Scaffold Q+S

Version 5.2

User's Guide

Release Information

The following release information applies to this version of the Scaffold Q+S User's Guide. This document is applicable for Scaffold Q+S, Release 5.2 or greater, and is current until replaced.

Copyright

Limit of Liability

Proteome Software, Inc. has made its best effort in preparing this guide. Proteome Software, Inc. makes no representations or warranties with respect to the accuracy or completeness of the contents of this guide and specifically disclaims any implied warranties of merchantability or fitness for a particular purpose. Information in this document is subject to change without notice and does not represent a commitment on the part of Proteome Software, Inc. or any of its affiliates.The accuracy and completeness of the information contained herein and the opinions stated herein are not guaranteed or warranted to produce any particular results, and the advice and strategies contained herein may not be suitable for every user.

The software described herein is furnished under a license agreement or a non-disclosure agreement. The software may be copied or used only in accordance with the terms of the agreement. It is against the law to copy the software on any medium except as specifically allowed in the license or the non-disclosure agreement.

Trademarks

The name *Proteome Software*, the Proteome Software logo, *Scaffold, Scaffold Q+*, *Scaffold Q+S*, *Scaffold LFQ, Scaffold PTM, Scaffold DIA, Scaffold perSPECtives* and *Scaffold Elements* logos are trademarks or registered trademarks of Proteome Software, Inc. All other products and company names mentioned herein may be trademarks or registered trademarks of their respective owners.

Customer Support

Customer support is available to organizations that purchase *Scaffold, Scaffold Q+, Scaffold Q+S, Scaffold LFQ,Scaffold PTM, Scaffold DIA or Scaffold Elements* and that have an annual support agreement. Contact Proteome Software at:

Proteome Software, Inc.
1340 SW Bertha Blvd
Suite 10
Portland, OR 97219
1-800-944-6027 (Toll Free)
1-928-244-6024 (Fax)
www.proteomesoftware.com

-

Table of Contents

-

Preface

Welcome to the Scaffold Q+S User's Guide. Its purpose is to answer users' questions and guide them through the procedures necessary for using Scaffold Q+S efficiently and effectively.

Using the manual

A Table of Contents and an Index are provided in this manual for the user's convenience. This Preface also provides a brief discussion of each chapter to further assist users in locating needed information.

Special information about the manual

This User's Guide has a dual-purpose design. It can be distributed electronically and printed on an as-needed basis, or it can be viewed on-line in its fully interactive capacity. If users print the document, for best results it is recommended that they print it on a duplex printer; however, single-sided printing is also possible. When the document is viewed on-line, a standard set of bookmarks appears in a frame on the left side of the document window for navigation through the manual. For better viewing, users can decrease the size of the bookmark frame and use the magnification box to adjust the display according to their viewing preferences.

If users do print the document using a single-sided printer, they might see a single blank page at the end of some chapters. This blank page has been added solely to ensure that the next chapter begins on an odd-numbered page. This blank page in no way indicates that the book is missing information.

Conventions used in the manual

The *User's Guide uses the following conventions:*

- Information that can vary in a command—variable information—is indicated by alphanumeric characters enclosed in angle brackets; for example, <ProteinAnalyte Name>.
- A new term, or term that must be emphasized for clarity of procedures, is *italicized*.
- Page numbering is "on-line friendly." Pages are numbered from 1 to x, *starting with the cover* and ending on the last page of the index.
- This manual is intended for both print and on-line viewing.
- If information appears in blue, it is a hyperlink. Table of Contents and Index entries are also hyperlinks. Click the hyperlink to advance to the referenced information.
- A sample set of Demo data, available for download from *http://www.proteomesoftware.com/products/demo-data* is used as the basis for most screen captures, examples, and data manipulations that are shown in the manual.

Assumptions in the manual

The Scaffold Q+S User's Guide assumes that*:*

- The user is familiar with Windows operating systems, and basic Windows navigational elements, content formatting and layout tools.
- The user has the appropriate licensing to run Scaffold Q+S.

-

Getting Started with Scaffold Q+S

System Requirements

For information about the system requirements for Scaffold Q+S, see:

https://support.proteomesoftware.com/hc/en-us/articles/213578086-Scaffold-Software-System-Requirements

Installing Scaffold Q+S

Scaffold Q+S runs on Windows, MAC or Linux systems. Follow these instructions to install the application on your system:

Request an evaluation by filling in the form found at ***http://www.proteomesoftware.com/products/scaffold/evaluate/.*** You will receive download instructions and a license key to activate the software via email.

1. Download and launch the installation executable.
2. Carefully follow the instructions provided in the installation wizard, accepting the user agreement when prompted and moving through the screens by clicking Next.

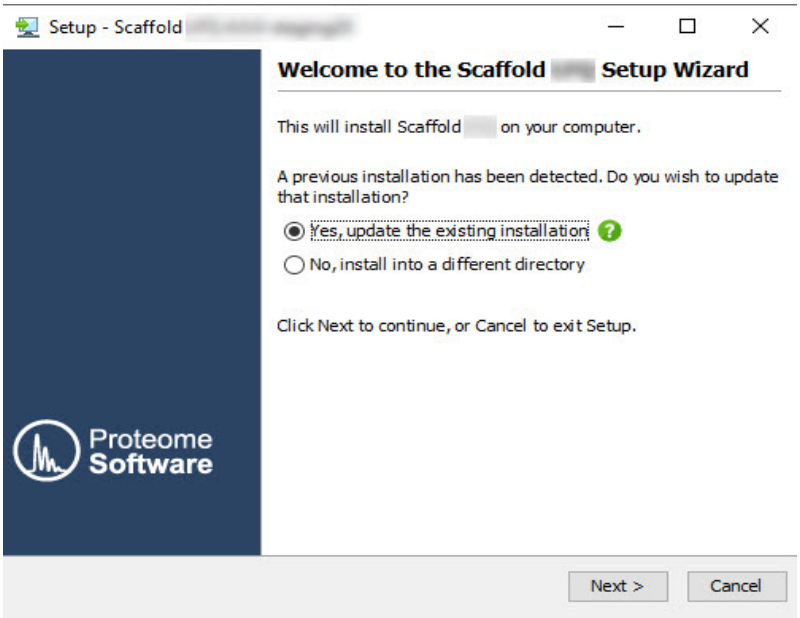


Figure 1-1: Scaffold Q+S installation Setup Wizard

-

3. The installer will then provide you an opportunity to allocate memory to Scaffold Q+S. We recommend that you set the Maximum Memory to approximately 80% of the amount of physical RAM on your system. Click “Next”.
4. You may select a Start Menu Folder for the application and choose whether or not to create shortcuts for all users of the system. The next screen allows you to set a file association between SF3 files and Scaffold Q+S, and the following screen allows creation of desktop icons. Clicking “Next” begins the installation.
5. Finally, Scaffold Q+S allows you to select the option to have the program open at the closing of the wizard. Click “Finish”..

For better performance you should allocate as much RAM as possible to Scaffold Q+S. The memory setting can be adjusted after installation by selecting the menu option ***Edit > Preferences - Memory tab****. You must close Scaffold Q+S and restart the program in order for the new memory setting to take effect.*

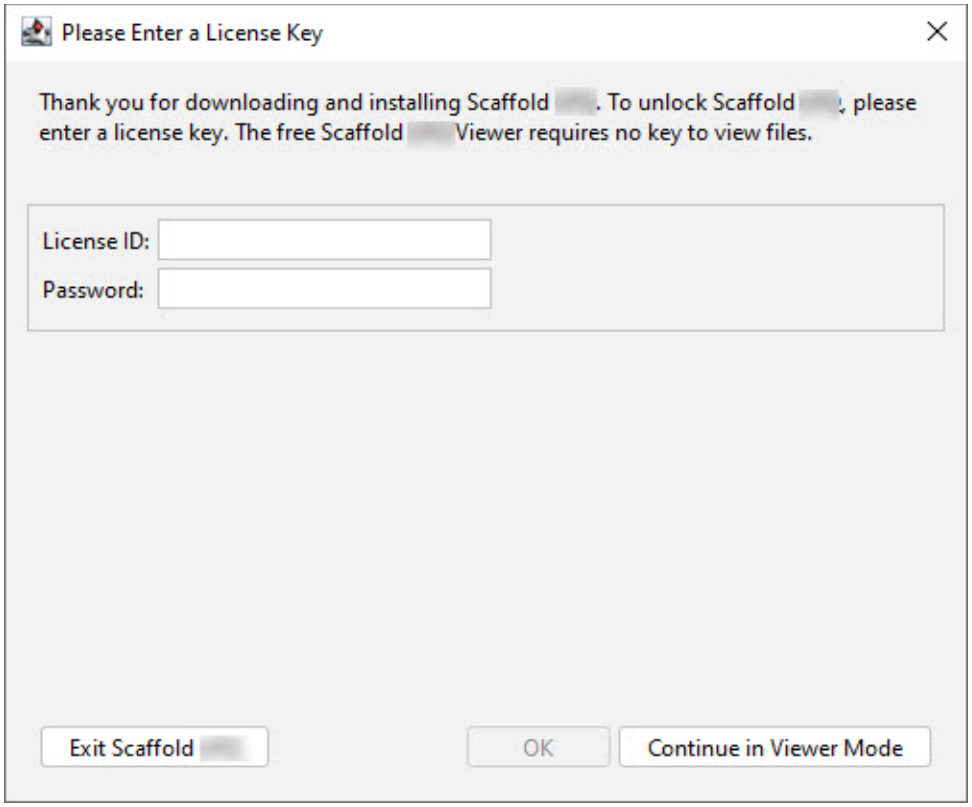
After Scaffold Q+S has been installed on a computer, a shortcut icon for the application is placed on the desktop. An option is also available from the Start menu. Double-clicking the desktop icon launches Scaffold Q+S, as does, for Windows computers, selecting the option from the Start menu (**Start > All Programs > Scaffold Q+S > Scaffold Q+S**)

Licensing

The first time Scaffold Q+S opens after installation, the Enter License Key dialog box opens.

Keys and passwords may be typed, pasted or dragged into the appropriate fields. Both items may be pasted or dragged together.

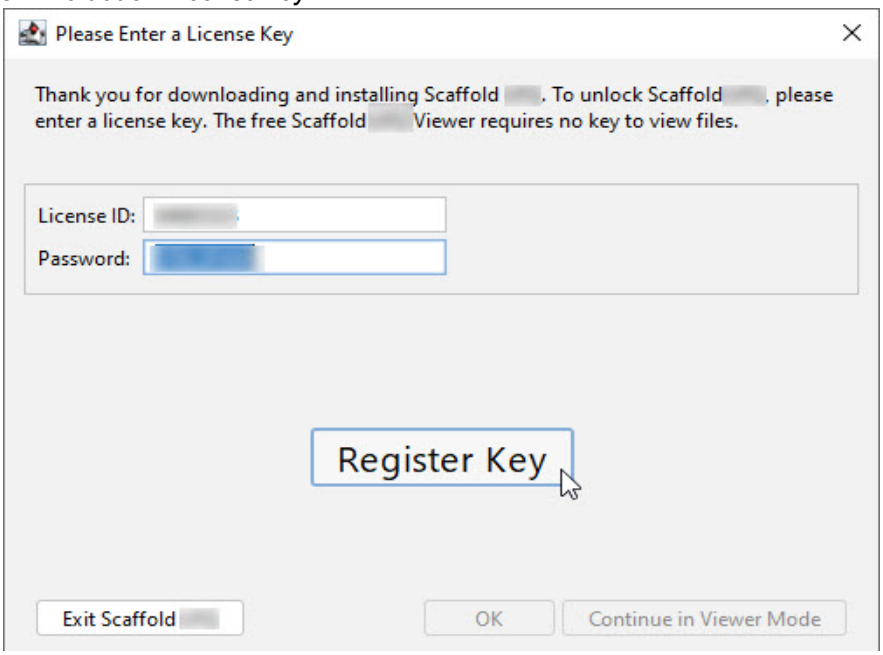
Figure 1-2: Scaffold License Key messages



Two kinds of keys are available to activate the software:

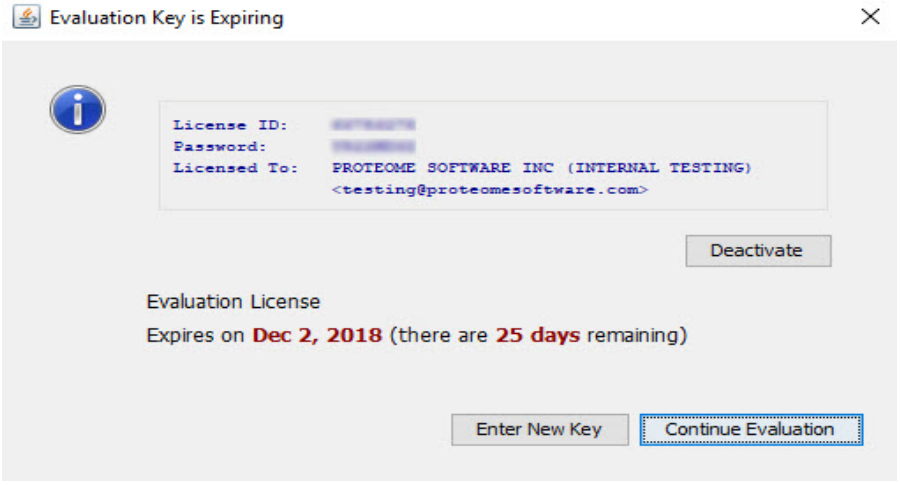
Evaluation key - An Evaluation key is valid for a limited period. A free evaluation key for Scaffold Q+S may be obtained through *www.proteomesoftware.com* An evaluation key may be used on two computers. Once the key and password have been copied and pasted into the license key dialog box, a message will appear below it, displaying confirmation of the key registration. Pressing OK starts the application.

Figure 1-3: Evaluation License key



Every time Scaffold Q+S is launched in evaluation mode, a message appears showing the remaining time available for evaluation and offering the option to enter a new key.

Figure 1-4: Message appearing when launching an evaluation copy of Scaffold Q+S

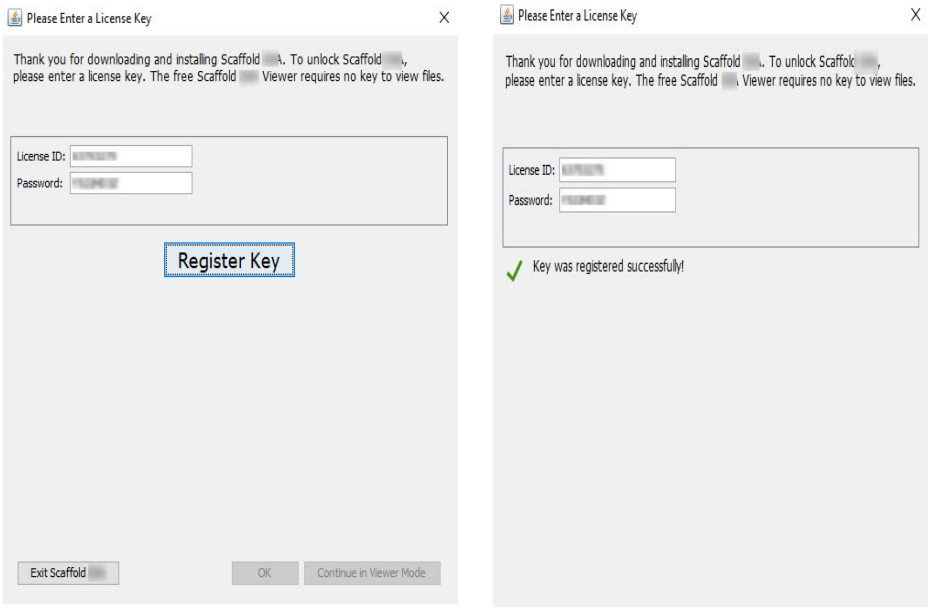


Time-Based License key—a Time-Based License key allows the user to access all features of the software permanently. It only allows upgrades within a certain time limit, however. The time tracks the length of the support contract. Once expired, Scaffold Q+S will continue to work beyond the expiration date, but no upgrades are allowed unless the support contract is renewed.

Contact*sales@proteomesoftware.com* to purchase the appropriate key.

A Time-Based License key is valid only for a single computer. If it is necessary to move the Scaffold Q+S installation to a different computer, see "Moving Scaffold Q+S to a different computer" for instructions to transfer the key at no charge.

Figure 1-5: Time-Based License key



When the Time-Based License key and password are entered, pressing **Register Key** verifies their validity and a message appears describing the status of the key.

Once the key is successfully registered, pressing OK closes the dialog box and a Scaffold Q+S Welcome message opens.

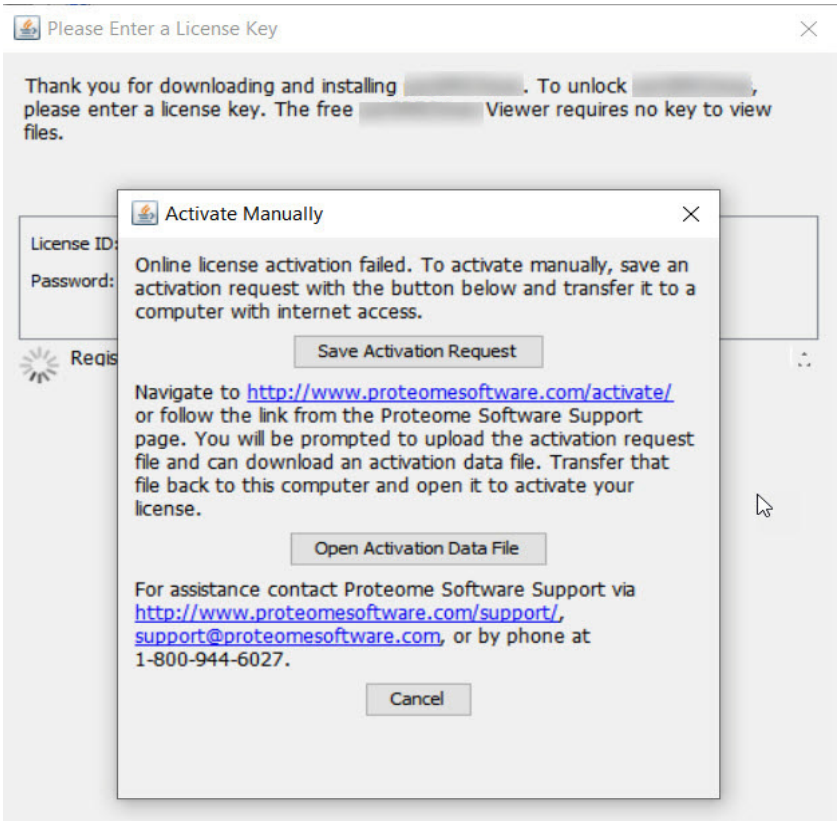
If the user is using an evaluation copy of Scaffold Q+S, then an Evaluation message opens, indicating the number of days left in the evaluation period. The user must click OK to close this message and then the Scaffold Q+S Welcome message opens.

From this window, the user may create a new experiment, open an existing experiment (*.SF3 file), or work with the demonstration data that is provided in the Scaffold Q+S installation.

Registering a Time-Based License key with no INTERNET connection

When a Time-Based License key is entered and the Register Key button is pressed, but no INTERNET connection is available, a dialog appears, providing instructions for manual activation.

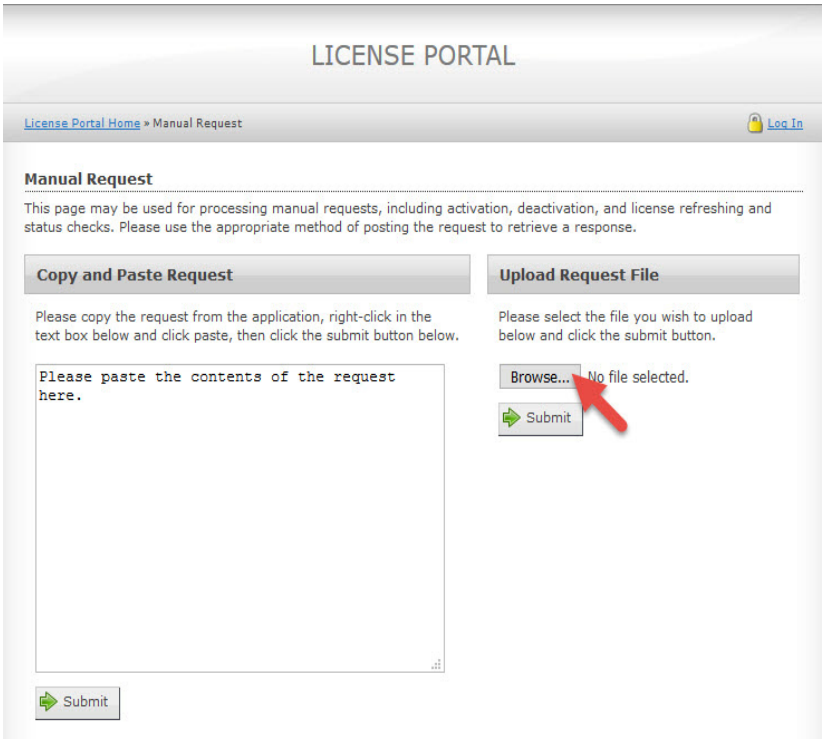
Figure 1-6: Manual or offline activation dialog



To activate Scaffold Q+S without an internet connection:

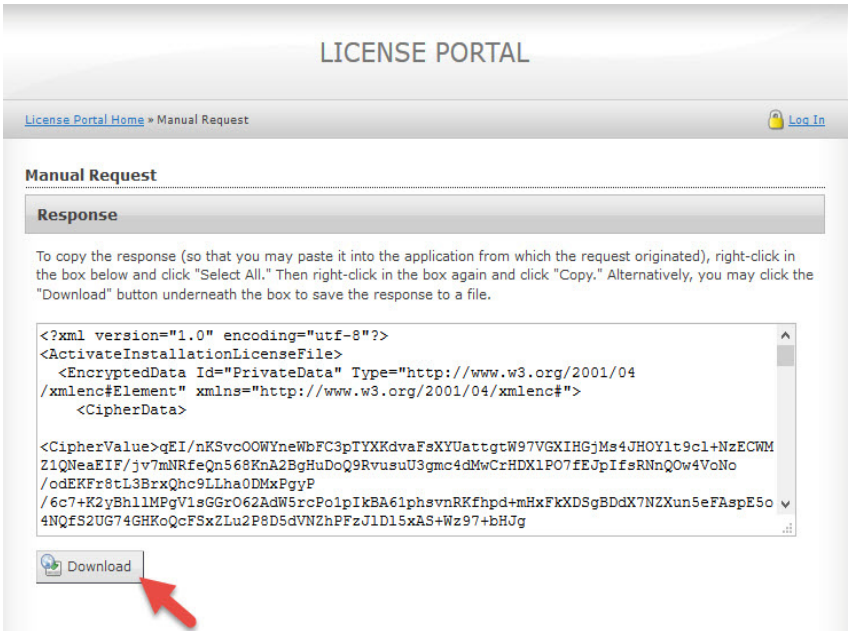
1. First, use the Save Activation Request button to create an activation request file.
2. Transfer this file to a computer with internet access (e.g. using a USB drive).
3. On the connected computer, navigate to */http://www.proteomesoftware.com/activate/*This link is also accessible from the Proteome Software Support page *(http://www.proteomesoftware.com/support/)* to make it easier to access from the internet-connected computer.
4. The License Portal will open. The Portal provides two different options for activating your software. Use the Browse button in the Upload Request File section on the right, and select the activation request file that was transferred from the offline computer (See Figure 1-7 below).

Figure 1-7: The Proteome Software License Portal



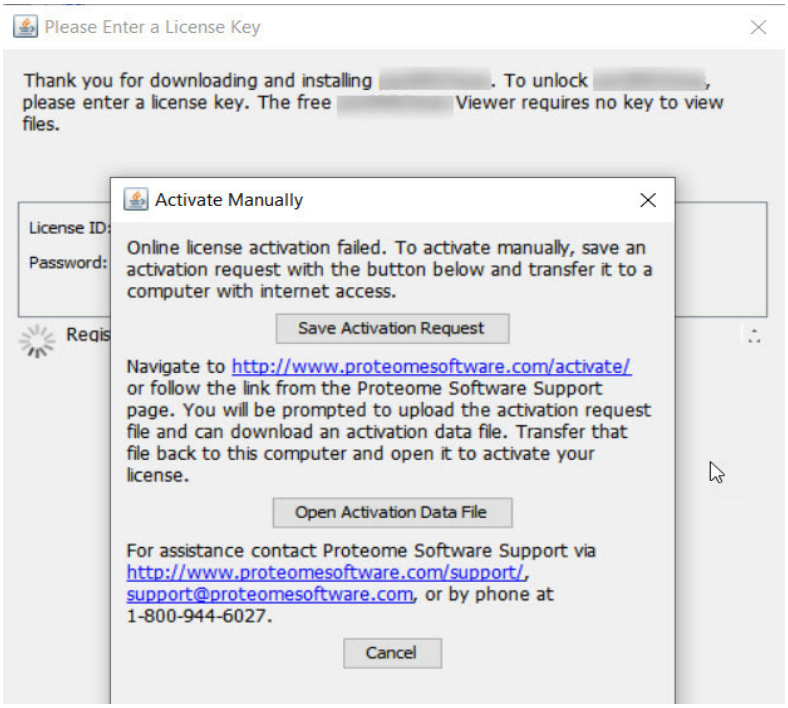
5. Click the Submit button just below the Browse button to upload the activation request file. The license portal will respond with a long text sequence (See Figure 1-8 below).

Figure 1-8: .License Portal Response to Activation Request



6. Click the Download button to save the response to a file named response.xml, which will be downloaded to the default download location.
7. Transfer the response.xml file to the computer on which Scaffold Q+S has been installed.
8. Return to Scaffold Q+S on the disconnected computer. Select Open Activation Data File.

Figure 1-9: Select Activation File returned by the License Portal



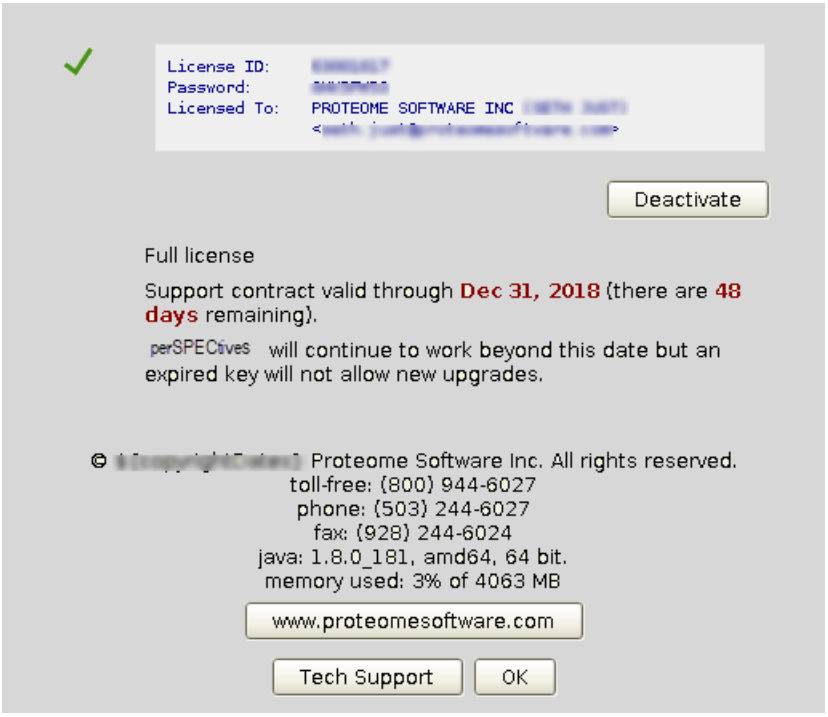
9. Browse to locate the response.xml file and click Open.
10. Scaffold Q+S should report that the key was registered successfully. If not, please contact Proteome Software Support for assistance.

Time based license key renewal

Time based license keys have time limits connected to the term of the user's support contract. When the support contract expires, Scaffold Q+S continues to work but upgrades are not allowed until the contract is renewed. The status of the Scaffold Q+S license key may be checked by selecting **Help > About Scaffold Q+S** from the main menu.

If the contract has expired and the user wishes to upgrade Scaffold Q+S, clicking the **Renew** button in the dialog opens the **Key reset Request** page on the Proteome Software website. The user should complete the request. A sales representative will promptly contact him/her providing further information.

Figure 1-10: About Scaffold Q+S dialog



Moving Scaffold Q+S to a different computer

Each permanent Scaffold Q+S key allows activation of the program on a single computer. If it becomes necessary to reinstall the program either on a different computer or on the same computer following an operating system upgrade or hardware replacement, the user may deactivate the key and then reactivate it on the new system. This may be done once per support contract period. If additional reinstallations are required within the same period, please contact Proteome Software Support.

To deactivate a key:

1. Be sure you have a record of your key and password. These were sent via email at the time of purchase, or may be copied from the Help>>About Scaffold Q+S dialog.
2. Select Help>>Update License Key and click the Deactivate button.

To reinstall Scaffold Q+S:

1. Download the program from the Proteome Software website to the new system and run the installation program.
2. Paste in the key and password and register as described in Installing Scaffold Q+S.

Scaffold Viewer

A free Scaffold Viewer may be downloaded from *www.proteomesoftware.com.* The Scaffold Viewer can open and display any *.SF3 file created by Scaffold Q+S, as well as regular Scaffold files, and allows users to distribute Scaffold Q+S results to colleagues, collaborators or reviewers.

The Viewer may be installed on any number of computers, and multiple instances of the Viewer may be run on a single computer simultaneously. It performs most of the functions of the full Scaffold Q+S program, but it cannot load search results files and analyze data.

Only a single fully-licensed instance of Scaffold Q+S may be run on a computer at one time. Additional instances will function as Viewers.

Licensing

-

-

Scaffold Suite Tiered Licensing

The Scaffold suite of applications consists of the core Scaffold product and Scaffold Q+S. The core Scaffold product is the basis for all installations. The licensing key that Proteome Software provides determines whether the User only has access to the features and functions of the core Scaffold product, or the features and functions of Scaffold Q+S.

Users who purchased a license for Scaffold Q+S, then also have access to all the features and functions of Scaffold.

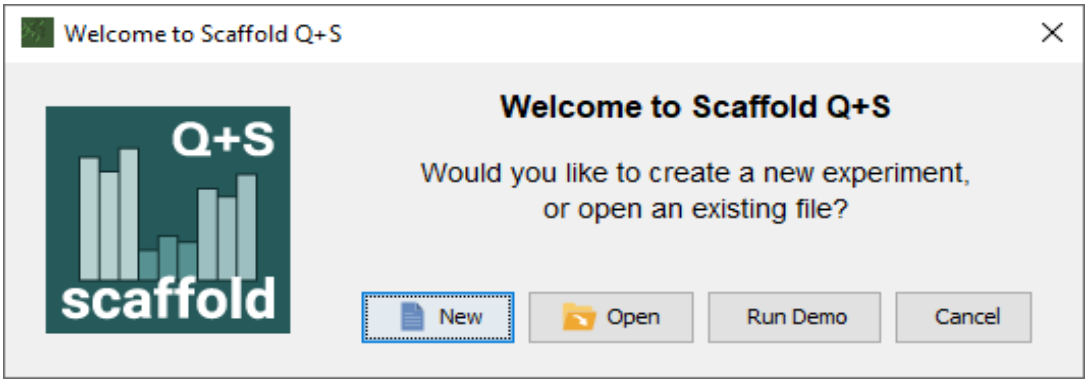
Application	Description
Scaffold	Visualize and validate MS/MS proteomics experiments.
Scaffold Q+S	Calculate and display relative protein expression levels in a sample determined by tandem mass spectrometry of iTRAQ- or TMT-labeled proteins, precursor intensity quantitation and stable isotopically-labeled (for example, SILAC) proteins.

After Scaffold Q+S has been installed on a computer, a shortcut icon for Scaffold is placed on the desktop. An option to start Scaffold is also available from the Start menu. The User can double-click the desktop icon to launch Scaffold, or select the option from the Start menu **Start > All Programs > Scaffold> Scaffold**.

Figure 1-1: Scaffold Q+S desktop icon

Note: Although the icon and startup options are for Scaffold, if the product is registered with a Scaffold Q+S license, the user will see the Scaffold Q+S Welcome screen when the program launches and will have access to all of the features of Scaffold Q+S.

Figure 1-2: Scaffold Q+S Welcome Screen



-

Loading Quantitative Data into Scaffold Q+S

The user begins by loading data into Scaffold using the normal **Scaffold Load Data View** or **New Experiment Wizard** in preparation for analysis in Scaffold Q+ or Scaffold Q+S. For detailed instructions, please consult the Scaffold Users Guide.

The following document, published on the Proteome Software website, provides detailed information on search engine data files compatible with Scaffold Q+ or Scaffold Q+S: *https://www.proteomesoftware.com/scaffold/file-compatibility* .

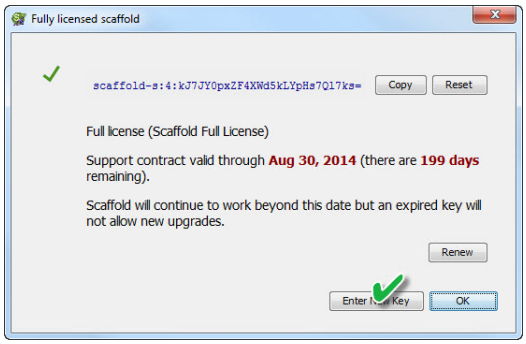
Upgrading Scaffold to Scaffold Q+ or Scaffold Q+S

To upgrade a core installation of Scaffold to Scaffold Q+ or Scaffold Q+S, contact our sales department at *sales@proteomesoftware.com.* When the purchase has been finalized the sales department will send an upgraded license key that will unlock the Scaffold Q+ or Scaffold Q+S features.

To input the new upgraded key:

1. If possible, be sure that the computer is connected to the internet.
2. Open the current copy of Scaffold installed on the computer by double clicking on the Scaffold icon found on the desktop or selecting the Scaffold application from the Start menu.
3. If a warning appears suggesting an upgrade to Scaffold perform the upgrade and reopen Scaffold after the upgrade.
4. When the “Welcome to Scaffold” dialog appears click cancel.
5. Go to the **Help** menu and select the option **Upgrade License Key...**
6. When the **Overwrite** dialog appears click Yes.
7. The **Fully licensed Scaffold** dialog opens showing the information related to the current license.

Figure 1-3: Fully Licensed Scaffold dialog



8. Click Enter New Key. The **Please Enter License Key** dialog opens, see Figure 1-3.

-

9. Copy and paste the license key. After verification of the key the Register Key button appears, click it.
10. If the key is valid the message “Key was registered successfully!” appears, click OK Scaffold is ready to go.
11. If the key is not valid for whatever reason contact *sales@proteomesoftware.com.*

Time based license key renewal

Time based license keys have time limits connected to the duration of the user’s support contract. When the support contract expires, Scaffold Q+S will still function but upgrades are not allowed until the support contract is renewed. The status of the Scaffold Q+S license key can be checked in the **About Scaffold...** dialog, opened by selecting **Help > About Scaffold** from the main menu.

To upgrade Scaffold Q+S when the support contract has expired, click the Renew button in the dialog to open the **Key reset Request** page on the Proteome Software website. Fill in the request and a sales representative will promptly reply providing further information.

How Scaffold and Scaffold Q+S structure data

Scaffold and Scaffold Q+S store all of the data related to an experiment in one single file. Each experiment file (*.sf3) can hold a large number of spectra and all associated data. The user can create, name, and save as many experiment files as disk space permits, but only one at a time can be opened with full Scaffold capabilities. Multiple experiments can be opened in the Viewer mode.

-

Chapter 2
Supported Quantitative Methods

Scaffold Q+ and Scaffold Q+S are software packages that perform quantitative analysis of labeled and label-free proteomic data. These types of quantitative techniques are typically used to monitor patterns of protein abundance in biological samples under various conditions and states.

The quantitative proteomic techniques supported by Scaffold Q+S are:

- Isobaric tags (multiplex): iTRAQ and TMT - either MS2 or MS3
- Stable Isotope labeling techniques (multiplex): SILAC and Stable Isotope Dimethylation
- Label-Free quantitation: Precursor intensity (Area Under the Curve)

To perform a quantitative analysis in Scaffold Q+S, the user must first load the quantitative experiment into Scaffold, see “Quantitative methods details” on page 18 and “Loading Quantitative Datasets in Scaffold” on page 23, then launch Scaffold Q+S to tease apart and analyze the quantitative information for each sample present in a multiplex dataset and to organize the data according to the original experimental design.

The list of search engines supported in Scaffold Q+ and Scaffold Q+S for quantitative analysis is provided in the File compatibility matrix .

-

Quantitative methods details

The first task in analyzing labeled or label-free quantitative data in Scaffold Q+S is to load search engine results in Scaffold. The user must be aware of the following technical information related to the various quantitative techniques supported by the application:

- Chapter 2, “Precursor Intensity Quantitative Data,” on page 18.
- “SILAC Data” on page 21

Precursor Intensity Quantitative Data

For label-free quantitation purposes Scaffold Q+ and Scaffold Q+S can load Precursor Intensity analysis results created in:

- MaxQuant
- Mascot Distiller
- Proteome Discoverer
- FragPipe

For any of these data files the user should select the **Precursor Intensity (Standard)** option in the **Select Quantitative Technique** page of the Scaffold Loading Wizard.

- *A minimum of two BioSamples are needed to perform the analysis: a reference sample and a treated samples, for example.*

MaxQuant

To load MaxQuant data into Scaffold, the user must load a directory that contains all the data and files that Scaffold needs for processing the MaxQuant results. To do so, the user must:

- Create a fresh, new directory for *each new* MaxQuant RUN (each search/analysis combination), using a naming convention of your choice. Proteome Software recommends a name that is descriptive for archiving purposes, such as:

  <date>_<descriptor><descriptor-params>_<further-descriptor-params>

  Example: **2013_0415_maxquant_3-plex-bob-jones_1**

- Copy the raw files that are to be analyzed into the newly created directory. This prevents existing MaxQuant analysis files from being overwritten.
- When prompted to “Queue Files for Loading,” load this directory directly into Scaffold.

*For information about loading MaxQuant data into Scaffold, see the **Loading Search Engine results in Scaffold** document at:*
www.proteomesoftware.com/pdf/loading_search_engine_results_into_scaffold.pdf.

Mascot Distiller

In Mascot Distiller the user needs to follow the directions given in the Mascot Wizard to set

-

up the search and select the Average[MD] as the quantitation method for the experiment. After using quantitate to analyze the experimental data in Mascot Distiller, the user has to select **Analysis > Quantitation report > Save as XML** and save the XML file and its corresponding .rov and .dat files in the same directory without renaming the .rov file. (Scaffold Q+S and Scaffold Q+ require both the .xml file and its corresponding .rov file.) When the Scaffold Wizard prompts to queue files for loading, the user has to select the exported .xml file. Scaffold then automatically reads the *.rov file and accesses the *.dat file directly from the Mascot server. If Scaffold is unable to obtain the *.dat file directly from the Mascot server, it looks for it in the directory that contains the *.xml file.

*For information about loading Precursor intensity data from Mascot Distiller into Scaffold, see the **Loading Search Engine Results in Scaffold** document at www.proteomesoftware.com/pdf/loading_search_engine_results_into_scaffold.pdf.*

Proteome Discoverer

When prompted to “Queue Files for Loading” the user needs to load search results files created using the precursor intensity template.

- **Version 1.3 and higher -** point Scaffold to the resulting MSF files**.**
- **Version 2.0 and higher** - When loading the resulting MSF files created in PD 2.x, the user needs to make sure that all files belonging to a study are located in the same folder containing the MSF file to be loaded into Scaffold.
  - Select the **Save All** option in PD 2.x before loading a new MSF file in Scaffold.
  - When running the Daemon, export the parameter file to the location where the MSF files are being saved.
  - See: www.proteomesoftware.com/documentation/configuring-proteome-discoverer/ .

FragPipe

Scaffold Q+S supports precursor intensity quantitation from FragPipe if MS1 Quantification was selected when the FragPipe search was configured. Either the default IonQuant or the optional FreqQuant algorithm may be used. To create the necessary input files for Scaffold Q+S, select the MSFragger option to output results as a TSV file (or as both TSV and pepXML files) and also check the option to write calibrated MGF files. Load the TSV file into Scaffold and launch Scaffold Q+S to analyze and display precursor intensity quantitative data.

Figure 2-1: Select an option to perform MS1 Quantification in FragPipe

-

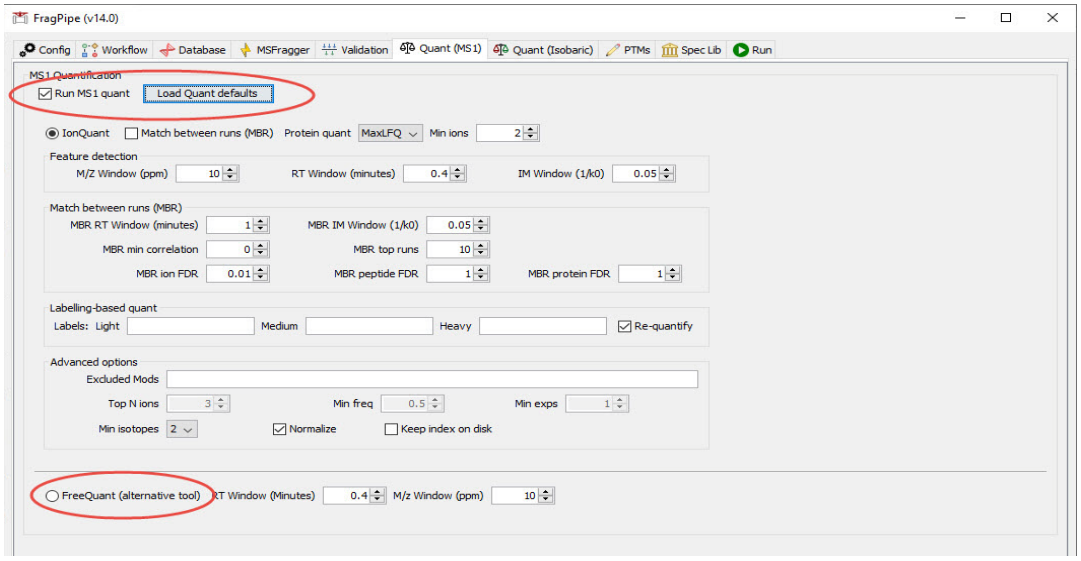
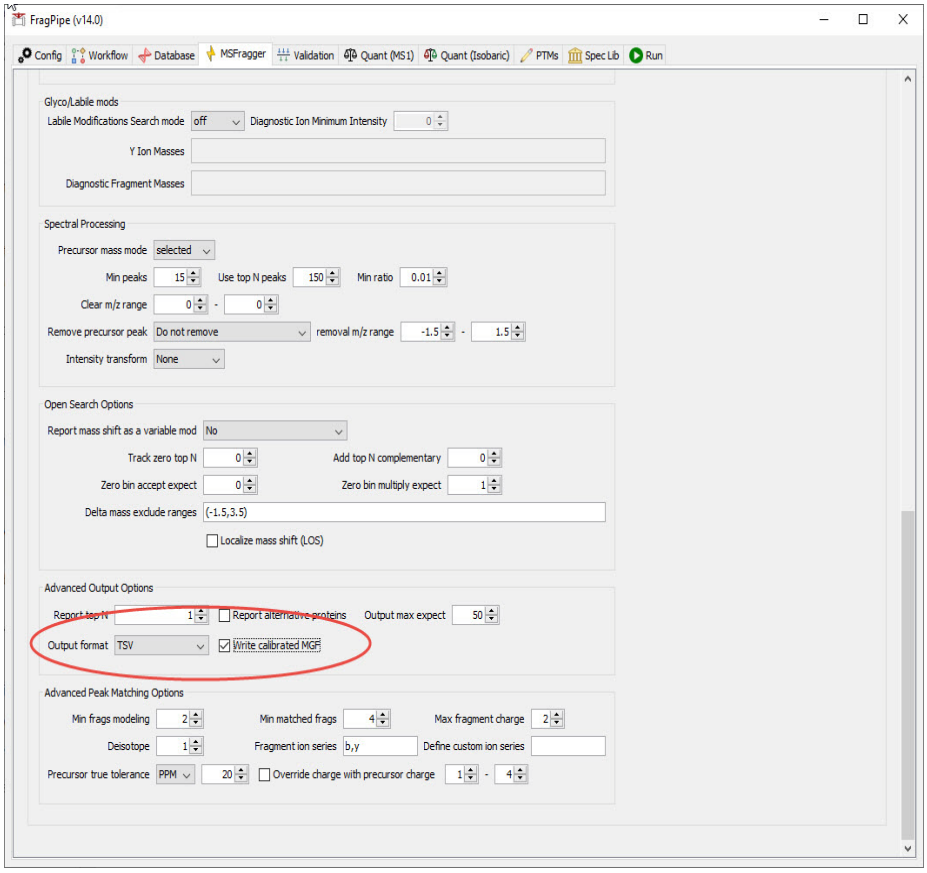


Figure 2-2: Select options to output a TSV file and to write calibrated MGFs



-

SILAC Data

Scaffold Q+S can load SILAC analysis results created in:

- MaxQuant
- Mascot Distiller
- Proteome Discoverer

For any of these data files the user needs to select the **SILAC (Multiplex)** option in the **Select Quantitative Technique** page of the Scaffold Loading Wizard.

MaxQuant

To load MaxQuant data into Scaffold, the user must load a directory that contains all the data and files that Scaffold needs for processing the MaxQuant results. To do so, the user must:

- Create a fresh, new directory for *each new* MaxQuant RUN (each search/analysis combination), using a naming convention of your choice. Proteome Software recommends a name that is descriptive for archiving such as:

  <date>_<descriptor><descriptor-params>_<further-descriptor-params>

  Example: **2013_0415_maxquant_3-plex-bob-jones_1**

- Copy the raw files that are to be analyzed into the newly created directory. This prevents existing MaxQuant analysis files from being overwritten.
- When prompted to "Queue Files for Loading," load this directory directly into Scaffold.

*For information about loading MaxQuant data into Scaffold, see the **Loading Search Engine results in Scaffold** document at www.proteomesoftware.com/pdf/loading_search_engine_results_into_scaffold.pdf*

Mascot Distiller

After using quantitate to analyze the experimental data in Mascot Distiller, the user should select **Analysis > Quantitation report > Save as XML** and save the XML file and its corresponding .rov and .dat files in the same directory without renaming the .rov file. (Scaffold Q+S and Scaffold Q+ require both the .xml file and its corresponding *.rov file.) When the Scaffold Wizard prompts to queue files for loading, the user has to select the exported *.xml file. Scaffold then automatically reads the *.rov file and accesses the .dat file directly from the Mascot server. If Scaffold is unable to obtain the *.dat file directly from the Mascot server, it looks for it in the directory that contains the *.xml file.

*For information about loading Stable Isotope Labeled search data from Mascot Distiller into Scaffold, see the **Loading Search Engine Results in Scaffold** document at www.proteomesoftware.com/pdf/loading_search_engine_results_into_scaffold.pdf.*

-

Proteome Discoverer

When prompted to “Queue Files for Loading,” load search results files created using the Stable Isotope search templates, like SILAC templates for example and select the SILAC modifications.

Version 1.3 and higher

The user needs to point Scaffold to the resulting MSF files.

Version 2.0 and higher

When loading the resulting MSF files created in PD 2.x, the user needs to make sure that all files belonging to a study are located in the same folder containing the MSF file to be loaded into Scaffold.

- Select the **Save All** option in PD 2.0 before loading a new MSF file in Scaffold.
- When running the Daemon export the parameter file to the location where the MSF files are being saved.
- Further information is provided in *www.proteomesoftware.com/documentation/configuring-proteome-discoverer/*

-

Loading Quantitative Datasets in Scaffold

The following procedure details the steps in this task, which are:

- Specifying the sample quantitative type. See "Specifying sample quantitation" below.
- Specifying the sample setup. See "If you selected iTRAQ or TMT multi-plex type of quantitation, specify the isobaric Purity Correction." on page 24.
- Queuing the sample files for loading. See "Queuing sample files for loading" on page 27.
- Analyzing the loaded data. See "Specifying analysis options and analyzing the data" on page 31.
- Specifying the fasta database that is associated with the sample files. See "Specifying the fasta database" on page 33.

Because the instructions given below address the use of Scaffold and this guide is dedicated to the use of Scaffold Q+ or Scaffold Q+S, the following procedure is presented at a high-level only. For detailed information about loading datasets in Scaffold, refer to the Scaffold User's Guide or the Scaffold Online Help.

The following procedure is written either from the perspective of loading the sample Mascot iTRAQ Demo Data file (iTRAQ_test.dat) and iTRAQ FASTA file (iTRAQ_test.fasta) or from the perspective of loading the sample data from Mascot Distiller (Distiller_Tutorial_SILAC.zip) and the SILAC FASTA file (SILAC-demo.fasta). Both sets of sample data are available for download at: http://www.proteomesoftware.com/products/demo-data#qplus.

To carry out this procedure using this sample data, the User must first extract the contents of the zip file

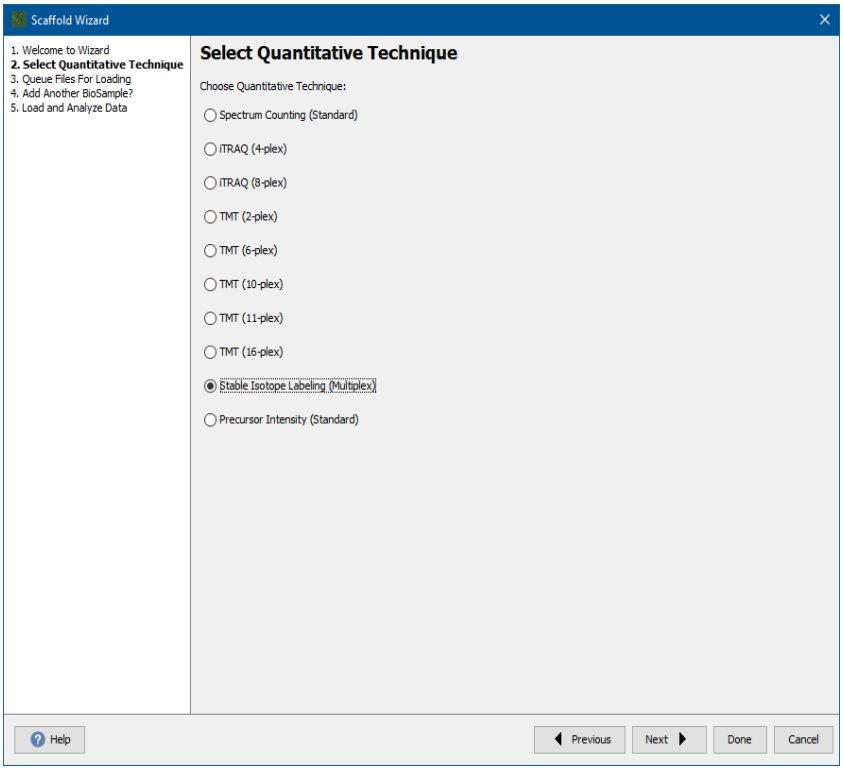
Specifying sample quantitation

1. Open Scaffold.
2. In the Welcome to Scaffold Q+S window, click New.

   The Scaffold Wizard, Welcome to Wizard page opens, click Next to go to the Select Quantitative Technique page.
3. Specify which quantitatiave method Scaffold Q+S is to use.

Figure 2-3: Scaffold Wizard, New Quantitative Technique page inScaffold Q+S

-



If you are carrying out this procedure using the sample iTRAQ data provided by Proteome Software, then select iTRAQ (4-plex).

If you are carrying out this procedure using the sample SILAC data provided by Proteome Software, then select Stable Isotope Labeling (Multiplex).

4. Click Next.

   If you have selected the iTRAQ or TMT multiplex options The Scaffold Wizard, the New iTRAQ or TMT multi-plex page opens to allow specification of a purity correction.

5. If you selected iTRAQ or TMT multi-plex type of quantitation, specify the isobaric Purity Correction.

   - For iTRAQ 4-Plex and 8-Plex data, the default purity corrections that are specified by AB SCIEX are provided. You can select this default correction scheme, you can select from a list of previously created correction schemes, or you can click Add Correction to create a new scheme or edit an existing one for use.

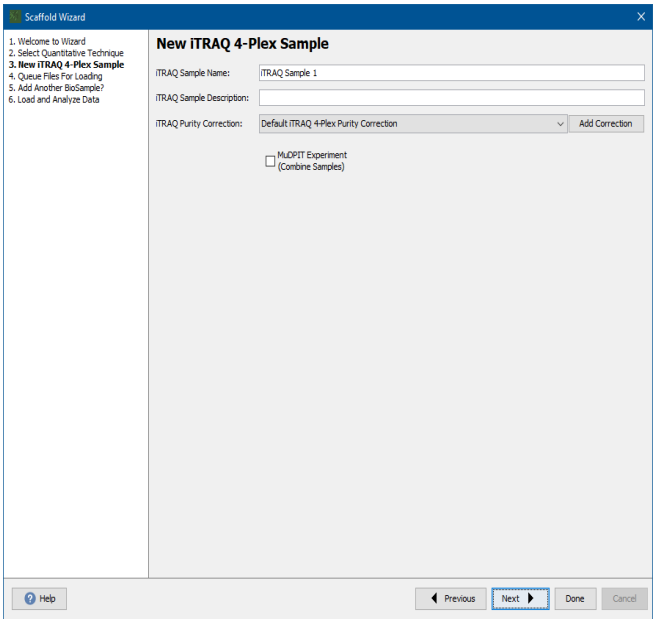
-

If you do not create a new correction scheme, or select an existing one, then the default scheme is used. If you are carrying out this procedure using the sample iTRAQ data provided by Proteome Software, then use the Default iTRAQ 4-Plex purity correction.

- For TMT 2-Plex, 6-Plex, 10-Plex, or 11-Plex, the default purity corrections are empty and you must enter the correct values according to the information that was provided in the Certificate of Analysis that came with the reagents. You can select from a list of previously created correction schemes, or you can click Add Correction to create a new scheme or edit an existing one for use.
- For TMTpro 16-Plex data, please see the detailed instructions provided in the Scaffold4 User's Guide for entering the purity corrections from the Certificate of Analysis. Once values have been entered for a specific lot, that correction scheme is saved and is available for reuse from the dropdown list. When a new lot of reagents is obtained, a new correction scheme must be entered by clicking Add Correction.

6. Click Next.

Figure 2-4: Scaffold Wizard, New BioSample setup page (iTRAQ-labeled samples)

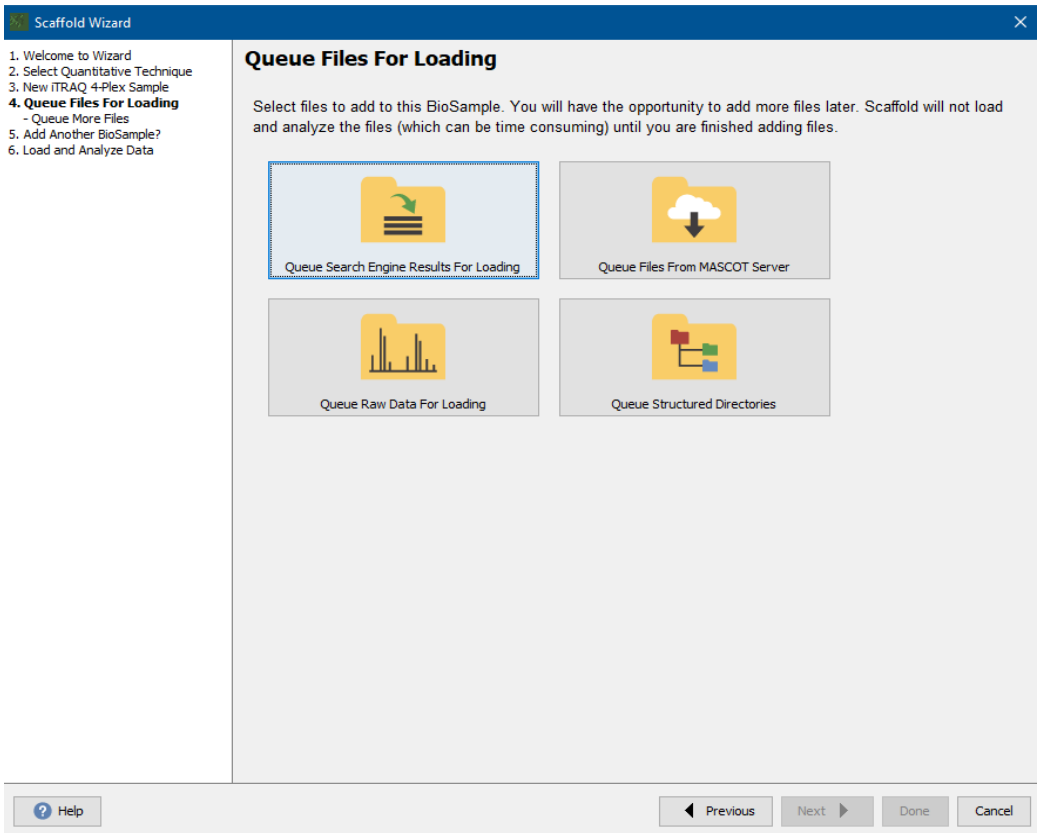


If you are carrying out this procedure using the sample iTRAQ data provided by Proteome Software, then the Sample Setup page is specific for an iTRAQ 4-Plex sample.

If you have selected the SILAC or Precursor Intensity option the Scaffold Wizard, the Queue Files for Loading page opens.

-

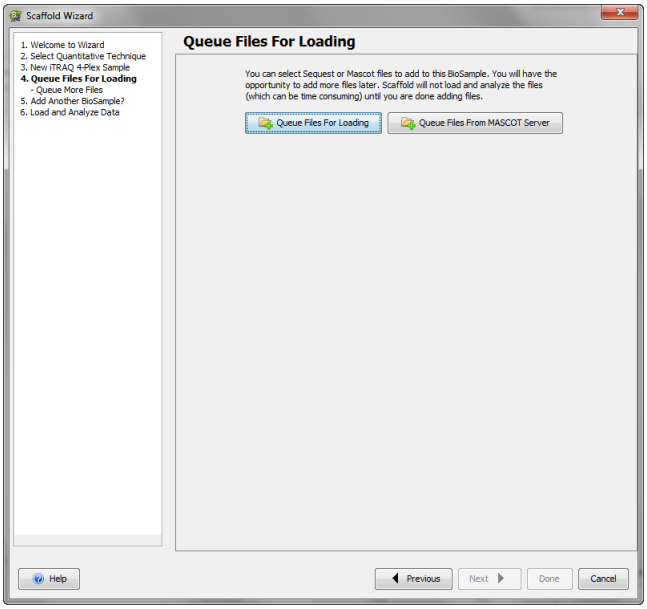
Figure 2-5: Scaffold Wizard,Queue Files for Loading



7. Scaffold Q+S currently only supports experiments created through the Queue Search Engine Results or Queue Structured Directories options.

8. Continue Loading as directed in the Scaffold Users Guide.

Figure 2-6: Scaffold Wizard, Queue Files for Loading page



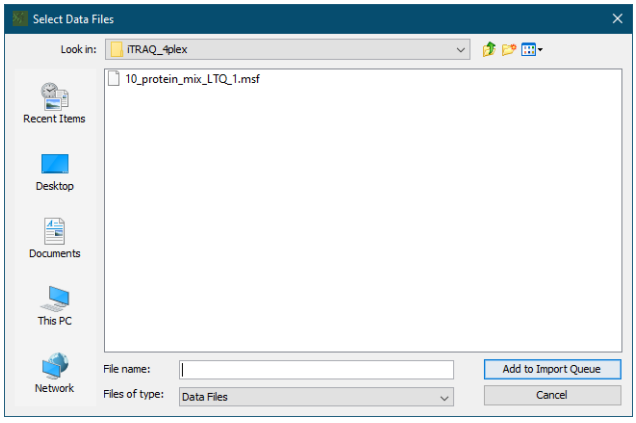
9. Continue to "Queuing sample files for loading" on page 27.

Queuing sample files for loading

1. Click Queue Files for Loading.

   The Select Data Files dialog box opens.

Figure 2-7: Select Data Files dialog box



2. Navigate to the directory in which you saved your sample data set and fasta database, select the sample data set, and then click Add to Import Queue.

-

If you are carrying out this procedure using the sample iTRAQ data provided by Proteome Software, then select the iTRAQ_test.dat file.

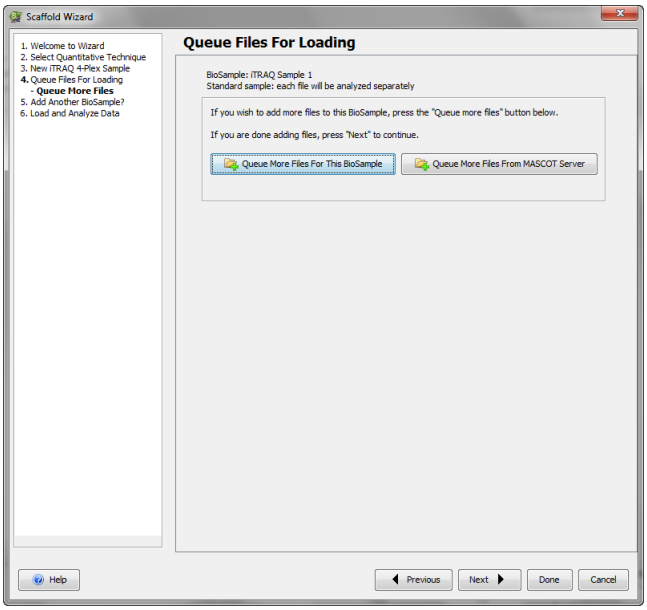
If you are carrying out this procedure using the sample SILAC data provided by Proteome Software, then select the Distiller-tutorial_SILAC.xml file. If you are carrying out this procedure using your own data, and you are uncertain as to which files to add, make sure to review *"SILAC Data" on page 21.*

The Select Data file dialog box closes, and you return to the Scaffold Wizard, Queue Files for Loading, Queue More Files.

3. Queue More Files

   The page prompts you to load additional data files for the current BioSample.

Figure 2-8: Scaffold Wizard, Queue More Files for Loading page



4. Continue to "Queuing more files for loading" on page 28.

Queuing more files for loading

1. If you have more data files to load for the current *BioSample*, then do the following for each set of these data files; otherwise, continue to Step 2.

If you are carrying out this procedure using either the sample iTRAQ data or the sample SILAC data provided by Proteome Software, then continue to Step 2..

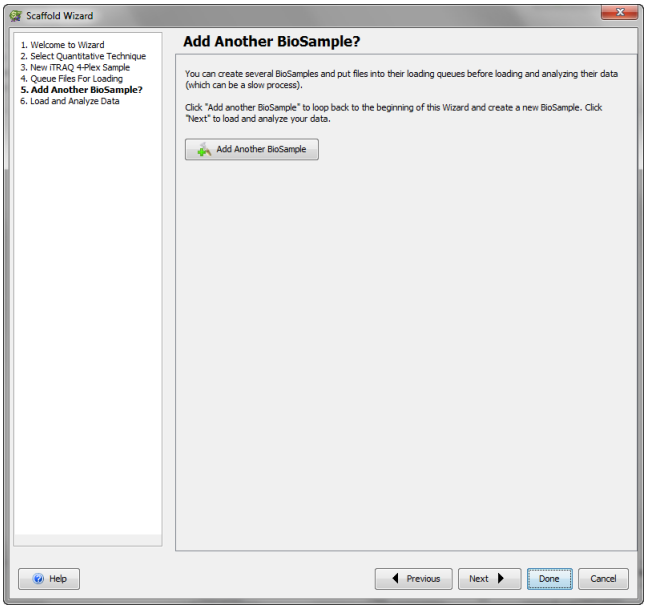
-

- Click Queue More Files For This BioSample.
- Repeat “Queuing sample files for loading”

2. Click Next.

   The Scaffold Wizard, Add Another Quant Sample? page opens.

Figure 2-9: Scaffold Wizard, Add Another BioSample? page



3. Continue to “Adding another BioSample” on page 30.

-

Adding another BioSample

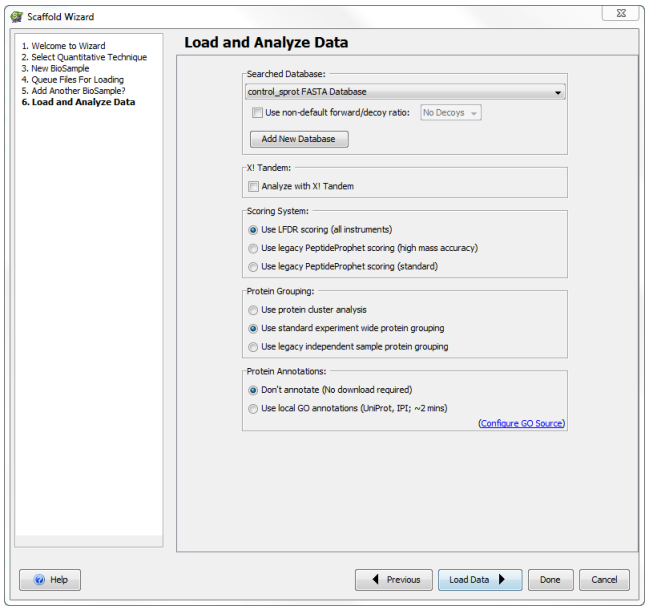
1. Do one of the following:
   - If you have other *BioSamples* that are to be analyzed, then for each of these BioSamples, click Add Another BioSample to return to page 2 of the Scaffold Wizard, cycle through the wizard to add the sample, and then click Next.
   - If you do not have other BioSamples that are to be analyzed, then click Next.

If you are carrying out this procedure using either the sample iTRAQ data or the sample SILAC data provided by Proteome Software, then click Next.

Load and Analyze Data

The Scaffold Wizard, Load and Analyze Data page opens. This Window is divided into various panes, each of them addressing specific analysis options available in Scaffold.

Figure 2-10: Scaffold Wizard, Load and Analyze Data page



2. Continue to "Specifying analysis options and analyzing the data" on page 31.

-

Specifying analysis options and analyzing the data

1. **Searched Database Pane** -- which allows to select or import the database used for the search data files loaded in Scaffold. Databases previously loaded will appear listed in the pull down menu. To add new databases click the Add New Database button. For more detailed information continue to "Specifying the fasta database" on page 33.
2. **Analyze with X!Tandem Pane --** Select this option to run an additional database search on your loaded data, which improves protein identifications, but significantly increases analysis times. For more information continue to "Running X!Tandem" on page 37

If you are carrying out this procedure using either the sample iTRAQ data or the sample SILAC data provided by Proteome Software, do not select run with X!Tandem.

3. **Scoring System Pane** -- This pane allows the user to select which post processing scoring algorithm is applied when Scaffold analyzes the imported data.
   - ***Use LFDR Scoring*** - Scaffold4 includes a new algorithm for assessing the confidence level of peptides. Based on a Bayesian approach to LFDR (Local False Discovery Rate), the novel algorithm is particularly effective for QExactive and high mass accuracy data
   - ***Use Legacy PeptideProphet Scoring (high Mass Accuracy)*** - This option will use the standard PeptideProphet algorithm developed in previous Scaffold versions together with the high mass accuracy option
   - ***Use Legacy PeptideProphet Scoring (Standard)*** - Standard PeptideProphet with no high mass accuracy.

When the data set to be analyzed was not searched using the decoy option or against a decoy concatenated database, the Legacy PeptideProphet Scoring Option will be automatically selected.

4. **Protein Grouping Pane --**It shows options related to the grouping analysis performed by Scaffold over the list of identified proteins.
   - ***Use Protein Cluster Analysis*** - Since Scaffold 4, a new hierarchical grouping level was added above the Scaffold standard protein grouping, PEG, also referred to, in Scaffold previous versions, as protein groups. While similar to Mascot's hierarchical family clustering, Scaffold 4 clusters are created using added stringencies that often succeed in separating proteins into sets of biologically meaningful isoforms. Each cluster showed in the Samples View can be expanded or collapsed. The clusters sub-menu contains options for expanding or collapsing all of the protein clusters displayed in the Samples View with a single click.

-

When the check box is unselected the list of proteins will be analyzed using only the regular Scaffold grouping algorithm.

- **Use standard experiment wide protein grouping** - When selected proteins are grouped across all MS and all BioSamples
- **Use legacy independent sample protein grouping** - When selected proteins are grouped only across each MS sample.Select this option to inform Scaffold *not* to group proteins across all MS samples. Instead, each MS sample appears as if it has been loaded independently

.

If you are carrying out this procedure using either the sample iTRAQ data or the sample SILAC data provided by Proteome Software, do not select this option

5. **Protein Annotations --** Included options for searching the Gene Ontology annotations, GO terms, during loading:
   - ***Don't Annotate (No download required)***
   - ***Use local GO Annotations (Uniprot, IPI; ~ 2mins)***. If the GOA database is not configured, the option will appear grayed out. For activation click the link *Configure GO Source* and select a GOA database from the GO Term configurations window, GO Annotation Databases pane. If the database you are searching is not available click New database and import the GOA database of your interest.

If you are carrying out this procedure using either the sample iTRAQ data or the sample SILAC data provided by Proteome Software, then select Don't Annotate.

6. If you have selected to run X!Tandem continue to "Running X!Tandem" on page 37
7. Once all the options have been properly checked, click:
   - Load and Analyze Data.

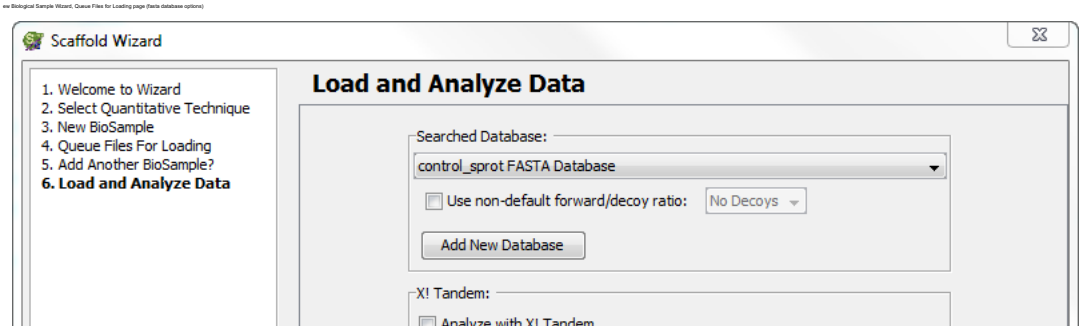
   A message opens, indicating that the data is being loaded and analyzed. After the analysis is complete, the data opens in the Samples View.

   Continue to Chapter 3, "Q+ Quantitation Module," on page 40.

-

Specifying the fasta database

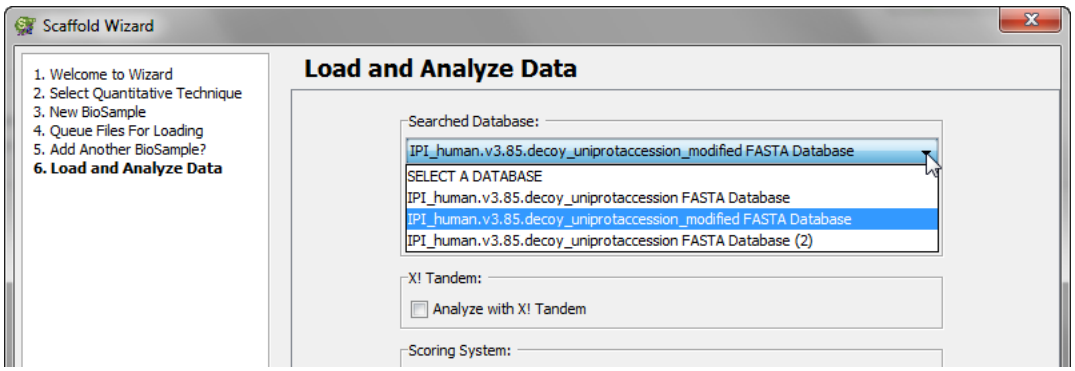
Figure 2-11: Searched Database Pane



1. Specify the fasta database that is associated with these sample files. You can select from a list of existing fasta databases shown in the pull down menu, or you can add a new fasta database.

   - Just select a database from the existing list .

Figure 2-12:



   - If you are adding a new database, continue to Step 2.

If you are carrying out this procedure using either the sample iTRAQ data or the sample SILAC data provided by Proteome Software, then you must add a new database.

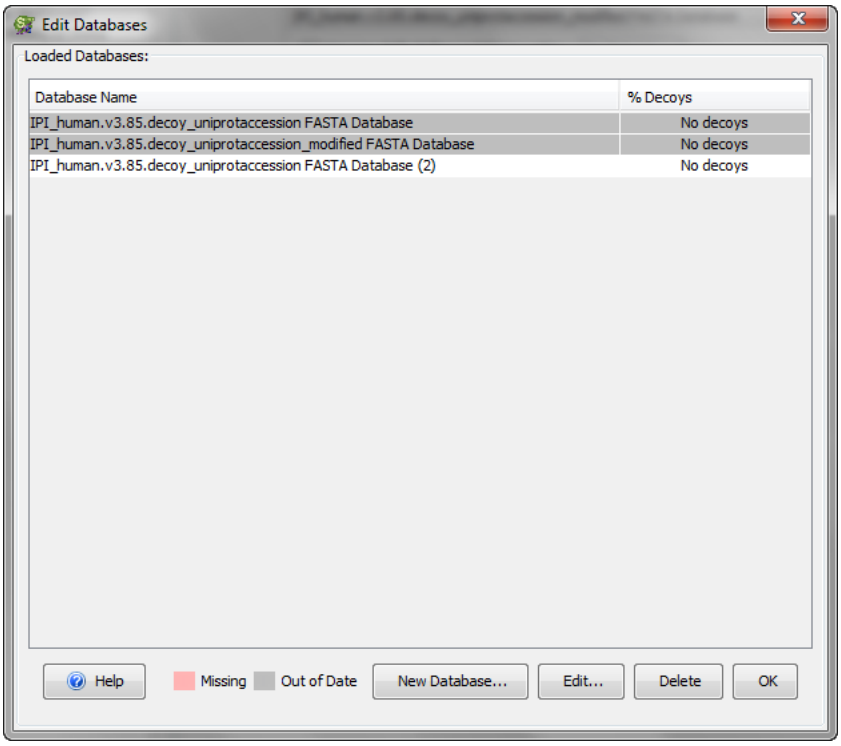
If the fasta database selected is not identical to the external protein database, including the version, that you used for searching your experimental data, then the protein sequence and molecular weight might not be available later in the Proteins View.

2. Click Add New Database.

   The Edit Databases dialog box opens.

-

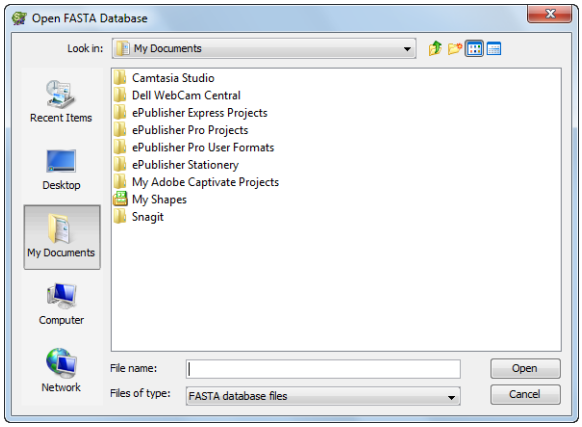
Figure 2-13: Edit Databases dialog box



3. On the Edit Databases dialog box, click New Database.

   The Open FASTA Database dialog box opens.

Figure 2-14: Open FASTA Database dialog box



4. Navigate to the directory in which you saved your sample data set and fasta database, select the fasta database, and then click Open.

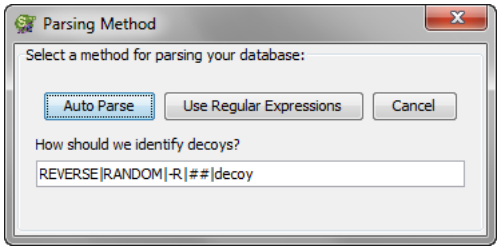
-

If you are carrying out this procedure using the sample iTRAQ data provided by Proteome Software, then select the iTRAQ_test.fasta database.

If you are carrying out this procedure using the sample SILAC data provided by Proteome Software, then select the SILAC-demo.fasta database.

The Parsing Method dialog box opens. You use the options on this dialog box to select the parsing rules that display protein accession numbers and protein descriptions in the correct format.

Figure 2-15: Parsing Method dialog box



5. Do one of the following:

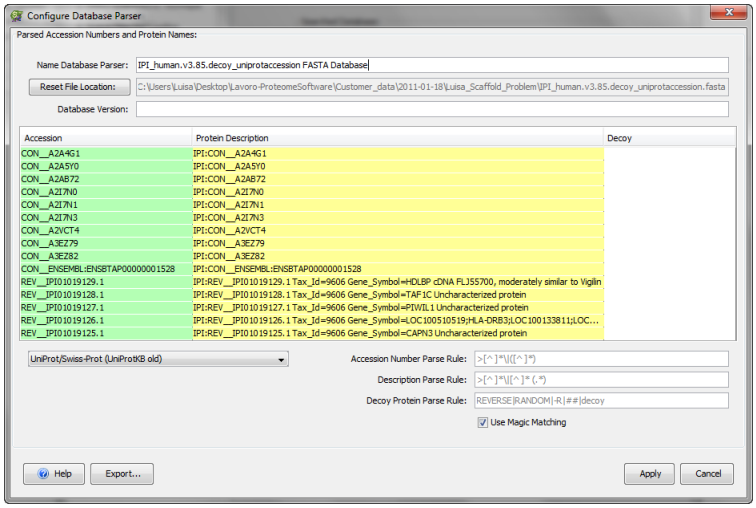
   - Click Auto Parse to have Scaffold decide on the parsing rules to use. If you are parsing a database that contains decoys make sure the decoy identification tag is included in the How should we identify Decoys? list shown in the dialog box.

Auto Parse is the preferred method for parsing the database. If you are carrying out this procedure using either the sample iTRAQ data or the sample SILAC data provided by Proteome Software, then select Auto Parse.

   - Click Use Regular Expressions to open the Configure Database Parser dialog box and select a specific pre-configured parsing rule to use, or create your own parsing rule. See Figure 2-16 on page 36.

-

Figure 2-16: Configure Database Parser dialog box



6. After the parsing rules are applied, you return to the Edit Databases dialog box with the correct database selected. Click OK.

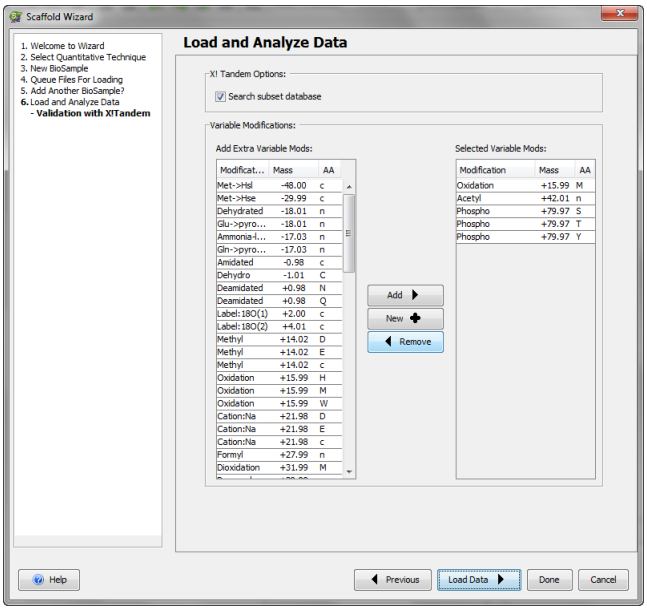
-

Running X!Tandem

To include X! Tandem results as part of your Scaffold run, check the box labeled Analyze with X! Tandem. X! Tandem runs with the same parameters as the original loaded search files. You may add additional variable modifications.

1. The Scaffold Wizard, Load and Analyze - Validation with X!Tandem page opens

Figure 2-17: Scaffold Wizard, Load and Analyze - Validation with X!Tandem



From the input files Scaffold reads the parameters used during the database search including instrument mass error tolerances, digestion enzymes, and expected fixed modifications. When X! Tandem is run, Scaffold passes these parameters to X! Tandem.

Scaffold lets you modify the list of variable modifications (PTMs) used in running X!Tandem on your data.

Selecting more variable modifications may increase the number of peptides identified. It will surely increase the run time for X!Tandem's analysis. If many modifications are chosen, it will take many times longer to execute -- think hours instead of minutes.

The modification tables can be sorted by clicking on the header for any column.

2. Add or remove modifications to or from the list already included in the original search data by using the arrows in the middle of the two lists.
3. Once all the options have been properly checked, click:
   - Load and Analyze Data.

-

A message opens, indicating that the data is being loaded and analyzed. After the analysis is complete, the data opens in the Samples View.

4. Continue to Chapter 3, “Q+ Quantitation Module,” on page 40.

-

-

Chapter 3
Q+ Quantitation Module

When labeled or precursor intensity quantitative data is loaded into Scaffold Q+S, the user has access to the Q+ Quantitation Module where quantitative data may be analyzed in depth. In the case of labeled data, multiplexes are teased apart into separate Quant Samples that the user can then organize to reflect the intended experimental design.

The Q+ Quantitation Module window launches when the user clicks the Q+ icon, available in Scaffold's Tool bar, or selects the command **Quant > Launch Q+ Quantitation Module** from Scaffold's main menu. These functionalities are accessible in Scaffold only when quantitative data are loaded.

The first time the user launches the Q+ Quantitation Module, the Experimental Design Wizard appears, to help the user set up the analysis. Then the **Q+ Quantitation Module** opens in the main Scaffold window. The window shows the quantitative data organized into various tables and views and contains tools for statistical analysis.

When invoked, the Q+ Quantitation Module analyzes the quantitative information for the proteins that meet thresholds in Scaffold. Note that both hidden and filtered proteins will still be considered for analysis.

If a particular protein should not be considered for quantitative analysis, Proteome Software suggests that the user examine the protein in the Scaffold Proteins View and deselect the peptides associated with the protein. This will lower the probability assigned to the protein and cause it to fall below the thresholds and be dropped from the Protein List.

To perform a correct analysis in Scaffold Q+S, it is important that the user be familiar with the fundamental concepts of statistical experimental design for mass spectrometry analysis and to design experiments based on statistically sound premises.

This chapter covers the following topics:

-

- “Log2 Fold Change” on page 56, which describes the way the Log2 Fold Change is cal.culated

-

Experimental Design for Mass Spectrometry Analysis

An *Experimental Design* specifies how samples are selected to provide meaningful information related to a particular hypothesis. Several terms are critical to understanding the fundamentals of statistical experimental design for mass spectrometry analysis of biological samples[1]. The table below defines these general industry-recognized terms and how they translate to Scaffold Q+S-specific terms.

Term	Definition	Scaffold Q+ Specific Term
Specimen	The biological sample material that is labeled for an experiment.	*Quant Sample* refers to a labeled specimen (specimen + tag).
Tag	The label that is applied to a specimen, for example, an iTRAQ or TMT label.	
Category	A distinct division that identifies some characteristic or property of the labeled sample, for example, diseased or healthy, or indicates how the labeled sample was manipulated, for example, treated and untreated.	*Quant Category* contains one or more Quant samples, which can be either technical replicates or biological replicates.
Experiment	A set of specimens that are subjected to mass spectrometry (MS) simultaneously.	Scaffold Experiment refers to the collection of Biosamples and MS Samples loaded in an SF3 file. Once the Q+ Quantitation Module is called upon, each BioSample is quantitatively analyzed and reorganized according to the Quant Samples that are included in the BioSamples. This collection is what is referred to as a Q+S Experiment".
Study	A collection of specific MS experiments that are used to test a specific hypothesis.	Experiment-An experiment in Scaffold can contain one or more MS Samples distributed in different BioSamples.

There are three main classes of experimental design for high-dimensional studies—the *Comparison* class, which compares groups of treated and untreated samples, the *Prediction* class, which involves the identification of prediction rules for the classification of new subjects, and the *Discovery* class, which involves the identification of unknown biological subtypes for a specific condition. Scaffold Q+S is designed to analyze experiments that fall into the Comparison class. Within this class are experiments that compare effects among different groups of samples and experiments that compare samples representing the same biological subjects under different treatments or different time frames (paired experiments).

1. Statistical Methods for Quantitative Mass Spectrometry Proteomic Experiments with Labeling, Ann L Oberg and Douglas W. Mahoney, BMC Bioinformatics 2012, 13(Suppl 16):S7 doi:10.1186/1471-2105-13-S16-S7. Published: 5 November 2012

-

The extent of the differences among the categories is expressed in terms of *fold change*, which is the ratio of the reporter ion or MS1 ion intensity to a reference ion intensity. The degree of up-regulation or down-regulation is indicated in the application with green or red color-coding, respectively.

To account for the differences between these two experimental designs, the reference ion intensity can be calculated in one of two ways—Average Protein References or Individual Spectrum Reference. See *Chapter 8, "Quantitative Settings," on page 111.*

By using replication, randomization, and blocking in the experimental design of quantitative MS studies, the user can minimize confounding of experimental and biological effects and can minimize variability. Moreover, a thorough understanding of these concepts prior to designing an experiment can ensure that a sufficient amount of data is collected and that the data will be collected in a manner that produces statistically sound conclusions.

Scaffold Q+S requires the user to classify each experiment as one of three possible experimental designs it supports:

- **Between-Subjects Design** - This type of experiment, also called Between-Group Design, refers to an experiment that has two or more groups of subjects which each group representing a different testing factor. The simplest Between-Subjects Design occurs with two groups or categories; one is generally regarded as the *treatment group*, which receives the 'special' treatment, (i.e. the group is treated with some variable) and the *control group*, which receives no variable treatment. The control group is used as a reference to support the claim that any deviation in results from application of the variable to the treatment group is, indeed, a direct result of the variable.

  The Between-Subjects Design is further subdivided in Scaffold Q+S into two different types depending on the reference groups:

  - **Independent Groups** - measurements in the reference group are taken from different subjects in the study. Here the goal is to assess difference between the control group and the treatment groups.
  - **Common/Pooled Reference** - in order to normalize measurements across different multiplexes, a group of samples or a pooled sample consisting of a combination of all samples in the reference group is/are included in each multiplex. In this case, the intent is to study differences betweem the treatment groups while using the reference samples as a means to control for experimental variation between MS runs.
- **Repeated measure/Time Course** - This type of design compares two or more different treatment conditions (or compares treatment and control) by measuring the same of individuals in all of the treatment conditions being compared. The Repeated measure/ Time Course or Within-Subject Design tests for changes from the reference state as a result of the treatment conditions. It allows for differences in the baseline levels of individuals and attempts to assess whether the treatments produce similar changes from the baseline in different individuals. It is called a repeated measure design because the

-

study repeats measurements of the same individuals under different conditions or investigates changes occurring over time. This type of design should be considered when the user can answer affirmatively the question: Have you taken samples from a given subject under different conditions?

-

Experimental Design Wizard

This wizard is launched whenever the user has loaded a new quantitative experiment in Scaffold or Scaffold Q+S and starts the Q+ Quantitation Module window for the first time either by clicking on the Q+ Quantitation icon  in the Scaffold Tool Bar or by using the command **Quant > Launch Q+ Quantitation Module**.

The wizard is composed of five different pages which guide the user through the process of organizing quantitative data according to the experimental design.

The left pane of the wizard's dialog lists the five pages in the experimental design wizard, while the right pane shows the design options available to the user in the current page.

- 1. Analysis Type
- 2. Experiment Type
- 3. Edit Samples Names and Categories
- 4. Organize Quant Samples
- 5. Approve Settings

The four functional buttons located at the bottom of the wizard's dialog, allow the user to navigate between dialog pages, cancel the operation or finalize the sample organization by clicking "Done". A fifth button opens the Help Online functionality to guide the user through the process.

1. Analysis Type

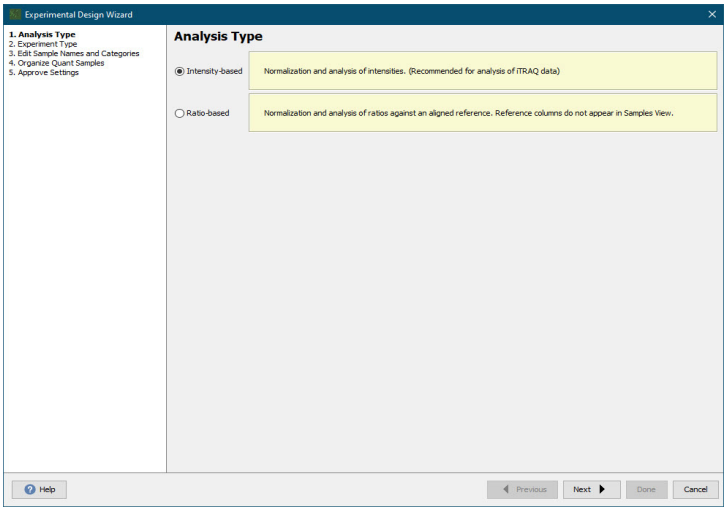
The first page of the Experimental Design Wizard asks the user to define the type of analysis which is most suitable for the experiment, Figure 3-1. The two options available are:

- **Intensity-based** - Normalizes and analyzes the Log_2 Intensity values of the quantitative raw data. Recommended for precursor intensity, iTRAQ and TMT data.
- **Ratio-based** - Normalizes and analyzes the Log_2 Fold Change of the quantitative raw data. Recommended when working with Stable Isotope labeling such as SILAC.

After selecting the desired Analysis Type, the user proceeds to page 2. Experiment Type of the Experimental Design Wizard by choosing the "Next" button at the bottom of the page.

-

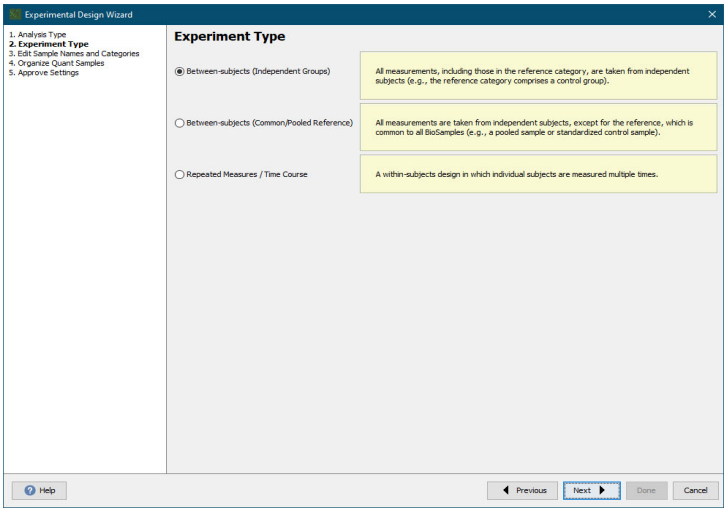
Figure 3-1: Experimental Design Wizard: Analysis Type



2. Experiment Type

The three experimental designs supported in Scaffold Q+ and Scaffold Q+S appear as selectable options in the Experimental Design Type pane on the second page of the Experimental Design Wizard, Figure 3-2.

Figure 3-2: Experimental Design Wizard: Experiment Type



Experimental Design Type pane

This pane contains the choices of experimental designs supported by the application, see Experimental Design for Mass Spectrometry Analysis 42.

- **Between-subjects (Independent Groups)**-- By “between-subjects” we mean that all measurements are taken from different subjects in the study. An example would be a study consisting of a control group and one or more treatment groups. Ratios are

calculated with respect to the average (median or mean) of all values in the Reference category.

- **Between-subjects (Common/Pooled Reference)** -- This design also implies measurements taken from different subjects, however the reference category consists of a common reference (i.e. a pooled sample or a standardized common sample)
- **Repeated measures / Time course** -- This is a design within subjects, where a measurement has been taken for different treatments applied to the same subject or at different times.

After selecting the desired experimental design option the user proceeds to page 3. Edit Samples Names and Categories of the Experimental Design Wizard by choosing the “Next” button at the bottom of the page.

3. Edit Samples Names and Categories

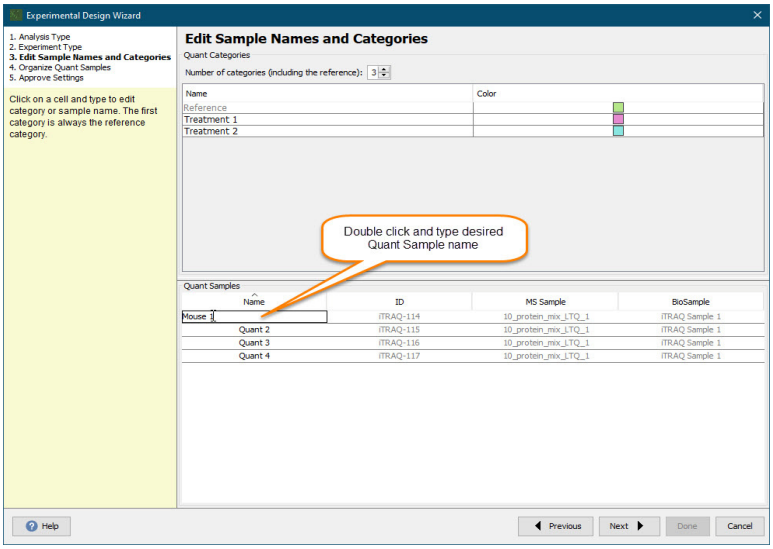
On this page the user defines the number of Categories or Time Points included in the experimental design and customizes the names assigned to the Categories and Quant Samples.

The page displays one number box and two tables: the “Category” table and the “Quant Samples” table.

- **Number of categories (including the reference)** - which defines the number of categories or time points included in the experiment.
- **Category table** - which is automatically populated with as many categories and their associated colors as the number appearing in the number box.
- **Quant Samples table** - which is already populated with default Quant Sample names, the Quant IDs, MS Sample and BioSample names contained in the loaded experiment. Depending on the experiment’s quantitation method, the Quant Ids will be labeled iTRAQ, TMT Tag, or Isotopic for SILAC data.

-

Figure 3-3: Experimental Design Wizard: Edit Samples Names and Categories.

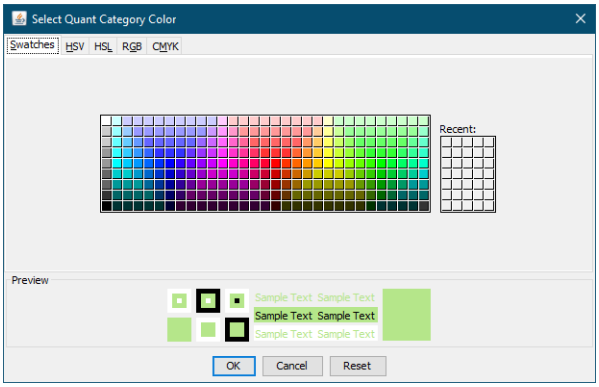


All Categories are automatically assigned a default name and color for easy identification in the Samples View. All Quant Samples appear with a default name as well . The user can leave the default names and colors or edit the names and change the colors at will. Double clicking on a Name cell in both tables allows editing of the names.

To select a different color:

1. Click on a color cell in the Category table to open the Select Quant Category Color dialog opens, Figure 3-4.
2. Click a color on the Swatches tab or select the HSB or RGB tab, to select the desired color value, and then click OK.

Figure 3-4: Select Quant Category Color Dialog



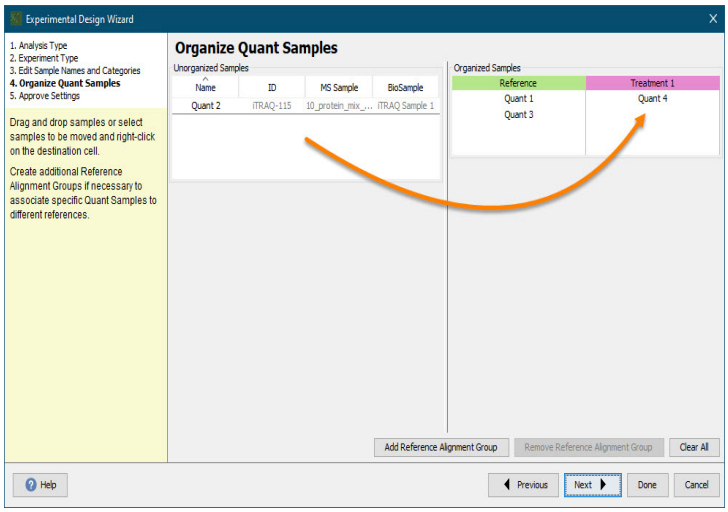
After completing any edits of Quant Sample or Quantitative Category names the user proceeds to page 4. Organize Quant Samples of the Experimental Design Wizard by choosing the “Next” button at the bottom of the page.

-

4. Organize Quant Samples

In this page of the wizard the user assigns Quant Samples to the Quantitative Categories defined in "3. Edit Samples Names and Categories" on page 47.

Figure 3-5: Experimental Design Wizard: Organize Quant Samples



The page includes two tables and three buttons:

- **Unorganized Samples Table** -- This table contains a list of the Quant samples not yet assigned to a Quantitative Category with their Quant IDs and BioSamples names.
- **Category Table** -- Each column of this table represents one of the previously defined Quantitative Categories and shows the list of Quant Samples included in each category. This table has a right click context menu, the Category Table Context menu, which contains various commands useful for assigning samples to the different categories.
- **Add Reference Alignment Group button** - This button allows the user to specify which reference should be used to calculate ratios for a number of specific Quant Samples. When clicked the button adds a new row to the Category table referred to as Reference Alignment Group (RAG). The user will distribute the different Quant Samples under the various categories along the new row. The ratios will then be calculated using the reference indicated in the row. Any number of new rows can be added to the table.
- **Remove Reference Alignment Group Button**- This button removes the last Reference Alignment Group (RAG) added. If the deleted group contained Quant Samples they will revert back to the Unorganized Samples table.
- **Clear All Button**- This button empties the Category tables. If Reference Alignment groups were added they will be removed.

By default, when the Experimental Design Wizard is first used, all the Quant Samples that have been loaded and analyzed for the current Scaffold session are displayed in the Unorganized Samples table, while the Category table just shows the Category headings over

-

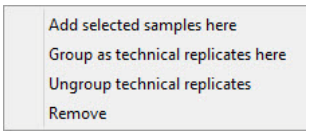
empty columns.

To organize Quant Samples into Quantitative Categories:

1. Add samples to a specific Quantitative Category by selecting one or more Quant Samples from the Unorganized Samples table. To select multiple reporter or MS1 ions or precursor intensity Quant Samples in consecutive rows, press and hold the SHIFT key. To select multiple reporter or MS1 ions or precursor intensity Quant Samples in non-consecutive rows press and hold the CTRL key. Then drag the selected Quant Samples to the chosen category column or select the “Add Selected Samples Here” command from the right click Category Table Context menu activated under the chosen category.
2. Within a RAG add Quantitative Samples to a Reference Category to compare them to the samples assigned to the other Quantitative Categories.
3. Group technical replicates Quant Samples when needed by selecting them and using the “Group as technical replicates here” command from the right click Category Table Context menu activated under the chosen category.
4. Remove one or more Quant Samples from the assigned Quantitative Category or categories by selecting assigned Quant Samples and dragging the selection back to the Unorganized Samples table or by selecting the “Remove” command from the right click Category Table Context menu. The selected Quant Samples will then reappear in the Unorganized Samples table.

Category Table Context menu

Figure 3-6: Category Table: right click context menu



This menu contains commands that help the user organize the Quant data in different categories:

- **Add Selected categories here** - Assigns whatever selected Quant samples in any of the two tables to the category where the context menu has been activated from.
- **Group as technical replicates here** -- Whatever Quant samples are selected they will be considered and grouped as technical replicates and assigned to the category from where the right click option has been activated from.
- **Ungroup Technical Replicates** - This command undoes the technical replicate assignment.
- **Remove** - Removes the selected Quant Samples and places them back to the Unorganized Quant samples table

After organizing the Quant Sample the user proceeds to page 5. Approve Settings of the Experimental Design Wizard by choosing the “Next” button at the bottom of the page.

-

5. Approve Settings

This page provides a tabular list of recommended quantitative settings that are based on the user's experimental design choices selected in the previous sections of the Experimental Design Wizard. The page includes a table listing the suggested selections, see Experimental Design Type Default Settings 159.

To change any design parameters, the user can open the Quantification Setup dialog by clicking the button "Edit settings..." which is located under the table of recommended settings. Changes to the settings are finalized by clicking "Apply". The Experimental Design Wizard will update the table and display the new settings.

Finally, clicking the "Finish" button applies the new organization and settings. When the calculations are complete the Q+ Quantitation Module window will appear to replace the Scaffold window, showing the Quantitative Samples View.

-

Setting up the Q+ Quantitation Module window

After the user has loaded and analyzed the data, the classic Scaffold Samples View opens. The user then should follow the procedure described below to set up the Q+ Quantitation Module.

1. On Scaffold's tool bar, click the Q+ Quantitation icon  . The Experimental Design Wizard opens to its first page, **1. Analysis Type**, see Figure 3-1.
2. Select the type of analysis you want Q+ to use for your experiment then click the "Next" button.

Proteome Software recommends that you select the analysis type as follows:

- iTRAQ or TMT dataset - Intensity-based analysis
- SILAC or dimethyl dataset - Ratio-based analysis
- Precursor Intensity dataset - Intensity-based analysis

3. From the **2. Experiment Type** page select the type of design that reflects your experiment from the list of available options, then click the "Next" button.

.

If using a dataset provided by Proteome Software, select the experiment type as follows:

- iTRAQ dataset - Between-Subjects (Independent Groups)
- SILAC dataset - Between-Subjects (Independent Groups)
- Precursor Intensity dataset - Repeated Measures / Time Course

4. From the **3. Edit Samples Names and Categories** page of the Experimental Design Wizard, assign the number of categories or time points appropriate to the experiment. If desired, rename the Quantitative Categories and/or Quant Samples and change colors assigned to any category. Once complete click the "Next" button.

If using a dataset provided by Proteome Software, assign the following number of categories or time points:

- iTRAQ dataset - Select two categories
- SILAC dataset - Select three categories
- Precursor Intensity dataset - Select 3 time points

5. From the **4. Organize Quant Samples** page of the Experimental Design Wizard assign samples to Quantitative Categories by either dragging and dropping the Quant Samples from the Unorganized Samples table or selecting them and then using the right click options to assign them to a Quantitative Category. When complete click the "Next" button.

-

If using a dataset provided by Proteome Software, assign Quant Samples to Quantitative Categories as follows:

- iTRAQ dataset - Quant1 andQuant3 are technical replicate references, select them and place them in the Reference category. Then select them both, right click and select “Group as technical replicate here”. Quant2 and Quant4 are technical replicates of the same Quant Sample. Place them in Category 1 and define them as technical replicates as selected for the references.
- SILAC dataset - Light is the untreated reference while medium and heavy are two different treatments.
- Precursor Intensity dataset -

6. The **5. Approve Settings** page lists the recommended settings assigned by the application. To adjust the settings click the “Edit Settings” button and modify any parameters in the “Quantification Set up” dialog under the Normalization tab. When complete click the “Apply” button. Review the list of updated parameters and finalize the organization of the experiment’s Quant Samples by clicking the “Finish” button.

If using datasets provided by Proteome Software do not adjust the recommended parameters.

7. The Scaffold Q+ Quantitation window opens in the Quantitative Samples View.
8. To make additional changes to the organization of the experiment while in the Q+ Quantitation Module window select the menu command Quant > Update Experimental Design... to reopen the Experimental Design Wizard. Using the Previous and Next arrows located at the right lower corner of the dialog, move through the dialog pages to modify the experiment’s parameters as needed. When complete click the “Finish” button to finalize the new organization of the experiment.

-

Normalization method used in Scaffold Q+ and Scaffold Q+S

Upon launching the Q+ Quantitation Module, the quantitative raw data intensities are Log-transformed and normalized for bias removal. At the set up of the experiment, the user can choose from two possible normalization methods.

1. Intensity-Based Normalization
2. Ratio Based Normalization

Intensity-Based Normalization

Intensity-based normalization is developed on a model for the analysis of iTRAQ data described by Ann Oberg *et al.*[2] [3]. In her analysis, Oberg constructs an ANOVA model to account for variability and biases present in multiple-samples iTRAQ experiments.

This model considers the effects that influence the intensity of the observed reporter ion peaks, in particular: (1) sample handling effects across MS acquisitions, (2) sample handling effects between channels, (3) Peptide, ionization effects and other errors due to the imprecision of measurements. It adjusts for these experimental effects through normalization so that the remaining differences may be considered to be biological effects and maybe analyzed by ANOVA.

Her method works quite well when dealing with Log normally distributed intensities with few missing values.

Since proteomics experiments may have many missing values, however, Scaffold Q+ and Scaffold Q+S allow the option of employing a non-parametric median-based variant of the Oberg's model (Tukey's median polish as described in the Encyclopedia of Statistical Science[4]). .This new method creates a more robust normalization scheme for non-normal type of data.

Within Scaffold Q+ or Scaffold Q+S normalization is accomplished by imputing missing values, see Pre-Processing, log transforming the raw data and applying intensity weighting, then applying an iterative normalization procedure at the following levels (where average is defined as either mean or median, depending on the parameter settings):

1. **Global normalization at the MS level** - Adjust the values within each MS Sample so that all MS Sample averages are equal.
2. **Normalization at the channel level** - Adjust values within each Quant Sample so that each Quant Sample average is equal to the average at the MS Sample level.

---

2. Ann L. Oberg, Douglas W. Mahoney, Jeanette E. Eckel-Passow, Christopher J. Malone, Russell D. Wolfinger, Elizabeth G. Hill, Leslie T. Cooper, Oyere K. Onuma, Craig Spiro, Terry M. Therneau and H. Robert Bergen, Statistical analysis of relative labeled mass spectrometry data from complex samples using ANOVA, J. Proteome Res. 2008 January; 7(1):225-233 DOI: 10.1021/pr700734f
3. Elizabeth G. Hill, John H. Schwacke, Susana Comte-Walters, Elizabeth H. Slate, Ann L. Oberg, Jeanette E. Eckel-Passow, Terry M. Therneau, and Kevin L. Schey, A STATISTICAL MODEL FOR iTRAQ DATA ANALYSIS, J Proteome Res. 2008 Aug; 7(8): 3091–3101. Published online 2008 Jun 26. DOI: 10.1021/pr070520u
4. Encyclopedia of Statistical Sciences, Copyright © 2006 John Wiley & Sons, Inc.

-

3. **Within protein normalization** - For each protein, adjust the partially normalized Log_2 Intensity values of each spectrum multiplex so that the average of all channels in the individual spectrum matches the average spectrum multiplex average of all spectra in that Quant Sample for the protein.

Ratio Based Normalization

Ratio based normalization operates on Log_2 intensity ratios rather than on the intensity themselves. Note that currently for ratio based analysis, only mean mode with individual spectrum reference is available.

Missing values are imputed, see Pre-Processing, then the ratio of each value to the reference value within that spectrum is calculated. These ratios are logged transformed, and for each Quant Sample the average log-ratio is determined.

The average Log_2 Ratio is then subtracted from each Log_2 Ratio in the Quant Sample with the result that the average Log_2 Ratio for each Quant Sample is zero. This allows Quant Samples from different MS Samples to be combined without further normalization.

For more details see Appendix “Normalization” on page 152.

-

Log_2 Fold Change

Log_2 Fold Change is a measure of the differential abundance of a protein between one sample and another. For example, the Log_2 Fold Change of +3 means that the intensity has increased by a factor of $2^3 = 8$, while a Log_2 Fold Change change of -1 means that intensity has decreased to $2^{-1} = ½$ its original value.

Notes on Logarithmic Values

When intensities have been transformed by applying a base 2 logarithm, ratios become differences:

$$Log_2(x/y) = Log_2 x - Log_2 y$$

For any x and y values. Furthermore it is important to note that means of logarithms correspond to the logarithm of the geometric mean:

$$(Log_2 r_1 + Log_2 r_2 + ... + Log_2 r_n)/n = Log_2((r_1 \cdot r_2 \cdot ... \cdot r_n)^{(1/n)})$$

Where r_j for j=1,n are ratios. However the same is not true for medians, where the median of the logs is (roughly) the log of the medians.

Computing the Log_2 Fold Change

The way the Q+ Quantitation Module calculates the Log_2 Fold Change depends on the type of analysis selected, on the way the Quant Samples are organized during the initial set up of the experiment in the Experimental Design Wizard and on the selection for the **Reference Type:** option listed in the The goal of quantitative analysis in Scaffold Q+ and Scaffold Q+S is to detect differential expression of a protein between Quantitative Categories. The quantitative settings that the user specifies in the Quantification Setup dialog determine the type of evidence that is produced to support the presence or absence of differential expression. On the main menu of the Q+ Quantitation Module window, selecting Quant > Quantitative Settings opens the Quantification Setup dialog. dialog.

The basic approach is to first compute Log_2 Fold Change at the spectrum-level and then use these values to compute Log_2 Fold Change at higher levels.

Spectrum Level Log_2 Fold Change

Each spectrum value (Log_2 Intensity) in a given Quant Sample has a set of associated reference values. The make up of this set depends on the selected Reference Type (Average Protein Reference or Individual Spectrum Reference) and the organization of Quant Samples into Reference Alignment Groups (RAGs).

- If **Average Protein Reference** has been selected, reference values are taken only from all spectra belonging to the same protein in all Quant Samples listed as a Reference in the same Reference Alignment Group.

-

- If **Individual Spectrum Reference** has been selected, reference values are taken only from the *same* spectrum. If more than one Quant Sample in the same MS Sample is designated as a reference, the values in these Quant Samples in the individual spectrum are averaged.

Once the set of reference values has been determined, these Log_2 Intensities are averaged, using the Kernel Density Average method, to determine the Reference Log2 Intensity.

Log_2 Fold Change at the spectrum level, for a given Quant Sample, is computed as:

$$Log_2(\text{Value/Reference Intensity}) = Log_2(\text{Value}) - Log_2(\text{Reference Intensity})$$

Peptide- or Protein-Level Log_2 Fold Change

The spectrum-level Fold Changes described above may be rolled up to provide a Log_2 Fold Change at the peptide-level (in the Proteins View) or the protein-level (in the Samples View). This roll up uses Kernel Density Average of the spectrum-level ratios in the selected mode (mean or median). For example, in median mode the Log_2 Fold Change displayed for a protein in the Samples View is the Kernel Density Average for median of the ratios over all spectrum-level ratios in the given Quant Sample.

When **Ratio-based** analysis is selected, Q+ calculates the fold changes in a similar manner as described for the Intensity based analysis but starts form the normalized Log_2 Fold Change values.

If the experiment includes technical replicates, a Kernel Density Average is first applied to the Log_2 Intensities or Log_2 Fold Changes for MS1 ion intensity values among the technical replicates present in each category within a RGA.

-

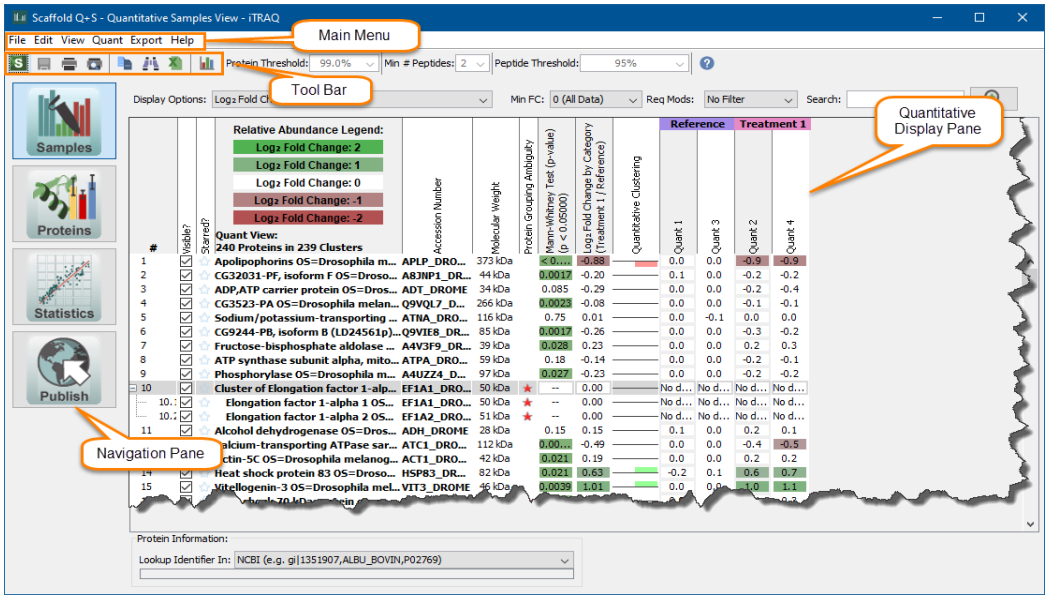
Chapter 4
Q+ Quantitation Module Main Window

The Q+ Quantitation Module main window is the user's starting point for analyzing quantitative data in Scaffold Q+S. This window provides quick access to all of the Q+S quantitative features and functions.

The window has the following major components:

- Title bar
- Main menu commands
- Toolbar
- Navigation pane
- Display pane

Figure 4-1: Q+ Quantitation Module window



-

Title bar

"Scaffold Q+S", depending on the type of license, is always displayed in the title bar at the top of the Q+ Quantitative Browser window. Additional text is displayed in the title bar depending on the actions that the user is currently carrying out in Q+ Quantitative Browser. For example, if the user has opened the Quantitative Samples View for a dataset, then "Quantitative Samples View - <dataset name> is displayed in the title bar.

Figure 4-2: Title bar

The version of

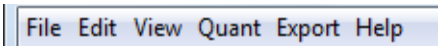
Scaffold Q+S that the user is running is not displayed in the Title bar. The Help > About option in the main menu includes the version number. See "Main menu commands" below.

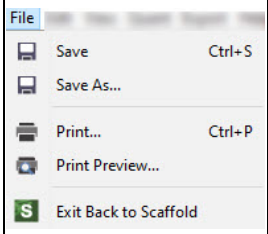
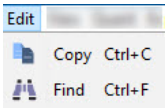
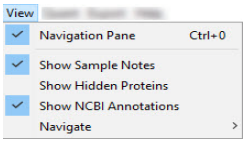
-

Main menu commands

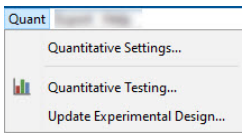
The Q+ Quantitative main menu is set up in a standard Windows menu format with menu commands grouped into menus (File, Edit, View, Quant, Export, and Help) across the menu bar. Some of these menu commands are available in other areas of the application.

Figure 4-3: Q+ Quantitative Browser main menu



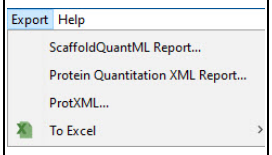
Menu	Menu Commands
File 	• **Save**—Standard Windows behavior. • **Save As**—Standard Windows behavior. • **Print**—Standard Windows behavior. • **Print Preview**—Standard Windows behavior. • **Exit Back to Scaffold**—Closes the Q+ Quantitative Browser window and returns to the Scaffold window.
Edit 	• **Copy**—Applicable for the Quantitative Samples View and The Scaffold Q+ Quantitative Proteins View provides an overview of the peptides quantitative detection for a selected protein group.. Copies the currently displayed Proteins table to the clipboard, and from there, you can paste it into a third-party program such as Excel or Microsoft Word. • **Find**—Applicable for the Quantitative Samples View and The Scaffold Q+ Quantitative Proteins View provides an overview of the peptides quantitative detection for a selected protein group..Searches for characters in the tables included in the views.
View 	• **Navigation pane**—Toggles the display (on or off) of the Navigation pane. • **Show Sample Notes**—Toggles the display (on or off) of the Protein Information pane that is displayed at the bottom of theQuantitative Samples View. • **Show Hidden Proteins**—Toggles the view of Hidden Proteins. • **Show NCBI Annotations**—Toggles the display (on or off) of Gene Ontology (GO) terms for the currently displayed proteins while in the Quantitative Samples View or The Scaffold Q+ Quantitative Proteins View provides an overview of the peptides quantitative detection for a selected protein group.. • **Navigate**—Select Previous Tab or Select Next Tab. Moves you through the various chart tabs that are displayed in the lower pane of the The Scaffold Q+ Quantitative Proteins View provides an overview of the peptides quantitative detection for a selected protein group..

-

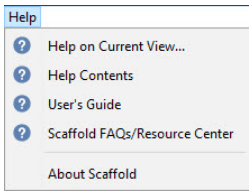
Menu	Menu Commands
Quant 	• **Quantitative Settings...**—Opens the Quantification Setup dialog dialog. You use the various options on this dialog box to specify the options for generating the quantitative data. • **Quantitative Testing...**—Opens the dialog box which displays a list of statistical tests that are available for the selected normalization scheme. • **Update Experimental Design**—Opens the Experimental Design Wizard where the user can modify the design initially set up when the Q+ Quantitation Module was first launched.

-

Menu	Menu Commands

Export
 
	• Scaffold QuantML Report...—Exports the protein, peptide identification data and reporter or MS1 ion quantitation data in a file format similar to mzIdentML called ScaffoldQuantML. Also exports the spectrum information as MGF files. The user can load the exported information into Scaffold PTM for protein translation modifications analysis. For more information, go to http://www.proteomesoftware.com/Scaffold_PTM/Scaffold_PTM.html.
	• Protein Quantitation XML Report...—Exports the protein quantitative expression to use for protein normalized quantitative PTM experiments analyzed in Scaffold PTM. For more information, go to http://www.proteomesoftware.com/Scaffold_PTM/Scaffold_PTM.htm
	• ProtXML...—Exports all quantitative data in the protXML format, which is an open XML file format for the storage of data at the raw spectral data, peptide, and protein levels. This format enables uniform analysis and exchange of MS/MS data generated from a variety of different instruments, and assigned peptides using a variety of different database search programs. See Molecular Systems Biology, 1:2005.0017 for more information.
	• To Excel:
	  • Publication report...—Generates the Publications report, which lists the data analysis information that is required for publications in many proteomics journals. The report is a copy of the Publish View.
	  • Samples report...—Generates the Samples Report, which mimics the Quantitative Samples View.
	  • Samples report with Isoforms...—Generates the Samples report, but also includes the expanded protein families in the Proteins list.
	  • Quantitative Peptide report...—Generates the Peptide report, which details all the peptides that pass the current filter settings.
	  • Quantitative Spectrum report...—Generates the Spectrum report, which details all the spectra that pass the current filter settings.
	  • Raw Data report...—Generates the Raw Data report, which contains the raw intensity data values for every protein in each sample as well as the normalized values that are used in the fold change and other calculations.
	  • Statistical Analysis report...—
	  • Protein Accession Number report...—Generates the Protein Accession Number report, which lists all the proteins that pass the current filter settings.
	  • Current View—Exports the information that is contained in the current view to a tab-delimited text file that can be opened and viewed in Excel.
	Note: For detailed information about each report, including a sample report, see Chapter 11, “Reports,” on page 143.

-

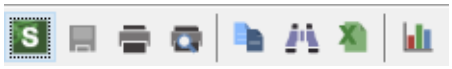
Menu	Menu Commands
Help 	• **Help on Current View**—Opens the Online Help that is specific for the currently displayed topic. • **Help Contents**—Opens the Contents page for the Online Help. • **User's Guide**—Opens the current Scaffold Q+ User's Guide. • **Scaffold FAQs/Resource Center**—Opens your default web browser to the Home page of the Scaffold Users' forum. • **About Scaffold**—Provides the release information for the current version of Scaffold Q+ or Scaffold Q+S, license information, contact information for Proteome Software, Inc.. It also reports information about the system where Scaffold is installed, the amount of memory available to the software and the percentage of memory used by the application.

-

Toolbar

The Q+ Quantitation Module tool-bar contains icons that represent equivalent commands for frequently used main menu options.

Figure 4-4: Q+ Quantitation Module tool-bar

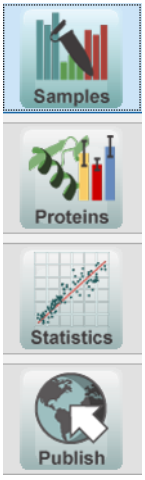


Icon	Function
	Scaffold—Closes the Q+ Quantitation Module and returns the user to Scaffold.
	Save—Standard Windows behavior.
	Print—Standard Windows behavior.
	Print Preview—Standard Windows behavior.
	Copy Data in Current View—Applicable for the "Quantitative Samples View" on page 72 and "The Scaffold Q+ Quantitative Proteins View provides an overview of the peptides quantitative detection for a selected protein group." on page 84. Copies the currently displayed Proteins table to the clipboard, and from there, the user can paste it into a third-party program such as Excel or Microsoft Word.
	Find—Find text in current view
	Excel—Exports the information that is contained in the current view to a tab-delimited text file that can be opened and viewed in Excel.
	Quantitative Testing—Opens the Mean-based Between-subjects Statistical Tests dialog
	Help—Opens the Scaffold Q+ Online Help.

-

Navigation pane

Figure 4-5: Q+ Quantitation Module Navigation pane



The Q+ Quantitation Module Navigation pane is a vertical bar that is displayed on the left side of the Q+ Quantitation Module window. The navigation bar contains buttons that toggle the among the four different views that are available for the Q+ Quantitation Module window—"Quantitative Samples View" on page 72,, "The Scaffold Q+ Quantitative Proteins View provides an overview of the peptides quantitative detection for a selected protein group." on page 84, "Quantitative Statistics View" on page 101, and "Quantitative Publish View" on page 141.

-

Display pane

The information included in the different views appears in the Q+ Quantitation Module Display pane. Depending on the view, the information reported might appear framed in one or more tables or graphs included in one or more sub-panes. All panes and tables included in the Q+ Quantitation Module share the following characteristics:

- Tool-tips
- Resizing of columns and panes
- Moving columns
- Column sorting feature
- Multi selection of rows in the Samples Table
- Mouse Right Click Context Menus
- Graph Features

Tool-tips

The user can view information about fields or columns in a View by just hovering the mouse pointer over the location of interest.This operation opens a collapsed tool-tip. Pressing F2 opens an expanded tool-tip. Pressing the Escape (ESC) key on the keyboard closes the expanded tool-tip,

Figure 4-6: Viewing information in a collapsed tool-tip

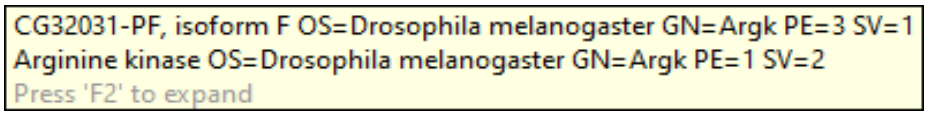
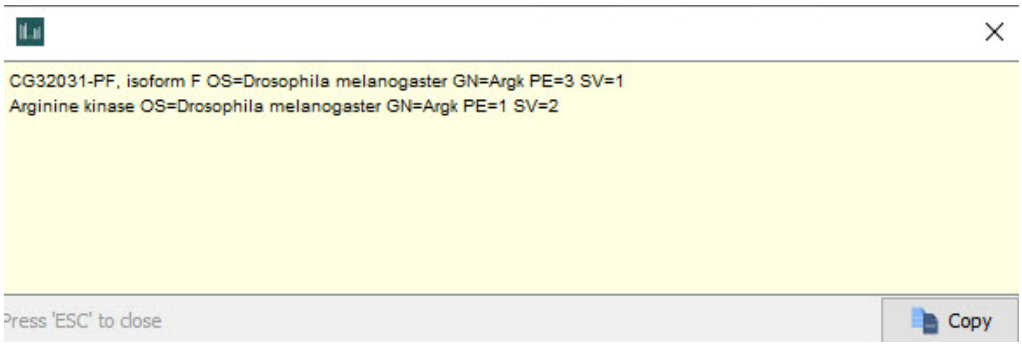


Figure 4-7: Viewing information in an expanded tool-tip



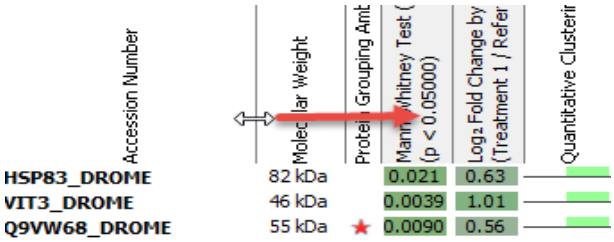
Resizing of columns and panes

The user can resize columns and different panes in each of the views to better suit his/her working needs. For example, in the Samples Table, the user can change the width of a column by resting the mouse pointer on the right side of a column heading until the pointer

-

changes to a double-headed arrow, and then dragging the boundary until the column is the width that he or she wants.

Figure 4-8: Changing the width of a column in the Quantitative Samples View

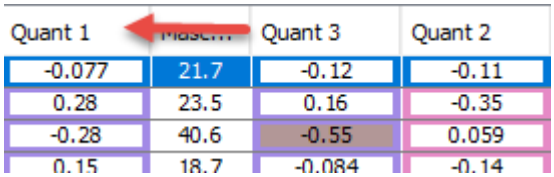


Moving columns

In all tables throughout Q+ Quantitation Module, except the Samples Table, every column can be moved from one position to another for easier access.

Simply click on the header of the column and drag it to the desired location. The new position is retained when the user switches to a different view and then returns.

Figure 4-9: Moving columns in tables



Column sorting feature

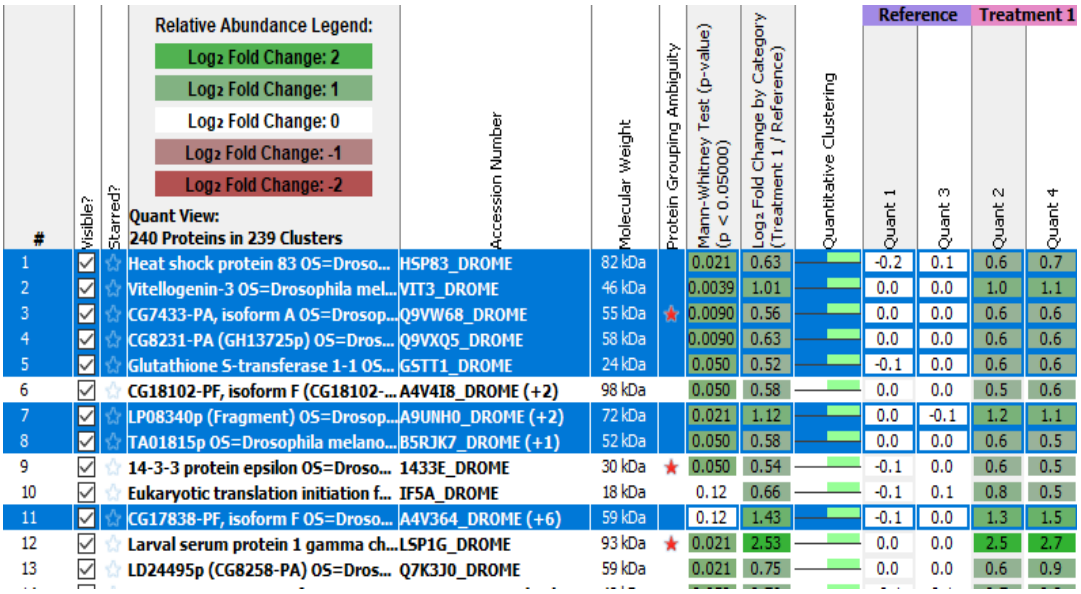
In all tables throughout Q+ Quantitation Module, the user can utilize the tri-state column sorting feature and sort the display by clicking on any column header. For example, to sort the proteins based on increasing molecular weight, the user can click the Molecular Weight column header once. To sort the proteins based on decreasing molecular weight, the user can click the Molecular Weight column header twice. To return to the default display, the user can click the Molecular Weight column header a third time.

Multi selection of rows in the Samples Table

In the Samples table the user can select multiple rows by using either the SHIFT or the CTRL key, depending on whether the desired selection has contiguous rows or not, and the click of the mouse in a pretty standard fashion. Other functions can then be applied, like assigning a star to the selected group of proteins in the Samples table, for example.

-

Figure 4-10: Rows multi-selection in the Samples View



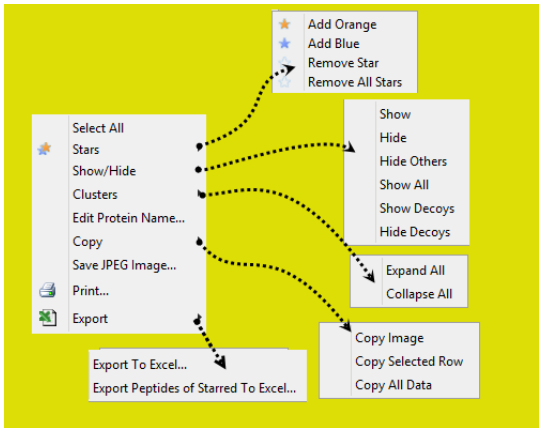
Mouse Right Click Context Menus

When the user right clicks the mouse while hovering over the Display Pane, a menu with various options appears. Depending on the selected view the list of options available in the menu varies. A description of the mouse right click command is provided in "Context menu Commands" on page 157.

Samples View

When the user right clicks anywhere over the list of proteins Context Menu A appears. It contains a number of sub-menus:

Context Menu A:

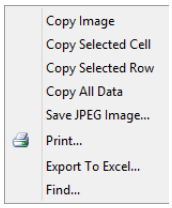


Proteins View

When the user right clicks over the top two tables appearing in the Proteins View, Context Menu B appears.

-

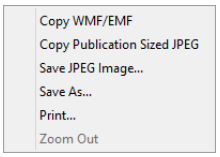
Context Menu B:



When the user right clicks on the content of the tabs appearing in the lower section of the Proteins View the following context menus appears:

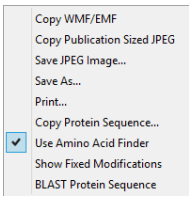
- General tabs context menu:

Context Menu C:



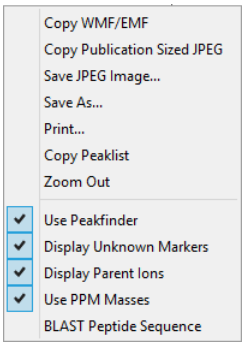
- Protein Sequence tab context menu:

Context Menu D:



- Spectrum Tab context menu:

Context Menu E:



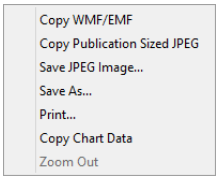
Statistics View

- Sample Wide Fold Changes pane - When right clicking over this pane Context Menu C appears.
- Intensity Weighting and Error Estimation:

-

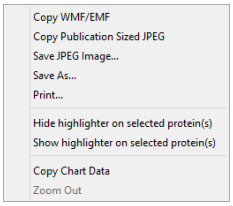
- Intensity Weighing tab: when right clicking over the content of this tab Context Menu C appears.
- Average CVs context menu:

Context Menu F:



- Normalized Intensity Scatterplot context menu appearing in both tabs:

Context Menu G:



- Raw Intensity Distribution: Context Menu F appears in all tabs.

Publish View

When mousing over the Experiment Methods tab Context Menu B appears.

Go to Display pane

Graph Features

Every graph appearing in any of the Scaffold's view contain the following tools:

- **Zoom Function** - Zooming in within a graph is done by holding down the left mouse button and dragging the pointer from left to right. A box, drawn from the upper left hand corner of the graph towards the lower right hand corner, is formed and highlighted in gray around the area that, after releasing the mouse button, is being enlarged for viewing within the graph plotting area. A single click of the mouse zooms out the image to the previous magnification. Clicking various times returns the graph to the initial100% magnification.
- **Context menu** - The user can right-click on a graph to open a context menu. The type menu might depend on the view where the graph appears, check Mouse Right Click Context Menus and Context menu Commands for more information.

-

Chapter 5
Quantitative Samples View

The Q+ Quantitative Samples View displays information about the proteins that were identified in the Scaffold analysis and tools to help the user make direct comparisons among the different Quant samples analyzed in the experiment.

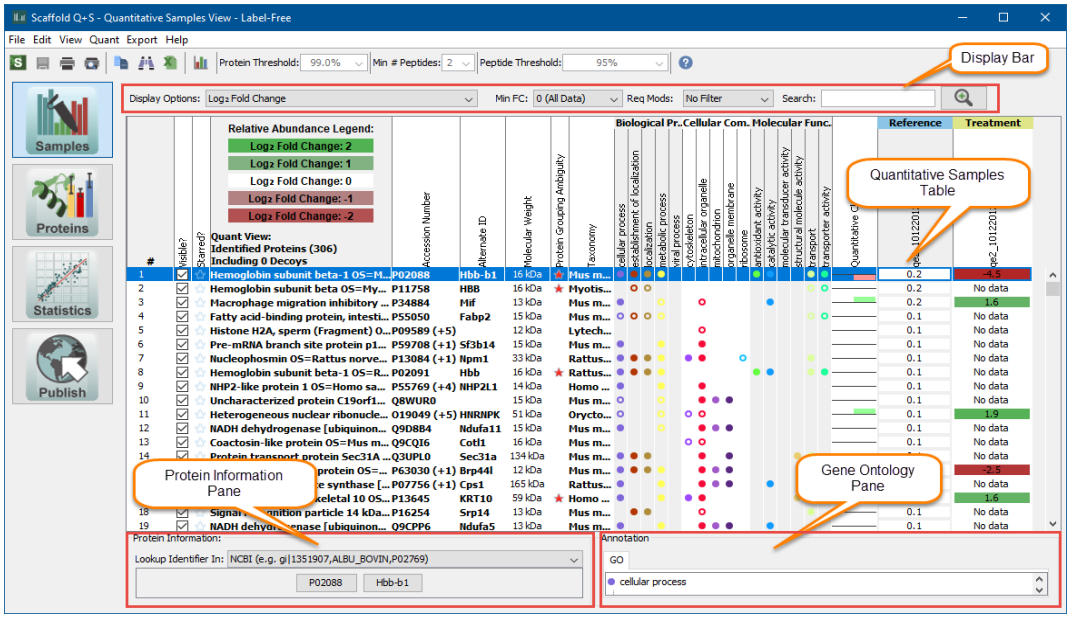
When the user calls the Q+ Quantitation Module from Scaffold, the Quantitative Samples View is the first view appearing in the Display pane. It provides a quantitative overview of the list of proteins appearing in the table.

The Quantitative Samples View is composed of the following elements:

- **Quantitative Samples Table** - A table showing, for each protein group included in the table, how a selected quantitative information is distributed among the different Quant Samples and Quant Categories. (See “Quantitative Samples Table” on page 74.)
- **Display Bar** - A bar containing a pull down list for modifying the quantitative information shown in the table, filters and search tools. (See “Display bar” on page 80.)
- **Protein Information Pane** - Displays information about the selected protein as identified in the database search. (See “Protein Information pane” on page 83.)
- **Gene Ontology Pane** - Present only when Gene Ontology Associations were searched in Scaffold and appear in the Quantitative Samples Table.

-

Figure 5-1: Q+ Quantitative Samples View



-

Quantitative Samples Table

The Quantitative Samples Table provides a list of the protein groups and clusters identified in a quantitative experiment. It also includes quantitative and qualitative information distributed among different Quant Samples and Quantitative Categories.

The table is a collection of Frequency Tables, one for each of the identified protein groups or protein clusters included in the Proteins List. This type of table is typically referred to in statistics as a Contingency table.

Each row displays quantitative information for a protein group or cluster in every Quant Sample represented as a column. The quantitative information shown depends on the option selected through the Display Options pull down list. The Quant Samples are grouped into Quantitative Categories that were defined during the experiment setup process, see "Experimental Design Wizard" on page 45.

The general characteristics of the Quantitative Samples Table are as follows:

- Quantitative Samples Table Features
- Thresholding
- Default sorting of columns
- Proteins List
- Relative Abundance Legend
- Quantitative value tags
- Quantitative Values

Quantitative Samples Table Features

The Quantitative Samples Table contains the same features and tools described in section "Display pane" on page 67 for the Q+ Quantitation Module, with one exception. Existing columns may not be moved in this table.

The same first eight columns found in Scaffold's Samples Table appear in the Quantitative Samples Table. Described below in order of their appearance, they provide basic information about the identified protein groups.

- **# --** Displays the order number of each row as specified by the current ordering conditions.
- **Visible? --** Shows a list of selected check boxes. Deselecting a box hides the corresponding row when the show hidden option in the menu command ***View > Show Hidden Proteins*** is turned off. See "Hidden Proteins" on page 77Hidden Proteins.
- **Starred?**-- Tags protein of interest. It initially displays a grayed out star for every row. By clicking over the star multiple times, the star will loop through four possible color states. The color of the star will go from gray to orange, to blue and to orange and blue then back to its original state. For more information see "Tagging Proteins of Interest: The Star Function" on page 76.

-

- **Protein Name** -- Lists the name of the protein group or cluster.
- **Accession Number** -- Displays the protein identification number as parsed when loading the database in Scaffold, see Scaffold User's Manual.
- **Molecular Weight** -- Provides the protein molecular weight calculated from the protein sequence.
- **Protein Grouping Ambiguity** -- Displays a star in any row containing a protein that shares some but not all peptides with other proteins. Specific functionalities are connected with this column in the Scaffold's Samples Table, for more information see the Scaffold User's Manual.
- **Quantitative Clustering** - Highlights *at a glance* the average level of up-regulation or down-regulation of a protein in each of the Quantitative Categories. See "Quantitative Clustering" on page 75.

Other columns appear after the Protein Grouping Ambiguity column when a Quantitative statistical test among the different Quantitative Categories is selected, see "Quantitative Testing" on page 121. Columns are also added when GO annotations are searched in Scaffold before invoking the Q+ Quantitation Module.

Quantitative Clustering

Every cell in this column is divided into color-coded segments, each of which corresponds by position to a Quantitative Category. An individual segment of the Quantitative Clustering column will display a color block if the level of up-regulation or down-regulation of the calculated protein ratio located in the corresponding Quantitative Category exceeds a pre-defined threshold.

- If the average Log_2 Ratio for a protein in a Quantitative Category is ≥ 0.5, a green color block appears in the corresponding segment of the Quantitative Clustering column.
- If the average Log_2 Ratio for a protein in a Quantitative Category is ≤ -0.5, a red color block appears in the corresponding segment of the Quantitative Clustering column.
- In all other cases undefined white space appears in the corresponding segment of the Quantitative Clustering Column.

The user can easily sort the up-regulated and down-regulated proteins in the display by using the tri-state column sorting feature. Clicking the Quantitative Clustering column header once sorts the protein display in order of decreasing relative abundance (up-regulated to down-regulated). Clicking the column header twice sorts the protein display in order of increasing relative abundance. (down-regulated to up-regulated). Clicking the column header a third time returns the default display.

Thresholding

The list of proteins appearing in the Quantitative Samples Table is determined by the thresholds applied in Scaffold. To adjust the length of the proteins list the user has to navigate back to the Scaffold window and release or tighten the thresholds. To return to

-

Scaffold the user may ether click on the Scaffold icon appearing in the Toolbar found in the Q+ Quantitation Module main window or select the menu command ***File > Exit Back to Scaffold***.

Default sorting of columns

When the Quantitative Samples View first opens, all the protein groups and, if the clustering option was selected in Scaffold, all the clusters that meet Scaffold thresholds appear in the Samples List.

If the clustering option was not selected in Scaffold, the displayed protein groups are first sorted based on a protein probability of 50% (Scaffold's calculated probability of a correct protein identification), and if any proteins have the same probability, then the proteins are sorted alphabetically based on their accession numbers.

To confirm the sort order of the proteins, click the Scaffold icon *appearing in the Toolbar to return to the classic Samples View in Scaffold. Then set the Min Protein value to 50% and the Min Peptide to 0% and select Number of Assigned Spectra for the Display Option.*

Proteins List

Scaffold uses two levels of hierarchy to aggregate the proteins in the Proteins List: Proteins groups and Proteins clusters. The Scaffold User's Manual provides details about this type of aggregation and how groups and clusters are formed.

Clusters are displayed in the Proteins List only if the clustering option was selected in Scaffold before launching the Q+ Quantitation module. If a row in the Proteins List includes a protein cluster, a small square appears before the row number containing either a + or a - sign to indicate whether the cluster is expanded or collapsed. An option is available in the right click context menu to show clusters either all collapsed or all expanded.

The following tools are also available to comfortably examine the protein list:

- Representation of Quantitative Categories
- Tagging Proteins of Interest: The Star Function
- Hidden Proteins

Representation of Quantitative Categories

Quantitative Categories are displayed as colored stripes above the corresponding Quant Samples columns. The user assigns colors to quantitative categories when setting up the experiment through the Experimental Design Wizard.

Tagging Proteins of Interest: The Star Function

The user can mark a protein of special interest by clicking the Star icon ☆ in the **Starred?** column cell for that protein. To select an orange, blue or orange-plus-blue star the user can either click multiple times on the star icon or choose the desired color option by activating the star icon right click context menu.

-

A combination of different stars allows the creation of four different sets of proteins of interest. The user can then bring these protein groups to the top of the display by clicking the **Starred?** column header. Clicking the column header twice more returns the Protein List to the default order.

Groups of selected proteins can be starred together by using the star option available in the right click context menu

Hidden Proteins

The user can hide any unwanted proteins and /or contaminants from the Proteins List display by clearing the Visible? check box for that entry.

For example, to eliminate Trypsin products from the view, the user can search for all the proteins that contain “Trypsin” in their names, and then clear the Visible? check box for all the proteins that meet this search criteria. Only those proteins that do not have “Trypsin” in their names are displayed.

To display the proteins that are hidden go to the View menu and toggle the menu entry Show Hidden Proteins

Relative Abundance Legend

The Relative Abundance Legend, at the top of the Protein Name column, defines the color coding for the Log_2 Fold Changes in label ratios. Proteins that show a decrease in quantitation when compared to the reference sample are color-coded in varying shades of red. Proteins that show an increase in quantitation when compared to the reference sample are color-coded in varying shades of green.

For example, if you are carrying out this procedure using the sample iTRAQ data provided by Proteome Software, note the decrease in protein detection in Quant Sample 2 and Quant Sample 4 for the protein Apolipophorins.

Quantitative value tags

When selecting Display Options that includes the Log_2 Fold Change of a quantity, three types of tags might appear in the samples table’s cells:

- **Value Missing**-- The “numerator” of the Fold change is marked as missing and the log of ratio goes to infinity. This case might happen when a protein group has not been detected in a particular Quant Sample.
- **No Values** -- Both the numerator and the reference are marked as missing values.
- **Reference Missing**-- The reference of the ratio is marked as a missing value.

Another tag might appear in the following circumstances is:

-

- **No Data** -- It appears for the "Log_2 Fold Change" and "Fold Change Ratio" display options when the selected protein has no exclusive quantifiable peptides assigned to it.

Quantitative Values

The samples table li

In the Samples Table the different Quant Samples are represented as columns and may show quantitative values for each protein in the list according to the selected Display Options.

The quantitative values displayed in the Samples Table are represented as columns

For iTRAQ- or TMT-labeled data, the different reporter ions (or channels) are identified by default as Quant 1, Quant 2, Quant 3, and so on. For isotopically labeled data, the three MS1 ions (or channels) are identified as Light, Medium, and Heavy. The normalized intensity values that are displayed for a protein in each of the quantitative categories are based on a roll up of the data. A weighted median of the reporter ion or MS1 ion intensity values for all the spectra that identify a specific peptide in a quantitative category is calculated. A Kernel Density Average or mean of the Log_2 intensity values for all the peptides in this same quantitative sample category is then calculated and assigned as the overall intensity for the protein in the category.

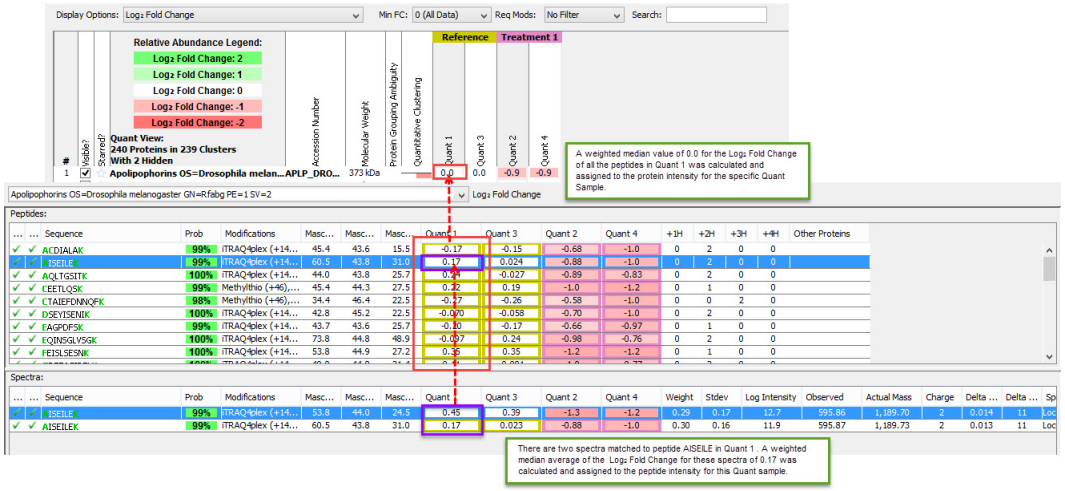
For example, if you are carrying out this procedure using the sample iTRAQ data provided by Proteome Software, then do the following:

1. Make sure that the first protein in the Proteins table, Apoliphorins OS, is selected.
2. Click Proteins to open the Proteins View.
3. In the Peptides pane, select the AISEILEK peptide.
4. Refer to Figure 5-2 below.

The Proteins View is discussed in Chapter 6, "The Scaffold Q+ Quantitative Proteins View provides an overview of the peptides quantitative detection for a selected protein group.," on page 84.

-

Figure 5-2: Example of data roll up to assign a protein intensity in a quantitative sample category

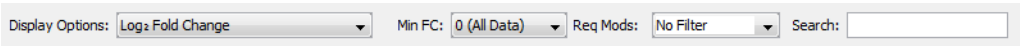


-

Display bar

Through the Display bar the user can vary the type of Quantitative values shown in the Quantitative Samples Table for the various Quant Samples for each protein.

Figure 5-3: Scaffold Q+ Display pane



The bar also contains filtering options for limiting the display to only those proteins that meet specific criteria.

- "Display Options" on page 80.
- "Min FC" on page 81.
- "Req Mods" on page 81.
- "Search" on page 81.
- "Filters" on page 82

Display Options

Specifies the value that is displayed for each protein in every Quant Sample in the Quantitative Samples table.

- **Protein Identification Probability** - Displays Scaffold's calculated probability (a percentage) that the protein identification is correct.
- **Quantitated Spectrum Count** - Number of spectra used for the quantitation.
- **Log_2 Fold Change** - The default value. Displays the Log base 2 (Log_2) of the Fold Change which is defined as the difference of the Log_2 Normalized Intensity for a quantitation value for a Quant sample in a Quantitative Category minus the Log_2 Normalized Intensity for a quantitation values for a Quant sample in the Reference category. See "Log2 Fold Change" on page 56.
- **Fold Change Ratio**-The ratio of a reporter ion intensity or an MS1 ion intensity or a precursor intensity to a reference ion intensity for a Quant Sample.

If any of the following values are displayed for fold change—Missing Values, No Values, or Reference Missing—see "Treatment of Missing Data by the Q+ Quantitation Module" on page 120.

- **Log_2 Normalized Intensity**- The normalized Log_2 of the intensities of all the Quant Samples.
- **Coefficient of Variation** - The normalized measure of the dispersion of the distribution of the intensities for all the Quant Samples. Note that this value is reported only when technical replicates are included in the experiment.

-

Min FC

Proteins that do *not* meet the minimum required Log_2 fold change threshold (the Minimum Fold Change filter) in at least one Quant Sample are hidden from view in the Quantitative Sample Table. Six threshold options are available, with 0 (All Data) being the default value.

- 0 (All Data)
- +/- 0.3 (1.2 Fold)
- +/-0.6 (1.5 Fold)
- +/- 1 (2 Fold)
- +/- 2 (4 Fold)
- +/- 3 (8 Fold)

These thresholds represent a departure from zero, and therefore, can be positive or negative values.

Req Mods

The Required Modifications filter lists all the post-translational modifications (PTMs) the user can select during the search phase of data processing. Selecting a modification on the dropdown list filters the display to only those proteins, peptides, and spectra that contain the selected modification.

- **No Filter** - No filtering is applied. All proteins, peptides, and spectra that meet all other display and filtering options are displayed.
- **Unmodified Only** - Displays only those proteins, peptides, and spectra that do not have any associated PTMs. If the user selects this option for iTRAQ- and TMT- labeled data, or SILAC data or Precursor Intensity data, proteins that were not labeled correctly, but were still identified in the search, are displayed.
- **Variable Modifications** - Displays only those proteins, peptides, and spectra that were identified as having the selected modification.
- **iTRAQ** or **TMT** - Labeling modifications for quantitation of proteins and peptides. This option is available only when the user has loaded iTRAQ- or TMT- labeled data. For example, if the user loaded iTRAQ (4-plex) data, then the available option is iTRAQ 4plex. If no proteins are displayed when this option is selected, then most likely, a labeling error occurred.

Search

The Search field accepts regular expressions and filters the results based on accession number or protein name. Only those proteins that meet all the search criteria are displayed. The search is limited to the exact order of the characters in the string, but the string is not case-sensitive and it can appear anywhere in the search results. For example, a search string

-

of **ATP** returns both **ATP** synthase and calcium-transporting **ATP**ase. In another example, a search string of **sodium**|**transport** returns all values that have sodium and/or transport in the protein name - **Sodium**/potassium **transport**ing... and Calcium-**transport**ing ATPase sar..., and so on.

Filters

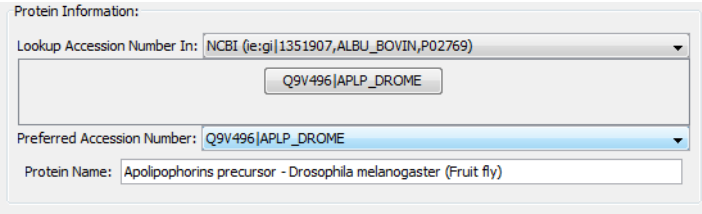
The Filter control provides access to the “Configure Advanced Protein Filters” dialog which also appears in Scaffold. This allows the user to filter on a variety of criteria, such as star color, GO terms and taxonomy.

-

Protein Information pane

The Protein Information pane is displayed at the bottom of the Samples View. After you select a protein in the Proteins Table, Scaffold retrieves and displays the information for the selected protein based on the database that you searched to identify the proteins in your biosamples.

Figure 5-4: Protein Information pane



- **Lookup Accession Number In**: The user can look up the protein accession number for a selected protein in an online protein database such as SwissProt or NCBI. The database can be selected from the dropdown list and then click the Preferred Accession Number option in the field beneath this dropdown list to open an Internet browser window that is linked directly to the accession number in the selected database. If more than one protein accession number is associated with the selected protein, then these values are displayed in the field beneath the Lookup Accession Number In dropdown list.

-

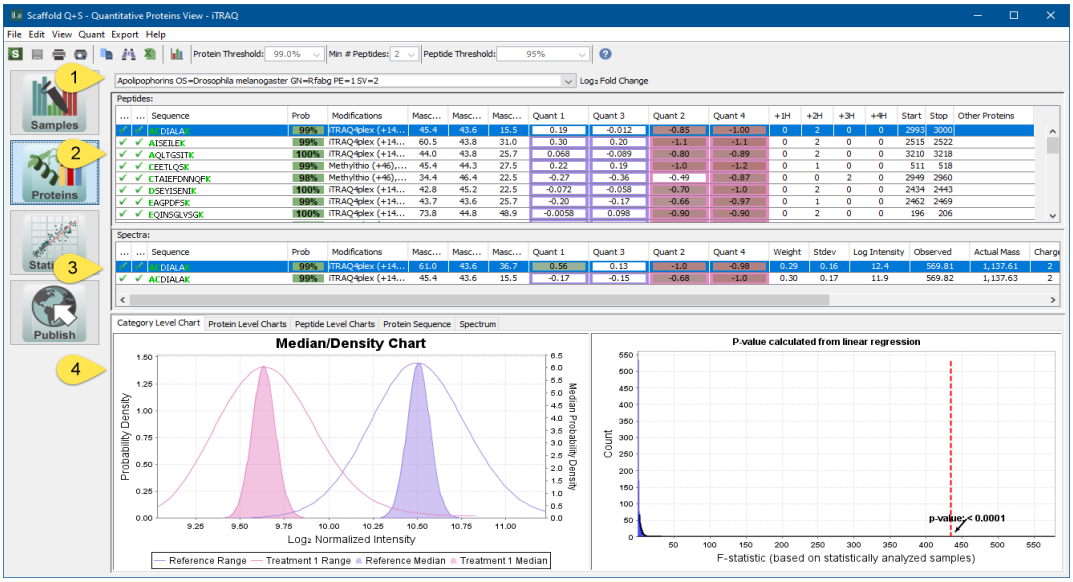
Chapter 6
Quantitative Proteins View

The Scaffold Q+ Quantitative Proteins View provides an overview of the peptides quantitative detection for a selected protein group.

The user can reach this view either by clicking the Proteins button included in the Navigation pane located on the left side of the main Q+ window or from the Samples View by selecting a protein listed in the Samples Table and then double clicking over it. The user can always select a different protein in the Quantitative Proteins View from the dropdown list that is displayed above the Peptides pane, see Figure 6-1, tag 1.

If you are carrying out this procedure using the sample iTRAQ data provided by Proteome Software, then select and double click over the first protein that is listed in the Proteins Table, Apolipophorins OS-Drosophila melanogaster.

Figure 6-1: Quantitative Proteins View



The Quantitative Proteins View displays information about the selected protein in three panes. The first two panes are organized into tables whose features and tools are more extensively described in section "Display pane" on page 67.

- The Peptides pane (2) displays all the *exclusive* peptides that were detected for the protein.
- The Spectra pane (3) displays all the spectra that matched to a selected peptide in the Peptides pane.

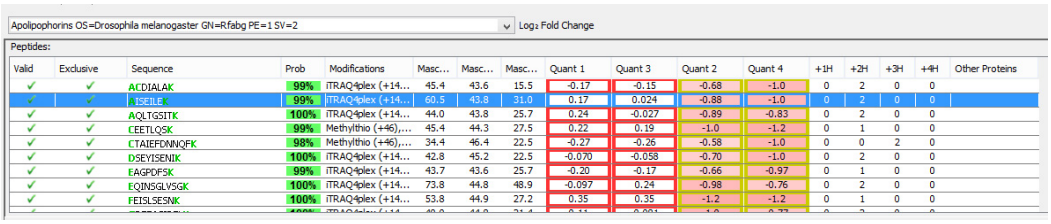
-

- The Visualization pane (4) contains five tabs—the Category Level Chart tab, the Protein Level Charts tab, the Peptide Level Charts tab, the Protein Sequence tab, and the Spectrum tab.

-

Peptides pane

Figure 6-2: Quantitative Proteins View, Peptides pane



The Peptides pane (top pane) contains a table that by default lists all the *exclusive* peptides that were detected for the selected protein and that met the threshold requirements specified in the Scaffold Samples view before launching the Q+ Quantitative module.

The columns preceding the Quant Samples values provide information about the identification of the peptide like its probability, its modifications and the assigned search engine scores.

The quantitative value displayed for each peptide and for every Quant Sample, depends on the Display Options selected in the Quantitative Samples View as specified in Table 6-1.

Table 6-1:

Display Options	Quant Value Displayed
Protein Identification Probability	Log $_2$ Fold Change
Quantitated Spectrum Count	Log $_2$ Fold Change
Log $_2$ Fold Change	Log $_2$ Fold Change
Fold Change Ratio	Fold Change Ratio
Log $_2$ Normalized Intensity	Log $_2$ Normalized Intensity
Coefficient of Variation (Appearing only if technical replicates are specified)	Log $_2$ Normalized Intensity

To highlight the different Quantitative Categories, each cell containing a Quant Sample value for the peptide is enclosed in a rectangle of the color assigned, during the set up of the experiment, to its Quantitative Category.

The Quant Sample values are directly related to the corresponding values appearing in the Spectra pane as explained in the "Quantitative Values" section on page 78. The background of the cells are colored according to the Relative Abundance Legend appearing in the Quantitative Samples View.

If you are carrying out this procedure using the sample iTRAQ data provided by Proteome Software, then select the peptide AISEILEK (second from the top) and review the Log$_2$ Fold Change scores of 0.24, -0.027, -0.89, and -0.83.

-

The Charge State columns (+1H, +2H, +3H, and +4H) display the number of spectra that were matched to the peptide depending on its charge state.

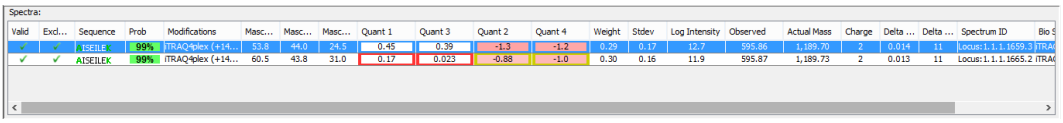
Information about each spectra that was matched to a peptide is displayed in the Spectra pane.

If the user changes the default quantitative settings to include non-exclusive peptides in the analysis, then these peptides are displayed in the Peptides pane, and the accession numbers of the other proteins in which a non-exclusive peptide can be found are displayed in the Other Proteins column for the peptide.

-

Spectra pane

Figure 6-3: Quantitative Proteins View, Spectra pane



The Spectra pane (2) contains a table that displays information about each of the spectra that were matched to a peptide that is selected in the Peptides pane and the quantitative values at the spectrum level for each Quant Sample in each Quantitative category. The quantitative value displayed for each spectrum and for every Quant Sample, depends on the Display Options selected in the Quantitative Samples View as specified in Table 6-1.

To highlight the different Quantitative Categories, each cell containing a Quant Sample value for the spectrum is enclosed in a rectangle of the color assigned, during the set up of the experiment, to its Quantitative Category.

From left to the right, the columns in the Spectra pane for a selected peptide are as follows:

- **Valid** - If the peptide contributes to the probability calculation for the selected protein, then a green checkmark is displayed in this column; otherwise, the column is blank.
- **Exclusive** - If the peptide is exclusive to the selected protein, then a checkmark is displayed in this column; otherwise, the column is blank.
- **Sequence** - The peptide sequence. Amino acids that have PTMs are displayed in green.
- **Prob** - The Peptide Identification Probability, which is Scaffold's calculated probability (a percentage) that the peptide identification is correct.
- **Other peptide data** - Some of the columns that are displayed are identical for iTRAQ- or TMT- labeled data and isotopically labeled data, while others are specific to the label used. To view what type of data is displayed in a column, rest your mouse pointer on the column.

If multiple spectra are assigned to the same peptide, then, by default, Scaffold Q+ calculates a weighted median value for the fold change for the peptide. Moreover, if multiple spectra are assigned to the same peptide and their profiles are consistent across the Quant Samples, then there is increased confidence in the detected change. The opposite, however, is also true. If the profiles of the spectra are not consistent across the Quant Samples, then there is decreased confidence in the detected change.

If you are carrying out this procedure using the sample iTRAQ data provided by Proteome Software, then note that two spectra are assigned to the peptide AISEILEK.

You can use the tri-state column sorting feature and sort the display by clicking on any column header. For example, to sort the peptide data based on increasing log intensity values, click the Log Intensity column header once. To sort the spectra based on decreasing

log intensity values, click the Log Intensity column header twice. To return to the default display, click the Log Intensity column header a third time.

-

Visualization pane

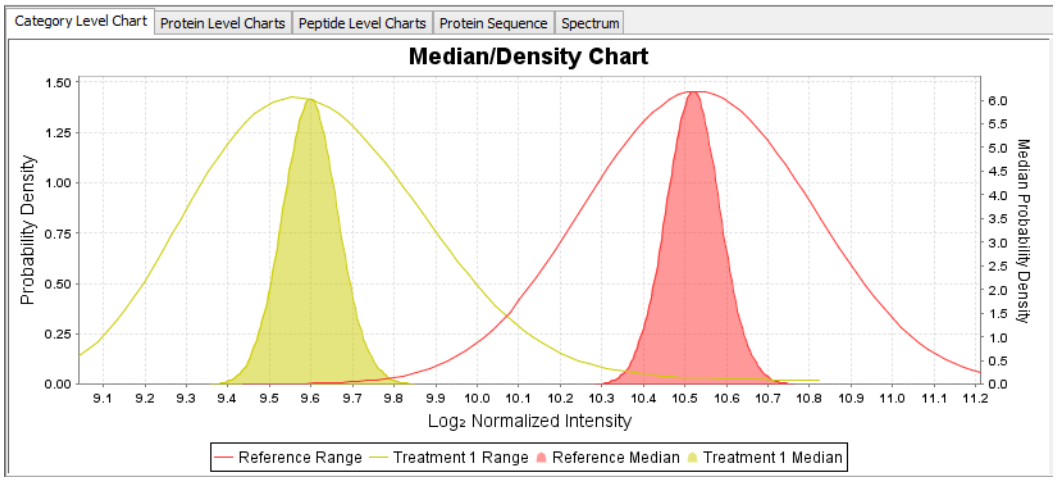
The lower half of the Proteins View or Visualization pane, contains five tabs. Each tab provides a different graphical representation of the quantitative data collected for the selected protein.

- The Category Level Chart tab, the Protein Level Charts tab, and the Peptide Level Charts tab display different charts that illustrate the various quantitative calculations carried out for the peptides and proteins.
- The Protein Sequence tab displays the sequence of the selected protein.
- The Spectrum tab displays the spectrum that is selected in the Spectra pane.

Category Level Chart tab

The Category Level Chart tab displays two different graphs—a Category Density Function graph (thin line) and a Sampling Distribution graph (solid graph)—to judge how much the proteins are differentially expressed in each Quantitative Category. Both graphs display the color-coding that was specified in the "Experimental Design Wizard" on page 45.

Figure 6-4: Density/Sampling Distribution graph

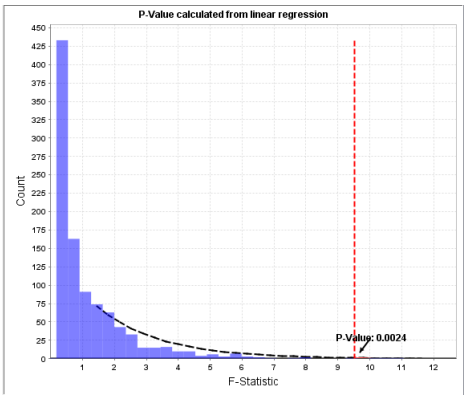


- **Category Density Function (Density Probability)**—The Category Density Function graph reflects the distribution of the Log_2 normalized intensities for a specific protein in each Quantitative Category. The Q+ Quantitation Module estimates the distribution using a Kernel Density function based on a finite number of measurements. These measurements come from all the spectra in all samples that belong to the protein in a specific category. The function also allows for the combining of data from different samples into one overall distribution.
- **Sampling Distribution (Median Probability)**—Depending on the selected Calculation Level, the Sampling Distribution graph represents the distribution of sample averages, medians, or means of randomly drawn samples from the Category Density function. The

degree of separation of the averages sampling distributions is an indication of the statistical significance of the differences in the protein levels in each of the Quantitative Categories. These distributions also help in clarifying the differences patterns that an ANOVA test might detect. If the median/mean sampling distributions of two categories are clearly separated, (as shown in the example in this manual), then the greater the likelihood that there is a meaningful difference between the protein levels in these categories.

The tab also displays a third graph—the Random Permutation Test graph—which is a histogram that shows the F-statistic distribution that was calculated using a randomized permutation procedure.

Figure 6-5: Randomized Permutation Test graph



For information about how Scaffold Q+ calculates randomized permutation tests, see https://proteomesoftware.com/statistics/permutation-test

Both the Density/Sampling Distribution graph and the Randomized Permutation Test graph have the following features in common:

- You can hold down the left mouse button and draw a box from the upper left hand corner of the graph towards the lower right hand corner. A box is formed around the area that being reduced for viewing. A single click of the mouse returns the zoom out magnification to 100%.
- You can right-click on either graph to open a context menu that has the following options:
  - **Copy WMF/EMF** - Copy a vector-based image to the clipboard which you can then paste into a third party tool such as Microsoft PowerPoint for easy editing and manipulation.
  - **Copy Publication Sized JPEG**—Copies the currently displayed spectrum in a JPEG format to the clipboard. You can then paste this copied image, which is appropriately pre-sized for publication, into a third-party application such as Microsoft Word.

-

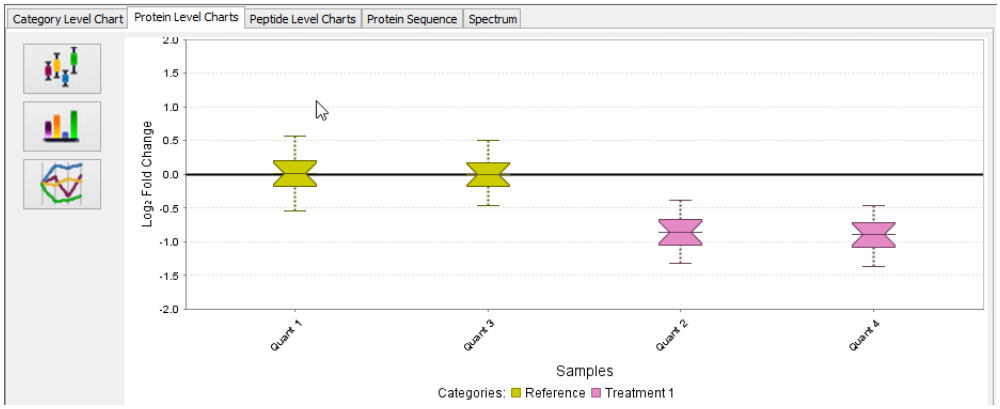
- **Save JPEG Image**—Saves the currently displayed spectrum in a JPEG format and opens the Write JPEG picture file dialog box in which you can specify the name and directory for this saved JPEG file.
- **Save As**—Opens a dialog box in which you can specify the name, format, and directory for saving the displayed spectrum. Options are also available for working with and/or modifying the specific file format chosen.
- **Print**—Opens the Print dialog box in which you can specify the options for printing (printer, number of copies, and so on) the currently displayed spectrum.
- **Zoom Out**—Returns the zoom out magnification to 100%.

Protein Level Charts tab

The Protein Level Charts tab displays a graph that show the relative quantitation information for the selected protein in each of the Quantitative Categories. Three formats are available for this graph—a Box plot, a Bar chart, and a Trend Line. All three formats display the color-coding that was specified for the Quantitative Category in the "Experimental Design Wizard" on page 45.

- **Box plot**—The default plot. The Box plot displays the relative median value and range for the Log_2 Fold Change. Place your cursor on any box plot to display information about the median and interquartile range for the corresponding quantitative sample category.

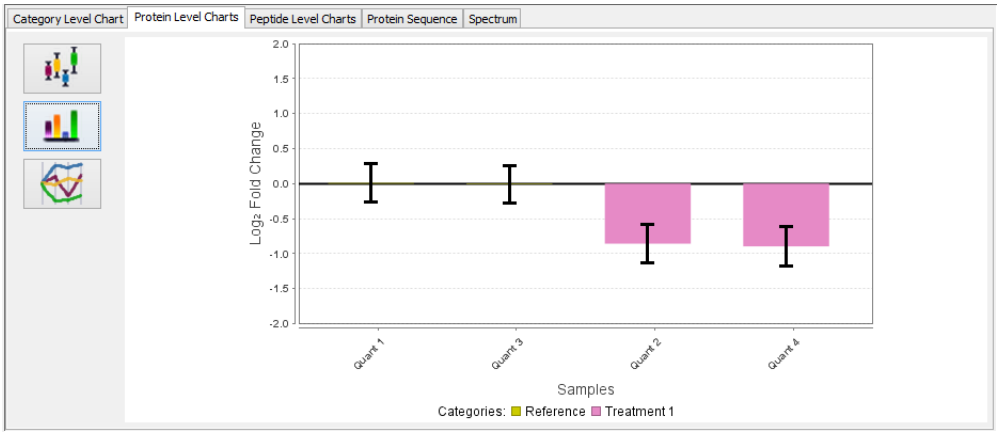
Figure 6-6: Protein Level Charts tab, Box plot



- **Bar chart**—The Bar chart displays the relative median value and range for the Log_2 Fold Change. Place your cursor on any bar to display information about the median and range for the corresponding quantitative sample category.

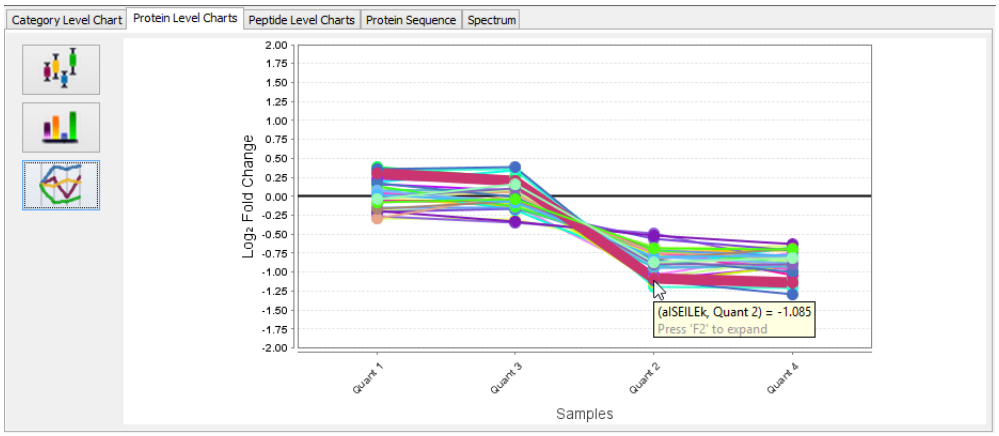
-

Figure 6-7: Protein Level Charts tab, Bar chart



- **Trend Line**—The Trend Line displays the relative Log_2 Fold Change for each peptide in the selected protein. (Accordingly, there is one Trend Line displayed per peptide.) Place your cursor on a data point in a Trend Line to highlight a specific peptide. The Trend Line is displayed in bold, and the corresponding peptide is automatically selected in the Peptides pane.

Figure 6-8: Protein Level Charts tab, Trend Line



-

All three formats have the following features in common:

- You can hold down the left mouse button and draw a box from the upper left hand corner of the graph towards the lower right hand corner. A box is formed around the area that being reduced for viewing. A single click of the mouse returns the zoom out magnification to 100%.
- You can right-click on either graph to open a context menu that has the following options:
  - **Copy WMF/EMF** - Copy a vector-based image to the clipboard which you can then paste into a third party tool such as Microsoft PowerPoint for easy editing and manipulation.
  - **Copy Publication Sized JPEG**—Copies the currently displayed spectrum in a JPEG format to the clipboard. You can then paste this copied image, which is appropriately pre-sized for publication, into a third-party application such as Microsoft Word.
  - **Save JPEG Image**—Saves the currently displayed spectrum in a JPEG format and opens the Write JPEG picture file dialog box in which you can specify the name and directory for this saved JPEG file.
  - **Save As**—Opens a dialog box in which you can specify the name, format, and directory for saving the displayed spectrum. Options are also available for working with and/or modifying the specific file format chosen.
  - **Print**—Opens the Print dialog box in which you can specify the options for printing (printer, number of copies, and so on) the currently displayed spectrum.
  - **Zoom Out**—Returns the zoom out magnification to 100%.

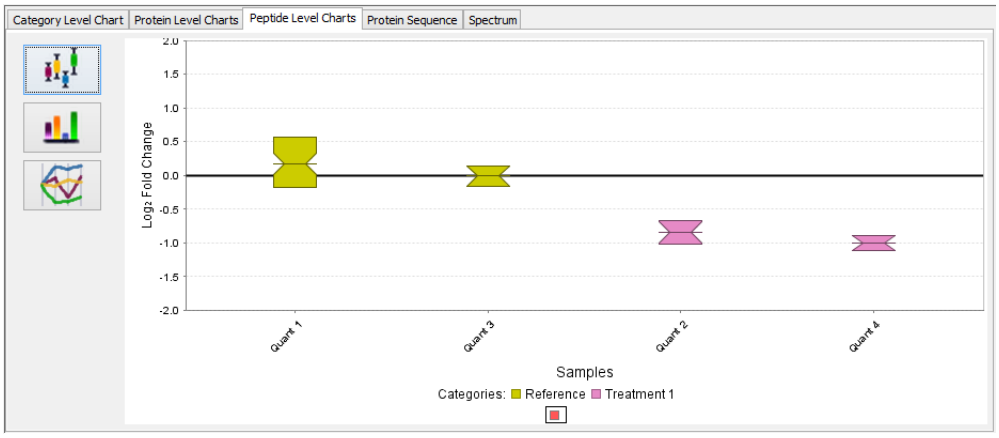
Peptide Level Charts tab

The Peptide Level Charts tab displays a graph that show the relative quantitation information for the selected peptide in each of the Quantitative Categories. Three formats are available for this graph—a Box plot, a Bar chart, and a Trend Line. All three formats display the color-coding that was specified for the Quantitative Category in the Experimental Design Wizard.

- **Box plot**—The default plot. The Box plot displays the relative median value and range for the Log_2 Fold Change. Place your cursor on any box plot to display information about the median and interquartile range for the corresponding quantitative sample category. See Figure 6-9.

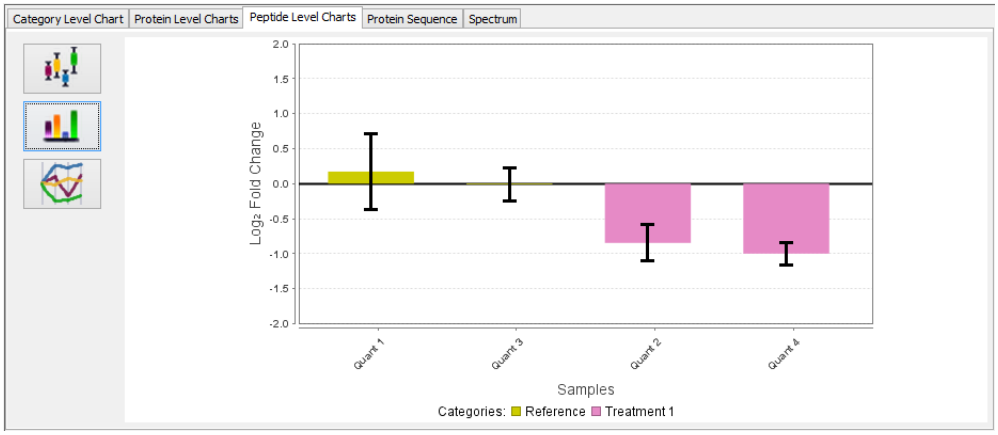
-

Figure 6-9: Peptide Level Charts tab, Box plot



- **Bar chart**—The Bar plot displays the relative median value and range for the Log_2 Fold Change. Place your cursor on any bar to display information about the median and range for the corresponding quantitative sample category.

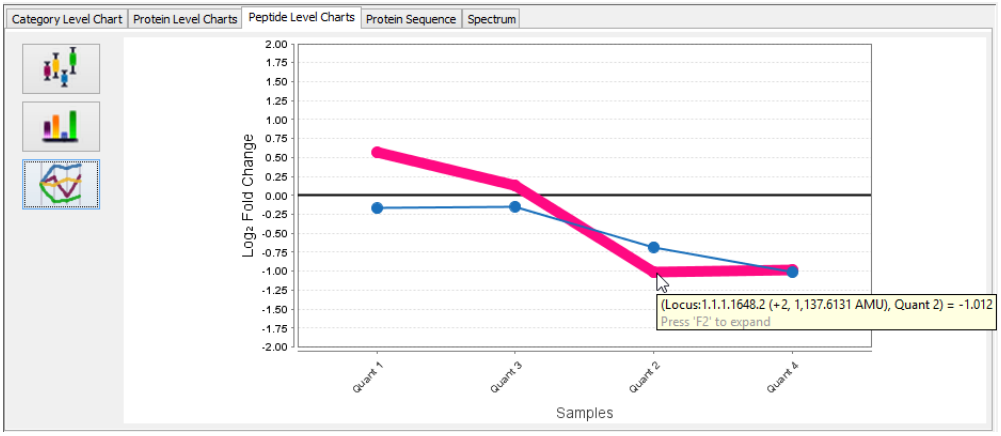
Figure 6-10: Peptide Level Charts tab, Bar chart



- **Trend Line**—The Trend Line graph displays the relative Log_2 Fold Change for each spectra for the selected peptide. (Accordingly, there is one Trend Line graph displayed per peptide spectrum.) Place your cursor on a data point in a Trend Line graph to highlight a specific spectrum. The Trend Line graph is displayed in bold, and the corresponding spectrum is automatically selected in the Spectrum pane. See Figure 6-11.

-

Figure 6-11: Peptide Level Charts tab, Trend Line



All three formats have the following features in common:

- You can hold down the left mouse button and draw a box from the upper left hand corner of the graph towards the lower right hand corner. A box is formed around the area that being reduced for viewing. A single click of the mouse returns the zoom out magnification to 100%.
- You can right-click on either graph to open a context menu that has the following options:
  - **Copy WMF/EMF** - Copy a vector-based image to the clipboard which you can then paste into a third party tool such as Microsoft PowerPoint for easy editing and manipulation.
  - **Copy Publication Sized JPEG**—Copies the currently displayed spectrum in a JPEG format to the clipboard. You can then paste this copied image, which is appropriately pre-sized for publication, into a third-party application such as Microsoft Word.
  - **Save JPEG Image**—Saves the currently displayed spectrum in a JPEG format and opens the Write JPEG picture file dialog box in which you can specify the name and directory for this saved JPEG file.
  - **Save As**—Opens a dialog box in which you can specify the name, format, and directory for saving the displayed spectrum. Options are also available for working with and/or modifying the specific file format chosen.
  - **Print**—Opens the Print dialog box in which you can specify the options for printing (printer, number of copies, and so on) the currently displayed spectrum.
  - **Zoom Out**—Returns the zoom out magnification to 100%.

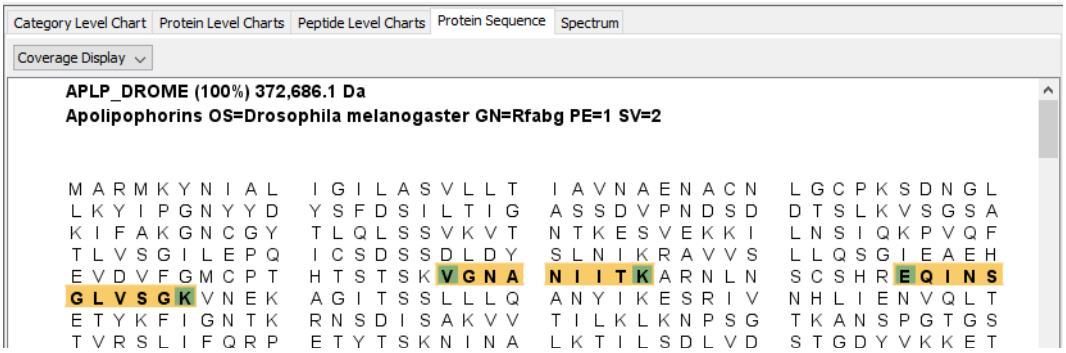
Protein Sequence tab

By default, when the Protein Sequence tab first opens, the full protein coverage map is displayed. The protein sequence and molecular weight might also be displayed. The sequence is derived from the fasta database that was loaded into Scaffold with the

quantitative sample files. Peptides that were identified in user's dataset are highlighted in yellow. Amino acids with PTMs are highlighted in green.

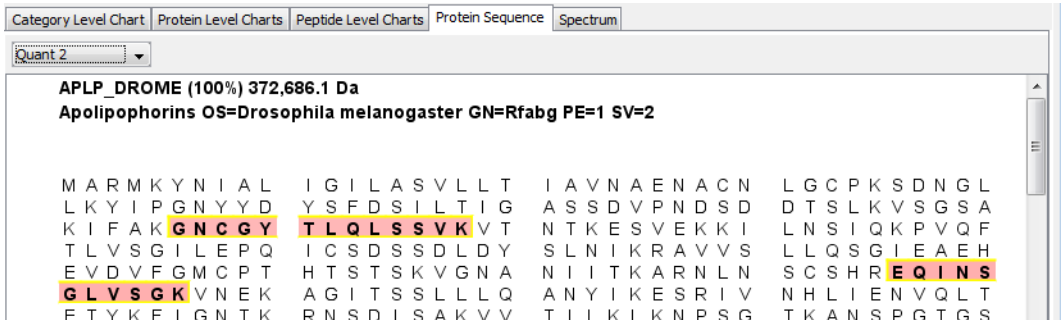
Remember, if this fasta database is not identical to the external protein database, including the version, that you used for searching your experimental data, then the protein sequence and molecular weight might not be available for display.

Figure 6-12: Protein Sequence tab, full coverage map



On the Coverage Display dropdown list, you can select a specific Quant Sample to display the coverage map for the sample. The identified peptides are outlined in yellow. The identified peptides that show a decrease in quantitation when compared to the reference sample are highlighted in red while those that show an increase in quantitation when compared to the reference sample are highlighted in green.

Figure 6-13: Protein Sequence tab, quantitative sample coverage map



The user can right-click on either graph to open a context menu that has the following options:

- **Copy WMF/EMF** - Copy a vector-based image to the clipboard which you can then paste into a third party tool such as Microsoft PowerPoint for easy editing and manipulation.
- **Save As**—Opens a dialog box in which you can specify the name, format, and directory for saving the displayed spectrum. Options are also available for working with and/or modifying the specific file format chosen.
- **Print**—Opens the Print dialog box in which you can specify the options for printing (printer, number of copies, and so on) the currently displayed spectrum.

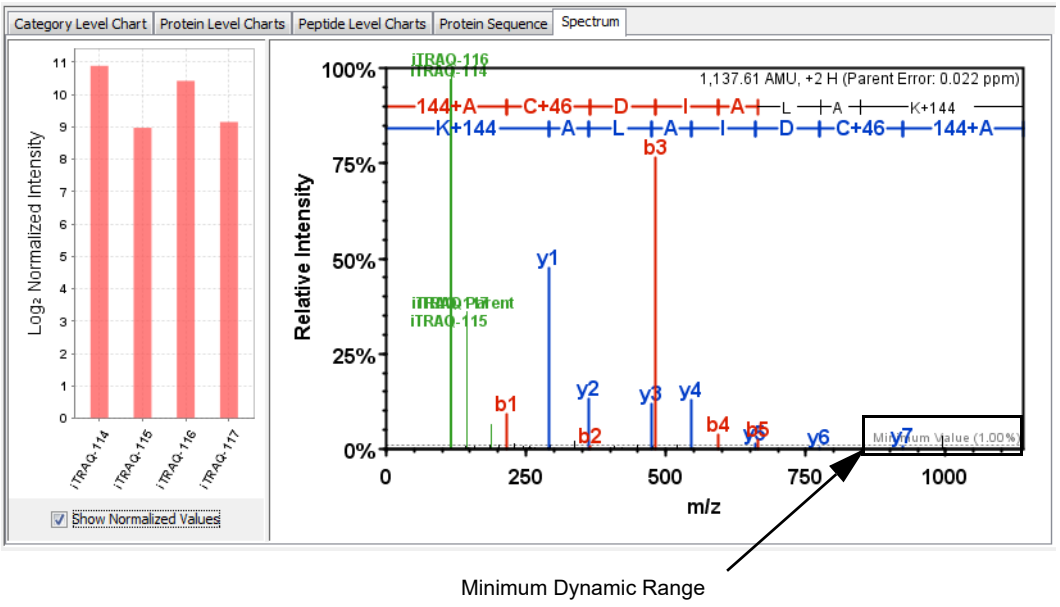
-

- **Copy Protein Sequence**—Copies the currently displayed protein sequence spectrum to the Clipboard which you can then paste into a third party tool such as Microsoft Word.
- **Use Amino Acid Finder**—Selected by default. Toggles the activiation of an amino acid tooltip. If on, when placing your cursor on the protein sequence, a tooltip opens that dipslays the full amino acid sequences for the underlying peptides, their IDs and probability.
- **Show Fixed Modifications**—Select this option to highlight amino acids with fixed modifications in green.
- **BLAST Protein Sequence**—Select this option to automatically open an Internet browser session and display the Standard Protein BLAST page (blastp) for the selected protein.

Spectrum tab

The Spectrum tab displays the spectrum that is selected in the Spectra pane. The dashed gray horizontal line that is labeled Minimum Dynamic Range is the *cutoff value* for the dynamic range of the spectrum. (The user can specify this value on the Minimum Dynamic Range tab of the Quantification Setup dialog box. See "Minimum Dynamic Range tab" on page 118.)

Figure 6-14: Spectrum tab



If you are carrying out this procedure using the sample iTRAQ data provided by Proteome Software, then in the Peptides pane, select the peptide CTAIEFDNNQFK to view its associated spectrum on the Spectrum tab.

The Spectrum tab is interactive:

- Click anywhere on the spectrum to display the M/z value for the position.

-

- Click and hold the left mouse button anywhere on the spectrum and then drag your mouse pointer to any position in the spectrum of your choosing. As you drag your mouse pointer, the Start and Stop M/z values for the segment are displayed as well as the length for the segment. Release the button to zoom in on the selected region. A single click of the mouse returns the zoom out magnification to 100%.
- Right-click anywhere on the spectrum to open a context menu that has the following menu options:
  - **Copy WMF/EMF** - Copy a vector-based image to the clipboard which you can then paste into a third party tool such as Microsoft PowerPoint for easy editing and manipulation.
  - **Save As**—Opens a dialog box in which you can specify the name, format, and directory for saving the displayed spectrum. Options are also available for working with and/or modifying the specific file format chosen.
  - **Print**—Opens the Print dialog box in which you can specify the options for printing (printer, number of copies, and so on) the currently displayed spectrum.
  - **Copy Peaklist**—Copies the peak list for the currently displayed spectrum to the Clipboard which you can then paste into a third party tool such as Microsoft Excel.
  - **Zoom Out**—Returns the zoom out magnification to 100%.

This is identical to a single click of the left mouse button after using the Click and Drag feature.

  - **Use Peakfinder**—Selected by default. When selected, a tooltip opens that displays the ion designation, M/z value, and M/z error (in ppm) for the daughter ion that is closest to the current cursor position.
  - **Display Unknown Markers**—Selected by default. If a peak has not been assigned to any identified ion, then the peak m/z value is shown with a question mark (?).
  - **Display Parent Ions**—Displays parent ions in the spectrum. If parent ions are much more prevalent than daughter ions, then clear this option so that you can more easily examine the daughter ions.
  - **Use PPM Masses**—Selected by default. Displays the M/z error in ppm in the tooltip for a daughter ion. Clear this option to display the M/z error in AMU.
  - **BLAST Peptide Sequence**—Select this option to automatically open an Internet browser session and display the Standard Protein BLAST page (blastp) for the selected peptide.

By default, the Spectrum tab also displays a Bar chart of the normalized values of the Log_2 Intensity values of the reporter ions or MS1 ions. The user can clear the Show Normalized Values option to display the non-normalized values for the data.

-

-

Chapter 7
Quantitative Statistics View

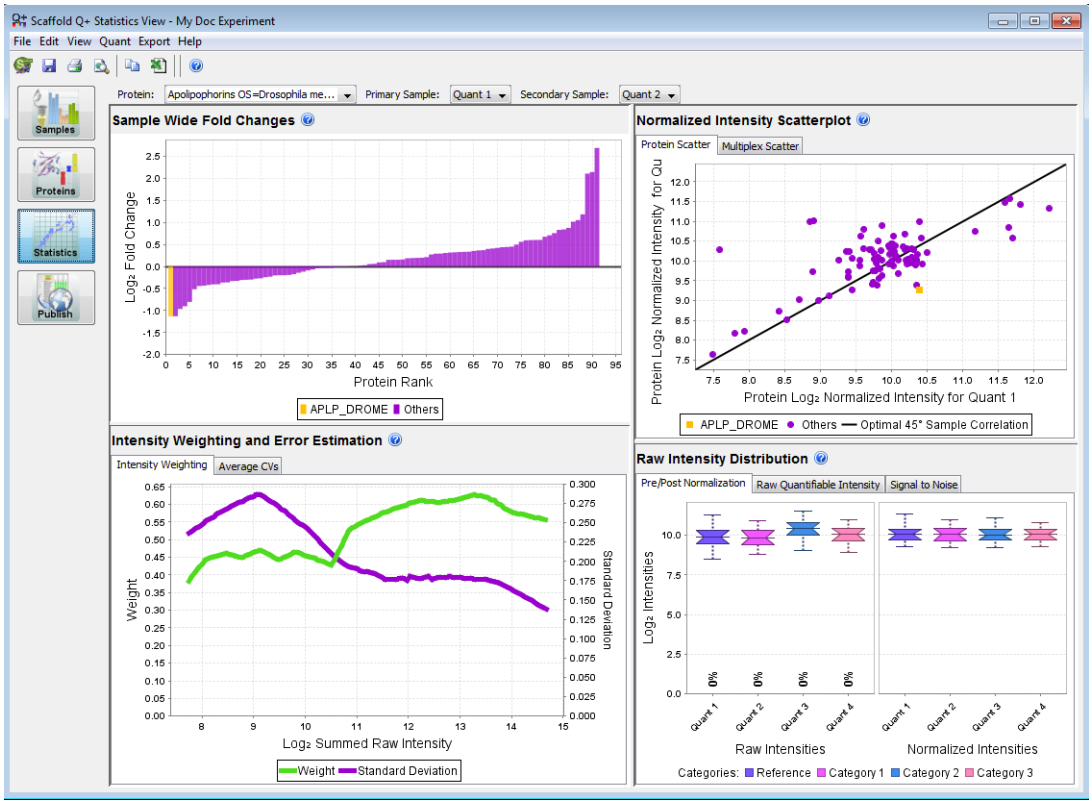
The Scaffold Q+ Statistics View demonstrates the relationship between a selected protein and the data set as a whole. This chapter details the Scaffold Q+ Statistics View, including the information that can be gleaned from this view as well as the features of the view.

Select a protein in the Proteins table in the Samples View, and then click Statistics View to open the corresponding Statistics View. The upper half of the Scaffold Q+ Statistics View—the Sample Wide Fold Changes graph and the Normalized Intensity Scatterplots—provides information about the relationship between a selected protein and the data set as a whole. The lower half of the Scaffold Q+ Statistics View—Intensity Weighting and Error Estimation graphs and Raw Intensity Distribution graphs—provides insight into the normalization calculations carried out by Scaffold Q+ for the data set as a whole.

If you are carrying out this procedure using the sample iTRAQ data provided by Proteome Software, then select the first protein that is listed in the Proteins table, Apolipophorins OS-Drosophila melanogaster.

-

Figure 7-1: Statistics View



All the graphs in the Statistics view have the following features in common:

- You can hold down the left mouse button and draw a box from the upper left hand corner of the graph towards the lower right hand corner. A box is formed around the area that being reduced for viewing. A single click of the mouse returns the zoom out magnification to 100%.
- You can right-click on any graph to open a context menu that has the following options:
  - Copy WMF/EMF - Copy a vector-based image to the clipboard which you can then paste into a third party tool such as Microsoft PowerPoint for easy editing and manipulation.
  - Save As—Opens a dialog box in which you can specify the name, format, and directory for saving the displayed spectrum. Options are also available for working with and/or modifying the specific file format chosen.
  - Print—Opens the Print dialog box in which you can specify the options for printing (printer, number of copies, and so on) the currently displayed spectrum.
  - Copy Chart Data—Available only for the Normalized Intensity Scatterplots and the Raw Intensity Distribution graphs. Copies the currently displayed chart data (x-axis

-

coordinate and y-axis coordinate for each data point) to the clipboard, which you can then paste into a third-party application such Microsoft Excel.

- Zoom Out—Returns the zoom out magnification to 100%.

See:

- “Sample Wide Fold Changes” on page 104.
- “Quantitative Scatterplots” on page 104.
- “Intensity Weighting and Error Estimation” on page 106.
- “Raw Intensity Distribution” on page 108.

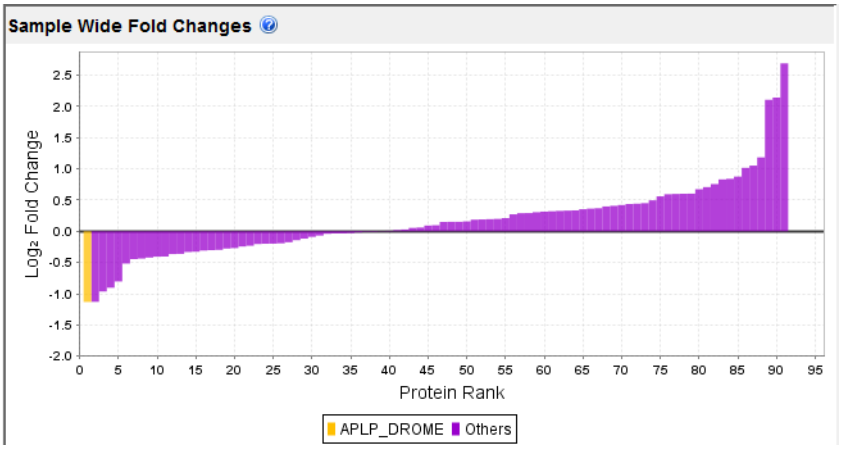
-

Sample Wide Fold Changes

The Sample Wide Fold Changes graph displays difference in the Log2 fold change for a selected protein in two quantitative samples relative to the differences in the Log2 fold change for all other proteins in the same samples.

$$\text{Log}_2 \text{ fold change (Secondary Sample)} - \text{Log}_2 \text{ fold change (Primary Sample)}$$

Figure 7-2: Sample Wide Fold Changes graph



The selected protein is displayed in yellow on the graph. the other proteins are displayed in purple. The proteins are ordered based on *increasing* fold change. If the difference in the Log2 fold change for the selected protein in the quant sample categories is close to zero, then the yellow protein indicator might not be displayed on the graph.

Quantitative Scatterplots

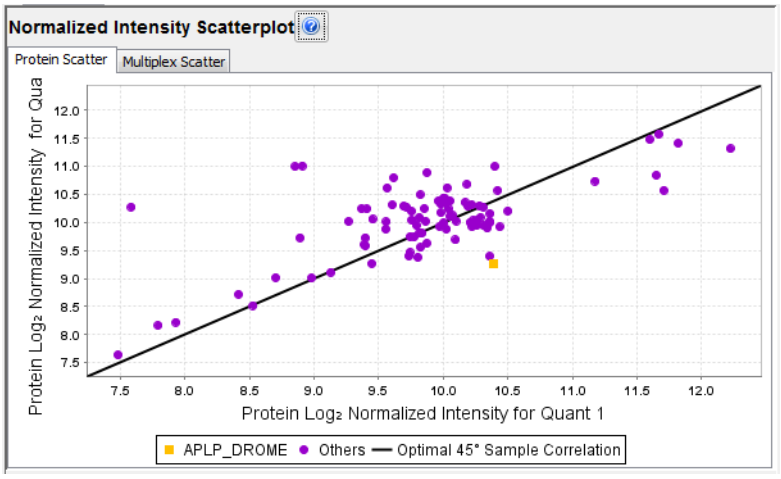
The QuantitativeScatterplot pane contains three graphs that provide information about the relationship between a selected protein and the data set as a whole.

For detailed information about normalization calculations in Scaffold Q+, see .https://www.proteomesoftware.com/docs/quant-normalization

- **Protein Scatterplot**—The Protein Scatterplot compares the Log2 normalized reporter ion or MS1 ion intensities for a selected protein (yellow) relative to the Log2 normalized reporter ion or MS1 ion intensities for all other proteins (purple) in two quantitative samples. The x axis is the primary quantitative sample value and the y axis is the secondary quantitative sample value. An optimal 45º reporter or MS1 ion intensity is overlaid on the scatterplot and it represents a 1:1 ratio between the selected quantitative samples. If the selected protein is up-regulated, then the yellow indicator falls above the optimal indicator. Likewise, if the selected protein is down-regulated, then the yellow indicator falls below the optimal indicator. See Figure 7-3 on page 105.

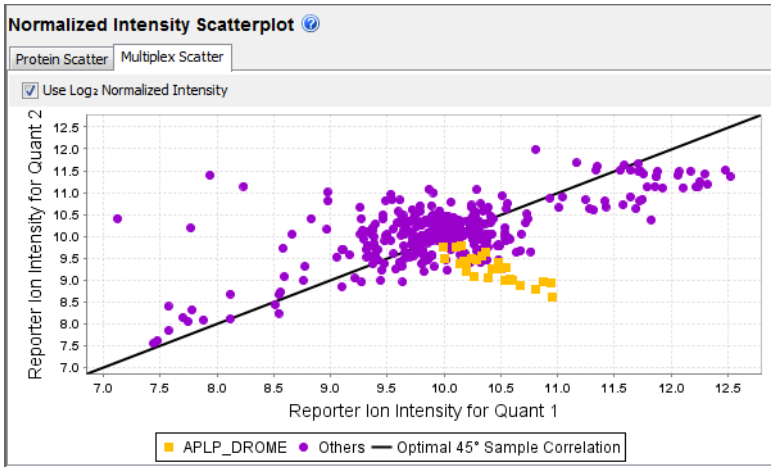
-

Figure 7-3: Normalized Intensity Scatterplot, Protein Scatter graph



- **Multiplex Scatterplot**—By default, the Multiplex Scatterplot compares the Log2 normalized reporter ion or MS1 ion intensities for all the spectra that matched to the selected protein (yellow indicators, one per spectrum) relative to the Log2 normalized reporter ion or MS1 ion intensities for all the spectra that matched to all the other proteins (purple indicators, one per spectrum) in two quantitative samples. To view normalized values, clear the Use Log2 Normalized Intensity option that is displayed at the top of the plot. The x axis is the primary quantitative sample value and the y axis is the secondary quantitative sample value. An optimal 45º reporter or MS1 ion intensity is overlaid on the scatterplot and it represents a 1:1 ratio between the selected quantitative samples. If the selected protein is up-regulated, then the yellow indicators fall above the optimal indicator. Likewise, if the selected protein is down-regulated, then the yellow indicators fall below the optimal indicator.

Figure 7-4: Normalized Intensity Scatterplot, Multiplex Scatterplot (Log2 Normalized Intensities)



-

- **Volcano Plot** - the Volcano Plot is only available when a quantitative test comparing exactly two categories of data is applied. A volcano plot is a type of scatterplot used to quickly identify changes in large datasets composed of multiple replicate samples. The graph plots the relation ship between p-value and fold change, both expressed in Log2. The Volcano Plot enables a quick visualization identification of those points that display large-magnitude changes that are also statistically significant. Plotting points in this way results in two regions of interest in the plot: those points near the top of the plot far to the left or to the right.

  Points that represent statistically significant p-values are colored green, while the remaining points are gray. Values are capped in each dimension and all points that would have fallen beyond the boundaries of the chart are plotted at the boundaries as triangles. Optionally, lines may be added to indicate the p-value level that separates the significant and insignificant p-values and the point of zero fold-change.

Intensity Weighting and Error Estimation

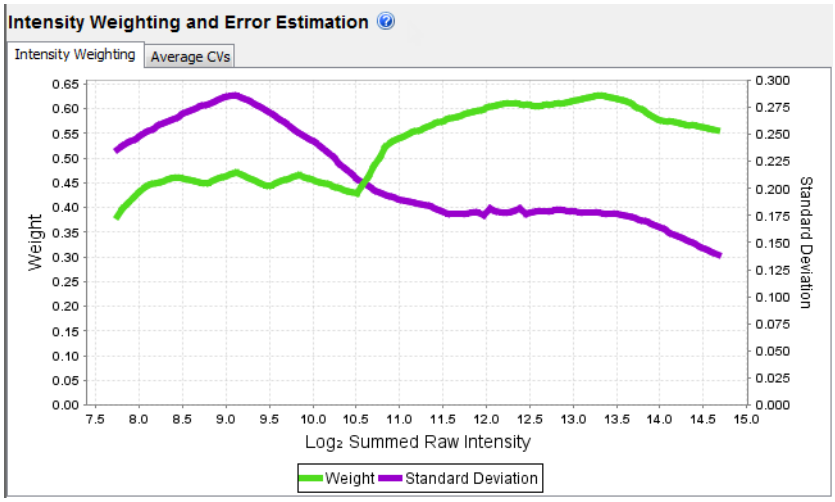
The Intensity Weighting and Error Estimation pane displays two graphs that provide insight into the normalization calculations carried out by Scaffold Q+S for the data set as a whole.

- **Intensity Weighting**—Accuracy in reporter ion and MS1 ion quantitative data is largely dependent on the intensity of the labeled peaks. Typically, measurements of higher intensity peaks tend to be more accurate than measurements of lower intensity peaks. Scaffold Q+ takes this dependency into consideration by using a weighting function in its normalization algorithm. This weighting function is based on the distribution of the deviations of the raw intensities from the median protein value for all proteins in the experiment. From the distribution, a confidence level is calculated that represents the weight. The Intensity Weighting graph shows the weight as a function of the Log2 of the *summed raw intensities* for each spectrum, where the summed raw intensity is defined as the sum of all the reporter ion intensities for iTRAQ- or TMT-labeled data or all the pairs or triplets for isotopically labeled data in the spectrum. The Intensity Weighting graph also displays the standard deviation of the Log2 of the summed raw at each intensity level. These standard deviations are calculated only after the normalization calculation is complete.

For detailed information about the calculation of the intensity weighting function in Scaffold Q+S, see page 7 of .https://www.proteomesoftware.com/docs/quant-normalization

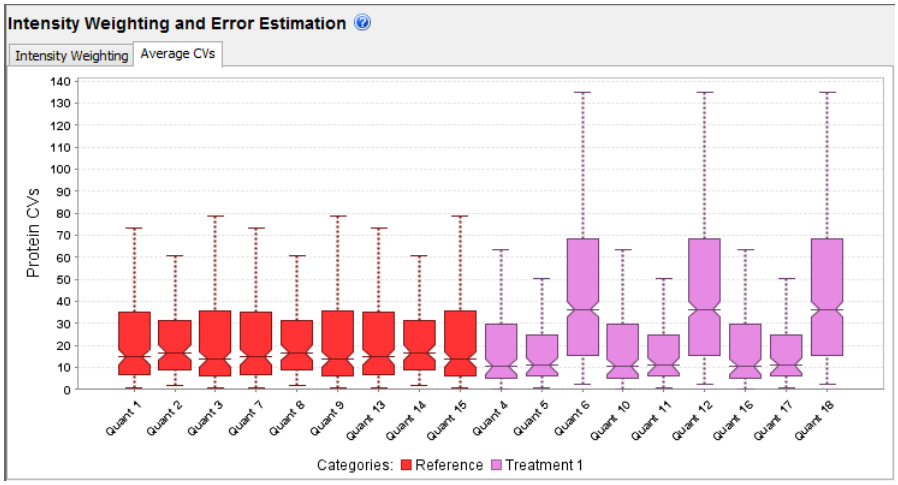
-

Figure 7-5: Intensity Weighting and Error Estimation, Intensity Weighting graph



- **Average CVs**—The Average CVs graph shows the distribution of the Coefficient of Variation (CV) for all the proteins in each Quant Sample. The graph displays the color-coding that was specified in the "Experimental Design Wizard" on page 45. This provides a quick visual way of comparing the differences in the CV distribution for all the quantitative samples in an experiment and noting if a Quant Sample shows a large deviation in its variance. If a Quant Sample has a significantly larger deviation in its CV in comparison to the other samples, then there is a higher likelihood that the sample data is bad or suspect. If all the deviations in the CVs for all samples seem relatively constant, then most likely, there is little error in the experimental data. Place your cursor on any box plot to display information about the median and interquartile range for the corresponding quantitative sample.

Figure 7-6: Intensity Weighting and Error Estimation, Average CVs graph



-

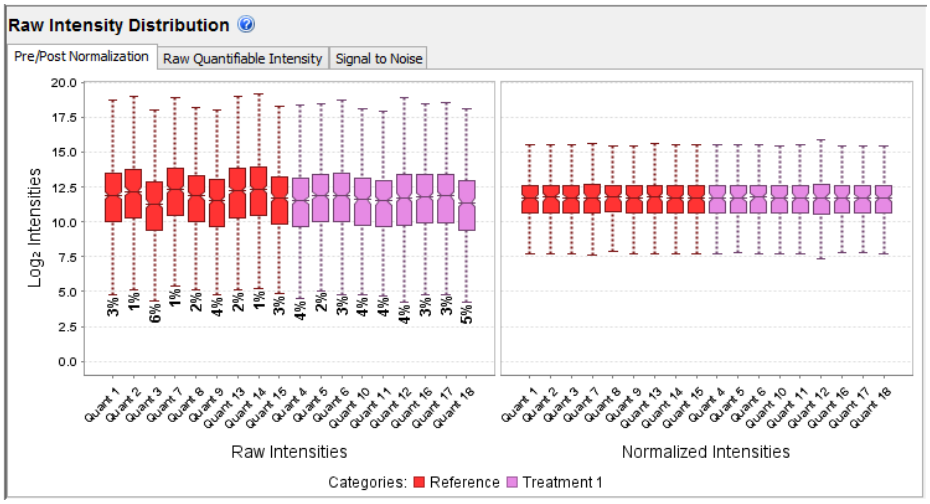
Raw Intensity Distribution

The Raw Intensity Distribution pane displays three graphs that provide insight into the normalization calculations carried out by Scaffold Q+ for the data set as a whole.

- **Pre/Post Normalization**—The Pre/Post Normalization box plot displays two distinct box plots to show the amount by which Scaffold Q+ normalized the raw intensities for all samples in each Quant Sample. The graph displays the color-coding that was specified in the "Experimental Design Wizard" on page 45. When placing the cursor on any box plot a tool tip displays information about the median and interquartile range for the corresponding Quant Sample. The percentage that is displayed below each Pre-normalization box plot indicates the percentage of missing value for the Quant Sample. Quantitative analysis of Quant Samples that have a high percentage of missing spectral data might be unreliable. See Figure 7-7 on page 108.

If you are carrying out this procedure using the sample iTRAQ data provided by Proteome Software, then note in Figure 7-7 on page 108, the samples in quantitative sample 3 had their intensities adjusted the most by normalization. Also, the sample dataset had no missing spectral data, which is not a typical scenario.

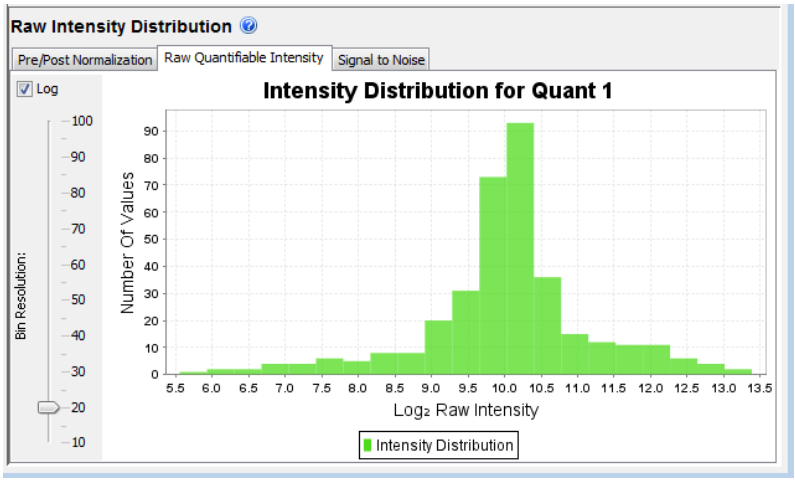
Figure 7-7: Raw Intensity Distribution, Pre/Post Normalization graph



- **Raw Quantifiable Intensity**—By default, the Raw Quantifiable Intensity graph is a histogram that reflects the distribution of the Log2 of all the raw intensities for a selected primary Quant Sample. Ideally, this histogram should indicate a normal (Gaussian) distribution. To view the distribution of non-transformed raw intensities, clear the Log option that is displayed in the top left corner of the graph. The Bin Resolution slider that is displayed to the left of the graph designates the number of *bins*, or data sections, of equal widths to display in the graph. For example, a Bin Resolution equal to 10 produces

a histogram with 10 distinct bins of equal widths, with the bars weighted based on the number of data points that they contain.

Figure 7-8: Raw Intensity Distribution, Raw Quantifiable Intensity histogram, Log2 Intensity Distribution

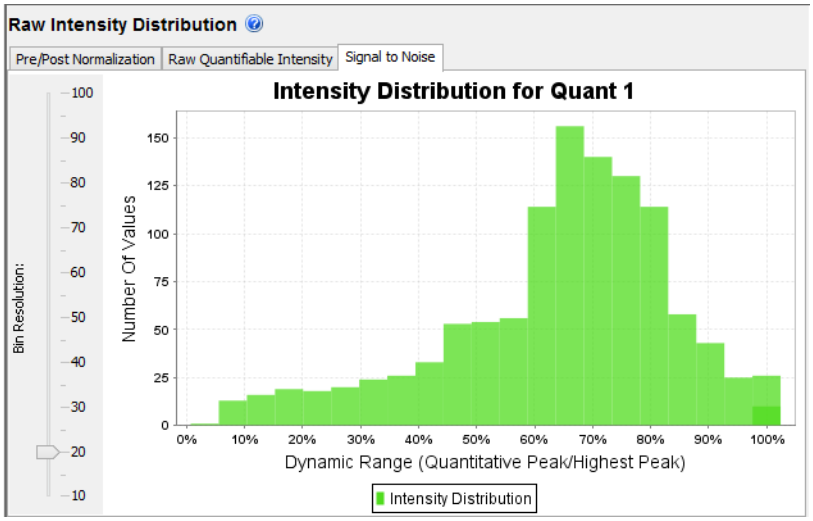


- **Signal to Noise**—The Signal to Noise graph is a histogram that shows the raw intensity of the reporter ions or MS1 ions compared to the largest peak, or *dynamic range*, in the spectrum for a selected primary Quant Sample. The Bin Resolution slider that is displayed to the left of the graph designates the number of *bins*, or data sections, of equal widths to display in the graph. For example, a Bin Resolution equal to 10 produces a histogram with 10 distinct bins of equal widths, with the bars weighted based on the number of data points that they contain. This distribution helps in understanding if the reporter ions' or MS1 ions' intensities for the selected quantitative sample have enough overall intensity to be considered. For example, in the case of iTRAQ-labeled samples, the largest peak is often one of the peaks that identifies a peptide rather than one of the reporter ions. If most of the reporter ion peaks are very small, this might indicate that the collision energy was tuned to fragment the peptide backbone rather than the reporter ions themselves. Low reporter ion peaks can be less accurate if the instrument has a limited dynamic range.

For both the Raw Quantifiable Intensity graph and the Signal to Noise graph, there is no "best" number of bins, and different bin sizes can reveal different features of the data.

-

Figure 7-9: Raw Intensity Distribution, Signal to Noise histogram



If you are carrying out this procedure using the sample iTRAQ data provided by Proteome Software, then select Quant 1 as the primary sample and note the good dynamic range for the data.

-

Chapter 8
Quantitative Settings

The goal of quantitative analysis in Scaffold Q+ and Scaffold Q+S is to detect differential expression of a protein between Quantitative Categories. The quantitative settings that the user specifies in the Quantification Setup dialog determine the type of evidence that is produced to support the presence or absence of differential expression. On the main menu of the Q+ Quantitation Module window, selecting **Quant > Quantitative Settings** opens the Quantification Setup dialog.

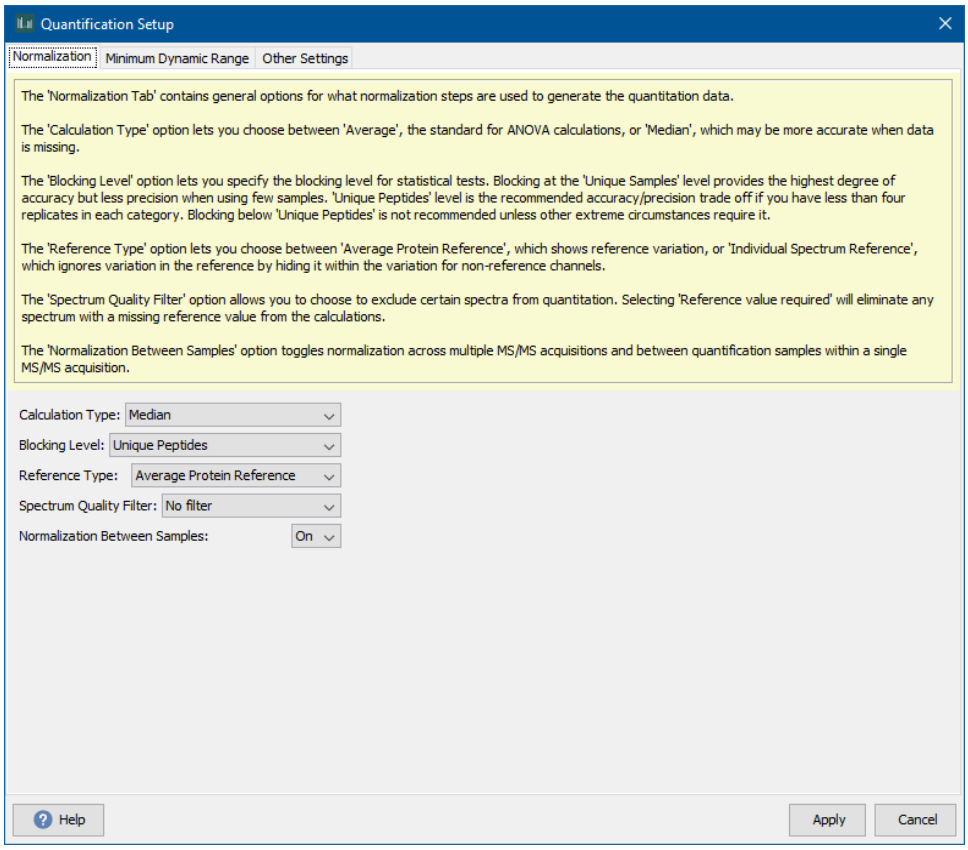
Quantification Setup dialog

The dialog contains three tabs. When it opens it shows the Normalization tab first.

- "Normalization tab" on page 113.
- "Minimum Dynamic Range tab" on page 118.
- "Other Settings tab" on page 119.

-

Figure 8-1: Quantification Setup dialog, Normalization tab



-

Normalization tab

The Q+ Quantitation Module calculates the normalization of the Log_2 of the raw quantitative data at the spectra level, rolls up the results from the spectra level to the peptide level, and then from the peptide level to the protein level using either the kernel density average or mean of the quantitative data based on what is selected for the Calculation Type. The scheme that is applied to the Log_2 of the quantitative raw data is based on a normalization model described in "Normalization method used in Scaffold Q+ and Scaffold Q+S" on page 54

The Normalization tab contains options for specifying how the normalization calculation is to be carried out for the quantitation data.

- **Calculation Type**—Two options, median (the default value) and mean, are available. Proteome Software recommends to use the Median option when analyzing isobaric tags (iTRAQ and TMT) or Precursor Intensity type of quantitation since it is the most robust option for that type of data. In the case of Stable Isotope labeling techniques (i.e. SILAC) the only calculation type available is the Mean.

  When the user selects Median for the calculation type, non-parametric statistical tests can be carried out on the quantitation data. Selecting Mean for the calculation type is more appropriate when the quantitation data is well-behaved in the following ways:

  - Little missing data.
  - Few outliers.
  - Little data below the detection limit.
  - Little distortion because of saturation at high intensities.
  - Little variation in the spread of data between the datasets.
  - Little skew (The data looks symmetrical).
  - The data has normal (Gaussian) distribution (No fat or thin tails)

  When the user selects Mean for the calculation type, standard statistical tests, the t-test and ANOVA, can be carried out on the quantitation data.

- **Blocking Level**— Blocking is a statistical tool used to guard against biases and minimize variances within a study. Specimen assayed within a block are more similar than specimen assayed between different blocks. Blocking isolates the variability attributable to the differences between blocks so that differences caused by a treatment, for example, become clearer.

  The Blocking Level affects the measurements for inclusion in statistical tests (like Permutation test, T-Test or ANOVA) and the Kernel Density Estimator (KDE) chart. It provides different levels of blocking that are used for computing, for example, the Random permutation test or when applying other inference tests, see Blocking Level example. Specifically, in the Permutation test the sampling distribution for any test statistic, under the Null hypothesis that no difference is present among categories, is computed by randomly shuffling the normalized intensity values in the different

-

categories. The blocking level determines the groups from which the random intensities are drawn from.

In the Q+ Quantitation Module the available levels of blocking are:

- **All Spectra** - no blocking
- **Unique spectra** - Consider only unique spectra
- **Unique peptides** - Consider only groups of unique peptides. This choice allows to compare measurements for each peptide. Most people running Proteomics experiments don't have a lot of data.In this case the user can consider the quantitative measurements for each peptide as a proxy for different samples. While this confounds answering a biological question about changing proteins with peptide level variability, it can provide a measurement to carry forward to future studies. In terms of the KDE chart, each peptide in each sample provides a single sub-distribution, resulting in many more measurements.
- **Unique samples** - Consider only samples. In a statistical test when analyzing multiple groups the user would usually compare the quantitative intensities for a protein of one group of biological samples versus another group. In doing so he/she would aggregate all of the peptide level quantitative values into a single measurement for each sample and then compare those measurements, this is exactly what the "Unique samples" option does. To use this level of blocking the user needs at least 4 to 8 samples in each group to say something statistically significant about a changing protein. In terms of the KDE chart, each sample provides a single sub-distribution.

When an experiment is conducted with a large number of replicates, we suggest to set the Blocking level at the samples level. Otherwise the peptide level provides a good level of blocking when a fairly average amount of data is available. While we support blocking below that point, it is not recommend unless the user has a very good reason (e.g. measuring isolated changes in specific peptides or phosphorylation sites) as each additional level below adds new confounding variability.

- **Reference Type**—Fold change is the ratio of the reporter ion intensity or the MS1 ion intensity to a reference ion intensity for a quantitative sample. The Reference Type provides two options for the denominator in this ratio, Average Protein Reference (the default value) or Individual Spectrum Reference. The *Average Protein Reference* is appropriate for experiments that are not time-based. If the Average Protein Reference is selected, then the median or mean value of the spectra intensity is calculated for all the spectra in the reference channels to give a single reference value. This means that the denominator is the same for all the samples, regardless of the n-plex (iTRAQ 4-plex, TMT 6-plex, SILAC 2-plex, and so on) that was used. It also means that the graphs of the quantitation data show the amount of variance in this single reference channel. The *Individual Reference Spectrum* is appropriate when the experimental design requires ratios that are derived from references within each biosample and not from a reference that is common to all biosamples, for example, a Before and After Treatment time-based

-

experiment. Consider the following experimental design—two mice, with Mouse #1 representing the Before Treatment and Mouse #2 representing After Treatment. Samples are taken from each mouse at four different time points. One iTRAQ 4-plex is used for all of the samples from Mouse #1. The Other iTRAQ 4-plex is used for Mouse #2. With Individual Reference Spectrum selected, all the measurements for Mouse #1 are compared back to the measurement taken at its first time point. Likewise, all the measurements for Mouse #2 are compared back to the measurement taken at its first time point.

- **Spectrum Quality Filter**—which allows to choose to exclude from quantitation spectra that have reference missing values, see .
- **Normalization Between Samples** —Regardless of the option that the user selects for Normalization Between Samples, Q+ Quantitation Module always carries out some normalization *within* a sample. The default Normalization Between Samples option is ON and it results in normalization being carried out *between* samples. The OFF option prevents normalization from being carried out between samples, and there are some experimental designs where turning off normalization between samples is appropriate, for example, if the experimental designed involved enriching for different proteins in different samples, or using a different dilution for each sample.

To understand how the Q+ Quantitation Module treats missing data in the Normalization, see "Treatment of Missing Data by the Q+ Quantitation Module" on page 120.

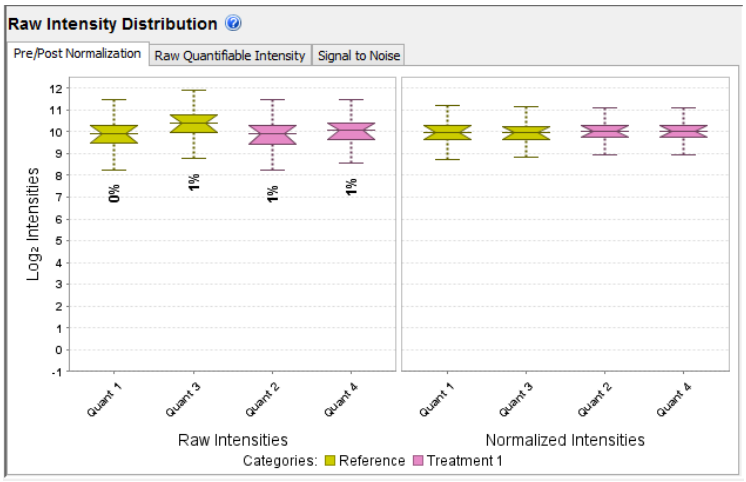
Familiarizing yourself with the Normalization tab

If you following along using the sample iTRAQ data provided by Proteome Software, then make sure that the first protein that is listed in the Proteins table, Apolipophorins OS-Drosophila melanogaster, is selected.

1. Click Cancel to close the Quantification Setup dialog box and return to Scaffold Q+.
2. Open the Statistics View for the data.
3. In the Raw Intensity Distribution pane, open the Pre/Post Normalization graph.

-

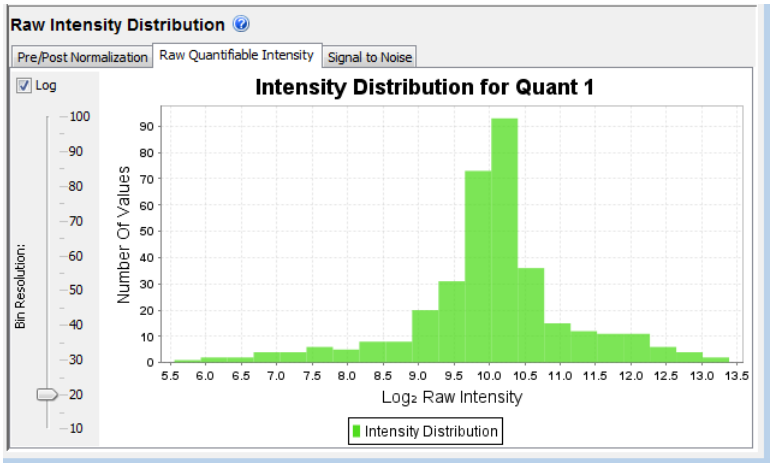
Figure 8-2: Raw Intensity Distribution, Pre/Post Normalization graph



4. Examine the graph and note that very little data is missing, that all the quantitative samples have approximately the same variance, and that relatively little adjustment would be required to align the samples.

5. Open the Raw Quantifiable Intensity graph for the Quant 1 primary sample.

Figure 8-3: Raw Intensity Distribution, Raw Quantifiable Intensity histogram, Log_2 Intensity Distribution



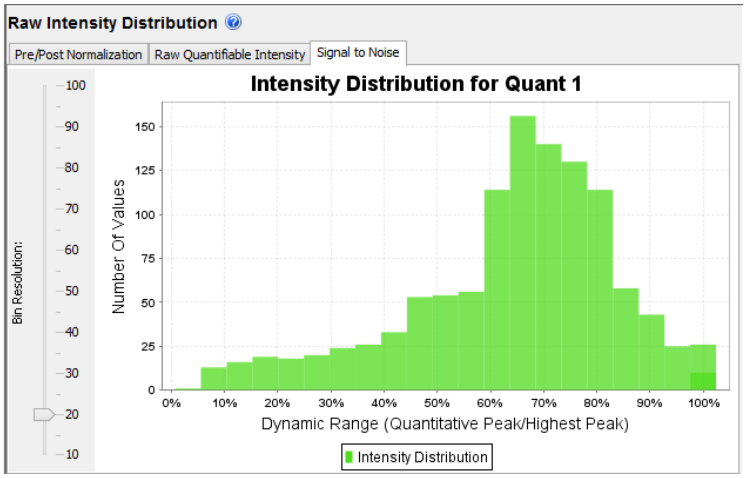
6. Examine the graph and note the apparently normal (Gaussian) distribution of data for the sample.

7. Select each Quant Sample in turn as the primary sample and examine the Raw Quantifiable Intensity histogram for the sample. Note how some of the samples do not exhibit a normal (Gaussian) distribution of data.

8. Open the Signal to Noise tab, and then cycle through the histograms for each of the Quant Samples. If the distribution displays a good dynamic range, (for example, as Quant 1 does) this indicates that the data is "well-behaved."

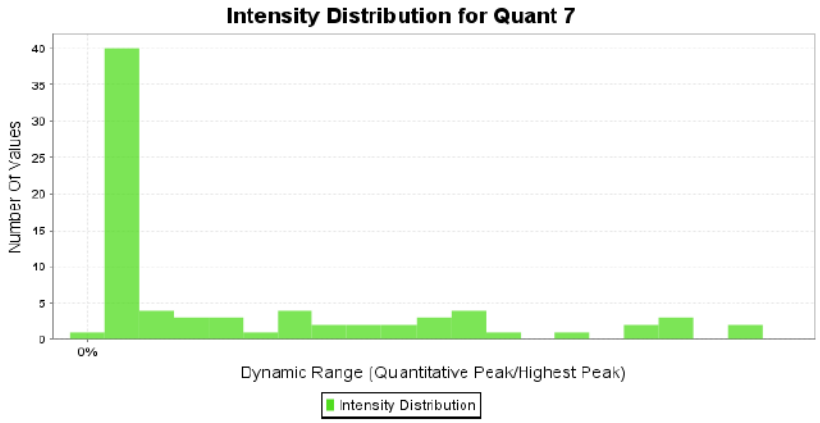
-

Figure 8-4: Raw Intensity Distribution, Signal to Noise histogram, Good dynamic range



A poor dynamic range—that is, many small reporter ion peaks—such as that shown in Figure 8-5 below (the data is from a different dataset than the sample iTRAQ dataset), might indicate that the collision energy was tuned to fragment the peptide backbone instead of the reporter ions. Low reporter ion peaks can be less accurate if the instrument used has a limited dynamic range.

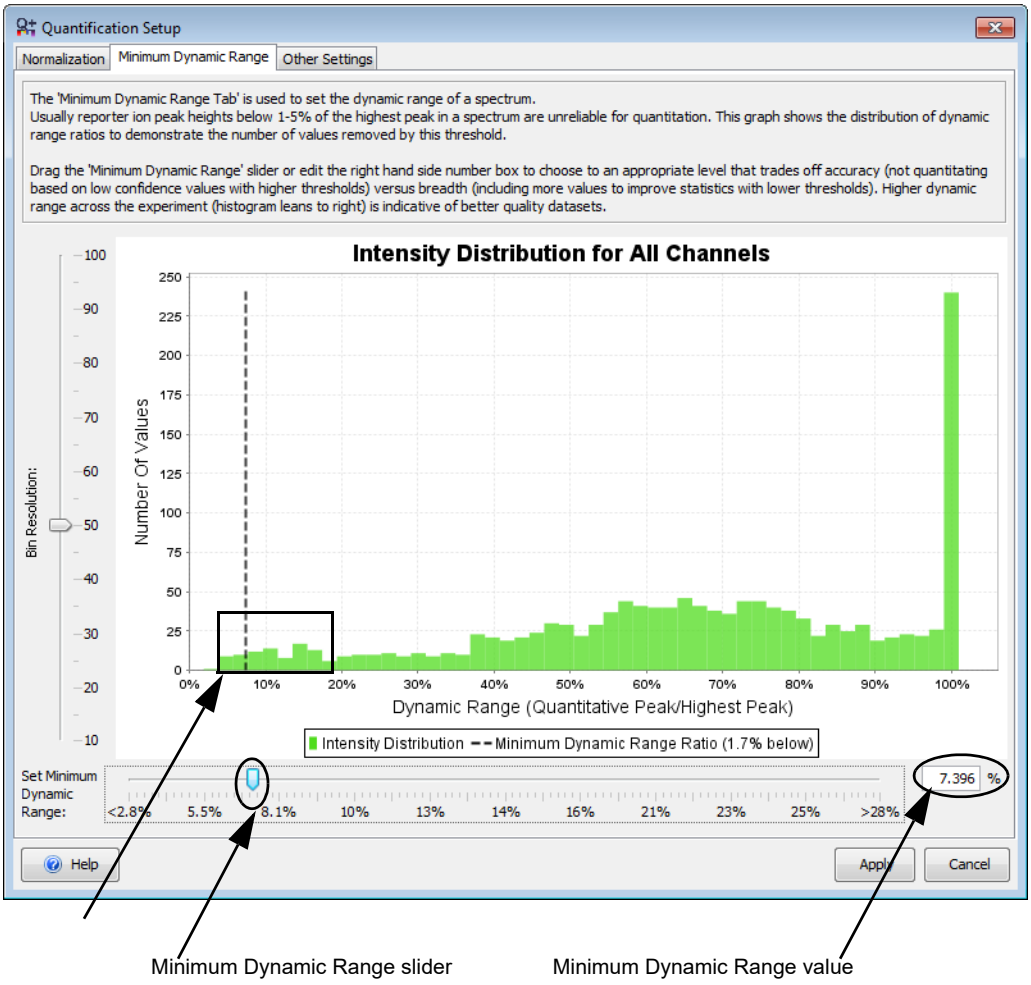
Figure 8-5: Raw Intensity Distribution, Signal to Noise histogram, Poor dynamic range



-

Minimum Dynamic Range tab

Figure 8-6: Quantification Setup dialog box, Minimum Dynamic Range tab



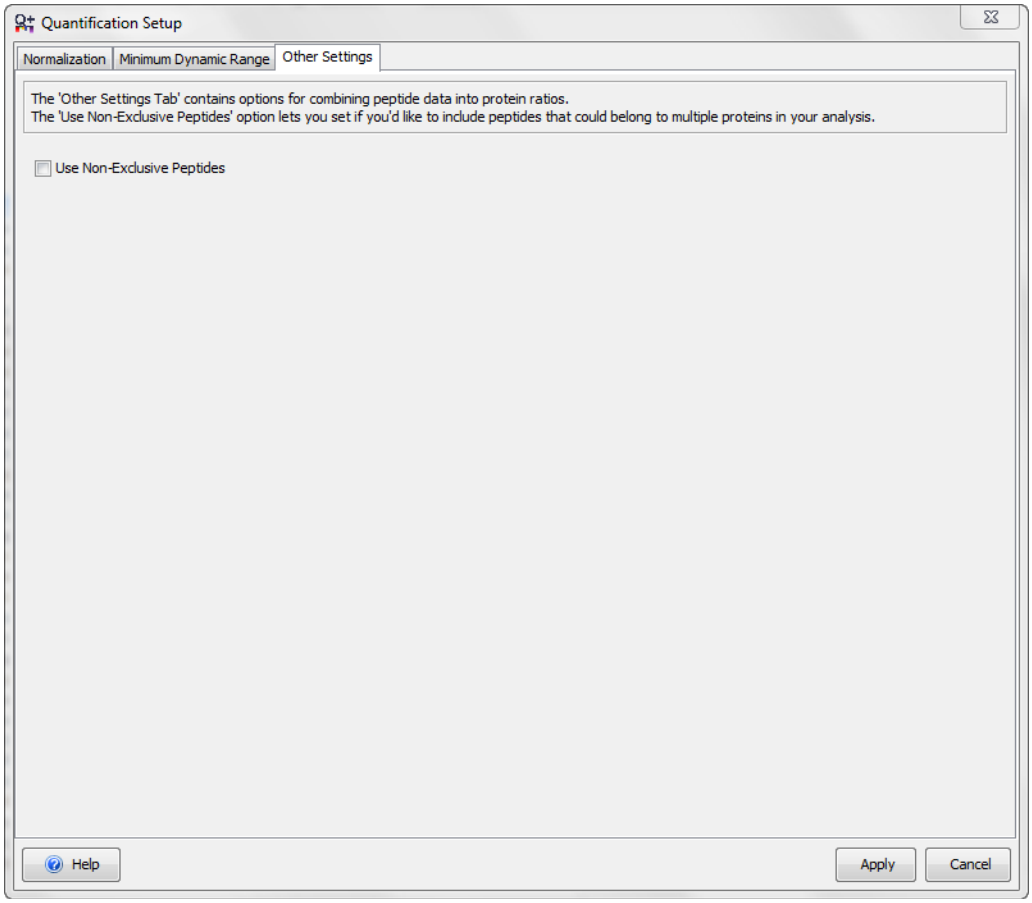
The Minimum Dynamic Range tab displays the Signal to Noise histogram, which plots the raw intensities of the reporter ions or MS1 ions compared to the highest peak (dynamic range) of the spectrum for all channels. The *minimum dynamic range* is defined as a percentage of the highest peak in the spectrum and it determines the value that the Q+ Quantitation Module uses in place of the intensity of reporter ions or MS1 ions when the intensity falls below a minimum threshold. The default value is 1% of the intensity value of the larger reporter ion or MS1 ion peak in the spectrum.

If you are following along using the sample iTRAQ data provided by Proteome Software, then drag the Minimum Dynamic Range slider at the bottom of the histogram and note how the minimum dynamic range value is adjusted in the display field to the right of the slider. Also note the vertical dotted line is adjusted accordingly, which indicates the values are being excluded.

-

Other Settings tab

Figure 8-7: Quantification Setup dialog box, Other Settings tab



Because reporter ion and MS1 ion quantitative measurements are made at the peptide level, the model that is used for normalizing data should reflect the relationship between peptide and the associated protein expression levels. The association might be ambiguous as some tryptic peptides can be associated with more than one protein. These peptides are referred to as “degenerate,” and they are typically eliminated from an analysis. Because the relative contributions of the different proteins to the intensity of a shared peptide cannot be determined, Scaffold Q+ considers quantitation to be reliable only when the proteins that are being analyzed include peptides that are not shared with other proteins. “No Values” is displayed in the various Quant columns for proteins that include peptides that are shared by other proteins and the Proteins view is not active for analysis. If you still want to consider these proteins in the analysis, then you must select “Use Non Exclusive Peptides.” Scaffold Q+ uses the full intensity value for each associated peptide when calculating the fold changes for the proteins. It does not apportion the values.

-

Treatment of Missing Data by the Q+ Quantitation Module

The Q+ Quantitation Module defines Quantitative values as *missing* when isobaric tags or isotopic labels or precursor intensity values have either a raw value of zero or the values fall below the “Minimum Dynamic Range.”

Within each MS-Sample a “missing value” is assigned the larger value between (a) the minimum positive logged intensity acquired and (b) the value whose z-score is -4 for the distribution of all logged values in the MS-Sample.

It is possible to remove from the calculation spectra with missing values in the reference channel through the Spectrum Quality filter option available in the Quantification Setup dialog.

-

Chapter 9
Quantitative Testing

The goal of quantitative analysis in Scaffold Q+S is to detect differential expression of a protein between Quantitative Categories. While a fold change value can indicate differential expression of a protein between sample categories, the user should carry out statistical testing to determine if the evidence that supports the presence or absence of differential expression is truly conclusive or happens by chance. Scaffold Q+S provides a variety of statistical tests for this purpose. Different tests are appropriate for different types of experiments.

Selecting the **Quant > Quantitative Testing** command from the Q+ Quantitation Module main menu opens the Quantitative Testing dialog where the user may select the tests to be carried out for the current data. By default, when the dialog first opens, all Quant Samples in all Quantitative Categories for the experiment are selected for testing, unless the user has selected the Between-Subject (Common/Pooled Reference) Experimental Type, in which case the Reference Samples are not available for analysis.

Depending on the Analysis Type (Intensity-based or Ratio-based) selected when setting up the experiment in the Experimental Design Wizard, the Quantitative tests are computed using the Log_2 Intensity for Intensity-based analyses and the Log_2 Ratio for Ratio-based analyses.

The tests that are available on the dialog for the experiment depend upon the following conditions:

1. The Experiment Type selected when setting up the experiment in the Experimental Design Wizard.
2. The Calculation Type (median or mean) selected from the Quantification Setup dialog.
3. The number of Quantitative Categories selected when setting up the experiment in the Experimental Design Wizard.

See:

- “Test assumptions” on page 123
- “Median-based Between-subjects Statistical Tests” on page 124
- “Mean-based Between-subjects Statistical Tests” on page 127
- “Median-based Repeated Measures Statistical Tests” on page 130
- “Mean-based Repeated Measures Statistical Tests” on page 133
- “Common Statistical Test Options” on page 136

-

- “Blocking and inferential tests” on page 139
- “Coloring of Test results” on page 139

-

Test assumptions

Normality Testing

Normality of the distribution of residuals in each category is an assumption of all parametric tests we support. For our test we use Kolmogorov-Smirnov to assess normality of all residuals grouped together. When the categories' residuals have equal variance and, when grouped together, are normally distributed it is extremely unlikely the individual distributions are not normally distributed.

Equal Variances

We use the Brown-Forsythe test to assess whether the variances of the residuals are equal across categories. This is an explicit assumption for the parametric tests.

The non-parametric tests other than the Permutation test assume that the distributions of residuals across categories are i.i.d. (independent and identically distributed). We do not test for this explicitly, but instead use this testing of equal variances as a partial proxy for this condition. If the variances are unequal, then certainly the distributions are not identically distributed.

Hence, we apply the Brown-Forsythe test to all our statistical tests except the permutation test.

Sphericity Correction

The Repeated Measures Analysis of Variance (rANOVA) test makes a further assumption that the distributions of residuals across categories satisfy a so-called sphericity condition. Rather than test this assumption directly, we automatically apply a correction known as the Greenhouse-Geisser epsilon. The error and model degrees of freedom for rANOVA are both multiplied by this epsilon before calculation of a p-value.

Interpretation

It is important to note that these assumption tests are only guides for interpretation and should not be accepted as a definitive decision criterion for accepting or rejecting a significance result for any specific protein. That is, for a significant protein whose p-value is flagged as orange, the analysis may indeed be valid. For example, the assumption tests themselves may be over sensitive to deviations from normality when n is large. Many statistical tests (both parametric and non-parametric) are relatively robust even when their theoretical assumptions are not met. Similarly, a green p-value does not imply that the analysis can be accepted without question because some test assumptions (such as independence of error deviations for ANOVA) are not measurable within Scaffold Q+S.

-

Median-based Between-subjects Statistical Tests

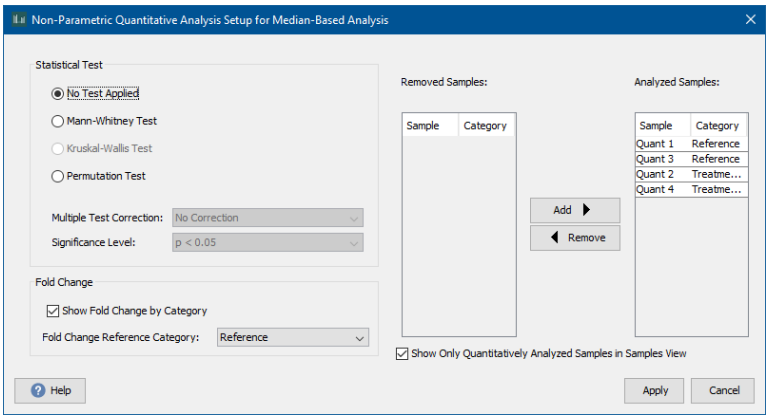
When the user specifies as the Experiment Type either the **Between-subjects (Independent Groups)** or the **Between-subjects (Common/Pooled Reference)** in the Experimental Design Wizard and chooses the median as the Calculation Type for the current analyzed data, the Quantitative Analysis dialog opens with the following non-parametric statistical tests:

- Mann Whitney test.
- Kruskal-Wallis test.

Other options—No Test Applied, Permutation Test, Multiple Test Correction and Significance Level, Fold Change by Category pane and Show only Quantitatively Analyzed Samples in Samples View—are also available. See Common Statistical Test Options.

Note that the Blocking Level selection determines the blocking group considered for computing the tests, see Blocking and inferential tests.

Figure 9-1: Non-Parametric Quantitative Analysis Setup for Median-Based Between-subject Analysis dialog



Mann Whitney test

The Mann Whitney test is a non-parametric statistical hypothesis test for assessing whether one of two samples of independent observations tends to have larger values than the other. It can also be defined as a distribution-free test of whether two medians are equal. The test uses the ranks of the data in the two samples. Although the Mann Whitney test compares well with a t-test, it is independent of the way the data is distributed. Because the Mann Whitney test is the non-parametric version of the t-test, it requires exactly two quantitative sample categories to be selected for testing. By default, when the Quantitative Testing dialog first opens, all available Quant Samples in all Quantitative Categories for the experiment are selected for testing. As a result, if the user's experiment contains more than two quantitative sample categories, then the Mann Whitney test option is not available. The user should select all samples in all but two of the quantitative sample categories, and then click Remove before the option becomes available.

After the user selects the Mann Whitney Test, and then clicks Apply, the Quantitative

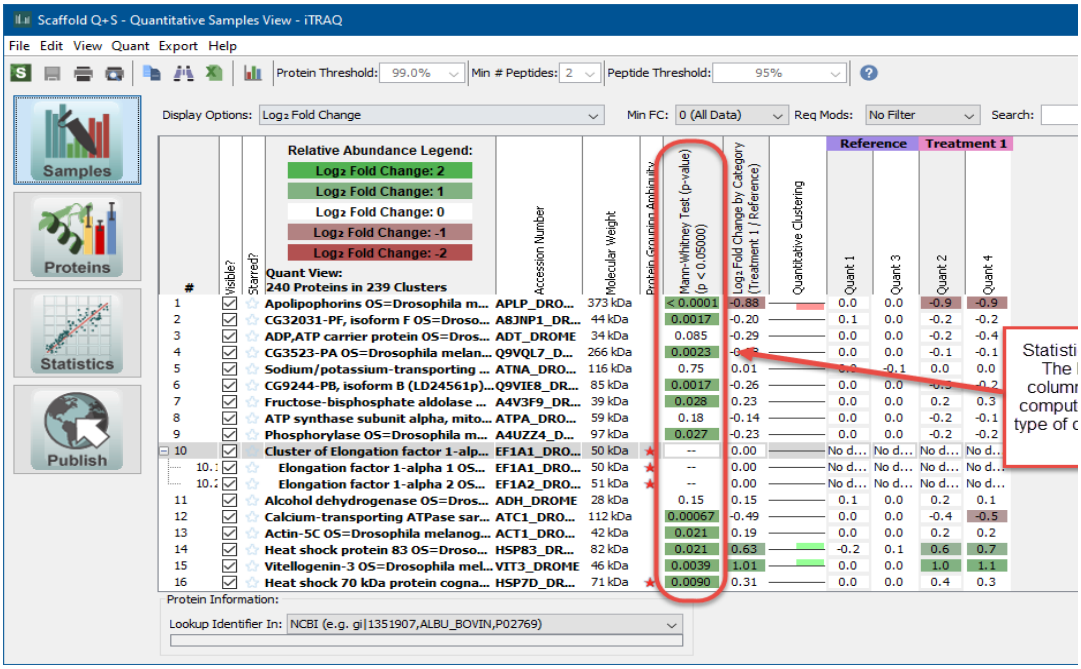
-

Samples View opens showing the experimental data. The Mann Whitney p-values are displayed in a Mann Whitney Test (P-Value) column. The statistically significant values, determined by the selected Significant Level are highlighted in green when the Test assumptions are met or flagged in orange when otherwise, see Coloring of Test results.

The user has always the option of applying a Multiple Test Corrections to the calculated p-values and the name of chosen correction is reported in the Mann Whitney Test (P-Value) column header with its corrected significant value.

The user can also select the option Show only Quantitatively Analyzed Samples in Samples View to hide from view the Quant Samples not selected for the test.

Figure 9-2: Quantitative Samples View, Mann Whitney test carried out for data



Kruskal-Wallis test

The Kruskal-Wallis one-way analysis of variance by ranks is a non-parametric method for testing whether samples originate from the same distribution. It is used for comparing more *three or more samples* that are independent, or not related. (The parametric equivalence of the Kruskal-Wallis test is the one-way analysis of variance (ANOVA)). The factual null hypothesis is that the populations from which the samples originate have the same median. When the Kruskal-Wallis test leads to significant results, then at least one of the samples is different from the other samples. The test does not identify where the differences occur or how many differences actually occur. Because it is a non-parametric method, the Kruskal-Wallis test does not assume a normal distribution (which is unlike the analogous one-way analysis of variance); however, the test does assume an identically-shaped and scaled distribution for each group, except for any difference in medians. After the user selects the three or more Quant Samples in different Quantitative Categories that are to be tested, he/she

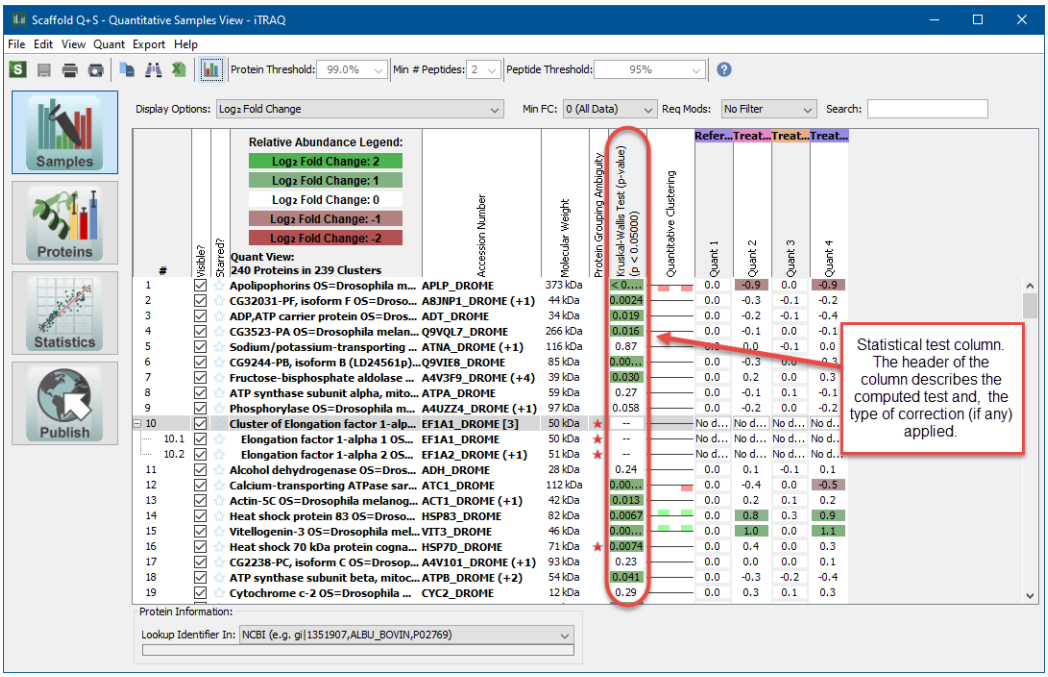
-

should select Kruskal-Wallis Test, and then click Apply. The Quantitative Samples View opens showing the experimental data. The Kruskal-Wallis p-values are displayed in a Kruskal-Wallis Test (P-Value) column. The statistically significant values, determined by the selected Significant Level, are highlighted in green when the Test assumptions are met or flagged in orange when otherwise, see Coloring of Test results.

The user has always the option of applying a Multiple Test Corrections to the calculated p-values and the name of the chosen correction is reported in the Kruskal-Wallis Test (P-Value) column header with its corrected significant value.

The user can also select the option Show only Quantitatively Analyzed Samples in Samples View to hide from view the Quant Samples not selected for the test.

Figure 9-3: Quantitative Samples View, Kruskal-Wallis test carried out for data



Because the Kruskal-Wallis test is an extension of the Mann-Whitney test for three or more categories, after you carry out the Kruskal-Wallis test for your data, you can then carry out the Mann Whitney test to analyze specific sample pairs for statistically significant differences in protein expression.

Mean-based Between-subjects Statistical Tests

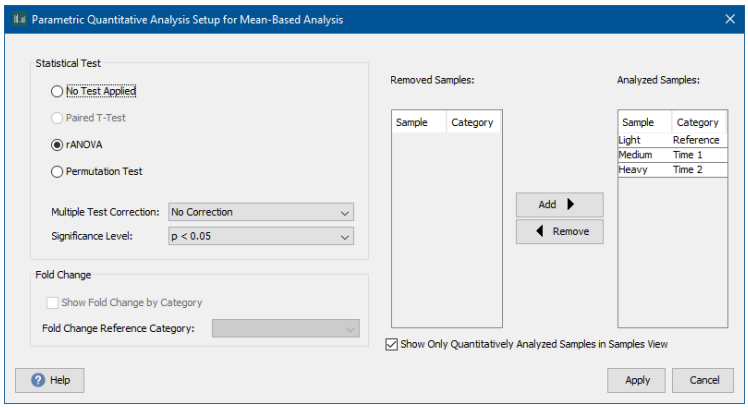
When the user specifies as the Experiment Type either the **Between-subjects (Independent Groups)** or the **Between-subjects (Common/Pooled Reference)** in the Experimental Design Wizard and chooses the Mean as the Calculation Type for the current analyzed data, the Quantitative Analysis dialog opens with the following parametric statistical tests:

- T-Test.
- ANOVA.

Other options—No Test Applied, Permutation Test, Multiple Test Correction and Significance Level, Fold Change by Category pane and Show only Quantitatively Analyzed Samples in Samples View—are also available. See Common Statistical Test Options.

Note that the Blocking Level selection determines the blocking group considered for computing the tests, see Blocking and inferential tests.

Figure 9-4: Parametric Quantitative Analysis Setup for Mean-based Between-subjects Analysis dialog



T-Test

The T-test is the parametric statistical hypothesis test for assessing whether the means of two groups are statistically different from one another. As a result, the T-test is useful for determining whether the difference in the expressions levels for a protein in two different Quantitative Categories is statistically significant. The smaller the p-value for a T-test, the more likely that the difference is indeed statistically significant. The T-test requires exactly two quantitative sample categories to be selected for testing and the test is reliable only if each category has at least three replicates. By default, when the Quantitative Testing dialog first opens, all available Quant Samples in all Quantitative Categories for the experiment are selected for testing. As a result, if the experiment contains more than two quantitative sample categories, then the T-Test option is not available. The user should select all samples in all but two of the quantitative sample categories, and then click Remove before the option is available.

After the user selects T-Test, and then clicks Apply, the Quantitative Samples View opens showing the experimental data. The T-Test p-values are displayed in a T-Test (P-Value)

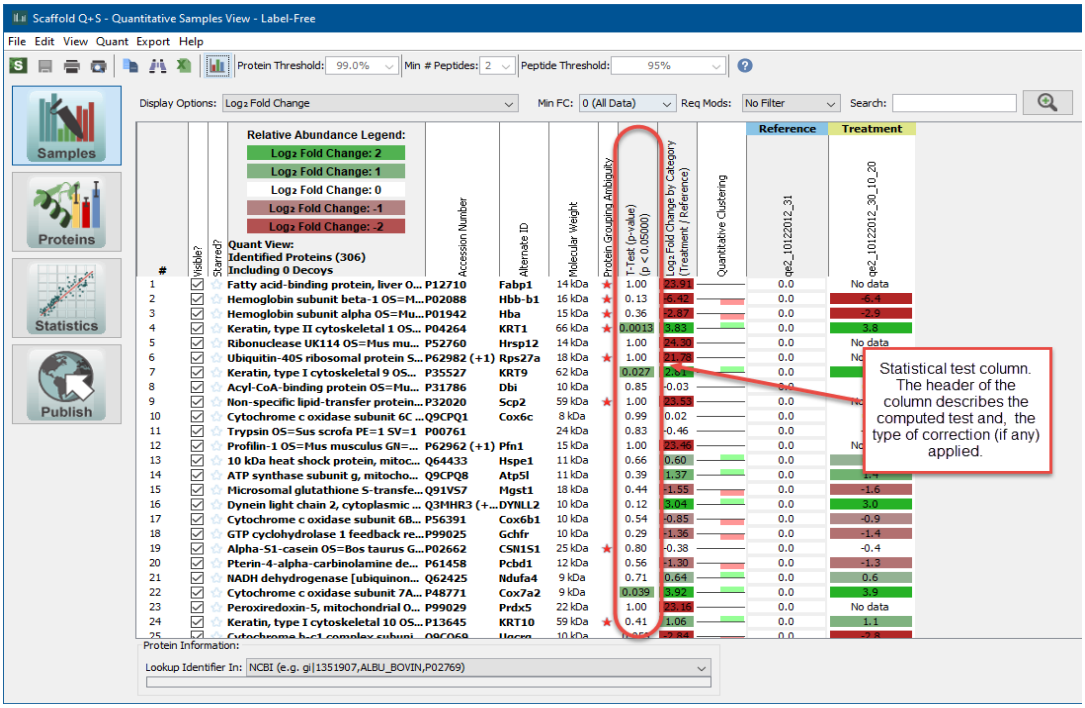
-

column. The statistically significant values, determined by the selected Significant Level, are highlighted in green when the Test assumptions are met or flagged in orange when otherwise, see “Coloring of Test results” on page 139.

The user has always the option of applying a Multiple Test Corrections to the calculated p-values and the name of the chosen correction is reported in the T-Test (P-Value) column header with its corrected significant value.

The user can also select the option Show only Quantitatively Analyzed Samples in Samples View to hide from view the Quant Samples not selected for the test.

Figure 9-5: Quantitative Samples View, T-Test carried out for data



ANOVA

ANOVA is one-way analysis of variance by ranks. It is a parametric method for testing whether samples originate from the same distribution. It is used for comparing *three or more samples* that are independent, or not related. (The non-parametric equivalence of the ANOVA is the Kruskal-Wallis test). The factual null hypothesis is that the populations from which the samples originate have the same mean. When ANOVA leads to significant results, then at least one of the samples is different from the other samples. ANOVA requires three or more quantitative sample categories to be selected for testing and the test is reliable only if each category has at least three replicates.

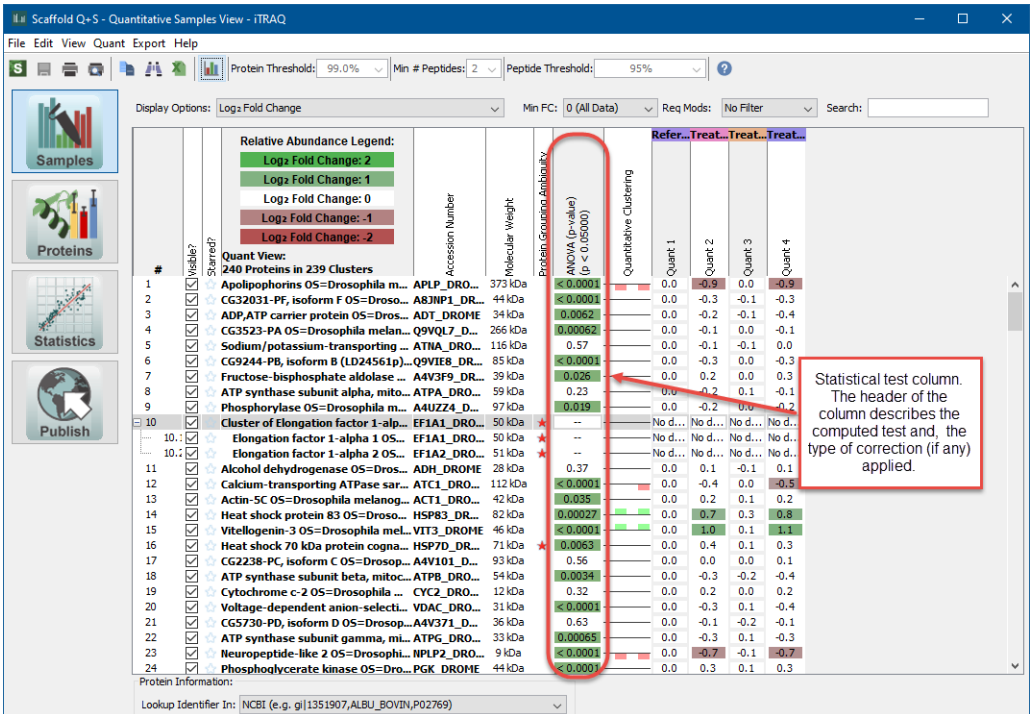
After the user selects ANOVA, and then clicks Apply, the Quantitative Samples View opens showing the experimental data. The ANOVA p-values are displayed in an ANOVA Test (P-Value) column. The statistically significant values, determined by the selected Significant Level, are highlighted in green when the Test assumptions are met or flagged in orange when

otherwise, see Coloring of Test results.

The user has always the option of applying a Multiple Test Corrections to the calculated p-values and the name of the chosen correction is reported in the ANOVA Test (P-Value) column header with its corrected significant value.

The user can also select the option Show only Quantitatively Analyzed Samples in Samples View to hide from view the Quant Samples not selected for the test.

Figure 9-6: Quantitative *Samples View, ANOVA carried out for data*



-

Median-based Repeated Measures Statistical Tests

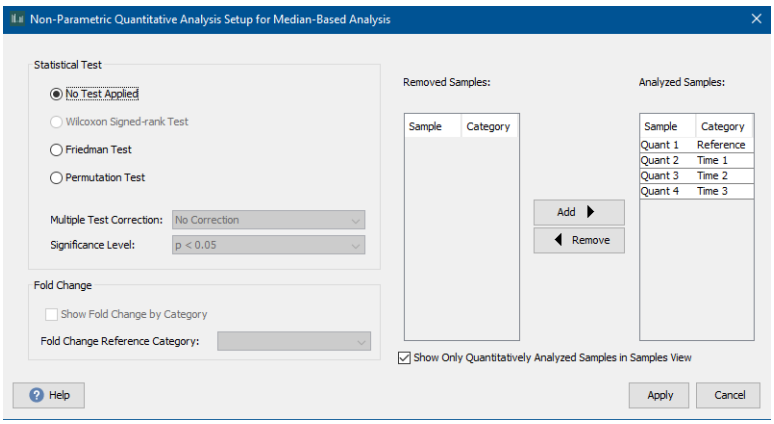
When the user specifies as an Experiment Type the **Repeated Measures / Time Course** in the Experimental Design Wizard and chooses the median as the Calculation Type for the current analyzed data, the Quantitative Analysis dialog opens with the following non-parametric statistical tests:

- Wilcoxon Signed-rank
- Friedman

Other options—No Test Applied, Permutation Test, Multiple Test Correction and Significance Level, Fold Change by Category pane and Show only Quantitatively Analyzed Samples in Samples View—are also available. See Common Statistical Test Options.

Note that the Blocking Level selection determines the blocking group considered for computing the tests, see Blocking and inferential tests.

Figure 9-7: Non-Parametric Quantitative Analysis Setup for Median-Based Repeated Measures Analysis dialog



Wilcoxon Signed-rank

The Wilcoxon Signed-rank test is a nonparametric statistical hypothesis test for assessing whether the population mean ranks of repeated measurements differ when the measurements have been taken under exactly two conditions (e.g., time points). It is a nonparametric alternative to the Paired T-Test, and may be used when the data being analyzed is not normally distributed. The Wilcoxon Signed-rank test does assume that the distributions in the two categories are independent and identically distributed (i.i.d.).

By default, when the Quantitative Testing dialog box first opens, all available Quant Samples in all Quantitative Categories for the experiment are selected for testing. As a result, if the user's experiment contains more than two quantitative sample categories, then the Wilcoxon Signed-rank Test option is not available. The user should select all samples in all but two of the quantitative sample categories, and then click Remove before the option is available.

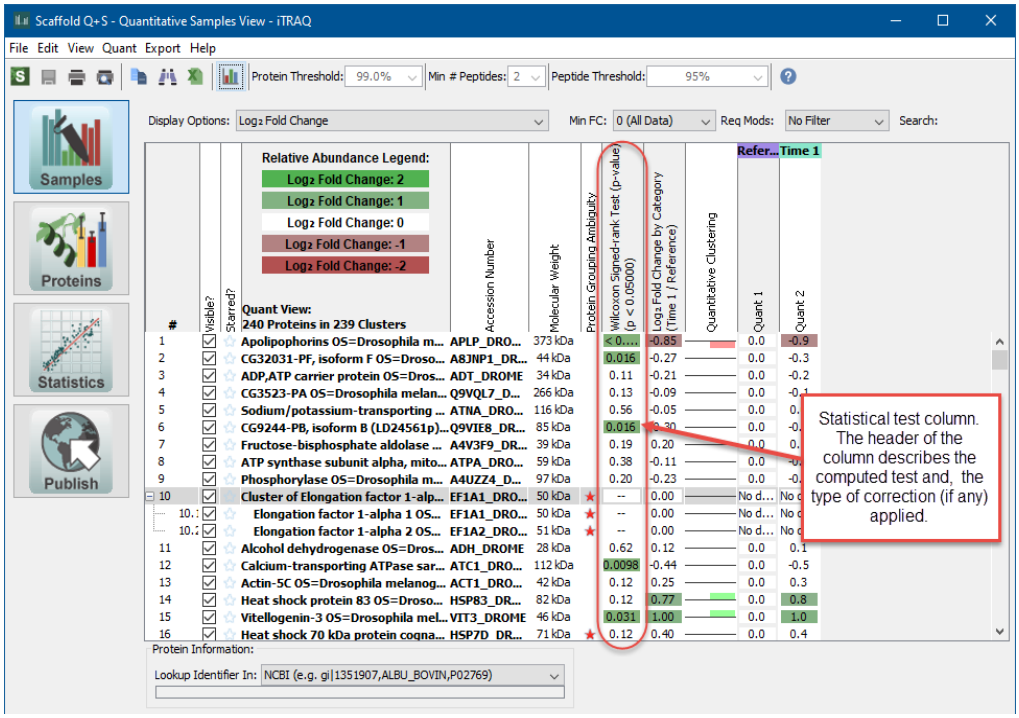
After the user selects the Wilcoxon Signed-rank Test, and then clicks Apply, the Quantitative Samples View opens showing the experimental data. The Wilcoxon Signed-rank p-values are

displayed in a Wilcoxon Signed-rank Test (P-Value) column. The statistically significant values, determined by the selected Significant Level, are highlighted in green when the Test assumptions are met or flagged in orange when otherwise, see Coloring of Test results.

The user has always the option of applying a Multiple Test Corrections to the calculated p-values and the name of chosen correction is reported in the Wilcoxon Signed-rank Test (P-Value) column header with its corrected significance value.

The user can also select the option Show only Quantitatively Analyzed Samples in Samples View to hide from view the Quant Samples not selected for the test.

Figure 9-8: Quantitative Samples View, Wilcoxon Signed-rank Test carried out for data



Friedman

The Friedman test is a nonparametric statistical hypothesis test for assessing whether the population mean ranks of repeated measurements differ when the measurements have been taken under at least three conditions (e.g., time points). It is a nonparametric alternative to the Repeated Measures Analysis of Variance (rANOVA), and may be used when the data being analyzed is not normally distributed. The Friedman test does assume that the distributions in the categories are independent and identically distributed (i.i.d.).

By default, when the Quantitative Testing dialog box first opens, all available Quant Samples in all Quantitative Categories for the experiment are selected for testing and the Friedman Test will be the default option. However, if the user's experiment contains only two quantitative sample categories, then the Friedman Test option will not be available.

After the user selects the Friedman Test, and then clicks Apply, the Quantitative Samples View opens showing the experimental data. The Wilcoxon Signed-rank p-values are

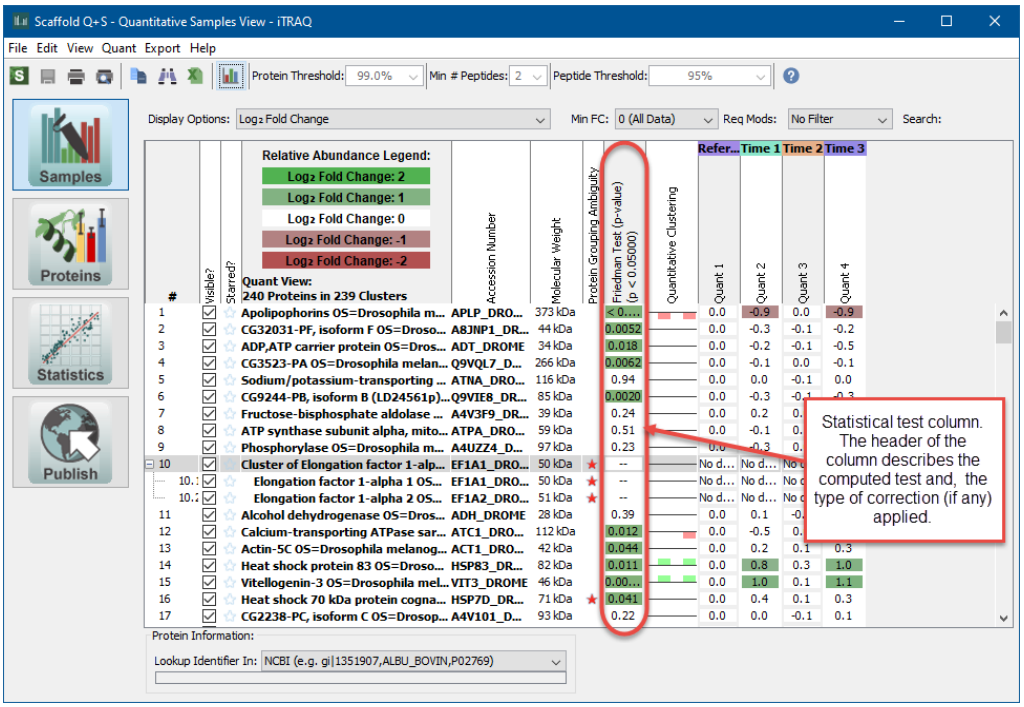
-

displayed in a Friedman Test (P-Value) column. The statistically significant values, determined by the selected Significant Level are highlighted in green when the Test assumptions are met or flagged in orange when otherwise, see Coloring of Test results.

The user has always the option of applying a Multiple Test Corrections to the calculated p-values and the name of the chosen correction is reported in the Friedman Test (P-Value) column header with its corrected significance value.

The user can also select the option Show only Quantitatively Analyzed Samples in Samples View to hide from view the Quant Samples not selected for the test.

Figure 9-9: Quantitative Samples View, Friedman Test carried out for data



Mean-based Repeated Measures Statistical Tests

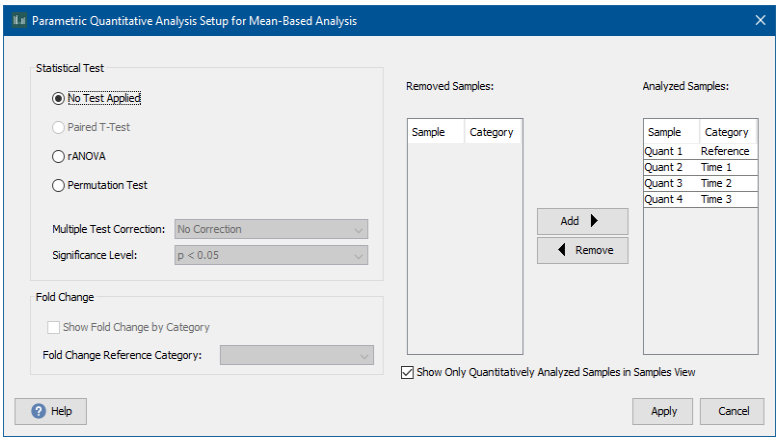
When the user specifies as an Experiment Type the **Repeated Measures / Time Course** in the Experimental Design Wizard and chooses the mean as the Calculation Type for the current analyzed data, the Quantitative Analysis dialog opens with the following parametric statistical tests:

- Paired T-Test
- Repeated Measures Analysis of Variance (rANOVA)

Other options—No Test Applied, Permutation Test, Multiple Test Correction and Significance Level, Fold Change by Category pane and Show only Quantitatively Analyzed Samples in Samples View—are also available. See Common Statistical Test Options.

Note that the Blocking Level selection determines the blocking group considered for computing the tests, see Blocking and inferential tests.

Figure 9-10: Parametric Quantitative Analysis Setup for Mean-Based Repeated Measures Analysis dialog



Paired T-Test

The Paired T-Test is a parametric statistical hypothesis test for assessing whether the population means of repeated measurements differ when the measurements have been taken under exactly two conditions (e.g.,time points). It is a parametric alternative to the Wilcoxon Signed-rank test, and may be used when the data being analyzed is normally distributed and has equal variances in the two categories. By default, when the Quantitative Testing dialog first opens, all available Quant Samples in all Quantitative Categories for the experiment are selected for testing. As a result, if the user's experiment contains more than two quantitative sample categories, then the Paired T-Test option is not available. The user should select all samples in all but two of the quantitative sample categories, and then click Remove before the option becomes available.

After the user selects the Paired T-Test, and then clicks Apply, the Quantitative Samples View opens showing the experimental data. The Paired T-Test p-values are displayed in a

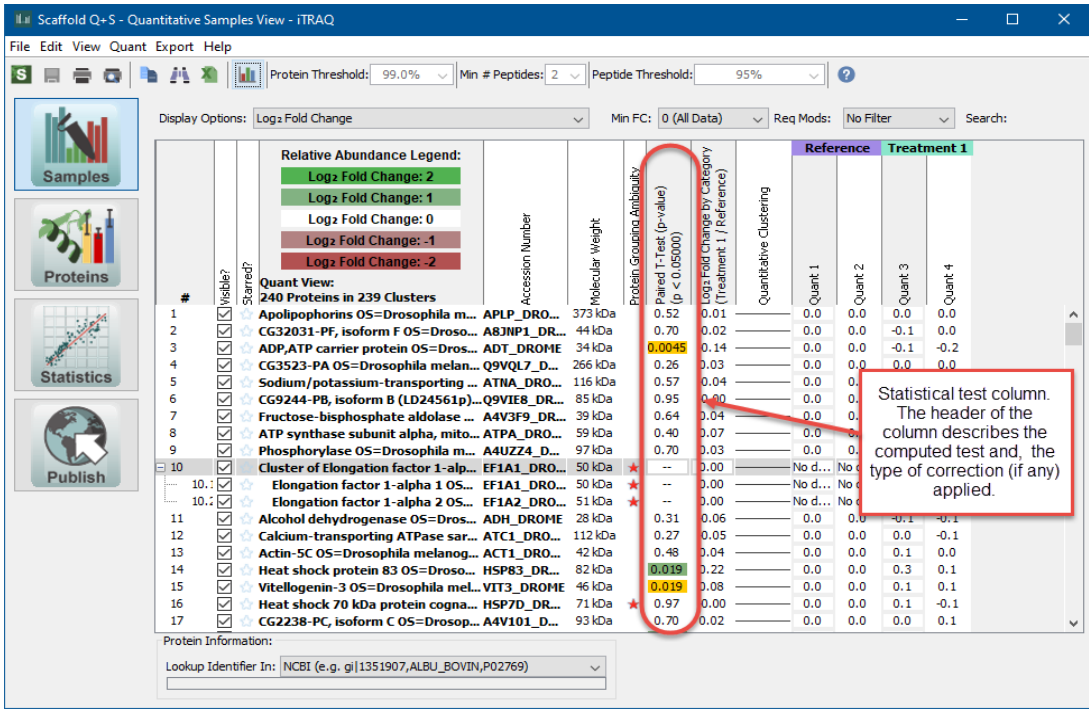
-

Paired T-Test (P-Value) column. The statistically significant values, determined by the selected Significant Level are highlighted in green when the Test assumptions are met or flagged in orange when otherwise, see Coloring of Test results.

The user has always the option of applying a Multiple Test Corrections to the calculated p-values and the name of the chosen correction is reported in the Paired T-Test (P-Value) column header with its corrected significance value.

The user can also select the option Show only Quantitatively Analyzed Samples in Samples View to hide from view the Quant Samples not selected for the test.

Figure 9-11: Quantitative Samples View, Paired T-Test *carried out for data*



Repeated Measures Analysis of Variance (rANOVA)

The Repeated Measures Analysis of Variance test (rANOVA) is a parametric statistical hypothesis test for assessing whether the population means of repeated measurements differ when the measurements have been taken under at least three conditions (e.g., time points). It is a parametric alternative to the Friedman test, and may be used when the data being analyzed is normally distributed and has equal variances across the categories.

By default, when the Quantitative Testing dialog box first opens, all available Quant Samples in all Quantitative Categories for the experiment are selected for testing and rANOVA will be the default option. However, if the user's experiment contains only two quantitative sample categories, then the rANOVA option will not be available.

After the user selects the rANOVA, and then clicks Apply, the Quantitative Samples View opens showing the experimental data. The rANOVA p-values are displayed in a rANOVA (P-Value) column. The statistically significant values, determined by the selected Significant

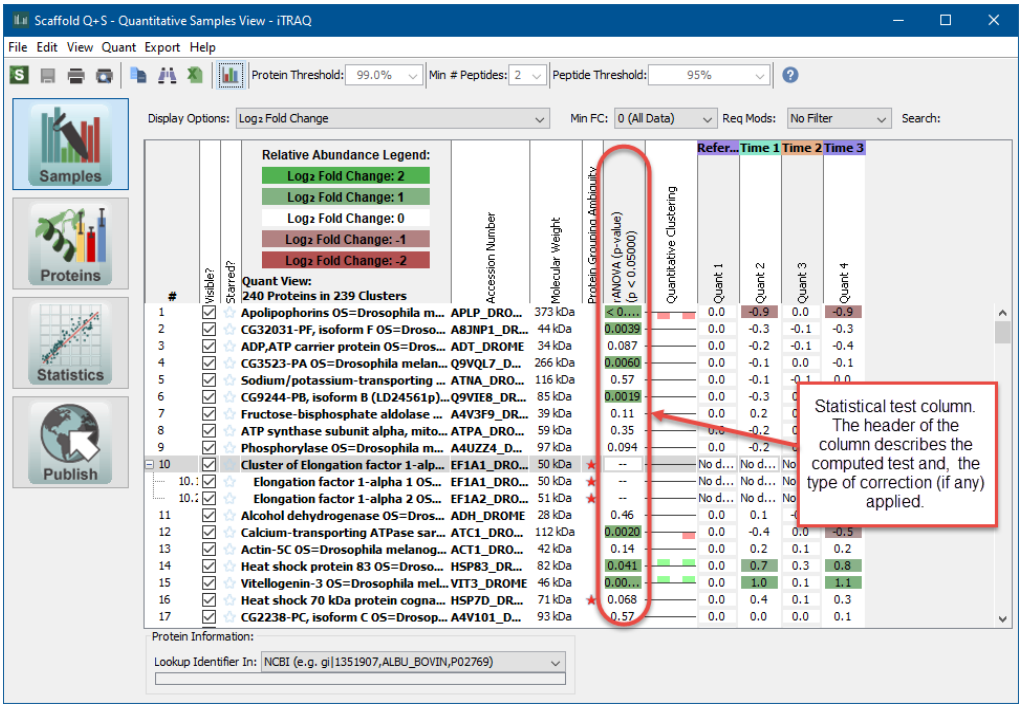
-

Level are highlighted in green when the Test assumptions are met or flagged in orange when otherwise, see Coloring of Test results.

The user has always the option of applying a Multiple Test Corrections to the calculated p-values and the name of the chosen correction is reported in the rANOVA (P-Value) column header with its corrected significance value.

The user can also select the option Show only Quantitatively Analyzed Samples in Samples View to hide from view the Quant Samples not selected for the test.

Figure 9-12: Quantitative Samples View, rANOVA *carried out for data*



-

Common Statistical Test Options

Regardless of the Experiment type, Calculation Type (median or mean) and number of Quantitative Categories selected the quantitative setup dialog always includes the following options:

- No Test Applied
- Permutation Test
- Multiple Test Correction and Significance Level
- Fold Change by Category pane
- Show only Quantitatively Analyzed Samples in Samples View

Details about how Scaffold Q+ or Scaffold Q+S compute the inferential tests are available in section Blocking and inferential tests.

No Test Applied

The user selects No Test Applied to return the Quantitative Samples View to its default display after he/she has carried out quantitative testing of the experimental data. The user may also select this option if he/she wishes to only view a Fold Change by Category.

Permutation Test

The Permutation test is a quantitative assessment of whether the protein abundance is differentially expressed between the different Quantitative Categories. The test assumes that there is no differential expression between the categories when the differential expression is calculated using the F-statistic and then calculates a p-value to determine the likelihood that the assumption is wrong. The approach that the Permutation Test uses to calculate the p-value is based on building a reference histogram of F-statistic values. This reference histogram is made by permutating the measurements between the Quantitative Categories.

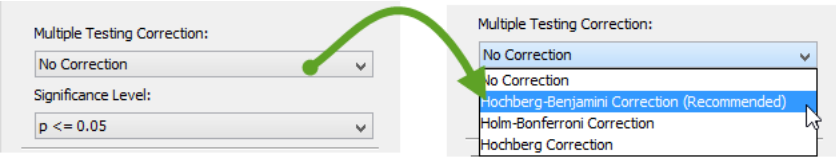
For information about how the Q+ Quantitation Module calculates randomized permutation tests, see https://proteomesoftware.com/statistics/permutation-test.

Multiple Test Correction and Significance Level

In the Quantitative Analysis Setup below the list of available statistical tests, there are two pull down lists: the Multiples Testing Correction and the Significance Level that allow the user to set a significance level for the selected inference test and choose methods to control the familywise error rate.

-

Figure 9-13: Multiple Test Correction and Significance Level lists



When considering a set of statistical inferences simultaneously and doing multiple comparisons the risk of making one or more false discoveries or a Type I error grows quite quickly. In these cases it is common to adjust p-values for the number of hypothesis tests performed. There are many different methods that provide a way to perform this adjustment. A common one is to control the familywise error rate, which is defined as the probability of making Type I errors.

One of the initial and still quite common methods used to control this error is provided by the Bonferroni correction where the significance level α for an individual test is found by dividing the familywise error rate (usually 0.05) by the number of performed tests. Thus when doing 100 statistical tests, the α level for an individual test would be 0.05/ 100=0.0005, and only individual tests with P<0.0005 would be considered significant.

The Bonferroni approach is a fairly conservative one and for a very large number of independent comparisons it may lead to a high rate of false negatives.

To address this issue Scaffold provides three different types of corrections:

- Benjamini-Hochberg correction (recommended)
- Bonferroni Correction
- Holm-Bonferroni correction - a step down method see Control FWER with Hochberg's step-up and Holm's step-down
- Hochberg correction - a step-up method Control FWER with Hochberg's step-up and Holm's step-down

Benjamini-Hochberg correction (recommended)

This method of controlling the error rate in multiple experiments is particularly useful in high-dimensional type of experiments where a more common goal is to identify as many true positive findings as possible, while incurring a relatively low number of false positives. The false discovery rate (FDR) is designed to quantify this type of trade-off, making it particularly useful for performing many hypothesis tests on high-dimensional data sets.

Scaffold computes the FDR using the Benjamini-Hochberg procedure as developed in the original paper[1].

Bonferroni Correction

The Bonferroni Correction is a statistical correction factor that is used to counteract the problem of multiple comparisons. It is considered the simplest and most conservative

1. Benjamini Y. and Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing Journal of the Royal Statistical Society, Series B (Methodological), 1995, Vol.57, No. 1: 289-300

-

method to control the rate of false positives, which in the case of Scaffold Q+ and Scaffold Q+S, correlates to establishing which proteins are truly differentially expressed in the different quantitative sample categories. Be aware, however, that the Bonferroni Correction can be overly stringent for some data and might result in some actual cases of differential expression being excluded from your results. To see the effect of the Bonferroni Correction on your experimental data, first carry out a Mann Whitney test and/or Kruskal-Wallis test for your experimental data without the Bonferroni Correction factor selected and review the data carefully. Next, select the Bonferroni Correction factor and rerun the tests. You should then compare the results with the correction factor selected to those generated without the correction factor selected to determine if the correction factor is appropriate for your data.

Control FWER with Hochberg's step-up and Holm's step-down

There are various methods described in the literature that control the Familywise error rate (FWER) using less conservative corrections than the Bonferroni one but are still based on the Bonferroni inequality. These methods are usually quite appropriate to control the FWER in control trial experiments in which a limited number of comparisons are of interest and where the use of the False Discovery Rate is inappropriate. In these cases the corrections guard against any false positive occurring.

For more information on how the two methods are developed in Scaffold please see the Techniques to Control the Familywise Error Rate.

Fold Change by Category pane

The Fold Change by category pane, included in the Quantitative Analysis Setup dialog, provides a way of comparing ratios between two categories other than the reference.

The pane includes:

- **Show Fold Change by Category check-box** - when checked a Fold change by Category column is added to the Quantitative Samples table even when a Statistical test has been computed.
- **Fold Change Reference Category: pull-down list** - Allows the selection of the denominator used to calculate the Ratio.

This option is available for selection only when the user adds samples belonging to only two categories to the **Selected Samples** table in the Quantitative Analysis Setup dialog.

The computation of Fold Change by Category operates on the same values used for statistical testing. In particular the definition of Fold Change by Category depends on the Analysis Type (Intensity-based or Ratio-based). The user specifies two categories to compare, let's say, Category A and Category B.

- In Ratio-Based mode, the Log_2 Ratios at the spectrum-level within each category are rolled up to the Blocking Level. The Fold Change by Category is defined as:

  Average(Log2 ratios from B) – Average(Log2 ratios from A).

- In Intensity-Based mode, Log_2 intensities at the spectrum-level are rolled up to the Blocking Level, and Fold Change by Category is defined as:

-

Average(Log2 intensities from B) – Average(Log2 intensities from A).

Note that ratios in log-space are expressed as differences of logged values, and that all averaging is done with Kernel Density Average where average can be mean or median as per the user's preferences.

The Fold Change is particularly useful when analyzing Super SILAC data, where Super SILAC can be seen as an extension of Spike-in SILAC.

Show only Quantitatively Analyzed Samples in Samples View

When selected, in the Quantitative Samples the columns of Quant Samples that are not included in the test are hidden from view.

Blocking and inferential tests

Depending on the selected Blocking Level described in section "Quantification Setup dialog", the data used to calculate the inferential tests varies. For example, if the selected **Blocking Level:** is "Unique peptides", the program uses the Log_2 Normalized Intensity values rolled up at the peptides level for computing the test. The data can be seen listed in the peptides table located in the Proteins view when the Log_2 Normalized Intensity option is selected from the Display Options pull down menu. On the other hand when the selected **Blocking Level:** is "All Spectra", Q+ Quantitation Module calculates the tests using the Log_2 Normalized Intensity values appearing in the Spectra table for each peptide included in the peptide list and in a category.

Note that all observations are considered independent, see Appendix "Blocking Level example," on page 154.

Coloring of Test results

When a statistical test has been applied the results shown in the p-value column are colored as follows:

- If the result is not statistically significant (above the adjusted alpha-level) the cell background is white.
- If the result is statistically significant (below the adjusted alpha-level), we color the cell green, unless any of the assumption-testing tests give a significant result in which case we color the cell orange.

Mousing-over the cell will give an extended tool-tip with information about the calculation of the test itself together with information about what test assumptions have been checked and those results.

The results of the statistical assumption checking also appear in the Statistical Analysis Export.

-

-

Chapter 10
Quantitative Publish View

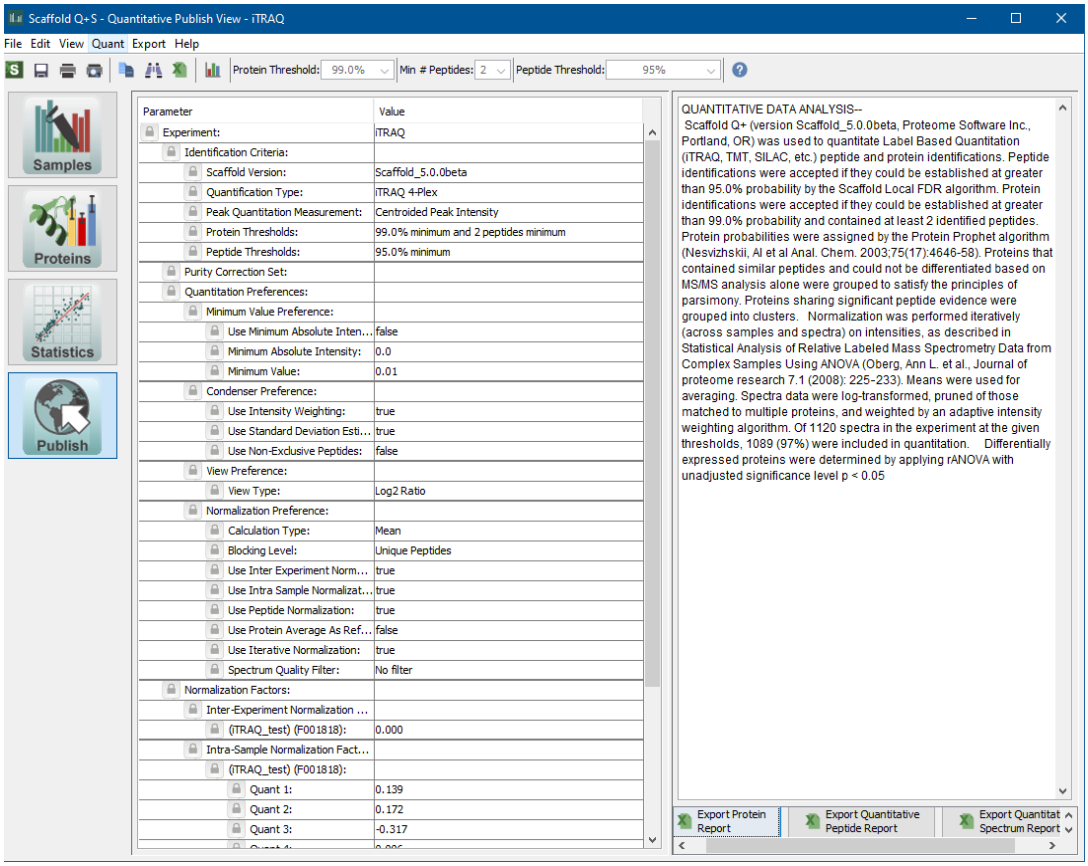
The Q+ Quantitation Module Publish View displays the information about the user's quantitative analysis data that is required for publication in a number of proteomics journals. The left pane of the view displays the following information in a structured list view:

- The type of quantitation used (iTRAQ, TMT, or isotopically labeled).
- The peptide and quantitation option used.
- Normalization information.

The right pane of the view provides the same information in a narrative format. You can use this narrative as a rough draft for the methods section of a journal article. At the bottom of the right pane are two options—Export Protein Supplemental Material and Export Peptide Supplemental Material. You can select these options to generate the Protein Supplemental report and Peptide Supplemental report, respectively, which can provide critical supplemental data for supporting publication in a proteomics journal.

-

Figure 10-1: Publish View



Chapter 11
Reports

A variety of reports are available in Scaffold Q+ and Scaffold Q+S to assist the user in interpreting and working with quantitative analysis data. All the reports are available from the Export option on the Scaffold Q+ and Scaffold Q+S main menu. Every report is saved in a predefined format, and in the same directory as its quantitative analysis data. The user cannot change the report format, but can always select a different location in which to save the report. When user saves a Scaffold QuantML report or ProtXML report, he/she must provide a name for the report. When the user saves an Excel report, a default name in the format <Report Name><Scaffold File name> is provided for the report, but the user can always change this value. Finally, the user can open and view any Excel report such as the Publication report in Excel or another spreadsheet application, but the user might need to specify that the report file is a tab-delimited file to do so.

Export commands:

Scaffold QuantML Report...

When the user generates the Scaffold QuantML report, the protein, peptide identification data and reporter or MS1 ion quantitation data is exported to a ScaffoldQuantML file, which has a file format similar to mzIdentML. The spectrum information is also exported as MGF files. The user can load the exported information into Scaffold PTM for protein translation modifications analysis.

For more information, go to http://www.proteomesoftware.com/Scaffold_PTM/Scaffold_PTM.html.

Protein Quantitation XML Report...

This report provides the quantitative information related to the protein expression utilized for protein normalized quantitative PTM experiments analyzed in Scaffold PTM. When selected the command opens a file browser so the user can locate a directory where to save the produced XML file.

ProtXML...

Exports all quantitative data in the protXML format, which is an open XML file format for the storage of data at the raw spectral data, peptide, and protein levels. This format enables uniform analysis and exchange of MS/MS data generated from a variety of different instruments, and assigned peptides using a variety of different database search programs.

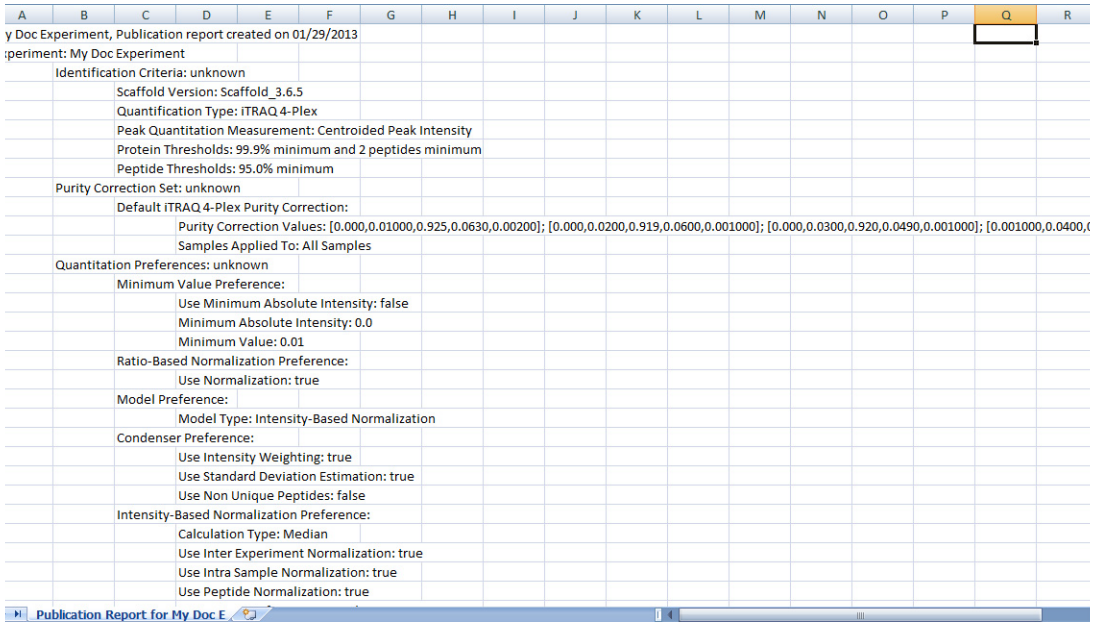
See Molecular Systems Biology, 1:2005.0017 for more information.

Export > To Excel

Publication report...

The Publication report lists the data analysis information that is required for publications in many proteomics journals. The report is a copy of the “The Q+ Quantitation Module Publish View displays the information about the user’s quantitative analysis data that is required for publication in a number of proteomics journals. The left pane of the view displays the following information in a structured list view:” on page 141.

Figure 11-1: Publication report example



y Doc Experiment, Publication report created on 01/29/2013
:periment: My Doc Experiment
Identification Criteria: unknown
Scaffold Version: Scaffold_3.6.5
Quantification Type: iTRAQ 4-Plex
Peak Quantitation Measurement: Centroided Peak Intensity
Protein Thresholds: 99.9% minimum and 2 peptides minimum
Peptide Thresholds: 95.0% minimum
Purity Correction Set: unknown
Default iTRAQ 4-Plex Purity Correction:
Purity Correction Values: [0.000,0.01000,0.925,0.0630,0.00200]; [0.000,0.0200,0.919,0.0600,0.001000]; [0.000,0.0300,0.920,0.0490,0.001000]; [0.001000,0.0400,(
Samples Applied To: All Samples
Quantitation Preferences: unknown
Minimum Value Preference:
Use Minimum Absolute Intensity: false
Minimum Absolute Intensity: 0.0
Minimum Value: 0.01
Ratio-Based Normalization Preference:
Use Normalization: true
Model Preference:
Model Type: Intensity-Based Normalization
Condenser Preference:
Use Intensity Weighting: true
Use Standard Deviation Estimation: true
Use Non Unique Peptides: false
Intensity-Based Normalization Preference:
Calculation Type: Median
Use Inter Experiment Normalization: true
Use Intra Sample Normalization: true
Use Peptide Normalization: true
Publication Report for My Doc E

Samples report...

The Samples report mimics the Samples View. The report header lines identify the data and how it was created, which is the same information that is contained in the Publication report....

If the selected display type is log2 fold change, the standard deviation is also shown. If the selected display type is fold change ratio, the geometric standard deviation is shown.

Figure 11-2: Samples report, report columns

#
Visible?
Starred?
Identified Proteins (607)
Accession Number
Molecular Weight
Protein Grouping Ambigui
Quantitative Variance
Taxonomy
Light
Medium
Heavy

Samples report with Isoforms...

The Samples Report with Isoforms is identical to the Samples report...but it also includes the expanded protein families in the Identified Proteins list.

Quantitative Peptide report...

The Peptide report details all the peptides that pass the current filter settings. The report header lines identify the data and how it was created, which is the same information that is contained in the Publication report.... There is a single report entry for each peptide in each sample, regardless of the number of spectra that matched to a peptide. For example, if there

are three samples and each sample contains 100 peptides, and three spectra matched to one of the peptides, the report still contains only 300 rows. You can view the Peptide report in Excel or another spreadsheet application.

Figure 11-3: Quantitative Peptide report, report columns

Quantitative Spectrum report...

The Spectrum report not only contains the same information as the Quantitative Peptide report..., but also details all the spectra that pass the current peptide and protein filter settings. There is a single report entry for each spectrum that matched peptide within the filter limit. The report header lines identify the data and how it was created, which is the same information that is contained in the Publication report....

Figure 11-4: Quantitative Spectrum report, spectra-specific report columns

Raw Data report...

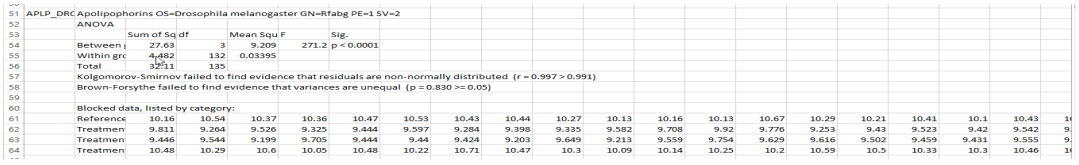
The Raw Data report contains the raw intensity data values for every protein in each sample as well as the normalized values that are used for the fold change and other calculations.

Figure 11-5: Raw Data report, report columns

Statistical Analysis report...

This report is created only when the user applies a statistical test. For each protein in the Quantitative Samples list, the report provides the summary for the selected statistical test and the assumption testing results. The data used to compute the test statistics is also included listed by category.

Figure 11-6: Statistical Analysis report: example for one protein



Protein Accession Number report...

The Protein Accession Number report lists all the proteins that pass the current filter settings. It also lists the name of the protein and the database that you used for searching your experimental data. There is a report entry for each protein in each sample. For example, if three samples have the same 12 proteins, then there will be 36 lines in the report.

Figure 11-7: Protein Accession Number report, report columns

Current View report...

The Current View report contains the information that is displayed in the current view. This report is applicable for the “Quantitative Samples View” on page 72, the “The Scaffold Q+ Quantitative Proteins View provides an overview of the peptides quantitative detection for a selected protein group.” on page 84, and the “Quantitative Publish View” on page 141.

-

Appendix

This Appendix provides information about the algorithms utilized in Scaffold Q+ and Scaffold Q+S.

- Appendix A.Terminology
- Appendix B.Normalization
- Appendix C.Blocking Level example
- Appendix D.Context menu Commands
- Appendix E. Experimental Design Type Default Settings
- Appendix F.Kernel Density Average
- Appendix G.Techniques to Control the Familywise Error Rate
- Appendix H.Coefficient of Variation

-

Appendix A.Terminology

Blocking

When groups of experimental units are similar, it is often a good idea to gather them together into blocks. By blocking the variability attributable to the differences between the blocks is isolated so that the differences caused by the treatments appear clearer.

BioSample

In Scaffold it refers to the data associated with a biological sample like a drop of blood or a biopsy, for example. A BioSample may contain one or more MS Sample.

Contingency table

In statistics, a contingency table (also referred to as cross tabulation or cross tab) is a type of table in a matrix format that displays the (multivariate) Frequency Table or distribution of the variables.

Experiment (a statistical definition)

An experiment manipulates factor levels to create treatments, randomly assigns subjects to these treatments levels, and compares the responses of the subject groups across treatment levels.

Experimental Design

The design of experiments (or experimental design) is the design of any task that aims to describe or explain the variation of information under conditions that are hypothesized to reflect the variation. The term is generally associated with true experiments in which the design introduces conditions that directly affect the variation, but may also refer to the design of quasi-experiments, in which natural conditions that influence the variation are selected for observation. In its simplest form, an experiment aims at predicting the outcome by introducing a change of the preconditions, which is reflected in a variable called the predictor. The change in the predictor is generally hypothesized to result in a change in the second variable, hence called the outcome variable. Experimental design involves not only the selection of suitable predictors and outcomes, but planning the delivery of the experiment under statistically optimal conditions given the constraints of available resources.(definition derived from Wikipedia)

Factor

A variable whose levels are manipulated by experimenters. It is a single biological or technical parameter that the user controls in an experiment like gender, diet, environment, stimulus, age and so on. Experiments attempt to discover the effects that differences in factor levels may have on the responses of the experimental units.

Frequency Table

In statistics, a frequency table is a table that displays the frequency of various outcomes in a sample. Each entry in the table contains the frequency or count of the occurrences of values within a particular group or interval, and in this way, the table summarizes the

-

distribution of values in the sample. Bivariate joint frequency distributions are often presented as (two-way) Contingency tables.

MS Sample

In Scaffold it designates the search data results from a mass spectrometry run. Biological samples might be separated using techniques such as 2D gels or liquid chromatography. The different fractions processed by a mass spectrometer and the resulting data searched for identification are part of what Scaffold calls a MS Sample. This means that BioSamples might include one or more MS Samples.

SILAC

Stable isotope labeling with amino acids in cell culture (SILAC) is a simple and straightforward approach for in vivo incorporation of a label into proteins for mass spectrometry (MS)-based quantitative proteomics. SILAC relies on metabolic incorporation of a given 'light' or 'heavy' form of the amino acid into the proteins. The method relies on the incorporation of amino acids with substituted stable isotopic nuclei (e.g. deuterium, 13C, 15N). Thus in an experiment, two cell populations are grown in culture media that are identical except that one of them contains a 'light' and the other a 'heavy' form of a particular amino acid (e.g. 12C and 13C labeled L-lysine, respectively). When the labeled analog of an amino acid is supplied to cells in culture instead of the natural amino acid, it is incorporated into all newly synthesized proteins.

Spike-in SILAC

Classical SILAC experiments are designed by differentially labeling different cell lines with combinations of heavy amino acids, processing the lysates together and carrying out mass spectrometry analysis. Another way to use SILAC technique is to generate a heavy labeled sample and spike it in known amounts into the experimental samples which have been grown in normal media. This serves as an internal control. Comparisons between two or more experimental states can then be made by comparing the light/heavy ratio for the same peptide in each case.

Super SILAC

Previously the use of SILAC was limited to cells grown in culture. However an extension of the spike-in approach called “Super-SILAC” has made it possible to use SILAC for tissues. As tissue proteome is much more diverse than cell line Proteome a spike in from many SILAC labeled cell lines can be used to create a “Super-SILAC” mix which can serve as an internal control. A known amount of this mix can be spiked into the tissue lysates to be compared.

Treatment

The process, intervention or other controlled circumstance applied to randomly assigned experimental units. Treatments are the different levels of a single factor or are made up of combinations of levels of two or more factors.

Bootstrapping

Statistical inference is usually based on the sampling distributions of a number of sample statistics. *Bootstrapping* uses a single sample to determine an approximate sampling

-

distribution. Typically, a sampling distribution relies on many random samples selected from the population of interest. The bootstrap distribution of a statistic instead collects its values from many re-samples with replacement and gives information about the original sampling distribution from which the samples, that were used for the bootstrap, were initially drawn.

Quant Category

Main structural unit used to organize an experiment. It may contain one or more Quant Samples, which can be either technical replicates or biological replicates. A Quant Category gets defined when first accessing the Scaffold PTM and can be readjusted according to need through the menu command: **Quant > Update Experimental Design**.

Quant Sample

Labeled specimen. A Quant Sample is assigned to a category when first accessing the Scaffold PTM. It is initially assigned a default name that the user can adjust either when setting up the experiment the first time or through the menu command: **Quant > Update Experimental Design**.

-

Appendix B.Normalization

Quantitative Raw data imported into the Scaffold PTM are processed for Bias removal with two possible different normalizations methods:

- Intensity-Based Normalization
- Ratio-Based Normalization

Intensity-Based Normalization

The method used is a non-parametric version of a statistical model proposed by Ann Oberg and others for analyzing iTRAQ data[1]. A technical description of the different steps included in the normalization algorithm developed in Scaffold Q+ or Scaffold Q+S is provided below.

Pre-Processing

1. The raw intensity data is acquired from the spectra and purity-corrected as appropriate.
2. Spectra not assigned exclusively to one protein are discarded (unless the user has indicated otherwise).
3. The data is transformed by applying a logarithm (base 2).
4. Within each MS-Sample a “missing value” is assigned as the larger between (a) the minimum logged intensity acquired and (b) the value whose z-score is -4 for the distribution of all logged values in the MS-Sample. These missing values are applied to all intensities which either had raw value zero, or fall below the “Minimum Dynamic Range.”
5. If the option is selected, spectra with missing values in the reference channel are removed.

Iterative Median Polish

A version of Tukey’s median polish is applied iteratively to normalized the data. In what follows, all computations are on the logged data, and “average” denotes either median or mean depending on the mode the user has selected. (Median is the default mode.) The normalization has four steps:

1. Inter-Sample Normalization: A normalization factor consisting of the global average minus the within-MS-Sample average is added to each data point.
2. Intra-Sample Normalization: A normalization factor consisting of the within-MS-Sample average minus the within-channel average is added to each data point.

1. Elizabeth G. Hill, John H. Schwacke, Susana Comte-Walters, Elizabeth H. Slate, Ann L. Oberg, Jeanette E. Eckel-Passow, Terry M. Therneau, and Kevin L. Schey, A STATISTICAL MODEL FOR iTRAQ DATA ANALYSIS, J Proteome Res. 2008 Aug; 7(8): 3091–3101. Published online 2008 Jun 26. doi: 10.1021/pr070520u

-

3. Peptide/Spectrum Normalization: For each protein, the averages across the values in each spectrum are brought into alignment by adding a per-spectrum normalization factor (average of these averages minus the particular spectrum's average).
4. Intensity Weighting: A weight is assigned to each spectrum based on a t-statistic derived from percent deviations from channel averages.

Steps (1)-(4) are repeated three times.

Post-Processing

A standard deviation estimate is derived for each spectrum, based on smoothed within-protein deviations of spectral values from averages binned by total intensity.

The weight of each spectrum is divided by the total number of spectra matched to its peptide, providing a form of intermediate peptide-level averaging when subsequently computing protein-level values.

Ratio-Based Normalization

Ratio-based normalization is used for stable isotope labeling experiments, in which each measurement consists of a pair of intensity values. In this type of quantitation, the intensities may not be consistent between spectra or between samples, but the ratios of the intensities of the different labeled states associated with a specific spectrum are considered more reliable.

In Scaffold Q+S, ratio-based normalization is only supported in mode with individual spectrum reference. Ratio based normalization operates on Log_2 intensity ratios rather than on the intensity values themselves.

Missing values are imputed, see Pre-Processing, then the ratio of each value to the reference value within that spectrum is calculated. These ratios are logged transformed, and for each Quant Sample the average log-ratio is determined.

The average Log_2 Ratio is then subtracted from each Log_2 Ratio in the Quant Sample with the result that the average Log_2 Ratio for each Quant Sample is zero. This allows Quant Samples from different MS Samples to be combined without further normalization.
For more details see Appendix "Normalization" on page 159 and "Advancing Statistical Analysis of Multiplexed MS/MS Quantitative Data with Scaffold Q+ (ASMS 2012)".

-

Appendix C.Blocking Level example

When using the Intensity-based normalization scheme, the selection of the Blocking level from the Quantitation set up dialog affects the way Q+ Quantitation Module computes inference tests. To show how the calculations are performed we show an hypothetical experiment as an example and indicate hereafter which group of intensities are used to calculate the test.

Let's consider a 4plex iTRAQ experiment, each sample (channel) equals a mouse under a treatment. The treatment has two levels (treatment -> T, control -> C). The corresponding two group comparison would be T(113,114) vs. C(115,116). Let's say that we choose to analyze the data using a mean as the calculation type so we apply a T-Test to asses the differences between the two groups or categories. Depending on the selection of the Uniqueness model the program will use different groups of intensities to calculate the test.

For a single protein the original matrix of observations can be summarized as shown in Table 1, which shows 3 peptides belonging to an hypothetical protein identified in the experiment. Peptide 1 and 2 have 6 different spectra assigned while peptide 3 has 4. As an example only two different charge states and two possible combination of modifications indicated a PT1 and PTM2 are considered here. The intensity of each iTRAQ peak is generically indicated by y. This does not mean that the intensities are identical.

Table 1: Blocking Level All spectra

Peptide	Charge	PTM	ID	113	114	115	116
pep1	+2	PTM1	1	y	y	y	y
pep1	+2	PTM1	2	y	y	y	y
pep1	+2	PTM2	3	y	y	y	y
pep1	+3	PTM1	4	y	y	y	y
pep1	+3	PTM2	5	y	y	y	y
pep1	+3	PTM2	6	y	y	y	y
pep2	+2	PTM1	7	y	y	y	y
pep2	+2	PTM1	8	y	y	y	y
pep2	+2	PTM2	9	y	y	y	y
pep2	+3	PTM1	10	y	y	y	y
pep2	+3	PTM2	11	y	y	y	y
pep2	+3	PTM2	12	y	y	y	y
pep3	+2	PTM1	13	y	y	y	y
pep3	+2	PTM1	14	y	y	y	y
pep3	+3	PTM1	15	y	y	y	y
pep3	+3	PTM1	16	y	y	y	y

When the Blocking Level is set to All Spectra, the T-Test is calculated using the $\log_2$ of all

intensities included in Table 1. The data used for the calculation in Scaffold Q+S would be listed in the Spectra table appearing in the Spectra pane when the selected Display Options is Log_2 Normalized Intensity.

When the Blocking Level is set to **Unique Spectra** the data shown in Table 1 is collapsed to Table 2 for calculation purposes; the values of the intensities y' are the mean, or median if the calculation type selected is Median, of the intensities in each of the spectra that have same charge states and PTMs.

Table 2: Blocking Level Unique Spectra

Peptide	Charge	PTM	ID	113	114	115	116
pep1	+2	PTM1	1,2	y'	y'	y'	y'
pep1	+2	PTM2	3	y'	y'	y'	y'
pep1	+3	PTM1	4	y'	y'	y'	y'
pep1	+3	PTM2	5,6	y'	y'	y'	y'
pep2	+2	PTM1	7,8	y'	y'	y'	y'
pep2	+2	PTM2	9	y'	y'	y'	y'
pep2	+3	PTM1	10	y'	y'	y'	y'
pep2	+3	PTM2	11,12	y'	y'	y'	y'
pep3	+2	PTM1	13,14	y'	y'	y'	y'
pep3	+3	PTM1	15,16	y'	y'	y'	y'

When the Blocking Level is set to Unique Spectra, the T-Test is calculated using the log_2 of all intensities included in Table 2. The data used for the calculation is not directly shown in the The Scaffold Q+ Quantitative Proteins View provides an overview of the peptides quantitative detection for a selected protein group. but can be inferred from the values appearing both in the Spectra and Peptides panes.

When the Blocking Level is set to **Unique Peptides** the data shown in Table 2 is collapsed to Table 3 for calculation purposes; the values of the intensities y" are the mean, or median if the calculation type selected is Median, of the intensities in each of the spectra that identify the same amino acid sequence regardless of the different charge states or combinations of PTMs.

Table 3: Blocking Level Unique Peptides

Peptide	Charge	PTM	ID	113	114	115	116
pep1	+2,+3	PTM1,2	1,2,3,4,5,6	y"	y"	y"	y"
pep2	+2,+3	PTM1,2	7,8,9,1011,12	y"	y"	y"	y"
pep3	+2,+3	PTM1	13,14,15,16	y"	y"	y"	y"

When the Blocking Level is set to Unique Peptides, the T-Test is calculated using the log_2 of all intensities included in Table 3. The data used for the calculation in Scaffold Q+S would be listed in the Peptides table appearing in the Peptides pane when the selected Display

-

Options is Log_2 Normalized Intensity.

When the Blocking Level is set to **Unique Samples** the data shown in Table 3 is collapsed to Table 4 for calculation purposes; the values of the intensities y’” are the mean, or median if the calculation type selected is Median, of the intensities in each of the peptides belonging to the considered protein.

Table 4: Blocking Level Unique Samples

Peptide	Charge	PTM	ID	113	114	115	116
Protein including pep1, pep2, pep3	+2,+3	PTM1,2	all	y’”	y’”	y’”	y’”

When the Blocking Level is set to Unique Samples, the T-Test is calculated using the log_2 of all intensities included in Table 4. The data used for the calculation in Scaffold Q+S would be listed in the Samples table appearing in the Quantitative Samples View when the selected Display Options is Log_2 Normalized Intensity.

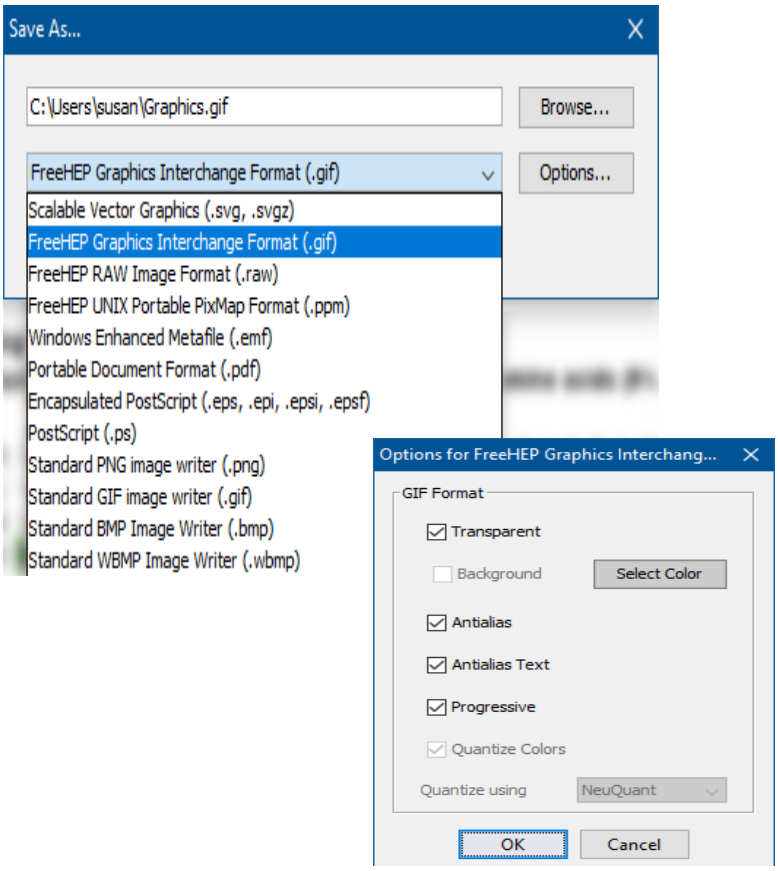
The **Unique Samples** level should be the correct blocking level for analyzing data when the unit of replication in the experiment is a mice and a a mice equals one iTRAQ channel. Unfortunately in order to do so the user would need to have at least 4 samples per category for the t-test to be statistically meaningful. Since it is not easy to plan experiments that include this amount of samples per category, we suggest and have it selected as the default option, to choose Unique Peptides as the preferred blocking level for analyzing the data. When doing so the biological variation, which is what the user is interested in assessing, gets confounded with the peptide technical variation. But, even if the peptide level blocking is in general inadequate to provide correct p-values, it is necessary because in most proteomics data sets very few samples are acquired. So using the peptide blocking is like trading off accuracy for sensitivity and in doing so we are hoping that the value we are using is not that different from the value we are trying to measure.

-

Appendix D.Context menu Commands

- **BLAST Peptide Sequence** - opens http://www.ncbi.nlm.nih.gov/blast/ site for current peptide
- **BLAST protein sequence** - opens http://www.ncbi.nlm.nih.gov/blast/ site for current protein.
- **Copy All Data -** copies all the data listed in the table shown in the current pane to the clipboard.
- **Copy Image -** copies the image of the current view and current pane to the clipboard.
- **Copy Peak List -** copies the pick list of the current spectrum to the clipboard.
- **Copy Protein Sequence** - copies the sequence to the clipboard.
- **Copy publication Sized JPEG** - for publication purposes.
- **Copy Selected Cell** - from the table copies selected cell to the clipboard.
- **Copy Selected row** - from the table copies selected row to the clipboard.
- **Copy WMF/EMF** - copies picture using Windows Meta-file formats which are portable between applications. They contain both vector graphics and bitmap components. Images can be edited and scaled without compromising their resolution.
- **Delete Biological Sample** - a window pops up asking to confirm deletion.
- **Display parent Ions -** toggle function.
- **Display unknown markers** - toggle function
- **Export to Excel** - export information in current tab table
- **Print** - print image of current view and pane.
- **Save as** - provides the option of saving pictures in a large variety of graphical formats.

-



- **Save JPEG Image** - saves image of current view and pane JPEG format.
- **Show Fixed Modifications** - it toggles the function of highlighting fixed modifications along the sequence.
- **Use Amino Acid Finder** - it toggles the activation of the tool tip that shows the peptides along the sequence.
- **Use Peak-finder** - displays the tool tip for the different peaks if checked.
- **Use PPM Masses** - toggle function.
- **Zoom Out** - zoom function.

Appendix E. Experimental Design Type Default Settings

Normalization settings		Multiplex Quantitation			Label Free Quantitation		
		Between-subjects (Independent groups)	Between-Subjects (Common/pooled reference)	Repeated measure/ Time Course	Between-subjects (Independent groups)	Between-Subjects (Common/pooled reference)	Repeated measure/ Time Course
Reference Type	Average Protein Reference	YES	NO	NO	YES	YES	YES
Spectrum Quality Filter	Reference Value required	NO	YES	NO	NO	NO	NO

-

Appendix F.Kernel Density Average

The Q+ Quantitation Module calculates averages of a set of quantitative data points by first creating a Probability Density Function (PDF) using a Kernel Density Estimator (KDE) and then from the distribution estimating either the median or the mean of the dataset depending on the calculation type selected in the The goal of quantitative analysis in Scaffold Q+ and Scaffold Q+S is to detect differential expression of a protein between Quantitative Categories. The quantitative settings that the user specifies in the Quantification Setup dialog determine the type of evidence that is produced to support the presence or absence of differential expression. On the main menu of the Q+ Quantitation Module window, selecting Quant > Quantitative Settings opens the Quantification Setup dialog. dialog.

The Kernel Density Estimator builds the PDF approximating each data point in the set with a Gaussian Kernel 3 Standard Deviations (StD) wide and with an area equivalent to the intensity weight or confidence level associated to each point.

The intensity weight calculation for multiplex data is based on the distribution of the deviations of the raw intensity values from the median protein value for all proteins in the experiment, while for precursor intensity data the weight is represented as a sigmoid function of the form[2]:

$$W(X) = \frac{2}{1 + e^{-X}} - 1$$

Where X is the Log_2 Raw Intensity.

The Standard deviation in both cases is calculated after the normalization is completed.

The values for both the Intensity weighting and the Standard deviation are plotted in the Quantitative Statistics view, see “Intensity Weighting and Error Estimation” on page 106.

Once the PDF is constructed, if the median mode was selected, the value of the point that corresponds to a 50% area under the PDF curve is assigned to the median. When the mean mode is selected, instead, the value of the mean is calculated as follows:

$$\mu = \frac{\sum w_i \cdot X_i}{\sum w_i}$$

2. Alessandra Tiengo, Lorenzo Pasotti, Nicola Barbarini, and Paolo Magni, “PhosphoHunter: An Efficient Software Tool for Phosphopeptide Identification,” Advances in Bioinformatics, vol. 2015, Article ID 382869, 12 pages, 2015. doi:10.1155/2015/382869

-

Where w_i is the weight and X_i is one of the points belonging to the considered set of quantitative data.

-

Appendix G. Techniques to Control the Familywise Error Rate

Scaffold supports different methods to control the familywise error rate, FWER. Among them it supports the Holm's step-down procedure and the Hochberg's step-up procedure, which are developed in the program as described in footnote [3].

The two methods make the same type of comparisons, but Holm starts at the smallest p-value and works down the list until one fails the bound, while Hochberg starts at the largest p-value and works up the list until one passes the bound (and then declares that everything below that passes. Hence the Holm bound is in general more conservative than the Hochberg.

For example, let's suppose we have (m=5) proteins A, B, C, D, E with p-values 0.030, 0.014, 0.013, 0.060, and 0.009 respectively, and want to reject the null hypothesis at alpha = 0.05. Let's sort the p-values and make the following table:

k	p-value	$\alpha/(m+1-k)$	p-value < $\alpha/(m+1-k)$
1	0.009	0.01	yes
2	0.013	0.0125	no
3	0.014	0.0167	yes
4	0.030	0.025	no
5	0.060	0.05	no

- The Holm step-down procedure would start at k=1 and reject $H_0(1)$ but it would stop at k=2 since the p-value is larger than the bound.
- The Hochberg step-up procedure would start at k=5, go to k=4, go to k=3, see that the bound passes and stop, accepting $H_0(1)$, $H_0(2)$, and $H_0(3)$.

The user should be aware of the fact that technically the Hochberg procedure should only be used if the hypothesis tests are independent (which they are certainly not for Fisher's Exact Test, and not usually really for the other tests as well).

3. Y. Huang and Hsu J.C. Hochberg's Step-Up Method: Cutting Corners Off Holm's Step-Down Method Biometrika, 2007, 94,4,pp.965–975. DOI: 10.1093/biomet/asm067

-

Appendix H.Coefficient of Variation

In Scaffold perSPECtives when three or more treatments are selected it is possible to choose the Coefficient of Variation test. In probability theory and statistics, the coefficient of variation (CV) is a normalized measure of the dispersion of a probability distribution. In a single variable setting it is defined as the ratio of the Standard Deviation over the Mean of the distribution.

The CV is typically used to describe the dispersion of the variable independently of its measurement unit. The higher the CV, the greater the dispersion in the variable. It is typically used in place of an ANOVA test when not enough replicates are available to give sufficient statistical power to apply ANOVA.

If, for example, the User selects four treatments- A, B, C, and D - the coefficient of variation outputs how dispersed are the values in respect to their mean. A small coefficient of variation means that the four treatments have values close together compared to their average value. If the coefficient of variation is big, then at least one of the four treatments is different, but it doesn't specify which of them.

Index

N

O

P

Q

R

S

T

U

W