MS data analysis using Peptide Mass Fingerprinting (PMF)

AIM: Get an understanding of the principles behind PMF and parameters that influence the results.

This section deals with the identification of proteins by Peptide Mass Fingerprints. A summary of available tools is given at the ExPASy server (http://www.expasy.org/proteomics/). The one that we will use today is:

Mascot - Matrix Science. (http://www.matrixscience.com)

Before you put your hands on the data, answer the following questions:

1. Why are not all proteases suitable for Peptide Mass Fingerprinting?
2. What is the most commonly used protease? Why?
3. What accidents that will influence the PMF may happen during digestion?
4. What types of amino acid modifications will influence the results?

Remark:

* to run the exercises, you need to download the file MS-Spectra.zip from

http://www.immun.lth.se/education_old/advanced_proteomics_courses/plant-proteomics/

to your machine and then unzip it in a directory that will be easy to find. All peak lists that will be used today will be in this directory.

A: A step by step PMF submission

Do a first submission with the Mascot tool. At the Mascot website, you will find the "Peptide Mass Fingerprint" form. Look at the various parameters (you may click on the labels and read the descriptions). Do you understand all of them? The help feature of Mascot is very complete. Don't hesitate to use it. Now try to identify the protein using the MS spectrum that is in the file ppw_A23.txt.

1. Look at the file format. What information does the file contain?

As a first try, you may use the default parameters (but limit the search to the UniProtSP/SwissProt to shorten the search time).

2. Look at your results. What is the most likely identification for the spectrum according to Mascot? Would you be confident with this result?
3. What does the field "Expect" mean?
4. What does the "Protein scores greater than XX are significant ($p<0.05$)." mean?

5. Give some possible reasons for non-confident matches.

6. Click on the identifier of the first scoring protein, to open up a detailed results window. You will see the matched peptides as well as observed mass errors for the matched peaks. What can you conclude from the observed mass errors?

Take a look at the parameters. Try to refine the settings. In fact, the sample has been alkylated using iodoacetamide, which modifies the mass of the cysteins. It is called "carbamidomethylation" and occurs on almost all cysteine residues (fixed modification then). The methionines may also have been oxidised (variable modification then). Set the parameter accordingly and run Mascot again.

7. What can you say about the results? Is the result correct?

8. Is the number of matched peptides high or low?

9. Now, return to the submission page and change the error from 1 Da to 400 ppm. By doing so, do you increase or decrease the error tolerance?

Run Mascot again. This time, the match is clearly significant. If you take a look at the error plot, the error range is small and constant, and the score and Expect values are good.

MS/MS data analysis using Peptide Fragmentation Fingerprinting

Before you start:

1. What are the advantages of PFF identification over PMF?
2. Enumerate possible reasons for non-significant identifications.
3. What is the difference between an a ion and a b ion?

The MS/MS tools that we are going to use are here:

X!Tandem http://www.thegpm.org/

Mascot http://www.matrixscience.com/

A: A step by step PFF submission

Use Mascot to identify the peptides in the file ppw_A23_MSMS.mgf that contains 15 PFF spectra. Select SwissProt as a database, set the modifications as for PMF (cystein carbamidomethylation, possible oxidations on methionine), choose 2 Da for the precursor mass filter, 0.4 Da for the fragment error and mgf for the file format.

1. Which Mascot parameters correspond to the precursor mass error tolerance and the fragment mass error tolerance?
2. Run Mascot. Which protein(s) have been identified? Please comment.
3. Visualise the quality of the spectrum-to-peptide assignment for some of the spectra with low and high score (to do this, click on the spectrum (query) numbers).
4. Which main differences do you see between a confident peptide hit and a non-confident one?

Perform the same search with X!Tandem (using the Cow – Bos taurus proteome). Does these pieces of software identify the same peptides and proteins?

B: Comparing different instruments and runs

1D-SDS gel was performed on a *Staphyloccocus aureus* (Firmicutes) sample; proteins are alkylated with iodoacetamide. Gel bands were submitted to trypsin digestion and digests were measured on both a MALDI TOF-TOF (ABI 4700) and a QqTOF (Micromass QTOF). The files are TFE_pellet...; file format for esi is sequest dta, file format for MALDI is mascot mgf)

We will use X!Tandem to explore these sets. Fill in the submission forms of X!Tandem and try to identify the proteins. Typical mass errors for the TOF/TOF instrument are 100 ppm for precursors and 0.1 Da for fragments. For the QTOF 0.3 Da for both.

1. Can you detect peptides that are specific to MALDI or ESI?

C: Exploring big data sets

Staphylococcus aureus (Firmicutes) sample has been submitted to Iodoacetamide and migrated on 1-D SDS gel. A band has been excised and processed with trypsin. The band was then analysed by LC/MS-MS, resulting in the file N2.mgf. The file contains 1780 MS/MS spectra of very good quality (Orbitrap). It is still a "small" set, since it is current to work with sets of thousands of spectra (and it is usual to confidently identify only a small part of them...). We will use X!Tandem.

Use the following parameters:

* Orbitrap
* Cysteine carbamidomethylation (as fixed modification)
* Oxidation of M, H and W (as variable)
* Parent tolerance 10 ppm, ion tolerance 0.5 Da
* SWISS-PROT database

How many proteins are confidently identified?