

MS & Proteomics Resource

Yale School of Medicine
Keck Biotechnology Resource Laboratory



Application Note 7: Using the Yale Protein Expression Database (YPED) to view MS & Proteomics results

YPED was designed to address the storage, retrieval, and integrated analysis of high throughput proteomic and small molecule analyses. For proteomics data, YPED handles data from the following analyses: LC-MS/MS protein identifications and protein posttranslational modification (i.e. phosphorylation, ubiquitinylation, acetylation, methylation, etc.); identification/quantitation results from label-based proteomics experiments (such as DIGE, iTRAQ, ICAT, and SILAC experiments); LCMS-based label-free quantitative proteomics; and targeted proteomics (MRM).

After logging into YPED (<http://yped.med.yale.edu/>), you need to click on the sample of interest from the table. This will pull up the Mascot search results which have been imported into YPED. The top header will tell you the date of the analysis, the Mascot version used, the database and taxonomy (i.e. tax: Homo sapiens (human)) and then formation such as the LC-MS/MS run number.

The next section will show you the **Protein Score Threshold**, the number of peptides matched above the **identity threshold**, and those above the **homology or identity threshold**. The protein score is the sum of the ions scores with the latter based on the calculated probability that the observed match between the experiment and the database is a random event. The **identity threshold** is calculated from the # of trials (or # of candidate peptides with the same precursor MW in the database). The **homology threshold** is an empirical measure of whether the match is an outlier. The Decoy database shows the number of hits to this randomized or reversed sequence and a False Discovery Rate (FDR) is calculated. The FDR can be set to 1% if requested.

The next section in YPED has links out to

- View the Sample information
- View the search parameters used for the Mascot database search such as methionine oxidation
- View information on the total list of peptides in the sample
- View proteins with sequences that are indistinguishable from other proteins in the database searched. This includes isomers and proteins with a high homology to other database entries
- Panther Summary – this links out to The Panther Classification system (<http://www.pantherdb.org/>) to assist the researcher in Pathway Analysis, Biological Process and Molecular Function

The main table summarizes the results from the mascot search. The **Score** and **Expectation** values in the first

2 columns have been determined by the Mascot search algorithm (see Application Note 4 for details on this calculation). The highest score (in the first column) and with a low expectation value (in the second column) will be the best matches. YPED is sortable so you can sort any column by protein ID, Protein Name, MW, or % coverage. When the file is first opened, it is sorted by score. Clicking on the Protein ID will link out to the database used for searching and this accession number. The % coverage shows where the peptides that were identified match to the indicated sequence. Clicking on the view link in the peptides column, will show the peptides identified with the *peptide* score (not the protein score) and the *peptide* expectation values.

In the **peptide view table**, the m/z column is the mass to charge ratio. Using electrospray mass spectrometry, the masses observed are typically +2 and +3 charge states. The charge state is found in the last column in this view. Hence, a m/z value of 679.35 which is doubly charged, has a peptide mass (found in the Ion Mass column) of M=1356.7 and an M+H= 1357.7. The Ion Mass (calc) column shows the theoretical mass. The Delta column is the mass difference between the Ion Mass (which is from the experiment) and the Ion Mass (calc) and is in Da. The ppm column shows you the mass difference as calculated in parts per million.

The peptides in the **peptide view table** are grouped based on the identity and homology scores. Mascot calculates these 2 scores for each peptide. The scores are not fixed and will change. The homology threshold helps determine if the peptide match is an outlier. Hence, the best peptide matches will be above the homology score. Those below this score are shown in the bottom of the table. Peptides with scores higher than the homology score but less than the calculated identity score are in the middle section of the table. The identity score is based on the probability of getting a false positive match. Hence, the best peptide matches are seen in the top part of this table with scores above homology and above the identity score.

Excel tables can be created by clicking on the Export options at the bottom the tables (for both the protein view and the peptide view).