

Scaffold Analysis

Scaffold is a computer program that integrates search results from three algorithms (Sequest, X! tandem and Mascot) to generate peptide identification and protein identification probabilities. Scaffold also displays an overview of the protein identifications that can be validated by probability scores. Protein information can also be used to detect false positives and also examine the peptide and spectral evidence used for identification.

Getting Started :

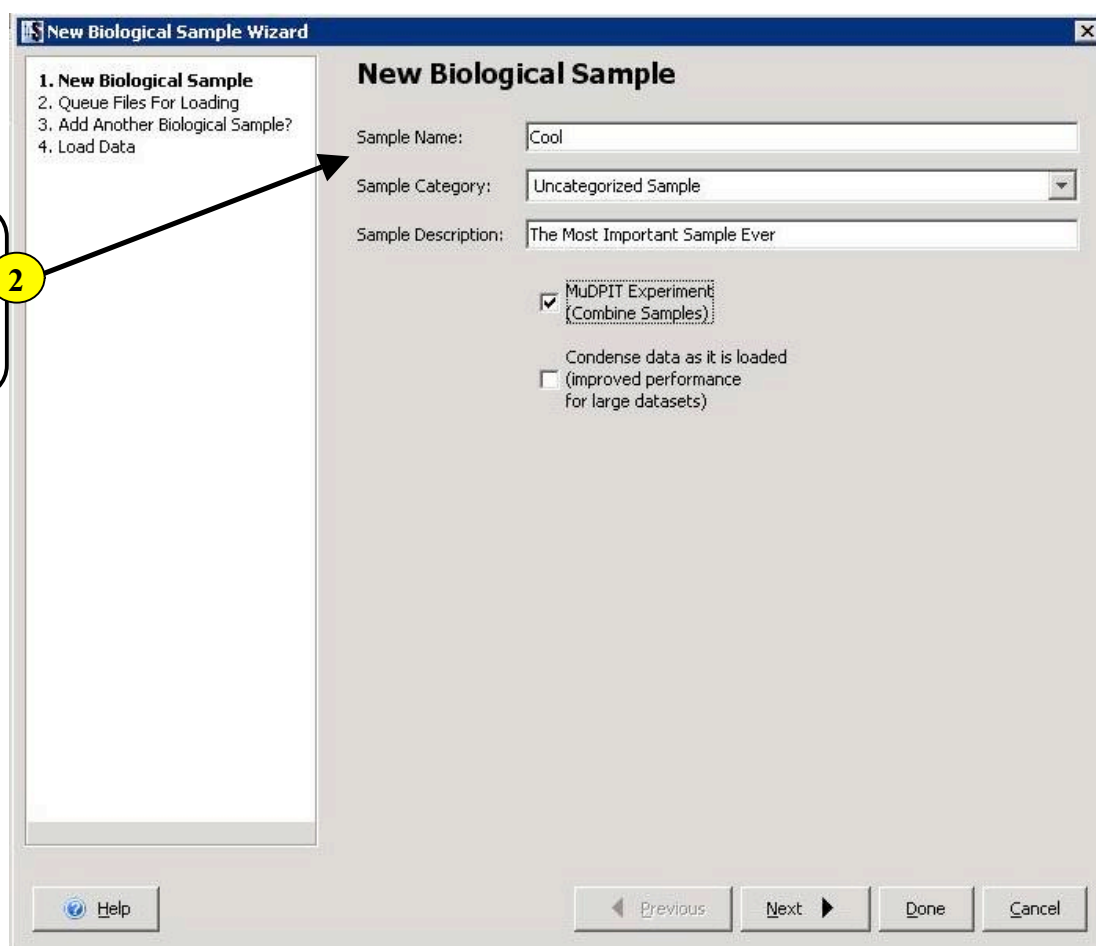
Click onto “New” for starting new Scaffold analysis. Click on “open” to analyze a previously saved Scaffold (.sfd) file.

1



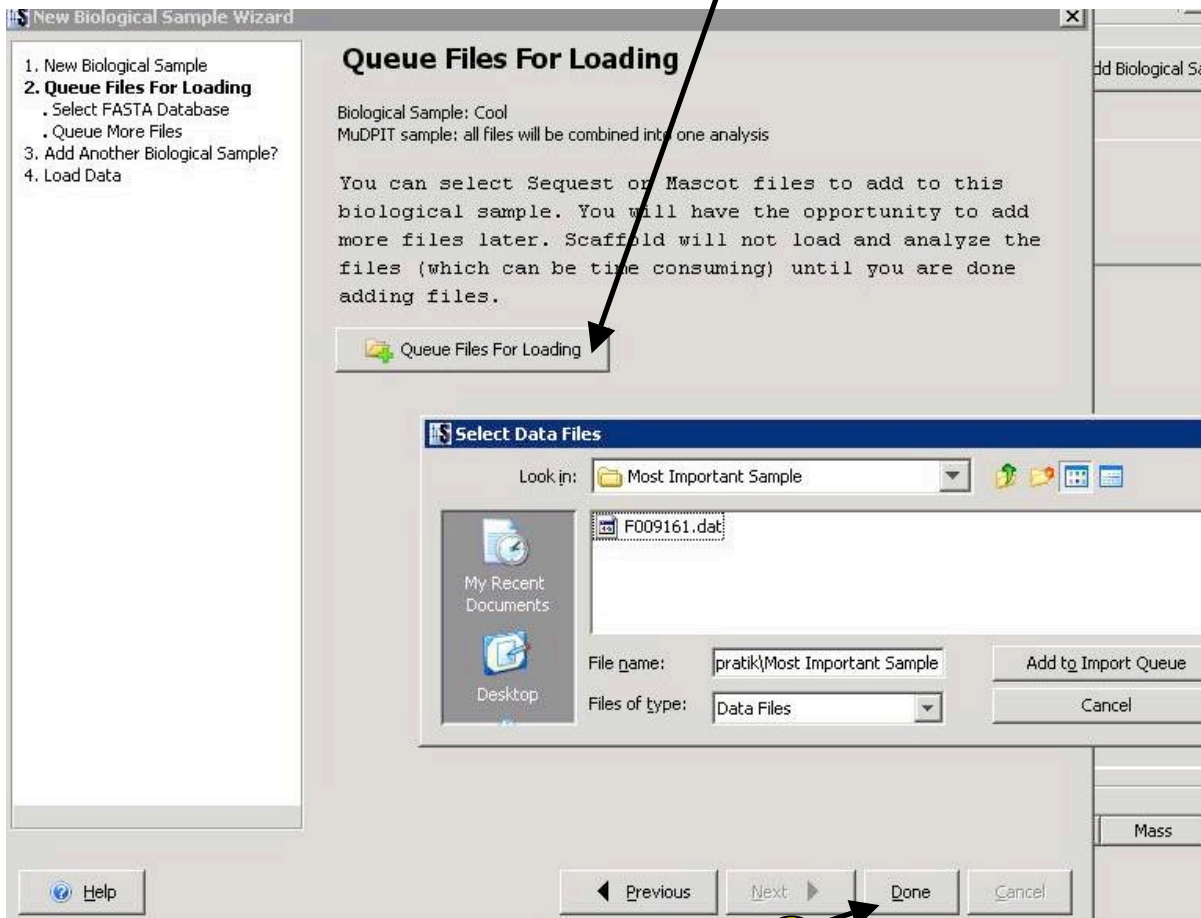
Enter sample name and description to set up a new analysis.

2



Add either .dat file that was stored from earlier Mascot search OR Zipped folder from .dta and .out files from Sequest search OR .xml file generated from X! tandem search.

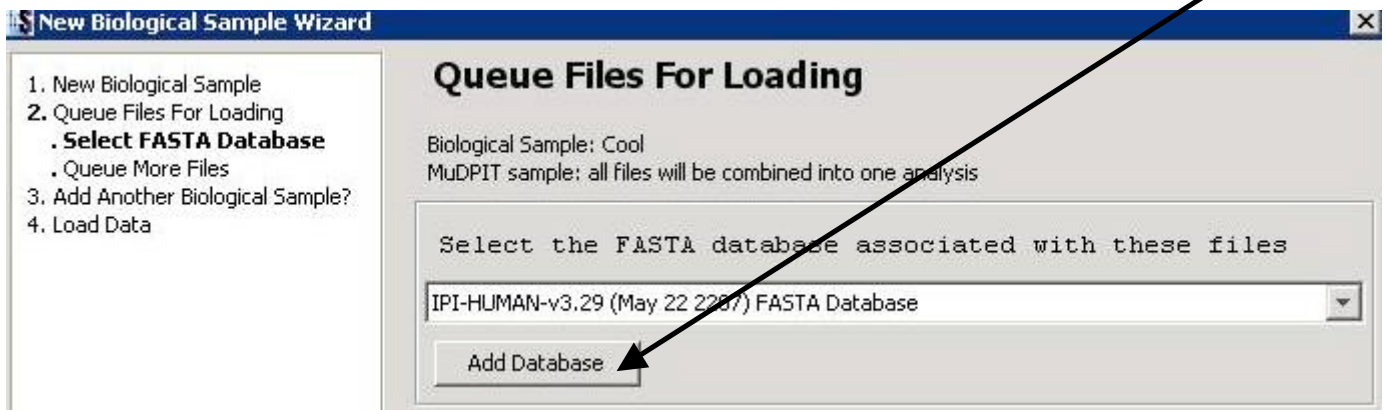
3

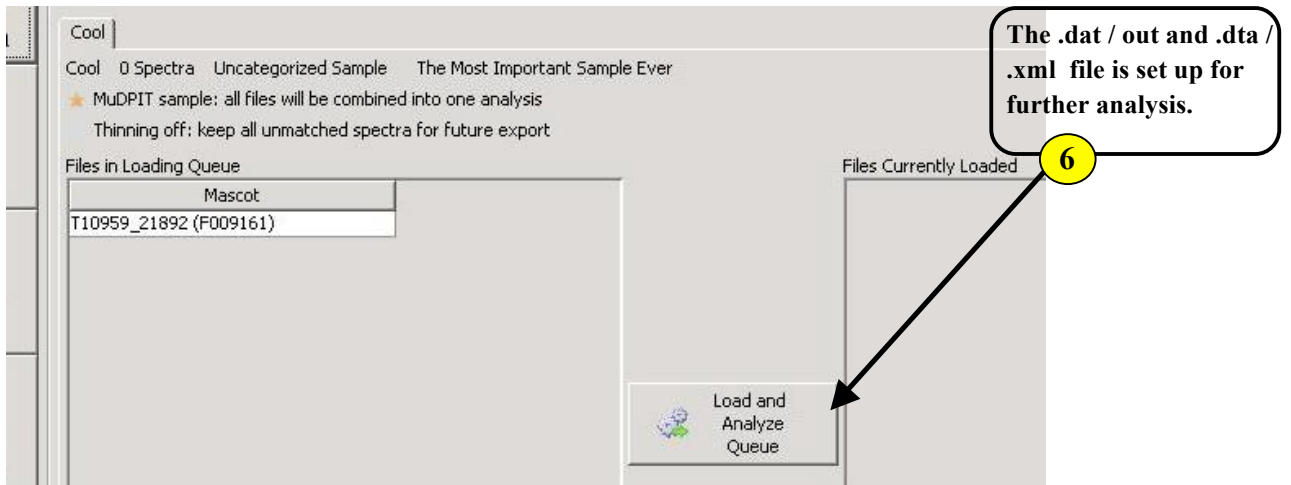


Click on done to queue files for analysis.

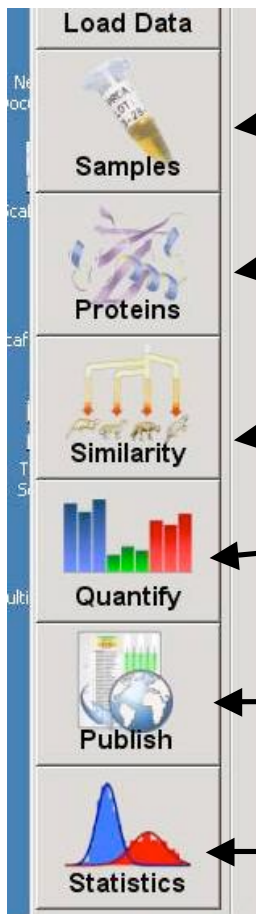
Add FASTA file that was used for earlier search to the database.

5





Analysis is initiated and the progress is displayed on the screen. Once the analysis is completed various tabs below can be clicked and information therein can be explored further.



7 "Samples" has information on probability of protein identification, name, accession number and biological processes and molecular function of proteins identified.

8 "Proteins" has information about protein coverage.

9 "Similarity" has information regarding peptides that are used for protein identification and spectral information that includes Spectrum / Model error and Fragmentation Table.

10 "Quantify" has information about spectral counts, quantitative scatterplots and Gene Ontology term distributions.

11 "Publish" has information that can be used for publishing manuscript with data that was analyzed.

12 "Statistics" has information regarding distribution of "correct" and "false" hits based on probability thresholds.

Once the analysis is completed, "Samples" tab opens and displays protein information. Most importantly, the Minimum Protein and Minimum Peptide thresholds can be adjusted. Moreover, number of peptides identified per protein can also

13

The screenshot shows the Scaffold software interface. At the top, there are menu options: Experiment, Export, Window, Help. Below the menu is a toolbar with icons for file operations and analysis. The main window displays a table of identified proteins with columns for #, Visible?, Protein Starred?, Gene Symbol, Accession Number, Molecular Weight, Protein Grouping Ambiguity, Taxonomy, Biological Processes, Cellular Compartments, and Molecular Functions. A 'Probability Legend' is shown on the left, with color-coded bars for probability ranges: over 95% (green), 80% to 94% (yellow), 50% to 79% (orange), 20% to 49% (red), and 0% to 19% (white). Below the table, there are sections for 'Protein Information' and 'Sample Information'. The 'Protein Information' section includes fields for 'Lookup Accession Number In' (set to NCBI), 'Preferred Accession Number' (IPI00418169), and 'Protein Name' (Gene_Symbol=ANXA2 annexin A2). The 'Sample Information' section includes fields for 'Biological Sample', 'Sample Category', 'Sample Description', 'MS/MS Sample', and 'MS/MS Sample Notes'. A yellow callout box with the number 13 points to the 'Min Protein' and 'Min # Peptides' settings in the top toolbar.

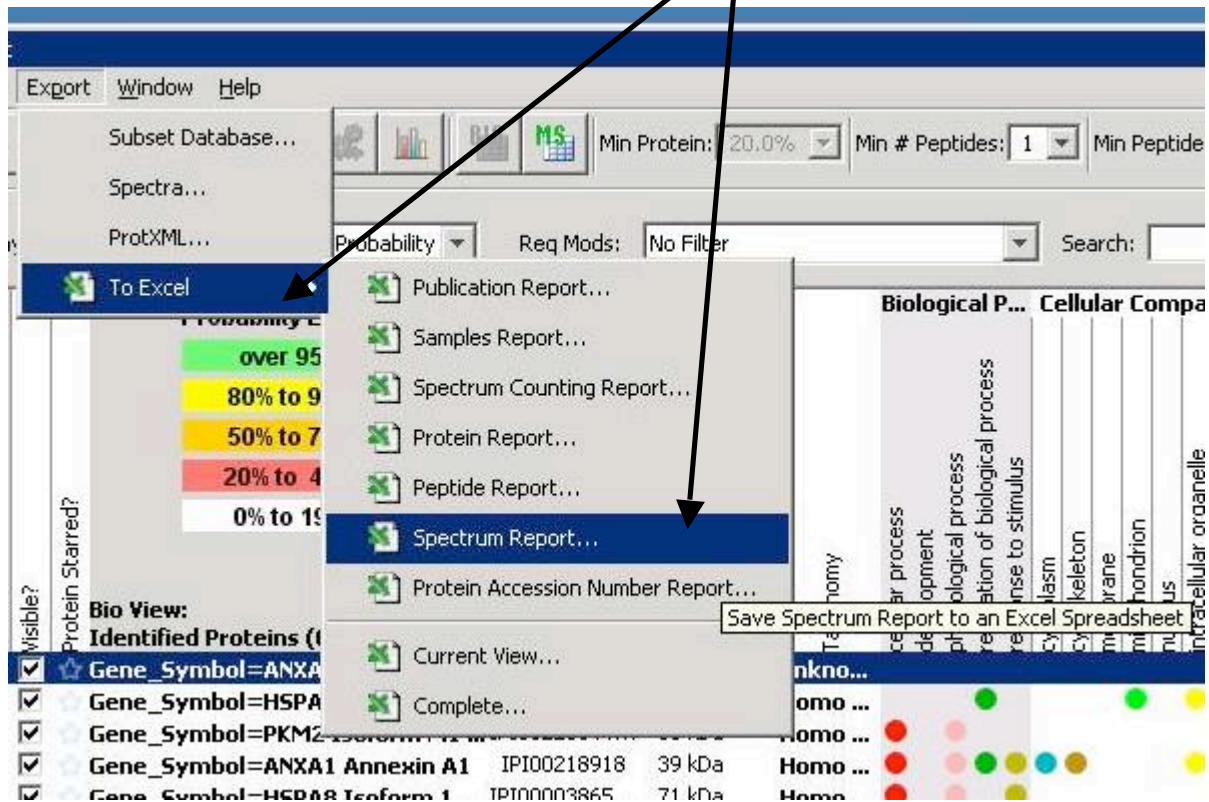
Before proceeding further it is generally recommended that the Scaffold file is saved in .sfd format, especially if you intend to use the analysis at a later stage.

14

The screenshot shows the 'File' menu of the Scaffold software. The menu items are: New (Ctrl+N), Open (Ctrl+O), Merge with SFD... (Ctrl+M), Close, Save (Ctrl+S), Save As..., Save the Current File With a New Name, Print... (Ctrl+P), Print Preview..., and Exit. A yellow callout box with the number 14 points to the 'Save As...' option.

Scaffold Export offers an opportunity to download various reports that are generated from analysis. One of this is “Spectrum Report” that carries information about peptide identification that we routinely use for further analysis.

15

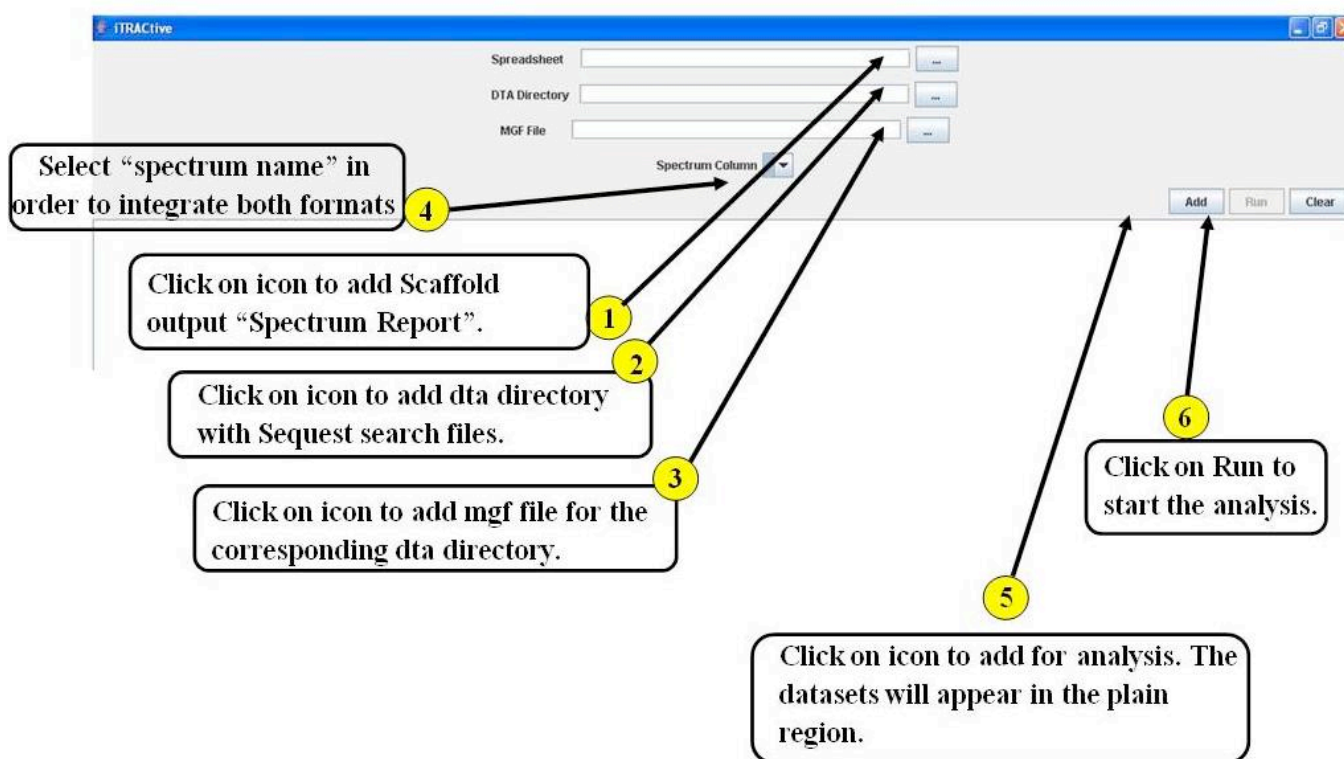


We generally use Scaffold analysis to generate Spectrum Report that contains information regarding peptide identification. In most cases, we use iTRAQ labeling for quantitative analysis. In such cases, we need to integrate peptide identification information from Scaffold spectrum report to iTRAQ peak areas associated with iTRAQ reporter groups for further quantitative analysis. For this we use a program developed in our lab by Panagiotis Papoulias called iTRAQtive. The following section deals with use of iTRAQtive for integrating quantitative information with peptide identification.

iTRAQtive Guide

iTRAQtive 1.0 is a tool used for Mascot and X! tandem peptide identifications from Scaffold spectrum report to merge quantitative information from the title line. It is a tool that was primarily developed for associating dta formatted peaklists used for Sequest search with its parent mgf file.

In Sequest search Scaffold outputs, iTRAQtive associates the reporter group areas stored in the mgf file with the dta formatted peaklists used in Sequest searches. An MD5 hash is calculated for each peaklist in the mgf file and for each dta peaklist used by Sequest. It then compares the MD5 hash of the individual peaklists in the mgf formatted file against the MD5 hash of the peaklist in the dta formatted file and integrates the result of the match into the Scaffold report.



Finally, this modified spectrum report that contains both quantitative information associated with peptide identification is subjected to normalization and quantitative analysis.