How to view Results with Scaffold 3.0

Source: Proteomics Shared Resource, Oregon Health & Science University
Adapted by Protein and Metabolite Analysis Facility (PMAF),UT-Austin
MBB 1.420, 471-2895

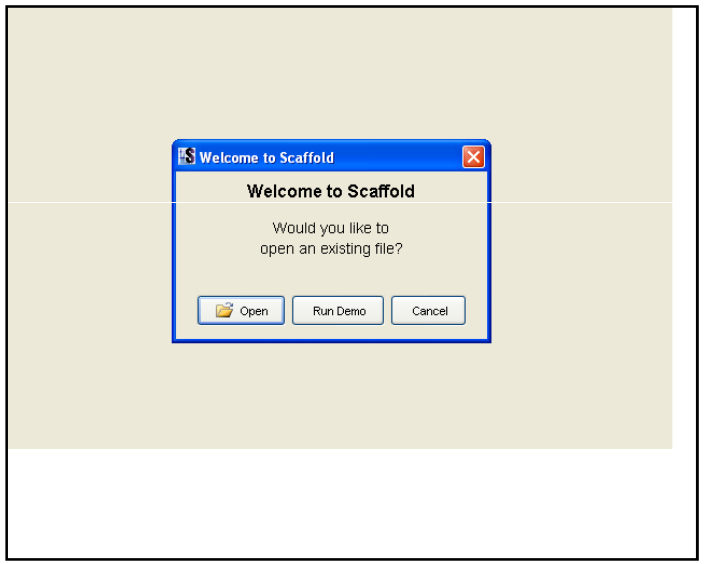
An overview

- This document is intended to walk you through Scaffold version 3.0.
- This is an introductory guide that goes over the basics needed to view your data.
- This guide will skim over several of the more in-depth features of the software.
- If you are interested in learning more about Scaffold you can view their official users guide here: http://www.proteomesoftware.com/pdf_files/Scaffold3_Users_Guide.pdf or contact PMAF and a member of the lab can sit down with you and go over the software in more detail.

Starting out

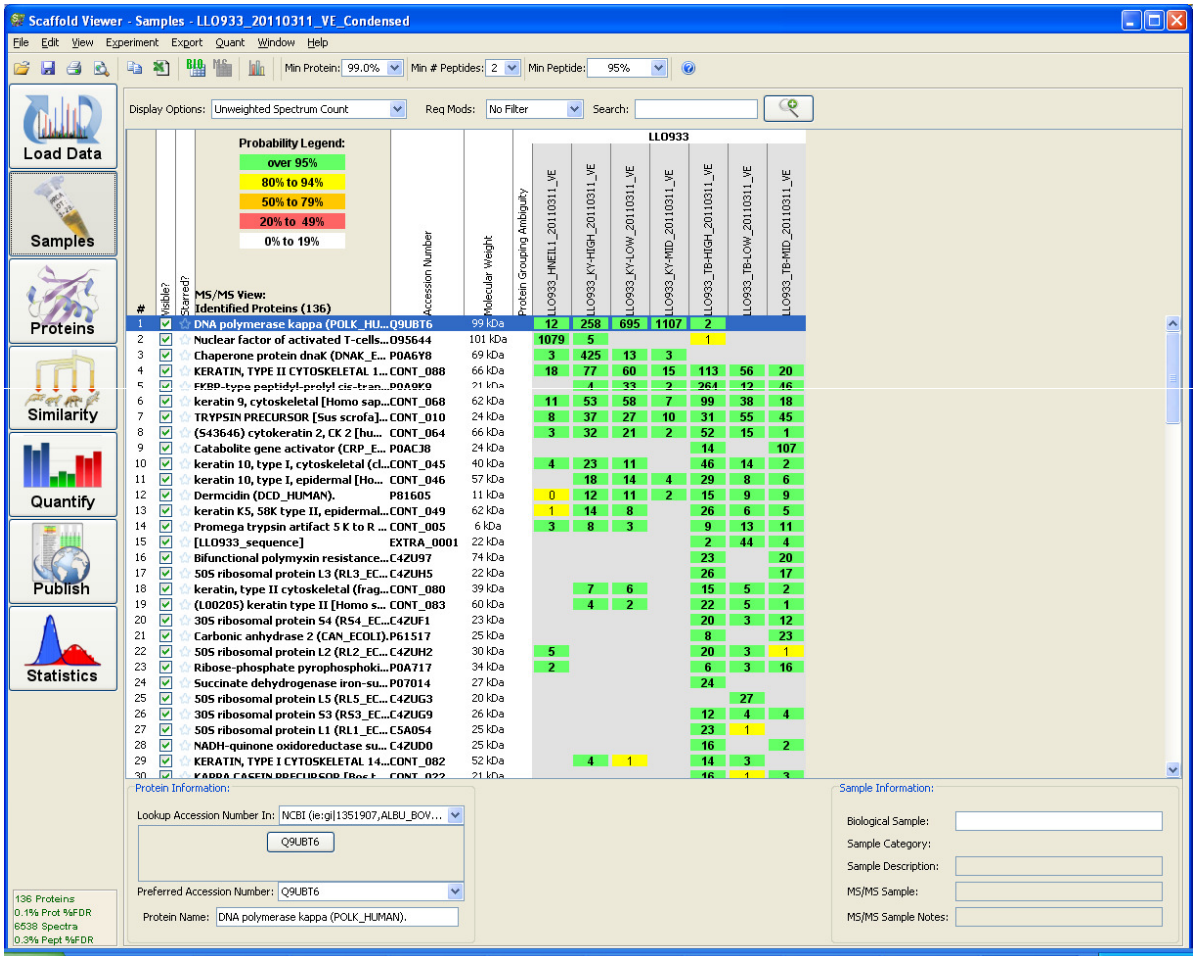
- Download Scaffold from http://www.proteomesoftware.com/Proteome_software_prod_Scaffold3_download-main.html
- Follow installation instructions on website, and install normally.
- When the installation is finished double-click on the Scaffold 3.0 icon to begin.
- When prompted to enter a Key select “Free Viewer” to use Scaffold for free to view your data

- Select Open and select the .sfd file containing the data of interest



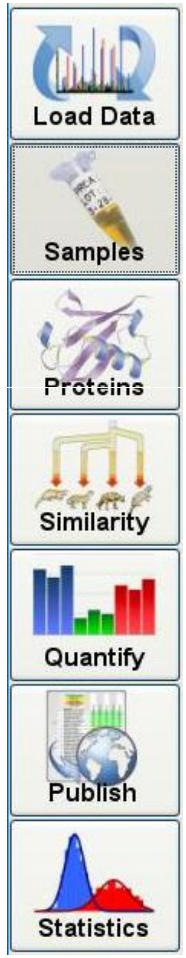
- The file will load in the viewer (this may take a minute)
- The opened file should be similar to what is on the next page

Scaffold Main Screen

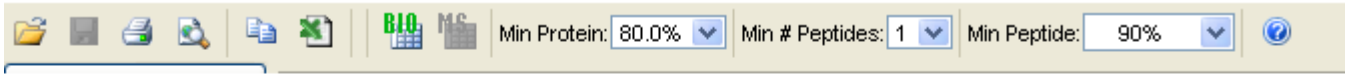


Left Toolbar

- The “Load Data” Tab isn’t used in the free viewer
- “Samples” displays a spread-sheet like format allowing you to sort your data
- “Proteins” shows the MS/MS spectra and % coverage information from a chosen protein
- The “Similarity” tab allows you to sort through proteins with shared peptides.
- “Quantify” gives you access to some basic tools for assessing differences in spectral counts.
- “Publish” creates a paragraph suitable for a methods Section from the settings on the “Samples” page
- The “Statistics” page shows statistical data created from the search algorithm(s) that processed the dataset

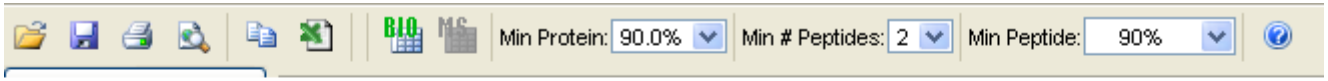


Top Toolbar



- On the far Left of the Top toolbar is the Open, Save, Print, and Print Preview commands respectively (note you can only have one Scaffold file displayed at a time)
- Next are tools for exporting the data to an excel spreadsheet
- “Copy data in current view” copies the displayed data so that you can paste it into an existing excel file and “Export to Excel Spreadsheet” exports all the data and creates a new file.
- The next two tabs switch between viewing the sample’s name and the name of the mass spectrometer raw data file.
- If there are multiple MS/MS samples in the same Biological switching to ‘Biological sample view’ will combine all the results into a single column and ‘MS sample view’ will separate out the samples so that you can see what proteins were identified in each individual MS/MS file.

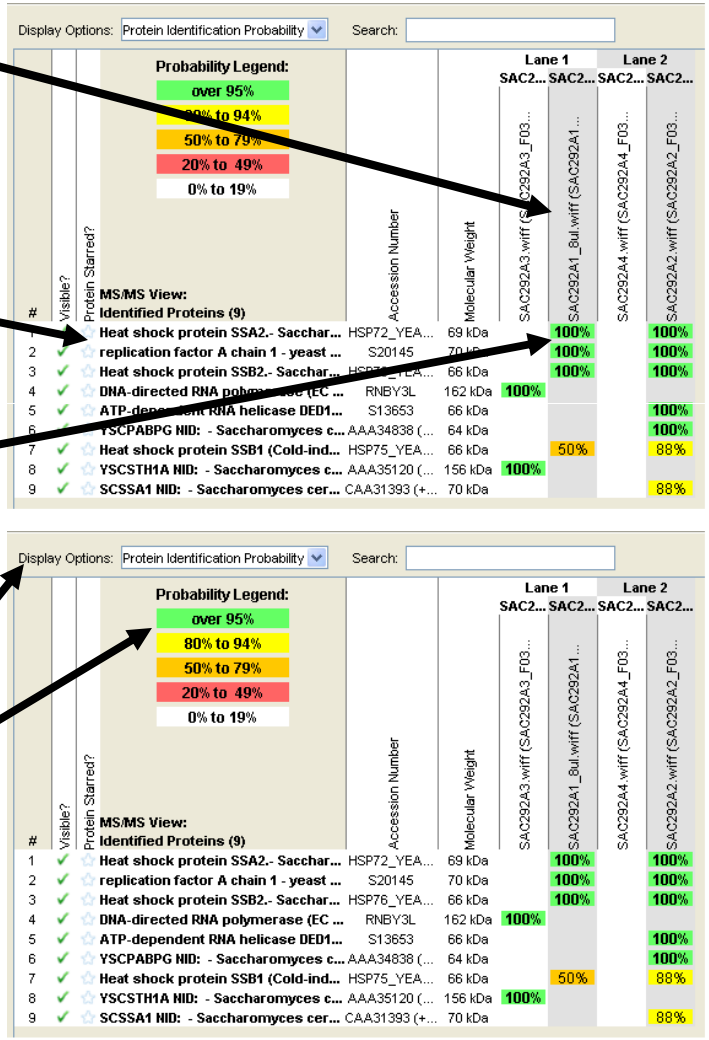
Top Toolbar 2



- "Min Protein," "Min # of Peptides," and "Min Peptide" determine what Proteins are displayed on the spreadsheet.
  - Note that Proteins displayed must meet all the criteria listed.
  - So with the criteria above a protein with 2 peptides, each peptide with >90% peptide prob. and 90% protein prob. would be displayed; but a protein with 1 peptide with peptide and protein probabilities of 95% each would not be in the list
- Also note that Protein probability is derived in part from peptide probability, so setting the protein probability much lower than the peptide probability likely won't display any more results
- For more details on how peptide and protein probabilities relate you can view the statistics tab
- The blue "?" access the help menu for scaffold

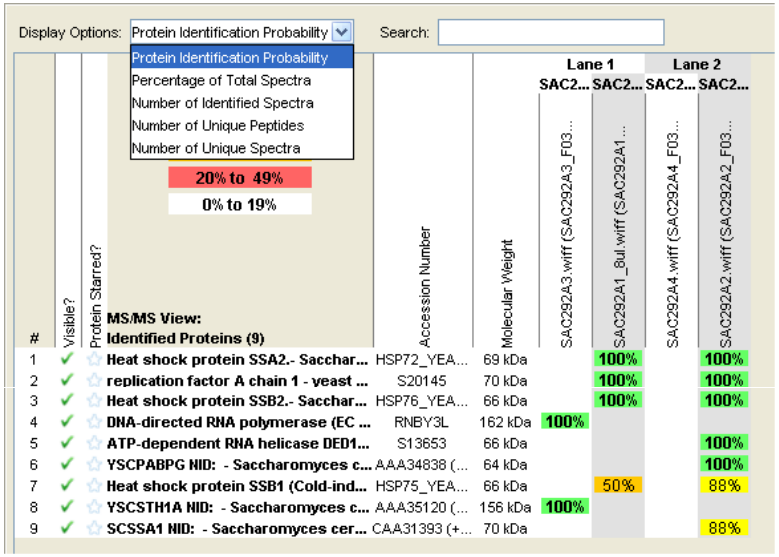
Samples Screen

- Displays a list of samples on the horizontal axis and proteins on the vertical axis
- Where these columns meet there is a value for that protein in the particular protein
- What value is displayed is determined by the display options tab
- Color corresponds to the Probability Legend referring to Protein Prob.



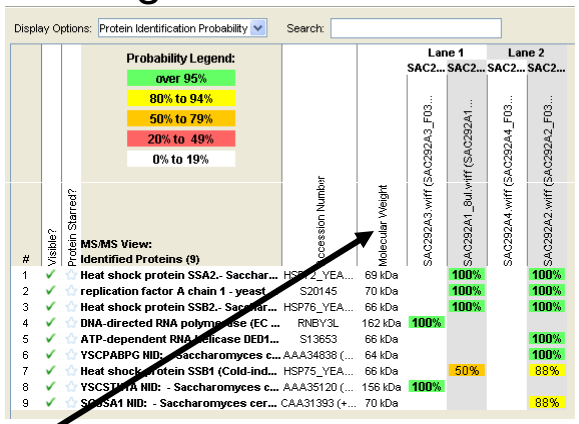
Display Options

- The Display Options drop-down menu determines what shows up where the two columns intersect
- “Protein Identification Probability” lists the Probability being present
- “% of total spectra” lists what % of the MS/MS spectra were assigned to that protein
- “# of Identified Spectra” gives the total # of spectra assigned to the protein
- “# of Unique Peptides” lists the # of Unique peptides in the identified protein (note that missed cleavages/degradation products are considered a different peptide)
- Finally “# of unique Spectra” is similar to the “# of Identified Spectra” but counts spectra of different charge states, but matching the same amino acid sequence, only once.

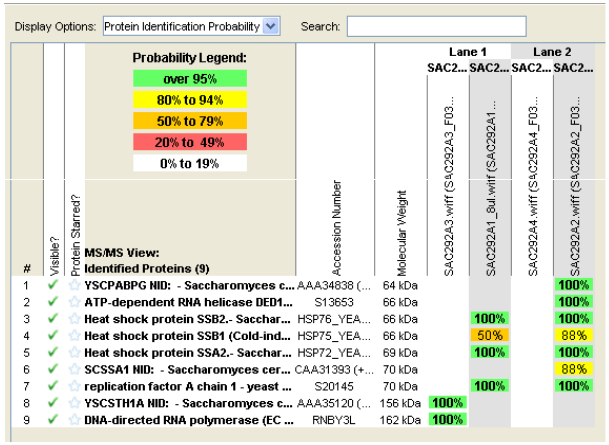


Sorting data

-Clicking on any of the horizontal axis columns sorts the data
-Clicking once sorts the data, clicking twice sorts it in the opposite order, and clicking a 3rd time returns the data to its original look

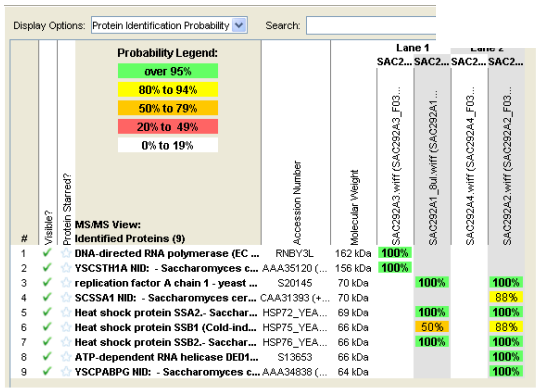


once



Click here

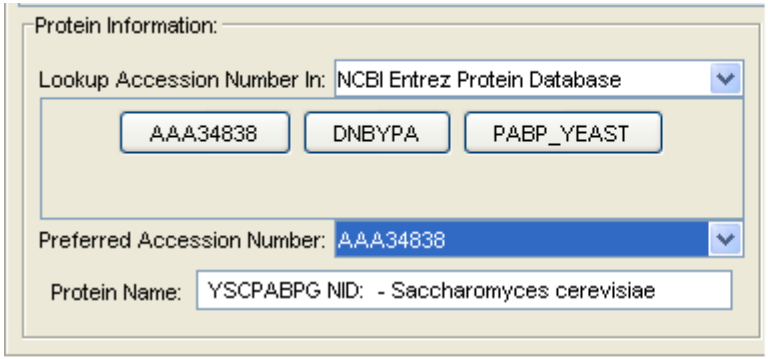
3 times



twice

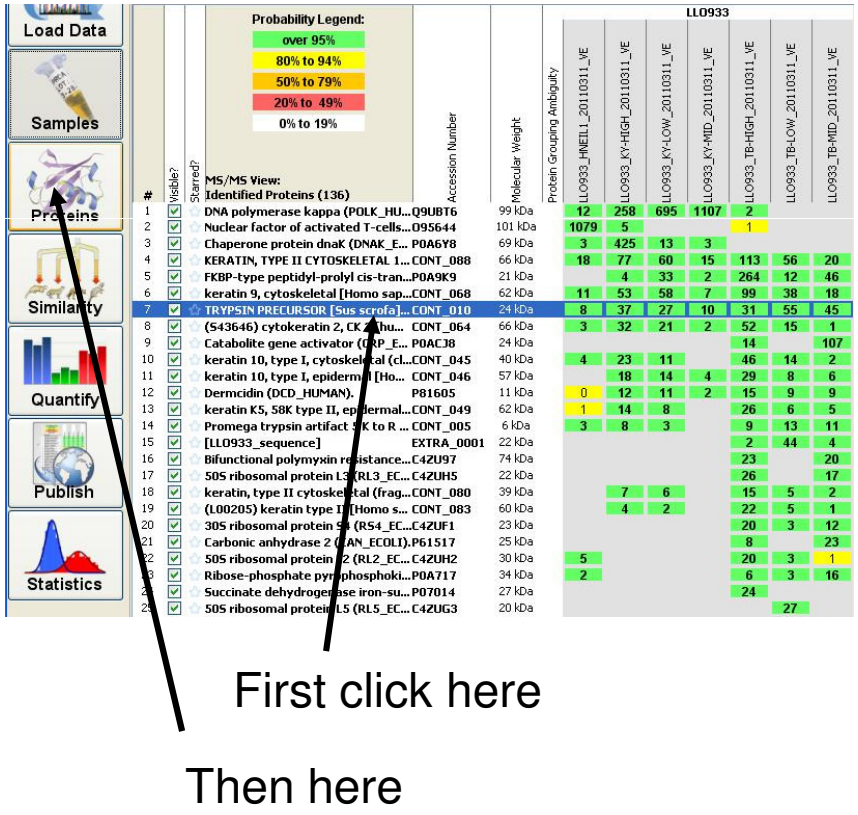
Other Sample Screen info

- Hovering your mouse over a value in the table will display more details
- When hovered over the Protein name this displays all proteins with which the identified peptides are a strong match. If more than one protein is listed here then you do not have enough sequence information to determine the protein your peptides belong to, but instead have one or more of the proteins listed in your sample.
  - This list can also be viewed under the “Proteins” Tab on the Left Scroll Bar
- At the bottom of the page is a protein information screen. This interface allows you to look up your protein on-line at various sites.
  - This will allow you to find more information on your protein, but what these screens reveal is beyond the scope of this guide to cover

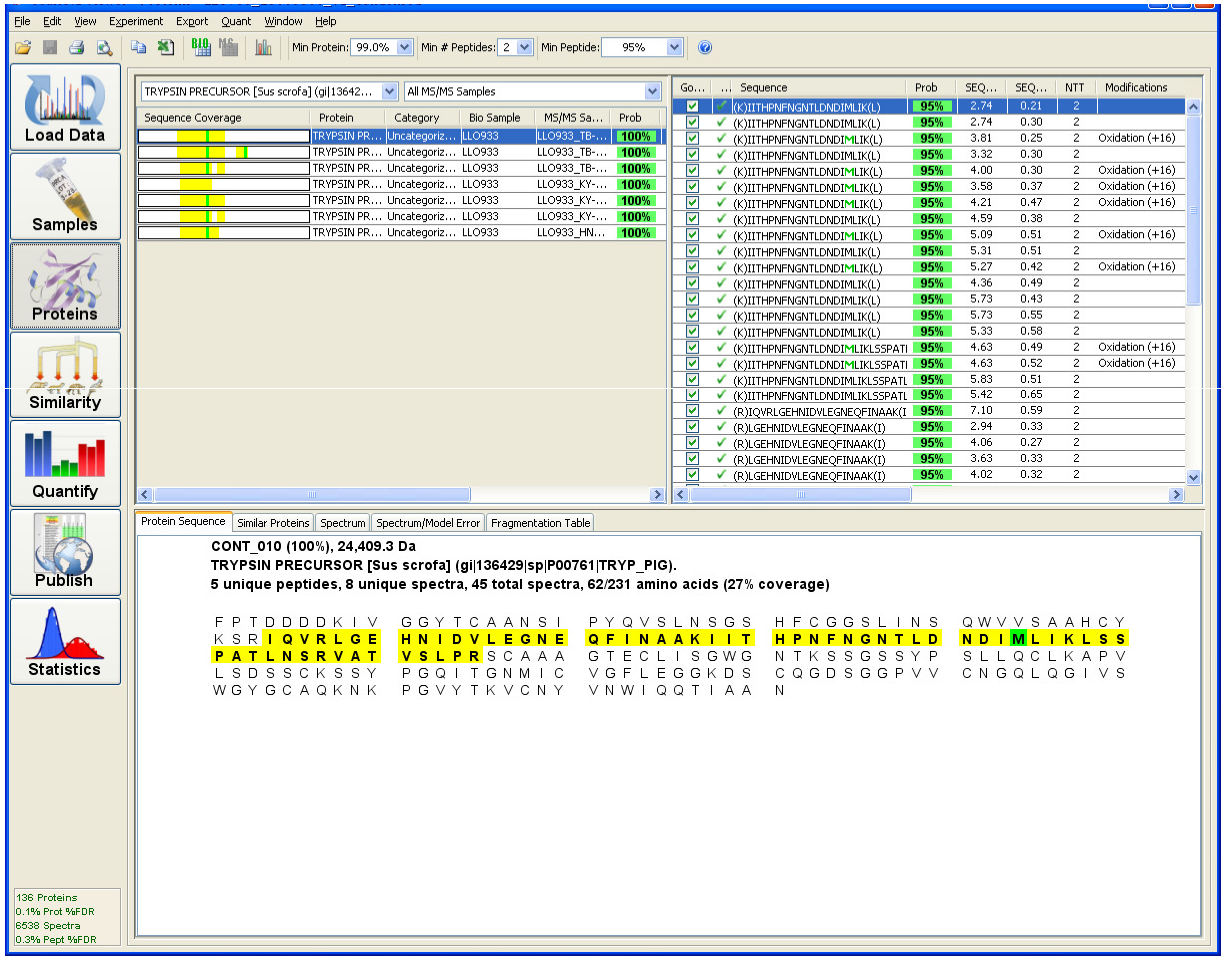


Protein Screen

- To view data in the “Proteins” Tab on the left tool bar first select a protein in the “Samples” screen by clicking on it.
- Then select the “Proteins” Tab on the left toolbar

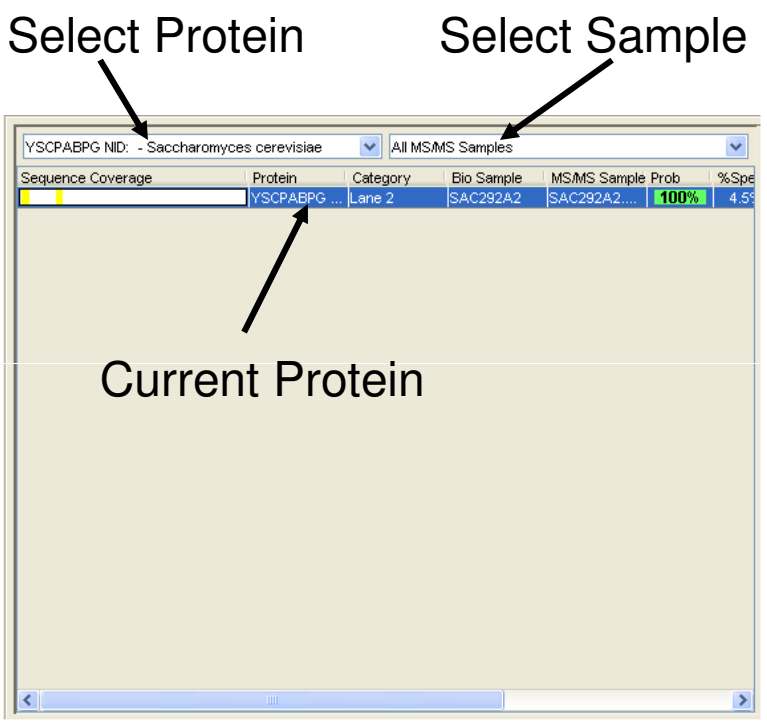


Main Protein Screen



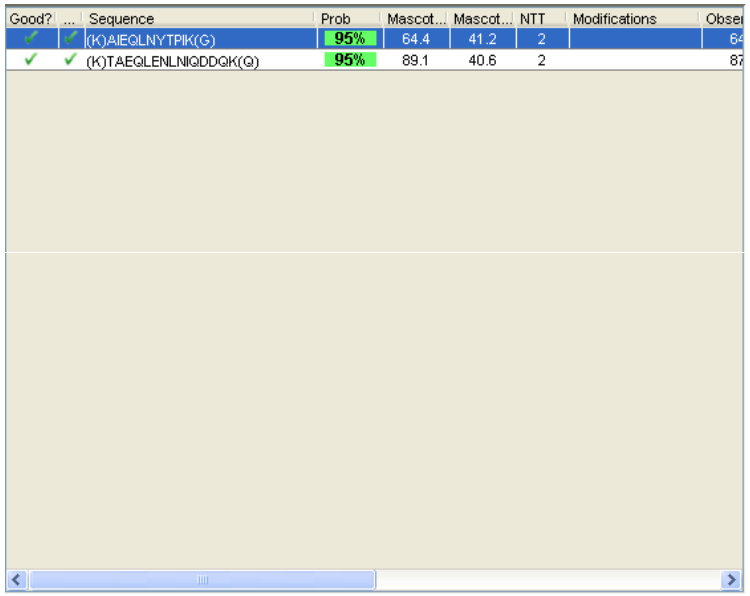
Upper Left Window

- The Upper left window contains much of the same information as the “Samples” Tab
- The Chosen Protein is listed
- Other Proteins can be chosen from various samples using the drop-down menus



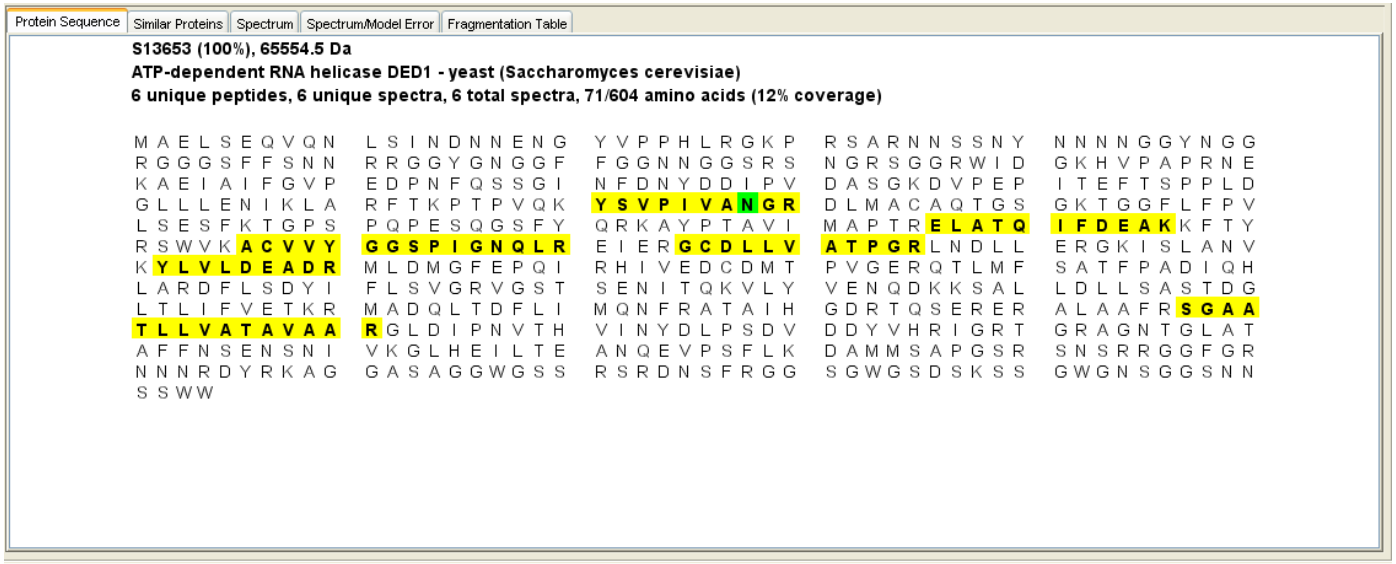
Upper Right Window

- The upper right window displays the peptides which have been assigned to the protein in the upper left window
- Values from the database search are included as well (i.e. mascot identity score)
- Modified amino acids are shown as a green letter in the sequence



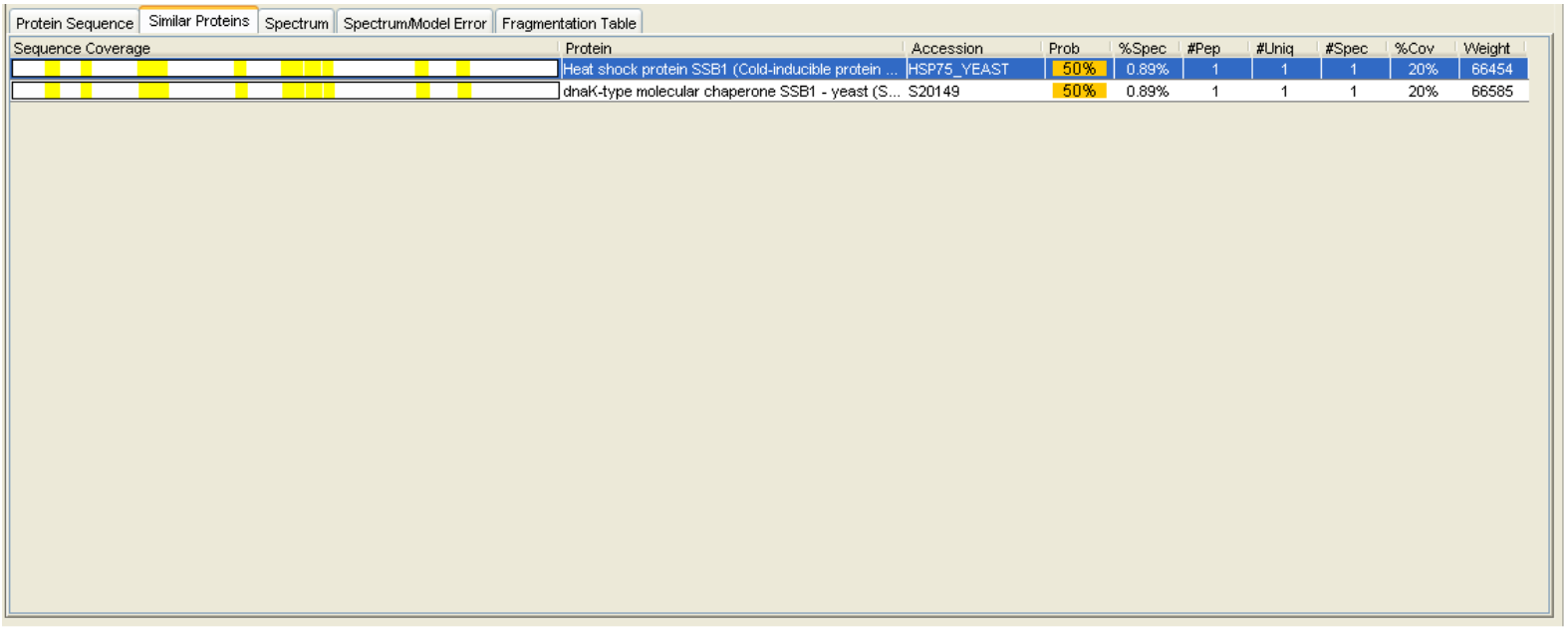
Lower window

- The lower window has 6 tabs to display information about the current protein
- The Protein Sequence tab shows the location of identified peptides on the protein
- Amino-acids matched to a MS/MS spectrum are in yellow. Amino-acids marked in green have a post-translational modification (i.e. phosphoylation)
- Hovering the mouse pointer over a yellow amino acid sequence will display a list of all the spectra matching that part of the sequence



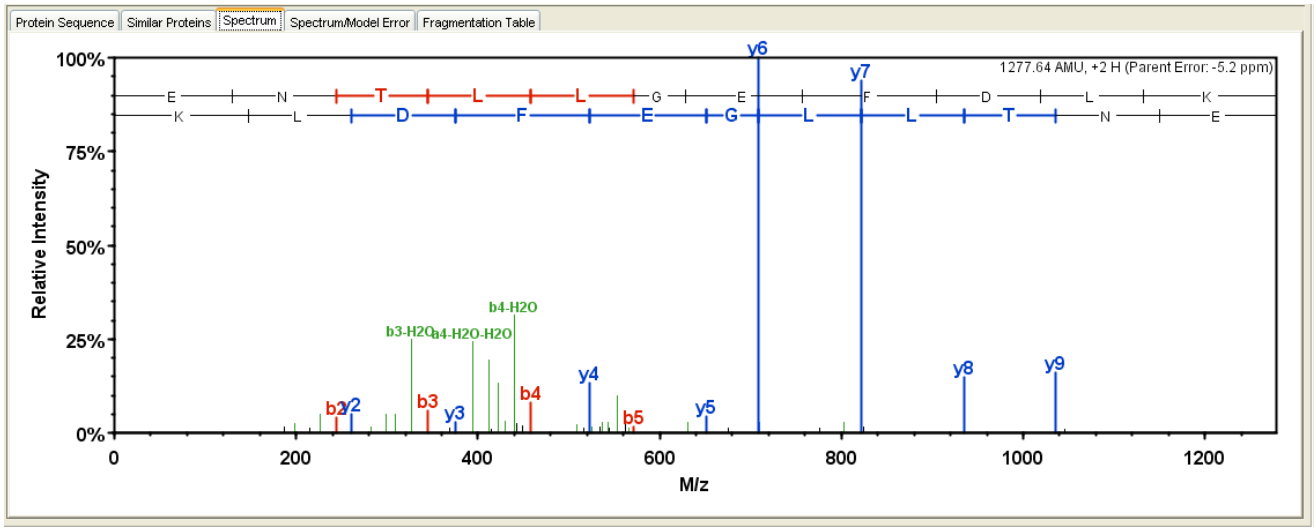
Lower Window 2

- The “similar proteins” tab lists all the protein which share the sequences identified (yellow/green) in the “protein sequence” tab
- If there is more than one protein listed here than there isn’t enough identified sequence information to distinguish between the proteins listed.
- This is common with genes that are heavily processed after transcription (i.e. exons and/or post translational modifications)



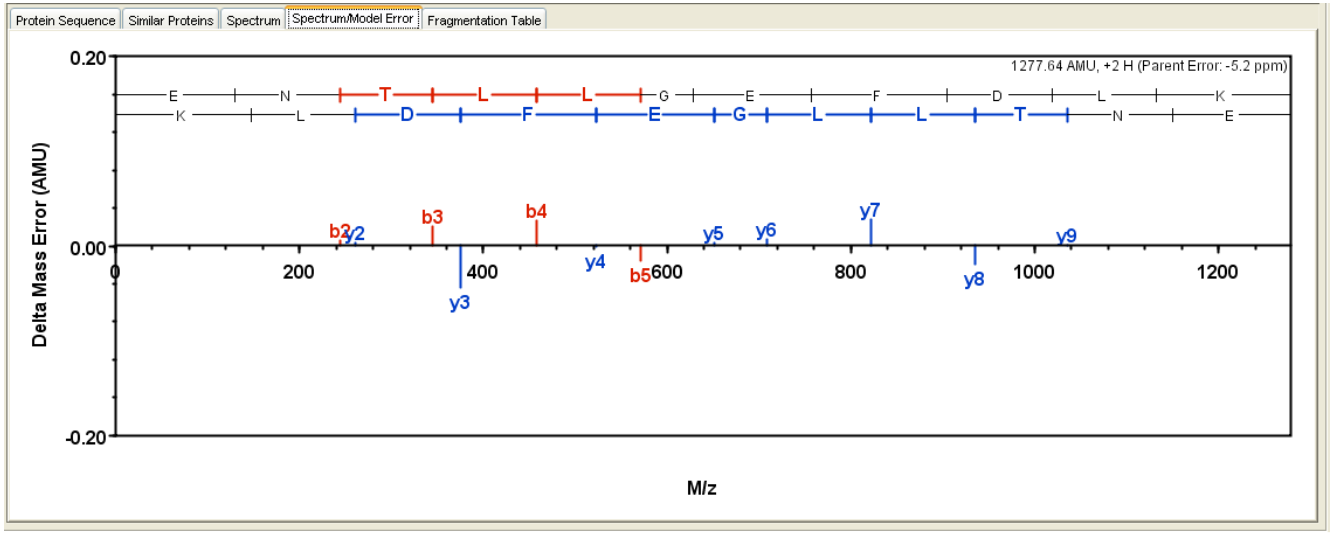
Lower Window 3

- The Spectrum tab displays the MS/MS spectra which the mass spectrometer generated, this is matched against the peptide in the database which lined up best with the fragmentation pattern
- B-ion and y-ion series are color-coded (red and blue) and the amino-acid sequence is across the top, and the parent ion mass is listed.
- Please note that this is a graphical representation and will differ in appearance slightly from the actual MS/MS spectra generated by the mass spectrometer



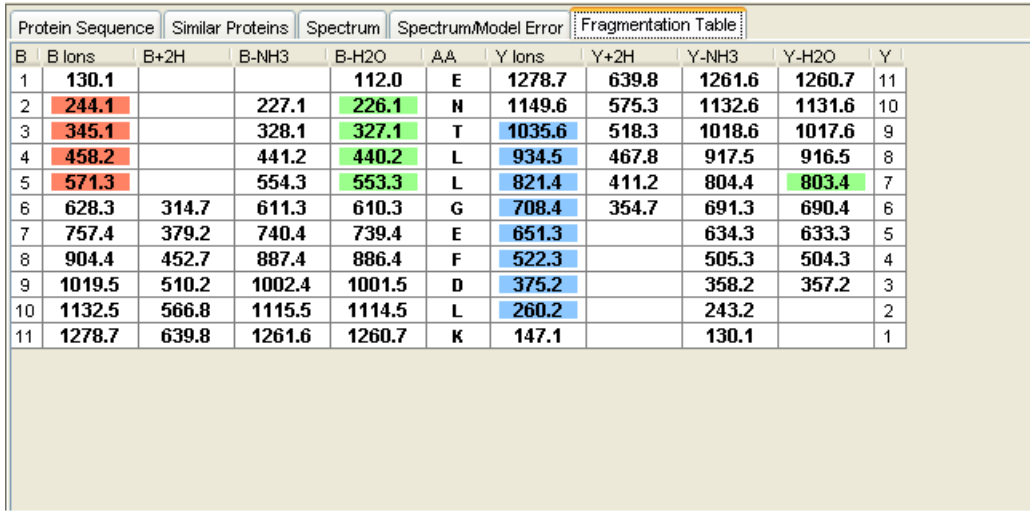
Lower Window 4

- This window displays the Spectrum/Model error
- The bars on the graph show how far the masses recorded by the mass spectrometer differ from the calculated masses.
- When a spectrum and peptide are matched correctly the error for the peaks should match up well to the mass accuracy of the mass spectrometer used.



Lower Window 5

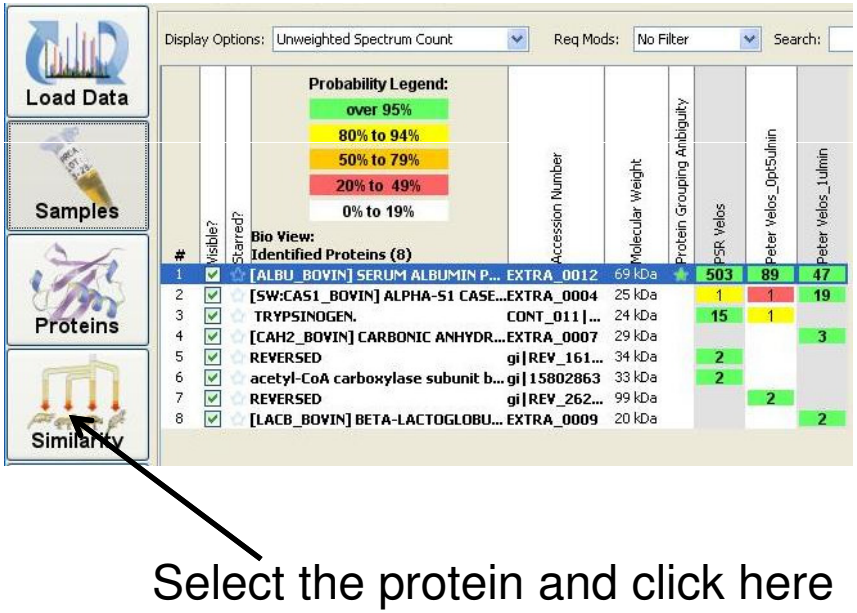
- The fragmentation table displays the same information as the spectrum window, but in a spreadsheet format.
- Potential ions which match the spectra are colored (these colored boxes are the lines in the spectra window)
- Green boxes refer to neutral loss or similar fragmentation patterns; this is the same as the green bars in the spectrum window



B	B Ions	B+2H	B-NH3	B-H2O	AA	Y Ions	Y+2H	Y-NH3	Y-H2O	Y
1	130.1			112.0	E	1278.7	639.8	1261.6	1260.7	11
2	244.1		227.1	226.1	N	1149.6	575.3	1132.6	1131.6	10
3	345.1		328.1	327.1	T	1035.6	518.3	1018.6	1017.6	9
4	458.2		441.2	440.2	L	934.5	467.8	917.5	916.5	8
5	571.3		554.3	553.3	L	821.4	411.2	804.4	803.4	7
6	628.3	314.7	611.3	610.3	G	708.4	354.7	691.3	690.4	6
7	757.4	379.2	740.4	739.4	E	651.3		634.3	633.3	5
8	904.4	452.7	887.4	886.4	F	522.3		505.3	504.3	4
9	1019.5	510.2	1002.4	1001.5	D	375.2		358.2	357.2	3
10	1132.5	566.8	1115.5	1114.5	L	260.2		243.2		2
11	1278.7	639.8	1261.6	1260.7	K	147.1		130.1		1

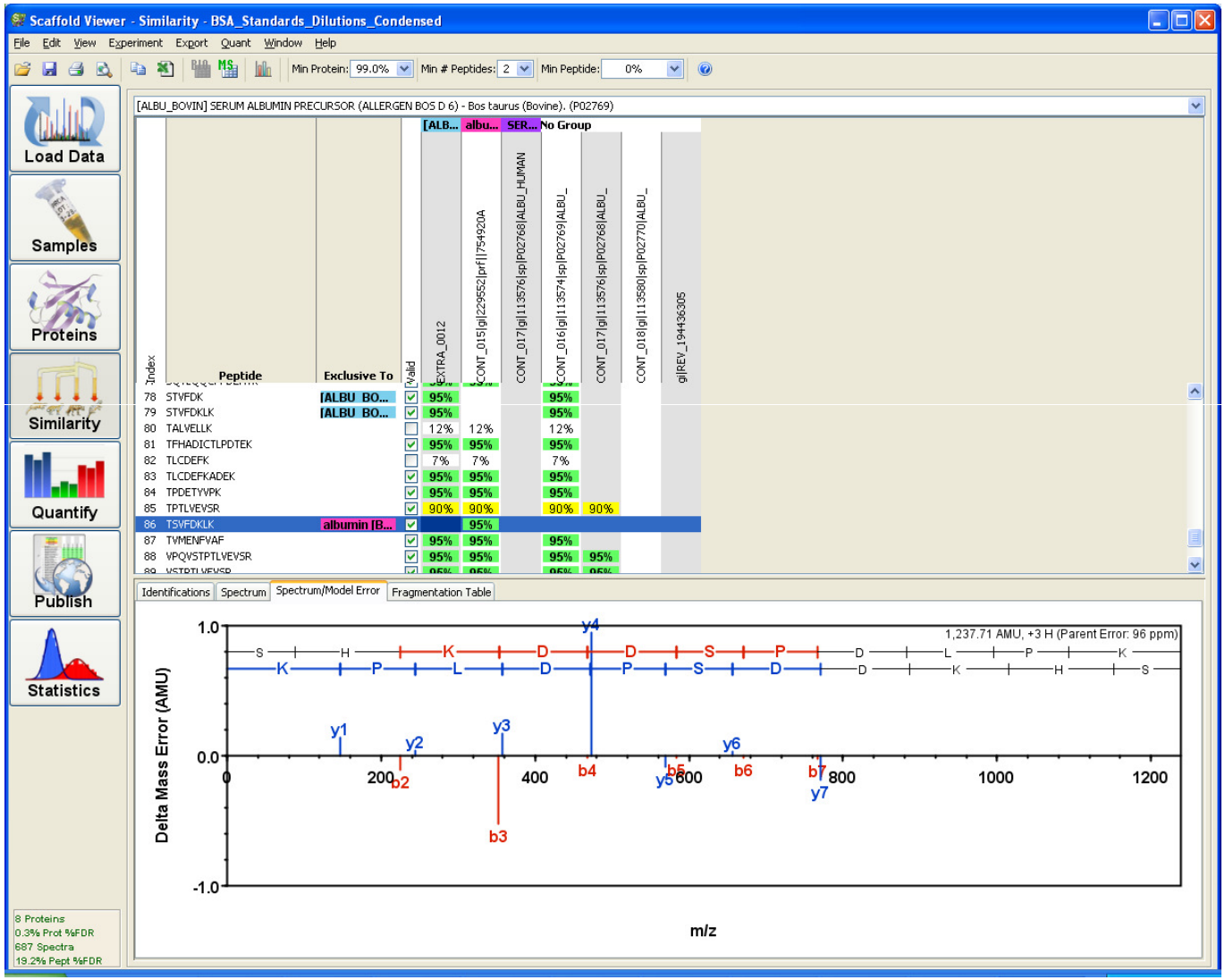
Protein Similarity

- A red star in the Protein Grouping Ambiguity means there are proteins that have shared peptides that haven't been examined in the similarity tab yet.
- Selecting that protein and clicking on the similarity tab will allow you to sort through the peptides.



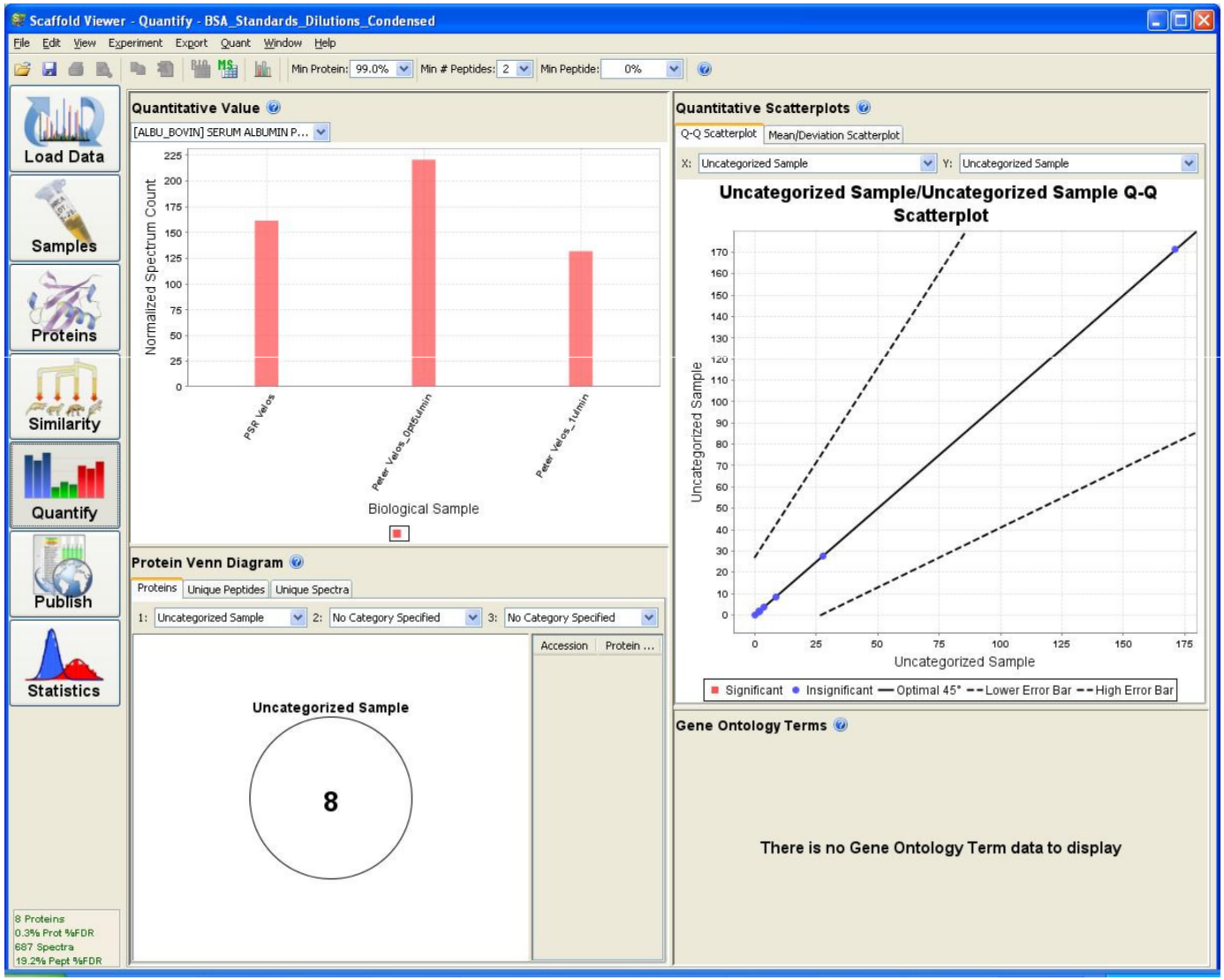
Similarity Tab

- Each peptide is listed
along with the proteins it is found in.
- The spectrum viewer at the bottom allows you to critique individual peptide identifications
- Checking or un-checking the “valid” box will add or remove that peptide from your data.
- If all the unique peptides from a protein are removed it will disappear from your list of identified proteins on the Samples tab.



Quantify Tab

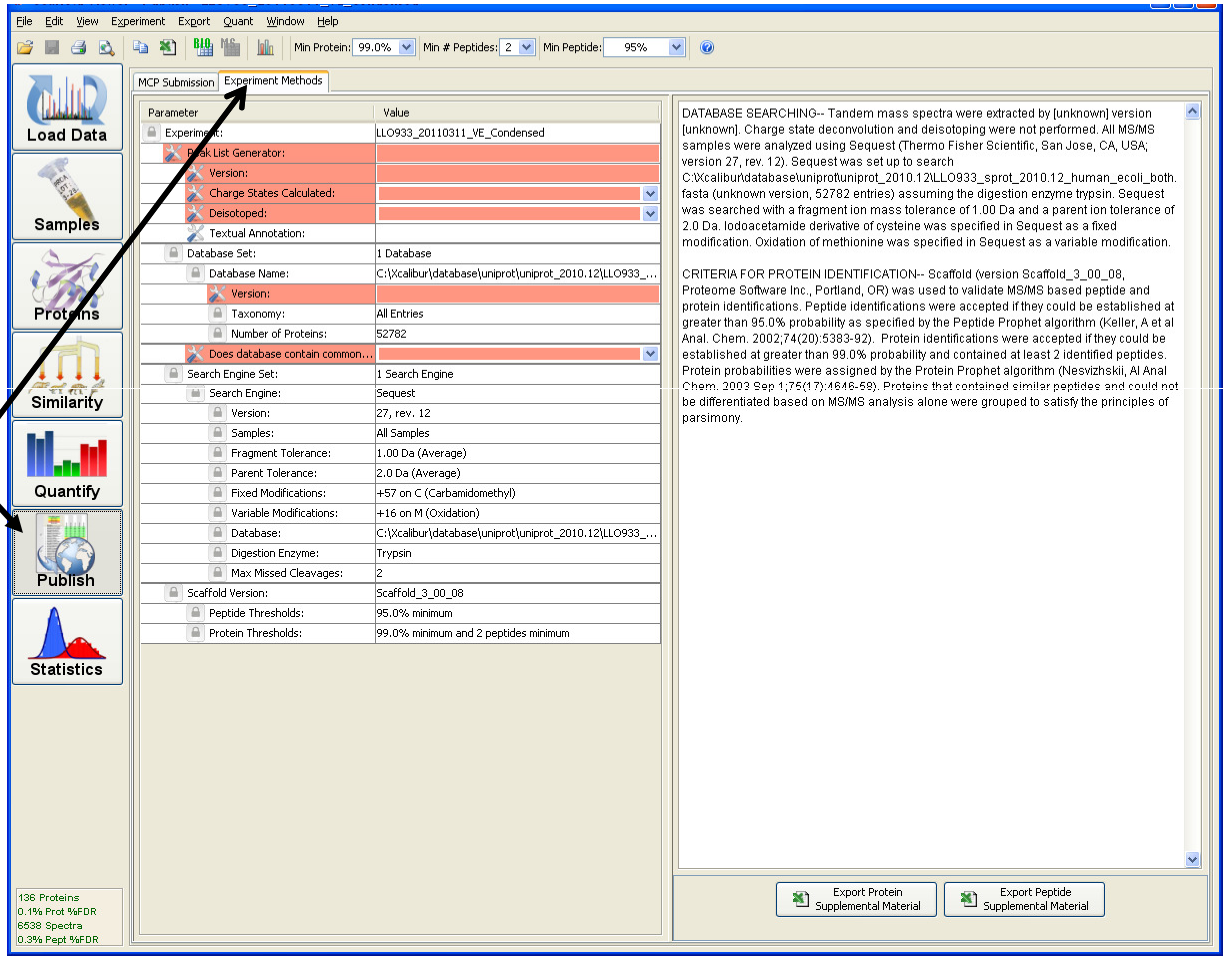
- The quantify tab has several options for analyzing your data.
-
- Contact PMAF if you are interested in a more in-depth statistical treatment of your data.



Publish Screen

Click here To get to Publish screen

Then click here to see the methods summary



Experimental Methods

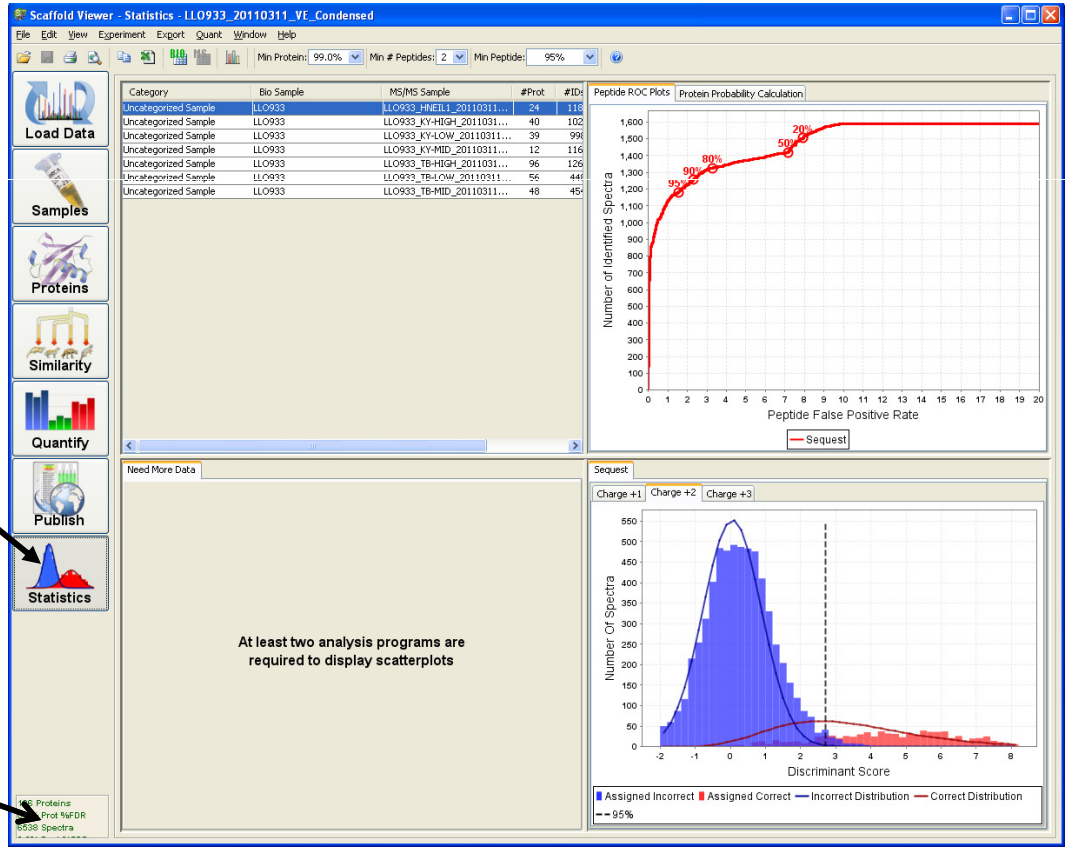
- The Experimental Methods tab contains a couple of short paragraphs suitable for the methods section of a paper using the settings that software was run with, the way data has been filtered in the “samples” tab, and variables entered into the lines on the left side of the screen.
- The paragraphs are written on the right side of the screen. This data can be transferred to a document program (i.e. MS Word) by highlighting it, right-clicking and selecting “copy.” Then you can “paste” the paragraphs into the document program with the same method.
- The corresponding data in the “samples” tab can be exported to excel using the tabs on the bottom of the screen
- While helpful the information in this screen is rather generic, and it’s a good idea to contact PMAF and request a methods summary if you are going to present/publish the results.

Statistics

-Displays information relating to the software used to match the MS/MS spectra to the amino-acid sequences in the database, and which make probability estimates based on this information. Note: this page can take a long time to open with larger datasets.

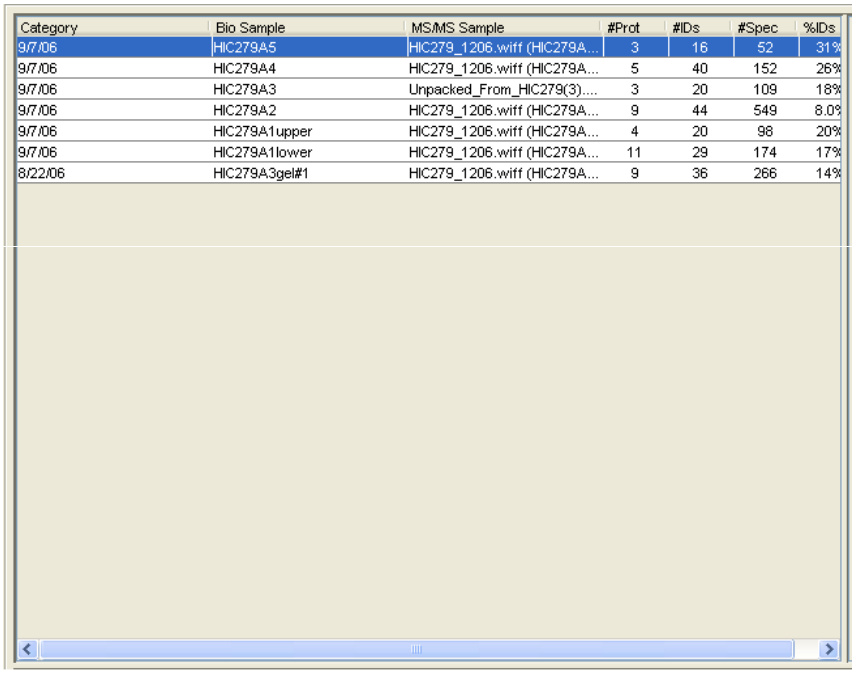
Click here to View Statistics

There is also a brief stats summary here that can be viewed from any page



Upper Left Window

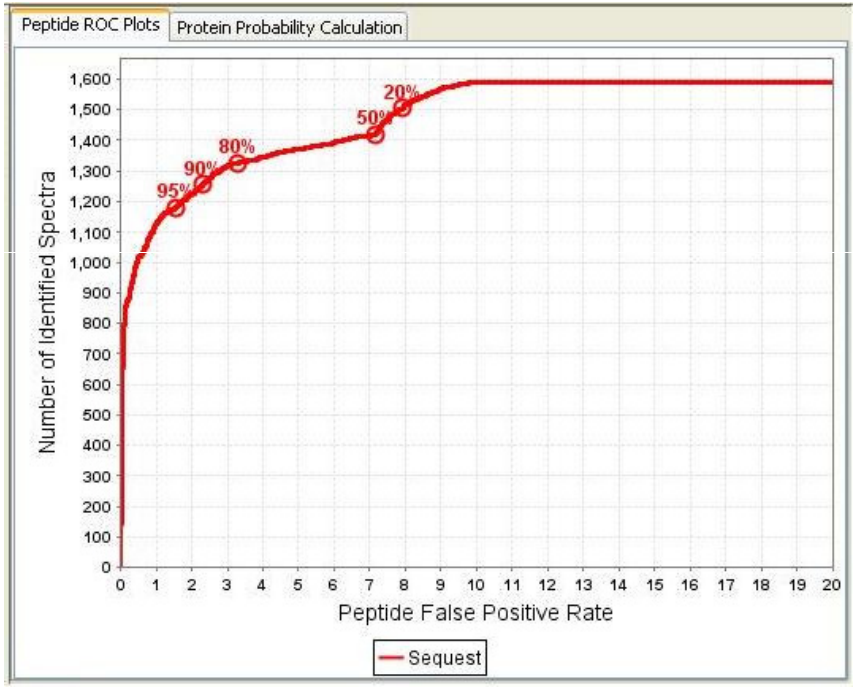
- This window lists the different samples from the “samples” tab
- The data displayed in the other 3 windows is for the sample highlighted in this window
- Note that any of the other fields may be blank if there is not enough data in the sample



Category	Bio Sample	MS/MS Sample	#Prot	#IDs	#Spec	%IDs
9/7/06	HIC279A5	HIC279_1206.wiff (HIC279A...	3	16	52	31%
9/7/06	HIC279A4	HIC279_1206.wiff (HIC279A...	5	40	152	26%
9/7/06	HIC279A3	Unpacked_From_HIC279(3)....	3	20	109	18%
9/7/06	HIC279A2	HIC279_1206.wiff (HIC279A...	9	44	549	8.0%
9/7/06	HIC279A1upper	HIC279_1206.wiff (HIC279A...	4	20	98	20%
9/7/06	HIC279A1lower	HIC279_1206.wiff (HIC279A...	11	29	174	17%
8/22/06	HIC279A3gel#1	HIC279_1206.wiff (HIC279A...	9	36	266	14%

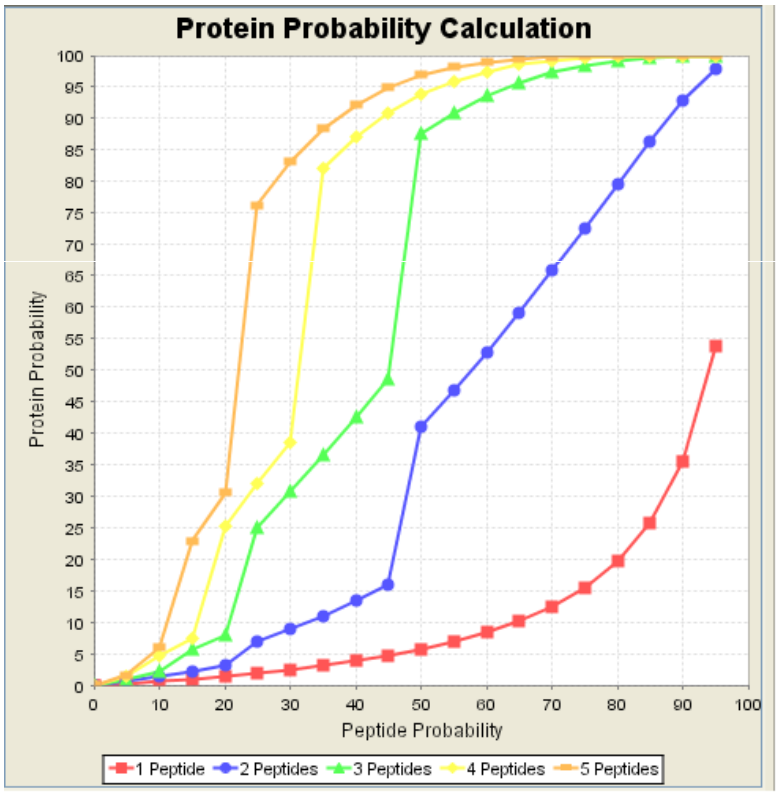
Upper Right Window

- This window displays both a ROC plot and an in-depth analysis of the protein probability calculation.
- The ROC plot displays an estimated peptide FDR against the number of identified spectra.
- Please note that this often greatly underestimates the amount of error, when compared to PSR's internal standards.
-



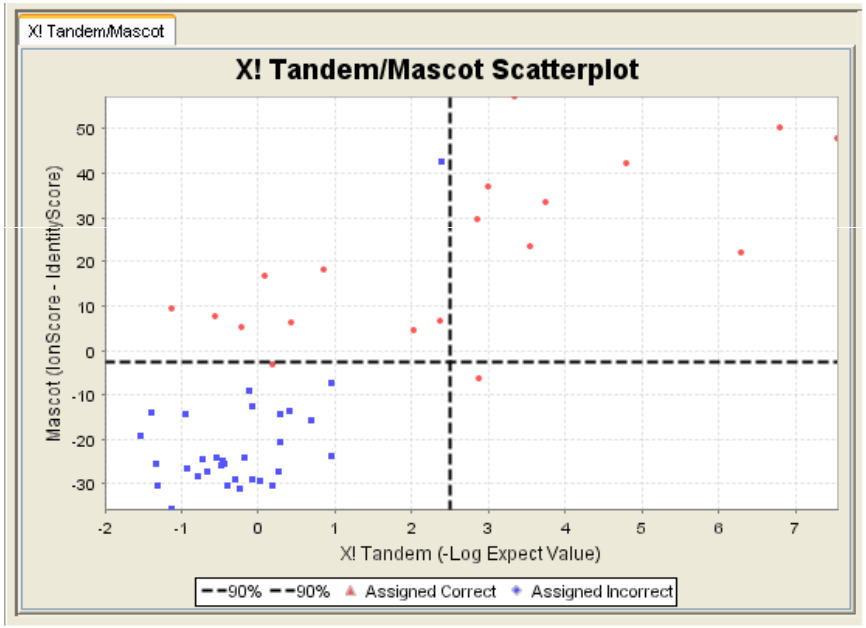
Upper Right Window

- The protein probability window displays the relationship between peptide probability, # of peptides and protein probability
- Note that the # of peptides found strongly affects the protein probability
- Also note that with 95% probability on a single peptide this only relates to about a 50% probability of the protein being present
- Often 2 to 3 high probability peptides are necessary to have a confident protein identification



Lower Left Window

- This window displays a scatter plot of the X! Tandem and other search engine scores for each identified peptide.
- This field is useful for comparing the search engines and evaluating how useful they are to your dataset
- Note that if X! Tandem was not run on your dataset then this field will be blank



Lower Right Window

- This displays the calculated curves which the peptide identification algorithm uses to calculate probabilities
- Scores are sorted by value and 2 curves are matched to the distributions
- The degree of overlap of the two curves relates to the peptide probability