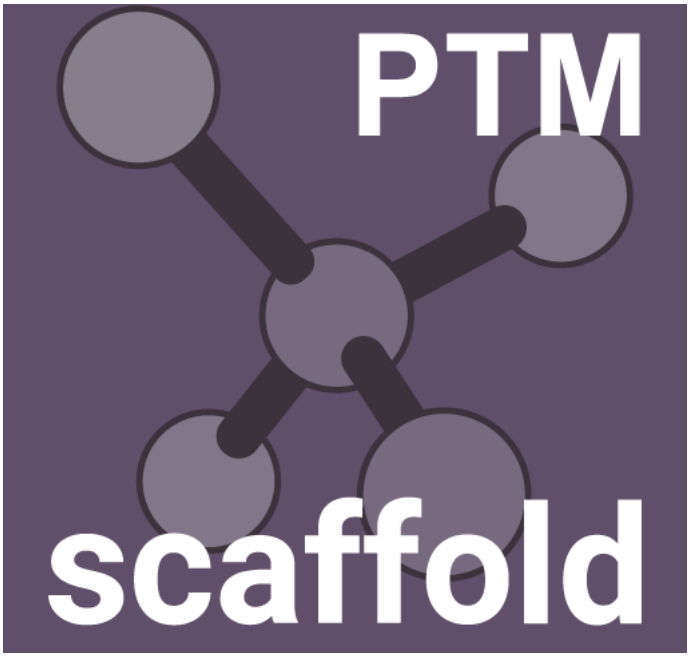


Scaffold PTM

Version 4.0

User's Guide

Release Information

The following release information applies to this version of the *Scaffold PTM*. This document is applicable for Scaffold PTM, Release 4.0 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 Scaffold DIA* 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 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)
1-503-245-4910 (Fax)
www.proteomesoftware.com

Table of Contents

Getting Started with Scaffold PTM

System Requirements

For information about the system requirements for Scaffold PTM, see:

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

Installing Scaffold PTM

Scaffold PTM 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.

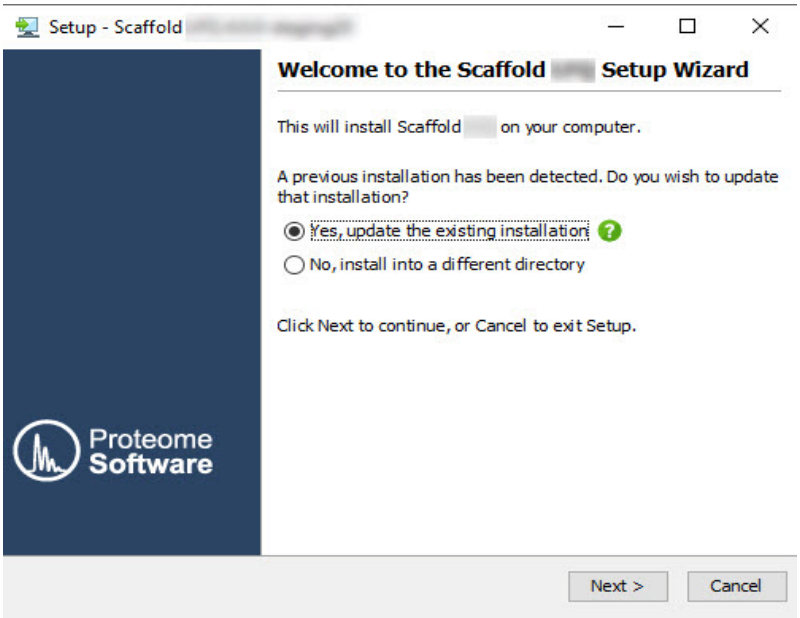


Figure 1-1: Scaffold PTM installation Setup Wizard

3. The installer will then provide you an opportunity to allocate memory to Scaffold PTM. 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 then 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 SPTM files and Scaffold PTM, and the following screen allows creation of desktop icons. Clicking “Next” begins the installation.
5. Finally, Scaffold PTM 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 PTM. The memory setting can be adjusted after installation by selecting the menu option ***Edit > Preferences - Memory tab****. You must close Scaffold PTM and restart the program in order for the new memory setting to take effect.*

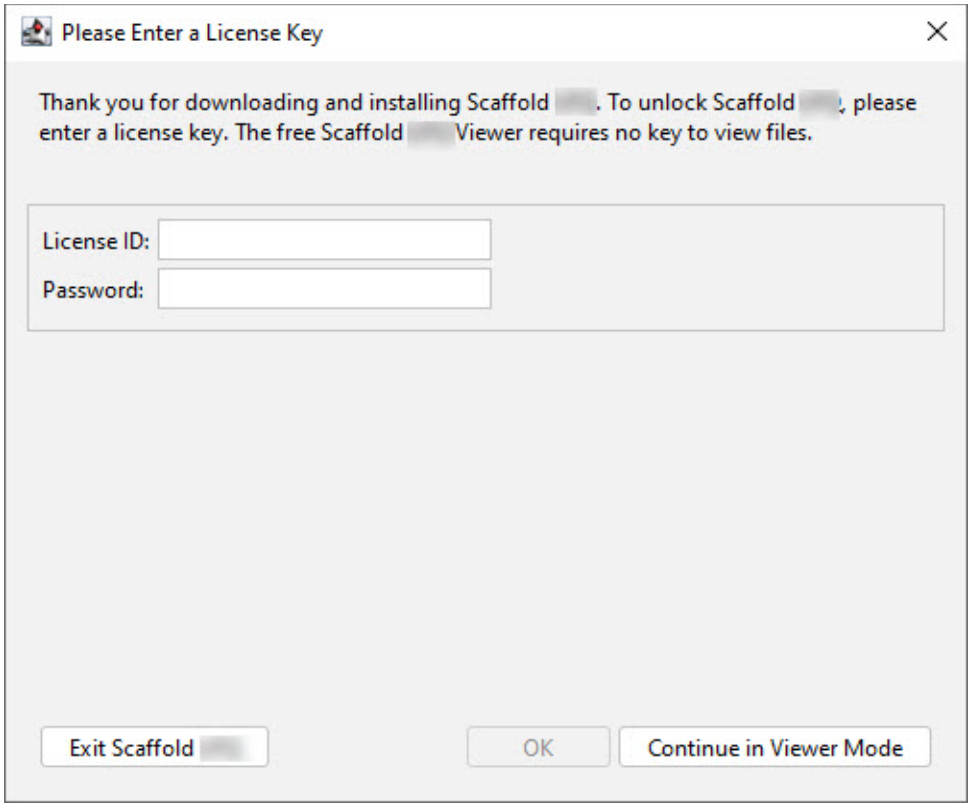
After Scaffold PTM has been installed on a computer, a shortcut icon for the application is placed on the desktop. An option is also available from the Start menu. Double-clicking the desktop icon launches Scaffold PTM, as does, for Windows computers, selecting the option from the Start menu (**Start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

Licensing

The first time Scaffold PTM 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.

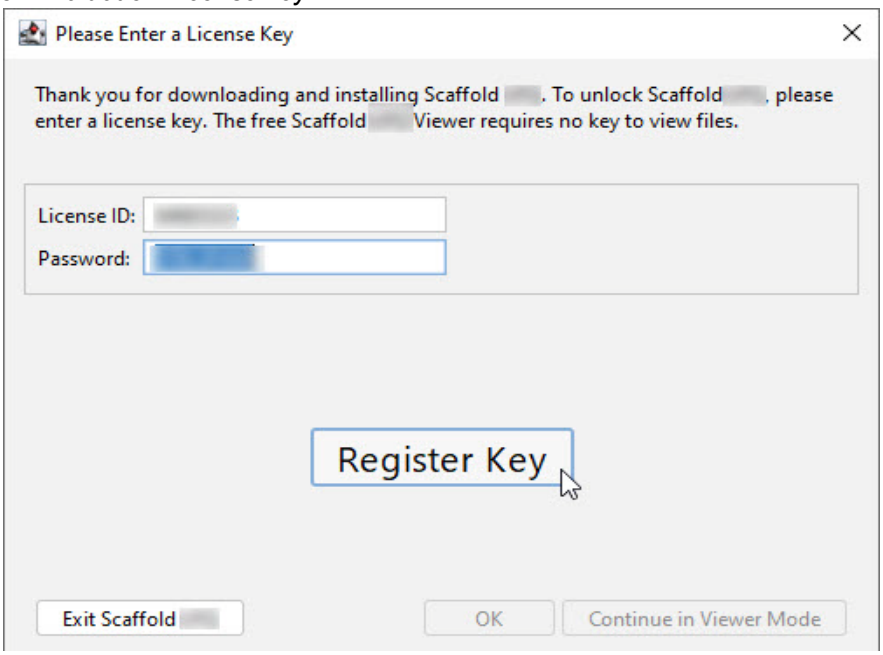
Figure 1-2: Scaffold License Key messages



Two kinds of keys are available to activate the software:

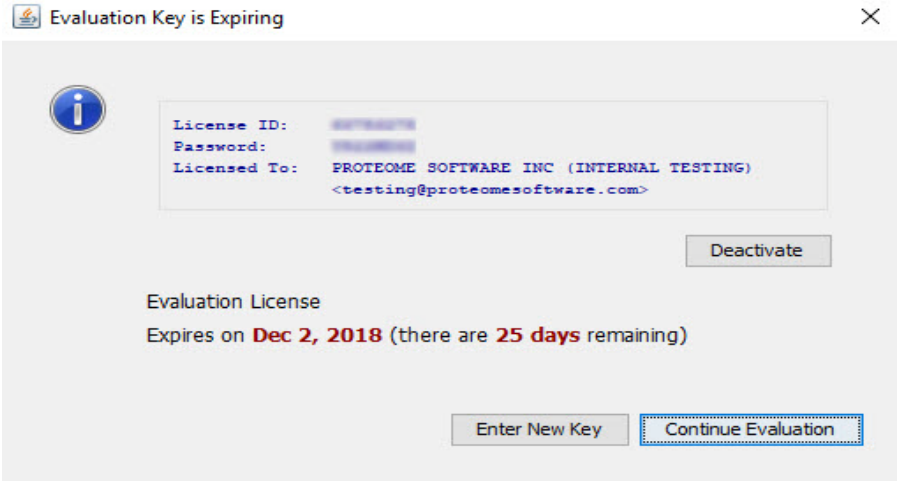
Evaluation key - An Evaluation key is valid for a limited period. A free evaluation key for Scaffold PTM may be obtained through *www.proteomesoftware.com* An evaluation key may be used on two computers. Once the key and password have been copied and pasted into the license key dialog box, a message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

Figure 1-3: Evaluation License key



Every time Scaffold PTM 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 PTM

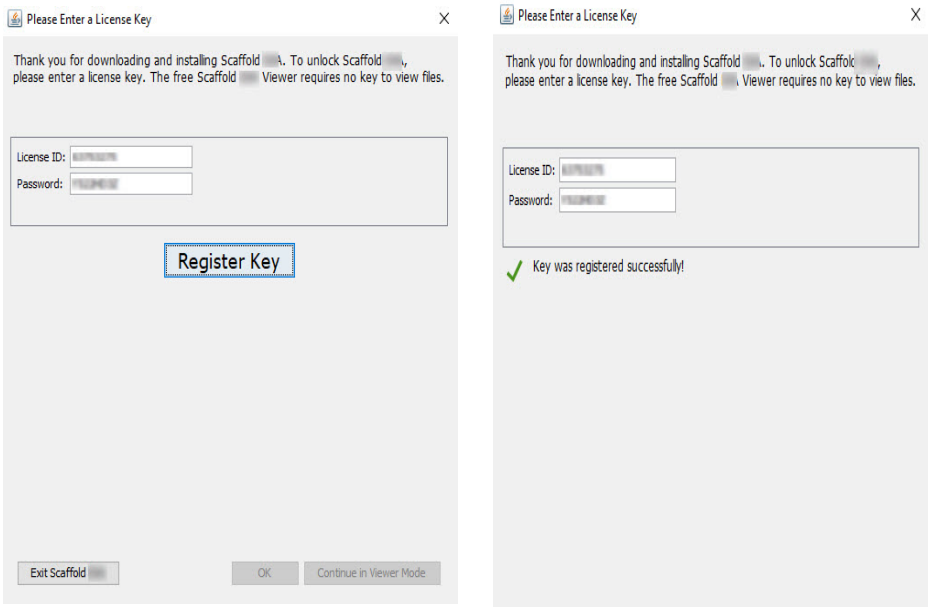


Time-Based License key—a Time-Based License key allows the user to access all features of the software permanently. It only allows upgrades within a certain time limit, however. The time tracks the length of the support contract. Once expired, Scaffold PTM will continue to work beyond the expiration 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 PTM installation to a different computer, see 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 PTM Welcome message opens.

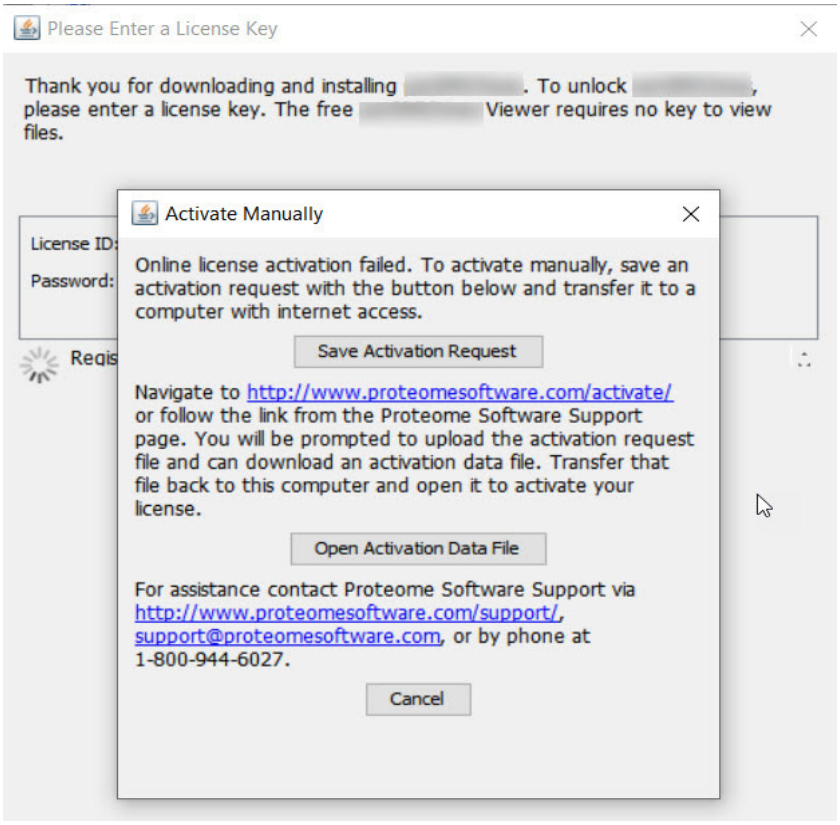
If the user is using an evaluation copy of Scaffold PTM, then an Evaluation message opens, indicating the number of days left in the evaluation period. The user must click OK to close this message and then the Scaffold PTM Welcome message opens.

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

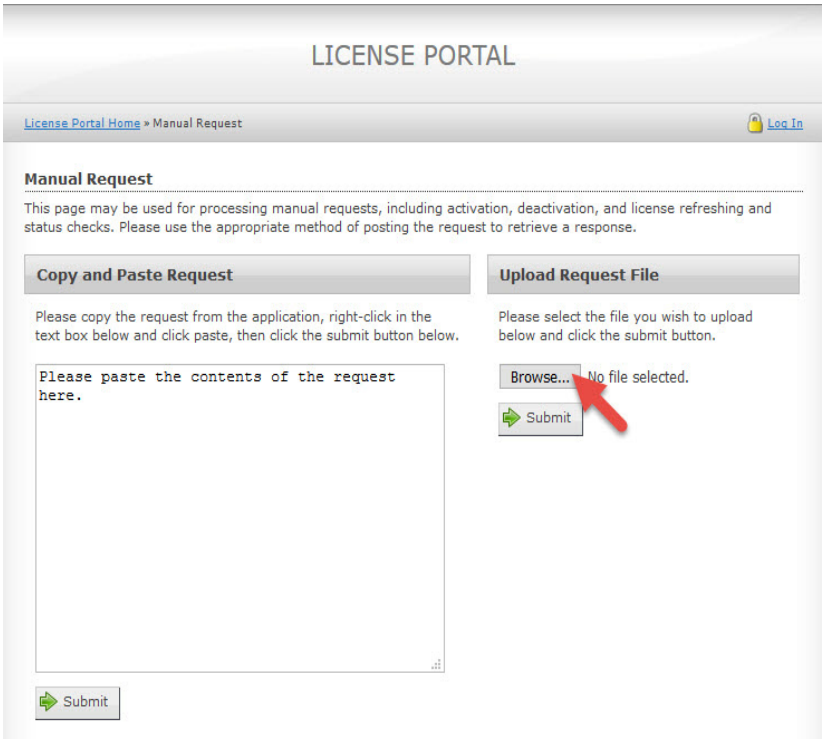
Figure 1-6: Manual or offline activation dialog



To activate Scaffold PTM without an internet connection:

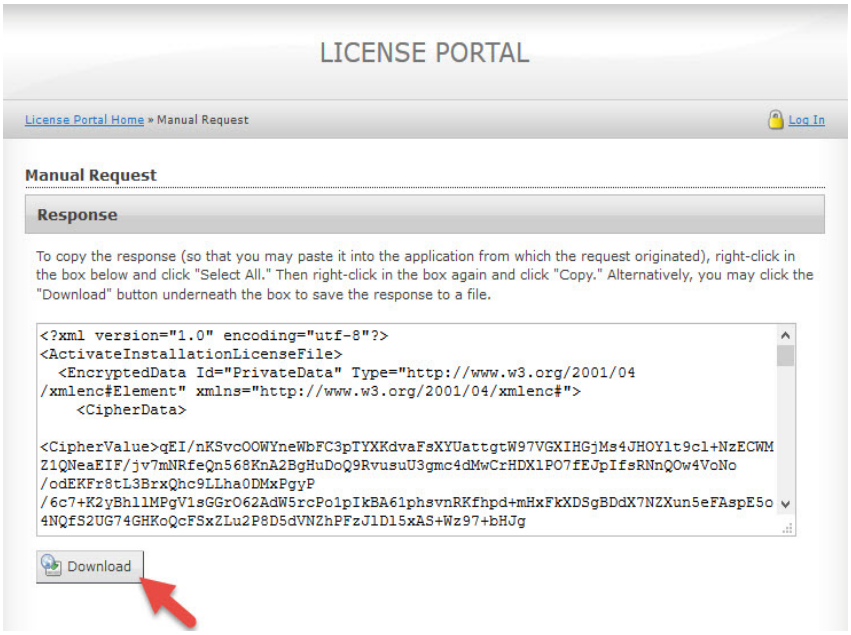
1. First, use the Save Activation Request button to create an activation request file.
2. Transfer this file to a computer with internet access (e.g. using a USB drive).
3. On the connected computer, navigate to*http://www.proteomesoftware.com/activate*This link is also accessible from the Proteome Software 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



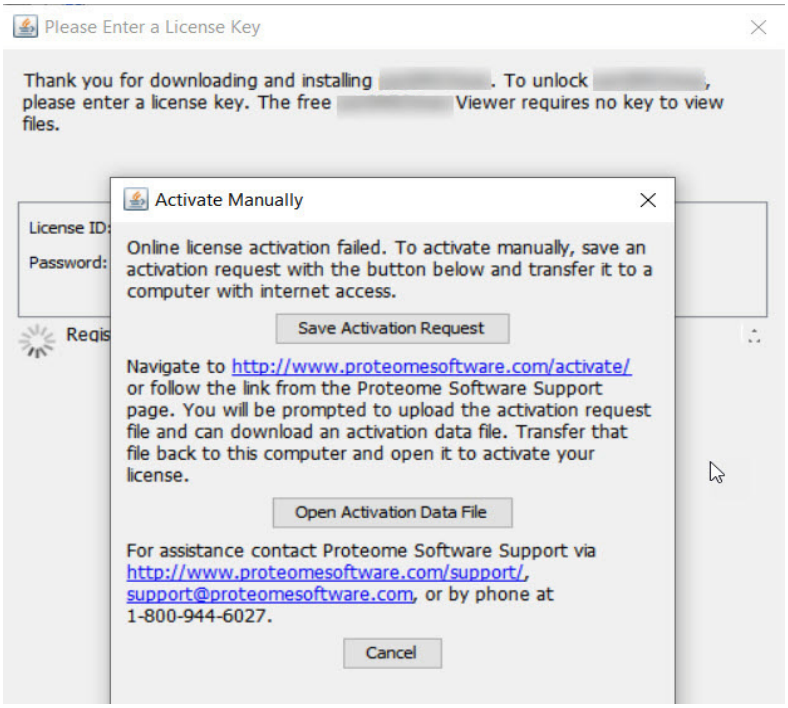
5. Click the Submit button just below the Browse button to upload the activation request file. The license portal will respond with a long text sequence.

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 PTM has been installed.
8. Return to Scaffold PTM on the disconnected computer. Select Open Activation Data File.

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



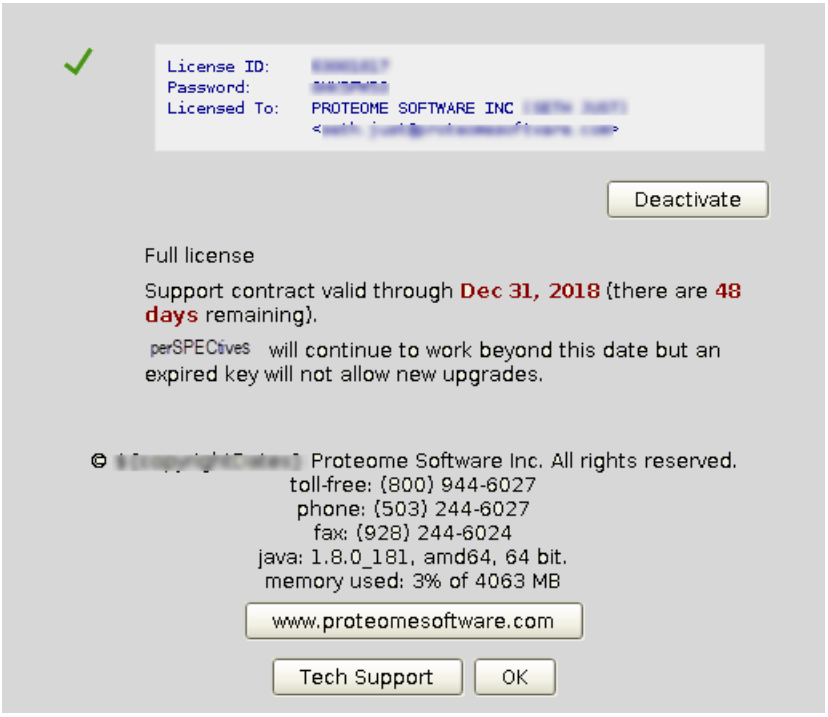
9. Browse to locate the response.xml file and click Open.
10. Scaffold PTM should report that the key was registered successfully. If not, please contact Proteome Software Support for assistance.

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 PTM continues to work but upgrades are not allowed until the contract is renewed. The status of the Scaffold PTM license key may be checked by selecting **Help > About Scaffold PTM** from the main menu.

If the contract has expired and the user wishes to upgrade Scaffold PTM, 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 PTM dialog



Moving Scaffold PTM to a different computer

Each permanent Scaffold PTM 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 PTM dialog.
2. Select Help>>Update License Key and click the Deactivate button.

To reinstall Scaffold PTM:

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 PTM.

Scaffold PTM Viewer

A free Scaffold PTM Viewer may be downloaded from *www.proteomesoftware.com.* The Viewer can open and display any *.SPTM file created by Scaffold PTM, and allows users to distribute Scaffold PTM 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 PTM program, but it cannot load search results files and analyze data.

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

Preface

Welcome to the Scaffold PTM User's Guide. Its purpose is to answer users' questions and guide them through the procedures necessary for using Scaffold PTM 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.

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, <ProteinName>.
- 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 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 PTM.

Scaffold PTM highlights

Scaffold PTM is a software tool designed to help researchers automate PTM site assignments and provide a measure of confidence of PTM site identification. It also identifies potential enzyme recognition sites by scanning the dataset for overrepresented patterns in the amino acids surrounding modification sites.

Graphical Views

Once the data is searched and analyzed the results are shown through various graphical views that are designed to help the user examine the list of Post Translational modifications protein by protein and perform visual inspections of the spectra as needed.

Organize

The Organize View lists all loaded MZID and SQML files and their associated MS Samples. From here, the user can add or delete files, toggle which of the MS Samples to analyze, and rename various identifying details.

Summarize

Three levels of summarization can be easily toggled from Scaffold PTM's main window, providing a way to better evaluate the PTM quantitative values.

Visualize

All identified proteins are listed in the PTM List table along with the number of PTM sites identified in each of them. Many specialized visualization tools are provided, along with the ability to:

- Group proteins
- In the Proteins view modifications are highlighted in different colors along the proteins sequence.
- In the Motifs view motifs are depicted in the sequence logo representation colored according to their chemical properties and their size representing the frequency of the amino acid in that position.

Statistical Tests

Statistical tests are included in the Quantify view when data is imported from Q+ and it is analyzed using protein normalization.

Proteome Software Products

Scaffold PTM belongs to the Scaffold Suite of products. While Scaffold Q+S is an add-on to the core Scaffold product, Scaffold PTM is a standalone application within the Scaffold Suite of applications and requires an independent license key provided by Proteome Software to get activated, see .

It processes mzIdentML (MZID) files that can be exported from all of the other Scaffold Suite of products and also may be produced directly from a number of search engines, like Mascot and PEAKS.

Scaffold PTM provides tools for analyzing MS raw data of relatively small molecules such as metabolites, lipids or glycans.

Suite	Application	Description
Scaffold	**Scaffold**	Visualize and validate MS/MS proteomics experiments.
	Scaffold Q+	Calculate and display relative protein expression levels in a sample determined by tandem mass spectrometry of iTRAQ- or TMT-labeled proteins.
	Scaffold Q+S	Calculate and display relative protein expression levels in a sample determined by tandem mass spectrometry of stable isotopically-labeled (for example, SILAC) proteins.
	Scaffold Scaffold DIA	Catalog, summarize and analyze complex large-scale experiments. Compare protein and peptide similarities and differences at any attribute group summarization level. Easily reorganize samples to compare the impact of tissue types, treatment types, demographic differences, experiment conditions and more.
	Scaffold PTM	Scaffold PTM is a computational tool that, starting from MS/MS spectra of identified Post Translational Modification (PTM), allows the user to derive biological relevant results in an automatic fashion reducing the amount of manual validation required to assure data integrity.
	Scaffold PTM	Scaffold PTM is a software tool designed to help the researchers in the field of metabolomics to search and identify metabolites included in samples analyzed using liquid chromatography-mass spectrometry (LC-MS1 and MS2).

Referencing Scaffold PTM Results

Users are free to copy, modify, and distribute the following examples for citing Scaffold PTM in their publications and reports.

Post Translational Modifications (PTM) Site Localization - Scaffold PTM (Proteome Software, Portland, Oregon, USA) was used to annotate PTM sites derived from MS/MS sequencing results obtained using <SEARCH_ENGINES>. Using the site localization algorithm developed by Sean A Beausoleil, Judit Villén, Scott A Gerber, John Rush & Steven P Gygi, Nature Biotechnology 24, 1285 - 1292 (2006), Scaffold PTM re-analyzes MS/MS spectra identified as modified peptides and calculates Ascore values and site localization probabilities to assess the level of confidence in each PTM localization. Scaffold PTM then combines localization probabilities for all peptides containing each identified PTM site to obtain the best estimated probability that a PTM is present at that particular site.

Motif Analysis - PTM were scanned for over-represented patterns in the amino acids surrounding the modification sites using the method described in Schwartz, D. & Gygi, SP (2005) Nature Biotechnology 23(11):1391-1398.

Chapter 2
PTM Analysis in Scaffold PTM

Scaffold PTM is a computational tool that automates post-translational modification (PTM) site assignment in proteomic experiments. It analyzes MS/MS spectral output and provides researchers with an objective measure of the confidence of PTM (e.g. phosphorylation) site localization.

The application also includes a motif extraction algorithm that identifies motifs surrounding PTM sites by looking for patterns that are over-represented in modified peptids as compared to a background set of sequences. This provides a means to identify potential enzyme recognition sites.

When used in conjunction with Scaffold Q+ or Q+S, Scaffold PTM also offers quantitative features to assess differential expression and modification among samples. When quantitative data exported from one of the Scaffold quantitative modules is loaded into Scaffold PTM, it computes the statistical significance of fold-change values for each modification site.

This chapter explains the following topics:

- "Automated PTM Site Localization" on page 24, which provides a brief description of the algorithm used to confidently assign PTM sites.
- "Motif Identification" on page 26, which briefly describes the algorithm used to identify motifs surrounding PTMs.
- "Quantitative Analysis" on page 27, which describes statistical features in the Quantify View.
- "Quantitating PTM dynamics" on page 28, which describes an approach to quantify PTM expression when protein levels change during the experiment.
- "Scaffold PTM Views" on page 30, which lists all of the data views included in Scaffold PTM.

Automated PTM Site Localization

Scaffold PTM uses ASCORE's probabilistic approach and scoring technique to annotate modification sites in MS/MS spectra. Sequest, Mascot and similar peptide search engines that identify peptides may, as a side effect, identify peptides in which certain amino acid sites are modified. However, in cases in which the peptide contains multiple potential sites for the modification, they do not generally measuare the likelihood that the PTM is located at one site rather than another. That task has traditionally been left to manual validation. As the size of experimental data sets has increased, and the importance of high-confidence PTM site recognition has grown, however, an automated validation tool has become essential.

In response to this need, the Ascore algorithm was developed in the Gygi lab at the Harvard Medical School's Department of Cell Biology[1] [2]. The algorithm measures the probability of correct PTM site localization based on the presence and intensity of site-determining ions in MS/MS spectra, and targets high-throughput PTM analysis and site localization.

Within the Scaffold PTM environment, Ascore becomes an effective tool for automating large-scale, post-translational studies.

Using Ascore, Scaffold PTM re-analyzes results of previous searches done with Sequest, Mascot or other Scaffold-compatible search engines. The new analysis attempts to determine the likelihood that the location selected by the search engine is the best match in the observed spectrum for a PTM site. It graphically displays a list of the reported PTM sites with estimates of the confidence of the assignment of the PTM to a specific site and presents the evidence supporting the assignment of each PTM to its site.

By comparing site-determining ions, Scaffold PTM produces Ascore results assigning an ambiguity score to each reported PTM site. The number of PTMs in each peptide can be determined by the precursor ion mass of the peptide's spectra. Scaffold PTM adds this knowledge to the Ascore to derive a site location probability and it then combines the site location probability estimates from all spectra matching peptides containing the site to obtain the best estimate of the probability that the PTM is at that site.

As a result, Scaffold PTM reduces the amount of manual validation required while improving data set integrity. Accurate determination of these sites removes an important bottleneck in proteomics and facilitates faster and more comprehensive analysis of PTM studies.

Critical for scientifically useful publication, the results derived from Scaffold PTM satisfy the requirements for acknowledgment of ambiguity.

Neutral Losses

The original Ascore calculation does not account for neutral losses that modified peptides may undergo. In some cases, particularly with modifications such as O-GlcNac, in which the

1. Sean A Beausoleil, JuditVillén, Scott A Gerber, John Rush,& Steven P Gygi. A probability-based approach for high-throughput protein phosphorylation analysis and site localization. Nature Biotechnology 24, 1285 - 1292 (2006)
2. Zhai B, Villén J, Beausoleil SA, Mintseris J, Gygi SP. Phosphoproteome analysis of Drosophila melanogaster embryos, J Proteome Res. 2008 Apr;7(4):1675-82

full mass of the modification may be lost, this results in treating peaks which are actually ambiguous as evidence for a particular localization. To better assess localization probability in the face of such ambiguity, Scaffold PTM offers the option to use an extended version of the Ascore algorithm in which potential neutral loss peaks are accounted for in the calculations.

When this option is selected, the predicted framentation patterns used in the Ascore calculation include potential neutral loss peaks.

Two types of losses are considered:

- Neutral losses from modifcations, as read from the modification specification in the input mzIdentML file.
- Water losses, only from unmodified S,T,E or D residues.

Doubly charged fragments are allowed if the precursor charge is greater than or equal to 2 and the fragment contains a basic residue.

Note that the neutral loss option does not affect ETD spectra, in which c and z ions are always considered, up to triply charged ions as the precursor charge state allows.

Motif Identification

Scaffold PTM's motif analysis tool allows detailed investigation of statistically over-represented motifs identified in the experiment. That is, it allows the discovery of sequence motifs that are found to be modified more than would be expected if modification sites were chosen randomly from all possible sites. Additionally, it allows for analysis of user-specified motifs to assess their prevalence.

Motif analysis algorithm

In order to identify potential enzyme recognition sites, Scaffold PTM scans the dataset in an experiment for overrepresented patterns in the amino acids surrounding modification sites. The motif analysis algorithm is based on the motif discovery and scoring algorithm implemented by Schwartz and Gygi[3], where protein sequences around identified PTM sites (the "foreground" dataset) are compared to protein sequences around possible PTM sites (the "background" dataset).

The prevalence of amino acids at each position in the motif sequences in each dataset are used to compute a binomial probability that the experimental prevalences would be observed if PTMs were randomly distributed among potential sites (that is, independently of the surrounding sequence). These probabilities are used, when discovering new motifs or analyzing known motifs, to compute a score for each motif, using the equation

$$S = \sum_{i} -\log(p_i)$$

where p_i is the binomial probability computed for the i_{th} position in the motif.

Motif Discovery

Scaffold PTM searches for motifs in the current dataset using the approach described in the section Motif analysis algorithm. For each modification type (including the modified residue, e.g. Phospho of S is considered separately from Phospho of T), this technique looks for the most significant (lowest binomial probability) combination of amino acid and sequence position and adds it to the currently considered motif. This process is repeated until no more residue/position pairs have a binomial probability greater or equal than 10^{-5}. The just-discovered motif is removed from both the foreground and background datasets to ensure that subsequent motif discovery is not biased by previously-discovered patterns in the data. The program then tries to discover another motif by the same process. This is repeated until no new motifs are discovered.

Motif search is activated when initially loading data or when the user clicks on the action icon Search for Motifs , see "Motif Tool bar" on page 95

3. Schwartz, D. & Gygi, SP (2005) An iterative statistical approach to the identification of protein phosphorylation motifs from large-scale data sets. Nature Biotechnology 23(11):1391-1398

Quantitative Analysis

Scaffold PTM can use quantitative values exported from Scaffold Q+ or Q+S to perform relative quantitation among samples or categories. The program performs statistical testing to assess the significance of fold change values calculated for each modification site. When available, these values are displayed in the PTM Quantitation Tab, the Peptide Quantitation Tab and the Volcano Plot.

Statistical Analysis

For any modification site with two or more ratios within a single MS Sample (or Biological Sample or Category, depending on the current Summary Level), Scaffold PTM computes a p-value that assesses the probability that you would observe a median fold change at least as extreme as the given observations even though the true fold change is zero. That is, lower p-values imply more confidence in the conclusion that a given site was up- or down- regulated in a given sample, while high p-values imply there is not enough evidence to conclude that the true fold change was non-zero.

Scaffold PTM does not perform statistical analysis of fold changes in the Reference category, as the definition of zero fold change is drawn from these measurements.

To compute the p-value the application uses a non-parametric technique called the Wilcoxon Signed Ranks test, which is similar to a t-test, but does not assume that observations are normally distributed. Thus it is appropriate to apply to (and continues to produce meaningful p-values for) any input, e.g. when combining dissimilar samples at higher Summary Levels.

The computed p-values are visible in the PTM Quantitation tab, where each non-Reference Category sample has a second column showing the p-value for each site in that sample. These values are colored blue when they are greater than 0.05, and gold when they are less than or equal to 0.05 (i.e. significant at 0.05). The p-values are also used in the Volcano Plot, where they are transformed (via the function $y=-\log10(p)$) to make significant values large, and insignificant values close to zero. The plot also shows a horizontal line corresponding to $p=0.05$, so that all sites appearing above the line are colored gold in the table.

Quantitating PTM dynamics

There are two major classes of phosphoproteomics experiments,: the first is an in vitro kinase assay, where the researcher takes a group of proteins or digested peptides, adds a particular kinase and observes which proteins or peptides it phosphorylates. Since the interest is in dynamics, the researcher usually samples at time points on the order of minutes) and consequently expects that the levels of underlying proteins will not change dramatically. This is the sort of quantitative experiment for which Scaffold PTM was originally designed.

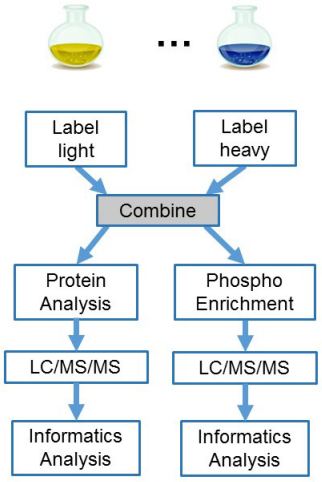
However, sometimes the primary interest is in interactions of kinases in vivo or in dynamic systems in which protein levels are expected to change. In this case, the phosphopeptide abundances reflect both changes in phosphorylation status and in protein expression levels simultaneously. In such experiments, separating these changes becomes important in order to be able to assess the true change in PTM expression.

If the protein level does not change, then increases in phosphopeptide abundance reflect an increase in phosphorylation, while decreases in phosphopeptide abundance indicate that phosphorylation has decreased. However, if the protein expression level has changed, changes in the abundance of phosphopeptides may not reflect changes in the level of phosphorylation but may merely reflect changes in overall peptide abundance.

The most common strategy for dealing with this problem was discussed in an MCP paper by the Gygi lab[4]. In this approach, samples are labeled, combined together, and then divided into two groups. One group is carried forward for protein analysis, and the other for phospho enrichment using IMAC, SCX fractionation, or titanium dioxide, see Figure 2-1. This results in two sets of quantitative data, those reflecting differences in the phosphopeptides, and those which can be used in assessing the overall protein levels.

4. Wu R1, Dephoure N, Haas W, Huttlin EL, Zhai B, Sowa ME, Gygi SP, Correct interpretation of comprehensive phosphorylation dynamics requires normalization by protein expression changes, 2011 Molecular & Cellular Proteomics, 10, M111.009654

Figure 2-1: PTM dynamic experiment



To accommodate this type of experiment, Scaffold Q+ and Scaffold Q+S offer an export that provides the protein level data, the Protein Quantitation XML Report. If this export is imported into a Scaffold PTM experiment that was created by loading a Scaffold Q+ or Q+S SQML file, Scaffold PTM will normalize the levels measured in the PTM-enriched samples to adjust for differences in protein expression. For more details about the calculations performed see "PTM dynamics - quantitative calculations" on page 126.

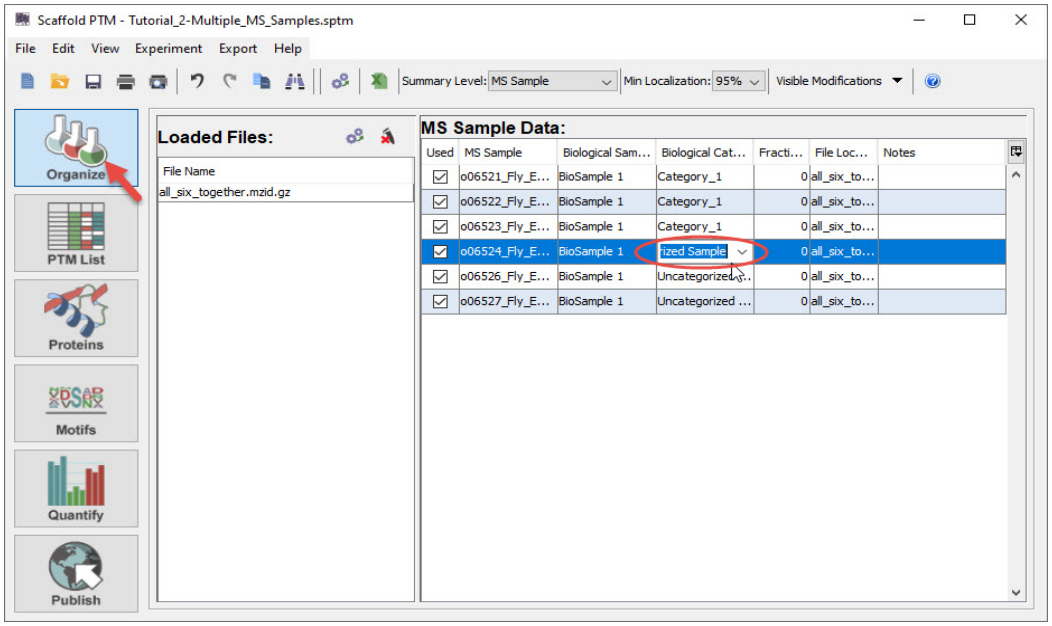
Scaffold PTM Views

Scaffold PTM offers both a high-level overview of the list of Post Translational Modifications, and a detailed look at the supporting data. Scaffold PTM presents the more detailed levels in a coherent structure, helping the user to verify critical findings. The information is organized through a series of views which can be easily accessed through the main Scaffold PTM window.

Organize View

This view shows the list of MS samples loaded into Scaffold PTM. Tools in the view help the user add new files or remove files that have already been loaded. The user may also edit the Sample names and add Category names to allow grouping of the samples.

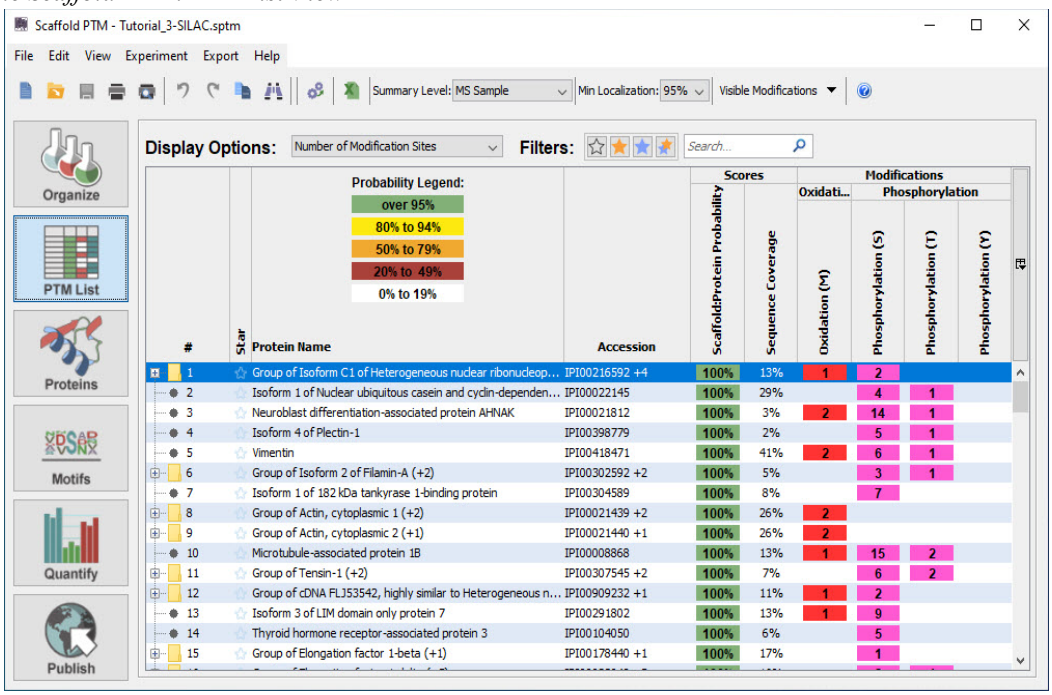
Figure 2-2: Scaffold PTM: Organize View



PTM List View

This View provides a list of identified proteins and shows the number and types of modification sites in each sample for each of them.

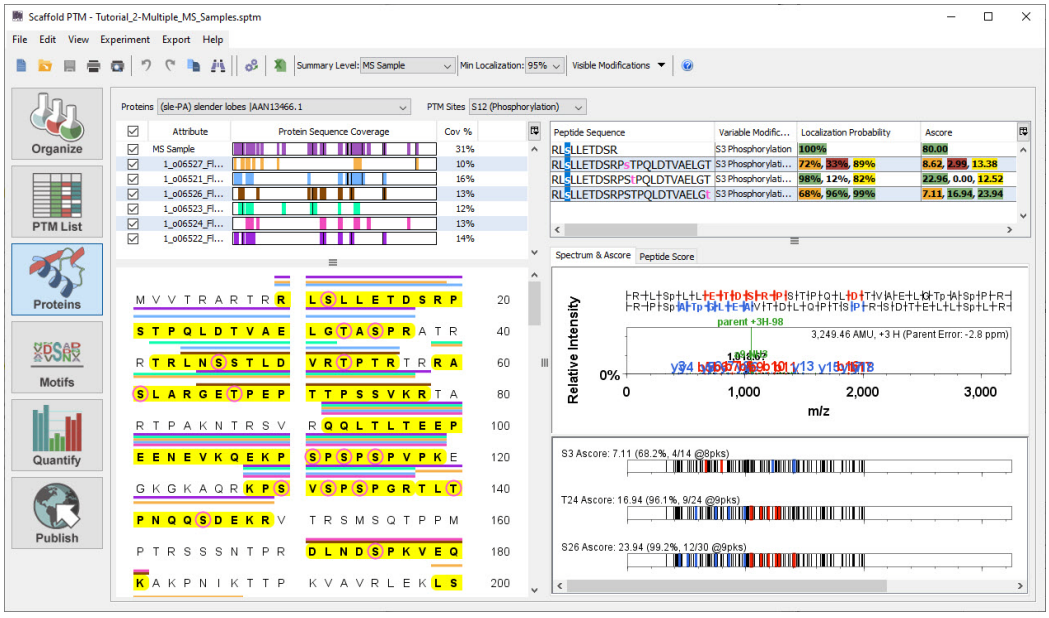
The Scaffold PTM: PTMList View



Proteins View

This view structures, in different graphical containers, a large amount of detailed information about the modifications and peptides present in a protein.

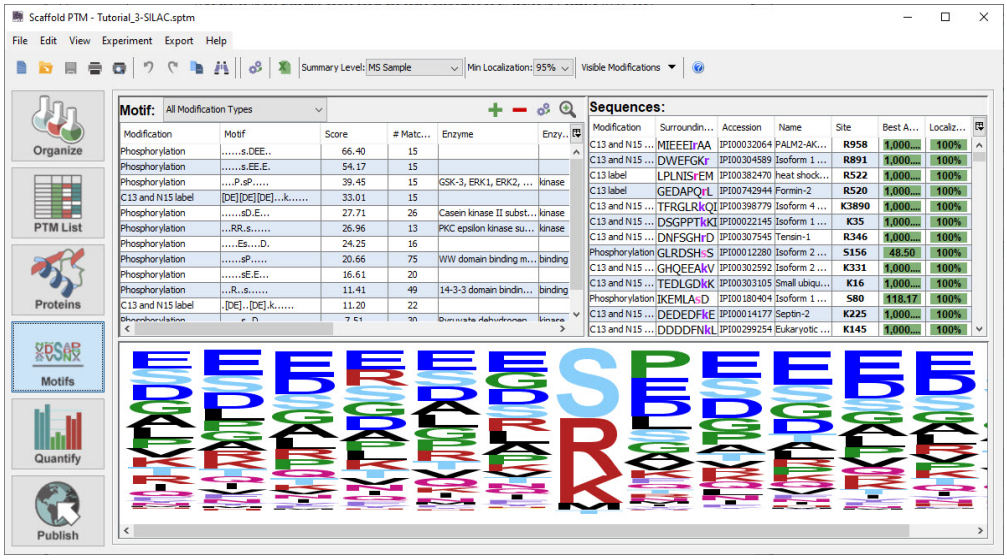
Figure 2-3: Scaffold PTM: Proteins View



Motifs View

Scaffold PTM performs a search for sequence patterns that are associated with the modifications represented in the experiment. This view provides the list of identified PTM motifs and provides information about them.

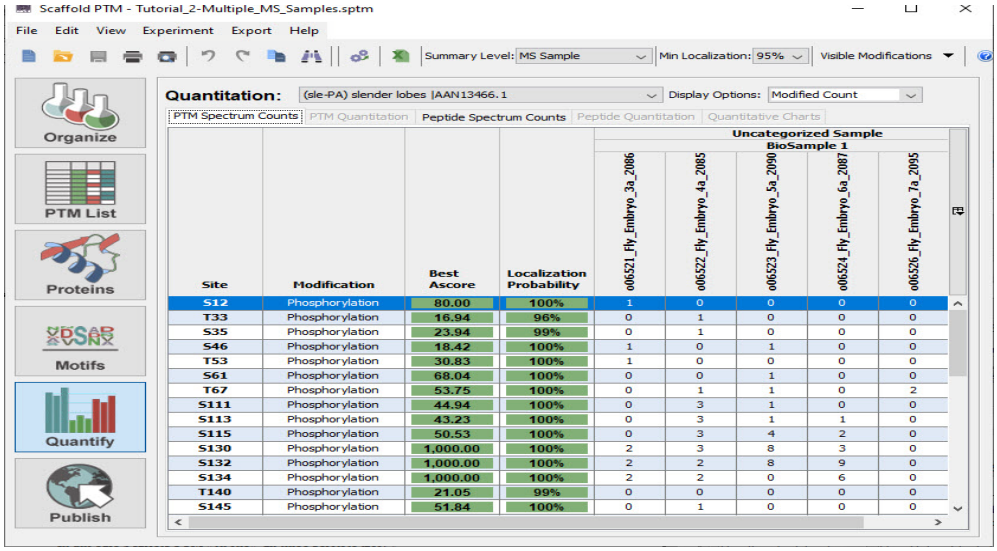
Figure 2-4: Scaffold PTM: Motifs view



Quantify View

This view provides quantitative site information. Depending on the type of quantitative data loaded, it can show spectral counts, isobaric labeling results or precursor intensity quantitation. In certain types of experiments, it may show quantitative values normalized to adjust for differences in protein level, providing a better measure of differential modification.

Figure 2-5: Scaffold PTM: Quantify View

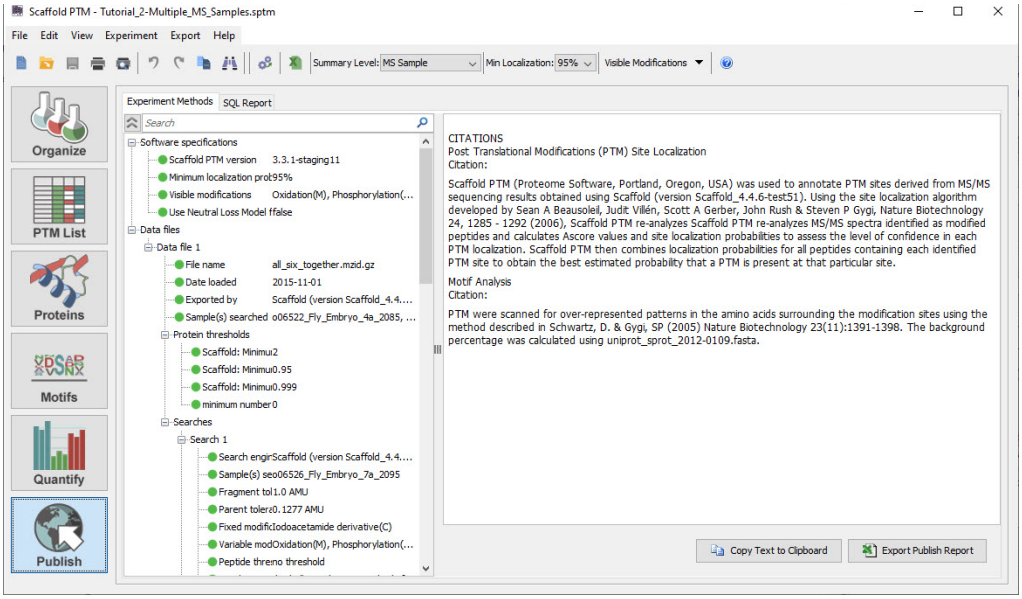


Publish View

This view contains two tabs: the Experiment Methods tab and the SQL Report tab. The Experiment Methods tab records the information needed to reproduce the analysis of the experiment. This provides the information needed for publication of results.

The SQL Report tab is an SQLite platform through which the user may view the data stored in the current SPTM file through SQL queries. Queries can be saved and reused to create custom reports.

Figure 2-6: Scaffold PTM: Publish View



Chapter 3
Loading data in Scaffold PTM

Files supported by Scaffold PTM

Scaffold PTM analyzes search engine results in the mzIdentML format from Scaffold, Mascot, or Peaks. It also analyzes quantitative data exported from Scaffold Q+ or Scaffold Q+S.

NOTE: Scaffold PTM does not support combining mzIdentML files produced by different programs. Results from different search engines may be processed through Scaffold and the resulting MZID files may be loaded together into Scaffold PTM.

mzIdentML file specifications

The mzIdentML standard format for proteomics data, developed by the HUPO Proteomics Standards Initiatives is a common output file format for many search engine applications. Typically mzIdentML exports create MZID files with one or more related MGF files.

- *A description of the standard specifications is available at the following website http://www.psidev.info/mzidentml.*

Scaffold PTM creates experiments by loading *.mzid or *.mzid.gz files version 1.1.0 or higher. It is important that the related *.mgf files are included in the same directory where the mzid files are stored.

*Note that PTM validation needs to have the related *.mgf files loaded along with the *.mzid files.*

mzIdentML exports from Scaffold

For data analyzed in **Scaffold**, **Scaffold Q+** and **Q+S,** an mzIdentMLand related MGF files may be exported by selecting ***Export > mzIdentML...*** from the main menu of the application**.** Three export types are offered in the export dialog and the user should select “Scaffold PTM analysis”. This will generally provide mzIdentML files suitable for loading int Scaffold PTM but clicking Advanced allows the user to further customize the export parameters if necessary.

Once the desired options are selected, clicking **OK** brings up a file browser for selecting a destination in which to save the MZID file export. Scaffold creates a new directory that contains MZID files and the related MGF files.

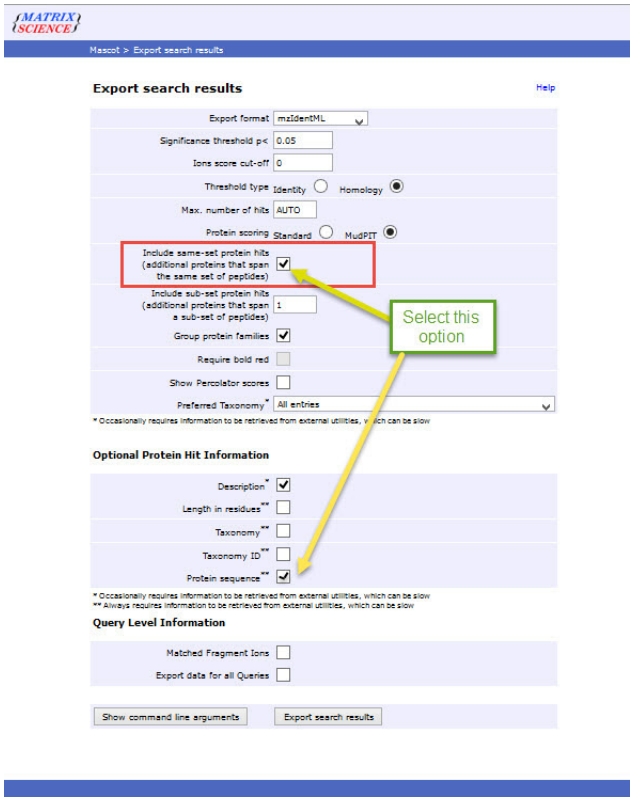
ScaffoldBatch also includes commands to create mzIdentML exports from existing Scaffold

files or from new files directly created in ScaffoldBatch.

mzIdentML exports from MASCOT

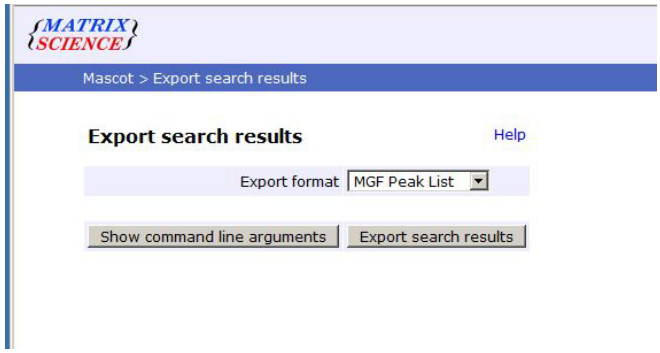
Mascot creates mzIdentML files from the MASCOT Search Results page. Among the various options available in **Export Search Results** pane, either “Group Protein Families” or “Required Bold Red” should be selected. Users should also ensure that proper homology information is reported by selecting “Include Same-set protein hits”, see Figure 3-1, and should select the option “Protein sequence” from the **Optional Protein Hit Information**, the created MZID files will include information that will allow Scaffold PTM to display sequences and coverage.

Figure 3-1: Mascot Server - Export mzIdentML



A separate export is required to produce the related MGF files, see Figure 3-2

Figure 3-2: Mascot Server Export Search Results: MGF Peak List



NOTE: When loading Mascot data, if Scaffold PTM does not find the related MGF files in the same directory where the MZID is saved, it will ask for the location of the MGF files. This might also happen when Mascot exports the MGF files under a different name than the one reported in the MZID file. During loading, a browser will open and the user should select the MGF file which corresponds to the MZID being loaded.

mzIdentML exports from PEAKS

Scaffold PTM loads and analyzes MZID files created by PEAKS. A description of how to export MZID and MGF files from PEAKS is provided in Loading Peaks Data into Scaffold.

ScaffoldQuantML exports

Scaffold Q+ and Scaffold Q+S are Proteome Software's labeled quantitation software packages. Scaffold Q+ can analyze iTRAQ (Applied Biosystems) or Tandem Mass Tagged (TMT, Thermo Scientific) labeled data and precursor intensity data while Scaffold Q+S can also load stable isotope labeled samples.

From the Q+ or Q+S quantitative Add-on window in Scaffold, the menu command **Export > ScaffoldQuantML** creates an SQML file and the related MZID and MGF files.The SQML file contains quantitative information, the MZID contains all of the protein and peptide identifications, and the MGF files contain the spectra.All of these must be present in the same directory, but the user should select the SQML file for loading into Scaffold PTM.

SQML files exported from Scaffold Q+ or Scaffold Q+S contain quantitative ratio values. These are the same values displayed in the Scaffold Q+ or Q+S Samples View when the Log2 Fold Change Display Option is selected. Scaffold PTM uses these value**s** to provide a measure of differential modification between samples or categories of samples.

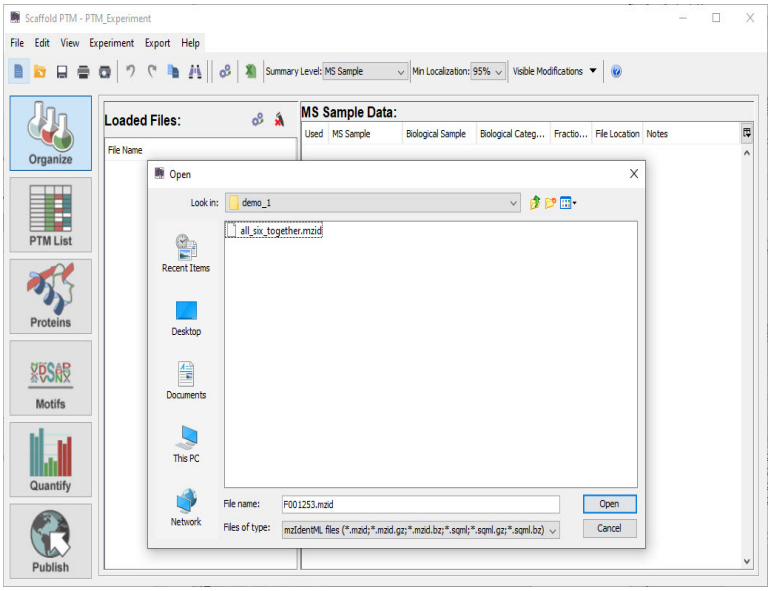
Scaffold Protein Quantitation XML exports

In order to normalize quantitative PTM results to adjust for differences in protein levels, Scaffold Q+ or Q+S also provides the Protein Quantitation XML Report, a *.ProteinQuantXML file. This file is exported from a Scaffold analysis of non-enriched samples and is used in conjunction with a SQML file from corresponding enriched samples. For more information see “Creating a Protein-Normalized Quantitative PTM Experiment” on page 40

Loading data into Scaffold PTM

1. To create a new experiment in Scaffold PTM the user may either select **File > New** or click on the “New” icon in the tool bar below the main menu in the Scaffold PTM main window. A dialog box appears, asking the user to navigate to the directory where the MZID or SQML file(s) is (are) located.

Figure 3-3: Select data files dialog box

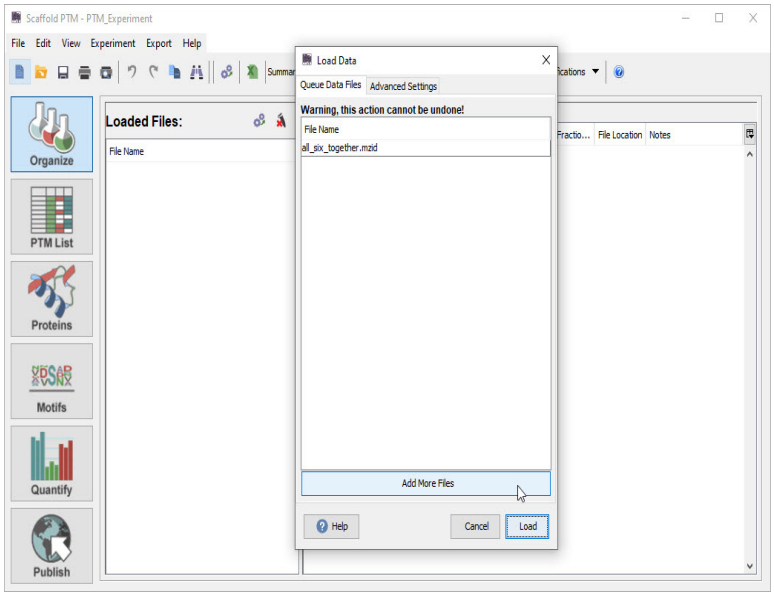


2. It is possible to select either a directory that contains the MZID files, a compressed directory MZID.GZ or a single MZID file. When analyzing quantitative data from Scaffold Q+ or Scaffold Q+S, the user should select the SQML file. The associated MZID and MGF files will be loaded automatically.

3. Clicking Open launches a new dialog, **Queue Data Files...**. From this dialog, additional files may be selected for loading by clicking the **Add More Files** button. All of the selected file names appear , listed in the dialog box and ready to be loaded into Scaffold PTM.

NOTE: It is not possible to delete a file from the Queue Data Files list. Once the files have been loaded the user can delete a file by clicking the delete button in the Loaded Files Pane in the Organize View.

Figure 3-4: Queue files for loading dialog box



4. When analyzing modifications which readily undergo neutral losses that complicate the localization calculations, the user may wish to open the Advanced Settings tab and select the option “Use Neutral Loss Model for Ascore”.
5. Clicking Load loads the listed files into Scaffold PTM.

- During loading, a wait dialog box provides a description of the ongoing operations and a rough estimate of the time left to completion.
- When the loading operation is completed, Scaffold PTM opens the Organize view which shows the list of loaded files.

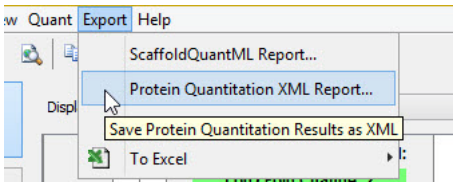
Creating a Protein-Normalized Quantitative PTM Experiment

For certain types of quantitative studies, Scaffold PTM includes a feature that enables quantitative analysis of PTM activity by simultaneously considering protein-level and site-level changes, see “Quantitating PTM dynamics” on page 28.

To create a protein-normalized PTM Quant experiment in Scaffold PTM, the user should:

1. Create two Scaffold Q+ or Scaffold Q+S experiments: one with modification-enriched Quant data and one with unenriched protein Quant data.
2. Export a SQML file from the modification enriched experiment using the command **Export > ScaffoldQuantML report...** located in the Q+ add-on window menu.
3. Export a Protein Quantitation XML file from the protein experiment using the command **Export > Protein Quantitation XML report...** located in the Q+ add-on window menu.

Figure 3-5: Q+ add-on window menu

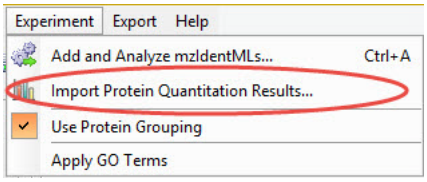


4. Load the SQML data file into Scaffold PTM, as instructed in Loading data into Scaffold PTM and save the file.

   The PTM Quantitation Tab, Peptide Quantiation Tab and Volcano Plot will show the loaded quantitative ratios for each PTM site in the experiment. Because the protein quantitation data has not yet been loaded, the PTM Quantitation Tab will have the title “PTM Quantitation”, and the Volcano Plot will show all points as triangles.

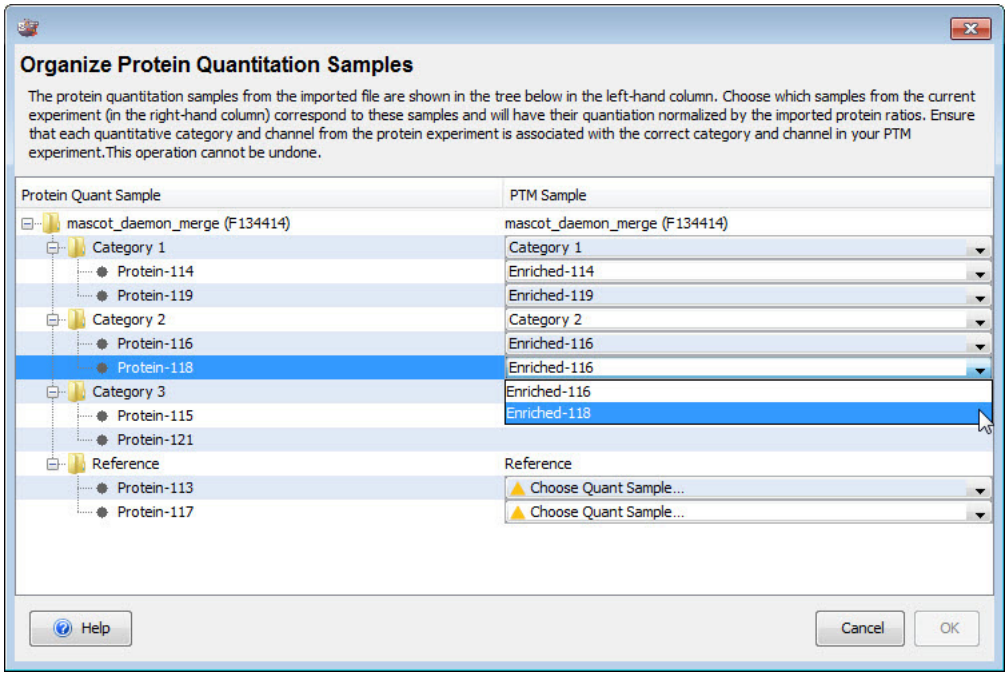
5. To normalize the data by protein quantitative expression, load the protein quantitation data into the Scaffold PTM experiment just created. This can be done by selecting the command **Experiment > Import Protein Quantitation Results...** from the Scaffold PTM main menu and choosing the XML file exported in Step 3.

Figure 3-6: Scaffold PTM Import Protein Quantitation results



6. After selection of the XML file a hierarchical list of the samples and quantitative channels in the imported file appears.
7. Match the samples in the PTM experiment with their corresponding protein measurements.

Figure 3-7: Organizing protein quantitation samples



8. After specifying the sample organization, click “OK” and the data will be imported.

The PTM Quantitation Tab will now display the title “PTM Quantitation (Protein-Normalized)” and the values shown will be normalized by the imported protein ratios. In any sample for which a protein’s ratio was not measured, the values shown will be un-normalized and surrounded with square brackets (“[<ratio>]”). The Volcano Plot will show all sites with protein-normalized ratios as circles, and will not show any un-normalized data.

To prevent un-normalized proteins, it may be advisable to export the Protein Quantiation report with lower thresholds than were applied when exporting the SQML file from the enriched experiment.

Scaffold PTM files

Scaffold PTM creates its own file type called SPTM, which stands for Scaffold Post Translational Modification. This file is an SQLite file, a light weight, high performance SQL database file with a great deal of flexibility. Indeed, in Scaffold PTM, it is possible to query the experiment using Structured Query Language (SQL) and to save these queries for future use. This direct access to the data structure gives Scaffold PTM users a unique capability to manipulate and analyze their data.

Inconsistent Ascore Warning

This message appears if a file created in a previous version is opened and the program detects that the displayed Ascore values, which are the values saved in the file, do not match the calculations depicted in the Spectrum & Ascore display, which is recomputed when the file is opened.

This is an unexpected situation, so if you encounter this warning, please contact support@proteomesoftware.com.

Chapter 4
The Scaffold PTM Main Window

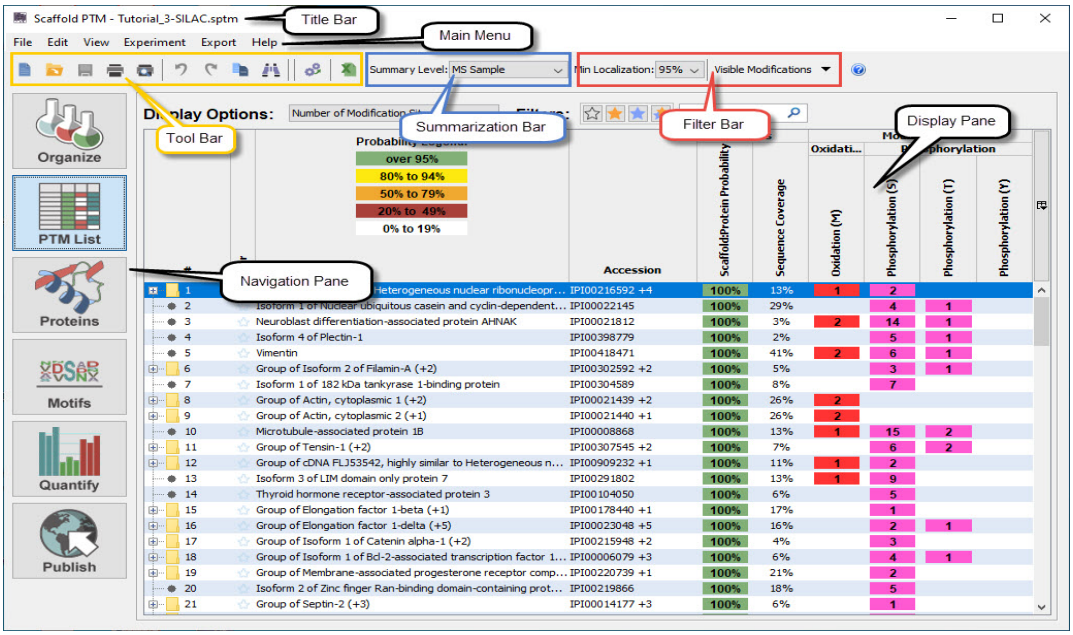
Scaffold PTM, like all of the applications in the Scaffold Suite, is built around a main window which contains a number of different views. In each view the experimental data is organized so that users can easily examine experimental results from different perspectives.

The Scaffold PTM main window provides quick access to all of the features and functions of the application.

The window has a number of components, including:

- Title bar
- Main menu commands
- Tool-bar
- Summarization bar
- Scaffold PTM Main Window Filters bar
- Navigation bar
- Display pane.

Figure 4-1: Scaffold PTM main window



Title bar

Figure 4-2: Title bar

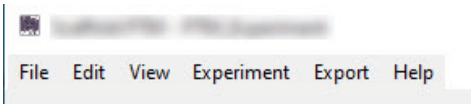
The “Scaffold PTM” logo and the program name are always displayed in the title bar at the top of the main window. If an experiment is open, the experiment name also appears in the title bar. When a new experiment is created, the default name “PTM Experiment” is displayed, and when the experiment is saved, the name of the SFDB file becomes the new experiment name.

---

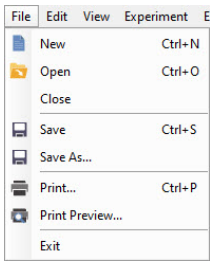
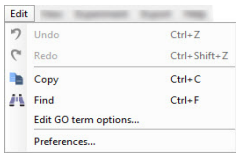
*The version of Scaffold PTM in use is not displayed in the Title bar. The user must go the **Help > About** option in the main menu to determine the version number. See “Main menu commands” below.*

Main menu commands

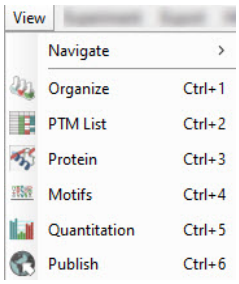
Figure 4-3: Scaffold PTM Main Menu

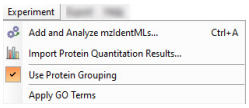
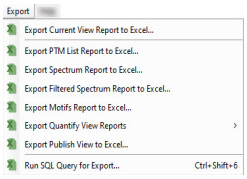


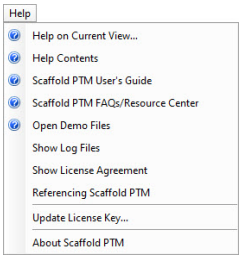
The Scaffold PTM main menu is set up in a standard Windows menu format with sub menu commands grouped into menus (File, Edit, View, Experiment, Export and Help) across the menu bar. Some of these menu commands are available in other areas of the application.

Menu	Menu Commands
File 	• **New (Ctrl+N)**—Starts a new experiment and opens a file browser to allow the selection of *.mzid files to be loaded in the application. Once selected, the Queue Data Files dialog opens, allowing the possibility to add more files to the list queued for loading. See “Files supported by Scaffold PTM” on page 35. • **Open (Ctrl+O)**—Opens a saved Scaffold PTM experiment file, SPTM, through a file browser. • **Close**—Closes the current experiment, standard Windows behavior. • **Save (Ctrl+S)**—Saves the current experiment, standard Windows behavior. • **Save As...**—Saves the current experiment offering the option to use a different name, standard Windows behavior. • **Print... (Ctrl+P)**—Prints the current view. • **Print Preview...**—Previews the current view with the option of printing the document. • **Exit**—Closes the Scaffold PTM window.
Edit 	• **Undo -** when active, allows the last operation to be reversed • **Redo -** when active, allows reapplication of an operation that has been reversed by Undo • **Copy (Ctrl+C)**—Copies the currently selected table to the clipboard. The user can then paste the copied information into a third-party program such as Excel or Microsoft Word. . • **Find (Ctrl+F)**—Opens the Find dialog that searches the first table present in the Current View • **Edit GO Terms Option...**—See “Edit GO Term Options” on page 52. • **Preferences**—see “Preferences” on page 48.

Menu	Menu Commands

View
 
	• **Navigate**— has two options active only when the view contains tabs:
	  • **Select Previous Tab (CTR+Page Up)**—
	  • **Select Next Tab (CTR+Page Down)**—
	• **Organize View (CTRL+1)**— see “The Organize View” on page 69
	• **PTM List View (CTRL+2)**—see “The PTM List View” on page 73
	• **Proteins View (CTRL+3)**—see “The Proteins View” on page 81
	• **Motifs View (CTRL+4)**—see “The Motifs View” on page 93
	• **Quantify View (CTRL+5)**—see “The Quantify View” on page 103
	• **Publish View (CTRL+6)**—see “The Publish View” on page 117

Experiment 	• **Add and Analyze mzIdentML...(CTRL+A)**—Opens a file browser to select MZID data files to add to the already loaded data files, see “Files supported by Scaffold PTM” on page 35. • **Import Protein Quantitation Results...**—Opens file browser to import Protein Quant XML files exported from Scaffold Q+ Add-on, see .“Creating a Protein-Normalized Quantitative PTM Experiment” on page 40 • **Use Protein Grouping**— it is a toggle and once selected it organizes the protein list in the PTM List View in groups. • **Apply GO Terms**— Active when at least one GO annotation database is loaded in Scaffold PTM. When selected, PTM searches for GO terms and then lists them in the PTM List table, see “Edit GO Term Options” on page 52
Export 	All exports included in this menu create Coma Separated Values (CSV) text files that can be opened and viewed in Excel. • **Export Current View report to Excel...**—Generates a CSV file of the current view as it appears. • **Export PTM List Report to Excel...**—Generates a CSV file of the PTM List table as it appears in the PTM List View. • **Export Spectrum Report to Excel...**—Generates a CSV file with the list of all the spectra included in the MZID files loaded in the experiment, see “Spectrum report” on page 123. • **Export Filtered Spectrum Report to Excel...**—Generates a CSV file with the list of spectra included in the MZID files loaded in the experiment that respect the filter settings, see “Spectrum report” on page 123. • **Export Motifs Report to Excel...**—Generates a CSV file of the information listed in the Motifs table. • **Export PTM Counts Report to Excel...**—Generates a CSV file of the PTM Spectrum Counts table for all proteins. • Run SQL Query for Export... -Opens the SQL Report tab of the Publish View, see “SQL Export tab” on page 119

Menu	Menu Commands
Help  	• **Help on Current View**—Opens the Online Help that is specific for the currently displayed view. • **Help Contents**—Opens the Contents page for the Online Help dialog. • **Scaffold PTM User's Guide**—Opens the current Scaffold PTM User's Guide. • **Scaffold PTM FAQs/Resource Center**—Opens the user's default web browser to the Home page of the Proteome Software resource center. • **Open Demo Files**—Opens the folder where Scaffold PTM demo files are stored. The user can choose any of the pre-loaded files to test Scaffold PTM capabilities. • **Show Log Files**—Opens a folder that contains Scaffold PTM error_log and output_log files • **Show License Agreement**—Opens a TXT document which describes the usage license agreement for Scaffold PTM • **Referencing Scaffold PTM**—Provides examples on how to reference the use of this application. • **Update License Key...**—Opens a dialog accepting inputs of a purchased license key to allow full use of the application. For more info see "Scaffold PTM Licensing" on page 8. When a full licensed application is in use this option is not going to be visible. • **About Scaffold PTM**—Provides the release information for the current version of Scaffold PTM, license information, contact information for Proteome Software, Inc.. It also reports information about the system where Scaffold PTM is installed, the amount of memory available to the application and the percentage of memory used by it.

Find Dialog

The **Find dialog** searches and highlights proteins in the PTM List table according to what has been typed in the box. The filter searches for the typed characters in the Protein Name and Accession Number columns in the PTM List Table. When the searched items are found the text box turns green, if not it turns red.

Preferences

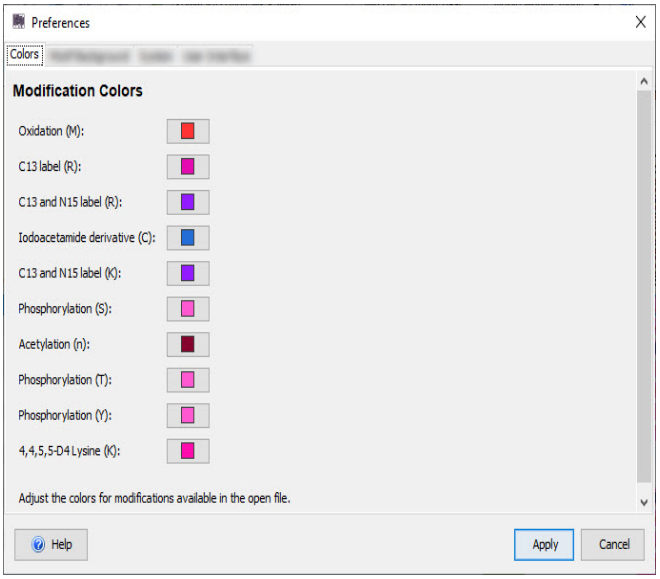
The main menu option: **Edit > Preferences** opens the **Preferences** dialog which contains the following tabs:

- "Colors tab" on page 48
- "Motif Background tab" on page 49
- "System" on page 50
- "Number of Processors" on page 51
- "Internet Settings" on page 51

Colors tab

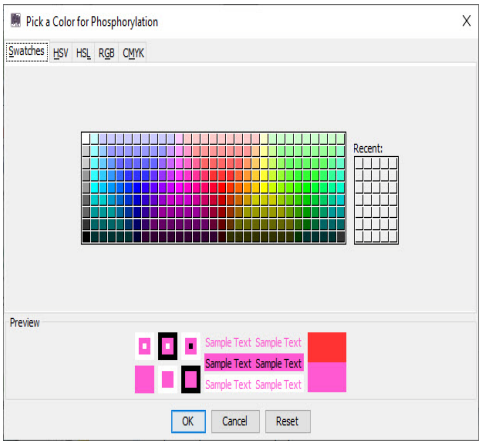
This tab allows the user to customize the color assigned to each PTM, see Figure 4-4.

Figure 4-4: Preferences: Colors tab"



Double-clicking on the colored box assigned to a PTM in the Colors tab opens the color selection dialog. Through this dialog, a different color may be assigned to the selected PTM using swatches, or the HBS or RGB methods, see Figure 4-5.

Figure 4-5: Color Selection Dialog



The **OK** command finalizes the new color selection, while **Reset** goes back to the original choice and **Cancel** voids the operation.

Motif Background tab

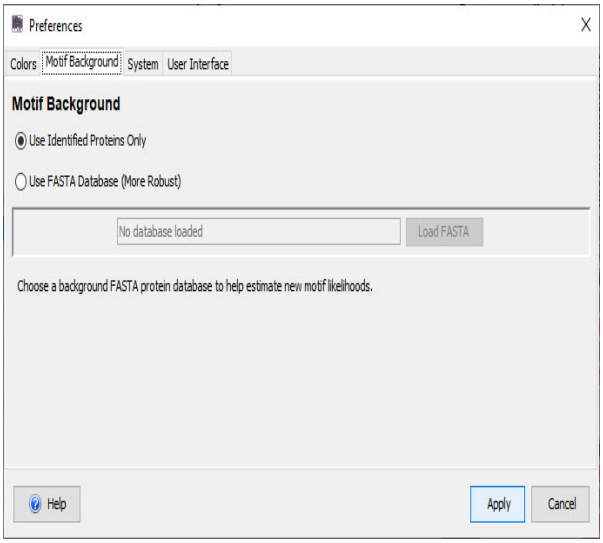
Scaffold PTM's motif analysis identifies sequence patterns that are over-represented in the amino acids surrounding modification sites. In order to assess this, the program uses a set of proteins as a background and measures the frequency with which a specific amino acid is surrounded by a particular sequence motif to establish a Background Percentage for comparison

When using "FASTA Database" or "Identified Proteins" as the Background source, Scaffold PTM will consider ALL sequences in the protein surrounding the given residue, as it has no information about which of these are truly PTM sites. This is typically the desired behavior when doing motif discovery, as the motifs then represent the sequences that are "responsible" for some sites being modified

The Motif Background tab allows the user to define the set of proteins from which the application will calculate the Background Percentages. The following choices are offered:

- **Use Identified Proteins Only** -Scaffold PTM calculates a Background Percentage based on only the proteins loaded into the application.
- **Use Fasta database** (More robust) - Scaffold PTM calculates the Background Percentage based on a FASTA database specified by the user.

Figure 4-6:



System

This tab allows the user to adjust a number of system-related settings:

Memory Usage

This control sets the maximum amount of memory that the system may allocate to Scaffold PTM. It is recommended that the memory allocation be set to a little less than the physical RAM available on the system. The first field accepts a number while the second provides a drop-down menu from which the user may select the units. .

- *The new memory setting will take effect only after the application has been closed and restarted.*

Number of Processors

This control allows the user to choose the maximum number of processors to be assigned to Scaffold PTM for computations. The default value is the maximum number of processors present in the system where the application is installed.

Internet Settings

This tab allows the user to enter a proxy server name or an IP address and a proxy port number. Through the check boxes in this dialog, the user may:

- **Allow Scaffold PTM to connect to the Internet** If this box is unchecked, then Scaffold PTM cannot access the Internet. A user might want to uncheck this box if organizational rules prevent connection to the Internet.
- **Use HTTP Proxy Server**
  - **Proxy Server name (or IP address)**
  - **Proxy port number**

  Proxy servers may be used by an organization's IT departments to filter communications to and from the Internet. If this is the case, the user needs to set the Proxy Server Name and Port Number. To determine whether proxy server settings are needed, a user may examine the way the user's web browser is connected to the internet.

User Interface

This tab contains settings that control the behavior of the user interface.

Messages

Some messages displayed by Scaffold PTM contain a checkbox that says "Do not show this message again." When this has been checked, the program disables the message dialog. This control allows the user to resume display of these messages.

Views

This control allows the user to choose which View will open automatically when new data is loaded into Scaffold PTM.

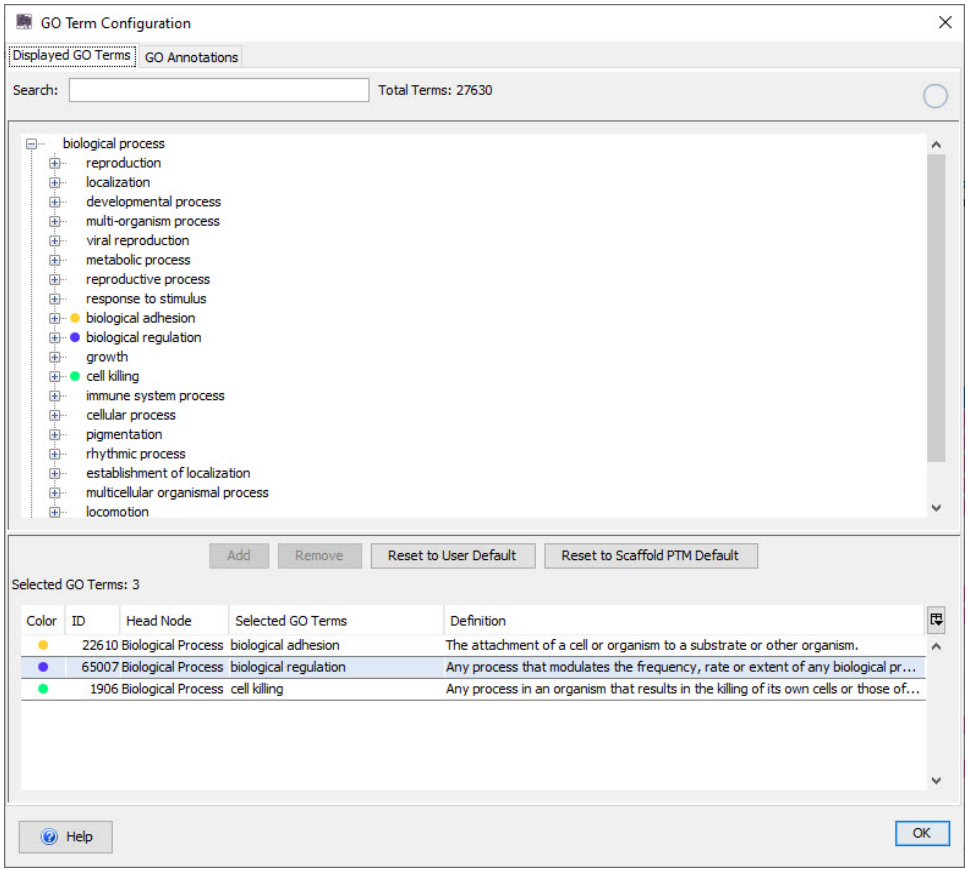
Edit GO Term Options

Selecting **Edit > Edit GO Term Options** from the main menu**,** the **GO Term Configuration** dialog opens. The dialog contains the following tabs:

- The Displayed GO Terms Tab
- GO Annotations Tab

The Displayed GO Terms Tab

Figure 4-7: GO Term Configuration dialog - Displayed GO Terms tab



Through this tab the user can create and modify a custom list of GO terms. The list is then displayed as extra columns in PTM List table whenever the terms are present in the experiment.

The **Display GO Terms** tab is divided into the following sections:

- **Search Field** - Searches GO terms available in the gene_ontology.obo file found in the parameters folder of the Scaffold PTM installation directory.
- **GO Tree list** - Hierarchical list of all available GO terms.

- **Add and Remove GO terms** - Provides tools for creating the custom Display list
- **Display List** - List of GO terms selected by the user that will be visible in PTM List table.
- **Save and Apply**- Allows the user to save the current Display List if changed

The user can create a new custom GO terms Display List by following these instructions:

1. If the **Display List** is not empty select all the rows and press delete.
2. Search and select any GO term of interest either by typing a name in the **Search Field** or by selecting a row in the **GO Tree List**.
3. Click **Add**; the selected term or group of terms is added to the **Display List**. Terms may be selected individually or by domain or group. If a group or domain is selected, all terms in that group will be added to the **Display List**.
4. To remove terms from the **Display List**, select a term or group of terms to be discarded then click **Remove**.
5. To save the current selections as *User Defaults* check the box **Save displayed GO terms as user default**.

When a Scaffold PTM experiment is saved, the displayed GO terms are saved within the SPTM file.

When a new file is created, or when Scaffold PTM is closed, the list of displayed GO terms is retained. To reset the list to the defaults, the user may click the **Reset to User Default** or the **Reset to Scaffold Default** button.

GO Annotations Tab

Scaffold PTM builds a table containing all GO annotations imported into the program from GO annotation database files. When GO annotations are applied to an experiment, the entire Scaffold PTM GO annotation table is searched for GO terms matching the proteins identified in the experiment.

The GO Annotations Tab contains a table listing the GO annotation databases already imported into Scaffold PTM. When no GO annotations have been imported and no GO database is available, a warning appears and the menu command Experiment>Apply GO Terms does not function. The user may populate the table with existing or custom-created GO term databases through the Import Annotations function.

The GO Annotations Tab also includes a search box that searches the list of imported GO terms.

Import Annotations

The **Import Annotations** button opens a dialog through which the user may import GO Annotation databases in Scaffold PTM. A pull-down menu directs Scaffold PTM to different locations from which GO Annotation Databases may be downloaded.

The pull-down list includes the following items:

- **Human Only** - provides a download of the human subset. It takes about 10 minutes to download.
- **Other Website** - the user can type in a website address from where a GO Annotations Database can be downloaded.
- **Other File** - the user can direct Scaffold PTM to a location in his/her computer where the GO Annotations database is stored.
- **Individual Taxonomy GOA Files** from //ftp.geneontology.org:/pub/go/gene-associations/

Selecting one of these options and clicking **Add** imports the selected GO Annotation database into Scaffold PTM and adds its name to the list of loaded databases

---

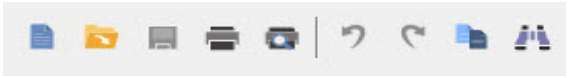
Because the full download of the Gene Ontology Database for all proteomes has grown so large, it is no longer included in the drop down list. Before loading the full database into Scaffold PTM, please check that the system's temp directory contains sufficient free space and note that the process may take many hours.

Clicking **OK** closes the dialog. The user may then annotate the protein list in the PTM List table with GO terms by choosing the now available option **Experiment > Apply GO Terms.**

The command Experiment > Apply GO Terms is available for use only when one or more GO Annotations databases are loaded into Scaffold PTM.

Tool-bar

Figure 4-8: Scaffold PTM tool bar



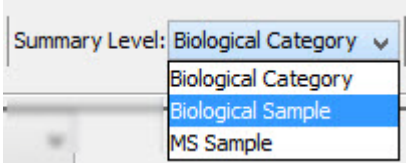
The Scaffold PTM tool bar contains icons that represent equivalent commands for frequently used main menu options.

Icon	Function
	New—Starts a new experiment by opening a file browser to locate MZID data file to be loaded in Scaffold PTM. See “Loading data into Scaffold PTM” on page 38
	Open—Opens a saved Scaffold PTM experiment file, an SPTM, through a file browser.
	Save—Standard Windows behavior.
	Print—Prints the current view.
	Print Preview—Previews current view with the option of printing the document.
	Redo—reapplies an operation that has been “undone”.
	Undo—reverse the last operation that has been done.
	Copy—For the current view, copies to the clipboard the first table appearing at the top of the view. The user can then paste it into a third-party program such as Excel or Microsoft Word.
	Find—Opens a find dialog box that searches the first table present in the current view, see “Find Dialog” on page 48.
	Excel—Exports the information that is contained in the current view to a CSV text file that can be opened and viewed in Excel.
	Help—Opens the Scaffold PTM Online Help.

Summarization bar

The summarization bar provides an easy way to switch between different levels of summarization of the MS samples loaded in the experiment. Data will be collapsed or expanded to the selected hierarchical level. The Summarization bar operates through a drop down menu containing a list of Attribute groups hierarchically ordered. While the Summarization bar is accessible throughout the program, it operates only on the values in the Quantify View.

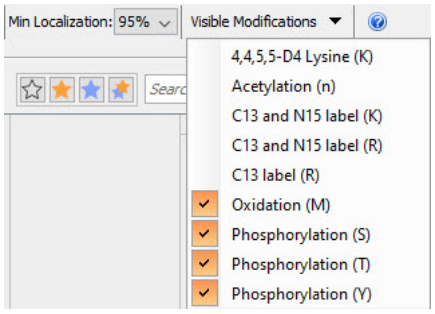
Figure 4-9: Scaffold PTM Summarization bar

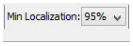


Scaffold PTM Main Window Filters bar

The Scaffold PTM Filters bar is located under the main menu bar at the right side of the Summarization bar. It contains two filters.

Figure 4-10: Scaffold PTM Filters bar



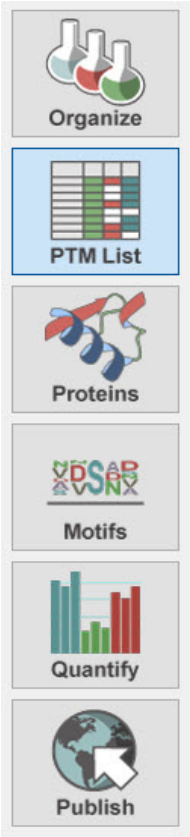
Icon	Function
	**Min Localization**—Filters PTM sites according to their Localization Probability.
	**Visible Modifications**— toggles visibility of modifications in all views.

Note that spectra whose modification sites do not meet the Min Localization threshold or whose modification types are not marked as Visible are still displayed in the PTM Modification Sites table in the Proteins View.

Modification Sites that do not have at least one spectrum which meets the Min Localization threshold or whose modification types are not marked Visible are removed from the Protein Sequence, however. Counts in the PTM List table are also adjusted based on these criteria.

Navigation bar

Figure 4-11: Scaffold PTM Navigation bar for View selection



The Scaffold PTM Navigation bar is a vertical bar displayed on the left side of the Scaffold PTM main window.

The bar contains buttons that toggle the six available views in the Scaffold PTM main window:

- The Organize View, see “The Organize View” on page 69.
- The PTM List View, see “The PTM List View” on page 73.
- The Proteins View, see “The Proteins View” on page 81.
- The Motifs View, see “The Motifs View” on page 93.
- The Quantify View, see “The Quantify View” on page 103.
- The Publish View, see “The Publish View” on page 117.

Display pane

Scaffold PTM's Display pane shows the View selected in the Navigation bar. Each View consists of one or more tables or graphs in one or more sub-panes. All panes and tables in Scaffold PTM's Display pane share certain characteristics:

Table Features

All tables in Scaffold PTM include the following tools:

- Tool Tips
- Resizing of columns and panes
- Column Control
- Moving columns
- Column sorting feature
- Multi selection of rows

Tool Tips

The user may view information about fields or columns in a View by hovering over the location of interest with the mouse pointer. Pressing F2 opens expands the displayed tool-tip and allows the user to copy the information contained in it. Pressing the Escape (ESC) key on the keyboard closes the expanded tool-tip.

Figure 4-12: Viewing information in a tool-tip

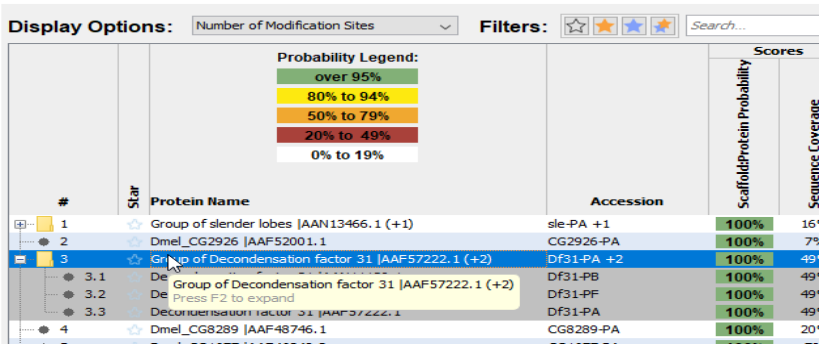
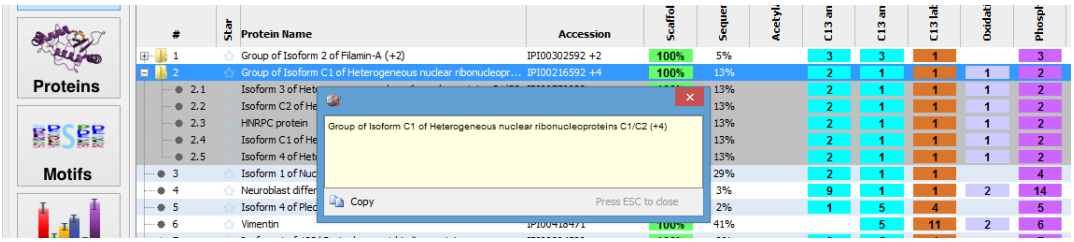


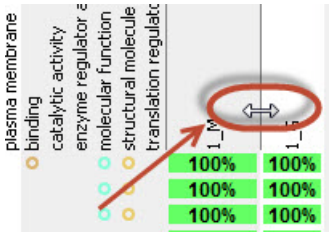
Figure 4-13: Viewing information in an expanded tool-tip



Resizing of columns and panes

The user may resize columns and different panes in each of the views. For example, in the Samples View, the width of a column may be changed by resting the mouse pointer on the right side of a column header until the pointer changes to a double-headed arrow, and then dragging the boundary until the column is the desired width.

Figure 4-14: Changing the width of a column in the Samples View



Column Control

All tables throughout Scaffold PTM have a feature called Column Control. It is a vertical button located to the right of the column headers. When the user clicks this button, a drop down menu containing a list of all columns in the table opens. Each column name has a check box and at the bottom of the list are three group commands.

Figure 4-15: Column Control button

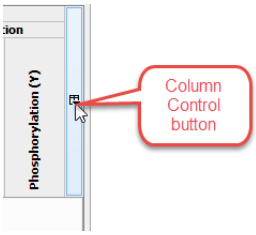
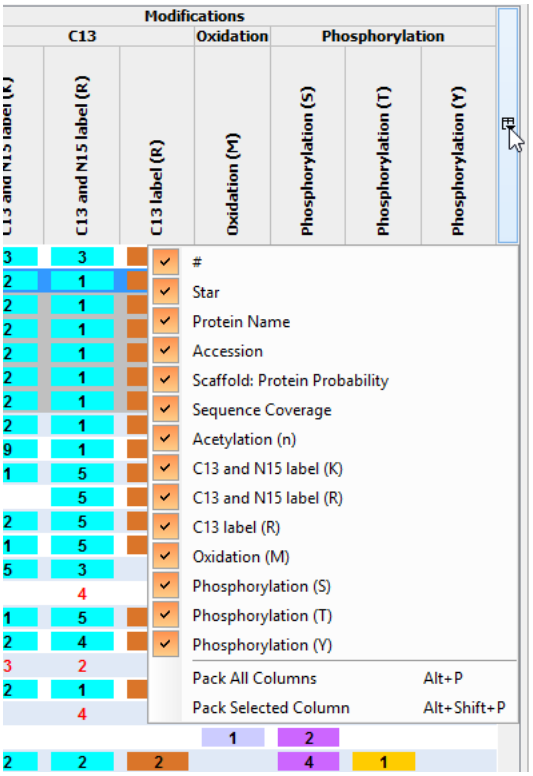


Figure 4-16: Column Control expanded menu



Unchecking columns from the list will hide them in the table. Note that columns can also be hidden by right clicking over the heading of a column and selecting the “Hide Column” option that appears. Columns must be hidden one by one.

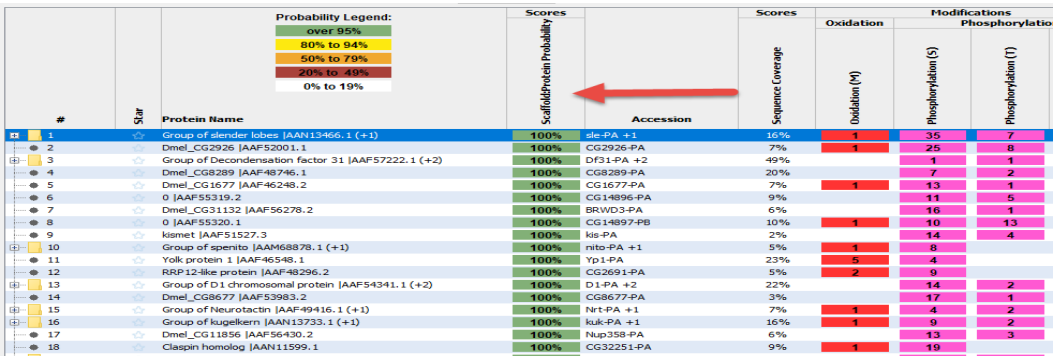
The Horizontal Scroll command, if checked, will add a scroll bar at the bottom of the table. Pack all columns, when selected, resizes each column to the width of the longest value in the column. Pack selected column is active when a specific column has been selected in the table before opening the Column Control.

Moving columns

In all tables throughout Scaffold PTM, every column can be moved from one position to another.

To move a column, click on the header of the column and drag it to the new location. The new column order will be maintained while switching views.

Figure 4-17: Moving columns in tables



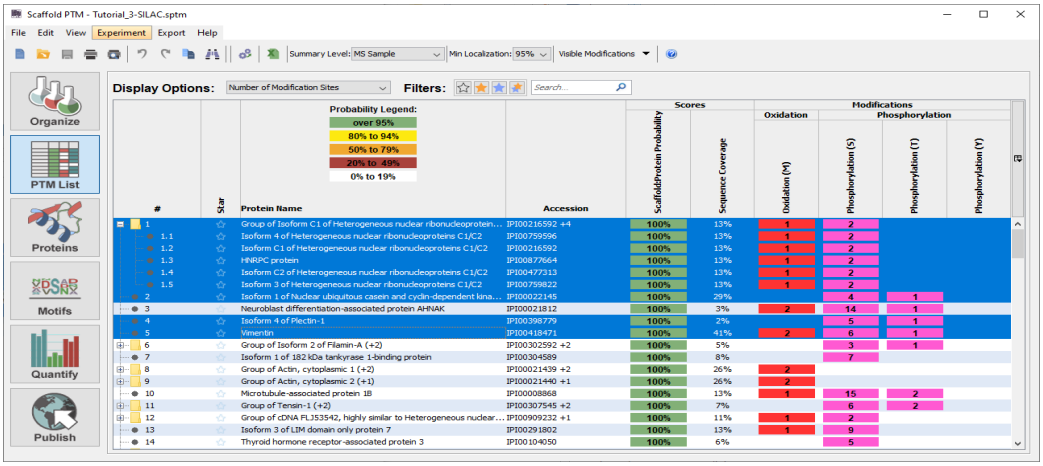
Column sorting feature

In all tables throughout Scaffold PTM, clicking on any column header activates a tri-state sorting function. For example, to sort the proteins based on increasing molecular weight, click the Molecular Weight column header once. To sort the proteins based on decreasing molecular weight, click the Molecular Weight column header twice. To return to the default display, click the Molecular Weight column header a third time.

Multi selection of rows

In all tables throughout Scaffold PTM, the user can select multiple rows by using either the SHIFT or the CTRL key, depending on whether or not the rows in the desired selection are contiguous, and a mouse-click. Other functions may then be applied to all selected rows simultaneously.

Figure 4-18: Selecting multiple rows in a table



Graph Features

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

- **Zoom Function** - Holding down the left mouse button and dragging the pointer from left to right zooms in on a graph. In some graphs, a two-dimensional area may be enlarged by

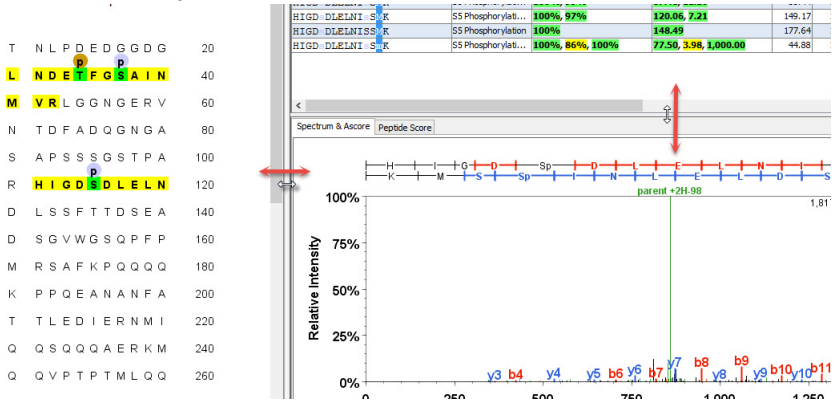
holding down the left mouse button and dragging to the right and down. Clicking anywhere in the graph returns the graph to the previous magnification level.

- **Context menu** - The user can right-click on a graph to open a context menu. The type of context menu might depend on the view in which the graph appears.

Pane Features

A view may contain one or more panes. The individual panes can be expanded or contracted by clicking and holding the mouse over their top or side edges until a double head arrow appears and then sliding the mouse either up and down or right and left, see Figure 4-19.

Figure 4-19: Pane size adjustments



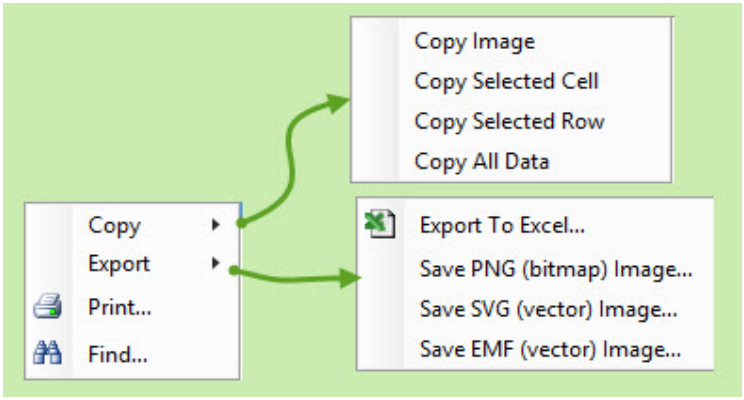
Mouse Right Click Context Menus

When the user right clicks the mouse while hovering over the Display Pane, a context menu with various options appears near the mouse pointer. The list of options available in the context menu varies depending on the selected View.

Organize View

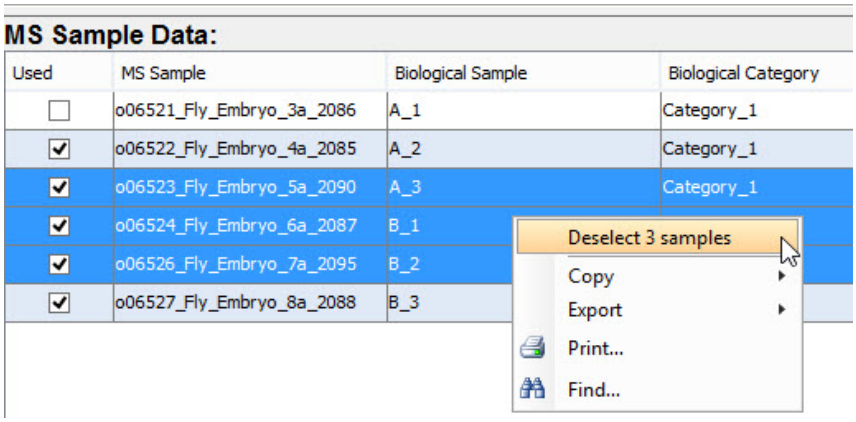
This view has two slightly different context menus depending on which pane in The Organize View the mouse hovers on. Context Menu A appears when the user right clicks in the Loaded Files pane.

Figure 4-20: Context Menu A



When right clicking in the MS Sample Data pane, Context Menu B becomes available. This menu includes an extra command that allows the deselection of the “Used” check box in all highlighted rows.

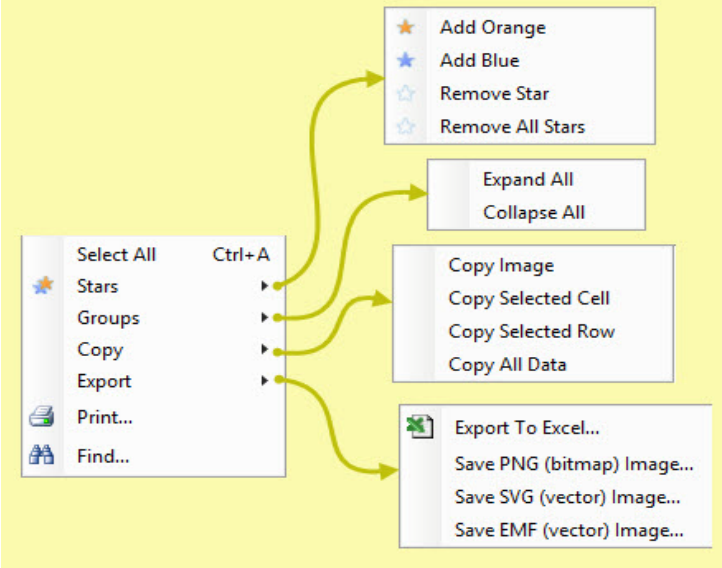
Figure 4-21: Context Menu B:



PTM List View

When the user right clicks anywhere in the Mod List table, Context Menu C: appears. The menu contains a number of different sub-menu options:

Figure 4-22: Context Menu C:

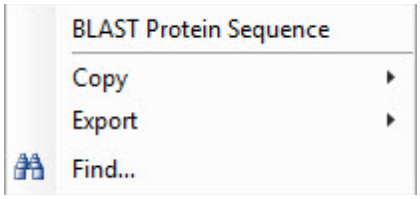


Proteins View

When the mouse hovers over the different panes in this view, a right click opens similar context menus, but with slight differences.

Sequence Coverage Pane > Sequence tab - Available context menu: Context Menu D:

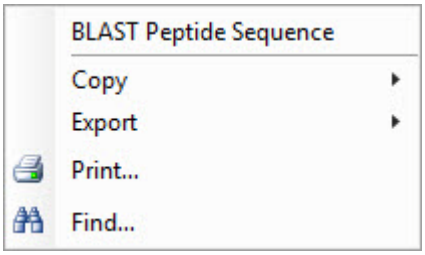
Figure 4-23: Context Menu D:



The same sub-menus are available as in Context Menu A.

- PTM Sites Pane -- Available context menu:

Figure 4-24: Context Menu E:

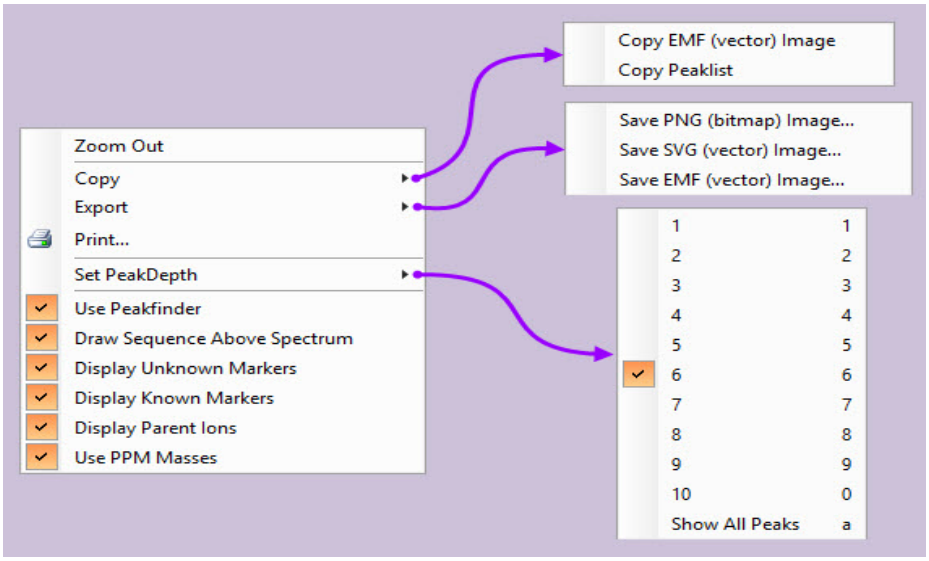


The same sub-menus are available as in Context Menu A.

- The Ascore Algorithm Pane - Available context menus:

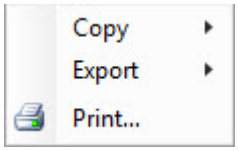
- Spectrum and Ascore tab - Context Menu F: appears when right clicking on the spectrum.

Figure 4-25: Context Menu F:



- Spectrum and Ascore tab - Context Menu G: appears when right clicking on a Bar graph:

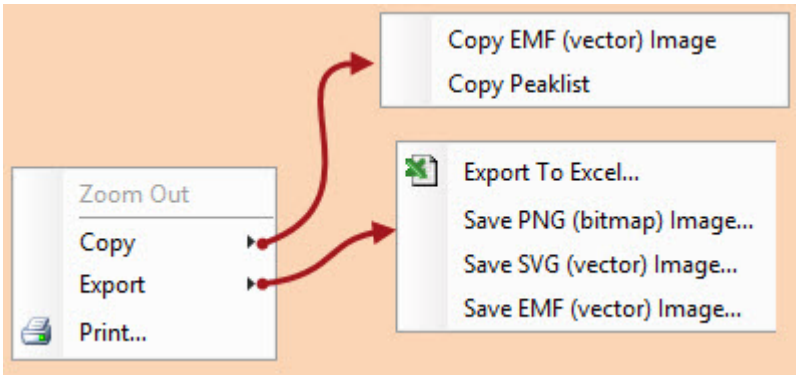
Figure 4-26: Context Menu G:



The same sub-menus are available as in Context Menu A.

- Peptide Score tab > Modification name tab - Available context menu:Context Menu H:

Figure 4-27: Context Menu H:



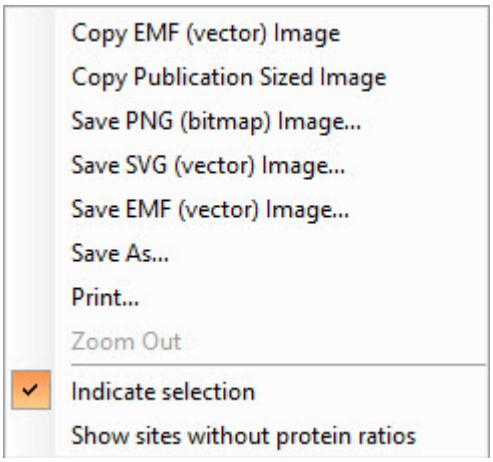
Motifs View

When hovering over the tables included in this view, a mouse right click opens Context Menu A. Clicking over the Motifs representation pane opens Context Menu G:.

Quantify View

Right clicking in the tables in the PTM Spectrum Counnts tab, The Peptide Spectrum Counts tab, the PTM Quantitation tab, and the Peptide Quantiation tab opens Context Menu A. In the Quantitative Charts tab, the context menu available is the following:

Figure 4-28: Context Menu I:



Publish View

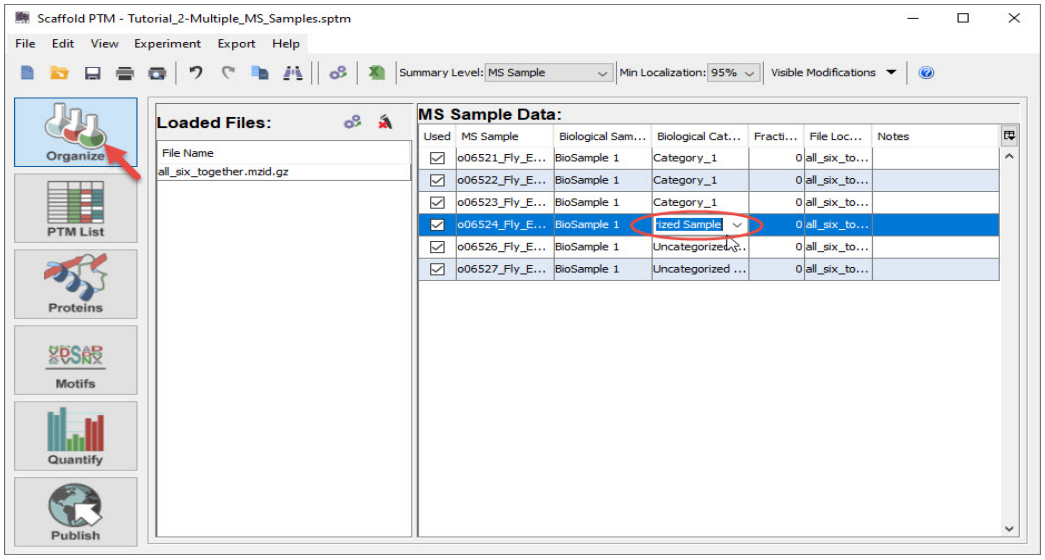
In the text pane in this view, a mouse right click opens the following context menu

Figure 4-29: Context Menu J:

Chapter 5
The Organize View

Scaffold PTM's Organize View appears after a new analysis is created or when a previously saved SPTM is opened. It provides the ability to group the loaded MS samples into Biological samples, define fractions and select or deselect MS samples for inclusion in the PTM analysis.

Figure 5-30: The Scaffold PTM Organize View



The Organize View contains:

- **The Loaded Files pane** -- which lists the loaded MZID data files.
- **The MS Sample Data pane** -- which shows more detailed information about the MS data in each loaded file.

Loaded Files pane

The Loaded Files pane contains a list of the MZID files loaded in the application and a tool bar placed on the top right side of the pane.

The tool bar contains two operational icons: the Add  and Delete  icon.

Clicking **Delete** excludes the selected files from the list.

A click of the **Add** icon opens a file browser allowing the user to navigate to and select MZID or SQML data files for loading. Once one or more files are selected and the **Open** button is clicked, the Queue Data Files dialog opens. This dialog offers the option to add more files to the loading file list.

When all desired files have been selected and analyzed, the table to the right in the MS Sample Data pane is populated with a list of the MS samples loaded from the input files.

MS Sample Data pane

This pane contains a table with information related to the MS files represented by the original MZID files loaded into Scaffold PTM. The program assigns the following parameters, shown as columns, to each MS Sample:

- **Used** - When checked, the MS Sample is included in the analysis, otherwise it is not. The default value is checked.
- **MS Sample** - Name of the MS sample, the program assigns a default value that can be edited.
- **Biological sample** - Name of the Biological Sample, the program assigns a default value that can be edited.
- **Biological category** - Name of the Biological Category, the program assigns a default value that can be edited.
- **Fraction #** - Number of the fraction corresponding to the MS sample. The cell is editable but accepts only numerical characters.
- **File Location** - Original location of the loaded file containing the spectral data for the MS sample.

Notes - This column allows the user to enter additonal information. If an error occurs during the load and analyze phase, Scaffold PTM will report it in this column.

It is possible to edit all of these assignments except the file location. Double clicking in one of the writable cells activates a cursor with all of the usual character editing features.

Many MS Samples can belong to a single BioSample and a Category may include multiple BioSamples. By modifying the names in the table, it is possible to specify the organization of the proteomic experiment as it was initially envisioned.

Queue Data Files dialog

Figure 5-31: Queue Data Files dialog

When a new experiment is created, either by selecting **File > New** or by clicking on the icon **New**, a file browser opens to let the user easily locate and select the MZID or SQML files to be loaded. Once selected, the file or files appear listed in the **Queue Data Files...** dialog and are ready to be loaded. The dialog provides options for adding more files to be analyzed in Scaffold PTM.

When the **Add More Files** button is clicked, the file browser is displayed again, allowing the user to locate more files to be loaded. Selecting the files and then clicking open, adds them to the existing list of files to be loaded. At this point the user can add more files, cancel the whole operation or initiate the loading procedure.

NOTE: It is not possible to delete specific files from the loading list. On the other hand, once the files are loaded, the user can select and delete one or more files from the **Loaded Files:** list by clicking the delete button in the Loaded Files pane.

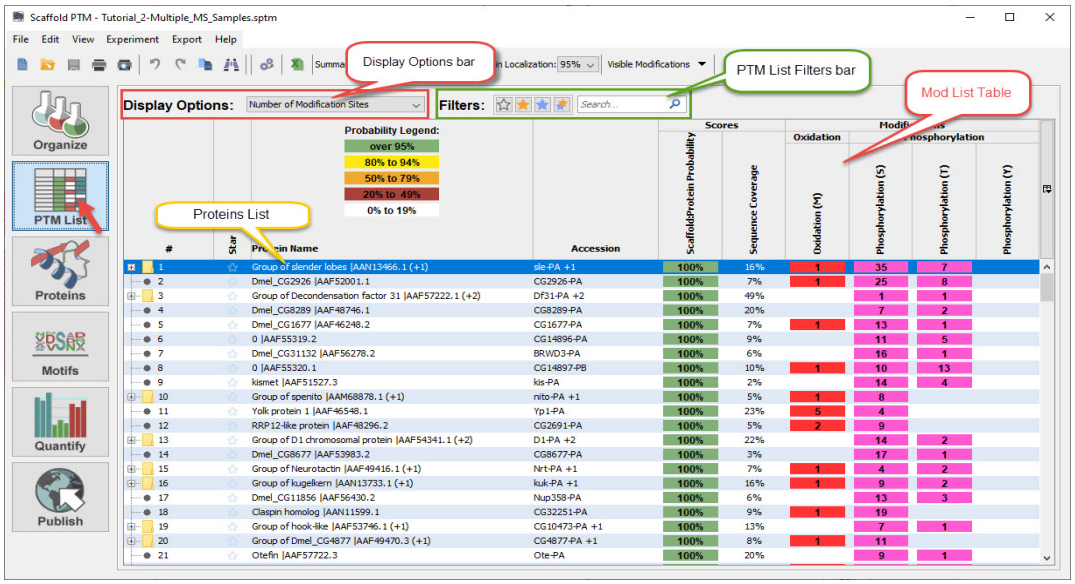
Chapter 6
The PTM List View

With the PTM List View, Scaffold PTM provides a protein-level look at Post Translational Modifications as they are distributed among the analyzed samples.

The PTM List View includes the following tools:

- **The Mod List Table** -- A table showing, for each protein or protein group in the list, counts of the Post Translational Modification (PTM) detected in the experiment. See "The Mod List Table" on page 74.
- **The Proteins list** -- A list of the identified proteins in the loaded files, see "The Proteins List" on page 76.
- **The Display Options: bar** -- A pull-down list of different counting options for displaying PTM abundance, see "Display Options bar" on page 79.
- **The Filters bar** -- Which provides tools to search and conveniently filter the protein list, see "Filters bar" on page 80.

Figure 6-1: The Scaffold PTM's PTM List View



The Mod List Table

In the Mod List Table, each row displays a protein group, with columns showing abundance values for each modification present in the protein group.

The **Display Options:** pull-down menu located in the Display Options bar above the Mod List Table, allows the selection of the type of count used to measure the abundance of the modifications reported in the table.

The user can apply thresholds and filters to the Mod List Table using the tools offered in the Threshold bar and the Filters bar located above the Mod List Table.

Samples Table features:

- "The Mod List Table Features" on page 74
- "Initial Filters applied" on page 75
- "Default Sorting of Columns" on page 75
- "Summary Level and the Mod List Table" on page 75
- "Color Legend" on page 75

The Mod List Table Features

Like any table in Scaffold PTM, the Mod List Table includes the features and tools described in "Display pane" on page 59.

The first four columns, initially ordered as shown below, appear in every experiment and provide the following information:

- **# --** Order number of each row at the initial ordering conditions.

Star -- Initially shows an empty star for every row. Clicking on a star changes its color, and clicking multiple times causes the star to loop through four possible states. The color goes from gray to orange, to blue, to orange and blue, then back to gray. Having multiple star states allows the user to group and filter proteins in complex ways . For more information see "Tagging Proteins of Interest, the star function" on page 77.

- **Protein Name** -- Protein or protein group name.
- **Accession** -- Protein identification number. The particular type of ID shown in this column depends on the parsing rules applied to the protein accession numbers during the database search or the Scaffold analysis.

The remaining columns are arranged into two groups:

- **Scores** -- Typically includes columns showing the main search engine scores associated with the protein or protein group. For example when Scaffold data is loaded, the columns Protein probability and Sequence Coverage are shown.

- **Modifications** -- This group includes a column for each variable modification included in the database search. Cells displaying abundance values are color coded according to the modification type.

The order of the columns can be changed and columns may be hidden. For instructions, see "Display pane" on page 59.

Initial Filters applied

When the PTM List view is selected for the first time after file loading completes, the default values of the thresholds and filters applied to the Modifications are set as follows:

- Min. Localization: 95%
- Visible Modifications: All

For more information about theses filters see "Scaffold PTM Main Window Filters bar" on page 57.

Default Sorting of Columns

When the PTM List View is visited for the first time, the protein list is sorted by:

1. Decreasing protein spectrum count.
2. Decreasing alphabetical order of the accession number

The tri-state sorting feature, activated by clicking on any column header, sorts the data according to the selected column. The first click sorts in ascending order, the second click in descending order, and clicking three time restores the initial order.

Summary Level and the Mod List Table

Adjusting the Summary level does not affect the Mod List table. It does, however affect the Quantify View.

Color Legend

The color legend appears only when data is analyzed by Scaffold and contains protein probabilities. It is located at the top of the Mod List Table in the Protein Name column header and it defines the color coding associated with the Scaffold Protein Probability.

The Proteins List

Proteins which share peptides can be displayed as groups using the command **Experiment > Use Protein Grouping.** This option toggles between grouped and individual protein modes, see "Representation of Protein Groups" on page 76. The application also offers tools to filter the list of proteins.

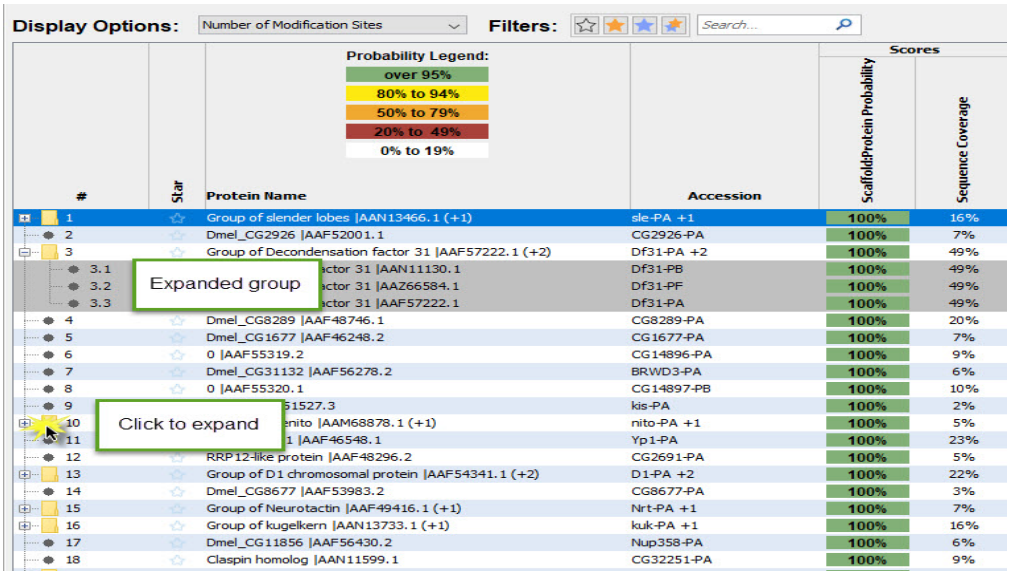
The following information may be useful when examining the Proteins List:

- "Representation of Protein Groups" on page 76
- "Proteins highlighted with red characters" on page 77
- "Tagging Proteins of Interest, the star function" on page 77
- "Applying filters to the Proteins List" on page 78

Representation of Protein Groups

If the grouping option is selected, see "Main menu commands" on page 46, each entry in the protein list will represent a protein group. Scaffold PTM merely displays protein grouping information read from the input files. Structures and names of protein groups are typically created by search engines or by Scaffold according to their grouping algorithms.

Figure 6-2: Protein groups in the PTM List View



If the group includes more than one protein it can be expanded or collapsed for ease of inspection, thus reducing the number of independent rows in the list. When a row displays a protein group with more than one protein, a clickable icon showing either a + or a - sign provides the ability to expand or collapse the group, see Figure 6-2.

- The icon is a + sign when the group is collapsed. To expand the group, click the icon and the sign shown will become a - sign. Clicking the minus sign will collapse the group.When the group is expanded the list of its proteins is visible and highlighted in

dark gray. Each row in the list is numbered as a subset of the protein group number. For example, see Figure 6-2, in the expanded protein group appearing in the third row of the Mod List table, we can see three proteins, which are numbered as 3.1, 3.2 and 3.3 and the numbers appear in a hierarchical structure in the # column.

- The accession number assigned to the group is followed by a plus and a number that indicates the number of additional proteins in the group.

Proteins highlighted with red characters

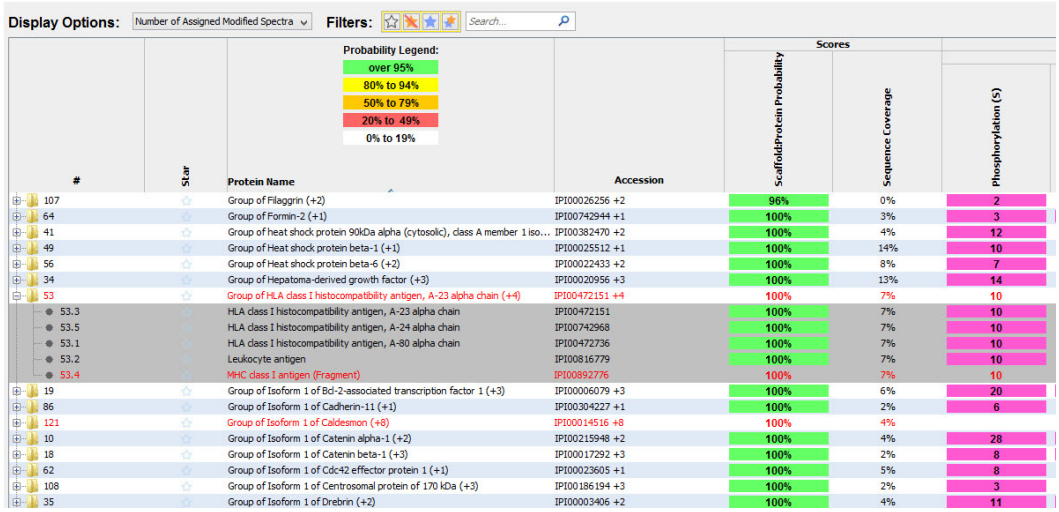
At times proteins or protein groups in the Mod List table appear in a red font. This occurs in two possible circumstances:

1. The protein sequence is not available. For example, this can happen if the data was exported from a Scaffold experiment in which an incorrect FASTA database had been applied, or if an mzIdentML file is imported from another program which does not supply protein sequences.
2. Peptide sequences cannot be localized along the protein sequence.

When this happens Scaffold PTM does not have enough information to assess the Ascores or Localization Probabilities for the modifications in the protein. Modification site counts may also be incorrect, since the program is unable to align the peptides with the sequence and identify overlapping peptides that may contain the same modification site. Highlighting the protein in the PTM List warns the user that for the highlighted protein, Scaffold PTM cannot correctly localize their modifications.

NOTE: a warning also appears in the Proteins View.

Figure 6-3: Example of proteins and protein groups highlighted in red.



Tagging Proteins of Interest, the star function

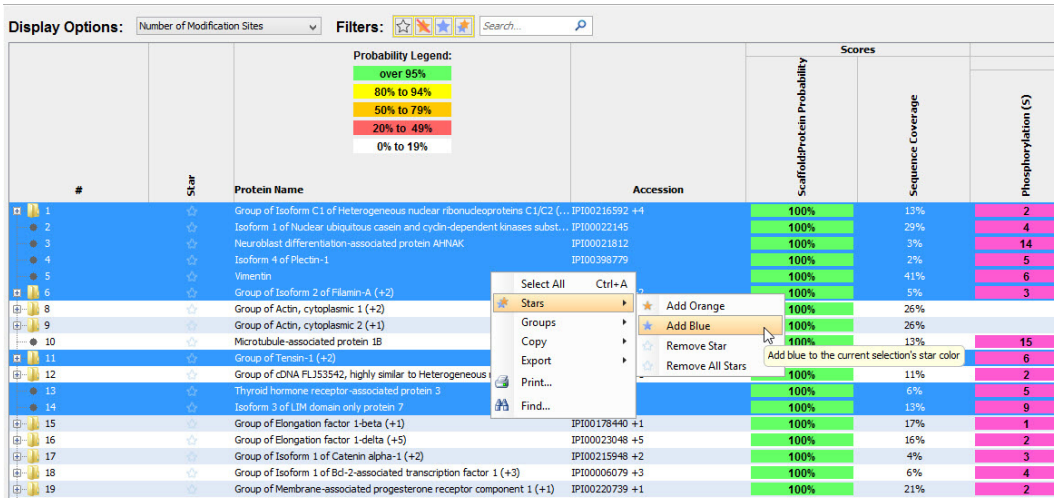
The user can mark proteins that are of special interest by simply clicking the protein star icon

☆ shown in the **Star** column. Two different colored stars, blue and orange, and a combination of an orange and a blue star are available by clicking multiple times on the same star or by selecting the star option in the right click menu.

By using a combination of different stars it is possible to create four different sets of proteins of interest.The user can then bring these proteins to the top of the display by clicking the **Star** column header and can return to the default protein order by clicking the column header twice more. The user can also filter based on stars using the Star filter available in the Filters bar, see "Star Filter" on page 80.

Groups of selected proteins can be starred together by using the star option in the right click menu see Figure 6-4.

Figure 6-4: PTM List View - Starring proteins



Applying filters to the Proteins List

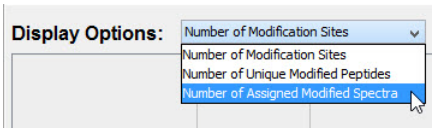
Scaffold PTM provides a Text Search box located in the Filters bar located above the Mod List Table to help the users both reduce the size of the protein list or search for specific proteins of interest, see "Text Search box" on page 80

Display Options bar

The Display Options pull-down list offers a range of statistics for displaying PTM abundance. The values shown in the Mod List Table depend on the Display Options selection.

The **Display Options** drop down list offers the following choices

Figure 6-5: :List of Display Options



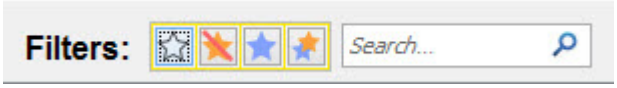
- **Number of Modification Sites** --Displays number of PTM sites in each protein.
- **Number of Unique Modified Peptides --** Displays number of identified peptides in the specified protein that contain modifications.
- **Number of Assigned Spectra --** Displays the non normalized spectra of each protein's PTM sites.

Filters bar

The **Filters:** bar includes two different tools that can be used to reduce the number of proteins shown in the proteins list appearing in the Mod List Table and also tag specific proteins of interest.

- “Star Filter” on page 80
- “Text Search box” on page 80

Figure 6-6: PTM List View: Filters: bar



Star Filter

The Star Filter box contains four toggle buttons. Each button is characterized by one of the four possible star states the user can trigger for a specific protein or group of proteins, by clicking the icon shown under the “Star” column in the Mod List Table. This action is referred to as starring a protein.

Each star filter button has two possible states:

- **Unfiltered** - The star appears in the icon. When a star color is unfiltered, proteins with stars of that color are displayed in the proteins list. Clicking the button changes the status to filtered.
- **Filtered** -The star appears with a red diagonal bar across it. When a star is filtered, proteins with stars of that color are not included in the proteins list. Clicking the star filter button again clears the filter and returns the proteins to the PTM List view.

It is possible to select one or more star filter buttons at the same time. The proteins tagged with the selected stars will be hidden from the proteins list. Selecting the uncolored star leaves only the starred proteins in the list. For more information about how to assign stars to proteins in the PTM List View: “Tagging Proteins of Interest, the star function” on page 77.

Text Search box

The **Text Search box** filters the list of proteins in the Mod List Table, displaying only proteins which contain the string that has been typed in the box. The filter searches for the typed characters in the Protein Name and Accession Number columns in the Mod List Table.

Chapter 7
The Proteins View

Scaffold PTM's Proteins View offers various graphical tools designed to help the user examine the evidence supporting the presence and location of modifications in a selected protein. The view can be reached either by clicking the Proteins icon on the "Navigation bar" on page 58 or from the PTM List view by double clicking in any row in the Mod List table.

The view consists of two controls and four panes, see Figure 7-1:

Controls:

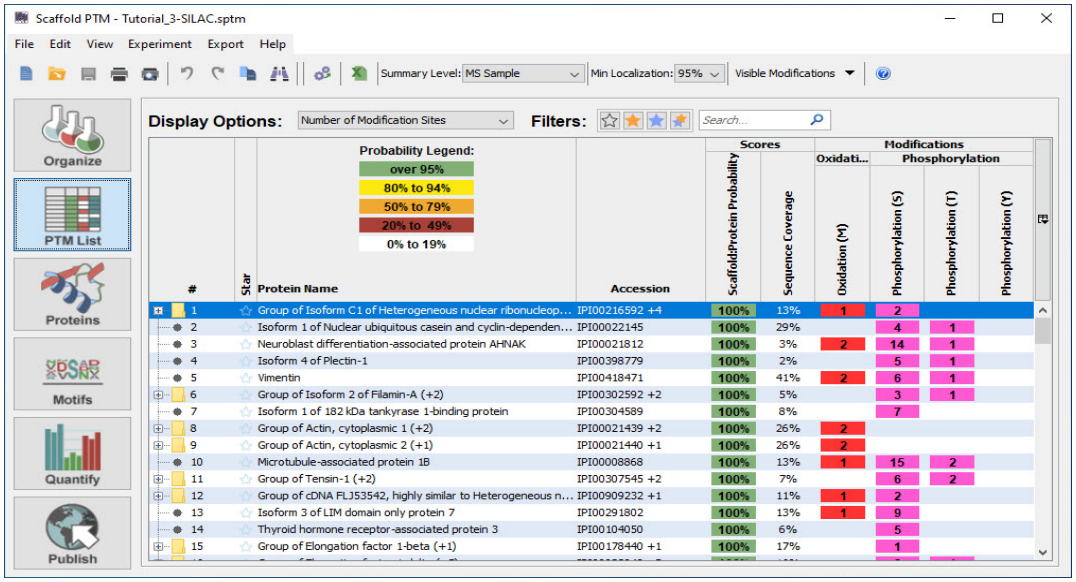
"Protein list" on page 82.

"PTM Sites List" on page 82.

Panes:

- "Sequence Coverage Pane" on page 83.
- "PTM Sites Pane" on page 83.
- " Sequence Pane" on page 85
- "The Ascore Algorithm Pane" on page 89.

Figure 7-1: Scaffold PTM Proteins View

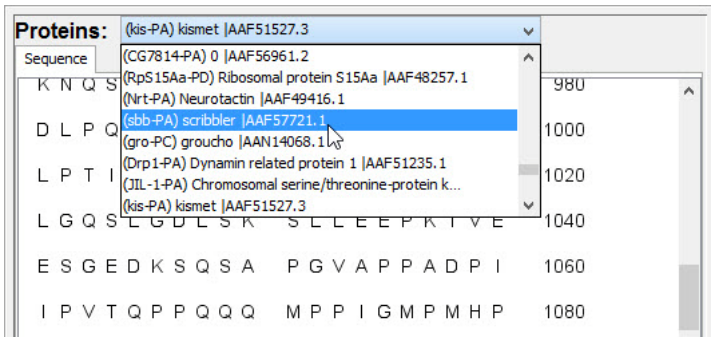


Note that all tables, graphs and panes present in the view include the features and tools described in section “Display pane” on page 59.

Protein list

Through this pull-down list it is possible to select a different protein without switching views.

Figure 7-2: Proteins pull-down list



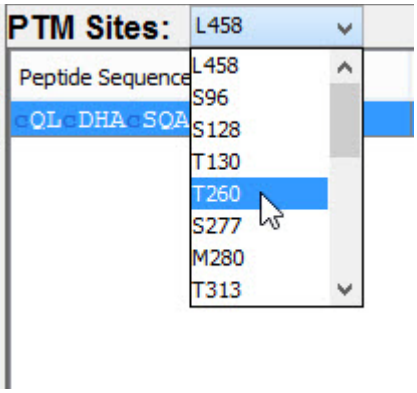
Clicking the currently selected protein name appearing on the right side of the tag **Proteins:** expands the list. Moving the mouse over the expanded list highlights each protein in the list. Clicking selects the highlighted protein and triggers the view to display a different protein. The right scroll bar in the list provides a faster way of going through the whole list of proteins. Once a protein is selected, the information shown in the other panes of the view is updated accordingly.

PTM Sites List

This pull-down menu lists all of the PTM sites identified by the search engine in the selected protein. Each entry consists of a letter identifying the modified amino acid and a number indicating the position of the modification in the protein sequence, along with the type of modification found at that site. Choosing one of these sites selects the peptide containing the amino acid in the Protein Sequence pane and populates the panes in the right side of the Proteins View with localization information for this PTM site.

Selecting a specific amino acid in the Protein Sequence pane also changes the selection in the PTM Sites List. If an amino acid with an identified modification is selected in the protein sequence, the corresponding modification site is selected in the PTM Sites control. If the selected amino acid does not have a modification assigned to it, the amino acid and its position are shown with no modification type.

Figure 7-3: PTM Sites List



Sequence Coverage Pane

In the upper left quadrant of the Proteins View, this pane depicts the sequence coverage of the protein in each sample, as well as the overall sequence coverage for all samples combined. Bars represent the full sequence of the protein. Areas corresponding to portions of the sequence for which peptides were detected are colored. Black lines indicate modification sites. The first bar represents the cumulative coverage including peptides from all samples, and the remaining bars are individual depictions of the coverage in each sample.

PTM Sites Pane

This table lists all peptides containing the selected amino acid and provides information about the modifications that have been identified by the search engine in them. Each row in the table represents a spectrum, and the columns provide information about the peptide-spectrum matches to help the user manually validate the spectra. Most importantly, each row provides the Localization Probabilities and Ascores for the modifications in the peptide represented in the row.

Note that spectra are shown in the PTM Sites Pane even if they do not meet the Min. Localization Probability Threshold or if their modification types are not Visible. Under these circumstances, the modification sites are removed from the protein sequence, but the spectra are not removed from the PTM Sites table.

The columns included in the table are:

- **Peptide Sequence**--Amino acids that carry modifications are designated by a lower-case letter. The selected amino acid is highlighted in each peptide in the table. The total number of highlighted lower-case letters in the table corresponds to the sum of the Number of Modifications identified in the samples in The Mod List Table.
- **Variable modifications**--List of modifications identified by the peptide spectrum. They are designated by the type of modification, the amino acid that carries them and their location along the peptide.
- **Localization Probability**--Probability assigned to the modification presthe peptide. The way the probability is calculated is explained in section "Automated PTM Site

Localization" on page 24.

Ascore-- Ambiguity score for the modification site assignement, see "Automated PTM Site Localization" on page 24.

NOTE: An Ascore of 1,000.00 means that there is no uncertainty in the position of the modification. This happens when there is only one amino acid where the modification could go, for this reason the linear spectrum appearing in the lower part of The Ascore Algorithm Pane", Spectrum and Ascore" tab will not be visible.

Peptide Score--Calculated as part of the Ascore calculation, the peptide score is a probability-based ion matching score. It is the cumulative binomial probability based on the potential number of b and y ions and the number matched. The peptide score reported is the maximal score calculated over all peak depths.

- **Search Engine scores**--one or more columns report search engine scores and the Scaffold peptide probabilities if the MZID were exported from Scaffold.
- **NTT**-- Number of termini consistent with the enzymatic cleavage rules for the enzyme used. NTT stands for number of tryptic termini, but the enzyme need not be trypsin.
- **Mass measurements**:
  - *Actual Mass* -- Peptide mass in Dalton obtained by multiplying the charge to the subtraction of one proton from the observed M/Z.
  - *Observed Mass* -- Mass over charge (M/Z) of the parent or precursor ion measured by the mass spectrometer.
  - *Charge* - Peptide charge.
  - *Delta AMU* -- (Actual Mass - Theoretical Peptide Mass) in Daltons, where the Theoretical Peptide Mass or Calculated peptide mass, is given by the sum of the masses of the amino acid residues in the peptide plus the mass of a water molecule.
  - *Delta PPM* -- (Actual Mass - Theoretical Peptide Mass) in PPM also referred to in the spectrum as the Parent error. It is calculated by dividing the delta mass expressed in Daltons by the Actual Mass and then multiplying by one million.
- **Start** --Peptide start index.
- **Stop** --Peptide stop index.
- **Fixed Modifications** -- List of fixed modifications identified in the peptide.
- **Spectrum Name** -- Name of the spectrum matched with the peptide.
- **MS Sample** -- Mas spec sample that includes the spectrum.

When a row is selected it is highlighted in blue and updates the information appearing in The Ascore Algorithm Pane is updated accordingly.

Sequence Pane

The Sequence Pane shows an overall view of the best scoring PTM assignments highlighted in the selected protein sequence. Color coded schemes are used to convey information about the type of each modification, its localization along the peptide chain and the degree of confidence with which it is localized.

Right-clicking in the Sequence Pane displays a context menu that allows the user to adjust the sequence display. Selecting the first menu item, Display, brings up a submenu that offers three options for displaying the positions of the modified peptides in the protein sequence.

Sequence Display Options

Spectral Coverage Mode

In Spectral Coverage mode, the pane displays the amino acid sequence of the selected protein, its identified peptides, which are highlighted in gray, and its identified PTM sites. The modifications appear as circles above their associated amino acids. The intensity of the color filling the circle reflects the modification's highest localization probability for the particular site. Hovering over the modification triggers a tool tip that contains the coordinates of the amino acid in the sequence, its highest localization probability, and the samples in which it is found.

Figure 7-4: Protein Sequence Tab in Spectral Coverage Mode

Stacked Mode

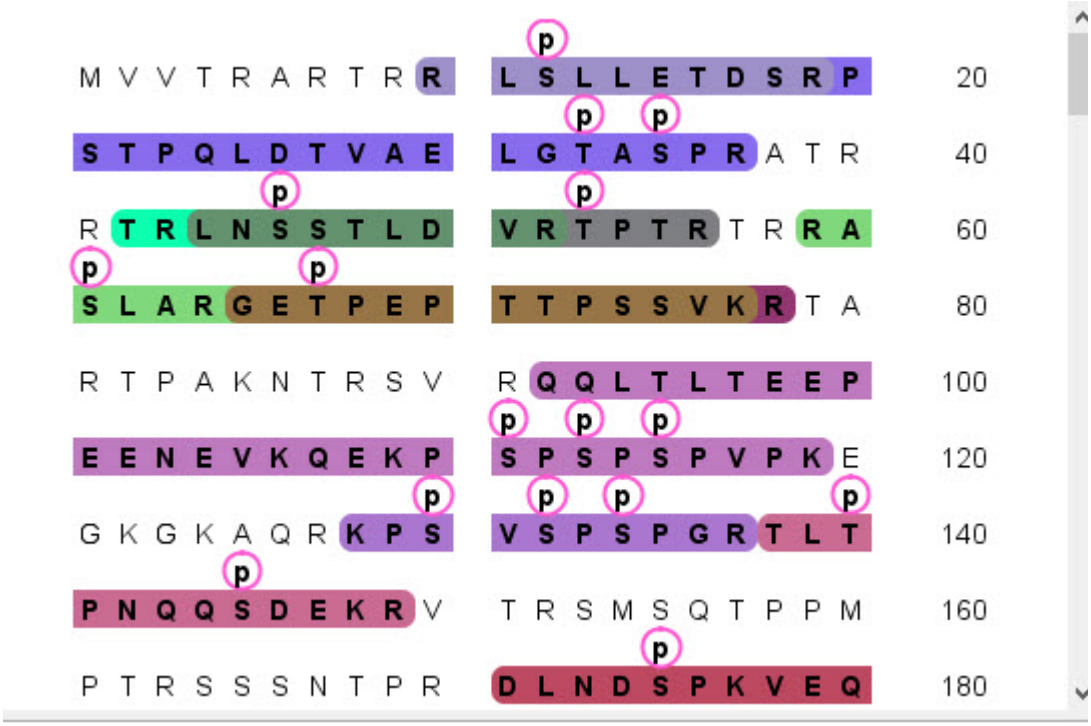
In stacked mode, color-coded bars representing the presence of a peptide in a specific MS Sample are placed above the protein sequence. This mode is helpful in comparing coverage in different samples. The colors correspond to the sample colors in the Sequence Coverage Pane above.

Figure 7-5: Protein Sequence Tab in Stacked Mode

Overlay Mode

In overlay mode, the identified peptides are highlighted with the colors corresponding to the MS Samples in which the peptides are identified. When a peptide is in more than one sample, the colors are blended.

Figure 7-6: Protein Sequence Tab in Overlay Mode



Hovering over an amino acid in the sequence gives its name, position and any modifications it might contain and lists each sample in which that amino acid was identified. When the Spectral Coverage display option is selected, it also shows the number of spectral matches at that point in the sequence.

The peptide selected in the Peptides Pane is indicated by brackets, and if the context menu option Use Blinking Cursor is enabled, it also blinks. Modifications are indicated by colored circles, either solid or outlined depending on another option in the context menu. Each modification is indicated by its own color, and these colors may be selected by the User. Clicking on an amino acid in the sequence display filters the tables above to show only peptides covering that region. Clicking on Show All Peptides clears the filter.

Selecting an amino acid in the sequence highlights the peptide which contains it and triggers an update of the other panes in the view.

- If the selected amino acid is not highlighted, it means that no peptides were associated with it and the rest of the panes are empty.
- When the selected amino acid belongs to a highlighted sequence, the other panes display information about the amino acid and its associated peptides

Whether a modification is displayed in the protein sequence and in the Proteins View in general, depends on the selected Min. Localization probability and on whether the particular type of modification is set as visible or not in the Visible Modifications pull-down list.

Context menu

The User can right-click on the sequence display to open a context menu that has the following options:

- Display - Selects the format in which to display the sequence coverage (see "Sequence Display Options" on page 85).
- Use Blinking Cursor - If Use Blinking Cursor is checked, the cursor blinks, and if Show Cursor is checked, brackets indicate the selected peptide. These two controls operate independently.
- Modification - Offers options to edit the modification colors and to either display modifications as an outlined (open) circle if Outline Modifications is checked or a fully colored circle if it is unchecked.
- Copy Image - 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 Sequence—Copies the protein sequence in text format to the Clipboard so it may be pasted into a third party tool such as Microsoft Word.
- Save PNG... —Saves the currently displayed spectrum in Portable Network Graphic format, which is a bitmap format, and opens the a file dialog box in which you can specify the name and directory for this saved PNG file.
- Save SVG... —Saves the currently displayed spectrum in Scalable Vector Graphic format and opens the a file dialog box in which you can specify the name and directory for this saved SVG file.
- Save EMF... —Saves the currently displayed spectrum in Windows Metafile format and opens the a file dialog box in which you can specify the name and directory for this saved PNG file.
- Print Protein...—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.
- 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.

The Ascore Algorithm Pane

This pane provides the core statistical information used to compute the Ambiguity score (Ascore), based on the Algorithm developed by Beausoleil et al., see http://www.ascore.med.harvard.edu.

The pane contains two tabs:

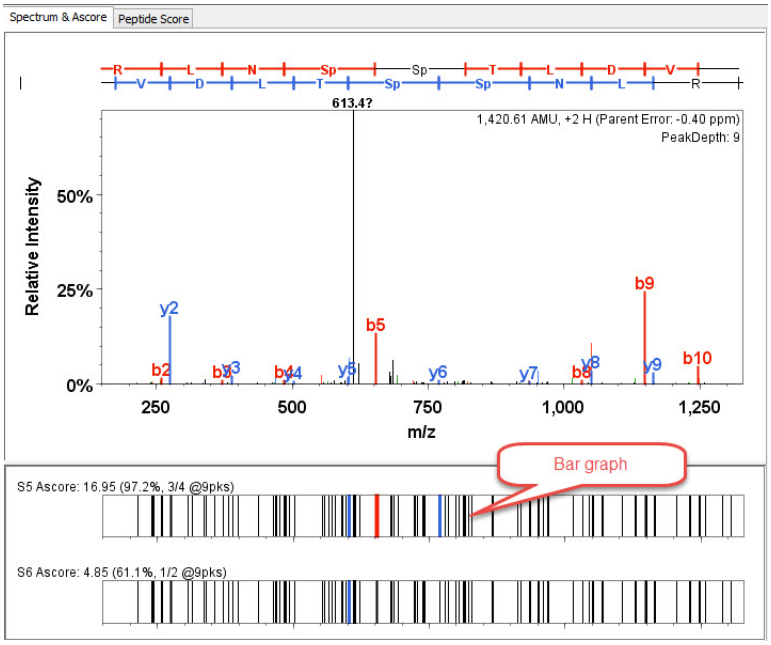
- Spectrum and Ascore tab
- Peptide Score tab

Spectrum and Ascore tab

This tab includes two graphical tools: the spectrum that identifies the peptide selected in PTM Sites Pane and below it one or more Bar graphs consisting of vertical lines that match the peaks in the spectrum. Each graph represents the spectrum at the peak depth which produced the highest peptide score. The lines that correspond to the site-determining ions used for the Ascore calculation are color coded by ion type. Blue lines correspond to the significant y-ions, while red lines correspond to significant b-ions, in the case of CID fragmentation.

The number of bar graphs present depend on the number of ambiguously localized modifications. Above each graph the modification identification and the calculated Ascore are listed along with the localization probability and the number of site-determining peaks present in the spectrum out of the total possible at a particular peak depth.

Figure 7-7: Spectrum and Ascore tab



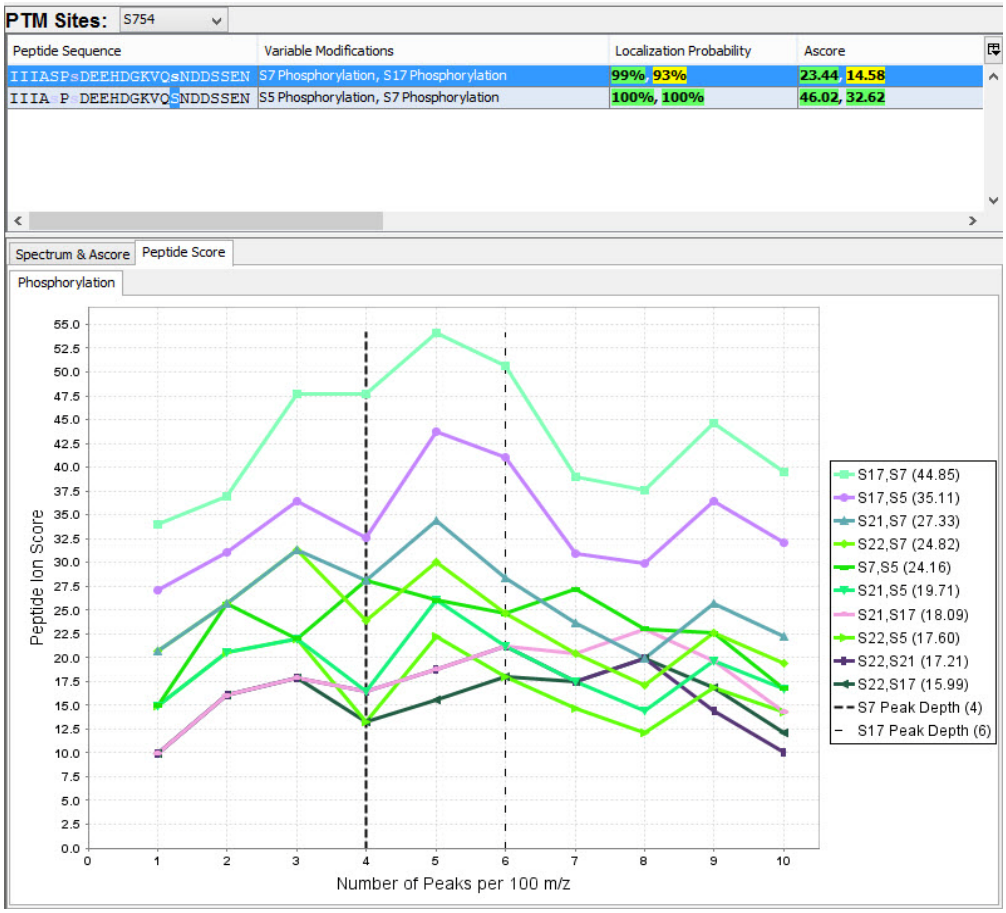
Peptide Score tab

The Peptide Score Tab provides visual confirmation of the calculations of the Peptide Score and the Ascore, which measure a modification's likelihood of being on one amino acid rather than another. If there is more than one ambiguously located modification in the peptide, this tab will contain a sub-tab for each modification

In order to calculate the Peptide Score, the spectrum is first divided into a series of windows. Then, a series of simplified spectra is formed by sampling the original spectrum at peak depths from 1 to 10. To construct a spectrum at peak depth N, the algorithm chooses the N most intense peaks from each window

The modification sub-tab shows a series of 2D plots of the Peptide Ions Score1 (Peptide Score) as a function of the peak depth used to calculate it. There is one plot for each potentially modified amino acid in the identified peptide. The earliest (in case of ties) peak depth that provides the largest separation between the highest-scoring and second highest-scoring modification sites is selected as the optimal peak depth for localization. In the graph, the optimal peak depth is indicated by a vertical dashed line. The legend on the right side of the graph identifies each curve with a corresponding peptide candidate that shows a particular combination of assigned modified amino acids.

Figure 7-8: Peptide Score tab



The Ascore is then calculated by computing the cumulative binomial probability using only the site-determining ions at the optimal peak depth.

Chapter 7
The Proteins View

Chapter 8
The Motifs View

The Motifs View identifies proteins' common sequence patterns or motifs, surrounding modified sites in order to determine the enzymes that interact with the protein at the identified modifications.

This view supports the motif identification approach proposed by Schwartz and Gygi, where the extraction of biologically relevant motifs is based on sequence information from mass spectrometry protein analysis datasets. For a brief description of the algorithm developed in Scaffold PTM see "Motif Identification" on page 26.

The Motifs View can be reached clicking the Motifs icon on the Navigation bar or from the main menu selecting the command **View > Motifs**.

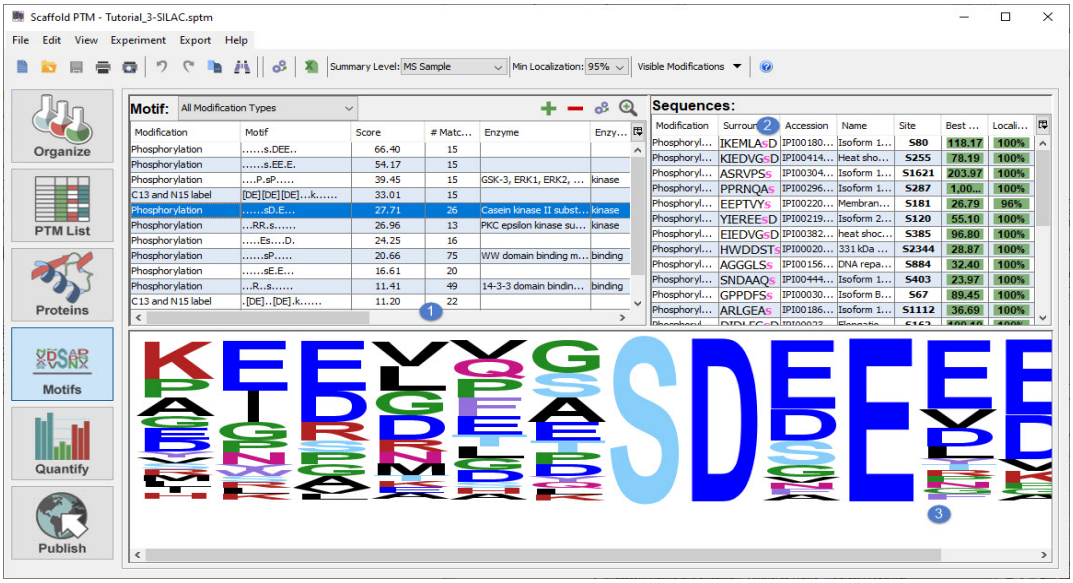
A motif search is run automatically when data is loaded if the motif background setting is "Use Identified Proteins Only" or if a FASTA Database has already been configured for use as a background database. Background settings are changed and FASTA files configured through the dialog opened by selecting Edit>Preferences>Motif Background.

Within the view, the available tools and features are grouped in three different panes, see Figure 8-1:

1. "Motif pane" on page 95
2. "Sequences: pane" on page 100
3. "Motifs representation pane" on page 101

The tables in the first two panes share the same properties as all tables in Scaffold PTM, see "Display pane" on page 59.

Figure 8-1: Motifs View



Motif pane

Scaffold PTM compares the sequences surrounding modified amino acids in the current experiment to the sequences surrounding the same amino acids in a background database. Sequences which appear significantly more often surrounding a modification than they do around the same amino acid in the background are identified as potential motifs. These motifs are reported in the Motif pane. Motifs are assigned scores based on how much more likely they are to appear in conjunction with modifications than would be expected if modification sites were chosen randomly from all possible sites in the background sequences. These scores are shown in the Motif pane, along with the number of modified peptides displaying the motif and the frequency of occurrence in the experiment and in the background. In addition, the Motif pane annotates motifs that have been previously reported in association with specific enzymes.

NOTE: Scaffold PTM can identify motifs based on one of two background database options selected in the “Motif Background tab” on page 49. The option available are: “Use Identified Proteins Only,” and “Use FASTA Database (More Robust).”

Scaffold PTM uses the Human Protein Resource Database in order to identify the enzymes that interact with the modification site associated with a motif.

The Motif pane includes the following tools:

- The Modification filter. - Filters a specific modification type.
- The Motif Tool bar --Provides tools to augment the motif search.
- The Identified Motifs table--Shows the list of identified motifs and related information.

Modification filter

This tool is used to help visualize the motifs and enzymes associated with a specific modification type. The filter is structured as a pull-down list.and includes all modifications identified in the current dataset, both fixed and variable modification. Each entry in the list shows the type of modification, the modified amino acid and the mass of the modification.

When a modification is selected, the tool filters all identified motifs based on the selected modification type. The other panes in the view are influenced by the filter’s selection.

Motif Tool bar

The Motif Tool bar is located at the top right corner of the Motif pane, see number 1 in Figure 8-1. It includes four action icons useful to enhance and customize the motif analysis performed by Scaffold PTM.

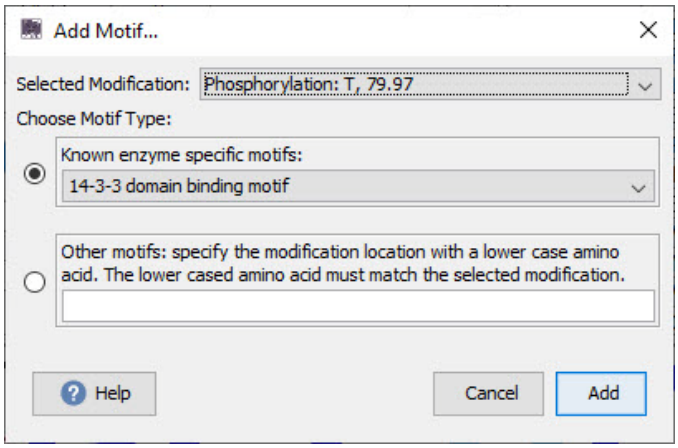
Figure 8-2: Motif Tool bar.

	• **Add Motif** - Adding a specific motif might be appropriate if there is outside information to implicate an enzyme in a modification. This action item opens the dialog Add Motif... which provides tools to specify the characteristics of the motifs the user wishes to add.
	Delete Selected Motifs -If a motif seems to have been improperly identified or defined, deleting it from the motifs list might be appropriate. This action icon when clicked deletes one or more selected motif from the Identified Motifs table. • This action can be undone using the command **Edit > Undo.** • Another way to restore a deleted motif is to rerun the motif search clicking over the Search for Motif icon. **Note:** This last action will NOT restore user-defined motifs, only those motifs that Scaffold PTM had automatically identified originally
	**Search for Motifs** - When the user clicks this action icon or when loading data, Scaffold PTM automatically searches for motifs in the current dataset using the approach described in “Motif Discovery” on page 26.
	**Search for Subset Motif** - This is a new feature introduced in Scaffold PTM 3.0. It searches for motifs present in defined subset of the experiment’s identified PTM site. Clicking this action icon opens the “Add Subset Motif...” on page 97 from which the foreground subset of identified PTM sites can be defined along with the search background.

Add Motif...

The **Add Motif...** dialog allows the user to add Motifs to the identified motifs list. This can be done either by adding a known enzyme-specific motif, which might be appropriate if there is outside information to implicate an enzyme in a modification or by adding motifs by hand.

Figure 8-3: Add Motifs... dialog.



The dialog can be reached by clicking the “Search Subset Motif” action icon located in the “Motif Tool bar” on page 95.

It includes the following components:

- **Selected Modification:** - pull-down list of the modifications present in the experiment
- **Choose Motif Type:** - provides the following two options for adding new motifs to the identified motifs list.
  - *Known enzyme specific motifs* - pull-down list of known motifs
  - *Other Motifs* - Text box for the manual input of a new motif sequence to be searched. NOTE: the input character sequence should have all modification sites lower-case and the flanking sequences should be UPPER-CASE. Bracketing a pair of amino acids, for example [DE], signals that the amino acid present could be either D or E.
- **Action buttons**:
  - *Cancel* - Cancels the operation and closes the dialog.
  - *Add* - Searches for new motifs as instructed and closes the dialog. When no motifs are associated with a specific modification and enzyme pair, a warning appears after **Add** is selected.

When adding a new motif the user should:

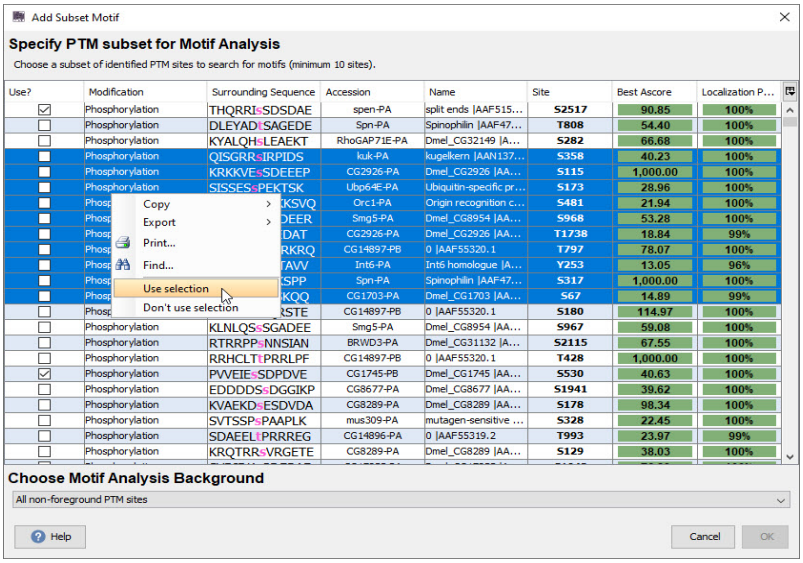
1. Choose the modification of interest from the selected modification list.
2. Either choose a type of enzymatic motif or add it manually.
3. Click Add.
4. A warning will appear if the selected pair does not have associated Motifs.

Add Subset Motif...

The **Add Subset Motif...** dialog allows the user to define motif searches performed with a custom subset of identified PTM sites, defined through selection from the PTM Sites table, as the foreground. An appropriate background dataset is defined through the Analysis Background list. When using "All non-foreground PTM sites" as a background (this only applies to subset motif analysis), the Background dataset will be filtered to include only sites that were identified as having the motif's modification on the correct residue.

Note that it is not possible to search for known motifs in a subset of PTM sites.

Figure 8-4: Add Subset Motif... dialog



This dialog can be reached by clicking the “Search Subset Motif” action icon located in the “Motif Tool bar” on page 95. It includes the following components:

- PTM Sites table
- Analysis Background list

PTM Sites table

The table provides a list of all identified PTM sites in the experiment from which the user can select the set of identified PTM sites of interest by checking appropriate rows, or by selecting several rows, right clicking, and choosing “Use selected” from the popup menu, see Figure 8-4.

Each row in the table represents an identified PTM site with a check box for selection and information to help the user characterize the site.

The list of columns included in the table is:

- **Use?** - Check box. All selected rows constitute the identified PTM sites subset
- **Modification** - Type of modification
- **Surrounding Sequence** - Representation of the modification site, shown in lower case, surrounded by two six amino acid flanking sequences one on each side shown in upper case letters
- **Accession** - Protein accession number where the PTM site is found
- **Name** - Protein name where the PTM site is found
- **Site** - PTM site location and type
- **Best Ascore** - Color coded according to ...

- **Localization Probability** - Color coded according to...

Analysis Background list

This is a pull-down list from which the user can select the background for the motif search. In addition to the option of using identified protein sequences, or (if configured) a FASTA file, when performing a subset motif search, the user may also use the remainder of identified PTM sites as the background dataset. This choice is appropriate for some experimental designs where it is desirable to control for statistical effects of the sample preparation and identification, for example, when combining (modified) proteins of interest with a complex background mixture.

When using "FASTA Database" or "Identified Proteins" as the Background source, Scaffold PTM will consider ALL sequences in the protein surrounding the given residue, as it has no information about which of these are truly PTM sites. This is typically the desired behavior when doing motif discovery, as the motifs then represent the sequences that are "responsible" for some sites being modified.

When using "unselected PTM sites" as a background (this only applies to subset motif analysis), the Background dataset WILL be filtered to only include sites that were identified as having the motif's modification on the same residue as the motif.

The list contains the following three possible Background choices:

- **All non-foreground PTM sites** - The remainder of identified PTM sites
- **Proteins identified in Search** - All proteins listed in The Mod List Table.
- **Fasta database** - Configure Fasta database through the Motif Background tab located in the **Edit > Preferences.**

Identified Motifs table

The table provides a list of all identified Motif searched in the experiment against a background database chosen from the Motif Background tab located in Edit > Preferences. For each identified motif, the table provides information to allow the user to assess the validity and biological relevance of the motif identification.

Note: By default, motifs are listed by their Motif Score from highest to lowest.

For each identified motif the table includes the following properties listed as columns:

- **Modification** - type of modification
- **Motif** - Motif description
- **Score** -
- **# Matches -**
- **Enzyme** -
- **Enzyme type -**
- **Citation -**

- **Dataset % -**
- **Background % -**
- **Foreground Origin -**
- **Background Origin -**

NOTE: Scaffold PTM identifies motifs based on one of two background database options displayed in the “Motif Background tab” on page 49. The option available are: “Use Identified Proteins Only,” and “Use FASTA Database (More Robust).”

Scaffold PTM uses the Human Protein Resource Database in order to identify the enzymes that interact with the modification site associated with a motif.

Sequences: pane

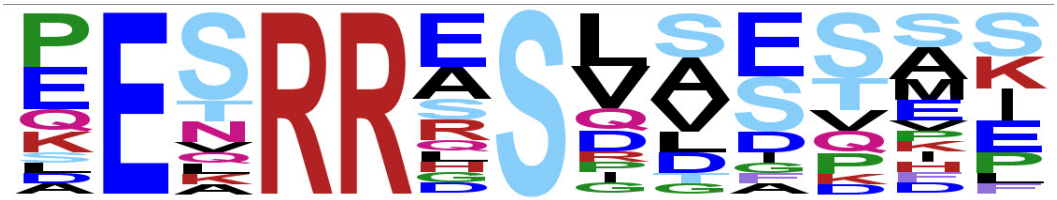
The Sequences Pane details, for the selected motif, each modification and its flanking sequences at the specific site in its protein.

Motifs representation pane

The Motifs Representation Pane creates a graphic representation of the probability that an amino acid might exist in the specific sites of a selected motif. Potentially significant modification trends that Scaffold PTM did not already list may become apparent here.

The visual representation centers the modification of the selected motif (serine, in Figure 8-5) and displays the 12 flanking amino acids in the sequence, 6 to each side of the modification.

Figure 8-5: Motif graphic representation



Scaffold PTM scales each representative letter by the probability that it might exist in the flanking sequence of a motif. The amino acids are color-coded by chemical property, see Figure 8-6.

Figure 8-6: Chemical Properties color code

Chemical Property	Amino Acids	Color
Hydrophobic	AILMV	Black
Basic	HKR	Red
Special AA	GP	Green
Alcohol	ST	Light Blue
Cysteine	C	Yellow
Acidic	DE	Blue
Polar	NQ	Pink
Aromatic	FWY	Purple

Chapter 9
The Quantify View

The Quantify view displays quantitative data for a single protein. The view can be reached either by double clicking on a protein in the The Mod List Table or clicking on the Quantify button on the left side of the Scaffold PTM window. When the Quantify View opens, it respects the protein selection of the other views. A different protein may be selected from the pull-down menu at the top of the pane. Next to the protein selection control is another pull-down that allows the user to select the Display Type.

When an MZID data file is exported from Scaffold or directly from a supported search engine, the quantitative information appearing in Scaffold PTM is based only on label free spectral counting methods and the only active tabs are the PTM Spectrum Count tab and the Peptide Spectrum Count tab, see Figure 9-1.

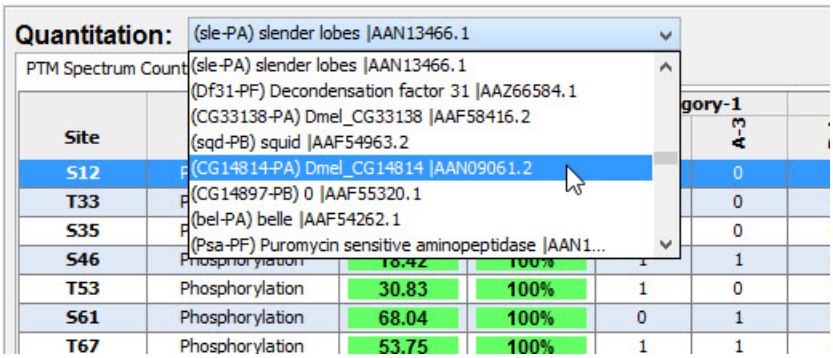
Depending on the type of quantitative data loaded into Scaffold PTM, however, the Quantify View may display three additional tabs. If quantitative data exported from Scaffold Q+ or Q+S in the SQML format has been loaded, Scaffold PTM offers quantitation based on isobaric or stable isotope labeling or on precursor intensity, and the PTM Quantitation, Peptide Quantitation and Quantitative Charts tabs are enabled. Scaffold PTM can also adjust quantitative ratios to account for differences in protein level and give a more accurate measure of differential modification if a ProteinQuant.xml file is loaded into the experiment. This file must be exported from Scaffold Q+ or Q+S and should be derived from unenriched samples. It can be imported through the Experiment>Import Protein Quantitation Results... option. When protein quantitative data is available an additional figure, the Protein/Modsite Scatterplot, is added to the Quantitative Charts tab.

List of tools and tabs in the Quantify View:

Protein List

The Quantify view shows information related to a specific protein. Initially, it respects the protein selection of the other views. The **Quantitation:** pull-down list provides a method for selecting a different protein without returning to the PTM List view.

Figure 9-1: Quantitation: Protein List.



Display Options

This pull-down allows the user to select the type of values to be displayed. The available options depend on the type of quantitative data loaded. The display types and the conditions under which they are supported are:

- Spectral Counting - These methods are available in the PTM Spectrum Counts and Peptide Spectrum Counts tabs, which are available for all types of experiments.
  - Modified Count - the number of spectra with modifications at the specified modification site or in the specified peptide.
  - Modified Count/Total - the fraction of spectra representing the specified modification site or peptide which contain a modification.
  - TIC - these methods approximate the true Total Ion Current by summing all of the fragment ion intensities in the MS/MS spectra. Three TIC methods are offered:
    - Top 3 TIC - sums the three highest TIC values among the spectra representing the modification site or Peptide.
    - Total TIC - sums the TIC values of all spectra representing the modification site or peptide.
    - Average TIC - calculates the mean of the TIC values of all spectra representing the modification site or peptide.

- Ratios - These methods are available only when quantitative data has been imported from Scaffold Q+ or Q+S. These ratios may be derived from precursor intensity, isobaric labeling or SILAC quantitation. They are available in the PTM Quantitation, Peptide Quantitation and Quantitative Charts tabs. Two display formats are available:
  - Log_2 Ratio - displays the log_2 ratio values exported from Scaffold Q+ or Q+S. In the PTM Quantitation Tab, the values represent the median of the log2 ratios of all of the peptides that contain modifications at that site.
  - Ratio - displays the antilog of the values described in Log2 Ratio.
- Protein-Normalized Ratios- These values are displayed when quantitative data has been imported from Scaffold Q+ or Q+S and has been normalized by importing protein quantitative information from a Scaffold Q=/Q+S Protein Quantiation XML Report (see "Creating a Protein-Normalized Quantitative PTM Experiment" on page 40). Protein-normalized ratios appear in the PTM Quantitation, Peptide Quantitation and Quantitative Charts tabs, whose titles change to reflect the fact that values are protein-normalized.

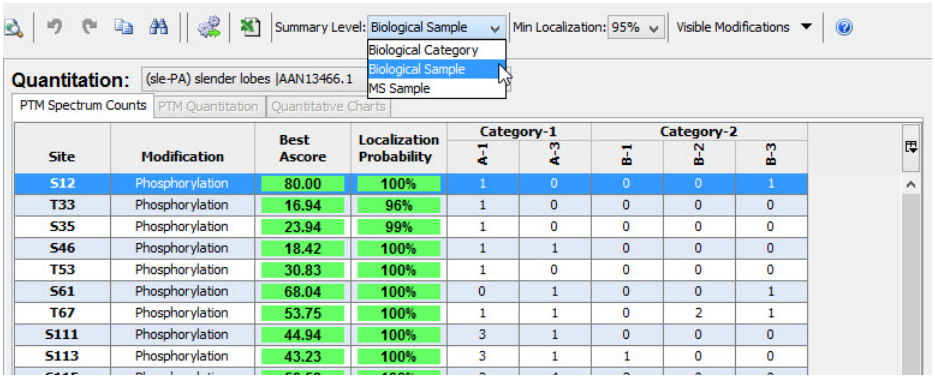
Summarization of Quantitative Values

Scaffold PTM offers the option to view spectral counts on the MS Sample, Biological Sample, or Biological Category levels. The default Summary Level is set to the "MS Sample".

The user may change the level of summarization through the Summary Level pull-down list in the Summarization bar.

When data is exported from Scaffold, the user can define Biosamples and Categories in that application before exporting the MZID. Otherwise the summarization levels can be defined and assigned in the "MS Sample Data pane" on page 71.

Figure 9-2: Spectrum Count tab change of summarization level



The data is summarized in a tablein which each row represents a PTM site or a peptide. A number of columns provide scores associated with the PTM site or peptide and the quantitative values are reported at the selected level of summarization. Spectral counts are

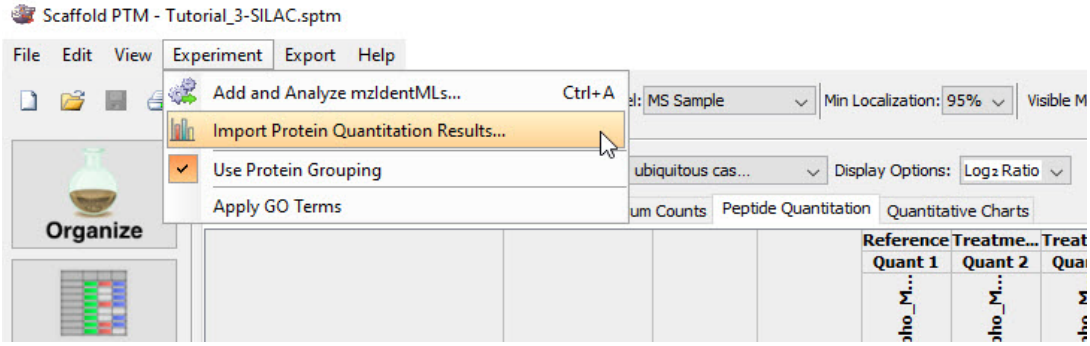
added across samples included in the summarized column, and medians are taken across all ratios in all of the samples.

Protein Level Normalization

Both the PTM Quantitation tab and the Peptide Quantitation tab support an experimental design that assists in separating the effects of changing protein levels from changing patterns or levels of modification.

In this type of experiment, MS samples are obtained under similar conditions from PTM-enriched and unenriched samples. The PTM-enriched samples are loaded into Scaffold Q+ or Q+S, and the results are exported as an SQML file. The unenriched samples are loaded into a separate Scaffold Q+ or Q+S experiment, and results are exported as a Protein Quantitation XML Report. The SQML is loaded into Scaffold PTM, then the Protein levels are imported through the **Import Protein Quantitation Results...** option in the Experiment menu.

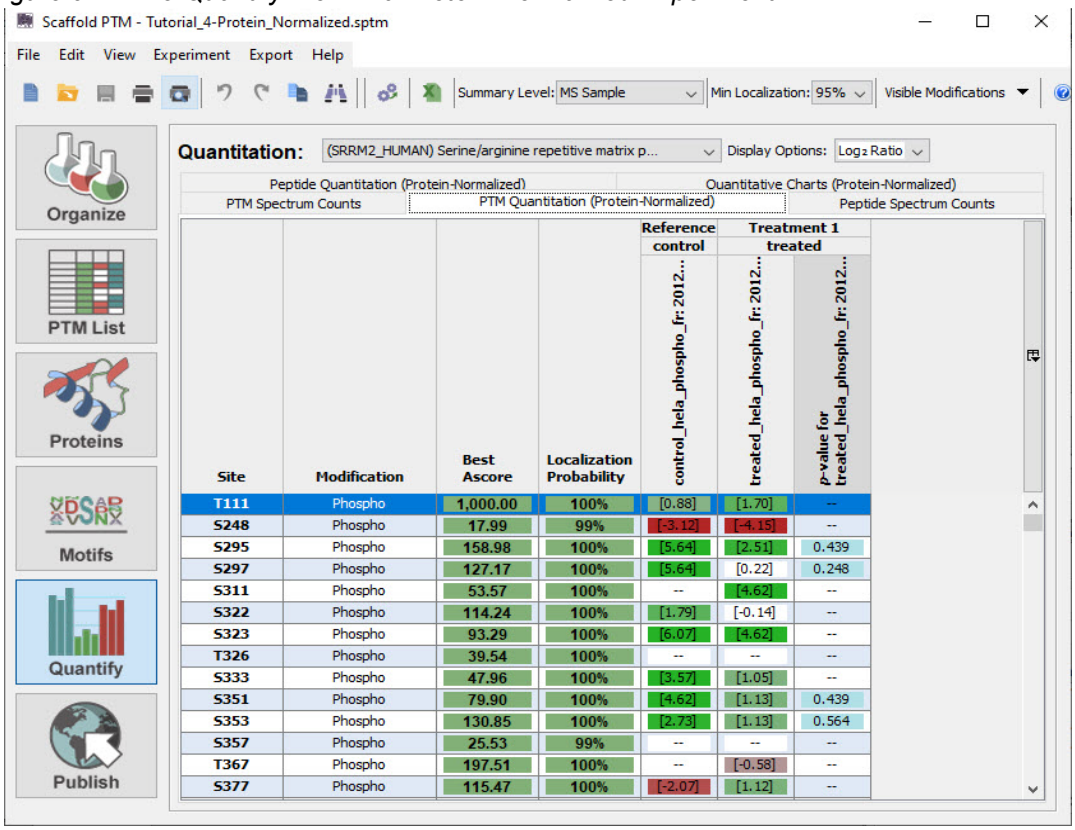
Figure 9-3: Importing Protein Quantitative Results



The program then prompts the user to align the protein quantitation samples with the enriched samples. The imported protein quantitative values are used to adjust the quantitative values in the PTM Quantitation tab or the Peptide Quantitation tab to better reflect differential modification by removing the effect of differential expression. For details, see “Creating a Protein-Normalized Quantitative PTM Experiment” on page 40.

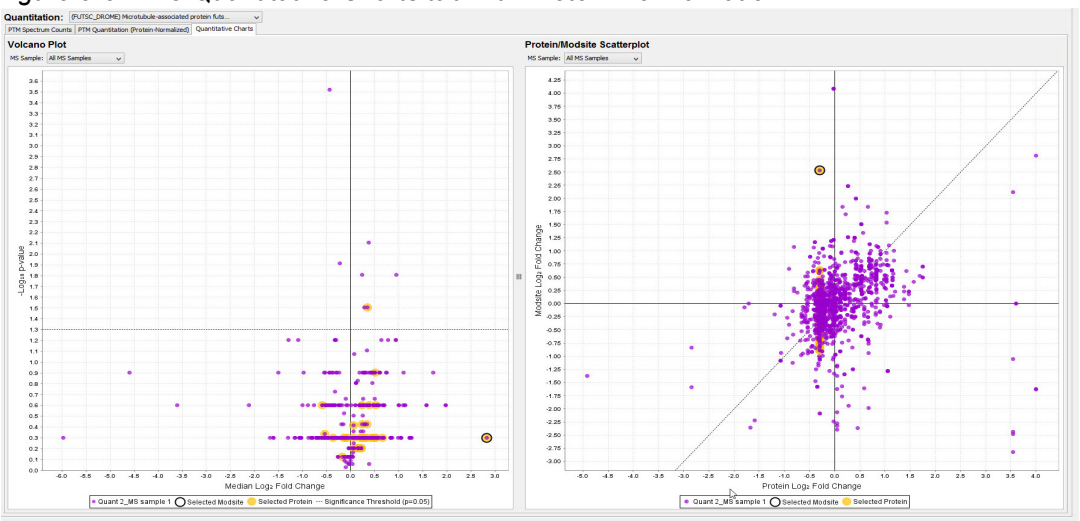
When Protein Quantitation information has been imported, the headers of the PTM Quantititation tab and the Peptide Quantiation tab add the notation: “(Protein Normalized), see Figure 9-4. The ratio values displayed in the tables and charts in these tabs are adjusted to account for differences in protein levels. In addition, the Protein/ModSite Scatterplot is displayed in the Quantitative Charts tab.

Figure 9-4: The Quantify View in a Protein-Normalized Experiment



In protein-normalized experiments, the Quantitative Charts tab also displays an additional graph. The Protein/Modsite Scatterplot appears. This graph plots the Log_2 Ratio of the Modsites vs. the Log_2 Ratio of the Protein Levels.

Figure 9-5: The Quantitative Charts tab with Protein Normalization

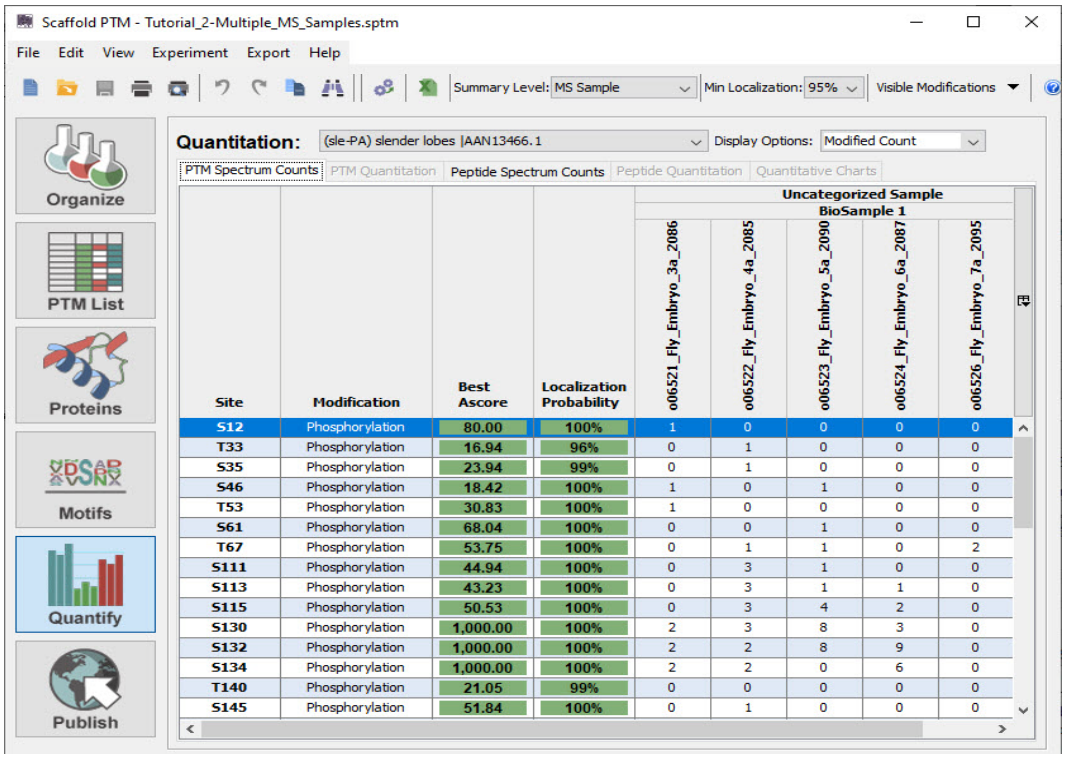


Tabs in the Quantify View

The tabs available in the Quantify View also depend on the type of data loaded into Scaffold PTM. There are three possible states. If data was loaded from:

- **MzIdentMLs** from Scaffold or a search engine - The quantitative information appearing in Scaffold PTM is only spectral counting or TIC and only the PTM Spectrum Count tab and the Peptide Spectrum Counts Tab are active, see Figure 9-7.

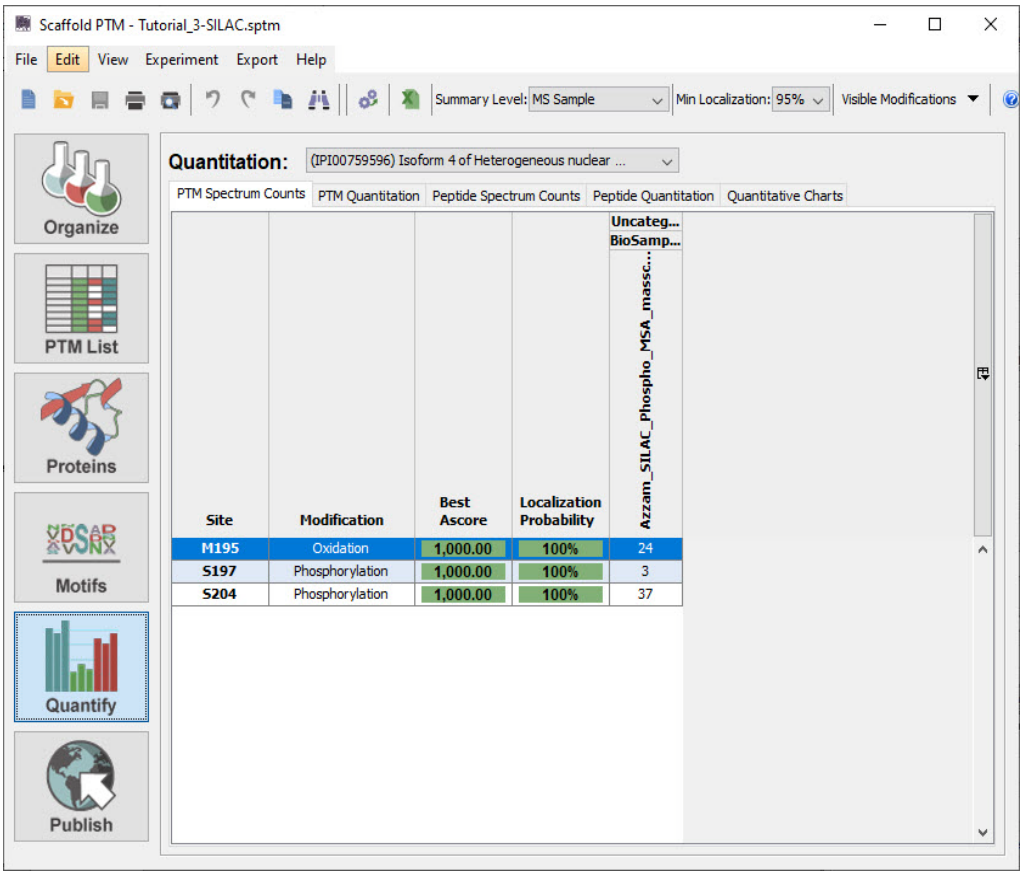
Figure 9-6: The Quantify View with Spectral Counting Data



- **SQML** files from Scaffold Q+ or Q+S - In this case Scaffold PTM will show all of the possible tabs, see Figure 9-7.

  Scaffold PTM offers statistical features to establish the significance of the fold change values assessed for each modification site. These values are displayed in the PTM Quantitation Tab, Peptide Quantitation Tab and the Quantitative Charts tab

Figure 9-7: Quantify View with quantitative values.



PTM Spectrum Counts tab

The PTM Spectrum Counts Tab offers semi-quantitative estimates of PTM abundances. It displays label-free, unweighted counts of spectra with assigned PTMs at a specific site for each MS Sample.The PTM Spectrum Counts tab gives an estimate of the relative occurrence of the PTM at the selected level of summary.

If the data loaded into Scaffold PTM is labeled or has precursor intensity data and is exported from Scaffold Q+, more reliable quantitative information is available through other tabs, but this view gives evidence for the values in the PTM Quantitation tab.

Five Display Types are available through the Display Options pull-down. By default, the Modified Count displays the number of spectra containing identified modifications of the specified type at the specified location. Another option, Modified Count/Total, shows the Modified Count, but also shows the total count of spectra spanning that location in the sequence. This provides an estimate of the relative degree of modification at a given site in various samples.

The remaining three display options show an estimate of the Total Ion Current (TIC). The reported TIC value of an MS/MS spectrum is the sum of the intensities of all ions in the spectrum. The Total TIC is the sum of the TIC values of all spectra containing the specified modification at the specified site. Average TIC is the mean of these values, and Top 3 TIC sums only the three highest TIC values of spectra containing the indicated modification. If there are fewer than three spectra with the modification, Top 3 TIC is the sum of TIC values of the spectra that do meet the criteria.

PTM Quantitation tab

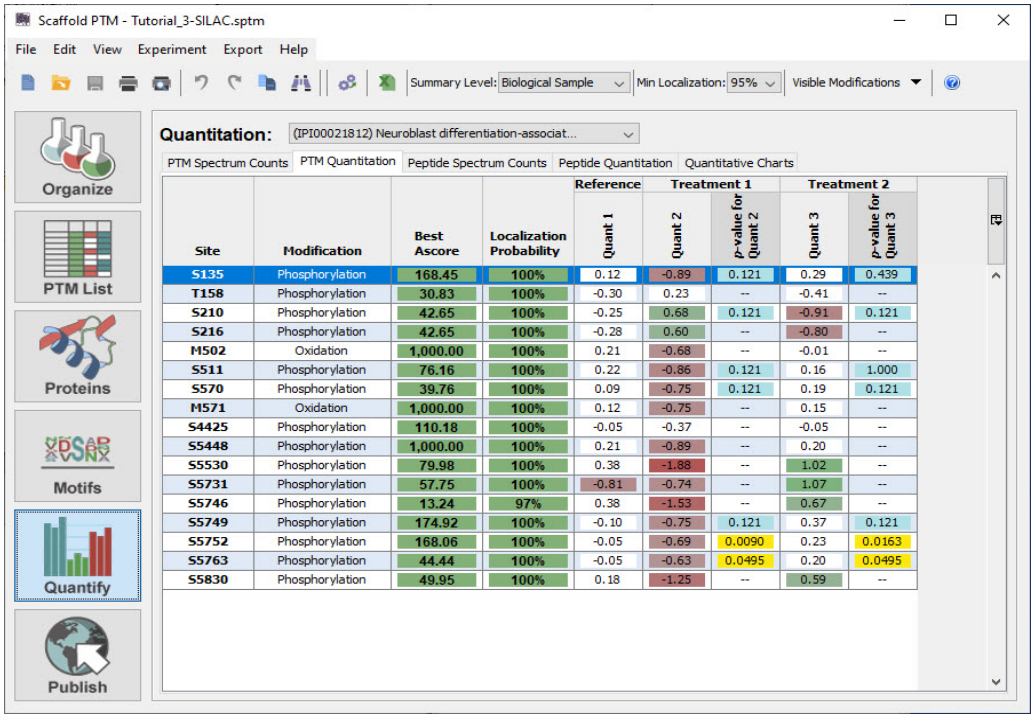
The PTM Quantitation Tab shows PTM abundance values imported from Scaffold Q+ or Scaffold Q+S analysis. The data is organized in a table containing identified PTM sites in the selected protein which meet the current filter conditions. The tab only appears if quantitative SQML data is loaded.

Scaffold Q+ and Scaffold Q+S calculate the reference values using quantitative options selected by the user before exporting SQML data. The PTM abundance is calculated by comparing the median *log2* of the fold change measurements for each PTM site in every sample to the reference value. The reference value is the average (median or mean) peptide abundance measurement in the reference sample. For more information on how to create SQML data files see "ScaffoldQuantML exports" on page 37.

Note: Scaffold PTM does not consider PTM sites that score below the selected Minimum Localization confidence level in this calculation.

Scaffold PTM also offers statistical measures to assess the significance of the fold change value computed for each modification site. These values are displayed under this tab and in the Volcano Plot.

Figure 9-8: Quantitation tab



When SQML data files from Scaffold Q+ or Scaffold Q+S are loaded, Scaffold PTM stores the quantitative ratios computed by Scaffold Q+ or Scaffold Q+S for every spectrum.

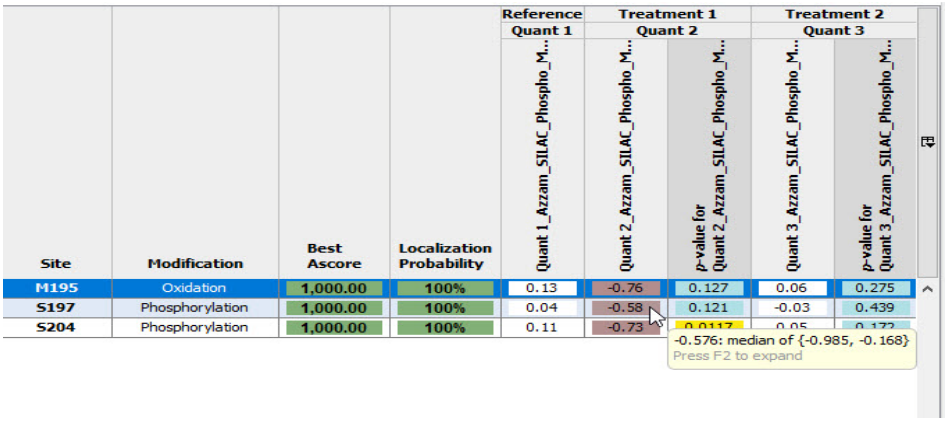
These values are then used to compute a fold change for each modification site in the experiment by taking the median fold change across all spectra in a given sample that were identified as a peptide containing the specific modification at that site.

These median fold change values are displayed in the table in this tab.

To see details of the calculation of modification site fold changes, the user can hover the mouse over a ratio value and see every underlying fold change associated with the specific site, see Figure 9-9.

When viewing data at the Summary Level of Biological Samples or Categories, the computation is analogous, but with a slight difference. The displayed values are the median fold changes for all spectra within that **Quantitative Sample** or Category. Note that spectral counts are summarized at the BioSample level, while quantities derived from Scaffold Q+ or Q+S are summarized by the Quantitative Channel or Sample .

Figure 9-9: Tooltip Showing Values that Contribute to Fold Change Calculation



Peptide Spectrum Counts Tab

The Peptide Spectrum Counts Tab is similar to the PTM Spectrum Counts Tab, but the table is organized by peptide, rather than by PTM site. Each row represents a modified peptide identified in the selected protein. Columns provide:

- Peptide Sequence - with modifications indicated as blue lowercase letters
- Sites - all of the modification sites in the peptide as a comma-delimited list. The numbers in the sites indicate the positions within the peptide of modified amino acids.
- Best Ascore - the highest Ascore among the modified spectra identifying the peptide.
- Localization Probability - the localization probability corresponding to the Best Ascore.
- Quantitative Values for each sample or category - the type of value is determined by the type selected in Display Options.

Peptide Quantitation tab

The Peptide Quantitation tab contains a table similar to the one found in the PTM Quantitation tab, but the rows represent peptides, rather than PTM sites. Columns include:

- Peptide Sequence - with modification sites indicated by color coded lower case letter.
- Modification Sites - a comma-separated list of modification sites identified in the peptide.
- Best Ascore - the highest Ascore among the modified spectra identifying the peptide.
- Localization Probability - the localization probability corresponding to the Best Ascore.
- Quantitative Values for each sample or category - the type of value is determined by the type selected in Display Options.

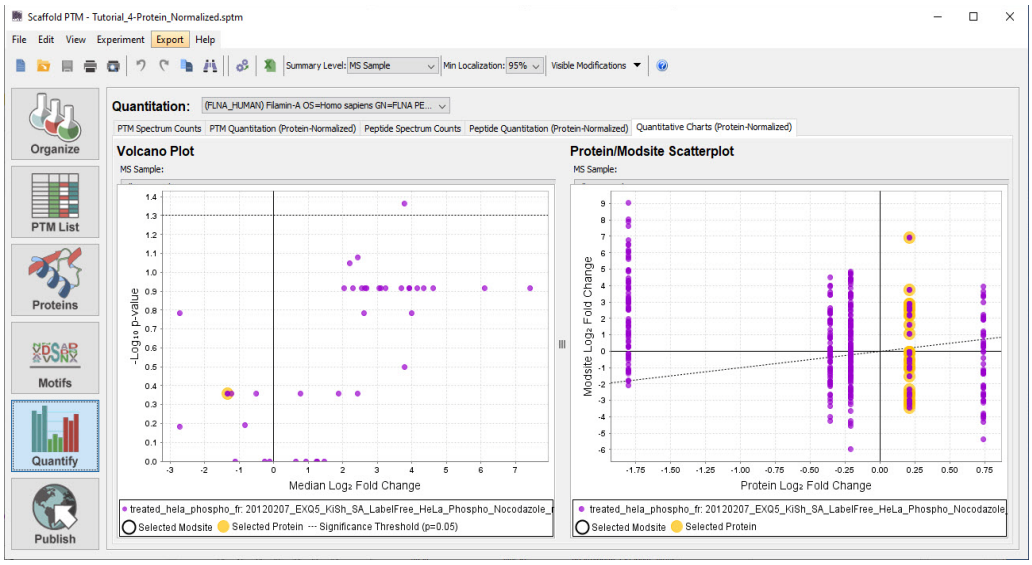
Quantitative Charts tab

This tab becomes available whenever the user loads SQML data. Depending on whether the loaded analysis includes the protein relative quantitation or not the tab will include one or two plots:

For any experiment containing quantitative data, the Quantitative Charts tab of the Quantify View contains a Volcano plot, showing the relationship between fold change and assessed statistical significance (computed as −log10(p)) for each modification site in the experiment. These values are computed for every sample at the current level of summarization, except for those in the Reference category, as described in Quantitative Statistics.

- Volcano plot
- Protein/Modsite Scatterplot

Figure 9-10: Quantitative Chat tab



Volcano plot

By default the volcano plot shows all samples simultaneously, but each sample can also be shown individually by selecting it from the drop-down menu above the plot. The plot also indicates the currently selected protein by surrounding the points for its sites with a yellow circle, and the selected site within that protein with a black outline. Selection can be changed by clicking points in the plot and will be reflected in the PTM Spectrum counts and PTM Quantitation tables, as well as the Proteins View. Selection indication can be disabled by right-clicking the plot and deselecting “Indicate selection”.

By default the volcano plot shows all samples simultaneously, but each sample can also be shown individually by selecting it from the drop-down menu above the plot.

Protein/Modsite Scatterplot

When protein quantitation data is present in a Scaffold PTM experiment, the Quantitative Charts Tab of the Quantify View will contain a scatterplot of PTM sites, showing the relationship between the site's quantitative ratio in each non-Reference-category sample (at the current level of summarization) and its protein-level quantitative ratio (both axes are in $\log_2$ space). By default the plot shows all samples simultaneously, but each sample can also be shown individually by selecting it from the drop-down menu above the plot. The diagonal dashed line indicates no change in PTM activity, while points far from this line have major changes in PTM quantity after controlling for protein expression.

The plot indicates the currently selected protein by surrounding the points for its sites with a yellow circle, and the selected site within that protein with a black outline. Selection can be changed by clicking points in the plot and will be reflected in the PTM Spectrum counts and PTM Quantitation tables, as well as the Proteins View. Selection indication can be disabled by right-clicking the plot and deselecting "Indicate selection".

Chapter 10
The Publish View

The Scaffold PTM Publish View displays information about the data in the experiment and the analysis performed on it. This information is required for replication of the experiment or publication in proteomics journals.

The Publish View contains two tabs:

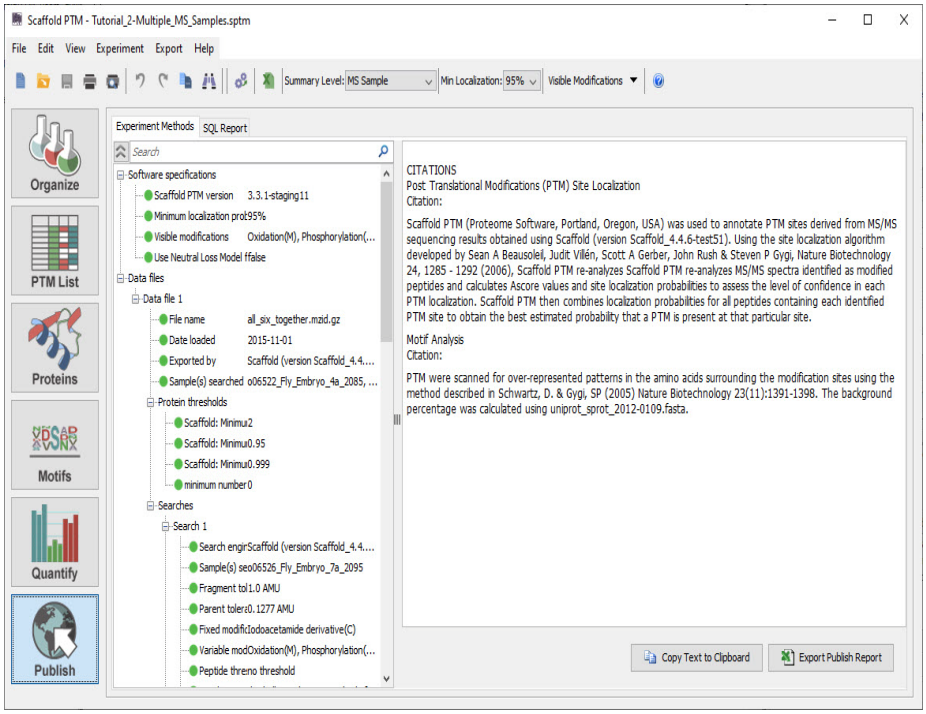
Experiment Methods tab

The **Experiment Methods** tab contains two panes:

- **Parameters pane** - located on the left side of the tab, this displays the parameters characterizing the current experiment. The information is organized in a tree-structured table that can be expanded or collapsed for ease of view. The pane also includes a text search tool for easily locating specific parameters of interest and a button that toggles the tree to an expanded or collapsed state.
- **Citations pane** - located on the right side of the tab, this pane provides the full citations for the key algorithms used in Scaffold PTM.

Below the text window there are buttons that allow the user to copy the information or to Export the Publish report. These reports may be useful as supplemental data to support publication in a Proteomics journal.

Figure 10-1: Experiment Methods Tab



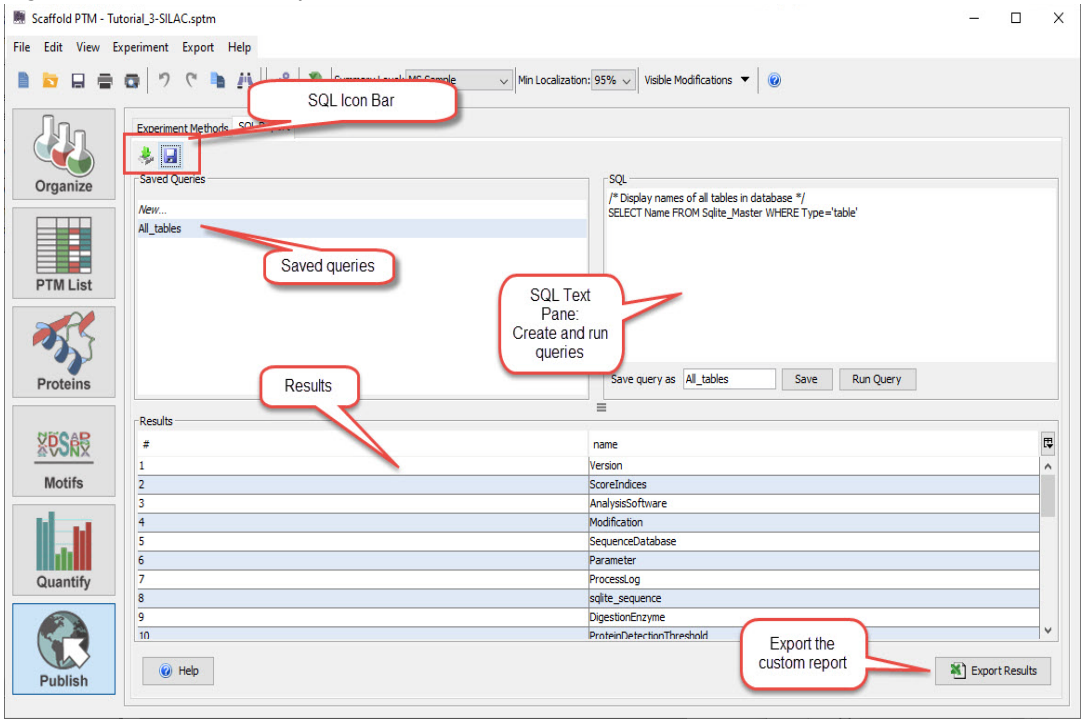
SQL Export tab

The experiment files created by Scaffold PTM, SPTM files, are SQLite databases.

The SQL export tab is an SQLite graphical interface where a Scaffold PTM experiment file can be searched as a database using SQLite commands. In this way, the user can create custom tables exportable to Excel.

A description of the schema of a SPTM file is shown in Structure of Scaffold PTM files (SPTM).

Figure 10-2: The SQL Export tab



The SQL Export tab contains four different panes:

- “The SQL pane” on page 119
- “The Saved Queries pane” on page 120
- “The Results pane” on page 120
- “The Icon bar” on page 120

The SQL pane

Through the SQL pane it is possible to directly explore the information stored in a Scaffold PTM file using SQLite queries.

- **The SQL text pane** - allows the user to enter, copy and paste SQL queries.

- **The SQL Icon bar** - which contains the **Run Query** button, the **Save query as:** text box and a **Save** button to save queries.

The results of the queries are shown in The Results pane. Saved queries are listed in the The Saved Queries pane

Examples:

List of tables available in *.SFDB files.

SELECT name FROM SQLite_master WHERE type='table' ORDER BY name;

The Saved Queries pane

When the user names and saves a query, it appears in this pane and it is conveniently available to be launched again whenever needed.

The Results pane

When the **Run Query** button is pressed, if there are no errors, a table containing the query results appears in this pane. Clicking the **Export Results** button or selecting the right click menu option **Export > Export to Excel...,** saves the table in a CSV text format file that can be easily opened in Excel.

The Icon bar

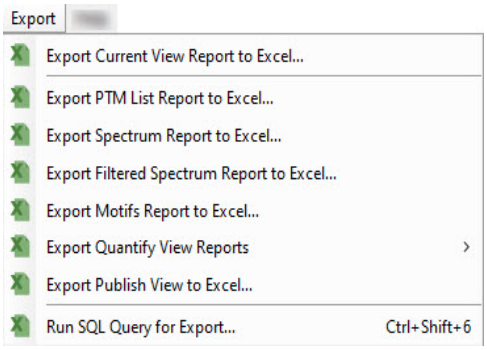
The icon bar contains an icon to save new queries to a file that can later be retrieved and an icon to import previously saved queries.

Chapter 11
Reports

A variety of reports are available in Scaffold PTM. All the reports are available through the Export option on the Scaffold PTM main menu, see Figure 11-1. Each report is saved as a CSV file which can be opened and edited in Excel or processed by other software.

The user cannot change the report format, but can select a different location in which to save the report. When the user saves an Excel-compatible report, a default name in the format <Report Name><Scaffold File name> is provided for the report, but the name may be changed.

Figure 11-1: Exports available in Scaffold PTM



Exports compatible with Excel

Scaffold PTM provides a number of reports created in the CSV (Coma Separated Values) text file format. Each report contains different types of information related to the analysis performed in the current Scaffold PTM experiment:

- Current View report
- Spectrum report
- Spectrum report
- Filtered Spectrum Report
- Motifs report
- Quantify View Reports
- Publish View report
- Run SQL Query for Export

Opening Scaffold PTM reports in Excel:

The exported reports can be viewed in Microsoft Excel for further analysis of the data they contain. A CSV file opens automatically in Excel if the application is installed on the computer system used.

To create an export that includes the GO annotations use Current View report from the PTM List View with GO terms visible.

Current View report

The Current View report contains the information that is displayed in the current view. This report is applicable for all Views.

PTM List report

The PTM List report mimics the PTM List View. Each row in the report represents a protein in the PTM List table. The modifications displayed in the columns depend on the filter settings specified in **Min Localization** and **Visible Modifications**. For example, the list of columns for Tutorial 3 is shown in Figure 11-2.

Figure 11-2: PTM List report columns

Spectrum report

The Spectrum report details all of the spectra in the experiment. Each row represents a spectrum matching a peptide.

Figure 11-3: Spectrum report columns

Notes about the Columns:

- The first 2 columns of the table provide information identifying the protein.
- Next, the peptide sequence is shown followed by the list of variable modifications with their Ascore and localization probabilities and the best scores for the spectrum matching and if data files were imported from Scaffold the peptide probabilities are also included.
- Number of enzymatic termini (NTT). When the digestion enzyme is trypsin, this tells if the peptide is tryptic (2) semi-tryptic (1) or non-tryptic (0).
- Several columns provide information about the mass and the delta mass.
- The start and stop positions of the peptide in the protein sequence.
- Fixed modifications.
- The spectrum name and the MS Sample in which it is found.

Filtered Spectrum Report

This report is similar to the Spectrum Report, but respects the thresholds and filters that are active in the experiment at the time the export is performed.

Publish View report

The Publish View report lists the data analysis information required for publication in a Proteomics journals. This report is a copy of the information reported in "The Publish View" on page 117.

Motifs report

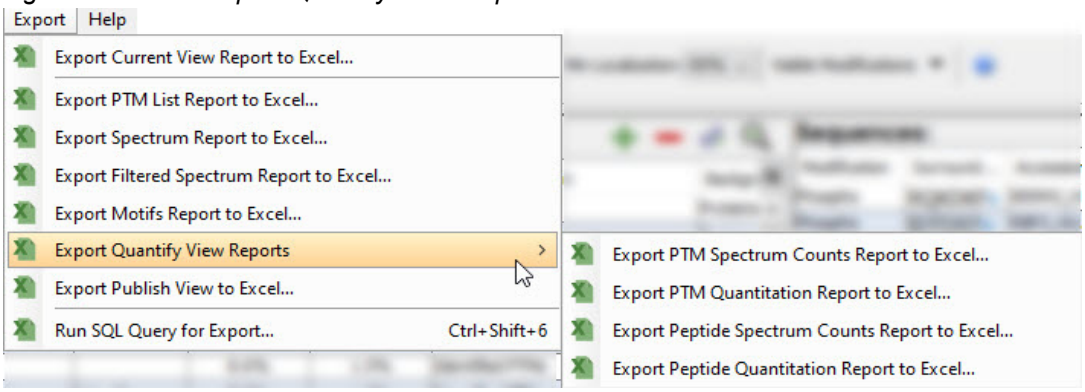
This option exports most of the information in the table in the **Motifs:** pane

Figure 11-4: Motifs report columns

Quantify View Reports

The Export Quantify Reports option opens a submenu offering a number of quantitative reports.

Figure 11-5: The Export Quantify View Reports submenu



PTM Spectrum Counts report

This option exports the data displayed in the PTM Spectrum Counts tab of the Quantify View for all proteins in the experiment. A dialog allows the user to select which of the PTM Spectrum Counts tab (Modified Count, Modified Count/Total, Top 3 TIC, Total TIC or Average TIC) should be reported.

PTM Quantitation Report

This report contains the same information as in the PTM Quantitation tab for all proteins in the experiment. The user may select either Log_2 Ratio or Ratio as the Display Option to be reported. If protein normalization information has been imported, the report shows Protein Normalized ratios.

Peptide Spectrum Counts Report

The Peptide Spectrum Counts Report exports the same information displayed in the Peptide Spectrum Counts Tab of the Quantify View for all proteins. The user may select from Modified Count, Modified Count/Total, Top 3 TIC, Total TIC or Average TIC as the PTM Spectrum Counts tab available for this export.

Peptide Quantitation Report

The Peptide Quantitation Report exports the information displayed in the Peptide Quantitation tab Tab of the Quantify View for all proteins in the experiment. The PTM Spectrum Counts tab available for this report are Log_2 Ratio or Ratio.If protein normalization information has been imported, the report shows Protein Normalized ratios.

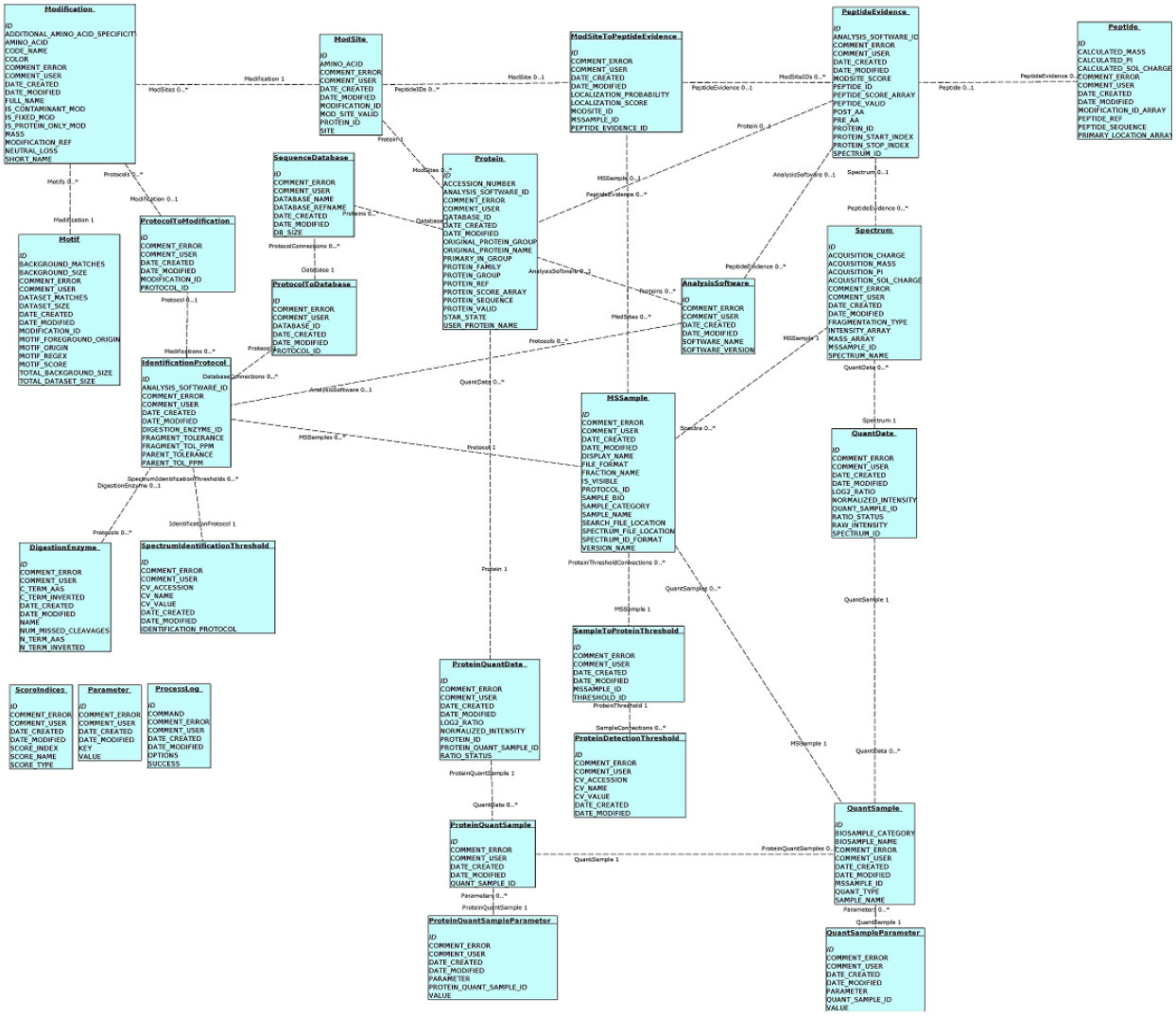
Run SQL Query for Export

Opens the SQL Exports Tab of the Publish View (see “SQL Export tab” on page 119).

Appendix A.Structure of Scaffold PTM files (SPTM)

Scaffold PTM experiment files, SPTM, are SQLite files. Figure 1 shows the schema of an SPTM file. SQlite queries of this database can be submitted through the SQL Export tab available in the Publish View or also reachable through the menu command **Export > Run SQL Query for Export...**.

Figure 1: Scaffold PTM database schema



Appendix B.PTM dynamics - quantitative calculations

Scaffold PTM contains features to enable quantitative analysis of PTM activity by simultaneously considering protein- and site-level changes within an experimental condition.

Scaffold PTM can import quantitative ratios computed by Scaffold Q+ and Scaffold Q+S and compute fold change ratios for individual PTM sites (see PTM Quantitation tab and Quantitative Analysis). However, these ratios reflect two distinct processes: up-/down-regulation of the whole protein (which affects quantitative measurements for all of the protein's peptides), and up-/down-regulation of a specific PTM site (which only affects measurements of the peptide(s) containing that site). This means that distinguishing which changes in PTM activity are present in an experiment requires a way of deconvolving these two effects. The simplest way of doing so is to directly measure any whole-protein quantitative changes, generally by analyzing a second preparation of the same biological samples without PTM enrichment. The difference between the measured PTM site ratios and whole-protein ratios then gives a measurement of the change in PTM activity.

Details of Quantitative Calculations

Consider that under some experimental condition, a protein P has a quantitative fold change $FC_P = p/q$, and that for some PTM site S in the protein we measure a fold change $FC_S = r/s$. Intuitively, this says that we observed *p* copies of the protein and *r* copies of the (modified) peptide under the experimental condition, while we measured *q* and *s* copies of the protein and peptide (respectively) in the reference sample. Thus, for every copy of the protein, we observed *r/p* copies of the peptide in the experimental condition, and *s/q* in the reference sample. We can then compute the change in PTM activity by the ratio:

$$\frac{r/p}{s/q} = \frac{r \cdot q}{p \cdot s} = \frac{r}{s} \cdot \frac{q}{p} = \frac{FC_S}{FC_P}$$

Because we do our calculations in log_2 space, the ratio of ratios may be computed as a difference:

$$\log 2\left(\frac{FC_S}{FC_P}\right) = \log 2(FC_S) - \log 2(FC_P)$$

When displaying quantitative ratios at higher levels of summarization, this normalization is applied to each measurement at the MS Sample level, and then the median of all the protein-normalized ratios in the Biological Sample or Category is computed for display. These pre-protein normalized values are also used when computing Quantitative Statistics.

Index

Q

R

S

T

U

V

W

Z