PEAKS Online 11 User Manual

PEAKS Online Team

PEAKS Online 11 User Manual

PEAKS Online Team

Publication date 2023

Table of Contents

1. Overview 1
    1. How to Use This Manual 1
    2. What is PEAKS Online ? 1
    3. What is New in PEAKS Online 11 ? 3
    4. Supported Data Types 4
    5. Quick Walkthrough 4
        5.1. PEAKS Online Main GUI 5
            5.1.1. PEAKS Online User Options 5
        5.2. Creating a PEAKS Online Project 6
        5.3. Conducting an Analysis 7
        5.4. Project Summary, Monitoring and Results 8
        5.5. Opening an Existing Project 9
        5.6. Batch Operation in Project listing 9
        5.7. Analysis Properties 10
2. Management Configuration 11
    1. Database Management and Configuration 11
    2. Library Management and Configuration 12
        2.1. Create a New Library 13
        2.2. Import a New Library 14
    3. Modification Management and Configuration 15
        3.1. Modifications 15
            3.1.1. Modification Types 16
        3.2. Create a New Modification 16
    4. Enzyme Management and Configuration 17
        4.1. Create a New Enzyme 17
    5. Instrument Management and Configuration 18
        5.1. Create a New Instrument 18
    6. TMT/iTRAQ and custom Label Q Methods 19
    7. SILAC Q Methods 21
    8. Workflows 23
    9. Account Settings 24
        9.1. My account 24
        9.2. Algorithm Parameters 25
            9.2.1. PEAKS PTM Algorithm Parameters 25
            9.2.2. DIA DeepNovo Peptidome 26
            9.2.3. De Novo Engine 26
3. Archive Project 27
    1. Overview 27
    2. Archive Project 27
        2.1. Create a new folder in Archive Repository 27
        2.2. Archive Project in Archive Repository 28
    3. Import Archived Project 28
        3.1. Import Archived Projects Page 29
4. Data View 31
    1. Overview 31
    2. Accessing the Data View 31
    3. Navigating the Data View 31
5. Peptide *De Novo* Sequencing 33
    1. Overview 33
        1.1. Standalone *DeepNovo/De Novo* Analysis 33
    2. *DeepNovo/De Novo* Analysis Workflow and Parameter Settings 34
        2.1. Precursor and Fragment Error Tolerance 36
        2.2. Enzyme Settings 36
        2.3. PTM settings 37
    3. Understanding PEAKS *De Novo* Sequencing Results 37

3.1. Result Filtration 37
3.2. Exporting *De Novo* Results 38
3.3. Summary View 38
3.3.1. Result Statistics 38
3.3.2. Low Quality Data Warnings 39
3.4. Peptide Table 39
3.5. Spectrum Annotation 40
3.6. Error Map 41
3.7. Ion Table 41
4. FAIMS Data 41
6. Peptide, PTM, and Mutation Identification (PEAKS DB, PEAKS PTM, and SPIDER) 43
1. Overview 43
2. Database Search Workflow and Parameter Settings 43
2.1. Data Refinement Parameters 46
2.2. Precursor and Fragment Error Tolerance 46
2.3. Enzyme Settings 47
2.4. Database Settings 47
2.5. PTM Settings 47
2.6. Find Unspecified PTMs with PEAKS PTM 48
2.6.1. PTM signature Ions 48
2.7. Find More Mutations with SPIDER 49
2.8. PSM Filter 49
3. Understanding PEAKS Search Results 49
3.1. Result Filtration 50
3.1.1. Confident PTM Site Filter 50
3.1.2. Show Results for All Samples 50
3.1.3. Show Results for Specific Sample 50
3.2. Result Exporting 51
3.3. Summary View 52
3.3.1. False Discovery Rate (FDR) Curve 53
3.3.2. Result Statistics Tables 53
3.3.3. Summary Figures 55
3.3.4. Low Quality Data Warnings 55
3.4. Protein View 56
3.4.1. Protein View Filters 56
3.4.2. Protein Table 57
3.4.3. Global Protein Comparative View 58
3.4.4. Protein Coverage 58
3.4.5. Supporting Peptides 59
3.4.6. PTM Profiling 60
3.5. Peptide View 63
3.5.1. Peptide View Filters 63
3.5.2. Peptide Table 64
3.5.3. Peptide Global Comparative View 65
3.5.4. Peptide Menu 65
3.5.5. Peptide-Spectrum Matches 66
3.6. De Novo Only View 66
4. Running PEAKS PTM and SPIDER Separately 66
5. FAIMS Data 66
7. DIA Streamlined Identification 68
1. Overview 68
2. Adding a Spectral Library from the Parameters Page 68
3. Spectral Library Search Workflow and Parameter Settings 68
3.1. Precursor and Fragment Error Tolerance 69
3.1.1. Optimize Tolerance 70
3.1.2. Optimize for Short Gradient 70
3.2. Spectral Library 70
3.3. Protein Inference Settings 70

3.4. Report Filter 70
4. Understanding PEAKS Spectral Library Results 70
4.1. Result Filtration 71
4.1.1. Show Results for All Samples 71
4.1.2. Show Results for Specific Sample 71
4.2. Result Exporting 71
4.3. Summary View 72
4.3.1. Summary Statistics 72
4.3.2. False discovery rate (FDR) curve 73
4.3.3. Statistics of filtered results 73
4.3.4. Precusor mass error 73
4.3.5. RT Calibration 74
4.3.6. Box plot for Rt change 74
4.3.7. Number of identified peptides in each sample by number of missed cleavages 74
4.3.8. PTM Profile 74
4.3.9. Low Quality Data Warnings 74
4.4. Protein View 75
4.4.1. Protein View Filters 75
4.4.2. Protein Table 76
4.4.3. Global Protein Comparative View 76
4.4.4. Protein Coverage 77
4.4.5. Supporting Peptides 78
4.5. Peptide View 78
4.5.1. Peptide View Filters 78
4.5.2. Peptide Table 79
4.5.3. Peptide Global Comparative View 80
4.5.4. Peptide Menu 81
4.5.5. Protein Selection 81
4.5.6. Peptide-Spectrum Matches 81
4.6. Manual Search 83
4.6.1. Performing a Manual Search 83
4.6.2. Manual Search Input 84
4.6.3. Viewing Manual Search Results 85
4.6.4. Peptide Table 85
4.6.5. Peptide Information 86
5. DIA Database Search Workflow and Parameter Settings 86
5.1. Precursor and Fragment Error Tolerance 87
5.1.1. Optimize Tolerance 87
5.2. Enzyme Settings 88
5.3. Database Settings 88
5.4. PTM Settings 88
5.5. PSM Filter 88
6. Understanding PEAKS DIA Database Search Results 89
6.1. Result Filtration 89
6.1.1. Show Results for All Samples 89
6.1.2. Show Results for Specific Sample 90
6.2. Result Exporting 90
6.3. Summary View 91
6.3.1. Summary Statistics 91
6.3.2. False discovery rate (FDR) curve 91
6.3.3. Statistics of filtered results 92
6.3.4. Precursor mass error 92
6.3.5. RT Calibration 92
6.3.6. Box plot for Rt change 92
6.3.7. Number of identified peptides in each sample by number of missed cleavages 92
6.3.8. PTM Profile 92
6.3.9. Low Quality Data Warnings 93
6.4. Protein View 93

6.4.1. Protein View Filters 94
6.4.2. Protein Table 95
6.4.3. Global Protein Comparative View 95
6.4.4. Protein Coverage 95
6.4.5. Supporting Peptides 97
6.5. Peptide View 97
6.5.1. Peptide View Filter 98
6.5.2. Peptide Table 98
6.5.3. Peptide Global Comparative View 99
6.5.4. Peptide Menu 100
6.5.5. Peptide-Spectrum Matches 100
6.5.6. DIA Database Search Mirror Plot 100
6.5.7. Error Map 101
6.5.8. Ion Table 101
6.5.9. Precursor Profile 101
6.5.10. Fragment Ion XIC 101
6.5.11. TimsTOF In Frame Mode 101
7. DIA *De Novo* Analysis Workflow and Parameter Settings 102
7.1. Precursor and Fragment Error Tolerance 102
7.1.1. Optimize Tolerance 103
7.2. Enzyme Settings 103
7.3. PTM Settings 103
7.4. *De Novo* Filter 103
8. Understanding PEAKS DIA *De Novo* Sequencing Results 104
8.1. Result Filtration 104
8.2. Exporting *De Novo* Results 104
8.3. Summary View 105
8.3.1. Result Statistics 105
8.3.2. Low Quality Data Warnings 106
8.4. Peptide Table 106
8.5. Spectrum Annotation 107
8.6. Spectrum View 108
8.6.1. Raw Scan 108
8.6.2. Mirror Plot 108
8.6.3. Pseudo Scan 109
8.7. Error Map 109
8.8. Ion Table 109
8.9. Precursor Profile 109
8.10. Fragment Ion XIC 109
8.11. TimsTOF In Frame Mode 110
9. FAIMS Data 110
8. Label Free Quantification (LFQ) 111
1. Overview 111
2. Setting Parameters 112
3. Understanding the LFQ Results 115
3.1. Protein Filters 116
3.1.1. Protein Filter 116
3.2. Summary View 117
3.2.1. Statistics of Filtered Results 117
3.2.2. The Volcano Plot for Proteins 117
3.2.3. Protein Profile Heatmap 117
3.2.4. RT Shift & M/Z Shift 117
3.2.5. Percentage of missing values in each sample with or without ID transfer 117
3.2.6. Protein Area Coefficient of Variance. 117
3.2.7. Low Quality Data Warnings 118
3.3. Protein View 118
3.3.1. Protein Table 118
3.3.2. Protein Volcano Plot 119

3.3.3. Coverage view 119
3.3.4. Peptides view 120
3.3.5. Protein View Filter 121
3.4. Peptide View 122
3.4.1. Peptide Table 122
3.4.2. Feature Details 122
4. Ion Mobility Spectrometry (IMS) Data 123
4.1. Result Exporting 124
9. PEAKS Quality Control 68
1. Overview 68
2. QC Parameters 126
2.1. Quality Control Attributes 127
2.1.1. DDA QC Attributes 127
2.1.2. DIA QC Attributes 127
3. PEAKS QC Results 127
3.1. QC Summary Tab 127
3.2. QC Data Tab 128
3.2.1. Data QC Figures 129
3.3. QC Identification Tab 130
3.3.1. Identification QC Figures 131
3.4. QC Quantification Tab 133
3.4.1. Quantification QC Figures 134
3.5. QC Standard Peptides Tab 135
3.6. QC Results Exporting 136
3.6.1. Result Page Exporting 136
3.6.2. Analysis Export Page 137
10. Isobaric Labelling Quantification (TMT/iTRAQ) 138
1. Overview 138
2. Setting Parameters 139
2.1. Spectrum Filter: 139
2.2. Protein Filter: 139
2.3. Experiment Setting: 139
3. Understanding the Results 141
3.1. Result Filtration 142
3.2. Result Exporting 142
3.3. Summary View 143
3.3.1. Statistics of Filtered Results 143
3.3.2. The Volcano Plot for Proteins 143
3.3.3. Protein Profile Heatmap 143
3.3.4. Low Quality Data Warnings 143
3.4. Protein View 144
3.4.1. Protein Table 145
3.4.2. Protein Volcano Plot 145
3.4.3. Protein Coverage 146
3.4.4. Support Peptides 146
3.4.5. Protein View Filter 147
3.5. Peptide View 148
3.5.1. Peptide Table 148
3.5.2. Ion Match 148
3.5.3. Reporter Ion 148
4. Ion Mobility Data 149
11. Stable Isotope Labeling by Amino acids in Cell culture (SILAC) 150
1. Overview 150
2. Setting Parameters 151
3. Understanding the SILAC Results 154
3.1. Result Filtration 155
3.1.1. Protein Filter 155
3.2. Summary View 156

3.2.1. Statistics of Filtered Results 156
3.2.2. Protein Profile Heatmap 156
3.2.3. Ratio Distribution 156
3.2.4. The Volcano Plot for Proteins 156
3.2.5. Low Quality Data Warnings 156
3.3. Protein View 157
3.3.1. Protein Table 157
3.3.2. Protein Volcano Plot 158
3.3.3. Protein Coverage 158
3.3.4. Support Peptides 159
3.3.5. Protein View Filter 159
3.4. Peptide View 160
3.4.1. Peptide Table 160
3.4.2. Peptide View Filters 161
3.4.3. Feature Details 161
3.5. Result Exporting 163
12. Fraciton Assisted ID Transfer SILAC 164
1. Overview 164
2. Setting Parameters 165
3. Understanding the SILAC Results 167
3.1. Protein Filter 168
3.2. Summary View 169
3.2.1. Statistics of Filtered Results 169
3.2.2. Protein Profile Heatmap 169
3.2.3. Protein Profile Heatmap 169
3.2.4. Low Quality Data Warnings 169
3.3. Protein View 170
3.3.1. Protein Table 170
3.3.2. Protein Volcano Plot 171
3.3.3. Protein Coverage 171
3.3.4. Support Peptides 172
3.3.5. Protein View Filter 172
3.4. Peptide View 173
3.4.1. Peptide Table 173
3.4.2. Peptide View Filters 174
3.4.3. Feature Details 174
3.5. Result Exporting 176
13. DeepNovo Peptidome Workflow 178
1. Overview 178
2. DDA Peptidome Workflow and Parameter Settings 178
2.1. Data Refinement Parameters 180
2.2. Precursor and Fragment Error Tolerance 181
2.3. Enzyme Settings 181
2.4. PTM Settings 181
2.5. Peptide Search Parameters 181
2.6. Report Filter 181
3. DIA Peptidome Workflow and Parameter Settings 178
3.1. Precursor and Fragment Error Tolerance 184
3.2. Enzyme Settings 185
3.3. PTM Settings 185
3.4. Peptide Search Parameters 185
3.4.1. Target Library 185
3.4.2. Target Database 185
3.5. Report Filter 185
4. Understanding Peptidome Results 186
4.1. Result Filtration 186
4.1.1. Show Results for All Samples 186
4.1.2. Show Results for Specific Sample 186

4.2. Result Exporting 187
4.3. DeepNovo Summary View 187
4.3.1. Result Statistics 187
4.4. DeepNovo Sequence Table 189
4.4.1. Peptide View Filters 189
4.4.2. DeepNovo Table 190
4.4.3. Annotated Spectrum 191
4.5. Peptide Search Summary View 192
4.5.1. Result Statistics 194
4.5.2. Summary View Figures 194
4.5.3. Low Quality Data Warnings 195
4.6. Peptide View 195
4.6.1. Peptide View Filters 196
4.6.2. Peptide Table 198
4.6.3. Peptide Global Comparative View 199
4.6.4. Annotated Spectrum 200
14. Instrument Daemon 201
1. What is an Instrument Daemon? 201
2. How to add Data using an Instrument Daemon 201
2.1. Data Page 201
2.2. Data Copy 204
2.3. Instrument Daemon Samples 205

Chapter 1. Overview

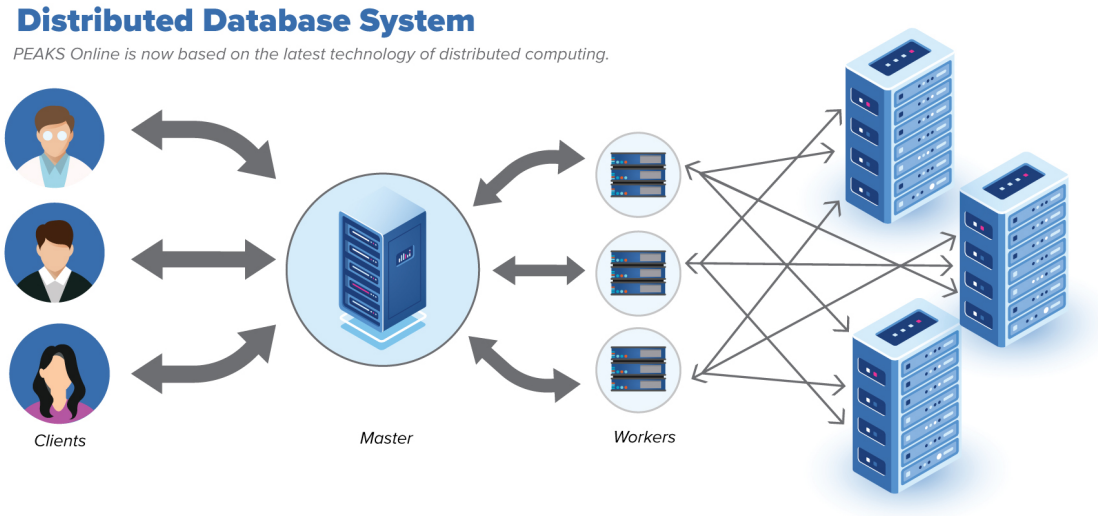
Welcome to PEAKS Online 11 .

1. How to Use This Manual

This chapter provides an overview of PEAKS Online's features and describes a typical data analysis work flow in PEAKS Online. It is strongly recommended to read this chapter to get a big picture of what PEAKS Online provides and how to use PEAKS Online.

2. What is PEAKS Online ?

PEAKS Online is a distributed computing platform for data dependent acquisition (DDA) and data independent acquisition (DIA) shotgun proteomic mass spectrometry data analysis. PEAKS Online users get detail-oriented results with the advantage of using a powerful, shared resource. This solution is designed to harness the hardware of a powerful server to analyze data using sophisticated algorithms on a high-volume of mass spectrometry data simultaneously, all while providing an easy-to-use and easy-to-access interface without the need to install a local client.



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

- **Complete Analysis**

  PEAKS Online provides a complete solution for shotgun proteomics data analysis. It is a vendor neutral computing platform 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) can also be performed. Intuitive result visualization tools are provided at every stage of analysis and 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 significance tests such as ANOVA or PEAKS Q.

- **Result Visualization**

  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. Sample comparative coverage

view visualizes the frequency of supporting peptides of a protein from selected samples. Modification and mutations can be visualized and filtered based on certain thresholds. Quantification results can be visualized using heat maps.

- **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 Online 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) of the precursor ion or maximum feature intensity. An Expectation-Maximization (EM) based algorithm is used for feature detection, deconvolution and refinement. PEAKS Online uses an optimization model for simultaneous feature matching and retention time alignment. This approach allows PEAKS Online to quantify peptide features independent of MS/MS identification.

- **Reporter Ion Quantification**

  Reporter ion quantification with chemical labels at the MS2 (MS3) level is one of the two quantification modes that are supported by the optional PEAKS Q module of PEAKS Online. 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.

- **SILAC (Stable Isotope Labeling by/with Amino acids in Cell culture)**

  In SILAC, 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 Online 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.

- **ID Trasnfer LFQ/SILAC (Stable Isotope Labeling by/with Amino acids in Cell culture)**

  Compared to the regular SILAC, ID transfer LFQ/SILAC can provide quantification work flow with high throughput and cov- erage. The unidentified precursors/pairs can be transferred from pre-run fractions in ID Transfer group.

- **Multiple Instrument Vendor Support**

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

- **Exporting Results**

  Analysis result can be exported to a variety of formats. The .csv format make it easy to view the results in spreadsheets and to analyze with in-house scripts or other third-party software. PEAKS Online also supports the ability to export data to standard result formats such as pepXML and mzIdentML.

3. What is New in PEAKS Online 11 ?

PEAKS Online 11 is a high throughput solution for proteomics allowing concurrent access from multiple users to support parallelism at project and data levels. It uses a distributed database to yield higher I/O performance and better fault tolerance. PEAKS online is ready to scale vertically and horizontally, add new worker nodes or database nodes to the system and see performance improvements right away.

PEAKS Online 11 offers the following improvements and added features:

- **Flexible User allocation.**

  PEAKS Online now allows flexible user allocation through deletion of previous users and username editing.

- **Increased Data Support.**

  We now support analysis of mgf files and Bruker .d folders containing .tsf files from timsTOF instruments operating in timsOFF mode.

- **Enhanced Password Security**.

  We now have optional enhanced password security features including required reset for temporary passwords, password history uniqueness, password expiry and password similarity strictness.

- **Quality Control.**

  PEAKS Online now has an automated quality control tool for DDA and DIA identification and label-free quantification workflows. Compared to the summary page from the database search of quantification results, the QC module in PEAKS Online provides more metrics to assess the quality of each LC-MS run.

- **DDA Deep Learning Boost.**

  New optional Deep learning boost for DDA data analysis can increase peptide identification over 10%.

- **Streamlined DIA Workflow.**

  PEAKS Online now offers a powerful and flexible solution for DIA analysis by allowing peptide identification using a spectral library search with or without a protein inference database direct database search or a combination of both for optimal peptide coverage. Then, unmatched spectra from the database search are analysed using *de novo* sequencing. By combining identification strategies we can provide increased sensitivity and and increased identification rate.

  Enhanced Deep learning-based technology for predicting spectra, retention time, and collision cross-section values for Direct database search to improve sensitivity and accuracy including DIA with short gradient optimization.

- **DeepNovo Peptidome.**

  The immunopeptidome is an attractive avenue for peptide-based vaccine and cancer immunotherapy development. As such, PEAKS Online now offers a specialized DeepNovo Peptidome workflow for both DDA and DIA data acquisition. De novo sequencing for these workflows is powered by a deep learning model trained specifically with large HLA peptide data set for retention time, fragmentation ion and ion mobility predictions. Furthermore, tailored second-round searches allow FDR to be estimated for *deep novo* peptides to increase accuracy of identification.

- **DeepNovo Standalone Workflow.**

  PEAKS Online now offers a standalone Deep Novo sequencing workflow

- **Spectral Library Generation improvements.**

In PEAKS Online 11, we have expanded spectral library generation to support more workflow types. Spectral libraries can now be generated from De Novo, DB, and Peptidome search results for both DDA and DIA acquisition modes.

- **PEAKS PTM improvements.**

  PEAKS provides comprehensive tools for PTM identification and PTM profiling. In PEAKS Online 11, we now offer two options for users to determine confident modification sites: minimal ion intensity and AScore. We also offer a new PTM profiling tool for direct visualization of quantitative information (eg. Abundance of modified and unmodified forms of the PTM sites identified).

  PEAKS PTM searches also contain a new "Modified Peptide" tab for direct comparison of modified and unmodified peptide identification parameters.

- **LFQ Normalization Improvements.**

  PEAKS Online 11 now uses a local RT normalization algorithm for TIC and internal standard protein normalization.

- **User-specific Algorithm Parameters.**

  PEAKS now allows individual algorithm parameters to be controlled on the user-level. These parameters include confident signature ion requirements for PEAKS PTM, DIA DeepNovo parameters and default *de novo* search engine.

4. Supported Data Types

- **mzXML** : Any instrument type converted to this format is accepted. MSConvert from ProteoWizard is suggested for conversion.
- **mzML** : Any instrument type converted to this format is accepted. MSConvert from ProteoWizard is suggested for conversion.
- **mgf**: Any instrument type converted to this format is accepted.
- **Thermo raw files** : Thermo .raw files are accepted.
- **Waters raw folders:** The master node must be on a Windows operating system. Otherwise, conversion to mzXML is necessary. If loading directly from your computer, compress the folder into a zip file before uploading.
- **Waters Mse:** Waters Mse files are now accepted. The master node must be on a Windows operating system. Otherwise, conversion to mzXML is necessary.
- **Bruker .d folders:** Folders containing .baf, .tdf and .tsf files from Impact, Maxis, or timsTOF instruments are accepted. If loading directly from your computer, compress the folder into a zip file before uploading.
- **Thermo DIA:** Thermo DIA files are now accepted.
- **Bruker DIA:** Bruker DIA files are now accepted.
- **WIFF DDA/DIA:** WIFF DDA/DIA files are now accepted.

5. Quick Walkthrough

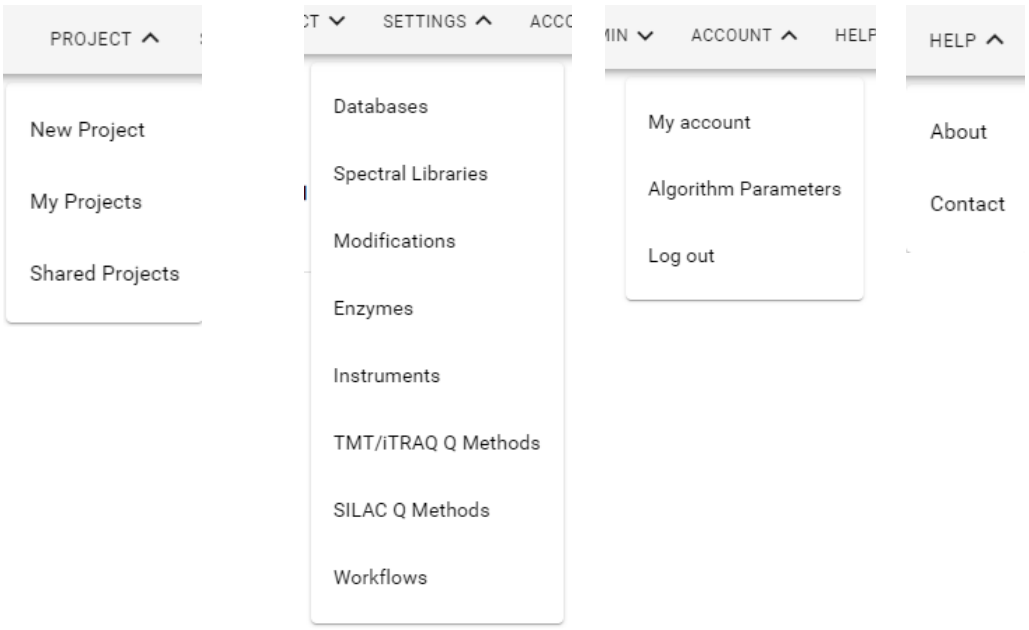
This section presents a quick walkthrough of a typical data analysis work flow. For step-by-step instructions and detailed explanation of the results, refer to Chapters 2-6.

5.1. PEAKS Online Main GUI

The main graphical user interface (GUI) of PEAKS Online can be accessed from a Web Browser. Upon entering PEAKS Online, you must sign in first. There are two types of users: Administrator or User. The user manual for Administrators is provided separately.

5.1.1. PEAKS Online User Options

PEAKS Online User accounts can only be created by the Administrator(s). After logging in, PEAKS Online users can access their completed projects or create a new project from the home page. For convenience, Import Project, Create New Project and Download an be accessed from the top left. Once logged in as a PEAKS Online user, the following items can be accessed from the top menu:



1. Project
   a. New Project: Users can create a new project by clicking here to upload data and set parameters.
   b. My Projects: Users can access all projects that were created and run on their user account.
   c. Archived Projects: Users can archive projects that were created by their account.
   d. Shared Projects: Users can access all projects that were shared to or by their account.
2. Settings
   a. Databases: Users can access all databases uploaded on their user account and databases configured by the administrator.
   b. Modifications: Users can view default modifications included in PEAKS Online and can also, access all custom modifications configured on their user account and modifications configured by the administrator.
   c. Enzymes: Users can view default enzymes included in PEAKS Online and can also, access all custom enzymes configured on their user account and enzymes configured by the administrator.
   d. Instruments: Users can view default instruments included in PEAKS Online and can also, access all custom instruments configured on their user account and instruments configured by the administrator.
   e. TMT/iTRAQ Q Methods: Users can view the default Label Q Methods included in PEAKS Online and can also, access all custom Label Q Methods configured on their user account and Label Q Methods configured by the administrator.

f. SILAC Q Methods: Users can view the default SILAC Q Methods included in PEAKS Online and can also, access all custom SILAC Q Methods configured on their user account and SILAC Q Methods configured by the administrator.

g. Workflows: Users can access all custom workflows made on their user account and workflows made by the administrator.

h. Libraries: Users can generate or import spectral libraries.Users can access all libraries configured on their user account and libraries configured by the administrator.

3. Account

   a. My Account: Users can manage their account name, password and associated email and set priority.

   b. Algorithm Parameters: Users can manage their individual algorithm parameters for confident signature ion requirements for PEAKS PTM, DIA DeepNovo parameters and default *de novo* search engine.

4. Help

   a. About: Users can view End-User PEAKS Online Software License Agreement.

   b. Contact: Links to Bioinformatics Solutions Inc. contact information.

5.2. Creating a PEAKS Online Project

To create a new PEAKS Online project from raw data files, perform the following steps:

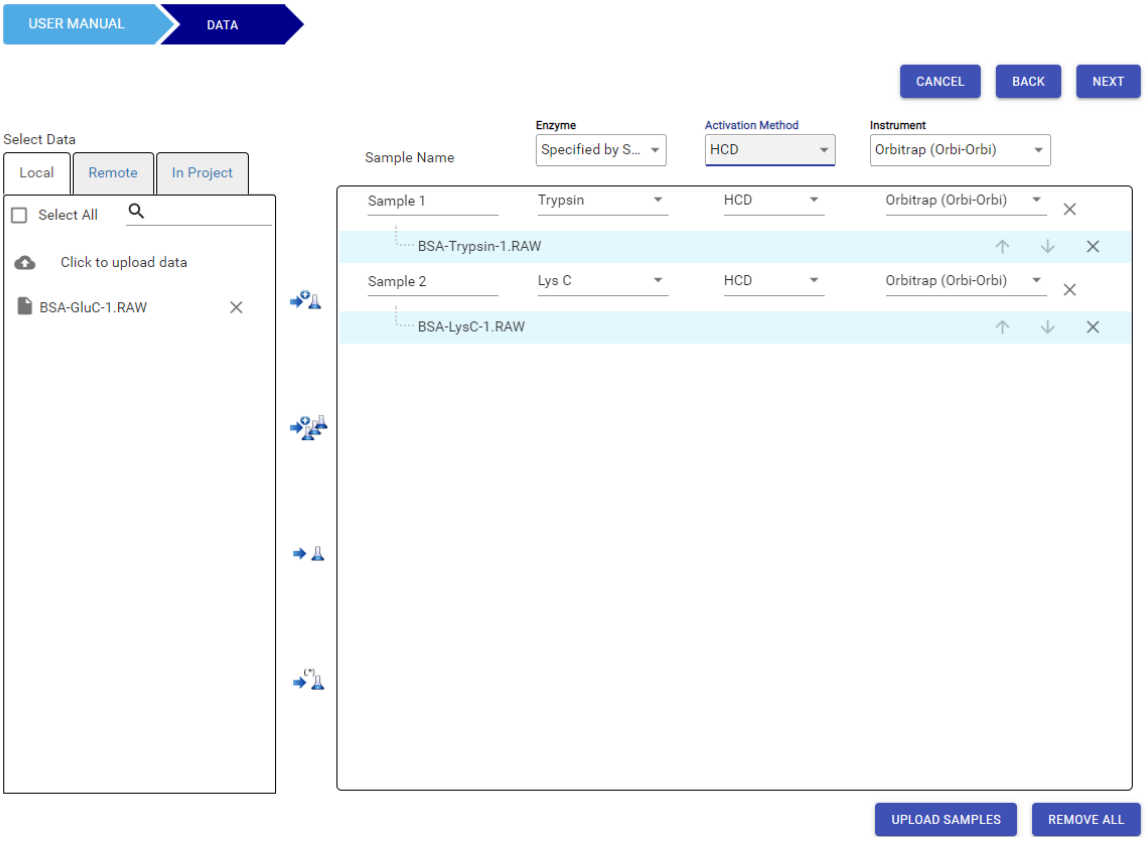
1. Select the icon from the top left or select "New Project" from the menu. This will prompt the “New Project” page where the project can be named and parameters set.

2. Name the new project and/or the description. Select either an existing work flow or create a project from scratch. Click Next button.

3. Select "Local" tab and click to bring up a browser window where files can be selected for uploading. If the administrator has specified folders on the server that can be accessed for data uploading they will be listed under the "Remote" tab. Click "In Project" tab to bring in data from the existing project if this is a new analysis for an existing project.

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

5. For each sample, specify the name and sample details, including "Enzyme" name, "Activation Method", and "Instrument" type, or specify the sample details on the top row if all samples have been prepared and acquired with the same method.



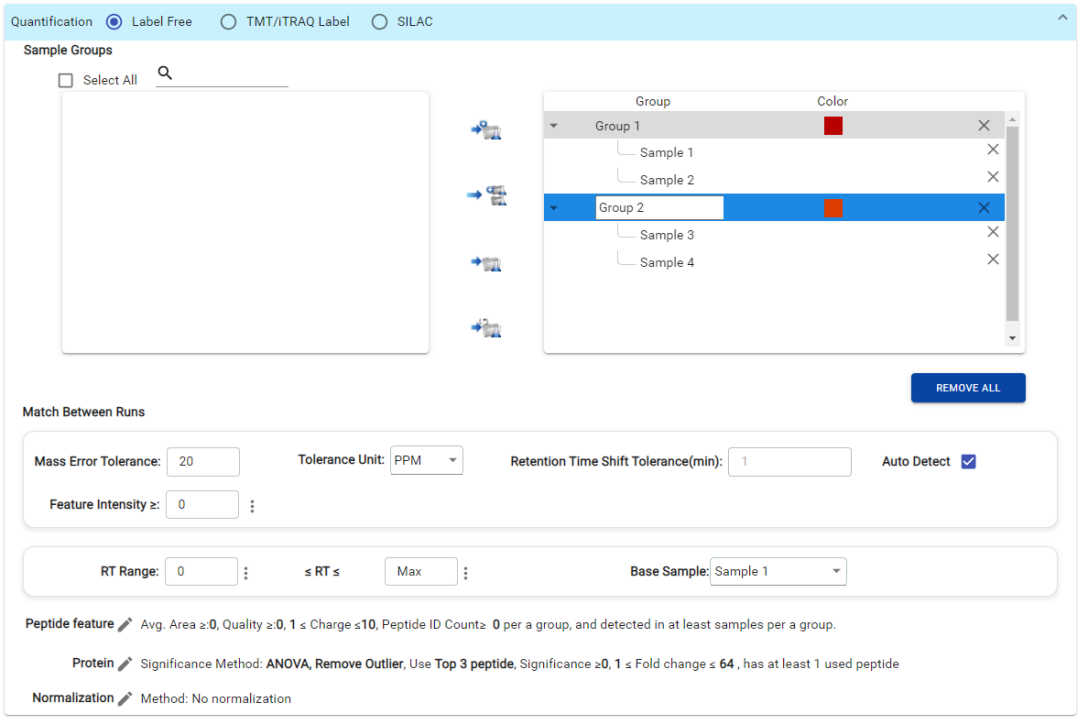
6. Click the Next button to select an Analysis for the project.

7. Populate the parameters for the analysis. If desired you can save this work flow for future projects.

8. Submit Project

5.3. Conducting an Analysis

To conduct a PEAKS Online Search, a Label-free quantification (LFQ) work flow for a complete identification plus quantification analysis is shown as an example. Refer to Chapter 8, *Label Free Quantification (LFQ)* for more details about conducting an LFQ analysis. Refer to Chapter 5, *Peptide De Novo Sequencing* and Chapter 6, *Peptide, PTM, and Mutation Identification (PEAKS DB, PEAKS PTM, and SPIDER)* for more details about conducting identification analyses.



The different options when setting up a full LFQ work flow are described below (see screenshot above):

1. Depending on the instrument used for data acquisition, set the parent and fragment error tolerance.

2. If the proteolytic enzyme was specified for each sample at the project creation step, users can select "specified by each sample". This makes it possible to use multiple enzymes in a single project and a single search.

3. Select a protein sequence database and the taxonomy for the database search. Select a contaminant database to be included in the database search. A decoy database will be created based on the target database by default. 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. Please see BSI's web tutorial (http://www.bioinfor.com/fdr-tutorial/) for more details.

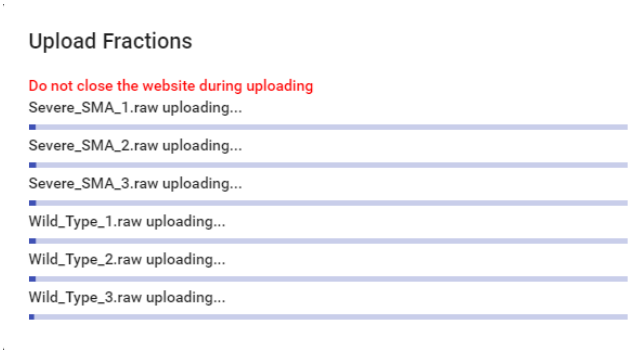
4. Specify the fixed PTMs and a few (no more than 5) common variable PTMs expected in the sample. To prevent long search times, select only the most frequent PTMs in the sample for PEAKS DB, and check the PEAKS PTM option to search for unspecified modifications.

5. *De novo* sequencing and PEAKS DB search will be performed using the same parameters listed above. See Chapter 5, *Peptide De Novo Sequencing* and Chapter 6, *Peptide, PTM, and Mutation Identification (PEAKS DB, PEAKS PTM, and SPIDER)* for more details on parameter setting for *de novo* sequencing and database searching.

6. Specify parameters and grouping for LFQ analysis. See Chapter 8, *Label Free Quantification (LFQ)* for more details.

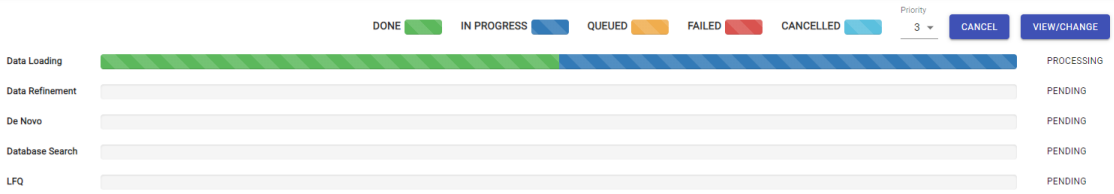
5.4. Project Summary, Monitoring and Results

When uploading data using a browser, after the submit button is clicked, the Upload Fractions page will appear. Be sure to keep the browser open as the data loads to server.



Once the data upload is complete, the progress page will appear, where the user can view the entire project analysis and monitor each individual step.

- User can change the priority for each task. The highest number means the highest priority. And the task with a high priority will finish before the task with a low priority.

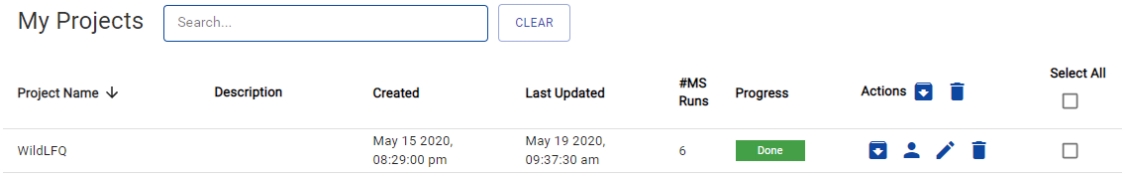


When the project is processed and tasks are completed, results will populate and can be viewed.

5.5. Opening an Existing Project

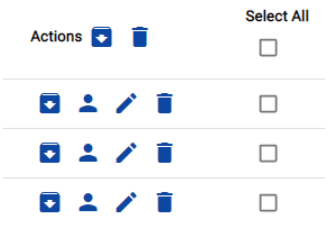
After launching PEAKS Online, existing projects can be accessed in "My Projects". Alternatively, projects that have been shared by the administrator or other users can be accessed from "Shared Projects".

From "My Projects" page, users can observe all their projects that have been completed and the projects currently being processed. Clicking on a project shows all analyses under this project. Click to archive the project. Click to share the project with others. means that the project is shared with other users. To edit the project name, click . To delete a project, click .



5.6. Batch Operation in Project listing

In the project list page, batch operations can be conducted. Multiple projects or all projects can be selected using the check box.



- User can achieve, share, edit or delete the selected projects suing the action buttons

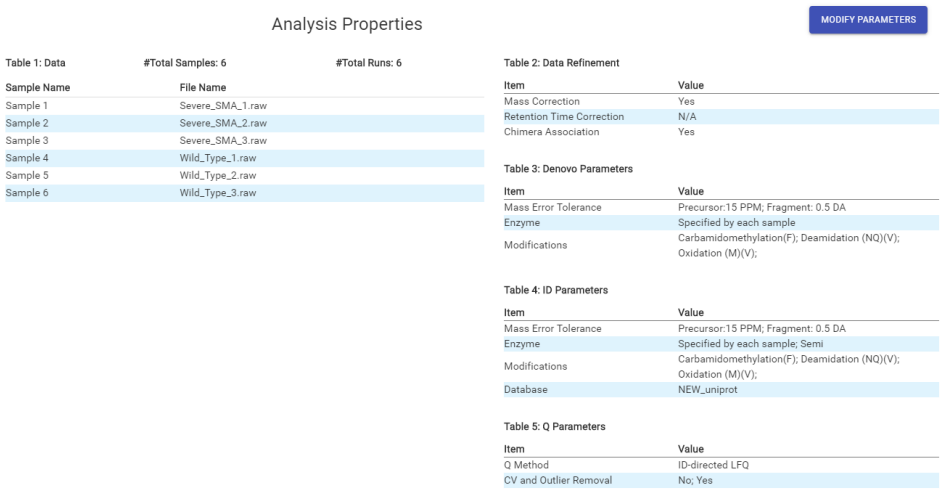
When the projects are processed and tasks are completed, results will populate and can be viewed.

5.7. Analysis Properties

Open any project and click one analysis, the "Analysis Properties" will show up and display the parameters used in the analysis.

The following tables will show up according to the different types of analysis:

- **Table 1:** Data samples used in the analysis
- **Table 2:** Data Refinement
- **Table 3** : DeNovo Parameters
- **Table 4** : ID Parameters
- **Table 5** : Q Parameters



Chapter 2. Management Configuration

From the top menu, within the Settings drop-down menu, PEAKS Online users can access options to configure databases, modifications, enzymes, and instruments for use in their PEAKS analyses. Users can only edit and delete the ones they create.

1. Database Management and Configuration

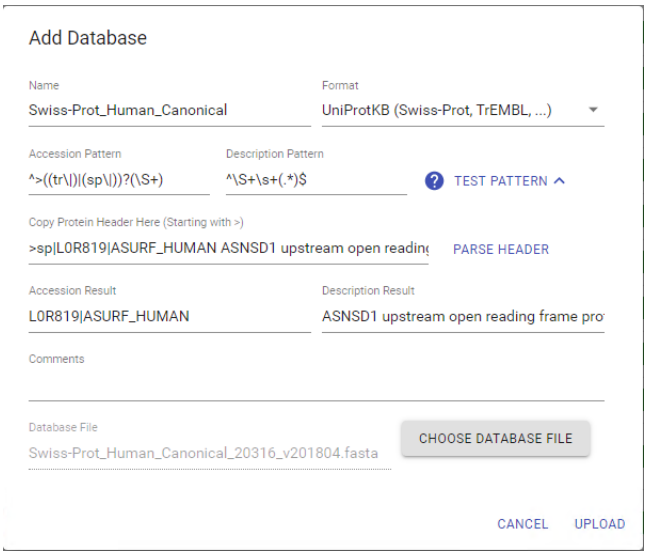
To use the PEAKS DB, PEAKS PTM and SPIDER functionalities to identify proteins, PEAKS Online must have access to a protein database in FASTA format. To add, edit or delete a database, users can access the "Manage Database" page by selecting "Database" from the Settings drop-down menu. Users can also add a database from within an analysis workflow by selecting the "Create New Database" option from the Target Database drop-down menu on the analysis parameters page.

- Add a database

  Open the "Add database" dialog by clicking the + on the "Manage Databases" page, or from the "Create New Database" option within a workflow . Give the new database a name, choose the format to match where the file came from (e.g. UniProtKB, Ensembl, etc.) , and then click on the "Choose Database File button" to specify the file path of the database's FASTA file. Finally, click Upload. The database will then be uploaded and added to the list of configured databases.

  ### Note

  Once uploaded, the FASTA file can't be changed. The database must be deleted and a new one created with the updated FASTA file



- Edit a database

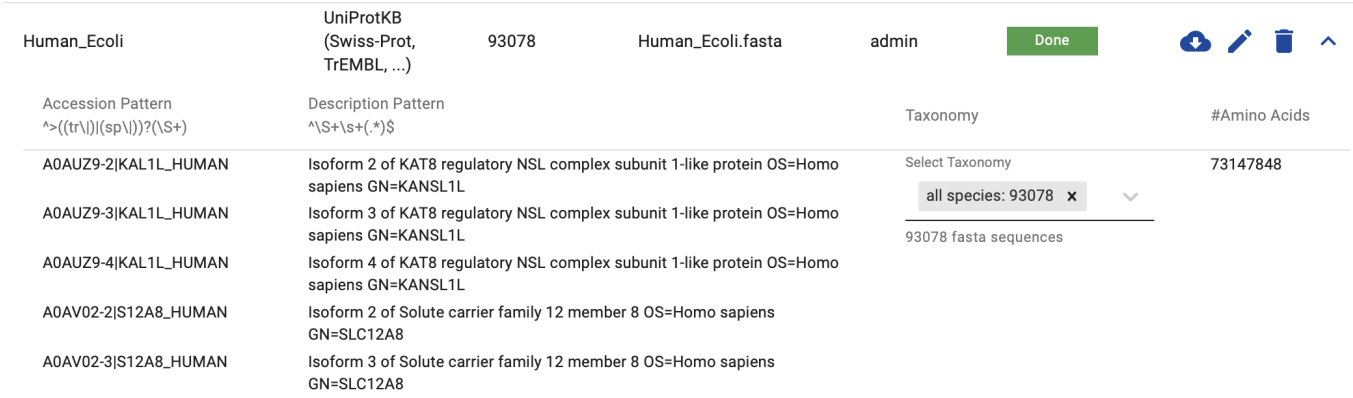
  From the "Manage Databases" page, click button in the Action column of a database to open the "Edit Database" dialog. Change the necessary parameters (e.g. name, format) and click Save.

- Delete a database

  From the "Manage Databases" page, click button in the Action column to delete a database permanently.

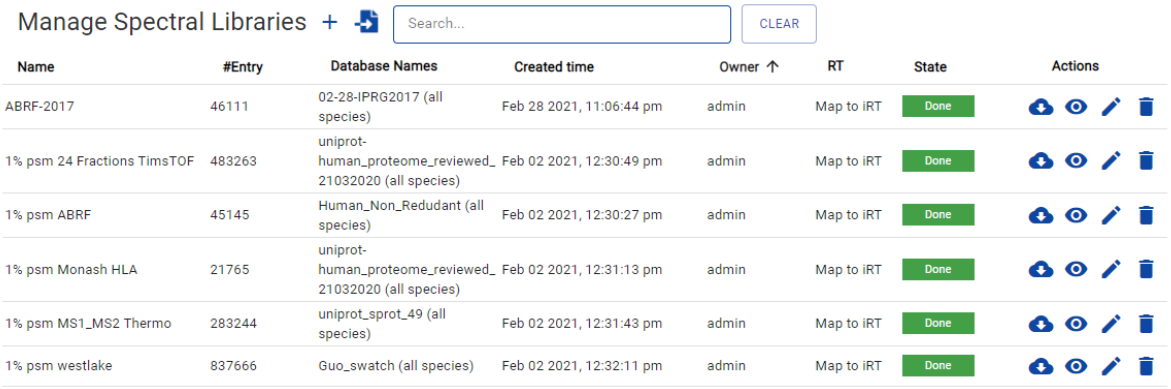
- View database Information

From the "Manage Databases" page, click ⌄ button in the Action column to view database information. By default all species are selected. But You can also select the species you want by clicking.



2. Library Management and Configuration

In order to complete a Spectral Library Search Workflow, users must first create or import a spectral library. Configuration of spectral libraries is accessed by selecting the "Spectral Libraries" option from the Settings drop-down menu, which will open the Manage Spectral Libraries page below. For more information on the use of Spectral Libraries please see the Chapter 7, DIA Streamlined IdentificationSection 2, “Adding a Spectral Library from the Parameters Page” .



From the Manage Spectral Libraries page, PEAKS Online Users can perform following actions:

- Create a new library

  Click ＋ next to "Manage Spectral Libraries" to open the "Create Libraries" dialog. Configure the Libraries accordingly and click Save. See Create New Library section below for more details on choosing settings.

- Download a Library

  Click ☁ button in the Action column of a customized library to download library.

- Edit a Library

  Click ✎ button in the Action column of a customized library to open the "Edit Libraries" dialog. Change the necessary parameters and click Save.

- Delete a Library

  Click 🗑 button in the Action column of a library to delete the library permanently.

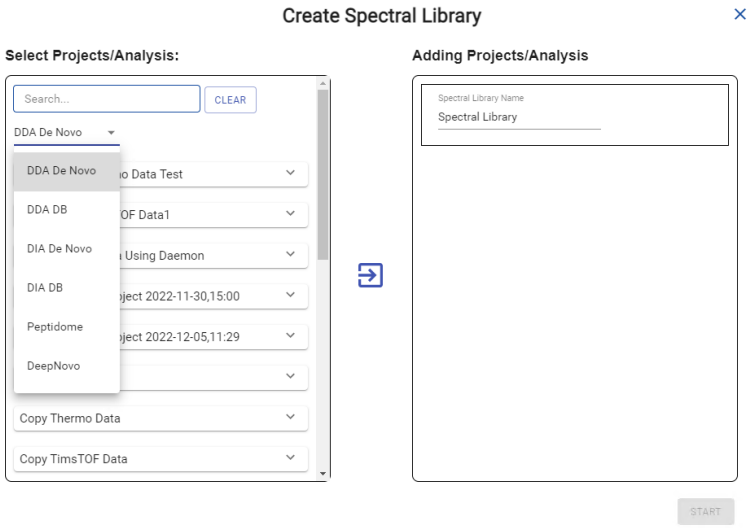
2.1. Create a New Library

Libraries can be created using pre-existing analysis results from a PEAKS De Novo DDA, PEAKS DB DDA, or a PEAKS DB DIA, PEAKS De Novo DIA, Peptidome and Deep Novo workflow.

In order to create a Spectral Library from an existing analysis follow these steps:

1. Select the "Spectral Libraries" option from the Settings drop-down menu.

2. Click the + button to open the "Create Library" dialog.



3. Select the workflow type from the workflow drop-down menu below the search bar. These workflow types include DDA De Novo, DDA DB, or DIA De Novo, DIA DB, Peptidome, and DeepNovo.

4. Find the analysis you wish to create a library from in the list below. You can use the search bar to further filter the list of analyses.

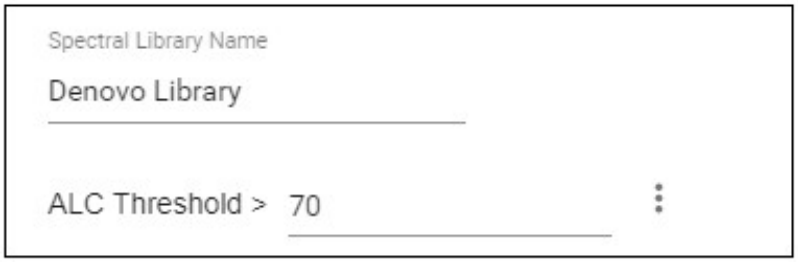
5. Click the button to move the analysis to the right side of the pane.

6. Enter a name for the new Spectral Library. This should be a descriptive name as it will be used to identify the library during project creation.

7. Set the parameters for library creation. These parameters will be different depending on the workflow being used to create the library.

   For DDA Denovo workflows the parameters are:

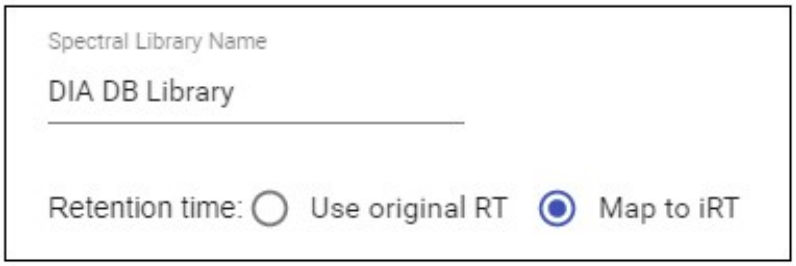
   - ALC Threshold: Only denovo candidates with an average local confidence score equal to or greater than this value will be used for library generation.



   For DDA DB Search the parameters are:

- Include DE NOVO Only: If this is checked, denovo only candidates will also be included in library generation.
- ALC Threshold: This parameter is only used if the "Include DE NOVO Only" checkbox is checked. Only denovo candidates with an average local confidence score equal to or greater than this value will be used for library generation.
- Retention time: The map to iRT function maps experimental retention times for peptides to an indexed RT (iRT) to account for chromatographic differences between runs (e.g. different gradient lengths).

- Retention time: The map to iRT function maps experimental retention times for peptides to an indexed RT (iRT) to account for chromatographic differences between runs (e.g. different gradient lengths).

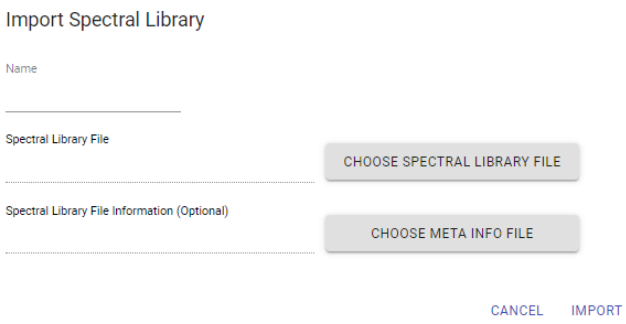


8. Click the "Start" button to begin library generation.

2.2. Import a New Library

From the Manage Spectral Libraries page, click the import button to open the "Import Library" dialog. The following information can be configured for a new custom library:

- **Name:** This name will appear in the library list for future use after it is saved.
- **Spectral Library File:** The spectral library file in TSV format.
- **Spectral Library File Information:** The meta information for the spectral library in the format of an INFO file (plain text information). This information is optional but can be added to describe the contents of the library for future reference.



3. Modification Management and Configuration

To view PTM details for default and customized modifications, select the "Modifications" option from the Settings drop-down menu.

From the Manage Modifications page, PEAKS Online Users can scroll through lists of default modifications using the "All", "Common", "Custom", etc., options at the top of the page. Users can also perform the following actions:

- Create a new modification

  Click + next to "Manage Modifications" to open the "Create Modification" dialog. Configure the PTM accordingly and click Save. See the Create New Modification section below for further details.

- Edit a modification

  Click button in the Action column of a customized modification to open the "Edit Modification" dialog. Change the necessary parameters and click Save. Default modifications cannot be edited.

- Delete a modification

  Click button in the Action column of a customized modification to delete the PTM permanently. Default modifications cannot be deleted.

- View a modification

  By clicking button in the Action column of any PTM, the parameters configured for that modification can be viewed.

3.1. Modifications

By default, PEAKS Online includes modifications in the Unimod database as built-in modifications. Unimod modifications are categorized into four lists:

- 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.

- Common

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

- Uncommon

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

- Glycosylation

The "Glycosylation" list contains commonly observed glycosylations.

Additional modification categories offered are "All" and 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 from this list.

- Customized

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

  Note

  Built-in modifications cannot be modified or deleted.

3.1.1. Modification Types

In PEAKS online, modifications can be classified as one of three types:

- PTM
- Isotopic
- Artifact

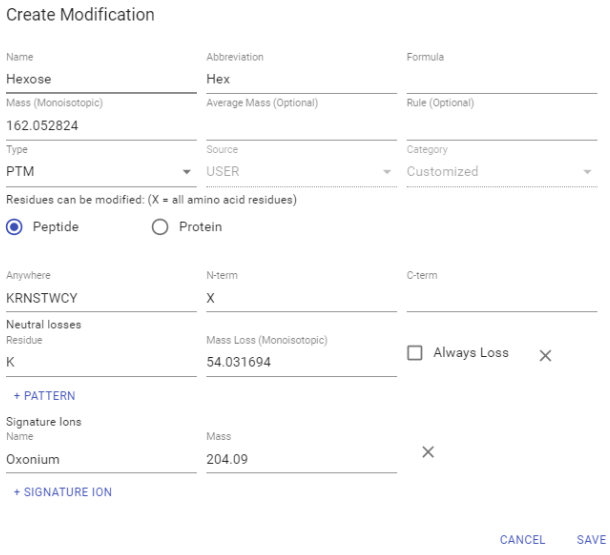
The new PTM categories Isotopic and Artifact describe types of artificial modifications such as those used for chemical labelling. For modifications which do not change the chemical properties and retention time of the unmodified peptide, Isotopic PTM should be chosen. Otherwise, the Artifact type should be chosen.

3.2. Create a New Modification

Click + button to open the "Create Modification" dialog.

Note: The other way is at the parameter page when submitting a new analysis, click "ADD/REMOVE MODIFICATION" then click "CREATE MODIFICATION".

The following information can be configured for a new custom, user-defined modification:



- **Name:** This name will appear in the modification list to specify the custom PTM for future use after it is saved.
- **Abbreviation:** The shortened, abbreviated form for the custom PTM (optional).
- **Formula:** The chemical formula of the modification (optional). This should correspond to the mass listed for the modification.
- **Mass (Monoisotopic):** The monoisotopic mass that a residue will gain or lose as a result of the modification.
- **Average mass** : The average mass that a residue will gain or lose as a result of the modification (optional).
- **Rule:** This field can be used to enter a comment about the PTM, to be used for your own reference (optional).
- **Type:** This field is used to set modification type. Type can be set to PTM, isotopic, artifact (see section 3.1.1 above for more information on modification types).
- **Residues can be modified** : This section is used to indicate which residues can be modified. Users must list the amino acids that can be modified anywhere, at the N-terminus, or at the C-terminus, and whether the modification occurs on the peptide or protein level. Residues are added using their single letter code without spaces or commas. See the example in the screenshot above.
- **Neutral Losses** : Users can define neutral loss patterns observed for the custom PTM (optional).
- **Signature Ions :** Users can define signature ion patterns observed for the custom PTM (optional). Through rescoring based on signature ions, these can increase the confidence of your PTM identification.

Note

Once a custom modification is saved, the name of that PTM cannot be changed.

Leucine (*L*) and Isoleucine (*I*) cannot be directly distinguished using traditional mass spectrometry techniques, since the two residues are isobaric, meaning they have exactly the same mass. For that reason, PEAKS Online uses *L* in *de novo* sequencing to represent either *L* or *I* . 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.

4. Enzyme Management and Configuration

To view and configure enzymes in PEAKS Online, select "Enzymes" from the Settings drop-down menu. This will bring you to the "Manage Enzymes" page where you will be able to view all built-in enzymes and create/modify/delete customized enzymes.

- Built-in enzymes

  All of the built-in enzymes included with PEAKS Online will be visible on this page. Built-in enzymes cannot be modified or deleted.

- Edit enzymes

  Click button in an enzyme Actions column to open the "Edit Enzyme" dialog.

- Delete enzymes

  Click button in an enzyme Actions column to delete an enzyme.

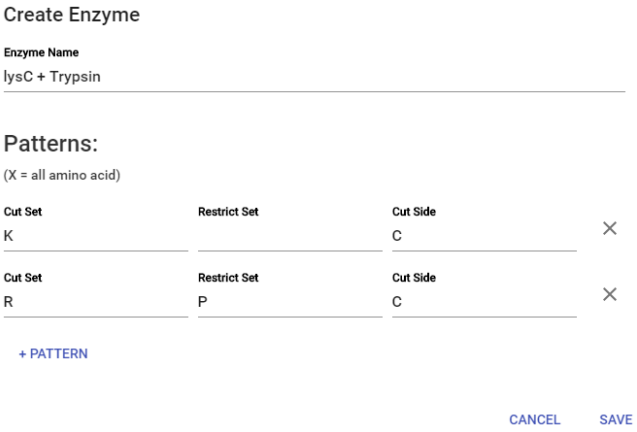
4.1. Create a New Enzyme

Click + button next to "Manage Enzymes" to open the "Create Enzyme" dialog.

Note: The other way is at the data loading page, click the drop-down window at "Enzyme", then click "Create a new Enzyme". This will open the same "Create Enzyme" dialog.

Provide the name of the new enzyme in the "Enzyme Name" field. Specify how the custom enzyme will cleave the protein between two amino acids to create peptides in the "Patterns" section. The letter, X, denotes any amino acid at this position. Add multiple amino acids to indicate that cleavage happens before or after any of the stated amino acids. For example, RK means after R or K, not R and K.

Choose where the cleavage sites are by specifying amino acids in the "cut set" and restricting the locations with the "restrict set". Set the side of the cut with the "cut side", which must be C or N. For example, "Cut Set: RK, Restrict Set: P, Cut Side: C" would denote a cut after R or K but not before P.



5. Instrument Management and Configuration

Instruments can be configured in PEAKS Online by selecting "Instruments" from the Settings drop-down menu. This will bring you to the "Manage Instruments" page where you will be able to view all built-in instruments and create/modify/delete customized instruments.

- Built-in instruments

  All of the built-in instruments included with Peaks Online will be visible on this page. You can view the instrument details by hovering over icon in the instrument Actions column. Built-in instruments cannot be modified or deleted.

- Edit custom instruments

  Click button in an instrument Actions column to open the "Edit Instrument" dialog.

- Delete custom instruments

  Click button in a custom instrument Actions column to delete an instrument permanently.

5.1. Create a New Instrument

To create a new instrument, click + button next to "Manage Instruments" to open the "Create Instrument" dialog. The following information can be added for a new instrument:

- **Instrument Type** : This name will appear in the instrument list for future use after it is saved.
- **Manufacturer:** Details can be added for future reference.
- **Model:** Details can be added for future reference.

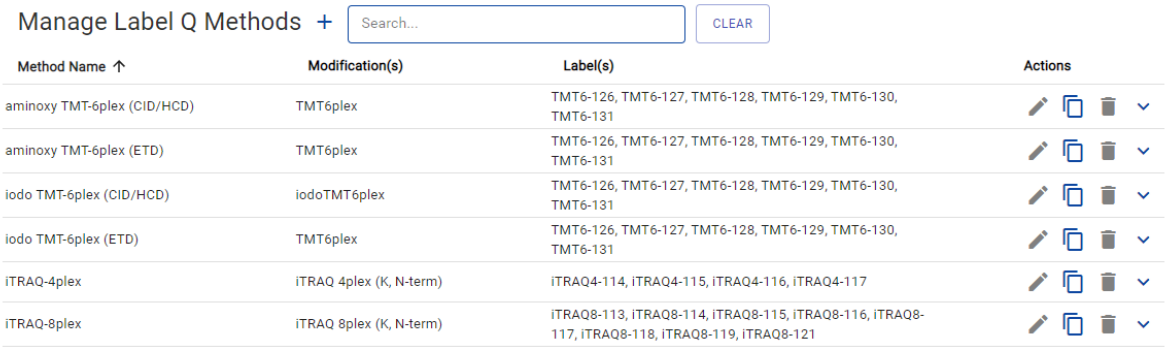
- **Ionization Source** : Select the ion source that was used: MALDI/SELDI or ESI (nano-spray). This will help the PEAKS Data Refine tool determine the charge of the ions.
- **Lock Mass** : set "-1" if instrument doesn't have lock mass. Enter the lock masses if used.
- **MS Precursor Scan Mass Analyzer** : PEAKS Online supports a single MS1 analyzer. Select the type and set the resolution and error tolerance. Check "Centroided" if applicable.
- **MS/MS Product Scan Mass Analyzer** : PEAKS Online supports multiple MS2 analyzers. Click "Add" to add as many as required. Fill in the information similar to MS1 analyzers.

6. TMT/iTRAQ and custom Label Q Methods

To run a labelled quantification analysis, the labelled quantification method must be setup beforehand. The Manage Label Q Methods window can be accessed by users and the administrator in the Settings drop-down menu.

PEAKS Online comes with a number of predefined labelled quantification methods for TMT and iTRAQ quantification. Custom label Q methods can also be defined by users.

Users are able to create their own label Q methods, which will only be available to them.



Built-in Label Q Methods. The built-in label Q methods within PEAKS Online come standard with all installations and can not be modified by users. They will be available to all users on the server.

To modify a built-in label Q method click the copy button in the label Q method's actions column. This will create a new label Q method (it will *not* modify the original method) with the same settings as the built-in method, which can then be modified.

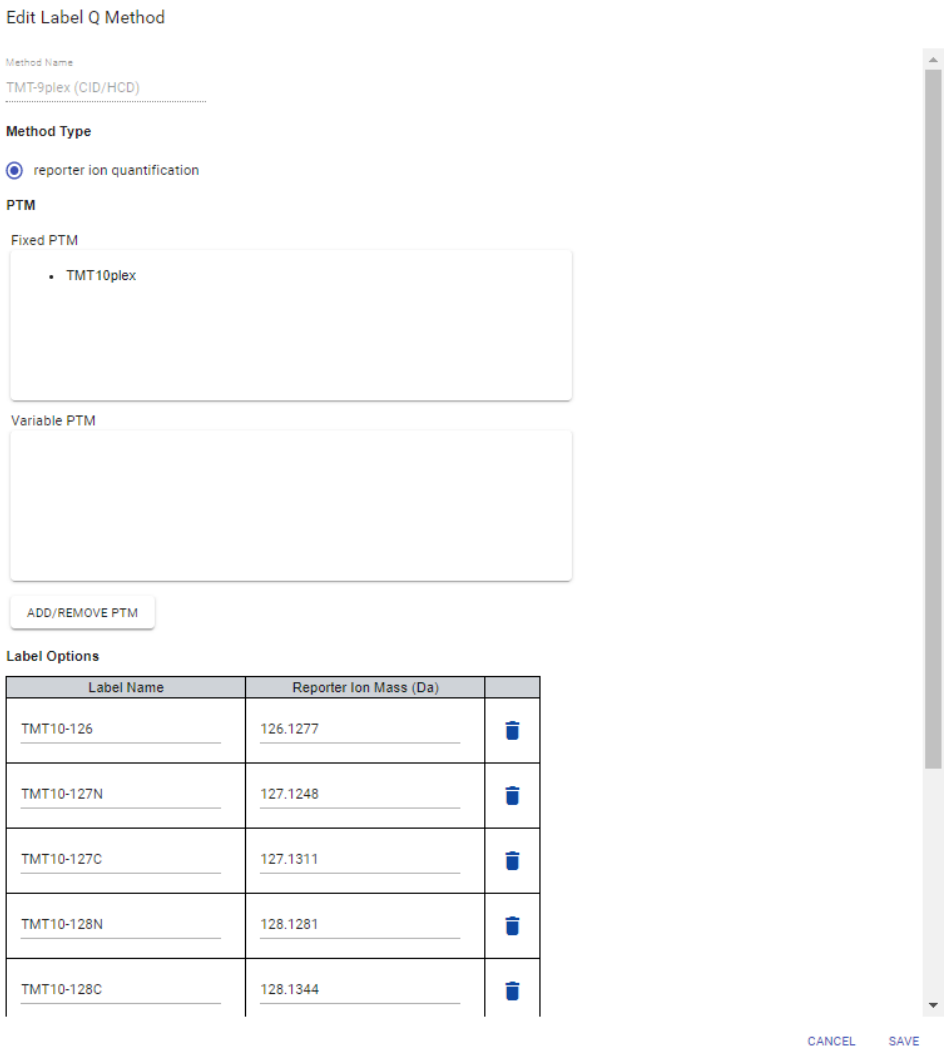
Note

Built-in label Q methods cannot be deleted or edited. They can only be copied.

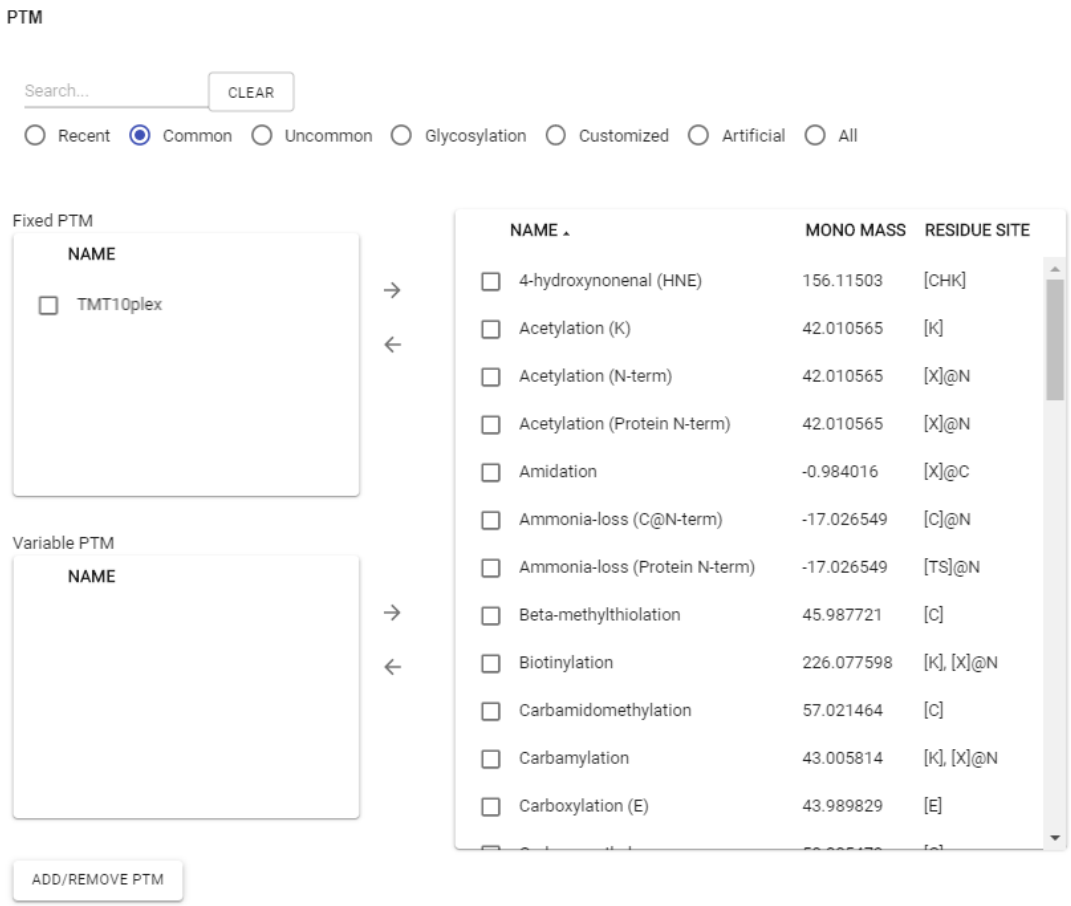
Create a new Label Q Method. Click the + button at the top of the Manage Label Q Methods page to create a new label Q method. The label Q method can then be configured in the "Create Label Q Method" dialog (see below). Click the save button to save the instrument for use on the server.

Note

This label Q method will be available to all users on the server.



- **Method Name:** This is the name that will appear in the labelled Q methods list for future use after it is saved. *This can not be modified after saving.*
- **Method Type:** At this time PEAKS Online only supports Reporter Ion Quantification (TMT / iTRAQ).
- **PTM:** The modifications used as labels for this label Q method. Click the Add/Remove PTM button to expand the PTM section (see below) and set the modifications for this label Q method.
  - **Fixed PTM:** PTMs to be used as fixed labels for this label Q method. Select PTMs from the right side, the search and headers can be used to locate specific PTMs, and then click the left arrow beside the Fixed PTM box to add them to the label Q method.
  - **Variable PTM:** PTMs to be used as variable labels for this label Q method. Select PTMs from the right side and then click the left arrow beside the Variable PTM box to add them to the label Q method.



- **Label Options:** The labels used by this label Q method. Each label must include a:
  - **Name:** (e.g.: iTRAQ-114). This will be the name associated with that channel in the labelled Q parameter page and on labelled Q result pages.
  - **Reporter Ion Mass:** (e.g.: 114.11 for iTRAQ-114). The mass for this channel's reporter ion.

Delete a custom Label Q Method. The user can delete the custom Label Q Methods, but not built-in label Q methods. To delete a label Q method find it in the list on the Manage Label Q Methods page and click the delete button in it's action column.

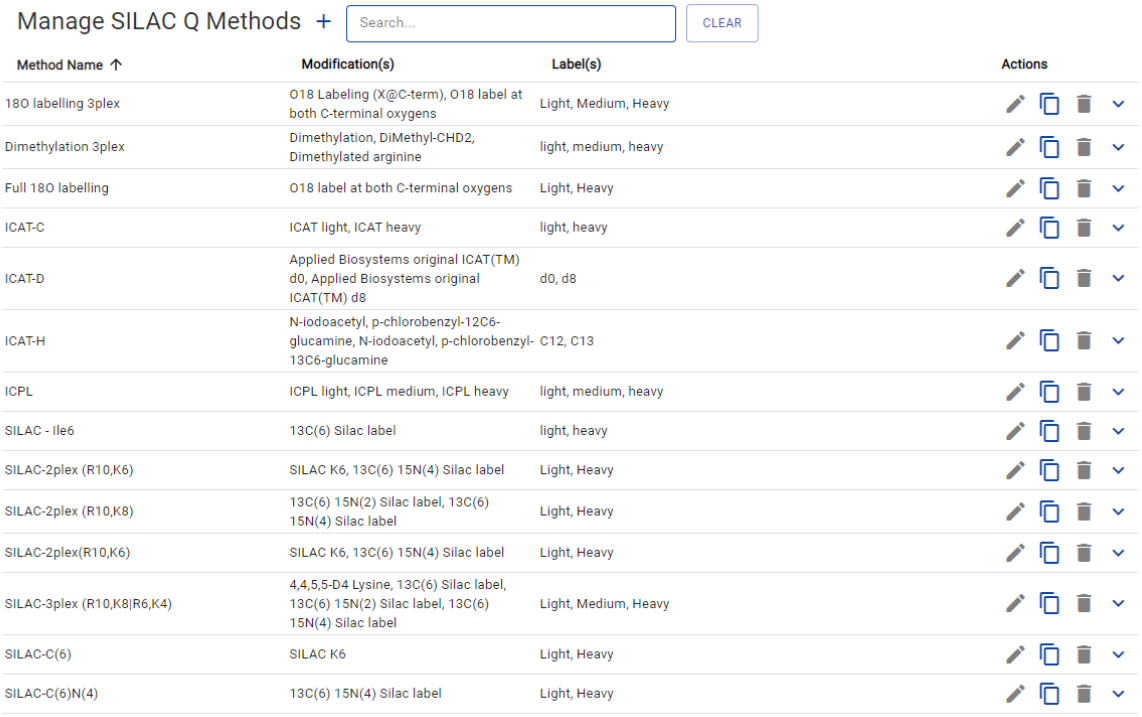
Edit a custom Label Q Method. To edit a label Q method find it in the list on the Manage Label Q Methods page and click the edit button in it's action column. All fields besides the method name are editable.

7. SILAC Q Methods

To run a SILAC quantification analysis, the SILAC quantification method must be setup beforehand. The Manage SILAC Q Methods window can be accessed by users in the Settings drop-down menu.

PEAKS Online comes with a number of predefined SILAC quantification methods for SILAC quantification. Custom SILAC Q methods can also be defined by users.

Users are able to create their own SILAC Q methods, which will only be available to them.



Method Name ↑	Modification(s)	Label(s)	Actions
18O labelling 3plex	O18 Labeling (X@C-term), O18 label at both C-terminal oxygens	Light, Medium, Heavy	
Dimethylation 3plex	Dimethylation, DiMethyl-CHD2, Dimethylated arginine	light, medium, heavy	
Full 18O labelling	O18 label at both C-terminal oxygens	Light, Heavy	
ICAT-C	ICAT light, ICAT heavy	light, heavy	
ICAT-D	Applied Biosystems original ICAT(TM) d0, Applied Biosystems original ICAT(TM) d8	d0, d8	
ICAT-H	N-iodoacetyl, p-chlorobenzyl-12C6-glucamine, N-iodoacetyl, p-chlorobenzyl-13C6-glucamine	C12, C13	
ICPL	ICPL light, ICPL medium, ICPL heavy	light, medium, heavy	
SILAC - Ile6	13C(6) Silac label	light, heavy	
SILAC-2plex (R10,K6)	SILAC K6, 13C(6) 15N(4) Silac label	Light, Heavy	
SILAC-2plex (R10,K8)	13C(6) 15N(2) Silac label, 13C(6) 15N(4) Silac label	Light, Heavy	
SILAC-2plex(R10,K6)	SILAC K6, 13C(6) 15N(4) Silac label	Light, Heavy	
SILAC-3plex (R10,K8|R6,K4)	4,4,5,5-D4 Lysine, 13C(6) Silac label, 13C(6) 15N(2) Silac label, 13C(6) 15N(4) Silac label	Light, Medium, Heavy	
SILAC-C(6)	SILAC K6	Light, Heavy	
SILAC-C(6)N(4)	13C(6) 15N(4) Silac label	Light, Heavy	

Built-in Silac Q Methods. The built-in SILAC Q methods within PEAKS Online come standard with all installations and can not be modified by users. They will be available to all users on the server.

To modify a built-in SILAC Q method click the copy button in the SILAC Q method's actions column. This will create a new silac Q method (it will *not* modify the original method) with the same settings as the built-in method, which can then be modified.

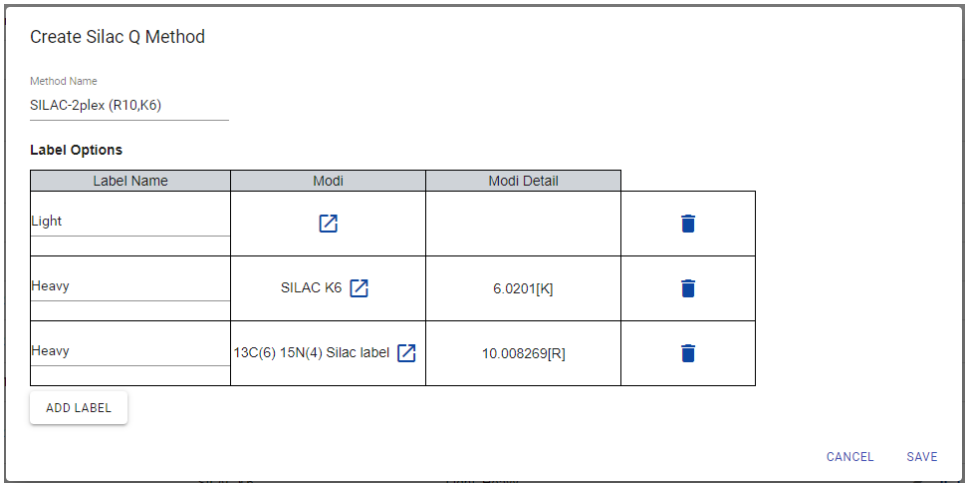
Note

Built-in SILAC Q methods cannot be deleted or edited. They can only be copied.

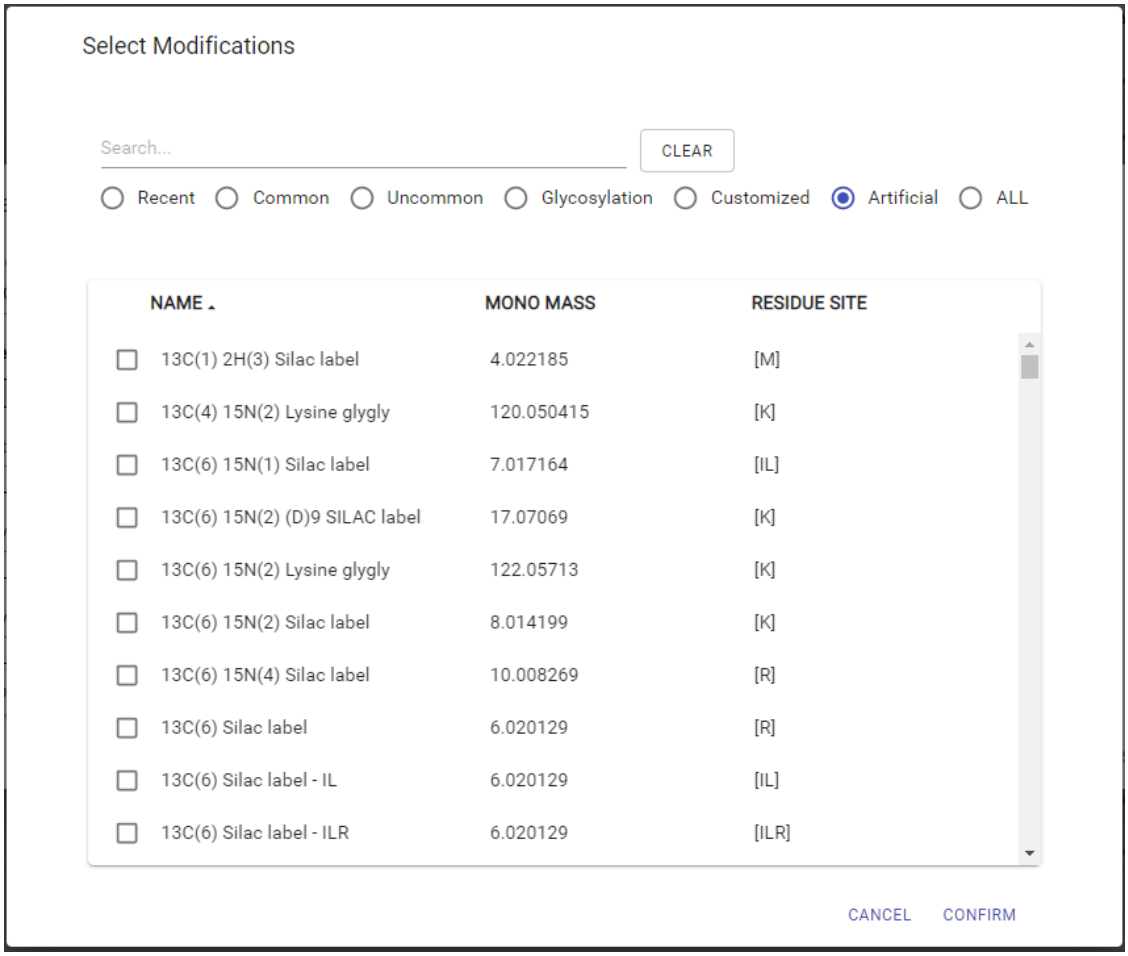
Create a new SILAC Q Method. Click the + button at the top of the Manage SILAC Q Methods page to create a new SILAC Q method. The SILAC Q method can then be configured in the "Create SILAC Q Method" dialog (see below). Click the save button to save the method for use on the server.

Note

This SILAC Q method will be available to all users on the server.



- **Method Name:** This is the name that will appear in the SILAC Q methods list for future use after it is saved. *This can not be modified after saving.*
- **Label Options:** The labels used by this label Q method. Each label must include a:
  - **Name:** (e.g.: Light). This will be the name associated with condition in the SILAC Q parameter page and on SILAC Q result pages.
  - **Modi:** The modifications used as a condition for this SILAC Q method. Click the button to open the Select Modifications dialog (see below) and set the modifications for this silac Q method.
  - **Modi Detail**: Shows the selected modification's monoisotopic mass and residue site.



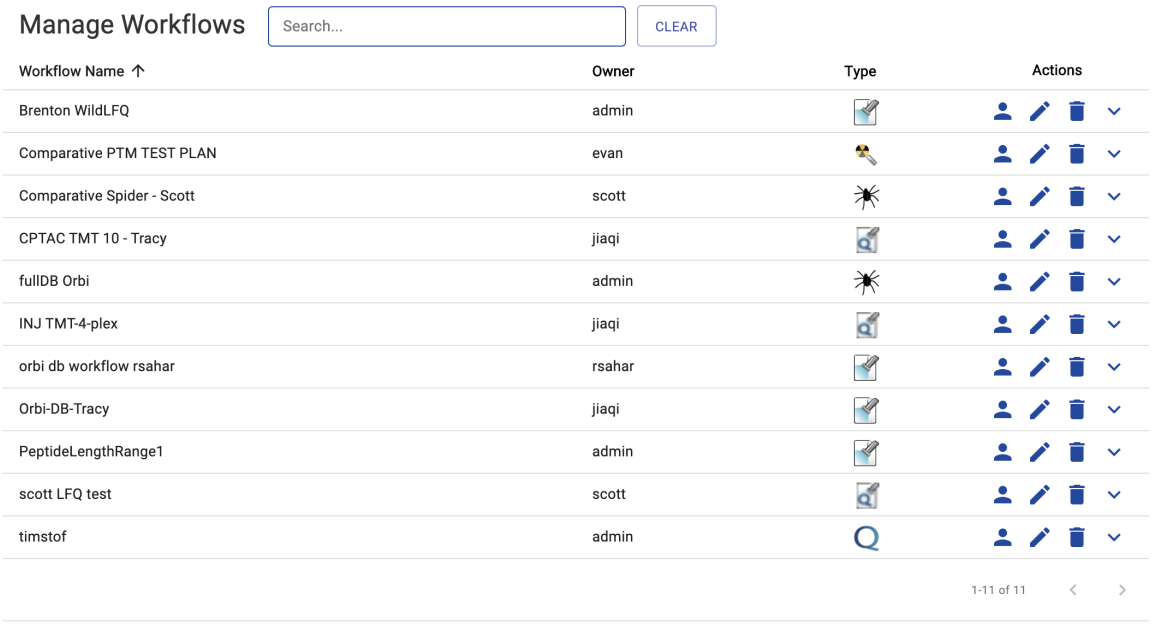
Delete a custom SILAC Q Method. The user can delete the custom SILAC Q Methods, but not built-in SILAC Q methods. To delete a SILAC Q method find it in the list on the Manage SILAC Q Methods page and click the delete button in it's action column.

Edit a custom Silac Q Method. To edit a SILAC Q method find it in the list on the Manage SILAC Q Methods page and click the edit button in it's action column. All fields besides the method name are editable.

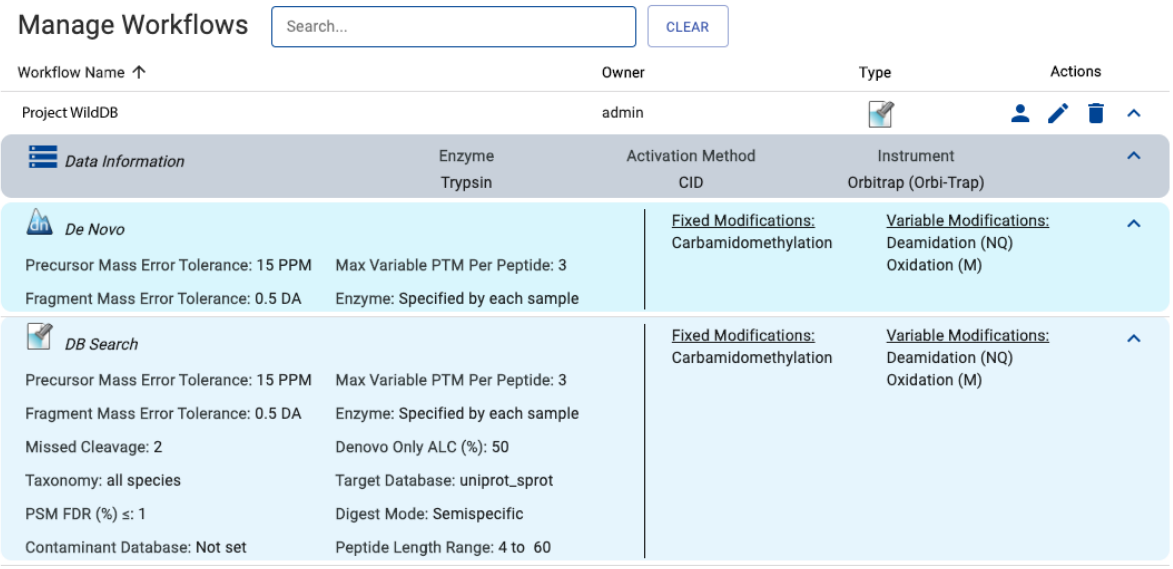
8. Workflows

To view or edit your current saved analysis work flows, you must go to the workflows tab. This section can be accessed by users and the administrator in the Settings drop-down menu.

Users are able to save workflows by selecting "Save Workflow As" on the parameters page when they are setting up an analysis within a project. These workflows will only be available to them and the Admin unless they share their workflow for other users.



To view the settings and parameters of a saved workflow click



To edit workflow settings click the button. This will take you to a dialog to change the parameters

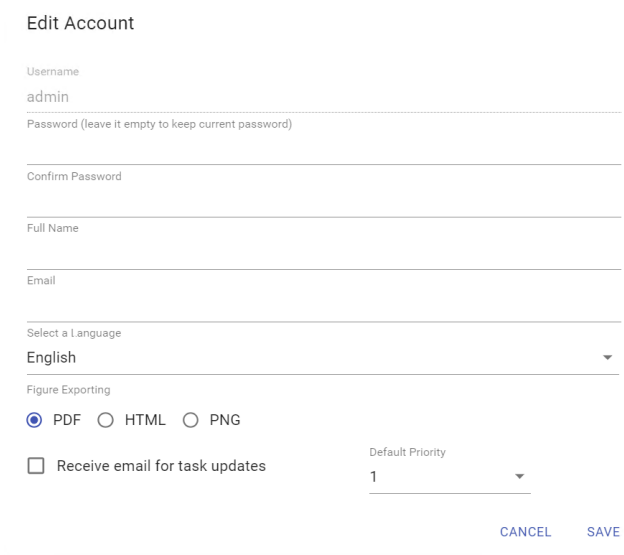
9. Account Settings

The account settings can be accessed from the top menu.

9.1. My account

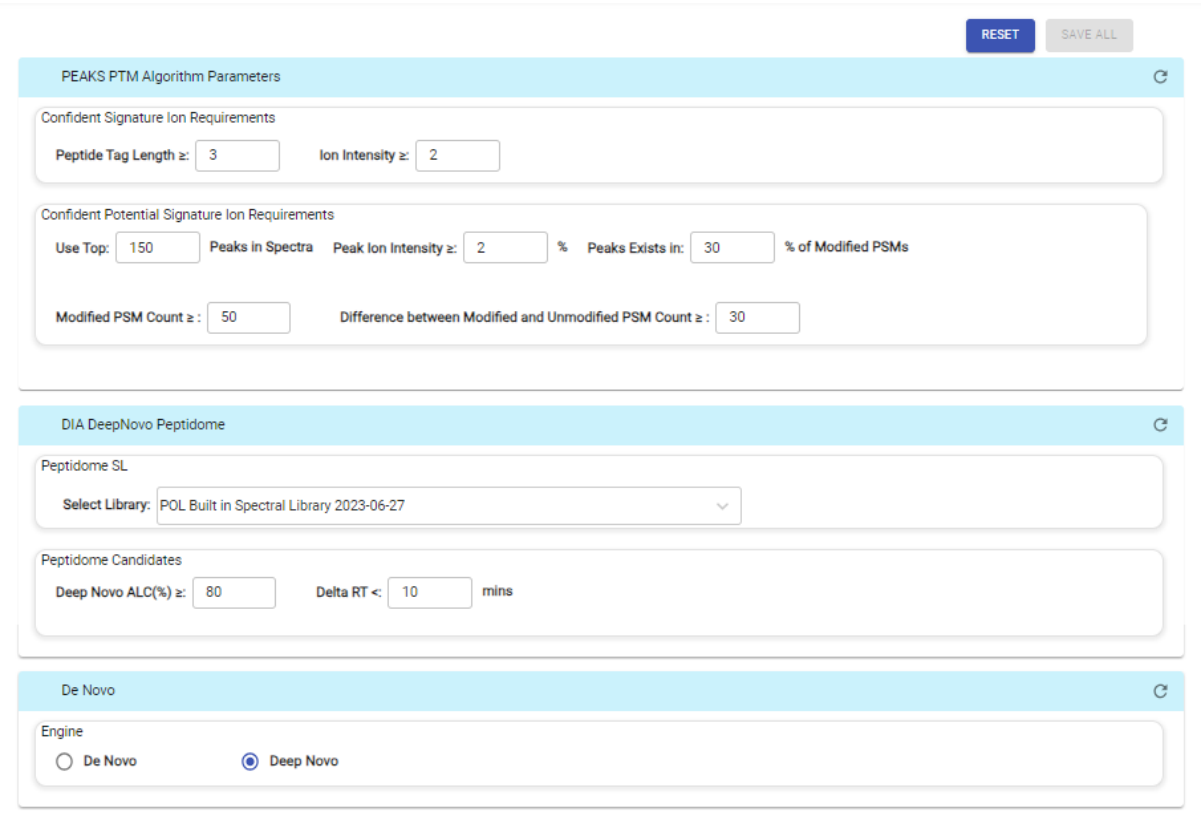
The user name and email address of the account currently logged in will be displayed. Click button to edit the account. This will bring up a page where the password, email address, and full name of the account can be changed.The user name and email address of the account currently logged in will be displayed.

- **Receive email for task updates:** Check this box, the user will receive the email when the analysis is finished.
- **Default Priority:** Setting the default priority of the task when the user create a new analysis.



9.2. Algorithm Parameters

User defined Algorithm Parameters are PEAKS PTM Algorithm Parameters and DIA DeepNovo Peptidome parameters.



9.2.1. PEAKS PTM Algorithm Parameters

Confident Signature Ions Requirements:

- Peptide Tag Length: peptide tag length of 3 or more amino acids is recommended for a confident PTM assignment. Peptide Tag Length $\geq$ X : The number of X consecutive ions at the modification site in the peptide that have ion intensity above the value Y set in the following algorithm parameter. If the PTM takes place at the N-term or C-term, the requirement is a number of consecutive ions + 1 (default would be 3 + 1)

- Ion Intensity: percent ion intensity for the fragment ions within the peptide tag compared to base peak for a confident PTM assignment. Ion Intensity $\geq$ Y: Defines the minimum ion intensity Y that is used in considering whether the PTM is confident

Confident Potential Signature Ion Requirements:

- Use Top # Peaks in Spectra: This top number of peaks (in intensity) in the spectra is considered when searching for potential signature ions. It is used in conjunction with the peak ion intensity parameter in determining how many peaks are considered when searching for potential signature ions.
- Peak Ion Intensity: The ion intensity for a peak must exceed this value for it to be considered as a potential signature ion. This is used in conjunction with the Use top # peaks in the spectra parameter to only select # peaks with ion intensity $\geq$ than a set value. If a signature ion does not meet these criteria, it will not display in the result
- Peak Exists in $\geq$ X % of modified PSMs: For the list of Modified PSMs, if the number of PSMs with this signature ion is greater than X % of the total, then the potential signature ion will be reported. Potential signature ions that do not pass this threshold will still be in the all potential signature ion export which is useful in determining how to tune this parameter for future runs.
- Modified PSM Count: minimal number of modified PSMs with the potential signature ion required to be reported in the result.
- Difference between Modified and Unmodified PSM count: percent difference between the numbers of modified and unmodified PSMs. Both the list of Modified PSMs with the modification of interest that has the detected signature ion as well as the list of Unmodified PSMs that has the detected signature ion are recorded. A % of the Modified PSMs will have the potential signature ion detected, and a % of Unmodified PSMs will have the potential signature ion detected. If the % of Modified PSMs is greater than X % compared to the % of Unmodified PSMs, then this signature ion is reported as a potential signature ion. Potential signature ions that do not pass this threshold will still be in the all potential signature ion export which is useful in determining how to tune this parameter for future runs.

9.2.2. DIA DeepNovo Peptidome

Peptidome SL:

- Select Library: Spectral library used for internal pre-search. PEAKS Online installations include a default Spectral Library generated by PEAKS Research team.

Peptidome Candidates:

- Deep Novo ALC(%) - minimum deep novo score cut off to be included in second round Peptide Search.
- Delta RT - minimum delta RT for the deep novo peptide to be included in the second round Peptide Search.

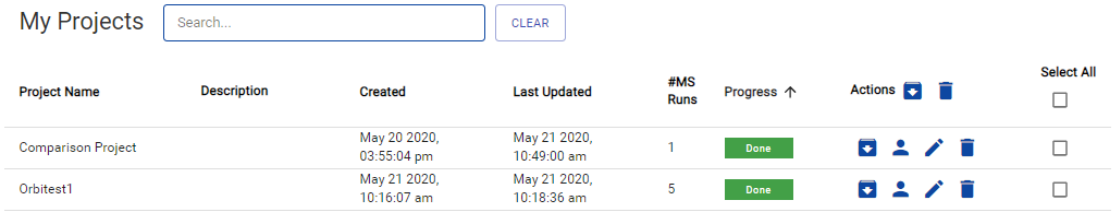
9.2.3. De Novo Engine

Each user has the option to set their default de novo search engine for the Deep Novo/De Novo standalone workflow. DeepNovo searching uses curated deep learning models. The speed of a DeepNovo search can be accelerated if a GPU is enabled with PEAKS Online.

Chapter 3. Archive Project

1. Overview

After the project is finished processing, users and admin can click button to archive the project or click button to delete the project to save the cassandra memory. When the user or admin need to view an archived project, the user or admin can click button on the top left of the menu to import the archived project from the Archived Repository back to my project list.



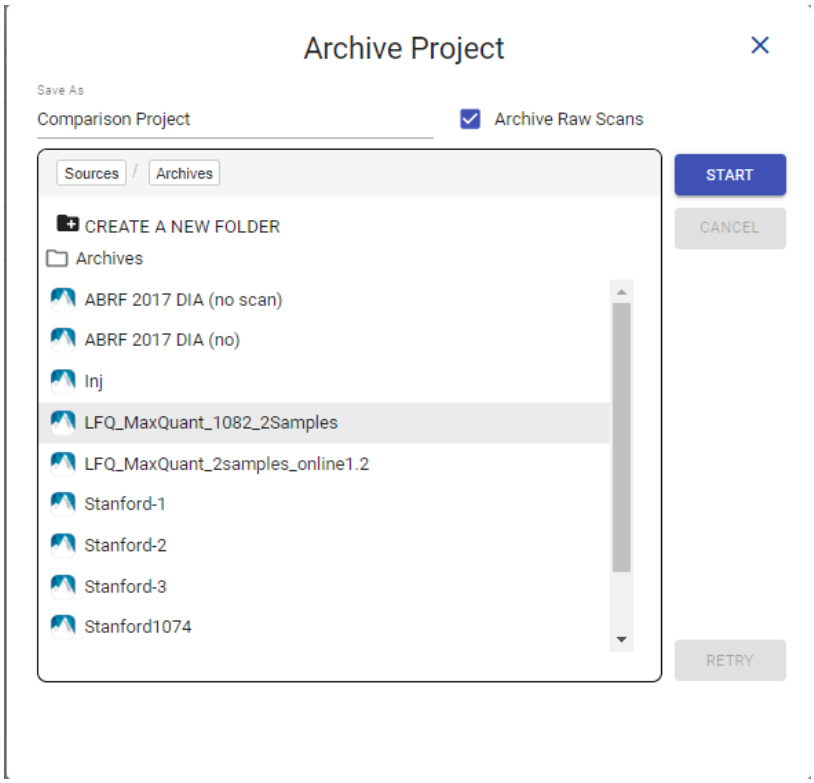
2. Archive Project

Click the archive button on the selected project to start archiving a project that has a Done progress status.

2.1. Create a new folder in Archive Repository

User and Admin can create a new folder under the archive repository to distinguish the archived project.

1. Go to the archive repository and click create a new folder Create a new folder button.
2. Fill in the folder name and press enter button on the keyboard.
3. The new folder will be created and you can click the folder to archive the project into this folder.



2.2. Archive Project in Archive Repository

1. Click the archive button on the selected project to start archiving the project that has a Done progress status.
2. The archive Project dialog pop up.
3. Choose the archive repository from the archive repository list.
4. User and Admin can create a new folder in the archive repository or use an existing folder.
5. User and Admin can edit the names of archived projects in the Save As name field.
6. Click the start START button to start archiving project.

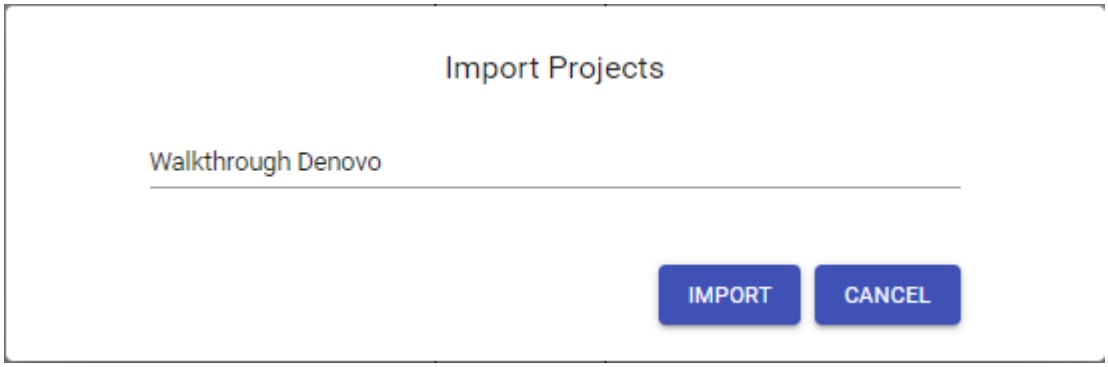
7. The archive progress bar will occur on the bottom. It displays the archiving details, which include the finished tasks and the total tasks.
8. User and Admin can click the button to hide archive project dialog and let the archiving run in the background.
9. There are three archive progress status: Archiving Archiving, Failed Archive Failed Archive and Done Done.
   - Archiving status: The project can not do any actions: open project, edit project, delete project. The user or Admin can click the archive button to view the archiving progress details.
   - Failed Archive status: The project can not do any actions: open project, edit project, delete project. The user or Admin needs click the archive button and see the following information: Failed Archiving... Please Retry. The user and admin can click cancel CANCEL button to cancel the archiving or click retry RETRY button to archive the project again. The retry action will start from the recently failed task.
   - Done status: The project can be viewed and edited. It also can be found in Archived project Page, if it done by the archiving.
10. The archive progress bar will occur on the bottom. It displays the archiving details which include the finished tasks and the total tasks.

3. Import Archived Project

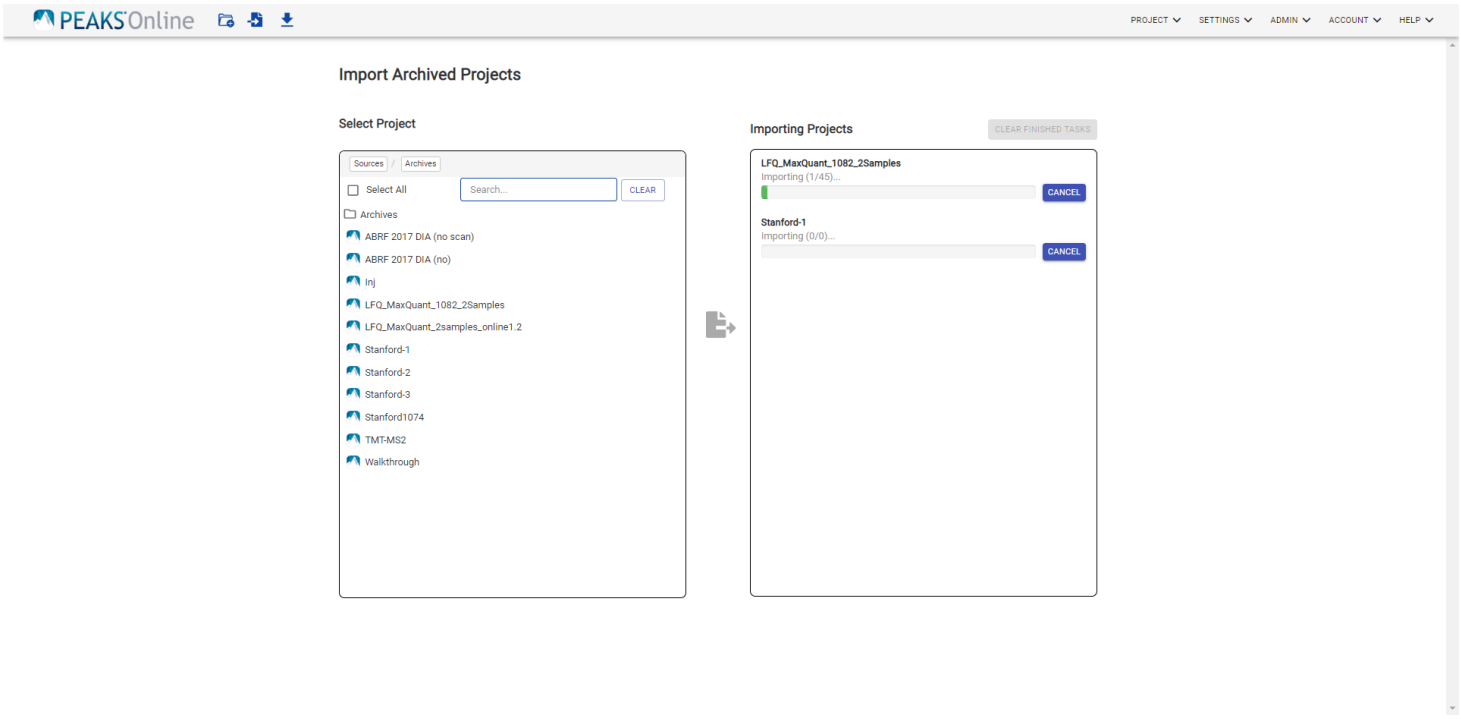
After archiving the projects in the Archive Repository, if the user or admin want to import the archived project back, they should delete the archived project in my project list. Otherwise they will get an error message that reads "This project is already in the project list".

3.1. Import Archived Projects Page

1. Using the button on the top left menu to go to Import Archived Project Page.
2. Select an Archived Project from the repository.
3. Click the import button in the middle of the import page.
4. The user or admin can edit the import project name if you needed.
5. Click import button to start import archived project or cancel button to cancel the importing.



6. The importing project will show on the right side with progress details, which include project name, finished tasks and total tasks.
7. The importing project has three status: importing, failed and finished.
   - The importing status: You can click cancel CANCEL button to cancel the importing project.
   - The failed status: User can click retry button to re-import the project.
   - The finished status: The importing project will show in my project list.
8. The user or admin can use clear all button to clear all finished importing project in the importing projects list. Or using clear button for each finished importing project to clear it in the importing projects lit.



9. User can click PeaksOnline logo on the top left of menu to leave the importing archived project page.

Chapter 4. Data View

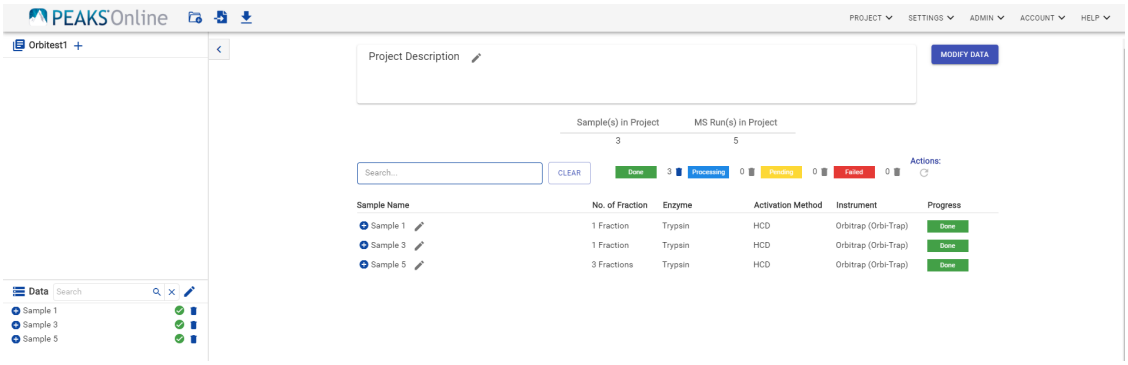
1. Overview

The data view gives a visualization of the mass spectrometry data loaded into the project. It allows for the inspection of LC-MS/MS details in an intuitive way for quality control and troubleshooting issues.

2. Accessing the Data View

The data view can be accessed by following the procedure below:

- Open an existing project/analysis. Refer to Section 5.5, “Opening an Existing Project” in Chapter 1, Overview for details.
- On the bottom left hand side of the project page, select "Data" to open the list of samples.
- At the top of the screen will be a summary listing "Sample(s) in Project", "Sample(s) in Analysis", "MS Run(s) in Project", and "MS Run(s) in Analysis".
- Click a sample row, and then click the ⊕ button to expand the details of the sample.
- After you click the ⊕ button, click the button to see data view.
- After you click the ⊕ button, click the button to download the MGF file.
- Type in a search word in the search box to list any samples with a matched sample name.
- Click the delete button beside the status icon to delete any samples in that status.
- Click the retry button below "Action" to have the failed samples start re-uploading.



3. Navigating the Data View

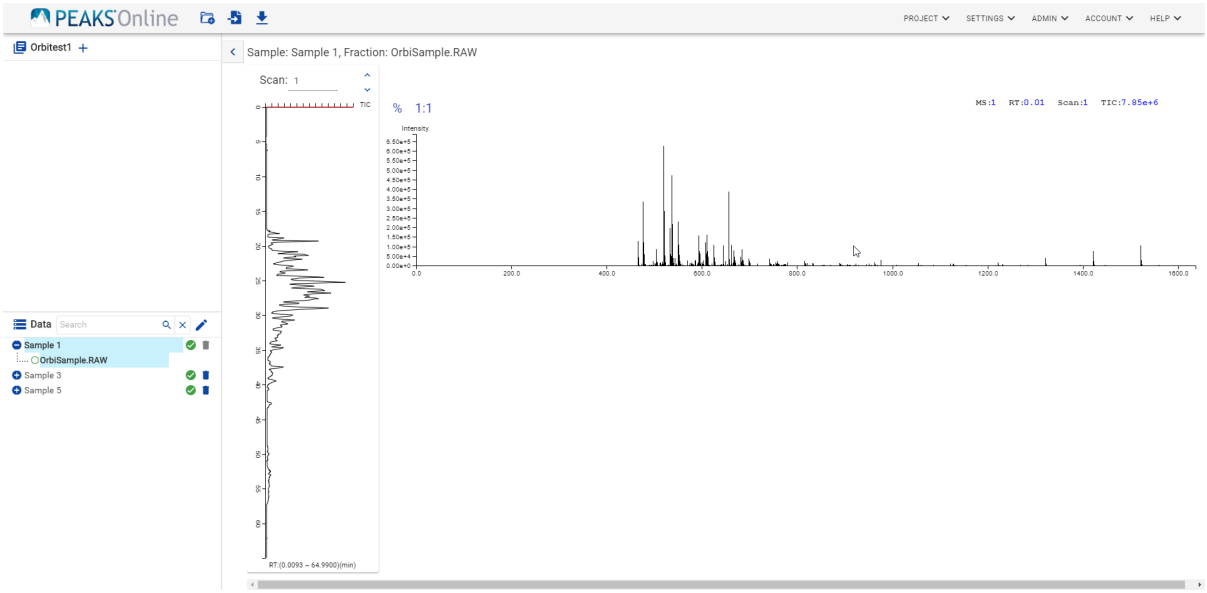
The data view is shown in three panes:

1. The TIC curve: shown vertically along the left side of the page. The total ion chromatogram (TIC) gives the total intensity of all signal acquired at each retention time. Use the Scan (or Frame Index with timsTOF) text box above the TIC display to search the data based on scan number. Also, it is possible to click on any point in the TIC curve to show details at that retention time. The arrow buttons beside the TIC graph can be used to scroll through the data based on retention time. The position of the red line indicates which retention time is currently being displayed.

2. The MS Scan: The MS scan at the given retention time is shown in the top right pane. The retention time (RT), scan number (Scan), and TIC intensity values are given in the top right corner of the MS scan display. The y-

axis of the spectrum can be switched between observed and relative intensity values using the # and % buttons located above the axis respectively. Hold the cursor within the spectrum and use your mouse wheel to zoom in and out on the m/z axis. Hold your cursor to the left of the y-axis to zoom in and out on the intensity axis. Use the 1:1 button to return to the default scale.

3. The MS/MS Scan: In the middle pane, m/z values of the precursors that were selected for MS2 and associated with the MS1 scan shown on top are displayed. Users can select the m/z and view its MS2 spectrum in the bottom right pane. The retention time (RT), scan number (Scan), TIC, mass over charge of the precursor ion (MZ, corrected by PEAKS mass correction), and charge (Z) are given in the top right corner. TimsTOF data will additionally show Ion Mobility and PrecursorID. The y-axis of the spectrum can be switched between observed and relative intensity values using the # and % buttons respectively. Hold the cursor within the spectrum and use your mouse wheel to zoom in and out on the m/z axis. Hold your cursor to the left of the y-axis to zoom in and out on the intensity axis. Use the 1:1 button to return to the default scale.



Chapter 5. 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 Online provides a reliable and comprehensive solution for automated peptide *de novo* sequencing. It features:

- **Accurate de novo sequencing**

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

- **Fast sequencing speed**

  PEAKS Online can perform *de novo* sequencing on over hundreds of spectra per second on servers.

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

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

- **Support of all major instrument types**

  PEAKS Online supports *de novo* sequencing of spectra generated by all instrument types (Ex. Orbitrap, FT-MS, Ion Trap, Time-of-Flight, Quadrupole, timsTOF) in common fragmentation modes (Ex. CID, HCD, ETD, ETHCD).

- **Automated result validation**

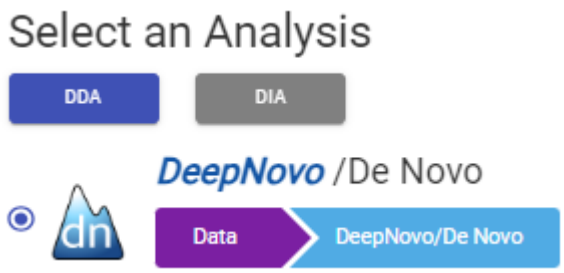
  PEAKS Online assigns a *local confidence* score to each amino acid in a *de novo* sequence. The local confidence score ranges from 0 to 99%, indicating how confident the algorithm considers a particular amino acid to be the correct assignment. Moreover, the peptide sequence is evaluated by an ALC (Average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

1.1. Standalone *DeepNovo/De Novo* Analysis

In the PEAKS 11 upgrade release, our standalone De Novo sequencing workflow has been modified to include an optional DeepNovo search engine. Deep Novo will only be available if the user has GPU enabled. If GPU is enabled, it will be used as default. Users can change their default De Novo engine under the Account > Algorithm Parameters menu. For more information, refer to Chapter 2, Section 9.



Other PEAKS Online tools, including PEAKS DB for database search; PEAKS PTM for unspecified PTM search; and SPIDER for homology search depend on the regular *de novo* sequencing result.

2. *DeepNovo/De Novo* Analysis Workflow and Parameter Settings

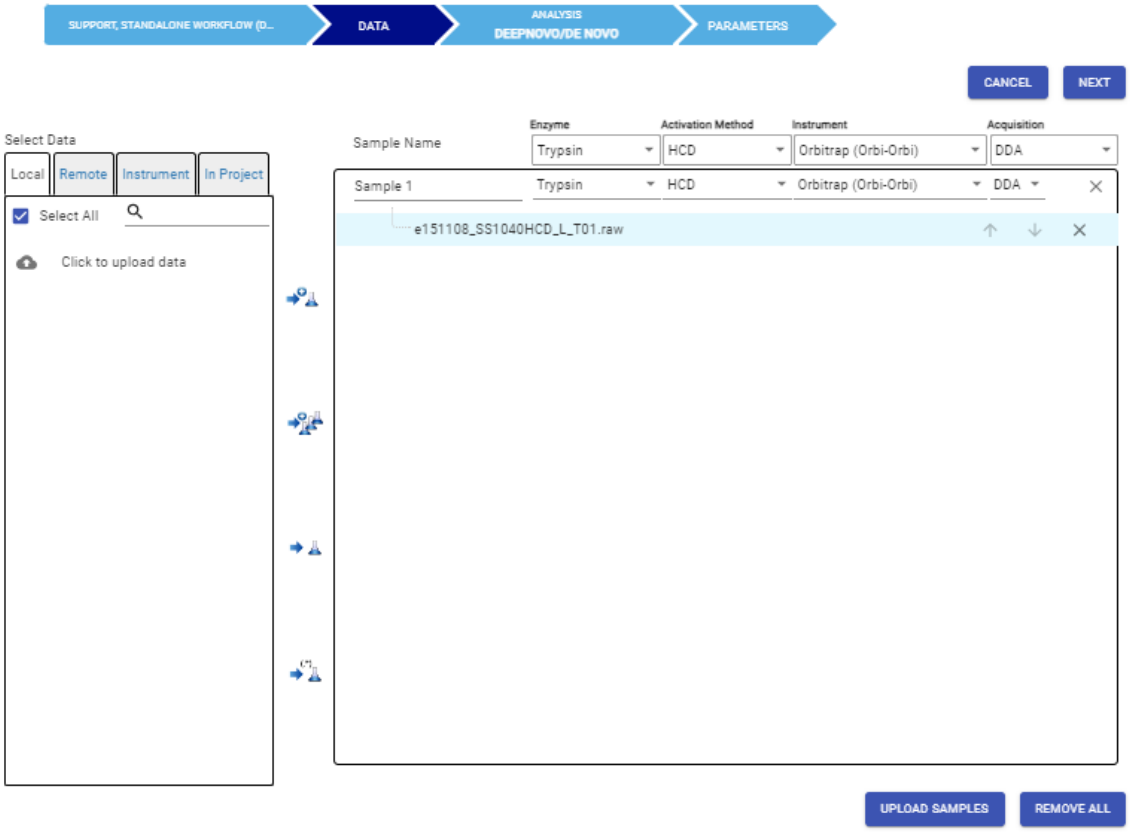
To conduct a PEAKS *de novo* sequencing analysis, follow the steps below:

1. Select "New Project" from the Project drop-down menu, or the Create New Project icon from the home page. This will prompt the "Project Creation" page where the project can be named, a description can be added, and the user can choose to create a new workflow or select a saved workflow.

2. Name the new project with optional description, select new or saved workflow, and click the Next button.

3. Select "Local" tab and click to bring up a File Explorer or Finder window where files can be selected for uploading. If the administrator has specified folders on the server that can be accessed for data uploading, they will be listed under the "Remote" tab. Click "In Project" tab to bring in data from the existing project if this is a new analysis of an existing project. To add data using the Instrument tab please refer to the Instrument Daemon chapter of the user manual.

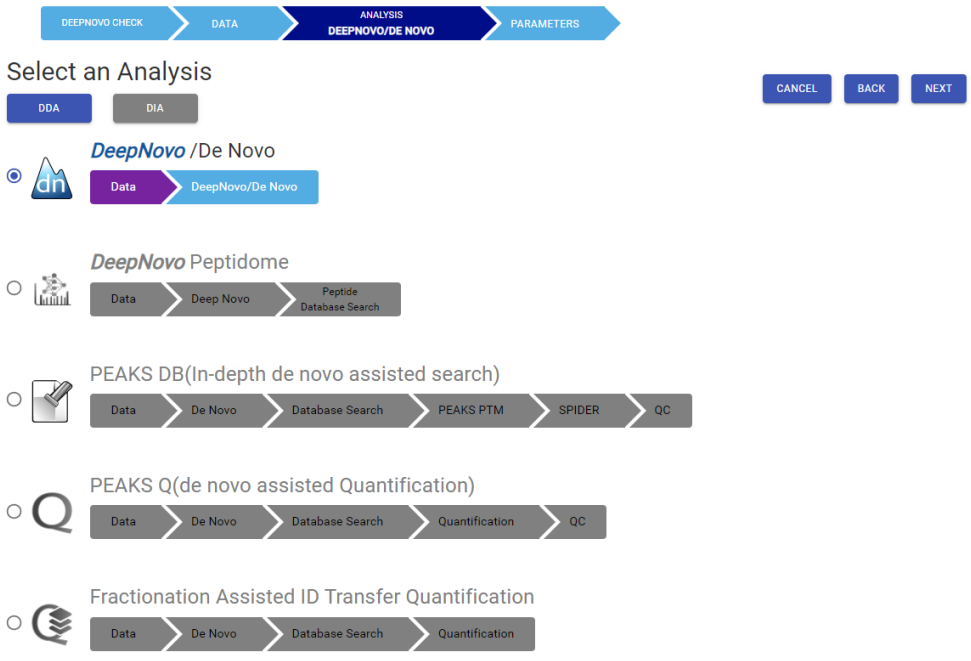
4. Newly added files will be selected and highlighted in grey. Use the icons to add the group of files as fractions or individual samples. Alternately, use the Select All box to de-select the group of files and then select individual files to add to the project using the different options. To place the selected data from the list into samples: use to place all files in an existing sample; use to create a new sample for selected files (i.e., add the files as fractions); put each file into an individual sample using ; or use to add samples by regular expression.



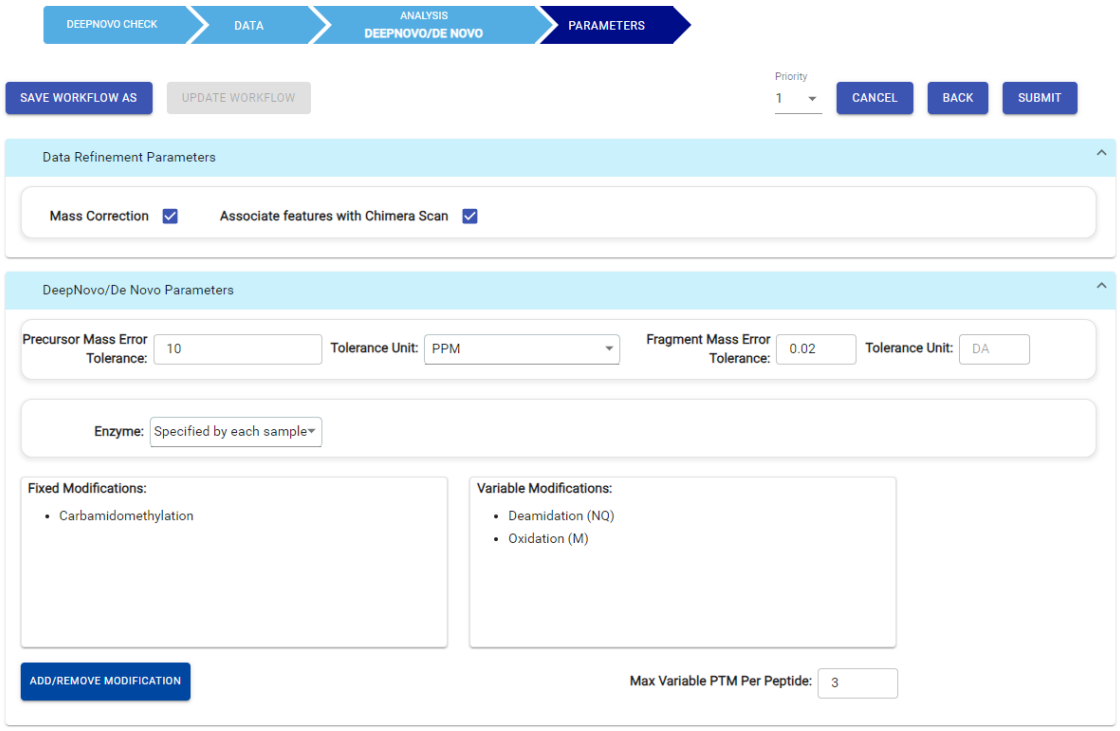
5. For each sample, specify the name and the implicit parameter details: "Enzyme" , "Activation Method", and "Instrument" type, or specify the project details using the drop-down menus at the top of the list if all samples have been prepared and acquired with the same method. If the required enzyme is not listed, it can be created at this page by selecting Create a New Enzyme from the drop down menu, then adding the pattern (cut set, restrict set, and cut side).

6. Click the Next button to continue choosing the analysis workflow.

7. For a DDA dataset, select the "De Novo" workflow.



8. Set the Denovo Parameters. See Section 2.1-2.3 below for details.



9. Click SUBMIT button to start the search.

2.1. Precursor and Fragment Error Tolerance

Precursor Mass Error Tolerance: The maximum mass difference between the *de novo* peptide and the monoisotopic mass of the precursor ion.

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

Fragment Mass Error Tolerance: PEAKS *de novo* uses this value when scoring *de novo* peptide sequences. PEAKS Online considers a fragment ion to be matched if the calculated m/z is within the specified error tolerance. Units are Daltons.

Both precursor and fragment error tolerance parameters need to be set consistently with the mass accuracy of the instrument.

2.2. Enzyme Settings

Select the enzyme used for protein digestion from the enzyme 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 no digestion enzyme was used or when the digestion enzyme exhibits a high degree of non-specificity.

2.3. PTM settings

Click "ADD/REMOVE MODIFICATION" button to open the "Select Modifications" dialogue and specify the fixed PTMs and a few common variable PTMs expected in the sample. Once the PTMs are set, select confirm. Once confirmed, the maximum number of variable PTMs per peptide can be specified and a value of less than 4 is recommended. A fixed modification forces all instances of applicable residues to be modified. A variable modification gives the option for the residues to be modified.

Note

For a specific residue, only one fixed modification is allowed.

Note

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

3. Understanding PEAKS *De Novo* Sequencing Results

When *de novo* sequencing is done, you can access the PEAKS Online project from the "My Projects" page. The analysis report for *de novo* search tool is presented in two tabs:

- **Summary:** This shows an outline of the PEAKS *de novo* search results with key statistics. The overall quality of the experiment can be examined and the filters for *de novo* ALC can be adjusted at the top of the page. This page is displayed after the search is done.
- **De Novo:** This shows a list of peptide sequences detected by *de novo* sequencing tool, which passed the de novo ALC filters set in the Summary page.

The *de novo* view displays the *de novo* sequencing results. The table on the top displays *de novo* sequences and their characteristics (length, m/z etc.), while the bottom section provides additional information about the peptide-spectrum match. The results can be filtered and exported.

3.1. Result Filtration

In the summary tab, the PEAKS Online *de novo* sequencing result can be filtered using Average Local Confidence (ALC) score. Low quality *de novo* sequences can be filtered out by specifying a minimal threshold of the ALC score. The purpose of filtering is to remove poor sequences in which residues are incorrectly identified. By default,

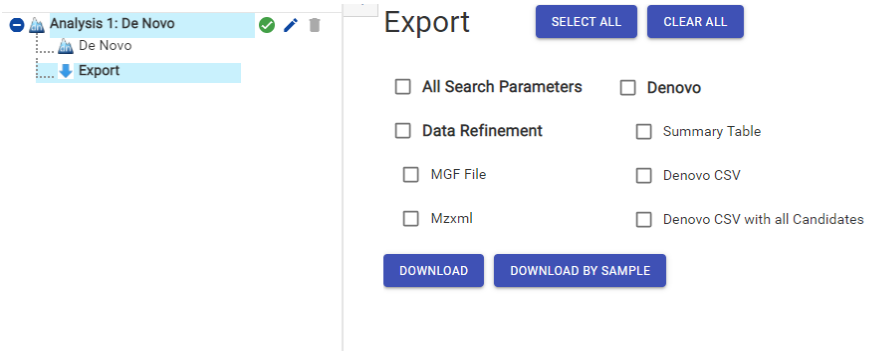
the ALC threshold is set to 50%. After changing the threshold, click APPLY to apply the new filter. The result in the De Novo result view will be updated accordingly.

In the De Novo tab, result for different samples can be viewed by selecting the sample using the "Show Results for Specified Sample" drop-down menu. The purpose of this filtering is to limit the quantity of *de novo* sequencing results displayed in the table. Results can also be filtered using the Filtered Peptides icon at the top left of the De Novo window. Here, results can be filtered by an amino acid sequence, scan #, precursor ID, m/z, or RT. Click APPLY to apply the new filter. The result will be updated accordingly.

3.2. Exporting *De Novo* Results

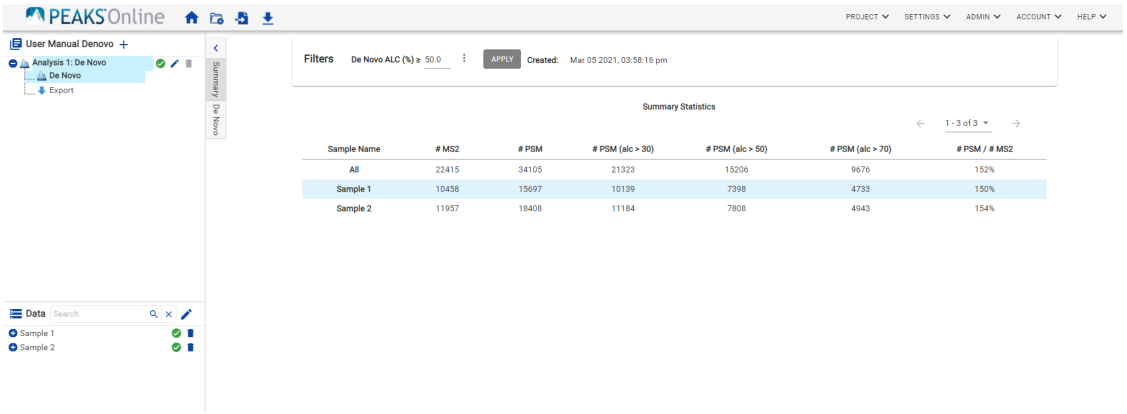
The *de novo* results can be exported from the De Novo results tab or from the Export page of the analysis. In the De Novo tab, above the peptide table there are two export options: "Denovo", and "Denovo with all Candidates". In the Export page, there are three files which can be exported: "Summary Table", "Denovo CSV", and "Denovo CSV with all Candidates".

- **Denovo CSV**: the filtered *de novo* results shown are exported. This file can also be downloaded from the Export page as shown in the following figure.
- **Denovo CSV with all candidates**: for each MS2 spectrum, denovo results of the 5 highest ALC scores are generated. *De novo* sequences that pass the ALC score filter are all included in this file. This file can also be downloaded from the Export page as shown in the following figure.
- **Summary Table:** shows the overview number of each sample for the fields MS2, PSM, PSM (alc>30), PSM (alc>50), PSM (alc>70), and notes.



3.3. Summary View

The "Summary" view reports key statistics as an overview of the result.



3.3.1. Result Statistics

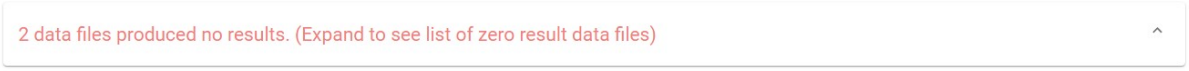
This table of the summary page shows the summary date for each sample for the following fields:

- **# MS2**: The number of MS/MS scans from the data.
- **# PSM**: The number of peptide-spectrum matches.
- **# PSM (alc>30)**: The number of peptide-spectrum matches with an ALC score greater than 30.
- **# PSM (alc>50)**: The number of peptide-spectrum matches with an ALC score greater than 50.
- **# PSM (alc>70)**: The number of peptide-spectrum matches with an ALC score greater than 70.

> **Note**
>
> ALC score is the average local confidence score. ALC is calculated as the total of the residue local confidence scores in the peptide divided by the peptide length.

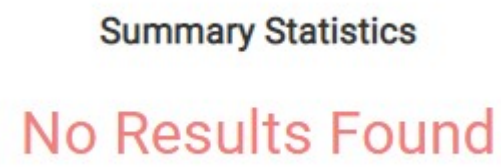
3.3.2. Low Quality Data Warnings

If PEAKS Online detects any data files producing very few or no results a warning section will appear at the top of the Summary Page indicating the number of data files producing poor or no results.



This section can be expanded to display a list of data files in the analysis that are producing poor or no results.

If there were no results found in any of your data files, a warning message stating "No Results Found" will be displayed instead of the usual Summary Page figures and tables.



3.4. Peptide Table

The peptide table shows the filtered *de novo* sequencing results. For each scan, the best candidate is displayed.

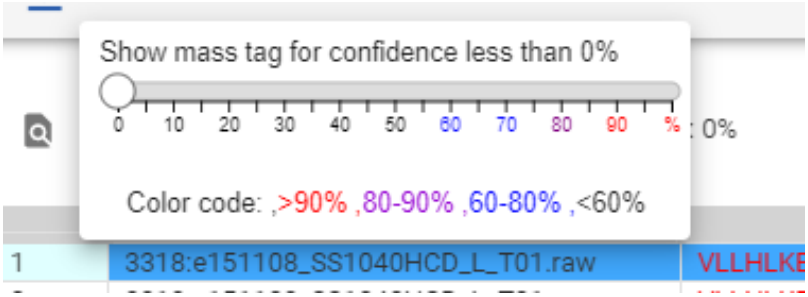
The following list describes the contents in each column:

- **Scan**: The scan number is a unique index for tandem mass spectra in the data. If the scan numbers were read from the data file, the scan numbers that match with the instrument data file will be shown.
- **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.
- **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 peptide sequence.
- **m/z**: The precursor mass-to-charge ratio.
- **z**: The precursor charge.
- **RT**: The retention time (elution time) at which the spectrum was 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 10E6 × (precursor mass - peptide mass) / peptide mass.
- **PTM**: The types and the numbers of the modifications present in the peptide shown as color-coded icons.
- **Mode**: The fragmentation mode in which *de novo* sequencing is performed by the algorithm.

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

Mass Tags: The low confidence residues can be displayed as mass tags by adjusting the local confidence score threshold by selecting the Set Residue Confidence Threshold icon and then using the button to adjust the threshold. 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. When low confidence residues are reduced to mass tags, the remaining residues become sequence tags. The length of the longest tag is called the tag length and reported in the export files.



Filter Peptides: The peptide table can be filtered by selecting the button and prompting the "Denovo Filter" window. The peptides can be searched by sub-sequence, scan number, precursor Id, m/z, and retention time.

3.5. Spectrum Annotation

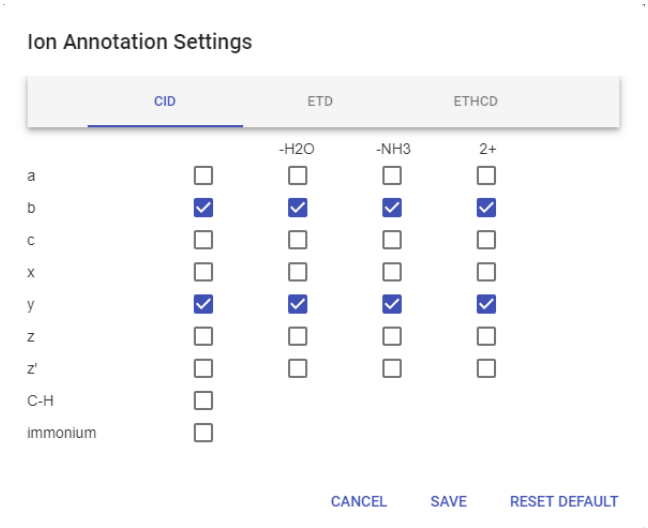
The spectrum annotation displays a graphical representation of the peptide-spectrum match.

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

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

- Zoom in/out on m/z: Place the cursor on the point in the spectrum you would like to zoom in on. Alternatively, place the cursor in the Error (da) figure and scroll the mouse wheel.
- Increase/Decrease peak intensity: Place the cursor on the Intensity axis and scroll the mouse wheel.
- See the whole spectrum, or reset a zoomed in panel: click the 1:1 button.

Click to open the Ion Annotation Settings dialogue and change the ions to be annotated in the spectrum. To reset the settings to the PEAKS Online defaults, use "RESET DEFAULT" button.



The button will switch the y-axis to relative intensity. Once selected, scrolling over peaks will display the relative intensity instead of the absolute intensity.

The button will switch the y-axis to absolute intensity. Once selected, scrolling over peaks will display the absolute intensity instead of the relative intensity.

3.6. Error Map

The "Error Map" shows the mass errors of the 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 "Ion Annotation Settings".

3.7. Ion Table

The "Ion Table" shows the calculated mass of possible fragment ions based on the ion types selected in the Ion Annotation Settings. If a fragment ion is found in the spectrum, its mass value is displayed in color. N-terminal ions are shown in blue and C-terminal ions are shown in red. A fragment ion is found when there is a matching peak within the mass error tolerance, as defined in the *de novo* sequencing parameters.

4. FAIMS Data

Field asymmetric waveform ion mobility spectrometry (**FAIMS**) analyzes complex mixtures and separates the ions according to their mobilities. PEAKS Online has the ability to load FAIMS data and analyze it using *de novo* sequencing, database search (PEAKS DB), unexpected modification (PEAKS PTM) and sequence variants (SPIDER) search, and labelled and label free quantification.

Note

There is no special procedure required to load FAIMS data. Refer to Chapter 1, *Overview* for more details on data loading.

The analysis result for FAIMS dataset contains the additional information for CV values compared to the non-FAIMS dataset. The CV information can be found at the following places for a Denovo analysis.

- Additional **CV** column in *de novo* table

  de novo table contains one additional column for CV values

- Additional **CV** filter in *de novo* table

  The filter for *de novo* table contains one additional option for CV, which selects the target scan based on the CV value

- Additional **CV** column in the exported *Denovo CSV*

  The exported *Denovo CSV* contains one additional column for CV values

- Additional **CV** column in the exported *Denovo with all candidates CSV*

  The exported Denovo with all candidates CSV contains one additional column for CV values

Chapter 6. Peptide, PTM, and Mutation Identification (PEAKS DB, PEAKS PTM, and SPIDER)

1. Overview

PEAKS Online 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 Online supports multiple enzyme digestion by which almost full sequence coverage can be achieved for single protein studies. PEAKS Online 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 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.*

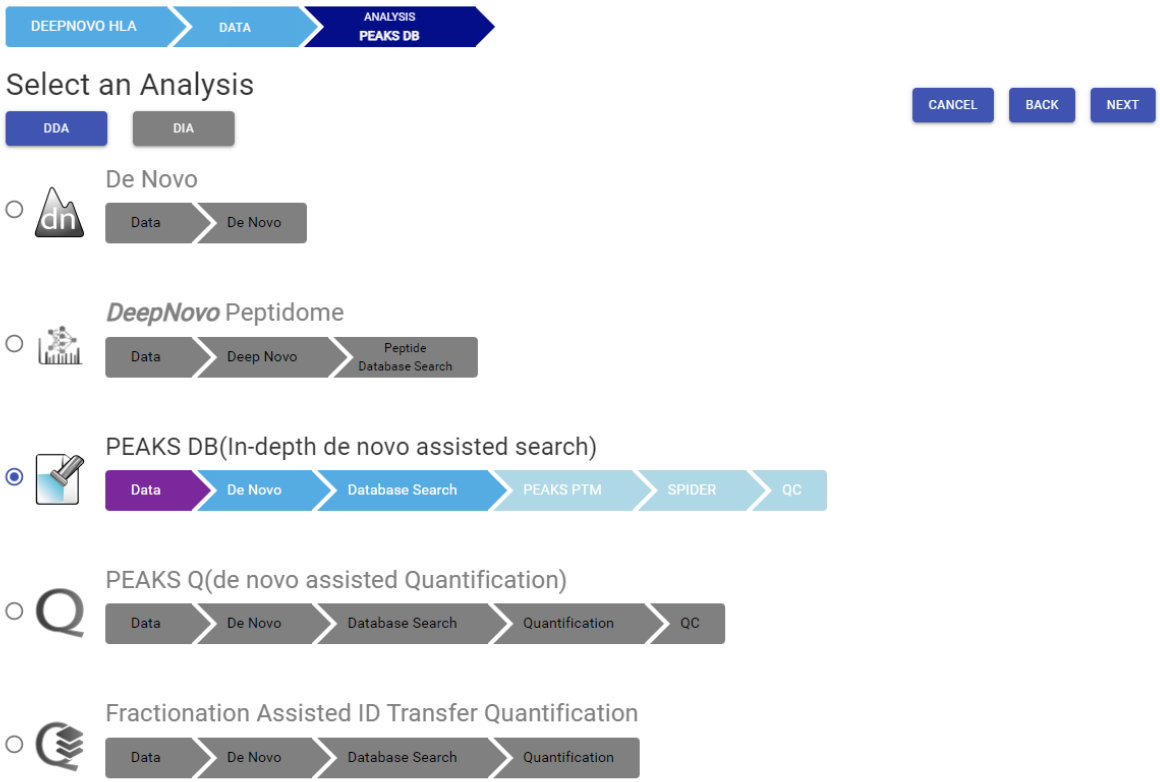
2. Database Search Workflow and Parameter Settings

PEAKS DB, PEAKS PTM, and SPIDER can be launched together as a work flow in a PEAKS Online search. An analysis report is generated to show the combined database search result. Users can also do PEAKS DB only, PEAKS DB plus PEAKS PTM, PEAKS DB plus SPIDER or PEAKS DB plus PEAKS PTM plus SPIDER.

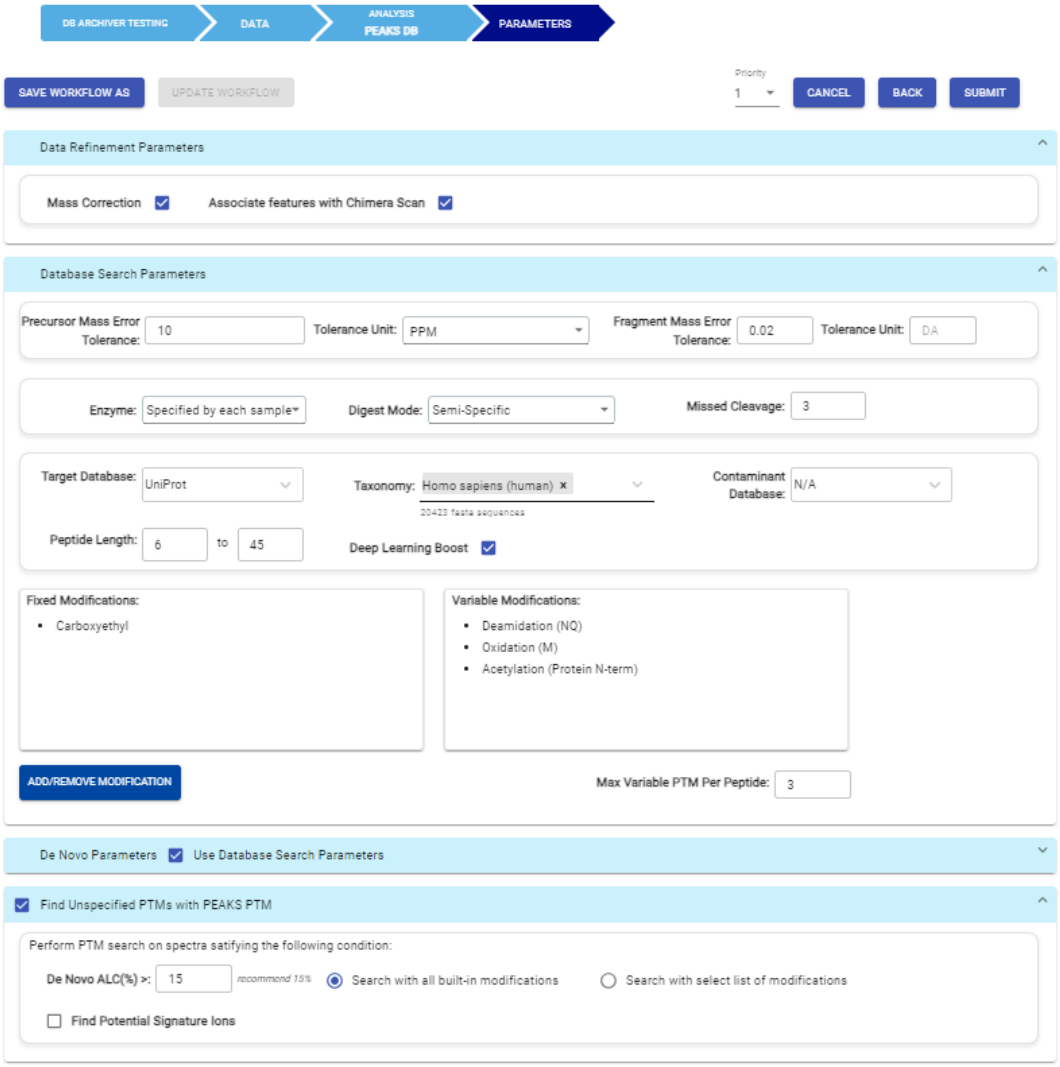
To conduct a database search workflow, follow the steps below:

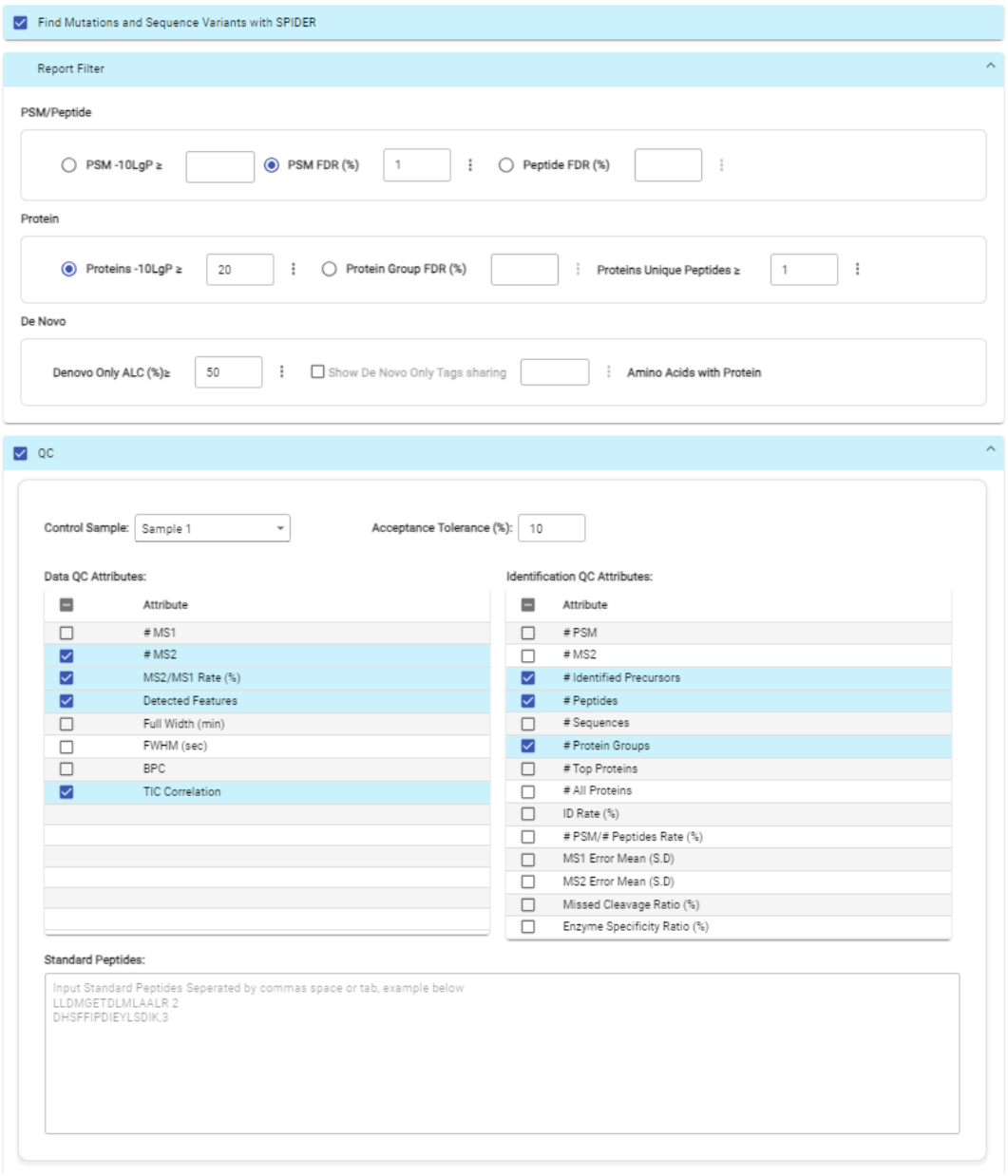
1. Refer to Chapter 5, Peptide De Novo Sequencing, Section 2, “*DeepNovo/De Novo* Analysis Workflow and Parameter Settings” , Step 1-6 to create a new project and add data.

2. Select the "Database Search" workflow.



3. Set the DB Search Parameters. Enable Deep Learning Boost algorithm for an enhanced number of PSMs identified (optional). Enable PEAKS PTM search to find more PTMs (optional). Enable SPIDER search to find more mutations (optional). See details in Section 2.1-2.6 below.


DB ARCHIVER TESTING
DATA
ANALYSIS
PEAKS DB
PARAMETERS
SAVE WORKFLOW AS
UPDATE WORKFLOW
Priority
1
CANCEL
BACK
SUBMIT
Data Refinement Parameters
Mass Correction
Associate features with Chimera Scan
Database Search Parameters
Precursor Mass Error Tolerance:
10
Tolerance Unit:
PPM
Fragment Mass Error Tolerance:
0.02
Tolerance Unit:
DA
Enzyme:
Specified by each sample
Digest Mode:
Semi-Specific
Missed Cleavage:
3
Target Database:
UniProt
Taxonomy:
Homo sapiens (human)
20423 fasta sequences
Contaminant Database:
N/A
Peptide Length:
6
to
45
Deep Learning Boost
Fixed Modifications:
Carboxyethyl
Variable Modifications:
Deamidation (NQ)
Oxidation (M)
Acetylation (Protein N-term)
ADD/REMOVE MODIFICATION
Max Variable PTM Per Peptide:
3
De Novo Parameters
Use Database Search Parameters
Find Unspecified PTMs with PEAKS PTM
Perform PTM search on spectra satifying the following condition:
De Novo ALC(%) >:
15
recommend 15%
Search with all built-in modifications
Search with select list of modifications
Find Potential Signature Ions



4. Click SUBMIT button to start the search.

2.1. Data Refinement Parameters

Mass Correction:: The precursor m/z values given by some instruments are often not of the monoisotopic ion. This creates problems in downstream analysis. By enabling mass correction, PEAKS will examine the isotope shapes in the corresponding MS scans and accurately adjust the precursor m/z to equal the monoisotopic ion m/z.

Associate Features with Chimera Scan: If enabled, PEAKS will assume that a tandem scan may contain fragments from two or more peptides, which will allow PEAKS to identify co-eluting peptides found within the acquisition window of the scan. PEAKS accomplishes this by searching for peptides that match the monoisotopic mass of any peptide feature in the acquisition window.

2.2. Precursor and Fragment Error Tolerance

Precursor mass : This enables the selection of a precursor mass error tolerance. PEAKS Online DB Search identifies peptides within the precursor mass tolerance, which is the allowable m/z shift between the theoretical value of the peptide versus the observed value of the precursor ion. Precursor mass tolerance can be specified in either Da or ppm.

Fragment ion : This enables the selection of a fragment ion m/z tolerance, which is the allowable m/z shift between the theoretical value of the fragment ions of the peptide versus the observed value. PEAKS Online DB Search uses this value when scoring peptide-spectrum matches. PEAKS Online considers that a fragment ion is matched if the calculated m/z is within the tolerance. Fragment mass tolerance can be specified in Da.

Error tolerance parameters need to be set consistently with the mass accuracy of the instrument.

2.3. Enzyme Settings

Enzyme : PEAKS Online digests the protein database *in silico* to generate peptide candidates. It is necessary to specify the enzyme for protein digestion from the Enzyme drop-down menu.

Note

When the selected dataset is digested with different enzymes, "Specified by each sample" allows samples to be analyzed separately using their respective enzymes specified during project creation.

Note

"None" is a special enzyme digest rule allowing non-specific cleavage at both ends of the peptide. The "None" digest rule can cut at every residue, generating peptides with lengths up to 65 amino acids. Its usage is recommended when no digestion enzyme was used or when the digestion enzyme exhibits a high degree of non-specificity.

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

Digest Mode : This specifies the type of *in silico* cleavages allowed at a peptide termini 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 rules.

Note

If the enzyme is specified as "None", then no matter which mode is selected for "Digest Mode", the "None" enzyme digest rule will be applied.

2.4. Database Settings

A protein database must be added first before creating a new project to enable its selection in the workflow. In the Target Database drop-down menu, select the appropriate database that have been configured in PEAKS Online and set the taxonomy, if applicable. A contaminant database can also be selected from the same list of databases that have been configured in PEAKS Online.The length for a peptide can be set when creating an analysis, where the default peptide length range is between 6 and 45. Enable Deep Learning Boost algorithm for an enhanced number of PSMs identified.

2.5. PTM Settings

Click "ADD/REMOVE MODIFICATION" button to open the "Select Modifications" dialogue and specify the fixed PTMs and a few (less than 11) common variable PTMs expected in the sample. If the required modification is not listed, the modification can also be created here. Additionally, the maximum number of variable PTMs per peptide can be specified and a value of less than 4 is recommended. A fixed modification forces all instances of applicable residues to be modified. A variable modification gives the option for the residues to be modified.

Note

For a specific residue, only one fixed modification is allowed.

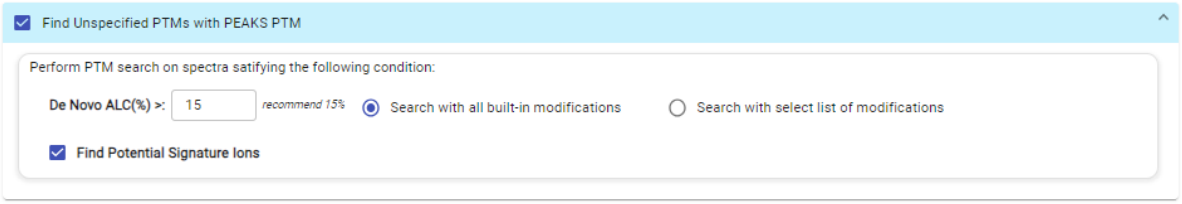
Note

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

2.6. Find Unspecified PTMs with PEAKS PTM

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

Alternatively, PEAKS PTM can search with a list of preferred modifications. Once "Search with preferred modifications" has been selected, click the "ADD/REMOVE MODIFICATION" button to select desired PTMs for the PTM search. Importantly, PEAKS PTM allows an unlimited number of variable modifications to be searched, however, limiting the number can improve search accuracy. By default, PEAKS PTM considers the same number of maximum number of variable PTMs per peptide as set in the DB search parameters. 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. In the PEAKS PTM result page, search results from PEAKS DB and PEAKS PTM are combined and shown.



2.6.1. PTM signature Ions

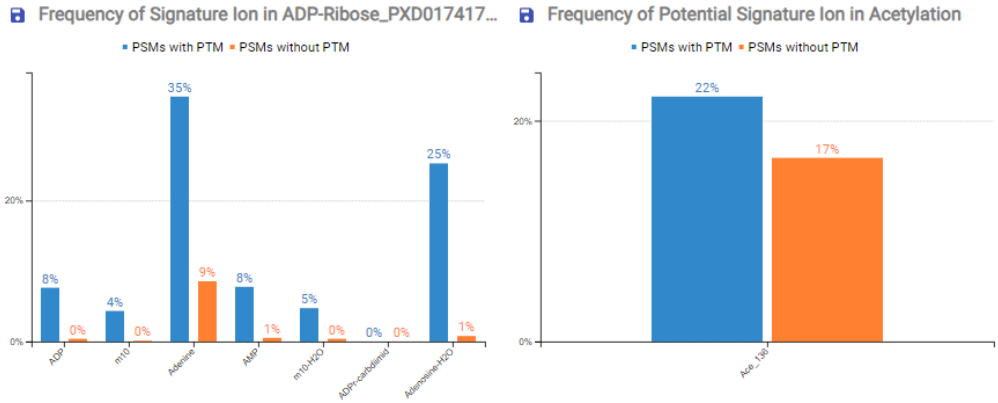
Some modified peptides give rise to signature ions (e.g. immonium ions of the modified residue) during fragmentation. These signature ions are included in PEAKS online to boost confidence of PTM identification. PEAKS uses the identification of these ions to perform re-scoring of identified peptide sequences. Users can specify expected signature ions for any custom modification. Alternatively, users can select "Find Potential Signature Ions" to identify potential signature ions in fragment scans.

PEAKS provides a summary of expected and potential signature ion identification on the PEAKS PTM summary page.

Table 4. Signature Ions Detected

1 - 2 of 2

Modification	Signature Ion name	Mass	#PSM with modification	#PSM with modification with signature ion	#PSM without modification	#PSM without modification with signature ion	Specificity (Expected/Potential)
ADP-Ribose_PXD017417-Sig	ADP	428.0366	44477	3396	55520	224	Expected
	m10	542.0683		1946		90	
	Adenine	136.0623		15463		4755	
	AMP	348.0703		3456		299	
	m10-H2O	524.0578		2108		198	
	ADPr-carbdiimide	584.0902		22		30	
	Adenosine-H2O	250.0940		11239		428	
Acetylation (Protein N-term)	Ace_136	136.0600	2383	527	97614	16245	Potential



2.7. Find More Mutations with SPIDER

Select this option to enable a SPIDER search. SPIDER performs homology searches on spectra with good *de novo* sequences (i.e. ALC>15%) that remain unidentified by either PEAKS DB and/or PEAKS PTM. SPIDER identifies novel peptides that are homologous to peptides in the searched protein database. SPIDER is good for cross-species searches and for finding point mutations of the protein. In the SPIDER result page, search results from PEAKS DB, PEAKS PTM (if applicable), and SPIDER are combined and shown.

2.8. PSM Filter

PSM filter section is a mandatory section to filter out some results under the user-defined filter. Any changes to the filter will create a new analysis with old parameters and update the results and statistics.

- **PSM -10lgP:** The PEAKS peptide score (-10lgP) is calculated for every peptide-spectrum match (PSM) reported by PEAKS DB, PEAKS PTM, and SPIDER. The score is derived from the p-value that indicates the statistical significance of the peptide-spectrum match. A peptide may be matched to many spectra, resulting in multiple PSMs. In that case, the peptide's score is calculated as the maximum 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* . A minimum requirement can be set and all identifications must pass this filter.

- **PSM FDR (%):** The PSM filter can also be set to use the false discovery rate. FDR is estimated using a "decoy-fusion" approach.

  ### Note

  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.

- **De Novo ALC (%):** Refer to Chapter 5, Peptide *De Novo* Sequencing, Section 3.4, “Peptide Table” for the full ALC description. This filter controls the minimum ALC score required for display in the *de novo* only result. A default of 50% is set.

- **De Novo Only Tags:** The "Denovo Only Tags" tab shows a table of "de novo only" peptides with sequence tags matched to the protein.

3. Understanding PEAKS Search Results

After a PEAKS Search is complete, all analyses selected during the project creation will be combined under the workflow drop-down menu. One is the *de novo* sequencing result, which includes *de novo* peptide sequences listed, spectrum annotation and other detailed information. 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.

The analysis results for database search tools are presented in four tabs:

- **Summary :** This shows an outline of the PEAKS Online 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. This page shows after the search is done.

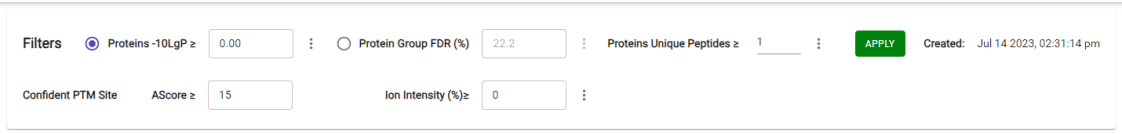
- **Protein :** This shows a list of protein identifications. This view also visualizes protein sequence coverage and helps with protein characterization.

- **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 Online database search.

3.1. Result Filtration

At the top of each summary result page, filters can be set. Identifications that fall below the filters will not be visible in the result pages or exports. After changing any of these filter parameters, click the APPLY button to apply and save the new filters. The results and statistical information will be updated accordingly.



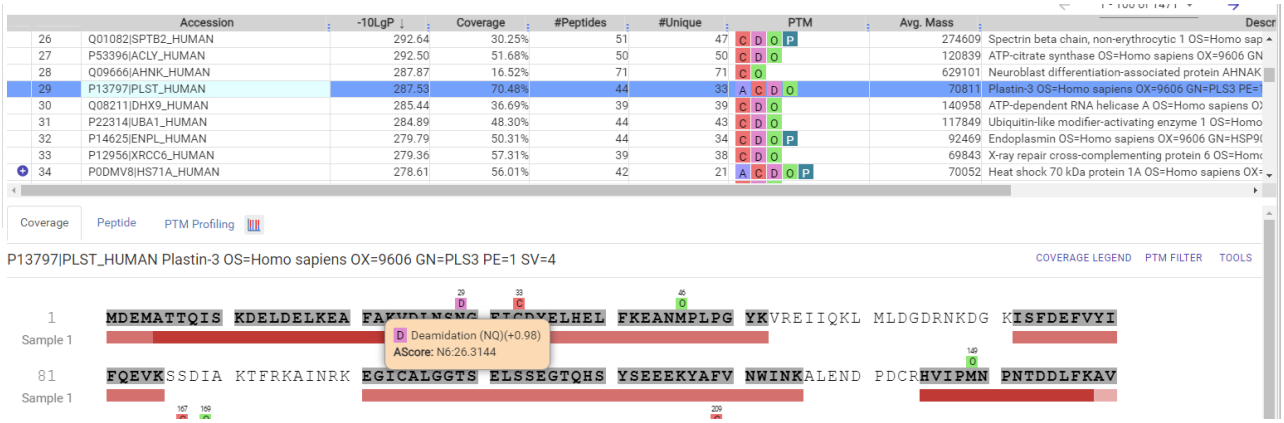
- **Proteins -10lgP:** The PEAKS protein score (-10lgP) is calculated as the weighted sum of the -10lgP scores of the protein's supporting peptides. After removing any redundant peptides, the supporting peptides are sorted by -10lgP scores in descending order, and the k-th ranked peptide contributes to the weighted sum with a weight of 1/k. A default threshold of 20 is set.
- **Proteins Unique Peptides:** The minimum number of unique supporting peptides for a protein identification. A unique peptide is defined as a peptide that can be mapped to only one protein group.

3.1.1. Confident PTM Site Filter

PEAKS PTM identification is site-specific allowing users to determine confident modification sites based on two parameters: minimal ion intensity and AScore.

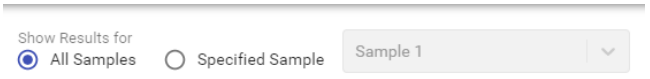
- **Minimal Ion Intensity:** Minimum ion intensity of each fragment ion pair, which define the modification within the peptide, compared to the base peak ion intensity. A default of 2% means that at least one pair of ions (e.g. ‘y’ ions) defining the modified amino acid has intensity of 2% or higher for both ions within the pair.
- **AScore:** A -10LgP score relating to positional confidence of the PTM. The p-value indicates the likelihood that the peptide is matched by chance. Therefore the higher the AScore the better.

Confident PTM Site filters are applied to the protein coverage, PTM profiling view and PTM-profiling CSV export. Modifications will only be annotated above the protein sequence if it passes the filters set by the user.



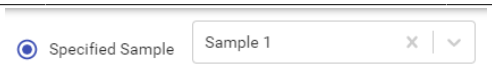
3.1.2. Show Results for All Samples

For the protein and the peptide tables, "All Samples" can be selected to view the combined search results from all the samples. In the Protein page, selecting "All Samples" will change the coverage panel displayed when a specific protein is highlighted. Supporting peptides from all samples instead of just a single sample will be shown.



3.1.3. Show Results for Specific Sample

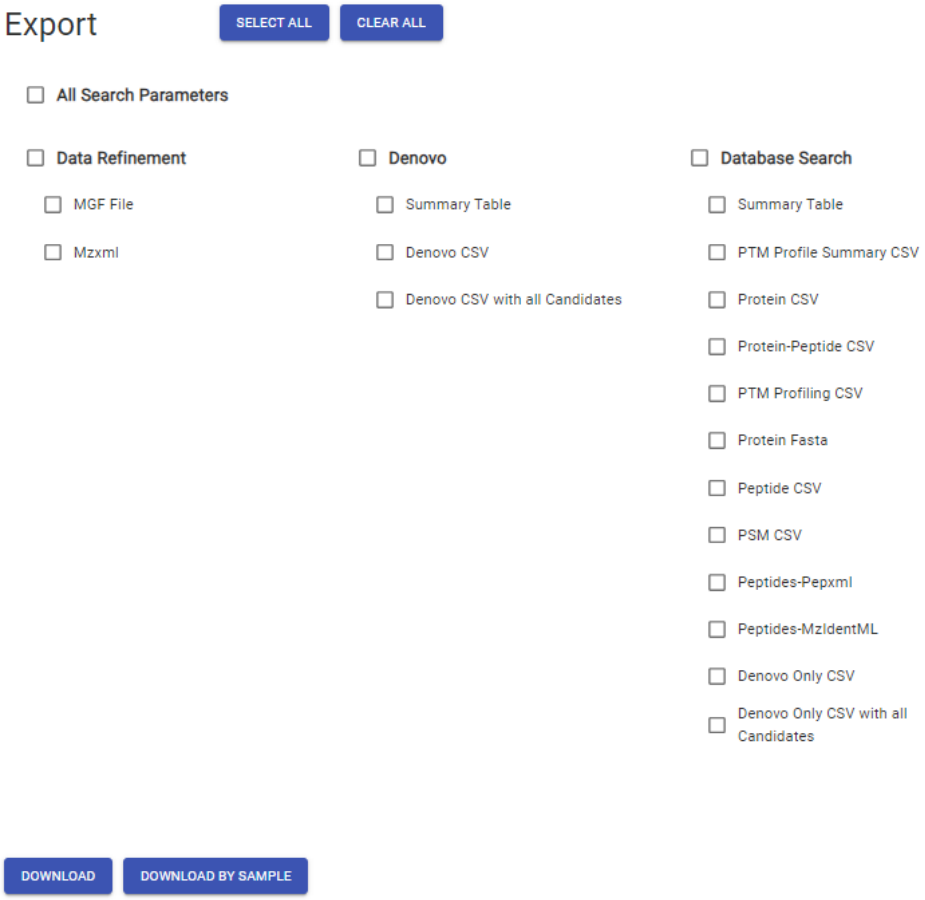
In the Protein, Peptide and De Novo page results, the filter can also be set to "Specified Sample" which will only show the results a single sample. By clicking in the text you can search for a specific sample by name.



Selecting "Specified Sample" will change the protein coverage view that is visible by clicking on a protein in the protein table. Peptides in the Protein, Peptide and De Novo Only pages will only be visible if they belong to the sample chosen. Once selected, the peptide tables will automatically add the "Area" column and the "# Spec" values will display the spectral count for the specified sample. For more information, refer to Section 3.5.2, “Peptide Table” . Similarly, the protein table will also include the columns "Coverage" and "Area" for the specified sample. For more information refer to Section 3.4.2, “Protein Table” .

3.2. Result Exporting

PEAKS Online database search results can be exported in the Export page:



- **Summary Table**: Content from the PEAKS DB/PTM/Spider summary statistics table saved in text CSV format.
- **PTM Profile Summary:** Content from the PEAKS DB/PTM/Spider summary PTM profile table (Table 3) saved in text CSV format.
- **Protein CSV:** A comma separated file containing all identified proteins and associated details from the protein table for all samples.
- **Protein-Peptide CSV:** A comma separated file containing all identified peptides and associated details from the peptide table found in the Protein page. Further details are provided regarding the protein the peptide supports. If a peptide supports multiple proteins, multiple rows are included for each supported protein for all samples.
- **PTM Profiling CSV:** Detailed information for the abundance of confidently identified PTMs versus unmodified peptides for all identified proteins. The PTM profile data are grouped together by protein, then by modification type, and then by the modification site on the protein sequence. The headers of the CSV column are similar to the PTM Profile Table.

- **Peptide CSV:** A comma separated file containing all identified peptides and associated details from the peptide table for all samples.

- **PSM CSV:** A comma separated file containing all identified peptide spectrum matches and associated details from the peptide table for all samples.

- **Pepxml:** An XML export of the identification results matching the pepxml schema. This can be uploaded to software that accept data in this format.

- **MzIdentML:** An XML export of the identification results matching the MZIdentML schema. This can be uploaded to software that accept data in this format.

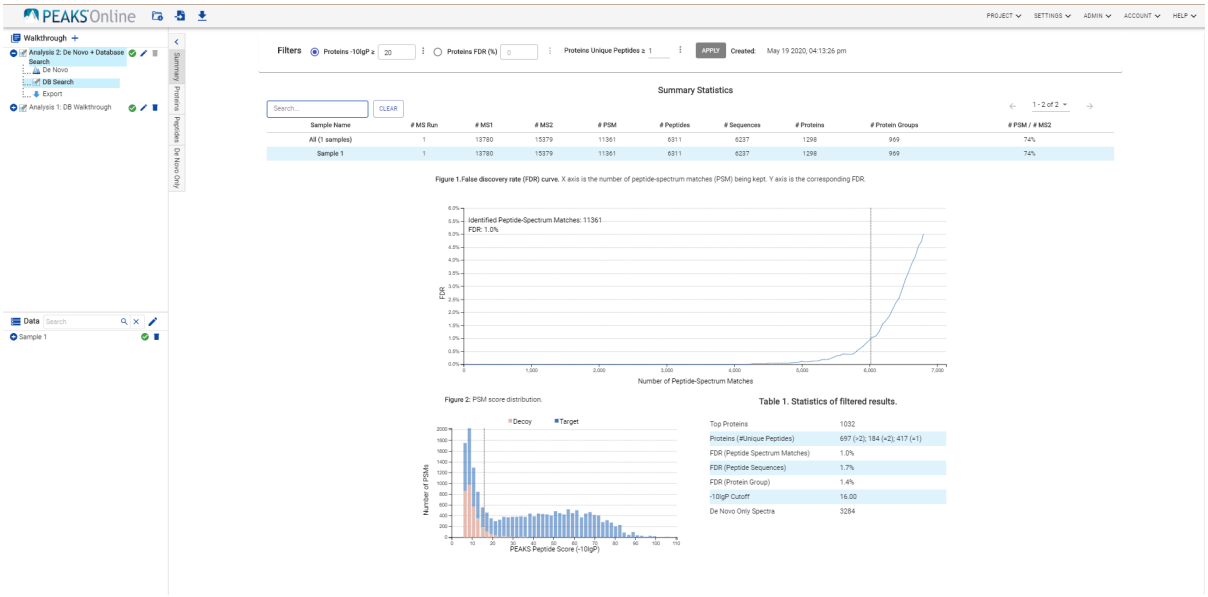
- **Denovo Only CSV:** Top scans from De Novo results that do not have an identification result by PEAKS DB, PEAKS PTM, or SPIDER that pass the set filters. Refer to Chapter 5, Peptide *De Novo* Sequencing, Section 3.2, “Exporting *De Novo* Results” for detailed definition. A separate CSV file is generated for each sample.

- **Denovo Only CSV with all Candidates:** De Novo results from scans that do not have an identification result by PEAKS DB, PEAKS PTM, or SPIDER that pass the set filters. Refer to Chapter 5, Peptide *De Novo* Sequencing, Section 3.2, “Exporting *De Novo* Results” for detailed definition. A separate CSV file is generated for each sample.

- **Protein Fasta** : A list of protein identifications will be saved to proteins.fasta.

3.3. Summary View

The Summary view reports key statistics as an overview of the result. In the report, several statistical charts can be examined to assess result quality.



3.3.1. False Discovery Rate (FDR) Curve

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

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

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

Note

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

Note

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

3.3.2. Result Statistics Tables

The first chart displays the summary information for each sample.

- **Sample Name:** This column lists the samples names with the exception of the first row, which shows the combination of all the samples.
- **# MS2:** The total MS/MS scans in the sample summed across all fractions.
- **# PSM:** Total number of peptide spectrum matches (PSMs) found in the sample.
- **# Peptides:** Total number of peptides found ignoring differences with leucine (L) and isoleucine (I) amino acids. Peptides with the same primary sequence but different PTMs are counted separately.
- **# Sequences:** Total number of different peptide backbone sequences found ignoring differences with leucine (L) and isoleucine (I) amino acids. Peptides with the same primary sequence but different PTMs are counted as one sequence.

- **# Proteins:** Provides the number of proteins given the parameters set for Protein Score (-10lgP) and number of unique peptides.
- **# Protein Groups:** PEAKS Online groups proteins identified by a common set of peptides. This number shows the number of protein groups in the filtered result based on All proteins.

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

- **Summary Statistics:** Presents the total number of MS and MS/MS scans from the data.
- **Table 1. Statistics of filtered results:** Presents the statistics from the data analysis and the 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:** PEAKS Online groups proteins identified by a common set of peptides. This number shows the number of protein groups in the filtered result based on All proteins.
  - **Top Proteins:** The number of proteins that 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 included here.
  - **Proteins:** The number of proteins given the parameters set for PEAKS Protein Score(-10lgP) and number of unique peptides.
  - **Proteins (#Unique Peptides):** In Table 2, 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. Peptides are counted based on their backbone sequence while considering I/L amino acids to be the same.
  - **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 Sequences):** 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.
  - **-10lgP Cutoff:** The value is determined by the estimated PSM FDR. Only the peptides with PSMs above the -10lgP score threshold are listed in the Peptide view.
  - **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 project.
- **Table 2. Number of identified peptides in each sample by the number of missed cleavages.** Presents the number of identified peptides in each sample with the number of missed cleavages, indicating the enzyme digestion efficiency.
- **Table 3. PTM Profile:** shows modifications identified in the filtered result and the number of PSMs containing each modification. Confident PTM site filters do not apply here.
  - **Name:** The given name of the PTM.

- **⊿Mass:** The monoisotopic change in mass caused by the PTM.
- **Position:** The amino acids modified by the PTM.
- **# PSM:** The number of PSMs with the given PTM.
- **-10lgP:** The highest -10lgP score achieved by a peptide modified by the PTM.
- **Area:** The area achieved by the modified peptide with highest -10lgP score.
- **AScore:** The AScore achieved by the modified peptide with highest -10lgP score.

3.3.3. Summary Figures

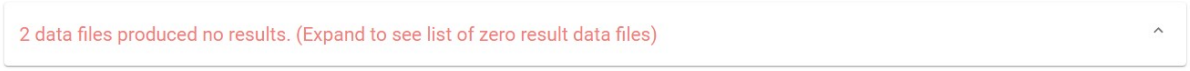
PSM Score Distribution: Figure 2 displays a histogram of the PSM -10LgP score distribution for target and decoy PSMs. This figure helps visually examine the FDR control in your search results**.**

Precursor mass error of peptide-spectrum matches: Figures 3(a) and 3(b) show the precursor mass errors of PSMs in filtered results. Figure 2(a) shows the distribution of the precursor mass errors in a histogram with and without software calibration. Figure 2(b) shows precursor m/z versus precursor mass error in a scatter plot after software calibration. Figures 2(a) and 2(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.

Data Completeness: Figure4 displays a histogram of data completeness based on the percentage of replicate protein groups identified across samples in the analysis. Overall data completeness is an average of the %replicates identified for all protein groups.

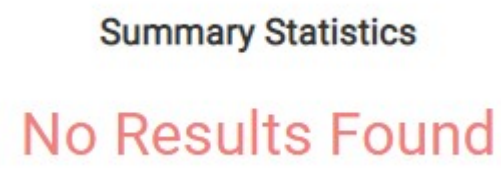
3.3.4. Low Quality Data Warnings

If PEAKS Online detects any data files producing very few or no results a warning section will appear at the top of the Summary Page indicating the number of data files producing poor or no results.



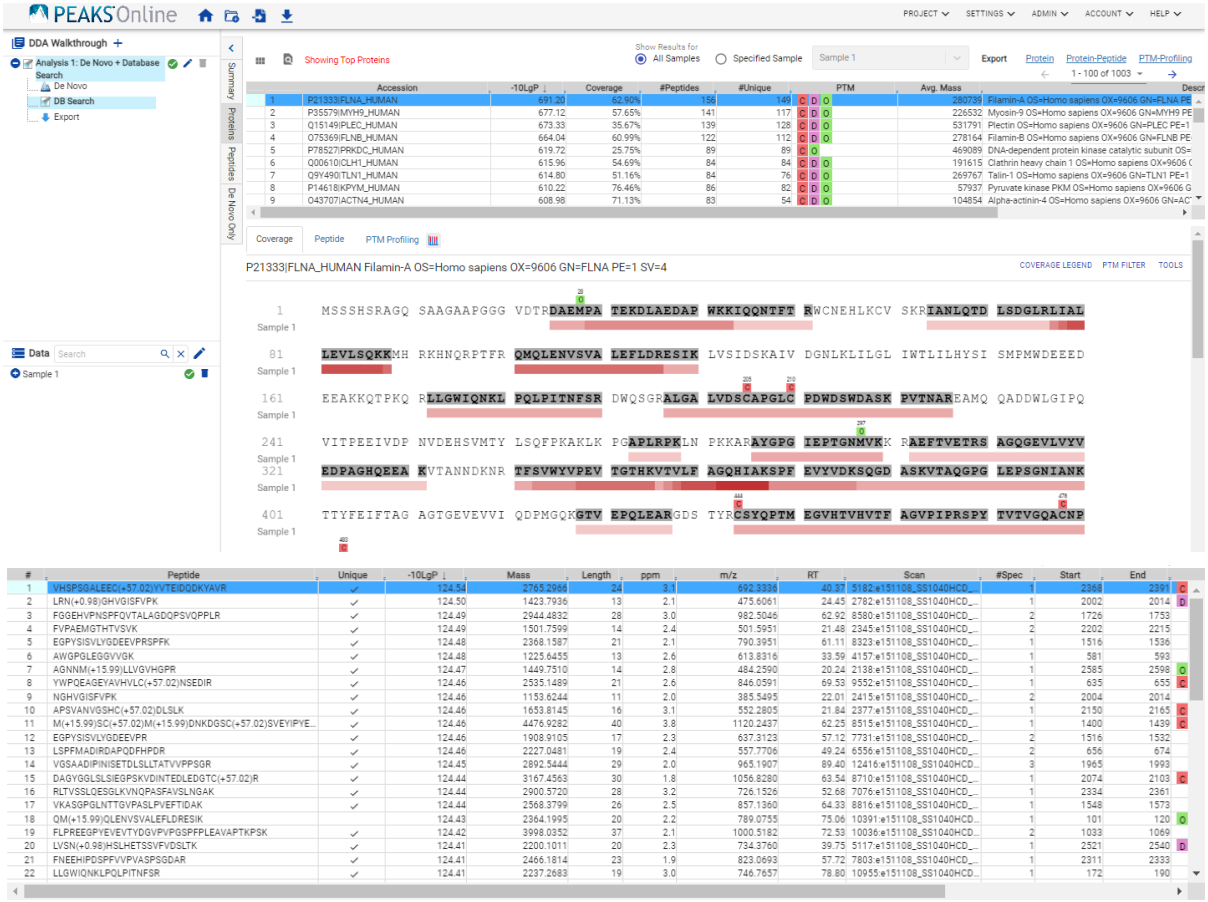
This section can be expanded to display a list of data files in the analysis that are producing poor or no results.

If there were no results found in any of your data files, a warning message stating "No Results Found" will be displayed instead of the usual Summary Page figures and tables.



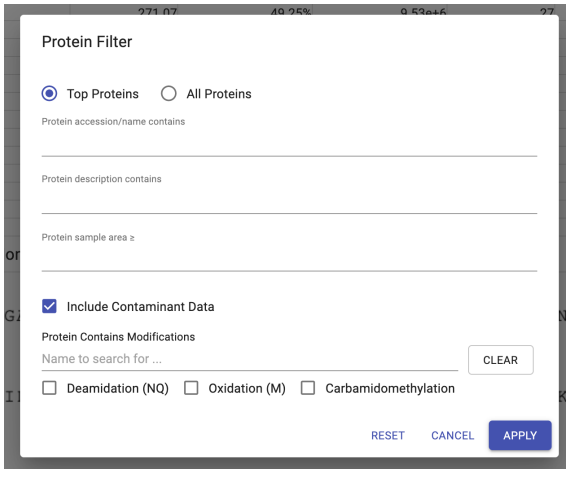
3.4. Protein View

The Protein view lists protein identifications that have been filtered by the current settings set for the project. It also visualizes the protein sequence coverage for identified proteins. After clicking the "Peptide" tab we can view the supporting peptide list, double click any peptide can lead to that peptide sequence in the peptide page.



3.4.1. Protein View Filters

By selecting the button, the Protein Filter will pop up. It controls which protein to display.



- **Show Top or All 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.
- **Protein accession/name contains:** If you add text here, only proteins with accessions that contain the text will be included in the protein table.
- **Protein description contains:** If you add text here, only proteins with descriptions that contain the text will be included in the protein table.
- **Protein sample area:** Filters 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.

  ### Note

  Only numeric digits can be typed in the protein sample area. For example, instead of 1.00e+4, 10000 must be typed.
- **Spider Mutations:** This filter is only visible in SPIDER result view filter. Check it to view proteins that have mutations.
- **Protein Contains Modifications** : Proteins that are supported by peptides containing selected PTMs will be included.

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 indicates that the group has multiple proteins. To expand the group, click that button.

The following columns are sortable: Accession, -10LgP, Coverage, Peptides, #Unique, Avg Mass, when All Samples are shown.

The columns in the Protein table are:

- **Accession:** The accession id of the protein as seen in the FASTA database.
- **-10lgP:** The protein confidence score. Refer to Section 3.1, “Result Filtration” for detailed explanation.
- **Coverage:** The percentage of the protein sequence that is covered by supporting peptides. This is the total coverage including results from all samples, unless results are only shown for a specified sample.
- **#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 the 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.

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

Global Protein Comparative View provides detailed comparisons among different samples regarding the identified protein. To Show/Hide Columns, click ▦ on top-left of the protein table. From the dialog, users can add and remove sample coverage, area, and spectral count columns. In the protein table, the selected columns will be displayed and can be exported.

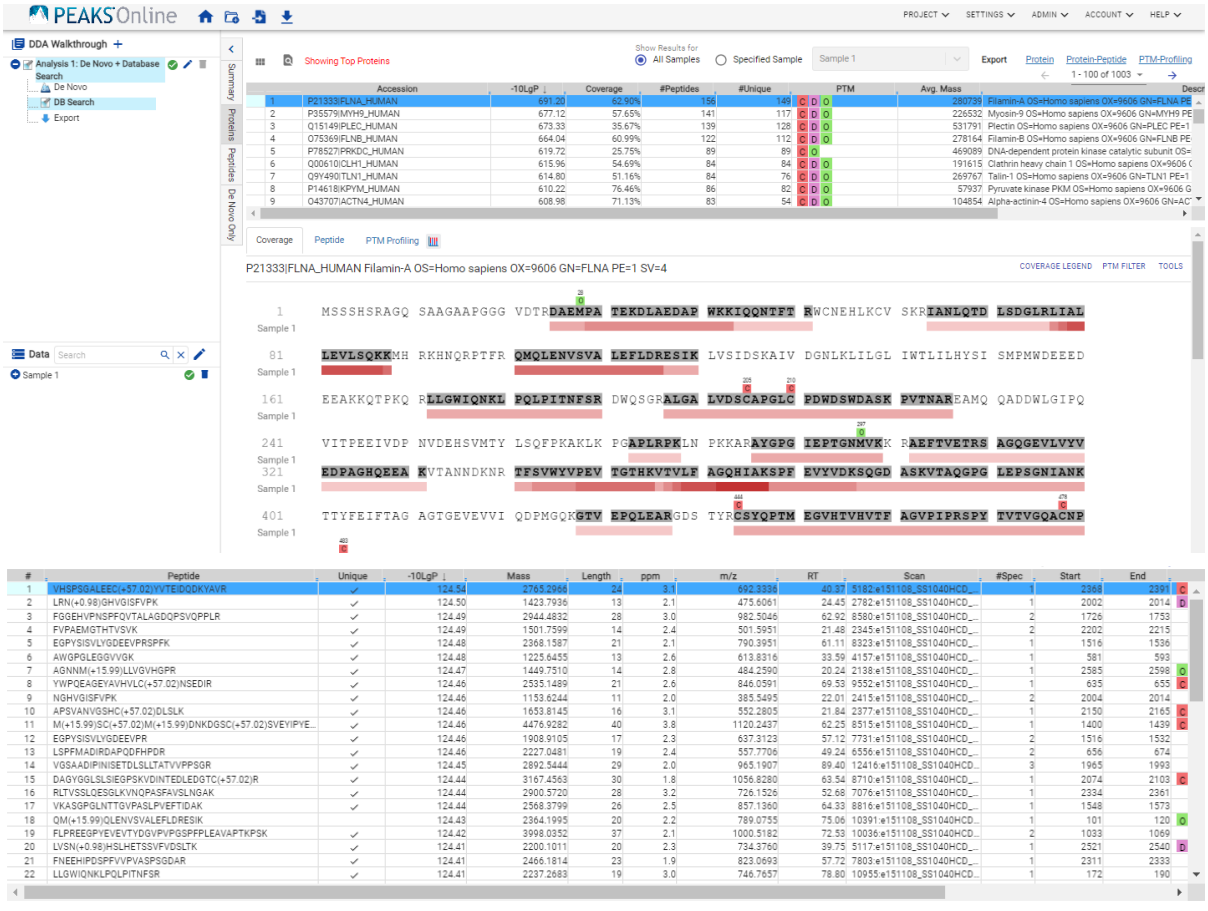
- **Coverage by Sample:** the coverage broken down by sample can be displayed if multiple samples were used. Only peptides that pass the filters are included in the percentage calculation.
- **Area by Sample:** total area of peptide features from unique supporting peptides in each sample are used for calculation. Peptides must pass the filters.
- **#Spec by Sample:** total number of spectra identified that support the given protein. Totals are given for each sample included in the search.

3.4.4. Protein Coverage

The Protein Coverage view visualizes the mapping of supporting peptides to the protein selected in the Protein table. It also shows all identified modification or mutation sites to assist with protein characterization at the amino acid level. The coverage view shows different views under different modes. The view shows a sample comparative coverage view in the all sample mode while it shows protein sequence coverage view in the specific sample mode.

3.4.4.1. Sample Comparative Coverage

This area visualizes the frequency of supporting peptide sequences from selected samples in the protein sequence. At the top of the protein sequence coverage view, the header information of the protein is shown. At the top-right corner, the COVERAGE LEGEND button will show a continuous color legend for the peptide frequency. Specific modifications can be selected in the PTM FILTER button and the template protein can be copied using the TOOLS button.

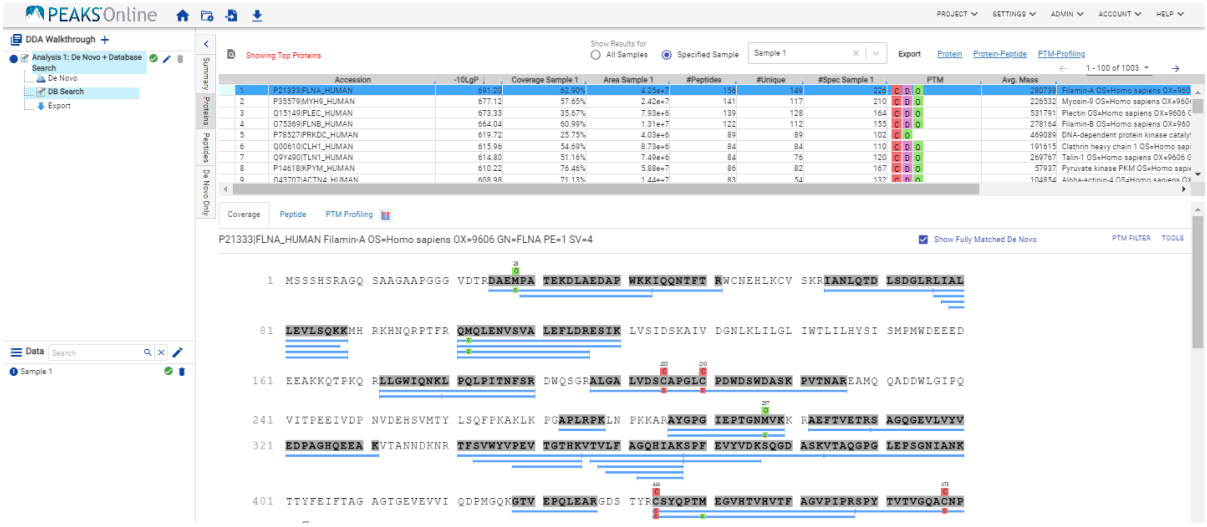


Regions in the protein sequence that are covered by supporting peptides are displayed in bold font with a grey background. Placing the cursor over a amino acid on the sequence shows the position of the amino acid in the sequence and the frequencies from different selected samples.

The frequency of supporting peptide sequences from different samples can be shown as colored bars under the protein sequence. The deeper the color, the higher frequency of that supporting peptide in the sample.

3.4.4.2. Protein Sequence Coverage

This area visualizes the coverage of the protein sequence in the specified sample mode. The header information of the protein, a PTM filter and a tool to copy the protein information are shown on the top of the protein sequence coverage view. Clicking the PTM filter button will pop up a menu to list the modifications identified in supporting peptides on the protein and the checkbox on the left controls whether to show the modification in the protein sequence coverage view above.



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

The de novo only peptides are the confident de novo sequence tags that remain unidentified by the identification search algorithms. To report a de novo only peptide, the De novo score must be equal to or better than the specified threshold. Meanwhile, the score of the spectrum's best identification search result should be no greater than the specified -10lgP threshold.If we enable the de novo only tag in the parameter setting, we can see the de novo only sequence with grey line below.

The supporting peptides 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 will pop up the annotated spectrum associated with the supporting peptide.

3.4.5. Supporting Peptides

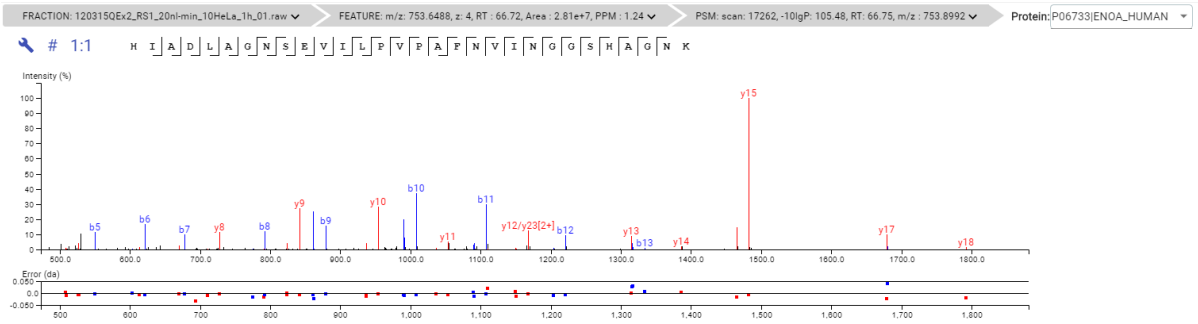
The Peptide table shows a list of supporting peptides for the selected protein. This table is similar to the peptide table in the Peptide View, except that "Accession" is excluded and the columns, "Unique", "Start", "End", and "Found by" are included.

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

Refer to Section 3.5.2, “Peptide Table” for the descriptions of other columns.

Double click a supporting peptide, it will go to that peptide sequence in the peptide page.

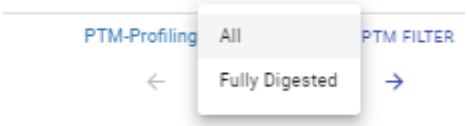


3.4.6. PTM Profiling

This tool calculates the difference in abundance between peptides with confidently identified PTMs versus unmodified peptides. Peptide feature areas are used for this comparison. For the Identification results (DB/PTM/SPIDER), the PTM Profiling tool provides a direct visualization of quantitative information for comparing modified peptides with unmodified peptides for the all modification sites in a protein across all MS samples.

3.4.6.1. PTM Profiling View Filters

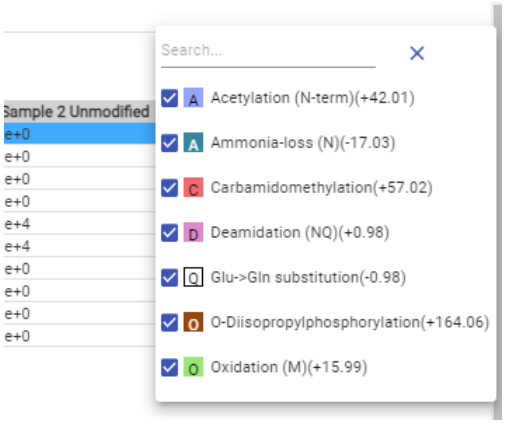
Peptide Filter: The PTM peptide filter controls what peptides contribute to modified and unmodified peptide areas. Users can select either All or Fully Digested Peptides, where fully digested peptides contain no missed cleavages.



PTM Filter: 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

Fixed PTMs are not used for PTM Profiling.

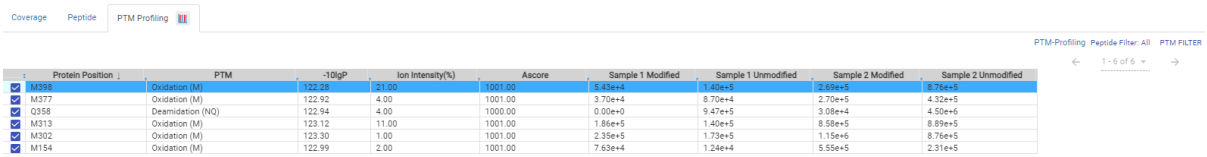


3.4.6.2. PTM Profile Table

The PTM Profile table displays a site-specific list of confident mutations on the selected protein. The table can be sorted by any column. The columns in this table are:

- **Protein Position:** Amino acid residue type and position # of the specified modification.
- **-10LgP:** Confidence score for the top PSM containing that modification that passes the confident site filters. Refer to Section 3.1, “Result Filtration” for detailed explanation.
- **Ion Intensity (%):** Ion intensity of the modification site-determining fragment ion pair, compared to the base peak ion intensity, for the top-scoring PSM that passes the confident site filters.
- **AScore:** The AScore achieved by the modified peptide with highest -10lgP score that passes the confident site filters.
- **Sample Area columns**: Cumulative area of modified and unmodified forms of the peptide on a per sample basis. Modified PSMs which do not pass the confident site filters will be considered as unmodified for area calculations.

This table is paginated showing 10 modification sites per page.

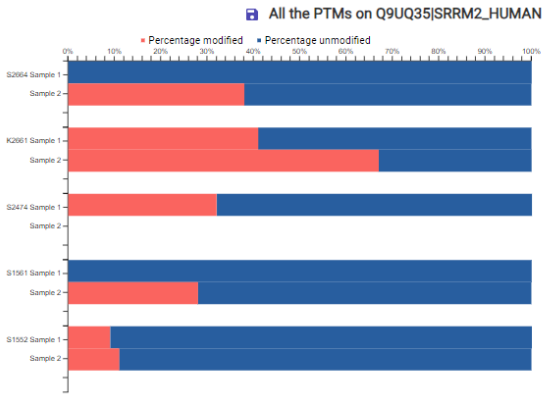


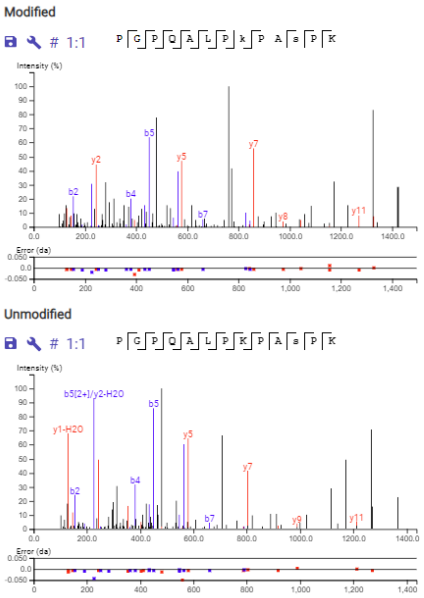
3.4.6.3. PTM Profile Figures

PTM Profile Chart: Below the PTM Profile table, a dynamic bar graph provides a visualization of the ratio between the modified vs. unmodified forms of the peptides detected at each modification site. . 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. By default, all modification sites listed in the PTM profiling table are displayed. Users may toggle which mutation sites are displayed by deselecting the check-box beside any modification site or using the PTM filter option.

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.

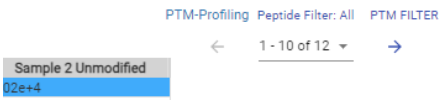




3.4.6.4. PTM Profile Export

The PTM profile data can be exported to a text file in CSV format for specific proteins or for all identified proteins (for all samples or single sample).

Single Protein: To export PTM profile information for a single protein, click the PTM-Profiling export link within the PTM Profiling tool. This will export all information displayed within the PTM profile table, meaning it will be filtered by the PTM profiling PTM and Peptide filters.

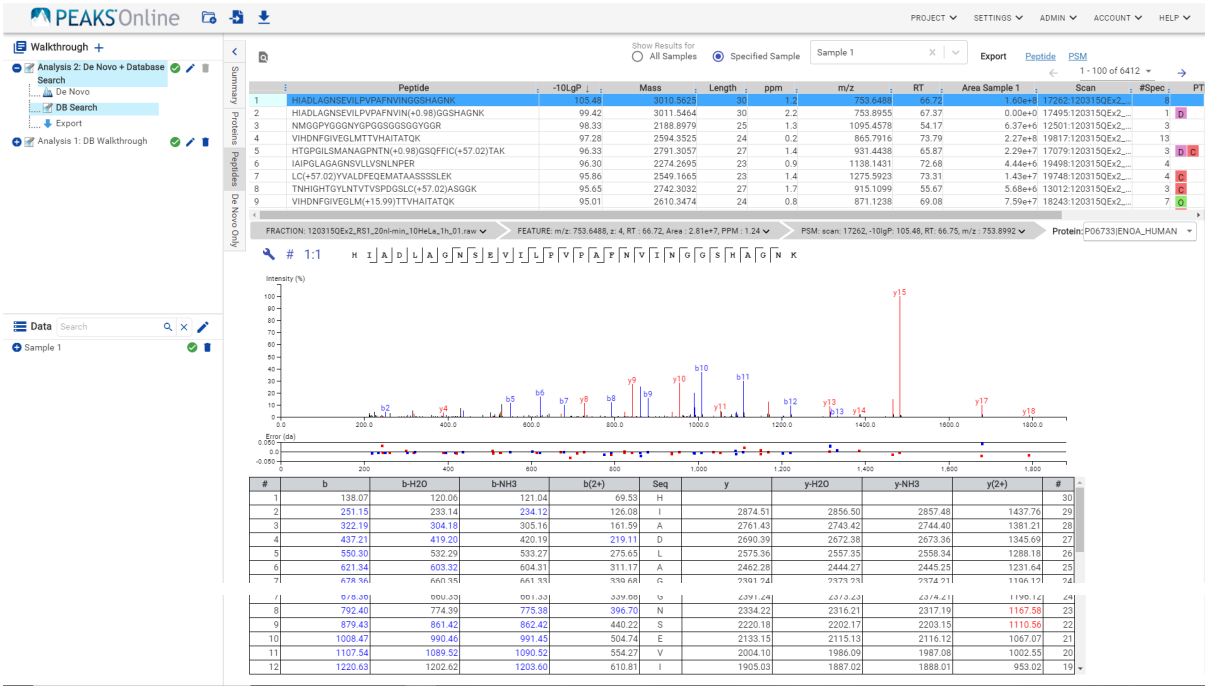


All Proteins: To export PTM profile information for all identified proteins click the PTM-Profiling export link above the Protein table. The exported file contains all necessary information to construct the PTM profile graph. The PTM profile data are grouped together by protein, then by modification type, and then by the modification site on the protein sequence. The headers of the CSV column are similar to the PTM Profile Table.

This export will adhere to all protein view filters applied to the protein table and can be filtered by specified or all samples.

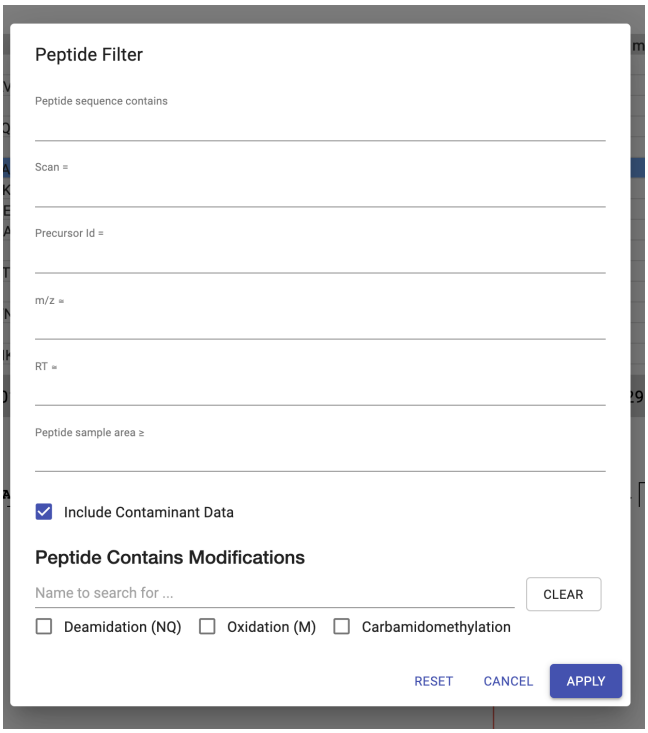
3.5. Peptide View

The Peptide Table shows the filtered peptide identification results. Click on a peptide row to show the spectrum below the peptide table.



3.5.1. Peptide View Filters

Click the button, the following Peptide Filter will pop up. It controls which peptides to display.



- **Peptide sequence contains:** Enter an amino acid sequence, only peptides that contain that sequence will be displayed in the peptide table.
- **Scan:** Filter for a single scan number based on the scan numbers read from the raw data file.

- **Precursor Id:** Filter is only relevant for TimsTOF data, for which it will filter based on the precursor id of the peptide.
- **m/z:** Filter based on the expected m/z, the last significant digit is rounded.
- **RT:** Filter based on the expected RT, the last significant digit is rounded.
- **Peptide sample area:** Filters based on total peptide area found in each sample. If a peptide is found in multiple samples, there should be at least one sample pass this filter for the peptide to be included.
- **Spider Mutations:** Filter only available for SPIDER result. When checked, peptides that have mutation identified with SPIDER are shown.
- **Peptide Contains Modifications:** Check boxes are provided for each identified modification. Peptides that contain the selected PTMs will be shown.

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 is sorted by the -10lgP value. In the Protein list drop-down window of each peptide, after clicking any of the protein, the protein page will be opened.

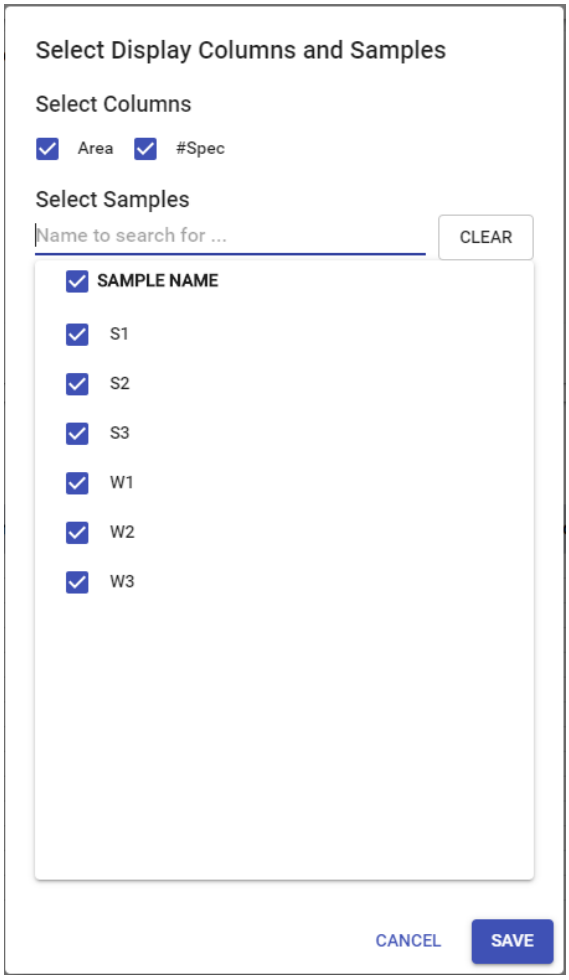
The columns Peptides, -10LgP, m/z, RT, Accession are sortable, i.e. Clicking any of these header can sort the whole table based on that column.

The columns in the Peptide table are:

- **Peptide:** The amino acid sequence of the peptide, as determined in PEAKS Search. A modified residue is followed by a pair of parentheses enclosing the modification mass.
- **-10lgP:** The peptide -10lgP score. The score indicates the scoring significance of a peptide-spectrum match.
- **Mass:** The monoisotopic mass of the peptide.
- **Length:** The number of amino acids in the peptide sequence.
- **ppm:** The precursor mass error, calculated as $10^6 \times$ (precursor mass - peptide mass) / peptide mass.
- **m/z:** The precursor mass-to-charge ratio.
- **RT:** The retention time (elution time) of the spectrum as recorded in the data.
- **Scan:** The scan number of the spectrum that matches the peptide sequence with the highest -10lgP.
- **#Spec:** The number of MS2 spectra assigned to the 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, Ascore.
- **Accession:** The accession number of the highest-scoring protein containing this peptide.
- **Precursor Id:** This column is only available for TimsTof data, it shows precursor's id from the raw data.
- **1/k0:** This column is only available for TimsTof data, it shows the ion mobility range of for the precursor.

3.5.3. Peptide Global Comparative View

"Peptide Global Comparative View" provides detailed comparisons among different samples regarding the identified peptides. To Show/Hide Columns, click on top-left of the peptide table. From the pop-up window, users can add and remove area and spectral count columns. In the peptide table, the selected columns will be displayed and can be exported. If "Specified Sample" is selected these columns are automatically displayed.



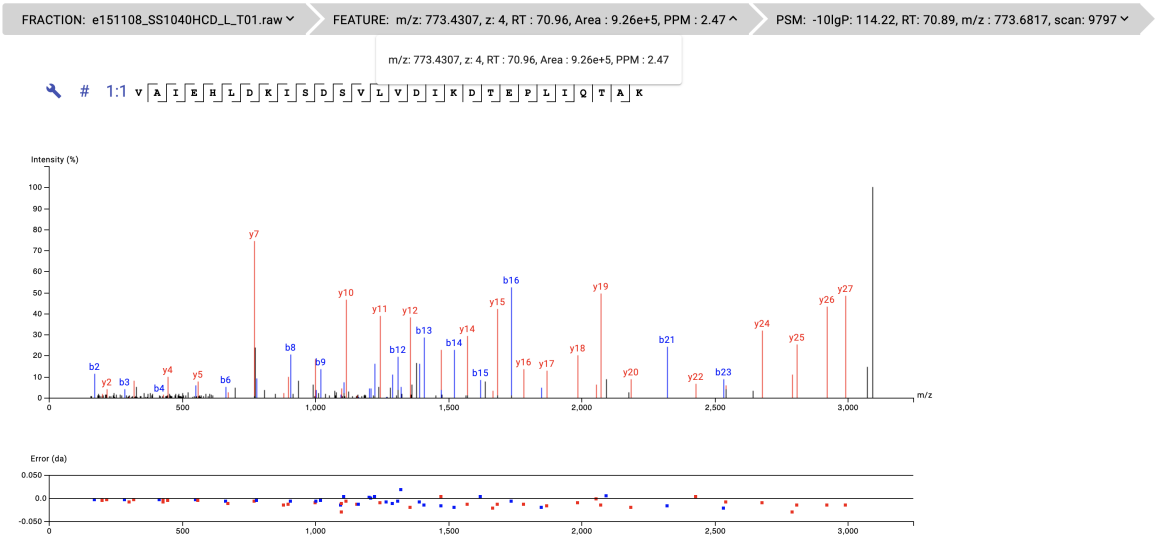
Note

PEAKS Online 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. The default can be changed by the administrator.

- **Area by sample:** The total area of all peptide features matched to spectra that identified the peptide. Area calculations are separated into multiple columns if multiple samples were included in the run.
- **#Peptide-spectrum matches by sample:** The number of peptide spectrum matches associated with a peptide separated by sample.

3.5.4. Peptide Menu

The peptide menu is only shown when "Show Results for Specified Sample" is selected in the filter section. The peptide menu displays 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 "FRACTION", “FEATURE” or “PSM” button to view the Spectrum Annotation of the other PSMs.



3.5.5. Peptide-Spectrum Matches

This section displays the spectrum annotation and other information about the peptide selected in the Peptide table. When multiple spectra are matched to the peptide, only the information about the top-scoring PSM is shown by default. The interface of this section is identical to the Spectrum Annotation in the PEAKS De Novo result. For detailed instructions, refer to Section 3.5, “Spectrum Annotation” in Chapter 5, Peptide De Novo Sequencing.

3.6. De Novo Only View

The De Novo Only view displays the list of quality peptide sequences detected by *de novo* sequencing that remain unidentified by the database search. 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 according to the current filters.

The table shows the filtered "de novo only" sequences by the current filters. "De novo only" sequences may suggest novel peptides, peptides with unknown modifications, or other interesting research subjects. The interface is identical to the "De Novo" view in a De Novo sequencing result page. For detailed instructions, refer to Section 3, “Understanding PEAKS *De Novo* Sequencing Results” in Chapter 5, Peptide De Novo Sequencing.In the setting parameter, the option "De novo only tags" is checked to show the locations of MS/MS spectra with de novo only sequences.

4. Running PEAKS PTM and SPIDER Separately

PEAKS PTM and SPIDER can be launched by going back to the WORKFLOW tab. Click the "MODIFY WORKFLOW" button and click Next to set parameters of PEAKS DB, PEAKS PTM, or SPIDER. They can also be launched by selecting the PARAMETERS tab and click "MODIFY PARAMETERS". Checking "Find unspecified PTMs with PEAKS PTM" and/or "Find more mutations with SPIDER" will launch PEAKS PTM and/or SPIDER.

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. For PTM, the search can be done either with all built-in modifications or with selected list of modifications.

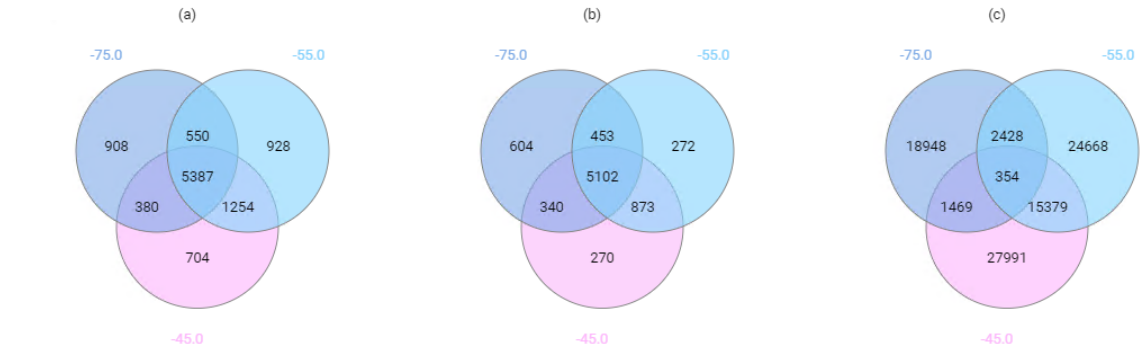
5. FAIMS Data

Field asymmetric waveform ion mobility spectrometry (**FAIMS**) analyzes complex mixtures and separates the ions according to their mobilities. PEAKS Online has the ability to load FAIMS data and analyze it using *de novo* sequencing, database search (PEAKS DB), unexpected modification (PEAKS PTM) and sequence variants (SPIDER) search, and labelled and label free quantification.

The analysis result for FAIMS dataset contains the additional information for CV values compared to the non-FAIMS dataset. The CV-related information can be found at the following places for a DB/PEM/Spider analysis.

A **Venn diagram** on the summary page of FAIMS projects.

Figure 3.FAIMS CV Venn diagram(up to 4 CVs). (a) All Proteins; (b) Top Proteins; (c) Peptides



- **Venn diagram** overlaps between different CVs in the search for all proteins, top proteins, and peptides are shown on the summary page.

With the introduction of FAIMS data, a CV column will appear on all peptide tables, as well as select exports:

- Additional **CV** column on the supporting peptide page

  The supporting peptide table on the protein page contains one additional column for CV values

- Additional **CV** column on the exported *Protein-Peptide CSV*

  The exported *Protein-Peptide CSV* contains one additional column for CV values

- Additional **CV** column on peptide page

  table on the peptide page contains one additional column for CV values

- Additional **CV** filter on peptide page

  The filter for peptide page contains one additional option for CV, which selects the target scan based on the CV value

- Additional **CV** column on the exported *Peptide CSV*

  The exported *Peptide CSV* contains one additional column for CV values

- Additional **CV** column on De Novo Only page

  The table on the De Novo Only page contains one additional column for CV values

- Additional **CV** filter on De Novo Only page

  The filter for De Novo Only page contains one additional option for CV, which selects the target scan based on the CV value

- Additional **CV** column on the exported *De Novo Only CSV*

  The exported *De Novo Only CSV* contains one additional column for CV values

- Additional **CV** column on the exported *De Novo Only with all candidates CSV*

  The exported *De Novo Only with all candidates CSV* contains one additional column for CV values

Chapter 7. DIA Streamlined Identification

1. Overview

PEAKS 11 offers a powerful and flexible solution for DIA identification using spectral library search with or without a protein inference database, direct database search, or a combination of both for optimal peptide coverage. In addition, de novo sequencing can be applied to ensure no good spectra remain unidentified.

2. Adding a Spectral Library from the Parameters Page

PEAKS Library Search offers an accurate and sensitive identification method for complex spectra, such as DIA spectra. In addition to annotated spectra along with precursor m/z and charge, PEAKS Library search also uses predicted retention time and ion- mobility collisional cross section (CCS) data for improved sensitivity.

To add a new Spectral Library from the Parameters page, open the "Select Library" drop down and select "**Create New Library**" or "**Import New Library**". See Chapter 2, section 2, *Library Management and Configuration Section 2, "Library Management and Configuration"* for details about creating and importing spectral libraries.

3. Spectral Library Search Workflow and Parameter Settings

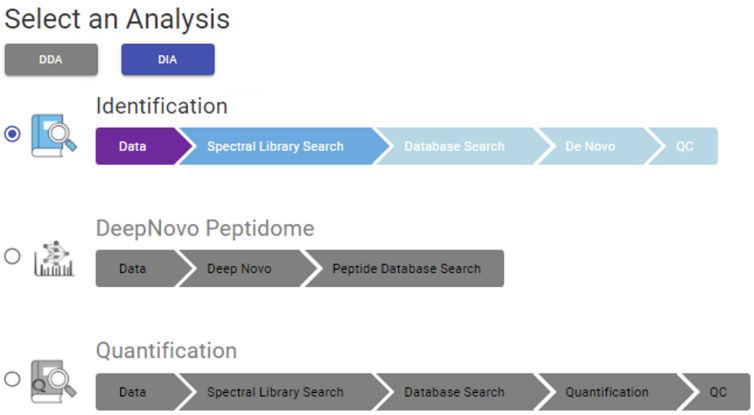
PEAKS Spectral Library Search can be launched as a workflow in a PEAKS Online search for DIA data (selected from the DIA section of the workflow selection during analysis creation). An analysis report is generated to show the search result.

To conduct a spectral library search workflow, follow the steps below:

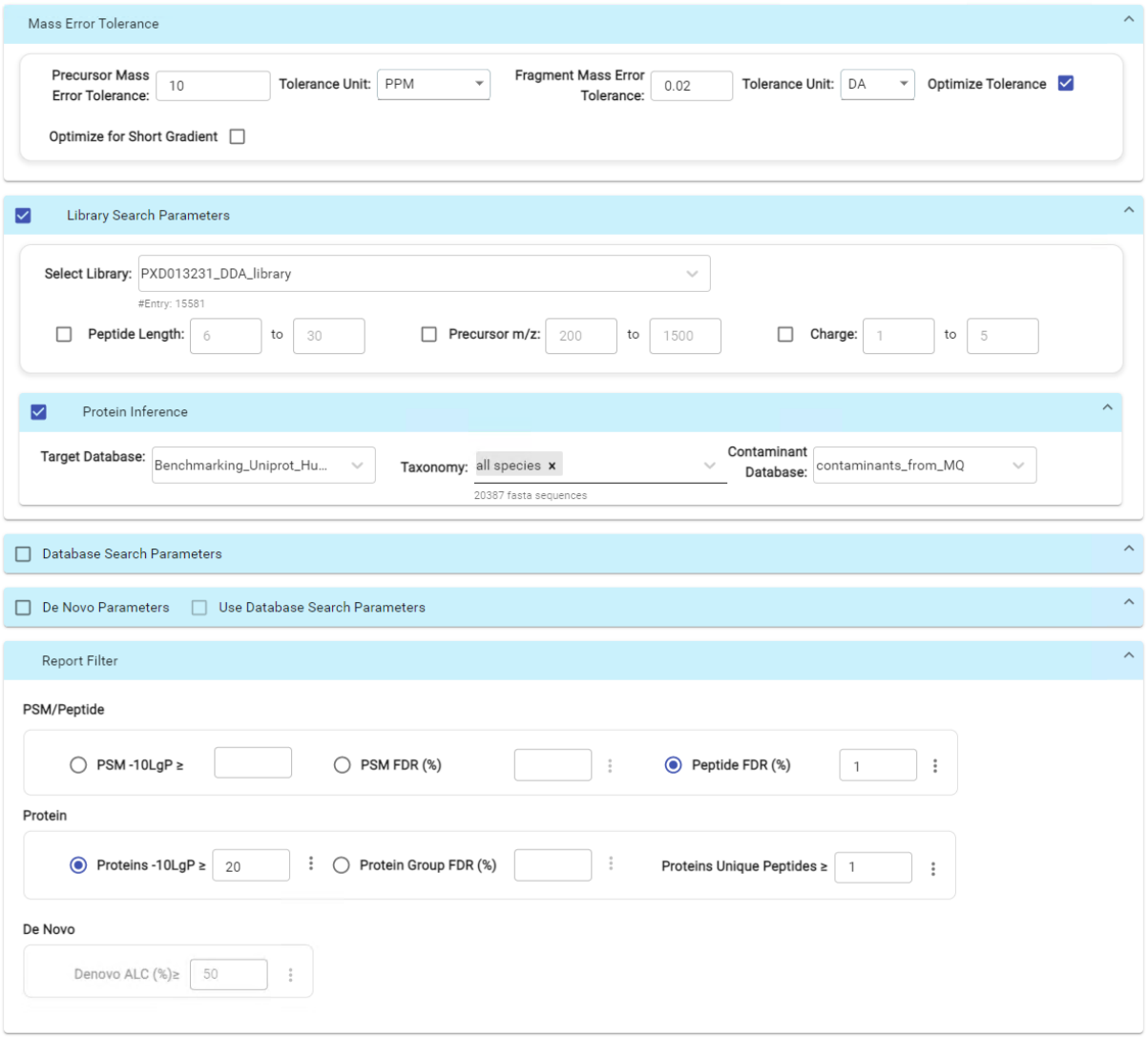
1. Refer to Chapter 5, Peptide De Novo Sequencing, Section 2, "*DeepNovo/De Novo* Analysis Workflow and Parameter Settings", Step 1-6 to create a new project and add data.

2. Select the "DIA" tab at the top of the Workflow step.

3. Select the "Identification" workflow.



4. Set the Spectral Library Search Parameters. See details in Section 3.1-3.4 below.



5. Click SUBMIT button to start the search.

3.1. Precursor and Fragment Error Tolerance

Precursor mass: This enables the selection of a precursor mass error tolerance. PEAKS Online DB Search identifies peptides within the precursor mass tolerance, which is the allowable m/z shift between the theoretical value of the peptide versus the observed value of the precursor ion. Precursor mass tolerance can be specified in either Da or ppm.

Fragment ion: This enables the selection of a fragment ion m/z tolerance, which is the allowable m/z shift between the theoretical value of the fragment ions of the peptide versus the observed value. PEAKS Online DB Search uses this value when scoring peptide-spectrum matches. PEAKS Online considers that a fragment ion is matched if the calculated m/z is within the tolerance. Fragment mass tolerance can be specified in Da.

Error tolerance parameters need to be set consistently with the mass accuracy of the instrument.

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.

3.1.1. Optimize Tolerance

When enabled, PEAKS Online will determine the optimal values to use for Precursor and Fragment Mass Error Tolerance for each fraction in your analysis. These values will replace the values provided in the previous Precursor and Fragment Mass Tolerance sections.

3.1.2. Optimize for Short Gradient

This selection will tune parameters specifically for short gradient runs.

3.2. Spectral Library

Library Selection.: Select a PEAKS-generated library for the search. Select one from the list of libraries that have been configured in PEAKS Online and filter for the peptide length, precursor m/z, and charge when necessary. To configure a new library, refer to Chapter 2 Section 2: Library Management and Configuration.

Library Filters.: The spectra that will be used from the library can be filtered based on peptide length, precursor m/z, and charge. This is optional. Each filter can be enabled by checking the checkbox to the left of the filter inputs.

3.3. Protein Inference Settings

Protein inference can optionally be enabled to search peptides found in the selected library against a provided database.

Target Database.: Select a reference database from the list of configured databases or create a new one, and set the taxonomy if applicable. This database is the database used for protein inference.

Contaminant Database.: A contaminant database can optionally be provided as well. #contam# will be added to the beginning of the accession of all identifications from the contaminant database.

3.4. Report Filter

The report filter section is a mandatory section to filter out some results under the user-defined filter. Any changes to the filter will create a new analysis with old parameters and update the results and statistics.

- **PSM -10lgP:** The PEAKS peptide score (-10lgP) is calculated for every peptide-spectrum match (PSM) reported by PEAKS Spectral Library Search. The score is derived from the p-value that indicates the likelihood that the identification was made by random chance alone.
- **PSM FDR (%):** The PSM filter can also be set to use the false discovery rate. FDR is estimated using a "decoy-fusion" approach. A decoy peptide/spectrum for each target peptide/spectrum in a spectral library and the target and decoy spectra are searched together. The decoy method is based on: https://doi.org/10.1021/pr900947u
- **Peptide FDR (%):** The filter can also be set to use the false discovery rate of the peptides instead of the PSMs. Redundant peptide hits found in multiple spectra are counted once.

4. Understanding PEAKS Spectral Library Results

After a PEAKS Spectral Library Search is complete, the result node will appear within the Analysis tree in the Analyses section of the left menu.

The analysis results for database search tools are presented in three pages:

- **Summary:** This shows an outline of the PEAKS Online spectral library search results with key statistics. The overall quality of the experiment can be examined and the filters for protein identifications can be adjusted. This page shows after the search is done.

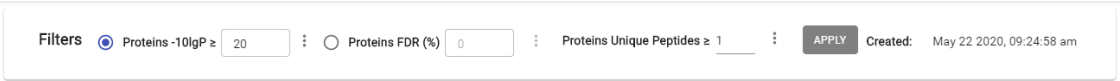
- **Protein:** This shows a list of protein identifications. This view also visualizes protein sequence coverage and helps with protein characterization.

  > **Note**
  >
  > This tab only appears if Protein Inference was enabled for the PEAKS Spectral Library Search.

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

4.1. Result Filtration

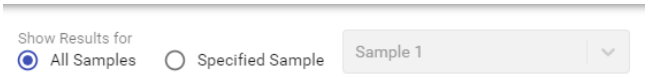
At the top of the summary result page, filters can be set. Identifications that fall below the filters will not be visible in the result pages or exports. After changing any of these filter parameters, click the APPLY button to apply and save the new filters. The results and statistical information will be updated accordingly.



- **Proteins -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. A default threshold of 20 is set.
- **Proteins FDR (%):** Optionally, the proteins false discovery rate can be selected to filter proteins instead of a -10lgP score.
- **Proteins Unique Peptides:** The minimum number of unique supporting peptides for a protein identification. A unique peptide is defined as a peptide that can be mapped to only one protein group. Proteins are grouped using parsimony.

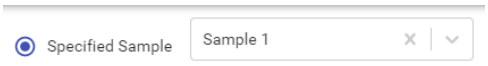
4.1.1. Show Results for All Samples

For the protein and the peptide tables, "All Samples" can be selected to view the combined search results from all the samples. In the Protein page, selecting "All Samples" will change the coverage panel displayed when a specific protein is highlighted. Supporting peptides from all samples instead of just a single sample will be shown.



4.1.2. Show Results for Specific Sample

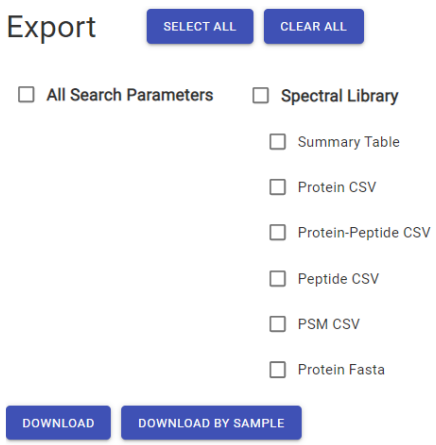
In the Protein, Peptide and De Novo page results, the filter can also be set to "Specified Sample" which will only show the results a single sample. By clicking in the text you can search for a specific sample by name.



Selecting "Specified Sample" will change the protein coverage view that is visible by clicking on a protein in the protein table. Peptides in the Protein, Peptide and De Novo Only pages will only be visible if they belong to the sample chosen. Once selected, the peptide tables will automatically add the "Area" column and the "# Spec" values will display the spectral count for the specified sample. For more information, refer to Section 3.5.2, “Peptide Table”. Similarly, the protein table will also include the columns "Coverage" and "Area" for the specified sample. For more information refer to Section 3.4.2, “Protein Table”.

4.2. Result Exporting

PEAKS Online spectral library search results can be exported in the Export page:



- **Summary Table:** A comma separated file containing the information from the Summary Statistics table on the Summary page.
- **Protein CSV:** A comma separated file containing all identified proteins and associated details from the protein table for all samples.
- **Protein-Peptide CSV:** A comma separated file containing all identified peptides and associated details from the peptide table found in the Protein page. Further details are provided regarding the protein the peptide supports. If a peptide supports multiple proteins, multiple rows are included for each supported protein for all samples.
- **Peptide CSV:** A comma separated file containing all identified peptides and associated details from the peptide table for all samples.
- **PSM CSV:** A comma separated file containing all identified peptide spectrum matches and associated details from the peptide table for all samples.
- **Protein Fasta:** All identified proteins and associated details from the protein table for all samples, formatted as a fasta file.

4.3. Summary View

4.3.1. Summary Statistics

The first chart displays the summary information for each sample.

- **Sample Name:** This column lists the samples names with the exception of the first row, which shows the combination of all the samples.
- **# MS Runs:** The total MS runs in the sample.
- **# MS1:** The total MS scans in the sample summed across all fractions.
- **# MS2:** The total MS/MS scans in the sample summed across all fractions.
- **# PSM:** Total number of peptide spectrum matches (PSMs) found in the sample.
- **# Peptides:** Total number of peptides found. Peptides with the same primary sequence but different PTMs are counted separately.
- **# Sequences:** Total number of different peptide backbone sequences found. Peptides with the same primary sequence but different PTMs are counted as one sequence.
- **# Proteins:** Provides the number of proteins given the parameters set for Protein Score (-10lgP) and number of unique peptides.
- **# Protein Groups:** PEAKS Online groups proteins identified by a common set of peptides. This number shows the number of protein groups in the filtered result based on All proteins.

- **# PSM / # MS2:** The percentage of MS2 scans in the sample that produced peptide spectrum matches.

4.3.2. False discovery rate (FDR) curve

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

PEAKS Spectral library search estimates the FDR using decoy peptides and spectra. A decoy peptide/spectrum for each target peptide/spectrum in a spectral library and the target and decoy spectra are searched together. The decoy method is based on: https://doi.org/10.1021/pr900947u

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

Note

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

Note

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

4.3.3. Statistics of filtered results

Table 1 shows the statistical numbers of the filtered results.

- **Top Proteins:** The number of proteins that 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 included here.
- **Proteins (#Unique Peptides):** In Table 2, 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. Peptides are counted based on their backbone sequence while considering I/L amino acids to be the same.
- **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.
- **-10lgP Cutoff:** The value is determined by the estimated PSM FDR. Only the peptides with PSMs above the -10lgP score threshold are listed in the Peptide view.

4.3.4. Precusor mass error

Figures 2(a) and 2(b) show the precursor mass errors of PSMs in filtered results. Figure 2(a) shows the distribution of the precursor mass error in a histogram with and without software calibration. Software calibration uses the observed distribution of mass error after identification to normalize the mass error around zero. Figure 2(b) shows precursor m/z versus precursor mass error in a scatter plot after software calibration. Figures 2(a) and 2(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.

4.3.5. RT Calibration

Plots the RT regression function used during the spectral library search. Peptides are plotted against their indexed retention time (iRT) predicted using machine learning.

4.3.6. Box plot for Rt change

A box plot showing the difference between the observed retention time and iRT peptides within set retention time bins across all fractions, showing the accuracy of retention time prediction.

4.3.7. Number of identified peptides in each sample by number of missed cleavages

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

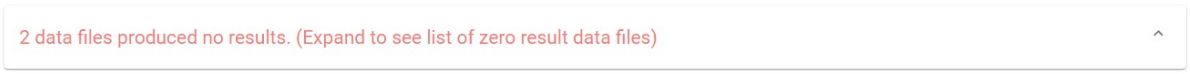
4.3.8. PTM Profile

Table 3 shows modifications identified in the filtered result and the number of PSMs containing each modification.

- **Name:** The given name of the PTM.
- **Δ Mass:** The monoisotopic change in mass caused by the PTM.
- **Position:** The amino acids modified by the PTM.
- **# PSM:** The number of PSMs with the given PTM.
- **-10lgP:** The highest -10lgP score achieved by a peptide modified by the PTM.
- **Area:** The area achieved by the modified peptide with highest -10lgP score.

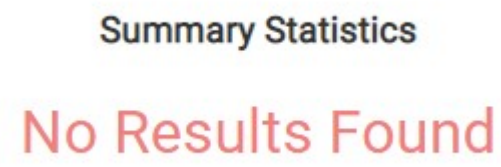
4.3.9. Low Quality Data Warnings

If PEAKS Online detects any data files producing very few or no results a warning section will appear at the top of the Summary Page indicating the number of data files producing poor or no results.



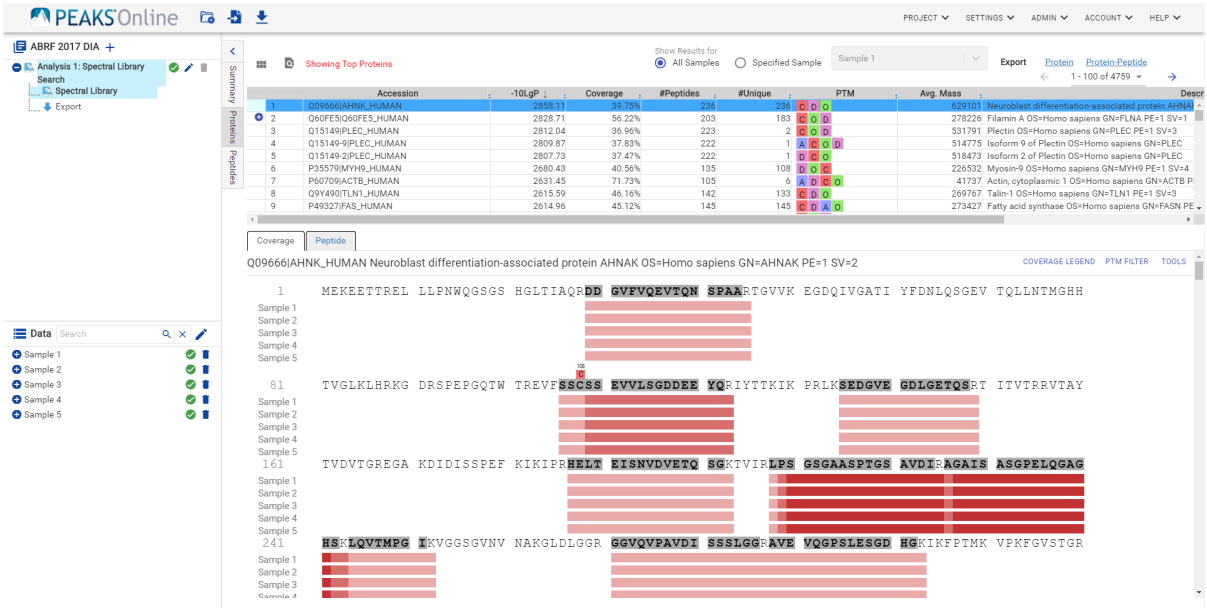
This section can be expanded to display a list of data files in the analysis that are producing poor or no results.

If there were no results found in any of your data files, a warning message stating "No Results Found" will be displayed instead of the usual Summary Page figures and tables.



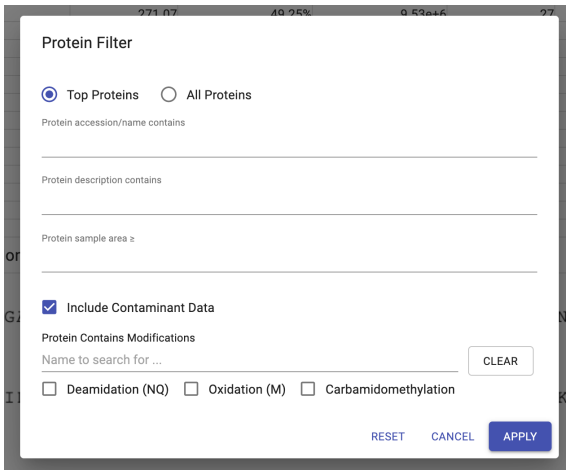
4.4. Protein View

The Protein view lists protein identifications that have been filtered by the current settings set for the project. It also visualizes the protein sequence coverage for identified proteins.



4.4.1. Protein View Filters

By selecting the button, the Protein Filter will pop up. It controls which proteins are displayed.



- **Show Top or All 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.
- **Protein accession/name contains:** If you add text here, only proteins with accessions that contain the text will be included in the protein table.
- **Protein description contains:** If you add text here, only proteins with descriptions that contain the text will be included in the protein table.

- **Protein sample area:** Filters 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.

  > **Note**
  >
  > Only numeric digits can be typed in the protein sample area. For example, instead of 1.00e+4, 10000 must be typed.

- **Spider Mutations:** This filter is only visible in SPIDER result view filter. Check it to view proteins that have mutations.
- **Protein Contains Modifications**: Proteins that are supported by peptides containing selected PTMs will be included.

4.4.2. Protein Table

Each row in the table represents a group of proteins that are supported by a common set of peptides. A ⊕ indicates that the group has multiple proteins. To expand the group, click that button.

The columns in the Protein table are:

- **Accession:** The accession id of the protein as seen in the FASTA database.
- **-10lgP:** The protein confidence score. Refer to Section 3.1, “Result Filtration” for a detailed explanation.
- **Coverage:** The percentage of the protein sequence that is covered by supporting peptides. This is the total coverage including results from all samples, unless results are only shown for a specified sample.
- **#Peptides:** The number of supporting peptides which pass the given peptide filters.
- **#Unique:** The number of supporting peptides which pass the given peptide filters that are mapped to only one protein group. Unique peptides with the 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.

> **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.

4.4.3. Global Protein Comparative View

Global Protein Comparative View provides detailed comparisons among different samples regarding the identified protein. To Show/Hide Columns, click ▦ on top-left of the protein table. From the dialog, users can add and remove sample coverage, area, and spectral count columns. In the protein table, the selected columns will be displayed and can be exported.

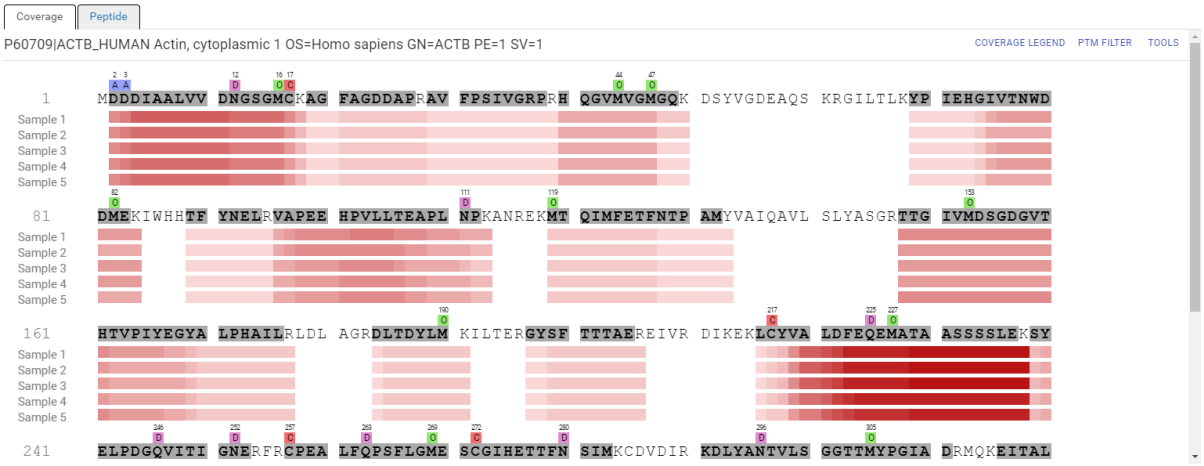
- **Coverage by Sample:** the coverage broken down by sample can be displayed if multiple samples were used. Only peptides that pass the filters are included in the percentage calculation.
- **Area by Sample:** the total area of peptide features from unique supporting peptides in each sample. Areas from peptides that don't pass the peptide filter won't be included in the total area.
- **#Spec by Sample:** total number of identified peptide spectrum matches which pass the peptide filters that support the given protein. Totals are given for each sample included in the search.

4.4.4. Protein Coverage

The Protein Coverage view visualizes the mapping of supporting peptides to the protein selected in the Protein table. It also shows all identified modification or mutation sites to assist with protein characterization at the amino acid level. The coverage view shows different views under different modes. The view shows a sample comparative coverage view in the all sample mode while it shows protein sequence coverage view in the specific sample mode.

4.4.4.1. Sample Comparative Coverage

This area visualizes the frequency of supporting peptide sequences from selected samples in the protein sequence. At the top of the protein sequence coverage view, the header information of the protein is shown. At the top-right corner, the COVERAGE LEGEND button will show a continuous color legend for the peptide frequency. Specific modifications can be selected in the PTM FILTER button and the template protein can be copied using the TOOLS button.



Regions in the protein sequence that are covered by supporting peptides are displayed in bold font with a grey background. Placing the cursor over a amino acid on the sequence shows the position of the amino acid in the protein sequence and the frequency of identification in each selected sample.

The frequency of supporting peptide sequences from different samples can be shown as colored bars under the protein sequence. The deeper the color, the higher frequency of that supporting peptide in the samples.

4.4.4.2. Protein Sequence Coverage

This area visualizes the coverage of the protein sequence in the specified sample. The header information of the protein, a PTM filter and a tool to copy the protein information are shown on the top of the protein sequence coverage view. Clicking the PTM filter button will pop up a menu to list the modifications identified in supporting peptides on the protein and the checkbox on the left controls whether to show the modification in the protein sequence coverage view above.

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

The supporting peptides 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 will pop up the annotated spectrum associated with the supporting peptide.

4.4.5. Supporting Peptides

The Peptide table shows a list of supporting peptides for the selected protein. This table is similar to the peptide table in the Peptide View.

- **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.
- **Found by**: The name of the PEAKS search algorithm that identified the peptide. This can be either PEAKS, PEAKS PTM, or SPIDER.

Refer to Section 3.5.2, “Peptide Table” for the descriptions of other columns.

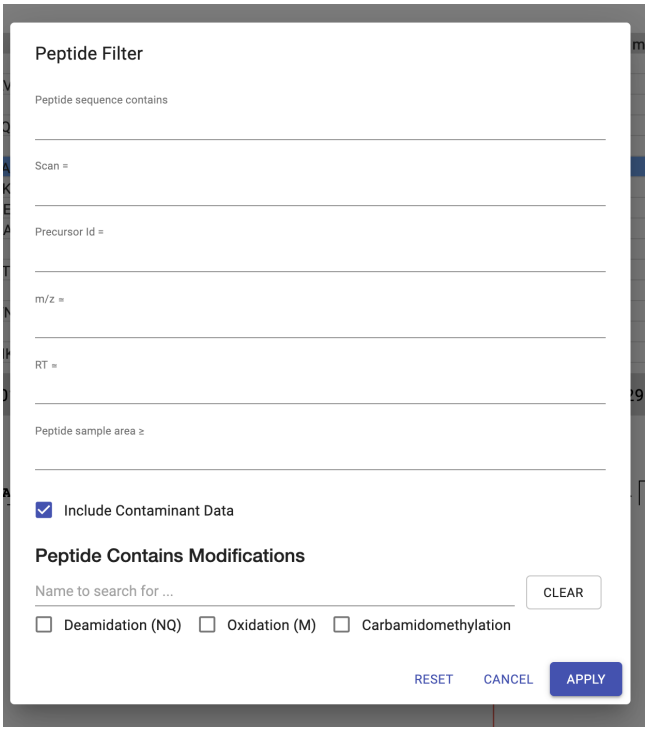
Double-click a row to navigate to the peptide in the peptide tab for more detailed information.

4.5. Peptide View

The Peptide Table shows the filtered peptide identification results. Click on a peptide row to show the spectrum below the peptide table. Information such as the scan number, retention time, isolation window and fragment error tolerance will be shown above the MS2 spectrum.

4.5.1. Peptide View Filters

Click the button, the following Peptide Filter will pop up. It controls which peptides to display.



- **Peptide sequence contains:** Enter an amino acid sequence, only peptides that contain that sequence will be displayed in the peptide table.
- **Scan:** Filter for a single scan number based on the scan numbers read from the raw data file.
- **Precursor Id:** Filter is only relevant for TimsTOF data, for which it will filter based on the precursor id of the peptide.
- **m/z:** Filter based on the expected m/z, the last significant digit is rounded.
- **RT:** Filter based on the expected RT, the last significant digit is rounded.
- **Peptide sample area:** Filters based on total peptide area found in each sample. If a peptide is found in multiple samples, there should be at least one sample pass this filter for the peptide to be included.
- **Spider Mutations:** Filter only available for SPIDER results. When checked, peptides that have an identified mutation using SPIDER are shown.
- **Peptide Contains Modifications:** Check boxes are provided for each identified modification. Peptides that contain the selected PTMs will be shown.

4.5.2. 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 is sorted by the -10lgP value.

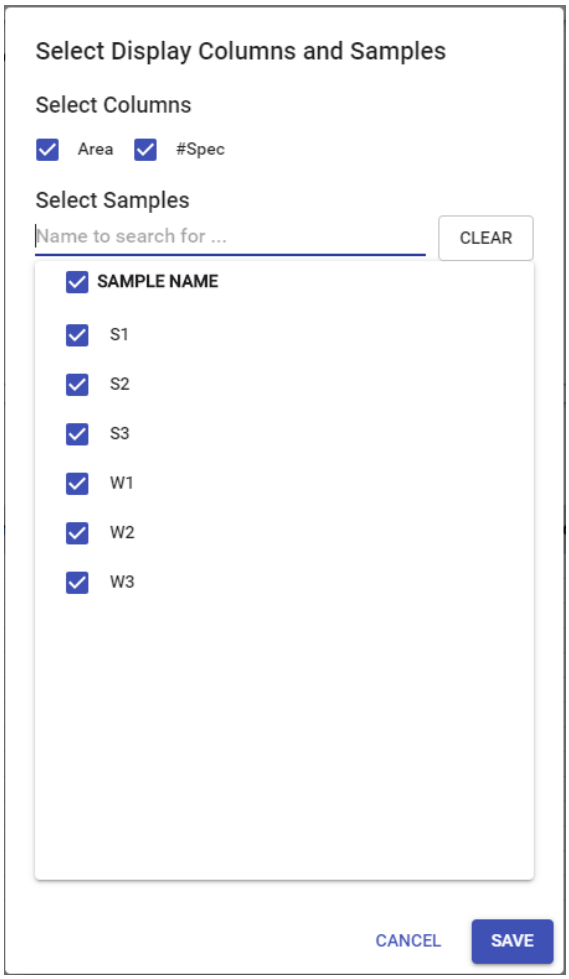
The columns in the Peptide table are:

- **Peptide:** The amino acid sequence of the peptide, as determined in PEAKS Search. A modified residue is followed by a pair of parentheses enclosing the modification mass.
- **-10lgP:** The peptide -10lgP score. The score indicates the scoring significance of a peptide-spectrum match.
- **Mass:** The monoisotopic mass of the peptide.
- **Length:** The number of amino acids in the peptide sequence.
- **ppm:** The precursor mass error, calculated as 10^6 × (precursor mass - peptide mass) / peptide mass.

- **m/z:** The precursor mass-to-charge ratio.
- **RT:** The retention time (elution time) of the spectrum as recorded in the data.
- **#RT:** Retention time difference between library and query spectrum.
- **Scan:** The scan number of the spectrum that matches the peptide sequence with the highest -10lgP.
- **#Spec:** The number of MS2 spectra assigned to the peptide.
- **PTM:** The types and the numbers of modifications present in the peptide shown in color-coded icons.
- **Accession:** The accession number of the highest-scoring protein containing this peptide.
- **Precursor Id:** This column is only available for TimsTof data, it shows precursor's id from the raw data.
- **1/k0:** This column is only available for TimsTof data, it shows the ion mobility range of the precursor.

4.5.3. Peptide Global Comparative View

"Peptide Global Comparative View" provides detailed comparisons among different samples regarding the identified peptides. To Show/Hide Columns, click on top-left of the peptide table. From the pop-up window, users can add and remove area and spectral count columns. In the peptide table, the selected columns will be displayed and can be exported. If "Specified Sample" is selected these columns are automatically displayed.



Note

PEAKS Online 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. The default can be changed by the administrator.

- **Area by sample:** The total area of all peptide features matched to spectra that identified the peptide. Area calculations are separated into multiple columns if multiple samples were included in the run.
- **#Peptide-spectrum matches by sample:** The number of peptide spectrum matches associated with a peptide separated by sample.

4.5.4. Peptide Menu

The peptide menu is only shown when "Show Results for Specified Sample" is selected in the filter section. The peptide menu displays 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 "FRACTION", “FEATURE” or “PSM” button to view the Spectrum Annotation of the other PSMs.

4.5.5. Protein Selection

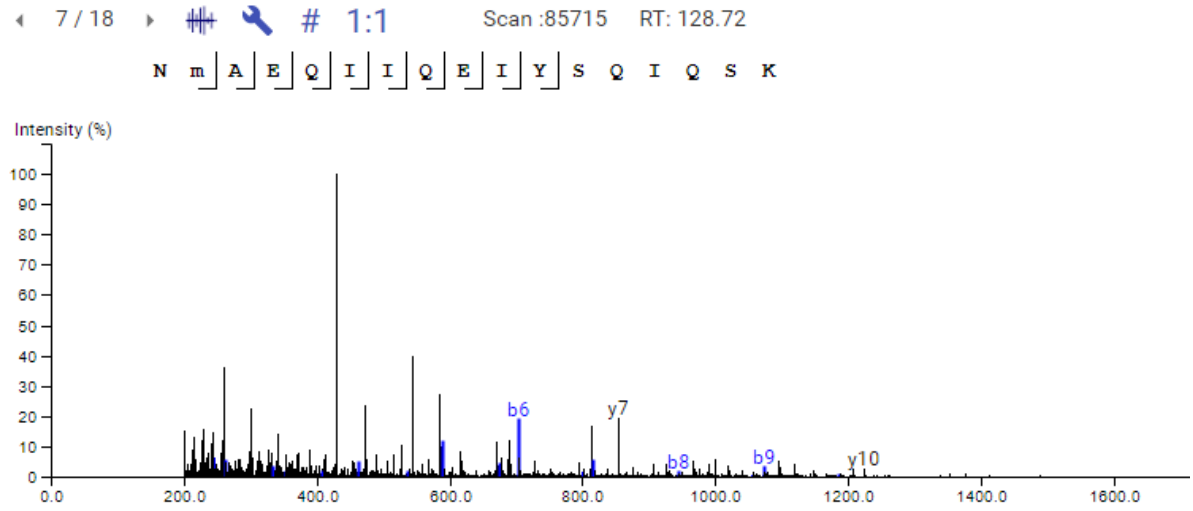
The protein drop-down in the top-right of the peptide information pane allows for quick navigation to any of the proteins the currently selected peptide supports.

4.5.6. Peptide-Spectrum Matches

This section displays the spectrum annotation and other information about the peptide selected in the Peptide table. When multiple spectra are matched to the peptide, only the information about the top-scoring PSM is shown by default.

4.5.6.1. Spectrum Annotation

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

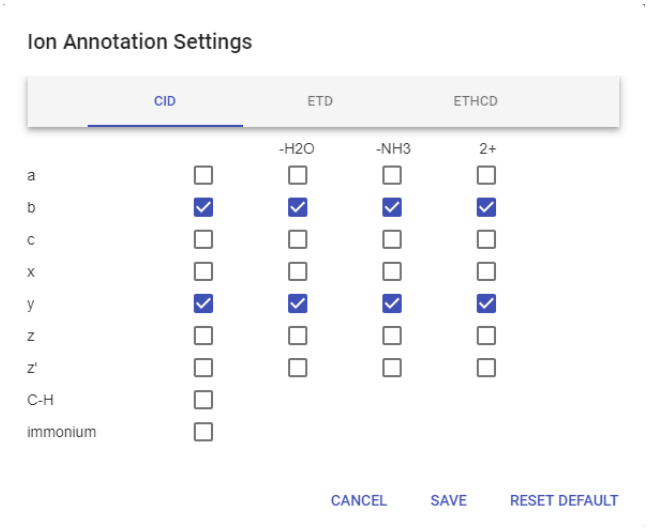


Moving the cursor over the peptide sequence in the spectrum will show the mass transitions for a particular amino acid residue.

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

- Zoom in/out on m/z: Place the cursor on the point in the spectrum you would like to zoom in on, or, place the cursor in the Error (da) figure and scroll the mouse wheel.
- Increase/Decrease peak intensity: Place the cursor on the Intensity axis and scroll the mouse wheel.
- See the whole spectrum: click the 1:1 button.

Click to open the Ion Annotation Settings dialogue and change the ions to be annotated in the spectrum. To reset the settings to the PEAKS Online defaults, use "RESET DEFAULT" button.

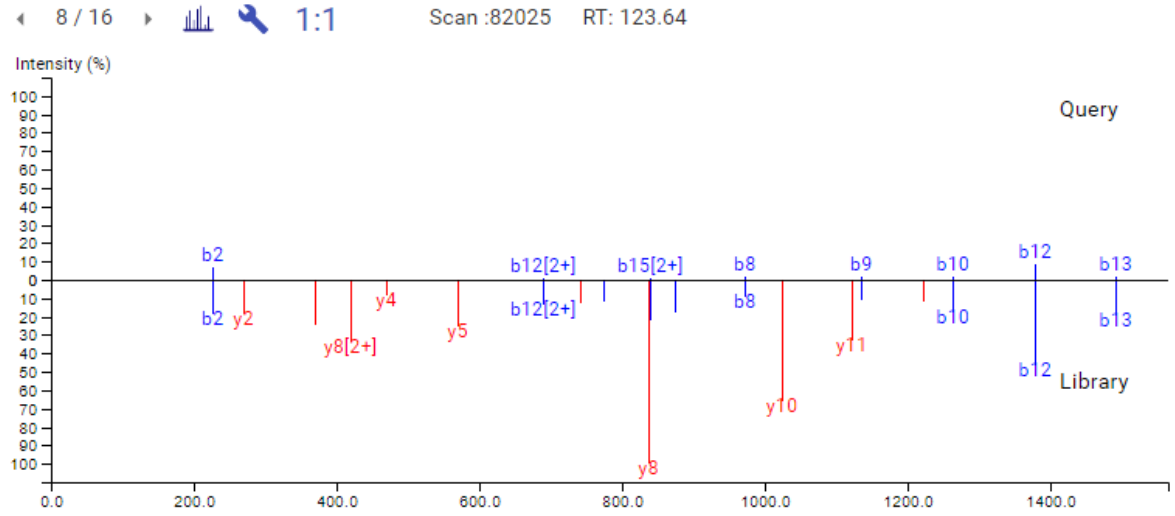


The % button will switch the y-axis to relative intensity. Once selected, scrolling over peaks will display the relative intensity instead of the absolute intensity.

The # button will switch the y-axis to absolute intensity. Once selected, scrolling over peaks will display the absolute intensity instead of the relative intensity.

4.5.6.2. Mirror Plot

You can press the button to swap the spectrum view to the mirror plot view.



This view shows the ions from the query spectrum compared to the spectrum it was matched to from the library.

4.5.6.3. Error Map

The "Error Map" shows the mass errors of the 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 "Ion Annotation Settings".

4.5.6.4. Ion Table

The "Ion Table" shows the calculated mass of possible fragment ions based on the ion types selected in the Ion Annotation Settings. If a fragment ion is found in the spectrum, its mass value is displayed in color. N-terminal ions are shown in blue and C-terminal ions are shown in red. A fragment ion is found when there is a matching peak within the mass error tolerance, as defined in the *de novo* sequencing parameters.

4.5.6.5. Precursor Profile

An eXtracted Ion Chromatogram (XIC) chart that displays the shape of the selected peptides precursor.

4.5.6.6. Fragment Ion XIC

An eXtracted Ion Chromatogram (XIC) chart that displays the shape of each identified fragment ion across all scans for the selected PSM. Below the XIC charts is a slider to adjust the correlation threshold of the fragment ion with the selected peptide precursor. Increasing the slider value will filter out fragment ions that are below the threshold.

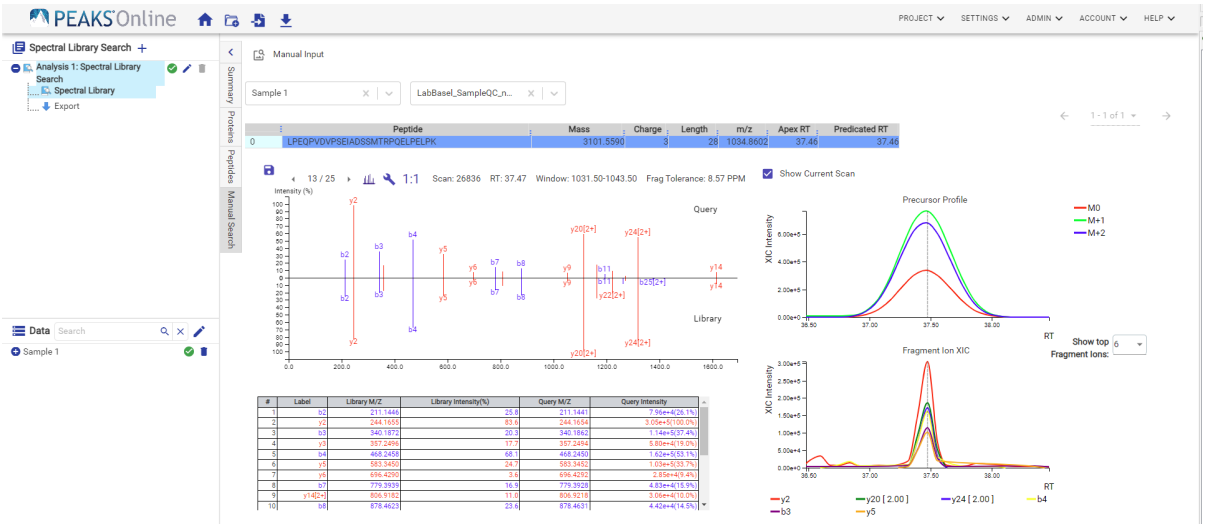
4.5.6.7. TimsTOF In Frame Mode

For timsTOF data only, there is an additional "In Frame Mode" button,  , above the spectrum view. Clicking this button will change the annotated spectrum view from "frame mode", the default view, to "in frame mode".

To switch back click the "Frame Mode" button,  . The standard, "Frame Mode" displays the scans within a frame merged over the CCS dimension. The "In Frame Mode" displays the scans merged over the retention time dimension.

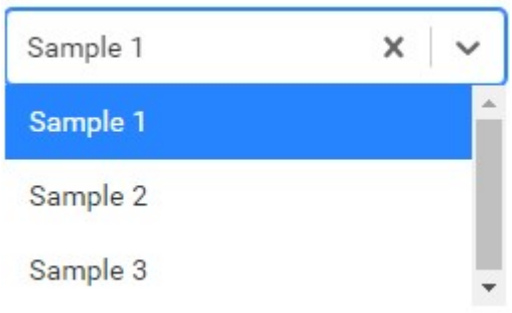
4.6. Manual Search

The Manual Search tab can be used to manually query your data for a specific library entry. If a peptide is found, PEAKS will display its peptide information and spectrum with scan, RT, isolation window and fragment tolerance information.



4.6.1. Performing a Manual Search

1. Select the sample you wish to search in from the Sample Dropdown, this should be the first dropdown displayed.



2. If your data is fractionated, select the fraction you wish to search within the selected sample form the Fraction Dropdown, this should be the second dropdown displayed.

3. Click the Input Library Entry Button,  , to specify the parameters of the library entry to be queried. *These parameters will be outlined in more detail in the following section.*

4. Click the Apply Button, 

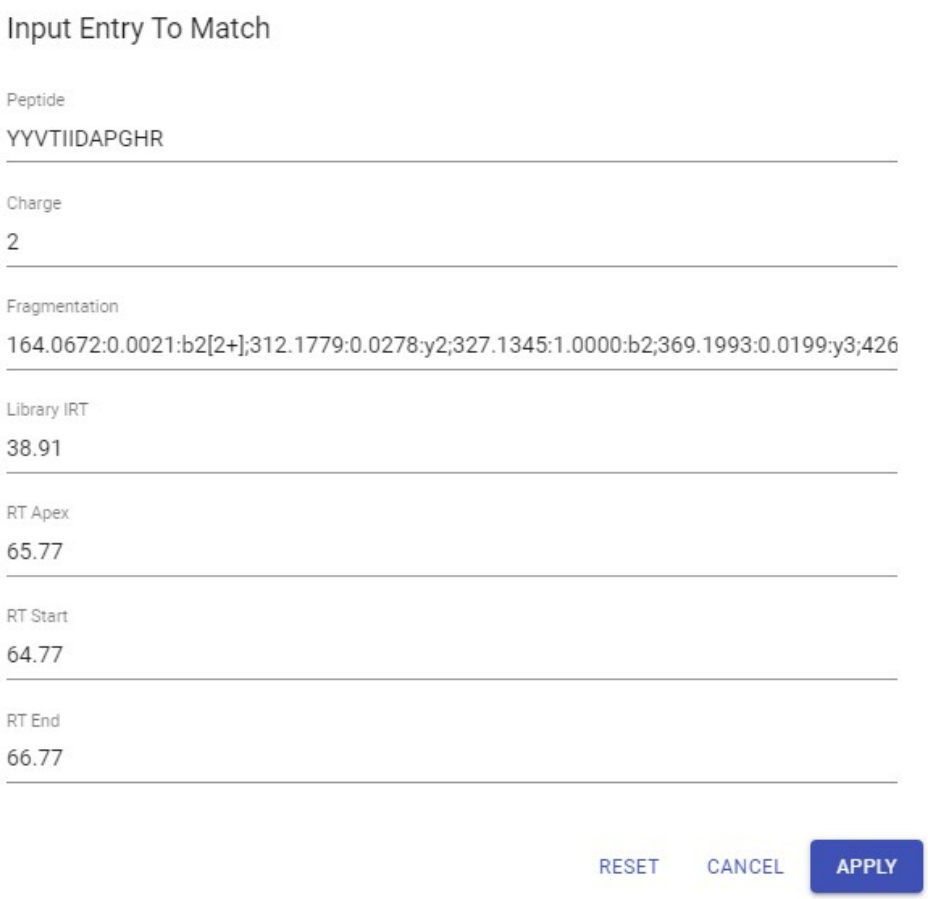
, to query the selected data file for the input librarD entry.

4.6.2. Manual Search Input

- **Peptide:** The peptide sequence that you wish to search for. This should include any modifications, for example LNIISNLDC(+57.02)VNEVIGIR.
- **Charge:** The precursor charge to search for.

> **Note**
>
> All optional parameters below will be filled in automatically if left blank with the best match found given the input parameters.

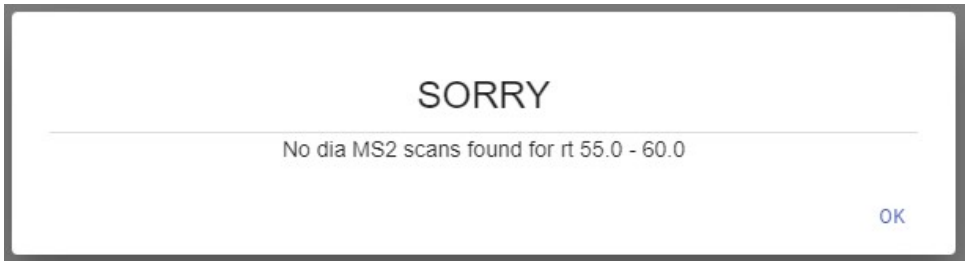
- **Fragmentation:** *(Optional)* The ion fragmentation information. This should be a list of ions with their m/z and intensity values in this format "<m/z>:<intensity>:<ion>", for example "327.13446,0.1547,b2". The intensity should be a relative intensity in decimal format, for example 0.1547 for 15.47% relative intensity.
- **Library iRT:** *(Optional)* The indexed retention time to search for, entered in minutes.
- **RT Apex:** *(Optional)* The retention time, in minutes, for the apex scan in the selected data file.
- **RT Start:** *(Optional)* The beginning retention time, in minutes, in the selected data file.
- **RT End:** *(Optional)* The end retention time, in minutes, in the selected data file.
- **CCS Start:** *(Optional)* The starting CCS value in the selected data file. Only applicable if searching TimsTOF Data.
- **CCS End:** *(Optional)* The ending CCS value in the selected data file. Only applicable if searching TimsTOF Data.



4.6.3. Viewing Manual Search Results

Once you have submitted your manual search, the results will be displayed after a brief delay.

If there were no results found in the selected data file that match the input spectra an error will appear indicating that no results could be found.



4.6.4. Peptide Table

The Peptide Table will list all possible identification results in the selected data file that meet the supplied criteria.

	Peptide	Mass	Charge	Length	m/z	Apex RT	Predicated RT
0	YYVTIIDAPGHR	1403.7197	2	12	702.8671	65.77	65.77

The columns in this peptide table are:

- **Peptide:** The amino acid sequence of the peptide, as input in the search parameters. A modified residue is followed by a pair of parentheses enclosing the modification mass.
- **Mass:** The monoisotopic mass of the peptide.

- **Charge:** The charge of the peptide, as input in the search parameters.
- **Length:** The number of amino acids in the peptide sequence.
- **m/z:** The precursor mass-to-charge ratio.
- **Apex RT:** The retention time (elution time) of the apex scan as recorded in the data.
- **Predicted RT:** The predicted retention time (elution time) of the input peptide.
- **1/k0:** This column is only available for TimsTof data, it shows the ion mobility range of the precursor.

4.6.5. Peptide Information

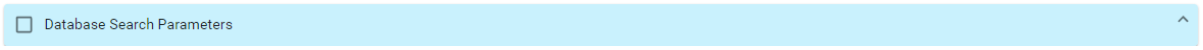
For more details on the provided figures refer to *Section 4.5.6*.

5. DIA Database Search Workflow and Parameter Settings

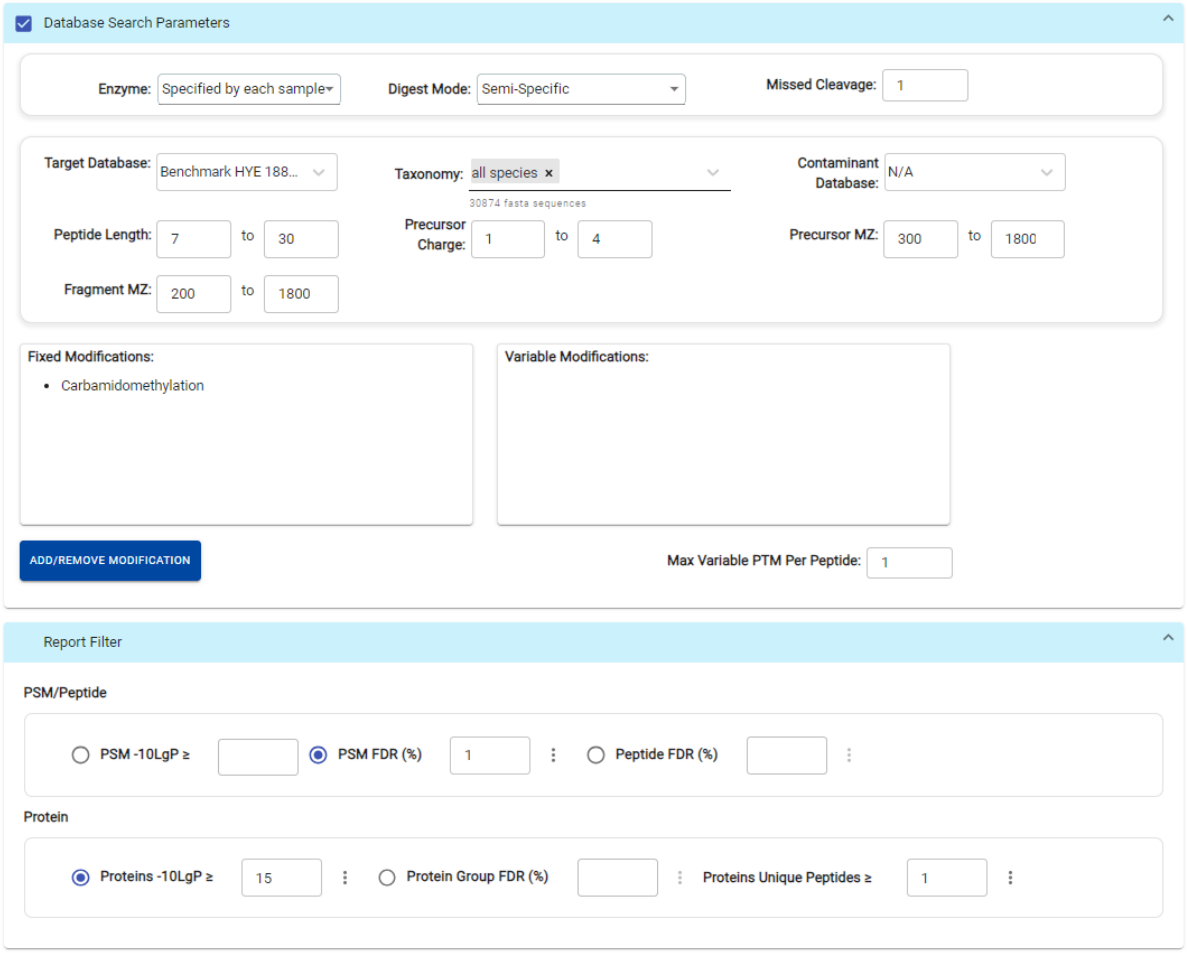
PEAKS DIA DB can be launched as a standalone workflow, or launched on top of a PEAKS Library Search workflow. An analysis report is generated to show the database search result, combined with the library search results if run on top of a PEAKS Library Search. Users can also run an optional PEAKS DIA Denovo search on top of the PEAKS DIA Database Search.

To conduct a database search workflow, follow the steps below:

1. Refer to Chapter 7, Section 3, Spectral Library Search Workflow and Parameter Settings, Section 3, “Spectral Library Search Workflow and Parameter Settings” , Step 1-3 to create a new project and add data. If you wish to run the Database Search on top of a Library Search, follow Step 4 as well.
2. Check the checkbox next to Database Search Parameters beneath Library Search. Uncheck the checkbox next to Library Search Parameters if you do not wish to run a Library Search before the Database Search.



3. Set the DB Search Parameters. See details in Section 5.1-5.6 below.



4. Click SUBMIT button to start the search.

5.1. Precursor and Fragment Error Tolerance

Precursor mass: This enables the selection of a precursor mass error tolerance. PEAKS Online DB Search identifies peptides within the precursor mass tolerance, which is the allowable m/z shift between the theoretical value of the peptide versus the observed value of the precursor ion. Precursor mass tolerance can be specified in either Da or ppm.

Fragment ion: This enables the selection of a fragment ion m/z tolerance, which is the allowable m/z shift between the theoretical value of the fragment ions of the peptide versus the observed value. PEAKS Online DB Search uses this value when scoring peptide-spectrum matches. PEAKS Online considers that a fragment ion is matched if the calculated m/z is within the tolerance. Fragment mass tolerance can be specified in Da.

Error tolerance parameters need to be set consistently with the mass accuracy of the instrument.

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.

5.1.1. Optimize Tolerance

When enabled, PEAKS Online will determine the optimal values to use for Precusor and Fragment Mass Error Tolerance for each fraction in your analysis. These values will replace the values provided in the previous Precursor and Fragment Mass Tolerance sections.

5.2. Enzyme Settings

Enzyme : PEAKS Online digests the protein database *in silico* to generate peptide candidates. It is necessary to specify the enzyme for protein digestion from the Enzyme drop-down menu.

Note

When the selected dataset is digested with different enzymes, "Specified by each sample" allows samples to be analyzed separately using their respective enzymes specified during project creation.

Note

"None" is a special enzyme digest rule allowing non-specific cleavage at both ends of the peptide. The "None" digest rule can cut at every residue, generating peptides with lengths up to 65 amino acids. Its usage is recommended when no digestion enzyme was used or when the digestion enzyme exhibits a high degree of non-specificity.

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

Digest Mode : This specifies the type of *in silico* cleavages allowed at a peptide termini 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 rules.

Note

If the enzyme is specified as "None", then no matter which mode is selected for "Digest Mode", the "None" enzyme digest rule will be applied.

5.3. Database Settings

A protein database must be added first before creating a new project to enable its selection in the workflow. In the Target Database drop-down menu, select the appropriate database that have been configured in PEAKS Online and set the taxonomy, if applicable. A contaminant database can also be selected from the same list of databases that have been configured in PEAKS Online.The length for a peptide can be set when creating an analysis, where the default peptide length range is between 6 and 45.

5.4. PTM Settings

Click "ADD/REMOVE MODIFICATION" button to open the "Select Modifications" dialogue and specify the fixed PTMs and a few (less than 11) common variable PTMs expected in the sample. If the required modification is not listed, the modification can also be created here. Additionally, the maximum number of variable PTMs per peptide can be specified and a value of less than 4 is recommended. A fixed modification forces all instances of applicable residues to be modified. A variable modification gives the option for the residues to be modified.

Note

For a specific residue, only one fixed modification is allowed.

Note

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

5.5. PSM Filter

PSM filter section is a mandatory section to filter out some results under the user-defined filter. Any changes to the filter will create a new analysis with old parameters and update the results and statistics.

- **PSM -10lgP:** The PEAKS peptide score (-10lgP) is calculated for every peptide-spectrum match (PSM) reported by PEAKS DB, PEAKS PTM, and SPIDER. The score is derived from the p-value that indicates the

statistical significance of the peptide-spectrum match. A peptide may be matched to many spectra, resulting in multiple PSMs. In that case, the peptide's score is calculated as the maximum 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* . A minimum requirement can be set and all identifications must pass this filter.

- **PSM FDR (%):** The PSM filter can also be set to use the false discovery rate. FDR is estimated using a "decoy-fusion" approach.

  ### Note

  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. Understanding PEAKS DIA Database Search Results

After a PEAKS Search is complete, all analyses selected during the project creation will be combined under the workflow drop-down menu. One is the *de novo* sequencing result, which includes *de novo* peptide sequences listed, spectrum annotation and other detailed information. 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.

The analysis results for database search tools are presented in four tabs:

- **Summary :** This shows an outline of the PEAKS Online 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. This page shows after the search is done.
- **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.

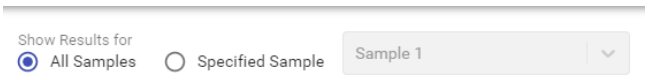
6.1. Result Filtration

At the top of each summary result page, filters can be set. Identifications that fall below the filters will not be visible in the result pages or exports. After changing any of these filter parameters, click the APPLY button to apply and save the new filters. The results and statistical information will be updated accordingly.

- **Proteins -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. A default threshold of 20 is set.
- **Proteins Unique Peptides:** The minimum number of unique supporting peptides for a protein identification. A unique peptide is defined as a peptide that can be mapped to only one protein group.

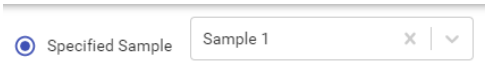
6.1.1. Show Results for All Samples

For the protein and the peptide tables, "All Samples" can be selected to view the combined search results from all the samples. In the Protein page, selecting "All Samples" will change the coverage panel displayed when a specific protein is highlighted. Supporting peptides from all samples instead of just a single sample will be shown.



6.1.2. Show Results for Specific Sample

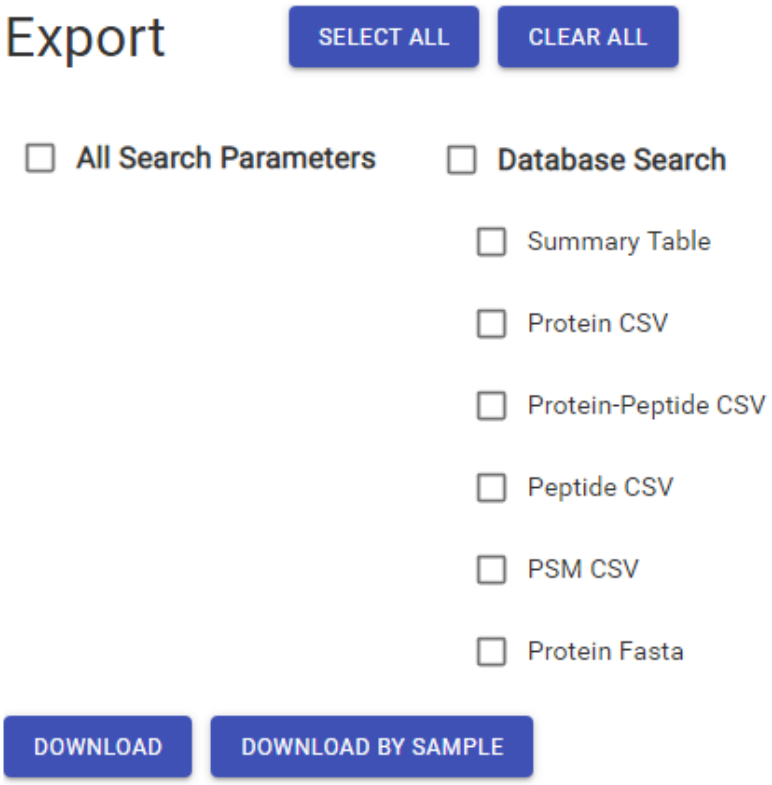
In the Protein, Peptide and De Novo page results, the filter can also be set to "Specified Sample" which will only show the results a single sample. By clicking in the text you can search for a specific sample by name.



Selecting "Specified Sample" will change the protein coverage view that is visible by clicking on a protein in the protein table. Peptides in the Protein, Peptide and De Novo Only pages will only be visible if they belong to the sample chosen. Once selected, the peptide tables will automatically add the "Area" column and the "# Spec" values will display the spectral count for the specified sample. For more information, refer to Section 3.5.2, “Peptide Table” . Similarly, the protein table will also include the columns "Coverage" and "Area" for the specified sample. For more information refer to Section 3.4.2, “Protein Table” .

6.2. Result Exporting

PEAKS Online database search results can be exported in the Export page:



- **Protein CSV:** A comma separated file containing all identified proteins and associated details from the protein table for all samples.
- **Protein-Peptide CSV:** A comma separated file containing all identified peptides and associated details from the peptide table found in the Protein page. Further details are provided regarding the protein the peptide supports. If a peptide supports multiple proteins, multiple rows are included for each supported protein for all samples.
- **Peptide CSV:** A comma separated file containing all identified peptides and associated details from the peptide table for all samples.
- **PSM CSV:** A comma separated file containing all identified peptide spectrum matches and associated details from the peptide table for all samples.

- **Pepxml:** An XML export of the identification results matching the pepxml schema. This can be uploaded to software that accept data in this format.
- **MzIdentML:** An XML export of the identification results matching the MZIdentML schema. This can be uploaded to software that accept data in this format.
- **Protein Fasta** : A list of protein identifications will be saved to proteins.fasta.

6.3. Summary View

6.3.1. Summary Statistics

The first chart displays the summary information for each sample.

- **Sample Name:** This column lists the samples names with the exception of the first row, which shows the combination of all the samples.
- **# MS Runs:** The total MS runs in the sample.
- **# MS1:** The total MS scans in the sample summed across all fractions.
- **# MS2:** The total MS/MS scans in the sample summed across all fractions.
- **# PSM:** Total number of peptide spectrum matches (PSMs) found in the sample.
- **# Peptides:** Total number of peptides found ignoring differences with leucine (L) and isoleucine (I) amino acids. Peptides with the same primary sequence but different PTMs are counted separately.
- **# Sequences:** Total number of different peptide backbone sequences found ignoring differences with leucine (L) and isoleucine (I) amino acids. Peptides with the same primary sequence but different PTMs are counted as one sequence.
- **# Proteins:** Provides the number of proteins given the parameters set for Protein Score (-10lgP) and number of unique peptides.
- **# Protein Groups:** PEAKS Online groups proteins identified by a common set of peptides. This number shows the number of protein groups in the filtered result based on All proteins.
- **# PSM / # MS2:** The percentage of MS2 scans in the sample that produced peptide spectrum matches.

6.3.2. False discovery rate (FDR) curve

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

PEAKS Spectral library search estimates the FDR using decoy peptides and spectra. A decoy peptide/spectrum for each target peptide/spectrum in a spectral library and the target and decoy spectra are searched together. The decoy method is based on: https://doi.org/10.1021/pr900947u

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

Note

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

Note

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

6.3.3. Statistics of filtered results

Table 1 shows the statistical numbers of the filtered results.

- **Top Proteins:** The number of proteins that 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 included here.
- **Proteins (#Unique Peptides):** In Table 2, 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. Peptides are counted based on their backbone sequence while considering I/L amino acids to be the same.
- **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.
- **-10lgP Cutoff:** The value is determined by the estimated PSM FDR. Only the peptides with PSMs above the -10lgP score threshold are listed in the Peptide view.

6.3.4. Precursor mass error

Figures 2(a) and 2(b) show the precursor mass errors of PSMs in filtered results. Figure 2(a) shows the distribution of the precursor mass error in a histogram with and without software calibration. Software calibration uses the observed distribution of mass error after identification to normalize the mass error around zero. Figure 2(b) shows precursor m/z versus precursor mass error in a scatter plot after software calibration. Figures 2(a) and 2(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.

6.3.5. RT Calibration

Plots the RT regression function used during the DIA database search. Peptides are plotted against their indexed retention time (iRT) predicted using machine learning.

6.3.6. Box plot for Rt change

A box plot showing the difference between the observed retention time and iRT peptides within set retention time bins across all fractions, showing the accuracy of retention time prediction.

6.3.7. Number of identified peptides in each sample by number of missed cleavages

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

6.3.8. PTM Profile

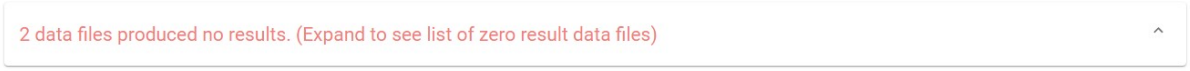
Table 3 shows modifications identified in the filtered result and the number of PSMs containing each modification.

- **Name:** The given name of the PTM.
- **Δ Mass:** The monoisotopic change in mass caused by the PTM.
- **Position:** The amino acids modified by the PTM.
- **# PSM:** The number of PSMs with the given PTM.
- **-10lgP:** The highest -10lgP score achieved by a peptide modified by the PTM.

- **Area:** The highest area achieved by a peptide modified by the PTM.

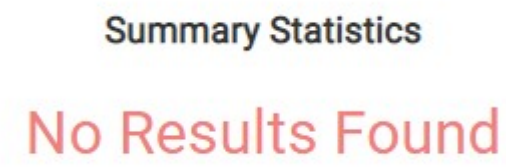
6.3.9. Low Quality Data Warnings

If PEAKS Online detects any data files producing very few or no results a warning section will appear at the top of the Summary Page indicating the number of data files producing poor or no results.



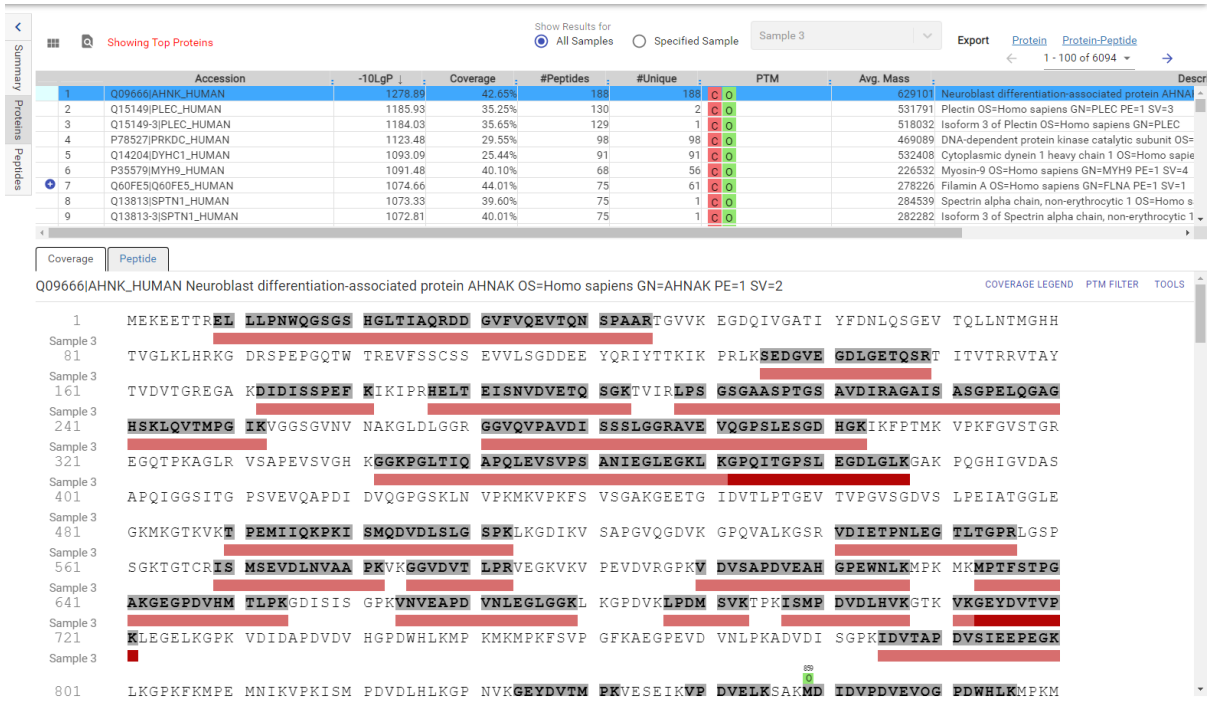
This section can be expanded to display a list of data files in the analysis that are producing poor or no results.

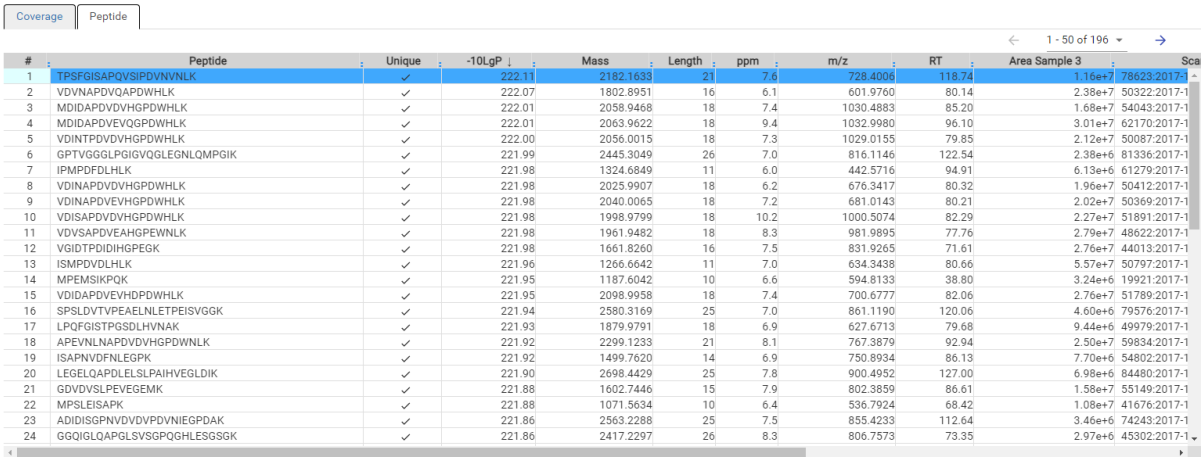
If there were no results found in any of your data files, a warning message stating "No Results Found" will be displayed instead of the usual Summary Page figures and tables.



6.4. Protein View

The Protein view lists protein identifications that have been filtered by the current settings set for the project. It also visualizes the protein sequence coverage for identified proteins. After clicking the "Peptide" tab we can view the supporting peptide list, double click any peptide can lead to that peptide sequence in the peptide page.





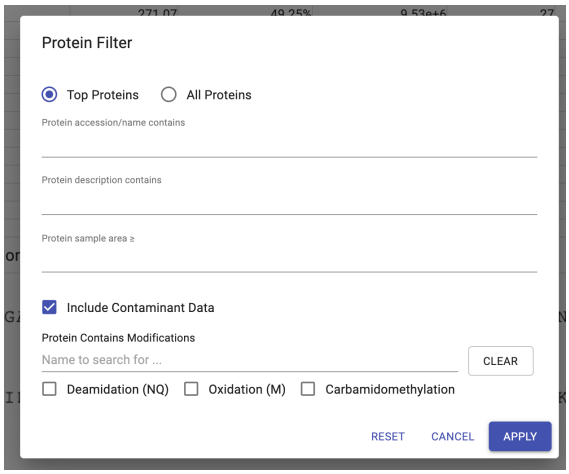
Coverage | Peptide

1 - 50 of 196

#	Peptide	Unique	-10LgP	Mass	Length	ppm	m/z	RT	Area Sample 3	Sca
1	TPSFGISAPQVSIPDVNVNLK	✓	222.11	2182.1633	21	7.6	728.4006	118.74	1.16e+7	78623:2017-1
2	VDVNAPDVQAPDWHLK	✓	222.07	1802.8951	16	6.1	601.9760	80.14	2.38e+7	50322:2017-1
3	MDIDAPDVDVHGPDWHLK	✓	222.01	2058.9468	18	7.4	1030.4883	85.20	1.68e+7	54043:2017-1
4	MDIDAPDVEVQGPDWHLK	✓	222.01	2063.9622	18	9.4	1032.9980	96.10	3.01e+7	62170:2017-1
5	VDINTPDVDVHGPDWHLK	✓	222.00	2056.0015	18	7.3	1029.0155	79.85	2.12e+7	50087:2017-1
6	GPTVGGGLPGIGVQGLEGNLQMPGIK	✓	221.99	2445.3049	26	7.0	816.1146	122.54	2.38e+6	81336:2017-1
7	IPMPDFDLHLK	✓	221.98	1324.6849	11	6.0	442.5716	94.91	6.13e+6	61279:2017-1
8	VDINAPDVDVHGPDWHLK	✓	221.98	2025.9907	18	6.2	676.3417	80.32	1.96e+7	50412:2017-1
9	VDINAPDVEVHGPDWHLK	✓	221.98	2040.0065	18	7.2	681.0143	80.21	2.02e+7	50369:2017-1
10	VDISAPDVDVHGPDWHLK	✓	221.98	1998.9799	18	10.2	1000.5074	82.29	2.27e+7	51891:2017-1
11	VDVSAPDVEAHGPEWNLK	✓	221.98	1961.9482	18	8.3	981.9895	77.76	2.79e+7	48622:2017-1
12	VGIDTPDIDIHGPEGK	✓	221.98	1661.8260	16	7.5	831.9265	71.61	2.76e+7	44013:2017-1
13	ISMPDVDLHLK	✓	221.96	1266.6642	11	7.0	634.3438	80.66	5.57e+7	50797:2017-1
14	MPEMSIKPQK	✓	221.95	1187.6042	10	6.6	594.8133	38.80	3.24e+6	19921:2017-1
15	VDIDAPDVEVHDPDWHLK	✓	221.95	2098.9958	18	7.4	700.6777	82.06	2.76e+7	51789:2017-1
16	SPSLDVTVPEAELNLETPEISVGGK	✓	221.94	2580.3169	25	7.0	861.1190	120.06	4.60e+6	79576:2017-1
17	LPQFGISTPGSDLHVNAK	✓	221.93	1879.9791	18	6.9	627.6713	79.68	9.44e+6	49979:2017-1
18	APEVNLNAPDVDVHGPDWNLK	✓	221.92	2299.1233	21	8.1	767.3879	92.94	2.50e+7	59834:2017-1
19	ISAPNVDFNLEGPK	✓	221.92	1499.7620	14	6.9	750.8934	86.13	7.70e+6	54802:2017-1
20	LEGELQAPDLELSLPAIHVEGLDIK	✓	221.90	2698.4429	25	7.8	900.4952	127.00	6.98e+6	84480:2017-1
21	GDVDVSLPEVEGEMK	✓	221.88	1602.7446	15	7.9	802.3859	86.61	1.58e+7	55149:2017-1
22	MPSLEISAPK	✓	221.88	1071.5634	10	6.4	536.7924	68.42	1.08e+7	41676:2017-1
23	ADIDISGPNVDVDVPDVNIEGPDAK	✓	221.86	2563.2288	25	7.5	855.4233	112.64	3.46e+6	74243:2017-1
24	GGQIGLQAPGLSVSGPQGHLESGSGK	✓	221.86	2417.2297	26	8.3	806.7573	73.35	2.97e+6	45302:2017-1

6.4.1. Protein View Filters

By selecting the button, the Protein Filter will pop up. It controls which protein to display.



- **Show Top or All 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.
- **Protein accession/name contains:** If you add text here, only proteins with accessions that contain the text will be included in the protein table.
- **Protein description contains:** If you add text here, only proteins with descriptions that contain the text will be included in the protein table.
- **Protein sample area:** Filters 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.

Note

Only numeric digits can be typed in the protein sample area. For example, instead of 1.00e+4, 10000 must be typed.

- **Spider Mutations:** This filter is only visible in SPIDER result view filter. Check it to view proteins that have mutations.
- **Protein Contains Modifications** : Proteins that are supported by peptides containing selected PTMs will be included.

6.4.2. Protein Table

Each row in the table represents a group of proteins that are supported by a common set of peptides. A ⊕ indicates that the group has multiple proteins. To expand the group, click that button.

The following columns are sortable: Accession, -10LgP, Coverage, Peptides, #Unique, Avg Mass, when All Samples are shown.

The columns in the Protein table are:

- **Accession:** The accession id of the protein as seen in the FASTA database.
- **-10lgP:** The protein confidence score. Refer to Section 3.1, “Result Filtration” for detailed explanation.
- **Coverage:** The percentage of the protein sequence that is covered by supporting peptides. This is the total coverage including results from all samples, unless results are only shown for a specified sample.
- **#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 the 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.

> **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.

6.4.3. Global Protein Comparative View

Global Protein Comparative View provides detailed comparisons among different samples regarding the identified protein. To Show/Hide Columns, click ▦ on top-left of the protein table. From the dialog, users can add and remove sample coverage, area, and spectral count columns. In the protein table, the selected columns will be displayed and can be exported.

- **Coverage by Sample:** the coverage broken down by sample can be displayed if multiple samples were used. Only peptides that pass the filters are included in the percentage calculation.
- **Area by Sample:** total area of peptide features from unique supporting peptides in each sample are used for calculation. Peptides must pass the filters.
- **#Spec by Sample:** total number of spectra identified that support the given protein. Totals are given for each sample included in the search.

6.4.4. Protein Coverage

The Protein Coverage view visualizes the mapping of supporting peptides to the protein selected in the Protein table. It also shows all identified modification or mutation sites to assist with protein characterization at the amino acid level. The coverage view shows different views under different modes. The view shows a sample comparative coverage view in the all sample mode while it shows protein sequence coverage view in the specific sample mode.

6.4.4.1. Sample Comparative Coverage

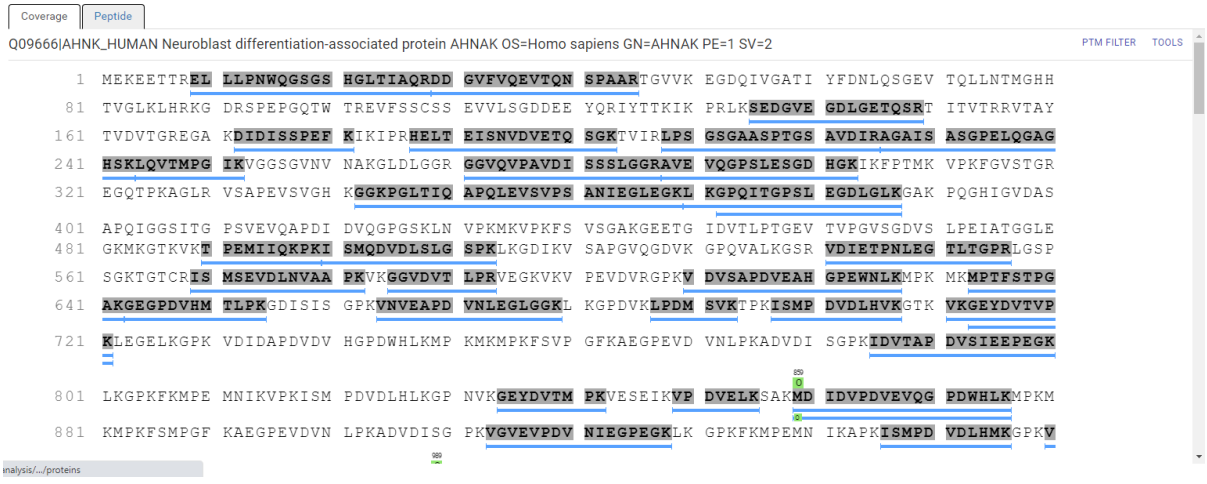
This area visualizes the frequency of supporting peptide sequences from selected samples in the protein sequence. At the top of the protein sequence coverage view, the header information of the protein is shown. At the top-right corner, the COVERAGE LEGEND button will show a continuous color legend for the peptide frequency. Specific modifications can be selected in the PTM FILTER button and the template protein can be copied using the TOOLS button.

Regions in the protein sequence that are covered by supporting peptides are displayed in bold font with a grey background. Placing the cursor over a amino acid on the sequence shows the position of the amino acid in the sequence and the frequencies from different selected samples.

The frequency of supporting peptide sequences from different samples can be shown as colored bars under the protein sequence. The deeper the color, the higher frequency of that supporting peptide in the sample.

6.4.4.2. Protein Sequence Coverage

This area visualizes the coverage of the protein sequence in the specified sample mode. The header information of the protein, a PTM filter and a tool to copy the protein information are shown on the top of the protein sequence coverage view. Clicking the PTM filter button will pop up a menu to list the modifications identified in supporting peptides on the protein and the checkbox on the left controls whether to show the modification in the protein sequence coverage view above.



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

indicates the position in the protein sequence. Placing the cursor over an amino acid on the sequence shows the position of the amino acid in the sequence.

The de novo only peptides are the confident de novo sequence tags that remain unidentified by the identification search algorithms. To report a de novo only peptide, the De novo score must be equal to or better than the specified threshold. Meanwhile, the score of the spectrum's best identification search result should be no greater than the specified -10lgP threshold.If we enable the de novo only tag in the parameter setting, we can see the de novo only sequence with grey line below.

The supporting peptides 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 will pop up the annotated spectrum associated with the supporting peptide.

6.4.5. Supporting Peptides

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

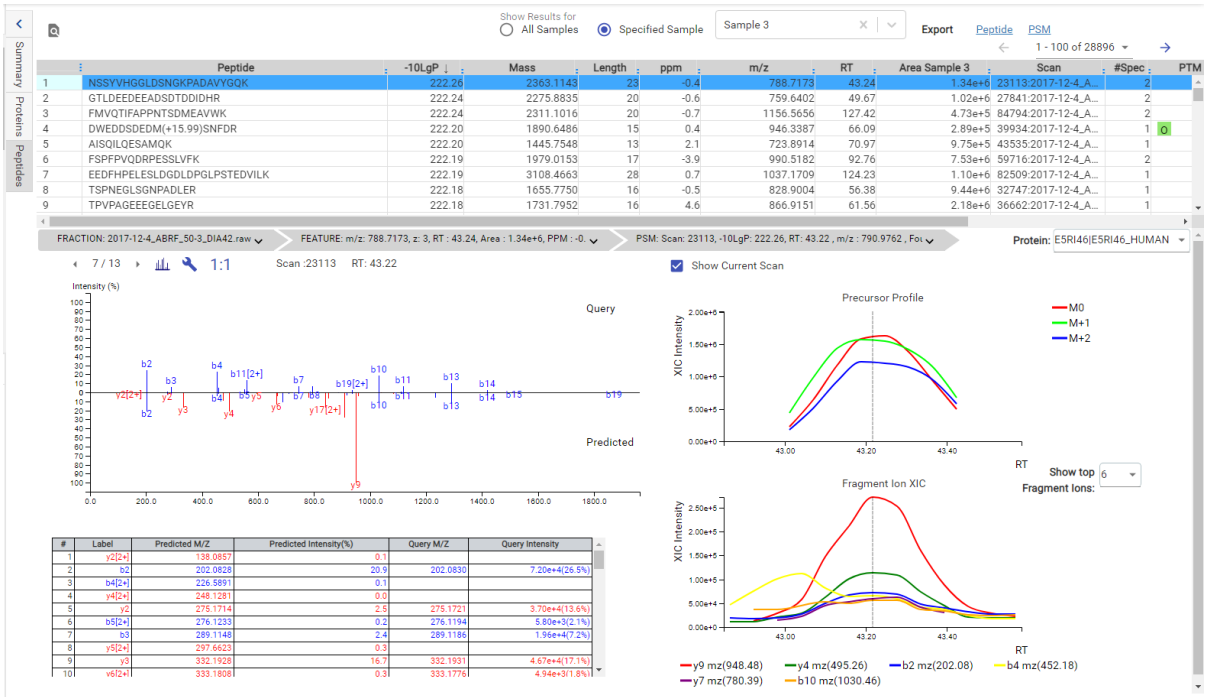
- **Unique:** This shows whether the peptide is a unique supporting peptide to the protein group.
- **Start:** This shows the peptide's starting position in the protein.
- **End:** This shows the peptide's ending position (inclusive) in the protein.
- **Found by** : The name of the PEAKS search workflow that identifies the peptide. This can be either DB Search or Library search.

Refer to Section 3.5.2, “Peptide Table” for the descriptions of other columns.

Double click a supporting peptide, it will go to that peptide sequence in the peptide page

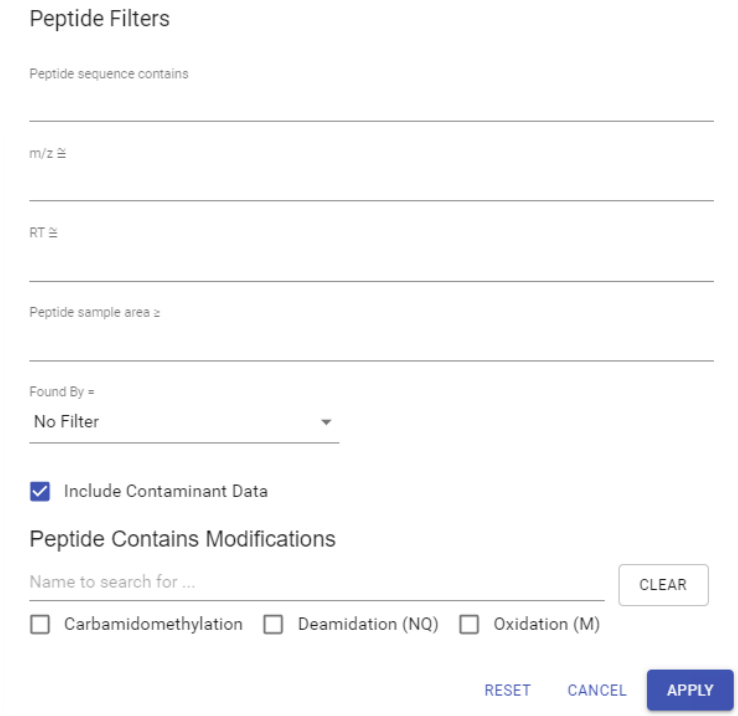
6.5. Peptide View

The Peptide Table shows the filtered peptide identification results. Click on a peptide row to show the spectrum below the peptide table.



6.5.1. Peptide View Filter

Click the button, the following Peptide Filter will pop up. It controls which peptides to display.



- **Peptide sequence contains:** Enter an amino acid sequence, only peptides that contain that sequence will be displayed in the peptide table.
- **m/z:** Filter based on the expected m/z, the last significant digit is rounded.
- **RT:** Filter based on the expected RT, the last significant digit is rounded.
- **Peptide sample area:** Filters based on total peptide area found in each sample. If a peptide is found in multiple samples, there should be at least one sample pass this filter for the peptide to be included.
- **Peptide Contains Modifications:** Check boxes are provided for each identified modification. Peptides that contain the selected PTMs will be shown.
- **Found by:** filter peptides found in DB search or Spectral Library search.

6.5.2. 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 is sorted by the -10lgP value. In the Protein list drop-down window of each peptide, after clicking any of the protein, the protein page will be opened.

The columns Peptides, -10LgP, m/z, RT, Accession are sortable, i.e. Clicking any of these header can sort the whole table based on that column.

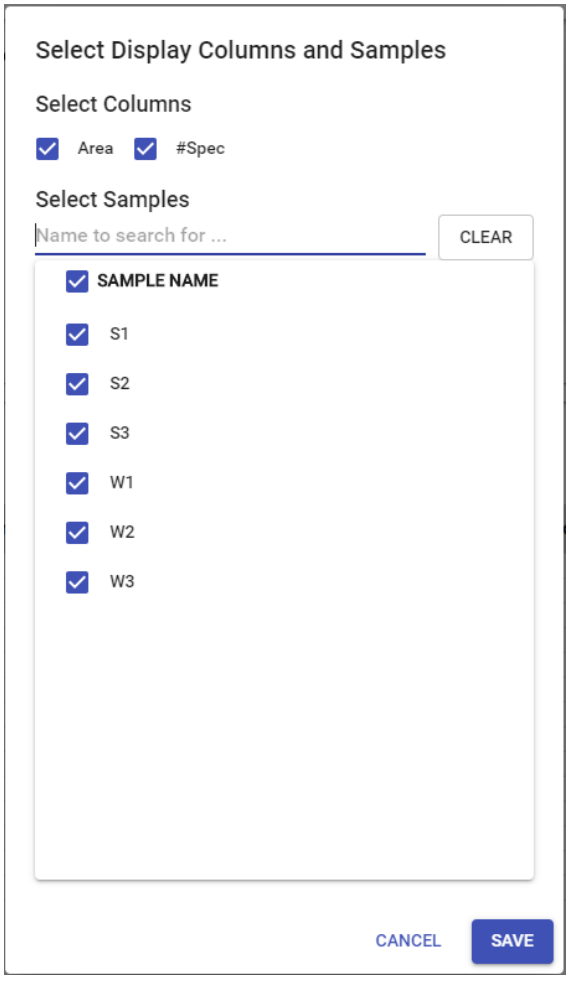
The columns in the Peptide table are:

- **Peptide:** The amino acid sequence of the peptide, as determined in PEAKS Search. A modified residue is followed by a pair of parentheses enclosing the modification mass.
- **-10lgP:** The peptide -10lgP score. The score indicates the scoring significance of a peptide-spectrum match.
- **Mass:** The monoisotopic mass of the peptide.

- **Length:** The number of amino acids in the peptide sequence.
- **ppm:** The precursor mass error, calculated as $10^6 \times$ (precursor mass - peptide mass) / peptide mass.
- **m/z:** The precursor mass-to-charge ratio.
- **RT:** The retention time (elution time) of the spectrum as recorded in the data.
- **#RT:** The retention time difference between the query and predicted spectrum.
- **Scan:** The scan number of the spectrum that matches the peptide sequence with the highest -10lgP.
- **#Spec:** The number of MS2 spectra assigned to the peptide.
- **PTM:** The types and the numbers of modifications present in the peptide shown in color-coded icons.
- **Ion Intensity:** The relative ion intensity (%) of site-determining fragment ions for modification sites.
- **Accession:** The accession number of the highest-scoring protein containing this peptide.
- **Precursor Id:** This column is only available for TimsTof data, it shows precursor's id from the raw data.
- **1/k0:** This column is only available for TimsTof data, it shows the ion mobility range of for the precursor.

6.5.3. Peptide Global Comparative View

"Peptide Global Comparative View" provides detailed comparisons among different samples regarding the identified peptides. To Show/Hide Columns, click on top-left of the peptide table. From the pop-up window, users can add and remove area and spectral count columns. In the peptide table, the selected columns will be displayed and can be exported. If "Specified Sample" is selected these columns are automatically displayed.



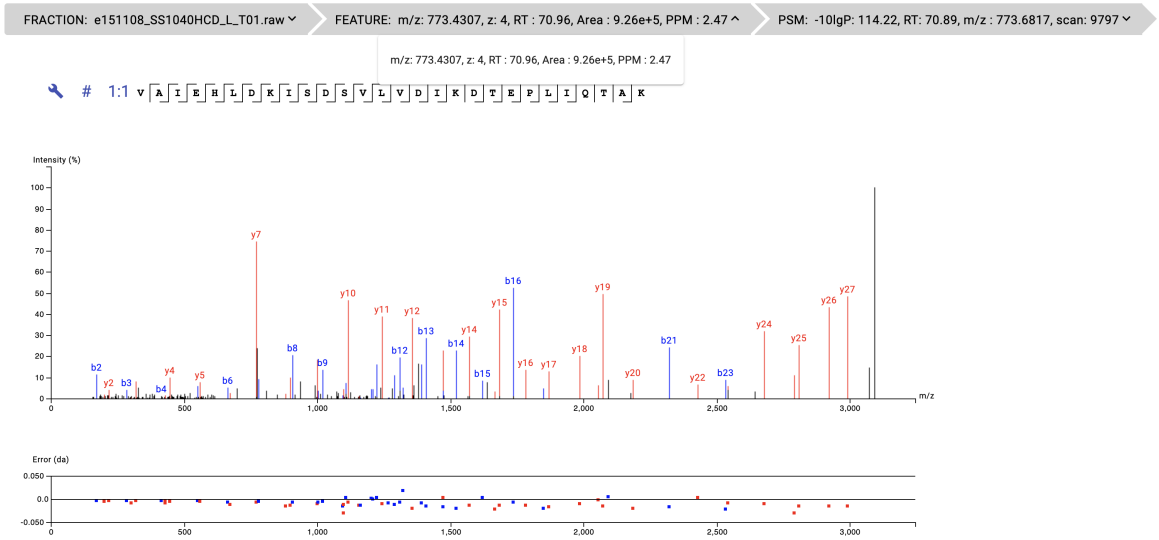
Note

PEAKS Online 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. The default can be changed by the administrator.

- **Area by sample:** The total area of all peptide features matched to spectra that identified the peptide. Area calculations are separated into multiple columns if multiple samples were included in the run.
- **#Peptide-spectrum matches by sample:** The number of peptide spectrum matches associated with a peptide separated by sample.

6.5.4. Peptide Menu

The peptide menu is only shown when "Show Results for Specified Sample" is selected in the filter section. The peptide menu displays 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 "FRACTION", “FEATURE” or “PSM” button to view the Spectrum Annotation of the other PSMs.

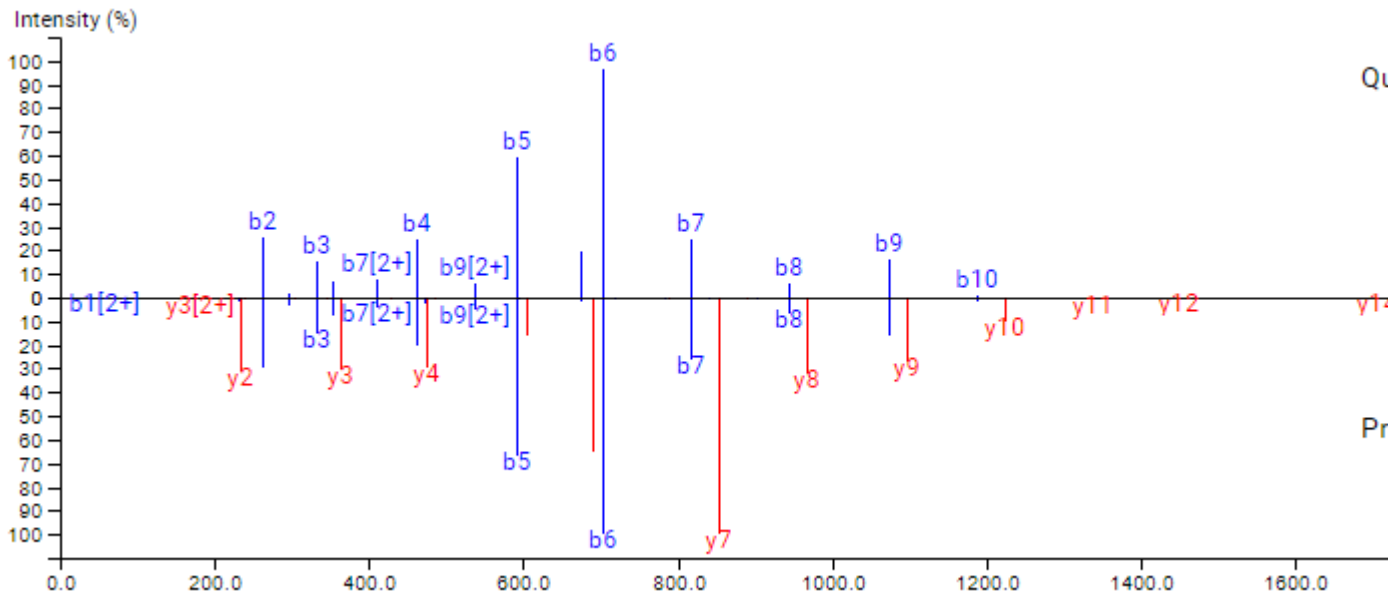


6.5.5. Peptide-Spectrum Matches

This section displays the spectrum annotation and other information about the peptide selected in the Peptide table. When multiple spectra are matched to the peptide, only the information about the top-scoring PSM is shown by default. The interface of this section is identical to the Spectrum Annotation in the PEAKS Spectral Library result. For detailed instructions, refer to Section 3.5, “Spectrum Annotation” in Chapter 5, Peptide De Novo Sequencing.

6.5.6. DIA Database Search Mirror Plot

Similar to Library Search, pressing the  button will display a mirror plot instead of the traditional spectrum view.



This view will display the ions from the query spectrum compared to the predicted spectrum generated from the database sequence.

6.5.7. Error Map

The "Error Map" shows the mass errors of the 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 "Ion Annotation Settings".

6.5.8. Ion Table

The "Ion Table" shows the calculated mass of possible fragment ions based on the ion types selected in the Ion Annotation Settings. If a fragment ion is found in the spectrum, its mass value is displayed in color. N-terminal ions are shown in blue and C-terminal ions are shown in red. A fragment ion is found when there is a matching peak within the mass error tolerance, as defined in the *de novo* sequencing parameters.

6.5.9. Precursor Profile

An eXtracted Ion Chromatogram (XIC) chart that displays the shape of the selected peptides precursor.

6.5.10. Fragment Ion XIC

An eXtracted Ion Chromatogram (XIC) chart that displays the shape of each identified fragment ion across all scans for the selected PSM. Below the XIC charts is a slider to adjust the correlation threshold of the fragment ion with the selected peptide precursor. Increasing the slider value will filter out fragment ions that are below the threshold.

6.5.11. TimsTOF In Frame Mode

For timsTOF data only, there is an additional "In Frame Mode" button,  , above the spectrum view. Clicking this button will change the annotated spectrum view from "frame mode", the default view, to "in frame mode". To switch back click the "Frame Mode" button,  . The standard, "Frame Mode" displays the scans merged over the CCS dimension. The "In Frame Mode" displays the scans merged over the retention time dimension.

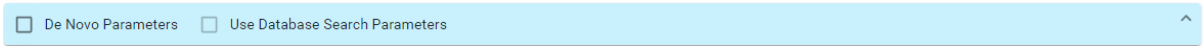
7. DIA *De Novo* Analysis Workflow and Parameter Settings

PEAKS DIA *De Novo* can be launched as a standalone workflow, or launched on top of a PEAKS Library Search, PEAKS DIA Database Search, or PEAKS Library Search + PEAKS Database Search workflow. An analysis report is generated to show the *De Novo* search results, combined with both the library search results and database search results if run on top of a PEAKS Library Search or PEAKS Database Search.

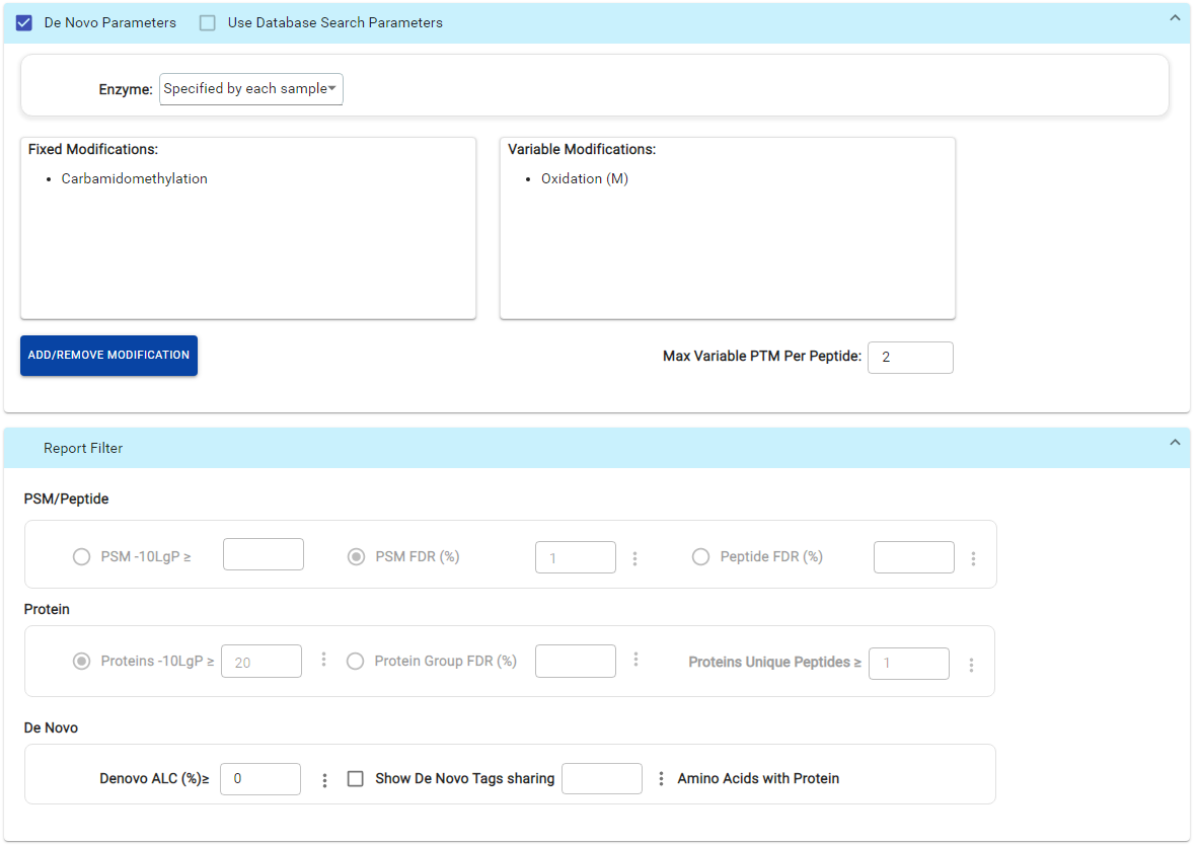
To conduct a DIA *De Novo* search workflow, follow the steps below:

1. Refer to Chapter 7, Section 3, Spectral Library Search Workflow and Parameter Settings, Section 3, “Spectral Library Search Workflow and Parameter Settings” , Step 1-3 to create a new project and add data. If you wish to run the De Novo Search on top of a Library Search, follow Step 4 as well. If you wish to run the De Novo Search on top of a DIA Database Search follow the steps in Section 2, “Database Search Workflow and Parameter Settings”.

2. Check the checkbox next to De Novo Parameters beneath Database Search. Uncheck the checkbox next to Library Search Parameters if you do not wish to run a Library Search before the De Novo Search.



3. Set the De Novo Parameters. See details in Section 7.1-7.6 below.



4. Click SUBMIT button to start the search.

7.1. Precursor and Fragment Error Tolerance

Precursor mass: This enables the selection of a precursor mass error tolerance. PEAKS Online DB Search identifies peptides within the precursor mass tolerance, which is the allowable m/z shift between the theoretical value of the peptide versus the observed value of the precursor ion. Precursor mass tolerance can be specified in either Da or ppm.

Fragment ion: This enables the selection of a fragment ion m/z tolerance, which is the allowable m/z shift between the theoretical value of the fragment ions of the peptide versus the observed value. PEAKS Online DB Search uses this value when scoring peptide-spectrum matches. PEAKS Online considers that a fragment ion is matched if the calculated m/z is within the tolerance. Fragment mass tolerance can be specified in Da.

Error tolerance parameters need to be set consistently with the mass accuracy of the instrument.

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.

7.1.1. Optimize Tolerance

When enabled, PEAKS Online will determine the optimal values to use for Precursor and Fragment Mass Error Tolerance for each fraction in your analysis. These values will replace the values provided in the previous Precursor and Fragment Mass Tolerance sections.

7.2. Enzyme Settings

Enzyme : PEAKS Online digests the protein database *in silico* to generate peptide candidates. It is necessary to specify the enzyme for protein digestion from the Enzyme drop-down menu.

Note

When the selected dataset is digested with different enzymes, "Specified by each sample" allows samples to be analyzed separately using their respective enzymes specified during project creation.

Note

"None" is a special enzyme digest rule allowing non-specific cleavage at both ends of the peptide. The "None" digest rule can cut at every residue, generating peptides with lengths up to 65 amino acids. Its usage is recommended when no digestion enzyme was used or when the digestion enzyme exhibits a high degree of non-specificity.

7.3. PTM Settings

Click "ADD/REMOVE MODIFICATION" button to open the "Select Modifications" dialogue and specify the fixed PTMs and a few (less than 11) common variable PTMs expected in the sample. If the required modification is not listed, the modification can also be created here. Additionally, the maximum number of variable PTMs per peptide can be specified and a value of less than 4 is recommended. A fixed modification forces all instances of applicable residues to be modified. A variable modification gives the option for the residues to be modified.

Note

For a specific residue, only one fixed modification is allowed.

Note

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

7.4. *De Novo* Filter

The report filter section is a mandatory section to filter out some results under the user-defined filter. Any changes to the filter will create a new analysis with old parameters and update the results and statistics.

- **Denovo ALC (%):** This filter will filter out any De Novo candidates with an average confidence less than the specified value.

8. Understanding PEAKS DIA *De Novo* Sequencing Results

When *de novo* sequencing is done, you can access the PEAKS Online project from the "My Projects" page. The analysis report for *de novo* search tools are presented in two tabs:

- **Summary:** This shows an outline of the PEAKS *de novo* search results with key statistics. The overall quality of the experiment can be examined and the filters for *de novo* ALC can be adjusted. This page shows after the search is done.
- **De novo:** This shows a list of peptide sequences that pass the ALC threshold detected by *de novo* sequencing.

The *de novo* view displays the *de novo* sequencing results. The table on the top displays *de novo* sequences, while the bottom section provides additional information about the peptide-spectrum match. The results can be filtered and exported.

8.1. Result Filtration

A PEAKS Online *de novo* sequencing result can be filtered by Average Local Confidence (ALC) score. Result for different samples can be specified at "Specified Sample" filter. Low quality *de novo* sequences can be filtered out by specifying a minimal threshold of the ALC score. The purpose of filtering is to remove poor sequences in which residues are incorrectly identified. By default, the ALC threshold is set to 50%. After changing the threshold, click APPLY to apply the new filter. The result in the DE NOVO result view will be updated accordingly.

A PEAKS Online *de novo* sequencing results display based on the sample selected in the "specified sample" drop-down menu. The purpose of this filtering is to limit the quantity of *de novo* sequencing results displayed in the table. The specified sample can be changed through the drop-down list, click APPLY to apply the new filter. The result will be updated accordingly.

8.2. Exporting *De Novo* Results

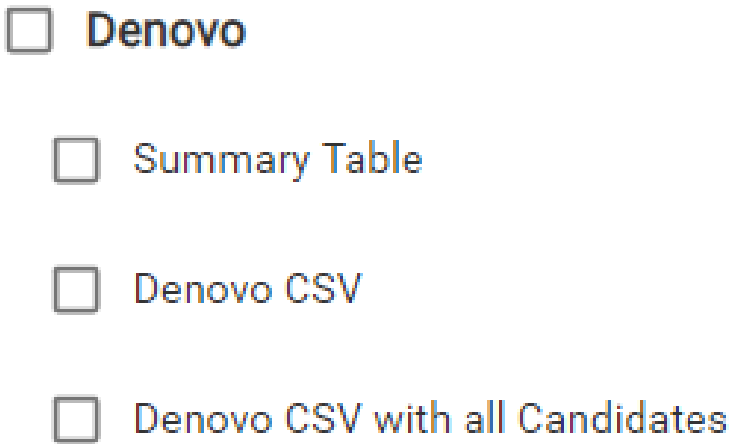
The *de novo* results can be exported as "Denovo.csv" or "Denovo CSV with all Candidates". These export files can be downloaded from either the Denovo Results page or the export page.

- **Denovo CSV**: the filtered *de novo* results shown are exported. It can also be downloaded from the export page as shown in the following figure.

- **Denovo CSV with all candidates**: for each MS2 spectrum, denovo results of the 5 highest ALC scores are generated. *De novo* sequences that pass the ALC score filter are all included in this file. It can also be downloaded from the export page shown in the following figure.

In the Denovo results page, above the peptide table there are two export options: "Denovo CSV", and "Denovo CSV with all Candidates".

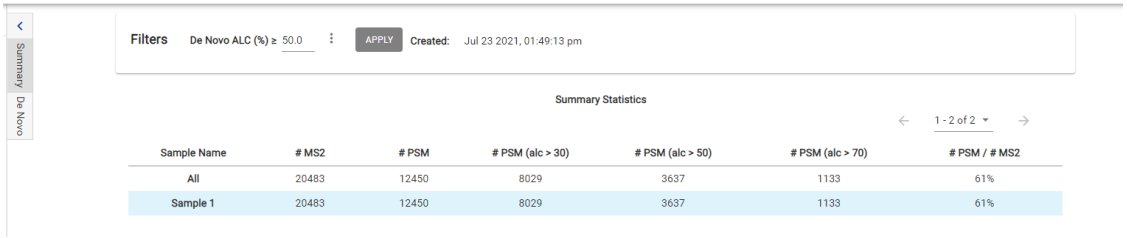
In the export page, there are 3 files which can be exported: "Summary Table", "Denovo CSV", and "Denovo CSV with all Candidates".



- Summary Table: the overview number of each sample for the fields MS2, PSM, PSM (alc>30), PSM (alc>50), PSM (alc>70), #PSM/#MS2 and notes.

8.3. Summary View

The "Summary" view reports key statistics as an overview of the result.



8.3.1. Result Statistics

This table of the summary page shows the overview number of each sample for the following fields:

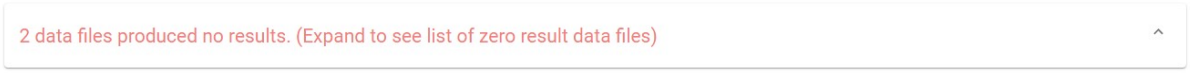
- **# MS2**: The number of MS/MS scans from the data.
- **# PSM**: The number of peptide-spectrum matches.
- **# PSM (alc>30)**: The number of peptide-spectrum matches with an ALC score greater than 30.
- **# PSM (alc>50)**: The number of peptide-spectrum matches with an ALC score greater than 50.
- **# PSM (alc>70)**: The number of peptide-spectrum matches with an ALC score greater than 70.
- **#PSM / #MS2:** The number of peptide-spectrum matches per MS2 scan.

Note

ALC score is the average local confidence score. ALC is calculated as the total of the residue local confidence scores in the peptide divided by the peptide length.

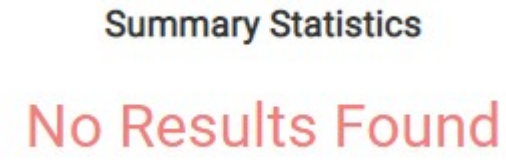
8.3.2. Low Quality Data Warnings

If PEAKS Online detects any data files producing very few or no results a warning section will appear at the top of the Summary Page indicating the number of data files producing poor or no results.



This section can be expanded to display a list of data files in the analysis that are producing poor or no results.

If there were no results found in any of your data files, a warning message stating "No Results Found" will be displayed instead of the usual Summary Page figures and tables.



8.4. Peptide Table

The peptide table shows the filtered *de novo* sequencing results. For each scan, the best candidate is displayed.

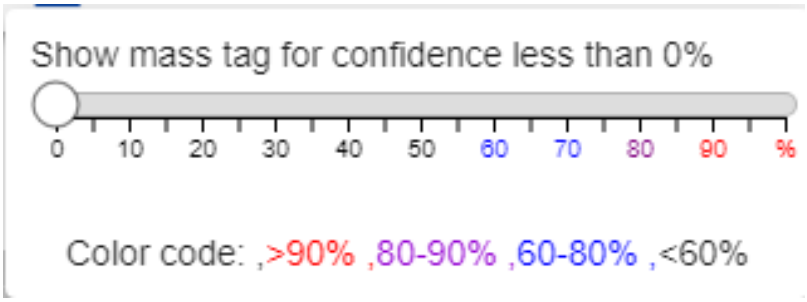
The following list describes the contents in each column:

- **Scan**: The scan number. The scan number is a unique index for tandem mass spectra in the data. If the scan numbers could be read from the data file, the scan numbers that match with the instrument data file will be shown.
- **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.
- **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 peptide 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 10E6 × (precursor mass - peptide mass) / peptide mass.
- **PTM**: The types and the numbers of the modifications present in the peptide shown as color-coded icons.
- **Mode**: The fragmentation mode in which *de novo* sequencing is performed by the algorithm.
- **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.

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

Mass Tags: The low confidence residues can be displayed as mass tags by adjusting the local confidence score threshold using the button. If the score is set at 0, all of the amino acids in the peptide sequences will be displayed. Increasing the threshold will reduce residues below the threshold to mass tags. When low confidence residues are reduced to mass tags, the remaining residues become sequence tags. The length of the longest tag is called the tag length and reported in the export files.



Filter Peptides: The peptide table can be filtered by selecting the button and prompting the "Denovo Filter" window. The peptides can be searched by sub-sequence, scan number, precursor Id, m/z, and retention time.

8.5. Spectrum Annotation

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

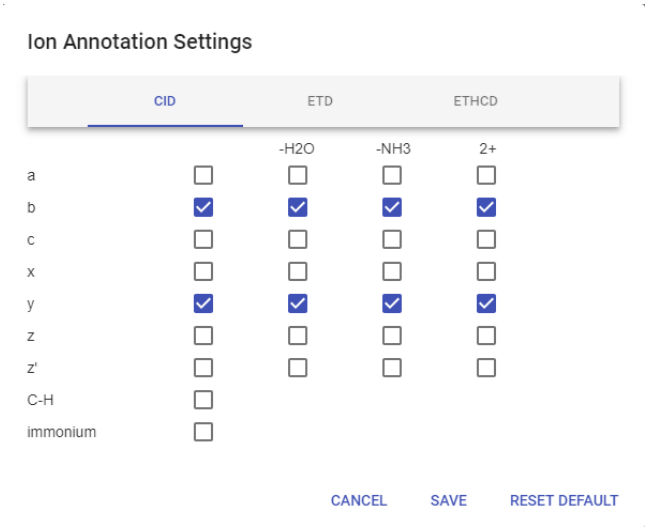
Moving the cursor over the peptide sequence in the spectrum will show the mass transitions for a particular amino acid residue.

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

- Zoom in/out on m/z: Place the cursor on the point in the spectrum you would like to zoom in on. Alternatively, place the cursor in the Error (da) figure and scroll the mouse wheel.

- Increase/Decrease peak intensity: Place the cursor on the Intensity axis and scroll the mouse wheel.
- See the whole spectrum: click the 1:1 button.

Click to open the Ion Annotation Settings dialogue and change the ions to be annotated in the spectrum. To reset the settings to the PEAKS Online defaults, use "RESET DEFAULT" button.



The % button will switch the y-axis to relative intensity. Once selected, scrolling over peaks will display the relative intensity instead of the absolute intensity.

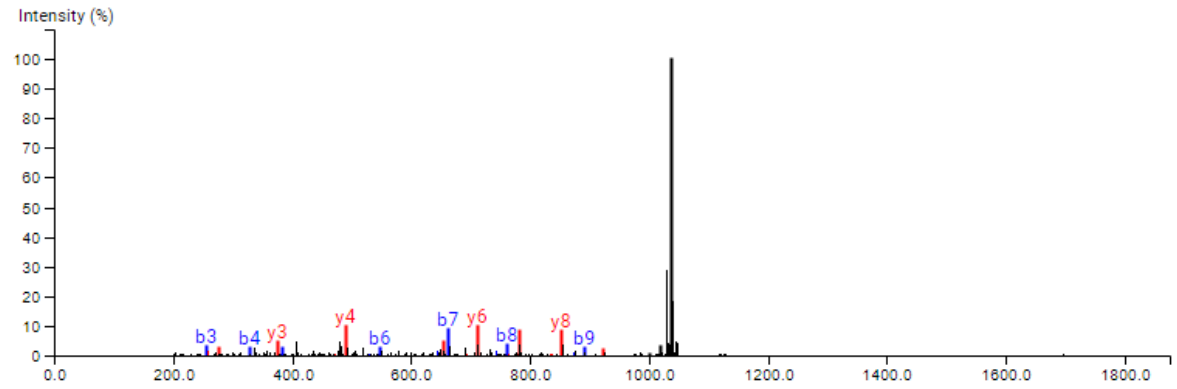
The # button will switch the y-axis to absolute intensity. Once selected, scrolling over peaks will display the absolute intensity instead of the relative intensity.

8.6. Spectrum View

For DIA De Novo results, the annotated spectrum view can be displayed in three different formats.

8.6.1. Raw Scan

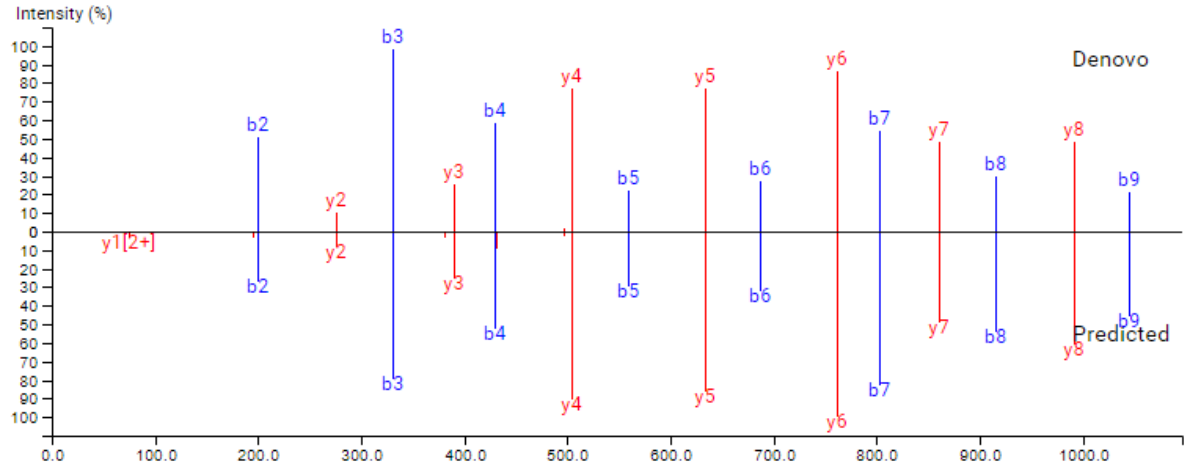
Click on the button to display the raw scan.



This view will display the raw scan from the data file, with the identified ions annotated.

8.6.2. Mirror Plot

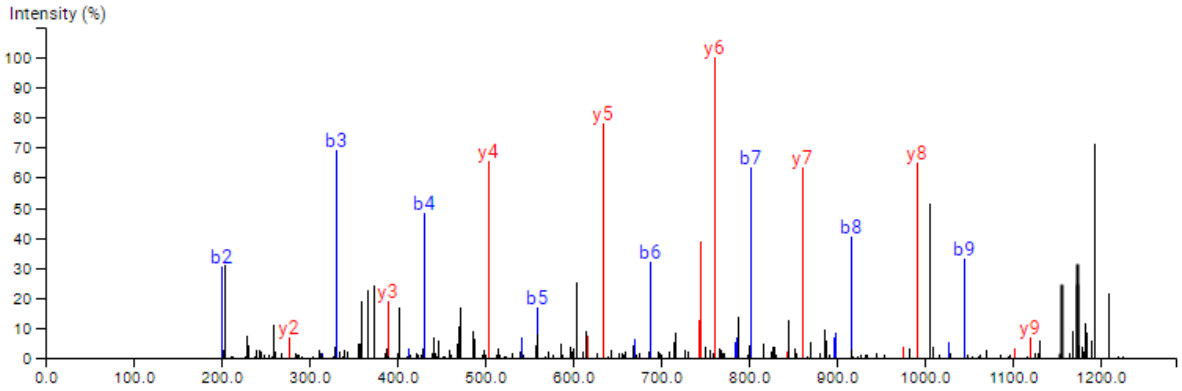
Click on the button to display the mirror plot.



This view will display the ions from the de novo candidate compared with the ions from the predicted spectrum.

8.6.3. Pseudo Scan

Click on the  button to display the pseudo scan.



8.7. Error Map

The "Error Map" shows the mass errors of the 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 "Ion Annotation Settings".

8.8. Ion Table

The "Ion Table" shows the calculated mass of possible fragment ions based on the ion types selected in the Ion Annotation Settings. If a fragment ion is found in the spectrum, its mass value is displayed in color. N-terminal ions are shown in blue and C-terminal ions are shown in red. A fragment ion is found when there is a matching peak within the mass error tolerance, as defined in the *de novo* sequencing parameters.

8.9. Precursor Profile

An eXtracted Ion Chromatogram (XIC) chart that displays the shape of the selected peptides precursor.

8.10. Fragment Ion XIC

An eXtracted Ion Chromatogram (XIC) chart that displays the shape of each identified fragment ion across all scans for the selected PSM. Below the XIC charts is a slider to adjust the correlation threshold of the fragment

ion with the selected peptide precursor. Increasing the slider value will filter out fragment ions that are below the threshold.

8.11. TimsTOF In Frame Mode

For timsTOF data only, there is an additional "In Frame Mode" button,  , above the spectrum view. Clicking this button will change the annotated spectrum view from "frame mode", the default view, to "in frame mode". To switch back click the "Frame Mode" button,  . The standard, "Frame Mode" displays the scans merged over the CCS dimension. The "In Frame Mode" displays the scans merged over the retention time dimension.

9. FAIMS Data

Field asymmetric waveform ion mobility spectrometry (**FAIMS**) analyzes complex mixtures and separates the ions based on their mobility. PEAKS Online has the ability to load FAIMS data and analyze it using *de novo* sequencing, database search (PEAKS DB), database search with unexpected modification identification (PEAKS PTM), sequence variant identification (SPIDER) search, spectral library search, and labelled and label free quantification.

The analysis FAIMS mass spectrometry datasets contains the CV values used when collecting the spectra. The CV-related information can be found at the following places for a spectral Library analysis.

- The **Venn diagram** on the summary page.

  Venn diagram overlaps between different CVs in the search for all proteins, top proteins, and peptides are shown on the summary page.

- The additional **CV** column on the supporting peptide page
- The additional **CV** column in the exported *Protein-Peptide CSV*
- The additional **CV** column on the peptide page
- The additional **CV** filter on peptide page

  The filter for peptide page contains one additional option for CV, which selects the displayed scans based on the selected CV values

- The additional **CV** column in the exported *Peptide CSV*

Chapter 8. Label Free Quantification (LFQ)

1. Overview

Label free quantification can be used as a quick profiling tool to examine the relative abundance of proteins in large proteomic datasets. In PEAKS Online this quantification method is based on the relative abundance of peptide features detected in multiple samples. PEAKS Online performs feature detection separately on each sample, and uses an EM (expectation-maximization) algorithm to detects and deconvolute overlapped features. Importantly, features of the same peptide from different samples are reliably aligned together (i.e. matched between runs) using a high performance retention time alignment algorithm to improve overall protein quantification rates.

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 workflow is outlined below.

1. Refer to Chapter 5, Peptide De Novo Sequencing, Section 2, “*DeepNovo/De Novo* Analysis Workflow and Parameter Settings” , Step 1-6 to create a new project and add data.

2. Label Free quantification can be run at the start of a new project, or on PEAKS Database Search results or PEAKS Spectral Library Search results.

   - To run at the start of a new project, select the appropriate Peaks Q workflow after you have added your DDA or DIA data

   - To run Label Free Quantification on existing PEAKS Database Search results, choose the Create New Analysis from within the project and then select the "Peaks Q" workflow in the DDA section

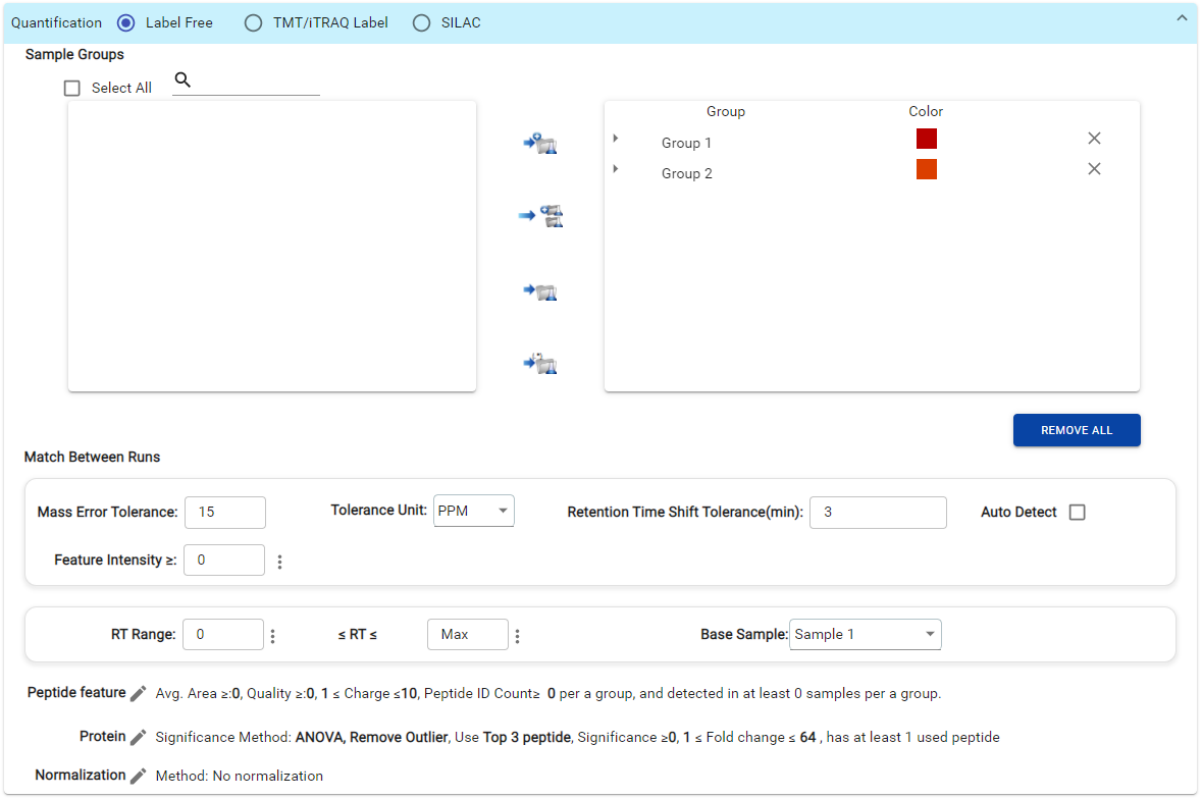
   - To run Label Free Quantification on an existing PEAKS Spectral Library Search results, choose the Create New Analysis from within the project and then select the "Quantification" workflow in the DIA section

3. Set Search Parameters.

   - If running LFQ on a DDA DB Search analysis, refer to Chapter 6, Peptide, PTM, and Mutation Identification (PEAKS DB, PEAKS PTM, and SPIDER), Section 2, “Database Search Workflow and Parameter Settings” for details.

   - If running LFQ on a DIA Spectral Library Search analysis, refer to Chapter 7, DIA Streamlined Identification, Section 3, “Spectral Library Search Workflow and Parameter Settings” for details.

4. Set Quantification Parameters: the LFQ-specific parameters will be shown at the bottom of the Project Submission page. Group your samples using the different add options (e.g. add to a new group, add to existing group, etc.). The names and colours for each groups can be edited by clicking on the name or coloured square, respectively. Next, set your Mass Error and Retention Time Shift tolerance, and other settings for match between runs. See Section 2 below for further details. Label-free quantification will be performed on the combined identification results from all samples in the analysis.



5. Thresholds and analysis parameters: Set Peptide Feature, Protein, and Normalization parameters by selecting the edit icon (pencil icon) next to each category. Refer to Chapter 6, Peptide, PTM, and Mutation Identification (PEAKS DB, PEAKS PTM, and SPIDER), Section 2.8, “PSM Filter” for details as well as Section 2 below.

6. Review project settings and parameters then click Submit button. Name the analysis (or use the default option) and click OK to start the analysis.

2. Setting Parameters

For more information on DB search parameters, refer to Chapter 6, Peptide, PTM, and Mutation Identification (PEAKS DB, PEAKS PTM, and SPIDER), Section 2, “Database Search Workflow and Parameter Settings” .

The following parameters are available when the Label Free Quantification workflow is selected.

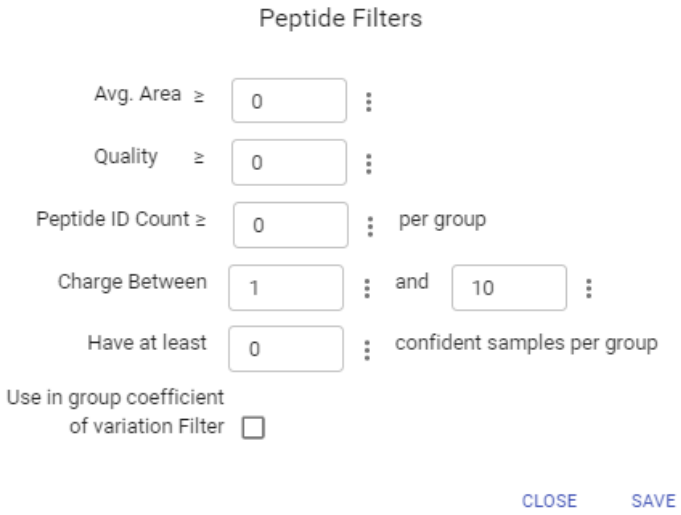
- **Mass Error Tolerance:** The mass shift of a precursor ion between different runs. This can be set similar to the precursor mass error tolerance used for database search.
- **Retention Time Shift Tolerance:** The retention time shift tolerance is the maximum elution time difference that is considered for the quantification of an identified peptide after RT alignment.
  - **Auto Detect:** Let PEAKS Online detect the optimal Retention Time Shift Tolerance for your data based on the observed trend.

- **CCS Error Tolerance (if applicable):** The maximum collision cross section (CCS) differential for features to transfer IDs.

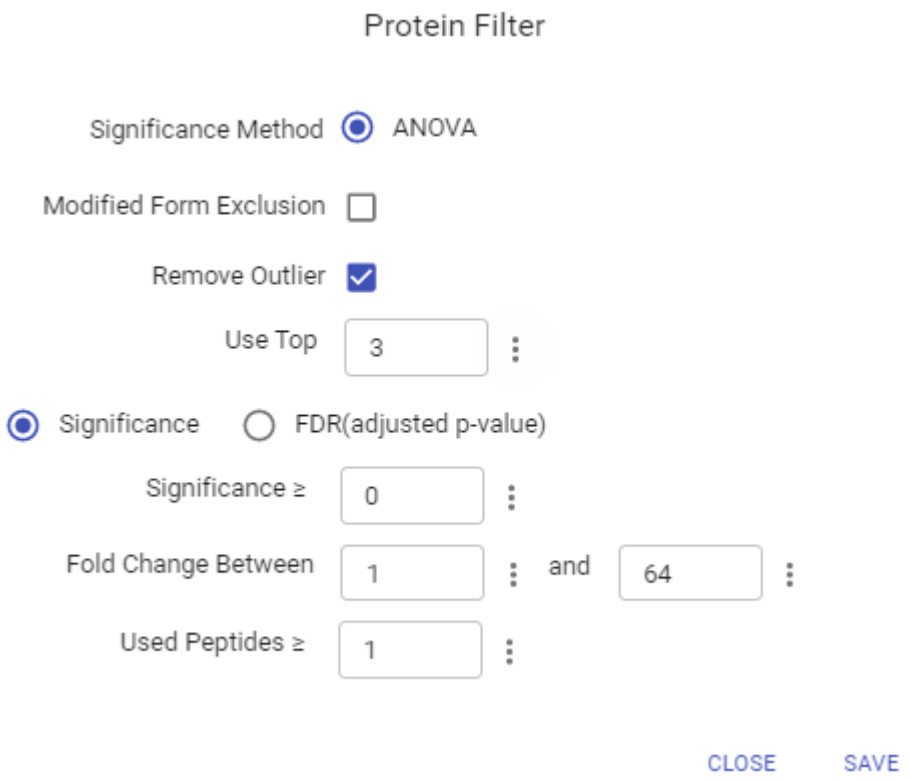
Note

This parameter is only available when the analysis contains timsTOF data.

- **Feature intensity**: Set a value for the minimum intensity a peptide feature must exhibit.
- **RT Range:** Feature vectors within the selected retention time range will be included in the results while everything else will be removed.
- **Base Sample:** The base sample will act as the reference. The peptide and protein abundance corresponding to this reference label will be the denominator in all of the abundance ratio calculations.
- **Peptide Feature** :



  - **Avg. Area:** The minimum average area of all peptide features associated with the peptide. Features with higher abundance have been shown to be more reproducible across replicates. Only peptide features with an average area above this threshold will be used for quantification.
  - **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.
  - **Peptide ID Count:** This filter sets the minimum limit for how many times a peptide must be identified within a group to be included in the LFQ results.
  - **Charge between:** Only peptides with feature vectors that all fall within this range will be used in protein quantification.
  - **Have at least ___ confident samples per group:** This filter sets the minimum number of sample(s) per group that the peptide feature has an abundance value. A peptide is more quantifiable when it is detected in more samples.
  - **Use in group Coefficient of Variation Filter** : For each feature vector (set of peptide features associated with a peptide), at least one group must have a coefficient of variation (CV) less than the CV threshold. The CV threshold is automatically selected by the software based on the current data set.
- **Protein Filter:**



- **Significance Method** : Protein abundance is used for the significance testing calculation. When there are replicates in each group, ANOVA is suggested to be used for significance testing. For moredetails on ANOVA**,** refer to the following paper: *"On the comparison of several mean values: an alternative approach ", Biometrika, 1951, 38(3/4): 330-336*

- **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 peptides for 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 for protein ratio calculation, 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

- **Remove Outliers** : This filter removes supporting peptides that have a different variation trend compared to others.

- **Use Top # peptides** : This is the number of quantifiable supporting peptide that will be used to calculate the protein profile. Three is the default setting.

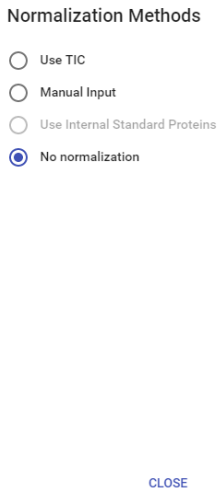
- **Significance:** Only protein groups with a significance above this threshold will be listed in the “Protein” view. The significance score is calculated as the -10log of the significance testing p-value (e.g. Significance score of 20 is equivalent to a p- value of 0.05). PEAKS provides the significance testing method: ANOVA.

- **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 below the lower threshold or above the upper threshold value will be listed in the "Protein" view.

- **Used Peptides:** Only protein groups that were quantified with this number of peptides (or more) will be listed in the "Protein" view.

- **Normalization** :

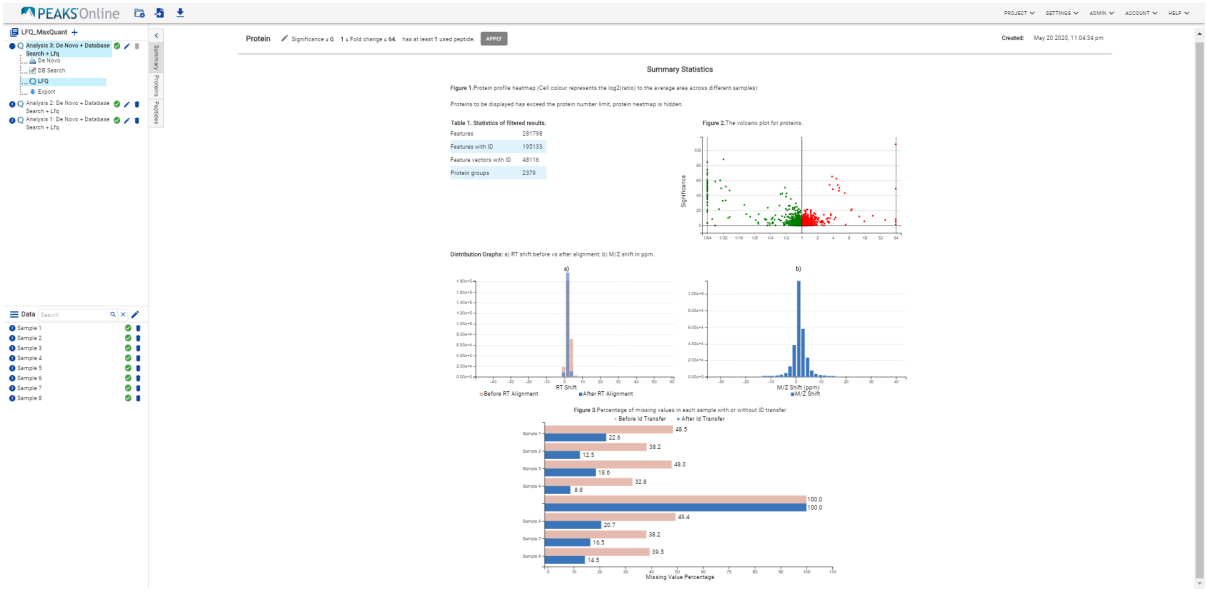


- **Use TIC:** By default, PEAKS Online uses a sliding-window method to calculate retention time-dependent normalization factors from the total ion current (TIC) of each sample. Normalized abundance for each sample is calculated from the raw abundance divided by the normalization factor for that sample.
- **Manual input:** This option performs the same default calculation as auto normalization ("Use TIC") but also allows for the manual change of each sample's expected ratios. This option should be used if the expected ratio is not 1:1 for each sample as the software attempts to normalize to a 1:1 ratio.
- **Use internal standard proteins:** This option displays a list of identified proteins (when a database or library search has been completed) that can be selected to create a normalization ratio. A search bar is included to enable the quick location of select proteins. Normalization by internal standard proteins is performed using local retention-time dependent normalization.
- **No Normalization:** All samples are automatically assigned Factors of 1.0.

- **Sample Groups:** LFQ requires sample grouping. All available samples are listed in the left window. After selecting a sample, it can be added to a new group by clicking the button or to an existing group by clicking the button and selecting that option from the drop-down menu. If you select multiple files, you can add each one individually to create multiple new group by clicking the button. If you want group files by delimiter or regEx, you can click the button. Samples can be removed individually or removed all at once by clicking the button.

3. Understanding the LFQ Results

After a PEAKS Online label free quantification is complete, the LFQ summary page will display automatically. There are also Protein and Peptide result tabs to view.



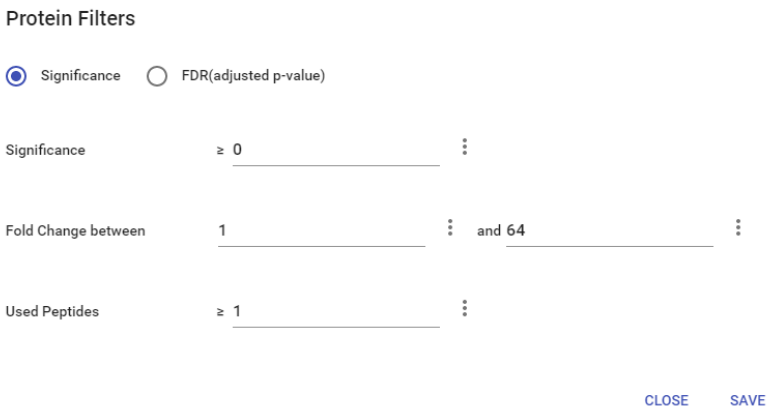
- **Summary:** The outline of the LFQ search results with statistics. This is the place to examine the overall performance of the experiment and adjust Protein filters.
- **Proteins:** The quantified proteins with a list of supporting peptide features for each protein.
- **Peptides:** The quantified peptides are listed in a table. Extracted ion chromatogram (XIC) and supporting peptide features are also displayed for each peptide.

3.1. Protein Filters

The results can be filtered using the Protein filter at the top of the Summary tab. Adjustments made here will affect the results displayed in the protein and peptide tabs.

3.1.1. Protein Filter

Click the ✎ button and the following parameter dialogue will pop up.



Filter settings used in the screenshot above means that no protein filter is applied, thus giving all quantifiable protein.

Alternately, parameters can be adjusted for additional analysis by selecting the analysis from the project list on the left hand side of the window, and then selecting Modify Parameters from the top right hand side of the Analysis Properties view. This will open all parameters for the analysis and allow for the modification of single or multiple settings. If changes are made using this method a new analysis is submitted and performed.

•

3.2. Summary View

The "LFQ Summary" view includes a result statistics table, a protein volcano plot, and an expression profile with candidate proteins clustered in a heat map. An "Analysis Description" can be added and displayed below the filter section.

3.2.1. Statistics of Filtered Results

Table 1 in the Summary page lists the filtered results of feature vectors and proteins.

3.2.2. The Volcano Plot for Proteins

A volcano plot combines a statistical test with the magnitude of the change enabling quick visual identification of those data-points that display large-magnitude changes that are also statistically significant. The volcano plot is displayed as Figure 2 in the results, and plots significance versus fold-change of the quantified proteins. Volcano plots show two thresholds:

- **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). 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.

Circles for the proteins that are above the set significance and fold change thresholds will be displayed in colour. Scrolling the cursor over a circle will show the protein info.

3.2.3. Protein Profile Heatmap

The heatmap is shown in Figure 1 of the results, and displays all the protein groups that pass 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 within a group. 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. If the number of proteins is too high, the heat map will not show in this case.

3.2.4. RT Shift & M/Z Shift

In Figure 3. Distribution Graphs: a) RT shift before vs after alignment. It displays the retention time shift before and after alignment

Distribution Graphs: b) plots the distribution of the m/ z shift (in ppm) of the data

3.2.5. Percentage of missing values in each sample with or without ID transfer

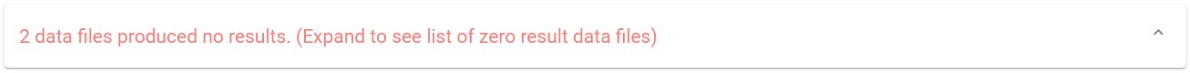
Figure 4 shows the percentage of missing values before and after ID transfer for each sample. A missing value before ID transfer is defined as a null value in a sample where a peptide was assigned a feature in another sample with an identification. Missing values after ID transfer are null values when no feature could be found in a sample based in identifications found in other samples. This data is influenced by the Match Between Runs settings selected by the user when the workflow was setup.

3.2.6. Protein Area Coefficient of Variance.

Figure 5 shows a histogram of all protein covariance values. For this chart, protein covariance is calculated within a group.

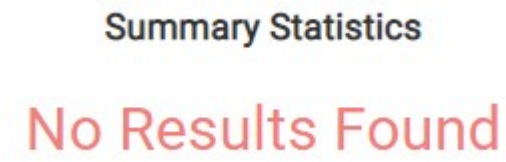
3.2.7. Low Quality Data Warnings

If PEAKS Online detects any data files producing very few or no results a warning section will appear at the top of the Summary Page indicating the number of data files producing poor or no results.



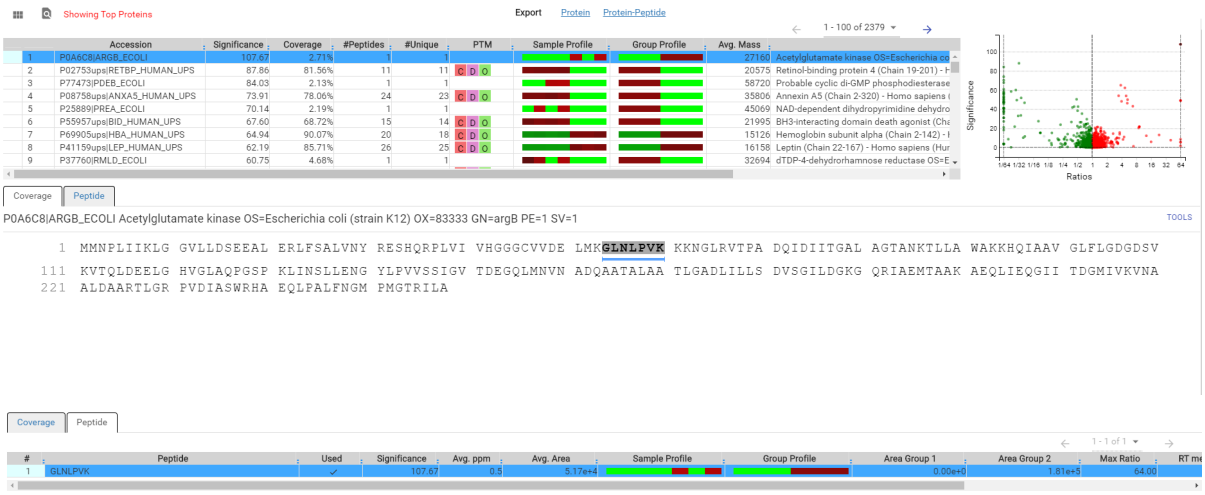
This section can be expanded to display a list of data files in the analysis that are producing poor or no results.

If there were no results found in any of your data files, a warning message stating "No Results Found" will be displayed instead of the usual Summary Page figures and tables.



3.3. Protein View

The LFQ Protein page lists all the quantified proteins that pass the filter set on the LFQ Summary page, characterizes each protein at the amino acid level, and lists the supporting peptide features of each protein.



3.3.1. Protein Table

The quantified proteins are listed in the protein table with parsimonious 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:

- **Significance:** The significance score is calculated as the -10log of the significance testing p-value (e.g. Significance score of 20 is equivalent to a p- value of 0.05). PEAKS provides the significance testing method: ANOVA.

- **Sample Profile:** The protein abundance among the samples is depicted as a heat map. Hold the cursor on a profile to view the sample channels, abundances, and ratios with respect to the base sample.

- **Group Profile:** The protein abundance among the groups is depicted as a heat map. The Group Profile is determined by calculated the total abundance (area) of supporting peptides within a group. Hold the cursor on a profile to view the group channels, abundances, and ratios compared to the group that contains the base sample.

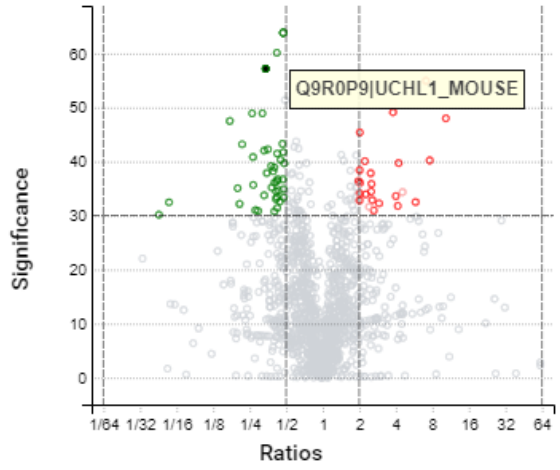
Refer to Chapter 6, Peptide, PTM, and Mutation Identification (PEAKS DB, PEAKS PTM, and SPIDER) Section 3.4.2, “Protein Table” for more details of other entries of the protein table.

3.3.2. Protein Volcano Plot

The volcano plot facilitates identification of the significant proteins. This interactive volcano plot enables the selection of a protein in the table. 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 less abundant versus the reference group while red points are more abundant versus the reference group. Grayed points represent proteins currently filtered out of the result table based on the Peptide Filters set on the Summay tab.

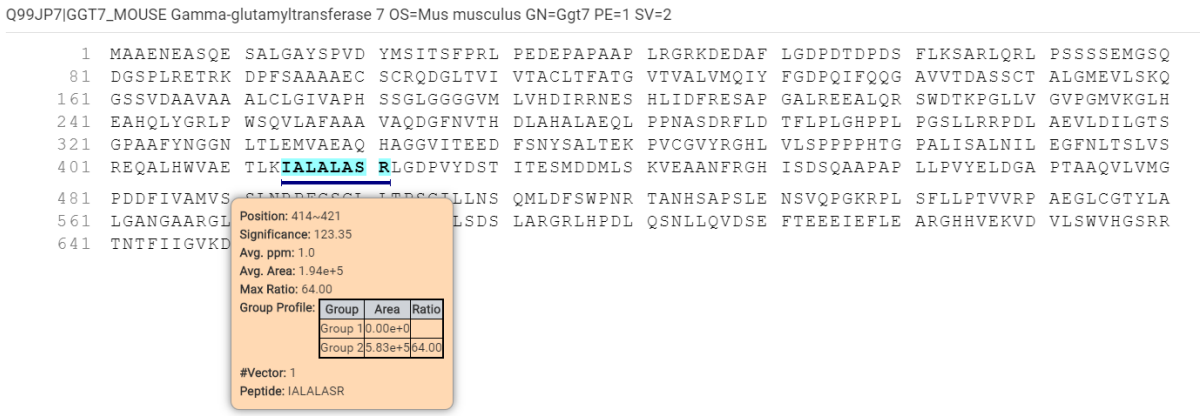
Red and green protein markers can be clicked and selected. Upon clicking a point on the volcano plot, the corresponding protein in the Protein Table will be highlighted and more details are displayed below the Protein Table.

- **Zoom in/out from Volcano Plot:** Place the cursor on the volcano plot and drag a rectangle over the area of interest to Zoom in. By double clicking on the volcano plot you can zoom out and view the volcano plot with the default dimensions.



3.3.3. Coverage view

The coverage view characterizes the selected protein sequences at the amino acid level. Refer to Chapter 6, Peptide, PTM, and Mutation Identification (PEAKS DB, PEAKS PTM, and SPIDER), Section 3.4.4, “Protein Coverage” for details.



3.3.4. Peptides view

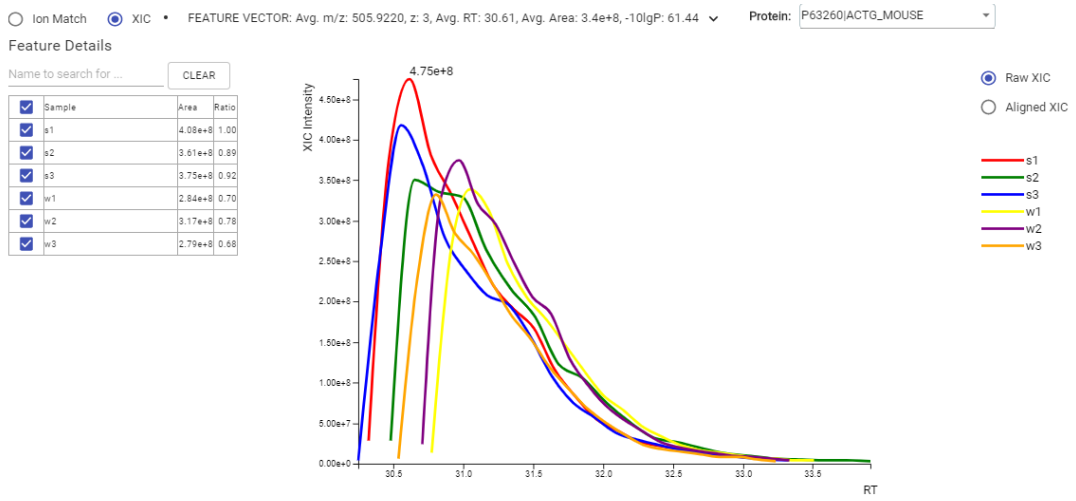
The supporting peptides assigned to the protein are shown in a separate tab beside the Protein Coverage view. The sequence can be clicked and a pop-up window will show up to display the feature details.

This table contains the following information of the quantifiable support peptides for the selected protein:

- **Peptide:** The amino acid sequence of the peptide, as determined in PEAKS Database or Spectral Library search. A modified residue is followed by a pair of parentheses enclosing the modification mass. All non-unique peptides and those not passing the filter will be excluded automatically.
- **Used:** Shows whether this peptide is used to calculate the protein abundance. The top three peptides with highest abundance are used to calculate the protein abundance.
- **Significance:** The -10lg of a p-value represents the likelihood that the observed change between conditions is caused by random chance. The peptides are first separated into groups based on similar quality score. Significance is then calculated for each quality score group, or bin, using a two-tail T-test that assumes log normal distribution but does not assume equal variance. The -10log p-value is then calculated from this result and then displayed in the significance 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 area under the XIC curve. The average area of a peptide is the sum of the average area of all features associated with that peptide.
- **Sample Profile:** The peptide abundance among the samples is 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**.**
- **Group Profile:** The peptide abundance among the groups is 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**.**
- **Group Area:** This column shows the peptide area for a group. It is calculated by summing up the areas of feature vectors within a group
- **Max Ratio:** This shows the maximum of the fold change values compared to the reference sample.
- **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.
- **#Vector:** This shows the number of quantifiable feature vectors of a peptide.
- **Start:** This shows the protein position of the first residue of the peptide.
- **End:** This shows the protein position of the last residue (inclusive) of the peptide.

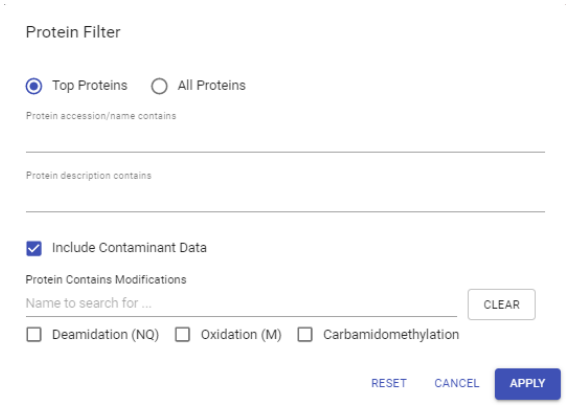
- **PTM:** The types and the numbers of modifications present in the peptide shown using color-coded icons. Scroll over each icon to see the modification name and mass.

- **Feature Detail:** Double click on a supporting peptide to go to the peptide page where the eXtracted Ion Chromatogram (XIC) curves are displayed (as seen below).



3.3.5. Protein View Filter

Click the button on top-left of the protein table to open the protein filter and control which proteins to display.



- **Top Proteins**: If this is selected only the proteins with the highest number of unique peptides supporting them (within their protein group) will be displayed.

- **All Proteins**: If this is selected all proteins that satisfy the parameters set for PEAKS Protein Score(-10lgP) and number of unique peptides will be displayed.

- **Protein Sequence contains**: Enter a protein sequence, only proteins that contain that sequence will be displayed in the protein table

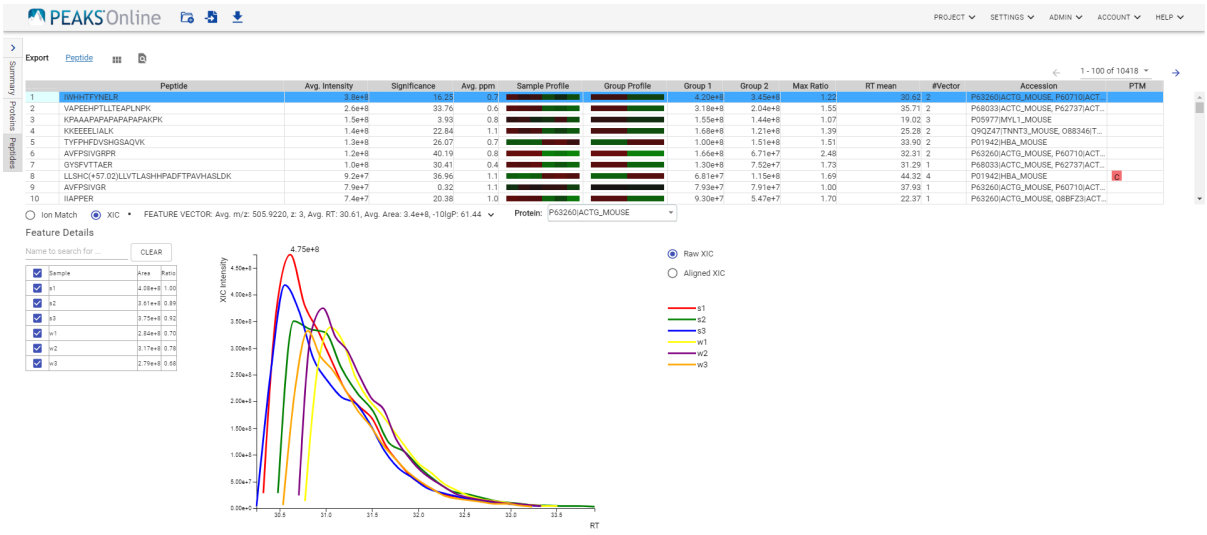
- **Protein description contains:** Enter a protein description, only proteins that contain that text within their description column will be displayed in the protein table

- **Include Contaminant Data:** Check this to display contaminants that were identified during a PEAKS DB search (if a contaminant database was specified).

- **Protein Contains Modifications**: Check boxes are provided for each identified modification.

3.4. Peptide View

Quantified proteins are displayed in the **Peptide View** along with their detailed information. The peptides are grouped together based on quantifiable feature vectors with the same sequence.



3.4.1. Peptide Table

The peptide table contains the list of quantifiable peptides along with relevant information including the accession information for the protein they support.

The view is similar to the support peptide table described in Section 3.3.4, “Peptides view” . The Peptide table entries may be filtered by using the **Peptide View Filters** on top of the peptide table, for more information refer to Section 3.5.1, “Peptide View Filters” in Chapter 6, Peptide, PTM, and Mutation Identification (PEAKS DB, PEAKS PTM, and SPIDER).

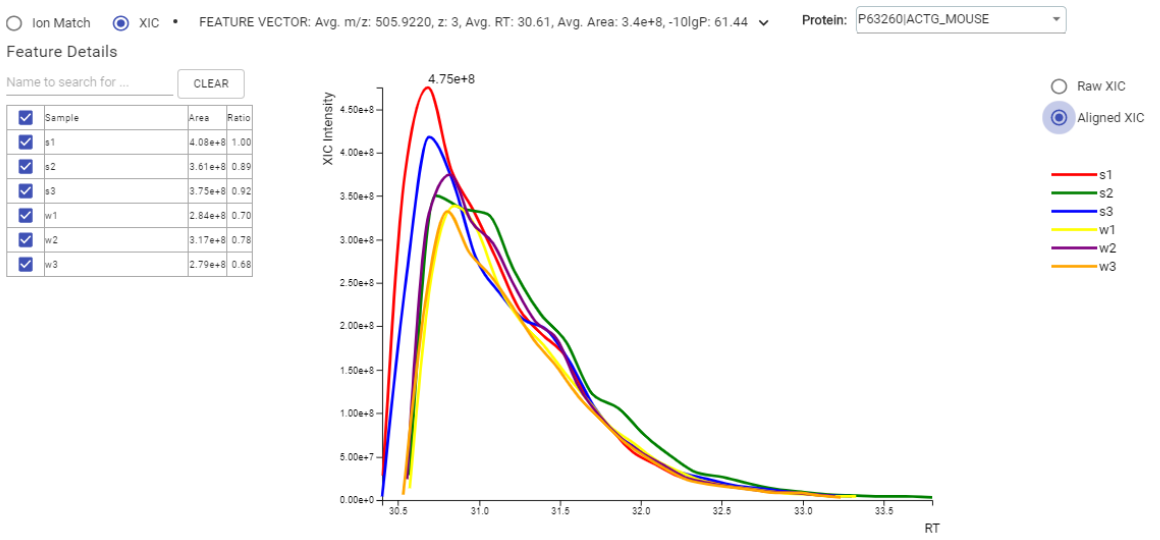
3.4.2. Feature Details

The **Feature Details** contains the eXtracted Ion Chromatogram (XIC) chart and Ion Match.

- **Peptide Menu:** It is under the peptide table. The peptide menu displays the information for the peptide that is being displayed in the spectrum. If a feature is associated to more than one tandem scan, click the "FRACTION" button to view the Spectrum Annotation of the other PSMs.

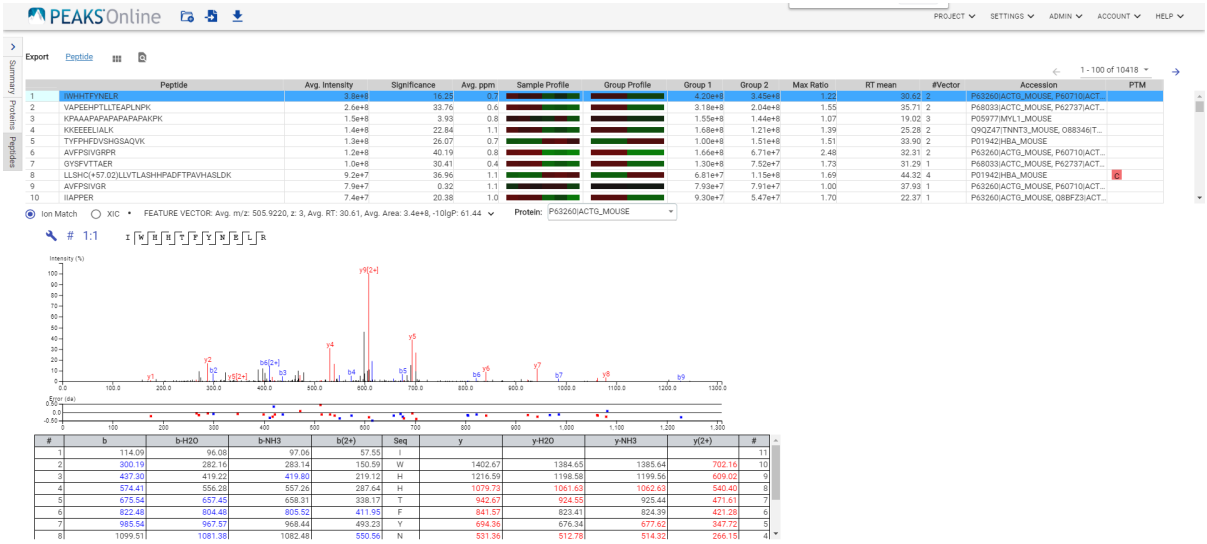
3.4.2.1. eXtracted Ion Chromatogram (XIC) chart

The eXtracted Ion Chromatogram (XIC) chart displays the shape of the selected peptide feature vector over the retention time range associated with the identification. The table on the left displays both the area and the ratio of the features in the highest average abundance feature vector. This table also works as a control to select the features that will be used to draw the XIC plot.



3.4.2.2. Ion Match

This section displays the spectrum annotation and other information about the peptide selected in the Peptide table. When multiple spectra are matched to the peptide, only the information about the top-scoring PSM is shown by default. The interface of this section is identical to the Spectrum Annotation in the PEAKS De Novo result. For detailed instructions, refer to Section 3.5, “Spectrum Annotation” in Chapter 5, Peptide De Novo Sequencing.



3.4.2.2.1. DIA Ion Match

When LFQ analysis are run on results from a DIA Identification search, there are two additional figures included in the Ion Match pane: The Precursor Profile XIC and the Fragment Ion XIC. For detailed instructions on these figures, refer to Section 4.5.6, “Peptide-Spectrum Matches” in Chapter 7, DIA Streamlined Identification.

4. Ion Mobility Spectrometry (IMS) Data

Ion mobility spectrometry (IMS), including Field asymmetric waveform ion mobility spectrometry (**FAIMS**) and Trapped ion mobility spectrometry (**tims**) analyzes complex mixtures and separates the ions based on their mobility. PEAKS Online has the ability to load IMS data and analyze it using *de novo* sequencing, database search (PEAKS DB), database search with unexpected modification identification (PEAKS PTM), sequence variant identification (SPIDER) search, spectral library search, and labelled and label free quantification.

The analysis result for IMS dataset contains the additional information for CV values compared to the non-IMS dataset. The CV-related information can be found at the following places for a LFQ analysis.

- The **Venn diagram** on the summary page.

  Venn diagram overlaps between different CVs in the search for all proteins, top proteins, and peptides are shown on the summary page.

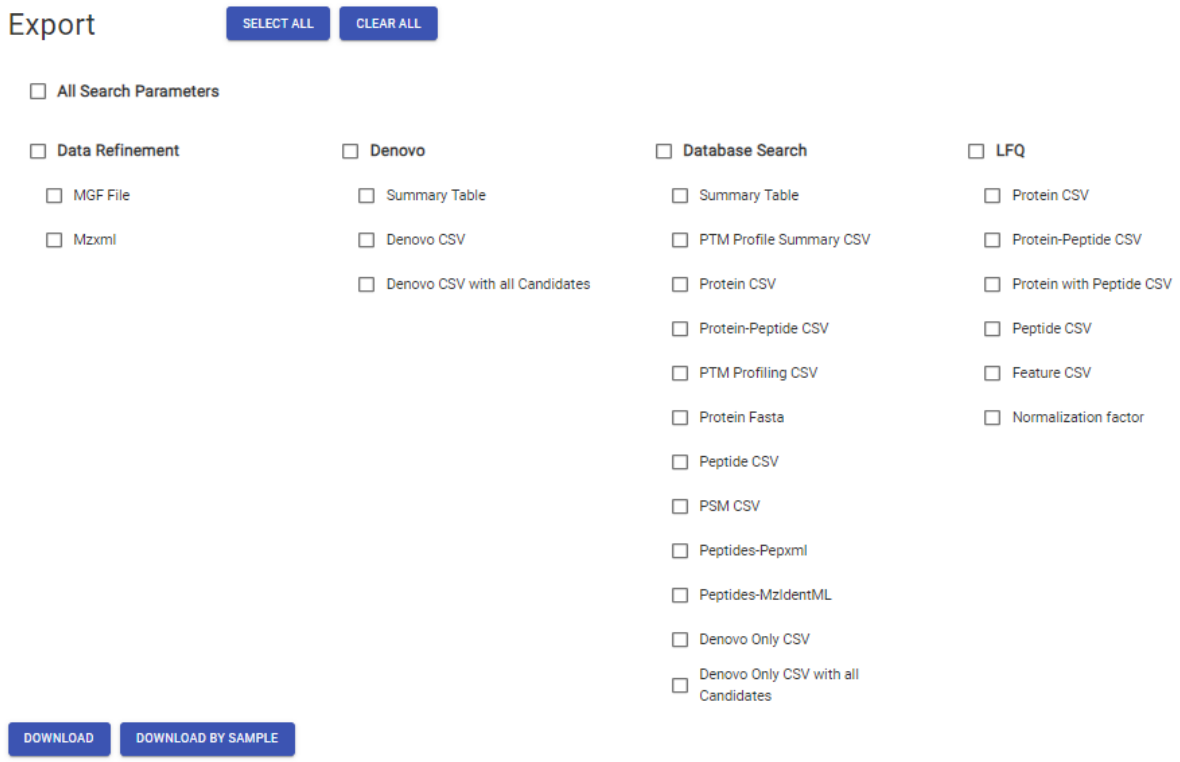
- The additional **CV** column on the supporting peptide page
- The additional **CV** column in the exported *Protein-Peptide CSV*
- The additional **CV** column on the peptide page
- The additional **CV** filter on peptide page

  The filter for peptide page contains one additional option for CV, which selects the displayed scans based on the selected CV values

- The additional **CV** column in the exported *Peptide CSV*

4.1. Result Exporting

PEAKS LFQ results can be exported to various text formats in the Export page



- **Proteins CSV:** These are the quantified proteins that pass the filters set on the "Summary" page and their associated details.
- **Protein-Peptide CSV:** These are the supporting peptides of the quantified proteins and their associated details.
- **Peptide CSV:** A complete list of all quantified peptides that pass the peptide feature filters set on the "Summary" page and their associated details.
- **Feature CSV:** All predicted peptide features and their associated quantification information. Both identified and unidentified peptide features are included. To include unidentified peptide features, set the Peptide Id Count equal to 0 in the Peptide feature filter.

- **Protein with Peptide CSV:** A list of supporting peptides for the quantified proteins and their associated details.
- N**ormalization Factor CSV:** A list of normalization factors used for LFQ normalization. For analyses using local normalization methods such as TIC or internal standard proteins, global normalization factors will display as N/A.

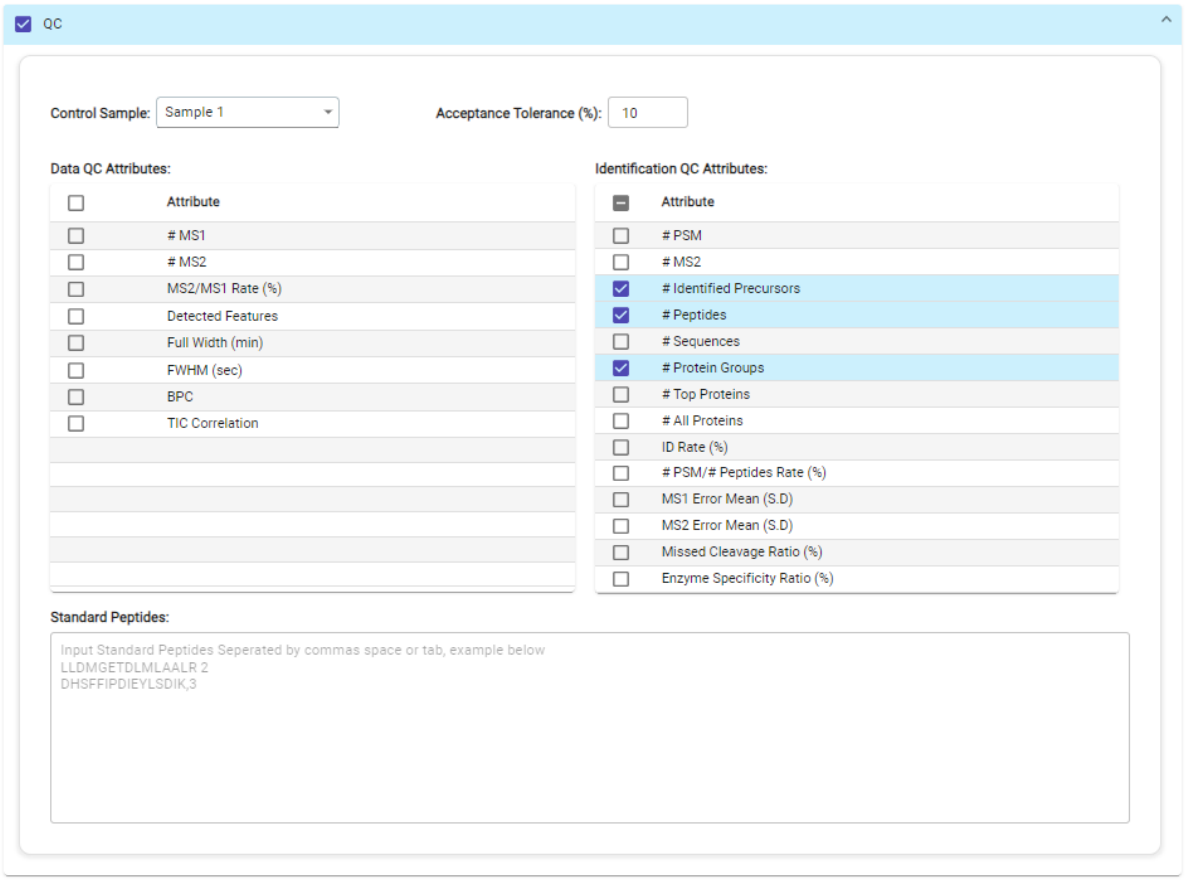
Chapter 9. PEAKS Quality Control

1. Overview

PEAKS Quality Control offers a sophisticated and systematic QC analysis. In order to help with verifying the integrity and validity of data being processed through PEAKS Online an optional QC (Quality Control) task can be run on top of PEAKS protein/peptide identification and label-free quantification workflows. This task will provide statistical information about your data to allow for informed decisions on the data's quality.

2. QC Parameters

To enable PEAKS QC to be run on top of your PEAKS identification or quantification results, ensure that the QC section is checked at the bottom of the Project Parameters page.



Before submitting the analysis ensure the following parameters are set properly:

- **Control Sample:** Specify the sample to be used as the Control Sample for quality control checking. All other samples will have their quality control values compared to the corresponding value from the control sample and will be considered failing if the variance is outside of the Acceptance Tolerance.
- **Acceptance Tolerance (%):** This is acceptable amount of variance a sample can have from the control sample (or average if specified). Any quality control metric that falls outside of this tolerance level will be labelled as failed (red) in the quality control result views. If Acceptance Tolerance is set to 0%, then each attribute will use the value from selected control sample as the threshold.
- **Standard Peptides:** Input amino acid sequences with analogue charge of your standard peptides separated by commas, spaces, or tabs.

2.1. Quality Control Attributes

Different quality control attributes can be analyzed in PEAKS online. These attributes are divided into data, identification and quantification sections. Quantification attributes are only visible if a PEAKS Q workflow is selected.

2.1.1. DDA QC Attributes

- **Data QC Attributes:** #MS1, #MS2, MS2/MS1 Rate (%), Detected Features, Full Width (min). FWHM (sec), BPC, TIC Correlation
- **Identification QC Attributes:** #PSM, #MS2, # Identified Features, #Peptides, #Sequences, #Protein Groups, #Top Proteins, #All Proteins, ID Rate (%), #PSM/# Peptides Rate (%), MS1 Error Mean (S.D), MS2 Error Mean (S.D), Missed Cleavage Ratio (%), Enzyme Specificity Ratio (%), Ion Mobility Error Mean (timsTOF only).
- **Quantification QC Attributes :** #Quantified Peptides, #Quantified Protein Groups, #Quantified Features, Missing Values, Feature Correlation.

2.1.2. DIA QC Attributes

- **Data QC Attributes:** #MS1, #MS2, Detected Precursors, Full Width (min). FWHM (sec), BPC, TIC Correlation
- **Identification QC Attributes:** #PSM, #MS2, # Identified Precursors, #Peptides, #Sequences, #Protein Groups, #Top Proteins, #All Proteins, ID Rate (%), MS1 Error Mean (S.D), MS2 Error Mean (S.D), Missed Cleavage Ratio (%), Enzyme Specificity Ratio (%), Ion Mobility Error Mean (timsTOF only).
- **Quantification QC Attributes :** #Quantified Peptides, #Quantified Protein Groups, #Quantified Precursors, Missing Values, Feature Correlation.

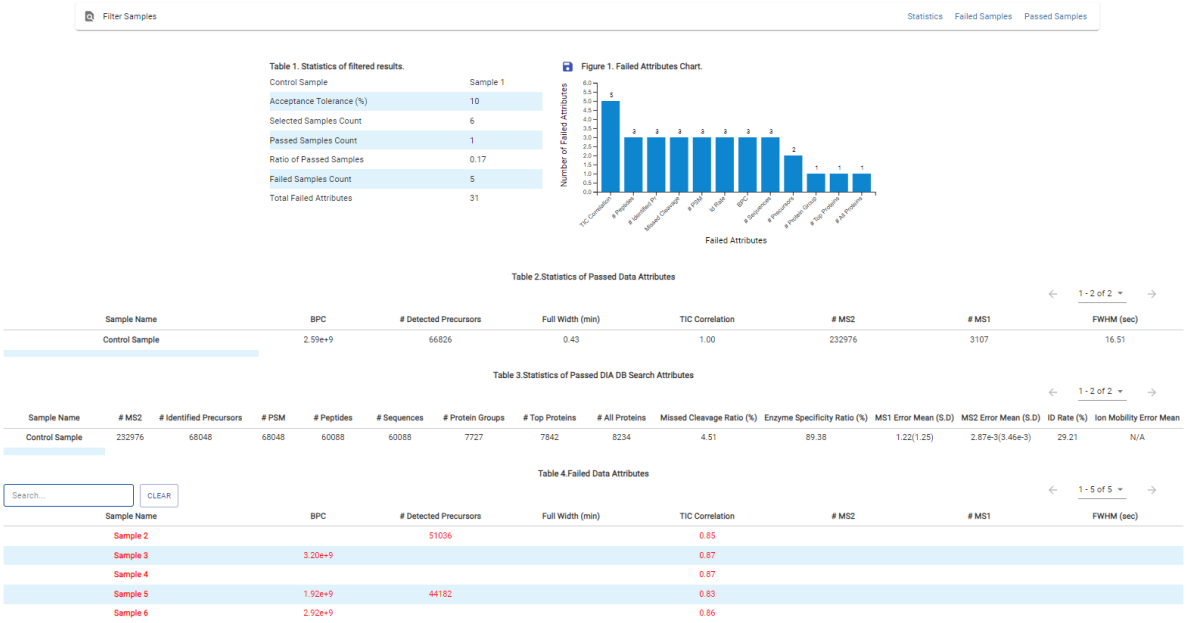
3. PEAKS QC Results

PEAKS QC results consist of Summary, Data, Identification, Quantification and Standard Peptides tabs.

3.1. QC Summary Tab

The QC summary tab contains:

- **Table 1: Statistics ofFiltered Results.** Contains the Statistic of filtered results including information of the Control sample, Acceptance tolerance, Total, Passed and Failed sample counts, Ratio of Passed samples and the total number of Failed Attributes. If no failed attributes present, the message will be displayed "All Attributes Pass!" and only Table 1 and 2 are displayed. Specific selection of QC attributes will also affect if sample is considered Passed or Failed.
- **Figure 1: Failed Attributes Chart.** For failed attributes, a bar chart will be displayed indicating the number of samples where specific attributes failed.
- **Table 2-4**: **Statistics of Passed Attributes**. Statistics for passed data, identification and quantification attributes are summarized in separate tables for Control Sample and for the Average of Passed Samples. Click on "Filter Samples" to select Samples and additional QC Attributes to be displayed and considered in average calculations. For a sample to be included in the average of passed samples, it must have no failing attributes in any category.
- **Table 5-7: Statistics of Failed Attributes.** Statistics for failed data, identification and quantification attributes are summarized in separate tables with entries for each failed sample. Click on "Filter Samples" to select Samples and additional QC Attributes to be displayed and considered in failing attributes tables.



3.2. QC Data Tab

This tab shows several figures that visualize the data quality control measures.

The Data QC summary table displays an overview of identification-related quality control measures. Any value in the table that exceeds the acceptance tolerance, set on the parameters page, will be coloured red.

1 - 20 of 21

SAMPLE NAME	# MS1	# MS/MS	# PRECURSORS	FULL WIDTH (MIN)	FWHM (SEC)	TOTAL BPC
P1-A10	925	30525	63528	0.15	8.28	2.105e+9
P1-G4	929	30657	62065	0.15	8.28	3.637e+9
P1-G6	927	30591	63148	0.15	8.27	3.396e+9
P2-A6	926	30558	62300	0.15	8.28	2.630e+9
P2-C11	928	30624	57903	0.15	8.28	2.928e+9
P2-C5	929	30657	60535	0.15	8.28	3.002e+9
P3-D10	931	30723	59086	0.15	8.28	4.474e+9
PG-G10	928	30624	59056	0.15	8.28	3.456e+9
P3-G3	927	30591	57269	0.15	8.28	3.314e+9
Sample 11	931	30723	57604	0.15	8.28	3.644e+9
P4-B4	930	30690	55501	0.15	8.28	4.049e+9
P4-C2	929	30657	53171	0.15	8.28	3.538e+9
P4-F10	928	30624	49567	0.19	8.29	2.238e+9
P4-F7	926	30558	49675	0.19	8.29	2.280e+9
P5-B12	929	30657	47300	0.19	8.29	1.674e+9

The columns in this table are:

- **Sample Name:** Name of the sample represented by this row. The table will always be sorted by this column (in sample order) with the control sample at the top and the average values at the bottom. The selected control sample will have its name displayed in bold in this column.
- **All Attributes:** This column displays whether samples either pass or fail quality control. This considers all attributes selected from the summary page filter menu, not only those in the data attribute list.
- **# MS1 Scans:** Total number of MS1 scans in the sample.
- **# MS2 Scans:** Total number of MS2 scans in the sample.
- **# MS2/MS1 Rate (%):** Total number of MS2 scans per MS1 scan from the sample. Displayed in percentage (%). (DDA Only).
- **Detected Precursors/Features:** Total number of precursors or features for DIA and DDA respectively.
- **Full Width (min):** Full peak width at base in minutes.

- **FWHM (sec):** Full width at Half Maximum of peak in seconds.
- **TIC Correlation:** The Pearson correlation coefficient with respect to points on the control sample TIC
- **BPC:** TIC at 80% of other TICs are above it.

3.2.1. Data QC Figures

Figure 1 will display a visualized summary of the selected column in the Data QC summary table, which will be highlighted in blue. By default the Detected Features/Precursors column will be selected, to select a new column click on the header row of the column you wish to select.

The dotted lines displayed on the chart represent the minimum and maximum values of the acceptance range (determined by the acceptance tolerance set on the parameters page). Any bar with a height that falls within these two dotted lines represents a value within the acceptance tolerance. Any bar with a height below or above these lines represents a value outside of the set acceptance tolerance.

Note

For analyses with greater than 30 samples, a line chart will be displayed instead of a bar chart. In this case, users may place mouse over the graph and zoom in and out to show more or less samples along the x-axes.

Figure 1. Features. Total number of features found in every sample

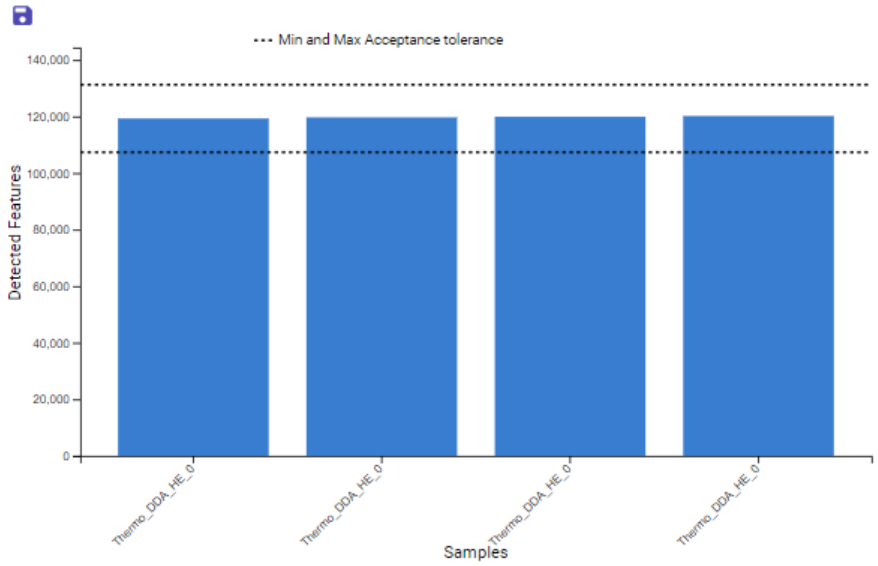


Figure 2. shows the TIC charts for selected samples side-by-side on the same axis. Up to 5 samples are selectable for display at once from a list beside the figure. By default the first 3 samples are selected.

Figure 2. Sample TICs.

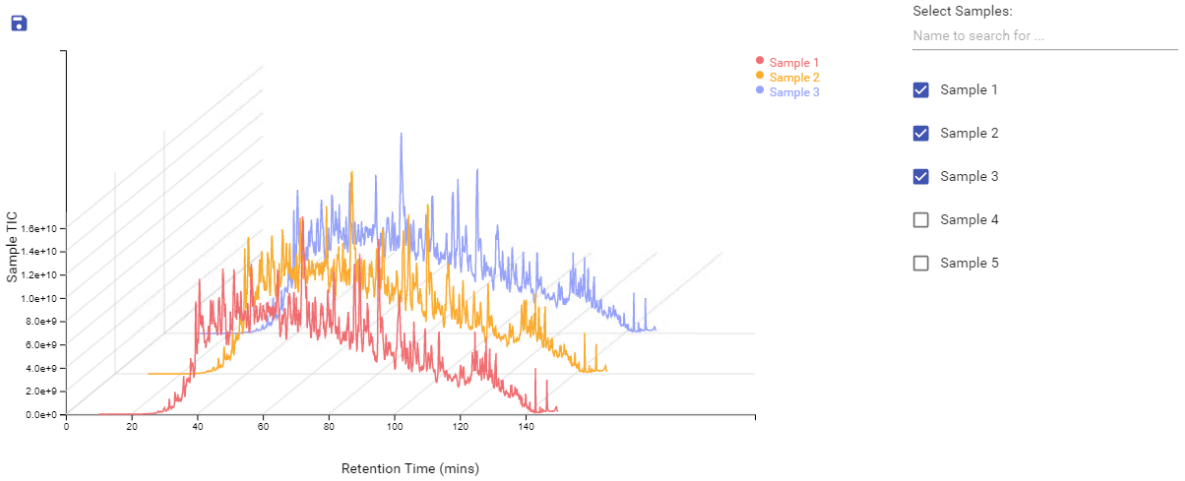
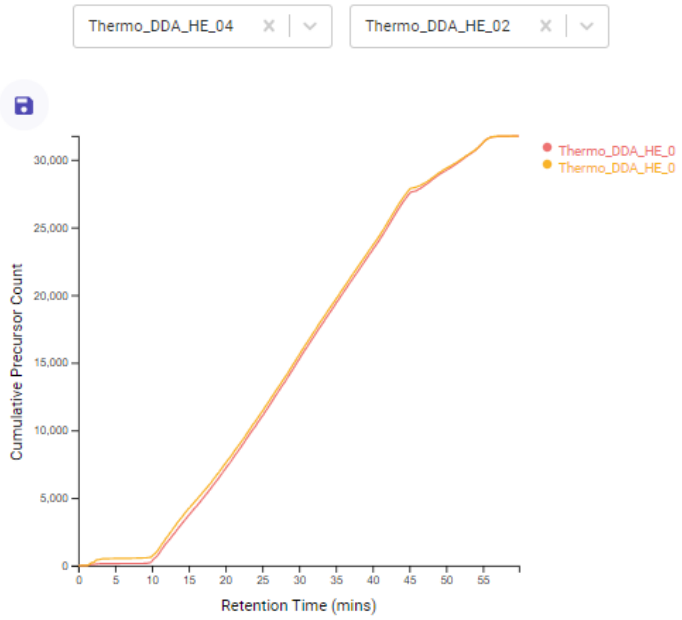


Figure 3. shows the cumulative precursor count over retention time. Only 2 samples can be displayed at once for pairwise comparison. These samples can be selected from the drop down menus above the figure.

Figure 3. Cumulative Precursor Count. Shows precursor count over retention time



3.3. QC Identification Tab

This tab shows several figures that visualize the data quality control measures.

The Data QC summary table displays an overview of identification-related quality control measures. Any value in the table that exceeds the acceptance tolerance, set on the parameters page, will be coloured red.

1 - 5 of 5

SAMPLE NAME	ALL ATTRIBUTES	# MS2	# IDENTIFIED FEATURES	# PSM	# PEPTIDES	# SEQUENCES	# PROTEIN GROUP	# TOP PROTEINS	# ALL PROTEINS	MISSED CLEAVAGE RATIO (%)	ENZYME SPECIFICITY RATIO (%)	# PSM/# PEPTIDES RATE (%)	MS1 ERROR MEAN (S.D)	MS2 ERROR MEAN (S.D)	ID RATE (%)
Thermo_DDA_HE_01	Pass	31611	21880	25710	19340	19207	3201	3280	3704	8.19	97.24	1.33	1.35(1.45)	2.75e-3(3.73e-3)	81.33
Thermo_DDA_HE_02	Pass	31759	22194	25970	19620	19472	3228	3311	3718	8.27	97.12	1.32	1.35(1.45)	2.76e-3(3.73e-3)	81.77
Thermo_DDA_HE_03	Pass	32058	22127	25881	19666	19538	3255	3333	3892	8.92	97.06	1.32	1.33(1.45)	2.86e-3(3.73e-3)	80.73
Thermo_DDA_HE_04	Pass	31767	22299	26025	19803	19665	3283	3350	3773	8.85	97.03	1.31	1.33(1.46)	2.85e-3(3.72e-3)	81.92
Average	Pass	31798	22125	25896	19607	19470	3241	3318	3771	8.56	97.11	1.32	1.34(1.45)	2.80e-3(3.73e-3)	81.44

The columns in this table are:

- **Sample Name:** Name of the sample represented by this row. The table will always be sorted by this column (in sample order) with the control sample at the top and the average values at the bottom. The selected control sample will have its name displayed in bold in this column.
- **All Attributes:** This column displays whether samples either pass or fail quality control. This considers all attributes selected from the summary page filter menu, not only those in the identification attribute list.
- **# MS2** Total number of MS2 scans in the sample.
- **#Identified Features/Precursors:** The number of identified features/precursors in the sample for DDA and DIA respectively.
- **# PSM:** The number of peptide-spectrum matches..
- **#Peptides:** Total number of peptides found ignoring differences with leucine (L) and isoleucine (I) amino acids. Peptides with the same primary sequence but different PTMs are counted separately.
- **#Sequences:** Total number of different peptide backbone sequences found ignoring differences with leucine (L) and isoleucine (I) amino acids. Peptides with the same primary sequence but different PTMs are counted as one sequence..
- **#Protein Groups:** PEAKS Online groups proteins identified by a common set of peptides. This number shows the number of protein groups in the filtered result based on All proteins.
- **#Top Proteins:** The number of proteins that 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 included here.
- **#All Proteins:** Total number of proteins identified in the sample.
- **Missed Cleavage Ratio (%):**The percentage of missed cleavages in a sample based on the number of expected cleavage sites as per the selected enzyme.
- **Enzyme Specificity Ratio (%):** The percentage of cuts that took place at expected cleavage sites as per the selected enzyme.
- **MS1 Error Mean (S.D):** PSM precursor error mean (ppm). Standard deviation is shown in brackets.
- **MS2 Error Mean (S.D):** Fragment Ion Error mean (Da). Standard deviation is shown in brackets.
- **ID Rate (%):** The percentage of #MS2 scans which produce a PSM.

Columns which are only present for DDA data:

- **#PSM/#Peptide Rate:** The #PSM divided by the number of identified peptides in a sample.

3.3.1. Identification QC Figures

Figure 1 will display a visualized summary of the selected column in the Identification QC summary table, which will be highlighted in blue. By default the # Protein Groups column will be selected, to select a new column click on the header row of the column you wish to select.

The dotted lines displayed on the chart represent the minimum and maximum values of the acceptance range (determined by the acceptance tolerance set on the parameters page). Any bar with a height that falls within these two dotted lines represents a value within the acceptance tolerance. Any bar with a height below or above these lines represents a value outside of the set acceptance tolerance.Place mouse over the graph and zoom in and out to show all or less samples/sample names on the x-axes

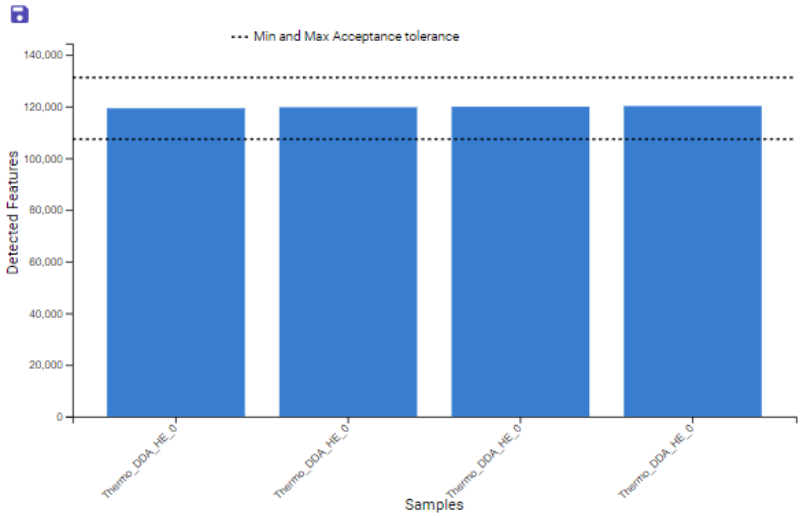
Note

For analyses with greater than 30 samples, a line chart will be displayed instead of a bar chart.

Note

For DIA spectral library searches without protein inference, #Peptides will be the default selected column.

Figure 1. Features. Total number of features found in every sample



Figures 2 and 3. shows sample reproducibility charts. These charts display overlap in Protein and Peptide identifications respectively. By default, only 2 samples are shown for pairwise comparison. Using the drop down menu above the figure, users can select different samples for comparison.

Figure 2. Protein Sample Reproducibility. Protein Overlap between selected Samples

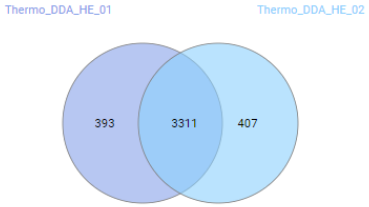


Figure 3. Peptide Sample Reproducibility. Peptide Sequence Overlap between selected Samples

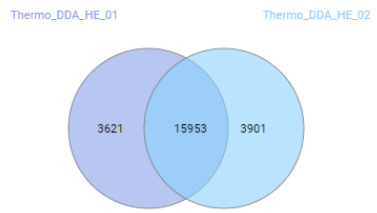


Figure 4. displays violin plots of the PSM precursor (ppm) and fragment ion error (Da) respectively.

Figure 4. MS1 & MS2 Mass Error Distribution.

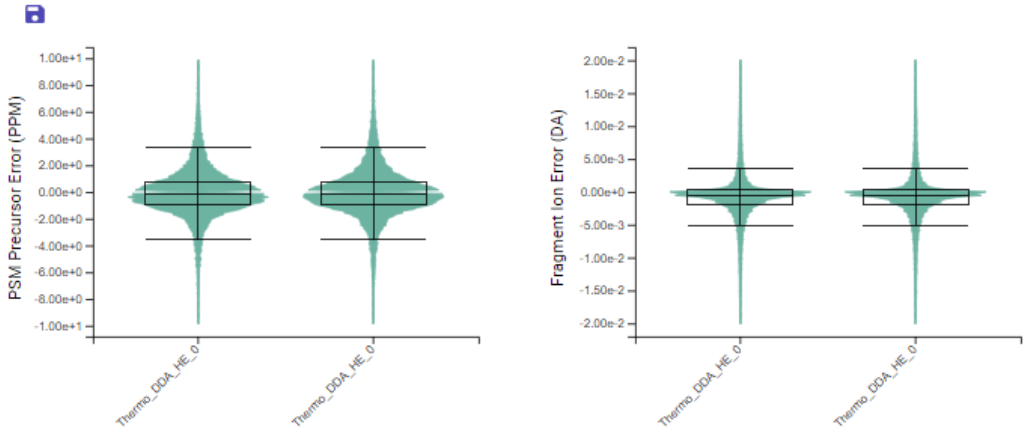
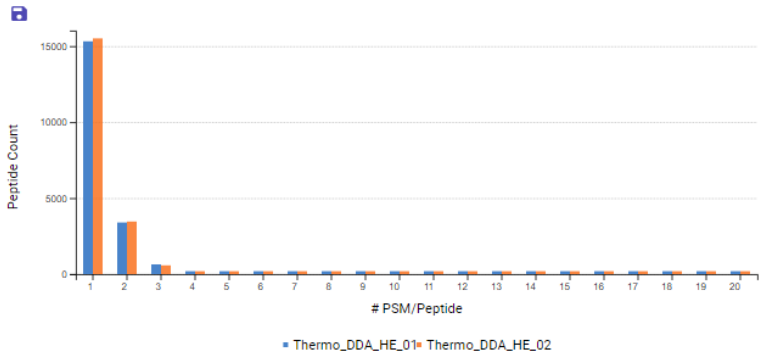


Figure 5. shows the peptide frequency distributions based on the number of PSMs per peptide identification at the individual sample level.

Figure 5. Peptide Frequency Distribution.



3.4. QC Quantification Tab

This tab shows several figures that visualize the label-free quantification quality control measures.

The LFQ quantification QC summary table displays an overview of identification-related quality control measures. Any value in the table that exceeds the acceptance tolerance, set on the parameters page, will be coloured red.

SAMPLE NAME	ALL ATTRIBUTES	# IDENTIFIED PRECURSORS	# QUANTIFIED PRECURSORS	MISSING VALUES (%)	PRECURSOR CORRELATION	# IDENTIFIED PEPTIDES	# QUANTIFIED PEPTIDES	# IDENTIFIED PROTEIN GROUP	# QUANTIFIED PROTEIN GROUPS
Sample 1	Pass	68241	0	0.00	0.00	60224	0	7331	0
Sample 2	Fail	55541	0	0.00	0.00	49329	0	6975	0
Sample 3	Fail	68385	0	0.00	0.00	60109	0	7305	0
Sample 4	Fail	56109	0	0.00	0.00	49400	0	7160	0
Sample 5	Fail	44625	0	0.00	0.00	40280	0	6313	0
Sample 6	Fail	65226	0	0.00	0.00	57082	0	7304	0
Average	Fail	59687	0	0.00	0.00	52737	0	7064	0

The columns in this table are:

- **Sample Name:** Name of the sample represented by this row. The table will always be sorted by this column (in sample order) with the control sample at the top and the average values at the bottom. If average is selected as the control sample it will be displayed at the top of the table instead of the bottom. The selected control sample will have its name displayed in bold in this column.

- **All Attributes:** This column displays whether samples either pass or fail quality control. This considers all attributes selected on the summary page, not only those in the quantification attribute list.

- **#Identified Features/Precursors:** The number of identified features/precursors in the sample for DDA and DIA respectively.
- **#Quantified Features/Precursors:** The number of quantified features/precursors in the sample for DDA and DIA respectively.
- **Missing Values (%):** The percent of peptides where no feature could be found in a sample based on identification in other samples.
- **Precursor Correlation (%):** Measures the pearson correlation coefficient of the MS1 features compared to the control sample.
- **#Identified Peptides:** Total number of peptides identified in the sample.
- **#Quantified Peptides:** Total number of peptides quantified in the sample.
- **#Identified Protein Groups:** The number of identified protein groups in the sample.
- **#Quantified Protein Groups:** Total number of proteins quantified in the sample.

3.4.1. Quantification QC Figures

Figure 1 will display a visualized summary of the selected column in the Quantification QC summary table, which will be highlighted in blue. By default the # Identified Features and #Quantified Features columns will be selected, to select a new column(s) click on the header row of the column you wish to select.

The dotted lines displayed on the chart represent the minimum and maximum values of the acceptance range (determined by the acceptance tolerance set on the parameters page). Any bar with a height that falls within these two dotted lines represents a value within the acceptance tolerance. Any bar with a height below or above these lines represents a value outside of the set acceptance tolerance.Place mouse over the graph and zoom in and out to show all or less samples/sample names on the x-axes.

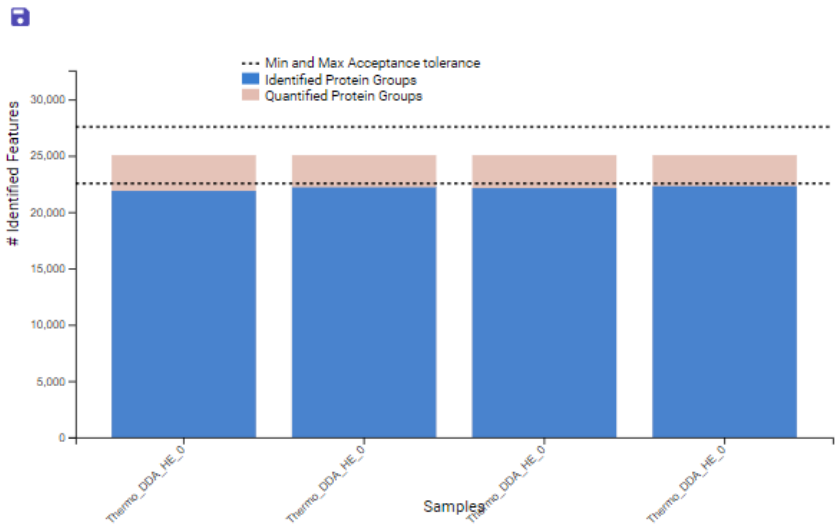
Note

For analyses with greater than 30 samples, a line chart will be displayed instead of a bar chart.

Note

For DIA spectral library searches without protein inference, #Identified peptides and #Quantified peptides will be the default selected column.

Figure 1. Identified and Quantified Precursors.



Figures 2, 3, and 4. show relative area correlation at the feature, peptide and protein level. Correlation values are displayed in log2 ratio. By default, only 2 samples are shown for pairwise comparison. Using the drop down menu above the figure, users can select different samples for comparison.

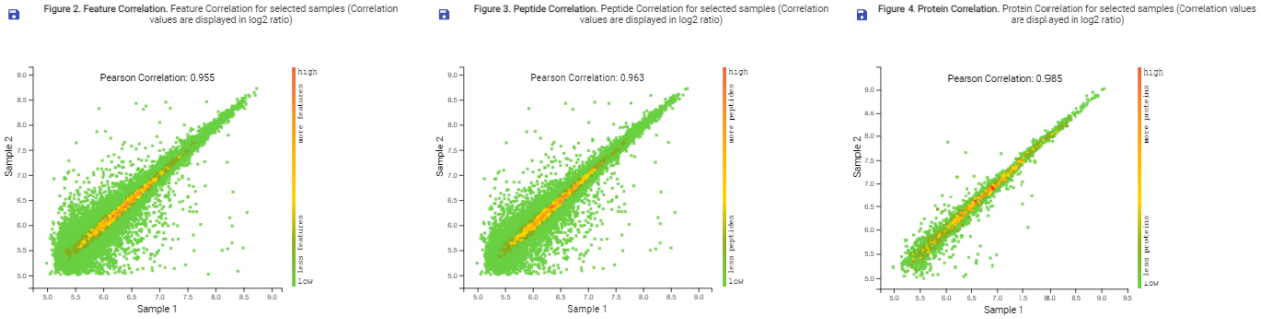
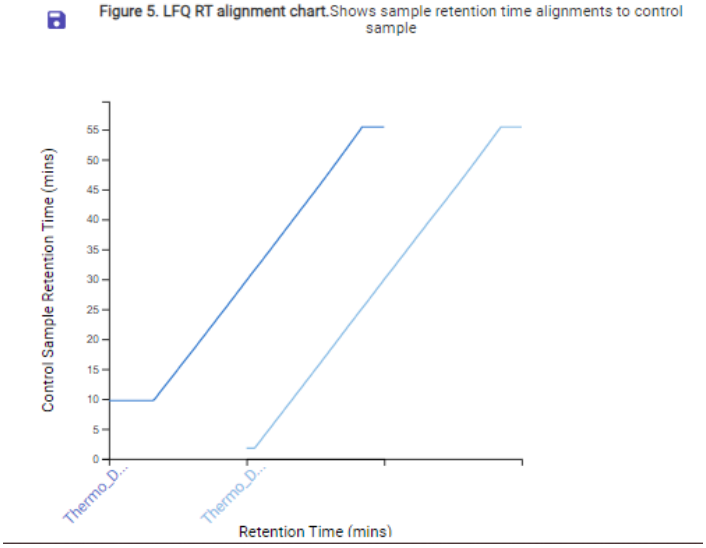


Figure 5. displays the LFQ RT alignment chart. Shows sample retention time alignments to the control sample. By default only 2 samples are shown. Using the drop down menu, users can select different samples for comparison to control. If one of the samples selected is the control sample, this chart will only display one line.

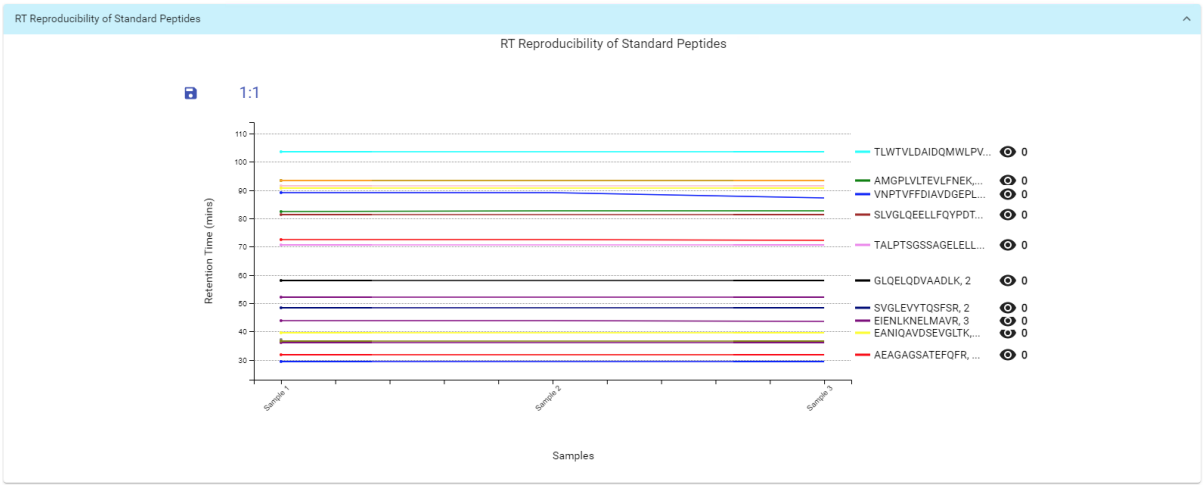


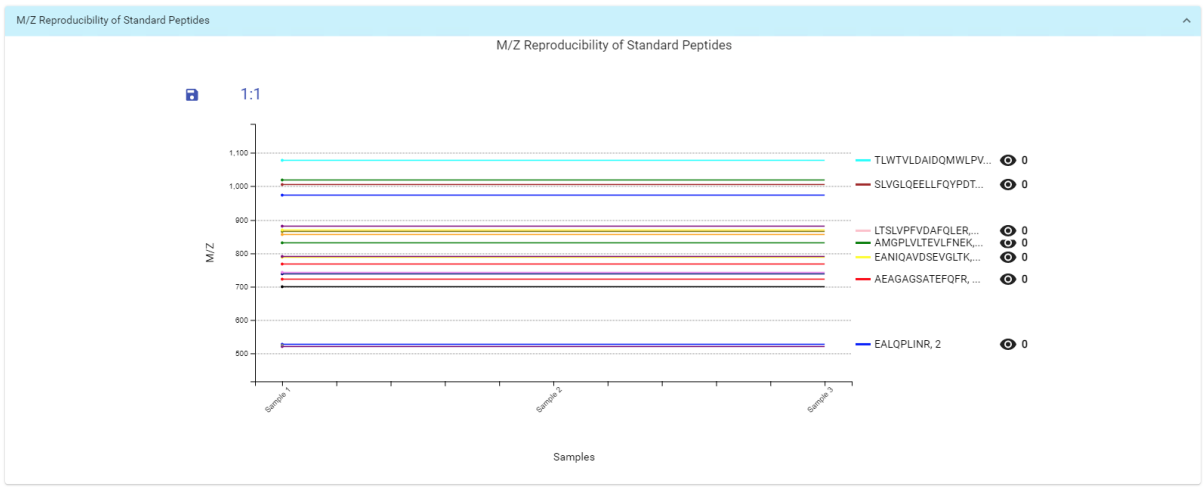
3.5. QC Standard Peptides Tab

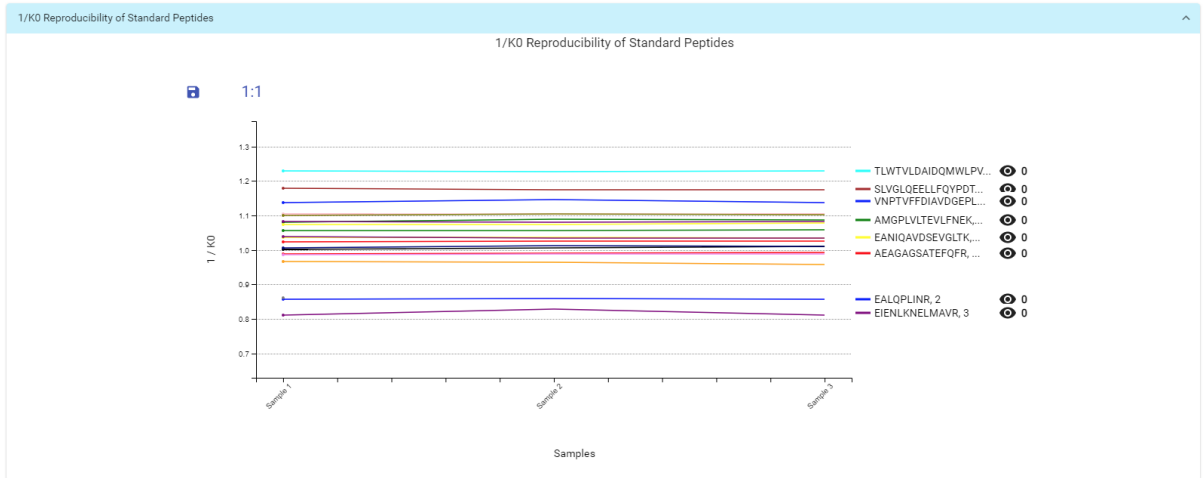
QC Standard Peptides tab displays RT and M/Z reproducibility graphs. Additionally, 1/k0 reproducibility of standard peptides will be displayed for timsTOF data.

Note

The eye icon is used to highlight samples with missing values. 1







3.6. QC Results Exporting

3.6.1. Result Page Exporting

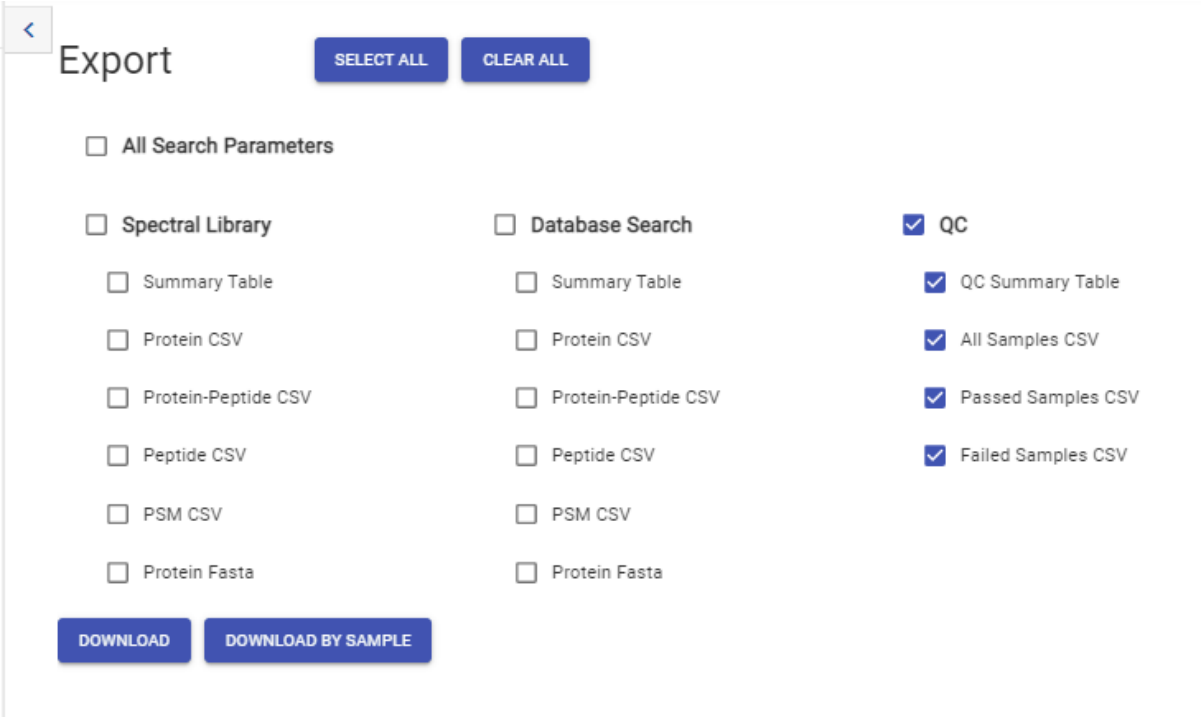
There are two export options available for quick exporting at the top right of the summary page. The exports are:

- **Statistics:** This export link will download a CSV containing summary of all information displayed in Table 1.
- **Failed Samples:** This export link will download a CSV containing summary of Failed Samples statistics as per selection made in "Filter Samples".

- **Passed Samples:** This export link will download a CSV containing summary of Passed Samples statistics as per selection made in "Filter Samples" and as displayed in Table 2.

3.6.2. Analysis Export Page

A QC checkbox in the Analysis Export page offers four options for result download. Selecting this option will add the All Samples QC export (as detailed above) to the resulting export zipped file. This option will also add the All Samples QC CSV for both the Download and Download by Sample export options.



Chapter 10. Isobaric Labelling Quantification (TMT/iTRAQ)

1. Overview

Isobaric labelling quantification with chemical labels at the MS2 (or MS3) level is one of the three quantification modes that are supported by the optional PEAKS Q module of PEAKS Online. In contrast to LFQ Quantification, TMT/iTRAQ Quantification is based on the relative intensities of the fragment peaks at fixed m/z values within an MS2 (or MS3) spectrum. To prepare for this, isobaric chemical labels are introduced into samples during the sample preparation stage. The samples are then combined and 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.

The usage of this workflow is outlined below.

1. Refer to Chapter 5, Peptide De Novo Sequencing, Section 2, “*DeepNovo/De Novo* Analysis Workflow and Parameter Settings” , Step 1-6 to create a new project and add data.

2. Select the DDA "Quantification" workflow.

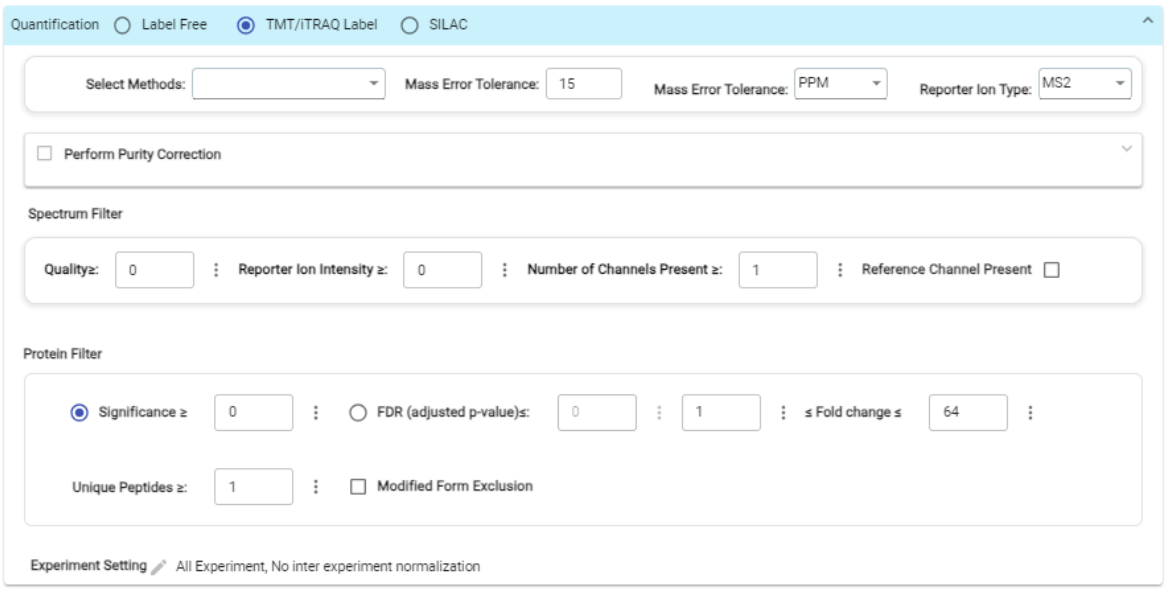
3. Set DB Search Parameters. Refer to Chapter 6, Peptide, PTM, and Mutation Identification (PEAKS DB, PEAKS PTM, and SPIDER), Section 2, “Database Search Workflow and Parameter Settings” for details.

   ### Note

   TMT/iTRAQ Modifications will be automatically added to DB search parameters when TMT/iTRAQ is selected.

4. Set the Report Filters. Refer to Chapter 6, Peptide, PTM, and Mutation Identification (PEAKS DB, PEAKS PTM, and SPIDER), Section 2.8, “PSM Filter” for details.

5. Set the Quantification parameters by first select "TMT/iTRAQ". This will automatically open the appropriate parameter settings. Set the TMT/iTRAQ specific Quantification settings: Label Q Method, Mass Error and Report Ion Type. See Section 2 below for further details.



6. Within the Quantification section: Set the Spectrum and Protein Filters. See Chapter 5 and below for further details.

7. Within the Quantification section: Edit the Experiment Settings. See below for further details.

8. Review project settings and parameters then click Submit button. Name the analysis and click OK to start the analysis.

2. Setting Parameters

For more information on DB search parameters, refer to Chapter 6, Peptide, PTM, and Mutation Identification (PEAKS DB, PEAKS PTM, and SPIDER), Section 2, “Database Search Workflow and Parameter Settings” .

The following parameters are available when the TMT/iTRAQ Quantification workflow is selected.

- **Select Methods:** From the Select Methods drop-down menu, choose the appropriate quantification method used in the experiment, e.g. iTRAQ-4plex

- **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.

- **Report Ion Type:** This parameter allows for the selection of MS2 or MS3 as the reporter ion type for quantification.

- **Perform Purity Correction:** Select the checkbox to edit correction factors for chemical labels (if applicable). Once the checkbox is selected, click the down arrow to open the dialogue that allows 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.

2.1. Spectrum Filter:

- **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.

- **Reporter Ion Intensity:** Peptides with at least one reporter ion above this value will be kept.

- **Number of Channels Present:** Only quantifiable peptides that contain at least the number of labels selected will be used in protein quantification.

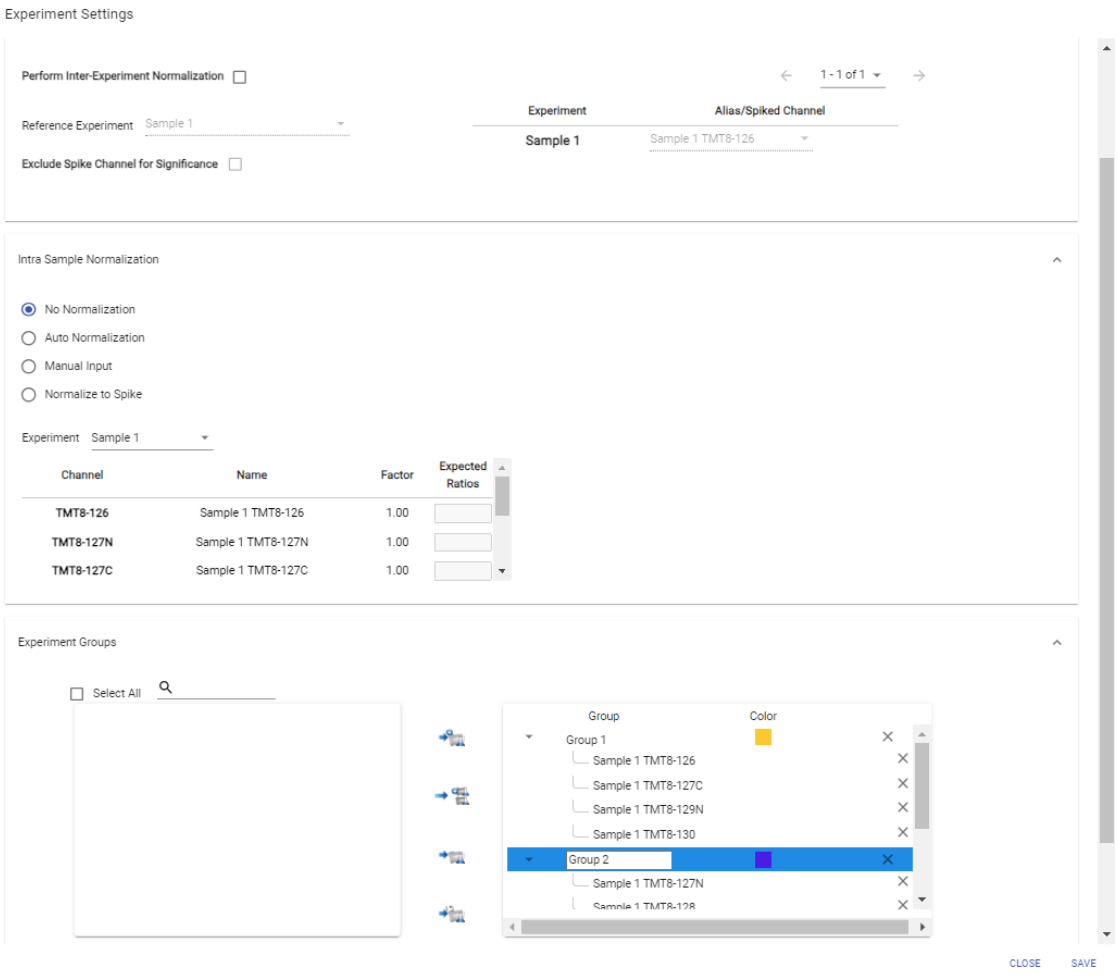
- **Reference Channel Present:** Only quantifiable peptides in which the reference label is present will be used in protein quantification. Note: The reference label is set in the experiment setting pop-up. See Experiment Setting section below

2.2. Protein Filter:

- Refer to Chapter 8, Label-Free Quantification (LFQ), Section 2, “Setting Parameters” , for detailed descriptions of quantification protein filters.

2.3. Experiment Setting:

The following parameter dialogue appears when the button is clicked.



- **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.
- **Setting Sample Channel Alias:** The name of the Sample Channel can be changed by using the alias name.
- **Inter-Experiment Normalization:** Check this box if a spike label or normalization channel was used in the experiment. This is used to enable quantification between experiments.
  - **Alias/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.
- **Intra Sample Normalization:**
  - No Normalization: If there are multiple samples in the quantification run and all samples are selected, No Normalization is chosen as the default option. It will be the only normalization option available for selection unless "Perform Inter Experiment Normalization" is check-marked in Experiment Settings.
  - Auto normalization: This option will calculate a global ratio from the total intensity of all labels in all quantifiable peptides. Factor values and Expected Ratios between the channels (i.e., the labels) will be presented. Selecting another sample from the Experiment drop-down menu will display the results of other samples.

  - 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.

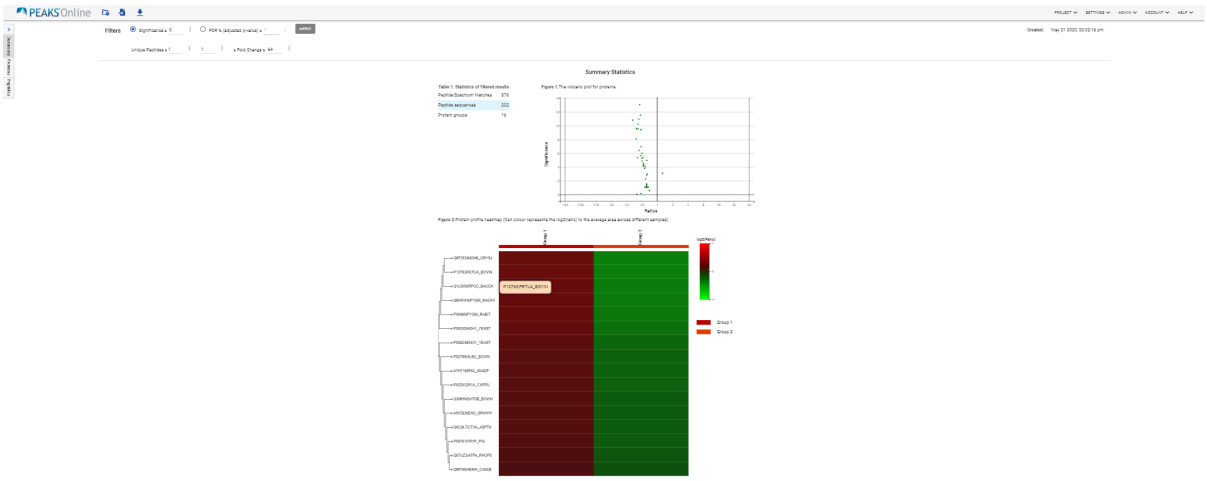
  - Normalize to spike: This option displays a list of identified proteins that can be selected to create a normalization ratio. Use the search bar to quickly locate specific proteins. Right-clicking in the Protein list will open a pop-up menu to select all highlighted rows or remove all selected proteins 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.

- **Experiment Groups** : TMT/iTRAQ Quantification requires sample grouping. All available samples are listed in the left window. 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 that appears when you choose the add to existing option . If you select multiple files, you can add each one to a new group by clicking the button. Samples groups can be removed by selecting the "X" next to the group name. Individual samples within a group can be removed by using the down arrow next to the group name to show all of the samples in a group, and then selecting the "X" next to the sample you want to remove from the group.

- **Reference Label:** Use this drop-down menu to select a reference label. The reporter ion intensity corresponding to this reference label will be the denominator in all ratio calculations. Only a label that is already listed in the Experiment Groups screen can be selected as a reference.

3. Understanding the Results

After a PEAKS Online TMT/iTRAQ quantification is complete, the isobaric labelling results will be displayed in a Summary, Protein, and Peptide page. The summary page will be the first page in view, and the others are accessed by clicking on the respective tabs.



- **Summary:** The outline of the isobaric labelling search results with statistics. This is the place to examine the overall performance of the experiment and adjust filters. If there are too many proteins, the heatmap will not show on the summary page.

- **Proteins:** The quantified proteins with a list of supporting peptide features for each protein.

- **Peptides:** The quantified peptides are listed in a table. Annotated spectrum chart, ion match table, and reporter ion spectrum are also displayed for each peptide.

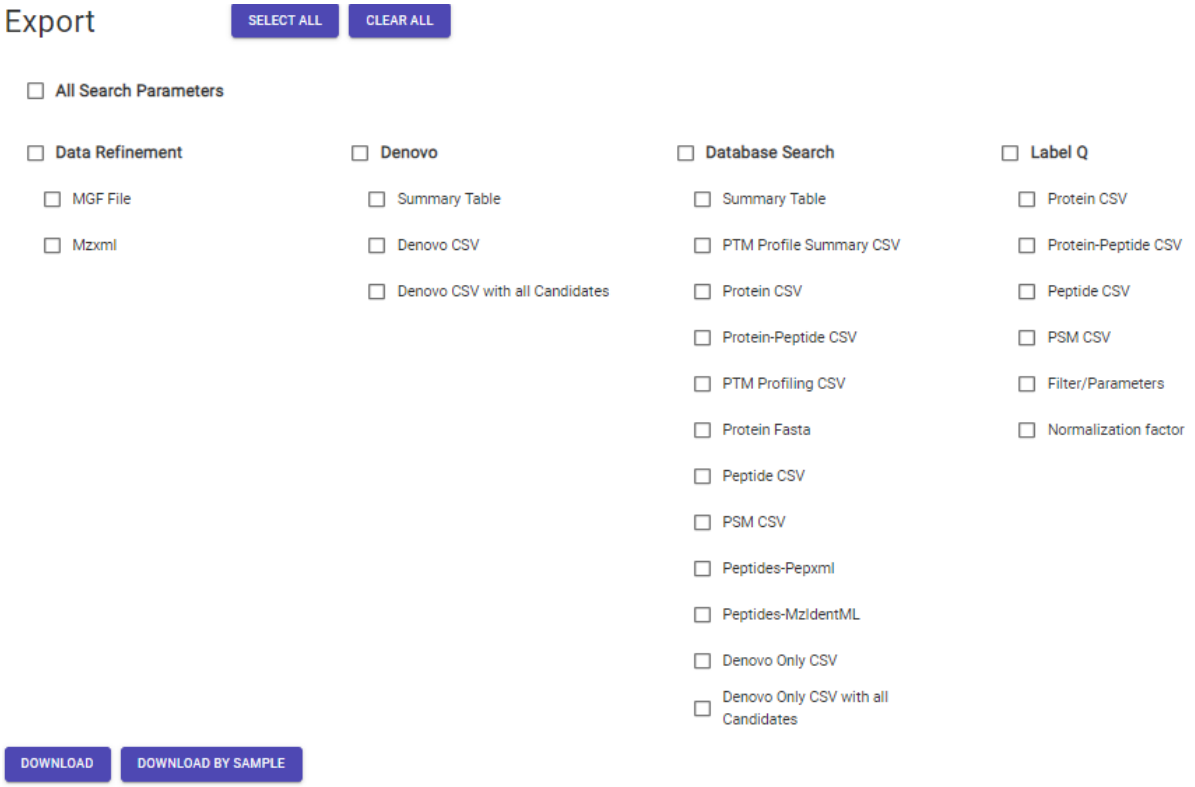
3.1. Result Filtration

On the top of each page, there are different options for users to filter isobaric labelling results that can be accessed by selecting the edit icon (pencil): First choose Significance or FDR % (adjusted p-value),then set your values. Next, enter a Fold Change and the number of Unique Peptides. After changing any of these filter parameters, click APPLY next to the filtration window to apply the new filters and update the results and statistics.

- **Significance:** Only protein groups with a significance above this threshold will be listed in the “Protein” view. The significance score is the -10lg of the significance testing p-value. 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. Significant is calculated using ANOVA as described in Chapter 8.
- **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.

3.2. Result Exporting

PEAKS TMT/iTRA Q results can be exported to various text formats in the Export option listed under the analysis (all export options), or from the top right of the Protein or Peptide pages (two export options).



- **Proteins CSV:** These are the quantified proteins that pass the filters set in the "Summary" page and their associated details.
- **Protein-Peptide CSV:** These are the supporting peptides of the quantified proteins and their associated details.
- **Peptide CSV:** A complete list of all quantified peptides that pass the peptide feature filters set in analysis parameters and their associated details.

- **PSM CSV:** All predicted peptide features and their associated quantification information. Both identified and unidentified peptide features are included. To include unidentified peptide features, set Peptide Id Count equals to 0 in the Peptide feature filter during selection of parameter settings.
- **Filter/Parameters:** Filter/Parameters for the current analysis
- **Normalization Factor:** Normalization Factors used in the current analysis

3.3. Summary View

The isobaric labelling view includes an expression profile with candidate proteins clustered in a heat map, result statistics table, and a protein volcano plot. An "Analysis Description" can be added and displayed below the filter section.

3.3.1. Statistics of Filtered Results

Table 1 in the Summary page lists the filtered results of feature vectors and proteins.

3.3.2. The Volcano Plot for Proteins

A volcano plot (Figure 1) 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 1 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). 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.

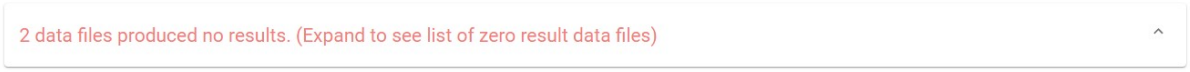
Circles for the proteins that are above the set significance and fold change thresholds will be displayed in colour. Scrolling the cursor over a circle will show the protein info.

3.3.3. Protein Profile Heatmap

The heatmap in Figure 2 displays the protein groups that pass the filters. The relative protein abundance data is represented as a heat map with a row for each protein group. 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.

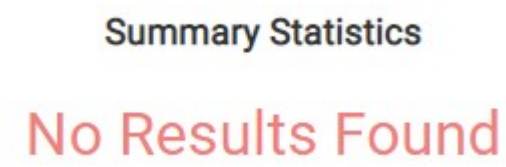
3.3.4. Low Quality Data Warnings

If PEAKS Online detects any data files producing very few or no results a warning section will appear at the top of the Summary Page indicating the number of data files producing poor or no results.



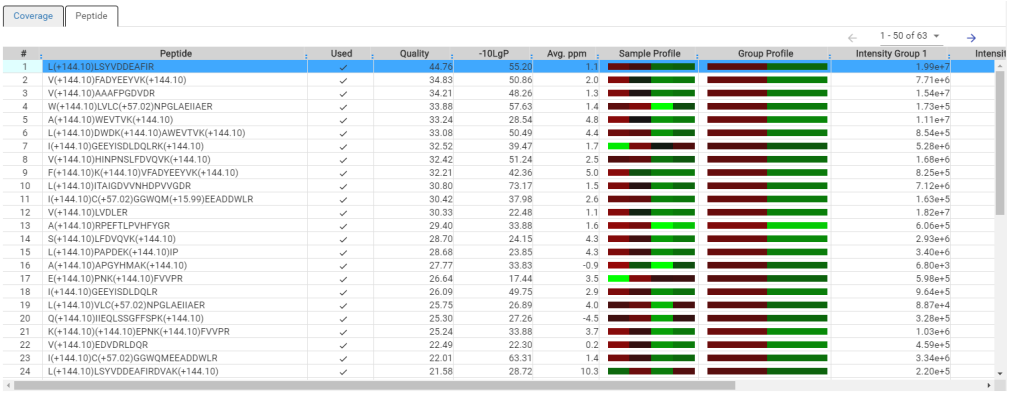
This section can be expanded to display a list of data files in the analysis that are producing poor or no results.

If there were no results found in any of your data files, a warning message stating "No Results Found" will be displayed instead of the usual Summary Page figures and tables.



3.4. Protein View

The isobaric labelling Protein page lists all the quantified proteins that pass the filter set on the Summary page, characterizes each protein at the amino acid level, and lists the supporting peptide features of each protein.



#	Peptide	Used	Quality	-10LgP	Avg. ppm	Sample Profile	Group Profile	Intensity Group 1	Intensit
1	L(+144.10)LSYVDDEAFIR	✓	44.76	55.20	1.1			1.99e+7	
2	V(+144.10)FADYEEYVK(+144.10)	✓	34.83	50.86	2.0			7.71e+6	
3	V(+144.10)AAAFPGDVDR	✓	34.21	48.26	1.3			1.54e+7	
4	W(+144.10)LVLC(+57.02)NPGLAEIIAER	✓	33.88	57.63	1.4			1.73e+5	
5	A(+144.10)WEVTVK(+144.10)	✓	33.24	28.54	4.8			1.11e+7	
6	L(+144.10)DWDK(+144.10)AWEVTVK(+144.10)	✓	33.08	50.49	4.4			8.54e+5	
7	I(+144.10)GEEYISDLDQLRK(+144.10)	✓	32.52	39.47	1.7			5.28e+6	
8	V(+144.10)HINPNSLFDVQVK(+144.10)	✓	32.42	51.24	2.5			1.68e+6	
9	F(+144.10)K(+144.10)VFADYEEYVK(+144.10)	✓	32.21	42.36	5.0			8.25e+5	
10	L(+144.10)ITAIGDVVNHDPVVGDR	✓	30.80	73.17	1.5			7.12e+6	
11	I(+144.10)C(+57.02)GGWQM(+15.99)EEADDWLR	✓	30.42	37.98	2.6			1.63e+5	
12	V(+144.10)LVDLER	✓	30.33	22.48	1.1			1.82e+7	
13	A(+144.10)RPEFTLPVHFYGR	✓	29.40	33.88	1.6			6.06e+5	
14	S(+144.10)LFDVQVK(+144.10)	✓	28.70	24.15	4.3			2.93e+6	
15	L(+144.10)PAPDEK(+144.10)IP	✓	28.68	23.85	4.3			3.40e+6	
16	A(+144.10)APGYHMAK(+144.10)	✓	27.77	33.83	-0.9			6.80e+3	
17	E(+144.10)PNK(+144.10)FVVPR	✓	26.64	17.44	3.5			5.98e+5	
18	I(+144.10)GEEYISDLDQLR	✓	26.09	49.75	2.9			9.64e+5	
19	L(+144.10)VLC(+57.02)NPGLAEIIAER	✓	25.75	26.89	4.0			8.87e+4	
20	Q(+144.10)IIEQLSSGFFSPK(+144.10)	✓	25.30	27.26	-4.5			3.28e+5	
21	K(+144.10)(+144.10)EPNK(+144.10)FVVPR	✓	25.24	33.88	3.7			1.03e+6	
22	V(+144.10)EDVDRLDQR	✓	22.49	22.30	0.2			4.59e+5	
23	I(+144.10)C(+57.02)GGWQMEEADDWLR	✓	22.01	63.31	1.4			3.34e+6	
24	L(+144.10)LSYVDDEAFIRDVAK(+144.10)	✓	21.58	28.72	10.3			2.20e+5	

3.4.1. Protein Table

The quantified proteins are listed in the protein table with homologous proteins grouped together. If proteins are grouped, the individual proteins within a group can be displayed by clicking on the plus sign next in the far left of the row.

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:

- **Significance:** The protein significance is calculated as explained above.
- **Sample Profile:** The protein abundance among the samples is depicted as a heat map. Hold the cursor on a profile to view the sample channels, abundances, and ratios with respect to the base sample. Channel intensities are the sum of all reporter ion intensities from supporting peptides that pass the filters.
- **Group Profile:** The protein abundance among the groups is 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, abundances, and ratios with respect to the group which contains the base sample. Intensity values are the average of the sample profile intensities based on the selected grouping.

Refer to Chapter 6, Peptide, PTM, and Mutation Identification (PEAKS DB, PEAKS PTM, and SPIDER) Section 3.4.2, “Protein Table” for more details of other entries of the protein table.

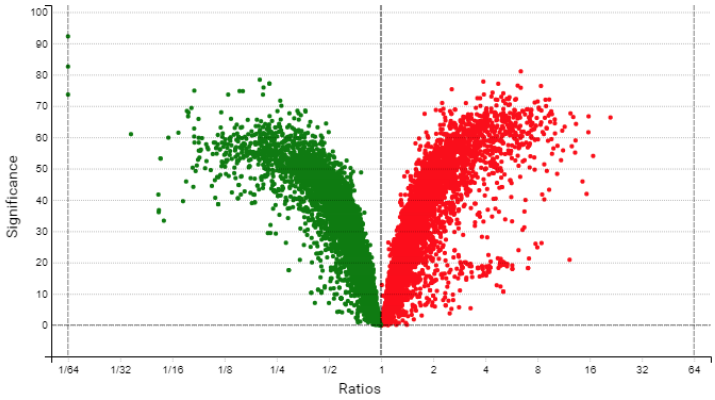
3.4.2. Protein Volcano Plot

The volcano plot facilitates identification of the significant proteins. This interactive volcano plot enables the selection of a protein in the table. 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 less intense relative to the reference group while red points are more intense relative to the reference group. Grayed points are filtered out as set in the filter. The volcano plot facilitates identification of significant proteins.

Protein markers that are in color can be clicked and selected. Upon clicking a point on the volcano plot, the corresponding protein in the Protein Table will be highlighted and more details are displayed below the Protein Table.

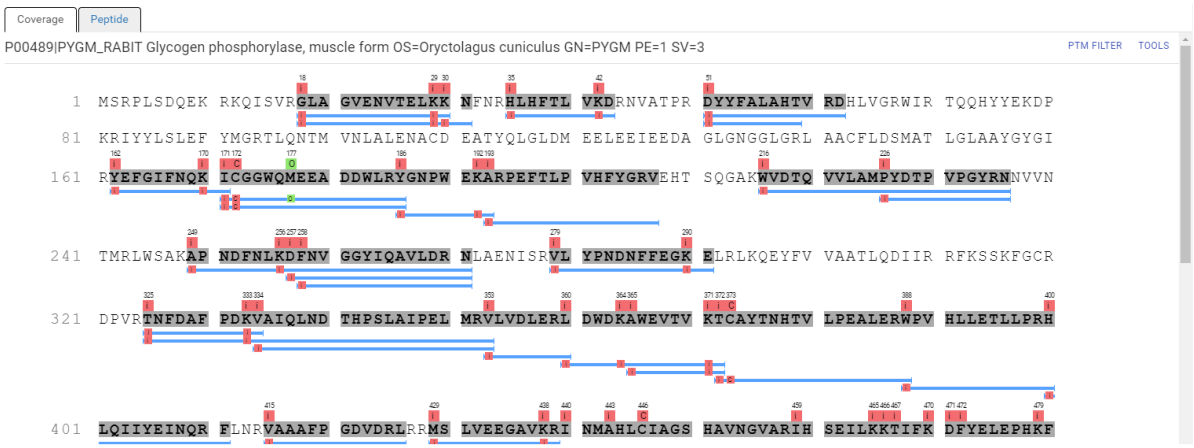
- **Zoom in/out from Volcano Plot:** Place the cursor on the volcano plot and drag a rectangle over the area of interest to Zoom in. By double clicking on the volcano plot you can zoom out and view the volcano plot with the default dimensions.

Figure 1.The volcano plot for proteins.



3.4.3. Protein Coverage

The coverage view characterizes the selected protein sequences at the amino acid level. Refer to Chapter 6, Peptide, PTM, and Mutation Identification (PEAKS DB, PEAKS PTM, and SPIDER), Section 3.4.4, “Protein Coverage” for details. Click the supporting peptide sequence to open an XIC pop-up window.



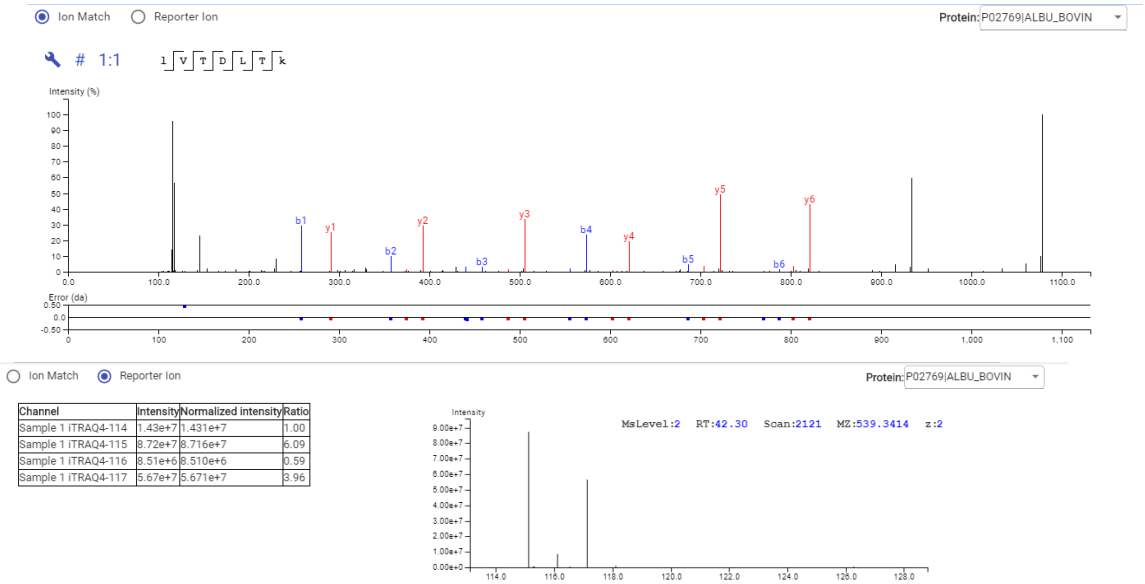
3.4.4. Support Peptides

The supporting peptides assigned to the protein are shown in the tab beside the Coverage tab.

This table contains the following information of the quantifiable support peptides for the selected protein in the protein table.

- **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. The top three peptides with highest abundance are used to calculate the protein abundance.
- **Quality:** This defines how well the peptide can be quantified. It is calculated by taking the average of the quality scores of the feature vectors.
- **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.
- **Sample Profile:** The peptide abundance among the samples is 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**.** The channel intensities are the sum of intensities from all PSMs that identified the peptide.

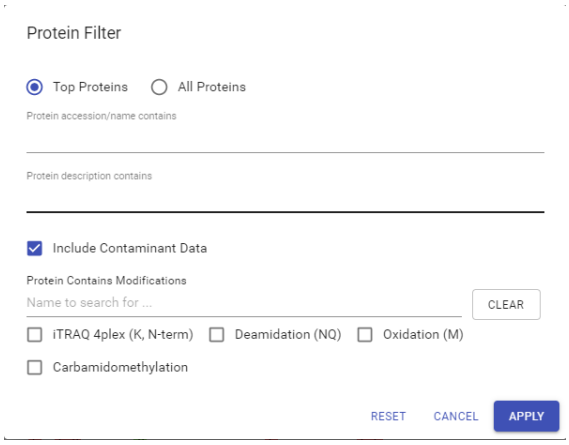
- **Group Profile:** The peptide abundance among the groups is depicted as a heat map. Hold the cursor on a profile to view the groups, the peptide areas in groups, and the corresponding ratios with respect to the group which contains the base sample **.** The given intensities are the average intensities of sample intensities included in each group.
- **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 the given groups and the reference group.
- **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.
- **Spectrum View:** The spectrum view for any peptide in the table can be displayed by clicking on the supporting peptide of interest, which will open a window displaying its Ion Match or Reporter Ion



Channel	Intensity	Normalized intensity	Ratio
Sample 1 iTRAQ4-114	1.43e+7	1.431e+7	1.00
Sample 1 iTRAQ4-115	8.72e+7	8.716e+7	6.09
Sample 1 iTRAQ4-116	8.51e+6	8.510e+6	0.59
Sample 1 iTRAQ4-117	5.67e+7	5.671e+7	3.96

3.4.5. Protein View Filter

Click the button on top-left of the protein table to open the protein filter and control which proteins to display.

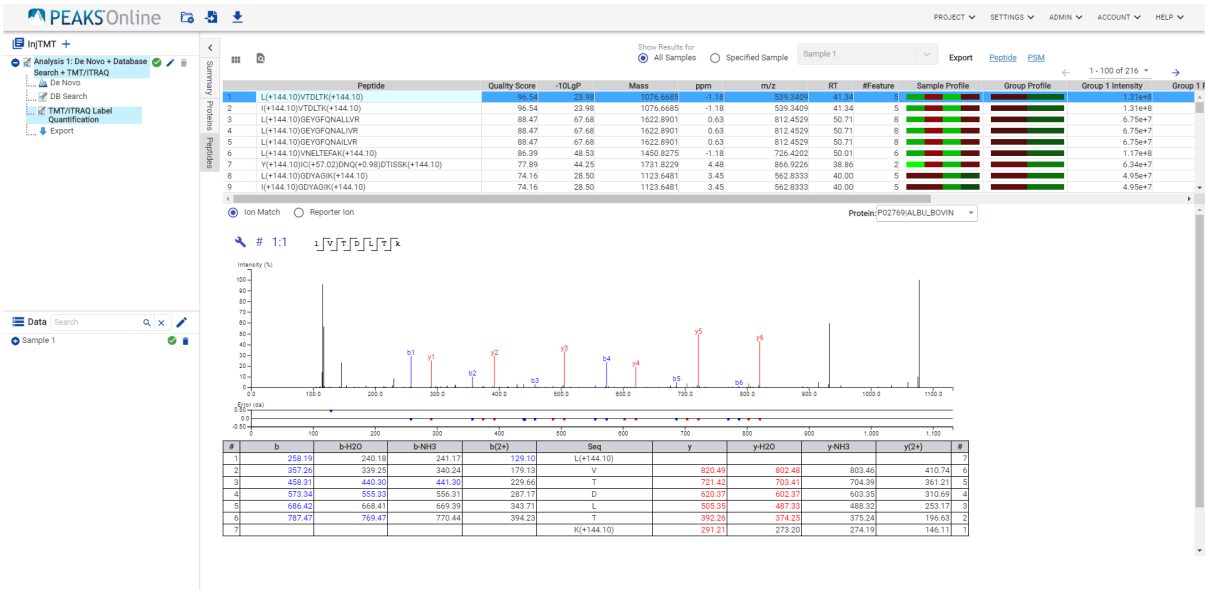


- **Top Proteins** : The number of proteins that have the highest number of unique peptides supporting them within their protein group.

- **All Proteins** : The number of proteins given the parameters set for PEAKS Protein Score (-10lgP) and number of unique peptides.
- **Protein Sequence contains** : Enter a protein sequence, only protein that contain that sequence will be displayed in the protein table
- **Protein description contains:** Enter a protein description, only protein that contain that description will be displayed in the protein table
- **Include Contaminant Data:** Check this to display contaminants that were identified during a PEAKS DB search.
- **Protein Contains Modifications** : Check boxes are provided for each identified modification.

3.5. Peptide View

The quantifiable peptide spectrum matches with the same sequence are grouped together to get the list of quantifiable peptides and are displayed in the **Peptide View** along with their detailed information.



3.5.1. 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 3.4.4, “Support Peptides”. The Peptide table entries may be filtered by using the **Peptide View Filters** on top of the peptide table, for more information refer to Section 3.5.1, “Peptide View Filters” in Chapter 6, Peptide, PTM, and Mutation Identification (PEAKS DB, PEAKS PTM, and SPIDER).

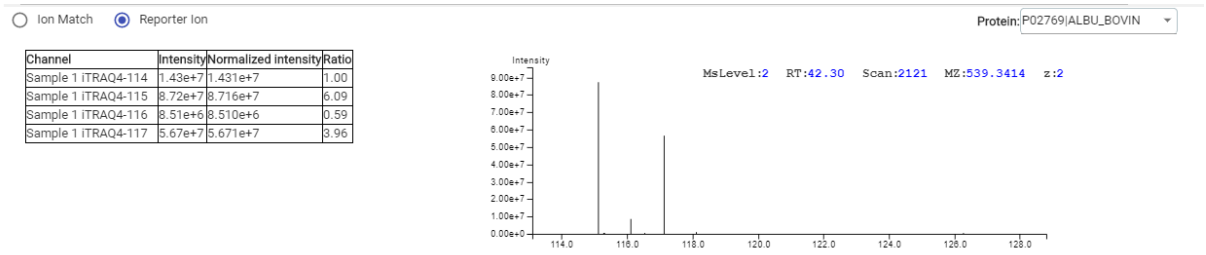
3.5.2. Ion Match

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. If the fragment ion is present in the spectrum, its mass 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”).

3.5.3. Reporter Ion

This view provides details on characteristics of the reporter ions used in the experiment. The table provides information on the normalized 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.



Channel	Intensity	Normalized intensity	Ratio
Sample 1 iTRAQ4-114	1.43e+7	1.431e+7	1.00
Sample 1 iTRAQ4-115	8.72e+7	8.716e+7	6.09
Sample 1 iTRAQ4-116	8.51e+6	8.510e+6	0.59
Sample 1 iTRAQ4-117	5.67e+7	5.671e+7	3.96

4. Ion Mobility Data

Ion mobility data, including Field asymmetric waveform ion mobility (**FAIMS**), and trapped ion mobility (**tims**) spectrometry analyzes complex mixtures and separates the ions based on their mobility. PEAKS Online has the ability to load IMS data and analyze it using *de novo* sequencing, database search (PEAKS DB), database search with unexpected modification identification (PEAKS PTM), sequence variant identification (SPIDER) search, spectral library search, and labelled and label free quantification.

The analysis result for IMS dataset contains the additional information for CV values compared to the non-IMS dataset. The CV-related information can be found at the following places for a TMT analysis.

- The **Venn diagram** on the summary page.

  Venn diagram overlaps between different CVs in the search for all proteins, top proteins, and peptides are shown on the summary page.

- The additional **CV** column on the supporting peptide page
- The additional **CV** column in the exported *Protein-Peptide CSV*
- The additional **CV** column on the peptide page
- The additional **CV** filter on peptide page

  The filter for peptide page contains one additional option for CV, which selects the displayed scans based on the selected CV values

- The additional **CV** column in the exported *Peptide CSV*

Chapter 11. Stable Isotope Labeling by Amino acids in Cell culture (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 Online. Typically, an unlabelled sample is mixed with a sample subjected to stable isotope labelling. 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 equal to the expected difference caused by the stable isotope label. 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.

The usage of this workflow is outlined below.

1. Refer to Chapter 5, Peptide De Novo Sequencing Section 2, “*DeepNovo/De Novo* Analysis Workflow and Parameter Settings”, step 1-6 to create a new project and add data.

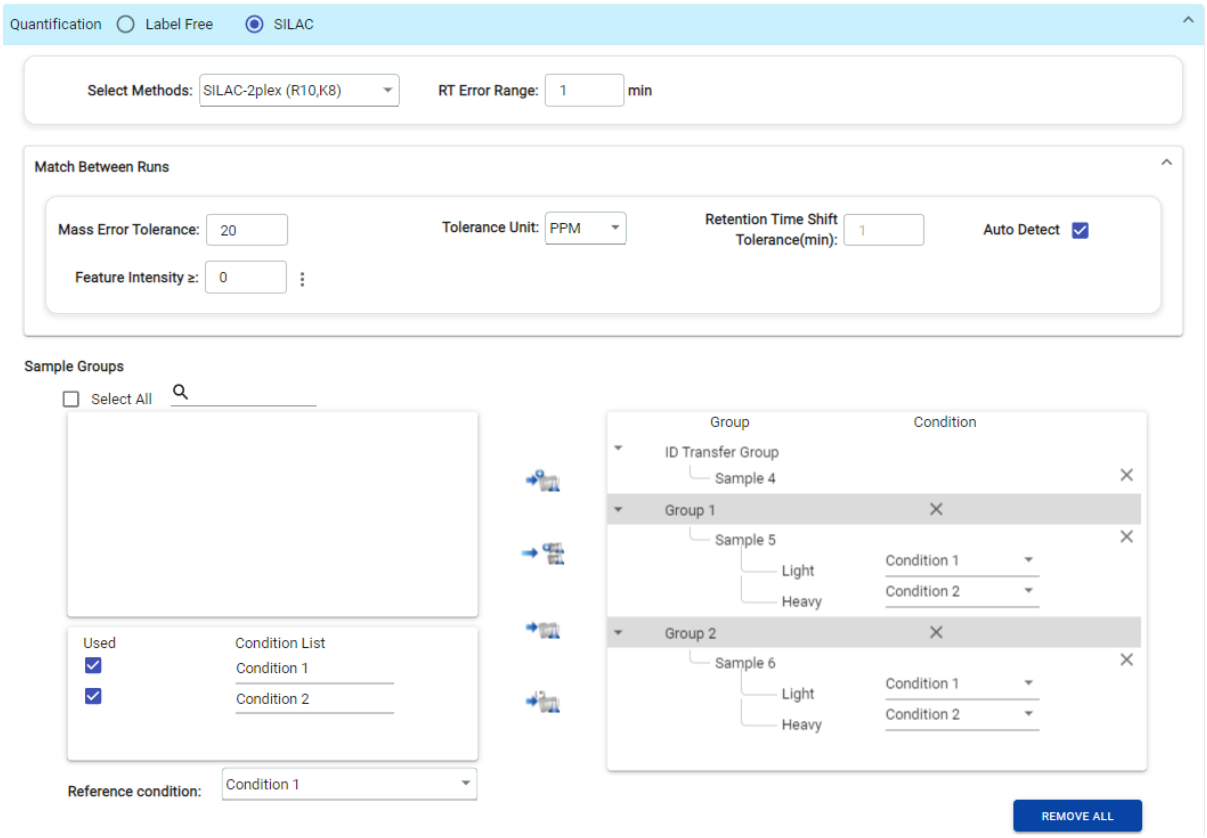
2. SILAC quantification can be run on PEAKS Database Search results. Select either “PEAKS Q” workflow in the DDA section of the Workflow step.

    - To run SILAC quantification on PEAKS Database Search results, select the “PEAKS Q” workflow in the DDA section.

3. Set the Identification Search Parameters.

    - If running SILAC quantification on a DDA DB Search analysis, refer to Chapter 6, Peptide, PTM, and Mutation Identification (PEAKS DB, PEAKS PTM, and SPIDER) Section 2, “Database Search Workflow and Parameter Settings”

4. For a SILAC workflow, the SILAC-specific parameters will be shown at the bottom of the Project Submission page. Set the SILAC sample grouping and Retention Time Error Tolerance. See Section 2 below for more details. SILAC quantification will be performed on the combined identification result.



5. Set PSM Filters. Refer to Chapter 6, Peptide, PTM, and Mutation Identification (PEAKS DB, PEAKS PTM, and SPIDER) Section 2.8, “PSM Filter”, for details.

6. Set the match between runs (MBR) parameters. MBR allows for identified MS/MS scans from one sample to be applied to LC-MS peptide features from other samples within the mass error and RT error tolerances.

   - Mass Error Tolerance: Controls the allowable difference in mass between the identified MS/MS scan in one sample versus LC-MS peptide features in other samples. Either Da or ppm can be selected as the mass unit.

   - Retention Time Shift Tolerance: The maximum allowable difference between the retention time of the identified MS/MS scan and the LC-MS peptide feature considered for MBR.

   - Feature intensity: Set a value for the minimum intensity a peptide feature must exhibit to be considered for MBR.

7. Review project settings and parameters, then click the Submit button at the top of the page.

8. Name the analysis, then click OK to start the analysis.

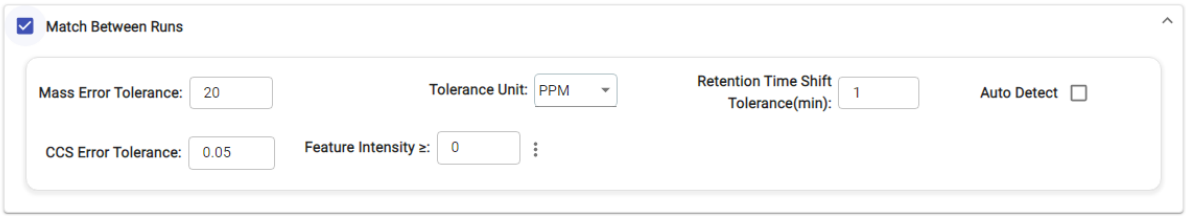
2. Setting Parameters

For more information on DB search parameters, refer to Chapter 5, Peptide, PTM, and Mutation Identification (PEAKS DB, PEAKS PTM, and SPIDER), Section 2, “Database Search Workflow and Parameter Settings”.

The following parameters are available when the SILAC Quantification workflow is selected.

- **Select Method:** From the Select Methods drop-down menu, choose the appropriate quantification method used in the experiment.

- **Match Between Runs (MBR):**



- **Mass Error Tolerance:** Controls the allowable difference in mass between the identified MS/MS scan in one sample versus features in other samples. This can be set similar to the precursor mass error tolerance used for database search.
- **Retention Time Shift:** The maximum allowable difference between the retention time of the identified MS/MS scan and the peptide feature considered for MBR.
  - **Auto Detect:** Let PEAKS Online detect the optimal Retention Time Shift Tolerance for your data. An initial quantification run will be performed to determine the retention time shift trend. The optimal error tolerance will then be used for the full search.
- **CCS Error Tolerance:** The maximum collisional cross section (CCS) differential between an identified MS/MS scan and peptide features in other samples to allow MBR.

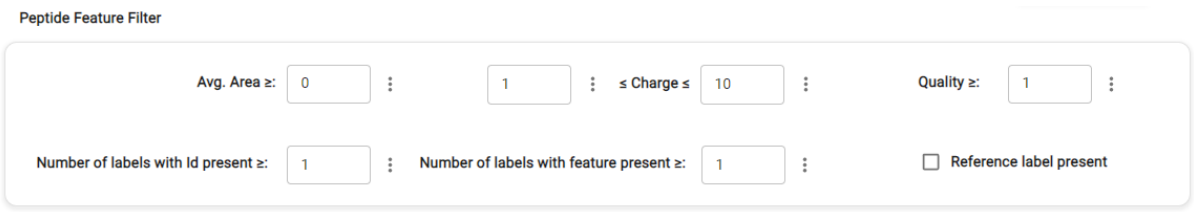
  ## Note

  This parameter is only available when the analysis contains TimsTOF data.

- **Feature Intensity:** Set a value for the minimum intensity a peptide feature must exhibit to be considered for MBR. By default, the integrated area under the curve of the peptide feature will be used to represent feature intensity.

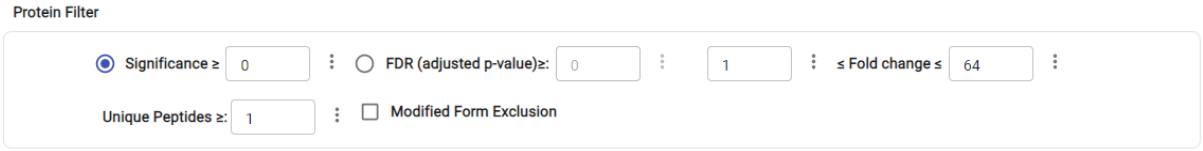
- **RT Error Range:** The retention time shift tolerance is the maximum elution time difference that is considered for the quantification of an identified peptide after RT alignment.
- **Samples:** The MS samples are listed for selection. After selecting a sample, it can be added to a new group by clicking 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.

  ID Transfer Group: If the "Fraction Assisted ID transfer" method was selected use the button to add the ID transfer sample to the ID Transfer group
- **Groups**: The samples can be arranged into one or more groups. A group represents a state of the experiment where quantitative ratios between groups 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 ‘X’ next to the sample or the Remove All button, respectively.
- **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 represents one unique condition. To remove a label from analysis, uncheck the corresponding condition. Condition names can be renamed.
- **Reference Condition**: The condition which is the denominator of the calculated ratios. The reference condition can be chosen from the selected conditions only.
- **Peptide Feature Filter:**



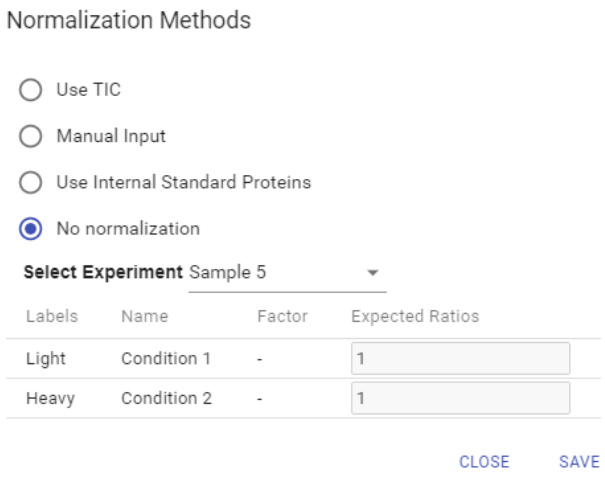
- **Average Area**: The minimum average area of all peptide features associated with the peptide to be included in the result. Features with higher abundance have been shown to be more reproducible across replicates.
- **Charge:** Only peptides with feature vectors that all fall within this range will be used in protein quantification.
- **Quality:** A quality value will be applied to each quantified peptide to represent the likely reproducibility of the quantification result. Set the minimum quality score cut-off. Factors that affect the quality score include m/z difference, RT difference, XIC shape similarity, and the feature intensities.
- **Number of labels with ID present**: Only peptides with an identified MS/MS scan in the number of given labels will be included in the peptide table.
- **Number of labels with feature present**: Only peptides with a predicted LC-MS peptide features in the number of given labels will be included in the peptide table.
- **Reference label present:**Only quantifiable peptides in which the reference label peptide feature is present will be included in the peptide table if this is checked.

- **Protein Filters:**



- **Significance:** Only protein groups with a significance above this threshold will be listed in the “Protein” view. The significance score is calculated as the -10lg of the significance testing p-value. Significance is calculated using welch's ANOVA.
- **FDR% (adjusted p-value):** Adopt the multiple hypothesis testing Benjamin-Hochberg method to adjust the p-value to the given false discovery rate. 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:** The required magnitude of change of abundance between conditions. The minimum value is one because the fold change is calculated as the magnitude of change where the lower value is always the denominator.
- **Unique Peptides:** Only protein groups with unique peptides at or above this number will be listed in the "Protein" view. Peptides are considered unique if they are only matched to one protein group.
- **Modified Form Exclusion:** When checked, peptides that are identified as both a modified and unmodified form will be excluded from the protein ration calculation. This does not apply to the quantification labels. When a peptide is split between an unmodified and modified form, the ratio of both forms may differ from the protein ratio. 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.

- **Normalization** :



- **Use TIC:** Using the total intensity of all quantifiable peptides from each condition, a global ratio is created. The global ratio is used as a normalization factor for all proteins and peptides.
- **Manual input:** This option performs the same default calculation as auto normalization ("Use TIC") but also allows for the manual change of each sample's expected ratios. This should be used whenever the expected ratio between conditions isn't 1:1.
- **Use internal standard proteins:** This option displays a list of identified proteins, which can be selected to create a normalization factor. A search bar is included to enable the quick location of select proteins.
- **No Normalization:** All samples are automatically assigned Factors of 1.0.

3. Understanding the SILAC Results

After a PEAKS Online SILAC quantification is complete, the SILAC summary page will display automatically. Beneath the Summary tab are the Protein and Peptide result tabs which are available to view.

- **Summary:** The outline of the SILAC search results with statistics. This is the place to examine the overall performance of the experiment and adjust filters.
- **Proteins:** The quantified proteins with a list of supporting peptide features for each protein.

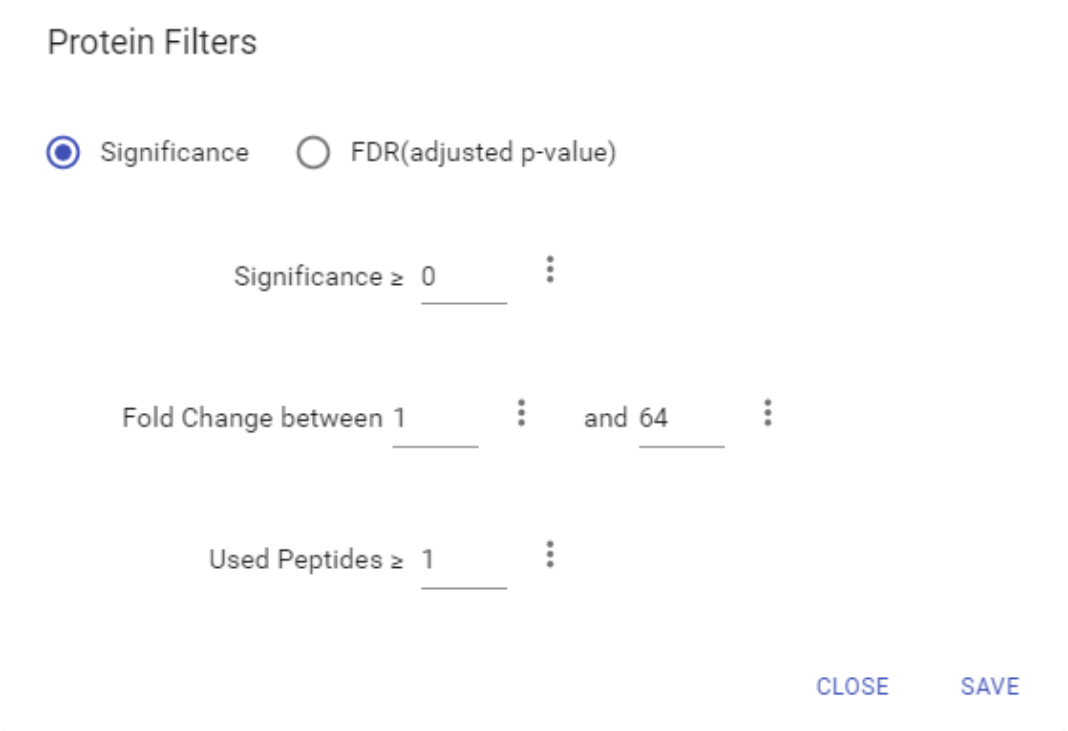
- **Peptides:** The quantified peptides are listed in a table. Extracted ion chromatogram (XIC) and supporting peptide features are also displayed for each peptide.

3.1. Result Filtration

Click the analysis name of a SILAC analysis, then click Modify Parameters, in the Quantification section, the following parameters are available.

3.1.1. Protein Filter

Click the button and the following parameter dialogue will pop up.



These filters determine what will be included in the protein group significance calculation, as well as what significance score thresholds will be set. The parameters used in the screenshot above indicate that no protein filter is applied, thus giving all quantifiable proteins.

- **Significance:** Only protein groups with a significance above this threshold will be listed in the “Protein” view. The significance score is calculated as the -10lg of the significance testing p-value. The significance is calculated using Welch's ANOVA.

- **FDR:** Adopt the Benjamin-Hochberg method for multiple hypothesis testing 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. Fold changes are always listed where the more intense sample is the denominator. So the fold change is always greater than 1.

- **Unique Peptides:** Only protein groups with unique peptides at or above this number will be listed in the "Protein" view. Unique peptides are peptides that only support one protein group.

3.2. Summary View

The "SILAC Summary" view includes an expression profile with candidate proteins sorted in a heat map, result statistics table, and protein volcano plots. An "Analysis Description" can be added and displayed below the filter section.

3.2.1. Statistics of Filtered Results

Table 1 in the Summary page lists the filtered results of feature vectors and proteins.

3.2.2. Protein Profile Heatmap

The heatmap in Figure 1 displays the protein groups that pass 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. If the number of proteins is too high, the heatmap will not be displayed.

3.2.3. Ratio Distribution

Figure 2 displays a histogram of the distribution of ratios relative to the reference condition for a specific sample. The sample displayed can be changed using the drop down menu.

3.2.4. The Volcano Plot for Proteins

A volcano plot (Figure 3 & Figure 4) 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 plots significance versus fold-change of the quantified proteins.

- **Horizontal dashed lines:** The selected significance threshold.
- **Vertical dashed lines:** The selected fold change thresholds.

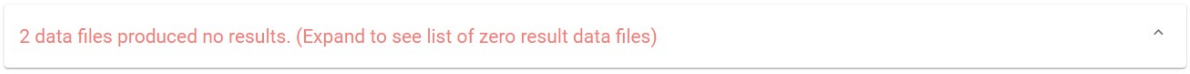
Plotting points in this way results in two regions of interest in the plot: those points that are found towards the top of the plot that are far to either the left or the right side. These represent values that display large magnitude fold changes as well as high statistical significance. The ratio is the group ratio set with respect to the base group.

Circles for the proteins that are above the set significance and fold change thresholds will be displayed in colour. Scrolling the cursor over a circle will show the protein info.

> **Note**
>
> Volcano plot cannot be displayed when multiple groups are selected.

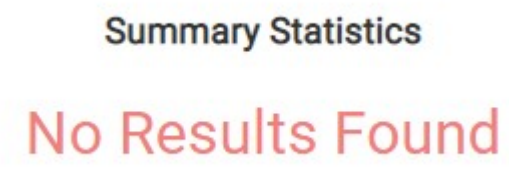
3.2.5. Low Quality Data Warnings

If PEAKS Online detects any data files producing very few or no results a warning section will appear at the top of the Summary Page indicating the number of data files producing poor or no results.



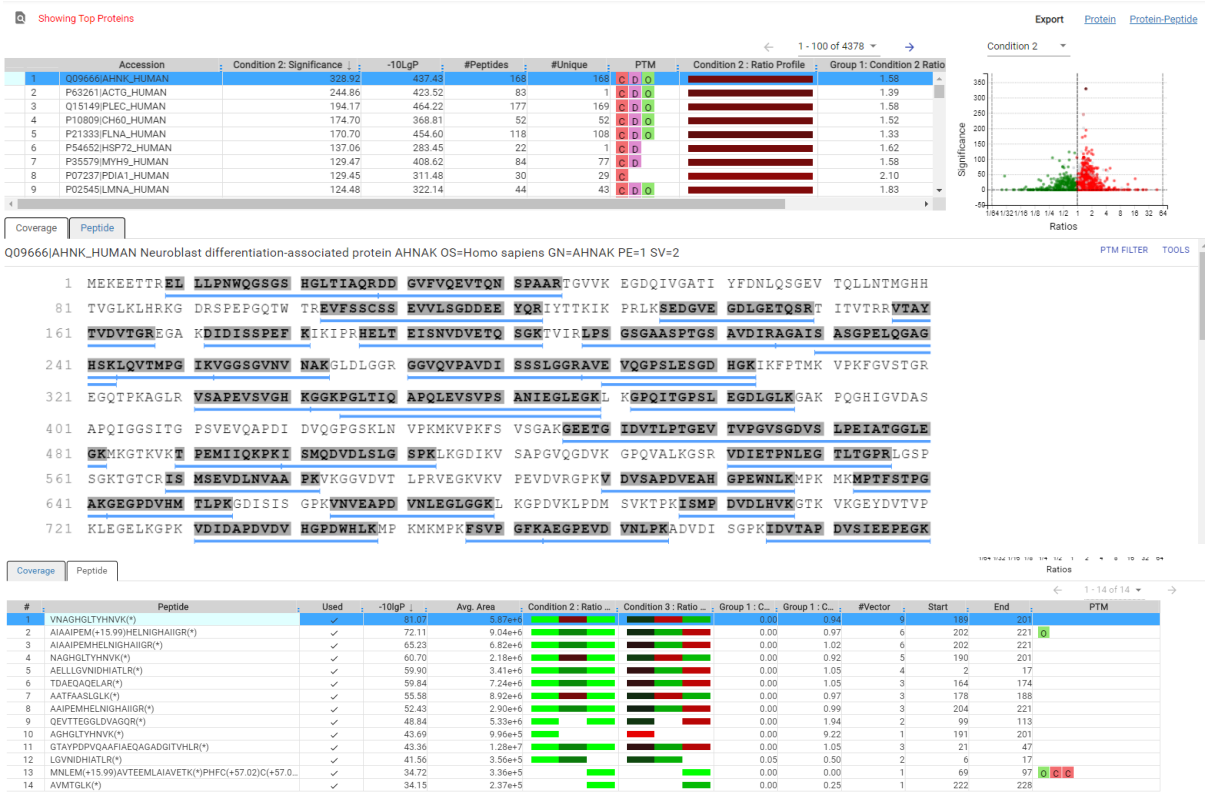
This section can be expanded to display a list of data files in the analysis that are producing poor or no results.

If there were no results found in any of your data files, a warning message stating "No Results Found" will be displayed instead of the usual Summary Page figures and tables.



3.3. Protein View

The SILAC Protein page lists all the quantified proteins that pass the filter set on the SILAC Summary page, characterizes each protein at the amino acid level, and lists the supporting peptide features of each protein.



3.3.1. 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:

- **Significance:** The protein significance is calculated as explained in Section 1, “Overview” in this chapter.

- **Condition Profile:** The protein abundance among the samples is depicted as a heat map. Hold the cursor on a profile to view the sample channels, abundances, and ratios with respect to the base sample.

- **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 all supporting feature vectors abundance within the filters. 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.

Refer to Chapter 6, Peptide, PTM, and Mutation Identification (PEAKS DB, PEAKS PTM, and SPIDER) Section 3.4.2, “Protein Table” for more details of other entries of the protein table.

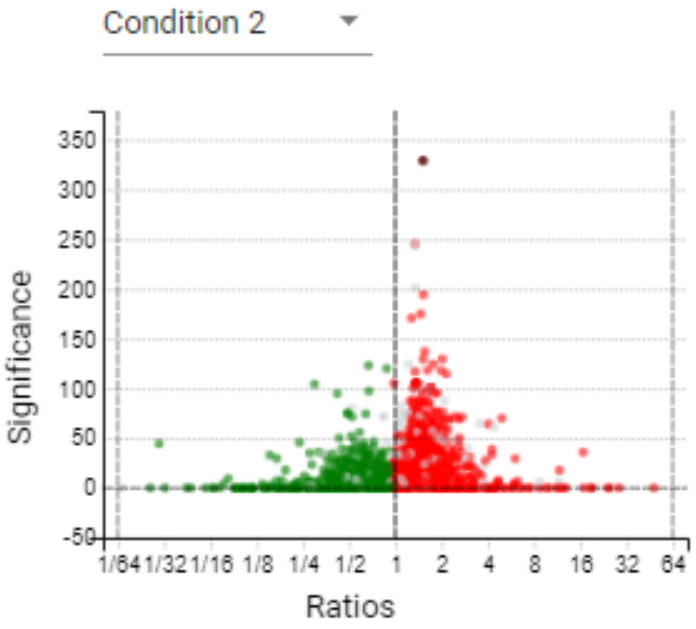
3.3.2. Protein Volcano Plot

The volcano plot facilitates identification of the significant proteins. This interactive volcano plot enables the selection of a protein in the table. 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 more abundant in the reference sample while red points are less abundant in the reference sample. Grayed points were filtered out.

Protein markers that are in color can be clicked and selected. Upon clicking a point on the volcano plot, the corresponding protein in the Protein Table will be highlighted and more details are displayed below the Protein Table.

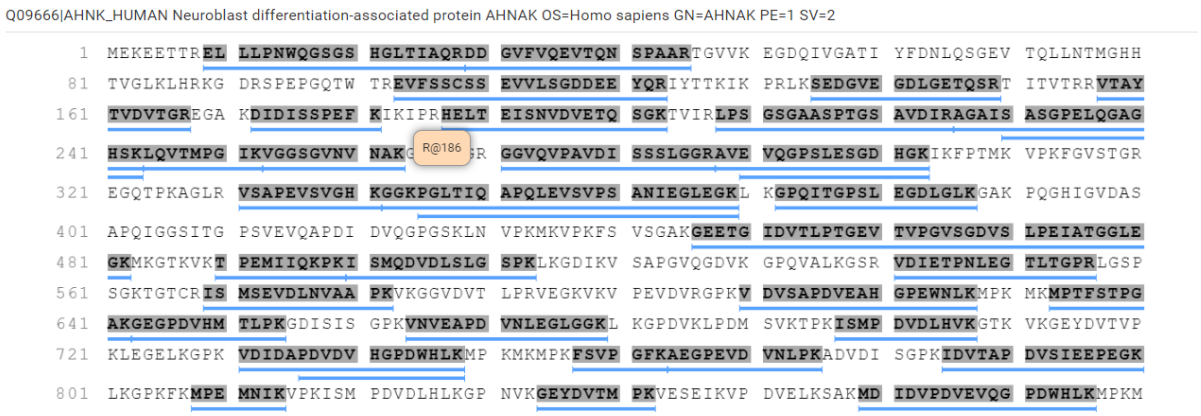
- **Selecting a Different Condition:** Click the condition dropdown above the volcano plot to display all volcano plot options, one for each experimental condition. Each volcano plot will show a different set of proteins that are relevant to that condition.

- **Zoom in/out from Volcano Plot:** Place the cursor on the volcano plot and drag a rectangle over the area of interest to Zoom in. By double clicking on the volcano plot you can zoom out and view the volcano plot with the default dimensions



3.3.3. Protein Coverage

The coverage view characterizes the selected protein sequences at the amino acid level. Refer to Chapter 6, Peptide, PTM, and Mutation Identification (PEAKS DB, PEAKS PTM, and SPIDER), Section 3.4.4, “Protein Coverage” for details.



3.3.4. Support Peptides

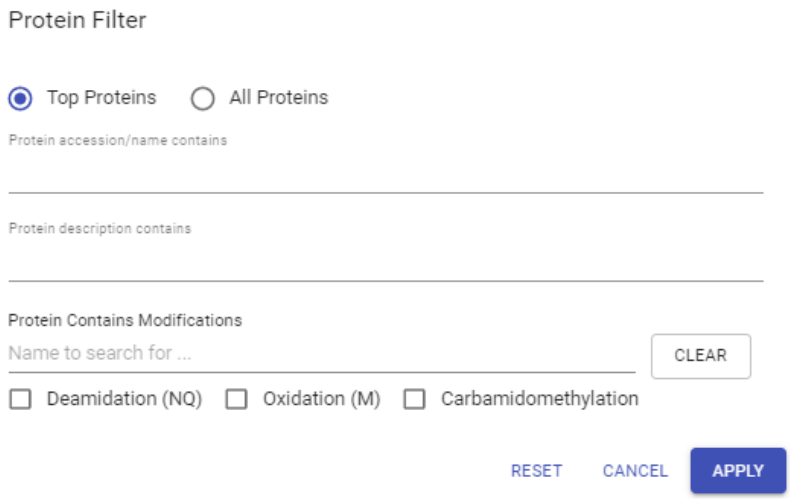
The supporting peptides assigned to the protein are shown in a separate tab beside the Protein Coverage. The sequence can be clicked and a pop-up window will show up to display the feature details.

This table contains the following information of the quantifiable support peptides for the selected protein in the protein table.

- **Peptide:** The amino acid sequence of the peptide, as determined by the 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. All peptides which pass the peptides filters are used to calculate the protein abundance.
- **-10LgP:** The peptide -10LgP score. The score indicates the likelihood that the peptide-spectrum match identification is correct.
- **Avg. Area:** Each peptide feature associated with the peptide has its own area under the XIC curve. The average area of is the average area of all the features associated with the peptide.
- **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 show 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 supporting peptide feature abundances within the peptide filters. 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 of the samples in that group.
- **#Vector:** This shows the number of quantifiable feature vectors of a peptide. A feature vector is a full set of peptide features used to calculate a ratio.
- **Start:** This shows the protein position of the first residue of the peptide that covers the protein.
- **End:** This shows the protein 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. Scroll over the modification icon to see the modification name and mass.
- **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.

3.3.5. Protein View Filter

Click the button on top-left of the protein table to open the protein filter and control which proteins to display.



- **Top Proteins**: The number of proteins that have the highest number of unique peptides supporting them within their protein group.
- **All Proteins**: The number of proteins given the parameters set for PEAKS Protein Score(-10lgP) and number of unique peptides.
- **Protein Sequence contains**: Enter a protein sequence, only proteins that contain that sequence will be displayed in the protein table
- **Protein description contains:** Enter a protein description, only proteins that contain that description will be displayed in the protein table
- **Include Contaminant Data:** Check this to display identifications from the contaminant database that were identified during a PEAKS DB search.
- **Protein Contains Modifications**: Check boxes are provided for each identified modification.

3.4. 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 detailed information.

3.4.1. 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 3.3.4, “Support Peptides”, with some new columns.

- **Quality Score:** The quality score represents the predicted reproducibility of the quantification result. Factors that affect the quality score include m/z difference, RT difference, XIC shape similarity, and the feature intensities.

- **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.

The Peptide table entries may be filtered by using the **Peptide View Filters** on top of the peptide table, for more information refer to Section 3.5.1, “Peptide View Filters” in Chapter 6, Peptide, PTM, and Mutation Identification (PEAKS DB, PEAKS PTM, and SPIDER).

3.4.2. Peptide View Filters

Click the  button on the upper left corner of the peptide table to open the peptide filter and control which peptides to display.

- **Peptide Sequence Contains:** Enter a peptide sequence, only peptides that contain that sequence will be displayed in the peptide table

- **Protein Accession Contains:** Enter a peptide accession, only peptides that contain that accession will be displayed in the peptide table

- **Show Transferred IDs Only:** Only display peptides where match between runs was used to include at least one peptide feature in the result.

- **Ratio Filter:** Only peptides where the ratio for the selected group and condition fall within the specified range will be displayed in the peptide table.

- **Peptide Contains Modifications:** Use the checkboxes to indicate which modified peptides to include in the result.

3.4.3. Feature Details

The **Feature Details** contains the eXtracted Ion Chromatogram (XIC) chart, Ion Match, and the Reporter Ion Spectrum.

- **Feature Menu:** It is under the peptide table. The feature menu displays the information of the feature that is being displayed in the spectrum.

  - **Samples:** If a peptide is quantified in more than one sample, click the "SAMPLE" drop-down menu to view the Spectrum Annotation of the other PSMs.

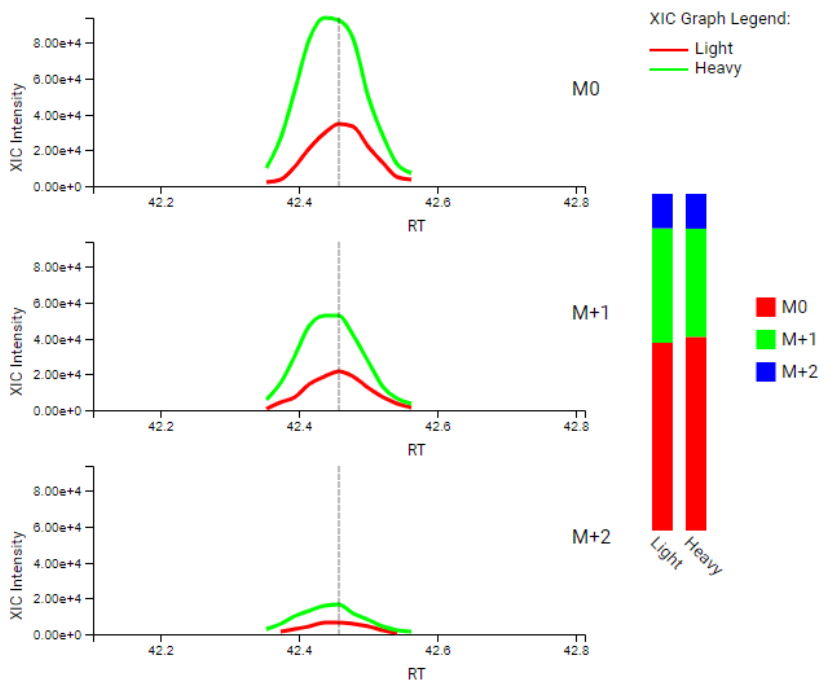
  - **Feature:** Click the Feature drop down menu to show all peptide features included in the result: -10lgP, Avg. Area, Condition Intensity, Condition Ratio, Id Count, m/z, z, RT, and ppm for each feature are given.

    - If the transferred column for the feature displays "true" that means the ID for at least one feature was transferred.

  - **PSM Label:** Click the PSM Label drop-down menu to show other identified MS/MS scans associated with the peptide.

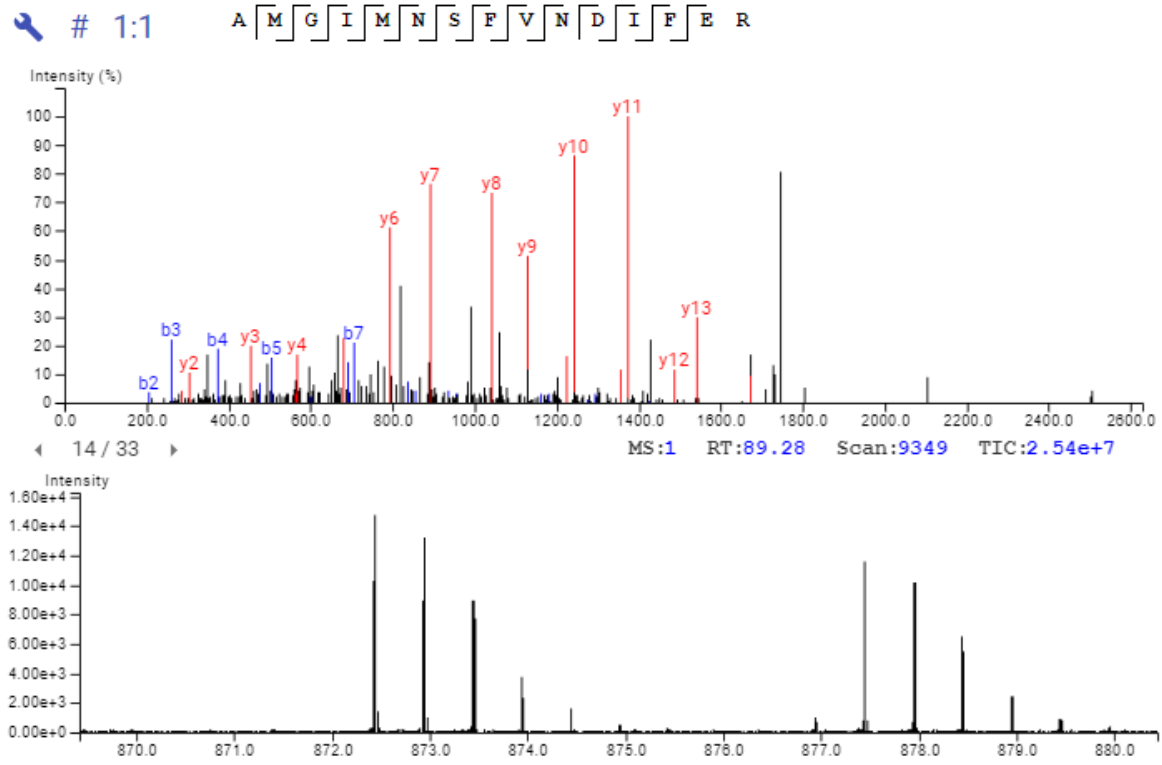
3.4.3.1. eXtracted Ion Chromatogram (XIC) chart

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, etc. – of each label is included, which shows the relative composition of each label.



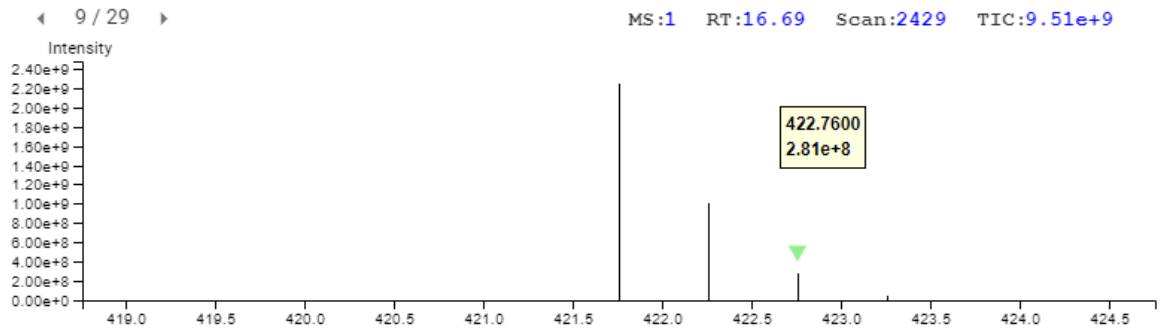
3.4.3.2. Ion Match

This section displays the spectrum annotation and other information about the peptide selected in the Peptide table. When multiple spectra are matched to the peptide, only the information about the top-scoring PSM is shown by default. The interface of this section is identical to the Spectrum Annotation in the PEAKS De Novo result. For detailed instructions, refer to Section 3.5, “Spectrum Annotation” in Chapter 5, Peptide *de novo* Sequencing.



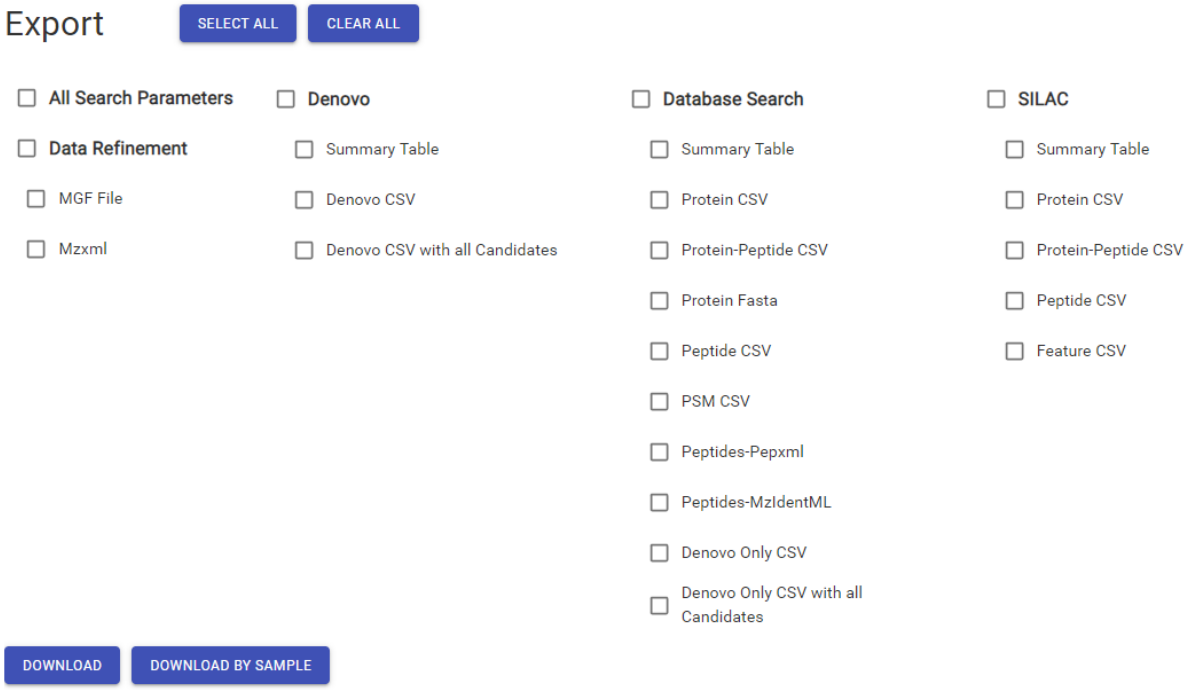
3.4.3.3. MS1 Spectrum

This section provides an MS1 spectrum zoomed in on the reporter ions in the experiment. This is an interactive spectrum, enabling zooming in and out using the scroll wheel of the mouse. Double-clicking will zoom in on an area of interest. Clicking and dragging will slide the spectrum to the left or right.



3.5. Result Exporting

PEAKS SILAC results can be exported to various text formats in the Export page. The protein and peptide results can also be exported from the Protein and Peptide page, respectively.



- **Proteins CSV:** These are the quantified proteins that pass the filters set in the "Summary" page and their associated details.
- **Protein-Peptide CSV:** These are the supporting peptides of the quantified proteins and their associated details.
- **Peptide CSV:** A complete list of all quantified peptides that pass the peptide feature filters set in the "Summary" page and their associated details.
- **Feature CSV:** All predicted peptide features and their associated quantification information. Both identified and unidentified peptide features are included. To include unidentified peptide features, set Peptide Id Count equals to 0 in the Peptide feature filter.

Note

Exporting by sample will begin downloading the same .zip folder, since this does not apply to SILAC quantification.

Chapter 12. Fraciton Assisted ID Transfer 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 Online. 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.

Compared to the regular SILAC, ID transfer can provide quantification workflow with high throughput and coverage. One group is designated as the ID Transfer group. It is used to associate peptide features from other samples to identified MS/MS scans in the ID transfer group. It is recommended that the ID Transfer group be a highly fractionated sample. The experimental design is outlined below.

- To get high coverage
  1. SILAC sample -> fractionation -> LC-MS runs (only once)
  2. Get a list of SILAC pairs by combining all fractions
- To get high throughput
  1. SILAC sample -> LC-MS (routine)
  2. Transfer SILAC pairs from pre-run fractions while performing data analysis

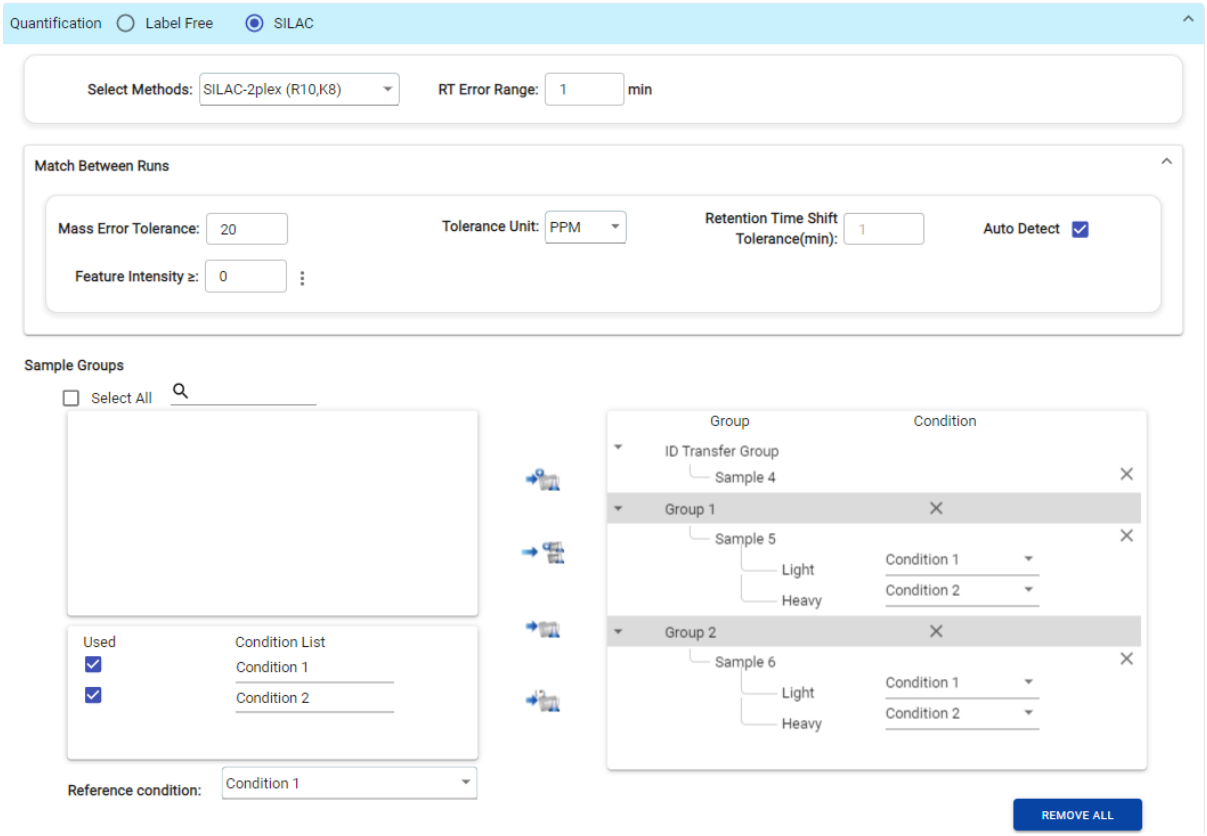
In the ID transfer SILAC, the analysis steps are outlined below.

- Get expected SILAC pairs from the fractionated sample
  1. Search against a target protein database using PEAKS DB.
  2. Generate a SILAC library for future ID transfer
- Routine SILAC quantification
  1. Find SILAC pairs directly from LC-MS/MS
  2. For remaining unidentified precursors/pairs, SILAC pairs are transferred from the ID transfer group run previously
  3. Combine 1 and 2 to get the final quantification results

The usage of this workflow is outlined below.

1. Refer to Chapter 4, Peptide De Novo Sequencing Section 2, “*DeepNovo/De Novo* Analysis Workflow and Parameter Settings”, step 1-6 to create a new project and add data.
2. SILAC quantification can be run on PEAKS Database Search results. Select the "Fractionation Assisted ID Transfer Quantification" workflow in the DDA section of the Workflow step.
   - To run SILAC quantification on PEAKS Database Search results, select the "Fractionation Assisted ID Transfer Quantification" workflow in the DDA section of the Workflow step.

3. Set Search Parameters.
    - If running SILAC quantification on a DDA DB Search analysis, refer to Chapter 5, Peptide, PTM, and Mutation Identification (PEAKS DB, PEAKS PTM, and SPIDER), LINK, for details. Section 2, “Database Search Workflow and Parameter Settings”
4. For an ID Transfer SILAC workflow, the SILAC-specific parameters will be shown at the bottom of the Project Submission page. Choose an ID Transfer Group and set the SILAC sample grouping as well as the Retention Time Error Tolerance. See Section 2 below for more details. ID Transfer SILAC quantification will be performed on the combined identification result.



5. Set the PSM Filters. Refer to Chapter 5, Peptide, PTM, and Mutation Identification (PEAKS DB, PEAKS PTM, and SPIDER) Section 2.8, “PSM Filter”, for details.
6. Review project settings and parameters, then click the Submit button at the top of the page.
7. Name the analysis, then click OK to start the analysis.

2. Setting Parameters

For more information on DB search parameters, refer to Chapter 5, Peptide, PTM, and Mutation Identification (PEAKS DB, PEAKS PTM, and SPIDER), Section 2, “Database Search Workflow and Parameter Settings”.

The following parameters are available when a SILAC Quantification workflow is selected.

- **Select Method:** From the Select Methods drop-down menu, choose the appropriate quantification method used in the experiment.
- **RT Error Range:** The retention time shift tolerance is the maximum elution time difference that is considered for the quantification of an identified peptide after RT alignment.
- **Match Between Runs** Match Between Runs is enabled for all ID Transfer SILAC. For more information on Match Between Runs, refer to Chapter 10, Stable Isotope Labeling by Amino acids in Cell culture (SILAC), Section 2, “Setting Parameters”.

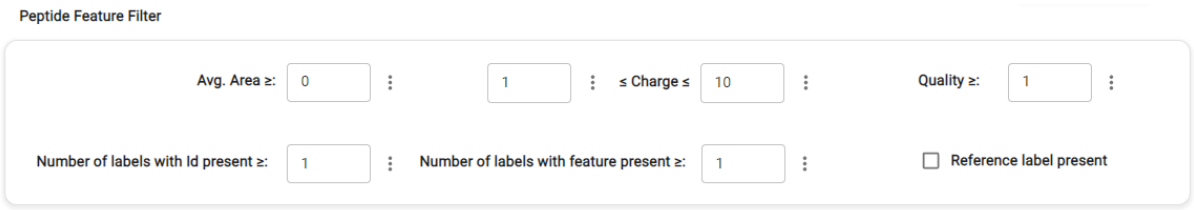
- **Samples:** The MS samples are listed for selection. After selecting a sample, it can be added to a new group by clicking the button or to an existing group by clicking the button and selecting that option from the drop-down menu. To add multiple samples into their own separate groups, click the button.

- **Groups**: The samples can be arranged into one or more groups. A group represents a state of the experiment where quantitative ratios between groups 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 'X' next to the sample or the Remove All button, respectively.

- **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 represents one unique condition. To remove a label from analysis, uncheck the corresponding condition.

- **Reference Condition**: The condition which is the denominator of the calculated ratios. The reference condition can be chosen from the selected conditions only.

- **Peptide Feature Filter:**



  - **Average Area**: The minimum average area of all peptide features associated with the peptide to be included in the result. Features with higher abundance have been shown to be more reproducible across replicates.

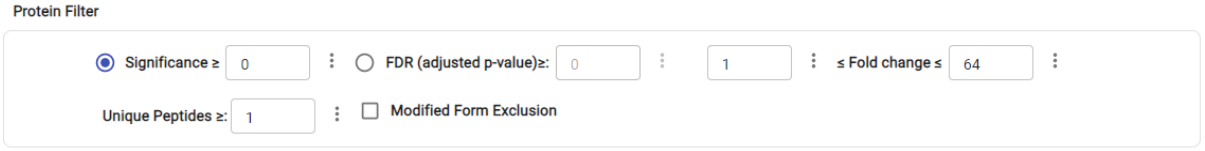
  - **Charge:** Only peptides with feature vectors that all fall within this range will be used in protein quantification.

  - **Reference Label:** Only quantifiable peptides in which the reference label is present will be displayed if this is checked.

  - **Number of labels with ID present**: Only protein groups with a number of labels with ID present at or above this number will be listed in the "Protein" view.

  - **Number of labels with feature present**: Only protein groups with a number of labels with feature present at or above this number will be listed in the "Protein" view.

- **Protein Filters:**

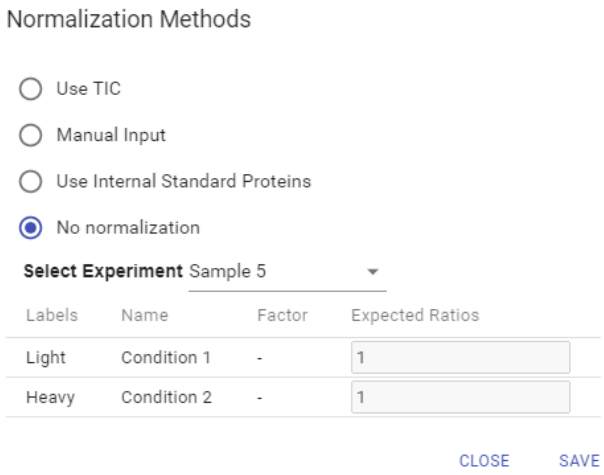


  - **Significance:** Only protein groups with a significance above this threshold will be listed in the "Protein" view. The significance score is calculated as the -10lg of the significance testing p-value. PEAKS provides the significance testing method: ANOVA.

  - **FDR% (adjusted p-value):** 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 unique peptides at or above this number will be listed in the "Protein" view. Peptides are considered unique if they are only matched to one protein group.
- **Modified Form Exclusion:** When checked, 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.

- **Normalization** :



- **Use TIC:** Using the total intensity of all quantifiable peptides from each condition, a global ratio is created. The global ratio is used as a normalization factor for all proteins and peptides.
- **Manual input:** This option performs the same default calculation as auto normalization ("Use TIC") but also allows for the manual change of each sample's expected ratios. This should be used whenever the expected ratio between conditions isn't 1:1.
- **Use internal standard proteins:** This option displays a list of identified proteins, which can be selected to create a normalization factor. A search bar is included to enable the quick location of select proteins.
- **No Normalization:** All samples are automatically assigned Factors of 1.0.

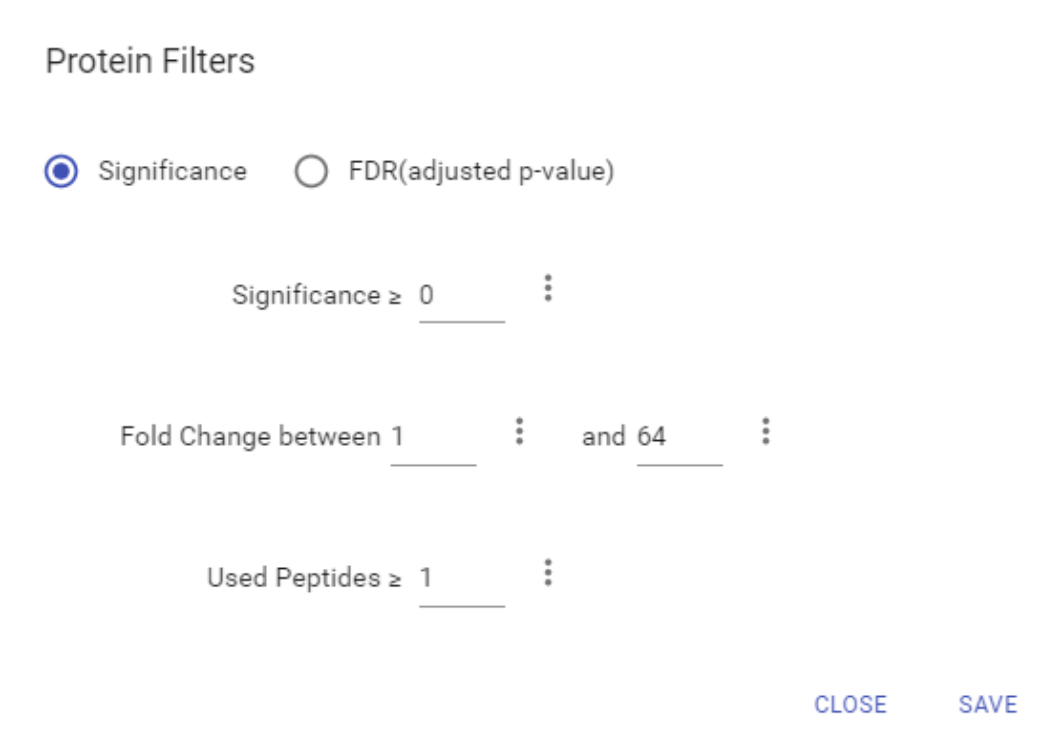
3. Understanding the SILAC Results

After a PEAKS Online SILAC quantification is complete, the SILAC summary page will display automatically. Beneath the Summary tab are the Protein and Peptide result tabs which are available to view.

- **Summary:** The outline of the SILAC search results with statistics. This is the place to examine the overall performance of the experiment and adjust filters.
- **Proteins:** The quantified proteins with a list of supporting peptide features for each protein.
- **Peptides:** The quantified peptides are listed in a table. Extracted ion chromatogram (XIC) and supporting peptide features are also displayed for each peptide.

3.1. Protein Filter

Click the ✎ button and the following parameter dialogue will pop up.



These filters determine what will be included in the protein group significance calculation, as well as what significance score thresholds will be set. The parameters used in the screenshot above indicate that no protein filter is applied, thus giving all quantifiable proteins.

- **Significance:** Only protein groups with a significance above this threshold will be listed in the “Protein” view. The significance score is calculated as the -10lg of the significance testing p-value. PEAKS provides significance testing method: ANOVA.
- **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. Fold changes are always listed where the more intense sample is the denominator. So the fold change is always greater than 1.
- **Used Peptides:** Only protein groups with unique peptides at or above this number will be listed in the "Protein" view. Unique peptides are peptides that only support one protein group.

3.2. Summary View

The "SILAC Summary" view includes an expression profile with candidate proteins sorted in a heat map, result statistics table, and protein volcano plots. An "Analysis Description" can be added and displayed below the filter section.

3.2.1. Statistics of Filtered Results

Table 1 in the Summary page lists the filtered results of feature vectors and proteins.

3.2.2. Protein Profile Heatmap

The heatmap in Figure 1 displays the protein groups that pass 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. If the number of proteins is too high, the heatmap will not be displayed.

3.2.3. Protein Profile Heatmap

A volcano plot (Figure 2 & Figure 3) 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 1 plots significance versus fold-change of the quantified proteins.

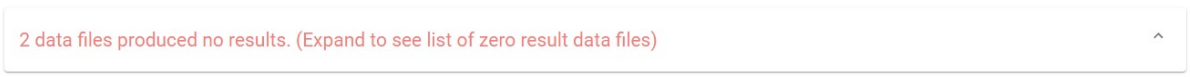
- **Horizontal dashed lines:** The selected significance threshold.
- **Vertical dashed lines:** The selected fold change thresholds.

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). 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.

Circles for the proteins that are above the set significance and fold change thresholds will be displayed in colour. Scrolling the cursor over a circle will show the protein info.

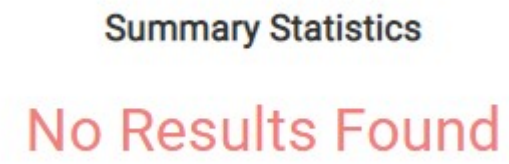
3.2.4. Low Quality Data Warnings

If PEAKS Online detects any data files producing very few or no results a warning section will appear at the top of the Summary Page indicating the number of data files producing poor or no results.



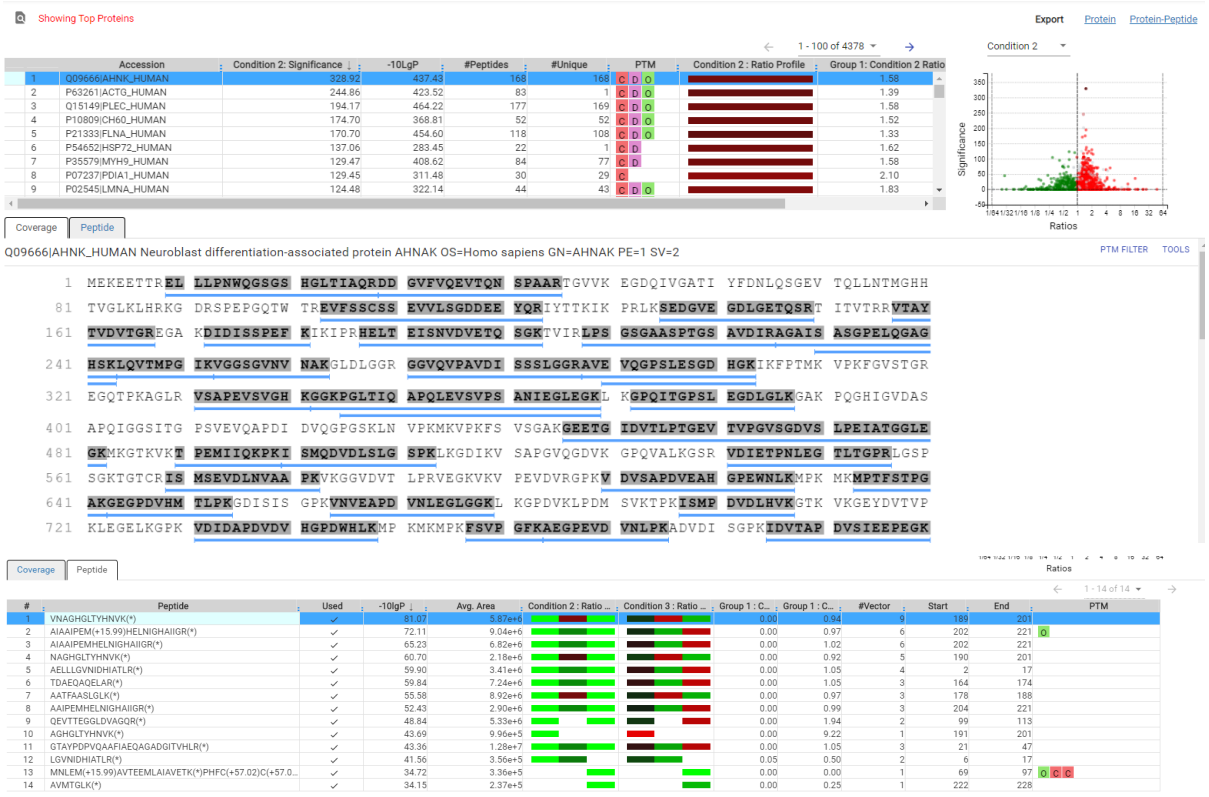
This section can be expanded to display a list of data files in the analysis that are producing poor or no results.

If there were no results found in any of your data files, a warning message stating "No Results Found" will be displayed instead of the usual Summary Page figures and tables.



3.3. Protein View

The SILAC Protein page lists all the quantified proteins that pass the filter set on the SILAC Summary page, characterizes each protein at the amino acid level, and lists the supporting peptide features of each protein.



3.3.1. 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:

- **Significance:** The protein significance is calculated as explained in Section 1, “Overview” in this chapter.

- **Condition Profile:** The protein abundance among the samples is depicted as a heat map. Hold the cursor on a profile to view the sample channels, abundances, and ratios with respect to the base sample.

- **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.

Refer to Chapter 5, Peptide, PTM, and Mutation Identification (PEAKS DB, PEAKS PTM, and SPIDER) Section 3.4.2, "Protein Table" for more details of other entries of the protein table.

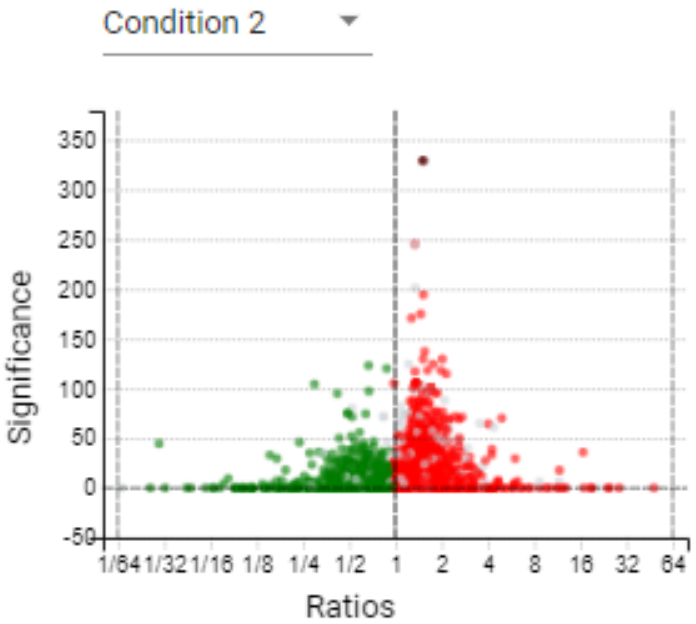
3.3.2. Protein Volcano Plot

The volcano plot facilitates identification of the significant proteins. This interactive volcano plot enables the selection of a protein in the table. 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 set in the filter. The volcano plot facilitates identification of significant proteins.

Protein markers that are in color can be clicked and selected. Upon clicking a point on the volcano plot, the corresponding protein in the Protein Table will be highlighted and more details are displayed below the Protein Table.

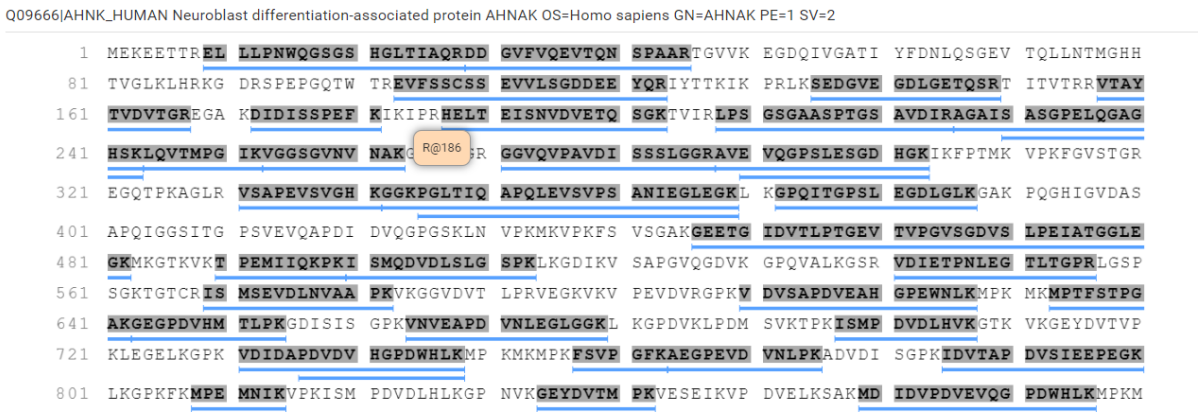
- **Selecting a Different Condition:** Click the condition dropdown above the volcano plot to display all volcano plot options, one for each experimental condition. Each volcano plot will show a different set of proteins that are relevant to that condition.

- **Zoom in/out from Volcano Plot:** Place the cursor on the volcano plot and drag a rectangle over the area of interest to Zoom in. By double clicking on the volcano plot you can zoom out and view the volcano plot with the default dimensions



3.3.3. Protein Coverage

The coverage view characterizes the selected protein sequences at the amino acid level. Refer to Chapter 5, Peptide, PTM, and Mutation Identification (PEAKS DB, PEAKS PTM, and SPIDER), Section 3.4.4, "Protein Coverage" for details.



3.3.4. Support Peptides

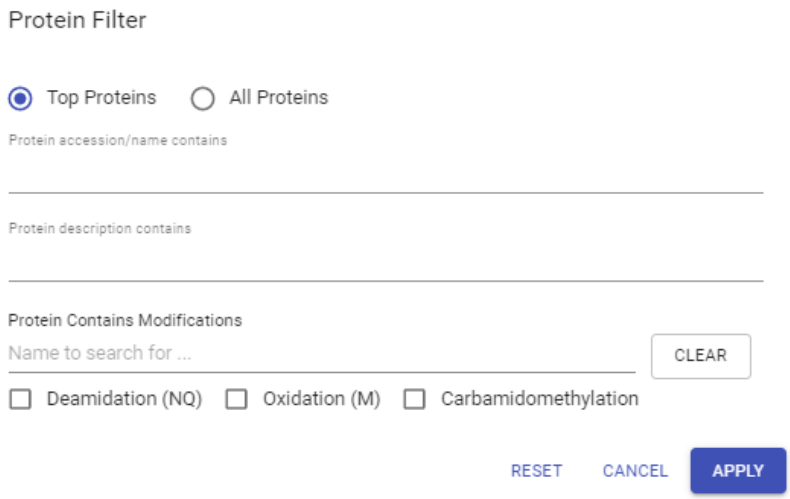
The supporting peptides assigned to the protein are shown in a separate tab beside the Protein Coverage. The sequence can be clicked and a pop-up window will show up to display the feature details.

This table contains the following information of the quantifiable support peptides for the selected protein in the protein table.

- **Peptide:** The amino acid sequence of the peptide, as determined by the 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. All peptides which pass the peptides filters are used to calculate the protein abundance.
- **-10LgP:** The peptide -10LgP score. The score indicates the likelihood that the peptide-spectrum match identification is correct.
- **Avg. Area:** Each peptide feature associated with the peptide has its own area under the XIC curve. The average area of is the average area of all the features associated with the peptide.
- **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 show 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:** This shows the number of quantifiable feature vectors of a peptide.
- **Start:** This shows the protein position of the first residue of the peptide that covers the protein.
- **End:** This shows the protein 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. Scroll over the modification icon to see the modification name and mass.
- **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.

3.3.5. Protein View Filter

Click the button on top-left of the protein table to open the protein filter and control which proteins to display.



- **Top Proteins**: The number of proteins that have the highest number of unique peptides supporting them within their protein group.
- **All Proteins**: The number of proteins given the parameters set for PEAKS Protein Score(-10lgP) and number of unique peptides.
- **Protein Sequence contains**: Enter a protein sequence, only proteins that contain that sequence will be displayed in the protein table
- **Protein description contains:** Enter a protein description, only proteins that contain that description will be displayed in the protein table
- **Include Contaminant Data:** Check this to display identifications from the contaminant database that were identified during a PEAKS DB search.
- **Protein Contains Modifications**: Check boxes are provided for each identified modification.

3.4. 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 detailed information.

3.4.1. 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 3.3.4, “Support Peptides”, with some new columns.

- **Quality Score:** 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.

- **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.

The Peptide table entries may be filtered by using the **Peptide View Filters** on top of the peptide table, for more information refer to Section 3.5.1, “Peptide View Filters” in Chapter 5, Peptide, PTM, and Mutation Identification (PEAKS DB, PEAKS PTM, and SPIDER).

3.4.2. Peptide View Filters

Click the  button on the upper left corner of the peptide table to open the peptide filter and control which peptides to display.

- **Peptide Sequence Contains:** Enter a peptide sequence, only peptide that contain that sequence will be displayed in the peptide table

- **Peptide Accession Contains:** Enter a peptide accession, only peptides that contain that accession will be displayed in the peptide table

- **Protein Contains Modifications:** Check boxes are provided for each identified modification.

3.4.3. Feature Details

The **Feature Details** contains the eXtracted Ion Chromatogram (XIC) chart, Ion Match, and the Reporter Ion Spectrum.

- **.Feature Menu:** It is under the peptide table. The feature menu displays the information of the feature that is being displayed in the spectrum.

  - **Samples:** If a peptide is quantified in more than one sample, click the "SAMPLE" drop-down menu to view the Spectrum Annotation of the other PSMs.

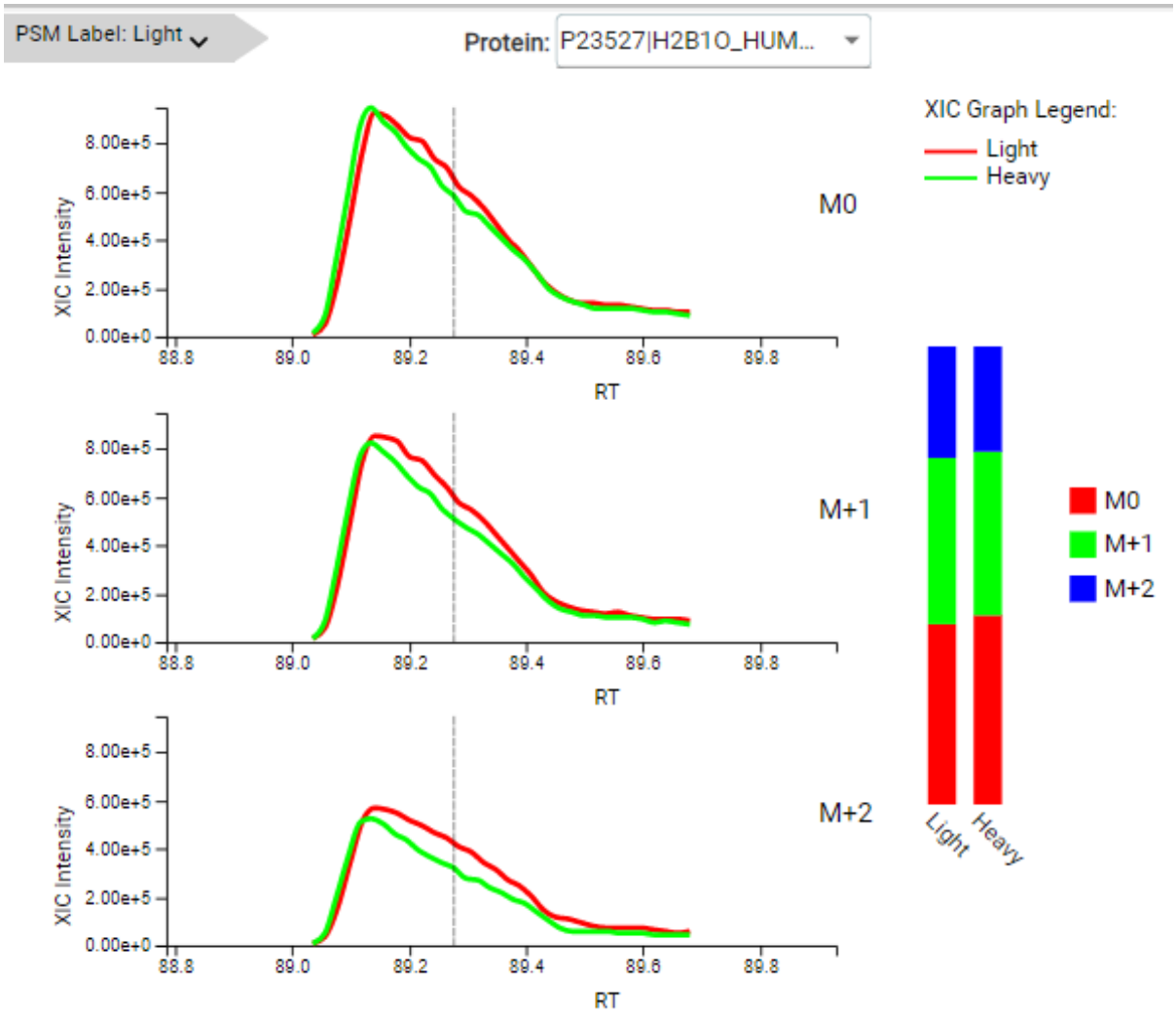
  - **Feature:** Click the Feature drop down menu to show all peptide features included in the result: -10lgP, Avg. Area, Condition Intensity, Condition Ratio, Id Count, m/z, z, RT, and ppm for each feature are given.

    - If the transferred column for the feature displays "true" that means the ID for at least one feature was transferred.

  - **PSM Label:** Click the PSM Label drop-down menu to show other identified MS/MS scans associated with the peptide.

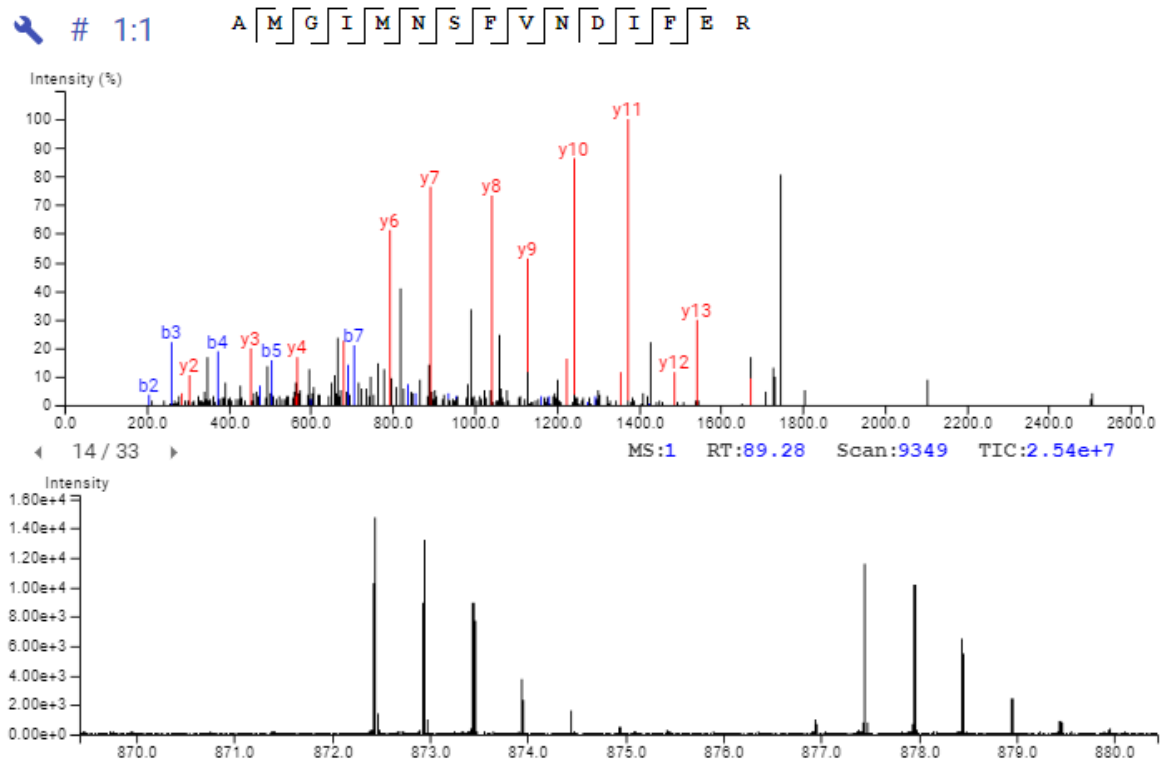
3.4.3.1. eXtracted Ion Chromatogram (XIC) chart

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, etc. – of each label is included, which shows the relative composition of each label.



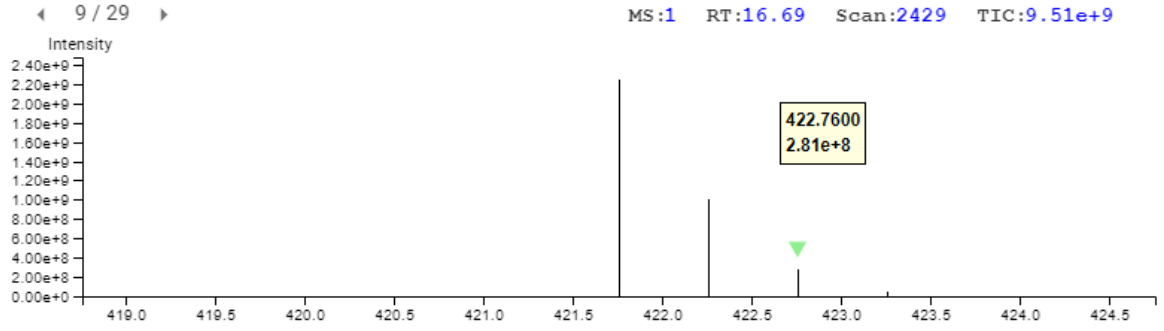
3.4.3.2. Ion Match

This section displays the spectrum annotation and other information about the peptide selected in the Peptide table. When multiple spectra are matched to the peptide, only the information about the top-scoring PSM is shown by default. The interface of this section is identical to the Spectrum Annotation in the PEAKS De Novo result. For detailed instructions, refer to Section 3.5, “Spectrum Annotation” in Chapter 4, Peptide *de novo* Sequencing.



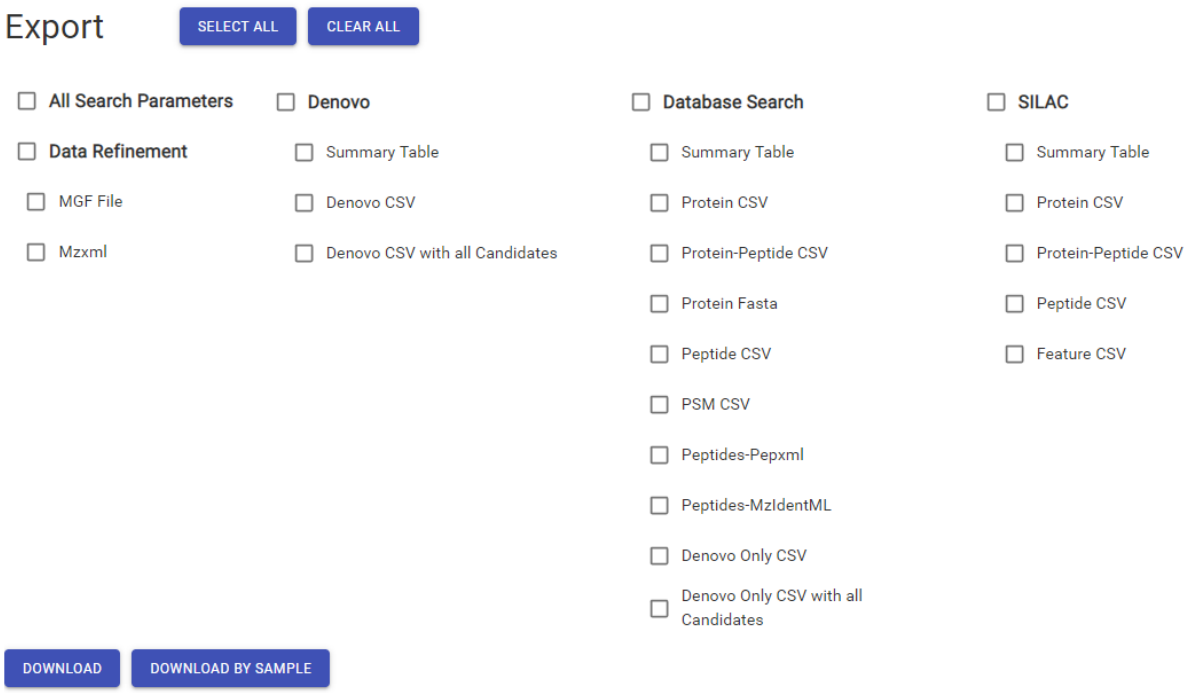
3.4.3.3. MS1 Spectrum

This section provides an MS1 spectrum zoomed in on the reporter ions in the experiment. This is an interactive spectrum, enabling zooming in and out using the scroll wheel of the mouse. Double-clicking will zoom in on an area of interest. Clicking and dragging will slide the spectrum to the left or right.



3.5. Result Exporting

PEAKS SILAC results can be exported to various text formats in the Export page. The protein and peptide results can also be exported from the Protein and Peptide page, respectively.



- **Proteins CSV:** These are the quantified proteins that pass the filters set in the "Summary" page and their associated details.
- **Protein-Peptide CSV:** These are the supporting peptides of the quantified proteins and their associated details.
- **Peptide CSV:** A complete list of all quantified peptides that pass the peptide feature filters set in the "Summary" page and their associated details.
- **Feature CSV:** All predicted peptide features and their associated quantification information. Both identified and unidentified peptide features are included. To include unidentified peptide features, set Peptide Id Count equals to 0 in the Peptide feature filter.

Note

Exporting by sample will begin downloading the same .zip folder, since this does not apply to SILAC quantification.

Chapter 13. DeepNovo Peptidome Workflow

1. Overview

The DeepNovo peptidome workflow takes advantage of the latest *de novo* technology *DeepNovo* developed using deep learning. It's designed for identification of undigested peptides from peptidomic datasets. The algorithm was trained with an immunopeptidomic dataset for accurate and sensitive identification of undigested peptides. PEAKS online now supports DeepNovo Peptidome workflows for both DDA and DIA data acquisition modes.

Note

For more details, refer to the paper: *"Deep learning enables de novo peptide sequencing from data-independent-acquisition mass spectrometry", Nat Methods 16, 63–66 (2019).*

.

2. DDA Peptidome Workflow and Parameter Settings

DeepNovo and Peptide Database Search can be launched together as a DeepNovo Peptidome workflow in a PEAKS Online search. An analysis report is generated to show the combined results. DeepNovo peptides are used to assist the following protein database search using PEAKS DB allowing for accurate and sensitive identification of undigested peptides present in the protein sequencing database.

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. 2012 Apr;11(4):M111.010587.*

Spectra unmatched by database search are re-analyzed using the *de novo* peptide sequence homology search SPIDER to allow for accurate identification of peptides with a single amino acid variation from the target database.

Note

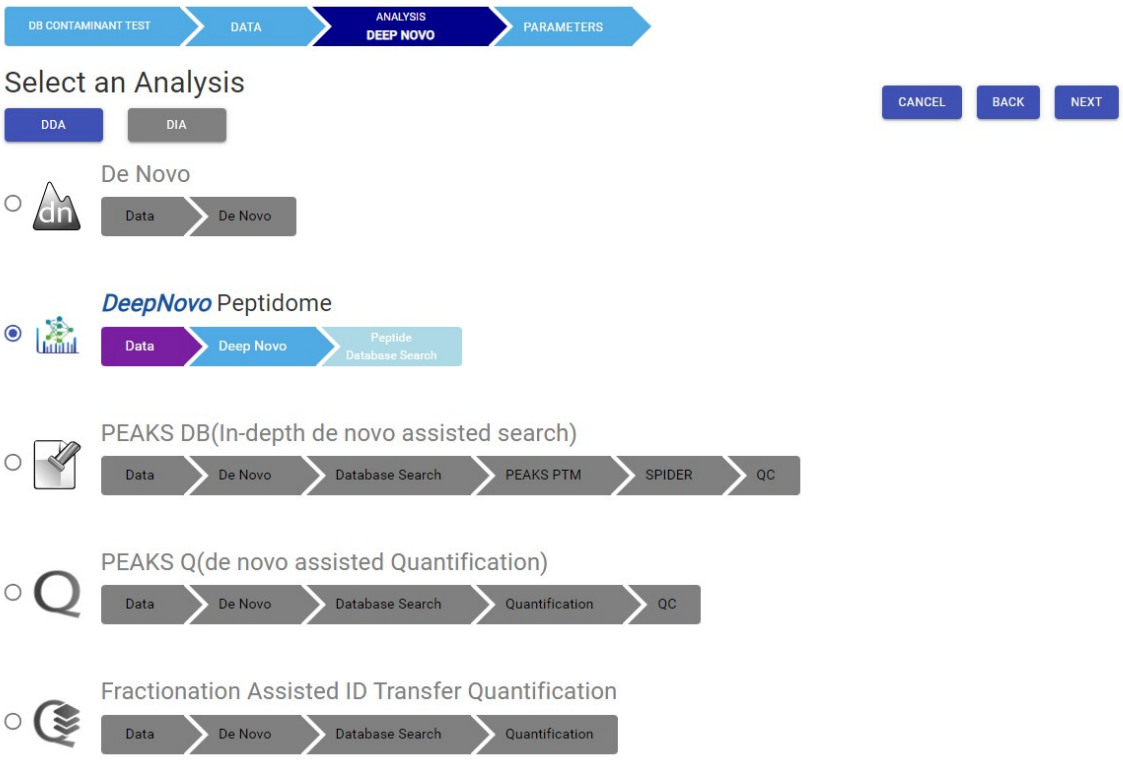
For more details, refer to the paper: *"SPIDER: software for protein identification from sequence tags with de novo sequencing error", J Bioinform Comput Biol . 2005 Jun;3(3):697-716.*

The combination of the results are then presented in the peptide table where the most confident identification that passes the given filters is given for each scan in the dataset.

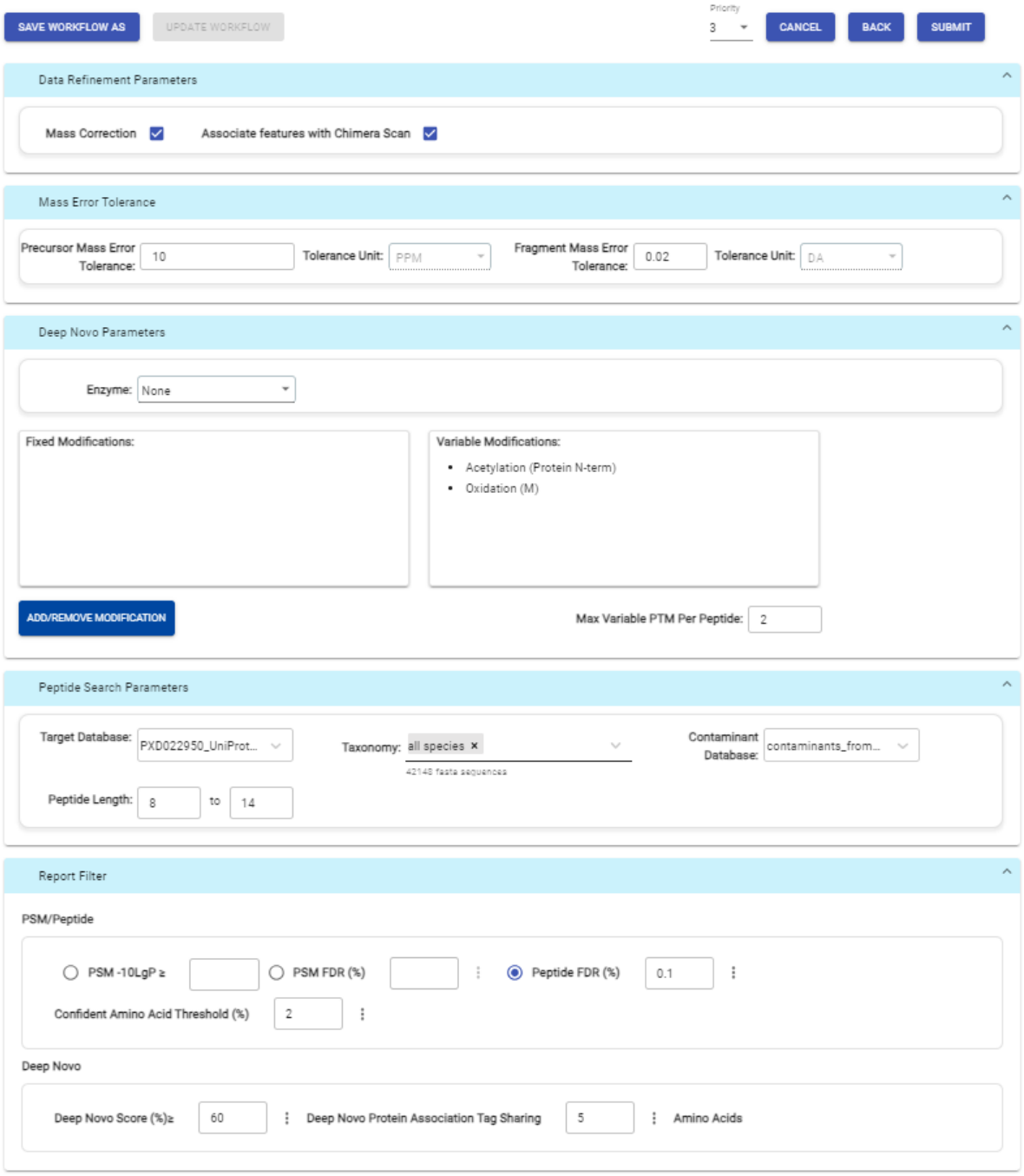
To conduct a database search workflow, follow the steps below:

1. Refer to Chapter 5, Peptide De Novo Sequencing, Section 2, “*DeepNovo/De Novo* Analysis Workflow and Parameter Settings” , Step 1-6 to create a new project and add data.

2. Select the "DeepNovo Peptidome" workflow.



3. Set the DeepNovo and Peptide Search parameters. See more information regarding the parameter settings in section 2.1-2.5 below.



4. Click SUBMIT button to start the search.

2.1. Data Refinement Parameters

Mass Correction:: The precursor m/z values given by some instruments are often not of the monoisotopic ion. This creates problems in downstream analysis. By enabling mass correction, PEAKS will examine the isotope shapes in the corresponding MS scans and accurately adjust the precursor m/z to equal the monoisotopic ion m/z.

Associate Features with Chimera Scan: If enabled, PEAKS will assume that a tandem scan may contain fragments from two or more peptides, which will allow PEAKS to identify co-eluting peptides found within the acquisition window of the scan. PEAKS accomplishes this by searching for peptides that match the monoisotopic mass of any peptide feature in the acquisition window.

2.2. Precursor and Fragment Error Tolerance

Precursor mass : This enables the selection of a precursor mass error tolerance. PEAKS Online DB Search identifies peptides within the precursor mass tolerance, which is the allowable m/z shift between the theoretical value of the peptide versus the observed value of the precursor ion. Precursor mass tolerance can be specified in either Da or ppm.

Fragment ion : This enables the selection of a fragment ion m/z tolerance, which is the allowable m/z shift between the theoretical value of the fragment ions of the peptide versus the observed value. PEAKS Online DB Search uses this value when scoring peptide-spectrum matches. PEAKS Online considers that a fragment ion is matched if the calculated m/z is within the tolerance. Fragment mass tolerance can be specified in Da.

Error tolerance parameters need to be set consistently with the mass accuracy of the instrument.

2.3. Enzyme Settings

Enzyme : The Deep Novo Peptidome workflow is designed for both undigested and Trypsin digested samples. The enzyme may be either none or Trypsin.

2.4. PTM Settings

Click "ADD/REMOVE MODIFICATION" button to open the "Select Modifications" dialogue and specify the fixed PTMs and common variable PTMs expected in the sample. It's recommended that no more than 11 variable modifications are selected. If the required modification is not listed, the modification can also be created here. Additionally, the maximum number of variable PTMs per peptide can be specified and a value of less than 4 is recommended. A fixed modification forces all instances of applicable residues to be modified. A variable modification gives the option for the residues to be modified.

Note

For a specific residue, only one fixed modification is allowed.

Note

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

2.5. Peptide Search Parameters

A peptide database must be added first before creating a new project to enable its selection in the workflow. In the Target Database drop-down menu, select the appropriate database that have been configured in PEAKS Online and set the taxonomy, if applicable. A contaminant database can also be selected from the same list of databases that have been configured in PEAKS Online.The length for a peptide can be set when creating an analysis, where the default peptide length range is between 6 and 45

2.6. Report Filter

PSM filter section is a mandatory section to filter out some results under the user-defined filter. Any changes to the filter will create a new analysis with old parameters and update the results and statistics.

- **PSM -10lgP:** The PEAKS peptide score (-10lgP) is calculated for every peptide-spectrum match (PSM) reported by PEAKS DB, PEAKS PTM, and SPIDER. The score is derived from the p-value that indicates the statistical significance of the peptide-spectrum match. A peptide may be matched to many spectra, resulting in multiple PSMs. In that case, the peptide's score is calculated as the maximum 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* . A minimum requirement can be set and all identifications must pass this filter.

- **PSM FDR (%):** The PSM filter can also be set to use the false discovery rate. FDR is estimated using a "decoy-fusion" approach.

  ### Note

  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.

- **Peptide FDR (%):** Peptide filter can also be set to use the false discovery rate. FDR is estimated using a "decoy-fusion" approach.

  ### Note

  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.

- **Confident Amino Acid Threshold (%):** Minimum ion intensity of each fragment ion pair, which define the amino acid position within the peptide, compared to the base peak ion intensity. A default of 2% means that at least one pair of ions (e.g. ‘y’ ions) defining the amino acid has intensity of 2% or higher for both ions within the pair..

- **Deep Novo ALC (%):** This filter controls the minimum ALC score required for display of *deep novo* results. A default of 50% is set.

3. DIA Peptidome Workflow and Parameter Settings

DeepNovo and Peptide Database Search can be launched together as a DeepNovo Peptidome workflow in a PEAKS Online search. An analysis report is generated to show the combined results. DeepNovo peptides are used to assist the following protein database search using PEAKS DB allowing for accurate and sensitive identification of undigested peptides present in the protein sequencing database. As with our streamlined DIA workflow, DIA DB search can be performed library-free or following an optional spectral library search.

Note

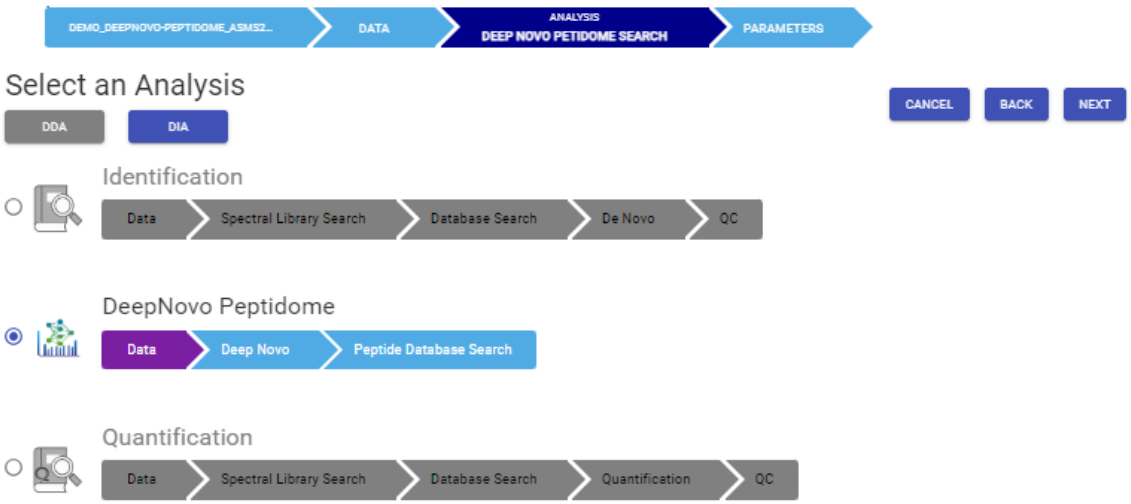
For more details please refer to the paper *"A streamlined platform for analyzing tera-scale DDA and DIA mass spectrometry data enables highly sensitive immunopeptidomics", Nat Commun. 2022 Jun, 13(3108):10.1038/s41467-022-30867-7*.

The combination of the results are then presented in the peptide table where the most confident identification that passes the given filters is given for each scan in the dataset.

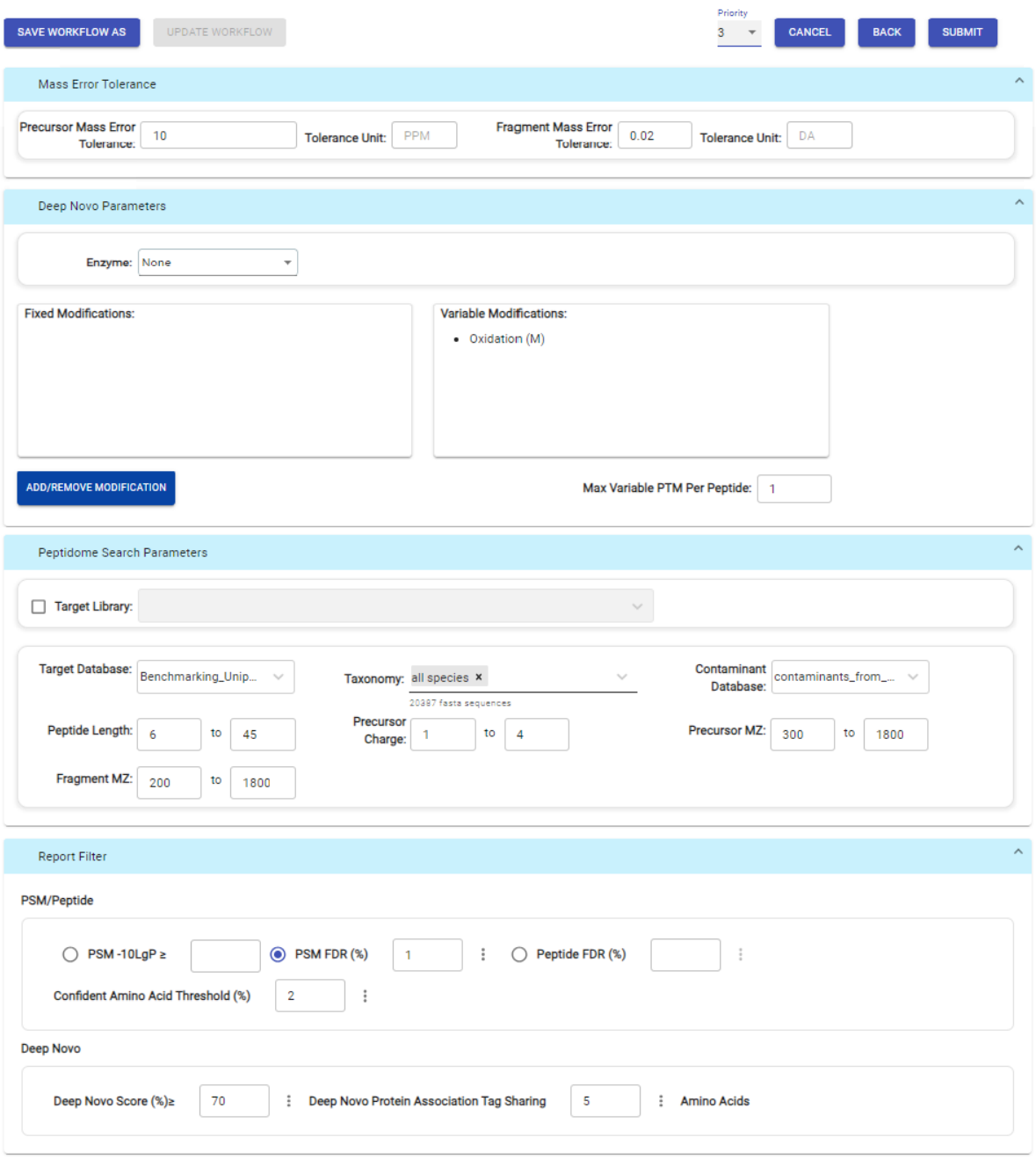
To conduct a database search workflow, follow the steps below:

1. Refer to Chapter 5, Peptide De Novo Sequencing, Section 2, “*DeepNovo/De Novo* Analysis Workflow and Parameter Settings” , Step 1-6 to create a new project and add data.

2. Select the "DeepNovo Peptidome" workflow.



3. Set the DeepNovo and Peptide Search parameters. See more information regarding parameter settings in sections 3.1-3.5 below.



4. Click SUBMIT button to start the search.

3.1. Precursor and Fragment Error Tolerance

Precursor mass : This enables the selection of a precursor mass error tolerance. PEAKS Online DB Search identifies peptides within the precursor mass tolerance, which is the allowable m/z shift between the theoretical value of the peptide versus the observed value of the precursor ion. Precursor mass tolerance can be specified in either Da or ppm.

Fragment ion : This enables the selection of a fragment ion m/z tolerance, which is the allowable m/z shift between the theoretical value of the fragment ions of the peptide versus the observed value. PEAKS Online DB Search uses this value when scoring peptide-spectrum matches. PEAKS Online considers that a fragment ion is matched if the calculated m/z is within the tolerance. Fragment mass tolerance can be specified in Da.

Error tolerance parameters need to be set consistently with the mass accuracy of the instrument.

3.2. Enzyme Settings

Enzyme : The Deep Novo Peptidome workflow is designed for both undigested and Trypsin digested samples. The enzyme may be either none or Trysin.

3.3. PTM Settings

Click "ADD/REMOVE MODIFICATION" button to open the "Select Modifications" dialogue and specify the fixed PTMs and common variable PTMs expected in the sample. It's recommended that no more than 11 variable modifications are selected. If the required modification is not listed, the modification can also be created here. Additionally, the maximum number of variable PTMs per peptide can be specified. For DIA workflows, setting this value greater than 1 will significantly increase search times. A fixed modification forces all instances of applicable residues to be modified. A variable modification gives the option for the residues to be modified.

Note

For a specific residue, only one fixed modification is allowed.

Note

To prevent long search times, select only the most frequent PTMs in the sample for PEAKS DB.

3.4. Peptide Search Parameters

3.4.1. Target Library

Target library can be enabled to enhance identification of expected (targeted) peptides.

Library Selection: Select a PEAKS-generated library for the search. Select one from the list of libraries that have been configured in PEAKS Online and filter for the peptide length, precursor m/z, and charge when necessary. To configure a new library, refer to Chapter 2 Section 2: Library Management and Configuration.

3.4.2. Target Database

A peptide database must be added first before creating a new project to enable its selection in the workflow. In the Target Database drop-down menu, select the appropriate database that have been configured in PEAKS Online and set the taxonomy, if applicable. A contaminant database can also be selected from the same list of databases that have been configured in PEAKS Online.The length for a peptide can be set when creating an analysis, where the default peptide length range is between 6 and 45. Peptidome results can also be filtered by precursor charge, m/z and fragment m/z.

3.5. Report Filter

PSM filter section is a mandatory section to filter out some results under the user-defined filter. Any changes to the filter will create a new analysis with old parameters and update the results and statistics.

- **PSM -10lgP:** The PEAKS peptide score (-10lgP) is calculated for every peptide-spectrum match (PSM) reported by PEAKS DB, PEAKS PTM, and SPIDER. The score is derived from the p-value that indicates the statistical significance of the peptide-spectrum match. A peptide may be matched to many spectra, resulting in multiple PSMs. In that case, the peptide's score is calculated as the maximum 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* . A minimum requirement can be set and all identifications must pass this filter.

- **PSM FDR (%):** The PSM filter can also be set to use the false discovery rate. FDR is estimated using a "decoy-fusion" approach.

Note

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.

- **Peptide FDR (%):** Peptide filter can also be set to use the false discovery rate. FDR is estimated using a "decoy-fusion" approach.

Note

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.

- **Confident Amino Acid Threshold (%):** Minimum ion intensity of each fragment ion pair, which define the amino acid position within the peptide, compared to the base peak ion intensity. A default of 2% means that at least one pair of ions (e.g. ‘y’ ions) defining the amino acid has intensity of 2% or higher for both ions within the pair.
- **Deep Novo ALC (%):** This filter controls the minimum ALC score required for display of *deep novo* results. A default of 50% is set.

4. Understanding Peptidome Results

After a PEAKS Search is complete, all analyses selected during the project creation will be combined under the workflow drop-down menu. One is the *deep novo* sequencing result, which includes *de novo* peptide sequences listed, spectrum annotation and other detailed information. The other result node contains DeepNovo, PEAKS DB, and optional spectral library search results.

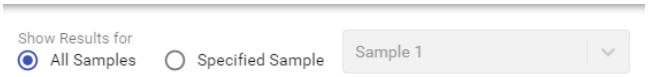
The analysis results for database search tools are presented in two tabs:

- **Summary :** This shows an outline of the PEAKS Online peptidome search results with key statistics and the overall quality of the experiment can be examined. This page shows after the search is done.
- **Peptides :** This shows a list of peptide identifications. This view also provides spectrum annotation and other detailed information for peptide precursor spectrum matches.

4.1. Result Filtration

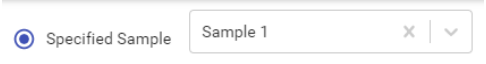
4.1.1. Show Results for All Samples

For the peptide tables, "All Samples" can be selected to view the combined search results from all the samples.



4.1.2. Show Results for Specific Sample

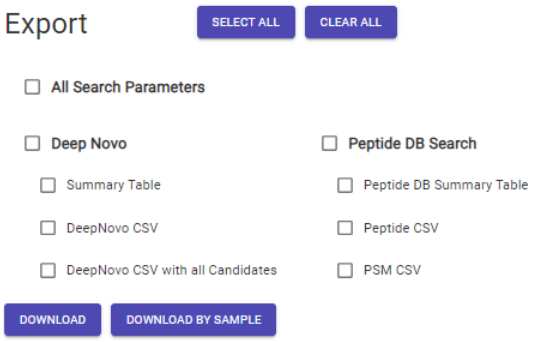
In the Peptides and DeepNovo page results, the filter can also be set to "Specified Sample" which will only show the results for a single sample. By clicking in the text you can search for a specific sample by name.



Peptides will only be visible if they belong to the sample chosen. Once selected, the peptide tables will automatically add the "Area" column and the "# Spec" values will display the spectral count for the specified sample. For more information, refer to Section 3.5.2, “Peptide Table” .

4.2. Result Exporting

PEAKS Online peptidome search results can be exported in the Export page:



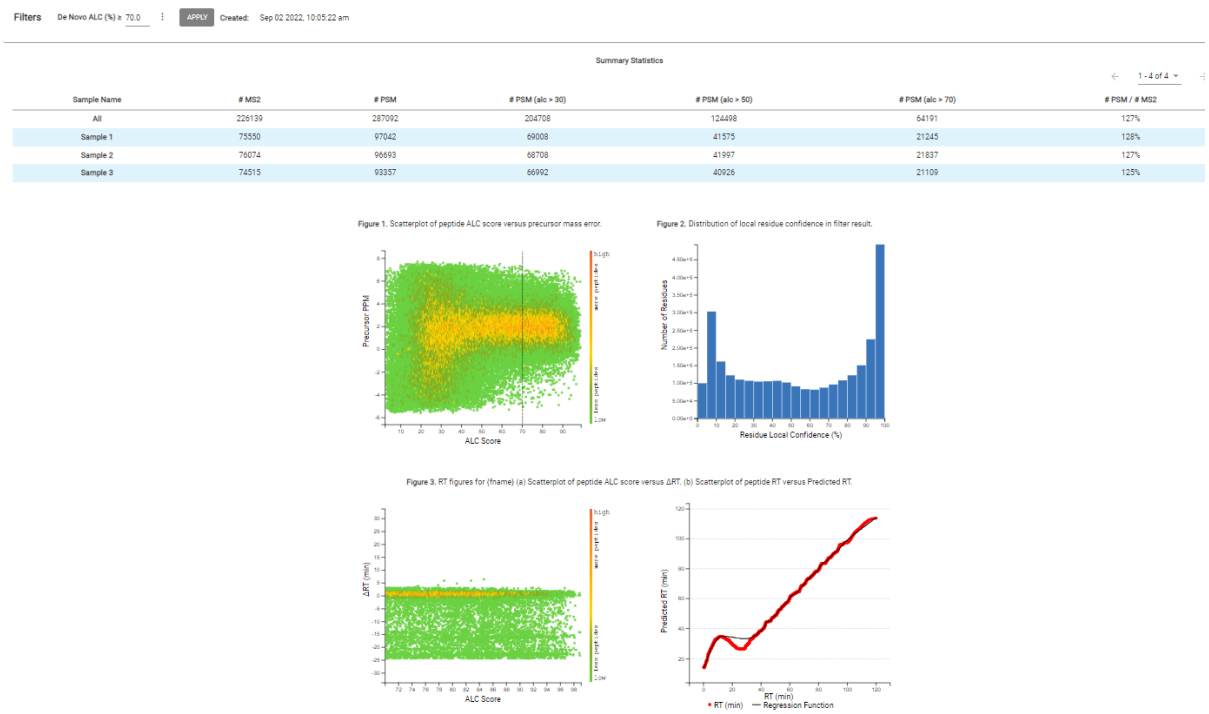
- **All search parameters:** A comma separated file containing all used search parameters.
- **DeepNovo Summary Table**: A comma separated file containing a summary of the number of peptides *de novo* sequenced in each sample.
- **DeepNovo CSV:** A comma separated file containing the details from the DeepNovo table
- **DeepNovo CSV with all Candidates:** A comma separated file containing all details from the DeepNovo table including the top 5 *de novo* candidates from each spectrum within the ALC threshold set in the search parameters.
- **Peptide CSV :** A comma separated file containing all identified peptides and associated details from the peptide table for all samples.
- **PSM CSV :** A comma separated file containing all identified peptide spectrum matches and associated details from the peptide table for all samples.

4.3. DeepNovo Summary View

The DeepNovo summary view reports the key statistics from DeepNovo peptide sequencing

4.3.1. Result Statistics

Details about the filtered result:



Filters De Novo ALC (%) ≥ 70.0 APPLY Created: Sep 02 2022, 10:05:22 am

Summary Statistics

1 - 4 of 4

Sample Name	# MS2	# PSM	# PSM (alc > 30)	# PSM (alc > 50)	# PSM (alc > 70)	# PSM / # MS2
All	226139	287092	204708	124498	64191	127%
Sample 1	75550	97042	69008	41575	21245	128%
Sample 2	76074	96693	68708	41997	21837	127%
Sample 3	74515	93357	66992	40926	21109	125%

Figure 1. Scatterplot of peptide ALC score versus precursor mass error.

Figure 2. Distribution of local residue confidence in filter result.

Figure 3. RT figures for (fname) (a) Scatterplot of peptide ALC score versus ΔRT. (b) Scatterplot of peptide RT versus Predicted RT.

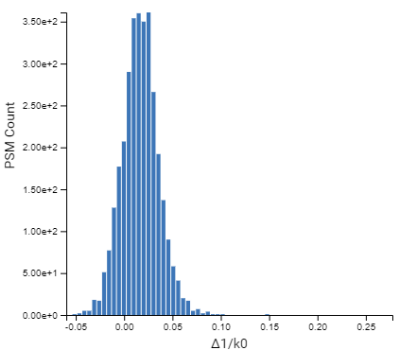
- **Sample Name**: The name of the sample given during the project setup
- **#MS2:** Number of MS2 scans in the sample
- **#PSM:** The 'precursor spectrum match' is the number of peptide features with matching DeepNovo results that pass the filters set in the search parameters
- **#PSM(alc > 30):** The is the number of peptide features with matching DeepNovo results above an average local confidence of 30
- **#PSM(alc > 50)**: The is the number of peptide features with matching DeepNovo results above an average local confidence of 50
- **#PSM(alc > 70)**: The is the number of peptide features with matching DeepNovo results above an average local confidence of 70
- **#PSM/#MS2**: The number of PSMs with DeepNovo results above the ALC filter set in the search parameters divided by the number of MS2 scans in the dataset. (DDA only).

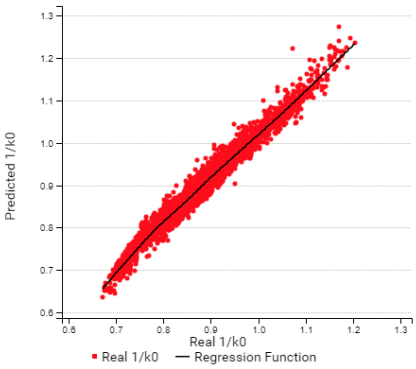
Table 1 provides the total number of MS2 scans and PSMs total as well as the number of MS2 scans and PSMs with different ALC thresholds for each sample. For DDA acquisition, the #PSM / #MS2 provides a de novo sequencing identification rate metric for each sample.

Figures 1 - 3 are used to visualize the distribution of the DeepNovo results

- Figure 1: Average local confidence score versus precursor PPM: The distribution of DeepNovo result average local confidence versus the precursor mass error in parts per million. A color gradient from green (low density) to orange (high density) is used to represent the density of results at each position on the dot plot
- Figure 2: Residue Local confidence histogram: Using the percentage local confidence assigned to each amino acid, a histogram is reported with the number of amino acids with a local confidence at the given local confidence score.
- Figure 3a: Average local confidence score versus the delta retention time: Compares the average local confidence to the difference between the predicted retention time of a peptide versus the actual retention time in minutes. This figure validates the quality of the retention time prediction

- Figure 3b: Scatterplot of retention time versus the predicted retention time for each DeepNovo result: Each point represents the retention time from the data file versus the predicted retention time. The regression function used to predict retention time is plotted against the data.
- Figure 4a: Histogram of peptide $_{\Delta 1}$/k0 for timsTOF data.
- Figure 4b: Regression plot of peptide 1/k0 versus predicted 1/k0 for timsTOF data.





4.4. DeepNovo Sequence Table

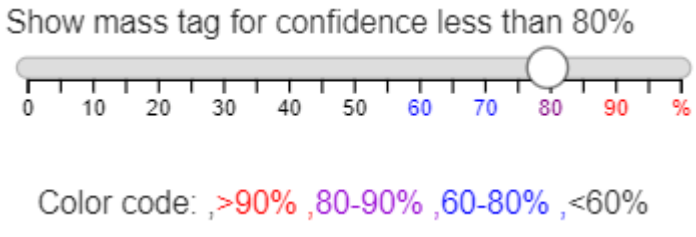
4.4.1. Peptide View Filters

Click the button, the following De Novo Filter will pop up. It controls which peptides to display.

- **Denovo sequence contains:** Enter a string of amino acids. The filter will return any result containing the string. Modification masses can be included as their mass within circle brackets, i.e (+15.99)
- **Scan:** Filter for a single scan number based on the scan numbers read from the raw data file.
- **Precursor Id:** Filter is only relevant for TimsTOF data, for which it will filter based on the precursor id of the peptide.
- **m/z:** Filter based on the expected m/z, the last significant digit is rounded.
- **RT:** Filter based on the expected RT, the last significant digit is rounded.
- **|#RT| <=:** Only results with an absolute difference in retention time between the predicted retention time and the observed retention time greater than or equal to the given value will be displayed.
- **Tag Length >= :** Only results with a tag length greater than the given value will be displayed.
- **Tag Length <= :** Only results with a tag length less than the given value will be displayed.

4.4.1.1. Local Confidence Slider

The local confidence slide accessed by clicking the  button to bring up the local confidence slider.



The slider is used to set the minimum local confidence for all amino acids. Amino acids with a local confidence score below the threshold will be reported as their mass in daltons. If multiple amino acids are found in a row below the cut-off the mass of each amino acid will be added to create a mass tag. This slider also controls the de novo tag that appears in the tag column of the export.

4.4.2. DeepNovo Table

The DeepNovo table shows the filtered DeepNovo results. Each row in the table is a precursor identification represented by its highest-scoring *de novo* sequencing result. The table is sorted by the ALC value.

The columns Scan, peptide, ALC, m/z, z, RT, and Delta RT, are sortable, i.e. Clicking any of these header can sort the whole table based on that column.

The columns in the Peptide table are:

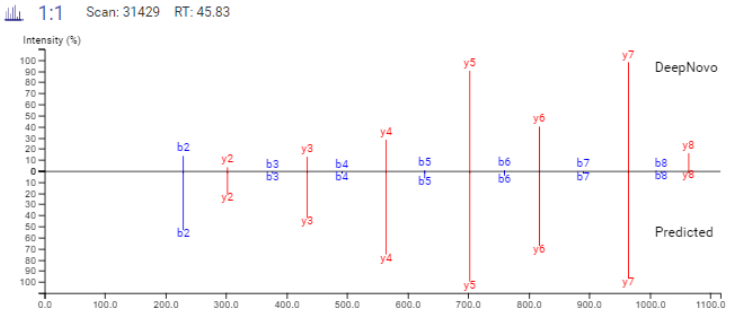
- **Scan:** The scan number, followed by the precursor ID (for timsTOF data), from the raw file followed by the raw file name
- **Peptide:** The predicted *de novo*sequence color coded by the local confidence score of each amino acid. The color legend can be found by clicking the local confidence slider button 
- **Score (%)**: A peptide confidence score which adds penalty to ALC (%) based on the the difference between predicted RT and observed RT. Typically Score(%) will be smaller than ALC(%) for the same peptide.
- **ALC(%):** The average local confidence of the peptide. It is the average of the individual local confidence scores. Local confidence scores are calculated by permutation of the possible peptide sequences that can match the precursor mass and comparing the sequence alignments. The local confidence represents the likelihood that the amino acid reported at the position is correct compared to the other possible matches.
- **Length:** The number of amino acids in the peptide sequence.
- **m/z:** The precursor mass-to-charge ratio.
- **z:** The charge of the peptide retrieved from the raw file
- **RT:** The retention time (elution time) of the spectrum as recorded in the data.
- **Delta RT:** The difference between the observed retention time and the retention time predicted during analysis.
- **Area:** The area of the peptide feature associated with the *de novo* result. If no area could be associated from feature detection, an empty cell is give.
- **Mass:** The theoretical mass of the predicted *de novo* sequencing result
- **ppm:** The precursor mass error, calculated as $10^{6} \times$ (precursor mass - peptide mass) / peptide mass.
- **PTM:** The types and the numbers of modifications present in the peptide shown in color-coded icons.
- **Mode:** The fragmentation mode of the scan associated with the *de novo* result

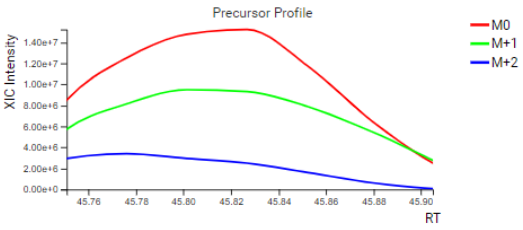
4.4.3. Annotated Spectrum

The annotated spectrum view by default will give a mirror plot comparing the DeepNovo result to the spectrum predicted by the deep learning model. To the right of the mirror plot, the precursor profile displays the extracted ion chromatogram of the MS1 signal associated with the M0, M+1, and M+2 ions of the intact peptide. Below the mirror plot, the ion details are given:

Fragment ion details table:

- **Label:** The type of predicted fragment ion
- **Predicted M/Z:** The theoretical mass to charge ratio of the amino acid predicted at the position
- **Predicted Intensity (%):** The predicted relative intensity of the fragment ion from the deep learning model
- **Query M/Z:** The observed mass to charge ratio of the fragment ion
- **Query Intensity:** The observed intensity of the fragment ion





#	Label	Predicted M/Z	Predicted Intensity(%)	Query M/Z	Query Intensity
1	b2	229.1188	53.4	229.1186	3.38e+5(16.3%)
2	y2	303.2139	21.2	303.2141	1.25e+5(6.0%)
3	b3	376.1872	2.4	376.1866	1.14e+4(0.5%)
4	y3	434.2544	41.9	434.2549	3.14e+5(15.1%)
5	b4	491.2142	2.5	491.2162	2.09e+4(1.0%)
6	y4	565.2949	75.3	565.2957	6.26e+5(30.1%)
7	b5	628.2731	6.5	628.2731	6.33e+4(3.0%)
8	y5	702.3538	100.0	702.3546	1.92e+6(92.6%)
9	b6	759.3136	4.6	759.3136	6.93e+4(3.3%)
10	y6	817.3807	68.1	817.3812	8.76e+5(42.2%)

Click the  button to switch to the annotated spectrum view. The annotated spectrum view is described in the Section 3.5, “Spectrum Annotation” section.

4.5. Peptide Search Summary View

The Summary view reports key statistics as an overview of the result. In the report, several statistical charts can be examined to assess result quality.

Summary Statistics

Search... CLEAR 1 - 5 of 5

Sample Name	# MS Run	# MS1	# Precursor	#PSM	#Peptides	# Sequences	# PSM / # Precursor
All (4 samples)	4	12454	153480	48193	13522	13225	31%
Sample 2	1	3069	41760	13151	6170	6057	31%
Sample 3	1	3035	46957	14865	6556	6534	32%
Sample 4	1	3206	42955	14001	6669	6493	33%
Sample 5	1	3144	21808	6176	1587	1582	28%

Figure 1.False discovery rate (FDR) curve.X axis is the number of peptides being kept. Y axis is the corresponding FDR.

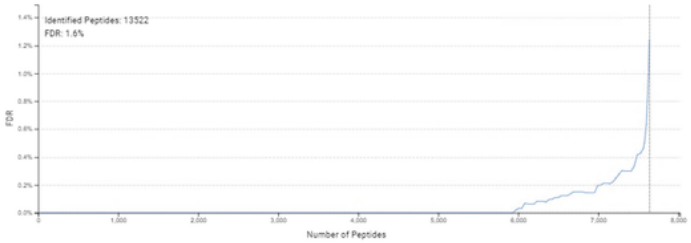


Figure 2: Distribution of precursor mass error of filtered PSM in ppm.

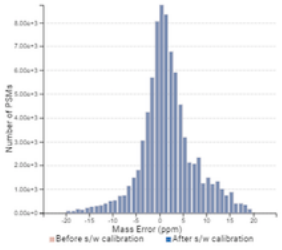
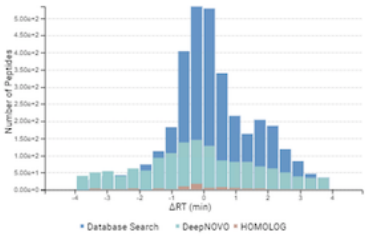


Table 1. Statistics of filtered results.

DB Peptides	7384
Homolog Peptides	153
DeepNovo Peptides	5984
FDR (Peptide Spectrum Matches)	0.6%
FDR (Peptide Sequences)	1.6%
-10LgP Cutoff (Peptides)	3.75

Figure 3. RT figures for Sara_Ligandome_murine_CT26_120320_RQ2_10of10ul_Slot2-20_20-04-19_1169.d with outliers removed (a) Histogram of peptide ΔRT. (b) Scatterplot of peptide RT versus Predicted RT.



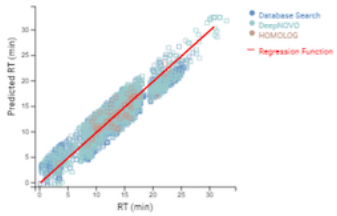
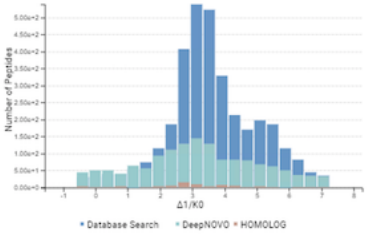


Figure 4. 1/K0 figures for Sara_Ligandome_murine_CT26_120320_RQ2_10of10ul_Slot2-20_20-04-19_1169.d with outliers and charge 1 peptides removed (a) Histogram of peptide Δ1/K0. (b) Scatterplot of peptide Real 1/K0 versus Predicted 1/K0.



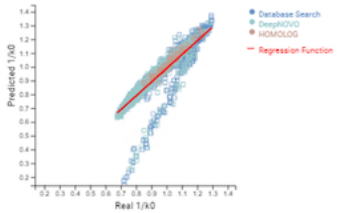
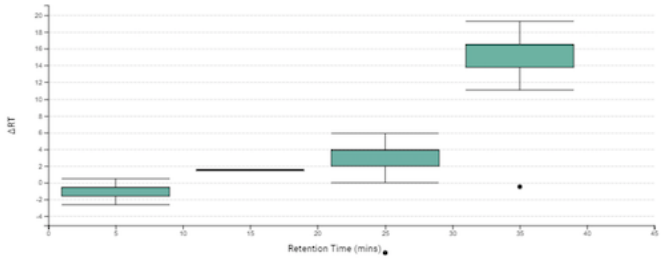


Figure 5. Boxplot for median ΔRT across all MS runs.



4.5.1. Result Statistics

The first chart displays the summary information for each sample.

- **Sample Name:** This column lists the samples names with the exception of the first row, which shows the combination of all the samples.
- **#MS Run:** The number of data files included in the sample
- **#MS1:** The number of MS scans in the sample
- **# MS2:** The total MS/MS scans in the sample summed across all data files.
- **# Precursor**: Total number of precursors (DIA only).
- **# PSM:** Total number of precursor spectrum matches (PSMs) found in the sample.
- **# Peptides:** Total number of peptides found ignoring differences with leucine (L) and isoleucine (I) amino acids. Peptides with the same primary sequence but different PTMs are counted separately.
- **# Sequences:** Total number of different peptide backbone sequences found ignoring differences with leucine (L) and isoleucine (I) amino acids. Peptides with the same primary sequence but different PTMs are counted as one sequence.
- **#PSM/#MS2:** The precursor spectrum matches identified divided by the total number of MS2 in the sample

4.5.2. Summary View Figures

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

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

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

Note

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

Note

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

Figure 2: Distribution of precursor mass error of filtered PSM in ppm: Shows a histogram of mass error of identifications bundled by the number of precursor spectrum matches in each group.

Table 1. Statistics of filtered results.

- **DB Peptides**: The number of peptides matched by PEAKS DB database search
- **Homolog Peptides:** The number of peptides matched by SPIDER homology search **(DDA Only)**
- **Targeted Peptides:** The number of peptides matched using the optional target library search **(DIA Only)**
- **DeepNovo Peptides:** The number of peptides matches by DeepNovo *de novo* sequencing

- **FDR (Peptide-spectrum matches):** The false discovery rate of the full list of target hits from the final round search versus hits from the decoy database at the precursor spectrum match level.
- **FDR (Peptide sequences):** The false discovery rate of the full list of target hits from the final round search versus hits from the decoy database at the peptide level.
- **-10LgP Cutoff (Peptides):** The score where the FDR cut-off is reached selected by the search parameters.

Figure 3a. Histogram of peptide delta RT: The difference in RT between the observed RT versus the predicted RT bundled by the number of peptides in each range for each individual algorithm that detected them eg. DeepNovo, DB, SPIDER (DDA), Target Library (DIA)

Figure 3b. Scatterplot of peptide RT versus Predicted RT: Each point represents a predicted peptide sequence. The distribution of predicted RT versus observed RT is shown for peptides for each individual algorithm that detected them eg. DeepNovo, DB, SPIDER (DDA), Target Library (DIA). The predicted regression function is also given.

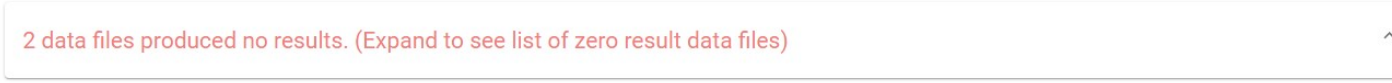
Figure 4a. Histogram of peptide $_{\Delta 1/}$k0 for timsTOF data.

Figure 4b. Regression plot of peptide 1/k0 versus predicted 1/k0 for timsTOF data.

Figure 5. Boxplot of median delta RT across all MS runs: The retention time plotted against delta RT showing the trend of different in predicted RT versus observed RT over time. This chart highlights times across all runs where RT prediction may be different from the average.

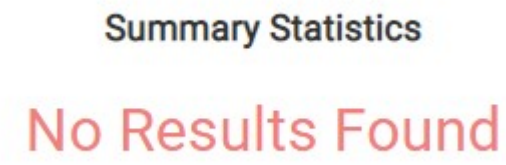
4.5.3. Low Quality Data Warnings

If PEAKS Online detects any data files producing very few or no results a warning section will appear at the top of the Summary Page indicating the number of data files producing poor or no results.



This section can be expanded to display a list of data files in the analysis that are producing poor or no results.

If there were no results found in any of your data files, a warning message stating "No Results Found" will be displayed instead of the usual Summary Page figures and tables.

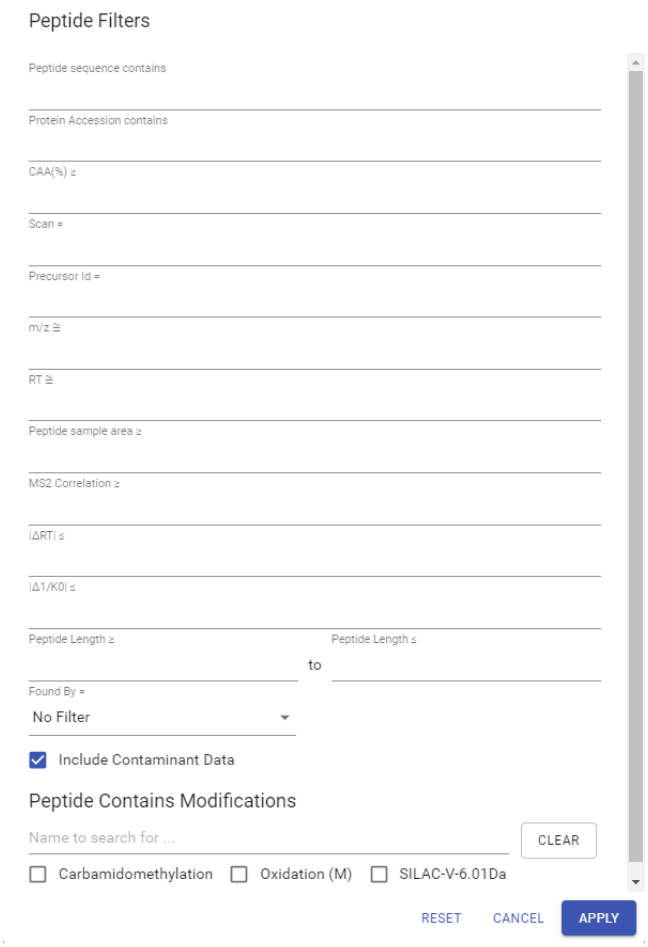


4.6. Peptide View

The Peptide Table shows the filtered peptide identification results. Click on a peptide row to show the spectrum below the peptide table.

4.6.1. Peptide View Filters

Click the button, the following Peptide Filter will pop up. It controls which peptides to display.



- **Peptide sequence contains:** Enter an amino acid sequence, only peptides that contain that sequence will be displayed in the peptide table.
- **Peptide Accession contains:** Enter an keyword, only peptides that are associated with protein accessions containing that keyword will be displayed in the table.
- **CAA (%):** Enter a percentage. This represents the percentage of residues which pass the CAA threshold set by the user in the project parameters. Only peptides with a percentage of passing residues greater than this will be displayed in the table.
- **Scan:** Filter for a single scan number based on the scan numbers read from the raw data file.

  ## Note

  Scan filter is only available for DDA samples.

- **Precursor Id:** Filter is only relevant for TimsTOF data, for which it will filter based on the precursor id of the peptide.
- **m/z:** Filter based on the expected m/z, the last significant digit is rounded.
- **RT:** Filter based on the expected RT, the last significant digit is rounded.
- **Peptide sample area:** Filters based on total peptide area found in each sample. If a peptide is found in multiple samples, there should be at least one sample pass this filter for the peptide to be included.
- **MS2 Correlation:** Enter a percentage.Only peptides with MS2 Correlation greater than this will be displayed in the table.

- **|#RT| <=:** Only results with an absolute difference in retention time between the predicted retention time and the observed retention time less than the given value will be displayed.
- **Peptide Length >=:** Only results with a difference in retention time between the predicted retention time and the observed retention time greater than the given value will be displayed.
- **Peptide Length <=:** Only results with a difference in retention time between the predicted retention time and the observed retention time less than the given value will be displayed.
- **Found By:** The algorithm used to find the peptide sequences
- **Include Contaminant Data:** Check this to display contaminants that were identified during the search (if a contaminant database was specified). Leaving this unchecked will only display peptides with at least one non-contaminant accession.
- **Peptide Contains Modifications:** Check boxes are provided for each identified modification. Peptides that contain the selected PTMs will be shown.

4.6.2. 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 is sorted by the -10lgP value. In the Protein list drop-down window of each peptide, after clicking any of the protein, the protein page will be opened.

The columns Peptides, -10LgP, m/z, RT, Accession are sortable, i.e. Clicking any of these header can sort the whole table based on that column.

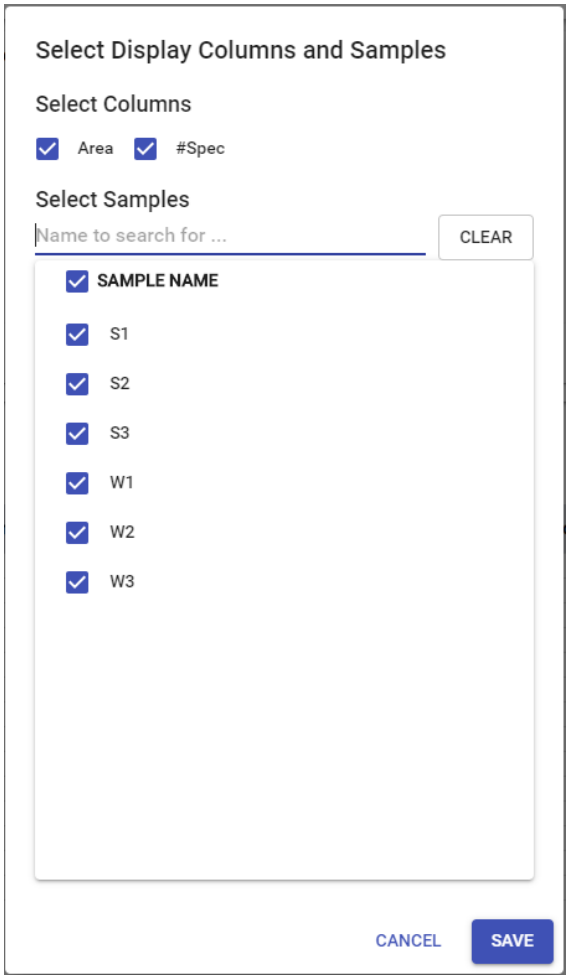
The columns in the Peptide table are:

- **Peptide:** The amino acid sequence of the peptide, as determined in PEAKS Search. A modified residue is followed by a pair of parentheses enclosing the modification mass.
- **-10lgP:** The peptide -10lgP score. The score indicates the scoring significance of a peptide-spectrum match.
- **CAA (%):** The percentage (%) of residues which have a CAA score passing the CAA threshold set by the user in the peptide view filters.
- **Mass:** The monoisotopic mass of the peptide.
- **Length:** The number of amino acids in the peptide sequence.
- **ppm:** The precursor mass error, calculated as $10^6 \times$ (precursor mass - peptide mass) / peptide mass.
- **m/z:** The precursor mass-to-charge ratio.
- **RT:** The retention time (elution time) of the spectrum as recorded in the data.
- **Delta RT:** The difference in retention time between the observed retention time and the predicted retention time.
- **MS2 Correlation:** Correlation between the predicted and query MS2 spectra (including ion match and intensity).
- **Scan:** The scan number of the spectrum that matches the peptide sequence with the highest -10lgP.
- **#Spec:** The number of MS2 spectra assigned to the 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, Ascore (DDA Only).

- **Ion Intensity:** Ion intensity of the modification site-determining fragment ion pair, compared to the base peak ion intensity, for the top-scoring PSM. Used as a measure of PTM site confidence (DIA Only).
- **Found By:** The algorithm the peptide was found by. For DDA the options are PEAKS DB, DeepNovo, or Spider (shown as homolog). For DIA the options are DeepNovo, Targeted Peptide (if optional spectral library search is enabled), or DB Search.
- **Accession:** The accession number of the highest-scoring protein containing this peptide.
- **1/k0:** This column is only available for TimsTof data, it shows the ion mobility range of for the precursor.

4.6.3. Peptide Global Comparative View

"Peptide Global Comparative View" provides detailed comparisons among different samples regarding the identified peptides. To Show/Hide Columns, click on top-left of the peptide table. From the pop-up window, users can add and remove area and spectral count columns. In the peptide table, the selected columns will be displayed and can be exported. If "Specified Sample" is selected these columns are automatically displayed.



Note

PEAKS Online 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. The default can be changed by the administrator.

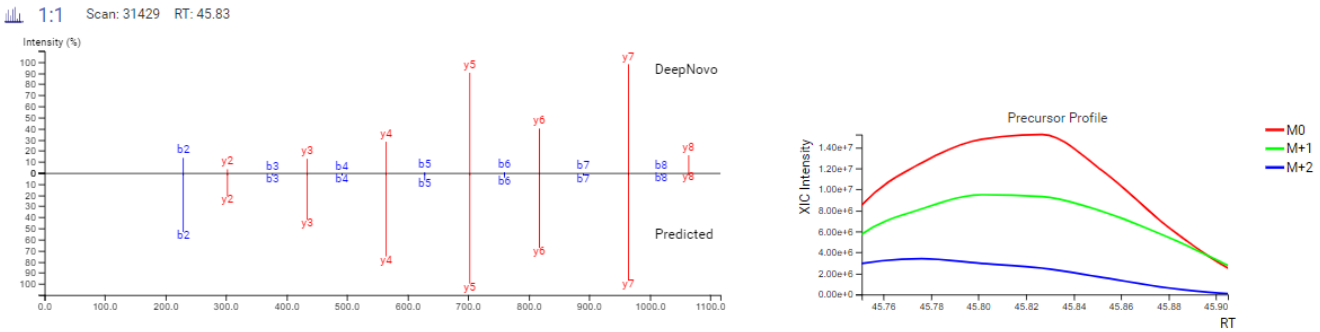
- **Area by sample:** The total area of all peptide features matched to spectra that identified the peptide. Area calculations are separated into multiple columns if multiple samples were included in the run.
- **#Peptide-spectrum matches by sample:** The number of peptide spectrum matches associated with a peptide separated by sample.

4.6.4. Annotated Spectrum

The annotated spectrum view by default will give a mirror plot comparing the result to the spectrum predicted by the deep learning model. To the right of the mirror plot, the precursor profile displays the extracted ion chromatogram of the MS1 signal associated with the M0, M+1, and M+2 ions of the intact peptide. Below the mirror plot, the ion details are given:

Fragment ion details table:

- **Label:** The type of predicted fragment ion
- **Predicted M/Z:** The theoretical mass to charge ratio of the amino acid predicted at the position
- **Predicted Intensity (%):** The predicted relative intensity of the fragment ion from the deep learning model
- **Query M/Z:** The observed mass to charge ratio of the fragment ion
- **Query Intensity:** The observed intensity of the fragment ion



#	Label	Predicted M/Z	Predicted Intensity(%)	Query M/Z	Query Intensity
1	b2	229.1188	53.4	229.1186	3.38e+5(16.3%)
2	y2	303.2139	21.2	303.2141	1.25e+5(6.0%)
3	b3	376.1872	2.4	376.1866	1.14e+4(0.5%)
4	y3	434.2544	41.9	434.2549	3.14e+5(15.1%)
5	b4	491.2142	2.5	491.2162	2.09e+4(1.0%)
6	y4	565.2949	75.3	565.2957	6.26e+5(30.1%)
7	b5	628.2731	6.5	628.2731	6.33e+4(3.0%)
8	y5	702.3538	100.0	702.3546	1.92e+6(92.6%)
9	b6	759.3136	4.6	759.3136	6.93e+4(3.3%)
10	y6	817.3807	68.1	817.3812	8.76e+5(42.2%)

Click the  button to switch to the annotated spectrum view. The annotated spectrum view is described in the Section 3.5, “Spectrum Annotation” section.

Chapter 14. Instrument Daemon

The PEAKS Online Instrument Daemon is a separate process included with your PEAKS Online server. It can be installed on an Instrument or Data machine to allow for an automatic upload and the analysis of raw data as the mass spectrometric acquisitions are complete.

This section of the manual covers the use of an Instrument Daemon that has already been installed and registered with your PEAKS Online server. For help with installing or configuring an Instrument Daemon please refer to the Instrument Daemon section of the Admin manual.

1. What is an Instrument Daemon?

An Instrument Daemon allows for real-time processing of data as it comes off the instrument. Instrument Daemons are separate processes run alongside PEAKS Online, typically on the data acquisition machine. Once installed, the Instrument Daemon will be pointed towards the directory on the machine that new data is written to. The Instrument Daemon will then scan the directory periodically for new data.

From the PEAKS Online Web Interface, projects can be scheduled using these Instrument Daemons. When the Instrument Data scans its directory and locates new data that fits the requirements for a scheduled project, it will automatically upload the data and submit it for processing.

Notes on Instrument Daemons:

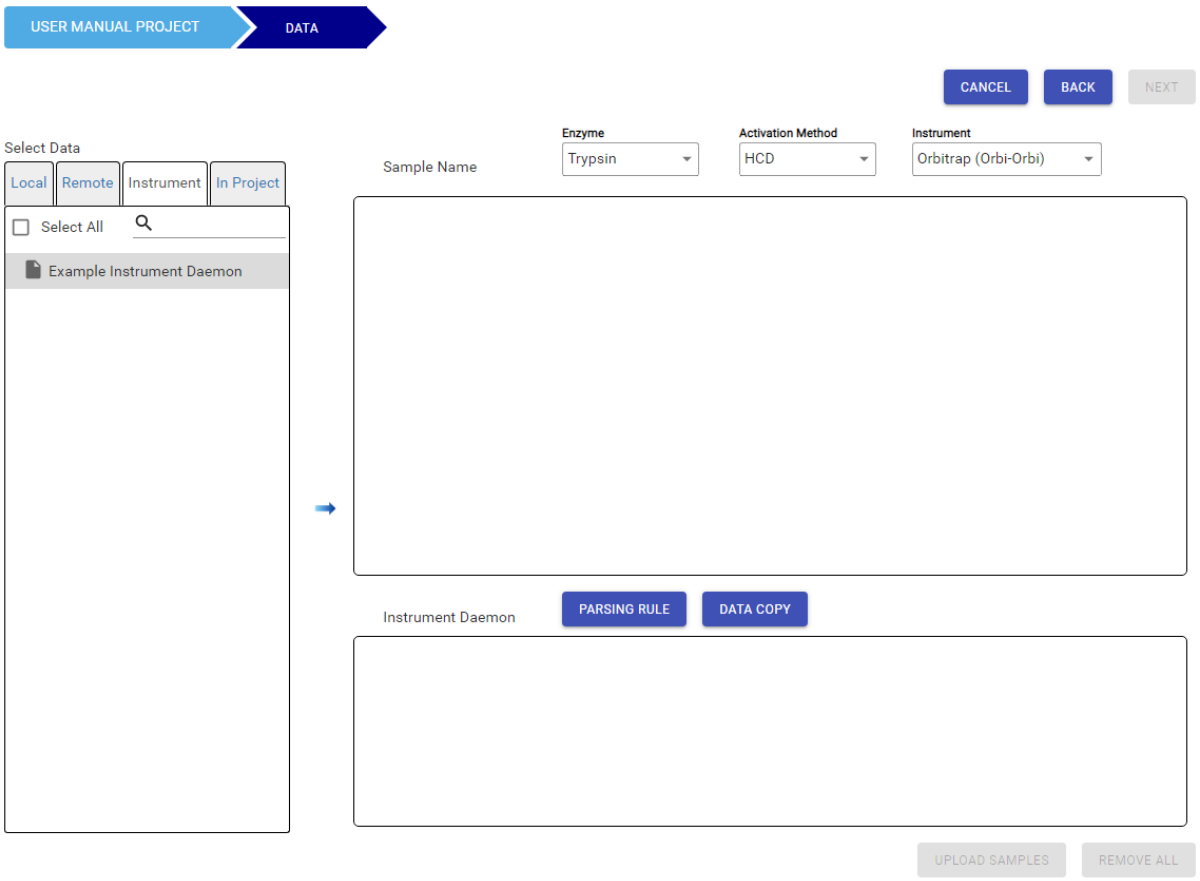
- Your PEAKS Online server can have any number of Instrument Daemons connected to it.
- A single project can request data from multiple Instrument Daemons at the same time.
- Data can be uploaded using an Instrument Daemon either pro-actively (before your experiment has run through the Mass Spectrometer) or retro-actively (after your experiment has been run through the Mass Spectrometer).

2. How to add Data using an Instrument Daemon

Adding data to a project, either a new project or an existing project, is very similar to adding data either from a Local or Remote data source. To review how to add data to a project refer to Chapter 1 Section 5.2 - Creating a PEAKS Online Project.

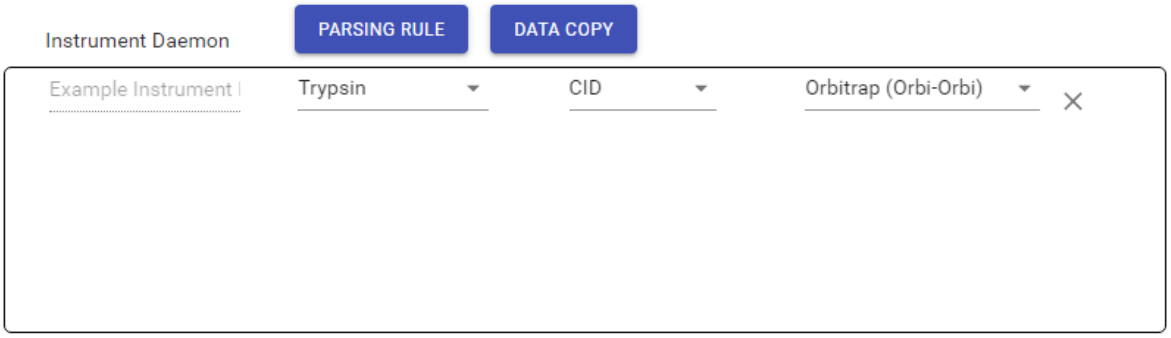
2.1. Data Page

To add data from an Instrument Daemon, instead of using existing data either Locally or from a Remote Data Repository, follow these steps:



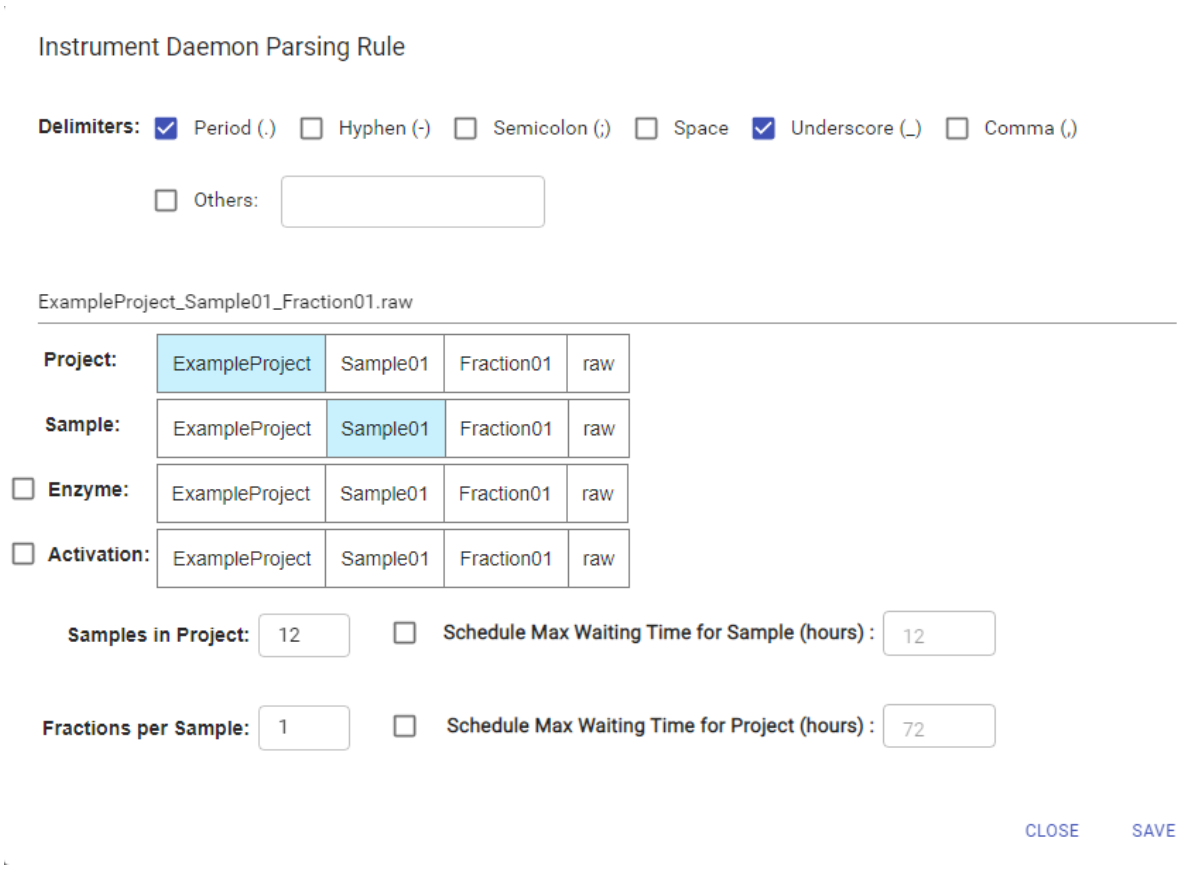
1. On the Data Page, select the "Instrument" tab instead of the "Local" or "Remote" tabs.

2. Select one, or more, Instrument Daemons from the list and click the  button to move them to the right side of the page. This will add them to the "Instrument Daemon" section at the bottom of the page, instead of the "Samples" section above.



3. Once added, you will need to specify the criteria for data on the selected Instrument Daemons to be added to this project. To do this, click on the 

button above the list of added Instrument Daemons. This will pop up a new dialog box.



4. In this new dialog you will need to complete the following steps:

   i. Type in, or copy and paste, an example filename that follows the format of files from the experiment you wish to analyze.

      ## Note

      It is best if this is an actual filename from your experiment (either from a file that has already been generated or an example of one that will be generated).

   ii. Select the delimiters, at the top of the dialog, that should be used to split your example filename. This will be used to cut your filename into sections, and these sections will then be used to determine if data should be added to this project or not as well as how to group files into samples within the project. For the example above, period (.) and underscore (_) were selected as delimiters, which breaks the example filename into 4 sections, "ExampleProject", "Sample01", "Fraction01", and "raw".

   iii. Beneath your example file name, select the sections of the filename that should be used to do the following grouping:

      a. Project. This should be a portion of the filename that is common across every file that will be generated for this project, but that is not used for any other files on that Instrument Daemon. In this example, the first section, "ExampleProject" was selected. This means any data on the selected Instrument Daemon(s) that starts with the words "ExampleProject" will be added to this project. This would include existing data on the Instrument Daemon.

      b. Sample. This should be a portion of the filename that is common across all fractions of a single sample. In this example, the second section, "Sample01", was selected. This means any data that shares the same text in this section will be grouped together in a sample. So all files with "Sample01" will be grouped as "Sample01" and all files with "Sample02" will be grouped as "Sample02".

      c. Enzyme. This should be a portion of the filename that relates to the Enzyme used for that Mass Spec run.

Important

This must match an Enzyme that exists on your PEAKS Online server exactly in order to work. For example, Trypsin.

d. Activation. This should be a portion of the filename that relates to the Activation Method used for that Mass Spec run.

Important

This must match one of the built-in Activation Methods on PEAKS Online exactly in order to work. For example, HCD.

e. Samples in Project. Here you should specify the number of samples you expect to run, or have already run, for this experiment.

f. Fractions per Sample. Here you should specify the number of fractions that you are expecting for each sample you are running as part of this experiment.

Important

All samples added to a project using Instrument Daemons must have the same number of fractions.

g. Schedule Max Waiting Time for Sample. This is an optional parameter that will set a max wait time for each sample. Once a sample has received its first fraction (in the case of fractionated samples) if it has not received all of its fraction files within the specified time period, that sample will be considered failed. *This is to prevent files that may have run into issues on the instrument from causing projects to hang indefinitely.*

h. Schedule Max Waiting Time for Project. This is an optional parameter that will set a max wait time for the entire project. After the specified time has elapsed, any samples that have not received all of their fractions will be considered failed.

iv. Once this has all been set, click the SAVE button to save the parsing rule.

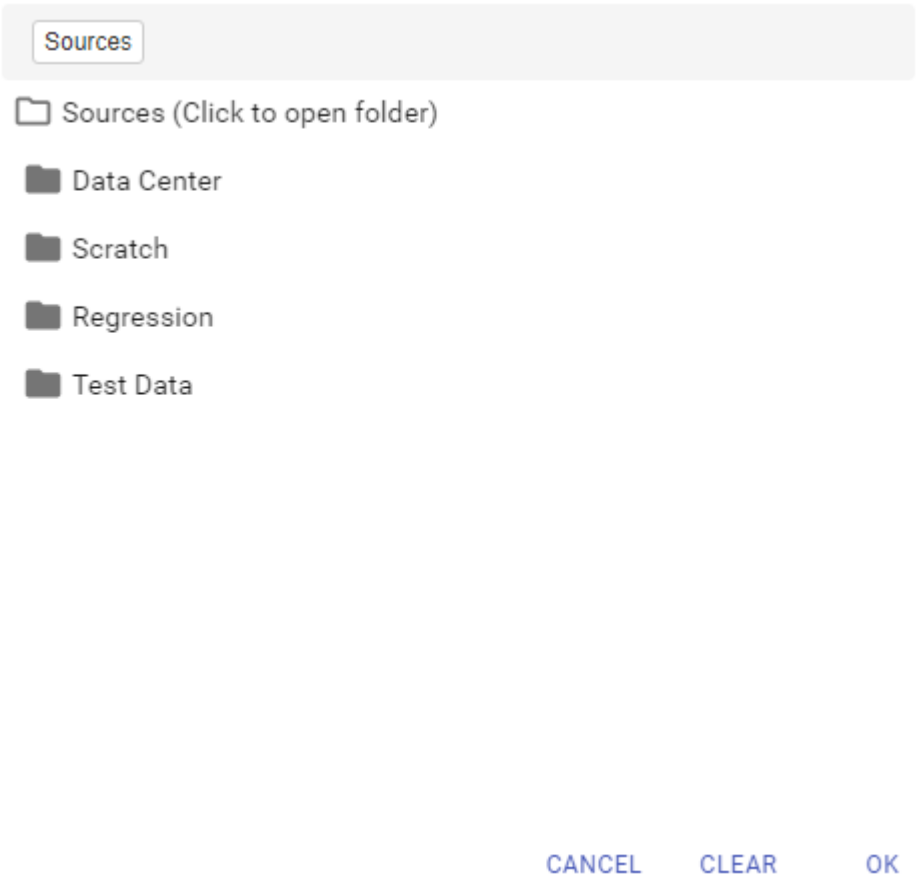
5. Once the parsing rule has been set, you can either click the NEXT button to continue to the analysis parameter settings or the UPLOAD SAMPLES button to load the data without performing an analysis.

2.2. Data Copy

And additional function that the Instrument Daemon has is the ability to copy data into a Data Repository. This is an optional step that can be requested when adding new data to a new or existing project using an Instrument Daemon.

In order to utilize this function, follow these steps:

1. After the Instrument Daemon has been added to the project on the Data Page, click the DATA COPY button at the top of the Instrument Daemon section. This will pop up a new dialog.



2. This new dialog will contain a list of all Remote Data Repositories attached to your PEAKS Online server, similar to selecting the "Remote" tab for data upload.

3. Select the Data Repository you wish to copy the data to and navigate to the location within that repository where you wish to save the data.

4. Click the OK button.

After a location has been specified, any data that is sent to the master from an Instrument Daemon as a part of this analysis will then also be copied to the specified Data Repository location.

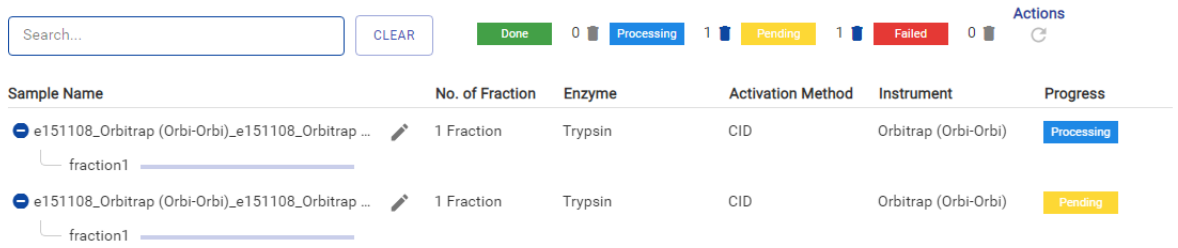
Note

Copying data to a Remote Repository will not modify or affect the original copy of the data on the Instrument Daemon machine.

2.3. Instrument Daemon Samples

Once added to a project, samples from an Instrument Daemon will appear like any other sample on the project data page.

Before the Instrument Daemon starts the upload for a sample, placeholder samples will be created and displayed on the data page.



When a sample finishes loading, the sample names and file names from the placeholder sample will be updated with the actual names.