

# **PEAKS X User Manual**

**PEAKS Team**

---

# **PEAKS X User Manual**

PEAKS Team

Publication date 2018

---

---

# Table of Contents

1. Overview .....	1
1. How to Use This Manual .....	1
2. What Is PEAKS? .....	1
3. What Is New in PEAKS X? .....	3
4. Quick Walkthrough .....	4
4.1. Opening an Existing Project .....	4
4.2. PEAKS Main GUI .....	5
4.3. Result Summary and Filtering .....	5
4.4. Result Visualization .....	7
4.5. Creating a PEAKS Project .....	8
4.6. Conducting an Analysis .....	9
2. Installation and Activation .....	11
1. Package Contents .....	11
2. System Requirements .....	11
3. Installation on a Windows Computer .....	11
4. Activation .....	12
4.1. Activate PEAKS with a Trial or Purchased License Key .....	13
4.2. Register to Get a Free Trial License Key .....	13
4.3. Use PEAKS as a Viewer .....	14
4.4. Activate PEAKS Manually .....	14
4.5. Re-Registering PEAKS .....	16
4.6. Common Errors During Registration .....	17
5. PEAKS Performance Configuration .....	17
5.1. PEAKS 64-bit Version .....	17
5.2. PEAKS 32-bit Version .....	18
6. What's Next .....	19
3. Configuration Wizard - Configure Instruments and Public Databases .....	20
1. Overview .....	20
2. Instrument Selection .....	20
3. Database Selection .....	21
4. Download Information .....	21
5. Vendor Instrument Software Options .....	22
6. Apply Changes .....	24
4. Project Wizard - Create a PEAKS Project, Load Data and Set up a Workflow .....	26
1. Overview .....	26
2. Create New Project .....	27
3. Data Refinement - Setting Parameters .....	28
4. DB Search - Setting Parameters .....	28
5. Quantification - Setting Parameters .....	29
6. Adding Data to an Existing Project .....	30
7. Supported Data Formats .....	30
8. Vendor-Specific Requirements .....	31
8.1. Agilent Data .....	31
8.2. Thermo Data .....	32
8.3. Sciex Data .....	33
8.3.1. TripleTOF, QSTAR, or QTRAP .....	33
8.3.2. TOF-TOF 4700/4800 .....	34
8.4. Bruker Data .....	35
8.5. Shimadzu Data .....	35
8.6. Varian .....	36
8.7. Waters/Micromass (MassLynx) Data .....	37
8.8. Older PEAKS Projects .....	38
9. Changing the Default Project Location .....	39
5. Data Visualization .....	40
1. Overview .....	40

---

2. MS View .....	40
3. MS/MS View .....	41
4. LC/MS View .....	42
4.1. Noise Level .....	43
4.2. Show/Hide Features .....	43
4.3. Locate M/Z and Retention Time .....	44
4.4. Synchronize LC/MS View Between LC Runs .....	44
4.5. Show/Hide MS/MS Spectra .....	44
4.6. Intensity View .....	45
4.7. Export LC/MS .....	46
6. Adding a Sequence Database .....	47
1. Configuring Sequence Databases .....	47
2. Databases to be Used in PEAKS' inChorus Function .....	49
7. Data Refinement .....	50
1. Overview .....	50
2. Data Refinement Parameters .....	51
2.1. Saving the Parameters for Future Use .....	52
8. Peptide <i>De Novo</i> Sequencing .....	53
1. Overview .....	53
2. <i>De Novo</i> Sequencing Parameters .....	54
2.1. Error Tolerance .....	55
2.2. Enzyme .....	55
2.3. PTM .....	55
2.4. Other Parameters .....	57
2.5. Saving the Parameters for Future Use .....	57
3. Understanding PEAKS <i>De Novo</i> Sequencing Results .....	58
3.1. Summary View .....	58
3.2. <i>De Novo</i> Peptide View .....	59
3.2.1. Peptide Table .....	60
3.2.2. Spectrum Annotation .....	62
3.2.3. Ion Table .....	63
3.2.4. Parent Scan .....	64
3.2.5. LC-MS Snapshot .....	64
3.2.6. Precursor Profile .....	65
3.3. Feature View .....	65
3.3.1. Feature Table .....	66
3.3.2. Spectrum Annotation .....	67
3.3.3. Ion Table .....	68
3.3.4. Parent Scan .....	68
3.3.5. LC-MS Snapshot .....	68
3.3.6. Precursor Profile .....	68
3.4. LC-MS View .....	68
4. Filtering <i>De Novo</i> Sequencing Results .....	69
5. Exporting <i>De Novo</i> Results .....	71
6. <i>De Novo</i> Sequencing using Spectral Pairs or Triplets .....	71
7. <i>De Novo</i> Sequencing on Single Spectrum .....	72
8. Manual <i>De Novo</i> Sequencing .....	73
8.1. Manual <i>De Novo</i> Graphical User Interface .....	73
8.2. Manual <i>De Novo</i> Operations .....	74
9. Peptide, PTM, and Mutation Identification (PEAKS DB, PEAKS PTM, and SPIDER) .....	80
1. Overview .....	80
2. Setting PEAKS Search Parameters .....	81
2.1. Saving the Parameters for Future Use .....	84
3. Understanding PEAKS Search Results .....	84
3.1. Peptide and Protein Scores .....	85
3.2. Summary View .....	85
3.2.1. False Discovery Rate (FDR) Curve .....	85
3.2.2. PSM Score Distribution .....	86

---



---

3.2.3. Distribution of Peptide Feature Detection .....	87
3.2.4. De Novo Result Validation .....	88
3.2.5. Statistics of Data and Results .....	89
3.2.6. Experiment Control .....	91
3.2.7. Other Information .....	91
3.3. Feature View .....	92
3.3.1. <b>Feature View Filters:</b> .....	93
3.3.2. Feature Table .....	94
3.3.3. Peptide Precursor Spectrum Matches .....	95
3.3.4. Ion Table .....	96
3.3.5. Parent Scan .....	96
3.3.6. LC-MS Snapshot .....	96
3.3.7. Precursor Profile .....	97
3.4. Protein View .....	97
3.4.1. Protein View Filters .....	98
3.4.2. Protein Table .....	99
3.4.3. Global Protein Comparative View .....	100
3.4.4. Protein Coverage .....	101
3.4.5. Peptide Mapping .....	104
3.4.6. Supporting Peptides .....	108
3.4.7. De novo Tags .....	108
3.5. Peptide View .....	109
3.5.1. Peptide View Search and Filter Options .....	109
3.5.2. Peptide Table .....	110
3.5.3. Peptide Global Comparative View .....	111
3.5.4. Peptide-PrecursorSpectrum Matches .....	112
3.6. De Novo Only View .....	112
3.7. LC-MS View .....	113
4. Filtering PEAKS Search Results .....	113
5. Exporting PEAKS Search Results .....	115
6. Running PEAKS PTM and SPIDER Separately .....	115
6.1. Running PEAKS PTM on a PEAKS DB Result .....	115
6.2. Running SPIDER on a PEAKS DB or PEAKS PTM Result .....	116
7. Comparison of PEAKS Results .....	117
7.1. Comparison Result .....	117
7.2. Peptide Comparison .....	117
7.3. Protein Comparison .....	118
7.4. Statistical Charts .....	118
7.5. Exporting Comparison Result .....	119
10. Multi-Round Search .....	120
1. Multi-Round Search Overview .....	120
11. PTM Profiling .....	121
1. Overview .....	121
2. PTM Profiling for DB/PTM/SPIDER Results .....	121
2.1. Input Selection .....	122
2.2. PTM Profile Table .....	122
2.3. PTM Profile Chart .....	123
2.4. Spectrum Annotation View .....	124
2.5. Result Export .....	125
3. PTM Profiling for Label Free Quantification and Reporter Ion Quantification Results .....	126
3.1. Input Selection .....	126
3.2. PTM Profile Table .....	126
3.3. PTM Profile Heat Map .....	126
3.4. Detail Pane .....	127
3.5. Result Export .....	127
4. PTM Profiling for Precursor Ion Quantification Results .....	127
4.1. PTM Profile Table .....	128
4.2. PTM Profile HeatMap .....	128

---

---

4.3. Detail Pane .....	128
4.4. Result Export .....	128
12. Combining Multiple Database Search Engines with PEAKS inChorus .....	130
1. PEAKS inChorus Overview .....	130
2. Understanding PEAKS inChorus Results .....	133
3. Filtering PEAKS inChorus Results .....	135
4. Exporting inChorus Results .....	135
13. PEAKS Q - MS Level .....	136
1. Overview .....	136
2. Setting Parameters .....	136
3. Understanding the Results .....	139
3.1. Summary View .....	139
3.1.1. Result Filtration .....	139
3.1.2. Summary Report .....	144
3.1.3. Result Exporting .....	147
3.2. Protein View .....	148
3.2.1. Protein Table .....	148
3.2.2. Protein Table View Filter .....	149
3.2.3. Volcano Plot .....	149
3.2.4. Protein Coverage .....	149
3.2.5. PTM Profile .....	149
3.2.6. Supporting Peptides .....	149
3.3. Peptide View .....	150
3.3.1. Peptide Table .....	150
3.3.2. Peptide View Filters .....	151
3.3.3. All Vectors Table .....	151
3.3.4. Peptide Details .....	151
3.4. LC/MS View .....	152
14. PEAKS Q - MS/MS Level .....	153
1. Overview .....	153
2. Setting Parameters .....	153
3. Understanding the Results .....	155
3.1. Summary View .....	155
3.1.1. Result Filtration .....	155
3.1.2. Summary Report .....	160
3.1.3. Result Exporting .....	164
3.2. Protein View .....	164
3.2.1. Protein Table .....	165
3.2.2. Protein Volcano Plot .....	166
3.2.3. Coverage Tab .....	166
3.2.4. Peptides Tab .....	166
3.3. Peptide View .....	166
3.3.1. Ion Match Table .....	167
3.3.2. Survey Scan .....	167
3.3.3. Reporter Ion .....	167
3.4. LC/MS View .....	168
15. PEAKS Q - Label Free .....	169
1. Overview .....	169
2. Setting Parameters .....	169
3. Understanding the LFQ Results .....	171
3.1. Summary View .....	171
3.1.1. Result Filtration .....	172
3.1.2. Summary Report .....	176
3.1.3. Result Exporting .....	180
3.2. Protein View .....	180
3.2.1. Protein Table .....	181
3.2.2. Protein Volcano Plot .....	181
3.2.3. Protein Coverage .....	181

---

---

3.2.4. Support Peptides .....	181
3.3. Peptide View .....	183
3.3.1. Peptide Table .....	183
3.3.2. All Vectors Table .....	183
3.3.3. Feature Details .....	184
3.3.4. Sample Features .....	184
3.3.5. RT Alignment .....	185
3.4. Features View .....	185
3.5. LC/MS View .....	185
16. Data Independent Acquisition (DIA) Data Analysis .....	186
1. Overview .....	186
2. Data Visualization .....	187
2.1. MS .....	187
2.2. MS/MS .....	188
2.3. LC-MS .....	189
3. Understanding PEAKS Results .....	190
3.1. Data Refine .....	190
3.2. <i>De Novo</i> .....	190
3.2.1. Summary View .....	191
3.2.2. <i>De Novo</i> Peptide View .....	191
3.2.3. Feature View .....	192
3.2.4. LC-MS .....	192
3.3. PEAKS DB .....	193
3.3.1. Summary View .....	193
3.3.2. Feature View .....	193
3.3.3. Protein View .....	193
3.3.4. Peptides View .....	193
3.3.5. <i>De Novo</i> Only View .....	195
3.3.6. LC-MS View .....	195
4. Label Free Quantification (LFQ) .....	195
5. Precursor Ion Quantification (e.g. SILAC and iCAT) .....	195
6. Reporter Ion Quantification (e.g. iTRAQ and TMT) .....	195
17. Ion Mobility Data Analysis .....	197
1. Overview .....	197
2. Data Visualization .....	198
2.1. IM-MS .....	198
2.2. IM-MS/MS .....	198
2.3. LC-IM/MS .....	199
3. Understanding PEAKS Results .....	200
3.1. Data Refine .....	200
3.2. <i>De Novo</i> .....	201
3.2.1. Summary View .....	201
3.2.2. <i>De Novo</i> Peptide View .....	201
3.2.3. LC-IM/MS .....	202
3.3. PEAKS DB .....	202
3.3.1. Summary View .....	202
3.3.2. Protein View .....	202
3.3.3. Peptides View .....	202
3.3.4. <i>De Novo</i> Only View .....	202
3.3.5. LC-IM/MS View .....	202
4. Label Free Quantification (LFQ) .....	203
4.1. Setting Parameters .....	203
4.2. Result Filtration .....	203
4.3. Peptide Features .....	203
4.3.1. Feature Details .....	204
4.3.2. Sample Features .....	204
5. Precursor Ion Quantification (eg. SILAC) .....	205
6. Reporter Ion Quantification (e.g. iTRAQ and TMT) .....	205

---

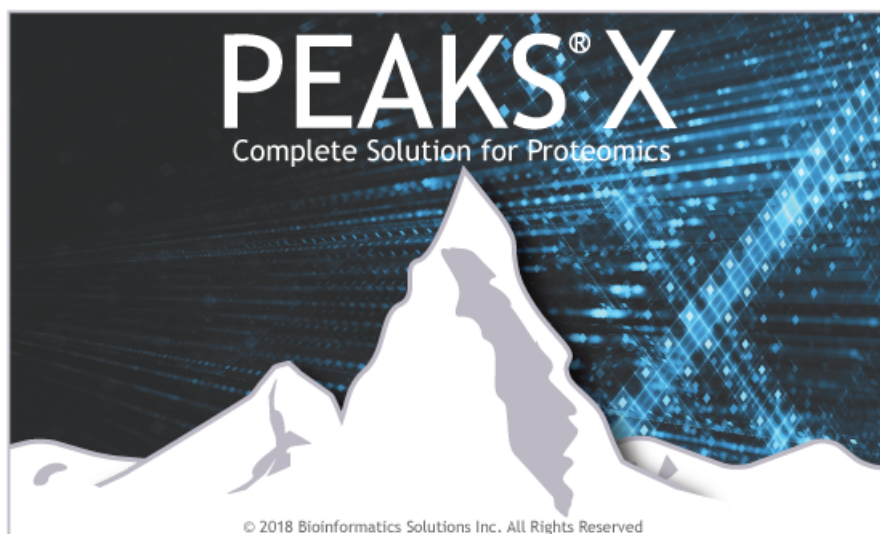
---

6.1. Result Display .....	205
6.1.1. Peptide View .....	205
18. Workflow .....	207
1. Identification Workflow .....	207
2. Quantification Workflow .....	208
3. inChorus Workflow .....	209
19. Exporting Data/Reports and Printing .....	210
1. Exporting Data .....	210
2. Exporting Result From Project View .....	210
3. Exporting <i>De Novo</i> Results .....	211
3.1. Exporting Summary, Peptides, and Features .....	211
3.2. Exporting Images .....	212
4. Exporting Database Search Results .....	213
4.1. Exporting Summary, Proteins, and Peptides .....	213
4.2. Exporting Images .....	215
4.2.1. Exporting Protein Coverage .....	215
4.2.2. Exporting LC-MS, IM-MS, and LC-IM-MS .....	215
5. Exporting Quantification Results .....	215
5.1. Exporting Precursor Ion Labelling Quantification Results .....	215
5.2. Exporting Reporter Ion Labelling Quantification Results .....	215
5.3. Exporting Label-Free Quantification Results .....	217
6. Export inChorus Result .....	218
20. Configuration and Preferences .....	219
1. PEAKS Environment Preferences .....	219
1.1. General Preferences .....	219
1.1.1. Display Options .....	220
1.1.2. RMI Connections .....	220
1.1.3. Derby Database .....	221
1.2. Raw File Converter Preferences .....	222
1.2.1. Thermo (.raw) .....	222
1.2.2. SCIEX (.wiff) .....	222
1.2.3. Bruker (.yep/baf/tdf, fid) .....	223
1.2.4. Shimadzu AXIMA (.run) .....	223
1.2.5. Varian (.xms) .....	223
1.2.6. Waters (.raw) .....	223
1.3. Search Engine Preferences .....	223
1.3.1. Mascot Settings .....	223
1.3.2. X!Tandem Settings .....	224
1.3.3. OMSSA Settings .....	225
1.4. Spectrum Annotation Preferences .....	226
2. PEAKS Configuration .....	227
2.1. Enzyme Configuration .....	228
2.2. PTM Configuration .....	229
2.3. Labeled Q Method Configuration .....	230
2.4. Database Configuration .....	232
2.5. Instrument Configuration .....	234

---

# Chapter 1. Overview

Welcome to PEAKS.



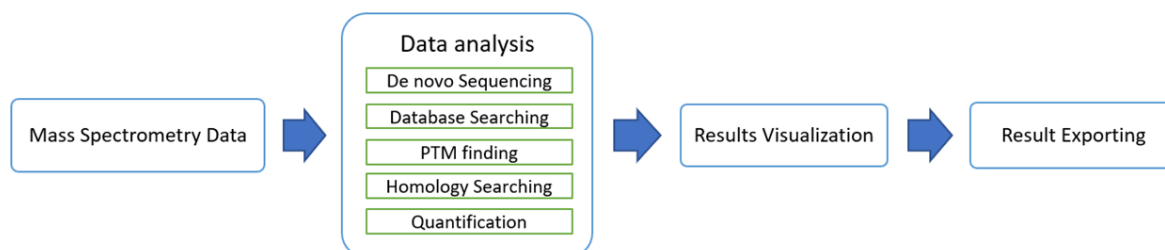
## 1. How to Use This Manual

This chapter provides an overview of the distinctive features in PEAKS and describes a typical data analysis work flow within. 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 purposes only and does not need to be read from cover to cover. The electronic (and most up-to-date) version of this manual can be found at <http://www.bioinfor.com/user-manual/>.

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

## 2. What Is PEAKS?

PEAKS is a complete software package for both data dependent (DDA) and data independent (DIA) acquisition shotgun proteomics data analysis. It is a vendor neutral software capable of reading in raw mass spectrometry data and public data formats. *De novo* sequencing, PEAKS DB (database searching)-based protein identification, PEAKS PTM (post translational modification) analysis, and SPIDER homology search are then used to identify the most likely peptide sequence that match the given spectra. Relative quantification by label free, isobaric labeling (e.g. TMT or iTRAQ), or metabolic labeling (e.g. SILAC) can also be performed. Intuitive result visualization tools are provided at every stage of analysis and analysis results can be exported.



The following is a brief overview of the distinctive features included in the PEAKS software:

- **Complete Analysis**

PEAKS is a complete software package for both data dependent (DDA) and data independent (DIA) acquisition shotgun proteomics data analysis. It is a vendor neutral software capable of reading in raw mass spectrometry data and public data formats. *De novo* sequencing, PEAKS DB (database searching)-based protein identification, PEAKS PTM (post translational modification) analysis, and SPIDER homology search are then used to identify the most likely peptide sequence that match the given spectra. Relative quantification by label-free quantification (LFQ), isobaric labeling (e.g. TMT or iTRAQ), or metabolic labeling (e.g. SILAC) can also be performed. Intuitive result visualization tools are provided at every stage of analysis and analysis results can be exported.

- **Built-in Result Validation**

Statistical tools are provided at every stage of analysis. *De novo* peptide sequencing results are confirmed by local confidence scores at the amino acid level. The "decoy-fusion" method is used to validate the peptide identifications automatically and a false discovery rate (FDR) is estimated and provided. Quantification results can also be analyzed by significant tests such as ANOVA or PEAKS Q.

- **Result Visualization**

LC-MS/MS heat maps allow for full visualization of peptide features, MS/MS spectra acquisition, and identification position relative to mass over charge ( $m/z$ ), retention time (RT), ion mobility ( $1/ko$ ), and signal intensity. Peptides are mapped to the identified proteins using an interactive protein coverage view. Clicking on peptides will bring up the peptide spectrum match that best explains the identification. Modification and mutations can be visualized and filtered based on certain thresholds. PTM profiling is included to quantify differences in modified and unmodified peptide abundance at different modification sites. Quantification results can be visualized using heat maps, correlation profiles, and extracted ion chromatograms (XICs).

- **Accuracy and Sensitivity**

The superiority of the PEAKS algorithms has been demonstrated by publications and third-party evaluations. By combining four complementary algorithms, including *de novo* sequencing, PEAKS DB, PEAKS PTM and SPIDER, the sensitivity is further improved.

- **Multiple-Enzyme Project**

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

- **Label Free Quantification**

Proteins are quantified based on peptide features detected from LC-MS/MS data by integrating the area under the curve (AUC) or maximum feature intensity. 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. This approach allows PEAKS to quantify peptide features independent of MS/MS identification.

- **Label Based Quantification**

PEAKS can analyze data generated from label-based methods. For isobaric labelling data such as iTRAQ and TMT, the intensity of peptides from individual samples are determined from reporter ion signals in MS2 or MS3 scans. For MS1-based labeling data such as SILAC, labeled and unlabeled features are paired and their abundance ratios are calculated.

- **Multiple Instrument Vendor Support**

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

- **Result Exporting**

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

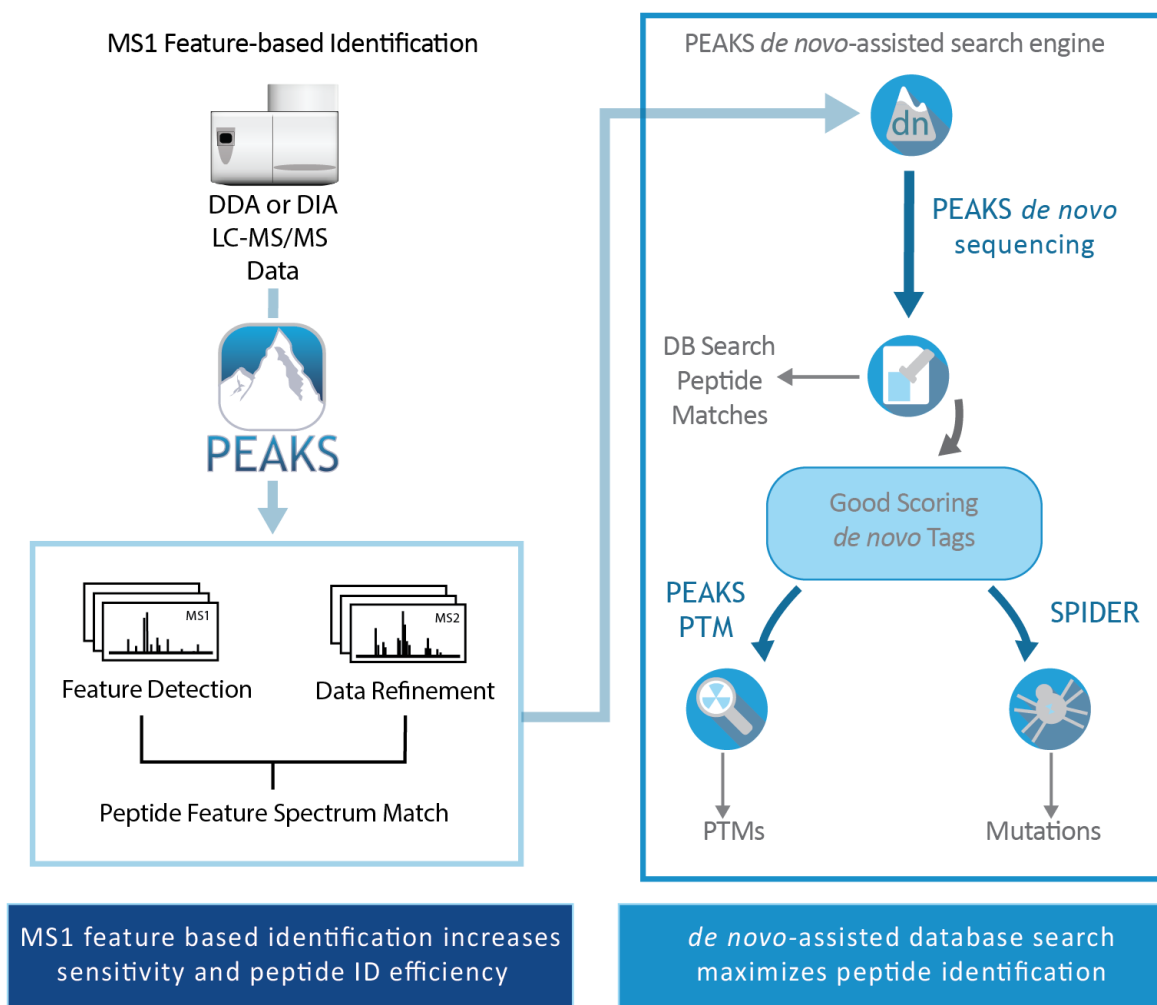
- **inChorus**

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

### 3. What Is New in PEAKS X?

PEAKS X introduces a feature-based identification method to its unique *de novo*-assisted identification work flow to increase sensitivity and maximize peptide identification efficiency for in-depth shotgun analysis of complex proteomes and peptidomes. This approach provides three main benefits:

1. Support both DDA and DIA technology to improve reproducibility;
2. Integrate database search and *de novo* sequencing to extend in-depth analysis; and
3. Resolve chimeric spectra to increase efficiency of peptide ID.



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

### Highlights.

- **Resolve Chimeric Spectra**

Using the new MS1 feature-based identification approach, one MS2 scan can be associated with multiple peptide features within its acquisition window, enabling multiple peptide identifications per chimeric MS2 spectrum.

- **Spectral-Library-Free DIA Support**

PEAKS X's new feature-based identification approach enables full DIA analysis support without the need of a spectral library. DIA spectra are analyzed directly to determine the peptide sequences of peptide features in their acquisition window, allowing peptide identifications beyond the information from a specific spectral library. With the combined *de novo* and feature-based technique, PEAKS removes biases found in commonly utilized approaches.

- **De Novo Sequencing for DIA Data**

PEAKS X introduces the first *de novo* sequencing solution for DIA data to provide non-biased results.

- **Improved Identification of Endogenous and Low Abundant Peptides**

PEAKS X improves analysis for endogenous peptide identifications by allowing variability at the C-terminus of peptides in no enzyme searching. This combined with feature-based peptide identification increases sensitivity when analyzing endogenous peptide datasets.

- **Ion-Mobility Spectrometry (IMS-MS) support**

IMS-MS data can be analyzed with PEAKS *de novo*, identification and quantification work flows and interactive data visualization tools allow data view projected on m/z-rt or m/z-1/k0 dimensions.

### Additional Improvements.

- Feature-based analysis for timsTOF data.
- Support for Thermo Fisher Scientific's Advanced Peak Determination (APD).
- Purity correction for TMT/iTRAQ.
- Support for TMT11-plex as a built-in method.
- Support for "No Digestion" samples using "Digest Mode" option in DB search.
- Protein group FDR in identification result.

## 4. Quick Walkthrough

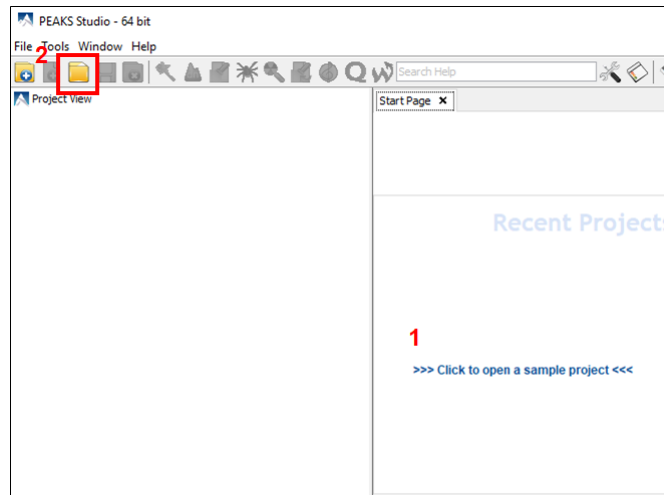
This section presents a quick walkthrough of a typical data analysis work flow. The PEAKS user interface is introduced by using the sample project included in PEAKS installation, filtration and visualization of the analysis result are showcased, which will help users to understand what can be accomplished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

### 4.1. Opening an Existing Project

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

1. For fresh installations, click the "Sample Project" in the "Recent Projects" list on the Start Page.
2. For all installations, click the Open Project icon, browse to the directory where PEAKS X was installed, select "SampleProject", and click the Open button in the file browser.

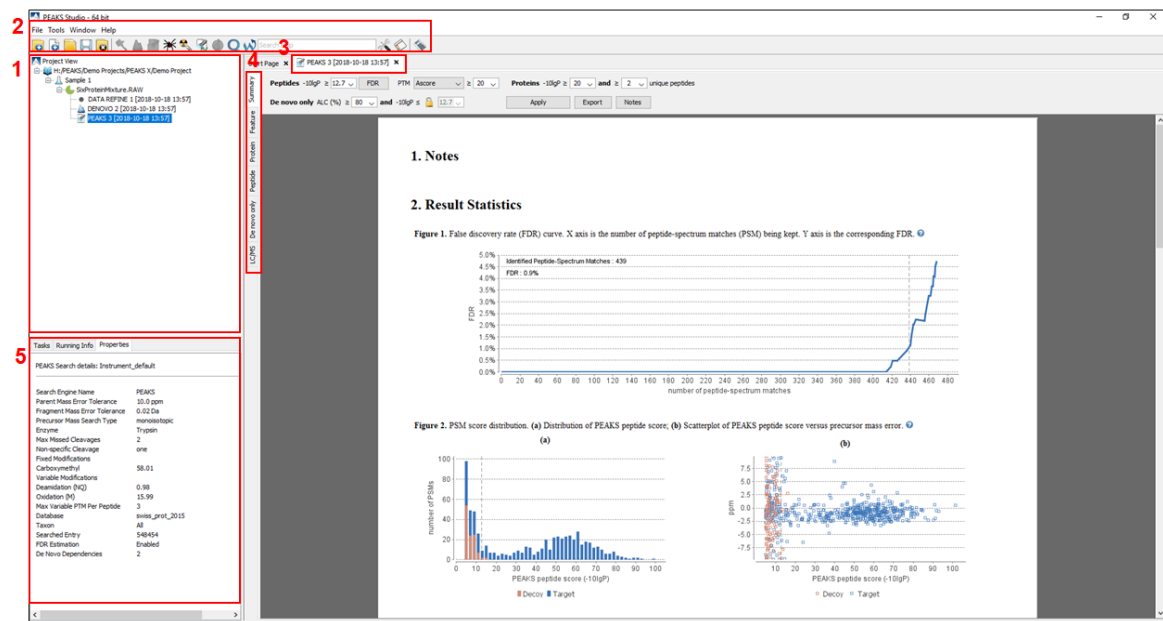




## 4.2. PEAKS Main GUI

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

1. The Project View pane shows all of the opened projects. Each project may include multiple samples and each sample may include multiple fractions (LC-MS runs). The analysis results are displayed as result nodes under the project.
2. The menu and tool bar contain all of the analysis tool icons. Selecting a node (project, sample, fraction, or result) in an opened project will highlight the analysis tool icons that are available to the selected node.
3. A result node in a project can be opened by double-clicking the node. Opened result nodes are shown in tabs.
4. Each opened result node provides several different "views" as different tabs. In particular, the "Summary View" shows the result statistics and is also the central tab in which to filter and export PEAKS results.
5. The information pane shows useful information such as the node properties and the progress of running tasks.



## 4.3. Result Summary and Filtering

After double-clicking to open a result node (eg. the PEAKS DB node in the sample project) the "Summary View" is shown by default and provides three main functions:

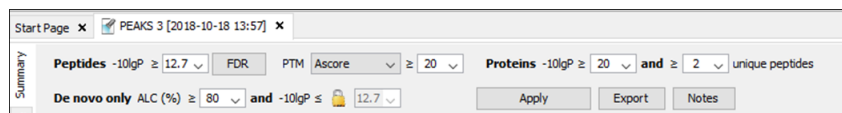
1. Specify score thresholds to filter the results.
2. Examine the result statistics.
3. Export the results.

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

- Low scoring peptide identifications are filtered out by setting a desired peptide-spectrum match  $-10\lg P$  threshold or a desired FDR (false-discovery rate), which can be specified by clicking the FDR button.
- Low scoring protein identifications are filtered by setting a desired protein  $-10\lg P$  score and a desired unique peptide count.
- The *de novo* only peptides are the confident *de novo* sequence tags that remain unidentified by the database search algorithms. To report a *de novo* only peptide, the ALC (average local confidence) scores must be equal to or better than the specified threshold. Meanwhile, the score of the spectrum's best database search result should be no greater than the specified  $-10\lg P$  threshold.

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

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



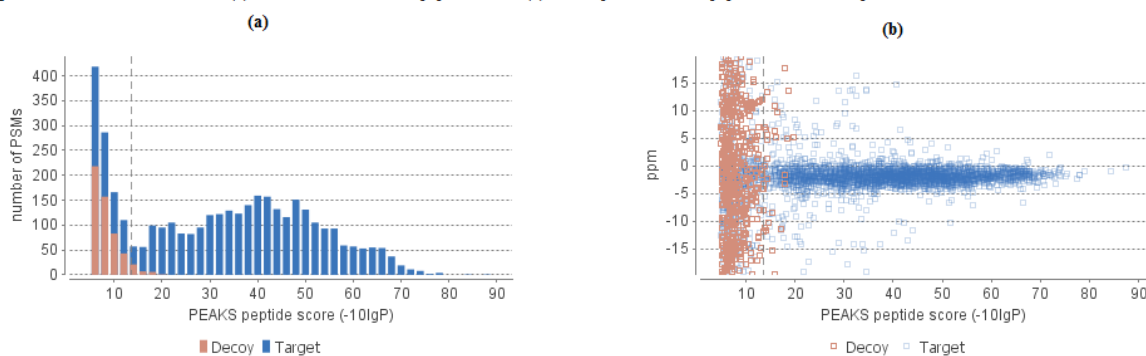
The top control pane has two additional buttons: Export and Notes. Clicking Export displays several options and file formats in which a user can export analysis results. Clicking Notes allows the user to add a text note about the project which will be displayed at the top of the "Summary" view page and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

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

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

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

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

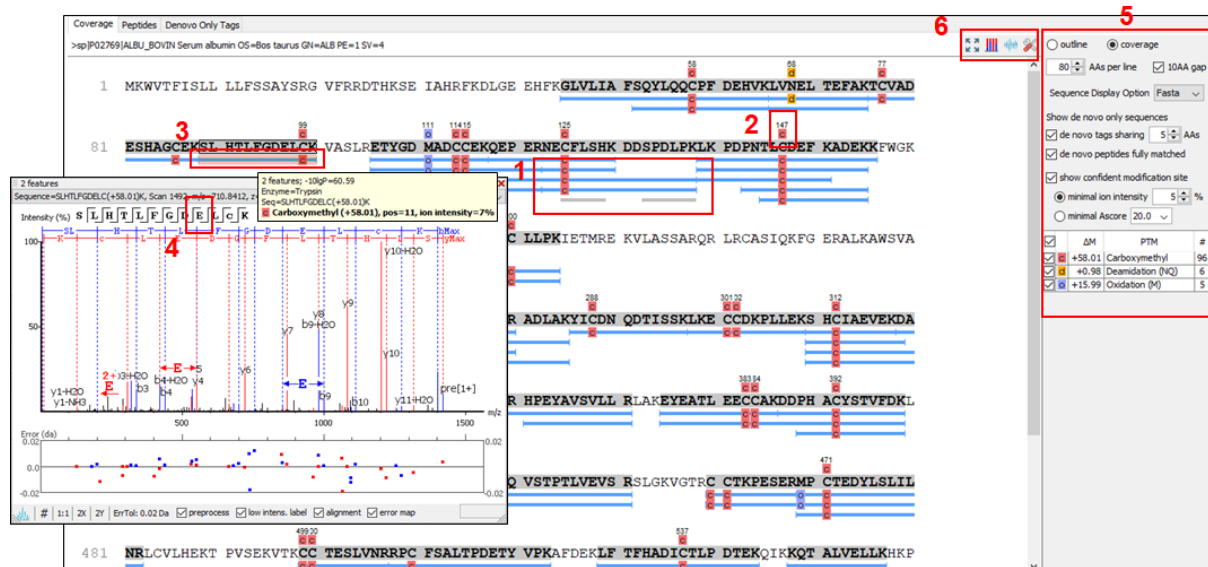


## 4.4. Result Visualization

In addition to the "Summary View", PEAKS organizes its results into four other sections: "Feature", "Protein", "Peptide", "De novo only", and "LC/MS."

- The "Feature View" contains a list of all detectable features that passed the filters. Each feature will list all the MS/MS scans associated with it.
- The "Protein View" contains a list of proteins that passed the filters. The proteins identified with the same set (or a subset) of peptides are grouped together.
- The "Peptide View" shows all the peptide identifications passing the filters. The multiple spectra that identified the same peptide sequence are grouped together.
- The "De Novo Only View" shows all of the peptides identified exclusively by *de novo* sequencing.
- The "LC/MS" view displays the LC-MS data as a heat map with highlighted MS/MS scans and detected features.

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



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

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

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

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







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

- The "coverage/outline" choice enables the display or the removal of the peptide bars.
  - The "sequence display option" highlights the coverage map based on either a fasta display or an enzyme display from one of the enzyme options selected during the project setup.
  - The "de novo tags sharing" specifies the minimum number of consecutive amino acid matches between a *de novo* only sequence and the protein before it can be displayed as a gray bar.
  - The "de novo peptides fully matched" check box allows a *de novo* peptide to be displayed if the sequence, regardless of its length, is fully matched to a sequence in the protein.
  - The "minimum ion intensity" specifies the minimum fragment ion relative intensity in one of the MS/MS spectra before a PTM location is regarded as confident and displayed on top of the protein sequence.
  - The check boxes in the PTM list allow users to view peptides with specific PTMs in the results. Click the colored boxes to change a color. Double-click a PTM name to see the PTM detail.
6. The full screen button, the PTM profile button, the peptide mapping button, and the tool box button.

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

## 4.5. Creating a PEAKS Project

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

1. Select New Project ... from the file menu or click the New Project icon  on the tool bar. The "Project Wizard" will appear.
2. Use the Add Data button to add the desired files to be loaded, and then click Open. All of the selected data files will be listed on the left side.
3. Place the selected data from the list into samples: use  to place all files in a new sample; use  to put them in an existing sample; or put them in individual samples for each file using .
4. Click the  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, "Enzyme" name, and Data "Acquisition".

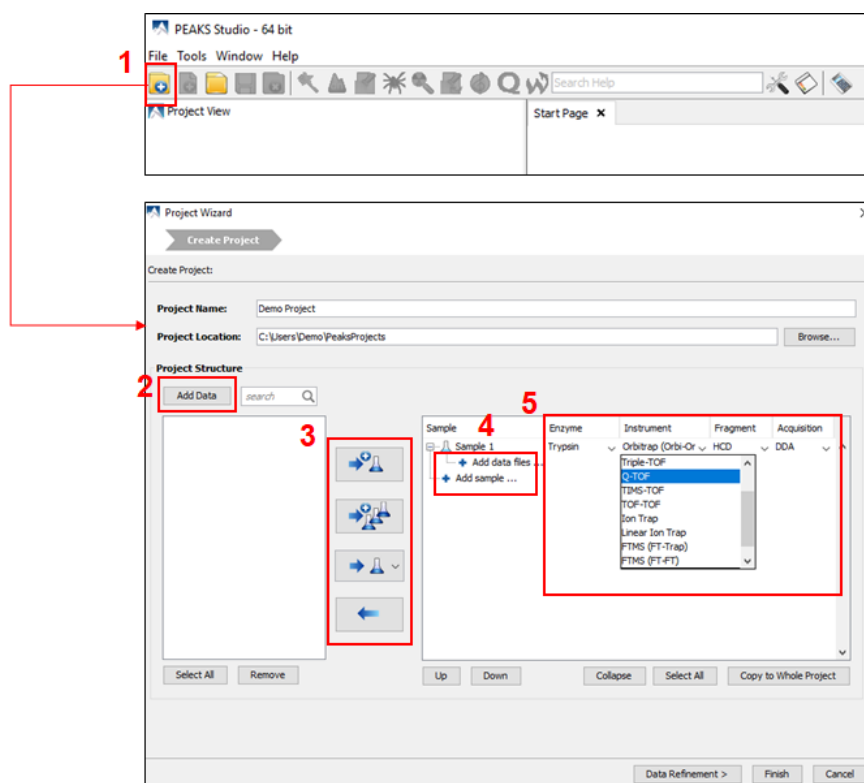
### Note

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

### Note

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

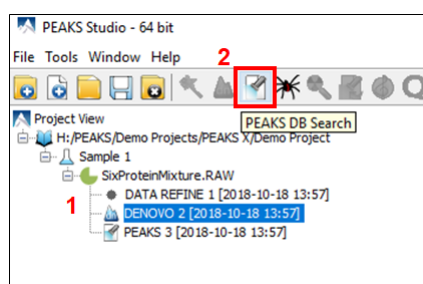
6. Click the Finish button to create the project.



By default, the peptide feature area is used for peptide and protein abundance calculation. Users can also select to use feature intensity as a calculation method. Set this parameter before creating a project. More details can be found in Chapter 4, *Project Wizard - Create a PEAKS Project, Load Data and Set up a Workflow* Section 2, “Create New Project”.

## 4.6. Conducting an Analysis

To conduct an analysis: 1) Select a project, a sample, or a result node in the Project View; 2) Click the desired analysis tool button. In the following example, the completed *de novo* result is selected from the Project View pane, and a PEAKS DB search will be employed using those results. The PEAKS Search is a database DB workflow will perform a database identification search.



A window will appear to allow users to specify the desired search parameters for the analysis. Most search options for PEAKS DB are standard and straightforward. More details are provided in the following:

1. If the proteolytic enzyme was specified for each sample at the project creation step, users can choose to use the enzyme already 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. Users can also select a contaminant database to be included in the database search. (optional)

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/fdr-tutorial/>) for more details.

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

By default, PEAKS PTM will perform an additional PTM search which considers all 313 naturally occurring modifications from Unimod. Users can instead specify a custom list of PTMs they wish to be used for the additional PEAKS PTM search by clicking the "Advanced Setting" button.

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

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

- 1 Enzyme:** A red box highlights the 'Enzyme' section, which includes a dropdown menu set to 'Specified by each sample' and a 'New' button.
- 2 PTM:** A red box highlights the 'PTM' section, which includes a list of modifications (Carboxymethyl, Deamidation (NQ), Oxidation (M)) with checkboxes, and buttons for 'Set PTM', 'Remove', and 'Switch type'.
- 3 Database:** A red box highlights the 'Database' section, which includes options for 'Select database' (UniProtKB/Swiss-Prot\_20180710), 'Paste sequence', 'Taxa' (all species), and 'Contaminant database' (cRAP 2017/7/13).
- 4 De Novo Tag Options:** A red box highlights the 'De Novo Tag Options' section, which includes a dropdown menu for 'Available de novo tags' set to 'de novo with current parameter'.
- 5 General Options:** A red box highlights the 'General Options' section, which includes a checked checkbox for 'Estimate FDR with decoy-fusion'.
- 6 Advanced Settings:** A red box highlights the 'Advanced Settings' section, which includes checkboxes for 'Find unspecified PTMs with PEAKS PTM' and 'Find more mutations with SPIDER', along with an 'Advanced Settings' button.

---

# 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:

- Quick Start Manual
- PEAKS Software USB Stick
- PEAKS Pen
- PEAKS Sticky Notes Pad
- License Certificate
- Quick reference sheet for mass spectrometry
- Printed Matter

## 2. System Requirements

PEAKS X is available in both 64-bit and 32-bit versions. By default, BSI distributes the 64-bit version for installation. For users with a 32-bit operating system, the 32-bit version can be downloaded only upon request. PEAKS X supports Windows operating systems. Although users are recommended to use Windows 10, previous versions of Windows are supported as well (Windows XP, Windows 7, and Windows 8). PEAKS program files will use ~700MB disk space, but it will require some extra disk space to store temporary files and intermediate results. The amount of disk space required for this purpose depends on the size of the user's datasets.

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

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

- **Recommended:** Quad-core processors and 16 GB of RAM.
- **Ideal:** Intel Core i7/i9 or Xeon processors with total 16 threads or more, and 32 GB of RAM or more.

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

## 3. Installation on a Windows Computer

### Important

Please uninstall any older version of PEAKS currently installed on the system before proceeding to prevent any licensing issues. Older versions of PEAKS can be used as a Viewer instead. Please contact <support@bioinfor.com> for additional assistance.

### Important

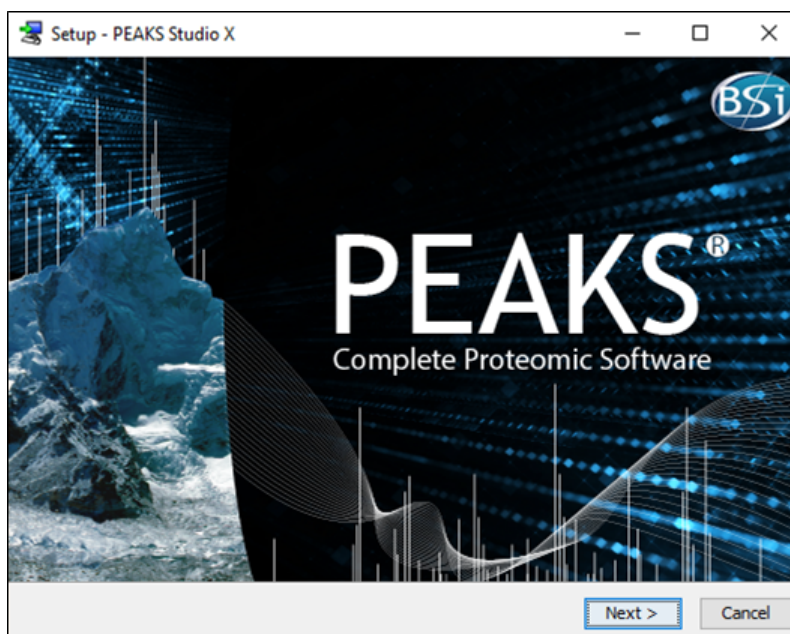
Avoid installing PEAKS in any directory that contains a white space (for example, the Program Files directory). PEAKS has some features that may not function correctly in such situations. Please make sure that the user account has full access permissions (read/write/execute) on the selected directory.



## Important

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

1. Close all programs that are currently running.
2. Insert the PEAKS USB stick into a USB port. 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 USB is inserted. If it does not, find the drive the USB was connected to and open it to access the contents. Double-click on `PEAKS_Studio.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.

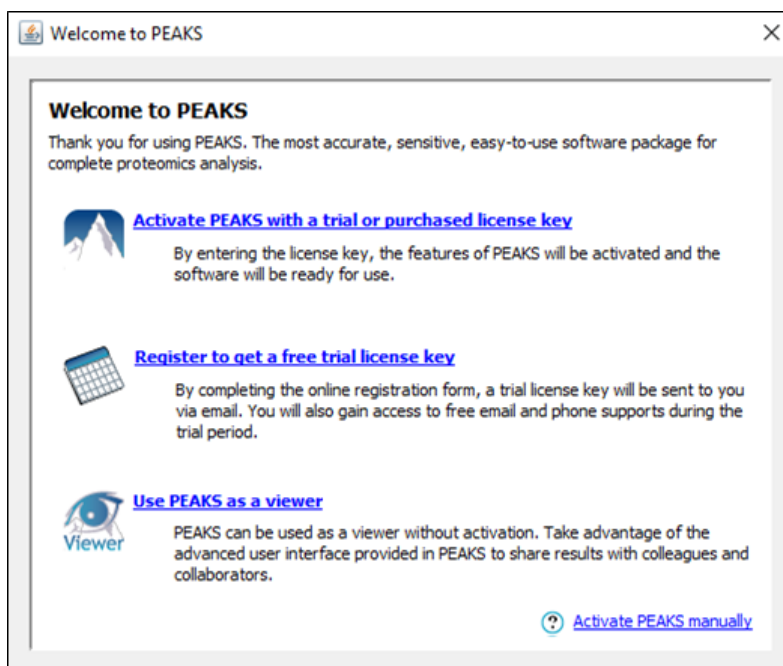
## Note

Microsoft's Visual C++ Redistributables 2013 and 2017 are required to load Thermo and Bruker data. Please make sure these two versions are installed if it is expected that Thermo and/or Bruker data will be used. Refer to Section 8, “Vendor-Specific Requirements” for details.

## 4. Activation

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



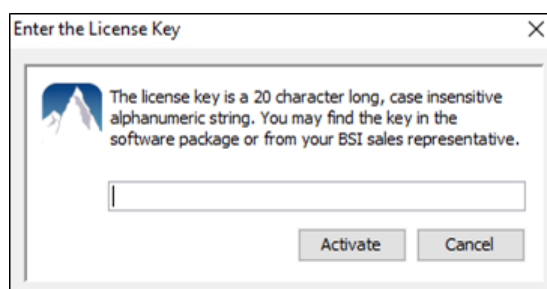


## Important

Keep the license key safe. After a computer hardware upgrade, it might be required to reactivate 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 fails with the message 'An error occurred while communicating with BSI licensing server ...'. Refer to Section 4.4, "Activate PEAKS Manually" for manual activation.



### 4.2. Register to Get a Free Trial License Key

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

## Important

Each computer can only have one free trial. Requesting a trial license repeatedly will not extend the trial period automatically. Please contact <support@bioinfor.com> if you have any questions.

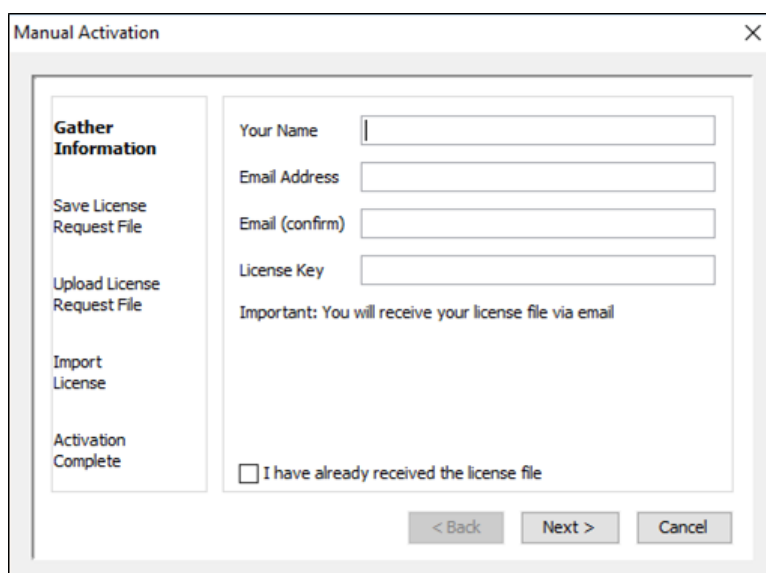
### 4.3. Use PEAKS as a Viewer

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

### 4.4. Activate PEAKS Manually

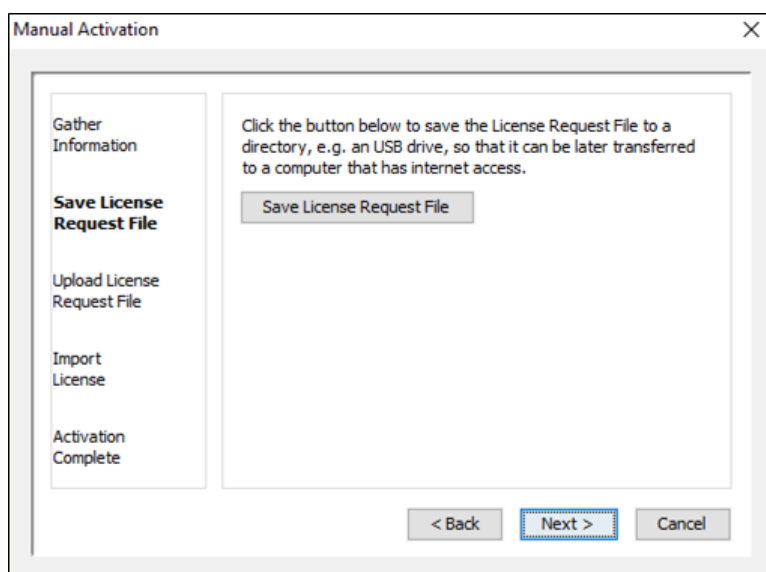
In certain situations, when the computer does not have an Internet connection (or is behind a firewall that blocks the activation), the activation process requires the assistance of another computer with an Internet connection (or outside the firewall). Manual activation can be accessed via the link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

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



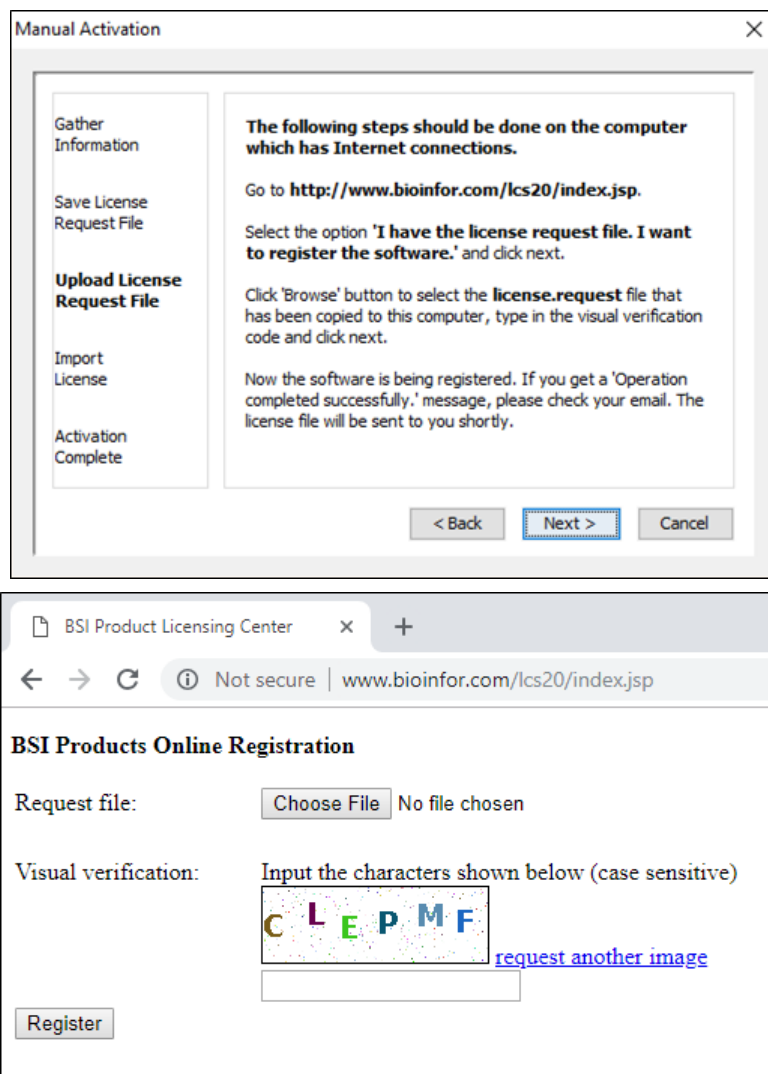
The screenshot shows the 'Manual Activation' wizard window. On the left is a vertical list of steps: 'Gather Information' (highlighted), 'Save License Request File', 'Upload License Request File', 'Import License', and 'Activation Complete'. The main area contains input fields for 'Your Name', 'Email Address', 'Email (confirm)', and 'License Key'. Below these fields is a note: 'Important: You will receive your license file via email'. At the bottom of the main area 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).

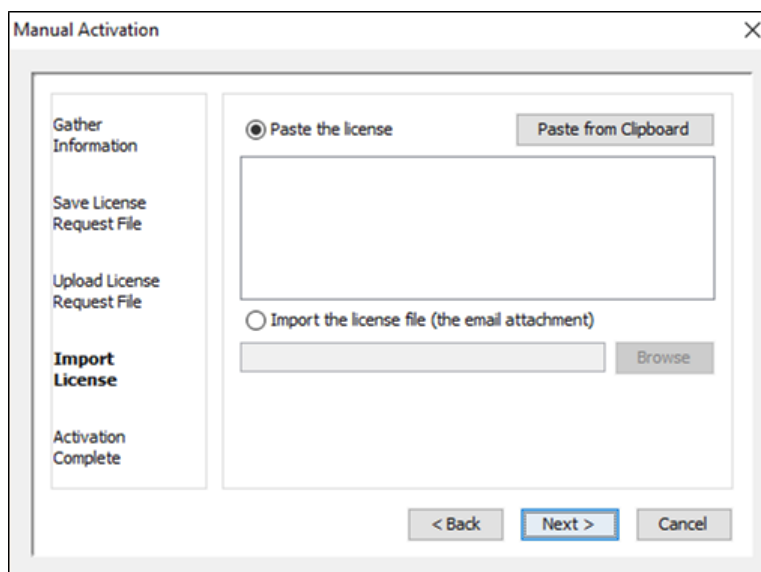


The screenshot shows the 'Manual Activation' wizard window at the second step. The left sidebar now highlights 'Save License Request File'. The main area contains the text: 'Click the button below to save the License Request File to a directory, e.g. an USB drive, so that it can be later transferred to a computer that has internet access.' Below this text is a button labeled 'Save License Request File'. At the bottom of the window are three buttons: '< Back', 'Next >' (which is highlighted with a blue border), and 'Cancel'.

- From another computer with an Internet connection, upload the license request file to BSI's license server as described in the following screen.

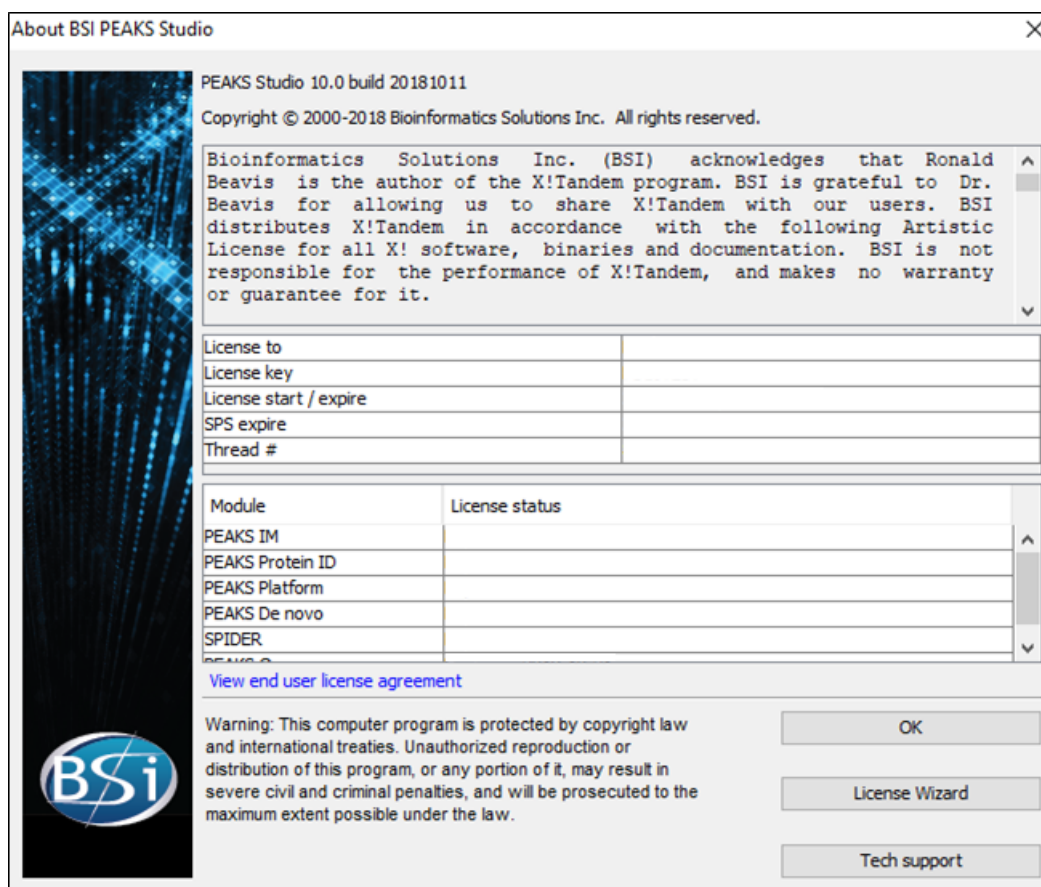


- If the license request is sent successfully, an automated BSI service will generate the license file (license.lcs) and send an email (from <support@bioinfor.com>) to the email address provided to the License Wizard. Either save the attached license file or copy the content between '====>' and '<====' in the email to the Windows clipboard.
- Transfer the license file to the computer running PEAKS and import the license file into the License Wizard.



## 4.5. Re-Registering PEAKS

Re-registering PEAKS may be necessary when an additional software module was purchased or SPS was renewed. BSI will modify the license information accordingly on the server side. A new license file is required to make the changes effective. Select About PEAKS from the Help menu. The “About BSI PEAKS Studio” dialogue box will appear:



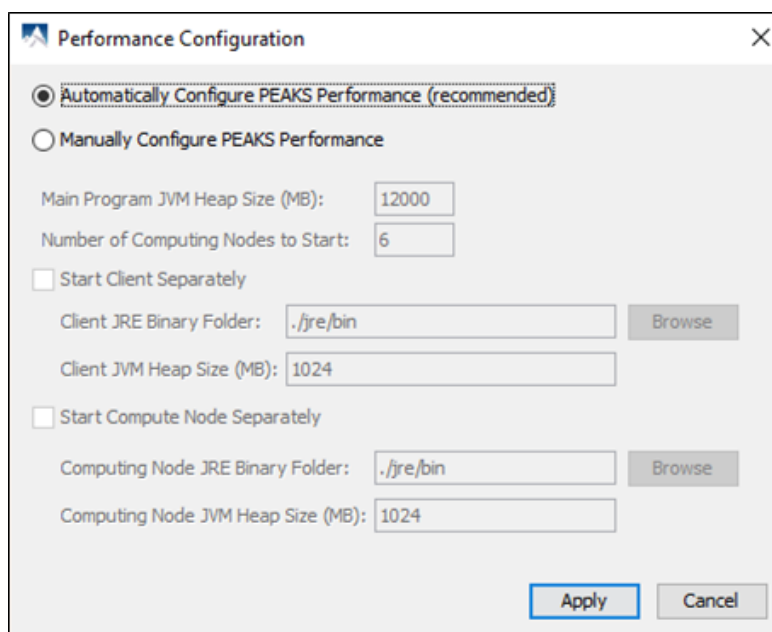
Click the License Wizard button to continue then follow the instructions in Section 4.1, “Activate PEAKS with a Trial or Purchased License Key” or Section 4.4, “Activate PEAKS Manually” to re-register PEAKS.

## 4.6. Common Errors During Registration

1. The license key contains only English letters and numbers. It is recommended to copy (Ctrl-C) / paste (Ctrl-V) the license key whenever possible.
2. The user information can only contain English characters (letters, digits and symbols). Characters from a non-English encoding may cause the activation to fail.
3. If the computer is behind a firewall or has other Internet connection problems, the activation may fail. Please follow the on-screen instructions or refer to Section 4.4, “Activate PEAKS Manually”.

## 5. PEAKS Performance Configuration

The PEAKS Performance Configuration tool can be accessed from the Windows Start Menu in the PEAKS Studio menu item. By default, PEAKS automatically determines its performance parameters to take full advantage of the processors and memory available on the computer. In most situations, the **Automatically Configure PEAKS Performance** option should be used. The **Manually Configure PEAKS Performance** option will come in handy in situations where advanced users want to adjust PEAKS memory allocation and threading.



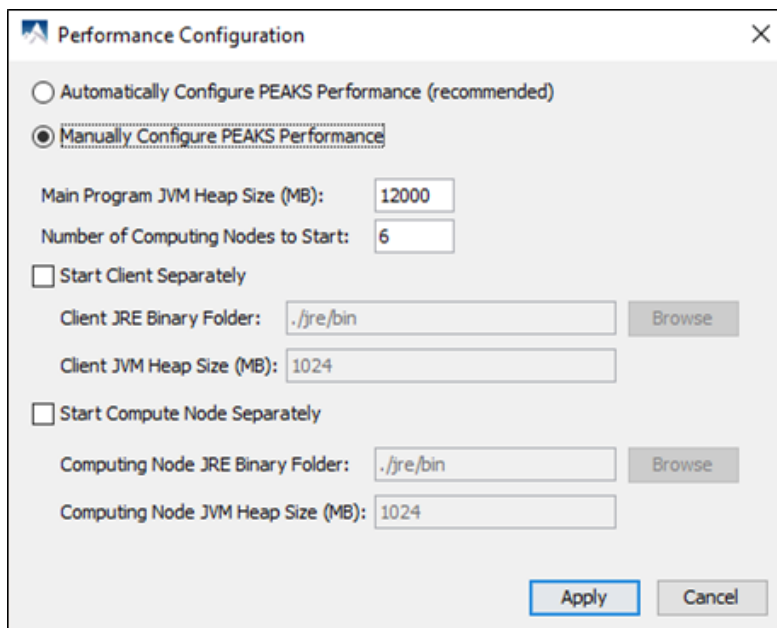
The **Number of Computing Nodes to Start** determines the number of parallel processing threads/processes that will be used by PEAKS. A higher number usually lets PEAKS run faster if the computer still has extra processing power. The maximum number of computing nodes PEAKS may start with is determined by the license; a *Desktop License* allows up to 16 nodes and a *Workstation License* allows up to 32 nodes.

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

### 5.1. PEAKS 64-bit Version

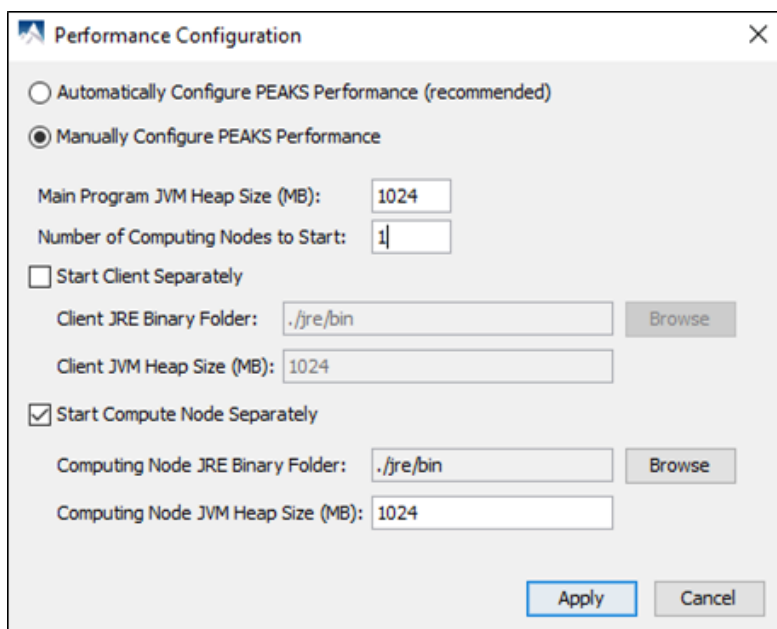
The 64-bit version of PEAKS can access a large amount of memory; therefore, it is highly recommended to deselect the options **Start Client Separately** and **Start Compute Node Separately**. The **Main Program JVM Heap Size (MB)** is the maximum amount of memory PEAKS can use. Although a higher number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

$16000 \times 80\% = 12800$  MB memory is configured to the main program JVM.  $12800/2000 = 6.4$ , therefore 6 computing nodes is configured.



## 5.2. PEAKS 32-bit Version

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



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

## 6. What's Next

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

- To read the instrument's raw data formats, installation of specific software libraries to support the instrument may be required. See 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*.

---

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

## 1. Overview

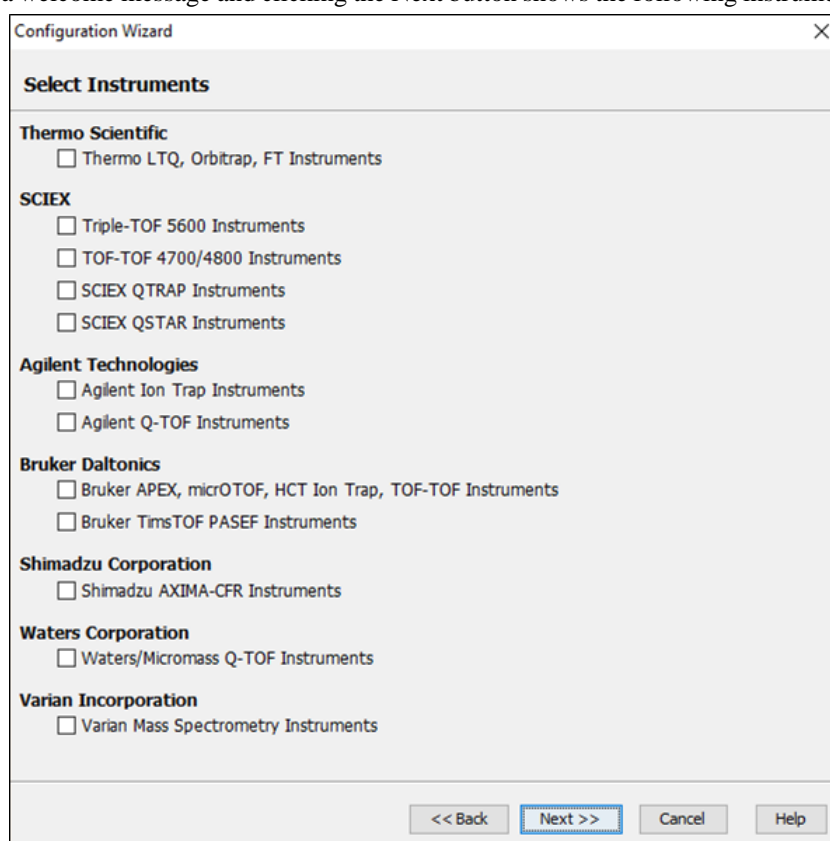
The Configuration Wizard guides users through easy-to-follow steps to configure PEAKS for instrument raw data support and database searching. The configuration wizard can be initiated from the menu Window > Config Wizard.

PEAKS supports different instrument vendors' raw data formats. A list of supported formats can be found in Chapter 4, Section 7, "Supported Data Formats". Some raw data file formats require the vendors' software to be installed on the same computer that PEAKS is running on. The Configuration Wizard helps to 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 to select the appropriate databases from a list of public sequence databases, and also downloads and configures them automatically. The list of public databases can be found in Section 3, "Database Selection".

## 2. Instrument Selection

The Configuration wizard can be run by selecting Window > Config Wizard from the menu. The Configuration Wizard displays a welcome message and clicking the Next button shows the following instrument selection panel.



The screenshot shows the 'Configuration Wizard' window with the 'Select Instruments' tab selected. The window lists various instrument vendors and their specific models, each with an unchecked checkbox. The vendors listed are Thermo Scientific, SCIEX, Agilent Technologies, Bruker Daltonics, Shimadzu Corporation, Waters Corporation, and Varian Incorporation. At the bottom of the window, there are four buttons: '<< Back', 'Next >>' (which is highlighted with a blue border), 'Cancel', and 'Help'.

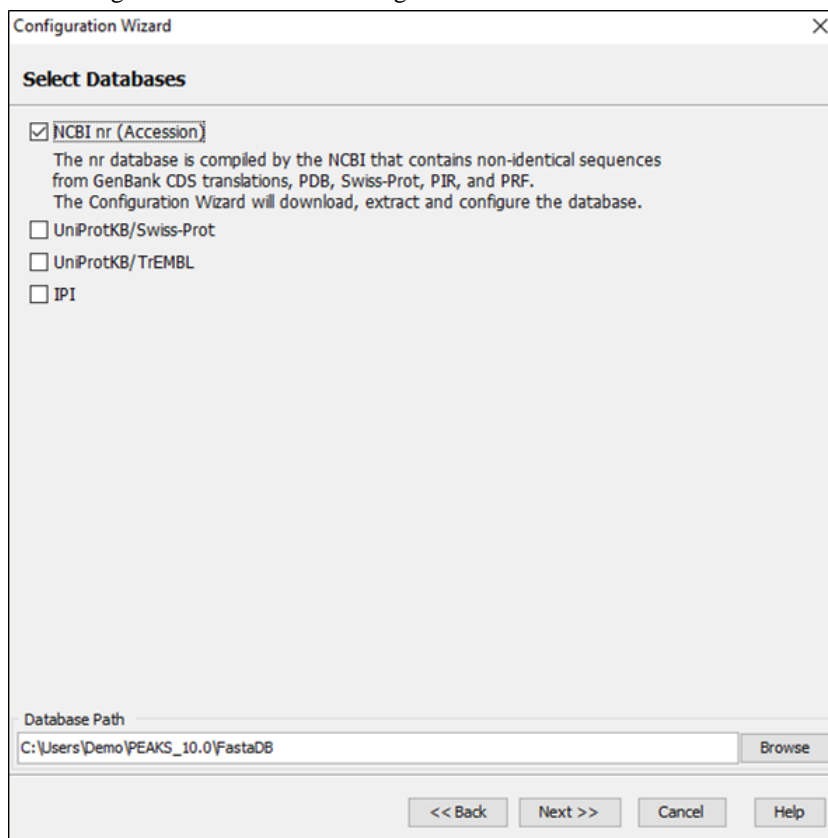
Vendor	Instrument Models
Thermo Scientific	Thermo LTQ, Orbitrap, FT Instruments
SCIEX	Triple-TOF 5600 Instruments, TOF-TOF 4700/4800 Instruments, SCIEX QTRAP Instruments, SCIEX QSTAR Instruments
Agilent Technologies	Agilent Ion Trap Instruments, Agilent Q-TOF Instruments
Bruker Daltonics	Bruker APEX, microTOF, HCT Ion Trap, TOF-TOF Instruments, Bruker TimsTOF PASEF 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

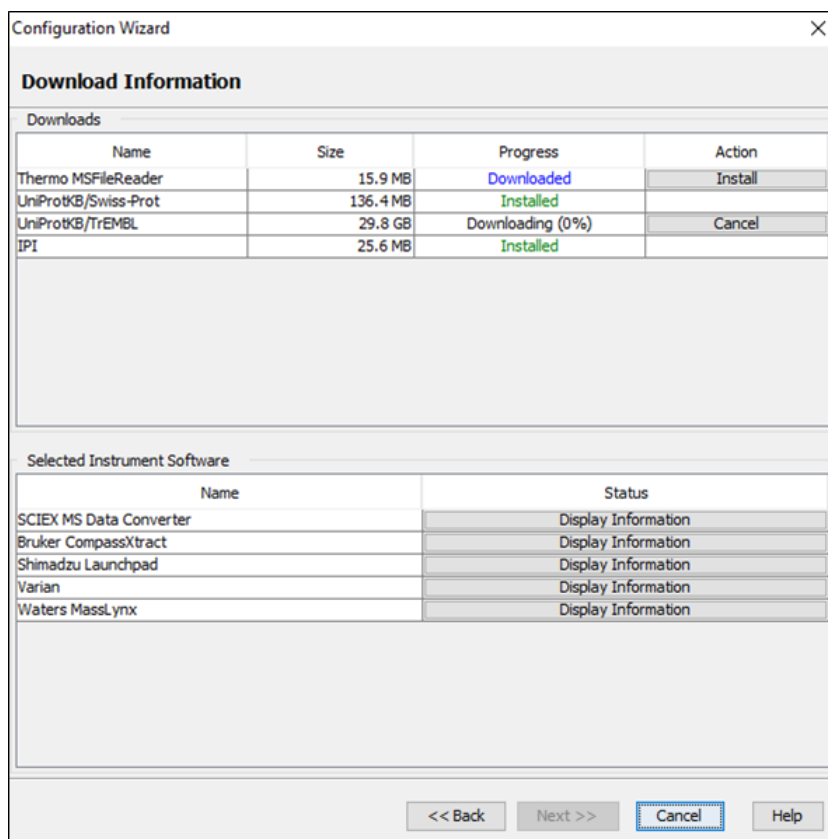
The database selection panel lists the public databases available for automatic download. The automatically downloadable databases are NCBI nr, Swiss-Prot, TrEMBL, and IPI human. Select the databases to be configured. A small description of the database and the action to be taken by the Wizard is displayed 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.



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

### 4. Download Information

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



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

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

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

## 5. Vendor Instrument Software Options

This panel allows users to set the raw file converter preferences of some vendor specific software. Refer to 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.

Configuration Wizard

### Vendor Instrument Software Preferences

**SCIEX (.wiff)**

☐ None  
PEAKS won't be able to load Sciex (.wiff) raw files.  
Please select a converter for raw file conversion to a standard format that PEAKS can load.

☐ ProteoWizard - MSConvert  
Location - msconvert.exe

☒ SCIEX MS Data Converter  
Location - AB\_SCIEX\_MS\_Converter.exe  
C:\Program Files (x86)\AB SCIEX\MS Data Converter\AB\_SCIEX\_MS\_Converter.exe

<< Back **Next >>** Cancel Help

Configuration Wizard

### Vendor Instrument Software Preferences

**Bruker (yep/baf/fid)**

☒ PEAKS Built-in Loader  
Options  
Bruker (.fid) data may contain several files. Do you want to merge them? ☐ yes ☒ no

☐ ProteoWizard - MSConvert  
Location - msconvert.exe

**Shimadzu (.run)**

☒ None  
PEAKS won't be able to load Shimadzu (.run) raw files.  
Please select a converter for raw file conversion to a standard format that PEAKS can load.

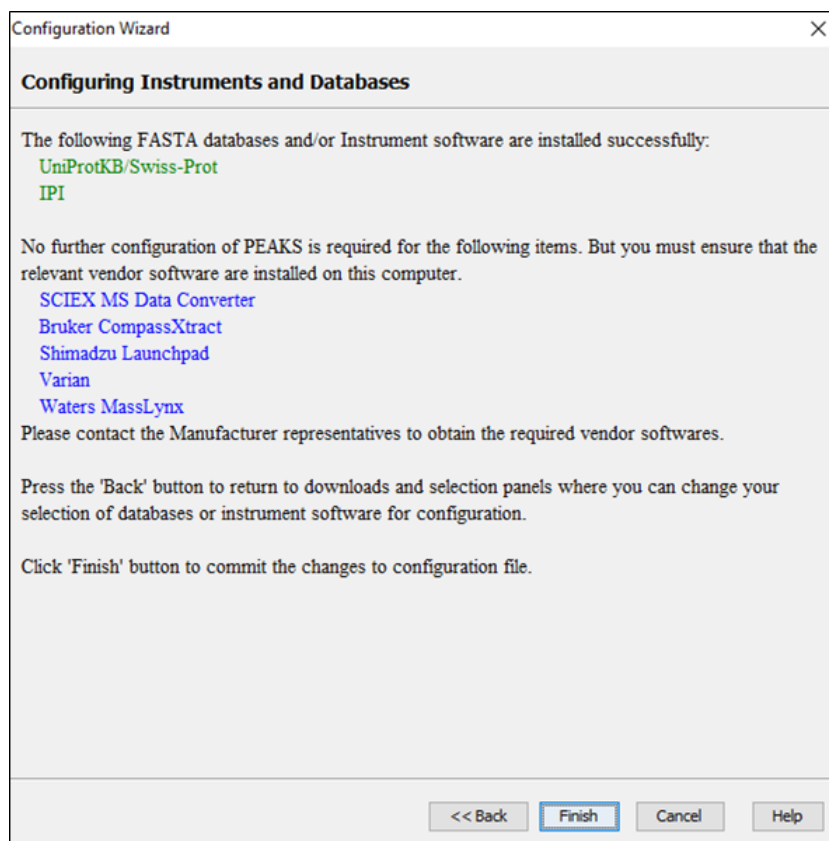
☐ Shimadzu Converter  
Location - run2xml.exe

<< Back **Next >>** Cancel Help

The screenshot shows a 'Configuration Wizard' window with the title bar 'Configuration Wizard' and a close button. The main heading is 'Vendor Instrument Software Options'. Under the 'Waters (.raw)' section, the 'PEAKS Built-in Loader' is selected. Below it, there is an 'Options' box containing a checkbox for 'Use lockmass calibration' (which is unchecked), a text field for 'Calibration mass' with the value '785.8426', and a note 'Calibration scans will be removed.'. Below this, the 'ProteoWizard - MSConvert' option is unselected, with a text field for 'Location - msconvert.exe' below it. The 'Varian (.xms)' section contains a paragraph: '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.' Below this is a text field for 'Default xmlrai.exe File Location' and a 'Browse' button. At the bottom of the window are four buttons: '<< Back', 'Next >>' (which is highlighted with a blue border), 'Cancel', and 'Help'.

## 6. Apply Changes

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




A summary of the actions performed by the Config Wizard will be displayed in the final screen. Click the Finish button to accept all changes and to complete the Config Wizard.

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

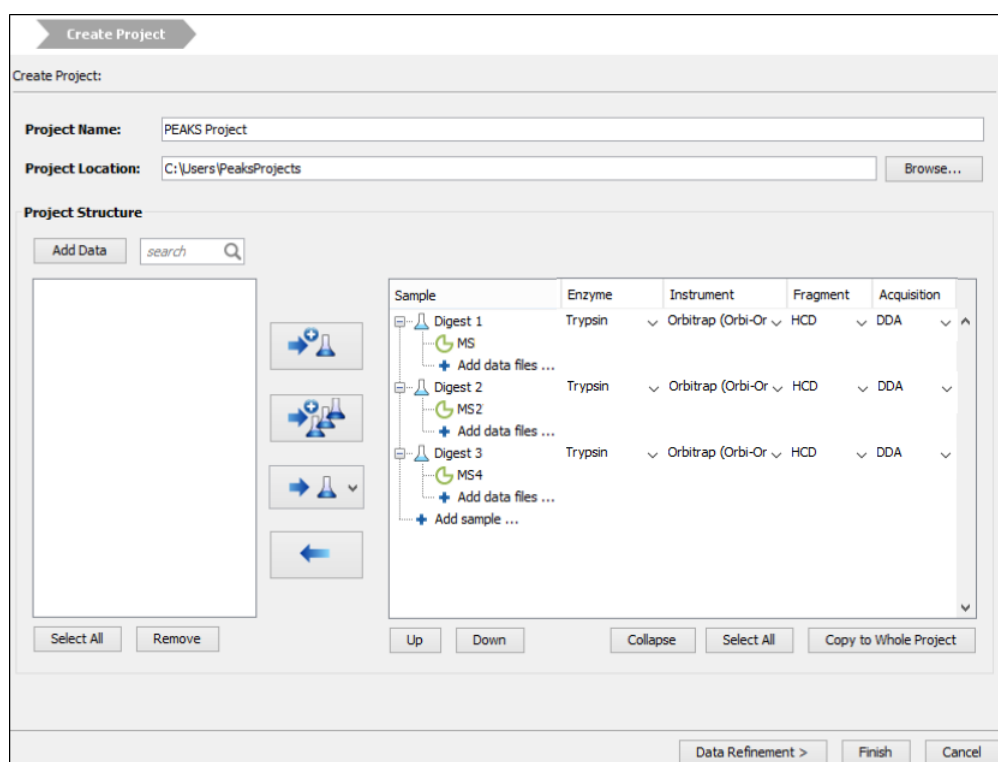
## 1. Overview

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 in the Project View pane. In the computer's file directory, a project is saved as a folder that contains multiple files, including the compressed spectral data and the analysis results. To share a PEAKS project, send the whole project folder to the shared location. The project can be opened in another copy of PEAKS Studio or the free PEAKS Viewer, given that it is the same version the project was created with.


The “Project Wizard” guides users through easy-to-follow steps to create a PEAKS project; set up project structure; load data into project; and set up the data analysis work flow for data refinement, identification, and quantification. To initiate the Project Wizard, simply click the New Project icon  on the tool bar or select File > New Project from the menu. The rest of this chapter discusses the details of the Project Wizard including: project creation, data loading and setting up a simple work flow.

### Note

The Project Wizard analyzes the data on a project level. Therefore, the work flow setup by the Project Wizard will be executed for all fractions of all samples in the project and included in a single result (ex. one result node for each *de novo*, PEAKS DB, PEAKS PTM, and SPIDER). To analyze the samples or fractions individually so that each will have their own result, click Finish after setting up the project structure and locating the source of the data files instead of moving onto the Data Refinement step. To automatically perform batch analysis on a multiple samples or fractions on an individual basis, please refer to Chapter 18, Chapter 18, *Workflow*







## 2. Create New Project

1. To create a new project, select File > New Project from the menu or click the New Project icon  on the tool bar. The “Project Wizard” window will appear.
2. Use the “Project Name” field to name your project. Click “Browse” to change the directory for saving the project. The selected directory will appear in the “Project Location” text box.



### Note



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

3. Use the Add Data button to browse to the directory of the data files, select the files you wish to load, and click Open. All of the selected data files will be listed on the left side.
4. This selected data can be placed into appropriate samples of the project structure using one of the following several options. All of the selected files can be placed in a new sample () , in an existing sample () , or in individual samples for each file () .
5. To add a sample to the project structure, click on the  Add sample ... button. From the drop-down menus, select the “Instrument” type, the “Fragmentation” method, and the “Enzyme” that was used to generate the experimental data from the drop-down lists. To apply the same sample details to the whole project, select the sample with the correct settings and click the Copy to whole project button.

### Note

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

6. Use the  Add data files ... button to add data files to the corresponding sample. Browse to the directory of the files, select the files to be loaded, and click Open.
7. To remove a sample or data file, select the appropriate node (sample or data file) in the project structure and press  button. To delete the file entries from the list on the left, select the files and click Remove.
8. To change the order of the samples within a project or the data files within a sample, use the Up and Down buttons.
9. Click the Finish button to exit the Project Wizard without specifying parameters for further analysis. Otherwise, click the Data Refinement button to proceed to the data refinement step. See Section 3, “Data Refinement - Setting Parameters” for details on data refinement parameters.

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


PEAKS supports all major raw data formats from vendor instruments. A list of supported formats can be found in Section 7, “Supported Data Formats”. Some formats may require that a specific software from the vendor be installed on the same computer on which PEAKS is running. Before creating a project, ensure that the vendor-specific requirements discussed in Section 8, “Vendor-Specific Requirements” are met. Once the data is loaded, it becomes part of the PEAKS project, and the original data files can be manipulated or deleted without affecting the analysis in PEAKS.

By default, the peptide feature area is used for peptide and protein abundance calculation. Users can also select to use feature intensity as the abundance calculation method. For intensity method, the highest signal intensities from

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

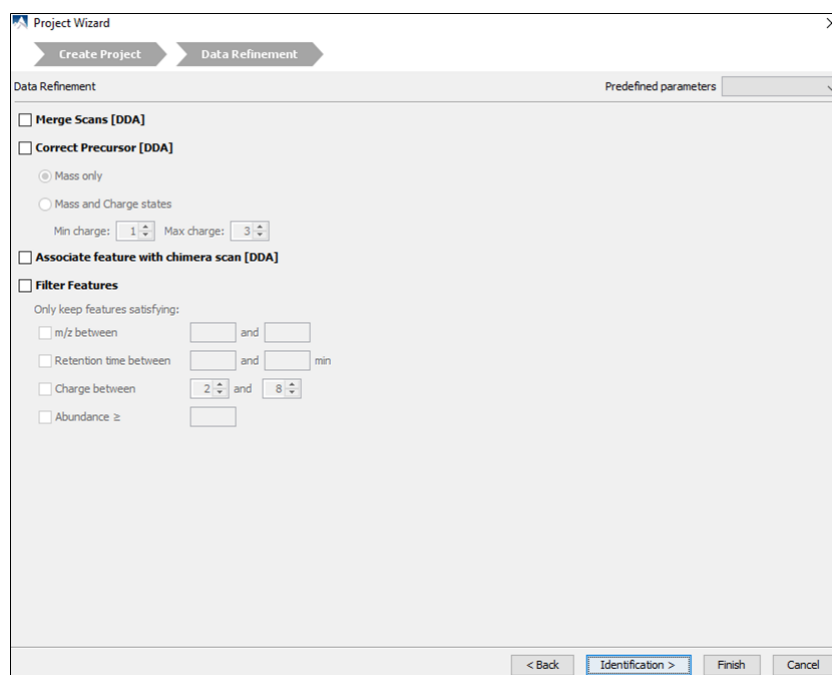
the feature's carbon isotopes are summed and used to calculate peptide and protein abundance. The abundance calculation method is set before creating a new project. One project can only be analyzed by one abundance calculation method. This information is shown in project properties when clicking the project name and looking for Abundance Calculation Method in the Properties tab. In the results, either Area or Intensity are displayed in the corresponding abundance columns.

Before creating a new project, go to C:\Users\[user name]\PEAKS\_10.0 folder and open PeaksInternal.properties file in Notepad. use\_intensity\_for\_abundance\_calculation=false means that **Area** is used for calculating abundance. To switch to use intensity instead, change false to true.

To close an opened project, select the project node in the Project View pane and perform any of the following options: from the menu bar, select File > Close Current Project, click the Close Project icon  from the tool bar, or right-click and choose Close Project from the pop-up menu. It is recommended to close unused projects to preserve computer memory.

### 3. Data Refinement - Setting Parameters

To run data refinement on each of the raw data files immediately after creating a project, click the Data Refinement button in the Project Wizard to move onto the “Data Refinement” step. See Chapter 7, Section 2, “Data Refinement Parameters” for more details on data refinement parameters.



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

### 4. DB Search - Setting Parameters

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



## Project Wizard - Create a PEAKS Project, Load Da- ta and Set up a Workflow

The Project Wizard dialog box is shown with the Identification tab selected. The tabs are Create Project, Data Refinement, and Identification. The PEAKS Search section includes a dropdown for Predefined parameters. The Error Tolerance section has fields for Precursor mass (10.0 ppm), using monoisotopic mass, and Fragment ion (0.05 Da). The Enzyme section has a dropdown for Specified by each sample, a New button, a Digest mode dropdown (Semispecific), and a Maximum missed cleavages per peptide field (2). The PTM section has a list of PTMs (Carbamidomethylation, Deamidation (NQ), Oxidation (M)) with buttons for Set PTM, Remove, and Switch type. The Database section has a Select database dropdown (UniProtKB/Swiss-Prot\_20180710), a View button, a Paste sequence field, a Taxa dropdown (Homo sapiens (human)), a Set/View taxa... button, a Contaminant database checkbox, and a View button. The General Options section has checkboxes for Estimate FDR with decoy-fusion (checked), Find unspecified PTMs with PEAKS PTM, and Find more mutations with SPIDER. At the bottom are buttons for Skip Identification, < Back, Quantification >, 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, load the data, run a data refinement on the fractions, and then conduct *de novo* sequencing and database searching on the data at the project level. See Chapter 8, Chapter 8, *Peptide De Novo Sequencing* and see Chapter 9, Chapter 9, *Peptide, PTM, and Mutation Identification (PEAKS DB, PEAKS PTM, and SPIDER)* for more details on *de novo* sequencing and database searching. Alternatively, click the Quantification button to set up parameters for quantification.

## 5. Quantification - Setting Parameters



To run quantification tools on the project immediately after setting database search parameters, click the Quantification button to open the “Quantification” panel. The optional add-on PEAKS Q module of PEAKS Studio supports both labeled and label-free quantification. With labeled quantification, PEAKS pre-configures commonly used reporter ion quantification (MS/MS-level) and precursor ion quantification (MS-level) methods. See Chapter 15, Section 2, “Setting Parameters”, Chapter 14, Section 2, “Setting Parameters”, and Chapter 13, Section 2, “Setting Parameters” for more details on quantification parameters.

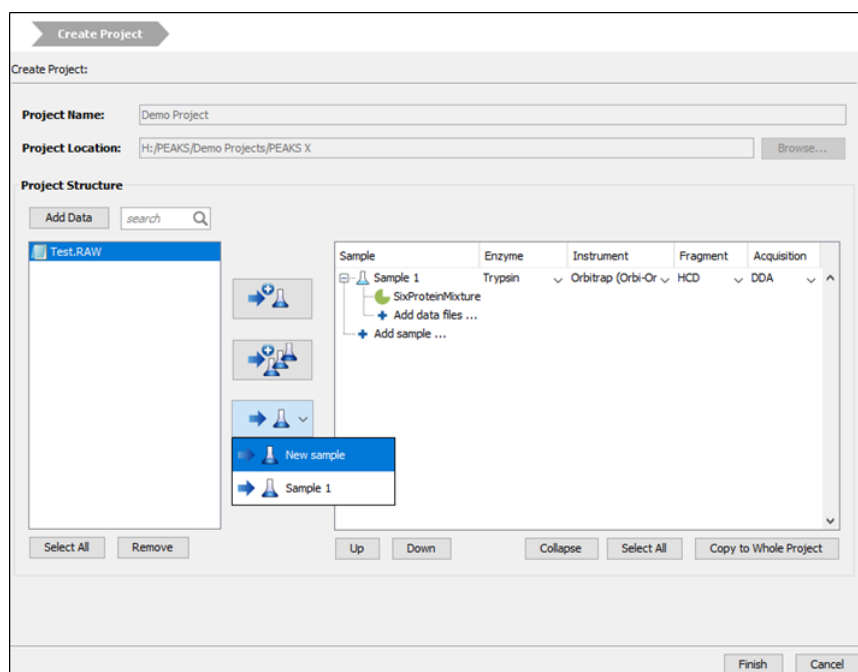
The Quantification panel is shown with the Label Free Quantification tab selected. The Tools section has radio buttons for Label Free, Reporter Ion, and Precursor Ion. The Basic Options section has fields for Mass Error Tolerance (20.0 ppm) and Retention Time Shift Tolerance (5.0 min). The Sample Groups section has a search sample list field, a list of sample groups (Digest 1, Digest 2, Digest 3), and buttons for Remove All, Collapse, and Attach Identification. The Attach Identification section has buttons for Attach... and Remove. At the bottom are buttons for < Back, Next >, Finish, and Cancel. A note at the bottom states: "The identification result from the previous step will be automatically attached."







## 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 data refinement, and then conduct *de novo* sequencing, database searching, and quantification at the project level. See Chapter 15, Chapter 15, *PEAKS Q - Label Free*, Chapter 14, Chapter 14, *PEAKS Q - MS/MS Level* and Chapter 13, Chapter 13, *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 from the menu, the File > Open Recent Project list, or click the Open Project icon  from the tool bar.
2. To add data to an open project, click on the project from the “Project View” pane and select File > Add Data from the menu, or use the Add Data icon  from the tool bar. The “Project Wizard” window will open.



3. Use the Add Data icon  to browse to the location of the data files. Select the files to be loaded and then click Open. All of the selected data files will be listed on the left side.
4. The selected data from the list on the left can be placed into an appropriate sample within the project structure. There are several options to enable this placement. All of the selected files can be placed in a new sample () , in an existing sample () , or in individual samples for each file () .
5. More files can be added to an existing sample by clicking the Add Data icon  . Additional samples can be created by clicking the Add Sample  Add sample ... button.
6. From the drop-down menus, select the “Enzyme”, “Instrument” type, and the “Fragmentation” method that was used to generate the experimental data. For more information on adding files/samples or setting up the instrument configuration refer to Section 2, “Create New Project”.

### Note

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

## 7. Supported Data Formats

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

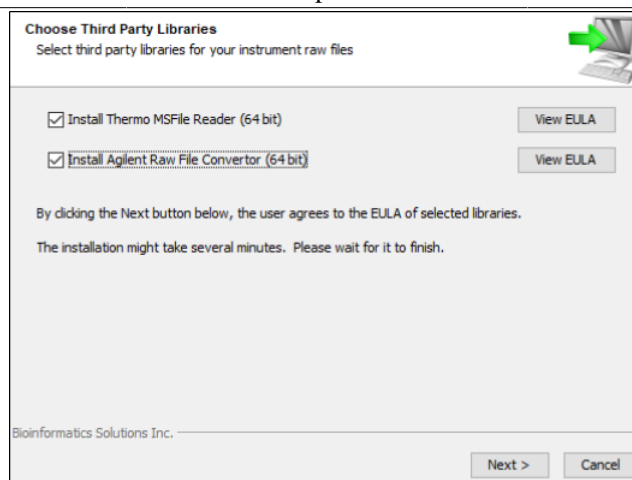
- Native Support - PEAKS can read the following files directly without any additional tools:
  - mzXML
  - mzData
  - mzML
  - .DTA file or a directory of .DTA files
  - .MGF
  - .PKL
  - PEAKS 8.5, PEAKS 8 and PEAKS 7.5 projects. See Section 8.8, “Older PEAKS Projects”.
- Library Level Support - The instrument vendor's software library must be installed on the same computer as PEAKS. PEAKS will call the software library to read the data directly.
  - .D directory (Agilent instruments). See Section 8.1, “Agilent Data”.
  - .RAW file (Thermo Fisher Scientific instruments). See Section 8.2, “Thermo Data”.
  - .LIFT or .D directory (Bruker instruments). See Section 8.4, “Bruker Data”.
  - .RAW directory (Waters QTOF instruments). See Section 8.7, “Waters/Micromass (MassLynx) Data”.
- Converter Level Support - Third party converters are required. Users need to install the required converters correctly and configure their locations in PEAKS. This only needs to be set up once. PEAKS will call the converter to convert the data to a native supported format before loading it. The actual conversion process is invisible to the user.
  - .RAW directory (Waters QTOF instruments). See Section 8.7, “Waters/Micromass (MassLynx) Data”.
  - .WIFF file (Sciex QSTAR and QTRAP instruments). See Section 8.3.1, “TripleTOF, QSTAR, or QTRAP”.
  - .T2D file (AB 4700/4800 series). See Section 8.3.2, “TOF-TOF 4700/4800”.
  - .RUN folders from Shimadzu instruments. See Section 8.5, “Shimadzu Data”.
  - .XMS files from Varian instruments. See Section 8.6, “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 preprocessing tools built into the software that are designed to maximize identification results. Listed below are the requirements to load raw data from each supported vendor.

### 8.1. Agilent Data

Agilent data can be loaded, provided that you have the Mass Hunter Data Access Component installed (MHDAC). Users can do so either by selecting the Agilent check box when first installing the software,



or through the PEAKS installation folder using the following steps:

1. Locate the PEAKS Studio installation directory. (default: C:\PEAKSStudioX\)
2. Access the folder, 'MHDAC\_x86 for 32-bit systems, or 'MHDAC\_x64' for 64-bit systems
3. Right-click on the Windows batch file 'RegisterMassHunterDataAccess' and click on "Run as administrator."
4. If this process was successful, Agilent data will load into PEAKS automatically.

## 8.2. Thermo Data

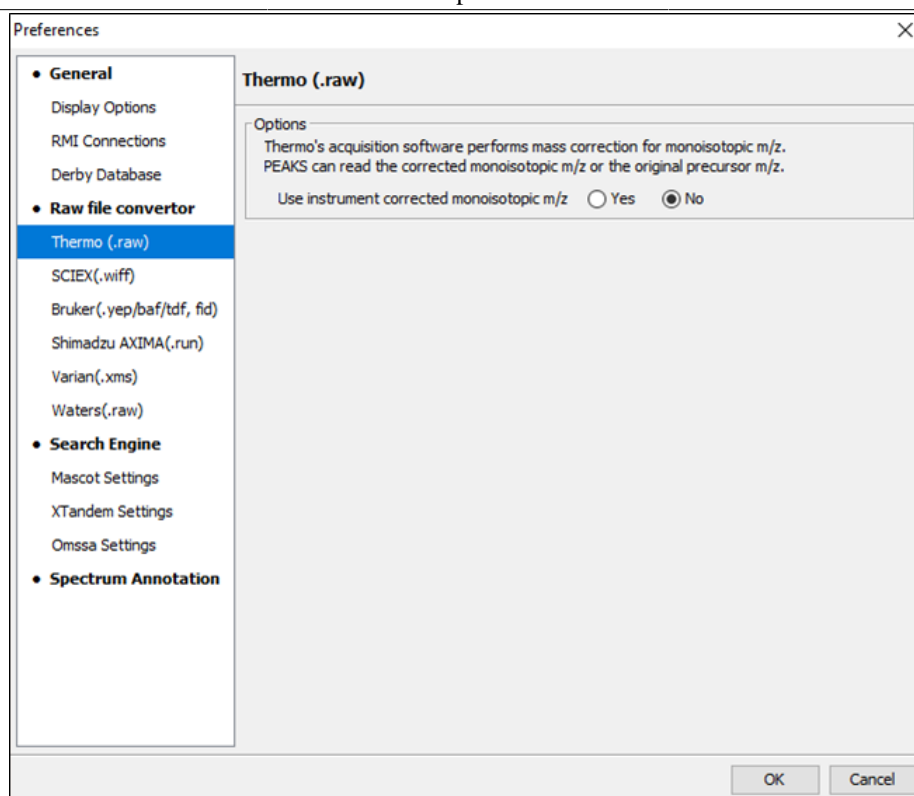
RAW data from Thermo Fisher Scientific mass spectrometers can be loaded, provided that the Thermo MSFileReader package is installed on the same computer as PEAKS. The PEAKS Configuration Wizard can download and install MSFileReader automatically (see Chapter 3, Section 2, "Instrument Selection"). MSFileReader is publicly available and can be found at the following link:

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

Thermo's acquisition software can perform monoisotopic precursor mass correction during data acquisition. Select "Yes" to load the corrected monoisotopic precursor mass from raw data. In such cases, users may skip precursor mass correction step in Data Refinement.

### Note

For raw data acquired using Advanced Peak Determination (APD) method, it is strongly suggested to select **Yes** for this option to load data accurately and get better identification results.

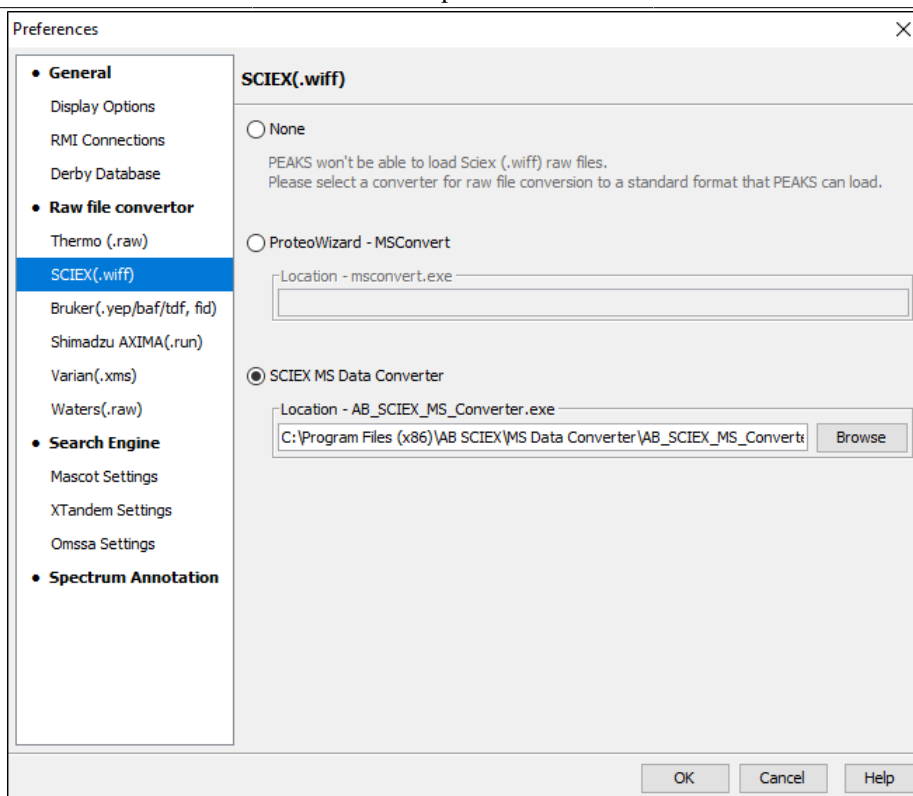


## 8.3. Sciex Data

### 8.3.1. TripleTOF, QSTAR, or QTRAP

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

- Sciex MS Data Converter can be used to load .wiff files into PEAKS directly. It can be downloaded from the following link, <https://sciex.com/software-support/software-downloads>. Once installed, use the browse button within the Sciex Data Converter option to find AB\_SCIEX\_MS\_Converter.exe. By default, it is installed in C:\Program Files (x86)\AB SCIEX\MS Data Converter\AB\_SCIEX\_MS\_Converter.exe
- ProteoWizard - MSConvert is the recommended method for loading .wiff files. It can be downloaded from <http://proteowizard.sourceforge.net/>. Once installed, select the ProteoWizard -MSConvert button and use the corresponding browse button to find msconvert.exe.



### 8.3.2. TOF-TOF 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 on another machine that has direct network access to the 4700 SERVER. There cannot be a firewall or proxy between the computers. Windows 7 or Windows XP is recommended for the use of this tool.

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

- **4700 SERVER Name or IP Address:** input “localhost” if the Extractor is running on the same computer as ABI 4700 Explorer (this is the default value). Otherwise, enter the IP address of the 4700 SERVER.
- **The socket used by the 4700 SERVER:** the port that the Oracle database listens to (the default is 1521).
- **Username to access the Oracle database:** most likely does not need to 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 that will be exported.


## 8.4. 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, .tdf (for timsTOF data) or fid file. Select the .D folder [containing .yep, .baf, and .tdf (if timsTOF data) 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 3.2 is readily available on the Bruker Daltonics website at the following link: <http://www.bruker.com/service/support-upgrades/software-downloads/mass-spectrometry.html>. Once you reach the website, Bruker may request a username and a password

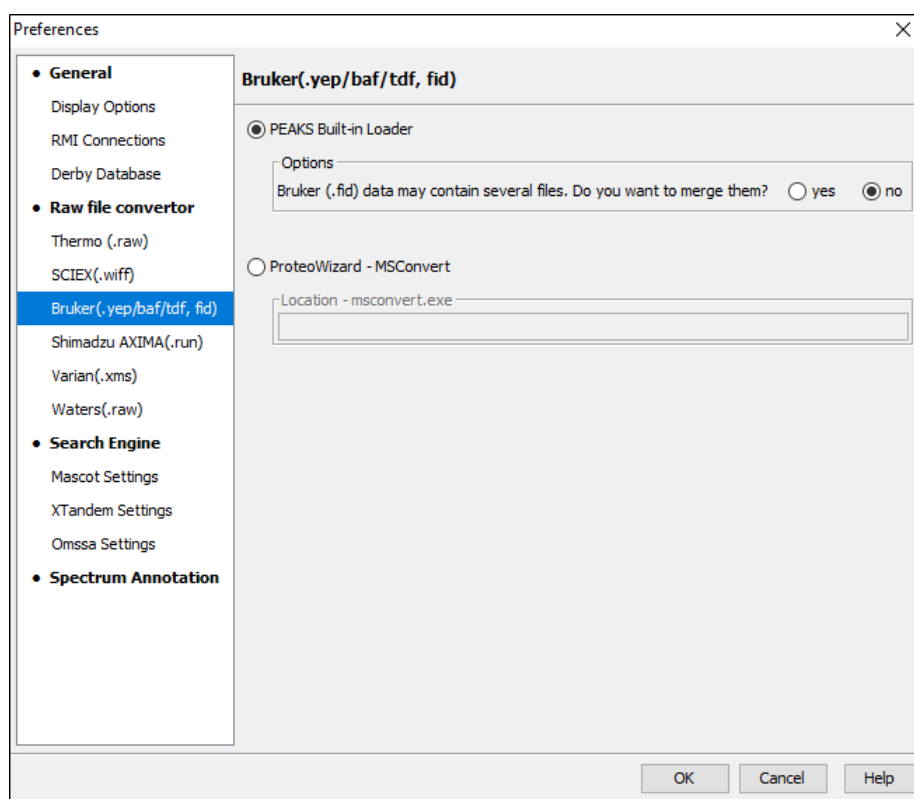
### Note

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

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


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

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

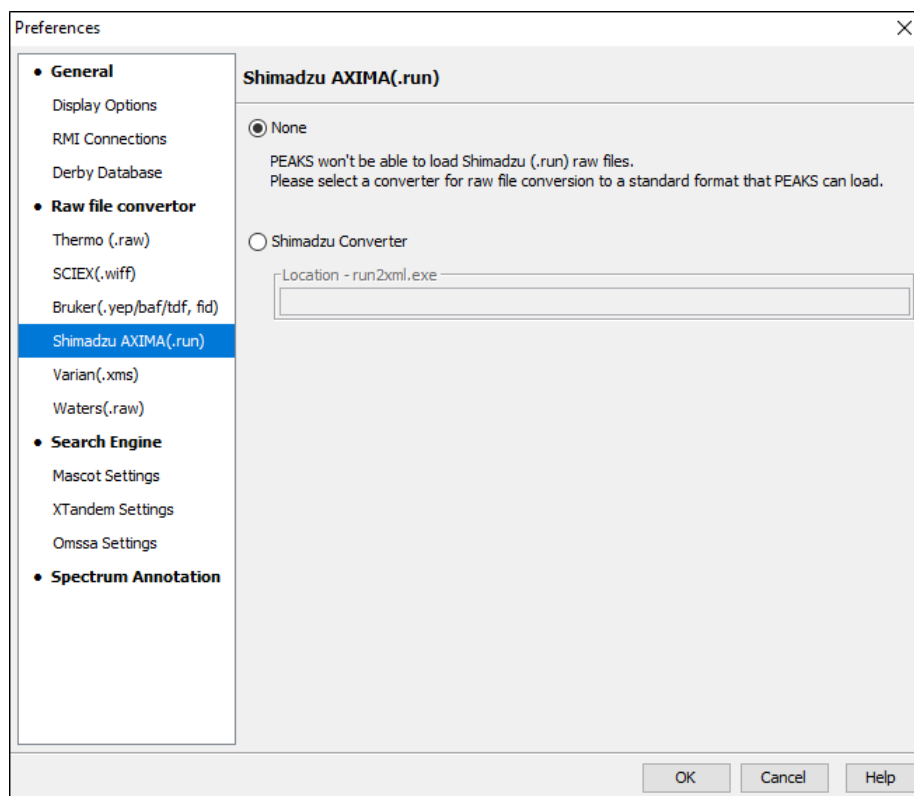


## 8.5. Shimadzu Data

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” tool bar icon  or by selecting Preferences from the Window menu. Click “Instrument” and then “Shimadzu AXIMA (.run)” in the menu on the left. This will show the Shimadzu instrument preferences on the right side. Select the Shimadzu Converter button and click Browse to tell PEAKS the location of the Shimadzu `run2xml.exe` file.

Select None to disable Shimadzu data loading



## Note


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

## 8.6. Varian

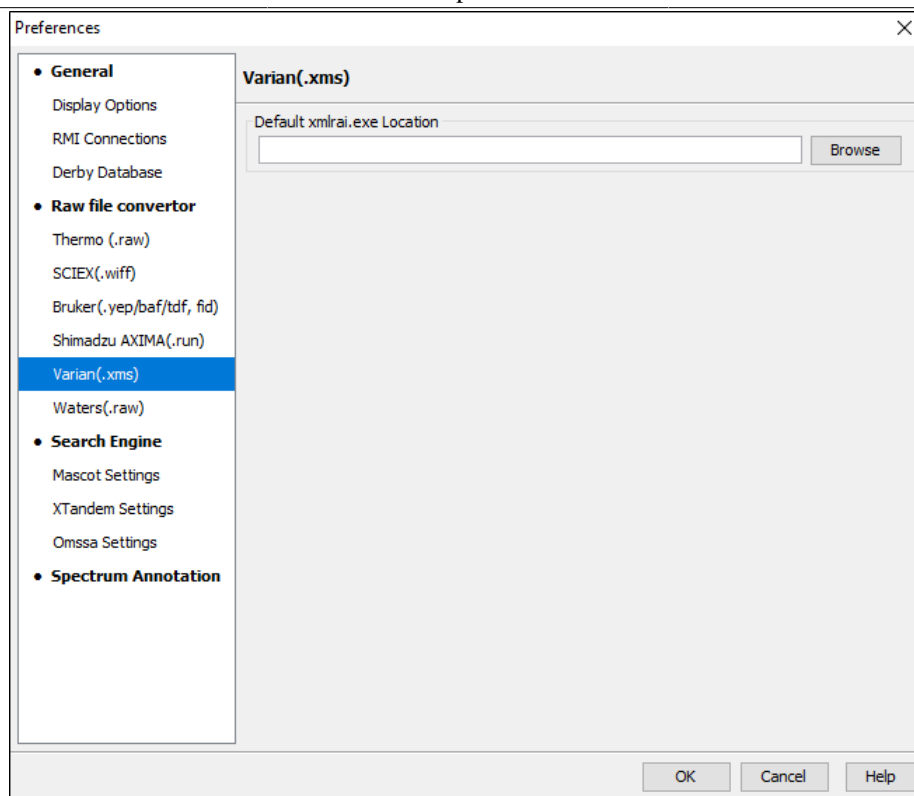
A conversion tool is embedded into Varian’s data acquisition software that allows for conversion of Varian raw data into `.pk1` files that can be immediately read by PEAKS.

The `.trans` data files can be converted in Varian programs by clicking **File > Save As** and selecting either the `.pk1` file format or by performing the following steps: clicking File, right-clicking Export, and then selecting `.pk1`. If you are viewing a chromatogram with the Varian Software, all the spectral data in the viewed chromatogram are converted to `.pk1` 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” tool bar 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.

## 8.7. Waters/Micromass (MassLynx) Data

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

- PEAKS Built-in Loader

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

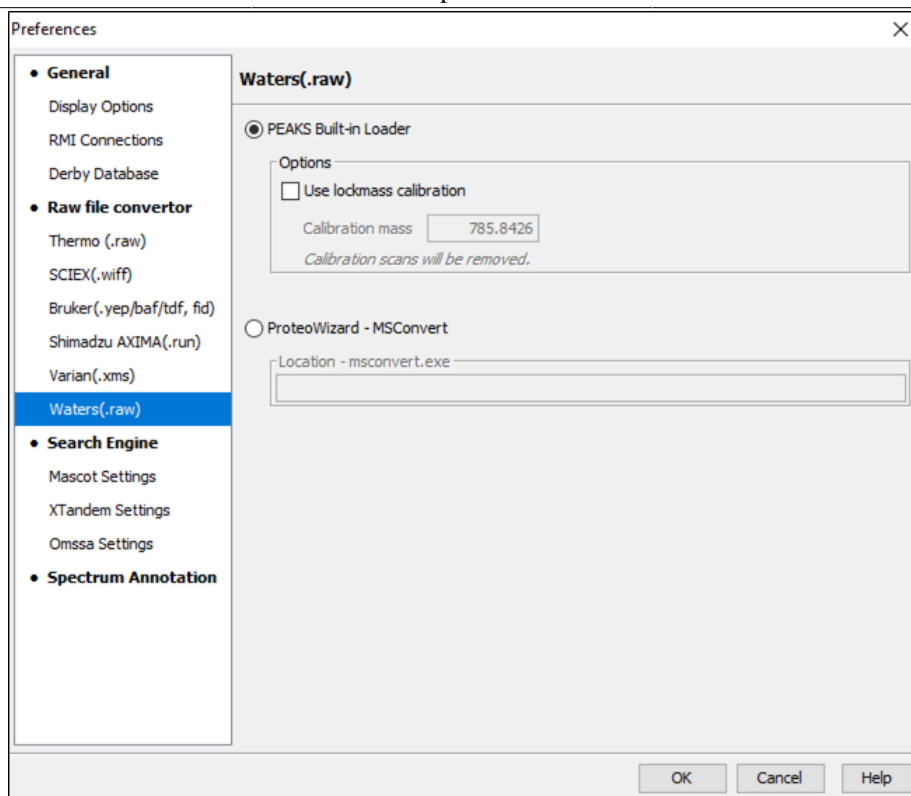
With the PEAKS Built-In Loader, users also have the option to specify whether or not PEAKS should consider lockmass calibration for their Waters `.raw` data files.

- ProteoWizard - MSConvert

Select this option if the built-in loader is unable to load some `.raw` files as an alternative. Browse to the location of `msconvert.exe`, which is located by default in:

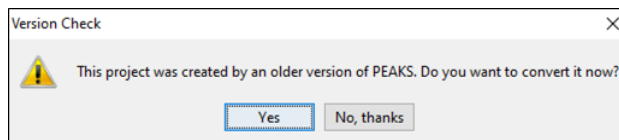
`C:\Program Files\ProteoWizard\ProteoWizard\msconvert.exe`

ProteoWizard can be downloaded from: <http://proteowizard.sourceforge.net/>

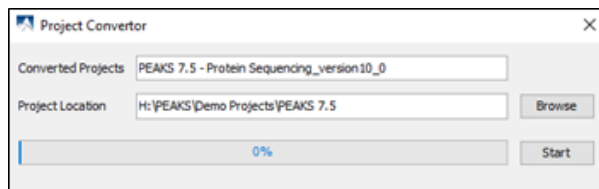


## 8.8. Older PEAKS Projects

Projects created in PEAKS 8.5, PEAKS 8, and PEAKS 7.5 software can be opened in PEAKS X. To convert the project to a PEAKS X project, open the project in the same way you would open any existing PEAKS project. The project will be recognized as a PEAKS project from an older version. The following confirmation dialogue box will pop up for PEAKS 7.5 projects (PEAKS 8 and PEAKS 8.5 projects will load automatically).



Choose Yes to convert the project and proceed. The following “Project Converter” dialogue 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.

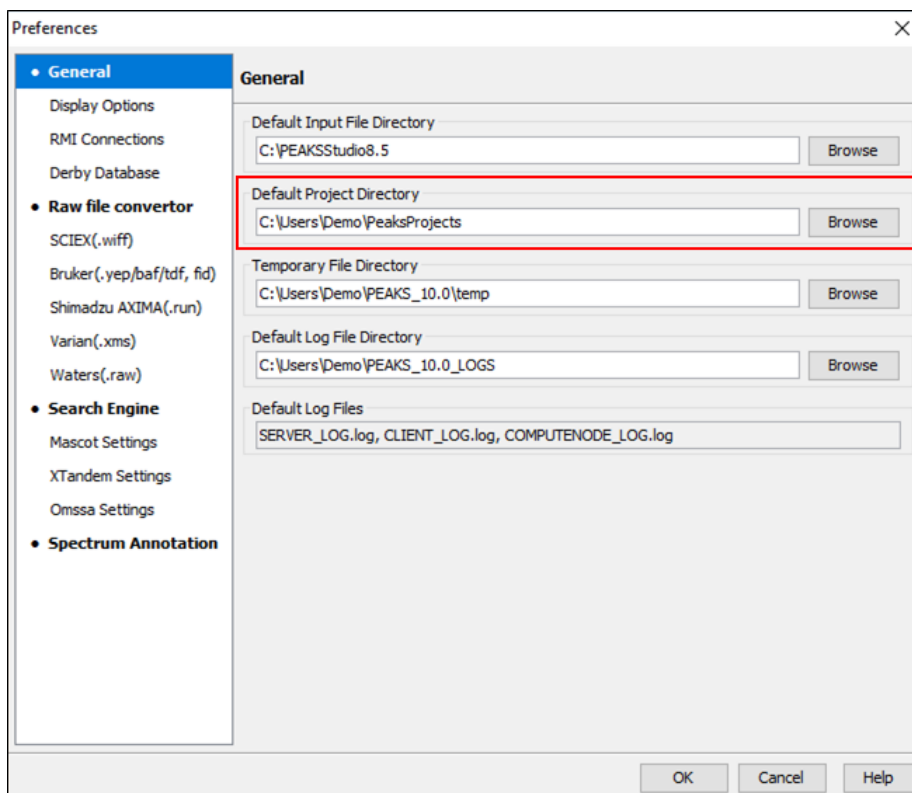
### Note

The project converter is designed to convert PEAKS 7.5 projects so that they can be viewed in PEAKS X, but some new features may not be available for display from older PEAKS projects (including PEAKS 8 and PEAKS 8.5). Further analysis on the converted projects are strongly discouraged.

## 9. Changing the Default Project Location

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

1. Click the Preferences icon  from the tool bar. The "Preferences" dialogue pops up.



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

---

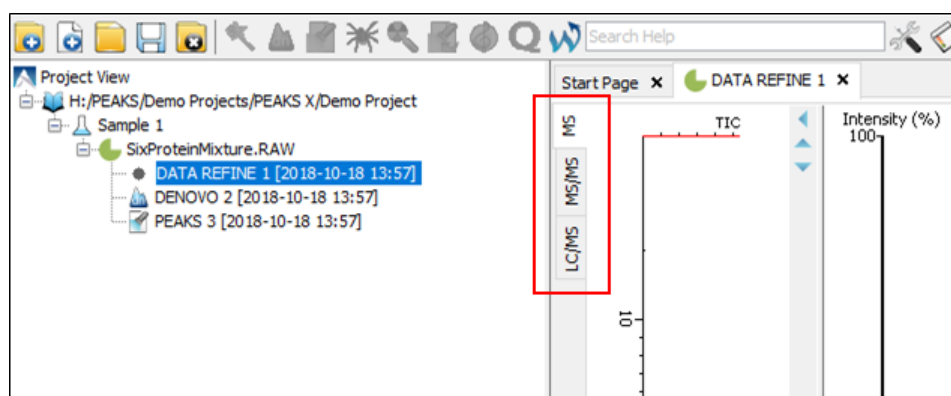
# Chapter 5. Data Visualization

## 1. Overview

After a project is created, the spectral data can be visually examined. The information of a typical DDA LC-MS/MS raw data file is organized into 3 sections:

- **MS** - This section shows the TIC (total ion chromatogram) plot and all of the MS scans. For each MS scan, the corresponding MS/MS scans are also displayed.
- **MS/MS** - This section lists all the MS/MS scans. For each MS/MS scan, the corresponding features are listed below and associated number of features is displayed next to the MS/MS scan information header.
- **LC/MS** - This section 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 from the Project View pane, the choice of different views can be made by choosing different tabs at the upper-left corner of the data view window.

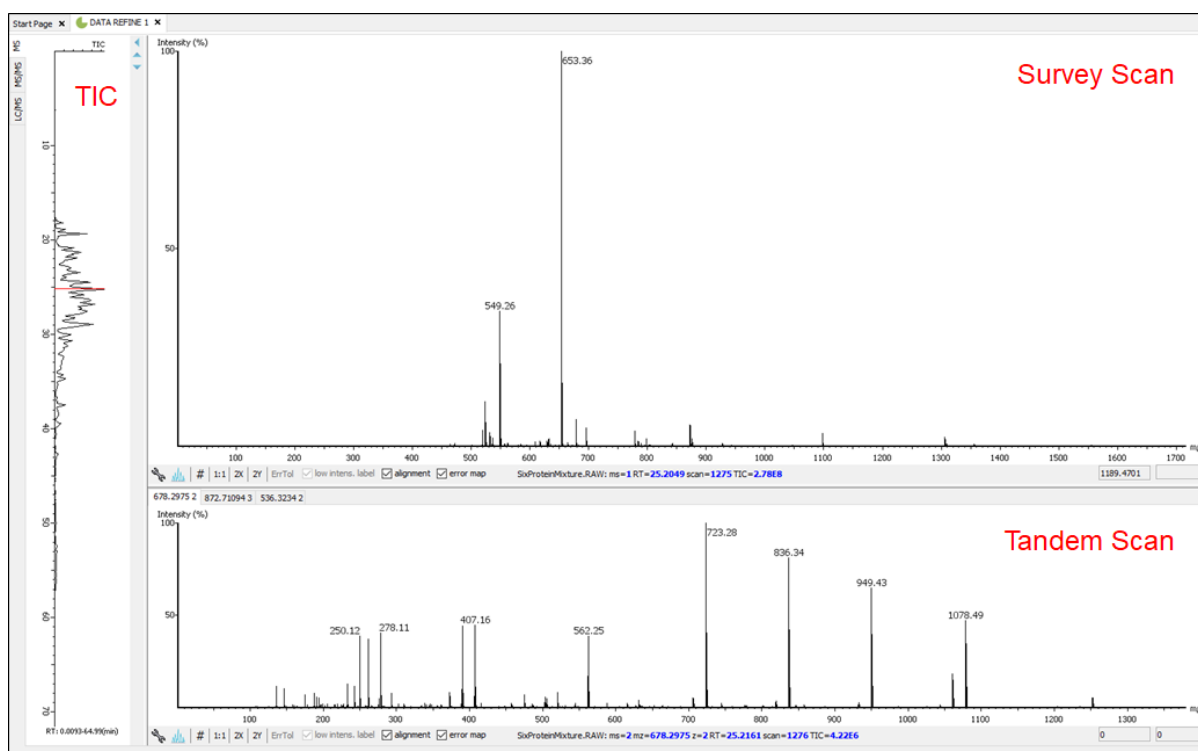


### Note

Please refer to Chapter 16, *Data Independent Acquisition (DIA) Data Analysis* and Chapter 17, *Ion Mobility Data Analysis* for data visualization details for DIA and IMS data respectfully.

## 2. MS View

The MS View contains the TIC and all of the scans. The total ion chromatogram (TIC) is displayed on the left of the MS view. The navigation buttons are highlighted in the figure. To collapse the TIC chart, click the left navigation button. Clicking on a specific position in TIC will display the corresponding survey (MS) scan. To navigate the MS scans, use the up and down navigation buttons. The MS scans can also be navigated through using the up and down arrows of the keyboard. The tandem (MS2) scans associated with the selected MS scan are shown in the panel below, where each can be viewed by selecting the different tabs identified by their "m/z z" information.



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.

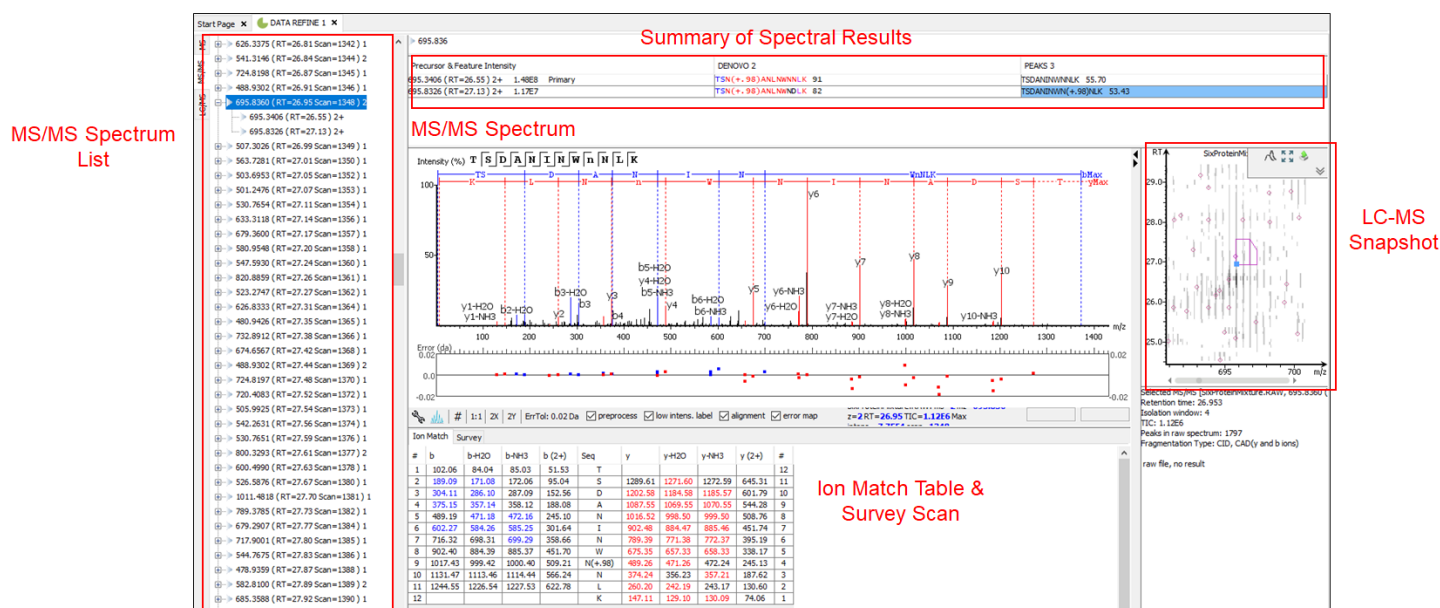
### 3. MS/MS View

The MS/MS View shows the list of tandem scans on the left. If data refinement has been performed on the data file, each tandem scan in the list will indicate the total number of associated features at the end of the spectrum header [ex. "695.8360 (RT=26.95 Scan=1348) 2" from the image below indicates that there are 2 peptide features associated with the selected tandem scan]. In addition, when data refinement has been performed, a tandem scan header can be expanded and the associated features and their details (e.g. m/z, RT, charge) are listed below. When a tandem scan is selected from the list, the following information is provided to the right of the list: a summary of the spectral results, the MS/MS spectrum, the zoomed in LC/MS view, the Ion Match Table, and its Survey Scan. Similarly, when an individual peptide feature is instead selected from the list, the results for that feature are displayed. Zoom options are the same as described in the "MS View" section.

- **Summary of Spectral Results:** A table with a summary of the spectral results are displayed here, including the associated feature information, *de novo* sequence (if *de novo* sequencing has already been performed), and database ID sequence (if a PEAKS search has been performed).
- **MS/MS Spectrum:** A view of the tandem scan will be displayed here. If *de novo* sequencing and/or a PEAKS search has been performed for the data file, the scan will be annotated. By clicking the sequence results from the Spectral Result table, the corresponding annotation will be displayed. Users will also have the ability to turn on the alignment and error map for the spectrum view when *de novo* sequencing and/or a PEAKS search is performed.
- **LC-MS Snapshot:** The LC-MS Snapshot displays a zoomed-in view of the precursor scan in the LC-MS heat map 2-D view. The MS/MS scan is indicated by a red framed square and the span of the yellow line designates

the isolation window. If data refinement has been completed, purple circles will also be displayed on the heat map, and selecting an individual feature from the list on the left will highlight the feature in purple.

- **Parent Scan:** The “Survey” tab displays the precursor scan with the best signal within the isolation window. The buttons that appear in this section are identical to those explained in Section 2, “MS View”.
- **Ion Table:** If *de novo* sequencing and/or a PEAKS search has been performed, an “Ion Match” tab will appear when one of the sequence results from the Spectral Results table is selected. The “Ion Match” tab in the bottom pane contains the displays a table with possible fragment ions for the selected MS/MS scan. Each ion in the table shows the calculated mass, where if the fragment ion is present in the spectrum, its mass value is displayed in color. N-terminal ions are shown in blue and C-terminal ions are shown in red. A fragment ion is found when there is a matching peak within the mass error tolerance, as defined in the *de novo* sequencing parameters, and when relative intensity of the matching peak is at least 2%. The ion types displayed in the table can be configured in “Spectrum Annotation Settings” (Section 3.2.2, “Spectrum Annotation”).

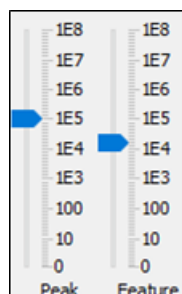


- **Move to a specific heat map area:** Click on the heat map, hold and drag the cursor until the desired area of the heat map is reached and then release the mouse button.
- **Zoom to a specific heat map area:** Right-click on the heat map, hold and drag the cursor to select the desired area of the heat map and then release the mouse button.
- **Move to a specific heat map area from thumbnail:** Current view range of the heat map is shown as a rectangle in the Thumbnail. Click on the thumbnail view range rectangle, hold and drag the cursor until the desired area of the heat map is reached and then release the mouse button.
- **Jump to a specific heat map area:** Provide approximate m/z and RT values separated by a space in the Locate: m/z RT box and then hit the Enter key to zoom in to an area centered at the specified m/z and RT value.
- **Zoom in/out from thumbnail:** Place the cursor on the Thumbnail and then scroll the mouse wheel button.
- **Zoom in/out smoothly:** Place the cursor at a particular (m/z value, RT) position and then scroll the mouse wheel button.
- **Move to a specific RT range:** Click the cursor on the TIC or the RT axis, and then hold and drag the cursor until the desired range is reached.
- **Zoom to a specific RT range:** Place the cursor on the TIC or the RT axis, and then scroll the mouse wheel button.
- **Move to a specific m/z range:** When the m/z axis is zoomed in, click the mouse on the m/z axis, and then hold and drag the cursor until the desired range is reached.
- **Zoom to a specific m/z range:** Place the cursor on m/z axis and then scroll the mouse wheel button.
- **Change contrast of the heat map:** Place the cursor on the heat map, press and hold the Ctrl key, and then scroll the mouse wheel button.
- **Change peak height in the 3D intensity mode:** Place the cursor on the heat map, press and hold the Ctrl key, and then scroll the mouse wheel button.
- **View the whole heat map:** Click the 1:1 button in the control.
- **Move back/forward to previously specified heat map area:** Use the Back and Forward buttons in the control to navigate the previously specified heat map areas.

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

## 4.1. Noise Level

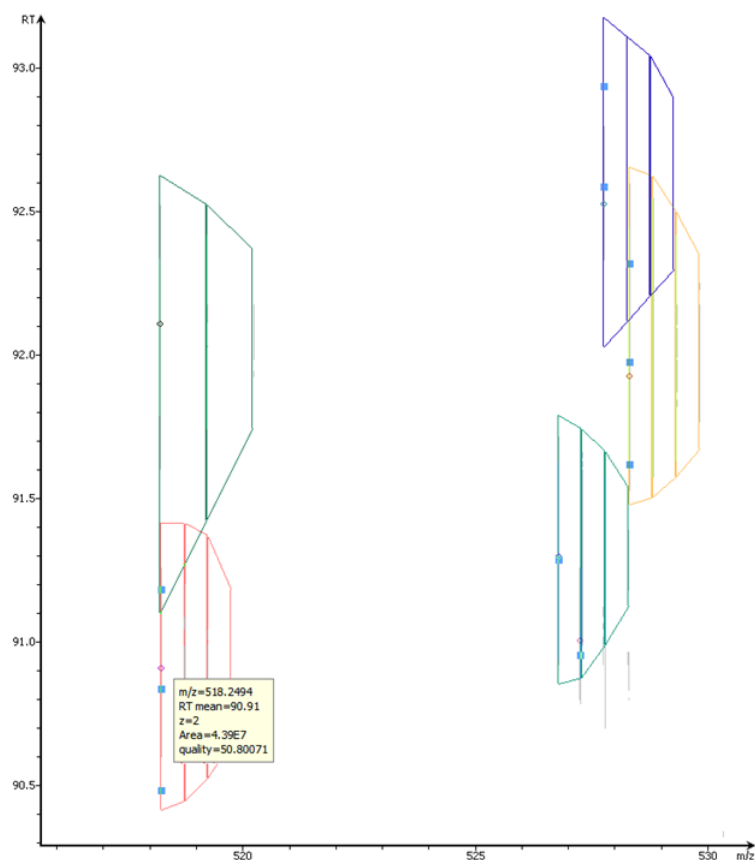
Select the appropriate threshold for noise filtering by peak and/or feature intensities. Once selected, the heat map view will reflect the changes dynamically.



## 4.2. Show/Hide Features

To view the detected LC/MS features, mark the Features checkbox in the control. The features detected during the data refinement steps will be highlighted by small, red, diamond-shaped points on the heat map. Placing the cursor

over such a feature point will show the boundary of the feature using a red polygon and display a pop-up window that includes the following information: the monoisotopic  $m/z$ , the mean RT, the charge state of the feature, and the total area under the XIC curves of the feature.




## Note

Multiple features can be selected and represented by different colors by holding Ctrl and selecting the feature points.

## 4.3. Locate M/Z and Retention Time

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

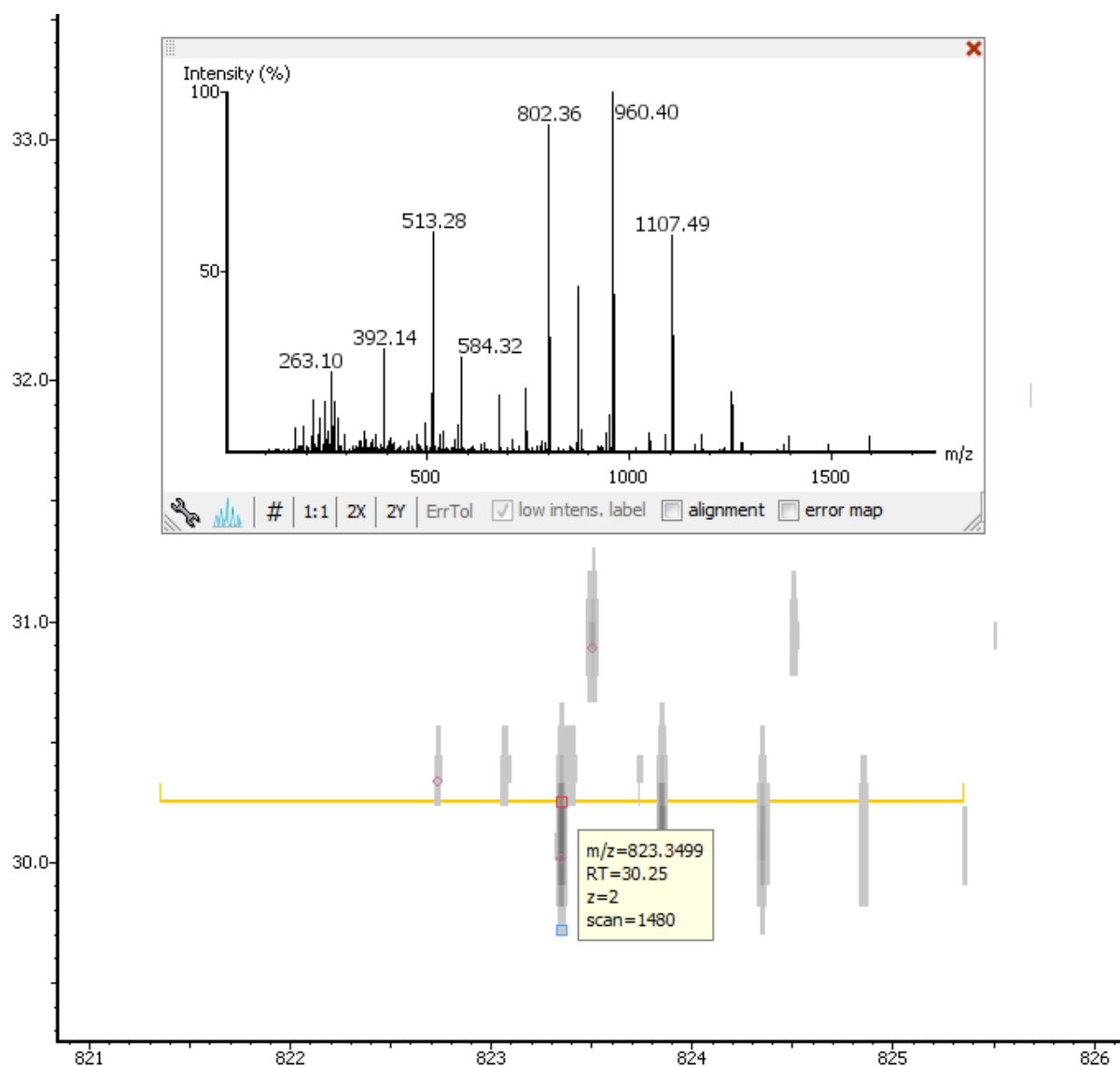
## 4.4. Synchronize LC/MS View Between LC Runs

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


## 4.5. Show/Hide MS/MS Spectra

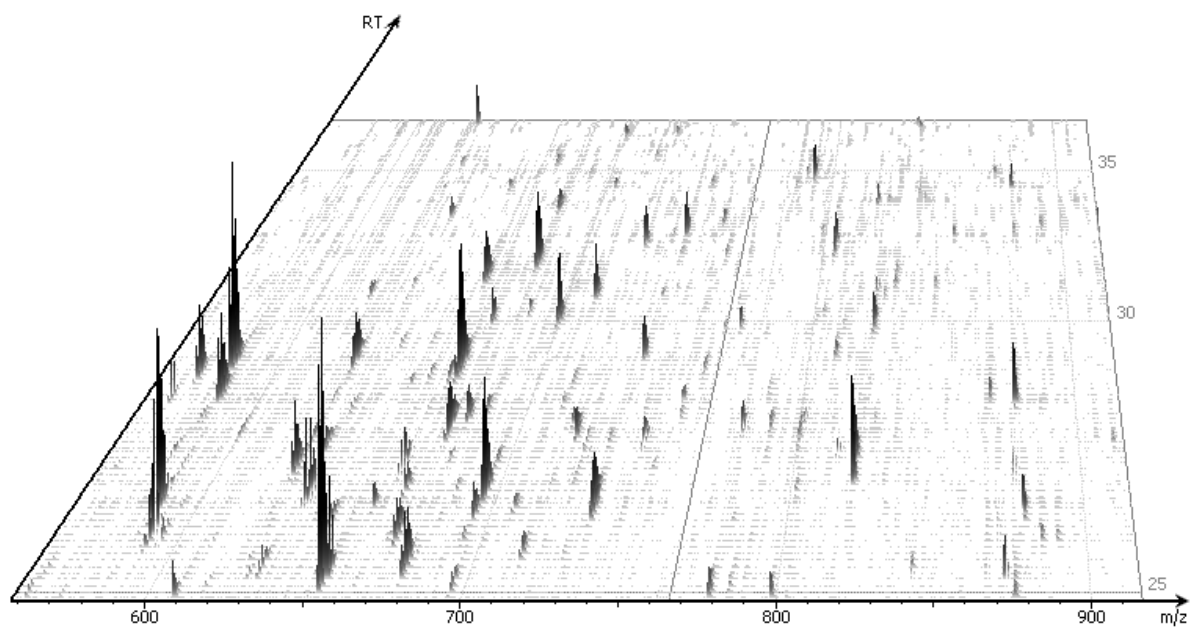
Select the MS/MS Spectra checkbox to indicate where a tandem scan has been taken. The location will be marked with blue-outlined squares on the heat map. Place the cursor over a square to view the precursor  $m/z$ , the RT, the precursor charge state, and the scan number of that spectrum. Click on the square to view the acquisition window (indicated by the yellow line), and the spectrum in a pop-up window. Once selected, the square will become red-outlined.





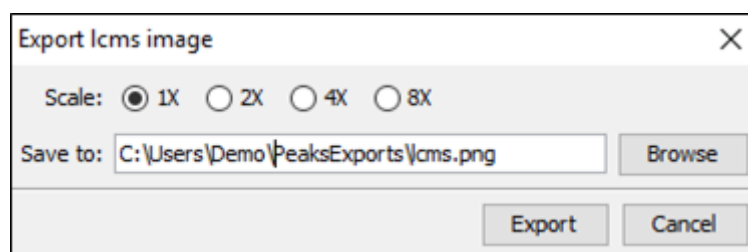
## 4.6. Intensity View

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



## 4.7. Export LC/MS


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

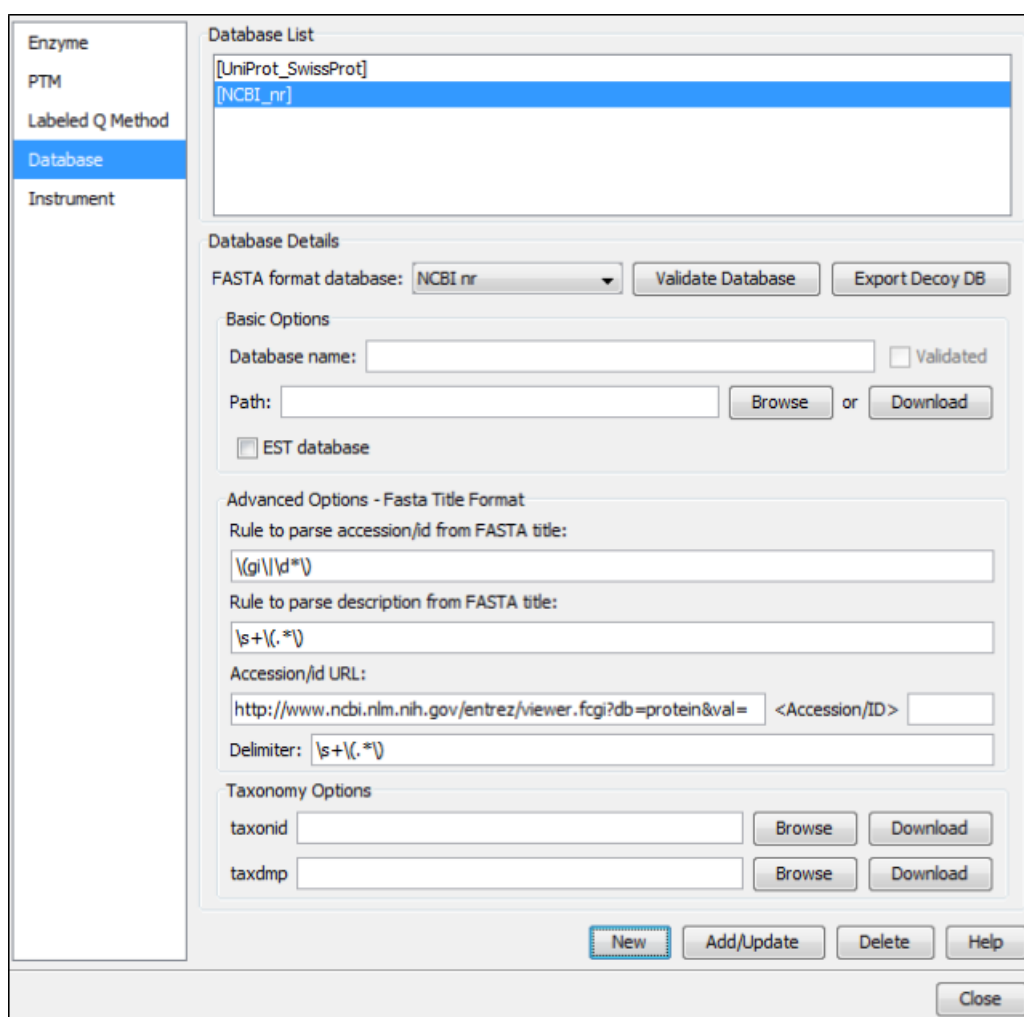


# Chapter 6. Adding a Sequence Database

## 1. Configuring Sequence Databases

PEAKS has the ability to search through a database to identify proteins using PEAKS DB to identify proteins of interest from a target database, or to filter out identifications from a contaminant database. PEAKS must have access to either a protein or an EST database in FASTA format to do so. PEAKS can be configured to use a publicly available sequence database or a custom database. Additionally, taxonomy options may be specified for certain databases so that users can focus their search using a specific species or group of species.

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



Follow these steps to configure a database:

1. Select the database format from the “FASTA Format Database” drop-down menu, or select “Other” if the desired format is not present. If “Other” is selected, you must enter custom parse rules. If you don't know what parse rules to use, the MSDB parse rules generally work with most databases.

2. In the “Basic Options” section, enter a name for the database. If the database FASTA file is already on the local system, skip to step 6. Otherwise, click Download. A window will appear confirming the database chosen to be downloaded along with the corresponding FTP or website.
3. Click OK to invoke the default FTP client software and download the database automatically. Click Cancel to copy the URL to the system clipboard. If Cancel was selected, click OK on the dialogue detailing the copy to the clipboard. Next, open a browser and paste the URL into the address bar. When the file download window opens, click Save.
4. Once the database has been downloaded, check to see if it is compressed. If so, extract the file using a program such as WinZip, or 7zip. The decompressed file should be a FASTA format text file (.fas or .fasta file).
5. Move the database file into a directory that PEAKS can access.
6. Click Browse in the “Basic Options” to inform PEAKS of the location of the database file.
7. If the selected database is an EST database, check the box labeled “EST database”. Otherwise, leave this unchecked.
8. To make sure PEAKS can read the database correctly, click the Validate Database button. The program will check for duplicates, extremely long sequences, and other irregularities. The number of proteins in the database will be listed along with example accessions, descriptions, and any errors encountered.
9. 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 text boxes in the “Advanced Options- Fasta Title Format” section below.  
  
If “Other” was selected in Step 2, enter the parsing parameters into the corresponding text boxes. Alternatively, if the database format is similar to one of the public databases such as *NCBI-nr*, the parsing rules can be filled by selecting a similar database from the drop-down list and edited to set the desired parsing rules.
10. 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. One deviation from Java Regular Expressions is that /( indicates the beginning of the text to be included and /) indicates the end of the text to be included.

To run PEAKS DB using a specific taxonomy, the corresponding taxon files must be downloaded and then directed to within PEAKS in the “Taxonomy Options” section. (Taxonomy files for *NCBI nr* database are: **gi\_taxid\_prot.dmp.gz** and **taxdmp.zip**; for *UniProt/Swiss-Prot*, they are **speclist.txt** and **taxdmp.zip**.)

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

## Note

Please do not rename the taxonomy files; otherwise PEAKS cannot recognize them. Also compressed files (for example, in .gz and .zip format) should be kept in their downloaded format.

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

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

- **X!Tandem** - At the time of writing, X!Tandem has difficulty searching through large databases and may crash during a search. Therefore, it's suggested that X!Tandem only be used with small databases. If used with a large database, a taxon should be specified. When using *NCBI* or *SwissProt* databases with X!Tandem it is best to use a subtaxon.
- **OMSSA** - At the time of writing, OMSSA cannot be used with databases that are not in NCBI or SwissProt format in a way that is available to inChorus. Also, a bug in OMSSA prevents easy use of databases with OMSSA when they are stored in a folder that contains spaces or non-English characters in its path. This creates problems when PEAKS creates temporary databases on your behalf. To avoid this, best practices suggest that all databases are put in a folder named `C:/peaksdatabases`. Note that the folder `C:/My Documents/databases` wouldn't work as it contains a space between 'My' and 'Documents'. Using spaces in the database file name causes the same problem. Once databases have been downloaded and extracted, save the database file as `ncbinr.fas`, or `ncbi_nr.fas` rather than `ncbi nr.fas`.
- **Mascot** - The database used by Mascot has to be identical to the database configured in PEAKS in order for inChorus to parse Mascot results correctly.

---


# Chapter 7. Data Refinement

## 1. Overview

Data refinement optimizes spectra for *de novo* peptide sequencing and database search algorithms resulting in improved identification and quantification results. All or some of the following functions can be applied to the data in a project according to the user's requirements.

- **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.
- **Feature association for chimera scans:** In some cases, multiple peptides may elute from the column at the same time so that they appear to be overlapped in the MS1 scan. As a result, the MS/MS scan may acquire multiple ionized peptides in the acquisition window. To solve this issue, the peptide features associated with the chimera scan will be differentiated and the MS/MS spectrum will store each feature's precursor mass for following *de novo* sequencing and identification steps.
- **Low quality peptide feature removal:** This function attempts to remove any false positive peptide features. Use this function with caution as it may also remove a small portion of identifiable features.
- **Background Refinement Processes:**
  - i. **Centroiding:** If the MS/MS scans are read in profile mode, each fragment ion area will be integrated and summed to the highest intensity in the distribution.
  - ii. **Charge Deconvolution:** The charge of fragment ions are predicted and fragment ions greater than charge +1 are relocated to their expected charge +1 m/z.
  - iii. **Isotope Deconvolution:** The isotopic distribution of fragment ions are determined and relocated to the monoisotopic mass of the fragment ion.
  - iv. **Peptide Feature Detection:** Signal from the LC-MS data is searched for features that resemble the expected retention time and m/z pattern of an ionized peptide. Peptide features that follow a Gaussian distribution across the retention time dimension and an expected carbon isotope distribution across the m/z dimension will be stored for quantification.

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 initiated. You can run the data refinement function separately by selecting a fraction, a sample, or a project on the project tree. All fraction(s) under the selected node will be refined. The use of this function is outlined in the following:

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

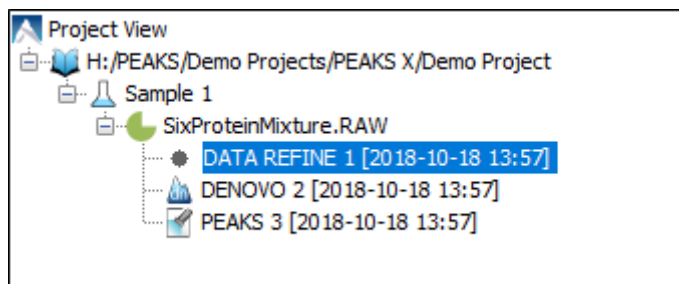


2. Specify the data refinement parameters in the pop-up dialogue and click OK. Most of the parameters are self-explanatory and the default parameters provide a good starting point for the analysis.


## Note

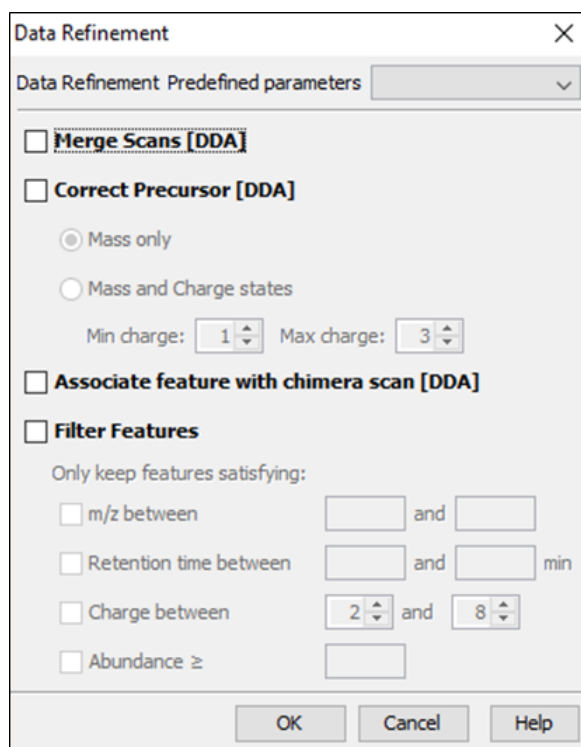
Default parameters are preconfigured in the Data Refinement parameters window. To use default parameters, select "Default" from the drop-down in the upper right-hand corner of the window,

- Wait for the analysis to complete. A new Data Refinement node will appear in the Project View pane. Additional downstream analyses using this fraction will be based on the refined data.



## 2. Data Refinement Parameters

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



- **Merge Scans (DDA)**
- **Correct Precursor (DDA)**

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

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

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

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

- Associate Feature with Chimera Scan (DDA): PEAKS will consider a tandem scan can contain two peptides and when selected, PEAKS will search for chimera spectra to enable correct inference of co-eluted isobaric peptides.
- **Filter Features**

**Precursor m/z Range:** The mass-to-charge range a feature must appear within to perform additional analysis.

**Retention Time Range:** The retention time region, in minutes, for features to be selected for further analysis.

**Charge Range:** The range of peptide charge a feature must exhibit to be selected for further analysis.

**Abundance Threshold:** The minimum intensity threshold a feature must have to be selected for further analysis.

## Note

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

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

## 2.1. Saving the Parameters for Future Use

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



---

# Chapter 8. Peptide *De Novo* Sequencing

## 1. Overview

*De novo* sequencing derives a peptide sequence directly 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 (Ex. Orbitrap/FTMS/Ion Trap/Time-of-Flight/Quadrupole) in common fragmentation modes (Ex. CID/HCD/ETD). PEAKS can also perform *de novo* sequencing on complementary CID/HCD/ETD spectral pairs or triplets, and mixed fragmentation methods (Ex. EThcD).

- **Automated result validation**

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

### Note

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

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

### Note

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

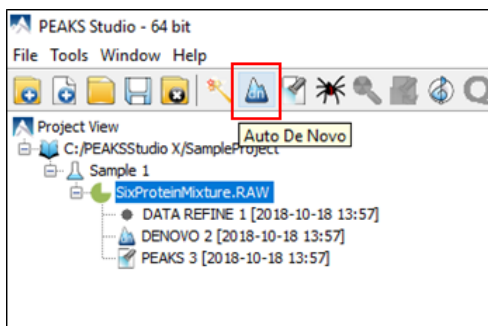
Other PEAKS tools, including PEAKS DB for database search; PEAKS PTM for unspecified PTM search; and SPIDER for homology search depend on the *de novo* sequencing result to achieve superior sensitivity and accuracy.

### Note

Refer to Section 1, "Overview"

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

1. Select a fraction, a sample, or a project node in the "Project View" pane. Depending on what is selected, PEAKS will analyze all fractions contained within selected node in the project tree. Launch PEAKS *de novo* sequencing by clicking the "Auto De Novo" button in the tool bar.



## Note

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

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

## Note

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

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

	Feature ID	Peptide
1	66798	DYLEKYYKF
2	53812	KFLDTTSKF
3	67581	KYFLWEEKF
4	44447	SPNAELHLL
5	67816	VYM(+15.99)DWYEFK
6	60990	KLTYLYLQK
7	49986	YSSEALKSY
8	61476	FYM(+15.99)DTSHLF
9	66280	KYQEEFEHF
10	67816	VYM(+15.99)DWYEFK
11	54888	VYKVPPFTF

## 2. De Novo Sequencing Parameters

After selecting a fraction, a sample, or a project node in the "Project View" pane, click the PEAKS *de novo* icon



from the tool bar. The PEAKS *de novo* parameters dialogue will appear:

## 2.1. Error Tolerance

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

**Fragment ion m/z:** 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 specified error 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 chosen by selecting "Default" from the "Predefined parameters" drop-down menu in the upper right-hand corner of the dialogue.

### Note

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

## 2.2. Enzyme

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

### Note

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

### Note

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

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

## 2.3. PTM

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

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-t...	42.0106	[X]@N
Amidation	-0.9840	[X]@C
Ammonia-loss (C@N-te...	-17.0265	[C]@N
Ammonia-loss (Protein ...	-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(KR)	28.0313	[KR]
Dimethylation(NP)	28.0313	[N], [P]@N
Dioxidation (M)	31.9898	[M]
FAD	783.1415	[CHY]
Farnesylation	204.1878	[C]
Formylation	27.9949	[K], [X]@N
Formylation (Protein N-...	27.9949	[X]@N
Geranyl-geranyl	272.2504	[C]
Guanidination	42.0218	[K]

Recent Common Uncommon Glycosylation Artificial Customized

New

Remove Remove All Switch Type

OK Cancel

**Selected Fixed PTM**

☐ Carbamidomethylation

**Selected Variable PTM**

☒ Deamidation (NQ)

☒ Oxidation (M)

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

- **Common**

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

- **Uncommon**

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

- **Glycosylation**

The "Glycosylation" list contains commonly observed glycosylation.

- **Artificial**

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

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

- **Recent**

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

- **Customized**

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

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

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

## Note

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

## Note

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

## Note

Leucine (*L*) and Isoleucine (*I*) cannot be directly distinguished in traditional mass spectrometry since the two residues are isobaric, meaning they 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	<input type="text"/> N-term
	<input type="text"/> C-term
Formula:	H(3) C(2) N(1) O(1)
Rule:	<input type="text"/>
<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” pop-up dialogue. The same information can be viewed by selecting a PTM from a list and clicking the View button. If a desired modification is not listed or is different than the listed modification, it can be defined as a customized modification. Click the New button to show the “New PTM” dialogue and enter the information of the particular modification. The newly defined modification will be added to the “Customized” list. Refer to Section 2.2, “PTM Configuration” for details on creating new PTMs and viewing PTM information.

## 2.4. Other Parameters

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

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

## 2.5. Saving the Parameters for Future Use

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

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

### 3. Understanding PEAKS *De Novo* Sequencing Results

When *de novo* sequencing is done, a new result node labeled DENOVO will appear in the project tree in the "Project View" pane. Double-click the node to load the analysis report. The analysis report is presented in three pages:

- **Summary:** The outline of PEAKS *de novo* sequencing result, which includes key statistics. It also provides functions to filter and export *de novo* sequences.
- **De novo:** *De novo* peptide sequences listed in a table. Spectrum annotation and other detailed information are provided.
- **Feature:** All features listed in a table. Spectrum annotation and other detailed information are provided.
- **LC/MS:** Interactive LC-MS heat map to visualize *de novo* sequenced peptides.

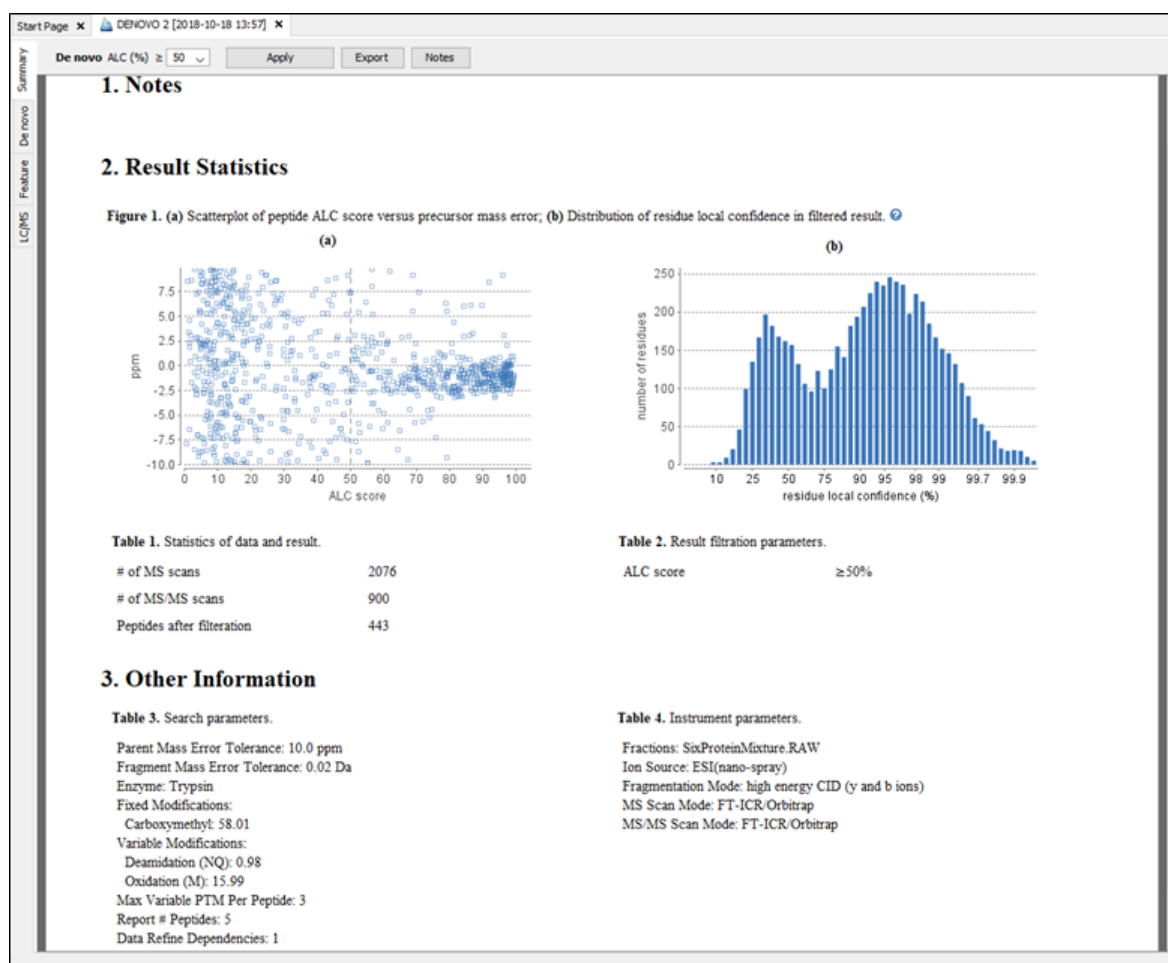
#### 3.1. Summary View

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

1. **Result filtering:** Filtering rules can be set up at the top of the Summary View. Low quality *de novo* sequences can be filtered out by specifying a minimal threshold of the ALC score. More detail about how to properly set the ALC threshold is discussed in Section 4, "Filtering *De Novo* Sequencing Results".
2. **Result exporting:** *De novo* sequencing results can be exported into various formats by clicking the Export button at the top of the Summary View. More detail is discussed in Section 5, "Exporting *De Novo* Results".

The report is composed of three sections:

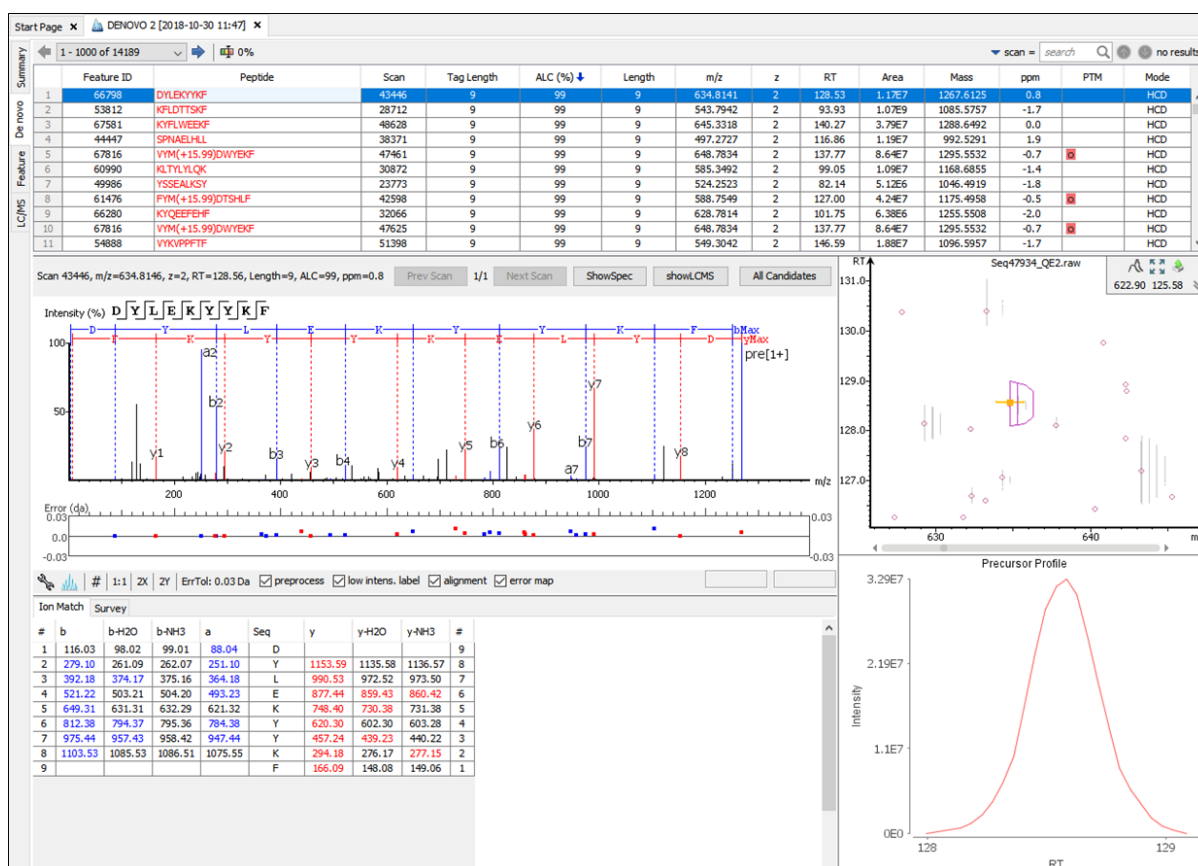
1. **Notes:** The header of the report. Experiment description or other information can be added to this section. The notes can be edited by clicking the Notes button.
2. **Result Statistics:** Key statistics shown in figures and tables.
3. **Other Information:** Search parameters and MS instrument information.



Statistical figures in this report are discussed in a later section Section 4, “Filtering *De Novo* Sequencing Results”.

## 3.2. *De Novo* Peptide View

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



## Note

Please refer to Chapter 16, *Data Independent Acquisition (DIA) Data Analysis* and Chapter 17, *Ion Mobility Data Analysis* for *de novo* peptide view details for DIA and IMS data.

### 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:


- **Feature:** An arbitrary number provided by PEAKS during data refinement used as a unique identifier.
- **Peptide:** The amino acid sequence of the peptide as determined by *de novo* sequencing. A modified residue is followed by a pair of parentheses enclosing the mass of that modification.
- **Scan:** The scan number. The scan number is a unique index for tandem mass spectra in the data.
- **Tag Length:** The length of the longest sequence tag in the peptide. When low confidence residues are reduced to mass tags, the remaining residues become sequence tags. The length of the longest tag is reported in this column. Refer to Mass Tags.
- **ALC (%):** The average local confidence. ALC is calculated as the total of the residue local confidence scores in the peptide divided by the peptide length.
- **Length:** The number of amino acids in the *de novo* sequence.
- **m/z:** The precursor mass-to-charge ratio.
- **z:** The precursor charge.

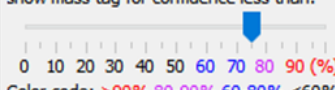


- **RT:** The retention time (elution time) for the spectrum as recorded in the data.
- **Area:** The area under the curve of the peptide feature found at the same m/z and retention time as the MS/MS scan. This can be used as an indicator of the abundance of the peptide.
- **Mass:** The monoisotopic mass of the peptide.
- **ppm:** The precursor mass error, calculated as  $10^6 \times (\text{precursor mass} - \text{peptide mass}) / \text{peptide mass}$ .
- **PTM:** The types and the numbers of the modifications present in the peptide shown as color-coded icons.
- **Mode:** The fragmentation mode in which *de novo* sequencing is performed by the algorithm.

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

	Feature ID	Peptide
990	55545	VQNGKHFKF
991	9484	AGSHLLQRM(+15.99)
992	43772	NATDLLKVL
993	56756	VYVYGDRF
994	67828	LYQDQ N A T D L L K V L
995	40664	QYGV 73 88 99 99 99 99 99 99 (%)
996	33433	LPL

**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.

1 - 1000 of 14189		75%
Feature ID	Peptide	show mass tag for confidence less than:
990	55545 VQNGKHFKF	 0 10 20 30 40 50 60 70 80 90 (%) Color code: >90% 80-90% 60-80% <60%
991	9484 AGSHLLQRM(+15.99)	
992	43772 [114.0]ATDLL	
993	56756 VYVYGDRF	

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

## Note

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

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

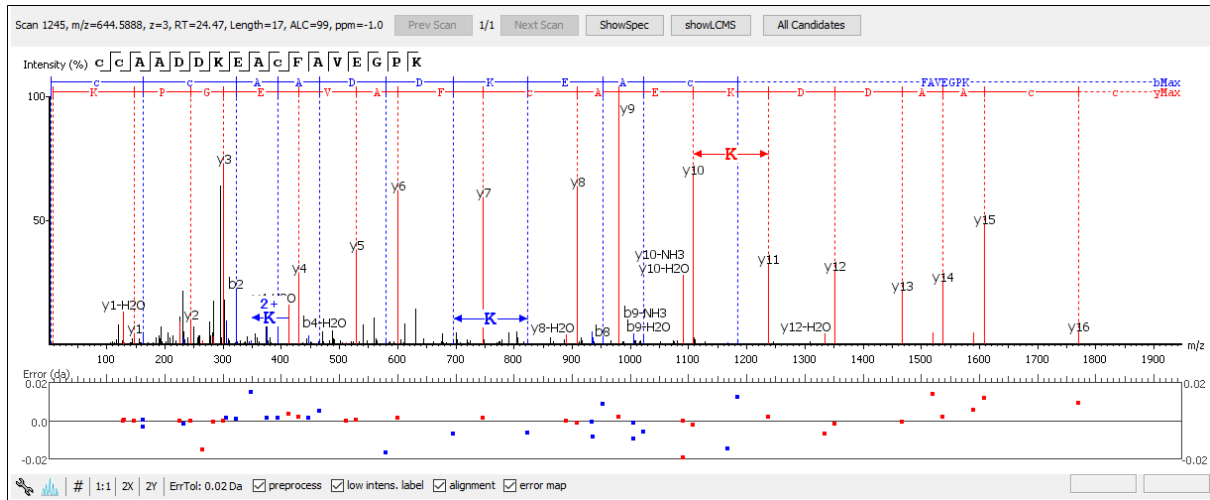
scan =	search	no results
scan =	ppm	...
seq contains	-3.6	
m/z ≈	-3.7	
RT ≈	-2.7	
PTM Δm ≈	-3.8	
	-1.5	
	-1.8	

## Note

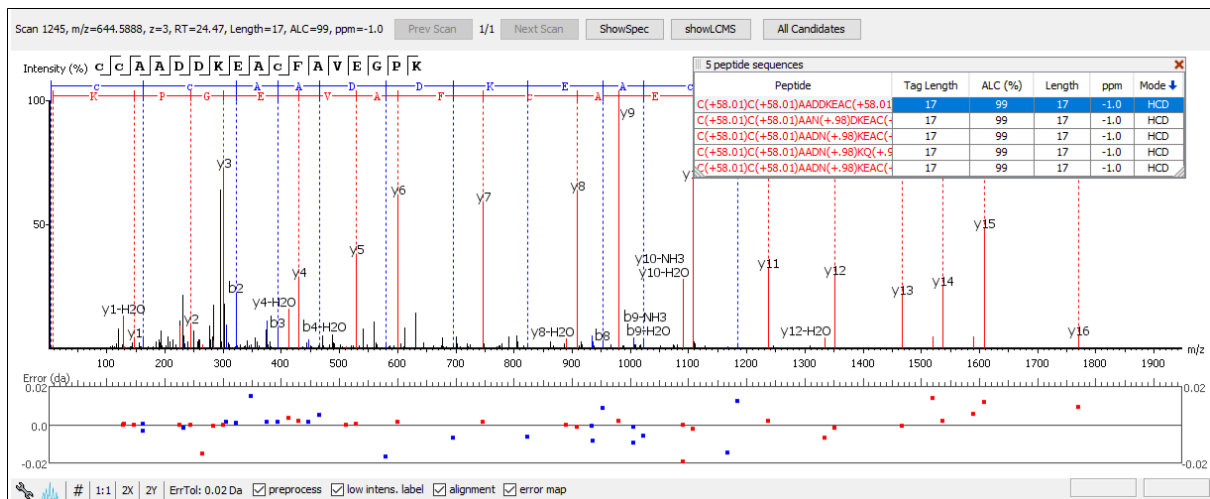
Precursor  $m/z$ , retention time, and modification mass are searched approximately. For example, 130.3 will match any value from 130.25 to 130.35 (exclusive). And 130 will match from 129.5 to 130.5 (exclusive).

### 3.2.2. Spectrum Annotation

The spectrum annotation displays a graphical representation of the peptide-spectrum match. Moving the cursor over the peptide sequence in the spectrum will show the mass transitions for a particular amino acid residue.



The title bar shows the information of the peptide sequence that is being displayed in the spectrum. Press the "Show Spec" button to open the raw MS/MS scan in the Data Refinement MS/MS View. Press the "Show LCMS" button to automatically view the location of the MS/MS scan and isolation window in the LC-MS heat map for the current peptide under analysis. Press the "All candidates" button to open a pop-up window that displays all alternative peptides. Click on a peptide sequence in the All Candidates pop-up window to select and display the annotation for that peptide spectrum.

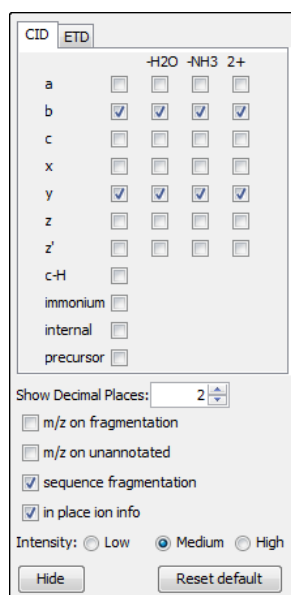




The spectrum annotation panel provides convenient ways to zoom and navigate within the spectrum:


- **Zoom to a  $m/z$  region:** Drag horizontally from the start  $m/z$  and to the end  $m/z$  with left mouse button.
- **Zoom in/out smoothly:** Place the cursor at a particular  $m/z$  value (right below the x-axis), zoom in/out by scrolling the mouse wheel.
- **Increase/Decrease peak intensity:** Place the cursor in the spectrum and scroll the mouse wheel.

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

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



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

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

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

The pre-process check box allows for transition between the spectrum after centroiding and deconvolution and the original spectrum.



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

The "alignment" check box allows user to display/hide 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.

The "error map" check box 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").

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

## Note

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

### 3.2.3. Ion Table

The "Ion Match" tab in the bottom pane contains a table with possible fragment ions for the selected MS/MS scan. Each ion in the table shows the calculated mass, where if the fragment ion is present in the spectrum, its mass

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

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

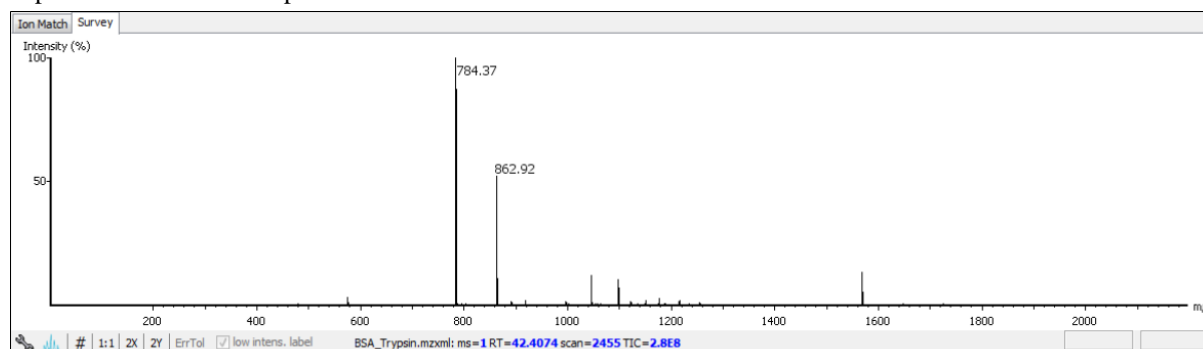
To view the fragment ions from the table highlighted in the spectrum annotation view and/or error map, select the column heading of the ion type of interest.

## Note




Uncheck the alignment checkbox in the spectrum annotation view to see the highlighted fragment ions more easily.

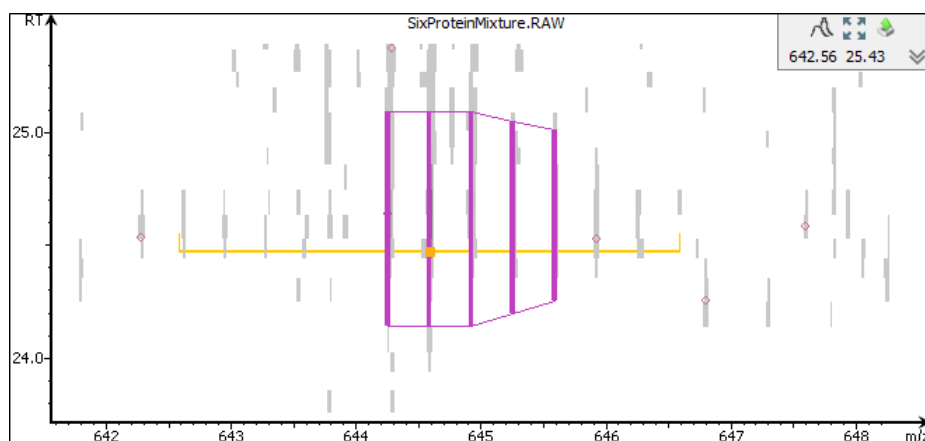
### 3.2.4. Parent Scan

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.



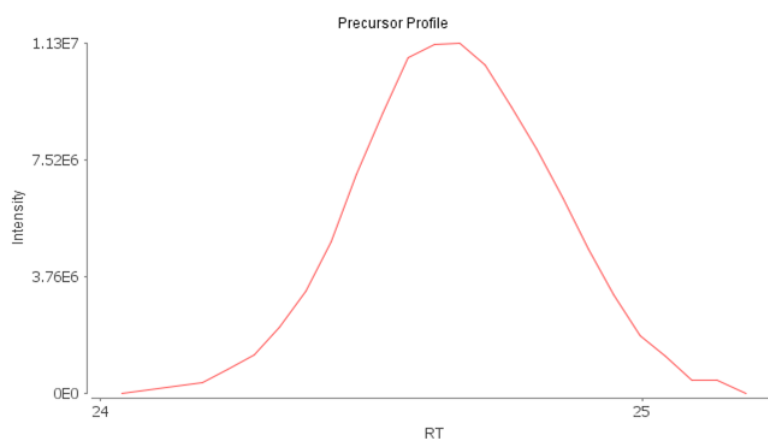
### 3.2.5. LC-MS Snapshot

The LC-MS Snapshot displays a zoomed-in view of the precursor scan in the LC-MS heat map 2-D view. The peptide feature is highlighted in purple and the *de novo* sequenced MS/MS scan is indicated by a yellow square. The span of the yellow line in the LC-MS Snapshot designates the isolation window. Users can change the view from the 2-D heat map to a 3-D intensity view by clicking the intensity view button, , in the upper right-hand corner of the LC-MS Snapshot, and enlarge the LC-MS view to cover the full screen by clicking the expand button, , in the upper right-hand corner of the LC-MS Snapshot. In addition, users can export the LC-MS heat map by clicking the export image button, , in the upper right-hand corner of the LC-MS Snapshot. See Chapter 19, Section 4.2.2. Section 4.2.2, "Exporting LC-MS, IM-MS, and LC-IM-MS".



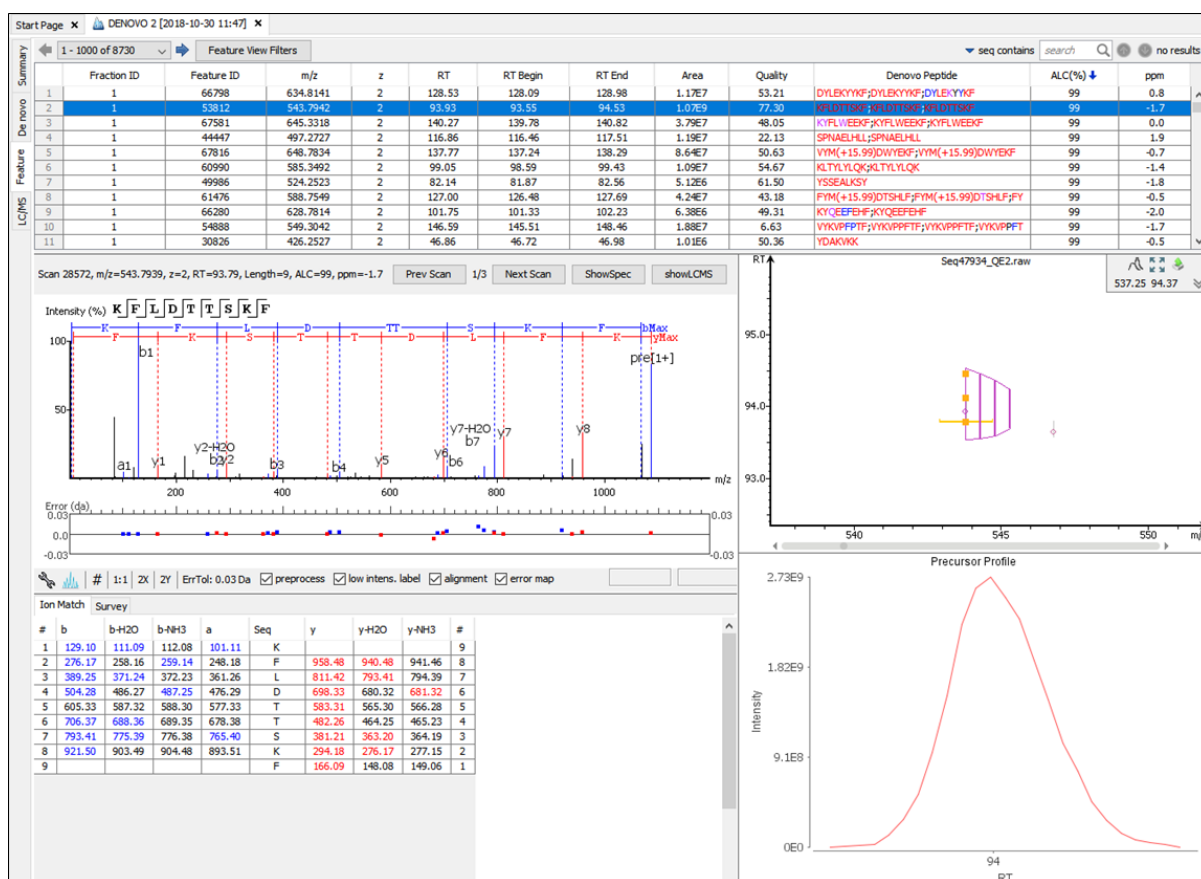
### 3.2.6. Precursor Profile

The Precursor Profile, or eXtracted Ion Chromatogram (XIC) of the precursor ion, graphs the elution profile of the selected peptide feature vector over the retention time range where it was identified. The graph is plotted with the retention time along the x-axis and the intensity along the y-axis. Hovering over the XIC curve will display the RT and intensity coordinates.



### 3.3. Feature View

The "Feature View" displays all the feature information in greater detail, as shown in the next figure. The table on the top displays features, while the bottom section provides additional information about the feature, precursor profile, and if available peptide-spectrum match.



## Note

Please refer to Chapter 16, *Data Independent Acquisition (DIA) Data Analysis* and Chapter 17, *Ion Mobility Data Analysis* for feature view details for DIA and IMS data.

### 3.3.1. Feature Table

The feature table shows the identified feature result. For each feature, the feature information is displayed and if available the *de novo* peptide sequence candidate. The result can be sorted by any column. For example, to sort the identified feature by the quality scorer, just click on the title of the “Quality” column.

The following list describes the contents in each column:

- **Fraction ID:** The scan number. The scan number is a unique index for tandem mass spectra in the data.
- **Feature ID:** An arbitrary number provided by PEAKS during data refinement for each identifiable feature.
- **m/z:** The precursor mass-to-charge ratio.
- **z:** The precursor charge
- **RT:** The retention time (elution time) for the spectrum as recorded in the data.
- **RT Begin:** The beginning retention time (elution time) for the feature as recorded in the data.
- **RT End:** The ending retention time (elution time) for the feature as recorded in the data.
- **Area:** The area under the curve of the peptide feature found at the same m/z and retention time as the MS/MS scan. This can be used as an indicator of the abundance of the peptide.
- **Quality:** The quality score for the peptide feature. Factors that affect the quality score include m/z difference, RT difference, XIC shape similarity, and the feature intensities.

- **Denovo Peptide:** The amino acid sequence of the peptide associated with the feature, as determined by *de novo* sequencing. A modified residue is followed by a pair of parentheses enclosing the mass of that modification. If multiple MS2 are associated with the feature, all *de novo* peptide results are listed.
- **ALC (%):** The maximum average local confidence of the *de novo* peptides. ALC is calculated as the total of the residue local confidence scores in the peptide divided by the peptide length.
- **ppm:** the mass error in ppm of the *de novo* peptide with the highest ALC score

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

## Note

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

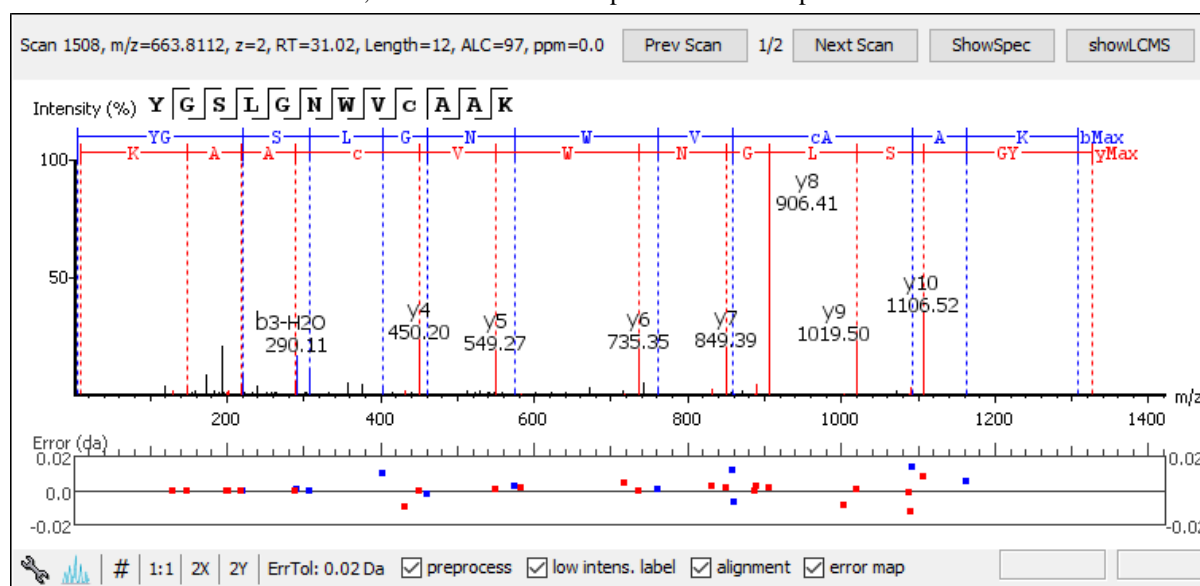
**Feature Search.** The feature table can be searched by entering the value in the search bar located in the top-right corner. The features can be searched by sequence, and precursor m/z. The matched features are highlighted in the table, and can be iterated by clicking the up and down buttons in the search bar.

## Note

Precursor m/z are searched approximately. For example, 430.3 will match any value from 430.25 to 430.35 (exclusive). And 430 will match from 429.5 to 430.5 (exclusive).

### 3.3.2. Spectrum Annotation

The spectrum annotation displays a graphical representation of the peptide feature spectrum match. If viewing features without an identification, there will be no PSM present to show Spectrum Annotation.



The title bar shows the information of the peptide sequence that is being displayed in the spectrum. If one feature contains two tandem scans, press the “Next Scan” button to view the Spectrum Annotation of the second PSM and “Prev Scan” to return to the first PSM. Based on the PSM selected, the LC-MS Snapshot will change dynamically to highlight PSM in the LC-MS view. Press the “Show Spec” button to open the MS/MS Scan in the Data Refinement MS/MS View. Press the “Show LCMS” button to automatically view the location of the MS/MS scan and isolation window in the LC-MS heatmap for the current peptide under analysis.

Spectrum annotation tool options are the same as described above in Section 3.2.2, “Spectrum Annotation”.

### 3.3.3. Ion Table

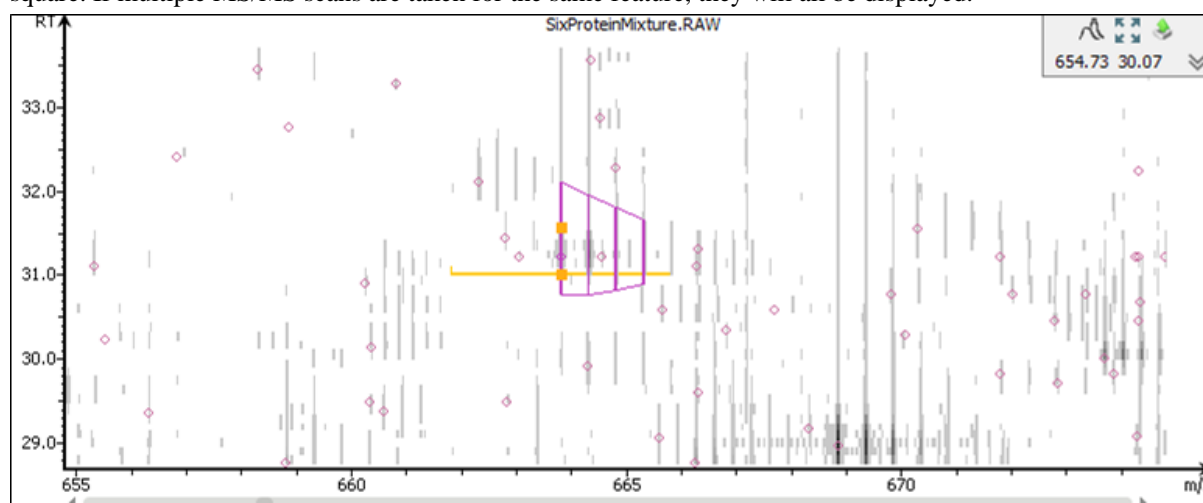
The “Ion Match” tab at the bottom panel of the feature view contains the “Ion Table” that shows the calculated mass of possible fragment ions. For more details, see Section 3.2.2, “Spectrum Annotation”.

### 3.3.4. Parent Scan

The “Survey” tab displays the precursor ion spectrum. The buttons that appear in this section are identical to those explained above in Section 3.2.2, “Spectrum Annotation”.

### 3.3.5. LC-MS Snapshot

LC-MS Snapshot displays a zoomed-in snapshot of the peptide precursor spectrum in the LC-MS heat map 2-D view. The peptide feature is highlighted in purple and the associated MS/MS scan(s) is indicated by a yellow square. If multiple MS/MS scans are taken for the same feature, they will all be displayed.



LC-MS Snapshot tool options are the same as described above in Section 3.2.2, “Spectrum Annotation”.

### 3.3.6. Precursor Profile

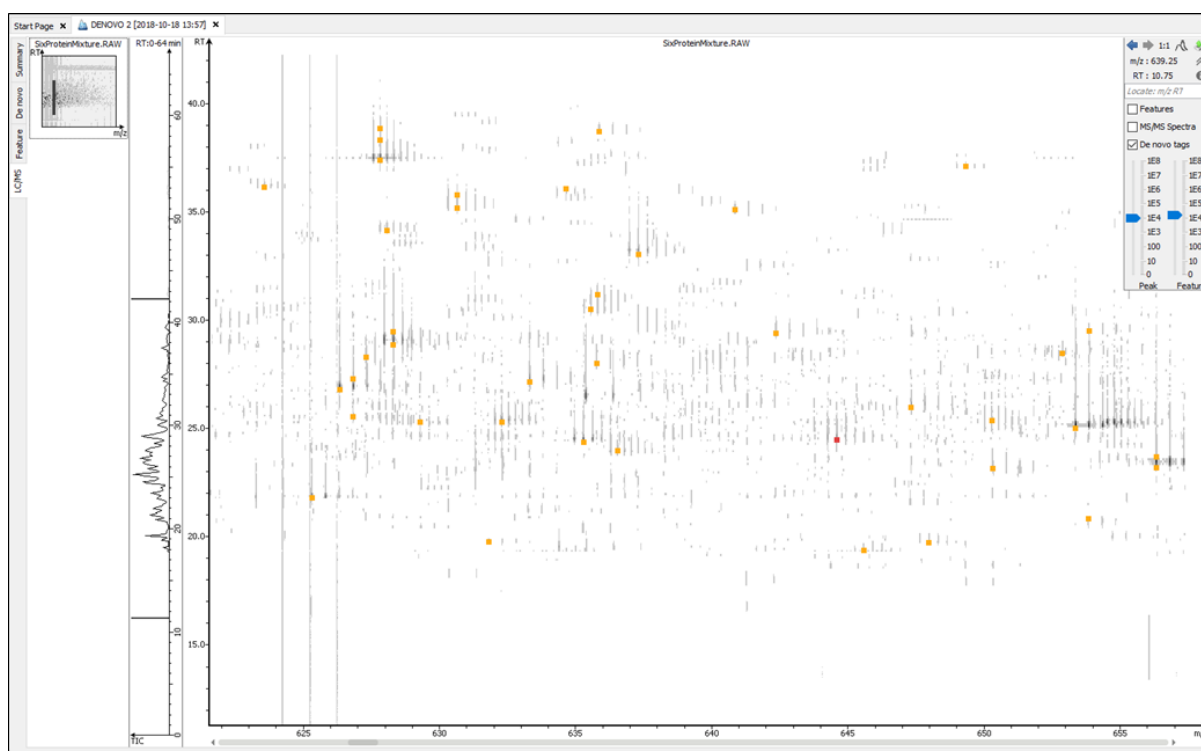
The Precursor Profile, or eXtracted Ion Chromatogram (XIC) graphs the elution profile of the selected peptide feature vector over the retention time range where it was identified. Refer to Section 3.2.2, “Spectrum Annotation”.

## 3.4. LC-MS View

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

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





## 4. Filtering *De Novo* Sequencing Results

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

De novo ALC (%)  $\geq$  50

Apply

Export

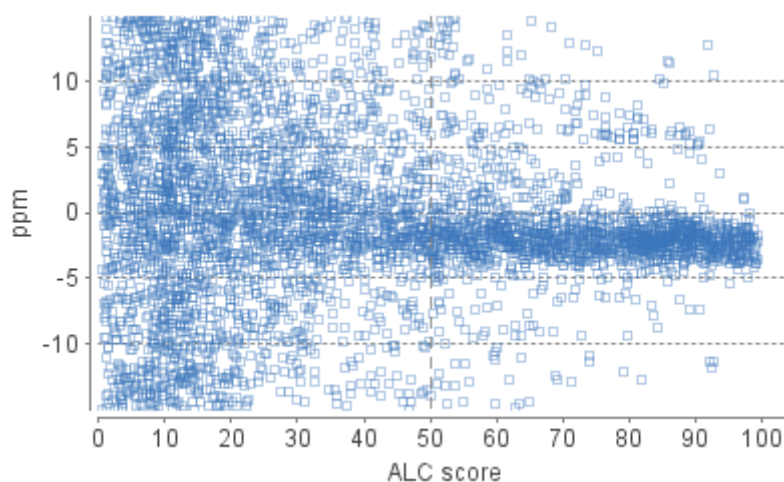
Notes

### Note

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

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

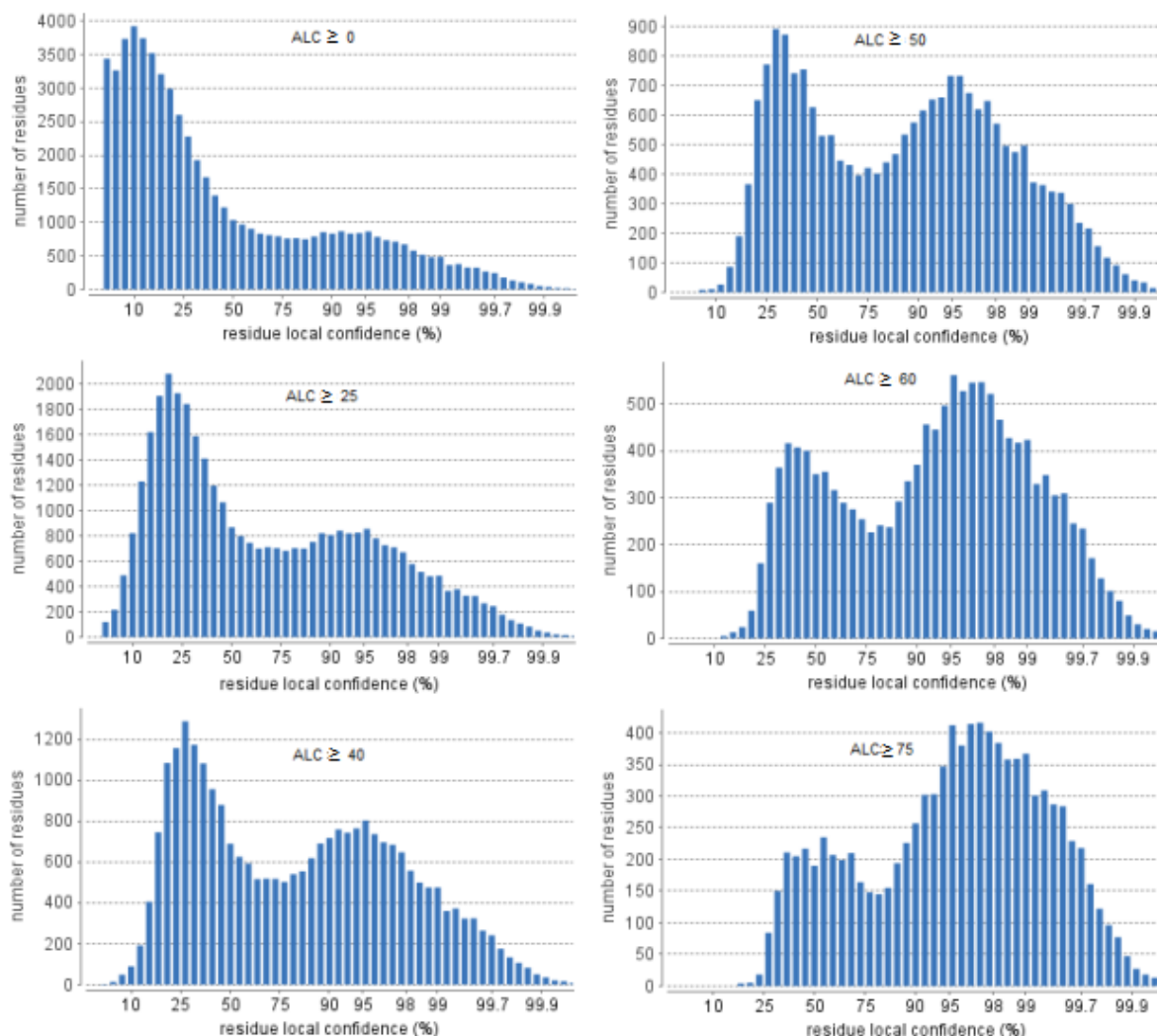
**Peptide ALC Score Scatter Plot.** Figure 1(a) is a scatter plot showing ALC score versus precursor mass error (in ppm) of the *de novo* sequences. In a typical *de novo* sequencing analysis, peptides with higher ALC scores are more likely to have a small mass error, indicating a higher chance that a peptide is correctly sequenced. This figure provides a rule-of-thumb for setting the ALC score threshold. The threshold can be set at the score where the precursor mass error starts to converge.



**Residue Local Confidence Distribution.** Figure 1(b) shows the local confidence score distribution of residues in filtered *de novo* sequences. It is essentially a mixed score distribution of both correct and incorrect residues. This figure allows you to examine the proportion of correct and incorrect residues when adjusting the ALC threshold. By increasing the ALC threshold, the number of incorrect residues will decrease because poor sequences are filtered out first. It is recommended to gradually raise the ALC threshold until the number of correct residues starts to significantly decrease.

## Note

In some cases, the local confidence score may not be able to distinguishably separate incorrect and correct residues as shown in the user manual.



## 5. Exporting *De Novo* Results

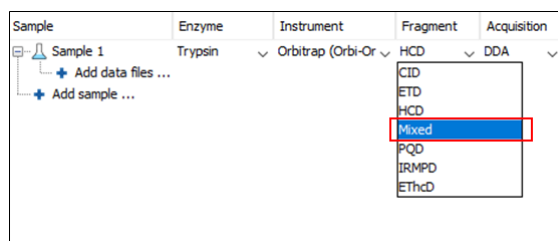
The Export button in the title bar of the "Summary" view allows exporting of the filtered results into various formats. It provides the opportunity to share the report in HTML and to also use the result as supplemental information in a publication. Please refer to Section 3, "Exporting *De Novo* Results"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

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

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

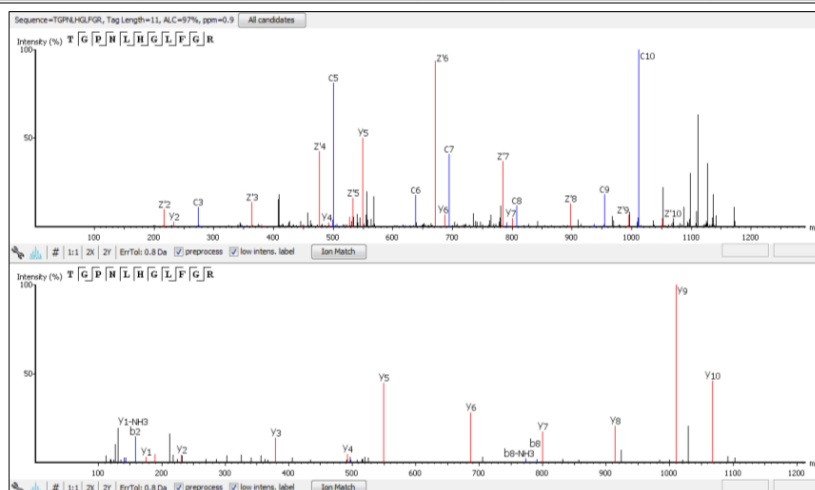
PEAKS automatically detects spectral pairs and triplets when doing *de novo* sequencing. The fragmentation type of data fractions containing spectral pairs or triplets has to be set to "Mixed" during the project creation. In "Mixed" mode, PEAKS loads the fragmentation type for each tandem spectrum from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

- The pair or triplet of spectra are under the same precursor scan and are fragmented from the same precursor ion.
- The pair or triplet of spectra have different fragmentation types. For example, HCD/ETD for spectral pairs, CID/HCD/ETD for spectral triplets.



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

Scan	Peptide	Tag Length	ALC (%)	Length	m/z	z	RT	Area	Mass	ppm	PTM	Mode
1	7949 7950	LUNNPYR	8	98	518.7825	2	107.00	3.46E5	1035.5500	0.3		ETD/HCD
2	4944 4945	VATVSPRL	8	98	421.7585	2	67.50	5.8E7	841.5021	0.3		ETD/HCD
3	4220 4221	LMVAHPEPR	9	97	372.1873	3	57.99	5.82E5	1113.5388	1.0		ETD/HCD
4	1252 1253	VTLTC(+57.02)K	6	97	361.1997	2	18.11	2.03E4	720.3840	1.1		ETD/HCD
5	8405 8406	TGPNLHG.FGR	11	97	390.2126	3	114.08	5.44E4	1167.6147	0.5		ETD/HCD
6	3379 3380	LASFLN(+98)K	7	97	405.2240	2	46.96	2.33E5	808.4330	0.6		ETD/HCD
7	6179 6180	FASFLN(+98)K	7	97	414.2186	2	83.73	4.07E4	826.4225	0.3		ETD/HCD
8	2028 2029	VTLTC(+57.02)K	6	96	375.2028	2	28.85		748.3901	1.1		ETD/HCD
9	6015 6016	YEQ(+98).LQ.TAGR	10	96	590.3042	2	81.59	2.63E5	1178.5930	0.7		ETD/HCD
10	1809 1810	VTLTC(+57.02)K	6	96	375.2024	2	25.85		748.3901	0.1		ETD/HCD
11	5805 5806	LQ(+98).SELR	8	96	487.2591	2	78.78		972.5240	-0.4		ETD/HCD

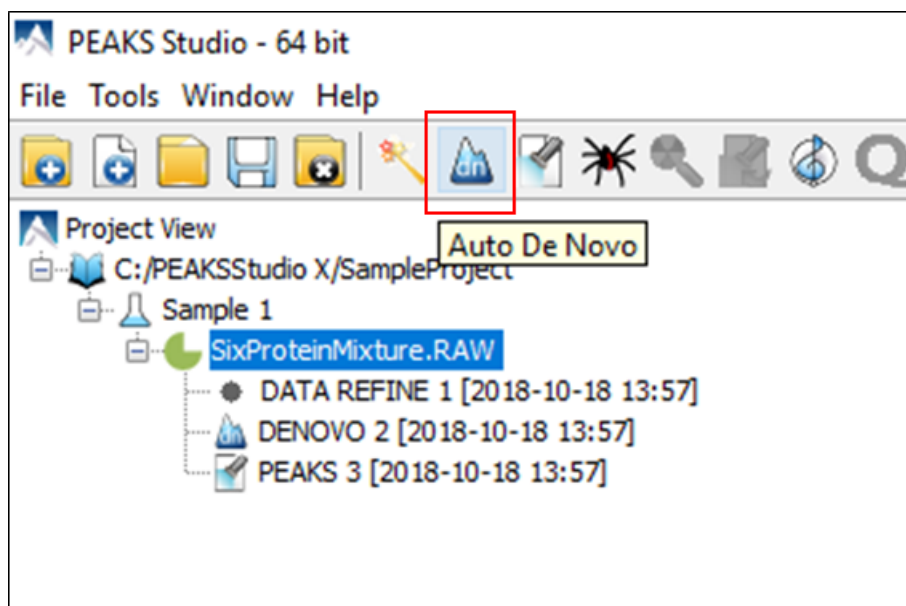


## 7. *De Novo* Sequencing on Single Spectrum

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

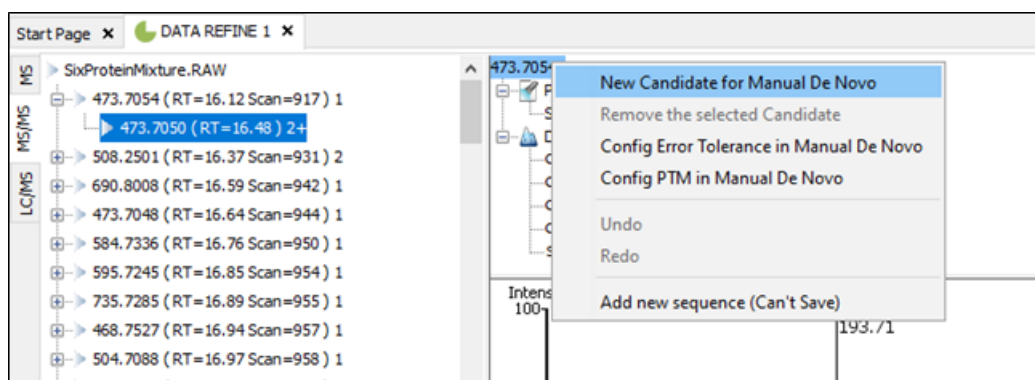
PEAKS provides a set of tools to help manually sequence a peptide from the spectrum.

### Note

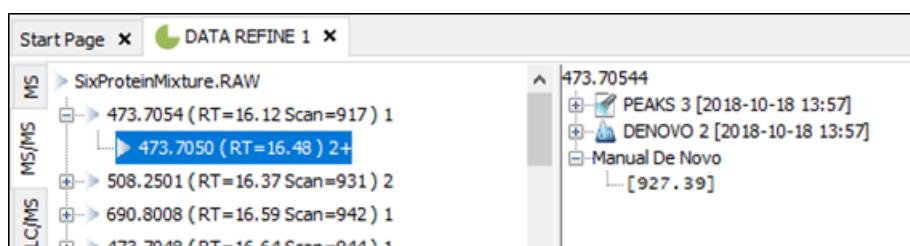
Manual *de novo* sequencing only supports b- and y- ions.

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

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



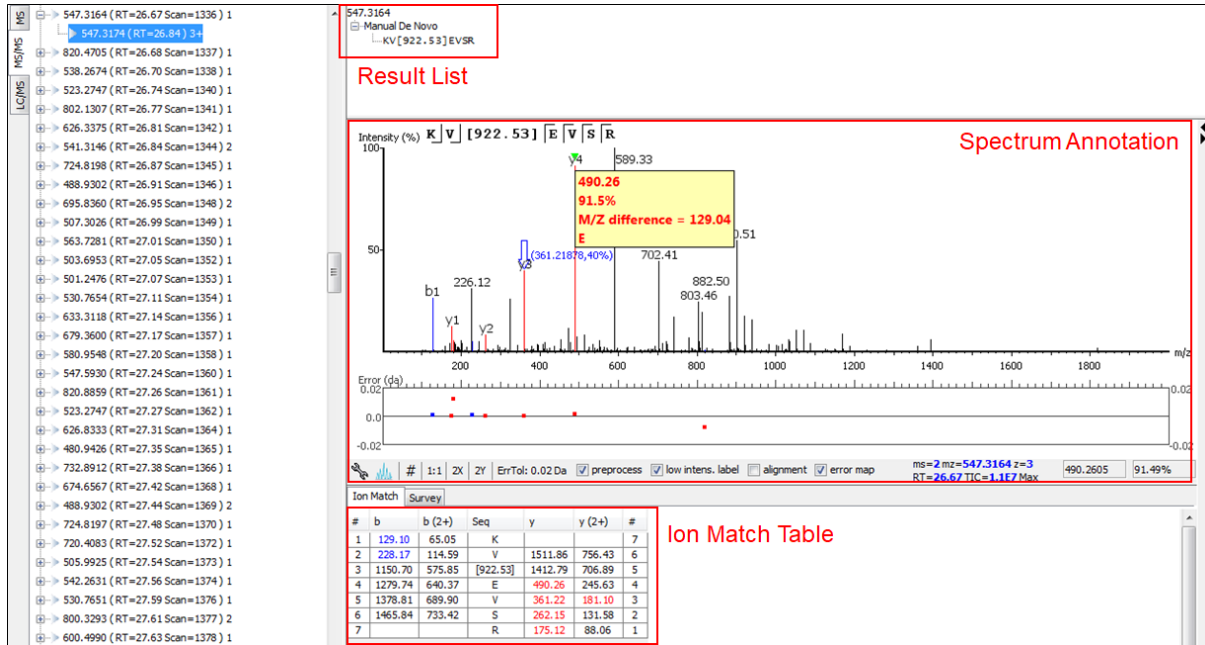
Select New Candidate for Manual De Novo from the pop-up menu. A new candidate will be created under the 'Manual De Novo' heading. The new candidate is initiated as unsequenced; so, it is represented by the total mass of the residues (as seen in the example below).



## Note

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

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

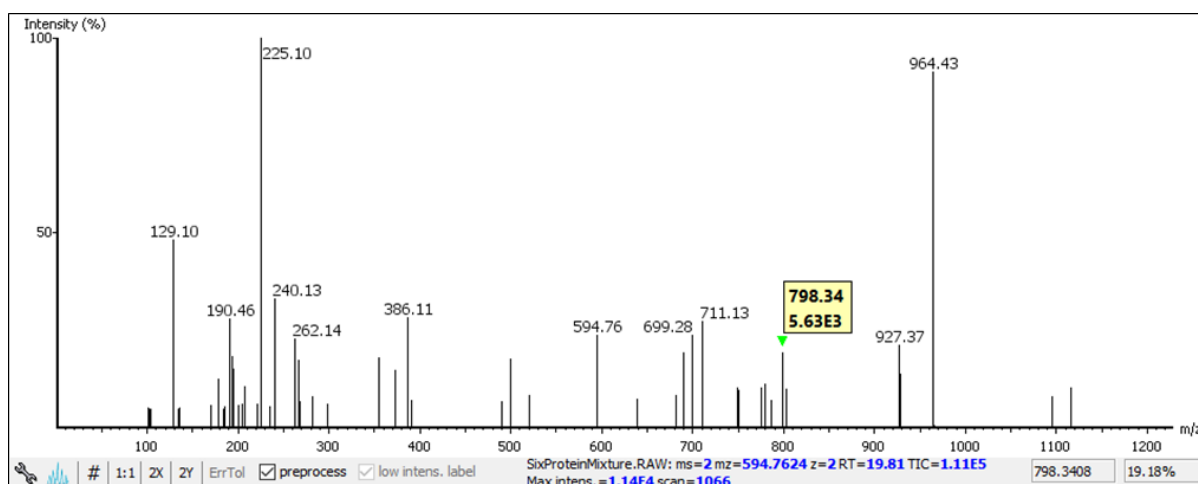


The panels are briefly described below:

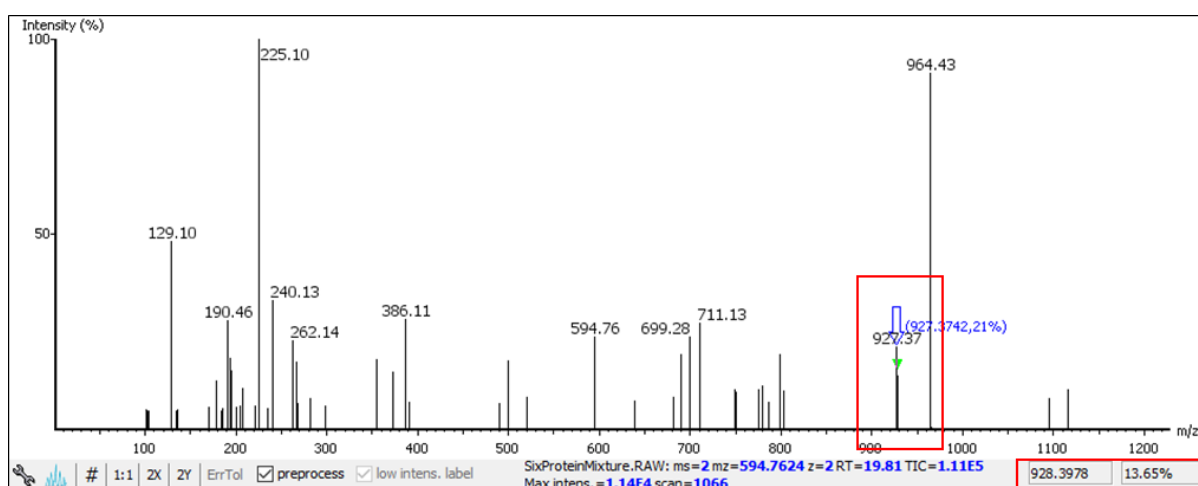
- **Result Panel:** The “Result Panel” shows all sequencing results of the selected peptide feature scan. The results of manual *de novo* are listed under “Manual *De Novo*”.
- **Spectrum Annotation Panel:** The “Spectrum Annotation” shows a graphical representation of the spectrum, the peaks in the spectrum, the user-selected peaks, and the assigned ions. Pick a peak on the panel with the cursor and assign ions or tags to it in manual *de novo*.
- **Ion Table Panel:** The “Ion Table” shows the proposed ions with their corresponding masses. The default Ion Table will display immonium, b, b-H<sub>2</sub>O, b-NH<sub>3</sub>, y, y-H<sub>2</sub>O, and y-NH<sub>3</sub> ions.

## 8.2. Manual *De Novo* Operations

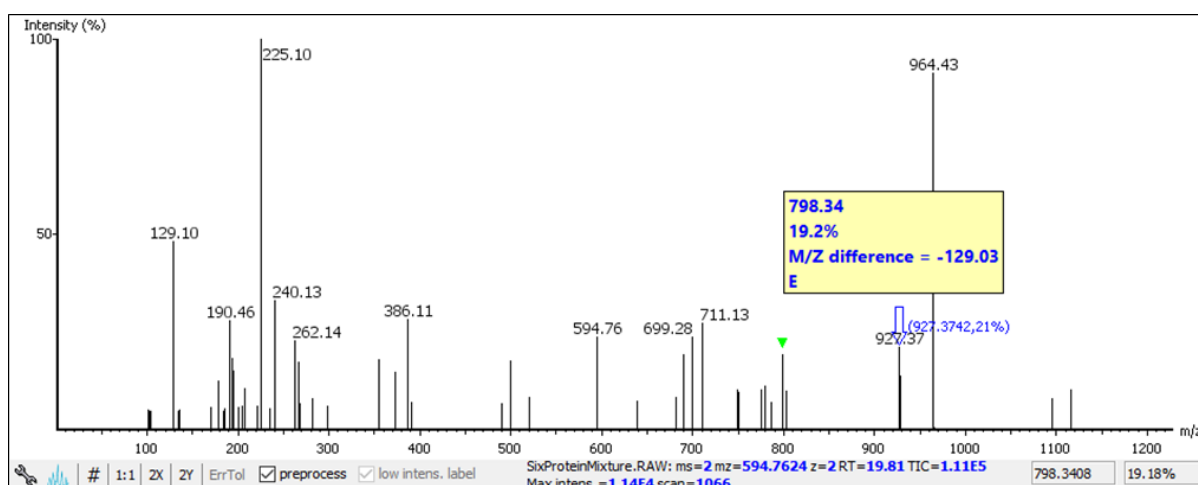
When the cursor is placed in the “Spectrum Annotation” panel, a green (by default) triangle follows the movement of the mouse. This is the *Position Bar* and it is used as a cursor for all manual *de novo* operations. The cursor's position on the m/z scale and its relative intensity are shown in a pop-up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



**Selecting a peak.** To select a peak, simply click on it. A blue (by default) arrow, called the *Freeze Bar*, indicates the selected peak. Alternatively, an ion peak can be selected by clicking on its corresponding cell in the Ion Table.



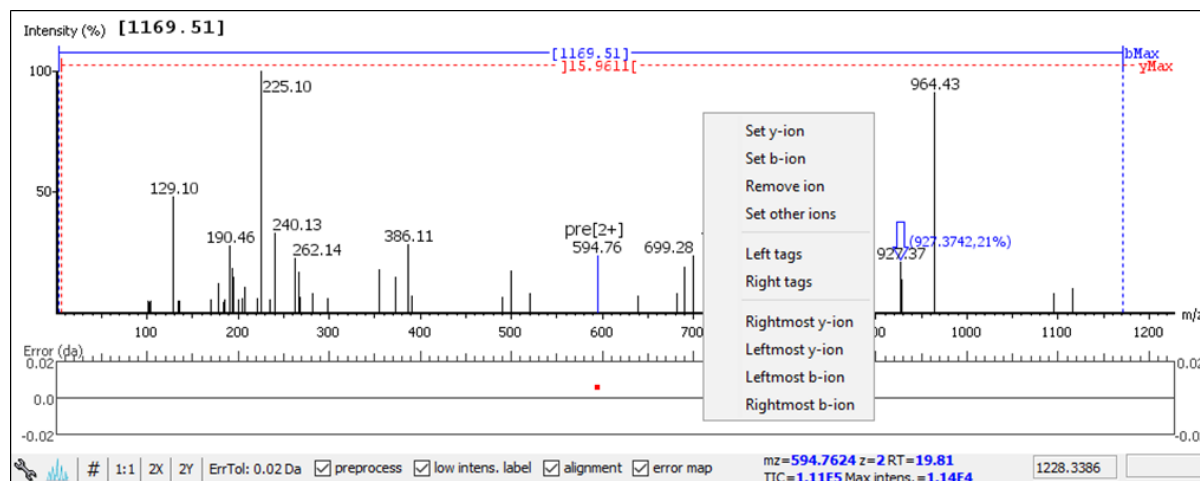
**Measure the m/z difference between two peaks.** Select a peak (blue arrow) with the *Freeze Bar* and move the mouse to the left or to the right. Hold the *Position Bar* (green triangle) above another peak. A pop-up window displays the difference between the two peaks (in the example below the difference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



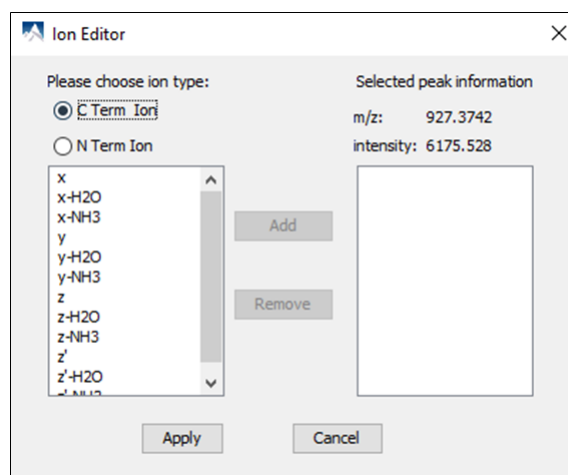
**Deselect a peak.** The *Freeze Bar* can be deselected by holding Ctrl and clicking anywhere in the spectrum.

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

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



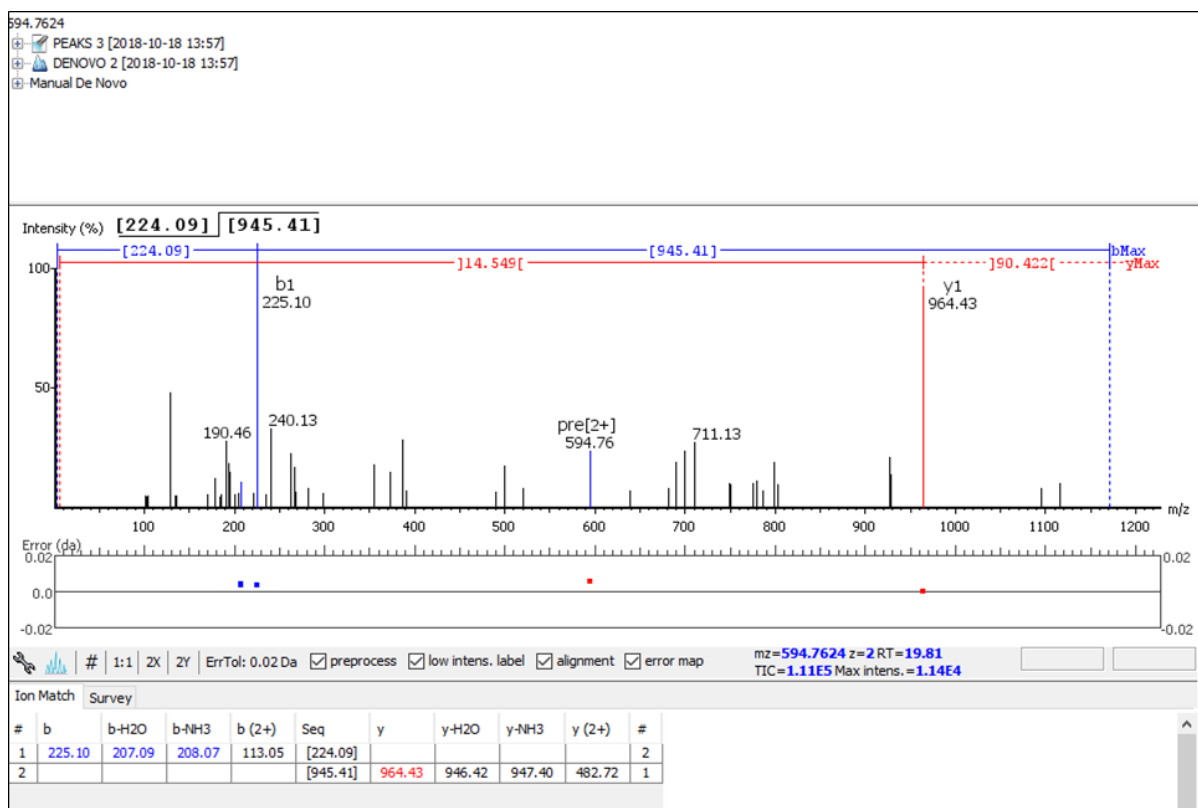
Select Set other ions from the pop-up menu to view the “Ion Editor” dialogue box. The “Ion Editor” dialogue box allows addition or removal of ion designations to/from a peak. Select either “C Term Ion” or “N Term Ion” to see the C- and N- terminal ions, respectively. Then select an ion from the ion list and press the Add button to add it to the selected ion list. Remove an ion from the selected ion list by selecting it and pressing the Remove button. Click the Apply button to apply the changes to the selected peak.



After setting an ion, the “Spectrum Annotation” panel, the “Spectrum Alignment and Error Map” panel, and the “Ion Table” panel will reflect the changes. The *peptide sequence candidate* name (as displayed in the “Result” panel and on the top of the “Spectrum Annotation” panel) will also change to reflect the mass remaining to be sequenced on either side of the ion.

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

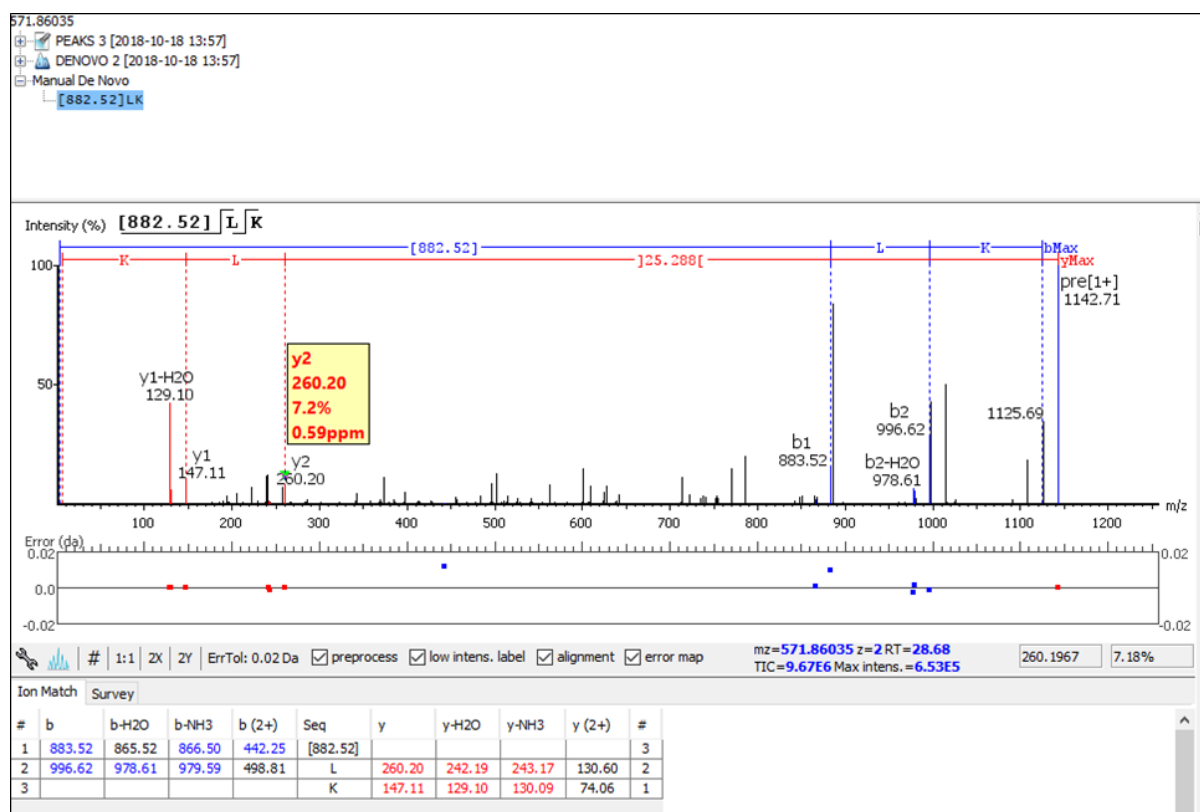




## Note

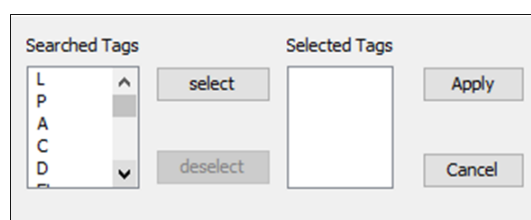
The manual *de novo* candidate information is updated in the “Result” panel, the “Ion Table” panel, and the “Spectrum Alignment and Error Map” panel. The selected ions are also annotated and color coded in the “Spectrum Annotation” panel.

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

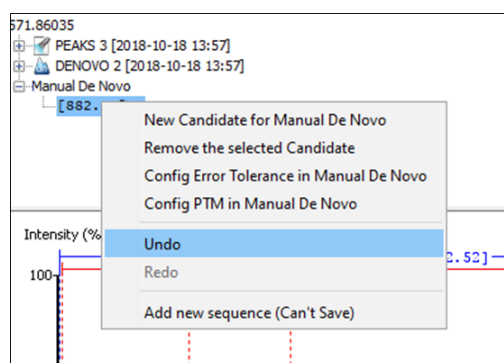


**Searching the left or right side of the spectrum for the first/last y or b ion (Search a sequence tag).** Select a peak and right-click anywhere in the Spectrum View to trigger the pop-up menu. From the menu, select either Left tags or Right tags. PEAKS will select the appropriate terminal tags and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

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



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



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

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

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

---

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

## 1. Overview

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

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

### Note

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

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

### Note

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

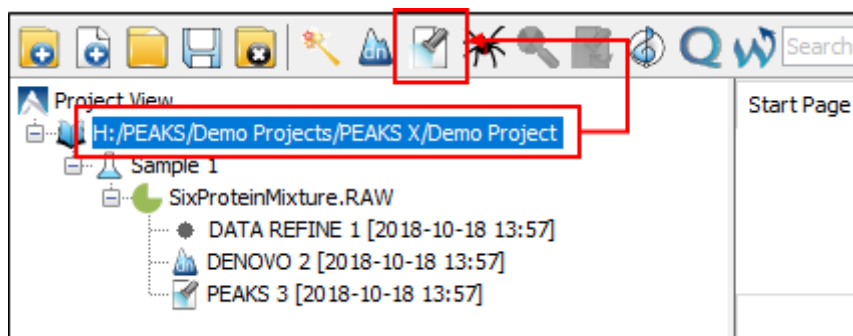
SPIDER is a dedicated search tool for finding novel peptides that are homologous to peptides in a given protein database.

### Note

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

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

1. Select a data node in "Project View". The selected node can be a fraction, a sample, or a project. PEAKS will analyze all fractions under the selected node. Launch a PEAKS search by clicking the "PEAKS DB Search" button on the tool bar.



## Note

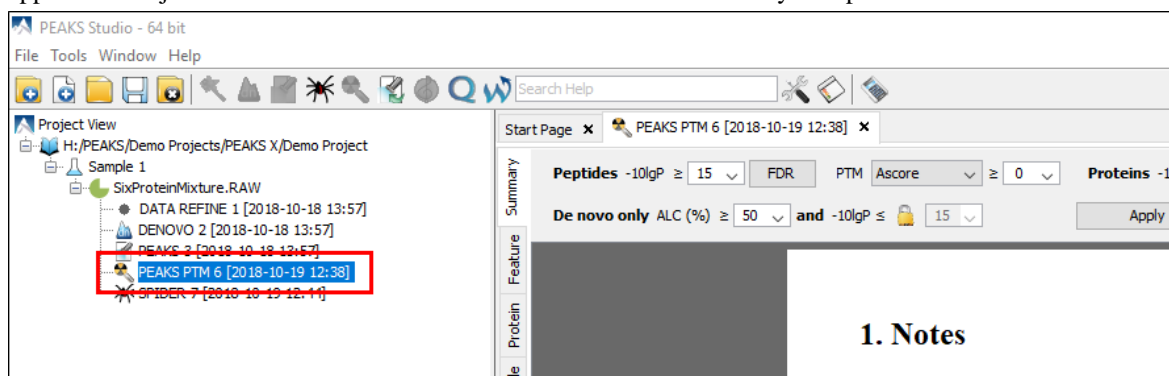
Refer to Chapter 4, *Project Wizard - Create a PEAKS Project, Load Data and Set up a Workflow* for details on creating a PEAKS project.

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


## Note

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

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



## 2. Setting PEAKS Search Parameters

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

**Error Tolerance. Precursor mass:** This enables the selection of a precursor mass error tolerance. PEAKS Search identifies peptides within the precursor mass tolerance. Precursor mass tolerance can be specified in either Da or ppm.

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

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

## Note

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

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

## Note

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

**Enzyme Selection:** This allows for the selection of the enzyme used for protein digestion from a drop-down menu. If the enzyme used for protein digestion is not in the list, a new enzyme can be added. Please refer to Chapter 8, *Peptide De Novo Sequencing* Section 2.2, “Enzyme” for further details.

## Note

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

**Digest mode:** This specifies the type of in silico cleavages allowed at a peptide terminus based on the method used for protein digestion. If “Specific” is selected, the specificity of the selected enzyme is strictly enforced at both termini of a peptide. “Semispecific” allows one terminus to disobey the enzyme specificity. “Unspecific” allows both peptide termini to disobey the enzyme specificity. “No digestion” specifies that no in silico cleavages will be applied to the peptide candidates. “No digestion” should only be used when searching a database containing peptide sequences, not intact protein sequences.

**Missed Cleavages:** This specifies the number of missed cleavage sites allowed in a peptide.

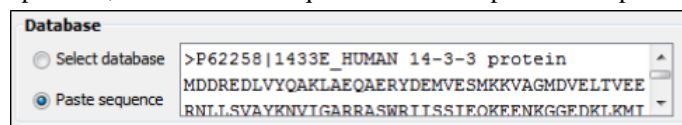
**PTM.** To select PTMs for a PEAKS DB search, click the Set PTM button to open the “PTM Options” dialogue. PEAKS DB search supports a limited number of variable PTMs. Refer to Chapter 8, *Peptide De Novo Sequencing* Section 2.3, “PTM” for details.

## Note

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

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

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



## Note

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

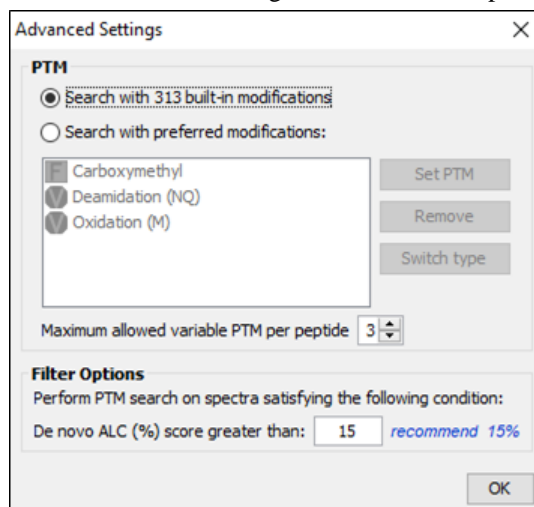
Select a contaminant database for the search from the list of databases that have been configured in PEAKS.

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

**Estimate FDR with decoy-fusion.** Select this option to enable search result validation with decoy-fusion, an enhanced target-decoy approach. When this option is selected, decoy sequences are automatically generated from the target database and are searched by PEAKS DB, PEAKS PTM, and SPIDER. This enables the estimation of the false discovery rate (FDR) in the analysis report. It is NOT recommended to de-select this option unless manual validation of the result using another approach is preferred.

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

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



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

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

## Note

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

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

## 2.1. Saving the Parameters for Future Use

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

## Note

"Advanced Settings" for PEAKS PTM are not saved.

## 3. Understanding PEAKS Search Results

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

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



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

### 3.1. Peptide and Protein Scores

**PEAKS Peptide Score (-10lgP).** The PEAKS peptide score (-10lgP) is calculated for every peptide precursor 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-precursor spectrum match. When a peptide-feature is associated with multiple spectra, the peptide score is the maximum among all PSM scores. For details of the scoring algorithm, please refer to the publication, "*PEAKS DB: De Novo sequencing assisted database search for sensitive and accurate peptide identification*" *Mol Cell Proteomics*. 2011 Dec 20.

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

### 3.2. Summary View

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

1. **Result filtering:** Filtering rules are located at the top of the "Summary" view. Low quality peptide and protein identifications can be filtered by -10lgP score. Peptide sequences identified only by *de novo* sequencing can be filtered by ALC score. More details about how to properly set the filters are discussed in Section 4, "Filtering PEAKS Search Results".
2. **Result exporting:** Identified proteins and peptides can be exported to various formats by clicking the Export button at the top of the "Summary" view. More details are discussed in Section 5, "Exporting PEAKS Search Results".

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

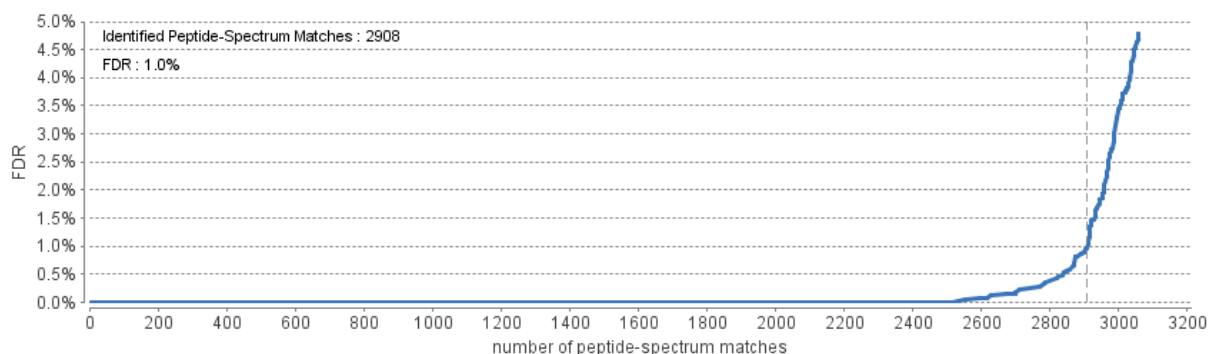
1. **Notes:** In the header of the report, information such as a description of the experiment can be entered. The notes can be edited by clicking the Notes button.
2. **Result Statistics:** Key statistics are shown in the figures and tables. The figures provide important information about target-decoy result validation, which can be used to filter the results appropriately.
3. **Experiment Control:** Figures show the distribution of precursor mass error, revealing how well the instrument is calibrated. A table shows the statistics of missed cleavages in peptides, revealing enzyme digestion efficiency of the protein sample.
4. **Other Information:** Search parameters and MS instrument information are shown.

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

#### 3.2.1. False Discovery Rate (FDR) Curve

Figure 1 shows the FDR curve with respect to the number of PSMs being kept after filtration by the peptide -10lgP score. By lowering the score threshold, more PSMs are kept in the filtered result. On the other hand, the FDR

increases because more false positives are kept. In this figure, the vertical line indicates the current score threshold. The number of PSMs and the corresponding FDR by the current score threshold are shown in the top-left corner.



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

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

## Note

In a chimeric spectrum, multiple peptide features are co-fragmented in DDA mode, and therefore may result in multiple PSMs.

## Note

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

### 3.2.2. PSM Score Distribution

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

Figure 2(a) is a stacked histogram showing the number of target and decoy matches in each score interval. There should be a similar number of target and decoy matches with low scores, but very few decoy matches with high scores. The vertical line indicates the current score threshold.

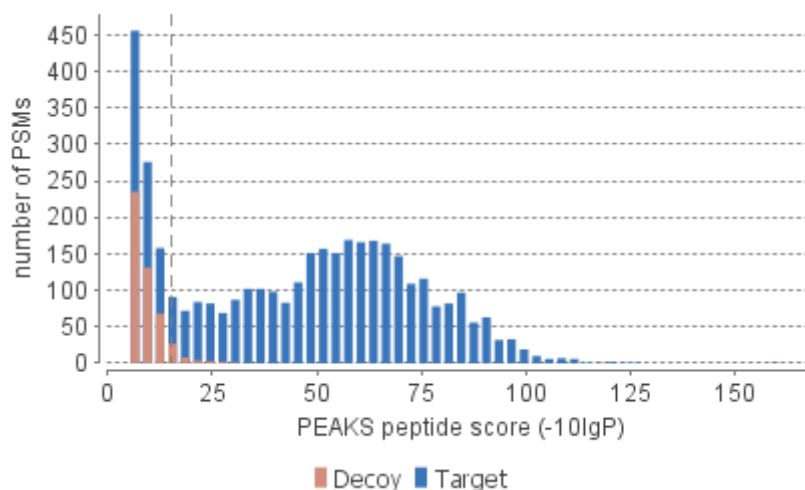
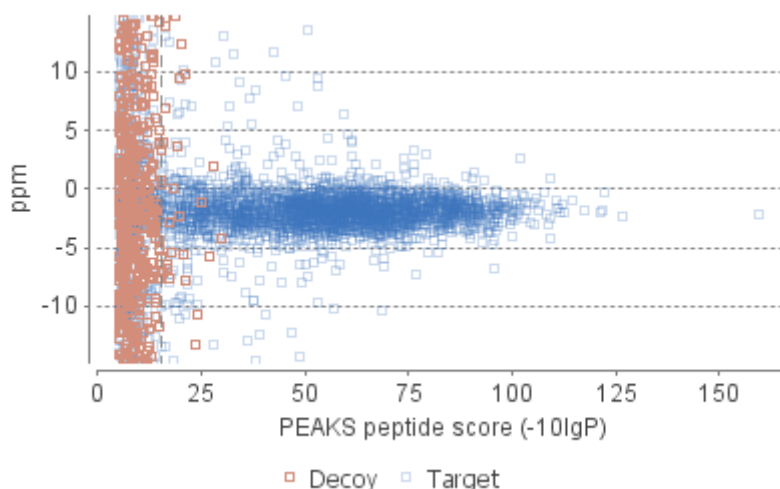


Figure 2(b) is a scatter plot showing the peptide score versus the precursor mass error in ppm for all PSMs. This figure is most useful for high resolution instruments. The PSMs with high scores should be centered near the mass error of 0, and the PSMs with low scores are usually scattered to larger mass error. The vertical dotted line indicates the current score threshold.



### 3.2.3. Distribution of Peptide Feature Detection

Figures 3(a), 3(b), and 4(a) in the "Summary" view show the distribution of peptide feature detection.

Figure 3(a) in the "Summary" view provides all detectable features distributed over the m/z range. This stacked histogram shows the number of detectable features in each m/z interval. In addition, the number of detectable features with database results are highlighted in orange and the number of detectable features with *de novo* results in addition to the database results are highlighted in light blue. In a typical dataset, it is often observed that a higher number of detectable features occur in the lower m/z range, which is reflected within the identification results.

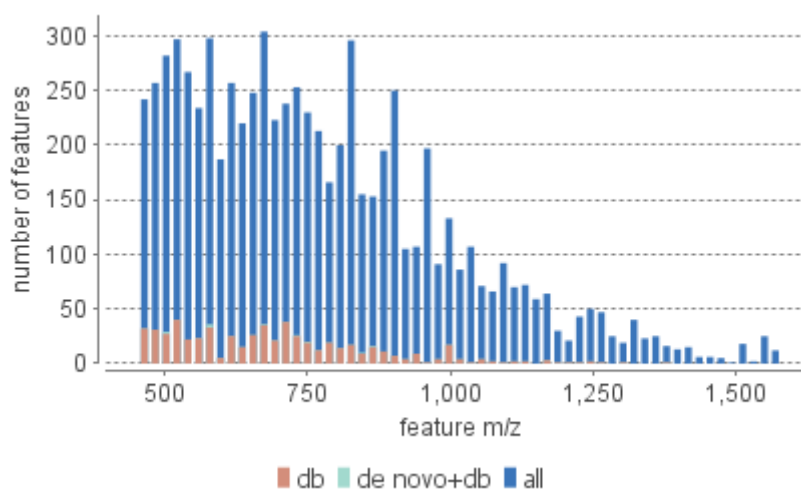


Figure 3(b) in the "Summary" view provides all detectable features distributed over the total retention time. This stacked histogram shows the number of detectable features in each RT interval. In addition, the number of detectable features with database results are highlighted in orange and the number of detectable features with database and *de novo* results are highlighted in light blue. In a typical dataset, it is often observed that a higher number of detectable features occur in the middle of the RT window, which aligns well with the identification results.

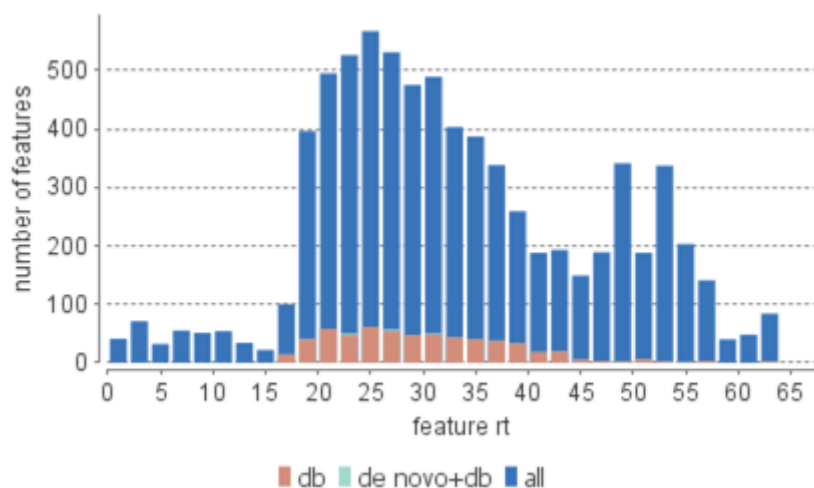
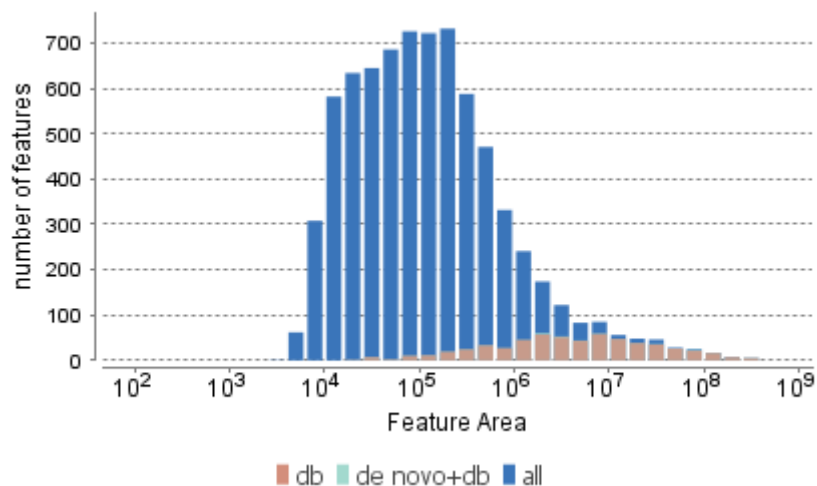
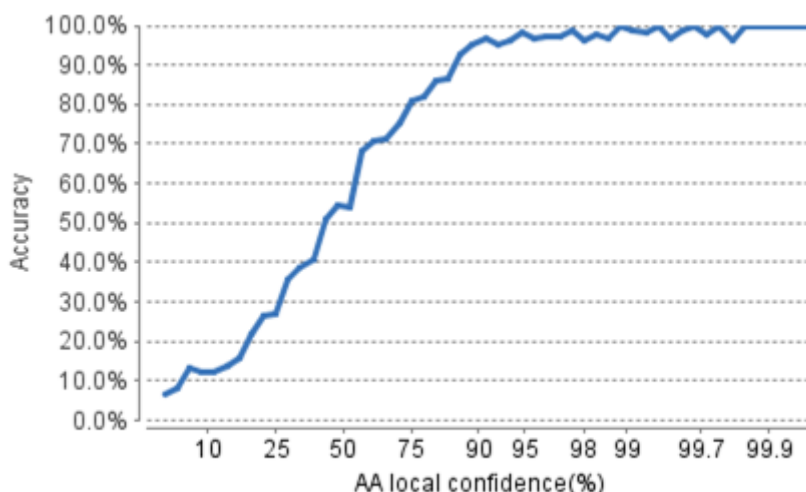


Figure 4(a) in the "Summary" view provides all detectable features distributed over the feature abundance measured as the area under the XIC curve. This stacked histogram shows the number of detectable features measured with a certain total area. In addition, the number of detectable features with database results are highlighted in orange and the number of detectable features with *de novo* results in addition to the database results are highlighted in light blue. In a typical dataset, it is often observed that more peptide features will be assigned a database ID when they exhibit a stronger LC-MS signal.



### 3.2.4. De Novo Result Validation

Figure 4(b) in the "Summary" view provides a means to estimate the expected accuracy of "de novo only" sequences for a given amino acid local confidence.



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 considered "*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  $-10\lg P$  scores above the current peptide filtering threshold.

As these residues cannot be directly validated using database peptides, their ratios of correctness are statistically estimated. By comparing the *de novo* results and the database results for all detectable tandem scans, PEAKS will statically calculate the expected accuracy. 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. By drawing this comparison, for a given amino acid local confidence, the expected accuracy of all "*De novo only*" sequences can be estimated.

### 3.2.5. Statistics of Data and Results

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

**Table 1.** Statistics of data.

# of MS scans	2076
# of MS/MS scans	900
# of Features	7427
# of Chimera scans	140

**Table 2.** Result filtration parameters.

Peptide -10lgP	≥ 15
Peptide AScore	≥ 0
Protein -10lgP	≥ 20
Proteins unique peptides	≥ 0
De novo ALC Score	≥ 50%

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

Peptide-Spectrum Matches	433
Peptide sequences	271
Protein groups	20
Proteins	42
Proteins (#Unique Peptides)	30 (>2); 1 (=2); 9 (=1);
FDR (Peptide-Spectrum Matches)	1.8%
FDR (Peptide Sequences)	2.6%
FDR (Protein Group)	0.5%
De Novo Only Spectra	107
# of identified Features	443

**Table 4.** PTM profile.

Name	ΔMass	Position	#PSM	-10lgP	Abundance	AScore
Carboxymethyl	58.01	C	197	86.04	1.18E7	1000.00
Deamidation	.98	NQ	45	70.96		46.13
Oxidation	15.99	M	16	65.74	3.62E6	1000.00

- **Table 1. Statistics of data:** Table 1 displays the total number of MS and MS/MS scans obtained from the raw data. This table also provides details for the total number of features detected and the total number of chimera scans, if "Associate feature with chimera scan (DDA)" was selected during Data Refinement.
- **Table 2. Result filtration parameters:** Table 2 summarizes the result filtration parameters set at the top of the "Summary" view page. This information provides a helpful reference for exported data.
- **Table 3. Statistics of filtered results:** Table 3 presents the statistics based on the data analysis results after setting the desired filters. See below for further explanation:
  - **Peptide-Precursor Spectrum Match:** The number of peptide features that had an associated spectrum which matched to a database sequence.
  - **Peptide Sequences:** This represents the number of distinct peptides in the filtered result. Peptides with the same primary sequence but different PTMs are counted separately. However, if several peptides differentiated only by leucine (L) and isoleucine (I), they are counted as one peptide. Since the same peptide may be identified from multiple spectra due to data redundancy and differing charge states, this number is usually smaller than the number of peptide-spectrum matches.
  - **Protein Groups:** In 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:** Provides the number of proteins given the parameters set for PEAKS Protein Score (-10lgP) and number of unique peptides.
  - **Proteins (#Unique Peptides):** This entry counts the number of proteins identified with unique peptides. A unique peptide is a peptide with a -10lgP score above the peptide filtering threshold that can be mapped to only one protein group.

- **FDR (Peptide-Spectrum Matches):** The total number of decoy database assignments to spectra relative to the total number of target database assignments to spectra represented as a percentage.
- **FDR (Peptide Sequences):** The total number of decoy database assignments to unique peptide sequences relative to the total number of target database assignments to unique peptide sequences represented as a percentage.
- **FDR (Protein Group):** The total number of decoy protein groups where the top hit in the protein group is a decoy database protein relative to the total number of target database protein groups where the top hit is from the target database.
- **De novo only Spectra:** Corresponds to the number of existing de novo sequencing results without a positive protein identification, given the filters set for the search results.
- **# of Identified Features:** The total number of peptide features which have an associating scan that matches to the database, given the filters set for the search results.
- **Table 4. PTM Profile:** shows modifications identified in the filtered result and the number of PSMs containing each modification

### 3.2.6. Experiment Control

Figures 5(a) and 5(b) in the "Summary" view show the precursor mass errors of PSMs in filtered results. Figure 5(a) shows the distribution of the precursor mass errors in a histogram with and without software calibration. Figure 5(b) shows precursor m/z versus precursor mass error in a scatter plot after software calibration. The precursor mass error is calculated in ppm as  $10^6 \times (\text{precursor mass} - \text{peptide mass}) / \text{peptide mass}$ . Calibration is for display purposes only and PEAKS corrects m/z value for each PSM to correct systematic biases and instrument measurement drift. Figures 5(a) and 5(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.

**Figure 5.** Precursor mass error of peptide-spectrum matches (PSM) in filtered result. (a) Distribution of precursor mass error in ppm; (b) Scatterplot of precursor m/z versus precursor mass error in ppm.

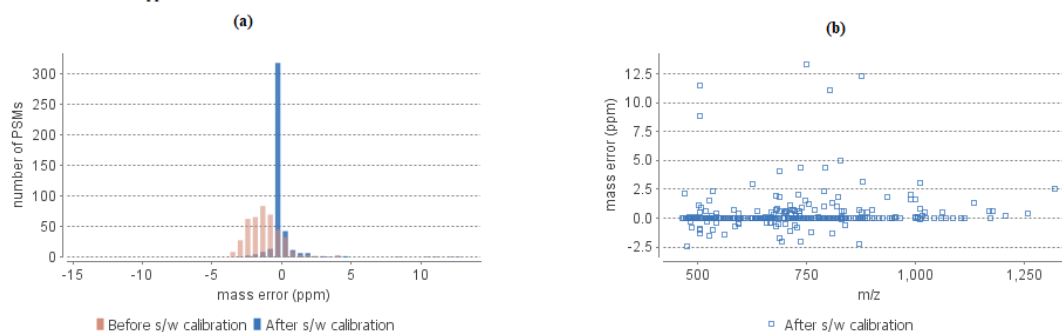


Table 5 shows the number of identified peptides in each sample with the number of missed cleavages, indicating the enzyme digestion efficiency.

**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

### 3.2.7. Other Information

Table 6 summarizes the parameters defined by the user to analyze the data. This information includes error mass tolerances, missed cleavages, database(s) used, etc.

Table 7 summarizes the instrument parameters defined by the user during project creation. This information includes ion source, fragmentation method, MS scan mode, MS/MS scan mode, etc.

**Table 6.** Search parameters.

Search Engine Name: PEAKS  
Parent Mass Error Tolerance: 15.0 ppm  
Fragment Mass Error Tolerance: 0.05 Da  
Precursor Mass Search Type: monoisotopic  
Enzyme: Specified by each sample  
Max Missed Cleavages: 3  
Non-specific Cleavage: one  
Fixed Modifications:  
  Carboxymethyl: 58.01  
Variable Modifications:  
  Deamidation (NQ): 0.98  
  Oxidation (M): 15.99  
Max Variable PTM Per Peptide: 3  
Database: UniProtKB/Swiss-Prot\_20180710  
Taxon: All  
Contaminant Database: cRAP 2017/7/13  
Searched Entry: 557829  
FDR Estimation: Enabled  
Merge Options: no merge  
Precursor Options: corrected  
Charge Options: no correction  
Filter Charge: 2 - 8  
Process: true  
Associate chimera: yes

**Table 7.** Instrument parameters.

Fractions: SixProteinMixture.RAW  
Ion Source: ESI(nano-spray)  
Fragmentation Mode: high energy CID (y and b ions)  
MS Scan Mode: FT-ICR/Orbitrap  
MS/MS Scan Mode: FT-ICR/Orbitrap

### 3.3. Feature View

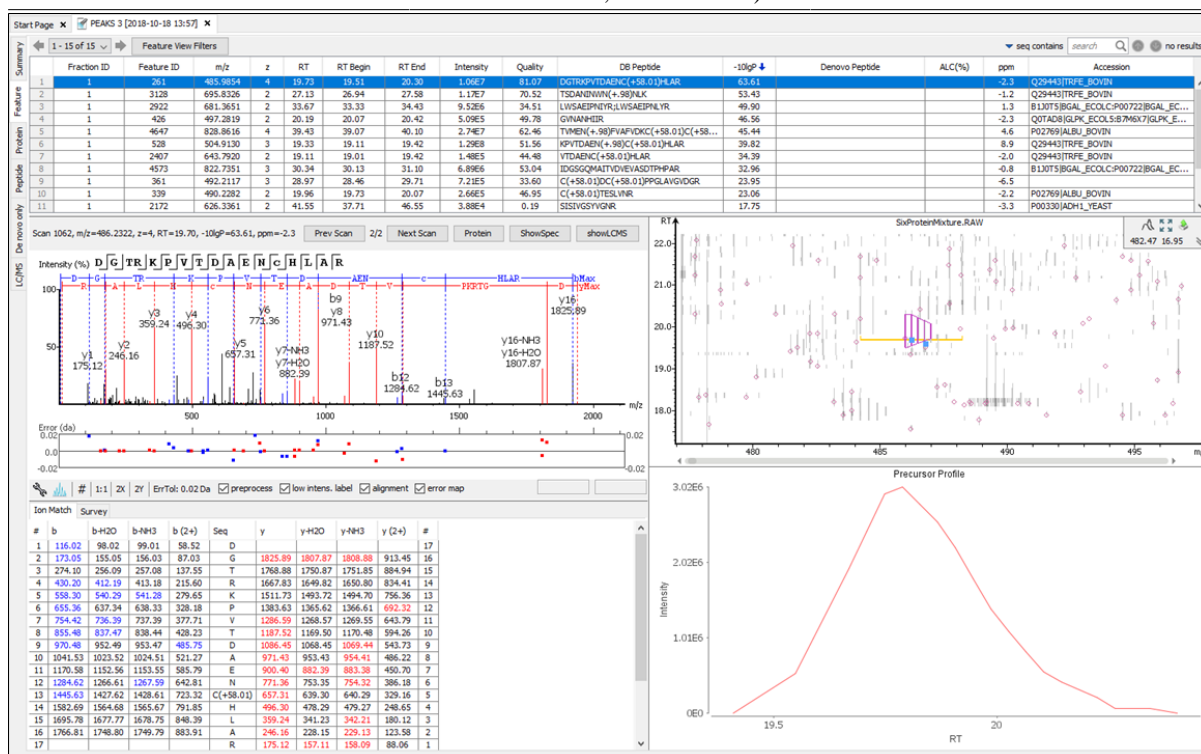
The "Feature View" lists all detectable features resulting from the filters set by the current settings in "Summary" view and "Feature" view filters. For each feature, the feature information is displayed and the database or *de novo* peptide result. The result can be sorted by any column.

#### Note

Please refer to Chapter 16, *Data Independent Acquisition (DIA) Data Analysis* and Chapter 17, *Ion Mobility Data Analysis* for *de novo* peptide view details for DIA and IMS data.



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

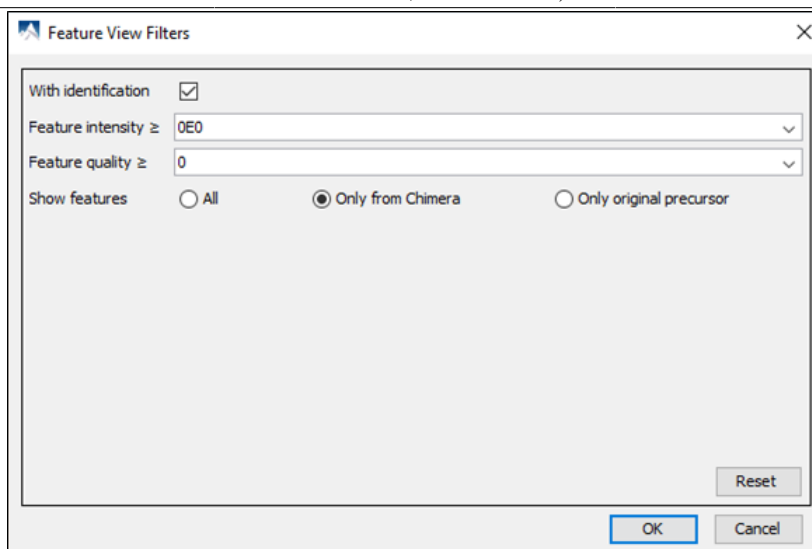


This view is composed of six main components:

- **Feature Table:** The feature table lists all features with a database and/or de novo assignment that have passed the filters set in the Summary View and the Feature View Filters found at the top of the feature table..
- **Spectrum Annotation:** The spectrum annotation displays a graphical representation of the peptide-precursor spectrum match. If viewing features without an identification, which can be specified in the Feature View Filters, there will be no PSM present to show Spectrum Annotation.
- **Ion Table:** The “Ion Match” tab at the bottom panel of the feature view contains the “Ion Table” that shows the calculated mass of possible fragment ions.
- **Parent Scan:** The “Survey” tab displays the precursor ion spectrum.
- **LC-MS Snapshot:** The LC-MS Snapshot shows a zoomed-in display of the peptide feature-spectrum match from the LC-MS view.
- **Precursor Profile:** The Precursor Profile, or eXtracted Ion Chromatogram (XIC) of the precursor ion graphs the elution profile of the selected peptide feature vector over the retention time.

### 3.3.1. Feature View Filters:

By selecting the “Feature View Filters” button, the peptide features displayed can be filtered with the following options:



The dialog box titled "Feature View Filters" contains the following controls:

- With identification:** A checked checkbox.
- Feature intensity  $\geq$ :** A dropdown menu showing "OE0".
- Feature quality  $\geq$ :** A dropdown menu showing "0".
- Show features:** Three radio buttons: "All" (unselected), "Only from Chimera" (selected), and "Only original precursor" (unselected).
- Buttons:** "Reset", "OK", and "Cancel".

- **With identification:** By selecting this check box, only features with a positive database result or a confident de novo result will be displayed in the feature table, given the thresholds filters set on the "Summary" view. If this check box is unselected, all detectable features will be presented in the feature table, including those without any peptide sequence assignment.
- **Feature intensity:** Set a value for the minimum intensity a peptide feature must exhibit.
- **Feature quality:** Set a value for the minimum quality score a peptide feature must have.
- **Show features:** If "All" is selected, all detectable features will be presented in the feature table. If "Only from Chimera" is selected, all detectable features associated with chimera spectra will be displayed. If "Only original precursor" is selected, only the original peptide feature is displayed, omitting all features which are associated with chimeric scans.

### 3.3.2. Feature Table

The Feature Table shows the detectable feature results. Each row in the table represents a peptide feature detected in the LC-MS. The table can be sorted by any column. A search function is also provided. When there are more than 1000 peptides, the table is split into multiple pages.

- **Page Drop-down and Scroll Arrows:** 1000 entries are listed per page. Click the drop down menu to quickly scroll through the table. Use the scroll arrows to move forward or back in the list.
- **Table Search Bar:** Search the table by sequence or m/z. If hits are found the total number of hits will be displayed beside the bar with blue arrows allowing you to scroll through the table. Hits will be highlighted in light blue.

1 - 15 of 15														
Feature View Filters														
	Fraction ID	Feature ID	m/z	z	RT	RT Begin	RT End	Intensity	Quality	DB Peptide	-10lgP	De novo Peptide	ALC(%)	Accession
1	1	3128	695.8326	2	27.13	26.94	27.58	1.1767	70.52	TSQANNNYIN(+58.0)HLAR	53.43			Q29443TRFE_BOVIN
2	1	2922	681.3651	2	33.67	33.33	34.43	9.5266	34.51	LWSAEPNRYR(LWSAEPNLYR	49.90			B1J075IBGAL_ECCLC-P00722BGAL_EC...
3	1	426	497.2819	2	20.19	20.07	20.42	5.0955	49.78	GVNANHHR	46.56			Q07A28IGLPK_ECCLC-SB7M6X7IGLPK_E...
4	1	4647	828.8616	4	39.43	39.07	40.10	2.7467	62.46	TVMEN(+58.0)FVAFVOKC(+58.0)C(+58...	45.44			P02769IALBU_BOVIN
5	1	528	504.9130	3	19.33	19.11	19.42	1.2688	51.56	KPTIDAEIN(+58.0)C(+58.0)HLAR	39.82			Q29443TRFE_BOVIN
6	1	2407	645.7920	2	19.11	19.01	19.42	1.4865	44.48	VTDAEKNC(+58.0)HLAR	34.39			Q29443TRFE_BOVIN
7	1	4573	822.7351	3	30.34	30.13	31.10	6.8966	53.04	IDSGGQMATYDVEASDTHPAR	32.96			B1J075IBGAL_ECCLC-P00722BGAL_EC...
8	1	361	492.2117	3	28.97	28.46	29.71	7.21E5	33.60	C(+58.0)DC(+58.0)PPGLAVGIDGR	23.95			
9	1	339	490.2282	2	19.96	19.73	20.07	2.66E5	46.95	C(+58.0)TESLVNR	23.06			P02769IALBU_BOVIN
10	1	2172	626.3361	2	41.55	37.71	46.55	3.88E4	0.19	SISIVGSIVGNR	17.75			P00330A0H1_YEAST

The following list describes the contents in each column:

- **Fraction ID:** The fraction number of the sample file.
- **Feature ID:** An arbitrary number provided by PEAKS during data refinement used as a unique identifier.
- **m/z:** The precursor mass-to-charge ratio.

- **z:** The precursor charge.
- **RT:** The retention time (elution time) for the spectrum as recorded in the data.
- **RT Begin:** The beginning retention time (elution time) for the feature as recorded in the data.
- **RT End:** The ending retention time (elution time) for the feature as recorded in the data.
- **Intensity:** The intensity of the peptide feature found at the same m/z and retention time as the MS/MS scan. This can be used as an indicator of the abundance of the peptide.
- **Quality:** The quality score for the peptide feature. Factors that affect the quality score include m/z difference, RT difference, XIC shape similarity, and the feature intensities.
- **DB 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.
- **DB 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.
- **Denovo Peptide:** The amino acid sequence of the peptide associated with the feature, as determined by de novo sequencing. A modified residue is followed by a pair of parentheses enclosing the mass of that modification.
- **ALC (%):** The average local confidence. ALC is calculated as the total of the residue local confidence scores in the peptide divided by the peptide length.
- **PPM:** The precursor mass error, calculated as  $10^6 \times (\text{precursor mass} - \text{peptide mass}) / \text{peptide mass}$ .
- **Accession:** The accession number of the highest-scoring protein containing this peptide. If a de novo only results surpasses the ALC% threshold set on the summary page.

## Tip

Right-click on a feature to copy the peptide sequence.

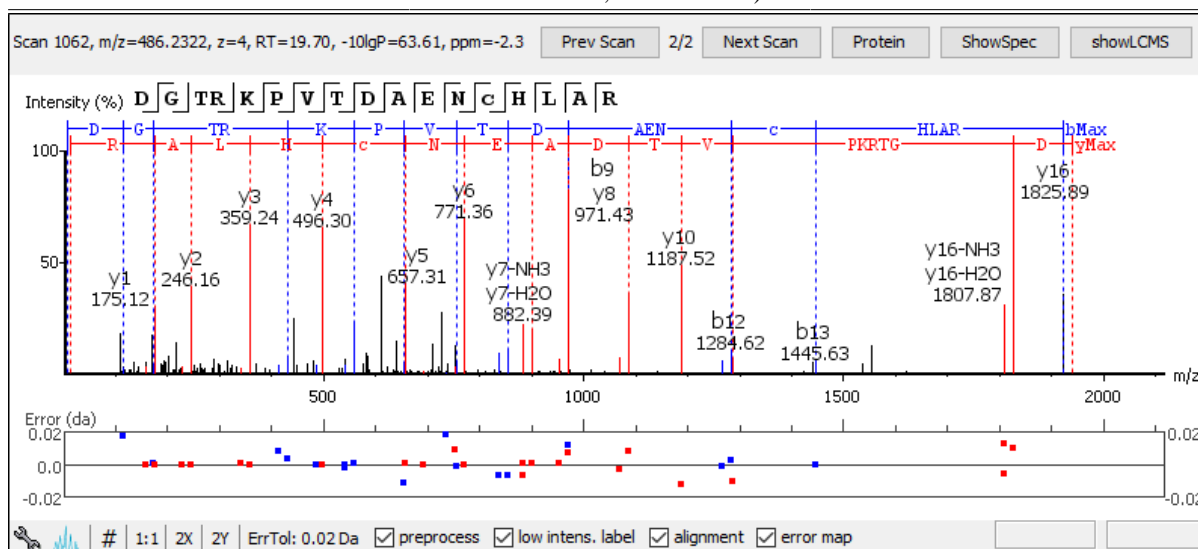
If a feature was able to match to a database sequence, the database peptide will be referenced in the "DB Peptide" column. However, if no database sequence was matched to the peptide feature, the *de novo* sequencing results will be displayed instead. Click the ALC column heading to sort by the ALC score.

	Fraction ID	Feature ID	m/z	z	RT	RT Begin	RT End	Intensity	Quality	DB Peptide	-10lgP	Denovo Peptide	ALC... ↓	ppm	Accession
1	1	913	530.7654	2	27...	26.83	28.25	1.18E8	35.71			LGWNPWKG(SAGWNPWKG	98	-0.8	Q29443 TRFE_BOVIN:P09571 T...
2	1	4044	769.4042	2	31...	31.66	32.32	2.29E6	68.08			LGEYGFQNELLYR	97	0.1	P02769 ALBU_BOVIN
3	1	3583	732.8905	2	27...	27.13	28.17	7.02E6	50.98			KLLQ(+98)SGPFVSC(+58.01)VK	97	-1.9	Q29443 TRFE_BOVIN:P09571 T...
4	1	3584	732.8905	2	26...	26.17	27.13	2.62E7	53.23			LLESGLGDDYR	96	3.1	Q29443 TRFE_BOVIN
5	1	3059	690.3191	2	34...	34.43	35.65	4.82E6	35.15			FELEGAGN(+98)VAFVK	95	-1.2	Q29443 TRFE_BOVIN:P09571 T...
6	1	450	499.3001	2	37...	37.15	38.07	6.67E6	64.40			DPALVELLK	93	0.2	P02769 ALBU_BOVIN
7	1	2598	659.3616	2	26...	26.08	26.73	2.51E7	74.12			YVDEPQNLK	91	8.4	P02769 ALBU_BOVIN:B1J0T5 BG...
8	1	1510	573.9226	3	19...	19.22	20.75	1.65E7	17.99			TPQPPDSSKDEC(+58.01)MK	90	-1.8	Q29443 TRFE_BOVIN:P09571 T...
9	1	2761	671.3469	2	35...	35.54	36.34	4.02E6	71.34			VTMQ(+98)NFVAFVK	89	-0.5	P02769 ALBU_BOVIN
10	1	2747	670.3270	2	24...	24.26	24.86	2.63E6	53.66			ELDPQEDIQR	88	-1.6	Q29443 TRFE_BOVIN
11	1	5040	860.3806	2	19...	18.90	20.65	3.8E6	11.57			QPTPPDSSKDEC(+58.01)MK	88	-1.4	Q29443 TRFE_BOVIN:P09571 T...

### 3.3.3. Peptide Precursor Spectrum Matches

The spectrum annotation displays a graphical representation of the peptide precursor spectrum match. If a feature without an identification is selected from the feature table above, which can be specified in the Feature View Filters, there will be no PSM present to show within the Spectrum Annotation window.

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



The title bar shows the information of the peptide sequence that is being displayed in the spectrum. If a feature is associated to more than one tandem scan, click the “Prev Scan” or “Next Scan” button to view the Spectrum Annotation of the other PSMs in the LC-MS view. Press the “ShowSpec” button to open the MS/MS Scan in the Data Refinement MS/MS View. Press the “ShowLCMS” button to automatically view the MS/MS and MS feature Scan in the LC-MS View of the current PEAKS Search result being viewed.

## Note

If a peptide feature spectrum match is associated to more than one protein as listed in the Accession column, clicking the “Protein” button will show a list of all the Accession names to select from.

For detailed instructions, please refer to Section 3.2.2, “Spectrum Annotation”.




### 3.3.4. Ion Table

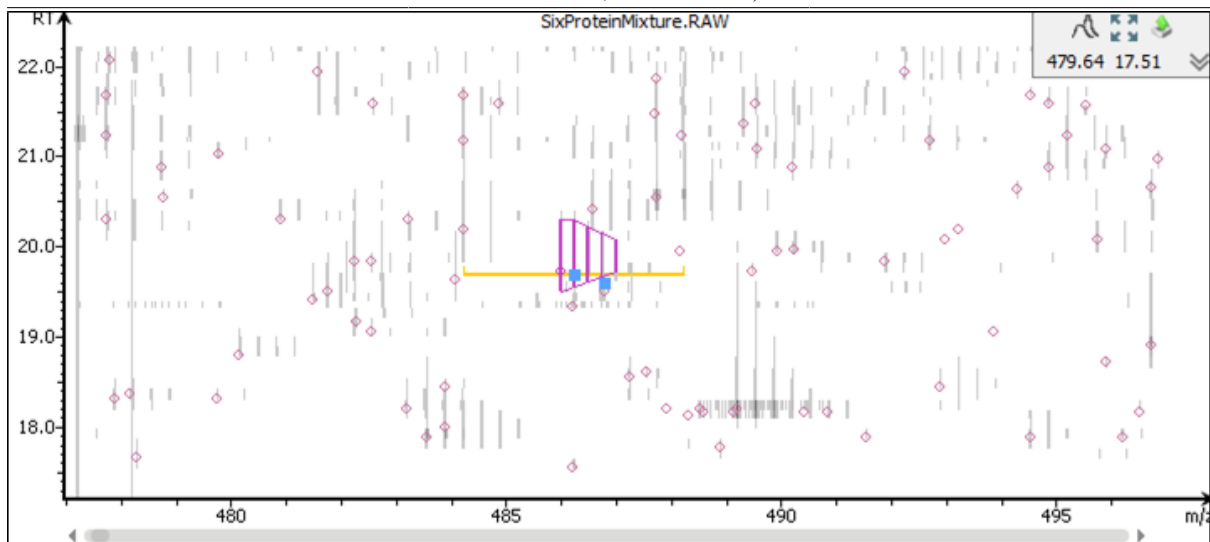
The “Ion Match” tab at the bottom panel of the feature view contains the “Ion Table” that shows the calculated mass of possible fragment ions. Ion table options are the same as described in Section 3.2.3, “Ion Table”.

### 3.3.5. Parent Scan

The “Survey” tab displays the precursor ion spectrum. The buttons that appear in this section are identical to those explained above in Section 3.2.4, “Parent Scan”.

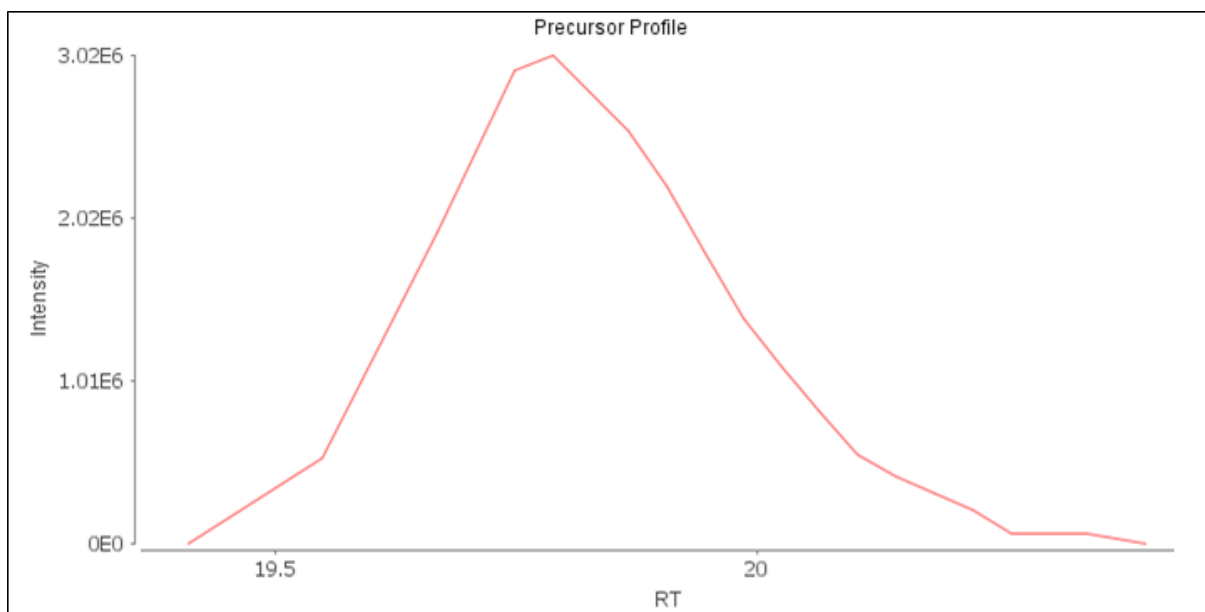
### 3.3.6. LC-MS Snapshot

LC-MS Snapshot displays a zoomed-in snapshot of the peptide precursor spectrum in the LC-MS heat map 2-D view. The peptide feature is highlighted in purple and the database matched MS/MS scan is indicated by a blue square, whereas a de novo sequenced MS/MS scan is indicated by a yellow square, and no square if no MS/MS is taken for the peptide feature. The span of the yellow line in the LC-MS Snapshot designates the isolation window. Users can change the view from the 2-D heat map to a 3-D intensity view by clicking the intensity view button  in the upper right-hand corner of the LC-MS Snapshot, and enlarge the LC-MS view to cover the full screen by clicking the expand button  in the upper right-hand corner of the LC-MS Snapshot. In addition, users can export the LC-MS heat map by clicking the export image button  in the upper right-hand corner of the LC-MS Snapshot. See Section 4.2.2, “Exporting LC-MS, IM-MS, and LC-IM-MS”.



### 3.3.7. Precursor Profile

The Precursor Profile, or eXtracted Ion Chromatogram (XIC) of the precursor ion graphs the elution profile of the selected peptide feature vector over the retention time. The graph is plotted with the retention time along the x-axis and the intensity along the y-axis. Hovering over the XIC curve will display the RT and intensity coordinates.

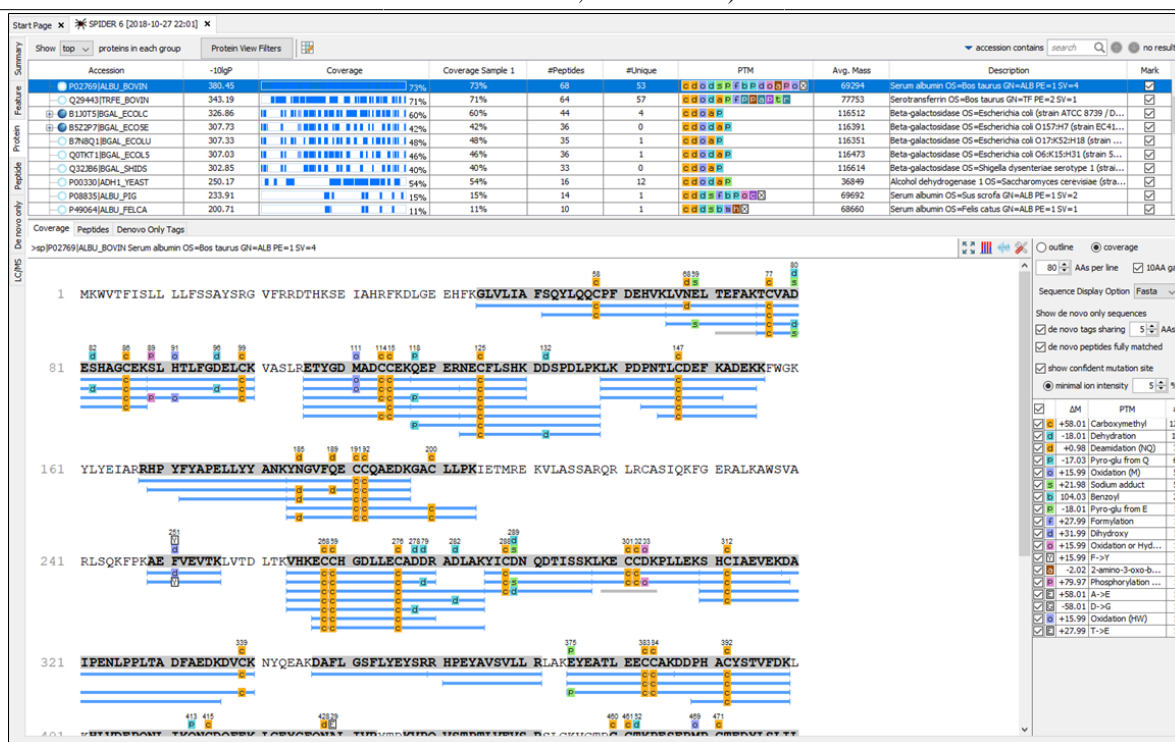


### 3.4. Protein View

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 main components:

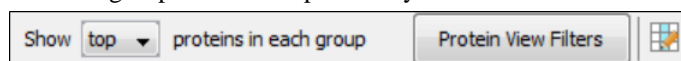
- **Protein table:** This lists protein identifications that have been filtered by the current settings in the Summary View and the "Protein View Filters".
- **Coverage:** This visualizes the protein sequence coverage of the selected protein, showing the mapping of support peptides and *de novo* tags to the protein sequence.
- **Peptides:** This lists supporting peptides mapped to the selected protein.
- **De novo Tags:** This lists "*de novo* only" peptides mapped to the selected protein.

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



### 3.4.1. Protein View Filters

- **Show Top, All, or First proteins in each group:** Adjust the protein list based on Protein grouping, top is selected by default. Proteins are grouped based on parsimony.



**Top:** Show only the top proteins in each group. These proteins are supported by the most unique peptides in the group. Proteins in the group that share a subset of the unique peptides that support the top protein (sub-proteins) will not be displayed.

**All:** Show all proteins in each group. Proteins are grouped together if they are supported by the same set or a sub-set of the top protein in the group. If a protein is supported by a peptide not supporting the top protein it will be added to a new group.

**First:** If there are more than one 'top' protein in a group, only show the first one.

- **Protein View Filters:** The following pop-up will appear when the "Protein View Filters" button is selected:

- **Protein accession/name contains:** Restrict the protein table to only display protein IDs which contain the entered text in its accession/name.
- **Protein description contains:** Only proteins with descriptions that contain the entered text will be included in the protein table if text is entered in this field.
- **Protein sample area:** Filter proteins based on total protein area found in each sample. Sample areas are calculated using the total of all peptide features from unique supporting peptides. If a protein is identified and has areas from multiple samples, only one of the samples must pass this filter to be included.
- **Spider Mutations:** Only show proteins that have mutated peptides identified with SPIDER.
- **PTM contains:** Filter protein results by the presence of selected PTMs. Proteins that are supported by peptides containing modifications that match the selected PTMs will be included. If multiple PTMs match the modification name, they will all be included in the table.

## Note

In all filters, multiple entries can be separated by a semi-colon.

- **Show/Hide Columns:** Please refer to Chapter 9, *Peptide, PTM, and Mutation Identification (PEAKS DB, PEAKS PTM, and SPIDER)* Section 3.4.3, “Global Protein Comparative View” for details.

## 3.4.2. 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.

## Note

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

Accession	-10lgP	Coverage	Coverage Sample 1	Intensity Sample 1	#Peptides	#Unique	#Spec Sample 1	PTM	Avg. Mass	Description	Mark
⊖ P02769 ALBU_BOVIN	372.77	71%	71%	3.67E9	61	59	130		69294	Serum albumin OS=Bos taurus GN=ALB PE=1 SV=4	
⊖ Q29443 TRFE_BOVIN	338.82	68%	68%	3.47E9	59	52	134		77753	Serotransferrin OS=Bos taurus GN=TF PE=2 SV=1	
⊖ B13075 BGAL_ECOLC	323.17	60%	60%	1.01E8	43	9	69		116512	Beta-galactosidase OS=Escherichia coli (strain ATCC 8739...	
⊖ P00330 ADH1_YEAST	248.36	54%	54%	3.56E8	16	12	29		36849	Alcohol dehydrogenase 1 OS=Saccharomyces cerevisiae [...]	
⊖ P00698 LYSC_CHICK	187.15	61%	61%	3.21E8	8	5	14		16239	Lysozyme C OS=Gallus gallus GN=LYZ PE=1 SV=1	
⊖ P62895 CYC_PIG	173.03	54%	54%	5.1E8	9	9	16		11704	Cytochrome c OS=Sus scrofa GN=CYC PE=1 SV=2	

The columns in the "Protein" table are:

- **Accession:** The accession number of the protein as seen in the FASTA database.
- **-10lgP:** The protein confidence score.




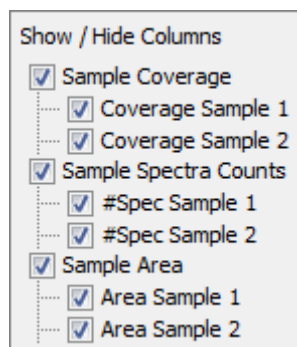
- **Coverage:** The percentage of the protein sequence that is covered by supporting peptides. The coverage is visualized by a colored bar. Light blue blocks indicate the parts of the sequence covered by low-confidence peptides. Dark blue blocks indicate the parts covered by high-confidence peptides. This is the total coverage including results from all samples.
- **#Peptides:** The number of high-confidence supporting peptides.
- **#Unique:** The number of high-confidence supporting peptides that are mapped to only one protein group. Unique peptides with same sequence but different modifications are only counted once in this number.
- **PTM:** The identified modifications displayed with color-coded icons.
- **Avg. Mass:** The protein mass calculated using the average mass.
- **Description:** The protein's header information as seen in the FASTA database.
- **Mark:** This allows the selection of specific proteins. The marked proteins are exported when using the exporting functions in the "Summary" view.

## Note

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

### 3.4.3. Global Protein Comparative View

The "Global Protein Comparative View" provides detailed information for users to draw semi-quantitative comparisons for each protein among different samples in the whole project. To Show/Hide the "Protein Comparative" view, click the  button to add and remove protein coverage, MS1 area/intensity, and spectral count columns. All samples included in the search will be displayed and named according to the user defined name.



## Note

PEAKS provides users the option to calculate peptide abundance using the MS1 feature area calculation or using the MS1 feature intensity. By default, area is used.

In the "Protein" table, one or more columns listed below will be displayed for the corresponding columns if checked.

- **Coverage by Sample:** The protein coverage as a percentage will be displayed for the individual sample when selected. Only peptides that pass the filters set in the "Summary" view are included in the percentage calculation.
- **#Spec by Sample:** The total number of spectra identified that support the given protein. Totals are given for each sample included in the search.



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

- **Area/Intensity by Sample:** The total area or intensity of peptide features from unique supporting peptides in each sample will be displayed. Peptides must pass the filters set in the "Summary" view to be included in the calculation.

Accession	Coverage	Coverage Sample 1	Coverage Sample 2	Area Sample 1	Area Sample 2	#Spec Sample 1	#Spec Sample 2
Proteins							
P04406 G3P_HUMAN		87%	82%	2.73E10	2.93E10	156	168
P30086 PEBP1_HU...		87%	87%	4.81E9	5.73E9	50	43
P23528 COF1_HUM...		86%	83%	9.37E9	9.94E9	42	46
P80349 FUC3_RAT		85%	85%	2.5E8	3.25E8	16	16
P04075 ALDOA_HU...		83%	73%	1.71E10	1.78E10	101	91
Q0ZQL2 PPIA_PANPA		83%	83%	1.7E10	1.85E10	60	62
Q01469 FABP5_HU...		80%	65%	2.48E9	3.29E9	18	20
P06733 ENO4_HUM...		79%	78%	1.06E10	1.22E10	149	173
P61604 CH10_HUM...		79%	79%	1.17E10	1.28E10	34	28
P09936 UCHL1_HU...		78%	72%	7.65E9	8.78E9	38	38
P60175 TPIS_PANTR		78%	77%	7.52E9	8.29E9	48	49
Q06830 PRDX1_HU...		78%	63%	6.9E8	8.03E8	57	63
Q4R561 ACTB_MA...		78%	78%	2.78E9	2.77E9	221	245

### 3.4.4. Protein Coverage

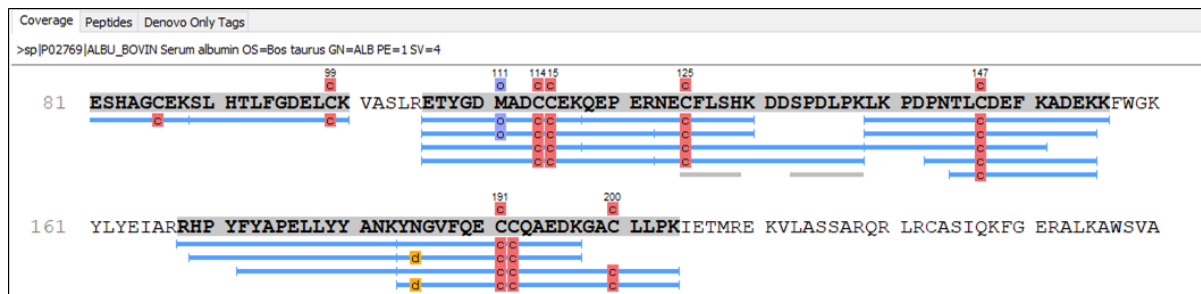
The "Protein Coverage" view visually maps the supporting peptides and *de novo* tags to the protein selected in the Protein table. It also shows all identified sites with modifications or mutations to assist with protein characterization at the amino acid level. The "Protein Coverage" view is composed of three major components - Protein Sequence Coverage, Coverage Control Panel, and Protein Tools.

#### 3.4.4.1. Protein Sequence Coverage

This area visualizes the coverage of the protein sequence.

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

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



#### 3.4.4.2. Coverage Control Panel

This panel controls what to display in the protein sequence coverage view.

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

☐ outline ☒ coverage

80 AAs per line ☒ 10AA gap

Sequence Display Option Fasta

Show de novo only sequences

☒ de novo tags sharing 5 AAs

☒ de novo peptides fully matched

☒ show confident modification site

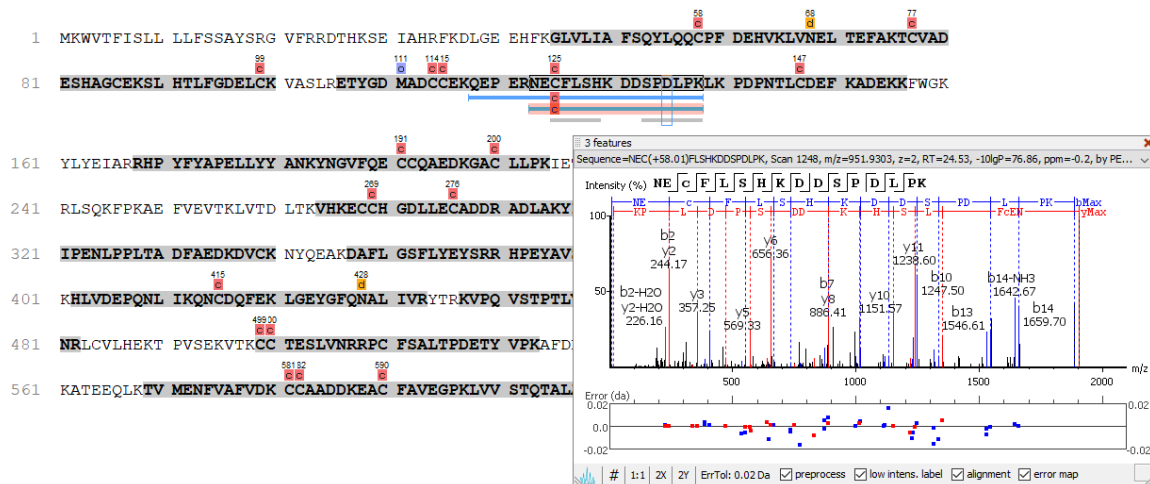
☒ minimal ion intensity 5 %

☐ minimal Ascore 20.0

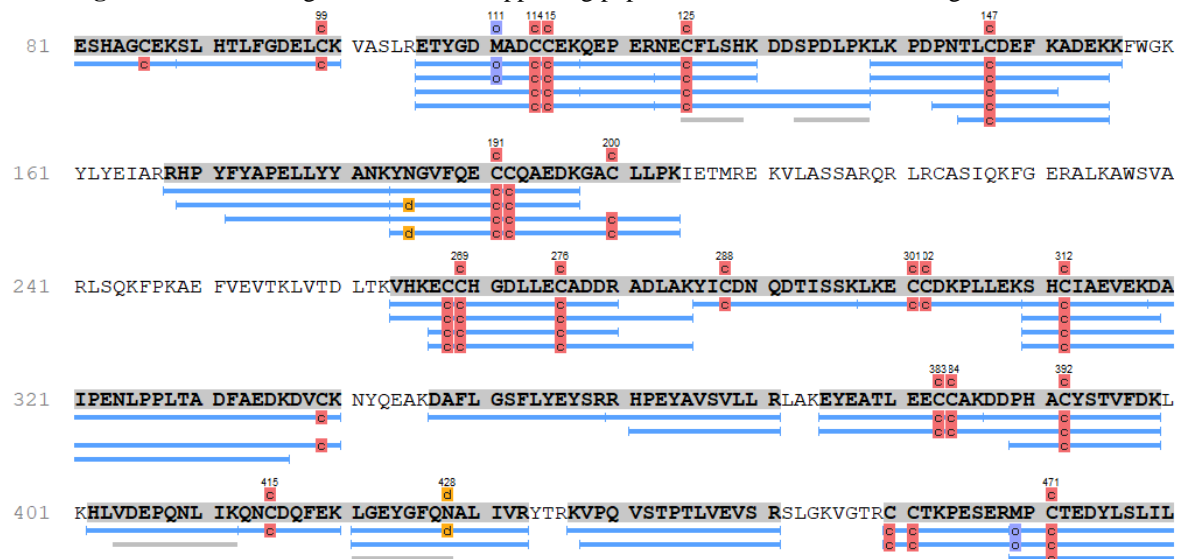
<input checked="" type="checkbox"/>	$\Delta M$	PTM	#
<input checked="" type="checkbox"/>	+58.01	Carboxymethyl	96
<input checked="" type="checkbox"/>	+0.98	Deamidation (NQ)	6
<input checked="" type="checkbox"/>	+15.99	Oxidation (M)	5

- **Sequence Display Mode:** 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.



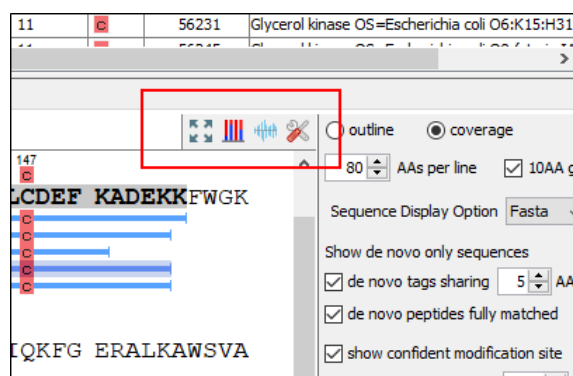
**Coverage mode:** In coverage mode, all the supporting peptides and matched *de novo* tags are shown.








- **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 in the protein sequence.
- **Sequence Display Option:** Select whether to view the protein sequence as Fasta or a specific enzyme used in the project. When a particular enzyme is selected, the protein sequence will be highlighted to emphasize expected peptides generated by the selected enzyme's cleavage specificity.
- **De novo tags sharing X AAs:** Show *de novo* tags with at least X consecutive amino acids matched to the protein sequence.
- **De novo peptides fully matched:** Show *de novo* tags with their full sequences matched to the protein sequence. This control is especially useful for showing short but fully matched *de novo* peptides.
- **Modification minimal ion intensity:** A modification or mutation is considered confident if there are fragment ions supporting the modified residue with the minimal peak intensity above the specified threshold. Only confident modifications and mutations are displayed as icons above the protein sequence
- **Modification table:** The modification table shows the modifications identified in supporting peptides on the protein. For each modification, the number of supporting PSMs with this particular modification is shown. The checkbox on the left controls whether to indicate the modification in the protein coverage. Double-clicking on the name of a modification shows detailed information about the modification. Right clicking a modification shows a pop-up menu for quick operations.

### 3.4.4.3. Protein Tools

The "Protein View" tool bar is at the upper-right corner of the protein sequence coverage tab and contains helpful protein tools.



- **Full screen** : When selected, the protein coverage window expands in full screen mode. Click  to return from full screen mode.
- **Tool box** : The tool box contains the following tools:
  - Copy template protein. This copies the original database protein into the system clipboard.
  - Copy mutated protein. This copies the protein with detected mutations into the system clipboard.
  - Save protein coverage as image. Refer to Section 4.2.1, "Exporting Protein Coverage" for more details.
  - Coverage statistics.
  - NCBI BLAST search.
  - NCBI Entrez search.
  - Multiple sequence alignment. Performs multiple sequence alignment for the selected proteins.










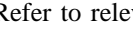
- **PTM Profiling**  Calculates the difference in abundance between peptides with confidently identified PTMs versus unmodified peptides. Peptide feature areas are used for this comparison. For a detailed walk through, see Chapter 11, *PTM Profiling* Section 2, “PTM Profiling for DB/PTM/SPIDER Results”.
- **Peptide Mapping**  Select this button to examine the detailed ion signal information generated for peptides between compared samples. See Chapter 9, *Peptide, PTM, and Mutation Identification (PEAKS DB, PEAKS PTM, and SPIDER)* Section 3.4.5, “Peptide Mapping” for more details.

### 3.4.5. Peptide Mapping

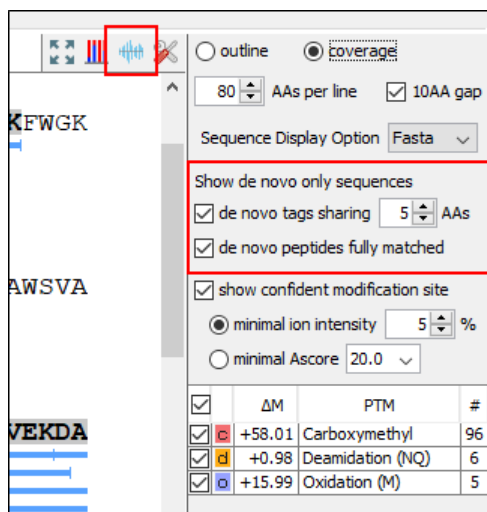
A peptide map is often referred to as a “fingerprint” of a protein. It is a powerful tool that provides a comprehensive understanding of the protein being analyzed. It gives a detailed picture of the ion signal of all peptides associated with the protein. In addition to generating the peptide map, PEAKS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

Follow the steps mentioned below to generate the peptide map:


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

Accession	-10lgP	Cc
 P02769 ALBU_BOVIN	372.77	
 Q29443 TRFE_BOVIN	338.82	
 B1J0T5 BGAL_ECOLC	323.17	
 P00722 BGAL_ECOLI	323.17	
 D00220 ADU1_YEAST	748.36	

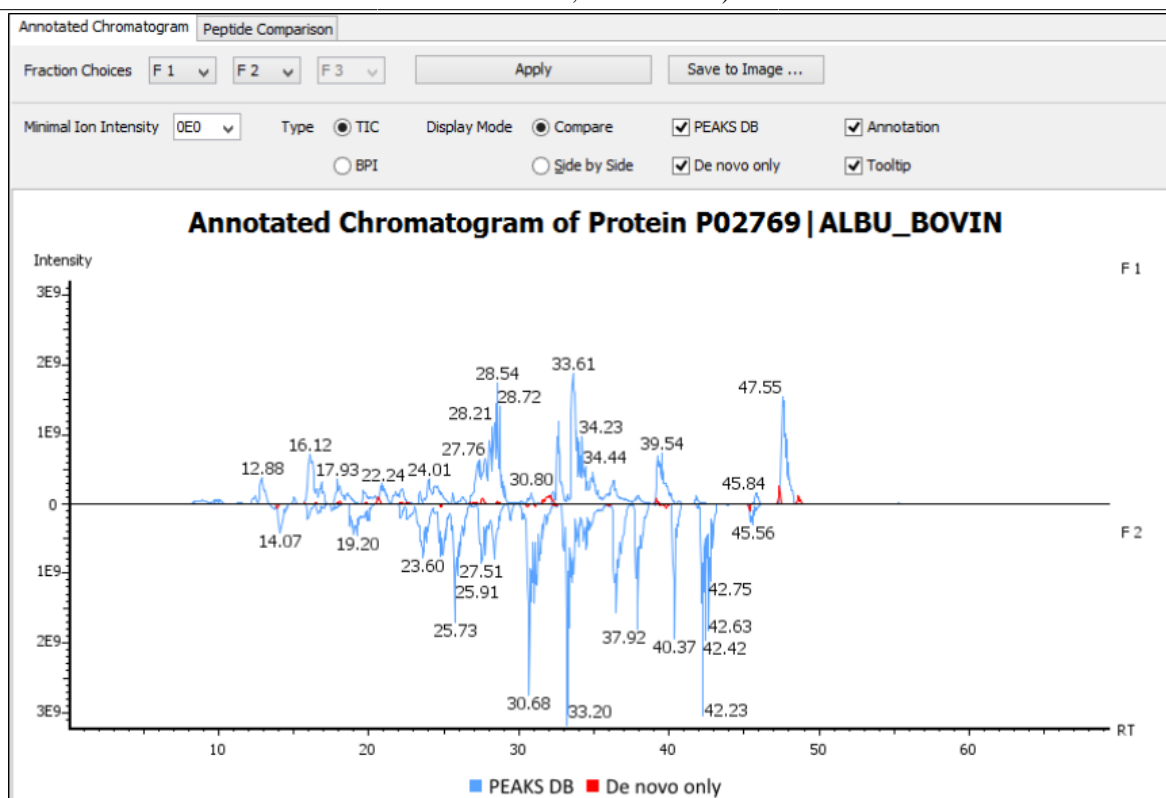
2. Select and set the desired *de novo* only peptide selection filters. Refer to relevant section of Section 3.4.4, “Protein Coverage” for details.



ΔM	PTM	#
+58.01	Carboxymethyl	96
+0.98	Deamidation (NQ)	6
+15.99	Oxidation (M)	5

3. Click the  button on the coverage pane tools bar. Once selected, a new dialogue 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.
- **Fraction Choices:** The data source used for peptide mapping. The “F#” represents the fraction to which the protein belongs, where # refers to the ID of the fraction, as indicated in the fraction node of the project tree

## Note

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

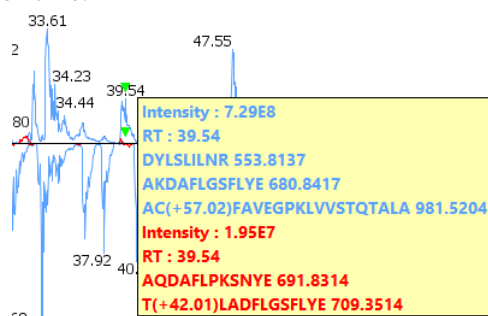
- **Minimal Ion Intensity:** A filter used to set an intensity threshold, where peptide features below this value will not be included in the display.
- **Type - TIC or BPI:** Configures the display to view the Total Ion Chromatogram (TIC) or Base Peak Intensity (BPI).
- **Display Mode:** PEAKS offers two modes to view multiple chromatograms for better comparison.
  - **Compare:** In compare mode, the protein chromatograms for the specified data samples will appear on opposite sides of the x-axis. This allows the user to directly compare the protein chromatograms between two samples.
  - **Side by Side:** In Side-by-Side mode, the chromatograms are displayed individually; each with their own set of x- and y-axes. This display mode allows for the comparison of up to three samples.
- **PEAKS DB:** Peptide features associated with peptides identified by database searching are highlighted in the annotated chromatogram.

- **De novo only:** This highlights the peptide features associated with *de novo* only peptides (see Chapter 9, *Peptide, PTM, and Mutation Identification (PEAKS DB, PEAKS PTM, and SPIDER)* Section 3.2.4, “De Novo Result Validation”).

## Note

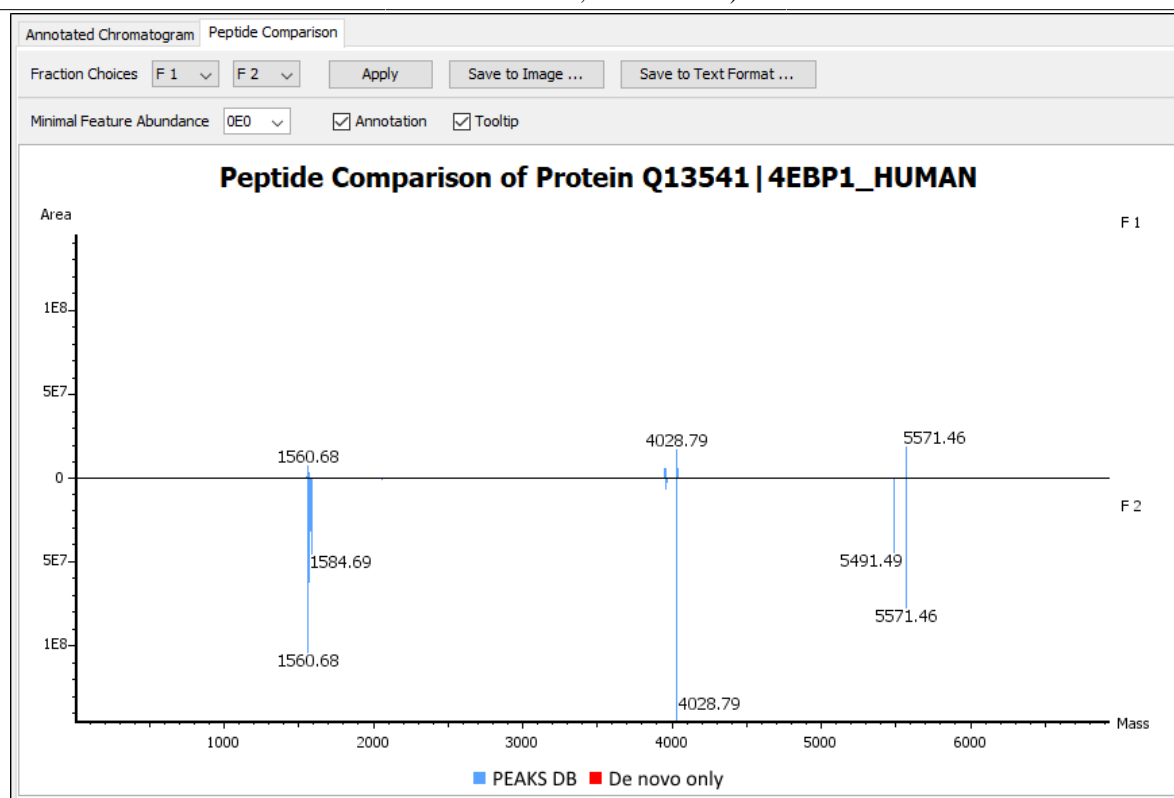
*De novo only* peptides displayed in the peptide map may be filtered using the *de novo* only peptide selection filters in the control panel of the coverage pane.

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



- **Annotated Chromatogram Chart:** This chart gives a visualization of the peptide map for the selected data. It provides a few convenient ways to zoom-in and navigate the chromatogram
  - Zoom to an RT region - Click the desired starting retention time, drag horizontally to the desired final retention time, and then release the mouse button.
  - Zoom in/out smoothly - Place the cursor at a particular RT value (right below the x-axis line), scroll the mouse wheel button.
  - Increase/decrease the ion intensity - Place the cursor on the chromatogram.
  - See the whole chromatogram - Double-click on the chromatogram.
  - Pan to an RT region: Place the cursor on the chromatogram, right-click, and then hold and drag to pan to an RT region.
- **Result Export:** The annotated chromatogram chart can be exported as an image in PNG format. Click the Save to Image... button to save the displayed chart.
- **Peptide Comparison:** Another option PEAKS offers its users to analyze the peptide map is in terms of a peptide comparison view. In this view, the total area of the features associated with a peptide is mapped to the mass of the peptide. The peptide comparison has its own input selection, result display, and result exporting options.

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

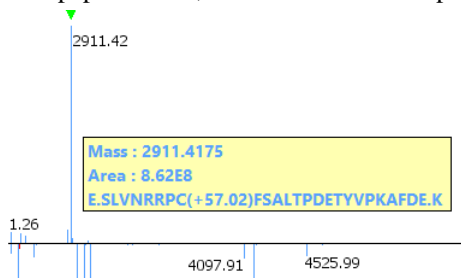


- **Inputs:** The inputs for the peptide comparison can be set using the drop-down menus at the top-left corner of the dialogue.
- **Fraction Choices:** The data source used for peptide mapping. The “F#” represents the fraction from which the protein originated, where # refers to the ID of the fraction, as indicated in the fraction node of the project tree. After making choices for the fraction selection, click the Apply button to update the chart.

## Note

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

- **Minimal Feature Abundance:** A filter used to set a feature area threshold, where peptide features below this value will not be included in the display.
- **Annotation:** When selected, the masses of the major peptides are displayed on the comparison chart.
- **Tool Tip:** When selected, a pop-up will appear when the cursor hovers over any peptide within the comparison chart. The pop-up contains the peptide mass, the total area for that peptide, and the peptide sequence.



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

- **Result Export:** The peptide comparison of a protein can be exported as an image, as well as in text format.
- **Image export:** Click the Save to Image... button to save the displayed graph as a PNG file.
- **Text format export:** Click the Save to Text Format... button to export the peptide comparison data to a text file in CSV format. The exported file contains all of the necessary information to construct the peptide comparison chart. The data are grouped together by peptide sequence and mass. The information for the same peptide are combined together if they belong to the same group, as mentioned earlier. The headers of the CSV column are self-explanatory. The following items are explained for clarity:
  - **Unique:** This specifies that the peptide is unique (Y) or not unique (N).
  - **-10lgP or ALC:** The significance score of the peptide identified by database searching or by *de novo* sequencing for *de novo* only peptides.
  - **Mass:** The mass of the peptide.
  - **ppm:** The error tolerance of the peptide.
  - **Feature Abundance:** This list of areas of the features associated with the corresponding peptide. The format of the feature areas is (area [space] +charge [semi-colon])\* . For example, 500 +2; 800 +2; 700 +3
  - **Total Abundance:** The sum of the areas of the features associated with the corresponding peptide in the specified fraction.
  - **Ratio:** When the data sources contain two fractions, the ratio of the total areas of the corresponding peptide between two fractions is calculated.

### Note

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

### Note

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

## 3.4.6. Supporting Peptides

The "Peptides" tab displays a table of supporting peptides for the protein of interest. This table is similar to the peptide table in the "Peptide" view, except that "Accession" is excluded and the columns "(Checkbox)", "Unique", "Start", and "End" are included. Descriptions of extra columns are given below. See Section 3.5, "Peptide View" for all descriptions.

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

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

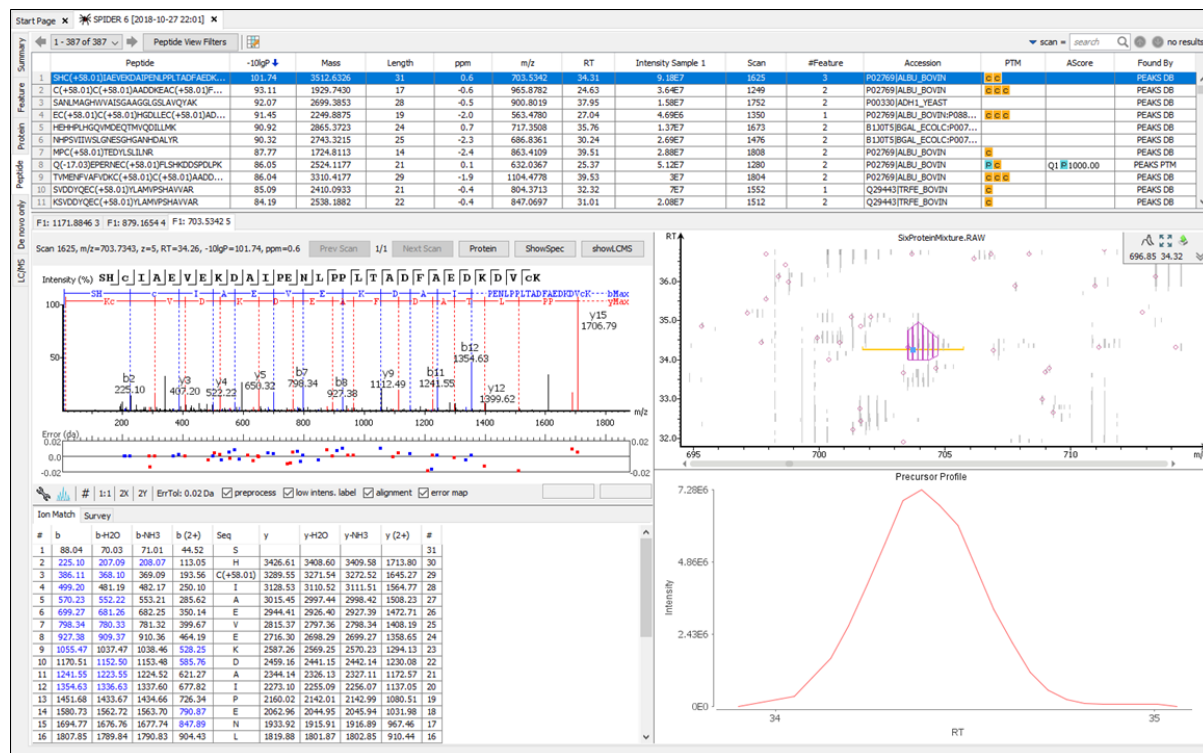
## 3.4.7. De novo Tags

The "Denovo Only Tags" tab shows a table of "*de novo* only" peptides with sequence tags matched to the protein. The "*de novo* only" peptides shown in the table are controlled by the **de novo tags sharing X AAs** and **de novo peptides fully matched** filters in the Coverage Control Panel (see Section 3.4.4.2, "Coverage Control Panel").



## 3.5. Peptide View

The "Peptide View" displays the resulting peptide identifications that have been filtered by the current settings in Summary View and "Peptide View Filters". 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.

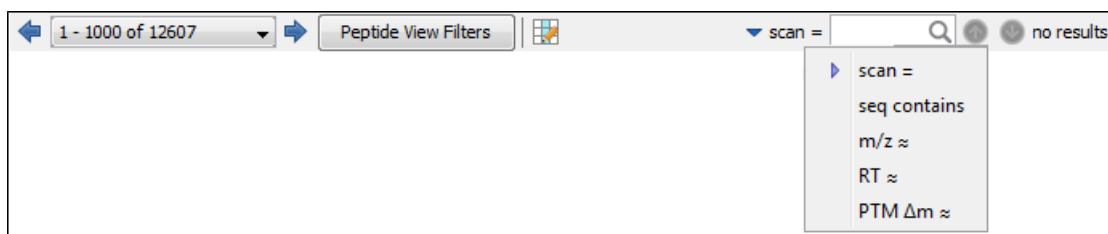


### Note

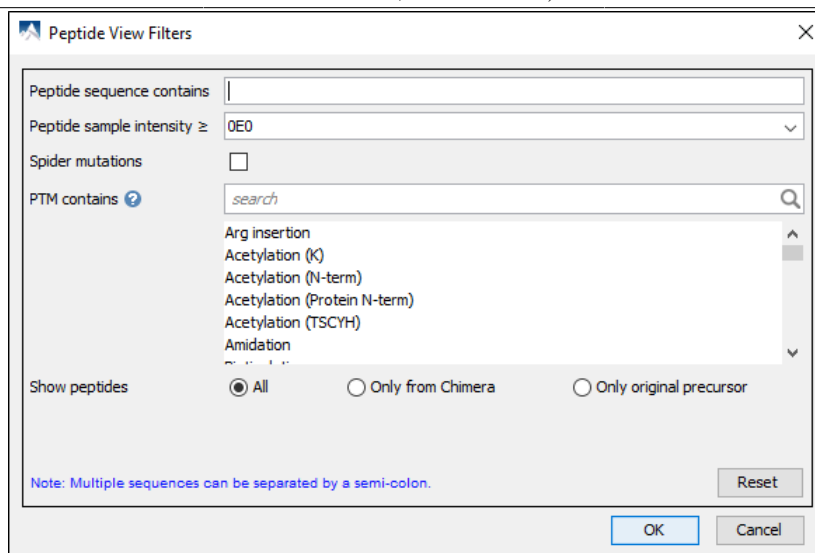
Please refer to Chapter 16, *Data Independent Acquisition (DIA) Data Analysis* and Chapter 17, *Ion Mobility Data Analysis* for feature view details for DIA and IMS data.

### 3.5.1. Peptide View Search and Filter Options

The menu above the peptide table provides options to scroll through the table of peptides, filter, and search for specific entries.



- **Page Drop-down and Scroll Arrows:** 1000 entries are listed per page. Click the drop down menu to quickly scroll through the table. Use the scroll arrows to move forward or back in the list.
- **Peptide View Filters:** Select the "Peptide View Filters" button to bring up the following pop-up to filter the peptide table.



- **Peptide sequence contains:** Enter an amino acid sequence, only peptides that contain that sequence will be displayed in the peptide table.
- **Peptide sample area/intensity:** Filters based on total peptide area found in each sample. Sample areas or intensities are calculated using the total of all peptide features from all spectra that identified the peptide. The peptide spectrum match's -10lgP score must pass the filters set in the summary view to be included in the area. If the peptide is found in multiple samples, only one sample must pass this filter for the peptide to be included.
- **Spider Mutations:** This option is available for SPIDER result only. When checked, proteins that have mutated peptides identified with SPIDER are shown.
- **PTM contains:** Type the names of PTMs you would like to include. Proteins that are supported by peptides containing PTMs that match the text will be included. If multiple PTMs match the text they will all be included in the table.

## Note

In all filters, multiple entries can be separated by a semi-colon.

- **Table Search Bar:** Several options for searching the table. If hits are found the total number of hits will be displayed beside the bar with blue arrows allowing you to scroll through the table. Hits will be highlighted in light blue. Options include:

**Scan:** Enter a specific scan number to find it in the table. If multiple data files are included in the project, include F#: before the scan number to specify

### 3.5.2. Peptide Table

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

## Note

If there are multiple PSMs matched to a peptide, the table displays the information for the top-scoring PSM. Peptides with the same sequence but different modifications are considered different and shown in different rows. The #Feature column shows the number of PSMs matched to the peptide. Other PSMs can be examined after selecting a peptide in the table. See Section 3.5.4, "Peptide-PrecursorSpectrum Matches" for details.

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

	Peptide	-10lgP	Mass	Length	ppm	m/z	RT	Intensity Sample 1	Scan	#Feature	#Feature Sampl...	Accession	PTM	AScore	Found By
1	SHC(+58.0)DAEVEDADPENPLTADFAED...	101.74	3512.6326	31	0.6	703.5342	34.31	9.18E7	1525	3	3	P02769 ALBU_BOVIN			PEAKS DB
2	C(+58.0)C(+58.0)AADOKEAC(+58.0)...	93.11	1929.7430	17	-0.6	965.8782	24.63	3.64E7	1249	2	2	P02769 ALBU_BOVIN			PEAKS DB
3	SANLMAGHWVAISGAAGLGLSLAVQYAK	92.07	2699.3853	28	-0.5	900.8019	37.95	1.58E7	1752	2	2	P00330 ADH1_YEAST			PEAKS DB
4	EC(+58.0)C(+58.0)MGDLLEC(+58.0)A...	91.45	2249.8875	19	-2.0	563.4780	27.04	4.69E6	1350	1	1	P02769 ALBU_BOVIN			PEAKS DB
5	HEHPLHQQWQEGTQWQDLIAK	90.92	2865.3723	24	0.7	717.3508	35.76	1.57E7	1673	2	2	B130T5 BGAL_ECOLC:P00...			PEAKS DB
6	NHPSIVISLGNESGSHGANHDALYR	90.32	2743.3215	25	-2.3	686.8361	30.24	2.69E7	1476	2	2	B130T5 BGAL_ECOLC:P00...			PEAKS DB
7	MPC(+58.0)TTEVLSLILNR	87.77	1724.8113	14	-2.4	863.4109	39.51	2.88E7	1808	2	2	P02769 ALBU_BOVIN			PEAKS DB
8	TVMEFVAFVQKC(+58.0)C(+58.0)AAD...	86.04	3310.4177	29	-1.9	1104.4778	39.53	3E7	1804	2	2	P02769 ALBU_BOVIN			PEAKS DB
9	SVDDHQQC(+58.0)YLAMPSPHAYVAR	85.09	2410.0933	21	-0.4	804.3713	32.32	7E7	1552	1	1	Q29443 TRFE_BOVIN			PEAKS DB
10	KSDVQQC(+58.0)YLAMPSPHAYVAR	84.19	2538.1882	22	-0.4	847.0697	31.01	2.08E7	1512	2	2	Q29443 TRFE_BOVIN			PEAKS DB
11	DVSLHPTTQISDFHAYR	82.16	2264.1914	20	-1.2	567.0544	28.25	7.87E7	1401	2	2	B130T5 BGAL_ECOLC:P00...			PEAKS DB

## Tip

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

The following list describes the contents in each column:

- **Peptide:** The amino acid sequence of the peptide, as determined by the PEAKS search workflow. A modified residue is followed by a pair of parentheses enclosing the modification mass.
- **-10lgP:** The peptide -10lgP score. The score indicates the scoring significance of a peptide-precursor spectrum match.
- **Mass:** The monoisotopic mass of the peptide.
- **Length:** The number of amino acids in the peptide sequence.
- **ppm:** The precursor mass error, calculated as  $10^6 \times (\text{precursor mass} - \text{peptide mass}) / \text{peptide mass}$ .
- **m/z:** The precursor mass-to-charge ratio.
- **RT:** The retention time (elution time) of the spectrum as recorded in the data.
- **Scan:** The scan number of the spectrum that matches the peptide sequence with the highest -10lgP.
- **#Feature:** The number of features assigned to the peptide sequence.
- **Accession:** The accession number of the highest-scoring protein containing this peptide.
- **PTM:** The types and the numbers of modifications present in the peptide shown in color-coded icons.
- **AScore:** Localization score assigned to modifications on the peptide. It is the -10 log of a p-value. In this case, the p-value is the probability that the modification occurs at the reported position compared to other possible positions. A -10lgP of 20 is equal to a p-value of 0.01. Scores are listed in the format AminoAcid, peptide position, modification, AScore. For example, serine at amino acid 3 is phosphorylated with an AScore of 54.01 will be reported as: S3P54.01.
- **Found by:** The name of the PEAKS search workflow that identifies the peptide; this can be either PEAKS DB, PEAKS PTM, or SPIDER.

## 3.5.3. Peptide Global Comparative View

"Peptide Global Comparative View" provides detailed information for users to draw semi-quantitative comparisons for each peptide among different samples in the whole project. To Show/Hide the "Peptide Global Comparative" view, click the button to add and remove peptide area/intensity, and spectral count columns. All samples included in the search will be displayed and named according to the user defined name.

## Note

PEAKS provides users the option to calculate peptide abundance using the MS1 feature area calculation or using the MS1 feature intensity. By default, area is used.

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

Show / Hide Columns

☒ Sample Intensities
☒ Intensity Sample 1
☒ Sample Features Counts
☒ #Feature Sample 1

In the peptide table, one or more columns listed below will be displayed if the corresponding columns are checked.

- **Area/Intensity by sample:** The total area/intensity of all peptide features matched to spectra that identified the peptide. Area/Intensity calculations are separated into multiple columns if multiple samples were included in the run.
- **#Feature by sample:** The number of MS/MS features associated with each peptide displayed on an individual sample basis.

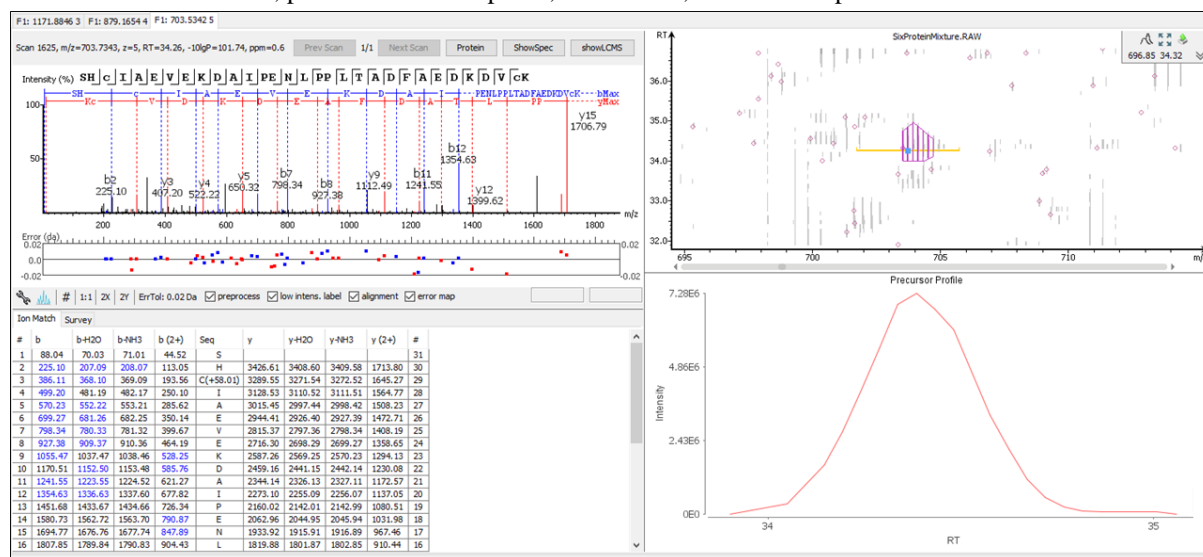
	Peptide	-10lgP ↓	Mass	Length	ppm	m/z	RT	Area Sample 1	Area Sample 2	Scan	#Feature	#Feature Sample 1	#Feature Sample 2
1	DTHEHDOT(+79.97S(+79.97)TEN...	107.83	4092.6423	34	1.6	1365.2236	102.28		7.7E7	F2:39863	1	0	1
2	MESEGGADDS(+79.97)AEEGOLDD...	100.45	4292.7041	38	3.4	1431.9135	93.28		7.7E7	F2:36223	3	0	2
3	FEESK(+8.01)EPVADEEEDS(+79...	99.05	4082.7175	34	0.6	1021.6873	106.91		9.3E7	F2:41689	4	0	2
4	MESEGGADDS(+79.97)AEEGOLDD...	97.62	3789.4844	34	1.1	1264.1702	111.32		1.57E9	F2:43521	23	0	10
5	QRLENRLPHAPGVQMQAIPEDA...	96.59	4565.0264	40	0.6	1142.2645	223.85	3.48E7	1.92E8	F1:80732	3	1	1
6	DSTLMQLRDLNLTWTS(+79.97)D...	95.97	5708.5874	50	-2.6	1428.1504	260.27	0		F1:94643	0	0	0
7	M(+15.99)ES(+79.97)EGGADDS(+...	94.73	3885.4456	34	1.3	1296.1575	108.73	3.39E7	3.52E8	F2:42837	2	1	1
8	C(+57.02)GSGPVHISGQHLVAEED...	94.30	3459.4297	31	0.7	1154.1514	157.08	5.13E8	2.06E8	F1:55046	14	7	3
9	RNPETRGYYNQDNDGLEEDES(+...	93.85	4136.6523	34	1.6	1035.1720	68.44	2.05E9	1.82E7	F2:26213	8	7	1
10	M(+15.99)EGGADDS(+79.97)AE...	93.56	3813.4934	34	1.4	1272.1736	103.56	6.1E5	3.65E7	F2:40385	5	1	4
11	GEAEQS(+79.97)EEADEEDK(+4.0...	93.41	3357.3291	28	0.7	1120.1178	49.55		9.49E6	F2:18464	2	0	1

### 3.5.4. Peptide-PrecursorSpectrum Matches

The title bar shows the information of the peptide sequence that is being displayed in the spectrum. If a feature contains two tandem scans, click the "Prev Scan" or "Next Scan" button to view the Spectrum Annotation of the other PSMs. Based on the PSM selected, the LC-MS Snapshot will change dynamically to highlight PSM in the LC-MS view. Press the "Protein" button to look at the protein view associated with the peptide feature spectrum match. Press the "Show Spec" button to open the MS/MS Scan in the "Data Refinement MS/MS" view. Press the "Show LCMS" button to automatically view the MS/MS and MS feature Scan in the LC-MS View of the current PEAKS Search result being viewed.

### Note

If a peptide feature spectrum match is associated to more than one protein as listed in the Accession column, clicking the "Protein" button will show a list of all the Accession names to select from. For detailed instructions, please refer to Chapter 8, Section 3.2, "De Novo Peptide View".



### 3.6. De Novo Only View

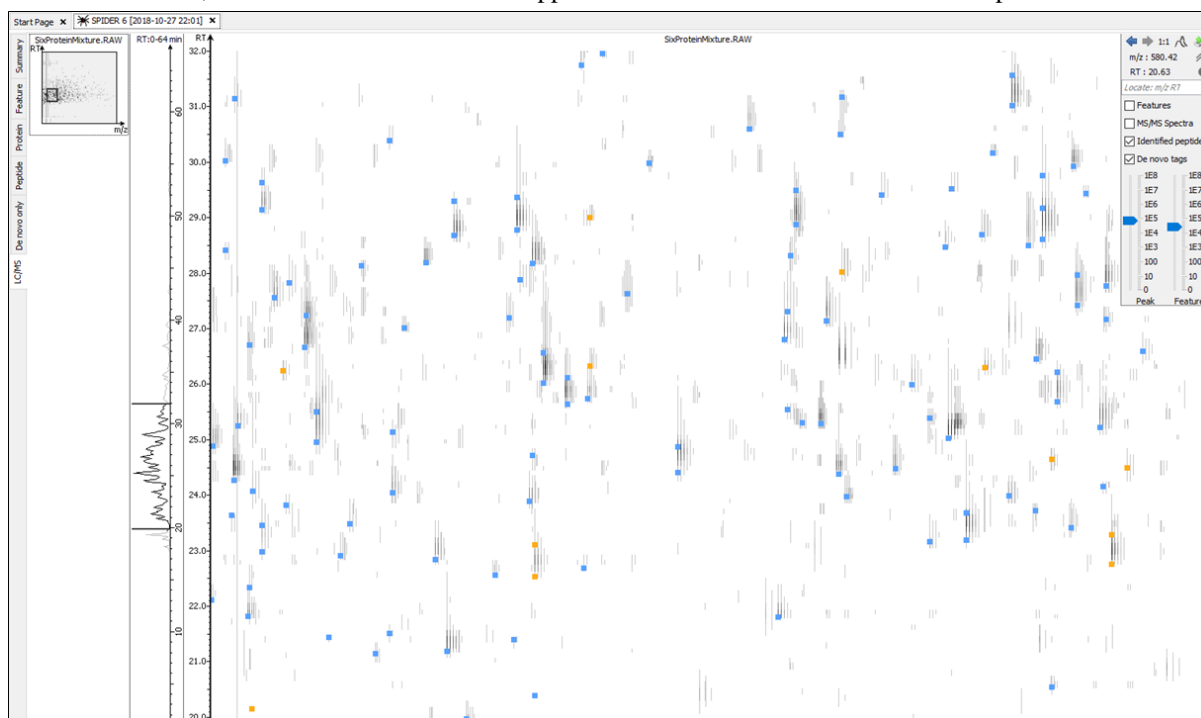
The "De novo Only View" displays the filtered *de novo* only sequences. A *de novo* sequence is *de novo* only if it is from an MS/MS spectrum that is not confidently assigned to any database peptide. The table shows the filtered *de*

*novo* only sequences by the current settings in the Summary View. "De novo only sequences may suggest novel peptides, peptides with unknown modifications, or other interesting research subjects. The interface is identical to the Peptide View in a *de novo* sequencing result node. For detailed instructions, please refer to Section 3.2, "De Novo Peptide View".

### 3.7. LC-MS View

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

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



## 4. Filtering PEAKS Search Results

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

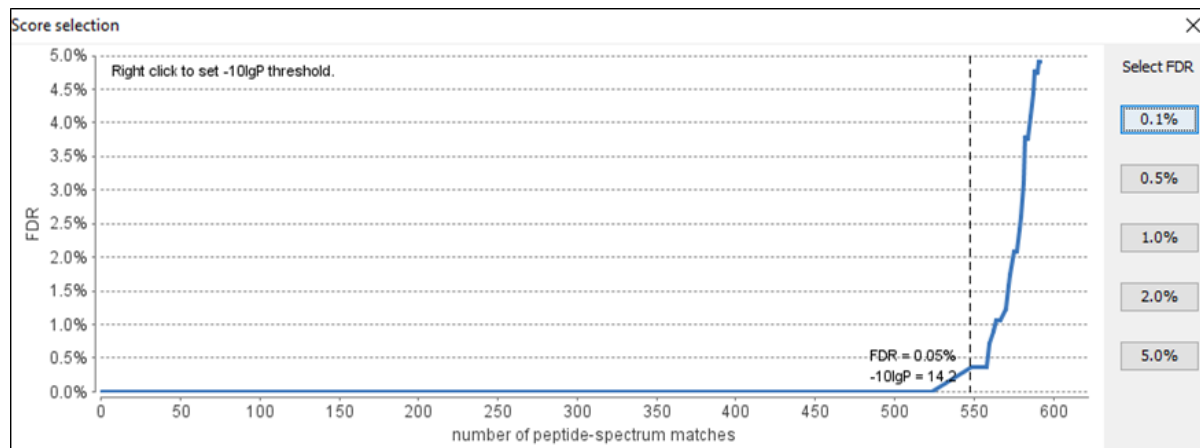
Peptides -10lgP ≥ 12	FDR	PTM Ascore ≥ 20	Mutation ion intensity ≥ 5 %	Proteins -10lgP ≥ 20 and ≥ 1 unique peptides
De novo only ALC (%) ≥ 50 and -10lgP ≤ 12	<div>Apply</div> <div>Export</div> <div>Notes</div>			

### Note

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

**Peptides.** Peptide-spectrum matches are filtered by peptide -10lgP scores. The peptide -10lgP threshold affects the results shown in both the "Peptide" and the "Protein" views. Only the peptides with PSMs above the -10lgP score threshold are listed in the "Peptide" view. These filtered peptides are used as supporting peptides to infer protein identifications; therefore, affect the proteins shown in "Protein" view.

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



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

## Note

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

**PTM.** Use the drop down menu to select **AScore** or **Ion intensity**, only modified peptides that include a modification that above this threshold will be included.

- **AScore:** Localization score assigned to modifications on the peptide. It is the -10 log of a p-value. In this case, the p-value is the probability that the modification occurs at the reported position compared to other possible positions. A -10lgP of 20 is equal to a p-value of 0.01. Scores are listed in the format AminoAcid, peptide position, modification, AScore. For example, serine at amino acid 3 is phosphorylated with an AScore of 54.01 will be reported as: S3P54.01. AScores are reported in the peptide table (Section 3.5, "Peptide View").
- **Ion intensity:** Minimum relative fragment ion intensity required for accurate localization. A pair of major fragment ions must be found showing fragmentation before and after the modified amino acid with at least the given minimum intensity. For HCD and CID fragmentation b and y ions are considered. For ETD fragmentation c and z' ions are considered. For example, if 5% is chosen and CID fragmentation is used, a pair of b or y ions must be found with at least 5% relative intensity showing fragmentation before and after the modified amino acid.

**Mutation. Ion intensity:** Minimum relative fragment ion intensity required for accurate localization. A pair of major fragment ions must be found showing fragmentation before and after the modified amino acid with at least the given minimum intensity. For HCD and CID fragmentation b and y ions are considered. For ETD fragmentation c and z' ions are considered. For example, if 5% is chosen and CID fragmentation is used, a pair of b or y ions must be found with at least 5% relative intensity showing fragmentation before and after the modified amino acid.

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



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

The ALC score threshold can be determined using Figures 3(a) and 3(b) in the Summary View. For details, please refer to Section 3.2.4, "De Novo Result Validation".



## 5. Exporting PEAKS Search Results

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


### Note

Projects can be shared by copying the PEAKS project directory. It can then be opened in the free PEAKS Viewer that has the same GUI as PEAKS Studio. Contact <support@bioinform.com> for additional details.

## 6. Running PEAKS PTM and SPIDER Separately

PEAKS PTM and SPIDER can be launched independently from a PEAKS DB result by selecting a PEAKS search result node and clicking the  button or the  button on the tool bar, respectively.

### 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 tool bar. 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 defined in PEAKS DB.

The screenshot shows the 'PEAKS PTM' window with the following settings:

- Predefined parameters:** A dropdown menu.
- Error Tolerance:** Precursor mass: 15.0 ppm using monoisotopic mass. Fragment ion: 0.02 Da.
- Enzyme:** Specified by each sample. Allow non-specific cleavage at: one end of the peptide. Maximum missed cleavages per peptide: 3.
- PTM:** Search with 313 built-in modifications (selected). Search with preferred modifications: Carboxymethyl, Deamidation (NQ), Oxidation (M). Buttons: 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%).
- Buttons:** OK, Cancel, Help.

## 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 screenshot shows the 'SPIDER Search' window with the following settings:

- Predefined parameters:** A dropdown menu.
- PTM:** Carboxymethyl, Deamidation (NQ), Oxidation (M). Buttons: Set PTM, Remove, Switch type.
- Filter:** Perform SPIDER search on spectra satisfying the following condition: De novo (ALC %) score greater than: 15 % (recommend 15%).
- Buttons:** OK, Cancel, Help.

### Note

For users of previous versions of PEAKS, SPIDER refers to the function previously called Homology Match.

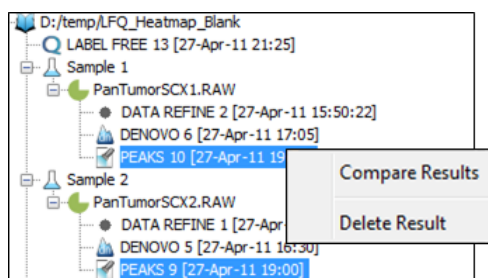
**PTM.** Clicking the Set PTM button will bring up a separate window for choosing modifications. The PTM configuration is the same as it is in *de novo* sequencing. Refer to Section 2.3, “PTM” for details.



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

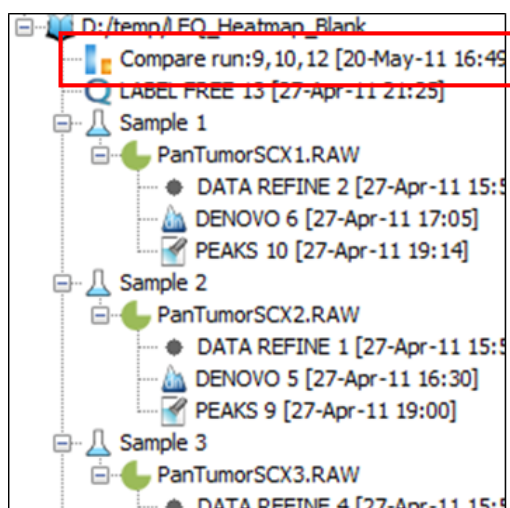
## 7. Comparison of PEAKS Results

It is possible to compare up to three search results in the same project. To compare results, hold the Ctrl key to select multiple result nodes and right-click. A popup menu will appear with the options to either Compare Results or Delete Result. Select Compare Results and PEAKS will start comparing the results.



### 7.1. Comparison Result

When the comparison analysis is completed, a comparison result node indicated as "Compare run:[project numbers]" is added to the project tree, as shown in the following screen shot.



The comparison result is opened automatically when the analysis completes. The result view consists of three main sections: peptide comparison, protein comparison, and statistical charts.

### Note

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

### 7.2. Peptide Comparison

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

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

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

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

Peptide Sequence	Cover Map	PEAKS 28 [2017-04-07 13:39]					PEAKS 29 [2017-04-07 13:39]					PEAKS 30 [2017-04-07 13:39]				
		M/Z	RT	-10logP	Z	MULTI-HIT	M/Z	RT	-10logP	Z	MULTI-HIT	M/Z	RT	-10logP	Z	MULTI-HIT
SVFPIQSYSGNSELQYVSYR	■■■	1162.5693	77.2056	74.93	2	false	1162.5612	78.8857	83.18	2	true	1162.5656	76.9528	66.54	2	false
TALNVN(+98)GSMAGTITSQLMK	■■■	1005.01	77.9901	74.68	2	false	1005.0038	79.3657	86.09	2	false	1005.0092	77.7151	66.96	2	false
APVIVQFSN(+98)GGAFFIAGK	■■■	874.47577	78.6667	74.38	2	true	874.47644	81.0035	88.28	2	true	874.47626	80.1806	67.11	2	true
LPSC(+57.02)GFSAQAVQALSAC(+...	■■■	1055.0002	67.956	74.01	2	false	1055.0004	69.6893	83.34	2	true	1054.999	67.3428	67.31	2	false
VGDYTVVAGNPGYGLGETATSGIVSALGR	■■■	1369.7029	99.185	73.67	2	true						1369.7054	99.0933	66.76	2	false
GDVLADGPSTDLGELALGQNMNR	■■■	1115.0481	81.5754	73.58	2	false	1115.046	83.3083	89.06	2	true	1115.0477	81.1737	66.32	2	true
VGDFAVAVGNPFGLGQTATSGIISALGR	■■■	1338.2137	104.2897	73.51	2	false						1338.2131	104.1433	63.7	2	true
YEEVTANC(+57.02)SC(+57.02)GN...	■■■	881.3677	30.1482	73.16	2	true	881.3686	28.9744	82.63	2	true	881.3655	29.4327	68.13	2	true
WTSQSSLDLGEPLSLITESVFAR	■■■	1268.6521	106.3125	72.59	2	true						1268.6447	106.1687	63.63	2	true
GPVEN(+98)DGFSPVEAIAAGVIER	■■■	1084.5527	85.1375	72.56	2	true	1084.5474	86.4413	84.74	2	false	1084.5515	84.6505	65.89	2	true
AGTHDVHGAALGAEEVAATR	■■■	937.978	31.5578	72.32	2	true	937.985	30.1532	90.09	2	true	937.98	31.2476	65.61	2	true
LPQVEGTGGDVQPSHELVR	■■■	1009.52313	43.7807	72.29	2	true	1009.5223	43.222	87.31	2	true	673.34985	42.7703	52.36	3	false

## 7.3. Protein Comparison

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

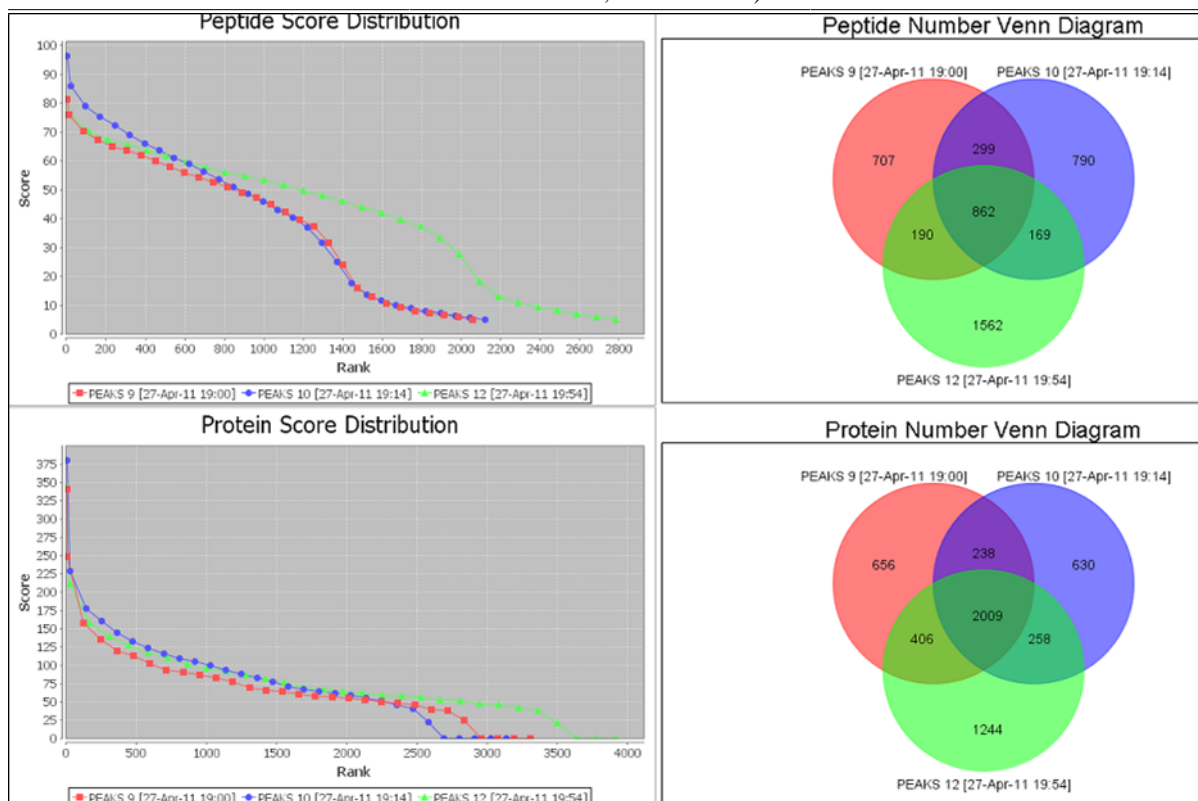
- **Score:** PEAKS protein -10lgP score
- **#Spec:** The number of PSMs matched to supporting peptides for the protein.
- **#Pep:** The number of supporting peptides for the protein
- **#Uniq:** The number of unique supporting peptides for the protein
- **%Spec:** The number of peptides detected for the protein over the number of peptides in theoretical digestion.
- **%Cov:** The sequence coverage of the protein

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

Protein_ID	Cover Map	PEAKS 28 [2017-04-07 13:39]						PEAKS 29 [2017-04-07 13:39]						PEAKS 30 [2017-04-07 13:39]					
		Score	#Spec	#Pep	#Uniq	%Spec	%Cov	Score	#Spec	#Pep	#Uniq	%Spec	%Cov	Score	#Spec	#Pep	#Uniq	%Spec	%Cov
A0A1C0F2L5 A0A1C0F2L5_SERMA	■■■	28.79	1	1	1	1.06	0.8	61.87	1	1	1	1.06	1.71	43.65	2	2	2	2.13	1.94
A0A1C0CIR9 A0A1C0CIR9_SERMA	■■■	24.36	1	1	1	1.28	1.25												
A0A1C0FH9 A0A1C0FH9_SERMA	■■■	33.27	1	1	1	2.04	1.26							51.96	2	2	2	4.08	3.61
A0A1C0DTA5 A0A1C0DTA5_SERMA	■■■	26.26	1	1	1	1.43	1.27							21.12	1	1	1	1.43	1.27
A0A1C0CA44 A0A1C0CA44_SERMA	■■■	21.01	1	1	1	1.28	1.34							25.19	1	1	1	1.15	1.42
A0A1C0NS42 A0A1C0NS42_SERMA	■■■	36.89	1	1	1	1.15	1.42	64.18	2	2	2	2.3	3.28	25.19	1	1	1	1.15	1.42
A0A1C0FUM1 A0A1C0FUM1_SERMA	■■■	24.74	1	1	1	1.09	1.42	41.1	2	2	2	2.17	3.41	23.93	1	1	1	1.09	2.13
A0A1C0KFA8 A0A1C0KFA8_SERMA	■■■	44.42	2	2	2	1.96	1.44	59.08	3	3	3	2.94	2.55	62.76	5	4	4	3.92	3.48
A0A1C0KU04 A0A1C0KU04_SERMA	■■■	36.53	1	1	1	1.52	1.48							37.73	1	1	1	1.52	1.48
A0A1C0GX91 A0A1C0GX91_SERMA	■■■	23.55	1	1	1	0.79	1.64	69.91	2	2	2	1.57	2.24	45.36	2	2	2	1.57	2.24
A0A1C0GSF2 A0A1C0GSF2_SERMA	■■■	23.51	1	1	1	2.22	1.65	25.42	1	1	1	2.22	1.65	27.06	1	1	1	2.22	1.65

## 7.4. Statistical Charts

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



## 7.5. Exporting Comparison Result

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

**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

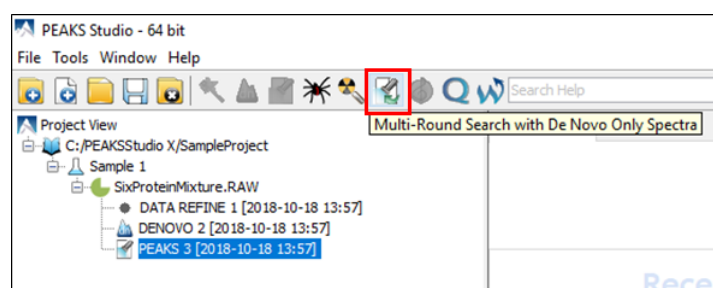
The Multi-Round Search is an additional search function that follows a completed search result to examine the remaining “de novo only” spectra. This additional search allows users to perform PEAKS DB, PTM and SPIDER searches using different parameters and/or to integrate a different database.

### Note

In a poster presented by Bioinformatics Solutions Inc. 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: Quantitative analysis of endogenous peptides.

Follow these steps to perform a multi-round search:

1. Select a PEAKS DB, PTM or SPIDER result node in the "Project View". Open the result and set filters. The *de novo* results found in the "de novo only" tab will be carried forward to the Multi-Round Search.
2. To open the search parameters window, select a PEAKS DB, PEAKS PTM or SPIDER result node in the "Project View". Launch the multi-round search by clicking the  icon on the tool bar.



3. Specify the search parameters in the pop-up dialogue. The parameter selection dialogue is similar to the parameter selection dialogue in Chapter 9, *Peptide, PTM, and Mutation Identification (PEAKS DB, PEAKS PTM, and SPIDER)* Section 2, “Setting PEAKS Search Parameters”.
4. Wait for the analysis to finish. New result nodes for multi-round PEAKS DB, PEAKS PTM, and SPIDER will appear in "Project View". Double-click the result nodes to examine the associated results.

# Chapter 11. PTM Profiling

## 1. Overview

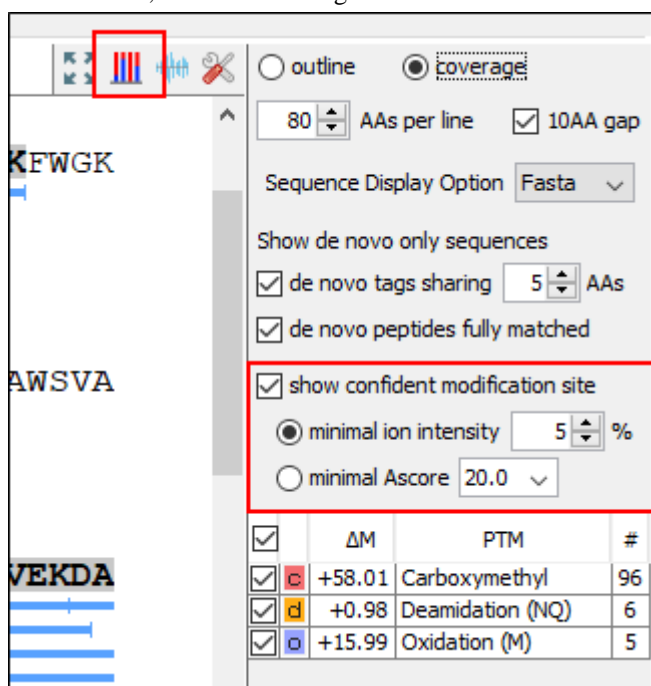
The PTM Profile tool provides a direct visualization and summary of the quantitative information such as the abundance of modified and unmodified forms containing the detected modification sites on proteins. The PTM Profile tool can compare the changes in the expression level of the modification sites by comparing the area under the XIC curve of the modified peptide features from the LC-MS data across the MS samples in the experiment with the peptide feature area of the unmodified peptides.

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.

Accession	-10lgP	Cc
Proteins		
P02769 ALBU_BOVIN	372.77	
Q29443 TRFE_BOVIN	338.82	
B1J0T5 BGAL_ECOLC	323.17	
P00722 BGAL_ECOLI	323.17	
P00330 ADHI1_YEAST	248.36	

2. Set minimal ion intensity or minimal AScore parameter threshold for confident modification site determination. Refer to Chapter 9, *Peptide, PTM, and Mutation Identification (PEAKS DB, PEAKS PTM, and SPIDER)* Section 3.4.4, “Protein Coverage” for details.

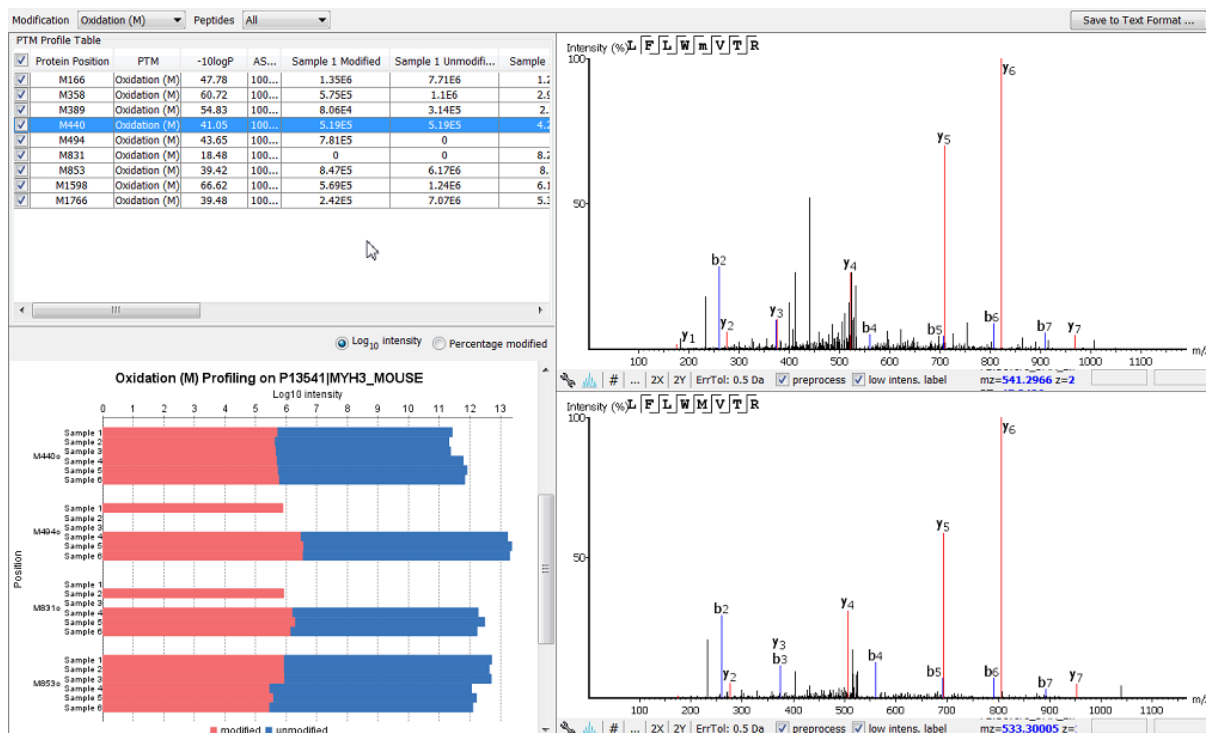


	$\Delta M$	PTM	#
<input checked="" type="checkbox"/>	+58.01	Carboxymethyl	96
<input checked="" type="checkbox"/>	+0.98	Deamidation (NQ)	6
<input checked="" type="checkbox"/>	+15.99	Oxidation (M)	5

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.

## 2. PTM Profiling for DB/PTM/SPIDER Results

For the Identification results (DB/PTM/SPIDER), the PTM Profiling tool provides quantitative information of modified peptides compared with unmodified peptides for the modification sites of the protein across all MS samples.




## 2.1. Input Selection

The inputs for the profiling can be set using the drop-down lists at the top-left corner of the dialog.



- **Proteins.** The accession ID of the selected protein to be analyzed for PTM profiling is shown in the title of the dialog.

### 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.

- **Modification.** The drop-down list contains the detected and selected confident PTMs for the protein. Select the type of modification to study a specific PTM or choose **All the PTMs** to visualize the profiling information of all PTM sites at together.

### Note

The fixed PTMs are not used for profiling.

- **Peptides.** Data source used for the analysis. **All** considers all the peptides for this protein. When **Fully digested** is selected, only the full enzyme digested peptides will be used to calculate the modification abundance.

## 2.2. PTM Profile Table

PTM Profile Table shows the detailed quantitative information for the modification sites. Each row in the table shows the modified and unmodified area for each sample of the selected PTM at each detected modification site. The table can be sorted by any column. The following list describes the contents in each column:

- **(Check box):** Allow the selection of each table row. Only the marked rows are used for the PTM Profile Chart.

- **Protein Position:** The detected modification sites in the protein. The modified amino acid symbol together with the position in the protein sequence is shown.
- **PTM:** The type of PTM being profiled.
- **-10logP:** The best -10logP identification score for the corresponding modified peptides.
- **AScore:** The best PTM AScore associated with the corresponding PTM at the specified site of the modified peptides.
- **Sample Modified:** Total area of the features associated with the corresponding modified peptides for each sample.
- **Sample Unmodified:** Total area of the features associated with the corresponding unmodified peptides for each sample.

PTM Profile Table							
<input checked="" type="checkbox"/>	Protein Position	PTM	-10logP	AScore	Sample 1 Modified	Sample 1 Unmodifi...	Sample 2 Modifi
<input checked="" type="checkbox"/>	M166	Oxidation (M)	47.78	1000.00	1.35E6	7.71E6	1.22E6
<input checked="" type="checkbox"/>	M358	Oxidation (M)	60.72	1000.00	5.75E5	1.1E6	2.98E6
<input checked="" type="checkbox"/>	M389	Oxidation (M)	54.83	1000.00	8.06E4	3.14E5	2.3E5
<input checked="" type="checkbox"/>	M440	Oxidation (M)	41.05	1000.00	5.19E5	5.19E5	4.24E5
<input checked="" type="checkbox"/>	M494	Oxidation (M)	43.65	1000.00	7.81E5	0	0
<input checked="" type="checkbox"/>	M831	Oxidation (M)	18.48	1000.00	0	0	8.29E5
<input checked="" type="checkbox"/>	M853	Oxidation (M)	39.42	1000.00	8.47E5	6.17E6	8.5E5
<input checked="" type="checkbox"/>	M1598	Oxidation (M)	66.62	1000.00	5.69E5	1.24E6	6.13E5
<input checked="" type="checkbox"/>	M1766	Oxidation (M)	39.48	1000.00	2.42E5	7.07E6	5.33E5

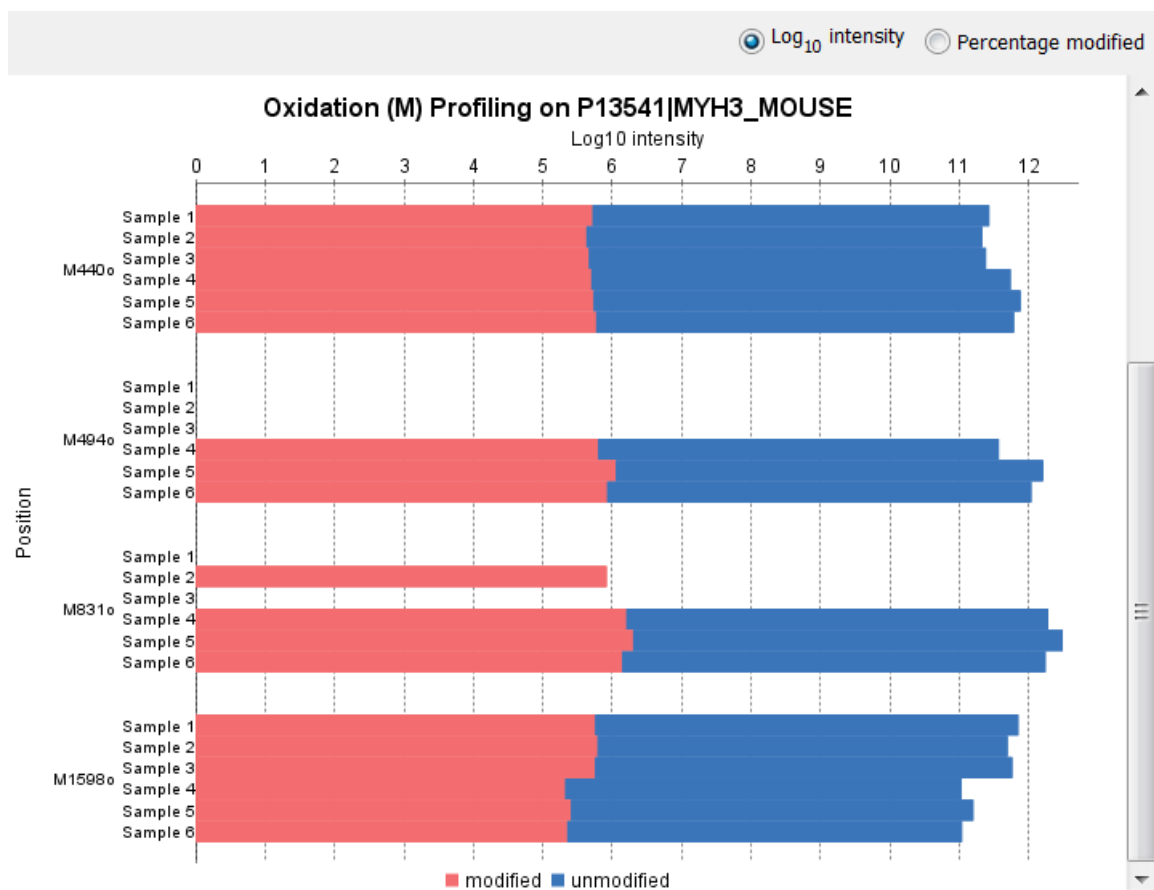
When **Fully digested** is selected, there are two more columns shown in the PTM Profile Table:

- **Peptide:** The fully digested peptide sequence.
- **Peptide Position:** Position of the modified amino acid in the peptide sequence.

## 2.3. PTM Profile Chart

The PTM Profile Chart is located at the bottom left corner in the dialog. It gives a visualization of the intensity or ratio between unmodified and modified amino peptides at each detected modification sites.

- **Log10 intensity.** When **Log10 intensity** is selected, the chart shows the intensities for modified and unmodified peptides at the modification sites in all MS samples. The x-axis indicates the Log10 intensity whereas the y-axis indicates the position in the protein sequence where a modification was identified for each individual sample. Blue bar represents the unmodified area and red bar represents the modified area.



- **Percentage modified.** When **Percentage modified** is selected, the chart shows the ratio between modified and unmodified peptides at each modification site for all MS samples. The x-axis indicates the percentage (%) of modified or unmodified peptide feature areas relative to the total abundance whereas the y-axis indicates the position in the protein sequence where a modification was identified for each individual sample.

### Note

Since the percentage is calculated over the sum of the modified and unmodified areas, the modified and unmodified portion always sum up to 100%.

### Note

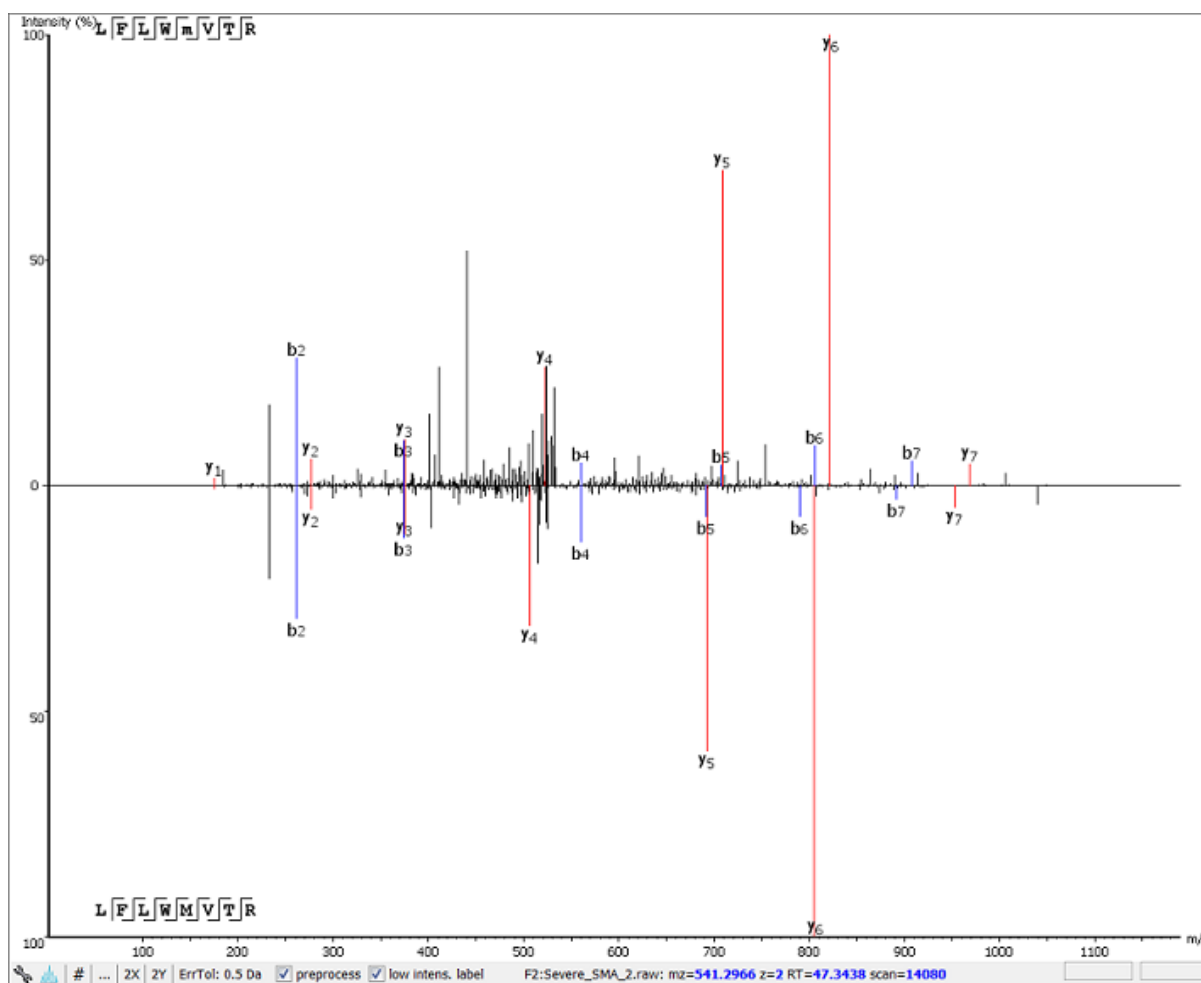
For samples with the specific PTM not detected, the corresponding bars are left as blank.

## 2.4. Spectrum Annotation View

The Spectrum Annotation View provides a graphical representation for the best modified peptide and the best unmodified peptide when **All** is selected from the **Peptides** drop-down list. The top spectrum corresponds to the best modified peptide, whereas the bottom spectrum shows the best unmodified peptide.

When **Fully digested** is considered, the best modified peptide and the best unmodified peptide are displayed in a compare view for quick visualization and validation.





## 2.5. Result Export

The PTM profile can be exported as an image as well as in text format.

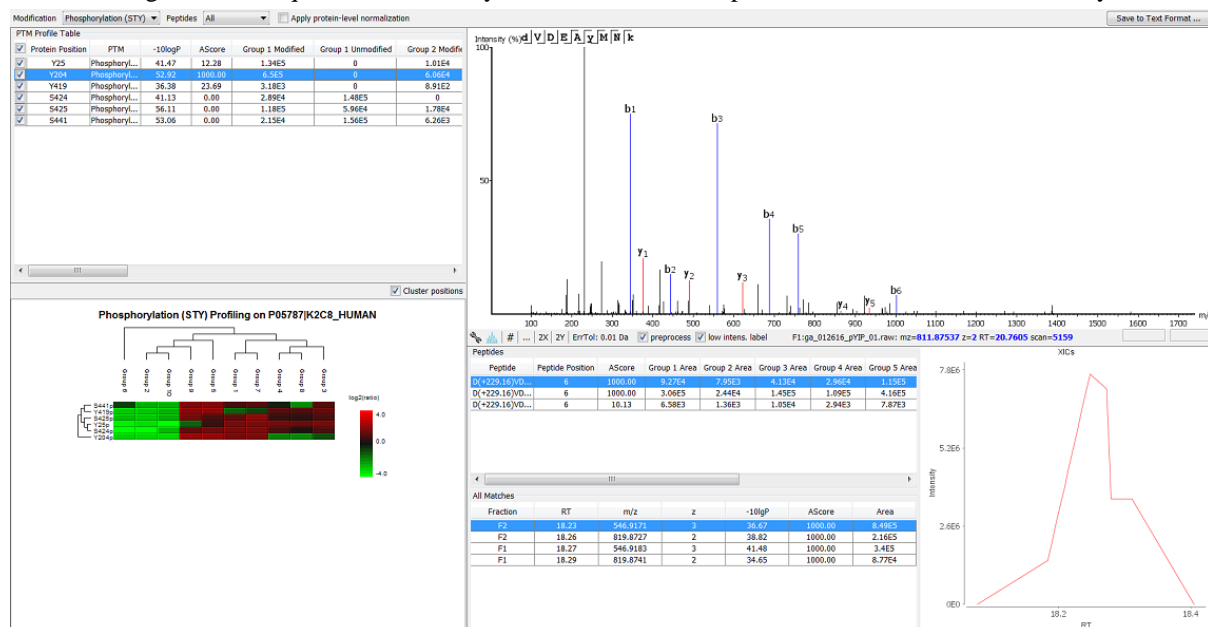
- **Image export.** To save the profile chart, right click and choose **Save to Image...** menu item to save the displayed graph as a PNG file.
- **Text format export.** Click **Save to Text Format...** button at the top-right corner to export the calculated PTM Profile data to a text file in CSV format. The same export function is available from the profile chart by right clicking and selecting **Save to Text Format...** menu item. The PTM Profile of the selected modification on the selected protein will be exported by default. In addition, profiles on all proteins for the selected modification may also be exported, if **Export ptm profile for all proteins** are checked. The exported file contains all necessary information to construct the PTM profile graph. The PTM profile data are grouped together by protein, then by modification type, and then by the the modification site on the protein sequence. The headers of the CSV column are similar to the PTM Profile Table and also self-explanatory. The following items are explained for clarity:
  - **Modified Peptide Sequence Window:** protein sequence containing 10 AA ahead of the modification site and 10 AA after the modification site with padding as necessary.
  - **Sample Best AScore:** best AScore for each sample.

### Note

When "Export PTM profile for all proteins" is not checked, only the profiling data that is marked in the PTM Profile Table will be exported.

### 3. PTM Profiling for Label Free Quantification and Reporter Ion Quantification Results

This section describes the PTM Profiling for Label Free Quantification (LFQ) and Reporter Ion Quantification (e.g. TMT/iTRAQ) results. PTM Profiling for these two types of Quantification results have similar interfaces. PTM Profiling for a TMT quantification analysis is used as an example to demonstrate the functionality.



#### 3.1. Input Selection

The input selection is similar to the input section of the PTM profiling on DB result, except the additional option **Apply protein-level normalization**. By default, the protein-level normalization is not applied. If checked, the displayed modified/unmodified area is obtained by dividing the true modified/unmodified area by the corresponding ratio of the selected protein.

#### Note

In addition to the fixed PTMs, the labeling modifications are not considered for profiling.

#### 3.2. PTM Profile Table

PTM Profile Table shows the peptide feature areas for modified and unmodified peptides for the selected PTM in all groups. The group area is the sum of the sample/channel areas in that group. If no grouping is defined, sample areas for modified and unmodified peptides are displayed.

#### Note

For TMT results, the peptide area in each channel is calculated by distributing the associated peptide feature area to each channel based on channel intensities.

#### 3.3. PTM Profile Heat Map

The relative abundance of modified peptides in all groups for the selected modification sites as a heat map. The color scheme is determined by the log2 ratios of the modified area of each group relative to the average abundance. For more details about the heat map, please refer to Chapter 15, *PEAKS Q - Label Free* Section 3.2.2, "Protein Volcano Plot".

### 3.4. Detail Pane

The right panel in the dialog provides detailed information for the selected profile data in the PTM Profile Table.

- **Spectrum View.** The Spectrum View provides a graphical representation of the best modified peptide.
- **Peptides Table.** Peptides table lists the modified peptides that contain the selected PTM at the specified modification site.
- **All Matches Table.** The supporting PSMs of the selected peptide in **Peptides Table** are listed in the **All Matches Table**. The headers are self-explanatory. The last column **Area** is the area of the feature that is associated with the corresponding PSM.
- **XICs.** It shows the XICs figure for the selected modified match in the **All Matches Table**.

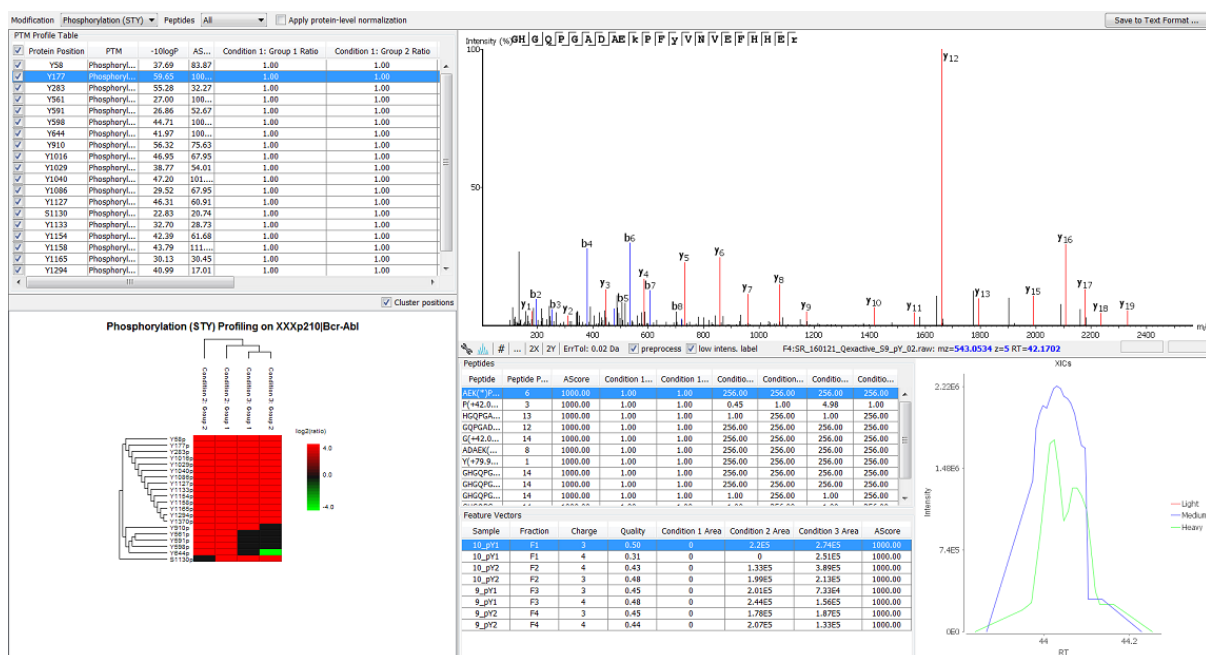
### 3.5. Result Export

Similar to DB PTM Profiling, the result can be exported as an image as well as in the text format. In the exported CSV file, the following items are explained for clarity:

- **Group Modified:** total area of the features associated with the corresponding modified peptides for each group.
- **Group unmodified:** total area of the features associated with the corresponding unmodified peptides for each group.
- **Group modified (normalized):** total area of the features associated with the corresponding modified peptides for each group, after applying the protein level normalization.
- **Group unmodified (normalized):** total area of the features associated with the corresponding unmodified peptides for each group, after applying the protein level normalization.
- **Ratio Group modified:** the ratio of the group modified area with respect to the average of the modified area for all groups.
- **Ratio Group modified (normalized):** the ratio of the group modified area with respect to the average of the modified area for all groups, after applying the protein level normalization.

## 4. PTM Profiling for Precursor Ion Quantification Results

This section describes the PTM Profiling for Precursor Ion Quantification (e.g. SILAC/ICAT) Results. The SILAC PTM Profiling focuses on the changes in abundance of the modified peptides among different conditions for each group. If no groups are set, the PTM abundance is compared among different conditions for each sample. The interface is similar to the TMT/Label Free Quantification PTM Profiling.



## 4.1. PTM Profile Table

The PTM Profile Table displays the profiling results for the selected modification. For Super SILAC result, the PTM profiling studies the changes in ratio of abundance of modified peptides for each condition with respect to the reference condition in each group. The ratio of the abundance for all condition with respect to the reference condition in each sample is calculated first. Then the median of the condition ratios of samples in a group is taken as the condition ratio of that group.

## 4.2. PTM Profile HeatMap

The relative abundance of the modified peptides that contain the selected PTM at the selected modification sites is represented as a heat map at condition level for each group. The color scheme is determined by the log2 ratio of each condition with respect to the reference condition for each group. The reference condition ratio is not shown for simplicity. For more details about the heat map, please refer to Chapter 15, *PEAKS Q - Label Free* Section 3.2.2, "Protein Volcano Plot".

## 4.3. Detail Pane

The detail pane provides detailed peptide information of the selected PTM site.

- **Spectrum View.** The Spectrum View provides a graphical representation of the best modified peptide.
- **Peptides Table.** Peptides Table lists the modified peptides that contain the selected PTM at specified modification site. The ratios of modified peptide abundance in each condition compared with that in reference condition are shown in the table.
- **Feature Vectors Table.** The supporting feature vectors of the selected peptide in the **Peptides Table** are listed in the **Feature Vectors Table**.
- **XICs.** It shows the XICs figure for the selected feature vector in the **Feature Vectors Table**.

## 4.4. Result Export

Similar to PTM Profiling for DB result, the SILAC profiling result can be exported as an image as well as in the text format. In the exported CSV file, the following items are explained for clarity:

- **Condition: Sample modified area:** total area of the features associated with the corresponding modified peptides in this condition for a certain sample.

- **Condition: Sample unmodified area:** total area of the features associated with the corresponding unmodified peptides in this condition for a certain sample.
- **Condition: Sample modified(normalized):** total area of the features associated with the corresponding modified peptides in this condition for a certain sample, after applying the protein level normalization,
- **Condition: Sample unmodified (normalized):** total area of the features associated with the corresponding unmodified peptides in this condition for a certain sample, after applying the protein level normalization,
- **Condition: Group Ratio:** the modification abundance ratio of each condition with respect to the reference condition for each group.
- **Condition: Group Ratio (normalized):** the modification abundance ratio of each condition with respect to the reference condition for each group, after applying the protein level normalization.

---

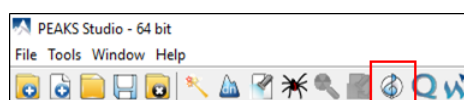
# Chapter 12. Combining Multiple Database Search Engines with PEAKS inChorus

## 1. PEAKS inChorus Overview

Properly combining the results from different database search engines can enhance the accuracy and sensitivity of peptide identifications. PEAKS inChorus is a tool for combining PEAKS results with search results from other engines. PEAKS inChorus can either launch a database search or import a search result. The other supported search engines are SEQUEST (Proteome Discoverer versions 1.3 and 1.4), Mascot (v2.5), 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 dialogue, there are drop-down menus for each supported database search engine. Using the drop-down menus, you can choose the search results to be combined. You can either:

- **Choose an existing result**

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

- **Import a result from a file**

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

- **Launch a new search**

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

## Combining Multiple Database Search Engines with PEAKS inChorus

**PEAKS** Please select  
Please select  
New Run ...  
Summary PEAKS 3

**Mascot** Please select

**X!Tandem** Please select

**Sequest** Please select

**OMSSA** Please select

Engine	Precursor Error	Fragment Error	Enzyme	Database	Missed Cleavages	PTM
--------	-----------------	----------------	--------	----------	------------------	-----

⚠ PEAKS is not selected, inchorus combination will not be performed.

OK Cancel

### Note

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

### Note

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

### Important

To enable the calculation of the inChorus FDR, the same target-decoy database needs to be searched by all of the engines. For PEAKS, X!Tandem, and OMSSA, this target-decoy database is generated automatically. For SEQUEST and Mascot, a target-decoy database needs to be exported from PEAKS and added to their database list. To export the target-decoy database, open the Configuration tool and select Database from the left-hand menu. The Export Decoy DB button is located just under the list of configured databases.

Combining Multiple Data-  
base Search Engines  
with PEAKS inChorus

Enzyme  
PTM  
Labeled Q Method  
**Database**  
Instrument

Database List

[UniProt\_SwissProt]  
**[NCBI\_nr]**

Database Details

FASTA format database: NCBI nr Validate Database **Export Decoy DB**

Basic Options

Database name:  ☐ Validated

Path:  Browse or Download

☐ EST database

Advanced Options - Fasta Title Format

Rule to parse accession/id from FASTA title:

Rule to parse description from FASTA title:

Accession/id URL:  
 <Accession/ID>

Delimiter:

Taxonomy Options

taxonid  Browse Download

taxdmp  Browse Download

New Add/Update Delete Help

Close

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



## Combining Multiple Database Search Engines with PEAKS inChorus

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 # <sup>13</sup>C: 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

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

**Peptide Score.** The first noticeable difference is that the inChorus peptide score is not the  $-10\lg P$  score used in PEAKS DB. Instead, a percentage confidence score is used to reflect the probability that this peptide-spectrum match is correct. The percentage score is calculated in accordance with the empirical calculation used in PeptideProphet<sup>TM</sup> (Keller et al. Anal. Chem. 2002, 74:5383-92).

1. The method of PeptideProphet<sup>TM</sup> is applied to each engine's result to estimate the probability of correctness for each peptide identification.
2. If a peptide-spectrum match is identified by multiple engines, the scores for all those engines are added up to a weighted sum.
3. The weighted sum scores of all peptides are converted to a probability by using the PeptideProphet<sup>TM</sup> method again.

**Protein Score.** The protein score is also a percentage score. It is calculated using the following approach. The scores of the peptides from a protein are added up to a weighted sum. The PeptideProphet<sup>TM</sup> 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

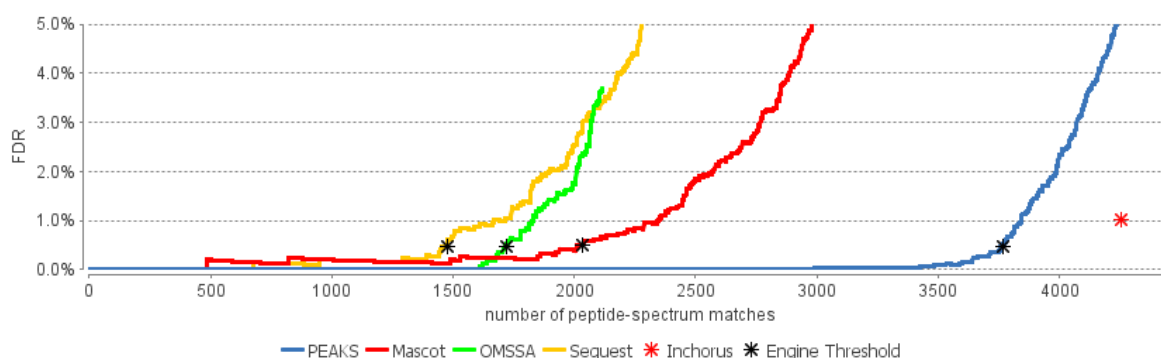
Also, the following appearances of the icon indicates:

- **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.

Summary										Engines						
Peptide										PEAKS	Sequest	Mascot	X!Tandem	Omssa	PTM	
1	AVARAYN(+98)TTK	533	365.8648	24.04	1094.5720	0.6	2	99.96	P 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 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 M - O	45.42	1.33	42.05	-	-		
4	LVAETEDRK	465	354.1932	23.07	1059.5559	1.7	3	99.92	P 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 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 - O	44.77	-	37.32	-	-		
7	Q(-17.03)AKHEEIDTK	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	PGGGQVEVK	561	435.7363	24.38	869.4607	-3.0	1	99.38	P S M X	29.06	0.90	21.06	7.1E-2	-		
10	NPTAFK	142	403.2322	17.77	804.4494	0.5	1	99.25	P S M -	20.29	0.91	35.37	-	-		
11	AALAAARDSK	405	516.2780	22.17	1030.5408	0.6	2	99.16	P S M -	20.95	0.83	32.96	-	-		
12	VISQDTQPHQQK	557	704.8610	24.33	1407.7107	-2.3	1	97.93	P - - - -	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.

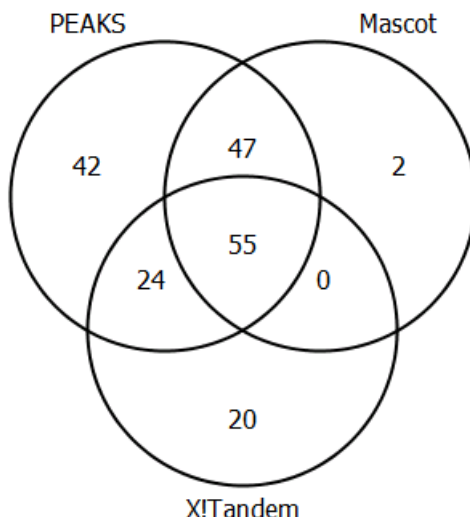
**FDR curves.** The FDR curves for all selected search engines are placed together in a single figure in the "Summary" view. The black \* marks on the search engines' FDR curves indicate the score thresholds of the corresponding engines. The red \* mark indicates the inChorus FDR threshold.



## 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.

The screenshot shows the control panel for filtering results. It includes the following elements:

- Peptides** section: Target inChorus FDR: 1% (with an **Edit...** button).
- Proteins** section: Score (%) ≥ 50 and ≥ 0 unique peptides.
- De novo only** section: ALC (%) ≥ 50.
- Buttons: **Apply**, **Export**, and **Notes**.

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

The screenshot shows the 'Set individual engine Threshold' dialog box. It has two radio buttons: 'Set target inchorus FDR:' (set to 1%) and 'Set individual engine Threshold' (selected). Below the selected option, it states 'At least one of the following conditions is true:' and lists five conditions with checkboxes and input fields:

- ☒ PEAKS -10lgP ≥ 11.4 [FDR]
- ☒ Sequest Xcorr ≥ 3.7 [FDR]
- ☒ Mascot Score ≥ 33.5 [FDR]
- ☒ X!Tandem e-value ≤ 2.4E-2 [FDR]
- ☒ Omssa e-value ≤ 9.89E-1 [FDR]

Buttons at the bottom: **OK** and **Cancel**.

### 4. Exporting inChorus Results

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


---

# Chapter 13. Precursor Ion Quantification (e.g. SILAC)

## 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. The ratio is calculated from the intensities of those peaks. PEAKS Q supports both user-defined labels and commercial quantification labels. PEAKS supports the analysis of Super-SILAC experiments that uses a mixture of SILAC-labeled cells as a spike-in standard for accurate quantification of other unlabeled or labeled samples.

Precursor Ion Quantification can be launched from either the Project Wizard work flow, or on an existing PEAKS DB result. When running quantification on an existing result, ensure that the specified isotopes are as **variable** PTMs when performing the database search. After the database search is complete, follow these steps:


1. Select a PEAKS DB result node in the "Project View" frame. Click the PEAKS Q icon, , from the tool bar or select the Tools > Quantification command from the menu.

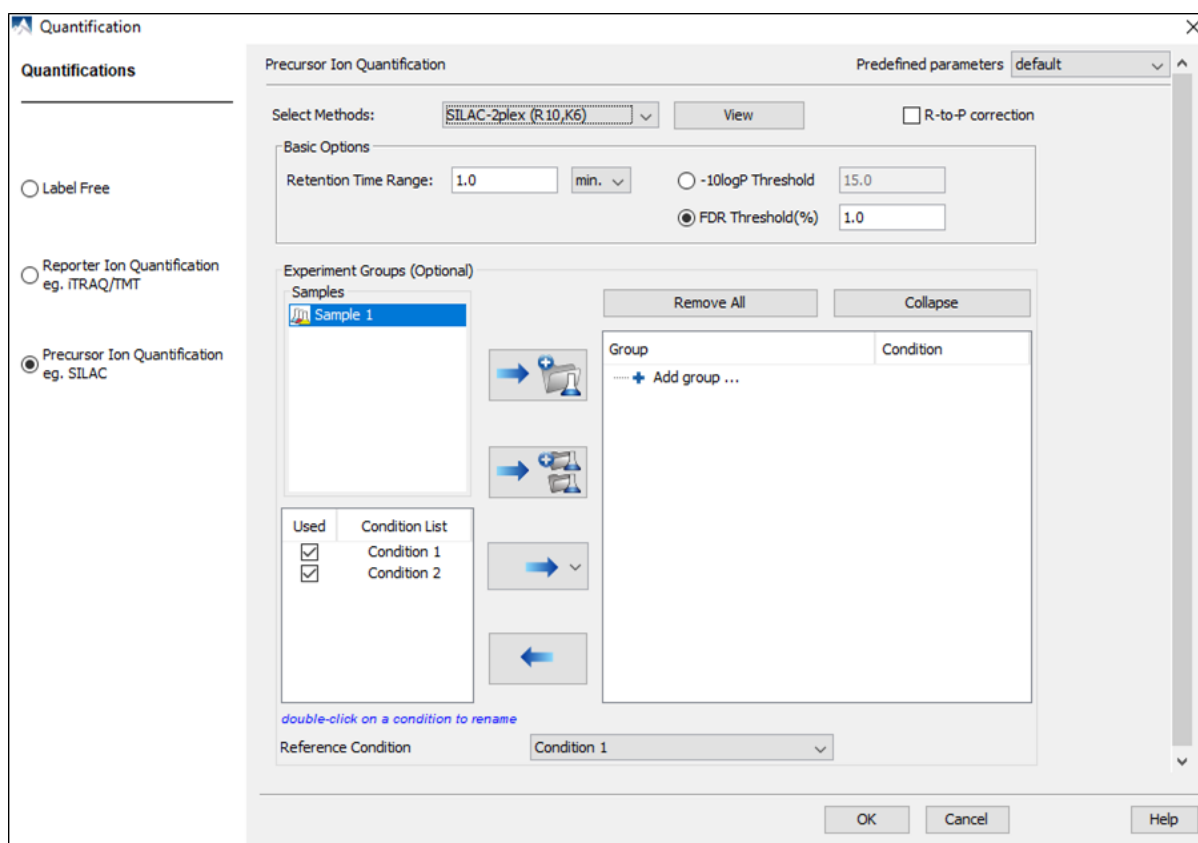
### Note

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

2. Select the quantification protocol, "Precursor Ion Quantification eg. SILAC" and specify the PEAKS quantification parameters in the dialogue box on the right; and click OK.
3. Wait for the analysis to finish. A new quantification result node, , will appear in the "Project View". Double click the node to open the result.

## 2. Setting Parameters

The following parameter dialogue appears when the quantification tool bar icon, , is clicked or the Tools > Quantification command from the menu is clicked. Select "Precursor Ion Quantification eg. SILAC" under the "Quantifications" options.



From the Select Methods drop-down menu, choose the appropriate quantification method used in experiment. To examine the details of a selected quantification method, click on the **View** button. To create a new quantification method, select **Add New Method** from the Select Methods drop-down menu. Refer to Chapter 13, *PEAKS Q - MS Level* Section 2.3, “Labeled Q Method Configuration” for how to create or edit a quantification method.

**R-to-P correction:** This will tell PEAKS to consider the possibility that free heavy Arginine(R) residues metabolically converted to Proline(P) residues before insertion into the peptide. The areas of both the normal heavy-labeled Arginine peptide feature and the heavy-labeled Proline feature will be summed to calculate the complete area of the heavy-labeled peptide feature.

## Note

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




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

- **Retention Time Range:** The retention time range is the maximum retention time difference between paired peptide features.
- **-10logP Threshold:** Only peptides with an identification score above this threshold are used for quantification.
- **FDR Threshold(%):** Only peptides with an identification score above this threshold are used for quantification.

## Note


Either -10logP or FDR threshold can be selected

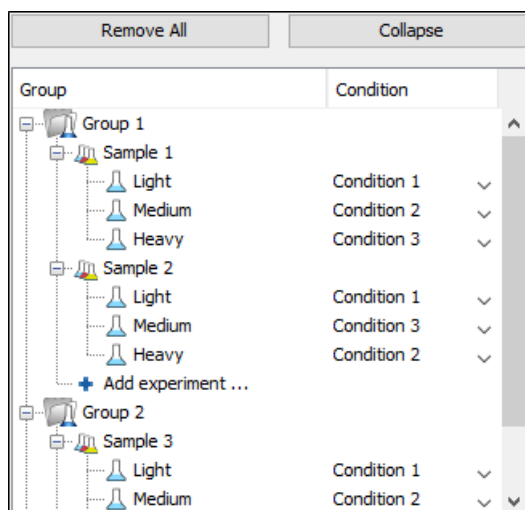
Quantification allows for the option of sample grouping. The samples can be grouped into one or more groups according to the quantification experiment design.

- **Samples:** The MS samples are listed for selection. After selecting a sample, it can be added to a new group by clicking the  button or to an existing group by clicking the  button and selecting that option from the drop-down menu. To add multiple samples into their own separate groups, click the  button.
- **Conditions:** The number of conditions depends on the quantification method selected. For example, *SILAC-2plex(R10,K8)* has 2 conditions and *SILAC-3plex(R10,K8|R6,K4)* has 3 conditions to study. Each label in a sample represent one unique condition. To remove a label from analysis, uncheck the corresponding condition. Double-click on a condition name to rename.

## Note

At least 2 conditions must be selected for quantitative analysis.

- **Groups:** The samples can be arranged into one or more groups. A group represents a state of the experiment where quantitative ratios between conditions are investigated. Replicate MS samples should be put together in one group. Double-click on the group node to rename the group. Display all samples in each group or just the individual groups by selecting the **Expand/Collapse** buttons, respectively. Samples can be removed either individually or all at once by clicking the  button or the **Remove All** button, respectively.



## Note

Same condition can not be assigned to more than one label of a Sample.


## Note

If no group is defined, then all samples are considered as replicates which is equivalent to put all samples in one group.

- **Reference Condition:** Relative abundance ratios at feature, peptide and protein level for each condition are calculated using the corresponding reference condition area as denominator of the ratio calculations. The reference condition can be chosen from the selected conditions only.

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

## 3. Understanding the Results

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

### 3.1. Summary View

The "Summary View" provides three main functions:

1. **Result Filtration:** This is achieved by specifying the filtration rules in the area at the top of the Summary View. The filtration function is discussed in Section 3.1.1, "Result Filtration".
2. **Summary Report:** Several statistical charts assist the user to get an overall picture of the results, assess the result quality, and examine the reliability and reproducibility of the experiment. This function is the focus of this section and is discussed in Section 3.1.2, "Summary Report".
3. **Result Exporting:** This is achieved by clicking the "Export" button at the top of the Summary View. The exporting function is discussed in Section 3.1.3, "Result Exporting".

The Summary View includes an expression profile with candidate proteins assorted in a heat map, result statistics in both figures and tables, and a list of filtration parameters. To add a summary note, click on the **Notes** button at the top-right of the Summary View to open a "Notes Entry" editor in which you can edit the notes. These notes will be displayed in the Summary View.

#### 3.1.1. Result Filtration

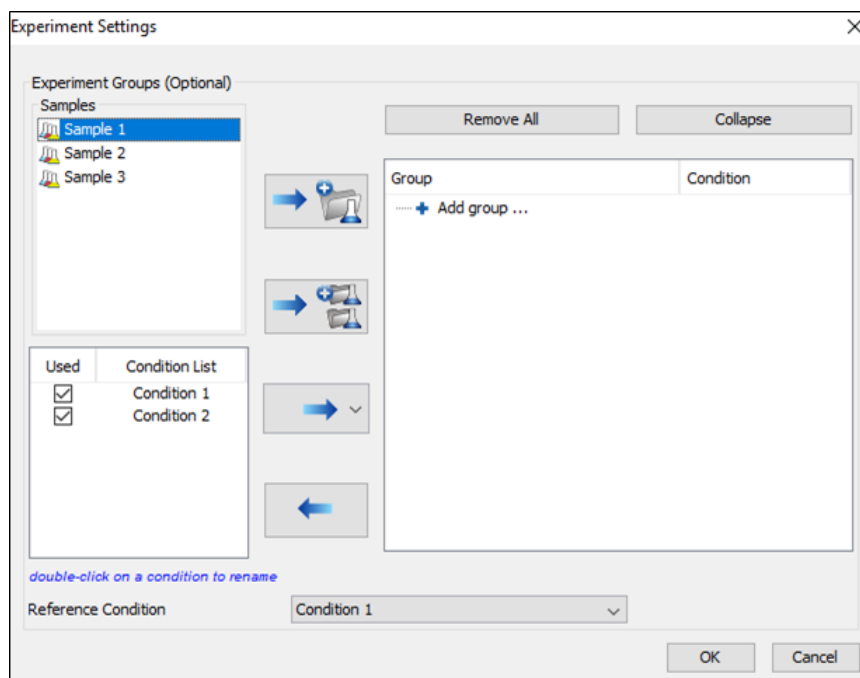
PEAKS provides an extensive set of tools to filter Precursor Ion Quantification results according to the users' requirements. There are three categories of options in the "Summary" view: **Experiment**, **Feature Vector Filter**, and **Protein Filter**.

<b>Experiment</b>	All experiments, No groups, No normalization	Experiment Settings	Normalization	Apply	Export	Notes
<b>Feature Vector Filter</b>	-10LgP $\geq$ 16.9, Quality $\geq$ 0, Avg. Area $\geq$ 0E0, 1 $\leq$ Charge $\leq$ 10, Peptide Hit Id Count $\geq$ 1, detected in at least 1 feature label					Edit...
<b>Protein Filter</b>	Significance $\geq$ 0, Fold change $\geq$ 1, Protein quantified in $\geq$ 0 sample, has at least 1 unique peptide					Edit... Heat Map Fold Changes

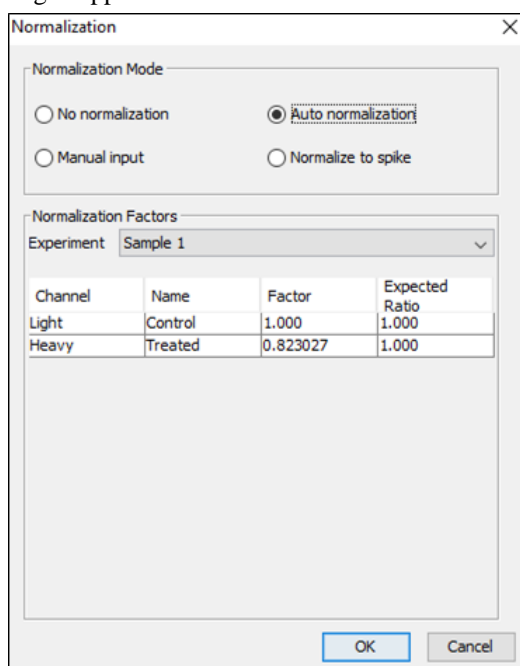
##### 3.1.1.1. Experiment

There are two buttons under this heading – **Experiment Settings** and **Normalization** – which allow for control over the analysis.

The following settings dialogue appears when the **Experiment Settings** button is selected. The MS samples can be arranged in groups according to experiment design. For details on grouping, please refer to Section 2, "Setting Parameters".



The following normalization dialogue appears when the **Normalization** button is selected.



## Note

**No normalization** is chosen as the default option.

- **Auto normalization:** This option will calculate a global ratio from the total intensity of all labels. Factor values and Expected Ratios between these labels will be presented. It is assumed that the total protein content is same in different labels as depicted by 1s in the 'Expected ratio' columns. Another sample can be selected from the **Experiment** drop-down menu to display the results of another sample. Normalized abundance is calculated from the raw abundance divided by the normalization factor.
- **Manual input:** This option performs the same default calculation as Auto normalization but also allows for the manual change of each label's Expected Ratios. After changing a ratio and hitting Enter on the keyboard,



PEAKS will recalculate the Factor for the non-referenced label(s). The Experiment drop-down menu can be opened to display the expected ratio and normalization factors for that sample.

- **Normalize to spike:** This option displays a list of identified proteins, which can be selected for normalization factor calculation. A search bar is included to enable the quick location of selected proteins. Right-clicking in the Protein screen will open a context menu that will allow for selection of the highlighted proteins or the removal of all check-marked proteins. The expected ratios can be manually modified. It is assumed that the total protein content of the selected proteins in different labels have the ratios entered in the 'Expected ratio' column.

**Normalization**

Normalization Mode

☐ No normalization ☐ Auto normalization

☐ Manual input ☒ Normalize to spike

Normalization Factors

Experiment: Sample 1

Channel	Name	Factor	Expected Ratio
Light	Control	1.000	1.000
Heavy	Treated	0.848605	1.000

Protein List

search

Prev Next

Protein

P06733|ENOA\_HUMAN ☒

Q98VA1|TB82B\_HUMAN ☒

P13639|EF2\_HUMAN ☒

P68371|TB84B\_HUMAN ☒

Q13885|TB82A\_HUMAN ☒

P06733|ENOA\_HUMAN ☒

P49321|NASP\_HUMAN ☒

P60709|ACTB\_HUMAN ☐

P11142|HSP7C\_HUMAN ☐

P10809|CH60\_HUMAN ☐

P08238|HS90B\_HUMAN ☐

P13010|XRCC5\_HUMAN ☐

P63261|ACTG\_HUMAN ☐

P55060|XPO2\_HUMAN ☐

P49321|NASP\_HUMAN ☐

3 proteins selected

(Right click to show popup menu)

OK Cancel

Select **Apply** within the filtration window of the “Summary” view to include all Experiment Settings and Normalization choices in the analysis.

### 3.1.1.2. Feature Vector Filter

The following parameter dialogue appears when the **Edit...** button is selected.

**Feature Vector Filters**

Feature Vector Filters

-10LgP  $\geq$  11.7

Quality  $\geq$  0

Ave. area  $\geq$  0E0

Charge between 1 and 10

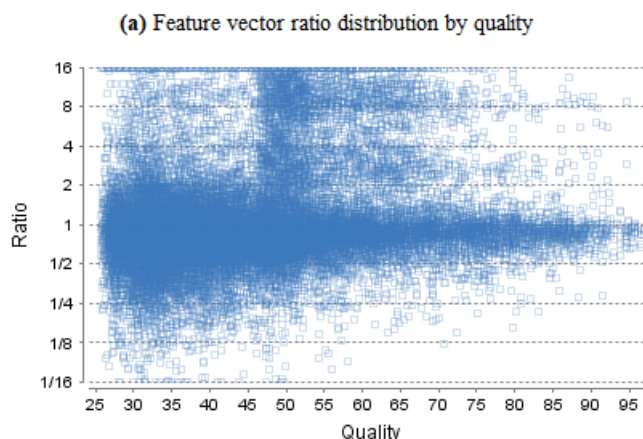
Reference label present ☐

Number of Labels with Id Present  $\geq$  0

Number of labels with feature present  $\geq$  1

OK Cancel

- **-10LgP:** Only peptides with an identification score above this threshold are used for quantification. PEAKS Q will not allow the user to set a value less than the value selected when setting up the search. Click the **FDR** button to set a -10LgP value equivalent to a selected FDR.
- **Quality:** A higher quality peptide indicates that the peptide is more quantifiable. Factors that affect the quality score include m/z difference, RT difference, XIC shape similarity, and the feature intensities. An appropriate threshold value can be selected by checking the ratio-quality figure.



- **Avg. area:** This filter is directly applied to the MS signal intensity. Higher intensity peptide features have been shown to be more reproducible across replicates. The rule of thumb is to choose the average area near a fold change of 8, as shown in this labeled version of Figure 3(b) of the Summary View. Only peptides with at least one label above this threshold will be used in protein quantification.
- **Charge between:** Only peptides with feature vectors that all fall within this range will be used in protein quantification.
- **Reference label present:** Only quantifiable peptides in which the reference label is present will be used in protein quantification.
- **Number of labels with Id present:** The quantifiable peptides that contain at least the number of labels selected with an Id attached will be used in protein quantification.

## Note

Feature vectors are aligned across different samples to find the missing features in a sample even if there is no peptide Id attached to the feature vector in that sample. To see the aligned feature vectors with no Id attached, set this filter threshold to '0'.

- **Number of labels with feature present:** The quantifiable peptides that contain at least the specified number of labels with detected feature will be used in protein quantification.

Select **Apply** within the filtration window of the Summary View to include all selected filters in the analysis.

### 3.1.1.3. Protein Filter

The following filter dialogue appears when the **Edit...** button is selected. These filters determine what will be included in a protein group significance calculation, as well as what significance score thresholds will be set. For a protein group to have a significance score, it must have been identified in a sample/label that has already been placed into a group (in either the Tools > Quantification window or in Experiment Settings).

**Protein Filters**

☒ Significance  $\geq$  0

☐ FDR (adjusted p-value)  $\leq$  1 %

Fold Change  $\geq$  1

Protein quantified in  $\geq$  1 samples

Unique Peptides  $\geq$  1

Modified Form Exclusion ☐

OK Cancel

- **Significance:** Only protein groups with a significance above this threshold will be listed in the “Protein” view. The significance score is calculated as the  $-10\log_{10}$  of the significance testing p-value. Paired T-test is used for significance calculation when all samples are put in a single group. The ANOVA method is used for significance calculation when there are more than one group in the experiment settings. A significance score threshold of 20 is recommended, which equals to a significance testing p value of 0.01. Either this or “FDR (adjusted p-value)” can be selected to set a significance threshold. Scores will only be calculated for protein groups identified from samples that are in groups.

### Note

If no group is defined, then all samples are considered replicates which is equivalent to put all samples in one group.

### Note

For statistical significance calculation, at least two samples must be included in each group.

### Note

For more details on ANOVA, refer to the following paper: “*On the comparison of several mean values: an alternative approach*” *Biometrika*. 1951, 38(3/4): 330-336.

- **FDR:** Adopt Benjamin–Hochberg method to adjust the p-value to the false discovery rate for all protein groups that have already passed the other filters. Only protein groups with significance scores passing the calculated FDR will be listed in the “Protein” view. Either this or “Significance” can be selected to set a significance threshold.
- **Fold Change:** For analysis with only one group, the fold change refers to the protein group ratio between any two conditions. For multiple group analysis, the fold change refers to the protein group ratios between any two groups. Only protein groups at or above this fold change threshold will be listed in the “Protein” view.
- **Protein quantified in # of samples:** when one group or no group is used, this filter will be shown and it requires that the protein must be quantified in at least # of samples to be included in the “Protein” view.
- **Unique Peptides:** Only protein groups with at least this many unique peptides will be listed in the “Protein” view.
- **Modified Form Exclusion:** Peptides that are modified with a variable PTM will not be included in the significance calculation for a protein group. This does not apply to the quantification labels. The expression level of modified peptides might be different than the overall expression level of a protein. Therefore, including modified peptides in the protein ratio calculation might lead to incorrect results. By checking this box, if a peptide has modified and unmodified forms, both will not be used as candidates, if a peptide only has one modified form, it is still a candidate, but if a peptide has more than one modified form, all forms will not be used as candidates.

Select Apply within the filtration window of the Summary View to include all selected filters in the analysis.

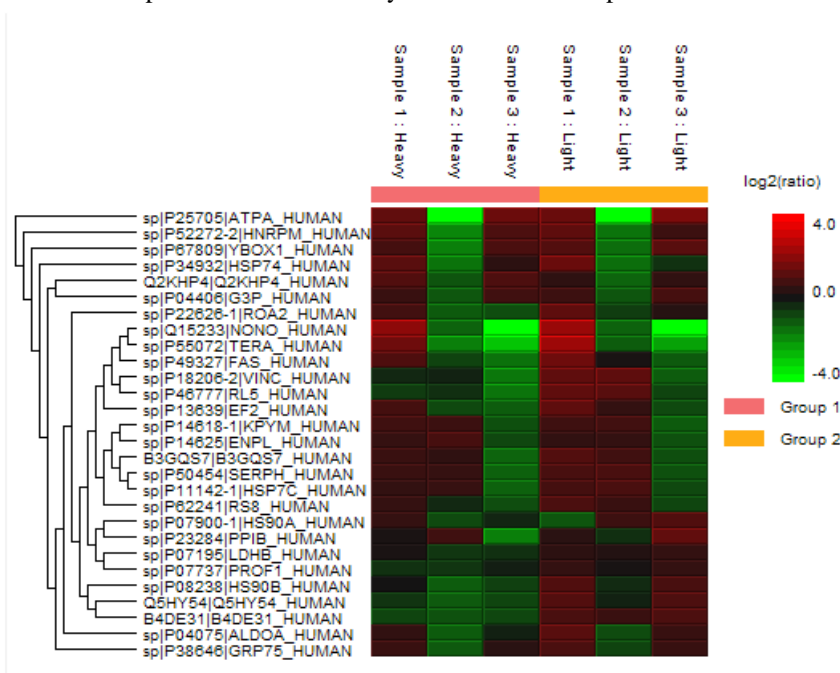
## Note

To view all quantifiable proteins, i.e. Do not use significance and fold change filters, choose Significance and set its threshold at 0, and set fold change at 1.

### 3.1.2. Summary Report

#### 3.1.2.1. Heat Map

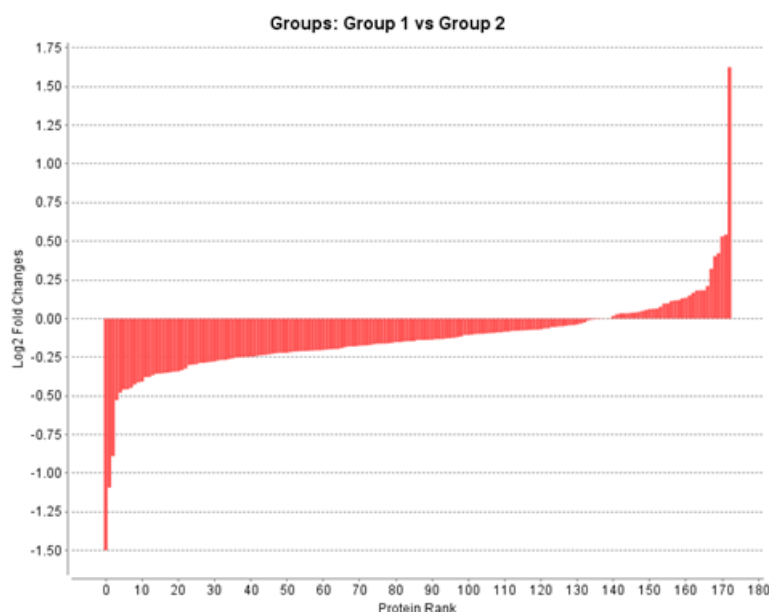
The heat map displays the relative protein abundance in different conditions for the protein groups that passed the filters. The relative protein abundance in each condition with respect to the abundance in reference condition is represented as a heat map of the representative proteins of each protein group. These representative proteins are clustered if they exhibit a similar expression trend across the samples. The hierarchical clustering is generated using a neighbour-joining algorithm with a Euclidean distance similarity measurement of the log<sub>2</sub> of the protein ratios of each sample (for one group) or each group (for multiple groups) relative to the average ratio. Similarly, the conditions in different samples are clustered if they exhibit a similar expression trend across the protein groups.



An interactive protein profile heat map is opened in a new window when the **Heat Map** button that is located in the filtration options. Right-click on the heat map in this window to save a copy of it.. In the interactive heat map (by clicking the Heat Map button), move the cursor over any point to highlight the protein accession name and the sample/label to which it matches. Right-click on the heat map and click **Save as...** to export a PNG file of the heat map.

#### 3.1.2.2. Interactive Expression Wide Fold Change Graph

The following bar graph displays relative log<sub>2</sub> ratio between two selected conditions in samples with respect to the rank of each protein. The log<sub>2</sub> function applies to the selected primary group condition with respect to the selected secondary group condition. Hold the cursor over any bar in this figure to display the associated fold change of each protein within the rank. Select the drop-down menus to change the primary and secondary groups. Click **Apply** to set the new log<sub>2</sub> ratio. Click **Save as...** to export a PNG file of the bar graph.



### 3.1.2.3. Volcano Plot

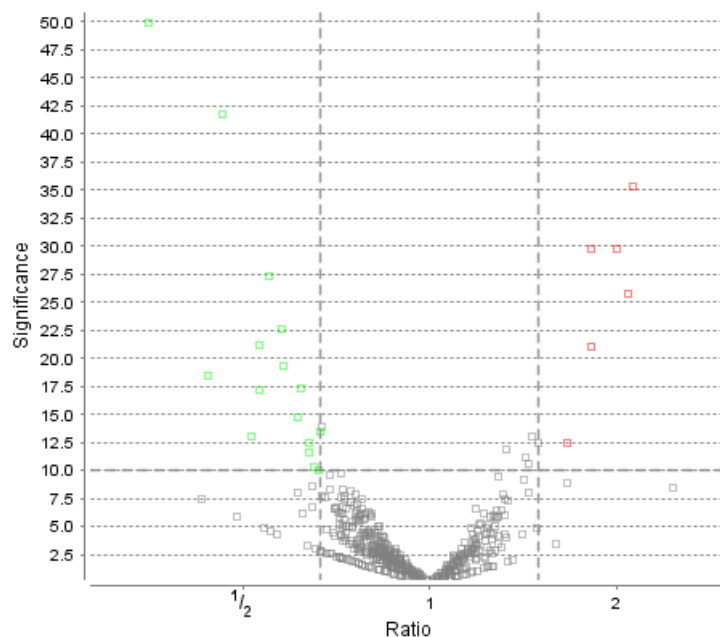
A volcano plot combines a statistical test with the magnitude of the change enabling quick visual identification of those data-points that display large-magnitude changes and that are also statistically significant. For each condition defined in the experiment settings, a volcano plot in Figure 2 plots significance versus fold-change of the quantified proteins for that condition compared to the reference condition.

- **Horizontal broken grey line:** The selected significance threshold
- **Vertical broken grey lines:** The selected fold change thresholds

### Note

No volcano plot is generated if the samples are arranged in more than one group.


Plotting points in this way results in two regions of interest in the plot: those points that are found towards the top of the plot that are far to either the left or the right side. These represent values that display large magnitude fold changes (hence being left or right of center) as well as high statistical significance (hence being towards the top). In this figure, the ratio is the condition ratio set with respect to the reference condition.

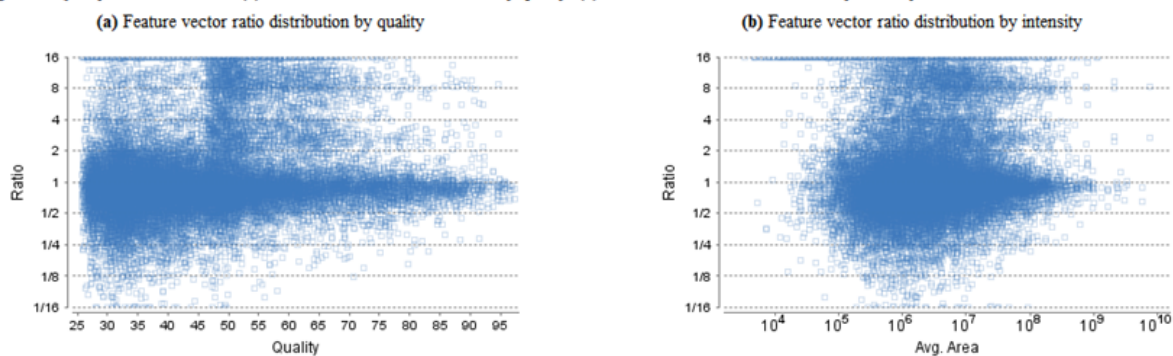


The “Protein View” contains an interactive volcano plot that can be hidden or displayed. Markers for the proteins that are above the set significance threshold will be displayed in color. Scrolling the cursor over a marker will highlight it and, subsequently, clicking on this marker will highlight the protein in the interface.

#### 3.1.2.4. Feature Vector Ratio Distribution

Figures 3(a) and 3(b) plot the distribution of feature vector ratios by quality and by average area, respectively. These plots help to determine the threshold values for the quality and intensity filters to obtain a reliable section of the analysis results.

**Figure 3.** Quality control statistics. (a) Feature vector ratio distribution by quality. (b) Feature vector ratio distribution by intensity. 



#### 3.1.2.5. Summary Tables

Tables 1 and 3 list the data and the filtered results, respectively. Table 2 lists the feature vector and protein result filtration parameters.

**Table 1.** Statistics of data.

# of MS Scans	31468
# of MS/MS Scans	106231

**Table 2.** Result filtration parameters.

-10LogP	≥ 15
Average area	≥ 1E4
Charge	[1, 3]
Quality	≥ 2
Reference label present	true
Number of labels present	1
Protein significance	≥ 0
Protein fold change	≥ 1
Unique peptides	≥ 3
Modified From Exclusion	false
Significance method	PEAKSQ

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

Feature vector groups	3538
Feature vectors	9699
Features	23021
Peptide Sequences	7155
Protein Groups	370

Tables 4 and 5 list the quantification and instrument parameters, respectively.

**Table 4.** Quantification parameters.

Quantification Type: ICAT/SILAC
Quantification Mass Tolerance: 15.0ppm
Quantification RT Range: 0.5min.
-10LogP Threshold: 15.0
Reporter Ion Type: MS2
R to P Convert enabled: true
Light: R (100.0%) K (100.0%)
Heavy: R(+10.01) (100.0%) K(+8.01) (100.0%)
Group Settings in Parameters:
Group 1: Sample 1 : Heavy
Group 1: Sample 2 : Heavy
Group 1: Sample 3 : Heavy
Group 2: Sample 1 : Light
Group 2: Sample 2 : Light
Group 2: Sample 3 : Light
DB Dependency: 5
Current Group Information:
Group 1: Sample 1 : Heavy
Group 1: Sample 2 : Heavy
Group 1: Sample 3 : Heavy
Group 2: Sample 1 : Light
Group 2: Sample 2 : Light
Group 2: Sample 3 : Light

**Table 5.** Instrument parameters.

Fractions: 20120812_ShAr_QE1_ARL_NX4_mix_sample_1.raw, 20120812_ShAr_QE1_ARL_NX4_mix_sample_2.raw, 20120817_ShAr_QE1_ARL_NX4_mix_sample_2_REP_3.raw
Ion Source: ESI(nano-spray)
Fragmentation Mode: CID, CAD(y and b ions)
MS Scan Mode: FT-ICR/Orbitrap
MS/MS Scan Mode: Linear Ion Trap

### 3.1.3. Result Exporting

PEAKS precursor ion quantification results can be exported to various HTML and text formats by clicking the Export button within the filtration options in the Summary View. The Summary View and the images from various views can also be exported to various HTML and text formats. Refer to Chapter 19, *Exporting Data/Reports and Printing* Section 5, “Exporting Quantification Results”, for details.

## 3.2. Protein View

The "Protein View" lists all the quantified proteins present in the sample, characterizes each protein at the amino acid level, and lists the supporting peptide features of each protein. It has four components:

- **Protein Table:** This lists all the quantified proteins present in the sample.
- **Volcano Plot:** The Volcano Plot plots significance versus fold-change of the quantified proteins for that condition compared to the reference condition.
- **Coverage:** This characterizes the protein sequence at the amino acid level. All the PTMs and mutations occurring on the protein sequence will be displayed together with the MS/MS data supporting the inference.
- **Peptides:** This lists the peptides identified from the selected protein in the Protein table.



### 3.2.1. Protein Table

The quantified proteins are listed in the protein table with homologous proteins grouped together. This protein table is the same as the other Protein tables found in the PEAKS DB, PEAKS PTM, and SPIDER results, but includes the following additions:

- **Significance:** The protein significance is set according to the rules set in the Summary View. Refer to Section 3.1.1.3, "Protein Filter".
- **Condition Profile:** The relative protein abundance in each condition among the samples are depicted as heat maps. Hold the cursor on a profile to view the samples, and ratios with respect to the reference condition specified in Experiment Settings of the Summary View.
- **Condition Ratio:** For a protein, the relative support peptide abundance in each condition is compared to the reference condition to obtain the peptide condition ratio in each sample separately. In each sample, the median of the support peptide condition ratios is the protein condition ratio in that sample. Finally, the protein condition ratio in a group is computed by calculating the median from the protein condition ratios in the samples in that group.
- **#Unique:** The number of quantifiable unique peptides of the protein.
- **Unique Coverage:** The protein coverage by the unique peptides only.



See Chapter 9, *Peptide, PTM, and Mutation Identification (PEAKS DB, PEAKS PTM, and SPIDER)* Section 3.4, “Protein View”, “Protein Table” for more details on other entries of the Protein table.

### 3.2.2. Protein Table View Filter

The quantifiable protein list can be further filtered by protein view filters. These filters change the protein list by adding or removing proteins based on the filter parameters and does not change the contents of other tabs or the protein information such as ratio. For details, please refer to Chapter 9, *Peptide, PTM, and Mutation Identification (PEAKS DB, PEAKS PTM, and SPIDER)* Section 3.4.1, “Protein View Filters”.

### 3.2.3. Volcano Plot


The Volcano Plot plots significance versus fold-change of the quantified proteins for that condition compared to the reference condition. When viewed in the Protein View, the selected protein from the Protein Table will be highlighted as blue solid square in the volcano plot. Users can easily select protein markers from the volcano plot to jump to that protein's information in the Protein Table. Refer to Section 3.1.2.3, “Volcano Plot” for details.

### 3.2.4. Protein Coverage

The Coverage tab characterizes the selected protein sequences at the amino acid level. See Chapter 9, *Peptide, PTM, and Mutation Identification (PEAKS DB, PEAKS PTM, and SPIDER)* Section 3.4.4, “Protein Coverage”, “Protein Coverage” for details.

### 3.2.5. PTM Profile

SILAC PTM Profile studies the changes in abundances of peptides with confidently identified PTMs in each condition with respect to the reference condition. Peptide feature areas of the modified peptides are used for this

comparison. Click on  to open PTM Profile dialog. Please refer to Chapter 11, *PTM Profiling* Section 4, “PTM Profiling for Precursor Ion Quantification Results” for details.

### 3.2.6. Supporting Peptides

The supporting peptides assigned to the protein are shown under the “Peptides” tab.

Coverage Peptides											
	Peptide	Used	Unique	-10lgP	Quality Score ↓	Condition 2 : Ratio Profile	Condition 1_Ratio	Condition 2_Ratio	#Vector	AScore	Start End
1	<input checked="" type="checkbox"/> R.C(+57.02)TGGEVGATSLAPK(*) <sub>1</sub>	<input checked="" type="checkbox"/>	•	89.15	2.79		1.00	0.92	3		17 31
2	<input checked="" type="checkbox"/> K.IGPLGLSPK(*) <sub>2</sub>	<input checked="" type="checkbox"/>	•	47.77	2.51		1.00	0.98	3		32 40
3	<input checked="" type="checkbox"/> K.EILGTASQSVGC(+57.02)NVDGR(*) <sub>3</sub>	<input checked="" type="checkbox"/>	•	81.53	2.11						
4	<input checked="" type="checkbox"/> R.QAQIEVPSASALIK(*) <sub>4</sub>	<input checked="" type="checkbox"/>	•	74.18	1.91						
5	<input checked="" type="checkbox"/> R.ELSGTIK(*) <sub>5</sub>	<input checked="" type="checkbox"/>	•	33.37	1.49						
6	<input checked="" type="checkbox"/> K.LTIQNR(*) <sub>6</sub>	<input checked="" type="checkbox"/>	•	24.39	1.30						
7	<input checked="" type="checkbox"/> K.VVYLR(*) <sub>7</sub>	<input checked="" type="checkbox"/>	•	31.92	1.28						
8	<input checked="" type="checkbox"/> K.K(*)VGDDIAK(*) <sub>8</sub>	<input checked="" type="checkbox"/>	•	35.47	1.06		1.00	1.08	1		41 48
9	<input checked="" type="checkbox"/> K.VGDDIAK(*) <sub>9</sub>	<input checked="" type="checkbox"/>	•	28.30	1.00		1.00	0.25	3		42 48
10	<input checked="" type="checkbox"/> K.HSGNITFDEIVNIAR(*) <sub>10</sub>	<input checked="" type="checkbox"/>	•	71.71	0.96		1.00	0.40	2		100 114

This table is similar to the peptide table in the Peptide View.

- **(Checkbox):** Select the peptides to appear within the Coverage Tab as blue bars.
- **Used:** Select whether this peptide is used to calculate the protein profile. Select one and right click on the column to show the pop-up menu. Use "Reset the used selection" to reset the used setting for all proteins.
- **Unique:** This shows which peptides are only assigned to the current protein group.
- **Quality Score:** The quality score of the peptide based on m/z difference, RT difference, XIC shape similarity and feature intensities. The higher the quality of a peptide, the more quantifiable it is.
- **Condition Profile:** The relative peptide abundance for each condition compared to the reference condition among the samples is depicted as a heat map. Hold the cursor on a profile to view the samples, the areas, and the ratios with respect to the reference condition.
- **Condition Ratio:** The ratio columns shows the relative peptide abundance for each condition with respect to the reference condition for all groups. A peptide's abundance in each condition is the sum of its feature vector(s)

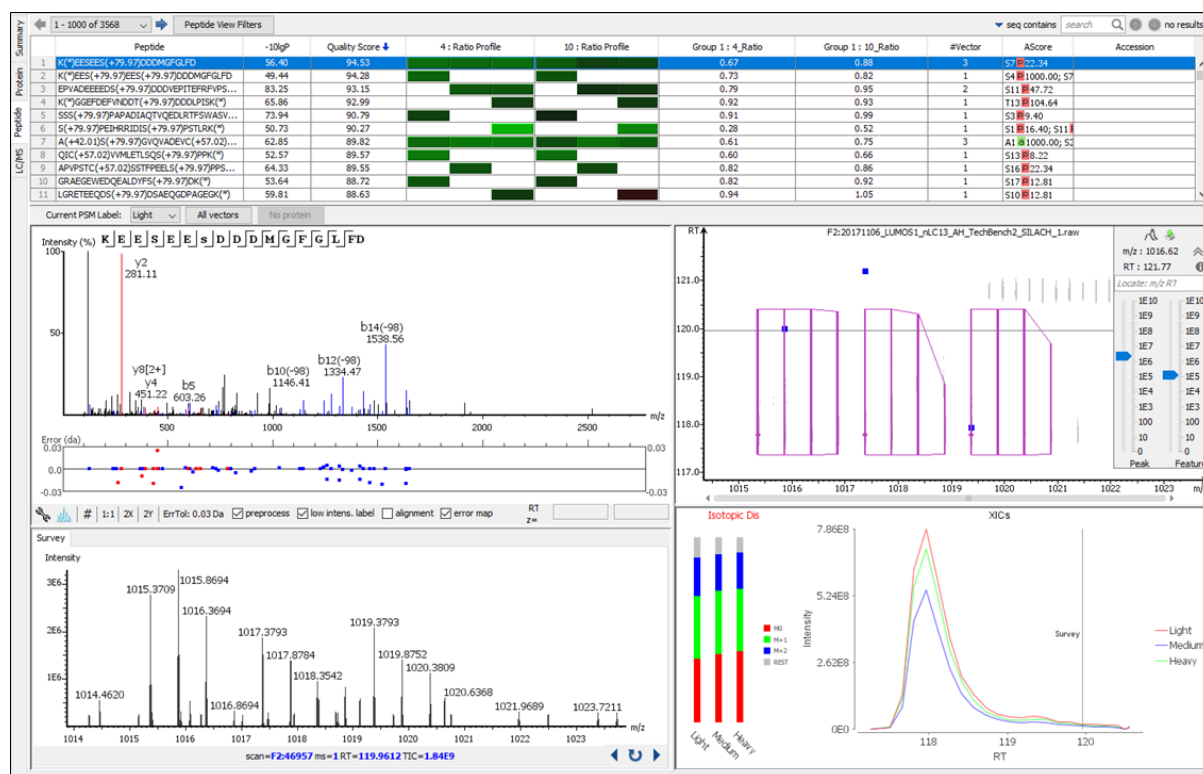
abundance. Then the peptide abundance in each condition is compared to the reference condition to obtain the peptide condition ratio in each sample separately. The peptide condition ratio of a group is computed by calculating the median from the peptide condition ratios in the samples in that group.

- **#Vector:** The number of quantifiable feature vectors included in the peptide ratio calculation. A vector is considered to be a group of mapped peptide features across the labels in the same MS run.
- **AScore:** Confidence score for a modified residue in the peptide.
- **Start:** The position of the first residue of the peptide that covers the protein.
- **End:** The position of the last residue (inclusive) of the peptide that covers the protein.

To view the actual abundance of the peptide features among the samples or groups, move the cursor over the corresponding profile. A tool-tip will display the peptide feature intensity along with the ratio. Double-click on the peptide feature to jump to the Peptide View. Right-click on the peptide feature to show a pop-up menu similar to the one in the PEAKS DB supporting peptides tab.

### 3.3. Peptide View

The "Peptide View" displays the quantified peptides and their detailed PSM matching and feature information. The interface is similar to the Peptide View in a PEAKS DB result, with the addition of an isotopic distribution heat map to the left of the XIC curve. By hovering over the different labels in the heat map, a popup will appear with the percent abundance. See Section 3.5, "Peptide View", for the remainder of the "Peptide View" details.



#### 3.3.1. Peptide Table

**Peptide Table:** The peptide table contains the important details and statistics associated with each quantified peptide that passes the filters in the Summary View.

#### Note

Right-click on a peptide in the table to 'copy the peptide sequence', 'show spectrum in LC/MS', and 'show original spectrum'.

11	ELISNASDALDK(*)	39.36	8.59		1.00	1.00	0.76	0.81	2
12	ELISNASDALDK(*)	39.36	8.59		1.00	1.00	0.88	0.88	2
13	AVFPSIVGR(*)	32.03	8.35				0.88	0.88	2
14	AVFPSIVGR(*)	32.03	8.35				0.65	0.67	2
15	AVFPSIVGR(*)	32.03	8.35				0.65	0.67	2
16	STGGAPTFRNVTVTK(*)	50.31	8.21				0.75	0.81	2
17	VIPELNGK(*)	25.00	7.99				0.77	0.77	2

- **-10lgP:** Identification -10lgP score that indicates the likelihood of a true identification. The score corresponds to the -10lgP score of the best PSM associated with a feature from all quantifiable feature vectors of this peptide.
- **Quality Score:** The higher the quality of a peptide indicates that the peptide is more quantifiable. Factors that affect the quality score include the m/z difference, RT difference, XIC shape similarity, and feature intensities.
- **Condition Profile:** The relative peptide abundances in each condition with respect to reference condition for all groups are depicted as heat map. Hold the cursor on a profile to view the samples, groups, the areas, and the ratios with respect to the reference condition.
- **Condition Ratio:** The ratio columns show the peptide abundance in each condition for all groups. A peptide's abundance in each condition is the sum of its feature vector(s)' abundance. Then the peptide abundance in each condition is compared to the reference condition to obtain the peptide condition ratio in each sample separately. The peptide condition ratio of a group is computed by calculating the median from the peptide condition ratios in the samples in that group.
- **#Vector:** The number of quantifiable vectors included in the peptide ratio calculation. A vector is considered to be a group of peptide features across labels in the same MS sample.
- **Accession:** The proteins supported by the peptide, separated by ';' if the peptide supports multiple proteins. If none of the proteins that the peptide supports pass the protein filters, the accession field will be empty.

### 3.3.2. Peptide View Filters

The quantifiable peptide list can be further filtered by the peptide view filters. These filters change the peptide list by adding or removing peptides based on the filter parameters and does not change the contents of other tabs or the peptide quantification information such as ratio. For details, please refer to Chapter 9, *Peptide, PTM, and Mutation Identification (PEAKS DB, PEAKS PTM, and SPIDER)* Section 3.5.1, "Peptide View Search and Filter Options".

### 3.3.3. All Vectors Table

Select **All Vectors** to open a window to display the feature vectors that correspond to the selected peptide above the Peptide table. The peptide -10lgP score is the the highest among the -10lgP scores of its feature vectors. The peptide quality score corresponds to the average of the quality scores of all feature vectors. The Id count of a feature vector is the number of features attached with a peptide Id.

6 feature vectors											
	Sample	Fraction	-10lgP	Quality Score ↓	Control_Area	Treated_Area	Id Count	m/z	z	ppm	RT
1	Bio Rep 2_2	F 3	70.48	2.26	9.16E7	1.05E8	1	595.9917	3	-1.1	103.05
2	Bio Rep 2_1	F 2	63.08	2.24	1.23E8	1.09E8	1	593.3220	3	1.8	101.64
3	Bio Rep 1	F 1	63.08	2.01	1.07E8	4.85E7	0	593.3220	3	1.8	101.64
4	Bio Rep 2_2	F 3	70.81	1.30	1.75E7	1.54E7	1	893.4855	2	0.7	102.96
5	Bio Rep 1	F 1	65.19	1.23	7.56E6	7.96E6	2	893.4918	2	7.8	101.82
6	Bio Rep 2_1	F 2	69.79	1.05	1.19E7	1.11E7	2	889.4800	2	2.5	101.49

Select the Protein button to go to the Protein tab. The protein to which this peptide corresponds will be highlighted in the tab. If this peptide identifies more than one protein, a new window will open and show all of the corresponding proteins. Select the desired protein to go to the Protein tab.

### 3.3.4. Peptide Details

The bottom screen of the Peptide View contains four interactive windows that are organized into the following sections:

- **MS/MS View:** The tandem scan that corresponds to the selected label set at the **Current PSM Label** drop-down menu. Click the drop-down window to change the PSM Label. This change will be reflected in the LC/

MS View, the XIC, and the MS View. See Chapter 5, *Data Visualization* Section 3, “MS/MS View”, “MS/MS View” for details.

- **LC/MS View:** This LC/MS screen is zoomed into the features that correspond to the isotopic labels, e.g. “Light” or “Heavy”. The  $m/z$  between the markers for these features should equate to the mass difference of the two labels in this sample. If R-to-P correction is applied (See Chapter 13, *PEAKS Q - MS Level* Section 2, “Setting Parameters” for details), peptides with heavy `Proline` labels will be highlighted to the right of the normal heavy label in this view. See Chapter 5, *Data Visualization* Section 4, “LC/MS View”, “LC/MS View”, for details.
- **MS View:** The survey scan represents the scan at a selected retention time on the LC/MS View. By default, the parent scan of the tandem scan (in the MS/MS view) is displayed. View other neighbouring survey scans by clicking on the horizontal arrows to move along the retention time axis. The retention time position will be adjusted in the LC/MS view and the XIC. Click the curved arrow to return to the default survey scan. See Chapter 5, *Data Visualization* Section 2, “MS View” for details.
- **XIC:** This map presents the changes in feature intensity over the retention time range of the peaks for all labels of the selected peptide. Additionally, the coloured isotopic distribution plot – M0, M+1, M+2, and REST – of each label is included, which shows the relative composition of each label.

### 3.4. LC/MS View

This view displays the distribution of LC-MS signals/features in the fractions. See Chapter 5, *Data Visualization* Section 4, “LC/MS View” for details on LC/MS View.


---

# Chapter 14. Reporter Ion Quantification (e.g. iTRAQ and TMT)

## 1. Overview


Reporter ion quantification with chemical labels at the MS2 (MS3) level is one of the three quantification modes that are supported by the optional PEAKS Q module of PEAKS Studio. It is based on the relative intensities of the fragment peaks at fixed  $m/z$  values within an MS2 (MS3) spectrum. In this mode, isobaric chemical labels 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 retention time, and are fragmented together. In the MS/MS (MS3) scans, labels from different samples will produce different reporter ions, which can then be used to calculate the quantification ratio between samples. Both user-defined and commercial labels (i.e. iTRAQ and TMT) are supported by PEAKS Q.

Reporter Ion Quantification can be launched from either the Project Wizard (see Chapter 4, *Project Wizard - Create a PEAKS Project, Load Data and Set up a Workflow*), by selecting an existing PEAKS DB result node (see Chapter 9, *Peptide, PTM, and Mutation Identification (PEAKS DB, PEAKS PTM, and SPIDER)*) or by using the PEAKS Workflow (see Chapter 18, *Workflow*). When running quantification on an existing result, ensure that the specified labels are listed as fixed PTMs on the database search result against which its referencing. After the database search is complete, follow these steps:

1. Select a PEAKS DB (PEAKS PTM or SPIDER) result node in the "Project View" pane. Click the Quantification icon from the toolbar, , or select Tools > Quantification command from the menu..


### Note

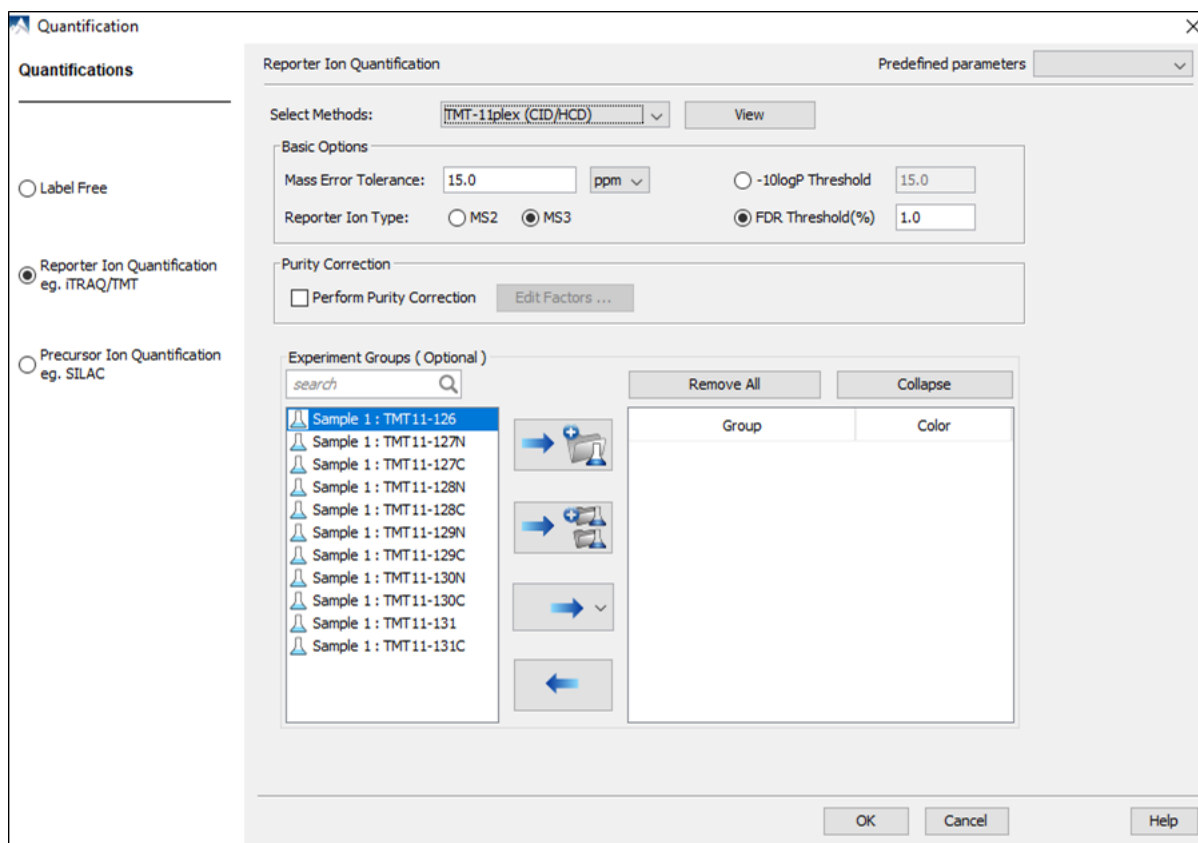
Refer to Chapter 4, *Project Wizard - Create a PEAKS Project, Load Data and Set up a Workflow* for how to perform PEAKS DB and quantification in a single easy workflow.

2. Select the quantification protocol, "Reporter Ion Quantification"; specify the PEAKS quantification parameters in the dialogue box on the right; and click OK.
3. Wait for the analysis to finish. A new quantification result node, , will appear in the "Project View". Double-click the node to open this result node.




The quantification results with labels are displayed in a format that is similar to the PEAKS DB result. A ratio is added to each quantifiable peptide and protein.

## 2. Setting Parameters


Select the Quantification toolbar icon, , or the Tools > Quantification command from the menu to create a quantification analysis workflow and to set up the corresponding parameters to perform the analysis. For labeled reporter ion quantification, select "Reporter Ion Quantification eg. iTRAQ" under the "Quantifications" options.



From the Select Methods drop-down menu, choose the appropriate quantification method used in the experiment. To examine the details of a selected quantification method, click on the View button. To create a new quantification method, select Add New Method from the Select Methods drop-down menu. Refer to Chapter 20, Section 2.3, Section 2.3, “Labeled Q Method Configuration” for how to create or edit a quantification method.

- **Mass Error Tolerance:** This parameter is used to locate the reporter ion peaks in the MS/MS scan. Either Da or ppm can be selected as the mass unit.
- **Reporter Ion Type:** This parameter allows for the selection of MS2 or MS3 as the reporter ion type for quantification.
- **-10lgP Threshold:** Only peptides with a score above this threshold are used to quantify the identified proteins. Either this option or the FDR Threshold can be selected.
- **FDR Threshold(%)**: Only peptides with a score above this threshold are used to quantify the identified proteins. Either this option or the -10logP Threshold can be selected.
- Quantification allows for the option of sample grouping. For experiments with multiple samples, the *search* bar can be used to look for a specific sample. After selecting a sample, it can be added to a new group by clicking the  button or to an existing group by clicking the  button and selecting that option from the drop-down menu. Display all samples in each group or just the individual groups by selecting the Expand and Collapse buttons, respectively. Samples can be removed individually or removed all at once by clicking the  button or the Remove All button, respectively.
- **Perform Purity Correction:** Select the checkbox to edit correction factors for chemical labels. Once the checkbox is selected, click the "Edit Factors..." button where a new dialogue will appear to allow users to setup the purity correction. The dialogue is setup to mimic a Certificate of Analysis, where each factor within the table can be modified by double-clicking the particular cell of interest.
- **Predefined parameters:** This drop-down menu enables the saving, as well as the previously saved parameters, for the ease to perform quantification regularly with the same parameters.

## 3. Understanding the Results

Once completed, the protein quantification result will be displayed in the "Project View" pane as a "QUANTIFICATION" result node, . Double-click on this node to open the result which have been divided into four categories: "Summary" view, "Protein" view, "Peptide" View, and "LC/MS" view. The "Summary" view will appear by default.

### 3.1. Summary View

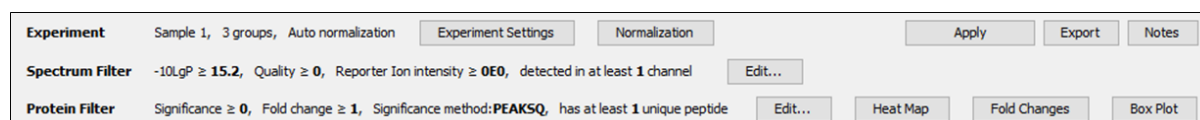
The Summary View provides three main functions:

- **Result Filtration:** This is achieved by specifying the filtration rules in the area at the top of the Summary View. The filtration function is discussed in Section 3.1.1, "Result Filtration".
- **Summary Report:** Several statistical charts are displayed to provide an overall picture of the results, assess the result quality, and examine the reliability and reproducibility of the experiment. This function is discussed in Section 3.1.2, "Summary Report".
- **Result Exporting:** This is achieved by clicking the "Export" button at the top of the Summary View. The exporting function is discussed in Section 3.1, "Summary View".

The Summary View includes an expression profile with candidate proteins assorted in a heat map, result statistics in both figures and tables, and a list of filtration parameters. To add a summary note, click on the Notes button at the top-right of the Summary View filters to open a "Notes Entry" editor. These notes will be displayed in the Summary View and Summary View export.

#### 3.1.1. Result Filtration

PEAKS provides an extensive set of tools to filter reporter ion labelled quantification results according to the users' requirements. These options are divided into three groups on the Summary View toolbar: Experiment, Spectrum Filter, and Protein Filter.

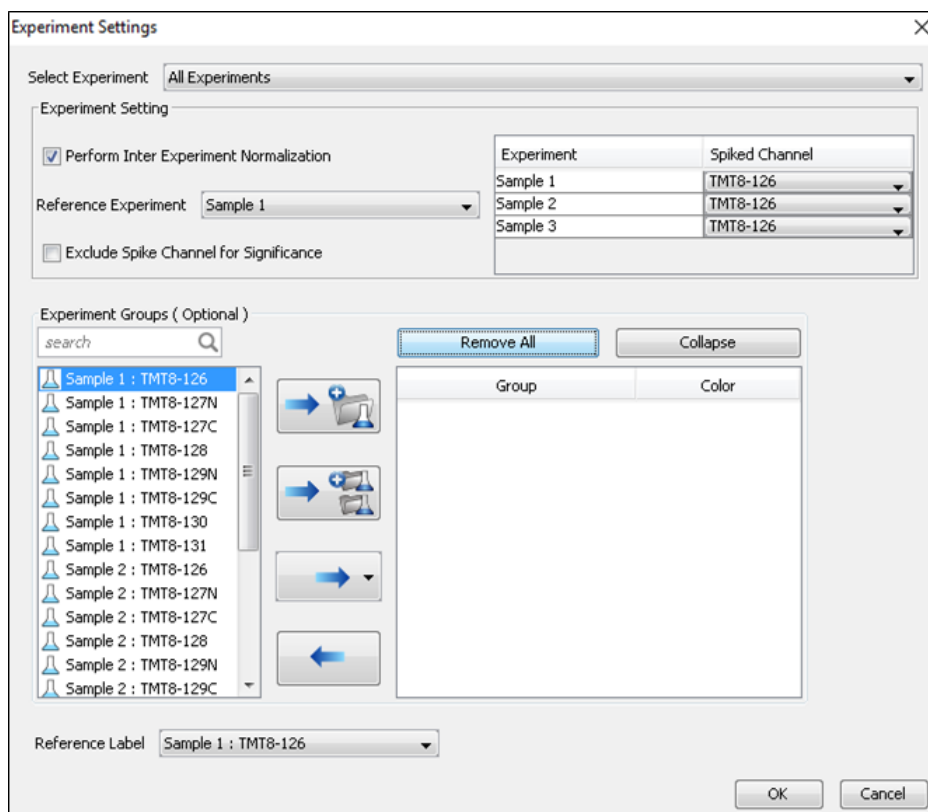





##### 3.1.1.1. Experiment

The experiment group provides additional tools to setup the project accordingly for data analysis. There are two buttons in this group: Experiment Settings and Normalization.

The following parameter dialogue appears when the Experiment Settings button is selected.





- Select Experiment:** When running quantification on an identification node that searched data within multiple samples (i.e. replicate experiments within different treatments), this drop-down menu allows for the selection of a single sample or every sample (All Experiments) for analysis. Every label of the sample(s) selected will appear in the screen located below the *search* bar.
- Experiment Groups (Optional):** This section enables sample grouping, which can be facilitated by typing a sample name (i.e. Sample 1) or a sample label (i.e. TMT-126) into the search bar. All samples/labels that are identified in the search will be listed together at the top of the screen for easy subsequent grouping into the screen on the right. After selecting the sample(s) to be grouped, add to a new group by clicking the  button or to an existing group by clicking the  button and selecting that existing group from the drop-down menu. All grouped samples can be displayed or hidden by selecting the Expand and Collapse buttons, respectively. Samples can be removed individually or all at once by clicking the  button or by clicking Remove All, respectively.
- Reference Label:** Use this drop-down menu to select a reference label. The peptide area corresponding to this reference label will be the denominator in all feature area ratio calculations. Only a label that is already listed in the Experiment Groups screen can be selected as a reference. Furthermore, if the samples/labels have been grouped together onto the right-hand side in the Group/Color window, the text "Base Group" will appear at the bottom-right of the experiment window and state the reference group, which is the group that contains the reference label.
- Experiment Setting:** Changes in this section can only be made when "All Experiments" is selected from the above drop-down menu. These options dictate which samples/labels the normalization options apply.
  - Perform Inter Experiment Normalization:** Check this box if a spike label was used in the experiment. This is used to enable quantification between experiments.
  - Spike Channel Drop-down Menus:** Select the channel in each sample that will be used for normalization. The software assumes that similar peptide features containing this label across experiments will approach a 1:1 ratio. Differences are assumed to be due to systematic error and all labeled channels are normalized to this spike label.



- **Exclude Spike Label for Significance:** Since the spike label is assumed to be in a 1:1 ratio between experiments, it should not be included in the significance calculation.

The following parameter dialogue appears when the Normalization button is selected.

Normalization

Normalization Mode

☐ No normalization ☒ Auto normalization

☐ Manual input ☐ Normalize to spike

Normalization Factors

Experiment: Sample 1

Channel	Name	Factor	Expected Ratio
TMT10-126	TMT10-126	1.000	1.000
TMT10-127N	TMT10-127N	1.042183	1.000
TMT10-127C	TMT10-127C	1.027791	1.000
TMT10-128N	TMT10-128N	0.98514	1.000
TMT10-128C	TMT10-128C	1.095197	1.000
TMT10-129N	TMT10-129N	1.027627	1.000
TMT10-129C	TMT10-129C	1.112901	1.000
TMT10-130N	TMT10-130N	0.971478	1.000
TMT10-130C	TMT10-130C	0.950701	1.000
TMT10-131	TMT10-131	0.952784	1.000
TMT11-131C	TMT11-131C	0.978411	1.000

OK Cancel

## Note

If there are multiple samples in the quantification run and all samples are selected, No Normalization is chosen as the default option and will be the only normalization option available for selection unless "Perform Inter Experiment Normalization" is check-marked in Experiment Settings.

## Note

- **Auto normalization:** This option will calculate a global ratio from the total intensity of all labels in all quantifiable peptides. Factor values and Expected Ratios between these channels (the labels) will be presented. The Experiment drop-down menu can be opened to display the results of another sample.
- **Manual input:** This option performs the same default calculation as Auto normalization but also allows for the manual change of each label's Expected Ratios. After changing a ratio and hitting "Enter" on the keyboard, PEAKS will recalculate the Factor for the non-referenced label(s). The Experiment drop-down menu can be opened to display the results of another sample.
- **Normalize to spike:** This option displays a list of identified proteins, which can be selected to create a normalization ratio. Use the search bar to quickly locate specific proteins. Right-clicking in the Protein list will open a popup menu to select all highlighted rows or remove all selected proteins instead. The expected ratios can be manually modified. It is assumed that the total protein content of the selected proteins in different labels have the ratios entered in the 'Expected ratio' column. If desired, sample names can be changed by double-clicking the chosen channel below the "Name" column and typing the new name.

**Normalization**

Normalization Mode

☐ No normalization
 ☐ Auto normalization
 ☐ Manual input
 ☒ Normalize to spike

Normalization Factors

Experiment: Sample 1

Channel	Name	Factor	Expected Ratio
TMT8-126	Sample 1 : T...	1.000	1.000
TMT8-127N	Sample 1 : T...	3.712786	1.000
TMT8-127C	Sample 1 : T...	0.075996	1.000
TMT8-128	Sample 1 : T...	3.288936	1.000
TMT8-129N	Sample 1 : T...	272.566437	1.000
TMT8-129C	Sample 1 : T...	258.64624	1.000
TMT8-130	Sample 1 : T...	406.862335	1.000
TMT8-131	Sample 1 : T...	212.58728	1.000

Protein List

search

Prev Next

Protein	
P31327 CP5M_HUMAN	<input checked="" type="checkbox"/>
P31327-3 CP5M_HUMAN	<input checked="" type="checkbox"/>
O75369 FLNB_HUMAN	<input checked="" type="checkbox"/>
tr E7EN95 E7EN95_HUMAN	<input checked="" type="checkbox"/>
O75369-7 FLNB_HUMAN	<input checked="" type="checkbox"/>
O75369-6 FLNB_HUMAN	<input checked="" type="checkbox"/>
P08	<input type="checkbox"/>
Q00	<input type="checkbox"/>
P31	<input type="checkbox"/>
O75369-8 FLNB_HUMAN	<input type="checkbox"/>
P14618 KPVM_HUMAN	<input type="checkbox"/>
O75369-3 FLNB_HUMAN	<input type="checkbox"/>
P22314 UBA1_HUMAN	<input type="checkbox"/>
O75369-6 FLNB_HUMAN	<input type="checkbox"/>
Q00610 CLH1_HUMAN	<input type="checkbox"/>
O75369-2 FLNB_HUMAN	<input type="checkbox"/>

3 proteins selected  
(Right click to show popup menu)

OK Cancel

Once changes have been made to the Experiment Settings or Normalization options, the Apply button from the Summary View toolbar will become red in text to indicate that the changes have not yet been applied to the project. Select Apply to make these changes come into effect.

### 3.1.1.2. Spectrum Filter

The following parameter dialogue appears when the Edit... button is selected.

**Spectrum Filters**

-10LgP  $\geq$  15.2

Quality 0

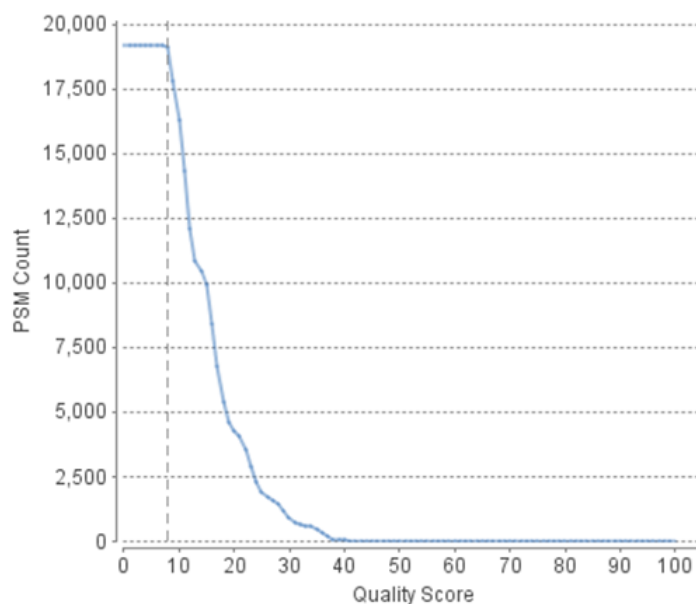
Reporter Ion Intensity  $\geq$  0E0

Reference Channel Present ☐

Number of Channels Present  $\geq$  1

OK Cancel

- **-10LgP:** Only peptides with a score above this threshold are used to quantify the identified proteins. PEAKS Q will not allow the user to set a value less than the value selected during the quantification setup (refer to Section 2, "Setting Parameters"). Click the FDR button to set a -10LgP value equivalent to the selected FDR.
- **Quality:** A higher quality peptide indicates that the peptide is more quantifiable. Factors that affect the quality score include the identification -10LgP score, the noise around the reporter ions, and the mass error of reporter ions. An appropriate threshold value can be selected by checking Figure 3(b): Accumulation curve for quality in the "Summary" view. The rule of thumb is to set the quality threshold near the region of the quality score axis at which the curves begin to significantly drop.



- **Charge between:** Only peptides with feature vectors that all fall within this range will be used in protein quantification.
- **Reference label present:** Only quantifiable peptides in which the reference label is present will be used in protein quantification. The reference label is set in the experiment settings pop-up.
- **Number of labels present:** Only quantifiable peptides that contain at least the number of labels selected will be used in protein quantification.

Select Apply within the filtration window of the "Summary" view to include all selected filters in the analysis.

### 3.1.1.3. Protein Filter

The following parameter dialogue appears when the Edit... button is selected. These filters determine what will be included in a protein group significance calculation, as well as what significance score thresholds will be set. For a protein group to have a significance score, it must have been identified in a sample/label that has already been placed into a group (in either the Tools > Quantification window or in Experiment Settings).

- **Significance:** Only protein groups with a significance above this threshold will be listed in the "Protein" view. Either this or "FDR" can be selected to set a significance threshold. The significance score is the  $-10\lg$  of the p-value calculated using either ANOVA or PEAKS Q significance. The p-value calculates the likelihood that the observed change is caused by random chance. A significance score threshold of 20 is recommended.
- **FDR:** Adopt the Benjamini–Hochberg method to adjust the p-value to the false discovery rate for all protein groups that have already passed the other filters. Only protein groups with significance scores passing the calculated FDR will be listed in the "Protein" view. Either this or "Significance" can be selected to set a significance threshold.

## Note

For more details on the Benjamini-Hochberg FDR method, refer to the following paper: *"Controlling the false discovery rate: a practical and powerful approach to multiple testing."* J. R. Statist. Soc. 1995, B57, 289-300.

- **Fold Change:** Only protein groups at or above this fold change threshold will be listed in the "Protein" view.
- **Unique Peptides:** Only protein groups with at least this many unique peptides will be listed in the "Protein" view.
- **Modified Form Exclusion:** Peptides that are modified with a variable PTM will not be included in the significance calculation for a protein group. This does not apply to the quantification labels. If a peptide is present as an unmodified and modified form, both the unmodified and modified form will be excluded from protein ratio calculation.
- **Significance Method:** One of two significance calculations can be applied to the list of protein groups: PEAKS Q or ANOVA. Scores will only be calculated for protein groups identified from samples that are in groups. PEAKS Q is similar to the Significance B method employed by the paper referenced below. Additionally, support for peptide similarity is considered. ANOVA requires that at least two samples be included in each particular group. The ANOVA method employed is also referenced below.

## Note

For more details on PEAKS Q, refer to the following paper: *"MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification."* Nat Biotechnol. 2008, 26(12): 1367-1372.

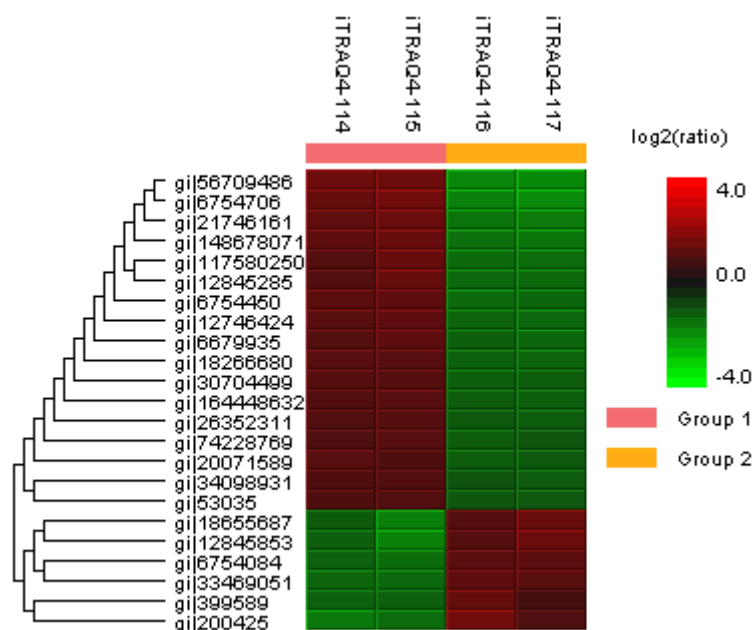
## Note

For more details on ANOVA, refer to the following paper: *"On the comparison of several mean values: an alternative approach"* Biometrika. 1951, 38(3/4): 330-336.

### 3.1.2. Summary Report

#### 3.1.2.1. Heat Map

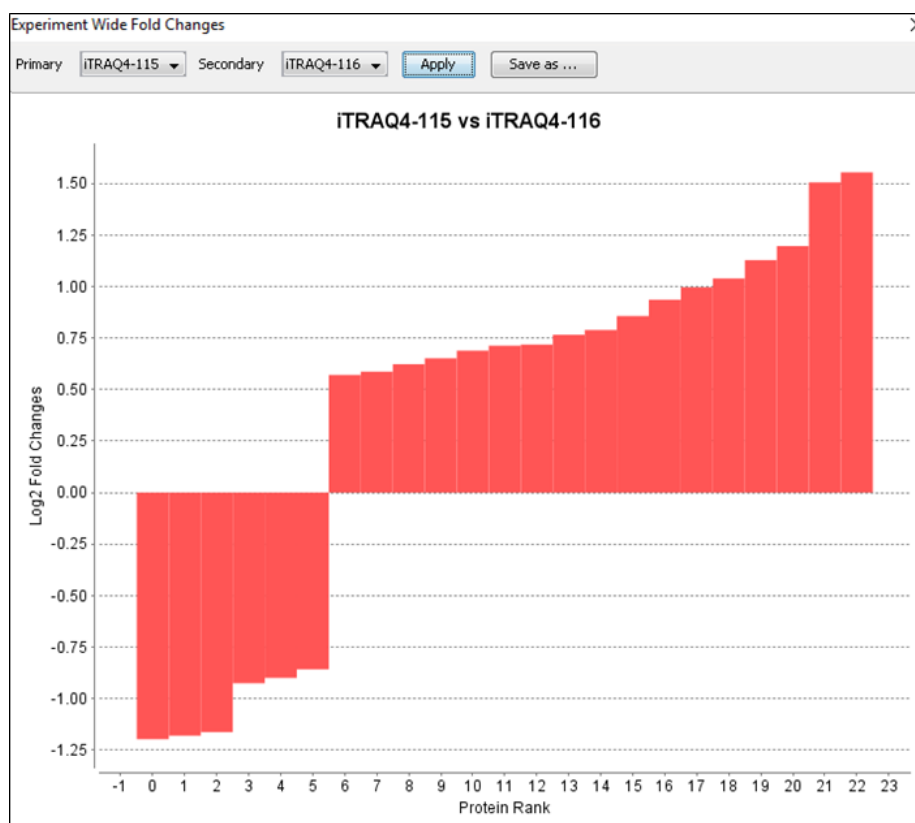
The heat map displays the protein groups that passed the filters. The relative protein abundance is represented as a heat map of the representative proteins of each protein group. These representative proteins are clustered if they exhibit a similar expression trend across the samples. The hierarchical clustering is generated using a neighbour-joining algorithm with a Euclidean distance similarity measurement of the log<sub>2</sub> ratios of the abundance of each sample relative to the average abundance. An interactive protein profile heat map is opened in a new window when the Heat Map button is selected. Right-click on the heat map in this window to save a copy of it.



In the interactive heat map (displayed by clicking the Heat Map button), move the cursor over any point to highlight the protein accession name and the sample/label to which it matches. Right-click on the heat map and click Save as... to export a PNG file of the heat map.

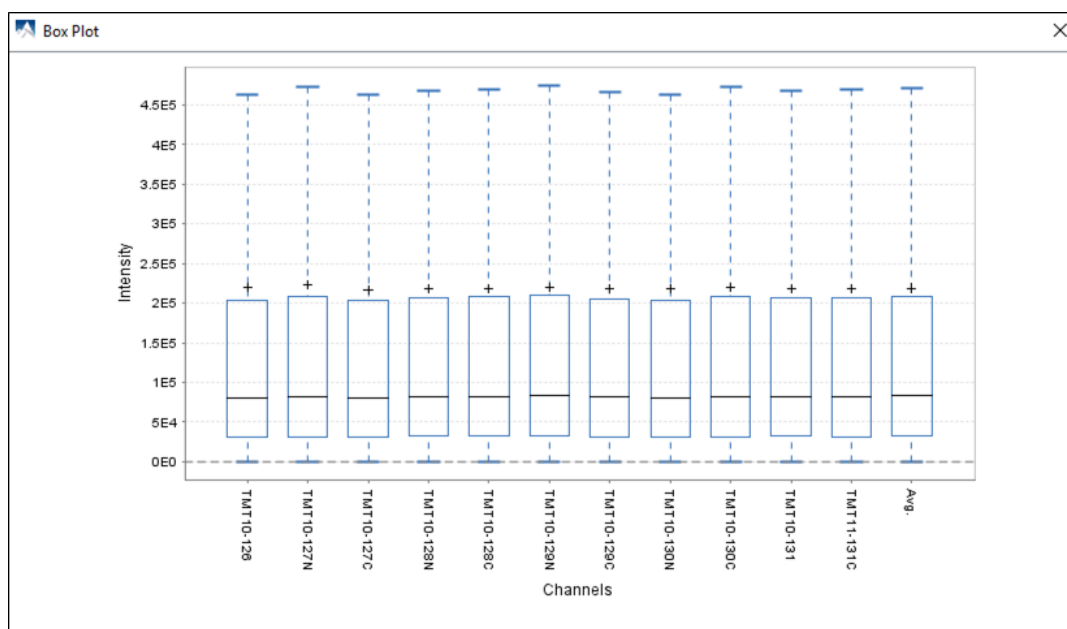
### 3.1.2.2. Interactive Expression Wide Fold Change Graph

The following bar graph displays the relative log2 ratio between two selected groups with respect the rank of each protein. The log2 function applies to the selected primary group with respect to the selected secondary group. Hold the cursor over any bar in this figure to display the associated fold change of each protein within the rank. Click the drop-down menus to change the primary and secondary groups. Click Apply to set the new log2 ratio. Click **Save as...** to export a .PNG file of the bar graph.



### 3.1.2.3. Box Plot

An interactive box plot (by clicking the **Box Plot** button) displays a distribution of the intensity values for each label in the experiment. A plot displaying the average distribution for all labels is also included.



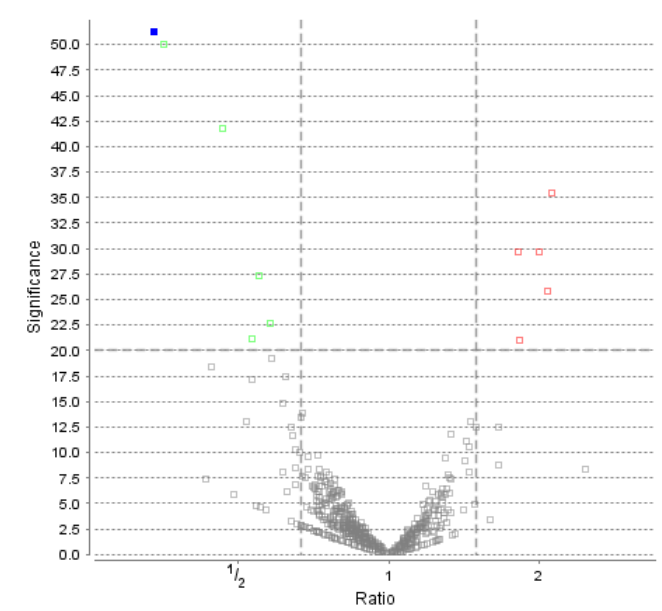
### 3.1.2.4. Volcano Plot

A volcano plot combines a statistical test with the magnitude of the change enabling quick visual identification of those data-points that display large-magnitude changes and those that are also statistically significant. The volcano plot in Figure 2(a) plots significance versus fold-change of the quantified proteins.

- **Horizontal broken grey line:** The selected significance threshold.

- **Vertical broken grey lines:** The selected fold change thresholds

Plotting points in this way results in two regions of interest in the plot: those points that are found towards the top of the plot that are far to either the left or the right side. These represent values that display large magnitude fold changes (hence being left or right of center), as well as high statistical significance (hence being towards the top). In this figure, the ratio is the group ratio set with respect to the base group. If more than two groups are present, the max ratio is used instead.

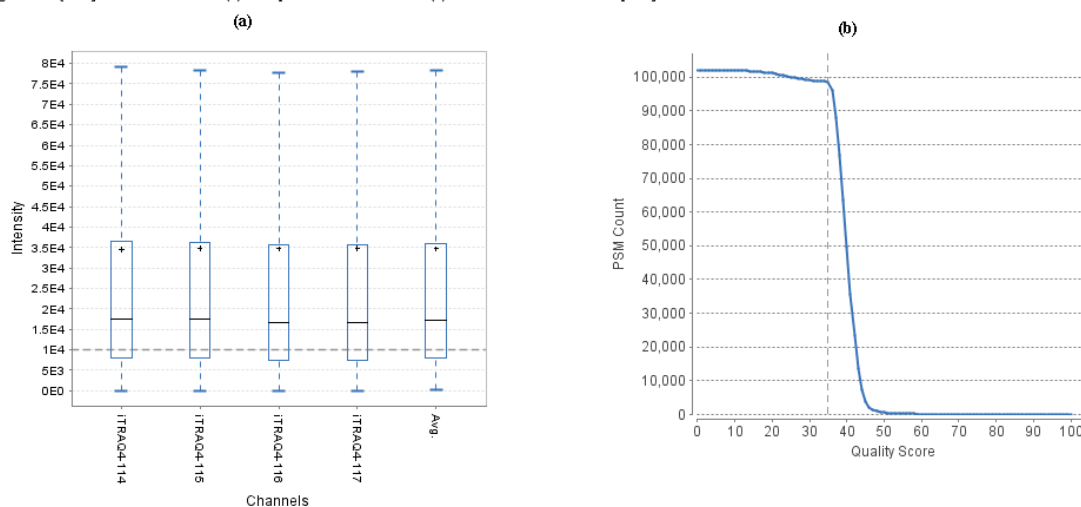


The “Protein” view contains an interactive volcano plot that be hidden or displayed. Markers for the proteins that are above the set significance threshold will be displayed in color. Scrolling the cursor over a marker will highlight it and, subsequently, clicking on this marker will highlight the protein in the interface.

### 3.1.2.5. Quality Control Statistics

Figure 3 displays the box plot and the accumulation quality curve described in Sections 3.1.2.3 and 3.1.1.2, respectively.

**Figure 3.** Quality control statistics. (a) Box plot for each channel. (b) Accumulation curve for quality. [?](#)



### 3.1.2.6. Summary Tables

Tables 1 and 3 list the data and the filtered results, respectively. Table 2 lists the feature vector and protein result filtration parameters.

**Table 1.** Statistics of data.

# of MS Scans	129931
# of MS/MS Scans	410708

**Table 2.** Result filtration parameters.

Peptide significance	$\geq 0$
Peptide fold change	$\geq 1$
Report ion intensity	$\geq 1E4$
Quality	$\geq 35$
Reference channel present	false
Number of channels present	1
Protein significance	$\geq 10$
Protein fold change	$\geq 1.5$
Unique peptides	$\geq 1$
Significance method	PEAKSQ

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

Peptide-Spectrum Matches	67565
Peptide Sequences	8650
Protein Groups	23

Tables 4 and 5 list the quantification and instrument parameters, respectively.

**Table 4.** Quantification parameters.

Quantification Type: iTRAQ-4plex  
 Quantification Mass Tolerance: 0.2Da  
 FDR Threshold(%): 1.0  
 Reporter Ion Type: MS2  
 iTRAQ4-114: 114.1107  
 iTRAQ4-115: 115.1077  
 iTRAQ4-116: 116.1111  
 iTRAQ4-117: 117.1144  
 Group Settings in Parameters:  
 Group 1: Sample 1: iTRAQ4-114  
 Group 1: Sample 1: iTRAQ4-115  
 Group 2: Sample 1: iTRAQ4-116  
 Group 2: Sample 1: iTRAQ4-117  
 DB Dependency: 24  
 Current Group Information:  
 Group 1: iTRAQ4-114  
 Group 1: iTRAQ4-115  
 Group 2: iTRAQ4-116  
 Group 2: iTRAQ4-117

**Table 5.** Instrument parameters.

Fractions: 100702V1RS1\_Kaja\_sample1\_frac\_1.raw, 100709V1RS1\_Kaja\_sample1\_frac\_2.raw, 100712V1RS1\_Kaja\_sample1\_frac\_3.raw, 100714V1RS1\_Kaja\_sample1\_frac\_4.raw, 100715V1RS1\_Kaja\_sample1\_frac\_5.raw, 100716V1RS1\_Kaja\_sample1\_frac\_6.raw, 100719V1RS1\_Kaja\_sample1\_frac\_7.raw, 100723V1RS1\_Kaja\_sample1\_frac\_8.raw, 100726V1RS1\_Kaja\_sample1\_frac\_9.raw, 100702V1RS1\_Kaja\_sample1\_frac\_10.raw, 100708V1RS1\_Kaja\_sample1\_frac\_11.raw, 100709V1RS1\_Kaja\_sample1\_frac\_12.raw, 100712V1RS1\_Kaja\_sample1\_frac\_13.raw, 100714V1RS1\_Kaja\_sample1\_frac\_14.raw, 100715V1RS1\_Kaja\_sample1\_frac\_15.raw, 100716V1RS1\_Kaja\_sample1\_frac\_16.raw, 100719V1RS1\_Kaja\_sample1\_frac\_17.raw, 100723V1RS1\_Kaja\_sample1\_frac\_18.raw, 100726V1RS1\_Kaja\_sample1\_frac\_19.raw, 100702V1RS1\_Kaja\_sample1\_frac\_20.raw, 100708V1RS1\_Kaja\_sample1\_frac\_21.raw, 100709V1RS1\_Kaja\_sample1\_frac\_22.raw  
 Ion Source: ESI(nano-spray)  
 Fragmentation Mode: CID/ECD  
 MS Scan Mode: FT-ICR/Orbitrap  
 MS/MS Scan Mode: Linear Ion Trap

### 3.1.3. Result Exporting


PEAKS reporter ion labeled quantification results can be exported to various HTML and text formats by clicking the Export button within the filtration options in the Summary View. The Summary View page and the images from the various Summary View toolbar functions can also be exported to various HTML and text formats. Refer to Section 5.1, “Exporting Precursor Ion Labelling Quantification Results” for details.

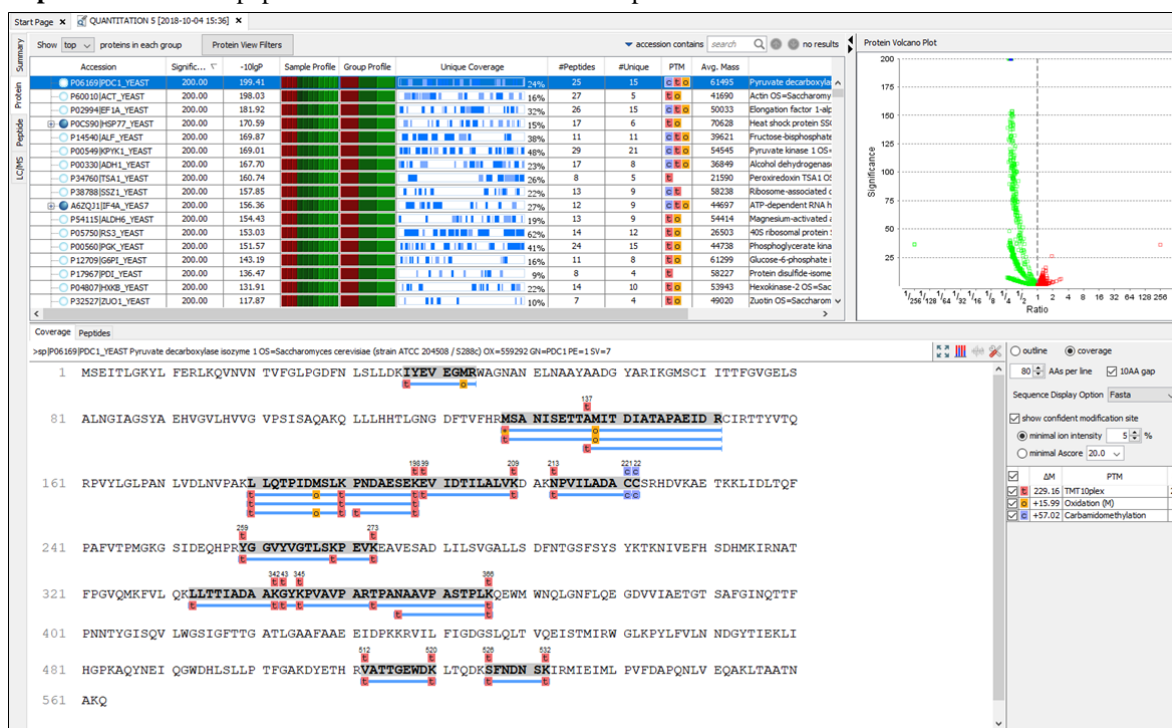
## 3.2. Protein View

The Protein View lists all the quantified proteins present in the sample, characterizes each protein at the amino acid level, and lists the supporting peptide features of each protein. It has six components:

- **Protein Table:** This lists all the quantified proteins present in the sample.



- **Protein View Filtering:** Allows for filtering the proteins. For more information refer to Section 3.4.1, “Protein View Filters”
- **Protein Volcano Plot:** Click on a marker that is above the significance threshold that was set in the “Summary” view. Upon selection, the corresponding protein in the protein table will be highlighted. Right click and drag is also supported in the volcano plot to zoom in on an area. Double click to return to the default scale.
- **Coverage:** This characterizes the protein sequence at the amino acid level. All the PTMs and mutations occurring on the protein sequence will be displayed together with the MS/MS data supporting the inference.
- **PTM Profiling:** Click on the  open PTM Profiling. For more information refer to Chapter 11, *PTM Profiling*.
- **Peptides:** The list of peptides identified from the selected protein in the Protein table.



### 3.2.1. Protein Table

The quantified proteins are listed in the protein table with homologous proteins grouped together. This Protein table is similar to the tables displayed in the PEAKS DB, PEAKS PTM, and SPIDER results, but includes the following additions:

- **Significance:** The protein significance is set according to the rules set in the Summary View, refer to Section 3.1.1.3, “Protein Filter”.
- **Sample Profile:** The protein abundance in each sample relative to the average is log2 transformed and depicted as a heat map. Hold the cursor on a profile to view the sample channels (labels), the areas, and the ratios with respect to the reference label selected in Experiment Settings of the Summary View.
- **Group Profile:** The protein abundance in each group relative to the average is log2 transformed and depicted as a heat map. Hold the cursor on a profile to view the group channels (labels), areas, and ratios with respect to the group which contains the reference label selected in Experiment Settings of the Summary View.

See Section 3.4, “Protein View” for more details on the other entries of the Protein table.

### 3.2.2. Protein Volcano Plot

This interactive volcano plot enables the selection of a protein in the table. Protein markers that are above the significance threshold can be selected. Upon selection, the corresponding protein in the Protein Table will be highlighted. Proteins located in the top-right and top-left sections of this plot (above the significance threshold and to the right and left of the fold change thresholds) are considered statistically significant. The volcano plot facilitates identification of these significant proteins. Proteins filtered out using the "Protein View Filter" will be grayed along with those described in Section 3.1.2.4, "Volcano Plot".

### 3.2.3. Coverage Tab

The Coverage tab characterizes the selected protein sequences at the amino acid level. See Section 3.4.4, "Protein Coverage" for details.

### 3.2.4. Peptides Tab

The supporting peptides assigned to the protein are shown below the "Peptides" tab.

Coverage	Peptides	Used	Unique	-10lgP	Quality Score	Sample Profile	Group Profile	Mass	ppm	m/z	RT	Scan	#Feature	Start	End	PTM	AScore
1	R.M(+229.16)SANSETTAM(+15.99)ITDI...	<input checked="" type="checkbox"/>	•	61.73	27.48			2766.3618	-0.1	1384.1880	166.51	F2:86130	3	128	151		M11 1000.00
2	R.M(+229.16)SANSETTAM(+15.99)ITDI...	<input checked="" type="checkbox"/>	•	51.32	19.60			2782.3567	-0.5	928.4590	143.76	F1:73009	1	128	151		M11 1000.00
3	R.Y(+229.16)GGVYVGLSK(+229.16)PEY...	<input checked="" type="checkbox"/>	•	48.04	23.60			2283.3447	1.9	762.1236	104.77	F2:54933	1	259	273		
4	K.E(+229.16)EDTLALVK(+229.16)D	<input checked="" type="checkbox"/>	•	47.41	24.86			1671.0587	2.1	836.5384	178.66	F1:88135	1	199	209		
5	K.L(+229.16)QTFIDMSLK(+229.16)PND...	<input checked="" type="checkbox"/>	•	45.44	26.89			2815.5610	1.7	704.8987	142.59	F1:72373	2	180	198		
6	R.V(+229.16)ATTGEVDK(+229.16)L	<input checked="" type="checkbox"/>	•	44.73	26.92			1463.8026	2.1	732.9101	64.01	F2:31602	1	512	520		
7	K.L(+229.16)QTFIDMSLK(+229.16)PND...	<input checked="" type="checkbox"/>	•	41.44	20.73			2831.5559	1.0	944.8602	118.80	F1:66807	1	180	198		
8	T.A(+229.16)METDIAAPAEIDR.C	<input checked="" type="checkbox"/>	•	39.36	22.76			1815.9603	0.5	908.9879	125.73	F2:65914	1	137	151		M8 1000.00
9	A.N(+229.16)AAVPASTPLK(+229.16)Q	<input checked="" type="checkbox"/>	•	39.01	22.18			1525.9233	2.3	763.9707	64.24	F2:31813	1	356	366		
10	K.G(+229.16)K(+229.16)PVAIPAR.T	<input checked="" type="checkbox"/>	•	38.37	23.26			1514.9338	1.8	758.4755	43.36	F2:20044	1	343	352		
11	R.Y(+229.16)PANAIVASTPLK(+229.16)Q	<input checked="" type="checkbox"/>	•	36.50	24.08			1795.0609	2.3	599.3623	77.82	F2:39583	3	353	366		
12	K.F(+229.16)QAESEK(+229.16)E	<input checked="" type="checkbox"/>	•	33.87	20.94			1246.7084	1.2	674.3622	20.03	F2:7170	1	191	198		
13	K.N(+229.16)PVLAADAC(+57.02)C(+57...	<input checked="" type="checkbox"/>	•	33.83	22.50			1360.6682	0.7	681.3419	97.31	F1:49956	1	213	222		
14	K.L(+229.16)TTIADAAK(+229.16)G	<input checked="" type="checkbox"/>	•	33.05	21.67			1473.9172	0.4	492.3132	133.18	F2:69745	1	333	342		
15	K.L(+229.16)QTFIDMSLK(+229.16)PND...	<input checked="" type="checkbox"/>	•	30.59	23.85			1732.0210	1.3	867.0189	129.93	F2:68056	1	180	190		M8 1000.00
16	K.L(+229.16)VEVSGK(+15.99)K.W	<input checked="" type="checkbox"/>	•	27.15	21.83			1240.6324	0.7	621.3239	40.06	F1:17968	1	37	44		M7 1000.00
17	K.S(+229.16)PNDMSK(+229.16)I	<input checked="" type="checkbox"/>	•	22.66	22.99			1268.6766	1.2	638.3464	25.43	F2:10243	1	526	532		

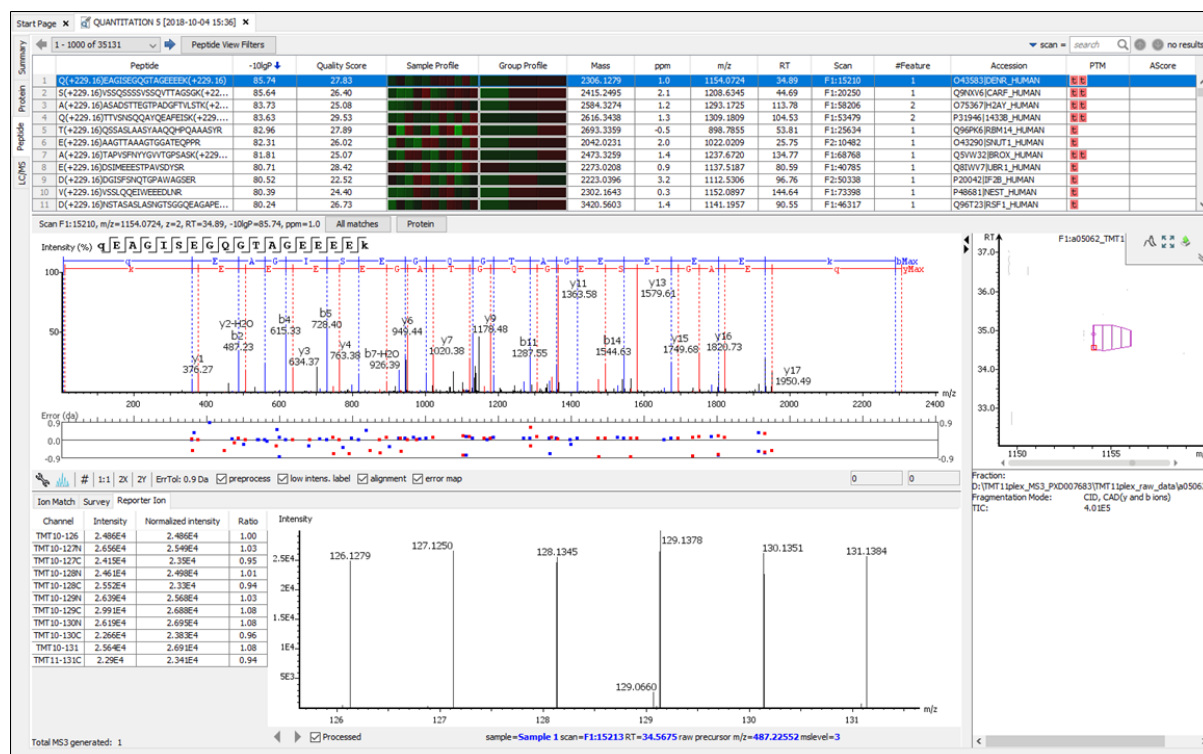
This table includes the following columns:

- **(Checkbox):** Select the peptides to appear within the Coverage Tab as blue bars.
- **Used:** Select whether this peptide is used to calculate the protein profile.
- **Unique:** This shows which peptides are only assigned to the current protein group.
- **Quality Score:** The higher quality of a peptide indicates that the peptide is more quantifiable. Factors that affect the quality score include the identification -10LgP score, the noise around the reporter ions, and the mass error of reporter ions. An appropriate threshold value can be selected by checking Figure 3(b): Accumulation curve for quality in the "Summary" view. The rule of thumb is to set the quality threshold near the region of the quality score axis at which the curves begin to significantly drop.
- **Sample Profile:** The peptide abundance in each sample relative to the average is log2 transformed and depicted as a heat map. Hold the cursor on a profile to view the sample channels (labels), the areas, and the ratios with respect to the reference label selected in Experiment Settings of the "Summary" view.
- **Group Profile:** The peptide abundance in each group relative to the average is log2 transformed and depicted as a heat map. Hold the cursor on a profile to view the group channels (labels), the areas, and the ratios with respect to the group that contains the reference label selected in Experiment Settings of the "Summary" view.
- **#Spec:** The number of quantification spectra included in the peptide ratio calculation. A vector is considered to be a group of peptide features across samples and replicates used to create a peptide ratio. For example, if there are three replicates in your experiment, you will have a maximum of three vectors.
- **Start:** This shows the position of the first residue of the peptide that covers the protein.
- **End:** This shows the position of the last residue of the peptide that covers the protein.

### 3.3. Peptide View

The "Peptide" view displays all the quantified peptides and their intensities. The interface is similar to the "Peptide" view in a PEAKS DB result. See Section 3.5.1, "Peptide View Search and Filter Options" for details. The Peptide

table may be filtered by clicking on the "Peptide View Filters" button, for more information refer to Chapter 9, *Peptide, PTM, and Mutation Identification (PEAKS DB, PEAKS PTM, and SPIDER)*. The average quality score of the top quantified spectra of a quantifiable peptide is listed in the spreadsheet. The Sample Profile displays a heatmap of the peptide abundance in each sample relative to the average after log<sub>2</sub> transformation. Hold the cursor on a profile to view the sample channels (labels), the areas, and the ratios with respect to the reference label selected in Experiment Settings of the "Summary" view.



### 3.3.1. Ion Match Table

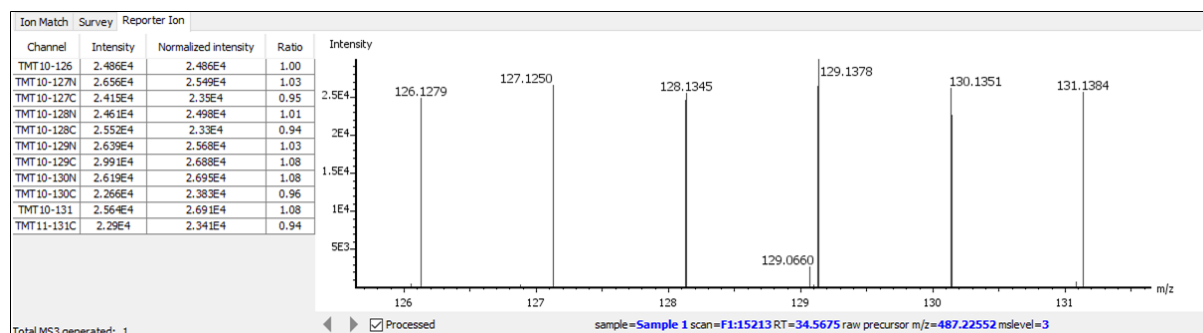
For details on the Ion Match Table, see Section 3.2.3, "Ion Table".

### 3.3.2. Survey Scan

For details on the survey scan, see Section 3, "MS/MS View".

### 3.3.3. Reporter Ion

This view provides details on characteristics of the reporter ions used in the experiment. The table provides information on the intensity, the raw intensity, and the ratio of each channel with respect to the set reference channel. Additionally, an associated MS2 or MS3 spectrum is included of the reporter ions in the experiment, determined by the selection made in the Quantification parameter settings, see Section 2, "Setting Parameters". This is an interactive spectrum, enabling zooming in and out using the scroll wheel of the mouse. Double-clicking will zoom out and reveal the entire MS3 spectrum.



### **3.4. LC/MS View**

This view displays the distribution of LC-MS signals/features in the fractions. See Section 4, “LC/MS View”.

---

# Chapter 15. Label Free Quantification (LFQ)

## 1. Overview


Label-free quantification is one of the three quantification methods supported by the PEAKS Q module. It serves as a fast profiling tool for differential studies of large scale proteomics. This quantification method is based on the relative abundance of peptide features detected in multiple samples. Feature detection is performed separately on each sample. By using the EM (expectation-maximization) algorithm, more overlapped features can be detected. The features of the same peptide from different samples are reliably aligned together using a high-performance retention time alignment algorithm.

The label-free quantification of PEAKS Q module does not have to be based on an identification result as label-free quantification is based on detected features. However, you can choose to attach an existing identification result as the last step of the label-free quantification set-up.

### Note

For details of the retention alignment algorithm, refer to the paper "*A Combinatorial Approach to the Peptide Feature Matching Problem for Label-Free Quantification*", *Bioinformatics*, 2013, 10.1093.

The usage of this function is outlined in the following overview.

1. Select a project node in the "Project View" frame. Click the PEAKS Quantification tool bar icon  or select the Tools > Quantification command from the menu.

### Note

Refer to Chapter 4, *Project Wizard - Create a PEAKS Project, Load Data and Set up a Workflow* for project creation, data loading, PEAKS DB search, and quantification in a single work flow.

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.

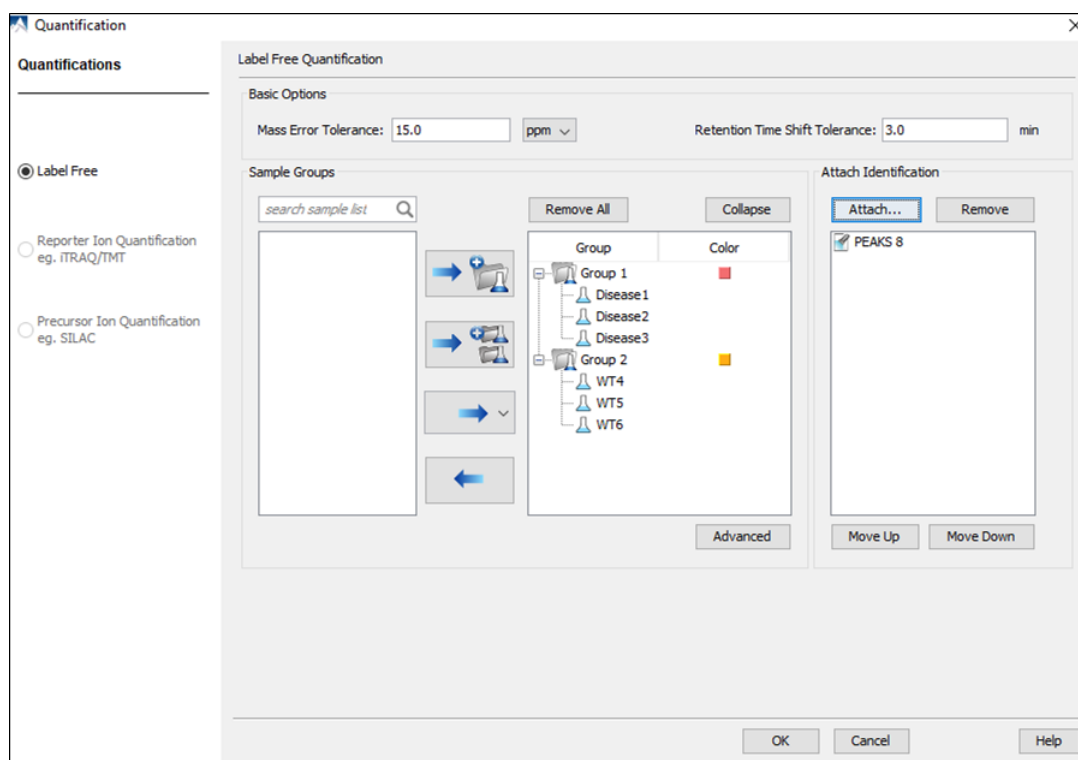
### Note

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 entitled "LABEL FREE #" will appear at the project tree. Double-click the node to open the result file.

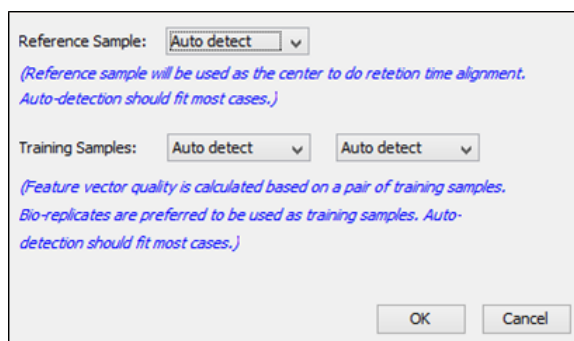
## 2. Setting Parameters

The following parameter dialogue appears when the quantification tool bar icon, , is clicked or Tools > Quantification command from the menu to create a quantification analysis work flow and to set up the corresponding parameters to perform the analysis. Select "Label Free" under the "Quantifications" options.

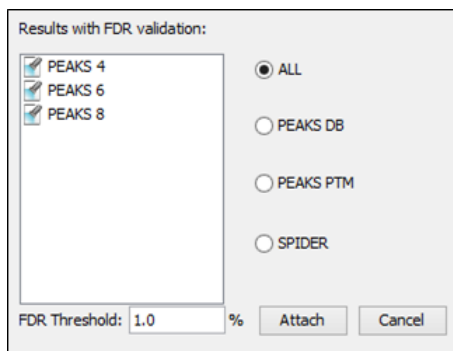


The following parameters are available in the “Basic Options” section of the quantification window:

- **Mass Error Tolerance:** The mass shift between different runs. Initially this should be set as twice as the mass error tolerance in database search. Later, by checking the M/Z shift distribution figure in the result, you can narrow the tolerance.
- **Retention Time Shift Tolerance:** The retention time range is the maximum elution time range that is considered for the quantification of an identified peptide. Initially, this should be set as the highest possible shift between runs. This tolerance can be later narrowed by checking the RT shift distribution figure in the result.
- LFQ requires sample grouping to generate a result node. The “Sample Groups” section allows for this group selection. All available samples are listed on the left window. For experiments with multiple samples, the search bar can be used to look for a specific sample. After selecting a sample, it can be added to a new group by clicking the  button or to an existing group by clicking the  button and selecting that option from the drop-down menu. If you select multiple files, you can add each group one to a new sample by clicking the  button. Display all samples in each group or just the individual groups by selecting the Expand and Collapse buttons, respectively. Samples can be removed individually or removed all at once by clicking the  button or the **Remove All** button, respectively.
- The **Advanced** option opens another window that enables the manual selection of a reference sample and training samples. The reference sample will be used as the center to do the retention time alignment. The training samples are used to calculate the feature vector quality. For training samples, biological replicates are preferred. Auto detection should fit most cases of both reference and training samples.



- To attach the identification results to the label free quantification result, click the Attach ... button in the Attach Identification section. The “Attach identification results” dialogue lists all available identification results. Select the identification results to be attached and the FDR threshold to be used, then click the Attach button. All selected results will be listed in the Attach Identification section. If multiple identification results for the same fraction are selected, then only the first identification result in the list will be used. Use the Move Up and Move Down buttons to set the order in which the identifications will be attached.



## Note

If identification results from multiple samples are attached for LFQ, PEAKS can combine the ID results and re-do the protein grouping and generate LFQ results.

## 3. Understanding the LFQ Results

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

- Summary:** The outline of the 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, *Data Visualization* 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.
- Peptides:** The quantified peptides listed in a table. Extracted ion chromatogram (XIC), sample features and retention time alignment also displayed for each peptide.

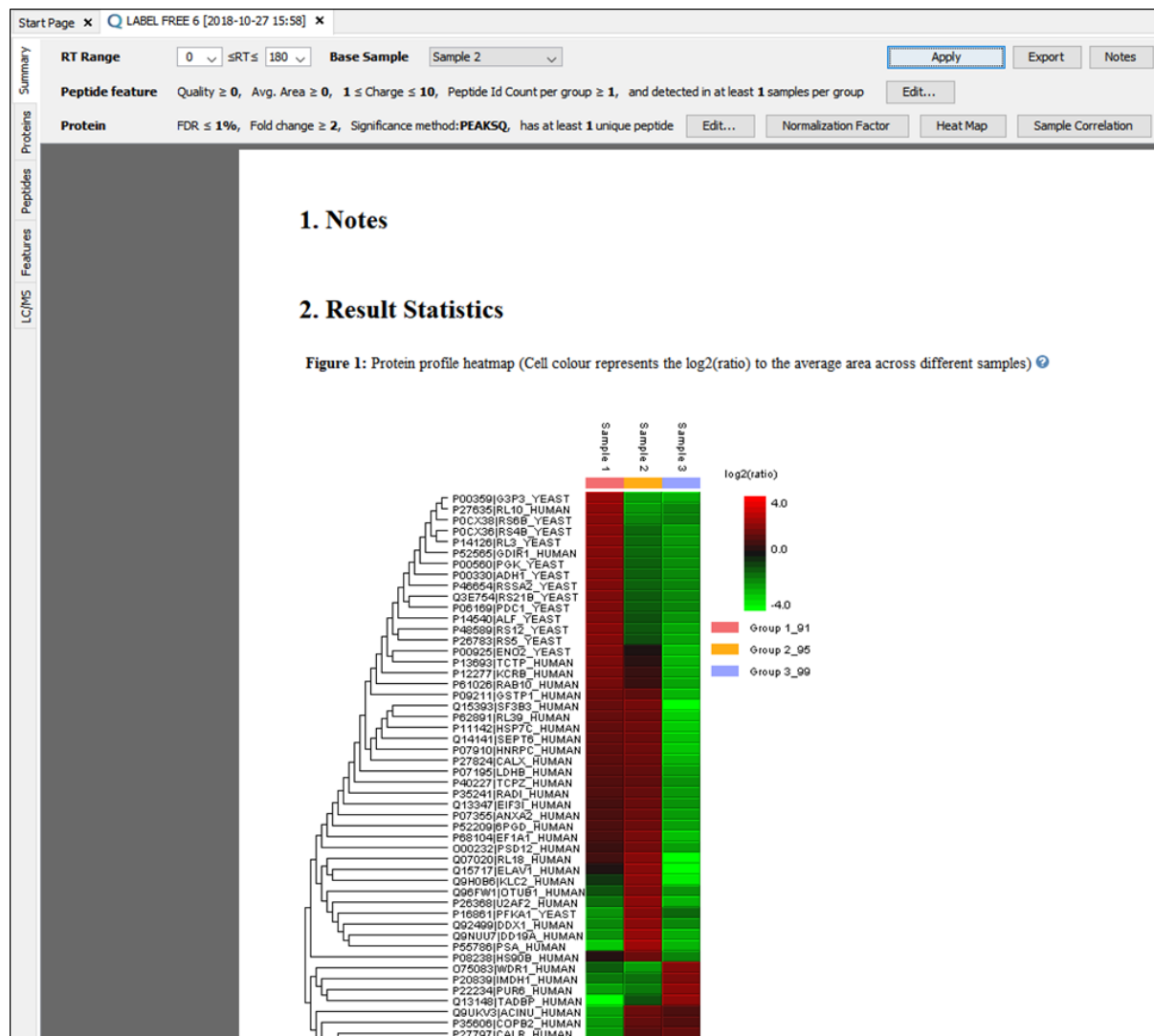
### 3.1. Summary View

The "Summary" view provides three main functions:



1. **Result Filtration:** This is achieved by specifying the filtration rules in the area at the top of the Summary View. The filtration function is discussed in Section 3.2.1, “Protein Table”.
2. **Summary Report:** Several statistical charts assist the user to get an overall picture of the results, assess the result quality, and examine the reliability and reproducibility of the experiment. This function is the focus of this section.
3. **Result Exporting:** This is achieved by clicking the "Export" button at the top of the summary view. The exporting function is discussed in Chapter 19, *Exporting Data/Reports and Printing* Section 5.3, “Exporting Label-Free Quantification Results”.

The "Summary" view includes an expression profile with candidate proteins assorted in a heat map, result statistics tables, and a list of filtration parameters. To add a summary note, click on the Notes button to open a “Notes Entry” editor in which you can edit the notes. These notes will be displayed in the “Summary” view.



### 3.1.1. Result Filtration

PEAKS provides an extensive set of tools to filter Label Free Quantification results according to the user's requirements. There are four categories of options in the “Summary” view: **RT Range**, **Base Sample**, **Peptide feature**, and **Protein**.



### 3.1.1.1. RT Range

Feature vectors within the selected retention time range will be included in the results while everything else will be removed. Select **Apply** within the filtration window of the “Summary” view to include the selected RT range in the analysis.

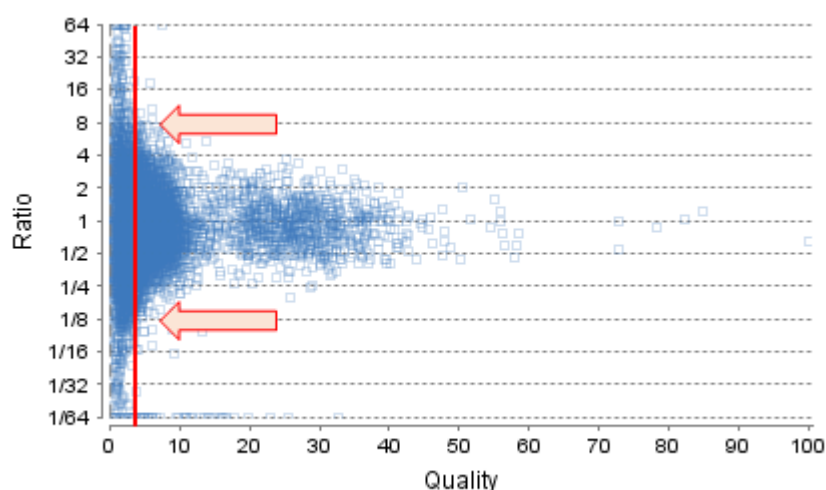
### 3.1.1.2. Base Sample

Select the sample that will act as the reference. The peptide area corresponding to this reference label will be the denominator in all of the feature area ratio calculations. Select **Apply** within the filtration window of the “Summary” view to include the selected base sample in the analysis.

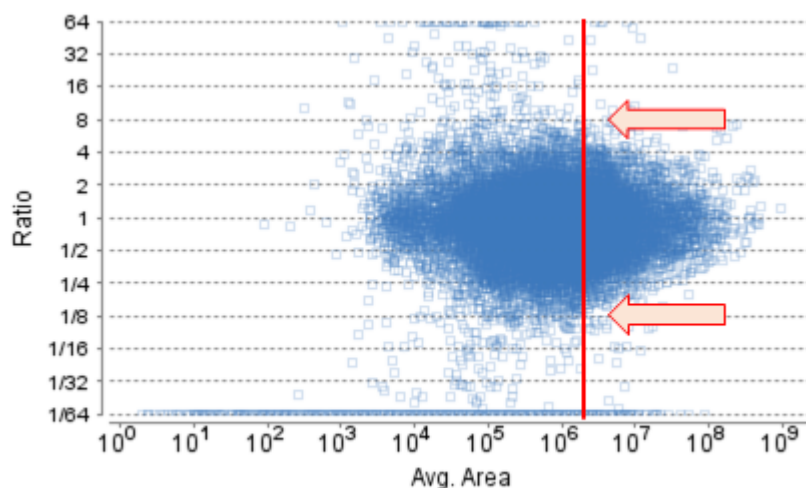
### 3.1.1.3. Peptide Feature

The following parameter dialogue appears when the **Edit...** button is selected.

- **Quality:** The higher quality of a peptide indicates that the peptide is more quantifiable. Factors that affect the quality score include m/z difference, RT difference, XIC shape similarity, isotope distribution and shape similarity, signal-to-noise ratios, frequencies and feature intensities. An appropriate threshold value can be selected by checking the ratio-quality figure. The rule of thumb is to set the quality threshold around a fold change of 8, as shown in this labelled version of Figure 3a) of the “Summary” view. Only feature vectors with a quality score above this threshold are used to calculate the quality score of a peptide.



- **Ave. Area:** This filter is directly applied to the MS signal intensity. Higher intensity peptide features have been shown to be more reproducible across replicates. The rule of thumb is to choose the average area near a fold change of 8, as shown in this labelled version of Figure 3b) of the “Summary” view. Only peptides with an average area above this threshold will be used in protein quantification.



- **Peptide ID Count:** This filter sets the minimum number of tandem spectra identified per group of the given peptide.
- **Charge between:** Only peptides with feature vectors that all fall within this range will be used in protein quantification.
- **Confidently detected sample:** A peptide is more quantifiable when it is detected in more samples. This filter sets the minimum number of sample(s) per group that the peptide has an abundance value. The rule of thumb is to set this value to half the number of samples per group.

Select **Apply** within the filtration window of the “Summary” view to include all selected filters in the analysis.

#### 3.1.1.4. Protein Filter

The following parameter dialogue appears when the **Edit...** button is selected. These filters determine what will be included in a protein group significance calculation, as well as what significance score thresholds will be set. For a protein group to have a significance score, it must have been identified in a sample/label that has already been placed into a group (in either the Tools > Quantification window or in Experiment Settings).

- **Significance:** Only protein groups with a significance above this threshold will be listed in the “Protein” view. The significance score is calculated as the  $-10\log_{10}$  of the significance testing p-value. PEAKS provides two significance testing methods: PEAKS Q and ANOVA. A significance score threshold of 20 is recommended, which equals to a significance testing p value of 0.01. Either this or “FDR (adjusted p-value)” can be selected to set a significance threshold.
- **FDR:** Adopt the Benjamin–Hochberg method to adjust the p-value to the false discovery rate for all protein groups that have already passed the other filters. Only protein groups with significance scores passing the calculated FDR will be listed in the “Protein” view. Either this or “Significance” can be selected to set a significance threshold.
- **Fold Change:** Only protein groups at or above this fold change threshold will be listed in the “Protein” view.

- **Unique Peptides:** Only protein groups with at least this many unique peptides will be listed in the “Protein” view.
- **Modified Form Exclusion:** The expression level of modified peptides might be different than the overall expression level of a protein. In such cases, including modified peptide in the protein ratio calculation might lead to incorrect results. By checking this box, if a peptide has modified and unmodified forms, both will not be used as candidates, if a peptide only has one modified form, it is still a candidate, but if a peptide has more than one modified form, all forms will not be used as candidates.
- **Significance Method:** One of two significance calculations can be applied to the list of protein groups: PEAKS Q or ANOVA. Scores will only be calculated for protein groups identified from samples that are in groups. PEAKS Q is similar to the Significance B method employed by the paper referenced below. Additionally, support peptide similarity is considered. When there is only one sample in each group, PEAKS Q is suggested to be used for significance testing. ANOVA requires that at least two samples be included in each particular group. The ANOVA method employed is also referenced below. When there are replicates in each group, ANOVA is suggested to be used for significance testing. Log10 (protein abundance) is used for the significance testing calculation.

## Note

For more **details on PEAKS Q**, refer to the following paper: “*MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification.*” *Nat Biotechnol.* 2008, 26(12): 1367-1372.

For more **details on ANOVA**, refer to the following paper: “*On the comparison of several mean values: an alternative approach*” *Biometrika.* 1951, 38(3/4): 330-336.

The following parameter dialogue appears when the **Normalization Factor** button is selected.

**Normalization**

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

Sample	Factor	Expected Ratio
Disease1	-	1.0
Disease2	-	1.0
Disease3	-	1.0
WT4	-	1.0
WT5	-	1.0
WT6	-	1.0

**Protein List**

search

Previous Next

Protein

- A2ASS6|TTIN\_MOUSE
- Q3URH1|ZSWM8\_MOUSE
- Q3URK3|TET1\_MOUSE
- Q618S1|FGFR3\_MOUSE
- Q3V0Q6|SPAG8\_MOUSE
- Q6ZQL4|WDR43\_MOUSE
- Q80TM6|R3HD2\_MOUSE
- Q03142|FGFR4\_MOUSE
- Q0VBN2|DSEL\_MOUSE
- Q01097|NMDE2\_MOUSE
- P13542|MYH8\_MOUSE
- P13541|MYH3\_MOUSE
- Q5SX40|MYH1\_MOUSE
- Q5SX39|MYH4\_MOUSE

0 protein selected.  
(Right click to show popup menu)

OK Cancel

- **Use TIC:** By default, PEAKS uses the total ion current (TIC) of the samples to calculate the normalization factors. Normalized abundance is calculated from the raw abundance divided by the normalization factor.
- **Use internal standard proteins:** This option displays a list of identified proteins, which can be selected to create a normalization ratio. A search bar is included to enable the quick location of select proteins. Right-clicking in the Protein screen will open a window that will allow for the selection of multiple highlighted rows or the removal of all check-marked proteins. The Experiment drop-down menu can be opened to display the results of another sample. If this option is chosen but no proteins are selected, no Factors will be listed for any of the samples,

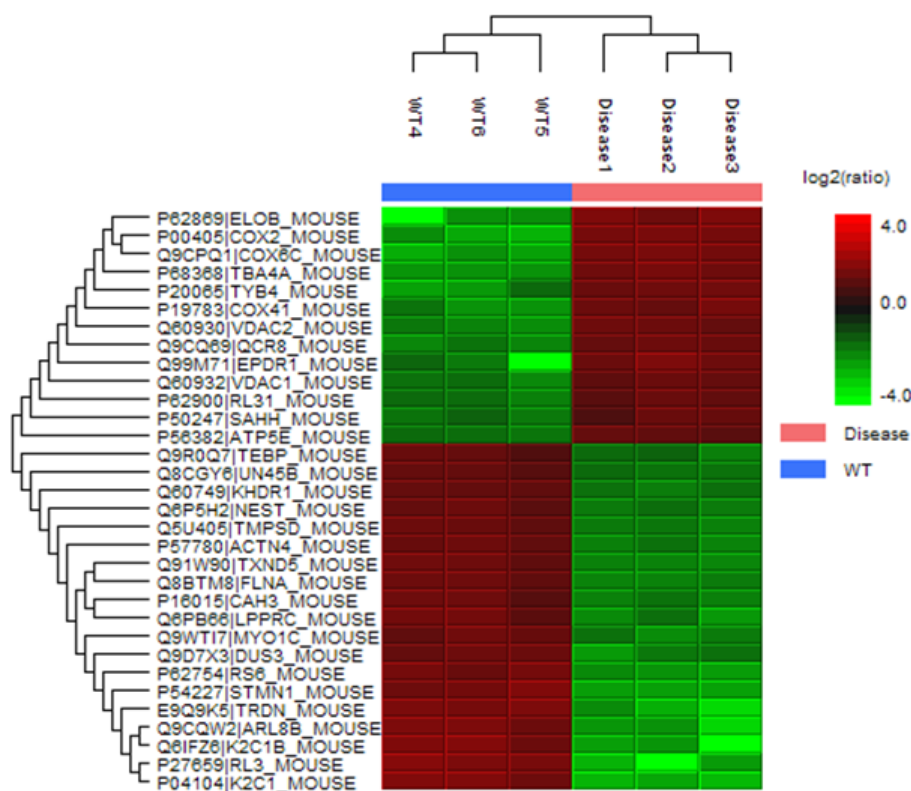
- **Manual input:** This option performs the same default calculation as auto normalization ('Use TIC') but also allows for the manual change of each label's expected ratios. After changing a ratio and hitting **Enter** on the keyboard, PEAKS will recalculate the Factor for the non-referenced label(s). The Experiment drop-down menu can be opened to display the results of another sample.
- **No Normalization:** All samples are automatically assigned Factors of 1.0.

Select **Apply** within the filtration window of the "Summary" view to include all selected filters and Normalization choices in the analysis.

### 3.1.2. Summary Report

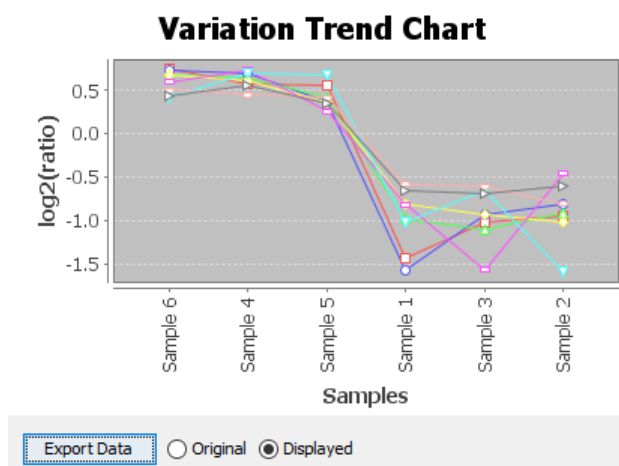
#### 3.1.2.1. Heat Map

The heat map displays the protein groups that passed the filters. The relative protein abundance is represented as a heat map of the representative proteins of each protein group. These representative proteins are clustered if they exhibit a similar expression trend across the samples. The hierarchical clustering is generated using a neighbour-joining algorithm with a Euclidean distance similarity measurement of the log2 ratios of the abundance of each sample relative to the average abundance. Similarly, the conditions in different samples are clustered if they exhibit a similar expression trend across the protein groups. An interactive protein profile heat map is opened in a new window when the **Heat Map** button that is located in the filtration options. Right-click on the heat map in this window to save a copy of it.



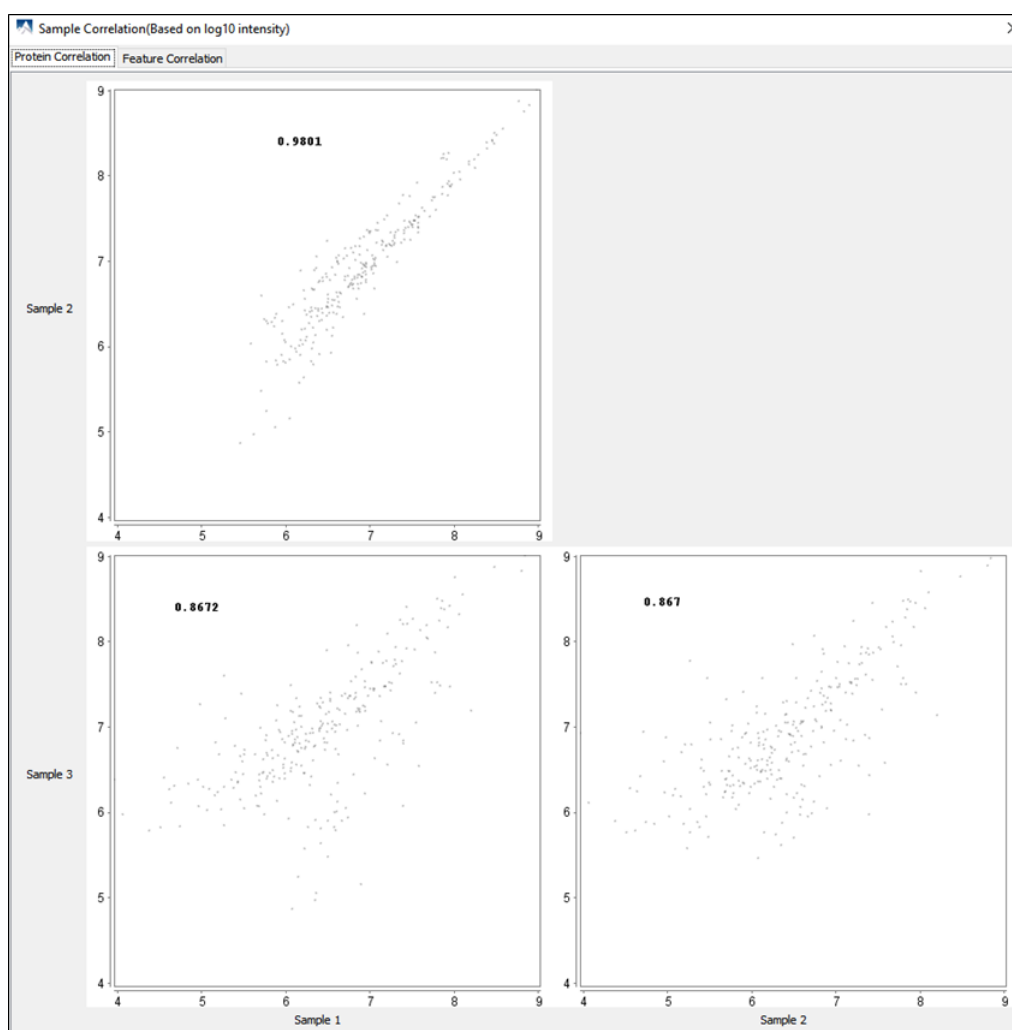
In the interactive heat map (by clicking the Heat Map button), move the cursor over any point to highlight the protein accession name and the sample/label to which it matches. Right-click on the heat map and click **Save as...** to export a PNG file of the heat map.

Move the cursor over a node in the protein cluster tree, to select a set of proteins and left click to open the **Variation Trend** chart that displays the expression level of the selected proteins in the samples. Right-click on the chart and select **Save as...** to export the chart as image. Click **Export Data** to save the expression variation as a CSV file.



### 3.1.2.2. Sample Correlation

To view the abundance correlation plot for the detected features and proteins of the two selected samples, click the Sample Correlation button from the “Summary” view title bar. Sample correlation plots will be displayed in a pop-up window. Correlation plots between all possible samples are shown. The Pearson’s correlation score indicates the reproducibility of the experiment based on a selected pair of samples. To export the image right click on pop-up window and select **Save as...**

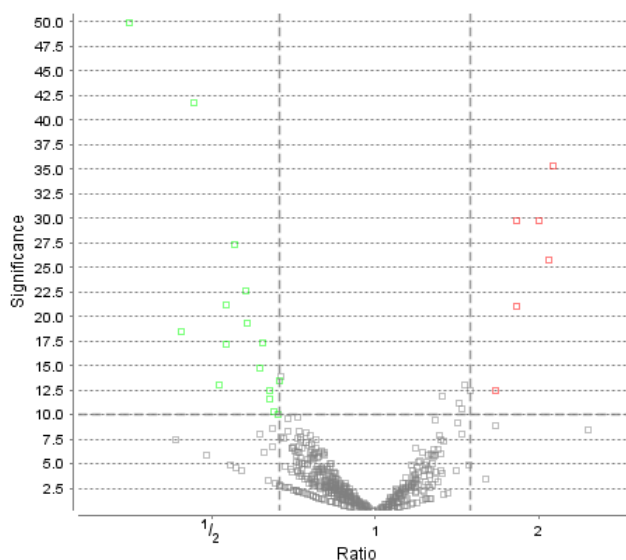


### 3.1.2.3. Volcano Plot

A volcano plot combines a statistical test with the magnitude of the change enabling quick visual identification of those data-points that display large-magnitude changes and that are also statistically significant. The volcano plot in Figure 2(a) plots significance versus fold-change of the quantified proteins.

- **Horizontal broken grey line:** The selected significance threshold
- **Vertical broken grey lines:** The selected fold change thresholds


Plotting points in this way results in two regions of interest in the plot: those points that are found towards the top of the plot that are far to either the left or the right side. These represent values that display large magnitude fold changes (hence being left or right of center) as well as high statistical significance (hence being towards the top). In this figure, the ratio is the group ratio set with respect to the base group. If more than two groups are present, the max ratio is used instead.

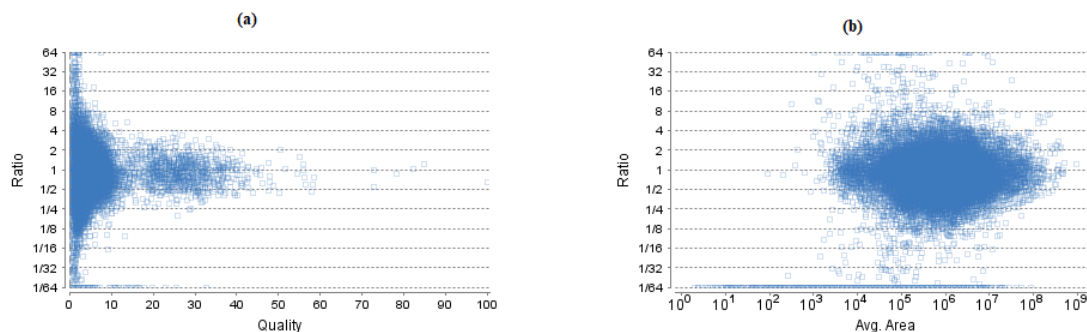


The “Protein” view contains an interactive volcano plot that can be hidden or displayed. Markers for the proteins that are above the set significance threshold will be displayed in colour. Scrolling the cursor over a marker will highlight it and, subsequently, clicking on this marker will highlight the protein in the interface.

### 3.1.2.4. Feature Vector Ratio Distribution

Figures 3(a) and 3(b) plot the distribution of feature vector ratios by quality and by average area, respectively. These plots help to determine the threshold values for the quality and intensity filters to obtain a reliable section of the analysis results. The rule of thumb is to set the quality threshold around an 8-fold change, as shown below in Figure 3(a). The quality filter is usually the better filter to use but under some circumstances, it may be better to apply the filter to the MS signal directly. The rule of thumb is still to choose the average area around an 8-fold change. In both figures, the ratio is the group ratio with respect to the base group. If more than two groups are present, each ratio value depicts a data point in the figures.

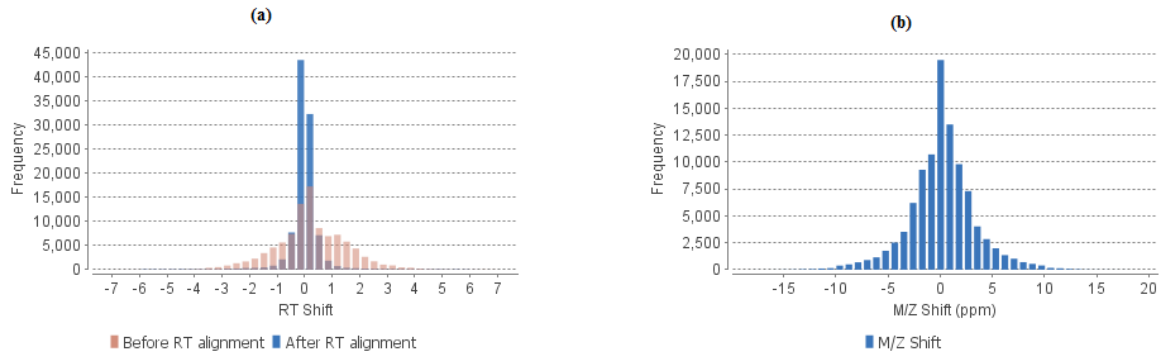
**Figure 3.** The distribution of feature vector ratio: (a) By quality. (b) By intensity. 



### 3.1.2.5. RT and m/z Shift Distribution

Figure 4(a) plots the retention time shift distribution to the base sample. Figure(b) plots the distribution of the m/z shift (in ppm) of the data with respect to the base sample. These two figures can be used to refine the parameter settings for label free quantification

**Figure 4. (a) RT shift distribution; (b) M/Z shift distribution.**



### 3.1.2.6. Summary Tables

Tables 1 and 3 list the data and the filtered results, respectively. Table 2 lists the feature vector and protein result filtration parameters.

**Table 1. Result filtration parameters.**

Retention time	$\geq 0$
Retention time	$\leq 140$
Feature fold change	$\geq 1$
Quality	$\geq 0$
Intensity	$\geq 0E0$
Charge	$\geq 1$
Charge	$\leq 10$
Confident sample number	$\geq 1$
With peptide ID	true
Normlization	Use TIC
Protein significance	$\geq 0$
Protein fold change	$\geq 1$
Confident unique supports	$\geq 1$
Top protein groups	50

**Table 2. Statistics of filtered result.**

Features	7184
Features with ID	2714
Feature vectors	1025
Feature vector with ID	1025
Protein groups	50

**Table 3. Search Parameters**

Quantification type: Label free quantification
Mass Error Tolerance: 20.0 ppm
Retention Time Shift Tolerance: 6.0 min
Dependent on PID: 10, FDR = 1%
Samples: Control1_1, Control1_2, Control2_1, Control2_2, Treatment12_1, Treatment12_2, Treatment24_1, Treatment24_2,
Reference Sample: Auto detect

Tables 4 and 5 list the quantification and instrument parameters, respectively.



**Table 4.** Quantification parameters.

Quantification Type: ICAT/SILAC  
 Quantification Mass Tolerance: 15.0ppm  
 Quantification RT Range: 0.5min.  
 -10LogP Threshold: 15.0  
 Reporter Ion Type: MS2  
 R to P Convert enabled: true  
 Light: R (100.0%) K (100.0%)  
 Heavy: R(+10.01) (100.0%) K(+8.01) (100.0%)  
 Group Settings in Parameters:  
 Group 1: Sample 1 : Heavy  
 Group 1: Sample 2 : Heavy  
 Group 1: Sample 3 : Heavy  
 Group 2: Sample 1 : Light  
 Group 2: Sample 2 : Light  
 Group 2: Sample 3 : Light  
 DB Dependency: 5  
 Current Group Information:  
 Group 1: Sample 1 : Heavy  
 Group 1: Sample 2 : Heavy  
 Group 1: Sample 3 : Heavy  
 Group 2: Sample 1 : Light  
 Group 2: Sample 2 : Light  
 Group 2: Sample 3 : Light

**Table 5.** Instrument parameters.


Fractions: 20120812\_ShAr\_QE1\_ARL\_NX4\_mix\_sample\_1.raw, 20120812\_ShAr\_QE1\_ARL\_NX4\_mix\_sample\_2.raw, 20120817\_ShAr\_QE1\_ARL\_NX4\_mix\_sample\_2\_REP\_3.raw  
 Ion Source: ESI(nano-spray)  
 Fragmentation Mode: CID, CAD(y and b ions)  
 MS Scan Mode: FT-ICR/Orbitrap  
 MS/MS Scan Mode: Linear Ion Trap

### 3.1.3. Result Exporting

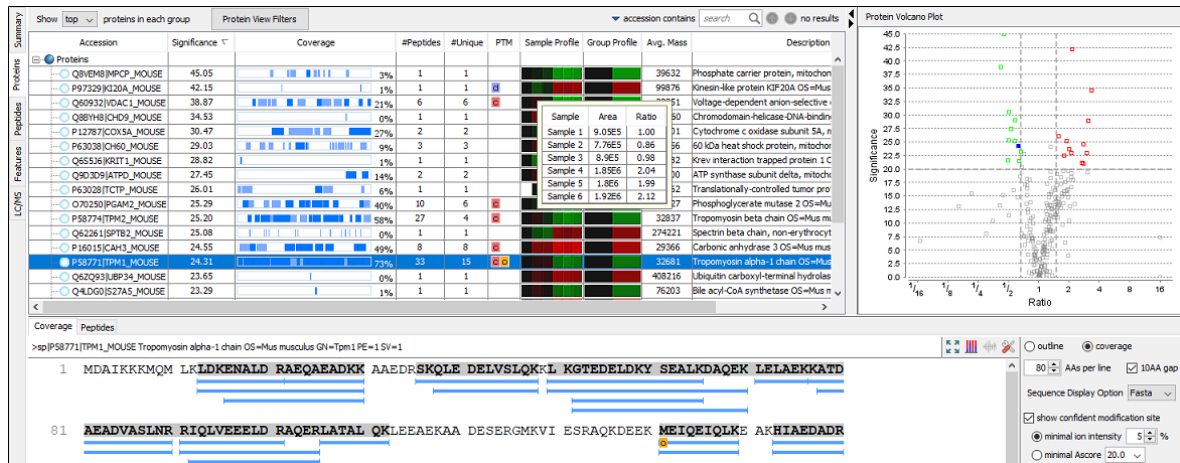
PEAKS label free quantification results can be exported to various HTML and text formats by clicking the **Export** button within the filtration options in the Summary View. The Summary View and the images from various views can also be exported to various HTML and text formats. Refer to Chapter 19, *Exporting Data/Reports and Printing* Section 5.3, “Exporting Label-Free Quantification Results” for details.

## 3.2. Protein View

The "Protein View" lists all the quantified proteins present in the sample, characterizes each protein at the amino acid level, and lists the supporting peptide features of each protein. It has six components:

- **Protein Table:** Lists all the quantified proteins present in the sample.
- **Protein View Filters:** Allows filtering the protein table to select specific proteins. For more information refer to Chapter 9, *Peptide, PTM, and Mutation Identification (PEAKS DB, PEAKS PTM, and SPIDER)* Section 3.4.1, “Protein View Filters”
- **Protein Volcano Plot:** Click on a marker that is above the significance threshold (the horizontal dashed line) set in the Summar View. Upon selection, the corresponding protein in the Protein Table will be highlighted.
- **PTM Profiling:** Click on  to open ptm profiling. For more information refer to Chapter 11, *PTM Profiling* Section 3, “PTM Profiling for Label Free Quantification and Reporter Ion Quantification Results”.
- **Coverage:** Characterizes the protein sequence at the amino acid level. All the PTMs and mutations occurred on the protein sequence will be displayed together with the MS/MS data supporting the inference.
- **Features:** Lists the features identified from the selected proteins in the protein table.





### 3.2.1. Protein Table


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

- **Significance:** The protein significance is set according to the rules set in the Summary View, refer to **Section 3.1.1.4, “Protein Filter”**.
- **Sample Profile:** The protein abundance in each sample relative to the average is log2 transformed and depicted as a heat map. Hold the cursor on a profile to view the sample channels (labels), areas, and ratios with respect to the base label selected in the Summary View.
- **Group Profile:** The protein abundance in each group relative to the average is log2 transformed and depicted as a heat map. The Group Profile is calculated by summing up the group areas of its used supporting peptides. Hold the cursor on a profile to view the group channels (labels), areas, and ratios with respect to the group which contains the base label selected in the Summary View.

See Chapter 9, *Peptide, PTM, and Mutation Identification (PEAKS DB, PEAKS PTM, and SPIDER)* Section 3.4.2, “Protein Table” for more details on other entries of the protein table

### 3.2.2. Protein Volcano Plot

This interactive volcano plot enables the selection of a protein in the table. Protein markers that are above the significance threshold can be selected. Upon selection, the corresponding protein in the Protein Table will be highlighted. Proteins located in the top-right and top-left sections of this plot (above the significance threshold and to the right and left of the fold change thresholds) are considered statistically significant. Green points are down-regulated while red points are up-regulated. Grayed points are filtered out as described in Section 3.2.2, “Protein Volcano Plot” and from the protein view filter. The volcano plot facilitates identification of these significant proteins.

Use the  button to expand or collapse the volcano plot.

### 3.2.3. Protein Coverage

The coverage tab characterizes the selected protein sequences at the amino acid level. See Chapter 9, *Peptide, PTM, and Mutation Identification (PEAKS DB, PEAKS PTM, and SPIDER)* Section 3.4.4, “Protein Coverage” for details.

### 3.2.4. Support Peptides

The supporting peptides assigned to the protein are shown under the “Peptides” tab.

Coverage		Peptides																	
		Peptide	Used	Candidates	Quality	Significance	Avg. ppm	Avg. Area ↓	Sample Profile	Group Profile	SMA_Area	Wild_Type_Area	Max Ratio	#Vector	Start	End	PTM		
1	<input checked="" type="checkbox"/>	K.LDKENALDRAEQAEADK...	•	<input checked="" type="checkbox"/>	2.37	60.00	1.8	1.2E8			1.71E8	7.79E7	2.19	2	13	30			
2	<input checked="" type="checkbox"/>	K.YSQMEDKYEEIK.V			2.06	54.81	1.3	6.1E7			7.28E7	4.98E7	1.46	3	214	226			
3	<input checked="" type="checkbox"/>	R.KYEEVAR.K			2.98	29.66	1.0	5.6E7			E7	6.37E7	1.31	1	161	167			
4	<input checked="" type="checkbox"/>	K.LDKENALDR.A		<input checked="" type="checkbox"/>	0.98	9.12	1.7	4.3E7			E7	4.58E7	1.26	4	13	21			
5	<input checked="" type="checkbox"/>	R.RIQLVEELDR.A			1.21	60.00	2.1	4.2E7			E7	5.41E7	1.77	2	91	101			
6	<input checked="" type="checkbox"/>	K.LKGTEDLDKYSEALDKA...			1.81	23.45	1.6	3.4E7			E7	3.01E7	1.26	2	50	70			
7	<input checked="" type="checkbox"/>	K.EKYYEEIK.V			1.74	0.23	1.9	3E7			E7	3.04E7	1.01	2	218	226			
8	<input checked="" type="checkbox"/>	K.KATDAEADVASNR.R			1.62	44.48	0.8	2.7E7			E7	2.26E7	1.43	2	77	90			
9	<input checked="" type="checkbox"/>	R.KLVIESDLER.A	•	<input checked="" type="checkbox"/>	1.81	6.22	1.1	2.5E7			E7	2.42E7	1.10	2	168	178			
10	<input checked="" type="checkbox"/>	K.HIAEDADR.K	•	<input checked="" type="checkbox"/>	2.11	60.00	1.3	2.2E7			E7	3.06E7	2.24	1	153	160			
11	<input checked="" type="checkbox"/>	R.IQLVEELDR.A			1.37	9.80	4.0	2E7			2.36E7	1.73E7	1.36	1	92	101			
12	<input checked="" type="checkbox"/>	R.LATALQK.L			2.27	60.00	1.2	2E7			2.61E7	1.39E7	1.88	1	106	112			

This table contains the following information on the quantifiable support peptides for the selected protein in the protein table.

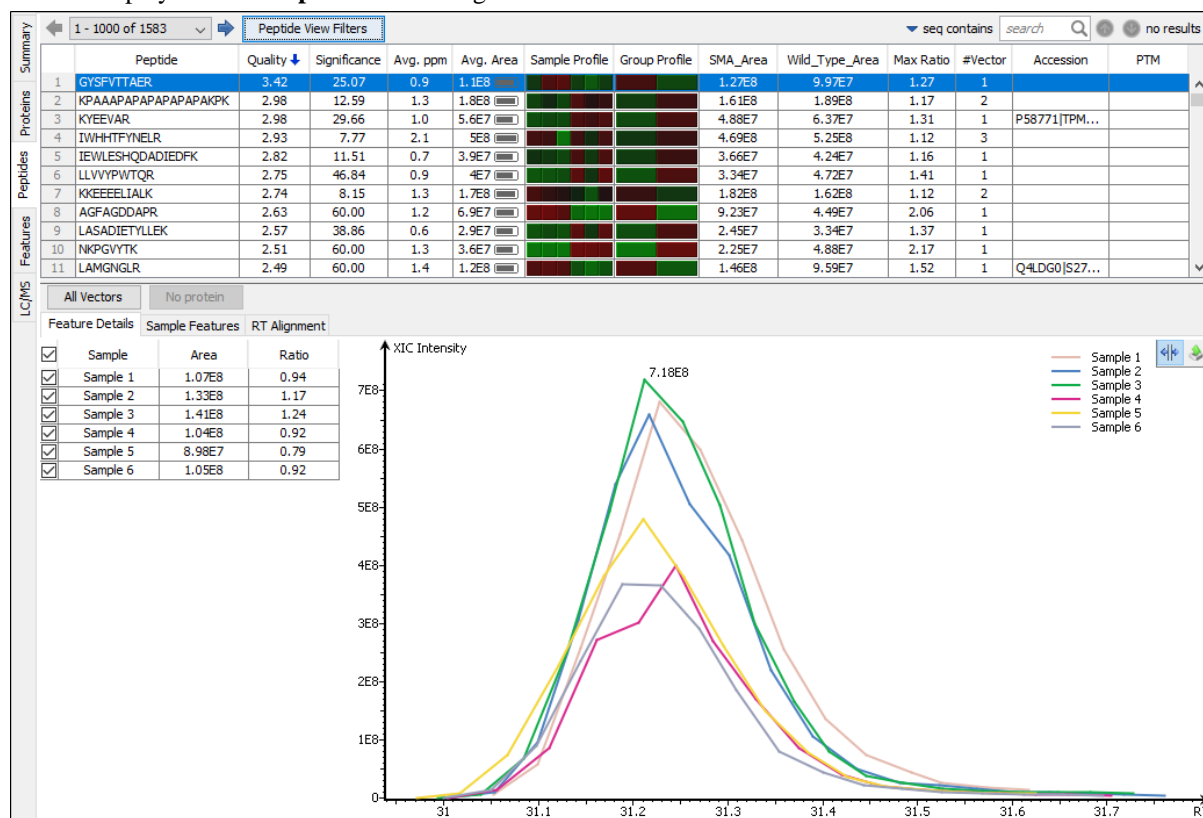
- **(Checkbox):** Select the peptides to appear within the Coverage Tab as blue bars.
- **Peptide:** The amino acid sequence of the peptide, as determined in PEAKS Search. A modified residue is followed by a pair of parentheses enclosing the modification mass. All non-unique peptides and ones not passing the filter will be excluded automatically.
- **Used:** Shows whether this peptide is used to calculate the protein profile.
- **Candidates:** Select which peptides will be used as candidates to quantify the protein. The top three peptides from the selected candidates will be involved in quantification. Click **Apply** button in the **Summary View** to make the change of selection effective. Select one row and right-click to show the pop-up menu, use **Reset the candidate selection** command to reset the candidate selection for all proteins.
- **Quality:** This defines how well the feature vector can be quantified. More details are explained in the Section 3.1.1.3, "Peptide Feature".
- **Significance:** It assumes that for all the peptides with the same quality, their ratios follow a log-normal distribution. Different qualities may have log-normal distributions with different variances. The two-tail p-value is calculated as the significance of the feature vector using the log-normal distribution corresponding to its quality and its maximum group ratio. The  $-10\lg p$  value is calculated from this and displayed as the significance in the column.
- **Avg. ppm:** A peptide can have one or more feature vectors. Each feature vector has its own ppm calculated from the mass error of each feature. The average mass error of a peptide is the average of mass errors of all the feature vectors. This column displays the average mass error in ppm.
- **Avg. Area:** Each feature in the feature vector has its own intensity area under the XIC curve. The average area of a feature vector is the the average area of all the features in the feature vector. This column shows the average area of a peptide by summing up the average areas of its feature vectors.
- **Sample Profile:** The peptide abundance in each sample relative to the average is log2 transformed and depicted as a heat map. Hold the cursor on a profile to view the samples, the peptide areas in samples, and the corresponding ratios with respect to the base sample (or average) selected in the **Summary View**.
- **Group Profile:** The peptide abundance in each group relative to the average is log2 transformed and depicted as a heat map. Hold the cursor on a profile to view the groups, the peptide areas in groups, and the corresponding ratios with respect to the group which contains the base sample selected in the **Summary View**.
- **Group Area:** This column shows the peptide area in a group that is calculated by summing up the group areas of its feature vectors.
- **Max Ratio:** This shows the maximum of the fold changes between a pair of groups.
- **#Vector:** This shows the number of quantifiable feature vectors of a peptide.
- **Start:** This shows the position of the first residue of the peptide that covers the protein.

- **End:** This shows the position of the last residue (inclusive) of the peptide that covers the protein.
- **PTM:** The types and the numbers of modifications present in the peptide shown in color-coded icons.

Right-click on the peptide feature to show the pop-up menu similar to the one in the PEAKS DB supporting peptides tab.

### 3.3. Peptide View

The quantifiable feature vectors with the same sequence are grouped together to get the list of quantifiable peptides and are displayed in the **Peptide View** along with their intensities.



#### 3.3.1. Peptide Table

The peptide table contains the list of quantifiable peptides along with relevant information including the protein accession. The view is similar to the support peptide table described in Section 3.2.4, “Support Peptides”. The Peptide table entries may be filtered by using the **Peptide View Filters** button, for more information refer to Chapter 9, *Peptide, PTM, and Mutation Identification (PEAKS DB, PEAKS PTM, and SPIDER)* Section 3.5.1, “Peptide View Search and Filter Options”.

#### 3.3.2. All Vectors Table


All quantifiable feature vectors on a selected peptide can be viewed by selecting the peptide from peptide table and then clicking the **All Vectors** button. A pop up table shows the details of the feature vectors that contains the following columns:

- **Quality:** This defines how well the feature vector can be quantified. More details are explained in the Section 3.1.1.3 Section 3.1.1.3, “Peptide Feature”.
- **ppm:** The mass error of a feature vector is calculated from the mass error of each feature in that vector.
- **m/z:** Each feature in the feature vector has its own precursor mass-to-charge ratio. This column displays the average precursor mass-to-charge ratio of all of the features in the feature vector.

- **RT range:** The retention time range of the features in the feature vector.
- **z:** The precursor charge of the feature vector.
- **Avg. Area:** Each feature in the feature vector has its own intensity area under the XIC curve. This column displays the average area of all the features in the feature vector.
- **Sample Profile:** The peptide abundance in each sample relative to the average is log2 transformed and depicted as a heat map.
- **Group Profile:** The peptide abundance in each group relative to the average is log2 transformed and depicted as a heat map.
- **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.
- **Id Count:** The total number of identification found for the peptide in fractions.

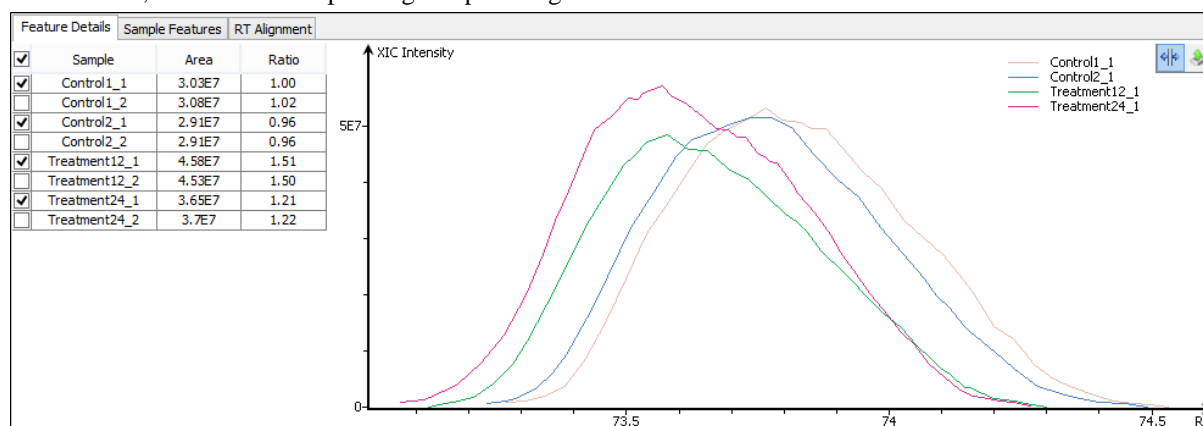
Selecting a feature vector from the table updates the **Feature Details**, **Sample Features**, and **RT Alignment** views accordingly.

### 3.3.3. Feature Details

The **Feature Details** tab contains the eXtracted Ion Chromatogram (XIC) chart that displays the shape of the selected peptide feature vector over the retention time range where they were identified. The table at the top-left corner displays both the area and the ratio of the features in the feature vector. This table also works as a control to select the features that will be used to draw the XIC plot and other plots in the “Sample Features” and “RT Alignment” tabs. The  button at the top-right corner of the tab can be used to toggle the plot data between, before, and after RT alignment.


#### Note

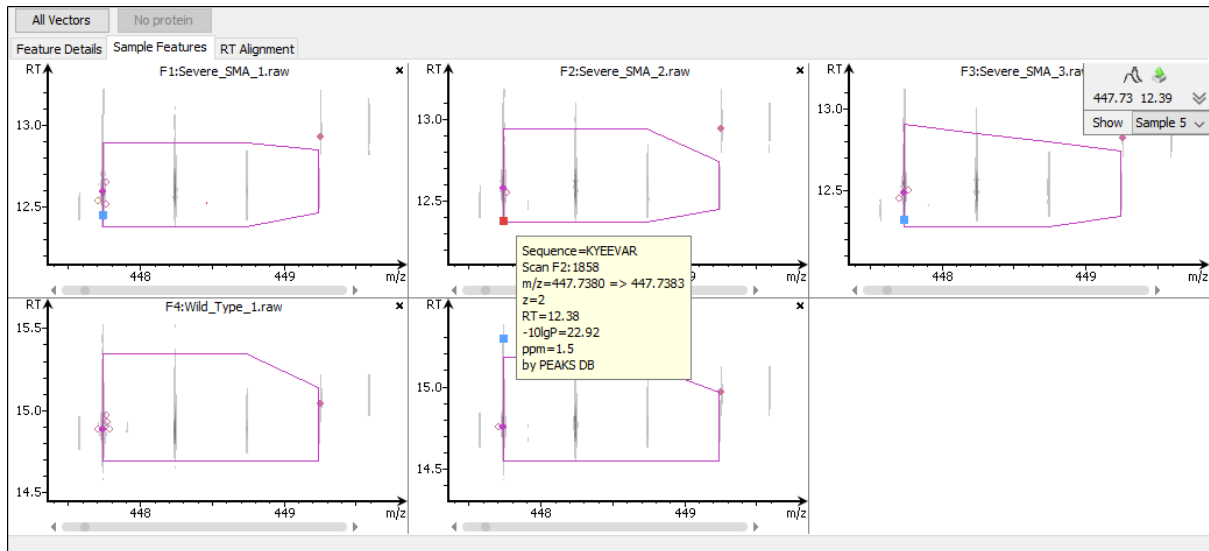
By default, the first features from each group are selected to be displayed in the plots. To display other features, select the corresponding sample using the check box in the table.




### 3.3.4. Sample Features

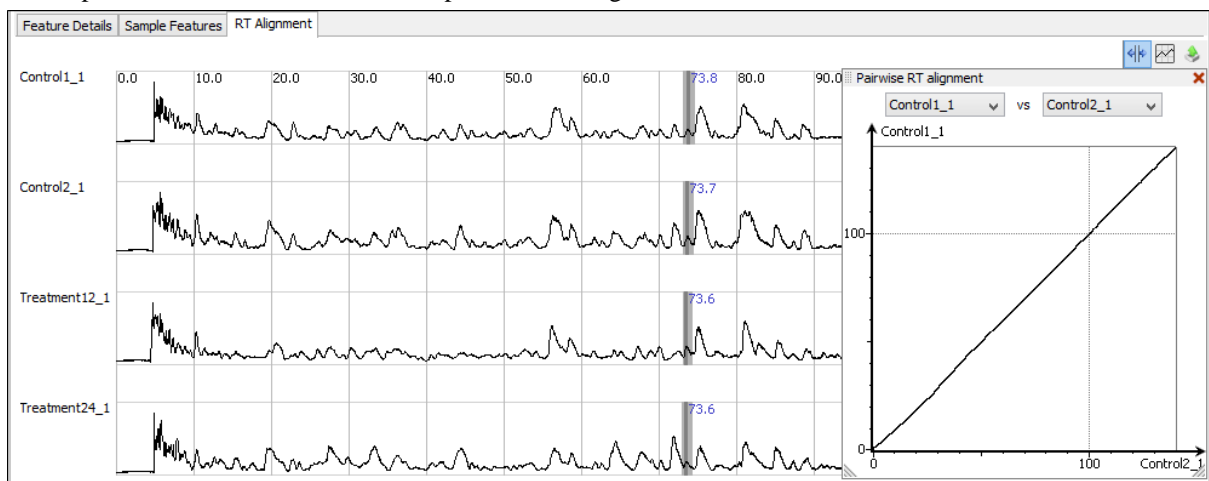
This tab displays a close up LC/MS view of the detected features from the selected samples. The feature boundary is shown by the bounding polygon. The center of the feature is denoted by a solid red dot on the monoisotopic peak cluster. Placing the cursor on this feature center displays a tool tip that shows the m/z value, the mean retention time, the charge state, and the intensity. The centers of the features are used to calculate the retention time alignment.

The  button at the top-right corner of the tab can be used to toggle between the 2D and the 3D LC/MS views. To open a closed sample select the sample using the drop down found on the right hand side.



### 3.3.5. RT Alignment

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



### 3.4. Features View

The **Features View** displays all of the quantifiable feature vectors and their intensities. The interface is similar to the **Peptide View** described earlier. See Section 3.3, “Peptide View” for details.

### 3.5. LC/MS View

This view displays the distribution of LC-MS signals/features in the fractions. See Chapter 5, *Data Visualization* Section 4, “LC/MS View” for details on LC/MS View.


# Chapter 16. Data Independent Acquisition (DIA) Data Analysis

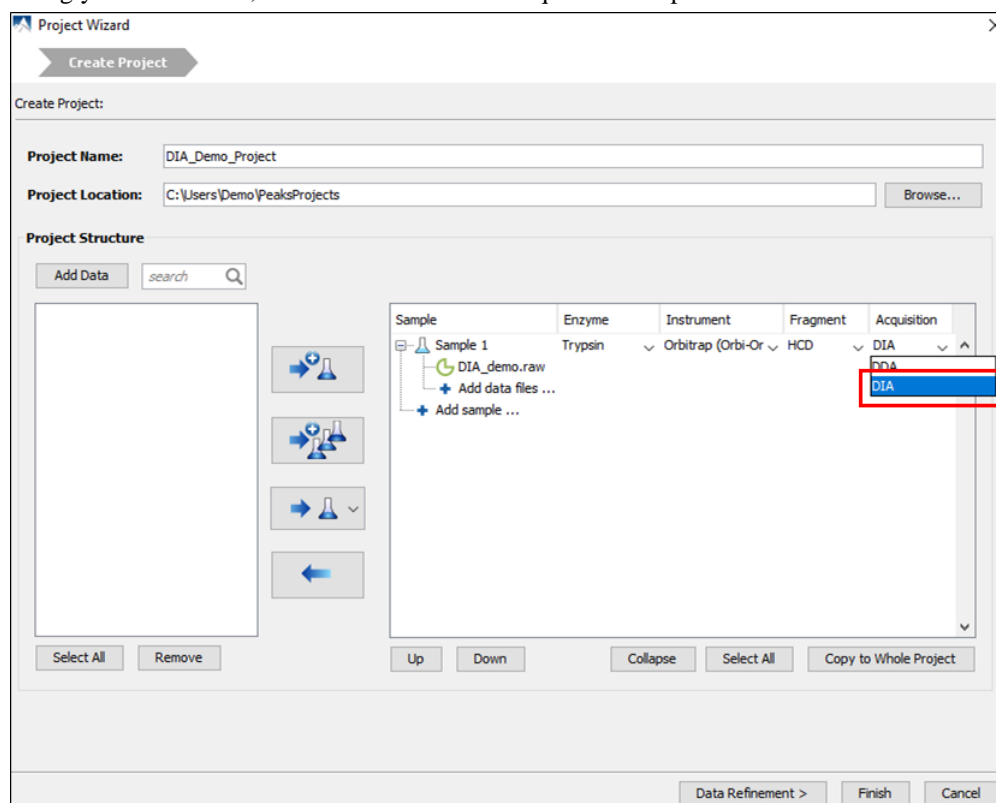
## 1. Overview

Data Independent Acquisition (DIA) is a relatively new method for generating MS/MS data. In recent years, DIA has become increasingly popular due to its parallel nature of acquiring fragment ions for all precursors within a selected  $m/z$  range. This untargeted extraction approach allows DIA to collect more information about fragment ions thus have better reproducibility comparing to the traditional data dependent acquisition.

With a feature-based identification approach, PEAKS X enables full DIA data analysis without the need of a spectral library. By analysing correlated precursor and fragment ion signals together from DIA scans, the peptide sequences are determined directly and this overcomes the limitations imposed by using a spectral library. PEAKS X supports the full work flow for DIA data, including *de novo* sequencing, database search, PTM and sequence variants identification, and quantification.

To create a new PEAKS project using DIA data, perform the following steps:

1. To create a new project, click the "New Project" button, , or select File > New Project from the general command menu to open the Project Wizard. Click the "Add Data" button and browse to the location of DIA data (Refer to Section 7, "Supported Data Formats" for configuration details to successfully load various raw data file types). The selected data from the list on the left pane can be setup into the PEAKS project when added to the right pane using the appropriate option. Ex. All of the selected files can be placed in a new sample using the "Create a new sample for selected files" button. For more details, refer to Section 6, "Adding Data to an Existing Project". Using the drop-down menus provided in each column, setup the parameters for each data file accordingly. For DIA data, select "DIA" from the Acquisition drop-down menu.




## Note

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


2. Click the “Data Refinement” button in the Project Wizard to continue with the data refinement step after loading the data; or select "Finish" to only load the data at this stage. Select the appropriate options based on the project. For details on data refinement, refer to Chapter 7, *Data Refinement*.

## Note

To perform data refinement on an existing project, click the Data Refinement icon, , from the tool bar after selecting a raw data node in the Project View pane.


3. Once the data refinement parameters have been set, click the "Identification" button in the Project Wizard to continue with a Database Search after the loaded data has been refined; or select "Finish" to only load the data and perform data refinement at this stage. For details on database searching, refer to Chapter 9, *Peptide, PTM, and Mutation Identification (PEAKS DB, PEAKS PTM, and SPIDER)*.

## Note

To run database search on an existing project, selecting either a project node, a sample node, or a *de novo* result node from the Project View pane and then click the PEAKS DB icon, , from the tool bar.

4. Once the identification parameter have been set, click the “Quantification” button in the Project Wizard to continue with a quantification analysis on the data after the identification search completes; or select "Finish" to perform all the steps up to the identification step. For details on the different quantification analyses, refer to Chapters 13, 14, and 15 for Chapter 13, *PEAKS Q - MS Level*, Chapter 14, *PEAKS Q - MS/MS Level* and Chapter 15, *PEAKS Q - Label Free* respectively.

## Note

To run quantification on an existing project, click the Quantification icon, , from the tool bar after selecting a project node, or a PEAKS search result node.

## 2. Data Visualization

After the project is created, the DIA data can be visually examined. To do so, double click the raw data file node or corresponding DATA REFINE node from the Project View pane. Like DDA data, PEAKS allows raw DIA data to be analyzed from three perspectives:

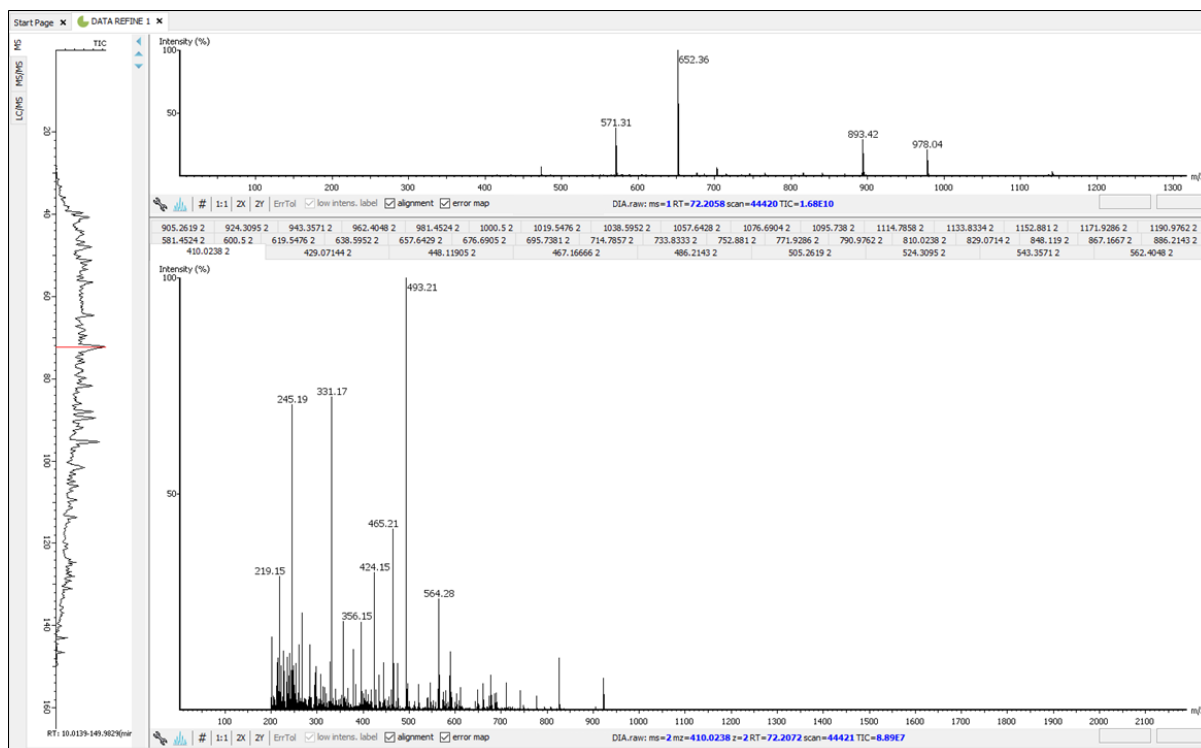
- **MS** - This section shows the TIC (total ion chromatogram) plot and all of the MS scans. For each MS scan, the corresponding MS/MS scans are also displayed.
- **MS/MS** - This section lists all the MS/MS scans. For each MS/MS scan, the corresponding features are listed below and associated number of features is displayed next to the MS/MS scan information header.
- **LC/MS** - This section provides a bird's eye view of the whole LC-MS dataset with detected features and MS/MS scans marked on refined data.

### 2.1. MS

The MS View contains the TIC and all of the scans. The total ion chromatogram (TIC) is displayed on the left of the MS view. The navigation buttons are highlighted in the figure. To collapse the TIC chart, click the left navigation button. Clicking on a specific position in TIC will display the corresponding survey (MS) scan. To navigate the



MS scans, use the up and down navigation buttons. The MS scans can also be navigated through using the up and down arrows of the keyboard. The tandem (MS/MS) scans associated with the selected MS scan are shown in the panel below, where each can be viewed by selecting the different tabs identified by their "m/z" information.



## Note

The navigation options within the MS and MS/MS view are the same as with the options to view DDA data. Refer to Chapter 5, Section 2 Section 2, "MS View" for details.

## 2.2. MS/MS

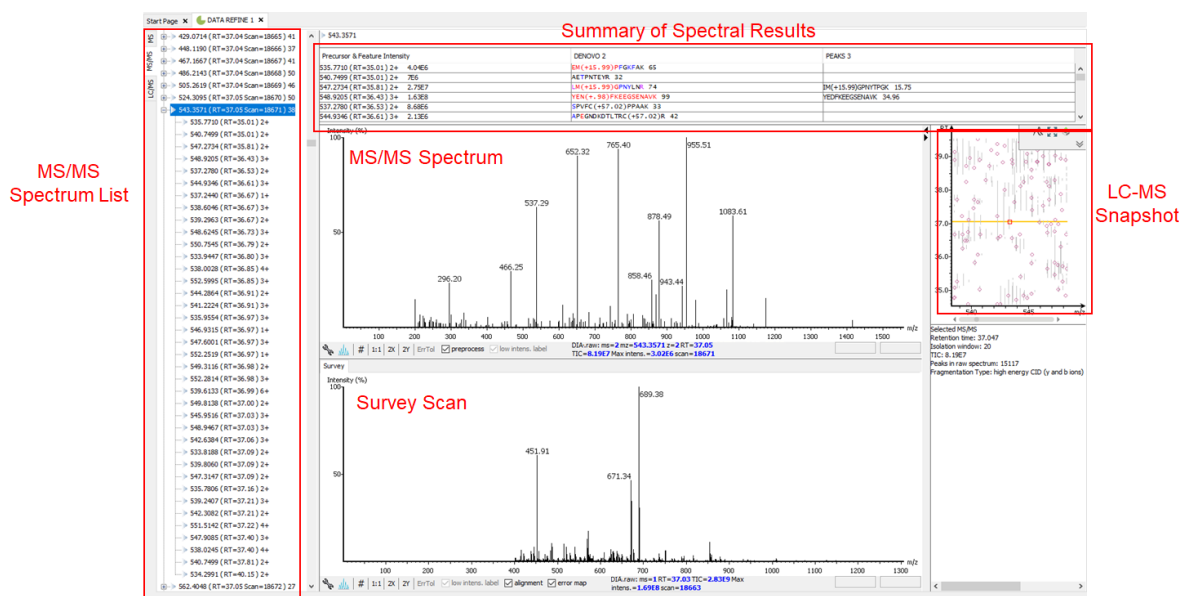
The MS/MS View shows the list of MS/MS spectra on the left. If data refinement has been performed on the data file, the number of associated features within the isolation window will be indicated at the end of the spectrum header [ex. "543.3571 (RT=37.05 Scan=18671) 38" from the image below indicates that there are 38 peptide features associated with the selected tandem scan]. In addition, when data refinement has been performed, an MS/MS isolation window header can be expanded and the associated features are listed below. When an MS/MS is selected from the list, the following information is provided to the right of the list: a summary of the spectral results, the MS/MS spectrum, the LC/MS snapshot, the Ion Match Table, and its Survey Scan. Similarly, when an individual peptide feature is instead selected from the list, the results for that feature are displayed. Zoom options are the same as described in the "MS View" section.

- **Summary of Spectral Results:** A table with a summary of the spectral results are displayed here, including the associated feature information, *de novo* sequence (if *de novo* sequencing has already been performed), and database ID sequence (if a PEAKS search has been performed).
- **MS/MS Spectrum:** A view of the tandem scan will be displayed here. If *de novo* sequencing and/or a PEAKS search has been performed for the data file, the scan will be annotated. By clicking the sequence results from the Spectral Result table, the corresponding annotation will be displayed. Users will also have the ability to turn on the alignment and error map for the spectrum view when *de novo* sequencing and/or a PEAKS search is performed.
- **LC-MS Snapshot:** The LC-MS Snapshot displays a zoomed-in view of the precursor scan in the LC-MS heat map 2-D view. The MS/MS scan is indicated by a red framed square and the span of the yellow line designates



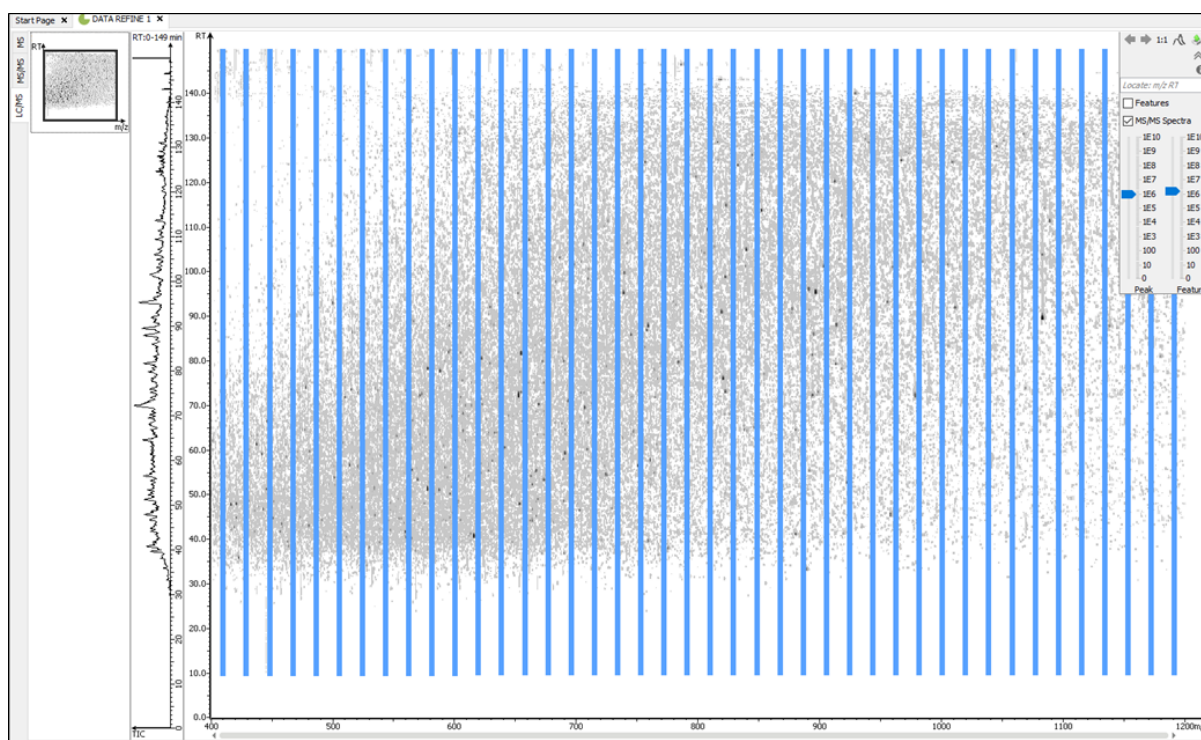
the isolation window. If data refinement has been completed, purple circles will also be displayed on the heat map, and selecting an individual feature from the list on the left will highlight the feature in purple.

- **Parent Scan:** The “Survey” tab displays the precursor scan with the best signal within the isolation window. The buttons that appear in this section are identical to those explained in Section 2, “MS View”.
- **Ion Table:** If *de novo* sequencing and/or a PEAKS search has been performed, an "Ion Match" tab will appear when one of the sequence results from the Spectral Results table is selected. The “Ion Match” tab in the bottom pane displays a table with possible fragment ions for the selected MS/MS scan. Each ion in the table shows the calculated mass, where if the fragment ion is present in the spectrum, its mass value is displayed in color. N-terminal ions are shown in blue and C-terminal ions are shown in red. A fragment ion is found when there is a matching peak within the mass error tolerance, as defined in the *de novo* sequencing parameters, and when relative intensity of the matching peak is at least 2%. The ion types displayed in the table can be configured in "Spectrum Annotation Settings" (Section 3.2.2, “Spectrum Annotation”).



## 2.3. LC-MS

The LC/MS view shows the distribution of LC-MS signals/features as a heat map. Placing the cursor on the heat map will show the m/z value and the retention time of that point.

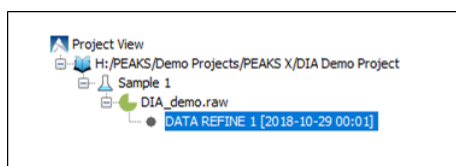


### 3. Understanding PEAKS Results

After analysis is complete, PEAKS will attached result nodes accordingly in the Project View pane . Double-click a result node to load the analysis.

#### 3.1. Data Refine

Data refinement can perform peptide precursor association for the MS/MS isolation windows . Once data refinement is complete, a “Data Refine” node will be generated in the project tree. Double click the data refine node to view the refined data. For details on data refinement, refer to Chapter 7, *Data Refinement*.



The data refinement analysis report uses the same views as the data visualization. The detected features can be pointed out in the views by selecting “Features” checkbox in the views. The feature centers are marked as purple circles. Placing the cursor over a feature center shows the feature details in a popup.

#### Note

Multiple features can be selected by holding the **Ctrl** key and left click on the feature center.

#### 3.2. De Novo

*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 from DIA data. For details on denovo sequencing, please refer to Chapter 8, *Peptide De Novo Sequencing*.

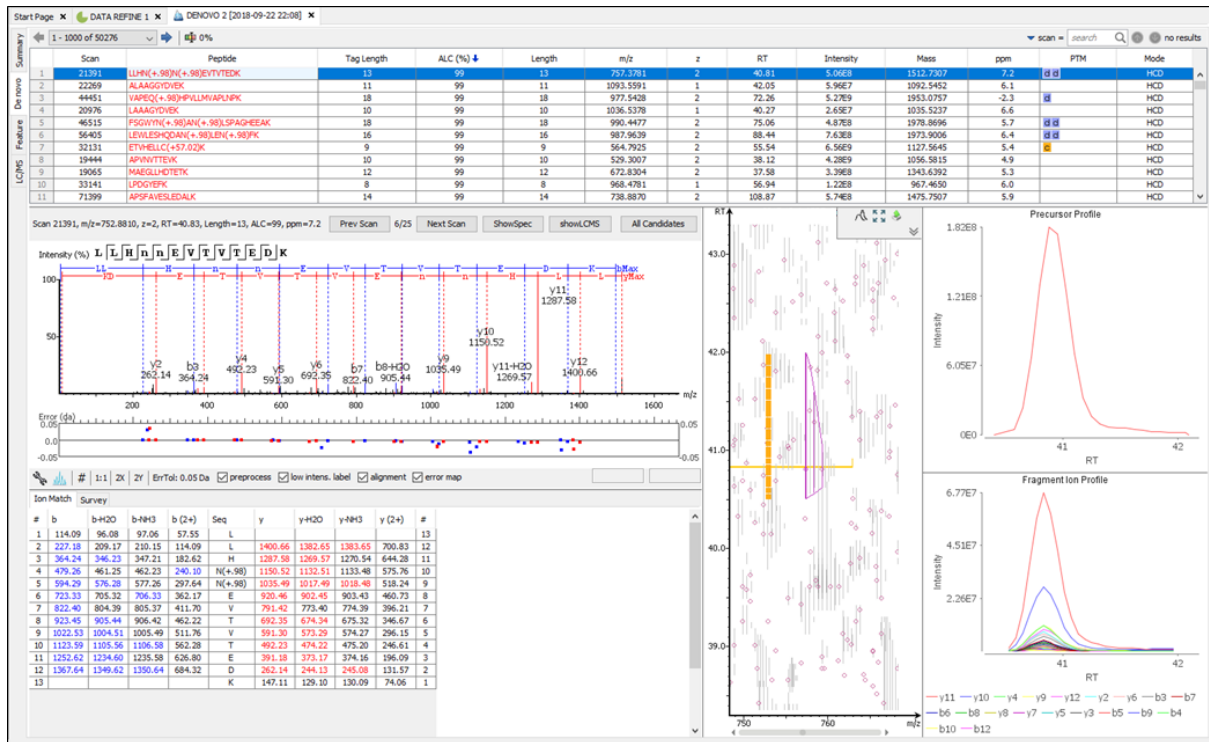
The *de novo* analysis results is organized into four sections:

### 3.2.1. Summary View

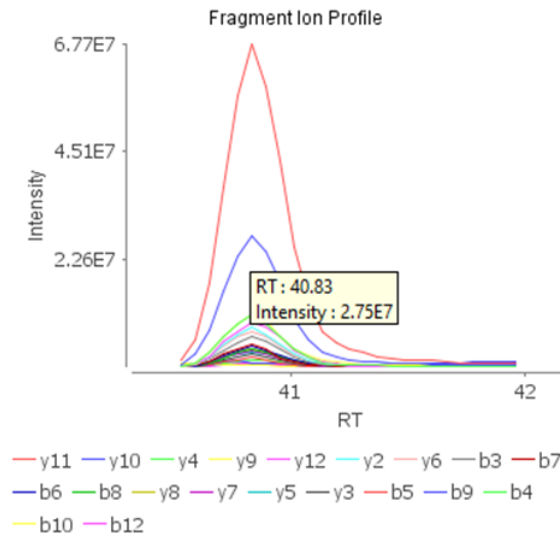
The "Summary View" reports key statistics as an overview of the result. For details on the *de novo* node Summary View, refer to Section 3.1, "Summary View"

### 3.2.2. De Novo Peptide View

The "De Novo View" displays the *de novo* sequencing results in greater detail, as shown in the next figure. The table on the top displays *de novo* sequences, while the bottom section provides additional information about the peptide-spectrum match. In comparison to a DDA *de novo* sequencing analysis result, there is one additional component to the De Novo Peptide View. For DIA data, a Fragment Ion Profile is displayed under the Precursor Profile and the LC-MS Snapshot is moved to the left and spans from the top to bottom under the *de novo* peptide table.



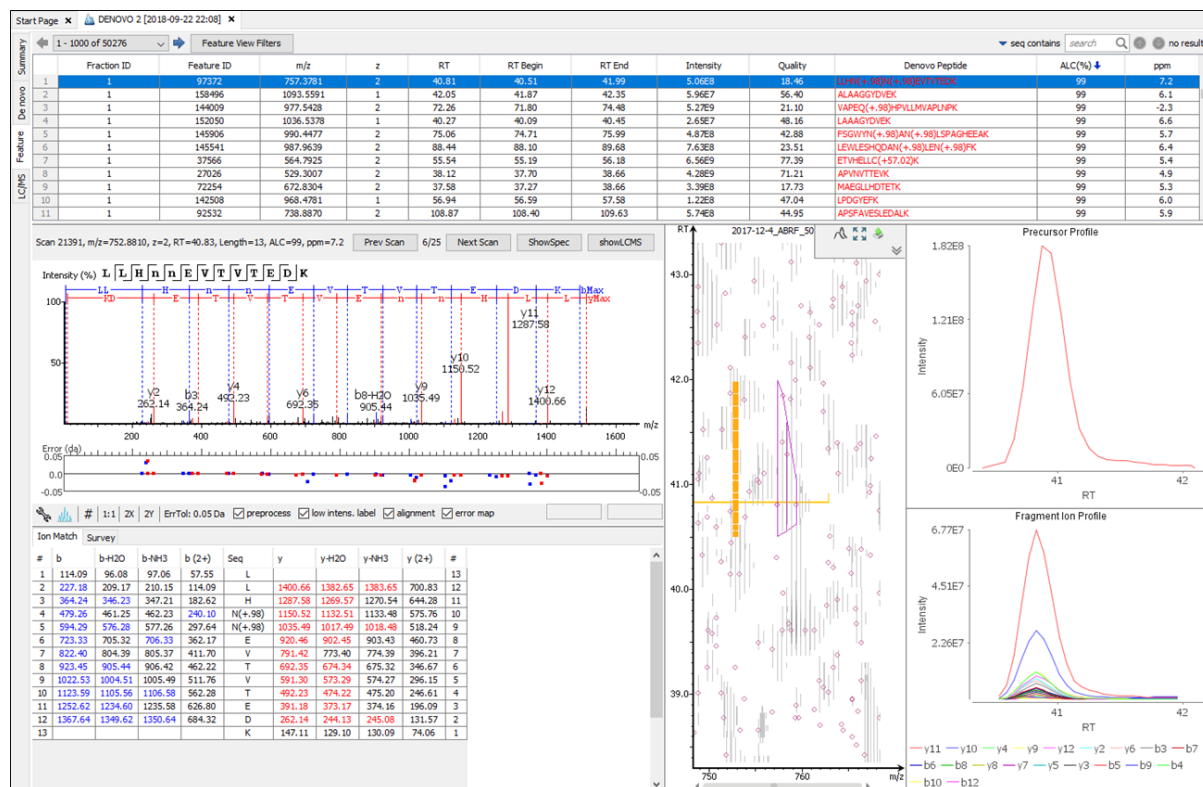
- Fragment Ion Profile:** The Fragment Ion Profile, or eXtracted Ion Chromatogram (XIC) of the fragment ions, graphs the elution profile of the top 20 fragment ion signals over the retention time. Like the Precursor Profile, the graph is plotted with the retention time along the x-axis and the intensity along the y-axis. Hovering over the XIC curve will display the RT and intensity coordinates.



For the remainder of the details about the De Novo Peptide View, please refer to Chapter 8, Section 3.2 Section 3.2, “De Novo Peptide View”.

### 3.2.3. Feature View

The "Feature View" displays all the feature information in greater detail, as shown in the next figure. The table on the top displays features, while the bottom section provides additional information about the feature, precursor profile, and if available peptide-spectrum match.

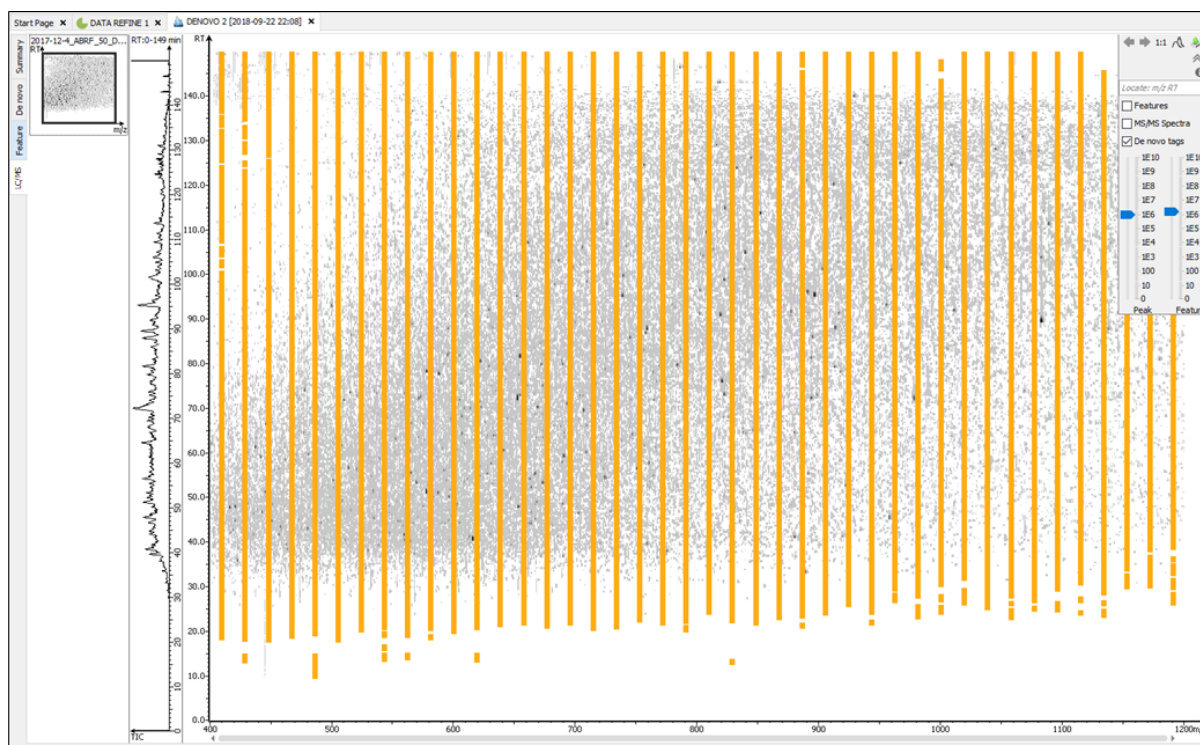


## Note

Please refer to Section 3.3, “Feature View” for details about the Feature View.

### 3.2.4. LC-MS

The view is same as in the data visualizations. For details, please refer to Section 2.3, “LC-MS”.



### 3.3. PEAKS DB

PEAKS DB node provides a complete set of database search tools for in-depth protein analysis. In this section, we present the differences in the DB result on ion mobility data from the traditional LC-MS data. To know more about PEAKS DB Searches and results, please refer to Chapter 9, *Peptide, PTM, and Mutation Identification (PEAKS DB, PEAKS PTM, and SPIDER)*.

The *PEAKS DB* analysis report is presented in five pages:

#### 3.3.1. 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. Please refer to Section 3.2, "Summary View" for details.

#### 3.3.2. Feature View

The "Feature View" lists all detectable features resulting from the filters set by the current settings in Summary View and Feature View Filters. For each feature, the feature information is displayed and the database or *de novo* peptide result. The result can be sorted by any column.

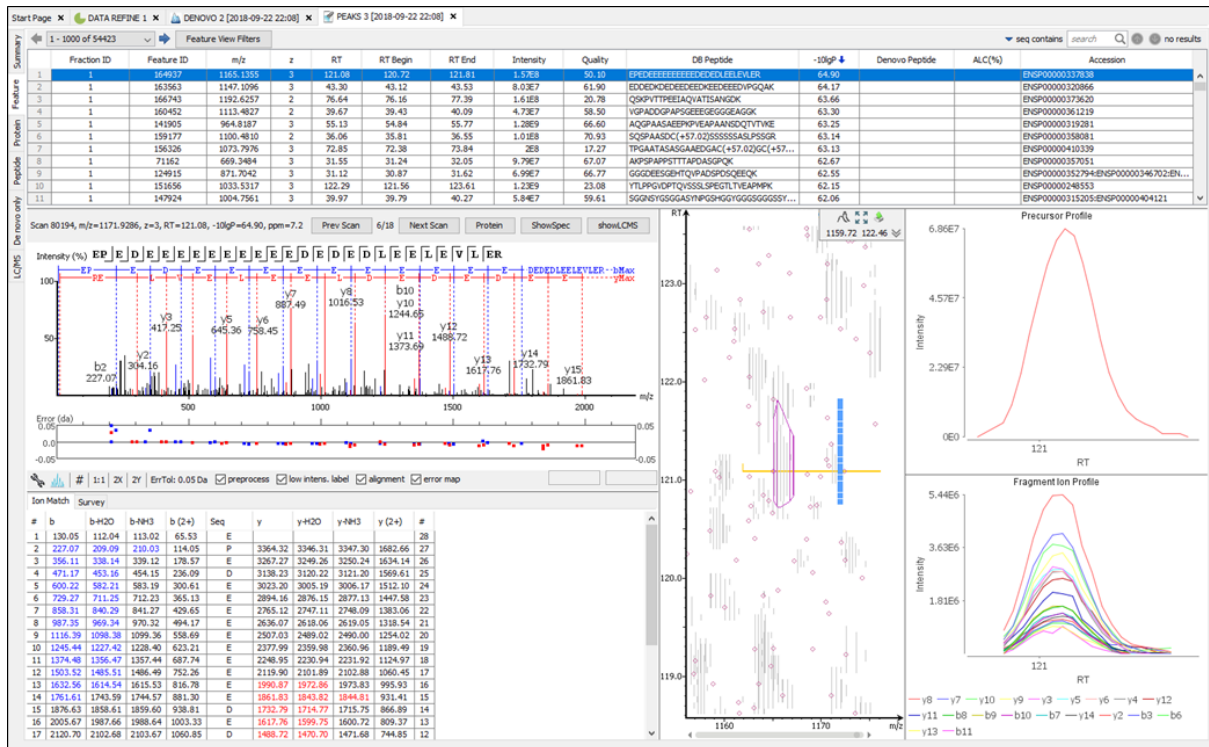
#### 3.3.3. Protein View

The "Protein View" presents the list of proteins identified in the project. For each protein, further information is displayed including protein sequence coverage and supporting peptide information. Please refer to Section 3.4, "Protein View" for details.

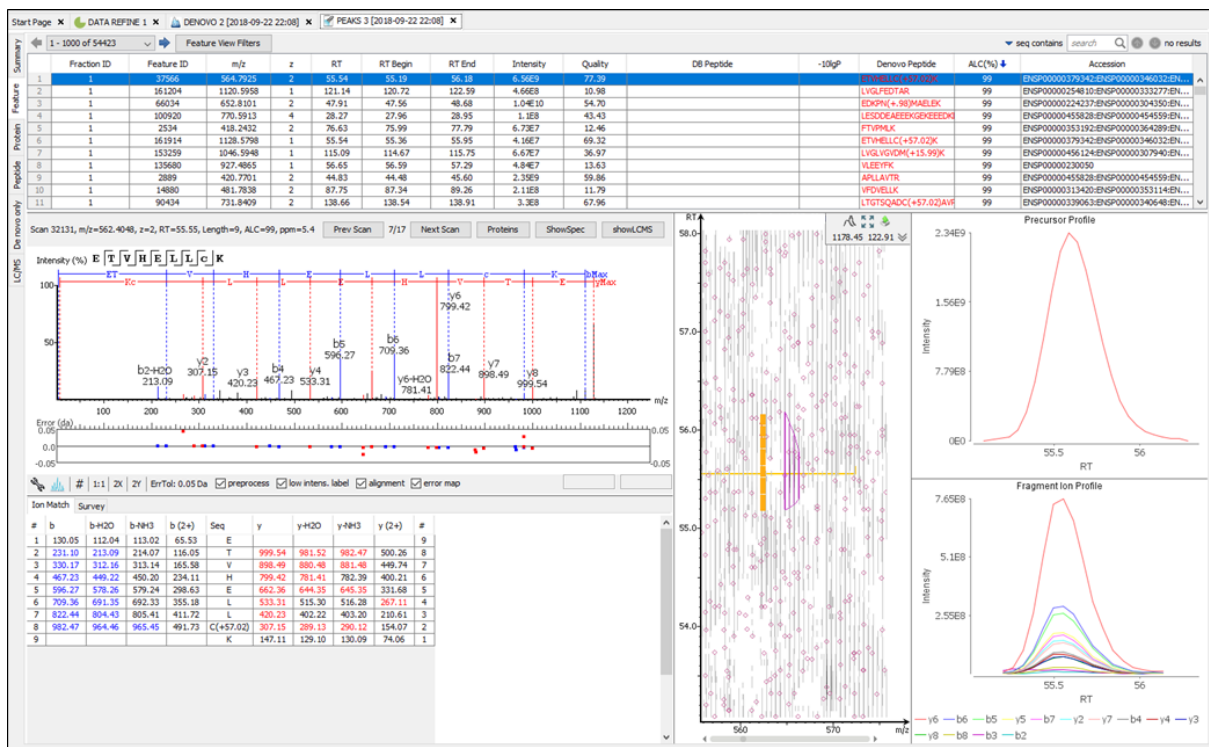
#### 3.3.4. Peptides View

Peptides table displays a list of peptide identifications. In comparison to a DDA database search result, there is one additional component to the Peptides View. For DIA data, a Fragment Ion Profile is displayed under the Precursor Profile and the LC-MS Snapshot is moved to the left and spans from the top to bottom under the peptides table. By default, the table is sorted by the  $-10\lg P$  score from highest to lowest.

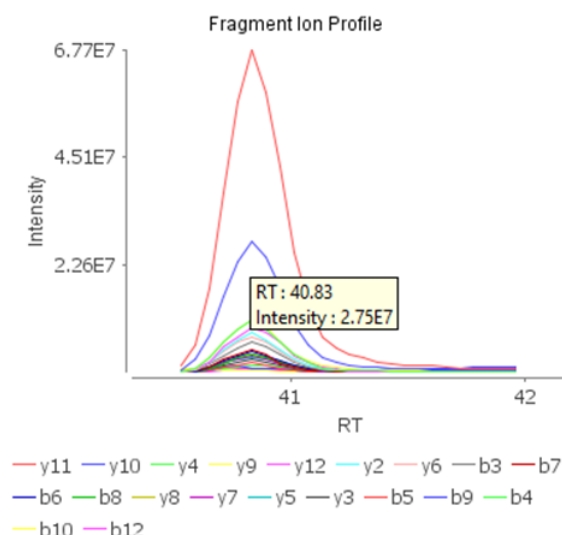




If a feature was able to match to a database sequence, the database peptide will be referenced in the "DB Peptide" column. However, if no database sequence was matched to the peptide feature, the de novo sequencing results will be displayed instead.



- Fragment Ion Profile:** The Fragment Ion Profile, or eXtracted Ion Chromatogram (XIC) of the fragment ions, graphs the elution profile of the top 20 fragment ion signals over the retention time. Like the Precursor Profile, the graph is plotted with the retention time along the x-axis and the intensity along the y-axis. Hovering over the XIC curve will display the RT and intensity coordinates.



## Note

For the remainder of the details about the Peptides View, please refer to Section 3.5, “Peptide View”

### 3.3.5. De Novo Only View


This shows a list of quality peptide sequences detected by *de novo* sequencing that remain unidentified by the PEAKS database search. Please refer to Section 3.6, “De Novo Only View” for details.

### 3.3.6. LC-MS View

The view is same as in the data visualization. For details, please refer to Section 2.3, “LC-MS”.


## 4. Label Free Quantification (LFQ)

PEAKS supports the Label Free Quantification (LFQ) analysis of DIA data. For details on LFQ analysis, please refer to Chapter 15, *PEAKS Q - Label Free*.

To perform an LFQ analysis, select a PEAKS project or data file node in the “Project View” pane. Click the PEAKS Q icon, , from the tool bar or select the Tools > Quantification command from the menu. Select “Label Free” check box to perform a LFQ analysis.

## 5. Precursor Ion Quantification (e.g. SILAC and iCAT)


PEAKS can perform precursor ion quantification analysis for DIA data with the optional PEAKS Q module. Both user-defined and commercial labels (eg. SILAC and iCAT) are supported by PEAKS Q. For details on precursor ion quantification setup and data analysis, please refer to Chapter 13, *PEAKS Q - MS Level*, as the setup and functions are similar to any other type of MS data.

To perform a precursor ion quantification analysis, select a PEAKS project in the “Project View” pane. Then, click the PEAKS Q icon, , from the tool bar or select the Tools > Quantification command from the menu. Select “Precursor Ion Quantification eg. SILAC” check box to perform precursor ion quantification on the PEAKS project.

## 6. Reporter Ion Quantification (e.g. iTRAQ and TMT)

PEAKS can perform reporter ion quantification analysis for DIA data with the optional PEAKS Q module. Both user-defined and commercial labels (eg. iTRAQ and TMT) are supported by PEAKS Q. For details on reporter

ion quantification setup and data analysis, please refer to Chapter 14, *PEAKS Q - MS/MS Level*, as the setup and functions are similar to any other type of MS data.

To perform a reporter ion quantification analysis, select a PEAKS DB/PTM/SPIDER result node in the “Project View” pane. Then, click the PEAKS Q icon, , from the tool bar or select the Tools > Quantification command from the menu. Select “Reporter Ion Quantification eg. iTRAQ/TMT” check box to perform reporter ion quantification.




---

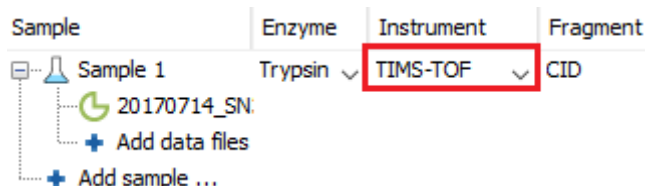
# Chapter 17. Ion Mobility Data Analysis

## 1. Overview

Ion Mobility Mass Spectrometry (IM-MS) analyzes complex mixtures and separates the ions according to their mobilities. Due to the superior separation capabilities, IM-MS can resolve ions that may be indistinguishable by traditional mass spectrometry. The PEAKS IM module of the PEAKS Studio software package has the ability to load ion mobility data from Bruker's timsTOF Pro instrument using PASEF acquisition procedure. The data can be analyzed using *de novo* sequencing, database search (PEAKS DB), unexpected modification (PEAKS PTM) and sequence variants (SPIDER) search, and labelled and label free quantification.

To create a new PEAKS Ion Mobility project and analyze the data, perform the following steps:

1. To create a new project, click the new project button  to open the project wizard. Click the “Add Data” button, browse to the location of timsTOF data and select the .d folder(s) that contain .tdf and .tdf\_bin files. Add data to samples using one of the add buttons (Refer to Section 6, “Adding Data to an Existing Project” for details for adding data to a project). Make sure to select “TIMS-TOF” from the Instrument list for each sample.



### Note


PEAKS IM module is preconfigured to load timsTOF data.

### Note

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


2. Click the “Data Refinement” button in the project wizard to continue with the data refinement step after loading the data, or "Finish" to only load the data at this stage. For details on data refinement, refer to Chapter 7, *Data Refinement*.

### Note

To perform data refinement on an existing project, click the data refinement tool bar icon  after selecting a data node in the Project View pane.


3. Click the "Identification" button in the project wizard to continue with a Database Search after loading the data, or "Finish" to only load the data and perform data refinement at this stage. For details on database searching, refer to Chapter 9, *Peptide, PTM, and Mutation Identification (PEAKS DB, PEAKS PTM, and SPIDER)*.

### Note

To run database search on an existing project, click the PEAKS DB icon  after selecting either a project node, a sample node, or a *de novo* result node in the Project View pane.

4. Click the “Quantification” button in the project wizard and select one of the quantification methods to perform that quantification analysis on the data after identification, or click "Finish" to perform all the steps up to the identification step. For details on the different quantification analyses, refer to Chapters 13, 14 and 15 Chapter 13, *PEAKS Q - MS Level*, Chapter 14, *PEAKS Q - MS/MS Level* for Chapter 15, *PEAKS Q - Label Free* respectively.

## Note

To run quantification on an existing project, click the Quantification icon  after selecting a PEAKS project, data file, or a search result node.

## 2. Data Visualization

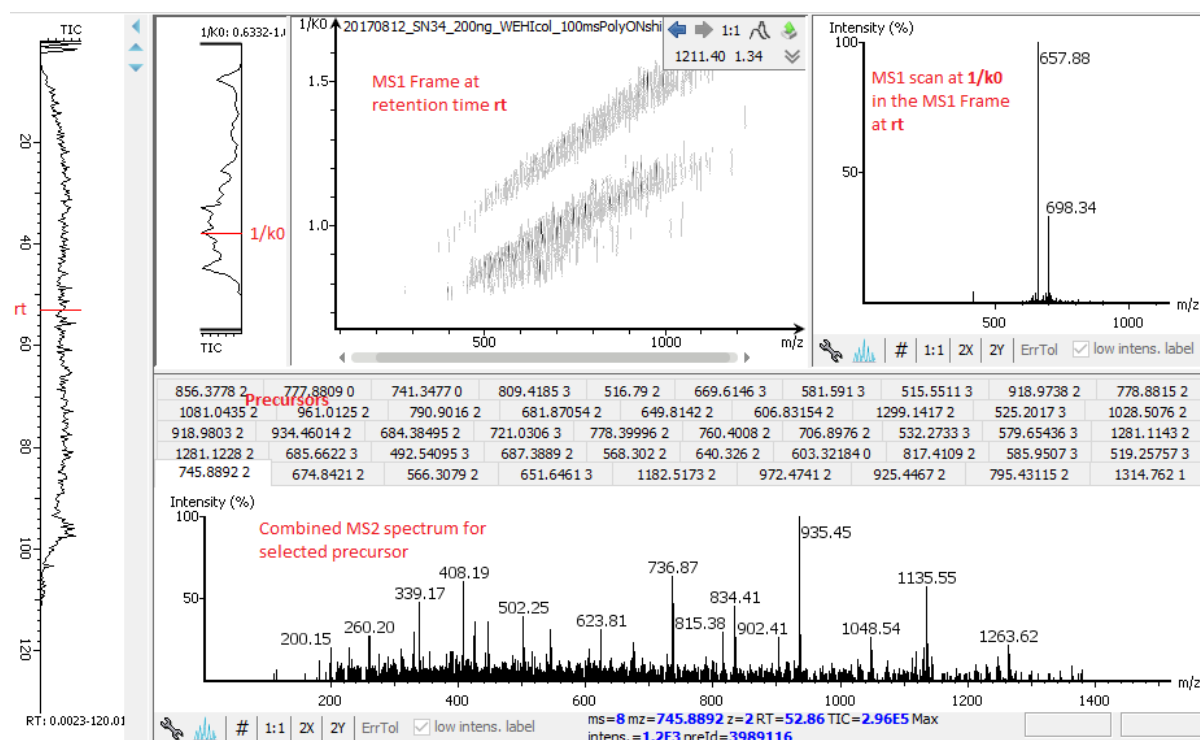
After the project is created, the ion mobility data can be visually examined. PEAKS IM module provides a view of the data at individual IM-MS level, individual IM-MS/MS level and overall view at LC-IM/MS level. Double click a data node to view the data.

- **IM-MS view:** This view presents the data focused at the MS1 level.
- **IM-MS/MS view:** This view presents the data focused at the MS2 level.
- **LC-IM/MS view:** This view provides a bird's eye view of the whole LC-IM/MS dataset with detected features and MS/MS scans marked on refined data.

### 2.1. IM-MS

The total ion chromatogram (TIC) is displayed on the left of the IM-MS view. A red bar on TIC curve points the current selected retention time. The top panel at right side shows the MS1 data at the selected retention time as an MS1 Frame - a heat map in  $m/z$ - $1/k_0$  dimension. Individual MS1 spectrum at the selected  $1/k_0$  is shown at the top-right panel.

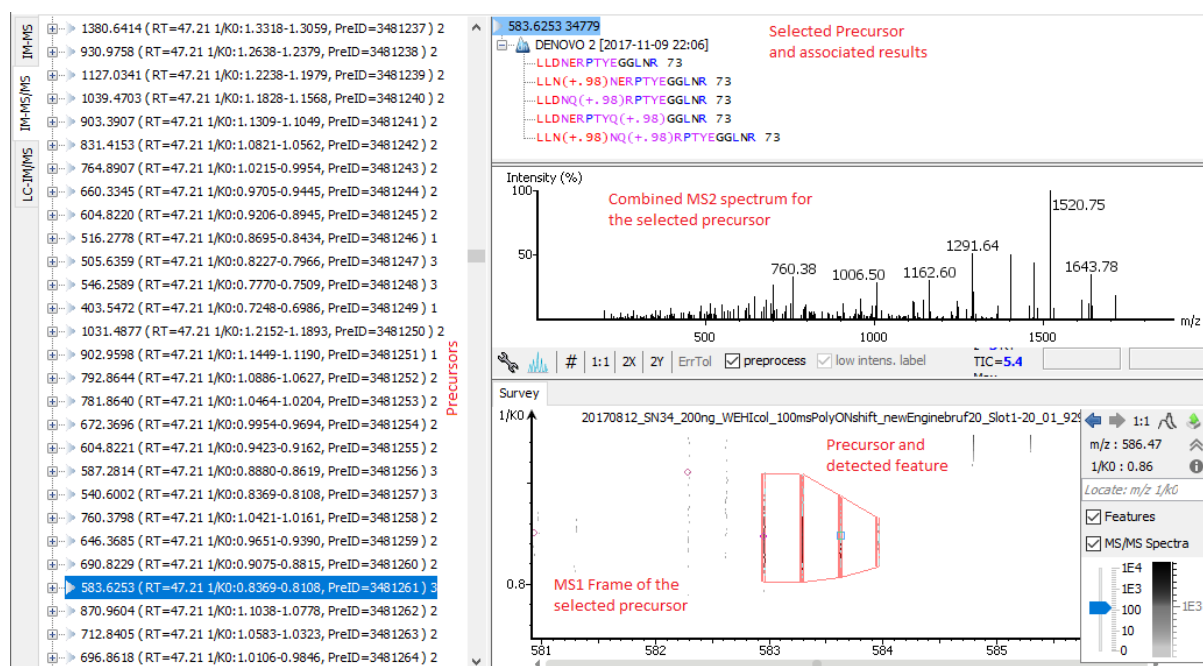
The precursors associated with the selected MS1 Frame are listed at the bottom panel - one tab per precursor. The MS2 data associated with a precursor are combined together to form a combined MS2 spectrum. Each precursor tab shows the corresponding combined MS2 spectrum.



### 2.2. IM-MS/MS


The IM-MS/MS View shows the list of precursors on the left panel. The right panel shows the details data associated with the selected precursor. The combined MS2 spectrum is shown in the right middle panel. The MS1

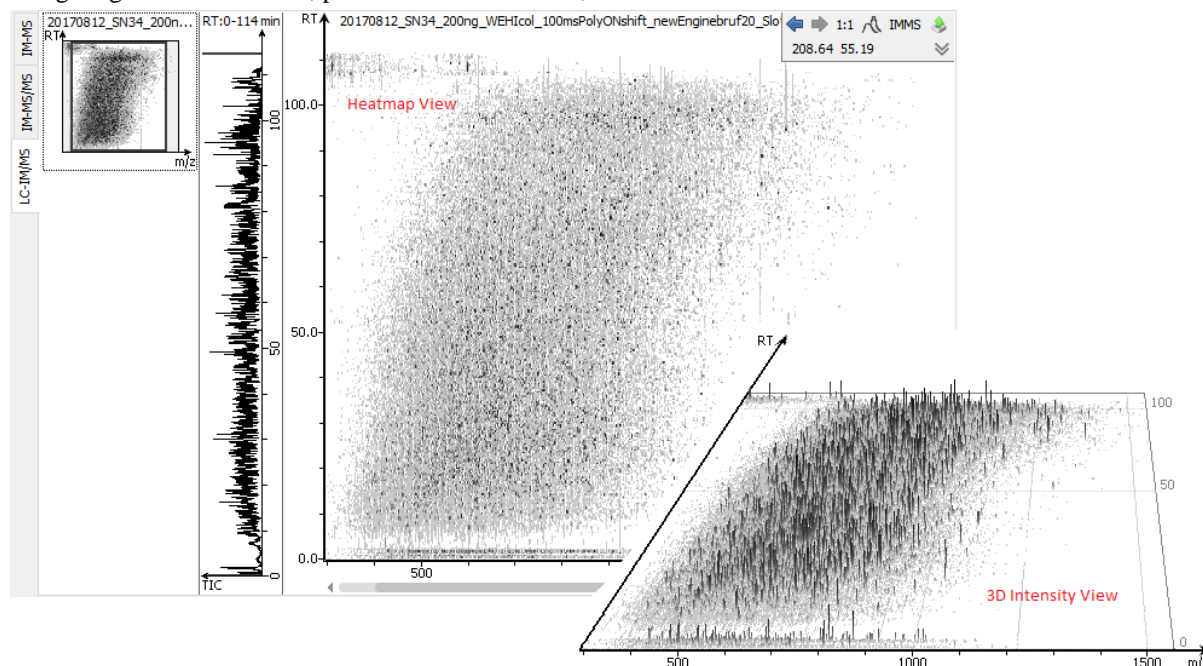
Frame, to whom the precursor belongs, is displayed at the bottom-right panel zoomed in to show the precursor and the associated feature.



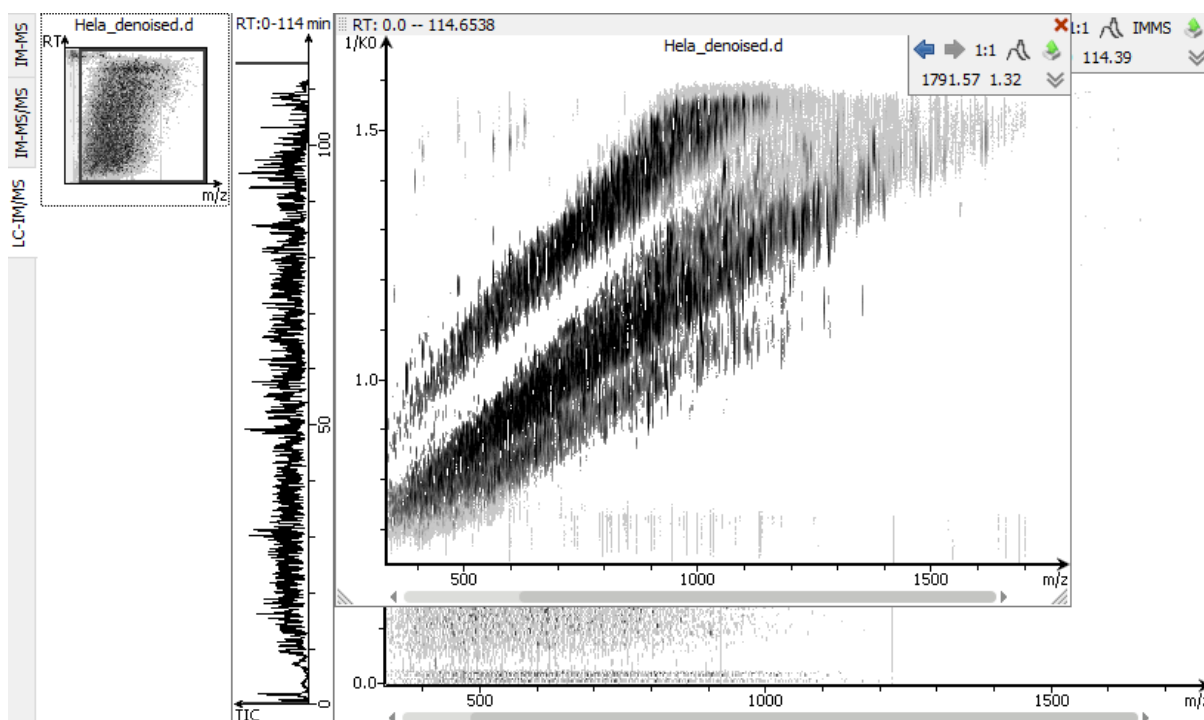
## 2.3. LC-IM/MS

The LC-IM/MS view shows the distribution of LC-IM/MS signals/features as a heat map projected on to  $m/z$ - $rt$  dimensions. PEAKS IM also provides a 3D intensity view for visual inspection of the LC-IM/MS data. To switch

between the heat map view and the intensity view, click the  button in the top-right control options. The view provides various convenient ways to zoom and navigate the LC-IM/MS features in the data. To learn more on navigating LC-IM/MS view, please refer to Section 4, "LC/MS View".



The distribution of LC-IM/MS data within the selected retention time range, can be viewed as a heat map projected on to  $m/z$ - $1/k0$  dimensions. Click "IM/MS" button to view the  $m/z$ - $1/k0$  heat map in a pop-up window.



## Note

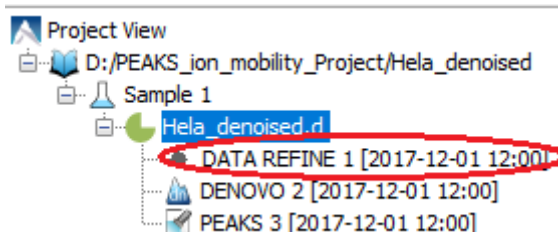
LC-IM/MS view requires some index files that will be generated in the background and the progress will be reported in the LC-IM/MS view.

## 3. Understanding PEAKS Results

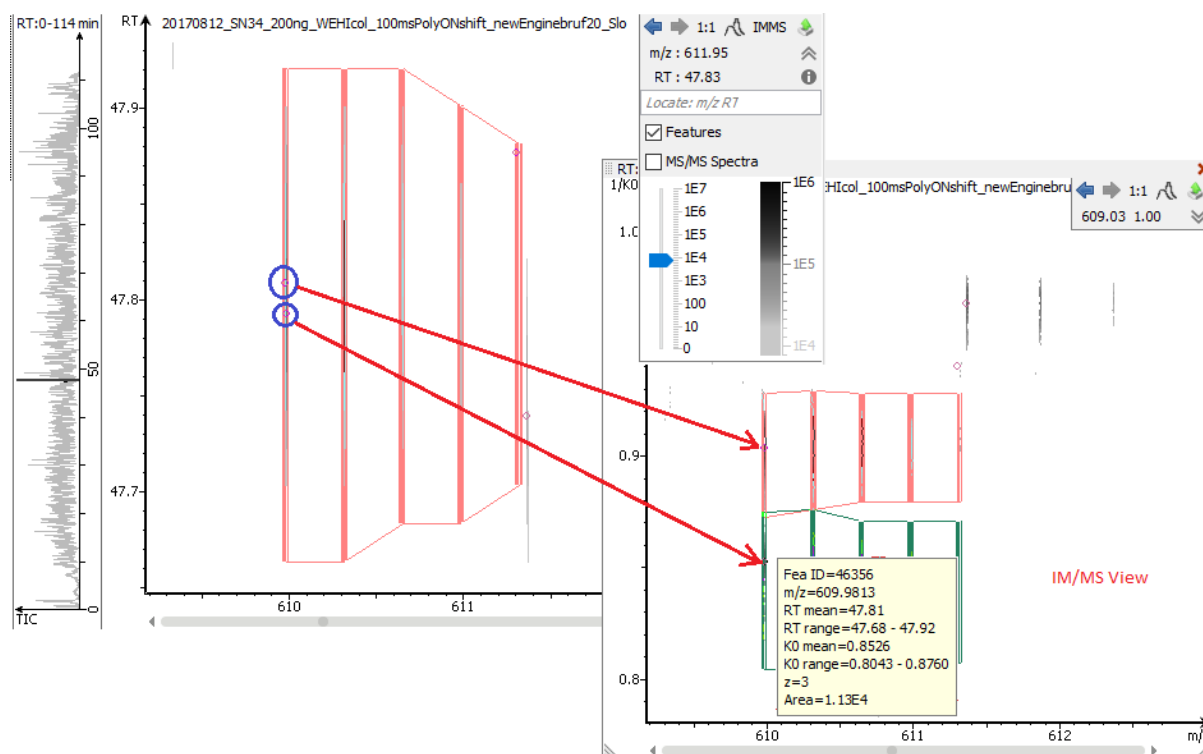
After analysis is complete, PEAKS IM module attach corresponding result nodes to the project tree. Double-click a result node to load the analysis report.

### 3.1. Data Refine

Data refinement can do precursor mass and charge correction as well as detect the MS1 4D features in the data. Default options should be good enough in the most cases. Once data refinement is complete, a “Data Refine” node will be generated in the project tree. Double click the data refine node to view the refined data. For details on data refinement, refer to Chapter 7, *Data Refinement*.



The data refinement analysis report uses the same views as the data visualization. The detected features can be pointed out in the views by selecting “Features” checkbox in the views. The feature centers are marked as red circles. Placing the cursor over a feature center shows the feature details in a popup.



## Note

Multiple features can be selected by holding the **Ctrl** key and left click on the feature center.

## 3.2. De Novo

*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 IM provides a reliable and comprehensive solution for automated peptide *de novo* sequencing from TimsTOF data. For details on *de novo* sequencing, please refer to Chapter 8, *Peptide De Novo Sequencing*.

The *de novo* analysis report is presented in three pages:

### 3.2.1. Summary View

The "Summary" view reports key statistics as an overview of the result. For details on *de novo* node summary view, refer to Section 3.1, "Summary View"

### 3.2.2. De Novo Peptide View

The *de novo* peptide table displays the *de novo* sequencing results in greater detail. For details on *de novo* result peptide view, refer to Section 3.2, "De Novo Peptide View". Here we mention the differences in the *de novo* peptide table from traditional LC-MS result.

1 - 1000 of 102253														
precursor Id = search no results														
Precursor Id	Peptide	Tag Length	ALC (%)	Length	m/z	z	RT	1/k0	Intensity	Mass	ppm	PTM	Mode	
3301049	DAPEEEDHVLRL	13	99	13	761.3893	2	45.29	1.0323-1.0583	1.78E4	1520.7471	11.1		CID	
4992302	LC(+57.02)DDELLLLK	10	99	10	616.3306	2	64.98	0.9151-0.9412	5.46E3	1230.6528	-5.1		CID	
4769667	LLGHWEAAHDLALAC(+57.02)K	17	99	17	645.3276	3	62.03	0.9238-0.9499	2.14E4	1932.9517	4.9		CID	
4937662	KVESLQEELAFK	13	99	13	767.4314	2	64.25	1.0572-1.0832	3.91E4	1532.8450	2.1		CID	
1424025	SYC(+57.02)AELAHNVSSK	13	99	13	733.3405	2	22.49	1.0161-1.0421	4.65E4	1464.6667	-0.3		CID	
7455032	WAVLGASGGLGQPLSLLK	19	99	19	897.0590	2	96.10	1.1395-1.1655	3.53E4	1792.0823	11.8		CID	
5159354	LLEEC(+57.02)LSQLVLHK	13	99	13	791.4473	2	67.20	1.0767-1.1027	3.58E3	1580.8596	13.0		CID	

- **Precursor Id:** This shows the Id associated with the precursor obtained from the raw data.
- **1/k0:** This shows the ion mobility range associated with precursor.

## Note

The *de novo* peptide table can be search for a specific precursor id.

### 3.2.3. LC-IM/MS

The view is same as in the data visualization with additional *de novo* tag marks. For details, please refer to Section 2.3, “LC-IM/MS”.

## 3.3. PEAKS DB

PEAKS DB node provides a complete set of database search tools for in-depth protein analysis. In this section, we present the differences in the DB result on ion mobility data from the traditional LC-MS data. To know more about PEAKS DB Searches and results, please refer to Chapter 9, *Peptide, PTM, and Mutation Identification (PEAKS DB, PEAKS PTM, and SPIDER)*.

The PEAKS DB analysis report is presented in five sections:

### 3.3.1. Summary View

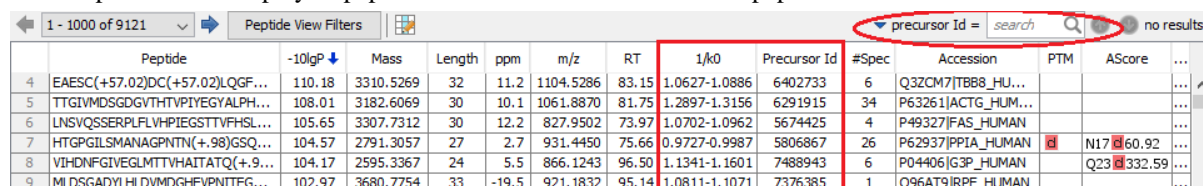
The "Summary View" reports key statistics as an overview of the result. It also provides functions to filter and export identified proteins and peptides. Please refer to Section 3.2, “Summary View” for details.

### 3.3.2. Protein View

The "Protein View" shows a list of protein identifications along with protein sequence coverage that helps in protein characterization. Please refer to Section 3.4, “Protein View” for details.

### 3.3.3. Peptides View

The "Peptide View" displays a peptides table which contains a list of peptide identifications.



	Peptide	-10lgP ↓	Mass	Length	ppm	m/z	RT	1/k0	Precursor Id	#Spec	Accession	PTM	AScore	...
4	EAESC(+57.02)DC(+57.02)LQGF...	110.18	3310.5269	32	11.2	1104.5286	83.15	1.0627-1.0886	6402733	6	Q3ZCM7 TBB8_HU...			...
5	TTGIVMDSGDGVTHTPVIYEGYALPH...	108.01	3182.6069	30	10.1	1061.8870	81.75	1.2897-1.3156	6291915	34	P63261 ACTG_HUM...			...
6	LNSVQSSERPLFLVHPVIEGTTVFHSL...	105.65	3307.7312	30	12.2	827.9502	73.97	1.0702-1.0962	5674425	4	P49327 FAS_HUMAN			...
7	HTGPGILSMANAGPNTN(+.98)GSQ...	104.57	2791.3057	27	2.7	931.4450	75.66	0.9727-0.9987	5806867	26	P62937 PPIA_HUMAN			...
8	VIHDFNGIVEGLMTTVHAITATQ(+.9...	104.17	2595.3367	24	5.5	866.1243	96.50	1.1341-1.1601	7488943	6	P04406 G3P_HUMAN			...
9	MLDSGADYLHLDVMDGHFVFNITFG...	102.97	3680.7754	33	-19.5	921.1832	95.14	1.0811-1.1071	7376385	1	Q96AT9 RPE_HUMAN			...

For details on Peptide View, please refer to Section 3.5, “Peptide View”. The following are the difference in the peptide table from the traditional LC-MS result:

- **1/k0:** This shows the ion mobility range associated with the best PSM of that peptide.
- **Precursor Id:** This shows the Id of the precursor associated with the best PSM of that peptide.

## Note

The peptide table can be search for a specific precursor id.

### 3.3.4. De Novo Only View

The "De Novo Only View" shows a list of quality peptide sequences detected by *de novo* sequencing that remain unidentified by the PEAKS database search. Please refer to Section 3.6, “De Novo Only View” for details.


### 3.3.5. LC-IM/MS View

The view is same as in the data visualization with additional “identified peptides” marks. For details, please refer to Section 2.3, “LC-IM/MS”.



## 4. Label Free Quantification (LFQ)

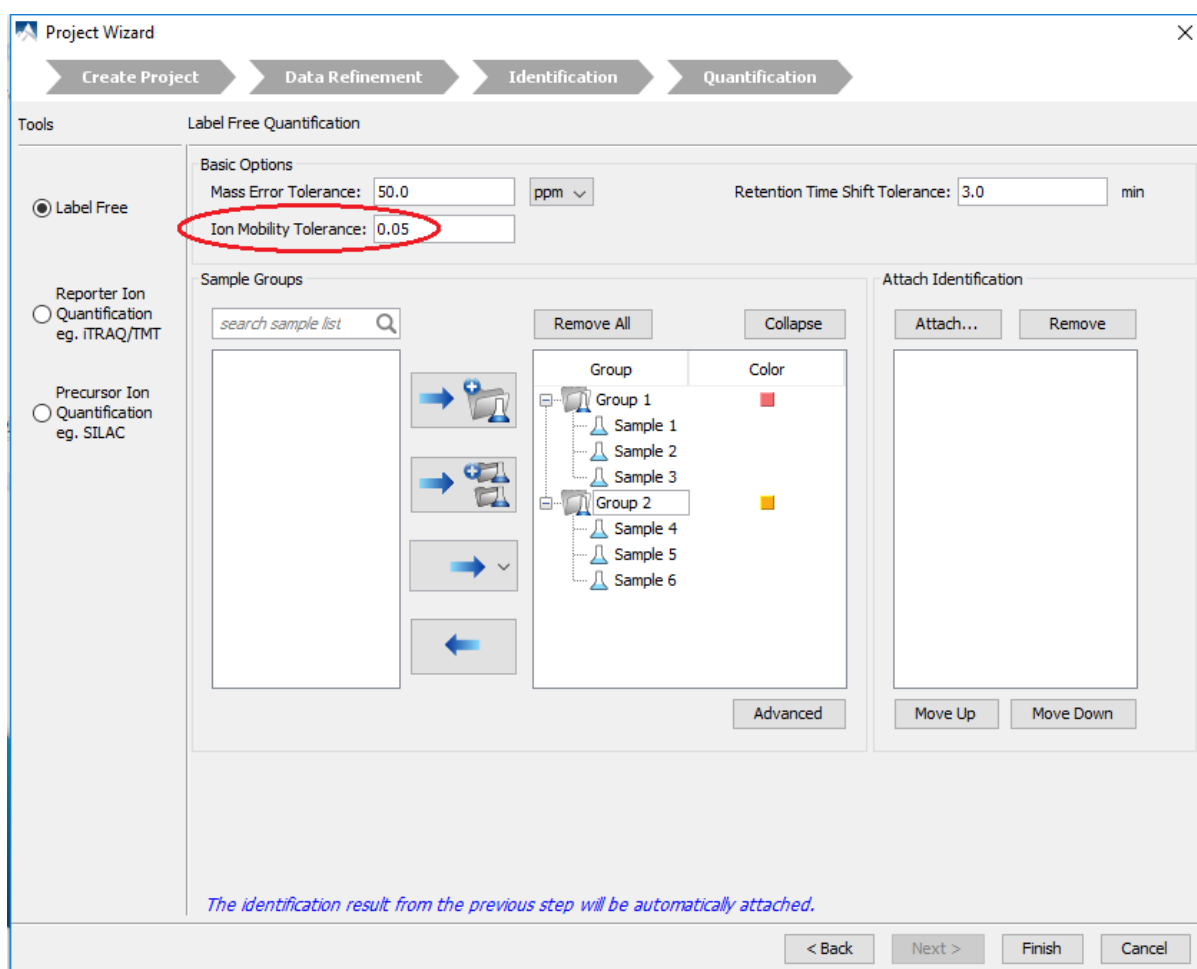
PEAKS IM module supports the Label Free Quantification (LFQ) analysis of the ion mobility data. In this section, we focus on the differences in LFQ analysis on ion mobility data from the LFQ on traditional LC-MS data. For details on LFQ analysis, please refer to Chapter 15, *PEAKS Q - Label Free*.

Select a project node in the “Project View” frame. Click the PEAKS Quantification tool bar icon  or select the Tools | Quantification command from the menu. Select “Label Free” check box to perform LFQ analysis.

### 4.1. Setting Parameters

Please refer to Section 2, “Setting Parameters” for details on the LFQ parameters settings. The following additional LFQ parameter is available for ion mobility data:

- **Ion Mobility Tolerance:** The allowable maximum shift in mean  $1/k_0$  of an identified peptide between different runs.



### 4.2. Result Filtration

PEAKS provides an extensive set of tools to filter Label Free Quantification results according to the user's requirements. Please refer to Section 3.1.1, “Result Filtration” for details on various filter options.

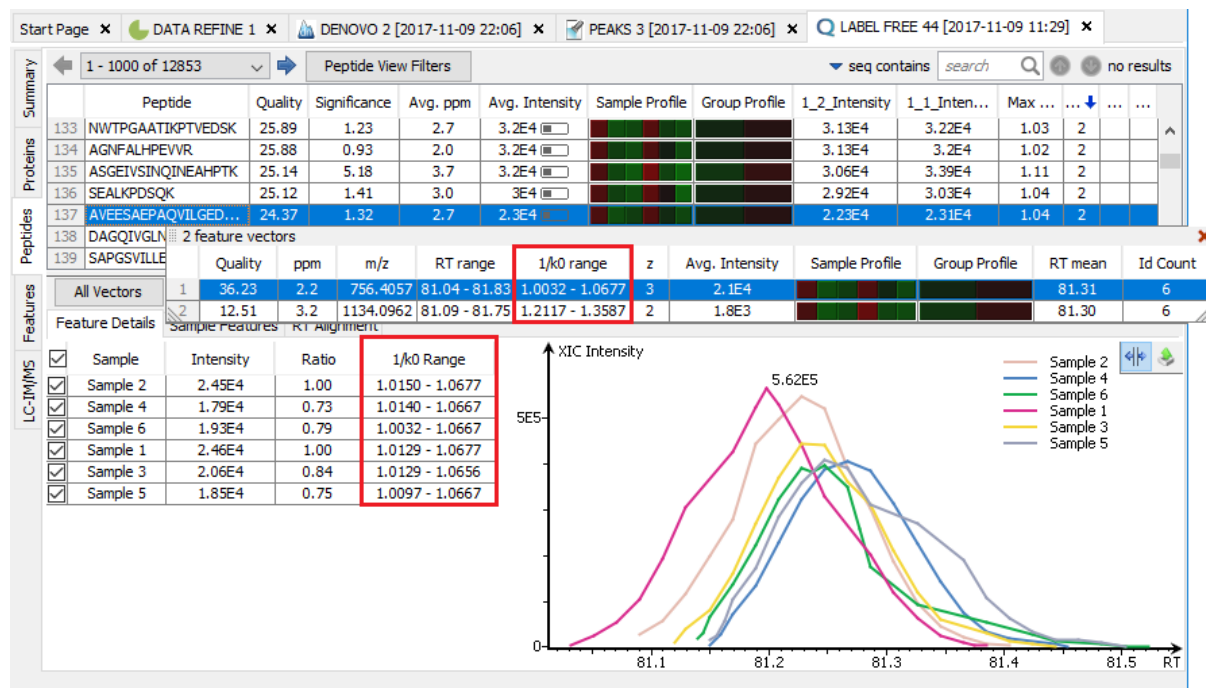
### 4.3. Peptide Features

The quantifiable feature vectors with the same sequence are grouped together to get the list of quantifiable peptides and are displayed in “Peptide View”. Please refer to Section 3.3, “Peptide View” for details on how to interpret LFQ peptide results. In this section, we highlight some differences for the ion mobility data.

### 4.3.1. Feature Details

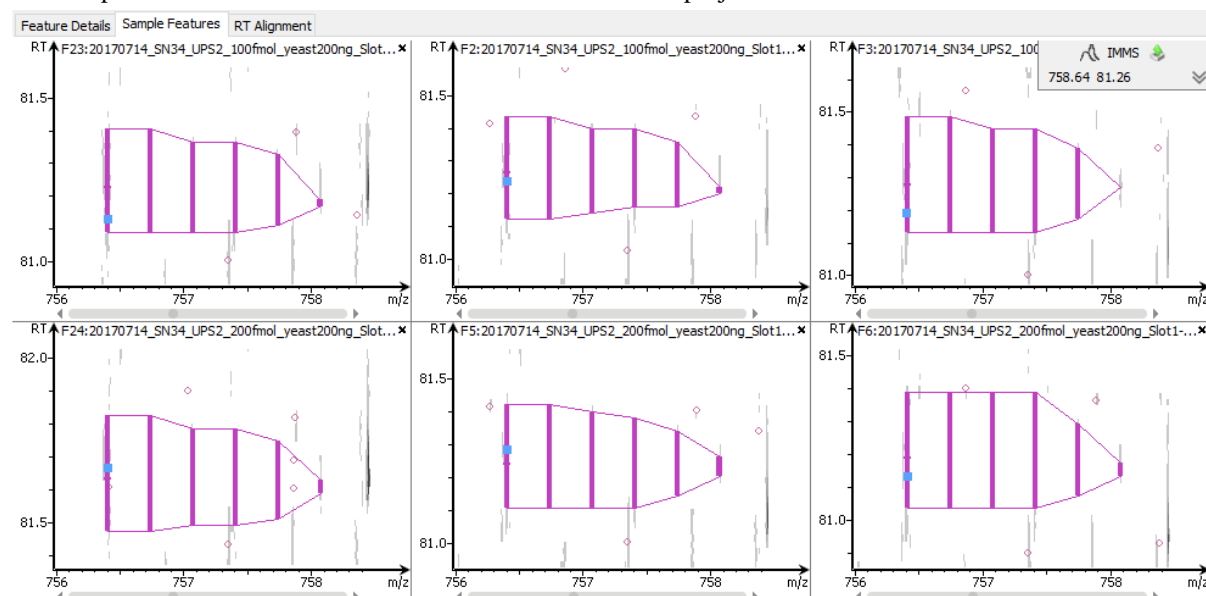
The feature details contain the following additional column:

- **1/k0 Range:** In “All Vectors” table this shows the ion mobility range associated with the feature vector. In “Feature Details” table, this shows the individual ion mobility range of the 4D features in the feature vector.



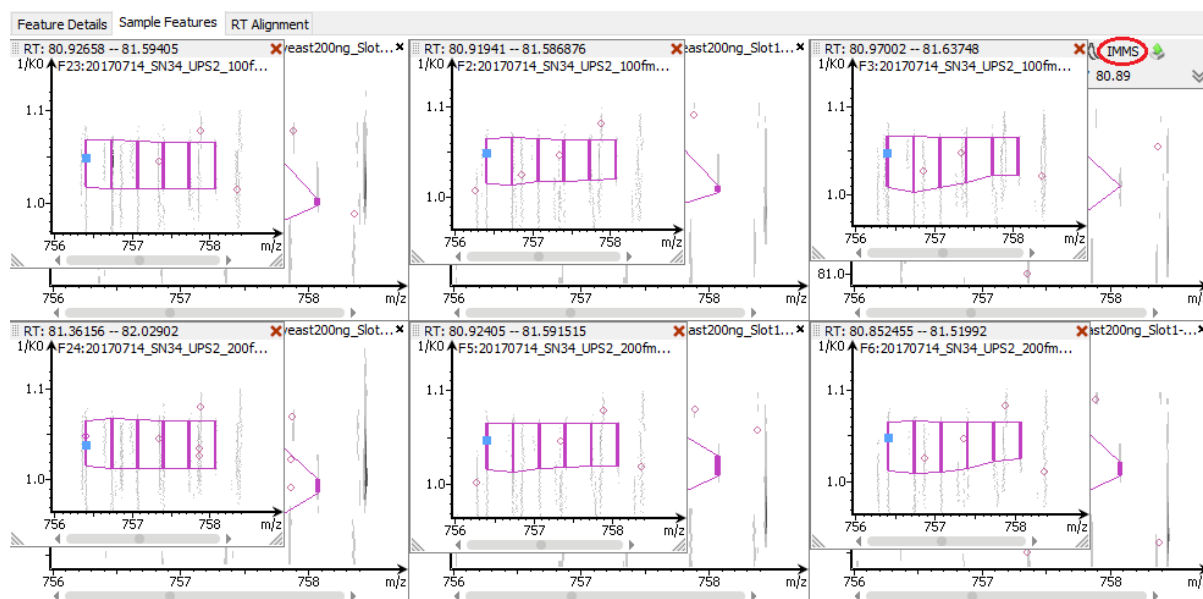
### 4.3.2. Sample Features

The sample features shows the 4D features from different runs projected on **m/z-rt** dimensions.




Click the “IMMS” button in the feature control panel, to view the 4D features projected on **m/z-1/k0** dimensions.






## 5. Precursor Ion Quantification (eg. SILAC)

PEAKS IM can perform precursor ion quantification analysis for ion mobility data with the optional PEAKS Q module. Both user-defined and commercial labels (eg. SILAC and iCAT) are supported by PEAKS Q. For details on precursor ion quantification setup and data analysis, please refer to Chapter 13, *PEAKS Q - MS Level*, as the setup and functions are similar to any other type of MS data.

To perform a precursor ion quantification analysis, select a PEAKS project in the “Project View” pane. Then, click the PEAKS Q icon, , from the tool bar or select the Tools > Quantification command from the menu. Select “Precursor Ion Quantification eg. SILAC” check box to perform precursor ion quantification for the PEAKS project.

## 6. Reporter Ion Quantification (e.g. iTRAQ and TMT)

Reporter ion quantification with chemical labels at the MS2 (MS3) level is supported by the optional PEAKS Q module. Both user-defined and commercial labels (i.e. iTRAQ and TMT) are supported by PEAKS Q. For details on reporter ion quantification setup and data analysis, please refer to Chapter 14, *PEAKS Q - MS/MS Level*, as the setup and functions are similar to any other type of MS data.

Select a PEAKS DB/PTM/SPIDER result node in the “Project View” frame. Then, click the PEAKS Q icon, , from the tool bar or select the Tools > Quantification command from the menu. Select “Reporter Ion Quantification eg. iTRAQ/TMT” check box to perform reporter ion quantification for the PEAKS search results.

### 6.1. Result Display

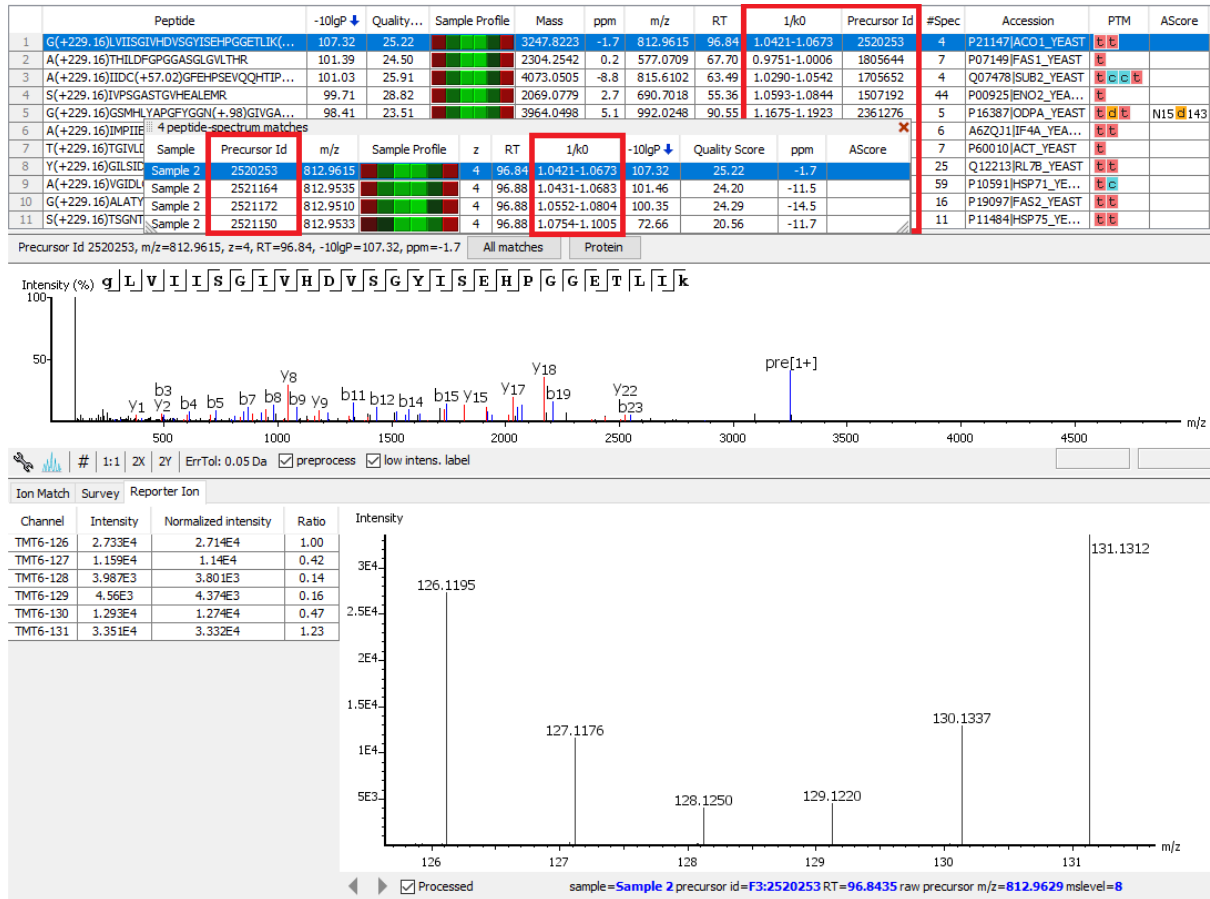
Please refer to Section 3, “Understanding the Results” for details on how to interpret Reporter Ion Quantification results. In this section, we highlight some differences for the ion mobility data.

#### 6.1.1. Peptide View

The peptide view contains the following additional columns:

- **1/k0 Range:** In “All matches” table this shows the ion mobility range associated with the corresponding PSM. In “Peptide” table, this shows the ion mobility range of best PSM according to the -10lgP score.
- **Precursor Id:** This shows the precursor id associated with the PSM.


# Ion Mobility Data Analysis

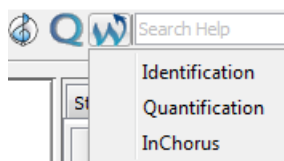


---

# Chapter 18. Creating a High-Throughput Workflow

PEAKS Workflow provides a simple setup to allow users to batch analyze multiple samples and/or fractions on an individual level. The Workflow function can be used to setup batch identification, quantification or inChorus (multi-engine protein ID) searches. This tool is especially helpful for analyses on multiple projects using the same parameters.


To setup the PEAKS Workflow, select the Workflow icon  from the toolbar. Users will be given 3 workflows to choose from in a pop-up menu: Identification (PEAKS DB, PEAKS PTM, SPIDER), Quantification (PEAKS Q) or inChorus. Once a specific workflow is selected, a dialogue appears to specify the necessary parameters for each step in the workflow.

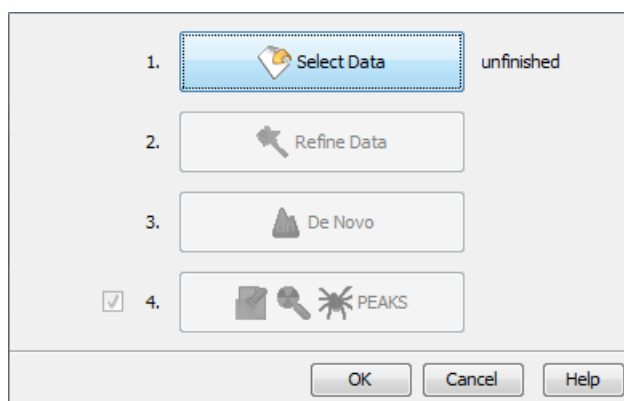


## Note

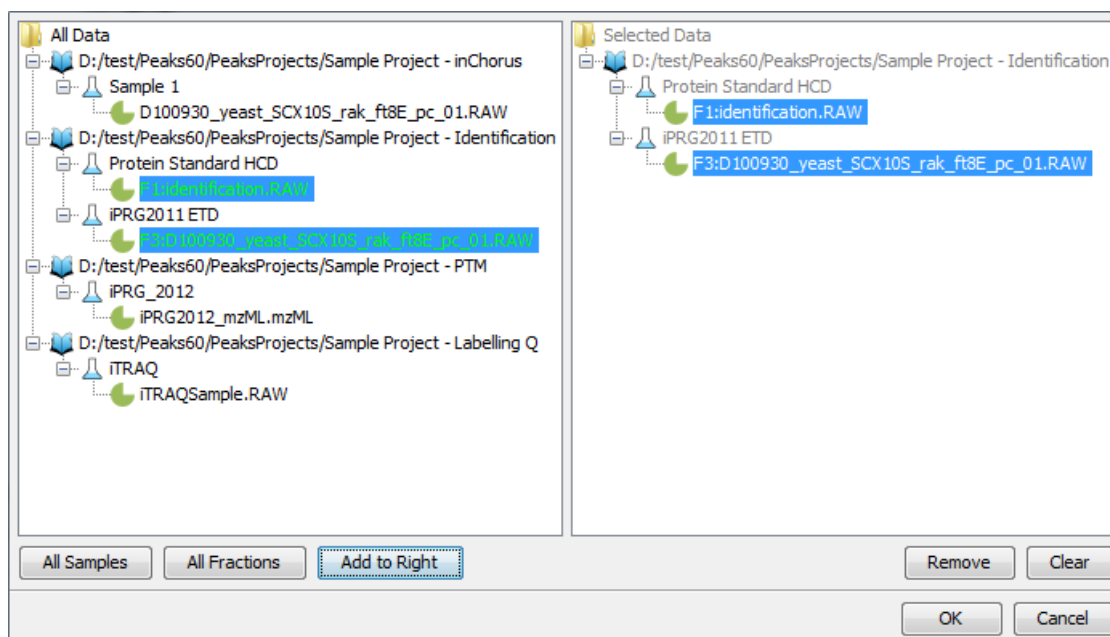
To use the PEAKS Workflow function, a PEAKS project must already be setup and opened with the data already loaded. See Chapter 4, Chapter 4, *Project Wizard - Create a PEAKS Project, Load Data and Set up a Workflow* for details on how to create a PEAKS project and load data.

## 1. Identification Workflow

To perform a batched identification workflow with PEAKS, select the Workflow icon  from 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. Select the desired samples or fractions for analysis from the left-hand side of the window and move them to the right-hand side using the "Add to Right" button. Multiple samples or fractions can be selected at once by holding the Ctrl button from the keyboard. To simply select all the samples displayed in the left-hand side of the window, select the "All Samples" button and then click "Add to Right". If instead all fractions need to be analyzed, users can select the "All Fractions" button and then click "Add to Right". Use the Remove and Clear buttons to remove the selected files/samples or all files/samples from the “Selected Data” list, respectively. Click OK to proceed to the next step.



## Note

All files that are loaded into a single workflow will be processed in exactly the same way, using the exact same parameters. To run some files differently than others, you must set up separate workflows.

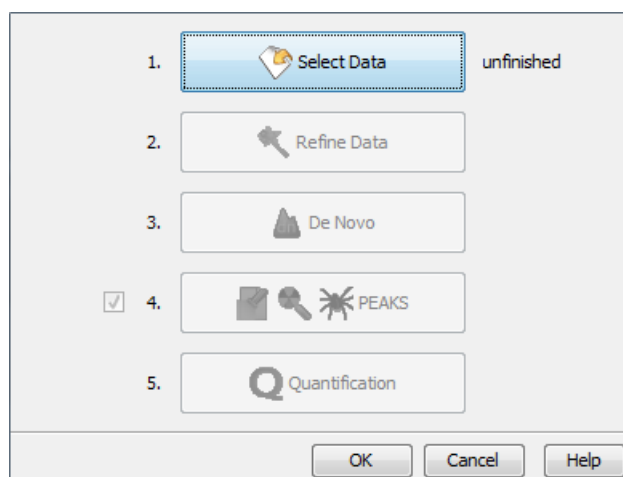
Once the data is selected, users can specify parameters for the following stages of the identification workflow one-by-one by following the order presented in the workflow dialogue (ex. Data refinement, de novo sequencing, PEAKS identification searches). Please refer to the chapters on each individual function for details on setting up the function-specific parameters.

## Note

PEAKS DB, PEAKS PTM, and SPIDER can be setup in the same step (see Chapter 9, *Peptide, PTM, and Mutation Identification (PEAKS DB, PEAKS PTM, and SPIDER)*), but are optional. Users can uncheck these choices if they do not wish to perform these searches.

## 2. Quantification Workflow

The quantification workflow is similar to the identification workflow, but includes 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.

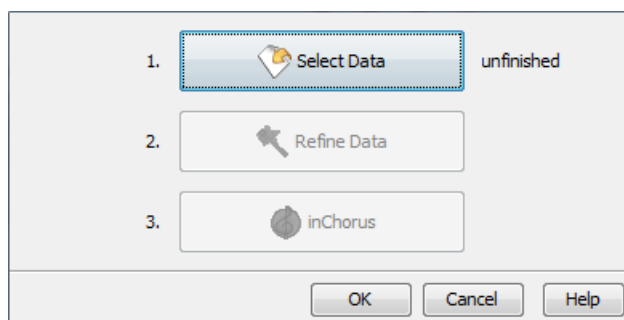


## Note

Only Reporter Ion Quantification (eg. iTRAQ/TMT) or Precursor Ion Quantification (eg. SILAC) workflows can be set in the quantification workflow. For the Label Free workflow, please use the Project Wizard or PEAKS Q button.

## 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 for all selected data. Refer to Chapter 12, *Combining Multiple Database Search Engines with PEAKS inChorus* for details on inChorus parameters.



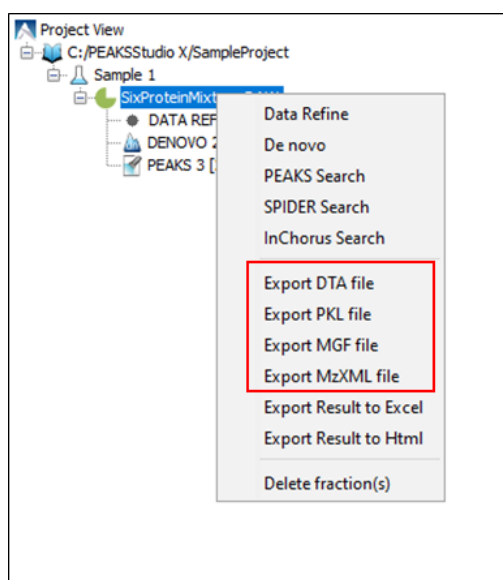
---

# Chapter 19. Exporting Data/Reports and Printing

PEAKS offers a rich collection of exporting functions to allow users to create reports and share the analysis results with collaborators, colleagues, and clients. The supported formats include HTML, Comma Separated Values (CSV), pepXML, mzIdentML, and various image formats for image exporting. Labs with in-house software can easily make use of the CSV files in their own analysis workflows. The exported results in HTML can be viewed with a web browser. The entire exported result directory can be zipped and emailed to colleagues, or the whole directory can be put directly on a website.

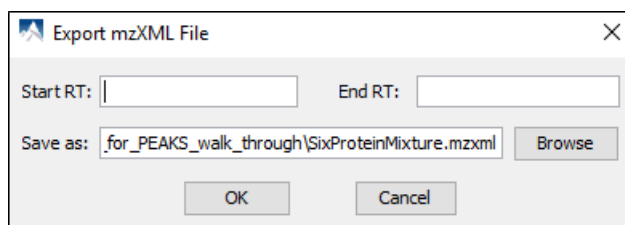
## 1. Exporting Data

Spectrum data can be exported to a number of file formats including mzXML, DTA, MGF, and PKL. To do so, right-click on the sample node or the data file that is to be exported and select the desired exporting format.



Selecting the Export DTA file command will open a dialogue prompting the folder name and location to which the DTA files will be exported. For MGF and PKL exports, select the Export MGF file and Export PKL file respectively, and the dialogue will ask for a name and a location for the file.

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

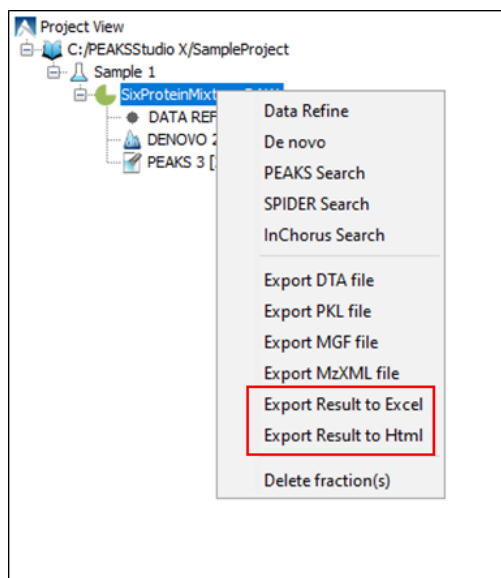


Enter the starting and ending retention times in the appropriate boxes. Then, click the Browse button to select a destination to save your file.

## 2. Exporting Result From Project View

Fraction information and peptide identification results can be exported to Excel or HTML format from the pop-up menu in the “Project View”. Right 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.



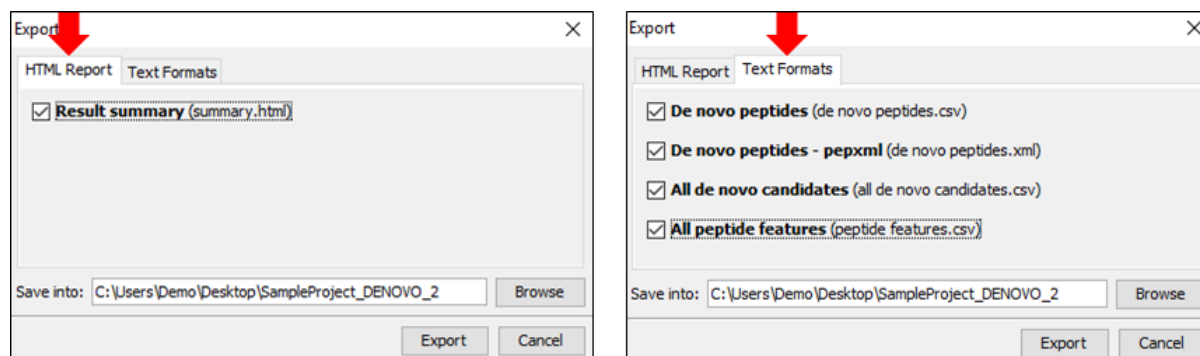
More function-specific export options are available from the “Summary” view of the respective results. The following sections describe in detail the function-specific exporting options available in various formats.

## 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, Peptides, and Features

To export results, press the Export button in the tool bar of the “Summary” view panel. The following export dialogue will appear. Select the items that you want to export from the available options. Select the output location and click the Export button to save the selected result components to the specified location.



The export options are grouped into “HTML Report” and “Text Formats” based on the output format.

**HTML Report.** This will generate a summary report in the specified location. After the completion of exporting, the result will be opened in the default browser automatically. The following exporting options are available:

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

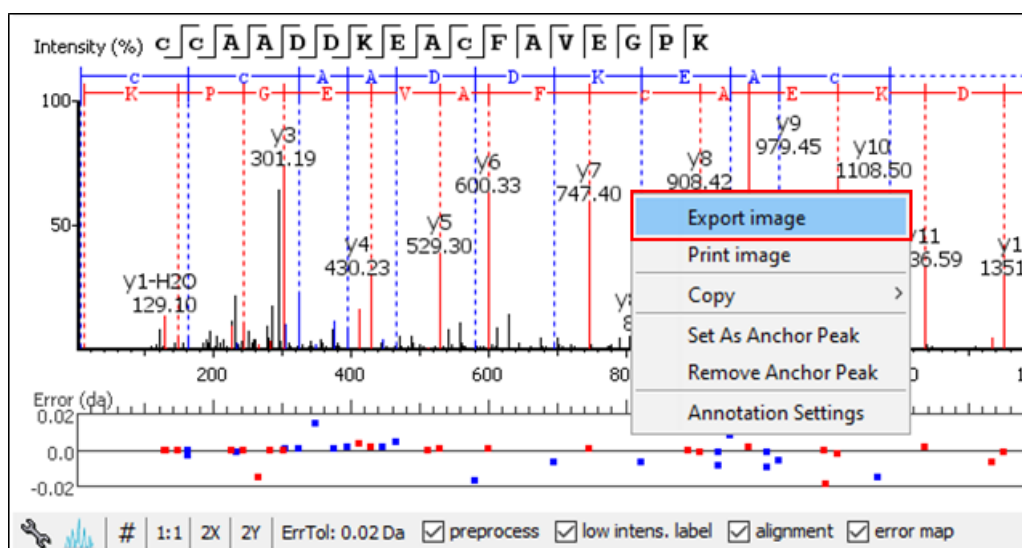
**Text Formats.** The following exporting options are available in various text formats:

- **De novo peptides :** The peptides identified by *de novo* sequencing will be saved to `de novo peptides.csv` file in Comma Separated Values (CSV) format in the specified folder.

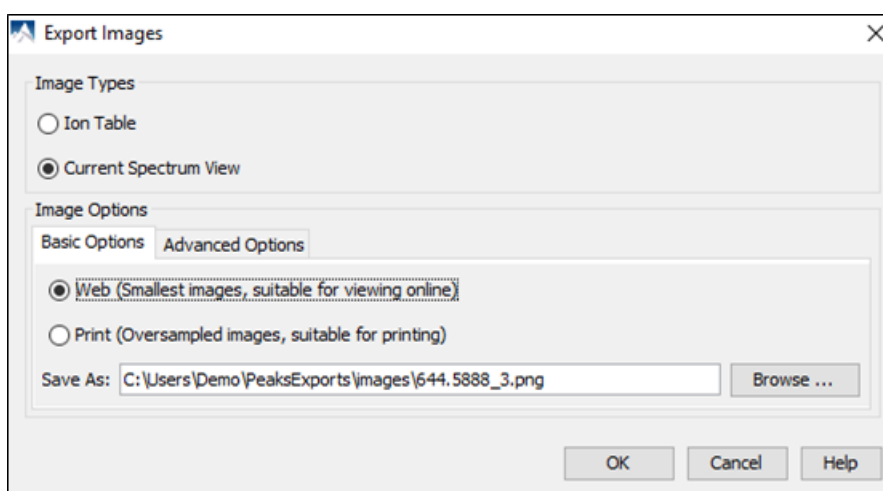
- **De novo peptides - pepxml** : In addition to CSV format, the peptides can be saved in pepXML format to `de novo peptides.xml`
- **All de novo candidates** : All *de novo* candidates for each spectrum will be saved to `all de novo candidates.csv` file in CSV format.
- **All peptide features**: All predicted peptide features found in the LC/MS information will be sorted by m/z and saved to `peptide features.csv`. If a *de novo* result is associated with the peptide feature above the ALC filter it will be displayed as well.

## 3.2. Exporting Images

The Ion Match Table and current spectrum view can each be exported to image files. The spectrum image will export as displayed in the interface. For example, if the error map and spectrum alignment is enabled in the spectrum view, then this information will be included in the exported image. To export, position the cursor on either the spectrum or ion match table, right-click to view the pop-up menu, and select the Export Image command from the menu.



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



Select the desired result elements from the “Image Types” list. The “Basic Options” tab offers choices for location and image size options to which can be saved at either “Web (Smallest images, suitable for viewing online)” or “Print (Oversampled images, suitable for printing)”.



The "Advanced Options" panel contains the following settings:

- Scale: 100 %
- Oversample: 1 X
- File Format: png
- Width: 1,398
- Height: 344

The “Advanced Options” panel offers choices for editing the scaling factor, the file format, the resolution, and the oversample factor. PEAKS supports BMP, GIF, JPEG, PNG, and SVG image formats. After setting all parameters, click the OK button to export the selected result item to an image.

## 4. Exporting Database Search Results

The exporting mechanism for PEAKS DB, PEAKS PTM, and SPIDER results are the same as that of a *de novo* result with the exception of the number and the type of available exporting options. All exporting functions are available through the “Summary” view panel.

### 4.1. Exporting Summary, Proteins, and Peptides

To export the result, press the Export button in the title bar of the “Summary” view panel. The following export dialogue will appear.

The 'Export' dialog box has three tabs: 'HTML Report', 'Text Formats', and 'For Third Party'. Each tab contains a list of exportable items and a 'Save into' field.

- HTML Report:**
  - Report proteins in HTML with:
    - ☒ summary view
    - ☒ protein coverage
    - ☒ supporting peptides
    - ☐ best unique PSM
  - ☒ Collect all the selected details above in a single html
  - ☒ Export peptide list in html
- Text Formats:**
  - ☒ 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)
  - ☐ Filters (filter.csv)
  - ☐ Proteins - fasta (proteins.fasta)
  - ☐ Peptide-spectrum match ions (PSM ions.csv)
  - ☐ Peptides - pepxml (peptides.pep.xml)
  - ☐ 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)
- For Third Party:**
  - For Skyline:
    - ☐ peptides - pepXML (peptides.pep.xml)
  - For Scaffold:
    - ☐ peptides - mzIdentML (peptides\_1\_1\_0.mzid)

Each tab has a 'Save into' field with the path 'C:\Users\Demo\Desktop\SampleProject\_PEAKS\_3' and a 'Browse' button. The 'Export' button is at the bottom of each tab.

The export options are grouped into “HTML Report”, “Text Formats”, and “For Third Party” based on the output format. Select the items that you want to export from the available options. Select the output location and click the Export button to save the selected result components to the specified location.

**HTML Report.** This will generate single or multiple HTML reports in the specified location. After the completion of result exporting, the index file for the reports, `protein.html`, will be opened in the default browser automatically. The following exporting options are available:

- **Summary view:** The “Summary” view page will be saved as a `summary.html` file in HTML format in the specified location.
- **Protein coverage:** The coverage pane will be saved for each protein.
- **Supporting peptides:** A list of supporting peptides will be saved for each protein.
- **Best unique PSM:** The best unique PSM will be saved for each protein.
- **Export peptide list in html:** creates a separate HTML-containing peptide identifications.

An individual protein will have its own HTML output file, where the corresponding protein coverage, the supporting peptides, and the best unique PSM are gathered. Select “Collect all the above details in a single html” to collect all the protein reports in a single `protein.html` file.

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

- **Proteins:** The list of protein identifications, filtered by the protein filters (in the "Summary" view) and the “Show” filter (in the "Protein" view above the Protein table), will be saved to `proteins.csv` in Comma Separated Values (CSV) format.
- **Supporting peptides:** A list of supporting peptides of each protein identification, filtered by the peptide filters and protein filters (in the "Summary" view) and the “Show” filter (in the "Protein" view above the Protein table), will be exported to `protein-peptides.csv`. This usually contains more entries than the Peptide table as a peptide can be identified in several proteins and, therefore, can be reported multiple times in this file.
- **DB Search Peptides:** All of the identified peptides and details will be saved to `peptide.csv`.
- **DB search peptide-spectrum matches:** The selected peptide-spectrum matches (PSMs), their display controlled by the peptide filters in the "Summary" view, will be exported to `DB_search_psm.csv`. Peptides differentiated with only I/L isoform are represented by separate entries. As a result, the number of entries in this file might be bigger than the number of PSMs mentioned in the "Summary" view.
- **De novo only peptides:** A list of good *de novo* sequences that do not have good, or any, database matches will be saved to `de_novo_only_peptides.csv`.
- **Filters:** The summary filter, protein view filters, as well as peptide view filters will be saved to `filter.csv`.
- **Proteins - fasta:** A list of protein identifications will be saved to `proteins.fasta`.
- **Peptide-spectrum match ions:** All theoretical and actual fragment ions associated with each peptide-spectrum match, including fragment mass error will be saved to `PSM_ions.csv`
- **Peptides - pepxml:** A list of peptide-spectrum matches, their display controlled by the peptide filters in the "Summary" view, will be saved to `peptides.pep.xml` in pepXML format. The number of the “spectrum\_query” elements will be the same as the number of PSMs mentioned in the "Summary" view, and the number of the “search\_hit” elements will be the same as the number of PSMs counted from Peptide table ( summing up #Spec column).
- **Peptides - mzidentml (version 1.1.0):** A list of peptide-spectrum matches will be saved to `peptides_1_1_0.mzid` in mzIdentML format version 1.1.0.
- **All peptide features:** For each data file in the search, a list of predicted peptide features is exported and sorted by mass and saved to `peptide_features.csv`. If a peptide or *de novo* only result is associated with it the peptide sequence will be given as well.
- **De novo only peptides - pepxml:** A list of good *de novo* sequences with either poor or non-existent database matches will be saved in `de_novo_only_peptides.xml` in pepXML format.
- **De novo only spectra - mgf:** All spectra missed by an identification search but with *de novo* results passing the ALC filter set on the "Summary" view and will be saved to `de_novo_only_spectra.mgf`

**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 the supported mzIdentML format, 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, such as PTM analysis.

## Note

Peptides identified by SPIDER may contain residue substitutions, insertions, and deletions. The pepXML format does not specify the encoding for substitution, insertion, and deletion, and the mzIdentML format only specifies the encoding for substitution. Therefore, when exporting from a SPIDER result, those peptides may not be correctly interpreted by third party software.

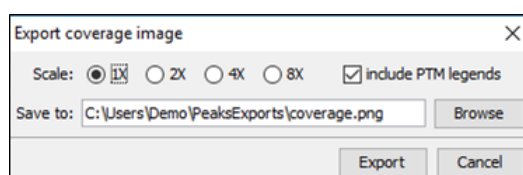
- **For Skyline: peptides - pepXML:** In addition to exporting the peptide-spectrum matches to pepXML format, PEAKS will export the spectrum data in mzXML format. The exported pepXML file and mzXML files can be loaded into Skyline for spectrum library generation.

## 4.2. Exporting Images


From the “Peptide” view, “Feature” view and the “*De novo* only” view, the Annotated Spectrum and/or Ion Match table can be exported to an image file. The Annotated Spectrum image will export as is displayed on the interface. For example, if the error map and spectrum alignment is enabled in the spectrum view, then this information will be included in the exported image. To do so, position the cursor on any of those items in the result panel, right-click, and then select the Export Image command from the menu. Refer to Section 3.2, “Exporting Images” for details.

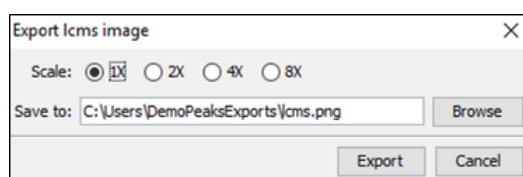
### 4.2.1. Exporting Protein Coverage

To export the protein coverage pane, click the tools  button in the coverage pane and select the Save image as... command from the menu. Select the resolution of the image and the location to save. Click the Export button. The PTM legends can be included in the exported image.



### 4.2.2. Exporting LC-MS, IM-MS, and LC-IM-MS

To export the LC-MS, IM-MS and/or the LC-IM-MS heatmaps, click the export image button  from the toolbar in the upper right-hand corner of the result view. Select the resolution of the image and location to save. Click the Export button.

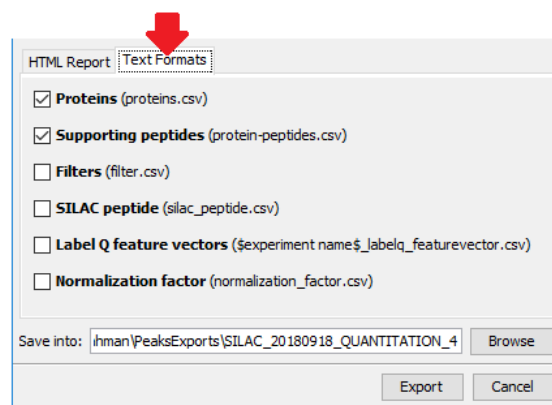
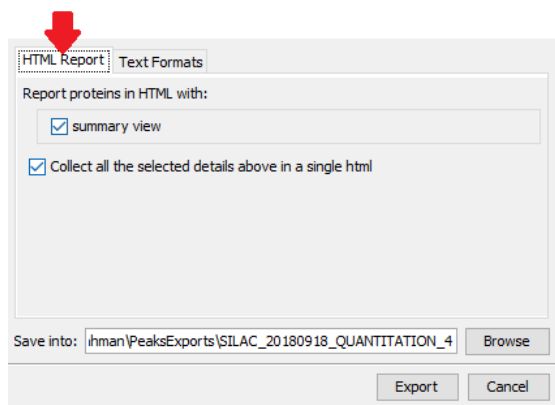


## 5. Exporting Quantification Results

PEAKS Q (labeled and label free quantification) exporting functions are also similar to that of *de novo* or PEAKS DB. All export functions are available through the “Summary” view panel.

### 5.1. Exporting Precursor Ion Labelling Quantification Results

PEAKS Q results can be exported to other supported formats. To export the quantification results, press the Export button in the title bar of the “Summary” view panel. The following export dialogue will appear.



**HTML Report:** All information found in the "Summary" view of the quantification result will be exported to html format.

#### Text Formats:

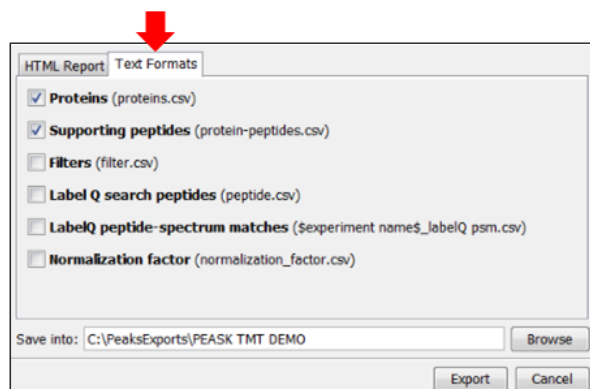
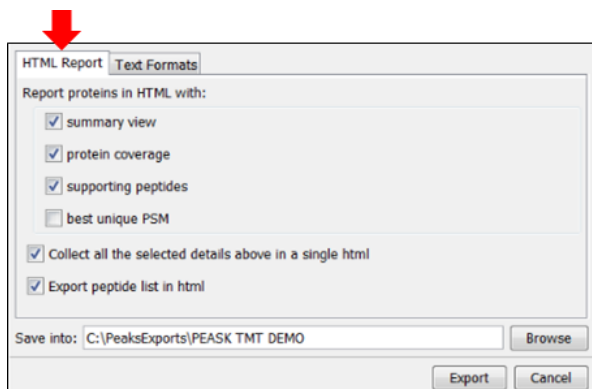
- **Proteins:** These are quantified proteins that pass the filters set in the "Summary" view and associated details
- **Supporting peptides:** These are supporting peptides of quantified proteins and associated details
- **Filters:** The summary filter, protein view filters, as well as peptide view filters will be saved to filter.csv.
- **SILAC peptide:** All details of all identified and quantified peptides passing the peptide filters are listed here. Duplicate peptides are grouped.
- **Label Q feature vectors:** All details of all identified and quantified peptides passing the peptide filters are listed here. Duplicates are not grouped. If several data files were used in the search a separate export will be created for each file.
- **Normalization factor:** Details of the normalization factor set in the "Summary" view are listed here.

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

**Export Images.** The Annotated Spectrum, the Ion Match table, the Error Map, the Spectrum Alignment, and the Protein Coverage can be exported to an image file. Refer to Section 3.2, "Exporting Images" for details.

## 5.2. Exporting Reporter Ion Labelling Quantification Results

PEAKS Q results can be exported to other supported formats. To export the quantification results, press the Export button in the title bar of the "Summary" view panel. The following export dialogue will appear.



**HTML Report:** The options are the same as those for PEAKS DB, except that only the quantified proteins are used in the exporting. All information found in the "Summary", "Protein", and "Peptide" view of the quantification result will be found in the export. See Section 4.1, "Exporting Summary, Proteins, and Peptides" for details.

**Text Formats:**

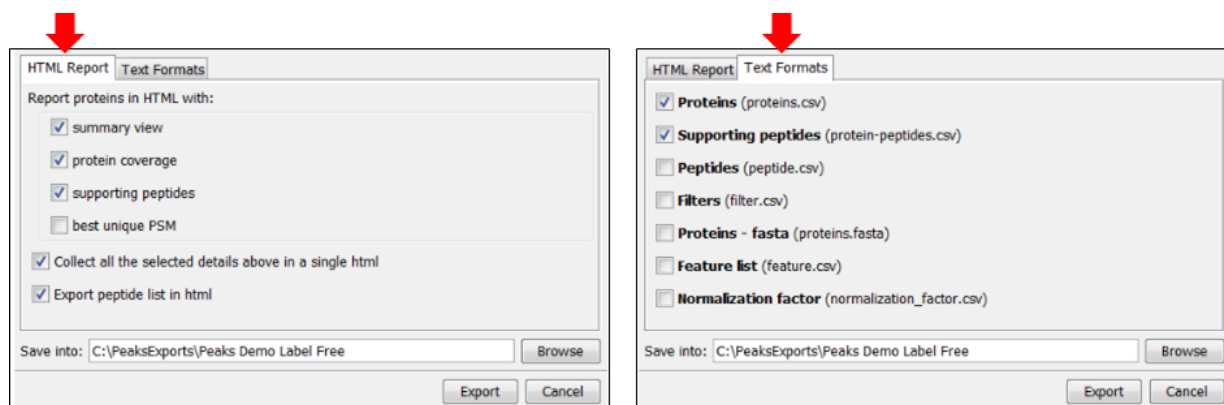
- **Proteins:** These are the quantified proteins that pass the filters set in the "Summary" view and their associated details.
- **Supporting peptides:** These are the supporting peptides of the quantified proteins and their associated details.
- **Filters:** The summary filter, protein view filters, as well as peptide view filters will be saved to filter.csv.
- **Label Q search peptides:** All details of all identified and quantified peptides passing the peptide filters are listed here. Duplicate peptides are grouped.
- **Label Q peptide-spectrum matches:** All details of all identified and quantified PSMs passing the spectrum filters are listed here. Duplicates are not grouped. If several data files were used in the search, a separate export will be created for each file.
- **Normalization factor:** Details of the normalization factor set in the "Summary" view are listed here.

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

**Export Images.** The Annotated Spectrum, the Ion Match table, the Error Map, the Spectrum Alignment, and the Protein Coverage can be exported to an image file. Refer to Section 3.2, "Exporting Images" for details.

### 5.3. Exporting Label-Free Quantification Results

PEAKS LFQ results can be exported to other supported formats. To export the quantification results, press the Export button in the title bar of the "Summary" view panel. The following export dialogue will appear.




**HTML Report:** The options are the same as those for labeled quantification. See Section 5.1, "Exporting Precursor Ion Labelling Quantification Results" for details.

**Text Formats:**

- **Proteins:** These are the quantified proteins that pass the filters set in the "Summary" view and their associated details.
- **Supporting peptides:** These are the supporting peptides of the quantified proteins and their associated details.
- **Peptides:** A complete list of all peptides.
- **Filters:** The summary filter, protein view filters, as well as peptide view filters will be saved to filter.csv.
- **Proteins - fasta:** This is a database file of all quantified protein sequences above the "Summary" view filters.
- **Feature List:** All predicted peptide features and their associated quantification information. Both identified and unidentified peptide features are included.

- **Normalization factor:** Details of the normalization factor set in the "Summary" view are listed here.

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

**Export Images.** The XIC intensity curve, the LC/MS view of the sample features, and the RT alignment figure of the samples can be exported to image files from the "Features" view. Click the  button of the corresponding tabs on the "Features" view. Select the resolution and the location to save the file.

## 6. Export inChorus Result


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

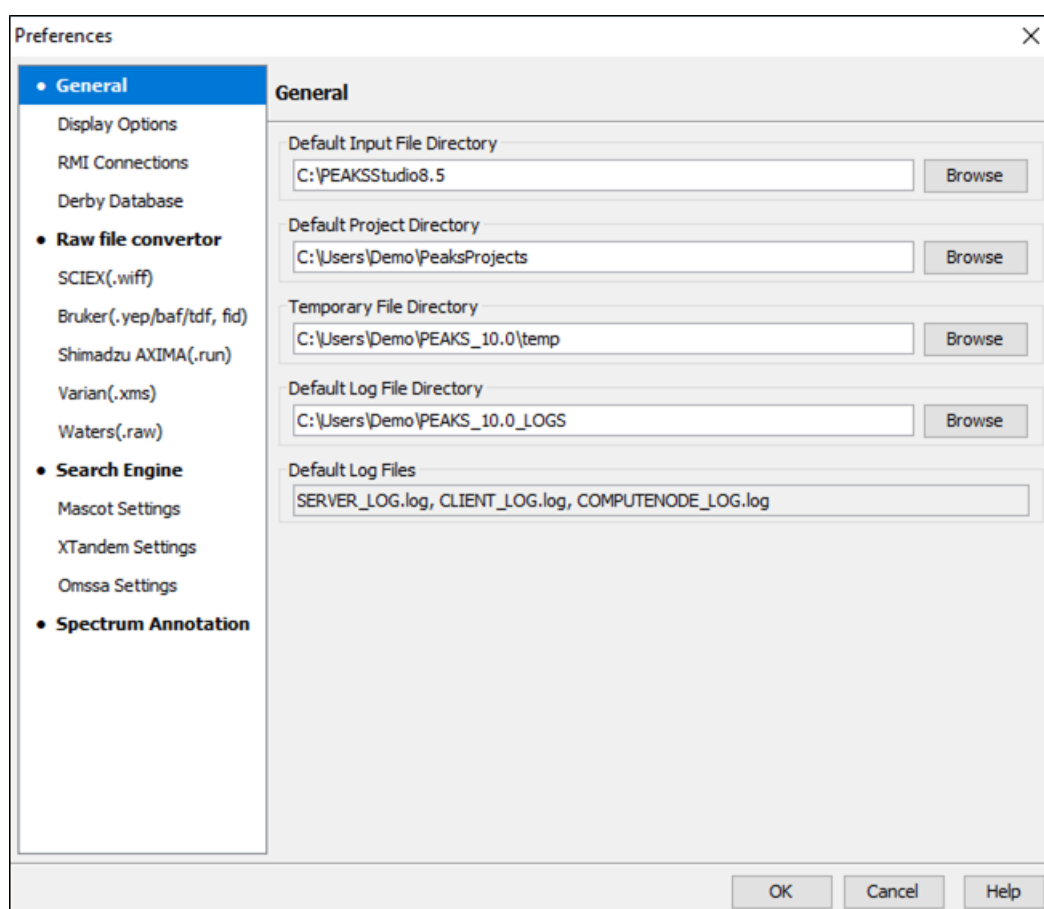
---

# Chapter 20. Advanced Configuration and Environment Preferences

## 1. PEAKS Environment Preferences

This section describes the settings of the environment preferences. These settings include general preferences, raw file converter preferences, search engine preferences, and spectrum annotation preferences.

To begin, click the “Preferences” toolbar icon  or select the Window > Preferences command from the menu bar to open the “Preferences” dialogue. Select the preferences category from the left to view the options available for that category.



### 1.1. General Preferences

**Default Input File Directory:** Select the Browse button to change the default directory, that PEAKS will locate to, when adding data to a project.

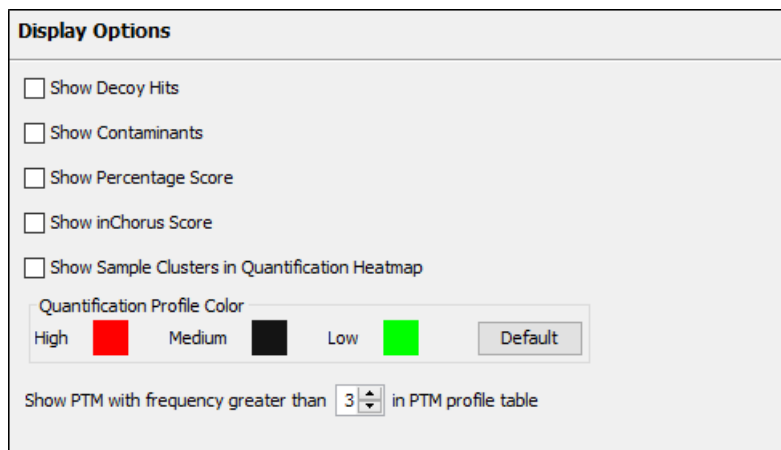
**Default Project Folder:** PEAKS uses [USER\_HOME]/PeaksProjects as the default folder for project files, where USER\_HOME is the user home directory in your system. Select the Browse button to change this location. Please make sure this directory is readable/writable by PEAKS.

**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 [USER\_HOME]/PEAKS\_10.0\_LOGS by default. The locations of these files cannot be changed.

### 1.1.1. Display Options

Select “Display Options” from the left-hand menu to display the interface preferences for PEAKS.



The screenshot shows the "Display Options" dialog box. It contains several checkboxes: "Show Decoy Hits", "Show Contaminants", "Show Percentage Score", "Show inChorus Score", and "Show Sample Clusters in Quantification Heatmap". Below these is a "Quantification Profile Color" section with three color swatches: "High" (red), "Medium" (black), and "Low" (green), along with a "Default" button. At the bottom, there is a text field "Show PTM with frequency greater than" followed by a spinner box set to "3" and the text "in PTM profile table".

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

**Show Contaminant:** Check this to display contaminants that were identified during a PEAKS DB search. To check for contaminants, ensure that a contaminant database is selected during the PEAKS search workflow setup.

**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 the "Feature", "Peptide" and "Protein" views. These changes are applied to the exported files of the PEAKS DB results.

**Show inChorus Score:** Check this to display the percentage score in both the "Peptide" and the "Protein" views of the inChorus results.

**Show Sample Clusters in Quantification Heatmap:** Check this to display the sample clustering in the "Summary" view protein heatmaps of the Quantification results.

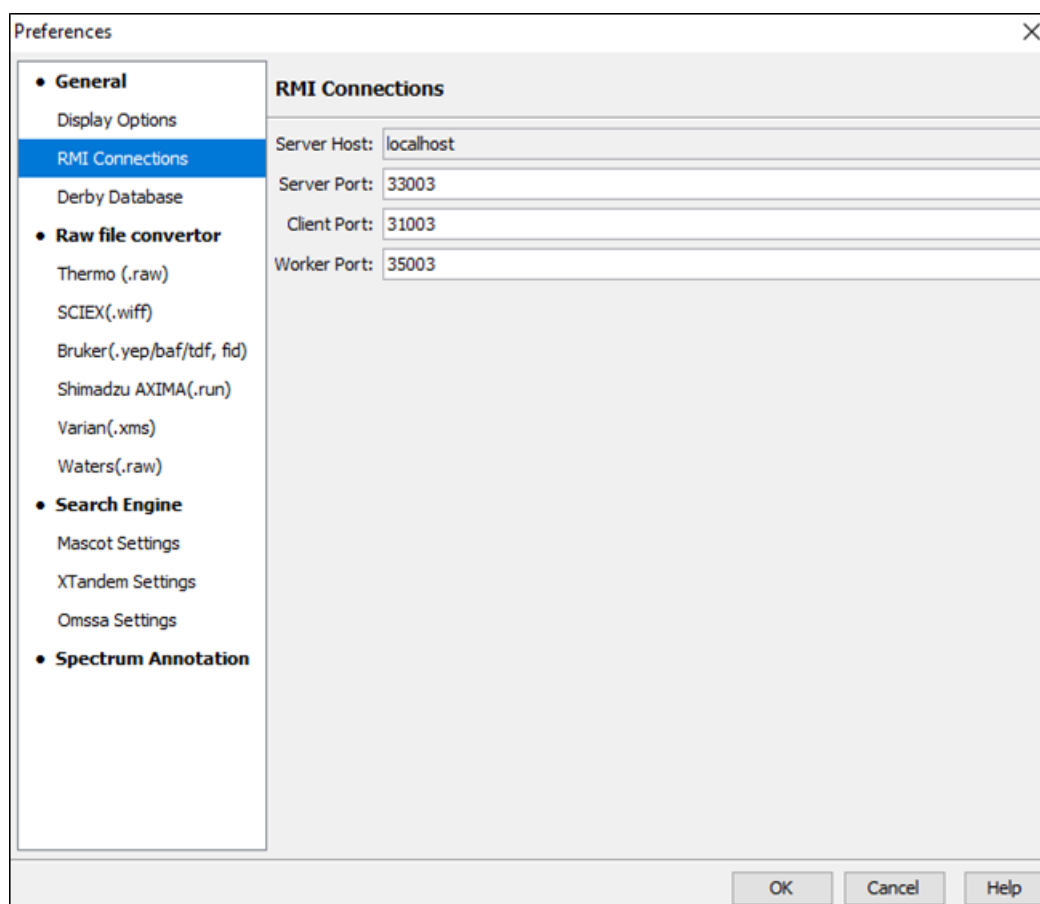
**Quantification Profile Color:** Quantification results shows the relative abundances of peptide and protein in samples/groups as a heatmap. Users can choose to change the colours used for the heatmap display.

**PTM Frequency:** This option sets the PTM display threshold by selecting the minimum number the PTM must appear in the PTM profile table. If there are fewer instances of a PTM identification, within a given protein, than the minimum number set, it will not be displayed in the PTM profile table.

### 1.1.2. RMI Connections

Select “RMI Connections” from the the left-hand menu to display the RMI (Java Remote Method invocation) connection preferences .

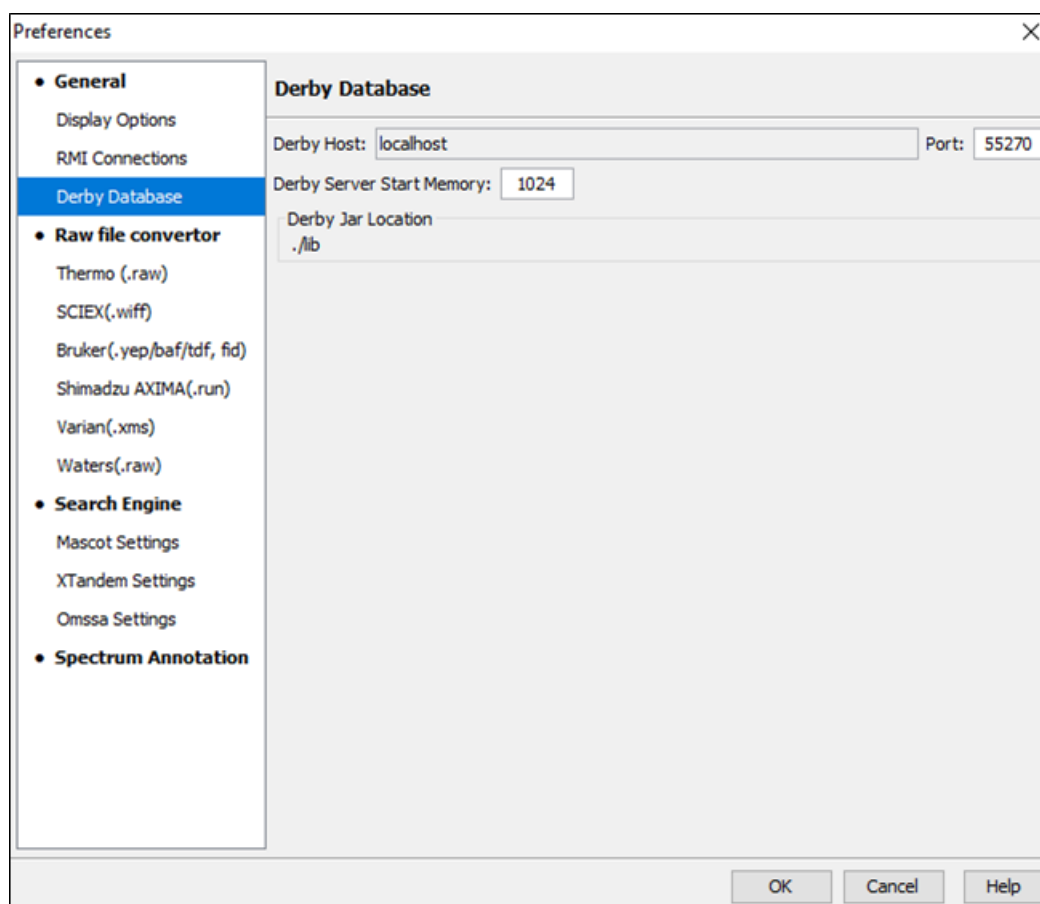




The default port numbers for the Server, Client and Worker will appear by default. If conflicts arise, the port numbers can be changed. Contact technical support at [<support@bioinfor.com>](mailto:support@bioinfor.com) for more information.

### 1.1.3. Derby Database

Select “Derby Database” from the left-hand menu to display the derby database preferences.



**Derby Host:** The name of the “Derby Host”, as well as the “Port” number, will come up by default. The port number can be changed.

**Derby Server Start Memory:** The amount of memory used to start the derby server will also come up by default but can be changed if more memory is available. However, it is not recommended to change this from the default setting. To increase performance, use the performance configuration utility (see Section 5, “PEAKS Performance Configuration”).

**Derby Jar Location:** The “Derby Jar Location” panel will list the location of the derby jar file by default. This is displayed to find its location. This location cannot be changed.

## 1.2. Raw File Converter Preferences

This section allows users to change the preferences for the raw file converters of the following instruments: Thermo, SCIEX, Bruker, Shimadzu, Varian, and Waters.

### 1.2.1. Thermo (.raw)

Click the "Thermo (.raw)" option under the "Raw file converter" section from the menu on the left side to show the Thermo instrument raw file converter preferences.

#### Note

Refer to Section 8.2, “Thermo Data” for details on Thermo (.raw) preferences.

### 1.2.2. SCIEX (.wiff)

Click the “SCIEX (.wiff)” option under the “Raw file converter” section from the menu on the left side to show the preferences for the SCIEX instrument raw file converter.

## Note

Refer to Section 8.3.1, “TripleTOF, QSTAR, or QTRAP” for details on SCIEX (.wiff) preferences.

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

Click the “Bruker (.yep/baf/tdf, fid)” option under the “Raw file converter” section from the menu on the left side to display the Bruker instrument raw file converter preferences.

## Note

Refer to Section 8.4, “Bruker Data” for details on Bruker instrument preferences.

### 1.2.4. Shimadzu AXIMA (.run)

Click the “Shimadzu AXIMA (.run)” option under the “Raw file converter” section from the menu on the left side to show the Shimadzu instrument raw file converter preferences.

## Note

Refer to Section 8.5, “Shimadzu Data” for details on Shimadzu instrument preferences.

### 1.2.5. Varian (.xms)

Click the “Varian (.xms)” option under the “Raw file converter” section from the menu on the left side to display the Varian instrument preferences.

## Note

Refer to Section 8.6, “Varian” for details on Varian instrument preferences.

### 1.2.6. Waters (.raw)

Click the “Waters (.raw)” option under the “Raw file converter” section from the menu on the left side to display the Waters instrument preferences.

## Note

Refer to Section 8.7, “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

Select “Mascot Settings” from the left-hand menu to display the Mascot preferences.

The screenshot shows the 'Preferences' dialog box with the 'Mascot Settings' section selected in the left-hand menu. The 'Mascot Settings' section contains the following fields and options:

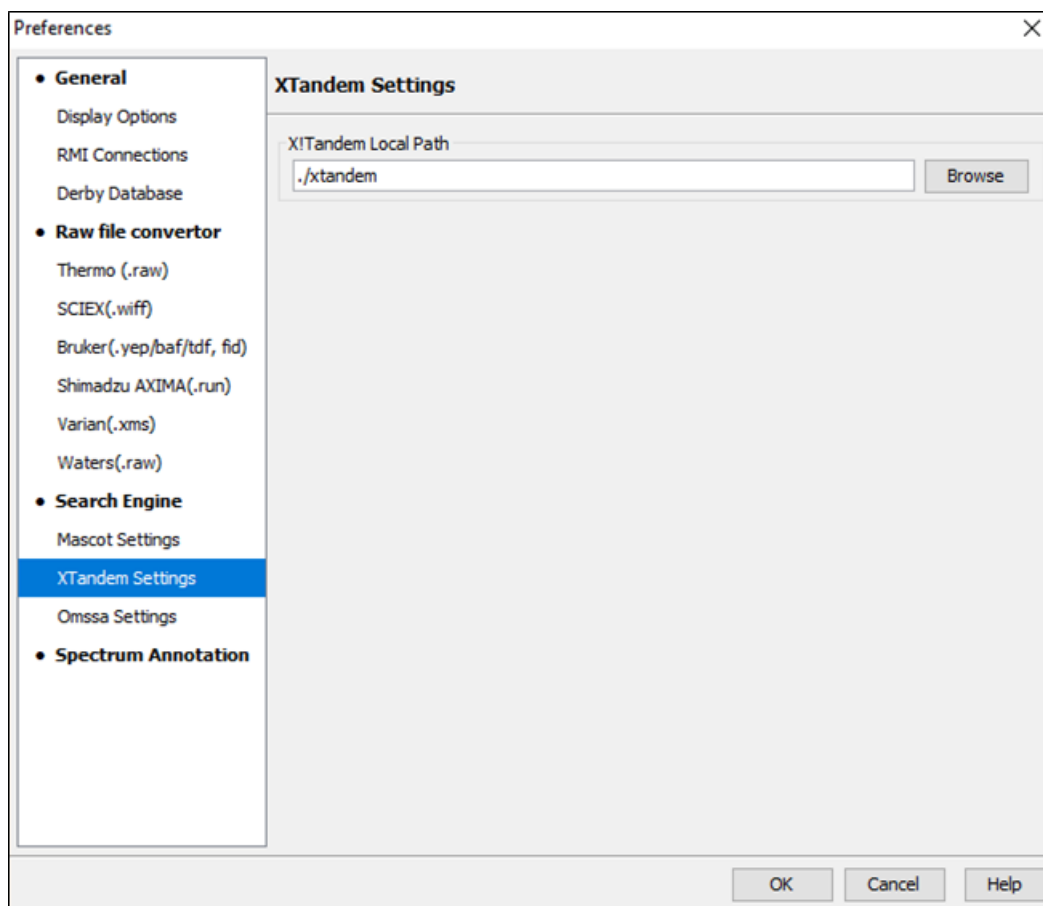
- Hostname (or IP address):** myMascot.mysite.com
- Port:** 80
- Virtual Directory:** /mascot/
- Version:** Mascot Server 2.5.x (dropdown menu)
- User name:** myMascotAccount
- Password:** (masked with dots)
- Email:** test@bioinform.com
- Test Connection:** (button)

At the bottom of the dialog box are three buttons: OK, Cancel, and Help.

These parameters specify how PEAKS accesses the Mascot server (if applicable). Enter the hostname (or an IP address), the port, the virtual directory, and the Mascot server version. In addition, users must also enter the username, password, and email address so that PEAKS can log into the user-defined Mascot server. To make sure that everything is entered correctly and that the server is working, click the Test Connection button. The port and the virtual directory match the above settings for most servers.

### 1.3.2. X!Tandem Settings

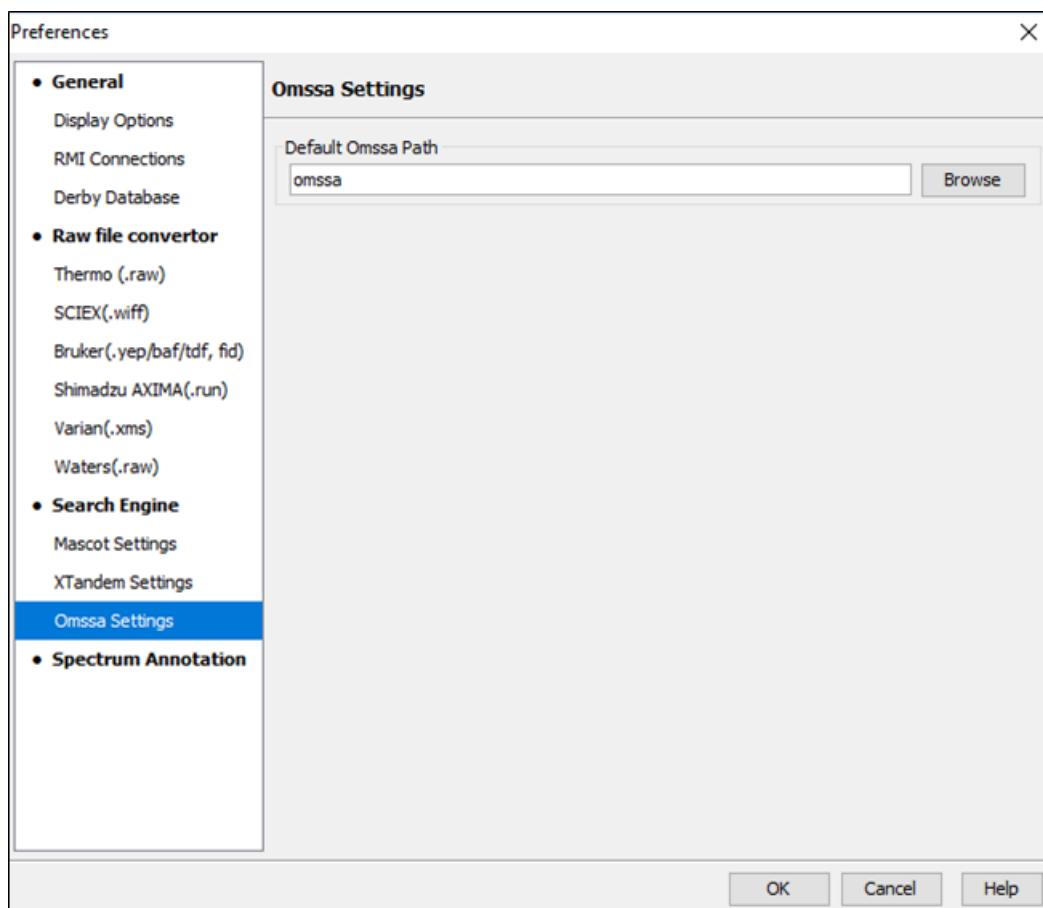
Select “X!Tandem Settings” from the left-hand menu to display the X!Tandem preferences.



As PEAKS provides a local copy of X!Tandem, upon installation, a default path will appear in the Local Settings section. To use another license/location for X!Tandem, click the Browse button to tell PEAKS where to find the search engine.

### 1.3.3. OMSSA Settings

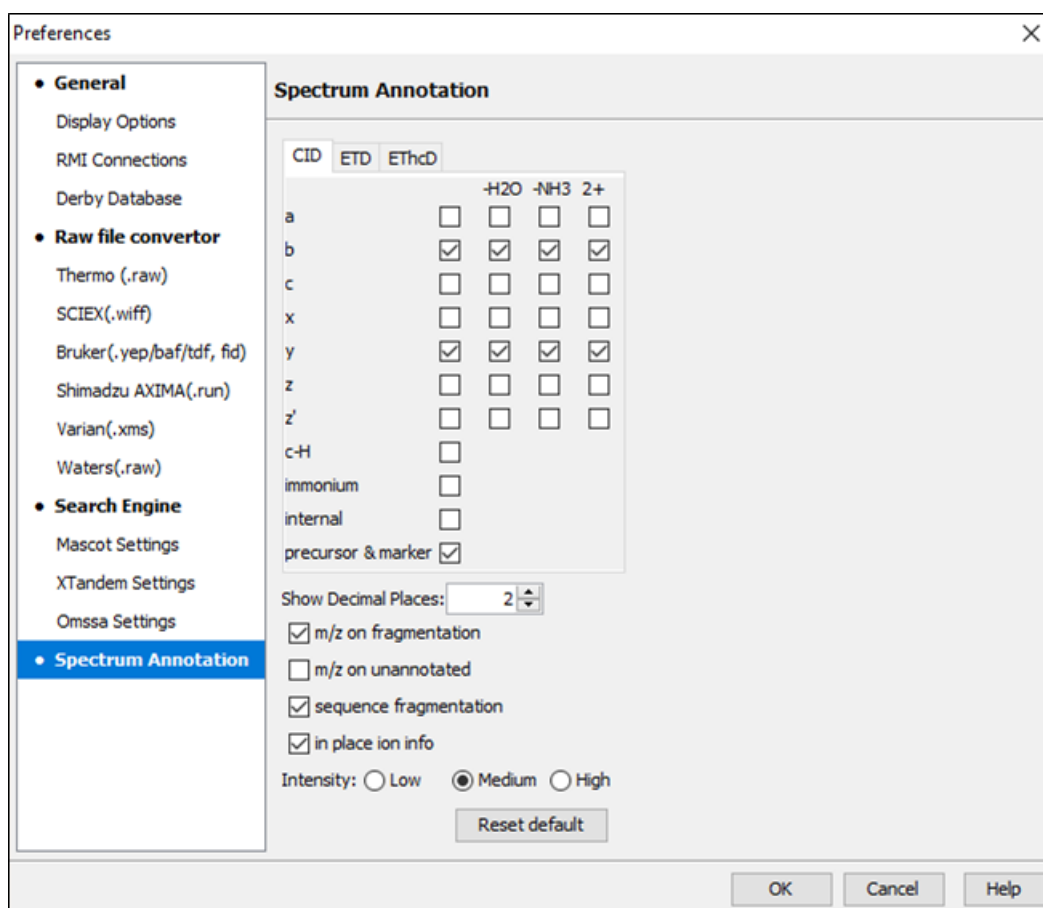
Select “OMSSA Settings” from the left-hand menu to display the OMSSA preferences.



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

Select “Spectrum Annotation” from the left-hand menu to open the spectrum annotation preferences:



The spectra in the PEAKS results can be annotated by selecting ion types from the thorough collection of ions that PEAKS offers. The selected ion types will be displayed in the “Ion Match” table, as well. It is possible to annotate the spectrum with various ions for both CID, EThcD and ETD. By default, y-ion, y-H<sub>2</sub>O, y-NH<sub>3</sub>, y(2+), b-ion, b-H<sub>2</sub>O, b-NH<sub>3</sub>, and b(2+) are selected.

## Note

To reset the settings to the PEAKS defaults, use the “Reset default” button.

**Show Decimal Places:** Select the number of decimal places that will appear in the ion table and the spectrum view. The default is set to two decimal places.

**m/z on Fragmentation:** Select this to display the m/z value on top of the annotated ions.


**m/z on Unannotated:** Select this to display the m/z value on top of the peaks without ion matches.

**Sequence Fragmentation:** Select this to display the sequence fragmentation on the top-left corner of the “Spectrum Annotation” view.

**In Place Ion Info:** Ion information, m/z value, and relative intensity are all displayed in a pop-up within the “Spectrum Annotation” view when this option is checked and the cursor is placed on a peak.

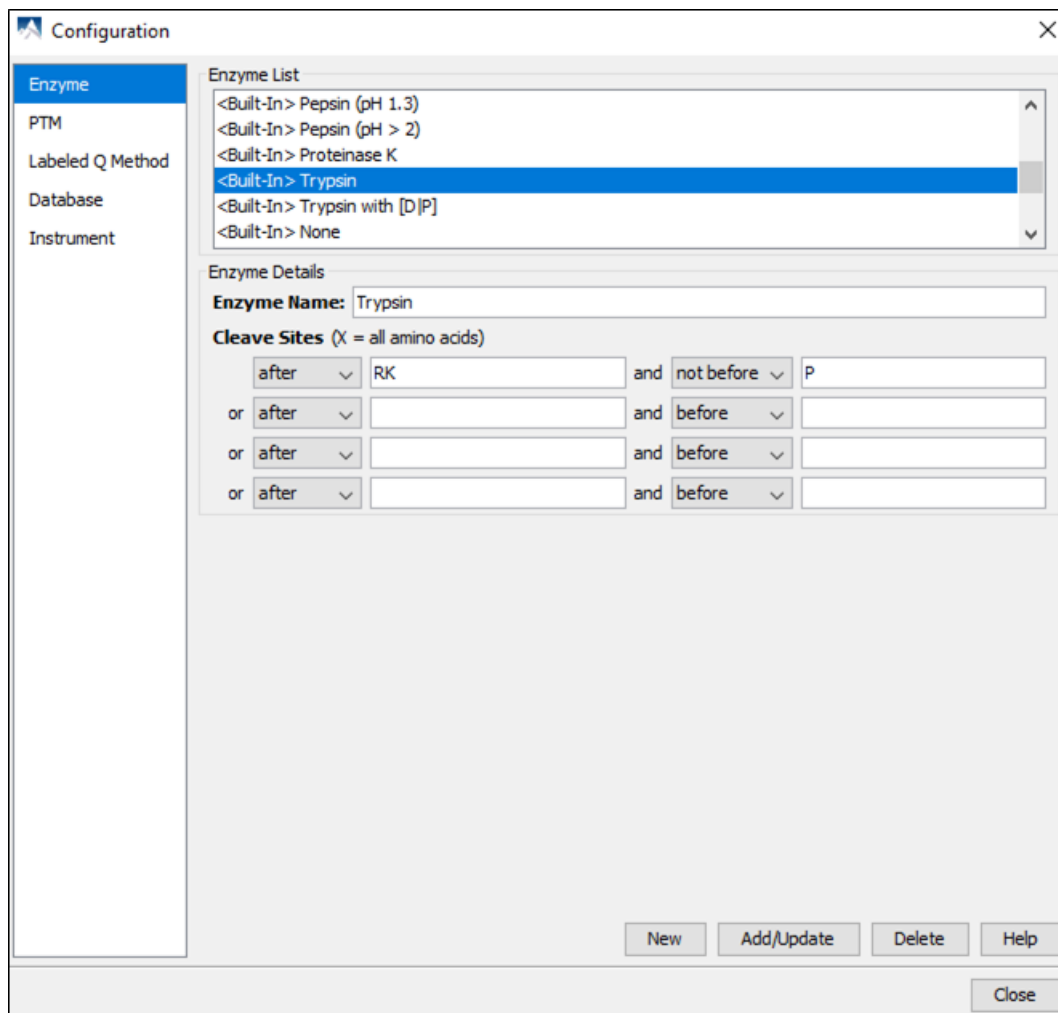
**Intensity -** Set the intensity threshold for spectrum annotation to low (2%), medium (5%), or high (10%). To apply this intensity threshold, select the “intensity threshold” checkbox in the “Spectrum Annotation” view (see Section 3.2.2, “Spectrum Annotation”).

## 2. PEAKS Configuration

This section describes the configuration of enzymes, PTMs, databases, instruments, and parameters. To begin, click the Configuration toolbar icon  or select “Configuration” from the “Window” menu.

## 2.1. Enzyme Configuration

PEAKS can use almost any single enzyme or combination of enzymes to analyze data. Select built-in enzymes from the extensive list provided in PEAKS or define a new one. From the “Configuration” window select “Enzyme” from the left side menu to change the enzyme configuration.



**Built-in Enzymes.** All of the built-in enzymes within PEAKS are listed in the “Enzyme List”. Click on one of these built-in enzymes to view the information about that enzyme in the “Enzyme Details” panel.

### Note

A built-in enzyme cannot be deleted or edited.

**Create a New Enzyme.** Provide the name of the new enzyme in the “Enzyme Name” field and specify how the custom enzyme will cleave the protein between two amino acids to create peptides in the “Enzyme Details” panel. The letter, X, denotes *any amino acid at this position*, while amino acids within the { and } brackets indicate any amino acid except for the one in the brackets. Choose where the cleavage sites are by selecting *after* or *not after* and *before* or *not before* to specify the range. Add multiple amino acids to indicate that cleavage happens before or after any of the stated amino acids. For example, after RK means after R or K not after R and K.

Click the Add/Update button to save the changes. The new enzyme will now appear in the “Enzyme List”, where it can be accessed later. To delete a customized enzyme, select the appropriate enzyme and click the Delete button.

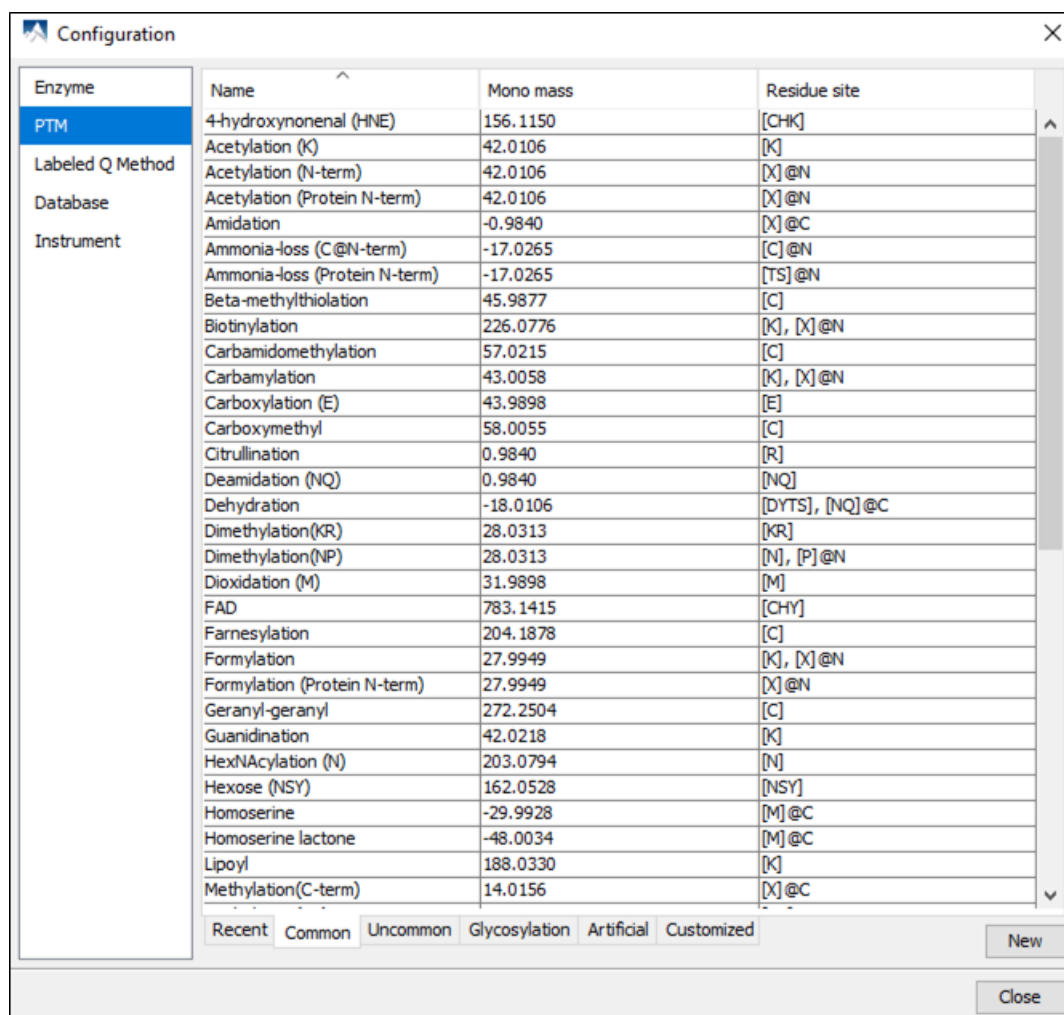


## Note

For information on defining new enzymes in the middle of a workflow, (ex. While setting up a PEAKS DB search), refer to Section 2.2, “Enzyme”.

## 2.2. PTM Configuration

From the “Configuration” window, select “PTM” from the menu on the left-hand side to change the PTM configuration.



**Built-in PTMs.** The built-in standard PTMs within PEAKS are organized into four categories: “Common”, “Uncommon”, “Glycosylation”, and “Artificial”. The “Common” list contains the most commonly used PTMs, the “Uncommon” list contains less frequently used PTMs, the “Glycosylation” list contains all the glycosylation modifications recorded in the UniMod database, and the “Artificial” list contains the PTMs that can only be induced artificially. In addition, users will find a “Recent” tab which contains PTMs that were used recently in PEAKS, and a “Customized” tab which contains all the user-defined PTMs. Double-clicking on any of these PTMs will display the PTM details in the “PTM Info” pop-up dialogue. The same information can be viewed by selecting a PTM from a list and clicking the View button. For details on PTMs, refer to Section 2.3, “PTM”.

PTM name:

PTM abbreviation:

Mass (Monoisotopic):

Residues that can be modified:  Anywhere

☒ peptide ☐ protein

N-term

C-term

Formula:

Rule:

OK Cancel Help

**Create a new PTM.** To create a new PTM, click on the New button in the bottom right-hand corner of the PTM Configuration window to display the “New PTM” dialogue. Provide the necessary information about the new PTM.

PTM name:

PTM abbreviation:

Mass (Monoisotopic):

Residues that can be modified:  Anywhere

☒ peptide ☐ protein

N-term

C-term

Formula:

Rule:

OK Cancel Help

- **PTM Name:** This name will appear in the PTM list for future use after it is saved.
- **PTM abbreviation:** The PTM name expressed in a shortened form.
- **Mass (Monoisotopic):** The mass that the residue gains or loses as a result of the PTM.
- **Residues that can be modified:** Enter residues that can be modified anywhere, or residues that can only be modified if they are at the N- or C-terminus. User will also need to specify if these residues can be modified on the peptide-level or protein-level.
- **Formula:** The chemical formula of the PTM. This should correspond to the mass listed above.
- **Rule:** This field is used for the user's reference only. Users can use the field to enter a comment about the PTM.

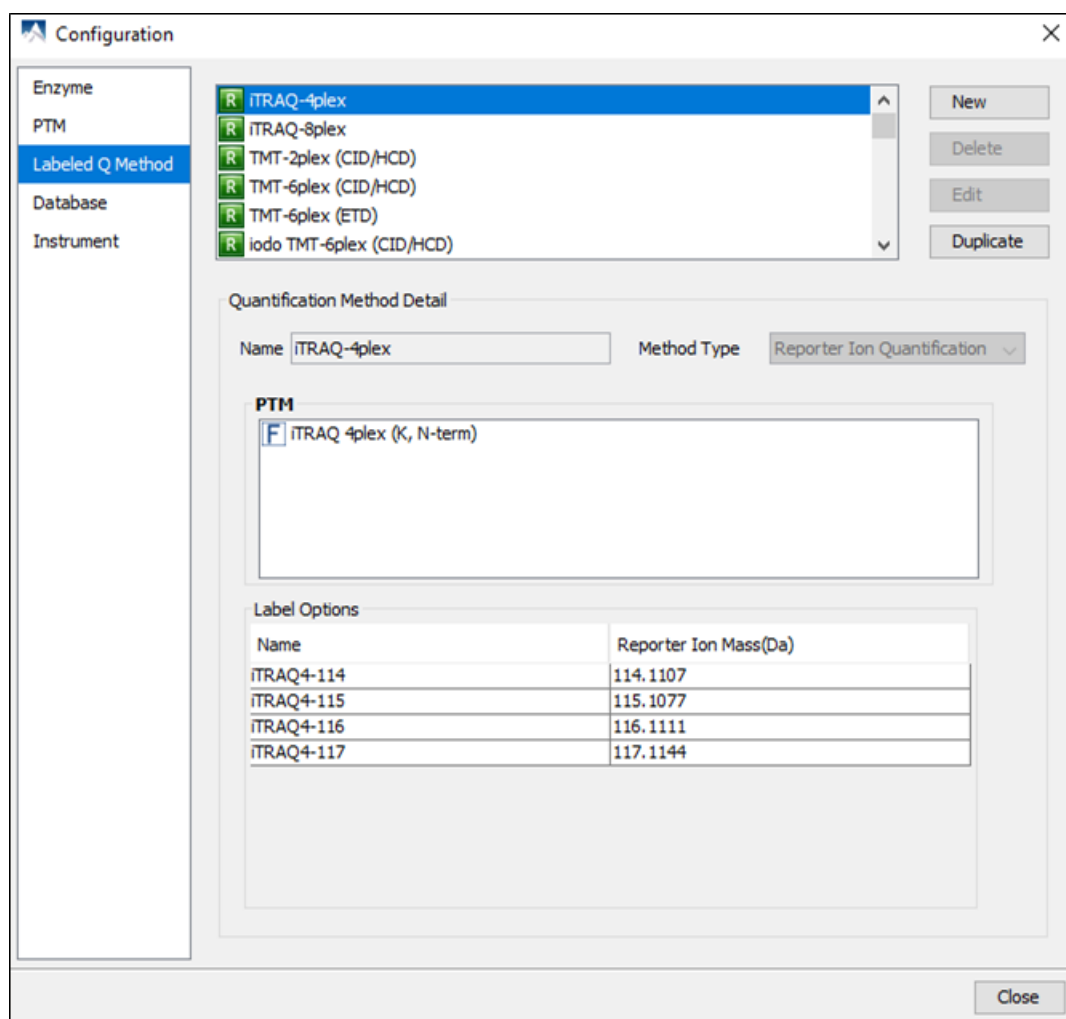
Click the OK button to save the changes. The new PTM will now appear in the “Customized” PTM list, where it can be accessed later. To delete a customized PTM, select the appropriate PTM from the list and click the Delete button.

## Note

For information on defining new PTMs in the middle of a workflow, (ex. While setting up a PEAKS DB search), refer to the section Section 2.3, “PTM”.

## 2.3. Labeled Q Method Configuration

From the “Configuration” window, select “Labeled Q Method” from the left-hand menu to change the labeled Q method configuration.



The built-in labeled Q methods are listed in the methods list. Select a method from the list to view detailed information in the “Quantification Method Detail” panel. Methods with the R icon beside the name are reporter level methods. Methods with the P icon beside the name are precursor level methods.

**Name and Method Type:** The name and the type of quantification method. The method type can be “Reporter Ion Quantification” or “Precursor Ion Quantification”.

**PTM:** The post-translational modification associated with the selected Quantification method; including the residue site and the location on the residue at which it is attached.

**Label Options:** List of quantifiable labels. For reporter ion quantification methods, the Label Options pane will contain the name and reporter ion mass. For precursor ion quantification methods, the Label Options pane will contain the sample name, the modification, and the modification details.

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

Quantification Method Detail

Name  Method Type **Reporter Ion Quantification** ▼

**PTM**

**Label Options**



Name	Reporter Ion Mass(Da)

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

Quantification Method Detail

Name  Method Type **Precursor Ion Quantification** ▼

**Label Options**

Name	Modi		Modi Detail	R to P Modi
Heavy	SILAC K6		6.0201@[K]	
Light	None			

Use Add row and Delete Row to respectively add or remove a label for a “Precursor Ion Quantification” method. Each label is defined by the sample name, the added mass, the target residue, and the labeling efficiency. If one sample has multiple labels with different mass shifts, a user can add multiple labels with the same sample name. These labels will contribute to the same sample in the ratio. The modification for each label must be selected from the PTM lists, which can be accessed by clicking the button in the “Modifications” column.

## 2.4. Database Configuration

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

From the “Configuration” window, select “Database” from the left menu to change the database configuration. The “Database List” on the top shows all the configured databases in the system. Select a database from the list to view the detailed information about the database in the “Database Details” panel.

The screenshot shows the 'Database' tab of the 'Configuration and Preferences' dialog. On the left is a sidebar with options: Enzyme, PTM, Labeled Q Method, Database (selected), and Instrument. The main area is divided into sections: 'Database List' with a list box containing '[UniProt\_SwissProt]' and '[NCBI\_nr]' (highlighted); 'Database Details' with a dropdown for 'FASTA format database' set to 'NCBI nr', and buttons for 'Validate Database' and 'Export Decoy DB'; 'Basic Options' with fields for 'Database name', 'Path' (with 'Browse' and 'Download' buttons), and a checkbox for 'EST database'; 'Advanced Options - Fasta Title Format' with fields for 'Rule to parse accession/id from FASTA title' (containing '\(gi\|d\*\)'), 'Rule to parse description from FASTA title' (containing '\s+\. \*'), 'Accession/id URL' (containing 'http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?db=protein&val=' and '<Accession/ID>'), and 'Delimiter' (containing '\s+\. \*'); and 'Taxonomy Options' with 'taxonid' and 'taxdmp' fields, each with 'Browse' and 'Download' buttons. At the bottom are buttons for 'New', 'Add/Update', 'Delete', 'Help', and 'Close'.

**Create a new database.** Click the New button to create a new sequence database entry or select a database from the “Database List”. The database can be configured in the “Database Details” area below. Click the Add/Update button to save any changes. Refer to Chapter 6, *Adding a Sequence Database* for details on adding a new database.

**Delete a previously configured database.** To delete a database file, select the database to be deleted from the “Database List” and click the Delete button at the bottom.

**Moving/Updating a database.** To move a database to another directory, the location must be updated in PEAKS. Select the database and specify the new location using the Browse button next to the “Path” field. Then, click Add/Update to save the new settings. If the database location is invalid, the database name will appear in *red* in the list of databases. Any protein identification using that database will fail. If an update is made to the database file (perhaps by downloading the latest database file and overwriting the old database file), PEAKS will show the database information in *light gray*. A light gray color could also mean that the database does not have header information.

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

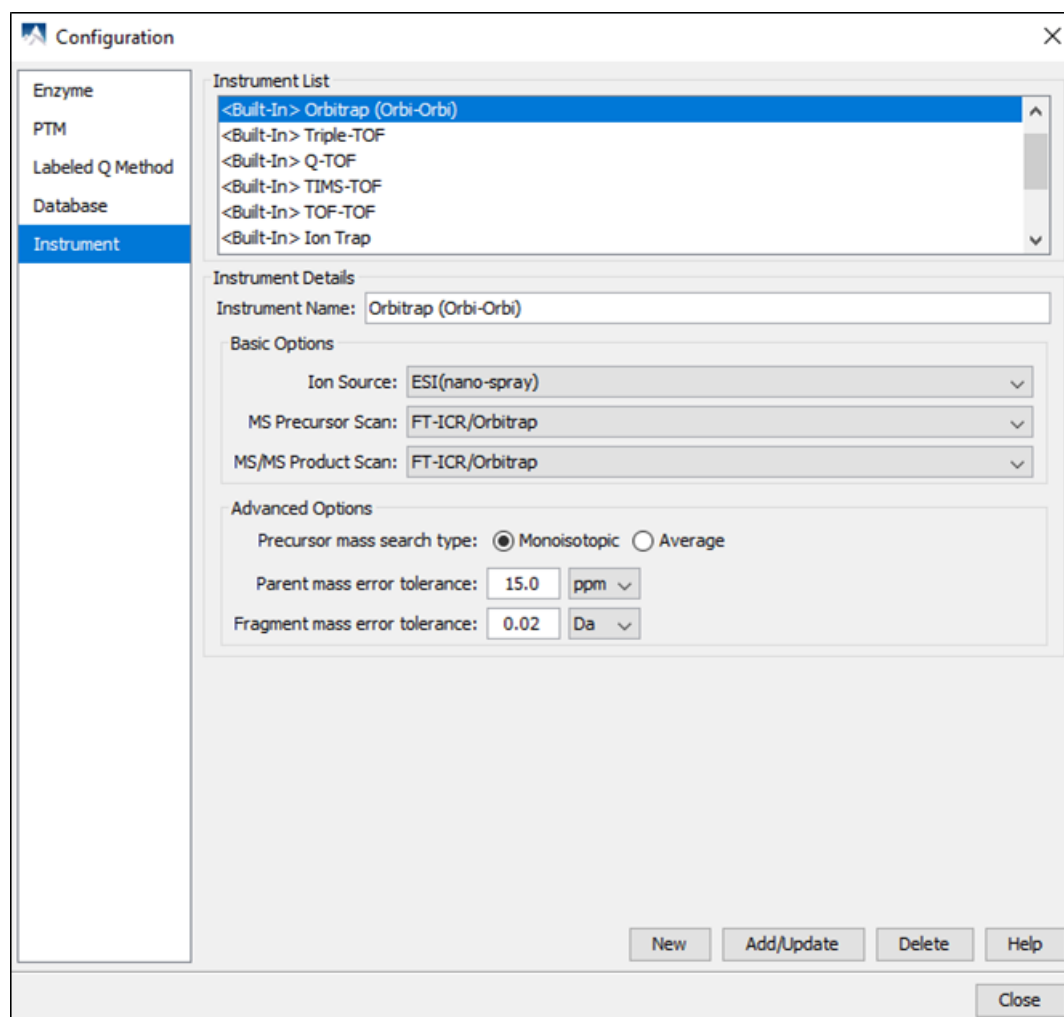
- **X!Tandem:** At the time of writing, X!Tandem has difficulty in searching through large databases and may crash during a search. Therefore, it is suggested that X!Tandem only be used with small databases. If used with a large database, a taxon should be specified. For example, NCBI-nr and SwissProt databases should be used with subtaxa selected when using X!Tandem.
- **OMSSA:** At the time of writing, OMSSA cannot be used with databases that are not in NCBI-nr, or SwissProt format in a way that is available to inChorus. Also, a bug in OMSSA results in a database, stored in a folder that

contains a space in its path, to be unusable. This creates problems when PEAKS creates temporary databases on your behalf. To avoid this, best practices suggest that all of the databases should be put in a folder with no spaces, such as `C:/peaksdatabases`. Note that the folder `C:/My Documents/databases` does not work as it contains a space between 'My' and 'Documents'. Using spaces in the database file name causes the same problem. For example, once the NCBI-nr database is downloaded and extracted, save the database as `ncbinr.fas`, or `ncbi_nr.fas` rather than `ncbi nr.fas`.

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

## 2.5. Instrument Configuration

From the “Configuration” window, select “Instrument” from the left-hand menu to change the instrument configuration.



**Built-in Instruments.** The names of the built-in instruments are provided in the “Instrument List”. Select an instrument to view the detailed instrument information in the “Instrument details” panel below.

### Note

The details of a built-in instrument cannot be deleted or edited.

#### Create a new instrument.

1. Click the New button and provide a name for the instrument in the “Instrument Details” panel.

2. Use the “Ion Source” drop-down menu to select the ion source that was used: MALDI/SELDI or ESI (nano-spray). This will help the PEAKS Data Refine tool determine the charge of the ions.
3. Use the “MS Precursor Scan” drop-down menu to select the type of MS scan that was performed. This selection will tell the PEAKS Data Refine tool whether the survey scan is of sufficient resolution to determine the charge and the monoisotopic peak during examination of the survey scan.
4. Use the “MS/MS Product Scan” drop-down menu to select the type of MS/MS scan that was performed. This selection will help PEAKS decide which internal parameters (for weighing fragments and amount of noise) to use during PEAKS auto *de novo* sequencing and a PEAKS DB search. Select LIT/FT if alternating high-res/low-res modes were used. This will allow the algorithm to determine the mass analyzer from the scan header.
5. Use the “Advanced Options” to specify additional parameters.
6. Select “Monoisotopic” or “Average” as “Precursor Mass Search Type”. For ion-trap instruments, it is usually beneficial to allow PEAKS DB search to use an average mass.
7. Specify the values for “Parent mass error tolerance” and “Fragment mass error tolerance” in Daltons or ppm. These will appear on the PEAKS *de novo* and PEAKS DB option screens when the instrument is selected.
8. Click the Add/Update button to save the changes. The new instrument will appear in the “Instrument List” where it can be accessed when creating a new project file. To delete a customized instrument, select the appropriate instrument from the “Instrument List” and click the Delete button.