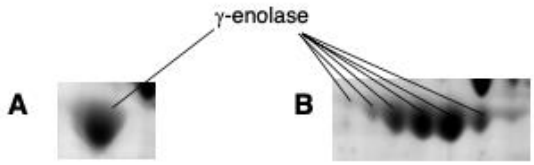
Limitation of 2D gels:

- Reproducibility: experiments are difficult to automate and samples must be run at least in triplicate to rule out effects from gel-to-gel variation
- Small dynamic range of protein staining as a detection technique and less abundant proteins might be missed.
- Co-migrating spots forming a complex region
- Difficulties with low MW proteins (<15 kDa): ultrafiltration is not powerful in terms of resolution.
- Requires Mass Spectrometry for the identification (needs proteins digested by trypsin)
- Weak spots and background; streaking and smearing
- Hydrophobic proteins are not compatible with the IEF--> marginal solubility leads to protein precipitation and degradation-smearing
- 2D electrophoresis is NEVER quantitative: it doesn't account for different expression in cells.

For most of the 2D electrophoresis limitations, multidimensional HPLC is the solution.

The pH gradient changes the resolution:

These two images belong to gel obtained from the same protocol but with a different pH gradient; in image A we have a wide gradient and we are not able to differentiate the protein isoforms while in image B we can see several spots due to the narrower pH gradient.



Gygi and Aebersold experiment:

Studying the relationship of gene expression (with mRNA transcripts) and protein levels (with incorporation of radiolabeled methionine) in yeast, they saw that of about 4000 expressed genes, 3000 were not detected in the 2D-SDS-PAGE analysis.

In order to improve the reproducibility/variability we can perform a:

2D Fluorescence difference gel electrophoresis (DIGE):

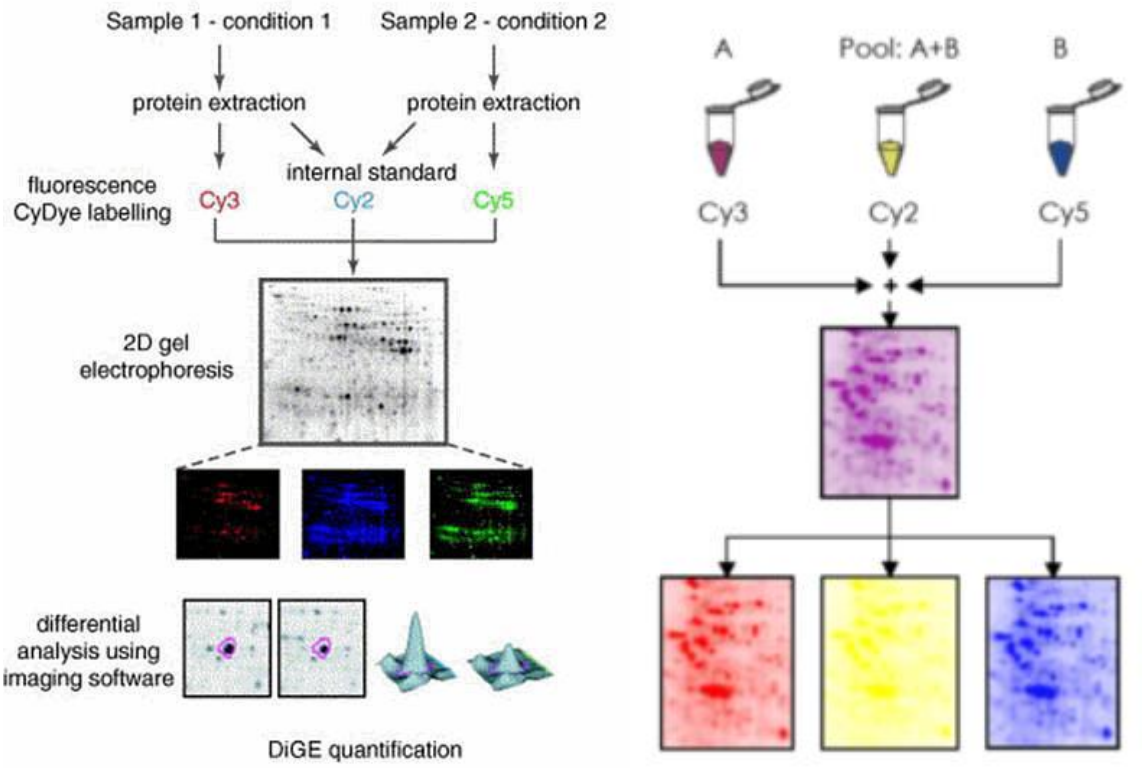
This variant allows simultaneous separation of up to three samples on one gel, bringing a new level of statistical confidence and reliability to 2D gel electrophoresis.

The system uses fluorescent **cyanine** (Cyn). Cyanines carry an N-hydroxysuccinimidyl ester reactive group that covalently binds the ε-amino groups of lysine residues in proteins. They are mass and charge matched but have unique fluorescent properties.

Proteins are labeled prior to running the first dimension (IEF) and then can migrate simultaneously on the 2D gel.

The methodology allows the inclusion of an **internal standard** that ideally is a pool of all the samples within the experiment. The internal standard is labeled with one of the cyanine dyes (typically Cy2) and runs on each gel in the experiment.

This creates an image that is the average of all experimental samples, with all proteins in the experiment represented. The presence of the internal standard in every DIGE gel provides an inherent link among samples that can be exploited to match and normalize the protein quantities across samples.



(*PubChem= database of compounds, for understanding the structure and the information available in the state of the art about that compound. You can find all the articles about each bioactivity that it has).

With differential gel electrophoresis (like 2D gel + using fluorescence marker, DIGE) we can do the comparison in the same gel and **rule out issues like gel-gel variations.**

There are alternative techniques for separating proteins.

1.Multidimensional LC

Combines orthogonally two different separations relying on different principles. Like in the 2D electrophoresis but improving resolution power thanks to chromatography. It's one of the most effective tools in analytical proteomics.

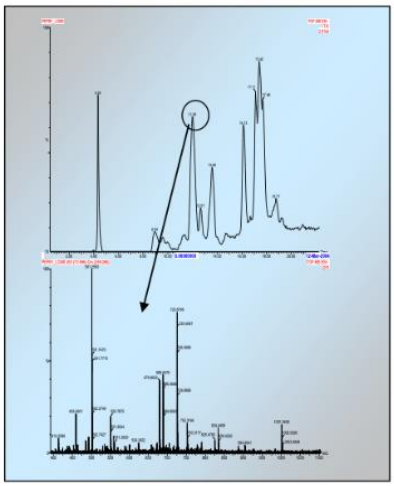
Several separation principles:

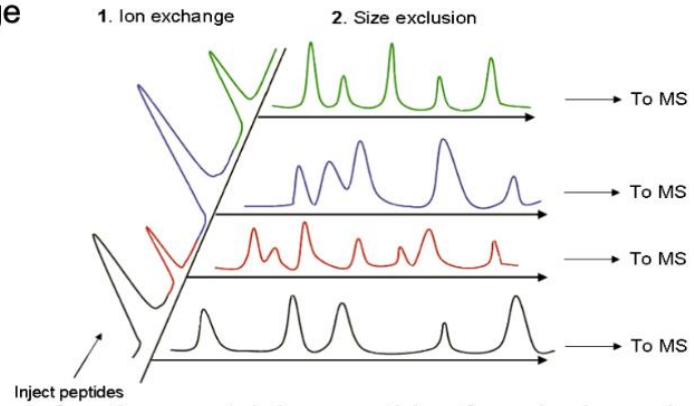
- Reversed phase, hydrophobicity
- Ion exchange, net positive/negative charge

- Size exclusion, peptide size, molecular weight
- Affinity chromatography, interaction with specific functional groups

Then each fraction will undergo a second separation. We combine separation modes in series (tandem HPLC) in order to obtain a greater resolution of peptides in a mixture. The idea behind tandem LC is that the combination of dissimilar separation modes allows a greater resolution of peptides in a mixture. Multidimensional Protein Identification Techniques (MudPIT) or Tandem HPLC.

First dimension: ionic exchange chromatography





Each fraction containing peptides from ionic exchage column is then separated on a reverse phase column and then revealed with a detector ESI-Q/TOF

2.Hydrophilic interaction liquid chromatography (HILIC)

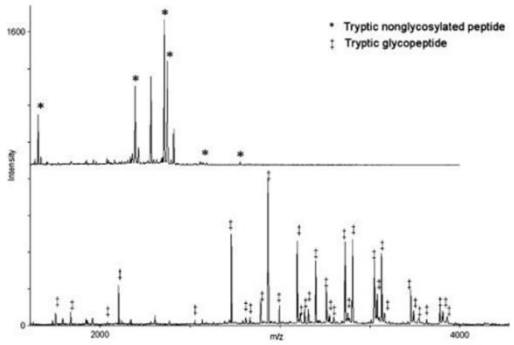
Based on hydrophilic interactions between the analytes and the hydrophilic stationary phase, with either highly polar, or hydrophilic compounds interacting most strongly.

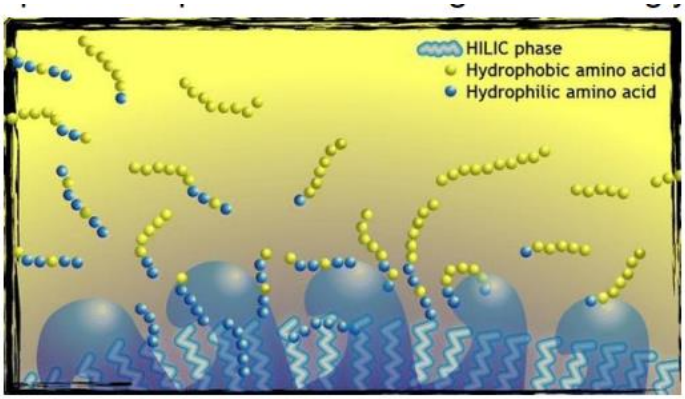
•A column with a hydrophilic stationary phase

•An eluent with water, buffer and a high concentration of water-miscible organic solvent.

The elution is achieved by a water gradient, hydrophilicity.

Optimal for glycosylated proteins, in the picture we can see the MALDI-TOF spectra of tryptic non-glycosylated peptide and tryptic glycopeptide and there is clearly a big difference.

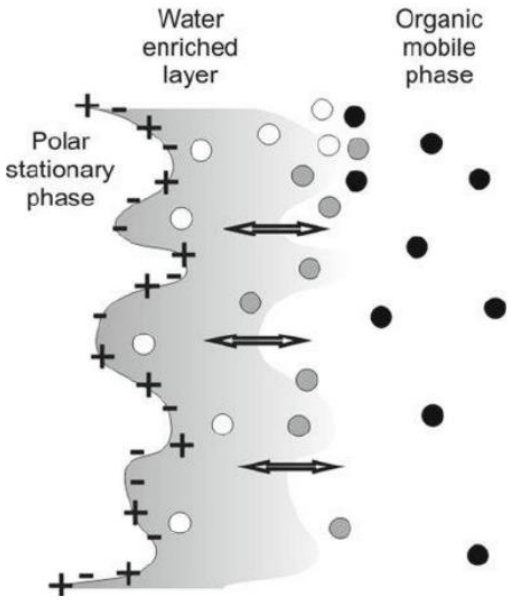




HILIC PRINCIPLE:

The organic mobile phase passing the polar stationary phase forms a **water enriched layer** next to the stationary phase. In the water enriched layer we have the interaction with a mix of H-bonds, ion exchange, dipole-dipole interactions. Therefore, the samples will be retained with a different distribution between stationary phase, water layer and mobile phase; polar compounds are then selectively eluted from the column as the mobile phase becomes more hydrophilic according to their hydrophilicity.

The hydrophilic analytes partition into this layer and as the mobile phase becomes more hydrophilic, they are eluted in order of increasing hydrophilicity (white circle = hydrophilic analyte; grey circle = semi-hydrophilic analyte, black circle = hydrophobic analyte).



Typical questions due to this part:

- How bidimensional electrophoresis works, main disadvantages, is it qualitative of quantitative?
- What are the alternatives?
- Having an idea of sample preparation.

After separating the protein what can we do?

- We can try to compare the gel to previous existing gel in database in order to find matches (Melanie)
- If we have a target protein, we can detect it with an antibody (western blotting);
- Edman degradation (N-terminal microsequencing)

- Mass spectrometry

MASS FINGERPRINTING WORKFLOW

Peptides obtained from cleavage are a **fingerprint** of our protein. It's a **bottom-up** approach because we start from a complex protein mixture, we separate, isolate and digest them in order to perform the mass spectrometry and obtain the protein ID.

This process requires the simplest instrumentation available but works only with proteins pre-existing in the database.

After the gel:

- cut single spots
- decoloration (stains can mess with MS analysis because dyes affect ionization process)
- in gel digestion with proteases with specific target sites
- analyze peptides with mass spectrometry
- we compare our data with databases containing peptides deriving from in-silico digested proteins

SPOTS RECOVERING

Can be done manually with the tip of a pipet --> risk of contamination with keratin of the researcher.

Big companies have automatic tools (spot pickers after you select the coordinates) but during the electrophoresis we could have a variation or deformation (correct coordinates but the spot is not there).

DIGESTION

The easiest way is to use **proteases** (it is important to remember that the amino acid residues are repetitive and the enzyme cuts quite efficiently).

The most used is **trypsin** (cuts the C-terminal of arginine and lysine) because:

• Robust, stable enzyme

• Works over a range of pH values & temperature

• Quite specific and consistent in cleavage (peptide fragments of between about 6–20 aminoacids are ideal for MS analysis and database comparisons)

• Cuts frequently to produce "ideal" MW peptides

• Inexpensive, easily available/purified

• Does produce "autolysis" peaks (which can be used in MS calibrations)

BUT it's difficult to obtain a complete digestion to the theory: the efficiency of the enzyme is never 100% and due to the shape of the protein some residues may not be available for the enzyme ---> when we are in the database, we must provide to the software the number of missed cleavage sites so that all the possible combinations with a missed cleavage are considered (peptides are also lost during different washes, some of them don't ionize or get lost during purification).

Sometimes we may have **proteins that contain several lysine and arginine**: treating them with trypsin would mean to have too many cleavages and obtaining too many peptides --> we can substitute it with **Glu-C** (endoproteinase that cleaves at the carboxyl side of glutamate residues in either ammonium acetate or ammonium bicarbonate buffer. In a sodium phosphate buffer, however, the enzyme cleaves at both glutamate and aspartate residues).

We can also use **combinations of different proteases** for achieving what we want in digestion.

Now we obtained the peptides that are the protein fingerprint and we need to detect the combination of them that leads to the protein.

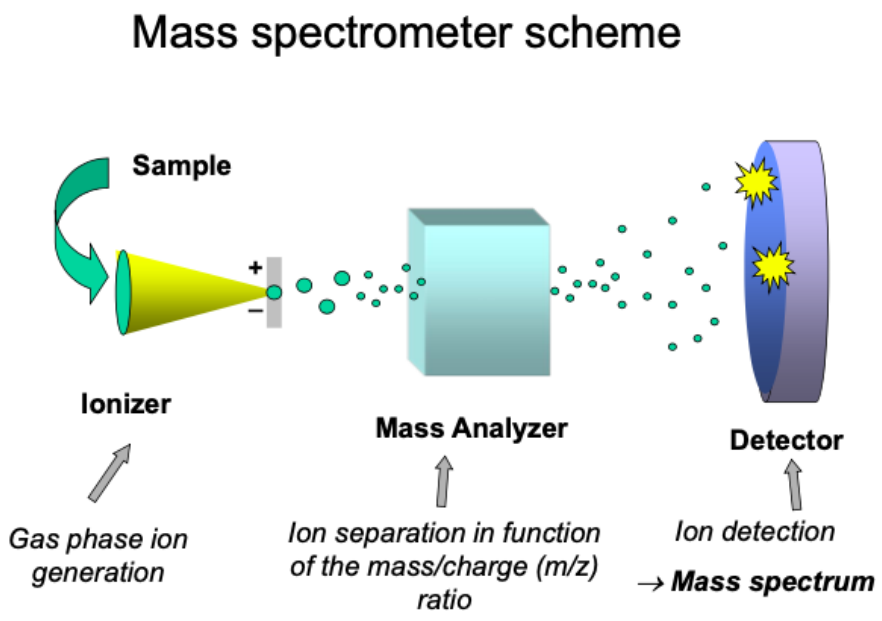
There is an **additional step** to get rid of **salt** and other contaminants--> REVERSE HPLC (everything is packed inside of the tips of a micropipette). In Reverse HPLC, stationary phase is hydrophobic or non-polar and the starting mobile phase must be more polar than the stationary phase.

Peptide mass fingerprinting

Now we can perform the ionization (mostly in gas phase), it can be:

- Protonation
- Cationization
- Deprotonation
- Electron release
- Electron capture

This provides us **the mass-to-charge ratio**, from which we infer **molecular weight** of the peptide.



Mass spectrum: intensity *vs* m/z ratio plot representing a chemical analysis

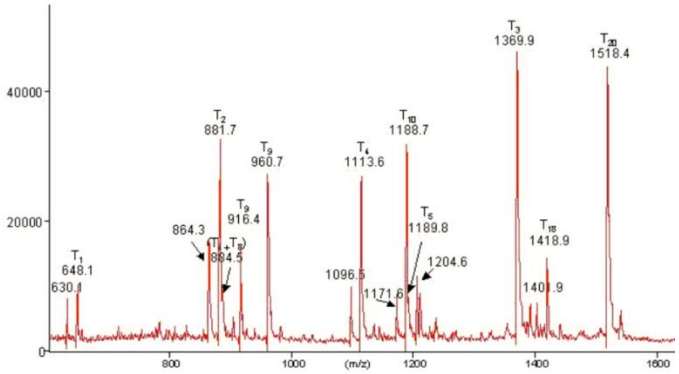
It's composed by three main building blocks:

- Source of ion
- Mass analyzer
- Detector

Samples can be added either manually (usually done with MALDI) or by direct introduction (HYPHENATION) where everything is performed automatically.

A typical mass spectrum is characterized by sharp, narrow peaks. In the x-axis we have m/z ratio of a given ion (for singly charged ions this corresponds to the mass of the ion) and the height of peaks indicate the relative abundance of a given ion. Peak intensity indicates the ion's ability to desorb or 'fly'. The intensity (is a relative abundance) doesn't mean that if the intensity is high the peptide existed in higher concentration.

relative abundance in respect of the other ions?



For the exam: we need to know how to draw a mass spectrum.

IONIZATION

The various methods mostly differ by the different energy provided:

- In **metabolomics** we are dealing with very small molecules and we need to break chemical bonds --> **hard** method
- For **proteomics** we are dealing with huge molecules that need to remain intact --> **soft** ionization

There are different ionization methods:

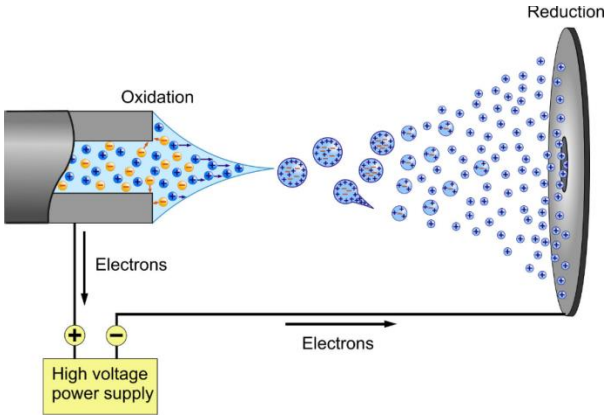
- **MALDI** (Matrix Assisted Laser Desorption): soft method. Used for peptides, proteins, DNA, up to 500 kDa.
- **ESI** (electrospray ionization): soft method. Proteins up to 200 kDa.
- **FAB** (Fast Atom Bombardment): semihard method; sometimes used for some little protein up to 5kDa). Sample is bombarded with high energy (4000 to 10,000 electron volts) beam of atoms (inert gas such as argon or xenon).
- **EI** (Electron Impact): hard method, up to 1 kDa.

possible question: what's the meaning of soft ionization? (That the molecule doesn't break, otherwise we'll have too many peaks)

ELECTROSPRAY IONIZATION (ESI)

Sample is in a liquid form, and it gets sprayed through a needle, in presence of high voltage and nitrogen there is the formation of charged droplets, who evaporate and shrink their size. Increasing of the charge until a limit, called the Rayleigh limit (the density of the charge increases) -->coulombic explosion.

At the end we have **de-solvated ions,** that can be both positive and negative (multiple charged ions, different from MALDI) that goes to the analyzer for the separation.



MAIN FEATURES:

- Can be coupled to HPLC
- Can be modified to "nanospray" system with flow rates < 1 μL/min

  normal ESI flow rates 1-500 μl/min
  normal analytical HPLC 0.1-1 ml/min
  nanoESI 1-500 nl/min
- Very sensitive technique, requires less than a picomole of material
- Strongly affected by salts & detergents
- Positive ion mode measures $(M + H)^+$ (add formic acid to solvent)if sample has amide and amino groups
- Negative ion mode measures $(M - H)^-$ (add ammonia to solvent) if sample has carboxylic acids and hydroxyls

LIMITATION:

The sample needs to be very pure (big limitation if we think about diagnostic).

ESI TRANSFORMATION:

Small samples of up to 1200 Da MW usually produce singly charged ions, while larger samples yield ions with multiple charges.

Multiple charged species form a gaussian distribution with those having the most charges showing up at a lower m/z value. Software can be used to convert these multiple spectra into a single (zero charge) profile which gives MW directly--> MS interpretation is much easier and it greatly increases the signal to noise.

PROTEIN STRUCTURE:

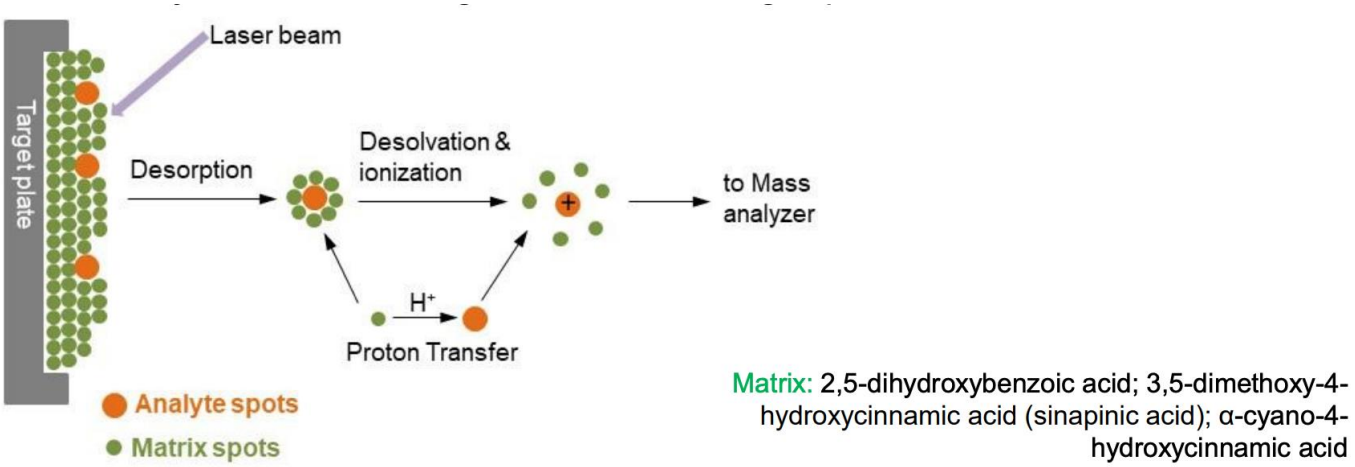
ESI spectra are sensitive to the conformation of the protein: folded, ligated or complexed proteins tend to display non gaussian peak distributions, with few observable peaks weighted toward higher m/z values. Denatured or open form which ionize easier tend to display many peaks with a classic gaussian distribution.

MALDI IONIZATION:

The energy is provided by a laser (UV radiation), the sample is **solid** and is **co-crystalized with a matrix**.

- The matrix is a **chromophoric acid** used to protect the analyte by absorbing the energy from the radiation and preventing the sample breaking. We will have a phase change, with the so called popping.
- The ions formed by protonation float from the analyte to the analyzer.
- Mostly produce **single-charged ions** (positive or negative depending on the nature of the sample)
- More tolerant to contamination, we can also have complex dirty samples

Samples are manually added on the target plate, the evaporation of residual water or other solvent allows the formation of a crystal lattice into which the peptide sample is integrated. Then the target is placed into the source--> the matrix absorbs photons from the beam and becomes electronically excited--> energy transferred to the peptides or proteins which are ejected from the target surface into the gas phase. Ionization process produces both positive and negative ions, but positive ones are almost always the species of interest. Only singly charge ion generated, more robust than ESI.



LIMITATION:

We have a co-crystal, a 3D structure, but the laser hits only the surface-->some peptides are not ionized because they are in the middle of the crystal --> more replicates in order to avoid artifacts. ESI is better for the analysis of complex mixtures.

SENSITIVITY, ACCURACY AND RESOLUTION

Good data should have:

- Sensitivity: we need to identify a very tiny amount
- Resolution: the capability to distinguish peaks close in the spectrum--> the narrower the peak, the best the resolution is
- Mass accuracy: closeness to the real value of the peptide (important for database matching)

ACCURACY: usually calculated as a part per million error. It's related to the calibration of the analyzer to properly assign the true m-z ratio to a detected ion and to the resolution of the detector response, which is in the form of an intensity spike or peak within a mass spectrum.

$$\text{Mass accuracy (ppm)} = \frac{m_{\text{observed}} - m_{\text{theoretical}}}{m_{\text{observed}}} \times 10^6$$

RESOLUTION: m/Δm, the former is the m/z value obtained from the spectrum and the latter is the full peak width at half height.

There are two known problems in the gaseous plume of analytes above the target that cause a decrease in the resolution of the detected analytes:

- Initial spatial distribution where not all of the desorbed analytes are at the exact same distance from the detector at the start of their flight.
- Not all of the analytes may have exactly the same velocity at the start of their flight toward the detector.

Question: how the MALDI works and the main differences with electrospray (there is a difference also in the sample addition)

MASS ANALYZER:

The mass analyzer is the heart of the mass spectrometric instrumentation--> separation of molecular ions (M+) and analyte ions (e.g., [M+H]+) in the gas phase.

The most fundamental aspect of the mass analyzer is the ability to **separate ions according to their mass-to-charge (m/z) ratio.**

MAGNETIC SECTOR ANALYZER (MSA):

When charged particles enter a magnetic field, they possess a circular orbit that is perpendicular to the poles of the magnet. The ions enter a flight tube (first field-free region) from a source through a source exit slit and travel into the magnetic field. The accelerating voltage in the source will determine the kinetic energy that is imparted to the ions. The magnetic field will deflect the charged particles according to the radius of curvature of the flight path, that is directly proportional to m/z of the ion--> You can calculate the mass-to-charge by calculating the bending.

QUADRUPOLE MASS ANALYZER (Q):

Four parallel metal rods with different charges, the applied voltages affect the trajectory of ions traveling down the flight path. For given dc (direct current, electrons go in the same direction) and ac (alternate current, electrons wiggle back and forth) voltages, only ions of a certain mass-to-charge ratio pass through the quadrupole filter and all other ions are thrown out of their original path. They need to have a finite amplitude of oscillation, so they are stable and can pass.

https://www.cif.iastate.edu/mass-spec/ms-tuto-rial

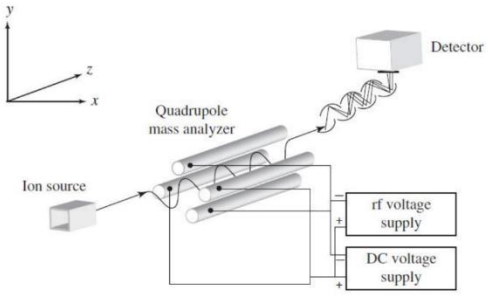


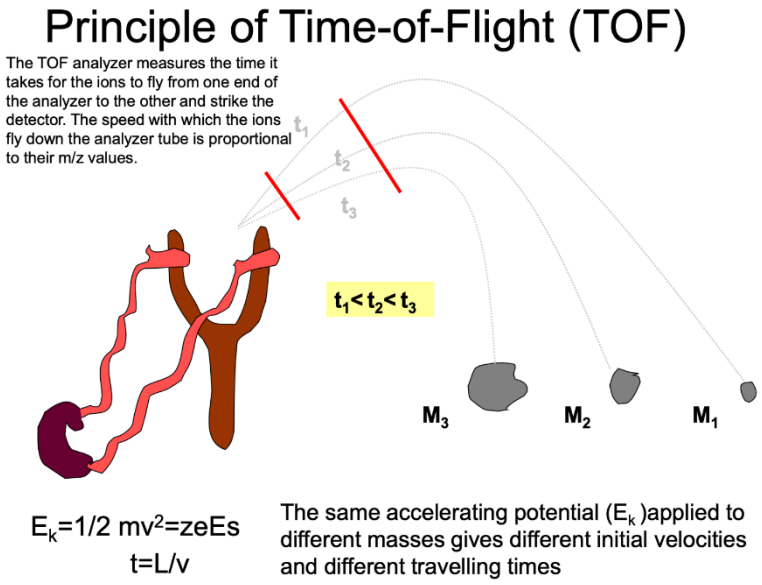
FIGURE 30.11 Quadrupole orientation and the configuration for the connections of the DC voltage (*U*) and radio frequency (rf) voltage (*V*). Ions are accelerated into the quadrupole by a small voltage of 5 eV, and under the influence of the combination of electric fields, the ions follow a complicated trajectory path.

ORBITRAP:

Ions are trapped in an electrostatic field between an inner and outer electrode: as the ions rotate around the inner, they process along its axis with a frequency characteristic of their mass to charge ratio. Acquisition of transients and the Fourier transformation of that signal yields frequency and their intensities. A simple relationship converts frequencies into m/z values. Ions can rotate around the central electrode, the frequency of the oscillation can be related to the ratio.

TOF:

Most common mass spectrometer that is coupled to the MALDI ionization technique is the time-of-flight (TOF). It separates compounds according to their mass-to-charge ratios through a direct relationship between a compound's drift time through a predetermined drift path length and the analyte ion's mass to charge ratio.



Initially, ions have similar E_k imparted to them from the draw-out pulse (representing time zero), which accelerates them into the flight tube. ---->different masses= different velocities according to their relationship between E_k and mass represented by $E_k = zeV = 1/2mv^2$ --> $m/z = 2eVt^2/L^2$

Ions with the same ratio but different kinetics are refocused by the reflectron therefore reducing differences in flight time due to different initial velocities--> the reflectron is used to increase the resolution.

Major components of TOF mass spectrometer

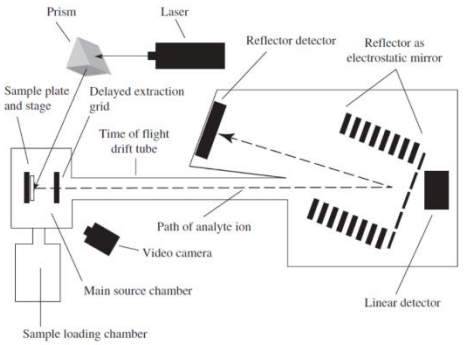


FIGURE 30.7 Components of a time-of-flight mass spectrometer illustrating the major sections including the source, drift tube, reflectron electrostatic mirror, and detector.

REFLECTRON

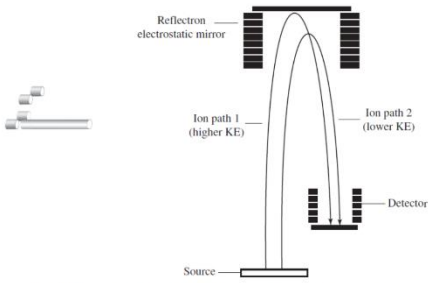


FIGURE 30.9 Electrostatic mirror focusing of two ions that have the same *m/z* value but slightly different kinetic energies. Ion path 1 possesses slightly higher kinetic energy in relation to ion path 2. Ion path 1 travels slightly farther to match that of ion path 2. The two ions are focused and arrive at the detector at the same time.

DETECTOR

Today's detectors (ion channel and electron multipliers) provide electronic signals via secondary electronic emission when struck by ion. A timing mechanism integrates these signals with scanning voltages to allow the instrument to report which m/z has struck the detector.

HYBRID INSTRUMENTS / TANDEM MS

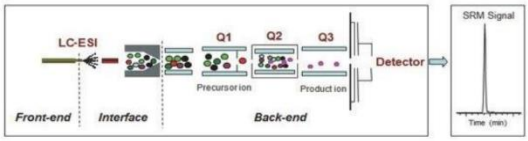
Combines two or more mass analyzers of the same or different types, the first one isolates the ion of interest (parent ion), between them there is a collision cell where the parent ion finds a gas and collide with it (it could be for e.g. argon), after the collision ions with high energy are produced. For decreasing the energy, we have the fragmentation (of selected parent ions, that is different from performing a hard ionization) and, after that, they pass through the second analyzer, which obtains the mass spectrum of the fragment ions (daughter ions spectrum)→ provides structural information about molecule and separation and identification of compounds in complex mixtures.

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4393390/

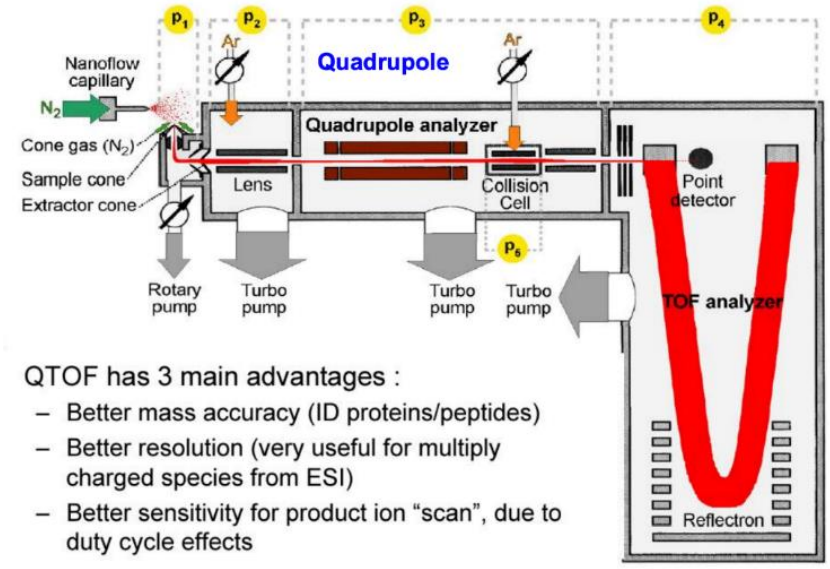
Tandem Mass Spectrometry (MS-MS) spectra reveal fragmentation patterns to provide structural information about a molecule.

Different MS-MS configuration:

- Quadrupole-quadrupole (low energy)
- Magnetic sector quadrupole (high)
- Quadrupole time of flight (low) (Q-TOF)
- Time of flight-time of flight (low)



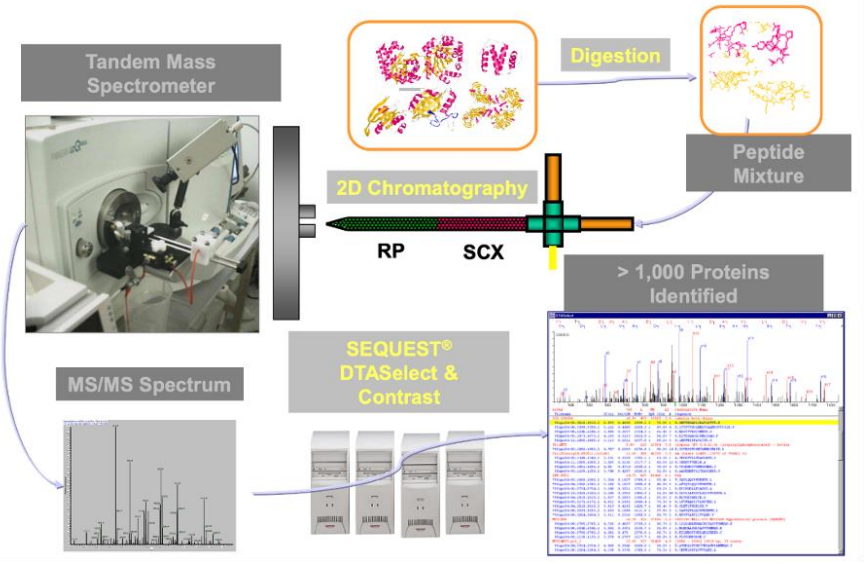
Q-TOF Mass Analyzer



QTOF has 3 main advantages :

- Better mass accuracy (ID proteins/peptides)
- Better resolution (very useful for multiply charged species from ESI)
- Better sensitivity for product ion "scan", due to duty cycle effects

Overview of Shotgun Proteomics:



SOFTWARE FOR PROTEIN IDENTIFICATION (examples in slides):

- SEQUEST: was one of the first tandem mass spectrometry database search program used for protein identification that correlates uninterpreted tandem mass spectra of peptides with amino acid sequences from protein and nucleotide databases.
- MASCOT: for identification, characterization and qualification of proteins using mass spectrometry data from primary sequence databases
- X! Tandem open source: software that can match tandem mass spectra with peptide sequences

- Proteome Discoverer Software: Multiple database search capability provides the option of applying multiple search algorithms and conveying their outputs to maximize and cross-validate results

PMF (Peptide Mass Fingerprinting)

Used to identify protein spots on gels or protein peaks from an HPLC run.

It's based on the fact that, if a peptide is cut up or fragmented in a known way, the resulting fragments (and resulting masses) are unique enough to identify the protein (in nature, only a minor portion of possible peptide combinations exist - a short peptide sequence is already highly protein-specific). Proteins with the same mass have different peptides with different mass spectra.

PMF requires a database of known sequences and uses software to compare observed masses with masses calculated from database.

MS performs differential profiling; protein identification is due to chemometrics techniques.

We are still unable to identify diseased vs. healthy profiles.

MASS LIST

In a typical MS, we obtain a list of peaks and experimental masses (this is the database mass list). After matching, masses are assigned to a protein. If there is no match it means that there are unknown masses missing in the database (we could have introduced ionization or purification issues). When we are matching with the database, we must provide some previous information such as missed cleavage sites during trypsin digestion (you need to consider all the situation deriving from this possible missing site). Or we can try to find PTMs (post translational modifications).

De-novo sequencing doesn't need a database

Possible question:

- what is a typical mass spectrometer

- Peptide mass fingerprinting workflow
- No mass accuracy and mass resolution
- main differences, main analyzers for proteomics
- How is the matching with the database performed

Proteomics is used most of the time in medical fields.

The first goal of proteomics is to make identity profiling. Distinguishing the peaks in diagnostic field which means different peptides.

Second goal is to identify the different proteins.

For the second goal the mass spectroscopy fingerprinting is used. Although when we digest different proteins we might get same mass we can differentiate them.