# PEAKS 8 User Manual

**PEAKS Team**

# PEAKS 8 User Manual

PEAKS Team

Publication date 2016

# Table of Contents

# Chapter 1. Overview

Welcome to PEAKS.
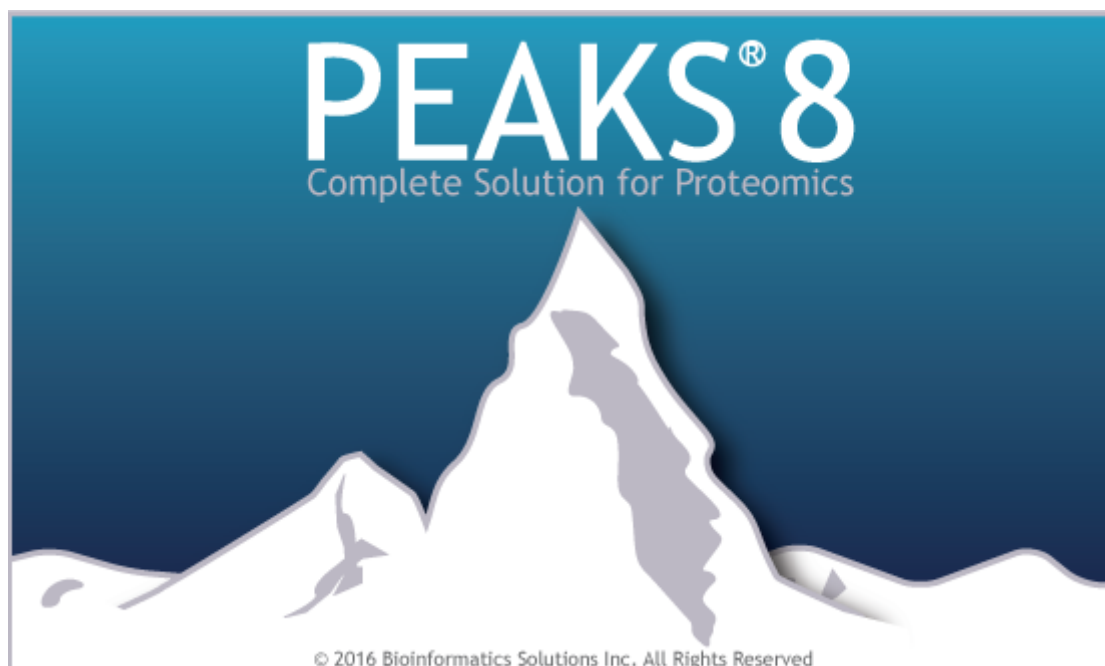


## 1. How to Use This Manual

This chapter provides an overview of PEAKS' distinctive features and describes a typical data analysis workflow in PEAKS. It is strongly recommended to read this chapter to get a big picture of what PEAKS provides and how PEAKS is used. Other parts of this manual are intended for reference and so it does not need to be read from cover to cover. Many contents of this manual can be read from the software's online help. The electronic (and most up-to-date) version of this manual can be found at `http://bioinfor.com/doc/peaks8/htmlmanual/index.html`.

The installation of the software is covered in a separate chapter : Chapter 2, *Installation and Activation*.

## 2. What Is PEAKS?

PEAKS is a complete software package for proteomic mass spectrometry data analysis. Starting from the raw mass spectrometry data, PEAKS takes care of every step in the analysis, including data conversion; peptide and protein identification; PTM and mutation characterization; as well as result validation, visualization, and reporting.
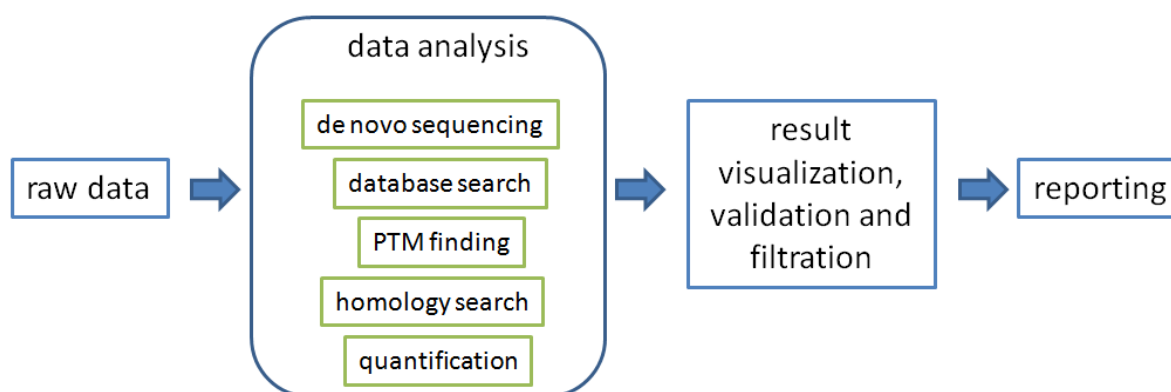


PEAKS helps in every step of the proteomics data analysis.

The following is a partial list of the distinctive features of PEAKS software:

- **Complete Analysis**

  PEAKS is a complete software package for *de novo* sequencing, database searching, characterizing unspecific PTMs, and detecting peptide mutations. The combined use of these four algorithms maximizes the number of identifications.

- **Built-in Result Validation**

  The "decoy-fusion" method is used to validate the peptide identifications automatically. The "decoy-fusion" method is an enhanced result validation method that avoids several pitfalls existing in the commonly used target-decoy method. With an informative result summary view, the results can be easily filtered by setting a false discovery rate (FDR).

- **Protein Coverage View**

  All the identified peptides are mapped to the identified proteins and displayed in a consolidated protein coverage view. All of the PTMs and the mutations in the protein sequence are highlighted. The interactive graphical user interface (GUI) allows for the in-depth and effortless examination of every amino acid in the identification results and every peak in the data.

- **Accuracy and Sensitivity**

  The superiority of the PEAKS algorithms has been demonstrated by publications and third-party evaluations. By combining four complementary algorithms, the sensitivity is further improved.

- **Multiple-enzyme Project**

  To maximize a protein's sequence coverage, it is advantageous to use different proteolytic enzymes to digest the protein sample. PEAKS supports the use of different enzymes in different samples of a single project. The software will automatically use the enzyme specified in each sample and then combine all sample results together.

- **Label Free Quantification**

  An Expectation-Maximization (EM) based algorithm is used for feature detection, deconvolution and refinement. PEAKS uses an optimization model for simultaneous feature matching and retention time alignment. The efficient and accurate algorithms used for Label Free Quantification can even handle significant time shifts and distortions. The quantification ratios can be normalized using total ion chromatogram (TIC) or selected proteins.

- **Multiple Instrument Vendor Support**

  PEAKS supports the raw file formats of most major instruments, which saves the trouble for file format conversion. More importantly, the algorithms have been tuned for different instrument types to ensure optimal accuracy and sensitivity.

- **Result Exporting**

  The analysis result can be exported to a variety of text formats. The HTML format makes it easy to share the results on a website, while other text formats such as .csv make it easy to view the results in spreadsheets and to post-analyze with either the in-house script or other third-party software. PEAKS also supports exporting to standard result formats such as pepXML and mzIdentML.

- **inChorus**

  PEAKS can import results from other database search engines (already acquired by the lab) and combine all the results together. PEAKS can estimate FDR for the combined result and can allow for result filtration. For each identified peptide, PEAKS also displays all the engines that identified that peptide, as well as each engine's identification score.

# 3. What Is New in PEAKS 8?

Here is a list of the noteworthy new features in PEAKS 8.

**Highlights.**

- **New TMT/iTRAQ Quantification module**

  For datasets generated from TMT/iTRAQ quantification experiments, PEAKS 8 supports MS2 and MS3 quantification, cross-experiment comparison through the use of spike channels, automatic protein FDR estimation, and the manual removal of low quality spectrum.

- **New SILAC Quantification module**

  For datasets generated from SILAC quantification experiments, PEAKS 8 analyses data a feature-based approach that supports easy sample grouping and replicate analysis, employs both interactive validation for feature details and an interactive volcano plot for proteins, and automatically corrects arginine to proline conversions.

- **Contaminant database support for identification**

  PEAKS 8 enables searching a target database together with a contaminant database and allows contaminant proteins to be shown or filtered out.

- **Enzyme-specific display for protein coverage**

  PEAKS 8 enables users to display protein sequences in different enzyme formats.

**Additional Improvements.**

- Improved feature detection.

- Improved feature and PSM association.

- Improved label free quantification

- Spectrum view supports absolute intensity and fragment ion error display.

- Support for ProteoWizard MS converter.

- Support for Thermo MS file reader 3.1 SP2.

- Improved PTM profile exporting.

- Support exporting for fragment ions.

- Better support for mzIdentML formats.

- Better multi-user support

# 4. Quick Walkthrough

This section presents a quick walkthrough of a typical data analysis workflow. The PEAKS user interface is introduced by using the sample project included in PEAKS installation, filtration and visualization of the analysis result are showcased, which will help users to understand what can be accomplished with PEAKS. After that, a PEAKS project created from the analysis of a sample of raw data is demonstrated.

## 4.1. Opening an Existing Project

The installation instructions for PEAKS can be found at Chapter 2, *Installation and Activation*. After launching PEAKS, the sample project can be opened using one of the following two methods (see screenshot below):

1. For fresh installations, click the "Sample Project" in the "Recent Projects" list on the Start Page.

2. For all installations, click the Open Project icon, browse to the directory where PEAKS 8 was installed, select "SampleProject", and click the Open button in the file browser.
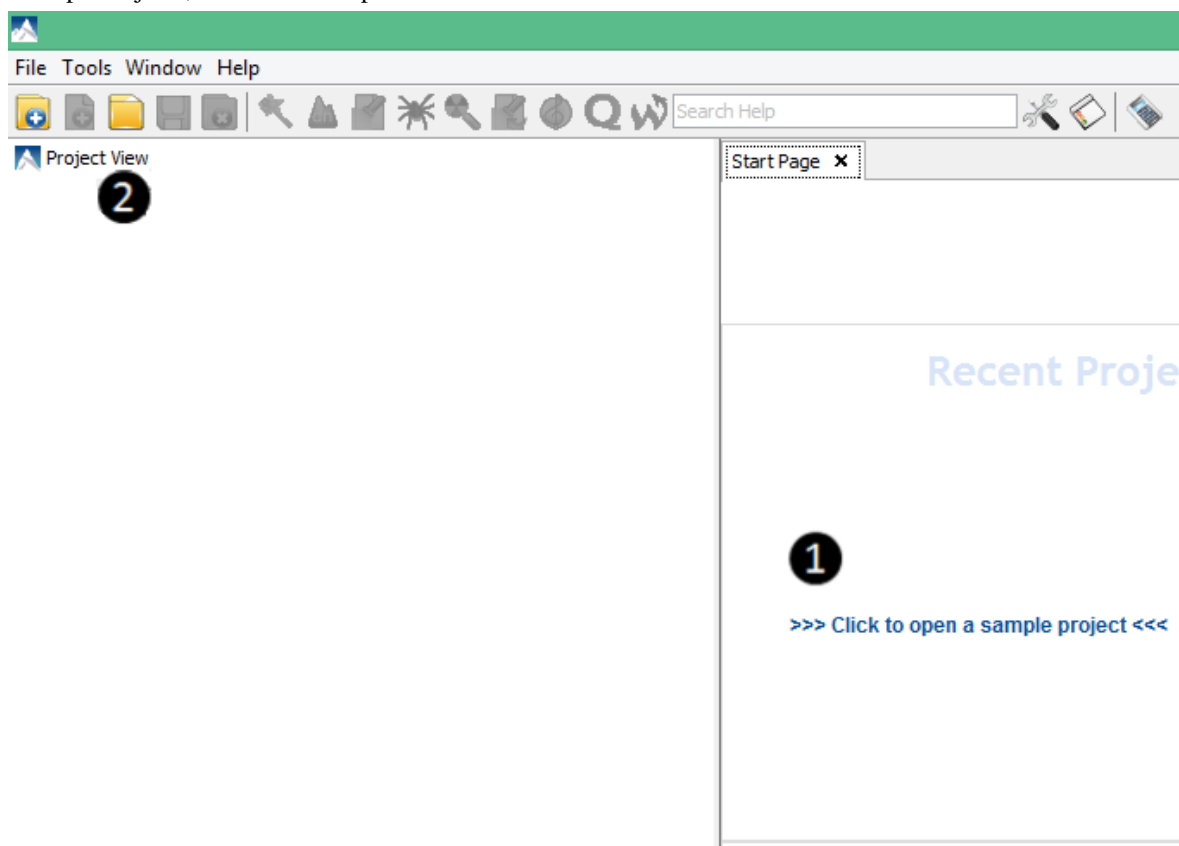


## 4.2. PEAKS Main GUI

The main graphical user interface (GUI) of PEAKS is divided into the following areas (see screenshot below):

1. The Project view shows all of the opened projects. Each project may include multiple samples and each sample may include multiple fractions (LC-MS runs). The analysis results are displayed as result nodes under the project.

2. The menu and toolbar contain all of the analysis tool icons. Selecting a node (project, sample, fraction, or result) in an opened project will highlight the analysis tool icons that are available to the selected node.

3. A result node in a project can be opened by double-clicking the node. Opened result nodes are shown in tabs.

4. Each opened result node provides several different "views" as different tabs. In particular, the "Summary" view shows the result statistics and is also the central tab in which to filter and export PEAKS results.

5. The information pane shows useful information such as the node properties and the progress of running tasks.
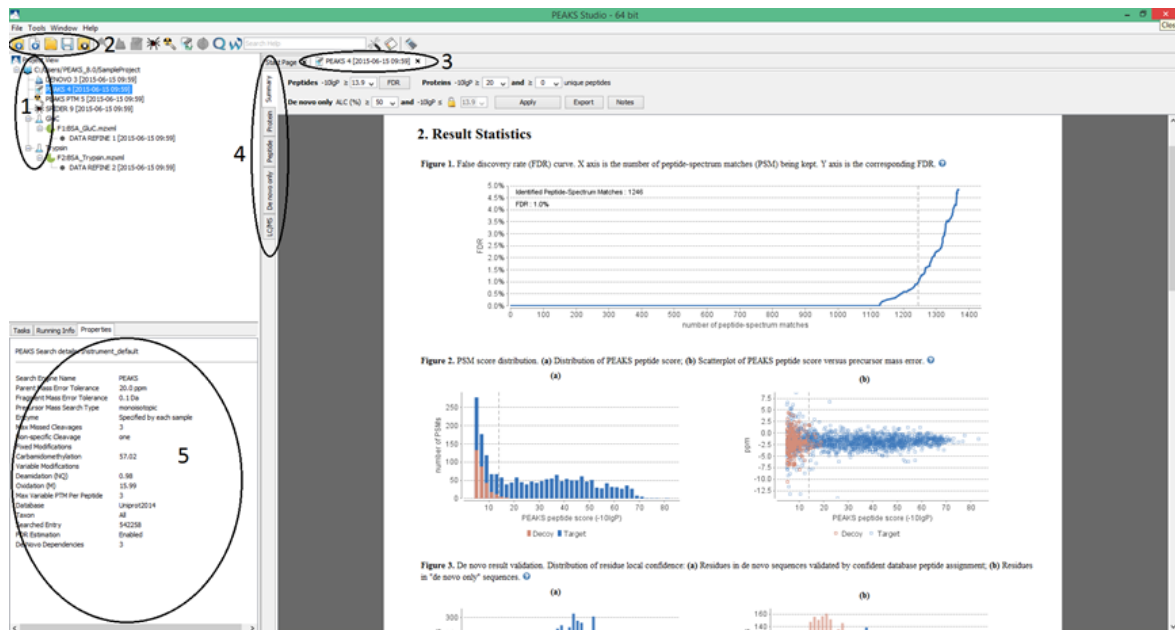
# 4.3. Result Summary and Filtering

After double-clicking to open a result node, i.e. the PEAKS DB node in the sample project, the "Summary" view is shown by default and provides mainly three functions:

1. Specify score thresholds to filter the results.

2. Examine the result statistics.

3. Export the results.

The top region of the "Summary" view is a control pane in which the result filters are controlled (see screenshot below), and the bottom region is a statistics report page:

- Low scoring peptide identifications are filtered out by setting a desired peptide-spectrum match -10lgP threshold or a desired FDR (false-discovery rate), which can be specified by clicking the FDR button.

- Low scoring protein identifications are filtered by setting a desired protein -10lgP score and a desired unique peptide count.

- The *de novo* only peptides are the confident *de novo* sequence tags that remain unidentified by the database search algorithms. To report a *de novo* only peptide, the ALC (average local confidence) scores must be equal to or better than the specified threshold. Meanwhile, the score of the spectrum's best database search result should be no greater than the specified -10lgP threshold.

    By default, the -10lgP threshold used for *de novo* only peptides is locked to be the same as the -10lgP threshold used for filtering peptides. To specify a different value, simply click the lock icon to unlock the filter.

After the filtering conditions are changed, the Apply Filters button will change to red. Click it to apply the new criteria.
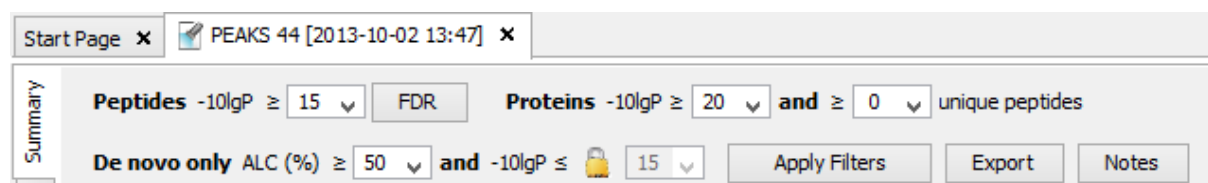


The top control pane has two additional buttons: Export and Notes. Clicking Export displays several options and file formats in which a user can export analysis results. Clicking Notes allows the user to add a text note about the project that will be displayed at the top of the "Summary" view page.

After applying the filters, the statistics report page at the bottom of the "Summary" view will be updated accordingly. Only two statistical charts are explained here (see screenshot below).

Figure 2(a) shows the PSM score distribution in a stacked histogram. If both he search result and the peptide -10lgP score threshold are of high confidence, very few decoy matches (brown) in the high score region should be observed. Additionally, if the FDR estimation method (decoy fusion) worked properly, then a similar or greater number of decoy (brown) matches should be observed in the low score region relative to the target (blue) matches.

Figure 2(b) plots the precursor mass error in ppm v.s. -10lgP peptide score for all the PSMs. This figure is the most useful for high resolution instruments. Generally, the high-scoring points should be centered around the mass error of 0. Notice that the data points start to scatter to larger mass errors when they are below a certain score threshold.



**Figure 2.** PSM score distribution. **(a)** Distribution of PEAKS peptide score; **(b)** Scatterplot of PEAKS peptide score versus precursor mass error.
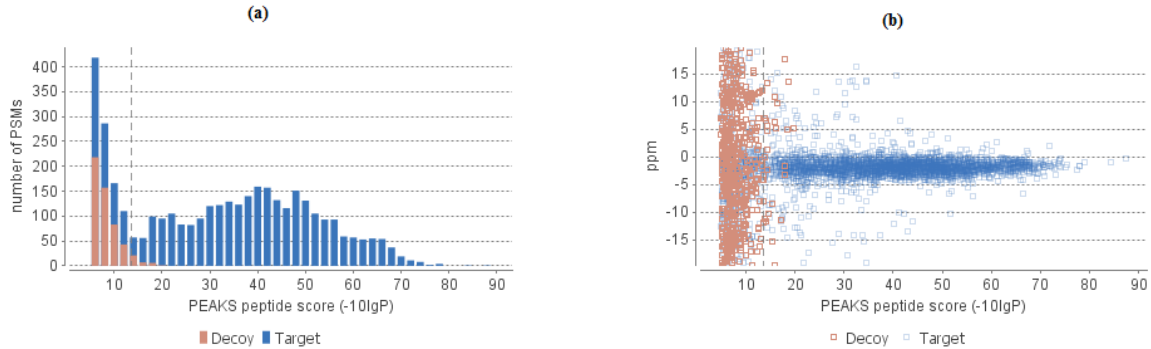
# 4.4. Result Visualization

Besides the "Summary" view, there are four other views for visualizing the data and the results: "Protein", "Peptide", "De novo only", and "LC/MS."

- The "Protein" view contains a list of proteins that passed the filters. The proteins identified with the same set (or a subset ) of peptides are grouped together.

- The "Peptide" view shows all the peptide identifications passing the filters. The multiple spectra that identified the same peptide sequence are grouped together.

- The "De novo only" view shows all of the peptides identified exclusively by *de novo* sequencing.

- The "LC/MS" view displays the LC-MS data as a heat map with highlighted MS/MS scans and detected features.

Here, the focus will be on the protein coverage view only. Click the "Protein" view tab and select one protein. The corresponding protein coverage map will be displayed at the bottom of the "Protein" view. The protein coverage view maps all peptide identifications of the selected protein onto the protein sequence. It enables the effortless examination of every PTM and mutation on each amino acid. Some of the most commonly used operations on the protein coverage view are listed (see screenshot below):

1. Each blue bar indicates an identified peptide sequence. Each gray bar indicates a *de novo* only tag match.

   Peptide identifications with the same amino acid sequence and the same PTMs are grouped together and displayed as a single bar.

2. PTMs and mutations are highlighted with colored icons and white letter boxes. Highly confident PTMs and mutations are displayed on top of the protein sequence.

   A PTM or mutation is regarded as confident if the two fragment ions at both sides of the modified residue have relative intensity values higher than the user-specified threshold in the display option (see item 5).

3. Click a peptide to show the spectrum annotation.

4. Hold the mouse over an amino acid to show the supporting fragment ion peaks.

5. These options to control the coverage view display.

   - The "coverage/outline" choice enables the display or the removal of the peptide bars.

   - The "sequence display option" highlights the coverage map based on either a fasta display or an enzyme display from one of the enzyme options selected during the project setup.

   - The "de novo tags sharing" specifies the minimum number of consecutive amino acid matches between a *de novo* only sequence and the protein before it can be displayed as a gray bar.

   - The "de novo peptides fully matched" checkbox allows a *de novo* peptide to be displayed if the sequence, regardless of its length, is fully matched to a sequence in the protein.

   - The "minimum ion intensity" specifies the minimum fragment ion relative intensity in one of the MS/MS spectra before a PTM location is regarded as confident and displayed on top of the protein sequence.

   - The checkboxes in the PTM list specify which PTMs are "interesting". Click the colored boxes to change a color. Double-click a PTM name to see the PTM detail.

6. The full screen button, the PTM profile button, the peptide mapping button, and the tool box button.

   The full screen button enables expansion of the coverage view. The PTM profile icon shows the position of and the modification ratio of a confidently modified peptide. The peptide mapping icon shows Annotated Chromatogram and Peptide Comparison displays between different fractions. The tool box icon provides common tools, such as exporting the coverage pane as a high-resolution image file.

## 4.5. Creating a PEAKS Project

To create a new PEAKS project from raw data files, perform the following steps (see screenshot below):

1. Select New Project ... from the file menu or click the New Project icon  on the toolbar. The "Project Wizard" will appear.

2. Use the Add Data button to add the desired files to be loaded, and then click Open. All of the selected data files will be listed on the left side.

3. Place the selected data from the list into samples: use  to place all files in a new sample; use  to put them in an existing sample; or put them in individual samples for each file using .

4. Click the  or the  buttons to add a sample to the project or data files to a sample, respectively.

5. For each sample, specify the sample details: "Instrument" type, "Fragmentation" method and "Enzyme" name.

## Note

Each sample can use a different proteolytic enzyme. Using multiple enzymes to analyze the same proteins can produce overlapping peptides, which will increase the protein coverage.

## Note

To apply the same sample details to the whole project, select the sample with the correct settings and click on the Copy to Whole Project button.

6. Click the Finish button to create the project.
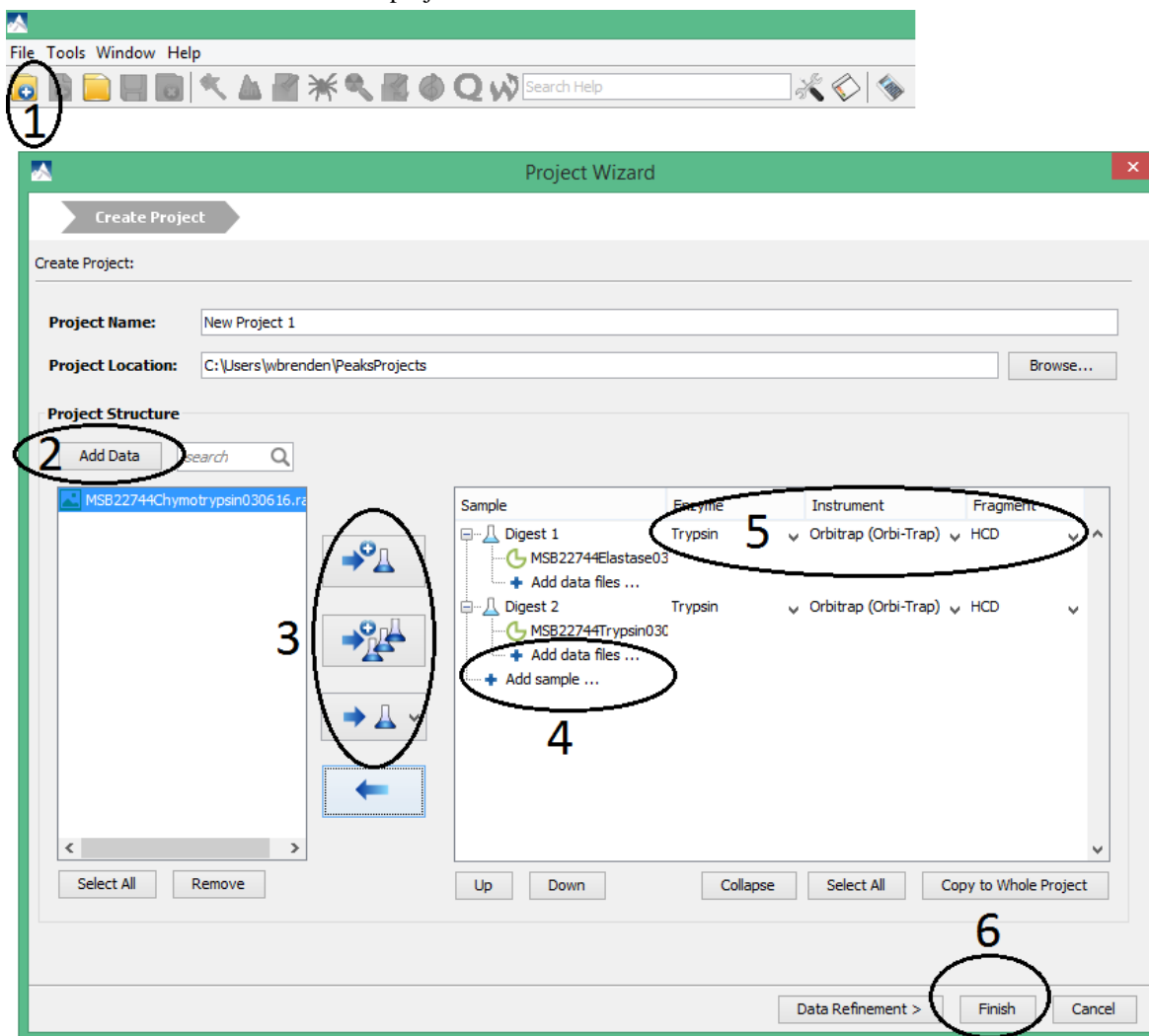


## 4.6. Conducting an Analysis

To conduct an analysis: 1) Select a project, a sample, or a result node in the Project View; 2) Click the desired analysis tool button. Here, PEAKS Search is shown as an example. PEAKS Search is a database search workflow for a complete identification analysis.
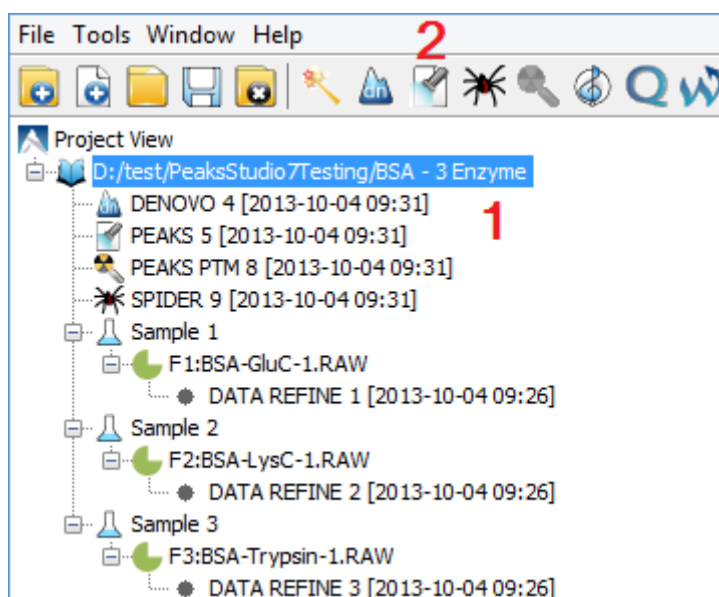
A search parameter pane will pop up. Most search options for PEAKS DB are standard and straightforward. More details are provided in the following (see screenshot below):

1. If the proteolytic enzyme was specified for each sample at the project creation step, one can choose to use the enzyme specified in each sample. This makes it possible to use multiple enzymes in a single project and a single search.

2. Specify the fixed PTMs and a few common variable PTMs expected in the sample.

3. Select a protein sequence database or copy and paste the protein sequences for the database search. Select a contaminant database to be included in the database search.

4. Conduct *de novo* sequencing using the same parameters or base the search on an existing *de novo* sequencing result node.

5. Estimate the false discovery rate (FDR) with the decoy fusion method.

   Decoy fusion is an enhanced target-decoy method for result validation with FDR. Decoy fusion appends a decoy sequence to each protein as the "negative control" for the search. See BSI's web tutorial (`http://www.bioinfor.com/peaks/tutorials/fdr.html`) for more details.

6. Enable PEAKS PTM and SPIDER algorithms after the PEAKS DB database search completes.

   By default, PEAKS PTM performs a blind search for additional PTMs in the data. Users can limit the PEAKS PTM search on a large number of PTMs by clicking the "Advanced Setting" button.

   SPIDER performs a homology search based on *de novo* sequencing tags. If selected, the SPIDER algorithm will be conducted on every confident *de novo* tag (ALC>15%) whose spectrum is not identified by PEAKS DB with high confidence (-10lgP<30). SPIDER will construct new peptide sequences by altering amino acids of database peptides. For each spectrum, the better sequence constructed by SPIDER or found by PEAKS DB will be used as the identified peptide. SPIDER is good for cross-species searches and for finding point mutations of the protein. SPIDER can either be invoked through this workflow or by clicking the SPIDER icon on the toolbar.

# Chapter 2. Installation and Activation

This section of the manual will guide users through the installation and activation of PEAKS.

## 1. Package Contents

The PEAKS package contains:

• This manual

• PEAKS Software

• Quick reference sheet for mass spectrometry

## 2. System Requirements

PEAKS 8 is available in both 64-bit and 32-bit versions. Installation of the 64-bit version, which is the default version that BSI distributes, is highly recommended for most users. For users with a 32-bit operating system, the 32-bit version can be downloaded upon request. PEAKS 8 supports Windows operating systems (Windows XP, Windows 7, Windows 8, and Windows 10). PEAKS program files will only use ~200MB disk space, but it will require some extra disk space to store temporary files and intermediate results. The amount of disk space required for this purpose depends on the size of the user's datasets.

The two main factors affecting PEAKS performance are CPU and RAM.

The following are the requirements for the default 64-bit version of PEAKS 8, the recommended and ideal requirements are :

• **Recommended**: Intel Core processor and 8GB of RAM.

• **Ideal**: Intel Core i7 or Xeon processors and 16GB of RAM or more (or 2GB per core).

For the 32-bit version of PEAKS 8, 4GB RAM is required.

## 3. Installation on a Windows Computer

### Important

Please uninstall any older version of PEAKS currently installed on the system before proceeding.

### Important

Avoid installing PEAKS in any directory that contains a white space. For example, the `Program Files` directory has some features that may not function correctly in such situations. Please make sure that the user account has full access permissions (read/write/execute) on the selected directory.

### Important

To open an instrument's raw data using PEAKS, it is necessary to install PEAKS on the same computer where the instrument vendors' own software is installed. Refer to Section 8, "Vendor-Specific Requirements" for the vendor-specific requirements load data.

1. Close all programs that are currently running.

2. Insert the PEAKS installation disc into the DVD drive. If an installation file was obtained elsewhere, double-click on the PEAKS installation file and move ahead to step 4.

3. The installation window should automatically appear after the DVD is inserted. If it does not, find the DVD drive and open it to access the disc. Double-click on `PEAKS_Studio_Installation.exe`.

4. A menu screen will appear. Select the top item, "PEAKS Installer". The installation utility will launch the installer. When the PEAKS installation dialogue appears, click the Next button.
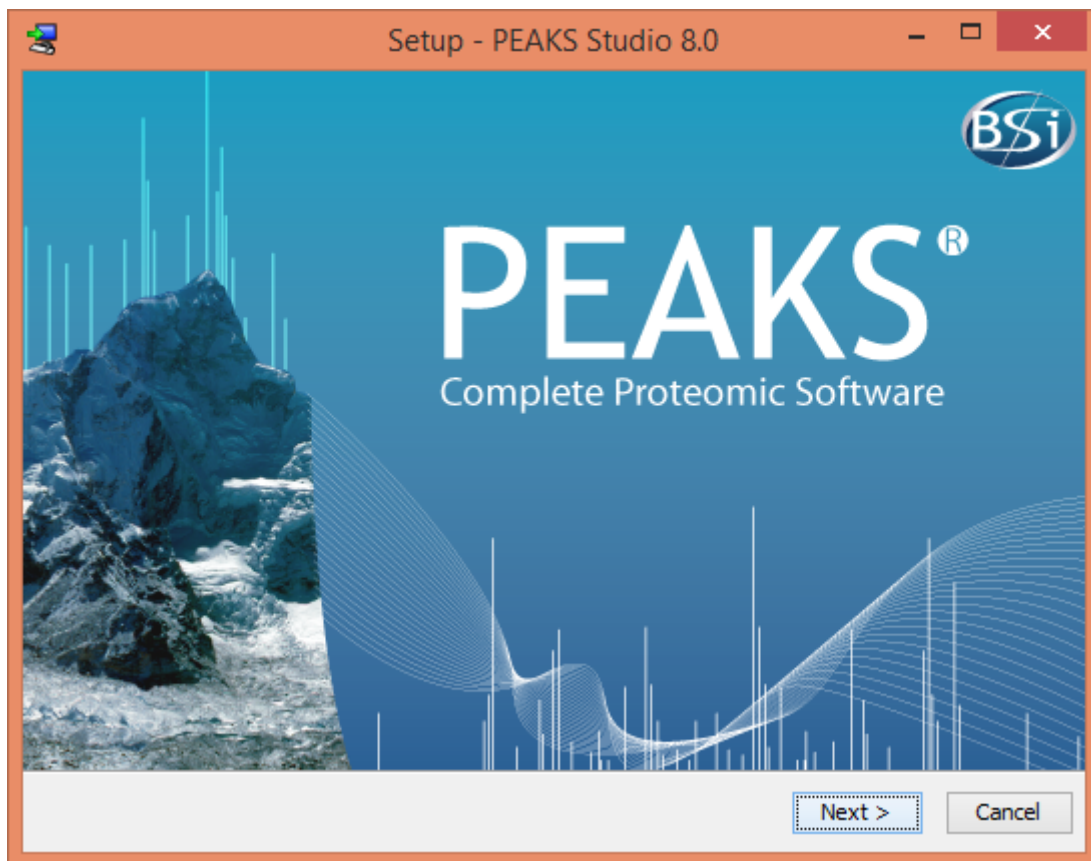


5. Follow the on-screen instructions to finish the installation.

# 4. Activation

All users are required to go through a software activation process in order to use PEAKS. A License Wizard will appear to guide users through the activation process the first time PEAKS is launched.
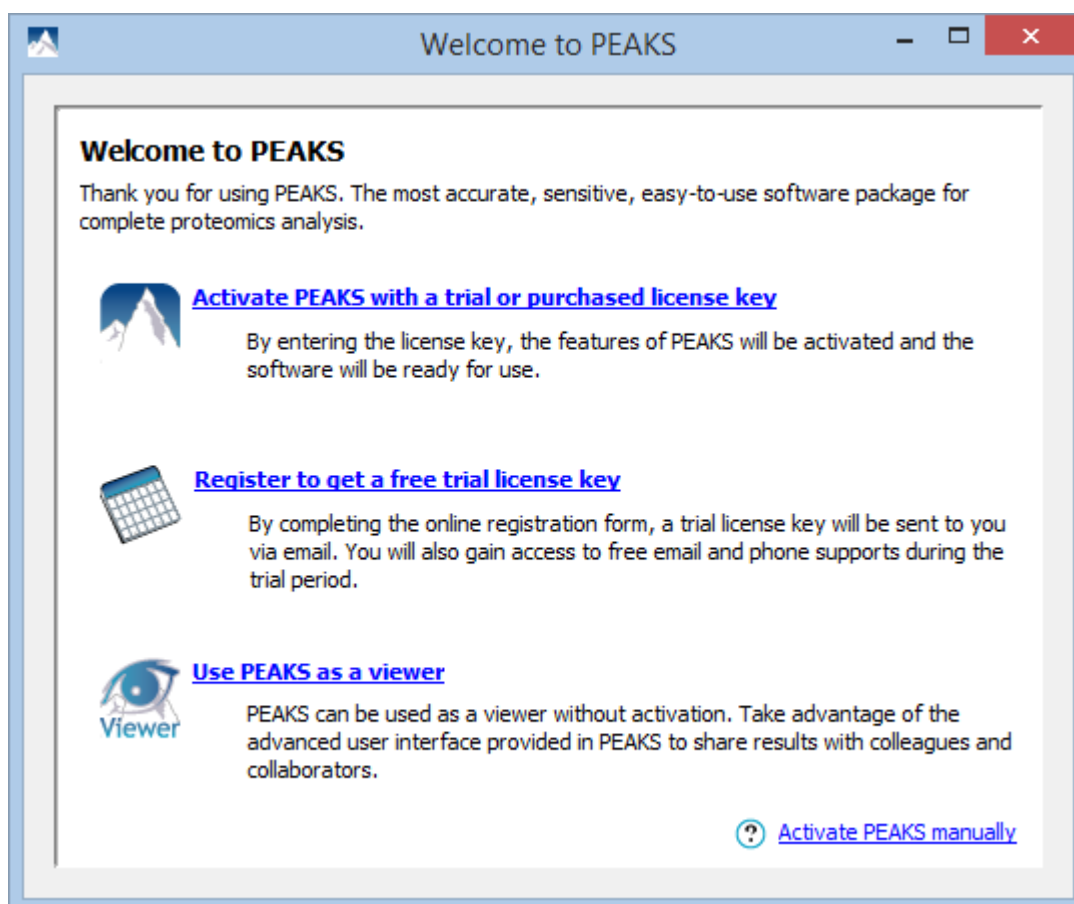
## Important

Keep the license key safe. After a computer hardware upgrade, it might be required to re-activate the software.

# 4.1. Activate PEAKS with a trial or purchased license key

The software activation process is very simple. If the computer is connected to the Internet, PEAKS can be activated by clicking on the first option in the wizard. In the 'Enter the License Key' dialogue, paste or type in the license key and click the Activate button. If the activation is successful, PEAKS will start normally. Manual activation is necessary when the activation failed with the message 'An error occurred while communicating with BSI licensing server ...'. Refer to Section 4.4, "Activate PEAKS manually" for manual activation.
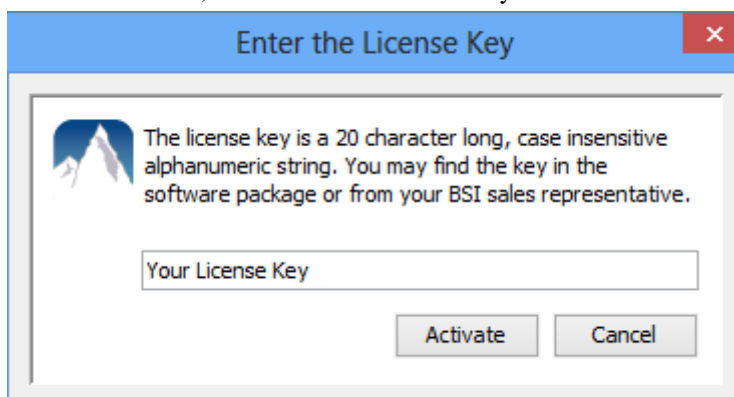
## 4.2. Register to get a free trial license key

This option allows new PEAKS users to evaluate the software before purchase. If the computer is connected to the Internet, clicking on this option will bring up a web form in the default web browser. Please provide your full name, institution, email address and phone number in the form. After the form is submitted, an email from `<support@bioinfor.com>` will be sent with the trial license key. This key can then be used to fully activate PEAKS for evaluation purposes.

> ### Important
>
> Each computer can only have one free trial. Requesting a trial license repeatedly will not extend the trial period automatically. Please contact `<support@bioinfor.com>` if you have any questions.
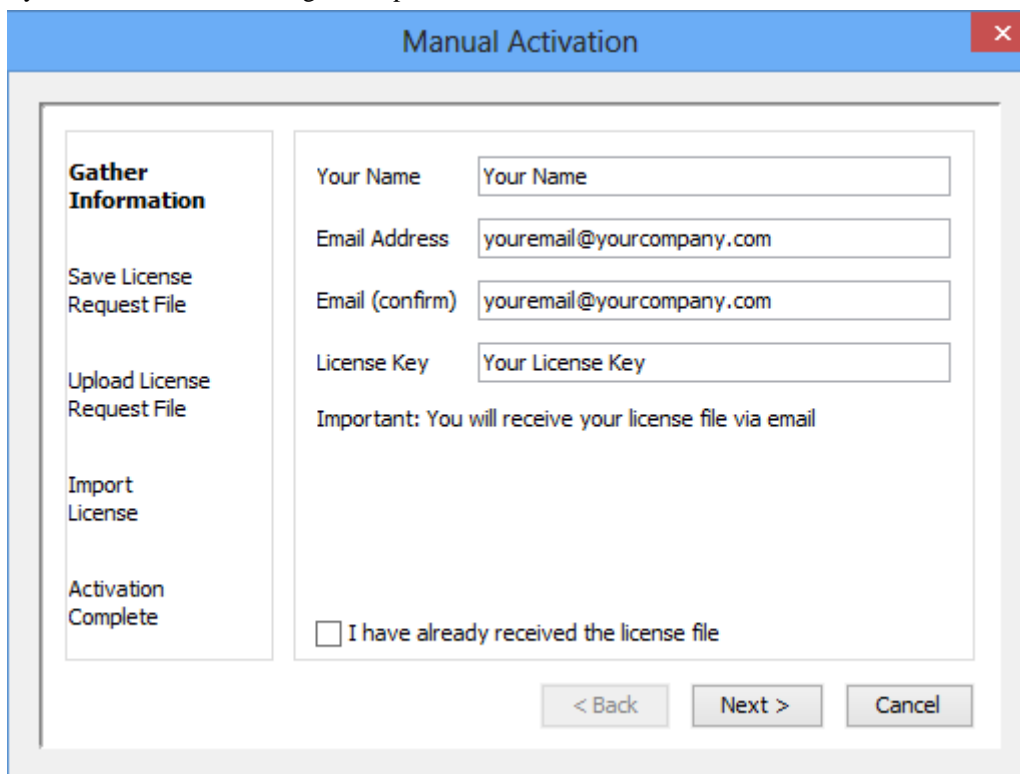
## 4.3. Use PEAKS as a viewer

BSI has consolidated the PEAKS product line. PEAKS Studio and Viewer are now one product. Unlicensed PEAKS Studio can be used in the viewer mode. Researchers around the world are now able to take advantage of PEAKS' most advanced user interface to share PEAKS results. In the viewer mode, all non-analytical features, such as read/display raw data, open/navigate through existing PEAKS results, exporting, etc..., work the same way as the full PEAKS Studio version.

## 4.4. Activate PEAKS manually

In certain situations, when the computer does not have an Internet connection (or is behind a firewall that blocks the activation), the activation process requires the assistance of another computer with an Internet connection (or outside the firewall). Manual activation can be accessed via the link on the bottom-right corner of the wizard. Manual activation consists of the following steps.

1. Provide the license key and user information required to the license wizard on the computer that will be running PEAKS. The option "I have already received the license file" should be check when the license file from BSI is already at hand. Click Next and go to step 5.



2. Save a generated request file to a removable storage device (e.g. a USB memory key).
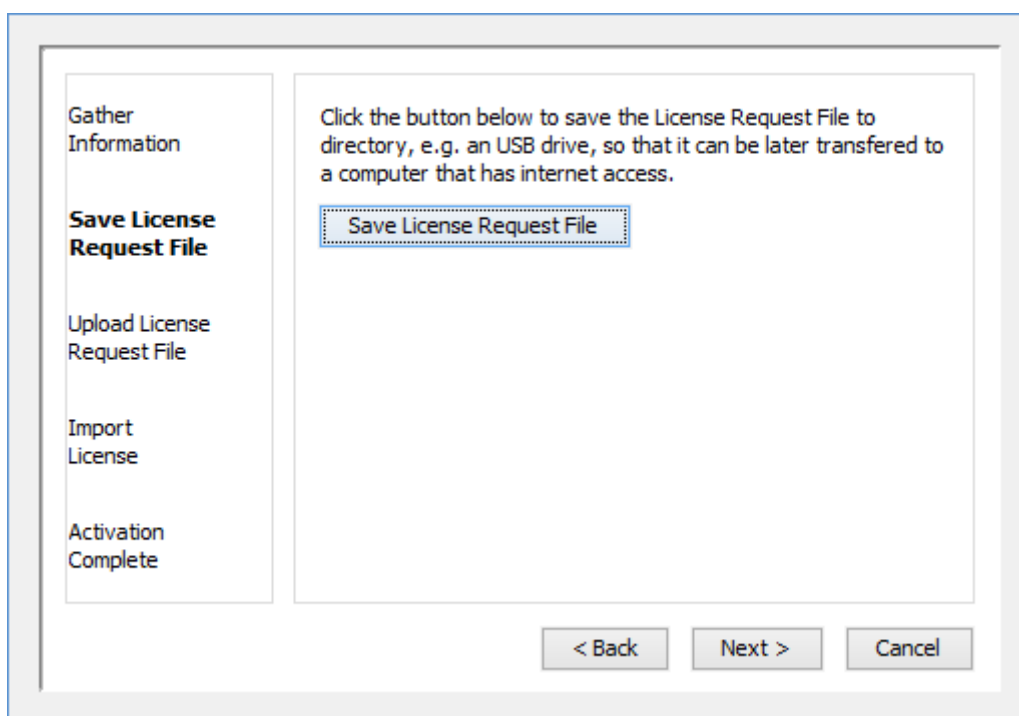
3. From another computer with an Internet connection, upload the license request file to BSI's license server as described in the following screen.
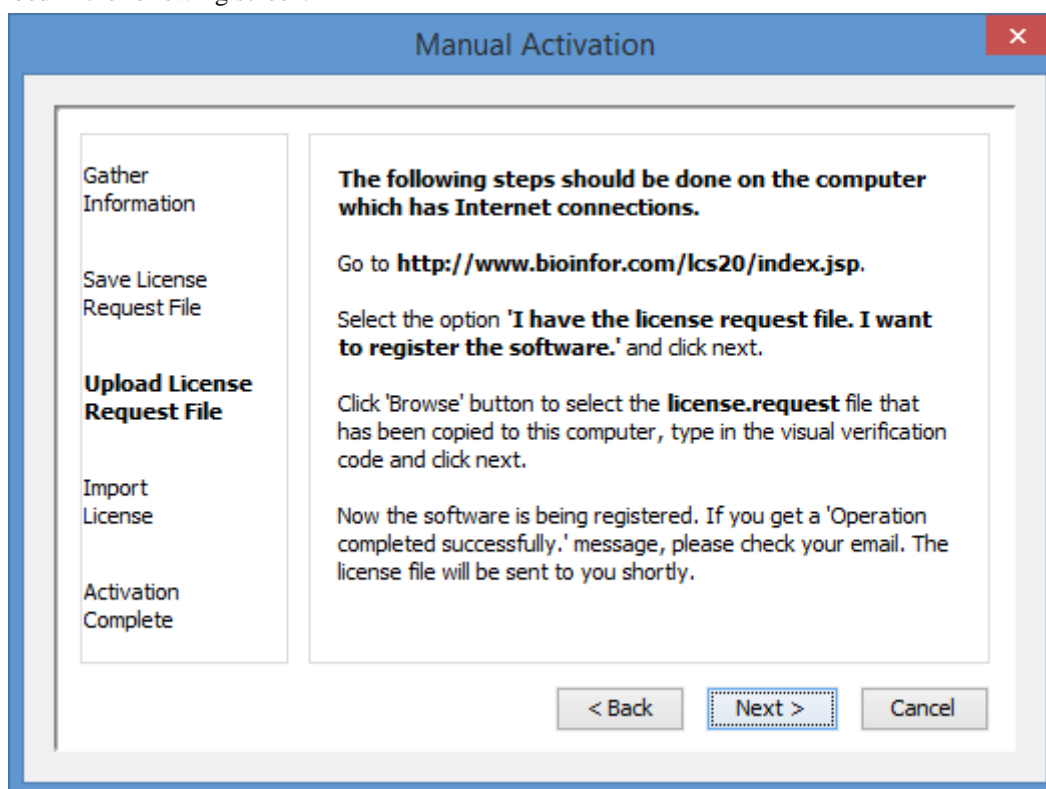
4. If the license request is sent successfully, an automated BSI service will generate the license file (`license.lcs`) and send an email (from `<support@bioinfor.com>`) to the email address provided to the License Wizard. Either save the attached license file or copy the content between '===>' and '<===' in the email to the Windows clipboard.

5. Transfer the license file to the computer running PEAKS and import the license file into the License Wizard.



## 4.5. Re-registering PEAKS

Re-registering PEAKS may be necessary when an additional software module was purchased or SPS was renewed. BSI will modify the license information accordingly on the server side. A new license file is required to make the changes effective. Select About PEAKS from the Help menu. The "About BSI PEAKS Studio" dialogue box will appear:
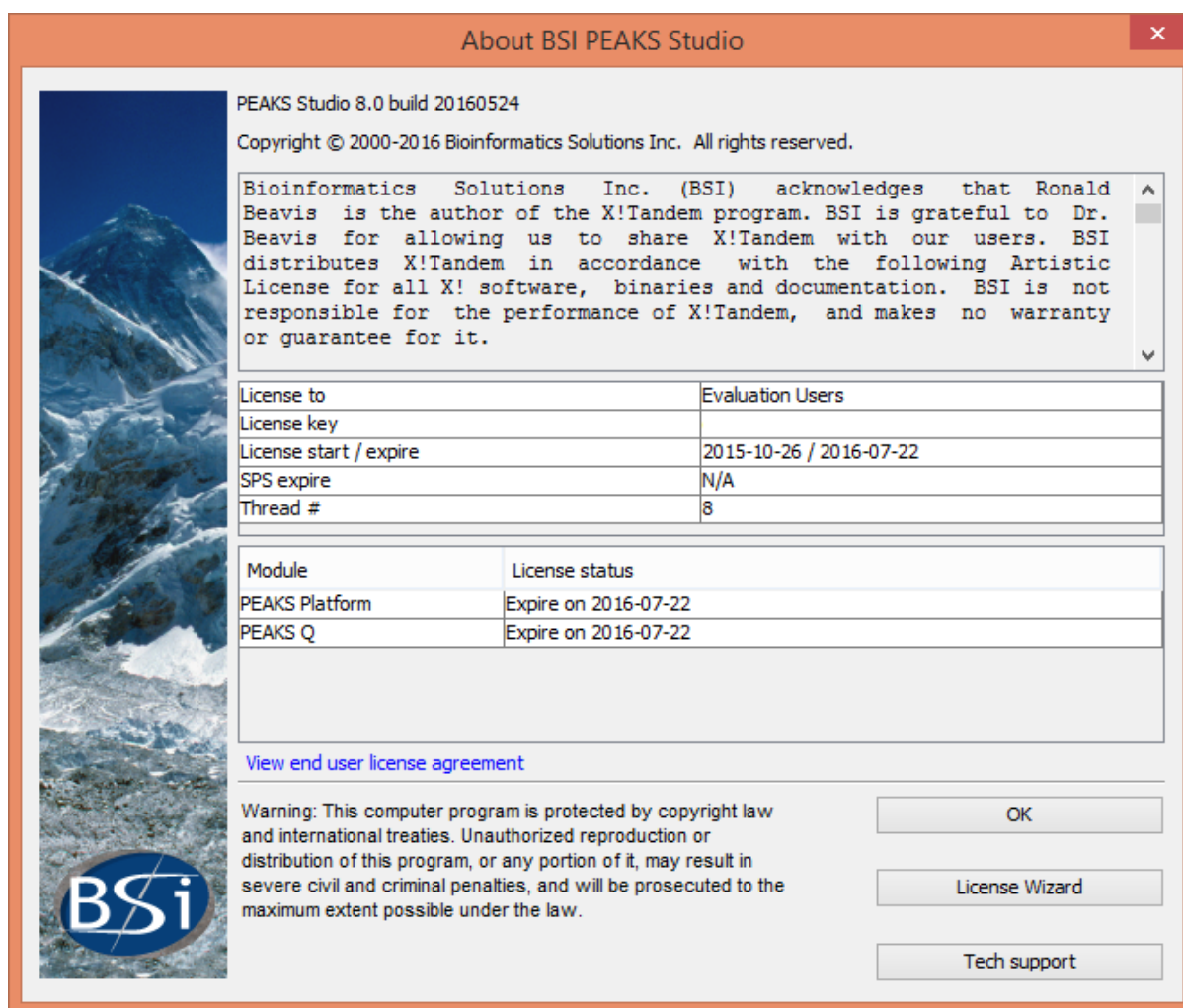
Click the License Wizard button to continue then follow the instructions in Section 4.1, "Activate PEAKS with a trial or purchased license key" or Section 4.4, "Activate PEAKS manually" to re-register PEAKS.

## 4.6. Common Errors during Registration

1. The license key contains only English letters and numbers. It is recommended to copy (Ctrl-C) / paste (Ctrl-V) the license key whenever possible.

2. The user information can only contain English characters (letters, digits and symbols). Characters from a non-English encoding may cause the activation to fail.

3. If the computer is behind a firewall or has other Internet connection problems, the activation may fail. Please follow the on-screen instructions or refer to Section 4.4, "Activate PEAKS manually".

# 5. PEAKS Performance Configuration

The PEAKS Performance Configuration tool can be accessed from the Windows Start Menu in the PEAKS Studio menu item. By default, PEAKS automatically determines its performance parameters to take full advantage of the processors and memory available on the computer. In most situations, the **Automatically Configure PEAKS Performance** option should be used. The **Manually Configure PEAKS Performance** option will come in handy in situations where advanced users want to adjust PEAKS memory allocation and threading.
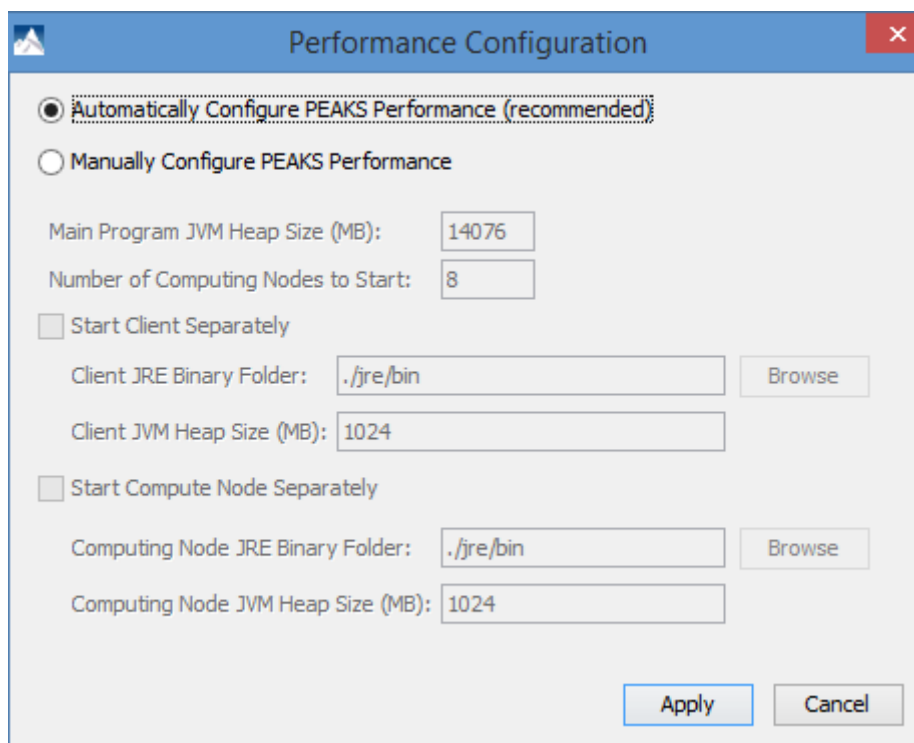
The *Number of Computing Nodes to Start* determines the number of parallel processing threads/processes that will be used by PEAKS. A higher number usually lets PEAKS run faster if the computer still has extra processing power. The maximum number of computing nodes PEAKS may start with is determined by the license; a *Desktop License* allows up to 8 nodes and a *Workstation License* allows up to 16 nodes.

The configuration is very different between the 64-bit version and 32-bit version due to the amount of memory Java can access. In both scenarios, if out-of-memory problems are experienced, the number of nodes in PEAKS should be reduced and the amount of memory per node should be increased.

## 5.1. PEAKS 64-bit Version

The 64-bit version of PEAKS can access a large amount of memory; therefore, it is highly recommended to deselect the options *Start Client Separately* and *Start Compute Node Separately*. The *Main Program JVM Heap Size (MB)* is the maximum amount of memory PEAKS can use. Although a higher number is better, it is recommended to use no more than 80% of the system memory. The *Number of Computing Nodes to Start* can now be calculated. For example, let **X** MB memory be assigned to the main program, the number of computing node **N** is the largest integer less than **X / 2000**. As shown in the figure, the system has 16000MB RAM, therefore 16000*80%=12800MB memory is configured to the main program JVM. 12800/2000=6.4, therefore 6 computing nodes is configured.

## 5.2. PEAKS 32-bit Version

PEAKS 32-bit version is produced only for compatibility on older 32-bit operating system. It is strongly recommended to use PEAKS 64-bit version on 64-bit operating systems. Due to the memory limitation for 32-bit Java, a typical configuration for a computer with 4GB of RAM looks like the following.



The total amount of memory used by PEAKS is **Computing Node JVM Heap Size (MB)** multiplied by the **Number of Computing Nodes to Start** plus the **Main Program JVM Heap Size (MB)**. In this case, 1024*1+1024 = 2048 MB. In some systems, PEAKS may not start with 1024MB assigned to the main program if there are other programs running on the system. Lowering the **Main Program JVM Heap Size (MB)** to 800 may help solve the issue.

# 6. What's Next

Depending on the data formats and the type of analysis needed, two additional configuration steps may still be needed before data analysis can be conducted:

- To read the instrument's raw data formats, installation of specific software libraries to support the instrument may be required. See Section 8, "Vendor-Specific Requirements" for this requirement.

- To conduct a database search for protein identification, a protein (or EST) sequence database must be configured. See Chapter 6, *Adding a Sequence Database*.

# Chapter 3. Configuration Wizard - Configure Instruments and Public Databases

## 1. Overview

The Configuration Wizard guides you through some easy-to-follow steps to configure PEAKS for instrument raw data support and database searching. The configuration wizard can be invoked from the menu Window | Config Wizard.

PEAKS supports different instrument vendors' raw data formats. A list of supported formats can be found in Chapter 4, Section 7, "Supported Data Formats". Some vendors' formats may require the vendors' specific software to be installed on the same computer that PEAKS is running on. The configuration wizard helps you select the proper instrument and install the appropriate vendor software.

A sequence database must be configured in PEAKS to identify peptides and proteins with the MS/MS spectra by database searching. The Configuration Wizard also helps you select the appropriate databases from a list of public sequence databases, and also downloads and configures them automatically. The list of public databases can be found in Section 3, "Database Selection".

## 2. Instrument Selection

When you select Window | Config Wizard from the menu, the Configuration Wizard will run and display a welcome message. Clicking the Next button shows the following instrument selection panel.

## Select Instruments

**Thermo Scientific**

☑ Thermo LTQ, Orbitrap, FT Instruments

    Thermo data can be loaded using Thermo MSFileReader library.
    The config wizard will download and install the MSFileReader automatically.

**AB SCIEX**

☐ AB 5600 Instruments

☐ AB 4700/4800 Instruments

☐ AB SCIEX QTRAP Instruments

☐ AB SCIEX QSTAR Instruments

**Agilent Technologies**

☐ Agilent Ion Trap Instruments

☐ Agilent Q-TOF Instruments

**Bruker Daltonics**

☐ Bruker APEX, micrOTOF, HCT Ion Trap, TOF-TOF Instruments

**Shimadzu Corporation**

☐ Shimadzu AXIMA-CFR Instruments

**Waters Corporation**

☐ Waters/Micromass Q-TOF Instruments

**Varian Incorporation**

☐ Varian Mass Spectrometry Instruments

       [<< Back]   [Next >>]   [Cancel]   [Help]

The instrument selection panel lists all the supported instruments and their vendor-specific software. Select the instruments as necessary. A small description about the instrument and possible action to be taken by the configuration wizard is displayed once it is selected. Click Next to proceed.

# 3. Database Selection

The database selection panel lists the public databases available for automatic download. The automatically downloadable databases are NCBI nr, SwissProt, TrEMBL, and IPI human. Select the databases to be configured. A small description of the database and the action to be taken by the Wizard is displayed once selected. If a public database is already configured using this wizard then "[configured]" text in green will appear beside its name in the list. It can be selected again to overwrite the configuration.

**Select Databases**

☑ NCBI nr
> The nr database is compiled by the NCBI that contains non-identical sequences from GenBank CDS translations, PDB, Swiss-Prot, PIR, and PRF.
> The Configuration Wizard will download, extract and configure the database.

☐ UniProtKB/Swiss-Prot - [configured]

☐ UniProtKB/TrEMBL

☐ IPI

Database Path

```
D:\FastaDB                                                          [ Browse ]
```

[ << Back ]  [ Next >> ]  [ Cancel ]  [ Help ]

The "Database Path" displays the location where the configured database will be stored. Click the Browse button to change the default database path. Click Back to go back to the "Instrument Selection panel. Click Next to proceed to the "Download Information" panel.

# 4. Download Information

The download information panel displays all the requested instrument software and database downloads in two tables: "Downloads" and "Selected Instrument Software".

**Downloads -** The "Downloads" table shows all of the downloadable requests, their status, and the available actions on them. The status of an ongoing download is displayed in the "Progress" column. To cancel an ongoing download, click the Cancel button in the "Action" column. The databases will be configured automatically after the download finishes. Once completely downloaded, the Install button appears in the "Action" column for the instrument software. The software can be installed by clicking the Install button. The Next button remains disabled until all the downloads and the installations are completed. Previously installed instrument software can be reinstalled by clicking the Reinstall button that appears in the "Action" column once the corresponding item is installed.

**Selected Instrument Software -** The Configuration Wizard cannot download some vendor-specific software; the corresponding vendor needs to be contacted to get that software. If any of those software packages were selected in the instrument selection panel, then their information appears in the "Selected Instrument Software" table. Clicking the Display Information button shows information on how to get the software from the vendor in a pop-up dialogue.

Click Back to go back to the instrument selection or the database selection panels to change the selected items. Click Cancel to cancel the Wizard at anytime. All ongoing downloads will be cancelled if the Wizard is cancelled. Click Next to set preferences for some vendor instrument software. A warning message box will pop up in case any downloaded item is not installed.

# 5. Vendor Instrument Software Options

This panel allows users to set the raw file converter preferences of some vendor specific software. Refer to Chapter 17, Section 1.2, "Raw File Converter Preferences" for details on raw file converter preferences. After making appropriate changes to the raw file converter preferences, click the Next button to finalize the changes.

## Vendor Instrument Software Preferences

### Bruker (.yep/baf, fid)

◉ PEAKS Built-in Loader

Options

Bruker (.fid) data may contain several files. Do you want to merge them?  ○ yes  ◉ no

○ ProteoWizard - MSConvert

Location - msconvert.exe

### Shimadzu (.run)

◉ None

PEAKS won't be able to load Shimadzu (.run) raw files.
Please select a converter for raw file conversion to a standard format that PEAKS can load.

○ Shimadzu Converter

Location - run2xml.exe

[ << Back ]  [ Next >> ]  [ Cancel ]  [ Help ]

# 6. Commit the Changes

After configuring the necessary downloaded vendor-specific software and public databases, click the Next button to get to the final panel.

## Configuring Instruments and Databases

The following FASTA databases and/or Instrument software are installed successfully:
 Thermo MSFileReader
 IPI

The following FASTA databases and/or Instrument software are downloaded but not installed:
 ABI mzWiff
A database or an instrument software will be configured for use with PEAKS only if you do so explicitly by clicking the 'Install' button in the 'Download Information' panel.

No further configuration of PEAKS is required for the following items. But you must ensure that the relevant vendor software are installed on this computer.
 ABI Analyst
 AB SCIEX MS Data Converter
Please contact the Manufacturer representatives to obtain the required vendor softwares.

Press the 'Back' button to return to downloads and selection panels where you can change your selection of databases or instrument software for configuration.

Click 'Finish' button to commit the changes to configuration file.

[ << Back ] [ Finish ] [ Cancel ] [ Help ]

A summary about the changes that the Wizard is going to make will be displayed. Click the Finish button to commit the changes to the configuration file. The changes will be immediately available to PEAKS.

# Chapter 4. Project Wizard - Create a PEAKS Project, Load Data and Set up a Workflow

## 1. Overview

Mass spectrometry data needs to be loaded into a PEAKS project before any analysis can be done. After creation, a PEAKS project is shown as a project node in the top left corner of the PEAKS user interface. On the computer's file system, a project is saved as a directory that contains multiple files, including the compressed spectral data and the analysis results. To share a PEAKS project, you can copy the whole project directory. The project can be opened in another copy of the same version of PEAKS Studio or the free PEAKS Viewer.

The "Project Wizard" guides you through easy-to-follow steps to create a PEAKS project; set up the project structure; load data into the project; and set up the workflow for data refinement, identification, and quantification. To run the Project Wizard, simply click the New Project icon  on the toolbar or select the File | New Project ... command. The rest of this chapter discusses the details of Project Wizard, project creation, data loading and setting up a simple workflow.

> ### Note
>
> Continue to Data Refinement only if you wish to have one result for all loaded data. After setting up a project by following the workflow through the Project Wizard, one node for each *de novo*, PEAKS DB, PEAKS PTM, and SPIDER will be generated if all options are selected. To create separate results, click Finish and run data from the project tree.
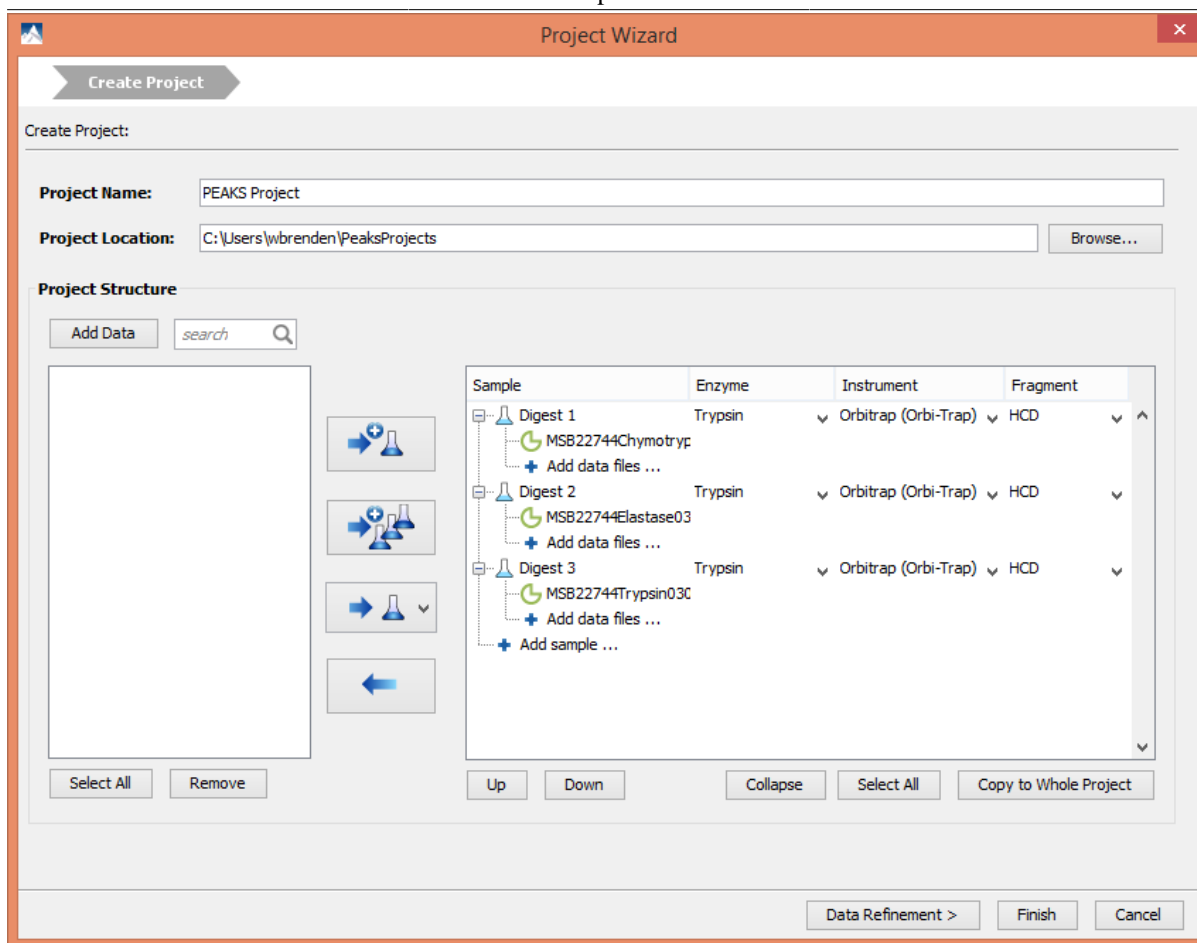
# 2. Create New Project

1. To create a new project, select New Project ... from the file menu or click the New Project icon  on the toolbar. The "Project Wizard" will appear.

2. Use the "Project Name" field to name your project. Click "Browse" to select a directory for saving the project. The selected directory will appear in the "Project Location" text box.

   ### Note

   Refer to Section 9, "Changing the Default Project Location" for changing the default save location for projects.

3. Use the Add Data button to browse to the directory of the data files, select the files you wish to load, and click Open. All of the selected data files will be listed on the left side.

4. This selected data can be placed into appropriate samples of the project structure using one of the following several options. All of the selected files can be placed in a new sample (  ), in an existing sample (  ), or in individual samples for each file (  ).

5. To add a sample to the project structure, click on the  button. From the drop-down menus, select the "Instrument" type, the "Fragmentation" method, and the "Enzyme" that that was used to generate the experimental data from the drop-down lists. To apply the same sample details to the whole project, select the sample with the correct settings and click the Copy to whole project button.

**Note**

Select "Mixed" as the fragmentation method if more than one fragmentation method was used and if each spectrum's fragmentation method is encoded in the data file. PEAKS loads the fragmentation mode for each spectrum during analysis. If the data file does not provide such information, these spectra will be ignored during analysis.

6. Use the **+ Add data files ...** button to add data files to the corresponding sample. Browse to the directory of the files, select the files to be loaded, and click Open.

7. To remove a sample or data file, select the appropriate node (sample or data file) in the project structure and press button. To delete the file entries from the list on the left, select the files and click Remove.

8. To change the order of the samples within a project or the data files within a sample, use the Up and Down buttons.

9. Click the Finish button to exit the Project Wizard without specifying parameters for further analysis. Otherwise, click the Data Refinement button to proceed to the data refinement step. See Section 3, "Data Refinement - Setting Parameters" for details on data refinement parameters.

Upon clicking the Finish button, the project will appear in the "Project View" panel. The outlined ![symbol] symbol indicates that the file is still loading. The solid ![symbol] symbol indicates that the file has finished loading.

PEAKS supports a variety of instrument vendor raw data formats. A list of supported formats can be found in Section 7, "Supported Data Formats". Some formats may require that the vendors' specific software be installed on the same computer on which PEAKS is running. Before creating a project with your own data, ensure that the vendor-specific requirements discussed in Section 8, "Vendor-Specific Requirements" are met. Once the data is loaded, it becomes a part of that project so that the original data files can be manipulated or deleted without affecting the analysis in PEAKS.

To close an open project, select the project node in the "Project View" and perform any of the following: choose the Close Current Project command from the file menu, click the Close Project icon ![icon] from the tool bar, or right-click and choose the Close Project command from the pop-up menu. It is recommended to close unused projects to preserve computer memory.

# 3. Data Refinement - Setting Parameters

To run a data refinement on the fractions immediately after creating a project, click the Data Refinement button on the Project Wizard to open the "Data Refinement" panel. See Section 2, "Data Refinement Parameters" for more details on data refinement parameters.

Click the Finish button to exit the Project Wizard without specifying parameters for further analysis. Upon clicking the Finish button, PEAKS will create the project and run the data refinement on the fractions. See Chapter 7, *Data Refinement* to learn more about data refinement. Alternatively, click the Identification button to proceed and to set up parameters for a database search.

# 4. DB Search - Setting Parameters

To perform *de novo* sequencing and to run a database search on the project immediately after setting data refinement parameters, click the Identification button on the Project Wizard. See Section 2, "Setting PEAKS Search Parameters" for more details on database search parameters.

Click the Finish button to exit the Project Wizard without specifying parameters for further analysis. Upon clicking the Finish button, PEAKS will create the project, load the data, run a data refinement on the fractions, and then conduct *de novo* sequencing and database searching on the data at the project level. See Chapter 8, *Peptide De Novo Sequencing* and Chapter 9, *Peptide, PTM, and Mutation Identification (PEAKS DB, PEAKS PTM, and SPIDER)* for more details on *de novo* sequencing and database searching. Alternatively, click the Quantification button to set up parameters for quantification.

# 5. Quantification - Setting Parameters

To run quantification tools on the project immediately after setting database search parameters, click the Quantification button to open the "Quantification" panel. The optional PEAKS Q module of PEAKS Studio supports Label Free quantification, Reporter Ion quantification with isotope labels at MS/MS level, and Precursor Ion quantification with isotope labels at the MS-level. See Section 2, "Setting Parameters", Section 2, "Setting Parameters", and Section 2, "Setting Parameters" for more details on quantification parameters.

Upon clicking the Finish button, PEAKS will create the project, load the data, run the data refinement on the fractions, and then conduct *de novo* sequencing, database searching, and protein quantification on the data at the project level. See Chapter 14, *PEAKS Q - Label Free*, Chapter 13, *PEAKS Q - MS/MS Level* and Chapter 12, *PEAKS Q - MS Level* for more details on the quantification tools available in PEAKS.

# 6. Adding Data to an Existing Project

1. To open a saved project, select the File | Open Project command, the File | Open Recent Project list, or click

   the ⬜ icon from the tool bar.

2. To add data to an open project, click on the project from the "Project View" panel and select the Add Data

   command from the File menu or use the Add Data icon 🗋 from the tool bar. The "Project Wizard" window will open.
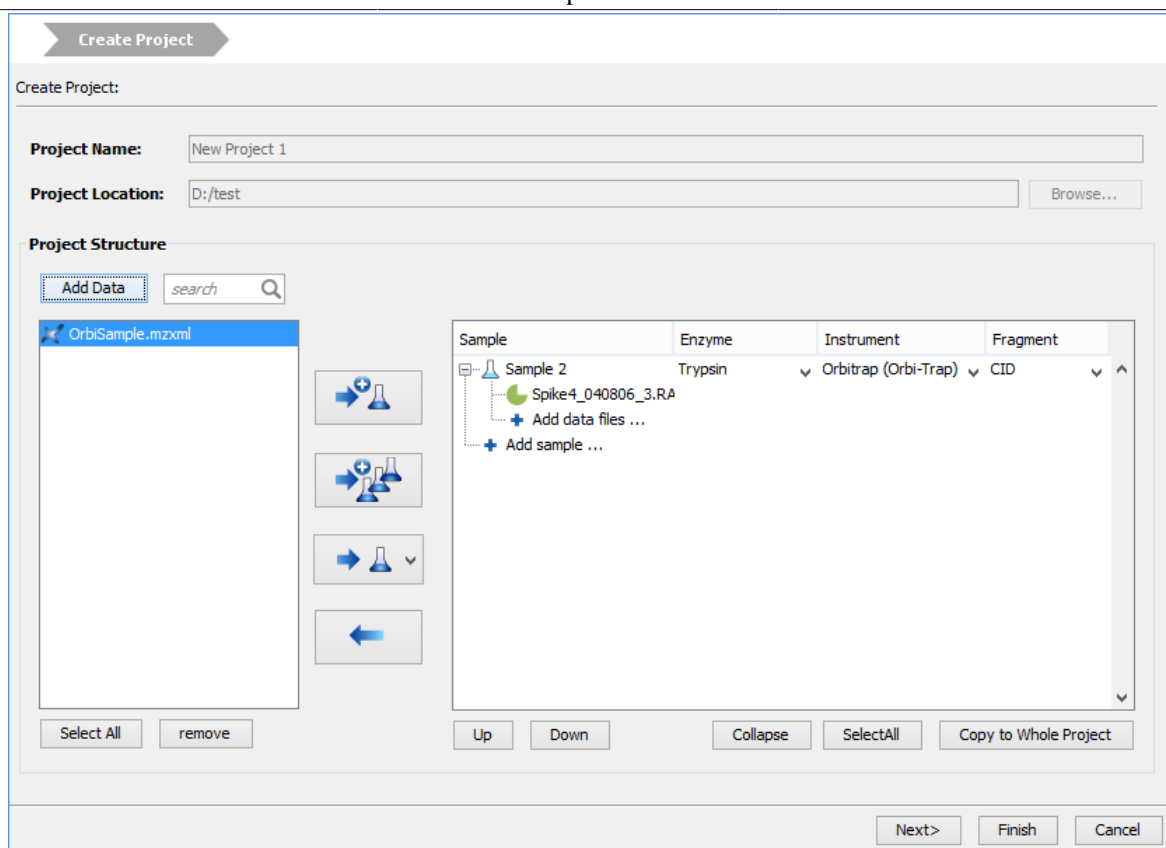
3. Use the Add Data icon to browse to the location of the data files. Select the files to be loaded and then click Open. All of the selected data files will be listed on the left side.

4. The selected data from the list on the left can be placed into an appropriate sample within the project structure. There are several options to enable this placement. All of the selected files can be placed in a new sample (  ), in an existing sample (  ), or in individual samples for each file (  ).

5. More files can be added to an existing sample by clicking the + Add data files ... button. Additional samples can be created by clicking the + Add Sample ... button.

6. From the drop-down menus, select the "Instrument" type, the "Fragmentation" method, and the "Enzyme" that was used to generate the experimental data. For more information on adding files/samples or setting up the instrument configuration refer to Section 2, "Create New Project".

### Note

It is discouraged to change the "Enzyme", the "Instrument", or the "Fragmentation" method for an existing sample. It may cause errors during further analysis or inconsistencies in the analysis results.

# 7. Supported Data Formats

The following is a list of supported data formats in PEAKS. PEAKS supports these formats at three different levels:

- Native Support - PEAKS can read the following files directly without any additional tools:

  - mzXML

  - mzData

  - mzML

- .DTA file or a directory of .DTA files

- .MGF

- .PKL

- PEAKS 7 projects. See Section 8.8, "PEAKS 7 Projects".

- Library Level Support - The instrument vendor's software library must be installed on the same computer as PEAKS. PEAKS will call the software library to read the data directly.

  - .RAW file (Thermo Fisher Scientific instruments). See Section 8.1, "Thermo Data".

  - .D directory (Agilent instruments). See Section 8.3, "Agilent Data".

  - .LIFT or .D directory (Bruker instruments). See Section 8.5, "Bruker Data".

  - .RAW directory (Waters QTOF instruments). See Section 8.2, "Waters/Micromass (MassLynx) Data".

- Converter Level Support - Third party converters are required. Users need to install the required converters correctly and and configure their locations in PEAKS. This only needs to be set up once. PEAKS will call the converter to convert the data to a native supported format before loading it. The actual conversion process is invisible to the user.

  - .RAW directory (Waters QTOF instruments). See Section 8.2, "Waters/Micromass (MassLynx) Data".

  - .WIFF file (AB/Sciex QSTAR and QTRAP instruments). See Section 8.4.1, "TrippleTOF, QSTAR, or QTRAP".

  - .T2D file (AB 4700/4800 series). See Section 8.4.2, "ABI 4700/4800".

  - .RUN folders from Shimadzu instruments. See Section 8.6, "Shimadzu Data".

  - .XMS files from Varian instruments. See Section 8.7, "Varian".

# 8. Vendor-Specific Requirements

Most vendors provide tools for MS analysis software to read their raw data format. PEAKS works best with the unprocessed raw data as PEAKS uses the data pre-processing tools built into the software that are designed to maximize identification results. Listed below are the requirements to load raw data from each supported vendor.

## 8.1. Thermo Data

RAW data from Thermo Fisher Scientific mass spectrometers can be loaded, provided that the XCalibur software or the Thermo MSFileReader package is installed on the same computer as PEAKS. The PEAKS Configuration Wizard can download and install MSFileReader automatically (see Section 2, "Instrument Selection"). MSFileReader is publicly available and can be found at the following link:

```
https://thermo.flexnetoperations.com/control/thmo/download?element=6306677
```

### Note

Install 64-bit MSFileReader with 64-bit version of PEAKS Studio, or 32-bit MSFileReader with the 32-bit version of PEAKS Studio.

## 8.2. Waters/Micromass (MassLynx) Data

Two methods can be used to load Waters's `.raw` data.

- `PEAKS Built-in Loader`

PEAKS is installed with the libraries needed to read Waters.raw files automatically. So, .raw files should be loaded into PEAKS without any configuration.

- ProteoWizard - MSConvert

Alternatively, select this option if the Built-in Loader is unable to load some .raw files as an alternative. Browse to the location of msconvert.exe, which is located by default in:

C:\Program Files\ProteoWizard\ProteoWizard\msconvert.exe

ProteoWizard can be downloaded form:

http://proteowizard.sourceforge.net/
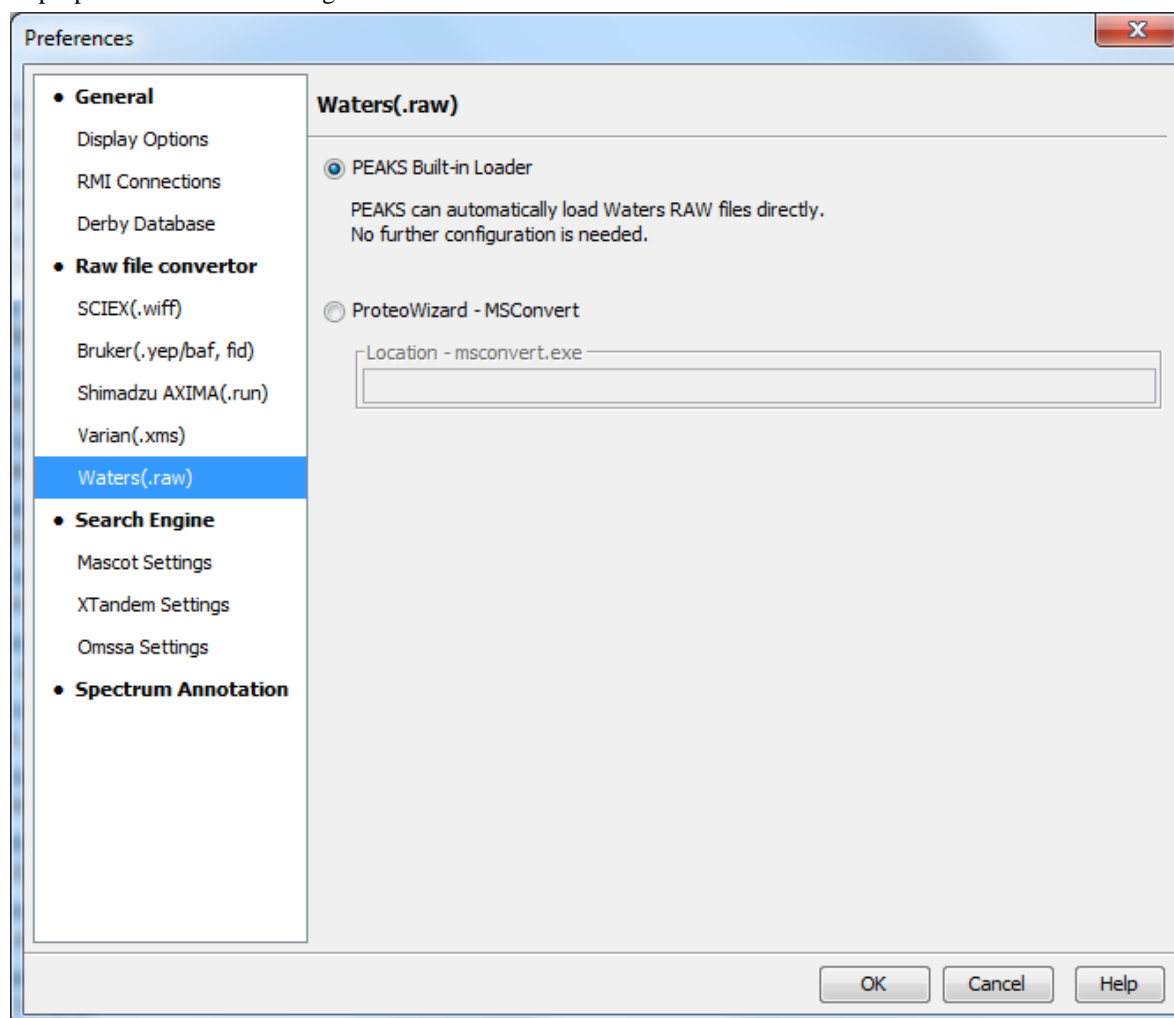


## 8.3. Agilent Data

**Agilent Q-TOF data** can be loaded, provided that you selected the Agilent checkbox when loading the software.
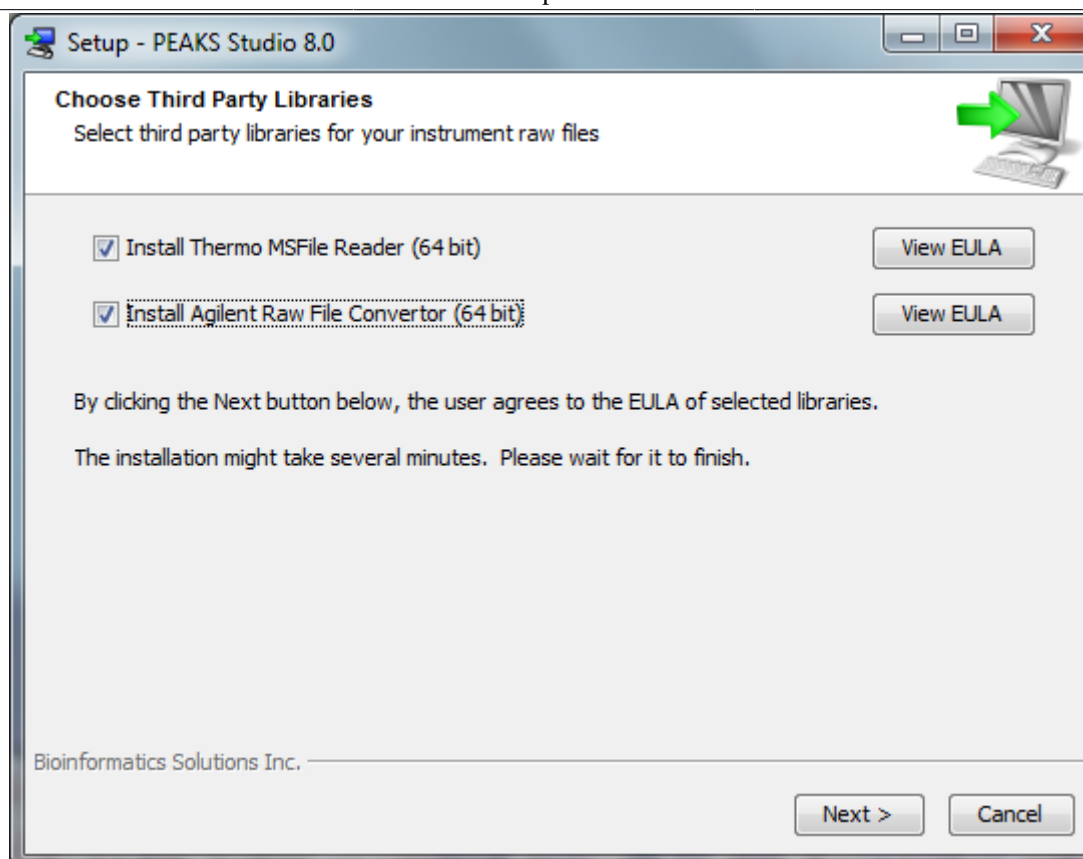
If you did not select that checkbox, it is still possible to install the libraries without reinstalling PEAKS.

1. Locate the PEAKS Studio installation directory. (default: c:\PEAKSStudio8)

2. Access the folder, 'MHDAC_x86 or 'MHDAC_x64', depending on your version of Windows

3. Right-click on the Windows batch file 'RegisterMassHunterDataAccess' and click on "Run as administrator."

4. If this process was successful, Agilent data should load into PEAKS automatically.

**Agilent Ion Trap data** can be loaded, provided that CompassXtract is installed on the same computer as PEAKS. The spectral data will be contained in the `.yep`, `.baf`, or `fid` file. Select the `.D` folder (containing `.yep` and `.baf` file) and the `.LIFT` folder (containing `fid` file) during the data loading step. To load all fid files at once, select the top level folder to load them.

CompassXtract 3.1 is readily available on the Bruker Daltonics web site at the following link: http://www.bruker.com/service/support-upgrades/software-downloads/mass-spectrometry.html. Once you reach the website, Bruker may request a username and password. See Section 8.5, "Bruker Data".

## 8.4. Sciex Data

### 8.4.1. TrippleTOF, QSTAR, or QTRAP

Sciex data can be loaded into PEAKS, provided one of the two options is installed.

• AB SCIEX MS Data Converter can be used to load .wiff files into PEAKS directly. It can be downloaded from the following link, http://www.absciex.com/downloads/software-downloads. Once installed, use the browse button within the Sciex Data Converter option to find AB_SCIEX_MS_Converter.exe. By default, it is installed in C:\Program Files (x86)\AB SCIEX\MS Data Converter\AB_SCIEX_MS_Converter.exe

• ProteoWizard - MSConvert is the recommended method for loading .wiff files. It can be downloaded from http://proteowizard.sourceforge.net/. Once installed, select the ProteoWizard -MSConvert button and use the associated browse button to fine msconvert.exe. By default it is located in:
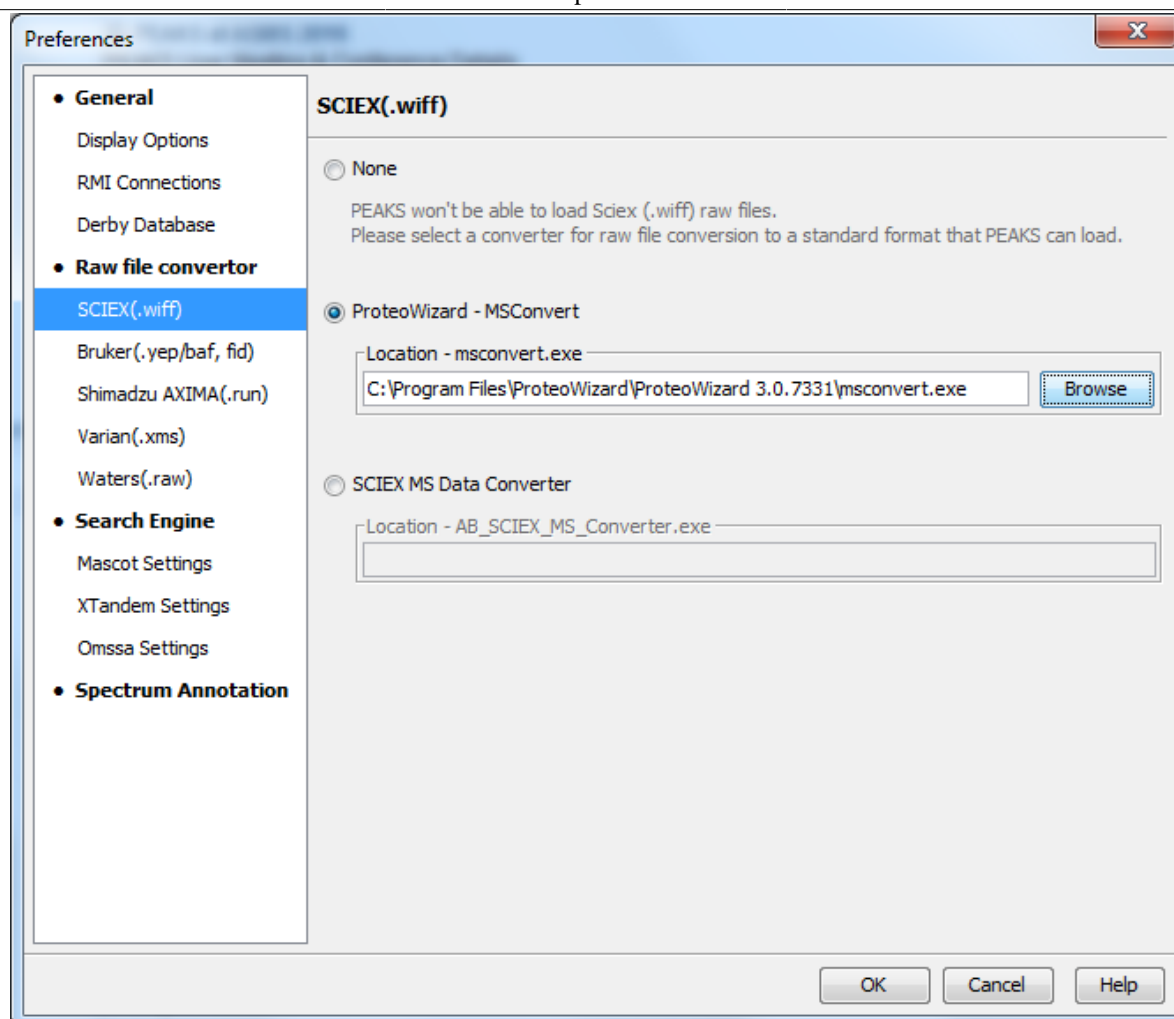
## 8.4.2. ABI 4700/4800

T2D files can be extracted and imported into PEAKS with a free tool created by BSI. The PEAKS Configuration Wizard can download and install the AB 4X00 Extractor automatically (see Section 2, "Instrument Selection").

**System Requirements.**    This extractor can be installed on the same machine as the ABI 4700 Explorer and the Oracle database or on another machine that has direct network access to the 4700 SERVER. There cannot be a firewall or proxy between the computers. Windows 7 or Windows XP is recommended for the use of this tool.

**Configuration.**    Start the ABI 4700 converter tool. Choose Settings from the File menu. Configuration requires the following:

- **4700 SERVER Name or IP Address:** input "localhost" if the Extractor is running on the same computer as ABI 4700 Explorer (this is the default value). Otherwise, enter the IP address of the 4700 SERVER.

- **The socket used by the 4700 SERVER:** the port that the Oracle database listens to (the default is 1521).

- **Username to access the Oracle database:** most likely does not need to changed (the default is "tsquared").

- **Password to access the Oracle database:** most likely does not need to changed.

**Data Extraction Procedure.**    The data extraction requires

1. **Load Spot Set List from the database:** (Do this via menu File | Load Spot Set List). The extractor will export the peak list of a spot set into a PKL file.

2. **Open a Spot Set:** (menu File | Open Spot Set) Spot Set Chooser will help the user choose a spot set. After selecting a spot set, click OK to open it. The job run information of the spot set will be shown.

3. **Select a job to run:** There is a button to select before each run. Only the MS/MS job run can be selected for export, as the precursor information is needed. Select a job run and click Convert to do the extraction.

4. **Choose a filename to save:** After clicking the Convert button, the user needs to input a file name and the peak lists of the selected job run that will be exported.

## 8.5. Bruker Data

`.D` and `.LIFT` directories from Bruker mass spectrometers can be imported provided that the CompassXtract Runtime library is installed on the same computer as PEAKS. The spectral data will be contained in the `.yep`, `.baf`, or `fid` file. Select the `.D` folder (containing `.yep` and `.baf` file) and the `.LIFT` folder (containing `fid` file) during the data loading step. To load all fid files at once, select a top level folder.

CompassXtract 3.1 is readily available on the Bruker Daltonics website at the following link: http://www.bruker.com/service/support-upgrades/software-downloads/mass-spectrometry.html. Once you reach the website, Bruker may request a username and a password

### Note

Install the 64-bit version if you are running PEAKS on a 64-bit operating system. Install the 32-bit version if you are running PEAKS on a 32-bit operating system.

**Instrument Preferences for Bruker Data.** To set Bruker data related preferences in PEAKS, open the "Preferences" window by clicking the "Preferences" toolbar icon  or by selecting Preferences from the Window menu. Click on "Instrument" and then "Bruker (.yep/baf, fid)" on the left side. This will show the Bruker instrument preferences on the right side. Select the PEAKS Built-in Loader to use CompassXtract to load Bruker files.

• Bruker (.fid) data may contain several files. Do you want to merge them? When loading .fid, select 'yes' to merge all the spectra contained in the file into one data file.

Alternatively, ProteoWizard - MSConvert can be used instead of the built-in data loader. It can be downloaded from http://proteowizard.sourceforge.net/. Once installed, select the ProteoWizard -MSConvert button and use the associated browse button to find msconvert.exe. By default, it is located in:
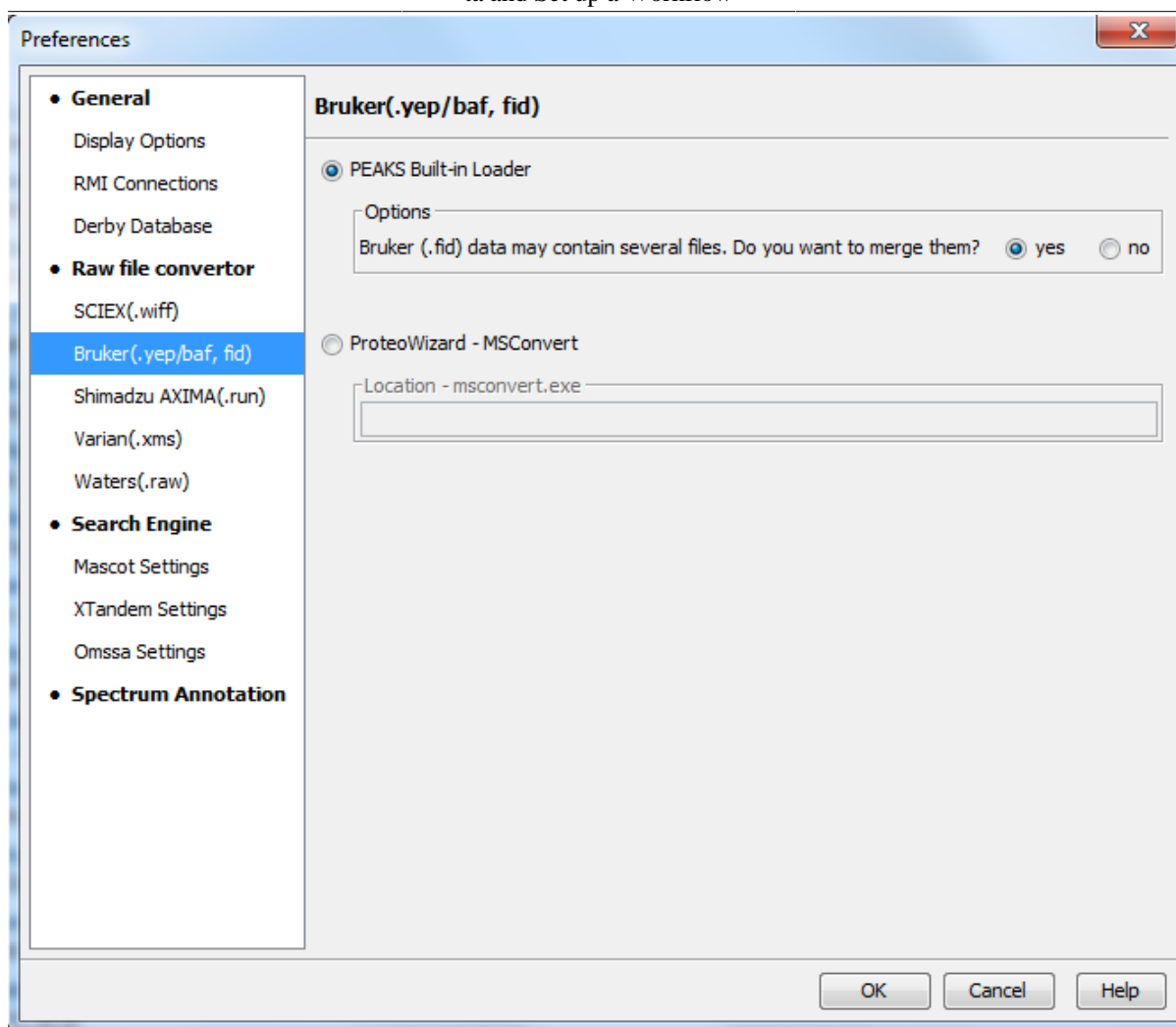
## 8.6. Shimadzu Data

RUN files from Shimadzu mass spectrometers can be loaded provided that the Shimadzu software is installed on the same computer as PEAKS.

**Instrument Preferences for Shimadzu Data.** To set Shimadzu data related preferences in PEAKS, open the "Preferences" window by clicking the "Preferences" toolbar icon or by selecting Preferences from the Window menu. Click "Instrument" and then "Shimadzu AXIMA (.run)" in the menu on the left. This will show the Shimadzu instrument preferences on the right side. Select the Shimadzu Converter button and click Browse to tell PEAKS the location of the Shimadzu `run2xml.exe` file.

Select None to disallow Shimadzu data loading
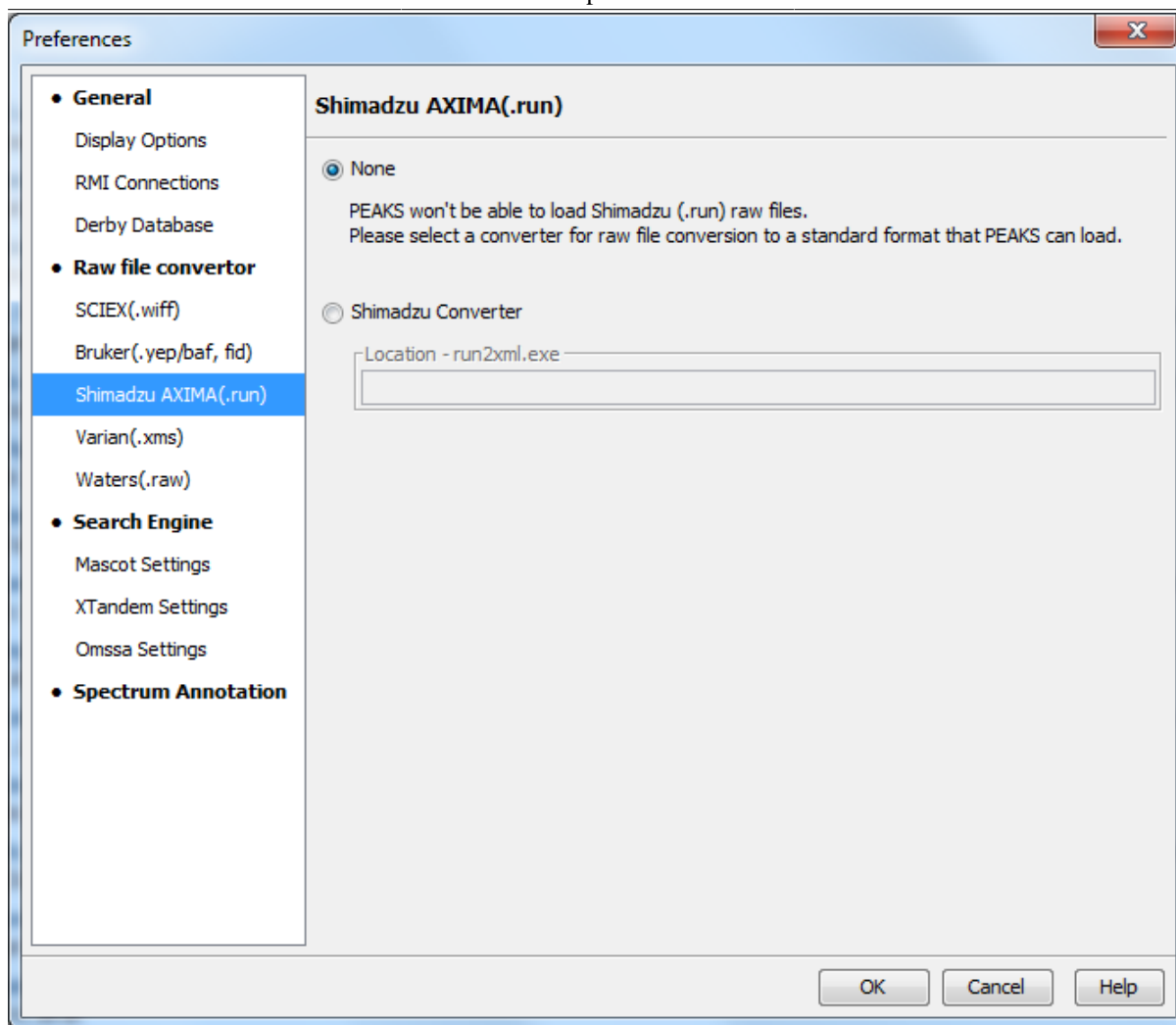
## Note

Shimadzu `run2xml.exe` can not read data files containing spaces or non-English characters in its path.

# 8.7. Varian

A conversion tool is embedded into Varian's data acquisition software that allows for conversion of Varian raw data into `.pkl` files that can be immediately read by PEAKS.

The `.trans` data files can be converted in Varian programs by clicking File | Save As and selecting either the `.pkl` file format or by performing the following steps: clicking File, right-clicking Export, and then selecting .pkl. If you are viewing a chromatogram with the Varian software, all the spectral data in the viewed chromatogram are converted to `.pkl` format. If you are viewing a single spectrum and choose to convert the data, only the viewed spectra will be converted.

Importing raw data that has not been preprocessed will produce better results when using the preprocessing options native to PEAKS.

**Instrument Preferences for Varian Data.** To set Varian data-related preferences in PEAKS, open the "Preferences" window by clicking the "Preferences" toolbar icon or by selecting Preferences from the Window menu. Click "Instrument" and then "Varian (.xms)" on the left. This will display the Varian instrument preferences on the right side. Click Browse to tell PEAKS the location of the `xmlrai.exe` file.

**Varian(.xms)**

Default xmlrai.exe Location

Browse

### Note

Varian `xmlrai.exe` can not read data files containing spaces or non-English characters in its path.

## 8.8. PEAKS 7 Projects

Projects created in PEAKS 7 and PEAKS 7.5 software can be opened in PEAKS 8. To convert the project to a PEAKS 8 project, open the project in the same way you would open any existing PEAKS project. The project will be recognized as a PEAKS project from an older version. The following confirmation dialogue box will pop up.

**Version Check**

⚠ This project was created by an older version of PEAKS. Do you want to convert it now?

Yes    No, thanks
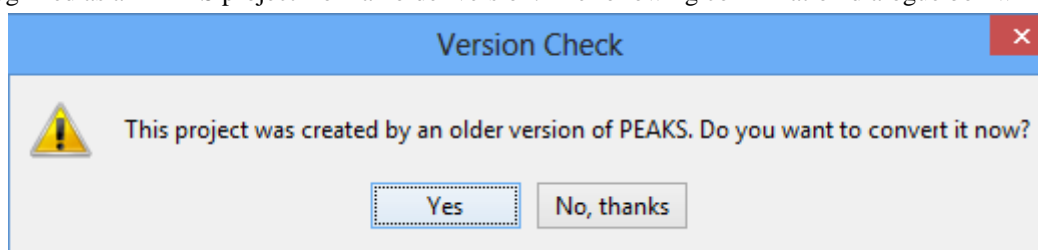
Choose Yes to convert the project and proceed. The following "Project Converter" dialogue will appear.

**Project Convertor**

Converted Projects    Project_version8_0

Project Location    D:\PeaksProjects    Browse
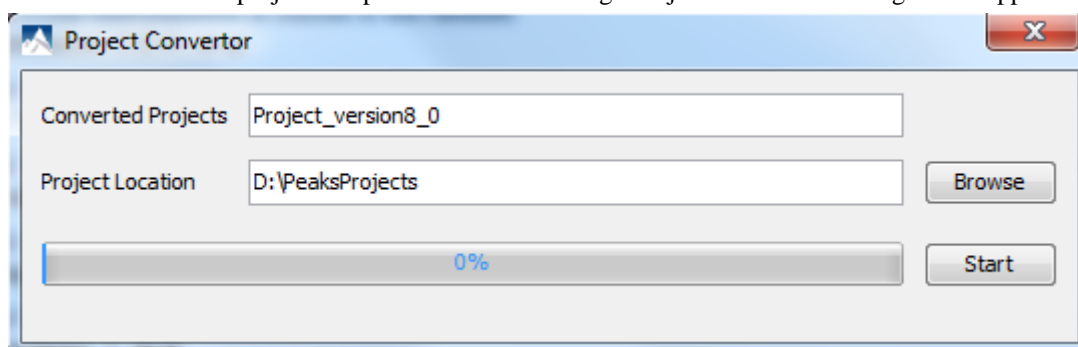
0%    Start

Choose the converted project name and location. Click Start to begin the conversion process. A new version of the project will be created at the new location. The old project will not be modified.

### Note

The project converter is designed to convert PEAKS 7 projects so that they can be viewed in PEAKS 8, but further analysis on the converted projects are strongly discouraged.

# 9. Changing the Default Project Location

If many projects are to be created, it is convenient to reconfigure the default project location to the directory where all the projects are stored. Please make sure this folder is readable/writable by PEAKS.

1.
   Click 🖉 from the toolbar. The following "Preference" dialogue pops up.
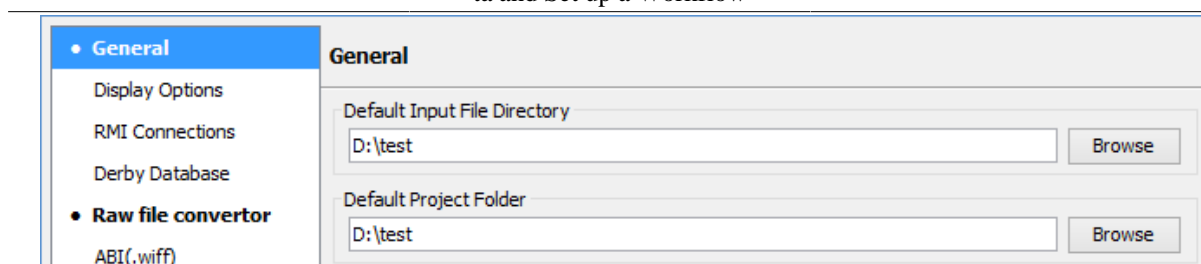
2. Select "General" in the in the "Preference" dialogue and click the Browse button below "Default Project Folder" to specify the default project location.

# Chapter 5. Data Visualization

## 1. Overview

After the project is created, the spectral data can be visually examined. For a typical LC-MS/MS fraction, three views are provided:

- **MS -** This view shows the TIC (total ion chromatogram) plot and all of the MS scans. For each MS scan, the corresponding MS/MS scans are also displayed.

- **MS/MS -** This view lists all the MS/MS scans. For each MS/MS scan, the corresponding MS scan is also displayed.

- **LC/MS -** This provides a bird's eye view of the whole LC-MS dataset with detected features and MS/MS scans marked on refined data.

After opening a data file by double-clicking the data node on the project tree, the choice of different views can be made by choosing different tabs at the upper-left corner of the data view window.
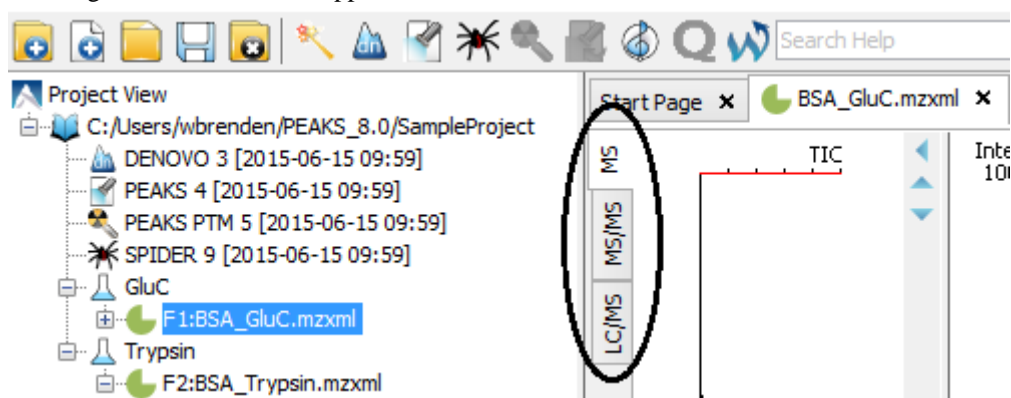


## 2. MS View

The MS View contains the TIC and all of the scans. The total ion chromatogram (TIC) is displayed on the left of the MS view. The navigation buttons are circled in the figure. To collapse the TIC chart, click the left navigation button. To navigate the survey (MS) scans, use the up and down navigation buttons. The survey scans can also be navigated by using the up and down arrows of the keyboard. Clicking on a specific position in TIC will display the corresponding survey scan. The tandem (MS2) scans associated with the current survey scan are shown in the bottom right panel.
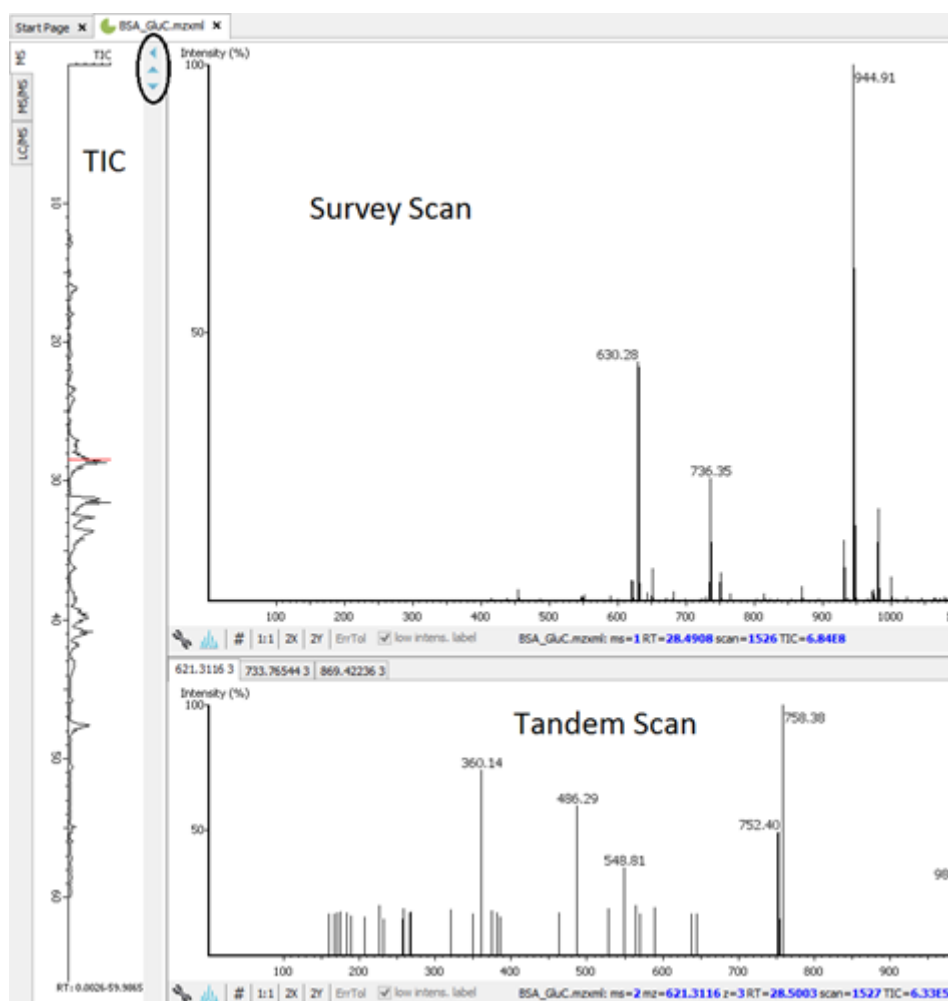
The survey scans and tandem scans provide a few convenient ways to zoom and navigate the spectrum:

- **Zoom to an m/z region -** Click the desired start m/z and drag horizontally to the desired end m/z, release the mouse button.

- **Zoom in/out smoothly -** Place the mouse pointer at a particular m/z value (right below the x-axis line), scroll the mouse wheel button.

- **Increase/decrease the peak intensity -** Place the mouse pointer on the spectrum, scroll the mouse wheel button.

- **See the whole spectrum -** Double-click on the spectrum or click the 1:1 button.

# 3. MS/MS View

The MS/MS View shows the list of tandem scans on the left. For each MS/MS scan, the list of identification results, the spectrum, the LC/MS view zoomed in to the area pointing to the Spectrum, the Ion Match Table, and its survey scans are shown on the right. Zoom options are the same as described in the "MS View" section.
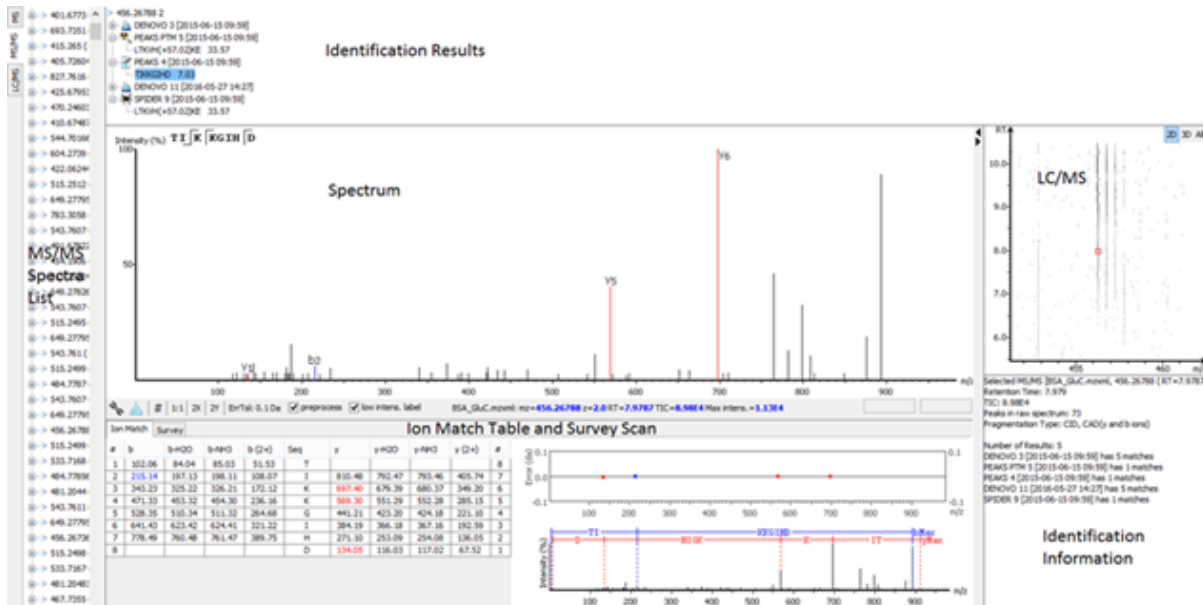
# 4. LC/MS View

The LC/MS view shows the distribution of LC-MS signals/features as a heat map. Placing the cursor on the heat map will show the m/z value and the retention time of that point.
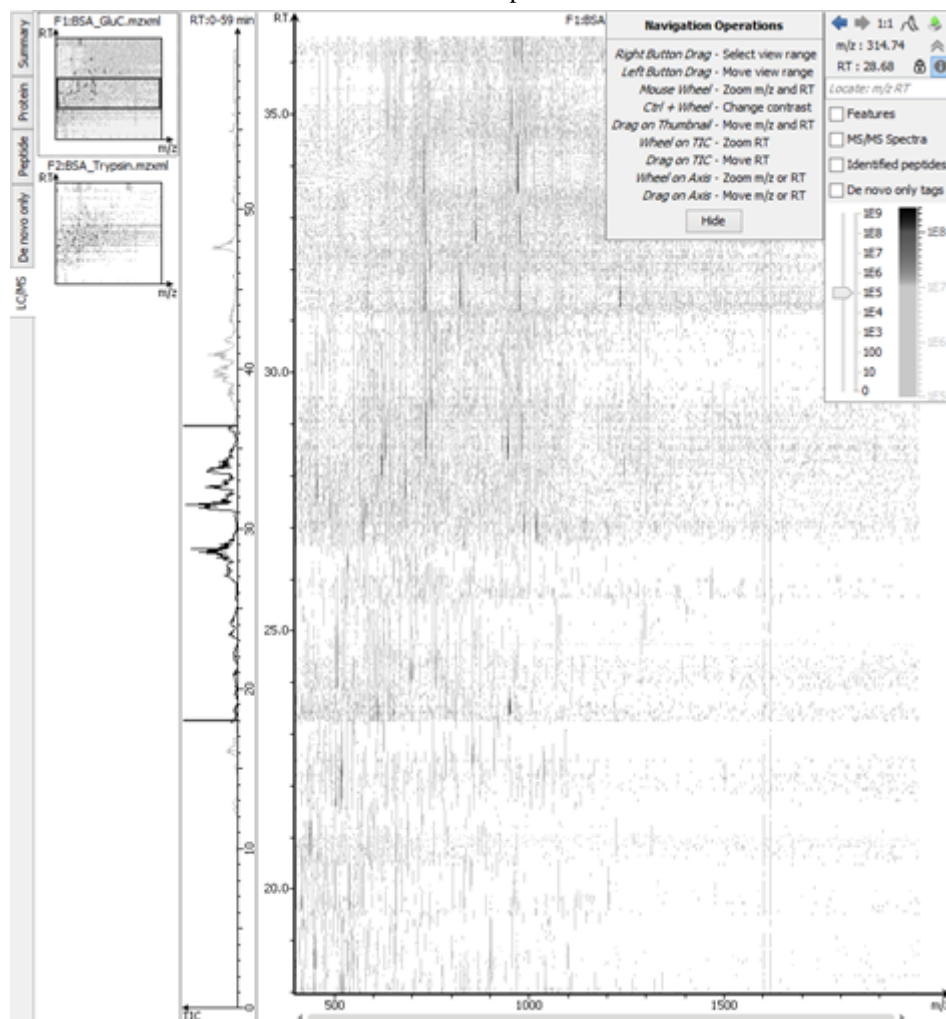
The LC/MS view provides various convenient ways to zoom and navigate the LC-MS features in the data:

- **Move to a specific Heat Map area:** Click on the heat map, hold and drag the cursor until the desired area of the heat map is reached and then release the mouse button.

- **Zoom to a specific Heat Map area:** Right-click on the heat map, hold and drag the cursor to select the desired area of the heat map and then release the mouse button.

- **Move to a specific Heat Map area from Thumbnail:** Current view range of the heat map is shown as a rectangle in the Thumbnail. Click on the thumbnail view range rectangle, hold and drag the cursor until the desired area of the heat map is reached and then release the mouse button.

- **Jump to a specific Heat Map area:** Provide approximate m/z and RT values separated by a space in the Locate: m/z RT box and then hit the Enter key to zoom in to an area centered at the specified m/z and RT value.

- **Zoom in/out from Thumbnail:** Place the cursor on the Thumbnail and then scroll the mouse wheel button.

- **Zoom in/out smoothly:** Place the cursor at a particular (m/z value, RT) position and then scroll the mouse wheel button.

- **Move to a specific RT range:** Click the cursor on the TIC or the RT axis, and then hold and drag the cursor until the desired range is reached.

- **Zoom to a specific RT range:** Place the cursor on the TIC or the RT axis, and then scroll the mouse wheel button.

- **Move to a specific m/z range:** When the m/z axis is zoomed in, click the mouse on the m/z axis, and then hold and drag the cursor until the desired range is reached.

- **Zoom to a specific m/z range:** Place the cursor on m/z axis and then scroll the mouse wheel button.

- **Change contrast of the Heat Map:** Place the cursor on the heat map, press and hold the Ctrl key, and then scroll the mouse wheel button.

- **Change peak height in the 3D intensity mode:** Place the cursor on the heat map, press and hold the Ctrl key, and then scroll the mouse wheel button.

- **View the whole Heat Map:** Click the 1:1 button in the control.

- **Move Back/Forward to previously specified Heat Map area:** Use the Back and Forward buttons in the control to navigate the previously specified heat map areas.

The Heat Map view offers various controls to study the LC-MS data features. The following sections describe the control options and their uses in detail.

## 4.1. Noise Level

Select the appropriate threshold for noise filtering. Once selected, the Heat Map view will reflect the changes.
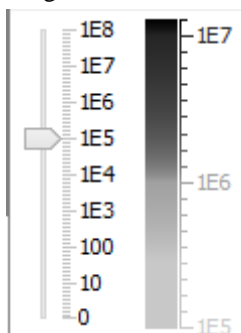
## 4.2. Show/Hide Features

To view the detected LC/MS features, mark the Features checkbox in the control. The features detected during the data refinement steps will be highlighted by small, red, diamond-shaped points on the heat map. Placing the cursor over such a feature point will show the boundary of the feature using a red polygon and display a pop-up window that includes the following information: the monoisotopic m/z, the mean RT, the charge state of the feature, and the total area under the XIC curves of the feature.



```
m/z=656.3986
RT mean=49.51
z=2
Area=8.09E6
```
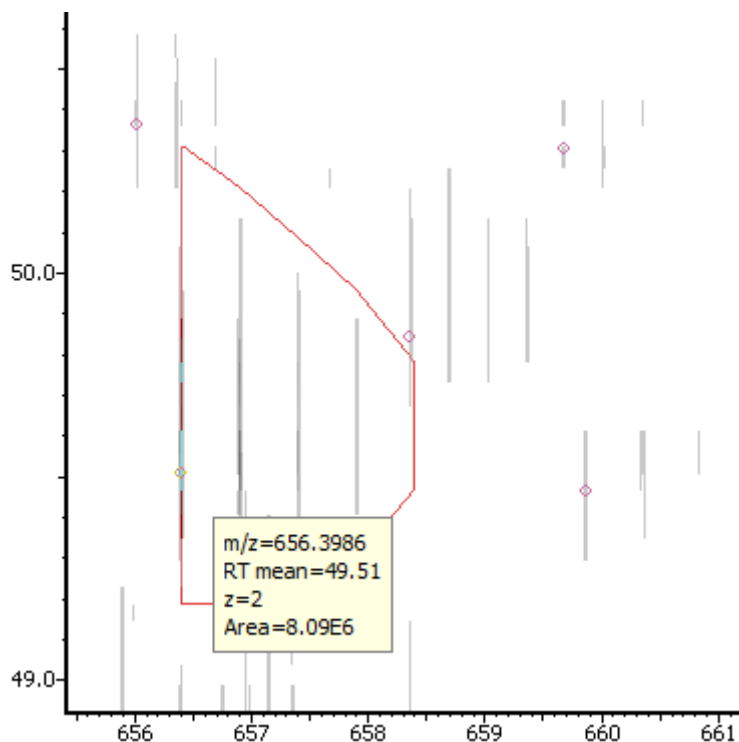
## 4.3. Locate M/Z and Retention Time

Type an m/z and an RT value, such as "618.88 50.5" in the following text box *Locate: m/z RT* to locate a (m/z, RT) position in the LC/MS view.

## 4.4. Synchronize LC/MS View between LC Runs

Use the 🔒 button to synchronize LC/MS view between different LC/MS runs. When the LC/MS views are synchronized, the (m/z, RT) position and noise level are locked as the same in each run.

## 4.5. Show/Hide MS/MS Spectra

Select the MS/MS Spectra checkbox to highlight the scans with associated tandem scans. They will be marked with tiny blue squares on the heat map. Place the cursor over an MS/MS spectra indicator to view a pop-up window that will contain the precursor m/z, the RT, the precursor charge state, and the scan number of that spectrum. Click an MS/MS spectra indicator to view the spectrum in a pop-up window. The indicator for that selected spectra will turn into a small red square.
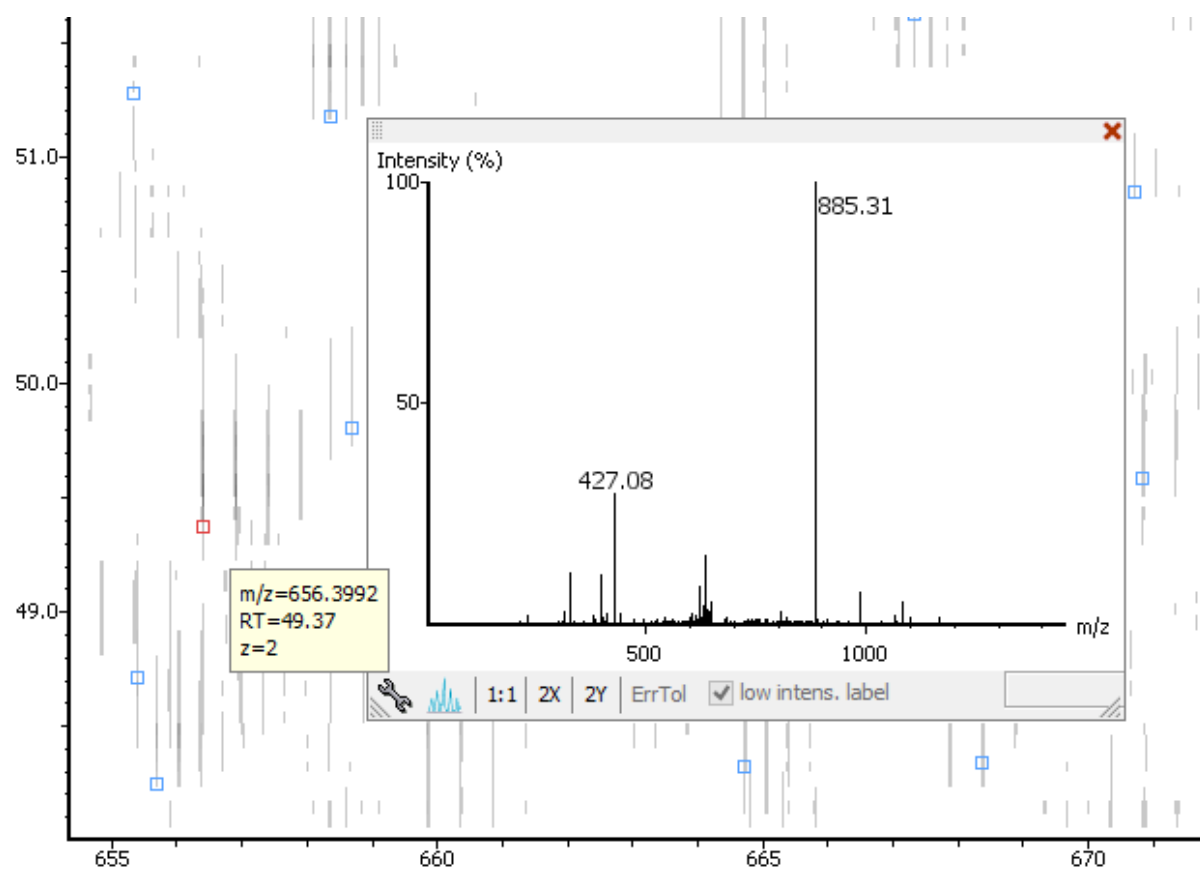
## 4.6. Intensity View

In addition to the heat map view, PEAKS also provides a 3D intensity view for visual inspection of the LC/MS data. To switch between the heat map view and the intensity view, click the  button in the top-right LC/MS control options. The intensity view has the same set of controls to easily zoom and navigate.
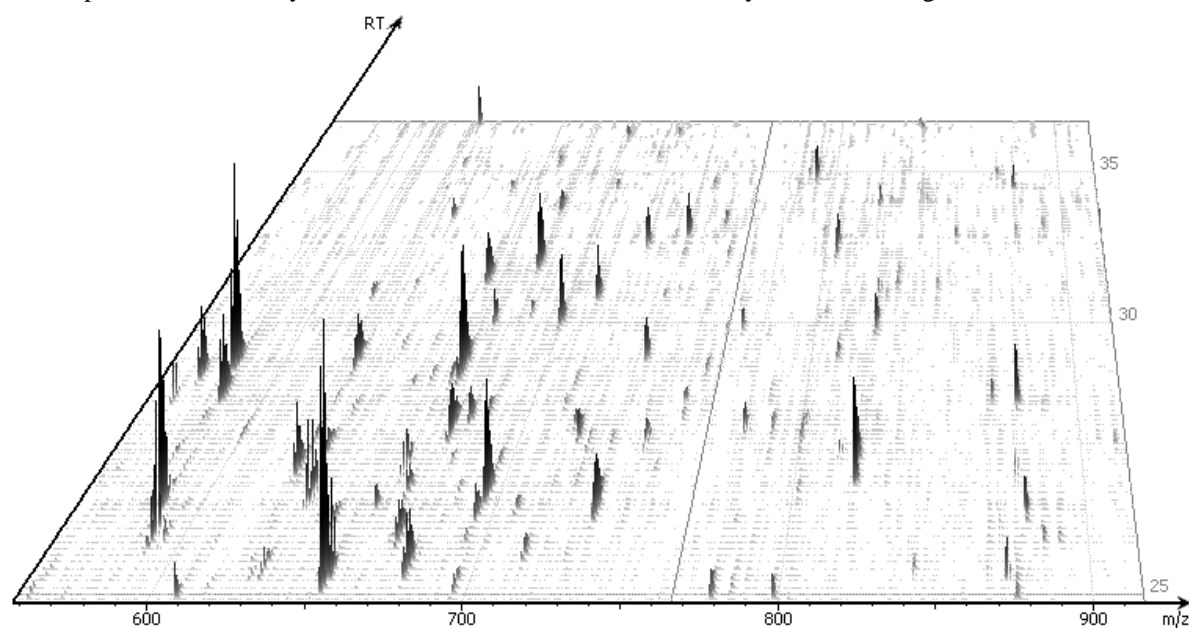
## 4.7. Export LC/MS

Click the ![icon] icon on the LC/MS control options to export the LC/MS heat map or intensity view in PNG format.
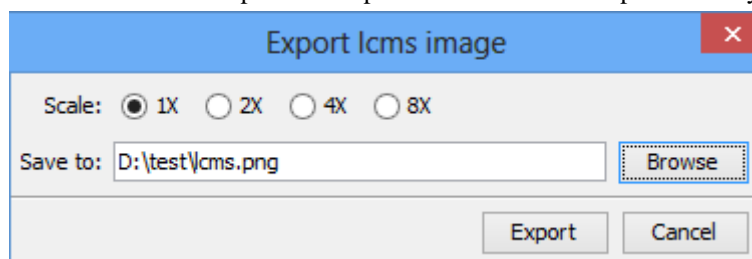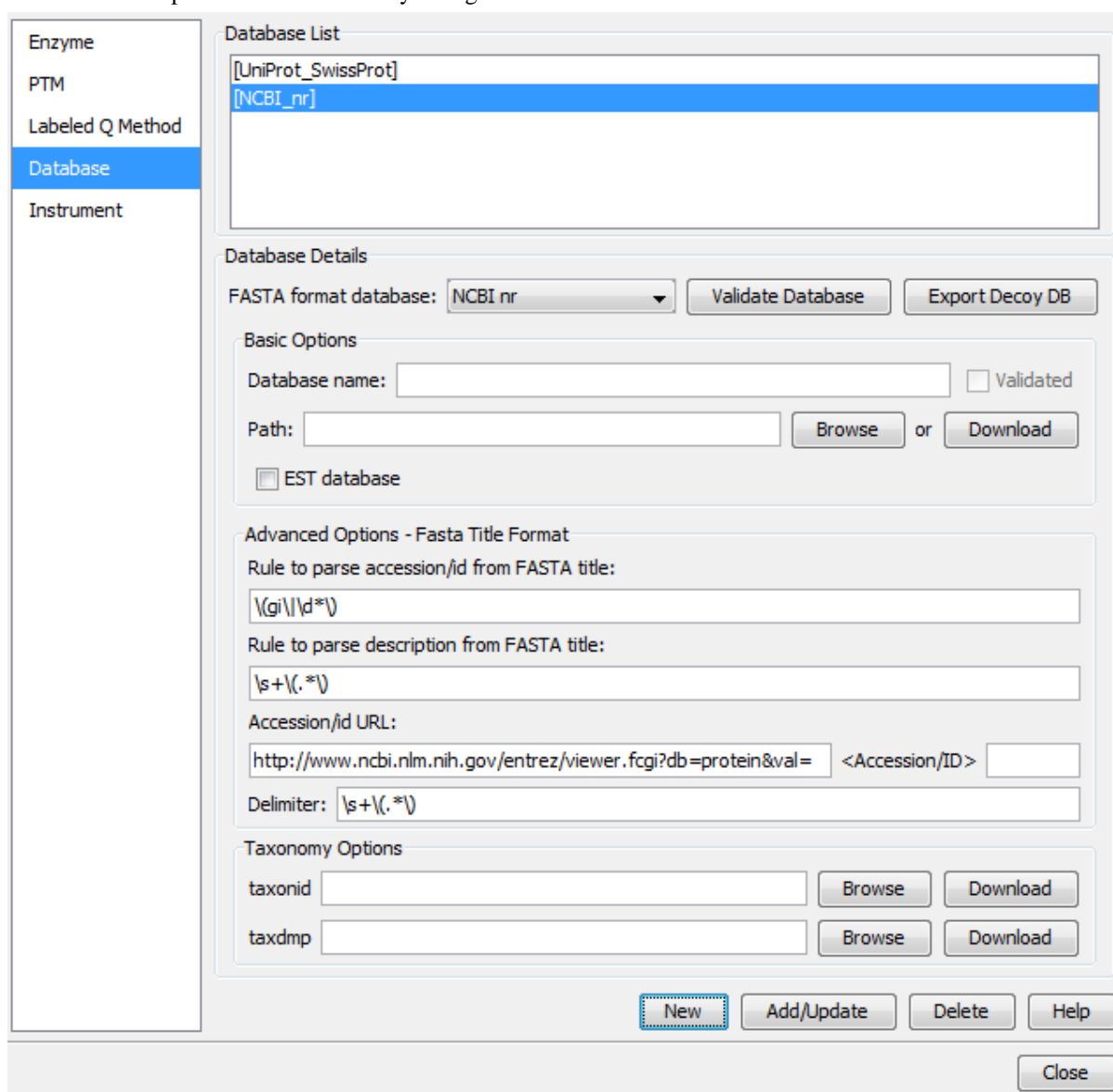
# Chapter 6. Adding a Sequence Database

## 1. Configuring Sequence Databases

PEAKS has the ability to search through a database to identify proteins using the PEAKS DB function. In order to use this function, PEAKS must have access to either a protein or an EST database in FASTA format (the standard format for popular public sequence databases). PEAKS can be configured to use existing databases on the system or from databases downloaded from public servers. Additionally, taxonomy options may be specified for certain databases.

To add a sequence database, click the ⚒ icon in the main toolbar and select "Database" from the left side. This will open the database configuration dialogue. Click the New button to create a new sequence database entry or select a database from the "Database List". The database can be configured in the "Database Details" area below. Click the Add/Update button to save any changes.



Follow these steps to configure a database:

1. Select the database format from the "FASTA Format Database" drop-down menu, or select "Other" if the desired format is not present. If "Other" is selected, you must enter custom parse rules.

2. In the "Basic Options" section, enter a name for the database. If the database FASTA file is already on the local system, skip to step 6. Otherwise, click Download. A window will appear confirming the database chosen to be downloaded along with the corresponding FTP or website.

3. Click OK to invoke the default FTP client software and download the database automatically. Click Cancel to copy the URL to the system clipboard. If Cancel was selected, click OK on the dialogue detailing the copy to the clipboard. Next, open a browser and paste the URL into the address bar. When the file download window opens, click Save.

4. Once the database has been downloaded, check to see if it is compressed. If so, extract the file using a program such as WinZip, or WinRar. The desired result is a FASTA format text file (`.fas` or `.fasta` file).

5. Move the database file into a directory that PEAKS can access.

6. Click Browse in the "Basic Options" to inform PEAKS of the location of the database file.

7. If the selected database is an EST database, check the box labeled "EST database". Otherwise, leave this unchecked.

8. Based on the selected format from the "FASTA Format Database" list in Step 2, the accession number information and parsing rules for the database headers are automatically entered in the textboxes in the "Advanced Options- Fasta Title Format" section below.

   If "Other" was selected in Step 2, enter the parsing parameters into the corresponding textboxes. Alternatively, if the database format is similar to one of the public databases such as *NCBI-nr*, the parsing rules can be filled by selecting a similar database from the drop-down list and edited to set the desired parsing rules.

9. Click the Add/Update button to save the configured database. The database name will appear in the "Database List".

   ## Note

   Apart from starting with a "greater than (>)" symbol, the precise syntax of the FASTA title line varies from database to database. For this reason, PEAKS uses Java Regular Expressions to define how the accession string and the description text should be parsed from the FASTA title line.

To be able to run PEAKS DB using a specific taxonomy, the corresponding files must be downloaded and then referenced by PEAKS in the "Taxonomy Options" section. (Taxonomy files for *NCBI nr* database are: **gi_taxid_prot.dmp.gz** and **taxdmp.zip**; for *UniProt/Swiss-Prot*, they are **speclist.txt** and **taxdmp.zip**.)

1. To download the "taxonid" file, click the Download button. A window will appear confirming the FTP or website URL that was identified as the location of the desired taxonomy file. To invoke the default FTP client software and download the file automatically, click OK. Click Cancel to copy the URL to the system clipboard. If Cancel was selected, click OK on the dialogue detailing the copy to the clipboard. Next, open a browser and paste the URL into the address bar. When the file download window opens, click Save. Use the same procedure to download the "taxdmp" file. Be sure to save the file to a location that is accessible by PEAKS. Please note that it is not necessary to decompress the taxonomy files.

2. Now that the taxonomy files are downloaded, PEAKS must be given access to them by clicking the Browse buttons and selecting the files.

   ## Note

   Please do not rename the taxonomy files; otherwise, PEAKS cannot recognize them.

# 2. Databases to be Used in PEAKS' inChorus Function

The database configured here will also be used in the PEAKS' inChorus function to call the X!Tandem and OMSSA search engines. However, Mascot search depends on Mascot's databases only. When using these third-party software tools, note the following with care:

- **X!Tandem** - At the time of writing, X!Tandem has difficulty searching through large databases and may crash during a search. Therefore, it's suggested that X!Tandem only be used with small databases. If used with a large database, a taxon should be specified. When using *NCBInr* or *SwissProt* databases with X!Tandem it is best to use a subtaxon.

- **OMSSA** - At the time of writing, OMSSA cannot be used with databases that are not in NCBI or SwissProt format in a way that is available to inChorus. Also, a bug in OMSSA prevents easy use of databases with OMSSA when they are stored in a folder that contains spaces or non-English characters in its path. This creates problems when PEAKS creates temporary databases on your behalf. To avoid this, best practices suggest that all databases are put in a folder named `C:/peaksdatabases`. Note that the folder `C:/My Documents/databases` wouldn't work as it contains a space between 'My' and 'Documents'. Using spaces in the database file name causes the same problem. Once databases have been downloaded and extracted, save the database file as `ncbinr.fas`, or `ncbi_nr.fas` rather than `ncbi nr.fas`.

- **Mascot** - The database used by Mascot has to be identical to the database configured in PEAKS in order for inChorus to parse Mascot results correctly.

# Chapter 7. Data Refinement

## 1. Overview

Raw LC-MS/MS data often contains noise, redundancy, as well as errors due to sample preparation and instrument approximation. The PEAKS Data Refinement tool can be used to improve the overall quality of the data. All or some of the following functions can be applied to the data in a project according to the user's requirements. Correct data refinement, especially the precursor m/z correction, can often result in significant improvement to in the final analysis result.

- **Merging scans :** MS/MS scans from the same precursor m/z and at a similar retention time will be merged together.

- **Precursor m/z correction:** The precursor m/z values given by some instruments are often not of the monoisotopic ion. This creates problems in downstream analysis. By examining the isotope shapes in the corresponding MS scans, this function can accurately adjust the precursor m/z to be equal to the monoisotopic ion m/z.

- **Precursor charge correction:** Occasionally the data provides wrong or no charge information for the precursor ions. This function attempts to correct the charge information.

- **Low quality spectrum removal:** This function attempts to remove the "junk" spectra. This will save some analysis time. Use this function with caution as it may also remove a small portion of identifiable spectra.

- **Centroiding, charge deconvolution, and isotope deconvolution:** Centroiding the peaks and deconvoluting multiply-charged ions to singly-charged ions in the MS/MS scans.

If the data is not refined within PEAKS, most analysis functions (such as *de novo* sequencing or PEAKS DB) will ask you to input the refinement parameters before the analysis is done. You can run the data refinement function separately by selecting a fraction, a sample, or a project on the project tree. All fraction(s) under the selected node will be refined. The use of this function is outlined in the following.

1. Select a project, a sample, or a fraction node. Click the Data Refinement icon ✎ on the tool bar.



2. Specify the data refinement parameters in the pop-up dialogue and click OK. Most of the parameters are self-explanatory and the default parameters provide a good starting point for the analysis.

3. Wait for the analysis to complete. A new Data Refinement node will appear in the project tree. Later analyses on this fraction will be based on the refined data.
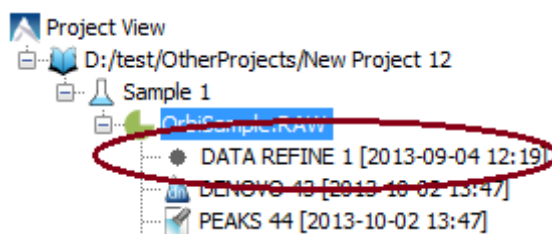


## 2. Data Refinement Parameters

After selecting a data node in the project tree, click the Data Refinement tool bar icon ✎ . The "Data Refinement" dialogue window will appear:

- **Merge Scans**

  **Retention Time Window:** The maximum retention time difference between two spectra to be merged.

  **Precursor m/z Error Tolerance:** The maximum difference in m/z between two spectra to be merged.

  **Merge CID and HCD scans together:** Scans generated from different fragmentation methods will not be merged. To enforce merging CID and HCD scans, please check this check box.

- **Correct Precursor**

  **Mass only:** Correct the precursor's mass.

  **Mass and Charge states:** Correct the precursor's mass and charge. It is suggested to use this option only when the precursor charge information is missing from the data files.

  **Min Charge:** The minimum charge to which a precursor ion can be corrected.

  **Max Charge:** The maximum charge to which a precursor ion can be corrected.

- **Filter Scans**

  **Precursor Mass Range:** The precursor mass region to select scans for further analysis.

  **Retention Time Range:** The retention time region, in minutes, to select scans for further analysis.

  **Quality Threshold:** The spectrum quality threshold to select scans for further analysis. The recommended percentage value is 0.65.

## Note

Data preprocessing (centroiding, deisotoping, deconvolution) is performed on all the data sets.

Once all parameters are set, press the OK button to initiate the data refinement process.

## 2.1. Saving the Parameters for Future Use

After setting up the desired parameters, you can save them for future use. Click the drop-down menu at the top-right of the window, select "Save as..." and type a name, when prompted, for these preferences for future use/reference when prompted. Any parameters that are saved will be available in the drop-down menu at the top of the window. To examine the contents, select a saved file and the parameter values will be automatically displayed.

# Chapter 8. Peptide *De Novo* Sequencing

## 1. Overview

*De novo* sequencing derives the peptide sequence from the tandem mass spectrum without the need of a sequence database. It is the preferred method for identifying novel peptides and the study of unsequenced organisms.

PEAKS provides a reliable and comprehensive solution for automated peptide *de novo* sequencing. It features:

- **Accurate de novo sequencing**

  The PEAKS *de novo* sequencing algorithm and scoring functions are constantly tuned for each instrument type to ensure optimal accuracy.

- **Fast sequencing speed**

  PEAKS can perform *de novo* sequencing on over 10 spectra per second on a moderate PC and even faster on servers.

- **Easy-to-use user interface**

  PEAKS generates a comprehensive result summary and provides interactive views of annotated spectra.

- **Support of all major instrument types**

  PEAKS supports *de novo* sequencing of spectra generated by all instrument types (Orbitrap/FTMS/Ion Trap/ Time-of-Flight/Quadrupole) in common fragmentation modes (CID/HCD/ETD). PEAKS can also perform *de novo* sequencing on complementary CID/HCD/ETD spectral pairs or triplets.

- **Automated result validation**

  PEAKS assigns a *local confidence* score for each amino acid in a *de novo* sequence. The local confidence score ranges from 0% to 99%, indicating how confident the algorithm considers a particular amino acid as correctly sequenced. Moreover, the peptide sequence is evaluated by ALC (Average of Local Confidence) score. ALC is the average of the local confidence score of all the amino acids in the sequence.

  ### Note

  For details about the PEAKS *de novo* sequencing algorithm, please refer to the initial publication: *"PEAKS: Powerful Software for Peptide De Novo Sequencing by Tandem Mass Spectrometry" Rapid Communication in Mass Spectrometry 17(20): 2337-2342 (2003).*

In addition to automated *de novo* sequencing, PEAKS also provides a manual *de novo* sequencing tool to assist in the manual interpretation of individual spectra.

### Note

Refer to Section 8, "Manual *De Novo* Sequencing"

Other PEAKS tools, including PEAKS DB for database search; PEAKS PTM for unspecified PTM search; and SPIDER for homology search depend on the *de novo* sequencing result to achieve superior sensitivity and accuracy.

### Note

Refer to Section 1, "Overview"

The use of automated *de novo* sequencing is outlined below. For detailed information, please refer to later sections.

1. Select a data node in the "Project View". The selected node can be a fraction, a sample, or a project. PEAKS will analyze all fractions under the selected node. Launch PEAKS *de novo* sequencing by clicking the "Auto De Novo" button in the toolbar.
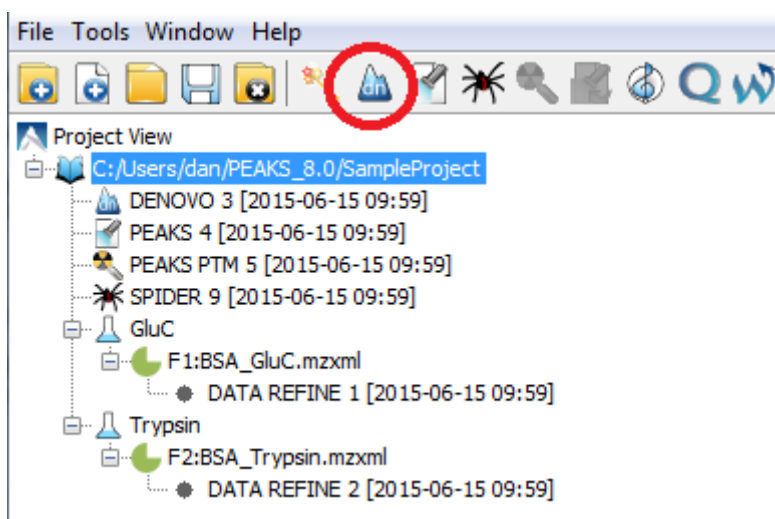


## Note

Refer to Chapter 4, *Project Wizard - Create a PEAKS Project, Load Data and Set up a Workflow* for how to create a project.

2. Specify the *de novo* sequencing parameters in the pop-up dialogue. Most of the parameters are self-explanatory and the default parameters provide a good starting point. Click OK to start automated *de novo* sequencing.

## Note

If data is not yet refined, you will be prompted to specify the data refinement parameters first. Refer to Chapter 7, *Data Refinement*.

3. Wait for the analysis to finish. A new result node will appear in "Project View". Double-click the node to examine the *de novo* sequencing result.



# 2. *De Novo* Sequencing Parameters

After selecting a data node in the "Project View", click the PEAKS *de novo* toolbar icon . The PEAKS *de novo* parameters dialogue will appear:

## 2.1. Error Tolerance

**Precursor mass tolerance :** PEAKS *de novo* generates peptides within the precursor mass tolerance. Precursor mass tolerance can be specified in either Daltons or ppm.

**Fragment ion m/z tolerance :** PEAKS *de novo* uses this value when scoring *de novo* peptide sequences. PEAKS considers a fragment ion to be matched if the calculated m/z is within the tolerance.

Error tolerance parameters need to be set consistently with the mass accuracy of the instrument. Default parameters for the instrument generating the data can be loaded by choosing "default" in "Predefined parameters".

### Note

When the selected data node is a project of multiple samples generated by different instruments, the default parameters of the first sample are loaded when choosing "default". It is recommended to set the tolerance parameters according to the least accurate instrument.

## 2.2. Enzyme

Select the enzyme used for protein digestion from a drop-down menu. PEAKS *de novo* respects the enzyme specificity at both ends when generating peptides.

### Note

When the selected data node is a project of multiple samples, "Specified by each sample" allows samples to be analyzed separately, using their respective specified enzymes during project creation.

### Note

"None" is a special enzyme allowing non-specific cleavage at both ends of the peptide. It is the recommended option when the digestion enzyme exhibits a high degree of non-specificity.

If the enzyme used for protein digestion is not in the list, a new enzyme can be added by clicking the New button. Refer to Section 2.1, "Enzyme Configuration" for details on creating new enzymes.

## 2.3. PTM

To select PTMs for *de novo* sequencing, click the Set PTM button to open the "PTM Options" dialogue.



PEAKS includes modifications in the Unimod database. Unimod modifications are categorized into three lists:

- **Common**

  The "Common" list contains commonly observed modifications, including naturally occurring PTMs, modifications induced in standard sample preparation, and common artifacts.

- **Uncommon**

  The "Uncommon" list contains less commonly observed modifications, including rare PTMs, residue substitutions, and artifacts.

- **Artificial**

  The "Artificial" list contains modifications that can only be artificially induced for a specific purpose, including various chemical and isotopic tags for quantification or other experiments.

In this dialogue, there are two other lists of modifications:

- **Recent**

  The "Recent" list keeps track of previously selected modifications. By default, modifications are sorted by the last time they were selected. You can conveniently select the modifications that are most frequently used for your analysis.

- **Customized**

  The "Customized" list shows all the user-defined modifications. User-defined modifications can be modified or deleted.

Modifications can be selected as either *fixed* or *variable* modifications:

- A **fixed modification** forces all instances of applicable residues to be modified. For example, when carbamidomethylation (C+57.02) is selected as a fixed modification, all cysteine residues in the result will be modified.

- A **variable modification** gives the option for the residues to be modified. If carbamidomethylation (C+57.02) is selected as a variable modification, only some of cysteine residues in the result will be modified.

## Note

For a specific residue, only one fixed modification is allowed. If more than one fixed modification is selected for the same residue, PEAKS will warn about conflicting fixed modifications. The warning message can be ignored, and only the first fixed modification on that residue will be used in the program.

## Note

The use of variable modifications increases the size of the computational search space for the *de novo* sequencing algorithm. It is recommend not to select too many variable modifications in PEAKS *de novo*.

## Note

Leucine (*L*) and Isoleucine (*I*) cannot be directly distinguished in mass spectrometry since the two residues have exactly the same mass. For that reason, PEAKS uses *L* in *de novo* sequencing to represent both *I* and *L*. If a user-defined modification modifies *I*, it should be defined on *L* instead. When *I* is defined as a modification site, the modification site will be ignored by the *de novo* sequencing algorithm.



Double-clicking any of the PTMs will display the PTM details in the "PTM Info" pop-up dialogue. The same information can be viewed by selecting a PTM from a list and clicking the View button. If a desired modification is not listed or is different than the listed modification, it can be defined as a customized modification. Click the New button to show the "New PTM" dialogue and enter the information of the particular modification. The newly defined modification will be added to the "Customized" list. Refer to Section 2.2, "PTM Configuration" for details on creating new PTMs and viewing PTM information.

## 2.4. Other Parameters

**Report up to (#peptides) candidates per spectrum** Displays how many peptide *de novo* sequences PEAKS will report per spectrum.

**Implicit Parameters.** **Instrument Type** and **Fragmentation Mode** were specified for each analyzed sample during project creation.

## 2.5. Saving the Parameters for Future Use

After setting up the parameters, you can save them for future use. Click the "Predefined parameters" drop-down menu, select "Save as ...", and enter a name for the current set of parameters to save it. Saved parameters are

available for future analysis. You can load a set of saved parameters by selecting from the "Predefined parameters" drop-down menu.

# 3. Understanding PEAKS *De Novo* Sequencing Results

When *de novo* sequencing is done, PEAKS attaches a result node to the analyzed data in "Project View". Double-click the node to load the analysis report. The analysis report is presented in three pages:

- **Summary**: The outline of PEAKS *de novo* sequencing result, which includes key statistics. It also provides functions to filter and export *de novo* sequences.

- **De novo**: *De novo* peptide sequences listed in the table. Spectrum annotation and other detailed information are provided.

- **LC/MS**: *De novo* peptide sequences visualized in an interactive LC-MS heat map.

## 3.1. Summary View

The "Summary" view reports key statistics as an overview of the result. It also provides functions to filter and export *de novo* sequences:

1. **Result filtering:** Filtering rules can be set up at the top of the "Summary" view. Low quality *de novo* sequences can be filtered out by specifying a minimal threshold of the ALC score. More detail about how to properly set the ALC threshold is discussed in Section 4, "Filtering *De Novo* Sequencing Results".

2. **Result exporting:** *De novo* sequencing results can be exported into various formats by clicking the Export button at the top of the "Summary" view. More detail is discussed in Section 5, "Exporting *De Novo* Results".

The report is composed of three sections:

1. **Notes:** The header of the report. Experiment description or other information can be added to this section. The notes can be edited by clicking the Notes button.

2. **Result Statistics:** Key statistics shown in figures and tables.

3. **Other Information:** Search parameters and MS instrument information.
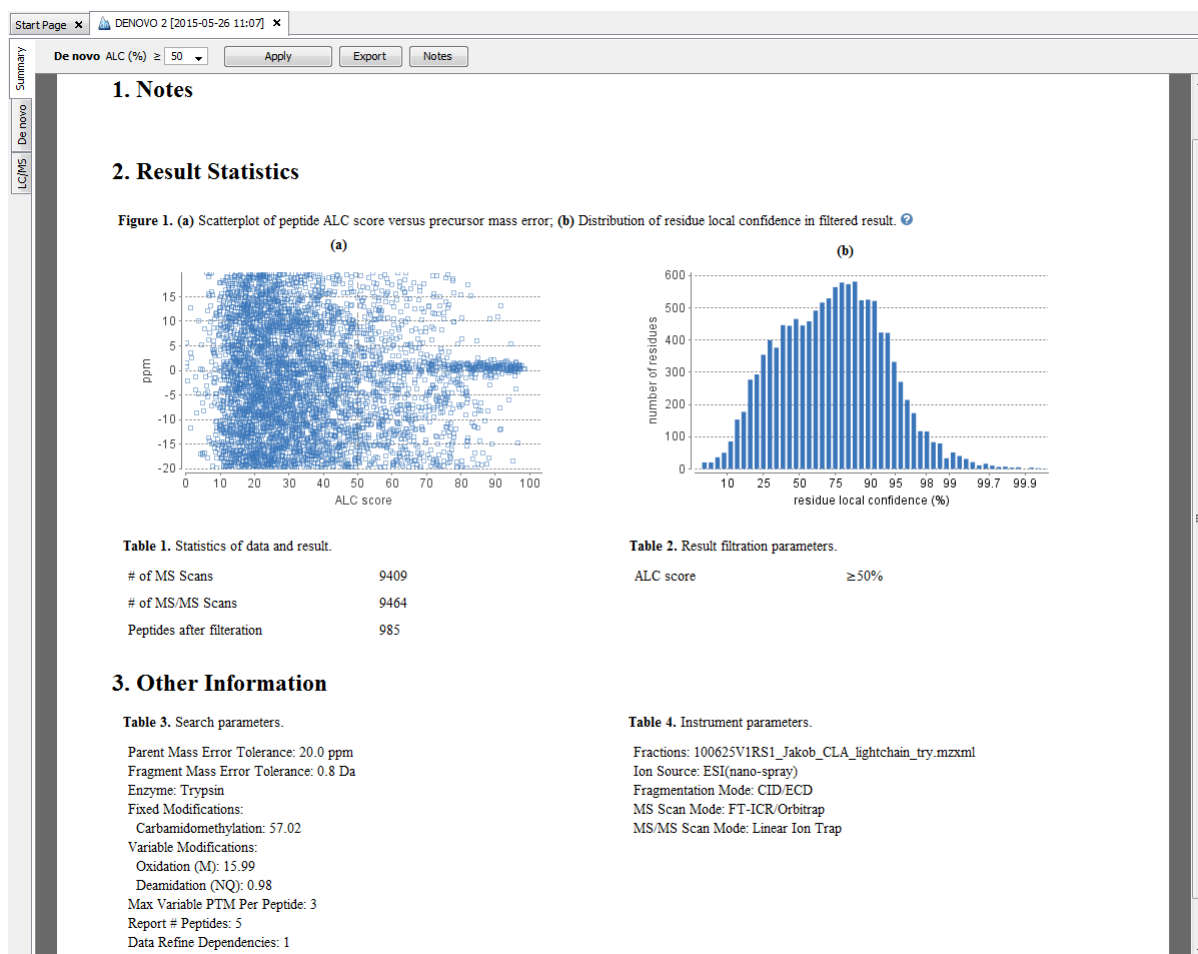
Statistical figures in this report are discussed in a later section Section 4, "Filtering *De Novo* Sequencing Results".

## 3.2. *De Novo* Peptide View

The *de novo* view displays the *de novo* sequencing results in greater detail, as shown in the next figure. The table on the top displays *de novo* sequences, while the bottom section provides additional information about the peptide-spectrum match.

| | Scan | Peptide | Tag Length | ALC (%) ↓ | Length | m/z | z | RT | Area | Mass | ppm | PTM | Mode |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 10 | F1:391 | KVTKC(+57.02)C(+57.02)TE | 8 | 97 | 8 | 513.2397 | 2 | 9.45 | 6.54E7 | 1024.4681 | -3.1 | | CID |
| 11 | F2:2488 | MPC(+57.02)TEDYLSLLLNR | 14 | 97 | 14 | 862.9193 | 2 | 42.96 | 2.81E8 | 1723.8274 | -2.0 | | CID |
| 12 | F1:535 | LLKHKPKATE | 10 | 96 | 10 | 582.8563 | 2 | 12.23 | 7.31E6 | 1163.7026 | -3.9 | | CID |
| 13 | F2:1840 | LVDQ(+.98)LTEFAK | 10 | 96 | 10 | 582.8109 | 2 | 32.48 | 1.76E6 | 1163.6074 | -0.2 | | CID |
| 14 | F1:1490 | KKFWGKYLYE | 10 | 96 | 10 | 681.3649 | 2 | 28.05 | 1.01E8 | 1360.7180 | -2.1 | | CID |
| 15 | F2:2919 | LVNELTEFAK | 10 | 96 | 10 | 582.3175 | 2 | 51.00 | 9.25E5 | 1162.6233 | -2.4 | | CID |
| 16 | F1:1057 | LAHRFKDLWE | 10 | 96 | 10 | 657.8417 | 2 | 21.11 | 3.8E7 | 1313.6880 | -14.6 | | CID |
| 17 | F1:2142 | VSVLLRLAKE | 10 | 96 | 10 | 564.3593 | 2 | 37.80 | 2.28E5 | 1126.7073 | -2.9 | | CID |
| 18 | F2:1344 | STVFDKLK | 8 | 96 | 8 | 469.2703 | 2 | 25.55 | 6.48E5 | 936.5280 | -2.0 | | CID |
| 19 | F2:583 | ATEEQLK | 7 | 96 | 7 | 409.7146 | 2 | 13.49 | 5.32E7 | 817.4182 | -4.3 | | CID |
| 20 | F2:794 | SHC(+57.02)LAEVEK | 9 | 96 | 9 | 536.7563 | 2 | 17.34 | 2.46E7 | 1071.5020 | -3.6 | | CID |

Sequence=MPC(+57.02)TEDYLSLLLNR, Tag Length=14, ALC=97%, ppm=-1.1

## 3.2.1. Peptide Table

The peptide table shows the filtered *de novo* sequencing result. For each scan, the best candidate is displayed. The result can be sorted by any column. For example, to sort the peptide sequence candidates by the scan number, just click on the title of the "Scan" column.

The following list describes the contents in each column:

- **Scan:** The scan number. The scan number is a unique index for tandem mass spectra in the data.

- **Peptide:** The amino acid sequence of the peptide as determined by *de novo* sequencing. A modified residue is followed by a pair of parentheses enclosing the mass of that modification.

- **Tag Length:** The length of the longest sequence tag in the peptide. When low confidence residues are reduced to mass tags, the remaining residues become sequence tags. The length of the longest tag is reported in this column. Refer to Mass Tags.

- **ALC (%):** The average local confidence. ALC is calculated as the total of the residue local confidence scores in the peptide divided by the peptide length.

- **Length:** The number of amino acids in the *de novo* sequence.

- **m/z:** The precursor mass-to-charge ratio.

- **z:** The precursor charge.

- **RT:** The retention time (elution time) for the spectrum as recorded in the data.

- **Area:** The area under the curve of the peptide feature found at the same m/z and retention time as the MS/MS scan. This can be used as an indicator of the abundance of the peptide.

- **Mass:** The monoisotopic mass of the peptide.

- **ppm:** The precursor mass error, calculated as $10^6 \times$ (precursor mass - peptide mass) / peptide mass.

- **PTM:** The types and the numbers of the modifications present in the peptide shown as color-coded icons.

- **Mode:** The fragmentation mode in which *de novo* sequencing is performed by the algorithm.

**Confidence Scores.** Amino acids in *de novo* sequences are color-coded according to their local confidence scores. Red represents a very high confidence (greater than 90%); purple represents a high confidence (80 to 90%); blue represents a medium confidence (60 to 80%); and black represents low confidence (less than 60%). Hover the cursor over the sequence to view local confidence scores of each individual amino acid.



**Mass Tags.** The low confidence residues can be displayed as mass tags by adjusting the local confidence score threshold using the button . If the score is set at 0, all of the amino acids in the peptide sequences will be displayed. Increasing the threshold will reduce residues below the threshold to mass tags. The Tag Length column in the peptide table is automatically updated when the threshold is adjusted.



**Modifications.** In the Peptide column, modifications are shown as brackets in peptide sequences. Consider the example, `SHM(+15.99)TNLNGNPEDR`, `(+15.99)` shows that the methionine (M) residue is modified with delta mass of 15.99 Da. In the PTM column, modifications in a sequence are shown as icons. Hover the cursor over the icons to show the modification names.

## Note

N-term modifications are shown after the first residue in the sequence, and C-term modifications are shown after the last residue.

**Peptide Search.** The peptide table can be searched by entering the value in the search bar located in the top-right corner. The peptides can be searched by scan number, sub-sequence, precursor m/z, retention time, and modification mass. The matched peptides are highlighted in the table, and can be iterated by clicking the up and down buttons in the search bar.



## Note

Precursor m/z, retention time, and modification mass are searched approximately. For example, 130.3 will match any value from 130.25 to 130.35 (exclusive). And 130 will match from 129.5 to 130.5 (exclusive).

## 3.2.2. Spectrum Annotation

The spectrum annotation displays a graphical representation of the peptide-spectrum match.
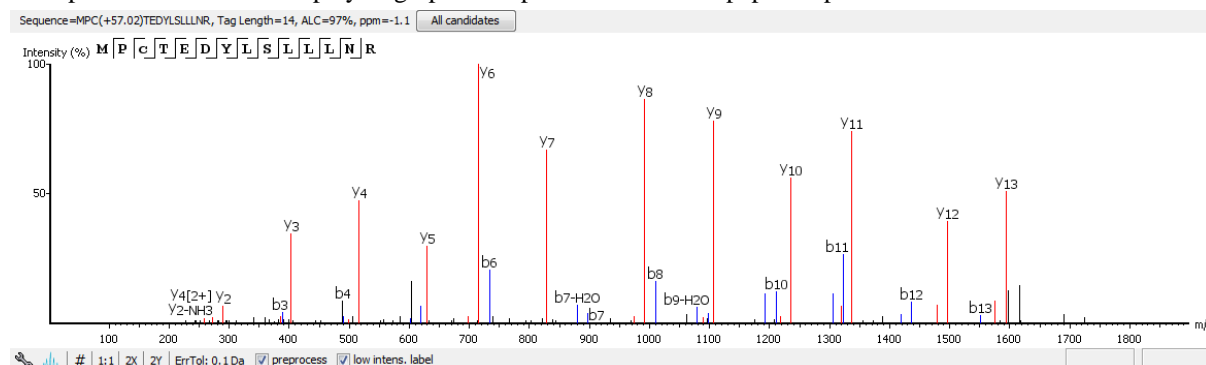


The title bar shows the information of the peptide sequence that is being displayed in the spectrum. Press the "All candidates" button to open a pop-up window that displays all alternative peptides. Click on a peptide sequence in the pop-up window to select and display the annotation.
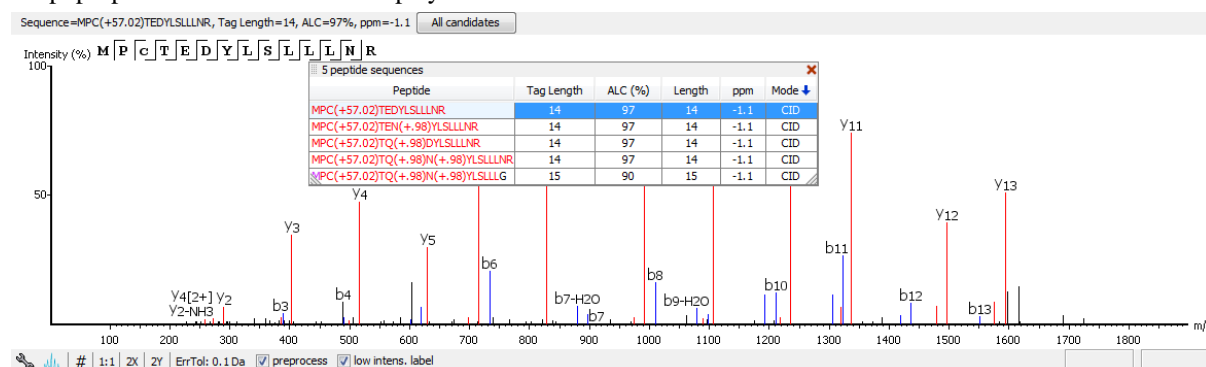


Moving the cursor over the peptide sequence in the spectrum will show the mass transitions for a particular amino acid residue.

The spectrum annotation panel provides convenient ways to zoom and navigate within the spectrum:

- **Zoom to a m/z region:** Drag horizontally from the start m/z and to the end m/z with left mouse button.

- **Zoom in/out smoothly:** Place the cursor at a particular m/z value (right below the x-axis), zoom in/out by scrolling the mouse wheel.

- **Increase/Decrease peak intensity:** Place the cursor in the spectrum and scroll the mouse wheel.

- **See the whole spectrum:** Double click in the spectrum or click the 1:1 button.

To change the "Spectrum Annotation Settings", click the ⚒ button to open the dialogue. Refer to Section 1.4, "Spectrum Annotation Preferences" for more details.

You can use the profile  and peak  buttons to switch the Spectrum View between the profile mode and the peak mode.

The  icon will switch the y-axis to absolute intensity. Once selected, scrolling over peaks will display the absolute intensity instead of the relative intensity. Spectrum exports will also display absolute intensity on the y-axis.

The "ErrTol" slider is used to adjust the error tolerance to view the display of matched ions.

The preprocess checkbox allows for transition between the spectrum after centroiding and deconvolution and the original spectrum

The "intensity threshold" checkbox provides an option to annotate lower intensity peaks.

Moving the cursor over a peak in the spectrum will show the annotation, as well as the peak's m/z, relative intensity, and ppm mass error. The m/z ratio and the relative intensity can also be found on the bottom-right corner of the spectrum annotation panel.

## Note

When the  is selected, scrolling over peaks will display the absolute intensity of the peak. When the  is selected, scrolling over peaks will display the relative intensity of the peak.

### 3.2.3. Ion Table

The "Ion Match" tab at the bottom panel of the *de novo* view contains the "Ion Table" that shows the calculated mass of possible fragment ions. If a fragment ion is found in the spectrum, its mass value is displayed in color.

N-terminal ions are shown in blue and C-terminal ions are shown in red. A fragment ion is found when there is a matching peak within the mass error tolerance, as defined in the *de novo* sequencing parameters, and when relative intensity of the matching peak is at least 2%. The ion types displayed in the table can be configured in "Spectrum Annotation Settings" (Section 3.2.2, "Spectrum Annotation").

Fragment ions can be selected by clicking the headers. Selected fragment ions are highlighted in both the spectrum annotation and the error map.

| # | b | b-H2O | b-NH3 | Seq | y | y-H2O | y-NH3 | # |
|---|---|---|---|---|---|---|---|---|
| 1 | 116.03 | 98.02 | 99.01 | D | | | | 14 |
| 2 | 230.08 | 212.07 | 213.05 | N | 1489.76 | 1471.77 | 1472.74 | 13 |
| 3 | 327.17 | 309.15 | 310.14 | P | 1375.71 | 1357.71 | 1358.71 | 12 |
| 4 | 455.19 | 437.18 | 438.16 | Q | 1278.69 | 1260.56 | 1261.66 | 11 |
| 5 | 556.24 | 538.23 | 539.20 | T | 1150.64 | 1132.62 | 1133.60 | 10 |
| 6 | 693.30 | 675.28 | 676.27 | H | 1049.60 | 1031.49 | 1032.55 | 9 |
| 7 | 856.36 | 838.35 | 839.33 | Y | 912.52 | 894.51 | 895.49 | 8 |
| 8 | 1019.43 | 1001.43 | 1002.40 | Y | 749.46 | 731.45 | 732.43 | 7 |
| 9 | 1090.47 | 1072.47 | 1073.43 | A | 586.44 | 568.29 | 569.32 | 6 |
| 10 | 1189.53 | 1171.51 | 1172.50 | V | 515.35 | 497.34 | 498.33 | 5 |
| 11 | 1260.56 | 1242.56 | 1243.54 | A | 416.29 | 398.28 | 399.26 | 4 |
| 12 | 1359.66 | 1341.62 | 1342.62 | V | 345.24 | 327.17 | 328.22 | 3 |
| 13 | 1458.70 | 1440.20 | 1441.67 | V | 246.18 | 228.17 | 229.15 | 2 |
| 14 | | | | K | 147.11 | 129.10 | 130.09 | 1 |

## 3.2.4. Error Map

The "Error Map" shows the mass errors of matched fragment ions. The m/z ratio is displayed on the x-axis and the error is listed on the y-axis in Daltons. Each matched fragment ion is represented by a dot. The ion types displayed can be configured in "Spectrum Annotation Preferences" (Section 3.2.2, "Spectrum Annotation").
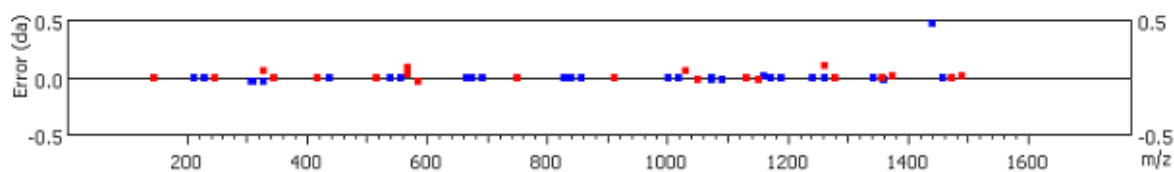


## 3.2.5. Spectrum Alignment

The "Spectrum Alignment" displays how the fragment ions generated from the peptide align with the spectrum. N-terminal ions are shown in blue and C-terminal ions are shown in red.
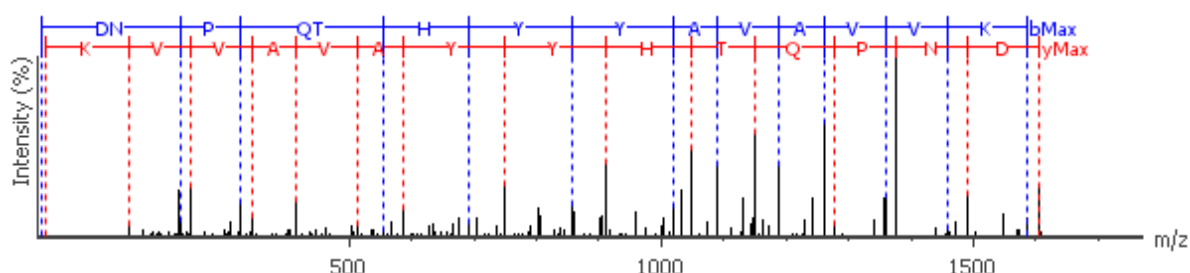


## 3.2.6. Parent Scan

The "Survey" tab displays the precursor ion spectrum. The buttons that appear in this section are identical to those explained above in the "Spectrum Annotation" section.
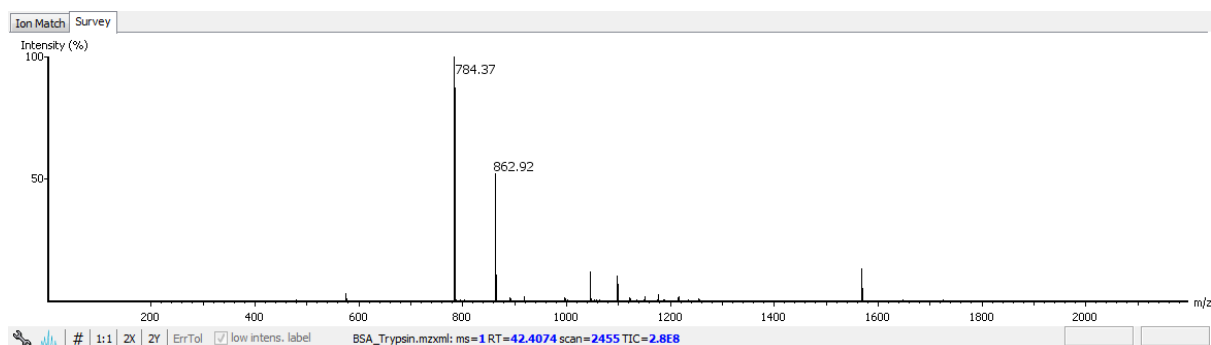
### 3.2.7. LC-MS Snapshot

LC-MS Snapshot displays a zoomed-in snapshot of the peptide-spectrum match in the LC-MS view. It has three modes of viewing that can be toggled using the selection bar in the top-right corner. In the *2D/3D* mode, the view focuses on the peptide feature containing the peptide-spectrum match, showing the area around the spectrum's precursor m/z and retention time. In the *All* mode, the view shows an overview of the LC-MS run. The spectrum is shown as a red square in the view at its precursor m/z and retention time.
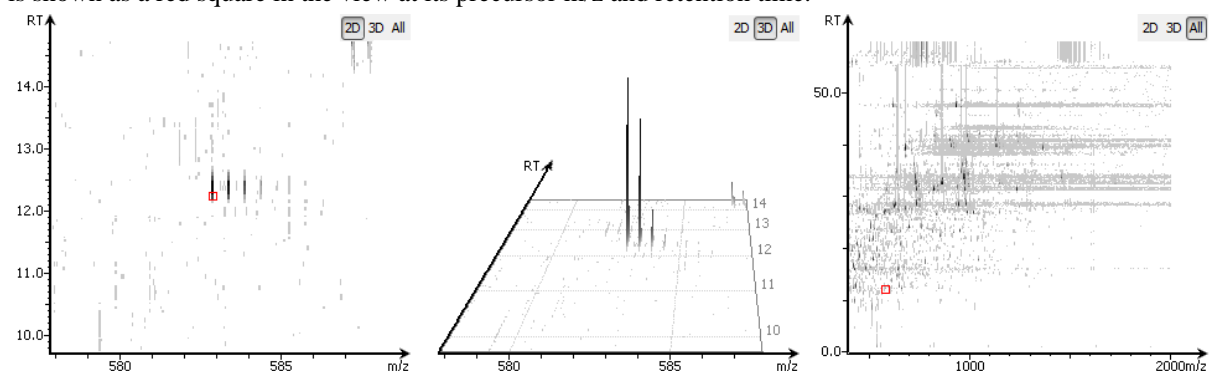


## 3.3. LC-MS View

LC-MS View visualizes *de novo* sequences in an interactive LC-MS heat map. Check the option "De novo tags" to show the locations of MS/MS spectra with *de novo* sequences. These MS/MS spectra are indicated by amber squares. Only the filtered *de novo* sequences are shown in the heat map.

Refer to Section 4, "LC/MS View" for a list of supported interactions in the LC-MS heat map.
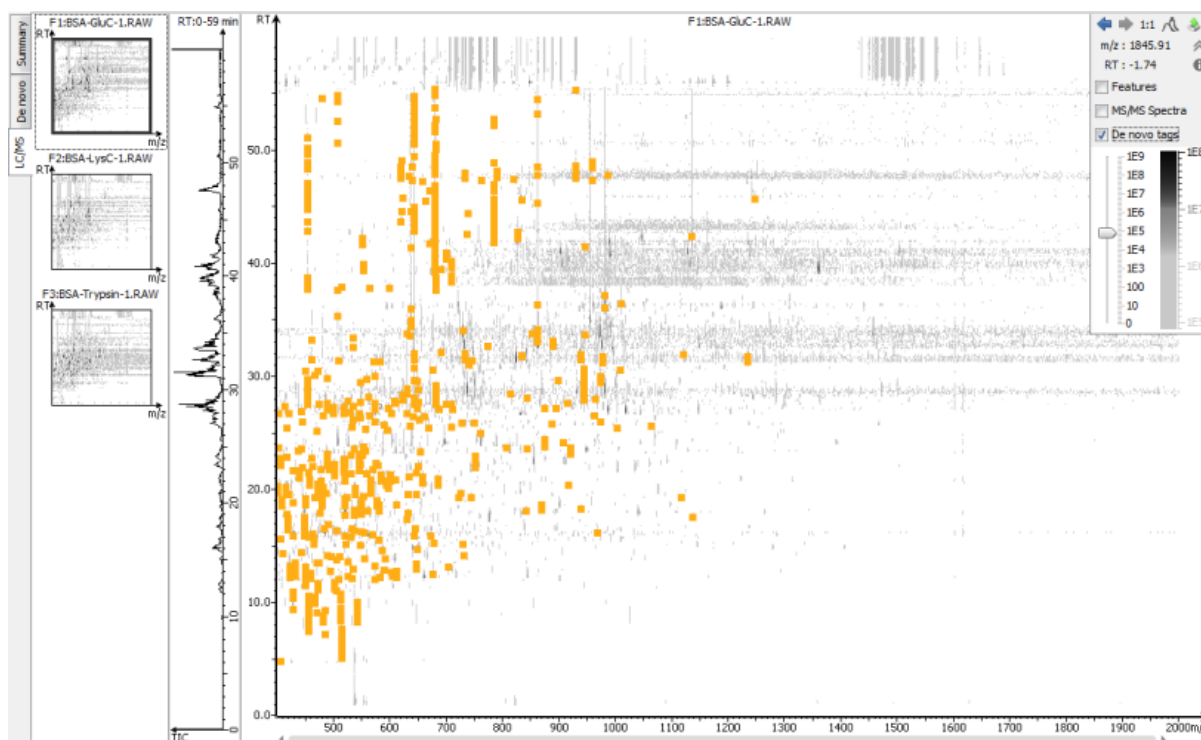
# 4. Filtering *De Novo* Sequencing Results

PEAKS *de novo* sequencing result can be filtered by ALC (Average Local Confidence) score. The ALC filtering threshold can be specified in the top panel of the "Summary" view. Click the Apply Filters button to filter the results. The result in the "Summary" view, "De Novo" view, and "LC/MS" view will be updated accordingly.



## Note

Whenever the score threshold is changed, the "Apply" button will be highlighted in red to remind you that the change has NOT taken effect yet.

The purpose of filtering is to remove poor sequences in which residues are incorrectly identified. By default, the ALC threshold is set to 50%. Statistical figures are shown to help evaluate the overall quality of the sequencing result. They also provide guidelines for determining a proper ALC threshold:

**Peptide ALC Score Scatter Plot.** Figure 1(a) is a scatter plot showing ALC score versus precursor mass error (in ppm) of the *de novo* sequences. In a typical *de novo* sequencing analysis, peptides with higher ALC scores are more likely to have a small mass error, indicating a higher chance that a peptide is correctly sequenced. This figure provides a rule-of-thumb for setting the ALC score threshold. The threshold can be set at the score where the precursor mass error starts to converge.
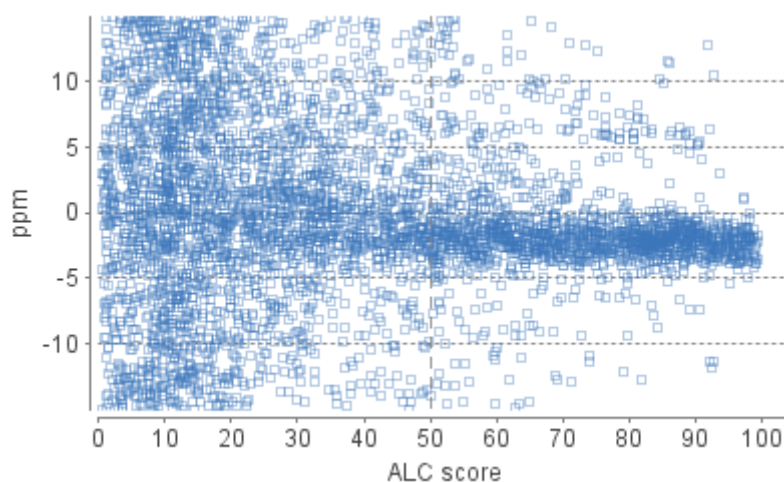
**Residue Local Confidence Distribution.**     Figure 1(b) shows the local confidence score distribution of residues in filtered *de novo* sequences. It is essentially a mixed score distribution of both correct and incorrect residues. This figure allows you to examine the proportion of correct and incorrect residues when adjusting the ALC threshold. By increasing the ALC threshold, the number of incorrect residues will decrease because poor sequences are filtered out first. It is recommended to gradually raise the ALC threshold until the number of correct residues starts to significantly decrease.

## Note

In some cases, the local confidence score may not be able to distinguishably separate incorrect and correct residues as shown in the user manual.
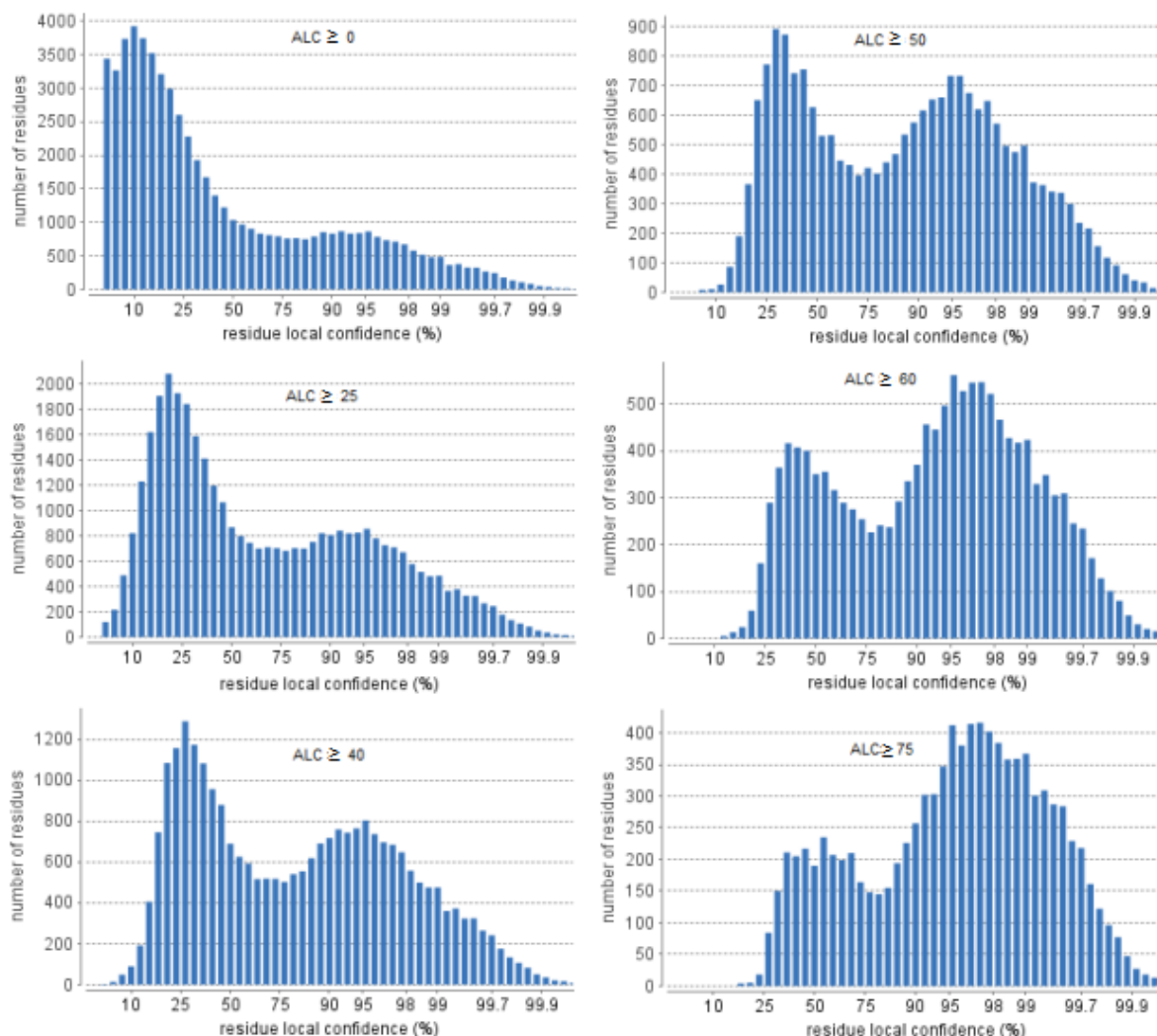
# 5. Exporting *De Novo* Results

The Export button in the title bar of the "Summary" view allows exporting of the filtered results into various formats. It provides the opportunity to share the report in HTML and to also use the result as supplemental information in a publication. Please refer to Section 3, "Exporting *De Novo* Results" for details.

# 6. *De Novo* Sequencing using Spectral Pairs or Triplets

PEAKS supports *de novo* sequencing using spectral pairs or triplets that are generated using different fragmentations. Mass spectrometers, like the LTQ Orbitrap Velos, are capable of generating multiple tandem mass spectra using different fragmentation types for the same precursor ion. Using complementary fragmentations such as CID/HCD and ETD, the accuracy and confidence of automated *de novo* sequencing can be significantly improved.

PEAKS automatically detects spectral pairs and triplets when doing *de novo* sequencing.The fragmentation type of data fractions containing spectral pairs or triplets has to be set to "Mixed" during the project creation. In "Mixed" mode, PEAKS loads the fragmentation type for each tandem spectrum from the meta information encoded in the data file. A spectral pair or triplet is detected if the following conditions are satisfied:

- The pair or triplet of spectra are under the same precursor scan and are fragmented from the same precursor ion.

- The pair or triplet of spectra have different fragmentation types. For example, HCD/ETD for spectral pairs, CID/HCD/ETD for spectral triplets.

For spectral pairs and triplets, PEAKS *de novo* sequencing algorithm analyzes each individual spectrum and then constructs a better sequence with a higher ALC score. In the Peptide Table, a *de novo* sequencing result of a spectral pair or triplet is presented in a single row. When viewing the sequence, PEAKS shows all spectra annotated by the sequence. Clicking the All candidates button will show the other *de novo* sequences generated for the individual spectrum.
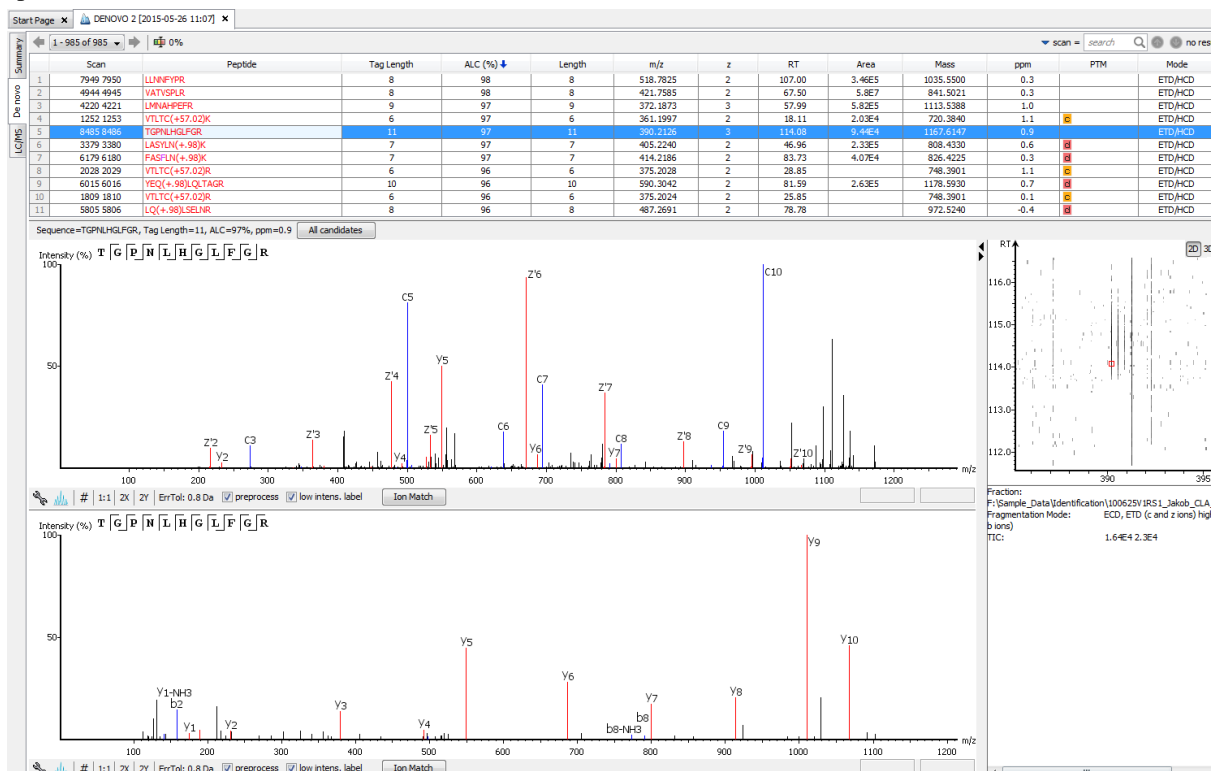
# 7. *De Novo* Sequencing on Single Spectrum

To run PEAKS *de novo* sequencing on a single spectrum, select the spectrum in the MS/MS view of the sample, click the right button of the mouse, and select the PEAKS Auto Denovo command from the pop-up menu.

## Note

Residue local confidence scores are NOT calculated for *de novo* sequences generated in this mode.
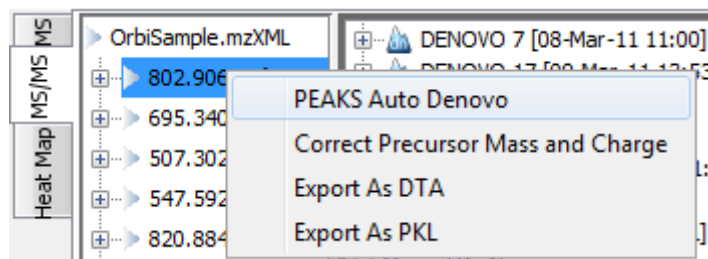


# 8. Manual *De Novo* Sequencing

PEAKS provides a set of tools to help manually sequence a peptide from the spectrum.

## Note

Manual *de novo* sequencing does NOT support ETD spectra.

## 8.1. Manual *De Novo* Graphical User Interface

To create a new peptide candidate for manual *de novo* sequencing, select the m/z value in the "Result Panel" and right-click to display up a pop-up menu.
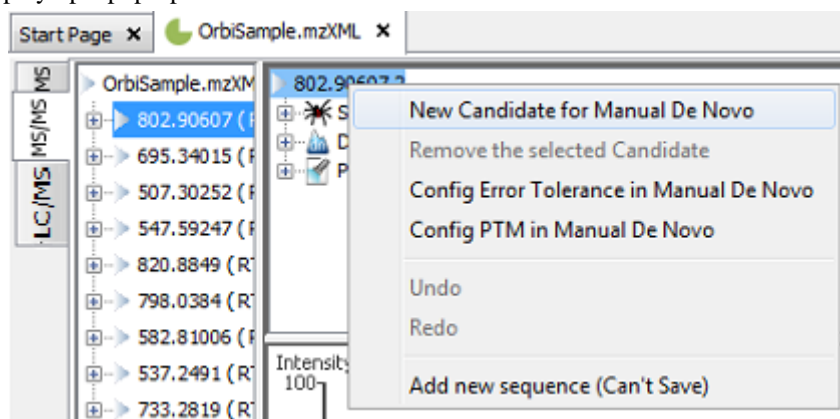


Select New Candidate for Manual De Novo from the pop-up menu. A new candidate will be created under the 'Manual *De Novo*' heading. The new candidate is initiated as unsequenced; so, it is represented by the total mass of the residues (as seen in the example below).
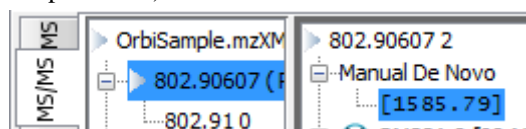


## Note

The pop-up menu will not be accessible if you have highlighted any of the results in the "Result Panel".

---

The figure below shows the main panels related to manual *de novo* sequencing. The five main panels are indicated in the figure below:
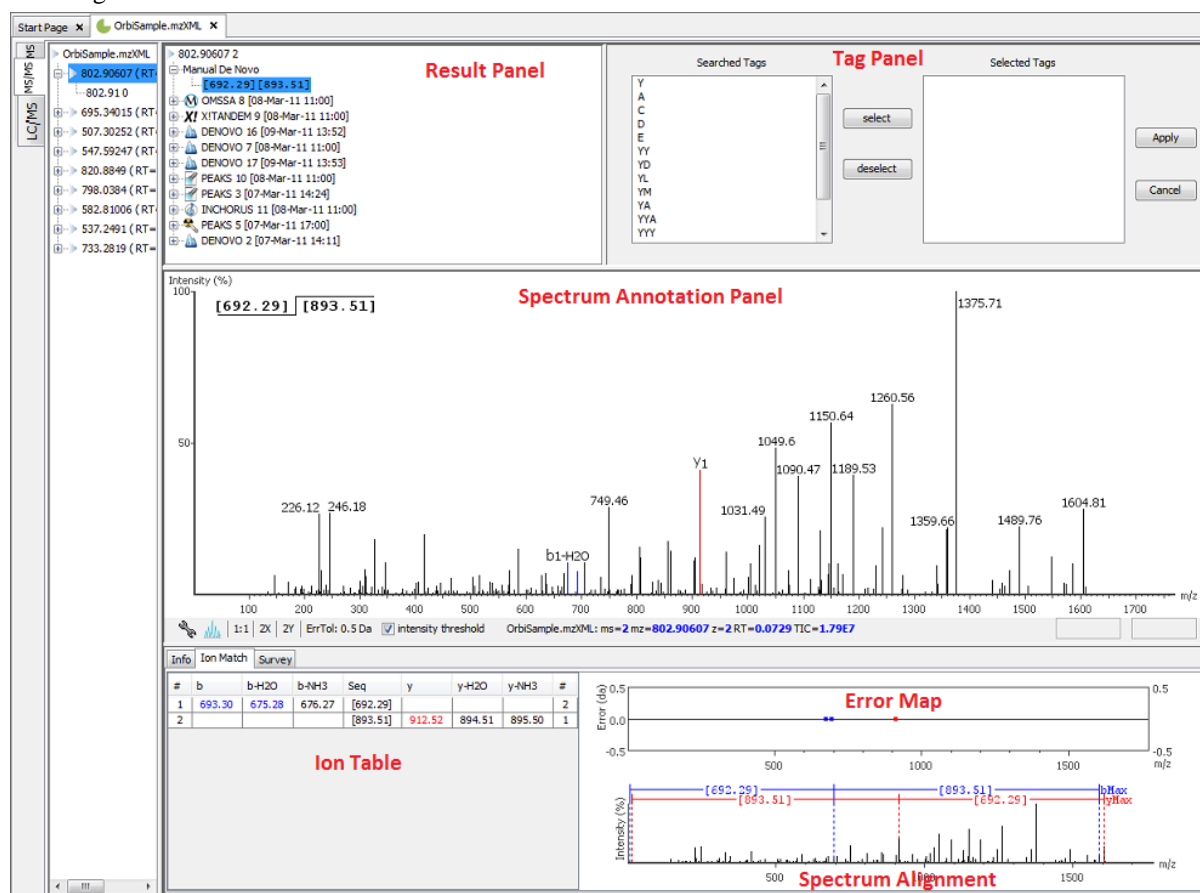


The panels are briefly described below:

- **Result Panel:** The "Result Panel" shows all sequencing results. The results of manual *de novo* are listed under "Manual *De Novo*".

- **Spectrum Annotation Panel:** The "Spectrum Annotation" shows a graphical representation of the spectrum, the peaks in the spectrum, the user-selected peaks, and the assigned ions. Pick a peak on the panel with the cursor and assign ions or tags to it in manual *de novo*.

- **Ion Table Panel:** The "Ion Table" shows the proposed ions with their corresponding masses. The default Ion Table will display immonium, b, b-H2O, b-NH3, y, y-H2O, and y-NH3 ions.

- **Spectrum Alignment and Error Map Panel:** The "Spectrum Alignment" shows how the proposed ions assigned in manual *de novo* align with the spectrum. By default, the "Spectrum Alignment" displays b-ions and y-ions. The b-ions are shown right-to-left in blue, while the y-ions are shown left-to-right in red. The "Error Map" displays the confidence assigned to each ion.

- **Tag Panel:** The "Tag Panel" will appear when you search tags or ions in the spectrum. You can select the tags in the list using the "Select" button. Clicking "Apply" will add the selected tags to the sequence candidate.

# 8.2. Manual *De Novo* Operations

When the cursor is placed in the "Spectrum Annotation" panel, a green (by default) triangle follows the movement of the mouse. This is the *Position Bar* and it is used as a cursor for all manual *de novo* operations. The cursor's position on the m/z scale and its relative intensity are shown in a pop-up window on top of the *Position Bar*.
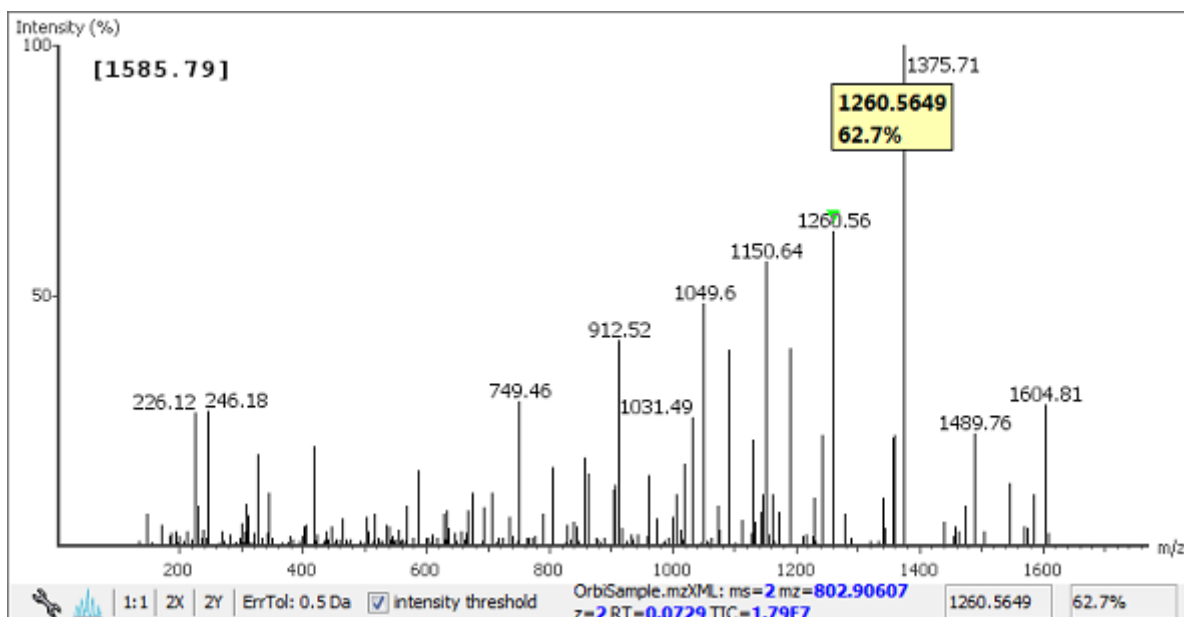
**Selecting a peak.**   To select a peak, simply click on it. A blue (by default) arrow, called the *Freeze Bar*, indicates the selected peak. Alternatively, an ion peak can be selected by clicking on its corresponding cell in the Ion Table.
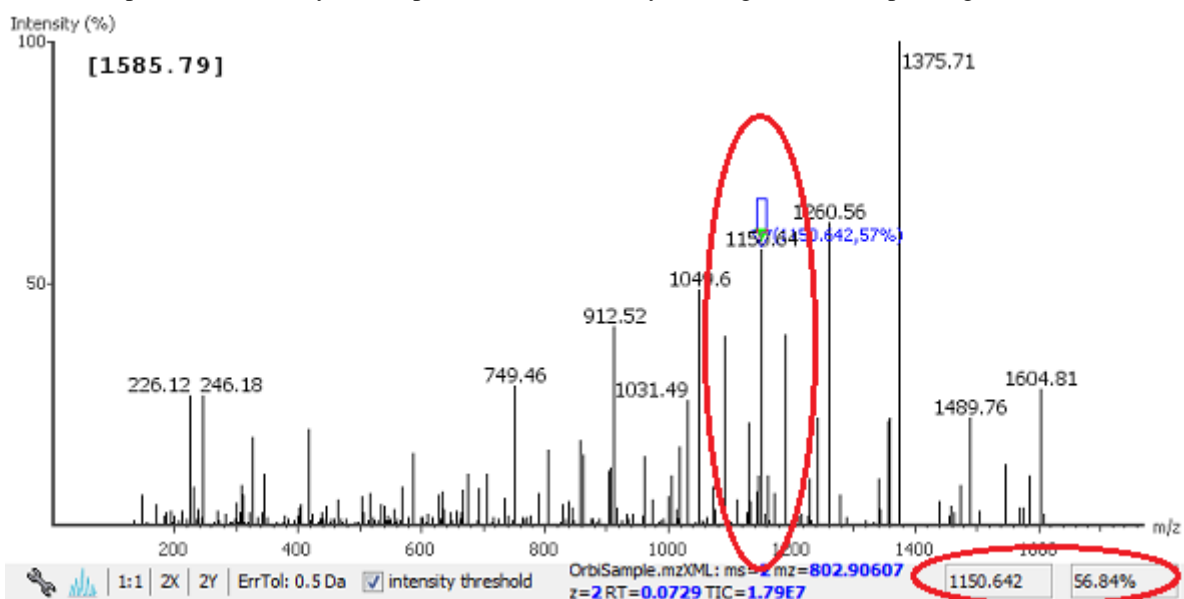


**Measure the m/z difference between two peaks.**   Select a peak (blue arrow) with the *Freeze Bar* and move the mouse to the left or to the right. Hold the *Position Bar* (green triangle) above another peak. A pop-up window displays the difference between the two peaks (in the example below the difference is 109.92297).
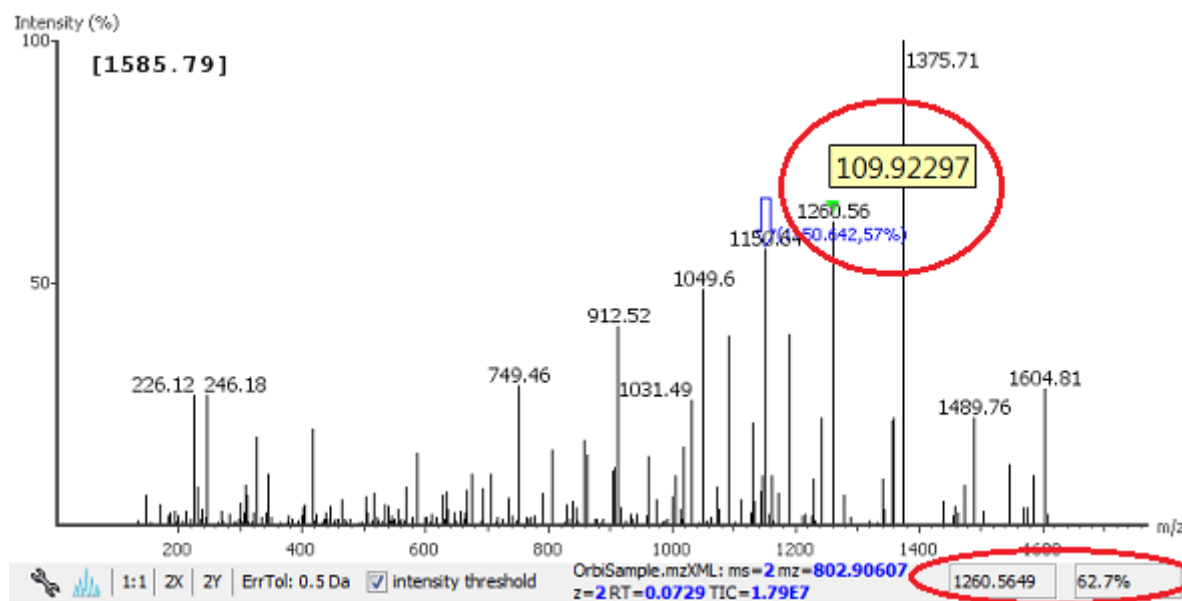
**Deselect a peak.** The Freeze Bar can be deselected by holding Ctrl and clicking anywhere in the spectrum.

**Zoom in on part of the spectrum.** In the "Spectrum Annotation" panel, click and drag the mouse horizontally. The selected area will be enhanced and shown in the "Spectrum Annotation" panel. Click the 1:1 button to return to the default view.

**Setting/removing ions to/from a peak.** Select a peak and right-click anywhere in the "Spectrum Annotation" panel. Select Set y-ion from the pop-up menu to designate the peak as a y-ion, or Set b-ion from the pop-up menu to designate the peak as a b-ion. Click on Remove ion to remove the ion that you have previously set.
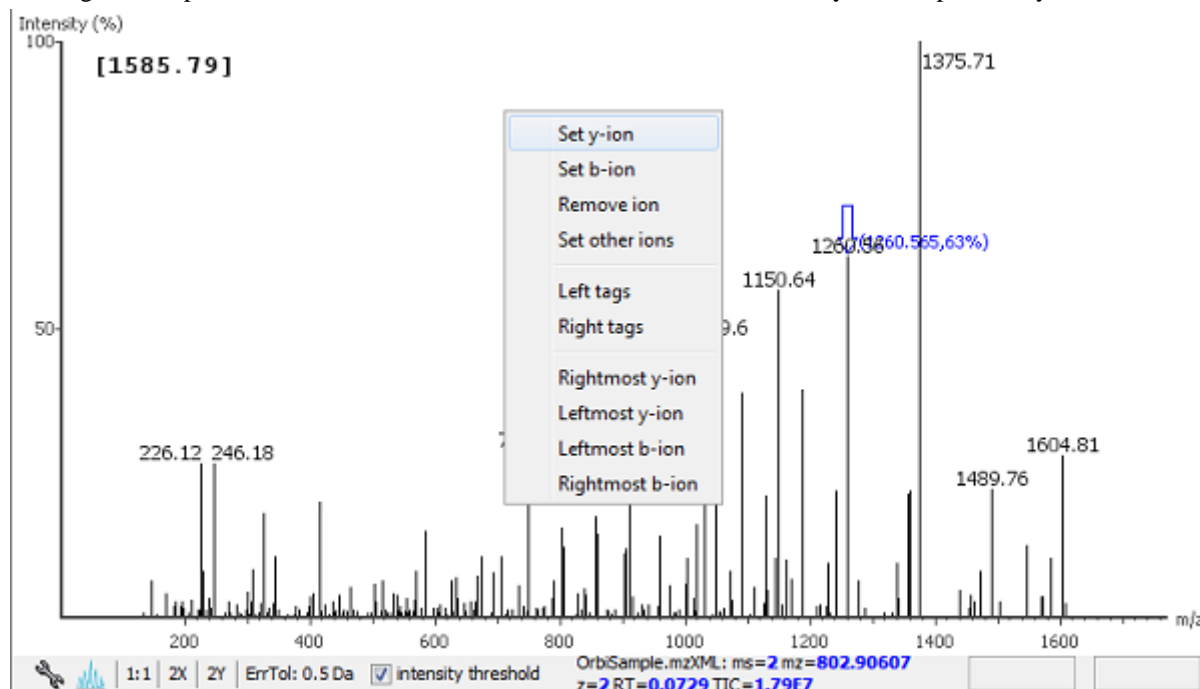


Select Set other ions from the pop-up menu to view the "Ion Editor" dialogue box. The "Ion Editor" dialogue allows addition or removal of ion designations to/from a peak. Select either "C Term Ion" or "N Term Ion" to see the C- and N- terminal ions, respectively. Then select an ion from the ion list and press the Add button to add it to the selected ion list. Remove an ion from the selected ion list by selecting it and pressing the Remove button. Click the Apply button to apply the changes to the selected peak.
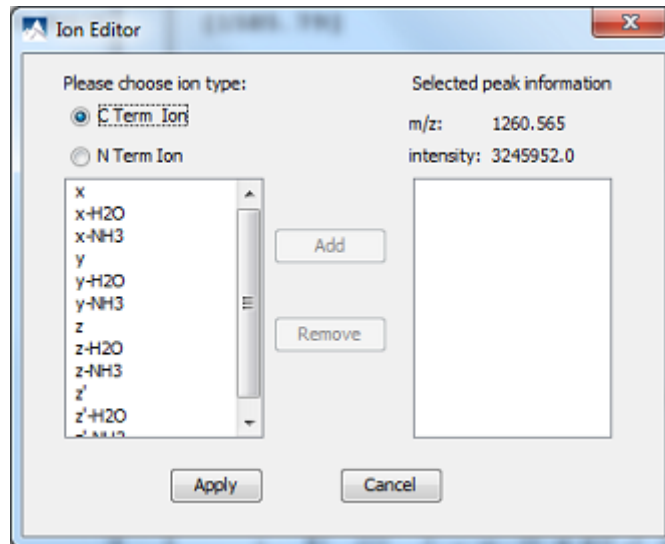
After setting an ion, the "Spectrum Annotation" panel, the "Spectrum Alignment and Error Map" panel, and the "Ion Table" panel will reflect the changes. The *peptide sequence candidate* name (as displayed in the "Result" panel and on the top of the "Spectrum Annotation" panel) will also change to reflect the mass remaining to be sequenced on either side of the ion.

In the example below, the selected peak at 1260.5649 m/z was designated as a y-ion.
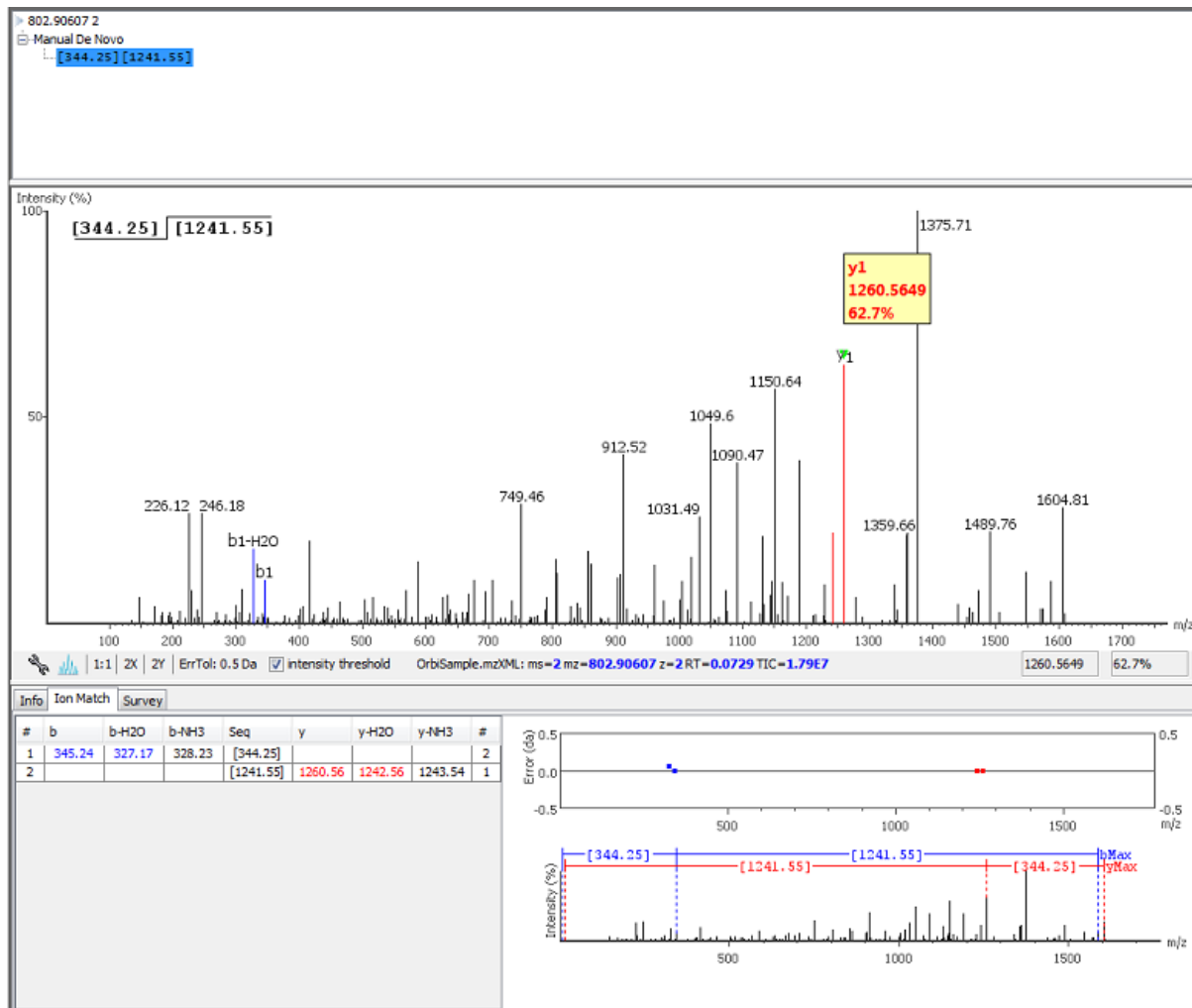
## Note

The manual *de novo* candidate information is updated in the "Result" panel, the "Ion Table" panel, and the "Spectrum Alignment and Error Map" panel. The selected ions are also annotated and color coded in the "Spectrum Annotation" panel.

After setting two ions, PEAKS will estimate the residue found between them (if a residue corresponds closely to the mass difference). The *peptide sequence candidate* name will change to show the residue and the mass remaining to be sequenced on either side of the residue. All other panels will also reflect the changes.
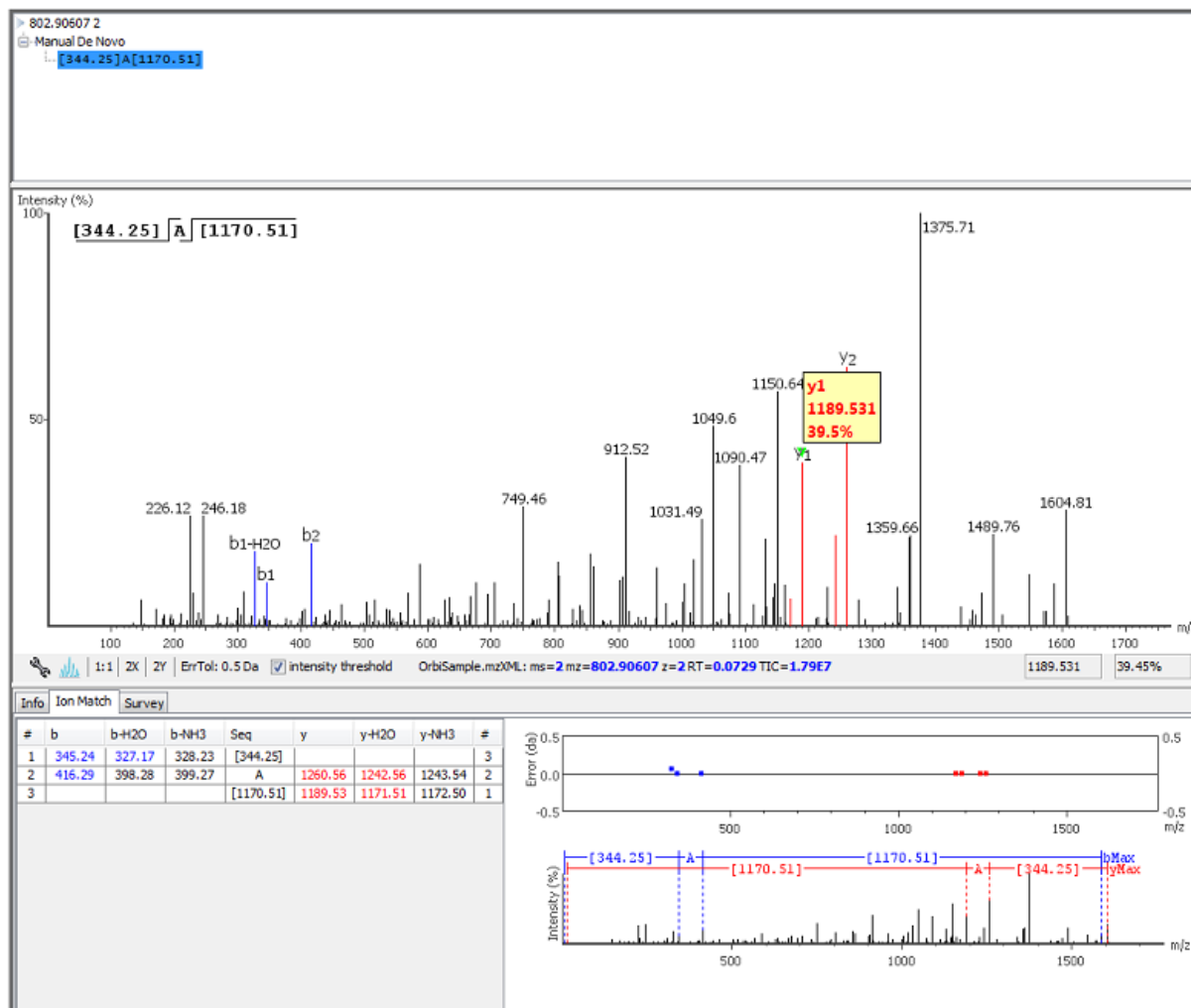


**Searching the left or right side of the spectrum for the first/last y or b ion (Search a sequence tag).** Select a peak and right-click anywhere in the Spectrum View to trigger the pop-up menu. From the menu, select either Left tags or Right tags. PEAKS will select the appropriate terminal tags and show them in the "Tag" panel (see below).

To test the suitability of a tag, highlight it in the "Searched Tags" list. The corresponding information for the tag will be shown in the "Spectrum Annotation" panel, the "Ion Table" panel, and the "Spectrum Alignment and Error Map" panel. One or more tags can be inserted by highlighting the desired tags, clicking Select to move them into the "Selected Tags" list, and then clicking the Apply button. Press the Cancel button at any time to exit the search and discard any changes.

**Undoing an edit.** Errors during the manual *de novo* sequencing can be undone easily. With the peptide candidate still selected in the "Result" panel, right-click and select the Undo command from the pop-up menu to return to the previous peptide sequence. This button can be used multiple times to return to previous edits.
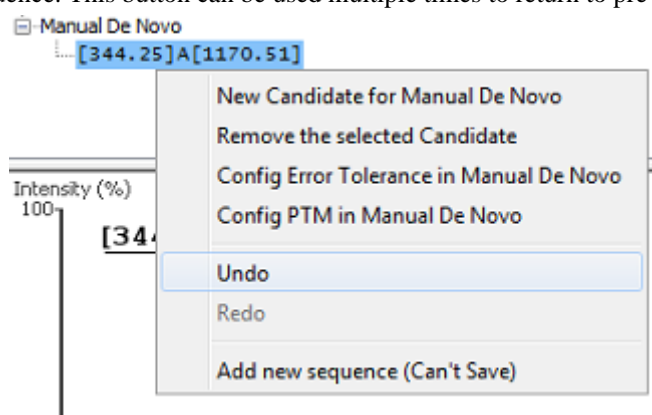


**Redoing an edit.** An undone edit during the manual *de novo* sequencing can be reapplied easily. Right-click and select the Redo command from the pop-up menu with the peptide candidate still selected in the "Result" panel. This button can be clicked multiple times to return to later stages in the edit.

**Error Tolerance.** To set the mass error tolerance in the manual *de novo* sequencing, select the peptide candidate in the "Result" panel, and then right-click and select the Config Error Tolerance in Manual De Novo command from the pop-up menu. This will open a dialogue where the error tolerance can be set.

**PTM Configuration.** To identify post-translational modifications (PTMs) while manually *de novo* sequencing a spectrum, start by selecting the peptide candidate in the "Result" panel. Right-click and select the Config PTM in Manual De Novo command from the pop-up menu to open the "PTM Setup" window. To know more about the PTM configuration, refer to Section 2.3, "PTM".

# Chapter 9. Peptide, PTM, and Mutation Identification (PEAKS DB, PEAKS PTM, and SPIDER)

## 1. Overview

PEAKS software provides a complete set of database search tools for in-depth protein analysis. With PEAKS DB, PEAKS PTM, and SPIDER, proteins in the sample can be identified with high sensitivity and accuracy. It is also possible to find potential sites for modification and mutation. PEAKS supports multiple enzyme digestion by which almost full sequence coverage can be achieved for single protein studies. PEAKS database search tools have built-in result validation using an enhanced target-decoy approach. The false discovery rate (FDR) is estimated to ensure that only valid results are reported.

PEAKS DB is a database search tool assisted by PEAKS *de novo* sequencing to achieve high sensitivity and accuracy.

### Note

For more details, refer to the paper: *"PEAKS DB: De Novo sequencing assisted database search for sensitive and accurate peptide identification" Mol Cell Proteomics. 2011 Dec 20.*

PEAKS PTM is a dedicated search tool for peptides with unspecified modifications and mutations. It can search with unlimited number of modifications, allowing searches with all modifications and mutations in the Unimod database.

### Note

For more details, refer to the paper: *"PeaksPTM: Mass Spectrometry Based Identification of Peptides with Unspecified Modifications" Journal of Proteomics Research, 2011, 10(7): 2930-2936.*

SPIDER is a dedicated search tool for finding novel peptides that are homologous to peptides in a given protein database.

### Note

For more details, refer to the paper: *" SPIDER: Software for Protein Identification from Sequence Tags Containing De Novo Sequencing Error." J Bioinform Comput Biol. 2005 Jun;3(3):697-716.*

PEAKS DB, PEAKS PTM, and SPIDER can be launched together as a workflow in a PEAKS search. An analysis report is generated to show the combined database search result. The steps to run a PEAKS search are:

1. Select a data node in "Project View". The selected node can be a fraction, a sample, or a project. PEAKS will analyze all fractions under the selected node. Launch a PEAKS search by clicking the "PEAKS DB Search" button on the tool bar.
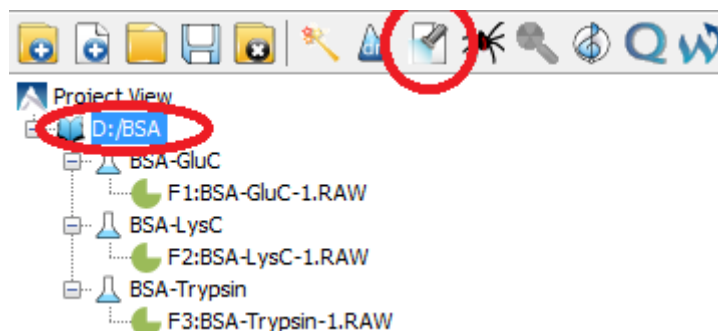
**Note**
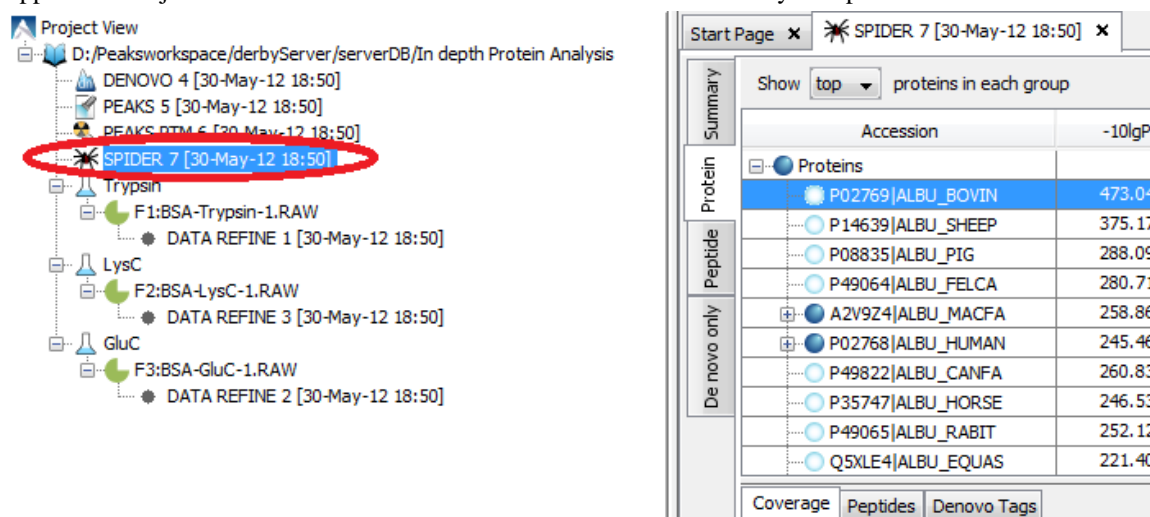
Refer to Chapter 4, *Project Wizard - Create a PEAKS Project, Load Data and Set up a Workflow* for details on creating a PEAKS project.

2. Specify the PEAKS DB parameters in the pop-up dialogue. Most of the parameters are self-explanatory and the default parameters provide a good starting point for the analysis. PEAKS PTM and SPIDER can be enabled in the dialogue to automatically run after PEAKS DB search. Click OK to start a PEAKS search.

**Note**

If data is not yet refined, a prompt to specify data refinement parameters appears. Refer to Chapter 7, *Data Refinement* for details on refinement.

3. Wait for the analysis to finish. Search times range from minutes to hours depending on several factors, including the dataset size and the computer speed. New result nodes for PEAKS DB, PEAKS PTM, and SPIDER will appear in "Project View". Double-click the result nodes to examine analysis reports.



# 2. Setting PEAKS Search Parameters

After selecting a project node, a sample node, or a fraction node in "Project View" and clicking the PEAKS DB Search toolbar icon , the PEAKS DB parameters dialogue will appear:

**Error Tolerance.** **Precursor mass** Precursor mass tolerance. PEAKS Search identifies peptides within the precursor mass tolerance. Precursor mass tolerance can be specified in either Da or ppm.

**Fragment ion** Fragment ion m/z tolerance. PEAKS Search uses this value when scoring peptide-spectrum matches. PEAKS considers that a fragment ion is matched if the calculated m/z is within the tolerance.

Error tolerance parameters need to be set consistently with the mass accuracy of the instrument. The default parameters for the instrument generating the data can be loaded by choosing "default" in "Predefined parameters".

## Note

When the samples in a project are generated by different instruments and "default" is chosen from the Predefined parameters drop-down menu, the default parameters of the first sample are loaded. However, it is recommended to set the tolerance parameters according to the least accurate instrument.

**Enzyme.** PEAKS search digests the searched protein database *in silico* to generate peptide candidates. It is necessary to specify the enzyme for protein digestion, whether or not non-specific cleavages are allowed, as well as the number of missed cleavage sites allowed in a peptide.

**Enzyme Selection** Select the enzyme used for protein digestion from a drop-down menu. If the enzyme used for protein digestion is not in the list, a new enzyme can be added. Please refer to Section 2.2, "Enzyme" for further details.

**Non-specific Cleavages** Specify whether or not non-specific cleavages are allowed at a peptide terminus. If "None" is selected, the specificity of the selected enzyme is strictly enforced at both termini of a peptide. "One" allows one terminus to disobey the enzyme specificity. "Both" allows both peptide termini to disobey the enzyme specificity.

**Missed Cleavages** Specify the number of missed cleavage sites allowed in a peptide.

## Note

When the selected data node is a project of multiple samples, "Use Sample Enzyme" allows samples to be analyzed separately using their respective enzymes specified during project creation.

## Note

"None" is a special enzyme allowing non-specific cleavage at both ends of the peptide. The "None" enzyme can cut at every residue, generating peptides with lengths up to 65 amino acids. Its usage is recommended when the digestion enzyme exhibits a high degree of non-specificity.

**PTM.** To select PTMs for a PEAKS DB search, click the Set PTM... button to open the "PTM Options" dialogue. PEAKS DB search supports a limited number of variable PTMs. Refer to Section 2.3, "PTM" for details.

## Note

To prevent long search times, select only the most frequent PTMs in the sample for PEAKS DB, and check the PEAKS PTM option to search for unspecified modifications.

**Database.** Select a protein sequence database for the search. Select one from the list of databases that have been configured in PEAKS and set the taxonomy, if applicable. To configure a new sequence database, refer to Chapter 6, *Adding a Sequence Database*.

For searching only a few proteins, choose "Paste sequence" and then paste in the protein sequences.
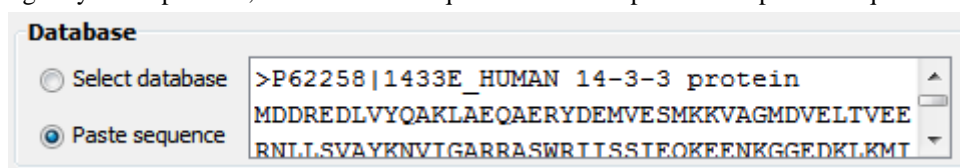


## Note

The pasted protein sequences should follow the FASTA format. Each sequence follows a description line which starts with '>'. PEAKS omits invalid characters, white spaces, and numbers in a sequence. A maximum of 1000 protein sequences can be pasted.

Select a contaminant database for the search from the list of databases that have been configured in PEAKS.

*De novo* **Tag Options.** PEAKS Search requires a *de novo* sequencing result to improve its search speed, sensitivity, and accuracy. By default, *de novo* sequencing runs on the fly with the current parameter settings. It is also possible to select an existing *de novo* sequencing result.

**Estimate FDR with decoy-fusion.** Select this option to enable search result validation with decoy-fusion, an enhanced target-decoy approach. When this option is selected, decoy sequences are automatically generated from the target database and are searched by PEAKS DB, PEAKS PTM, and SPIDER. This enables the estimation

of the false discovery rate (FDR) in the analysis report. It is NOT recommended to de-select this option unless manual validation of the result using another approach is preferred.

**Find unspecified PTMs and common mutations with PEAKS PTM.**    Select this option to enable a PEAKS PTM search after a PEAKS DB search finishes. PEAKS PTM analyzes spectra with good *de novo* sequences that remain unidentified by PEAKS DB. The default setting for PEAKS PTM is to search with all the built-in modifications in the "Common" and "Uncommon" lists, which include all of the natural modifications and mutations in the Unimod database.

Alternatively, PEAKS PTM can search with a list of preferred modifications. In "Advanced Settings", preferred modifications from the "PTM Options" dialogue can be chosen. PEAKS PTM allows an unlimited number of variable modifications to be searched, However, limiting the number can improve search accuracy.

Additionally, the maximum number of variable PTMs per peptide can be specified and a value of less than 4 is recommended. A threshold on the *de novo* ALC score can be specified so that only the spectra with good *de novo* sequences are analyzed by PEAKS PTM.

When PEAKS PTM search is enabled, a PEAKS PTM result node will be automatically generated after the search. In the PEAKS PTM report, search results from PEAKS DB and PEAKS PTM are merged.

## Note

Modifications selected for the PEAKS DB search are automatically added as preferred modifications to the PEAKS PTM search and cannot be removed. However, any number of additional preferred modifications can be added.



**Find more mutations with SPIDER.**    Select this option to enable a SPIDER search. SPIDER performs homology searches on spectra with good *de novo* sequences that remain unidentified by either PEAKS DB or PEAKS PTM. SPIDER identifies novel peptides that are homologous to peptides in the searched protein database. When SPIDER is enabled, a SPIDER result node will be automatically generated after the search. In the SPIDER report, search results from PEAKS DB, PEAKS PTM, and SPIDER are merged.

# 2.1. Saving the Parameters for Future Use

After setting up the search parameters, they can be saved for future use. Click the "Predefined parameters" drop-down menu, select the Save as ... command, and enter a name for the current set of parameters. Saved parameters are available for future analyses. A set of saved parameters can be loaded by selecting from the "Predefined parameters" drop-down menu.

**Note**

"Advanced Settings" for PEAKS PTM are not saved.

# 3. Understanding PEAKS Search Results

After PEAKS Search is complete, several result nodes are generated. One is the *de novo* sequencing result (assuming the default option for on-the-fly *de novo* sequencing is enabled). The other result nodes are identification results from the database search tools: PEAKS DB, PEAKS PTM, and SPIDER. If more than one database search tool is enabled, the results from previous search tools are automatically merged. Double-click the node to examine the analysis report. The analysis report is presented in five pages:

- **Summary**: This shows an outline of the PEAKS database search results with key statistics. The overall quality of the experiment can be examined and the filters for peptide and protein identifications can be adjusted.

- **Protein**: This shows a list of protein identifications. This view also visualizes protein sequence coverage and helps with protein characterization.

- **Peptide**: This shows a list of peptide identifications. This view also provides spectrum annotation and other detailed information for peptide-spectrum matches.

- **De novo only**: This shows a list of quality peptide sequences detected by *de novo* sequencing that remain unidentified by the PEAKS database search.

- **LC-MS**: This shows peptide-spectrum matches visualized in an interactive LC-MS heat map.

## 3.1. Peptide and Protein Scores

**PEAKS Peptide Score (-10lgP).**    The PEAKS peptide score (-10lgP) is calculated for every peptide-spectrum match (PSM) reported by PEAKS DB, PEAKS PTM, and SPIDER. The score is derived from the p-value that indicates the statistical significance of the peptide-spectrum match. A peptide may be matched to many spectra, resulting in multiple PSMs. In that case, the peptide's score is calculated as the maximum of all the PSMs. For details of the scoring algorithm, please refer to the publication, *"PEAKS DB: De Novo sequencing assisted database search for sensitive and accurate peptide identification" Mol Cell Proteomics. 2011 Dec 20.*

**PEAKS Protein Score (-10lgP).**    The PEAKS protein score (-10lgP) is calculated as the weighted sum of the -10lgP scores of the protein's supporting peptides. After removing any redundant peptides, the supporting peptides are sorted by -10lgP scores in descending order, and the k-th ranked peptide contributes to the weighted sum with a weight of 1/k.

## 3.2. Summary View

The "Summary" view reports key statistics as an overview of the result. It also provides functions to filter and export identified proteins and peptides:

1. **Result filtering:** Filtering rules are located at the top of the "Summary" view. Low quality peptide and protein identifications can be filtered by -10lgP score. Peptide sequences identified only by *de novo* sequencing can be filtered by ALC score. More details about how to properly set the filters are discussed in Section 4, "Filtering PEAKS Search Results".

2. **Result exporting:** Identified proteins and peptides can be exported to various formats by clicking the Export button at the top of the "Summary" view. More details are discussed in Section 5, "Exporting PEAKS Search Results".

In the report, several statistical charts can be examined to assess result quality. The report is composed of four sections:

1. **Notes:** In the header of the report, information such as a description of the experiment can be entered. The notes can be edited by clicking the Notes button.

2. **Result Statistics:** Key statistics are shown in the figures and tables. The figures provide important information about target-decoy result validation, which can be used to filter the results appropriately.

3. **Experiment Control:** Figures show the distribution of precursor mass error, revealing how well the instrument is calibrated. A table shows the statistics of missed cleavages in peptides, revealing enzyme digestion efficiency of the protein sample.

4. **Other Information:** Search parameters and MS instrument information are shown.

Statistical figures and tables in the report are discussed in the rest of this section.

## 3.2.1. False Discovery Rate (FDR) Curve

Figure 1 shows the FDR curve with respect to the number of PSMs being kept after filtration by the peptide -10lgP score. By lowering the score threshold, more PSMs are kept in the filtered result. Conversely, the FDR increases because more false positives are kept. In this figure, the vertical line indicates the current score threshold. The number of PSMs and the corresponding FDR by the current score threshold are shown in the top-left corner.
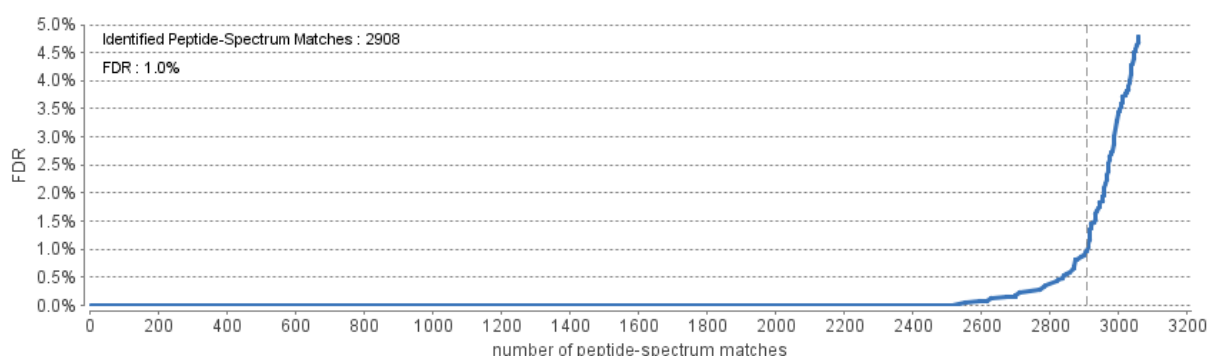


The FDR curve is available when FDR estimation is enabled in the PEAKS Search parameters. PEAKS Search estimates FDR using the "decoy-fusion" approach. Decoy-fusion is an enhanced target-decoy approach that makes more conservative FDR estimations. For details of the "decoy-fusion" approach, please refer to the publication, *"PEAKS DB: De Novo sequencing assisted database search for sensitive and accurate peptide identification" Mol Cell Proteomics. 2011 Dec 20.*

It is commonly recommended to set FDR at <1% by adjusting the score threshold. If a rapid growth of the FDR around the 1% FDR threshold is noticed, some PSMs may be sacrificed to significantly reduce the FDR.

> ## Note
>
> When counting the number of PSMs, PEAKS keeps one peptide per spectrum at most. Thus, the number of PSMs is actually the number of spectra with assigned peptides.

> ## Note
>
> Decoy matches are not counted in the number of PSMs. Unless otherwise specified, decoy matches are also excluded from the other statistical numbers shown in the "Summary" view.

## 3.2.2. PSM Score Distribution

Figures 2(a) and 2(b) in the "Summary" view show the peptide score distribution of PSMs. When FDR estimation is enabled in the PEAKS Search parameters, the target and decoy matches are respectively shown in two different colors.

Figure 2(a) is a stacked histogram showing the number of target and decoy matches in each score interval. There should be a similar number of target and decoy matches with low scores, but very few decoy matches with high scores. The vertical line indicates the current score threshold.
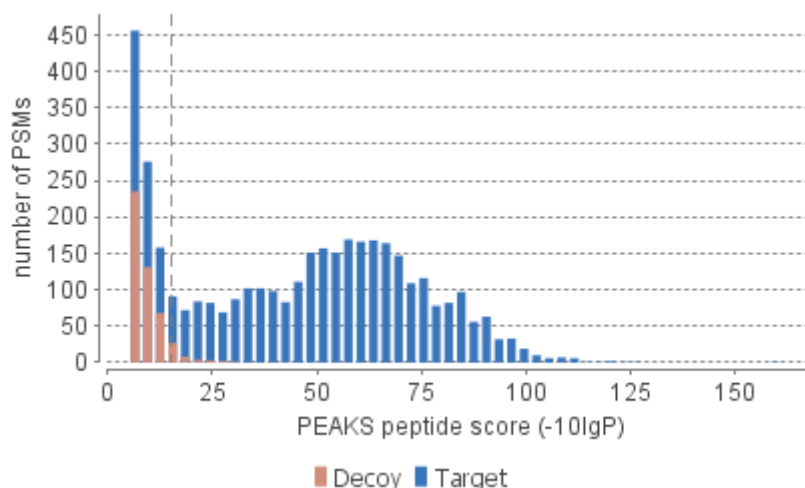
Figure 2(b) is a scatterplot showing the peptide score versus the precursor mass error in ppm for all PSMs. This figure is most useful for high resolution instruments. The PSMs with high scores should be centered near the mass error of 0, and the PSMs with low scores are usually scattered to larger mass error. The vertical dotted line indicates the current score threshold.
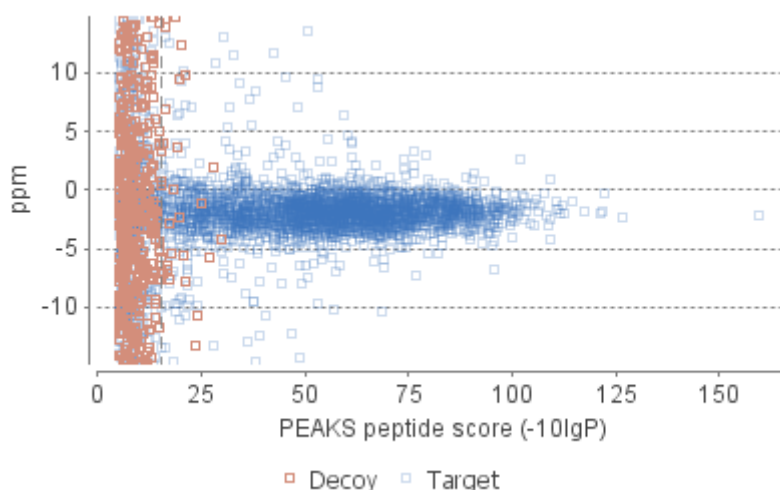


### 3.2.3. De Novo Result Validation

Figures 3(a) and 3(b) in the "Summary" view help to determine a proper score threshold to filter "*de novo* only" sequences.

After the database search, *de novo* sequences can be categorized as:

1. **Verifiable de novo sequences:** A *de novo* sequence is verifiable if the associated MS/MS spectrum is confidently matched to a database peptide. Residues in a verifiable *de novo* sequence can be validated using the database peptide as a reference.

2. **"De novo only" sequences:** A *de novo* sequence is "*de novo* only" if the associated MS/MS spectrum is not confidently matched to any database peptide. "*De novo* only" sequences may suggest novel peptides, peptides with unknown modifications, or other interesting research subjects.

> **Note**
>
> An MS/MS spectrum is considered to be confidently matched to a database peptide if there are PSMs to the spectrum with peptide -10lgP scores above the current peptide filtering threshold.

"*De novo* only" sequences are crucial for a complete proteomic analysis. However, it is often necessary to remove the low quality sequences. "*De novo* only" peptides can be filtered by *de novo* ALC score, which is the average local confidence score of residues in the *de novo* sequence.

Figures 3(a) and 3(b) provide a guideline for determining a proper ALC score threshold. The two figures show the local confidence score distribution of residues in *de novo* sequences that are filtered by the current ALC score threshold.

Figure 3(a) shows the score distribution of residues in verifiable *de novo* sequences. These residues are validated by aligning the *de novo* sequence with the database peptide. A residue is considered correct if it is consistent with the database peptide. Otherwise, the residue is considered incorrect. The figure shows the score distributions of correct residues and incorrect residues in two different colors.

Figure 3(b) shows the score distribution of residues in "*de novo* only" sequences. As these residues cannot be directly validated using database peptides, their ratios of correctness are statistically estimated using the distributions in Figure 3(a). The figure shows the estimated score distributions of correct residues and incorrect residues in two different colors.

As a guideline, the ALC score threshold should be increased gradually until the score distributions of correct and incorrect residues are similar, as seen in figures 3(a) and 3(b). In the following example, the ALC threshold is gradually increased to 80. This ensures that the filtered "*de novo* only" sequences are generated from MS/MS spectra that have the same spectral quality as the MS/MS spectra confidently matched in the database search. Figure 3(b) also allows estimation of the proportion of incorrect residues in the filtered "*de novo* only" sequences.
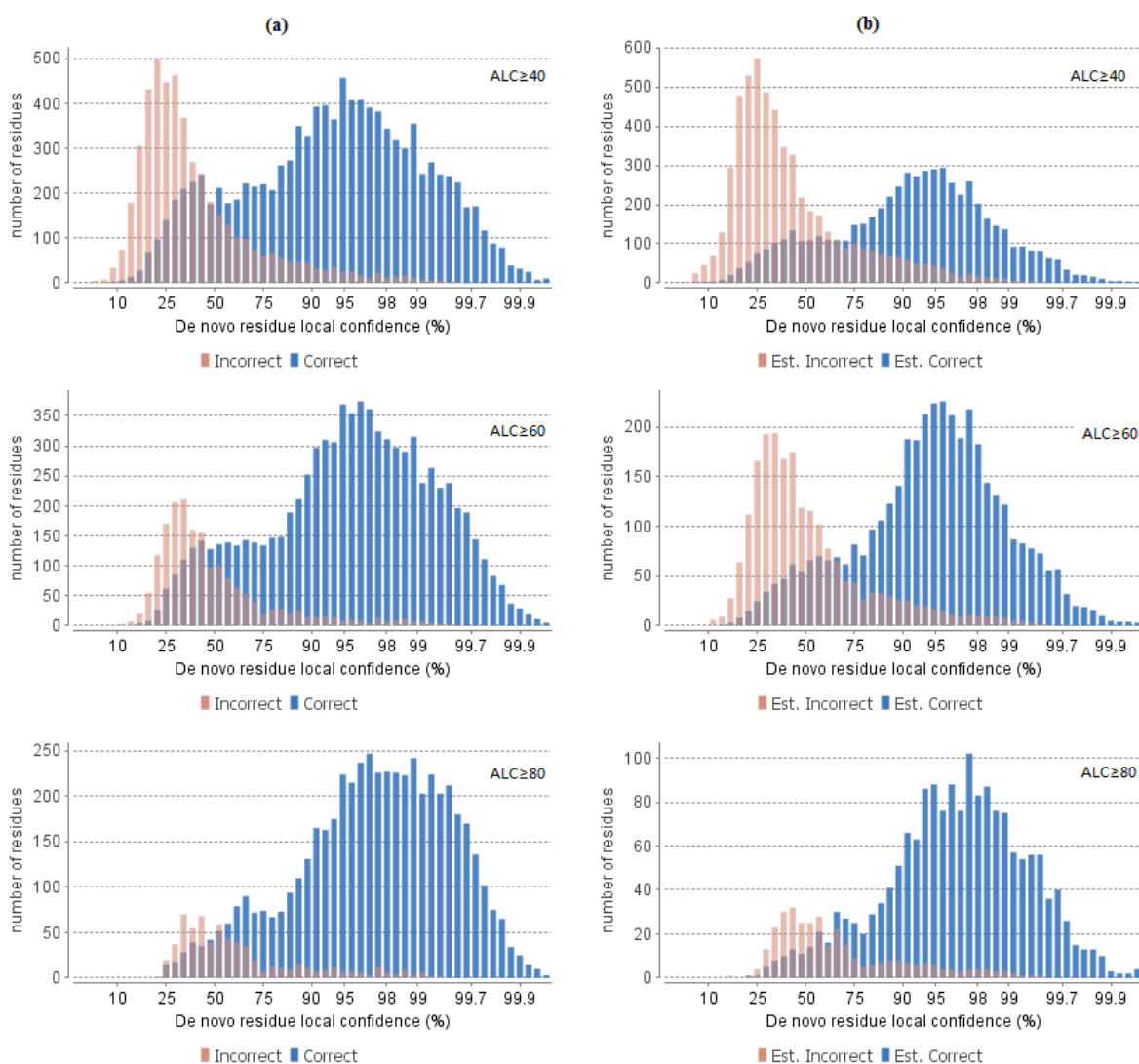
## 3.2.4. Statistics of Data and Results

Tables 1-4 show the statistical numbers of the data and results.

**Table 1.** Statistics of data.

| | |
|---|---|
| # of MS Scans | 4751 |
| # of MS/MS Scans | 5152 |

**Table 2.** Result filtration parameters.

| | |
|---|---|
| Peptide -10lgP | ⩾15.2 |
| Protein -10lgP | ⩾20 |
| Proteins unique peptides | ⩾0 |
| De novo ALC Score | ⩾50% |

**Table 3.** Statistics of filtered result.

| | |
|---|---|
| Peptide-Spectrum Matches | 2908 |
| Peptide Sequences | 1025 |
| Protein Groups | 27 |
| Proteins | 27 |
| Proteins (#Unique Peptides) | 6 (>2); 5 (=2); 6 (=1); |
| FDR (Peptide-Spectrum Matches) | 1.0% |
| FDR (Peptide Sequences) | 2.4% |
| FDR (Protein) | 58.8% |
| De Novo Only Spectra | 165 |

**Table 4.** PTM Profile.

| Name | ΔMass | #PSM | Position |
|---|---|---|---|
| Carbamidomethyl | 57.02 | 1548 | C |
| Carbamidomethyl | 57.02 | 585 | DEHK,N-term |
| Deamidation | .98 | 102 | NQ |
| Dehydration | -18.01 | 57 | DSTY |
| Dethiomethyl | -48.00 | 30 | M |
| Carbamylation | 43.01 | 29 | K,N-term |
| Oxidation | 15.99 | 24 | M |
| Sodium | 21.98 | 22 | DE,C-term |
| Ammonia loss | -17.03 | 21 | N |
| Pyro-glu from Q | -17.03 | 14 | N-term |
| Acetylation | 42.01 | 10 | N-term |
| Ubiquitin | 114.04 | 9 | ST |
| Hex | 162.05 | 9 | KT,N-term |
| Dihydroxy | 31.99 | 8 | KPW |
| Acetylation | 42.01 | 6 | K |
| Carboxymethyl | 58.01 | 5 | N-term |
| Formylation | 27.99 | 5 | K,N-term |
| Methyl ester | 14.02 | 4 | DE,C-term |
| Phosphorylation | 79.97 | 4 | SY |
| Glu | 129.04 | 4 | E |
| Pyro-glu from E | -18.01 | 4 | N-term |
| Carboxylation | 43.99 | 4 | DK |

Most entries in these tables are self-explanatory. The following entries are explained:

- **Peptide Sequences** (Table 3): This represents the number of distinct peptides in the filtered result. Peptides with the same primary sequence but different PTMs are counted separately. However, if several peptides differentiated only by leucine (L) and isoleucine (I), they are counted as one peptide. Since the same peptide may be identified from multiple spectra due to data redundancy and differing charge states, this number is usually smaller than the number of peptide-spectrum matches.

- **Protein Groups** (Table 3): PEAKS groups proteins identified by a common set of peptides. This number shows the number of protein groups in the filtered result.

- **Proteins (#Unique Peptides)** (Table 3): This entry counts the number of proteins identified with unique peptides. A unique peptide is a peptide with a -10lgP score above the peptide filtering threshold that can be mapped to only one protein group.

- **PTM Profile** (Table 4): The table shows modifications identified in the filtered result and the number of PSMs containing each modification.

## 3.2.5. Experiment Control

Figures 4(a) and 4(b) in the "Summary" view show the precursor mass errors of PSMs in filtered results. Figure 4(a) shows the distribution of the precursor mass errors in a histogram. Figure 4(b) shows precursor m/z versus precursor mass error in a scatter plot. The precursor mass error is calculated in ppm as $10^6 \times$ (precursor mass - peptide mass) / peptide mass. Figures 4(a) and 4(b) help examine whether or not the mass spectrometer is properly calibrated. For a well-calibrated instrument, precursor mass errors should center at 0 ppm across the range of m/z.
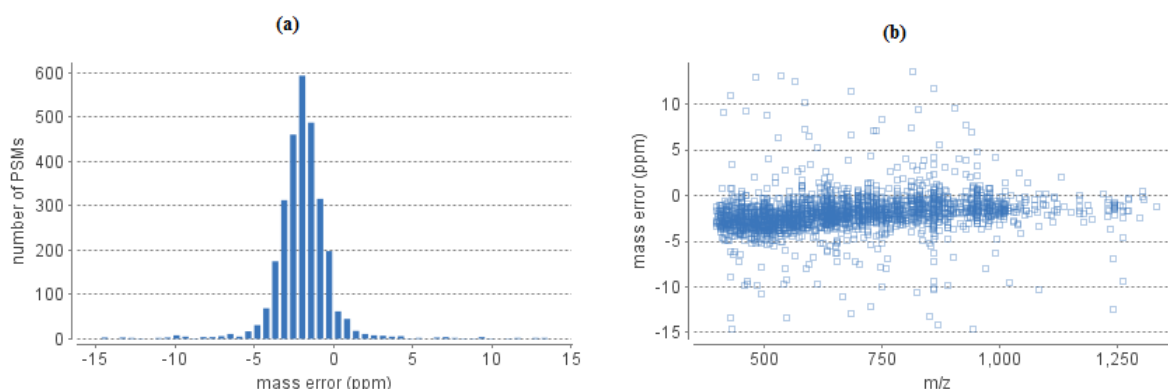


Table 5 shows the number of identified peptides in each sample with the number of missed cleavages, indicating the enzyme digestion efficiency.

**Table 5.** Number of peptides by number of missed cleavages

| Missed Cleavages | 0 | 1 | 2 | 3 | 4+ |
|---|---|---|---|---|---|
| Sample 1 | 78 | 56 | 23 | 16 | 0 |
| Sample 2 | 62 | 38 | 15 | 4 | 0 |
| Sample 3 | 107 | 55 | 15 | 6 | 0 |

# 3.3. Protein View

The "Protein" view lists protein identifications that have been filtered by the current settings in Summary view. It also visualizes the protein sequence coverage for identified proteins. The view has four components:

- **Protein table:** This lists protein identifications that have been filtered by the current settings in the "Summary" view

- **Coverage:** This visualizes the protein sequence coverage of the selected protein, showing the mapping of supporting peptides and *de novo* tags to the protein sequence.

- **Peptides:** This lists supporting peptides mapped to the selected protein.

- **De novo Tags:** This lists "*de novo* only" peptides mapped to the selected protein.
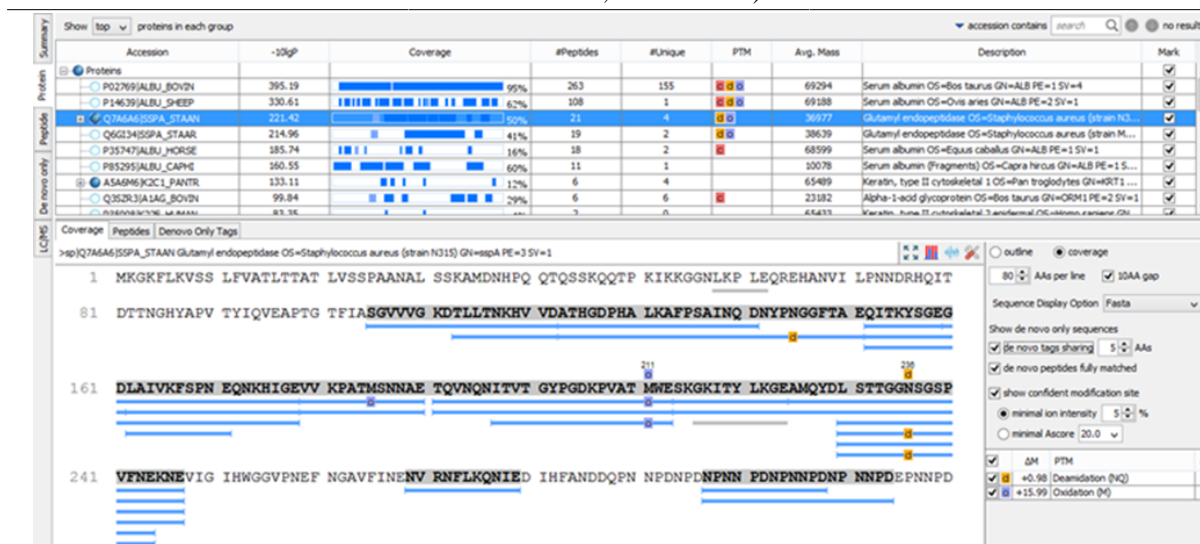
## 3.3.1. Protein Table

Each row in the table represents a group of proteins that are supported by a common set of peptides. A dark blue node indicates that the group has multiple proteins. To expand the group, click the "+" button at the left. The Show drop-down menu, located above the Protein table, specifies the way the proteins in each group are displayed. "All" shows every protein in the group, "Top" shows the top proteins in the group determined by the number of high-confidence supporting peptides, and "First" shows only the first of the top proteins in the group.

### Note

A supporting peptide is considered highly confident if it has a peptide -10lgP score above the set threshold.

The columns in the Protein table are:

- **Accession:** The accession number of the protein as seen in the FASTA database.

- **-10lgP:** The protein confidence score.

- **Coverage:** The percentage of the protein sequence that is covered by supporting peptides. The coverage is visualized by a colored bar. Light blue blocks indicate the parts of the sequence covered by low-confidence peptides. Dark blue blocks indicate the parts covered by high-confidence peptides.

- **#Peptides:** The number of high-confidence supporting peptides.

- **#Unique:** The number of high-confidence supporting peptides that are mapped to only one protein group. Unique peptides with same sequence but different modifications are only counted once in this number.

- **PTM:** The identified modifications displayed with color-coded icons.

- **Avg. Mass:** The protein mass calculated using the average mass.

- **Description:** The protein's header information as seen in the FASTA database.

- **Mark:** This allows the selection of specific proteins. The marked proteins are exported when using the exporting functions in the "Summary" view.

### Note

For #Peptides and #Unique, two peptides with the same starting and ending positions in the protein are counted as one, regardless of their PTM forms. This is to follow the MCP (Molecular & Cellular Proteomics) guidelines.

## 3.3.2. Protein Coverage

The Protein Coverage view visualizes the mapping of supporting peptides and *de novo* tags to the protein selected in the Protein table. It also shows all identified sites with modifications or mutations to assist with protein characterization at the amino acid level. The coverage view has three major components:

- **Protein sequence coverage.** This area visualizes the coverage of the protein sequence.

  Regions in the protein sequence that are covered by supporting peptides are displayed in bold font with a grey background. Confident modifications and mutations identified in supporting peptides are displayed as icons above the protein sequence. Modifications are represented by colored icons with the initial letter of its modification name. If a residue is modified by more than one modification in the same supporting peptide, "*" is used instead of a letter. Mutations are represented by framed white icons with a letter indicating the post-modification residue. Placing the cursor over a mutation icon shows the full name of the modification/mutation. The number above a mutation icon indicates the position in the protein sequence. The header information of the protein is shown on the top of the protein sequence.

  The supporting peptides and matched *de novo* tags can be shown as colored bars under the protein sequence. Placing the cursor over a bar shows detailed information of the peptide. Left clicking on a bar shows the annotated spectrum from which the supporting peptide or *de novo* tag is identified. Right clicking on a bar shows a pop-up menu for quick operations, including "Remove peptide" which hides the peptide from the protein coverage view. Hidden peptides can be restored by re-selecting them in the supporting peptide table or the *de novo* tag table.



- **Control panel.** This panel controls what to display in the protein sequence coverage.

- 

○ outline    ◉ coverage

80 ⬍  AAs per line    ☑ 10AA gap

Sequence Display Option  Fasta  ⌄

Show de novo only sequences

☑ de novo tags sharing  5 ⬍ AAs

☑ de novo peptides fully matched

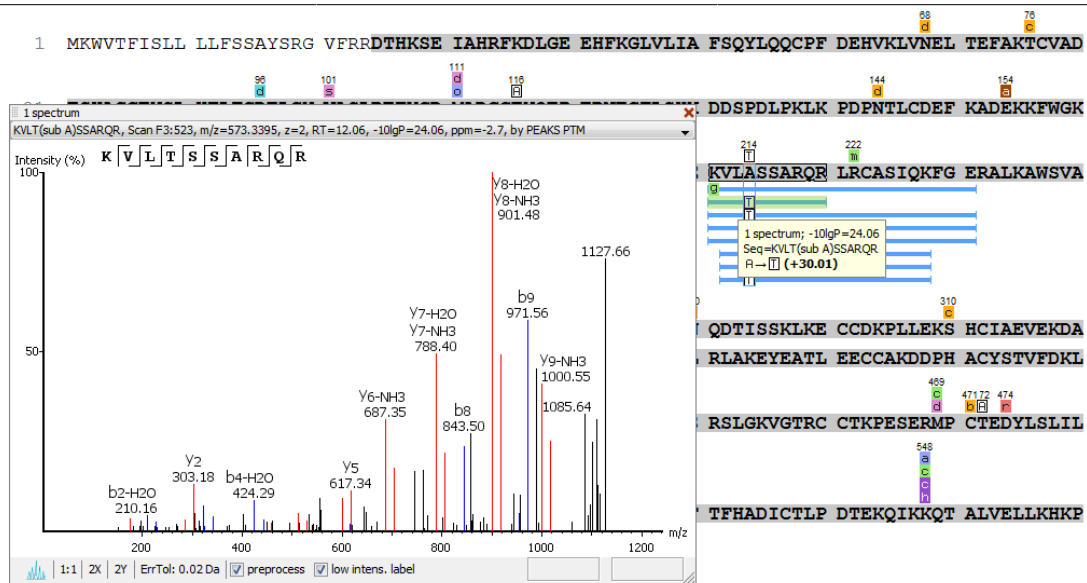☑ show confident modification site

◉ minimal ion intensity  5 ⬍ %

○ minimal Ascore  20.0 ⌄

| ☑ | | ΔM | PTM | # | |
|---|---|---|---|---|---|
| ☑ | c | +57.02 | Carbamidomethylation | 837 | ⌃ |
| ☑ | c | +57.02 | Carbamidomethylation (... | 400 | |
| ☑ | d | +0.98 | Deamidation (NQ) | 37 | |
| ☑ | d | -18.01 | Dehydration | 37 | |
| ☑ | c | +43.01 | Carbamylation | 20 | |
| ☑ | d | -48.00 | Dethiomethyl | 20 | |
| ☑ | o | +15.99 | Oxidation (M) | 9 | |
| ☑ | u | 114.04 | Ubiquitin | 6 | |
| ☑ | a | -17.03 | Ammonia-loss (N) | 6 | |
| ☑ | T | +30.01 | A->T | 6 | |
| ☑ | p | -17.03 | Pyro-glu from Q | 5 | |
| ☑ | h | 162.05 | Hexose | 5 | |
| ☑ | r | +37.95 | Replacement of 2 proto... | 4 | |
| ☑ | A | -30.01 | T->A | 4 | |
| ☑ | s | +21.98 | Sodium adduct | 3 | |
| ☑ | c | +58.01 | Carboxymethyl (KW, X... | 3 | |
| ☑ | d | +31.99 | Dihydroxy | 3 | |
| ☑ | G | -58.01 | D->G | 3 | |
| ☑ | a | +42.01 | Acetylation (N-term) | 2 | |
| ☑ | c | +58.01 | Carboxymethyl | 2 | |
| ☑ | f | +27.99 | Formylation | 2 | |
| ☑ | m | +14.02 | Methyl ester | 2 | |
| ☑ | c | +43.99 | Carboxylation (DKW) | 2 | |
| ☑ | m | 129.04 | Monoglutamyl | 2 | |
| ☑ | E | +58.01 | A->E | 2 | ⌄ |

**Mode control.** The protein sequence can be displayed in either "outline" or "coverage" mode.

- **Outline mode:** In outline mode, the protein sequence is displayed without supporting peptides and matched *de novo* tags. However, the sequence coverage at a specific position can be examined by left-clicking a residue.

```
  1   MKWVTFISLL LLFSSAYSRG VFRRDTHKSE IAHRFKDLGE EHFKGLVLIA FSQYLQQCPF DEHVKLVNEL TEFAKTCVAD
```

1 spectrum                                                                        ✕

KVLT(sub A)SSARQR, Scan F3:523, m/z=573.3395, z=2, RT=12.06, -10lgP=24.06, ppm=-2.7, by PEAKS PTM   ▼

Intensity (%)   K V L T S S A R Q R

DDSPDLPKLK PDPNTLCDEF KADEKKFWGK

KVLASSARQR LRCASIQKFG ERALKAWSVA

1 spectrum; -10lgP=24.06
Seq=KVLT(sub A)SSARQR
A → T (+30.01)

Y8-H2O
Y8-NH3
901.48

1127.66

Y7-H2O
Y7-NH3
788.40

b9
971.56

Y9-NH3
1000.55

Y6-NH3
687.35

b8
843.50

1085.64

Y2
303.18   b4-H2O
424.29

Y5
617.34

b2-H2O
210.16

m/z

200      400      600      800      1000      1200

1:1  2X  2Y  ErrTol: 0.02 Da  ☑ preprocess  ☑ low intens. label

QDTISSKLKE CCDKPLLEKS HCIAEVEKDA

RLAKEYEATL EECCAKDDPH ACYSTVFDKL

RSLGKVGTRC CTKPESERMP CTEDYLSLIL

TFHADICTLP DTEKQIKKQT ALVELLKHKP

- **Coverage mode:** In coverage mode, all the supporting peptides and matched *de novo* tags are shown.

```
  1   MKWVTFISLL LLFSSAYSRG VFRRDTHKSE IAHRFKDLGE EHFKGLVLIA FSQYLQQCPF DEHVKLVNEL TEFAKTCVAD
```

```
 81   ESHAGCEKSL HTLFGDELCK VASLRETYGD MADCCEKQEP ERNECFLSHK DDSPDLPKLK PDPNTLCDEF KADEKKFWGK
```

- **AAs per line and 10AA gap.** Specify the number of amino acids per line and whether to show a gap for every 10 amino acids in the protein sequence.

- **Sequence Display Option.** In a multi-enzyme experiment, select an enzyme that was chosen during either the project creation or the PEAKS search stage. The coverage map will be highlighted based on the peptides generated by the selected enzyme.

- **de novo tags sharing X AAs.** Show *de novo* tags with at least X consecutive amino acids matched in the protein sequence.

- **de novo peptides fully matched.** Show *de novo* tags with their full sequences matched in the protein sequence. This control is especially useful for showing short but fully matched *de novo* peptides.

- **modification minimal ion intensity.** A modification or mutation is considered confident if there are fragment ions supporting the modified residue with the minimal peak intensity above the specified threshold. Only confident modifications and mutations are displayed as icons above the protein sequence

- **modification table.** The modification table shows the modifications identified in supporting peptides on the protein. For each modification, the number of supporting PSMs with this particular modification is shown. The checkbox on the left controls whether to indicate the modification in the protein coverage. Double-clicking on the name of a modification shows detailed information about the modification. Right clicking a modification shows a pop-up menu for quick operations.

- **Tools bar.** The tools bar is at the upper-right corner of the protein sequence coverage. It has two icons:
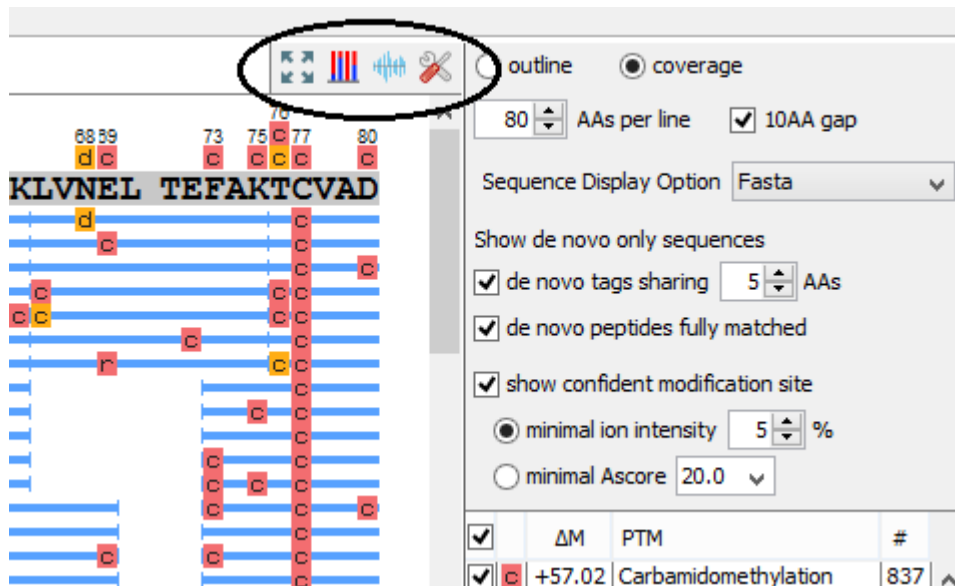


- **Full screen** ⁂. Left-clicking this icon shows the protein coverage in full screen mode. Click ⁂ to return from full screen mode.

- **Tool box** ✂. The tool box contains the following tools:

  - Copy template protein. This copies the original database protein into the system clipboard.

  - Copy mutated protein. This copies the protein with detected mutations into the system clipboard.

  - Save protein coverage as image.

  - Coverage statistics.

  - NCBI BLAST search.

  - NCBI Entrez search.

  - Multiple sequence alignment. Performs multiple sequence alignment for the selected proteins.

- **PTM Profiling** ▥ Select this button to examine the characteristics of modifications that have been confidently identified based on the options selected under "show confident modification sites". See Section 3.3.3 for more details.

- **Peptide Mapping** ▥ Select this button to examine the detailed ion signal information generated for peptides between compared samples. See Section 3.3.4 for more details.
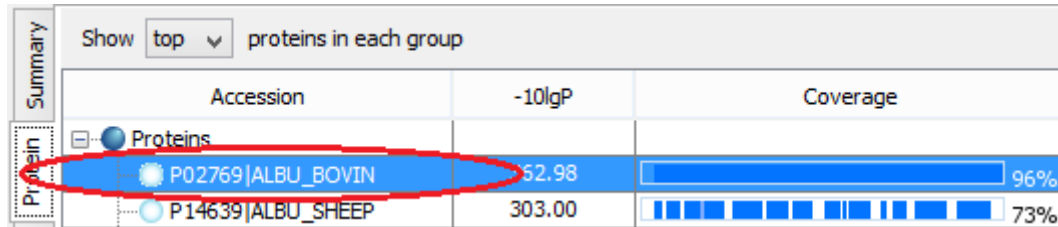
## 3.3.3. PTM Profiling

The PTM Profile tool provides the ability to compare and quantify detected modification sites on proteins in relation to unmodified sites. In many cases, a PTM only occurs on a portion of the sample, leaving the remainder unmodified. With this feature, PEAKS determines the protein's modification changes by comparing the area from

the LC-MS of the confidently modified peptide feature to the area from the associated unmodified (native) or non-confidently modified peptide feature, as well as to the area from peptide features modified by other modifications.

Follow the steps mentioned below to run the PTM profiling tool.

1. Select a protein of interest from the Protein table. By default, the first entry in the Protein table is selected
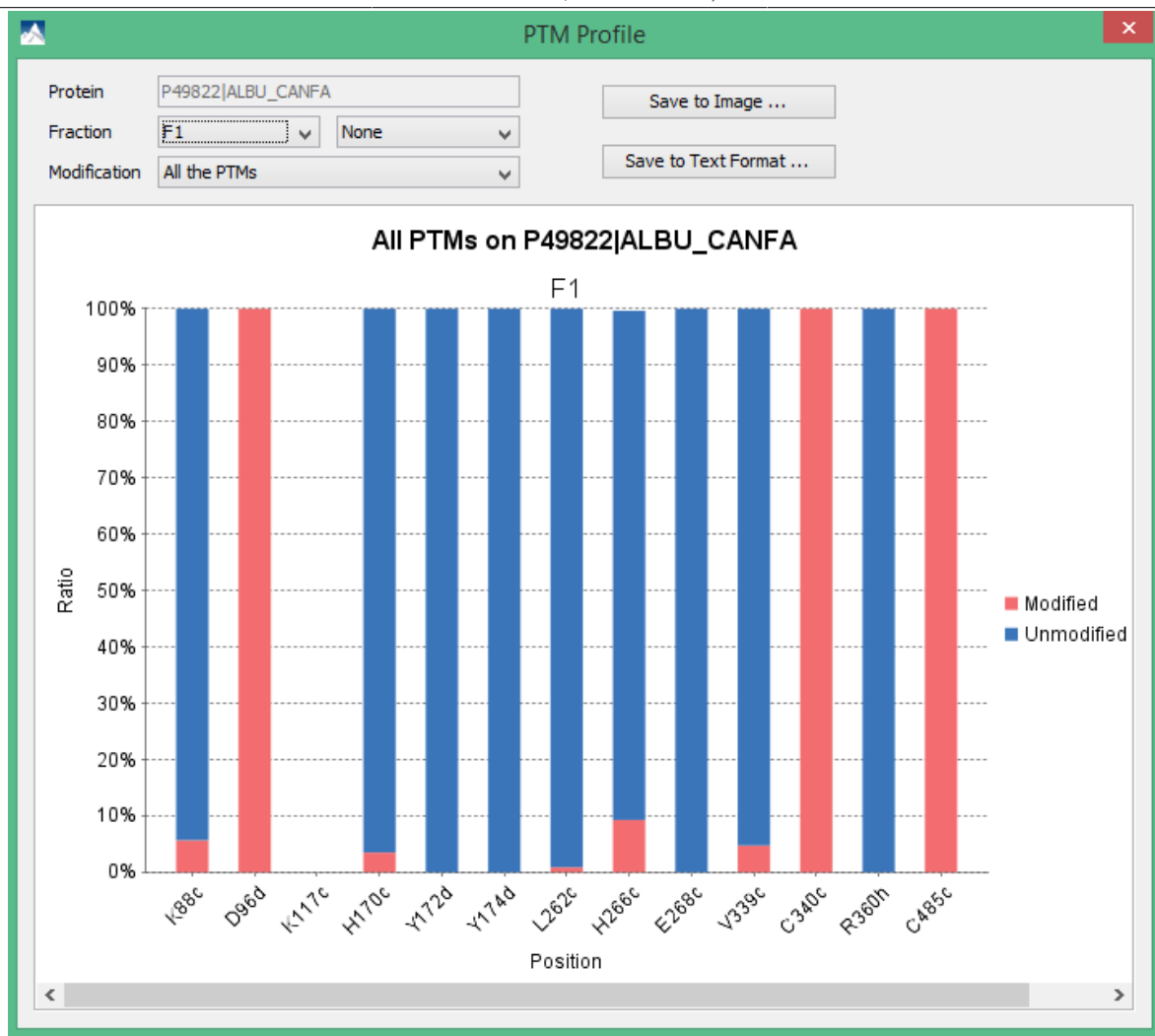


2. Set the "minimal ion intensity" or the "minimal AScore" parameter threshold for confident modification site determination. Refer to the relevant description in Section 3.3.2, "Protein Coverage" for details.
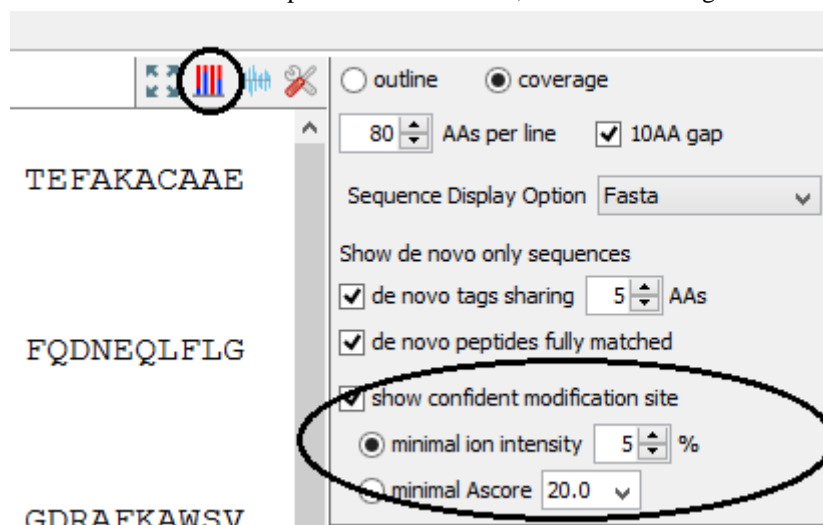


3. Click the ▥ button on the coverage pane tools bar. Once selected, a new dialogue will appear to display the PTM profiling.

The input selection, result display and result exporting options are combined together in the PTM profile dialog.

- **Inputs.** The inputs for PTM profiling can be set using the drop-down menus at the top-left corner of the dialogue.

  - **Protein.** The accession ID of the selected protein to be analyzed for PTM profiling.

  ## Note

  The selected protein cannot be changed within the dialogue. To run profiling on another protein, close the dialogue, select the protein of interest from the Protein table and run the profiling by clicking the ▥ button.

  - **Fraction.** The data source used for the analysis. The "F#" represents the fraction from which the protein was identified, where # refers to the ID of the fraction, as indicated in the fraction node of the project tree.
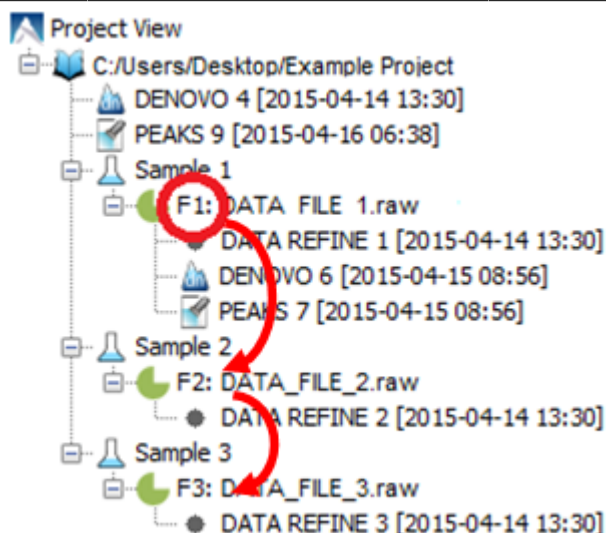
# Note

When the search result involves multiple fractions, a comparison of PTM profiles among the fractions can be performed by selecting the second fraction from the second fraction selection drop down menu.

- **Modification.** The PTM being profiled.

- **PTM Profile Chart:** The chart gives a visualization of the ratio between unmodified and modified amino acids at each position. Only the fully enzyme-digested peptides will be used to calculate the modification abundance. The x-axis indicates the position in the protein sequence where a modification was identified while the y-axis indicates the percentage (%) of modified or unmodified peptide features relative to the total abundance. A blue bar represents the unmodified portion and the red bar represents the modified portion. For some position, except for the specified modification, some other modification may exist. Thus, the combined modified and unmodified portions do not sum up to 100%.

- **Result Export.** The PTM profile can be exported as an image, as well as in text format.

  - **Image export.** Click the Save to Image... button to save the displayed graph as a PNG file.

  - **Text format export.** Click the Save to Text Format... button to export the calculated PTM Profile data to a text file in CSV format. The PTM Profile of the selected modification on a chosen protein may be exported for a specific fraction, or across all fractions. In addition, profiles of all proteins for the selected modification may also be exported; again, for either a specific fraction or across all fractions. The exported file contains all of the necessary information to construct the PTM profile graph. The PTM profile data are grouped together by fraction, then by protein, then by modification, and then by the position of the modification on the protein sequence. The profile information for the same peptide is combined together if they belong to the same group, as mentioned earlier. The headers of the CSV column are self-explanatory. The following items are explained for clarity:

    - **Peptide position:** The position of the modified amino acid in the peptide sequence.

    - **Unique:** This specifies that the peptide is unique (Y) or not unique (N).

    - **Total Area:** The total area of the features associated with the corresponding peptide.

    - **Modified Area**: The total area of the features associated with the modified version of the corresponding peptide.

    - **Unmodified Area:** The total area of the features associated with the unmodified version of the corresponding peptide.
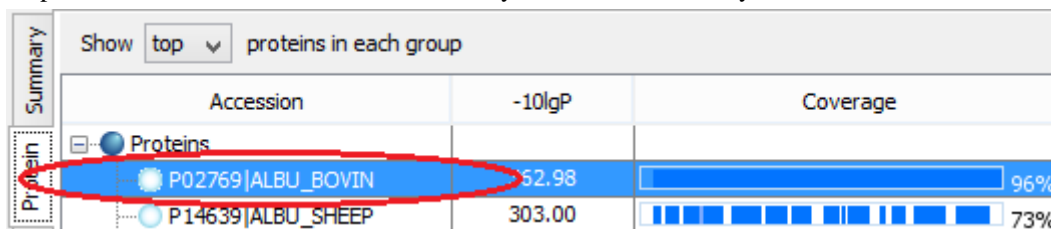
**Note**

The scans, the m/z ratios, the RTs, the charges, and the -10lgP values are all from the best PSM for the corresponding peptide.

## 3.3.4. Peptide Mapping

A peptide map is often referred to as a "fingerprint" of a protein. It is a powerful tool that provides a comprehensive understanding of the protein being analyzed. It gives a detailed picture of the ion signal of all peptides associated with the protein. In addition to generating the peptide map, PEAKS has integrated an annotated tool-tip which allows for easy identification of the peptide shown at any point on the map. If multiple data files are searched together in the analysis, then multiple peptide maps may be compared to one another. This allows users to easily determine the consistency of the protein between samples.
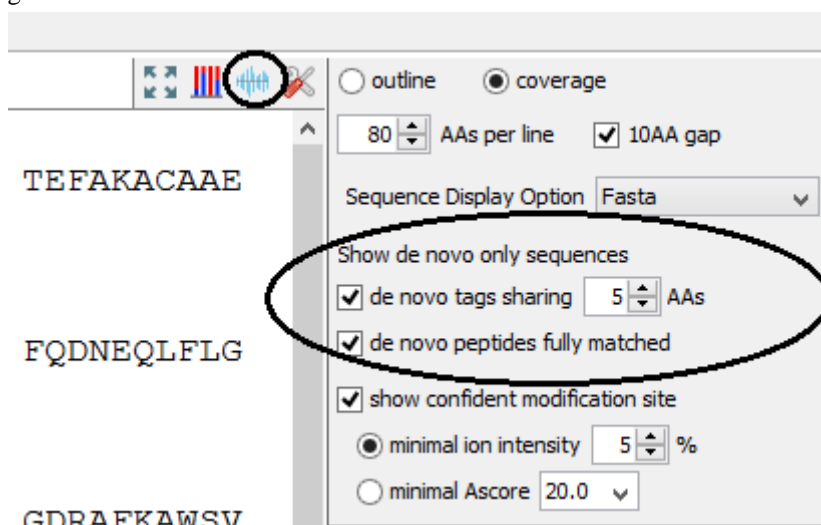
Follow the steps mentioned below to generate the peptide map:

1.  Select a protein of interest from the Protein table. By default, the first entry in the Protein table is selected.



2.  Select and set the desired *de novo* only peptide selection filters. Refer to relevant section of Section 3.3.2, "Protein Coverage" for details.



3.  Click the button on the coverage pane tools bar. Once selected, a new dialogue will appear to display the peptide mapping.
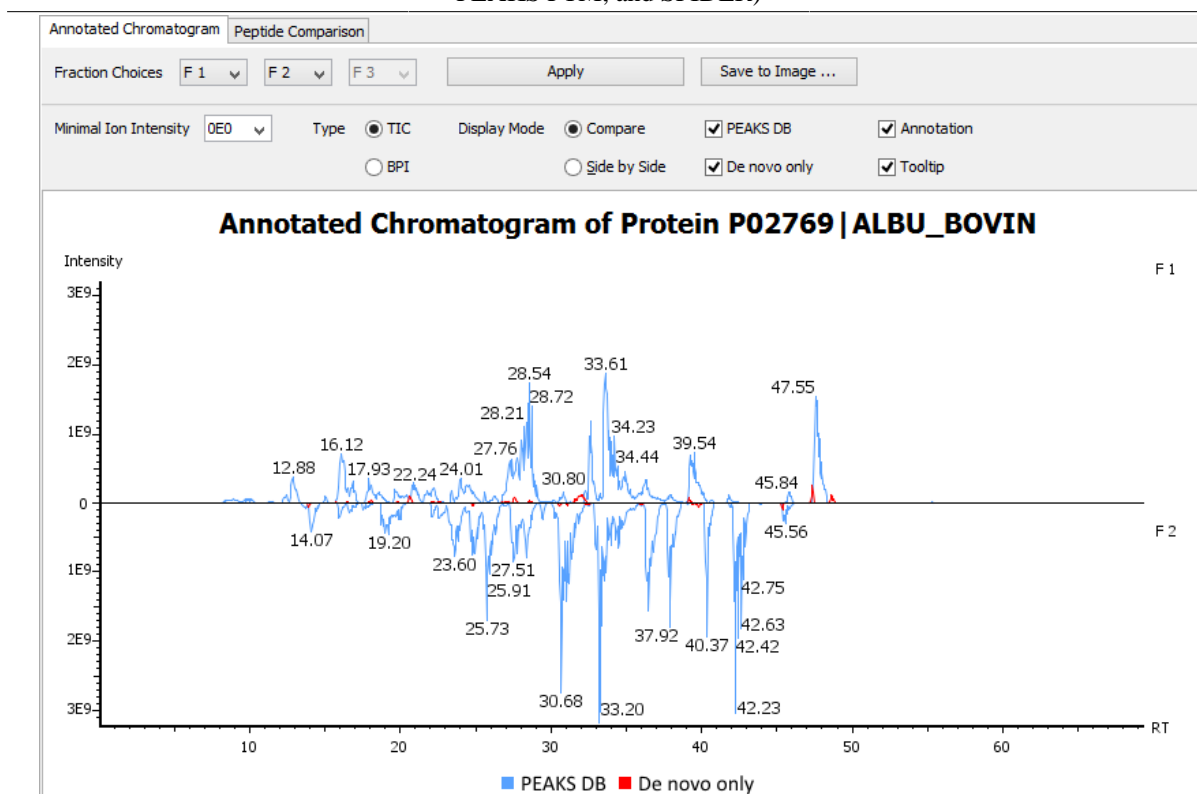
PEAKS provides two different approaches to analyze peptide mapping of a protein.

- **Annotated Chromatogram:** The annotated chromatogram of the selected protein is displayed in profile mode. The intensity of each peptide feature associated to the protein is mapped to its retention time (RT). The annotated chromatogram has its own input selection, result display, and result exporting options.

  - **Fraction Choices.** The data source used for peptide mapping. The "F#" represents the fraction to which the protein belongs, where # refers to the ID of the fraction, as indicated in the fraction node of the project tree

    ## Note

    When the search result involves multiple fractions, a comparison of annotated chromatograms for the selected protein among (a maximum of three) fractions, can be done by selecting the second and third fraction from the corresponding drop down menu.

  - **Minimal Ion Intensity:** A filter used to set an intensity threshold, where peptide features below this value will not be included in the display.

  - **Type - TIC or BPI:** Configures the display to view the Total Ion Chromatogram (TIC) or Base Peak Intensity (BPI).

  - **Display Mode:** PEAKS offers two modes to view multiple chromatograms for better comparison.

    - **Compare:** In compare mode, the protein chromatograms for the specified data samples will appear on opposite sides of the x-axis. This allows the user to directly compare the protein chromatograms between two samples.

    - **Side by Side:** In Side-by-Side mode, the chromatograms are displayed individually; each with their own set of x- and y-axes. This display mode allows for the comparison of up to three samples.

  - **PEAKS DB:** Peptide features associated with peptides identified by database searching are highlighted in the annotated chromatogram.

- **De novo only:** This highlights the peptide features associated with *de novo* only peptides (see Section 3.2.3, "De Novo Result Validation").

## Note

*De novo only* peptides displayed in the peptide map may be filtered using the de novo only peptide selection filters in the control panel of the coverage pane.

- **Annotation:** When selected, the retention time of major peptide features are displayed on the peptide map.

- **Tool Tip:** When selected, a pop-up will appear when the cursor hovers over any peptide within the peptide map. The pop-up contains the retention time, the intensity (TIC or base peak), and the peptide sequences found at that particular retention time.
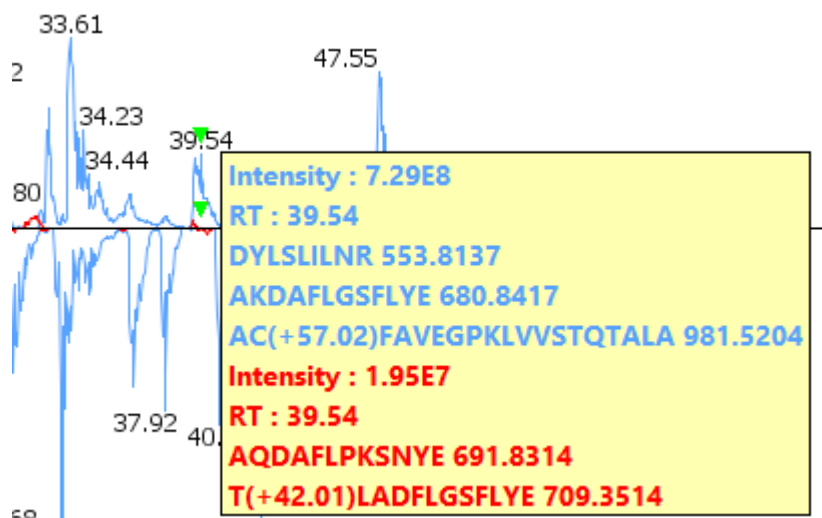


- **Annotated Chromatogram Chart:** This chart gives a visualization of the peptide map for the selected data. It provides a few convenient ways to zoom-in and navigate the chromatogram

  - Zoom to an RT region - Click the desired starting retention time, drag horizontally to the desired final retention time, and then release the mouse button.

  - Zoom in/out smoothly - Place the cursor at a particular RT value (right below the x-axis line), scroll the mouse wheel button.

  - Increase/decrease the ion intensity - Place the cursor on the chromatogram.

  - See the whole chromatogram - Double-click on the chromatogram.

  - Pan to an RT region: Place the cursor on the chromatogram, right-click, and then hold and drag to pan to an RT region.

- **Result Export.** The annotated chromatogram chart can be exported as an image in PNG format. Click the Save to Image... button to save the displayed chart.

- **Peptide Comparison:** Another option PEAKS offers its users to analyze the peptide map is in terms of a peptide comparison view. In this view, the total area of the features associated with a peptide is mapped to the mass of the peptide. The peptide comparison has its own input selection, result display, and result exporting options.
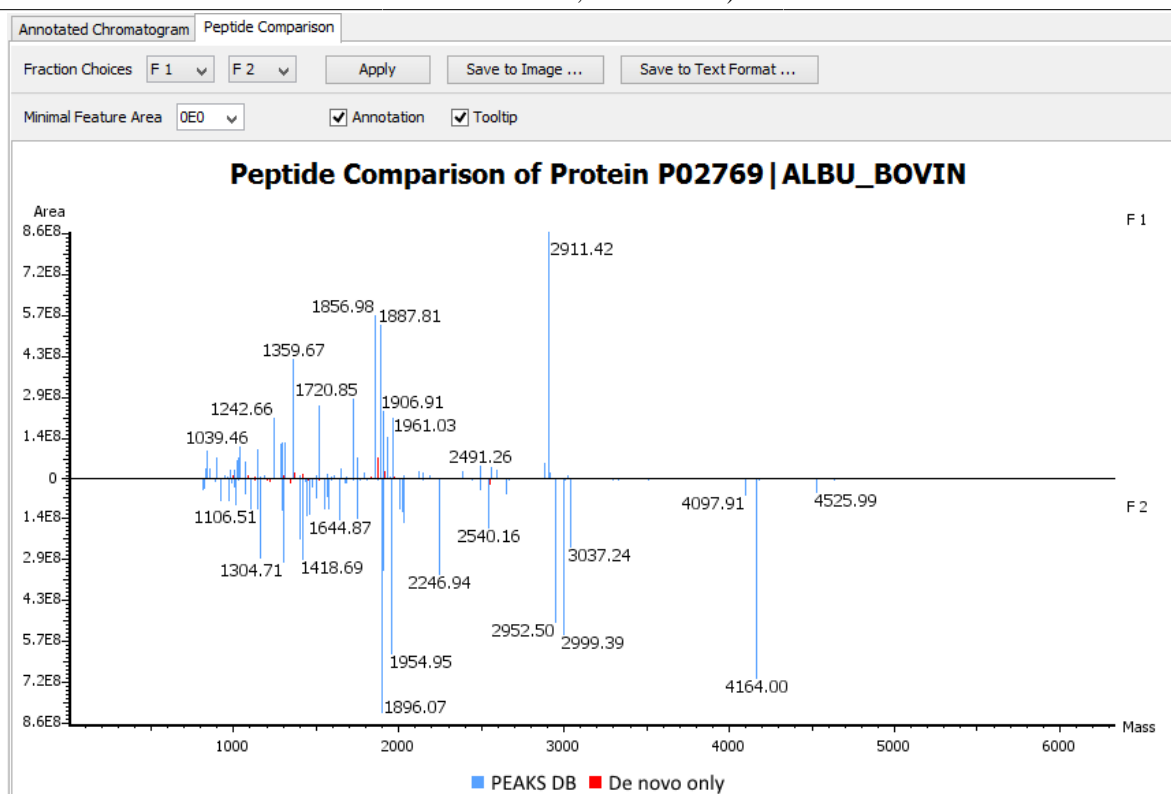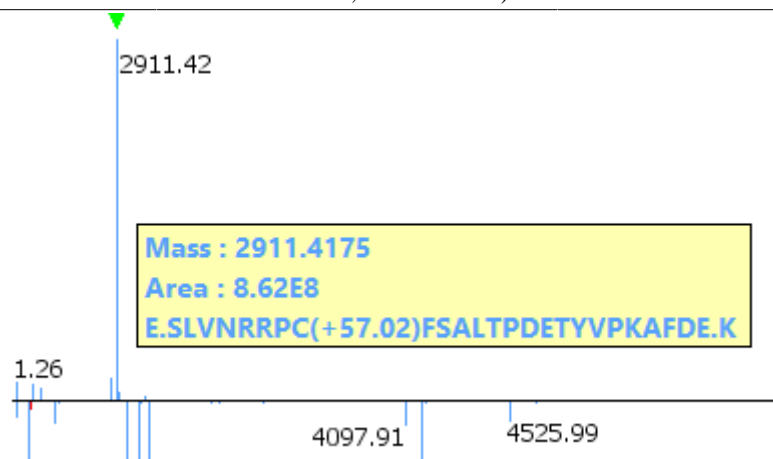
- **Inputs.** The inputs for the peptide comparison can be set using the drop-down menus at the top-left corner of the dialogue.

  - **Fraction Choices.** The data source used for peptide mapping. The "F#" represents the fraction from which the protein originated, where # refers to the ID of the fraction, as indicated in the fraction node of the project tree. After making choices for the fraction selection, click the Apply button to update the chart.

    ## Note

    Peptide comparison allows a maximum of two fractions as the source.

  - **Minimal Feature Area:** A filter used to set a feature area threshold, where peptide features below this value will not be included in the display.

  - **Annotation:** When selected, the masses of the major peptides are displayed on the comparison chart.

  - **Tool Tip:** When selected, a pop-up will appear when the cursor hovers over any peptide within the comparison chart. The pop-up contains the peptide mass, the total area for that peptide, and the peptide sequence.

- **Peptide Comparison Chart:** This chart gives a visualization of the peptide comparison for the selected data. The chart is displayed in Peak Mode, where each peptide is represented as a bar and the height of that bar represents the total area of the associated peptide features in the specified fraction. The chart provides similar convenient ways to zoom-in and navigate the comparison chart.

- **Result Export.** The peptide comparison of a protein can be exported as an image, as well as in text format.

  - **Image export.** Click the Save to Image... button to save the displayed graph as a PNG file.

  - **Text format export.** Click the Save to Text Format... button to export the peptide comparison data to a text file in CSV format. The exported file contains all of the necessary information to construct the peptide comparison chart. The data are grouped together by peptide sequence and mass. The information for the same peptide are combined together if they belong to the same group, as mentioned earlier. The headers of the CSV column are self-explanatory. The following items are explained for clarity:

    - **Unique:** This pecifies that the peptide is unique (Y) or not unique (N).

    - **-10lgP or ALC:** The significance score of the peptide identified by database searching or by *de novo* sequencing for *de novo* only peptides.

    - **Mass:** The mass of the peptide.

    - **ppm:** The error tolerance of the peptide.

    - **Feature Areas:** This list of areas of the features associated with the corresponding peptide. The format of the feature areas is (area [space] +charge [semi-colon])*. For example, 500 +2; 800 +2; 700 +3

    - **Total Area:** The sum of the areas of the features associated with the corresponding peptide in the specified fraction.

    - **Ratio:** When the data sources contain two fractions, the ratio of the total areas of the corresponding peptide between two fractions is calculated.

    ## Note

    The data is arranged in separate -10lgP, ALC(%), Mass, ppm, Feature Area, and Total Area columns for each fractions.

    ## Note

    -10lgP, ALC(%), and ppm values are from the best PSM of the corresponding peptide.

## 3.3.5. Supporting Peptides

The "Peptide" view shows a table of supporting peptides for the protein. This table is similar to the peptide table in the Peptide View, except that "Accession" is excluded and the columns "(Checkbox)", "Unique", "Start", and "End" are included.

- **(Checkbox):** This controls whether a peptide is shown in the protein coverage as a blue bar.

- **Unique:** This shows whether the peptide is a unique supporting peptide to the protein group.

- **Start:** This shows the peptide's starting position in the protein.

- **End:** This shows the peptide's ending position (inclusive) in the protein.

The supporting peptides with peptide -10lgP score below the peptide filtering threshold are listed in the table as well. By default, they are shown in grey and are not displayed in the protein coverage.

## 3.3.6. De novo Tags

The "Denovo Only Tags" tab shows a table of "*de novo* only" peptides with sequence tags matched to the protein. The "*de novo* only" peptides shown in the table are controlled by the **de novo tags sharing X AAs** and **de novo peptides fully matched** filters in the control panel.

# 3.4. Peptide View

The "Peptide" view displays peptide identifications that have been filtered by the current settings in Summary View. The table on the top shows the details of each peptide identification. The bottom section provides additional information about the peptide-spectrum matches for the peptide selected in the table.

## 3.4.1. Peptide Table

The Peptide Table shows the filtered peptide identification results. Each row in the table is a peptide identification represented by its highest-scoring PSM. The table can be sorted by any column. A search function is also provided. When there are more than 1000 peptides, the table is split into multiple pages.

## Note

If there are multiple PSMs matched to a peptide, the table displays the information for the top-scoring PSM. Peptides with the same sequence but different modifications are considered different and shown in different rows. The #Spec column shows the number of PSMs matched to the peptide. Other PSMs can be examined after selecting a peptide in the table. See Section 3.4.2, "Peptide-Spectrum Matches" for details.

| | Peptide | -10lgP ↓ | Mass | Length | ppm | m/z | RT | Area | Scan | #Spec | Accession | PTM | AScore | Found By |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | IWNVHSVLNVLHSLVDKSNINR | 57.14 | 2556.3926 | 22 | 0.6 | 640.1058 | 106.03 | 6.05E6 | 20003 | 10 | Q14697\|GANAB_HUMAN | | | PEAKS DB |
| 2 | DVHNIYGLYVHMATADGLR | 56.71 | 2144.0474 | 19 | 0.1 | 715.6898 | 85.06 | 2.48E6 | 15900 | 5 | Q14697\|GANAB_HUMAN | | | PEAKS DB |
| 3 | VDDK(+4.03)VNFHFILFNNVDGHLYELDG... | 56.35 | 2985.5017 | 25 | -1.0 | 996.1735 | 91.41 | 1.24E6 | 17145 | 10 | P09936\|UCHL1_HUMAN:Q60HC8\|UCHL1_MACFA:... | k r | k4 k 1000.00; R25 r 1000.00 | PEAKS DB |
| 4 | VDDKVNFHFILFNNVDGHLYELDGR | 55.79 | 2975.4565 | 25 | 0.5 | 992.8266 | 91.47 | 1.19E6 | 17158 | 8 | P09936\|UCHL1_HUMAN:Q60HC8\|UCHL1_MACFA:... | | | PEAKS DB |
| 5 | SLTGVVNAQALTSAFSPHTKPWIGLAEALG... | 55.23 | 3552.8762 | 34 | -2.6 | 889.2240 | 113.83 | | 21536 | 4 | O43175\|SERA_HUMAN:A5A6P1\|SERA_PANTR | c | M33 c 1000.00 | PEAKS DB |
| 6 | VADPDHDHTGFLTEY(+79.97)VATR | 54.80 | 2222.9634 | 19 | 0.4 | 1112.4894 | 54.98 | 7.99E5 | 10019 | 10 | P63086\|MK01_RAT:P63085\|MK01_MOUSE:P2848... | P | Y15 P 20.41 | PEAKS DB |
| 7 | FGLALAVAGGVVNSALYNVDAGHR | 53.93 | 2370.2444 | 24 | 0.4 | 791.0891 | 104.43 | 2.41E6 | 19696 | 2 | Q3T165\|PHB_BOVIN:P67779\|PHB_RAT:P35232\|P... | | | PEAKS DB |
| 8 | ELFDDPSY(+79.97)VNVQNLDKAR | 53.72 | 2201.9993 | 18 | 0.3 | 1102.0072 | 75.55 | 9.03E5 | 14041 | 4 | P29353\|SHC1_HUMAN:Q5R7W7\|SHC1_PONAB | P | Y8 P 17.01 | PEAKS DB |
| 9 | DVHNIYGLYVHMATADGLR(+6.02) | 53.58 | 2150.0674 | 19 | 0.4 | 717.6967 | 85.18 | 2.46E6 | 15923 | 4 | Q14697\|GANAB_HUMAN | r | R19 r 1000.00 | PEAKS DB |
| 10 | QFGFIVLTTSAGIMDHEEAR | 53.50 | 2221.0837 | 20 | 0.8 | 741.3691 | 92.93 | 1.59E6 | 17447 | 2 | Q90YQ8\|RS15A_ICTPU:Q5R938\|RS15A_PONAB:... | | | PEAKS DB |
| 11 | K(+8.01)SDIYVC(+57.02)MISFAHNVAA... | 53.47 | 2254.1208 | 20 | 0.3 | 752.3811 | 83.54 | 3.72E5 | 15610 | 2 | Q61598\|GDIB_MOUSE:P50399\|GDIB_RAT:P5039... | k c k | k1 k 1000.00; K20 k 1000.00 | PEAKS DB |

## Tip

Right-click on a peptide to show a pop-up menu, and select "Show original spectrum". This would jump to the spectrum in the data view.

The following list describes the contents in each column:

- **Peptide:** The amino acid sequence of the peptide, as determined in PEAKS Search. A modified residue is followed by a pair of parentheses enclosing the modification mass.

- **-10lgP:** The peptide -10lgP score. The score indicates the scoring significance of a peptide-spectrum match.

- **Mass:** The monoisotopic mass of the peptide.

- **Length:** The number of amino acids in the peptide sequence.

- **ppm:** The precursor mass error, calculated as $10^6 \times$ (precursor mass - peptide mass) / peptide mass.

- **m/z:** The precursor mass-to-charge ratio.

- **RT:** The retention time (elution time) of the spectrum as recorded in the data.

- **Area:** The area of the peptide feature associated with the best PSM.

- **Scan:** The scan number of the spectrum that matches the peptide sequence with the highest -10lgP.

- **#Spec:** The number of spectra assigned to the peptide.

- **Accession:** The accession number of the highest-scoring protein containing this peptide.

- **PTM:** The types and the numbers of modifications present in the peptide shown in color-coded icons.

- **Found by:** The name of the PEAKS search tool that identifies the peptide; can be either PEAKS DB, PEAKS PTM, or SPIDER.

### 3.4.2. Peptide-Spectrum Matches

This section displays the spectrum annotation and other information about the peptide selected in the Peptide table. When multiple spectra are matched to the peptide, the information about the top-scoring PSM is shown by default. Other PSMs can be viewed by choosing from the pop-up list shown after clicking the All matches button. The list of proteins that contain the selected peptide can be examined by clicking the Protein button. Jump to the protein in the "Protein" view by clicking a protein in the pop-up list. The interface of this section is identical to the Spectrum Annotation in the PEAKS *de novo* result. For detailed instructions, please refer to Section 3.2, "*De Novo* Peptide View".
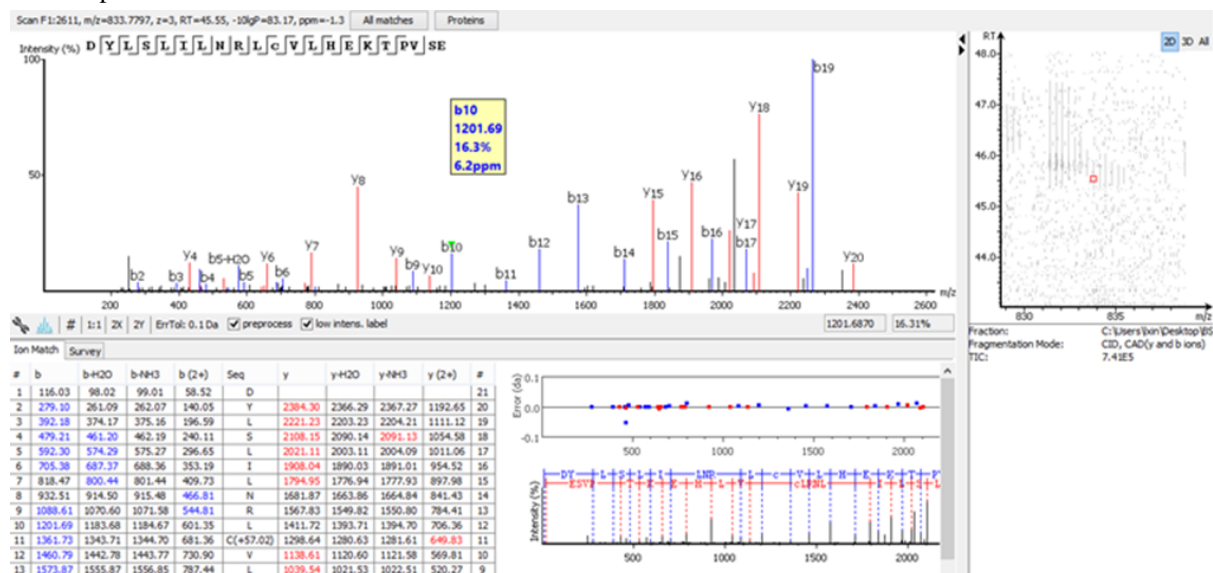


# 3.5. De Novo Only View

The *De novo* Only View displays the filtered "*de novo* only" sequences. A *de novo* sequence is "*de novo* only" if it is from an MS/MS spectrum that is not confidently assigned to any database peptide. The table shows the filtered "*de novo* only" sequences by the current settings in the "Summary" view. "*De novo* only" sequences may suggest novel peptides, peptides with unknown modifications, or other interesting research subjects. The interface

is identical to the "Peptide" view in a *de novo* sequencing result node. For detailed instructions, please refer to Section 3.2, "*De Novo* Peptide View".

## 3.6. LC-MS View

The LC-MS View visualizes identified peptides and "*de novo* only" sequences in an interactive LC-MS heat map. Check the option "Identified peptides" to show the locations of MS/MS spectra with peptide identifications, which are indicated by blue squares. Check the option "De novo only tags" to show the locations of MS/MS spectra with "*de novo* only" sequences, which are indicated by amber squares. Only the filtered peptide identifications and "*de novo* only" sequences are shown on the heat map.

Refer to Section 4, "LC/MS View" for a list of supported interactions in the LC-MS heat map.
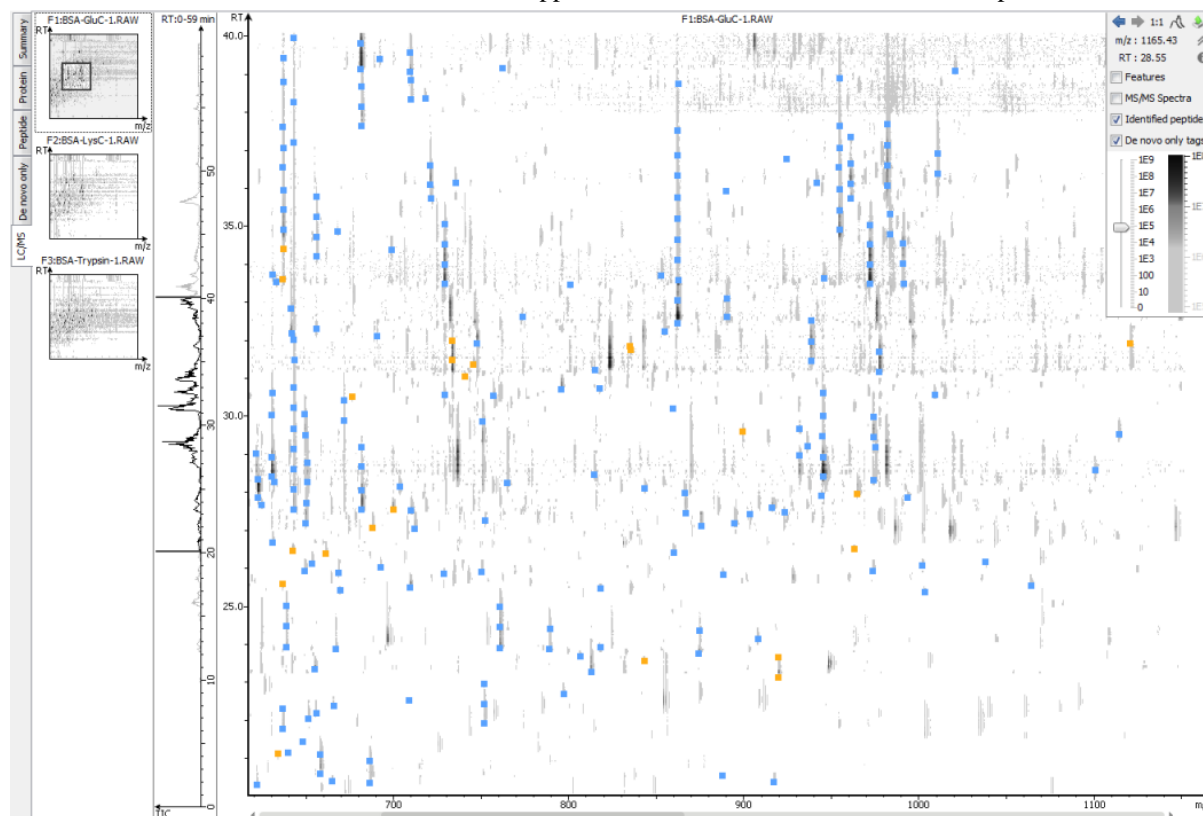


# 4. Filtering PEAKS Search Results

PEAKS Search results can be filtered to keep only the high quality peptides, proteins, and "*de novo* only" peptides. Filtering parameters can be specified using the control panel in the "Summary" view. Click the Apply Filters button to filter the results that will be updated accordingly in the Summary, Protein, Peptide, and LC/MS views. The filtering options in the control panel are described in the rest of this section.



### Note

Whenever a filtering parameter is changed, the Apply Filters button will be highlighted in red as a reminder that the changes have NOT taken effect yet.

**Peptides.**    Peptide-spectrum matches are filtered by peptide -10lgP scores. The peptide -10lgP threshold affects the results shown in both the "Peptide" and the "Protein" views. Only the peptides with PSMs above the -10lgP

score threshold are listed in the "Peptide" view. These filtered peptides are used as supporting peptides to infer protein identifications; therefore, affect the proteins shown in "Protein" view.

If the "Estimate FDR with decoy-fusion" option is selected in the PEAKS search parameters, the peptide -10lgP score threshold can be easily set using the estimated FDR. Clicking the FDR button shows an interactive FDR curve. Moving the cursor along the curve will show the estimated FDR at a specific score threshold. At a desired FDR, the corresponding peptide -10lgP score threshold can be set by right-clicking and selecting Copy score threshold. FDR selection from a list of predefined FDR values is also possible, which will also set the score threshold.
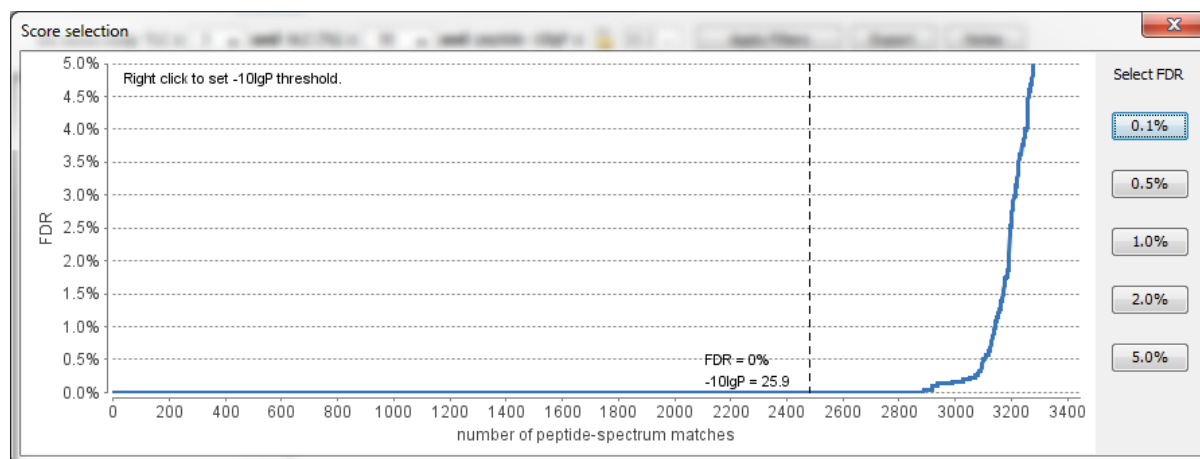


If the "Estimate FDR with decoy-fusion" option is not selected, the peptide -10lgP score threshold will need to be set empirically. A score threshold set at 20 is a good choice to start with as its equivalent p-value will be 0.01.

## Note

The concepts of P-value and FDR are different. In PEAKS Search, p-value is defined as the probability that a false identification in the current search achieves the same or better matching score. A 1% p-value does not automatically correspond to a 1% FDR. For more details, please see `http://www.bioinfor.com/peaks/tutorials/peaksdbscore.html`

**Proteins.**    Protein identifications are filtered by protein -10lgP scores and by the number of unique peptides. It is important to empirically set a protein -10lgP score threshold. A protein score threshold of 20 or higher is recommended. Unique peptides are supporting peptides that are only linked to one protein group. A confident protein identification normally requires at least two unique peptides. These protein filtering options affect only the "Protein" view.

***De novo Only.***    "*De novo* only" peptides are filtered by ALC score. The ALC score is the average of local confidence score of all of the residues in a *de novo* sequence. When the highest peptide -10lgP score of all PSMs matched to a spectrum falls below the peptide threshold, the *de novo* sequencing result on that spectrum is considered a "de novo only" peptide. By default, the threshold is automatically set, and is locked as the same value as the -10lgP score threshold for filtering peptides.

The ALC score threshold can be determined using Figures 3(a) and 3(b) in the "Summary" view. For details, please refer to Section 3.2.3, "De Novo Result Validation".

# 5. Exporting PEAKS Search Results

The "Export" button in the title bar of the "Summary" view allows for the exporting of filtered results into various formats. It provides the opportunity to share the report in HTML, use the result as a supplemental figure/table in a publication, or further validate and analyze the result in a third party software platform. Please refer to Section 4.1, "Exporting Summary, Proteins, and Peptides" for details.

**Note**

Projects can be shared by copying the PEAKS project directory. It can then be opened in the free PEAKS Viewer (`http://www.bioinfor.com/peaks/viewer/index.php`) that has the same GUI as PEAKS Studio.

# 6. Running PEAKS PTM and SPIDER Separately

PEAKS PTM and SPIDER can be launched independently from a PEAKS DB result by selecting a PEAKS DB result node and clicking the 🔧 button or the 🕷 button on the toolbar, respectively.
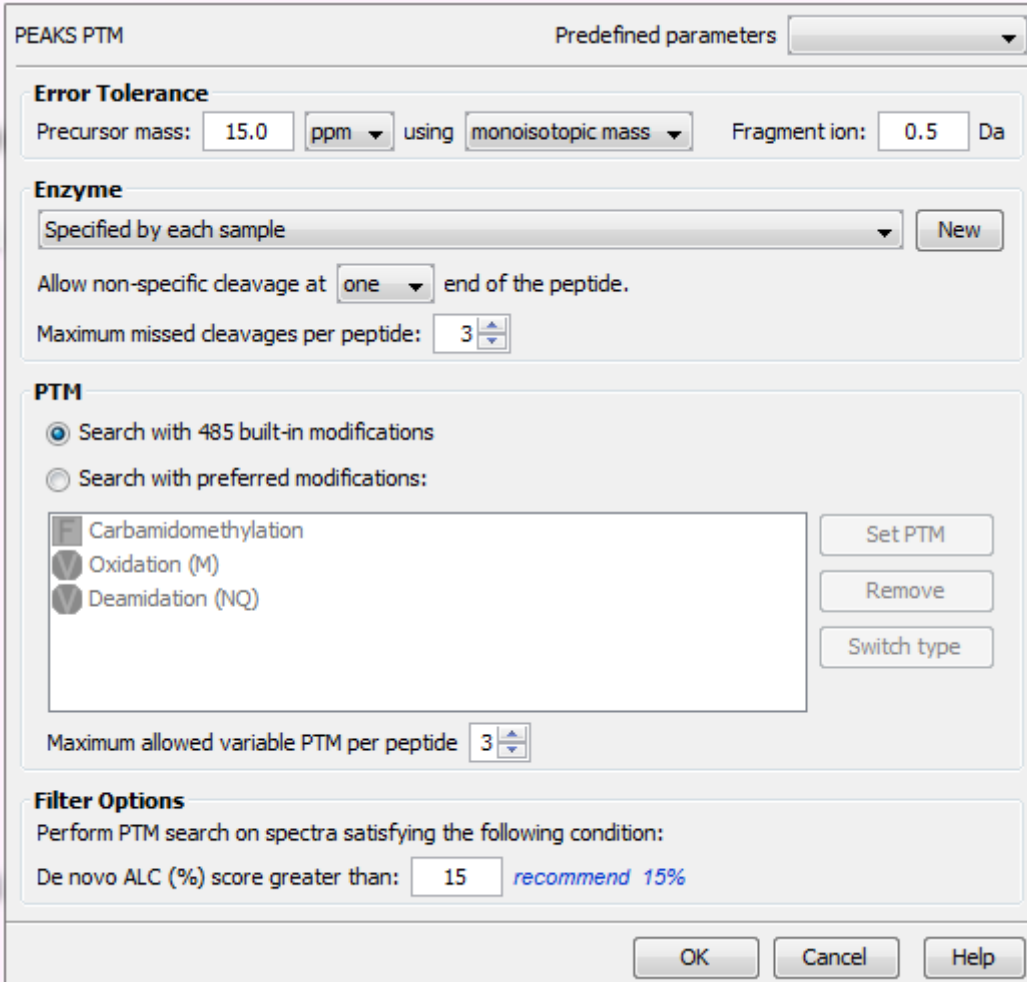
**Note**

SPIDER can also be launched from a PEAKS PTM result. In this case, select a PEAKS PTM result node and click the 🕷 button on the toolbar.

## 6.1. Running PEAKS PTM on a PEAKS DB Result

Launch PEAKS PTM by selecting a PEAKS DB result and clicking the PEAKS PTM icon on the toolbar 🔧. Running PEAKS PTM on a PEAKS DB result is functionally equivalent to running both searches together in an integrated search.

The parameters used in a PEAKS PTM search are similar to a PEAKS DB search. Selecting a protein database is not necessary as PEAKS PTM automatically uses the database searched in PEAKS DB.

## 6.2. Running SPIDER on a PEAKS DB or PEAKS PTM Result

Launch SPIDER by selecting a PEAKS DB or a PEAKS PTM result and clicking the SPIDER icon on the toolbar
✳. Running SPIDER on a PEAKS DB or a PEAKS PTM result is functionally equivalent to running both searches together in an integrated search. However, the parameters panel, in this case, appears as follows and allows for a few additional options:
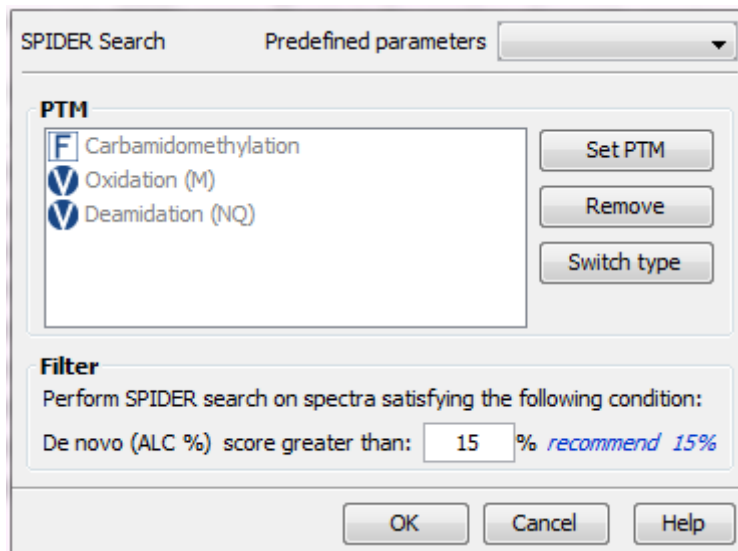


### Note

For users of previous versions of PEAKS, SPIDER in the current version runs in the mode that was previously called Homology Match.

**PTM.**    Clicking the Set PTM button will bring up a separate window for choosing modifications. The PTM configuration is the same as it is in *de novo* sequencing. Refer to Section 2.3, "PTM" for details.

**Filter.**    The filtering option asks for the minimum *de novo* tag score (ALC) for a spectrum to be analyzed in SPIDER. If the ALC score of a spectrum is too low, the spectrum is unlikely to provide a significant hit.

# 7. Comparison of PEAKS Results

It is possible to compare up to three search results in the same project. To compare results, press Ctrl to select two or three result nodes, right-click, and then choose Compare Results from the pop-up menu. PEAKS will then start to compare the results.
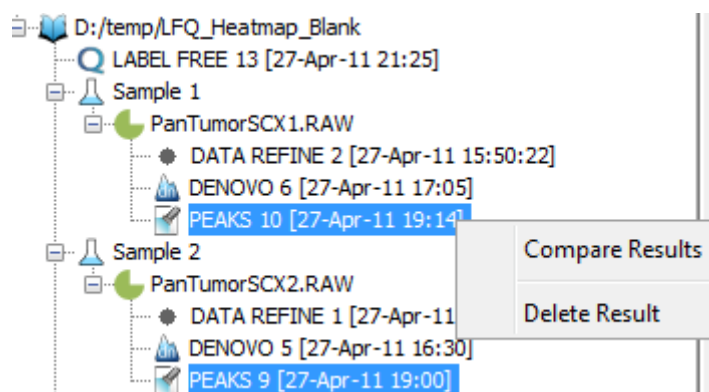
# 7.1. Comparison Result

When comparison is completed, a comparison result node is added to the project, as shown in the following screenshot. The comparison result is opened automatically. The result consists of three parts: peptide comparison, protein comparison, and statistical charts.
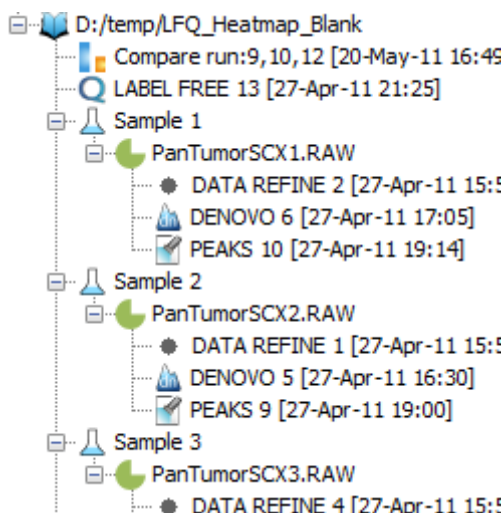


## Note

The comparison is done on the fly and the result is not saved with the PEAKS project. It is suggested to export the comparison result before closing the project. For details on exporting, please refer to Section 7.5, "Exporting Comparison Result".

# 7.2. Peptide Comparison

Peptide identifications from the compared search results are displayed in the table. For each peptide, the table shows m/z, retention time, peptide -10lgP score, charge, and whether there are multiple PSM hits. The Coverage Map column graphically illustrates from which results the peptide is identified. A solid icon indicates the peptide is identified.

Using the control panel in the bottom, each result can be filtered separately by setting the peptide -10lgP score threshold. It is possible to show only the common peptides shared by all results or the unique peptide in one result.

The following screenshot is the peptide table in a typical comparison result:

# 7.3. Protein Comparison

The top protein identifications from the compared search results are displayed in this table. The layout of the protein table and the control panel at the bottom is similar to the Peptide Comparison pane. The following information is also displayed for each protein:

- **Score:** PEAKS protein -10lgP score

- **#Spec:** The number of PSMs matched to supporting peptides for the protein.

- **#Pep:** The number of supporting peptides for the protein

- **#Uniq:** The number of unique supporting peptides for the protein

- **%Spec:** The number of peptides detected for the protein over the number of peptides in theoretical digestion.

- **%Cov:** The sequence coverage of the protein

The following screenshot shows the protein table displaying a typical comparison result:

| Protein_ID | Cover Map | PEAKS 9 [27-Apr-11 19:00] | | | | | | PEAKS 10 [27-Apr-11 19:14] | | | | | | PEAKS 12 [27-Apr-11 19:54] | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Score | #Spec | #Pep | #Uniq | %Spec | %Cov | Score | #Spec | #Pep | #Uniq | %Spec | %Cov | Score | #Spec | #Pep | #Uniq | %Spec | %Cov |
| P42656\|RAD24_SCHPO | | | | | | | | | | | | | | 64.13 | 8 | 2 | 0 | 6.9 | 6.67 |
| P0C7J6\|LRFN1_RAT | | | | | | | | 0 | 1 | 1 | 1 | 1.41 | 1.83 | | | | | | |
| Q6B4U9\|PRDX1_MYOLU | | 125.49 | 16 | 6 | 5 | 26.09 | 33.17 | 131.11 | 16 | 6 | 5 | 26.09 | 33.17 | 139.03 | 25 | 7 | 0 | 30.43 | 37.19 |
| Q9NY65\|TBA8_HUMAN | | 90.82 | 6 | 2 | 0 | 5 | 7.13 | 133.01 | 6 | 3 | 0 | 7.5 | 11.8 | 112.14 | 5 | 3 | 0 | 7.5 | 11.8 |
| A2A3N6\|PIPSL_HUMAN | | 63.27 | 1 | 1 | 1 | 1.2 | 1.97 | | | | | | | | | | | | |
| Q75BU6\|VID21_NEUCR | | | | | | | | 0 | 1 | 1 | 1 | 0.62 | 0.69 | | | | | | |
| Q8SQH5\|ADT2_BOVIN | | 53.64 | 3 | 1 | 1 | 2.38 | 4.36 | 81.82 | 4 | 2 | 1 | 4.76 | 7.38 | 51.5 | 3 | 1 | 1 | 2.38 | 4.36 |
| P10096\|G3P_BOVIN | | 59.15 | 2 | 1 | 1 | 2.78 | 4.2 | 118.76 | 6 | 2 | 2 | 5.56 | 10.51 | | | | | | |
| Q5ZKH4\|NDEL1_CHICK | | | | | | | | | | | | | | 51.43 | 2 | 1 | 1 | 2.22 | 3.5 |
| P30084\|ECHM_HUMAN | | 47.77 | 2 | 2 | 2 | 6.06 | 13.45 | 37.47 | 1 | 1 | 1 | 3.03 | 6.21 | 19.87 | 2 | 2 | 2 | 6.06 | 12.76 |
| P35998\|PRS7_HUMAN | | | | | | | | | | | | | | 48.99 | 2 | 2 | 2 | 3.33 | 4.85 |
| A1JM81\|RNFC_YERE8 | | | | | | | | 0 | 1 | 1 | 1 | 1.18 | 1.58 | | | | | | |
| P24724\|HSP90_THEPA | | 56.02 | 2 | 1 | 0 | 0.92 | 1.94 | 57.79 | 1 | 1 | 0 | 0.92 | 1.94 | 83.52 | 3 | 2 | 0 | 1.83 | 3.33 |
| Q8PD23\|CH60_XANCP | | | | | | | | 0 | 2 | 2 | 2 | 3.23 | 5.68 | | | | | | |
| A4WW89\|DNAK_RHOS5 | | | | | | | | | | | | | | 65.32 | 5 | 1 | 0 | 1.3 | 2.52 |
| A7EDF3\|IRS4_SCLS1 | | | | | | | | 0 | 1 | 1 | 1 | 0.97 | 2.28 | | | | | | |
| Q8CGP0\|H2B3B_MOUSE | | 89.94 | 10 | 2 | 1 | 7.14 | 19.05 | | | | | | | | | | | | |
| P09103\|PDIA1_MOUSE | | 138.91 | 12 | 5 | 5 | 7.94 | 12.38 | 153.41 | 20 | 5 | 5 | 7.94 | 11 | 155.03 | 22 | 7 | 7 | 11.11 | 14.34 |
| Q28DR4\|H4_XENTR | | 50.05 | 5 | 1 | 1 | 4 | 11.65 | 104.49 | 10 | 3 | 3 | 12 | 31.07 | 77.91 | 4 | 2 | 2 | 8 | 19.42 |
| O77819\|ROCK1_RABIT | | | | | | | | | | | | | | 40.73 | 1 | 1 | 1 | 0.47 | 0.81 |
| Q25538\|1433_NEOCA | | | | | | | | | | | | | | 64.13 | 8 | 2 | 0 | 6.06 | 6.77 |
| Q8NDF8\|PAPD5_HUMAN | | 16.93 | 2 | 1 | 1 | 1.52 | 3.5 | | | | | | | | | | | | |
| Q6VAF4\|TB69_GOSHI | | | | | | | | 55.2 | 1 | 1 | 0 | 2.7 | 3.15 | | | | | | |
| P22492\|H1T_HUMAN | | 65.87 | 4 | 1 | 1 | 1.92 | 5.31 | 73.17 | 3 | 1 | 1 | 1.92 | 5.31 | 65.17 | 4 | 1 | 1 | 1.92 | 5.31 |
| P12303\|TTHY_SHEEP | | | | | | | | | | | | | | 0 | 1 | 1 | 1 | 6.67 | 10.88 |
| Q92597\|NDRG1_HUMAN | | 63.45 | 1 | 1 | 1 | 3.7 | 3.55 | 67.78 | 2 | 2 | 2 | 7.41 | 7.61 | 64.4 | 1 | 1 | 1 | 3.7 | 3.55 |
| Q00799\|RBP2_PLAVB | | 0 | 1 | 1 | 1 | 0.23 | 0.59 | | | | | | | | | | | | |
| Q64L94\|PSME1_PIG | | 124.48 | 8 | 4 | 4 | 10 | 20.48 | 119.77 | 9 | 3 | 3 | 7.5 | 15.26 | 137.5 | 16 | 5 | 5 | 12.5 | 23.69 |
| P13491\|LDHA_RABIT | | | | | | | | 39.91 | 2 | 1 | 1 | 2.63 | 2.71 | 43.13 | 2 | 2 | 2 | 5.26 | 5.42 |
| P63112\|HBA_PAPCY | | 55.93 | 2 | 1 | 1 | 6.25 | 8.45 | 62.37 | 2 | 1 | 1 | 6.25 | 8.45 | 85.95 | 7 | 2 | 2 | 12.5 | 14.79 |
| Q3U962\|CO5A2_MOUSE | | | | | | | | | | | | | | 69.02 | 3 | 2 | 2 | 1.6 | 2.74 |
| P00739\|HPTR_HUMAN | | 59.47 | 3 | 1 | 1 | 2.86 | 3.45 | 72.06 | 5 | 2 | 2 | 5.71 | 7.18 | 83.49 | 6 | 2 | 2 | 5.71 | 6.32 |
| P08814\|PTMS_BOVIN | | 51.66 | 6 | 1 | 1 | 5.88 | 10.78 | 66.86 | 7 | 1 | 1 | 5.88 | 10.78 | 51.35 | 3 | 1 | 1 | 5.88 | 10.78 |
| Q5R887\|MYL3_PONAB | | | | | | | | | | | | | | 57.09 | 3 | 1 | 1 | 4 | 6.15 |
| Q4QQU6\|SPF30_RAT | | 36.46 | 2 | 1 | 1 | 2.86 | 4.2 | | | | | | | | | | | | |
| O43103\|SID2_USTMA | | 0 | 1 | 1 | 1 | 0.26 | 0.32 | | | | | | | | | | | | |
| A8YUJ9\|ATPA_LACH4 | | 39.42 | 1 | 1 | 1 | 1.92 | 2.39 | | | | | | | | | | | | |
| Q7VJ82\|RPOBC_HELHP | | | | | | | | 0 | 1 | 1 | 1 | 0.26 | 0.31 | | | | | | |

# 7.4. Statistical Charts

PEAKS provides a number of statistical charts that can be easily exported for usage in publications. The peptide score distribution, the protein score distribution, the peptide number Venn diagram, and the protein number Venn diagram can help validate the results. The charts can be exported from the pop-up menu after right-clicking the comparison result node.
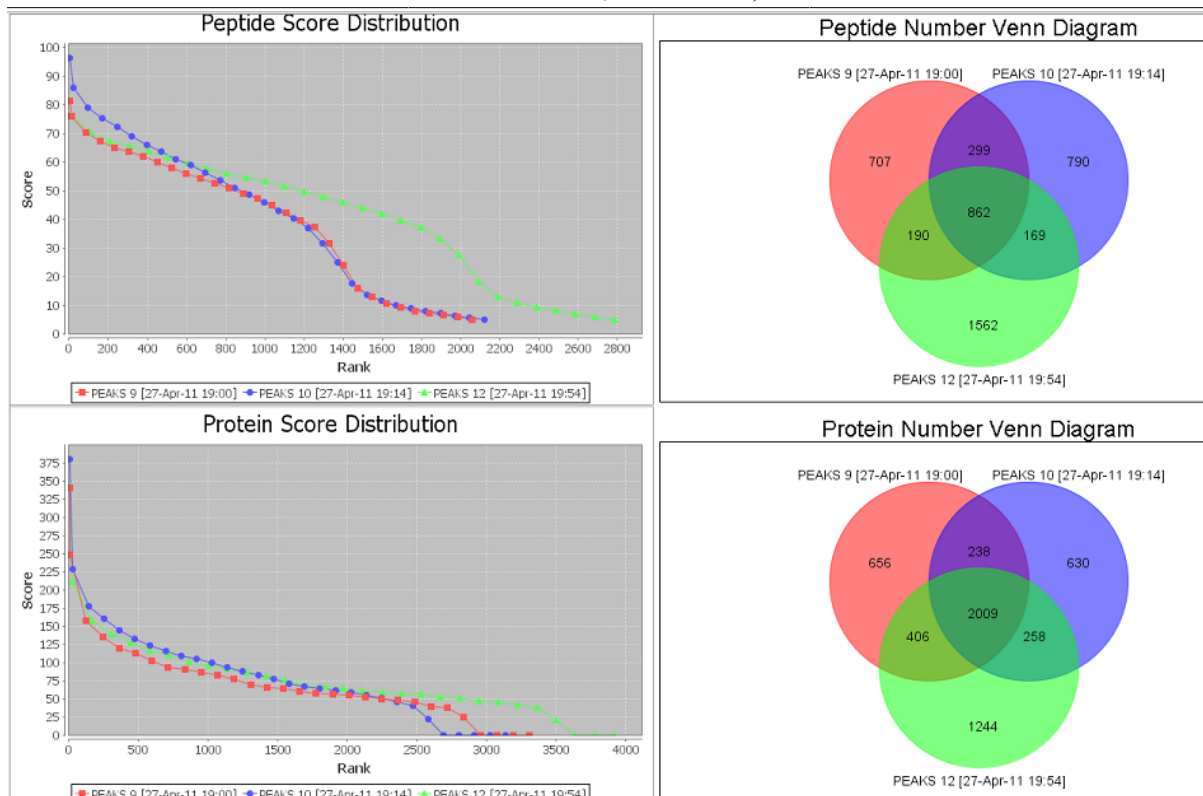
## 7.5. Exporting Comparison Result

The peptide table, the protein table, and the statistical charts in the comparison result can be exported into a single spreadsheet. Right-click on the comparison result node and choose Export Excel. Choose the items to be exported and set the image quality for statistical charts.

# Chapter 10. Multi-Round Search

## 1. Multi-Round Search Overview

The Multi-Round Search is an additional search function that follows a completed search result to examine the remaining "de novo only" spectra. This additional search allows users to perform PEAKS DB, PTM and SPIDER searches using different parameters and/or to integrate a different database.

### Note

In a recent poster presented by PEAKS and Delft University of Technology, the Multi-Round Search tool allowed a direct analysis of endogenous peptides in complex secretomes without physical removal of abundant proteins. For more details, refer to the poster: http://www.bioinfor.com/images/stories/pdf/PEAKSposters/peaks%20-%20quantitative%20analysis%20of%20endogenous%20peptides%20web.pdf - quantitative analysis of endogenous peptides web.pdf

Follow these steps to perform a multi-round search:

1. Select a PEAKS DB, PTM or SPIDER result node in the "Project View". Open the result and set filters. The *de novo* results found in the "de novo only" tab will be carried forward to the Multi-Round Search.

2. To open the search parameters window, select a PEAKS DB, PEAKS PTM or SPIDER result node in the "Project View". Launch the multi-round search by clicking the ![icon] icon on the tool bar.
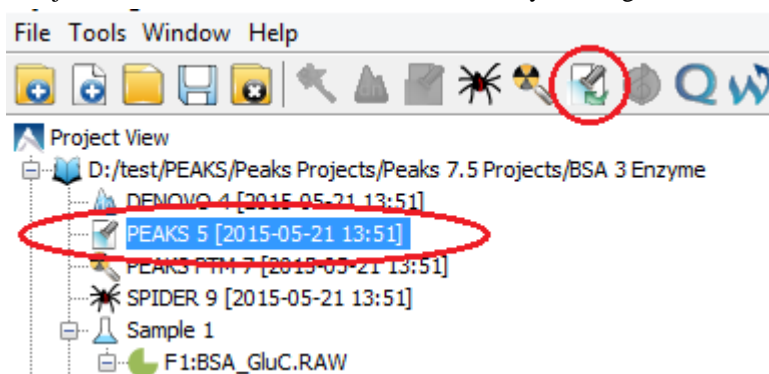


3. Specify the search parameters in the pop-up dialogue. The parameter selection dialogue is similar to the parameter selection dialogue in Chapter 9, *Peptide, PTM, and Mutation Identification (PEAKS DB, PEAKS PTM, and SPIDER)*.

4. Wait for the analysis to finish. New result nodes for multi-round PEAKS DB, PEAKS PTM, and SPIDER will appear in "Project View". Double-click the result nodes to examine the associated results.

# Chapter 11. Combining Multiple Database Search Engines with PEAKS inChorus

## 1. PEAKS inChorus Overview

Properly combining the results from different database search engines can enhance the accuracy and sensitivity of peptide identifications. PEAKS inChorus is a tool for combining PEAKS results with search results from other engines. PEAKS inChorus can either launch a database search or import a search result. The other supported search engines are SEQUEST (Proteome Discoverer versions 1.3 and 1.4), Mascot (v2.4), X!Tandem (v2010.12.01.1), and OMSSA (v2.1.8). A uniform FDR, also known as inChorus FDR, is estimated to assist the filtering of the combined result.

In this chapter, it is assumed that you are already familiar with PEAKS database search tools (Chapter 9, *Peptide, PTM, and Mutation Identification (PEAKS DB, PEAKS PTM, and SPIDER)*). The usage of PEAKS inChorus is outlined in the overview. For details of each step, please refer to later sections in this chapter.

1.
   Select a project, a sample, or a fraction node. Click the PEAKS inChorus button ⊘ on the tool bar.



2. In the inChorus parameter dialogue, there are drop-down menus for each supported database search engine. Using the drop-down menus, you can choose the search results to be combined. You can either:

- **Choose an existing result**

   The list shows search results that already exist in the PEAKS project. Simply choose an existing result from the list.

- **Import a result from a file**

   If the result by the search engine is exported to a file, that file can imported by choosing Import from the list. Specify the file and the searched database used from the pop-up dialogue.

- **Launch a new search**

   A new database search can be launched using the particular search engine by choosing New Run... from the list. Specify the search parameters in a pop-up dialogue similar to the search engine's native interface. Please refer to the particular search engine's user manual for instructions. For the PEAKS database search, please refer to Section 2, "Setting PEAKS Search Parameters".

## Note

A PEAKS result must be selected to combine results using inChorus.

## Note

When launching a new search, if the data in the PEAKS project is not yet refined, data refinement will be performed first with default parameters.

## Important

To enable the calculation of the inChorus FDR, the same target-decoy database needs to be searched by all of the engines. For PEAKS, X!Tandem, and OMSSA, this target-decoy database is generated automatically. For SEQUEST and Mascot, a target-decoy database needs to be exported from PEAKS and added to their database list. Use the tool from Configuration->Database to export the target-decoy database.

Then, check the "Search decoy database from PEAKS" checkbox for each third-party engine in their parameter settings or importing dialogues (for PEAKS check the "Estimate FDR with decoy-fusion" checkbox).

3. Wait for PEAKS inChorus to complete the analysis. A new inChorus result node will appear in the Project View. Double-click the node to open the result.

4. The result presentation is similar to a PEAKS Search result, with additional columns in the Peptide table, which show the search engine(s) that identified the peptide and their corresponding scores.

# 2. Understanding PEAKS inChorus Results

An inChorus result is displayed in a very similar format of a PEAKS Search results (Section 3, "Understanding PEAKS Search Results"). This section only highlights the differences.

**Peptide Score.**   The first noticeable difference is that the inChorus peptide score is not the -10lgP score used in PEAKS DB. Instead, a percentage confidence score is used to reflect the probability that this peptide-spectrum match is correct. The percentage score is calculated in accordance with the empirical calculation used in Peptide-Prophet[TM] (Keller et al. Anal. Chem. 2002, 74:5383-92).

1. The method of PeptideProphet[TM] is applied to each engine's result to estimate the probability of correctness for each peptide identification.

2. If a peptide-spectrum match is identified by multiple engines, the scores for all those engines are added up to a weighted sum.

3. The weighted sum scores of all peptides are converted to a probability by using the PeptideProphet[TM] method again.

**Protein Score.**   The protein score is also a percentage score. It is calculated using the following approach. The scores of the peptides from a protein are added up to a weighted sum. The PeptideProphet[TM] method is then applied to the weighted sum scores of all proteins to convert to a probability score.

**Engines Icon.**   For each peptide in the "Peptide" view, the "Engines" column shows the engine(s) that identified the peptide. Each engine is represented by a icon showing the initial letter of the engine's name:

- **P:** PEAKS Search (PEAKS DB, PEAKS PTM, and SPIDER)

- **M:** Mascot

- **S:** SEQUEST

- **X:** X!Tandem

- **O:** OMSSA

Also, the following appearances of the icon indicates:

- **Blue background** The engine identified the peptide with high confidence (above the engine's own score threshold. See Section 3, "Filtering PEAKS inChorus Results").

- **White background** The engine identified the peptide with low confidence (below the engine's own score threshold).

- **Dash** The engine did not identify the peptide.



**Individual Engine's Score.** In the "Peptide" view, each engine's own score is displayed. A dash symbol indicates that the peptide is not found by the engine.

**FDR curves.** The FDR curves for all selected search engines are placed together in a single figure in the "Summary" view. The black * marks on the search engines' FDR curves indicate the score thresholds of the corresponding engines. The red * mark indicates the inChorus FDR threshold.
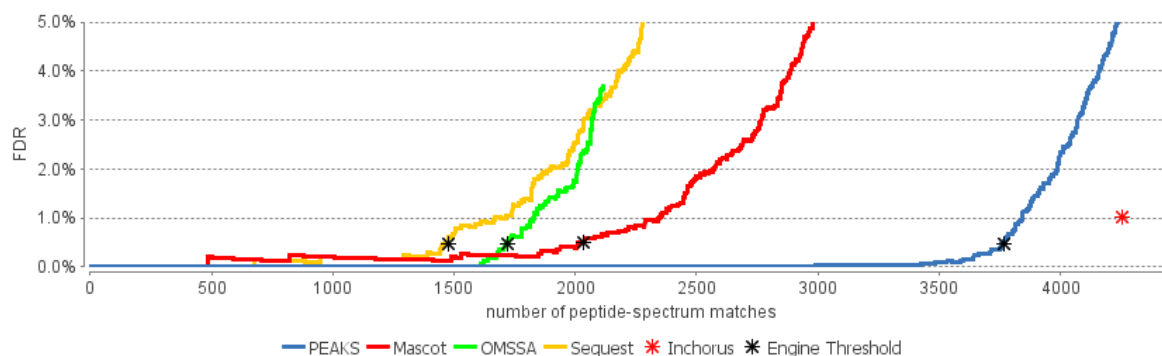


# Note

Since there are overlaps in the peptides identified by different engines, the FDR threshold for the inChorus result is higher than the respective FDR threshold for each individual engine.

**Venn Diagram.** A Venn diagram shows the number of identified PSMs by the three engines that identified the most PSMs under the current FDR threshold setting.
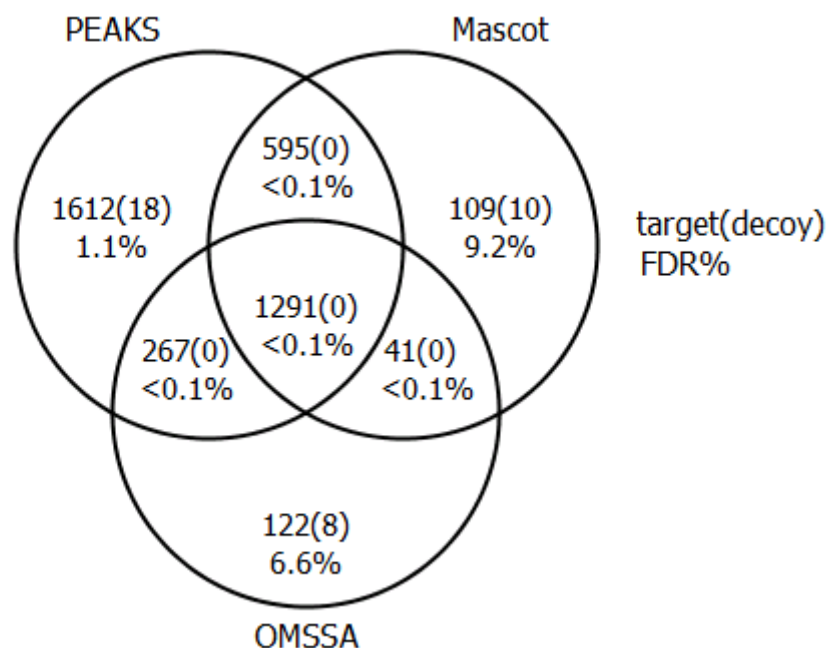
# 3. Filtering PEAKS inChorus Results

The filtration parameters can be specified using the control panel in "Summary" view. Click the Apply Filters button after changing the parameter values.



Filtering peptides in the inChorus result is different to that in the PEAKS Search result. Click the Edit filters button in the "Summary" view to specify the peptide filtering rules. Peptides can be filtered by either the inChorus FDR or by each individual engine's score. If filtering by the inChorus FDR, PEAKS calculates the score threshold for each search engine to automatically reach the target inChorus FDR. If filtering by each individual engine's score, a peptide is kept as long as at least one engine's score is above the specified threshold.

# 4. Exporting inChorus Results

The inChorus exporting function contains the same exporting options as PEAKS DB. Unlike PEAKS DB, the exported results will contain inChorus scores and individual search engine scores for the supporting peptides of the protein identifications. See Section 4, "Exporting Database Search Results" for details.

# Chapter 12. Precursor Ion Quantification (e.g. SILAC and ICAT)

## 1. Overview

Precursor ion quantification with isotope labels at the MS level is one of the three quantification modes that are supported by the optional PEAKS Q module for PEAKS Studio. In this mode, the isotope labels with different mass values are introduced as two or more samples. The samples are then analyzed together in an LC-MS/MS experiment. The same peptide from different samples is recognized by a set of precursor ion peaks with similar retention time and mass differences within the retention time window and the mass error tolerance set by the user. The ratio is calculated from the intensities of those peaks. PEAKS Q supports both user-defined labels and commercial quantification labels.

Precursor Ion Quantification can be launched from either the Project Wizard workflow Chapter 4, *Project Wizard - Create a PEAKS Project, Load Data and Set up a Workflow*, or on an existing PEAKS DB result. When running quantification on an existing result, ensure that the specified isotopes are labeled as variable PTMs when performing the database search. After the database search is complete, follow these steps:

1. Select a PEAKS DB result node in the "Project View" frame. Click the PEAKS Quantification toolbar icon,

   , or select the **Tools|Quantification** command from the menu.

   ### Note

   > Refer to Chapter 4, *Project Wizard - Create a PEAKS Project, Load Data and Set up a Workflow* for how to perform PEAKS DB and quantification in a single easy workflow.

2. Select the quantification protocol, "Precursor Ion Quantification"; specify the PEAKS quantification parameters in the dialogue box on the right; and click OK.

3. Wait for the analysis to finish. A new quantification result node, , will appear in the "Project View". Double click the node to open the result.

## 2. Setting Parameters

The following parameter dialogue appears when the quantification tool bar icon, , is clicked or the **Tools| Quantification** command from the menu is clicked. Select **Precursor Ion Quantification eg. SILAC** under the **Quantifications** heading.
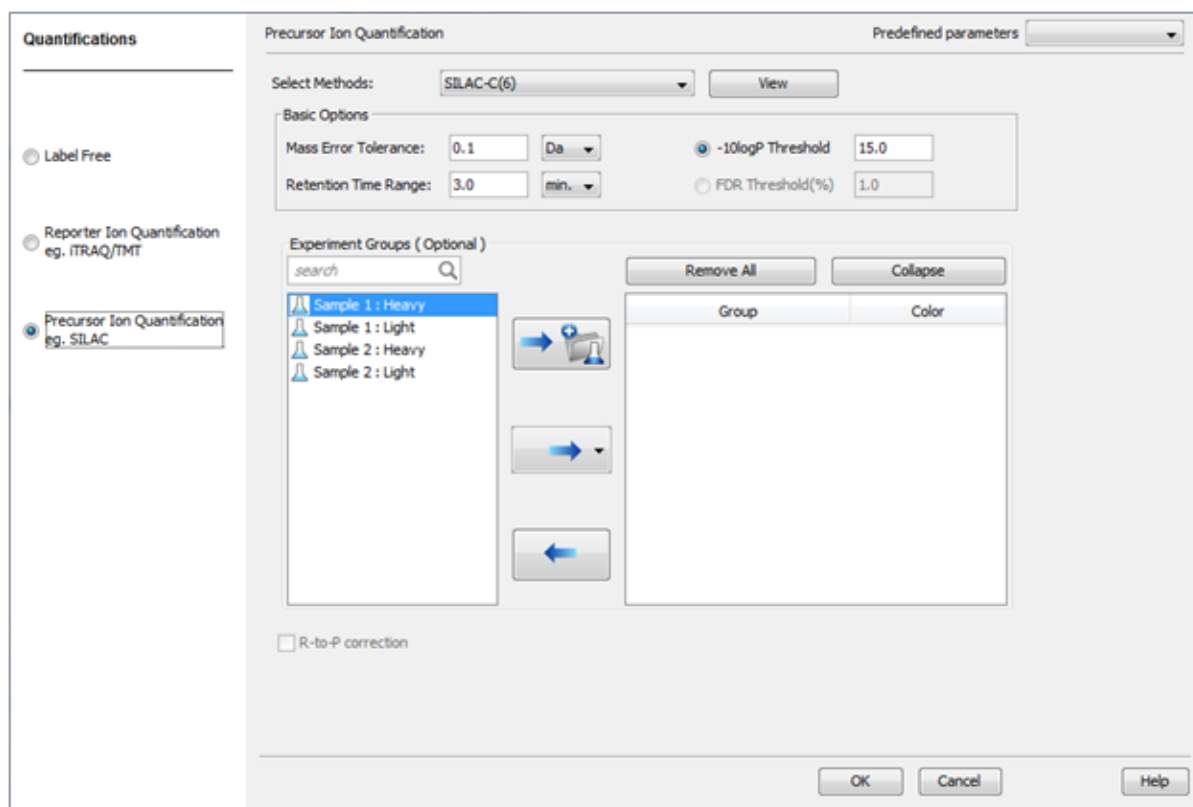
From the Select Methods drop-down menu, choose the appropriate quantification method used in experiment. To examine the details of a selected quantification method, click on the View button. To create a new quantification method, select Add New Method from the Select Methods drop-down menu. Refer to Section 2.3, "Labeled Q Method Configuration" for how to create or edit a quantification method.

The following parameters are available in the "Basic Options" section:

- **Mass Error Tolerance:** This parameter is used to locate the precursor ion peak group of an identified peptide in the survey scan. It is recommended to set the mass error tolerance at the same precursor ion error tolerance used in the PEAKS DB search . Either Da or ppm can be selected as the mass unit.

- **Retention Time Range:** The retention time range is the maximum retention time difference between paired peptide features.

- **-10logP Threshold:** Only peptides with a score above this threshold are used to quantify the identified proteins.

- **FDR Threshold(%):** Only peptides with a score above this threshold are used to quantify the identified proteins.

### Note

Either -10lgP or FDR threshold can be selected

Quantification allows for the option of sample grouping. For experiments with multiple samples, the search bar can be used to look for a specific sample. After selecting a sample, it can be added to a new group by clicking the ➡️ button or to an existing group by clicking the ➡️ button and selecting that option from the drop-down menu. To add multiple files into their own separate groups, click the ➡️ button. Display all samples in each group or just the individual groups by selecting the **Expand** and **Collapse** buttons, respectively. Samples can be removed either individually or all at once by clicking the ⬅️ button or the **Remove All** button, respectively.

**R-to-P correction:** This will tell PEAKS to consider the possibility that free heavy arginine residues metabolically converted to proline residues before insertion into the peptide. The areas of both the normal heavy-labelled arginine

peptide feature and the heavy-labelled proline feature will be summed to calculate the complete area of the heavy-labelled peptide feature.

## Note

If a quantification experiment is set separately from the complete workflow, select "R to P (13C5-15N1)" from the PTM options window in an Identification workflow and set it as a variable modification. This step is unnecessary if a complete workflow is run from a New Project as the modification will automatically be included.

**Predefined parameters:** This drop-down menu enables the saving, as well as the access to previously saved parameters, for the ease to perform quantification regularly with the same parameters.

# 3. Understanding the Results

Once completed, the protein quantification result will be displayed in the quantification node  in the Project View. Double-click on this node to open the result that contains four views: "Summary" view, "Protein" view, "Peptide" view, and "LC/MS" view. The "Summary" view will appear by default.
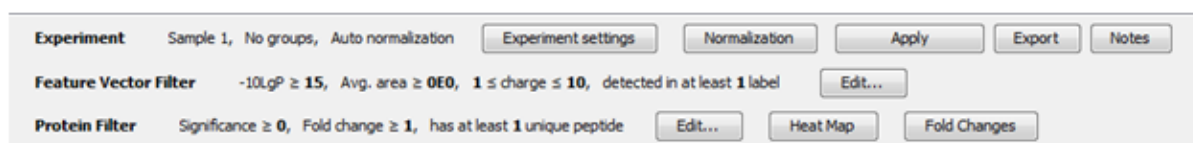
## 3.1. Summary View

The "Summary" view provides three main functions:

1. **Result Filtration:** This is achieved by specifying the filtration rules in the area at the top of the "Summary" view. The filtration function is discussed in Section 3.1.1., "Result Filtration".

2. **Summary Report:** Several statistical charts assist the user to get an overall picture of the results, assess the result quality, and examine the reliability and reproducibility of the experiment. This function is the focus of this section and is discussed in Section 3.1.2., "Summary Report".

3. **Result Exporting:** This is achieved by clicking the "Export" button at the top of the "Summary" view. The exporting function is discussed in Section 3.1.3, "Exporting Quantification Results".

The "Summary" view includes an expression profile with candidate proteins assorted in a heat map, result statistics in both figures and tables, and a list of filtration parameters. To add a summary note, click on the **Notes** button at the top-right of the "Summary" view to open a "Notes Entry" editor in which you can edit the notes. These notes will be displayed in the "Summary" view.

### 3.1.1. Result Filtration

PEAKS provides an extensive set of tools to filter Precursor Ion Quantification results according to the users' requirements. There are three categories of options in the "Summary" view: **Experiment**, **Feature Vector Filter**, and **Protein Filter**.



#### 3.1.1.1. Experiment

There are two buttons under this heading – **Experiment Settings** and **Normalization** – which allow for control over the analysis.

The following parameter dialogue appears when the **Experiment Settings** button is selected.

- **Select Experiment:** When running quantification on an identification node that searched multiple samples (i.e. replicate experiments within different treatments), this drop-down menu allows for the selection of a single sample or every sample (All Experiments) for an analysis. Every label of the sample(s) selected will appear in the screen located below the *search* bar.

- **Experiment Groups:** This section enables sample grouping, which can be facilitated by typing a sample name (i.e. Sample 1) or a sample label (i.e. Light) into the *search* bar. All samples/labels that are identified in the search will be listed together at the top of the screen for easy subsequent grouping into the screen on the right.

  After selecting the sample(s) to be grouped, add to a new group by clicking the [icon] button or to an existing group by clicking the [icon] button and selecting that existing group from the drop-down menu. All grouped samples can be displayed or hidden by selecting the **Expand** and **Collapse** buttons, respectively. Samples can be removed individually or all at once by clicking the [icon] button or by clicking the **Remove All**, respectively.

- **Reference Label:** Use this drop-down menu to select a reference label. The peptide area corresponding to this reference label will be the denominator in all feature area ratio calculations. Only a label that is already listed in the Experiment Groups screen can be selected as a reference. Furthermore, if the samples/labels have been grouped in the right-side screen, a **Base Group** title will appear at the bottom-right of the experiment window and state the reference group, which is the group that contains the reference label.

- **Experiment Setting:** Changes in this section can only be made when "All Experiments" is selected from the above drop-down menu. These options dictate which samples/labels the normalization options apply.

- **Perform Inter Experiment Normalization:** Check this box if a spike label was used in the experiment. This is used to allow for quantification between experiments.

- **Spike Channel Drop-down Menus:** Select the channel in each sample that will be used for normalization. The software assumes that similar peptide features containing this label across experiments will approach a 1:1 ratio. Differences are assumed to be due to systematic error and all labeled channels are normalized to this spike label.

- **Exclude Spike Label for Significance:** Since the spike label is assumed to be in a 1:1 ratio between experiments, it should not be included in the significance calculation.

The following parameter dialogue appears when the **Normalization** button is selected.



## Note

If there are multiple samples in the quantification result and all samples are selected, **No normalization** is chosen as the default option and will be the only normalization option unless **Perform Inter Experiment Normalization** is check-marked in Experiment Settings.

- **Auto normalization:** This option will calculate a global ratio from the total intensity of all labels. Factor values and Expected Ratios between these channels (the labels) will be presented. The Experiment drop-down menu can be opened to display the results of another sample.

- **Manual input:** This option performs the same default calculation as Auto normalization but also allows for the manual change of each label's Expected Ratios. After changing a ratio and hitting Enter on the keyboard. PEAKS will recalculate the Factor for the non-referenced label(s). The Experiment drop-down menu can be opened to display the results of another sample.

- **Normalize to spike:** This option displays a list of identified proteins, which can be selected for normalization factor calculation. A search bar is included to enable the quick location of selected proteins. Right-clicking in the Protein screen will open a context menu that will allow for selection of the highlighted proteins or the removal of all check-marked proteins. The expected ratios can be manually modified. It is assumed that the total protein content of the selected proteins in different labels have the ratios entered in the 'Expected ratio' column.
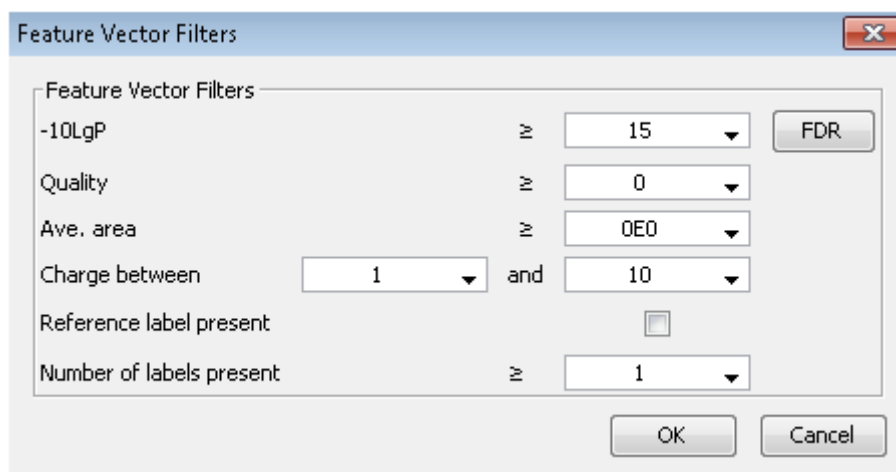


If desired, sample names can be changed by double-clicking the chosen channel below the "Name" column and typing the new name.

Select **Apply** within the filtration window of the "Summary" view to include all Experiment Settings and Normalization choices in the analysis.

### 3.1.1.2. Feature Vector Filter

The following parameter dialogue appears when the **Edit…** button is selected.
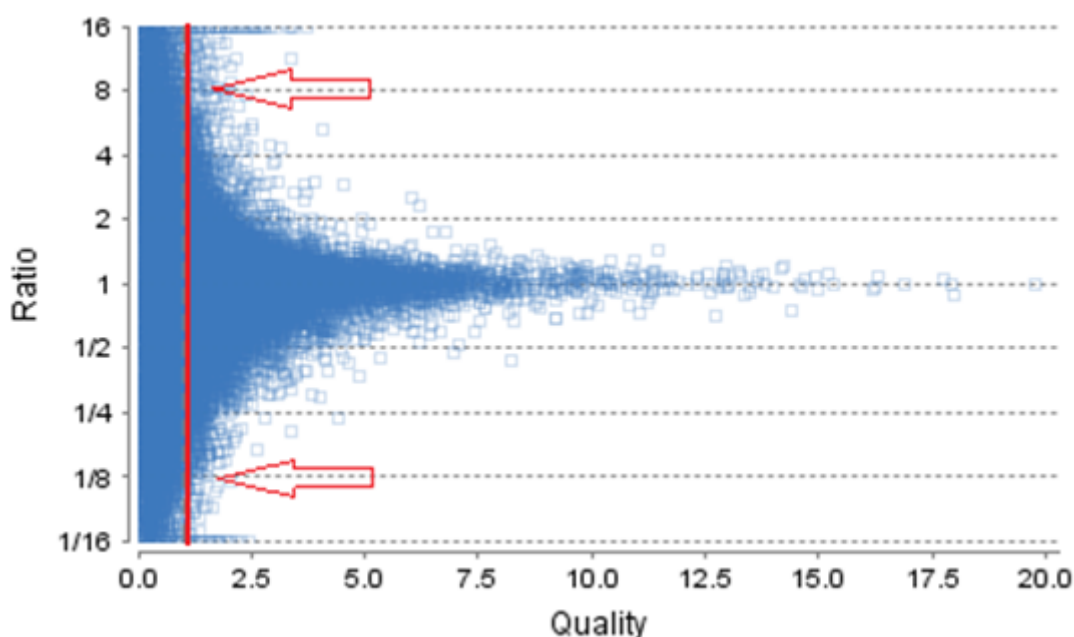
- **-10LgP:** Only peptides with a score above this threshold are used to quantify the identified proteins. PEAKS Q will not allow the user to set a value less the value selected when setting up the search. Click the FDR button to set a -10LgP value equivalent to a selected FDR.

- **Quality:** A higher quality peptide indicates that the peptide is more quantifiable. Factors that affect the quality score include m/z difference, RT difference, XIC shape similarity, and the feature intensities. An appropriate threshold value can be selected by checking the ratio-quality figure. The rule of thumb is to set the quality threshold around a fold change of 8, as shown in this labelled version of Figure 3(a) of the "Summary" view. Only feature vectors with a quality score above this threshold are used to calculate the quanlity score of a peptide



.

- **Ave. area:** This filter is directly applied to the MS signal intensity. Higher intensity peptide features have been shown to be more reproducible across replicates. The rule of thumb is to choose the average area near a fold change of 8, as shown in this labelled version of Figure 3(b) of the "Summary" view. Only peptides with at least one label above this threshold will be used in protein quantification.

- **Charge between:** Only peptides with feature vectors that all fall within this range will be used in protein quantification.

- **Reference label present:** Only quantifiable peptides in which the reference label is present will be used in protein quantification.

- **Number of labels present:** Only quantifiable peptides which contain at least the number of labels selected will be used in protein quantification.

Select **Apply** within the filtration window of the "Summary" view to include all selected filters in the analysis.

### 3.1.1.3. Protein Filter

The following parameter dialogue appears when the **Edit…** button is selected. These filters determine what will be included in a protein group significance calculation, as well as what significance score thresholds will be set. For a protein group to have a significance score, it must have been identified in a sample/label that has already been placed into a group (in either the Tools | Quantification window or in Experiment Settings).



- **Significance:** Only protein groups with a significance above this threshold will be listed in the "Protein" view. Either this or "FDR" can be selected to set a significance threshold. A significance score threshold of 20 is recommended.

- **FDR:** Adopt Benjamin–Hochberg method to adjust the p-value to the false discovery rate for all protein groups that have already passed the other filters. Only protein groups with significance scores passing the calculated FDR will be listed in the "Protein" view. Either this or "Significance" can be selected to set a significance threshold.

- **Fold Change:** Only protein groups at or above this fold change threshold will be listed in the "Protein" view.

- **Unique Peptides:** Only protein groups with at least this many unique peptides will be listed in the "Protein" view.

- **Modified Form Exclusion:** Peptides that are modified with a variable PTM will not be included in the significance calculation for a protein group. This does not apply to the quantification labels.

- **Significance Method:** One of two significance calculations can be applied to the list of protein groups: PEAKS Q or ANOVA. Scores will only be calculated for protein groups identified from samples that are in groups. PEAKS Q is similar to the Significance B method employed by the paper referenced below. Additionally, support for peptide similarity is considered. ANOVA requires that at least two samples be included in each particular group. The ANOVA method employed is also referenced below.

## Note

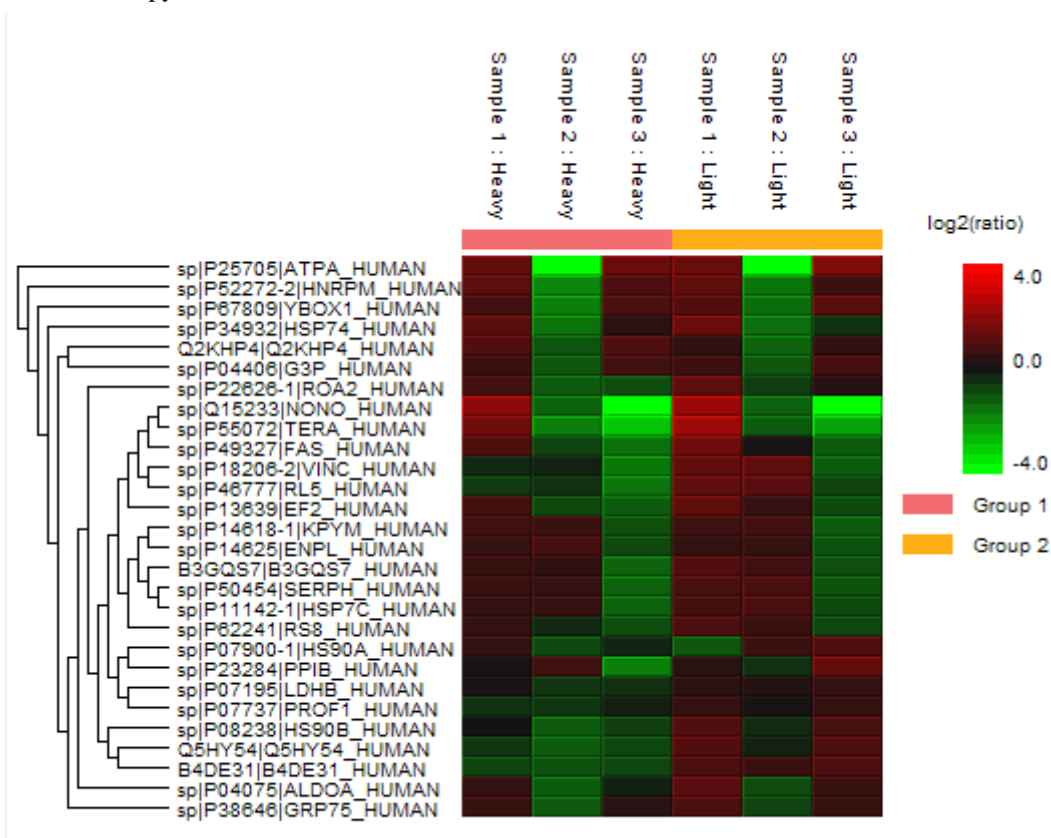For more **details on PEAKS Q**, refer to the following paper: *"MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol.* 2008, 26(12): 1367-1372.

For more **details on ANOVA**, refer to the following paper: *"On the comparison of several mean values: an alternative approach" Biometrika.* 1951, 38(3/4): 330-336.
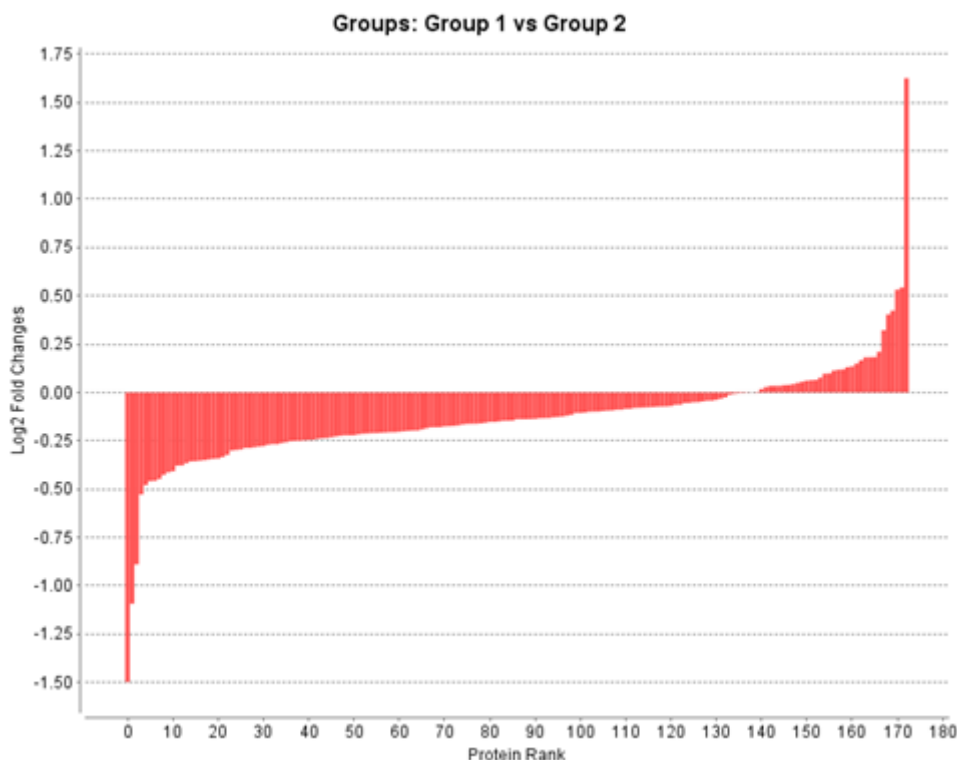
## 3.1.2. Summary Report

### 3.1.2.1. Heat Map

The heat map displays the protein groups that passed the filters. The relative protein abundance is represented as a heat map of the representative proteins of each protein group. These representative proteins are clustered if they exhibit a similar expression trend across the samples. The hierarchical clustering is generated using a neighbour-joining algorithm with a Euclidean distance similarity measurement of the log2 ratios of the abundance of each sample relative to the average abundance. An interactive protein profile heat map is opened in a new window when the **Heat Map** button that is located in the filtration options. Right-click on the heat map in this window to save a copy of it.



In the interactive heat map (by clicking the Heat Map button), move the cursor over any point to highlight the protein accession name and the sample/label to which it matches. Right-click on the heat map and click **Save as…** to export a PNG file of the heat map.

### 3.1.2.2. Interactive Expression Wide Fold Change Graph

The following bar graph displays relative log2 ratio between two selected groups with respect to the rank of each protein. The log2 function applies to the selected primary group with respect to the selected secondary group. Hold the cursor over any bar in this figure to display the associated fold change of each protein within the rank. Select the drop-down menus to change the primary and secondary groups. Click **Apply** to set the new log2 ratio. Click **Save as…** to export a PNG file of the bar graph.
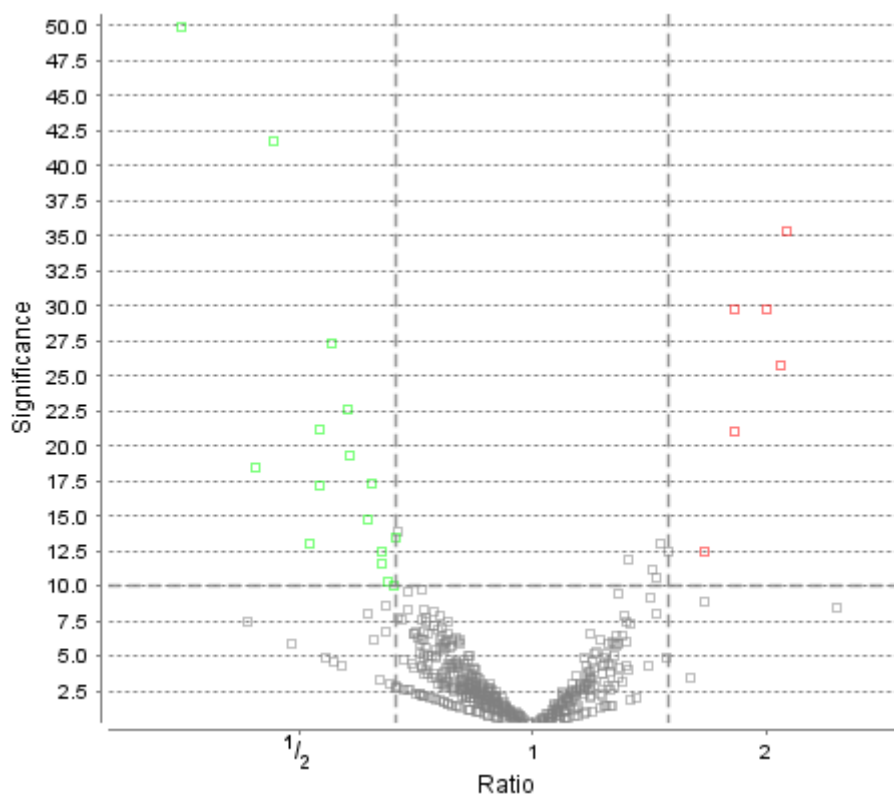
### 3.1.2.3. Volcano Plot

A volcano plot combines a statistical test with the magnitude of the change enabling quick visual identification of those data-points that display large-magnitude changes and that are also statistically significant. The volcano plot in Figure 2(a) plots significance versus fold-change of the quantified proteins.

- **Horizontal broken grey line:** The selected significance threshold

- **Vertical broken grey lines:** The selected fold change thresholds

Plotting points in this way results in two regions of interest in the plot: those points that are found towards the top of the plot that are far to either the left or the right side. These represent values that display large magnitude fold changes (hence being left or right of center) as well as high statistical significance (hence being towards the top). In this figure, the ratio is the group ratio set with respect to the base group. If more than two groups are present, the max ratio is used instead.
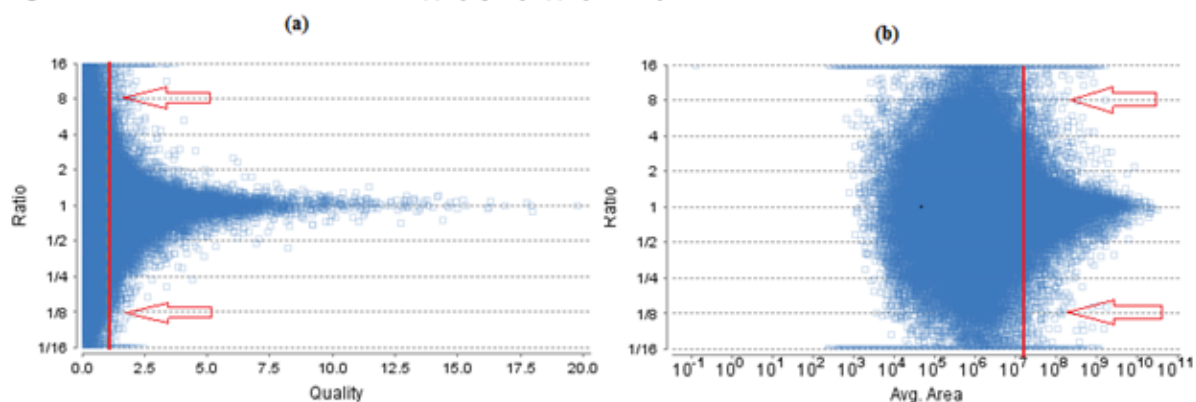
The "Protein" view contains an interactive volcano plot that can be hidden or displayed. Markers for the proteins that are above the set significance threshold will be displayed in color. Scrolling the cursor over a marker will highlight it and, subsequently, clicking on this marker will highlight the protein in the interface.

### 3.1.2.4. Feature Vector Ratio Distribution

Figures 3(a) and 3(b) plot the distribution of feature vector ratios by quality and by average area, respectively. These plots help to determine the threshold values for the quality and intensity filters to obtain a reliable section of the analysis results. The rule of thumb is to set the quality threshold around an 8-fold change as shown below in Figure 3(a). The quality filter is usually the better filter to use but under some circumstances, it may be better to apply the filter to the MS signal directly. The rule of thumb is still to choose the average area around an 8-fold change. In both figures, the ratio is the group ratio with respect to the base group. If more than two groups are present, each ratio value depicts a data point in the figures.

**Figure 3.** The distribution of feature vector ratio: **(a)** By quality. **(b)** By intensity. ❓



### 3.1.2.5. Summary Tables

Tables 1 and 3 list the data and the filtered results, respectively. Table 2 lists the feature vector and protein result filtration parameters.

**Table 1.** Statistics of data.

| | |
|---|---|
| # of MS Scans | 31468 |
| # of MS/MS Scans | 106231 |

**Table 2.** Result filtration parameters.

| | |
|---|---|
| -10LgP | $\geq 15$ |
| Average area | $\geq 1E4$ |
| Charge | [1, 3] |
| Quality | $\geq 2$ |
| Reference label present | true |
| Number of labels present | 1 |
| Protein significance | $\geq 0$ |
| Protein fold change | $\geq 1$ |
| Unique peptides | $\geq 3$ |
| Modified From Exclusion | false |
| Significance method | PEAKSQ |

**Table 3.** Statistics of filtered result.

| | |
|---|---|
| Feature vector groups | 3538 |
| Feature vectors | 9699 |
| Features | 23021 |
| Peptide Sequences | 7155 |
| Protein Groups | 370 |

Tables 4 and 5 list the quantification and instrument parameters, respectively.

**Table 4.** Quantification parameters.

Quantification Type: ICAT/SILAC
Quantification Mass Tolerance: 15.0ppm
Quantification RT Range: 0.5min.
-10LogP Threshold: 15.0
Reporter Ion Type: MS2
R to P Convert enabled: true
Light: R (100.0%) K (100.0%)
Heavy: R(+10.01) (100.0%) K(+8.01) (100.0%)
Group Settings in Parameters:
Group 1: Sample 1 : Heavy
Group 1: Sample 2 : Heavy
Group 1: Sample 3 : Heavy
Group 2: Sample 1 : Light
Group 2: Sample 2 : Light
Group 2: Sample 3 : Light
DB Dependency: 5
Current Group Information:
Group 1: Sample 1 : Heavy
Group 1: Sample 2 : Heavy
Group 1: Sample 3 : Heavy
Group 2: Sample 1 : Light
Group 2: Sample 2 : Light
Group 2: Sample 3 : Light

**Table 5.** Instrument parameters.

Fractions: 20120812_ShAr_QE1_ARL_NX4_mix_sample_1.raw, 20120812
_ShAr_QE1_ARL_NX4_mix_sample_2.raw, 20120817_ShAr_QE1_ARL_NX4
_mix_sample_2_REP_3.raw
Ion Source: ESI(nano-spray)
Fragmentation Mode: CID, CAD(y and b ions)
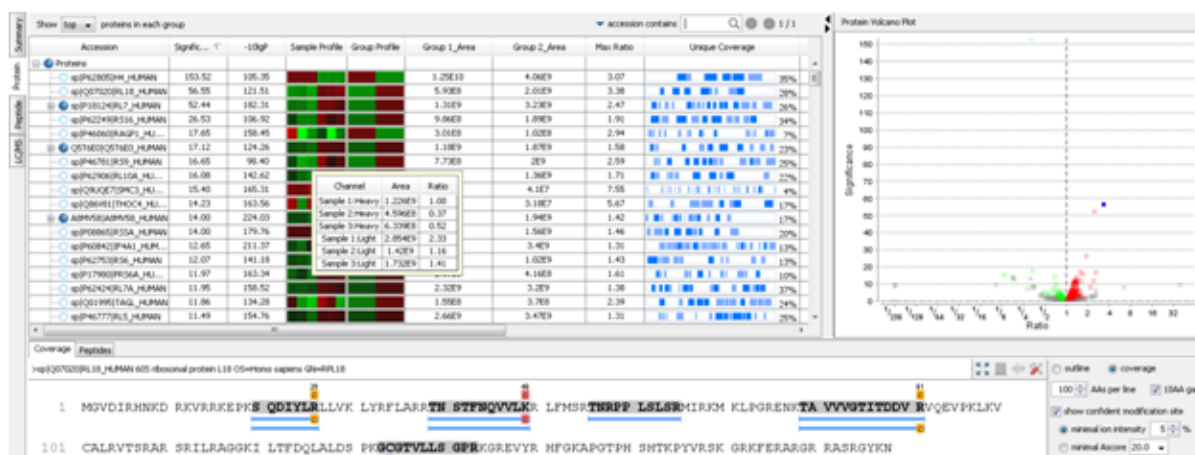MS Scan Mode: FT-ICR/Orbitrap
MS/MS Scan Mode: Linear Ion Trap

## 3.1.3. Result Exporting

PEAKS precursor ion quantification results can be exported to various HTML and text formats by clicking the Export button within the filtration options in the "Summary" view. The "Summary" view and the images from various views can also be exported to various HTML and text formats. Refer to Section 5, "Exporting Quantification Results", for details.

# 3.2. Protein View

The "Protein" view lists all the quantified proteins present in the sample, characterizes each protein at the amino acid level, and lists the supporting peptide features of each protein. It has three components:

- **Protein Table:** This lists all the quantified proteins present in the sample.

- **Coverage:** This characterizes the protein sequence at the amino acid level. All the PTMs and mutations occurring on the protein sequence will be displayed together with the MS/MS data supporting the inference.

- **Peptides:** This lists the peptides identified from the selected protein in the Protein table.



## 3.2.1. Protein Table

The quantified proteins are listed in the protein table with homologous proteins grouped together. This protein table is the same as the other Protein tables found in the PEAKS DB, PEAKS PTM, and SPIDER results, but includes the following additions:

- **Significance:** The protein significance is set according to the rules set in the "Summary" view. Refer to Section 3.1.1.3.

- **Sample Profile:** The absolute protein abundance among the samples is depicted as a heat map. Hold the cursor on a profile to view the sample channels (labels), areas, and ratios with respect to the reference label selected in Experiment Settings of the "Summary" view.

- **Group Profile:** The absolute protein abundance among the groups is depicted as a heat map. Hold the cursor on a profile to view the group channels (labels), areas, and ratios with respect to the group which contains the reference label selected in Experiment Settings of the "Summary" view.

See Section 3.3, "Protein View", "Protein Table" for more details on other entries of the Protein table.

## 3.2.2. Coverage Tab

The Coverage tab characterizes the selected protein sequences at the amino acid level. See Section 3.3.2, "Protein Coverage" , "Protein Coverage" for details.

## 3.2.3. Peptides Tab

The supporting peptides assigned to the protein are shown under the "Peptides" tab.

This table is similar to the peptide table in the "Peptide" view.

- **(Checkbox):** Select the peptides to appear within the Coverage Tab as blue bars.

- **Used:** Select whether this peptide is used to calculate the protein profile. Select one and right click on the column to show the pop-up menu. Use "Reset the used selection" to reset the used setting for all proteins.

- **Unique:** This shows which peptides are only assigned to the current protein group.

- **Quality Score:** The quality score of the peptide based on m/z difference, RT difference, XIC shape similarity and feature intensities. The higher the quality of a peptide, the more quantifiable it is.

- **Sample Profile:** The absolute peptide abundance among the samples is depicted as a heat map. Hold the cursor on a profile to view the sample channels (labels), the areas, and the ratios with respect to the reference label selected in Experiment Settings of the "Summary" view.

- **Group Profile:** The absolute peptide abundance among the groups is depicted as a heat map. Hold the cursor on a profile to view the group channels (labels), areas, and ratios with respect to the group which contains the reference label selected in Experiment Settings of the "Summary" view.

- **Group Area:** The average area of peptide features in the group. The number of group area columns will equal the number of groups.

- **Max Ratio:** Maximum calculated ratio with respect to the given base group.

- **#Vector:** The number of quantifiable feature vectors included in the peptide ratio calculation. A vector is considered to be a group of peptide features across samples and replicates used to create a peptide ratio. For example, if there are three replicates in your experiment,you will have a maximum of three vectors.

- **Start:** This shows the position of the first residue of the peptide that covers the protein.

- **End:** This shows the position of the last residue (inclusive) of the peptide that covers the protein.

To view the actual abundance of the peptide features among the samples or groups, move the cursor over the corresponding profile. A tool-tip will display the actual peptide feature intensity along with the ratio. Double-click on the peptide feature to jump to the "Peptide" view. Right-click on the peptide feature to show a pop-up menu similar to the one in the PEAKS DB supporting peptides tab.

## 3.3. Peptide View

The "Peptide" view displays the quantified peptides and their detailed PSM matching and feature information. The interface is similar to the "Peptide" view in a PEAKS DB result. See Section 3.4, "Peptide View", "Peptide View" for details.

A. **Peptide Table:** This contains the important details and statistics associated with each quantified peptide that passes the filters in the "Summary" view. The peptide table is associated with the corresponding letter A. located in the previous screenshot.

> ### Note
>
> Right-click on a peptide in the table to 'copy the peptide sequence', 'show spectrum in LC/MS', or 'show original spectrum'

- **-10lgP:** Identification -10lgP score that indicates the likelihood of a true identification. The score corresponds to the -10LgP score of the best PSM associated with a feature from all quantifiable feature vectors of this peptide.

- **Quality Score:** The higher the quality of a peptide indicates that the peptide is more quantifiable. Factors that affect the quality score include the m/z difference, RT difference, XIC shape similarity, and feature intensities.

- **Sample Profile:** The absolute peptide abundance among the samples is depicted as a heat map. Hold the cursor on a profile to view the sample channels (labels), the areas, and the ratios with respect to the reference label selected in Experiment Settings of the "Summary" view. Hold the cursor on a profile to view the sample channels (labels), areas, and ratios with respect to the reference label selected in Experiment Settings of the "Summary" view.

- **Group Profile:** The absolute peptide abundance among the groups is depicted as a heat map. Hold the cursor on a profile to view the group channels (labels), areas, and ratios with respect to the group which contains the reference label selected in Experiment Settings of the "Summary" view. Hold the cursor on a profile to view the sample channels (labels), areas, and ratios with respect to the group that contains the reference label selected in Experiment Settings of the "Summary" view.

- **Group Area:** The average area of peptide features in the group. The number of group area columns will equal the number of groups.

- **Max Ratio:** The maximum calculated ratio with respect to the given base group.

- **#Vector:** The number of quantifiable vectors included in the peptide ratio calculation. A vector is considered to be a group of peptide features across samples and replicates used to create a peptide ratio. For example, if there are three replicates in your experiment, you will have a maximum of three vectors.

- **Accession:** The protein with the highest -10lgP that the peptide supports. If no proteins that the peptide supports pass the protein filters, the accession field will be empty.

The bottom screen of the "Peptide" view contains four interactive windows that are organized according to their letters in the previous screenshot.

B. **MS/MS View:** The tandem scan that corresponds to the selected label set at the Current PSM Label drop-down menu. Click the drop-down window to change the PSM Label. This change will be reflected in the LC/MS View, the XIC, and the MS View. See Section 3, "MS/MS View", "MS/MS View" for details.

C. **LC/MS View:** This LC/MS screen is zoomed into the features that correspond to the isotopic labels, i.e. "Light" and "Heavy". The m/z between the markers for these features should equate to the mass difference of the two labels in this sample. If R-to-P correction is applied (See Chapter 12, Section 2 for details), peptides with heavy proline labels will be highlighted to the right of the normal heavy label in this view. See Section 4, "LC/MS View", "LC/MS View", for details.

D. **MS View:** The survey scan represents the scan at a selected retention time on the LC/MS View. By default, the parent scan of the tandem scan (in the MS/MS view) is displayed. View other neighbouring survey scans by clicking on the horizontal arrows to move along the retention time axis. The retention time position will be adjusted in the LC/MS view and the XIC. Click the curved arrow to return to the default survey scan. See Section 2, "MS View" for details.

E. **XIC:** This map presents the changes in feature intensity over the retention time range of the peaks for all labels of the selected peptide. Additionally, the coloured isotopic distribution plot – M0, M+1, M+2, and REST – of each label is included, which shows the relative composition of each label.

Select **All Vectors** to open a window that displays the feature vectors that correspond to a quantifiable peptide above the Feature Vector Filters assigned in the "Summary" view. The -10lgP in the upper screen of the "Peptide" view corresponds to the feature vector with the highest -10lgP score above the filters. The quality score in the upper screen corresponds to the average of the quality scores for all feature vectors above the filters. Scroll the cursor over the Sample Profile to see the m/z, the RT, the area, and ratio of the labels for each feature. The Max Ratio corresponds to the ratio of the largest labelled area to the smallest labelled area, regardless of which label was set as the reference in the "Summary" view.

| | Sample | Fraction | -10lgP | Quality Score ↓ | Sample Profile | Light_Area | Heavy_Area | Max Ratio | m/z | z | ppm | RT | AScore |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Sample 2 | F 2 | 86.44 | 3.47 | | 8.36E7 | 8.18E7 | 1.02 | 868.4636 | 3 | 0.7 | 106.62 | |
| 2 | Sample 3 | F 3 | 81.25 | 3.29 | | 1.05E8 | 9.68E7 | 1.09 | 868.4612 | 3 | -2.0 | 106.83 | |
| 3 | Sample 1 | F 1 | 101.76 | 2.54 | | 1.8E7 | 1.35E7 | 1.33 | 1298.1837 | 2 | 0.1 | 106.62 | |
| 4 | Sample 2 | F 2 | 92.54 | 2.31 | | 2.1E7 | 1.81E7 | 1.16 | 1298.1854 | 2 | 1.4 | 106.63 | |

Select the Protein button to go to the Protein tab. The protein to which this peptide corresponds will be highlighted in the tab. If this peptide identifies more than one protein, a new window will open and show all of the corresponding proteins. Select the desired protein to go to the Protein tab.

# 3.4. LC/MS View

This view displays the distribution of LC-MS signals/features in the fractions. See Section 4, "LC/MS View" for details on LC/MS View.

# Chapter 13. Reporter Ion Quantification (e.g. iTRAQ and TMT)

## 1. Overview

Reporter ion quantification with chemical labels at the MS2 (MS3) level is one of the three quantification modes that are supported by the optional PEAKS Q module of PEAKS Studio. It is based on the relative intensities of the fragment peaks at fixed m/z values within an MS2 (MS3) spectrum. In this mode, chemical labels with the same mass are introduced to several samples. The samples are then analyzed together in an LC-MS/MS experiment. The same peptides from different samples will have the same precursor m/z and are fragmented together. In the MS/MS (MS3) scans, labels from different samples will produce different reporter ions, which can then be used to calculate the quantification ratio between samples. Both user-defined and commercial labels (i.e. iTRAQ and TMT) are supported by PEAKS Q.

Reporter Ion Quantification can be launched from either the Project Wizard Workflow (see Chapter 4) or on an existing PEAKS DB result (See Chapter 9, *Peptide, PTM, and Mutation Identification (PEAKS DB, PEAKS PTM, and SPIDER)*). When running quantification on an existing result, ensure that the specified labels are listed as fixed PTMs on the database search result against which its referencing. After the database search is complete, follow these steps:

1. Select a PEAKS DB (PEAKS PTM or SPIDER) result node in the Project view. Click the Quantification toolbar

   icon, , or select **Tools | Quantification** command from the menu..

   ### Note

   > Refer to Chapter 4, *Project Wizard - Create a PEAKS Project, Load Data and Set up a Workflow* for how to perform PEAKS DB and quantification in a single easy workflow.

2. Select the quantification protocol, "Reporter Ion Quantification"; specify the PEAKS quantification parameters in the dialogue box on the right; and click OK.

3. Wait for the analysis to finish. A new quantification result node, , will appear in the "Project" view. Double click the node to open this result node.

The quantification results with labels are displayed in a format that is similar to the PEAKS DB result, A ratio is added to each quantifiable peptide and protein.

## 2. Setting Parameters

The following parameter dialogue pops up when the quantification tool bar icon, , is clicked or the **Tools | Quantification** command from the menu is clicked. Select "**Reporter Ion Quantification eg. iTRAQ**" under the "**Quantifications**".
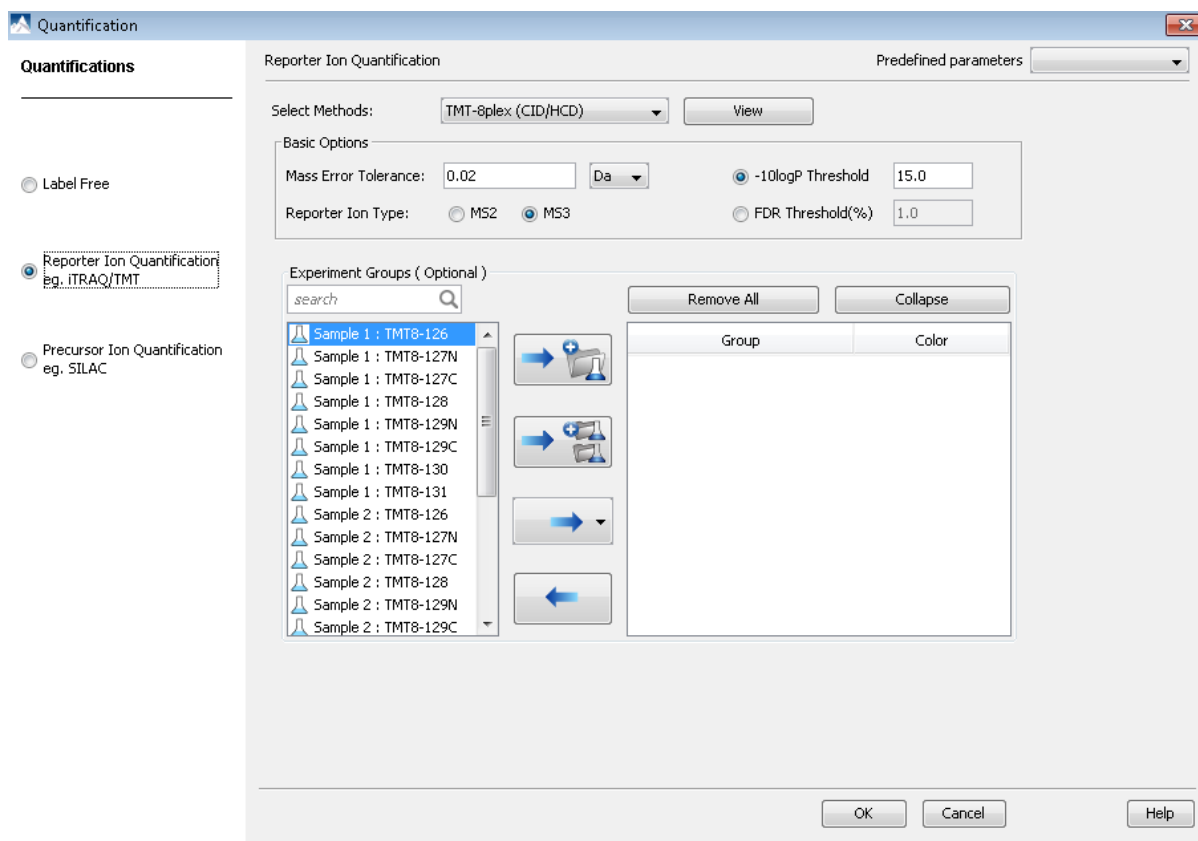
From the **Select Methods** drop-down list, choose the appropriate quantification method used in the experiment. To examine the details of a selected quantification method, click on the **View** button. To create a new quantification method, select **Add New Method** from the **Select Methods** drop-down menu. Refer to Chapter 17, Section 2.3, "Labeled Q Method Configuration" for how to create or edit a quantification method:

- **Mass Error Tolerance:** This parameter is used to locate the reporter ion peaks in the MS/MS scan. It is recommended to set the mass error tolerance a little wider than the fragment ion error tolerance used in the PEAKS DB search. Either Da or ppm can be selected as the mass unit.

- **Reporter Ion Type:** This parameter allows for the selection of MS2 or MS3 as the reporter ion type for quantification.

- **-10lgP Threshold:** Only peptides with a score above this threshold are used to quantify the identified proteins. Either this option or the FDR Threshold can be selected.

- **FDR Threshold(%):** Only peptides with a score above this threshold are used to quantify the identified proteins. Either this option or the -10logP Threshold can be selected.

- Quantification allows for the option of sample grouping. For experiments with multiple samples, the *search* bar can be used to look for a specific sample. After selecting a sample, it can be added to a new group by clicking the ![button] button or to an existing group by clicking the ![button] button and selecting that option from the drop-down menu. Display all samples in each group or just the individual groups by selecting the **Expand** and **Collapse** buttons, respectively. Samples can be removed individually or removed all at once by clicking the ![button] button or the **Remove All** button, respectively.

- **Predefined parameters:** This drop-down menu enables the saving, as well as the previously saved parameters, for the ease to perform quantification regularly with the same parameters.

# 3. Understanding the Results

Once completed, the protein quantification result will be displayed in the "Project View" as a quantification node, ⚿ . Double-click on this node to open the result that contains four views: "Summary" view, "Protein" view, "Peptide" View, and "LC/MS" view. The "Summary" view will appear by default.

## 3.1. Summary View

The "Summary" view provides three main functions:

- **Result Filtration:** This is achieved by specifying the filtration rules in the area at the top of the "Summary" view. The filtration function is discussed in Section 3.1.1, "Result Filtration".

- **Summary Report:** Several statistical charts are displayed to provide an overall picture of the results, assess the result quality, and examine the reliability and reproducibility of the experiment. This function is discussed in Section 3.1.2, "Summary Report".

- **Result Exporting:** This is achieved by clicking the "Export" button at the top of the "Summary" view. The exporting function is discussed in Section 3.1.3, "Exporting Quantification Results".

The "Summary" view includes an expression profile with candidate proteins assorted in a heat map, result statistics in both figures and tables, and a list of filtration parameters. To add a summary note, click on the Notes button at the top-right of the "Summary" view to open a "Notes Entry" editor in which you an edit the notes. These notes will be displayed in the "Summary" view.

### 3.1.1. Result Filtration

PEAKS provides an extensive set of tools to filter Reporter Ion Quantification results according to the users' requirements. There are three categories of options in the "Summary" view: **Experiment**, **Spectrum Filter**, and **Protein Filter**.



#### 3.1.1.1. Experiment

There are two buttons under this heading - **Experiment Settings** and **Normalization** - which allow for control over the analysis.

The following parameter dialogue appears when the **Experiment Settings** button is selected.

- **Select Experiment:** When running quantification on an identification node that searched data within multiple samples (i.e. replicate experiments within different treatments), this drop-down menu allows for the selection of a single sample or every sample (All Experiments) for an analysis. Every label of the sample(s) selected will appear in the screen located below the search bar.

- **Experiment Groups (Optional):** This section enables sample grouping, which can be facilitated by typing a sample name (i.e. Sample 1) or a sample label (i.e. TMT-126) into the search bar. All samples/labels that are identified in the search will be listed together at the top of the screen for easy subsequent grouping into the screen on the right. After selecting the sample(s) to be grouped, add to a new group by clicking the  button or to an existing group by clicking the  button and selecting that existing group from the drop-down menu. All grouped samples can be displayed or hidden by selecting the **Expand** and **Collapse** buttons, respectively. Samples can be removed individually or all at once by clicking the  button or by clicking **Remove All**, respectively.

- **Reference Label:** Use this drop-down menu to select a reference label. The peptide area corresponding to this reference label will be the denominator in all feature area ratio calculations. Only a label that is already listed in the Experiment Groups screen can be selected as a reference. Furthermore, if the samples/labels have been grouped in the right-side screen, a **Base Group** title will appear at the bottom-right of the experiment window and state the reference group, which is the group that contains the reference label.

- **Experiment Setting:** Changes in this section can only be made when "All Experiments" is selected from the above drop-down menu. These options dictate which samples/labels the normalization options apply.

- **Perform Inter Experiment Normalization:** Check this box if a spike label was used in the experiment. This is used to enable quantification between experiments.

- **Spike Channel Drop-down Menus:** Select the channel in each sample that will be used for normalization. The software assumes that similar peptide features containing this label across experiments will approach a 1:1 ratio. Differences are assumed to be due to systematic error and all labeled channels are normalized to this spike label.

- **Exclude Spike Label for Significance:** Since the spike label is assumed to be in a 1:1 ratio between experiments, it should not be included in the significance calculation.

The following parameter dialogue appears when the **Normalization** button is selected.



**Note** If there are multiple samples in the quantification run and all samples are selected, **No Normalization** is chosen as the default option and will be the only normalization option unless **Perform Inter Experiment Normalization** is check-marked in Experiment Settings.

- **Auto normalization:** This option will calculate a global ratio from the total intensity of all labels. Factor values and Expected Ratios between these channels (the labels) will be presented. The **Experiment** drop-down menu can be opened to display the results of another sample.

- **Manual input:** This option performs the same default calculation as Auto normalization but also allows for the manual change of each label's Expected Ratios. After changing a ratio and hitting "Enter" on the keyboard. PEAKS will recalculate the Factor for the non-referenced label(s). The **Experiment** drop-down menu can be opened to display the results of another sample.

- **Normalize to spike:** This option displays a list of identified proteins, which can be selected to create a normalization ratio. A *search* bar is included to enable the quick location of select proteins. Right-clicking in the

Protein screen will open a window that will allow for the selection of multiple highlighted rows or the removal of all check-marked proteins. The expected ratios can be manually modified. It is assumed that the total protein content of the selected proteins in different labels have the ratios entered in the 'Expected ratio' column. If desired, sample names can be changed by double-clicking the chosen channel below the "Name" column and typing the new name.



Select **Apply** within the filtration window of the "Summary" view to include all Experiment Settings and Normalization choices in the analysis.

### 3.1.1.2. Spectrum Filter

The following parameter dialogue appears when the **Edit…** button is selected.



- **-10LgP:** Only peptides with a score above this threshold are used to quantify the identified proteins. PEAKS Q will not allow the user to set a value less than he value selected when setting up the search. Click the FDR button to set a -10LgP value equivalent to a selected FDR.

- **Quality:** A higher quality peptide indicates that the peptide is more quantifiable. Factors that affect the quality score include the identification -10LgP score, the noise around the reporter ions, and the mass error of reporter ions. An appropriate threshold value can be selected by checking Figure 3(b): Accumulation curve for quality in the "Summary" view. The rule of thumb is to set the quality threshold near the region of the quality score axis at which the curves begin to significantly drop.



- **Charge between:** Only peptides with feature vectors that all fall within this range will be used in protein quantification.

- **Reference label present:** Only quantifiable peptides in which the reference label is present will be used in protein quantification.

- **Number of labels present:** Only quantifiable peptides that contain at least the number of labels selected will be used in protein quantification.

Select **Apply** within the filtration window of the "Summary" view to include all selected filters in the analysis.

### 3.1.1.3. Protein Filter

The following parameter dialogue appears when the Edit… button is selected. These filters determine what will be included in a protein group significance calculation, as well as what significance score thresholds will be set. For a protein group to have a significance score, it must have been identified in a sample/label that has already been placed into a group (in either the Tools | Quantification window or in Experiment Settings).

- **Significance:** Only protein groups with a significance above this threshold will be listed in the "Protein" view. Either this or "FDR" can be selected to set a significance threshold. A significance score threshold of 20 is recommended.

- **FDR:** Adopt the Benjamin–Hochberg method to adjust the p-value to the false discovery rate for all protein groups that have already passed the other filters. Only protein groups with significance scores passing the calculated FDR will be listed in the "Protein" view. Either this or "Significance" can be selected to set a significance threshold.

**Note**

For more details on the Benjamin-Hochberg FDR method, refer to the following paper: *"Controlling the false discovery rate: a practical and powerful approach to multiple testing." J. R. Statist. Soc. 1995, B57, 289-300.*

- **Fold Change:** Only protein groups at or above this fold change threshold will be listed in the "Protein" view.

- **Unique Peptides:** Only protein groups with at least this many unique peptides will be listed in the "Protein" view.

- **Modified Form Exclusion:** Peptides that are modified with a variable PTM will not be included in the significance calculation for a protein group. This does not apply to the quantification labels.

- **Significance Method:** One of two significance calculations can be applied to the list of protein groups: PEAKS Q or ANOVA. Scores will only be calculated for protein groups identified from samples that are in groups. PEAKS Q is similar to the Significance B method employed by the paper referenced below. Additionally, support for peptide similarity is considered. ANOVA requires that at least two samples be included in each particular group. The ANOVA method employed is also referenced below.

**Note**

For more details on PEAKS Q, refer to the following paper: *"MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification." Nat Biotechnol. 2008, 26(12): 1367-1372.*
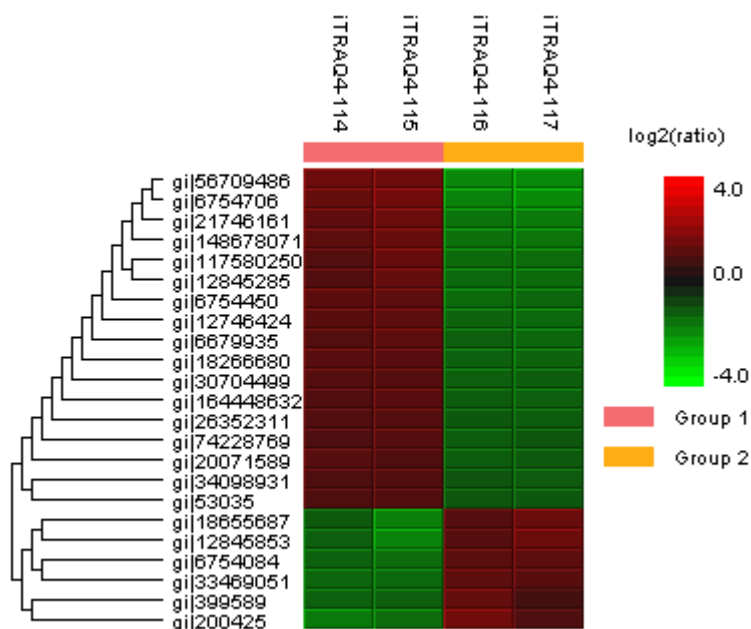
For more details on ANOVA, refer to the following paper: *"On the comparison of several mean values: an alternative approach" Biometrika. 1951, 38(3/4): 330-336.*

## 3.1.2. Summary Report

### 3.1.2.1. Heat Map

The heat map displays the protein groups that passed the filters. The relative protein abundance is represented as a heat map of the representative proteins of each protein group. These representative proteins are clustered

if they exhibit a similar expression trend across the samples. The hierarchical clustering is generated using a neighbour-joining algorithm with a Euclidean distance similarity measurement of the log2 ratios of the abundance of each sample relative to the average abundance. An interactive protein profile heat map is opened in a new window when the Heat Map button that is located in the filtration options. Right-click on the heat map in this window to save a copy of it.



In the interactive heat map (by clicking the Heat Map button), move the cursor over any point to highlight the protein accession name and the sample/label to which it matches. Right-click on the heat map and click Save as… to export a PNG file of the heat map.

### 3.1.2.2. Interactive Expression Wide Fold Change Graph

The following bar graph displays the relative log2 ratio between two selected groups with respect the rank of each protein. The log2 function applies to the selected primary group with respect to the selected secondary group. Hold the cursor over any bar in this figure to display the associated fold change of each protein within the rank. Click the drop-down menus to change the primary and secondary groups. Click Apply to set the new log2 ratio. Click **Save as…** to export a .PNG file of the bar graph.
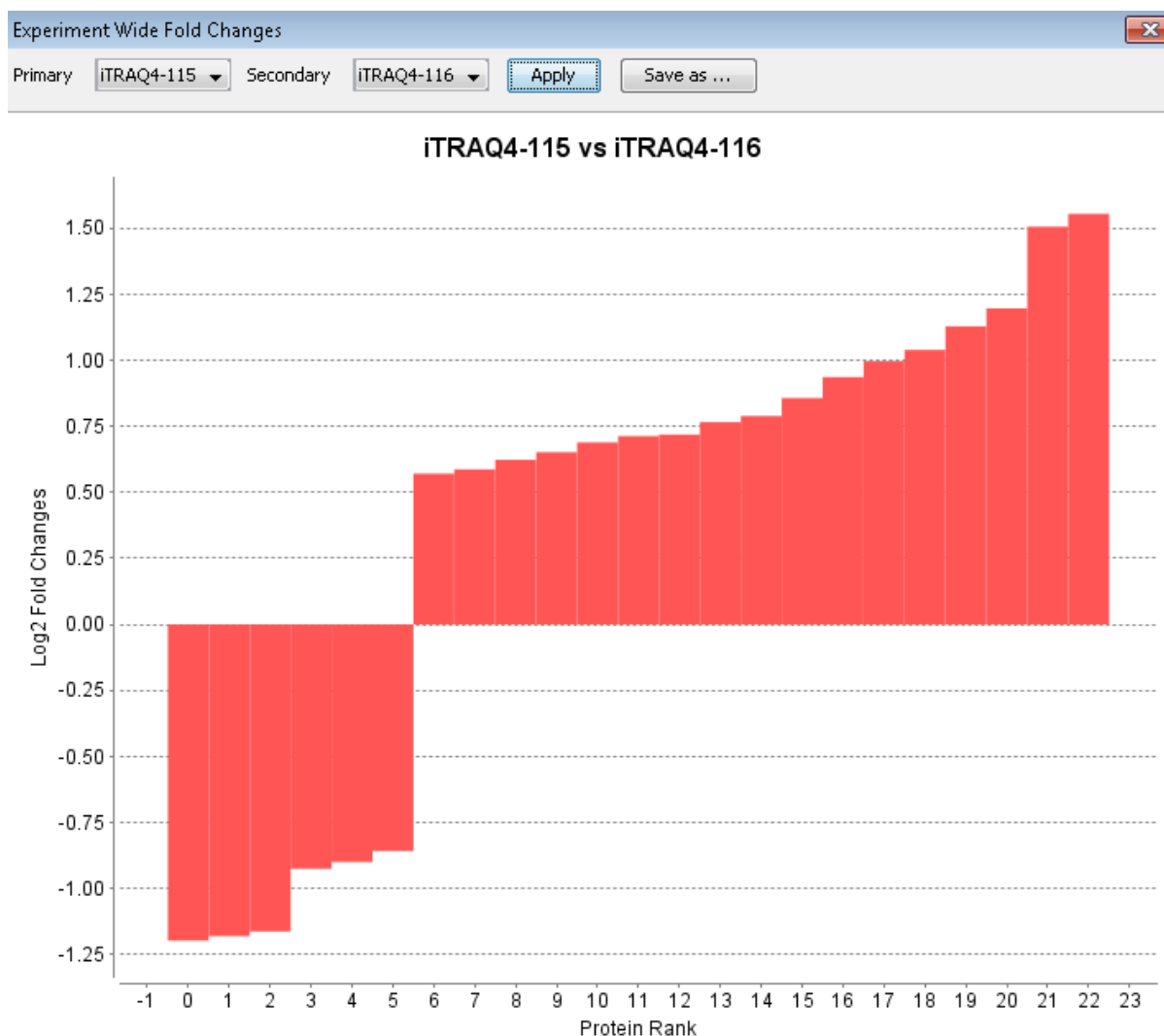
### 3.1.2.3. Box Plot

An interactive box plot (by clicking the Box Plot button) displays a distribution of the intensity values for each label in the experiment. A plot displaying the average distribution for all labels is also included.

### 3.1.2.4. Volcano Plot

A volcano plot combines a statistical test with the magnitude of the change enabling quick visual identification of those data-points that display large-magnitude changes and that are also statistically significant. The volcano plot in Figure 2(a) plots significance versus fold-change of the quantified proteins.

- **Horizontal broken grey line:** The selected significance threshold.

- **Vertical broken grey lines:** The selected fold change thresholds

Plotting points in this way results in two regions of interest in the plot: those points that are found towards the top of the plot that are far to either the left or the right side. These represent values that display large magnitude fold changes (hence being left or right of center), as well as high statistical significance (hence being towards the top). In this figure, the ratio is the group ratio set with respect to the base group. If more than two groups are present, the max ratio is used instead.

The "Protein" view contains an interactive volcano plot that be hidden or displayed. Markers for the proteins that are above the set significance threshold will be displayed in color. Scrolling the cursor over a marker will highlight it and, subsequently, clicking on this marker will highlight the protein in the interface.

### 3.1.2.5. Quality Control Statistics

Figure 3 displays the box plot and the accumulation quality curve described in Sections 3.1.2.3 and 3.1.1.2, respectively.

**Figure 3.** Quality control statistics. **(a)** Box plot for each channel. **(b)** Accumulation curve for quality.
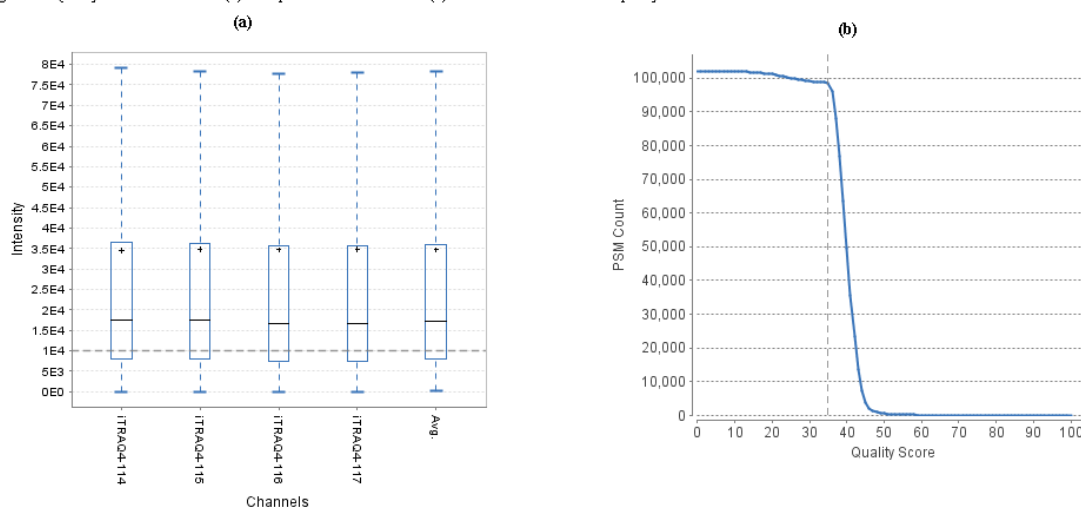
### 3.1.2.6. Summary Tables

Tables 1 and 3 list the data and the filtered results, respectively. Table 2 lists the feature vector and protein result filtration parameters.

**Table 1.** Statistics of data.

| # of MS Scans | 129931 |
|---|---|
| # of MS/MS Scans | 410708 |

**Table 2.** Result filtration parameters.

| Peptide significance | $\geq 0$ |
|---|---|
| Peptide fold change | $\geq 1$ |
| Report ion intensity | $\geq 1E4$ |
| Quality | $\geq 35$ |
| Reference channel present | false |
| Number of channels present | 1 |
| Protein significance | $\geq 10$ |
| Protein fold change | $\geq 1.5$ |
| Unique peptides | $\geq 1$ |
| Significance method | PEAKSQ |

**Table 3.** Statistics of filtered result.

| Peptide-Spectrum Matches | 67565 |
|---|---|
| Peptide Sequences | 8650 |
| Protein Groups | 23 |

Tables 4 and 5 list the quantification and instrument parameters, respectively.

**Table 4.** Quantification parameters.

Quantification Type: iTRAQ-4plex
Quantification Mass Tolerance: 0.2Da
FDR Threshold(%): 1.0
Reporter Ion Type: MS2
iTRAQ4-114: 114.1107
iTRAQ4-115: 115.1077
iTRAQ4-116: 116.1111
iTRAQ4-117: 117.1144
Group Settings in Parameters:
Group 1: Sample 1 : iTRAQ4-114
Group 1: Sample 1 : iTRAQ4-115
Group 2: Sample 1 : iTRAQ4-116
Group 2: Sample 1 : iTRAQ4-117
DB Dependency: 24
Current Group Information:
Group 1: iTRAQ4-114
Group 1: iTRAQ4-115
Group 2: iTRAQ4-116
Group 2: iTRAQ4-117

**Table 5.** Instrument parameters.

Fractions: 100702V1RS1_Kaja_sample1_frac_1.raw, 100709V1RS1_Kaja_sample1_frac_2.raw, 100712V1RS1_Kaja_sample1_frac_3.raw, 100714V1RS1_Kaja_sample1_frac_4.raw, 100715V1RS1_Kaja_sample1_frac_5.raw, 100716V1RS1_Kaja_sample1_frac_6.raw, 100719V1RS1_Kaja_sample1_frac_7.raw, 100723V1RS1_Kaja_sample1_frac_8.raw, 100726V1RS1_Kaja_sample1_frac_9.raw, 100702V1RS1_Kaja_sample1_frac_10.raw, 100708V1RS1_Kaja_sample1_frac_11.raw, 100709V1RS1_Kaja_sample1_frac_12.raw, 100712V1RS1_Kaja_sample1_frac_13.raw, 100714V1RS1_Kaja_sample1_frac_14.raw, 100715V1RS1_Kaja_sample1_frac_15.raw, 100716V1RS1_Kaja_sample1_frac_16.raw, 100719V1RS1_Kaja_sample1_frac_17.raw, 100723V1RS1_Kaja_sample1_frac_18.raw, 100726V1RS1_Kaja_sample1_frac_19.raw, 100702V1RS1_Kaja_sample1_frac_20.raw, 100708V1RS1_Kaja_sample1_frac_21.raw, 100709V1RS1_Kaja_sample1_frac_22.raw
Ion Source: ESI(nano-spray)
Fragmentation Mode: CID/ECD
MS Scan Mode: FT-ICR/Orbitrap
MS/MS Scan Mode: Linear Ion Trap

## 3.1.3. Result Exporting

PEAKS reporter ion quantification results can be exported to various HTML and text formats by clicking the Export button within the filtration options in the "Summary" view. The "Summary" view and the images from various views can also be exported to various HTML and text formats. Refer to Chapter 16, Section 5.2, "Exporting Reporter Ion Quantification Results" for details.

# 3.2. Protein View

The "Protein" view lists all the quantified proteins present in the sample, characterizes each protein at the amino acid level, and lists the supporting peptide features of each protein. It has four components:

- **Protein Table:** This lists all the quantified proteins present in the sample.

- **Protein Volcano Plot:** Click on a marker that is above the significance threshold that was set in the "Summary" view. Upon selection, the corresponding protein in the protein table will be highlighted.

- **Coverage:** This characterizes the protein sequence at the amino acid level. All the PTMs and mutations occurring on the protein sequence will be displayed together with the MS/MS data supporting the inference.

- **Peptides:** This lists the peptides identified from the selected protein in the protein table.



## 3.2.1. Protein Table

The quantified proteins are listed in the protein table with homologous proteins grouped together. This Protein table is similar to the tables displayed in the PEAKS DB, PEAKS PTM, and SPIDER results, but includes the following additions:

- **Significance:** The protein significance is set according to the rules set in the "Summary" view, refer to Section 3.1.1.3.

- **Sample Profile:** The absolute protein abundance among the samples is depicted as a heat map. Hold the cursor on a profile to view the sample channels (labels), the areas, and the ratios with respect to the reference label selected in Experiment Settings of the "Summary" view.

- **Group Profile:** The absolute protein abundance among the groups is depicted as a heat map. Hold the cursor on a profile to view the group channels (labels), areas, and ratios with respect to the group which contains the reference label selected in Experiment Settings of the "Summary" view.

See Chapter 9, Section 3.3.1, "Protein Table" for more details on the other entries of the Protein table.

## 3.2.2. Protein Volcano Plot

This interactive volcano plot enables the selection of a protein in the table. Protein markers that are above the significance threshold can be selected. Upon selection, the corresponding protein in the Protein Table will be highlighted. Proteins located in the top-right and top-left sections of this plot (above the significance threshold and to the right and left of the fold change thresholds) are considered statistically significant. The volcano plot facilitates identification of these significant proteins.

### 3.2.3. Coverage Tab

The Coverage tab characterizes the selected protein sequences at the amino acid level. See Chapter 9, Section 3.3.2, "Protein Coverage" for details.

### 3.2.4. Peptides Tab

The supporting peptides assigned to the protein are shown below the "Peptides" tab.



This table includes the following columns:

- **(Checkbox):** Select the peptides to appear within the Coverage Tab as blue bars.

- **Used:** Select whether this peptide is used to calculate the protein profile.

- **Unique:** This shows which peptides are only assigned to the current protein group.

- **Quality Score:** The higher quality of a peptide indicates that the peptide is more quantifiable. Factors that affect the quality score include the identification -10LgP score, the noise around the reporter ions, and the mass error of reporter ions. An appropriate threshold value can be selected by checking Figure 3(b): Accumulation curve for quality in the "Summary" view. The rule of thumb is to set the quality threshold near the region of the quality score axis at which the curves begin to significantly drop.

- **Sample Profile:** The absolute peptide abundance among the samples is depicted as a heat map. Hold the cursor on a profile to view the sample channels (labels), the areas, and the ratios with respect to the reference label selected in Experiment Settings of the "Summary" view.

- **Group Profile:** The absolute peptide abundance among the groups is depicted as a heat map. Hold the cursor on a profile to view the group channels (labels), the areas, and the ratios with respect to the group that contains the reference label selected in Experiment Settings of the "Summary" view.

- **#Spec:** The number of quantification spectra included in the peptide ratio calculation. A vector is considered to be a group of peptide features across samples and replicates used to create a peptide ratio. For example, if there are three replicates in your experiment, you will have a maximum of three vectors.

- **Start:** This shows the position of the first residue of the peptide that covers the protein.

- **End:** This shows the position of the last residue of the peptide that covers the protein.

## 3.3. Peptide View

The "Peptide" view displays all the quantified peptides and their intensities. The interface is similar to the "Peptide" view in a PEAKS DB result. See Chapter 9, Section 3.4, "Peptide View" for details. The average quality score of the top quantified spectra of a quantifiable peptide is listed in the spreadsheet. The Sample Profile displays the absolute peptide abundance among the samples as a depicted heat map. Hold the cursor on a profile to view the sample channels (labels), the areas, and the ratios with respect to the reference label selected in Experiment Settings of the "Summary" view.
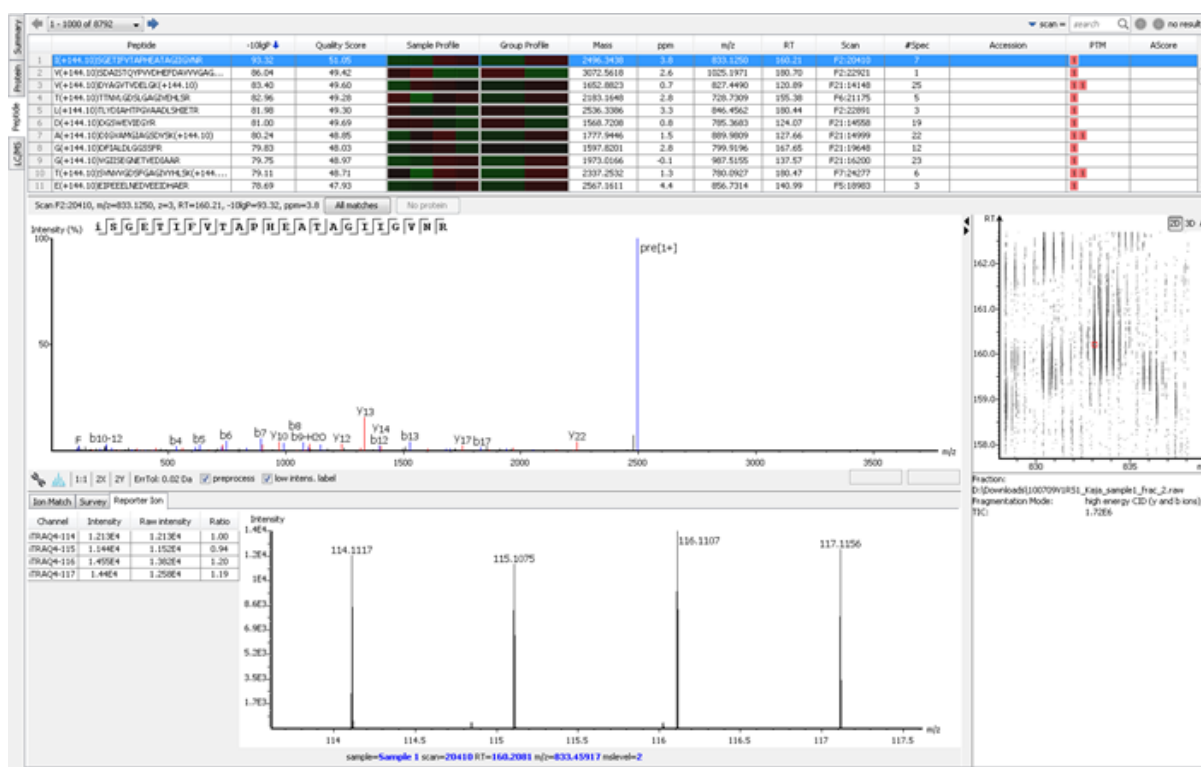
### 3.3.1. Ion Match Table

For details on the Ion Match Table, see Chapter 8, Section 8.1.

### 3.3.2. Survey Scan

For details on the survey scan, see Chapter 5, Section 2.

### 3.3.3. Reporter Ion

This view provides details on characteristics of the reporter ions used in the experiment. The table provides information on the intensity, the raw intensity, and the ratio of each channel with respect to the set reference channel. Additionally, an associated MS2 or MS3 spectrum is included of the reporter ions in the experiment, determined by the selection made in the Quantification parameter settings, see Section 2. This is an interactive spectrum, enabling zooming in and out using the scroll wheel of the mouse. Double-clicking will zoom out and reveal the entire MS3 spectrum.
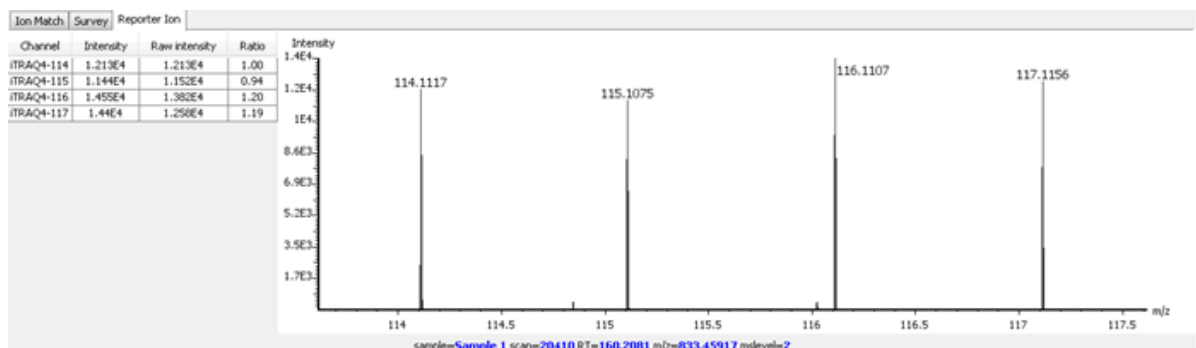


## 3.4. LC/MS View

This view displays the distribution of LC-MS signals/features in the fractions. See Chapter 5, Section 4, "LC/MS View" for details on LC/MS View.

# Chapter 14. Label Free Quantification (LFQ)

## 1. Overview

Label-free quantification is one of the three quantification methods supported by the PEAKS Q module. It usually serves as a fast profiling tool for differential studies of large scale proteomics. This quantification method is based on the relative intensities of peptide features detected in multiple samples. Feature detection is performed separately on each sample. By using the EM (expectation-maximization) algorithm, more overlapped features can be detected. The features of the same peptide from different samples are reliably aligned together using a high-performance retention time alignment algorithm.

The label-free quantification of PEAKS Q module does not have to be based on an identification result as label-free quantification is based on detected features. However, you can choose attach an existing identification result as the last step of the label-free quantification set-up.

### Note

For details of the retention alignment algorithm, refer to the paper "*A Combinatorial Approach to the Peptide Feature Matching Problem for Label-Free Quantification", Bioinformatics, 2013, 10.1093*.

The usage of this function is outlined in the following overview.

1. Select a project node in the "Project View" frame. Click the PEAKS Quantification tool bar icon **Q** or select the Tools | Quantification command from the menu.

   ### Note

   Refer to Chapter 4, *Project Wizard - Create a PEAKS Project, Load Data and Set up a Workflow* for project creation, data loading, PEAKS DB search, and quantification in a single workflow.

2. Select the quantification method as "Label Free" and specify the PEAKS Label Free quantification parameters in the right panel of the window.

3. Arrange the samples into groups, attach Identification results to the analysis, and click OK.

   ### Note

   LFQ requires at least 2 groups with at least 1 sample in each group.

   ### Note

   Currently, only identification results from PEAKS are supported in quantification. inChorus and all other third party engine results are not supported.

4. Wait for the analysis to finish. A new LFQ result node **Q** will appear at the project tree. Double-click the node to open the result file.
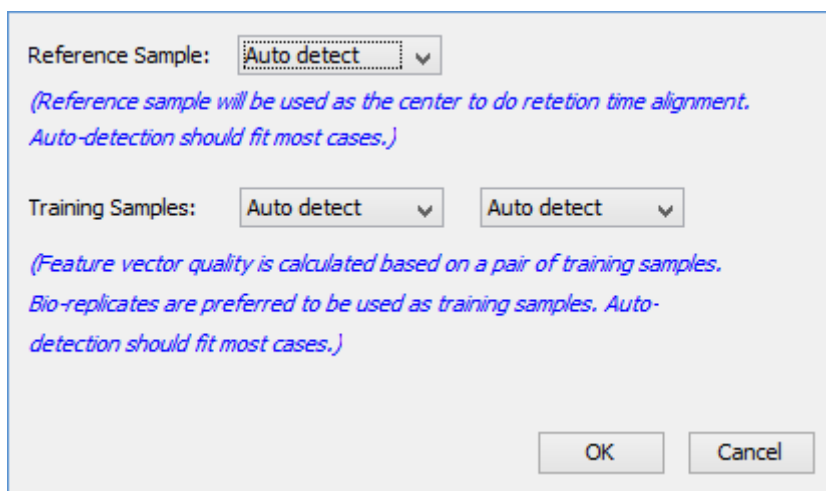
## 2. Setting Parameters

The following parameter dialogue appears when the quantification tool bar icon, **Q** , is clicked or **Tools|Quantification** command from the menu is clicked. Select **Label Free** under the **Quantifications** heading. Enter the relevant MS/MS level labeling quantification parameters on the right side of the window.
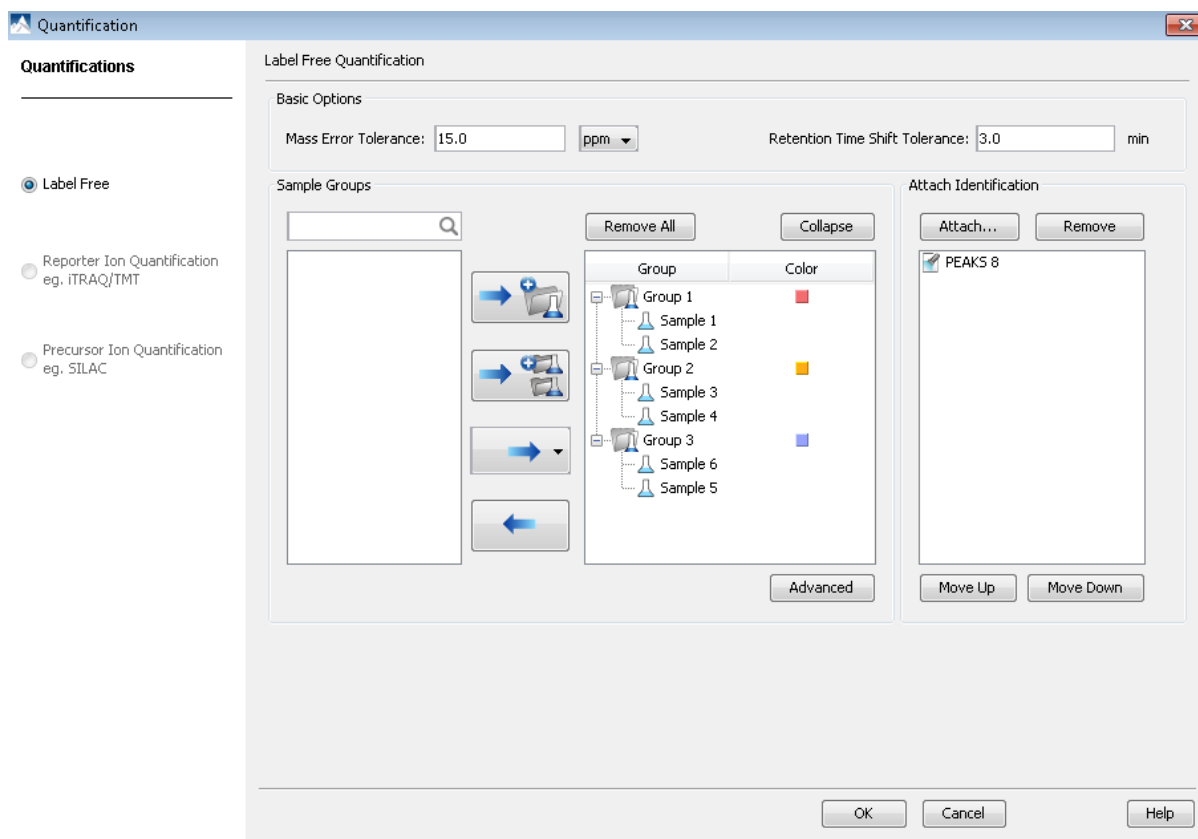
The following parameters are available in the "Basic Options" section of the quantification window:

- **Mass Error Tolerance:** The mass shift between different runs. Initially this should be set as twice as the mass error tolerance in database search. Later, by checking the M/Z shift distribution figure in the result, you can narrow the tolerance.

- **Retention Time Shift Tolerance:** The retention time range is the maximum elution time range that is considered for the quantification of an identified peptide. Initially, this should be set as the highest possible shift between runs. This tolerance can be later narrowed by checking the RT shift distribution figure in the result.

- LFQ requires that sample grouping to generate a result node. The "Sample Groups" section allows for this group selection. All available samples are listed on the left window. For experiments with multiple samples, the search bar can be used to look for a specific sample. After selecting a sample, it can be added to a new group by clicking the  button or to an existing group by clicking the  button and selecting that option from the drop-down menu. If you select multiple files, you can add each one to a new sample by clicking the  button. Display all samples in each group or just the individual groups by selecting the Expand and Collapse buttons, respectively. Samples can be removed individually or removed all at once by clicking the  button or the **Remove All** button, respectively.

- The **Advanced** option opens another window that enables the manual selection of a reference sample and training samples. The reference sample will be used as the center to do the retention time alignment. The training samples are used to calculate the feature vector quality. For training samples, biological replicates are preferred. Auto detection should fit most cases of both reference and training samples.
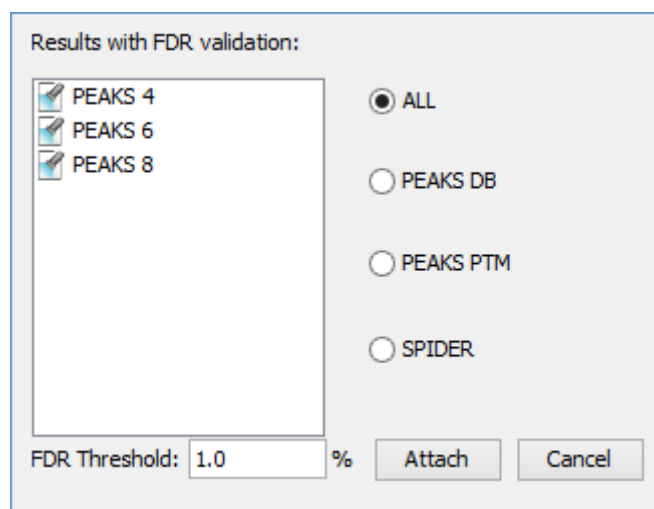
- To attach the identification results to the label free quantification result, click the Attach ... button in the Attach Identification section. The "Attach identification results" dialogue lists all available identification results. Select the identification results to be attached and the FDR threshold to be used, then click the Attach button. All selected results will be listed in the Attach Identification section. If multiple identification results for the same fraction are selected, then only the first identification result in the list will be used. Use the Move Up and Move Down buttons to set the order in which the identifications will be attached.



# 3. Understanding the LFQ Results

Once completed, a label free quantification result node Q will be added to the project tree. Double-click on this node to open the LFQ result that contains multiple tabs as described below:

- **Summary**: The outline of thePEAKS LFQ search results with statistics. This is the place to examine the overall performance of the experiment and adjust filters.

- **LC/MS**: The overall mass signal view for each MS run, marked with detected features and MS/MS spectra. SeeSection 4, "LC/MS View" for details on LC/MS view.

- **Features**: The quantified peptide feature vectors listed in a table. One feature vector is a list of features from different samples aligned together for one peptide. Extracted ion chromatogram (XIC), 2D/3D LC/MS view and retention time alignment are also displayed for each feature vector.

- **Proteins**: The quantified proteins with a list of supporting peptide features for each protein.

# 3.1. Summary View

The "Summary" view provides three main functions:

1. **Result Filtration:** This is achieved by specifying the filtration rules in the area at the top of the Summary View. The filtration function is discussed in Section 3.2.1, "Protein Table".

2. **Summary Report:** Several statistical charts assist the user to get an overall picture of the results, assess the result quality, and examine the reliability and reproducibility of the experiment. This function is the focus of this section.

3. **Result Exporting:** This is achieved by clicking the "Export" button at the top of the summary view. The exporting function is discussed in Section 5.3, "Exporting Label-Free Quantification Results".

The "Summary" view includes an expression profile with candidate proteins assorted in a heat map, result statistics tables, and a list of filtration parameters. To add a summary note, click on the Notes button to open a "Notes Entry" editor in which you can edit the notes. These notes will be displayed in the "Summary" view.
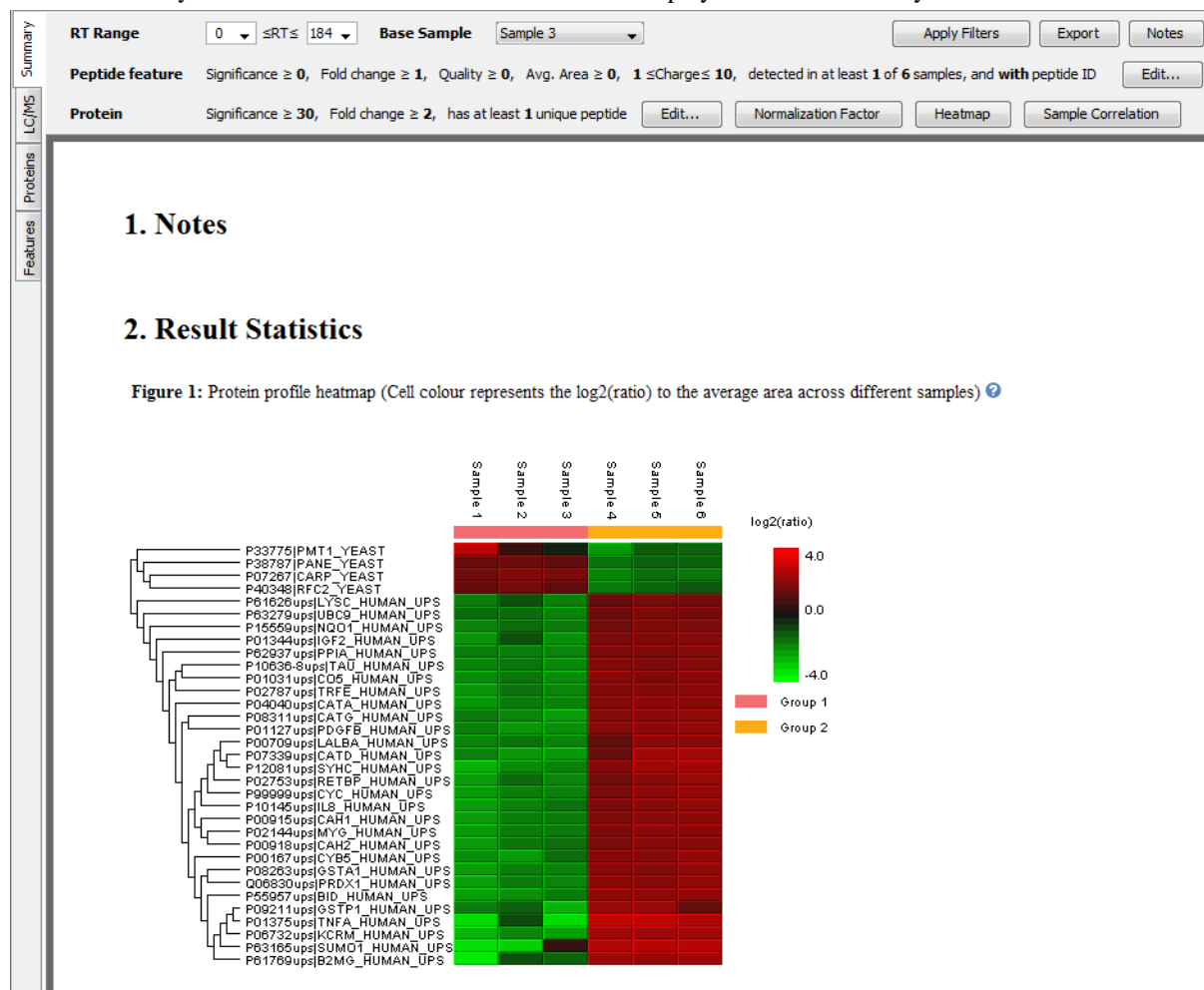


## 3.1.1. Result Filtration

PEAKS provides an extensive set of tools to filter Label Free Quantification results according to the user's requirements. There are four categories of options in the "Summary" view: **RT Range**, **Base Sample**, **Peptide feature**, and **Protein**.
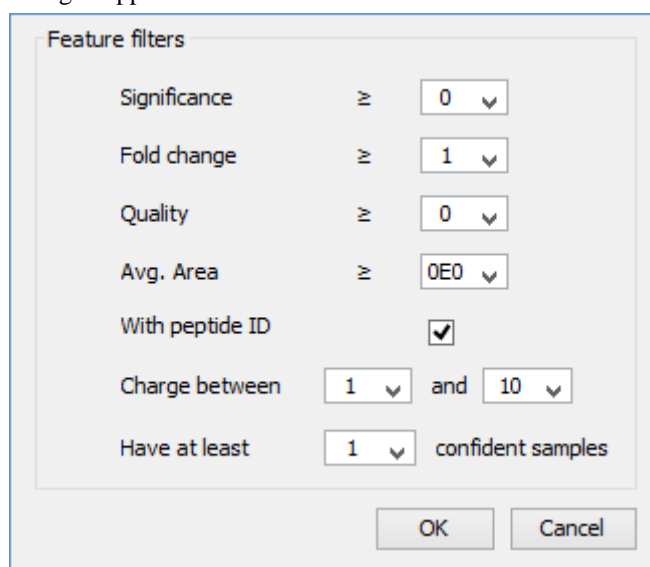
### 3.1.1.1. RT Range

Feature vectors within the selected retention time range will be included in the results while everything else will be removed. Select **Apply** within the filtration window of the "Summary" view to include the selected RT range in the analysis.

### 3.1.1.2. Base Sample

Select the sample that will act as the reference. The peptide area corresponding to this reference label will be the denominator in all of the feature area ratio calculations. Select **Apply** within the filtration window of the "Summary" view to include the selected base sample in the analysis.

### 3.1.1.3. Peptide Feature

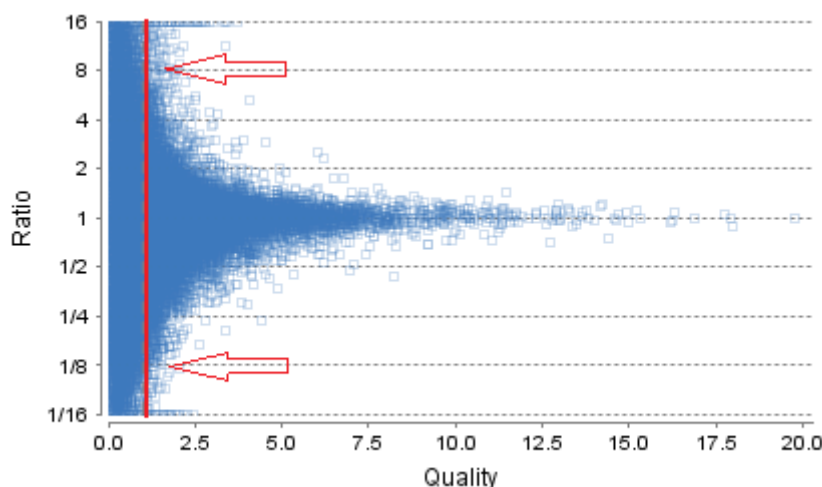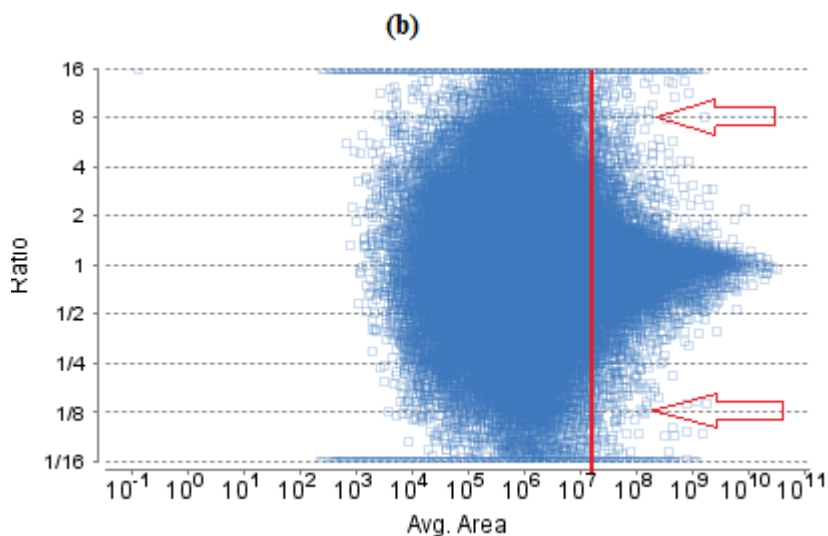The following parameter dialogue appears when the **Edit…** button is selected.



- **Significance:** Only peptides with a score above this threshold are included in the analysis.

- **Quality:** The higher quality of a peptide indicates that the peptide is more quantifiable. Factors that affect the quality score include m/z difference, RT difference, XIC shape similarity, and feature intensities. An appropriate threshold value can be selected by checking the ratio-quality figure. The rule of thumb is to set the quality threshold around a fold change of 8, as shown in this labelled version of Figure 3a) of the "Summary" view. Only feature vectors with a quality score above this threshold are used to calculate the quanlity score of a peptide.



- **Ave. area:** This filter is directly applied to the MS signal intensity. Higher intensity peptide features have been shown to be more reproducible across replicates. The rule of thumb is to choose the average area near a fold

change of 8, as shown in this labelled version of Figure 3b) of the "Summary" view. Only peptides with at least one label above this threshold will be used in protein quantification.
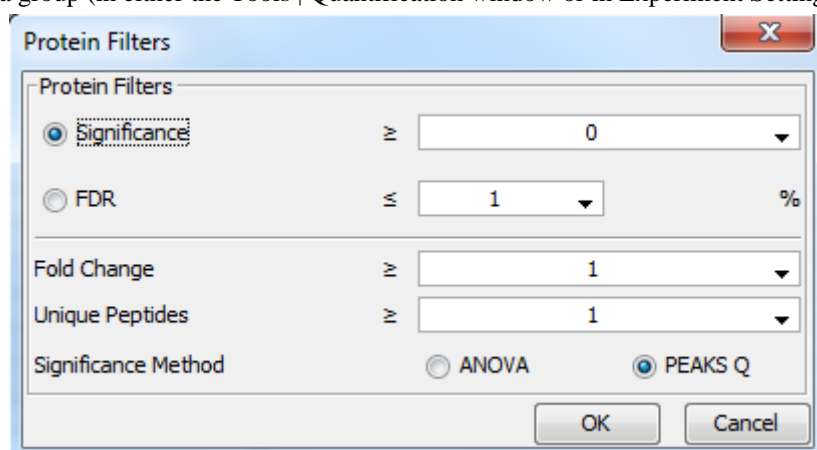


- **With peptide ID:** Uncheck this box to exclude peptide IDs determined from the attached identification results; check this box to include peptide IDs.

- **Charge between:** Only peptides with feature vectors that all fall within this range will be used in protein quantification.

- **Confidently detected sample:** A peptide is more quantifiable when it is detected in more samples. The rule of thumb is to set this filter to half the number of total samples.

Select **Apply** within the filtration window of the "Summary" view to include all selected filters in the analysis.

### 3.1.1.4. Protein Filter

The following parameter dialogue appears when the **Edit…** button is selected. These filters determine what will be included in a protein group significance calculation, as well as what significance score thresholds will be set. For a protein group to have a significance score, it must have been identified in a sample/label that has already been placed into a group (in either the Tools | Quantification window or in Experiment Settings).



- **Significance:** Only protein groups with a significance above this threshold will be listed in the "Protein" view. Either this or "FDR" can be selected to set a significance threshold. It's recommended that you set a significance score threshold at 20.

- **FDR:** Adopt the Benjamin–Hochberg method to adjust the p-value to the false discovery rate for all protein groups that have already passed the other filters. Only protein groups with significance scores passing the cal-

culated FDR will be listed in the "Protein" view. Either this or "Significance" can be selected to set a significance threshold.

- **Fold Change:** Only protein groups at or above this fold change threshold will be listed in the "Protein" view.

- **Unique Peptides:** Only protein groups with at least this many unique peptides will be listed in the "Protein" view.

- **Significance Method:** One of two significance calculations can be applied to the list of protein groups: PEAKS Q or ANOVA. Scores will only be calculated for protein groups identified from samples that are in groups. PEAKS Q is similar to the Significance B method employed by the paper referenced below. Additionally, support peptide similarity is considered. ANOVA requires that at least two samples be included in each particular group. The ANOVA method employed is also referenced below.

## Note

For more **details on PEAKS Q**, refer to the following paper: *"MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol.* 2008, 26(12): 1367-1372.

For more **details on ANOVA**, refer to the following paper: *"On the comparison of several mean values: an alternative approach"* Biometrika. 1951, 38(3/4): 330-336.

The following parameter dialogue appears when the **Normalization Factor** button is selected.



- **Use TIC:** By default, PEAKS uses the total ion current (TIC) of the samples to calculate the normalization factors.

- **Use internal standard proteins:** This option displays a list of identified proteins, which can be selected to create a normalization ratio. A *search* bar is included to enable the quick location of select proteins. Right-

clicking in the Protein screen will open a window that will allow for the selection of multiple highlighted rows or the removal of all check-marked proteins. The Experiment drop-down menu can be opened to display the results of another sample. If this option is chosen but no proteins are selected, no Factors will be listed for any of the samples except the first sample, which will display 0.0.
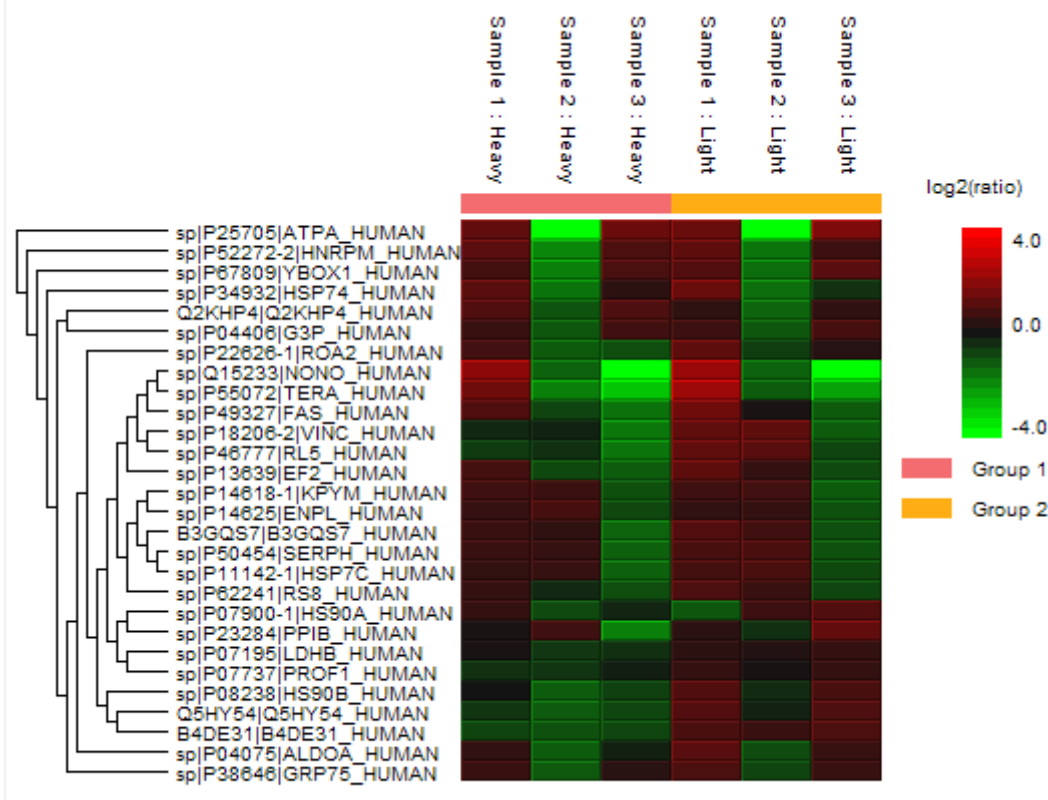
- **Manual input:** This option performs the same default calculation as Auto normalization but also allows for the manual change of each label's Expected Ratios. After changing a ratio and hitting Enter on the keyboard, PEAKS will recalculate the Factor for the non-referenced label(s). The Experiment drop-down menu can be opened to display the results of another sample.

- **No Normalization:** All samples are automatically assigned Factors of 1.0.

Select Apply within the filtration window of the "Summary" view to include all selected filters and Normalization choices in the analysis.

## 3.1.2. Summary Report
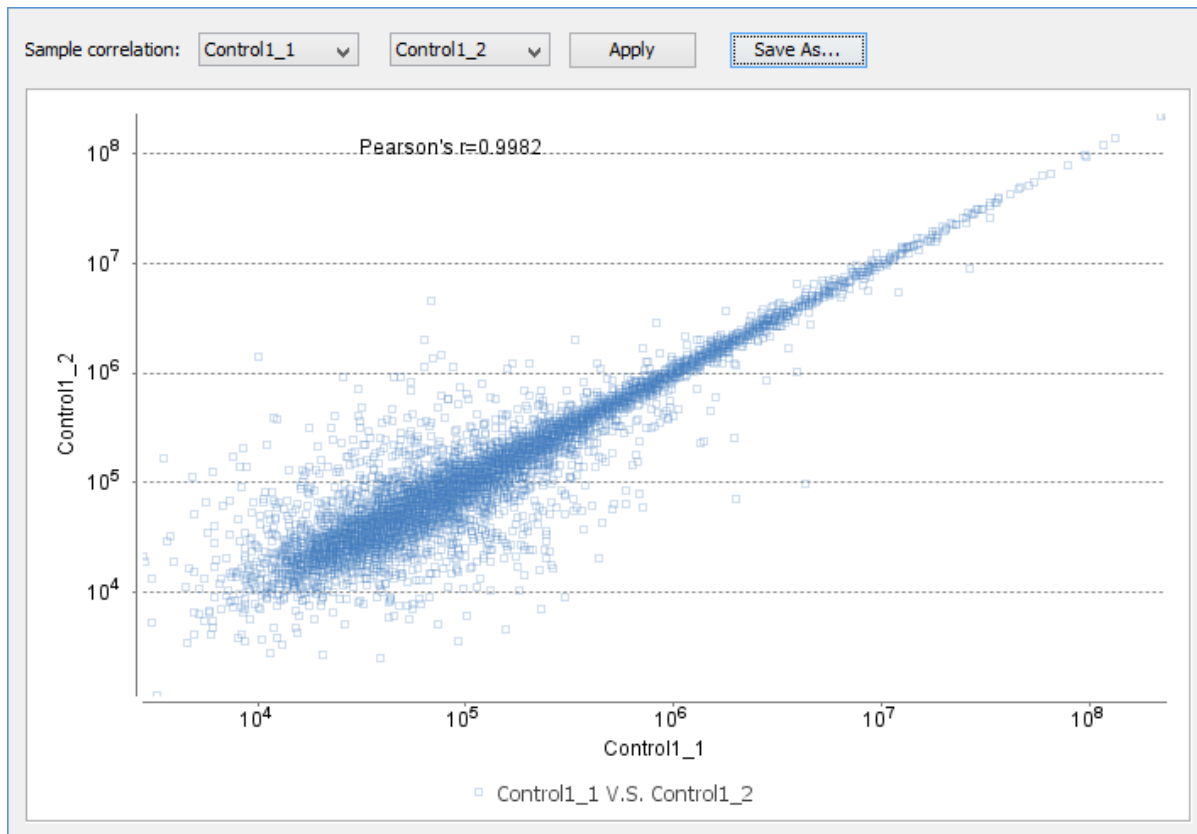
### 3.1.2.1. Heat Map

The heat map displays the protein groups that passed the filters. The relative protein abundance is represented as a heat map of the representative proteins of each protein group. These representative proteins are clustered if they exhibit a similar expression trend across the samples. The hierarchical clustering is generated using a neighbour-joining algorithm with a Euclidean distance similarity measurement of the log2 ratios of the abundance of each sample relative to the average abundance. An interactive protein profile heat map is opened in a new window when the **Heat Map** button that is located in the filtration options. Right-click on the heat map in this window to save a copy of it.



In the interactive heat map (by clicking the Heat Map button), move the cursor over any point to highlight the protein accession name and the sample/label to which it matches. Right-click on the heat map and click **Save as…** to export a PNG file of the heat map.

### 3.1.2.2. Sample Correlation

To view the intensity correlation plot for the detected features of the two selected samples, click the **Sample Correlation** button from the "Summary" view title bar. The correlation plot will be displayed in a pop-up window. Select the samples from the drop-down menus beside "Sample correlation" to plot the intensity correlation graph of different sample pairs. The Pearson's correlation score indicates the reproducibility of the experiment based on a selected pair of samples. Click Save as… to export the correlation plot to a PNG file.
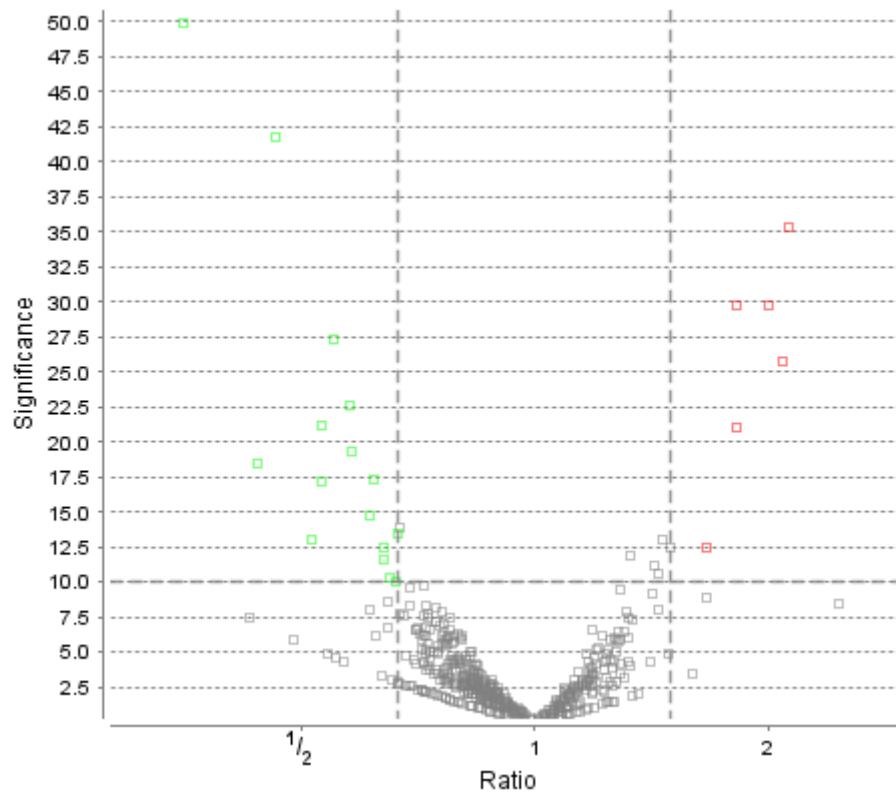


### 3.1.2.3. Volcano Plot

A volcano plot combines a statistical test with the magnitude of the change enabling quick visual identification of those data-points that display large-magnitude changes and that are also statistically significant. The volcano plot in Figure 2(a) plots significance versus fold-change of the quantified proteins.

- **Horizontal broken grey line:** The selected significance threshold

- **Vertical broken grey lines:** The selected fold change thresholds

Plotting points in this way results in two regions of interest in the plot: those points that are found towards the top of the plot that are far to either the left or the right side. These represent values that display large magnitude fold changes (hence being left or right of center) as well as high statistical significance (hence being towards the top). In this figure, the ratio is the group ratio set with respect to the base group. If more than two groups are present, the max ratio is used instead.
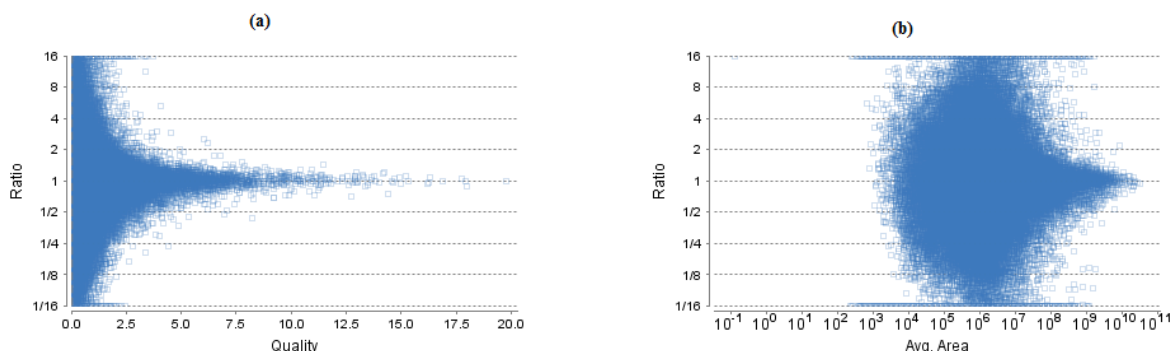
The "Protein" view contains an interactive volcano plot that be hidden or displayed. Markers for the proteins that are above the set significance threshold will be displayed in colour. Scrolling the cursor over a marker will highlight it and, subsequently, clicking on this marker will highlight the protein in the interface.

### 3.1.2.4. Feature Vector Ratio Distribution

Figures 3(a) and 3(b) plot the distribution of feature vector ratios by quality and by average area, respectively. These plots help to determine the threshold values for the quality and intensity filters to obtain a reliable section of the analysis results. The rule of thumb is to set the quality threshold around an 8-fold change, as shown below in Figure 3(a). The quality filter is usually the better filter to use but under some circumstances, it may be better to apply the filter to the MS signal directly. The rule of thumb is still to choose the average area around an 8-fold change. In both figures, the ratio is the group ratio with respect to the base group. If more than two groups are present, each ratio value depicts a data point in the figures.

**Figure 3.** The distribution of feature vector ratio: **(a)** By quality. **(b)** By intensity. ❷



### 3.1.2.5. RT and m/z Shift Distribution

Figure 4(a) plots the retention time shift distribution to the base sample. Figure(b) plots the distribution of the m/z shift (in ppm) of the data with respect to the base sample. These two figures can be used to refine the parameter settings for label free quantification

**Figure 4. (a)** RT shift distribution; **(b)** M/Z shift distribution. ⊘
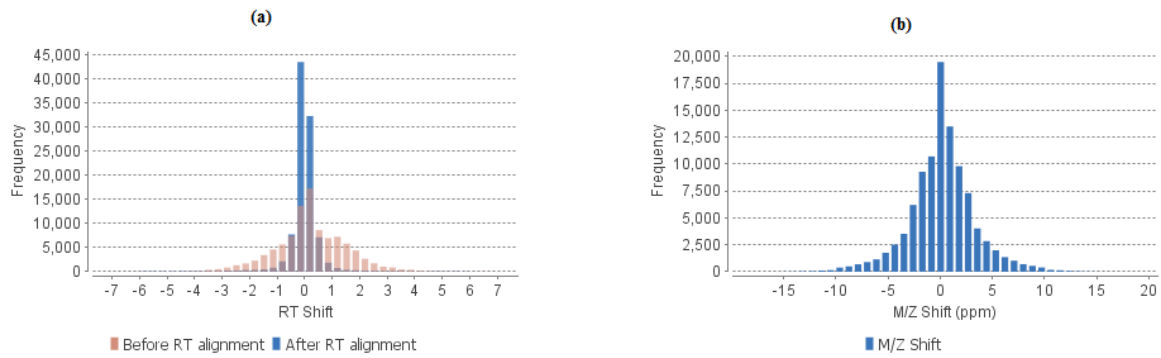


## 3.1.2.6. Summary Tables

Tables 1 and 3 list the data and the filtered results, respectively. Table 2 lists the feature vector and protein result filtration parameters.

**Table 1.** Result filtration parameters.

| | |
|---|---|
| Retention time | $\geq 0$ |
| Retention time | $\leq 140$ |
| Feature fold change | $\geq 1$ |
| Quality | $\geq 0$ |
| Intensity | $\geq 0E0$ |
| Charge | $\geq 1$ |
| Charge | $\leq 10$ |
| Confident sample number | $\geq 1$ |
| With peptide ID | true |
| Normlization | Use TIC |
| Protein significance | $\geq 0$ |
| Protein fold change | $\geq 1$ |
| Confident unique supports | $\geq 1$ |
| Top protein groups | 50 |

**Table 2.** Statistics of filtered result.

| | |
|---|---|
| Features | 7184 |
| Features with ID | 2714 |
| Feature vectors | 1025 |
| Feature vector with ID | 1025 |
| Protein groups | 50 |

**Table 3.** Search Parameters

Quantification type: Label free quantification
Mass Error Tolerance: 20.0 ppm
Retention Time Shift Tolerance: 6.0 min
Dependent on PID: 10, FDR = 1%
Samples: Control1_1, Control1_2, Control2_1, Control2_2, Treatment12_1, Treatment12_2, Treatment24_1, Treatment24_2,
Reference Sample: Auto detect

Tables 4 and 5 list the quantification and instrument parameters, respectively.

**Table 4.** Quantification parameters.

Quantification Type: ICAT/SILAC
Quantification Mass Tolerance: 15.0ppm
Quantification RT Range: 0.5min.
-10LogP Threshold: 15.0
Reporter Ion Type: MS2
R to P Convert enabled: true
Light: R (100.0%) K (100.0%)
Heavy: R(+10.01) (100.0%) K(+8.01) (100.0%)
Group Settings in Parameters:
Group 1: Sample 1 : Heavy
Group 1: Sample 2 : Heavy
Group 1: Sample 3 : Heavy
Group 2: Sample 1 : Light
Group 2: Sample 2 : Light
Group 2: Sample 3 : Light
DB Dependency: 5
Current Group Information:
Group 1: Sample 1 : Heavy
Group 1: Sample 2 : Heavy
Group 1: Sample 3 : Heavy
Group 2: Sample 1 : Light
Group 2: Sample 2 : Light
Group 2: Sample 3 : Light

**Table 5.** Instrument parameters.

Fractions: 20120812_ShAr_QE1_ARL_NX4_mix_sample_1.raw, 20120812_ShAr_QE1_ARL_NX4_mix_sample_2.raw, 20120817_ShAr_QE1_ARL_NX4_mix_sample_2_REP_3.raw
Ion Source: ESI(nano-spray)
Fragmentation Mode: CID, CAD(y and b ions)
MS Scan Mode: FT-ICR/Orbitrap
MS/MS Scan Mode: Linear Ion Trap

### 3.1.3. Result Exporting

PEAKS precursor ion quantification results can be exported to various HTML and text formats by clicking the Export button within the filtration options in the "Summary" view. The "Summary" view and the images from various views can also be exported to various HTML and text formats. Refer to Section 5.3, "Exporting Label-Free Quantification Results" **(reference)** for details.

## 3.2. Protein View

The "Protein" view lists all the quantified proteins present in the sample, characterizes each protein at the amino acid level, and lists the supporting peptide features of each protein. It has four components:

• **Protein Table:** Lists all the quantified proteins present in the sample.

• **Protein Volcano Plot:** Click on a marker that is above the significance threshold (the horizontal dashed line) set in the "Summary" view. Upon selection, the corresponding protein in the Protein Table will be highlighted.

• **Coverage:** Characterizes the protein sequence at the amino acid level. All the PTMs and mutations occurred on the protein sequence will be displayed together with the MS/MS data supporting the inference.

• **Features:** Lists the features identified from the selected proteins in the protein table.
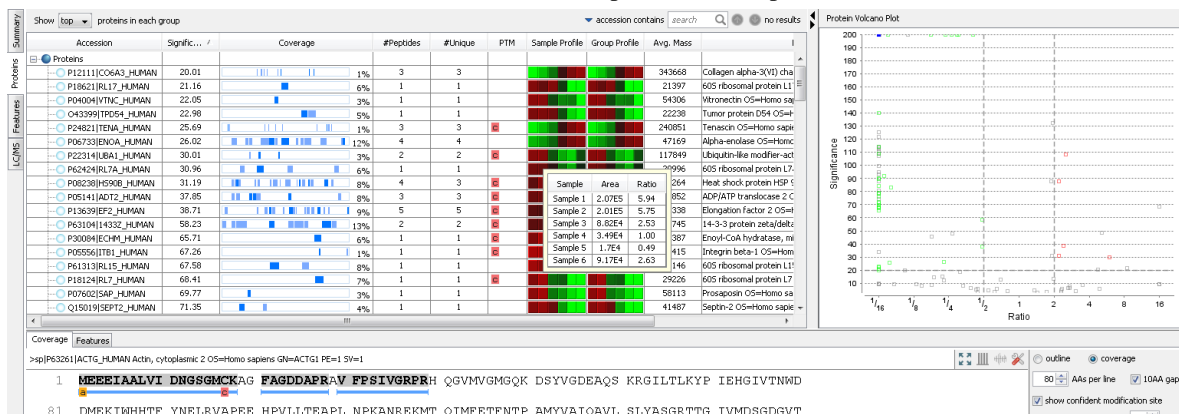
## 3.2.1. Protein Table

The quantified proteins are listed in the protein table with homologous proteins grouped together. This protein table is the same as the other protein tables found in the PEAKS DB, PEAKS PTM, and SPIDER results except the following columns mentioned below:

- **Significance:** The protein significance is set according to the rules set in the "Summary" view, refer to Section 3.1.1.3 **Section 3.2, "Protein View"**.

- **Sample Profile:** The absolute protein abundance among the samples is depicted as a heat map. Hold the cursor on a profile to view the sample channels (labels), areas, and ratios with respect to the base label selected in the "Summary" view.

- **Group Profile:** The absolute protein abundance among the groups is depicted as a heat map. Hold the cursor on a profile to view the group channels (labels), areas, and ratios with respect to the group which contains the base label selected in the Summary View.

See Section 3.3.1, "Protein Table" for more details on other entries of the protein table

## 3.2.2. Protein Volcano Plot

This interactive volcano plot enables the selection of a protein in the table. Protein markers that are above the significance threshold can be selected. Upon selection, the corresponding protein in the Protein Table will be highlighted. Proteins located in the top-right and top-left sections of this plot (above the significance threshold and to the right and left of the fold change thresholds) are considered statistically significant. Green points are down-regulated while red points are up-regulated. The volcano plot facilitates identification of these significant proteins.

Use the ◀ button to expand or collapse the volcano plot.

## 3.2.3. Coverage Tab

The coverage tab characterizes the selected protein sequences at the amino acid level. See Section 3.3.2, "Protein Coverage" for details.

## 3.2.4. Features Tab

The supporting peptides assigned to the protein are shown under the "Peptides" tab.



This table is similar to the peptide table in the "Peptide" view.

- **(Checkbox):** Select the peptides to appear within the Coverage Tab as blue bars.

- **Used:** Select whether this peptide is used to calculate the protein profile.

- **Candidates:** Select which peptide features will be used as the top three candidates involved in quantification. All non-unique peptide and ones not passing the filter will be excluded automatically. Click "Apply" button in the "Summary" view to make the selection effective. Select one row and right-click on the Candidates column to show the pop-up menu. Use "Reset the candidate selection" to reset the candidate selection for all proteins.

- **Quality:** This defines how well the peptide can be quantified. A higher quality peptide indicates that the peptide is more quantifiable. If a peptide's abundance does not change across the samples, then the measured intensity should not change too much, even though it may have some random variation. Small variation means the peptide can be well quantified. Let $\sigma$ be the estimated intensity fold change of the peptide across the samples when the

peptide abundance does not change. The quality is calculated as 1/log2#. Therefore, if a peptide has an expected two-fold change, then its quality will be 1; if a peptide has an expected 4-fold change, then its quality will be 0.5. An appropriate threshold value can be selected by checking the ratio-quality figure. The rule of thumb is to set the quality threshold around a fold change of 8, as shown in this labelled version of Figure 3(a) of the "Summary" view. Only feature vectors with a quality score above this threshold are used to calculate the quality score of a peptide.

- **Significance:** It assumes that for all the peptides with the same quality, their ratios follow a log-normal distribution. Different qualities may have log-normal distributions with different variances. The two-tail p-value is calculated as the significance of the feature vector using the log-normal distribution corresponding to its quality and its maximum group ratio. The -10lgP value is calculated from this and displayed as the significance in the column.

- **m/z:** Each feature in the feature vector has its own precursor mass-to-charge ratio. This column displays the average precursor mass-to-charge ratio of all the features in the feature vector.

- **RT range:** The retention time range of the feature vector.

- **z:** The precursor charge of the feature vector.

- **Avg. Area:** Each feature in the feature vector has its own intensity area under the XIC curve. This column displays the average area of all the features in the feature vector.

- **Sample Profile:** The absolute peptide abundance among the samples is depicted as a heat map. Hold the cursor on a profile to view the sample channels (labels), the areas, and the ratios with respect to the base label selected in the "Summary" view. Double-click the sample

- **Group Profile:** The absolute peptide abundance among the groups is depicted as a heat map. Hold the cursor on a profile to view the group channels (labels), the areas, and the ratios with respect to the group which contains the base label selected in the "Summary" view.

- **RT mean:** Each feature in the feature vector has its own retention time center. This column displays the average retention time center of the feature vector.

- **Start:** This shows the position of the first residue of the peptide that covers the protein.

- **End:** This shows the position of the last residue (inclusive) of the peptide that covers the protein.

Right-click on the peptide feature to show the pop-up menu similar to the one in the PEAKS DB supporting peptides tab.

# 3.3. Features View

The "Features" view displays all of the identified peptides and their intensities. The interface is similar to the "Peptide" view in a PEAKS DB result. See Section 3.4.1, "Peptide Table" for details.

## 3.3.1. Feature Table

Contains the important details and statistics associated with each quantified feature that passes the filters in the "Summary" view. This table contains the following columns:

- **(Checkbox):** Select the peptides to appear within the Coverage Tab as blue bars.

- **Used:** Select whether this peptide is used to calculate the protein profile.

- **Candidates:** This shows which peptides are unique to the current protein group. The top three unique peptides, in terms of average area, are selected by default. If a candidate is deselected, click **Apply** in the "Summary" view. Select one row and right-click on the Candidates column to show the pop-up menu. Use "Reset the candidate selection" to reset the candidate selection for all proteins.

- **Quality:** This defines how well the peptide can be quantified. A higher quality peptide indicates that the peptide is more quantifiable. If a peptide's abundance does not change across the samples, then the measured intensity

should not change too much, even though it may have some random variation. Small variation means the peptide can be well quantified. Let $\sigma$ be the estimated intensity fold change of the peptide across the samples when the peptide abundance does not change. The quality is calculated as 1/log2#. Therefore, if a peptide has an expected two-fold change, then its quality will be 1; if a peptide has an expected 4-fold change, then its quality will be 0.5. An appropriate threshold value can be selected by checking the ratio-quality figure. The rule of thumb is to set the quality threshold around a fold change of 8, as shown in this labelled version of Figure 3a) of the "Summary" view. Only feature vectors with a quality score above this threshold are used to calculate the quality score of a peptide.

- **Significance:** It assumes that for all the peptides with the same quality, their ratios follow a log-normal distribution. Different qualities may have log-normal distributions with different variances. The two-tail P-value is calculated as the significance of the feature vector using the log-normal distribution corresponding to its quality and its maximum group ratio. The -10lgP value is calculated from this and displayed as the significance in the column.

- **m/z:** Each feature in the feature vector has its own precursor mass-to-charge ratio. This column displays the average precursor mass-to-charge ratio of all of the features in the feature vector.

- **RT range:** The retention time range of the feature vector.

- **z:** The precursor charge of the feature vector.

- **Avg. Area:** Each feature in the feature vector has its own intensity area under the XIC curve. This column displays the average area of all the features in the feature vector.

- **Sample Profile:** The absolute peptide abundance among the samples is depicted as a heat map. Hold the cursor on a profile to view the sample channels (labels), the areas, and the ratios with respect to the base label selected in the "Summary" view. Double-click the sample

- **Group Profile:** The absolute peptide abundance among the groups is depicted as a heat map. Hold the cursor on a profile to view the group channels (labels), the areas, and the ratios with respect to the group which contains the base label selected in the "Summary" view.

- **RT mean:** Each feature in the feature vector has its own retention time center. This column displays the average retention time center of the feature vector.

- **Accession:** This accession is of the protein from which the peptide arises. Protein with the highest -10lgP that the peptide supports. If no proteins that the peptide supports pass the protein filters, the accession field will be empty.

- **PTM:** The post-translational modifications that appear within the peptide.

### 3.3.2. Feature Details

The "Feature Details" tab contains the eXtracted Ion Chromatogram (XIC) chart that displays the shape of the selected peptide feature vector over the retention time range where they were identified. The table at the top-left corner displays both the area and the ratio of the features in the feature vector. This table also works as a control to select the features that will be used to draw the XIC plot and other plots in the "Sample Features" and "RT Alignment" tabs. The  button at the top-right corner of the tab can be used to toggle the plot data between, before, and after RT alignment.

> # Note
>
> By default, the first features from each group are selected to be displayed in the plots. To display other features, select the corresponding sample using the check box in the table.
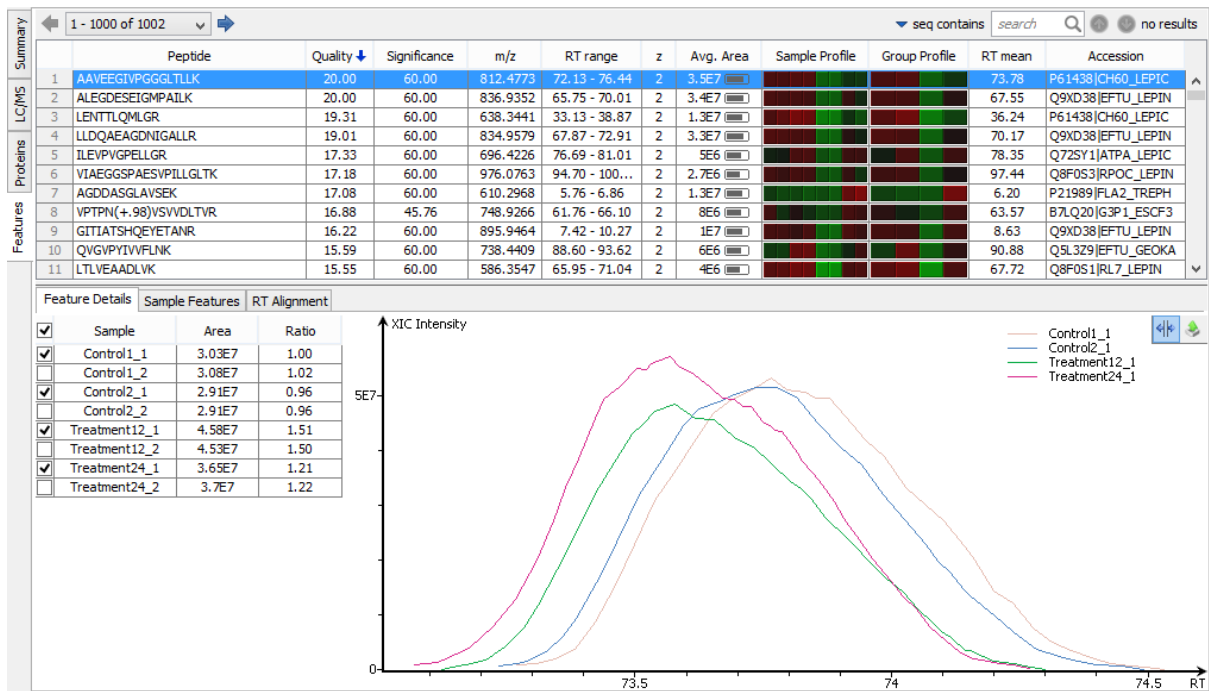
## 3.3.3. Sample Features

This tab displays a close up LC/MS view of the detected features from the selected samples. The feature boundary is shown by the bounding polygon. The center of the feature is denoted by a solid red dot on the monoisotopic peak cluster. Placing the cursor on this feature center displays a tool tip that shows the m/z value, the mean retention time, the charge state, and the intensity. The centers of the features are used to calculate the retention time alignment.

The  button at the top-right corner of the tab can be used to toggle between the 2D and the 3D LC/MS views.
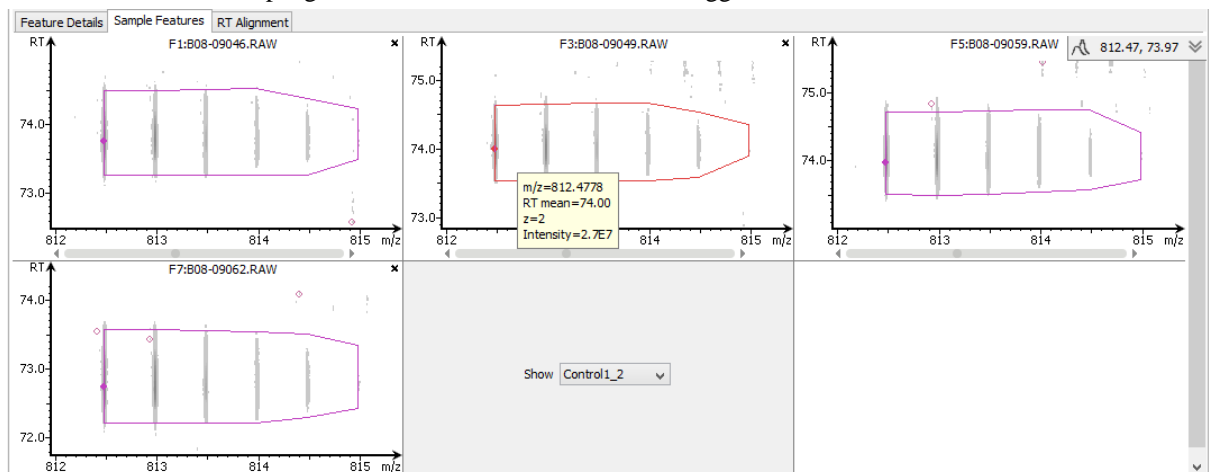


## 3.3.4. RT Alignment

This tab displays the aligned TIC curves of the selected samples with the highlighted retention time range of the selected feature. Use the toggle switch in the top-right corner to switch between the before and after RT alignment view. Click the  button to view the pairwise RT alignment in the pop-up window. Select the sample pair from the drop-down menu to view the desired pairwise RT alignment.
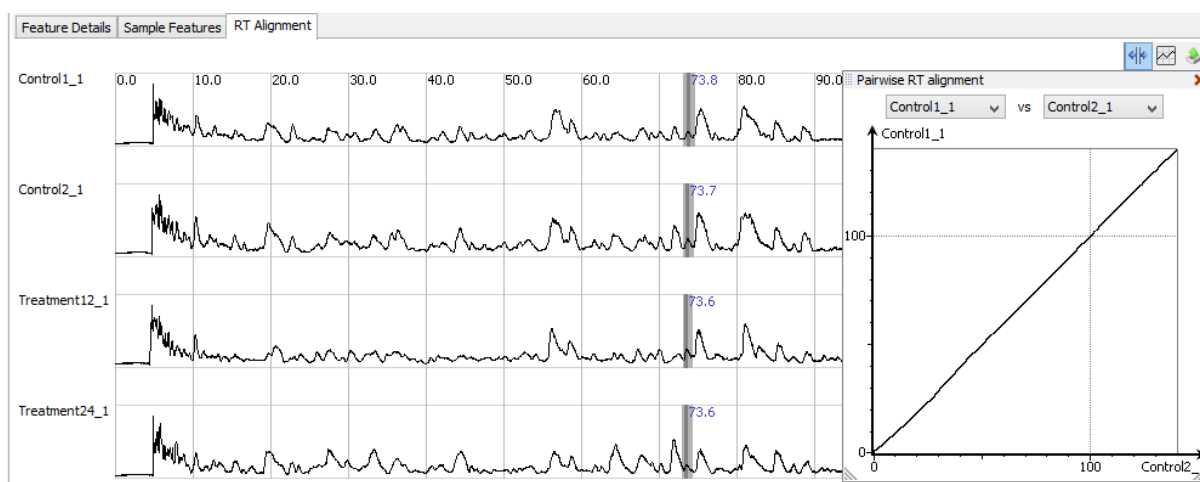
## 3.4. LC/MS View

This view displays the distribution of LC-MS signals/features in the fractions. See Section 4, "LC/MS View" for details on LC/MS View.

# Chapter 15. Creating a High-Throughput Workflow

PEAKS provides a simple mechanism to define an easy workflow while creating a new project using the Project Wizard. The workflow will be applied to the whole project and the results will be generated at the project level. Refer to Chapter 4, *Project Wizard - Create a PEAKS Project, Load Data and Set up a Workflow* for details on setting up an easy workflow.

For convenience, PEAKS also provides another mechanism to define workflows for protein identification, protein quantification, and inChorus (multi-engine protein ID) searching. This workflow can be applied on multiple projects that can perform PEAKS analyses at the fraction level, the sample level, or the project level. This tool is specially helpful for analyses on multiple projects using the same parameters.
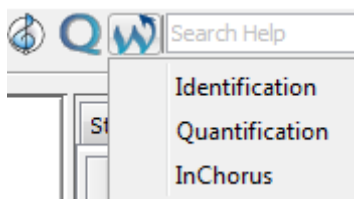


Once a specific workflow is selected, a dialogue pops up to specify the analysis steps and the parameters to use in each step.

## 1. Identification Workflow

Click the workflow icon on the toolbar and select Identification. The identification workflow setup window will appear:
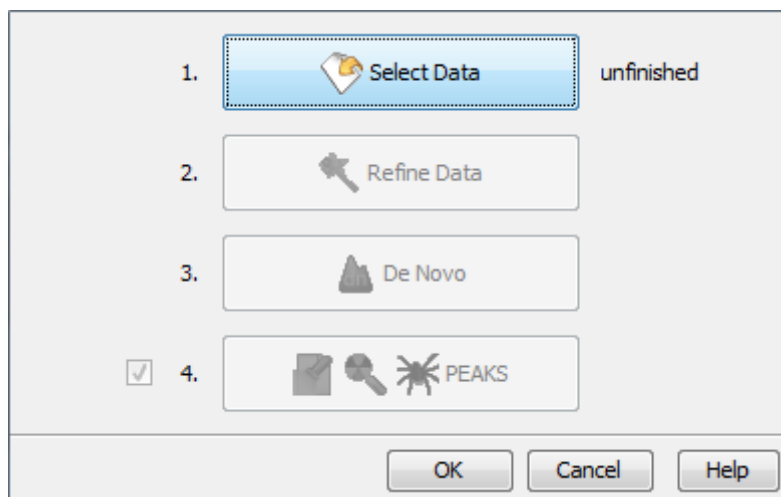


Click the Select Data button to navigate and locate files to be used for identification analysis. Only projects that are open in the "Project View" panel can be selected for analysis. To select the files/samples to be analyzed, either select the individual file/sample or click either the All Samples or the All Fractions buttons. Then, click the Add to Right to transfer the samples/files to the "Selected Data" list on the right side. Use the Remove and Clear buttons to remove the selected files/samples or all files/samples from the "Selected Data" list, respectively. Click OK to proceed to the next step.
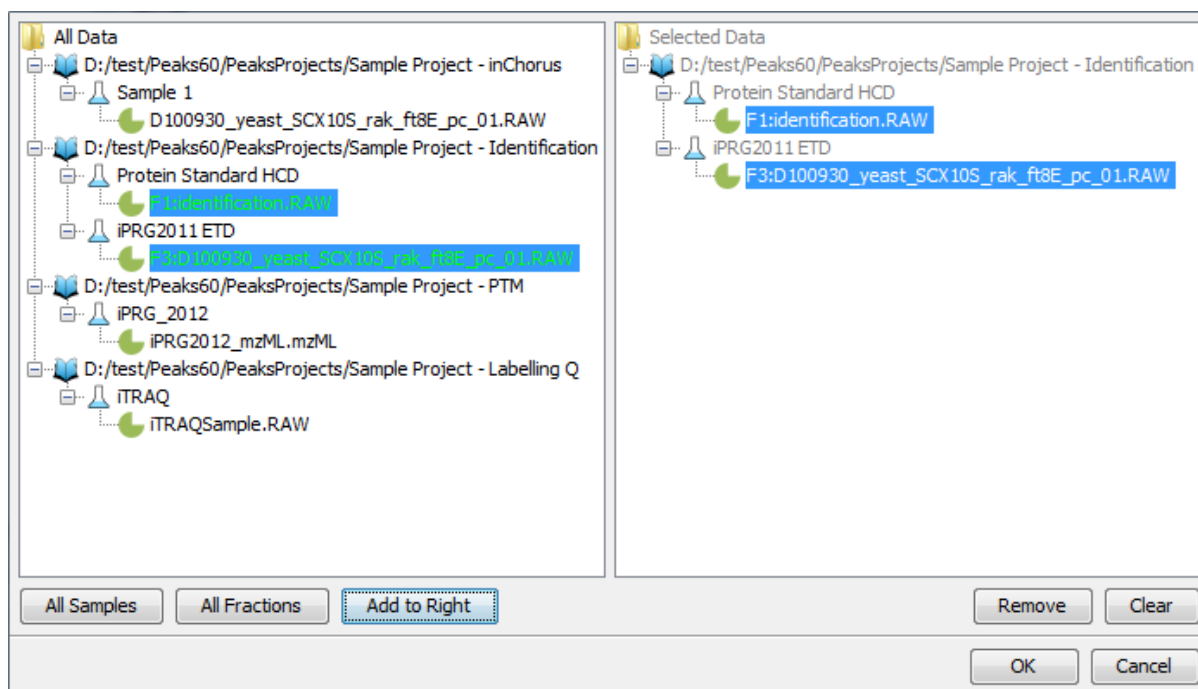
## Note

All files that are loaded into a single workflow will be processed in exactly the same way, using the exact same parameters. To run some files differently than others, you must set up separate workflows.

Once the data is selected, you can specify parameters for the identification analysis steps one-by-one by clicking the other buttons in the workflow dialogue. Please refer to the chapters on each individual function for details on setting up the function-specific parameters.

## Note

PEAKS DB, PEAKS PTM, and SPIDER are now combined together (see Chapter 9, *Peptide, PTM, and Mutation Identification (PEAKS DB, PEAKS PTM, and SPIDER)*) and are optional in the Identification Workflow. You can uncheck them if you do not want to perform those functions.

# 2. Quantification Workflow

The quantification workflow is similar to the identification workflow, but it has an additional step for quantification, where the quantification parameters to perform labeled quantification can be defined. Please refer to the chapters on each individual function for details on setting up the function-specific parameters.
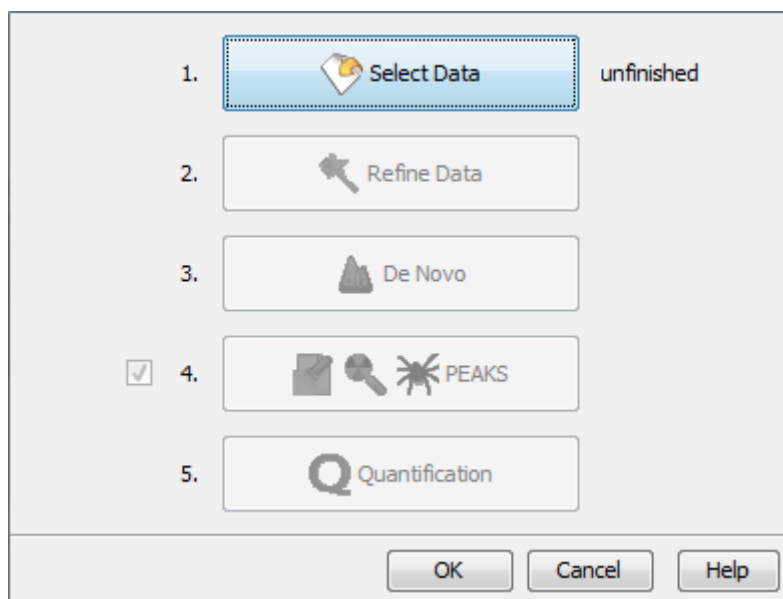
### Note

Only Reporter Ion Quantification (eg. iTRAQ) or Precursor Ion Quantification (eg. SILAC) workflows can be set in the quantification workflow. For the Label Free workflow, use the Project Wizard.

# 3. inChorus Workflow

The inChorus workflow is similar to the identification workflow, but offers the ability to specify inChorus parameters and invoke multiple search engines. Refer to Chapter 11, *Combining Multiple Database Search Engines with PEAKS inChorus* for details on inChorus parameters.
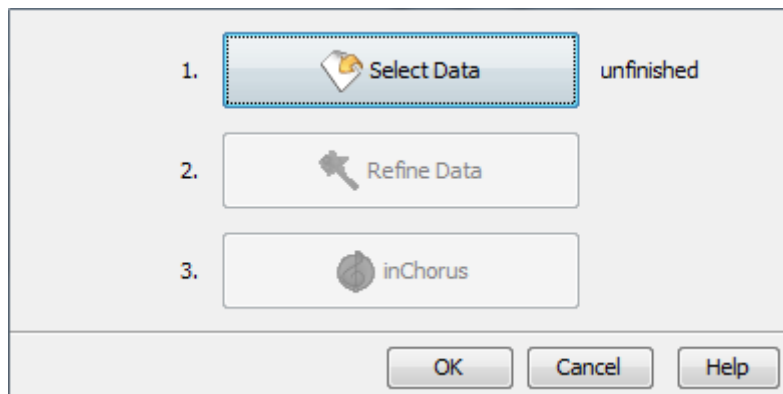
# Chapter 16. Exporting Data/Reports and Printing

PEAKS offers a rich collection of exporting functions to allow users to create reports and share the analysis results with collaborators, colleagues, and clients. The supported formats include HTML, Comma Separated Values (CSV), pepXML, mzIdentML, and various image formats for image exporting. Labs with in-house software can easily make use of the CSV files in their own analysis workflows. The exported results in HTML can be viewed with a web browser. The entire exported result directory can be zipped and emailed to colleagues, or the whole directory can be put directly on a website.

## 1. Exporting Data

Spectrum data can be exported to a number of file formats including mzXML, DTA, MGF, and PKL. To do so, right-click on the sample node or the data file that is to be exported and select the desired exporting format.
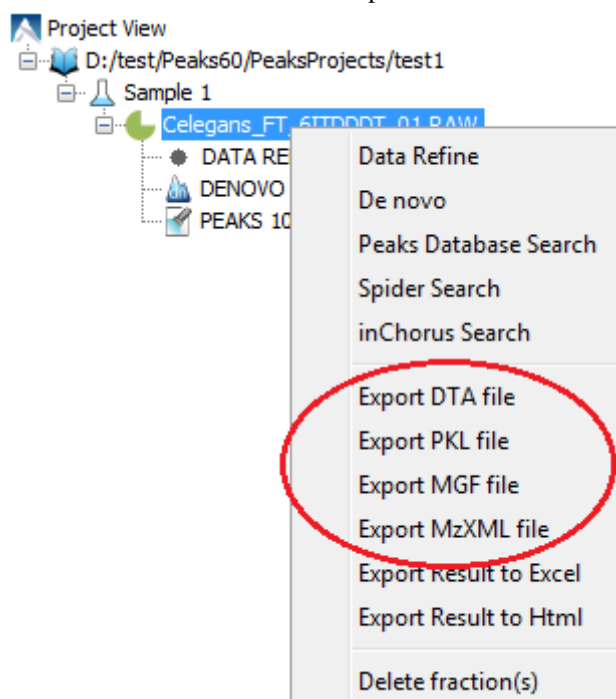


Selecting the Export DTA file command will open a dialogue prompting the folder name and location to which the DTA files will be exported. For MGF and PKL exports, the dialogue will ask for a name and a location for the file.

Selecting Export MzXML File command will open the "Export mzXML File" dialogue.
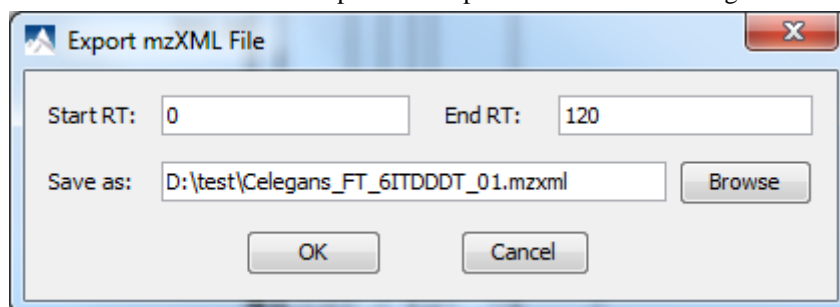


Enter the starting and ending retention times in the appropriate boxes. Then, click the Browse button to select a destination to save your file.

# 2. Exporting Result From Project View

Fraction information and peptide identification results can be exported to Excel or HTML format from the pop-up menu in the "Project View". Right -lick on a project node, a sample node, or a fraction node, and select the options Export Result to Excel or Export Result to Html. The exported result contains fraction information, PTM information, and a list of peptide identifications of the selected fractions.
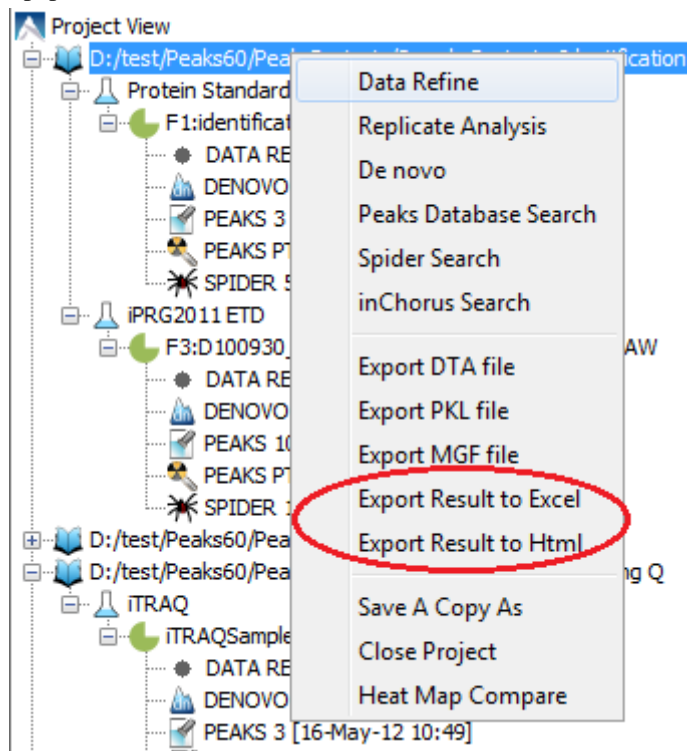


More function-specific export options are available from the "Summary" view of the respective results. The following sections describe in detail the function-specific exporting options available in various formats.

# 3. Exporting *De Novo* Results

The PEAKS *de novo* sequencing result can be exported to CSV, HTML, and pepXML formats. All exporting functions are available through the "Summary" view panel.

## 3.1. Exporting Summary and Peptides

To export results, press the Export button in the title bar of the "Summary" view panel. The following export dialogue will appear. Select the items that you want to export from the available options. Select the output location and click the Export button to save the selected result components to the specified location.
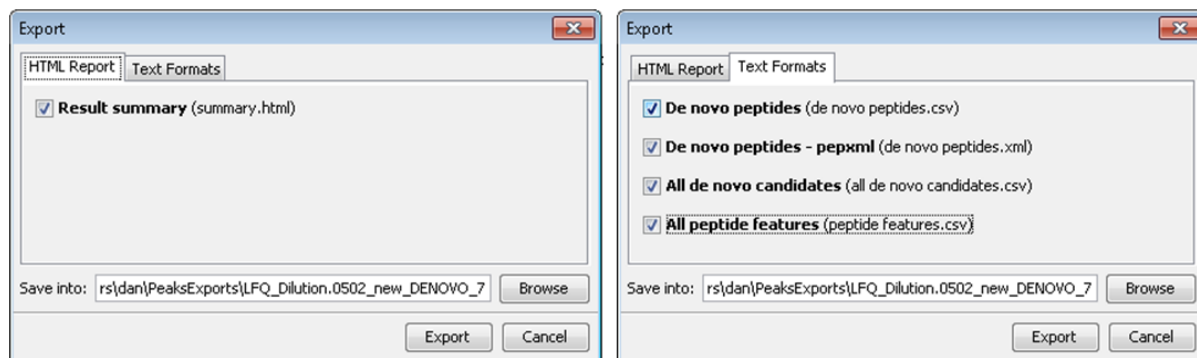


The export options are grouped into "HTML Report" and "Text Formats" based on the output format.

**HTML Report.** This will generate a summary report in the specified location. After the completion of exporting, the result will be opened in the default browser automatically. The following exporting options are available:

• **Result summary :** The "Summary" view page will be saved as `summary.html` file in HTML format in the specified location.

**Text Formats.** The following exporting options are available in various text formats:

• **De novo peptides :** The peptides identified by *de novo* sequencing will be saved to `de novo peptides.csv` file in Comma Separated Values (CSV) format in the specified folder.

• **De novo peptides - pepxml :** In addition to CSV format, the peptides can be saved in pepXML format.

• **All de novo candidates :** All *de novo* candidates for each spectrum will be saved to `all de novo candidates.csv` file in CSV format.

• **All peptide features:** All predicted peptide features found in the LC/MS information will be sorted by m/z. If a *de novo* result is associated with the peptide feature above the ALC filter it will be displayed as well.

# 3.2. Exporting Images

The Annotated Spectrum, Ion Match table, Error Map, and Spectrum Alignment can each be exported to image files. To do so, position the cursor on any of those items in the result panel, right-click to view the pop-up menu, and select the Export Image command from the menu.
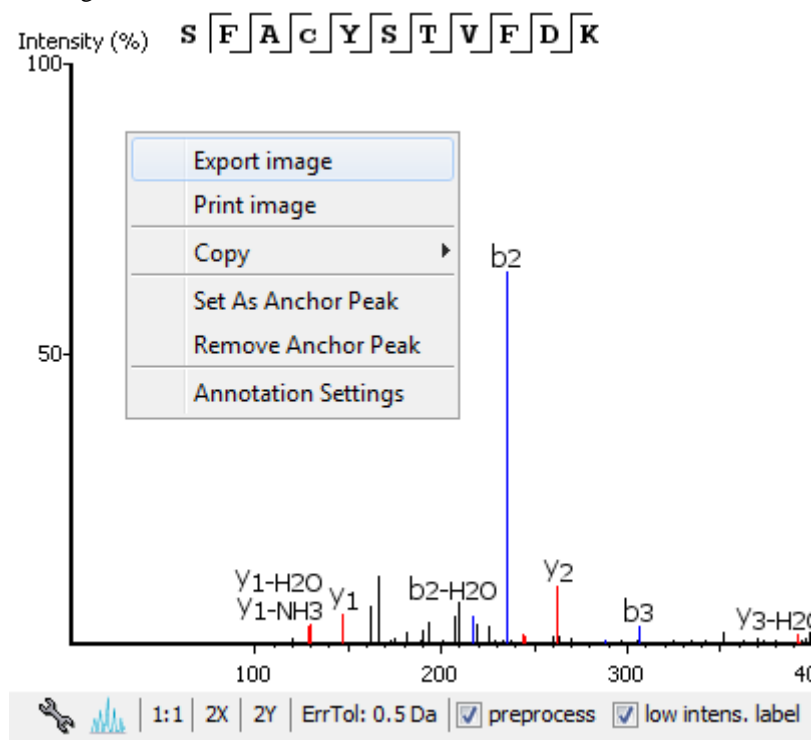


This will bring up the "Export Images" dialogue for selecting the result items to export.

Select the desired result elements from the "Image Types" list. The "Basic Options" tab offers choices for location and image size options to which can saved at either "Web (Smallest images, suitable for viewing online)" or "Print (Oversampled images, suitable for printing)".



The "Advanced Options" panel offers choices for editing the scaling factor, the file format, the resolution, and the oversample factor. PEAKS supports BMP, GIF, JPEG, PNG, and SVG image formats. After setting all parameters, click the OK button to export the selected result item to an image.

# 4. Exporting Database Search Results

The exporting mechanism for PEAKS DB, PEAKS PTM, and SPIDER results are the same as that of a *de novo* result with the exception of the number and the type of available exporting options. All exporting functions are available through the "Summary" view panel.

## 4.1. Exporting Summary, Proteins, and Peptides

To export the result, press the Export button in the title bar of the "Summary" view panel. The following export dialogue will appear.

The export options are grouped into "HTML Report", "Text Formats", and "For Third Party" based on the output format. Select the items that you want to export from the available options. Select the output location and click the Export button to save the selected result components to the specified location.

**HTML Report.**    This will generate single or multiple HTML reports in the specified location. After the completion of result exporting, the index file for the reports, `protein.html`, will be opened in the default browser automatically. The following exporting options are available:

- **Summary view:** The "Summary" view page will be saved as a `summary.html` file in HTML format in the specified location.

- **Protein coverage:** The coverage pane will be saved for each protein.

- **Supporting peptides:** A list of supporting peptides will be saved for each protein.

- **Best unique PSM:** The best unique PSM will be saved for each protein.

- **Export peptide list in html:** creates a separate HTML-containing peptide identifications.

An individual protein will have its own HTML output file, where the corresponding protein coverage, the supporting peptides, and the best unique PSM are gathered. Select "Collect all the above details in a single html" to collect all the protein reports in a single `protein.html` file.

**Text Formats.**     The following exporting options are available for various text formats:

- **Proteins:** The list of protein identifications, filtered by the protein filters (in the "Summary" view) and the "Show" filter (in the "Protein" view above the Protein table), will be saved to `proteins.csv` in Comma Separated Values (CSV) format.

- **Supporting peptides:** A list of supporting peptides of each protein identification, filtered by the peptide filters and protein filters (in the "Summary" view) and the "Show" filter (in the "Protein" view above the Protein table), will be exported to `protein-peptides.csv`. This usually contains more entries than the Peptide table as a peptide can be identified in several proteins and, therefore, can be reported multiple times in this file.

- **DB Search Peptides:** All of the identified peptides and details.

- **DB search peptide-spectrum matches:** The selected peptide-spectrum matches (PSMs), their display controlled by the peptide filters in the "Summary" view, will be exported to `DB search psm.csv`. Peptides differentiated with only I/L isoform are represented by separate entries. As a result, the number of entries in this file might be bigger than the number of PSMs mentioned in the "Summary" view.

- **De novo only peptides:** A list of good *de novo* sequences that do not have good, or any, database matches will be saved to `de novo only peptides.csv`.

- **Proteins - fasta:** A list of protein identifications will be saved to `proteins.fasta`.

- **Peptide-spectrum match ions:** All theoretical and actual fragment ions associated with each peptide-spectrum match, including fragment mass error.

- **Peptides - pepxml:** A list of peptide-spectrum matches, their display controlled by the peptide filters in the "Summary" view, will be saved to `peptides.pep.xml` in pepXML format. The number of the "spectrum_query" elements will be the same as the number of PSMs mentioned in the "Summary" view, and the number of the "search_hit" elements will be the same as the number of PSMs counted from Peptide table ( summing up #Spec column).

- **Peptides - mzidentml (version 1.0.0):** A list of peptide-spectrum matches will be saved to `peptides_1_0_0.mzid` in mzIdentML format version 1.0.0.

- **Peptides - mzidentml (version 1.1.0):** A list of peptide-spectrum matches will be saved to `peptides_1_1_0.mzid` in mzIdentML format version 1.1.0.

- **All peptide features:** For each data file in the search, a list of predicted peptide features is exported and sorted by mass. If a peptide or *de novo* only result is associated with it the peptide sequence will be given as well.

- **De novo only peptides - pepxml:** A list of good *de novo* sequences with either poor or non-existent database matches will be saved in `de novo only peptides.xml` in pepXML. format.

- **De novo only spectra - mgf:** All spectra missed by an identification search but with *de novo* results passing the ALC filter set on the "Summary" view.

**For Third Party.**     PEAKS analysis results can be used in other third party software packages for validation or further analysis of the data. In addition to the result files, some third party software packages also require spectrum data files in some specified format. The "For Third Party" pane provides the following exporting options to support such usage of PEAKS' results.

- **For Scaffold: peptides - mzIdentML:** In addition to exported results in supported mzIdentML formats, versions 1.0.0 and 1.1.0, PEAKS will export the spectrum data in MGF format. The exported mzIdentML file and MGF files can be loaded into Scaffold for further analysis, such as PTM analysis.

## Note

Peptides identified by SPIDER may contain residue substitutions, insertions, and deletions. The pepXML format does not specify the encoding for substitution, insertion, and deletion, and the mzIdentML format only specifies the encoding for substitution. Therefore, when exporting from a SPIDER result, those peptides may not be correctly interpreted by third party software.

• **For Skyline: peptides - pepXML:** In addition to exporting the peptide-spectrum matches to pepXML format, PEAKS will export the spectrum data in mzXML format. The exported pepXML file and mzXML files can be loaded into Skyline for spectrum library generation.

## 4.2. Exporting Images

From the "Peptide" view and the "*De novo* only" view, the Annotated Spectrum, Ion Match table, Error Map, or Spectrum Alignment can be exported to an image file. To do so, position the cursor on any of those items in the result panel, right-click, and then select the Export Image command from the menu. Refer to Section 3.2, "Exporting Images" for details.

### 4.2.1. Exporting Protein Coverage

To export the protein coverage pane, click the tools ✎ button in the coverage pane and select the Save image as... command from the menu. Select the resolution of the image and the location to save. Click the Export button. The PTM legends can be included in the exported image.
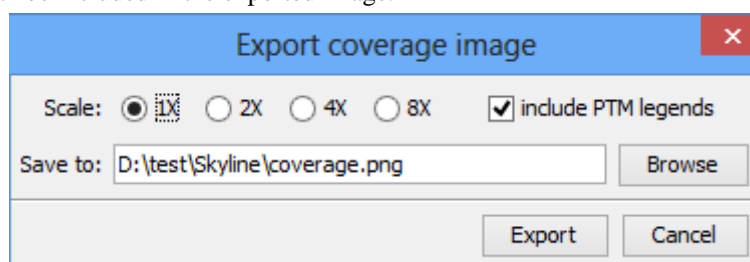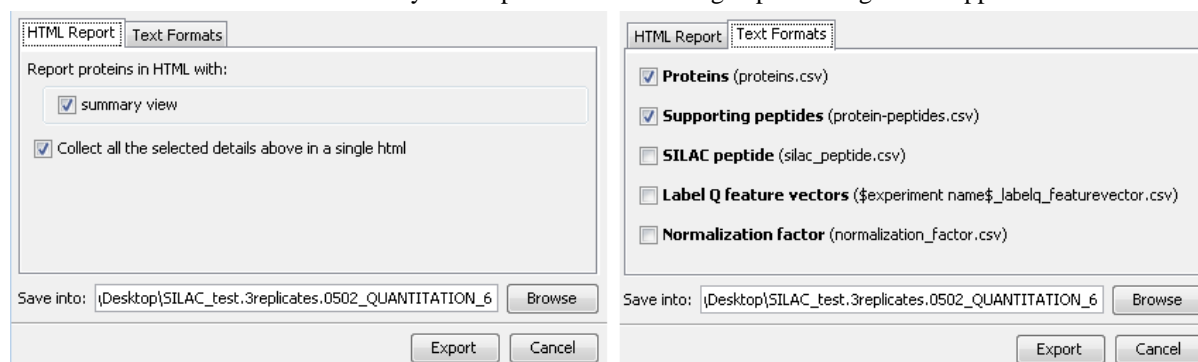


# 5. Exporting Quantification Results

PEAKS Q (labeled and label free quantification) exporting functions are also similar to that of *de novo* or PEAKS DB. All export functions are available through the "Summary" view panel.

## 5.1. Exporting Precursor Ion Labelling Quantification Results

PEAKS Q results can be exported to other supported formats. To export the quantification results, press the Export button in the title bar of the "Summary" view panel. The following export dialogue will appear.



**HTML Report:** The options are the same as those for PEAKS DB. For PEAKS Q, except that only the quantified proteins are used in the exporting. All information found in the "Summary", "Protein", and "Peptide" views of the

quantification result will be found in the export. See Section 4.1, "Exporting Summary, Proteins, and Peptides" for details.

**Text Formats:**

- **Proteins:** These are quantified proteins that pass the filters set in the "Summary" view and associated details

- **Supporting peptides:** These are supporting peptides of quantified proteins and associated details

- **SILAC peptide:** All details of all identified and quantified peptides passing the peptide filters are listed here. Duplicate peptides are grouped.

- **Label Q feature vectors:** All details of all identified and quantified peptides passing the peptide filters are listed here. Duplicates are not grouped. If several data files were used in the search a separate export will be created for each file.

- **Normalization factor:** Details of the normalization factor set in the "Summary" view are listed here.

Select the output location and click the Export button to save the selected result components to the specified location.

**Export Images.** The Annotated Spectrum, the Ion Match table, the Error Map, the Spectrum Alignment, and the Protein Coverage can be exported to an image file. Refer to Section 3.2, "Exporting Images" for details.

## 5.2. Exporting Reporter Ion Labelling Quantification Results

PEAKS Q results can be exported to other supported formats. To export the quantification results, press the Export button in the title bar of the "Summary" view panel. The following export dialogue will appear.



**HTML Report:** The options are the same as those for PEAKS DB, except that only the quantified proteins are used in the exporting. All information found in the "Summary", "Protein", and "Peptide" view of the quantification result will be found in the export. See Section 4.1, "Exporting Summary, Proteins, and Peptides" for details.

**Text Formats:**

- **Proteins:** These are the quantified proteins that pass the filters set in the "Summary" view and their associated details.

- **Supporting peptides:** These are the supporting peptides of the quantified proteins and their associated details.

- **Label Q search peptides:** All details of all identified and quantified peptides passing the peptide filters are listed here. Duplicate peptides are grouped.

- **Label Q feature vectors:** All details of all identified and quantified peptides passing the peptide filters are listed here. Duplicates are not grouped. If several data files were used in the search, a separate export will be created for each file.

- **Normalization factor:** Details of the normalization factor set in the "Summary" view are listed here.

Select the output location and click the Export button to save the selected result components to the specified location.

**Export Images.** The Annotated Spectrum, the Ion Match table, the Error Map, the Spectrum Alignment, and the Protein Coverage can be exported to an image file. Refer to Section 3.2, "Exporting Images" for details.

## 5.3. Exporting Label-Free Quantification Results

PEAKS LFQ results can be exported to other supported formats. To export the quantification results, press the Export button in the title bar of the "Summary" view panel. The following export dialogue will appear.
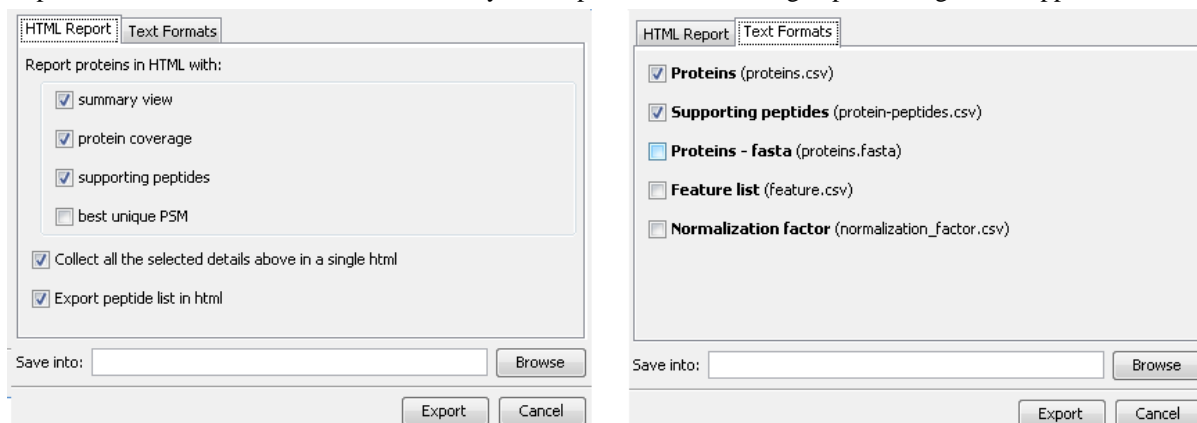


**HTML Report:** The options are the same as those for labeled quantification. See Section 5.1, "Exporting Precursor Ion Labelling Quantification Results" for details.

**Text Formats:**

- **Proteins:** These are the quantified proteins that pass the filters set in the "Summary" view and their associated details.

- **Supporting peptides:** These are the supporting peptides of the quantified proteins and their associated details.

- **Proteins - fasta:** This is a database file of all quantified protein sequences above the "Summary" view filters.

- **Feature List:** All predicted peptide features and their associated quantification information. Both identified and unidentified peptide features are included.

- **Normalization factor:** Details of the normalization factor set in the "Summary" view are listed here.

Select the output location and click the Export button to save the selected result components to the specified location.

**Export Images.** The XIC intensity curve, the LC/MS view of the sample features, and the RT alignment figure of the samples can be exported to image files from the "Features" view. Click the ⬇ button of the corresponding tabs on the "Features" view. Select the resolution and the location to save the file.

# 6. Export inChorus Result

The inChorus exporting function contains the same exporting options as PEAKS DB, but unlike PEAKS DB, the exported results will contain inChorus scores and individual search engine scores for the supporting peptides of the protein identifications. See Section 4, "Exporting Database Search Results" for details.

# Chapter 17. Advanced Configuration and Environment Preferences

## 1. PEAKS Environment Preferences

This section describes the settings of the environment preferences. These settings include general preferences, raw file converter preferences, search engine preferences, and spectrum annotation preferences.

To begin, click the "Preferences" toolbar icon  or select the Window | Preferences command from the menubar to open the "Preferences" dialogue. Select the preferences category from the left to view the options available for that category.
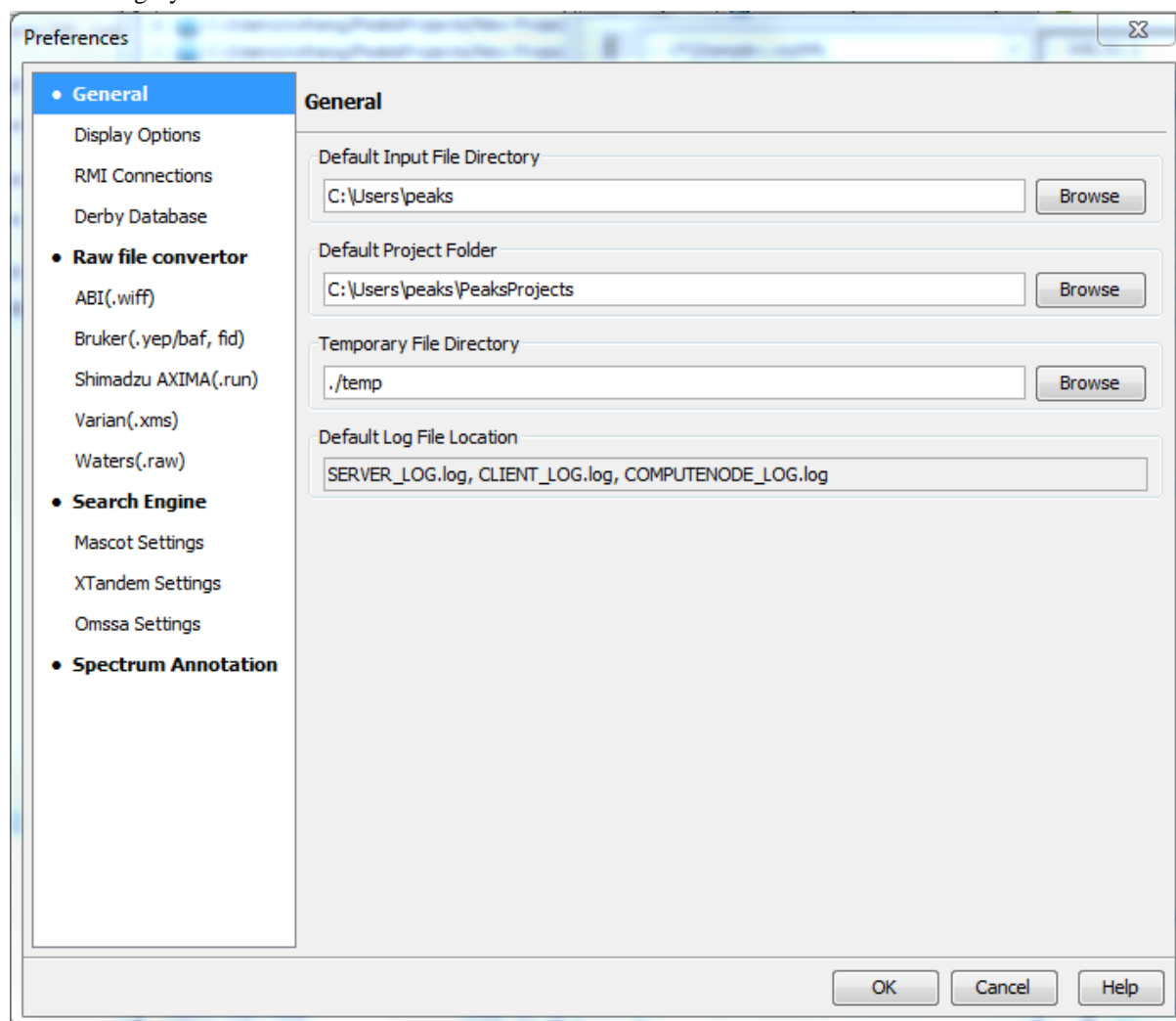


## 1.1. General Preferences

**Default Input File Directory:** Select the Browse button to change the directory that will appear when adding data to a project.
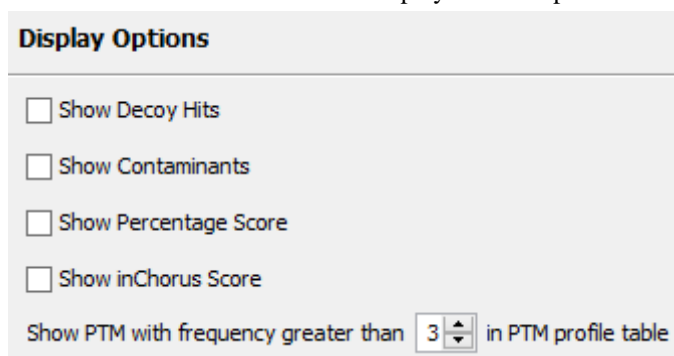
**Default Project Folder:** PEAKS uses `[USER_HOME]/PeaksProjects` as the default folder for project files, where `USER_HOME` is the user home directory in your system. Select the Browse button to change this location. Please make sure this directory is readable/writable by PEAKS.

**Temporary File Directory:** PEAKS uses `[PEAKS_HOME_DIRECTORY]/temp` as the default temporary file output directory, where `PEAKS_HOME_DIRECTORY` is the location where PEAKS is installed. Select the Browse button to change this location. Please make sure this directory is readable/writable by PEAKS.

**Default Log File Location:** The log files for PEAKS can be found at `[PEAKS_HOME_DIRECTORY]` by default. The locations of these files cannot be changed.

## 1.1.1. Display Options

Click "Display Options" from the menu on the left side to display interface preferences on the right side.



**Show Decoy Hits:** Check this to display protein and peptide hits from the decoy database within the PEAKS DB results.

**Show Contaminant:** Check this to display contaminants that were identified during a PEAKS DB search. To check for contaminants, ensure that a contaminant database is selected in the PEAKS Search window when setting up a search.

**Show Percentage Score:** PEAKS uses -10lgP to display its results by default. Check this to view the percentage score along with the -10lgP score in both the "Peptide" and "Protein" views. These changes are applied to the exported files of the PEAKS DB results.

**Show inChorus Score:** Check this to display the percentage score in both the "Peptide" and the "Protein" views of the inChorus results.

Set the PTM display threshold by selecting the minimum PTM frequency in the PTM profile table. If there are fewer instances of a PTM identification in a protein identification than the minimum, it will not be displayed in the PTM profile table.
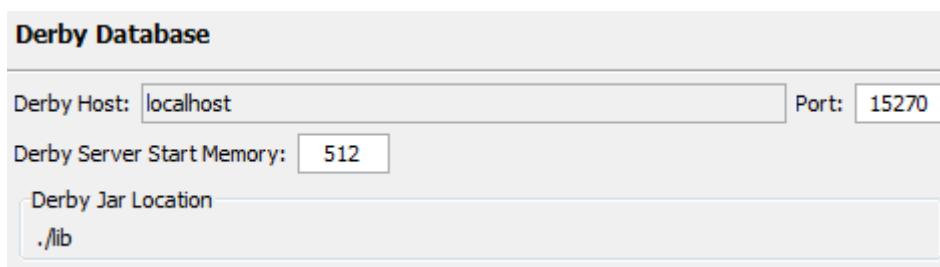
## 1.1.2. RMI Connections

Click "RMI Connections" from the menu in the left side to show the RMI (Java Remote Method invocation) connections preferences on the right side.



The default port numbers for the Server, Client and Worker will appear. The port numbers can be changed if conflicts arise. Contact technical support at BSI `<support@bioinfor.com>` for more information.

## 1.1.3. Derby Database

Click "Derby Database" from the menu on the left side to show the derby database preferences on the right side.

**Derby Host:** The name of the "Derby Host", as well as the "Port" number, will come up by default. The port number can be changed.

**Derby Server Start Memory:** The amount of memory used to start the derby server will also come up by default but can be changed if more memory is available. However, it is not recommended to change this from the default setting. To increase performance, use the performance configuration utility (see Section 5, "PEAKS Performance Configuration").

**Derby Jar Location:** The "Derby Jar Location" panel will list the location of the derby jar file by default. This is displayed to find its location. This location cannot be changed.

# 1.2. Raw File Converter Preferences

This section allows users to change the preferences for the raw file converters of the following instruments: AB SCIEX, Bruker, Shimadzu, Varian, and Waters.

## 1.2.1. ABI (.wiff)

Click the "ABI SCIEX (.wiff)" option under the "Raw file converter" section from the menu on the left side to show the preferences for the AB SCIEX instrument raw file converter.

### Note

Refer to Section 8.4.1, "TrippleTOF, QSTAR, or QTRAP" for details on AB SCIEX (.wiff) preferences.

## 1.2.2. Bruker (.yep/baf, fid)

Click the "Bruker (.yep/baf, fid)" option under the "Raw file converter" section from the menu on the left side to display the Bruker instrument raw file converter preferences.

### Note

Refer to Section 8.5, "Bruker Data" for details on Bruker instrument preferences.

## 1.2.3. Shimadzu AXIMA (.run)

Click the "Shimadzu AXIMA (.run)" option under the "Raw file converter" section from the menu on the left side to show the Shimadzu instrument raw file converter preferences.

### Note

Refer to Section 8.6, "Shimadzu Data" for details on Shimadzu instrument preferences.

## 1.2.4. Varian (.xms)

Click the "Varian (.xms)" option under the "Raw file converter" section from the menu on the left side to display the Varian instrument preferences.

## Note

Refer to Section 8.7, "Varian" for details on Varian instrument preferences.

### 1.2.5. Waters (.raw)

Click the "Waters (.raw)" option under the "Raw file converter" section from the menu on the left side to display the Waters instrument preferences.
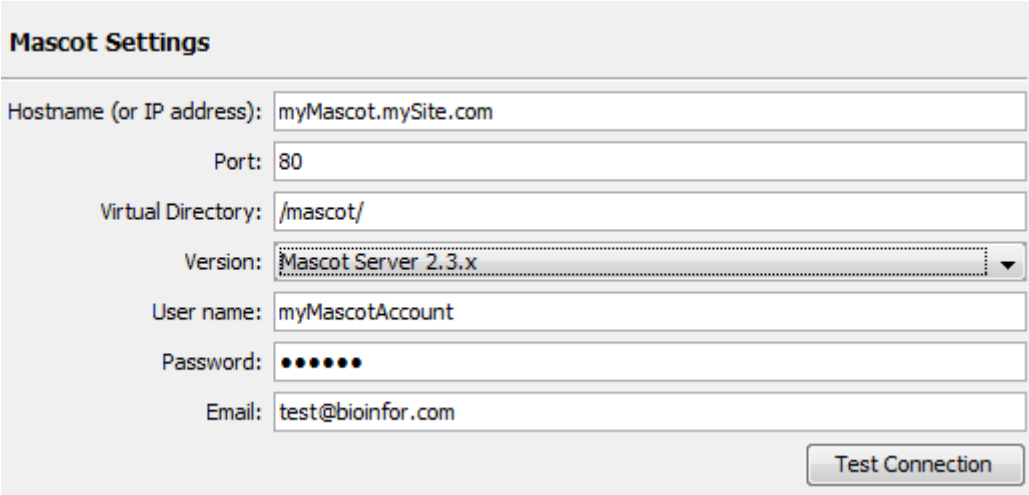
## Note

Refer to Section 8.2, "Waters/Micromass (MassLynx) Data" for details on Waters instrument preferences.

# 1.3. Search Engine Preferences

This section allows users to configure preferences for the following search engines: Mascot, X!Tandem, and OMSSA.

### 1.3.1. Mascot Settings

Click "Mascot Settings" on the left side to display the Mascot preferences.
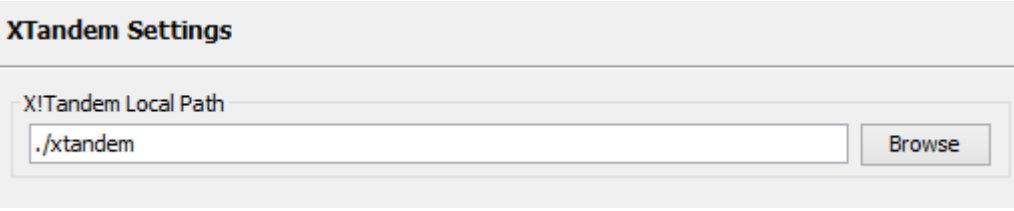


These parameters specify how PEAKS accesses the Mascot server (if applicable). Enter the hostname (or an IP address), the port, the virtual directory, and the Mascot server version. As well, enter the username, password, and email address. To make sure that everything is entered correctly and that the server is working, click the Test Connection button. The port and the virtual directory match the above settings for most servers.

### 1.3.2. X!Tandem Settings

Click "X!Tandem Settings" on the left side to display the X!Tandem preferences.



As PEAKS provides a local copy of X!Tandem, upon installation, a default path will appear in the Local Settings section. To use another license/location for X!Tandem, click the Browse button to tell PEAKS where to find the search engine.

### 1.3.3. OMSSA Settings

Click "OMSSA Settings" on the left side to display the OMSSA preferences.



As PEAKS provides a local copy of OMSSA upon installation, a default path will appear here. To use another license/location for OMSSA, click the Browse button to tell PEAKS where to find the desired search engine.

# 1.4. Spectrum Annotation Preferences

Click "Spectrum Annotation" from the left side to open the spectrum annotation preferences:



The spectra in the PEAKS results can be annotated by selecting ion types from the thorough collection of ions that PEAKS offers. The selected ion types will be displayed in the "Ion Match" table, as well. It is possible to annotate the spectrum with various ions for both CID and ETD. By default, y-ion, y-H2O, y-NH3, y(2+), b-ion, b-H2O, b-NH3, and b(2+) are selected.

## Note

To reset the settings to the PEAKS defaults, use the "Reset default" button.

**Show Decimal Places:** Select the number of decimal places that will appear in the ion table and the spectrum view. The default is set to two decimal places.

**m/z on Fragmentation:** Select this to display the m/z value on top of the annotated ions.

**m/z on Unannotated:** Select this to display the m/z value on top of the peaks without ion matches.

**sequence Fragmentation:** Select this to display the sequence fragmentation on the top-left corner of the "Spectrum Annotation" view.

**In Place Ion Info:** Ion information, m/z value,and relative intensity are all displayed in a pop-up within the "Spectrum Annotation" view when this option is checked and the cursor is placed on a peak.
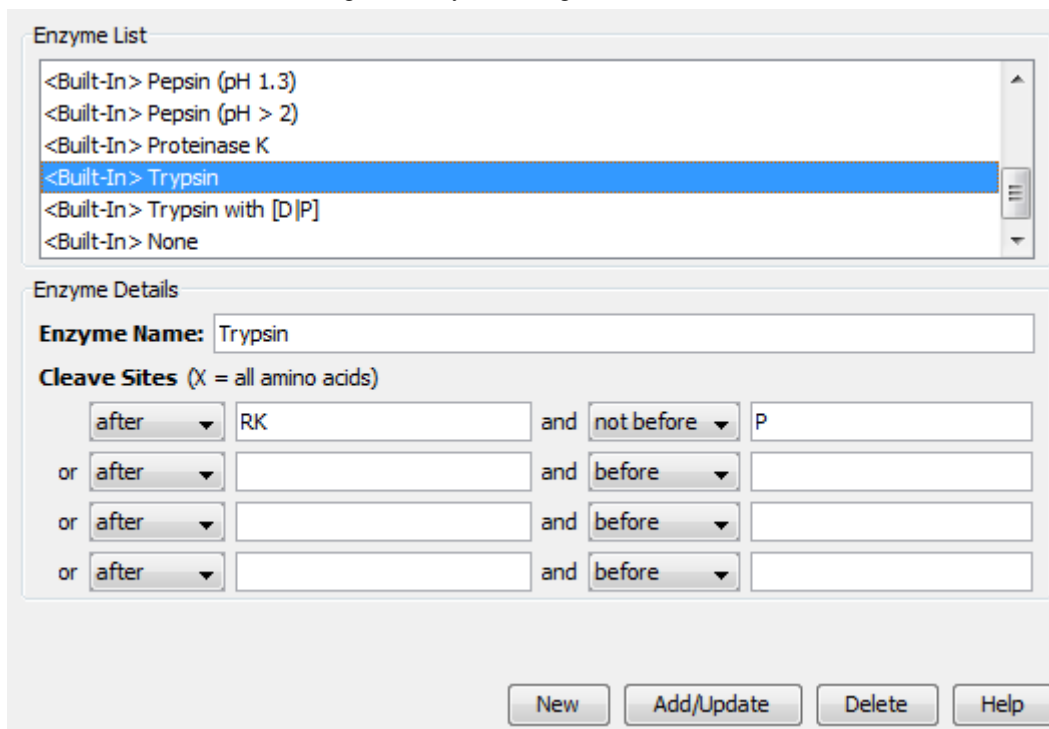
**Intensity -** Set the intensity threshold for spectrum annotation to low (2%), medium (5%), or high (10%). To apply this intensity threshold, select the "intensity threshold" checkbox in the "Spectrum Annotation" view (see Section 3.2.2, "Spectrum Annotation").

# 2. PEAKS Configuration

This section describes the configuration of enzymes, PTMs, databases, instruments, and parameters. To begin, click the Configuration toolbar icon ⚒ or select "Configuration" from the "Window" menu.

## 2.1. Enzyme Configuration

PEAKS can use almost any single enzyme or combination of enzymes in your analysis. Select built-in enzymes from the extensive list provided in PEAKS or define a new one. From the "Configuration" window select "Enzyme" from the left side menu to change the enzyme configuration.

**Built-in Enzymes.** All of the built-in enzymes within PEAKS are listed in the "Enzyme List". Click on one of these built-in enzymes to view the information about that enzyme in the "Enzyme Details" panel.

## Note

A built-in enzyme cannot be deleted or edited.

**Create a New Enzyme.** Provide the name of the new enzyme in the "Enzyme Name" field and specify how the custom enzyme will cleave the protein between two amino acids to create peptides in the "Enzyme Details" panel. The letter, X, denotes *any amino acid at this position*, while amino acids within the { and } brackets indicate any

amino acid except for the one in the brackets. Choose where the cleavage sites are by selecting *after* or *not after* and *before* or *not before* to specify the range. Add multiple amino acids to indicate that cleavage happens before or after any of the stated amino acids. For example, after RK means after R or K not after R and K.

Click the Add/Update button to save the changes. The new enzyme will now appear in the "Enzyme List", where it can be accessed later. To delete a customized enzyme, select the appropriate enzyme and click the Delete button.

## Note

For information on defining new enzymes "on the fly" for PEAKS *de novo* or PEAKS DB, refer to Section 2.2, "Enzyme".

# 2.2. PTM Configuration

From the "Configuration" window, select "PTM" from the left side menu to change the PTM configuration.

| Name | Mono mass | Residue site |
|---|---|---|
| 4-hydroxynonenal (HNE) | 156.1150 | [CHK] |
| Acetylation (K) | 42.0106 | [K] |
| Acetylation (N-term) | 42.0106 | [X]@N |
| Acetylation (Protein N-term) | 42.0106 | [X]@N |
| Amidation | -0.9840 | [X]@C |
| Ammonia-loss (C@N-term) | -17.0265 | [C]@N |
| Ammonia-loss (Protein N-term) | -17.0265 | [TS]@N |
| Beta-methylthiolation | 45.9877 | [C] |
| Biotinylation | 226.0776 | [K], [X]@N |
| Carbamidomethylation | 57.0215 | [C] |
| Carbamylation | 43.0058 | [K], [X]@N |
| Carboxylation (E) | 43.9898 | [E] |
| Carboxymethyl | 58.0055 | [C] |
| Citrullination | 0.9840 | [R] |
| Deamidation (NQ) | 0.9840 | [NQ] |
| Dehydration | -18.0106 | [DYTS], [NQ]@C |
| Dimethylation | 28.0313 | [KRN], [X]@N |
| Dioxidation (M) | 31.9898 | [M] |
| FAD | 783.1415 | [CHY] |
| Farnesylation | 204.1878 | [C] |
| Formylation | 27.9949 | [K], [X]@N |
| Formylation (Protein N-term) | 27.9949 | [X]@N |
| Geranyl-geranyl | 272.2504 | [C] |
| Guanidination | 42.0218 | [K] |
| HexNAcylation (N) | 203.0794 | [N] |
| Hexose (NSY) | 162.0528 | [NSY] |
| Homoserine | -29.9928 | [M]@C |
| Homoserine lactone | -48.0034 | [M]@C |
| Hydroxylation | 15.9949 | [PKDNRY] |
| Lipoyl | 188.0330 | [K] |
| Methylation | 14.0156 | [TSCKRHDE], [X]@N |

Recent | Common | Uncommon | Artificial | Customized          View | New

**Built-in PTMs.** The built-in standard PTMs within PEAKS are listed under three separate PTM lists: "Common", "Uncommon", and "Artificial". The "Common" list contains the most commonly used PTMs, the "Uncommon" list contains less frequently used PTMs, and the "Artificial" list contains the PTMs that can only be induced artificially. The most recently used PTMs are listed in the "Recent" tab. All of the user-defined PTMs are listed in the "Customized" tab. Double-clicking on any of these PTMs will display the PTM details in the "PTM Info" pop-up dialogue. The same information can be viewed by selecting a PTM from a list and clicking the View button. For details on PTMs, refer to Section 2.3, "PTM".

**Create a new PTM.**    Click on the New button to display the "New PTM" dialogue. Provide the information about the PTM. The required information about the PTM is listed below:



- **PTM Name:** This name will appear in the PTM list for future use after it is saved.

- **PTM abbreviation:** The PTM expressed in a shortened form.

- **Mass (Monoisotopic):** The mass that the residue gains or loses as a result of the PTM.

- **Residues that can be modified:** Enter residues that can be modified anywhere, or residues that can only be modified if they are at the N- or C-terminus.

- **Formula:** The chemical formula of the PTM. This should correspond to the mass listed above.

- **Rule:** This field can be used to enter a comment about the PTM, to be used for your own reference.

Click the OK button to save the changes. The new PTM will now appear in the "Customized" PTM list, where it can be accessed later. To delete a customized PTM, select the appropriate PTM from the list and click the Delete button.

## Note

For information on defining new PTMs "on the fly" for PEAKS *de novo* or PEAKS DB, refer to the section Section 2.3, "PTM".

## 2.3. Labeled Q Method Configuration

From the "Configuration" window, select "Labeled Q Method" from the left side menu to change the labeled Q method configuration.



The built-in labeled Q methods are listed in the methods list. Select a method from the list to view detailed information in the "Quantification Method Detail" panel. Methods with the R icon beside the name are reporter level methods. Methods with the P icon beside the name are precursor level methods.

**Name and Method Type:** The name and the type of quantification method. The method type can be "Reporter Ion Quantification" or "Precursor Ion Quantification".

**PTM:** The post-translational modification associated with the selected Quantification method; including the residue site and the location on the residue at which it is attached.

**Label Options:** List of quantifiable labels. For reporter ion quantification methods, the Label Options pane will contain the name and reporter ion mass. For precursor ion quantification methods, the Label Options pane will contain the sample name, the modification, and the modification details.

**Create a New Method.** To create a new labeled Q method, click the New button to open the "New/Edit Quantification Method" dialogue, where the quantification method details can be specified.

The modification target, the modification mass, and the PTMs for a "Reporter Ion Quantification" method. Use Add Label and Delete Current Line to add or remove a label, respectively. Each label is defined by the sample name and the reporter ion mass.



Use Add row and Delete Row to respectively add or remove a label for a "Precursor Ion Quantification" method. Each label is defined by the sample name, the added mass, the target residue, and the labeling efficiency. If one sample has multiple labels with different mass shifts, a user can add multiple labels with the same sample name. These labels will contribute to the same sample in the ratio. The modification for each label must be selected from the PTM lists, which can be accessed by clicking the button in the "Modifications" column.

## 2.4. Database Configuration

To use the PEAKS DB function to identify proteins through a database search, PEAKS must have access to a protein or an EST database in FASTA format (the standard format for popular public sequence databases). PEAKS can be configured to use existing databases on the system or to use databases downloaded from a public server. Additionally, taxonomy options may be specified with certain databases.

From the "Configuration" window, select "Database" from the left menu to change the database configuration. The "Database List" on the top shows all the configured databases in the system. Select a database from the list to view the detailed information about the database in the "Database Details" panel.

**Create a new database.** Click the New button to create a new sequence database entry or select a database from the "Database List". The database can be configured in the "Database Details" area below. Click the Add/Update button to save any changes. Refer to Chapter 6, *Adding a Sequence Database* for details on adding a new database.

**Delete a previously configured database.** To delete a database file, select the database to be deleted from the "Database List" and click the Delete button at the bottom.

**Moving/Updating a database.** To move a database to another directory, the location must be updated in PEAKS. Select the database and specify the new location using the Browse button next to the "Path" field. Then, click Add/Update to save the new settings. If the database location is invalid, the database name will appear in *red* in the list of databases. Any protein identification using that database will fail. If an update is made to the database file (perhaps by downloading the latest database file and overwriting the old database file), PEAKS will show the database information in *light gray*. A light gray color could also mean that the database does not have header information.

**Configure databases for use with other search engines in PEAKS inChorus.** The databases configured here will also be used in PEAKS' inChorus to call the X!Tandem and OMSSA search engines. However, Mascot search depends on Mascot's databases only. When using these third-party software tools, note the following with care:

- **X!Tandem**: At the time of writing, X!Tandem has difficulty in searching through large databases and may crash during a search. Therefore, it is suggested that X!Tandem only be used with small databases. If used with

a large database, a taxon should be specified. For example, NCBI-nr and SwissProt databases should be used with subtaxa selected when using X!Tandem.

- **OMSSA**: At the time of writing, OMSSA cannot be used with databases that are not in NCBI-nr, or SwissProt format in a way that is available to inChorus. Also, a bug in OMSSA results in a database, stored in a folder that contains a space in its path, to be unusable. This creates problems when PEAKS creates temporary databases on your behalf. To avoid this, best practices suggest that all of the databases should be put in a folder with no spaces, such as `C:/peaksdatabases`. Note that the folder `C:/My Documents/databases` does not work as it contains a space between 'My' and 'Documents'. Using spaces in the database file name causes the same problem. For example, once the NCBI-nr database is downloaded and extracted, save the database as `ncbinr.fas`, or `ncbi_nr.fas` rather than `ncbi nr.fas`.

- **Mascot** : The database used by Mascot will have to be identical to the database configured in PEAKS in order for inChorus to parse Mascot results correctly.

# 2.5. Instrument Configuration

From the "Configuration" window, select "Instrument" from the left side menu to change the instrument configuration.



**Built-in Instruments.** The names of the built-in instruments are provided in the "Instrument List". Select an instrument to view the detailed instrument information in the "Instrument details" panel below.

## Note

The details of a built-in instrument cannot be deleted or edited.

**Create a new instrument.**

1. Click the New button and provide a name for the instrument in the "Instrument Details" panel.

2. Use the "Ion Source" drop-down menu to select the ion source that was used: MALDI/SELDI or ESI (nano-spray). This will help the PEAKS Data Refine tool determine the charge of the ions.

3. Use the "MS Precursor Scan" drop-down menu to select the type of MS scan that was performed. This selection will tell the PEAKS Data Refine tool whether the survey scan is of sufficient resolution to determine the charge and the monoisotopic peak during examination of the survey scan.

4. Use the "MS/MS Product Scan" drop-down menu to select the type of MS/MS scan that was performed. This selection will help PEAKS decide which internal parameters (for weighing fragments and amount of noise) to use during PEAKS auto *de novo* sequencing and a PEAKS DB search. Select LIT/FT if alternating high-res/low-res modes were used. This will allow the algorithm to determine the mass analyzer from the scan header.

5. Use the "Advanced Options" to specify additional parameters.

6. Select "Monoisotopic" or "Average" as "Precursor Mass Search Type". For ion-trap instruments, it is usually beneficial to allow PEAKS DB search to use an average mass.

7. Specify the values for "Parent mass error tolerance" and "Fragment mass error tolerance" in Daltons or ppm. These will appear on the PEAKS *de novo* and PEAKS DB option screens when the instrument is selected.

8. Click the Add/Update button to save the changes. The new instrument will appear in the "Instrument List" where it can be accessed when creating a new project file. To delete a customized instrument, select the appropriate instrument from the "Instrument List" and click the Delete button.