Introduction:

Exam: present a poster (that is the first form of presentation of data) in January (choose an original paper), present it as it's your research. Highlight very brief materials.

The exam is the last two days of the course.

15 min to present it to the class, then there is a question of curiosity about the poster and a question about the course.

Glossary:

Physiome: description of its functional behavior.

Metabolome: complete set of small-molecule metabolites (such as metabolic intermediates, hormones and other signaling molecules).

Phenome: set of all phenotypes expressed by a cell, tissue, organ, organism, or species.

Morphome: to map and classify all the morphological features of species

Glycome: the entire complement of sugars, whether free or present in more complex molecules

Ribonome: total cellular complement of RNAs and their regulatory factors functioning dynamically in time and space

Orfeome: he totality of open reading frames (ORF) in biology

Regulome: whole set of regulation components in a cell (those components can be genes, mRNAs, proteins, and metabolites)

Cellome: whole set of biological entities within cells and their interactions in the cell, and the totality of biological cells

Operome: promoter map

Transportome*:* membrane transporters and channels governing cellular influx and efflux of ions, nutrients, and drugs

Functome: claimed as the next step of metabolome and regulome. It represents biological functions rather than chemical of cells (only 1 article in PubMed)

PROTEOMICS:

Most of proteomics deals with **personalized medicine**, achieved integrating the different omics. Personalized medicine is owned principally by startups and little companies, while the so called big-pharma wants something that fits with the highest number of people simultaneously. Different types of omics are integrated and analyzed by AI algorithms to extract patient-specific information.

New drugs are discovered most of the times by startups because they need to fit for a smaller population. The system biology group in Seattle coined the term **P4 medicine**:

- Predictive
- Preventive
- Personalized
- Participatory: the patient is at the core of the research, in order to understand the real needs

Proteomics can be divided in three main branches:

(proteome = proteins encoded by the genome)

- The first part is the historical more investigated (50 years and more): the **structural proteomics,** it has the goal of mapping the 3D structure of protein and proteins complex;
- With the structural proteomics we don't have enough information and we need **functional proteomics** to study protein-protein interaction, 3D structures cellular localization and PTMS in order to understand the physiological function of the whole set of proteomes.
- **expression proteomics** is the quantitative study of protein expression between samples that differ by some variables.

We have 3 principal approaches of proteomics:

- **top-down:** top is the full protein and bottom are the peptides. We consider the intact protein (we don't digest it) and we use tandem mass spectrometry.

- **bottom-up:** we need to perform the separation of proteins (most of the times with 2D electrophoresis) in order to obtain a single protein to digest (most of the times with trypsin) and obtain peptides on which we perform the mass spectrometry. This approach is the standard workflow for peptide mass fingerprinting (potential question at the exam)

- **shotgun**: it is the most powerful but for this reason requires the most sophisticated instrumentation. It starts from a complex mixture, after an enzymatic digestion we'll obtain peptides coming from different proteins. We perform a separation with liquid chromatography and then do a bidimensional electrophoresis.

These kinds of experiments are so complex and can easily fail without knowing the cause: therefore, we need to use standard experimental design techniques, incorporate quality control and perform statistical analysis in order to increase the chances of success.

Complexity of Proteomics

The study of structure and function of a protein is complicated due to the **intrinsic variability of the proteome.** Variability that occurs from cell to cell but also through biochemical interactions with the genome and the environment (that particularly affects the expression). While genome is a relatively constant entity, for the proteome we have to consider both the spatial and temporal connotations.

The identification of the whole number of proteins in a cell is impossible: most of these proteins are expressed at relatively low levels ($10–10^2$ per cell), although some are expressed at much higher levels ($10^4–10^6$ per cell, for example albumin).

Regardless of the absolute level of expression, most proteins exist in multiple **post-translationally modified** (PTMs) forms. Moreover, there is also **splicing**, **protein-protein interactions** (PPIs) and **subcellular localization:** we must find ways to detect a large number of distinct molecular species.

The complexity increases if we consider the problem of amplification and sensitivity to damage and degradation.

Sequence Alignment:

It is a way of arranging primary sequences (of DNA, RNA, or proteins) in such a way as to align areas sharing common properties. The degree of relatedness, similarity between the sequences is predicted computationally or statistically

- ClustalW
- Benchling
- Blast

• FASTA SEQUENCE

Begins with a single-line description followed by lines of sequence data

(SEQUENCES ARE EXPECTED TO BE REPRESENTED IN THE STANDARD IUB/IUPAC AMINO ACID AND NUCLEIC ACID CODES, WITH THESE EXCEPTIONS: LOWER-CASE LETTERS ARE ACCEPTED AND ARE MAPPED INTO UPPER-CASE; A SINGLE HYPHEN OR DASH CAN BE USED TO REPRESENT A GAP OF INDETERMINATE LENGTH; AND IN AMINO ACID SEQUENCES, U AND * ARE ACCEPTABLE LETTERS (SEE BELOW). BEFORE SUBMITTING A REQUEST, ANY NUMERICAL DIGIT IN THE QUERY SEQUENCE SHOULD EITHER BE REMOVED OR REPLACED BY APPROPRIATE LETTER CODES (E.G., N FOR UNKNOWN NUCLEIC ACID RESIDUE OR X FOR UNKNOWN AMINO ACID RESIDUE).

• BARE SEQUENCE

Just sequence of data. It can be interspersed with numbers and/or spaces, such as the sequence portion of a GenBank/GenPept flatfile report.

Functional Families:

(Proteins that carry out related functions)

Principal protein family classification databases:

- PROSITE: database of domains families and functional sites, one of the most used
- SMART

In this way, threading tries to **predict the three-dimensional structure** starting from **a given protein sequence**. Useful when comparisons based on sequences or sequence profiles alone fail to a too low similarity. (The complication comes from the interactions, the PTMs, differences of the proteins)

Proteomics aims and applications

- Protein sample identification/confirmation
- Protein sample purity determination
- Detection of post-translational modifications
- Monitoring protein-ligand complexes/structure
- De novo peptide sequencing
- Biomarker discovery by detection of differences in protein expression among different classes of sample
- Identification of new targets for drugs
- Detection of amino acid substitutions
- Mass fingerprint identification of proteins
- Monitoring protein folding

All of them need biochemical, bioanalytical, biomolecular and bioinformatic knowledges

Workflow

We need to simplify samples as much as possible, then after performing a mass spectrometry we can identify proteins exploiting and matching the sequence with our databases.

The aim is to identify the proteins present in our sample and the workflow change depending on the method we need to use.

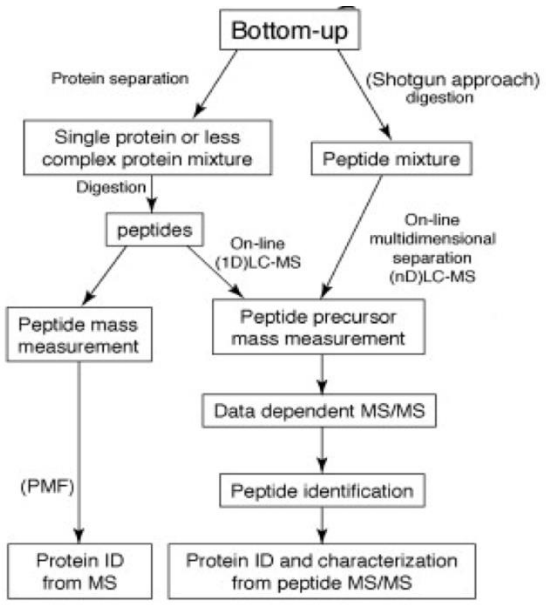
BOTTOM-UP APPROACH:

spezzetto le proteine tramite digestion, prima della massa

Complex mixture of proteins gets separated before enzymatic (or chemical) digestion--> direct peptide mass fingerprinting-based acquisition or further peptide separation on-line coupled to tandem mass spectrometry.

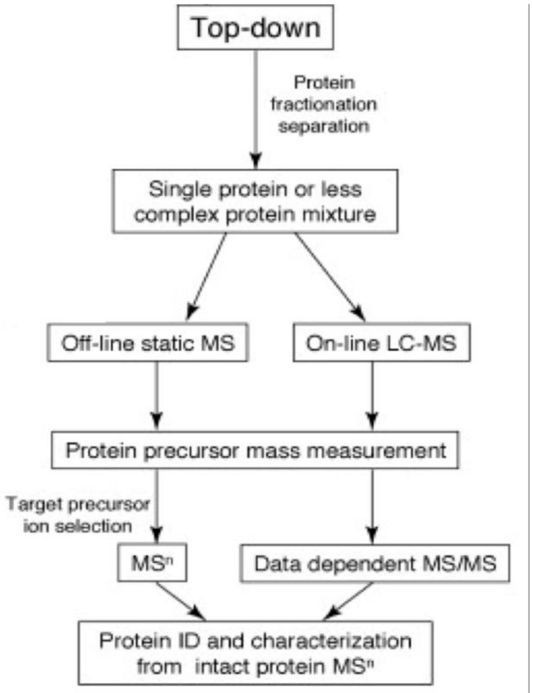
Alternatively, we have a direct digestion into a collection of peptides (**shotgun** approach), which are then separated by multidimensional chromatography on-line coupled to tandem mass spectrometric analysis.

This approach may cause the loss of some post-translational modifications.



TOP-DOWN APPROACH: non spezzetto le proteine prima della massa e posso rivelare PTMs

Complex mixtures are fractionated and separated into pure single proteins or less complex mixtures--> off-line static infusion of the samples into the mass spectrometer for intact protein mass measurement and intact protein fragmentation. An on-line LC–MS (liquid chromatography-mass spectrometry) strategy can also be used for large-scale protein interrogation.


Top-down
Protein fractionation separation
Single protein or less complex protein mixture
Off-line static MS
On-line LC-MS
Protein precursor mass measurement
Target precursor ion selection
MS^n
Data dependent MS/MS
Protein ID and characterization from intact protein MS^n

Current Opinion in Chemical Biology

SEPARATION TECHNIQUES:

For intact proteins:

- 1D- and 2D-SDS PAGE
- Preparative IEF isoelectric focusing
- HPLC (*High Pressure and performance Liquid Chromatography)*

For peptides:

- MS-MS (tandem mass spectrometry)
- Multidimensional HPLC (MudPIT)
- SELDI (Surface-enhanced laser desorption/ionization)

Then there are also other additional ways to obtain the less complex structure at the end such as difference gel electrophoresis (DIGE) and Isotope-coded affinity tagging (ICAT).

Separation techniques depend on the different **properties that you choose** in order to separate

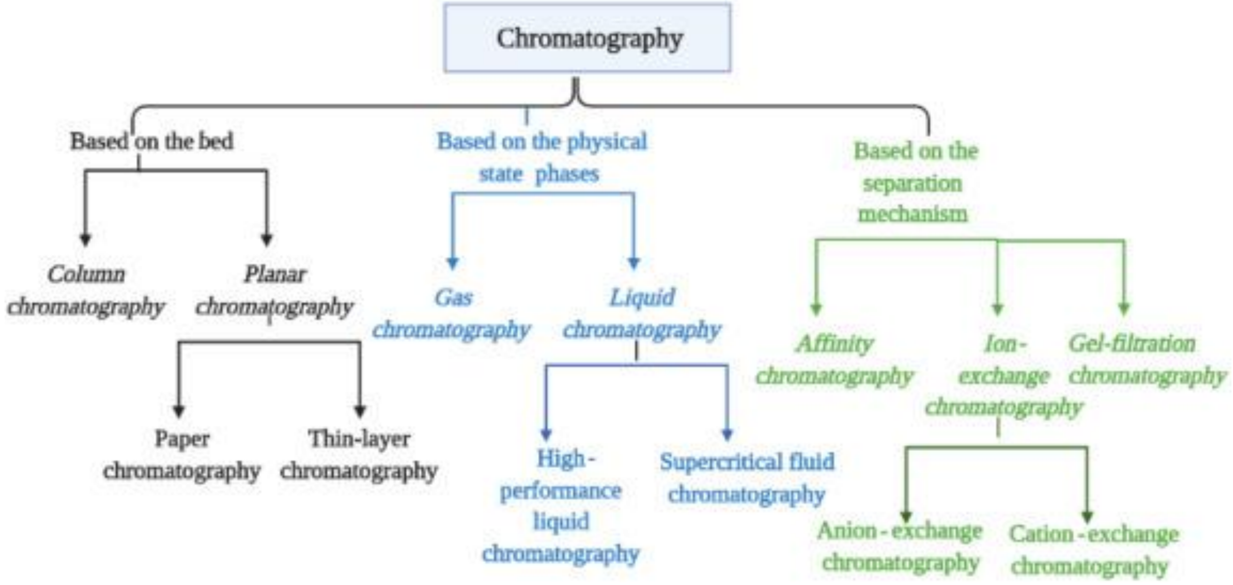
- By size:
  1. gel electrophoresis
  2. gel filtration chromatography
  3. ultracentrifugation
  4. dialysis

- By charge:
  1. isoelectric focusing
  2. ion exchange chromatography

- By polarity:
  1. paper and reverse-phase chromatography
  2. hydrophobic interaction chromatography

- By specificity:
  1. affinity chromatography
  2. Immuno-precipitation assay

Chromatography:


Chromatography
Based on the bed
Column chromatography
Planar chromatography
Paper chromatography
Thin-layer chromatography
Based on the physical state phases
Gas chromatography
Liquid chromatography
High-performance liquid chromatography
Supercritical fluid chromatography
Based on the separation mechanism
Affinity chromatography
Ion-exchange chromatography
Gel-filtration chromatography
Anion-exchange chromatography
Cation-exchange chromatography

The sample gets separated in different fractions depending on:

- Size of proteins: bigger proteins elute first (they cannot enter the beads)
- Presence of specific functional groups, so on the affinity (principle for the 6His tag)
- Basis of charge (considering the isoelectric point of the protein)
- Basis of net charge

Generally, we deal with very low amounts of protein, and in order to detect low-abundance proteins we can:

- Enrich from larger volumes, increase the concentration of our sample (only if we have very powerful tools to separate my protein from the others, like selective precipitation or centrifugation, preparative approaches)
- Combination of 2DE with LC;
- Multidimensional LC

Eg. Microdialysis is quite easy to perform (Diffusion through a semi-permeable cellulose membrane. Different pore sizes allow removal of molecules smaller than specific MW)

We can also remove specific proteins eg. Albumin (present in large amount): very useful kits available based on the use of columns with packed antibodies that recognize albumin.

Electrophoresis:

• Filter paper: proteins are easily denatured due to the high absorbance of filter paper. Works for small peptides or amino acids.

• Thin layer (TLC): chemically modified cellulose

• gel: starch, agarose, polyacrylamide (PAGE)

PAGE:

•**Native**:

- Enzyme activities are retained after electrophoresis, so enzymatic assays can be performed on separated proteins

- Factors affecting mobility: charge; molecular weight and shape of proteins

•SDS:

- SDS (sodium dodecyl sulfate) coats the surface of proteins with a balanced negative charge, in order to let run also the positively-charged proteins.

- Proteins are denatured, so enzyme activities are lost after SDS-PAGE.

- Factors affecting mobility: molecular weight

Immunological techniques:

Polyclonal and monoclonal antibodies have been developed to almost all known human proteins. Immunological techniques are based on the specificity of antibody-antigen interaction.

- Immunoprecipitation assay (IP)
- Immunoblotting (Western blotting)
- ELISA (Enzyme-Linked ImmunoSorbent Assays)

IP (Immunoprecipitation assay):

Procedure that permits the purification of the protein of interest with the use of specific antibody (poly- or monoclonal)

Applied to:

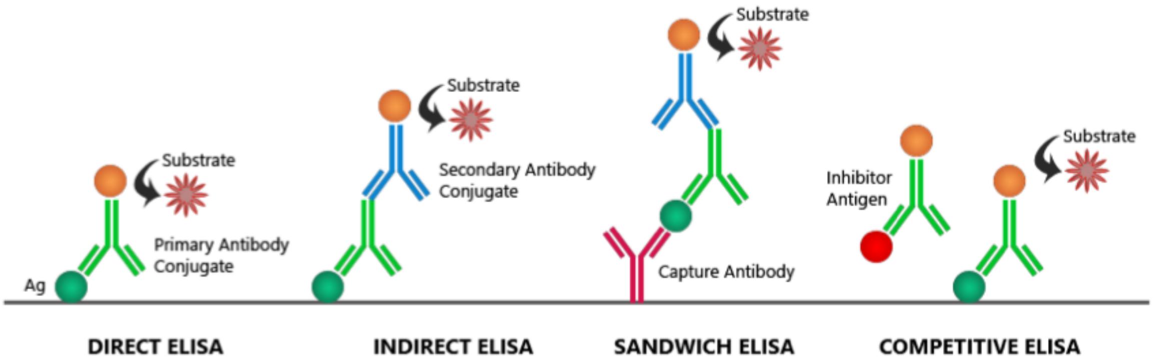
- analyze the activity of immunoprecipitated proteins
- prove the interactions between two proteins
- isolate multienzyme complexes and to identify their components
- analyze protein-DNA interactions (chromatin immunoprecipitation assay)

ELISA:

This technique combines the specificity of the antibody-antigen interaction with the sensitivity of enzyme assays using either an antibody or an antigen conjugated to an enzyme.

The activity of the enzyme is measured by adding an appropriate chromogenic substrate, which is converted to a colored product.

Automation of the assay is straightforward, so it is often used in drug discovery programs and, thanks to its high sensitivity, is often used in diagnostic kits.



CLASSICAL MS PROTEOMICS APPROACH:

•Sample solubilization with detergents (UREA 7M, TIOUREA 2M, CHAPS 4%)

•Bidimensional separation:

- 1st dimension: isoelectric focusing
- 2nd dimension (the second separation is orthogonal to the first one): SDS-PAGE

•Result visualization (immunoblotting, Coomassie blue or silver staining)

•From gel analysis we can try to achieve information about protein ID

•From mass spectrometry (MALDI) we will have more information

SAMPLE PREPARATION:

An efficient sample preparation should:

•Solubilize in a reproducible way all classes of proteins, including the hydrophobic ones

•Prevent protein aggregation and keep solubility during IEF

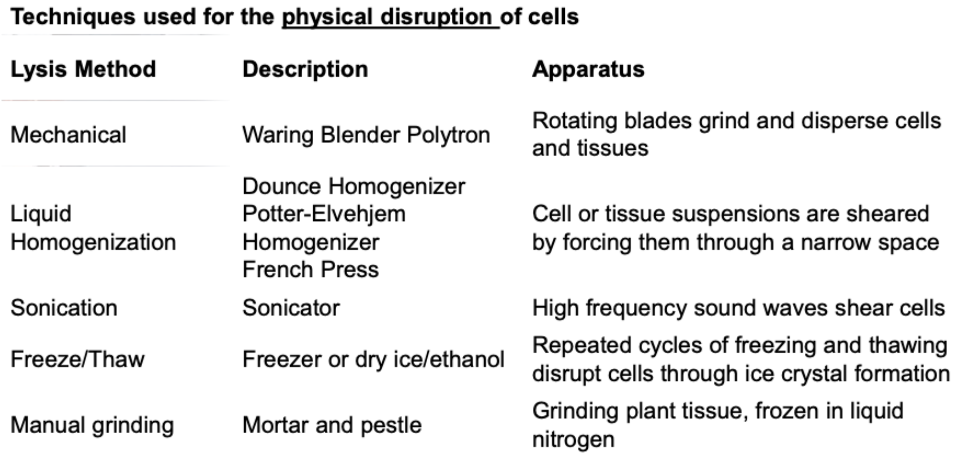
•Prevent chemical and enzymatic modifications during extraction process

•Digest or remove nucleic acids and other molecules that can interfere with the analysis

•Enrich target proteins (e.g., by eliminating most abundant proteins like albumin)

The first step (unless we are dealing secretomics) is to **break the cell membrane**.

Physical methods are by far the most used, they are less prone to artifacts because you don't introduce chemicals from outside, so for mass spectrometry they are a little less invasive, but are less reproducible.



Techniques used for the physical disruption of cells

Lysis Method	Description	Apparatus
Mechanical	Waring Blender Polytron	Rotating blades grind and disperse cells and tissues
Liquid Homogenization	Dounce Homogenizer Potter-Elvehjem Homogenizer French Press	Cell or tissue suspensions are sheared by forcing them through a narrow space
Sonication	Sonicator	High frequency sound waves shear cells
Freeze/Thaw	Freezer or dry ice/ethanol	Repeated cycles of freezing and thawing disrupt cells through ice crystal formation
Manual grinding	Mortar and pestle	Grinding plant tissue, frozen in liquid nitrogen

After breaking the membrane, we perform the **denaturation** that can be done with several denaturing agents (there are also commercial cocktails):

➢ Urea (denaturing agent): Chaotropic agents of choice for disrupting hydrogen bonds **Urea 8M** or mixtures of thiourea 2M and urea 5-8M.

➢ Thiol (reducing agent)

➢ Detergents

➢ Protease inhibitors

DETERGENTS

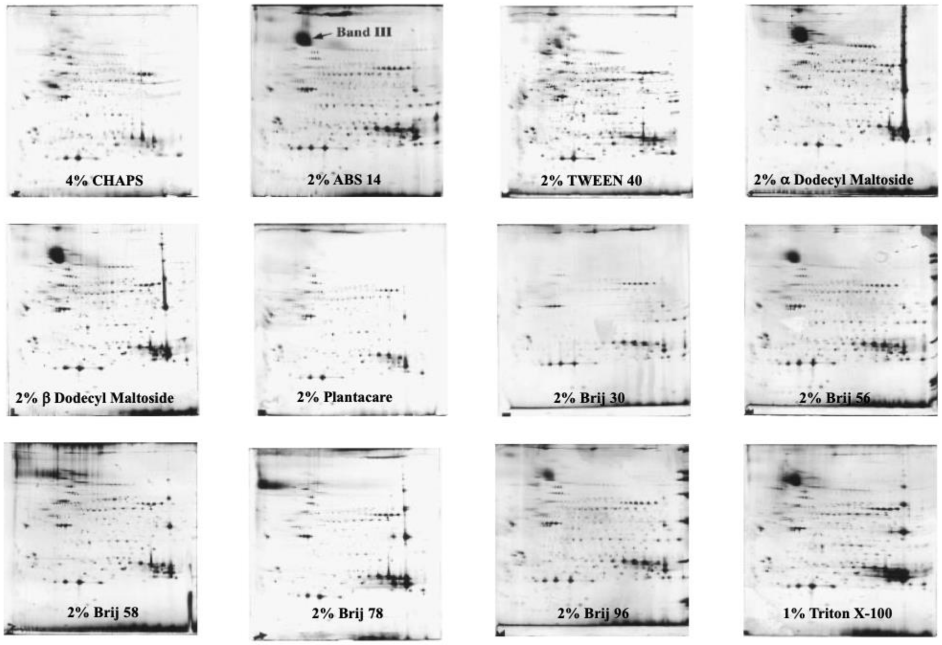
Detergents are a class of molecules whose unique properties enable manipulation (disruption or formation) of hydrophobic-hydrophilic interactions among molecules in biological samples. They are used to:

- lyse cells (release soluble proteins)
- solubilize membrane proteins and lipids
- control protein crystallization
- prevent non-specific binding in affinity purification and immunoassay procedures
- as additives in electrophoresis.

Non-ionic (uncharged) or zwitterionic (having both positively and negatively charged groups but with a net charge of zero) detergents must be used to allow protein migration according to its net charge (SDS cannot be used).

In order to break disulfide bonds (and analyze separate proteic subunits): dithiotreitol (DTT) or tributil-phosphine (TBP). A solution of DTT 50mM is enough for the majority of proteins, for difficult cases TBP must be used.

The problem of these procedures is that **depending on detergents and condition**, you'll have **different outputs**, different resolution in different spots--> in bidimensional electrophoresis we have a **low reproducibility of the results**.



2D Gel Electrophoresis

Now we need to perform the **separation**:

- Isoelectric point (IEF)
- Molecular weight (SDS PAGE)

These are two **orthogonal** separations, for every spot we have two c3oordinates from which we derive the info for IE point and MW.

First dimension separation:

IEF (Isoelectric focusing):

We need a pH gradient; then a voltage is applied and the proteins migrate. After the migration you'll have different sized protein with the same IE point.

Output--> strip with bands, in each there is probably more than one protein.

Second dimension separation:

We apply the strip on the SDS page for the **second dimensional** separation.

At the end we will have a map in which (hopefully) each spot represents a protein

STAIN:

It should have a high sensitivity and compatibility with mass spectrometry analysis.

Most used stains:

- Coomassie Blue R-250 (5-25 ng protein)

- Silver Stain Plus (0.25 ng protein) MOST SENSITIVE!

- Fluorescent molecules (Sypro RubyStain 0.25 ng protein)

The stain procedure needs to NOT INTERFERE with the next analysis.

After we obtain the map with spots:

We could skip MS analysis and try to identify protein from the gel using softwares.

E.g. Melanie found a match with previously identified spots (in previous gels). We import the gel (taking a picture) and modify some parameters like the contrast of the picture.

When we compare two gels and we notice a difference, there could be several reasons for the difference noticed (experiment conditions, another protein expressed, same protein but modified…). But it’s useful to combine different gels from the same sample deriving from same condition in a master gel. I will have a typical profile.