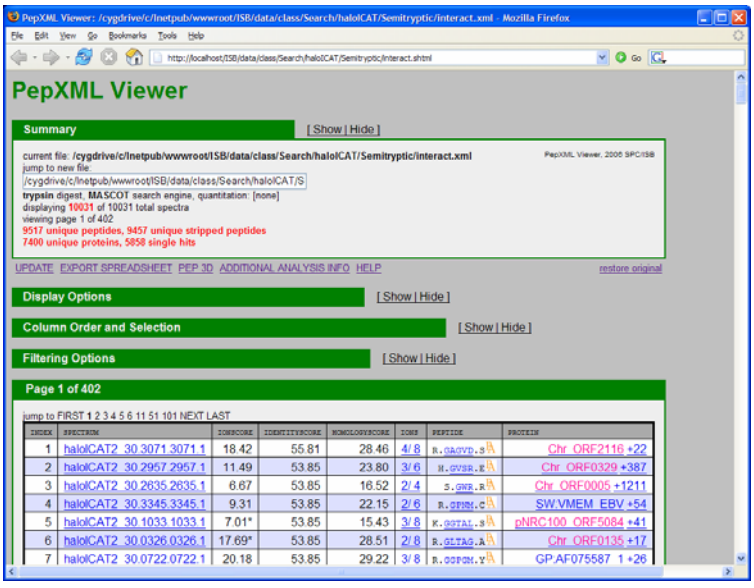
10/16/2006
ISB/SPC Software Proteomics Course, Day 1 Exercises

Manually inspect peptide identifications.

Can you tell a good database search result versus a bad one? Here's your test. Go to the home page of the browser if you're not already there. Click on the "Spectrum Exercise" link. Look at the peptide sequence, search scores, and spectrum view to determine whether or not the identifications are correct. Then look at the drop-down box the correct answer for each entry. For this first exercise, the class will go through the first 10 or so search results together. You can finish the rest of the exercise individually, time permitting; shout out any perplexing examples.

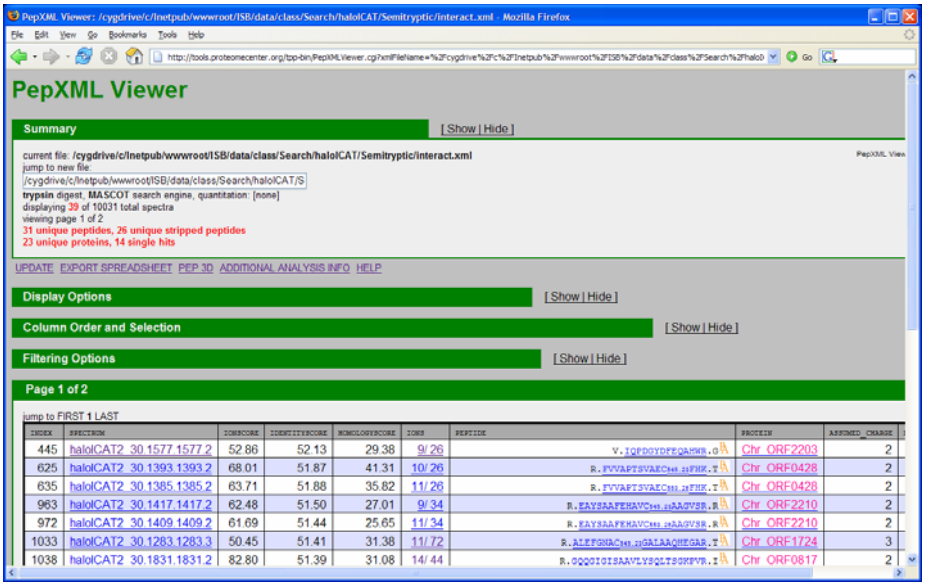
Using the pepXML viewer

1. Click on the home page icon in your browser to go to the URL http://localhost/tpp-bin/tpp_gui.pl. Log in to the TPP graphical user interface using the username "guest" and password "guest". Under the "tools" tab, use the File Browser to browse to the "class\Search\haloICAT\Semitryptic\" directory. Click on the "View" link for the file "interact.shtml". This will open up a pepXML viewer in a new browser window with Macot search results. The search is of a halobacterium dataset that was isotopically labeled with the original ICAT reagent.



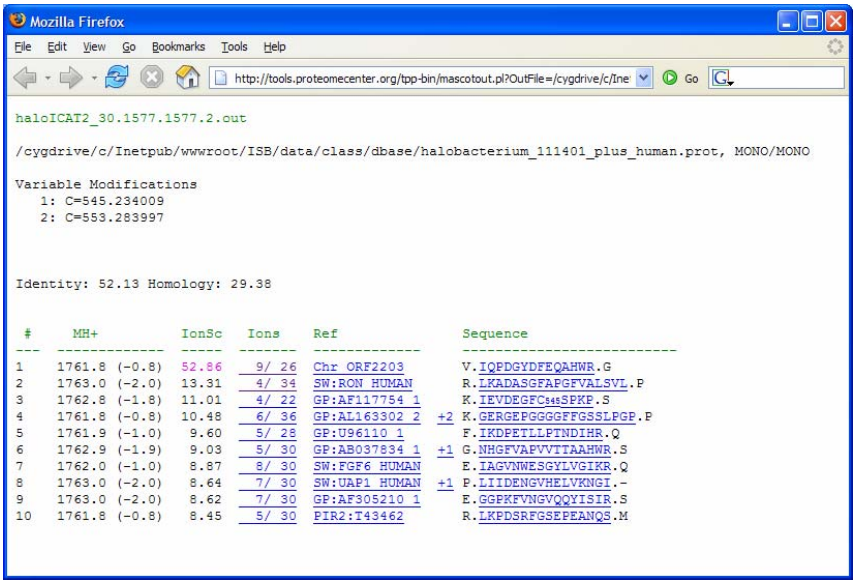
2. First, let's add "assumed_charge", "num_tol_termini" and "calc_neutral_pep_mass" columns to the page view using the "Column Order and Selection" pane.

3. The first filtering we will apply is a minimum ion score of 50. Note that this reduces the number of search results from 10031 down to just 39.



4. These 39 entries are relatively high confidence peptide hits based on an ion score cutoff of 50. As this is a trypsin-digested sample searched with a semi-tryptic constraint, find the 2 remaining search results where the number of tolerable termini is 1 e.g. show semi-tryptic behavior. Which of the 2 do you believe are correct?

5. For the search result associated with spectrum "haloICAT2_30.1577.1577.2", click on the spectrum name to display the top 10 peptide hits for this spectrum. Click on the "ions" link for the first hit and compare it with the "ions" link with the second hit. Does the first hit appear to be a significantly better match then the second hit? Note the "Ion Score" of the top his is 52.86 compared with 13.31 for the second best hit.

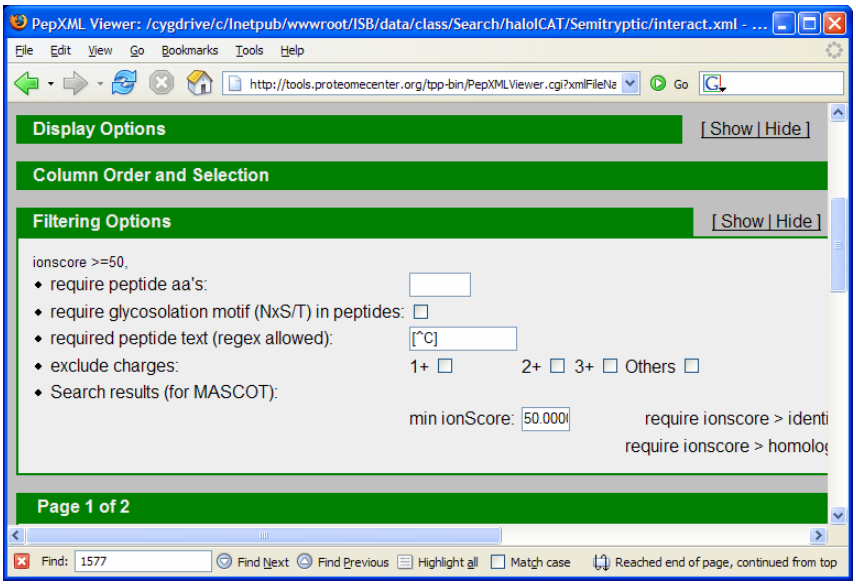


6. Now sort this list in ascending order based on the "protein" column using the "Display Options" pane. In the display options, also increase the "rows per page:" option to "All" to

display all 39 entries. Of the 14 single hit proteins on the list, how many do you believe are correct identifications?

7. Since the ICAT protocol includes a biotin affinity purification step to isolate/enrich for cysteine containing peptides, we should only see cysteine containing peptides in the search hits. First, let's highlight all cysteine residues by placing a "C" in the "Display Options → highlight peptide text (regex)" field. Note that all cysteine residues are now highlighted in pink. This is a useful visualization option to look at specific peptide residues or residue motifs. For example, the regular expression "[N].[ST]" will highlight the peptides that contain the N-link glycosite consensus motif NxS/T.

8. In the "Filtering Options" pane, apply the peptide regular expression "^[^C]*$" to exclude every peptide that contains a cysteine. You should be left with the 3 peptides non-cysteine containing peptides. Do these appear to be correct identifications?



9. Click on the "restore original" link. This link stores you to the base state before any of your filters, sort, display options have been applied. Effectively, this is a "start over" link.

10. Note the "Export Spreadsheet", "Pep3D", "Additional Analysis Info" and "Help" links.

Go ahead and explore the interface as desired. Take a look at the available columns and how you can re-order them. Explore your ability to sort/filter based on various properties of the table. You will be using this view for the next few days so you should definitely become comfortable navigating the interface.