Peptide mass fingerprinting exercise

In the course of a human platelet proteomic project, a protein band was excised from a one-dimensional electrophoresis gel at the top of the gel (protein markers indicate that the protein is larger than 220'000 Da. The protein was excised from the gel, reduced with DTT and alkylated with iodoacetamide, digested with porcine trypsin, peptides were extracted from the gel piece and spotted on a MALDI plate with CHCA matrix (α-Cyano-4-hydroxycinnamic acid). Peptides were analysed on a MALDI-Q_TOF instrument to identify the protein.

The mass list obtained from the MS spectrum is appended in the file Prot_spot.txt.

**First identification**

1. Go the the [Aldente web page](#). Copy/paste the mass list in the peak list box.
2. Choose UniProt/SwissProt as the database
3. Choose homo sapiens as the taxon.
4. Extend the search mass range to be sure to include the protein (higher than 220'000 Da).
5. Leave the pI range 0-14.
6. Untick Varsplices (search for known mutations) and Fragments (search for known proteolytic fragments).
7. Check the following options: Trypsin, 1 miscleavage, monoisotopic resolution, [M+H] positive ion mode. Uncheck the PTM (post-translation modifications).
8. Choose the appropriate modification corresponding to iodoacetamide alkylation (as fixed modification). Add methionine oxidation (MSO) as variable modification. Leave the Max and Score parameters unchanged.
9. In the Threshold section, increase the Spectrometer shift max to 0.5 Da, and the Spectrometer internal error max to 50 ppm.
10. Launch the search.

There should be only one significant identification (score highlighted in green). Report the identification indicators (score, number of peptide identified, sequence coverage, average shift in mass, slope) together with the identification parameters.
Second round of identification

When you inspect the details of the identification for the first match, you see the list of matched peptides. Some of them are marked with grey or red asterisks. Search the help section for the signification of these asterisks and explain in detail what they mean.

If you go to the bottom of the first identification details, there are links to other tools. Click on the FindPept tool on the recalibratd spectrum tool. Aldente is able to recalibrate the data (compensate for experimental deviations between the theoretical and observed masses of the peptides). **Briefly explain the theory of the recalibration procedure.** The Findpept tool allows you to search for explanation of unidentified peaks in your spectrum (trypsin autoproteolysis, known classical contaminants, unspecific cleavages).

On the Findpept page, check that cysteines are modified with iodoacetamide, tick the methionine oxidation modification, tick [M+H] positive ion mode, and monoisotopic resolution. In the enzyme section, select porcine trypsin, and tick all three boxes (specific cleavage by the enzyme, autolytic cleavage of the enzyme, specific cleavage of human keratins). Launch the search.

In the result page of FindPept, you first find the sequence of the protein identified by Aldente, then the Findpept search parameters, and the additionally identified peaks (first from keratins, then from trypsin autolysis, then peptides from the identified protein cleaved specifically by trypsin, then peptides from the identified protein cleaved unspecifically). Manually remove from the original peak list the masses corresponding to keratins and trypsin autolysis.

Search the internet for the masses of CHCA matrix clusters (the matrix used to perform MALDI ionization). Manually remove these masses from the peak list.

Submit the curated peak list to Aldente again using the same parameters. Report the identification results (score, sequence coverage, number of hits, average shift in mass, slope).

Third round of identification

In the previous round, you have removed all contaminant peaks, which should have increased the identification score. Then you will try to optimize the search parameters in Aldente: in the threshold tab, modify the spectrometer internal error (range 10-200 ppm) and report the identification results. You should realize that there is a lower limit below which the number of identified peptides diminishes. Do this lower limit corresponds to the known mass accuracy of a MALDI-Q-TOF instrument (provide background information for this parameter)?

You have now defined an optimal mass accuracy (spectrometer internal error max) that corresponds to the mass accuracy of the mass spectrometer in this
experiment that is a compromise between search stringency and realistic MS accuracy.

**Fourth round of identification**

Using the optimal spectrometer internal max error, go to the peptide tab and tick the PTM box to allow the search for post-translation modifications in your peak list. Launch the search, and report the additional peptide identifications with post-translational modifications.

By clicking on the protein accession number in the Aldente result page, you will access the Uniprot protein database summarising all known information about the identified protein. Check if the additional identification of PTMs correspond to known PTMs. Report the final identification results (score, sequence coverage, number of peptide identified, identified PTMs....)