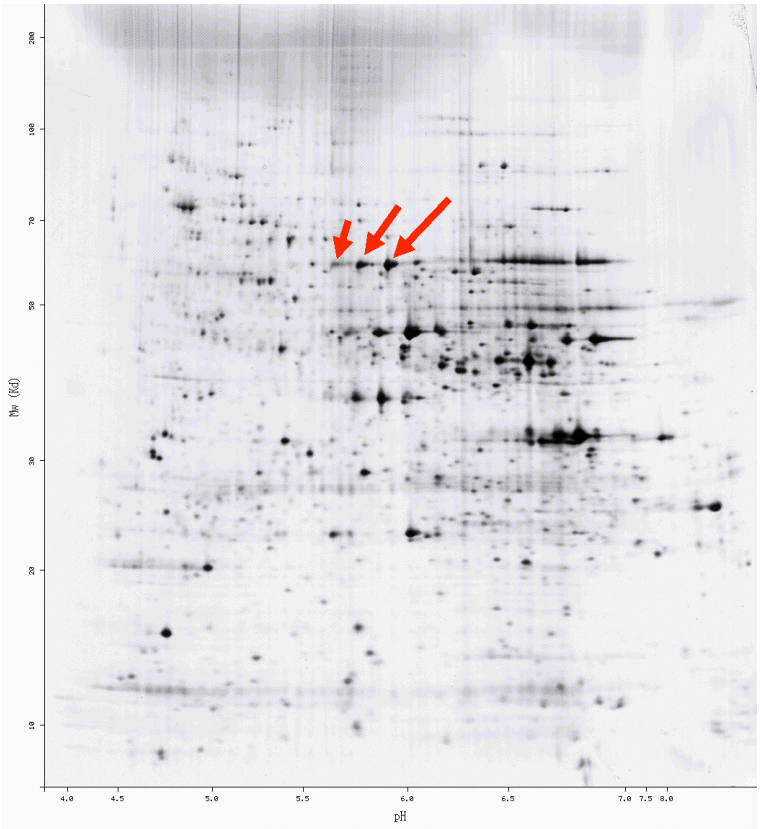
Exercise 1: Protein identification from two-dimensional gel electrophoresis and peptide mass fingerprinting

The figure below shows a two-dimensional electrophoresis map of the proteome of Saccharomyces Cerevisiae:



Question 1: Give a short description (max 20 lines) of Saccharomyces Cerevisiae, what it is used for in industry and in biological research.

Question 2: Using the Sequence Retrieval System, indicate how many proteins from Saccharomyces Cerevisae have been documented to date in SwissProt? In TrEMBL? Explain the difference between the two databases.

We want to identify the three spots highlighted in the gel image. The third spot on the right has the following properties measured from the gel:
Molecular weight: 58400 Da
Isoelectric point: 5.9

Question 3: From the gel image above, estimate the uncertainty on molecular weight and isoelectric point. Using the tool TagIdent from Expasy, give the number of possible proteins from SwissProt and/or TrEMBL with these molecular weight and isoelectric

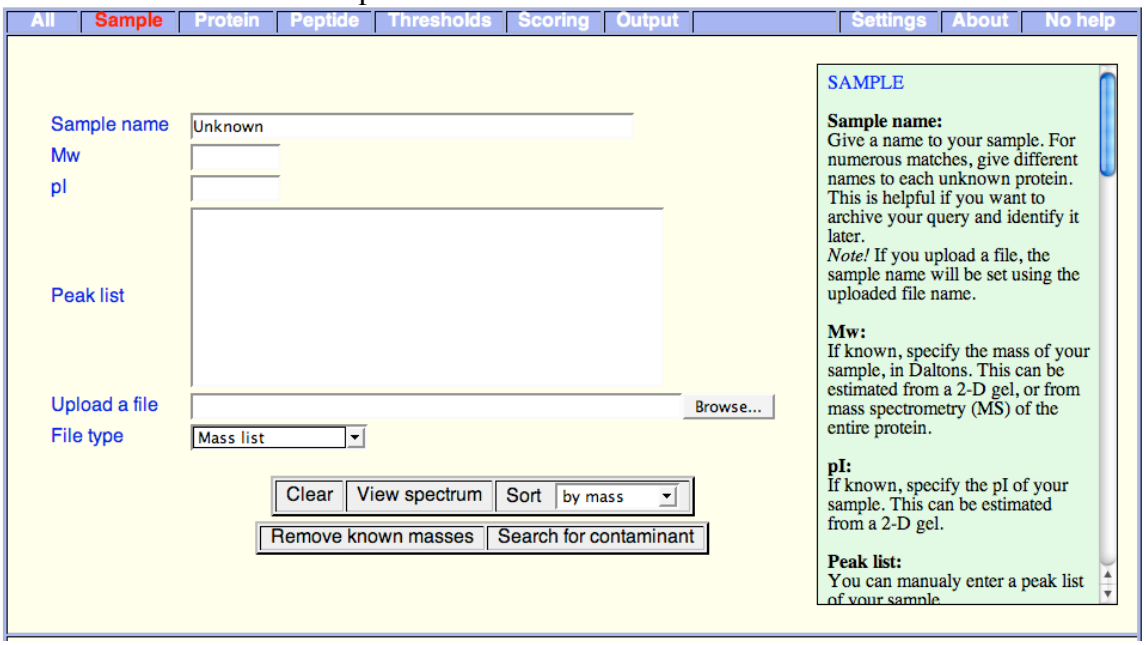
point, depending on the uncertainties you estimated. Report on the influence of measurement uncertainties on the number of candidate proteins.

Question 4: from what you obtained on question 3, it should be clear that the protein identity cannot be deduced from the 2D-GE image alone. Suggest possible techniques (including mass spectrometry or not) to confirm or establish the identity of the three proteins of interest (there are at least three classical techniques).

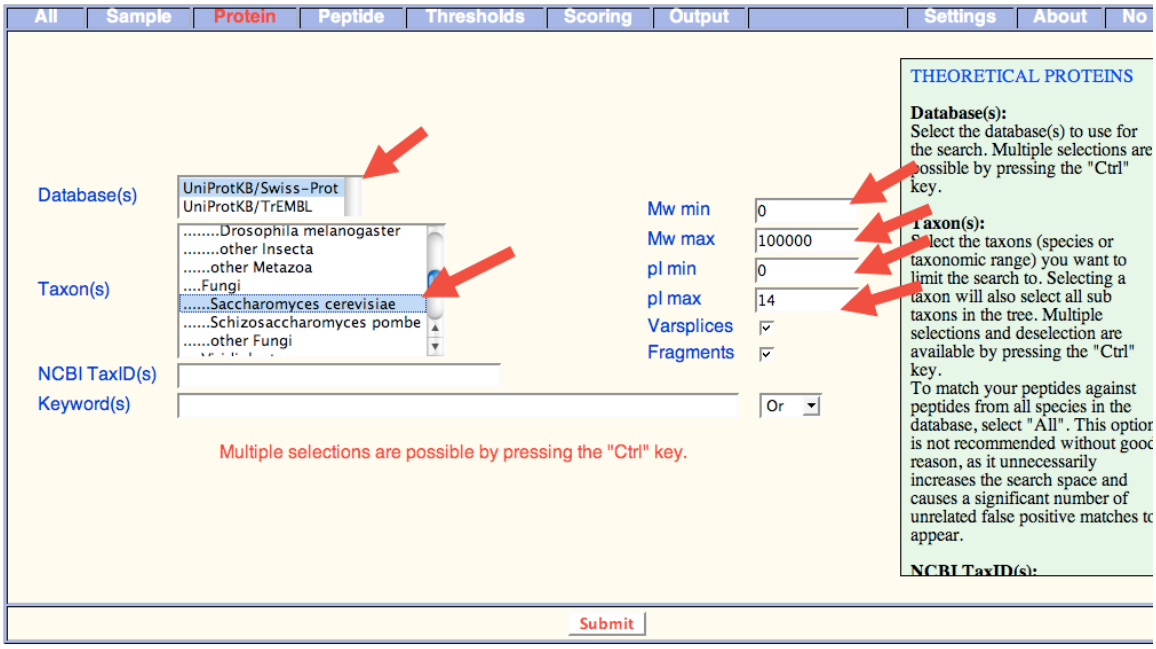
The three protein spots are excised from the gel, and incubated with trypsin. The resulting peptides are eluted from the gel piece, desalted and submitted to MALDI-TOF mass spectrometry. The list of measured masses $(M+H)^+$ are listed in the file 'spot right.txt' for the first spot on the right.

Question 5: Use Aldente to perform the protein identification. Follow the following steps:

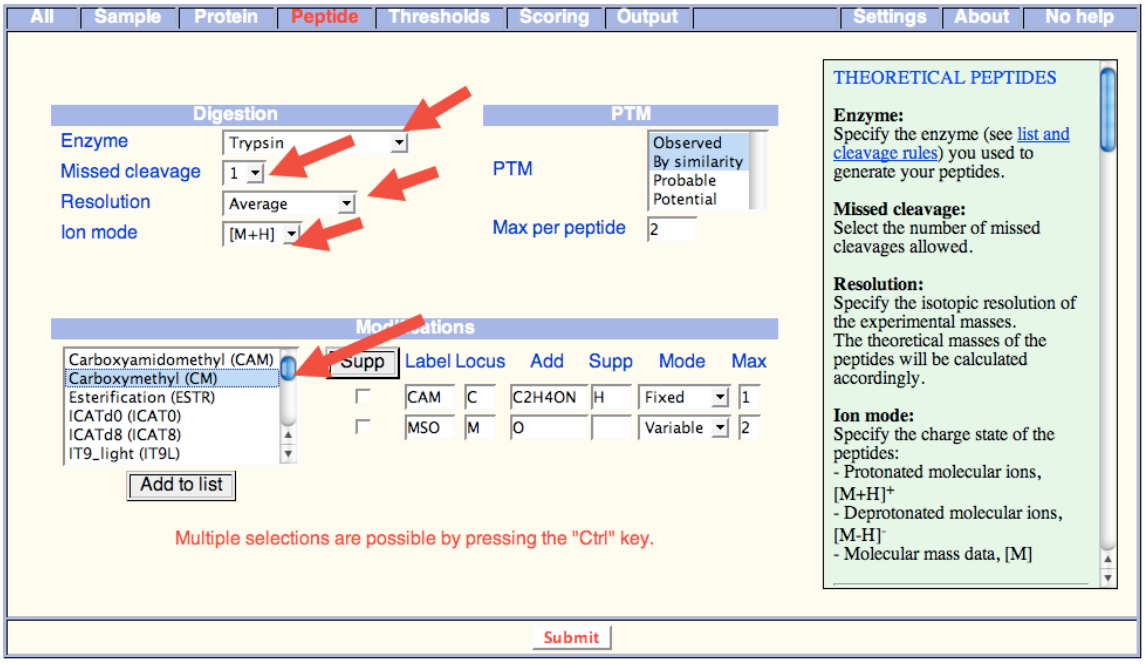
1. Go to the 'Sample' tab. Enter the protein molecular weight and isoelectric point. Copy/paste the peptide mass list from the 'spot right.txt'file (or alternatively load the file and tick the 'peak list' file format.



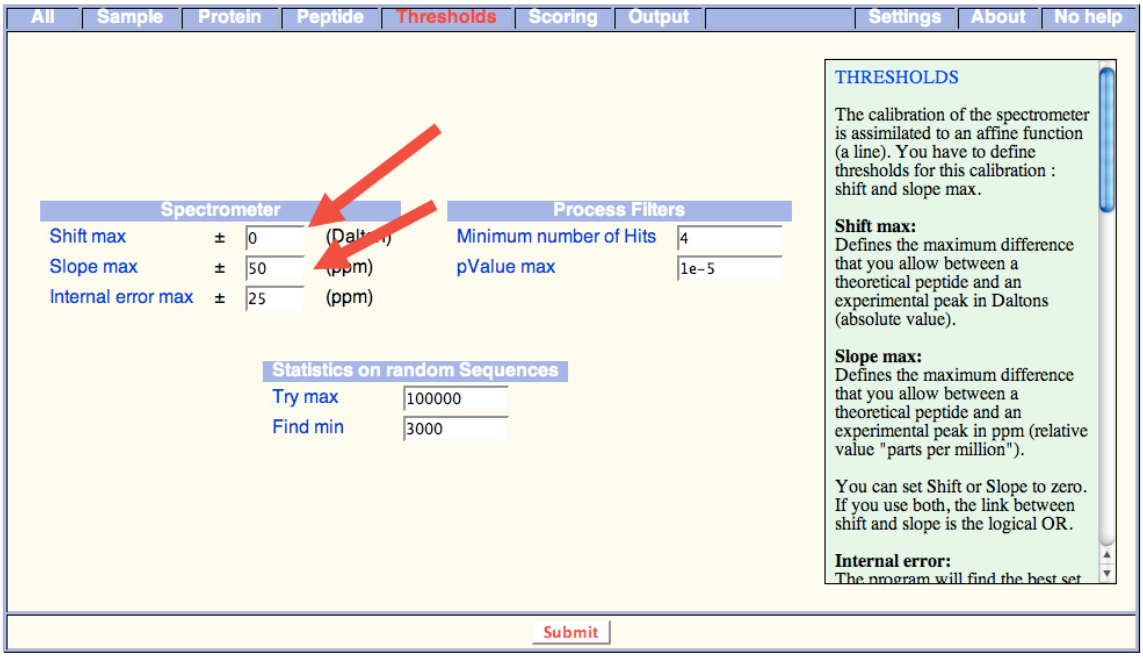
2. Go to the 'Protein' tab. Select SwissProt as the search database, Saccharomyces Cerevisae as the specie under study, and fill in the molecular weight and isoelectric point limits.



3. In the 'Peptide' tab, choose trypsin as protease, allow 1 miscleavage, choose 'average' as resolution, and [M+H] as ion polarity. Select Carboxymethlyation of cysteines as a permanent modification. Leave other fields unchanged.



4. In the Thresholds tab, enter a shift max (absolute mass accuracy) of 0, a slope max (relative mass accuracy in ppm) of 50. Leave other parameters unchanged.



5. Submit the identification. Report the result in terms of number of peptide matches, sequence coverage, and pValue.

In a second stage, change sequentially the following parameters and report the effect on the identification:

- On the Protein tab, Mw min, Mw max, pI min and pI max.
- On the Thresholds tab, change the 'Slope max' (relative mass accuracy in ppm)

Question 6: When the same operation is performed on the two other spots highlighted on the gel image, the very same protein is identified (though with less sequence coverage). Could you suggest why the same protein is identified from three different spots?