

Bioanalytical Proteomics and Interactomics

Saul Pierotti

November 18, 2019

Introduction

- Most proteomic studies are based on mass spectrometry
- The program will be finished in December, in January we will only do exercises
- The examination is composed in a poster presentation and an oral part
 - A poster is a self-sufficient work
 - Usually it is prepared at the beginning of a scientist's career
 - We should base our poster on selected papers
 - The poster is not mandatory, but if we do it we will only be asked 1 question in the exam
- Proteomics is one of the main omics, but also interactomics and secretomics are becoming really interesting
- The data produced by omics are on the order of zetta and yottabytes
- Personalized medicine is important, and now we have the tools to implement it
- Proteomics can be structural, functional and differential (expression)
- Three main approaches: top-down, bottom-up and shotgun
- Many funding sources now require researchers to share raw data
- In silico-cloning: we need to know at least one software
- Other useful softwares are Pymol, Graphpad Prism, ImageJ
 - Graphpad Prism is a spreadsheet used to elaborate data and do statistics
- Proteomics is the large scale study of proteins, the proteome is the total amount of proteins coded in the genome
- Proteomics involves a spatial and temporal dimension, while the genome is static
 - It is really noisy (!)
 - Proteins are really diverse, and their processing is complex
 - They are present in a really broad concentration range
 - There is no PCR for proteins (!)
- Unnatural amino acids are really interesting, and there are many patents about them
 - They are used frequently in cosmetics (!)
- Proteases are everywhere (!)
- Aromatic amino acid tend to be in β conformations
- MALEK tend to be in α helices
- 1/3 of drugs target GPCR receptors
- Membrane proteins are hard to study
- An antibody is around 150 kDa, it is a big protein
- Is of huge clinical interest to find fluorescent proteins that emit in the red spectrum
 - Most tissues absorb really well green light, and red is the least absorbed part of the spectrum
- Mutagenesis can be performed rationally or randomly
 - Random mutagenesis is made with kits that make an imprecise PCR
 - We can substitute any possible amino acid in a specific position that we know to be important
- The top-down approach starts from the protein
- The bottom-up approach uses peptides deriving from digestion of the protein
- The shotgun approach uses different peptides deriving from many proteins

- The less complex the protein mixture, the better (!)
- Informations on chemical compounds can be retrieved in PubChem
- PubMed does not indicize articles outside of biology and medicine (!)
 - Better to use Scopus or Web of Science for chemistry and other fields
 - For patents, Espacenet is the main database
- A volcano plot has the fold change in expression in the x axis and the significance in the y axis
 - It is used to rapidly identify differentially expressed proteins (it is also applicable to other fields)

Protein separation

- Proteins can be separated with different chromatographies or electrophoresis, microdialysis
- We can use depletion techniques to remove typical high-abundance proteins
 - There are kits to remove albumin with spin columns
- Protein electrophoresis can be done in native conformation or by SDS-PAGE
- Immunological techniques are Western blot, ELISA, immunoprecipitation
 - Antibodies have affinity constants ranging from 10^9 to 10^{12} M
- The upper limit for the number of resolvable proteins in a 2D-PAGE is 10000
- Spot identification in 2D-PAGE can be done by MS or by image analysis
 - Image analysis can be done by comparing our image with gel images of identified proteins present in databases
 - I usually work on master images created by combining 5-10 images of gels run in the same condition, to reduce noise
- Databases for 2D-PAGE, like ExPasy proteomic server, allow to search for images of specific tissues or cells, or conditions
 - I can then compare my spots with the ones already identified
- Differential PAGE (DIGE) runs 2 samples on the same gel, with different labels, so to be able to compare expression differences between them
 - It is a good practice to repeat the experiment by switching labels
 - Labels called cyanine derivatives (CyDyes) are available with many different emission spectra
 - * Labelling is done prior to IEF
 - * We can use many samples at the same time (multiplexing)
 - * They are really small and bind to the ϵ amino group of Lys in proteins
- The staining method for 2D-PAGE should be chosen considering the sensitivity required
- 2D-PAGE is generally speaking non-quantitative
- We can perform WB also on 2D-PAGE (!)
- HILIC (Hydrophilic interaction liquid chromatography) uses a polar stationary phase and a polar mobile phase
 - It is really useful for glycosylated proteins
 - Elution is done in water gradient
- N-terminal microsequencing is based on Edman degradation and can be used for spot identification, but it is obsolete
 - It has been replaced by MS
 - It uses phenyl-isothiocyanate that reacts with the N-termini forming thiazolinone derivatives
 - The derivative is then released and identified by chromatography or electrophoresis, before repeating the cycle
 - Amino acids are read at 269 nm

Mass spectrometry

- A mass spectrometer contains an ion source, a mass filter and a detector
- Some ion sources can be interfaced with a separation technique, others must be operated manually
- Soft ionization methods do not break chemical bonds, while hard methods do that
 - In proteomics we use soft ionization: MALDI or ESI

- There is also a semi-soft technique: fast atom bombardment (FAB)
 - * It is used for small proteins and peptides
 - * It uses typically a glycerol matrix
 - In ionization techniques, the matrix is used in order to prevent sample fragmentation
- In ESI the sample is in solution and it is sprayed by a small tube in a really strong electric field in presence of a hot nitrogen flow
 - The ionization is performed at atmospheric pressure
 - The sample forms droplets that evaporate concentrating charge, until they undergo coulombic explosion
 - The produced ions are multiple-charged, and this is useful for detecting big molecules
 - We can operate both in positive or negative ionization
 - * In positive mode I need to add formic acid to the solvent, in negative mode ammonia
 - * If I have acidic groups in my analyte I want to operate in negative modality, while with basic groups I want to operate in positive modality
 - It is possible to form protonated ions, deprotonated ions or cation adducts
 - * It is very sensitive to salts and detergents
 - Since I have multiple-charged ions, I have many peaks for each analyte
 - * With molecules under 1200 Da I tend to have single ions
 - * The multiple peaks are normally distributed and can be deconvoluted via software to derive the MW of the ion
 - Native and complexed proteins tend to have non-normal distribution of ions
 - It is very sensitive, but I need really pure samples
- In MALDI the solid sample is heated by a laser while embedded in the matrix
 - The matrix becomes ionized and transfer charge to the analyte, while exploding at supersonic speed
 - We can operate in negative or positive modality
 - It generates single-charged ions $M+H$ and $M-H$
 - It is more tolerant to contaminants
 - It can directly analyze cells (!)
 - MALDI resolution can be influenced by the differential time of ionization of analyte molecules, and differential initial velocity of the ions before the TOF acceleration
- A mass spectrum is a plot with m/z on the x axis and relative abundance in the y axis
 - Peaks are typically really sharp
 - From each ion I get a characteristic pattern of peaks, due to spontaneous fragmentation
 - I have spontaneous fragmentation only if my ionizing source is strong enough
- The accuracy of the mass spectrometer is measured in ppm
- The resolution of a mass spectrometer is obtained by dividing height of the peak by its width at half height
- A mass filter is a device that separates analyte ions based on their m/z
- A magnetic sector mass analyzer (MSA) deflects the ions by different radii depending on their m/z
- The quadrupole mass filter allows a stable path only for a selected m/z , and we can get a spectrum by scanning all the target m/z values
 - It is used also in tandem MS
- Orbitrap (Makarov, 2000) is the newest and most sensitive analyzer
 - There is a spindle-like electrode around which the analyte ions spin and a barrel-like electrode
 - The ions oscillate with a frequency related to their m/z , giving a signal
 - The signal is deconvoluted with the Fourier transform to yield the fundamental harmonics, which are related to m/z
- The time of flight analyzer (TOF) accelerates ions with an electric field in a way that is proportional to m/z
 - The time that is required for the ions to reach the detector is proportional to m/z
- The reflectron is a variant of TOF that improves resolution by compensating for different initial velocities
- MS detectors are diodes that generate secondary electrons when struck by an ion
 - Electron multipliers produce many secondary electrons for each ion. and therefore are more sensitive, but require extensive maintenance

- Tandem MS combines 2 MS in the same instrument
 - I avoid fragmentation in the first step using a soft ionization of the molecular ion
 - A collision cell between the 2 MS filled with inert gas allows to fragment the analite

Peptide mass fingerprinting

- It is the main bottom-up approach
- The standard workflow starts with 2D-PAGE that allows to recover unique spots
- It is important to chose a staining method that is compatible with MS
- We can also use multidimensional HPLC as an alternative to 2D-PAGE
- Spots are then cut and destained
- Spot picking can be done manually or with a robot
 - Manual picking is susceptible to keratin contamination
 - There is a risk for gel deformation
- The protein is then digested by trypsin to yeld peptides
 - Trypsin cuts after K or R, but only if not followed by P
 - It is important to have a complete digestion to avoid missing cleavage sites
 - Better to use volatile buffers to eliminate them easily afterwards
 - Trypsin can also self-digest (!) and the deriving peptides are really useful as a standard for MS
 - * It is an internal calibrator
- Peptides are then purified by reverse chromatography or Zip tips
 - RC uses apolar stationary phase on polar mobile phase
 - Zip tips are a miniature RC column (!)
- We perform MS on the peptides, getting a fingerprint of the protein
 - In MS peptides must be ionized in a gas phase
 - MS measures the m/z ratio of the peptide ions
- I can then identify the protein by searching for my fingerprint in databases
- Masses can be reported as monoisotopic or average
 - Entry level instrument cannot differentiate isotopes, therefore report only average masses
- To match a fingerprint with a database, I check how many hits on the same protein I have with my masses
 - I choose the protein with the maximum number of hits
- Main PMF databases are Mascot and ProFound

Peptide de novo sequencing

- Peptide de novo sequencing uses MS/MS spectra to determine the sequence of a protein without using any previous knowledge
- This is in contrast with database search, that identifies the peptide using databases
- The fragmentation process produces different kinds of ions
 - Give the low collision energy employed, most fragmentations involve peptide bonds
 - If the charge is retained in the N-terminal fragment, the ion is termed a, b or c
 - If the charge is retained in the C-terminal fragment, the ion is termed x, y or z
 - Fragmentation of the peptide bond produces y or b fragments
- I can recover the mass of the residues by analyzing the mass difference between ions
- Given a MS/MS spectrum, the software Peaks can give the protein sequence, with a confidence for each residue