

Lab Exercise 8: Analysis of MS/MS data (12/03/04)

Tools:

1. PeptideSearch tools from the Bioanalytical Research Group at EMBL:
<http://www.narrador.embl-heidelberg.de/GroupPages/Homepage.html>
(Contains the tools we need to use in the homework for peptide and protein identification.)
2. ExPASy Proteomics Server:
<http://us.expasy.org/>
(Comprehensive collection of tools useful in proteomics.)

Questions:

1. Swiss-2DPAGE database:
As we learned in the lecture, 2-D PAGE is a technique that are often used for separating mixture of proteins based on their pI and MW. ExPASy has a link to the Swiss-2DPAGE database (<http://us.expasy.org/ch2d/>), which contains information on proteins identified on 2-D PAGE reference maps.

Now, let's try to get access of these data. Click on "by clicking on a spot" under "Access to SWISS-2DPAGE". Here you will see a collection of reference maps for different tissues and species. Select the map for *Saccharomyces cerevisiae*.

Here we have a reference 2-D PAGE for budding yeast. Note that some of the points on the gel are marked with red cross. Those are the identified proteins. Click on one of the marked spots that have a pI ~6.0 and Mw slightly less than 50 Kd. You will get a list of data on this protein. In the section "2D PAGE maps for identified protein, there are a list of map locations corresponding to this protein. It's often the case that you would get more than one map location for the same protein. Click on one of the locations listed. You will see the locations of this protein marked on the reference map together with a rectangle box around the theoretical pI and Mw values for that protein. The box indicates the region where one would expect to see the protein on the gel.

Now, suppose you are interested in the yeast protein Glyceraldehyde-3-phosphate dehydrogenase 3 (TDH3p) and want to locate it on the yeast reference 2-D PAGE. Retrieve the protein sequence from NCBI. Use the tool from ExPASy (http://us.expasy.org/tools/pi_tool.html) to calculate the pI and Mw value for this protein. Can you find the spot(s) corresponding to this protein on the reference map?

2. Identification of proteins from mass spectrum data:

a) By amino acid sequence of a peptide:

As we can see, identifying proteins just based on their location on the 2D-PAGE is not easy. But suppose we know the amino acid sequence of a fragment of it, say a small peptide produced by Trypsin digestion of the protein. It would be much more straightforward to identify the protein.

Suppose we learnt the amino acid sequence of a peptide from DeNovo-sequencing of MS/MS data (or some other sequencing method). The sequence contains 12 amino acids: RHPEYAVSVLLR. Try to use the “amino acid sequence” option in the PeptideSearch homepage. Search for proteins with mass between 0 and 300 kDa that has the above peptide in unspecified organisms. How many results did you get from this search?

Retrieve the protein sequence from one of the results and use the PeptideCutter at ExPASy (<http://us.expasy.org/tools/peptidecutter/>) to calculate theoretical Trypsin digestion products of this protein. Set the options such that only the Trypsin digestion is predicted and only sites with cleavage probability $\geq 80\%$ are displayed. Can you find the peptide with the above sequence in the predicted product?

b) By sequence tag method:

DeNovo-sequencing of MS/MS data requires sophisticated software. However, oftentimes, you need not to know the entire sequence of the peptide in order to identify the peptide and the protein. Suppose that we learned from Full Scan and Zoom Scan of a protein that the average Mw value of a peptide is ~1440. And from the MS/MS spectrum of that peptide, we identified three neighboring amino acids, YAV, which locates between two peaks with m/z value of 520.6 and 941.0. We could then use the information available and the sequence tag method to identify the peptide and protein.

Click on the “sequence tag” option of the PeptideSearch homepage. Enter 1440 for peptide mass (Average mass). Enter (520.6)YAV(941.0) for peptide sequence tag. Set the Match regions for “1 and 2 and 3”. Allow for only one missed cleavage sites per peptide. Try to search first by “Y-type sequence ions”, then by “B-type sequence ions”. How many results you got in each case? Which of them find the same peptide and protein as in a)?

c) By mass fingerprinting:

Suppose that we do not have any MS/MS spectrum data available. We can still get a list of the Mw values of Trypsin digested peptides from the Full Scan data. Let's try to use these data and the mass fingerprinting method to identify the protein.

Here is a list of the Mw values, which I selected from the predicted Trypsin digestion products from b):

477.6, 711.8, 817.9, 885.9, 927.2, 974.0, 977.1, 1002.2, 1050.2, 2435.8, 1163.3, 1284.5, 1349.5, 1388.5, 1440.7, 1633.8, 2003.4

Try to use the “list of peptide mass” option on the PeptideSearch homepage to identify the protein. Set minimum number of peptide required for protein match to be 5. Allow for at most 1 missed cleavage sites per peptide. Set peptide mass accuracy to be 0.3 Da using Average mass. Select Neutral as peptide charge state.

How many protein candidates does this tool return? Are the proteins predicted by a) and b) included?

Based on this, what do you think about the sensitivity and specificity of the 3 methods in a), b) and c)?