

PEAKS 7.5 User Manual

PEAKS Team

PEAKS 7.5 User Manual

PEAKS Team

Publication date 2015

Table of Contents

1. Overview	1
1. How to Use This Manual	1
2. What Is PEAKS?	1
3. What Is New in PEAKS 7.5?	3
4. Quick Walkthrough	4
4.1. Opening an Existing Project	4
4.2. PEAKS Main GUI	4
4.3. Result Summary and Filtering	5
4.4. Result Visualization	6
4.5. Creating a PEAKS Project	8
4.6. Conducting an Analysis	9
2. Installation and Activation	12
1. Package Contents	12
2. System Requirements	12
3. Installation on a Windows Computer	12
4. Activation	13
4.1. Activate PEAKS with a trial or purchased license key	14
4.2. Register to get a free 30-day trial license key	14
4.3. Use PEAKS as a viewer	15
4.4. Activate PEAKS manually	15
4.5. Re-registering PEAKS	17
4.6. Common Errors during Registration	18
5. PEAKS Performance Configuration	18
5.1. PEAKS 64-bit Version	19
5.2. PEAKS 32-bit Version	20
6. What's Next	21
3. Configuration Wizard - Configure Instruments and Public Databases	22
1. Overview	22
2. Instrument Selection	22
3. Database Selection	23
4. Download Information	24
5. Vendor Instrument Software Options	25
6. Commit the Changes	26
4. Project Wizard - Create a PEAKS Project, Load Data and Set up a Workflow	28
1. Overview	28
2. Create New Project	28
3. Data Refinement - Setting Parameters	30
4. DB Search - Setting Parameters	30
5. Quantification - Setting Parameters	31
6. Adding Data to an Existing Project	32
7. Supported Data Formats	33
8. Vendor-Specific Requirements	34
8.1. Thermo Data	34
8.2. Waters/Micromass (MassLynx) Data	34
8.3. Agilent Data	35
8.4. Applied Biosystems/Sciex Data	35
8.4.1. QSTAR or QTRAP	35
8.4.2. Converters for WIFF	36
8.4.3. ABI 4700/4800	37
8.5. Bruker Data	37
8.6. Shimadzu Data	38
8.7. Varian	38
8.8. PEAKS 6 Projects	39
8.9. PEAKS 7 Projects	39
9. Changing the Default Project Location	39

5. Data Visualization	41
1. Overview	41
2. MS View	41
3. MS/MS View	42
4. LC/MS View	43
4.1. Noise Level	44
4.2. Show/Hide Features	44
4.3. Locate M/Z and Retention Time	45
4.4. Synchronize LC/MS View between LC Runs	45
4.5. Show/Hide MS/MS Spectra	45
4.6. Intensity View	46
4.7. Export LC/MS	47
6. Adding a Sequence Database	48
1. Configuring Sequence Databases	48
2. Databases to be Used in PEAKS' inChorus Function	50
7. Data Refinement	51
1. Overview	51
2. Data Refinement Parameters	51
2.1. Saving the Parameters for Future Use	53
8. Peptide <i>De Novo</i> Sequencing	54
1. Overview	54
2. <i>De Novo</i> Sequencing Parameters	55
2.1. Error Tolerance	56
2.2. Enzyme	56
2.3. PTM	57
2.4. Other Parameters	58
2.5. Saving the Parameters for Future Use	58
3. Understanding PEAKS <i>De Novo</i> Sequencing Results	59
3.1. Summary View	59
3.2. <i>De Novo</i> Peptide View	60
3.2.1. Peptide Table	61
3.2.2. Spectrum Annotation	62
3.2.3. Ion Table	64
3.2.4. Error Map	65
3.2.5. Spectrum Alignment	65
3.2.6. Parent Scan	65
3.2.7. LC-MS Snapshot	66
3.3. LC-MS View	66
4. Filtering <i>De Novo</i> Sequencing Results	67
5. Exporting <i>De Novo</i> Results	69
6. <i>De Novo</i> Sequencing using Spectral Pairs or Triplets	69
7. <i>De Novo</i> Sequencing on Single Spectrum	71
8. Manual <i>De Novo</i> Sequencing	71
8.1. Manual <i>De Novo</i> Graphical User Interface	72
8.2. Manual <i>De Novo</i> Operations	73
9. Peptide, PTM, and Mutation Identification (PEAKS DB, PEAKS PTM, and SPIDER)	79
1. Overview	79
2. Setting PEAKS Search Parameters	80
2.1. Saving the Parameters for Future Use	83
3. Understanding PEAKS Search Results	83
3.1. Peptide and Protein Scores	84
3.2. Summary View	84
3.2.1. False Discovery Rate (FDR) Curve	84
3.2.2. PSM Score Distribution	85
3.2.3. <i>De Novo</i> Result Validation	86
3.2.4. Statistics of Data and Results	87
3.2.5. Experiment Control	88
3.3. Protein View	89

3.3.1. Protein Table	90
3.3.2. Protein Coverage	91
3.3.3. PTM Profiling	94
3.3.4. Peptide Mapping	98
3.3.5. Supporting Peptides	102
3.3.6. <i>De novo</i> Tags	103
3.4. Peptide View	103
3.4.1. Peptide Table	103
3.4.2. Peptide-Spectrum Matches	104
3.5. <i>De Novo</i> Only View	104
3.6. LC-MS View	105
4. Filtering PEAKS Search Results	105
5. Exporting PEAKS Search Results	106
6. Running PEAKS PTM and SPIDER Separately	107
6.1. Running PEAKS PTM on a PEAKS DB Result	107
6.2. Running SPIDER on a PEAKS DB or PEAKS PTM Result	108
7. Comparison of PEAKS Results	108
7.1. Comparison Result	108
7.2. Peptide Comparison	109
7.3. Protein Comparison	110
7.4. Statistical Charts	110
7.5. Exporting Comparison Result	111
10. Multi-Round Search	112
1. Multi-Round Search Overview	112
2. Setting Multi-Round Search Parameters	112
3. Understanding Multi-Round Search Results	113
4. Filtering and Exporting Multi-Round Search Results	114
11. Combining Multiple Database Search Engines with PEAKS inChorus	115
1. PEAKS inChorus Overview	115
2. Understanding PEAKS inChorus Results	117
3. Filtering PEAKS inChorus Results	119
4. Exporting inChorus Results	120
12. PEAKS Q - MS Level	121
1. Overview	121
2. Setting Parameters	121
3. Understanding the Results	122
3.1. Summary View	123
3.2. Protein View	123
3.3. Peptide View	124
3.4. LC/MS View	124
3.5. Filtering Quantification Results	124
4. Exporting Quantification Results	125
13. PEAKS Q - MS/MS Level	126
1. Overview	126
2. Setting Parameters	126
3. Understanding the Results	127
3.1. Summary View	127
3.2. Protein View	128
3.3. Peptide View	129
3.4. LC/MS View	129
3.5. Filtering Quantification Results	129
4. Exporting Quantification Results	130
14. PEAKS Q - Label Free	131
1. Overview	131
2. Setting Parameters	131
3. Understanding the LFQ Results	133
3.1. Summary View	134
3.1.1. Result Filtration	134

3.1.2. Summary Report	138
3.1.3. Exporting Quantification Results	142
3.1.4. Additional Functions	142
3.2. LC/MS View	143
3.3. Feature View	143
3.3.1. Feature Table	144
3.3.2. Feature Details	145
3.3.3. Sample Features	145
3.3.4. RT Alignment	145
3.4. Protein View	146
3.4.1. Protein Table	146
3.4.2. Coverage Tab	147
3.4.3. Features Tab	147
15. Workflow	148
1. Identification Workflow	148
2. Quantification Workflow	149
3. inChorus Workflow	150
16. Exporting Data/Reports and Printing	151
1. Exporting Data	151
2. Exporting Result From Project View	152
3. Exporting <i>De Novo</i> Results	152
3.1. Exporting Summary and Peptides	152
3.2. Exporting Images	153
4. Exporting Database Search Results	154
4.1. Exporting Summary, Proteins, and Peptides	154
4.2. Exporting Images	157
4.2.1. Exporting Protein Coverage	157
5. Exporting Quantification Results	157
5.1. Exporting Labeled Quantification Results	157
5.2. Exporting Label-Free Quantification Results	158
6. Export inChorus Result	158
17. Configuration and Preferences	159
1. PEAKS Environment Preferences	159
1.1. General Preferences	159
1.1.1. Display Options	160
1.1.2. RMI Connections	160
1.1.3. Derby Database	160
1.2. Raw File Converter Preferences	161
1.2.1. ABI (.wiff)	161
1.2.2. Bruker (.yep/baf, fid)	161
1.2.3. Shimadzu AXIMA (.run)	161
1.2.4. Varian (.xms)	161
1.2.5. Waters (.raw)	162
1.3. Search Engine Preferences	162
1.3.1. Mascot Settings	162
1.3.2. X!Tandem Settings	162
1.3.3. OMSSA Settings	163
1.4. Spectrum Annotation Preferences	163
2. PEAKS Configuration	164
2.1. Enzyme Configuration	164
2.2. PTM Configuration	165
2.3. Labeled Q Method Configuration	167
2.4. Database Configuration	168
2.5. Instrument Configuration	170

Chapter 1. Overview

Welcome to PEAKS.



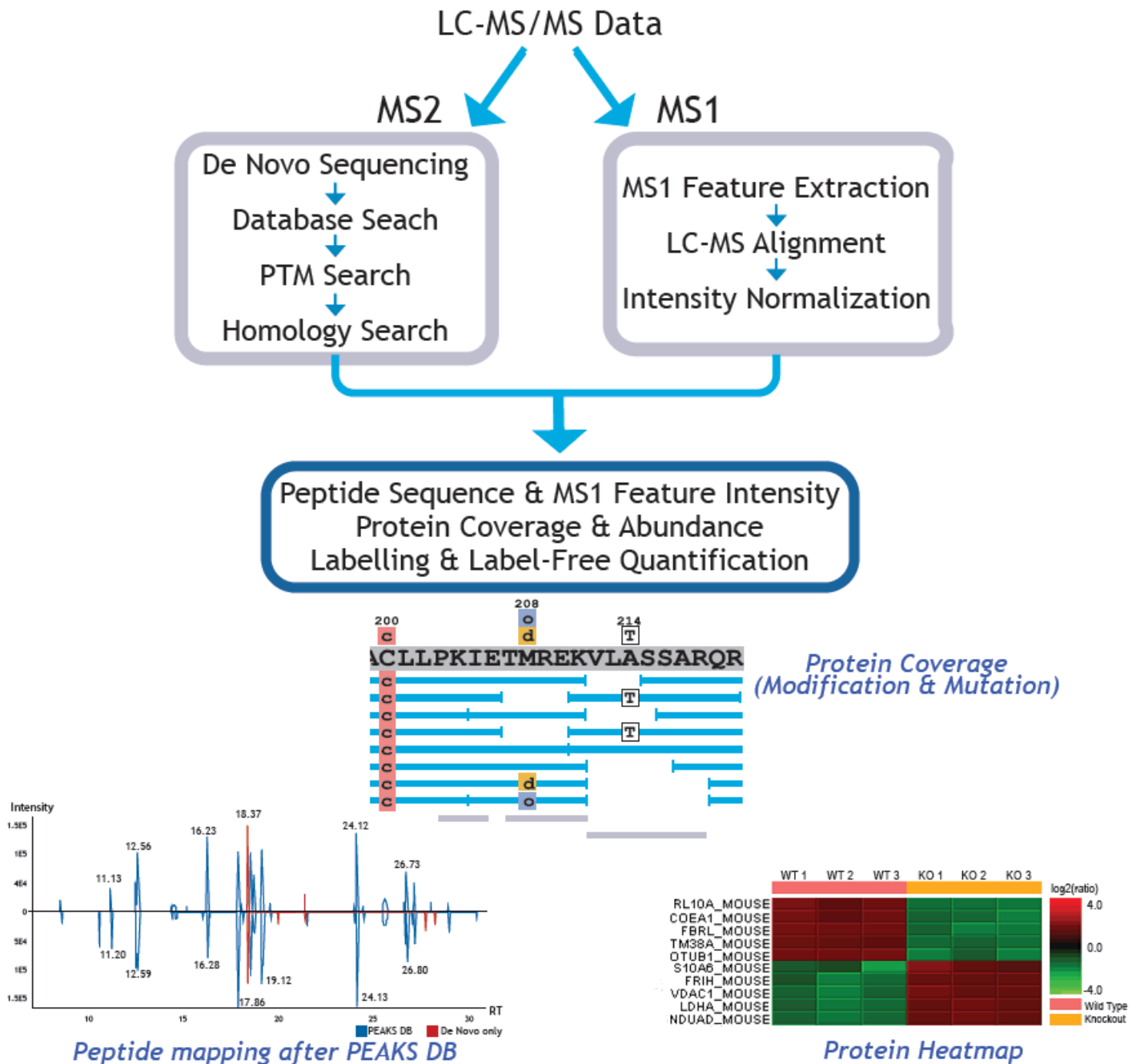
1. How to Use This Manual

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 inline help. The electronic (and most up-to-date) version of this manual can be found at <http://bioinfor.com/doc/peaks75/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 of data conversion; peptide and protein identification; PTM and mutation characterization; as well as result validation, visualization and reporting.



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 search, characterizing unspecific PTMs, and detection of peptide mutations. The combined use of 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 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 the PTMs and mutations on the protein sequence are highlighted. The interactive graphical user interface (GUI) allows 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 PEAKS algorithms has been demonstrated by publications and third-party evaluations. By combining four complementary algorithms together, the sensitivity is further improved.

- **Multiple-enzyme Project**

To maximize a protein's sequence coverage, it is advantageous to use different proteolysis 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 combine all samples' 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 TIC or selected protein(s).

- **Multiple Instrument Vendor Support**

PEAKS supports most major instruments' raw file formats, which saves the trouble for file format conversion. More importantly, the algorithms have been tuned for different instrument type, 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; whereas other text formats such as .csv make it easy to view the result in spreadsheets and post-analyze with 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 the 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 the peptide, as well as each engine's identification score.

3. What Is New in PEAKS 7.5?

Here is a list of the noteworthy new features in PEAKS 7.5.

- **PTM Profiling**

A tool to compare and quantify detected modifications sites on proteins. Determines the protein's PTM changes by comparing the “area-under-the-curve” from the LC-MS of the modified peptide features to the associated, unmodified (native) peptide features.

- **Peptide Mapping**

A powerful comparison tool, which provides an intensity map of all the peptides associated with any particular protein ID. The maps can be analyzed in terms of the retention time vs intensity, or the peptide mass vs intensity. PEAKS allows users to compare peptide maps of up to 3 different data files.

- **Multi-Round Search**

A tool to launch multiple rounds of a database search using *de novo* only spectra. *De novo* only spectra are high quality spectra with high scoring *de novo* results that could not be identified in the initial database search.

- **Feature Intensity**

One button function to fully analyze and quantitate peptides across all samples.

- **AScore**

Modification site localization score.

- **Better community support**

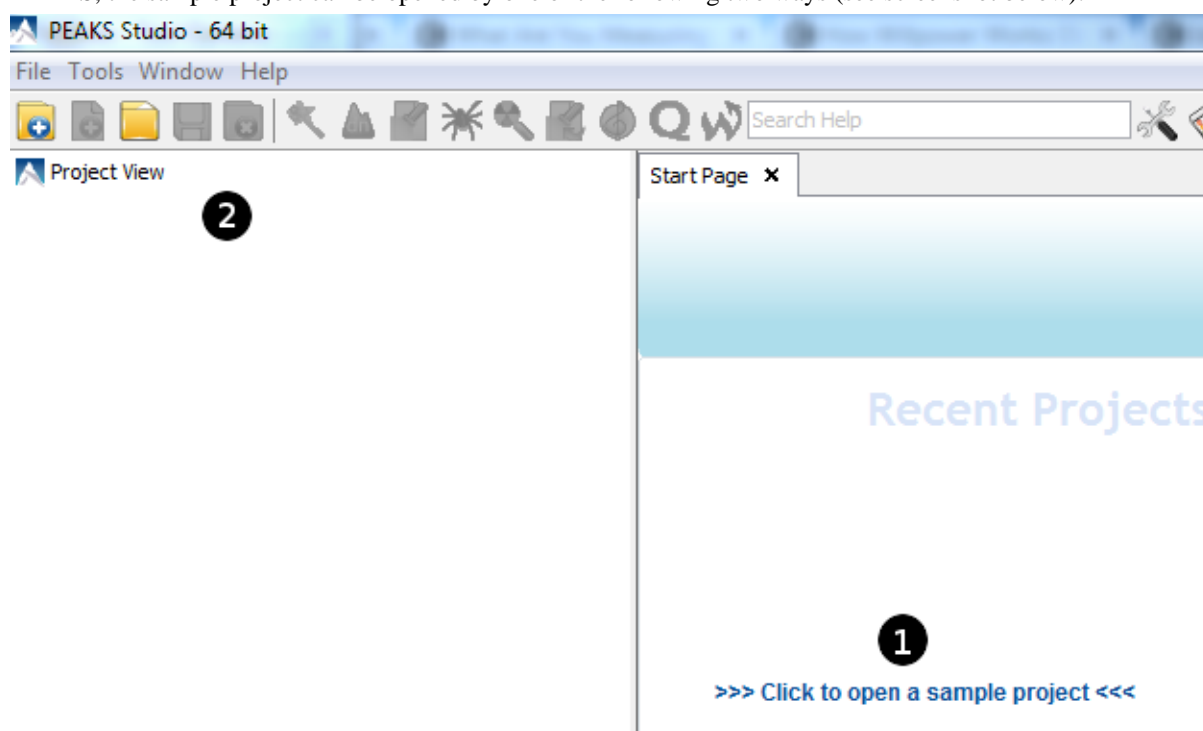
Better support for Waters raw files.


4. Quick Walkthrough

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

4.1. Opening an Existing Project

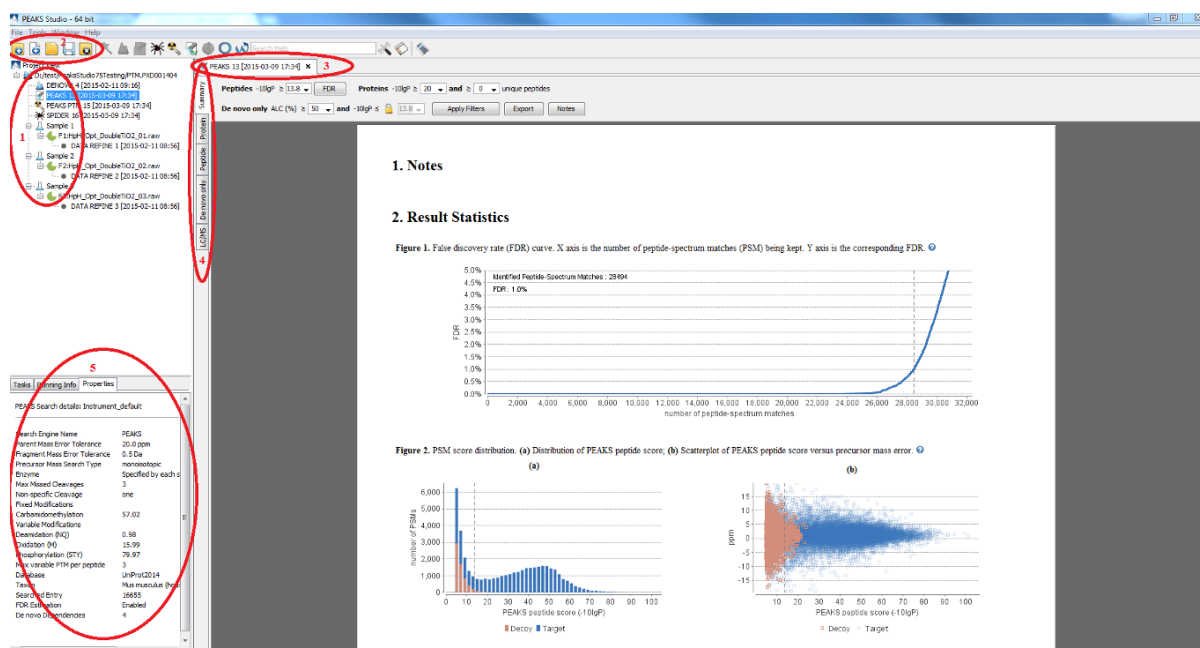
The installation instructions for PEAKS can be found at Chapter 2, *Installation and Activation*. After launching PEAKS, the sample project can be opened by one of the following two ways (see screenshot below):



1. If this is a fresh installation, click the "Sample Project" in the "Recent Projects" list of the Start Page.
2. Click the open project button , and browse to the directory where PEAKS 7.5 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):



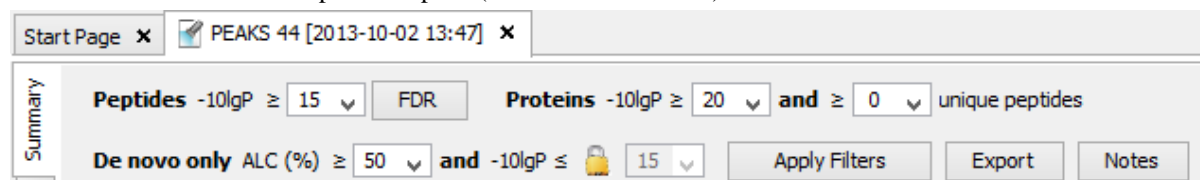
1. The project view shows all the opened projects. Each project may include multiple samples and each sample may include multiple fractions (LC-MS runs). The analysis results are also displayed as result nodes under the project.
2. The menu and toolbar. Selecting a node (project, sample, fraction, or result) in an opened project will highlight the analysis tool icons 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. The summary view is also the central place to filter and export the 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 opening a result node by double clicking it, i.e. the PEAKS DB node in the sample project, the summary view is shown by default. The summary view mainly provides three functions:

1. Specify score thresholds to filter the results.
2. Examine the result statistics.
3. Export results.

The top region of the summary view is a control pane and the bottom region is a statistics report page. The result filters are controlled at the top control pane (see screenshot below):



- The peptide identification is filtered by the peptide-spectrum match's $-10\lg P$ score. Or, by the desired FDR (false-discovery rate) specified by clicking the FDR button.
- The protein identification is filtered by the protein's $-10\lg P$ score and the number of unique peptides the protein contains.

- 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 better than or equal to the specified threshold. Meanwhile, the spectrum's best database search result's score should be no greater than the specified -10lgP threshold.

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

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. The result can be exported by clicking the Export button. The Notes button allows the user to add a text note about the project that will be displayed in the result summary report.

After applying filters, the statistics report page at the bottom of the summary view will be updated accordingly. We only explain two statistical charts here (see screenshot below).

Figure 2. PSM score distribution. (a) Distribution of PEAKS peptide score; (b) Scatterplot of PEAKS peptide score versus precursor mass error. [?](#)

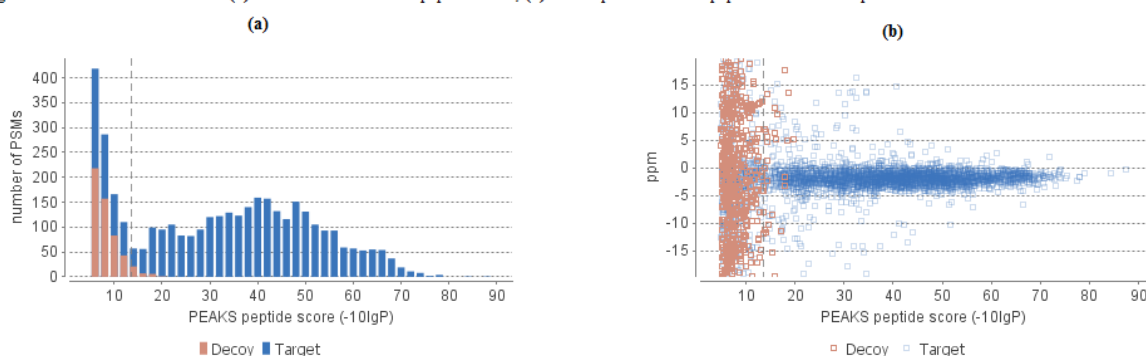


Figure 2 (a) shows the PSM score distribution in a stacked histogram. If the search result and the peptide -10lgP score threshold is of high confidence, then very few decoy matches (brown) in the high score region should be observed. Additionally, if the FDR estimation method (decoy fusion) worked properly,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

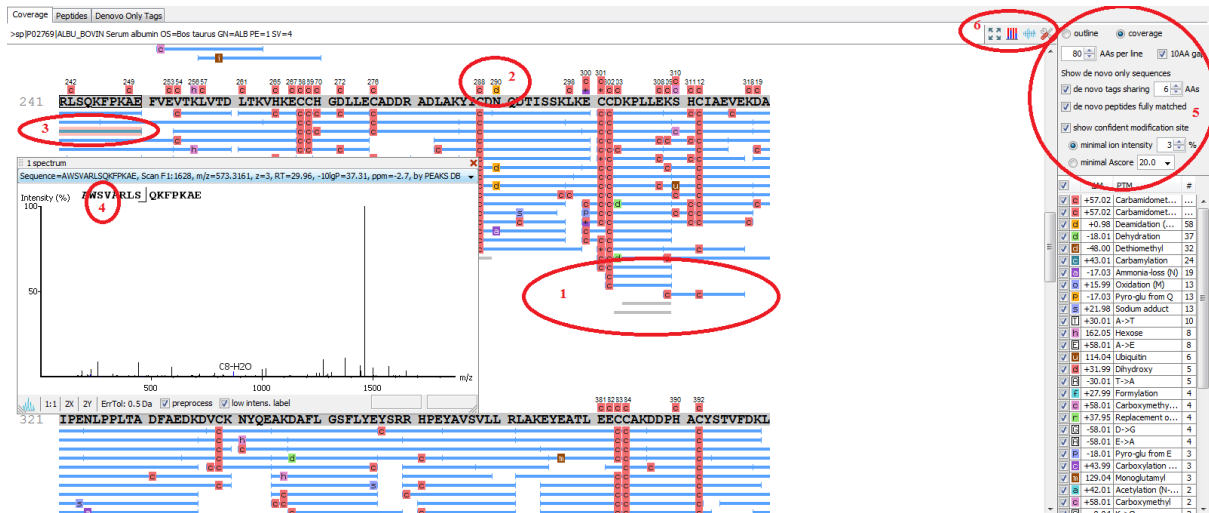
Figure 2 (b) plots the precursor mass error in ppm vs. -10lgP peptide score for all the PSMs. This figure is the most useful for high resolution instruments. Generally, the high-scoring points should be centered around the mass error 0. Notice that the data points start to scatter to larger mass errors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

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 passing 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 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 will be displayed at the bottom of the Protein view. The protein coverage view maps all peptide identifications passing the filter of the selected protein onto the protein sequence. It enables the effortless examination of every PTM and mutation on each amino acid. Some most commonly used operations on the protein coverage view are listed (see screenshot below):



1. Each blue bar indicates an identified peptide sequence passing the filter. A 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 either the two fragment ions at both sides of the modified residue have relative intensity higher than the user-specified threshold in the display option or the Ascore of this PTM is higher than the user-specified threshold in the display option. (see item 5).

3. Click a peptide to show the spectrum annotation.
4. Mouse over an amino acid to show the supporting fragment ion peaks.
5. Options to control the coverage view display.

- The "coverage/outline" choice turns on/off the peptide bars.
- 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 in the protein.
- The "show confident modification site" contains two types of filters to define confident PTMs to be displayed on top of the protein sequence. The "minimal ion intensity" specifies the minimum fragment ion relative intensity in one of the MS/MS spectra before a PTM location is regarded as confident; The "minimal Ascore" specifies the minimal Ascore in one of the MS/MS spectra before a PTM location is regarded as confident.
- The checkboxes in the PTM list specifies which PTMs are "interesting". Click the color boxes to change a color. Double-click a PTM name to see the PTM details.

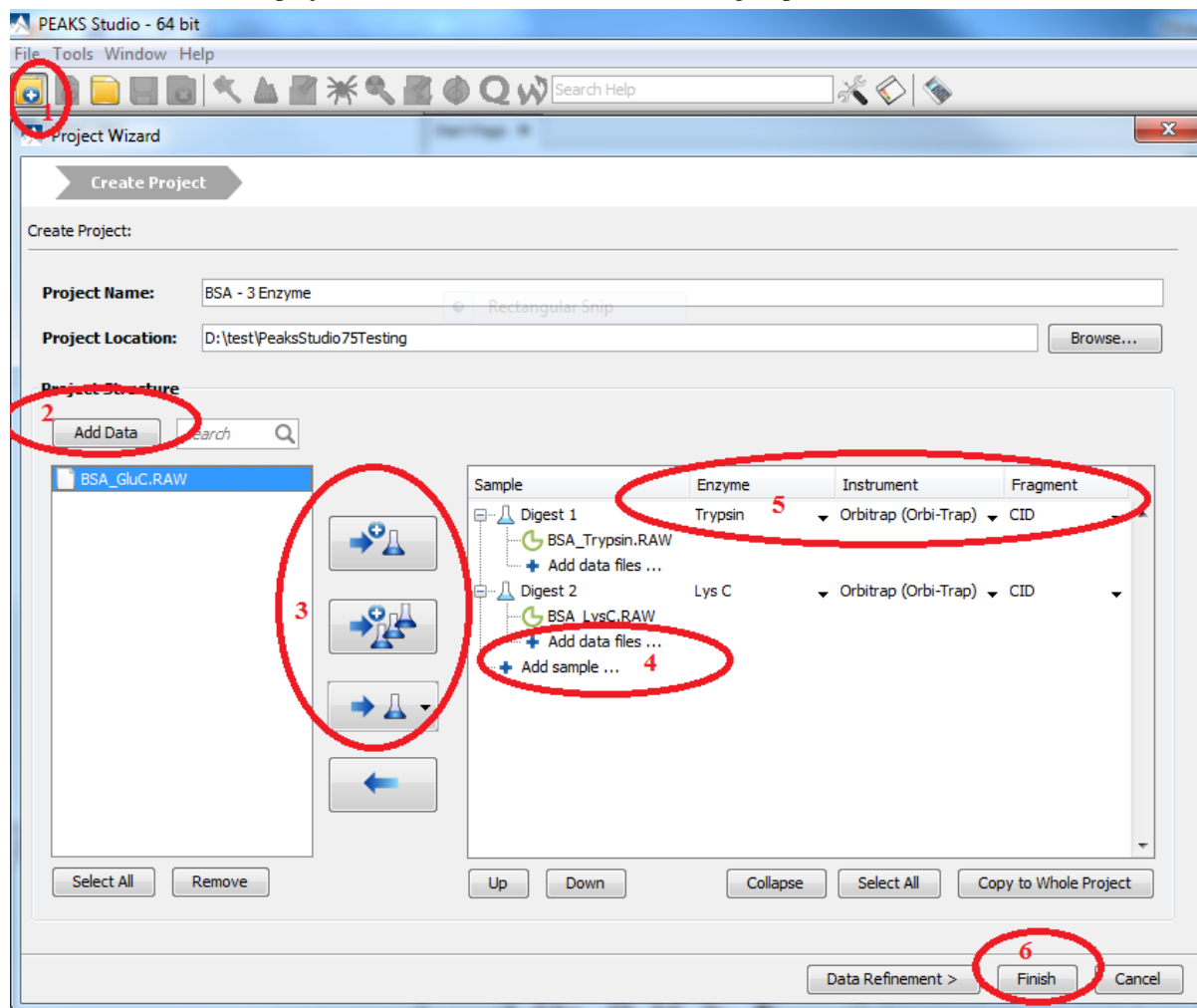
6. The full screen, PTM profile, peptide mapping, and toolbox buttons.





Full screen provides a larger view of the coverage. The toolbox provides common tools, such as exporting the coverage pane as a high-resolution image file.

PTM profile is a tool to compare and quantify detected modification sites on proteins. Peptide mapping provides an intensity map of all the peptides associated with any particular protein ID.

4.5. Creating a PEAKS Project

To create a new PEAKS project from raw data files, do 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 the files you wish to load and click Open. All 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 place each file in an individual sample using .
4. Click the **+ Add sample ...** or the **+ Add data files ...** 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 proteolysis enzyme. Using multiple enzymes to analyze the same proteins can produce overlapping peptides and therefore increase the protein coverage.

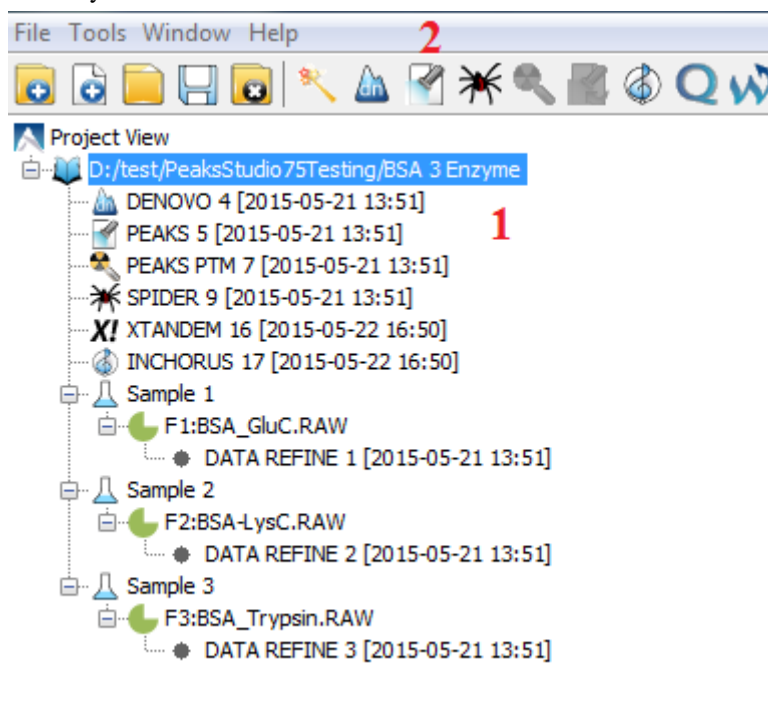
Note

To apply the same sample details to the whole project, select the sample with 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, sample, or 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):

The screenshot shows the 'PEAKS Search' dialog box with the following sections and annotations:

- Error Tolerance:** Precursor mass: 10.0 ppm using monoisotopic mass. Fragment ion: 0.5 Da.
- Enzyme:**
 - Annotation 1: 'Specified by each sample' dropdown menu.
 - Allow non-specific cleavage at: one end of the peptide.
 - Maximum missed cleavages per peptide: 3.
- PTM:**
 - Annotation 2: List of PTMs: Carbamidomethylation (F), Oxidation (M), Deamidation (NQ).
 - Buttons: Set PTM, Remove, Switch type.
 - Maximum allowed variable PTM per peptide: 3.
- Database:**
 - Annotation 3: 'Select database' radio button.
 - Database: [UniProt_SwissProt] View.
 - Paste sequence Taxa: all species Set/View taxa...
- De Novo Tag Options:**
 - Annotation 4: 'Available de novo tags: de novo with current parameter' dropdown.
- General Options:**
 - Annotation 5: 'Estimate FDR with decoy-fusion.' checkbox (checked).
 - Annotation 6: 'Find unspecified PTMs and common mutations with PEAKS PTM' checkbox (unchecked).
 - 'Find more mutations with SPIDER' checkbox (unchecked).
 - 'Advanced Setting' button.

At the bottom are 'OK', 'Cancel', and 'Help' buttons.

1. If the proteolysis 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.
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 PEAKS DB database search.

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 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 ($-10\lg P < 30$). SPIDER will construct new peptide sequences by altering the 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. It makes no difference whether SPIDER is 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 7.5 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 7.5 supports Windows Operation Systems (Windows XP, Windows 7, and Windows 8). PEAKS program files will only use ~200MB of disk space. Extra disk space is required 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.

For the default 64-bit version of PEAKS 7.5, the recommended and ideal requirements are :

- **Recommended:** Intel Core processor and 8GB of RAM.
- **Ideal:** Intel Core i7 or Xeon processor and 16GB of RAM or more (or 2GB per core).

For the 32-bit version of PEAKS 7.5, 4GB of RAM is required.

3. Installation on a Windows Computer

Important

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

Important

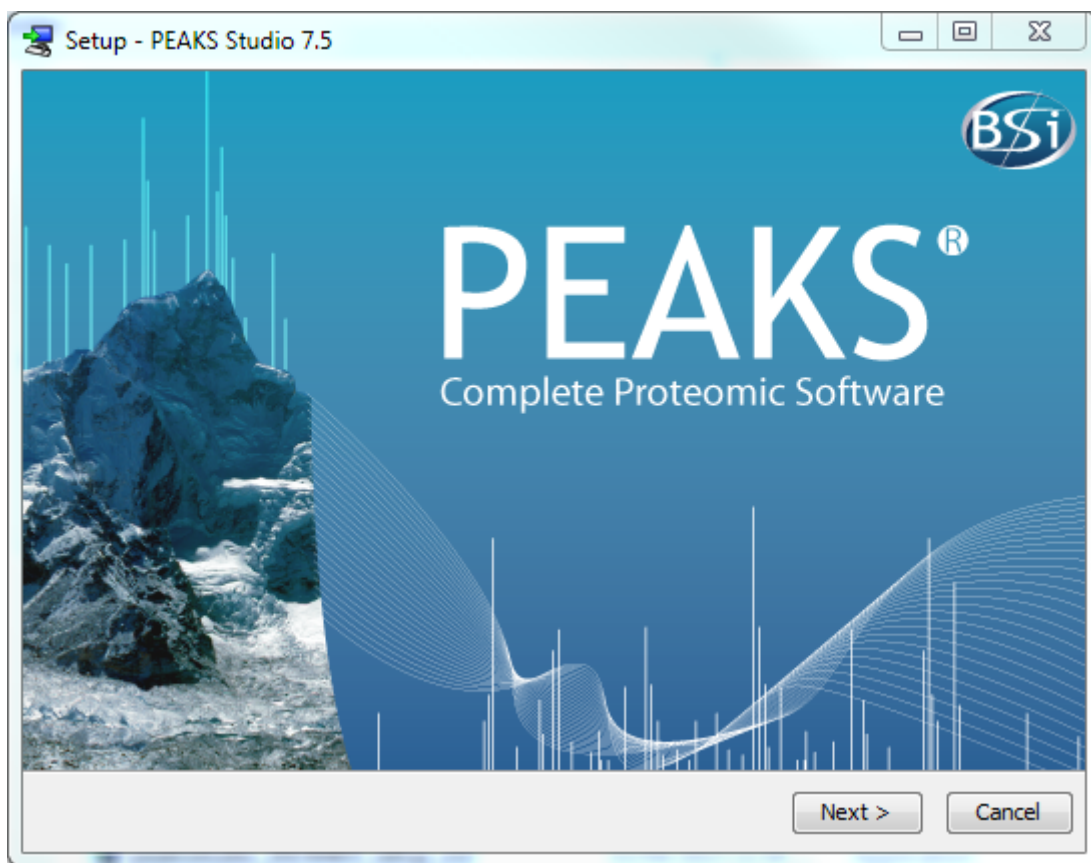
Avoid installing PEAKS in any directory that contains white space, for example the `Program Files`, directory as some features may not function correctly in such situations. Please make sure that the user account has full access permissions (read/write/execute) on the selected directory.

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 Chapter 4, Section 8, "Vendor-Specific Requirements" for the vendor-specific requirements for raw data loading.

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.

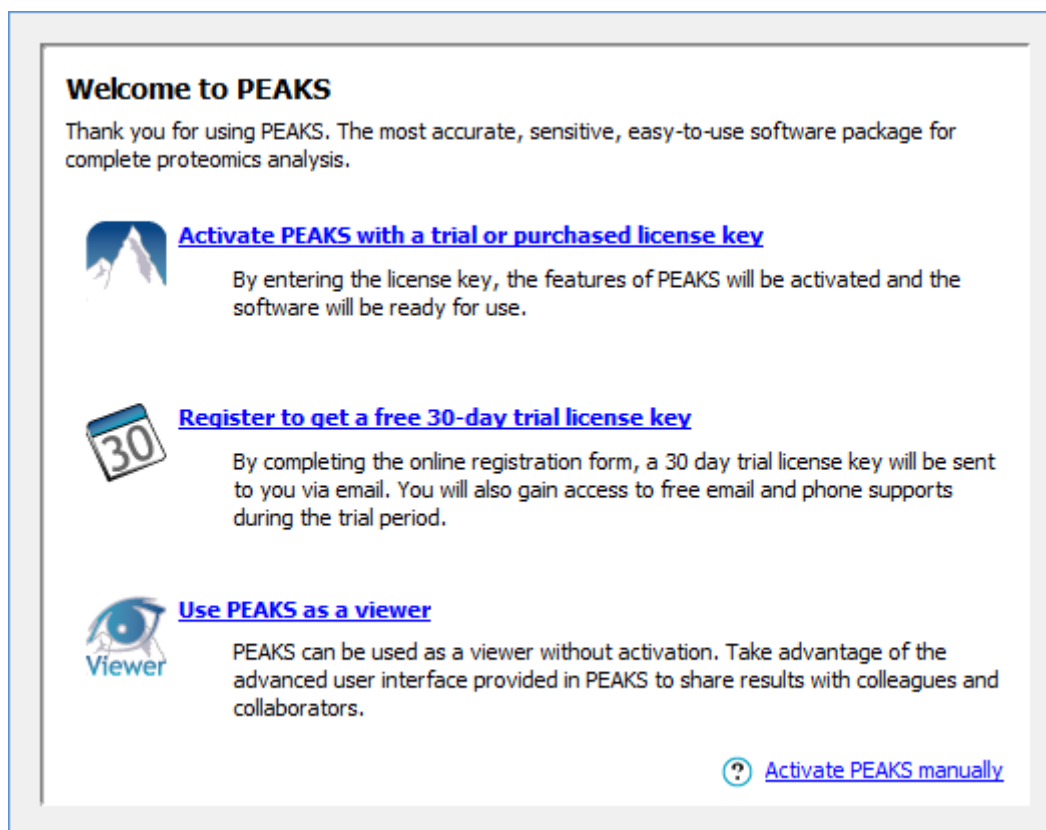
3. The installation window should automatically appear after the DVD is inserted. If it does not, find the DVD drive and open it to access the disc. Double-click on `PEAKS_Studio_Installation.exe`.
4. A menu screen will appear. Select the top item “PEAKS Installer”. The installation utility will launch 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.

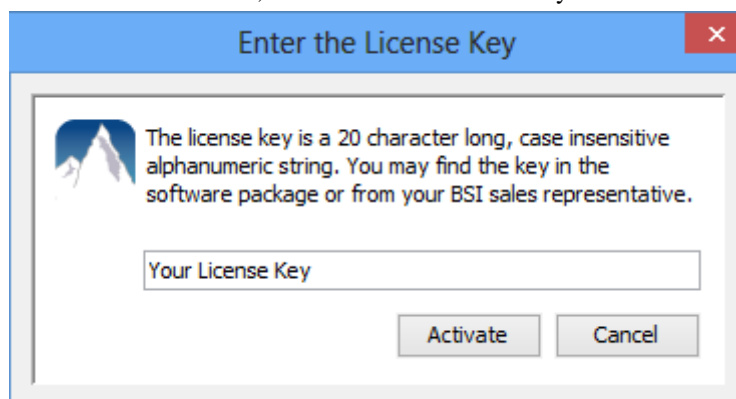


Important

Keep the license key safe. After a computer hardware upgrade, it might be necessary 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 30-day 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 repetitively will not extend the trial period automatically. Please contact <support@bioinfor.com> to discuss a trial license extension.

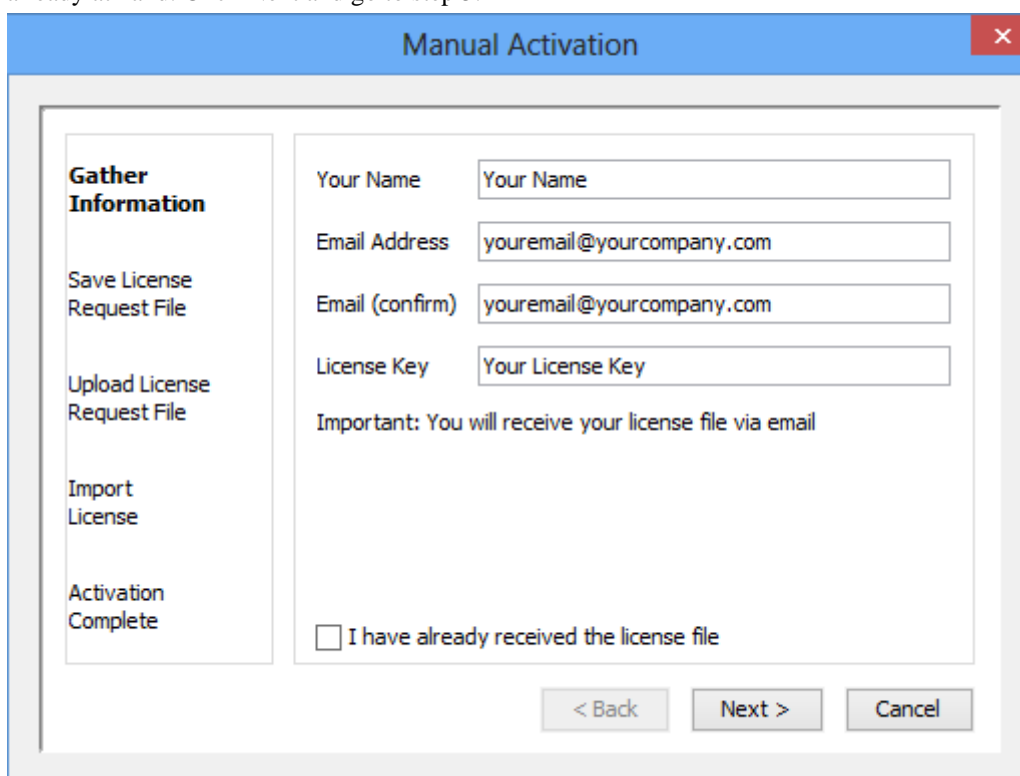
4.3. Use PEAKS as a viewer

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

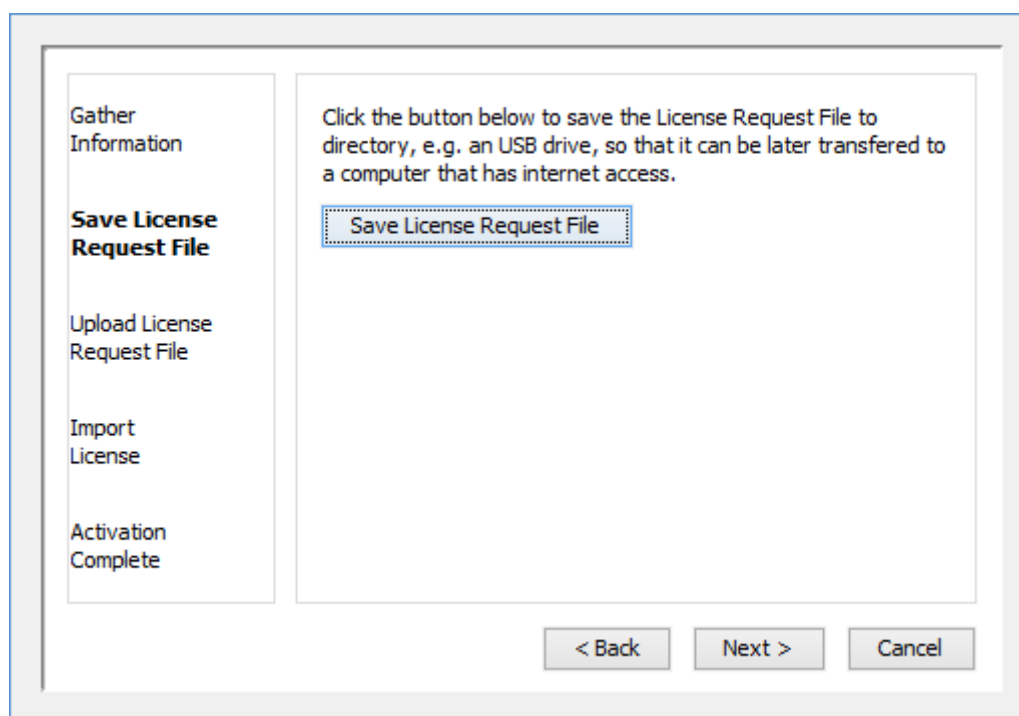
In certain situations, when the computer does not have an Internet connection (or is behind a firewall that blocks the activation), the activation process requires the assistance of another computer with an Internet connection (or outside the firewall). Manual activation can be accessed via the link 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 checked when the license file from BSI is already at hand. Click Next and go to step 5.

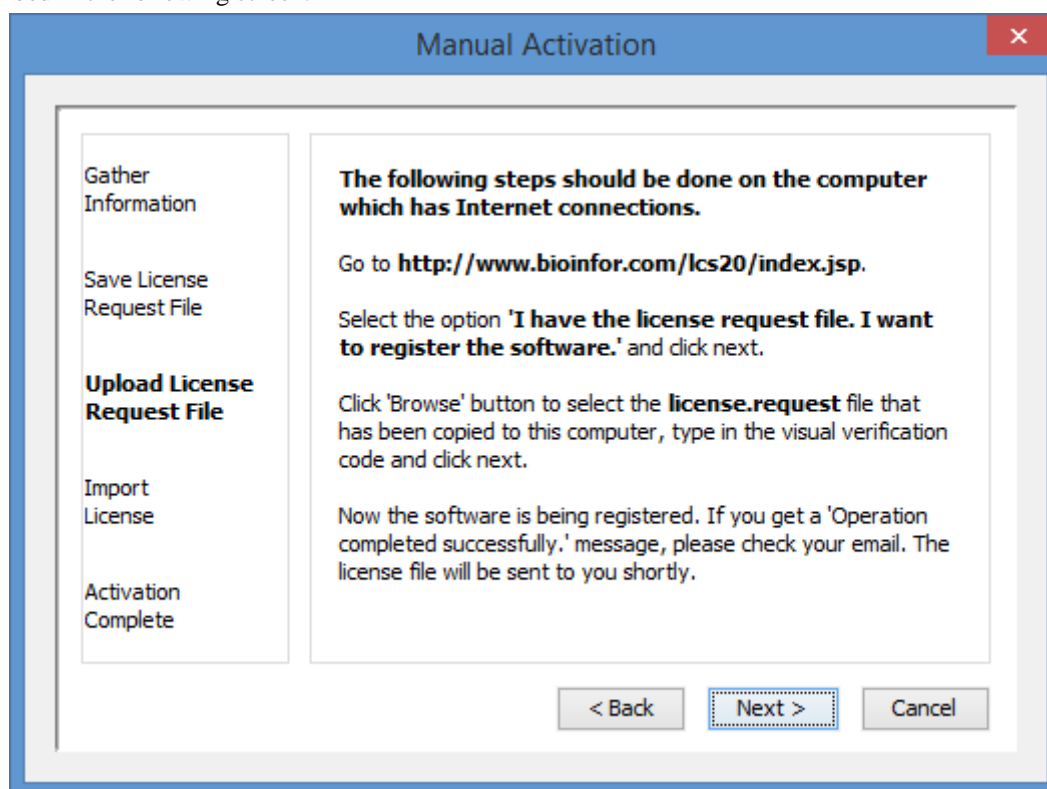


The screenshot shows a window titled "Manual Activation" with a blue header bar and a red close button. On the left is a sidebar with a tree view containing: "Gather Information" (selected), "Save License Request File", "Upload License Request File", "Import License", and "Activation Complete". The main area contains a form with the following fields: "Your Name" (text box with placeholder "Your Name"), "Email Address" (text box with placeholder "youremail@yourcompany.com"), "Email (confirm)" (text box with placeholder "youremail@yourcompany.com"), and "License Key" (text box with placeholder "Your License Key"). Below these fields is the text "Important: You will receive your license file via email". At the bottom of the form is a checkbox labeled "I have already received the license file". At the bottom of the window are three buttons: "< Back", "Next >", and "Cancel".

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.



The screenshot shows a web browser window with the address bar displaying `www.bioinfor.com/lcs20/index.jsp`. The page title is "BSI Product Licensing Center". The main heading is "BSI Products Online Registration". Below this, there is a "Request file:" section with a "Browse..." button and the text "No file selected.". The "Visual verification:" section displays a CAPTCHA image with the characters "B D R 9 9 A" and a "request another image" link. A "Register" button is located at the bottom left of the form.

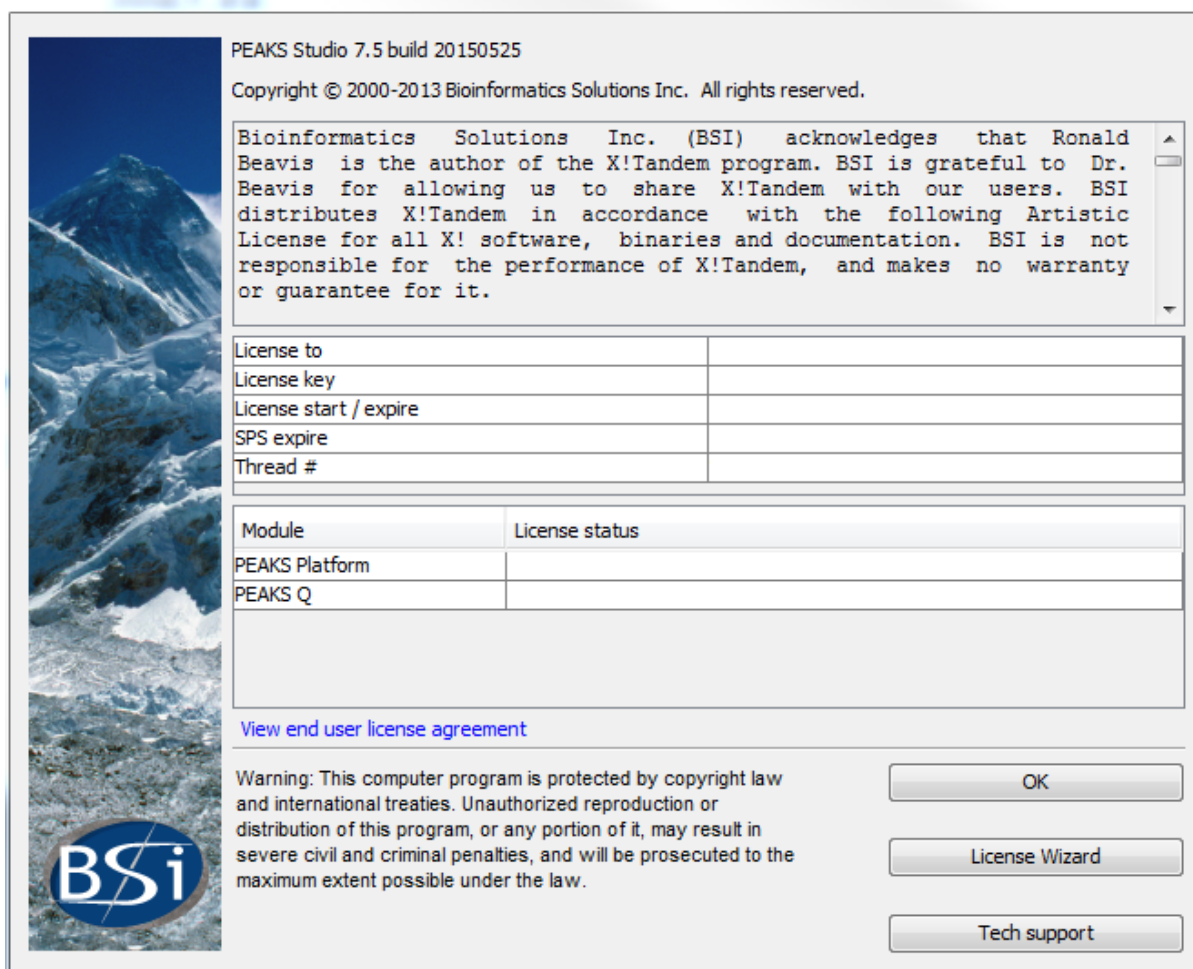
4. If the license request is sent successfully, an automated BSI service will generate the license file (`license.lcs`) and send an email (from `<support@bioinfor.com>`) to the email address provided to the License Wizard. Either save the attached license file or copy the content between '====>' and '<====' 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.

The screenshot shows the "Import License" step of the PEAKS license wizard. On the left, a vertical list of steps includes "Gather Information", "Save License Request File", "Upload License Request File", "Import License" (which is currently selected and bolded), and "Activation Complete". The main area has two radio button options: "Paste the license" (which is selected) and "Import the license file (the email attachment)". The "Paste the license" option has a "Paste from Clipboard" button next to it. The "Import the license file" option has a text input field and a "Browse" button. At the bottom, there are three buttons: "< Back", "Next >" (which is highlighted with a dashed border), and "Cancel".

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:

About BSI PEAKS Studio



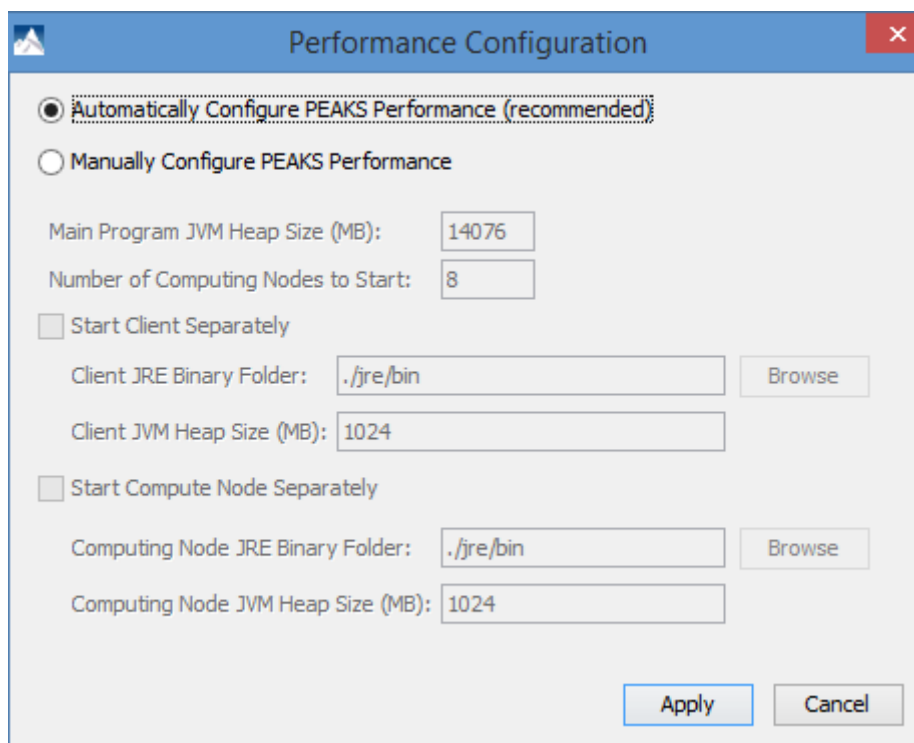
Click the License Wizard button to continue then follow the instructions in Section 4.1, “Activate PEAKS with a trial or purchased license key” or Section 4.4, “Activate PEAKS manually” to re-register PEAKS.

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. 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 de-select 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, if **X** MB of memory are assigned to the main program, the number of computing nodes **N** is the largest integer less than $X / 2000$. In the figure below, the system has 16000MB of RAM, therefore $16000 * 80\% = 12800\text{MB}$ of memory are configured to the main program JVM. $12800 / 2000 = 6.4$, therefore 6 computing nodes are configured.

☐ Automatically Configure PEAKS Performance (recommended)
☒ Manually Configure PEAKS Performance

Main Program JVM Heap Size (MB):
 Number of Computing Nodes to Start:

☐ Start Client Separately
 Client JRE Binary Folder:
 Client JVM Heap Size (MB):

☐ Start Compute Node Separately
 Computing Node JRE Binary Folder:
 Computing Node JVM Heap Size (MB):

5.2. PEAKS 32-bit Version

The PEAKS 32-bit version is produced only for compatibility on older 32-bit operating systems. It is strongly recommended to use the 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 RAM looks like the following.

☐ Automatically Configure PEAKS Performance (recommended)
☒ Manually Configure PEAKS Performance

Main Program JVM Heap Size (MB):
 Number of Computing Nodes to Start:

☐ Start Client Separately
 Client JRE Binary Folder:
 Client JVM Heap Size (MB):

☒ Start Compute Node Separately
 Computing Node JRE Binary Folder:
 Computing Node JVM Heap Size (MB):

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, $1024 \times 1 + 1024 = 2048\text{MB}$. In some systems, PEAKS may not start with 1024MB assigned to the main program due to 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 Chapter 4, 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*.

If you are eager to try PEAKS now, set these two configuration steps aside for a while, and try out the 15-minute walkthrough to get familiar with PEAKS' GUI and basic operations. See Chapter 1, Section 4, “Quick Walkthrough”.

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 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 to select, download, and configure the appropriate databases from a list of public sequence databases 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 install the MSFileReader v3.0.31 SP3 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

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”.

Download Information

Downloads

Name	Size	Progress	Action
Thermo MSFileReader	30.3 MB	Installed	Reinstall
ABI mzWiff	281.9 kB	Downloaded	Install
IPI	25.6 MB	Downloading (63%)	Cancel

Selected Instrument Software

Name	Status
ABI Analyst	Display Information
AB SCIEX MS Data Converter	Display Information

<< Back
Next >>
Cancel
Help

Downloads - The “Downloads” table shows all the downloadable requests, their status and 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 popup dialog.

Click Back to go back to instrument selection or the database selection panel 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 Options

Bruker (.yeap / .baf / fid)
No further configuration of PEAKS is needed.
But you must ensure that Bruker CompassXtract Runtime is installed on this computer.
[Download CompassXtract Runtime](#) from Bruker website.

Raw File Converter Options

Bruker file reader will export ☒ Raw data ☐ Line spectra

Bruker .fid file may contain several files, do you want to merge them? ☐ yes ☒ no

Shimadzu AXIMA (.run)
Shimadzu MS data (.run) can be converted to mzXML format using Shimadzu Launchpad.
The mzXML data can be loaded into PEAKS readily.
Please contact your Shimadzu representative to obtain Shimadzu Launchpad.

Shimadzu run2xml.exe File Location

C:\Program Files (x86)\Programs\run2xml.exe

Browse

Varian (.xms)
Varian MS data (.xms or .sms) can be converted to PKL format using VarianMS tool.
The PKL data can be loaded into PEAKS readily.
Please contact your Varian representative to obtain VarianMS tool.

Default xmlrai.exe File Location

Browse

<< 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 finalize 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.

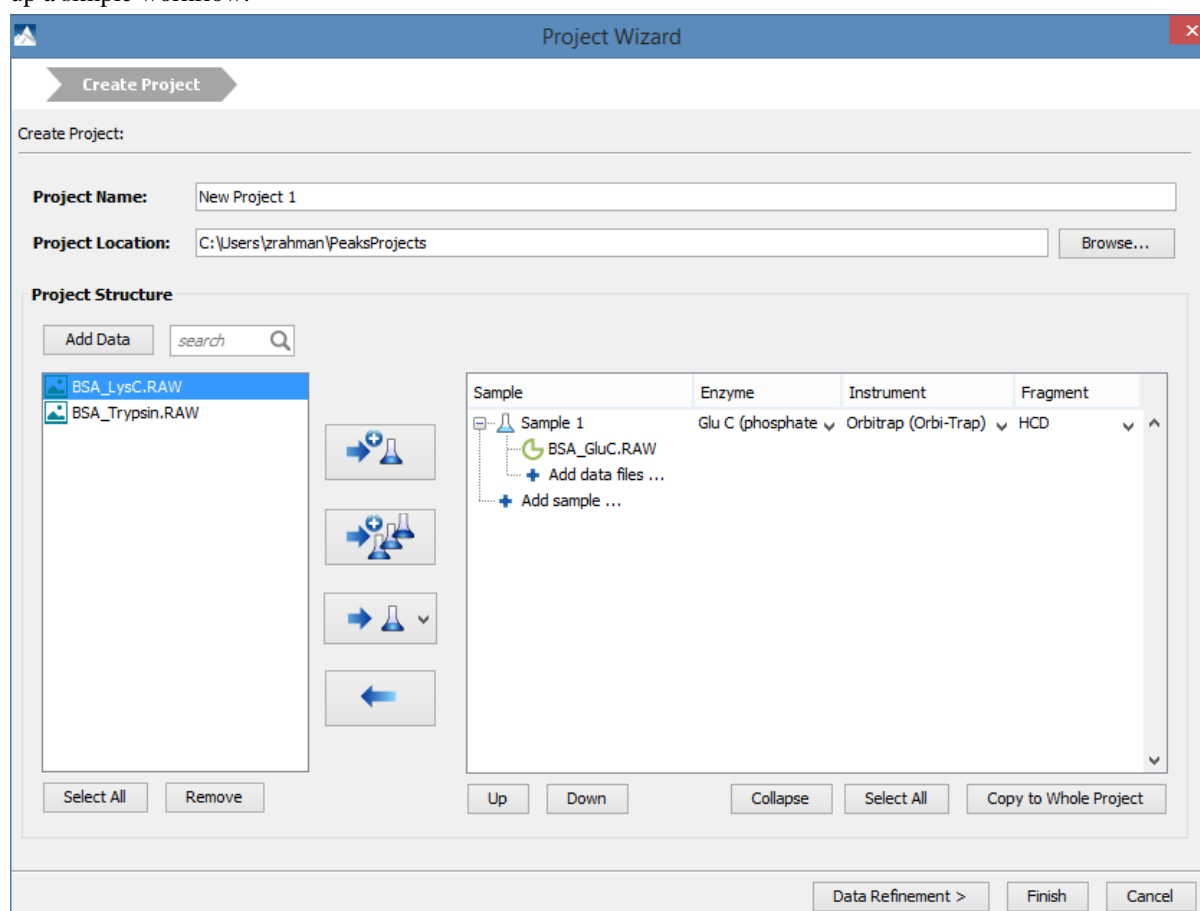
Chapter 4. Project Wizard - Create a PEAKS Project, Load Data and Set up a Workflow

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 with 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 button  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.



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 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 appropriate samples in the project structure. There are several options to place the selected data files into the project structure. All of the selected files can be



placed in a new sample (), in an existing sample (), or they can be placed in individual samples for each file ().

5. To add a sample to the project structure, click on the  Add sample ... button. Select the “Instrument” type, “Fragmentation” method and “Enzyme” name 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 mode if more than one fragmentation mode is used and the information about the fragmentation mode of each spectrum 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 you wish to load 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 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 set up parameters for data refinement step. See Section 3, “Data Refinement - Setting Parameters” for more details on data refinement parameters.

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

PEAKS supports a variety of instrument vendors' raw data formats. A list of supported formats can be found in 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. 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 choose the Close Current Project command

from the file menu or use the Close Project icon  from the tool bar or right-click and choose the Close Project command from the pop-up menu. It is recommended to close the 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 Chapter 7, Section 2, “Data Refinement Parameters” for more details on data refinement parameters.

The screenshot shows the 'Data Refinement' panel of the PEAKS Project Wizard. At the top, there are two tabs: 'Create Project' and 'Data Refinement', with 'Data Refinement' being the active tab. Below the tabs, the panel is titled 'Data Refinement' and has a 'Predefined parameters' dropdown menu set to 'default'. The panel contains three main sections: 'Merge Scans', 'Correct Precursor', and 'Filter Scans'. The 'Merge Scans' section has a checkbox that is unchecked. The 'Correct Precursor' section has a checkbox that is checked, and it contains two radio buttons: 'Mass only (recommended)' (selected) and 'Mass and Charge states'. Below these are 'Min charge' and 'Max charge' spinners set to 1 and 3 respectively. The 'Filter Scans' section has a checkbox that is unchecked and contains three sub-sections: 'Only keep scans satisfying:', 'Precursor mass between' (with two empty text boxes and 'Da' as a unit), 'Retention time between' (with two empty text boxes and 'min' as a unit), and 'Quality value greater than' (with an empty text box and '(suggest 0.65)' as a note). At the bottom of the panel, there are four buttons: '< Back', 'Identification >' (highlighted with a dashed border), 'Finish', and 'Cancel'.

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. Alternately, click the Identification button to proceed to set up parameters for database search.

4. DB Search - Setting Parameters

To run *de novo* sequencing and database search on the project immediately after the project creation and data refinement, click the Identification button on the Project Wizard. See Chapter 9, Section 2, “Setting PEAKS Search Parameters” for more details on database search parameters.

Project Wizard - Create a PEAKS Project, Load Data and Set up a Workflow

PEAKS Search Predefined parameters

Error Tolerance
Precursor mass: 10.0 ppm using monoisotopic mass Fragment ion: 0.5 Da

Enzyme
Specified by each sample New
Allow non-specific cleavage at one end of the peptide.
Maximum missed cleavages per peptide: 3

PTM
☒ Carbamidomethylation
☒ Oxidation (M)
☒ Deamidation (NQ)
Set PTM
Remove
Switch type
Maximum allowed variable PTM per peptide 3

Database
☒ Select database Database: [UniProt_SwissProt] View
☐ Paste sequence Taxa: Homo sapiens (human) Set/View taxa...

General Options
☒ Estimate FDR with decoy-fusion. ?
☐ Find unspecified PTMs and common mutations with PEAKS PTM Advanced Setting
☐ Find more mutations with SPIDER

Skip Identification < Back Quantification > Finish Cancel

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 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 with PEAKS. Alternatively, click the Quantification button to set up parameters for quantification.



5. Quantification - Setting Parameters

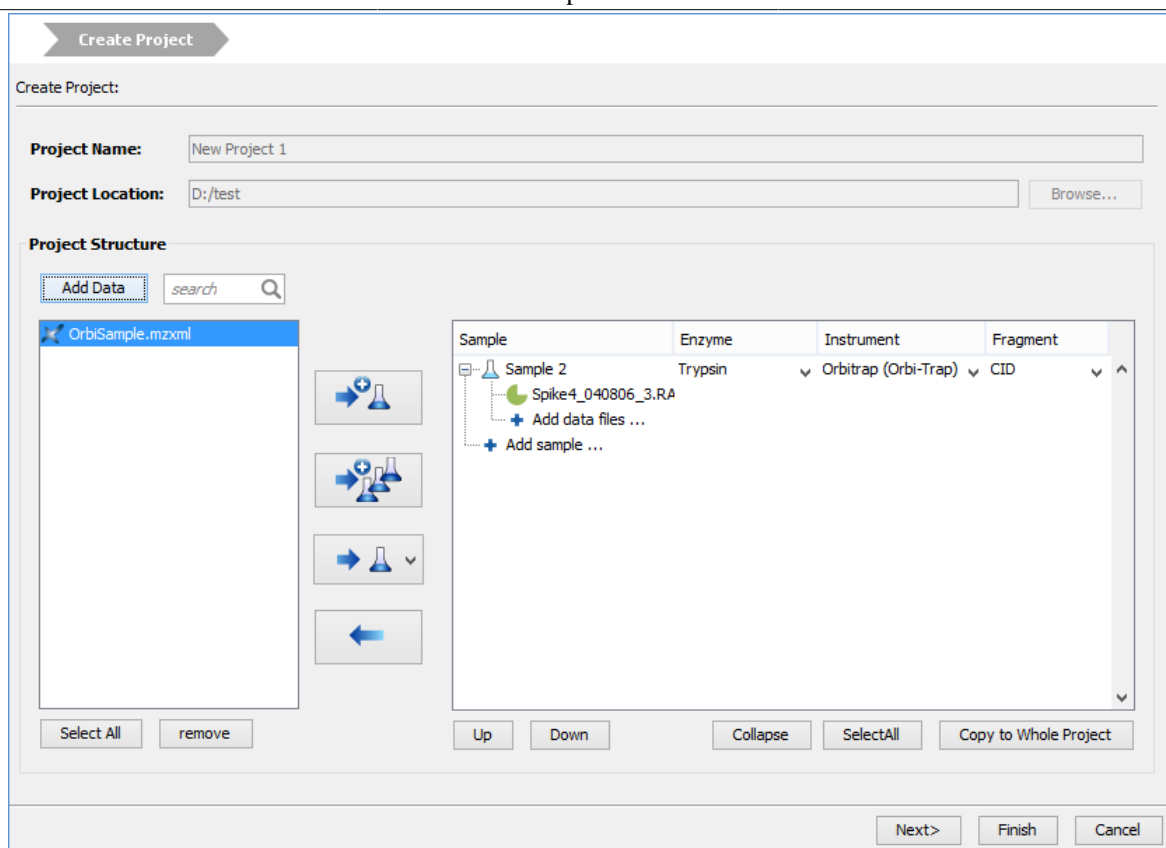
To run quantification tools on the project immediately after the project creation, data refinement, and identification, click the Quantification button to open the “Quantification” panel. The optional PEAKS Q module of PEAKS Studio supports Label Free quantification, Reporter ion quantification with isotope labels at MS/MS level, and Precursor ion quantification with isotope labels at the MS-level. See Chapter 14, Section 2, “Setting Parameters”, Chapter 13, Section 2, “Setting Parameters”, and Chapter 12, Section 2, “Setting Parameters” for more details on quantification parameters.




Project Wizard - Create a PEAKS Project, Load Data and Set up a Workflow

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 quantification on the data at 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 File | Open Project command or select from the File | Open Recent Project list, or add the  button from the toolbar.
2. To add data to an open project, choose the project from the “Project View” panel and select the Add Data command from the file menu or use the Add data button  from the toolbar. The “Project Wizard” window will open.



3. Use the Add Data button to browse to the location of the data files, select the files you wish to load and click Open. All 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 in the project structure. There are several options to place the selected data files into the project structure. All of the selected files can be placed in a new sample () , in an existing sample () , or they can be placed 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. Select the “Instrument” type, “Fragmentation” method and “Enzyme” name that was used to generate the experimental data from the drop-down lists. 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", "Instrument", or "Fragmentation" setting for an existing Sample. It may cause errors during further analysis or inconsistencies in the analysis result.

7. Supported Data Formats

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

1. 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 6 projects. See Section 8.8, “PEAKS 6 Projects”.
2. Library Level Support - The instrument vendor's software library is required to 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”.
3. Converter Level Support - Third party converters are required. Users need to install the required converters correctly and let PEAKS know their locations. This only needs to be set up once. PEAKS will call the converter to convert the data to a native supported format before loading. 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, “QSTAR or QTRAP”.
 - .T2D file (AB 4700/4800 series). See Section 8.4.3, “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 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 Thermo MSFileReader library is installed on the same computer as PEAKS. The MSFileReader 3.0 SP3 is bundled with PEAKS and can be installed during PEAKS' installation. The PEAKS Configuration Wizard can install the MSFileReader 3.0 SP3 automatically (see Chapter 3, Section 2, “Instrument Selection”). MSFileReader is publicly available and can be downloaded from Thermo website at the following link:

<https://thermo.flexnetoperations.com/control/thmo/RegisterMemberToAccount>

Note

Install 64-bit MSFileReader with 64-bit PeaksStudio, and 32-bit MSFileReader with 32-bit PeaksStudio.

8.2. Waters/Micromass (MassLynx) Data

Two ways can be used to load Waters' .raw data.


- Library Level Support

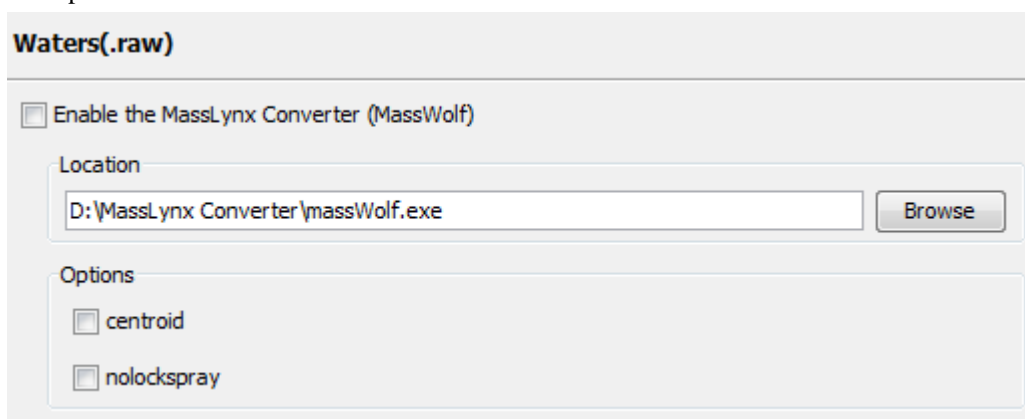
PEAKS is configured to load RAW data from Waters instruments out of the box. No further configuration is required.

- Converter Level Support - `massWolf.exe`

Note

If you want to use `massWolf` in PEAKS, please make sure `massWolf` is installed correctly and works from the command line on your computer.

If you have `massWolf` installed on your computer and prefer to use `massWolf.exe` to convert the `.raw` data, define the converter in the Preferences dialog. Click the “Preferences” toolbar icon  or select Preferences from the Window menu to open the “Preferences” dialog. Clicking the “Waters(.raw)” option under the “Raw file convertor” section on the left will show the preferences for `massWolf`. Point to the location of `massWolf.exe` on your computer and enable it.



8.3. Agilent Data

Agilent Q-TOF data can be loaded in PEAKS provided that MassHunter Data Access (MHDAC) library is installed on the same computer as PEAKS. The MHDAC library is bundled with PEAKS and can be installed during PEAKS installation.

Note

Install 64-bit MHDAC library with 64-bit PeaksStudio, and 32-bit MHDAC library with 32-bit PeaksStudio.

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 3.1.7 is readily available on the Bruker Daltonics web site. You may need to contact your Bruker representative to obtain CompassXtract 3.1.7. See Section 8.5, “Bruker Data”.

8.4. Applied Biosystems/Sciex Data

8.4.1. QSTAR or QTRAP

ABI data can be loaded into PEAKS provided the required converters are installed. Analyst QS is required for QSTAR data, Analyst 1.4 is required for QTRAP data, and AB SCIEX MS Data Converter is required for ABI 5600 data. PEAKS Configuration Wizard can download and install `mzWiff` automatically. AB SCIEX MS Data Converter can convert MS data from any AB SCIEX instruments to `mzML` or `MGF` format that PEAKS can load

readily. AB SCIEX MS Data Converter can be downloaded from the following link. Please see the specified software package to find out the requirements for it.


<http://www.absciex.com/downloads/software-downloads>

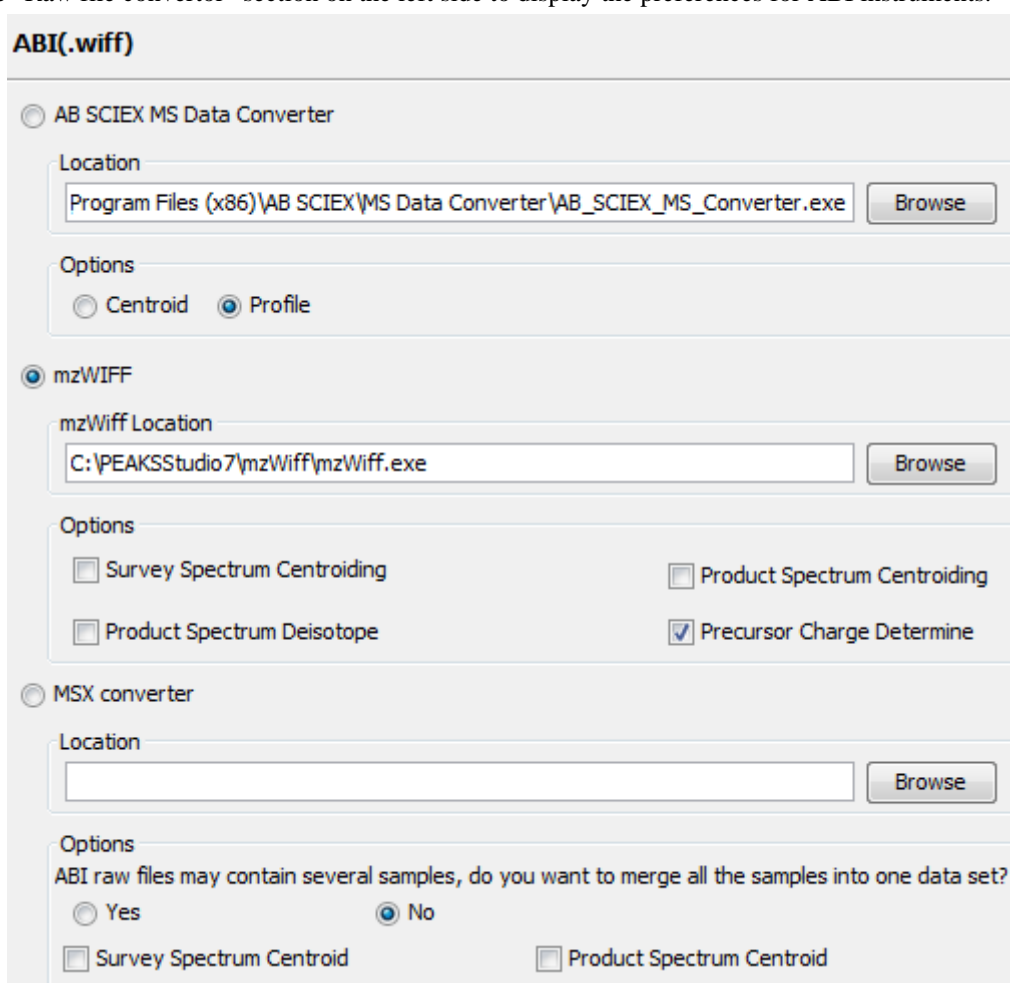
8.4.2. Converters for WIFF

Three Wiff converters are supported by PEAKS: AB SCIEX MS Data Converter, mzWiff, and MSX.

Note

Before defining the converters in PEAKS, please make sure these software packages are installed correctly on your computer and that they convert successfully from the command line on your computer.

To set WIFF related preferences in PEAKS, open the “Preferences” window by clicking the “Preferences” toolbar icon  or selecting Preferences from the Window menu to open the “Preferences” window. Select “ABI (.wiff)” from the “Raw file convertor” section on the left side to display the preferences for ABI instruments.



ABI(.wiff)

☐ AB SCIEX MS Data Converter

Location
Program Files (x86)\AB SCIEX\MS Data Converter\AB_SCIEX_MS_Converter.exe

Options
☐ Centroid ☒ Profile

☒ mzWIFF

mzWiff Location
C:\PEAKSStudio7\mzWiff\mzWiff.exe

Options
☐ Survey Spectrum Centroiding ☐ Product Spectrum Centroiding
☐ Product Spectrum Deisotope ☒ Precursor Charge Determine

☐ MSX converter

Location

Options
ABI raw files may contain several samples, do you want to merge all the samples into one data set?
☐ Yes ☒ No
☐ Survey Spectrum Centroid ☐ Product Spectrum Centroid

Select one of the converters as the default converter for WIFF file loading.

AB SCIEX MS Data Converter. Click Browse to tell PEAKS the location of the raw file converter. Select the preferred method of raw data to load. This is a useful option for ABI 5600 instruments.

mzWiff. Click Browse to tell PEAKS the location of the raw file converter. Select “Survey Spectrum Centroiding” and/or “Product Spectrum Centroiding”, if centroiding has been performed on the survey spectrum and the product spectrum, respectively, before loading it into PEAKS. This is important to ensure PEAKS' optimal performance.

MSX. Click Browse to tell PEAKS the location of the raw file converter. Select “Survey Spectrum Centroid” and/or “Product Spectrum Centroid” if centroiding has been performed on the survey spectrum or the product spectrum, respectively, before loading it into PEAKS.

8.4.3. 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 Chapter 3, 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 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 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 be changed (the default is “tsquared”).
- **Password to access the Oracle database:** most likely does not need to be 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 will be exported.


8.5. Bruker Data

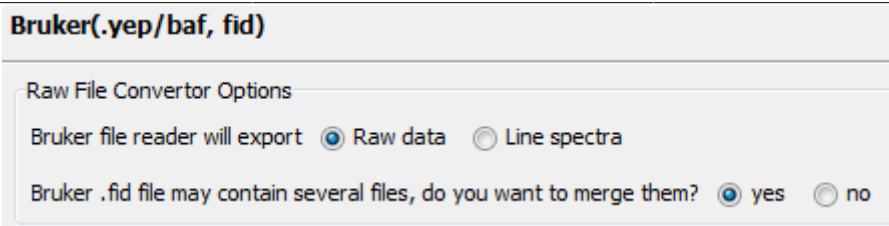
.D and .LIFT directories from Bruker mass spectrometers can be imported provided that the CompassXtract Runtime library is installed on the same computer as PEAKS. The spectral data will be contained in the .yep, .baf, or fid file. Select the .D folder (containing .yep and .baf file) and the .LIFT folder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

CompassXtract 3.1.7 can be downloaded from the Bruker Daltonics web site. You may need to contact your Bruker representative to obtain CompassXtract 3.1.7.

Note

Install 64-bit CompassXtract with 64-bit PeaksStudio, and 32-bit CompassXtract with 32-bit PeaksStudio.


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 hand side. This will show the Bruker instrument preferences on the right hand side.



CompassXtract by default will export raw data. If the attempt to load raw data results in no spectra, then choose “Line spectra”. A Bruker .fid file may contain several samples. By default, these samples are not merged into one data set. Select “Yes” to merge all the samples into one data set.

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. Click Browse to tell PEAKS the location of the Shimadzu run2xml.exe file.



Note


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

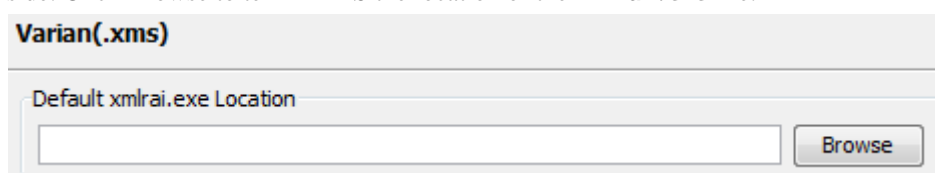
8.7. Varian

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

The .trans data files are converted in Varian programs by clicking File | Save As and selecting the .pkl file format or by clicking File, right clicking Export and 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.

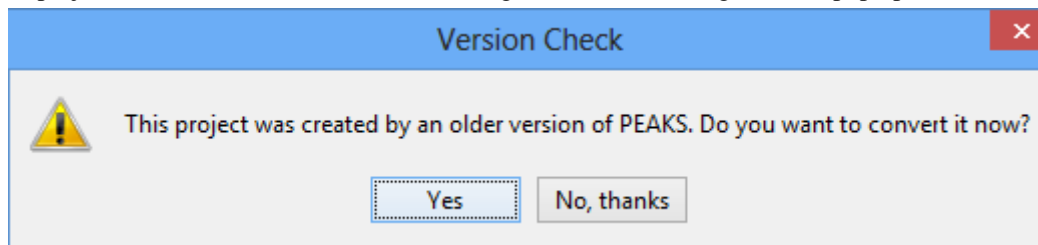


Note

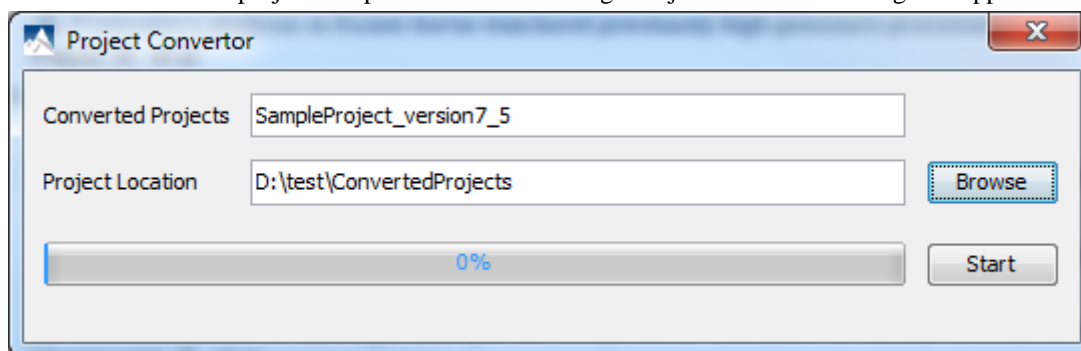
Varian `xmlrai.exe` can not read data files containing spaces or non-english characters in its path.

8.8. PEAKS 6 Projects

Projects created in PEAKS 6 software can be opened in PEAKS 7.5. To convert the project to a PEAKS 7.5 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 dialog box will pop up.



Choose Yes to convert the project and proceed. The following "Project Converter" dialog will appear.



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

For PEAKS 5.x projects, you need PEAKS 6 to convert these projects into PEAKS 6 projects first, then use PEAKS 7.5 to convert the PEAKS 6 projects to PEAKS 7.5 projects.

Note

The project converter is designed to convert PEAKS 6 projects so that they can be viewed in PEAKS 7.5, but further analysis on the converted projects is strongly discouraged.

8.9. PEAKS 7 Projects

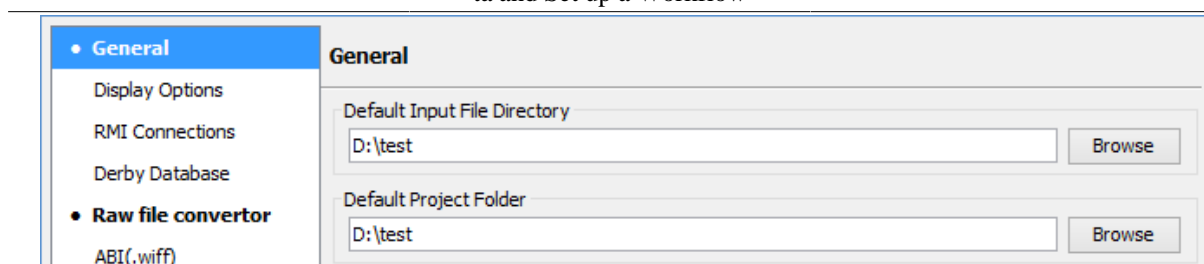
No conversion is necessary to open PEAKS 7 projects in PEAKS 7.5. However, the new features implemented in PEAKS 7.5 will not be available.

9. Changing the Default Project Location

If many projects are to be created, it is convenient to change 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 "Preferences" dialogue pops up.

Project Wizard - Create a
PEAKS Project, Load Data
and Set up a Workflow



2. Select "General" in the "Preferences" dialogue and click the Browse button to the right of "Default Project Folder" to specify the default project location.

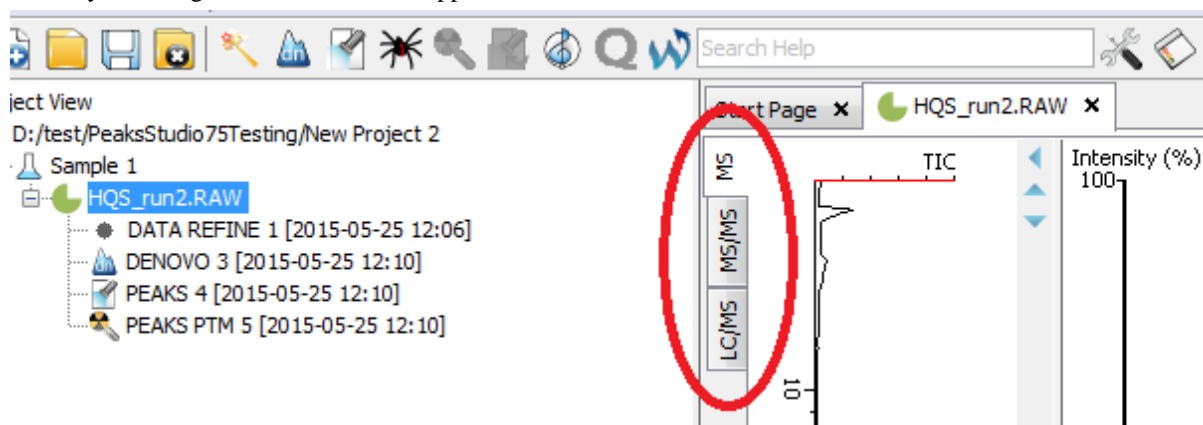
Chapter 5. Data Visualization

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:

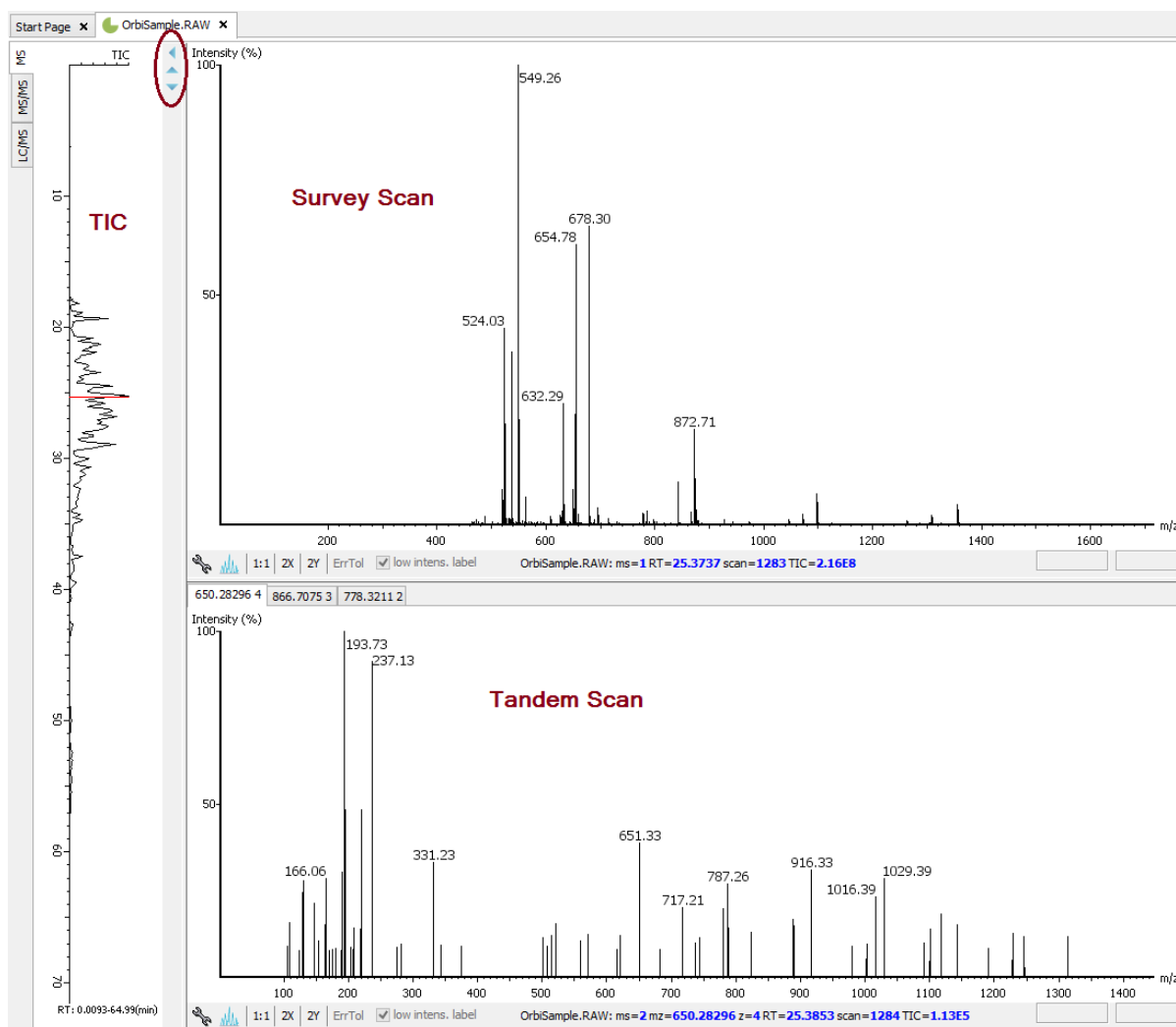
- **MS** - this view shows the TIC (total ion chromatogram) plot and all the MS scans. For each MS scan, the corresponding MS/MS scans are also displayed.
- **MS/MS** - this view lists all the MS/MS scans. For each MS/MS scan, the corresponding MS scan is also displayed.
- **LC/MS** - this provides a 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 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 scans, use the up and down navigation buttons. The survey scans can also be navigated by using the up and down arrow keys on the keyboard. Clicking on a specific position in TIC will display the corresponding survey scan. The tandem scans associated with the current survey scan are shown in the bottom right panel.

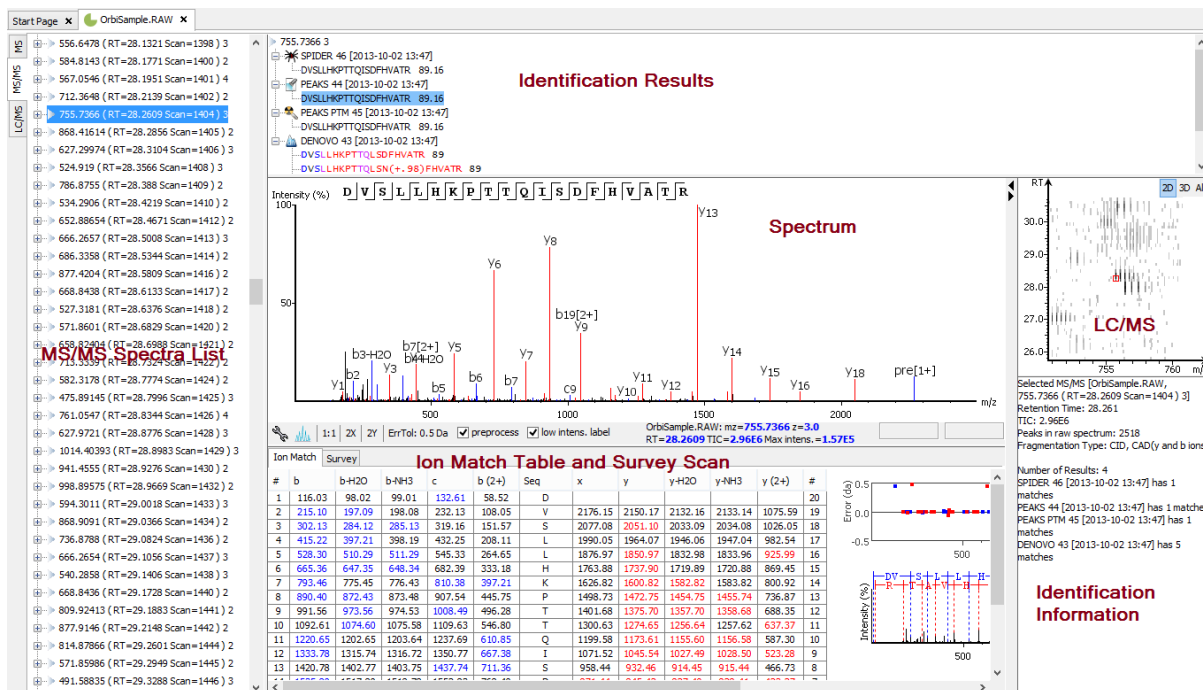


The survey scans and tandem scans provide a few convenient ways to zoom and navigate the spectrum:

- **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.

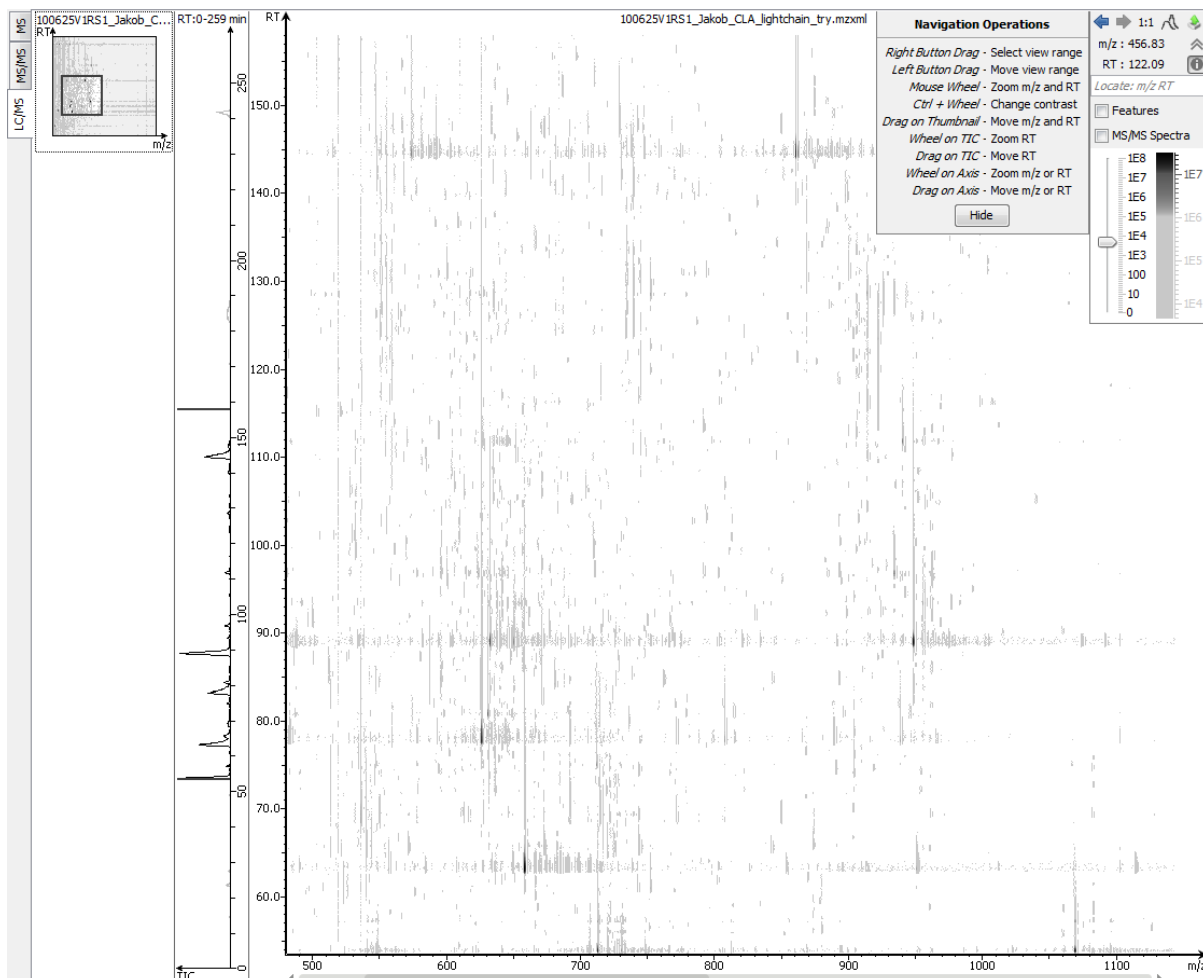
3. MS/MS View

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, 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 retention time of that point.



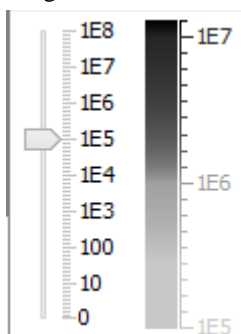
The LC/MS view provides various convenient ways to zoom and navigate the LC-MS features in the data:

- **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, release the mouse button.
- **Zoom in 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, release the mouse button.
- **Move to a specific Heat Map area from Thumbnail:** the 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, 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 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 Thumbnail, scroll the mouse wheel.
- **Zoom in/out smoothly:** place the cursor at a particular (m/z value, RT) position, scroll the mouse wheel button.
- **Move to a specific RT range:** click on the TIC or RT axis, hold and drag the cursor until the desired range is reached.
- **Zoom to a specific RT range:** place the mouse cursor on TIC or RT axis, scroll the mouse wheel.
- **Move to a specific m/z range:** when m/z axis is zoomed in, click the on the m/z axis, hold and drag the cursor until the desired range is reached.
- **Zoom to a specific m/z range:** place the mouse cursor on m/z axis, scroll the mouse wheel.
- **Change contrast of the Heat Map:** place the cursor on the heat map, press and hold the Ctrl key, and scroll the mouse wheel.
- **Change peak height in the 3D intensity mode:** place the cursor on the heat map, press and hold the Ctrlkey, and scroll the mouse wheel.
- **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

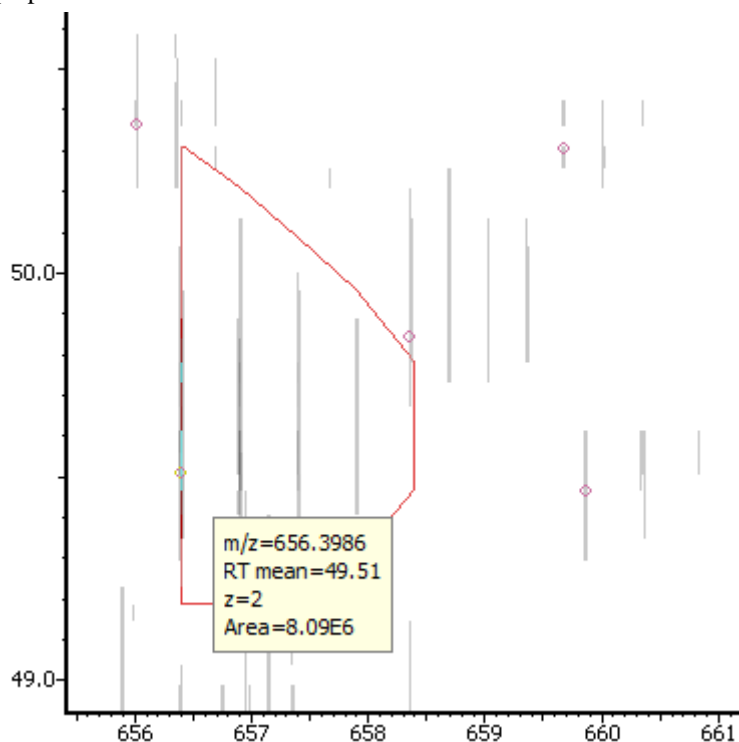
Select the appropriate threshold for noise filtering. Once selected, the Heat Map view will reflect the changes.



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


on such a feature point will show the boundary of the feature using a red polygon and display the information regarding the monoisotopic m/z , mean RT, charge state of the feature, and the total area under the XIC curves of the feature in a pop up window.



4.3. Locate M/Z and Retention Time

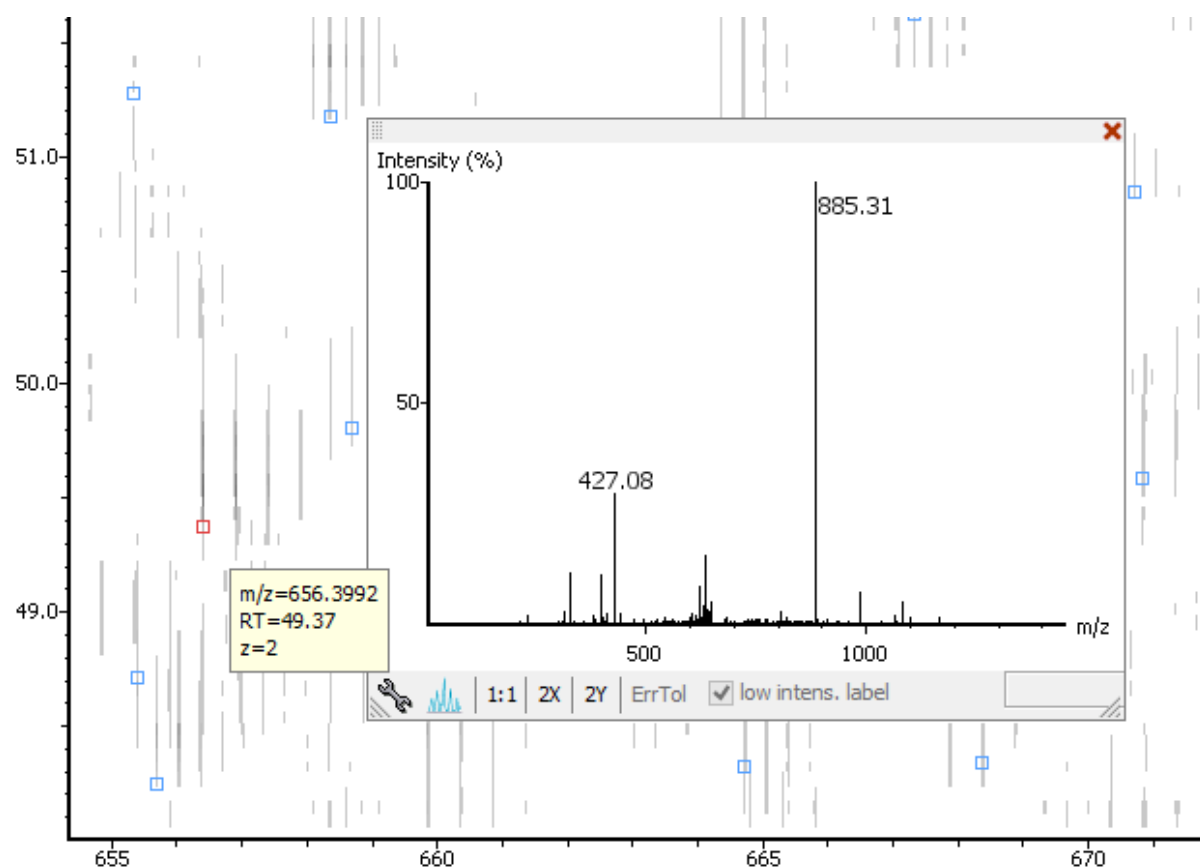
Input M/Z and retention time such as "618.88 50.5" in the following text box to locate a (M/Z, retention time) position in the LC/MS view.

4.4. Synchronize LC/MS View between LC Runs


Use the  button to synchronize LC/MS view between different LC runs. When the LC/MS views are synchronized, the (M/Z, retention time) position and noise level are locked as the same in each run.

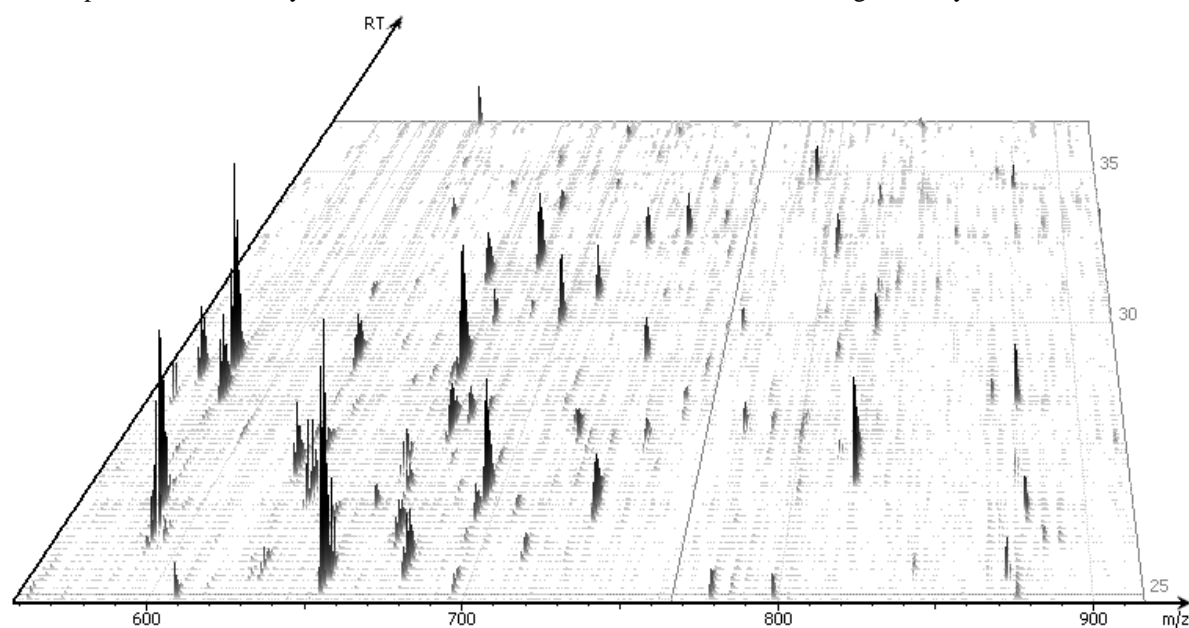
4.5. Show/Hide MS/MS Spectra

Select the MS/MS Spectra checkbox to highlight the scans with associated tandem scans by marking them with tiny blue squares on the heat map. Place the cursor on an MS/MS spectra indicator to view the precursor m/z , RT and precursor charge state information in a pop up window. 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.



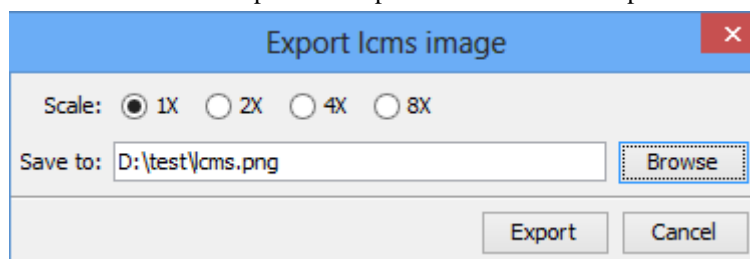
4.6. Intensity View

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 zoom and navigate easily.



4.7. Export LC/MS


Click the  button on the LC/MS control options to export the LC/MS heat map or intensity view in PNG format.

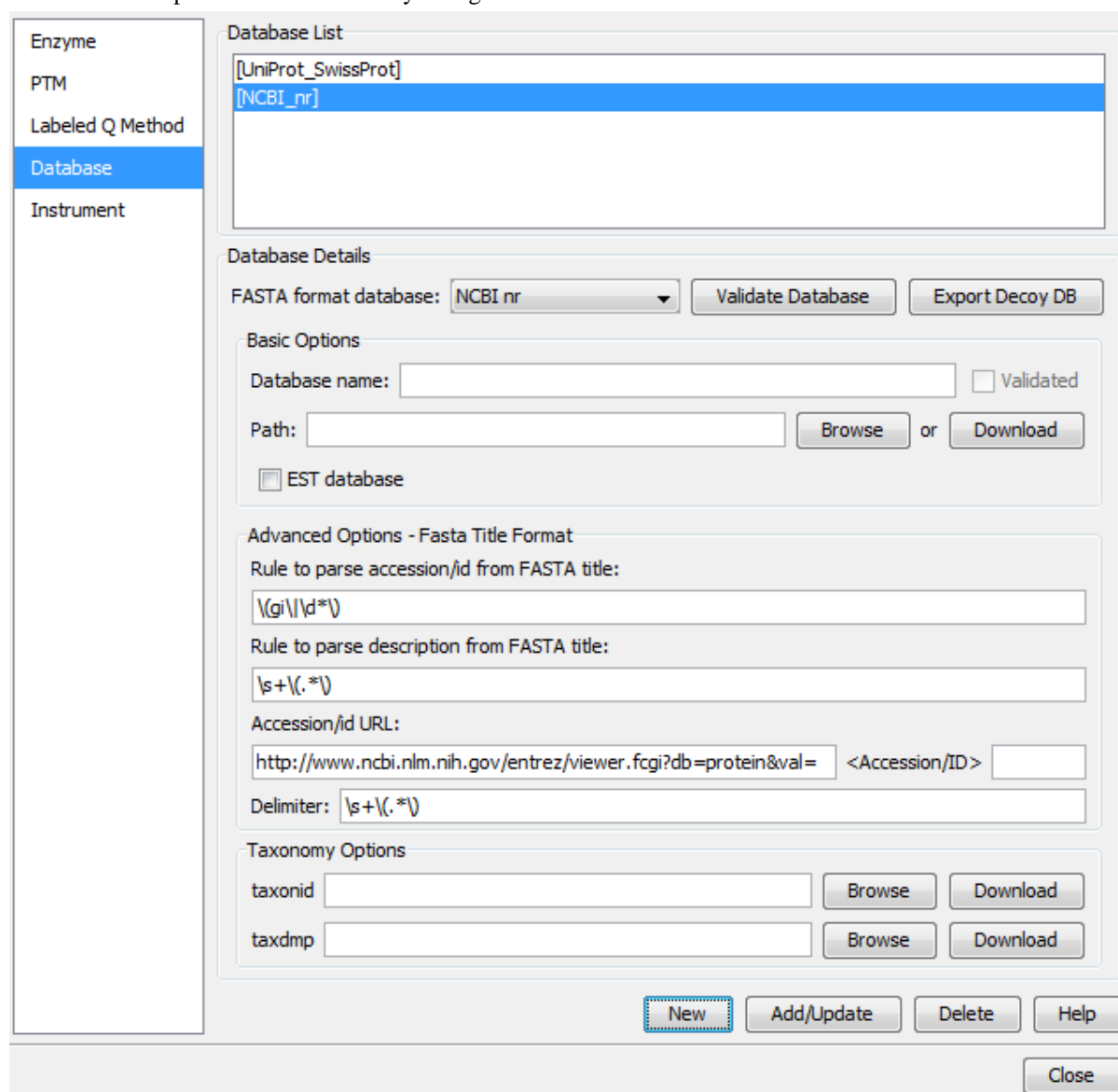


Chapter 6. Adding a Sequence Database

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 a protein or 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 download them 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-hand side. This will open the database configuration dialog. 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.



The screenshot shows the 'Database Configuration' dialog box. On the left is a sidebar with options: Enzyme, PTM, Labeled Q Method, Database (selected), and Instrument. The main area is titled 'Database List' and contains a list box with 'UniProt_SwissProt' and 'NCBI_nr' (highlighted). Below this is the 'Database Details' section. It includes a 'FASTA format database:' dropdown set to 'NCBI nr', with 'Validate Database' and 'Export Decoy DB' buttons. Under 'Basic Options', there are fields for 'Database name:' and 'Path:', a 'Validated' checkbox, and 'Browse' or 'Download' buttons. An 'EST database' checkbox is also present. The 'Advanced Options - Fasta Title Format' section contains three text fields: 'Rule to parse accession/id from FASTA title:' (containing '\(gi\|d*\)'), 'Rule to parse description from FASTA title:' (containing '\s+\(.*\)'), and 'Accession/id URL:' (containing 'http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?db=protein&val= <Accession/ID>'). A 'Delimiter:' field contains '\s+\(.*\)'. The 'Taxonomy Options' section has 'taxonid' and 'taxdmp' fields, each with 'Browse' and 'Download' buttons. At the bottom are buttons for 'New', 'Add/Update', 'Delete', 'Help', and 'Close'.

Follow these steps to configure a database:

1. Select the database format from the “FASTA Format Database” drop-down list, 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 pressed, 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” section 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”. If not, ensure that it is left 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 up by selecting the 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, 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 which has been 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 have been 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 the files.

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 in searching through large databases and may crash. It is therefore suggested that X!Tandem only be used with small databases. If used with a large database, a taxon should be specified. When using *NCBI nr* or *SwissProt* databases with X!Tandem it is best to use a sub-taxonomy.
- **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 a space 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 in the final analysis result.

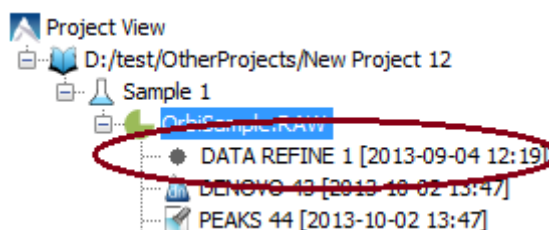
- **Merging scans** : MS/MS scans from the same precursor m/z and at a similar retention time will be merged together.
- **Precursor m/z correction** : the precursor m/z values given by some instruments are often not of the monoisotopic ion. This creates problems in downstream analysis. By examining 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 deconvolution of multiple charge 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, sample or project on the project tree. All the fraction(s) under the selected node will be refined. The use of this function is outlined in the following.

1. Select a project, sample or a fraction node. Click the data refinement button  on the tool bar.



2. Specify the data refinement parameters in the popup dialog 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 analysis 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 toolbar icon . The "Data Refinement Parameters" dialogue window will appear:

Data Refinement Predefined parameters default

☐ **Merge Scans**

Retention time window (for raw files only): min

Precursor m/z error tolerance: ☒ Da ☐ ppm

☐ Merge CID and HCD scans together

☒ **Correct Precursor**

☒ Mass only (recommended)

☐ Mass and Charge states

Min charge: Max charge:

☐ **Filter Scans**

Only keep scans satisfying:

☐ Precursor mass between and Da

☐ Retention time between and min

☐ Quality value greater than (suggest 0.65)

- **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: When merging scans, the scans of different fragment type will not be merged. If users want 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 in data files.

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

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

- **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 value is 0.65. This is a percentage.

Note

Data pre-processing (centroiding, deisotope, 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 list at the top-right of the window, select "Save as..." and define a name for these preferences for future use/reference when prompted. Any parameters that are saved will be available in the drop-down list at the top of the window. To examine the contents, select a saved parameters 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 *de novo* sequences. The local confidence score ranges from 0% to 99%, indicating how confident the algorithm considers a particular amino acid is 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 spectrum.

Note

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

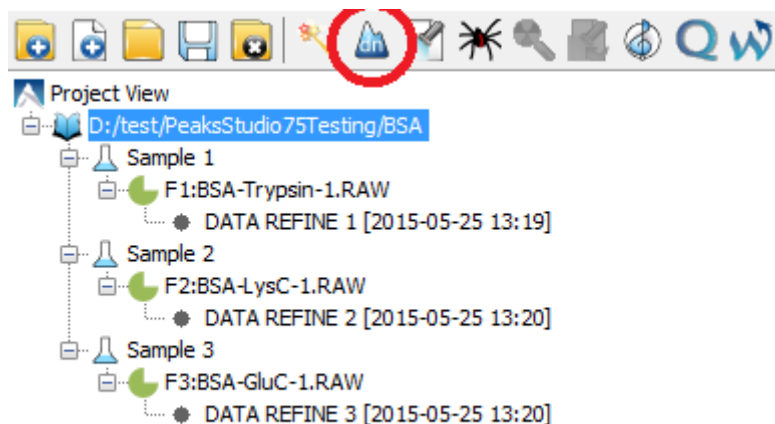
Other PEAKS tools, including PEAKS DB for database search; PEAKS PTM for unspecified PTM search; and SPIDER for homology search depend on the *de novo* sequencing result to achieve superior sensitivity and accuracy.

Note

Refer to Chapter 9, 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 "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* 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 dialog. 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.

The screenshot shows the PEAKS Project View window with the 'DENOV0 4 [2015-05-25 13:25]' node selected and circled in red. To the right, the 'De novo' results table is displayed, showing a list of scans and their corresponding peptides.

Scan	Peptide
1 F1:1432	YLIELAR
2 F3:441	LTKVHKE
3 F3:959	ELDTLSSKLQE
4 F1:1840	LVDELTEFAK
5 F1:1075	KFDLGEEHFK
6 F1:583	ATEEQLK
7 F1:2558	WFLGSFLYEYSR
8 F1:617	ATEEQLK

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 dialog will appear:

De Novo Predefined parameters

Error Tolerance
 Precursor mass: 15.0 ppm Fragment ion: 0.05 Da

Enzyme
 Specified by each sample New

PTM
☒ Carbamidomethylation
☒ Oxidation (M)
☒ Deamidation (NQ)
 Set PTM
 Remove
 Switch type
 Maximum allowed variable PTM per peptide 3

General Options
 Report up to 5 candidates per spectrum

OK Cancel Help

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, default parameters of the first sample is 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 list. 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 enzyme specified during project creation.

Note

"None" is a special enzyme allowing non-specific cleavage at both ends of the peptide. It is recommended 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 Chapter 17, 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” dialog.

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]

Recent Common Uncommon Artificial Customized View New Remove Remove All Switch Type OK Cancel

Selected Fixed PTM
☒ Carbamidomethylation

Selected Variable PTM
☒ Oxidation (M)
☒ Deamidation (NQ)

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 post-translational modifications, modifications induced in standard sample preparation, and common artifacts.

- **Uncommon**

The "Uncommon" list contains less commonly observed modifications including rare post-translational modifications, residue substitutions, and rare 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 dialog, 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 recommended to use a limited number of 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.

PTM name:	Carbamidomethylation		
PTM abbreviation:	Carbamidomethyl		
Mass (Monoisotopic):	57.021465		
Residues that can be modified:	C	Anywhere	
<input checked="" type="radio"/> peptide <input type="radio"/> protein		N-term	
		C-term	
Formula:	H(3) C(2) N(1) O(1)		
Rule:			
<input type="button" value="Close"/> <input type="button" value="Help"/>			

Double clicking any of the PTMs will display the PTM details in the “PTM Info” popup dialog. 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 to the listed modification, it can be defined as a customized modification. Click the New button to show the “New PTM” dialog and enter the information of the particular modification. The newly defined modification will be added to the “Customized” list. Refer to Chapter 17, 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 list, 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 list.

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:** Outline of PEAKS *de novo* sequencing result with 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.

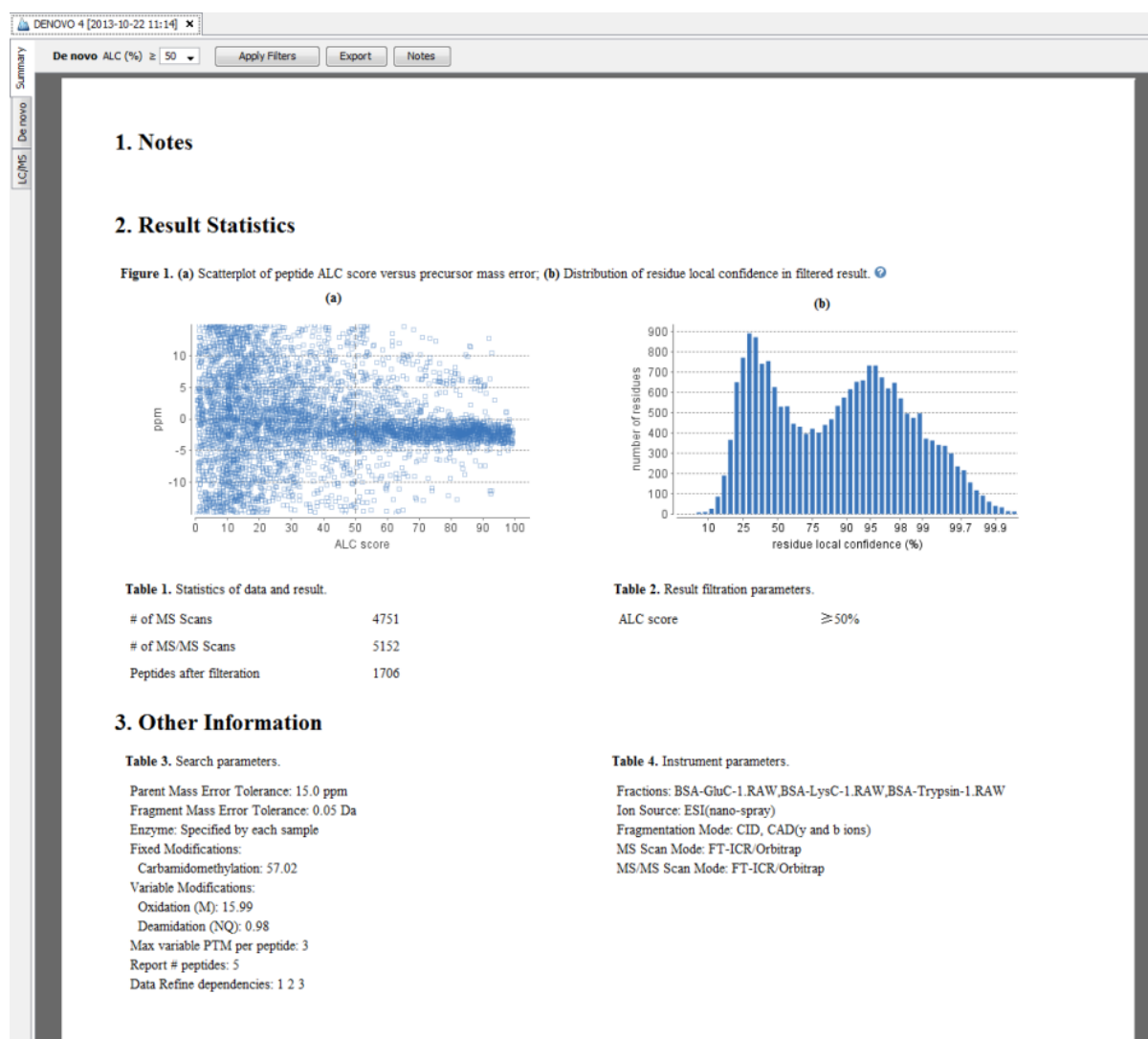
3.1. Summary View

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 to 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:** 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, and the bottom section provides additional information about the peptide-spectrum match.



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.

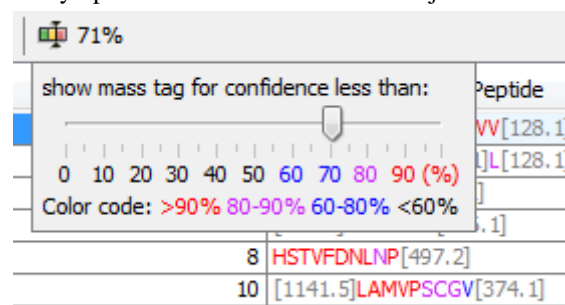
The following list describes the contents in each column:

- **Scan:** Scan number. The scan number is a unique index for tandem mass spectra in the data.
- **Peptide:** Amino acid sequence of the peptide as determined by *de novo* sequencing. A modified residue is followed by a pair of parentheses enclosing the modification mass.
- **Tag Length:** 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 (%):** 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 length of the amino acid sequence of the peptide.
- **m/z:** Precursor mass-to-charge ratio.
- **z:** Precursor charge.
- **RT:** Retention time (elution time) for the spectrum as recorded in the data.
- **Area:** Peptide feature area associated with the PSM.
- **Mass:** Monoisotopic mass of the peptide
- **ppm:** Precursor mass error, calculated as $10^6 \times (\text{precursor mass} - \text{peptide mass}) / \text{peptide mass}$.
- **PTM:** Types and numbers of modifications present in the peptide shown as color-coded icons.
- **Mode:** 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 mouse cursor over the sequence to view local confidence scores of each individual amino acid.

F3:1921	GHPEYAVSVLLR
F1:1559	FKDDAAC(+57.02)C(+57.02)KDVFARGE
F3:1207	TLC
F3:2471	EGF
F3:1717	LSH
F3:2254	LGE

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 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 mouse 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, subsequence, 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.

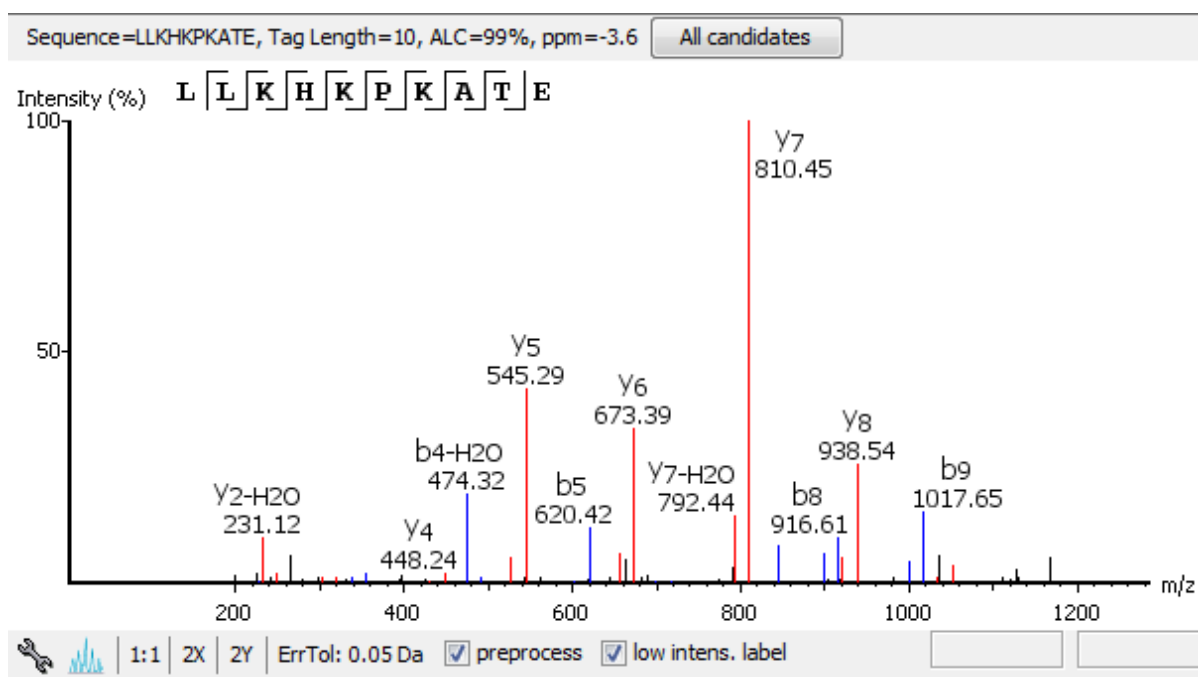
scan =	search	no results
seq contains	-3.6	
m/z ≈	-3.7	
RT ≈	-2.7	
PTM Δm ≈	-3.8	
	-1.5	
	-1.8	

Note

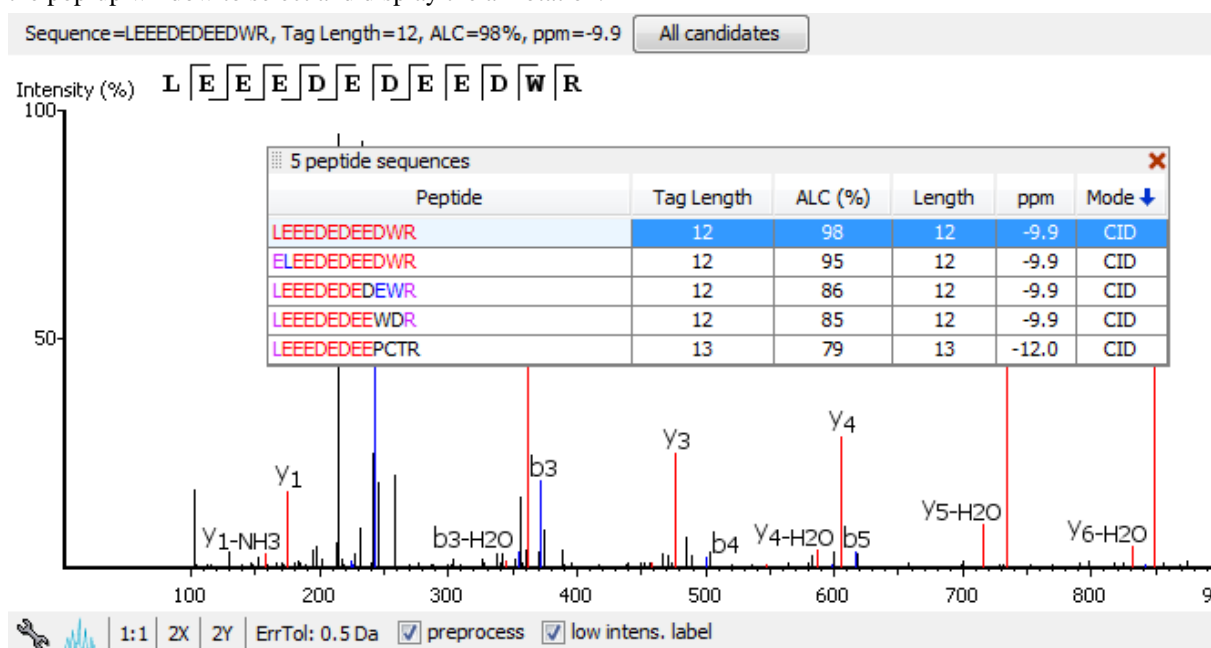
Precursor m/z, retention time, and modification mass are searched approximately. For example, 130.3 will match any value from 130.25 to 130.35 (exclusive). And 130 will match from 129.5 to 130.5 (exclusive).

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



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

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

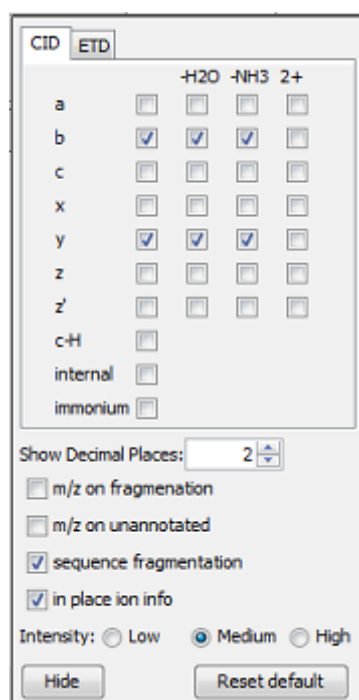
- **Increase/Decrease peak intensity:** Place the cursor in the spectrum and scroll the mouse wheel.
- **See the whole spectrum:** Double click in the spectrum or click the 1:1 button.

The “ErrTol” is used to adjust the error tolerance to view the display of matched ions.

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

The “intensity threshold” check box provides an option to annotate lower intensity peaks.

To change the "Spectrum Annotation Preferences", click the  button to open the dialog. Refer to Chapter 17, Section 1.4, “Spectrum Annotation Preferences” for more details.



3.2.3. Ion Table

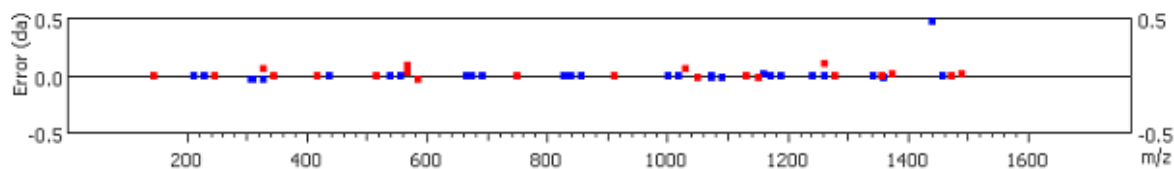
The “Ion Match” tab at the bottom panel of the *de novo* view contains the “Ion Table” that shows the calculated mass of possible fragment ions. If a fragment ion is found in the spectrum, its mass value is displayed in color. N-terminal ions are shown in blue, and C-terminal ions are shown in red. A fragment ion is found when there is a matching peak within the mass error tolerance, as defined in the *de novo* sequencing parameters, and relative intensity of the matching peak is at least 2%. The ion types displayed in the table can be configured in "Spectrum Annotation Preferences" (Section 3.2.2, “Spectrum Annotation”).

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

#	b	b-H ₂ O	b-NH ₃	Seq	y	y+H ₂ O	y-NH ₃	#
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

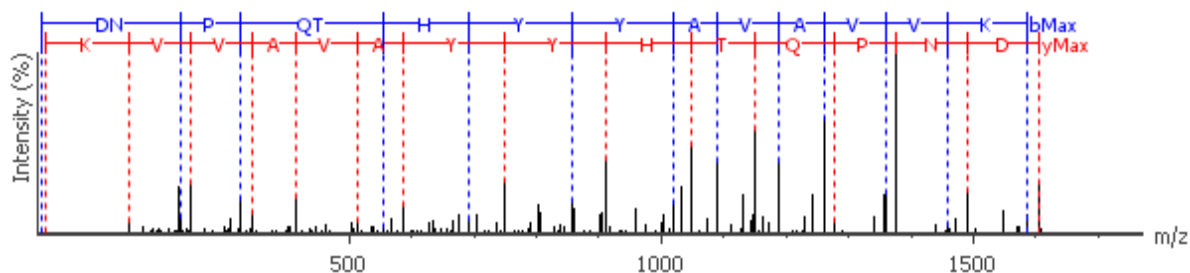
3.2.4. Error Map

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”).



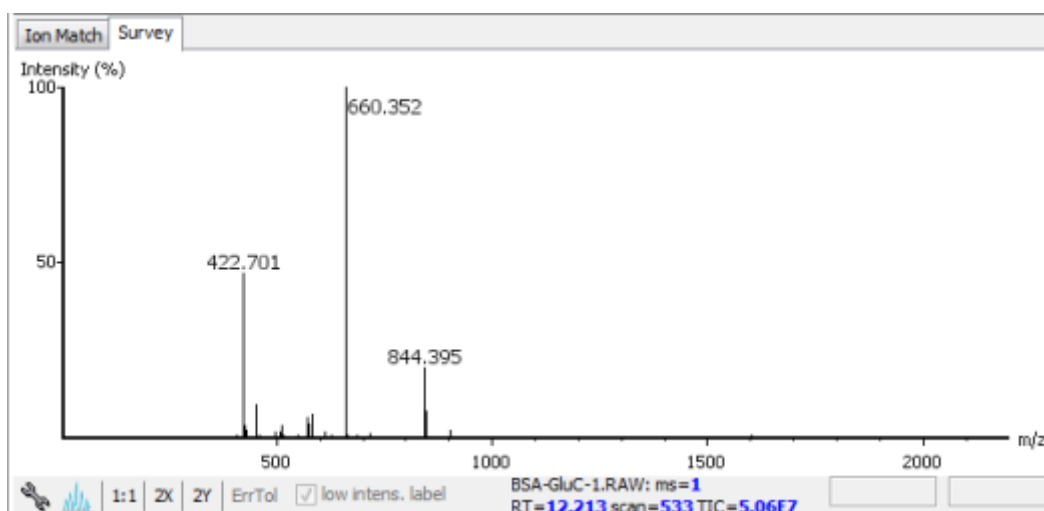
3.2.5. Spectrum Alignment

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.



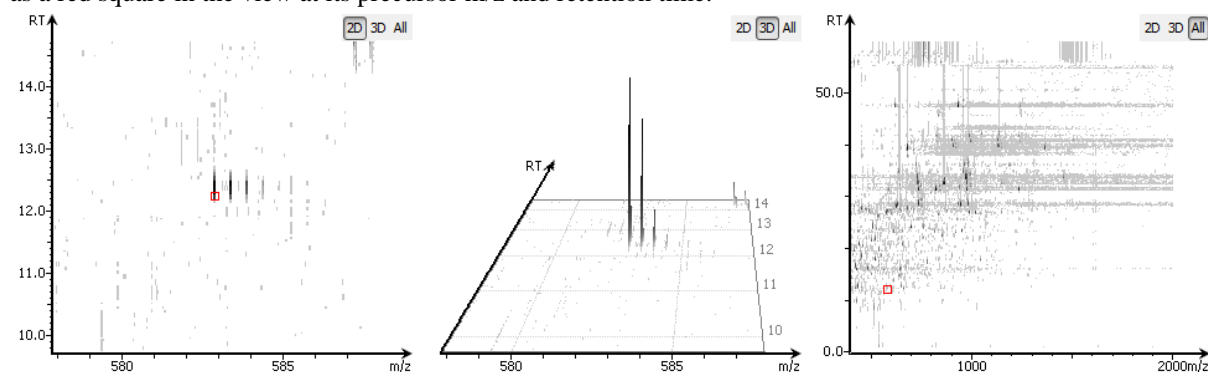
3.2.6. Parent Scan

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

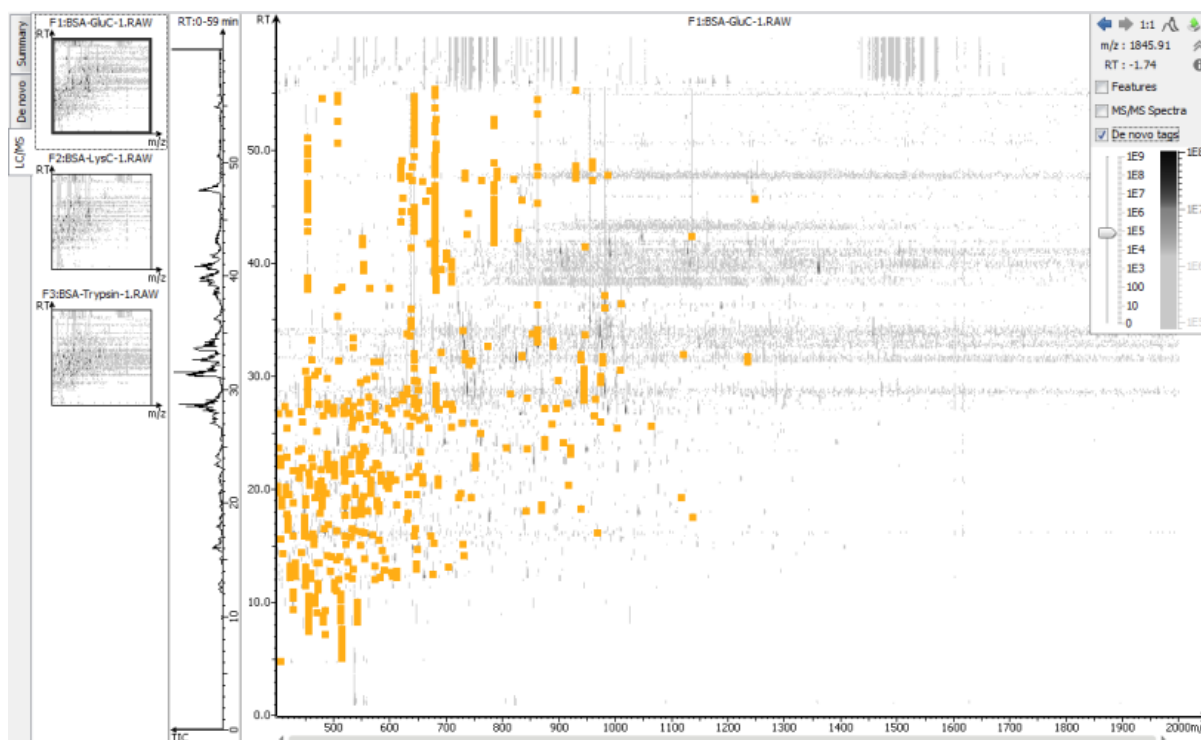
LC-MS Snapshot displays a zoomed-in snapshot of the peptide-spectrum match in LC-MS view. It has three modes of viewing that can be toggled using the selection bar in the top right corner. In the *2D/3D* mode, the view focuses on the peptide feature containing the peptide-spectrum match, showing the 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

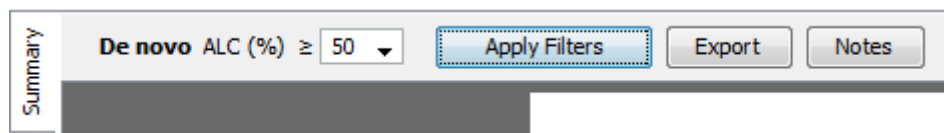
LC-MS View visualizes *de novo* sequences in an interactive LC-MS heat map. Check the option "De novo tags" to show the locations of MS/MS spectra with *de novo* sequences. Those MS/MS spectra are indicated by amber squares. Only the filtered *de novo* sequences are shown in the heat map.

Refer to Chapter 5, 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 Peptide View, and LC/MS View will be updated accordingly.

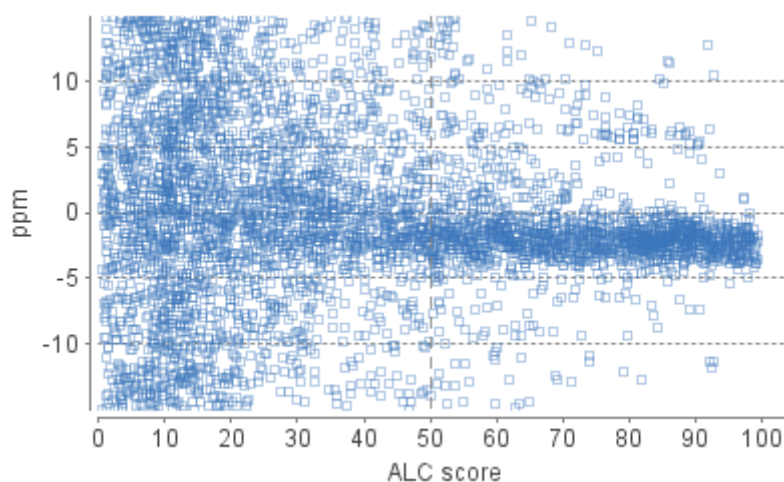


Note

Whenever the score threshold is changed, the "Apply Filters" 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 sequenced. 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:

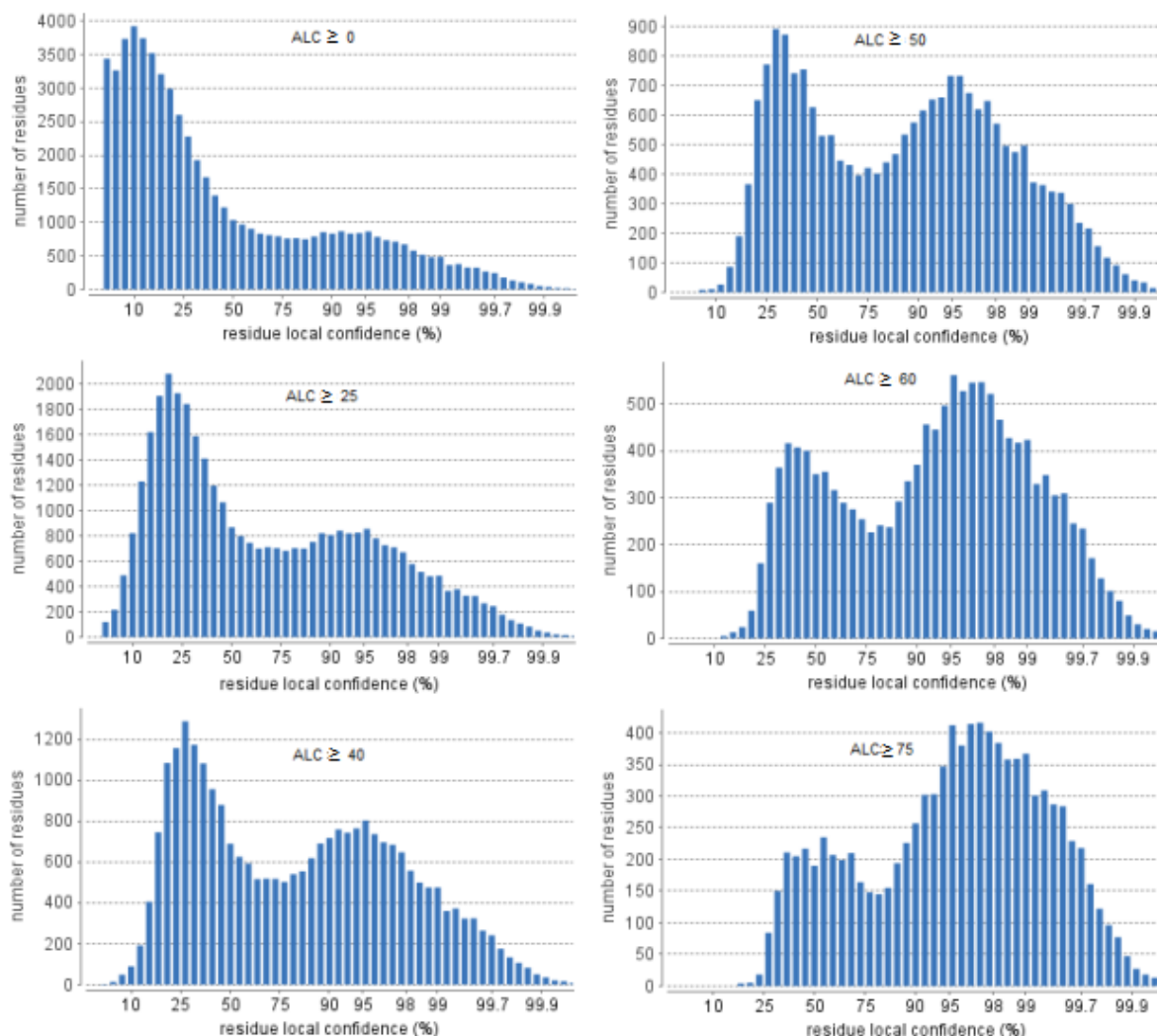
Peptide ALC Score Scatter Plot. Figure 1(a) is a scatter plot showing *de novo* sequences' ALC score versus precursor mass error in ppm. In a typical *de novo* sequencing analysis, peptides with higher ALC scores are more likely to have a small mass error, indicating a higher chance for the peptide to be 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

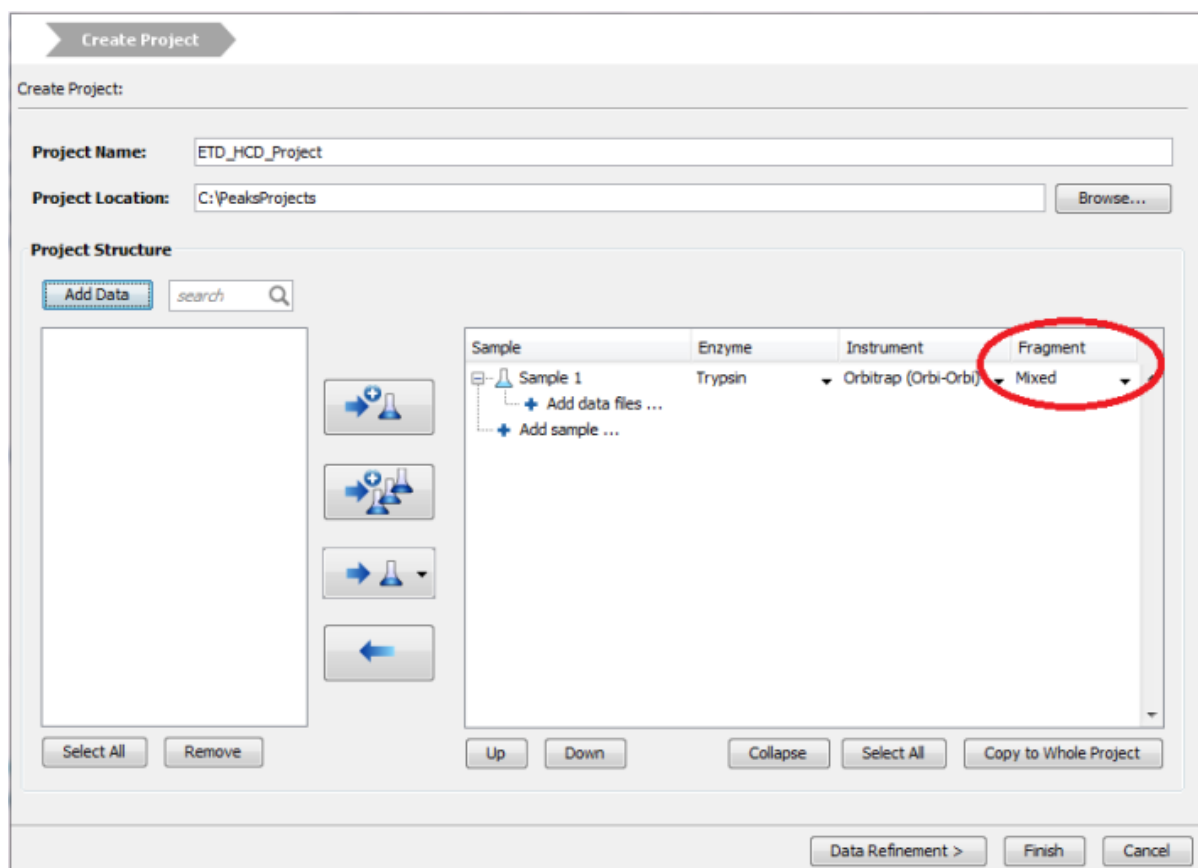
The Export button in the title bar of the Summary View allows exporting of the filtered results into various formats. It provides the opportunity to share the report in HTML and to also use the result as supplemental information in a publication. Please refer to Chapter 16, Section 3, “Exporting *De Novo* Results” for details.

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

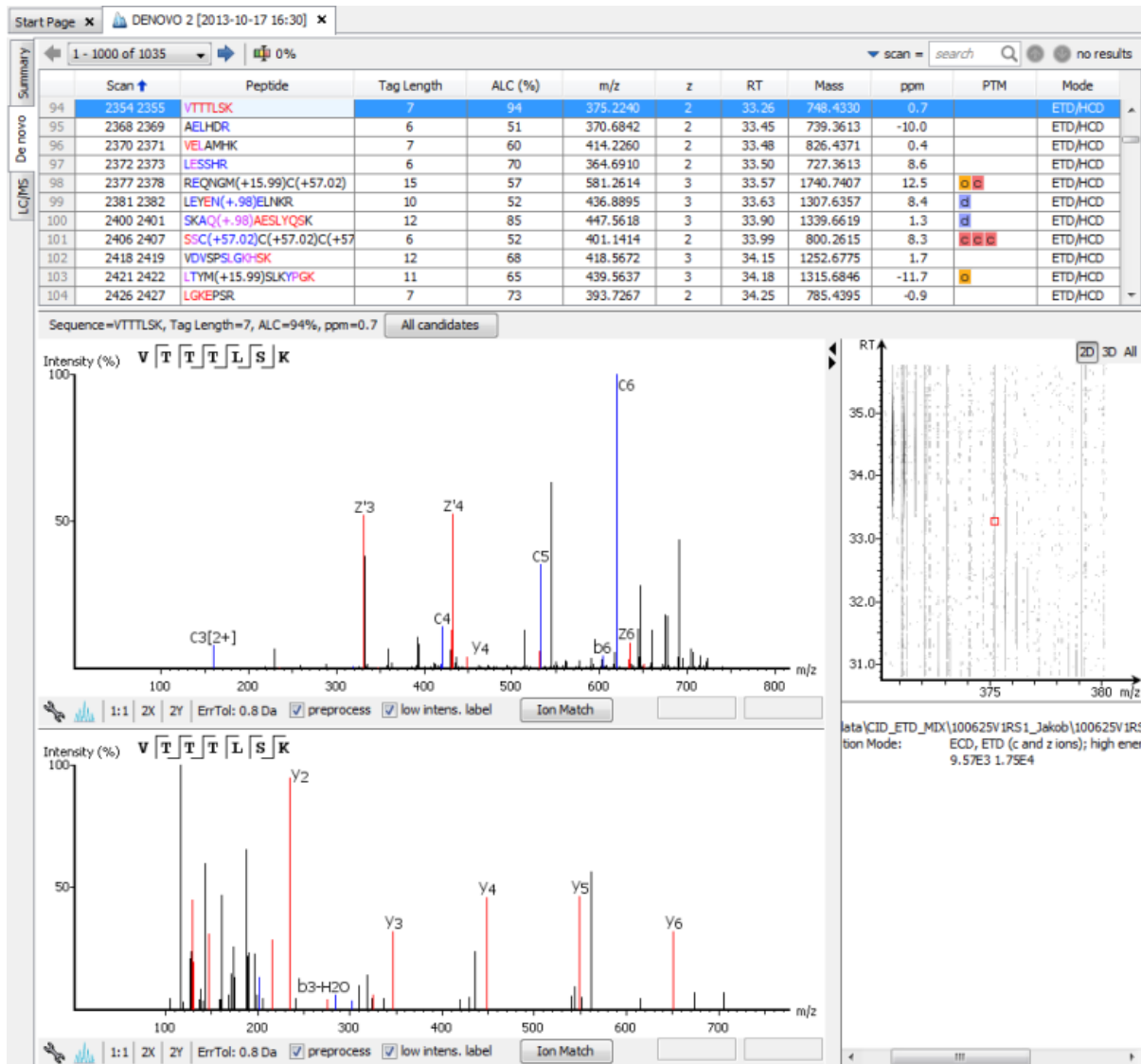
PEAKS supports *de novo* sequencing using spectral pairs or triples that are generated using different fragmentations. Mass spectrometers like the LTQ Orbitrap Velos is capable of generating multiple tandem mass spectra using different fragmentations for the same precursor ion. Using complementary fragmentations 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.

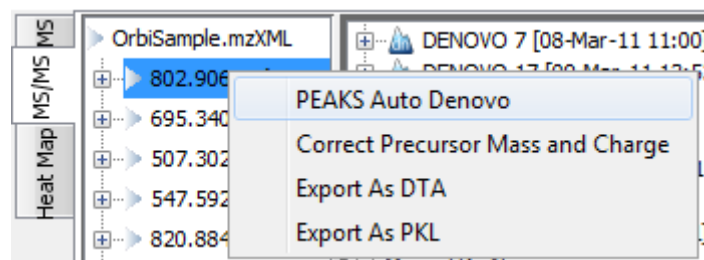


7. *De Novo* Sequencing on Single Spectrum

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

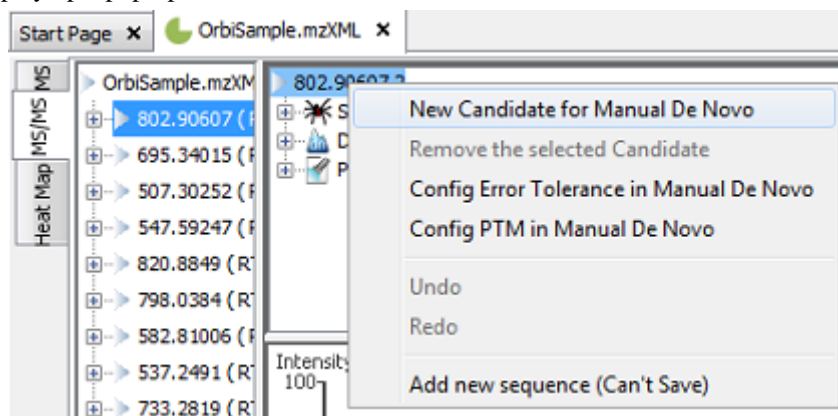
PEAKS provides a set of tools to help manually sequence a peptide from the spectrum.

Note

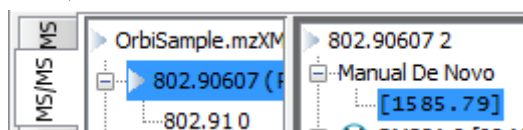
Manual *de novo* sequencing does NOT support ETD spectra.

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.



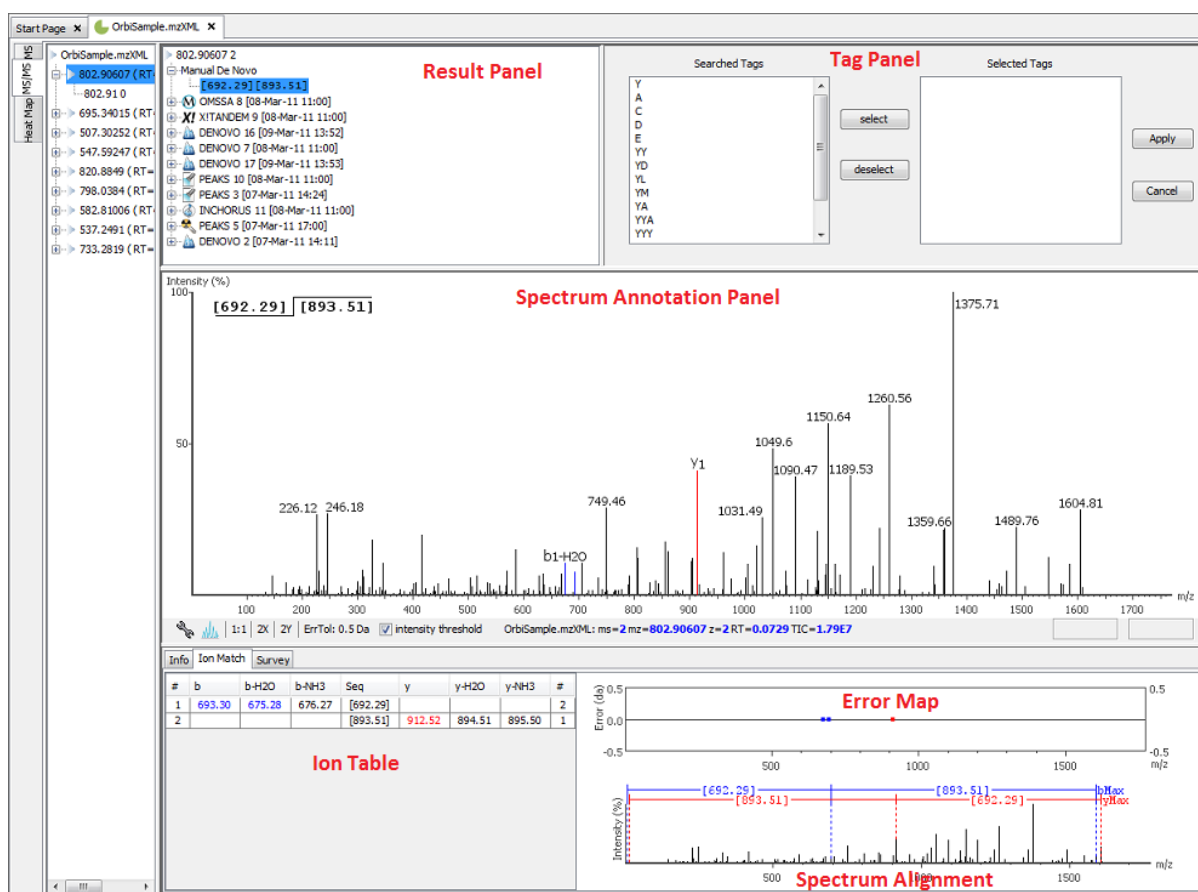
Select New Candidate for Manual De Novo from the pop-up menu. A new candidate will be created under the 'Manual *De Novo*' heading. The new candidate is initiated as unsequenced, so it is represented by the total mass of the residues(see an example below).



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:

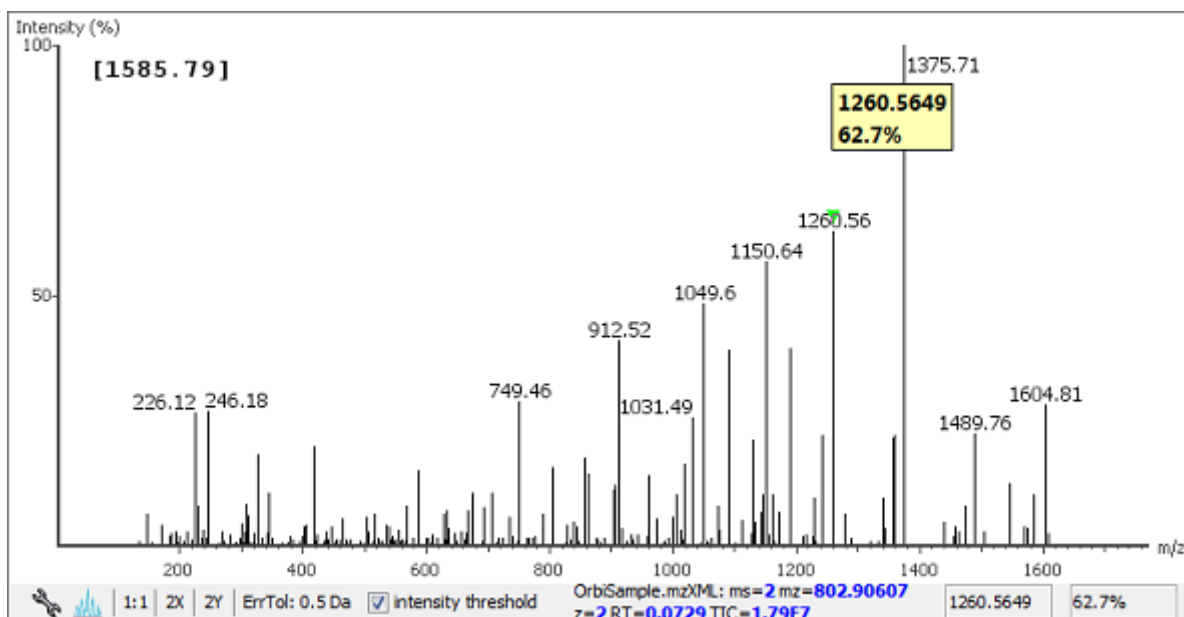


The panels are briefly described below:

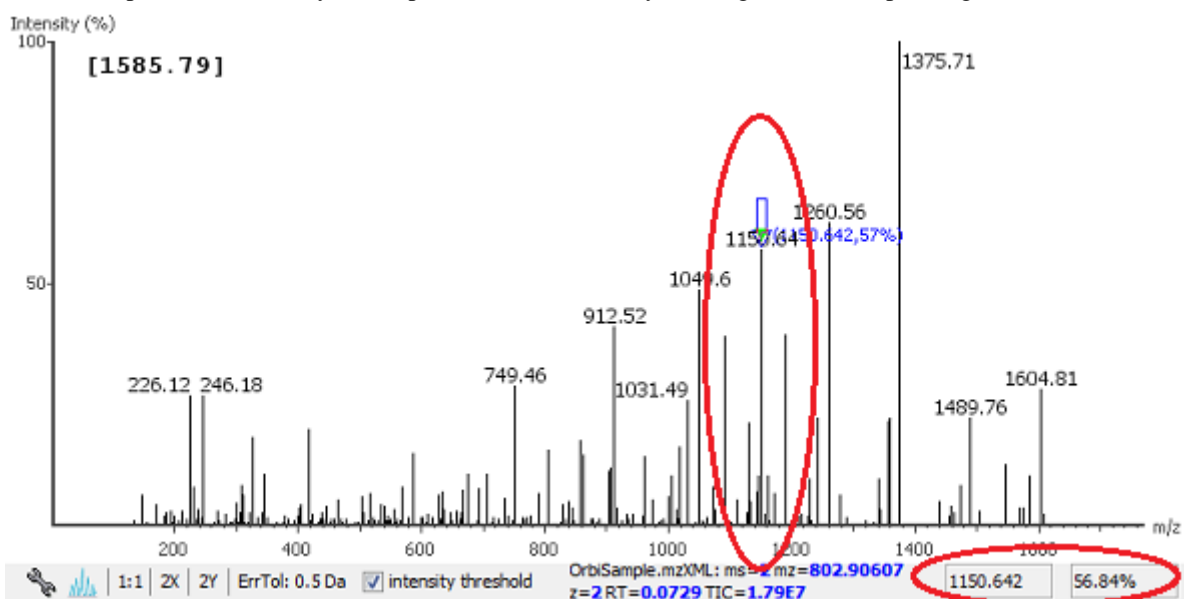
- **Result Panel:** The “Result Panel” shows all sequencing results. The results of manual *de novo* are listed under “Manual *De Novo*”.
- **Spectrum Annotation Panel:** The “Spectrum Annotation” shows a graphical representation of the spectrum, the peaks in the spectrum, the user-selected peaks and 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

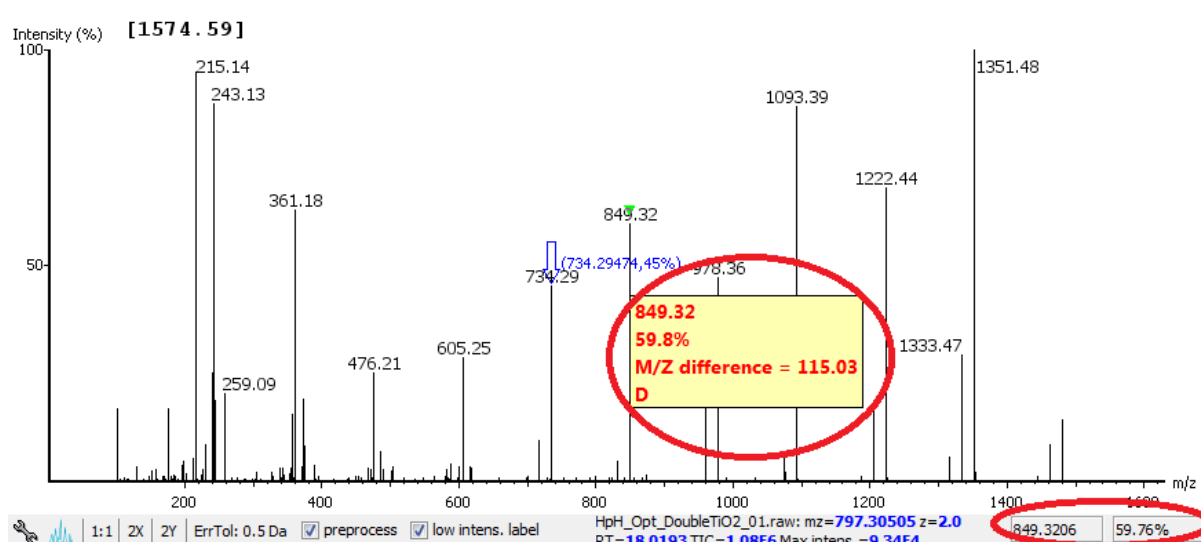
When the mouse cursor is placed in the “Spectrum Annotation” panel, a green (by default) triangle follows the movement of the mouse. This is the *Position Bar* and it is used as a cursor for all manual *de novo* operations. The cursor's position on the m/z scale and its relative intensity are shown in 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.



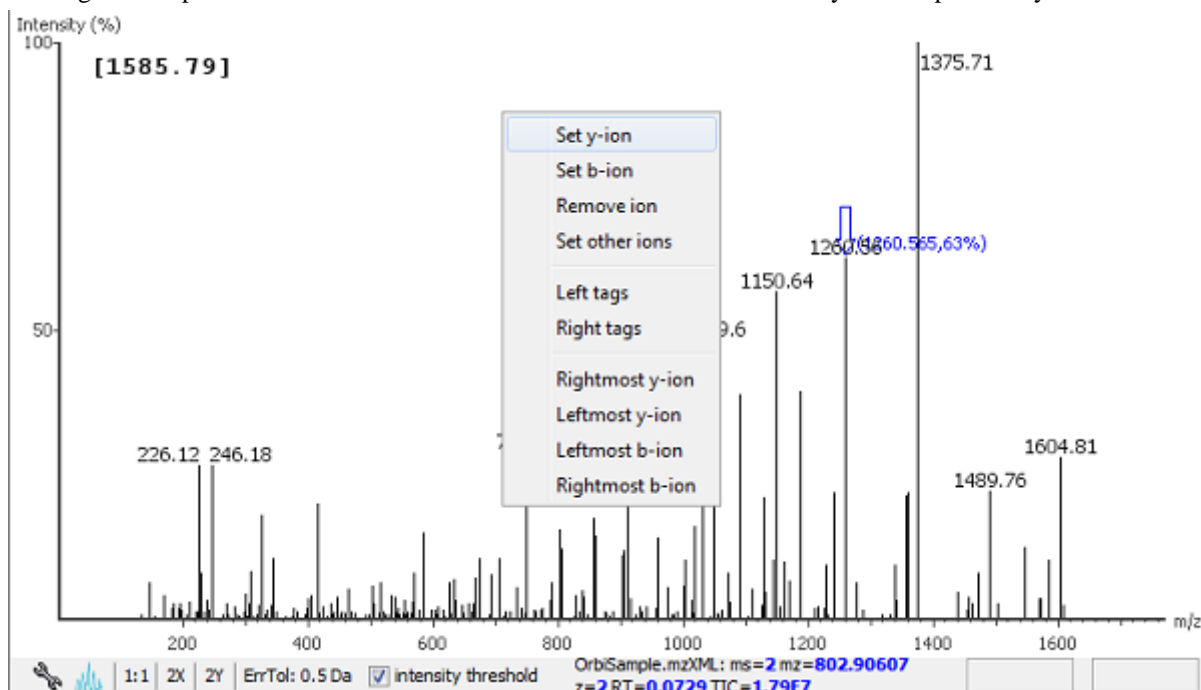
Measure the m/z difference between two peaks. Select a peak (blue arrow) with the *Freeze Bar* and move the mouse to the left or right. Hold the *Position Bar* (green triangle) above another peak. A tool tip displays the difference between the two peaks (in the example below the difference is 109.92297), relative intensity, and the m/z difference. If the mass difference matches the mass of an amino acid, it also shows the corresponding single letter amino acid code.



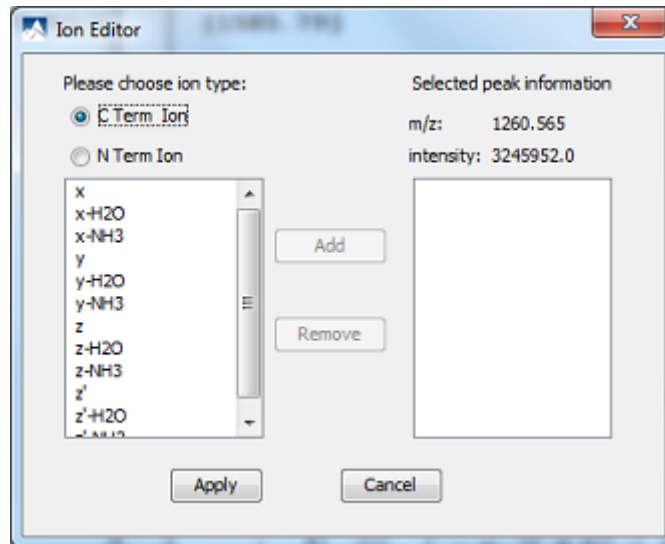
Deselect a peak. The Freeze Bar can be deselected by holding Ctrl and clicking anywhere in the spectrum.

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.

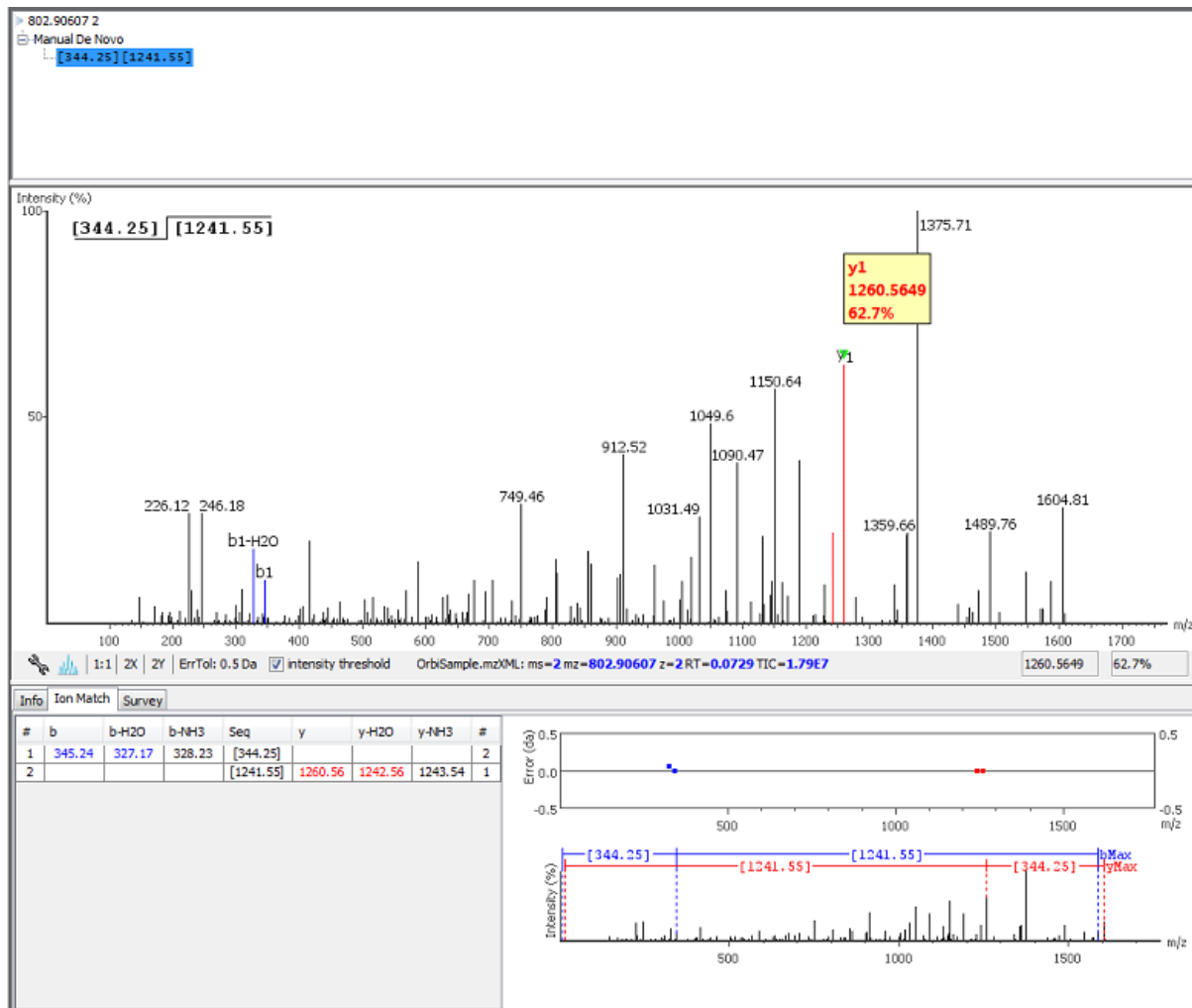


Select Set other ions from the pop-up menu to view the “Ion Editor” dialog box. The “Ion Editor” dialogue allows addition or removal of ion designations to/from a peak. Select either “C Term Ion” or “N Term Ion” to see the C and N terminal ions respectively. Then select an ion from the ion list and press 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.

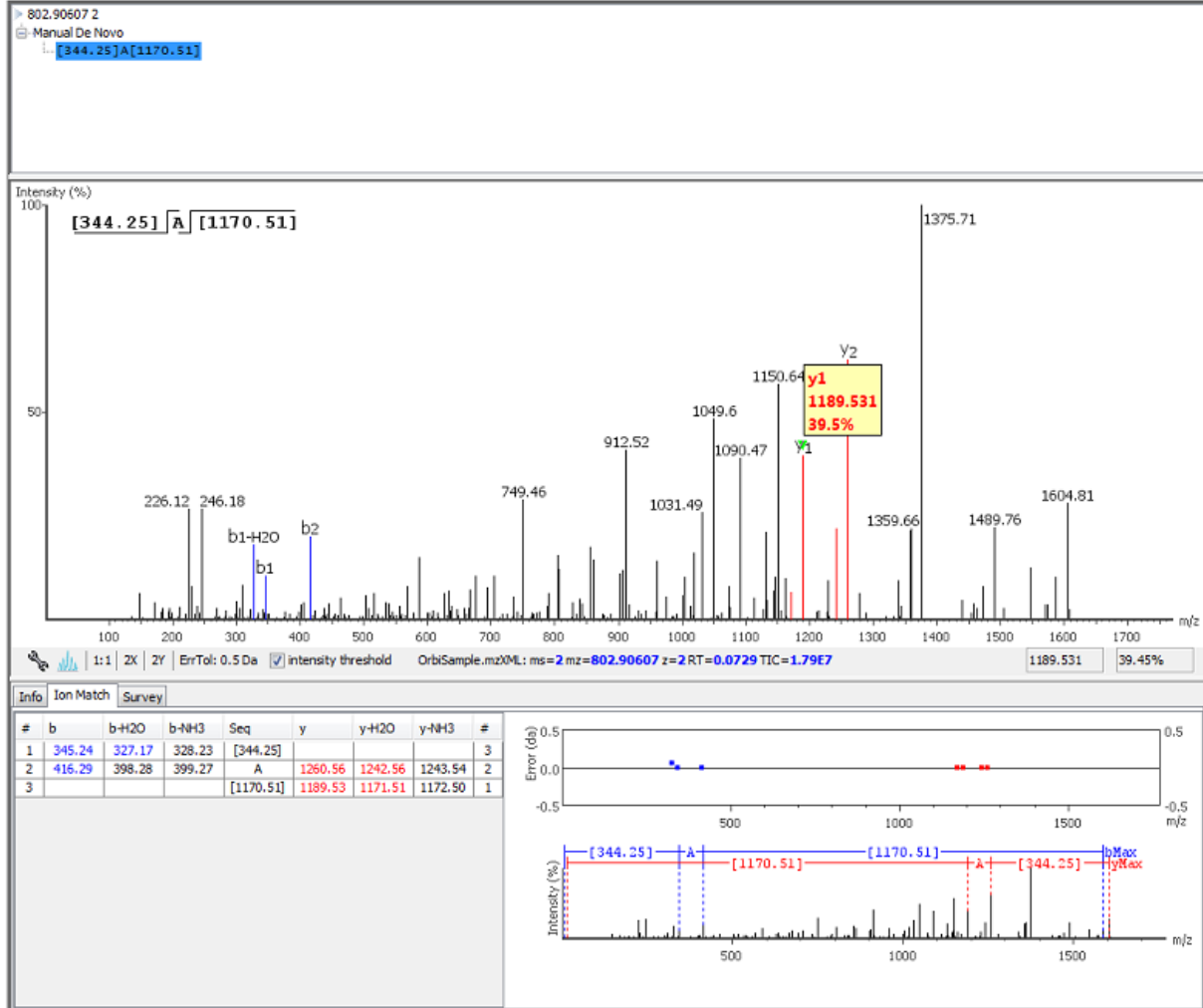
In the example below, the selected peak at 1260.5649 m/z was designated as a y-ion.



Note

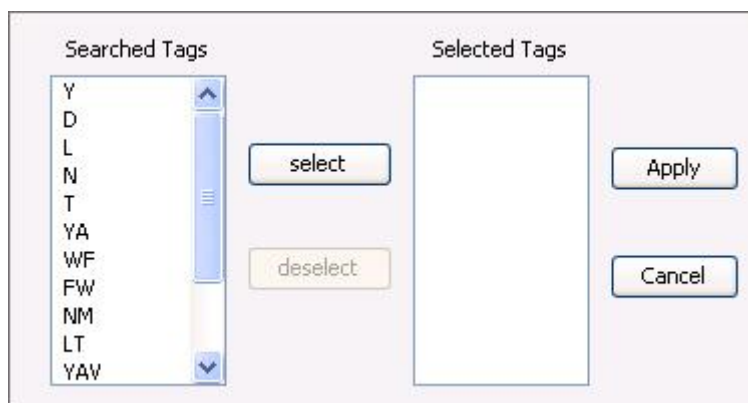
The manual *de novo* candidate information is updated in the “Result” panel, “Ion Table” panel and “Spectrum Alignment and Error Map” panel. The selected ions are also annotated and color coded in the “Spectrum Annotation” panel.

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.

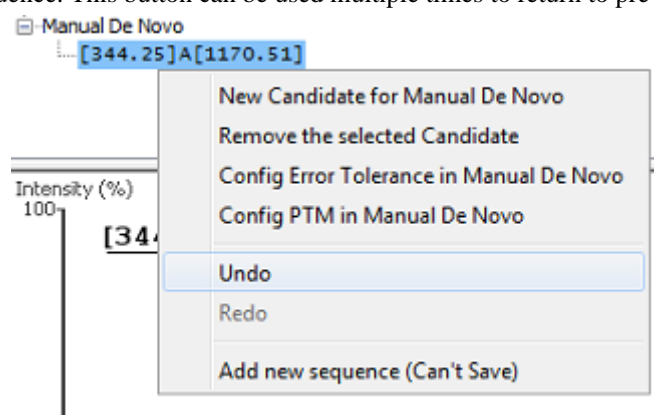


Searching the left or right side of the spectrum for the first/last y or b ion (Search a sequence tag). Select a peak and right-click anywhere in the Spectrum View to trigger the pop-up menu. From the menu, select either Left tags or Right tags. PEAKS will select the appropriate terminal tags and show 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, right-click and select the Config Error Tolerance in Manual De Novo command from the pop-up menu. This will open a dialog where the error tolerance can be set.

PTM Configuration. To identify post-translational modifications (PTM) 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 of modifications and mutations. 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 search 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.

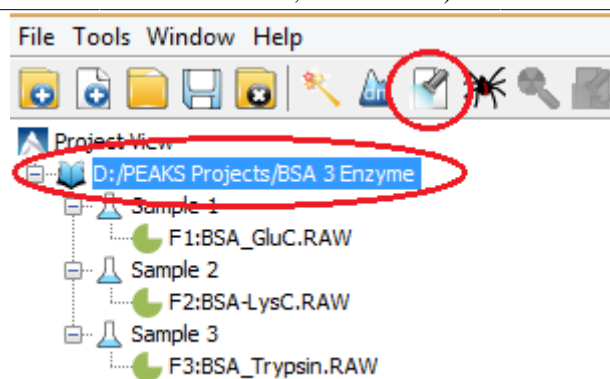
SPIDER is a dedicated search tool for finding novel peptides that are homologous to peptides in a given protein database.

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

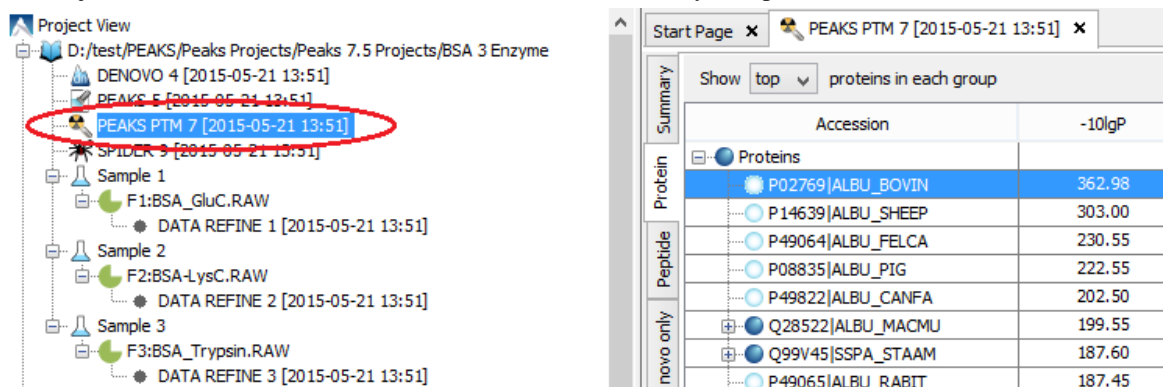
Refer to Chapter 4, *Project Wizard - Create a PEAKS Project, Load Data and Set up a Workflow* for details on creating a PEAKS project.

- Specify the PEAKS DB parameters in the pop-up dialog. 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 dialog to automatically run after PEAKS DB search. Click OK to start PEAKS search.

Note

For unrefined data, a prompt appears to specify data refinement parameters. Refer to Chapter 7, *Data Refinement* for details on refinement.

- Wait for the analysis to finish. Search times range from minutes to hours, depending on several factors including the dataset size and 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 dialog will appear:

The screenshot shows the 'PEAKS Search' dialog box with the 'Predefined parameters' dropdown set to 'default'. The 'Error Tolerance' section has 'Precursor mass' at 15.0 ppm using 'monoisotopic mass' and 'Fragment ion' at 0.05 Da. The 'Enzyme' section is set to 'Specified by each sample' with a 'New' button, allowing non-specific cleavage at one end of the peptide, and a maximum of 3 missed cleavages. The 'PTM' section lists 'Carbamidomethylation' (selected), 'Deamidation (NQ)', and 'Oxidation (M)', with buttons for 'Set PTM', 'Remove', and 'Switch type'. The 'Maximum allowed variable PTM per peptide' is set to 3. The 'Database' section has 'Select database' with 'Swiss-Prot' selected and a 'View' button, and 'Paste sequence' with 'Taxa' set to 'Bos taurus (cow)' and a 'Set/View taxa...' button. The 'De Novo Tag Options' section has 'Available de novo tags' set to 'de novo with current parameter'. The 'General Options' section has three checked checkboxes: 'Estimate FDR with decoy-fusion.', 'Find unspecified PTMs and common mutations with PEAKS PTM' (with an 'Advanced Setting' button), and 'Find more mutations with SPIDER'. At the bottom are 'Prev', 'OK', 'Cancel', and 'Help' buttons.

Error Tolerance. PEAKS search uses two levels of mass error tolerance: precursor mass and fragment ion mass.

Precursor mass Precursor mass tolerance. PEAKS Search identifies peptides within the precursor mass tolerance. Precursor mass tolerance can be specified in either Daltons 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 the "Predefined parameters".

Note

When the samples in a project are generated by different instruments and "default" is chosen from the Predefined parameters drop-down list, 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, and the number of missed cleavage sites allowed in a peptide.

Enzyme Selection Select the enzyme used for protein digestion from a drop-down list. If the enzyme used for protein digestion is not in the list, a new enzyme can be added. Please refer to Chapter 8, 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 enzyme 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 length up to 65 amino acids. Its usage is recommended when the digestion enzyme exhibits a high degree of non-specificity.

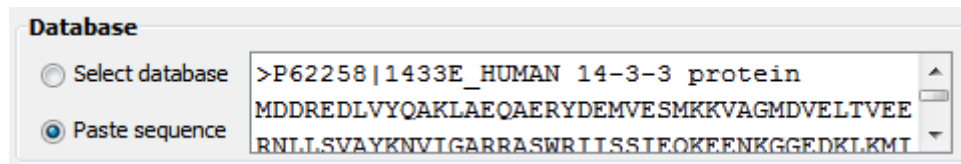
PTM. To select PTMs for PEAKS DB search, click the Set PTM... button to open the “PTM Options” dialog. PEAKS DB search supports a limited number of variable PTMs. Refer to Chapter 8, Section 2.3, “PTM” for details.

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 with 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 space and numbers in a sequence. A maximum of 1000 protein sequences can be pasted.

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 PEAKS PTM search after 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 the natural modifications and mutations in

the Unimod database. Refer to Chapter 17, Section 2.2, "PTM Configuration" for more information on built-in PTMs in PEAKS.

Alternatively, PEAKS PTM can search with a list of preferred modifications. In "Advanced Settings", preferred modifications from the "PTM Options" dialog 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 result 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.

PTM

☒ Search with 485 built-in modifications

☐ Search with preferred modifications:

- Carbamidomethylation
- Deamidation (NQ)
- Oxidation (M)

Set PTM

Remove

Switch type

Maximum allowed variable PTM per peptide 3

Filter Options

Perform PTM search on spectra satisfying the following condition:

De novo ALC (%) score greater than: 15 [recommend 15%](#)

OK

Find more mutations with SPIDER. Select this option to enable a SPIDER search. SPIDER performs homology search 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 list, select 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 list.

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:** 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:** shows a list of protein identifications. This view also visualizes protein sequence coverage and helps with protein characterization.
- **Peptide:** shows a list of peptide identifications. This view also provides spectrum annotation and other detailed information for peptide-spectrum matches.
- **De novo only:** shows a list of quality peptide sequences detected by *de novo* sequencing that remain unidentified by the PEAKS database search.
- **LC-MS:** 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:

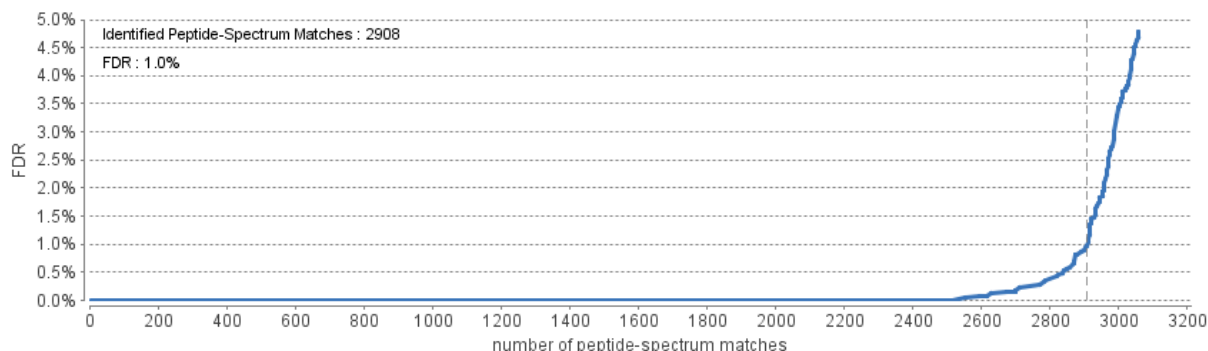
1. **Notes:** In the header of the report, information such as a description of the experiment can be entered. The notes can be edited by clicking the Notes button.
2. **Result Statistics:** Key statistics are shown in the figures and tables. The figures provide important information of target-decoy 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 filtering PSMs by peptide -10lgP score. By lowering the score threshold, more PSMs are kept in the filtered result. In the meantime the FDR

increases, because more false positives are being kept as well. 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 to be <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 page 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.

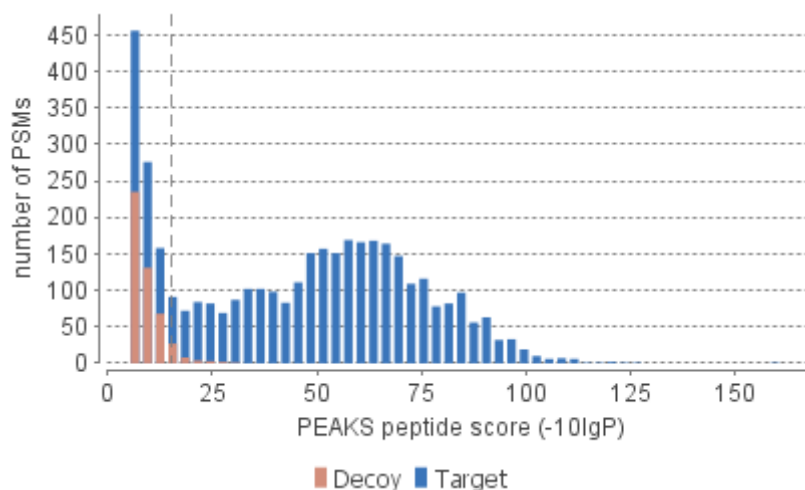
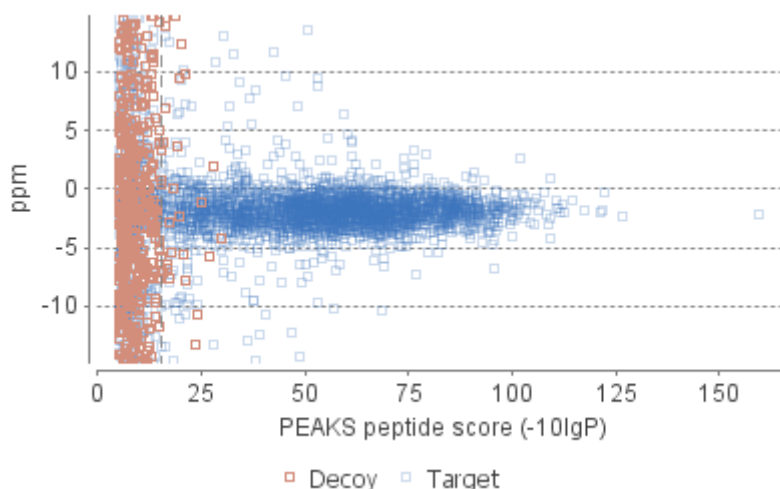


Figure 2(b) is a scatterplot showing the peptide score versus the precursor mass error in ppm for all PSMs. This figure is the most useful for high resolution instruments. The PSMs with high scores should be centered near the mass error of 0, and the PSMs with low scores are usually scattered to larger 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 page help 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 $-10\lg P$ score above the current peptide filtering threshold.

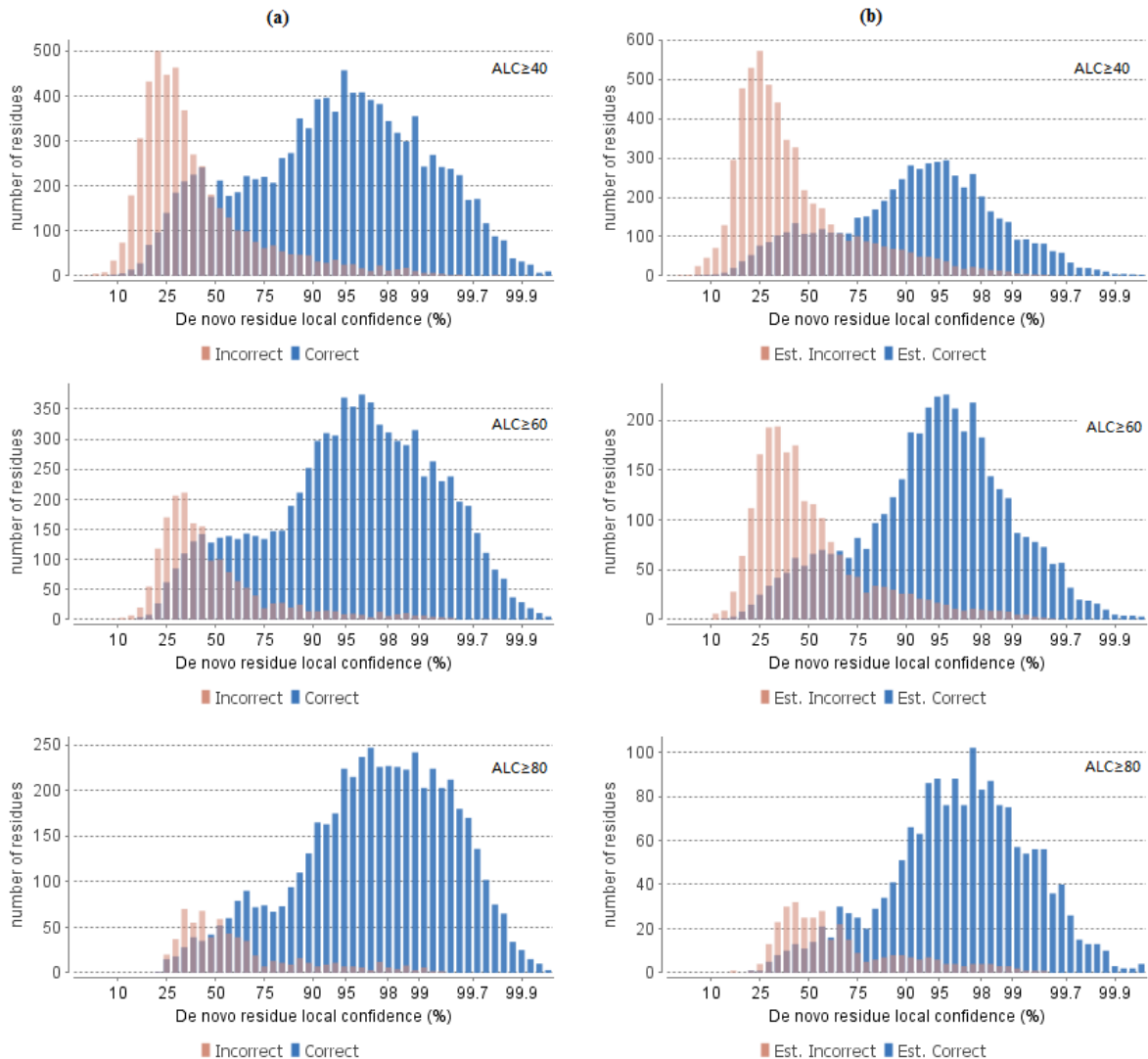
"*De novo* only" sequences are crucial for a complete proteomic analysis. However, it is often necessary to remove the low quality sequences. "*De novo* only" peptides can be filtered by *de novo* ALC score, which is the average local confidence score of residues in the *de novo* sequence.

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 is 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 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.

Peptide, PTM, and Mutation
Identification (PEAKS DB,
PEAKS PTM, and SPIDER)

Table 1. Statistics of data.

# of MS Scans	4751
# of MS/MS Scans	5152

Table 2. Result filtration parameters.

Peptide -10lgP	≥10.1
Protein -10lgP	≥20
Proteins unique peptides	≥0
De novo ALC Score	≥50%

Table 3. Statistics of filtered result.

Peptide-Spectrum Matches	2705
Peptide Sequences	1000
Protein Groups	29
Proteins	40
Proteins (#Unique Peptides)	15 (>2); 1 (=2); 13 (=1);
FDR (Peptide-Spectrum Matches)	1.0%
FDR (Peptide Sequences)	2.4%
FDR (Protein)	7.5%
De Novo Only Spectra	157

Table 4. PTM profile.

Name	ΔMass	Position	#PSM	-10lgP	Area	AScore
Carbamidomethyl	57.02	C	1330	68.55		1000.00
Carbamidomethyl	57.02	DEHK,N-term	580	62.99	9.55E6	87.62
Deamidation	.98	NQ	85	58.00	3.13E5	1000.00
Methyl ester	14.02	DE,C-term	58	42.98	7.32E7	149.94
Dehydration	-18.01	DSTY	40	49.86	1.31E6	0.00
Dethiomethyl	-48.00	M	32	55.96	4.56E6	1000.00
Carbamylation	43.01	K,N-term	26	61.23	7.06E7	1000.00
Oxidation	15.99	M	26	61.47	4.29E6	1000.00
Ammonia loss	-17.03	N	19	53.43	6.92E5	1000.00
Sodium	21.98	DE,C-term	14	36.88	2.61E6	1000.00
Pyro-glu from Q	-17.03	N-term	14	62.12	1.05E7	1000.00
Formylation	27.99	K,N-term	10	42.98	7.32E7	23.15
Ala->Thr	30.01	T	10	30.31	4.96E6	1000.00
Acetylation	42.01	N-term	8	41.38	1.29E5	1000.00
Ala->Glu	58.01	E	8	39.37		1000.00
Hex	162.05	KT,N-term	8	33.76	3.13E6	11.06
Hydroxylation	15.99	KPR	7	27.89	1.08E5	45.01
Ubiquitin	114.04	KST	7	44.34		19.05
Carboxylation	43.99	DK	6	41.62	8.53E4	178.84
Acetylation	42.01	K	5	48.34	1.1E7	1000.00

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

- **Peptide Sequences** (Table 3): This is 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 are differentiated only by leucine(L) and isoleucine(I), they are counted as one. 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 its -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, the number of PSMs containing each modification along with -10lgP score, area and the PTM AScore of the best PSM.

3.2.5. Experiment Control

Figures 4(a) and 4(b) in the Summary page 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 $10^6 \times (\text{precursor mass} - \text{peptide mass}) / \text{peptide mass}$. Figure 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.

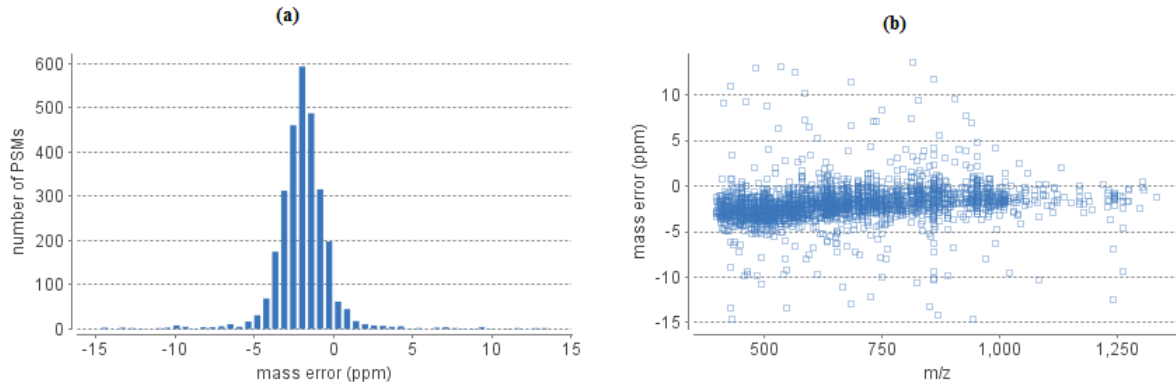


Table 5 shows the number of identified peptides in each sample with the number of missed cleavages, indicating the enzyme digestion efficiency.

Table 5. Number of peptides by number of missed cleavages

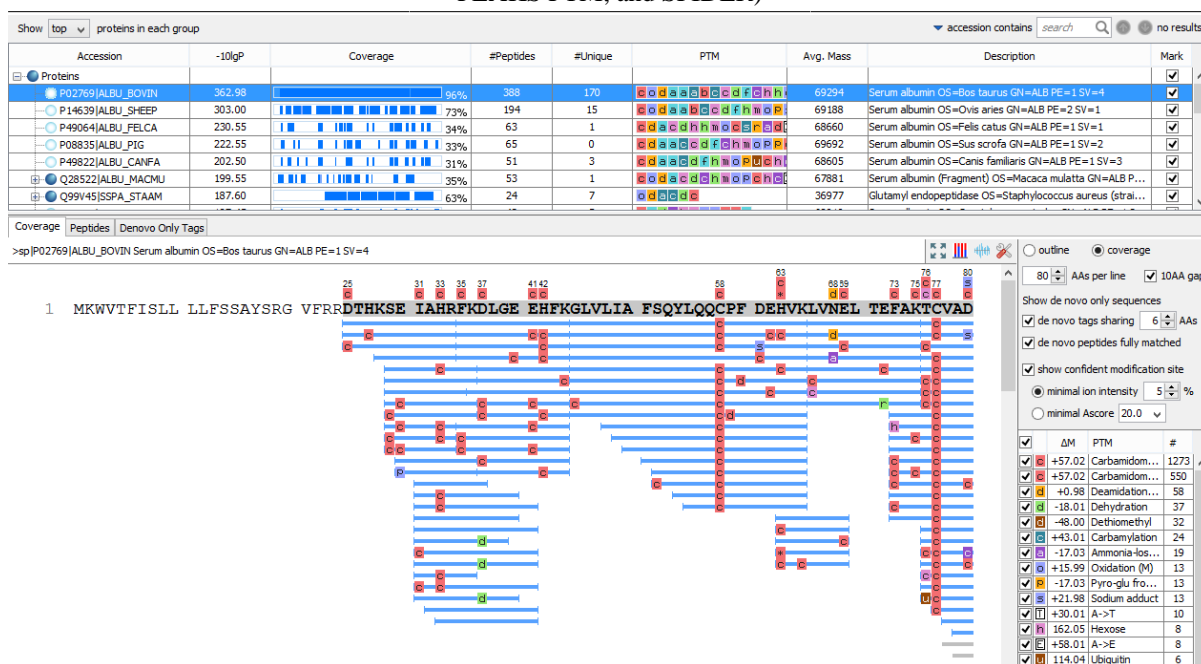
Missed Cleavages	0	1	2	3	4+
Sample 1	78	56	23	16	0
Sample 2	62	38	15	4	0
Sample 3	107	55	15	6	0

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:** Lists protein identifications that have been filtered by the current settings in Summary view
- **Coverage:** Visualizes the protein sequence coverage of the selected protein, showing the mapping of supporting peptides and *de novo* tags to the protein sequence.
- **Peptides:** Lists supporting peptides mapped to the selected protein.
- **De novo Tags:** Lists "*de novo* only" peptides mapped to the selected protein.

Peptide, PTM, and Mutation Identification (PEAKS DB, PEAKS PTM, and SPIDER)



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 list, 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; "first" shows only the first of the top proteins in the group.

Note

A supporting peptide is considered as of high-confidence if it has a peptide -10lgP score above the filtering threshold.

The columns in the protein table are:

- **Accession:** The accession number of the protein as in FASTA database.
- **-10lgP:** The protein confidence score.
- **Coverage:** The percentage of the protein sequence covered by supporting peptides. The coverage is visualized by a color 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 uniquely mapped to the 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:** Protein mass calculated using average mass.
- **Description:** Protein's header information as in FASTA database.
- **Mark:** 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 visualizes the mapping of supporting peptides and *de novo* tags to the protein selected in the protein table. The coverage view 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 the 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 which pass -10LogP threshold 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 (-10LogP is from the best PSM and AScore is the best AScore from all PSMs). 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.

>sp|P02769|ALBU_BOVIN Serum albumin OS=Bos taurus GN=ALB PE=1 SV=4



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

☐ outline
 ☒ coverage

80 AAs per line
 ☒ 10AA gap

Show de novo only sequences

☒ de novo tags sharing 6 AAs

☒ de novo peptides fully matched

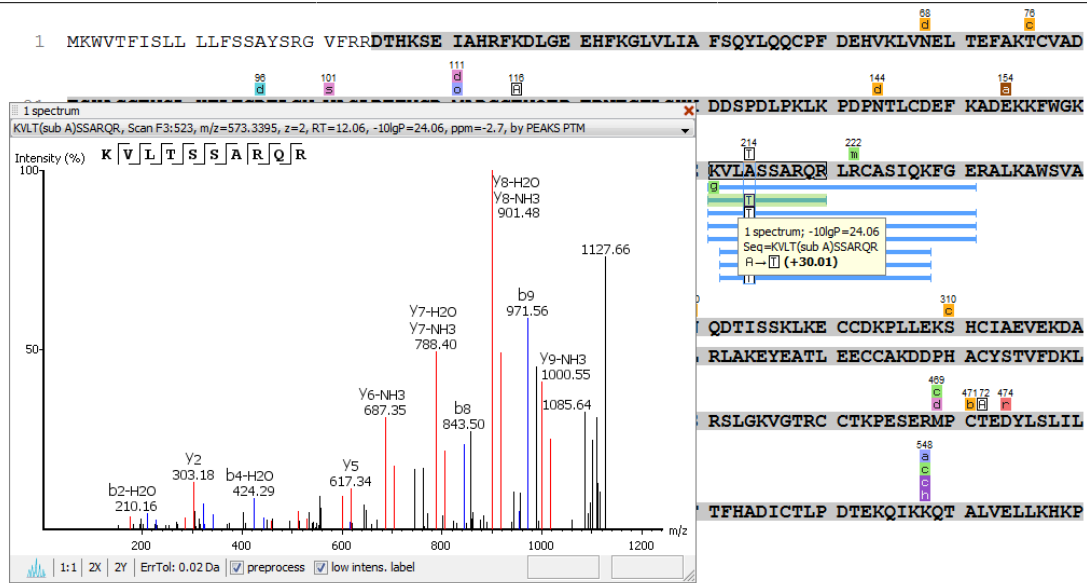
☒ show confident modification site

☒ minimal ion intensity 5 %

☐ minimal Ascore 20.0

<input checked="" type="checkbox"/>	ΔM	PTM	#
<input checked="" type="checkbox"/>	+57.02	Carbamidomethylation	1273
<input checked="" type="checkbox"/>	+57.02	Carbamidomethylation (DHKE,...	550
<input checked="" type="checkbox"/>	+0.98	Deamidation (NQ)	58
<input checked="" type="checkbox"/>	-18.01	Dehydration	37
<input checked="" type="checkbox"/>	-48.00	Dethiomethyl	32
<input checked="" type="checkbox"/>	+43.01	Carbamylation	24
<input checked="" type="checkbox"/>	-17.03	Ammonia-loss (N)	19
<input checked="" type="checkbox"/>	+15.99	Oxidation (M)	13
<input checked="" type="checkbox"/>	-17.03	Pyro-glu from Q	13
<input checked="" type="checkbox"/>	+21.98	Sodium adduct	13
<input checked="" type="checkbox"/>	+30.01	A->T	10
<input checked="" type="checkbox"/>	162.05	Hexose	8
<input checked="" type="checkbox"/>	+58.01	A->E	8
<input checked="" type="checkbox"/>	114.04	Ubiquitin	6
<input checked="" type="checkbox"/>	+31.99	Dihydroxy	5
<input checked="" type="checkbox"/>	-30.01	T->A	5
<input checked="" type="checkbox"/>	+27.99	Formylation	4
<input checked="" type="checkbox"/>	+58.01	Carboxymethyl (KW, X@N-term)	4
<input checked="" type="checkbox"/>	+37.95	Replacement of 2 protons by ...	4
<input checked="" type="checkbox"/>	-58.01	D->G	4
<input checked="" type="checkbox"/>	-58.01	E->A	4
<input checked="" type="checkbox"/>	-18.01	Pyro-glu from E	3
<input checked="" type="checkbox"/>	+43.99	Carboxylation (DKW)	3
<input checked="" type="checkbox"/>	129.04	Monoglutamyl	3
<input checked="" type="checkbox"/>	+42.01	Acetylation (N-term)	2
<input checked="" type="checkbox"/>	+58.01	Carboxymethyl	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.

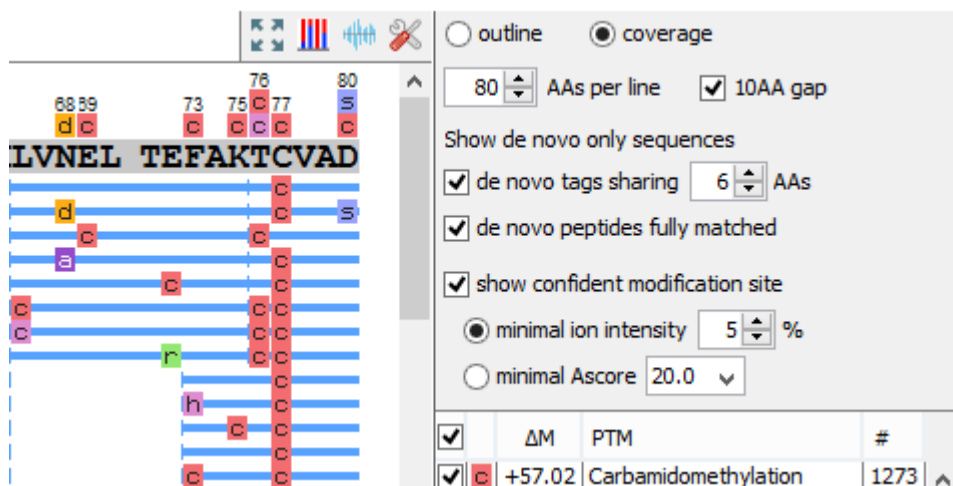


- **Coverage mode:** In coverage mode, all the supporting peptides and matched *de novo* tags are shown.



- **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.
- ***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 its full sequence matched in the protein sequence. This control is especially useful for showing short but fully matched *de novo* peptides.
- **show confident modification site.** Only confident modifications and mutations are displayed as icons above the protein sequence
 - **minimal ion intensity:** when chosen, 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.
 - **minimal AScore:** when chosen, a modification at a protein position is considered confident if the PTM AScore for that modification at that position is above the specified threshold.

- **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 four icons:



- **Full screen** . Left clicking this icon shows the protein coverage in full screen mode. Click to return from full screen mode.
- **PTM profile** . Left clicking this icon shows the PTM profiling of the selected protein in a dialog. Refer to Section 3.3.3, “PTM Profiling” for details on PTM profiling.
- **Peptide mapping** . Left clicking this icon shows the supporting peptide and *de novo* only peptide mapping of the selected protein in a dialog. Refer to Section 3.3.4, “Peptide Mapping” for details on peptide mapping.
- **Tool box** . The tool box contains the following tools:
 - Copy template protein. Copies the original database protein into the system clipboard
 - Copy mutated protein. Copies the protein with detected mutations into the system clipboard
 - Save image as... Saves the protein coverage as an image
 - Coverage statistics Shows protein coverage statistics such as protein length, percentage of peptide coverage, number of peptides, *de novo* tags and PTMs.
 - NCBI BLAST search Opens the NCBI BLAST search page in browser
 - NCBI Entrez search Opens the NCBI Entrez search page in browser
 - Multiple sequence alignment. Performs multiple sequence alignment for the selected proteins

3.3.3. PTM Profiling

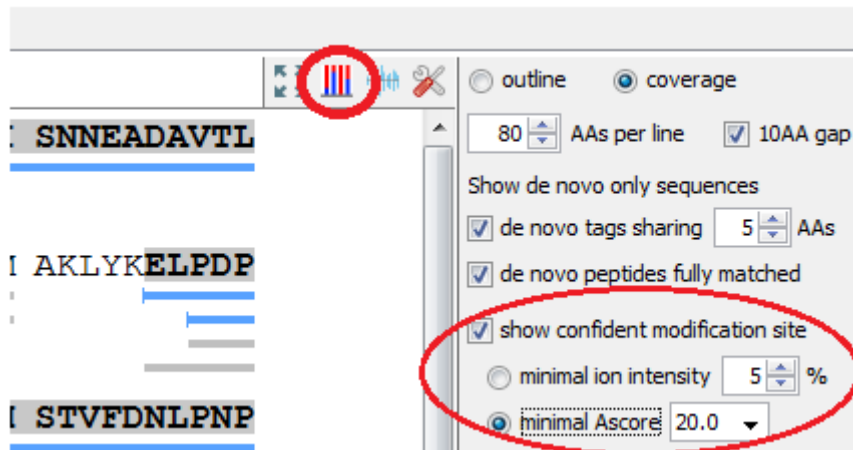
The PTM Profile tool provides the ability to compare and quantify detected modification sites on proteins in relation to unmodified sites. In many cases a PTM only occurs on a portion of the sample, leaving the remainder unmodified. With this feature, PEAKS determines the protein's modification changes by comparing the area from the LC-MS of the confidently modified peptide feature to the area from associated unmodified (native) or non-confidently modified peptide feature, and 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.

Summary	Show top proteins in each group		
	Accession	-10lgP	Coverage
	Proteins		
Protein	P02769 ALBU_BOVIN	52.98	96%
	P14639 ALBU_SHEEP	303.00	73%

2. Set minimal ion intensity or minimal AScore parameter threshold for confident modification site determination. 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 dialog will appear to display the PTM profiling.



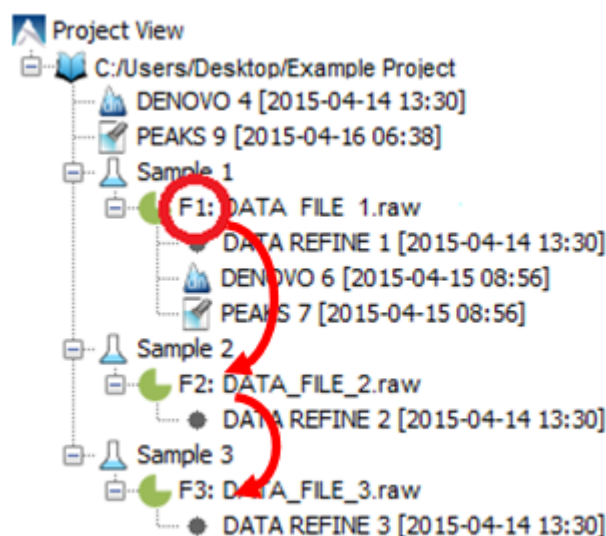
The input selection, result display and result exporting options are combined together in the PTM profile dialog.

- **Inputs.** The inputs for the PTM profiling can be set using the drop-down lists at the top-left corner of the dialog.
- **Protein.** The accession ID of the selected protein to be analyzed for PTM profiling.

Note

The selected protein can not be changed within the dialog. To run profiling on another protein, close the dialog, select the protein of interest from the protein table and run the profiling by clicking  button.

- **Fraction.** Data source used for the analysis. The “F#” represents the fraction the protein is from, 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 done by selecting the second fraction from the second fraction selection drop down list.

- **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 full 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 whereas the y-axis indicates the percentage (%) of modified or unmodified peptide features relative to the total abundance. Blue bar represents unmodified portion and red bar represents the modified portion. For some position, except for the specified modification, some other modification may exist. Thus, the modified and unmodified portion 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 Save to Image... button to save the displayed graph as a PNG file.
- **Text format export.** Click 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 the selected protein may be exported for a specific fraction, or across all fractions. In addition, profiles on all proteins for the selected modification may also be exported, again for a specific fraction or across all fractions. The exported file contains all necessary information to construct the PTM profile graph.

The PTM profile data are grouped together by fractions, 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 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:

- **Protein position:** Position of the modified amino acid in the protein sequence.
- **Peptide position:** Position of the modified amino acid in the peptide sequence.
- **AScore:** The probability of correct modification site localization. The “AScore” column shows the PTM AScore obtained from the best PSM.

Note

The “Best AScore” column shows the best PTM AScore associated with the corresponding peptide

- **Ion intensity:** intensity of fragment ions supporting the modified residue with the minimal peak intensity above the specified threshold.
- **Total Area:** total area of the features associated with the corresponding peptide

Note

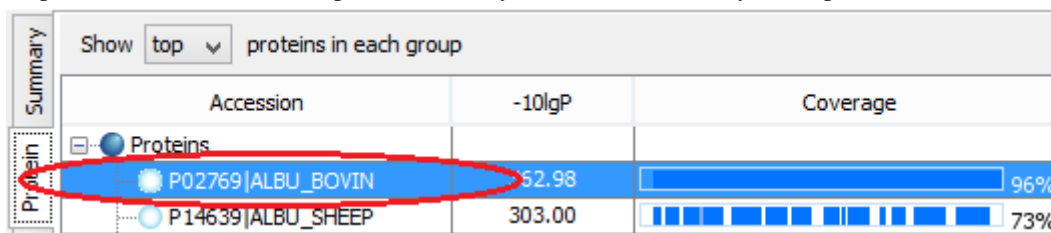
scan, m/z, RT, z, -10lgP values are from the best PSM for the corresponding peptide.

3.3.4. Peptide Mapping

A peptide map is often referred to as a “fingerprint” of a protein. It is a powerful tool which 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 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.

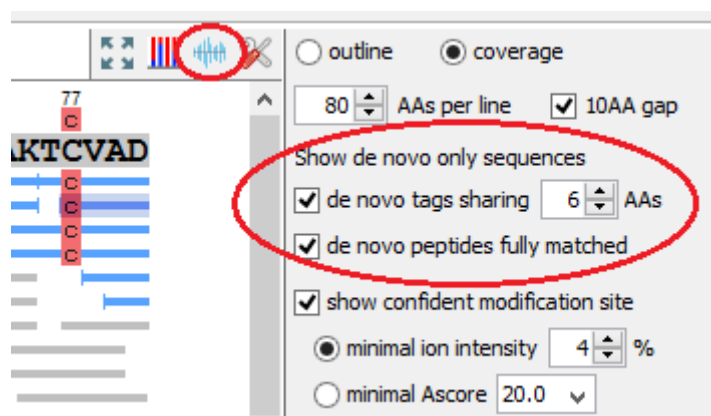
Follow the steps mentioned below to generate the peptide mapping:


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



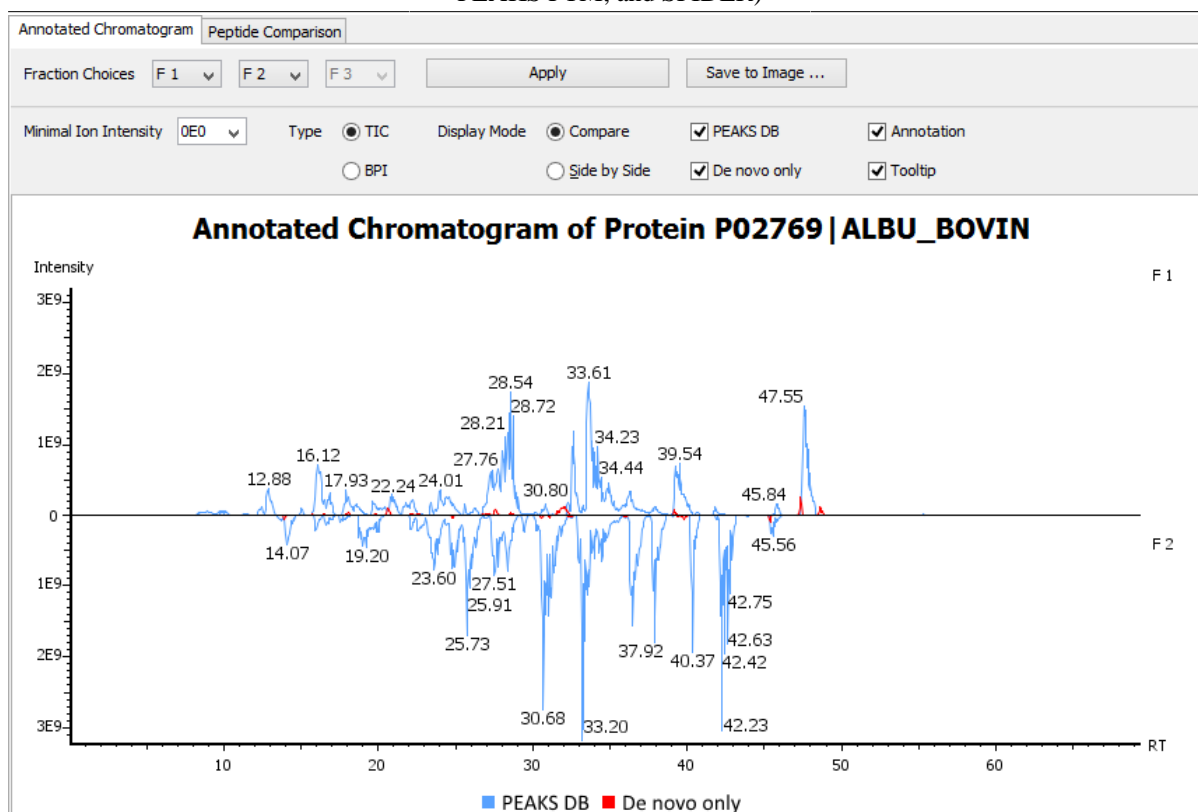
Accession	-10lgP	Coverage
P02769 ALBU_BOVIN	52.98	96%
P14639 ALBU_SHEEP	303.00	73%

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 dialog will appear to display the peptide mapping.

Peptide, PTM, and Mutation
Identification (PEAKS DB,
PEAKS PTM, and SPIDER)



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.
- **Inputs.** The inputs for the annotated chromatogram can be set using the drop-down lists at the top-left corner of the dialog.
- **Fraction Choices.** Data source used for peptide mapping. The “F#” represents the fraction the protein is from, 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 the fractions, up to three fractions, can be done by selecting the second and third fraction from the corresponding drop down list.

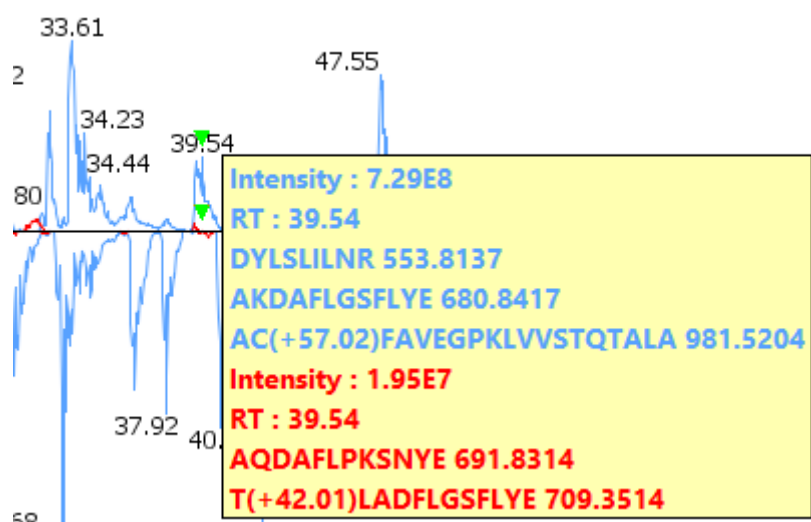
- **Minimal Ion Intensity:** A filter used to set an intensity threshold, where peptide features below this value will not be included in the display.
- **Type - TIC or BPI:** Configures the display to view the Total Ion Chromatogram (TIC) or Base Peak Intensity (BPI).
- **Display Mode:** PEAKS offers 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 search are highlighted in the annotated chromatogram.
- **De novo only:** Highlights the peptide features associated with *de novo* only peptides (see Section 3.5, “*De Novo* Only View”).

Note

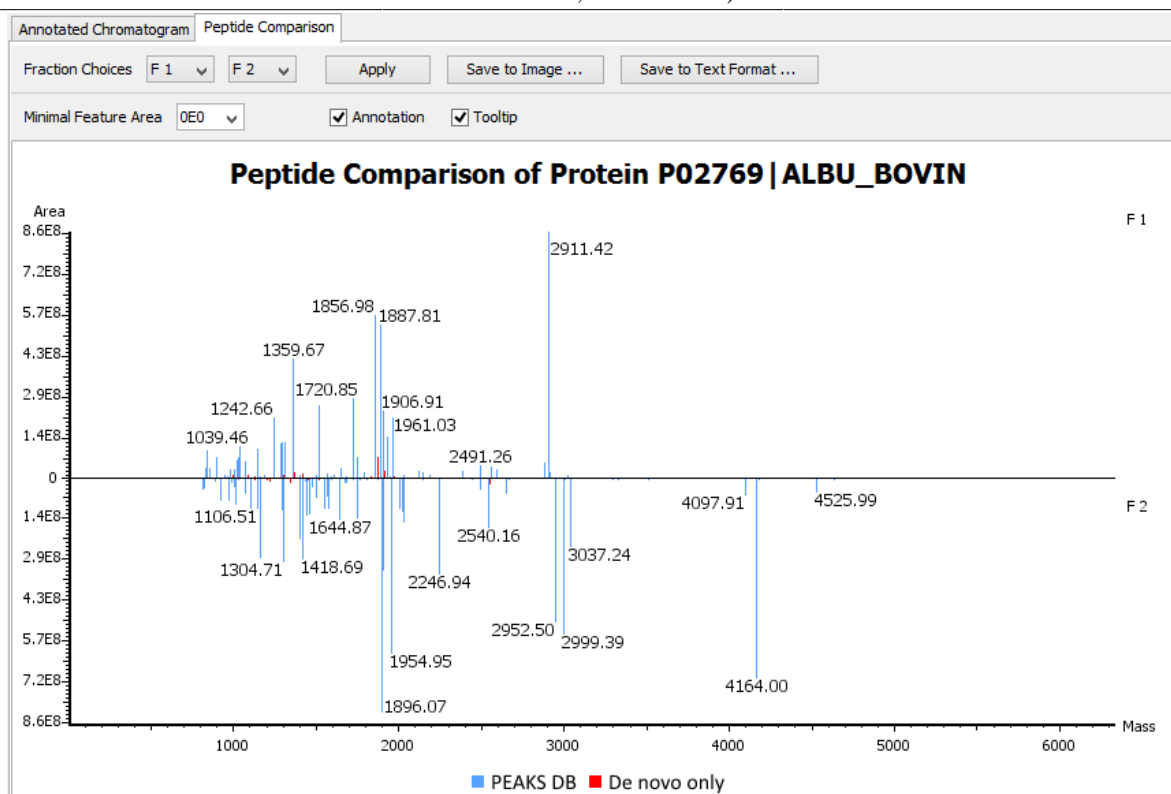
De novo only peptides displayed in the peptide map may be filtered using the *de novo* only peptide selection filters in the control panel of the coverage pane.

- **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, intensity (TIC or base peak), and peptide sequences found at that particular retention time.



- **Annotated Chromatogram Chart:** The chart gives a visualization of the peptide map for the selected data. It provides a few convenient ways to zoom and navigate the chromatogram
- **Zoom to an RT region** - click the desired start retention time and drag horizontally to the desired end retention time, release the mouse button.
- **Zoom in/out smoothly** - place the mouse pointer at a particular RT value (right below the x-axis line), scroll the mouse wheel button.
- **Increase/decrease the ion intensity** - place the mouse pointer on the chromatogram.
- **See the whole chromatogram** - double click on the chromatogram.
- **Pan to an RT region:** Place mouse cursor on chromatogram, right click, hold and drag for panning to an RT region.
- **Result Export.** The annotated chromatogram chart can be exported as an image in PNG format. Click Save to Image... button to save the displayed chart.
- **Peptide Comparison:** Another approach 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.

Peptide, PTM, and Mutation
Identification (PEAKS DB,
PEAKS PTM, and SPIDER)

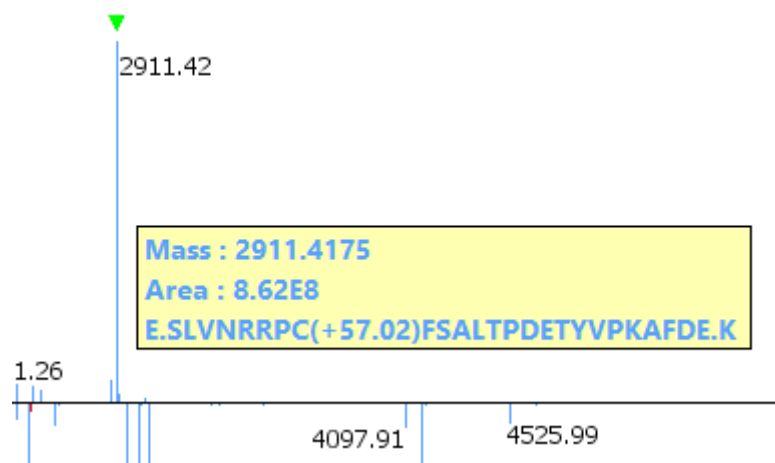


- **Inputs.** The inputs for the peptide comparison can be set using the drop-down lists at the top-left corner of the dialog.
- **Fraction Choices.** Data source used for peptide mapping. The “F#” represents the fraction the protein is from, where # refers to the ID of the fraction, as indicated in 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 at most 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 mass 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 peptide mass, total area for that peptide, and the peptide sequence.



- **Peptide Comparison Chart:** The 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 the bar represents the total area of the associated peptide features in the specified fraction. The chart provides similar convenient ways to zoom 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 Save to Image... button to save the displayed graph as a PNG file.
 - **Text format export.** Click Save to Text Format... button to export the peptide comparison data to a text file in CSV format. The exported file contains all 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:

- **-10lgP or ALC:** the significance score of the peptide identified by database search or by *de novo* sequencing for *de novo* only peptides.
- **Feature Areas:** 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:** 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 are calculated as well.

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 for the corresponding peptide.

3.3.5. Supporting Peptides

The “Peptides” tab shows a table of supporting peptides for the protein. This table is similar to the peptide table in the Peptide View, except that three additional columns are added:

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

- **End:** Shows the peptide's ending position (inclusive) in the protein.
- **(Checkbox):** Controls whether a peptide is shown in the protein coverage as a blue bar.
- **Unique:** Shows whether the peptide is a unique supporting peptide to the protein group.

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 color 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. 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 "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

Peptide Table shows the filtered peptide identification result. Each row in the table is a peptide identification represented by its highest-scoring PSM. The table can be sorted by any column except the AScore 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 about 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 the peptide. Other PSMs can be examined after selecting a peptide in the table. See Section 3.4.2, "Peptide-Spectrum Matches" for details.

1 - 1000 of 4371												
	Peptide	-10lgP	Mass	Length	ppm	m/z	RT	Area	Scan	#Spec	Accession	Found By
1	IWNHVSILNHLVSKNINR	57.14	2556.3926	22	0.6	640.1058	106.03	6.0566	20003	10	Q14697 GANAB_HUMAN	PEAKS DB
2	DVHNHGLYVHMATADGLR	56.71	2144.0474	19	0.1	715.6898	85.06	2.4866	15900	5	P09936 UCHL1_HUMAN:Q60HC8 UCHL1_MACFA:...	PEAKS DB
3	VDDK(+4.03)VNHFILFNNVDGHLDELG...	56.35	2985.5017	25	-1.0	996.1735	91.41	1.2466	17145	10	P09936 UCHL1_HUMAN:Q60HC8 UCHL1_MACFA:...	PEAKS DB
4	VDDKYNFHFILFNNVDGHLDELGR	55.79	2975.4565	25	0.5	992.8266	91.47	1.1966	17158	8	P09936 UCHL1_HUMAN:Q60HC8 UCHL1_MACFA:...	PEAKS DB
5	SLTGIVNAQALTSAPSHKIPWIGLAELG...	55.23	3552.8762	34	-2.6	889.2240	113.83		21536	4	O43179 SERP1_HUMAN:Q56P1 SERP1_PANTR	PEAKS DB
6	VADPDHDTGFTYH(+79.97)IATR	54.80	2222.3634	19	0.4	1112.4894	54.98	7.9965	10019	10	P63088 MKO1_RAT:P63088 MKO1_MOUSE:P2848...	PEAKS DB
7	FLALAVAGGVNSALVNDAGHR	53.93	2370.2444	24	0.4	791.0891	104.43	2.4166	19696	2	Q3T165 PHB_BOVIN:P67779 PHB_RAT:P35232 P...	PEAKS DB
8	ELFDPSY(+79.97)VNNQDKAR	53.72	2201.9993	18	0.3	1102.0072	75.55	9.0365	14041	4	P29353 SHC1_HUMAN:Q5R7W7 SHC1_PONAB	PEAKS DB
9	DVHNHGLYVHMATADGLR(+6.02)	53.58	2150.0674	19	0.4	717.6967	85.18	2.4666	15923	4	Q14697 GANAB_HUMAN	PEAKS DB
10	QFGVILTSAGIMDHHEAR	53.50	2221.0837	20	0.8	741.3691	92.93	1.5966	17447	2	Q90YQ8 RS15A_ICTPU:Q5R938 RS15A_PONAB:...	PEAKS DB
11	K(+8.01)SDTYVC(+57.02)MISFAHNAVA...	53.47	2254.1208	20	0.3	752.3811	83.54	3.72E5	15610	2	Q61598 GDIH_MOUSE:P50399 GDIH_RAT:P5039...	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.

Tip

Right click on a peptide to show a pop-up menu, and select Show Spectrum in LC/MS. This would jump to the LC/MS view, zoom in to the relevant LC/MS region and mark the MS/MS spectrum of the best PSM.

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:** Peptide -10lgP score. The score indicates the scoring significance of a peptide-spectrum match.
- **Mass:** Monoisotopic mass of the peptide
- **Length:** The number of amino acids in the peptide sequence.

- **ppm:** Precursor mass error, calculated as $10^6 \times (\text{precursor mass} - \text{peptide mass}) / \text{peptide mass}$.
- **m/z:** Precursor mass-to-charge ratio.
- **RT:** 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:** Scan number.
- **#Spec:** Number of spectra assigned to the peptide.
- **Accession:** List of accession number of the identified proteins containing this peptide. The list is sorted according to the protein score in descending order.
- **PTM:** Types and numbers of modifications present in the peptide shown in color-coded icons.
- **AScore:** List of the site localization scores of all modification sites in the peptide. The AScore measures the probability of correct modification site localization based on the presence and intensity of site-determining ions in MS/MS spectra. The scores are separated by semi-colon ";" and each score is represented by the single letter amino acid code followed by the position of the amino acid in the peptide sequence, the color-coded icon of the modification, and the AScore.

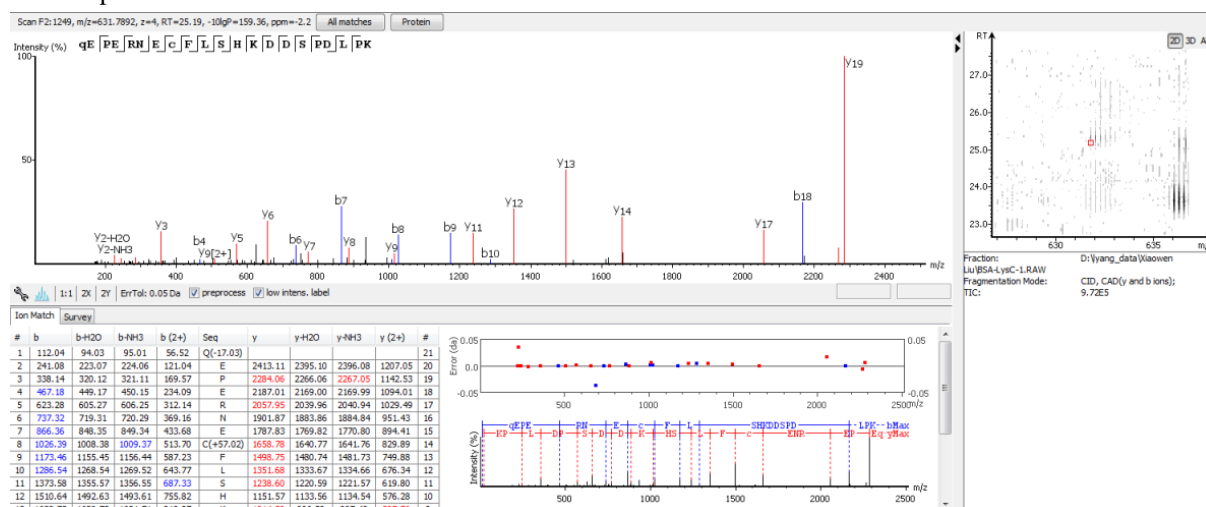
Note

The Maximum AScore is 1000. The AScores of the fixed PTMs are not displayed.

- **Found by:** The name of 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 Chapter 8, Section 3.2, "De Novo Peptide View".



3.5. De Novo Only View

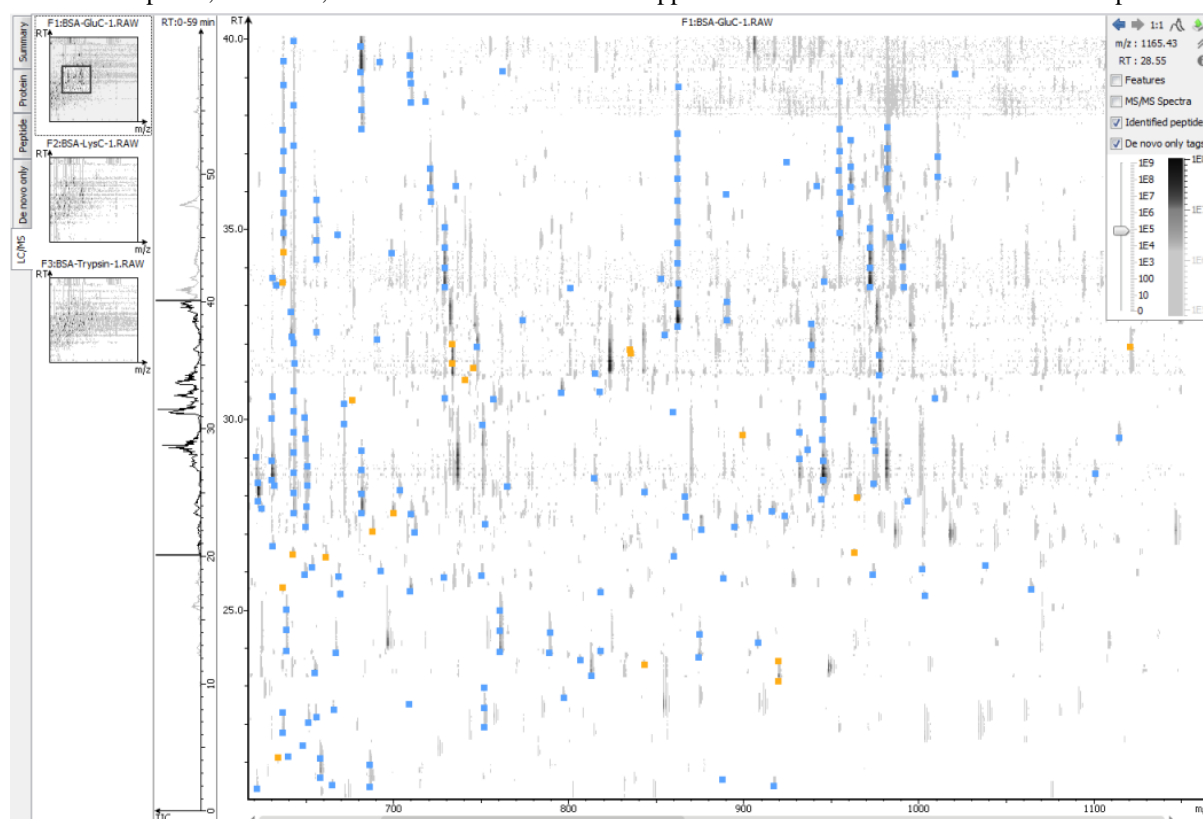
The De novo Only View displays the filtered "de novo only" sequences. A de novo sequence is "de novo only" if it is from an MS/MS spectrum that is not confidently assigned to any database peptide. The table shows the

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 Chapter 8, 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 Chapter 5, Section 4, "LC/MS View" for a list of supported interactions in the LC-MS heat map.



4. Filtering PEAKS Search Results

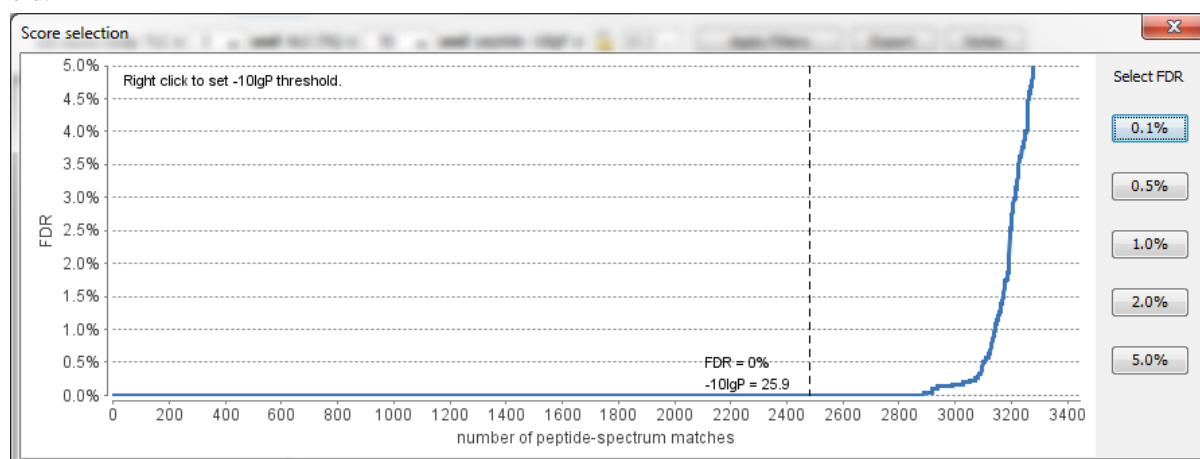
PEAKS Search results can be filtered to keep only the high quality peptides, proteins, and "de novo only" peptides. Filtering parameters can be specified using the control panel in the Summary View. Click the Apply Filters button to filter the results that will be updated accordingly in the Summary, 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 score. The peptide -10lgP threshold affects the results shown in Peptide and 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, and 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 needs to be set empirically. A score threshold set at 20 is a good choice to start with - the equivalent p-value will be 0.01 with a -10lgP score of 20.

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 score and 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 unique to a group of proteins (not found in other protein groups). 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 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 it 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 Summary View allows exporting of the filtered results into various formats. It provides the opportunity to share the report in HTML, use the result as a supplemental in a publication, or further validate and analyze the result in a third party software. Please refer to Chapter 16, 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.bioinform.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.

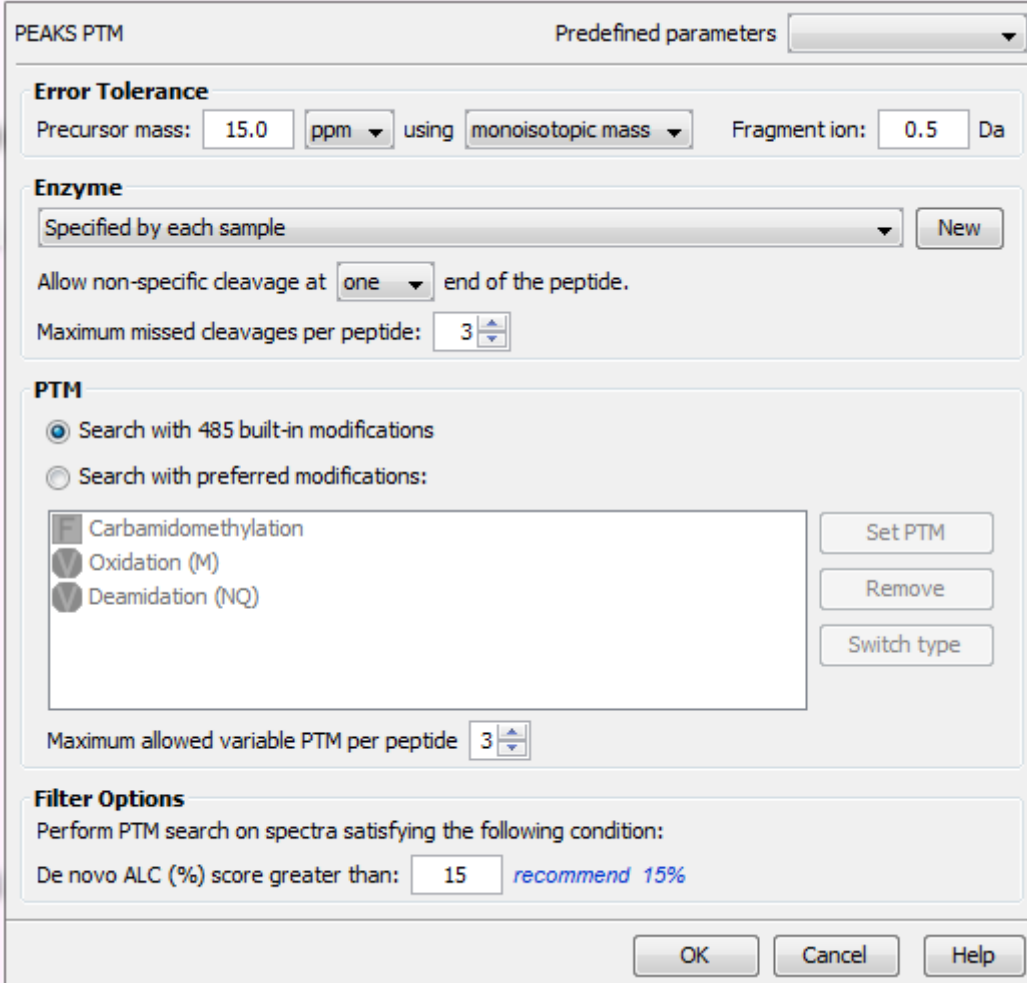
Note

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.

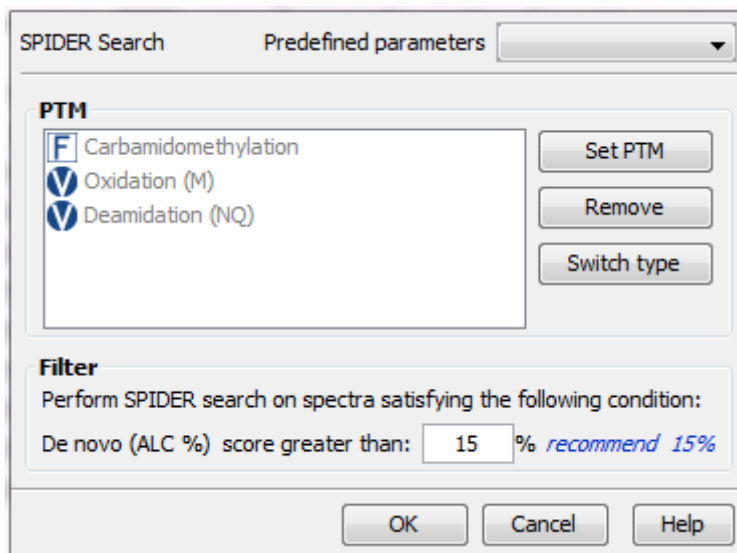


The screenshot shows the 'PEAKS PTM' configuration window. At the top, there is a 'Predefined parameters' dropdown menu. The 'Error Tolerance' section includes 'Precursor mass' (15.0 ppm) and 'Fragment ion' (0.5 Da). The 'Enzyme' section has a dropdown set to 'Specified by each sample' and a 'New' button. Below this, 'Allow non-specific cleavage at' is set to 'one' end of the peptide, and 'Maximum missed cleavages per peptide' is set to 3. The 'PTM' section has two radio buttons: 'Search with 485 built-in modifications' (selected) and 'Search with preferred modifications:'. Under the preferred modifications, a list contains 'Carbamidomethylation', 'Oxidation (M)', and 'Deamidation (NQ)'. To the right of this list are buttons for 'Set PTM', 'Remove', and 'Switch type'. Below the list, 'Maximum allowed variable PTM per peptide' is set to 3. The 'Filter Options' section states 'Perform PTM search on spectra satisfying the following condition:' and 'De novo ALC (%) score greater than: 15', with a 'recommend 15%' link. At the bottom are 'OK', 'Cancel', and 'Help' buttons.

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:



The SPIDER Search dialog box features a 'Predefined parameters' dropdown menu. It is divided into two main sections: 'PTM' and 'Filter'. The 'PTM' section contains a list of modifications: Carbamidomethylation (selected with a blue 'F' icon), Oxidation (M) (selected with a blue 'V' icon), and Deamidation (NQ) (selected with a blue 'V' icon). To the right of this list are three buttons: 'Set PTM', 'Remove', and 'Switch type'. The 'Filter' section contains the text 'Perform SPIDER search on spectra satisfying the following condition:' followed by 'De novo (ALC %) score greater than: 15 % recommend 15%'. At the bottom of the dialog are 'OK', 'Cancel', and 'Help' buttons.

Note

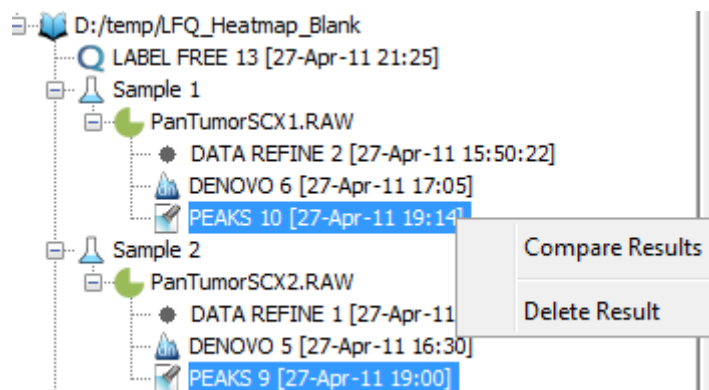
For users of previous versions of PEAKS, SPIDER in the current version runs in the mode previously called Homology Match.

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 Chapter 8, 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

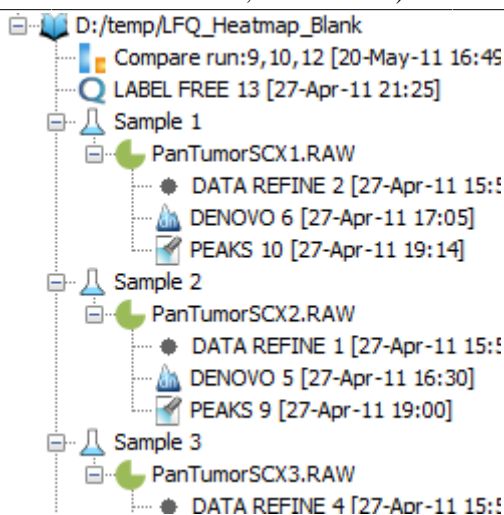
It is possible to compare up to three search results in the same project. To compare results, press Ctrl to select two or three result nodes, right-click, and choose Compare Results from the pop-up menu. PEAKS will then start to compare the results.



7.1. Comparison Result

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

Peptide, PTM, and Mutation Identification (PEAKS DB, PEAKS PTM, and SPIDER)



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 which results the peptide is identified from. 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 peptide table in a typical comparison result is shown below:

Peptide Sequence	Cover Map	PEAKS 9 [27-Apr-11 19:00]					PEAKS 10 [27-Apr-11 19:14]					PEAKS 12 [27-Apr-11 19:54]				
		M/Z	RT	-10lgP	Z	MULTI-HIT	M/Z	RT	-10lgP	Z	MULTI-HIT	M/Z	RT	-10lgP	Z	MULTI-HIT
VFLNVIR							495.29135	39.7717	44.76	2	true	542.26984	28.4963	47.67	2	true
SPFEVYDK												611.82666	35.0541	48.84	2	true
PSMVVQGVNK												633.822	43.2472	41.94	2	false
ALDTMFPDVK												865.50085	82.1928	61.86	2	true
EVDSYLPVLDLK		865.5011	82.2145	60.6	2	false						690.4143	68.3736	49.89	2	false
AEELGLFILGVIR		690.4142	69.7953	47.29	2	false	785.90485	26.8903	85.12	2	true	785.90533	25.4252	71.49	2	false
GAGTGLGLAIEGSPSEAK		785.9052	26.8499	72.26	2	true	651.3209	19.5598	44.54	2	false	651.3207	18.5345	64.55	2	true
ISVYNEATGGK		651.3205	19.51	35.03	2	false	630.3677	33.0669	45.33	2	false	630.3655	31.5186	39.08	2	false
LQTSVLSQGLR		630.368	33.1726	36.7	2	false	760.92554	47.5334	48.38	2	true	760.9249	46.4147	55.76	2	true
VELALPSVLPQVVK		760.9248	47.423	48.51	2	false	953.9895	73.1017	66.44	2	false	953.9897	73.3267	71.69	2	true
ETDILLDSLVISFGNR												695.3538	37.3655	53.56	2	false
GYSPYFINTSK												716.8505	35.3759	15.28	2	true
ELALDNRSENGK		716.8508	36.088	16.28	2	true	716.8515	35.6881	13.25	2	true	473.7618	16.3971	32.03	2	false
LTVLSEER												466.7362	8.1263	46.34	2	true
ILLEGESR							466.73563	8.4792	32.5	2	true					
FSEAEVVPVVK							650.86554	49.3616	19.02	2	false					
IMAKSTLK												502.81247	41.7395	12.43	2	false
DLTNVQNLQK												586.80884	17.3177	21.29	2	false
LAADIDASATNESSWNGR												967.94336	53.1444	5.04	2	false
IIAPVGSVPVEDNEK		759.3939	34.552	23.71	2	true	759.39325	34.4955	30.52	2	true					
SLGEISALTQK		553.3064	30.595	55.37	2	true	553.30646	30.5446	63.31	2	true	553.3069	28.9616	55.37	2	true
PLPLPDR							454.2539	47.8436	24.63	2	true					
VESLQEEAFVK		703.379	53.16	54.9	2	true	703.38	52.7723	60.77	2	true	703.38043	51.3037	56.8	2	true
WATLNSEESDPPTTK		890.4166	16.7659	64	2	false						618.84375	42.7779	36.53	2	false
ISVIELTHLQK												487.5689	12.6798	12.37	3	false
MGVEIETISPGDGR												640.80743	18.2499	33.24	2	false
TFDQLPQESK		640.8053	19.1792	39.87	2	true	640.8045	19.2152	43.74	2	true	812.9061	41.0045	59.46	2	false
IVSEDFLQDVSAATK		812.90485	42.137	69.46	2	true						573.31366	40.4774	5.31	2	false
ILLADDMGLQK												489.7619	46.8365	38.74	2	true
ISDFDEPK												956.86176	83.1193	45.79	3	false
TALLDAAGVASLTTAEVVTETPKEEK																
AVLVDELPGTM(+15.99)DSVR		809.41	34.4498	57.27	2	true	809.4083	34.5133	69.37	2	false					
YLQDLIAWVEENQHR		638.6531	64.9527	24.41	3	true	638.6536	64.3657	19.18	3	true					
AEQPDVAIDLK		605.8175	28.2688	63.5	2	true	605.8171	28.0168	66.79	2	false	605.8166	26.8709	60.6	2	false
YKSPKPAKSPVSVK		570.3503	59.1537	12.75	3	true										
IVFGVSKR		453.2843	35.105	14.74	2	false						523.77545	9.944	38.51	2	false
TQDQISDK												621.8541	37.9902	56.72	2	true
ADLINLGTAK		621.8556	39.11	58.2	2	true	621.8543	39.0524	58.05	2	true					

7.3. Protein Comparison

The top protein identifications from the compared search results are displayed in the table. The layout of the protein table and 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

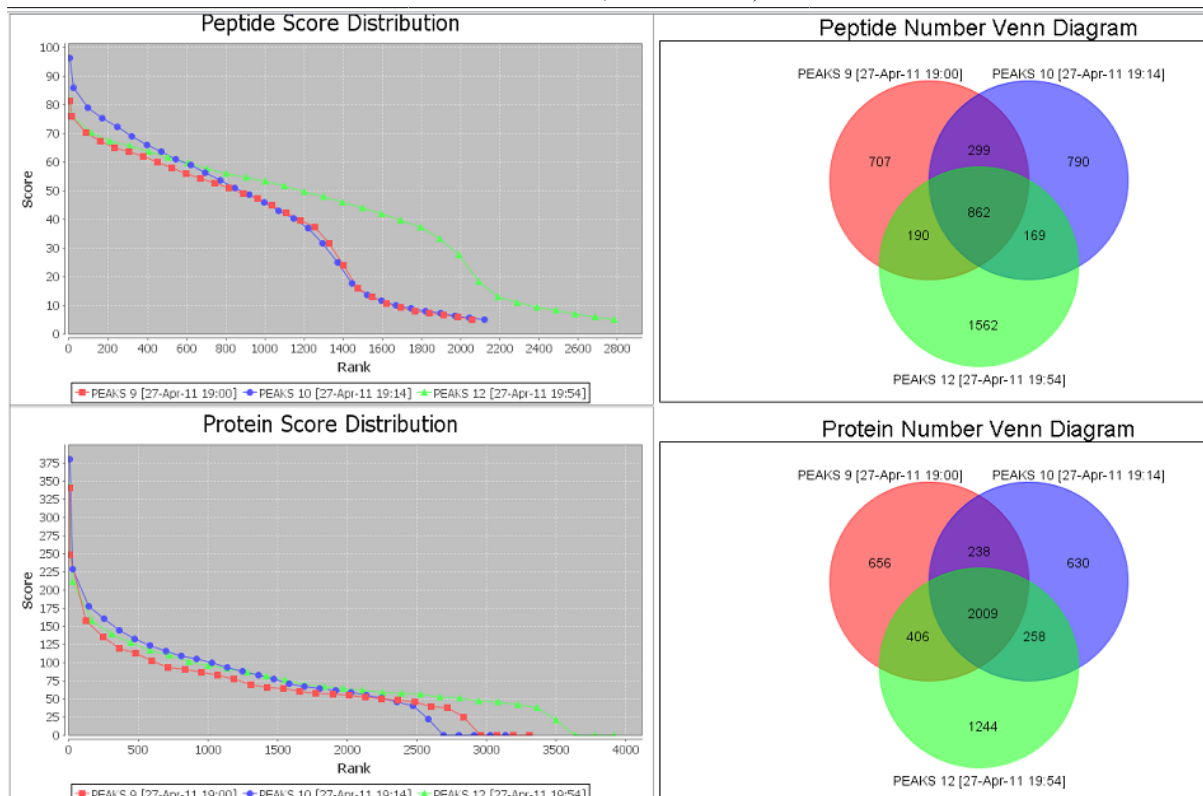
The following screenshot is the protein table showing a typical comparison result:

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]								0	1	1	1	1.41	1.83	64.13	8	2	0	6.9	6.67
P0C736[LRPN1_RAT]																			
Q684U9[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[TBAB_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[PISP_L_HUMAN]		63.27	1	1	1	1.2	1.97												
Q758U6[VID21_NEUCR]								0	1	1	1	0.62	0.69						
Q85QH5[AOT21_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						
Q52KH4[NOEL1_CHICK]														51.43	2	1	1	2.22	3.5
P30084[IECHM_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[PKS7_HUMAN]														46.99	2	2	2	3.33	4.85
A13M81[IRPFC_YEREB]								0	1	1	1	1.18	1.58						
P24724[HSPP90_THERP]		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						
A4VW89[ONAK_RHOS5]														65.32	5	1	0	1.3	2.52
A7EDF3[IR54_SOLS1]								0	1	1	1	0.97	2.28						
Q8CGP0[H2B36_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[H_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
Q77819[ROCK1_RABIT]														40.73	1	1	1	0.47	0.81
Q25538[1433_NEOCA]														64.13	8	2	0	6.06	6.77
Q8NDF8[IPAPD5_HUMAN]		16.93	2	1	1	1.52	3.5												
Q6VAF4[ITB89_GOSHI]								55.2	1	1	0	2.7	3.15						
P22492[HIT_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[ITTHY_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[IRP2_FLAVB]		0	1	1	1	0.23	0.59												
Q64L94[ISME1_FIG]		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[LDAH_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[PTTR_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
Q4QU6[ISPF30_RAT]		36.46	2	1	1	2.86	4.2												
O43103[ISID2_USTMA]		0	1	1	1	0.26	0.32												
ABYUJ3[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 which can be easily exported for usage in publications. The peptide score distribution, protein score distribution, peptide number Venn diagram, and 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.

Peptide, PTM, and Mutation
Identification (PEAKS DB,
PEAKS PTM, and SPIDER)



7.5. Exporting Comparison Result

The peptide table, protein table, and 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.

Export Options

- ☒ Export peptide result
- ☒ Export protein result
- ☒ Export statistics graph

Image Options

Scale: %

File Format: Width: Height:

Save As:

Chapter 10. Multi-Round Search


1. Multi-Round Search Overview

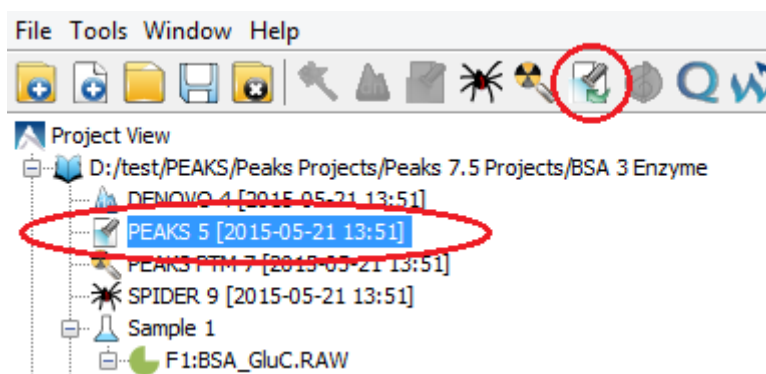
The Multi-Round Search is an additional search function which follows a completed search result, to examine the remaining “*de novo* only” spectra. This additional search allows users to perform PEAKS DB, PTM and SPIDER searches using different parameters and/or to integrate a different database.

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/PEAKSpsters/peaks - quantitative analysis of endogenous peptides web.pdf>

The steps to perform a multi-round search are:

1. Select a PEAKS DB, PTM or SPIDER result node in the "Project View". Launch the multi-round search by clicking the  button on the tool bar.



Note


The PEAKS DB, PTM or SPIDER result must be opened first to run a multi-round search on it.

Note

Refer to Chapter 9, *Peptide, PTM, and Mutation Identification (PEAKS DB, PEAKS PTM, and SPIDER)* for details on PEAKS DB search.

2. Specify the search parameters in the pop-up dialog. The parameter selection dialog is similar to the parameter selection dialog of PEAKS DB search. Specify new search parameters or select a different database to search the remaining unidentified spectra from the existing result. Most of the parameters are self-explanatory. PEAKS PTM and SPIDER can be enabled in the dialog to automatically run after PEAKS DB search. Click OK to start multi-round search.
3. Wait for the analysis to finish. Search times range from minutes to hours, depending on several factors including the dataset size and computer speed. New result nodes for multi-round PEAKS DB, PEAKS PTM, and SPIDER will appear in "Project View". Double click the result nodes to examine analysis reports.

2. Setting Multi-Round Search Parameters

After selecting a PEAKS search analysis node in the "Project View" and clicking the multi-round search toolbar icon , the multi-round search parameters dialog will appear.

PEAKS Search Predefined parameters ▼

Error Tolerance

Precursor mass: ppm ▼ using monoisotopic mass ▼ Fragment ion: Da

Enzyme

None ▼ View

Allow non-specific cleavage at both ▼ ends of the peptide.

In None Enzyme mode, PEAKS Search will process all peptides with length up to 65.

PTM

☒ Carbamidomethylation
☒ Oxidation (M)
☒ Acetylation (Protein N-term)
☒ Deamidation (NQ)

Set PTM
Remove
Switch type

Maximum allowed variable PTM per peptide ▲▼

Database

☒ Select database Database: [UniProt_SwissProt] ▼ View

☐ Paste sequence Taxa: all species Set/View taxa...

Spectra Source

Spectra from: de novo only of PEAKS5 ▼

General Options

☒ Estimate FDR with decoy-fusion. ?

☒ Find unspecified PTMs and common mutations with PEAKS PTM Advanced Settings

☐ Find more mutations with SPIDER

OK Cancel Help

The parameter selection dialog is similar to the parameter selection dialog of PEAKS DB search. The only difference is the Spectra Source that shows that the *de novo* only peptides spectra will be searched. Refer to Chapter 9, Section 2, “Setting PEAKS Search Parameters” for details on parameter settings.

Once the parameters have been specified, the Multi-Round Search analyzes only those spectra which were determined to be *de novo* only from the first round search. In addition, the *de novo* spectra to be analyzed in the multi-round search follow the average local confidence (ALC) filters that were set in the first round search.

3. Understanding Multi-Round Search Results

After multi-round search is complete, several result nodes are generated including a *de novo* result node and identification result from database search tools: PEAKS DB, PEAKS PTM and SPIDER.

The screenshot shows the 'Summary' view of a multi-round search result. On the left is a vertical sidebar with tabs: 'Summary' (selected), 'Protein', 'Peptide', and 'De novo only'. The main content area has a filter bar at the top with the following settings: 'Peptides -10lgP' is set to 15 with a dropdown arrow, followed by 'FDR'; 'Proteins -10lgP' is set to 20 with a dropdown arrow, followed by 'and' and '≥ 0'; 'De novo only ALC (%)' is set to 50 with a dropdown arrow, followed by 'and -10lgP' with a lock icon and a value of 15 with a dropdown arrow. An 'Apply Filters' button is on the right. Below the filter bar is a section titled '1. Notes' containing the text 'Multi-round search:' followed by a table of search parameters.

Multi-round search:	
Spectra source	Denovo only of run 5
Denovo only ALC%	≥50.0
Denovo only -10logP	≤15.0

Additional information is added to the “Notes” section of the Summary views of all the multi-round result. The information includes the spectra source, and parameters used to select the spectra for multi-round search. The same information is displayed in the properties tab, when a multi-round search node is selected in the “Project View”.

Note

Refer to Chapter 8, Section 3, “Understanding PEAKS *De Novo* Sequencing Results” for details on *de novo* results.

Note

Refer to Chapter 9, Section 3, “Understanding PEAKS Search Results” for details on PEAKS DB, PEAKS PTM, and SPIDER results.

4. Filtering and Exporting Multi-Round Search Results

The multi-round search results offer the same options for result filtering and exporting that are available for PEAKS search results. The filtering parameters and exporting options are available in the Summary view of the result.

Note

Refer to Chapter 9, Section 4, “Filtering PEAKS Search Results” for details on result filtering.

Note

Refer to Chapter 9, Section 5, “Exporting PEAKS Search Results” for details on result exporting.

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 result 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 (v2.5), 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 dialog, there are drop-down lists for each supported database search engine. Using the drop-down lists, you can choose the search results to be combined. You can either:

- **Choose an existing result**

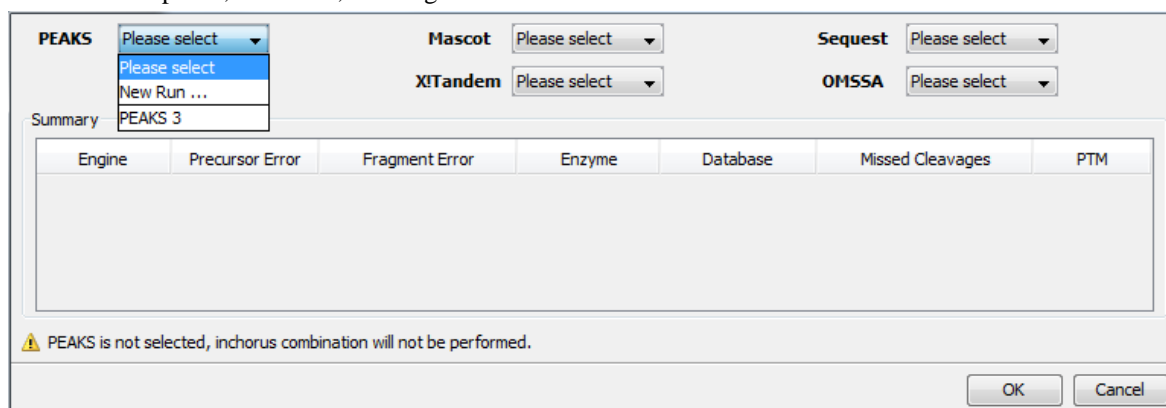
The list shows search results that already exist in the PEAKS project. Simply choose an existing result from the list.

- **Import a result from a file**

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

- **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 dialog 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 Chapter 9, Section 2, "Setting PEAKS Search Parameters".

A screenshot of the PEAKS inChorus parameter dialog box. The dialog has a title bar and a main area with several sections. At the top, there are four groups of search engines: PEAKS, Mascot, Sequest, and OMSSA. Each group has a 'Please select' dropdown menu. Below these, there is a 'Summary' section with a dropdown menu showing 'PEAKS 3'. The main area contains a table with columns: Engine, Precursor Error, Fragment Error, Enzyme, Database, Missed Cleavages, and PTM. At the bottom, there is a warning message: 'PEAKS is not selected, inchorus combination will not be performed.' and two buttons: 'OK' and 'Cancel'.

Note

A PEAKS result must be selected for combining results in 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 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.

The screenshot shows the 'Database List' configuration window. On the left, a sidebar contains 'Enzyme', 'PTM', 'Labeled Q Method', 'Database' (selected), and 'Instrument'. The main area is titled 'Database List' and contains a table with the following entries:

Database List
[UniProt_SwissProt]
Viridiplante
[IPI]

Below the table is the 'Database Details' section. It includes a dropdown for 'FASTA format database' set to 'UniProtKB/Swiss-Prot', and buttons for 'Validate Database' and 'Export Decoy DB' (which is circled in red). The 'Basic Options' section has a 'Database name' field with 'UniProt_SwissProt', a 'Path' field with 'D:\FastaDB\uniprot_sprot.fasta', and a 'Browse' button. There is also a 'Download' button and an 'EST database' checkbox. The 'Advanced Options - Fasta Title Format' section contains fields for 'Rule to parse accession/id from FASTA title' (set to '>(sp\|)*\(\S*\)'), 'Rule to parse description from FASTA title' (set to '\s+\(.*\)'), 'Accession/id URL' (set to 'http://www.uniprot.org/uniprot/'), and a 'Delimiter' field. The 'Taxonomy Options' section has fields for 'taxonid' (set to 'D:\FastaDB\spedist.txt') and 'taxdmp' (set to 'D:\FastaDB\taxdmp.zip'), each with 'Browse' and 'Download' buttons. At the bottom, there are buttons for 'New', 'Add/Update', 'Delete', 'Help', and 'Close'.

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

Mascot

Predefined parameters

Database

Mascot Database: SampleWithDecoy Taxa:

Local Database: SampleDB

Enzyme

Enzyme: Trypsin Allow up to 0 missed cleavages

PTM

Fixed Modification: Carboxymethyl (C)

Variable Modification: Deamidated (NQ), Oxidation (M)

Acetyl (K), Acetyl (N-term), Acetyl (Protein N-term), Amidated (C-term), Amidated (Protein C-term), Ammonia-loss (N-term C), Biotin (K), Biotin (N-term), Carbamidomethyl (C)

☐ Display all modifications

Error Tolerance

Peptide Tol: 20 ppm # ¹³C: 0 MS/MS Tol: 0.8 Da

Peptide Charge: 2+, 3+ and 4+ ☒ Monoisotopic ☐ Average

Misc

Instrument: ESI-TRAP Report top: AUTO hits ☐ Decoy ☒ Search decoy database from PEAKS

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

The format of an inChorus result is very similar to the format of a PEAKS Search result (Chapter 9, 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 the corresponding peptide-spectrum match is correct. The percentage score is calculated in accordance with the empirical calculation used in PeptideProphetTM (Keller et al. Anal. Chem. 2002, 74:5383-92).

1. The method of PeptideProphetTM 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 with a weighted sum.
3. The weighted sum scores of all peptides are converted to a probability by using the PeptideProphetTM 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 with a weighted sum. The PeptideProphetTM 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:

Combining Multiple Database Search Engines with PEAKS inChorus

- **P:** PEAKS Search (PEAKS DB, PEAKS PTM and SPIDER)
- **M:** Mascot
- **S:** SEQUEST
- **X:** X!Tandem
- **O:** OMSSA

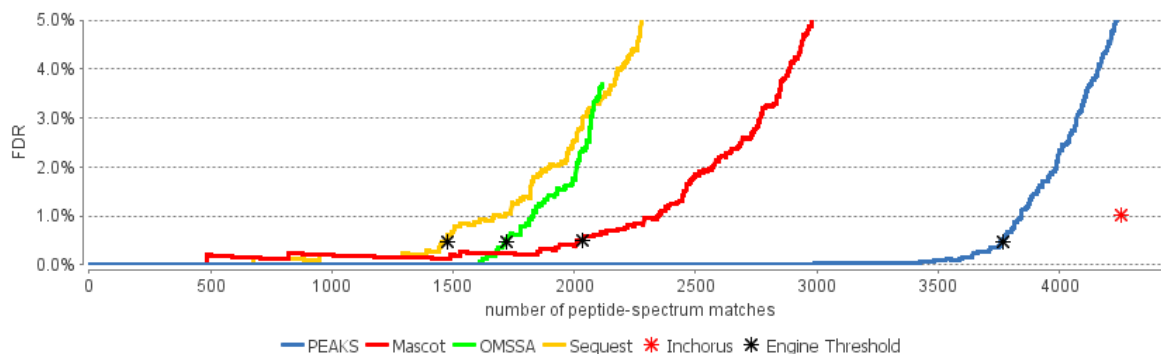
Also, the following appearances of the icon indicates:

- **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.

1 - 28 of 28												scan = <input type="text" value="search"/>		no results		
		Peptide	Scan	m/z	RT	Mass	ppm	#Spec	Score (%) ↓	Engines	PEAKS	Sequest	Mascot	X!Tandem	Omssa	PTM
De novo only	1	AVARYN(+98)TTK	533	365.8648	24.04	1094.5720	0.6	2	99.96	P B M - O	41.02	2.73	40.92	-	7.31E-5	
	2	ASEALKPDSQK	194	391.8754	18.59	1172.6038	0.5	8	99.95	P B M - O	56.42	2.43	39.99	-	5E-4	
	3	SEVAHRFK	414	487.2649	22.28	972.5142	1.1	2	99.92	P B M - O	45.42	1.33	42.05	-	-	
	4	LVAETEDRK	465	354.1932	23.07	1059.5559	1.7	3	99.92	P B M - O	28.04	2.27	45.01	-	1.01E-2	
	5	QGVTEAAEK	167	466.7378	18.14	931.4611	-0.1	3	99.92	P B M - O	27.83	1.11	40.80	-	5.2E-3	
	6	IGYGSN(+98)KK	220	434.2328	18.96	866.4498	1.4	2	99.91	P - M - -	44.77	-	37.32	-	-	
	7	Q(-17.03)AKHHEIDTK	568	394.5337	24.46	1180.5725	5.7	1	99.66	P - M - -	40.05	-	24.58	-	-	
	8	EAHEIVSK	17	456.7428	15.30	911.4713	-0.2	1	99.54	P - M - -	24.01	-	30.76	-	-	
	9	PGGGQVEIK	561	435.7363	24.38	869.4607	-3.0	1	99.38	P B M X	29.06	0.90	21.06	7.1E-2	-	
	10	NPTAFKK	142	403.2322	17.77	804.4494	0.5	1	99.25	P B M - -	20.29	0.91	35.37	-	-	
	11	AALAEARDSK	405	516.2780	22.17	1030.5408	0.6	2	99.16	P B M - -	20.95	0.83	32.96	-	-	
	12	VISQDTQPHQQK	557	704.8610	24.33	1407.7107	-2.3	1	97.93	P - M - -	34.71	-	-	-	-	

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.

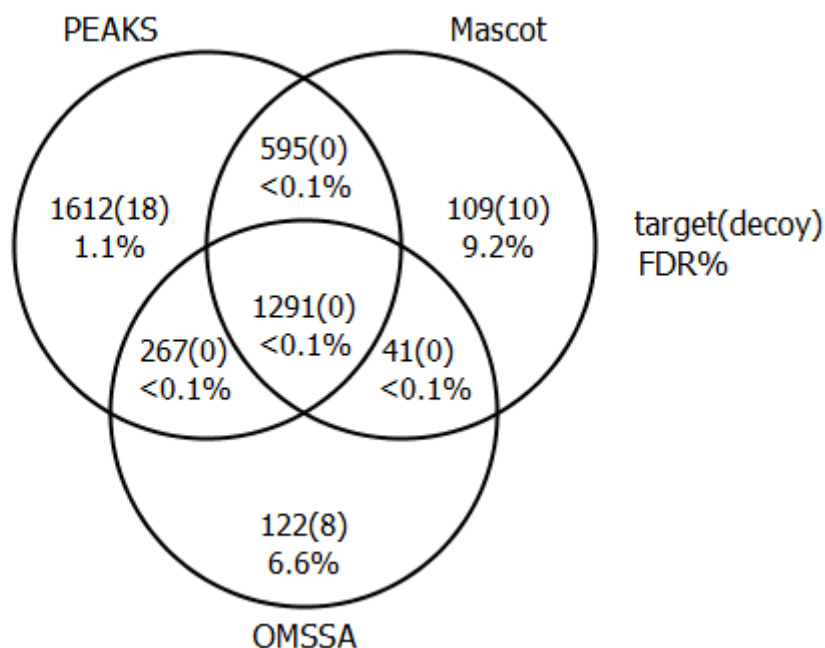
FDR curves. The FDR curves for all selected search engines are placed together in a single figure in the Summary View. The black * marks on the search engines' FDR curves indicate the score thresholds of the corresponding engines. The red * mark indicates the inChorus FDR threshold.



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 inChorus FDR or by each individual engine's score. If filtering by inChorus FDR, PEAKS calculates the score threshold for each search engine to automatically to 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 Chapter 16, 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 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.

The quantification analysis is based on a PEAKS DB result (See Chapter 9, *Peptide, PTM, and Mutation Identification (PEAKS DB, PEAKS PTM, and SPIDER)*). Ensure that the specified isotopes are labeled as PTMs when performing the database search. After database search is complete, follow these steps:

1. Select a PEAKS DB result node in the Project view. Click the PEAKS Quantification toolbar icon .


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 dialog; 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.

The quantification result display is similar to that of the PEAKS DB result. However, the ratio and standard deviation columns are 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 **Tools | Quantification** command from the menu is clicked. Select **Precursor Ion Quantification eg. SILAC** from the left hand side.

Quantifications

☐ Label Free

☐ Reporter Ion Quantification
eg. iTRAQ/TMT

☒ Precursor Ion Quantification
eg. SILAC

Predefined parameters ▼

Precursor Ion Quantification

Basic Options

Mass Error Tolerance: ppm Upper Bound of Precursor Charge:

Retention Time Range: min. Peptide Score Threshold (-10logP):

Select Method: SILAC-3plex (R10,K8)R6,K4 ▼ New

Quantification Method Detail

Name: SILAC-3plex (R10,K8)R6,K4 Method Type: Precursor Ion Quantification ▼

Label Options

Sample Name	Modification	Modification Detail
Heavy	13C(8) 15N(2) Silac label	10.020909@[R]
Heavy	13C(6) 15N(2) Silac label	8.014199@[K]
Medium	13C(6) Silac label	6.020129@[R]
Medium	4,4,5,5-D4 Lysine	4.025107@[K]
Light	None	

OK Cancel Help


The following parameters are available in the “Basic Options” section of the quantification window:

- **Mass Error Tolerance:** This parameter is used to locate the precursor ion peak group of an identified peptide in the survey scan. In a SILAC or ICAT experiment, we are usually dealing with non-centroided survey scans. Therefore, the mass error tolerance should be set a little wider than the parent ion error tolerance used in the PEAKS DB search.
- **Upper Bound Precursor Charge:** The precursor ion of an identified peptide may produce sibling ions of different charge states. Only those sibling ions with a charge less than the upper bound precursor charge will be considered for quantification of the identified peptide.
- **Retention Time Range:** The retention time range is the maximum elution time range that is considered for the quantification of an identified peptide.
- **Peptide Score Threshold (-10lgP):** Only peptides with a score above this threshold are used to quantify the identified proteins.

Choose the appropriate quantification method used in experiment from the **Select Method** drop-down list that contains predefined quantification methods. The details of the selected quantification method will appear in the “Quantification Method Detail” panel. To create a new quantification method, click the **New** button to display the “New/Edit Quantification Method” dialog. Refer to Chapter 17, Section 2.3, “Labeled Q Method Configuration” for how to create or edit a quantification method.

Selecting the **Save as...** command from the **Predefined parameters** drop-down list allows the users save parameters for the ease of use 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 tab will appear by default.

3.1. Summary View

The MS level quantification result summary is presented in the "Summary" view. The quantified proteins are clustered into a tree structure in the Heatmap. Proteins are clustered according to their exhibition of a similar expression trend across samples. Move the cursor to the tree to select a cluster and left click to show the variation trend chart for that cluster. Other information displayed on the summary page is self-explanatory.

1. Heatmap View



Cell colour represents the log₂ ratio to the Control Sample (Control Sample is marked with (*))

2. Notes

3. Result Statistics

Table 1. Statistics of data and unfiltered result.

of MS Scans 5591
of MS/MS Scans 12443

Table 2. Result filtration parameters.

Protein fold change ≥ 5

4. Other Information

Table 4. Search parameters.

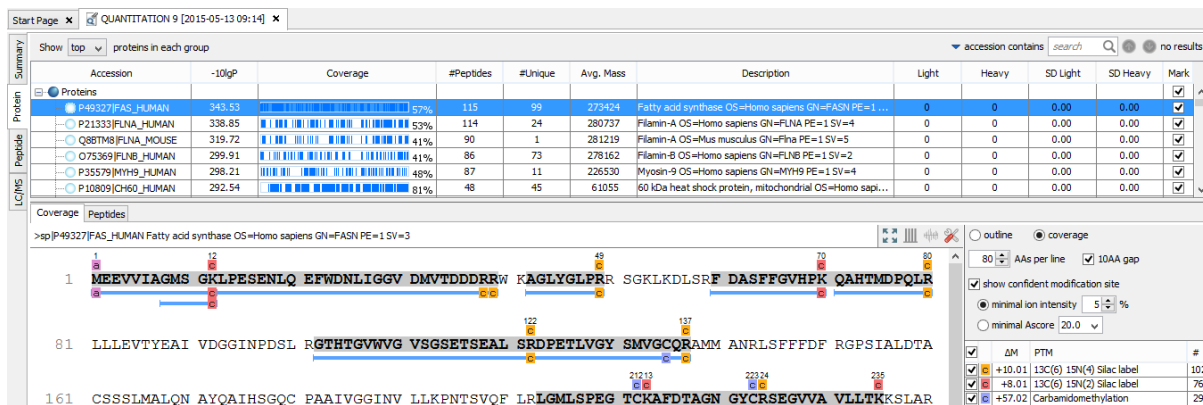
Quantification Type: ICAT/SILAC
Quantification Mass Tolerance: 0.1Da
Quantification RT Range: 1.0min.
Upper Bound Charge: 4
Peptide Score Threshold:15.0
No Label: K (100.0%) R (100.0%)
Light: K(+4.03) (100.0%) R(+6.02) (100.0%)
Heavy: K(+8.01) (100.0%) R(+10.01) (100.0%)

Table 5. Instrument parameters.

Fractions: VPS4B1J.RAW
Ion Source: ESI (nano-spray)
Fragmentation Mode: CID, CAD, IRMPD (y and b ions)
MS Scan Mode: FT-ICR/Orbitrap
MS/MS Scan Mode: Linear Ion Trap

3.2. Protein View

The "Protein" view shows a list of proteins that are identified in the database search, together with the protein coverage of their identified peptides in the window below. The relative intensity of those quantifiable proteins are displayed in the ratio columns with label names incorporated into the header (e.g. "Light", "Medium", or "Heavy"). The ratio is calculated from the unique peptides of the protein. Proteins with no unique peptides will not be assigned a ratio. SD represents the standard deviation of the peptide ratios in the protein. The sample on which the ratio is based on can be changed in the **Show ratio to** drop-down menu in the "Summary" view. Normalization factors can be selected in the "Normalization factor" drop-down list in the "Summary" view. The coverage of the selected protein and the peptides of the selected protein, together with their ratios, are displayed at the bottom half of the Protein view. Refer to Chapter 9, Section 3.3.2, "Protein Coverage" for details on protein coverage pane.



Note

The “Peptide Mapping” and the “PTM Profile” functions are available only for PEAKS DB, PEAKS PTM and SPIDER results.

3.3. Peptide View

The Peptide view displays all the identified peptides and their intensities. The interface is similar to the Peptide view in a PEAKS DB result. Refer to Chapter 9, Section 3.4, “Peptide View” for details. The relative intensities of the quantifiable peptides are displayed in the intensity columns with label names incorporated into the header (e.g. “Light”, “Medium”, or “Heavy”). The reference sample can be changed from the **Show ratio to** drop-down menu in the “Summary” view. The mode for the normalization factor can also be selected from the “Normalization factor” list in the “Summary” view.

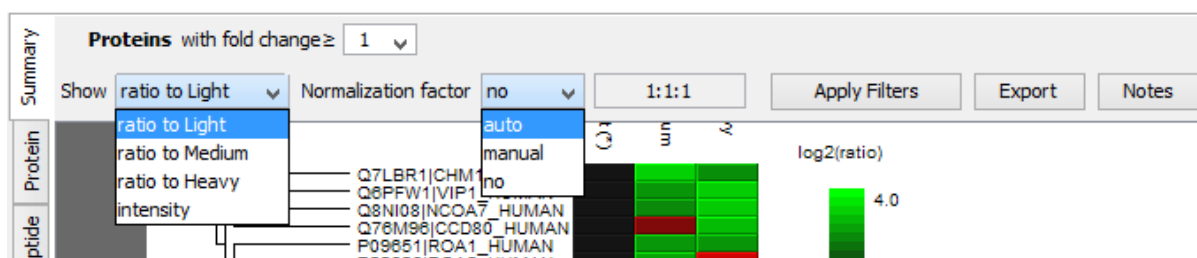
1 - 1000 of 1893												scan = <input type="text" value="search"/>		no results	
Summary		Peptide	-10lgP ↓	Mass	ppm	m/z	RT	Scan	#Spec	Accession	Light	Medium	Heavy	PTM	
Protein	1	GC(+57.02)TATLGNF(+4.03)AK(+4.03)	70.90	1146.5942	3.2	574.3062	47.98	6072	2	P15880/RS2_HUMAN	1	1.189	0.016	<div><div>c</div><div>d</div><div>d</div></div>	
	2	SAGPQSLLASVI(+6.02)AEK	65.20	1603.8876	0.2	802.9512	98.57	13921	2	P35612/ADD8_HUMAN	-	-	-	<div><div>c</div></div>	
	3	ATVAGDQWVEY(+4.03)K(+4.03)	60.66	1313.7067	7.8	657.8657	42.11	5097	1	Q13642/FHL1_HUMAN	1	1.591	0.16	<div><div>d</div><div>d</div></div>	
	4	LIIVEGCI(+57.02)QR	55.96	1086.5856	-3.5	544.2982	48.10	6093	2	Q13733/AT1A4_HUMAN	1	0.137	0.061	<div><div>c</div></div>	
	5	KVNPLPEK	54.68	1072.6281	-7.5	537.3173	49.42	6300	1	P62424/JL7A_HUMAN	1	0.216	0.181	<div><div>c</div></div>	
	6	SGVSLAALK(+4.03)	53.67	848.5269	3.2	425.2721	55.14	7216	1	P10412/H14_HUMAN	1	0.551	0.645	<div><div>d</div></div>	
	7	TK(+8.01)DIEDVFI(+4.03)K(+4.03)	52.68	1272.6932	6.3	637.3579	91.59	12840	14	Q07955/SRSF1_HUMAN	-	-	-	<div><div>c</div><div>d</div><div>d</div></div>	
	8	GGGGNFGPGGSNF(+4.03)R(+6.02)	50.96	1386.6674	9.3	694.3474	48.16	6103	1	P22626/ROA2_HUMAN	-	-	-	<div><div>d</div><div>c</div></div>	
	9	SHEGETAY(+4.03)IR(+6.02)	50.49	1171.5867	-7.7	586.7961	24.40	2433	3	Q07955/SRSF1_HUMAN	1	1.074	0.123	<div><div>d</div><div>c</div></div>	
	10	DGYDYDGY(+4.03)R(+6.02)	48.23	1132.4706	-3.6	567.2405	44.78	5544	3	Q07955/SRSF1_HUMAN	-	-	-	<div><div>d</div><div>c</div></div>	

3.4. LC/MS View

The LC-MS view shows the distribution of LC-MS signals/features and identified peptides in an interactive LC-MS heat map. Check the option “Identified peptides” to show the locations of the MS/MS spectra with peptide identifications as indicated by blue squares. Refer to Chapter 5, Section 4, “LC/MS View” for a list of supported interactions in the LC-MS heat map.

3.5. Filtering Quantification Results

Quantification results can be filtered based on the number of fold changes between samples. To do this, use the **Proteins with fold change#** drop-down list in the title bar of the “Summary” view panel. To change the reference sample, select the appropriate sample (eg. “ratio to light”) from the **Show** drop-down list. To change the normalization factor of the protein ratio, select “auto”, “manual”, or “no” from the **Normalization factor** drop-down list. For manual normalization, enter the normalization factors in the textbox to the right. Click the **Apply Filters** button to filter the results. The results will be updated in the “Summary” view, the “Protein” view and the “Peptide” view accordingly.



4. Exporting Quantification Results


PEAKS Q results can be exported to other supported formats. All export functions are available through the “Summary” view panel. To export the quantification results, click the **Export** button in the title bar of the “Summary” view panel. Refer to Chapter 16, Section 5.1, “Exporting Labeled Quantification Results” for details.

Chapter 13. Reporter Ion Quantification (e.g. iTRAQ and TMT)

1. Overview


Reporter ion quantification with isotope labels at MS/MS level is one of the three quantification modes that are supported by the optional PEAKS Q module of PEAKS Studio. This is based on the relative intensities of fragment peaks at fixed m/z values within an MS/MS spectrum. In this mode, isotope 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 scans, labels from different samples will produce different reporter ions; which can then be used to calculate the quantification ratio between samples. User-defined labels are supported in PEAKS Q as well as commercial labels such as iTRAQ and TMT.

The quantification analysis is based on a PEAKS DB result (See Chapter 9, *Peptide, PTM, and Mutation Identification (PEAKS DB, PEAKS PTM, and SPIDER)*). Ensure that you have specified the isotope labels as PTMs in the database search. After database search is complete, follow these steps:

1. Select a PEAKS DB result node in the Project view. Click the Quantification toolbar icon .


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 as “Reporter Ion Quantification”; specify the PEAKS quantification parameters in the dialog box on the right; and click **OK**.
3. Wait for the analysis to finish. A new quantification result node  will appear at the Project view. Double click the node to open the result file.

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 along with standard deviations for protein ratios.

2. Setting Parameters

The following parameter dialogue pops up when clicking the quantification tool bar icon . Select **Reporter Ion Quantification** eg. **iTRAQ** from the left side under the **Quantifications** heading in the quantification window. Enter the relevant MS/MS level labeling quantification parameters on the right side of the window.

Quantifications

☐ Label Free
☒ Reporter Ion Quantification
 eg. iTRAQ/TMT
☐ Precursor Ion Quantification
 eg. SILAC

Reporter Ion Quantification Predefined parameters

Basic Options
 Mass Error Tolerance: ppm Peptide Score Threshold (-10logP):

Select Method: New

Quantification Method Detail

Name: Method Type:

Modification Target
☒ N-Terminal Modification ☐ C-Terminal Modification
☐ Side Chain Modification at Modification Mass

Label Options

Name	Reporter Ion Mass (Da)
114	114.1107
115	115.1077
116	116.1111
117	117.1144

OK Cancel Help


The following parameters are available in the **Basic Options** section of the quantification window:

- **Mass Error Tolerance:** This parameter is used to locate the reporter ion peaks in the MS/MS spectrum. A little wider tolerance than the fragment ion error tolerance set in PEAKS DB is recommended.
- **Peptide Score Threshold (-10lgP):** Peptides with a score above this threshold are used to quantify the identified proteins. Peptide level quantification results are still shown for peptides with a score below this threshold.

Choose the appropriate quantification method used in the experiment from the **Select Method** drop-down list that contains predefined quantification methods. The details of the selected quantification method will appear in **Quantification Method Detail** panel. To create a new quantification method, click the **New** button to display the “New/Edit Quantification Method” dialog. Refer to Chapter 17, Section 2.3, “Labeled Q Method Configuration” for how to create or edit a quantification method.

Selecting the **Save as...** command from the **Predefined parameters** drop-down list allows the users save parameters for ease of use when regularly performing quantification with the same parameters.

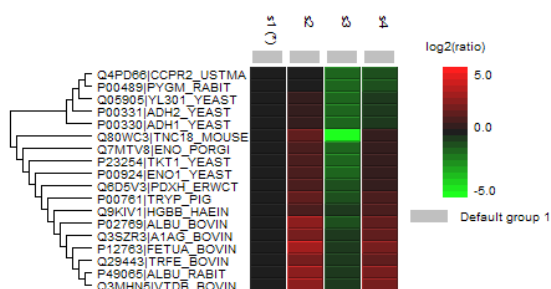
3. Understanding the Results

Once completed, the protein quantification result will be displayed in a quantification node  in the “Project View” panel. Double click on this node to open the result that contains three views: Summary view, Protein view and Peptide view. The “Summary” view tab will appear by default.

3.1. Summary View

The MS/MS level quantification result summary is presented in the “Summary” view. The quantified proteins are clustered into a tree structure in the Heatmap. Proteins are clustered according to their exhibition of a similar expression trend across samples. Move the cursor to the tree to select a cluster and left click to show the variation trend chart for that cluster. Other information displayed on the summary page is self explanatory.

1. Heatmap View



Cell colour represents the log₂ ratio to the Control Sample (Control Sample is marked with (*))

2. Notes

3. Result Statistics

Table 1. Statistics of data and unfiltered result.

of MS Scans 11982
of MS/MS Scans 22175

Table 2. Result filtration parameters.

Protein fold change ≥ 2

4. Other Information

Table 4. Search parameters.

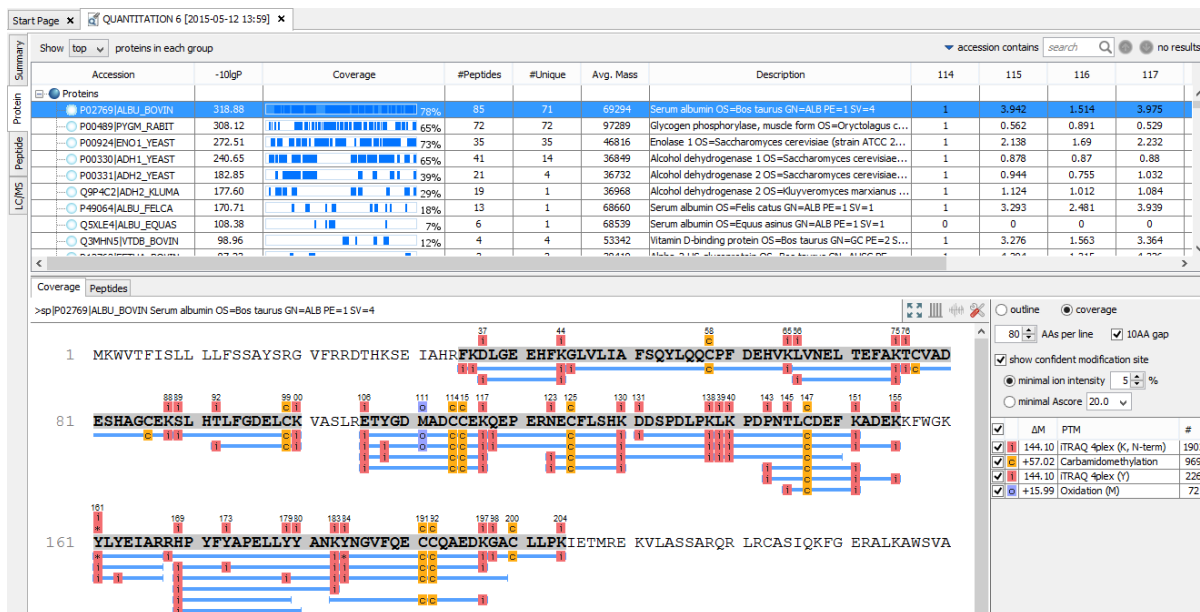
Quantification Type: ITRAQ
Quantification Mass Tolerance: 0.2Da
Quantification RT Range: 0.0min.
Upper Bound Charge: 0
Peptide Score Threshold: 20.0
Label 0: 114.11 (100.0%)
Label 1: 115.11 (100.0%)
Label 2: 116.11 (100.0%)
Label 3: 117.11 (100.0%)

Table 5. Instrument parameters.

Fractions: inj1.RAW, inj2.RAW, inj3.RAW, inj4.RAW, inj5.RAW
Ion Source: ESI(nano-spray)
Fragmentation Mode: CID, CAD, IRMPD (y and b ions)
MS Scan Mode: FT-ICR/Orbitrap
MS/MS Scan Mode: Linear Ion Trap

3.2. Protein View

The “Protein” view shows a list of proteins that are identified in the database search along with protein coverage of their identified peptides in the window below. The quantification ratios of the quantifiable proteins are displayed in the ratio columns with their label names as the header (eg. "114"). A protein is considered quantifiable when identified by a unique peptide above the peptide score threshold set in the parameters. The ratio is calculated from the unique peptides of the protein. Proteins with no unique peptides are not considered to be assigned a ratio. The denominator sample can be changed from the “Show” drop-down menu in the “Summary” view. The normalization mode can also be selected in the “Summary” view (see Section 3.5, “Filtering Quantification Results”). The SD is the standard deviation of the peptide ratios in the protein. The peptides of the selected protein, along with their ratios, are displayed at the bottom half of the Protein view.



3.3. Peptide View

The "Peptide" view displays all the identifiable peptides and their relative intensities. The interface is similar to the peptide table in a PEAKS DB result (see Chapter 9, Section 3.4.1, "Peptide Table"). The relative intensities of the quantifiable peptides are displayed in the intensity columns with label names as the headers (eg. "114").

Peptide	-10lgP	Mass	ppm	m/z	RT	Scan	#Spec	Accession	114	115	116	117	PTM
I(+144.10)EEETIAAAAGAGGFAFHHEK(+144.10)	140.04	2778.3926	1.3	927.1393	54.95	F23:3748	511	Q08655 ASR1_SOLLIC	-	-	-	-	-
V(+144.10)LVNEEQSPDIAQGVHGLTK(+144.10)	129.70	2670.4656	1.5	891.1638	47.96	F7:3051	20	A7PR36 METK4_VITVI	1	2.555	0	0	-
A(+144.10)LVNHEEDDLGK(+144.10)GGHSLTTGNAGGR	126.92	3132.6365	1.8	1045.2213	52.98	F20:3469	28	O65199 SODCP_VITVI	1	1.267	0	1.666	-
L(+144.10)SGGDHAGTVVGK(+144.10)	122.31	1734.9579	2.7	579.3281	38.65	F38:1909	25	P56648 RBL_VITVI	1	1.29	0.125	0.111	-
F(+144.10)DVYINDEYSVRPK(+144.10)	120.28	2234.1057	2.0	745.7106	46.99	F1:2393	102	P43311 PPO_VITVI	1	1.22	1.832	1.699	-
G(+144.10)PLQISWNY(+144.10)NYGPAGR	118.95	2080.0693	3.3	694.3660	48.52	F2:2881	18	P29023 CHIB_MAIZE	1	2.805	9.038	12.614	-
I(+144.10)WFGIATDFDSEDDITEER	118.95	2632.2295	1.2	878.4182	54.65	F11:3578	14	Q06FW4 PSAB_PELHO	1	0.821	0.158	0.162	-
I(+144.10)NEPTAAALAYGLDK(+144.10)K(+144.10)	117.58	2219.2830	1.4	740.7693	50.78	F12:2965	20	P22953 MD37E_ARATH	1	2.66	0	0.891	-
I(+144.10)NEPTAAALAYGLDK(+144.10)K(+144.10)	117.58	2219.2830	1.4	740.7693	50.78	F12:2965	20	Q9VCS1 DNAAK_RHOMR	1	2.66	0	0.891	-
I(+144.10)NEPTAAALAYGLDK(+144.10)K(+144.10)	117.58	2219.2830	1.4	740.7693	50.78	F12:2965	20	Q9VST6 DNAAK_ALCBS	1	2.66	0	0.891	-
G(+144.10)PAPSLALAHADIDGK(+144.10)	117.18	2048.1104	2.8	683.7126	50.73	F2:3107	4	Q69GY7 UCRIA_SOLTU	1	1.612	1.091	1.337	-

Note

Select a peptide and zoom in to the reporter ion region of the MS/MS to examine the reporter ions.

3.4. LC/MS View

The LC-MS view shows the distribution of LC-MS signals/features and identified peptides in an interactive LC-MS heat map. Check the option "Identified peptides" to show the locations of the MS/MS spectra with peptide identifications as indicated by blue squares. Refer to Chapter 5, Section 4, "LC/MS View" for a list of supported interactions in the LC-MS heat map.

3.5. Filtering Quantification Results

The Quantification result can be filtered based on the number of fold changes in the proteins. The appropriate values for the filters can be set by changing the filtration parameter values from the drop-down lists in the title bar of the "Summary" view panel and clicking on the **Apply Filters** button. The result will be updated in the "Summary" view, the "Protein" view and the "Peptide" view accordingly. The intensity columns of the "Protein" and "Peptide" views display the absolute intensity or relative intensity of the quantifiable proteins and peptides. To change the reference sample, select the appropriate sample (eg. "ratio to 114") from the **Show** drop-down list in the "Summary" view. PEAKS also supports changing the normalization factor of the protein ratio. Select **"auto"**, **"manual"**, or **"no"** from the **Normalization factor** drop-down list. For manual normalization, provide the normalization factors in the textbox to the right.

Summary	Proteins with fold change \geq 1				
	Show	ratio to 114	Normalization factor	auto	1:1.6:0.96:0.99
	Apply Filters		Export	Notes	

Note

When filtration parameters are changed, the **Apply Filters** button changes color as a reminder to apply the filter by clicking it.

4. Exporting Quantification Results

PEAKS Q results can be exported to other supported formats. All export functions are available through the “Summary” view panel. To export the quantification results, click the **Export** button in the title bar of the “Summary” view panel. Refer to Chapter 16, Section 5.1, “Exporting Labeled Quantification Results” for details.

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 can be performed without any identification result. But you can also choose to attach an existing identification result as the last step of the label free quantification.

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  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 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.

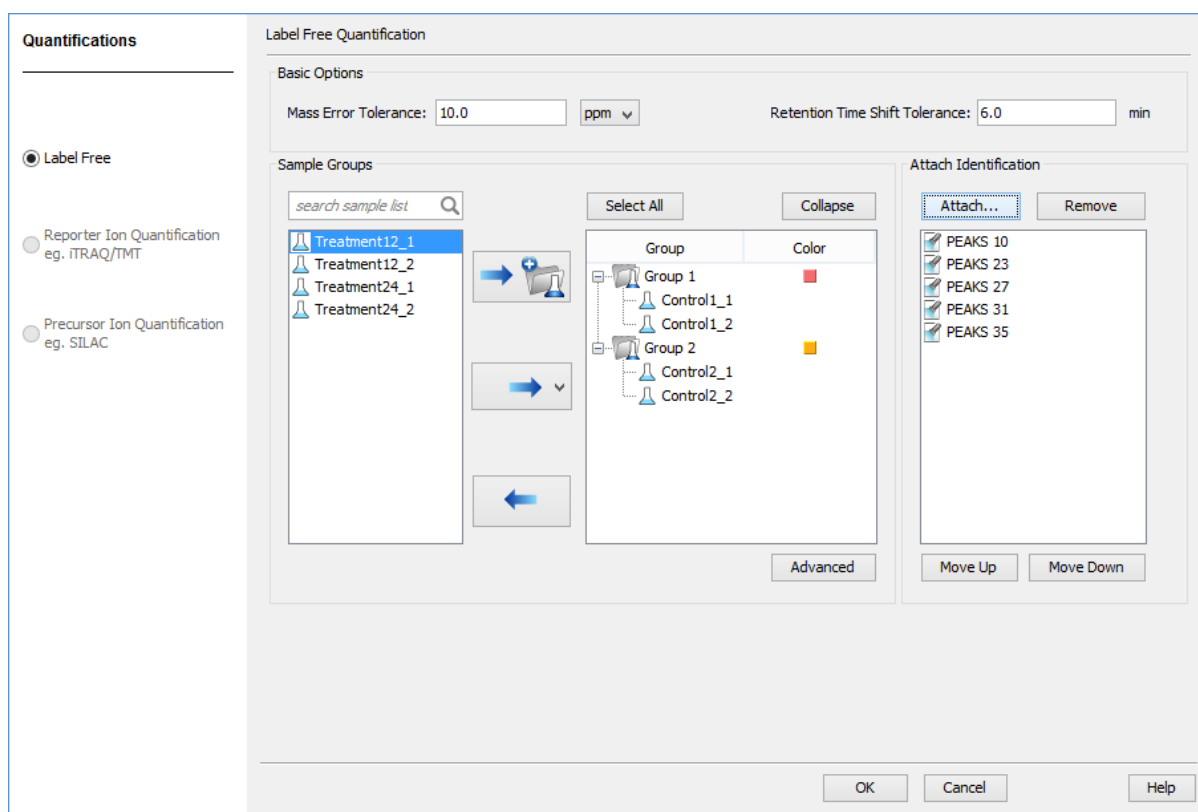
Note

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  will appear at the project tree. Double click the node to open the result file.




2. Setting Parameters

Select "Label Free" from the left side under the "Quantifications" heading in the quantification window to view the label free quantification parameters on the right side.



The following parameters are available in the “Basic Options” section of the quantification window:

- **Mass Error Tolerance:** Mass shift between different runs. Initially this should be set as twice 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 shift between different runs. Initially this should be set as the highest possible shift between runs. Later, by checking the RT shift distribution figure in the result, you can narrow the tolerance.

The “Sample Groups” section allows the users to arrange the samples into groups according to different experiment conditions. All available samples are listed on the left side of the panel. All of the selected samples in the left side list can be added to a new group in the group table by clicking , or can be added to an existing group by clicking  and selecting the group from the drop-down list. Selected groups or selected samples in the group table can be removed by clicking  for rearrangement of the samples. The groups are color-coded by the color column in the group table. These colors will be used in the heat map of the Summary view to distinguish the groups.

Click the **Advanced** button to open the advanced settings dialog where the reference sample and training samples can be selected manually. 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. In both cases, auto detection should fit most cases.

Reference Sample: Auto detect ▼

(Reference sample will be used as the center to do retention time alignment. Auto-detection should fit most cases.)

Training Samples: Auto detect ▼ Auto detect ▼

(Feature vector quality is calculated based on a pair of training samples. Bio-replicates are preferred to be used as training samples. Auto-detection should fit most cases.)

OK Cancel

To attach the identification results to the label free quantification result, click on **Attach ...** button in the Attach Identification section. The “Attach identification results” dialog 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.


Results with FDR validation:

<input checked="" type="checkbox"/>	PEAKS 4
<input checked="" type="checkbox"/>	PEAKS 6
<input checked="" type="checkbox"/>	PEAKS 8

☒ ALL
☐ PEAKS DB
☐ PEAKS PTM
☐ SPIDER

FDR Threshold: 1.0 % Attach Cancel

3. Understanding the LFQ Results

Once completed, a label free quantification result node  will be added to the project tree. Double click on this node to open the LFQ result that contains multiple tabs as described below:

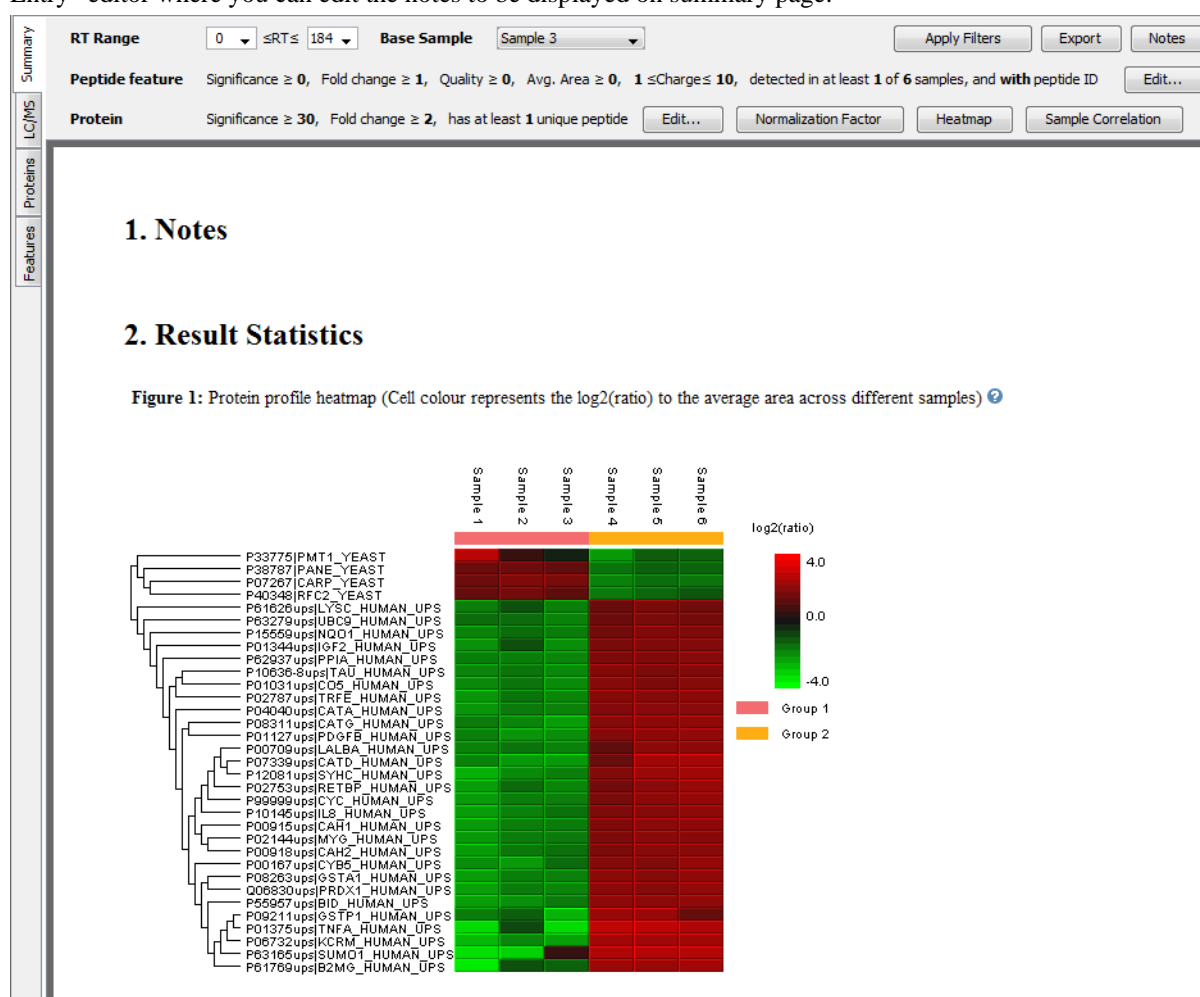
- **Summary:** Outline of PEAKS LFQ search results with statistics. This is the place to examine the overall performance of the experiment and adjust filters.
- **LC/MS:** The overall mass signal view for each MS run, marked with detected features and MS/MS spectra. See Chapter 5, Section 4, “LC/MS View” 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:

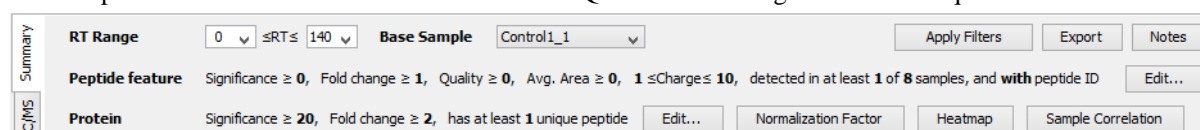
1. **Result Filtration:** This is achieved by specifying the filtration rules in the area at the top of the Summary View. The filtration function is discussed in Section 3.1.1, “Result Filtration”.
2. **Summary Report:** Several statistical charts assist the user to get an overall picture of the results, assess the result quality, and examine the reliability and reproducibility of the experiment. This function is the focus of this section.
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 report 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 where you can edit the notes to be displayed on summary page.



3.1.1. Result Filtration

PEAKS provides an extensive set of tools to filter LFQ results according to the users' requirements.



Use the **Edit...** button to pop up the corresponding filter dialogs to configure filters:

Feature filters

Significance	≥	0	▼
Fold change	≥	1	▼
Quality	≥	0	▼
Avg. Area	≥	0E0	▼
With peptide ID		<input checked="" type="checkbox"/>	
Charge between		1 ▼ and 10 ▼	
Have at least		1 ▼ confident samples	

OK Cancel

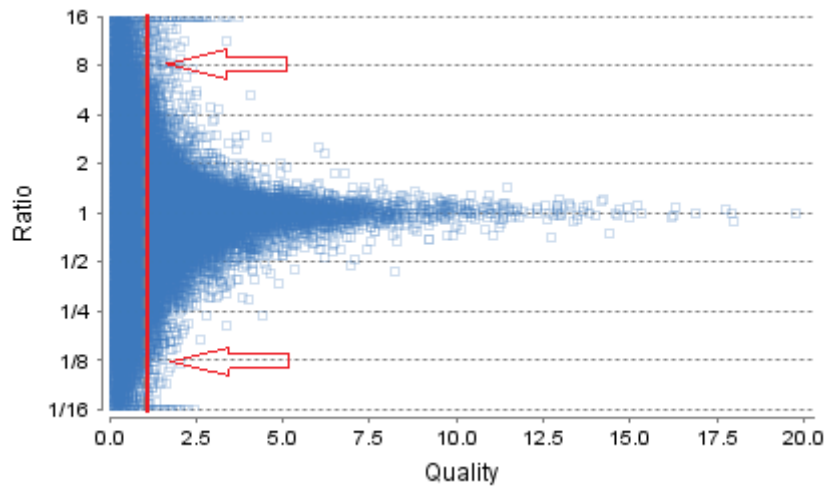
Protein filters

Significance	≥	20	▼
Fold change	≥	2	▼
Unique peptide	≥	1	▼

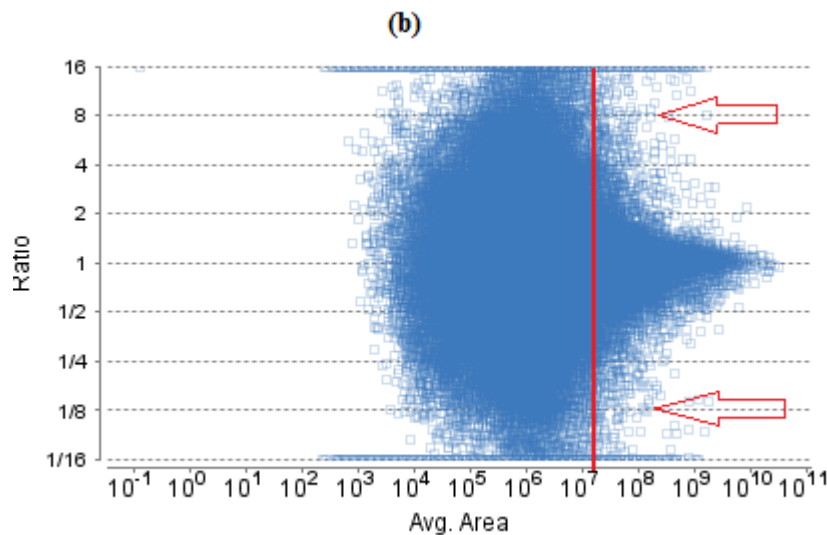
OK Cancel

Usually three steps of filtration need to be performed to obtain meaningful results:

- **Step 1.** The measured intensity of a quantifiable peptide does not change (random variations may be present) when its abundance remains unchanged across the samples. Several filters are designed to effectively locate quantifiable peptide feature vectors:
- **RT range filter:** Feature vectors within the retention time range will be included in the results and everything else will be removed. This filter can be used to remove useless head and tail in an MS runs.
- **Quality filter:** The higher quality of a peptide indicates that the peptide is more quantifiable. A proper cutoff value for quality 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 the figure.



- **Avg. Area filter:** This filter is applied directly to the MS signal intensity. Using the Quality filter is preferred over this filter. However, under some circumstances, it is desirable to apply this filter directly to the MS signal. The rule of thumb is to choose the average area around a fold change of 8 as shown in the figure:



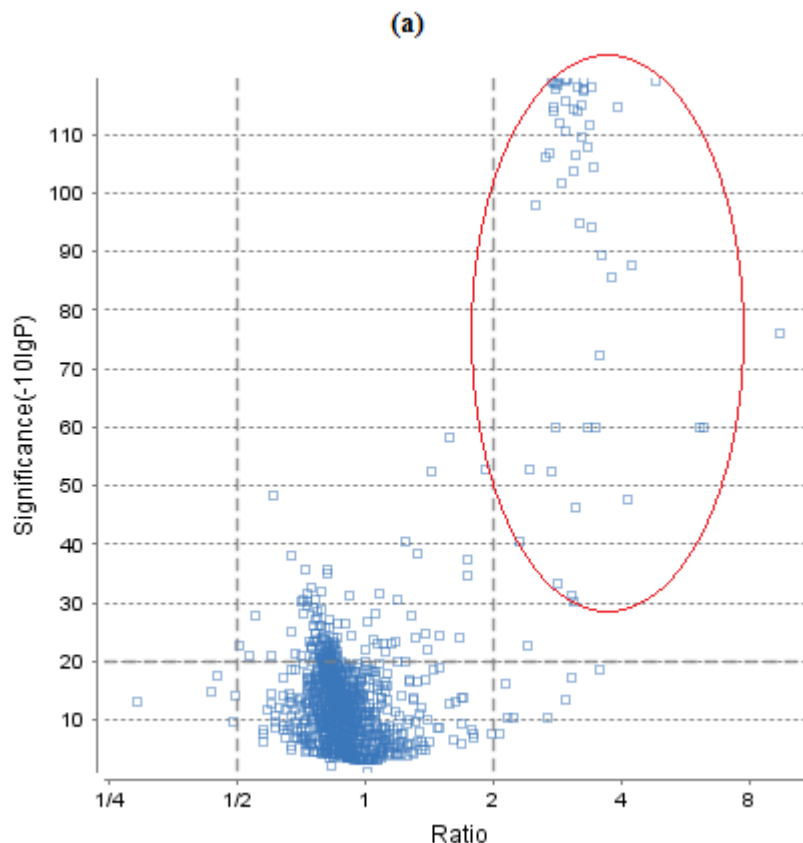
- **Charge filter:** Some low charge states such as charge 1 or high charges like 8 or 9 are more prone to error in feature detection. They can be removed using the charge filter.
- **Confidently detected sample filter:** A peptide is more quantifiable when it is detected in more samples. The rule of thumb is to set this filter to the half the number of total samples.
- **With peptide ID filter:** Set this filter to show only the feature vectors with peptide identification. It is recommended to set this filter for quantification results with attached identification results unless it is desirable to investigate the feature vectors having significant fold changes with no peptide identification attached to them.
- **Step 2A.** The following filters can be used to focus on the peptides with significant changes. These filters are of most use when the purpose is to study the peptide changes.
 - **Peptide feature significance filter:** The significance of a peptide is denoted by its $-10\log P$ score. The rule of thumb is to set the cutoff value at 20 which is equivalent to a P-value of 0.01.
 - **Peptide feature fold change:** This filter is highly dependent on the data set. Usually, a 2-fold change is used as a convention.

- **Use volcano plot to choose proper cutoff:** The volcano plot can be used to choose proper cutoff values for significance and fold change filters. The basic principle is to choose the cutoff values to make sure that only the outliers stay above the threshold in the top left or the top right corner of the volcano plot.

Note

These two filters do not influence protein significance and ratio calculation.

- **Step 2B.** The following filters can be used to focus on the quantifiable proteins. These filters are of most use when the purpose is to study protein changes or differential gene expression.
- **Protein unique peptide filter:** A protein can be quantified more confidently when it has more unique peptides. The convention is to set this filter to 2.
- **Step 3.** The following filters can be used to pick up the proteins (groups) with significant changes.
- **Protein significance filter:** The rule of thumb is to set the cutoff value to 20 as in the case of the peptide significance filter.
- **Protein fold change filter:** Usually a 2-fold change is often used as convention.
- **Use volcano plot to choose proper cutoff:** Volcano plot can be used to choose proper cutoff values for significance and fold change filters. The basic principle is to choose the cutoff values to make sure that only the outliers stay above the threshold in the top left or the top right corner of the volcano plot as shown in the figure.

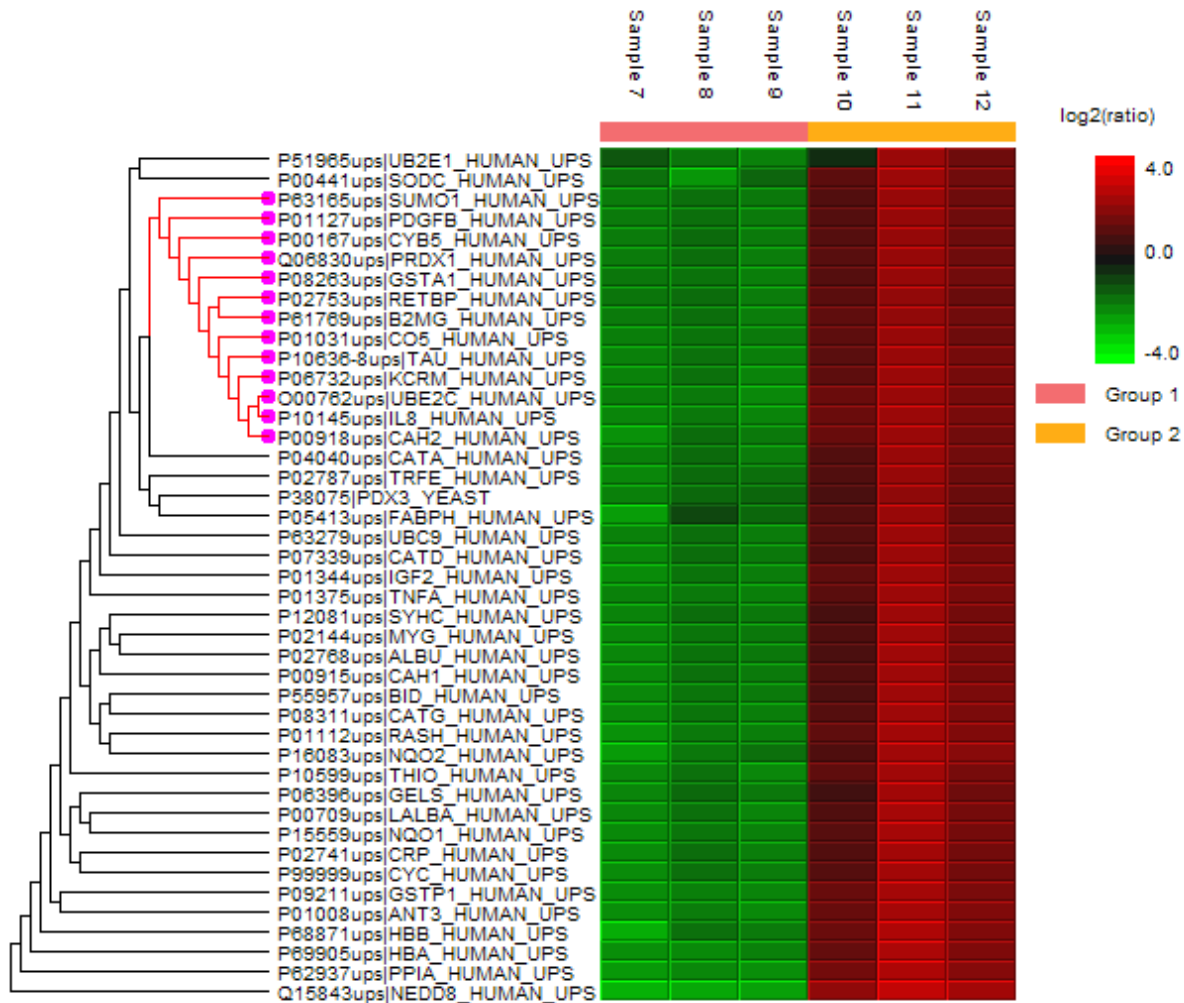


Whenever a filter value is changed, the **Apply Filters** button changes its color as a reminder. Click it to apply the filters.

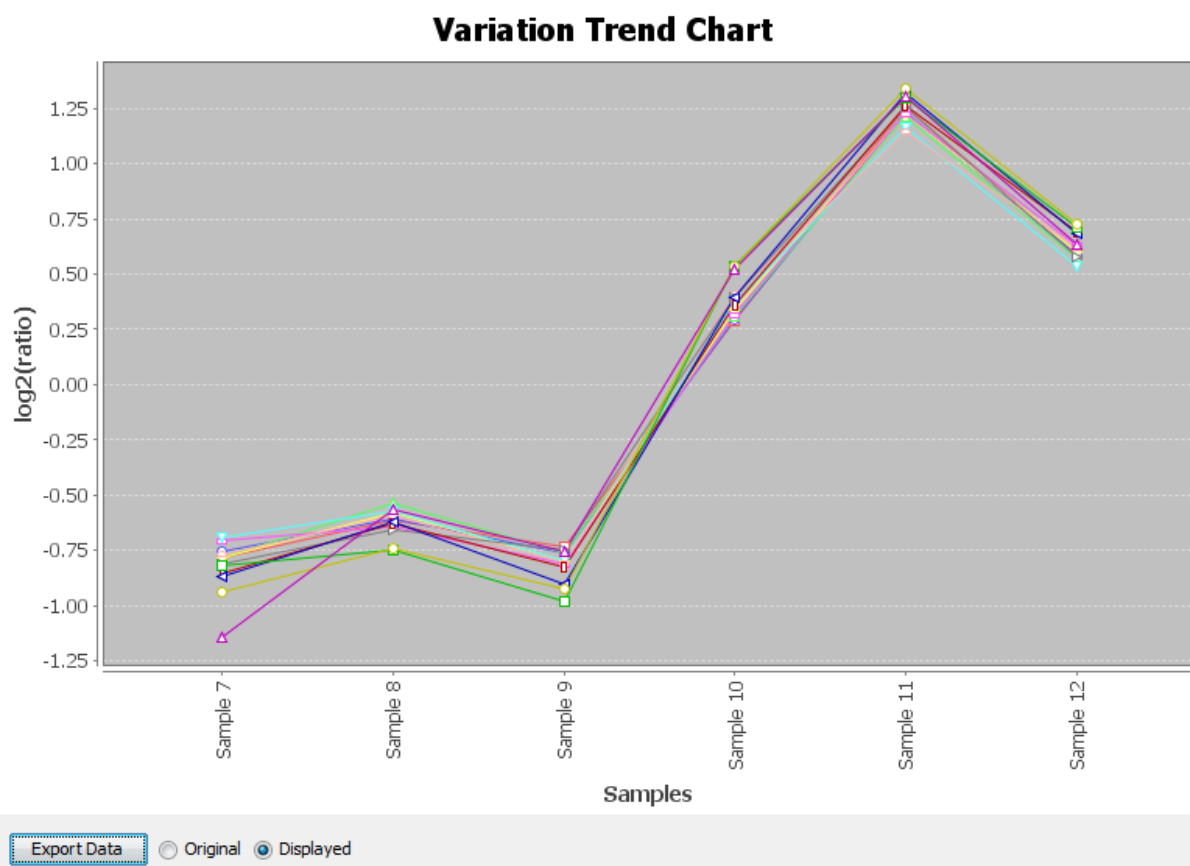
3.1.2. Summary Report

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. The representative proteins are clustered if they exhibit a similar expression trend across the samples. The hierarchical clustering is generated using neighbor-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 pops up in a window when the **Heat Map** button is clicked. Right click on the heat map to save a copy of it.

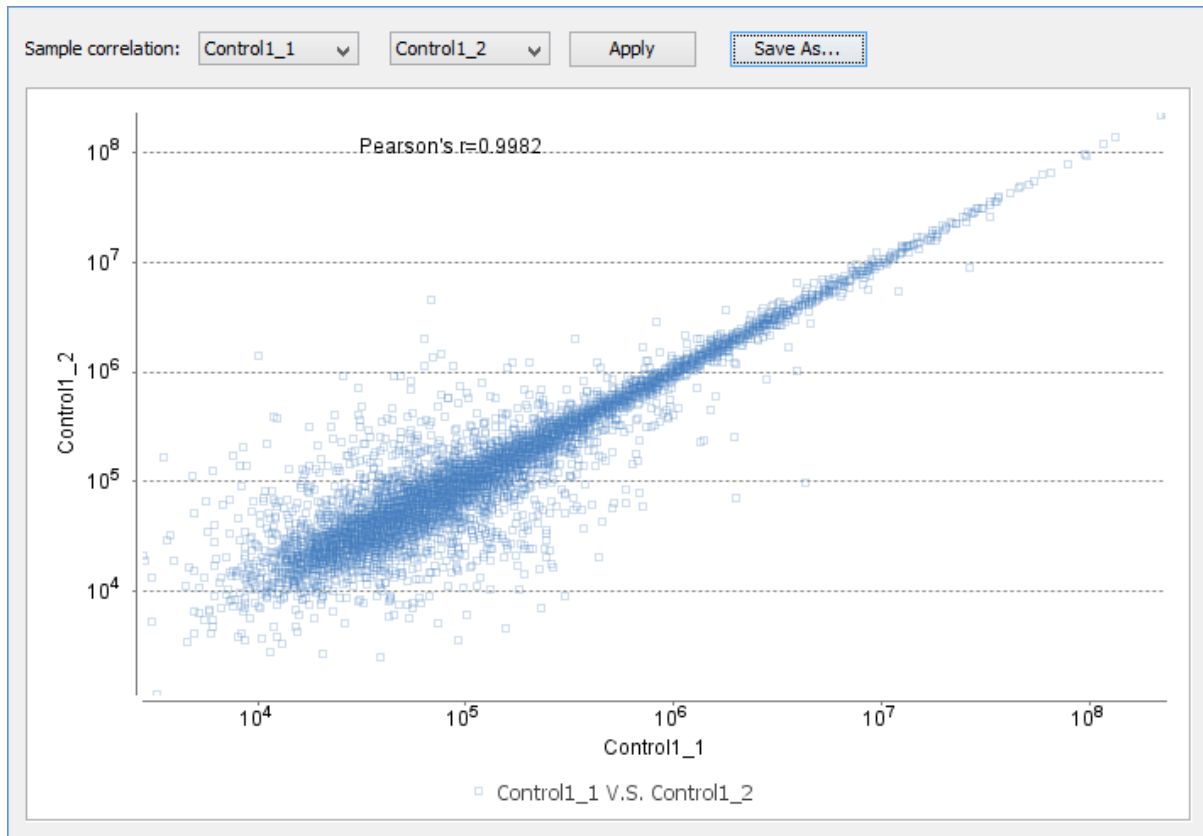


In the interactive heat map, move the mouse to the tree to select a cluster and left click to show the variation trend chart for that cluster. Place the cursor on any point on the variation trend chart to show the protein accession, sample name and the corresponding ratio information in a pop-up window. Click the **Export Data** button to export the variation trend in a comma separated text file. "Original" option exports the original abundance of proteins. Choosing "Displayed" exports the log2 ratio of the sample average.




3.1.2.2. Sample Correlation

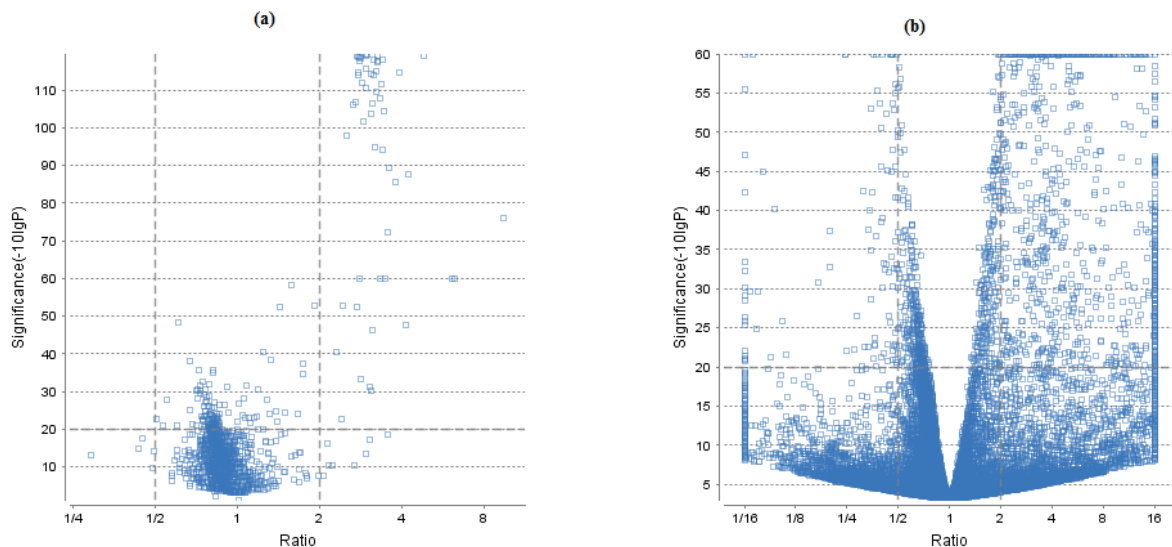
To view the intensity correlation plot for the detected features of the two selected samples, click the **Sample Correlation** button from the Summary view title bar. The correlation plot will be displayed in a pop-up window. Select the samples from the drop down lists beside “Sample correlation” to 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 the **Save As...** button to export the correlation plot to an image 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 that are also statistically significant. The volcano plots in Figures 2(a) and (b) plot significance versus fold-change for proteins and peptides, respectively. These plots help quickly identify the most meaningful changes. There are one horizontal and two vertical threshold lines which divide the plot area into six quadrants. 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 either to the far-left or the far-right. 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 both figures, the ratio is the group ratio to the base group. If more than two groups are present, then the maximum ratio is used in the figures.

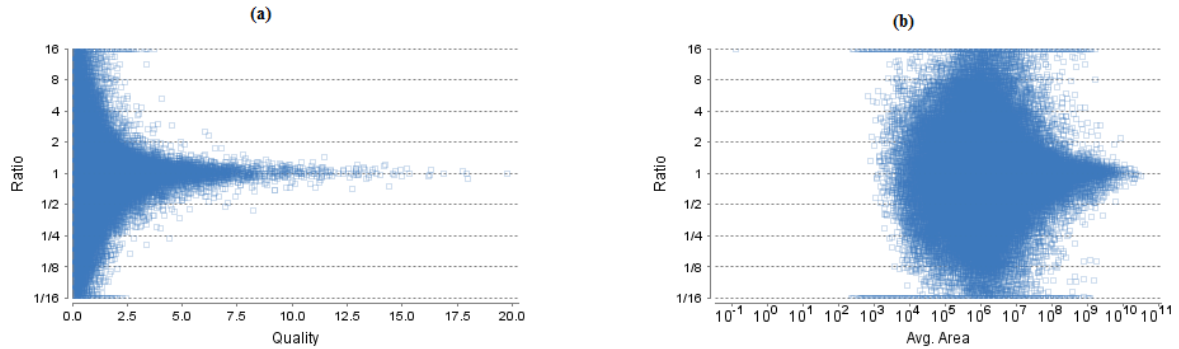
Figure 2. The volcano plots: (a) For proteins. (b) For peptides. 



3.1.2.4. Feature Vector Ratio Distribution


Figures 3(a) and 3(b) plot the distribution of feature vector ratios by quality and average area, respectively. In both figures, the ratio is the group ratio to the base group. If more than two groups are present, then each ratio value depicts a data point in the figures. These plots help to determine the threshold values for the quality and intensity filters to obtain reliable information for the analysis results.

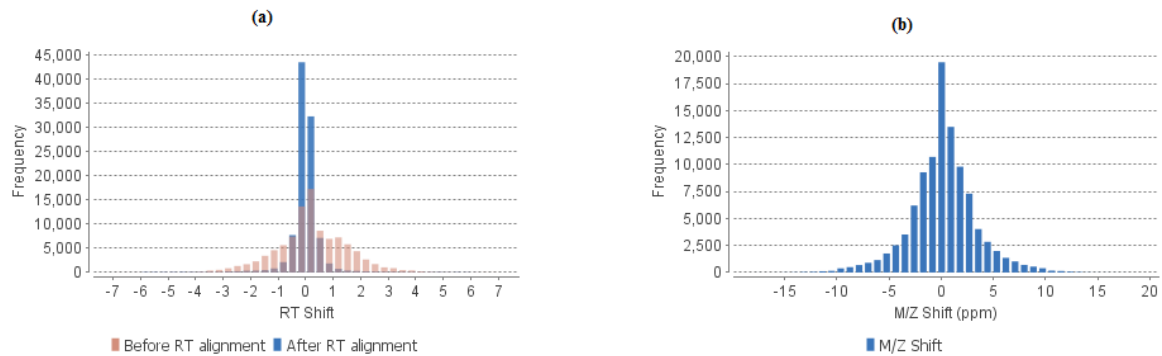
Figure 3. The distribution of feature vector ratio: (a) By quality. (b) By intensity. 



3.1.2.5. RT and m/z Shift Distribution

Figure 4(a) plots the retention time shift distribution to the base sample. The distribution of the m/z shift (in ppm) of the data with respect to the base sample is plotted in Figure 4(b). 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 filtration parameters and the label free quantification analysis parameters, respectively. Table 2 contains statistics regarding the feature vectors and protein groups in the filtered result.

Table 1. Result filtration parameters.

Retention time	≥ 0
Retention time	≤ 75
Feature fold change	≥ 1
Quality	≥ 0
Avg. Area	$\geq 0E0$
Charge	≥ 1
Charge	≤ 10
Confident sample number	≥ 1
With peptide ID	true
Normalization	Use TIC
Protein significance	≥ 20
Protein fold change	≥ 2
Confident unique supports	≥ 1

Table 2. Statistics of filtered result.

Features	12277
Features with ID	6802
Feature vectors	2913
Feature vectors with ID	2913
Protein groups	226

Table 3. Search Parameters

Quantification type:	Label free quantification
Mass Error Tolerance:	20.0 ppm
Retention Time Shift Tolerance:	3.0 min
Dependent on PID:	9
FDR Threshold:	1%
Samples:	6 samples in 6 groups
	: Sample 1 Sample 2 Sample 3 Sample 4
	: Sample 5 Sample 6
Reference Sample:	Sample 5 (auto detected)
Training Samples:	Sample 1, Sample 3 (auto detected)

3.1.3. Exporting Quantification Results

PEAKS label-free quantification results can be exported to various supported formats. The summary page and the images from various views can also be exported to various supported formats. Refer to Chapter 16, Section 5.2, “Exporting Label-Free Quantification Results” for details.

3.1.4. Additional Functions

3.1.4.1. Base Sample Selection

Use the drop down box to choose the base sample. The base sample will be used as the base when calculating sample ratios. The group in which the base sample belongs to will be used as the base for group ratio calculations. The base sample will also influence the retention time shift and M/Z shift distribution figures.

The screenshot shows the PEAKS software interface with the following elements:

- RT Range:** 0 ≤ RT ≤ 184
- Base Sample:** Sample 8 (highlighted with a red circle)
- Buttons:** Apply Filters, Export, Notes
- Peptide feature:** Significance ≥ 20, Fold change ≥ 2, Quality ≥ 2, Avg. Area ≥ 0, 1 ≤ Charge ≤ 10, detected in at least 1 of 6 samples, and with peptide ID. Button: Edit...
- Protein:** Significance ≥ 20, Fold change ≥ 2, has at least 2 unique peptides. Buttons: Edit..., Normalization Factor, Heatmap, Sample Correlation

3.1.4.2. Normalization

PEAKS offers several options to define the normalization factors. When calculating sample or group ratios, the area of each feature will multiply its corresponding normalization factor. By default, PEAKS uses total ion current (TIC) of the samples to calculate the normalization factors. To change the normalization factors, click on the **Normalization Factor** button on the summary page title bar. The table shows the normalization factor that will be used to display the LFQ results. To set the normalization factors manually, select the “Manual input” option and enter the factor values in the table. To use the original feature areas without normalization, select the “No normalization” option. To use internal standard protein list to calculate normalization factors, select the “Use internal standard proteins” option and then select a set of proteins by checking the check box from the list. Right click to check all selected protein or clear selections.

Normalization

☐ Use TIC
 ☒ Use internal standard proteins
 ☐ Manual input
 ☐ No normalization

Sample	Factor
Sample 7	1.0
Sample 8	1.0
Sample 9	1.0
Sample 10	1.0
Sample 11	1.0
Sample 12	1.0

Protein List

search

Previous Next

Protein	
P00549 KPYK1_YEAST	<input type="checkbox"/>
P52489 KPYK2_YEAST	<input type="checkbox"/>
P10592 HSP72_YEAST	<input type="checkbox"/>
P10591 HSP71_YEAST	<input type="checkbox"/>
P00359 G3P3_YEAST	<input type="checkbox"/>
P00925 ENO2_YEAST	<input type="checkbox"/>
P16521 EF3A_YEAST	<input type="checkbox"/>
P53978 EF3B_YEAST	<input type="checkbox"/>
P27466 KCC1_YEAST	<input type="checkbox"/>
P00358 G3P2_YEAST	<input type="checkbox"/>
P07259 PYR1_YEAST	<input type="checkbox"/>
P00560 PGK_YEAST	<input type="checkbox"/>
P15108 HSC82_YEAST	<input type="checkbox"/>
P02829 HSP82_YEAST	<input type="checkbox"/>

0 protein selected.

(Right click to show popup menu)

OK Cancel

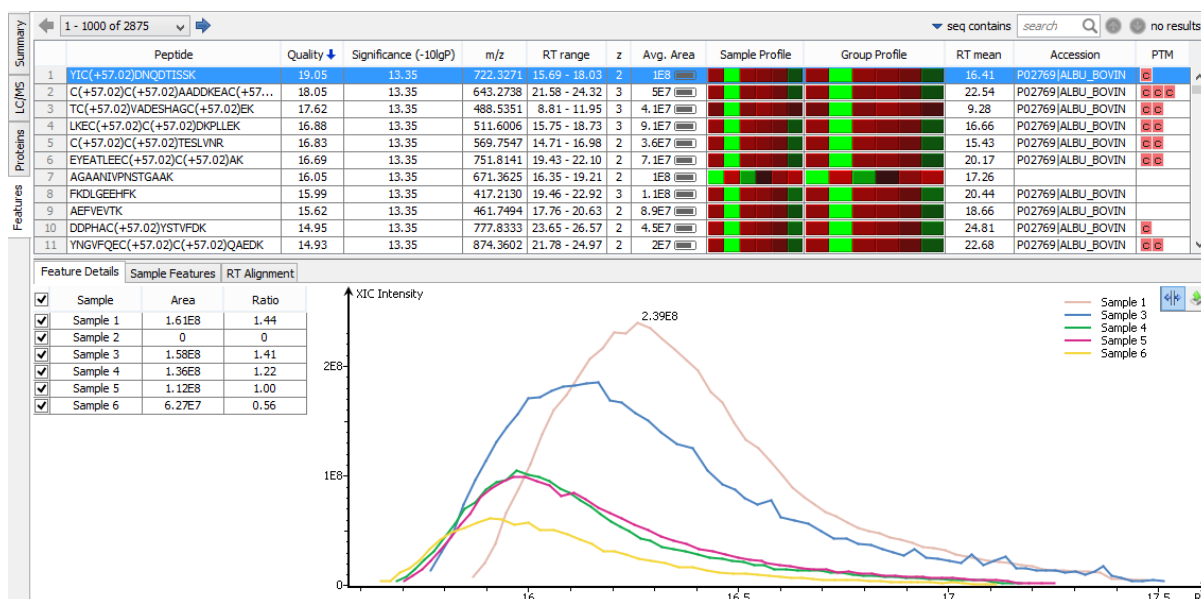
3.2. LC/MS View

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.

3.3. Feature View

The Feature view lists detailed information about all the quantified feature vectors. It has four components:

- **Feature Table:** Lists all the quantified feature vectors present in the samples
- **Feature Details:** Provides detailed information about the selected feature vector.
- **Sample Features:** Displays detected features in each sample at the LC/MS level.
- **RT Alignment:** Shows retention time alignment among the samples highlighting the RT range of the selected feature.



3.3.1. Feature Table

The quantified features are listed in the features table. This table contains the following columns:

- **Peptide:** The amino acid sequence of the peptide. For a PTM on an amino acid, the amino acid code is followed by a pair of parentheses enclosing the delta mass of the PTM.
- **Quality:** This defines how well the peptide can be quantified. 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 σ to be the estimated intensity fold change of the peptide across the samples when the peptide abundance does not change. The quality is calculated

$$\frac{1}{\log_2 \sigma}$$

as $\frac{1}{\log_2 \sigma}$. Thus, if a peptide has an expected 2-fold change then its quality will be 1; if a peptide has an expected 4-fold change then its quality will be 0.5. A high quality value implies that the peptide can be well quantified.

- **Significance:** It is assumed 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 2-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. $-10\log_{10}P - \text{value}$ is finally 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:** Precursor charge.
- **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 relative abundance is color coded as a heat map. The tooltip displays the ratio to the base sample. Press Ctrl to fix the tool tip window in case the sample list is too long and scrolling is needed.
- **Group Profile:** The relative abundance is color coded as a heat map. The tooltip displays the ratio to the base group.

- **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:** The accession of the protein from which the peptide comes. If the peptide has more than one protein associated with it, only one will be displayed.
- **PTM:** The post-translational modifications appear in 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 the area and 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”

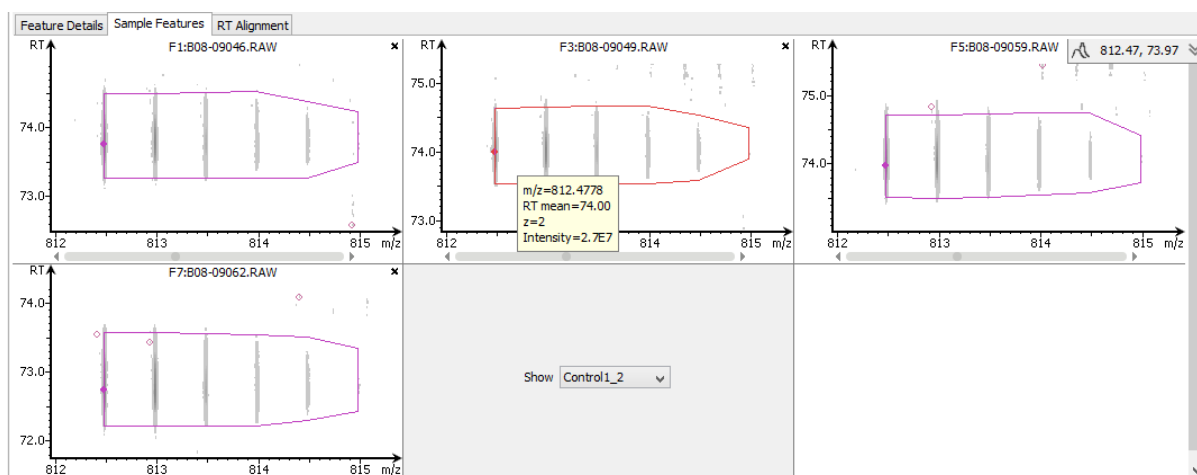
tabs. The  button at the top-right corner of the tab can be used to toggle the plot data between before and after RT alignment.

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.

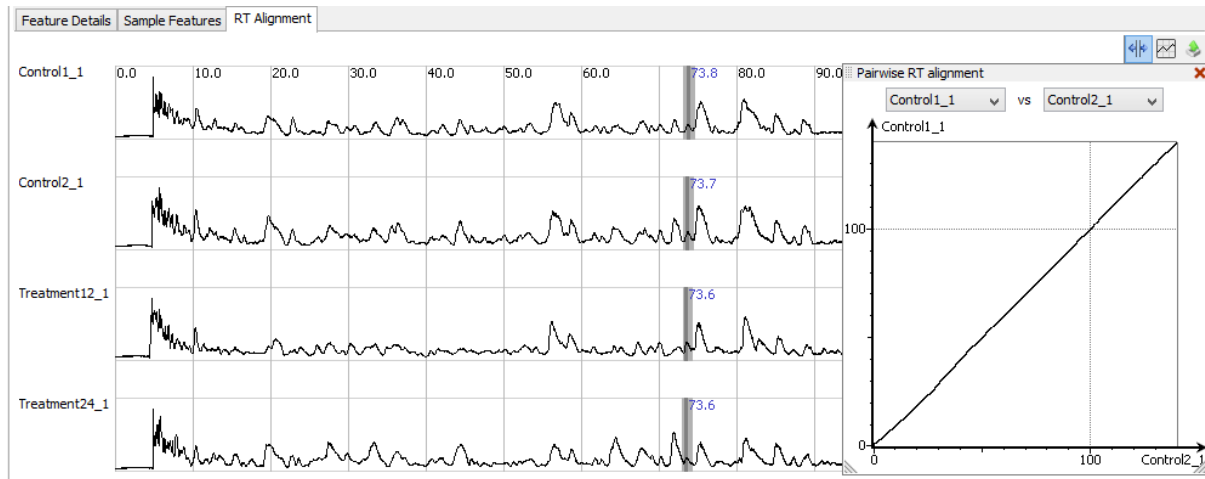
3.3.3. Sample Features

This tab displays a close up LC/MS view of the detected features from the selected samples. The feature boundary is shown by a bounding polygon. The center of the feature is denoted by a solid red dot on the mono-isotopic peak cluster. Placing the mouse cursor on this center of a feature pops up a tool tip that shows the m/z value, mean retention time, charge state, and 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 LC/MS view and the 3D LC/MS view.



3.3.4. RT Alignment

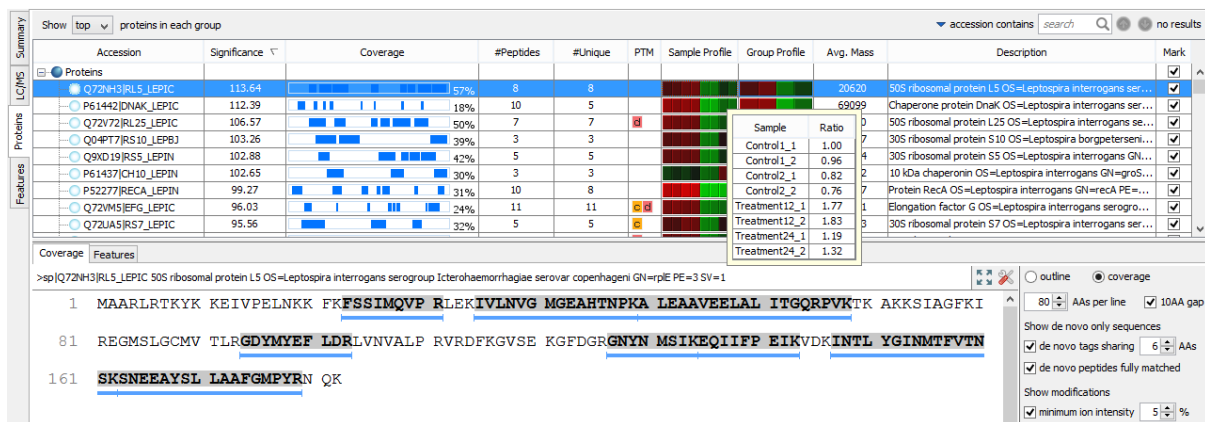
This tab displays the aligned TIC curves of the selected samples with the highlighted retention time range of the selected feature. Use the toggle switch in the top-right corner to switch between the before and after RT alignment view. Click the  button to view the pairwise RT alignment in a pop-up window. Select the sample pair from the drop-down list to view the desired pairwise RT alignment.



3.4. 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:** Lists all the quantified proteins present in the sample.
- **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 peptide features identified from the selected protein in the protein table.



3.4.1. Protein Table

The quantified proteins are listed in the protein table with homologous proteins clustered together. This protein table is same as other protein tables found in the PEAKS DB, PEAKS PTM, and SPIDER results except the following columns as mentioned below:

- **Significance:** The protein significance is inferred directly from its supporting peptides significance weighted by two factors: the intensity rank and the correlation between the relative abundance of the protein and the relative abundance of its supporting peptides.
- **Sample Profile:** The relative protein abundance among the samples is depicted as a heat map. Place the cursor on a profile to view the abundance ratios in the tool tip window. Press Ctrl to fix the tool tip window in case the sample list is too long and scrolling is needed.
- **Group Profile:** Similar to the sample profile, the group profile depicts the relative protein abundance between the groups.

See Chapter 9, Section 3.3.1, “Protein Table” for more details on other entries of the protein table.

3.4.2. Coverage Tab

The coverage tab characterizes the selected protein sequences at the amino acid level. Refer to Chapter 9, Section 3.3.2, “Protein Coverage” for details.

3.4.3. Features Tab

The supporting peptide features assigned to the protein are shown under the “Features” tab.

Coverage		Features													
		Peptide	Unique	Quality	Significance	m/z	RT range	z	Avg. Area	Sample Profile	Group Profile	RT mean	Start	End	Used
1	<input checked="" type="checkbox"/>	R.IINEPTAAALAYGFDK.K		11.14	60.00	847.4564	65.39 - 69.41	2	4.4E6			67.02	168	183	
2	<input checked="" type="checkbox"/>	K.FETSAGEFTPQEISAR.V	•	6.67	60.00	885.4336	40.05 - 45.68	2	3.8E6			43.11	103	118	•
3	<input checked="" type="checkbox"/>	K.NQAITNAVNTIR.S		11.88	60.00	657.8650	19.87 - 25.86	2	2.7E6			23.21	58	69	
4	<input checked="" type="checkbox"/>	K.NELDTLTYSLEK.T	•	9.50	60.00	713.3655	57.26 - 61.82	2	2.4E6			59.51	533	544	•
5	<input checked="" type="checkbox"/>	K.AVITVPAYFNDEQR.Q	•	9.47	60.00	811.9212	54.10 - 60.31	2	1.6E6			56.66	138	151	•
6	<input checked="" type="checkbox"/>	K.HLDMTLTR.A		7.89	60.00	493.7610	8.54 - 11.26	2	8.2E5			9.66	291	298	
7	<input checked="" type="checkbox"/>	R.IPAVQELVK.Q		6.37	60.00	498.8108	30.12 - 37.13	2	6.7E5			33.85	342	350	
8	<input checked="" type="checkbox"/>	R.FDLIGIPPAPR.G	•	6.55	59.23	598.3488	64.29 - 71.27	2	6E5			67.53	453	463	
9	<input checked="" type="checkbox"/>	K.GETLIGQFAK.N	•	5.06	60.00	532.2953	32.36 - 38.39	2	3E5			35.78	48	57	
10	<input checked="" type="checkbox"/>	R.TTPSIVAFPAK.G		2.83	11.42	568.3251	30.94 - 40.80	2	2.9E5			35.63	37	47	

This table is similar to the feature table in the feature view. The extra columns are:

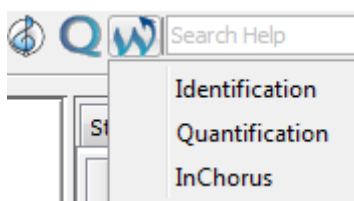
- **(Checkbox):** Select the peptides to appear within the Coverage Tab as blue bars.
- **Unique:** whether this peptide is unique to the current protein group.
- **Start:** the start position of the peptide in the protein.
- **End:** the end position (inclusive) of the peptide in the protein.
- **Used:** whether this peptide is used to calculate the protein profile.

To view the actual abundance of the peptide features among the samples or groups, move the cursor on 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 feature view. Right click on the peptide feature to show the pop-up menu similar to the one in the PEAKS DB supporting peptides tab.

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.

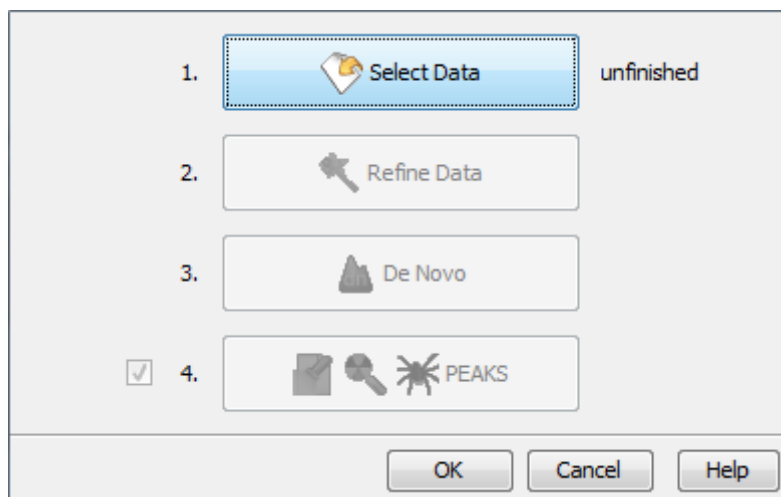
For convenience, PEAKS also provides another mechanism to define workflows for protein identification, quantification and inChorus search (multi-engine protein ID). This workflow can be applied to multiple projects that can perform PEAKS analyses at the fraction, sample or project levels. This tool is especially helpful for analyses on multiple projects using the same parameters.



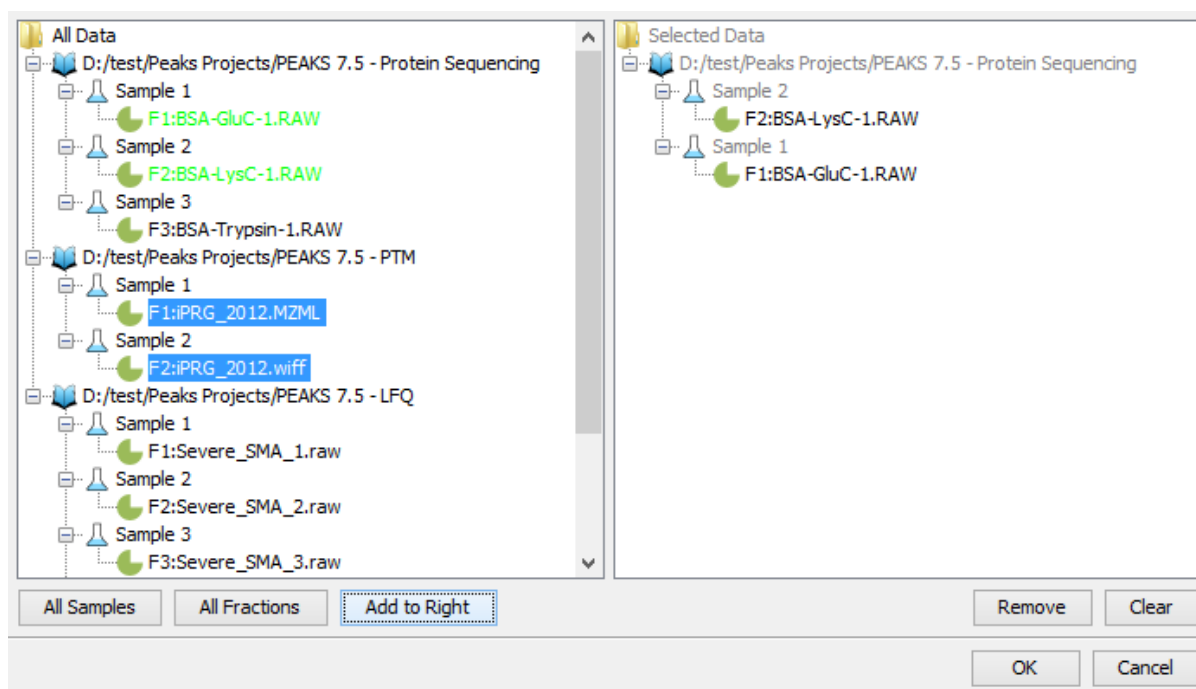
Once a specific workflow is selected, a dialog pops up to specify the analysis steps and the parameters to use in each step.

1. Identification Workflow

Click on 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 the All Samples or the All Fractions buttons. Then, click the Add to Right button to transfer the samples/files to the “Selected Data” list on the right hand 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 loaded in a single workflow will be processed in exactly the same way, using the exact same parameters. If you want to run some differently than others, then you must set up separate workflows.

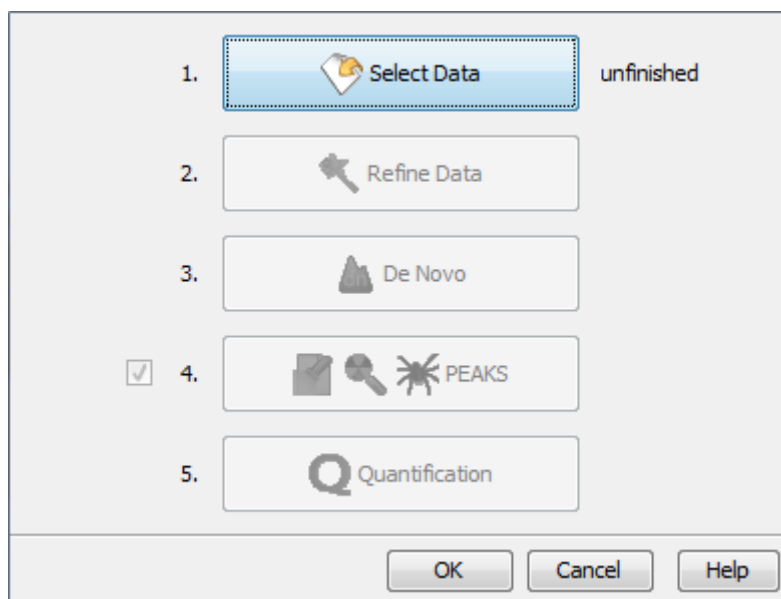
Once the data is selected, the parameters for the identification analysis steps can be specified one by one by clicking the other buttons in the workflow dialogue. Please refer to the chapters on each individual function for details on setting up the function-specific parameters.

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 with 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.

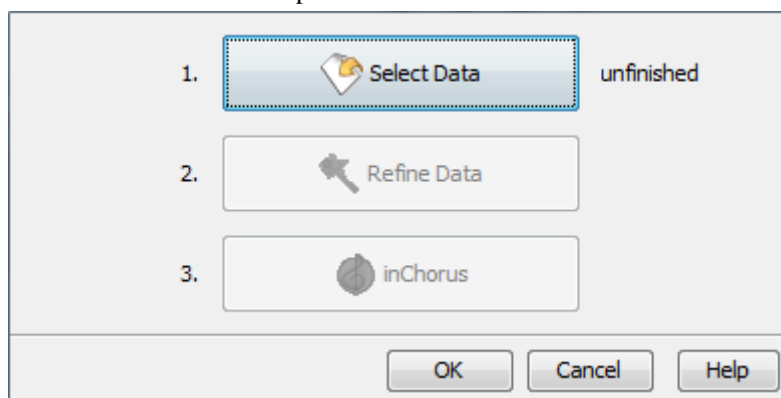


Note

Only Reporter Ion Quantification (eg. iTRAQ) or Precursor Ion Quantification (eg. SILAC) workflows can be set in the quantification workflow. For Label Free Quantification 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.



Note

The inChorus workflow allows only new database searches using the supported search engines. To run an inChorus using an existing search result in the project or to import a search result, use the inChorus

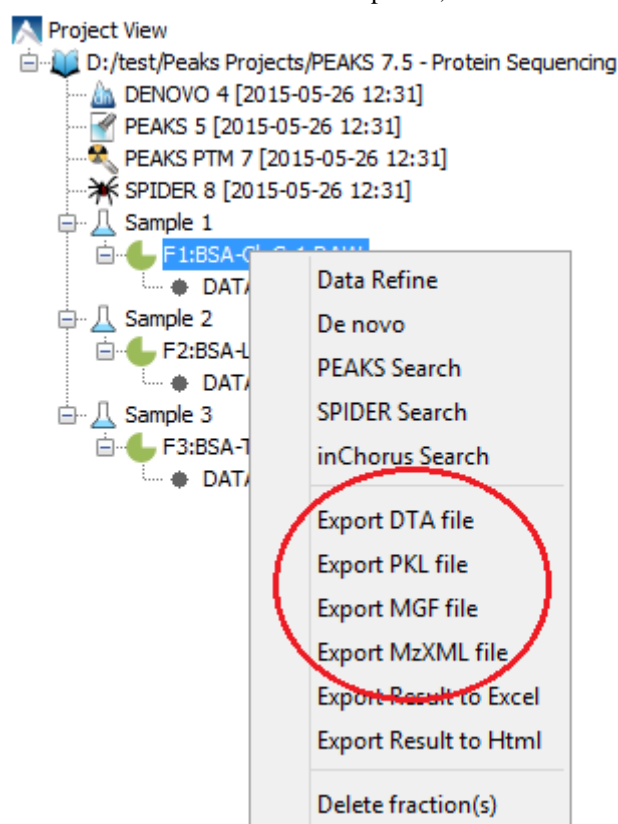
 search.

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 workflow. 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 it 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 dialog prompting for the folder name and location to which the DTA files will be exported. For MGF and PKL exports, the dialog will ask for a name and a location for the file.

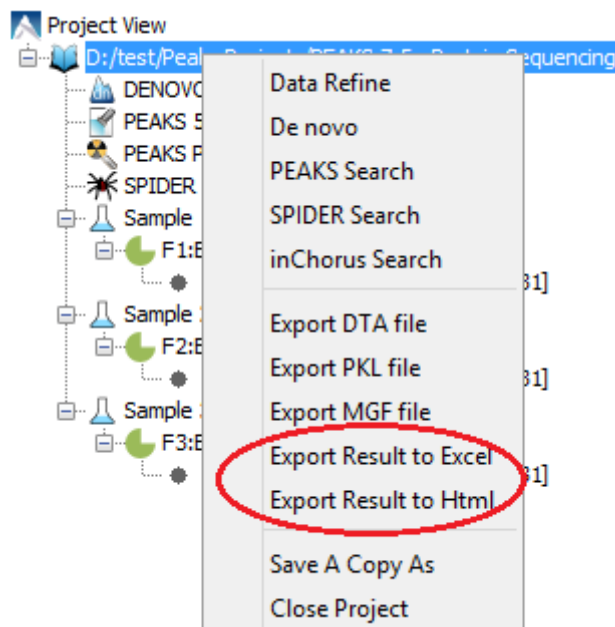
Selecting Export MzXML File command will open the “Export mzXML File” dialog.

A screenshot of the 'Export mzXML File' dialog box. It has a light gray background. At the top, there are two text input fields: 'Start RT: 0' and 'End RT: 120'. Below these is a 'Save as:' label followed by a text input field containing 'D:\test\BSA-GluC-1.mzxml' and a 'Browse' button. At the bottom, there are two buttons: 'OK' and 'Cancel'.

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 click 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.



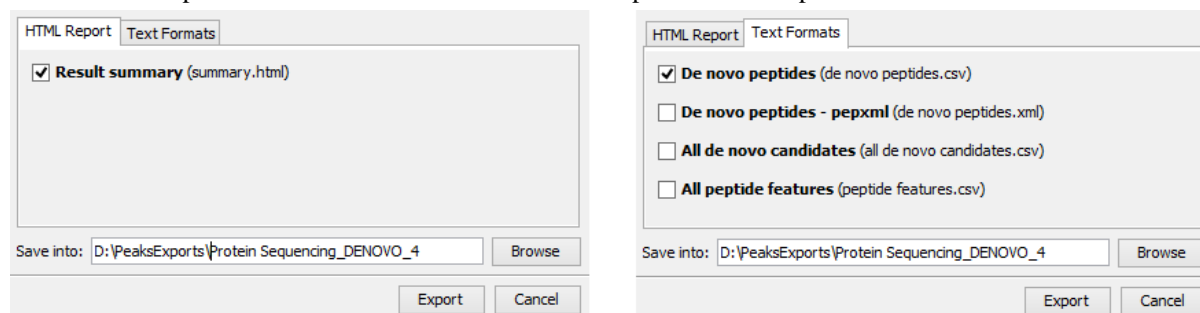
More function specific export options are available from the “Summary” view of the respective results. The following sections describe in detail the function-specific exporting options available in various formats.

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 dialog 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:

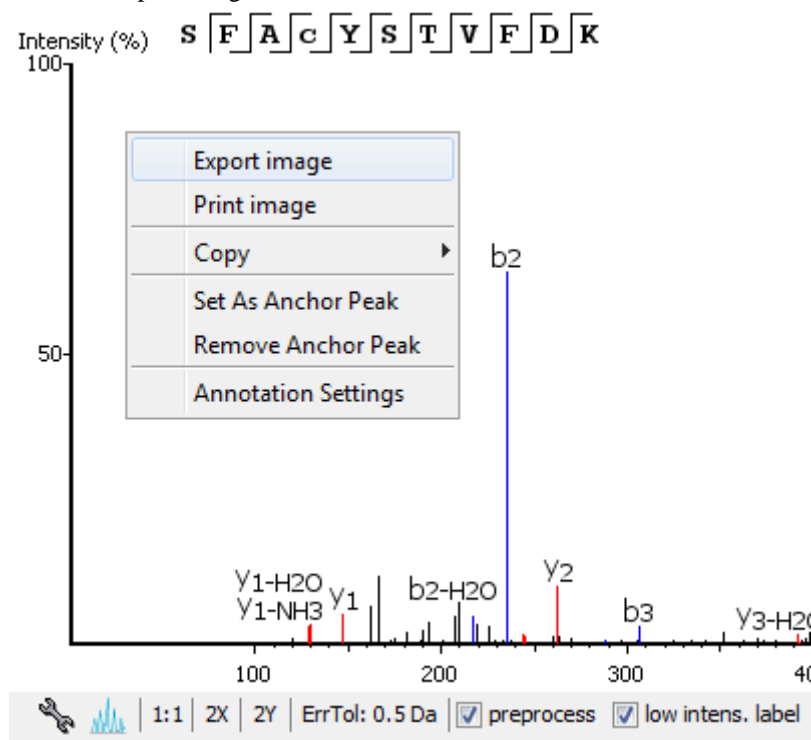
- **Result summary** : The “Summary” view page will be saved as `summary.html` file in HTML format in the specified location.

Text Formats. The following exporting options are available in various text formats:

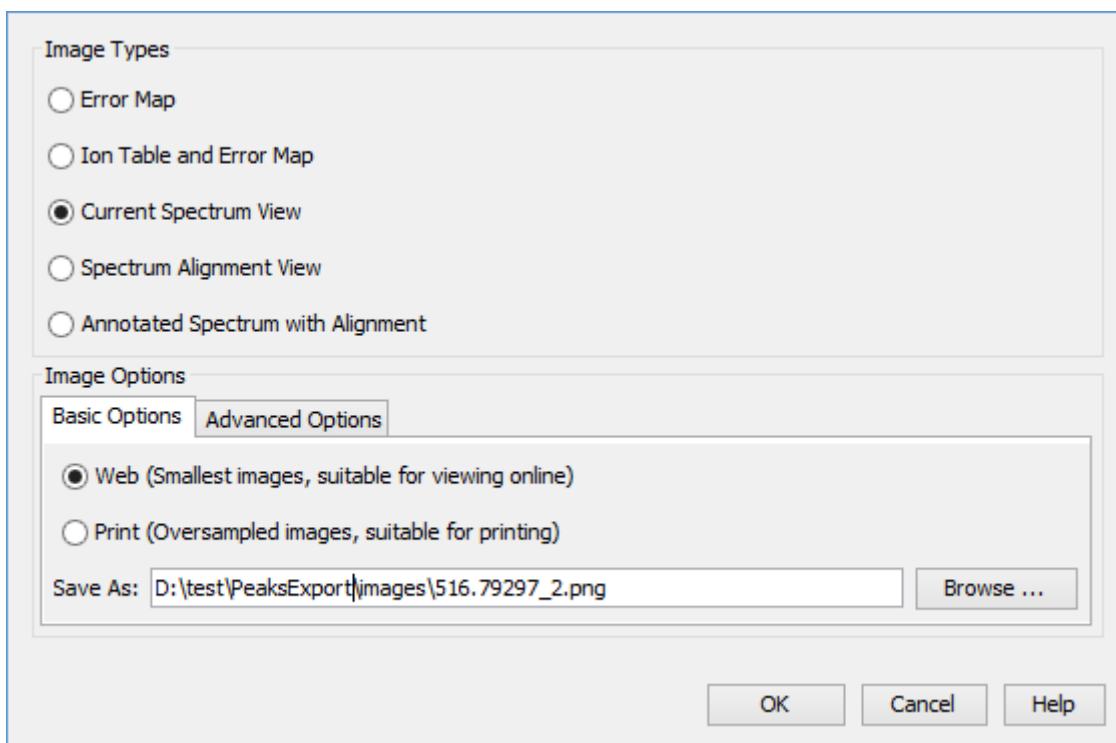
- **De novo peptides** : The peptides identified by *de novo* sequencing will be saved to `de novo peptides.csv` file in Comma Separated Values (CSV) format in the specified folder.
- **De novo peptides - pepxml** : In addition to CSV format, the peptides can be saved in pepXML format.
- **All de novo candidates** : All *de novo* candidates for each spectrum will be saved to `all de novo candidates.csv` file in CSV format.
- **All peptide features** : Identified peptide features for each fraction will be saved to `peptide features_??`.csv file in CSV format, where “??” is replaced by the corresponding fraction id.

3.2. Exporting Images

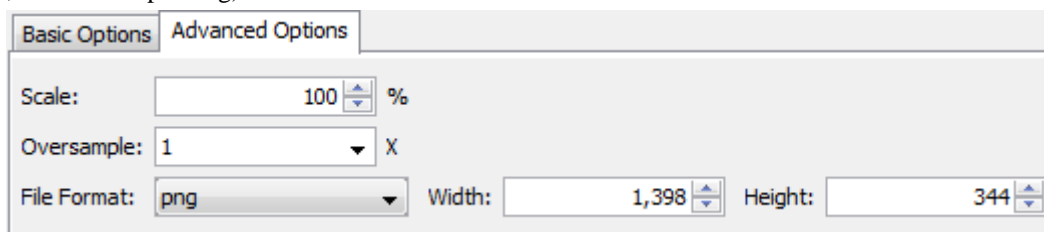
The annotated Spectrum, Ion Match table, Error Map, and Spectrum Alignment can be exported to image files. To do so, position the cursor on any of those items in the result panel and click the right mouse button to view the pop-up menu and select the Export Image command from the menu.



This will bring up the “Export Images” dialog for selecting the result items to export.



Select the desired result elements from the “Image Types” list. The “Basic Options” tab offers choices for the location to save and the image size: “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, file format, resolution, and 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 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 dialog will appear.

The image displays three screenshots of the PEAKS export dialog box, showing different tabs and their respective options.

HTML Report Tab: This tab allows users to select the content to include in the HTML report. The options are:

- ☒ summary view
- ☒ protein coverage
- ☒ supporting peptides
- ☐ best unique PSM
- ☒ Collect all the above details in a single html

 The "Save into:" field is set to "D:\PeaksExports\Protein Sequencing_DBSEARCH_5".

Text Formats Tab: This tab allows users to select the output format for the data. The options are:

- ☒ Proteins (proteins.csv)
- ☒ Supporting peptides (protein-peptides.csv)
- ☒ DB search peptides (peptide.csv)
- ☒ DB search peptide-spectrum matches (DB search psm.csv)
- ☒ De novo only peptides (de novo only peptides.csv)
- ☐ Proteins - fasta (proteins.fasta)
- ☐ Peptides - pepxml (peptides.pep.xml)
- ☐ Peptides - mzidentml (1.0.0) (peptides_1_0_0.mzid)
- ☐ Peptides - mzidentml (1.1.0) (peptides_1_1_0.mzid)
- ☐ All peptide features (peptide features.csv)
- ☐ De novo only peptides - pepxml (de novo only peptides.xml)
- ☐ De novo only spectra - mgf (de novo only spectra.mgf)

 The "Save into:" field is set to "D:\PeaksExports\Protein Sequencing_DBSEARCH_5".

For Third Party Tab: This tab allows users to select the output format for the data. The options are:

- ☐ peptides - pepXML (peptides.pep.xml)
- ☐ peptides - mzIdentML (peptides_1_0_0.mzid)
- ☐ peptides - mzIdentML (peptides_1_1_0.mzid)

 The "Save into:" field is set to "D:\PeaksExports\Protein Sequencing_DBSEARCH_5".

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.

An individual protein will have its own HTML output file where the corresponding protein coverage, 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 page) 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 page) 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 hence can be reported multiple times in this file.
- **DB search peptides:** A list of peptides from all fractions involved in the database search, filtered by the peptide filters, will be exported to `peptide.csv`. The features that belong to the same peptide within the same fraction are grouped together to calculate peptide areas in the corresponding fractions.
- **DB search peptide-spectrum matches:** The selected peptide-spectrum matches (PSM), filtered by the peptide filters in the Summary page, 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 page.
- **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`.
- **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.
- **Peptides - pepxml:** A list of peptide-spectrum matches, filtered by the peptide filters in the Summary page, 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 page. 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).
- **All peptide features :** Identified peptide features for each fraction will be saved to `peptide features_???.csv` file in CSV format, where “??” is replaced by the corresponding fraction id.
- **De novo only peptides - pepxml:** A list of good *de novo* sequences with no good database matches or no matches at all will be saved in `de novo only peptides.xml` in pepXML format.
- **De novo only spectra - mgf:** The collection of de novo only spectra (tandem scans with good de novo sequences but no good database matches or no database matches at all) will be saved in `de novo only spectra.mgf` in MGF format. This MGF file can be loaded in another PEAKS project or other search engines for further analysis.

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 version 1.0.0 and version 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 like PTM verification, for example.

Note

Peptides identified by SPIDER may contain residue substitution, insertion, and deletion. 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.

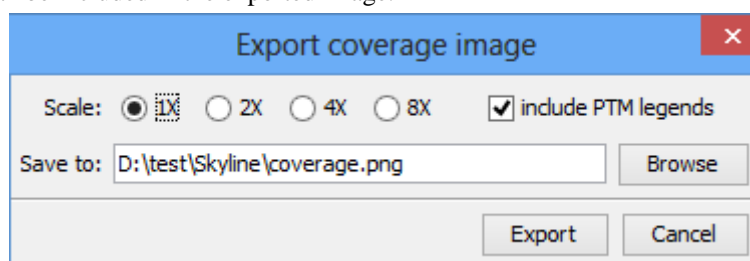
- **For Skyline: peptides - pepXML:** In addition to exporting the peptide-spectrum matches to pepXML format, PEAKS will export the spectrum data in mzXML format. The exported pepXML file and mzXML files can be loaded into Skyline for spectrum library generation.

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 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 of 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.

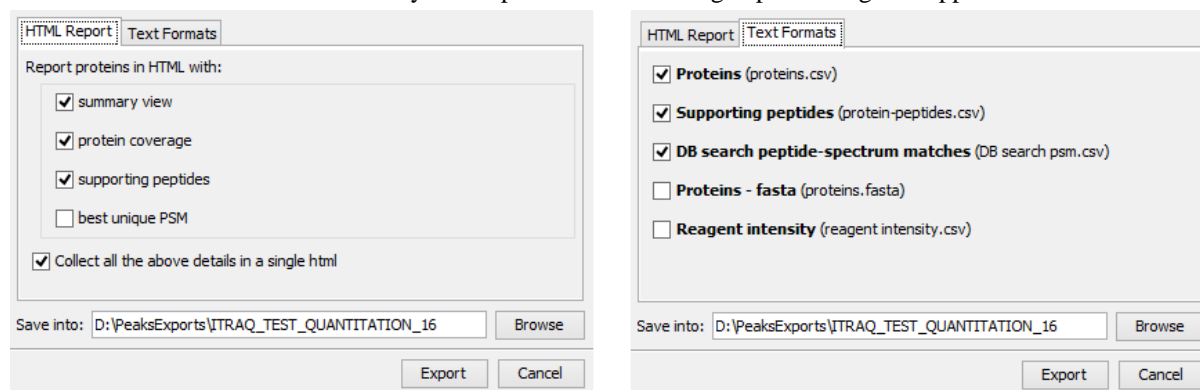


5. Exporting Quantification Results

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 Labeled 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 dialog 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. See Section 4.1, “Exporting Summary, Proteins, and Peptides” for details.

Text Formats: The text format exporting options are similar to those available for PEAKS DB. For PEAKS Q, except that only the quantified proteins and peptides are used in the exporting. The `reagent intensity.csv` file saves the absolute reagent intensity while `DB search psm.csv` exports reagent ratio or intensity based on

option selected from the Show dropdown list. See Section 4.1, “Exporting Summary, Proteins, and Peptides” for details of other options.

Select the output location and click the Export button to save the selected result components to the specified location.

Export Images. The Annotated Spectrum, Ion Match table, Error Map, 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 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 dialog will appear.

The image shows two side-by-side screenshots of the PEAKS LFQ export dialog. The left screenshot shows the 'HTML Report' tab with the following options: 'Report proteins in HTML with:' (summary view, protein coverage, supporting peptides, best unique PSM), 'Collect all the above details in a single html' (checked), and 'Save into: D:\test\PeaksExport\LFQ sample project_LABELFREE_39'. The right screenshot shows the 'Text Formats' tab with the following options: 'Proteins (proteins.csv)' (checked), 'Supporting peptides (protein-peptides.csv)' (checked), 'Proteins - fasta (proteins.fasta)' (unchecked), 'Feature list (feature.csv)' (unchecked), 'Normalization factor (normalization_factor.csv)' (unchecked), and 'Save into: D:\test\PeaksExport\LFQ sample project_LABELFREE_39'. Both dialogs have 'Export' and 'Cancel' buttons.

HTML Report: The options are the same as those for labeled quantification. See Section 5.1, “Exporting Labeled Quantification Results” for details.

Text Formats: The detected features will be saved as the file `feature.csv` file in CSV format. The normalization factor used to calculate the quantification ratios can be exported to the file `normalization_factor.csv` file in CSV format. The other options are the same as for Labeled Quantification. See Section 4.1, “Exporting Summary, Proteins, and Peptides” for details.

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, LC/MS view of the sample features, and 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

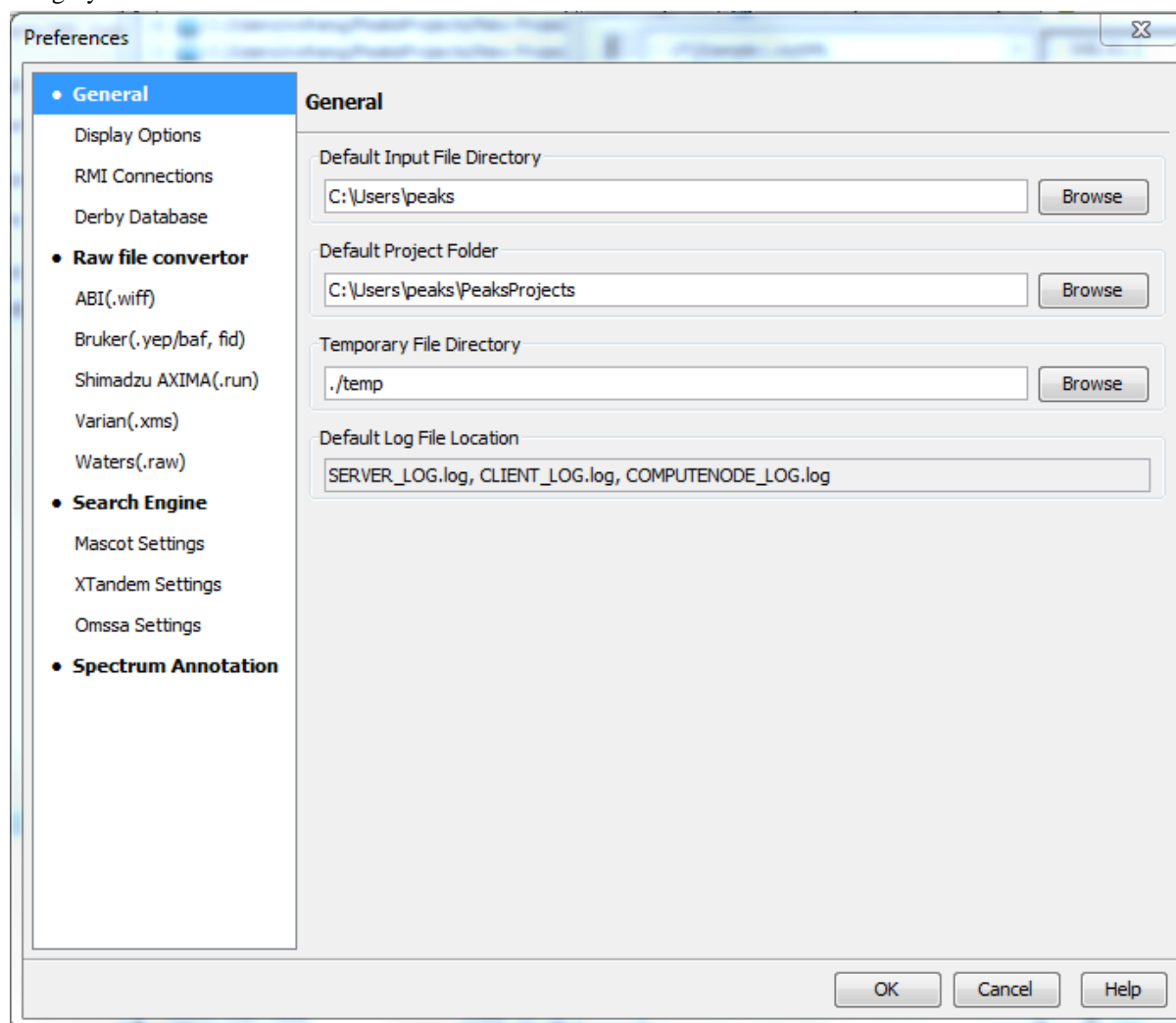
The inChorus exporting function contains the same exporting options as PEAKS DB. Unlike PEAKS DB, the exported results will contain inChorus scores and individual search engine scores for the supporting peptides of the protein identifications. See Section 4, “Exporting Database Search Results” for details.

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 Window | Preferences command from the menubar to open the “Preferences” dialog. Select the preferences category from 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.

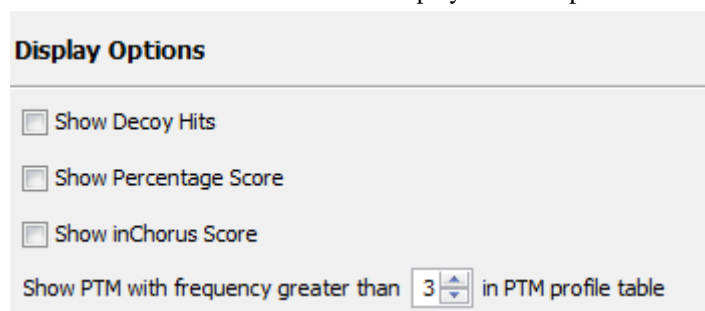
Default Project Folder: PEAKS uses [USER_HOME]/PeaksProjects as the default folder for project files, where USER_HOME is the user home directory in your system. Select the Browse button to change this location. Please make sure this directory is readable/writable by PEAKS.

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. These files' locations 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.



The screenshot shows a window titled "Display Options". It contains four checkboxes: "Show Decoy Hits", "Show Percentage Score", and "Show inChorus Score", all of which are currently unchecked. Below these is a label "Show PTM with frequency greater than" followed by a spin box containing the number "3", and then the text "in PTM profile table".

Show Decoy Hits: Check this to display protein and peptide hits from the decoy database in PEAKS DB results.

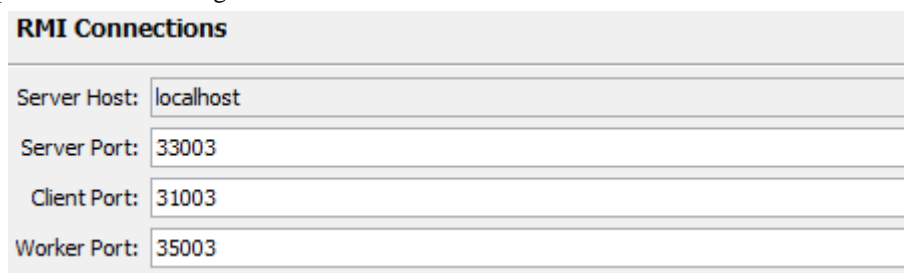
Show Percentage Score: PEAKS uses -10lgP to display its results by default. Check this to view the percentage score along with the -10lgP score in peptide and protein view. These changes are applied to the exported files of PEAKS DB results.

Show inChorus Score: Check this to display the percentage score in the peptide and protein views of 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.

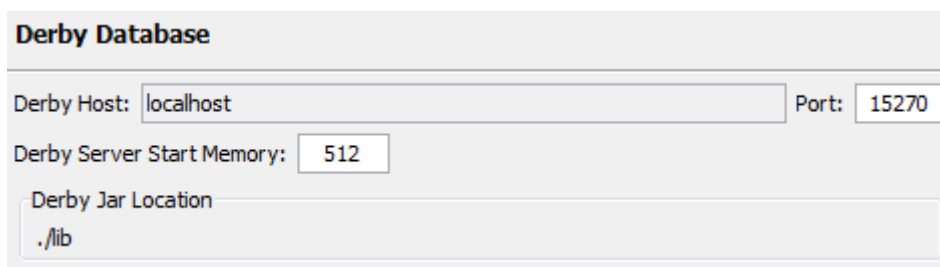


The screenshot shows a window titled "RMI Connections". It contains four text input fields: "Server Host" with the value "localhost", "Server Port" with the value "33003", "Client Port" with the value "31003", and "Worker Port" with the value "35003".

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 Database

Derby Host: Port:

Derby Server Start Memory:

Derby Jar Location

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 Chapter 2, 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 Chapter 4, Section 8.4.1, “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 Chapter 4, 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 Chapter 4, 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 Chapter 4, Section 8.7, “Varian” for details on Varian instrument preferences.

1.2.5. Waters (.raw)

PEAKS can readily load Waters raw file without further configuration. However, this option provides an alternative loading option using `MassWolf` converter. 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

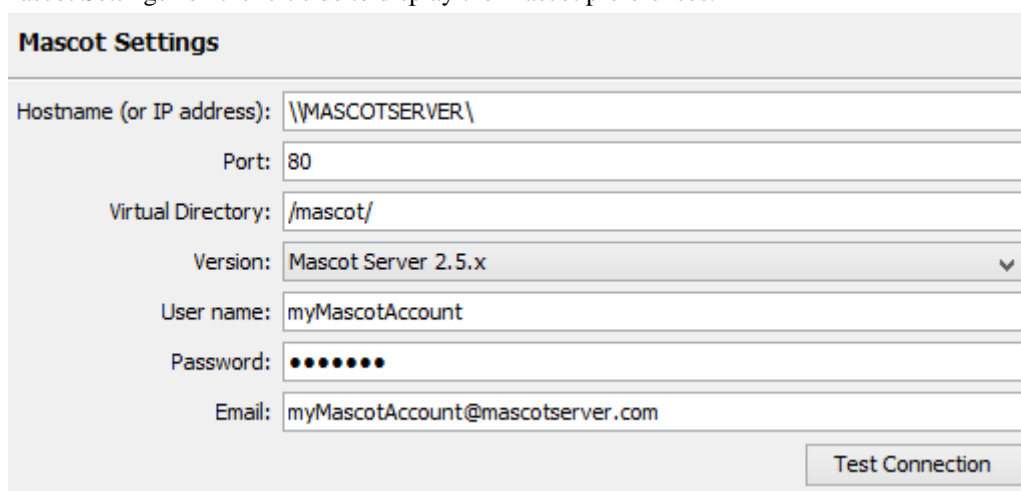
Enabling the MassLynx converter overwrites the default loading preference for the Water's raw data. Refer to Chapter 4, Section 8.2, “Waters/Micromass (MassLynx) Data” for details on Waters instrument preferences.

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.

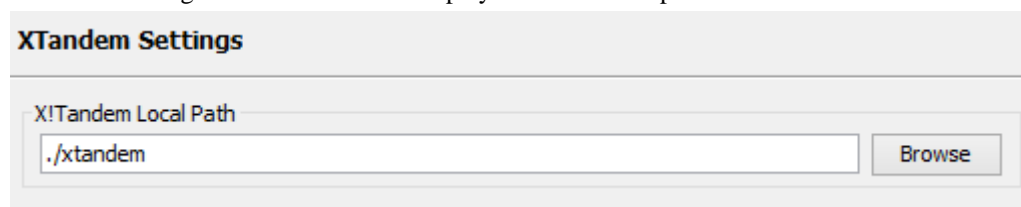


The screenshot shows the "Mascot Settings" dialog box. It contains several input fields: "Hostname (or IP address):" with the value "\\MASCOTSERVER\", "Port:" with the value "80", "Virtual Directory:" with the value "/mascot/", "Version:" with a dropdown menu showing "Mascot Server 2.5.x", "User name:" with the value "myMascotAccount", "Password:" with masked characters "••••••", and "Email:" with the value "myMascotAccount@mascotserver.com". A "Test Connection" button is located at the bottom right.

These parameters specify how PEAKS accesses the Mascot server (if applicable). Enter the hostname (or an IP address), port, virtual directory, and 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 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.



The screenshot shows the "X!Tandem Settings" dialog box. It contains a single input field labeled "X!Tandem Local Path" with the value "./xtandem". A "Browse" button is located to the right of the input field.

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:

	+H2O	-NH3	2+
a	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
b	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
c	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
x	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
y	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
z	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
z'	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
c-H	<input type="checkbox"/>		
immonium	<input type="checkbox"/>		
internal	<input type="checkbox"/>		
precursor	<input checked="" type="checkbox"/>		

The spectra in 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 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 displayed in a pop-up in the “Spectrum Annotation” view when this option is checked and the cursor is placed on a peak.

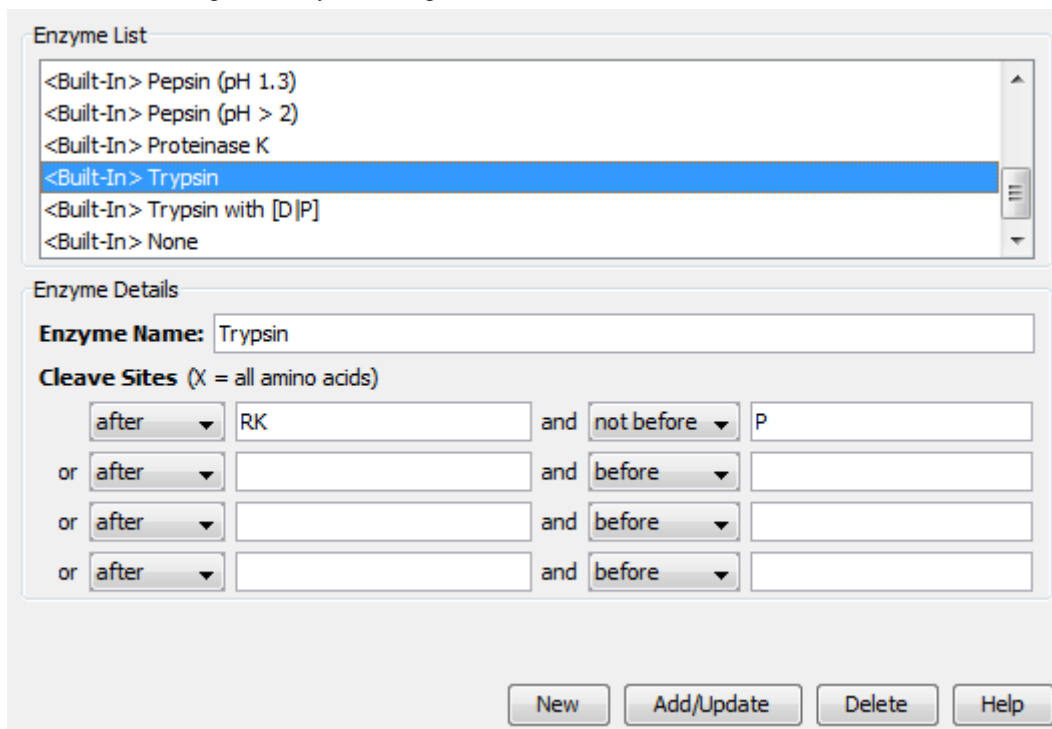
Intensity - Set the intensity threshold for spectrum annotation to low (2%), medium (5%), or high (10%). To apply this intensity threshold, select the “intensity threshold” checkbox in the “Spectrum Annotation” view (see Chapter 8, Section 3.2.2, “Spectrum Annotation”).

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 | Window menu.

2.1. Enzyme Configuration

PEAKS can use almost any 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.



Enzyme List

- <Built-In> Pepsin (pH 1.3)
- <Built-In> Pepsin (pH > 2)
- <Built-In> Proteinase K
- <Built-In> Trypsin
- <Built-In> Trypsin with [D]P
- <Built-In> None

Enzyme Details

Enzyme Name: Trypsin

Cleave Sites (X = all amino acids)

after	RK	and	not before	P
or	after		and	before
or	after		and	before
or	after		and	before

New Add/Update Delete Help

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 Chapter 8, 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 in three separate PTM lists under “Common”, “Uncommon”, and “Artificial” tabs. 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 be induced only artificially. The most recently used PTMs are listed in the “Recent” tab and the “Customized” tab lists all the user-defined PTMs. Double clicking on any of these PTMs will display the PTM details in the “PTM Info” popup dialog. The same information can be viewed by selecting a PTM from a list and clicking the View button. For details on PTMs, refer to Chapter 8, Section 2.3, “PTM”.

PTM name:	Carbamidomethylation		
PTM abbreviation:	Carbamidomethyl		
Mass (Monoisotopic):	57.021465		
Residues that can be modified:	C	Anywhere	
<input checked="" type="radio"/> peptide <input type="radio"/> protein		N-term	
		C-term	
Formula:	H(3) C(2) N(1) O(1)		
Rule:			
<div>Close</div> <div>Help</div>			

Create a new PTM. Click on the New button to display the “New PTM” dialog. Provide the information about the PTM. The required information about the PTM are listed below:

PTM name:			
PTM abbreviation:			
Mass (Monoisotopic):			
Residues that can be modified:		Anywhere	
<input checked="" type="radio"/> peptide <input type="radio"/> protein		N-term	
		C-term	
Formula:			
Rule:			
<div>OK</div> <div>Cancel</div> <div>Help</div>			

- **PTM Name:** This name will appear in the PTM list for future use after it is saved.
- **PTM abbreviation:** PTM expressed in 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 Chapter 8, 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 screenshot shows the 'Labeled Q Method Configuration' window. On the left, a list of methods is displayed with icons: R for ITRAQ-4plex, ITRAQ-8plex, and TMT-6plex; and P for ICPL, ICAT-C, and ICAT-D. The 'ITRAQ-4plex' method is selected. To the right of the list are buttons for 'New', 'Delete', 'Edit', and 'Duplicate'. The main panel, titled 'Quantification Method Detail', contains the following fields:

- Name:** ITRAQ-4plex
- Method Type:** Reporter Ion Quantification (dropdown)
- Modification Target:**
 - ☒ N-Terminal Modification
 - ☐ C-Terminal Modification
 - ☐ Side Chain Modification at []
 - Modification Mass:** 144.10207
- Label Options:**

Name	Reporter Ion Mass (Da)
114	114.1107
115	115.1077
116	116.1111
117	117.1144

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”.

Modification Target: The modification target and mass of the unfragmented modification for “Reporter Ion Quantification”. The modification targets can be N-terminal, C-terminal or Side Chain.

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, modification, and modification details.

Create a New Method. To create a new labeled Q method, click the New button to open the “New/Edit Quantification Method” dialog where the quantification method details can be specified.

Quantification Method Detail (Left):

Name: My RIQ Method Method Type: Reporter Ion Quantification

Modification Target:

☒ N-Terminal Modification ☐ C-Terminal Modification

☐ Side Chain Modification at: Modification Mass: 144.10207

Label Options:

Name	Reporter Ion Mass (Da)
114	114.1107

Buttons: Add Label, Delete Current Line, OK, Cancel

Quantification Method Detail (Right):

Name: My PIQ Method Method Type: Precursor Ion Quantification

Label Options:

Sample Name	Modification	Modification Detail
Heavy	SILAC K6	6.0201@K

Buttons: Add Row, Delete Row, OK, Cancel

Specify the modification target, modification mass, and label options for a “Reporter Ion Quantification” method. Use Add Label and Delete Current Line to add or remove a label. Each label is defined by the sample name and reporter ion mass.

Use Add row and Delete Row to add or remove a label for a “Precursor Ion Quantification” method. Each label is defined by the sample name, added mass, target residue, and 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 search through a database to identify proteins, PEAKS must have access to a protein or 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 download them from public servers. 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.

The screenshot shows the 'Database' tab of the 'Configuration and Preferences' dialog. On the left is a sidebar with buttons: 'Enzyme', 'PTM', 'Labeled Q Method', 'Database' (selected), and 'Instrument'. The main area is divided into sections:

- Database List:** A list box containing '[UniProt_SwissProt]', 'Viridiplante', '[IP1]', and 'ABRF_iPRG_2012_target'. '[UniProt_SwissProt]' is selected.
- Database Details:**
 - FASTA format database:** A dropdown menu showing 'UniProtKB/Swiss-Prot', with 'Validate Database' and 'Export Decoy DB' buttons to its right.
 - Basic Options:**
 - Database name:** A text field with '[UniProt_SwissProt]' and a 'Validated' checkbox.
 - Path:** A text field with 'D:\FastaDB\uniprot_sprot.fasta', 'Browse', and 'Download' buttons.
 - EST database:** An unchecked checkbox.
 - Advanced Options - Fasta Title Format:**
 - Rule to parse accession/id from FASTA title:** A text field with '>(sp|)|*(|S*|)'.
 - Rule to parse description from FASTA title:** A text field with '|s+|(|. *|)'.
 - Accession/id URL:** A text field with 'http://www.uniprot.org/uniprot/' and a '<Accession/ID>' placeholder.
 - Delimiter:** An empty text field.
 - Taxonomy Options:**
 - taxonid:** A text field with 'D:\FastaDB\specdist.txt', 'Browse', and 'Download' buttons.
 - taxdmp:** A text field with 'D:\FastaDB\taxdmp.zip', 'Browse', and 'Download' buttons.

At the bottom right are four buttons: 'New', 'Add/Update', 'Delete', and 'Help'.

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. It is therefore 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 sub-taxa 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 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 been 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 list 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 list 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 from the examination of the survey scan.

4. Use the “MS/MS Product Scan” drop-down list 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 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.