PEAKS® 5.2

User Manual

Chapter

1

1 WELCOME TO

1.1 MAIN FEATURES

PEAKS is an innovative software system designed to derive amino acid sequences and identify proteins using tandem mass spectrometry data from all major mass spectrometry vendors. PEAKS incorporates *de novo* sequencing results into the database search process for peptide/protein identification. It does this by generating sequence tags which are used in conjunction with fragment ion mass matching, to speed up the search, remove false positive matches, and ultimately find peptides with interesting sequence variations or modifications that would prevent them otherwise being identified. Our multi-engine protein identification search tool, inChorus allows users to utilize multiple search engines (PEAKS, X!Tandem, OMSSA, Sequest and Mascot) to expand sequence coverage and increase confidence. A key feature unique to PEAKS is SPIDER, a homology search tool for *de novo* sequencing. SPIDER is able to reconstruct the correct sequence using the *de novo* sequence and a homologous peptide.

1.2 NEW FEATURES

Throughout this manual, new and existing features will be discussed in considerable detail. Below is a short list of features implemented in PEAKS 5.2:

- Increased de novo and database search accuracy for ETD data
- CID/HCD, CID/PQD support in iTRAQ quantification with Thermo instruments
- Automatic precursor mass and charge correction for high resolution data
- MCP complaint result exporting
- New GUI for result comparison
- Added replicate analysis for label free quantification
- Improved workflows targeting different application needs such as Identification and Quantification
- Overall improvement of software capacity, stability and speed

- Improved spectrum display
- mzML/mzXML3.0/mzData/pepXML file format support
- Standardized PTM display
- Better support for Thermo LTQ RAW file format
- Adjusted SPIDER scoring function
- Reduced the memory overhead for SPIDER, improving its capacity and speed
- Search parameters of other search engines used in inChorus can now be saved and used at a later time
- PTM support in manual de novo
- Options to display assigned as well as unassigned de novo peptides in PEAKS Protein ID results
- Added FASTA database validation, important for custom databases

1.3 Guidelines for Using this Manual

This manual is intended to assist in the use of PEAKS 5.2. It outlines functionalities, provides instruction on how to customize PEAKS to a specific application, provides a task based reference, and offers troubleshooting recommendations. It is highly recommend that users begin by going through the included Walkthrough (Chapter 3) and use the sample data provided to quickly become familiar with most commonly used features of PEAKS.

1.4 Scope

It is assumed that PEAKS users are familiar and comfortable with using computers and their respective operating systems. Given this, it is beyond the scope of this manual to instruct the user on the use of windows, dialogue boxes, menus, file storage etc. Please refer to the operating system's manual or one of the numerous computer help books for such information. Similarly, PEAKS users are expected to be familiar with mass spectrometry, standard operating practices, data acquisition and analysis.

1.5 Service and Support

In addition to reading this manual, it is recommended that users take the time to view the accompanying training videos that explain the main features of PEAKS visually and in detail.

http://www.bioinfor.com/products/peaks/support/tutorials.php

Please send technical questions to support@bioinfor.com.

We strongly encourage users to provide BSI with any suggestions or comments, as BSI is consistently improving and updating PEAKS to meet the future needs of the scientific community.

http://www.bioinfor.com/corporate/contactform.php

Chapter

2

2 GETTING STARTED WITH PEAKS® 5.2

This section of the manual will guide users through the installation and configuration of PEAKS 5.2.

2.1 PACKAGE CONTENTS

The PEAKS 5.2 package contains:

- This manual
- PEAKS 5.2 software

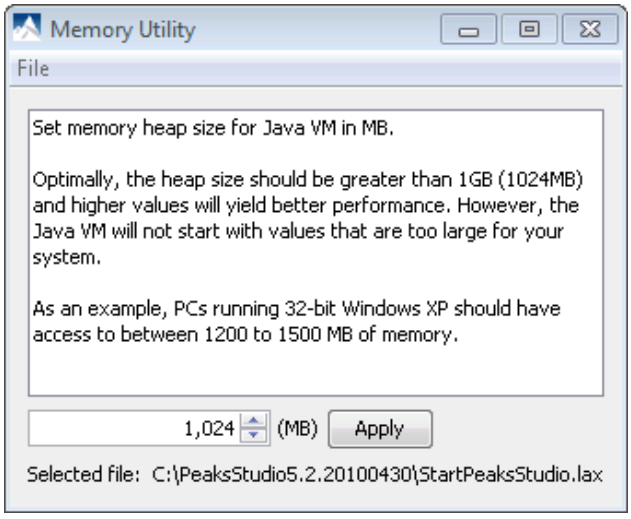
2.2 SYSTEM REQUIREMENTS

PEAKS 5.2 will run on most platforms with the following minimum requirements:

- Processor: Equivalent or superior processing power to an Intel™ Pentium 4 Processor 1.6GHz (we recommend an Intel™ Core2 Duo processor/AMD Phenom™ processor)
- Memory: 1 GB memory (we recommend 2GB memory)
- Operating System: PEAKS runs and has been tested on Windows XP, Vista and 7.

2.2.1 ADJUSTING THE AMOUNT OF MEMORY UTILIZED BY PEAKS

The PEAKS directory (e.g., C:\PeaksStudio5.2) contains a file called Memory Utility.exe. Click to open this file (it also can be accessed from the Windows Start Menu) and the following window will open:



The default of 1,024 MB tells the Java Virtual Memory (which runs PEAKS) to execute with 1,024 MB of memory. You can increase the Java VM to determine the highest optimal value for your computer. Keep in mind that trial and error may be needed as Java will not start if you set the value to be too large for your system. Please do not hesitate to contact BSI if you have any questions about your own system.

2.3 INSTALLATION FOR WINDOWS USERS

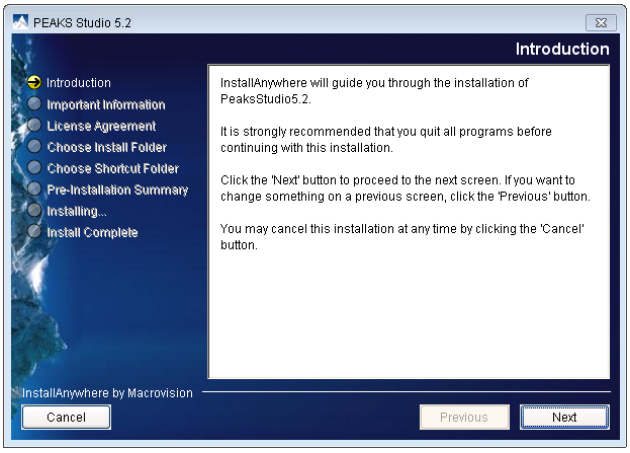
Note: If you already have an older version of PEAKS currently installed on your system, please uninstall it before proceeding.

1) Close all programs that are currently running.

2) Insert the PEAKS 5.2 disc into the CD-ROM drive. If you are installing PEAKS via the download link, move ahead to step 4, after downloading and running the file.

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

4) A menu screen will appear. Select the top item "**PEAKS Installer**." The installation utility will launch the installer. When the "PEAKS 5.2" installation dialogue appears, click the "Next" button.



5) Basic system requirements will be presented. Please note that while the minimum requirement is 1 GB of RAM, the preferred configuration is 2 GB of RAM. Click "Next".

6) Read the license agreement. If you agree, select "I accept the terms of the License Agreement" and click "Next".

7) Choose the folder/directory in which you would like to install PEAKS. The default location is simply "C:\PeaksStudio5.2". To change this location press the "Choose…" button to browse your system directory and make your selection, or type a folder name into the textbox. Avoid installing PEAKS in any directory that contains a white space, for example the "Program Files" directory as some features may not function correctly in such situation. Please make sure that the user account has full access permissions (read/write/execute) on the selected directory. Click "Next".

8) Choose where you would like to place icons for PEAKS 5.2. The default location for the icons is in the Programs section of your Start Menu. A common user preference is on the desktop. Click "Next".

9) The Pre-Installation Summary will allow you to review the selections you have made. You can click "Previous" if you would like to make any changes or click "Next" if you are satisfied with all your specifications.

10) PEAKS 5.2 will now install on your system. You may cancel the process at any time by simply pressing the "Cancel" button located in the lower left corner.

11) When the installation is complete, click “Done”. The PEAKS 5.2 menu screen should still be open. You may view tutorial movies and other materials from here. To access this menu at a future date, simply insert the disc in your CD-ROM drive.

2.4 REGISTERING PEAKS

To complete the registration, the user should have the following information ready.

- A valid email address – The license email will be sent to this address.
- A registration key – The key usually is a 20-character-long alphanumeric string and it is case-sensitive. The registration key is not needed when registering for a 30-day evaluation license.

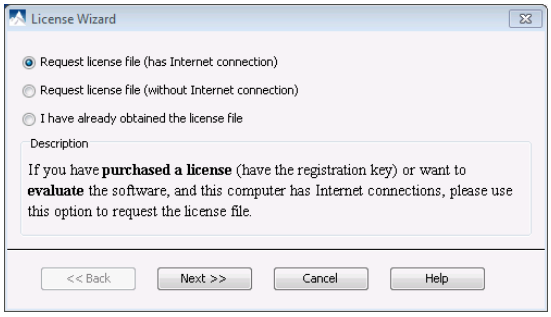
Also during the registration, it is recommended to input ONLY English characters as other character encodings may corrupt the license file.

The first time PEAKS is run, the license wizard will appear automatically. There are three options. The first two options should be used to start the registration process. The third option should be used when a user reinstalls/re-registers the software on the same computer and still has the license file (license.lcs) retrieved from the license email.

2.4.1 REGISTRATION INSTRUCTIONS (WITH INTERNET CONNECTION)

If the computer has internet connection, please follow the following instructions.

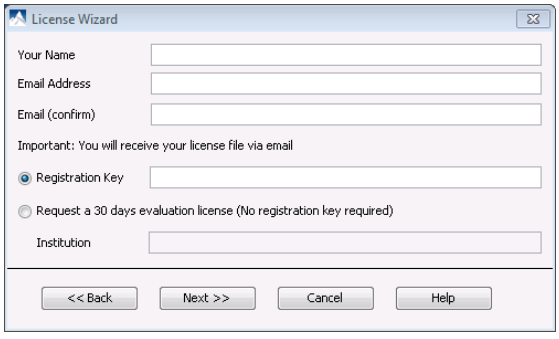
1) Select “Request license file (has Internet connection)” and click “Next”.



   a. The following window will appear:

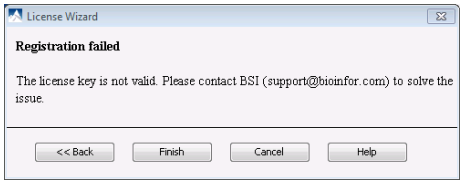
   b. If you have purchased PEAKS and have a registration key, select “Registration Key”. Enter your registration key as well as your name and email address and click “Next”.

   OR



   c. If you are trying a demo of PEAKS and do not have a registration key, select “Request a 30 day evaluation license (No registration key required)”. Enter your name, email address, as well as your institution. Click “Next”.

2) If the license file is requested successfully, the following window will appear:

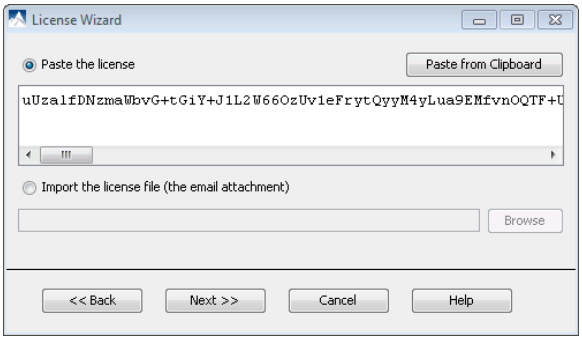


3) An automated BSI service will generate the license file (license.lcs) and send an email (from support@bioinfor.com) to the email address provided from the License Wizard. You can either save the attached license file you receive to a local directory or copy the content between '===>' and '<===' in the email. Click “Next” and go to step 4.

4) If the following error message appears, it means the input registration key is not correct. It is usually caused by a typo in the registration key field in step 2. The registration key is case-sensitive and it is recommended to copy (ctrl-C) / paste (ctrl-V) the registration key whenever possible.

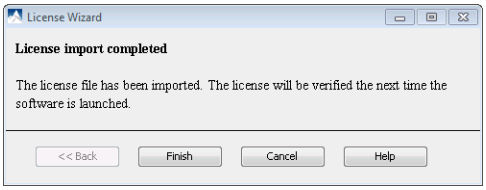
5) If the registration failed with other error messages, please follow the on screen instructions or refer to section 2.4.2 Registration Instructions (Without Internet Connection).

6) The following window will appear:



7) Select “Paste the license” and click on the “Paste from Clipboard” button to paste the copied license information between '===>' and '<===' from the license email or select “Import the license file (the email attachment)” and browse to locate the license file (license.lcs). Click “Next”.

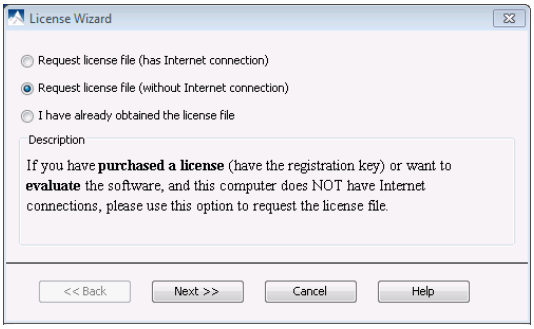
8) The following window will open:



9) Click “Finish” if you receive a message that the license has been imported.

10) Restart PEAKS to begin using the software.

2.4.2 Registration Instructions (Without Internet Connection)



If the computer does not have internet connection, or it is behind a proxy/firewall, or failed using the first option, please follow these instructions.

1) Select "Request license file (without Internet connection)" and click "Next". The following screen will describe the registration process:

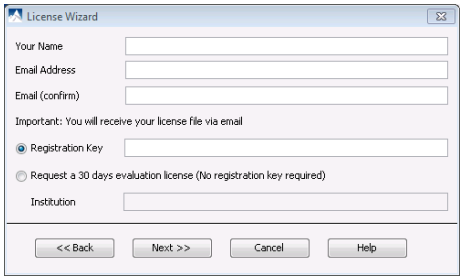
    a. Obtaining a license.request file (using the onscreen guide)

    b. Uploading the request file to a web address via a computer with internet access

    c. Retrieving the license file (license.lcs) via email and importing it to activate the software.

    d. Click Next.

2) The following window will open: If you have purchased PEAKS and have a registration key, select "Registration Key". Enter your registration key as well as your name and email address and click "Next".

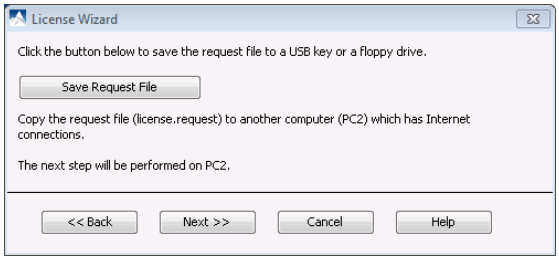


    OR

    If you are trying a demo of PEAKS and do not have a registration key, select "Request a 30 day evaluation license (No registration key required)". Enter your name, email address, as well as your institution. Click "Next".

3) The following window will appear:

    a. Select the "Save Request File" button to save license.request to your computer (PC1). Click "Next".



4) Transfer the license.request file from your computer (PC1) onto a computer with an Internet connection (PC2) using a USB key or a removable storage device. On PC2, go to http://www.bioinfor.com/lcs20/index.jsp

5) On this website, select "I have the license request file. I want to register the software" and click "Next".

6) On the following window click the "Browse" button to select the license.request file on your USB key, then type in the visual verification code and click "Next". The web page may display:

    a. Operation completed successfully. Go to step 7.

b. Invalid product key. It means the input registration key is not correct. It is usually caused by a typo in the registration key field in step 2. The registration key is case-sensitive and it is recommended to copy (ctrl-C) / paste (ctrl-V) the registration key whenever possible. The user will have to go through the registration wizard again to generate a new license.request file with the correct registration key.

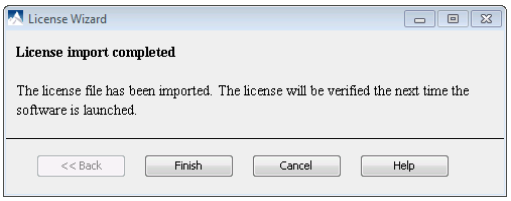
c. Invalid verification code. Sometimes the characters in the image are not very clear therefore the visual verification failed. In this case, the user can click the Back button and start from step 5. The visual verification code is case-sensitive.

d. Unknown error occurred. If this message is displayed, please contact support@bioinfor.com for additional assistance.

7) After the license email is received on PC2, save the attachment, license.lcs, as is and transfer the file to your computer with PEAKS (PC1). If you do not receive the license.lcs file in your inbox, please check your junk mail folder.

8) In the license wizard on your computer (PC1), click the "Browse" button to select the license.lcs file and click "Next".

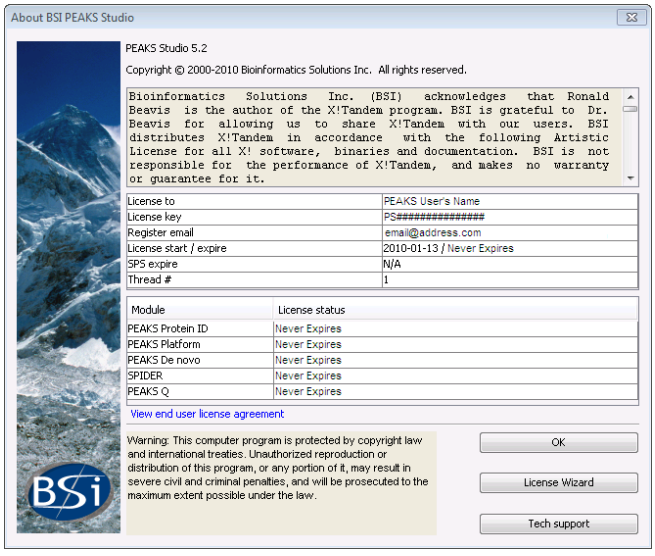
9) Click "Finish" you should receive a message that the license has been imported successfully.



10) Restart PEAKS to begin using the software.

2.4.3 Re-registration Instructions

Re-registering PEAKS may be necessary if you purchased additional software module or renewed SPS. BSI will modify the license information accordingly on the server side. You will need to obtain a new license file in order to make the changes effective. Select "About PEAKS" from the Help menu. The "About BSI PEAKS Studio" dialogue box will appear:



Click the "License Wizard" button to continue. Then follow the instructions previously listed above for re-registering PEAKS.

2.5 Set up PEAKS Preferences

Before running your data, you must set up search engine preferences. For an explanation please refer to section 16.1.3 Search Engine Preferences.

It is possible to analyze your data with PEAKS without changing any of the preferences but the analysis will use the default settings of PEAKS. For more information on changing these default settings refer to section 16.1 PEAKS Environment Preferences.

2.6 Set up PEAKS Configuration

Before running your data, you must configure the databases. For detailed instructions on the configuration process please refer to section 16.2.3 Database Configuration.

For more information on changing these default settings refer to section 16.2 PEAKS Configuration.

Chapter

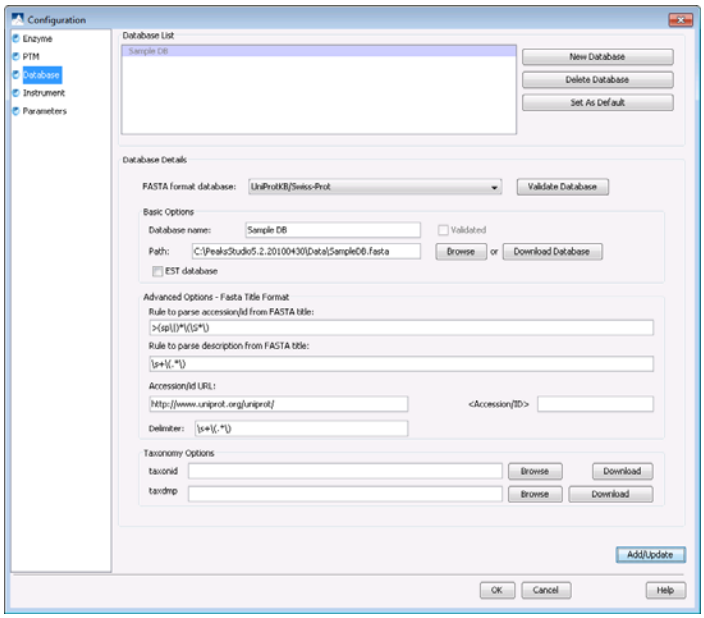
3

3 QUICK WALKTHROUGH

This section of the manual will walk you through the basic workflow of PEAKS 5.2. After completing this section you will see how easy it is to load, view a data file, perform data refinement, de novo sequencing, and protein identification via database searches.

Please note that version upgrades of PEAKS as well as upgrades to the databases may result in small changes to the screenshots displayed in this chapter.

3.1 CREATE A SAMPLE DATABASE

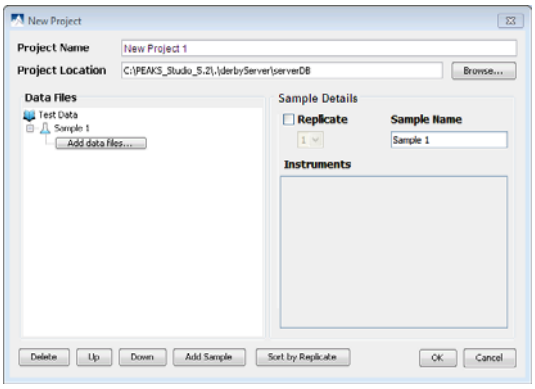


Before beginning with the walkthrough, you need to set up a database. We have provided you with a sample FASTA database called "SampleDB.fasta" in your PEAKS program folder (C:\PeaksStudio5.2\Data).

Click on the configuration toolbar icon or select "Configuration" from the "Tools" menu. Select "Database" from the left hand side of the window. Under "Database Details" select UniProtKB/SwissProt from the dropdown list, and then call this database Sample DB. Locate the database using the browse button (C:/PEAKSStudio5.2/Data/SampleDB.fasta).

You do not need to change any of the other information listed. Click the "Add/Update" button and then click "OK".

3.2 CREATE A PROJECT



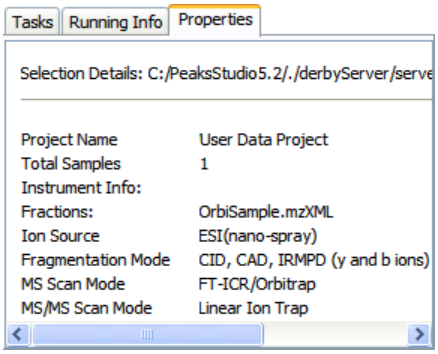
This will be a simple project as it will only contain one sample; however the same process will be used for projects with multiple samples and files. Click on the "Create new project" button or select "New Project" from the "File" menu. The following window will appear:

Enter a name for your project. In this case, we can call it User Data Project. By default the first sample will be named "Sample 1". Click the "Add data files" button to add a data file for Sample 1. Select the "OrbiSample.mzxml" file from your PEAKS program folder (for example "C:\PEAKSStudio5.2\Data") and click "Open". Leave the instrument vendor as "All Instruments" and select "FT-trap" from the "Instrument Type" drop-down list. Click "OK".

As this project is created, we can see the project load in the "Project View" panel, which is located in the upper left hand corner of the screen. The icon to the left of the sample will change from outlined to a solid fill when the sample is ready to be analyzed.

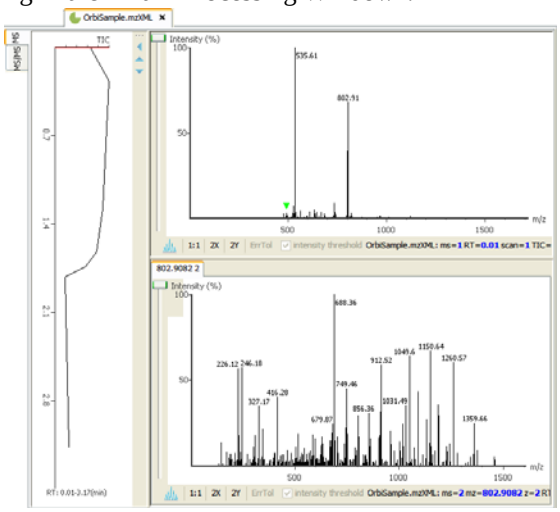
Use the '+' and '-' boxes to expand and collapse the project in order to access the data file in the "Project View". Make sure that you select the data file when choosing data to be analyzed, rather than just the project. As this is just a single sample project, there is no confusion, however, if multiple samples are loaded, clicking on the project will process them all. Selecting a single data file will ensure only the desired file is processed.

PEAKS reads and tracks information about the experiment and data for use in the current analysis and for future reference. Click on the "Properties" tab in the bottom left hand corner to view these features:



By double clicking on the loaded data file, we can see the following in the "Main Processing Window":

The information that is displayed by default pertains to the precursor scan. To the left of the window is the "Total Ion Chromatogram" (TIC). The graph in the upper right corner displays a survey scan with its corresponding tandem scans below. Click on the MS/MS tab to see the spectra that were generated from the tandem scan. For more information on the functions and tools found in these windows, refer to section 4.7 Orienting Yourself.



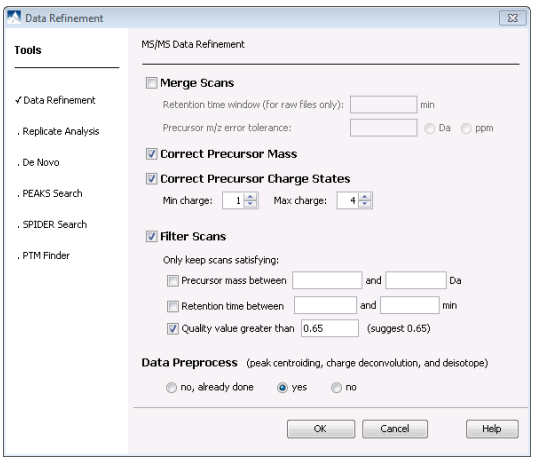
3.3 Perform Data Refinement

1) With the OrbiSample.mzXML file selected, click the Data Refine toolbar icon or select "Data Refine" from the "Tools" menu.

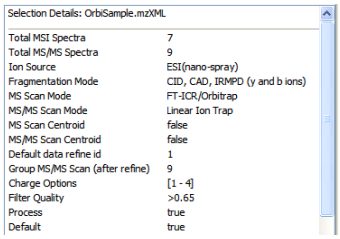
2) Enter the following settings:

- Correct Precursor Mass
- Correct Precursor Charge States (min 1, max 4)
- Filter Scans: quality value greater than 0.65)
- Data Preprocess: yes

For more details on setting up the parameters for data refinement refer to section 5.2 Data Refinement Parameters.



Here a quality filter is used to remove spectra with a quality value lower than 0.65. As all of the spectra in this data set are of good quality, no data will be removed using this data filter. After executing data refinement, there will be new information listed in the "Properties" file.

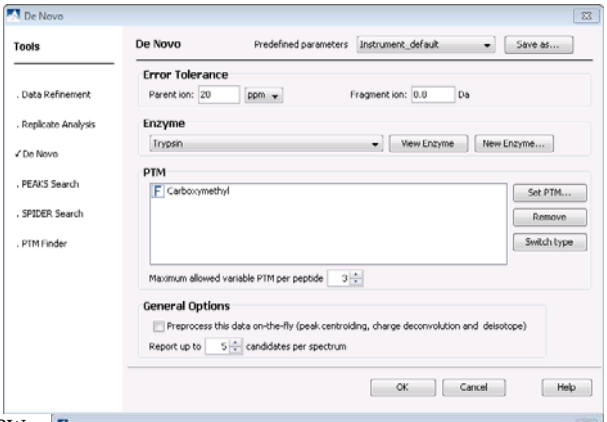


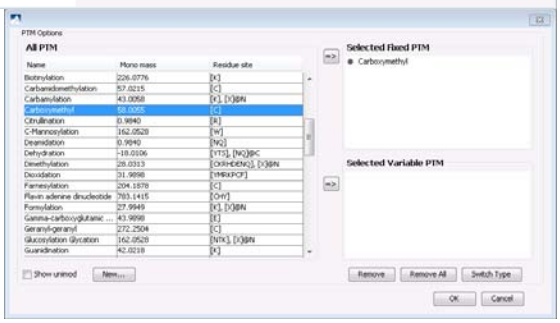
3.4 RUN *DE NOVO* SEQUENCING

1) Click the de novo sequencing toolbar icon or select "de novo" from the "Tools" menu.

2) Enter these settings:

- Error Tolerance:
  - Parent Ion to 20ppm
  - Fragment ion to 0.8 Da.
- Enzyme: Trypsin
- PTM: To enter Carboxymethyl as a fixed PTM click the "Set PTM…" button and a new window will appear. Select Carboxymethyl from the list on the left and click the arrow on the right that corresponds with "Selected Fixed PTM". Then click OK.
- Max allowed variable PTM: 3
- Note that there is no need to preprocess this data set "on the fly" as we have already preprocessed the data during the data refinement stage.

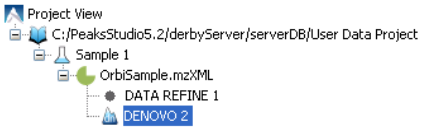




You can save the parameters that you used for future reference by clicking "Save As". For more information on setting up de novo parameters refer to section 6.1 Setting up Auto de novo Sequencing Parameters. Click "OK" to commence analysis.

The PEAKS auto de novo algorithm derives sequence candidates for each of the nine spectra in our example data file. Once the de novo sequencing is completed the results file will appear in the "Project View Panel":

Double click on the de novo file and the results will appear in the "Main Processing Window":



Take a look at spectra ID 6. Notice that the (+58.01) refers to the modification which in this case is Carboxymethyl.

At the top of the screen you will see the peptide candidates in the "Peptide Candidates Frame". The peptide candidates are sorted by "ID". Right next to the proposed sequence, the auto de novo "Total Local Confidence" (TLC) and "Average Local Confidence" (ALC%) confidence scores are shown. You will also see the m/z ratio, mass, retention time etc. listed for each peptide sequence. For information on the "Peptide Candidates Frame including color coding, refer to section 6.2.1 Peptide Candidates Frame.

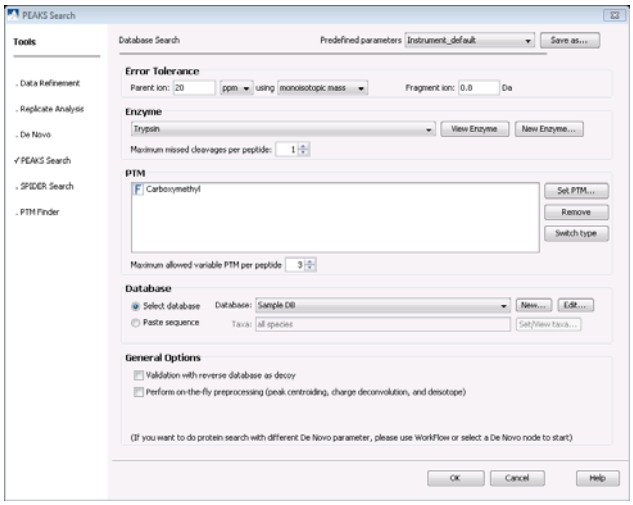
Below the "Peptide Candidates Frame" is the "Ion Table Frame". Each amino acid is color coded by confidence level (refer to section 6.2.2 Ion Table Frame) with the masses for matched a, b and c ions listed in blue and for the matched x, y and z ions listed in red.

Below the "Ion Table Frame" is the "Spectrum View Frame". This frame is useful for seeing the strengths of ms/ms peaks that PEAKS has identified as ions. Here the alignment of the assigned b (blue) and y (red) ions with the entire spectrum corresponding to the selected peptide can be observed. For more information on the "Spectrum View Frame", refer to section 6.2.3 Spectrum View Frame.

By clicking on the tabs are the bottom of the page, you can view the corresponding survey scan, the spectrum alignment and the error map. For more information, refer to sections 6.2.4 Survey Scan, 6.2.5 Spectrum Alignment Frame and 6.2.6 Error Map.

3.5 Run Protein Identification

1) With the de novo result selected, click the PEAKS Search toolbar icon  or select "PEAKS Search" from the "Tools" menu.



2) Enter the following settings:

- Error Tolerance:
  - Parent Ion to 20ppm
  - Fragment ion to 0.8 Da.
- Enzyme: Trypsin
- PTM: To enter Carboxymethly as a fixed PTM click the "Set PTM…" button and a new window will appear. Select Carboxymethyl from the list on the left and click the arrow on the right that corresponds with "Selected Fixed PTM". Then click OK.
- Max allowed variable PTM: 3
- Note that there is no need to preprocess this data set "on the fly" as we have already preprocessed the data during the data refinement stage.

As with de novo sequencing, to enter Carboxymethyl as a fixed PTM click "Set PTM…" from the right and a new window will appear. Select Carboxymethyl from the list on the left and click the arrow on the right that corresponds with "Selected Fixed PTM". Then click OK.

Parameters can be saved and used for future reference by clicking on the "Save Parameter" button. For more information on setting up protein identification parameters refer to section 7.1 Setting up Protein Identification Parameters.

3) Click "OK" to commence analysis.

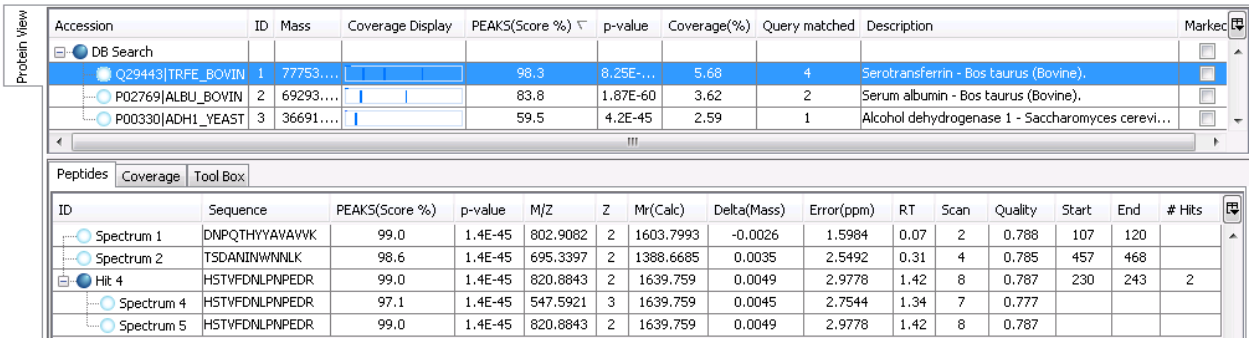
Once the Protein ID file appears in the "Project View Panel", double click on the file and the results will appear in the "Main Processing Window":

The "Peptide View" will appear in the "Main processing" window by default. The "Peptide ID List" shows each spectrum for which PEAKS found a matching peptide. Since there may be more than one spectrum that matches a

peptide, these spectra would be listed together under a Hit node. Use the '+' and '-' boxes to expand and collapse the node to see the spectra that are listed together. With this dataset, spectra 8 and 9 can both be found under one hit.

The "Peptide Alignment Panel" contains an "Ion Table", "Spectrum View Frame", "Spectrum Alignment Frame", "Survey Scan" and "Error Map". Click on the "Peptide Details" tab. If the spectrum has a de novo hit, then a simple alignment between the original de novo sequence and the peptide found in the database will be shown. At the bottom of the "Peptide Details" panel you will see where the selected peptide matches the protein. For more information about these sections refer to sections 6.2.2 Ion Table Frame - 6.2.6 Error Map.

Click on the "Protein View" tab on the upper left hand side. The following window will appear:

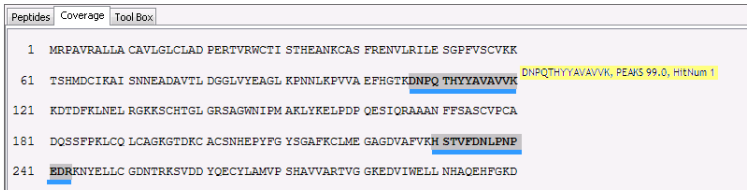


Accession	ID	Mass	Coverage Display	PEAKS(Score %)	p-value	Coverage(%)	Query matched	Description
DB Search								
Q29443|TRFE_BOVIN	1	77753....		98.3	8.25E-...	5.68	4	Serotransferrin - Bos taurus (Bovine).
P02769|ALBU_BOVIN	2	69293....		83.8	1.87E-60	3.62	2	Serum albumin - Bos taurus (Bovine).
P00330|ADH1_YEAST	3	36691....		59.5	4.2E-45	2.59	1	Alcohol dehydrogenase 1 - Saccharomyces cerevi...

Peptides | Coverage | Tool Box

ID	Sequence	PEAKS(Score %)	p-value	M/Z	Z	Mr(Calc)	Delta(Mass)	Error(ppm)	RT	Scan	Quality	Start	End	# Hits
Spectrum 1	DNPQTHYYAVAVVK	99.0	1.4E-45	802.9082	2	1603.7993	-0.0026	1.5984	0.07	2	0.788	107	120	
Spectrum 2	TSDANINWNNLK	98.6	1.4E-45	695.3397	2	1388.6685	0.0035	2.5492	0.31	4	0.785	457	468	
Hit 4	HSTVFDNLPNPEDR	99.0	1.4E-45	820.8843	2	1639.759	0.0049	2.9778	1.42	8	0.787	230	243	2
Spectrum 4	HSTVFDNLPNPEDR	97.1	1.4E-45	547.5921	3	1639.759	0.0045	2.7544	1.34	7	0.777			
Spectrum 5	HSTVFDNLPNPEDR	99.0	1.4E-45	820.8843	2	1639.759	0.0049	2.9778	1.42	8	0.787			

PEAKS presents a list of proteins, ranked in descending order from highest to lowest score. Clicking on any protein will display the peptides matched to that protein in the bottom pane. In this case the highest scoring protein match is Serotransferrin precursor from bovine. This protein has two matching peptides, which you can see in the "Peptide ID" list when you select the "Peptides" tab. To see exactly where the peptide corresponds to the protein sequence, select the "Coverage" tab. The entire sequence of the protein is shown and the matched parts are underlined in blue and highlighted. In this case the total matched portions accounts for 5.68% of the protein. Place mouse on the blue line, the hit information of that peptide will be shown.



3.6 RUN PTM FINDER

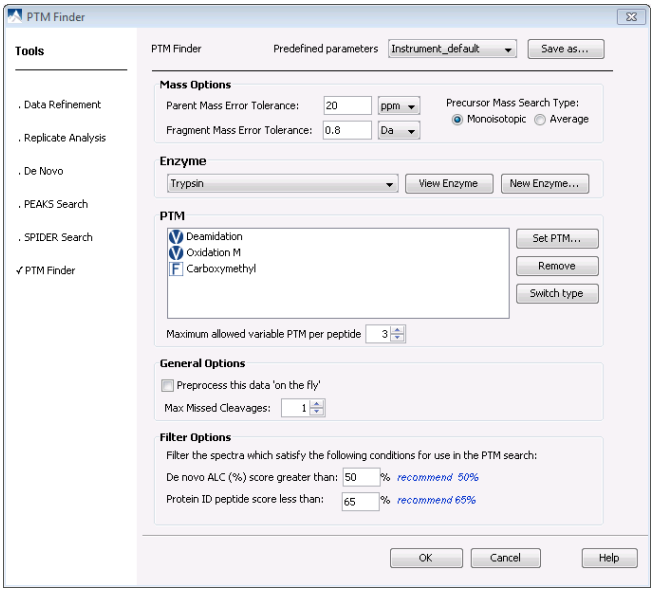
Using the PTM finder, you can identify any additional PTMs and increase the coverage of the proteins that we have found. It is important to note that the PTM Finder can only be applied to a Protein ID result file. As it is very time consuming to run Protein ID with many PTMs, this allows searching for more PTMs in less time. Make sure that you click on a Protein ID result before performing a PTM Finder search.

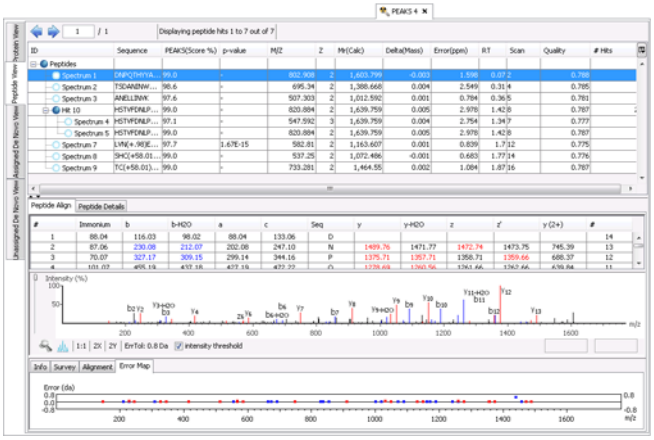
1) Click the PTM Finder toolbar icon or select "PTM Finder" from the "Tools" menu.

2) Enter the following settings:

   - Error Tolerance:
     - Parent Ion to 20ppm
     - Fragment ion to 0.8 Da.
   - Enzyme: Trypsin
   - PTM: Note that Carboxymethyl is fixed. Deamidation and Oxidation M are variable modifications.
   - Max allowed variable PTM: 3
   - Note that there is no need to preprocess this data set "on the fly" as we have already preprocessed the data during the data refinement stage.
   - Parameters can be saved and used for future reference by clicking on the "Save Parameter" button. For more information on setting up PTM Finder parameters refer to section 8.1 Setting up PTM Finder Parameters.

3) Click "OK" to commence analysis.





After the PTM Finder search is complete, double click on the result node. The "Peptide View" window will appear:

The results will be displayed in the same format as was seen for Protein ID. Recall that the Protein ID search identified spectra 1, 2, 3, 4, 5, 8 and 9. The PTM finder search also displayed spectra 7 with the addition of deamidation on N.

3.7 RUN AN INCHORUS SEARCH

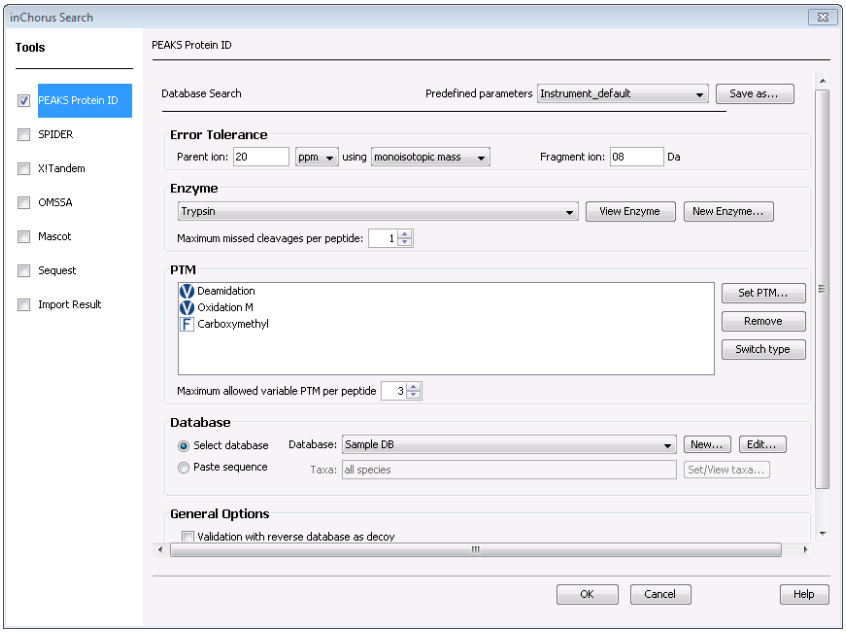
Performing a search with the same data via different search engines is useful both for finding new proteins and confirming others. You can perform an inChorus search using PEAKS Protein ID, X!Tandem, OMSSA, Mascot and Sequest. For this example we will be performing a local search using the X!Tandem and OMSSA search engines. If you have not already set up your search engine preferences, refer to section 16.1.3 Search Engine Preferences for more instructions.

1) Click on the Orbisample.mzxml file. Then click the inChorus Search toolbar icon or select "inChorus Search" from the "Tools" menu.

2) The inChorus Search window will open. "PEAKS Protein ID" box will initially be checked. Enter the following settings:

- Error Tolerance:
  - Parent Ion to 20ppm
  - Fragment ion to 0.8 Da.
- Enzyme: Trypsin
- PTM: Note that Carboxymethyl is fixed. Deamidation and Oxidation M are variable modifications.
- Max allowed variable PTM: 3

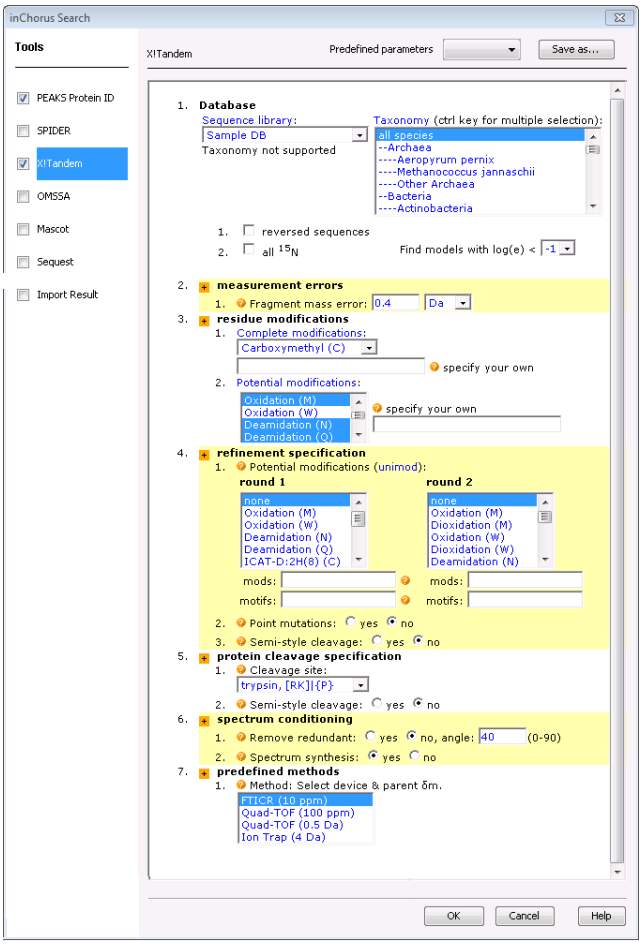
Note that there is no need to preprocess this data set "on the fly" as we have already preprocessed the data during the data refinement stage.



3) Next, check the "X!Tandem" box.

4) Enter the following settings:

- Database: Sample DB, all species
  - Fragment mass error: 0.8 Da.
- Complete modifications: Carboxymethly (C)
- Potential modifications: Oxidation (M), Deamidation (N) and (D)
- Refinement Specification: Round 1 (none), Round 2 (none)
- Cleavage site: trypsin, [RK]|{P}
- FTICR (10 ppm)

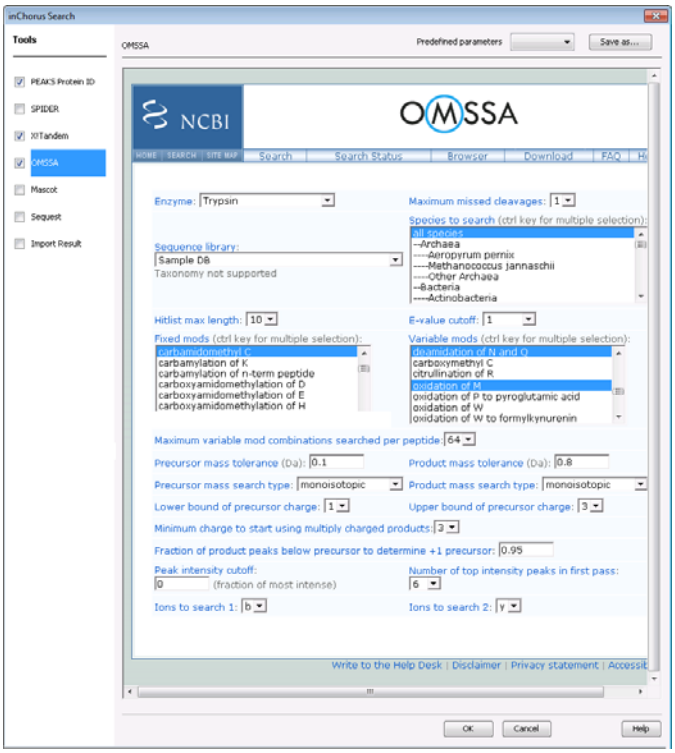


5) Finally, check the “OMSSA” box.

6) Enter the following settings:

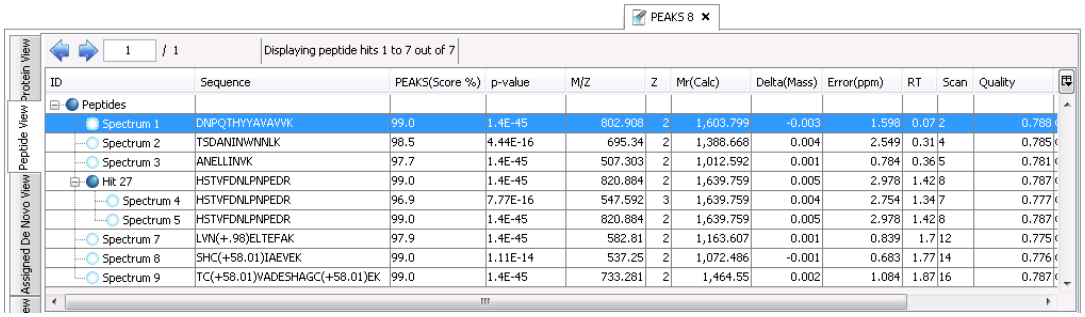
- Enzyme: Trypsin
- Max missed cleavages: 3
- Sequence library: Sample DB, all species
- Fixed modifications: carboxymethly C
- Variable modifications: Deamidation N and D, Oxidation of M
- Precursor mass tolerance (Da): 0.1
- Product mass search type (Da): 0.8

7) Click the “OK” button.



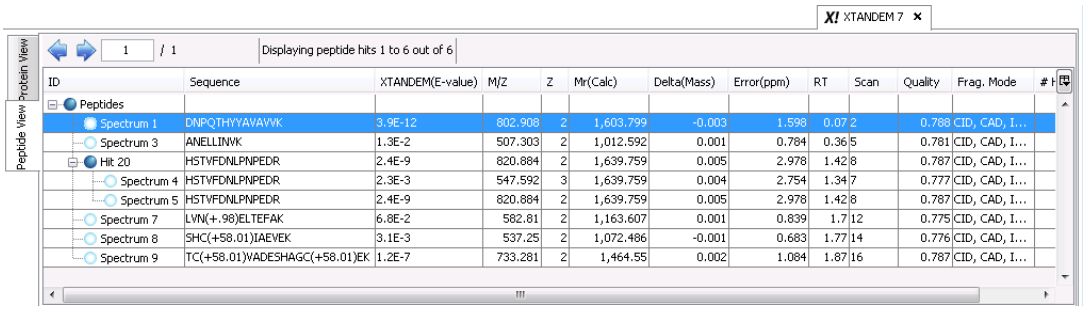
When the inChorus search is complete you should see the following new additions in the “Project View” panel. Presented here are individual reports for PEAKS, X!Tandem and OMSSA as well as an inChorus report that compares the individual reports. To see each of these reports, double click on the report that you would like to see in the “Project View” panel.

The “Peptide View” results for the PEAKS Protein ID search can be seen below:



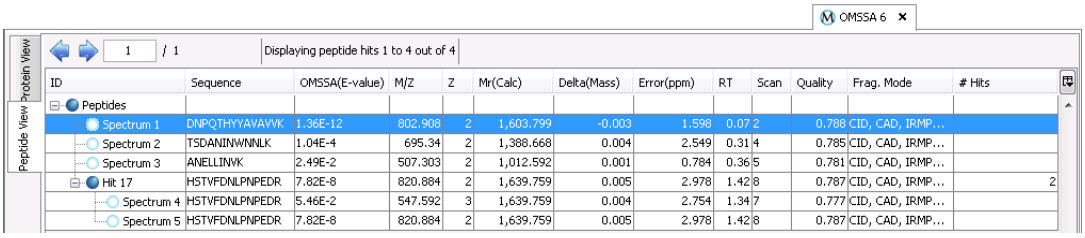
ID	Sequence	PEAKS(Score %)	p-value	M/Z	Z	Mr(Calc)	Delta(Mass)	Error(ppm)	RT	Scan	Quality
Peptides											
Spectrum 1	DNPQTHYYAVAVVK	99.0	1.4E-45	802.908	2	1,603.799	-0.003	1.598	0.07	2	0.788
Spectrum 2	TSDANINWNNLK	98.5	4.44E-16	695.34	2	1,388.668	0.004	2.549	0.31	4	0.785
Spectrum 3	ANELLINVK	97.7	1.4E-45	507.303	2	1,012.592	0.001	0.784	0.36	5	0.781
Hit 27	HSTVFDNLPNPEDR	99.0	1.4E-45	820.884	2	1,639.759	0.005	2.978	1.42	8	0.787
Spectrum 4	HSTVFDNLPNPEDR	96.9	7.77E-16	547.592	3	1,639.759	0.004	2.754	1.34	7	0.777
Spectrum 5	HSTVFDNLPNPEDR	99.0	1.4E-45	820.884	2	1,639.759	0.005	2.978	1.42	8	0.787
Spectrum 7	LVN(+.98)ELTEFAK	97.9	1.4E-45	582.81	2	1,163.607	0.001	0.839	1.7	12	0.775
Spectrum 8	SHC(+58.01)IAEVEK	99.0	1.11E-14	537.25	2	1,072.486	-0.001	0.683	1.77	14	0.776
Spectrum 9	TC(+58.01)VADESHAGC(+58.01)EK	99.0	1.4E-45	733.281	2	1,464.55	0.002	1.084	1.87	16	0.787

The “Peptide View” results for the X!Tandem search can be seen below:



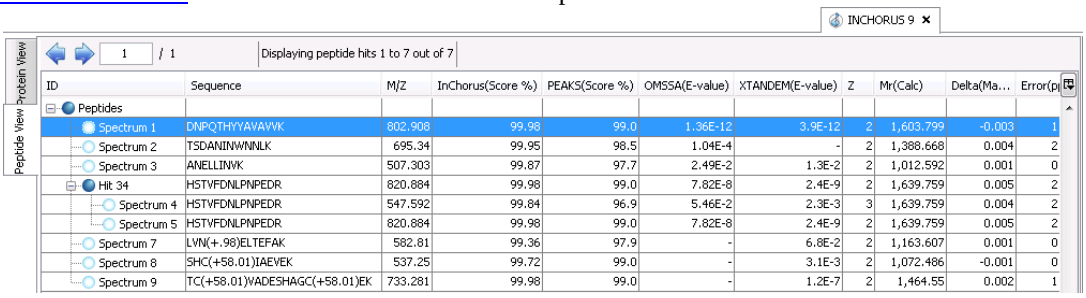
ID	Sequence	XTANDEM(E-value)	M/Z	Z	Mr(Calc)	Delta(Mass)	Error(ppm)	RT	Scan	Quality	Frag. Mode
Peptides											
Spectrum 1	DNPQTHYYAVAVVK	3.9E-12	802.908	2	1,603.799	-0.003	1.598	0.07	2	0.788	CID, CAD, I...
Spectrum 3	ANELLINVK	1.3E-2	507.303	2	1,012.592	0.001	0.784	0.36	5	0.781	CID, CAD, I...
Hit 20	HSTVFDNLPNPEDR	2.4E-9	820.884	2	1,639.759	0.005	2.978	1.42	8	0.787	CID, CAD, I...
Spectrum 4	HSTVFDNLPNPEDR	2.3E-3	547.592	3	1,639.759	0.004	2.754	1.34	7	0.777	CID, CAD, I...
Spectrum 5	HSTVFDNLPNPEDR	2.4E-9	820.884	2	1,639.759	0.005	2.978	1.42	8	0.787	CID, CAD, I...
Spectrum 7	LVN(+.98)ELTEFAK	6.8E-2	582.81	2	1,163.607	0.001	0.839	1.7	12	0.775	CID, CAD, I...
Spectrum 8	SHC(+58.01)IAEVEK	3.1E-3	537.25	2	1,072.486	-0.001	0.683	1.77	14	0.776	CID, CAD, I...
Spectrum 9	TC(+58.01)VADESHAGC(+58.01)EK	1.2E-7	733.281	2	1,464.55	0.002	1.084	1.87	16	0.787	CID, CAD, I...

"Peptide View" results for the OMSSA search can be seen below:



ID	Sequence	OMSSA(E-value)	M/Z	Z	Mr(Calc)	Delta(Mass)	Error(ppm)	RT	Scan	Quality	Frag. Mode	# Hits
Peptides												
Spectrum 1	DNPQTHYYAVAVVK	1.36E-12	802.908	2	1,603.799	-0.003	1.598	0.07	2	0.788	CID, CAD, IRMP…	
Spectrum 2	TSDANINWNNLK	1.04E-4	695.34	2	1,388.668	0.004	2.549	0.31	4	0.785	CID, CAD, IRMP…	
Spectrum 3	ANELLINVK	2.49E-2	507.303	2	1,012.592	0.001	0.784	0.36	5	0.781	CID, CAD, IRMP…	
Hit 17	HSTVFDNLPNPEDR	7.82E-8	820.884	2	1,639.759	0.005	2.978	1.42	8	0.787	CID, CAD, IRMP…	2
Spectrum 4	HSTVFDNLPNPEDR	5.46E-2	547.592	3	1,639.759	0.004	2.754	1.34	7	0.777	CID, CAD, IRMP…	
Spectrum 5	HSTVFDNLPNPEDR	7.82E-8	820.884	2	1,639.759	0.005	2.978	1.42	8	0.787	CID, CAD, IRMP…	

The inChorus report contains most of the information that is seen in a PEAKS Protein ID results file, refer to section 7.2 Protein Identification Results for more details. Click on the "Peptide View" tab.

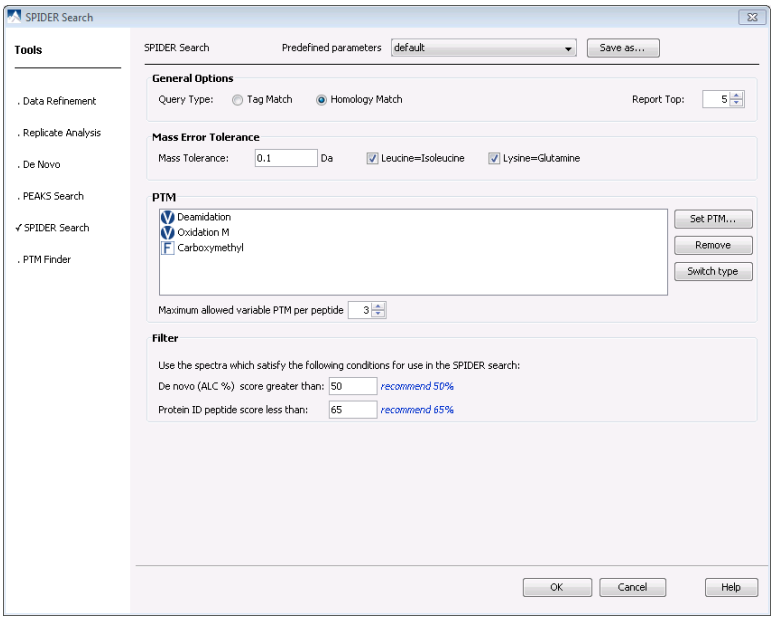


ID	Sequence	M/Z	InChorus(Score %)	PEAKS(Score %)	OMSSA(E-value)	XTANDEM(E-value)	Z	Mr(Calc)	Delta(Ma…	Error(p…
Peptides										
Spectrum 1	DNPQTHYYAVAVVK	802.908	99.98	99.0	1.36E-12	3.9E-12	2	1,603.799	-0.003	1
Spectrum 2	TSDANINWNNLK	695.34	99.95	98.5	1.04E-4	-	2	1,388.668	0.004	2
Spectrum 3	ANELLINVK	507.303	99.87	97.7	2.49E-2	1.3E-2	2	1,012.592	0.001	0
Hit 34	HSTVFDNLPNPEDR	820.884	99.98	99.0	7.82E-8	2.4E-9	2	1,639.759	0.005	2
Spectrum 4	HSTVFDNLPNPEDR	547.592	99.84	96.9	5.46E-2	2.3E-3	3	1,639.759	0.004	2
Spectrum 5	HSTVFDNLPNPEDR	820.884	99.98	99.0	7.82E-8	2.4E-9	2	1,639.759	0.005	2
Spectrum 7	LVN(+.98)ELTEFAK	582.81	99.36	97.9	-	6.8E-2	2	1,163.607	0.001	0
Spectrum 8	SHC(+58.01)IAEVEK	537.25	99.72	99.0	-	3.1E-3	2	1,072.486	-0.001	0
Spectrum 9	TC(+58.01)VADESHAGC(+58.01)EK	733.281	99.98	99.0	-	1.2E-7	2	1,464.55	0.002	1

The "Peptide View" of an inChorus report contains the scores received by each search engine involved in the inChorus search. A "–" indicates that the search engine did not find that a protein sequence for that particular spectrum. Notice that while PEAKS Protein ID found spectra 1-5, 7-9 X!Tandem found 1, 3-9 and OMSSA confirmed spectra 1-5. In this case, the X!Tandem and OMSSA searches did not generate any extra results that PEAKS did not find but did help to confirm quality matches.

3.8 Perform a SPIDER Search

In this example, spectrum 6 has not been identified with database searching tools. In order to gain more information from our data, we will run a SPIDER search next. For more information about the SPIDER search refer to section 9 SPIDER Search.



1) Click on the Protein ID Result node.

2) Next, click the SPIDER Search toolbar icon or select "SPIDER Search" from the "Tools" menu.

3) Enter the settings as shown:

- Query Type: Homology Match
  - Mass Tolerance: 0.1 Da
- Leucine=Isoleucine
- Lysine=Glutamine
- PTM: Carboxymethyl as fixed. Deamidation and Oxidation M as variable modifications.

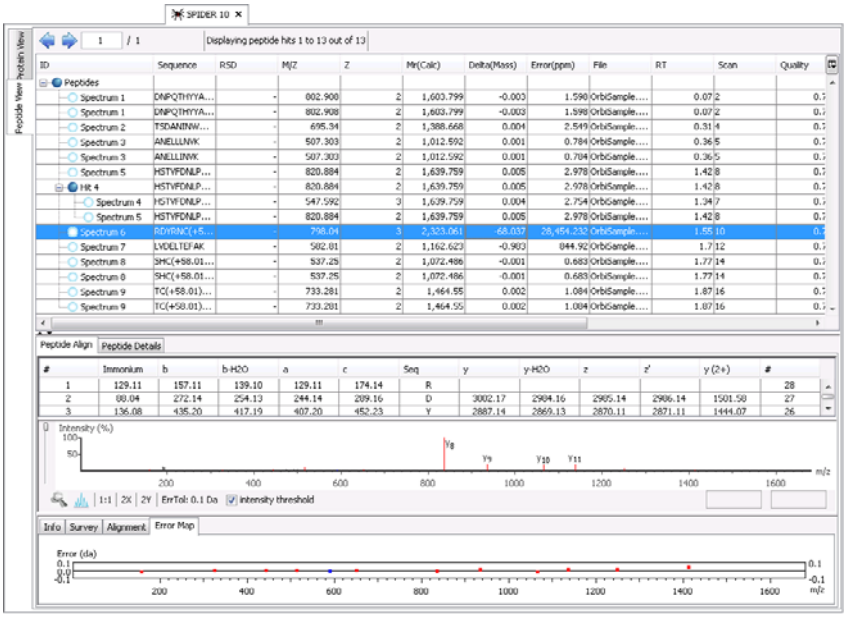
- Max allowed variable PTM: 3
- De Novo (ALC%) score greater than: 50
- Protein ID peptide score less than: 65

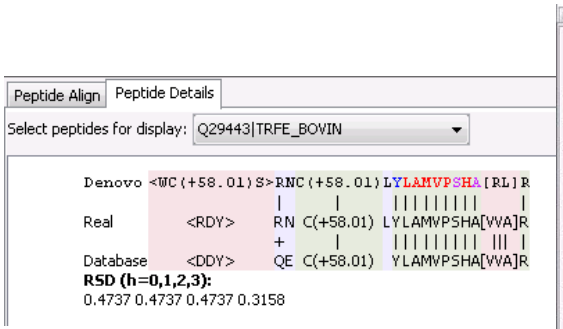
Parameters can be saved and used for future reference by clicking on the “Save Parameter” button. For more information on setting up SPIDER Search parameters refer to section 9.1 Setting up SPIDER Parameters. Click “OK” to commence analysis.

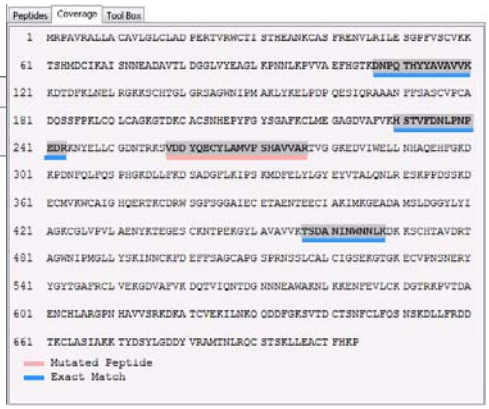
After the SPIDER search has completed, double click on the SPIDER results node. The “Peptide View” window will appear. The format is similar to what was seen in the results of a Protein ID search with the addition of a few columns of information such as Homolog used, PEAKS result used and SPIDER score. Note that all spectra can now be identified by the SPIDER search. Spectrum 6 is finally identified.

Clicking on the “Peptide Details” tab will display the protein with its matched peptides in red. SPIDER will also display a reconstructed sequence. Refer to section 9.2 SPIDER Results View for more information.

Click on the “Protein View” tab, select a result and then click on the “Coverage” tab to see where the SPIDER results overlap with the protein sequence. Note that mutated sequences identified by SPIDER are shown in red while exact matches found by PEAKS protein ID or SPIDER are in blue.







Chapter

4

4 LOADING SPECTRUM DATA

4.1 DATA FORMATS

Before loading data files into PEAKS, you must make sure that the data is in an accessible format. PEAKS handles data files in the following formats:

- .RAW files from Thermo Fisher Scientific instruments
- .D directories from Agilent QTOF instruments
- .BAF, .YEP and folders of .FID files from Bruker instruments
- .RAW directories from Waters QTOF instruments
- .WIFF files from AB/Sciex QSTAR and QTRAP instruments
- .RUN folders from Shimadzu instruments
- mzXML format
- mzData format
- mzML format
- .DTA files
- .MGF files
- .PKL files
- .XMS files from Varian instruments
- PEAKS 5.1 project directories
- PEAKS 5 project directories
- .ANZ – the zip compressed XML based file format associated with PEAKS 4.5 and earlier

4.2 DATA CONVERSION

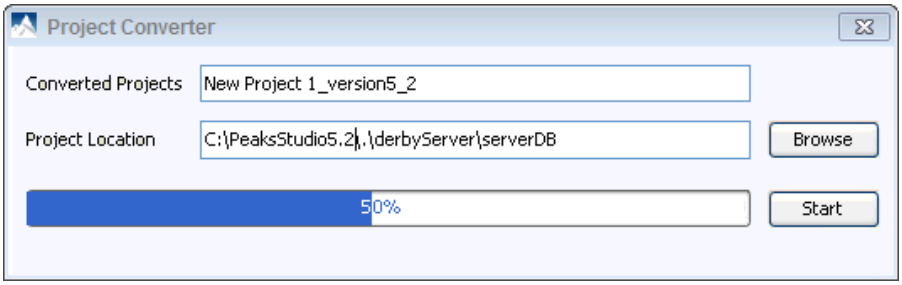
It is best to import RAW data directly, so that PEAKS can access the complete, unprocessed experimental data including the MS survey scans and corresponding retention time information. This will ensure that the PEAKS analysis does not suffer from poor preprocessing.

In order to load RAW data from different vendors, PEAKS may require third-party software to be installed. Please consult the following instructions for third-party software requirements.

4.2.1 PEAKS PROJECT CONVERTER

Convert from PEAKS 5.1 to 5.2

Projects created in PEAKS 5.1 can be opened in PEAKS 5.2. To convert the project to a PEAKS 5.2 project, open the PEAKS 5.1 project in PEAKS 5.2. The project will be recognized as a PEAKS project from an older version. A window will appear stating, "This project was created by an older version of PEAKS. Do you want to convert it now?" Choose Yes. The following dialog box will appear.



In the new window choose the converted project name and location. If a location is not chosen, the PEAKS 5.2 derby server is the default location. A new version of the project will be created at the new location and the old project will be preserved in its original location.

Convert from PEAKS Studio 5.0 to 5.2

A direct conversion from PEAKS 5.0 projects to 5.2 projects is only supported in command line. PEAKS provides a straight-forward project conversion tool that will allow users to quickly convert existing 5.0 projects into PEAKS 5.2 format. The resulting PEAKS 5.2 project can be loaded PEAKS Studio 5.2. Users comfortable with command line, will find the tool very easy. This section will cater to those unfamiliar with command line, but who wish to convert their projects promptly.

There are two methods that can be used to generate a PEAKS 5.2 project from an existing 5.0 project.

1) The first method converts a 5.0 project and stores the 5.2 project in the same directory as the original file, with the addition of "52" to the end of the file. Just enter ProjectConverter followed by the PEAKS 5.0 project location and name. For example, if the project is found in your C:\PeaksStudio5.0\derbyServer\serverDB\My5.0Project folder, enter

```
ProjectConverter "C:\PeaksStudio5.0\derbyServer\serverDB\My5.0Project"
```

This will produce a new 5.2 project in this same directory, with the name My5.0Project.52.

The quotation marks (" ") surrounding the location are not necessary in this case, but are necessary if your path contains any spaces, such would be found in use of "C:/My Documents/Projects…".

2) To produce a new file in a different directory, the only change needed is the addition of a blank space, followed by entering the output directory, such as "C:\PeaksStudio5.2\derbyServer\serverDB\My5.2Project"

Thus the line will look like (in one line)

```
ProjectConverter  "C:\PeaksStudio5.0\derbyServer\serverDB\My5.0Project"
"C:\PeaksStudio5.2\derbyServer\serverDB\My5.2Project"
```

If the new directory does not already exist, the program will create the directory for you. If the directory already exists, entering a dash and the letter '-f' before the original project will force the ProjectConverter to overwrite any existing projects in this location.

For example:

```
ProjectConverter  -f "C:\PeaksStudio5.0\derbyServer\serverDB\My5.0Project"
```

Or

```
ProjectConverter  -f "C:\PeaksStudio5.0\derbyServer\serverDB\My5.0Project"
"C:\PeaksStudio5.2\derbyServer\serverDB\My5.2Project"
```

You are now ready to call up these projects into PEAKS 5.2.

4.2.2 AGILENT DATA

PEAKS 5.2 can load native data from Agilent QTOF, provided that MassHunter software is installed on the same computer.

Agilent Ion Traps produce a .d folder which contains all data associated with that data set. The spectral data is contained in the .yep file. CompassXport can be used to read the .yep file and convert it to an mzXML file inside the .d folder. This file is then loaded into PEAKS. CompassXport must be installed on the same computer as PEAKS to perform this conversion automatically.

4.2.3 APPLIED BIOSYSTEMS (WIFF) DATA

WIFF data from Applied Biosystems/Sciex QSTAR (or QTRAP) mass spectrometers can be loaded, provided that the Analyst QS (Analyst 1.4.1 for QTRAP) software and the MSX plug-in are installed on the same computer as PEAKS 5.2. The MSX tool is produced and sold by Infochromics Ltd., and is available (at cost) from Bioinformatics Solutions Inc. Please contact a BSI sales representative to obtain an evaluation or full license.

4.2.4 APPLIED BIOSYSTEMS (4700/4800) DATA

BSI has created a converter to extract the data from an ABI-Oracle database. If you require this separate, free tool, contact your sales representative. Once installed, you can start up the ABI 4700 Data Extractor from the Start menu.

System Requirements

This extractor can be installed on the same machine as ABI 4700 Explorer and the Oracle database (we will call this machine the 4700 SERVER in the following instructions) or another machine that has direct network access (we require that there is no firewall or proxy) to the 4700 SERVER. Windows 2000 or Windows XP is recommended for use with this tool.

Configuration

Before using the ABI 4700 Data Extractor, it must be configured. To do so, choose "Settings" from the "File" menu. Configuration requires the following:

- 4700 SERVER Name or IP Address: input "localhost" if the Extractor is running on the 4700 SERVER (this is the default value), otherwise enter the IP address of the 4700 SERVER.
- The socket used by the 4700 SERVER: this is the port that the Oracle database listens to (the default is 1521).

- Username to access the Oracle database: most likely we do not need to change this (the default is "tsquared").

- Password to access the Oracle database: mostly likely you do not need to change this one either.

Data extraction procedure

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

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

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

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

4.2.5 Bruker Data

PEAKS 5.2 can load data from Bruker mass spectrometers provided that the CompassXport software is installed on the same computer. If loading fid files, select the top level folder to load them all at once (make sure the merge option is set in the Bruker instrument preferences refer to section 16.1.2.2 Bruker (.yep/baf, fid)). CompassXport 3.0 is readily available on the Bruker Daltonics web site. Autoflex and Ultraflex Tof/Tof instrument data files may require CompassXport 1.3. If spectra cannott be loaded into PEAKS with CompassXport 3.0 contact Customer Support. You may need to contact your Bruker representative to obtain CompassXport 1.3.

4.2.6 Shimadzu Data

Shimadzu mass spectrometer data can be loaded, provided that the Shimadzu software is installed on the same computer as PEAKS 5.2.

4.2.7 Thermo Data

RAW data from Thermo Fisher Scientific mass spectrometers can be loaded, provided that the XCalibur software or the Thermo MSFileReader package is installed on the same computer as PEAKS 5.2.

4.2.8 Varian Data

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

The .trans type data (raw) is converted in Varian programs by clicking "File", "Save As" and selecting the .pkl file format or by clicking "File", right clicking "Export" and selecting ".pkl". If you are viewing a chromatogram with the Varian software, all the spectral data in the viewed chromatogram is converted to the .pkl format. If you are viewing a single spectrum and choose to convert the data, only the viewed spectra will be converted.

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

4.2.9 Waters/Micromass (MassLynx) Data

PEAKS 5.2 can import RAW data from Waters/MicroMass QTOF instruments, provided that MassLynx 4.1 software is installed on the same computer. For MassLynx 4.0 users, a utility called wolf.exe (originally created as

part of the Sashimi Project) can be used to convert the raw files to mzXML format in command line. If you encounter difficulty loading Waters data into PEAKS you may require a different version of wolf, please visit:

www.bioinfor.com/products/peaks/support/watersmicromass.php

4.3 Create a New Project

1) To create a new project, select "New Project" from the file menu or using the "New project" icon  on the toolbar. A new window will appear.

2) Create a name for your project. The default project location will be automatically displayed. Click "Browse" to select where you would like to save the project.

3) To add data files to sample 1, click the "Add data files" button. Browse for the data file and click the "Open" button. Once the data file appears, select the "Instrument Vendor" and "Instrument type" that was used to generate the experimental data from the drop-down lists. Selecting the "All Instruments" option from the "Instrument Vendor" drop-down list will display some general instrument types in the "Instrument Type" drop-down list. If no fragmentation mode is specified in the instrument type name (in brackets), the default setting is CID. If you would like to apply the same instrument configuration to all of the files in the sample, or to all of the samples in the project click on the "whole sample" button or "whole project" button, respectively.

4) To add another sample, click on the "Add Sample" button. To add a data file to Sample 2 click on the "Add data files" button. Select the instrument vendor and type from the drop-down menus unless you had previously applied the instrument configuration to the whole project in step 3.

5) To use the new Replicate Analysis feature you can set up to 3 samples to be replicates of the same experiment. To do this, select the sample node that is the first replicate, click the replicate check box and choose '1' from the drop down menu. Follow this process for the second and third replicate.

6) To delete a sample or data file, select the appropriate node (sample or data file) and click the "Delete" button. You can also change the order of the samples within a project or data files within a sample using the "Up" and "Down" buttons.

7) Click the "OK" button once all data files and samples are added to the project.

8) The project will appear in the "Project View" window. The outlined "  " symbol indicates that the file is still loading. The solid "  " symbol indicates that the file has finished loading .

4.4 Add Data to an Existing Project

Select "Add data" from the file menu or "  " from the task bar. The original project window will open. You can add more files to an existing sample or create additional samples. You will need to select the instrument vendor type. For more information on adding files/samples or setting up the instrument configuration refer to section 4.3 Create a New Project.

4.5 Open / Close a Project

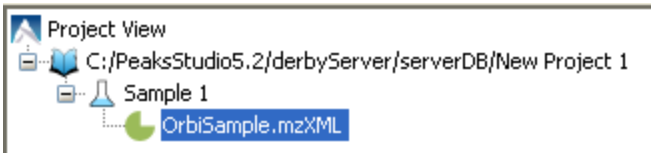
To open a stored project, select "Open Project" or "Open Recent Project" from the file menu or "  " from the task bar. To close a project that you are working on, select "Close Project" from the file menu or "  " from the task bar.

4.6 Change the Default Location of Projects

Projects are saved in the location that is listed in your "Preferences" window. To modify your preferences, select the "Preferences" toolbar icon "  " or select "Preferences" from the "Window" menu. Select "General" on the left hand side of the window. The "Default Output Directory" and "Project Folder" locations are listed in the "Default Output Directory" panel. Click on the "Browse" buttons to change the default output directory.

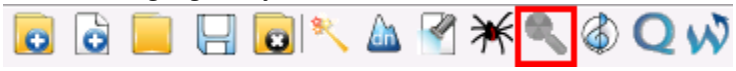
4.7 Orienting Yourself

4.7.1 Project View Panel



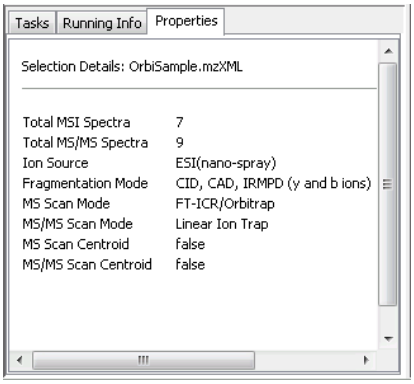
This frame appears in the upper left hand corner, displays the organization of a particular project (if applicable) or simply of a data file. Use the '+' and '-' boxes to expand and collapse the project in order to access the data file you want to analyze. Make sure the data file to be analyzed is selected. If you select a project, all of the samples will be analyzed and displayed in a single result file. If you select a sample, all of the files within that sample will be analyzed and displayed in a single result file.

Note: Any features that cannot be performed when a specific node is selected will be grayed out. In the example below, you cannot run the PTM finder when the raw file is highlighted; you must have a PEAKS Protein ID result selected in order to execute the PTM finder.

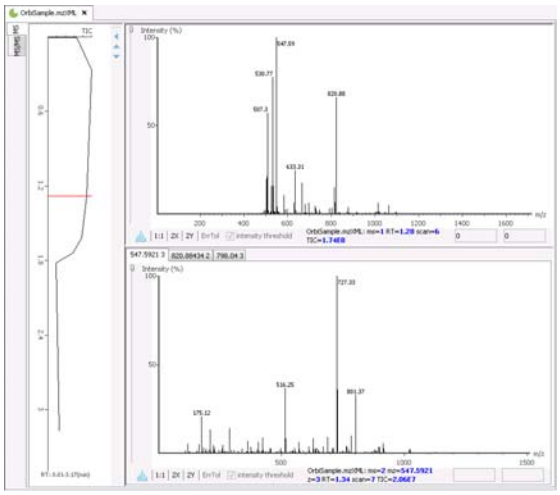


4.7.2 Properties Panel

PEAKS reads and tracks information about the experiment for use in the analysis and for future reference. Once the data file has loaded, click on the properties tab in the bottom left hand corner.



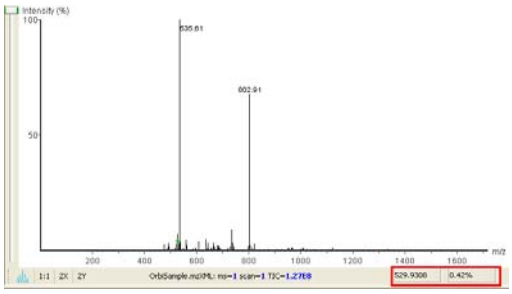
4.7.3 Raw Spectrum View



Opening the raw file in PEAKS will display the following graphs in the "Main Processing Window".

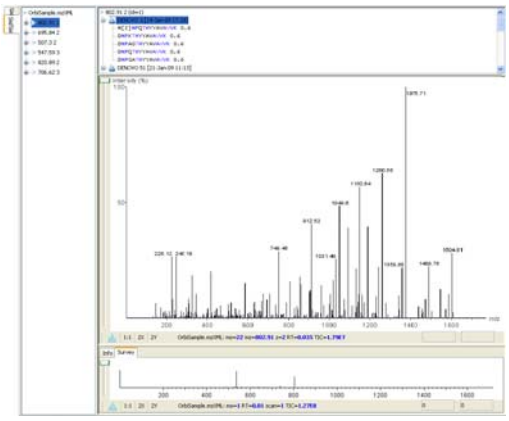
The "MS" tab is selected by default and represents the precursor scan. On the left hand side of the screen is the Total Ion Current (TIC). Depending on how the file was generated there may be simply a list of spectra and not a TIC graph. The retention time is plotted along the vertical axis. Clicking on the TIC graph will move the red line and display the MS spectra to the right of the TIC graph that corresponds to the selected retention time. Alternatively, use the up and down arrows, found on the keyboard, to move through the TIC. If the MS2 scan is available, it will be displayed below the corresponding MS1 scan.

To zoom either on the X or Y axes, select the "2X" or "2Y" buttons, respectively. To zoom in further, click and drag the mouse arrow to the side. To zoom into the intensity axis use slide the scroll bar on the left hand side. Selecting the "1:1" button will bring you back to the original image ratio where the entire spectrum is visible.



Scrolling over the spectrum will display the m/z ratio and the height/intensity (as a percentage of 100) of the particular peak under the spectrum view on the bottom right hand side (see the box highlighted in red below). Features above an intensity threshold will have their m/z displayed above them.

You can use the profile  and peak  buttons to switch the spectrum view from profile mode to peak mode and vice versa.



The MS/MS tab gives detailed information about each tandem spectrum:

Each of the spectra in the data file will be listed in the left most panel under the name of the data file. The precursor mass of the fragment is displayed followed by the charge. If data preprocessing was executed, the processed data will appear. Clicking on the '+' beside the fragment will display the raw data node as the second entry in the tree. Clicking on one of the spectra will display any results that have been generated for that spectrum in the top right hand panel.

More information can be found about the spectra under the "Info" tab. You will find information about the retention time, where to find the spectra on the TIC graph, the number of peaks and the fragmentation type (if available). You will also find an overview of the results that were found for that spectrum in the results files.

The largest panel displays the MS/MS and below it you will find the corresponding MS spectra located in the "Survey" tab.

Chapter

5

5 DATA REFINEMENT

Since mass spectrometry data most often contains noise and redundant data, it is logical to filter the data before analysis. This will increase the quality of the results, while saving time spent on database searching and de novo sequencing.

5.1 RUN DATA REFINEMENT

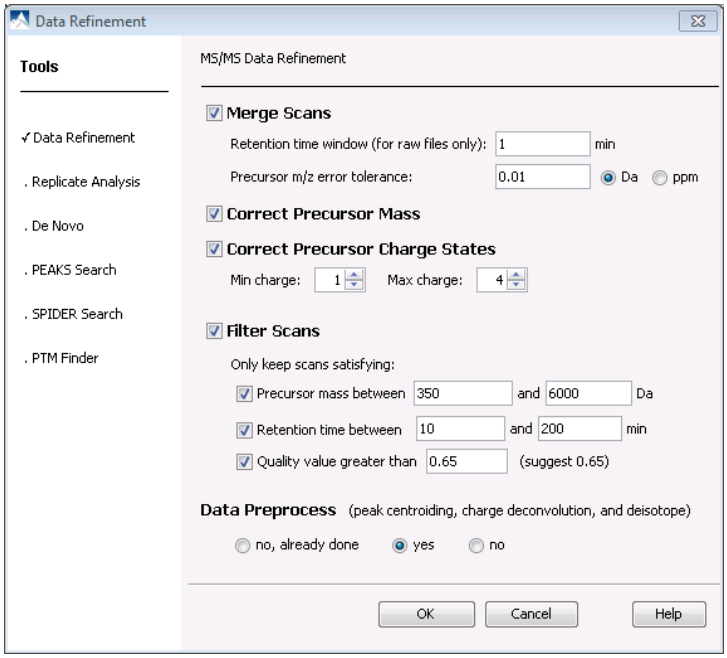
Data refinement is executed on each fraction file. You can run data refinement by selecting fraction, sample or project on the project tree and all the fraction(s) under the selected node will be refined. To begin the refinement of data from a whole MS/MS:

1) In the "Project View Frame", select a node containing the data file(s) that you wish to refine.

2) Click the Data Refine toolbar icon  Or Select "Data Refine" from the "Tools" menu.

The Data refinement options window will appear:

3) Choose the data refinement tools you wish to use by selecting the checkbox next to each refinement option. See the information below to help you decide on appropriate refinement parameters.



5.2 DATA REFINEMENT PARAMETERS

5.2.1 MERGING SCANS

In Data Dependent Acquisition (DDA) mode, a mass spectrometer will often produce several tandem (MS/MS) scans of the same peptide. To increase the intensity of real signal peaks within these scans and to reduce the size of the whole data set, it makes sense to merge MS/MS scans of the same peptide together. To avoid improper merging (of MS/MS scans of different peptides) we make sure that the measured parent ion masses of these peptides are very close and that they have similar retention times in the LC column. The units here are m/z values in Daltons or ppm. For retention time, we use the units that are recorded in the data file (usually minutes or seconds).

5.2.2 Mass Correction

In many cases, Thermo Fisher Scientific LTQ-FT and LTQ-Orbitrap instruments record a fragmented precursor mass, which is offset from the correct mono-isotopic mass of the precursor by one or more isotopes. As a result, a significant number of spectra will not be identified unless. PEAKS provides a mass correction function to accurately determine the parent charge state and mono-isotopic mass for high resolution data.

5.2.3 Precursor Charge Correction

Since a mass spectrometer measures mass-to-charge ratios, we must know the charge on a peptide before we can determine its mass. The standard method of finding the charge is to look at the spacing of the isotope ladder in the survey scan. However many Ion-Trap instruments do not have high enough resolution to allow for this type of analysis. Thus, PEAKS analyses the MS/MS data to determine whether it's charged 1+, 2+ or 3+. For data where the survey scan is available, PEAKS will examine the precursor ion's isotope distribution to confirm or correct the charge assignment. The dropdown boxes will allow you to set a range of charge values. Only spectra that fit in this range will be considered for analysis.

5.2.4 Filtering MS/MS Scans

Scans of contaminants and electrical noise should not be included in your analysis. Removing them from the data set will save time and reduce the risk of random matches to the database. PEAKS offers an effective tool for filtering these low quality MS/MS scans. The fields located under "Filter Scans" will allow you to define ranges of retention time and m/z ratio. Only peaks located within these defined ranges will be considered for further analysis. Additionally, PEAKS can examine the MS/MS spectrum to determine its quality. The quality filter is based on four characteristics: signal to noise ratio over MS/MS, number of peaks after pre-processing, sum of all peak intensities and length of the longest simple sequence tag that can be generated. You can choose a quality score threshold (a value from 0 to 1) for accepting a given scan. We recommend a quality score threshold of 0.65.

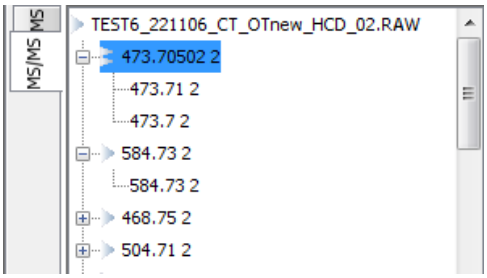
5.2.5 Preprocessing MS/MS Scans

This section deals with deconvolution (de-isotoping), centroiding and noise filtering within the MS/MS data. Preprocessing can save hard disk space or upload time but we recommend having the original spectral data available in the event you need to refer to it at a later time.

To see how your data has changed after data refinement, refer to the data properties window.

5.3 Data Preprocessing Results

To view the result of data pre-processing, click on the MS/MS tab on the spectrum view. In following example, the spectrum (m/z = 473.70502) results from raw spectra (m/z = 473.71 and m/z = 473.7).



Chapter

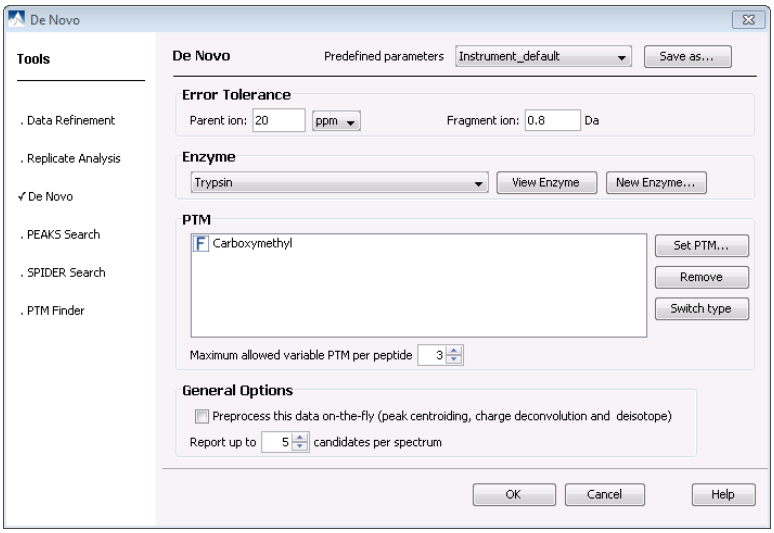
6

6 *DE NOVO* SEQUENCING

6.1 SETTING UP AUTO *DE NOVO* SEQUENCING PARAMETERS

In the "Project View Frame", select the data file(s) or project containing the spectra that you wish to sequence by auto de novo. Note that users can run a database search on a fraction or sample level by selecting the fraction node or sample node respectively.

Click the automatic de novo toolbar icon  Or Select "auto de novo" from the "Tools" menu. The auto de novo parameters dialogue window will appear.



You now have the option of changing any of the de novo parameters.

6.1.1 MASS OPTIONS

Parent mass error tolerance: This determines the degree of random and systematic experimental error on the parent/precursor ion that PEAKS will allow for during the analysis. Since the instrument used for the experiment has been previously selected, PEAKS will provide the suggested error tolerances for the selected instrument. Custom error tolerances can be defined by entering a tolerance in the textbox and choosing the desired units from the drop-down list. PEAKS provides two options for the value of mass error tolerance: Da and ppm.

Fragment mass error tolerance: This determines the degree of random and systematic experimental error on the fragment/daughter ion that PEAKS will allow for during the analysis. Again, PEAKS will provide suggested error tolerances based on the instrument selected. Custom error tolerances can be defined by entering a tolerance in the textbox and choosing the desired units from the drop-down list.

6.1.2 ENZYME OPTIONS

Enzyme: This informs PEAKS as to what type of enzyme was used to digest the sample. Choose from a drop-down list of enzymes, or if your enzyme (or combination of enzymes) is not in the list, click the "New Enzymes" button. You can then input the name of the new enzyme and define custom cleavage rules/sites and select if you would like to allow up to one end of a peptide to disobey the cleavage rule.

6.1.3 PTM Options

Selecting fixed and variable PTMs: The "PTM Options" list informs PEAKS as to what types of post-translational modifications to include in its analysis. To view additional modifications, select the "Show Unimod" box in the PTM dialog box. If a desired PTM does not appear on the list or is different than what is listed, select the "New PTM" button and the "PTM Editing" window will open. Fill in the information pertaining to your particular PTM. To select a PTM as Fixed or Variable, click the PTM and click the arrow beside the Fixed Modification or Variable Modification box. If you move over an incorrect PTM, simply drag it back to the "PTM Options" list.

Max variable PTM per peptide: To reduce uncertainty, limit PEAKS' de novo sequencing 'vocabulary' by restricting the number of variable PTM found on a peptide. Specify a number by typing it into the box.

6.1.4 General Options

Report up to (# peptides): Set how many peptide sequences PEAKS will report from its de novo sequencing analysis per spectrum.

Preprocess this data "on the fly": PEAKS has its own built-in preprocessor for removing noise, centroiding and deconvolution. Check this box to turn preprocessing on

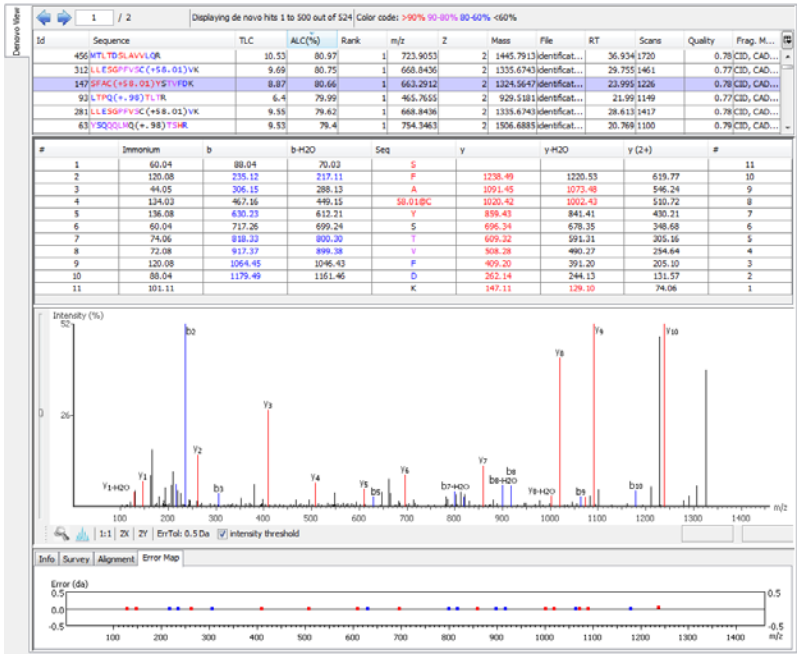
Note: If you have already pre-processed your data in the data refinement step, you do not need to do this again.

After setting up the desired parameters, you can save them for future use. Click the "Save Parameters" button at the top of the window, and define a name for these preferences for future use/reference when prompted. Any parameters that are saved will be available in the drop-down list at the top of the window. To examine the contents, select a saved parameters file and the parameters boxes will be automatically populated.

Press the "OK" button to initiate de novo sequencing.

6.2 *De novo* Sequencing Results

Once de novo sequencing is completed, the following window will appear:



6.2.1 Peptide Candidates Frame

PEAKS displays the peptide sequence candidates at the top of the screen in the "Peptide Candidates Frame". You can sort the results by clicking on any of the titles of the columns. For example, to sort the peptide sequence candidates by ID click on "ID". Note that all of the peptides that have the same ID have the same mass, charge, retention time and quality score (Refer to section 5.2.4 Filtering MS/MS Scans for more information on how the quality score is generated).

The following table describes the contents of the columns in the "Peptide Candidates Frame":

ID	A unique identifier for the MS/MS spectrum. This differs from a scan number since we may have merged several scans together.
Sequence	The sequence of the peptide (including modifications if present) as determined by de novo sequencing.
TLC	Total local confidence (the confidence that we have in the peptide sequence). It is calculated by adding the positional confidence for each amino acid in the peptide sequence.
ALC	Average local confidence (the confidence that we have in the peptide sequence). It is calculated by adding the positional confidence for each amino acid in the peptide sequence and dividing by the total number of amino acids.
Rank	The sequences for a particular spectrum (ID) as sorted by score (TLC).
m/z	The measured mass/charge value, in Daltons, for the peptide.
Z	The calculated charge value for the peptide.
Mass	Calculated using the measured m/z and calculated z, we use this as the experimental mass of the peptide.
File	The name of the file.
RT	Retention time (elution time) for the peptide as recorded in the scan header.
Scans	The scan number.
Quality	A value from 0 to 1 estimated from the spectrum to refer to spectrum quality. Attributes like signal to noise, total intensity, and spectrum tagging are used.
Frag. Mode	Mode that the fragmentation step was performed in.

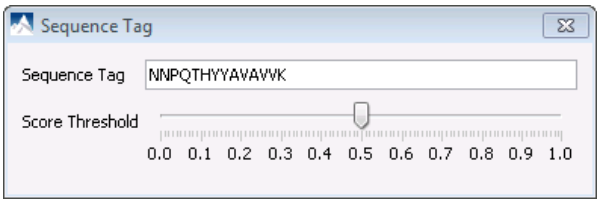
The columns themselves are customizable. To change the columns that are displayed click on the "  " button.

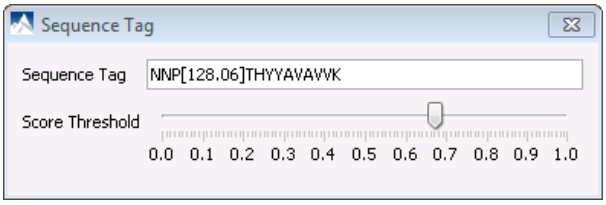
Confidence Scores

Next to the proposed sequence candidates, the auto de novo "Total Local Confidence" (TLC) and "Average Local Confidence" (ALC) confidence scores are shown. The confidence scores for each amino acid (that is, confidence that the correct residue in each position has been identified) are represented by color coding. 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 a low confidence (less than 60%). For more detailed positional confidence, place the cursor over the sequence of interest and right click "Show Positional Confidence". A "Position Confidence Table" will appear, showing the confidence that each amino acid/pair of amino acids are correct.

Sequence Tags

Right click on a peptide in the "Peptide Candidates Frame" and select "Show Sequence Tag". If the score threshold is set at 0.0, all of the amino acids in the peptide sequence will be displayed.





Increasing the "Score Threshold" will display a mass in square brackets if the amino acids do not satisfy the score threshold.

Modifications

Consider the following sequence: H(+15.99)ATVFDNLNPPEDR

The +15.99 in brackets refers to a position where a modification may have occurred. If you forget what possible modifications you specified before running de novo, click the "Properties" tab.

The fixed modification is set to Oxidation HW 15.99. In the sequence above, the modification has been made to the H. The colors assigned to the (+15.99) follow the same confidence scores as the amino acids themselves.

Refer to the above section on "Confidence Scores" for more information on color coding.

Denovo details: Trypsin(f: Oxidation HW)

Parent Mass Error Tolerance	0.1 Da
Fragment Mass Error Tolerance	0.6 Da
Enzyme	Trypsin
PTMS	
Fixed Modifications	
Name	Delta
Oxidation HW	15.994915
Max variable PTM per peptide	2
Report # peptides	5
Preprocess data	false
Data Refine dependencies	1

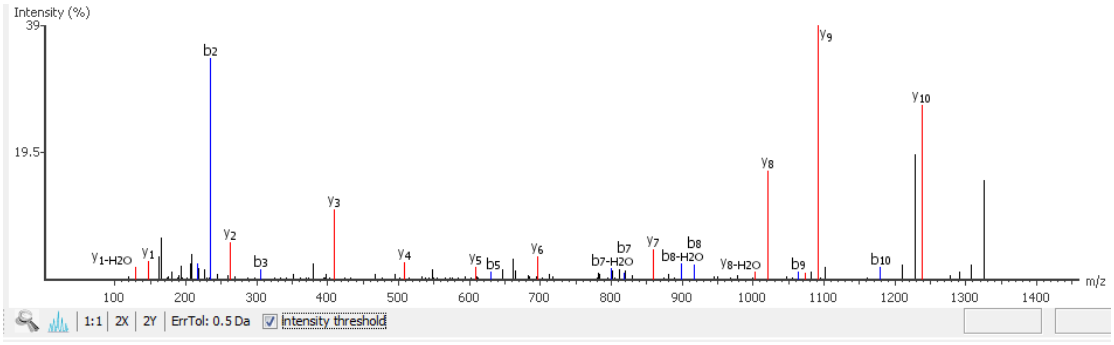
6.2.2 ION TABLE FRAME

#	Immonium	b	b-H2O	a	c	Seq	y	y-H2O	z	z'	y (2+)	#
1	88.04	116.03	98.02	88.04	133.06	D						14
2	87.06	230.08	212.07	202.08	247.10	N	1489.76	1471.77	1472.74	1473.75	745.39	13
3	70.07	327.17	309.15	299.14	344.16	P	1375.71	1357.71	1358.71	1359.66	688.37	12
4	101.07	455.19	437.18	427.19	472.22	Q	1278.69	1260.56	1261.66	1262.66	639.84	11
5	74.06	556.24	538.23	528.24	573.26	T	1150.64	1132.62	1133.60	1134.60	575.81	10
6	110.07	693.30	675.28	665.29	710.32	H	1049.60	1031.49	1032.55	1033.55	525.29	9
7	136.08	856.36	838.35	828.36	873.39	Y	912.52	894.51	895.49	896.49	456.76	8
8	136.08	1019.43	1001.43	991.43	1036.45	Y	749.46	731.45	732.43	733.43	375.23	7
9	44.05	1090.47	1072.47	1062.46	1107.49	A	586.44	568.29	569.32	570.37	293.70	6
10	72.08	1189.53	1171.51	1161.51	1206.55	V	515.35	497.34	498.33	499.33	258.18	5
11	44.05	1260.56	1242.56	1232.57	1277.59	A	416.29	398.28	399.26	400.26	208.64	4

The "Ion Table" shows the proposed ions with their corresponding masses. To add additional ions to the ion table, refer to section 16.1.4 Ion Editor Preferences.

If an ion is found in the corresponding spectrum, it must first pass two criteria before being displayed in a specific color (blue for N-terminal ions and red for C-terminal ions). It must be found within the mass error tolerance chosen by the user and must have an intensity of greater than 2% of the ion with the greatest intensity.

6.2.3 SPECTRUM VIEW FRAME



The “Spectrum View Frame” is found below the “Ion Table” and displays a graphical representation of the spectrum. The peptide that corresponds to the spectrum in the “Spectrum View Frame” is displayed in the “Input Sequence” box. Use the drop-down to select other peptides that have the same ID.

Scrolling over the spectrum will display a “tooltip” in the new window that will display the m/z ratio and the height/ intensity (as a percentage of 100) of that particular peak. Both the m/z ratio and the height of the peak can also found under the spectrum view on the right hand side.

The “Intensity threshold” check box provides an option to display lower intensity peaks in the “Spectrum View Frame”

More options are provide for view the spectrum.

- Ion Display: Immonium, b, b-H2O, y, y-H2O, y (2+), Internal Ion
- m/z Display: show m/z with ions, show m/z on peaks without ions
- Intensity Display: High, Medium, Low

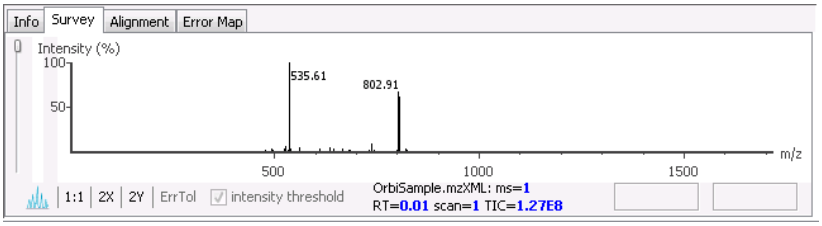
To zoom either on the X or Y axes, select the “2X” or “2Y” buttons, respectively and then use your mouse to move navigate the graph. Selecting the “1:1” button will restore settings the default view which displays the entire spectrum.

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

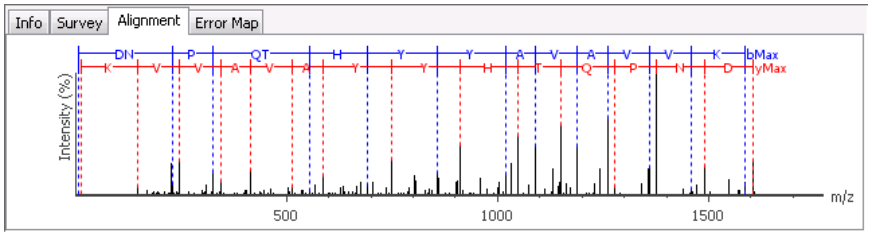
You can use the profile  and peak  buttons to switch the spectrum view from profile mode to peak mode and vice versa. The scrollbar on the left allows you to zoom into the intensity axis of the peaks.

6.2.4 SURVEY SCAN

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

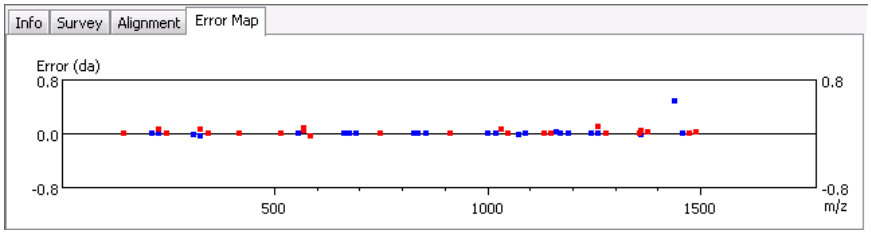


6.2.5 Spectrum Alignment Frame



The "Alignment" tab displays the entire spectrum and is used as a tool to help us navigate the spectrum view frame. The blue bar along the horizontal m/z axis of the alignment view indicates the range of the spectrum view in the Spectrum View Frame. This frame displays how the proposed ions align with the spectrum. By default, the Spectrum Alignment Frame displays b-ions and y-ions. The b ions are shown right to left in blue, while the y ions are shown left to right in red.

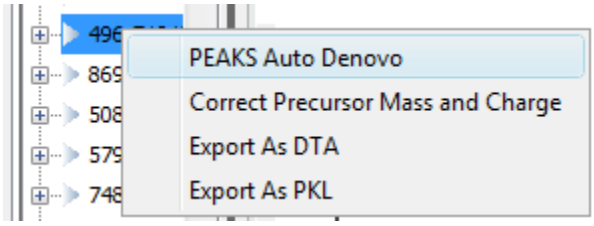
6.2.6 Error Map



Click on the "Error Map" tab. The m/z ratio is displayed on the x-axis and the error is listed on the y-axis in Daltons. The "Error Plot" displays the confidence that is assigned to each ion. The most confident results lie on the centerline. Clicking a cell or column in the Ion Table highlights the corresponding points on the error plot and corresponding peaks on the spectrum.

6.3 Auto de novo on a single spectrum.

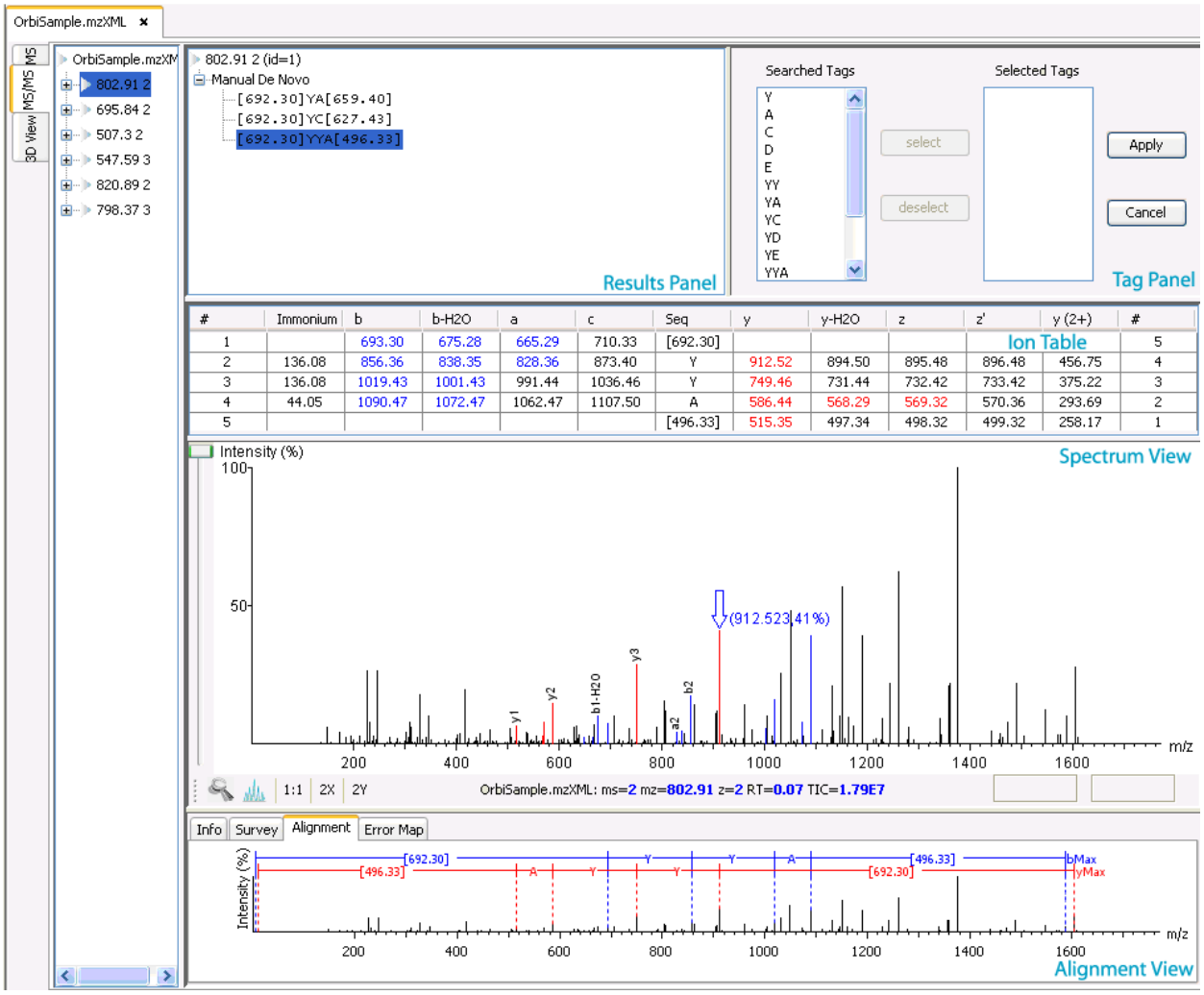
Auto de novo sequencing on a single spectra can be performed on the MS/MS view as displayed below.



6.4 Manual *de novo* Sequencing

PEAKS 5.2 provides a set of tools to help you manually sequence a peptide, using graphic cues from the spectrum.

6.4.1 MANUAL DE NOVO GRAPHICAL USER INTERFACE



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

The panels are briefly described below:

The "**Result Panel**" shows all the sequencing results. The results of manual de novo are listed in the sub-tree with root "Manual de novo".

The "**Spectrum View Panel**" shows a graphical representation of the spectrum, the peaks in the spectrum, the user-selected peaks and assigned ions. Users can pick a peak on the panel and assign ions or tags to it in Manual de novo.

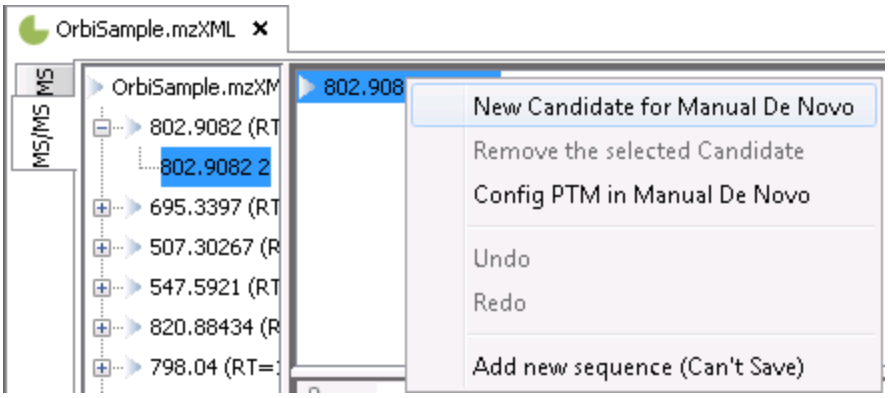
The "**Alignment View Panel**" shows how the proposed ions assigned in manual de novo align with the spectrum. By default, the "Spectrum Alignment Frame" displays b-ion and y-ion. The b ions are shown right to left in blue, while the y ions are shown left to right in red.

The "**Ion Table Panel**" shows the proposed ions with their corresponding masses. The default Ion Table will display immonium, b, b-H2O, a, c, y, y-H2O, z, z' and y(2+) ions.

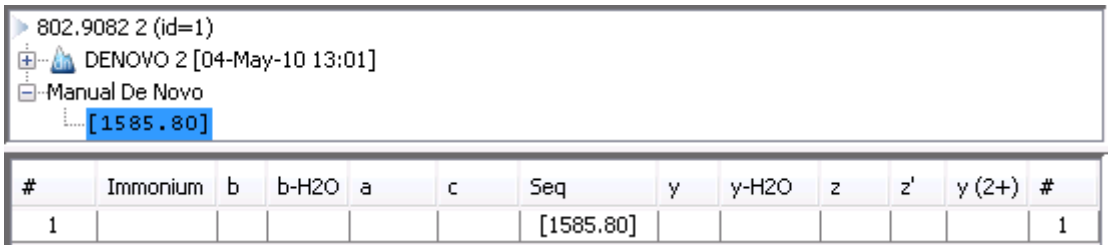
The "**Tag Panel**" will appear when you search tags or ions in the spectrum. You can select the tags in the list using the "Select" button. Clicking "Apply" will add the selected tags to the sequence candidate.

To create a new peptide candidate for sequencing:

Highlight the m/z value in the "Results Panel" and right click. This will bring up a pop-up menu.



Select "New Candidate for Manual De Novo" from the pop-up menu. A new candidate will be created under the 'Manual de novo' heading. The new candidate will not have been sequenced, so it will be represented by the mass of the peptide less the mass of water (see an example below).



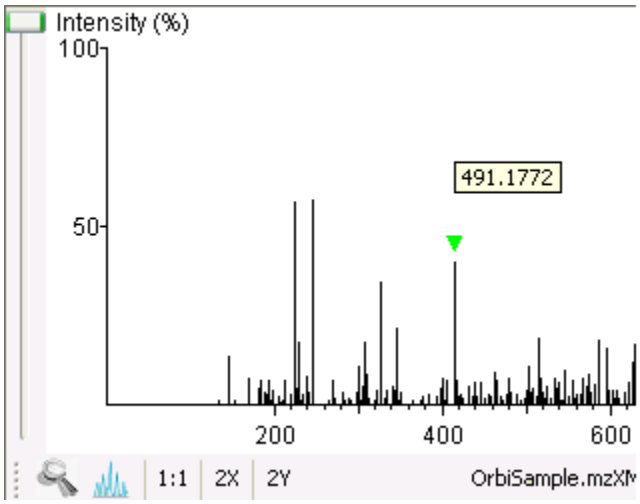
#	Immonium	b	b-H2O	a	c	Seq	y	y-H2O	z	z'	y (2+)	#
1						[1585.80]						1

Note: The popup menu will not be accessible if you have highlighted any of the results in the "Results Panel".

6.4.2 Manual de novo Operations

Note: All operations occur in the Spectrum View Frame of the Main Processing Window.

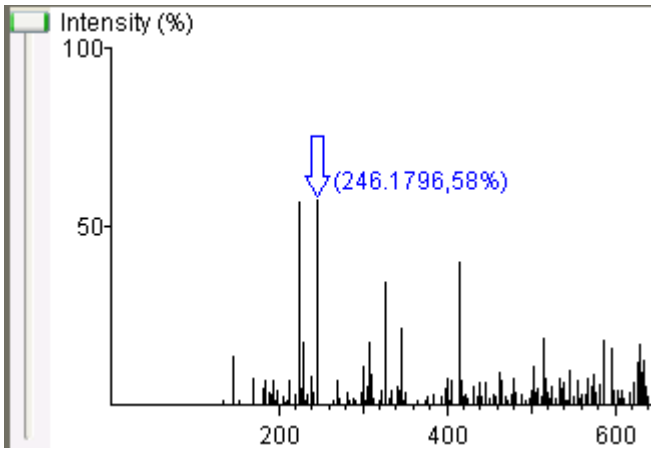
When the mouse cursor is placed in the Spectrum View Frame, a green (by default) arrow follows the movement of the mouse. This is the Position Bar and it is used as a cursor for all manual de novo operations. The cursor's position on the m/z scale is enumerated on the top of the Position Bar.



Selecting a peak

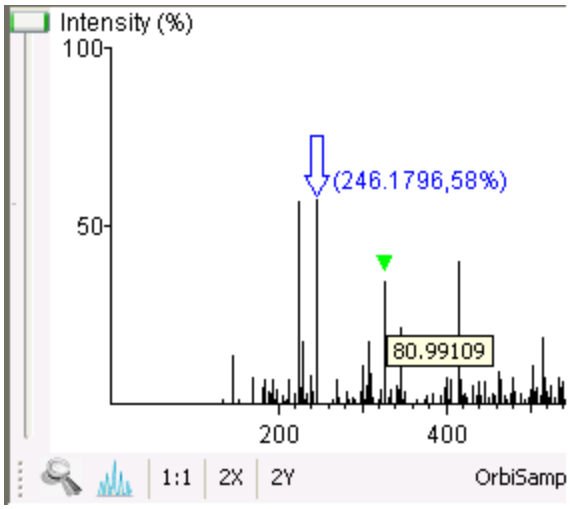
To select a peak, simply click on it. A blue (by default) arrow, called the Freeze Bar, indicates the selected peak.

Alternatively an ion peak can be selected by clicking on its corresponding cell in the Ion Table.



Measure the m/z difference between two PEAKS

Select a peak (blue arrow) with the Freeze Bar and move the mouse to the left or right. Hold the Position Bar (green arrow) above another peak. The number in the Position Bar is the difference between the two PEAKS (in the example below the difference is 80.99109).



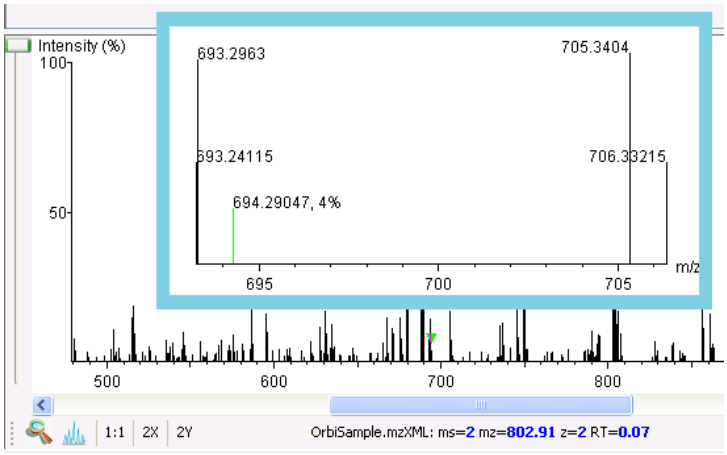
Deselect a peak

Double click anywhere in the Spectrum View Frame to deselect a peak.

Zoom in on part of the spectrum

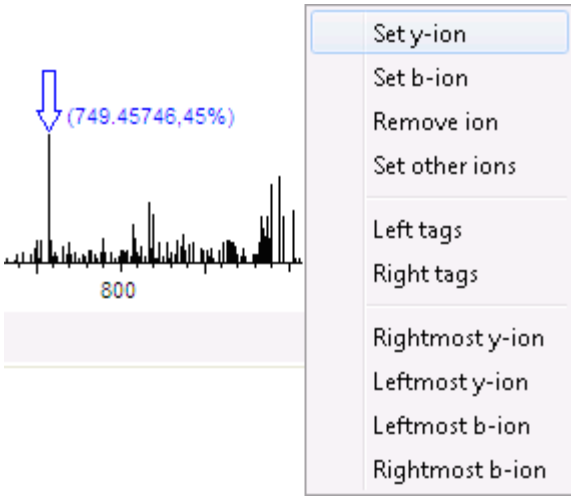
In the Spectrum View Frame, click and drag the mouse horizontally. The selected area will be enhanced and shown in the Spectrum View Frame. Click on the magnifying tool "  " to display a panel with the magnified spectrum. The peak which is selected by the Position bar will be highlighted in the zoomed panel in green.

Click on the magnifying tool again to return to the default view.



Setting/removing ions to/from a peak

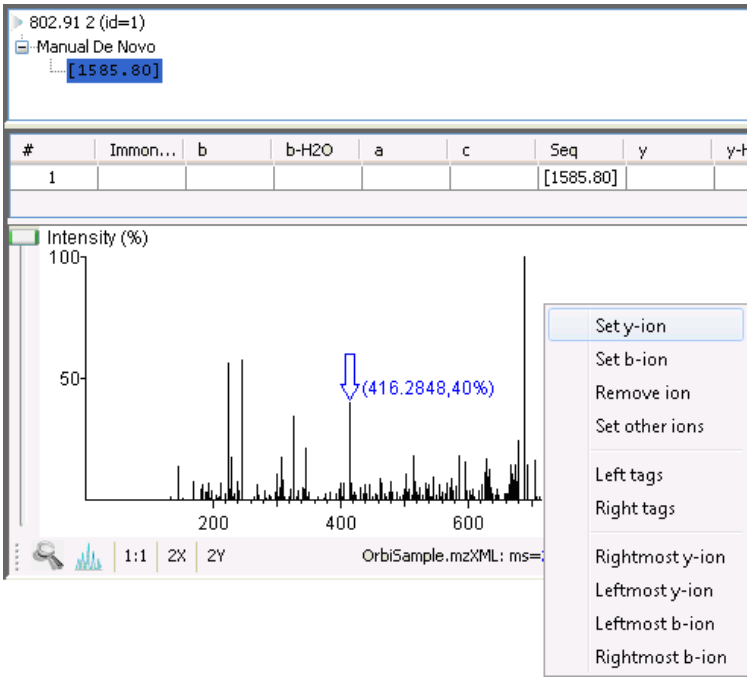
Select a peak, and then right click the mouse anywhere in the Spectrum View Frame. Select “Set y-ion” from the pop-up menu to designate the peak as a y ion, or “Set b-ion” from the pop-up menu to designate the peak as a b ion. Click on “Remove ion” to remove the ion that you have previously set.



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

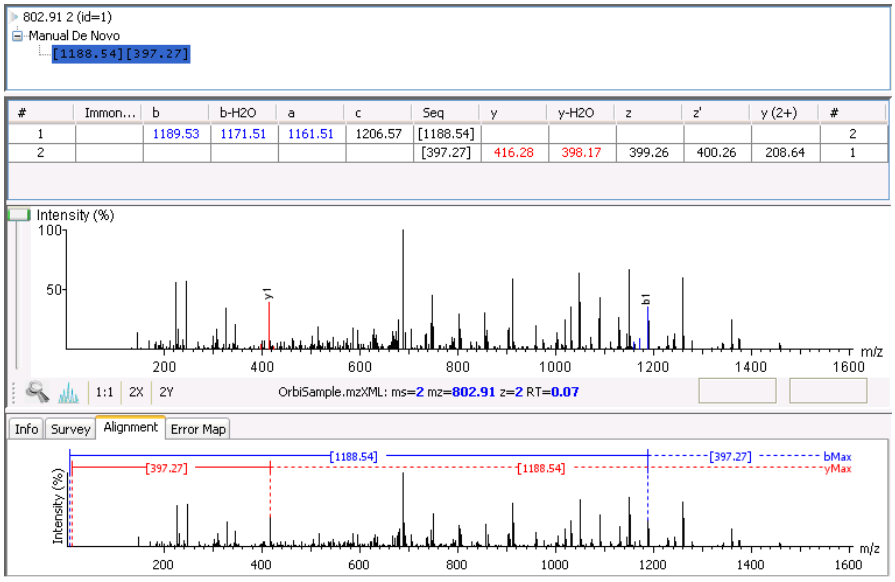
After setting an ion, both the alignment view and the peptide sequence candidate name (as displayed in the peptide candidate frame) will change to reflect the mass remaining to be sequenced on either side of the ion.

In the example below, the selected peak at 416.2848 m/z was designated as a y-ion.



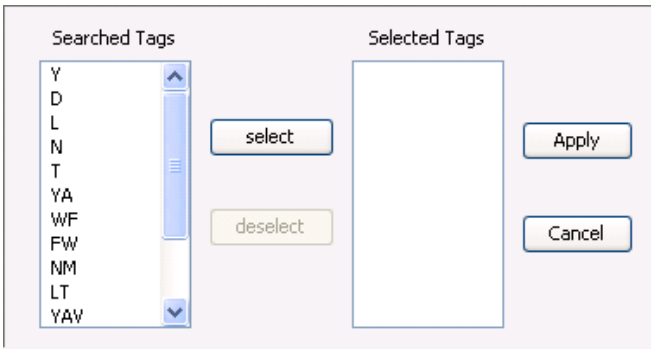
Note that the manual de novo candidate information is updated in the results panel, ion table and alignment view panel. The selected ions are also highlighted in the spectrum view panel.

After setting two ions, PEAKS 5.2 will estimate the residue found between them (if a residue corresponds closely to the mass difference). The peptide sequence candidate name (as displayed in the peptide candidate frame) will change to show the residue and the mass remaining to be sequenced on either side of the residue.



Searching the left or right side of the spectrum for the first/last y or b ion (Search a sequence tag)

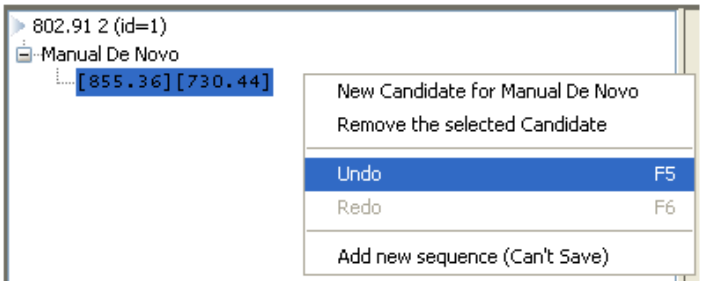
Select a peak, and then right click the mouse anywhere in the Spectrum View Frame to trigger the popup menu. From the menu, select either "Left tags" or "Right tags". PEAKS will select the appropriate terminal tags and show them in the "Tag Panel" frame (see below).



You can test the suitability of a tag by highlighting it in the "Searched Tags" panel; the corresponding information for the tag will be shown in the Spectrum View panel, the Ion table and the Alignment frame panel. You can insert one or more tags by highlighting the desired tags, clicking "Select" to move them into the "Selected Tags" panel and then clicking "Apply". Press the "Cancel" button at any time to exit the search and discard any changes.

Undoing an edit

If you have made an error in your sequencing it is possible to undo the change. With the Peptide candidate still selected in the "Results panel", right click the mouse and select "Undo" to return to the previous peptide sequence. You can use this button multiple times to return to earlier stages in your edit.



Redoing an edit

With the Peptide candidate still selected in the "Results panel", right click the mouse and select "Redo" if you have undone one too many changes. You can click this button multiple times to return to later stages in your edit.

Chapter

7

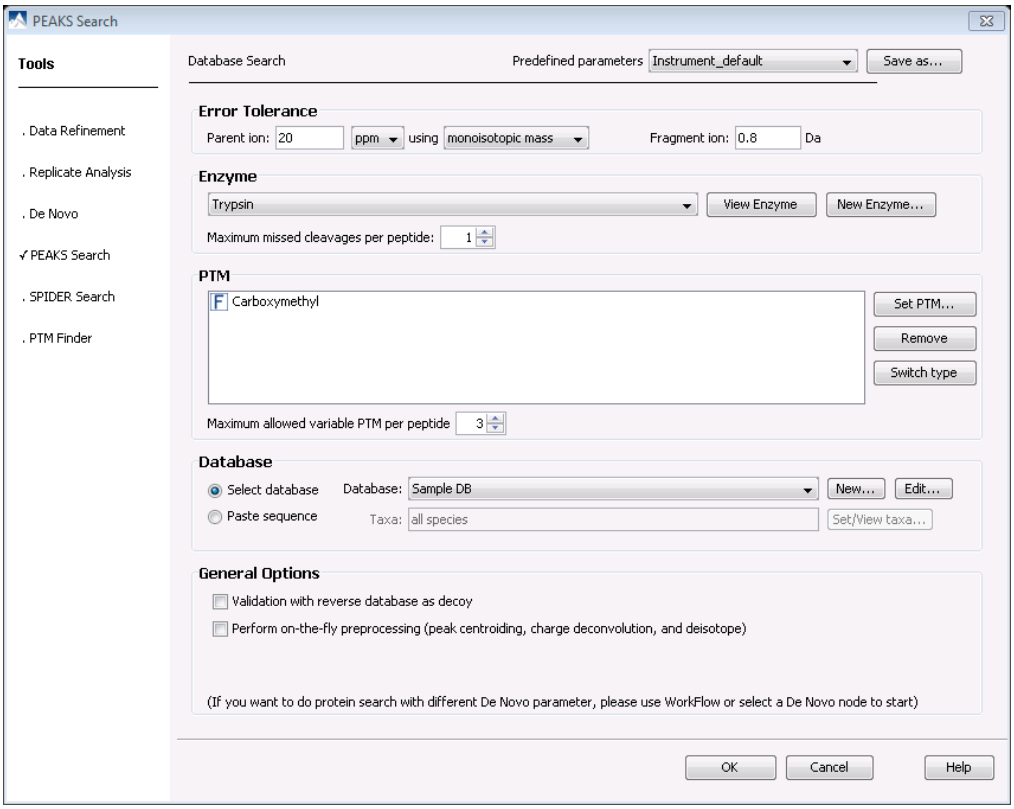
7 DATABASE SEARCH

7.1 SETTING UP PROTEIN IDENTIFICATION PARAMETERS

In the “Project View Frame”, select the data file(s) or project containing the spectra that you wish to identify using database search. Note that we can run database search on a fraction or sample level by selecting the fraction node or sample node respectively.

Click the PEAKS Protein ID toolbar icon  Or Select “PEAKS Protein ID” from the “Tools” menu. The Protein Identification Parameters dialogue window will appear:

You now have the option to change any of the protein identification search parameters.



7.1.1 Mass Options

Parent and Fragment mass error tolerance: refer to section 6.1.1 Mass Options.

Precursor mass search type: If the precursor mass is monoisotopic value, select monoisotopic from the dropdown list next to “using". Otherwise, select average.

7.1.2 Enzyme Options

Enzyme: refer to section 6.1.2 Enzyme Options.

Maximum missed cleavages: determines the greatest number of missed cleavages to allow.

7.1.3 PTM Options

PTM options: refer to section 6.1.3 PTM Options.

Max variable PTM per peptide: refer to section 6.1.3 PTM Options.

7.1.4 Database Options

Database: From the drop-down list select one of the FASTA databases configured in PEAKS. To edit an existing database, click the “Edit…” button. If the desired database is not in this list, click the “New…” button.

Note that you can also set up a new database in the “Database Configuration” window. The configuration window is the only place where you have the option deleting databases which you have created. For more information on setting up new databases refer to section 16.2.3 Database Configuration.

Set/View Taxa: This list displays the taxa you have chosen for your search. If the database selected has taxon information available, you can click on the “Set/View taxa…” button. Otherwise, the whole database will be searched. The selections correspond to established hierarchy -- i.e. selecting 'Mammalia' will search all of 'horse, cow, rat, mouse, human, etc.

Paste sequence: If you already know the sequence of the protein(s) you are looking for, select “Paste sequence” and paste the sequence in the space provided in fasta format. Alternatively, if you want search the same sequence regularly, it is recommended to simply create a small text file and configure it as a database for PEAKS.

7.1.5 General Options

Validate with reverse database as decoy: This executes a Protein ID search on a database that is physically reversed. The result gives an indication of how many random matches we should expect.

Perform on-the-fly processing: PEAKS has its own built-in processor for removing noise, centroiding and deconvolution. Check this box to turn processing on.

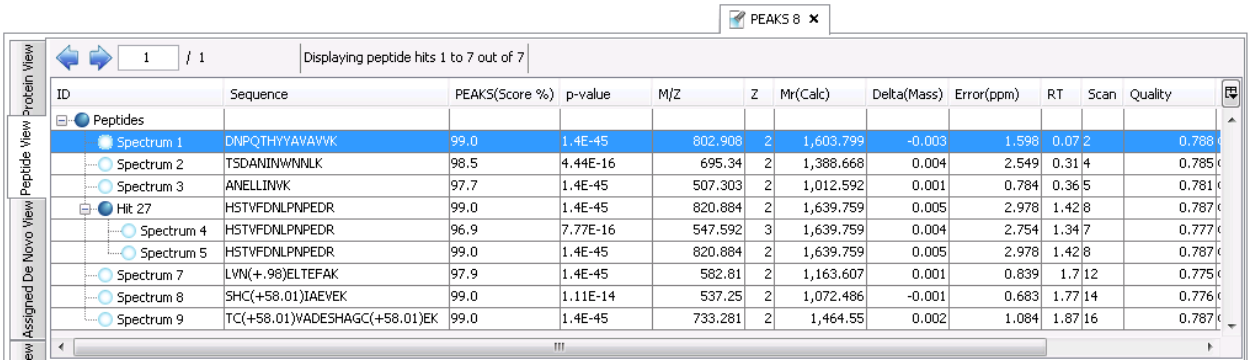
After setting up the desired parameters, we can save them for future use. Click the "Save Parameters" button, and define name for these preferences for future use/reference, when prompted. Any parameters that you save will be available in the drop-down list at the top of the window. To examine the contents, select a saved parameters file and parameters boxes will be populated.

Press the “OK” button. If you have already performed de novo sequencing, the database search will commence automatically. If you have not previously performed de novo sequencing, the auto de novo process will appear first in the task queue. Once de novo sequencing is finished the database search will begin.

7.2 PROTEIN IDENTIFICATION RESULTS

7.2.1 PEPTIDE VIEW

Once PEAKS has finished searching the database, the "Peptide View" window will open by default:



ID	Sequence	PEAKS(Score %)	p-value	M/Z	Z	Mr(Calc)	Delta(Mass)	Error(ppm)	RT	Scan	Quality
Peptides											
Spectrum 1	DNPQTHYYAVAVVK	99.0	1.4E-45	802.908	2	1,603.799	-0.003	1.598	0.07	2	0.788
Spectrum 2	TSDANINWNNLK	98.5	4.44E-16	695.34	2	1,388.668	0.004	2.549	0.31	4	0.785
Spectrum 3	ANELLINVK	97.7	1.4E-45	507.303	2	1,012.592	0.001	0.784	0.36	5	0.781
Hit 27	HSTVFDNLPNPEDR	99.0	1.4E-45	820.884	2	1,639.759	0.005	2.978	1.42	8	0.787
Spectrum 4	HSTVFDNLPNPEDR	96.9	7.77E-16	547.592	3	1,639.759	0.004	2.754	1.34	7	0.777
Spectrum 5	HSTVFDNLPNPEDR	99.0	1.4E-45	820.884	2	1,639.759	0.005	2.978	1.42	8	0.787
Spectrum 7	LVN(+.98)ELTEFAK	97.9	1.4E-45	582.81	2	1,163.607	0.001	0.839	1.7	12	0.775
Spectrum 8	SHC(+58.01)IAEVEK	99.0	1.11E-14	537.25	2	1,072.486	-0.001	0.683	1.77	14	0.776
Spectrum 9	TC(+58.01)VADESHAGC(+58.01)EK	99.0	1.4E-45	733.281	2	1,464.55	0.002	1.084	1.87	16	0.787

The "Peptide View" window summarizes the results for each MS/MS spectrum. All peptides that match to each spectrum are displayed. By default the spectra are listed by ID in the "ID" column with the corresponding peptide sequence in the "Sequence column". In certain cases, one peptide can correspond to more than one spectrum. These spectra are then listed in the "ID" column under a heading entitled "Hit". Click on "+" to expand the view to see all of the spectra that can be matched by the same peptide.

The table below describes the contents of the columns in the "Peptide View Window":

ID	A unique identifier for the MS/MS spectrum. This differs from a scan number since we may have merged several scans together.
Sequence	The amino acid sequence of the peptide. PTMs are listed in [square brackets].
PEAKS(Score %)	PEAKS' probability score.
p-value	The p-value of the peptide
M/Z	The measured mass/charge value, in Daltons, for the peptide.
Z	The calculated charge value for the peptide
Mr(Calc)	The sum of the theoretical mass of the residues that form the identified peptide sequence from the database.
Delta(Mass)	The difference between Mr(Calc) and Mass, in Daltons.
Error(ppm)	The difference between Mr(Calc) and Mass, ppm.
File	The name of the file.
RT	Retention time (elution time) for the peptide as recorded in the scan header.

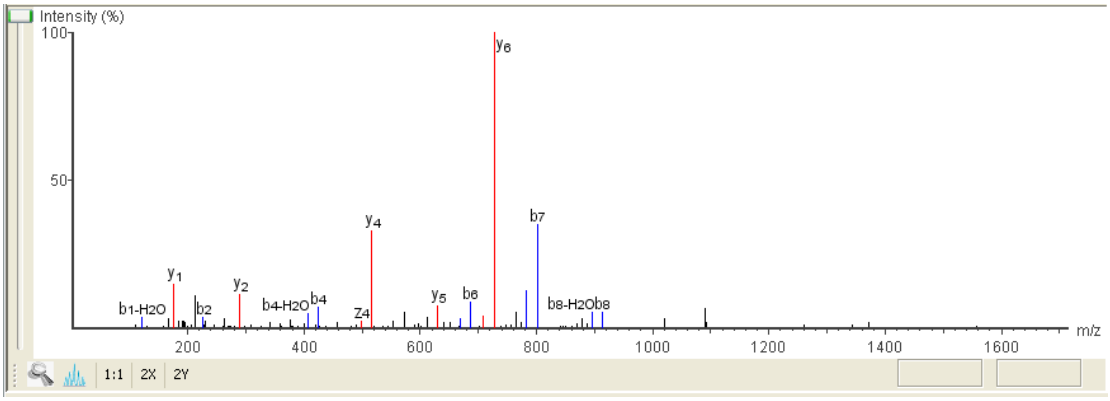
Scan	The scan number.
Quality	A value from 0 to 1 estimated from the spectrum to refer to spectrum quality. Attributes like signal to noise, total intensity, and spectrum tagging are used.
Frag. Mode	Mode that the fragmentation step was performed in.
# Hits	Number of hits corresponding to peptides with multiple spectra.

The columns themselves can be customized. To change the columns displayed click on the "  " button.

Peptide Alignment

Click on the "Peptide Align" window. Similar to the de novo results window, the "Ion Table" displays proposed ions with their corresponding masses. Below the "Ion Table" is the 'Spectrum View Frame" and below this is the "Error Map" which displays confidence assigned to each ion. The most confident results lie on the centerline. Clicking a cell or column in the Ion Table highlights the corresponding points on the error plot and corresponding peaks on the spectrum.

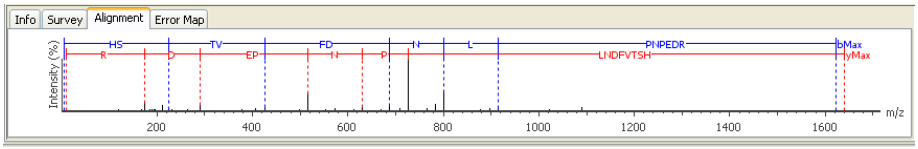
The "Spectrum View Frame" located below the "Ion Table" displays a graphical representation of the spectrum. The peptide that corresponds to the spectrum in the "Spectrum View Frame" is displayed in the "Input Sequence" box. Note that this is a drop-down menu so that you can select other peptides that have the same ID (if applicable). Scrolling over the spectrum will display a "tooltip" that will display the m/z ratio and the height/ intensity (as a percentage of 100) of that particular peak. Both the m/z ratio and the height of the peak can also found under the spectrum view on the right hand side.



To zoom either onto the X or Y axes, select the "2X" or "2Y" buttons, respectively. Then use your mouse to navigate the graph. Selecting the "1:1" button will restore settings the default view which displays the entire spectrum.

You can use the profile  and peak  buttons to switch the spectrum view from profile mode to peak mode and vice versa. The scrollbar on the left allows you to zoom into the intensity axis of the peaks.

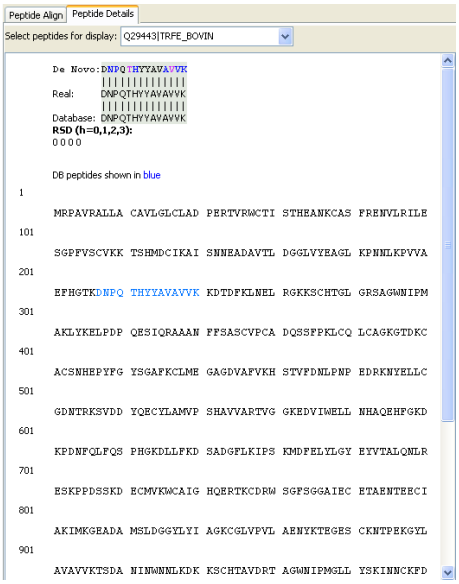
Clicking the "Alignment" tab will display the "Spectrum Alignment Frame". This frame will always display the entire spectrum and is used as a tool to help us navigate the spectrum view frame. The blue bar



along the horizontal m/z axis of the alignment view indicates the range of the spectrum view in the Spectrum View Frame. This frame displays how the proposed ions align with the spectrum. By default, the Spectrum Alignment Frame displays b-ions and y-ions. The b ions are shown right to left in blue, while the y ions are shown left to right in red.

Clicking on the "Survey" tab will display the corresponding precursor ion spectrum. The buttons that appear in this section are identical to those explained above in the "Spectrum View Frame" section. Click on the "Error Map" tab. The m/z ratio is displayed on the x-axis and the error is listed on the y-axis in Daltons. The "Error Plot" displays the confidence that is assigned to each ion. The most confident results lie on the centerline. Clicking a cell or column in the Ion Table highlights the corresponding points on the error plot and corresponding peaks on the spectrum.

Peptide Details



At the top of the "Peptide Details" frame is the accession number of the protein that corresponds to the peptide that you selected in the "Peptide View" window. If more than one protein matches a single peptide, you will be able to select and view these additional proteins using the drop-down menu.

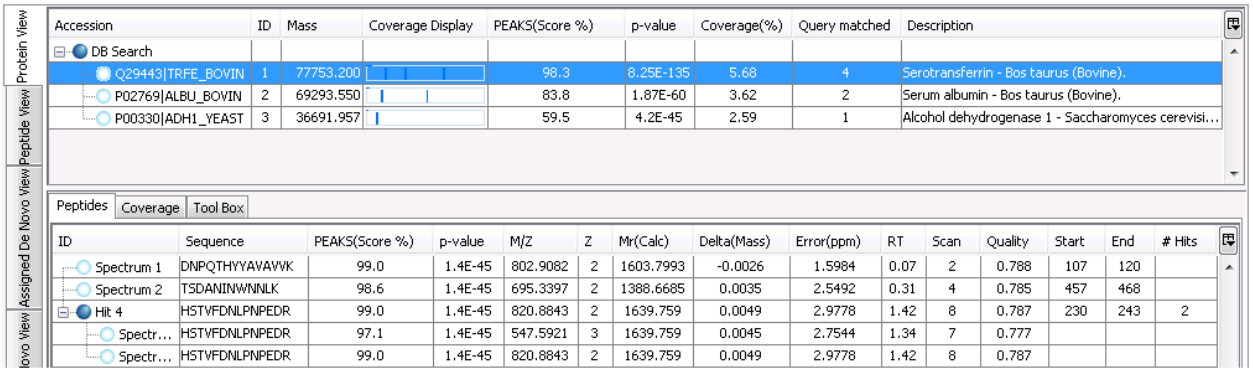
Below this you will see a simple alignment between the original de novo sequence for this spectrum (if available), the peptide found in the database and the reconstructed sequence. Letters on a green background, and with vertical bars, indicate agreement. Color codes on the de novo sequence letters still indicate positional confidence.

Click on the "Peptide Details" tab. The following window will appear:

At the bottom of the window you will see the sequence of the selected protein and in blue you will see where the selected peptide matches the protein. The darker the blue, the more confident the match. The matched peptides will be shown in red if you have performed a SPIDER search which is discussed in the next chapter.

7.2.2 PROTEIN VIEW

Clicking on the "Protein View" tab will produce the following window:



The "Protein View" collects all the peptide identifications together, summarizing which proteins were present in the sample and groups homologous proteins together. The same information is displayed in the Peptide View as in this Protein View; however, the results here are organized to best enable us to evaluate the results at the protein level.

This view is helpful when building a summary that can be sent to a customer/collaborator. Refer to 15 Exporting Data/Reports and Printing for more details on exporting whole files or proteins of interest to an Excel file.

Index

The top section of this view (shown above) behaves like an index, listing each protein found in the sample. Very similar proteins, containing the same set or a subset of matched peptides, are clustered together. To expand and collapse the full list of proteins within each cluster click the '+' or '-' sign respectively.

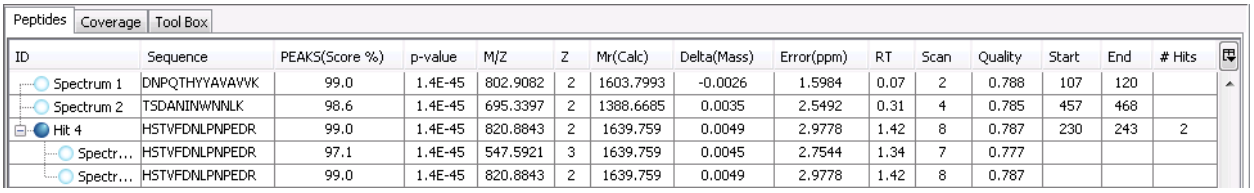
The table below describes the contents of the columns in the index:

Accession	The GI, accession or other unique identifier for this protein as recorded in the database that was searched.
Mass	The calculated mass of this protein
Coverage Display	A graphical coverage map. Blue areas represent parts of the sequence that have been explained by the identified peptides.
PEAKS (Score %)	A value from 1 to 99 representing the confidence we have in this protein identification – calculated from the confidence on the ten best peptide hits for this protein, and normalized against the other identified proteins.
Coverage (%)	The number of amino acids in the protein sequence that have been explained by the identified peptides. Expressed as a percentage of the total length of the protein.
Query Matched	The number of spectra explained by matching to a peptide from this protein.
Description	The part of the protein's header information as parsed from the database. Usually it contains the name of the protein.
Marked	A multi-function checkbox. By default unchecked, but we can use this to select proteins for export or multiple sequence alignment.

The columns themselves can be customized. Right click anywhere in the report and choose "Toggle Column" from the pop-up menu. The sub-menu that appears shows a checkmark in each of the columns that are currently displayed. Click any of them to show or hide a column. These settings will apply to all your reports.

Peptides

The "Peptides" tab displays information about the peptides that match the selected protein. This list is identical to the "Peptide View" panel. Please refer to section 7.2.1 Peptide View for more details.



ID	Sequence	PEAKS(Score %)	p-value	M/Z	Z	Mr(Calc)	Delta(Mass)	Error(ppm)	RT	Scan	Quality	Start	End	# Hits
Spectrum 1	DNPQTHYYAVAVVK	99.0	1.4E-45	802.9082	2	1603.7993	-0.0026	1.5984	0.07	2	0.788	107	120	
Spectrum 2	TSDANINWNNLK	98.6	1.4E-45	695.3397	2	1388.6685	0.0035	2.5492	0.31	4	0.785	457	468	
Hit 4	HSTVFDNLPNPEDR	99.0	1.4E-45	820.8843	2	1639.759	0.0049	2.9778	1.42	8	0.787	230	243	2
Spectr...	HSTVFDNLPNPEDR	97.1	1.4E-45	547.5921	3	1639.759	0.0045	2.7544	1.34	7	0.777			
Spectr...	HSTVFDNLPNPEDR	99.0	1.4E-45	820.8843	2	1639.759	0.0049	2.9778	1.42	8	0.787			

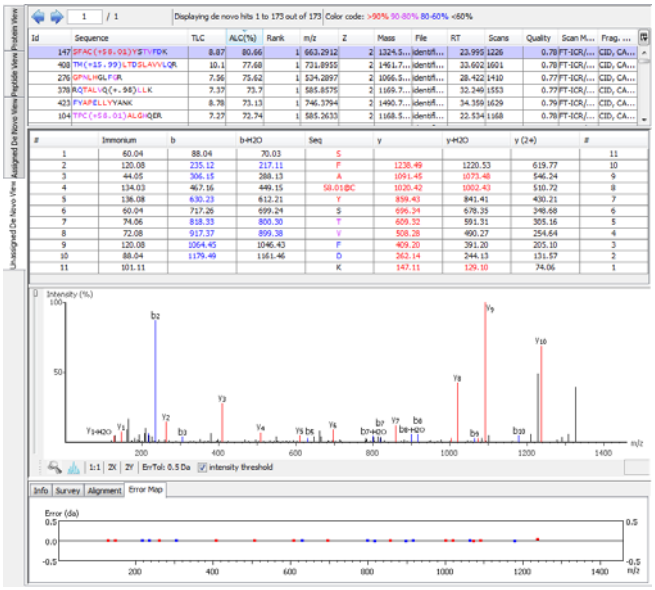
Coverage

Beside the "Peptide" tab, you will see the coverage tab, which contains the protein sequence with the matching peptide sequences underlined in blue. The darker the blue underline, the more confident the match.

```
  1  MRPAVRALLA CAVLGLCLAD PERTVRWCTI STHEANKCAS FRENVLRILE SGPFVSCVKK
 61  TSHMDCIKAI SNNEADAVTL DGGLVYEAGL KPNNLKPVVA EFHGTKDNPQ THYYAVAVVK
121  KDTDFKLNEL RGKKSCHTGL GRSAGWNIPM AKLYKELPDP QESIQRAAAN FFSASCVPCA
181  DQSSFPKLCQ LCAGKGTDKC ACSNHEPYFG YSGAFKCLME GAGDVAFVKH STVFDNLPNP
241  EDRKNYELLC GDNTRKSVDD YQECYLAMVP SHAVVARTVG GKEDVIWELL NHAQEHFGKD
301  KPDNFQLFQS PHGKDLLFKD SADGFLKIPS KMDFELYLGY EYVTALQNLR ESKPPDSSKD
361  ECMVKWCAIG HQERTKCDRW SGFSGGAIEC ETAENTEECI AKIMKGEADA MSLDGGYLYI
421  AGKCGLVPVL AENYKTEGES CKNTPEKGYL AVAVVKTSDA NINWNNLKDK KSCHTAVDRT
481  AGWNIPMGLL YSKINNCKFD EFFSAGCAPG SPRNSSLCAL CIGSEKGTGK ECVPNSNERY
541  YGYTGAFRCL VEKGDVAFVK DQTVIQNTDG NNNEAWAKNL KKENFEVLCK DGTRKPVTDA
601  ENCHLARGPN HAVVSRKDKA TCVEKILNKQ QDDFGKSVTD CTSNFCLFQS NSKDLLFRDD
661  TKCLASIAKK TYDSYLGDDY VRAMTNLRQC STSKLLEACT FHKP
```

Assigned/ UnAssigned De Novo View

The assigned de novo view displays de novo sequences that do have corresponding database matches. The assigned de novo view displays the de novo sequences that do have a corresponding database match. The two together are equivalent to the standalone de novo results.

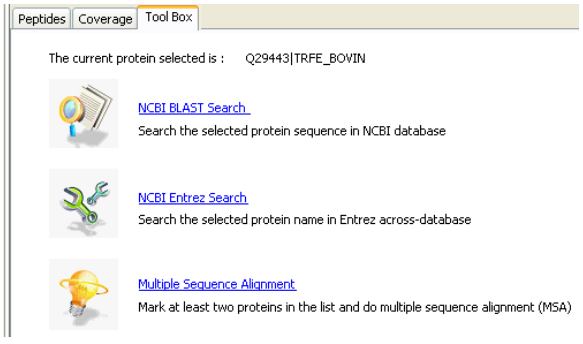


Tool Box

Click on the "Tool Box" tab:

The selected protein will be displayed at the top of this page. Click on NCBI Blast Search or NCBI Entrez Search to search for the selected protein sequence in the NCBI database or the Entrez database, respectively.

To perform a multiple alignment search mark two or more proteins in the above list of proteins, by clicking in their checkboxes (see example below).



Accession	ID	Mass	Coverage Display	PEAKS(Score %)	p-value	Coverage(%)	Query matched	Description	Marked
DB Search									
Q29443|TRFE_BOVIN	1	77753.200		98.3	8.25E-135	5.68	4	Serotransferrin - Bos taurus (Bovine).	☑
P02769|ALBU_BOVIN	2	69293.550		83.8	1.87E-60	3.62	2	Serum albumin - Bos taurus (Bovine).	☑
P00330|ADH1_YEAST	3	36691.957		59.5	4.2E-45	2.59	1	Alcohol dehydrogenase 1 - Saccharomyces cerevis...	☐

Then click on the "Multiple Sequence Alignment" link in the Toolbox window to generate the multiple sequence alignment in your web browser.

Identified peptides are highlighted in blue letters on the sequence. A more intense blue indicates a more confident match. The background colors indicate similarity between the sequences. A dark background indicates regions where residues or nucleotides are identical in all sequences, a light background indicates similarity across some sequences and lowercase letters on white background highlight differences. A dash '-' is displayed where a gap had to be introduced in one sequence to complete the alignment.

```
Q29443|TRFE_BOVIN      mrpavrallacavlglcladpertvrwctiSthEanKcasFrENvlRIlesgpfvscvkk
P00330|ADH1_YEAST      ------------------------------SipEtqKgviFyEShgKL------------

Q29443|TRFE_BOVIN      tsHmDcikaIsnnEADaVtLDgglVyeAGLkpnNLkpvvaeFHGtkdnPqthyyavavvk
P00330|ADH1_YEAST      -eHkDi--pVpkpKANeLlIN---VkySGVchtDLha----WHGdwplP------------

Q29443|TRFE_BOVIN      kdtdfKLneLrGkKsch--tGLGrSa-GWnIpmaklykelpdpqesiqrAaanFfSaSCV
P00330|ADH1_YEAST      ----vKLplVgGhEgagvvvGMGeNvkGWkIg---------------dyAgikWlNgSCM

Q29443|TRFE_BOVIN      pCadqssfpKlCQLcagkgtDKcaCsnHepyfGYS--GAF-KclmegAgdvAfVkhsTvf
P00330|ADH1_YEAST      aC-------EyCEL-----gNEsnCp-HadlsGYThdGSFqQyatadAvqaAhIpqgTdl

Q29443|TRFE_BOVIN      dnLpnpedrknyeLLCgdnTrksvddyqecYlAMvpSHaVvartvggkedviWelLNhAq
P00330|ADH1_YEAST      aqVa--------pILCagiT---------vYkALksANlMagh---------WvaISgAa
```

7.2.3 COMBINE DATASETS FOR ANALYSIS

PEAKS provides a function to combine datasets for protein/peptide identification.

Sample Level Identification

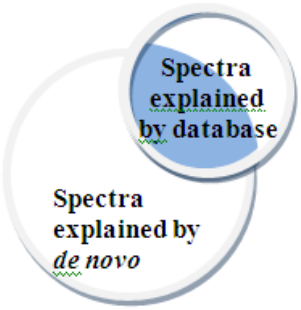
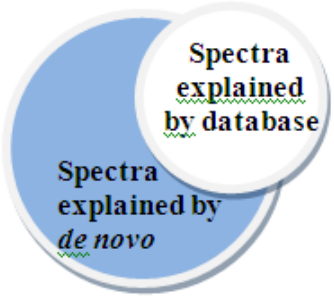
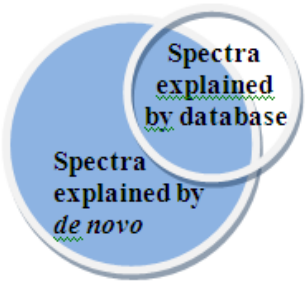
When you select a sample to perform PEAKS DB search, PEAKS will combine all datasets (files) within this sample together and generate a single result report.

Project Level Identification

When you select a whole project to perform PEAKS DB search, PEAKS will combine all samples in this project together and generate a single result report.

7.2.4 Filter Options

Previous users may recall the Filter Options pane. These options have been removed and re-displayed with greater prominence as part of the Assigned De Novo View and Unassigned De Novo View.

View	Graphical Depiction
Assigned De Novo View	
Unassigned De Novo View	
Independent De Novo View	

Chapter

8

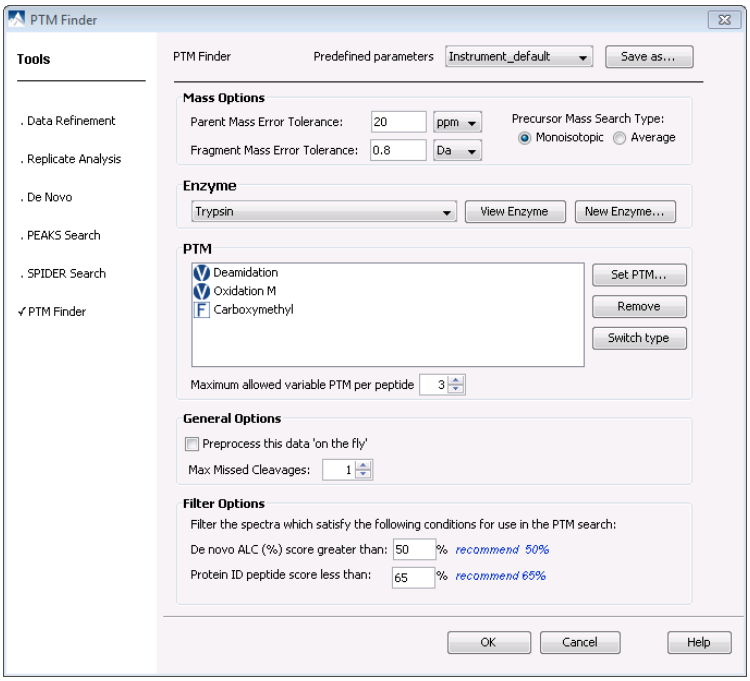
8 PTM FINDER

8.1 SETTING UP PTM FINDER PARAMETERS

Select a Protein ID results file to perform a PTM finder search.

Note that you cannot execute PTM Finder on a raw file or de novo results.

Click the PTM icon on the toolbar  Or Select "PTM Finder" from the Tools menu. The "PTM Finder Options" window will appear:



The presented parameters are identical to those used when performing protein ID (Refer to section 7.1 Setting up Protein Identification Parameters) with the exception of the filter options found at the bottom of the window. As PTM Finder searches tend to be computationally intensive, PEAKS will only look at de novo sequencing results that are above the amino acid score threshold and below the peptide score threshold that you input.

De novo ALC (%) score greater than: The PTM Finder requires a good quality sequence tag from de novo to be able to find good quality homologous proteins. Enter a value for the de novo score threshold. The recommended threshold is 50%.

Protein ID peptide score less than: As there is no need to run the PTM Finder on peptides that were already found to have a good match during PEAKS Protein ID, it is helpful to enter a peptide score threshold so that PTM Finder will only be performed on peptides below the threshold. The recommended threshold is 65%.

8.2 PTM FINDER RESULTS VIEW

The results from a PTM finder search are presented in a similar format to those seen in a PEAKS Protein ID search. Please refer to section 7.2 Protein Identification Results for more information on the PEAKS Protein ID search results.

Chapter

9

9 SPIDER SEARCH

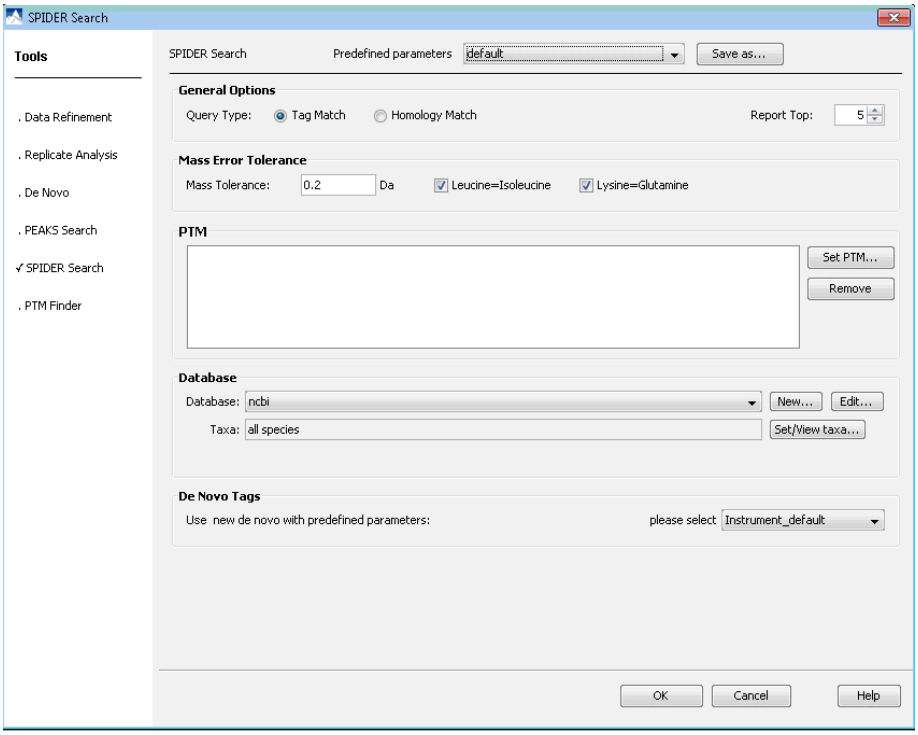
After having obtained de novo sequences for peptides that are not in the database, it is important to look for a homologous peptide in the database (Han et al., 2005). This will help you to learn more about the proteins in your sample. To search with SPIDER you must first have good quality de novo sequences.

9.1 SETTING UP SPIDER PARAMETERS

The configuration panel of SPIDER Search will be invoked by selecting a data file, a de novo result or a protein id result and clicking the SPIDER icon on the toolbar  or choosing SPIDER Search from the Tools menu. If SPIDER search is launched on a data file, de novo sequencing will be performed automatically with selected parameters. The configuration panel is as follows:

If SPIDER search is launched from a de novo result, PEAKS will use the existing tags in the de novo results.

If SPIDER search is launched from a PEAKS Protein ID result, we do not need to select a protein database or de novo sequence. A SPIDER search performed on a Protein ID result will produce a search based on the proteins identified in the Protein ID result, whereas a SPIDER search done on a data file will create new de novo results and search then against an entire database, as specified by the user.



The following section will describe the different options that you have when setting up the parameters for your SPIDER search.

9.1.1 General Options

Choose a Query Type. They are, in order of increasingly rigorous analysis:

Tag Match: This is not a true mutation search; instead, it will insist that the mass of the peptide returned is the same as that of the de novo sequence. This search mode allows for the use of fixed modifications.

Homology Match: this is a more rigorous (and a more resource intensive) search mode, taking into account all types of mutations and the positional confidence scores. This search mode also creates reconstructed peptides that use information from both the de novo sequences and the database sequence in order to characterize the real sequence. This search mode allows for the use of both fixed and variable modifications.

You can choose how many of the best homologous peptides should be displayed after searching by changing the number in the "Report Top" box.

9.1.2 Mass Error Tolerance

Mass tolerance: Enter the range of error (in Daltons) that PEAKS will allow for when determining the peptide sequences.

Choose if you would like PEAKS to consider Leucine equal to Isoleucine without a penalty in the score and also whether Lysine should be equal to Glutamine without penalty by selecting the appropriate checkboxes.

9.1.3 PTM

Set PTM: Clicking this button will bring up the "PTM Options" list, which informs PEAKS what types of post-translational modifications to include in its analysis. To view additional modifications, select the "Show unimod" box. If the desired PTM does not appear on the list or is different than what is listed, you can select the "New…" button and the "PTM Editing" window will open. Fill in the information pertaining to your particular PTM. For a more in depth explanation of creating a new PTM, refer to section 7.1.3 PTM Options.

To select a PTM as Fixed or Variable, click the desired PTM to highlight it and then click the button corresponding to its preferred designation, "Fixed Modification" or "Variable Modification". If you incorrectly place a PTM in one of these sections, simply reselect the PTM, and drag it out to have it removed from the list of selected modifications, or click "Switch Type" to have it set as the other modification type. For example, using the Switch Type button, a PTM currently listed as fixed, will automatically be removed from the fixed modification list and enter the variable modification list, or vice versa.

Note that in previous versions of PEAKS, only fixed PTMs were allowed; however, PEAKS now allows variable PTMs, while executing a homology search.

Max variable PTMs: To reduce uncertainty, PEAKS' de novo sequencing 'vocabulary' can be limited by restricting the number of variable PTM found on a peptide. Specify a number by typing it into the box. To lift such restrictions, type a very large number (longer than the length of the peptide).

9.1.4 Filter

As the SPIDER search is computationally intensive, it is not recommended that you run all of your de novo sequencing peptides against the database; only those that cannot be well explained.

De novo (ALC %) score greater than: The SPIDER search requires a good quality sequence tag from de novo to be able to find good quality homologous proteins. Enter a value for the de novo score threshold. The recommended threshold is 50%.

Protein ID peptide score less than: As there is no need to run SPIDER on peptides that have already been found to have a good match during PEAKS Protein ID, it is helpful to enter a peptide score threshold so that SPIDER will only search through peptides below the threshold. The recommended threshold is 65%.

9.1.5 Database Options

Note that these options are only visible if you choose to run a SPIDER search on a data file rather than a PEAKS result file. Selecting a database provides sequence variations for SPIDER to search against de novo sequences.

Database to search: From the drop-down list select one of the FASTA databases configured in PEAKS. To edit an already existing database, click on the "Edit Database" button. If the desired database is not in this list, click the "New Database" button. Note that you can also set up a new database in the "Database Configuration" window. The configuration window is the only place where you have the option of deleting databases which you have created. For more information on setting up new databases refer to section 16.2.3 Database Configuration.

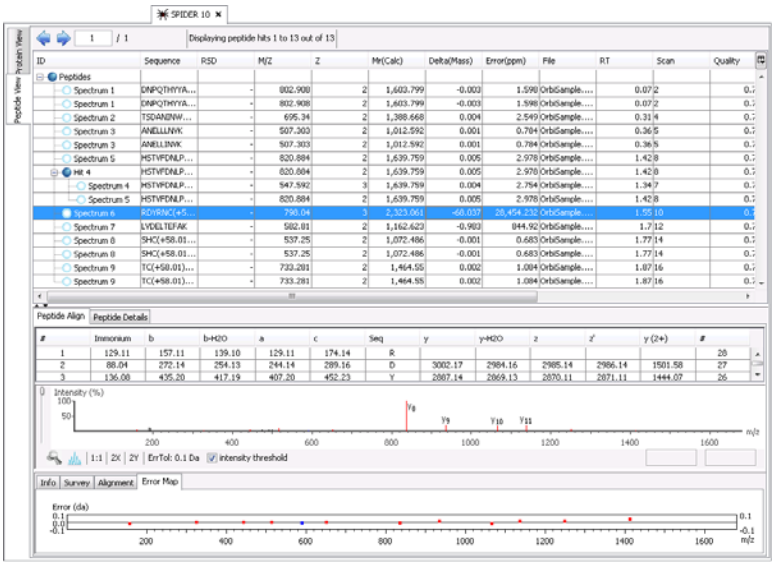
Taxonomy selection: This list displays the taxa you have chosen for your search. If the database selected has taxon information available, you can click on the "Set/View taxa…" button. Otherwise, the whole database will be searched. The selections correspond to established hierarchy -- i.e. selecting 'Mammalia' will search all of 'horse, cow, rat, mouse, human, etc.

After setting up the desired parameters, we can save them for future use. Click the "Save Parameters" button, and define name for these preferences for future use/reference, when prompted. Any parameters that you save will be available in the drop-down list at the top of the window. To examine the contents, select a saved parameters file and parameter boxes will be populated.

Press the "OK" button and the SPIDER search will begin.

9.2 SPIDER Results View

SPIDER will search the database for homologous peptides and at the same time attempt to consolidate these into protein hits. The result report will look much like the results for PEAKS Protein ID or inChorus searching.



Clicking on the "Peptide View" tab will display results that look very much like the results for PEAKS Protein ID. Refer to section 7.2.1 Peptide View for more details. Click on the "Peptide details" tab to see the SPIDER matches shown in red. Note that only SPIDER homology search results will display a reconstructed or "real" sequence.

Letters on a green background with vertical bars, indicate agreement. Letters on a red background indicate sequencing error. Color codes on the de novo sequence letters still indicate positional confidence. Letters on a blue

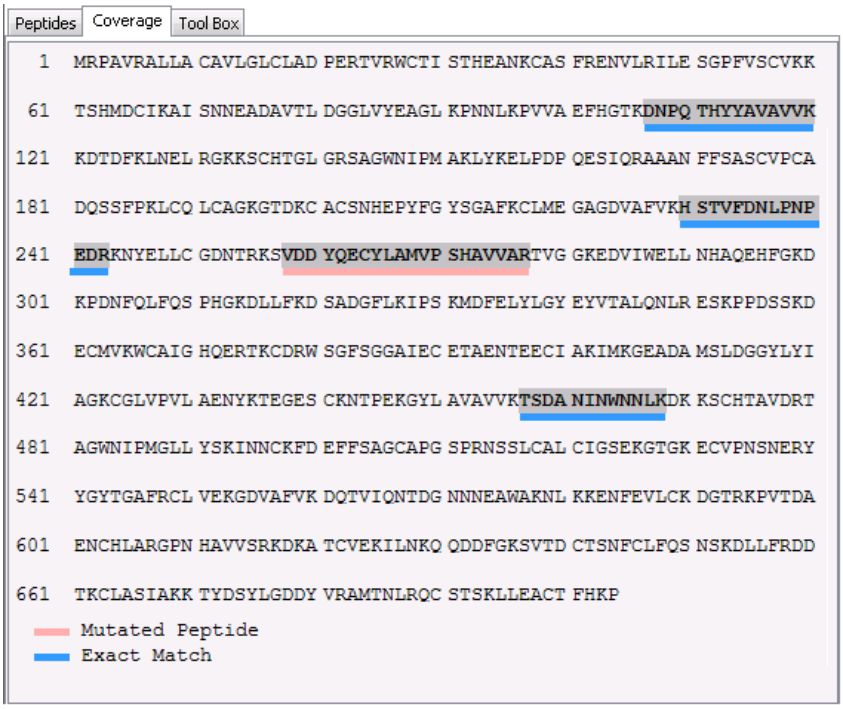
background indicate uncertainty or mutation. "+" signs represent more likely mutations. [Square brackets] indicate an equal mass substitution, common non-critical de novo errors. While <Angle brackets> indicate an equal mass substitution and a mutation.

When simply identifying exact peptides from the database using PEAKS Protein ID, SPIDER tag match or inChorus, there is no need to reconstruct the 'real' sequence.

Clicking on "Protein View" will yield a similar display as was seen for PEAKS Protein ID (refer to section 7.2.2 Protein View). In the "Coverage Display" column, red region indicate areas of homology and potential mutation identified by SPIDER. Blue regions indicate areas of homology from the protein ID search, and purple indicates that the spectrum was identified by both SPIDER and protein ID.

Accession	ID	Mass	Coverage Display	PEAKS(Score %)	p-value	Coverage(%)	Query matched	Description
DB Search								
Q29443|TRFE_BOVIN	1	77753....		98.3	8.25E-...	8.52	7	Serotransferrin - Bos taurus (Bovine).
P02769|ALBU_BOVIN	2	69293....		83.8	1.87E-60	5.27	5	Serum albumin - Bos taurus (Bovine).
P00330|ADH1_YEAST	3	36691....		59.5	4.2E-45	2.59	2	Alcohol dehydrogenase 1 - Saccharomyces cere

Click on the "Coverage" tab. The highlighted colors are the same as those shown in the "Display" column above.



Chapter

10

10 inChorus (Multi-Engine Protein ID)

The inChorus protein identification tool can call upon several search engines (PEAKS Protein ID, X!Tandem, OMSSA, Mascot and Sequest) for protein identification and will then compare and summarize the results from the different search engines in one unified report.

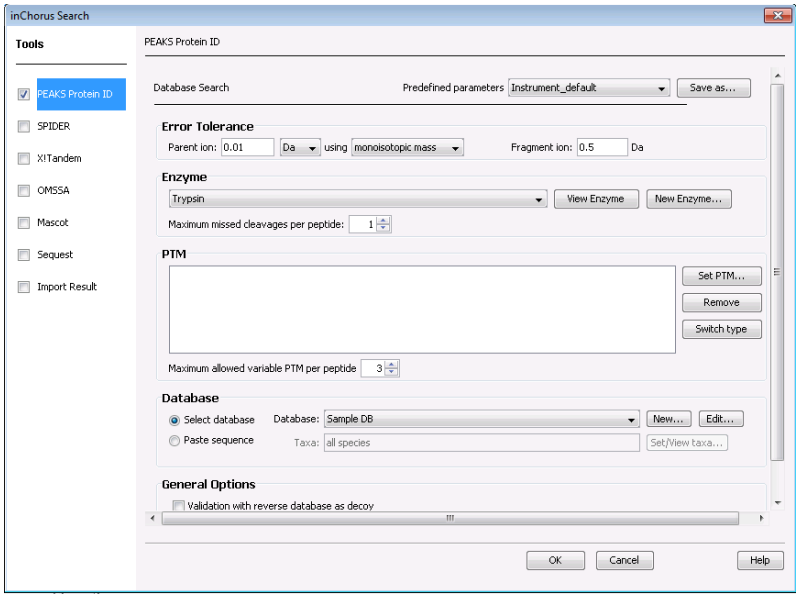
Please note that if you wish to launch either Mascot or Sequest for an inChorus search, these engines are not internally provided with PEAKS and may need to be purchased from their respective vendors; however, licenses of these engines are not required if the result files have already been obtained. To set up your search engine preferences, refer to section 16.1.3 Search Engine Preferences.

10.1 Setting up inChorus Parameters

Select the project, sample, or data file node to be run through inChorus.

Click the "inChorus Search" icon on the toolbar  Or Select "inChorus Search" from the Tools menu.

The "inChorus Options" window will appear:



First select each of the protein identification tools that you would like to use by checking the respective checkboxes. Search parameters for each program can be set by selecting the name of the search engine.

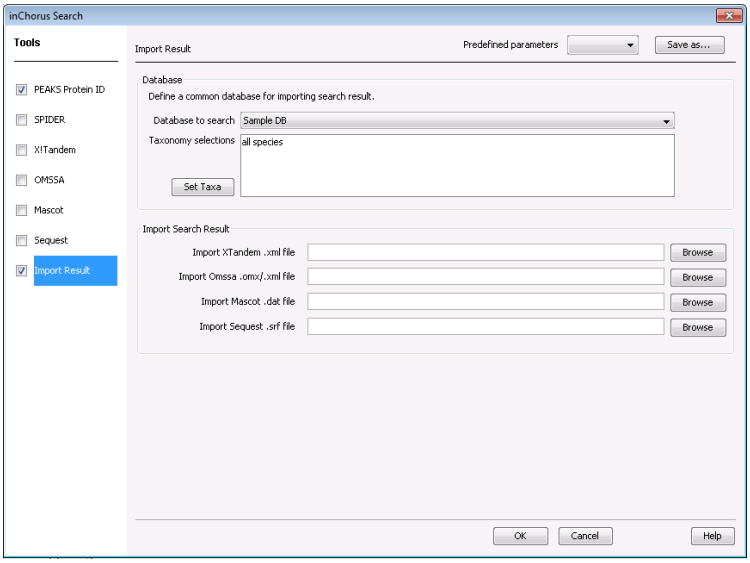
The option screens for each of the search engines available to inChorus are designed to work in the same way as options screens from the original programs. For help in setting search parameters for each program, please refer to that program's user manual. For help with PEAKS Protein ID, please refer to section 7.1 Setting up Protein Identification Parameters.

10.1.1 Importing Existing Results

PEAKS inChorus reads X!Tandem .xml files, OMSSA .omx files, Mascot .dat files and Sequest .srf files. When importing third party results files please make sure that the scan number in the results file is consistent with the one in PEAKS. PEAKS uses original data information to compute the inChorus score.

When you run an X!Tandem search with command line, you need to turn on the option of "-w" in order to export data information into X!Tandem .xml files.

To import existing results, check the "Import Result" checkbox and select "Import Result". The following window will open.



Although it is not necessary for the various search engines to use the same database in an inChorus search, it is necessary to have a unified database for an inChorus search that includes imported results. Select the database that you would like to use from the drop-down list. The inChorus search will be performed on all species in the database unless specified by the user. If this database does not appear in this list, refer to 16.2.3 Database Configuration to configure your databases. To specify which taxa you would like to search, click on the "Set Taxa" button.

To import your file, click the "Browse" button that is found beside the appropriate search engine. Find the file that you would like to import and click "Open". Once you have selected the file(s) that you would like to import and have selected the options for the other search engine searches you would like to perform, click "OK".

10.2 inChorus Results View

When the inChorus search is complete the "Project View" panel should contain a separate result node for each search engine that you selected as well as an inChorus report that combines the results from the multiple search engines. See an example below:

10.2.1 de novo, Peptide and Protein Views

Each result node for the 3rd party search engines looks very similar to the PEAKS Protein ID result node (Protein Identification Results) with a few small differences. Firstly, there is no "de novo View" or "Chart View", and secondly, the scoring will be specific to that search engine. For example, the score for OMSSA is listed as an E-value. For more information about different scoring methods, refer to the user manual of the third party search engines.

The inChorus search report also looks very similar to PEAKS Protein ID results file. The de novo sequencing results that are found in the "de novo View" are only those that correspond to results that have been identified by one of the search engines in the inChorus search.

Chapter

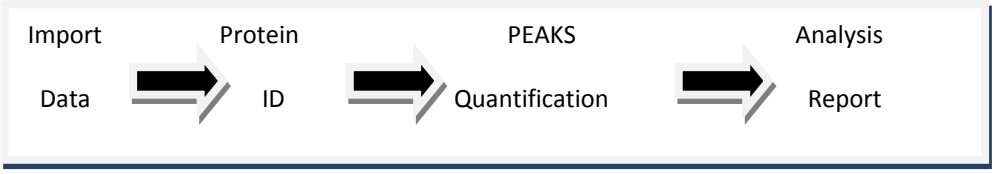
11

11 PEAKS® QUANTIFICATION MODULE

Many approaches for quantifying proteins using mass spectrometry have been described in the literature (ref). In terms of their implementation, most of them can be classified into three protocols.

- **MS**: Quantification based on the relative intensities of extracted ion chromatograms (XICs) for precursors within a single data set. This is the most widely used approach, which can be used with any chemistry that creates a precursor mass shift. For example, **ICAT** and **SILAC**.
- **MS/MS**: Quantification based on the relative intensities of fragment peaks at fixed m/z values within an **MS/MS** spectrum. For example, **iTRAQ** and Tandem Mass Tags (**TMTs**).
- **Label free**: This quantification type is based on the relative intensities of extracted ion chromatograms (XICs) for precursors in multiple data sets aligned using mass and elution time. Please note PEAKS requires the survey scan to be run in profile mode.

All three protocols are fully implemented within PEAKS Q. The flow chart is shown below:



11.1 SETTING UP PEAKS Q PARAMETERS

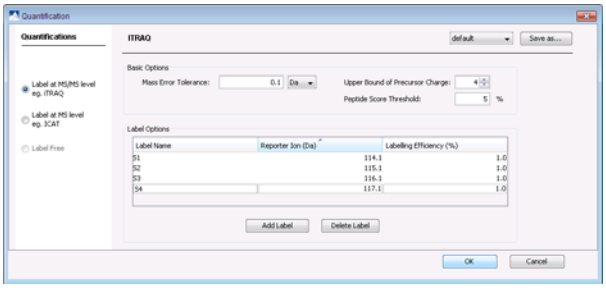
11.1.1 PARAMETERS FOR LABELED QUANTIFICATION

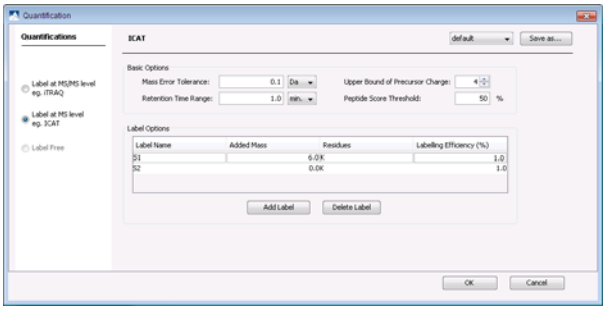
1) In the "Project View Frame", select a PEAKS Search result file.

2) Click the PEAKS quantification toolbar icon. Or Select "Quantification" from the "Tools" menu.

The following window will open, displaying the labeled quantification parameters:

Slightly different options will be available if you select labeling at the MS level:





Once a quantification type is selected, many quantification parameter options are included.

These options include the following:

Basic Options

- Mass Error Tolerance: Quantification is based on the feature of a peptide that identifies its origin in the sample mixture. For example, in a SILAC experiment, one feature is unmodified peptides and the other is peptides modified with Label:13C(6) on Arginine or Lysine. For iTRAQ, the feature would be reporter ion m/z value. The mass error tolerance is for pairing up features.
- Upper Bound of Precursor Charge: The peptide may be present in different charge states. Upper bound of precursor charge defines the maximum charge of peptides which are used for counting quantity.
- Peptide Score Threshold: Only those identified peptides with score above this threshold will be used in quantification.
- Retention Time Range: The retention time range is for pairing up features.

Labeling Tools

- Labeling occurs at the MS/MS level e.g. iTRAQ: It is for quantification based on the relative intensities of fragment peaks at fixed m/z values within an MS/MS spectrum.
- Labeling occurs at the MS level e.g. SILAC: It is for quantification based on the relative intensities of extracted ion chromatograms (XICs) for precursors within a single data set.
- ***MS Label Options***
- Label Name: It is for specifying label name which cannot be left empty
- Reporter Ion: It is for specifying mass of reporter ion.
- Labeling Efficiency: It is for specifying efficiency of chemical reaction.
- Add Label: Click on this button will add a default label.
- Delete Label: Select a label and Press this button, the selected label will be deleted.

MS/MS Label Options

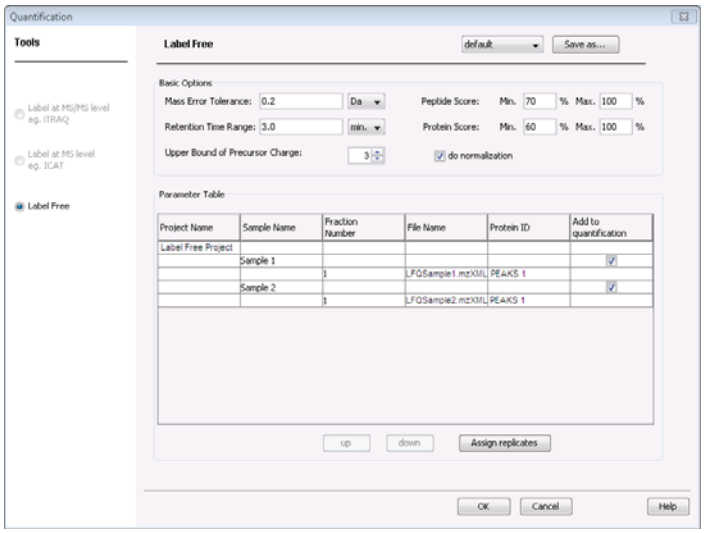
- Label Name: It is for specifying label name which cannot be left empty
- Added Mass: The modified mass of a residue.
- Residues: The residues to be modified given in single codes of amino acids, i.e. A or AK
- Labeling Efficiency: It is for specifying efficiency of chemical reaction.
- Add Label: Click on this button will add a default label.
- Delete Label: Select a label and Press this button, the selected label will be deleted.

11.1.2 PARAMETERS FOR LABEL FREE QUANTIFICATION

1) In the "Project View Frame", select any node of the project you want to do label free quantification. The project should have at least one PEAKS Search result.

2) Click the PEAKS quantification toolbar icon  . Or Select "Quantification" from the "Tools" menu.

The following window will open, displaying the label free quantification parameters:



Basic Options

- Mass Error Tolerance: The mass error tolerance is for pairing up features.
- Retention Time Range: The retention time range is for pairing up features. Only the feature retention time within this range may pairing up.
- Peptide Score: Only those identified peptides with score within the given range will be used in quantification.
- Protein Score: Only those identified proteins with score within the given range will be quantified
- Upper Bound of Precursor Charge: The peptide may be present in different charge states. Upper bound of precursor charge defines the maximum charge of peptides which are used for counting quantity.
- Do normalization: If this option is selected, normalization to protein ratio which based on total ion intensity will be done automatically.

Label-Free Parameter Table

- Project Name: Gives the name of the projects selected to do quantification
- Sample Name: Gives the names of samples in the project
- Fraction Number: Gives the number of the fraction in the sample
- File Name: Gives the name of the data file
- Protein ID: Gives the name of the PEAKS node will be used in quantification
- Add to Quantification: Check/uncheck to add the sample to the quantification. There must be at least two samples in label free quantification and the number of fraction within each sample must be the same.
- Up/Down: If the order of fractions in the sample is not consistent with that in the experiment, these two buttons can be used to adjust the order.

11.1.3 Saving Parameters

Clicking the "Save As" button at the top right allows the user to save parameters for ease of use when regularly performing quantification with the same parameters. All the parameters in quantification will be saved except the project table in label free quantification which will change from one project to another.

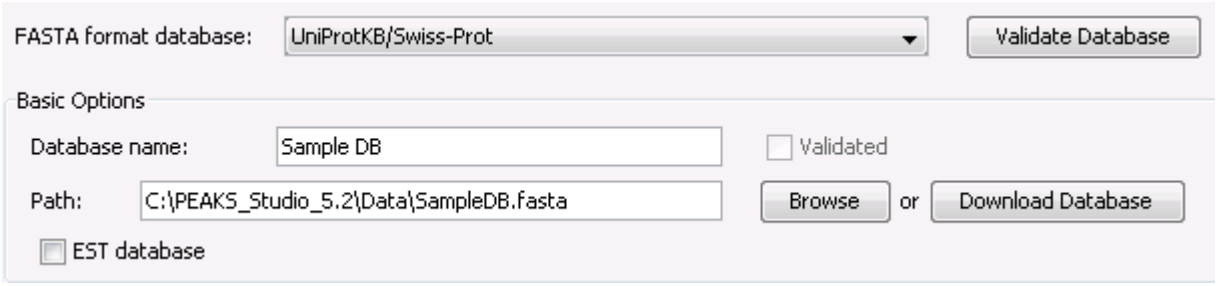
11.2 iTRAQ Walkthrough

Isobaric Tagging for Relative and Absolute Quantification (iTRAQ) uses isotopic labeling to enable relative quantitative comparisons. Up to eight different proteomic samples can be labeled using eight different isobaric tags.

11.2.1 Setting up a Database

Before running the walkthrough data, you need to set up a database. So that this can be a quick process, we have provided you with a sample fasta database called "SampleDB.fasta" in your PEAKS program folder (C:\PeaksStudio 5.2\Data).

Click on the configuration toolbar icon  or select "Configuration" from the "Window" menu. Select "Database" from the left hand side of the window. Under "Database Details" enter the following information:



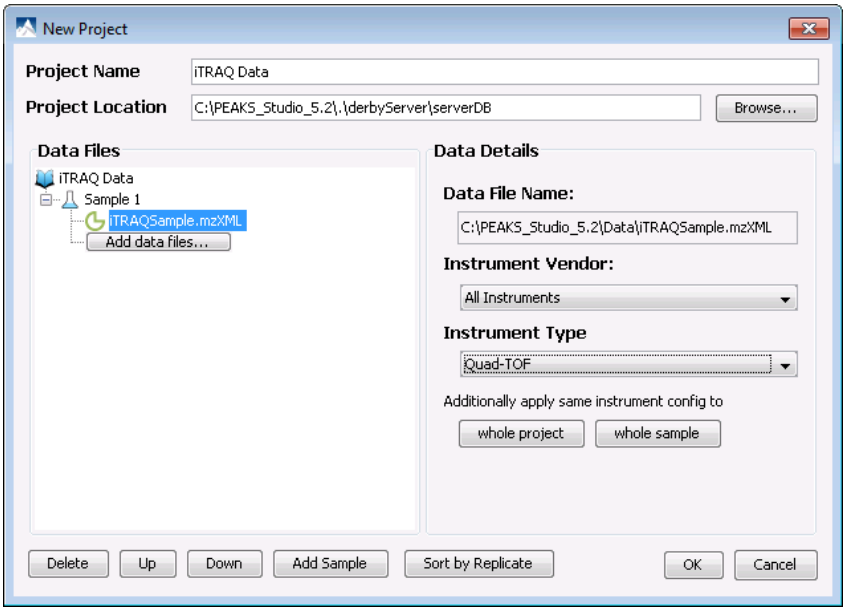
You do not need to change any of the other information listed. Click the "Add/Update" button and then click "OK".

11.2.2 Creating a Project

Click on the "Create new project" icon  or select "New project" from the "File" menu. Give a name to your project, for example "iTRAQ Data". Click Add data files. Locate "iTRAQSample.mzxml" from C:\PeaksStudio5.2\Data.

With the file highlighted, select, Quad-TOF from the "Instrument Type" drop-down list.

Click "OK". Wait until the file has

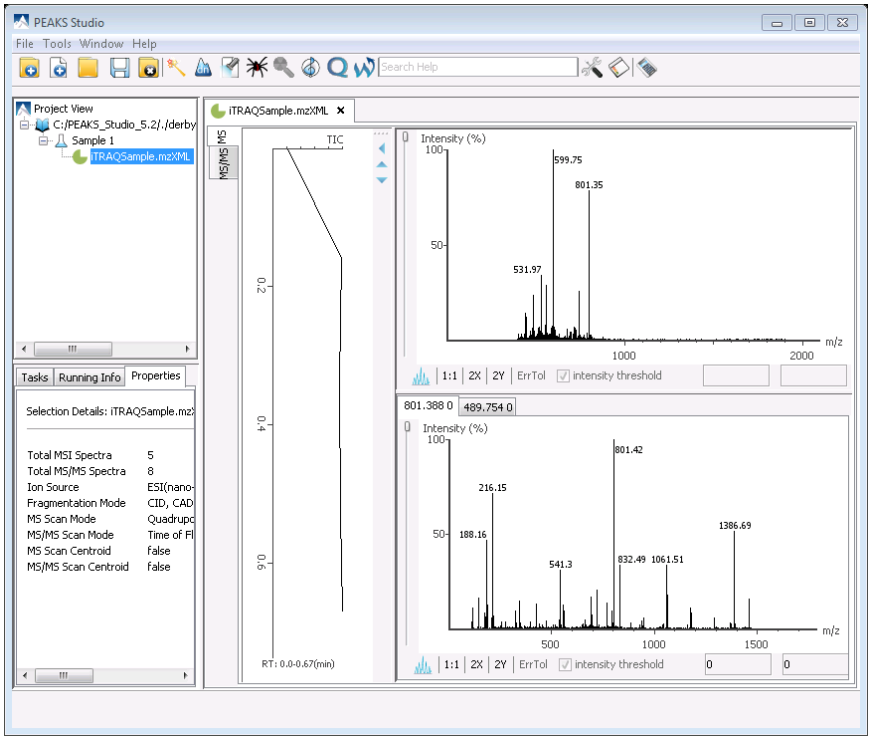


loaded.

Double click on the iTRAQ sample file.

Your “Main Processing Screen” should look something like this:

For ease of setup, this walkthrough will use the workflow tool in order to set all of the parameters at once.



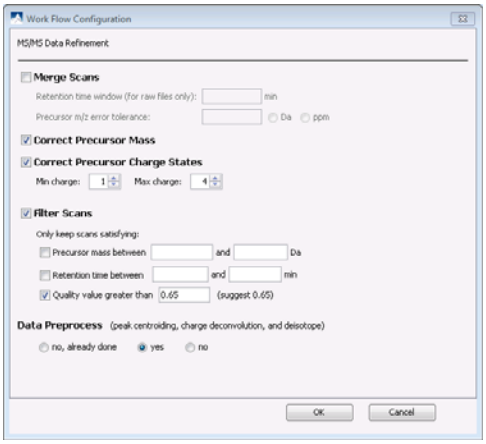
11.2.3 Setting up the Workflow Parameters

Click on the workflow button “” and select “Quantification”. The following window will appear:

Click “Select Data”. Select the iTRAQSample.mzXML file and then click “Add to Right”. The samples should now appear in the “Selected Data” panel.

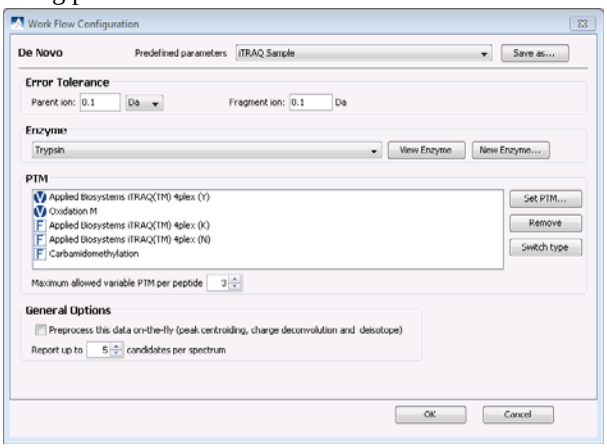
Click “OK”. The workflow window should now display “finished” beside the “Selected Data” button then click the “Refine data” button.

Enter the following parameters:



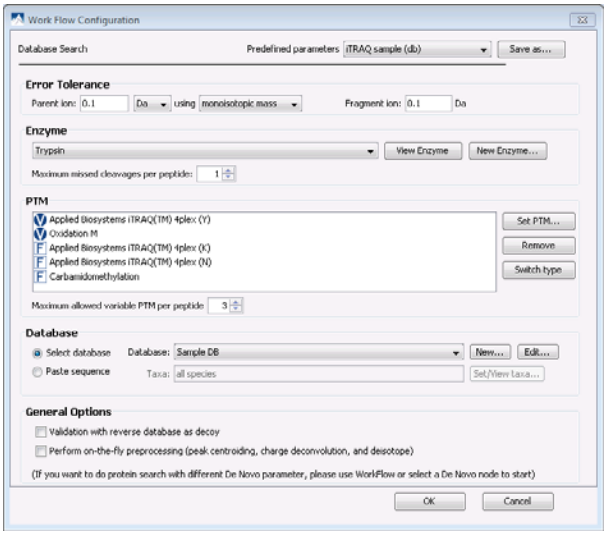
- Correct Precursor Mass
- Correct Precursor Charge State:
  - Min 1, Max 4
- Filter Scan:
  - Quality value greater than 0.65
- Data Preprocess: Yes

Click “OK”. Then click the “de novo” button. Enter the following parameters:



- Parent ion: 0.1 Da
- Fragment ion: 0.1 Da
- Enzyme: Trypsin
- PTM:
    - AB 4plex (Y) and Oxidation M as variable.
    - AB 4plex (K), (N) and Carbamidomethylation as fixed.
- Min allowed variable PTM: 3

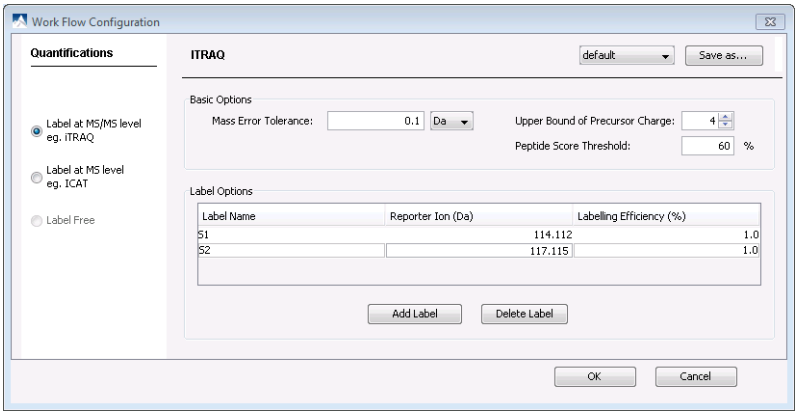
Click “OK” and then click the “PEAKS Search” button.



Enter the following parameters:

- Parent ion: 0.1 Da
- Fragment ion: 0.1 Da
- Enzyme: Semi Trypsin
- PTM:
    - AB 4plex (Y) and Oxidation M as variable.
    - AB 4plex (K), (N) and Carbamidomethylation as fixed.
- Min allowed variable PTM: 3

Click the “Quantification” button. Select “Label at the MS/MS level” from the left hand panel and enter the following parameters:

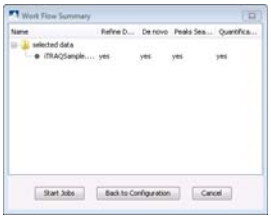


- Mass Error Tolerance: 0.1 Da
- Upper Bound of Precursor Charge: 4
- Peptide Score Threshold: 60

Note: To add information into the “Label Options” panel, click the “Add Label” button to add a new sample.

- Add labels S1: 114.112 & S2: 117.115

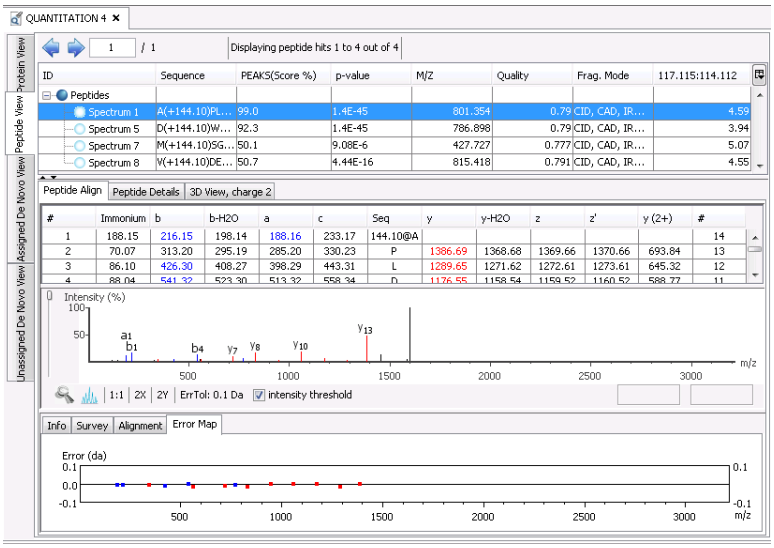
Click “OK” and then click “Start” in the workflow configuration panel. The following window will open:



Click “Start Jobs”.

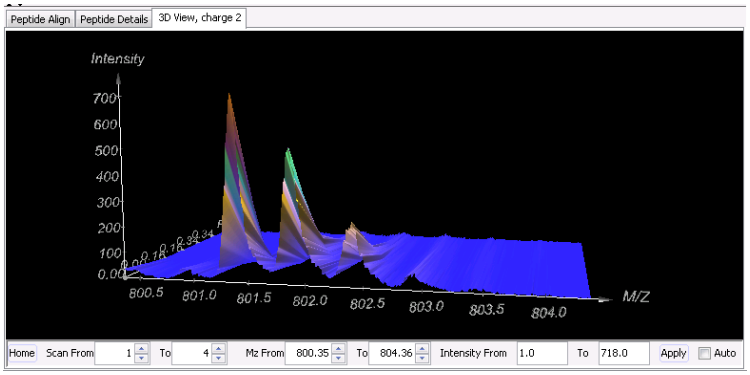
11.2.4 ITRAQ RESULTS

Once completed, the protein quantification result will be displayed in the Quantification node  . Double click on this node and the “Peptide View” tab will appear by default. The quantification results are listed as a “Ratio of 117.115:114.112”:

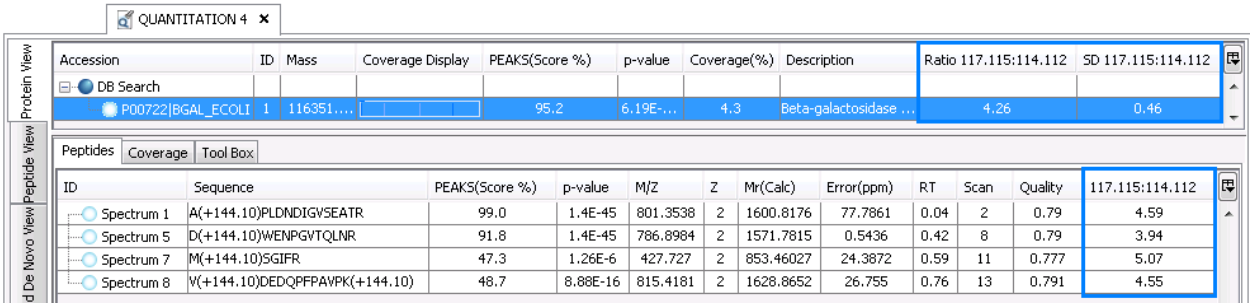


Click on the “Peptide Details” tab to see a simple alignment between the original de novo sequence, the peptide found in the database and the reconstructed sequence. At the bottom of the “Peptide Details” panel you will see where the selected peptide matches the protein highlighted in blue.

PEAKS provides the 3D View of each peptide feature for visual validation. Click on the 3D View tab. The panel along the bottom allows you to focus in on the peptides that you would like to examine. You can specify a particular scan number range, m/z range or intensity range. Click the “Apply” button to change the 3D View to your specified values.



Select the “Protein View” tab. The quantification results are listed as a “Ratio of 117.115:114.112” and as “Standard Deviation of 117.115:114.112”. They are highlighted in the blue box below. For example the relative protein ratio for the top ranked protein (Beta-galactosidase) is 4.26 with a standard derivation of 0.46.



11.3 SILAC WALKTHROUGH

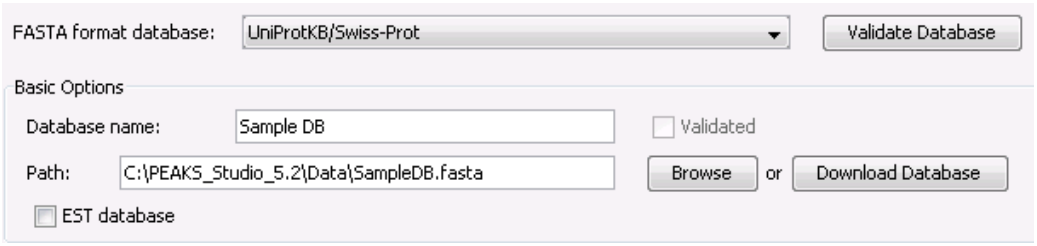
Stable Isotope Labeling with Amino Acids in Cell Culture (SILAC) is a method to metabolically label proteins for relative quantitative comparison. One cell population is fed amino acids of normal isotopic composition; the other cell population is fed amino acids labeled with heavier isotopes. The heavy amino acids are incorporated into newly synthesized proteins, eventually completely replacing the cells' proteins, such that labeling efficiency is near 100%. The cell populations are then mixed together and digested for MS analysis to determine differential protein abundances.

11.3.1 SETTING UP A DATABASE

Before running the walkthrough data, you need to set up a database. So that this can be a quick process, we have provided you with a sample fasta database called "SampleDB.fasta" in your PEAKS program folder (C:\PEAKS Studio 5.2\Data).

Click on the configuration toolbar icon or select "Configuration" from the "Tools" menu. Select "Database" from the left hand side of the window. Under "Database Details" enter the following information:

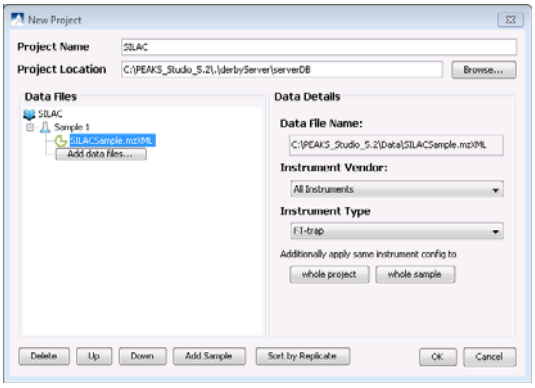
You do not need to change any of the other information listed. Click the "Add/Update" button and then click "OK".



11.3.2 CREATING A PROJECT

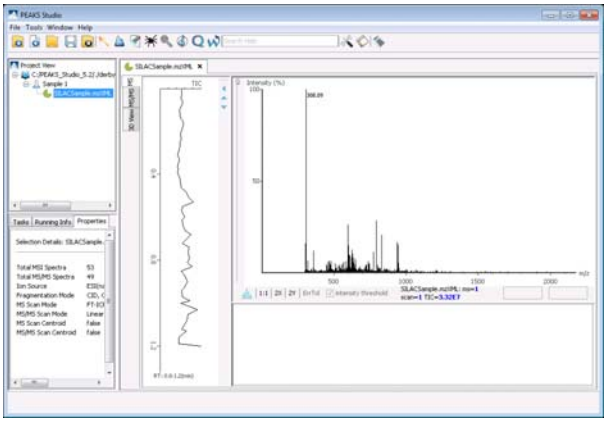
Click on the "Create new project" icon or select "New project" from the "File" menu. Give a name to your project, for example "SILAC Sample". Click Add data files. Locate "SILACSample.mzxml" from C:\PeaksStudio5.2\Data.

With the file highlighted, select, FT-trap from the "Instrument Type" drop-down list. Click "OK".



Double click on the SILAC sample file. Your "Main Processing Screen" should look something like this:

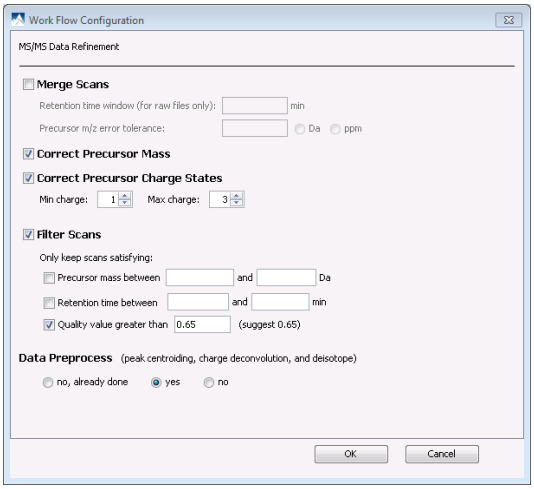
For ease of setup, this walkthrough will use the workflow tool in order to set all of the parameters at once. You will need to wait until the file has loaded (i.e. the file node looks like this:).



11.3.3 Setting up the Workflow Parameters

Click on the workflow button "  ". The following window will appear.

Click "Select Data". Select the SILACSample.mzXML file and then click "Add to Right". The samples should now appear in the "Selected Data" panel. Click "OK".

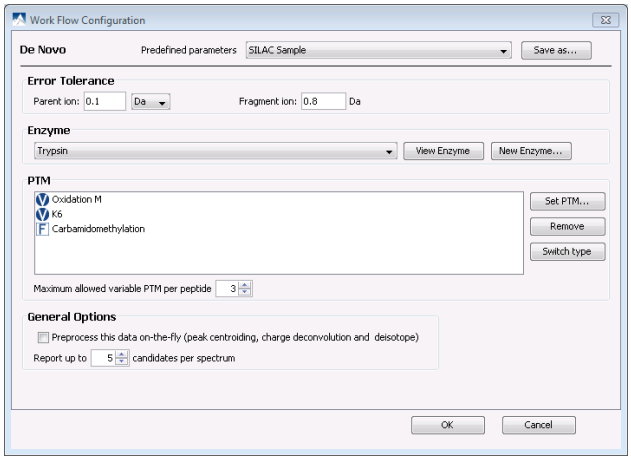


Check the 2nd box and then click the "Refine data" button. Enter the following parameters:

- Correct Precursor Mass
- Correct Precursor Charge State:
    - Min 1, Max 3
- Filter Scan:
    - Quality value greater than 0.65
- Data Preprocess: Yes

Click "OK".

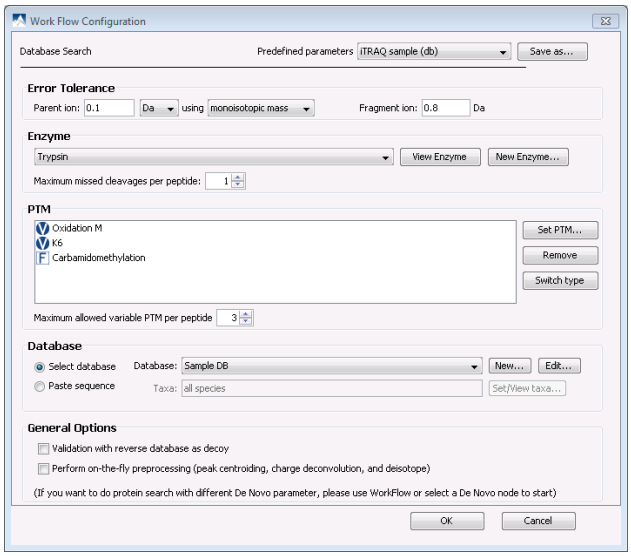
Check the 3rd box and then click the "de novo" button. Enter the following parameters:



- Parent ion: 0.1 Da
- Fragment ion: 0.8 Da
- Enzyme: Trypsin
- PTM:
    - K6 and Oxidation M as variable.
    - Carbamidomethylation as fixed.
- Min allowed variable PTM: 3

Click "OK".

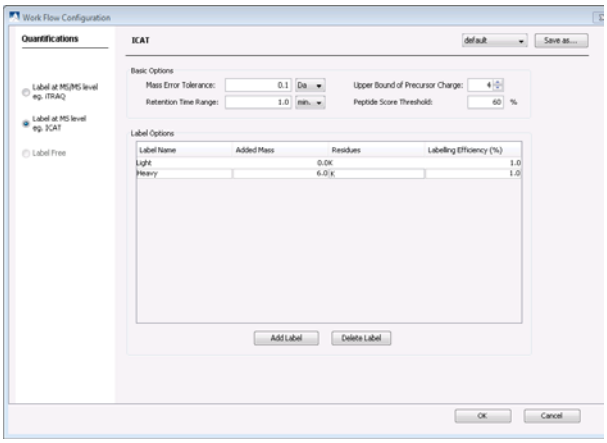
Check the 4th box and then click the "PEAKS Search" button. Enter the following parameters:



- Parent ion: 0.1 Da
- Fragment ion: 0.8 Da
- Enzyme: Trypsin
- PTM:
    - K6 and Oxidation M as variable.
    - Carbamidomethylation as fixed.
- Min allowed variable PTM: 3

Click "OK".

Check the 5th box and then click the "Quantification" button. Select "Label at the MS level" from the left hand panel and enter the following parameters:



- Mass Error Tolerance: 0.1 Da
- Upper Bound of Precursor Charge: 4
- Retention Time Range: 1.0 min
- Peptide Score Threshold: 60

Note: To add information into the "Label Options" panel, click the "Add Label" button to add a new sample.

Label: Light, with added mass 0.0 on residue K.

Label: Heavy, with added mass 6.0 on residue K

Click "OK" and then click "Start" in the workflow configuration panel. The following window will open:

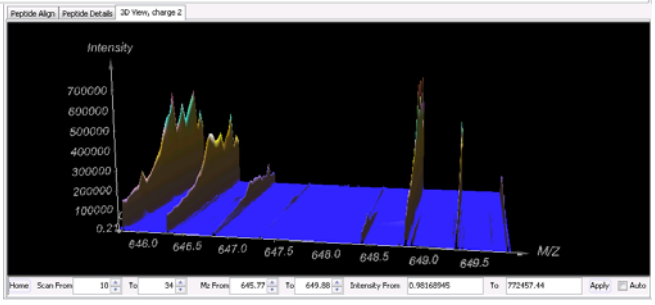
Click "Start Jobs".

11.3.4 SILAC RESULTS

Once completed, the protein quantification result will be displayed in the PEAKS Protein ID node "". Double click on this node and the "Peptide View" results will appear by default:

Clicking on the "Peptide Details" tab will display selected peptide matches in the protein highlighted in blue.

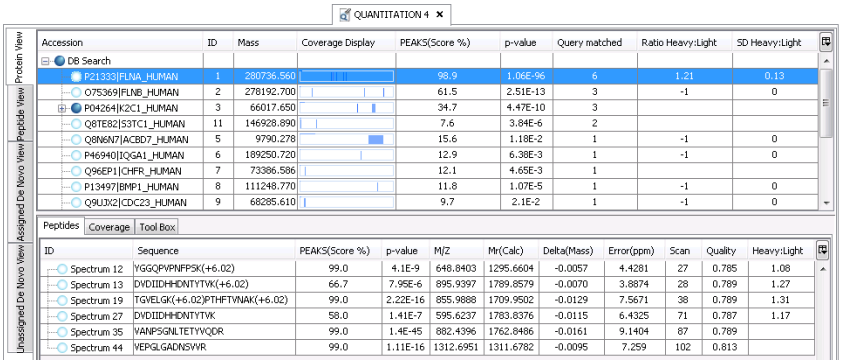
PEAKS provides the 3D View of each peptide feature for visual validation. Select the "3D View" tab. The panel along the bottom allows you to focus in on the peptides that you would like to examine. You can specify a particular scan number range, m/z range or intensity range. Click the "Apply" button to change the 3D View to your specified values.

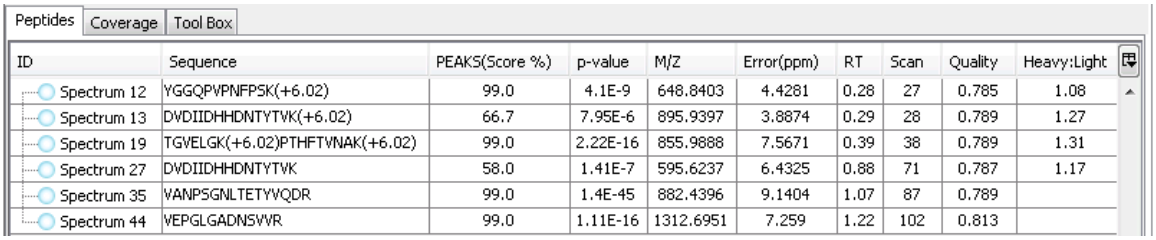


Select the “Protein View” tab. The SILAC quantification results are listed as a “Ratio Heavy: Light” and “Standard Deviation Heavy: Light”.

Note that the top protein result is Human Filamin A, with a score of 98.89%. The ratio of Heavy: Light is highlighted for each protein in the red box below. For example, the highest ranked protein, Human Filamin-A has a ratio of 1.2 and a standard deviation of 0.13.

A ratio of -1 means that while a reporter ion was found for the selected peptide, a ratio could not be calculated for the protein as the peptide did not meet the required score criteria (in this case we specified that the peptide score had to be greater than 0.6). A blank box in the ratio column means that a reporter ion for the selected peptide was not found. There are four peptide features identified and calculated for quantification of Human Filamin-A. Two other peptides (Spectrum 35 and 44) were used to identify the protein but were not identified as features.





11.4 Label Free Quantification

Label Free quantification relies on the changes in analyte signals directly reflecting their concentrations in one sample relative to another. This technology employs overall spectral intensity normalization by interpreting signals of molecules that do not change concentration from sample to sample. By comparing two or more spectra, PEAKS can determine the constant intensity ratio between the unchanging analytes. These unchanging analytes form the basis for identifying the non-changing concentrations, making spiking unnecessary.

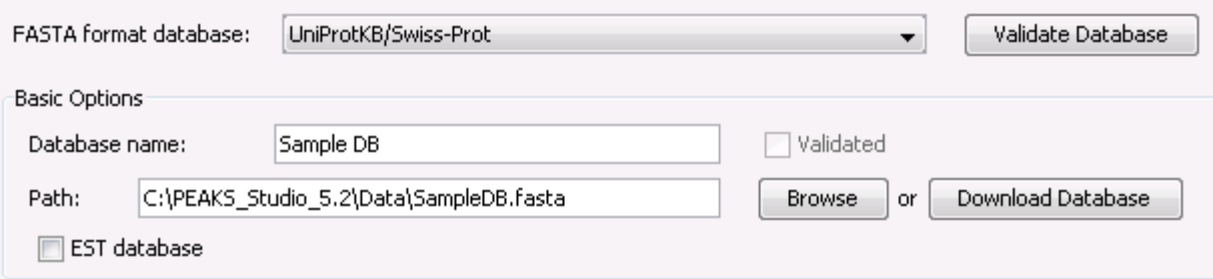
Note: PEAKS requires the survey scan to be run in profile mode.

Below are instructions to take you through a label free dataset. Please be aware that the data files are fairly large and thus the computing time will be increased compared to the other walkthrough datasets explained in this manual.

11.4.1 Setting up a Database

Before running the walkthrough data, you need to set up a database. So that this can be a quick process, we have provided you with a sample fasta database called “SampleDB.fasta” in your PEAKS program folder (C:\PEAKS Studio 5.2\Data).

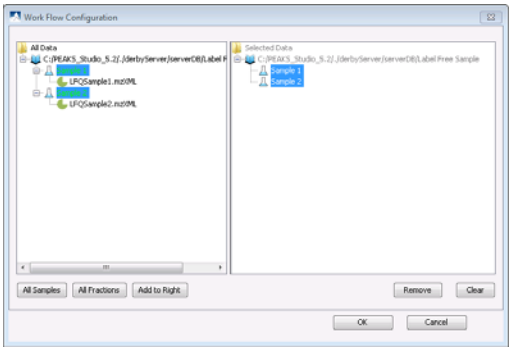
Click on the configuration toolbar icon or select “Configuration” from the “Tools” menu. Select “Database” from the left hand side of the window. Under “Database Details” enter the following information:



You do not need to change any of the other information listed. Click the "Add/Update" button and then click "OK".

11.4.2 Creating a Project

Click on the "Create new project" icon or select "New project" from the "File" menu. Give a name to your project, for example "Label free walkthrough". Click Add data files. Locate LFQSample1.mzXML. Then click "Add Sample" and Sample 2 will appear. Click "Add data files" to add "LFQSample2.mzXML" to Sample 2.



Note: For label free quantification to function in PEAKS, you need to have at least 1 sample with at least 1 file/fraction in each sample.

With one of the raw files highlighted, select, FT-trap from the "Instrument Type" drop-down. Then click "whole project" to apply these parameters to the rest of the project.

Click "Ok".

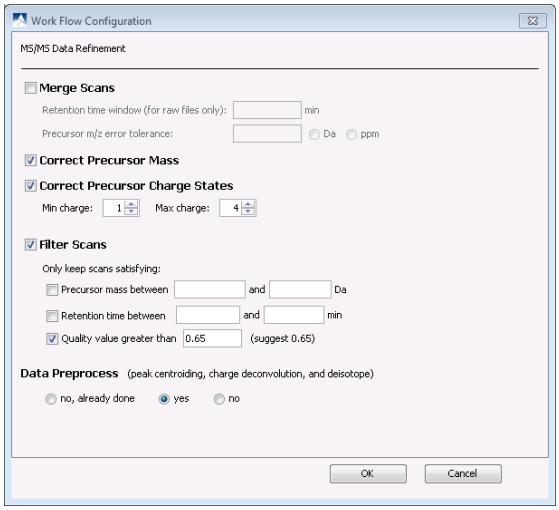
For ease of setup, this walkthrough will use the workflow tool in order to set all of the parameters at once. You will need to wait until the files have loaded (i.e. the file node looks like this:).

11.4.3 Setting up the Workflow Parameters

Click on the workflow button " ". The following window will appear:

Click "Select Data". The following window will open:

Click "All Samples" and then "Add to Right" or select the project node in the left tree and add to right. In work flow, label free quantification can only be done at the project level or at the sample level. Click "OK".

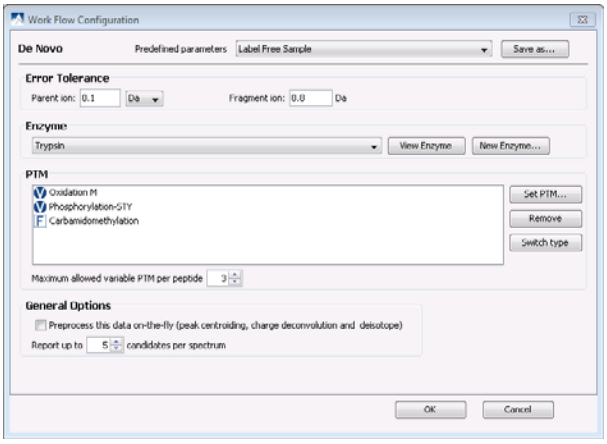


Then click the "Refine data" button. Enter the following parameters:

- Correct Precursor Mass
- Correct Precursor Charge State:
  - Min 1, Max 4
- Filter Scan:
  - Quality value greater than 0.65
- Data Preprocess: Yes

Click “OK”.

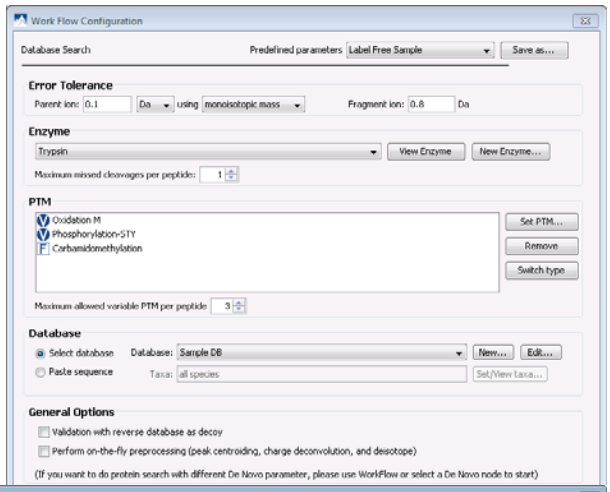
Click the “De novo” button. Enter the following parameters:



- Parent ion: 0.1 Da
- Fragment ion: 0.8 Da
- Enzyme: Trypsin
- PTM:
    - Oxidation M and Phosphorylation-STY as variable
    - Carbamidomethylation as fixed.
- Min allowed variable PTM: 3

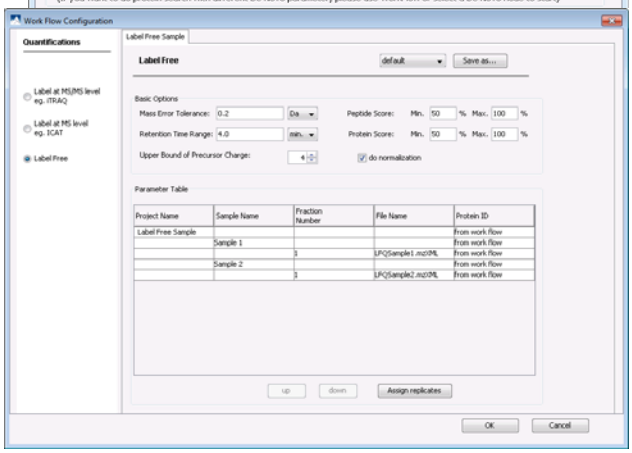
Click “OK”.

Click the “PEAKS Search” button. Enter the following parameters:



- Parent ion: 0.1 Da
- Fragment ion: 0.8 Da
- Enzyme: Trypsin
- PTM:
    - Oxidation M and Phosphorylation-STY as variable
    - Carbamidomethylation as fixed.
- Min allowed variable PTM: 3
- Click OK.

Next click the “Quantification” button. Select “Label Free” from the left hand panel and enter the following parameters:



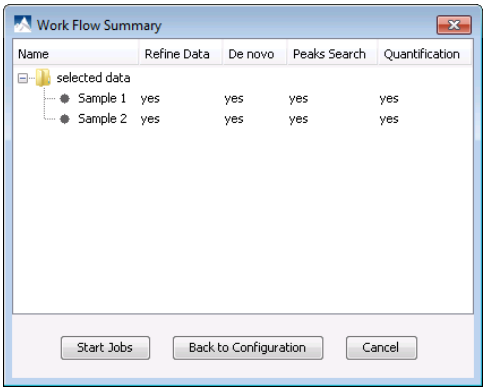
- Mass Error Tolerance of 0.2 Da
- Peptide Score: 50 – 100
- Retention Time Range: 4.0 min
- Protein Score; 50-100
- Upper Bound of Precursor Charge: 4
- Do Normailization: Check

Automatically the parameter table will be configured for both samples and fraction numbers.

If it is a project for replicate experiments, you can click “Assign replicate” to add replicate information. Each label free quantification needs at least two samples and a replicate project contains at least four samples. You can also change the number of replicates. Currently the maximum number of allowed replicates is three.

After assigning replicates, PEAKS will generate parameter tables for each replicate automatically. The default setting is the same for each replicate and you can change the settings for a replicate by click the name of the replicates and modify the parameter table.

Click “OK” and then click “Start” in the workflow configuration panel. The following window will open:



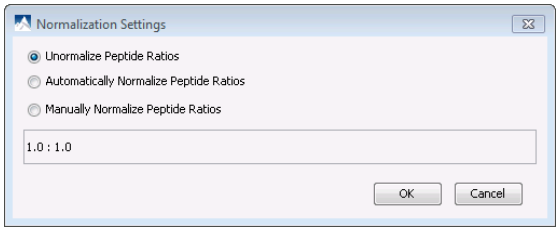
Click “Start Jobs”.

11.4.4 Label Free Results

Once the workflow analysis is complete, a label free quantification result node will appear on the project view.

You can double click on the label free node to open the result panel or right click on it to find more operations supported on a label free quantification result.

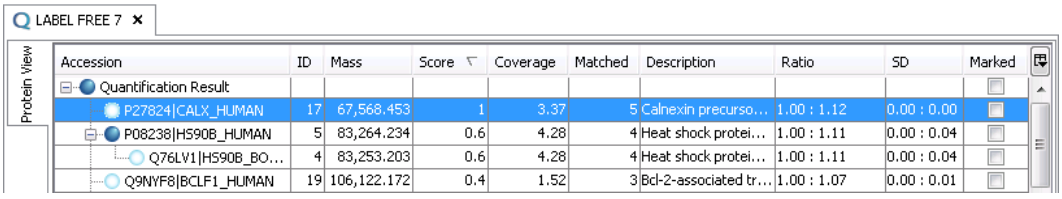
We support exporting the label free quantification results to Excel or HTML file by right clicking the result node and choosing the corresponding function. Please refer to section 15 Exporting Data/Reports and Printing for details. We also support changing the normalization factor in calculating the protein ratio. After clicking “Normalization settings” in the right click pop up menu, the following dialog will pop up:



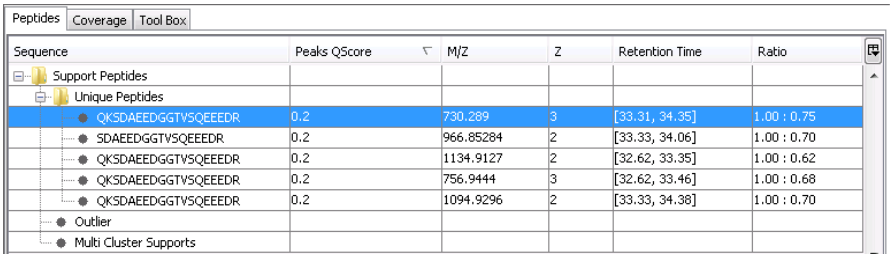
If you uncheck the “do normalization” option in configuring label free quantification, the protein ratio will be calculated from peptide ratios without normalization. Otherwise, we will normalize it and the normalization factor will be displayed. You can also set the normalization factor manually by clicking the “Manually Normalize Peptide Ratios” and input the ratios in the text field. The format of ratios should be digital numbers separated by colon and the number ratios should be the same as the number of samples in the quantification.

After double clicking the result panel, the quantified proteins, supporting peptides of each protein and peptide features in the spectrum from each sample will be displayed in the result panel. The quantified proteins will appear in the top panel, with homologous proteins clustered together. The ratio of Sample 1: Sample 2 appears in the “Ratio” column and the standard deviation of Sample 1: Sample 2 appears in the “SD” column.

In the example below the standard deviation is 0 as there is only one supporting peptide for the protein shown.



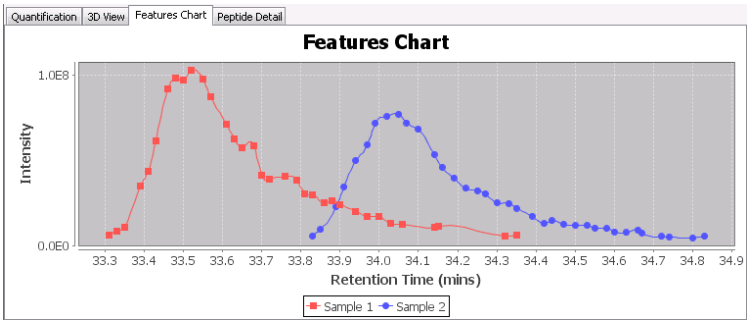
The supporting peptide is shown under the “Peptides” tab. The retention time is shown for the specific peptide as well as the peptide ratio from Sample 1: Sample 2.



Click on the “+” beside the “Outlier” folder to see the peptides that were not included in the ratio.

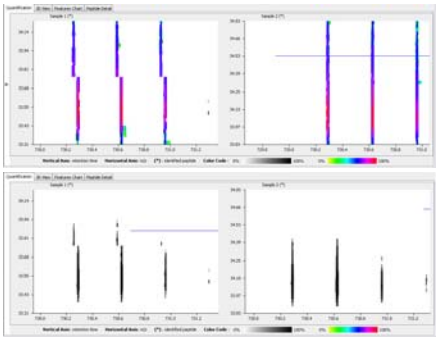
To see which peptides were used to identify the protein during the PEAKS protein ID search, select the “Coverage” tab. The entire sequence of the protein is shown and the matching peptides are highlighted in blue. In this example the total matched part accounts for 9.392% of the protein. This information can be found in the “Coverage” column in the “Protein View” panel.

The features chart will appear by default in the bottom panel:



To see exactly where the selected supporting peptide corresponds to the protein sequence, select the “Peptide detail” tab. The entire sequence of the protein is shown and the selected supporting peptide is highlighted in blue.

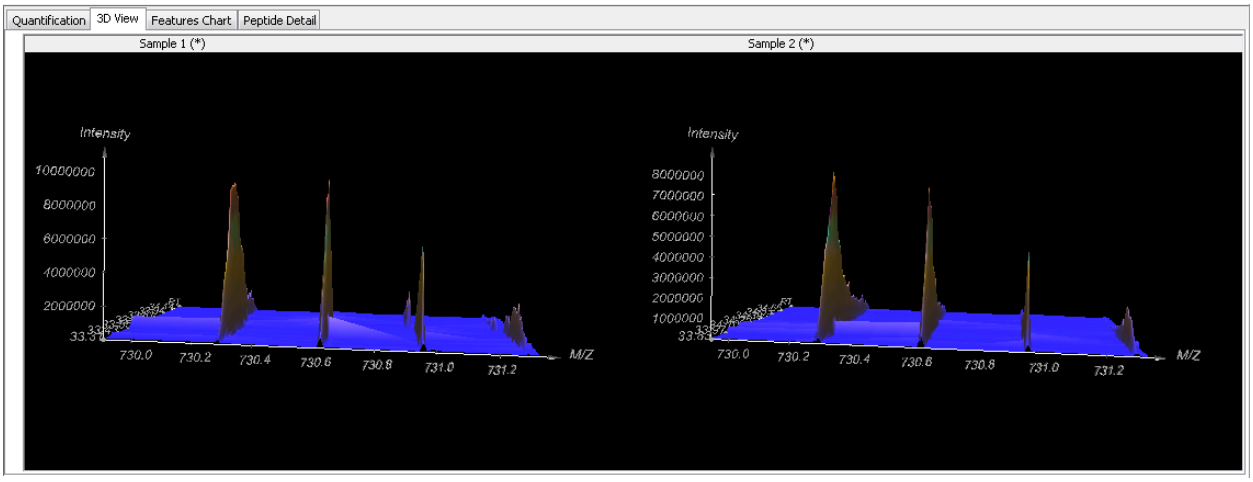
Clicking on the “Quantification” tab will display a 2D view of the features. Move your cursor around the map to display the m/z ratio and retention time.



Right click your mouse and select “Show features” to display a 2D heat map. When viewing the heat map in color, red represents high intensity and yellow represents low intensity. The grayscale heat map displays high intensity in black while white represents low intensity. If the peptide is identified in the PEAKS Protein ID, there will be a star after the sample name.

When there are more than nine samples in the project, we will display features of the first nine samples as default. If you want to display features of specific samples, click the triangle on the left of the quantification panel and a selection menu will pop up as follows. You can check the samples you are interested in viewing.

Click on the “3D View” tab to display a 3D View of the peptide features for sample 1 and sample 2. Intensity is displayed on the y-axis, m/z on the x-axis and retention time on the z-axis. Click and move the cursor to rotate around the image. Notice that as you move one sample image the other sample moves to the same location.



Chapter

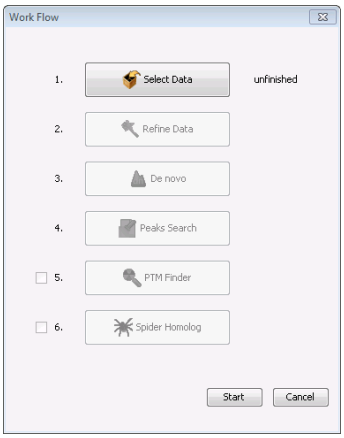
12

12 Creating a High-Throughput Workflow

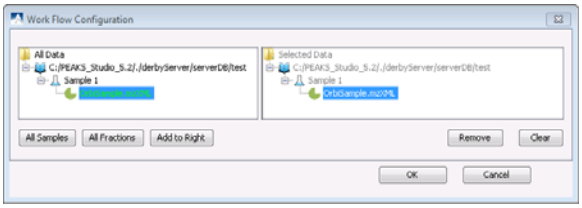
For your convenience, PEAKS provides workflows for protein identification, quantification and inChorus search (multi-engine protein ID).

12.1 Identification Workflow

Click the workflow icon on the toolbar and select "identification", the identification workflow configuration window will appear:



Click "Select Data" to select the data you wish to perform identification analysis. Select a data file, and click "Add to Right" to transfer the samples/files to the "Selected Data" panel on the right hand side. Use the "Remove" and "Clear" buttons to remove the selected file, respectively from the "Selected Data" panel. Click "OK" to proceed to the next step.



Check the boxes beside the functions that you would like to perform within the workflow. Note that PEAKS requires de novo sequencing results to perform Protein ID.

Then, you can define the parameters for data refinement, de novo, PEAKS search, PTM Finder and SPIDER. Please refer to the chapters on each individual function if you require more details on setting up the parameters.

Note that PTM Finder and SPIDER are optional functions. You can uncheck them if you do not want to perform those functions.

12.2 Quantification Workflow

The quantification workflow is similar to the identification workflow. Except after Protein ID, you can define quantification parameters to perform labeled or label free quantification.

12.3 inChorus Workflow

inChorus workflow is similar to identification workflow except that at the third step, you can specify inChorus parameters to start a inChorus search.

12.4 Batch Workflow

In some situations, you may have many data sets that you wish to process all at once, with the same parameters. PEAKS allows you to do this type of work using a workflow. By setting up different workflows, you can start a batch process of several projects and perform de novo, protein ID, SPIDER searches, as well as quantification without having to run each step individually.

It is important to note that all the files you load in a single workflow will be processed in exactly the same way, using exactly the same parameters. If you want to do some differently than others, you must set up separate workflows.

Click “Select Data” button.

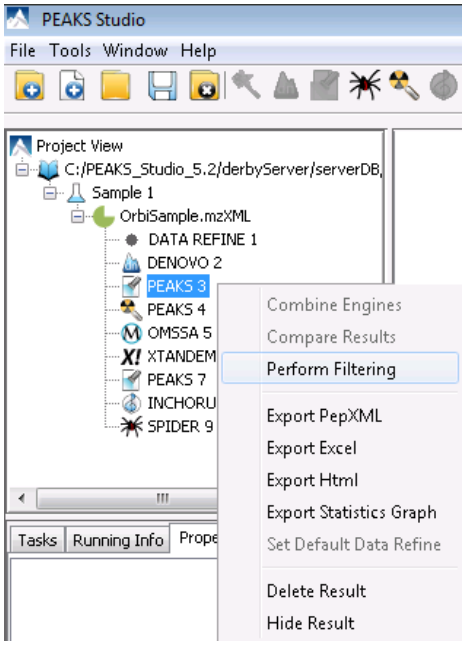
Only projects that are open in the “Project View panel” can be selected for analysis with the workflow function. To select which files/samples you would like to analyze, select the individual file/sample or click the “All Samples” or the “All Fractions” buttons and then click “Add to Right” to transfer the samples/files to the “Selected Data” panel on the right hand side. Use the “Remove” and “Clear” buttons to remove selected files/samples or all files/samples, respectively from the “Selected Data” panel. Click “OK” to proceed to the next step.

Chapter

13

13 Filtering Your Results

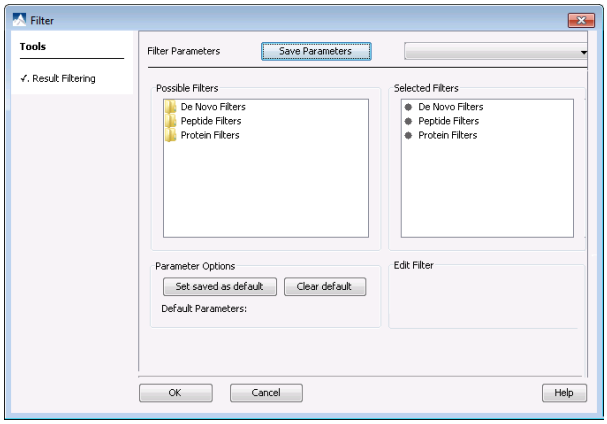
PEAKS 5.2 provides you with an exhaustive list of all proteins and peptides that can be identified in a sample. However, as many researchers have their own criteria of what information is required in a report as well as what is an acceptable result, PEAKS 5.2 provides the necessary tools that enable filtering out the less critical information, leaving just the essentials.



13.1 Setting Filter Parameters

Double-click on the time and date stamp associated with the result that you would like the filter. Once the report loads, click the right button on your mouse and select "Perform Filtering".

The following window will appear:



PEAKS will automatically create a new filtered results node in the project view that will save your filter choice every time you click "OK".

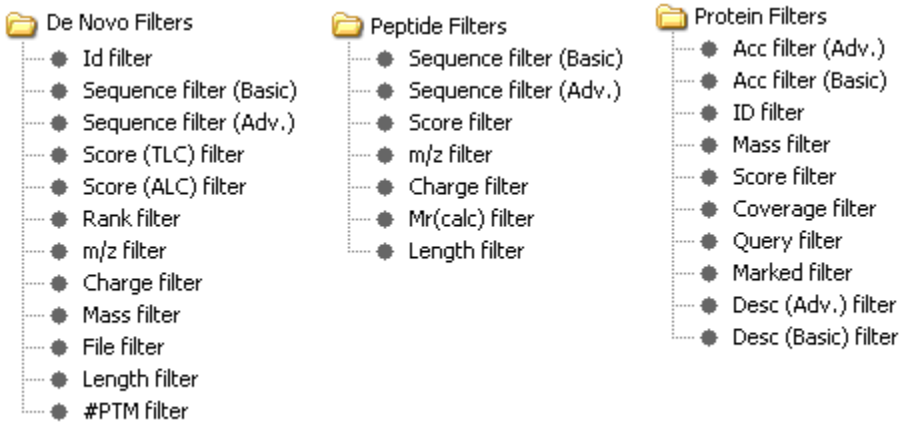
Note: The filter node relies upon the original result so DO NOT DELETE the original result in order to keep all associated filtered result nodes.

13.1.1 Possible Filters/ Selected Filters/ Edit Filter

The filters are grouped into three basic types. Each type identifies the level on which the filter will be performed:

- de novo filters act to remove proposed de novo sequences
- Peptide Filters act to remove peptides found in the database from the report
- Protein filters act to remove proteins from the report

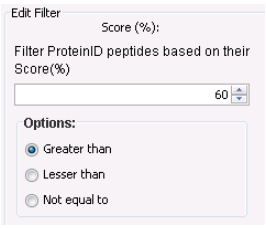
To see the available filters for each level of filtering, double-click on the appropriate folder in the "Possible Filters" frame. Here you can see a large variety of built-in filters:

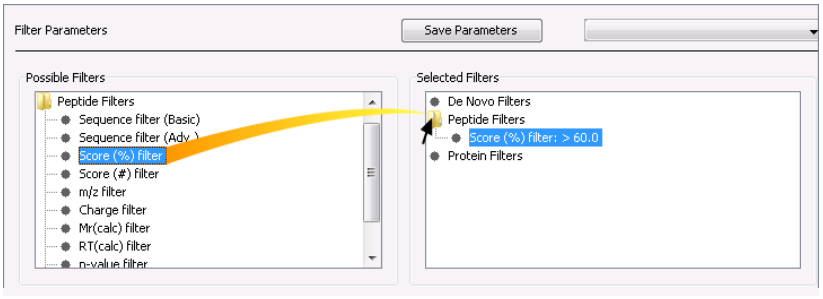


Choose a filter from the 'Possible Filters' list on the left by clicking on it. Options for this filter will appear in the "Edit Filter" frame. Once you have set the options for the filter in the "Edit Filter" frame, click and drag the filter into the "Selected Filter" list on the right hand side. Ensure that you move the filter to the appropriate de novo, peptide or protein node in the selected filters pane. Click "OK" to apply the filter(s) that you have selected to the current file.

You can add multiple filters by simply repeating the process outlined above. In this way it is also possible to have two filters with the same property; i.e. we can set a range of protein mass by applying one filter on the upper bound of the mass and adding a second filter to be the lower bound of the mass.

Example: Let's say that you want to show only proteins with more than one high scoring peptide (greater than 60%) a standard requirement for publication. Double click on the "Peptide Filters" folder. Select "Score Filter". Edit the filter to select peptides that have a score that is greater than 60% in the "Edit filter" frame. Now click and drag "Score filter" from the "Possible Filters" frame to the "Selected Filters" frame.





13.1.2 Parameter Options

Filter sets can be saved and re-used between sessions by clicking the "Save Parameters" button that is found at the top right hand corner of the "Filter Parameters" window.

You may prepare your results the same way each time; in which case it makes sense to set up a filter that will be automatically applied each time we load a report. Select a filter from the list of saved filters from the drop-down menu found at the top right hand corner of the "Filter Parameters" window. Click on the "Set saved as default" button. This filter will be displayed in the "Parameter Options" frame (as seen below) and will be applied automatically just after a report is loaded. Be careful, if your default filter is very stringent, it can sometimes remove everything! To remove a default filter, press the Clear Default Button at the bottom of the Filter Pane.

Peptide filters and protein filters can be applied to a protein id result at the same time, potentially making filtered results confusing. The basic principle for two different types of filters working together is:

- If a protein is filtered out by the given protein filter, all of its unique supporting peptides will also be filtered out
- If all of the supporting peptides of a protein have been filtered by the peptide filter, this protein will also filtered out.

We will use some samples to illustrate the usage of complex filters:

1) Goal: Show proteins that have two high-scoring hits:

- Add the Protein Filter called "Query" and in the "Edit Filter" section choose 'greater than' and type '1' in the box (without the quotes). This will remove any 'one hit wonders'.
- Add the Peptide Filter called 'Score' and in the "Edit Filter" section choose 'greater than' and type '50' in the box (without the quotes).

2) Goal: Find a protein that contains the word 'human' or 'rat' in the database entry's description, but not Keratin or Trypsin.

- Add the Protein Filters called 'Desc'
- In the "Edit filter" section, you are required to type in a regular expression (regex). This allows you to use wildcards.

Wildcard	Meaning	Example
.*	*"Anything of any length"*	**.*human.*** Will find anything that contains the word 'human', with anything before and anything after.
|	*"Or" (use brackets)*	**.*(human|rat|RAT).*** Will find anything that contains the word 'human', or the word 'rat' or the word 'RAT', with anything before and anything after.
?!	*"Not" (use brackets)*	**(?!.*(Keratin|Trypsin).*).*(human|rat).*** will find anything containing human or rat but not Keratin or Trypsin
[ ]	*"Any of these characters"*	**.*([Hh]uman|[Rr]at).*** will find anything containing the words Human, human, Rat or rat.

- Type in the regex: (?!.*(Keratin|Trypsin).*).*(human|rat).* and press the Enter key. If PEAKS confirms that this is a valid regular expression, it will put a check in the 'Valid Java Regex' box .

3) Goal: Setting a protein mass range

If we know the approximate mass of the proteins you are interested in, you can eliminate all proteins that are not close in mass.

- Add two filters: "Protein Filters: Mass >12000" and "Protein Filters: Mass < 32000".

Chapter

14

14 Comparison Functions

14.1 Overview

In liquid chromatography-mass spectrometry (LC-MS) based proteomics, multiple samples from different groups are analyzed in parallel. Tools that access the quality of proteomics data based on sound statistical principles are needed in this field. In PEAKS 5.2, comparison functions are provided in three levels:

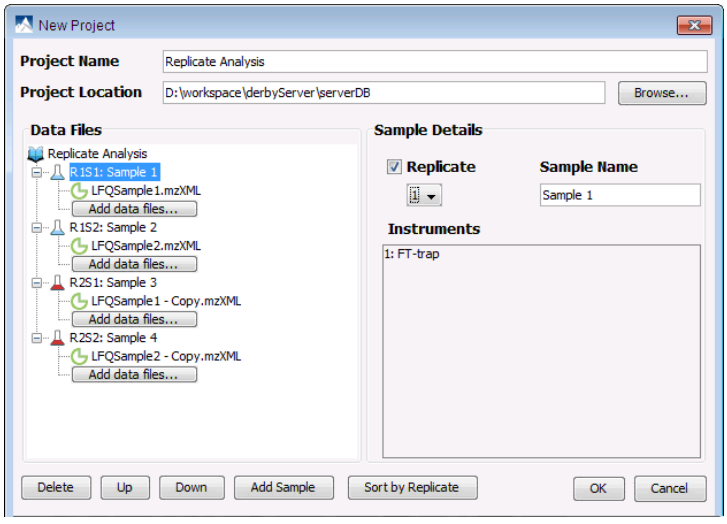
- Assess the reproducibility of MS data from technical replicates
- Perform compare analysis of peptide and proteins
- Assess the reproducibility of protein quantification from biological/technical replicates

This chapter is organized to first introduce how to create a replicate project. The replicate analysis of MS data comparisons and label free quantification are done together and so each function will be introduced together in one section. Secondly, the comparison function of peptide and proteins will be given later in this chapter.

14.2 Creating Replicate Project

Replicate analysis of sample data and label free quantification results in PEAKS 5.2 can be done by assigning replicate samples when initially setting up the project. Alternatively, the replicate information can be assigned to samples and fractions when you set parameters for label free quantification.

As show in the screenshots below, if you check “Replicate” in the sample details and select 1, the selected sample will be assigned to the first replicate. After adding replicate information, a label will appear before the samples name indicating the replicate and the index of the sample. For example, R1S1 means the first sample in replicate 1. In this way, the samples with the same index (S1) of different replicates (R1, R2, etc.) should be technical or biological replicates of each other.



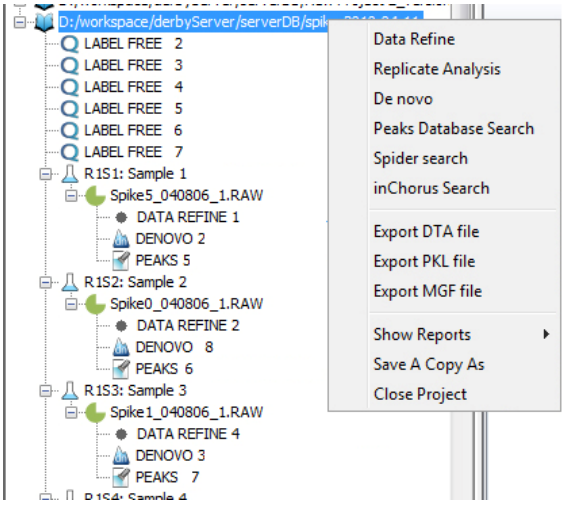
Currently, three replicates are allowed in a replicate project and the number of samples within different replicates should be the same. Different replicates can also be distinguished from the color of icons associated with each sample.

14.3 Compare Data and Label Free Quantification

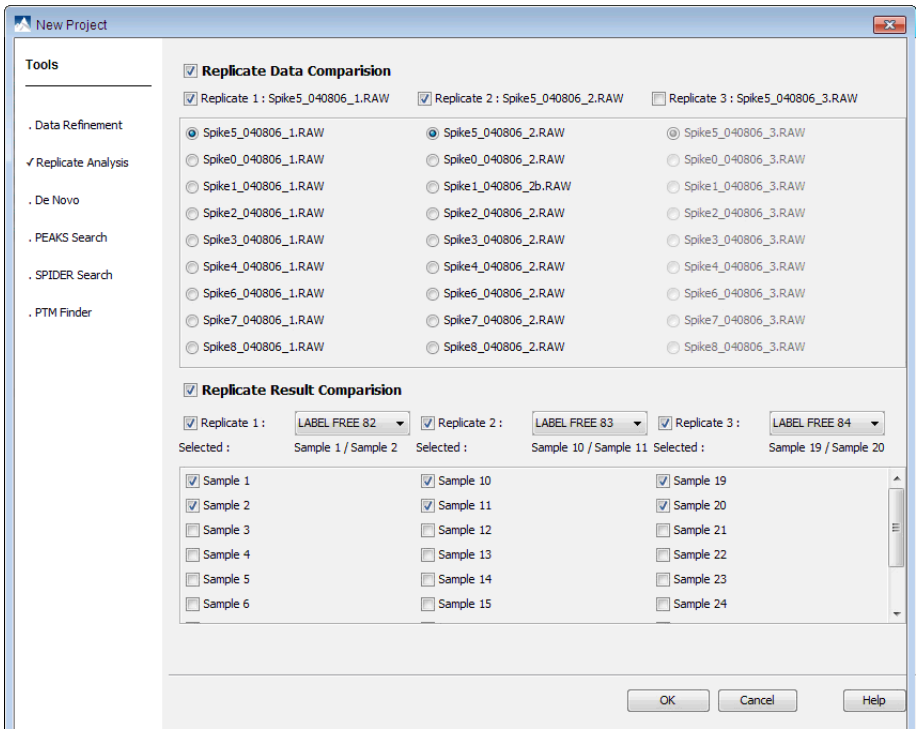
Replication is a standard for validating proteomics data and findings. The samples utilized within an experimental design often include technique and biological replicates. In PEAKS 5.2, we provide powerful replicate analysis tools for data and label free quantification results.

14.3.1 Parameter Setting

After acquiring label free quantification results, right click on the project node and select “Replicate Analysis”.



The following dialog will appear:



The analysis consists of replicate data comparison and replicate result comparison. We support analysis on at most two replicates in data comparison and three replicates in result comparison.

For replicate data comparison, you need to select one data file you want to compare from each replicate.

For replicate result comparison, you need to select one label free quantification result for each replicate and two samples you want to compare. Once you select a sample, all the samples with the same index in other replicates will be selected automatically.

14.3.2 Replicate Analysis Result

When a replicate analysis is done, a “replicate analysis” node will be added to the project and the results panels will open directly. The analysis will include the following information:

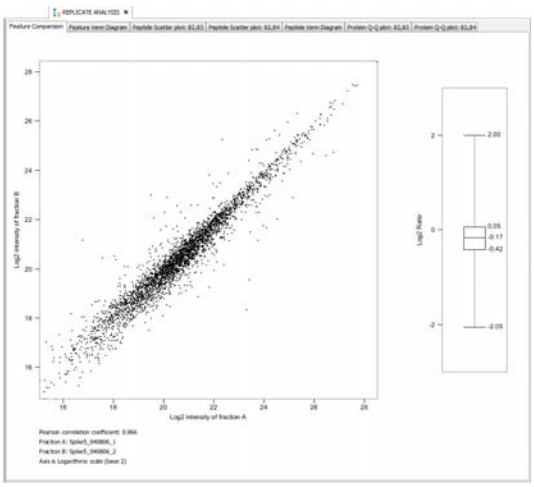
Replicate Data Comparison:

- Feature Comparison
- Feature Venn Diagram

Replicate Result Comparison:

- Peptide Scatter Plot
- Peptide Venn Diagram
- Protein Q-Q plot

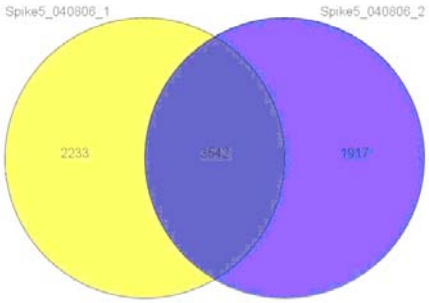
Feature Comparison Scatterplot:



The feature comparison scatter plot represents each feature vector which consists of two features detected in the two data files you want to analyze and aligned in the label free quantification. The x-axis is the log intensity of the feature detected in the first data file and the y-axis is the log intensity of the feature detected in the second data file. The Pearson Correlation Coefficient is calculated and listed under the chart. The standard box plot is shown on the right side of scatter plot.

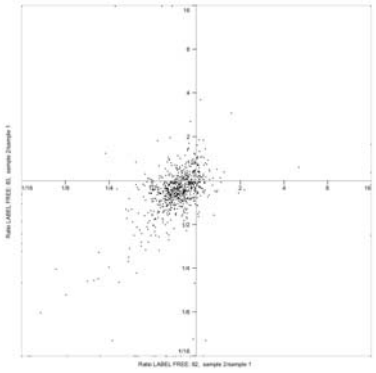
Feature Venn diagram:

The feature venn diagram is a standard venn diagram showing the number of common peptide features and unique peptide features of the two data files.



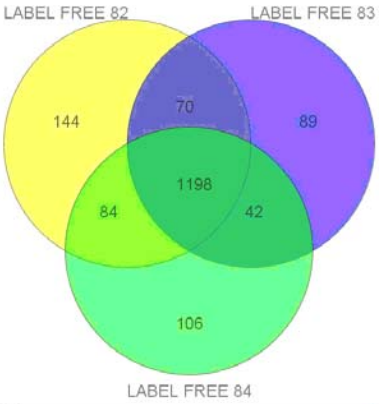
Peptide scatter plot:

The peptide scatter plot represents the peptides quantified in two label free quantification results in comparison. The x-axis is the ratio of the peptide of label free quantification result 1 and the y-axis is the ratio of the same peptide relative intensity ratios in corresponding samples of label free quantification result B.



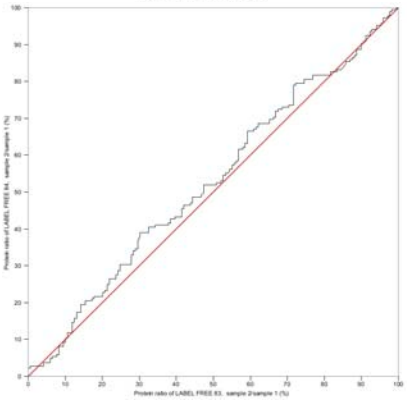
Peptide Venn diagram:

The peptide venn diagram is a standard venn diagram showing the number of quantified common peptides and unique peptides of label free quantification results in comparison



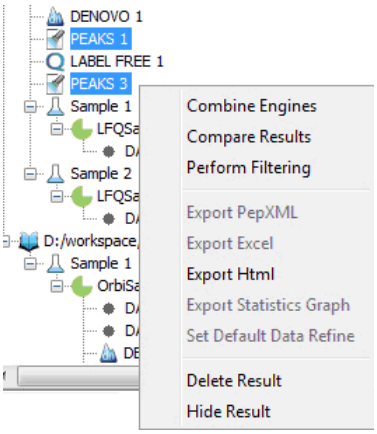
Q-Q Plot

The protein Q-Q plot is a standard quantile plot of protein ratios from selected samples of label free quantification results in comparison. The ratios of the proteins in the first sample is plotted against the ratios of the proteins in the second ratio, both in ascending order of size, and scaled from 0 to 100. In the ideal case, both replicates should result in the same ratios for the proteins, and thus the expected result is represented by the diagonal line in red.



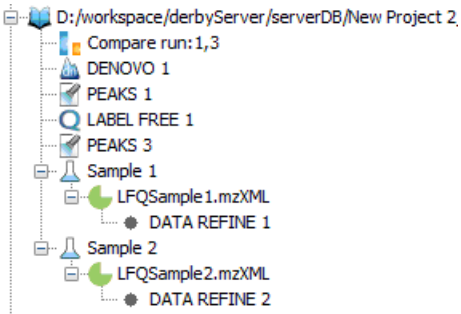
14.4 Compare Protein ID results

In PEAKS 5.2, we support comparisons on at most three PEAKS Protein ID results (including filtered results) in one project. To do such a comparison, select those PEAKS Protein ID nodes and right click. Click on “Compare Results” and the comparison will be done automatically.



14.4.1 Comparison Result

After comparison is finished, a comparison node will be added to the project as shown in the following picture.



The result panel will be opened automatically after completing the comparison. Since the comparison run is done on the fly it won’t be saved, it is suggested to export the results before closing the result panel. The details of exporting will be given in the next subsection.

The result consists of three parts: peptide comparison, protein comparison and statistical charts. Below is an outline of each.

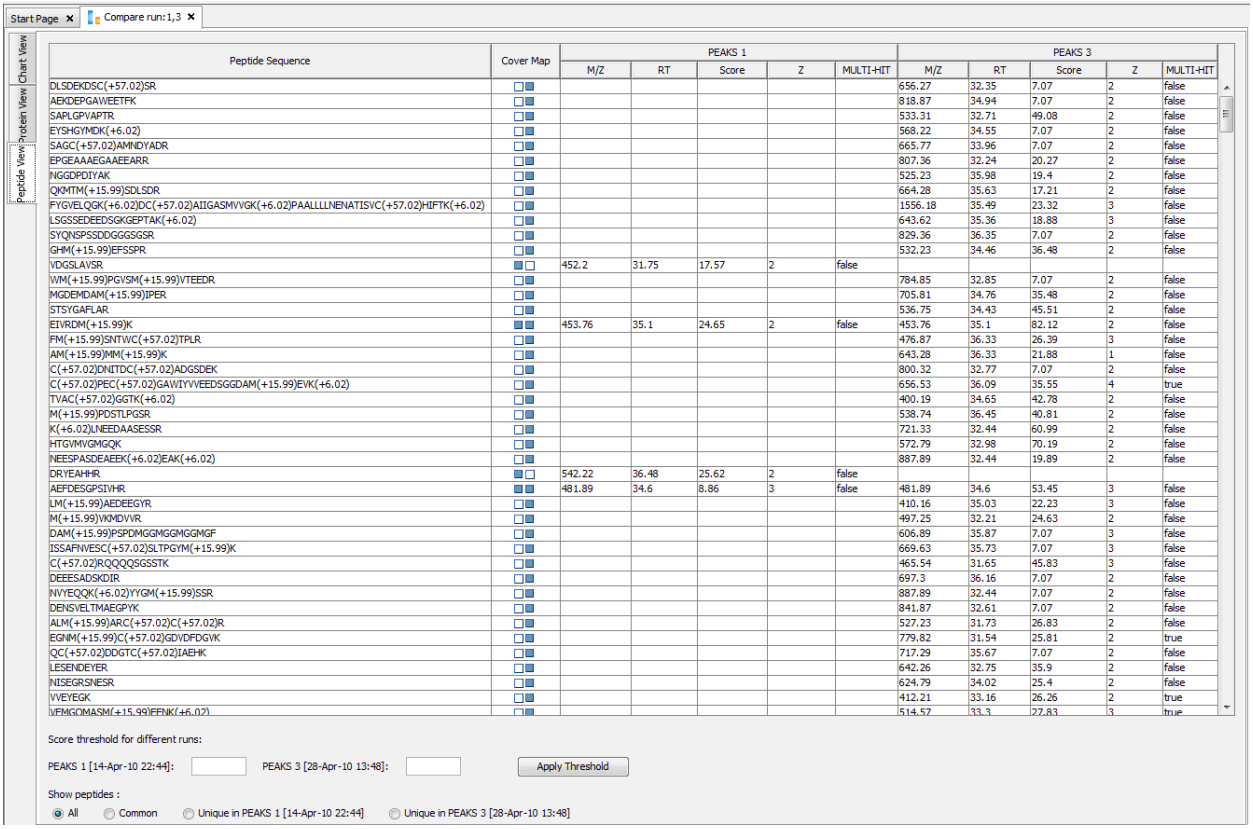
14.4.1.1 PEPTIDE COMPARISON

All the peptides identified in the two PEAKS Protein ID searches are displayed in the table. We show m/z, retention time, peptide score, charge and whether there are multiple hits for each peptide. The 'cover map' is a quick graphical illustration of the presence of the given peptide in one or both protein ID results. A solid icon indicates a successful detection of the peptide.

You can also select to show only the common peptides of those protein ID results, or the unique peptides of each protein ID result by changing the display settings at the bottom of the panel.

We also provide filters on the peptide comparison results. After inputting the PEAKS score threshold on each protein ID result and clicking "Apply Threshold" button, those peptides below the threshold will be filtered out.

The following screenshot is a typical peptide comparison result:



14.4.1.2 PROTEIN COMPARISON

The top proteins identified in the PEAKS Protein IDs are displayed in the table. The display setting, score filter and cover map function the same as in the peptide comparison frame. The following information is also displayed for each protein:

Score: PEAKS protein score

#Spec: the number of spectrum on which this protein has been detected.

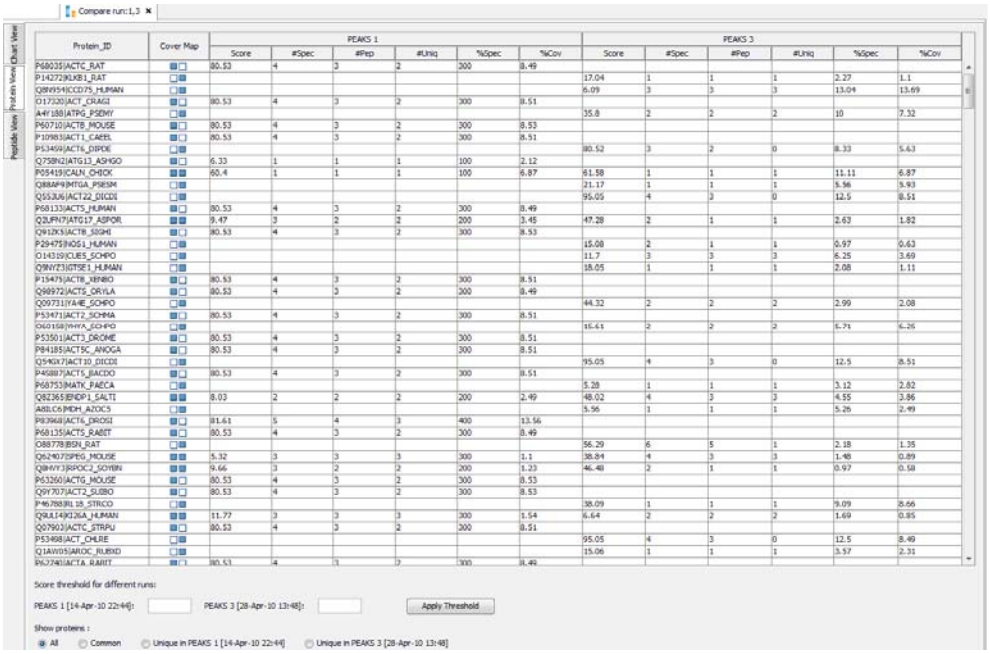
#Pep: the number of supporting peptides of the protein

#Uniq: the number of unique peptides of the protein

%Spec: the ratio of detected peptides to the theoretical numbers

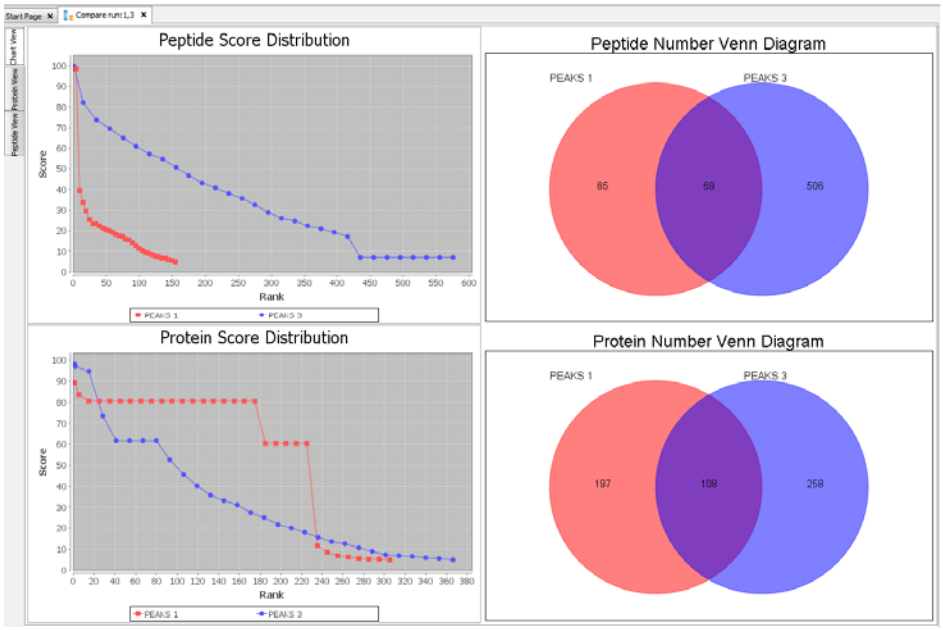
%Cov: the peptide coverage of the protein

The following screenshot is a typical results tab for protein comparisons:



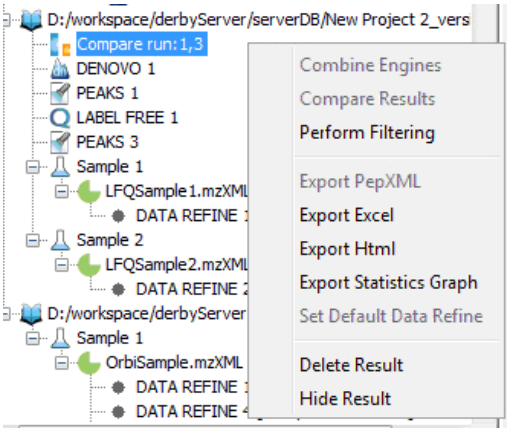
14.4.1.3 Statistical Charts:

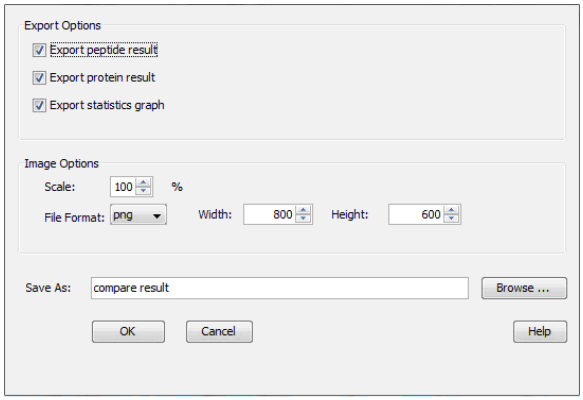
PEAKS 5.2 provides a number of statistical charts which are able to be exported for use in publications. The peptide score distribution, protein score distribution, peptide number venn diagram and protein number venn diagram for users to validate their results.



14.4.2 Exporting Comparison Results

To export the comparison results of PEAKS Protein IDs, please right click on the comparison run node and choose to export to Excel/HTML file. Here you can choose image quality and filter the content you want to export.





Chapter

15

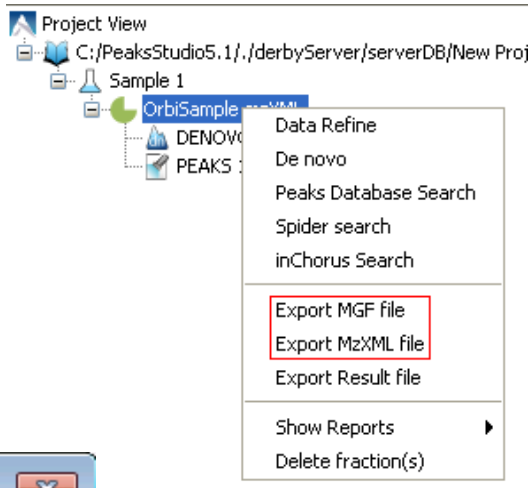
15 Exporting Data/Reports and Printing

PEAKS 5.2 allows you to create reports to share with collaborators, colleagues and clients. The reports are available in HTML or Microsoft Excel (.xls) formats and largely follow a 'What You See Is What You Get' philosophy while still adding information required to meet MCP (Bradshaw et al. 5 (5): 787. 2006) requirements. It is important that we complete results filtering and toggling columns before exporting a report.

Protein view exporting for PEAKS Protein ID and PEAKS LFQ both now have MCP exporting modes that attempt to collect as much information as possible to fit the MCP guidelines and make publication easier.

15.1 Export Sample Data

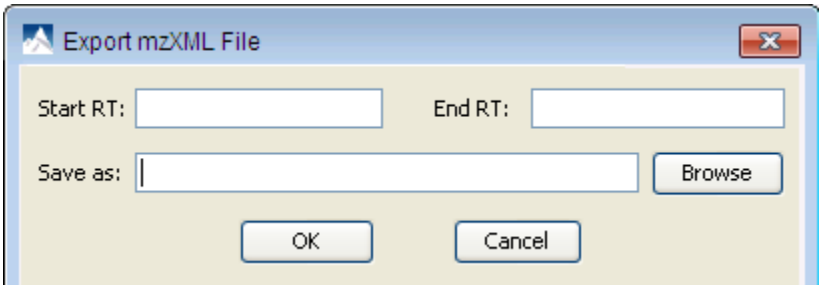
Data can be exported to a number of file formats including .mzxml, .mgf, DTA, and PKL. To do so, right click on the data file that you wish to export.



For example, click "Export MGF File". A window will open that will prompt you to enter a name and a location for the file. Click "Export".

OR

Click "Export MzXML File". The following window will open:



Enter the starting and ending retention times in the appropriate boxes. Then click the "Browse" button to select a destination to save your file.

15.2 Export Results of a fraction in Excel or HTML Format

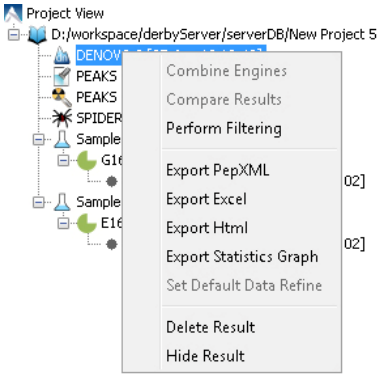
To export the results for a selected data file, right click on the data file and select "Export Result file". All of the PEAKS results file generated for the selected data file will be displayed in Excel or HTML format. This will create an output with the highest scoring result for all search engines sorted by spectrum.

15.3 Export Results in Excel or HTML Format

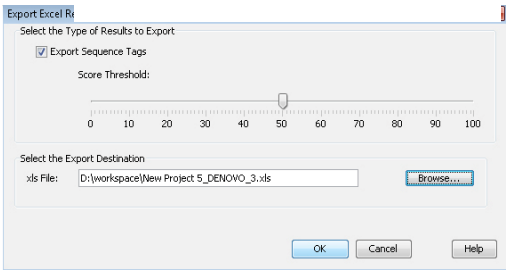
In PEAKS 5.2, all the de novo results, database search results, including PEAKS Protein ID, PTM finder, Spider, other search engines from inChorus, and quantification results can be exported to Excel or HTML Format.

15.3.1 Export de novo Results

To export the de novo results, right click on a de novo result node and choose “Export HTML” or “Export Excel”



The following dialog will appear:

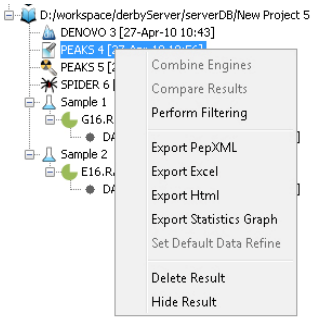


If the ‘export sequence tags’ option is selected, you can set a threshold for exporting your de novo sequences. Essentially, de novo residues with local confidences scores lower than the threshold will be displayed as mass gaps while de novo residues with local confidence scores higher than the threshold will be displayed as expected.

A typical exporting file of de novo results contains all the de novo peptide information: id, sequence, TLC, ALC, rank, m/z, charge (Z), mass, file, retention time, scans, quality and fragmentation mode. The fraction information and parameters during the de novo run will append at the end of the file.

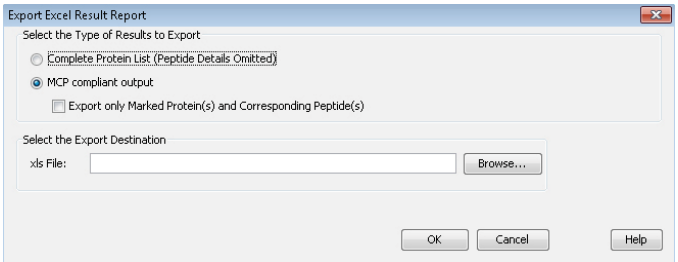
15.3.2 Export Database Search Results

To export the database search results are similar, only with slight differences in the results contents. Here we use PEAKS Protein ID as an example to demonstrate the exporting results.



For example, in order to export your PEAKS results file to Excel or HTML, right click on the results file that you wish to export and select “Export Excel”.

The following window will appear:



Here we provide two types of exporting functions: ‘complete protein list’ without peptide details, or ‘MCP compliant output’. When you select MCP compliant output, you can check the “Export only Marked Protein(s) and Corresponding Peptides(s)” if you are only interested in some proteins and previously marked them in the result table.

The output of “Complete Protein List” consists of two major sections, one is the representations table which display a representative protein for each cluster, and another is the whole protein table which lists all the clustered proteins.

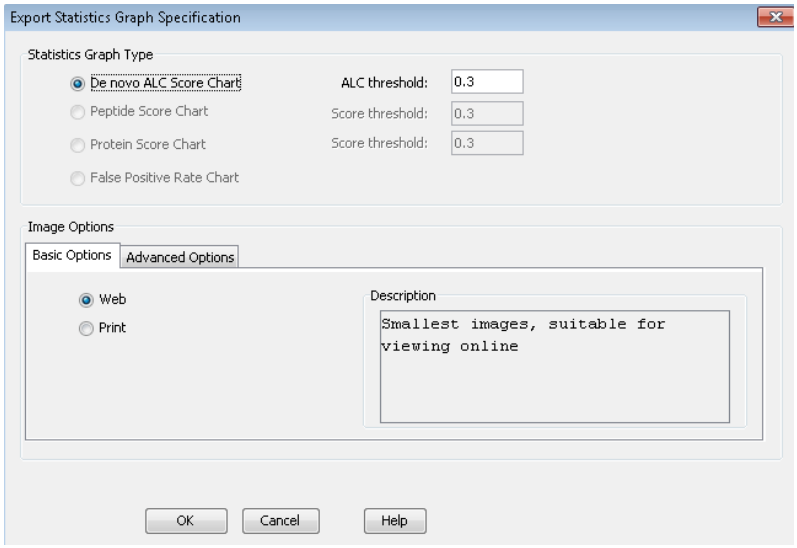
The MCP compliant output contains the two tables described above, however, it also provides more information in the whole protein table. These additions include all of the supporting peptides and their coverage within the protein. False discovery rate (FDR) estimation is also displayed if you have run PEAKS Protein ID search with decoy database. We also provide the ‘Single-Peptide Based Protein’ results table which contains all the proteins with only one supporting peptide detected.

15.4 Exporting Statistical Graphs

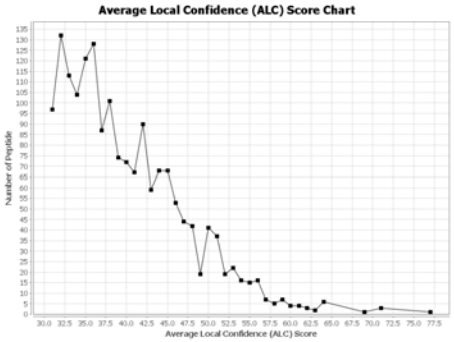
We provide exporting statistical graphs for de novo and PEAKS Protein ID which is useful in results validation and analysis.

15.4.1 Exporting Statistical Graphs on De novo result

Select a de novo results file. Right click and select “Export Statistics Graph”. The following window will appear:

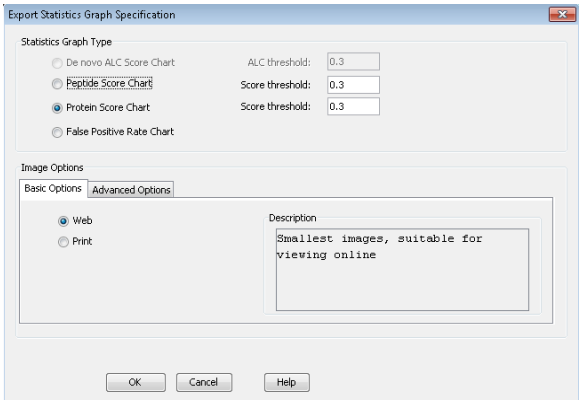


Currently PEAKS provides the de novo ALC score chart where the user can input the ALC score threshold and select the image quality for export. PEAKS 5.2 will export a broken line graph reflecting ALC distribution. The x-axis is the ALC score and the y-axis is the number of peptides. Each data point (x, y) in the graph means there are y peptides at score x.

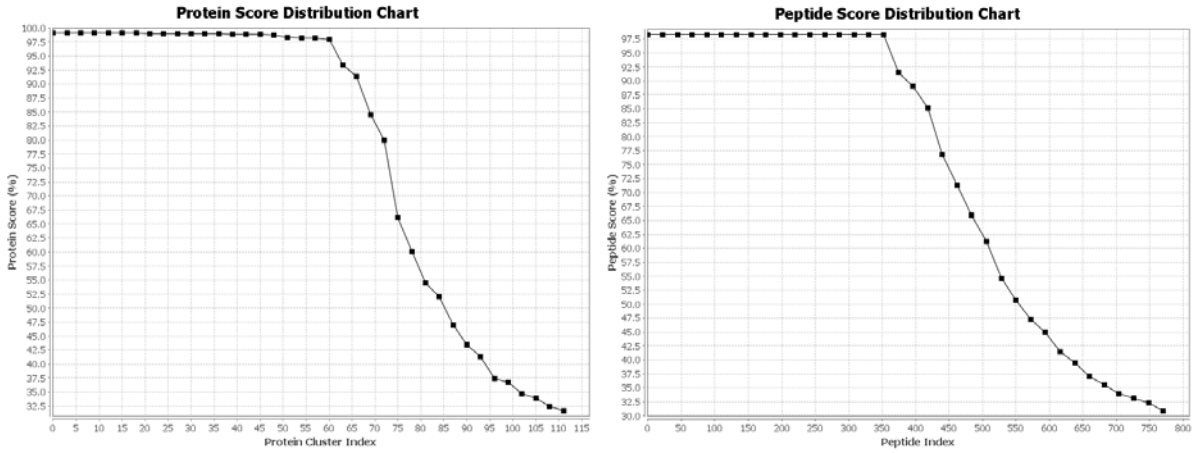


15.4.2 Exporting Statistical Graphs on Protein ID result

Select a Protein ID results file. Right click and select "Export Statistics Graph". The following window will appear, giving you the option of exporting a "Protein Score Chart", a "Peptide Score Chart" or a "False Positive Rate Chart":



Below are examples of the "Protein Score Chart" and the "Peptide Score Chart", respectively:



In the protein score distribution chart each data point (x, y), x is the rank of the representative of each cluster according to its PEAKS Score and y is the corresponding PEAKS score.

In the protein score distribution chart, for each data point (x, y), x is the rank of a peptide according its PEAKS Score and y is the corresponding PEAKS score.

Chapter

16

16 ADVANCED CONFIGURATION AND ENVIRONMENT PREFERENCES

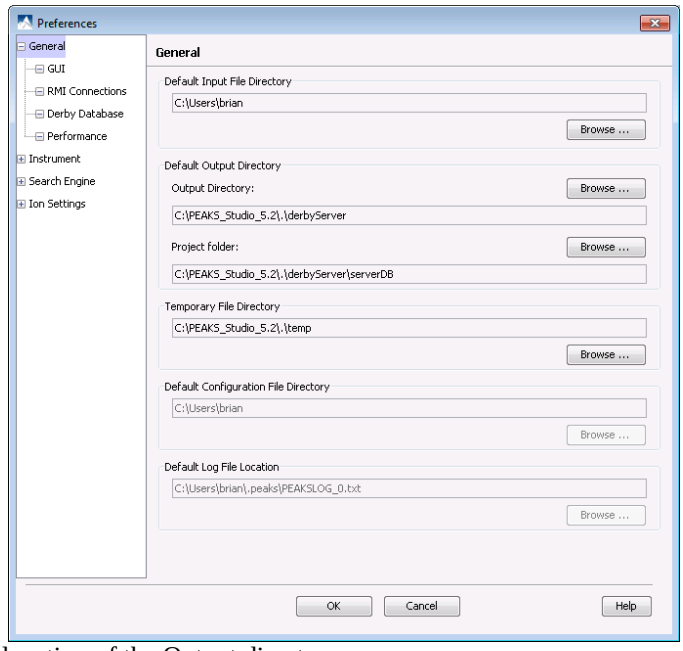
16.1 PEAKS ENVIRONMENT PREFERENCES

This section will describe the settings of the environmental preferences configuration including general, instrument, search engine and ion editor configurations.

To begin, click the Preferences toolbar icon  Or Select “Preferences” from the “Window” menu.

The following window will open:

Use the ‘+’ and ‘–’ boxes to expand and collapse the nodes.



16.1.1 GENERAL PREFERENCES

Default Input File Directory- Select the “Browse” button to change the directory that will appear when adding data to a project.

Default Output Directory- PEAKS outputs your results to C:\PEAKSStudio5.2\derbyServer by default. Select the “Browse” button to change this location. This directory will be used for all derbyServer data processing.

Project folder- PEAKS uses C:\PEAKSStudio5.2\derbyServer\serverDB as the default output folder for project files. Select the “Browse” button to change this location. This must be a subfolder of the Output Directory. If you wish to change the location where projects are saved you must also change the location of the Output directory.

Temporary File Directory- PEAKS uses C:\PeaksStudio5.2\temp as the default temporary file output directory. Select the “Browse” button to change this location.

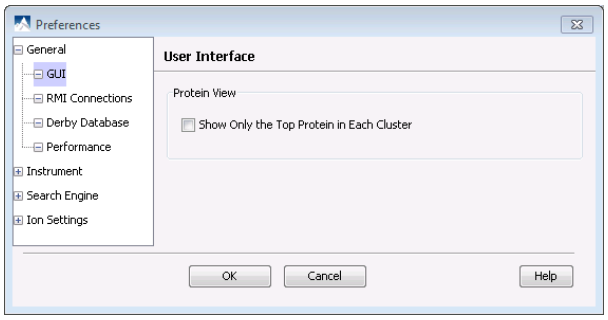
Default Configuration File Directory- Your configuration files for PEAKS can be found at C:\Documents and Settings\bsi\.peaks\peaksconf.xml by default. You cannot change this file’s location.

Default Log File Location- Your log file for PEAKS can be found at C:\Documents and Settings\bsi\.peaks\peaks.log by default. You cannot change this file's location.

16.1.1.1 GUI

Clicking on "GUI" on the menu on the left hand side will open the following window:

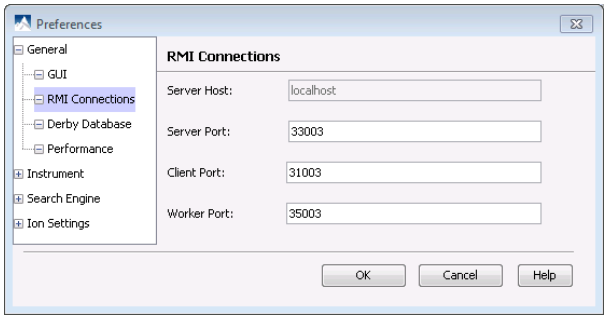
Choose "Show Only the Top Protein in Each Cluster" to remove subset protein identifications from the protein display.



16.1.1.2 RMI CONNECTIONS

Clicking on "RMI connections" on the menu on the left hand side will open the following window:

The default port numbers for the Server, Client and Worker will appear. The port numbers can be changed if conflicts arise. Contact technical support at BSI for more information.

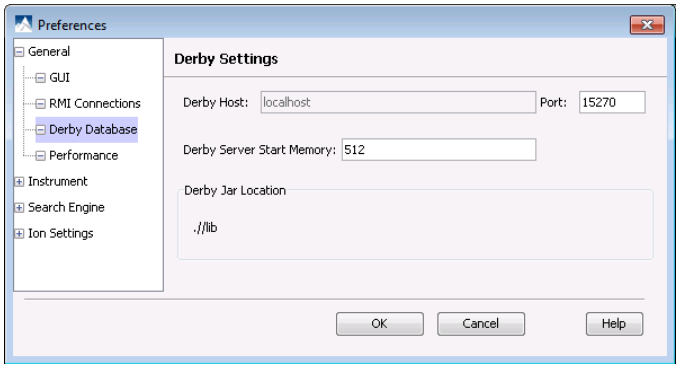


16.1.1.3 DERBY DATABASE

Clicking on "Derby Database" on the menu on the left hand will open the following window:

Derby Host

The name of the "Derby Host" as well as the "Port" number will come up by default. The port number can be changed.



Derby Server Start Memory

The amount of memory used to start derby server will also come up by default but can be changed if more memory is available. It is not recommended to change this from the default setting. If you would like to increase performance use the performance settings or the memory utility.

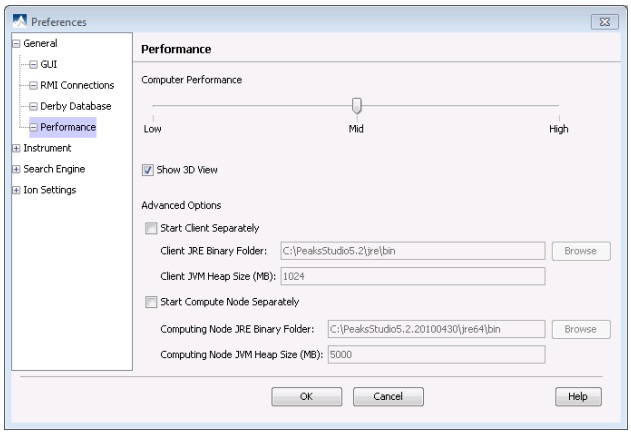
Derby Jar Location

The "Derby Jar Location" panel will list the location of the Derby Jar file by default. This is displayed to find its location. Its location cannot be changed.

16.1.1.4 PERFORMANCE

Clicking on “Performance” on the menu on the left hand will open the following window:

Computer Performance



Select low if you have between 512 MB and 1 GB of RAM. Select Medium if you have between 1 GB and 2 GB of RAM. Select High if you have 2 GB or more RAM.

3D View

PEAKS will display a 3D view with your quantification results. Check the “Show 3D View” box to enable this function. PEAKS 5.2 comes with the Java3D program to support the viewing of 3D images. If you don’t require this feature, deselect it to increase performance.

Advanced Options

These options are present for users who want to take full advantage of their 64 bit Windows operating system.

- Start client separately: Select this option to load the Client JRE separately. Select browse to choose the location of the Client JRE Binary folder. It is recommended that the Client JVM be a 32 bit JRE. The default setting is the 32 bit JRE which comes with PEAKS.

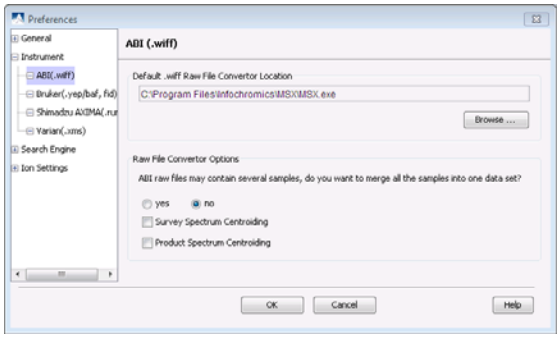
- Start Compute Node Separately: Select this option to load the Compute Node separately. For this, you can choose a 64 bit JRE downloaded from Java Sun by choosing the browse button. Select the location of the JRE binary folder. With a 64 bit JRE you can then toggle the Heap Space to any amount within the maximum physical memory of the computer.

16.1.2 INSTRUMENT PREFERENCES

This section will allow you to change preferences for the following instruments: ABI, Bruker, Shimadzu and Varian.

16.1.2.1 ABI (.WIFF)

Clicking on “Instrument” and then “ABI (.wiff)” on the menu on the left hand side will open the following window:



Default .wiff raw file convertor location

Click “Browse” to tell PEAKS the location of the Default .wiff raw file converter. MSX is a conmmercial .wiff converter developed by Infochromics is available from BSI.

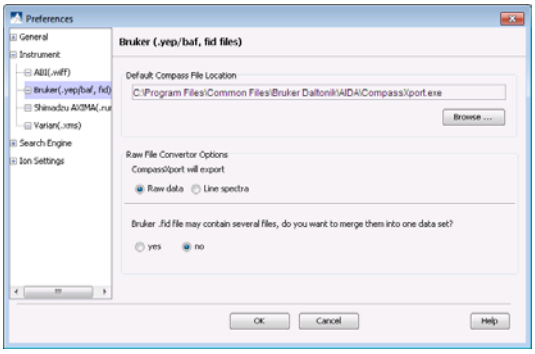
Raw file converter options

ABI raw files may contain several samples. By default, these samples are not merged into one data set. Select “yes” if you would like PEAKS to merge all the samples into one data set.

Select Survey Spectrum Centroiding if centroiding has been performed before loading the data into PEAKS. Select Product Spectrum Centroiding if centroiding has been performed on the Product spectrum before loading into PEAKS. This is important to insure PEAKS performs optimally.

16.1.2.2 Bruker (.yep/baf, fid)

Clicking on "Instrument" and then "Bruker (.yep/baf, fid)" in the menu on the left hand side will open the following window:



Default compass file location

Click "Browse" to tell PEAKS the location of the CompassXport file converter.

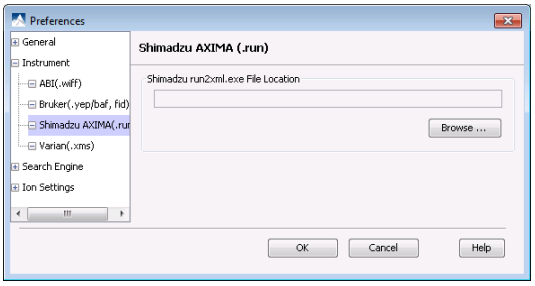
Raw file converter options

CompassXport by default will export raw data. If you attempt to load raw data and no spectra are displayed choose to export line spectra.

Bruker .fid files may contain several samples. By default, these samples are not merged into one data set. Select "yes" if you would like PEAKS to merge all the samples into one data set.

16.1.2.3 Shimadzu AXIMA (.run)

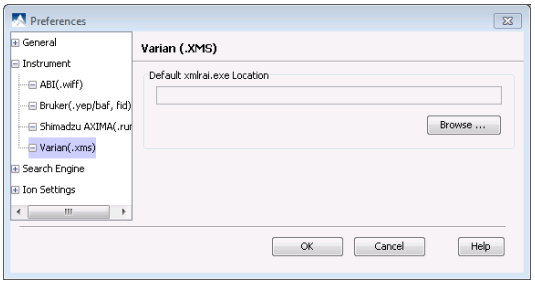
Clicking on "Instrument" and then "Shimadzu AXIMA (.run)" in the menu on the left hand side will open the following window:



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

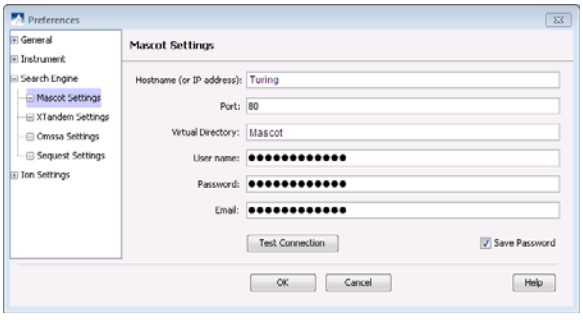
16.1.2.4 Varian (.xms)

Clicking on "Instrument" and then "Varian (.xms)" in the menu on the left hand side will open the following window:



Click "Browse" to tell PEAKS the location of the xmlrai.exe file.

16.1.3 Search Engine Preferences

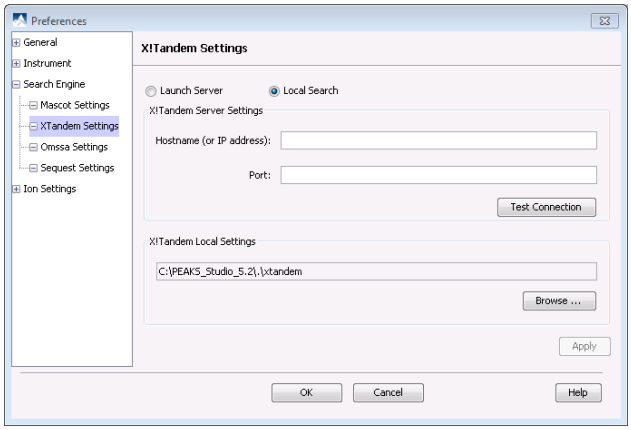


16.1.3.1 Mascot Settings

Clicking on "Search Engine" and then "Mascot Settings" on the menu on the left hand side will open the following window:

In this window you can tell PEAKS how to access your Mascot server (if applicable). Enter the Host name (or an IP address), Port, Virtual Directory as well as your user name, password and email address. To make sure that you entered everything correctly and that the server is working, click the "Test Connection" button. If you would like to save your password so that you don't have to enter it every time, check the "Save Password" box.

16.1.3.2 X!TANDEM SETTINGS

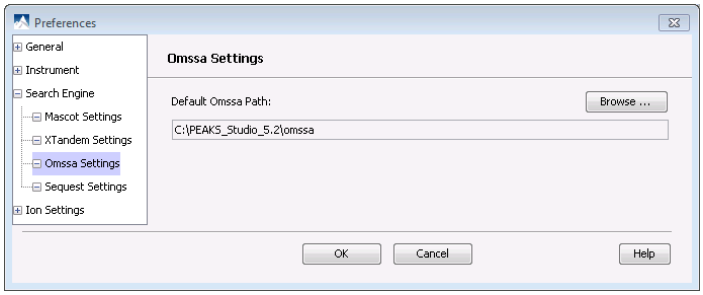


Clicking on "Search Engine" and then "XTandem Settings" in the menu on the left hand will open the following window:

Start by selecting whether you would like PEAKS to access a server or local version of X!Tandem. If you select the server version, you must enter the Host name (or IP address) as well as the port. To make sure that you entered everything correctly and that the server is working, click the "Test Connection" button.

As PEAKS provides a local copy of X!Tandem, upon installation, a default path will appear in the Local Settings section. If you wish to use another license/location for X!Tandem, click the "Browse" button to tell PEAKS where to find the search engine.

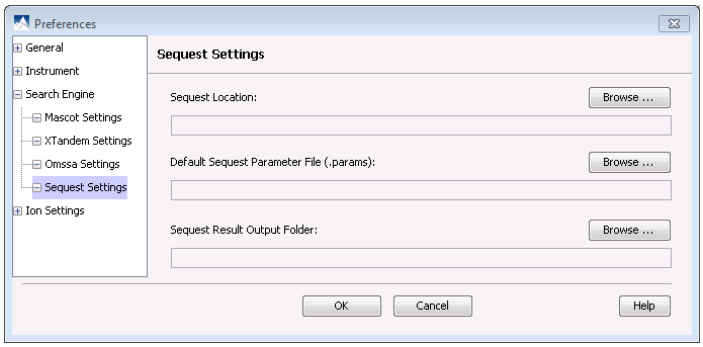
16.1.3.3 OMSSA SETTINGS



Clicking on "Search Engine" and then "OMSSA Settings" in the menu on the left hand will open the following window:

As PEAKS provides a local copy, upon installation, a default path will appear here. If you wish to use another license/location for OMSSA, click the "Browse" button to tell PEAKS where to find the desired path.

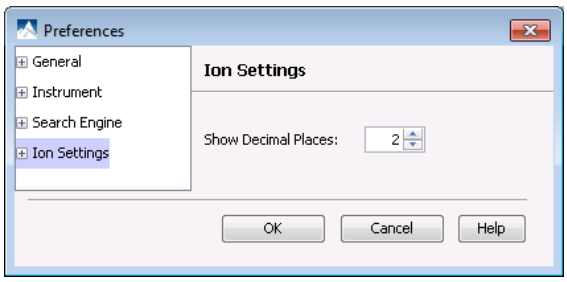
16.1.3.4 SEQUEST SETTINGS



Clicking on "Search Engine" and then "Sequest Settings" in the menu on the left hand will open the following window:

To use Sequest, click the "Browse" button to tell PEAKS where to find the search engine. You must also browse your computer to find the location of the "Default Sequest Parameter File (.params)" as well as the "Sequest Result Output Folder".

16.1.4 ION EDITOR PREFERENCES



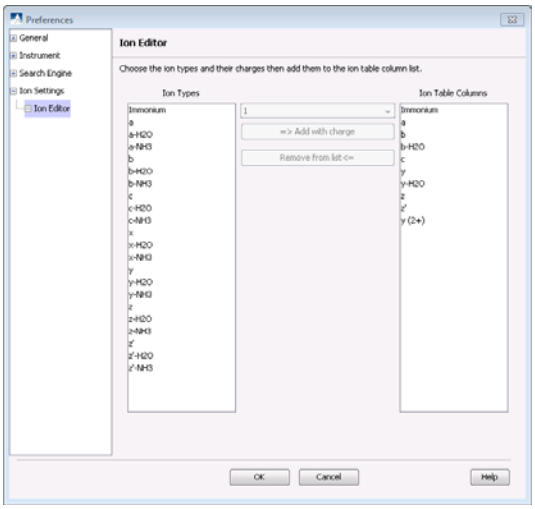
Clicking on “Ion Editor” on the menu on the left hand will open the following window:

Decimal places

Select the number of decimal places you would like to appear in the ion table. The default is set to two decimal places.

Ion Editor

Clicking on “Ion Settings” and then “Ion Editor” in the menu on the left hand will open the following window:



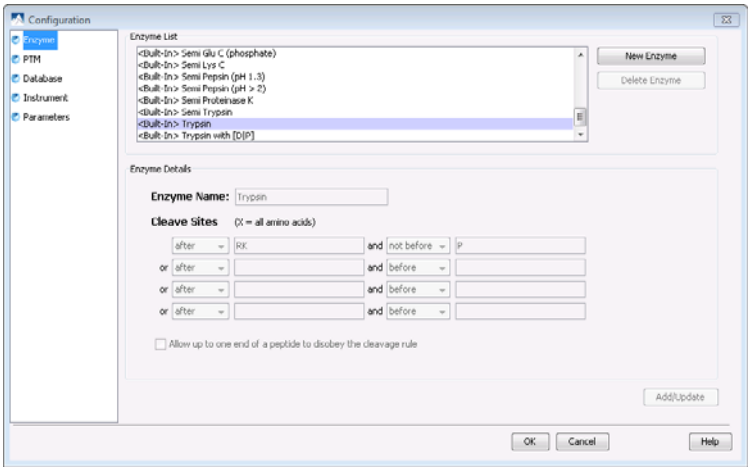
To select an ion type to be viewed in the ion table, click on the ion type in the “Ion Type” list found near the left hand side of the window. You now need to select the charge for that ion type from the drop-down menu. Once you have done this, click the “=>Add with charge” button and the ion type will now appear in the “Ion Table Columns” list on the right hand side of the window. To remove an ion type from the “Ion Table Column” list, select the ion type and click the “Remove from list <=” button. The ion type will now appear in the “Ion types” list.

16.2 PEAKS CONFIGURATION

This step includes configuration of enzymes, PTMs, databases, instruments, and parameters.

To begin click the Configuration toolbar icon  . Or Select “Configuration” from the Windows” menu.

16.2.1 ENZYME CONFIGURATION



PEAKS can use almost any enzyme or combination of enzymes in your analysis. You can select built-in enzymes from the extensive list provided in PEAKS or define your own. From the “Configuration” window select “Enzyme” from the left hand to change your enzyme configuration.

Built-in enzymes

All of the built-in enzymes within PEAKS are listed in the “Enzyme list”. Clicking on one of these built-in enzymes will display the information about that enzyme in the “Enzyme Details” panel. Note that you cannot delete or change the details of a built-in enzyme and therefore the “Delete Enzyme” button will be grayed out.

Create a new enzyme

Click on the “New Enzyme” button to create a new enzyme.

Enzyme Details: This is how you specify where the custom enzyme will cleave the protein between two amino acids to create peptides. The letter X denotes ‘any amino acid in this position’, while {set brackets} indicate any

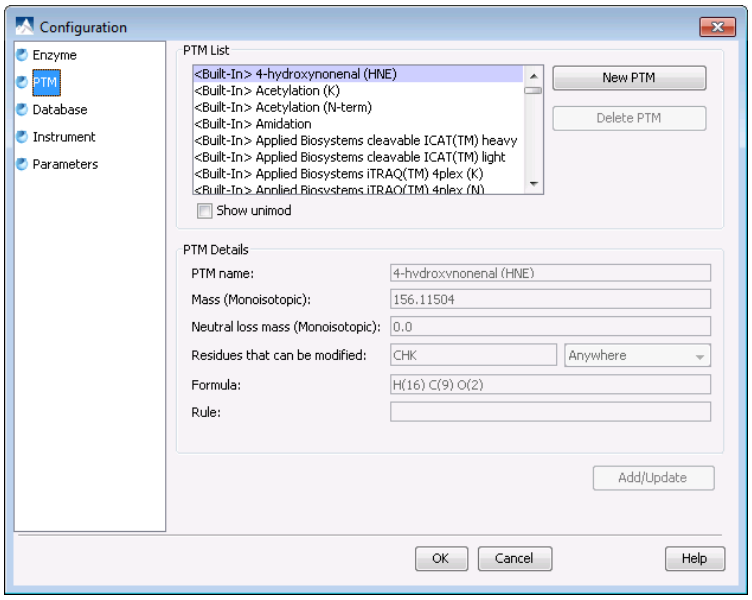
amino acid except the one in the brackets. You can also choose where the cleave sites are by choosing 'after' or 'not after' and 'before' or 'not before' to specify the range. There is also an option to "Allow up to one end of a peptide to disobey the cleavage rule".

You must click the "Add/Update" button for the changes to be saved. Your new enzyme will now appear in the "Enzyme List" where you can access it later. If you wish to delete an enzyme that you created, select the appropriate enzyme and click the "Delete Enzyme" button.

Note: For information on defining new enzymes "on the fly" for PEAKS de novo or PEAKS Protein ID, refer to section 6.1.2 Enzyme Options or 7.1.2 Enzyme Options, respectively.

16.2.2 PTM CONFIGURATION

From the "Configuration" window select "PTM" from the left hand to change your PTM configuration.



Built-in PTMs

The built-in PTMs within PEAKS are listed in the "PTM List". To see additional built-in PTMs from the Unimod library, click the "Show Unimod" box. Clicking on one of these built-in PTMs will display the information listed about that PTMs in the "PTM Details" panel. Note that you cannot delete or change the details of a built-in PTM and therefore the "Delete PTM" button and the "PTM Details" panel will be grayed out.

Create a new PTM

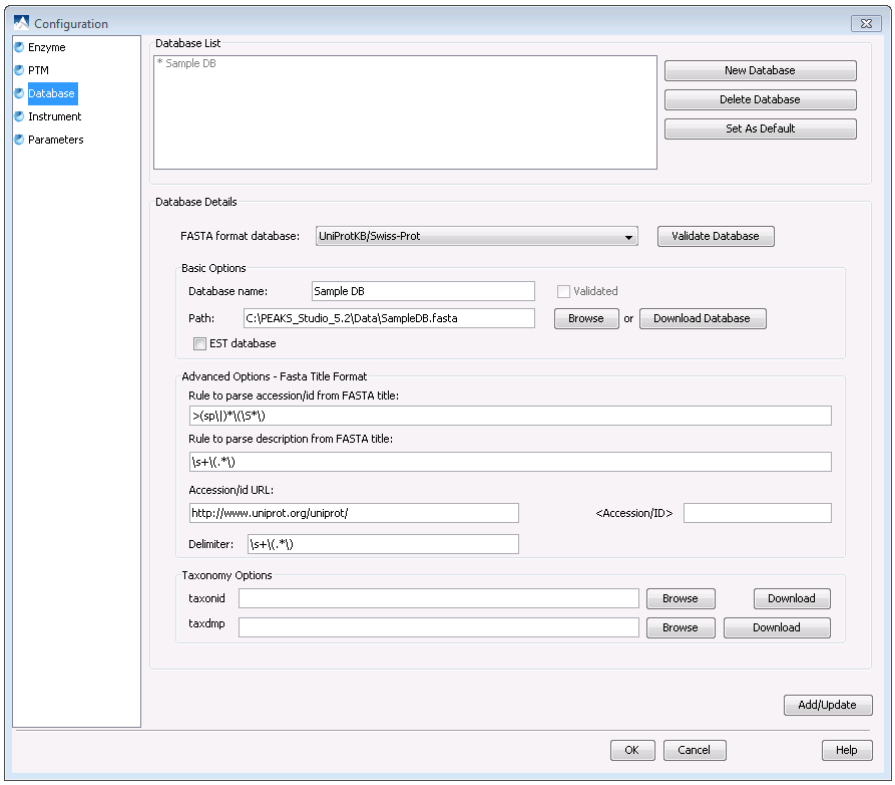
Click on the "New PTM" button. Now simply enter the information about your PTM in the "PTM Details" panel.

- PTM Name: this name will appear in the PTM list for future use after it is saved.
- Monoisotopic mass: the mass that the residue gains or looses as a result of the PTM.
- Neutral loss mass: the mass that the modified residue looses as a result of fragmentation in Daltons. For example, 28 would signify a loss of 28 Daltons.
- Residues that can be modified: Enter residues that can be modified anywhere, residues that can only be modified if they are at the N-or C-terminus or in the middle only.
- Chemical formula: the chemical formula of the PTM. This should correspond to the mass listed above.
- Rule: you can enter a comment for your reference.

You must click the "Add/Update button for the changes to be saved. Your new PTM will now appear in the "PTM List" where you can access it later. If you wish to delete a PTM that you created, select the appropriate PTM and click the "Delete PTM" button.

Note: For information on defining new PTMs "on the fly" for PEAKS de novo or PEAKS Protein ID, refer to section 6.1.3 PTM Options or 7.1.3 PTM Options, respectively.

16.2.3 DATABASE CONFIGURATION



In addition to de novo sequencing of peptides, PEAKS 5.2 also has the ability to search through a database search to identify proteins. In order to use this function, PEAKS must have access to a protein or EST database in FASTA format or an EST database of DNA sequences. You can point PEAKS to an existing database on your system, or download one. Additionally, you can specify taxonomy with certain databases.

From the "Configuration" window select "Database" from the left hand side to change your database configuration. The "Database List" at the top of the screen will show you databases that you have already configured. Select one of these files to see the details in the "Database Details" panel below.

Configure a new database

1) Select the "New Database" button on the right hand side of the "Database List". You will now be filling in the specifics for your database in the "Database Details" panel below.

2) FASTA Format Database: Select your database from this drop-down menu, or select "Other" if your database is not in the list or if you would like to submit your own database.

3) In the basics option panel, enter a name for your database and select "Download Database". A window will appear confirming the database chosen to be downloaded from the respective FTP or website.

4) If you would like to invoke your default FTP client software and download automatically, click "Yes". If you select "No" the following window will appear telling you that the URL will be copied to your system clipboard. Click "Ok".

Open your Internet Explorer and paste the URL into the address bar. A file download window will open. Click Save.

5) Once the database is downloaded, you need to make sure that you decompress the file if it is compressed using a program such as WinZip, or WinRar to extract its contents. The file inside the compressed file will be a FASTA format text file (a .fas or a .fasta file).

6) Finally put the database file into a directory that PEAKS can access.

7) Click "Browse" to tell PEAKS where the database file is located.

8) If the database that you have selected is an EST database, check the box labeled "EST Database". If not, leave it blank.

9) Since you have already selected a "FASTA Format Database" in Step 2, the Accession number information and the parsing rules for the database headers are shown in the textboxes below in the "Advanced Options- Fasta Title Format" panel.

If you chose "Other" (in step 2) you must enter parsing parameters by completing the textboxes. Alternatively, if the database format is the same as one of the public databases such as NCBI-nr, you can choose to apply that database's parsing rules when PEAKS reads the database. To do this, simply select the database that is similar to yours from the drop-down list to fill the textboxes with the appropriate parsing rules.

A note on parsing rules

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

A note on using a delimiter

Some databases use one entry to represent multiple protein entries. The FASTA headers are concatenated with a delimiter. Since some of these databases use unprintable control codes as delimiters, PEAKS will use the equivalent ASCII decimal code to represent them. For example the NCBI NR database uses CTRL-A as a delimiter so the user should input "1" as its equivalent decimal delimiter as listed here.

10) To be able to do PEAKS Protein ID using a specific taxonomy, you will need to download some files and refer PEAKS where to find them in the "Taxonomy Options" panel.

11) To download the taxonID file, click the "Download" button. A window will appear confirming the ftp or website which has been identified as to the location of the desired database.

If you would like to invoke your default FTP client software and download automatically, click "Yes". If you select "No" the following window will appear telling you that the URL will be copied to your system clipboard. Click "Ok".

Open your Internet Explorer and paste the URL into the address bar. A file download window will open. Click Save. Be sure to save the file in a location that is accessible by PEAKS. Please note that you do not have to decompress the taxonomy files.

12) Now that you have downloaded the taxonomy files, you must tell PEAKS where to access them by clicking the Browse button and selecting the file.

13) To save the database to your “Database List”, you must click the “Add/Update” button before clicking the “Ok” button.

Delete a previously saved database

If you would like to delete a database file, select the database that you wish to delete and click on the “Delete Database” button.

Set a database as default

Select the file and click the “Set as default” button which is located to the right of the “Database List”. This database will now be used by PEAKS when you run PEAKS Protein ID.

Moving/Updating a Database

If you choose to move a database to another directory or delete it entirely, you need to notify PEAKS. You must remove the database from the list and re-load it. Until you do so, the database name will appear in red in the list of databases and any protein identification using that database will fail.

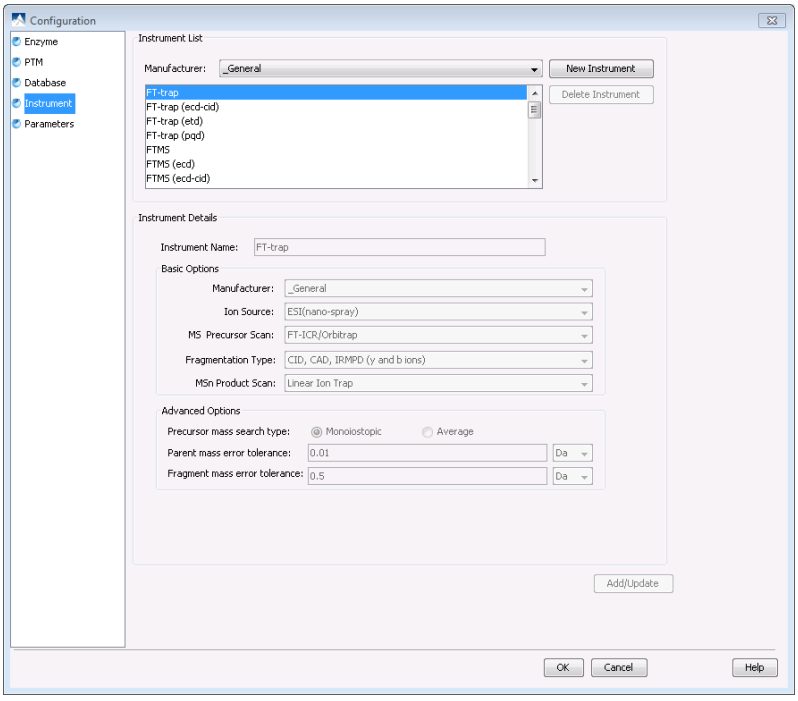
If you choose to update the database (perhaps by downloading the latest database file and overwriting the old database file), PEAKS will show the database information in light gray. A light grey color could also mean that the database does not have header information.

Best practices: configuring databases for use with X!Tandem

At the time of this writing, X!Tandem had trouble searching through large databases, and would crash. It is therefore suggested that X!Tandem only be used with small databases; or if used with a large database, a taxon should be specified. The NCBInr and SwissProt databases are ideal for this purpose.

Best practices: configuring databases for use with OMSSA

At the time of this writing, we could not use OMSSA with databases that were not in NCBI format, or SwissProt format, and have those results available to inChorus.



Also, a bug in OMSSA prevents us from easily using databases with OMSSA when they are stored in a folder that contains a space in its path. This creates problems when PEAKS creates temporary databases on our behalf. To avoid this, best practices suggest that you put all our databases in a folder “c:\peaksdatabases”. The folder “c:\my documents\databases” wouldn’t work because it contains a space between ‘my’ and ‘documents’. Using spaces in the database file name causes the same problem. So after you download and extract our database you should call the database file “ncbinr.fas”, or “ncbi_nr.fas” rather than “ncbi nr.fas”.

16.2.4 INSTRUMENT CONFIGURATION

From the "Configuration" window select "Instrument" from the left hand to change your instrument configuration.

Built- in Instruments

Select the manufacturer of your instrument from the drop-down list. The names of the instruments will then appear in their vendor specific formats. Select your instrument and you will be able to view the information on your instrument in the "Instrument details" panel below. You can also select "General" in the manufacturer list and the instruments will be listed in a general format.

Note: You cannot delete or change the details of a built-in instrument and therefore the "Delete PTM" button and the "Instrument Details" panel will be grayed out.

Create a new instrument

Click on the "New Instrument" button.

In the "Instrument Details" panel, create a name for your instrument.

Next, fill in your details in the "Basic Options" panel. In the manufacturer drop-down list, select a specific vendor or "General".

Ion Source: Use the drop-down list to select what ion source that was used; MALDI/SELDI or ESI (nano-spray). This will help the PEAKS Data Refine tool to decide the charge of the ions.

MS- Precursor Scan: Use the drop-down list to select what type of MS scan was performed. This selection will tell the PEAKS Data Refine tool if the survey scan is of sufficient resolution to determine the charge and the monoisotopic peak from the examination of the survey scan.

Fragmentation type: Use the drop-down list to select the method of fragmentation that was used. This selection will tell PEAKS what type of ion-series to expect for PEAKS auto de novo sequencing and PEAKS Protein ID database search. Select CID/ECD if alternating fragmentation is used to allow the algorithm to determine the type of fragmentation from each scan header.

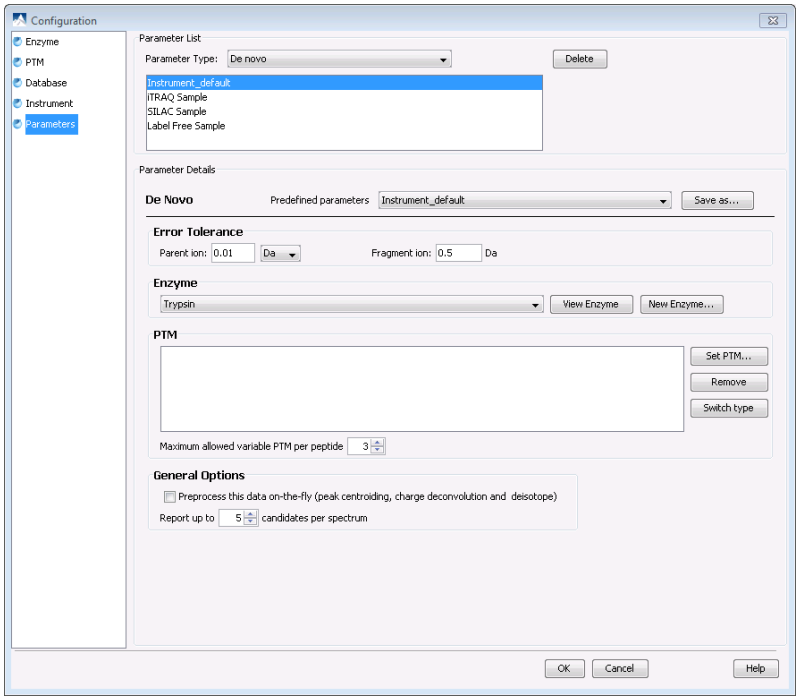
MSn Product Scan: Use the drop-down list to select what type of MSn scan was performed. This selection will help PEAKS decide which internal parameters (for weighing fragments and amount of noise) to use during PEAKS auto de novo sequencing and PEAKS Protein ID database search. Select LIT/FT if alternating hi-res/low-res modes were used. This will allow the algorithm to determine the mass analyzer from the scan header. You can also use the "Advanced Options" to specify additional parameters.

Precursor Mass Search Type: Select "Monoisotopic" or "Average". For ion-trap instruments, it is usually beneficial to allow the PEAKS Protein ID database search to use an average mass.

Parent and Fragment error mass tolerance: User specified values. These will appear on the PEAKS de novo and PEAKS Protein ID options screens when the instrument is selected.

You must click the "Add/Update button for the changes to be saved. Your new instrument will now appear in the "Instrument List" where it can be accessed when creating a new project file. If you wish to delete an instrument that you created, select the appropriate instrument and click the "Delete Instrument" button.

16.2.5 PARAMETER CONFIGURATION



From the "Configuration" window select "Parameters" from the left hand side to change your parameter configurations. Please note that you can only view and delete parameters from within this parameter window. From the "Parameter type" drop-down list at the top of the screen you can select de novo, PEAKS parameters, SPIDER parameters or other parameter categories. The parameters that you have saved within these categories will be displayed below in the list. Select the parameter file that you would like to view.

Creating a new parameter

If you would like to create and save new parameters, you can do this during or before you set up auto de novo sequencing (refer to section 6.1 Setting up Auto de novo Sequencing Parameters), PEAKS Protein ID (refer to section 7.1 Setting up Protein Identification Parameters) or SPIDER (refer to section 9.1 Setting up SPIDER Parameters). These references will provide you with an explanation of all of the parameters.

Deleting a previously saved parameter

If you would like to delete a parameter file, select the file that you wish to delete and click on the "Delete" button.

Viewing a previously saved parameter

Selecting a file will display the details of that file below. For an explanation of the parameters, please see the pages listed in the "Creating a new parameter file" section above.

Chapter

17

17 About Bioinformatics Solutions Inc.

BSI provides advanced software tools for analysis of biological data.

Bioinformatics Solutions Inc. develops advanced algorithms based on innovative ideas and research, providing solutions to fundamental bioinformatics problems. This small, adaptable group is committed to serving the needs of pharmaceutical, biotechnological and academic scientists and to the progression of drug discovery research. The company, founded in 2000 in Waterloo, Canada, comprises a select group of talented, award-winning developers, scientists, computer and biological scientists as well as valuable consultants and beta-testers.

At BSI, groundbreaking research and customer focus go hand in hand on our journey towards excellent software solutions. We value an intellectual space that fosters learning and an understanding of current scientific knowledge. With an understanding of theory, we can focus our talents on providing solutions to difficult, otherwise unsolved problems that have resulted in research bottlenecks. At BSI, we are not satisfied with a solution that goes only partway to solving these problems; our solutions must offer something more than existing software.

The BSI team recognizes that real people will use our software tools. As such, we hold in principle that it is not enough to develop solely on theory; we must develop with customer needs in mind. We believe the only solution is one that incorporates quality and timely results, a satisfying product experience, customer support and two-way communication. So then, we value market research, development flexibility and company-wide collaboration, evolving our offerings to match the market/user's needs.

Efficient and concentrated research, development, customer focus and market analysis have produced: PEAKS software for protein and peptide identification from tandem mass spectrometry data, RAPTOR and PROSPECT Pro software for threading based 3D protein structure prediction, PatternHunter software for all types of homology search sequence comparison and ZOOM software for next generation sequencing.

Chapter

18

18 PEAKS® SOFTWARE LICENSE

This is the same agreement presented on installation. It is provided here for reference only.

If we are evaluating a time limited trial version of PEAKS and we wish to update the software to the full version, we must purchase PEAKS and obtain a full version registration key.

1. License. Subject to the terms and conditions of this Agreement, Bioinformatics Solutions (BSI) grants to you (Licensee) a non-exclusive, perpetual, non-transferable, personal license to install, execute and use one copy of PEAKS (Software) on one single CPU at any one time. Licensee may use the Software for its internal business purposes only.

2. Ownership. The Software is a proprietary product of BSI and is protected by copyright laws and international copyright treaties, as well as other intellectual property laws and treaties. BSI shall at all times own all right, title and interest in and to the Software, including all intellectual property rights therein. You shall not remove any copyright notice or other proprietary or restrictive notice or legend contained or included in the Software and you shall reproduce and copy all such information on all copies made hereunder, including such copies as may be necessary for archival or backup purposes.

3. Restrictions. Licensee may not use, reproduce, transmit, modify, adapt or translate the Software, in whole or in part, to others, except as otherwise permitted by this Agreement. Licensee may not reverse engineer, decompile, disassemble, or create derivative works based on the Software. Licensee may not use the Software in any manner whatsoever with the result that access to the Software may be obtained through the Internet including, without limitation, any web page. Licensee may not rent, lease, license, transfer, assign, sell or otherwise provide access to the Software, in whole or in part, on a temporary or permanent basis, except as otherwise permitted by this Agreement. Licensee may not alter, remove or cover proprietary notices in or on the Licensed Software, or storage media or use the Licensed Software in any unlawful manner whatsoever.

4. Limitation of Warranty. THE LICENSED SOFTWARE IS PROVIDED AS IS WITHOUT ANY WARRANTIES OR CONDITIONS OF ANY KIND, INCLUDING BUT NOT LIMITED TO WARRANTIES OR CONDITIONS OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. LICENSEE ASSUMES THE ENTIRE RISK AS TO THE RESULTS AND PERFORMANCE OF THE LICENSED SOFTWARE.

5. Limitation of Liability. IN NO EVENT WILL LICENSOR OR ITS SUPPLIERS BE LIABLE TO LICENSEE FOR ANY INDIRECT, INCIDENTAL, SPECIAL, OR CONSEQUENTIAL DAMAGES WHATSOEVER, EVEN IF THE LICENSOR OR ITS SUPPLIERS HAVE BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGE OR CLAIM, OR IT IS FORESEEABLE. LICENSOR'S MAXIMUM AGGREGATE LIABILITY TO LICENSEE SHALL NOT EXCEED THE AMOUNT PAID BY LICENSEE

FOR THE SOFTWARE. THE LIMITATIONS OF THIS SECTION SHALL APPLY WHETHER OR NOT THE ALLEGED BREACH OR DEFAULT IS A BREACH OF A FUNDAMENTAL CONDITION OR TERM.

6. Termination. This Agreement is effective until terminated. This Agreement will terminate immediately without notice if you fail to comply with any provision of this Agreement. Upon termination, you must destroy all copies of the Software. Provisions 2,5,6,7 and 10 shall survive any termination of this Agreement.

7. Export Controls. The Software is subject at all times to all applicable export control laws and regulations in force from time to time. You agree to comply strictly with all such laws and regulations and acknowledge that you have the responsibility to obtain all necessary licenses to export, re-export or import as may be required.

8. Assignment. Customer may assign Customer's rights under this Agreement to another party if the other party agrees to accept the terms of this Agreement, and Customer either transfer all copies of the Program and the Documentation, whether in printed or machine-readable form (including the original), to the other party, or Customer destroy any copies not transferred. Before such a transfer, Customer must deliver a hard copy of this Agreement to the recipient.

9. Maintenance and Support. BSI will provide technical support for a period of thirty (30) days from the date the Software is shipped to Licensee. Further maintenance and support is available to subscribers of BSI's Maintenance plan at BSI's then current rates. Technical support is available by phone, fax and email between the hours of 9 am and 5 pm, Eastern Time, excluding statutory holidays.

10. Governing Law. This Agreement shall be governed by and construed in accordance with the laws in force in the Province of Ontario and the laws of Canada applicable therein, without giving effect to conflict of law provisions and without giving effect to United Nations Convention on contracts for the International Sale of Goods.