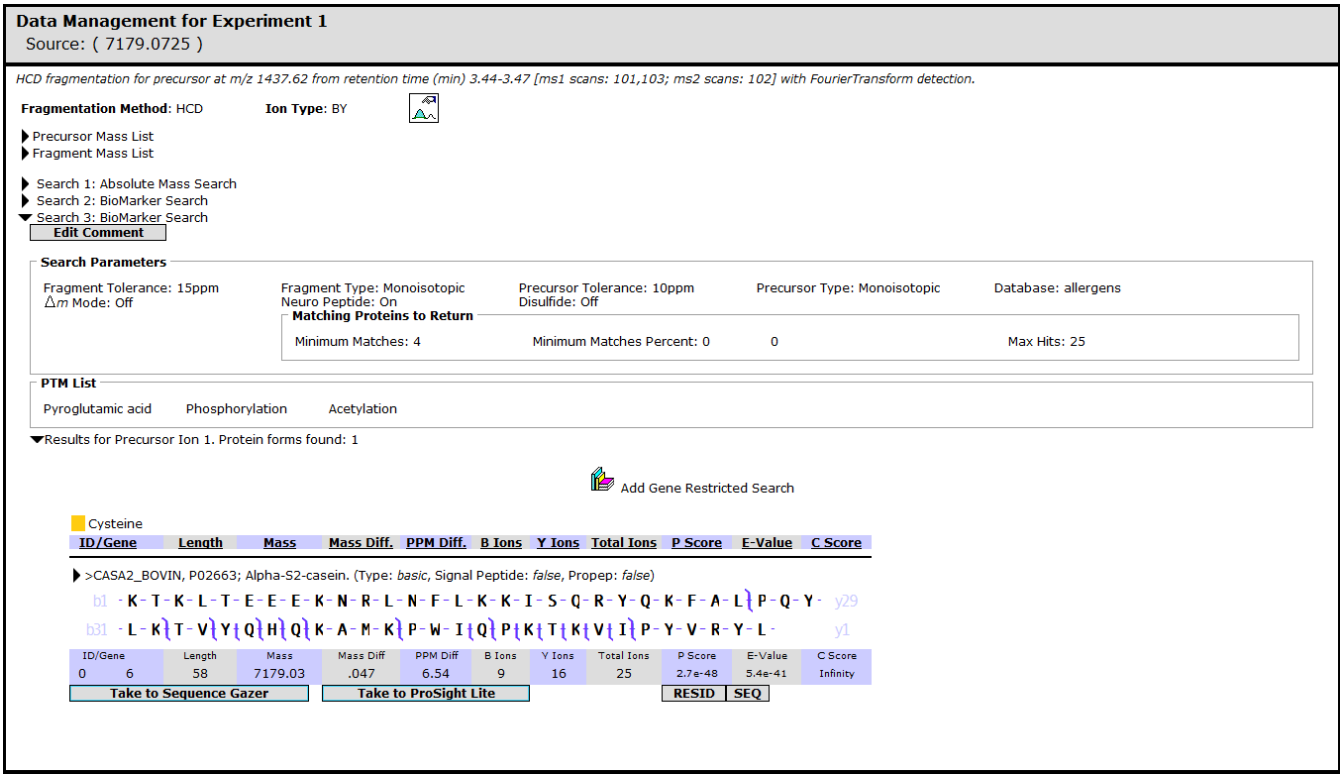
The Data Manager, shown in Figure 85, provides a visual representation of all the information related to a single MS/MS experiment. It appears when you double-click an experiment in the data grid. You can use it to view all information for a single experiment. The context-sensitive controls help you determine what information is displayed.

Figure 85. Data Manager



Each side arrow indicates hidden information related to a search. Expand the view to display detailed information, as shown in Figure 86.

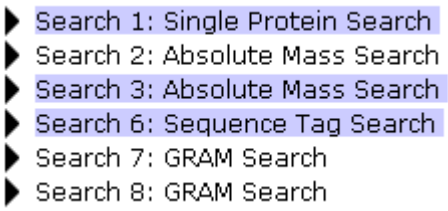
Figure 86. Expanded view in the Data Manager



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. Figure 87 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 87. Search numbers

When you generate an MS/MS experiment, the ProSightPC application declares the fragmentation method used to generate the MS/MS data. From this input, it determines which of the following ion types to use during searches:

- Electron capture dissociation (ECD and ETD) fragmentation is analyzed using c and z ions.
- Collision-induced dissociation (CID, HCD, and IRMPD) fragmentation is analyzed using b and y ions.
- Ultraviolet photodissociation (UVPD) fragmentation is analyzed using a, a+, x, x+, b, y, y–, c, and z˙ ions.
- Electron transfer higher-energy collision dissociation (EThcD) fragmentation is analyzed using b, c, y, and z˙ ions.

Opening a Data Manager Window

❖ **To open a Data Manager window**

Do one of the following:

Double-click an experiment displayed in the data grid.

–or–

Select the experiment and choose **View > Open Data Manager**.

–or–

Click the **Open Data Manager** icon,  .

Closing a Data Manager Window

❖ **To close a Data Manager window**

Choose **View > Close Data Tab**.

The experiment page and all pages related to it, such as the Sequence Gazer, disappear from the screen.

❖ **To close all Data Manager windows**

Choose **View > Close All Data Tabs**.

❖ **To close all Data Manager windows except for the currently selected window**

Choose **View > Close All Data Tabs But Selected**.

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,  .
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.

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. Do one of the following:
   - Choose **Experiment Tools > Edit Masses**.

   –or–

   - 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 an Edit Masses page in the application window, as shown in Figure 88.

Figure 88. Editing mass values in the Edit Masses page

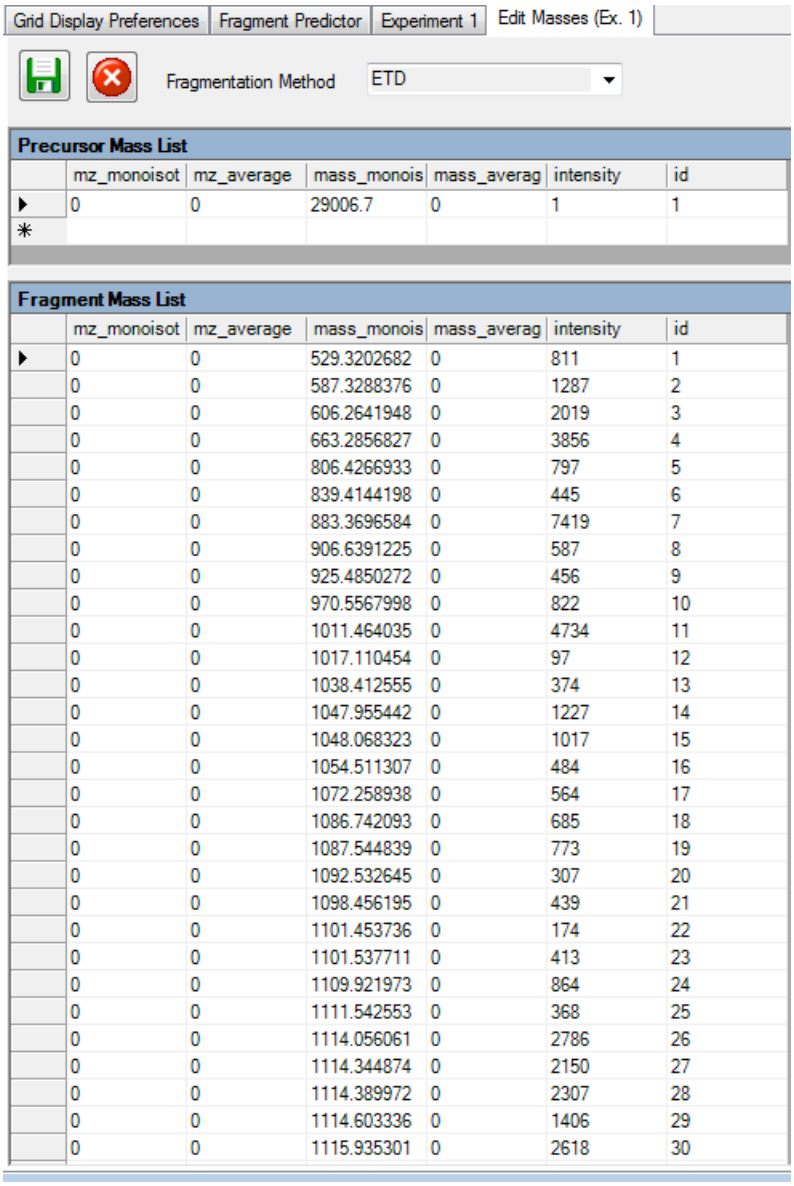


Table 42 lists the parameters and icons at the top of the Edit Masses *experiment_number* page.

Table 42. Edit Masses *experiment_number* page parameters and icons

Parameter	Description
	Saves the edits that you made to the mass values.
	Does not save any of the edits that you 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 9.

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.

Table 43 lists the columns in the Precursor Mass List area.

Table 43. Precursor Mass List area columns

Column	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.

Table 44 lists the columns in the Fragment Mass List area.

Table 44. Fragment Mass List area columns (Sheet 1 of 2)

Column	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.

Table 44. Fragment Mass List area columns (Sheet 2 of 2)

Column	Description
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  to save the changes.
5. In the Save Masses Confirmation box, click **Yes**.

   The ProSightPC application 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 appropriate values.
3. Click  to save the changes.
4. In the Save Masses Confirmation box, click **Yes**.

   The ProSightPC application 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  to save the changes.
4. In the Save Masses Confirmation box, click **Yes**.

   The ProSightPC application 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 applicable row, and then click in the leftmost column of the last applicable row.
   - To select disparate rows, hold the CTRL key down and click in the leftmost column of each applicable row.
2. Press CTRL+C and select the external application to paste.

❖ To change the fragmentation method

1. In the Fragmentation Method list, select the new fragmentation method.

   For more information on fragmentation methods, see Fragmentation Methods.
2. Click  to save the changes.
3. In the Save Masses Confirmation box, click **Yes**.

   The ProSightPC application returns you to the Data Manager.

❖ To return to the Data Manager without applying any changes to the mass values

Click  .

The ProSightPC application does not save any of the edits that you made to the mass values. It closes the Edit Masses *experiment_number* page and returns you to the Data Manager.

Running a Pending Search

❖ To run a pending search in the Data Manager

1. Expand the view of a pending search to reveal the search parameters and a Run Search button.
2. Click **Run Search** to run the search.

For additional information on search parameters, see "Searching Databases" on page 103.

A completed search generates a results list in the Data Manager, as shown in Figure 62 on page 168. For information on these results, see "Viewing the Results in the Data Manager" on page 167.

8

Using Proteome Databases

This chapter describes the proteome warehouse and how to create, manipulate, and modify proteome databases.

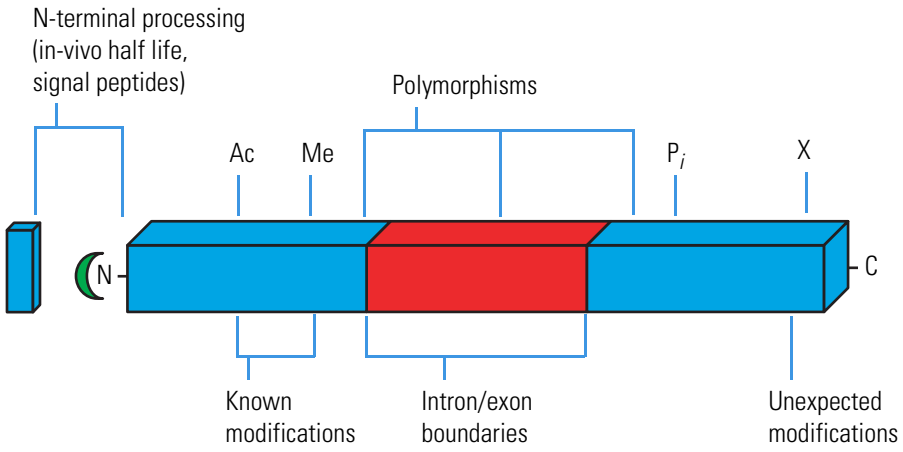
Contents

Proteome Warehouse

The ProSightPC application searches require sequence information to identify and characterize proteins. This sequence information and modification information are stored in proteome databases in the ProSightPC proteome warehouse. The sequence and PTM information are combined using shotgun annotation, as explained in "Introduction to the ProSightPC Application" on page 1.

Figure 89 shows all the known modifications, such as SNPs and sequence variants, that can be applied to a basic sequence.

Figure 89. Known modifications applied to a basic sequence



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.

Importing Data into the Proteome Warehouse

Use either of the following two methods to import data into the proteome warehouse:

- Load databases from PSCW or XML-format files. See "Importing a Proteome Database or Repository" on page 216 for details.
- Create databases from UniProtKB or FASTA-formatted text files. See "Creating a Proteome Database" on page 217 for details.

Importing a Proteome Database or Repository

You can load an existing proteome database in PSCW or XML file format or a repository in PSPH format into the ProSightPC application proteome warehouse.

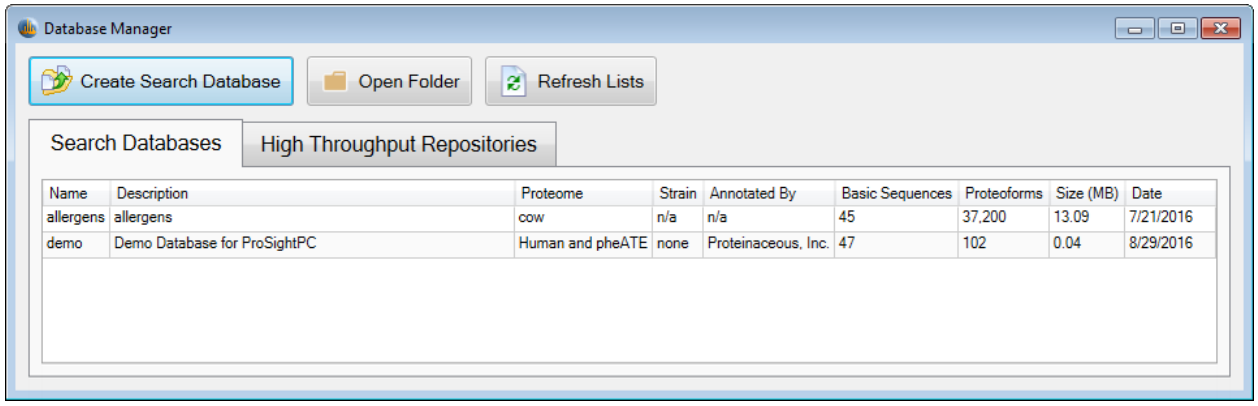
❖ To import a proteome database or repository

1. Copy a database (*database_name*.pscw file or *database_name*.xml file) or a repository (*database_name*.psph file) from an outside source.
2. Choose **Databases > Database Manager**, or click the **View Database Info** icon,  .

   The Database Manager window opens with the Search Databases page displayed by default, as shown in Figure 90.

   If the Search Databases page is not displayed, click the **Search Databases** tab.

Figure 90. Database Manager window



3. Click Open Folder in the toolbar.

   The default database folder storage location is as follows:

   C:\Users*your_name_folder*\Downloads

4. If you do not want to place the imported PSCW, XML, or PSPH file in the default folder, navigate to the appropriate folder.

5. Right-click the selected folder and choose **Paste**.

> **Note** Importing databases as PSCW files is faster than recreating a proteome database with the Create Proteome Database option, a process detailed in "Creating a Proteome Database."

Creating a Proteome Database

The ProSightPC application 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 three input file formats:

- (Recommended) UniProtKB XML files, which store a large amount of known modification information. UniProtKB is a curated biological database of protein sequences provided by the UniProt consortium.
- UniProtKB flat files, which contain information similar to that in UniProtKB XML files. The flat file format is included for backward compatibility, but UniProtKB XML files are recommended for the most faithful annotation of proteome databases.

  All PTMs listed in RESID are available for shotgun annotation. However, the ProSightPC application can only process the PTM information in the source UniProtKB flat files or XML files into a 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 initiated by a > character, followed by lines of sequence data. This format uses single-letter codes to represent base pairs or amino acids.

IMPORTANT Creating a proteome database can take several minutes or, in some extreme cases, hours to complete, depending on your hardware and the data being processed by the shotgun-annotation method.

❖ To create a proteome database with the Database Manager

1. Choose **Databases > Create a Custom Database**, or click 

in the Database Manager.

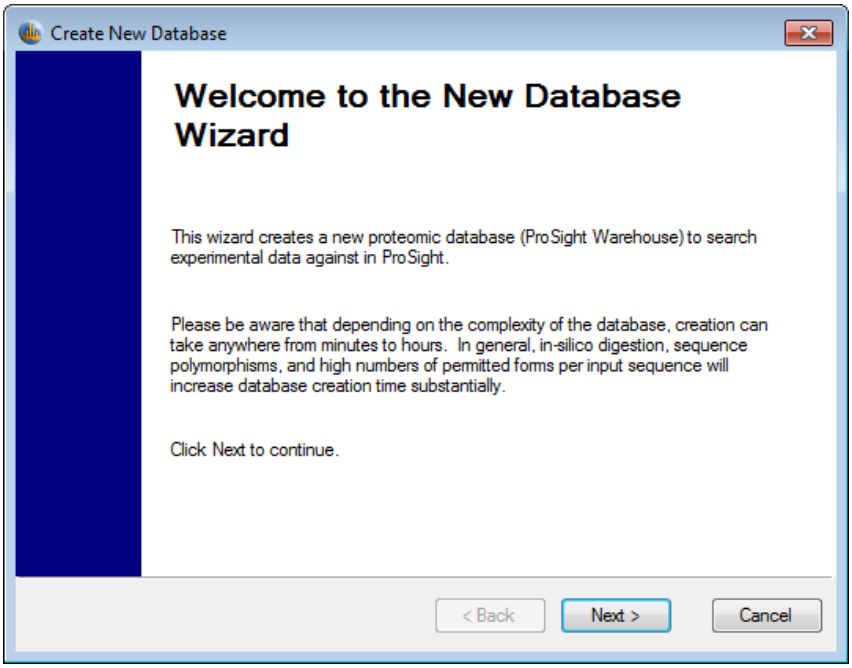
   –or–

   Choose **Databases > Database Manager**, and then click 

.

   The Welcome to the New Database Wizard page of the Create New Database Wizard opens, as shown in Figure 91.

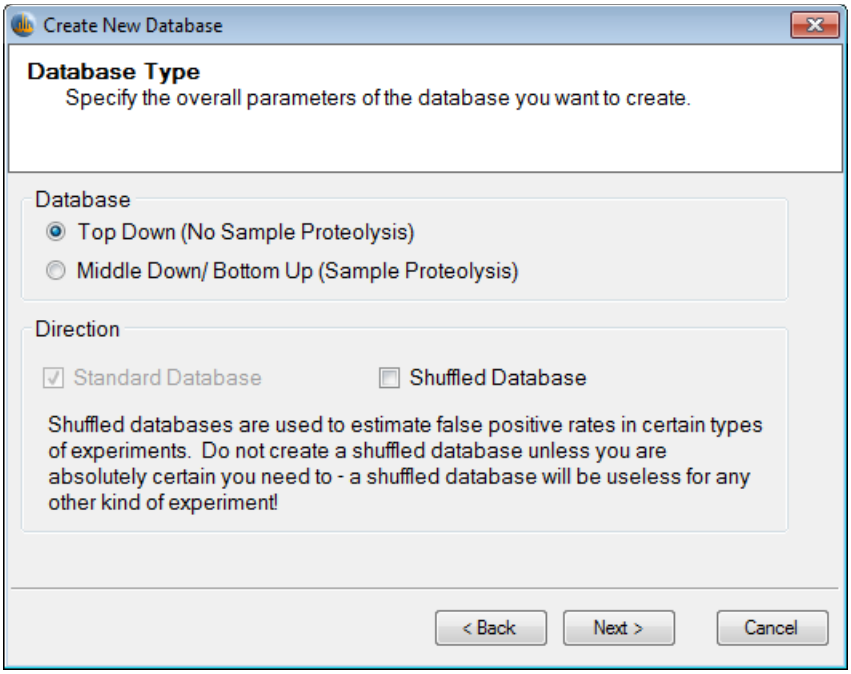
Figure 91. Welcome to the New Database Wizard page of the Create New Database Wizard



2. Click **Next**.

   The Database Type page of the Create New Database Wizard appears, as shown in Figure 92.

   Figure 92. 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.

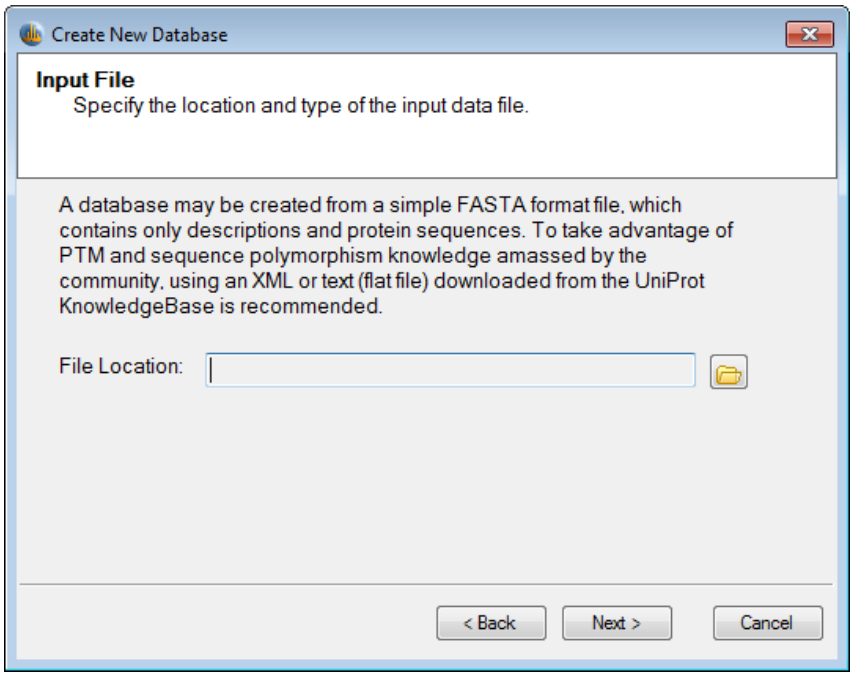
4. In the Direction area, select the direction of the database to build. You can select either of the following options:

   - Standard Database: Creates a database consisting of correct masses and forward sequences. A standard database is a typical protein database.
   - Shuffled Database: Creates a nonsense database consisting of correct masses but sequences with randomized letters. 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 93.

Figure 93. Input File page of the Create New Database Wizard



6. In the File Location box, type the name and path of the file containing the sequence information, or browse to it by clicking the **Browse Folder** icon.

   In the Open dialog box activated by the Browse Folder icon, you can select a FASTA file, a UniProtKB XML file, or a UniProtKB flat file for the input file. For descriptions of these files, see "Creating a Proteome Database" on page 217.

   The ProSightPC application 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 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 different text editor.

   The name of the input file must be unique.

   Once you select the file, the Input File page might display one or more of the annotations shown in Table 45. Figure 94 gives an example.

Figure 94. Annotated Input File page

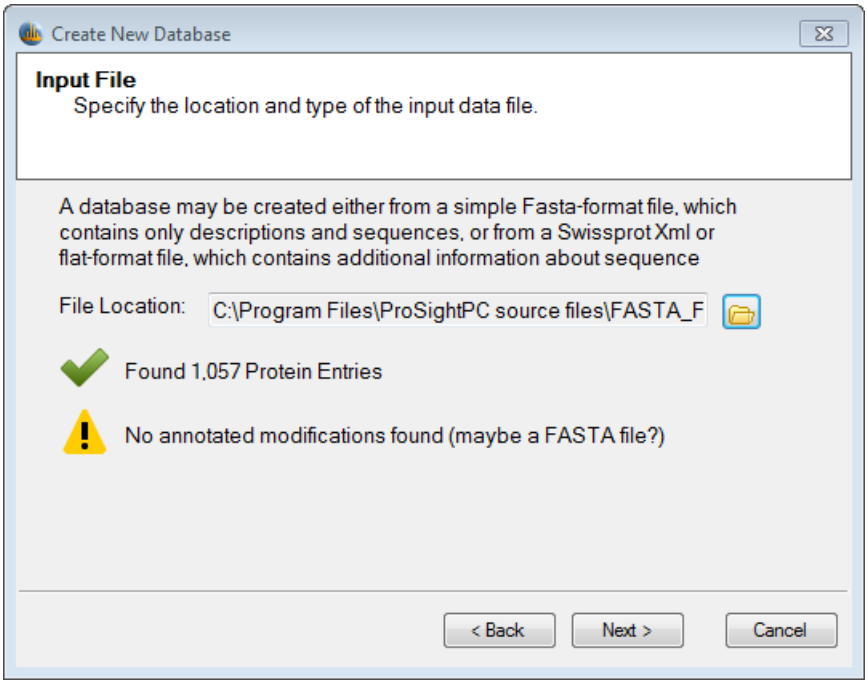


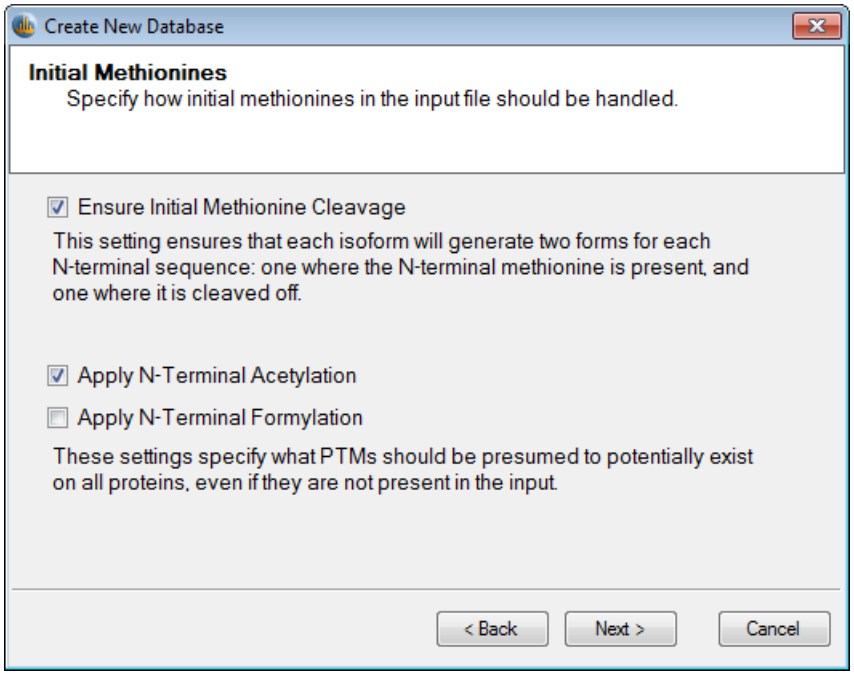
Table 45. Symbols on the Input File page

Symbol	Meaning
	The input file contains no sequence information and cannot be used to create a database.
	The input FASTA, UniProtKB XML, or UniProtKB flat file contains enough sequence information to build a database. The number of protein entries following the check mark indicates the size of the database.
	The input FASTA or UniProtKB file contains no annotation modifications, which might indicate a formatting issue. In the case of a FASTA file, this symbol indicates that the modifications are not recorded in the database. This symbol is only a warning and does not prevent you from proceeding to the next page of the Create Database Wizard.

7. Click **Next**.

The Initial Methionines page of the Create New Database Wizard appears, as shown in Figure 95.

Figure 95. Initial Methionines page of the Create New Database Wizard



8. Select the method of handling initial methionines:

   - Ensure Initial Methionine Cleavage (recommended): Determines whether each isoform generates two proteoforms for each N-terminal sequence: one where the N-terminal methionine is present and one where it is cleaved off.
     - (Default) Selected: Each isoform generates two proteoforms for each N-terminal sequence.
     - Clear: Each isoform generates only the proteoform where the N-terminal methionine is present.

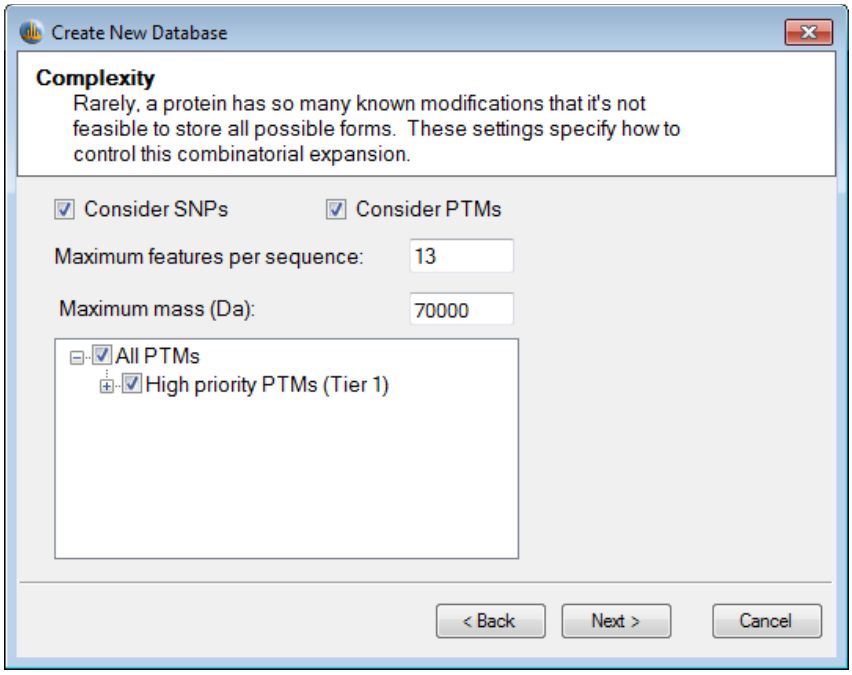
9. Specify the PTMs that are presumed to exist on all proteins, even if the input does not contain them:

   - Apply 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.
   - Apply N-Terminal Formylation: Adds N-terminal formylation. Select this option if you are building a prokaryotic database. Prokaryotes use N-formylmethionine for initiation.

10. Click **Next**.

    The Complexity page of the Create New Database Wizard appears, as shown in Figure 96.

Figure 96. Complexity page of the Create New Database Wizard



As indicated on the Complexity page, a protein might have so many known modifications that it is not feasible to store all possible proteoforms. 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.

11. If you want to include potential genetic variation as annotated in the UniProt database, select the **Consider SNPs** check box.

    This option enables you to incorporate sequence polymorphisms into the database.

12. If you want to annotate known post-translational modifications (PTMs) onto a protein, select the **Consider PTMs** check box.

13. In the Maximum Features Per Sequence box, type the maximum number of features per input sequence.

    This option sets the maximum number of database proteoforms produced from a particular entry. If a protein has, for example, four PTMs, the ProSightPC application puts 2^4, or 16, proteoforms 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 proteoforms 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). Proteoforms 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 proteoforms, the proteoforms 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 proteoform 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 proteoforms are produced: one where the site *is* modified and one where the site *is not* modified. If more known sites were known, database proteoforms would be produced with all combinations.

14. In the Maximum Mass to Annotate Features (Da) box, enter the upper limit for which PTMs are included in the database.

    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.

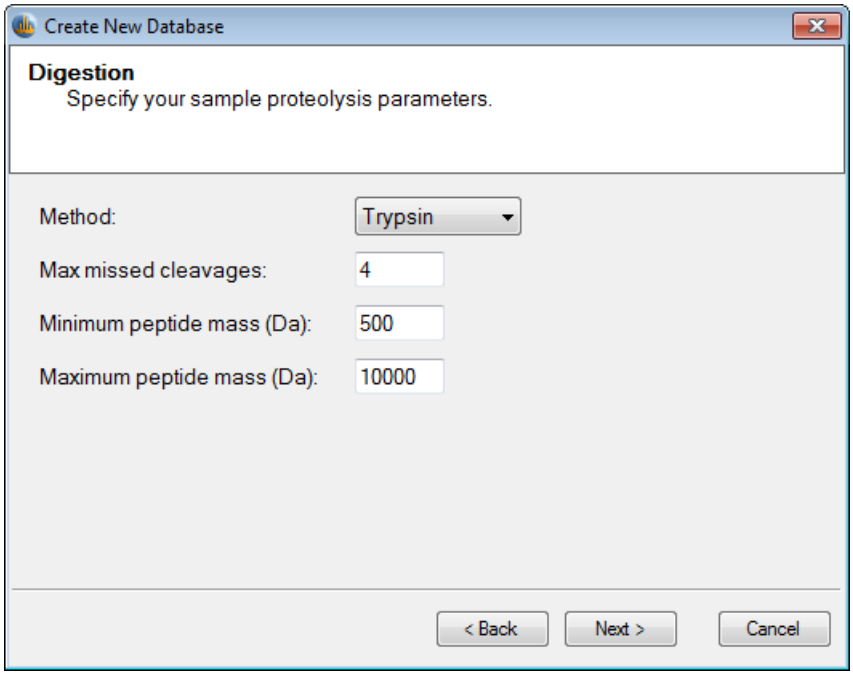
15. In the PTM Selection area, specify which PTMs should be considered for inclusion in the database.

    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 UniProtKB flat-formatted input data, because the standard FASTA format cannot encode information about PTMs.

16. Click **Next**.

    If you selected the Middle Down (Sample Proteolysis) option on the Database Type page of the Wizard, shown in Figure 92 on page 219, the Digestion page of the Create New Database Wizard opens, as shown in Figure 97. Use it to specify the parameters for a sample proteolysis.

Figure 97. Digestion page of the Create New Database Wizard



17. In the Method list, select the proteolytic method used to catalyze the breakdown of proteins into peptides.

18. 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

19. In the Minimum Peptide Mass (Da) box, type the minimum mass that a peptide must have, in daltons, before the application puts it 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.

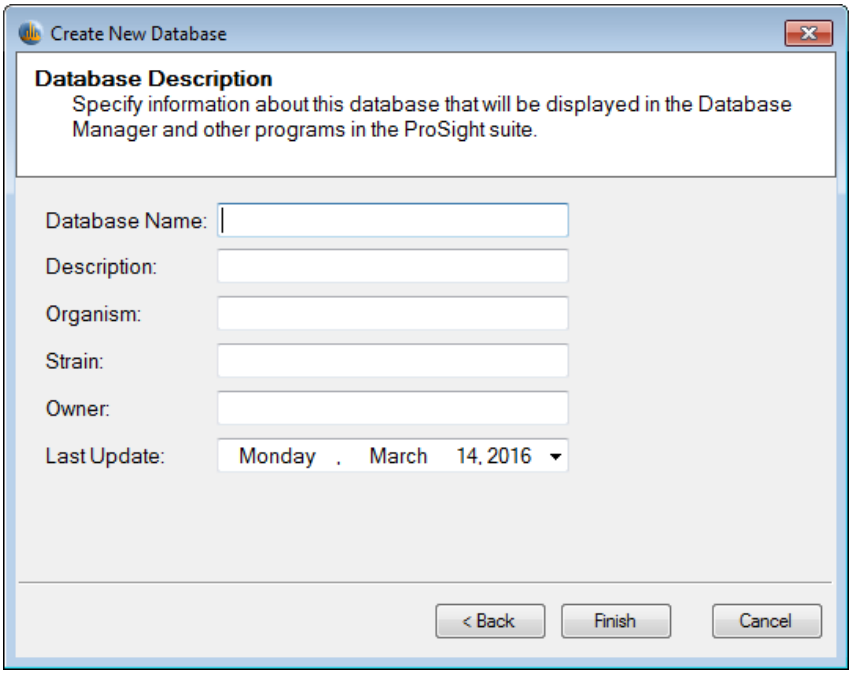
20. In the Maximum Peptide Mass (Da) box, type the maximum mass that a peptide can have, in daltons, before the application puts it into the database.

    No peptide greater than this mass is put into the database.

21. Click **Next**.

    The Database Description page of the Create New Database Wizard opens, as shown in Figure 98, so that you can enter identifying information about the database that you want to create.

Figure 98. 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.

22. Click **Finish**.

23. 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

Table 46 lists the parameters on the Database Type page of the Create New Database Wizard, shown in Figure 92 on page 219.

Table 46. Database Type page parameters

Parameter	Description
Database	Specifies the type of database to create.
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/ Bottom Up (Sample Proteolysis)	Builds a database around peptide sequences that arose from ex vivo proteolysis.
Direction	
Standard Database	Creates a database consisting of correct masses and forward sequences. A standard database is a typical protein database.
Shuffled Database	Creates a nonsense database consisting of correct masses but sequences with randomized letters. Do not select this option unless it is absolutely necessary. You cannot use a reverse database in any other kind of experiment.

Input File Page Parameters

Table 47 describes the sole parameter on the Input File page of the Create New Database Wizard, shown in Figure 93 on page 220.

Table 47. Input File page parameters

Parameter	Description
File Location	Specifies the name and path of the file containing the sequence information. The Browse Folder icon opens the Open dialog box so that you can browse for the input file. For the file type, you can select a FASTA file, a UniProKB XML file, or a UniProtKB flat file. For a description of the file types, see the beginning of "Creating a Proteome Database" on page 217.

Initial Methionines Page Parameters

Table 48 lists the parameters on the Initial Methionines page of the Create New Database Wizard, shown in Figure 95 on page 222.

Table 48. Initial Methionines page parameters (Sheet 1 of 2)

Parameter	Description
Ensure Initial Methionine Cleavage	Determines whether each isoform generates two proteoforms for each N-terminal sequence: one where the N-terminal methionine is present and one where it is cleaved off. • (Default) Selected: Each isoform generates two proteoforms for each N-terminal sequence. • Cleared: Each isoform generates only the proteoform where the N-terminal methionine is present.

Table 48. Initial Methionines page parameters (Sheet 2 of 2)

Parameter	Description
Apply N-Terminal Acetylation	Determines whether the ProSightPC application adds acetylation to N-terminal methionines whenever it is possible, regardless of whether the input includes it. • (Default) Selected: Adds acetylation to N-terminal methionines. • Cleared: Does not add acetylation to N-terminal methionines.
Apply N-Terminal Formylation	Determines whether the ProSightPC application adds formylation to N-terminal methionines. • Selected: Adds formylation to N-terminal methionines. • (Default) Cleared: Does not add formylation to N-terminal methionines.

Complexity Page Parameters

Table 49 lists the parameters on the Complexity page of the Create New Database Wizard, shown in Figure 96 on page 223.

Table 49. Complexity page parameters

Parameter	Description
Consider SNPs	Includes known potential genetic variations as annotated in the UniProt database.
Consider PTMs	Annotates known post-translational modifications (PTMs) onto a protein.
Maximum Features Per Sequence	Specifies the maximum number of features per input sequence.
Maximum Mass to Annotate Features (Da)	Specifies the cutoff point above which PTMs are not annotated in the database.
PTM Selection area	Specifies which PTMs should be considered for inclusion in the database.

Digestion Page Parameters

Table 50 lists the parameters on the Digestion page of the Create New Database dialog box, shown in Figure 97 on page 225.

Table 50. Digestion page 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.
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

Table 51 lists the parameters on the Database Description page of the Create New Database Wizard, shown in Figure 98 on page 226.

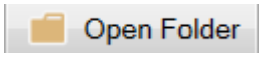
Table 51. Database Description page parameters

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.

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 or repository**

1. Choose **Databases > Database Manager** to open the Database Manager window.
2. Click  in the toolbar.
3. In the system folder that opens, right-click the proteome database (*database_name*.pscw file) or repository file (*database_name*.psph) to remove and choose **Delete**.
4. In the Delete File dialog box, click **Yes**.

> **IMPORTANT** Removing a proteome database or 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 or repository.

Database Manager Window Parameters

The Database Manager window contains the parameters and toolbar icons shown in Table 52.

Table 52. Database Manager window parameters (Sheet 1 of 2)

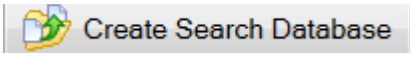
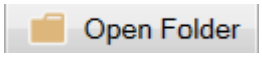
Parameter	Description
	Activates the Create New Database Wizard so that you can create a new proteome database.
	Opens a folder where you can export (copy) and import (paste) proteome database (PSCW) or repository (PSPH) files.
	Refreshes the proteome database data or repository data displayed in the top half of the Database Manager window.
Search Databases page	
Name	Displays the name of the proteome database. This name must be unique.
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.

Table 52. Database Manager window parameters (Sheet 2 of 2)

Parameter	Description
Basic Sequences	Lists the number of unmodified proteoforms in the proteome database.
Proteoforms	Lists the total number of shotgun-annotated proteoforms 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.
High Throughput Repositories page	
Name	Displays the name of the repository.
Description	Displays a brief description of the repository.
# of Projects	Displays the number of project categories in a repository.
# of Files	Displays the number of files included in a repository.
# of Experiments	Displays the number of experiments included in a repository.

Linking to the UniProt Database

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 database

Choose **Databases > Link to Uniprot**.

This command opens a Web browser with the appropriate UniProt address.

9

Using ProSightPC Tools

This chapter describes the utilities included in the ProSightPC application: the PTM Tier Editor, the Fragment Predictor, and the Font Converter.

Contents

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

The ProSightPC application 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. The ProSightPC application 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, the ProSightPC application comes with a preset list of included PTMs. It excludes any PTM not listed. You select which PTMs are available to be included in or excluded from analyses conducted by the ProSightPC application.

  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 the ProSightPC application analyzes MS data.

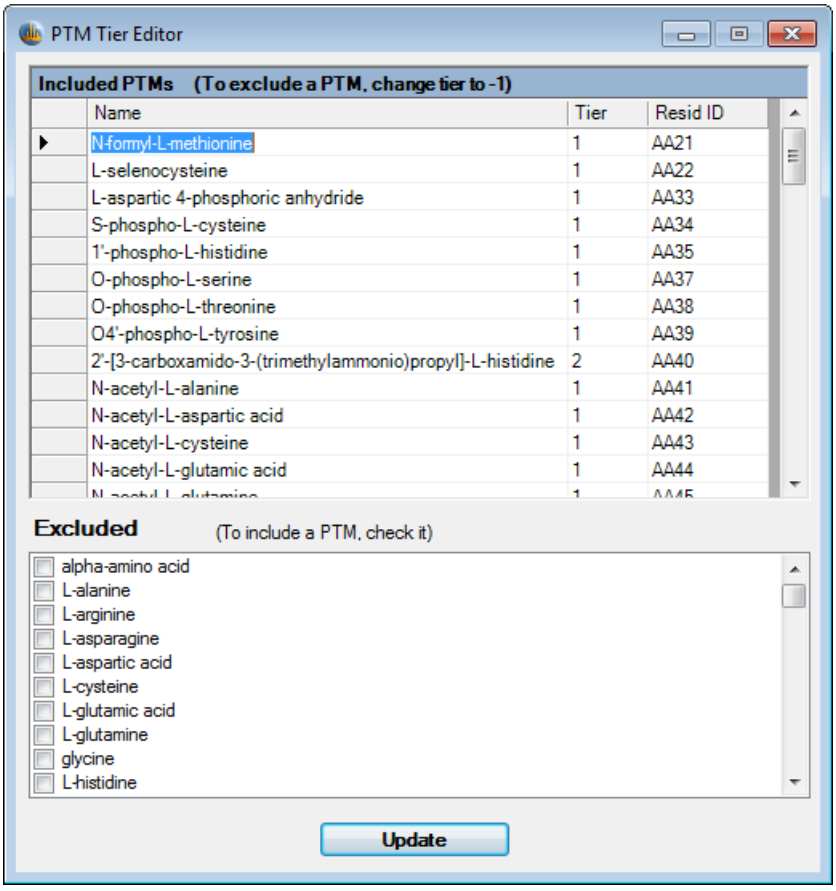
Accessing the PTM Tier Editor

❖ **To access the PTM Tier Editor**

Choose **Tools > PTM Tier Editor**.

The PTM Tier Editor dialog box opens, as shown in Figure 99.

Figure 99. 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

Table 53 lists the parameters in the PTM Tier Editor dialog box, shown in Figure 99 on page 234.

Table 53. PTM Tier Editor dialog box parameters

Parameter	Description
Included PTMs	Lists the included PTMs.
Name	Specifies the RESID name of the included PTM.
Tier	Specifies the current tier assignment of the included PTMs.
Resid ID	Specifies the RESID identifier of the included PTMs.
Excluded	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.

Excluding PTMs

Note Reincluded PTMs are automatically designated as Tier 1.

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 the ProSightPC application 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.

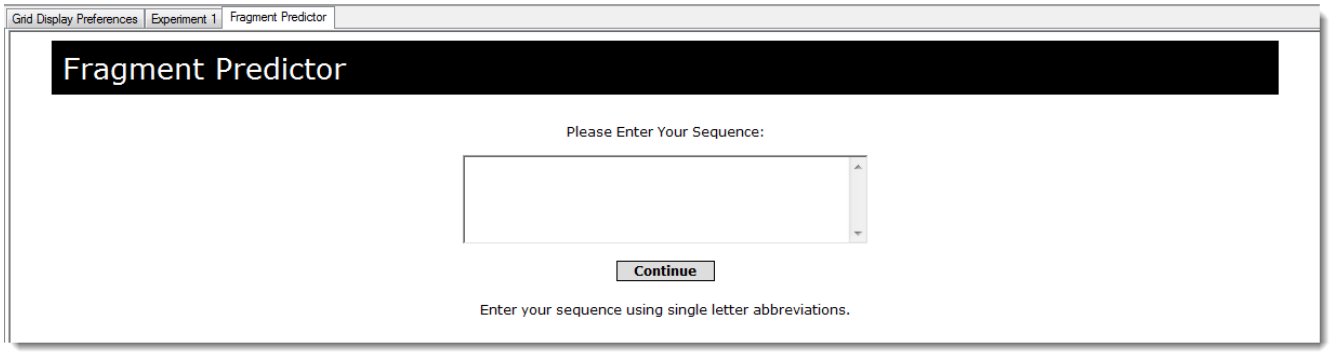
The Fragment Predictor does not directly handle cross-linked proteins, such as trans-peptide bonds or cyclized species like disulfides. Compute these by using the Enter Custom Mass function.

❖ To view the fragment ions for a sequence

1. Choose **Tools > Fragment Predictor**.

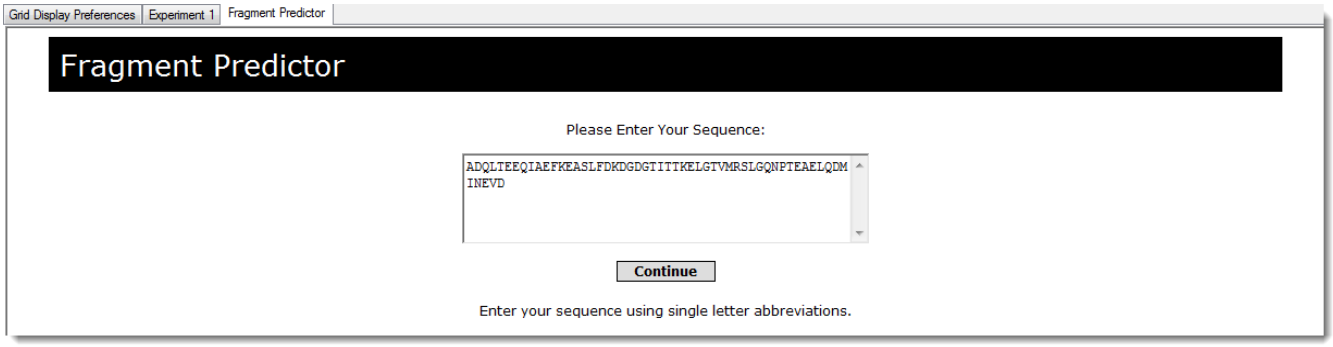
   The Fragment Predictor window opens, as shown in Figure 100.

Figure 100. Fragment Predictor window



2. Enter a protein sequence in the protein sequence box, as shown in Figure 101.

Figure 101. Entering a sequence in the Fragment Predictor window



You can use two methods of entering a protein sequence in the Protein Sequence box:

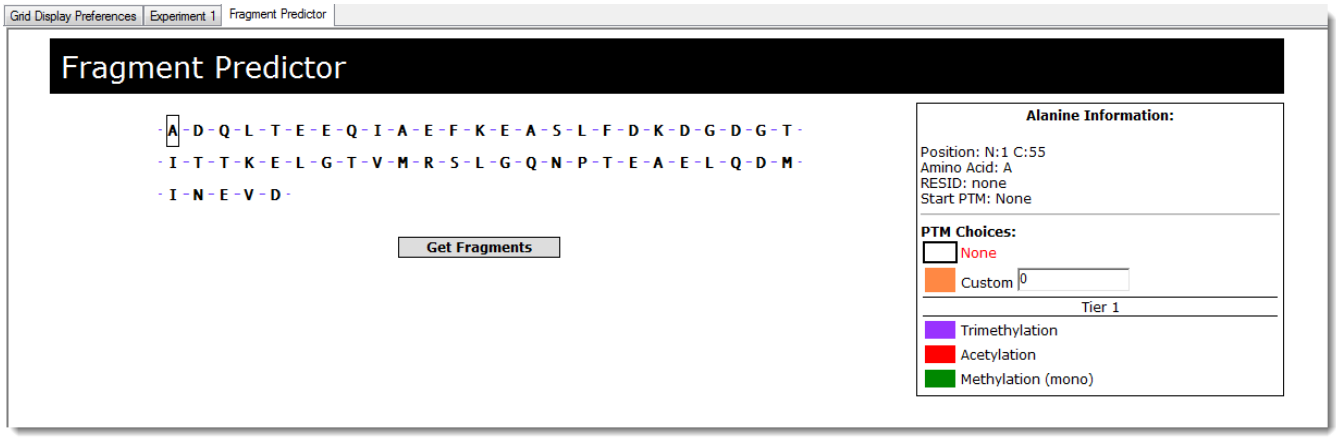
- Manually enter a protein sequence.
- Copy a protein sequence from another source.

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 102.

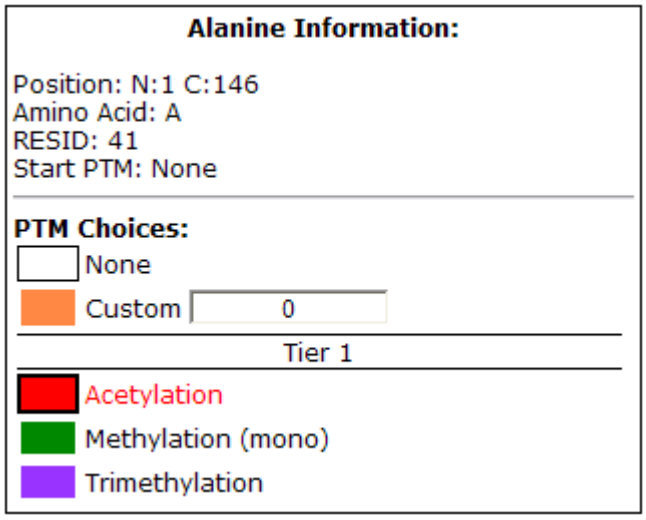
Figure 102. Interactive sequence map in the Fragment Predictor window



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 103.

   Figure 103. 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 233.

5. Click the name of the PTM.

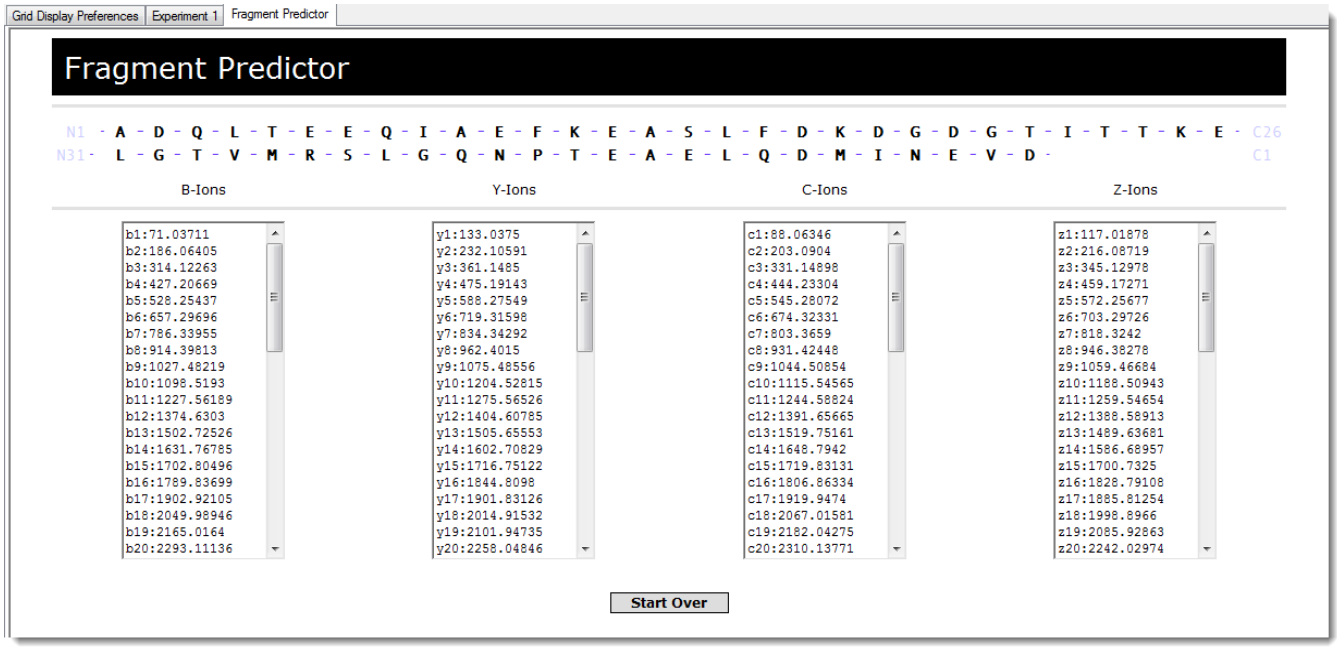
   The designated amino acid changes to match the color of the PTM selected.

   If applicable, 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 104. All theoretical fragment ion masses are arranged in ascending order and are classified as either b, y, c, or z•.

Figure 104. Results window



7. (Optional) Click **Start Over** to return to step 2.

Fragment Predictor Window Parameters

The Fragment Predictor window contains the parameters shown in Table 54.

Table 54. Fragment Predictor window parameters

Parameter	Description
Please Enter Your Sequence	Displays the protein sequence where 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

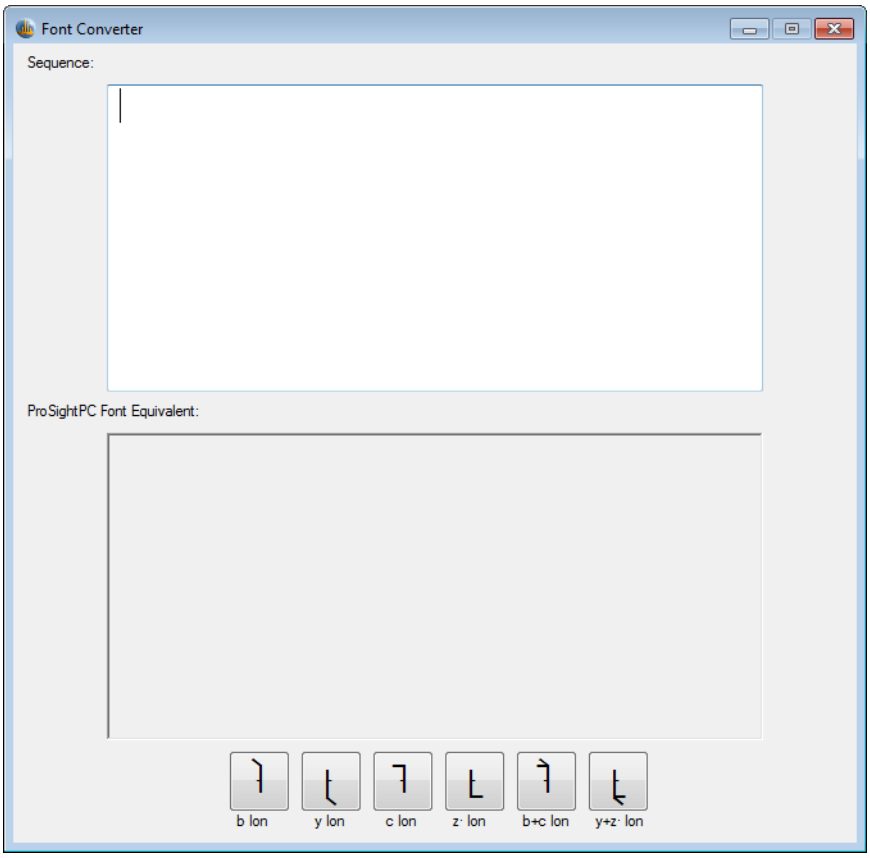
You can use the ProSightPC Font Converter to convert text into the ProSightPC fragment map font used to display N-terminal and C-terminal fragments. You can also use it to generate fragment maps to include in publications and presentations.

❖ To convert text to ProSightPC fragment map font

1. Choose **Tools > Font Converter**.

   The Font Converter dialog box opens, as shown in Figure 105.

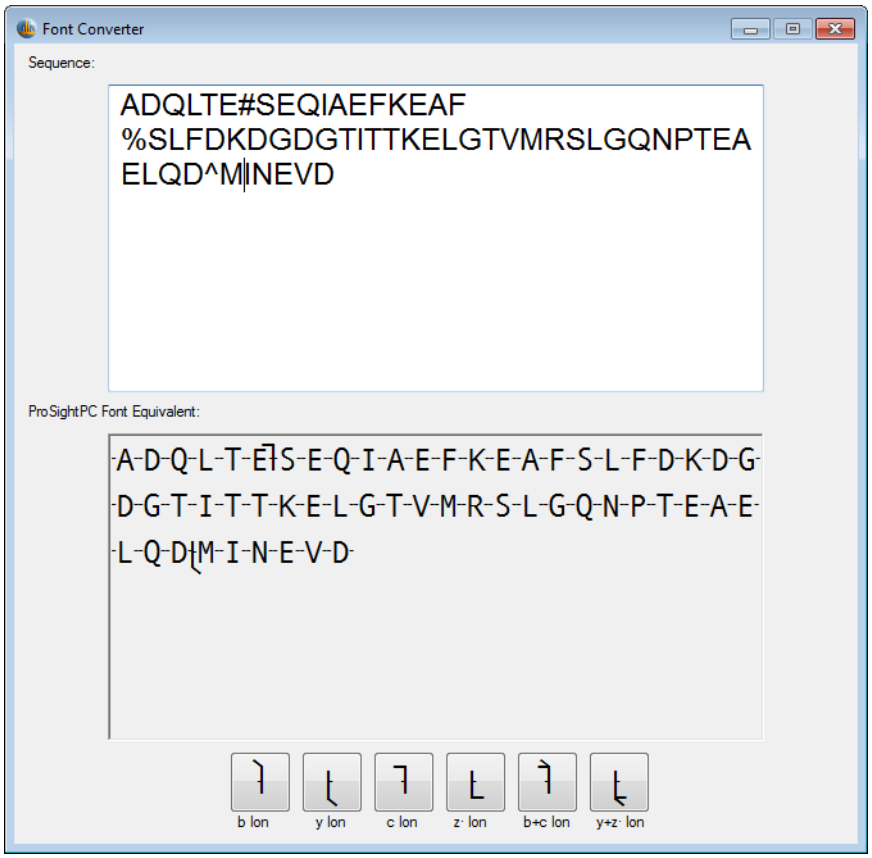
   Figure 105. Font Converter dialog box



2. In the Sequence box, enter the amino acid sequence to be converted, as shown in Figure 106.

   You can either type the amino acid sequence in the box or paste it from another source.

Figure 106. Font Converter dialog box with an 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, 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 the ProSightPC application.

Font Converter Dialog Box Parameters

The Font Converter dialog box contains the parameters described in Table 55.

Table 55. Font Converter dialog box parameters

Parameter	Description
Sequence	Specifies the amino acid sequence to be converted.
ProSightPC Font Equivalent	Displays the ProSightPC application font equivalent of the sequence displayed in the Sequence box.
	Adds the sign for a b ion.
	Adds the sign for a y ion.
	Adds the sign for a c ion.
	Adds the sign for a z˙ ion.
	Adds the sign for a bc ion.
	Adds the sign for a yz˙ ion.

A

ProSightPC Reference

This appendix describes the commands on the ProSightPC menus. They are listed in the order in which they appear in the menus.

Contents

File Menu

Table 56 lists the commands in the File menu.

Table 56. File menu commands (Sheet 1 of 2)

Command	Description
File > New	Clears the data grid so that you can create 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.

Table 56. File menu commands (Sheet 2 of 2)

Command	Description
File > Import Data from Repository	Imports experiments from a repository into the ProSightPC application 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 33 on page 72.
File > Export Data to Repository	Exports the experiments in the ProSightPC application data grid into a repository. It opens the Export Data to Repository dialog box, shown in Figure 34 on page 74. 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 data file using one of the following analyses to infer mass (AIMs): • (Default) 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 35 on page 77. • 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 38 on page 84.
Four most recently opened PUF files	Click the file name to open the file.
File > Exit	Closes the ProSightPC application.

Edit Menu

Table 57 lists the commands in the Edit menu.

Table 57. Edit menu commands

Command	Description
Edit > Copy	Copies text.
File > Paste	Pastes text.

View Menu

Table 58 lists the commands in the View menu.

Table 58. View menu commands (Sheet 1 of 2)

Command	Description
View > Grid Display Preferences	Determines whether the application displays the Grid Display Preferences page in the ProSightPC window. "Using Filters in the Show Columns Area" on page 265 describes the function of each of the options on the Grid Display Preferences page.
View > Start	Determines whether the application displays the Start page in the ProSightPC window.
View > Job Queue	Determines whether the job queue pane is visible. See Figure 107 on page 256 for the location of the job queue and "Job Queue" on page 263 for a description of the job queue.
View > Toolbar	Determines whether the application displays the toolbar in the ProSightPC window. See "Toolbar" on page 257 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.

Table 58. View menu commands (Sheet 2 of 2)

Command	Description
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.
View > Open Data Manager	Opens the Data Manager i 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 207 for detailed information on the Data Manager.

Experiment Tools Menu

Table 59 lists the commands in the Experiment Tools menu, which is only available when an experiment is open in the Data Manager.

Table 59. Experiment Tools menu commands

Command	Description
Experiment Tools > Append Predefined Search	Opens the Append Predefined Searches to Experiment *X* dialog box, shown in Figure 45 on page 108, so you can add a new predefined search to an experiment. For information on how to select options in this dialog box, see "Adding Predefined Searches to an Experiment" on page 108.
Experiment Tools > Edit Masses	Opens a new Edit Masses page in the ProSightPC window shown in Figure 88 on page 211, so you can review and edit mass values in an experiment.
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

Table 60 lists the commands in the Databases menu.

Table 60. Databases menu commands

Command	Description
Databases > Database Manager	Opens the Database Manager window, shown in Figure 90 on page 217. 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 91 on page 218, so you can manually create a custom database.
Databases > Download ProSightPC Databases	Downloads databases from the Proteinaceous website.
Databases > Link to UniProt	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

Table 61 lists the commands in the ProSightHT menu.

Table 61. ProSightHT menu commands (Sheet 1 of 2)

Command	Description
ProSightHT > High Throughput Wizard	Opens the Process a Dataset page of the High Throughput Wizard, shown in Figure 14 on page 28, so you can import data from a raw data 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 32 on page 70, so you can edit an existing repository or add a new one.

Table 61. ProSightHT menu commands (Sheet 2 of 2)

Command	Description
ProSightHT > Edit/Add Search Tree	Opens the Running High Throughput Logic page of the High Throughput Wizard, shown in Figure 16 on page 36. For information on the options on this page, see "Selecting or Creating a Repository" on page 35.
ProSightHT > Repository Report	Generates a repository report that lists all the experiments that a repository contains. Figure 69 on page 177 shows an example of this report. This command opens the Repository Report dialog box shown in Figure 68 on page 175.

Tools Menu

Table 62 lists the commands in the Tools menu.

Table 62. Tools menu commands (Sheet 1 of 3)

Command	Description
Tools > Experiment Adder	Imports experiment data into the ProSightPC application and displays it in the data grid. It opens the Experiment Adder dialog box, shown in Figure 30 on page 56.
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 43 on page 105.
Tools > PTM Tier Editor	Opens the PTM Tier Editor dialog box, as shown in Figure 99 on page 234, so you can manually reassign a PTM to another tier.
Tools > Fixed Modification Editor	Opens the Fixed Modification Editor so that you can add fixed modifications, which apply the same specific mass to all occurrences of the named amino acid.

Table 62. Tools menu commands (Sheet 2 of 3)

Command	Description
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 105 on page 240.
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˙ fragment ion masses. It opens the Fragment Predictor window, shown in Figure 100 on page 237.
Tools > Experiment Manager	Opens the Experiment Manager, shown in Figure 41 on page 96, 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.

Table 62. Tools menu commands (Sheet 3 of 3)

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 the ProSightPC application. See "Setting Default Options" on page 21 for instructions on setting default values.

Help Menu

Table 63 lists the commands in the Help menu.

Table 63. Help menu commands (Sheet 1 of 2)

Command	Description
Help > Help	Opens the Help for the ProSightPC application.
Help > Manage License	Opens the License Information dialog box so that you can request a a new license activation code.
Help > About	Opens a dialog box that displays the release version, the release date, and the trademark information.

Table 63. Help menu commands (Sheet 2 of 2)

Command	Description
Help > Open Log	Opens a log.txt file in Notepad for troubleshooting purposes. If you encounter an error, it appears in the log.txt file so that you can easily view it or share it with the Proteinaceous team.
Install Font	Opens the Windows Font Viewer so that you can install the ProSightPC font. This procedure only needs to be done once.
Manuals	Opens the PDF file of the selected ProSightPC user manual.

Data Grid Shortcut Menu

Table 64 lists the commands in the data grid shortcut menu, which appears when you right-click an experiment in the data grid.

Table 64. Data grid shortcut menu commands (Sheet 1 of 2)

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.
Append Predefined Searches	Opens the Append Predefined Searches to Experiment *X* dialog box, shown in Figure 45 on page 108, so you can add more than one predefined search to an experiment. For information on how to select options in this dialog box, see "Adding Predefined Searches to an Experiment" on page 108.
Append Predefined Search	Opens a submenu with all of the predefined searches. Clicking one of them adds it to the selected experiment.

Table 64. Data grid shortcut menu commands (Sheet 2 of 2)

Command	Description
Edit Search *x*	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). For information on how to edit a search, see "Editing a Predefined Search" on page 109. This command is only available when the Pending Search column displays "yes" for the appropriate search.
Edit Mass List	Opens a new Edit Masses page in the ProSightPC window, showing the Precursor Mass List and the Fragment Mass List. For information on displaying these two lists, see "Editing Mass Values" on page 210.
Remove Results	Removes search results from a search that has already been run. This command is useful if you want to rerun a search with different parameters. This command is only available when search results are present—that is, when the Pending Search column displays "no" for the appropriate search.
Run Search *x*	Runs a pending predefined search. This command is only available when the Pending Search column displays "yes" for the appropriate search. For information on how to run a predefined search, see "Running a Predefined Search" on page 111.
Remove Search *x*	Removes the specified search from the experiment. For more information on removing searches, see "Removing a Predefined Search" on page 112.
Remove Experiment *x*	Removes the specified experiment from the data grid. For more information on removing experiments, see "Removing an Experiment from the Data Grid" on page 113.

Table 65 describes the command in the secondary data grid shortcut menu, which appears when you right-click the area to the right of the columns in the data grid, as shown in Figure 110 on page 260.

Table 65. Data grid shortcut menu command

Command	Description
Columns	Determines which columns appear in the data grid.

B

Using the ProSightPC Application Window

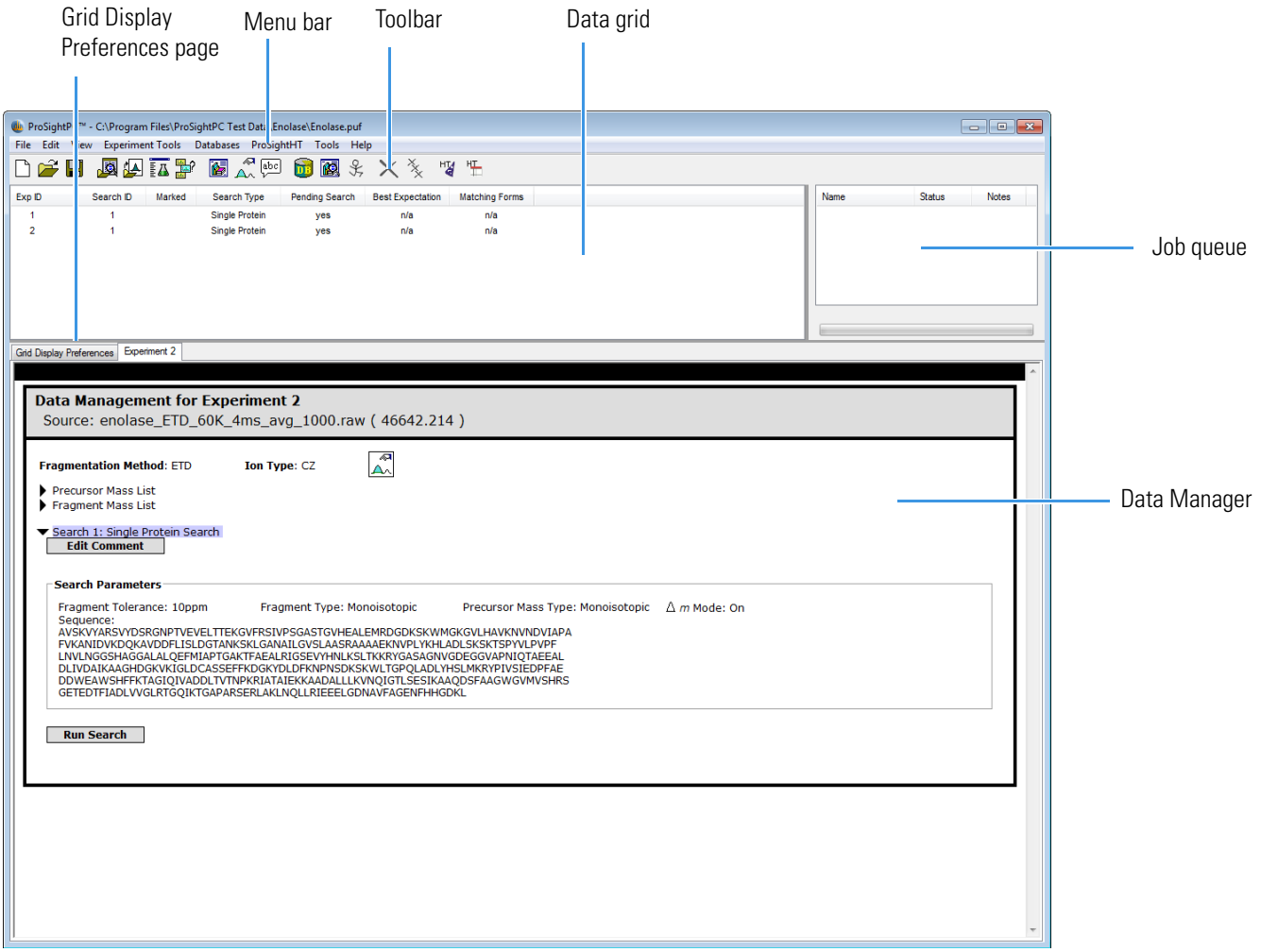
The ProSightPC application has a unique approach to organizing elements in the application window. This appendix describes these elements.

Contents

ProSightPC Application Window

Figure 107 shows some of the features of the ProSightPC application window.

Figure 107. ProSightPC window



The following topics explain the features of the ProSightPC main window in detail.

- Menu Bar
- Toolbar
- Data Grid
- Job Queue
- Pages in the ProSightPC Application Window
- Data Manager
- Grid Display Preferences Page

Menu Bar

The ProSightPC menu bar, shown in Figure 107, appears at the top of the ProSightPC application window. It contains the menus shown in Table 66. For detailed descriptions of the commands in these menus, see "ProSightPC Reference" on page 243.

Table 66. Menus in the ProSightPC menu bar

Menu	Description
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 application window features, such as grid preferences, start screen, job queue, and toolbar.
Experiment Tools	Use these commands to perform operations on experiments. 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, install the ProSightPC font, and access the Help.

Toolbar

The ProSightPC toolbar, pictured in Figure 108, appears directly below the menu bar of the ProSightPC apaplication window.

Figure 108. ProSightPC toolbar

❖ **To display the toolbar**

Choose **View > Toolbar**.

Table 67 describes each of the icons in the toolbar.

Table 67. ProSightPC toolbar (Sheet 1 of 2)

Icon	Menu equivalent	Function
	File > New	Clears the data grid so that you can create 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.
	View > Open Data Manager	Opens the last experiment using the open PUF file if there are no Experiment tabs open in the ProSightPC window.
	Experiment Tools > Append Predefined Search	Adds a predefined search to the selected experiment. This icon is available only when an experiment is open and showing in the ProSightPC window.
	Experiment Tools > Edit Masses	Changes MS and MS/MS data in the ProSightPC application. 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 ProSightPC window.
	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.
	Tools > Batch Run	Queues and runs a number of searches over any number of experiments.

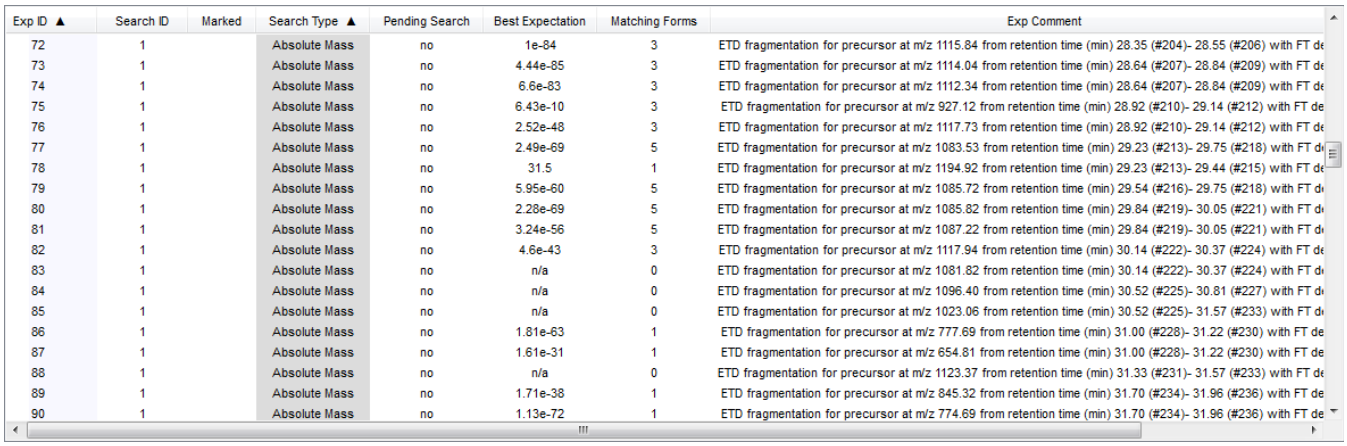
Table 67. ProSightPC toolbar (Sheet 2 of 2)

Icon	Menu equivalent	Function
	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 > High Throughput 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.

Data Grid

The data grid, shown in Figure 109, displays summary information about each search in the open PUF file, organized into columns. You can use the data grid to perform and modify searches.

Figure 109. Data grid



Exp ID ▲	Search ID	Marked	Search Type ▲	Pending Search	Best Expectation	Matching Forms	Exp Comment
72	1		Absolute Mass	no	1e-84	3	ETD fragmentation for precursor at m/z 1115.84 from retention time (min) 28.35 (#204)- 28.55 (#206) with FT de
73	1		Absolute Mass	no	4.44e-85	3	ETD fragmentation for precursor at m/z 1114.04 from retention time (min) 28.64 (#207)- 28.84 (#209) with FT de
74	1		Absolute Mass	no	6.6e-83	3	ETD fragmentation for precursor at m/z 1112.34 from retention time (min) 28.64 (#207)- 28.84 (#209) with FT de
75	1		Absolute Mass	no	6.43e-10	3	ETD fragmentation for precursor at m/z 927.12 from retention time (min) 28.92 (#210)- 29.14 (#212) with FT de
76	1		Absolute Mass	no	2.52e-48	3	ETD fragmentation for precursor at m/z 1117.73 from retention time (min) 28.92 (#210)- 29.14 (#212) with FT de
77	1		Absolute Mass	no	2.49e-69	5	ETD fragmentation for precursor at m/z 1083.53 from retention time (min) 29.23 (#213)- 29.75 (#218) with FT de
78	1		Absolute Mass	no	31.5	1	ETD fragmentation for precursor at m/z 1194.92 from retention time (min) 29.23 (#213)- 29.44 (#215) with FT de
79	1		Absolute Mass	no	5.95e-60	5	ETD fragmentation for precursor at m/z 1085.72 from retention time (min) 29.54 (#216)- 29.75 (#218) with FT de
80	1		Absolute Mass	no	2.28e-69	5	ETD fragmentation for precursor at m/z 1085.82 from retention time (min) 29.84 (#219)- 30.05 (#221) with FT de
81	1		Absolute Mass	no	3.24e-56	5	ETD fragmentation for precursor at m/z 1087.22 from retention time (min) 29.84 (#219)- 30.05 (#221) with FT de
82	1		Absolute Mass	no	4.6e-43	3	ETD fragmentation for precursor at m/z 1117.94 from retention time (min) 30.14 (#222)- 30.37 (#224) with FT de
83	1		Absolute Mass	no	n/a	0	ETD fragmentation for precursor at m/z 1081.82 from retention time (min) 30.14 (#222)- 30.37 (#224) with FT de
84	1		Absolute Mass	no	n/a	0	ETD fragmentation for precursor at m/z 1096.40 from retention time (min) 30.52 (#225)- 30.81 (#227) with FT de
85	1		Absolute Mass	no	n/a	0	ETD fragmentation for precursor at m/z 1023.06 from retention time (min) 30.52 (#225)- 31.57 (#233) with FT de
86	1		Absolute Mass	no	1.81e-63	1	ETD fragmentation for precursor at m/z 777.69 from retention time (min) 31.00 (#228)- 31.22 (#230) with FT de
87	1		Absolute Mass	no	1.61e-31	1	ETD fragmentation for precursor at m/z 654.81 from retention time (min) 31.00 (#228)- 31.22 (#230) with FT de
88	1		Absolute Mass	no	n/a	0	ETD fragmentation for precursor at m/z 1123.37 from retention time (min) 31.33 (#231)- 31.57 (#233) with FT de
89	1		Absolute Mass	no	1.71e-38	1	ETD fragmentation for precursor at m/z 845.32 from retention time (min) 31.70 (#234)- 31.96 (#236) with FT de
90	1		Absolute Mass	no	1.13e-72	1	ETD fragmentation for precursor at m/z 774.69 from retention time (min) 31.70 (#234)- 31.96 (#236) with FT de

❖ To change the columns displayed in the data grid

Do one of the following:

- In the data grid, right-click the area to the right of the columns and choose **Columns** > ***column_name***, as shown in Figure 110.

–or–

- Follow this procedure:

  i. Click the **Grid Display Preferences** tab.

The Show Columns area of the Grid Display Preferences page displays check boxes whose names correspond to the names of the data grid columns.

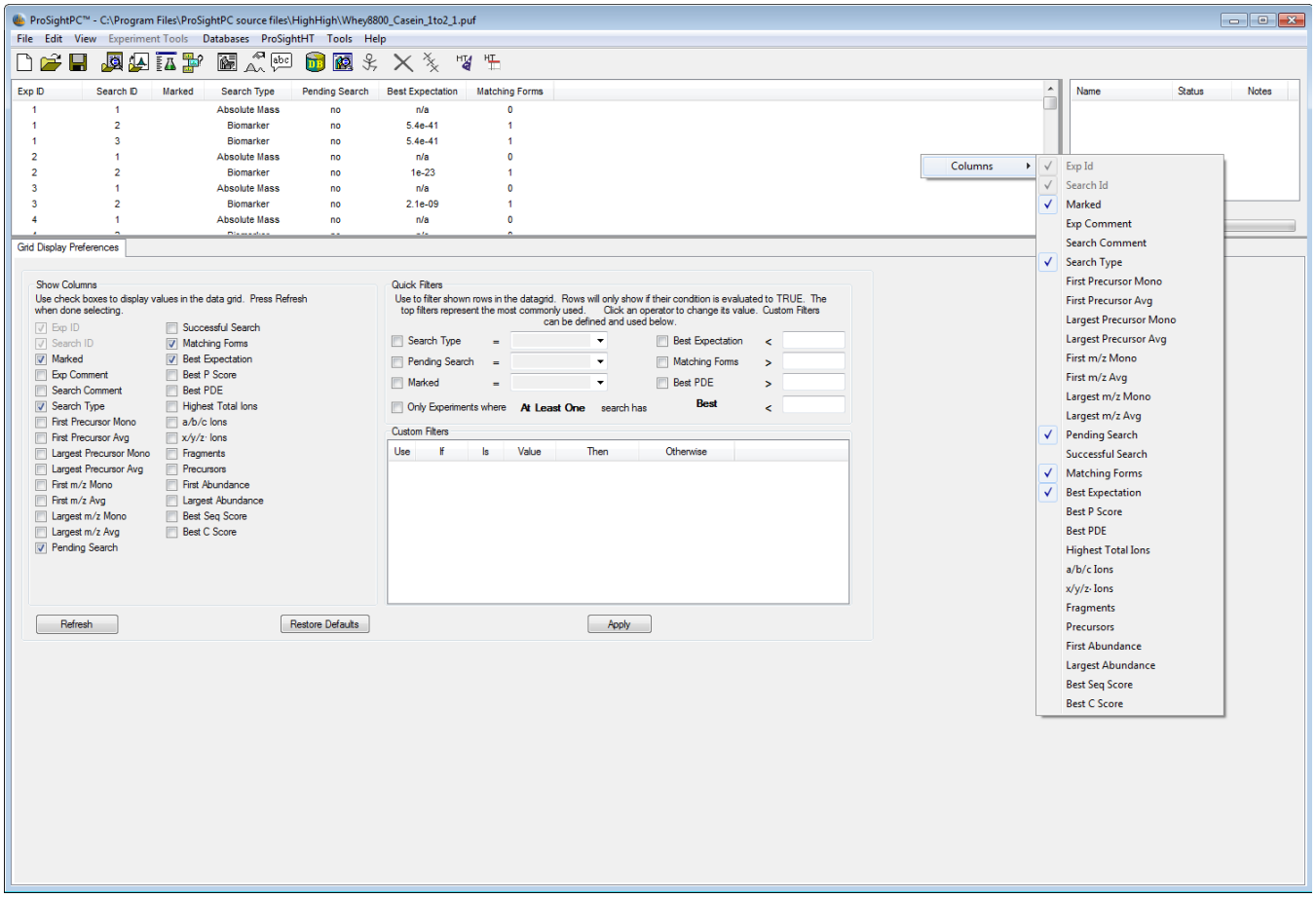
ii. Select the appropriate check boxes.

iii. Click **Refresh**.

To redisplay the default columns, click **Restore Defaults**.

For information on the columns available in the data grid and the filters available to refine the data displayed, see "Grid Display Preferences Page" on page 265.

Figure 110. Choosing data grid columns from the data grid shortcut menu



❖ To sort column data in the data grid

Click the appropriate column title to sort the data from lowest value to highest value, or highest to low.

❖ To select rows in the data grid

- To select contiguous rows, click the name of the first experiment, hold down the SHIFT key, and click the last row that you want to select.

- To select noncontiguous rows, click the name of the first experiment, hold down the CTRL key, and click each separate row.

❖ **To import data into the data grid**

See "Working with Experiments" on page 95.

❖ **To perform and modify a search**

See "Searching Databases" on page 103.

❖ **To open the relevant Data Manager from the data grid**

Double-click a search in the data grid.

For more information about the Data Manager, see "Displaying Data in the Data Manager" on page 207.

Data Grid Shortcut Menu Commands

Table 68 describes the commands available in the main data grid shortcut menu.

Table 68. Data grid main shortcut menu (Sheet 1 of 2)

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.
Append Predefined Searches	Opens the Append Predefined Searches to Experiment *X* dialog box, shown in Figure 45 on page 108, so that you can add more than one predefined search to the experiment. For information on how to select options in this dialog box, see "Adding Predefined Searches to an Experiment" on page 108.
Append Predefined Search	Opens a submenu with all of the predefined searches. Clicking one of them adds it to the selected experiment.

Table 68. Data grid main shortcut menu (Sheet 2 of 2)

Command	Description
Edit Search *x*	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). For information on how to edit a search, see "Editing a Predefined Search" on page 109. This command is only available when the Pending Search column displays "yes" for the appropriate search.
Edit Mass List	Opens a new Edit Masses page in the ProSightPC window, showing the Precursor Mass List and the Fragment Mass List. For information on displaying these two lists, see "Editing Mass Values" on page 210.
Remove Results	Removes search results from a search that has already been run. This command is useful if you want to rerun a search with different parameters. This command is only available when the Pending Search column displays "no" for the appropriate search.
Run Search *x*	Runs a pending predefined search. This command is only available when the Pending Search column displays "yes" for the appropriate search. For information on how to run a predefined search, see "Running a Predefined Search" on page 111.
Remove Search *x*	Removes the specified predefined search from the experiment. For more information on removing searches, see "Removing a Predefined Search" on page 112.
Remove Experiment *x*	Removes the specified experiment from the data grid. For more information on removing experiments, see "Removing an Experiment from the Data Grid" on page 113.

Table 69 describes the command on the secondary data grid shortcut menu, which appears when you right-click the area to the right of the columns in the data grid, as shown in Figure 110 on page 260.

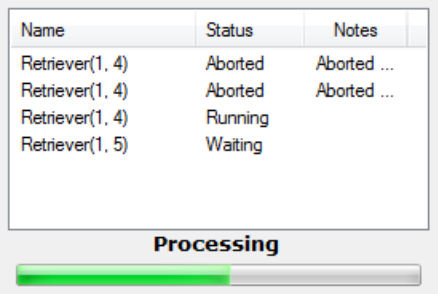
Table 69. Data grid secondary shortcut menu

Command	Description
Columns	Determines the columns that appear in the data grid.

Job Queue

The job queue, shown in Figure 111, 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.

Figure 111. Job queue



❖ To display the job queue pane

Choose **View > Job Queue**.

❖ To rerun a search in the job queue

If a job in the job queue has finished but you want to rerun it, right-click the search in the job queue and choose **Run** from the shortcut menu.

❖ To cancel a search in the job queue

- If a search is running, right-click the search in the job queue, and choose **Abort** or click the **Abort Running Job** icon, , in the toolbar.

  The search ends and the status changes to Failed.

- If you want to cancel all running searches, click the **Abort All Jobs** icon, .

Job Queue Parameters

The job queue contains the areas shown in Table 70.

Table 70. Job queue areas

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

Right-click the job queue pane to display the commands described in Table 71.

Table 71. Job queue shortcut menu

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

For information on performing searches, see Searching Databases.

Pages in the ProSightPC Application Window

Many of the more complex elements of the ProSightPC application appear in the pages of the ProSightPC window. Double-click an experiment in the data grid to display the experiment in the Data Manager in the ProSightPC window.

- 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 207.

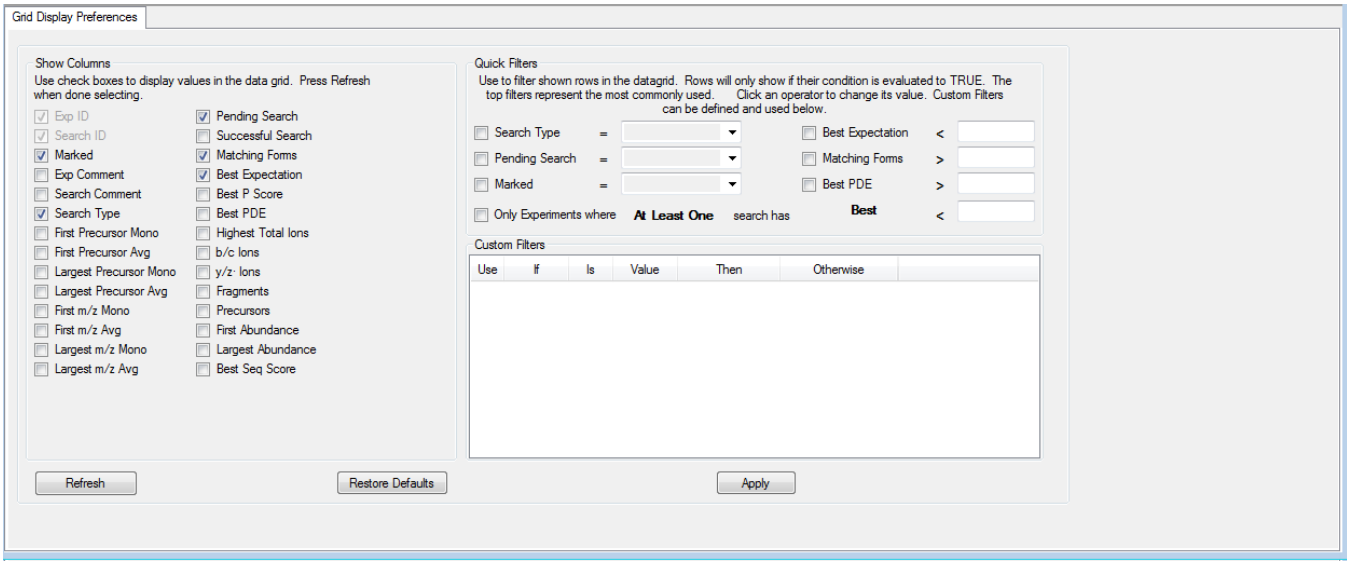
Data Manager

The Data Manager appears in the ProSightPC application window 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 207 describes the functions of the Data Manager in detail.

Grid Display Preferences Page

Use the Grid Display Preferences page, shown in Figure 112, 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 112. Grid Display Preferences page



- Using Filters in the Show Columns Area
- Using the Filters in the Quick Filters Area
- Using the Filters in the Custom Filters Section

Using Filters in the Show Columns Area

Use the Show Columns area to display or hide columns in the data grid. Each of the parameters shown in Table 72 controls the appearance of a column in the data grid.

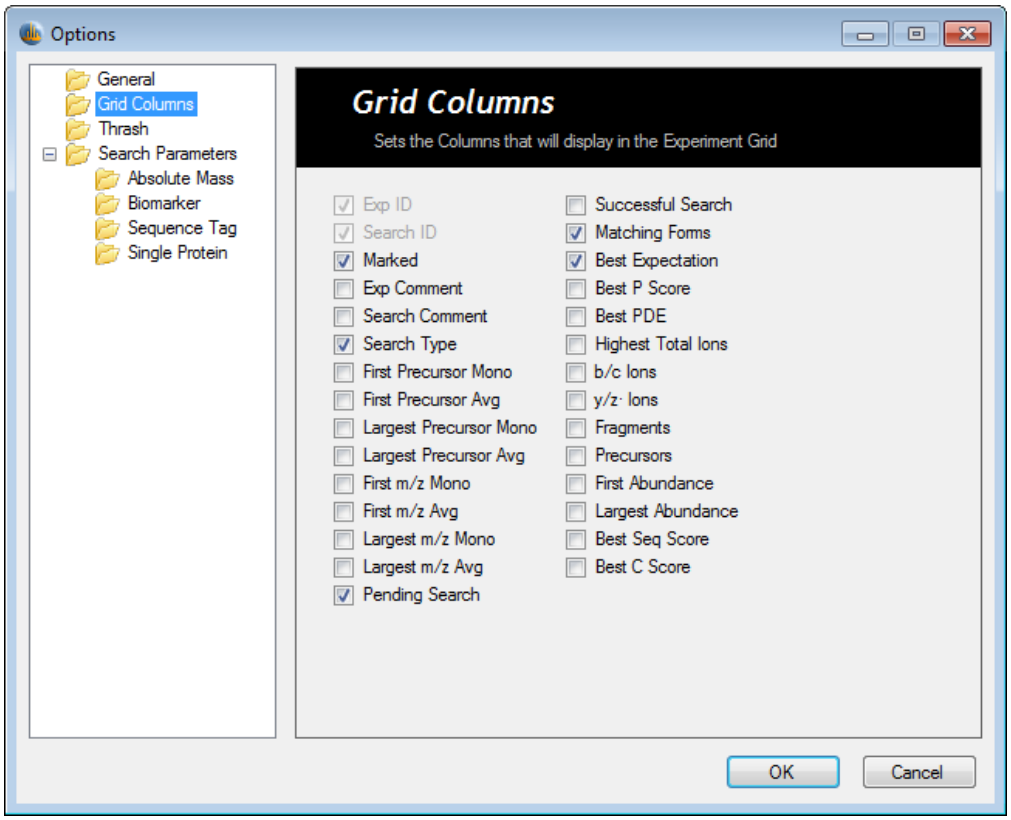
❖ **To access the Grid Display Preferences page**

Click the **Grid Display Preferences** tab.

❖ To set the default columns displayed in the data grid

1. Choose **Tools > Options** to open the Options dialog box.
2. From the left pane of the Options dialog box, click the **Grid Columns** folder to open the Grid Columns page, shown in Figure 113.

   Figure 113. Grid Columns page of the Options dialog box



3. Select the check boxes next to the names of the columns to be displayed by default in the experiment grid.

   For information on these columns, see "Show Columns Area Parameters" on page 267.

4. Click **OK**.

❖ To temporarily change the columns displayed in the data grid

1. Click the **Grid Display Preferences** tab.
2. In the Show Columns area, select the check boxes next to the names of any columns that you want to display in the data grid.
3. Click **Refresh** to display the columns that you selected in the data grid.

   The columns that you selected appear in the data grid.

❖ To remove a column from the data grid

1. Click the **Grid Display Preferences** tab.
2. Clear the check box for any of the columns that you want to hide in the data grid.
3. Click **Refresh** to hide the columns.

❖ To restore default columns

To reinstate the default settings, click **Restore Defaults**.

Show Columns Area Parameters

Table 72 describes the parameters available in the Show Columns area of the Grid Display Preferences page and on the Grid Columns page of the Options dialog box.

Table 72. Show Columns area parameters (Sheet 1 of 3)

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	(Default) 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	(Default) 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.

Table 72. Show Columns area parameters (Sheet 2 of 3)

Parameter	Description
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.
Successful Search	Displays a column indicating whether a match in the database was found.
Matching Forms	(Default) Displays a column showing the number of matching proteoforms.
Best Expectation	(Default) 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 196.
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 198.
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.
Fragments	Displays a column with the total number of theoretical fragments present.
Precursors	Displays a column with the total number of theoretical precursors present.
First Abundance	Displays a column with the abundance of the first precursor.

Table 72. Show Columns area parameters (Sheet 3 of 3)

Parameter	Description
Largest Abundance	Displays a column with the abundance of the precursor with the largest abundance.
Best Seq Score	Displays a column with the best sequence tag score.
Best C Score	Displays the C score, which is a measure of the confidence in the characterization of the proteoform.

Using the Filters in the 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.

❖ To define quick filters for a search

1. Access the **Grid Display Preferences** page.
2. (Optional) In the Quick Filters area of the Grid Display Preferences page, select the **Search Type**, **Pending Search**, or **Marked** check box.

   For information on these parameters, see Table 73.
3. (Optional) Select the **Best Expectation**, **Total Fragments**, or **Best PDE** check box, click the corresponding operator to set it, and enter the appropriate value in the box to the right of the option.
4. (Optional) Select the **Only Experiments Where *number* search has *option operator value*** check box. Click *number*, *option*, and *operator* to display the choices available. For information on these parameters, see Table 73. Type the value in the box to the right of the operator.
5. When you have set all the filters that you want, click **Apply**.

❖ To remove quick filters

1. Access the **Grid Display Preferences** page.
2. Clear the check box next to the name of the filter that you want to remove.

Quick Filters

Table 73 describes the quick filters available in the Quick Filters area of the Grid Display Preference page.

Table 73. Quick filters area parameters (Sheet 1 of 2)

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 196.
Matching Forms	Filters the search by the number of matching proteoforms.
Best PDE	Filters the search by PDE (McLuckey) score. For information about this scoring method, see "PDE (McLuckey) Score" on page 198.

Table 73. Quick filters area parameters (Sheet 2 of 2)

Filter	Description
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 value (e value). For information about the expectation value, see "Expectation Value (e value)" on page 196. • Best PDE: McLuckey score. For more information on the calculation of this score, see "PDE (McLuckey) Score" on page 198. • Best P Score: P score. For more information on the calculation of this score, see "P Score" on page 196. • Total Ions: Highest total number of ions that matched the ions in the database. • Matching Forms: Number of matching proteoforms. *Operator* can be • = Equal to • NOT = Not equal to • < Less than • > Greater than • <= Less than or equal to • >= Greater than or equal to
Refresh	Displays the columns selected in the Show Columns area in the data grid.
Restore Defaults	Reinstates the default settings in the Show Columns area.
Apply	Executes the filters that you set in the Quick Filters and Custom Filters areas.

Using the Filters in the Custom Filters Section

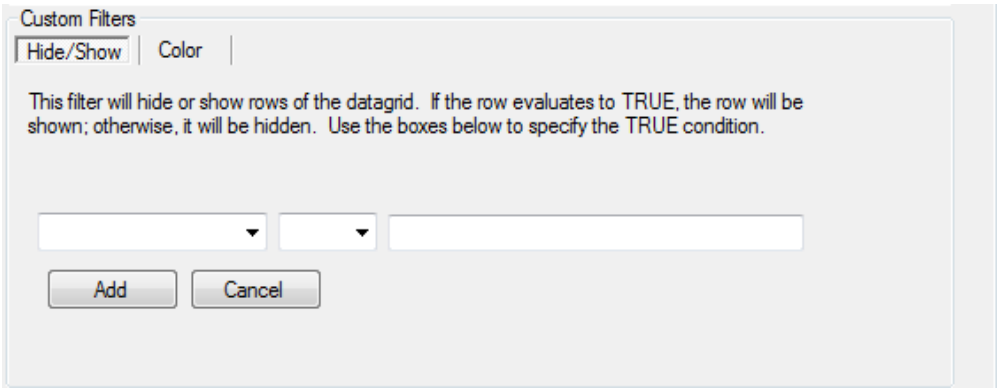
You can use the Custom Filters section of the Grid Display Preferences page to define completely custom conditions with which to filter the searches displayed in the data grid. Select one or more of the criteria to filter (hide) certain data grid rows. Click an operator to change its value.

❖ To define custom filters for a search

1. Access the **Grid Display Preferences** page.
2. Right-click the Custom Filters section and choose **New** from the shortcut menu.

   The appearance of the Custom Filters section changes to the default configuration shown in Figure 114.

   Figure 114. Custom Filters section of the Grid Display Preferences page

3. Specify the way that the data is displayed in the data grid. Do one of the following:

   - If you want a search to be displayed in the data grid if it meets the specified condition and not to be displayed if it does not meet the specified condition, click **Show/Hide** if it is not already selected.

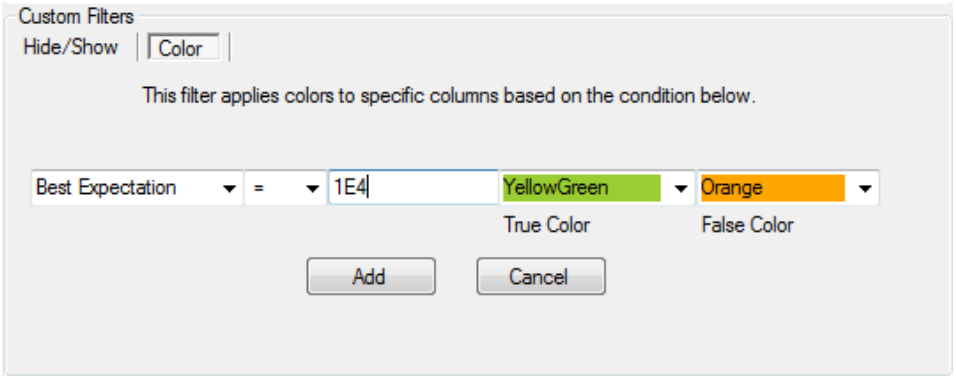
   –or–

   - If you want to apply colors to specific columns on the basis of the condition that you just set, follow this procedure:

     i. Click **Color**.

     Two new boxes with drop-down lists appear, as shown in Figure 115.

     Figure 115. Color selection lists

     ii. In the True Color list, select a color for columns containing data that meets your condition.

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

4. In 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 "Using Filters in the Show Columns Area" on page 265.

5. In the middle list, select an operator.

   The operators available in this list are the same as those given for the check box labeled Only Experiments Where *number* Search Has *option operator value* in "Using the Filters in the Quick Filters Area" on page 269. For information on these operators, see Table 73.

6. In the rightmost list, type an appropriate value.

   Here are some examples:

   largest precursor mono > 1000

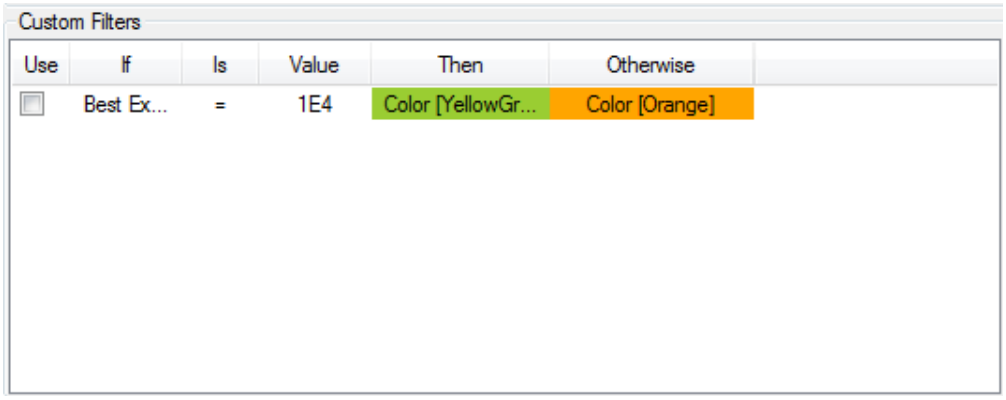
   b/c ions > 7

   y/z· ions < 20

7. Click **Add**.

   The filter appears with a small check box to the left, as shown in Figure 116.

   Figure 116. Filter added to list of custom filters



8. 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.

Custom Filters

Table 74 describes the filters available in the Custom Filters section of the Grid Display Preferences page.

Table 74. Custom Filters section parameters

Filter	Description
Hide/Show	Determines whether a search is displayed in the data grid if it meets the specified condition and is hidden if it does not meet the specified condition.
Color	Applies colors to specific columns on the basis of the conditions that you set.
Left list	Specifies the filter.
Middle list	Specifies the operator.
Right list	Specifies a value.
Add	Adds a newly defined filter.
Cancel	Resets the Custom Filters section to the configuration shown in Figure 112.
Use	Selects the filter to apply to a search.
If	Specifies the filter.
Is	Specifies the operator.
Value	Specifies a value.
Then	Specifies what happens when a search meets the specified condition.
Otherwise	Specifies what happens when a search does not meet the specified condition.
Refresh	Displays the columns selected in the Show Columns area in the data grid.
Restore Defaults	Reinstates the default settings in the Show Columns area.
Apply	Executes the filters that you set in the Quick Filters and Custom Filters sections.

Table 75 describes the commands on the menu that appears when you right-click in the Customer Filters section.

Table 75. Custom Filters shortcut menu

Filter	Description
New	Changes the configuration of the Custom Filters section to that shown in Figure 114 on page 272 so that you can set a new custom filter.
Remove	Removes the selected custom filter.
Check All	Selects all the defined custom filters.
Uncheck All	Clears all the defined custom filters.

Setting Default Options

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

In the Options dialog box, you can set the preferences shown in Table 76.

Table 76. Options dialog box parameters

Parameter	Location of information
General	Setting Default Options
Grid Columns	Grid Display Preferences Page
THRASH	Setting THRASH Preferences
Search Parameters	
Absolute Mass	Setting Absolute Mass Search Preferences
Biomarker	Setting Biomarker Search Preferences
Sequence Tag	Setting Sequence Tag Search Preferences
Single Protein	Setting Single-Protein Search Preferences

❖ **To access the Options dialog box**

Choose **Tools > Options**.

Index

Symbols

A

B

C

D

E

F

G

H

I

J

L

M

N

O

P

Q

R

S

T

U

V

W

X

Y

Z