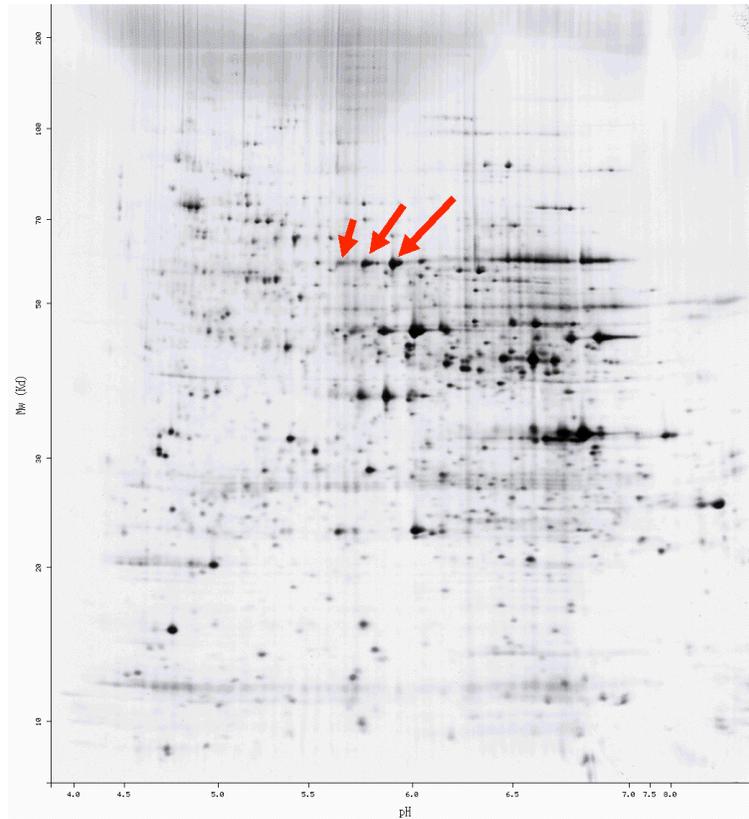


Exercise 1: Protein identification from two-dimensional gel electrophoresis and peptide mass fingerprinting

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 Cerevisiae* 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.

The screenshot shows the Aldente software interface. The 'Sample' tab is active. The left sidebar has the following elements: 'Sample name' (text input with 'Unknown'), 'Mw' (text input), 'pI' (text input), 'Peak list' (text area), 'Upload a file' (text input with 'Browse...' button), and 'File type' (dropdown menu with 'Mass list' selected). Below these are buttons for 'Clear', 'View spectrum', 'Sort' (dropdown with 'by mass'), 'Remove known masses', and 'Search for contaminant'. The right panel, titled 'SAMPLE', contains the following text: 'Sample name: Give a name to your sample. For numerous matches, give different names to each unknown protein. This is helpful if you want to archive your query and identify it later. Note! If you upload a file, the sample name will be set using the uploaded file name.' 'Mw: If known, specify the mass of your sample, in Daltons. This can be estimated from a 2-D gel, or from mass spectrometry (MS) of the entire protein.' 'pI: If known, specify the pI of your sample. This can be estimated from a 2-D gel.' 'Peak list: You can manually enter a peak list of your sample.'

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

THEORETICAL PROTEINS

Database(s):
Select the database(s) to use for the search. Multiple selections are possible by pressing the "Ctrl" key.

Taxon(s):
Select the taxons (species or taxonomic range) you want to limit the search to. Selecting a taxon will also select all sub taxons in the tree. Multiple selections and deselection are available by pressing the "Ctrl" key.

To match your peptides against peptides from all species in the database, select "All". This option is not recommended without good reason, as it unnecessarily increases the search space and causes a significant number of unrelated false positive matches to appear.

NCBI TaxID(s):

Multiple selections are possible by pressing the "Ctrl" key.

Submit

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

THEORETICAL PEPTIDES

Enzyme:
Specify the enzyme (see [list and cleavage rules](#)) you used to generate your peptides.

Missed cleavage:
Select the number of missed cleavages allowed.

Resolution:
Specify the isotopic resolution of the experimental masses. The theoretical masses of the peptides will be calculated accordingly.

Ion mode:
Specify the charge state of the peptides:
- Protonated molecular ions, [M+H]⁺
- Deprotonated molecular ions, [M-H]⁻
- Molecular mass data, [M]

Multiple selections are possible by pressing the "Ctrl" key.

Submit

- 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.

All	Sample	Protein	Peptide	Thresholds	Scoring	Output	Settings	About	No help																
							THRESHOLDS																		
							The calibration of the spectrometer is assimilated to an affine function (a line). You have to define thresholds for this calibration : shift and slope max.																		
							Shift max: Defines the maximum difference that you allow between a theoretical peptide and an experimental peak in Daltons (absolute value).																		
							Slope max: Defines the maximum difference that you allow between a theoretical peptide and an experimental peak in ppm (relative value "parts per million").																		
							You can set Shift or Slope to zero. If you use both, the link between shift and slope is the logical OR.																		
							Internal error: The program will find the best set																		
<table border="1"> <thead> <tr> <th colspan="2">Spectrometer</th> <th colspan="2">Process Filters</th> </tr> </thead> <tbody> <tr> <td>Shift max</td> <td>± 0 (Dalton)</td> <td>Minimum number of Hits</td> <td>4</td> </tr> <tr> <td>Slope max</td> <td>± 50 (ppm)</td> <td>pValue max</td> <td>1e-5</td> </tr> <tr> <td>Internal error max</td> <td>± 25 (ppm)</td> <td></td> <td></td> </tr> </tbody> </table>							Spectrometer		Process Filters		Shift max	± 0 (Dalton)	Minimum number of Hits	4	Slope max	± 50 (ppm)	pValue max	1e-5	Internal error max	± 25 (ppm)					
Spectrometer		Process Filters																							
Shift max	± 0 (Dalton)	Minimum number of Hits	4																						
Slope max	± 50 (ppm)	pValue max	1e-5																						
Internal error max	± 25 (ppm)																								
<table border="1"> <thead> <tr> <th colspan="2">Statistics on random Sequences</th> </tr> </thead> <tbody> <tr> <td>Try max</td> <td>100000</td> </tr> <tr> <td>Find min</td> <td>3000</td> </tr> </tbody> </table>							Statistics on random Sequences		Try max	100000	Find min	3000													
Statistics on random Sequences																									
Try max	100000																								
Find min	3000																								
<input type="button" value="Submit"/>																									

- 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?