AFTER 2D GEL

1. Gel-Matching with databases (es: Expasy proteomics Server);

2. Western-Blot with specific antibodies;

3. N-terminal microsequencing;

4. Mass spectrometry: MALDI-TOF, MS-MS

AFTER 2D GEL

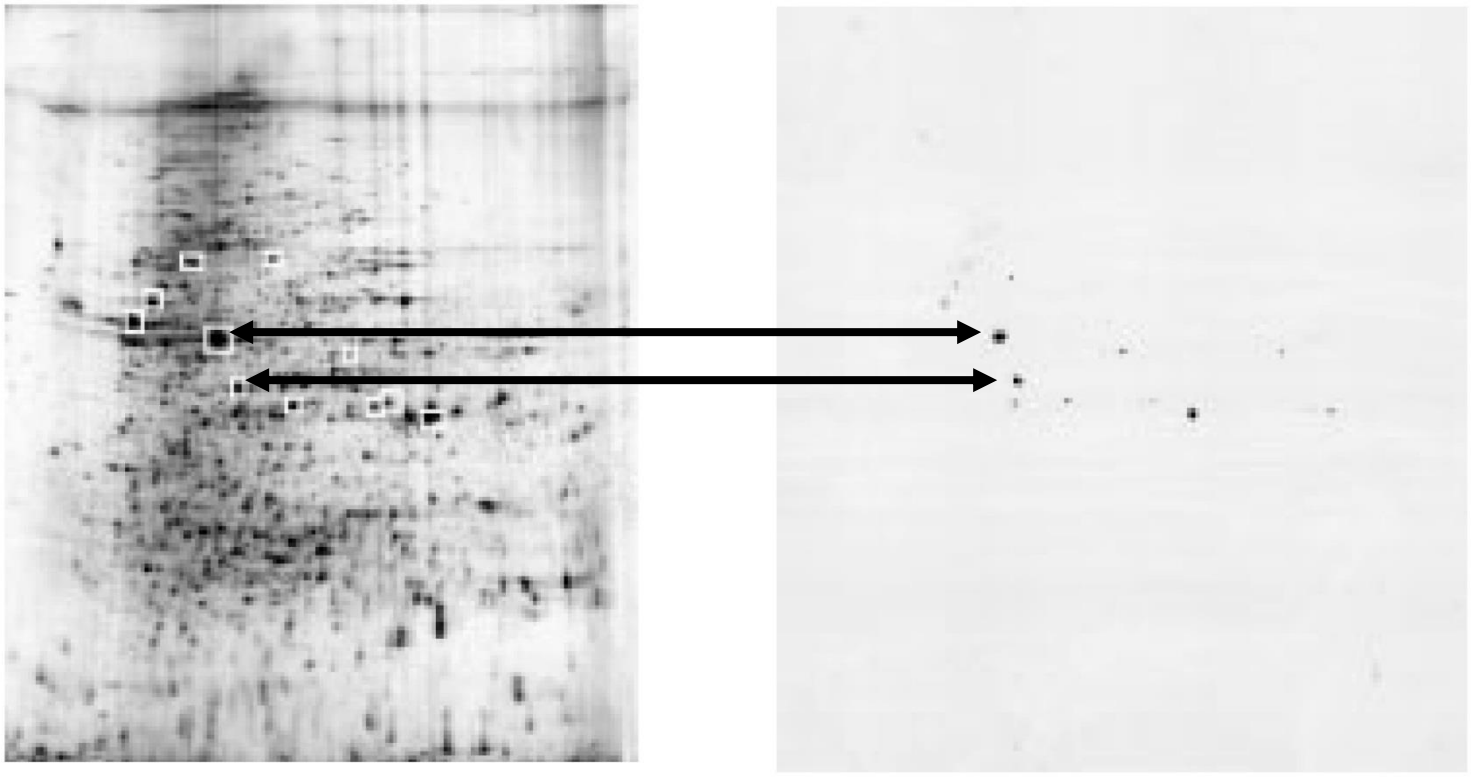
1. Gel-Matching with databases (es: Expasy proteomics Server);

2. Western-Blot with specific antibodies;

3. N-terminal microsequencing;

4. Mass spectrometry: MALDI-TOF, MS-MS

Detection via Western Blot



Glioma Silver Stain

Glioma Western Blot (anti-p53 antibody)

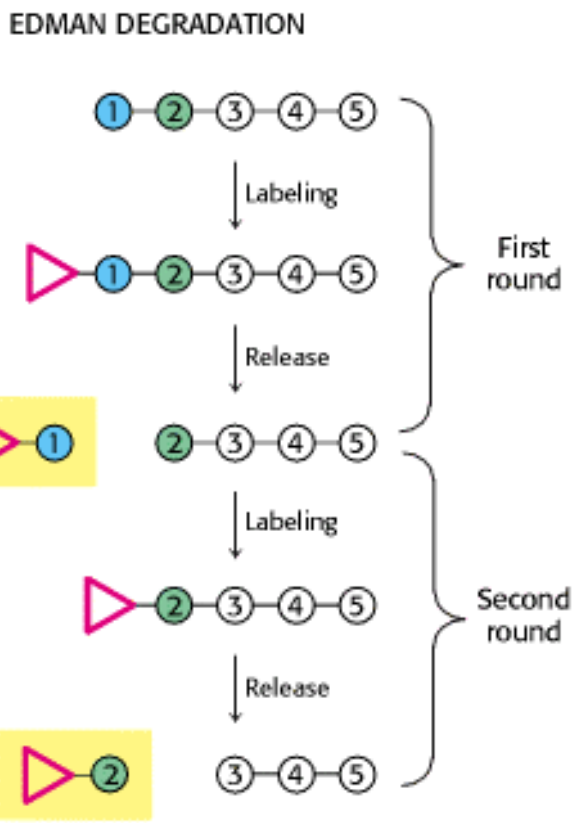
1. Gel-Matching with databases (es: Expasy);

2. Western-Blot with specific antibodies;

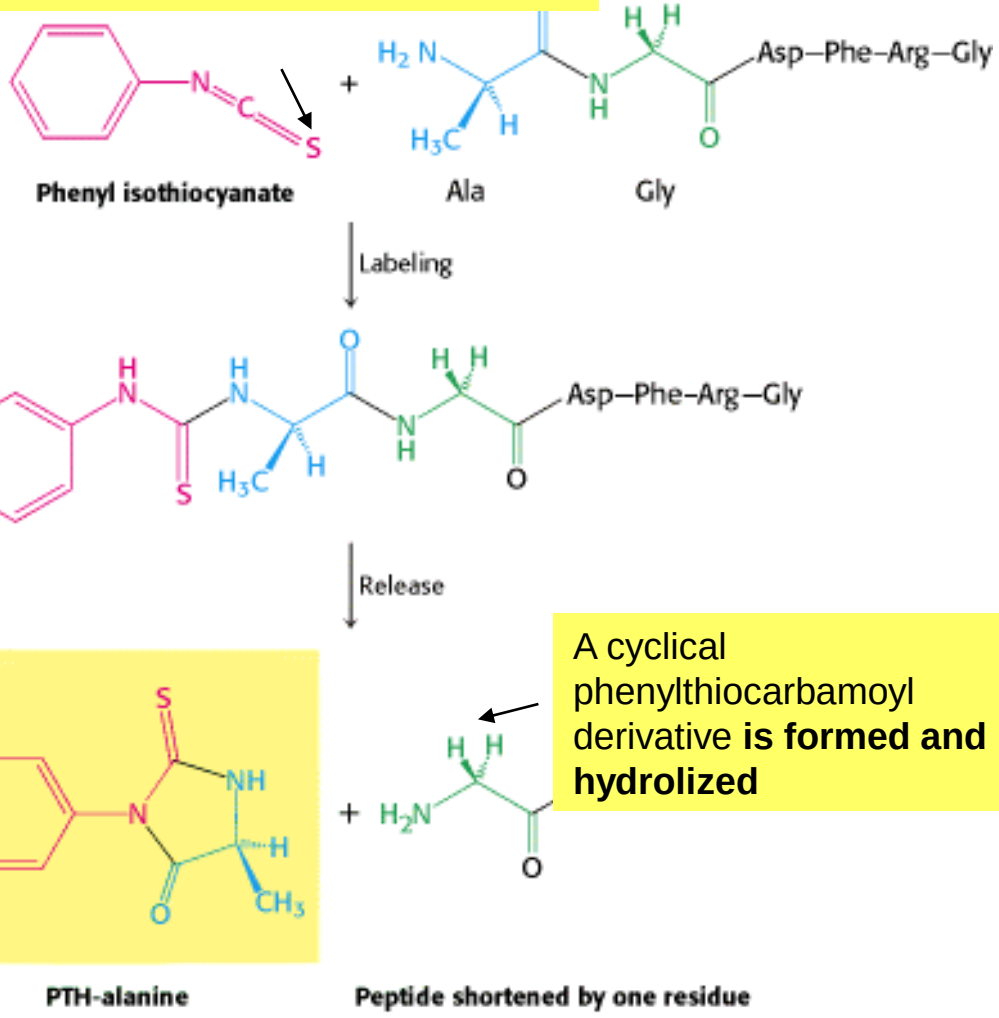
3. N-terminal microsequencing;

4. Mass spectrometry: MALDI-TOF, MS-MS

Edman degradation



Phenylisothiocyanate (**PTC**) reacts with N-terminal residues



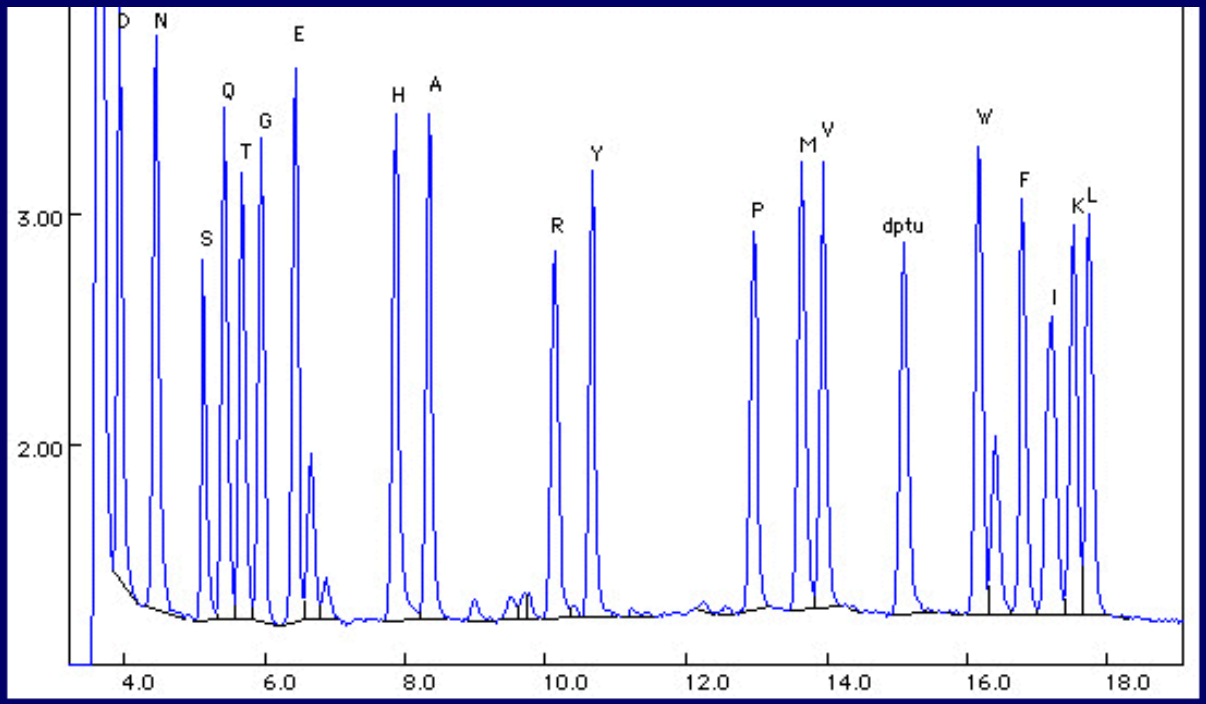
A cyclical phenylthiocarbamoyl derivative **is formed and hydrolized**

The thiazolinone amino acid is then selectively extracted and treated with acid to form the more stable phenylthiohydantoin (PTH)- amino acid derivative that can be identified by using chromatography or electrophoresis

Automatic sequencer

Sequencing

Protein sequence analysis employing Edman degradation chemistry commonly uses HPLC separation as the means for identification of the PTH-amino acid (phenylthiohydantoin amino acid) produced at each cycle. Each aminoacid is red at wavelength 269 nm and is identified by a specific retention time



After 2D gel

1. Gel-Matching with databases (es: Expasy);
2. Western-Blot with specific antibodies;
3. N-terminal microsequencing;
4. Mass spectrometry: MALDI-TOF, MS-MS

Spot identification: peptide mass fingerprint

Spot cutting

↓

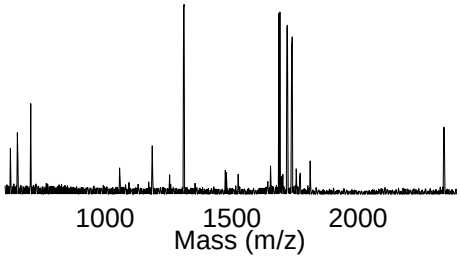
Decoloration

↓

“in gel” proteolytic digestion

↓

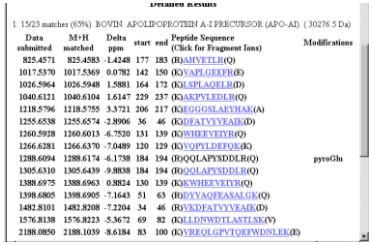
MS identification



YAEK	609.31
GLQQPVR	797.45
AYMESEK	857.38
SEEIDLPK	957.54

↓

Database searching

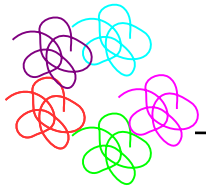


Detailed Results

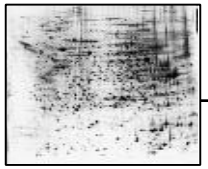
1. 15/23 matches (65%). BOVIN. APOLIPOPROTEIN A-I PRECURSOR (APO-AI). (30276.5 Da)

Data submitted	M+H matched	Delta ppm	start	end	Peptide Sequence (Click for Fragment Ions)	Modifications
825.4571	825.4583	-1.4248	177	183	(R)AHVETLR(Q)	
1017.5370	1017.5369	0.0782	142	150	(K)YAPLGEEPR(E)	
1026.5964	1026.5948	1.5881	164	172	(K)LSPLAQELR(D)	
1040.6121	1040.6104	1.6147	229	237	(K)AKPVLEDLR(Q)	
1218.5796	1218.5755	3.3721	206	217	(K)EGGGSLAEYHAK(A)	
1255.6538	1255.6574	-2.8906	36	46	(K)DFATVYVDAIK(D)	
1260.5928	1260.6013	-6.7520	131	139	(K)WHEEVEIYR(Q)	
1266.6281	1266.6370	-7.0489	120	129	(K)YQPVLDEFQK(K)	
1288.6094	1288.6174	-6.1738	184	194	(R)QQLAPYSDDLR(Q)	pyroGlu
1305.6310	1305.6439	-9.8838	184	194	(R)QQLAPYSDDLR(Q)	
1388.6975	1388.6963	0.8824	130	139	(K)WHEEVEIYR(Q)	
1398.6805	1398.6905	-7.1643	51	63	(R)DYVAQFEASALGK(Q)	
1482.8101	1482.8208	-7.2204	34	46	(R)VKDFATVYVEAIK(D)	
1576.8138	1576.8223	-5.3672	69	82	(K)LLDNWDTLASTLSK(V)	
2188.0850	2188.1039	-8.6184	83	100	(K)VREQLGPVTQEFWDNLEK(E)	

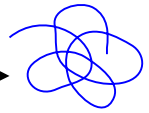
Peptide Mass Fingerprinting



Complex Protein Mixture



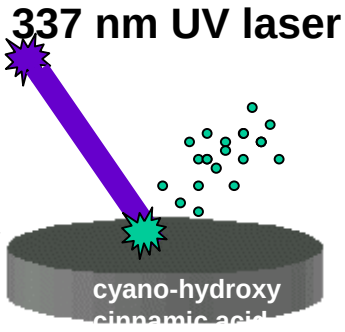
2D Gel Separation



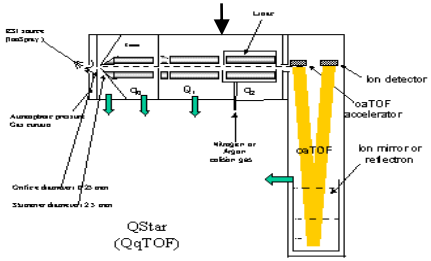
Purified Protein

Proteolysis

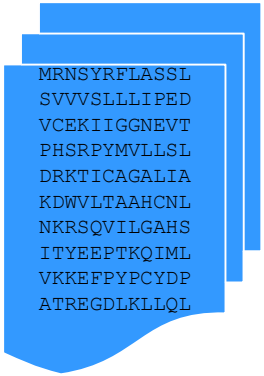
Peptide Digest



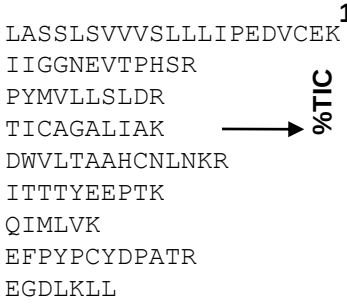
MALDI



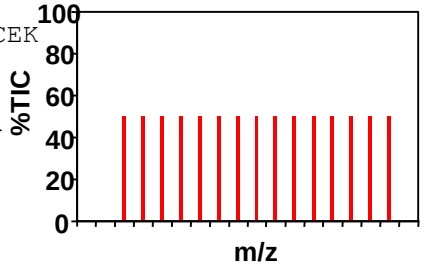
Mass Spec



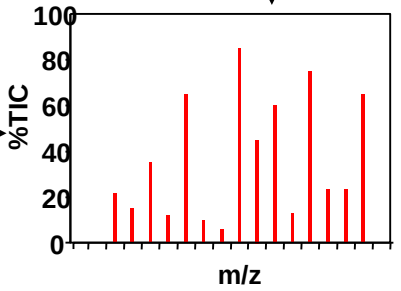
Protein Database



In Silico Digestion



Theoretical MS

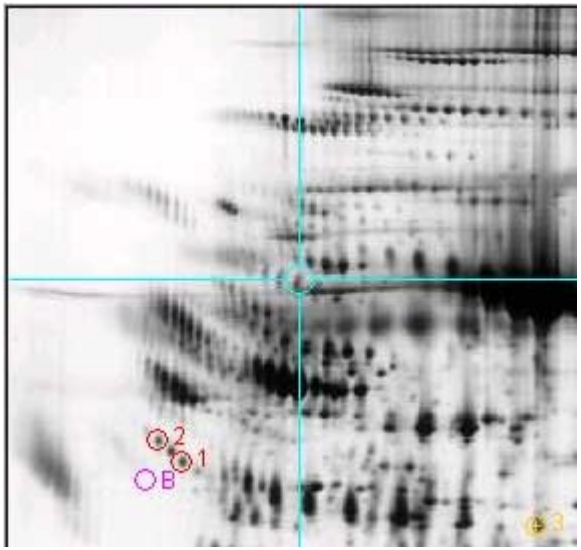


Experimental MS

Spot recovery

Manual

Keratin contaminations



Automatic

Gel deformation!

Enzymatic digestion

Aim: obtain peptide fragments from the protein of interest

a) **peptide mass fingerprinting**

b) **MS peptide sequencing**

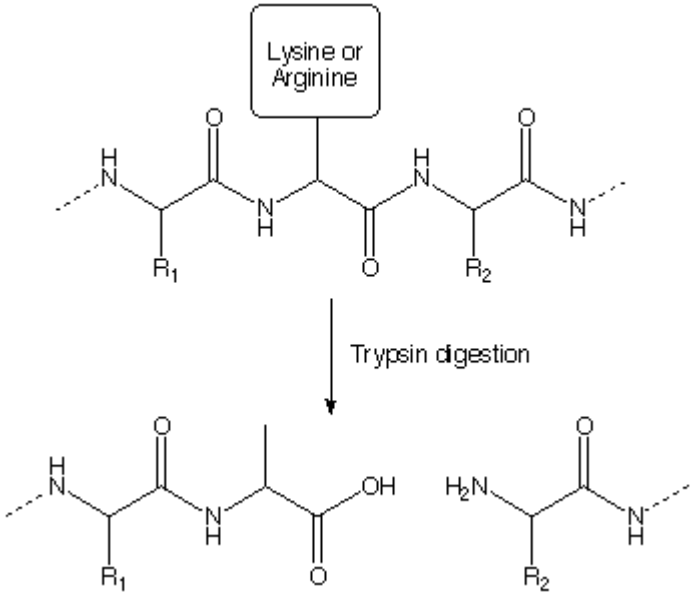
Method of protein cleavage	*Site of cleavage*	*Exception*	*pH range*
Trypsin	C-terminus of R-X, K-X	If X = P	7–9
Endoproteinase Glu-C (V8-DE)	C-terminus of E-X, D-X	If X = P	4–8
Chymotrypsin	C-terminus of F,Y,W,L,I,V,M	If X = P	7.5–8.5
Endoproteinase Lys-C	C-terminus of lysine, K-X	If X = P	8.5–8.8
Arg-C	C-terminus of arginine, R-X	If X = P	7.5–8.5
Elastase	Not very specific. C-terminal side of G, A, S, V, L and I.		8–8.5
Pepsin	C-terminus of F,L and E		2–4
Pronase	Pronase is a mixture of endo- and exo-proteinases. It cleaves almost any peptide bond.		7–8, dependent on proteases present

Complete digestion to obtain the minimum number of missing cleavage sites

Use volatile buffers (for easy elimination after enzymatic digestion)

Why Trypsin?

- Robust, stable enzyme
- Works over a range of pH values & Temp.
- Quite specific and consistent in cleavage
- Cuts frequently to produce “ideal” MW peptides
- Inexpensive, easily available/purified
- Does produce “autolysis” peaks (which can be used in MS calibrations)
  - 1045.56, 1106.03, 1126.03, 1940.94, 2211.10, 2225.12, 2283.18, 2299.18



PEPTIDE MASS FINGERPRINT

Peptide digested with trypsin

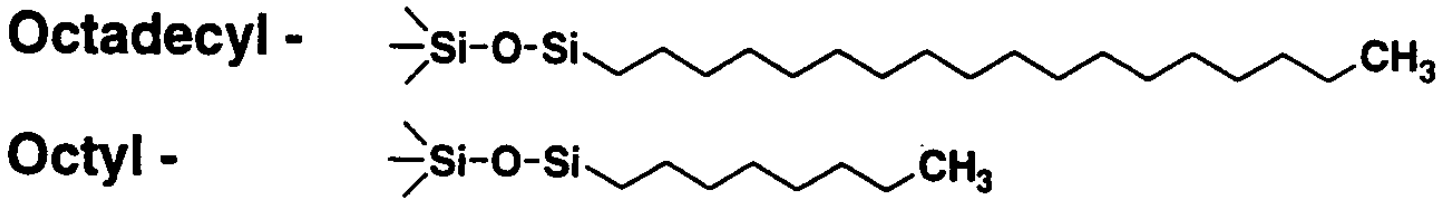
The 4 derived peptides have a well defined mass

YAEK	609.31
GLQQPVR	797.45
AYMESEK	857.38
SEEIDLPK	957.54

These 4 MW exactly identify the original peptide

After trypsin digestion: peptides are purified by reverse phase chromatography

The high resolving power of Reverse Phase [RP] Chromatography has made it the dominant mode of HPLC for both analytical and preparative separation of peptides and proteins, as well as other biomolecules.

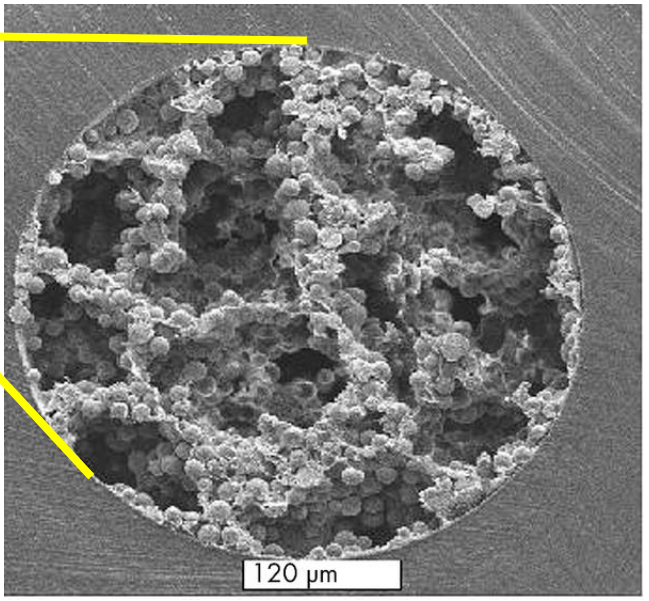


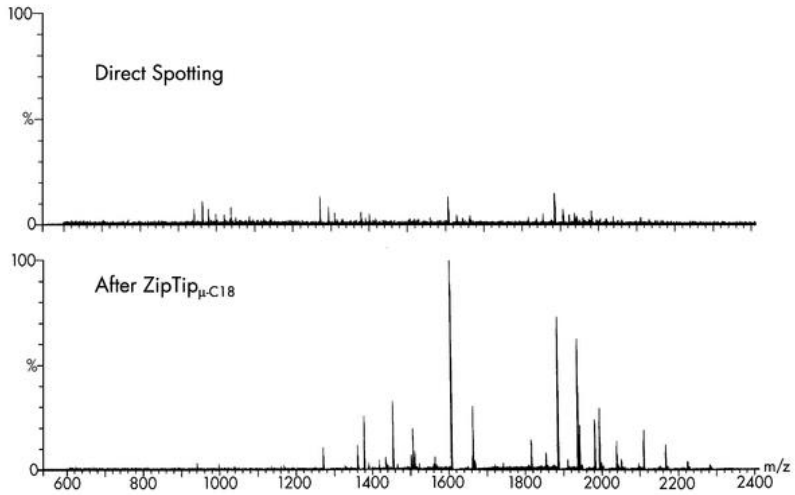
The term “Reverse Phase Chromatography” was used because RP is a form of partition chromatography where **chemically bonded phase** is **hydrophobic or non-polar** (e.g. octadecyl group), and the starting mobile phase (e.g. water) must be more polar than the stationary phase.

This is “reversed” from normal phase chromatography, where the stationary phase is polar or hydrophilic and the starting mobile phase is more non-polar or hydrophobic than the stationary phase, hence the term “Reverse Phase Chromatography”.

After trypsin digestion: peptides are purified with Zip Tip (tips packed with Reversed Phase (C18) resin

The ZipTip pipette tip is a 10 µL pipette tip with a bed of chromatography media fixed at its end. It is intended for concentrating and purifying peptide, protein or oligonucleotide samples.
C_{18} and C_4reversed-phase media for desalting and concentrating peptides and proteins.





How to detect peptides?

After digestion sample is ionized by:

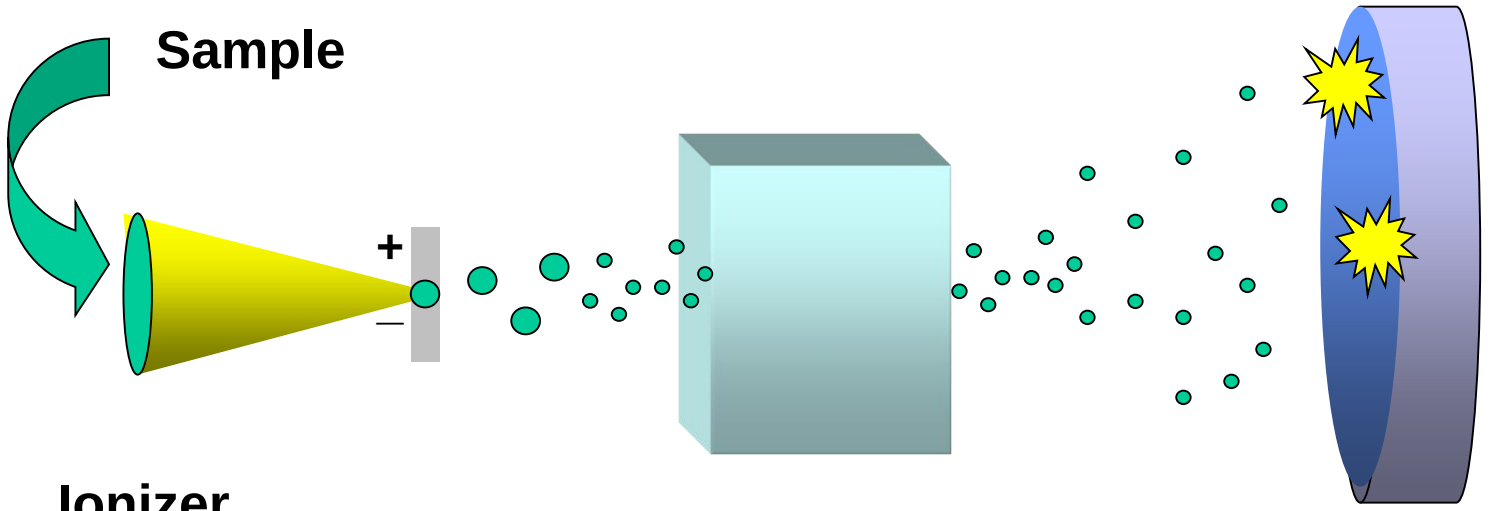
- Protonation $M + H^+ = MH^+$
- Cationization $M + Cat^+ = MCat^+$
- Deprotonation $MH = M^- + H^+$
- Eletron release $M = M^+ + e^-$
- Electron capture $M + e^- = M^-$

Analytes are ionized in gaseous phase

~~"A MASS SPECTROMETER MEASURES THE MW …"~~

"...A MS ANALYSIS GIVES THE MASS-TO-CHARGE RATIO (*m/z*) OF IONS…IN GAS PHASE".

Mass spectrometer scheme



Ionizer

Gas phase ion generation

Mass Analyzer

Ion separation in function of the mass/charge (m/z) ratio

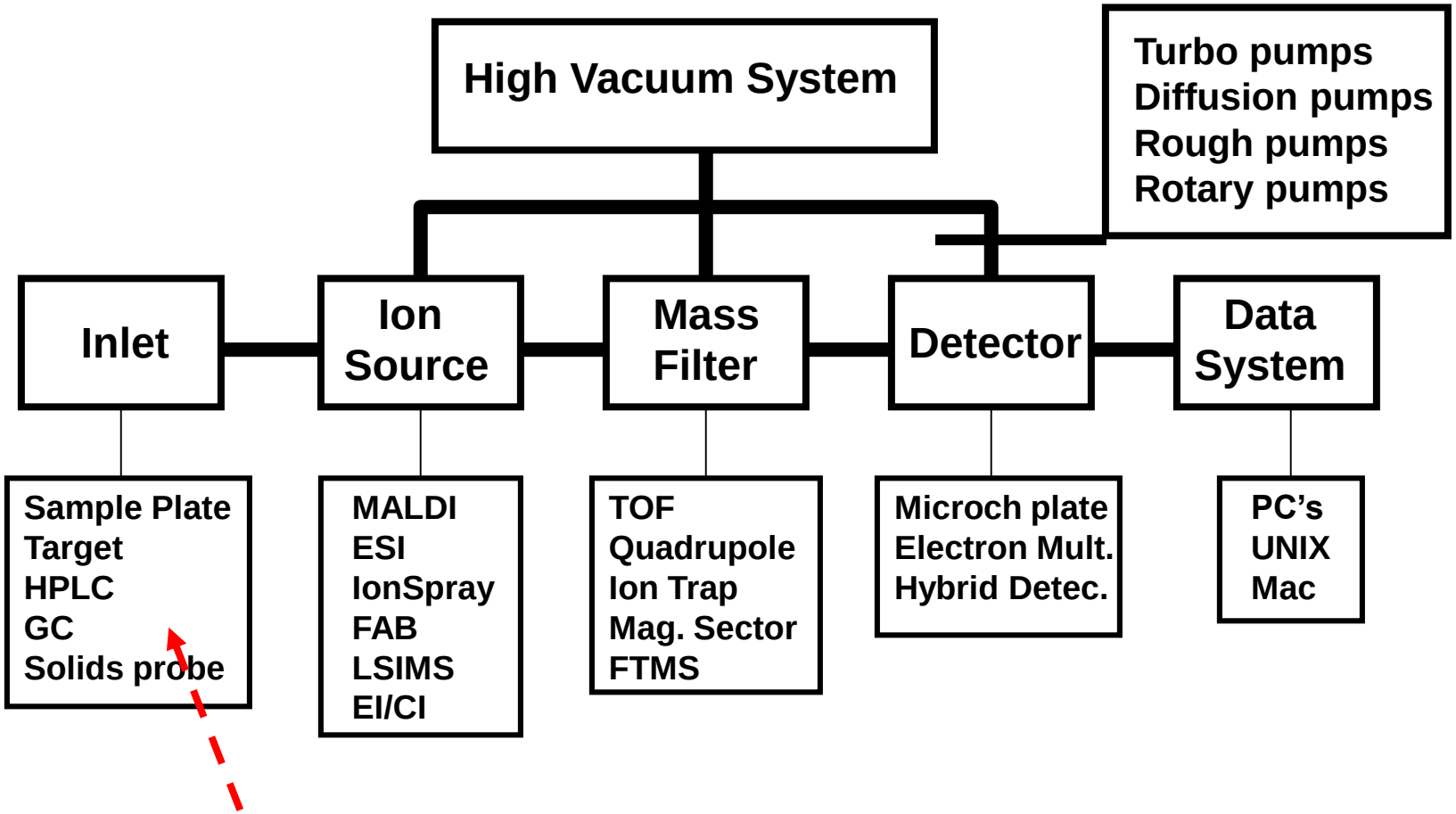
Detector

Ion detection

→ ***Mass spectrum***

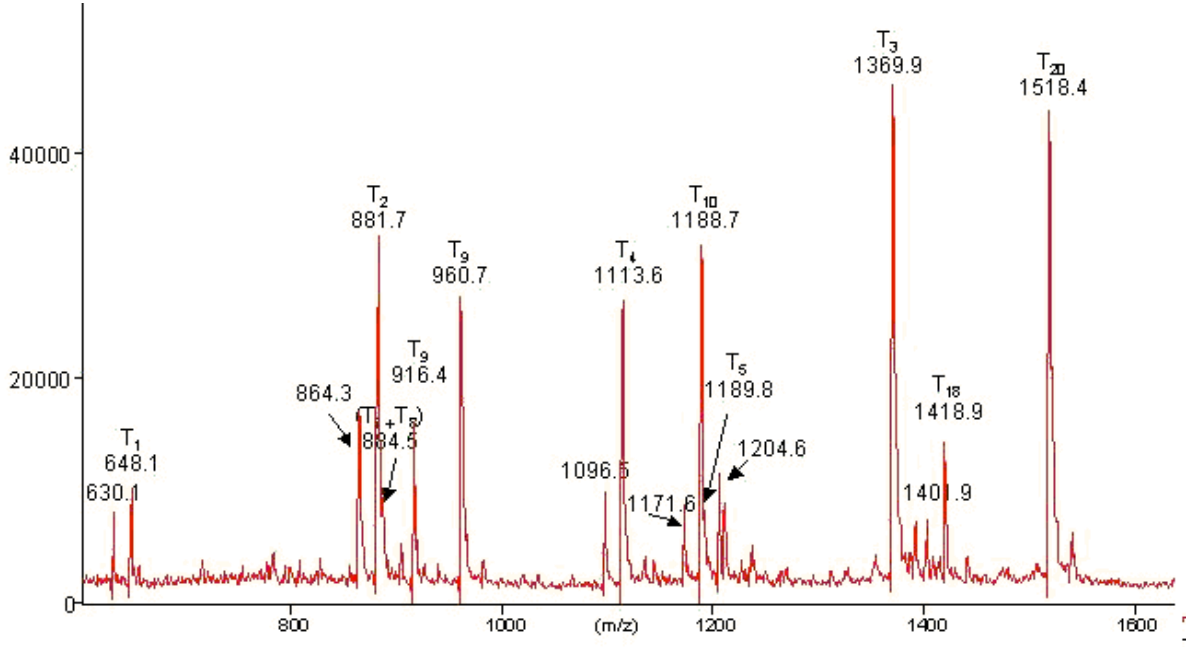
Mass spectrum: intensity *vs* m/z ratio plot representing a chemical analysis

Mass spectrometer scheme



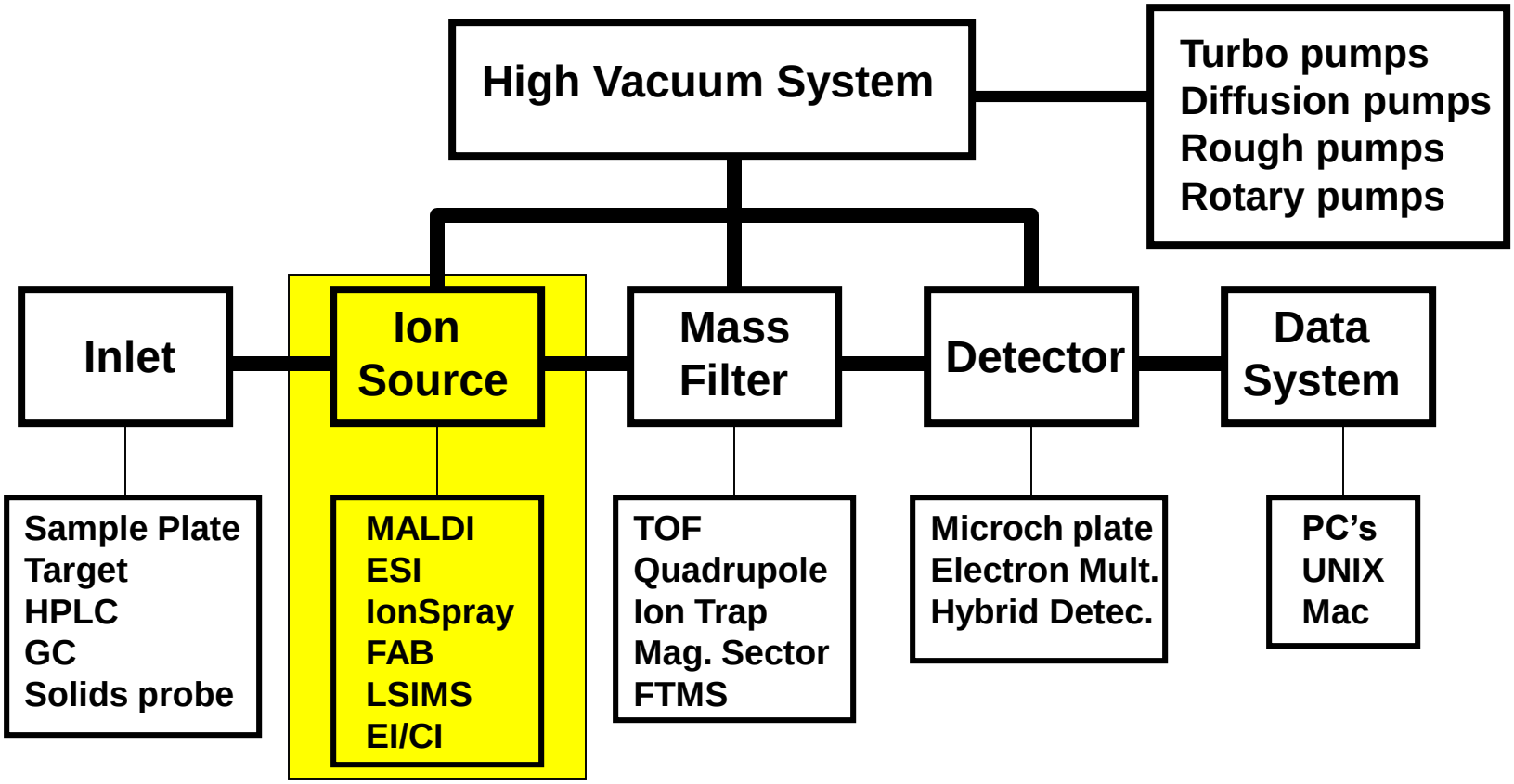
Typical Mass Spectrum

- Characterized by sharp, narrow peaks
- X-axis position indicates the m/z ratio of a given ion (for singly charged ions this corresponds to the mass of the ion)
- Height of peak indicates the relative abundance of a given ion (not reliable for quantitation)
- Peak intensity indicates the ion's ability to desorb or "fly"



Maldi Tof mass spectrum of the tryptic digest of a high molecular weight (85.000) glutenin (DY10)

Mass Spectrometer Schematic



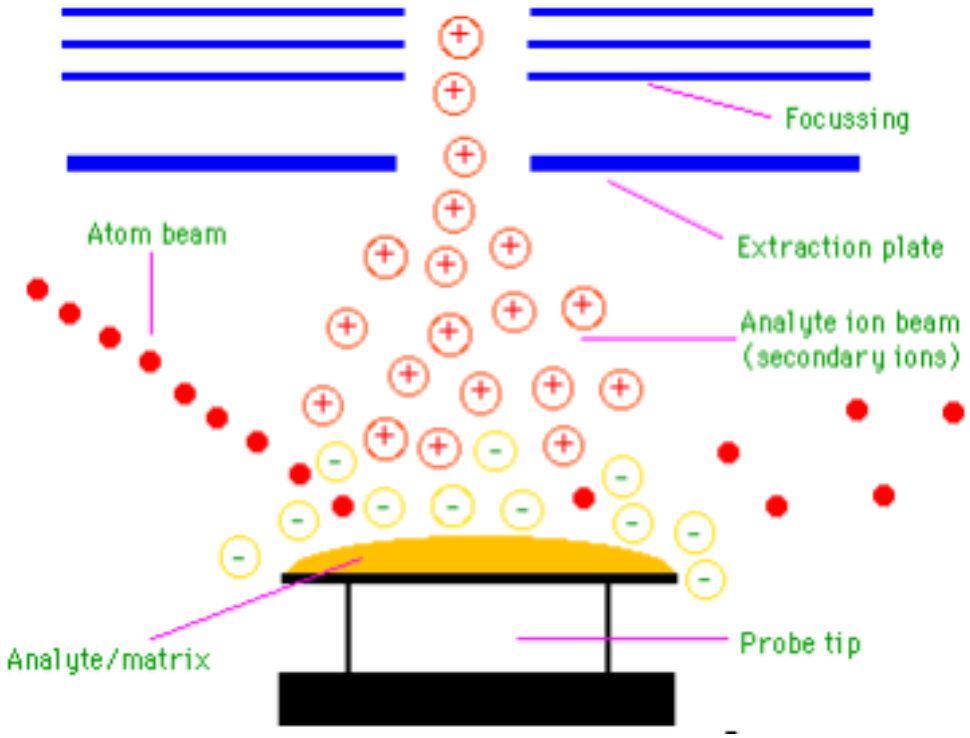
Different Ionization Methods

- Electron Impact (EI - Hard method)
  - small molecules, 1-1000 Daltons, structure
- Fast Atom Bombardment (FAB – Semi-hard)
  - peptides, sugars, up to 6000 Daltons
- Electrospray Ionization (ESI - Soft) → 1980's
  - peptides, proteins, up to 200,000 Daltons
- Matrix Assisted Laser Desorption (**MALDI**-Soft) → 1985
  - peptides, proteins, DNA, up to 500 kD

✓ **Soft ionization techniques keep the molecule of interest fully intact**

✓ Made it possible to analyze large molecules via inexpensive mass analyzers such as quadrupole, ion trap and Time-of-Flight (TOF)

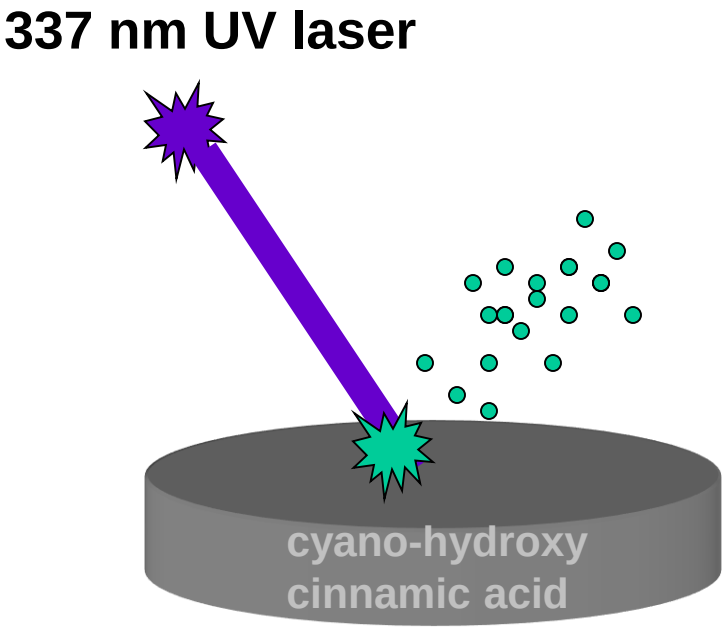
FAB - Fast Atom Bombardment



The sample is mixed with a matrix (non-volatile chemical protection environment) and is bombarded under vacuum with a high energy (4000 to 10,000 electron volts) beam of atoms. The atoms are typically from an inert gas such as argon or xenon. Common matrices include glycerol, thioglycerol.

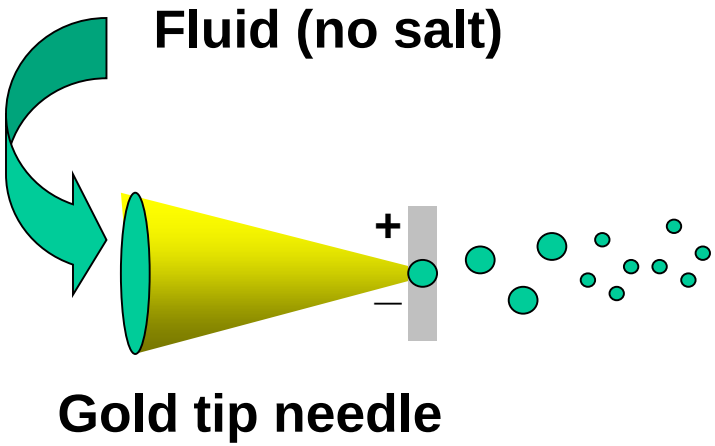
FAB is a relatively soft ionization technique and produces primarily intact protonated molecules denoted as [M+H]+ and deprotonated molecules such as [M-H]-. **Suitable for peptides, small proteins (up to 5000 Da**)

Soft Ionization Methods



MALDI

Matrix Assisted Laser Desorption

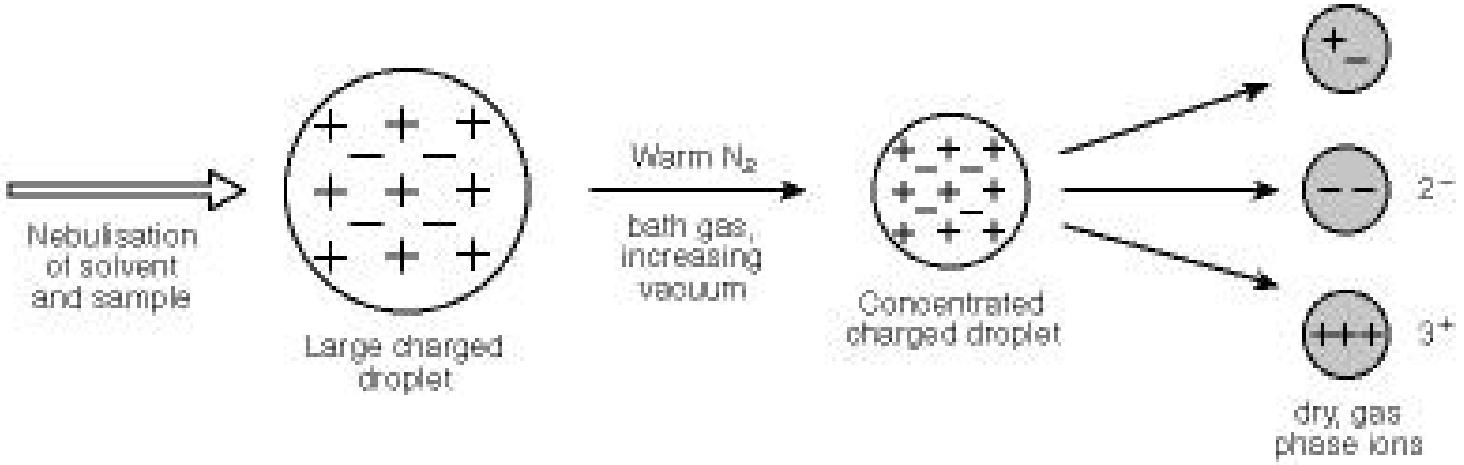


ESI

Electrospray Ionization

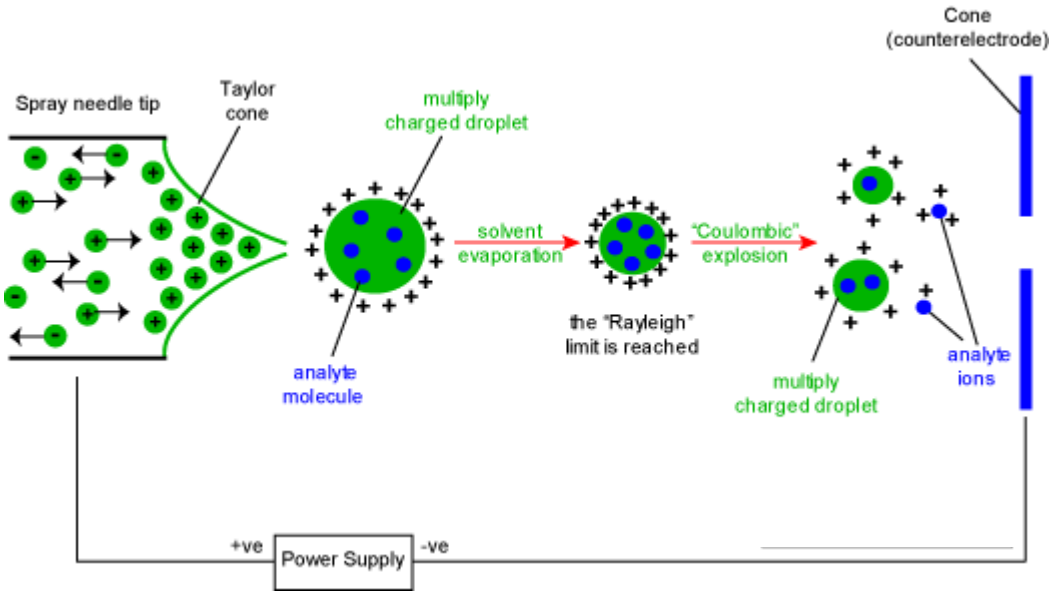
ESI

The ESI source operates at atmospheric pressure. A sample solution is sprayed from a small tube into a strong electric field in the presence of a flow of warm nitrogen to assist desolvation. The droplets formed evaporate in a region maintained at a vacuum of several torr causing the **charge to increase on the droplets**. The multiply charged ions then enter the analyzer.



ions carry multiple charges, which reduces their mass-to-charge ratio compared to a singly charged species. This allows mass spectra to be obtained for large molecules

ESI suitable for combination with HPLC and CE (Capillary Electrophoresis)

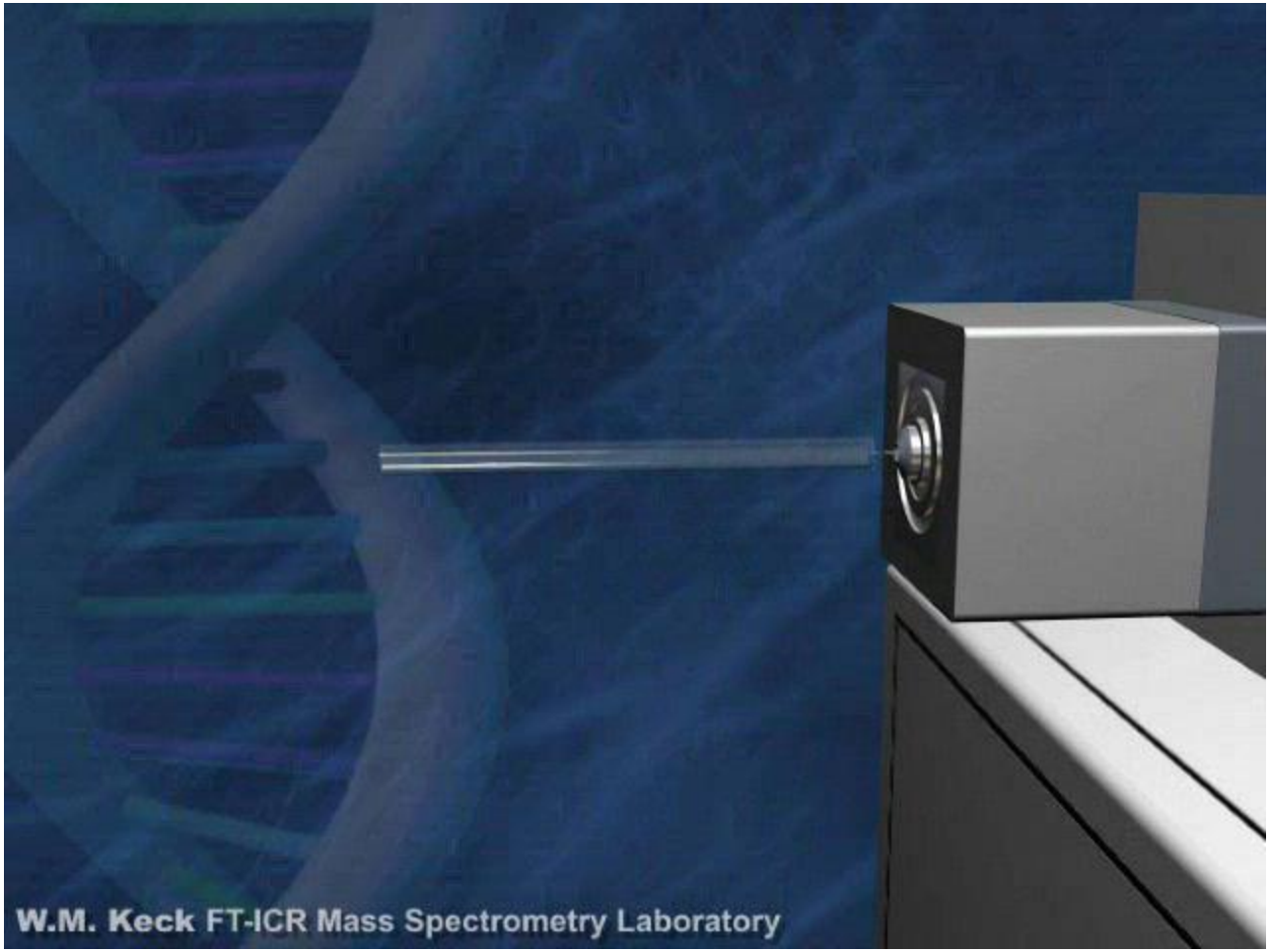


When the charge density at the droplet surface reaches a critical value (the Rayleigh limit), a so called **Coulombic explosion** occurs and several even smaller droplets are formed, each carrying some fraction of the original droplets surface charge.

Solvent evaporation
Droplet contraction
Coulombic explosions

} repeated until the molecular adducts are released from the final droplet.

If a positive voltage is applied to the capillary, then the droplets will carry positive charges and finally positive ions are formed, such as $[M+H]^+$ and $[M+Na]^+$ adducts. In the negative-ion mode, the base peak is typically the $[M-H]^-$ ion.


W.M. Keck FT-ICR Mass Spectrometry Laboratory

Electrospray Ionization

- 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)
- Negative ion mode measures $(M - H)^-$ (add ammonia to solvent)

Positive or Negative Ion Mode?

- If the sample has functional groups that readily accept H+ (such as amide and amino groups found in peptides and proteins) then positive ion detection is used –

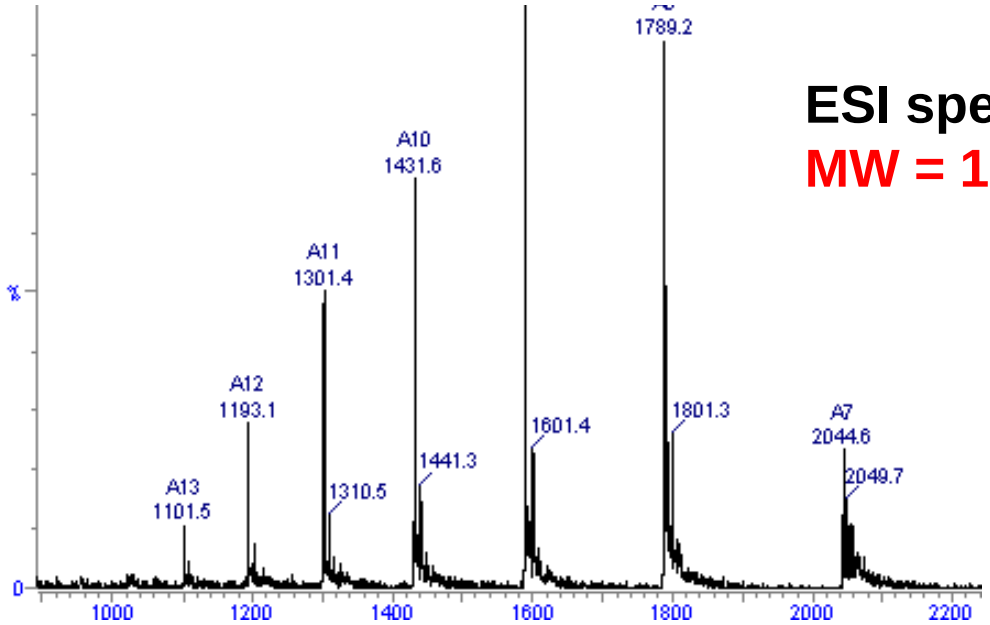
  For positive-ion mode, 0.1% formic acid or acetic acid is usually added into the analyte solution to enhance protonation and increase sensitivity-

- If a sample has functional groups that readily lose a proton (such as carboxylic acids and hydroxyls as found in nucleic acids and sugars) then negative ion detection is used

  For negative-ion mode, 0.3% NH_4OH is usually added into the analyte solution to help deprotonation and increase sensitivity -

Multiply Charged Ions

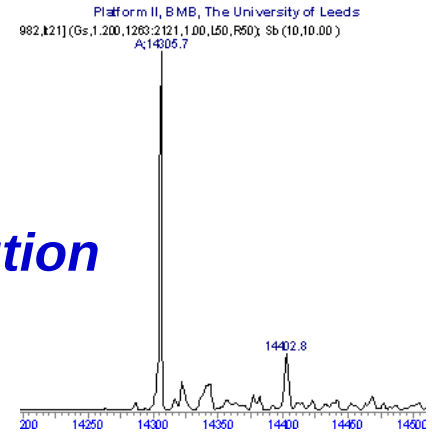
- Samples of MW up to 1200 Da usually produce singly charged ions (parent mass + H)
- Larger samples (typically peptides) yield ions with multiple charges (from 2 to 20 +)
- Multiply charged species form a Gaussian distribution with those having the most charges showing up at lower m/z values



ESI spectrum of Lysozyme

MW = 14,305.14

deconvolution



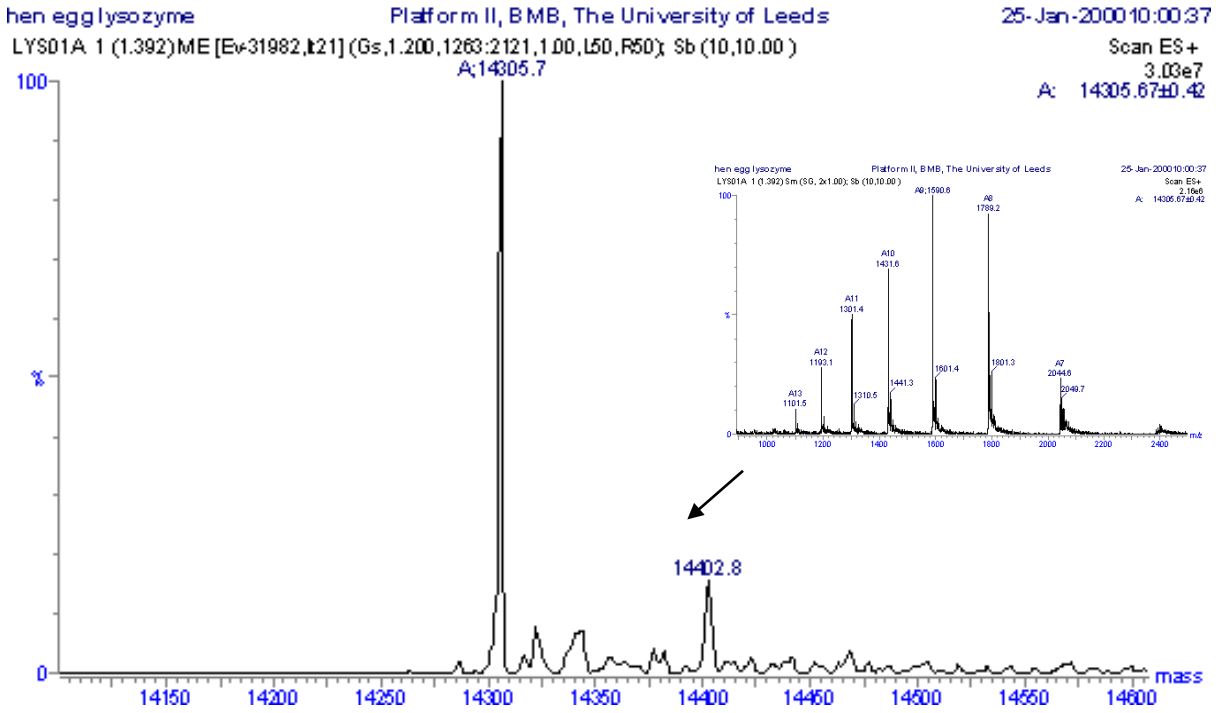
ESI Transformation

- Software can be used to convert these multiplet spectra into single (zero charge) profiles which gives MW directly
- This makes MS interpretation much easier and it greatly increases signal to noise

Two methods are available

- Transformation (requires prior peak ID)
- Maximum Entropy (no peak ID required)

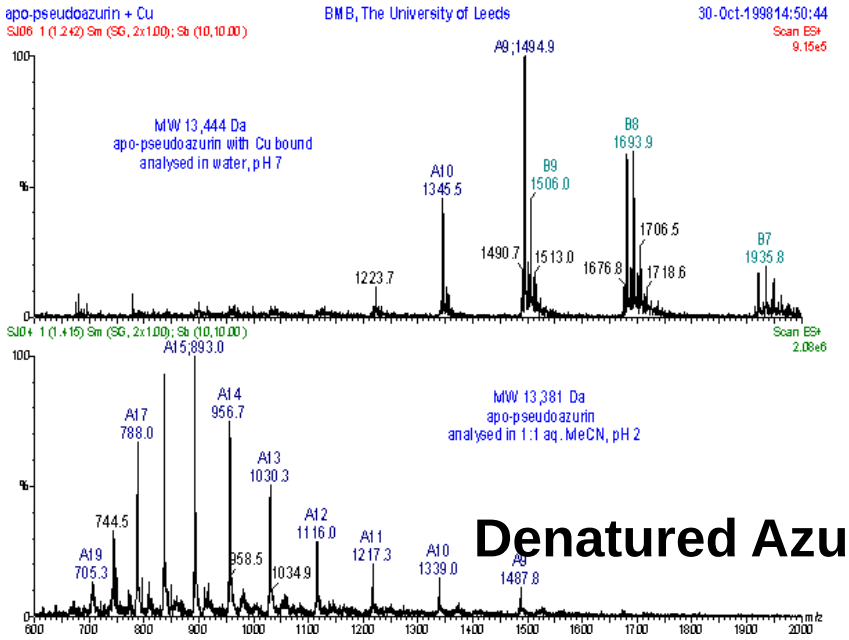
Maximum Entropy



ESI and Protein Structure

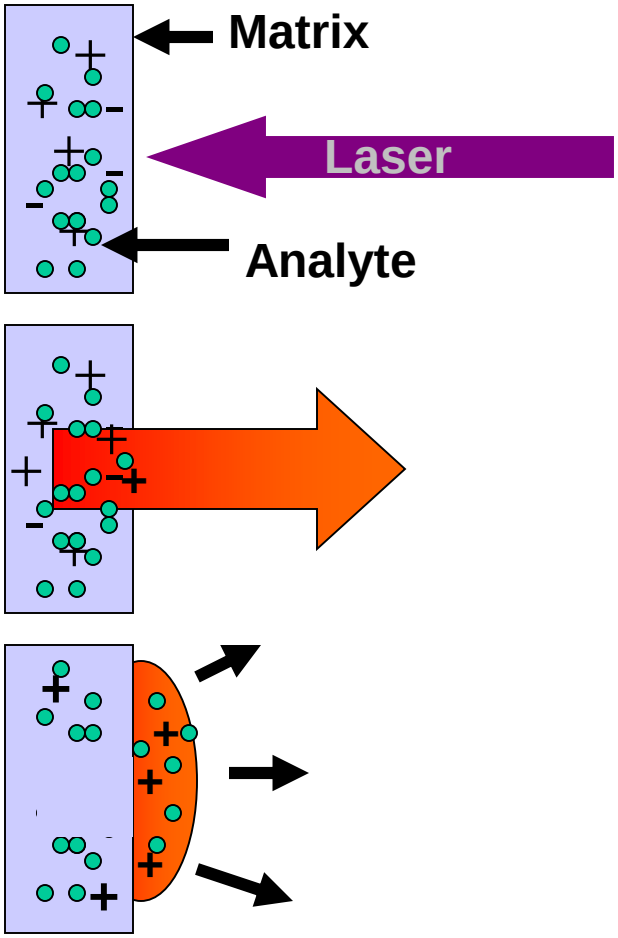
- ESI spectra are actually quite sensitive to the **conformation** of the protein
- Folded, ligated or complexed proteins tend to display **non-gaussian peak distributions**, with few observable peaks weighted toward higher m/z values
- Denatured or open form proteins/peptides which ionize easier tend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

Native Azurin



Denatured Azurin

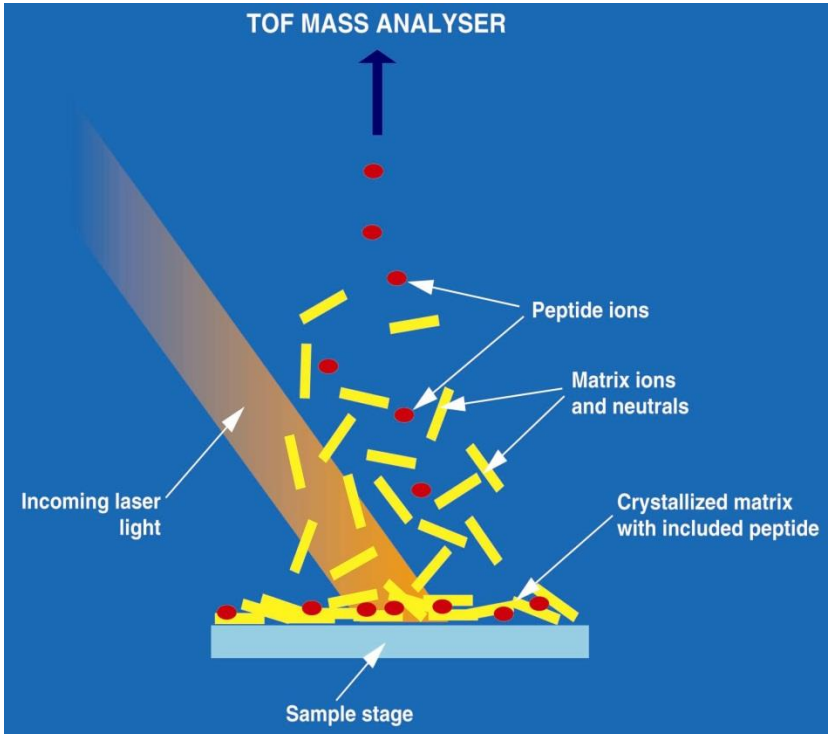
MALDI Ionization

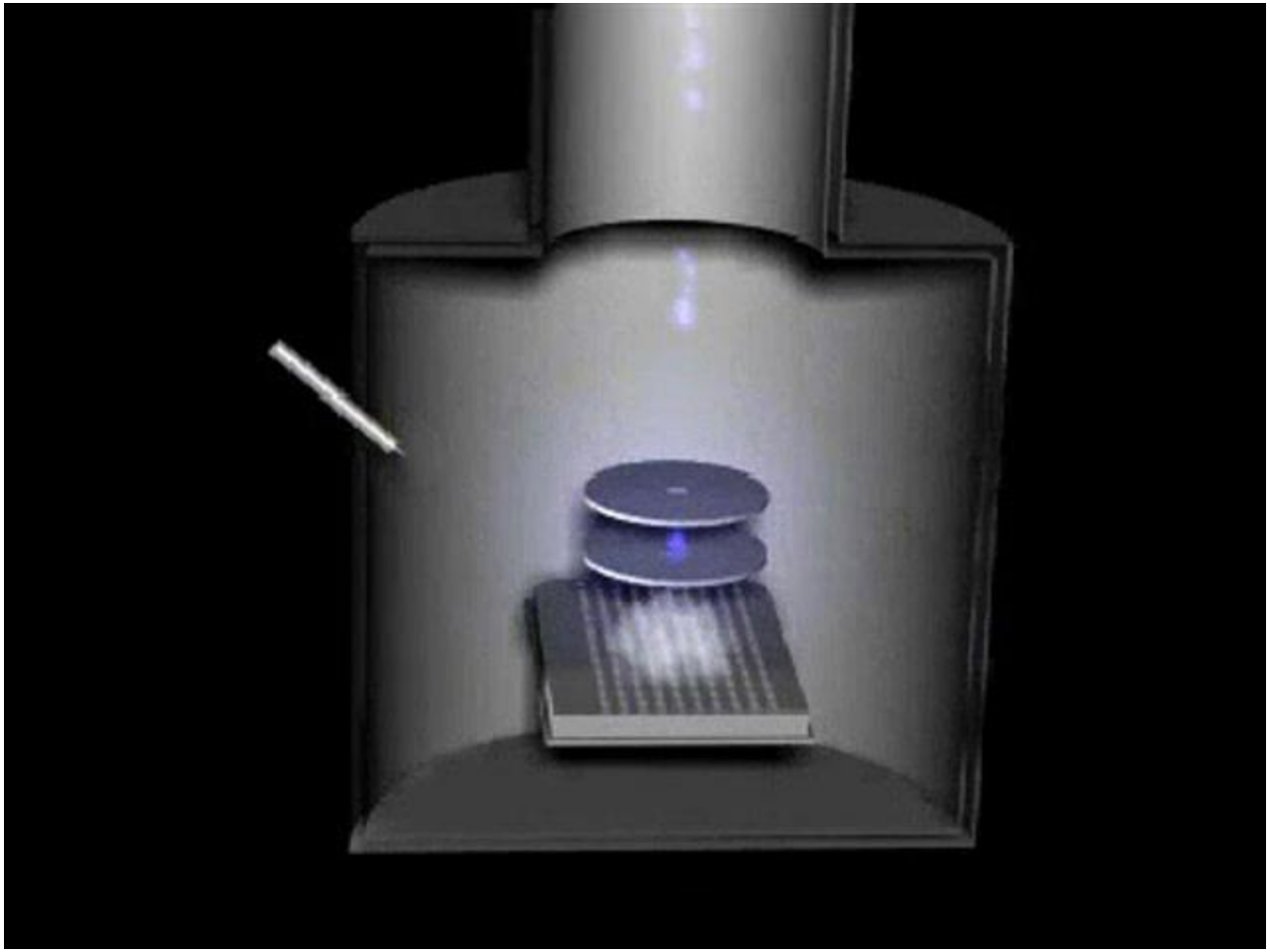


- Absorption of UV radiation by chromophoric matrix and ionization of matrix
- Dissociation of matrix, phase change to super-compressed gas, charge transfer to analyte molecule
- Expansion of matrix at supersonic velocity, analyte trapped in expanding matrix plume (explosion/”popping”)

MALDI

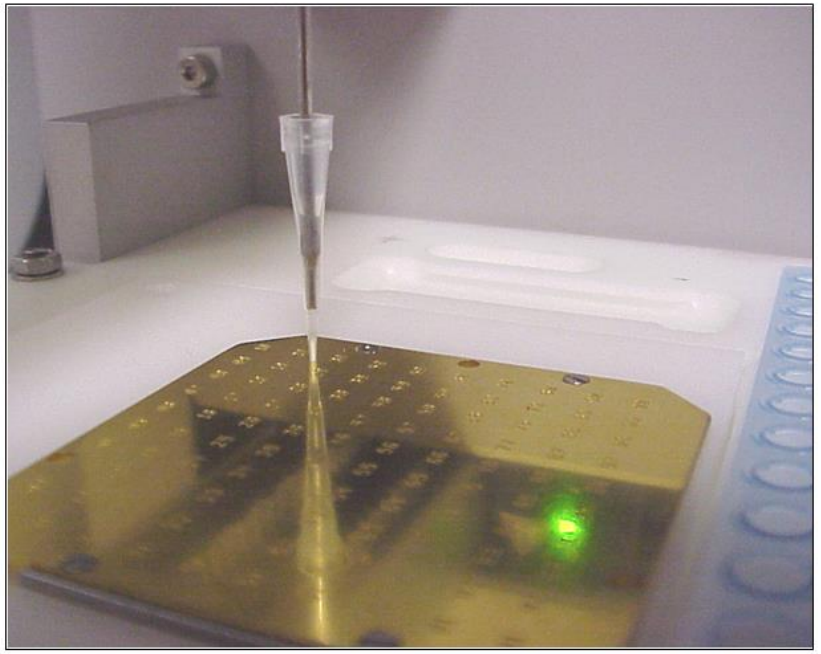
- Unlike ESI, MALDI generates spectra that have just a singly charged ion
- Positive mode generates ions of M + H
- Negative mode generates ions of M - H
- Generally more robust that ESI (tolerates salts and nonvolatile components)
- Easier to use and maintain, capable of higher throughput
- Requires 10 μL of 1 pmol/μL sample





MALDI Sample Limits

- Phosphate buffer < 50 mM
- Ammonium bicarbonate < 30 mM
- Tris buffer < 100 mM
- Guanidine (chloride, sulfate) < 1 M
- Triton < 0.1%
- SDS < 0.01%
- Alkali metal salts < 1 M
- Glycerol < 1%



….MALDI or Electrospray ?

ESI is **better for the analysis of complex mixture** as it is directly interfaced to a separation technique (i.e. HPLC or CE)

MALDI is more “flexible” (MW from 200 to 400,000 Da)

MALDI is easier to use and maintain and more robust (allows analysis of bacteria, cells and tissues without purification)

	MALDI	ESI
Sensitivity =	femtomole 10^{-15} M (...attomole 10^{-18} M)	
Simplicity =	very easy	training required
$$$ =	70 to 650 k$	120 to 650 k$
Speed (high throughput)=	~10^4/day	dynamic system

MASS ACCURACY AND RESOLUTION

Mass accuracy in mass spectrometry is related to the **calibration of the mass analyzer** to properly assign the true mass-to-charge 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. In calibrating a mass spectrometer, a series of standard compounds, usually in the form of a multicomponent standard, are measured and related to the fundamental properties of the mass analyzer.

The mass accuracy in mass spectrometry is usually calculated as a parts-per-million (ppm) error where the theoretical mass (calculated as the monoisotopic mass*) is subtracted from the observed mass, divided by the observed mass and multiplied by a 10^6 factor. Mass accuracy (ppm error) can thus be represented as

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

*Monoisotopic Mass: the sum of the exact or accurate masses of the lightest stable isotope of the atoms in a molecule

Mass error

- Mass error = (exact mass) – (accurate mass)
- Mass error in parts per million (ppm) =

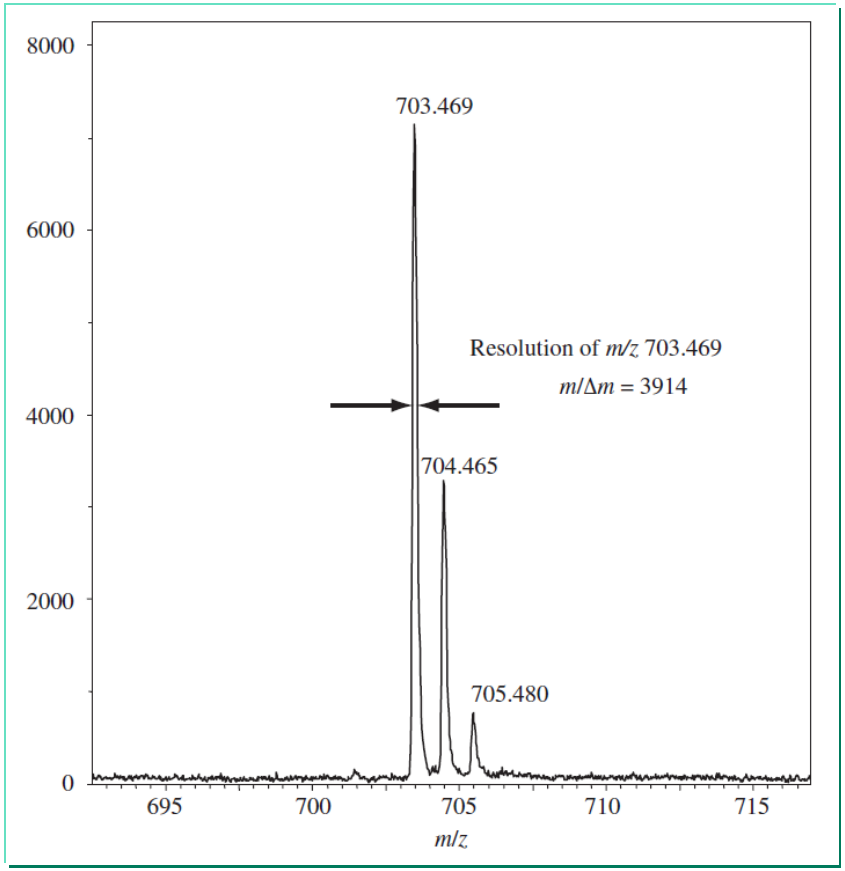
$$\frac{\text{(mass error)}}{\text{(exact mass)}} \times 10^{6}$$

MASS RESOLUTION

In mass spectrometry, mass resolution is generally calculated as ***m/Δm*** where m is the *m/z* value obtained from the spectrum and Δ*m* is the full peak width at half maximum (FWHM).

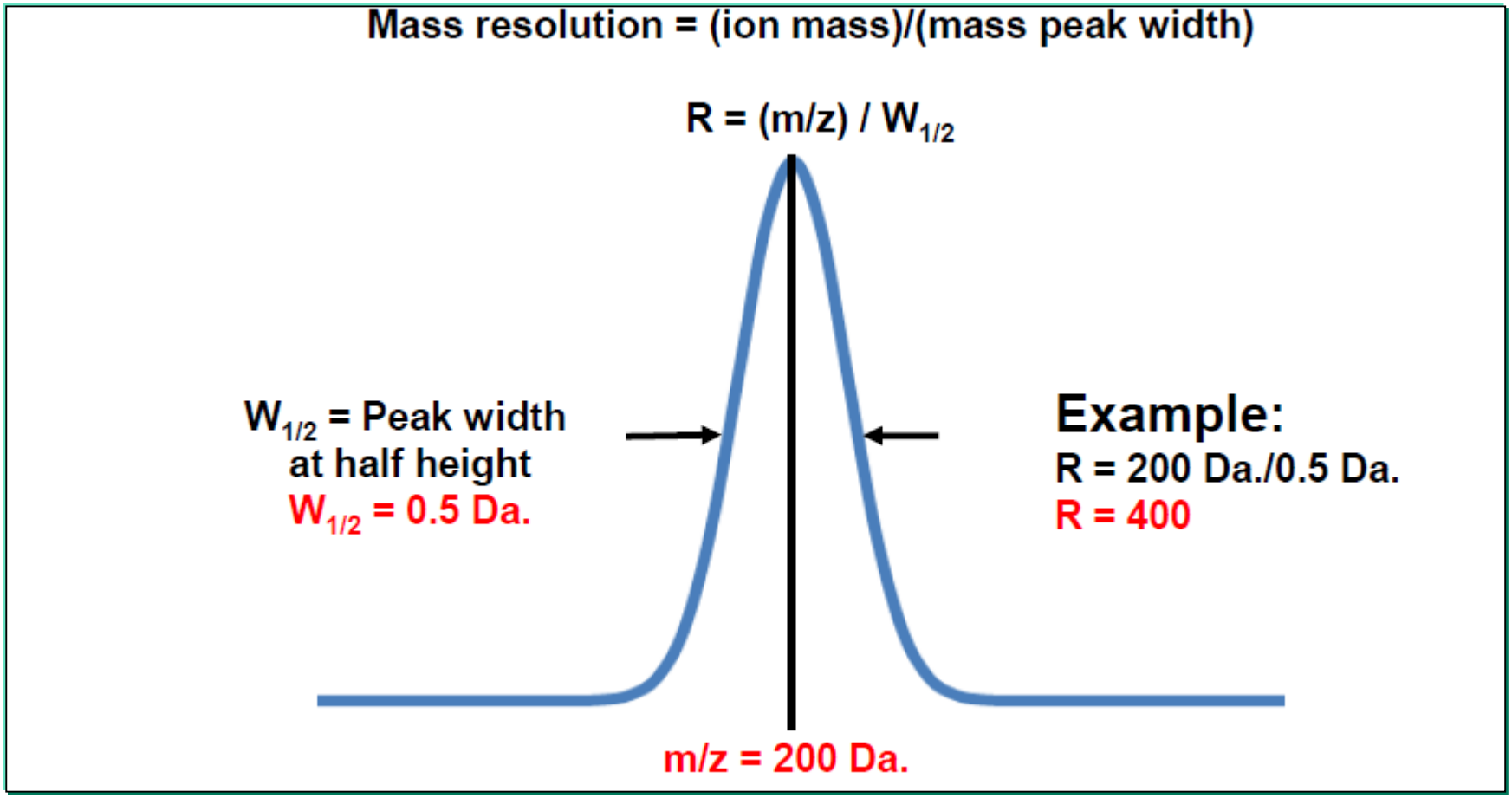
In the figure for the *m/z* 703.469 ion, which has a full peak width at half maximum of 0.1797 *m/z*, the resolution of the *m/z* 703.469 would be calculated as *m*/Δ*m* = 703.469/0.1797 = 3914.

The narrower and sharper a peak is will result in a higher-resolution value and a better estimate of the apex of the Gaussian-shaped peak, which thus results in a better estimate of the true value of the mass-to-charge ratio



Full width at half maximum (FWHM) calculation of the resolution of *m/z* 703.469 as *m*/Δ*m* = 3914.

Mass resolution



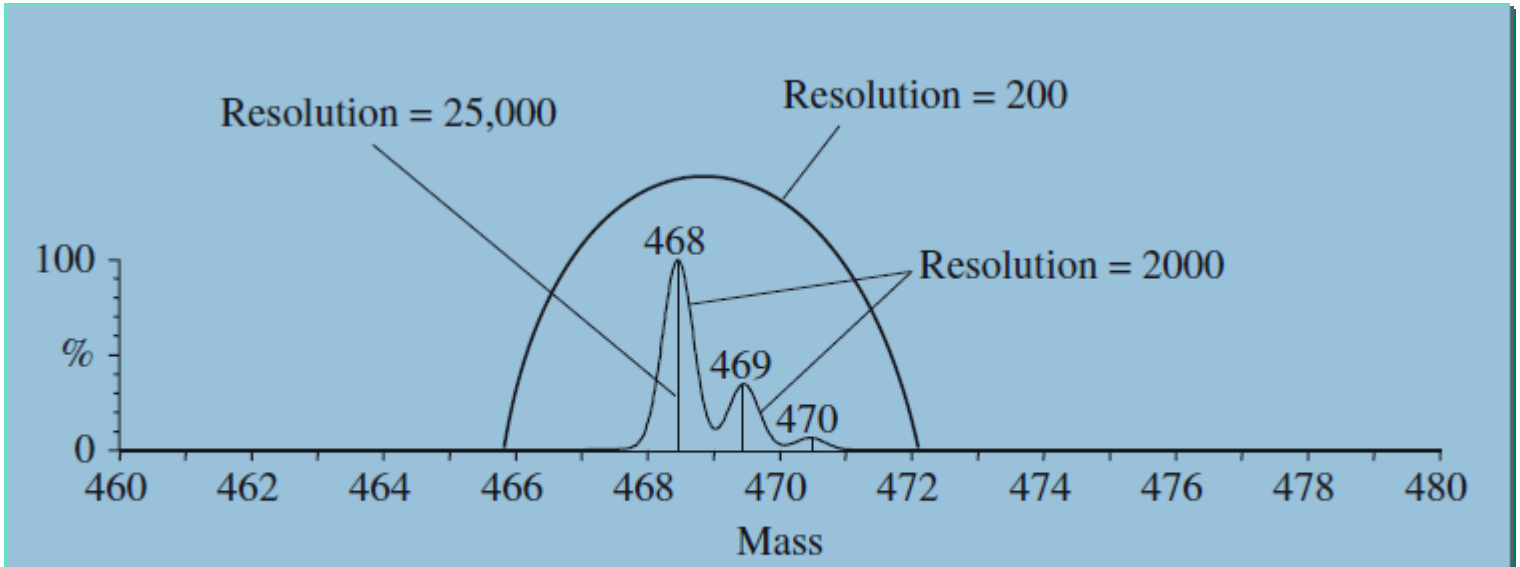


FIGURE 28.15 Effect of mass resolving power on the mass spectrum.

A mass resolution of approximately 200 for the overall curve where none of the isotopic peaks have been resolved. A resolution of 200 will give an average mass value for the species being measured with a high degree of error for the exact mass (>500 ppm).
The three isotope peaks are resolved at a resolution of 2000, and an intermediate accurate exact mass is obtainable.

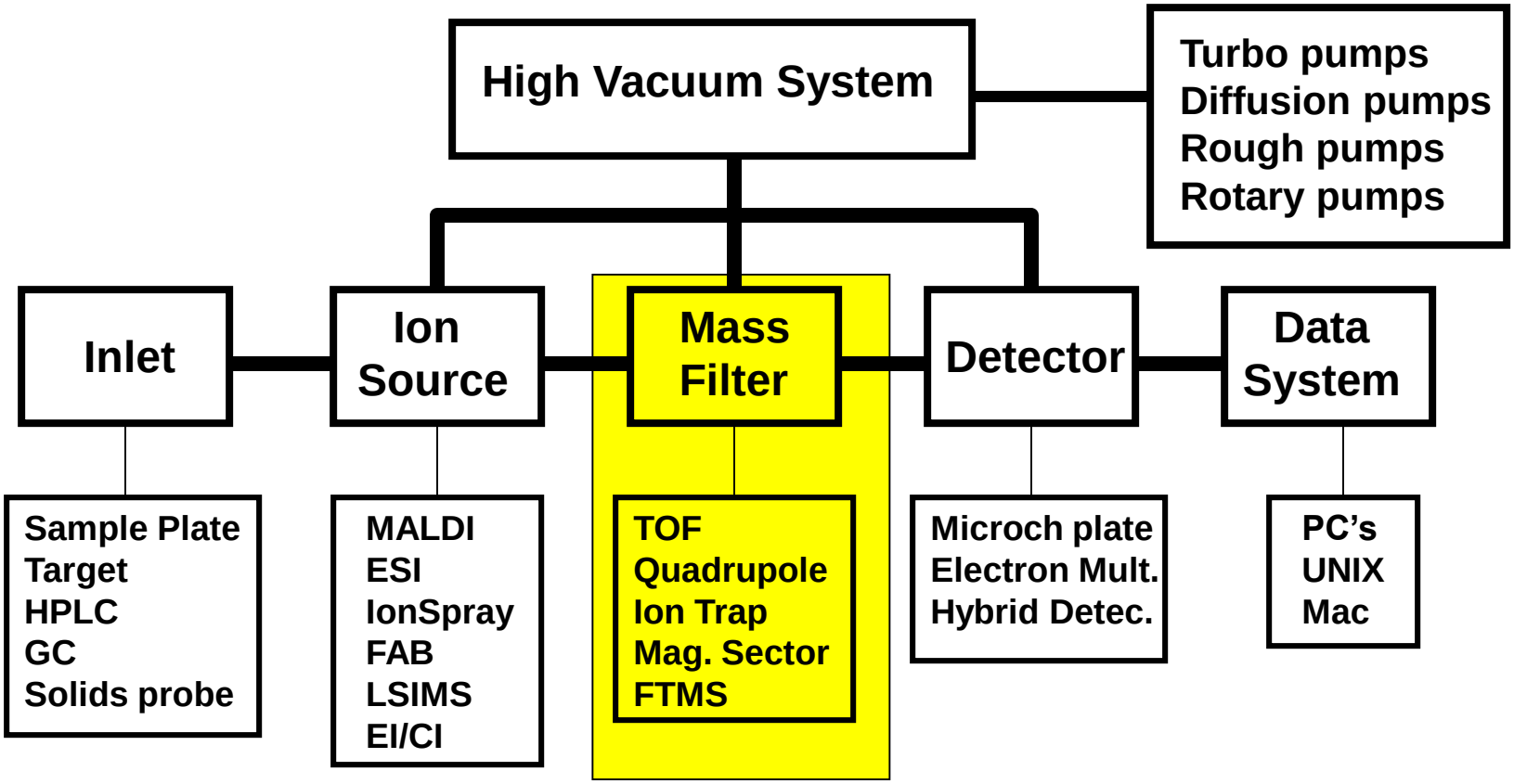
At a resolution of 25,000, the three isotopic peaks are completely baseline resolved, and a mass accuracy typically less than 5 ppm can be obtained

MALDI resolution issues

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

1. 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 (placement in crystal structure may attribute to this).
2. Not all of the analytes may have exactly the same velocity at the start of their flight toward the detector (*see slides on TOF, REFLECTRON, improvements from* $m/\Delta m$ = 500 to $m/\Delta m$=10,000 or higher).

Mass Spectrometer Schematic



Mass analyzers

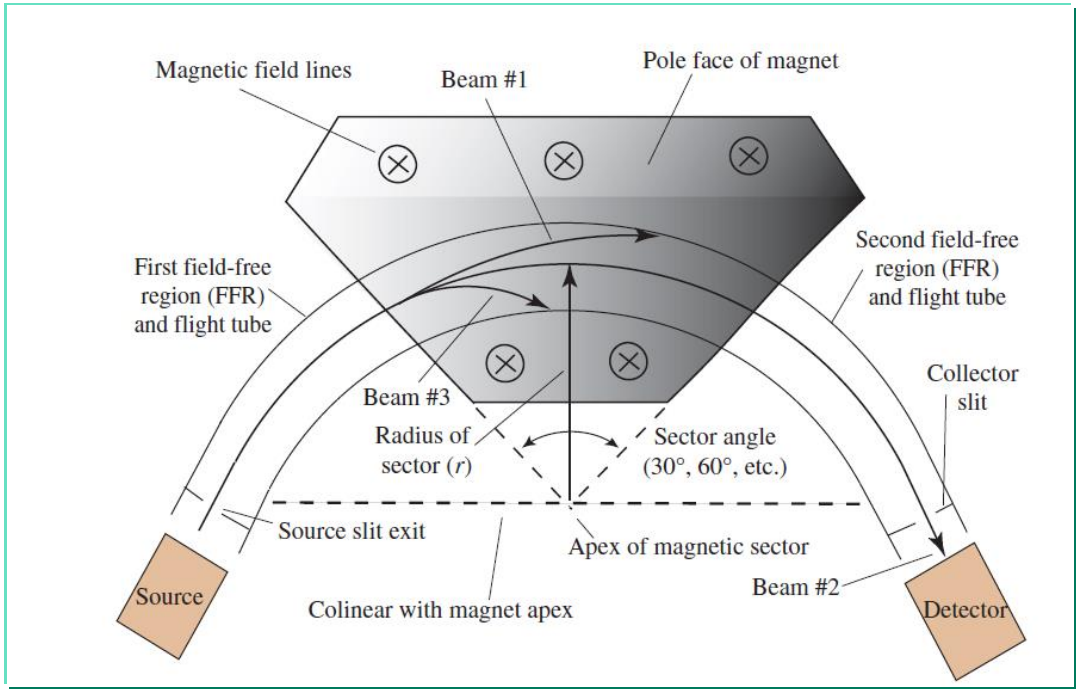
The mass analyzer is the heart of the mass spectrometric instrumentation used in the separation of molecular ions (M+) and analyte ions (e.g., [M+H]+) in the gas phase.
The most fundamental aspect of the mass analyzer is the ability to separate ions according to their mass-to-charge (*m/z*) ratio

Mass Analyzers

- Magnetic Sector Analyzer (MSA)
  - High resolution, exact mass, original MA
- Quadrupole Analyzer (Q)
  - Low resolution, fast, cheap
- Time-of-Flight Analyzer (TOF)
  - No upper m/z limit, high throughput
- ORBITRAP
  - high sensitivity, resolution and mass accuracy
- Ion Trap Mass Analyzer (QSTAR)
  - Good resolution, all-in-one mass analyzer
- Ion Cyclotron Resonance (FT-ICR)
  - Highest resolution, exact mass, costly

Magnetic sector mass analyzers

When charged particles enter a magnetic field, they will possess a circular orbit that is perpendicular to the poles of the magnet



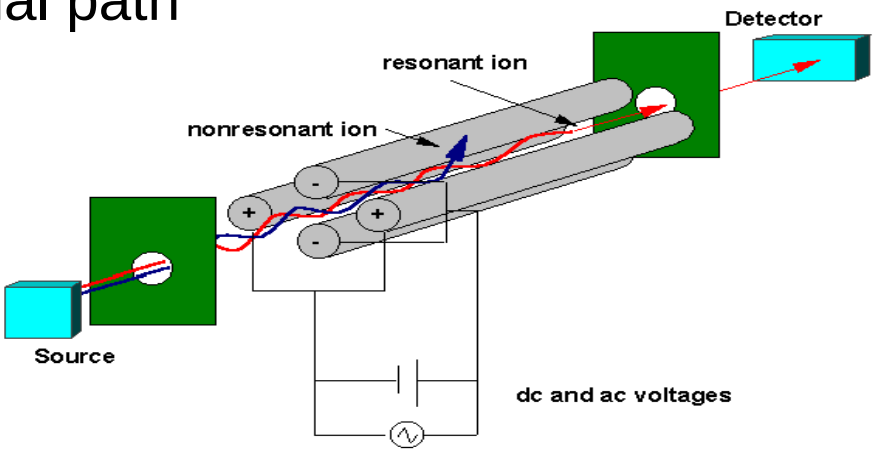
The ions enter a flight tube (first field-free region)
from a source through a source exit slit and travel into the magnetic field. The accelerating voltage in the source will determine the kinetic energy (KE) that is imparted to the ions:

$$\mathrm{KE} = zeV = \frac{1}{2}mv^2$$

V is the accelerating voltage in the source, *e* is the fundamental charge of an electron (1.60×10^{-19} C), *m* is the mass of the ion, *v* is the velocity of the ion, and *z* is the number of charges. The magnetic field will deflect the charged particles according to the radius of curvature of the flight path (*r*) that is directly proportional to *m/z* of the ion

Quadrupole Mass Analyzer

- A quadrupole mass filter consists of four parallel metal rods with different charges
- Two opposite rods have an applied potential of (U+Vcos(wt)) and the other two rods have a potential of -(U+Vcos(wt))
- The applied voltages affect the trajectory of ions traveling down the flight path
- For given dc and ac 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



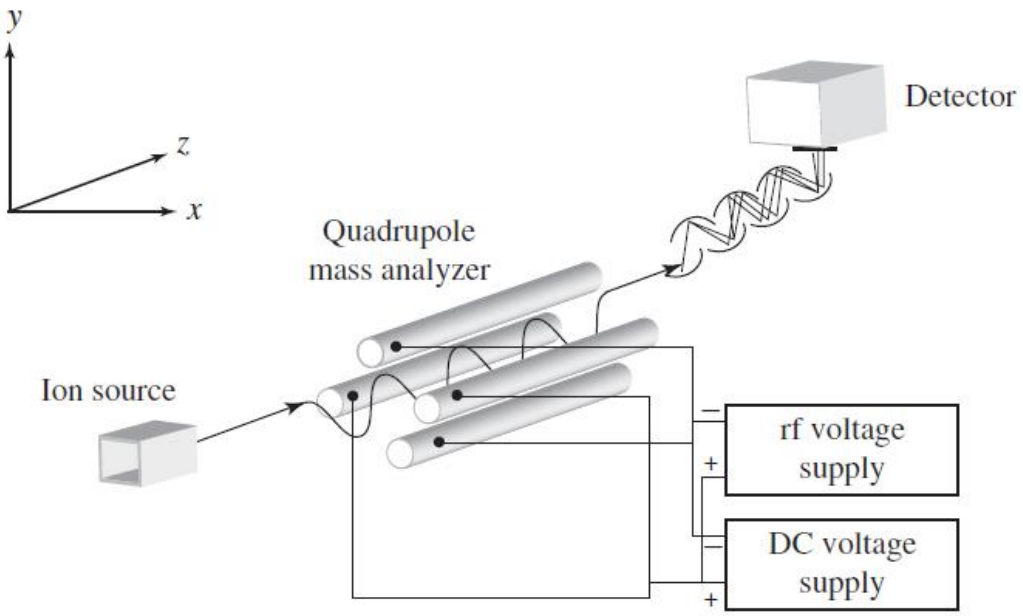
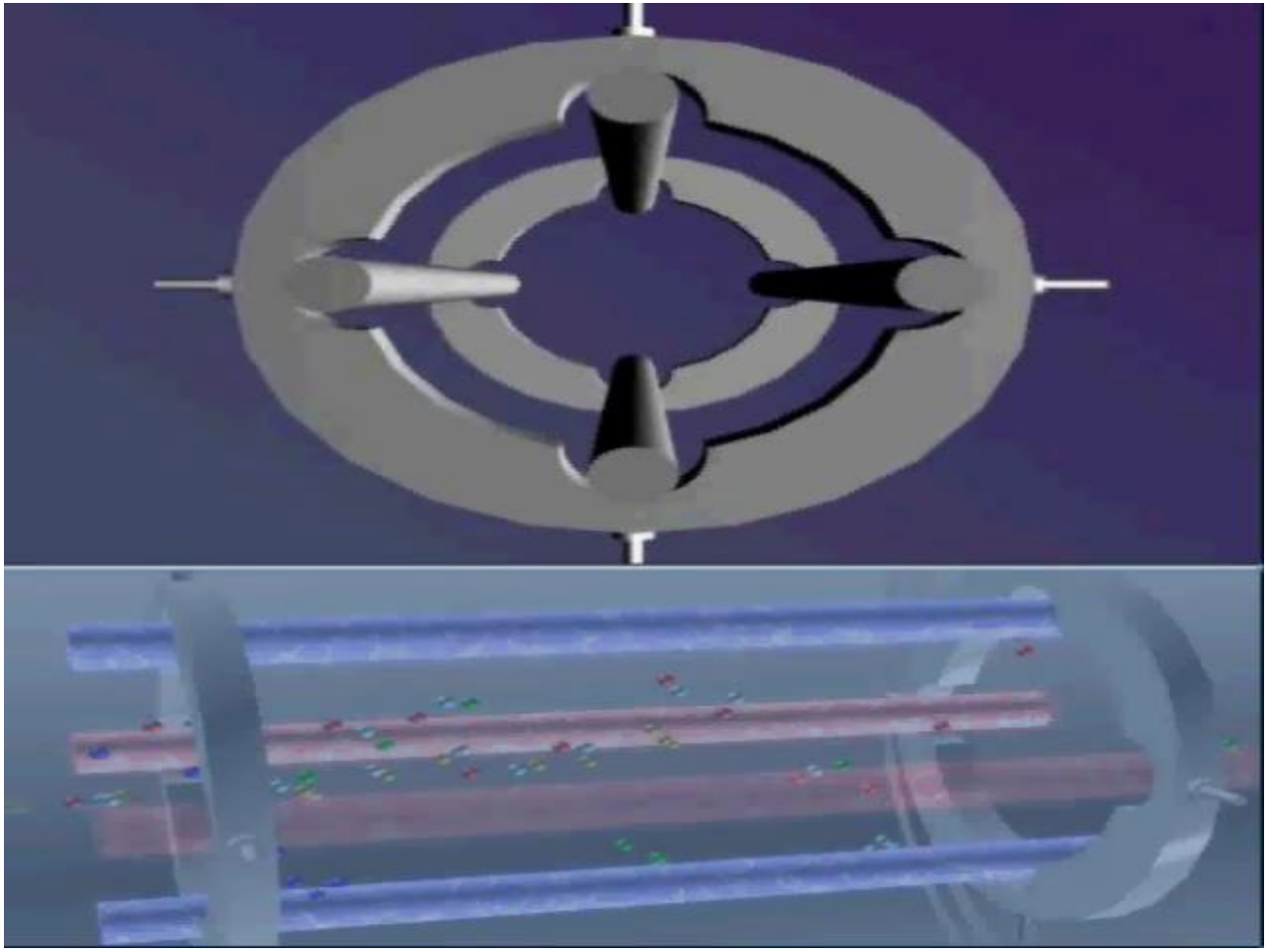


FIGURE 30.11 Quadrupole orientation and the configuration for the connections of the DC voltage (U) and radio frequency (rf) voltage (V). Ions are accelerated into the quadrupole by a small voltage of 5 eV, and under the influence of the combination of electric fields, the ions follow a complicated trajectory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

Quadrupole Mass Analyzer



Quadrupole Mass Analyzer

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. If the oscillation of the ions in the quadrupole has finite amplitude, it will be stable and pass through. If the oscillations are infinite, they will be unstable and the ion will collide with the rods.

The DC and rf voltages have been selected to give an *m/z* value of 100 a stable trajectory through the quadrupoles. An *m/z* value of 10, which is a less massive ion, will have a very unstable trajectory and will collide with the quadrupole rods at an early stage. An *m/z* 1000 species will tend to travel further through the quadrupole field but still possesses an unstable trajectory and will also suffer collision with the rods and thus also be effectively filtered out.

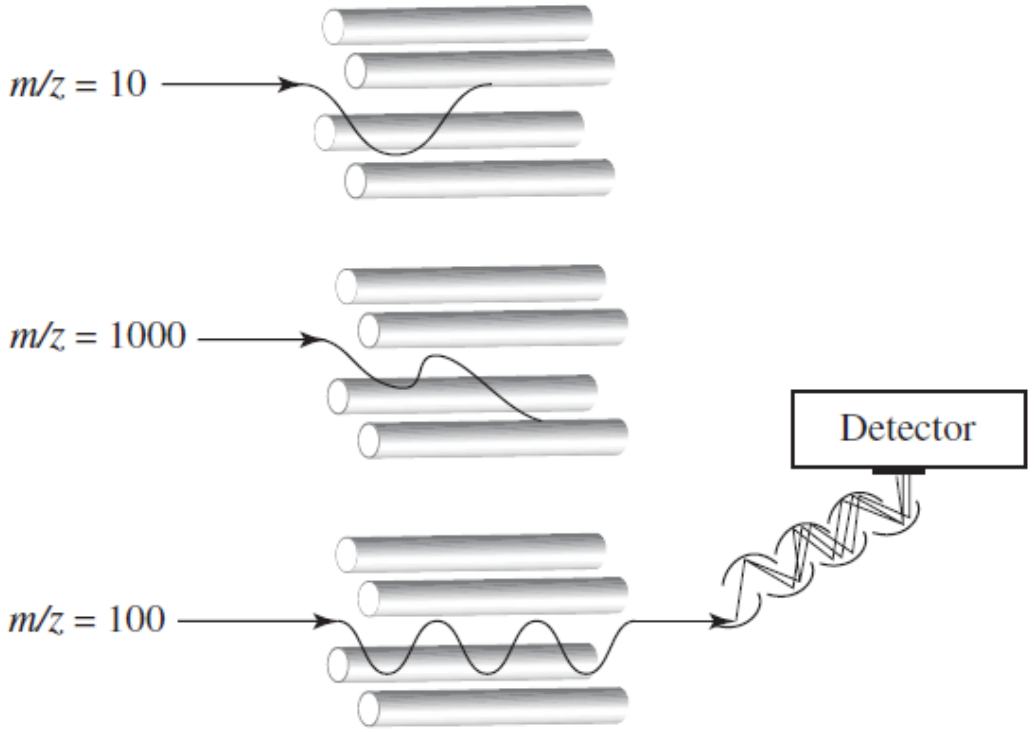


FIGURE 30.12 Stable and unstable trajectories of ions through the quadrupole. The *m/z* 100 species has been selected for stable path and transmission through the quadrupole for detection.

ORBITRAP

Ions are trapped in an electrostatic field between an inner and outer electrode (Makarov, 2000).

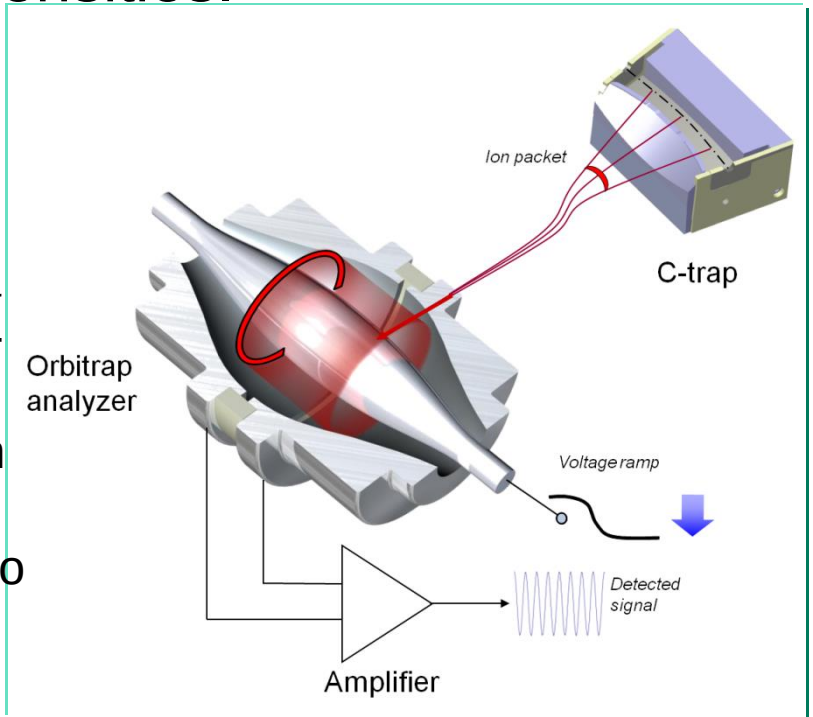
As the ions rotate around the inner electrode, they precess along its axis with a frequency characteristic of their mass-tocharge (m/z) ratio.

Acquisition of transients and the Fourier transformation of that signal yields frequencies and their intensities.

A simple relationship converts

frequencies into m/z values.

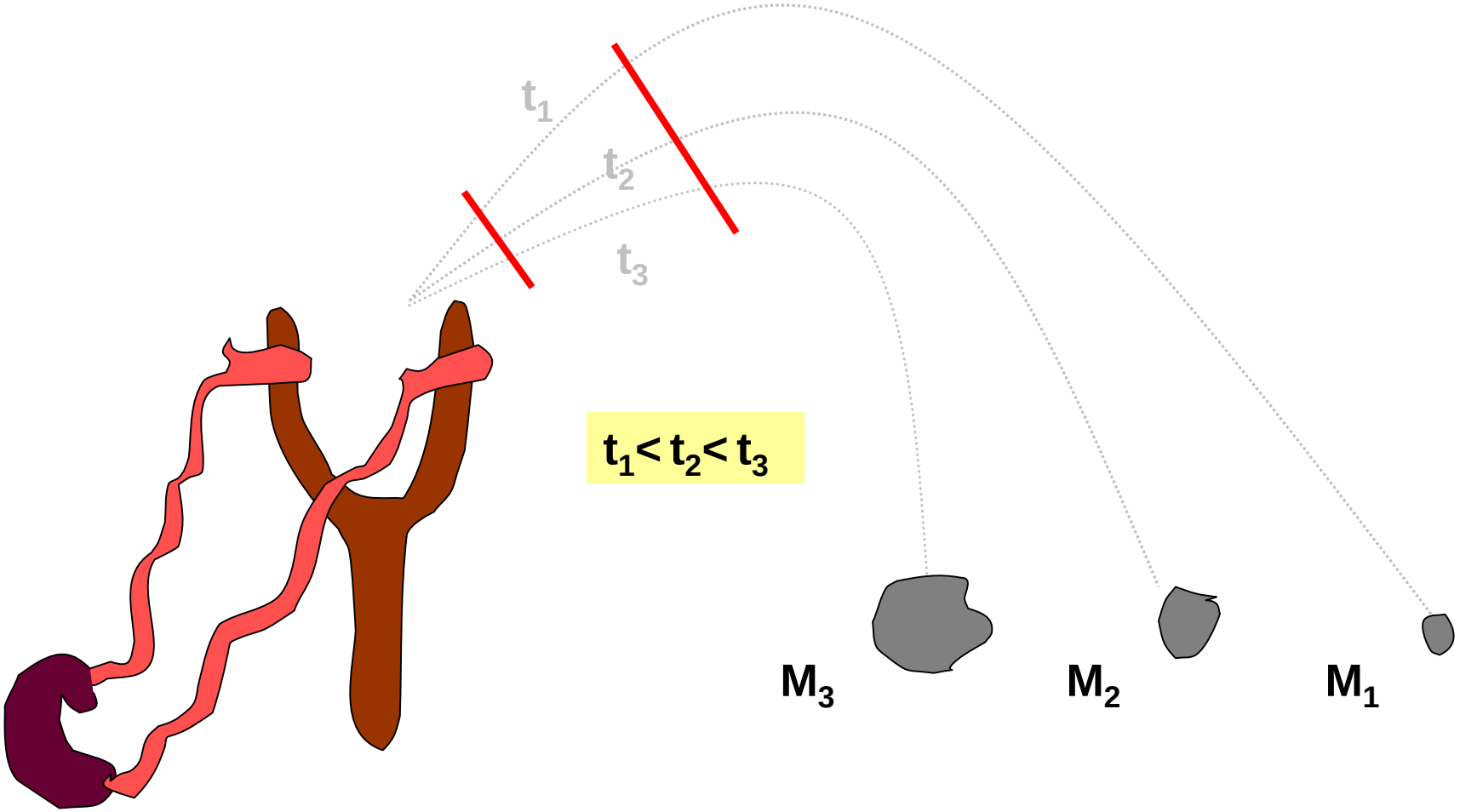
It features a pair of axially symmetric electrodes: a central spindle-like electrode and an outer «barrel-like electrode». In this electric field ions rotate around the central electrode while oscillating down the length of the electrode.
The frequency of such oscillations is proportional to $(m/z)^{-1/2}$



TOF

The most common mass spectrometer that is coupled to the MALDI ionization technique is the time-of-flight (TOF) mass spectrometer (TOF/MS). The TOF mass spectrometer separates compounds according to their mass-to-charge (*m/z*) 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 (*m/z*) ratio.

Principle of Time-of-Flight (TOF)



$E_k = 1/2\ mv^2 = zeEs$

$t = L/v$

The same accelerating potential (E_k)applied to different masses gives different initial velocities and different travelling times

Initially, all the ions have similar KE imparted to them from the draw-out pulse (representing time zero), which accelerates them into the flight tube. Because the compounds have different masses, their velocities will be different according to the relationship between KE and mass represented by KE = $zeV = 1/2mv_2$. From this expression, the mass-to-charge ratio is related to the ion's flight time by the following expression: $m/z = 2eVt_2/L_2$:

$$\mathrm{KE} = zeV = \frac{1}{2}mv^2$$

$$v = \left(\frac{2zeV}{m}\right)^{\frac{1}{2}}$$

$$t = \frac{L}{v}, \quad L = \text{length of drift tube}$$

$$t = L\left(\frac{m}{2zeV}\right)^{\frac{1}{2}}, \quad V \text{ and } L \text{ are fixed}$$

Solving for m/z,

$$t^2 = L^2\left(\frac{m}{2zeV}\right)$$

$$\frac{m}{z} = \frac{2eVt^2}{L^2}$$

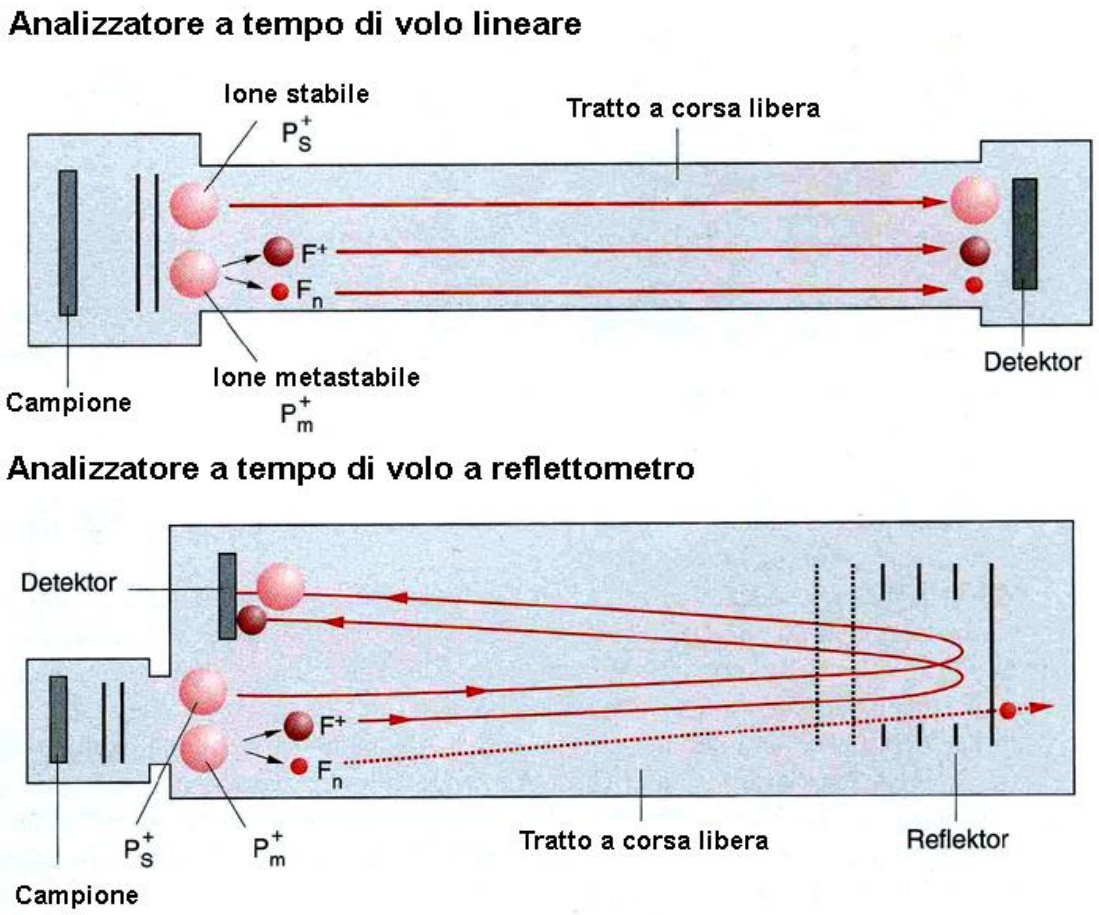
Mass Spec Equation (TOF)

$$\frac{m}{z} = \frac{2Vt^2}{L^2}$$

m = mass of ion
z = charge of ion
V = voltage
L = drift tube length
t = time of travel

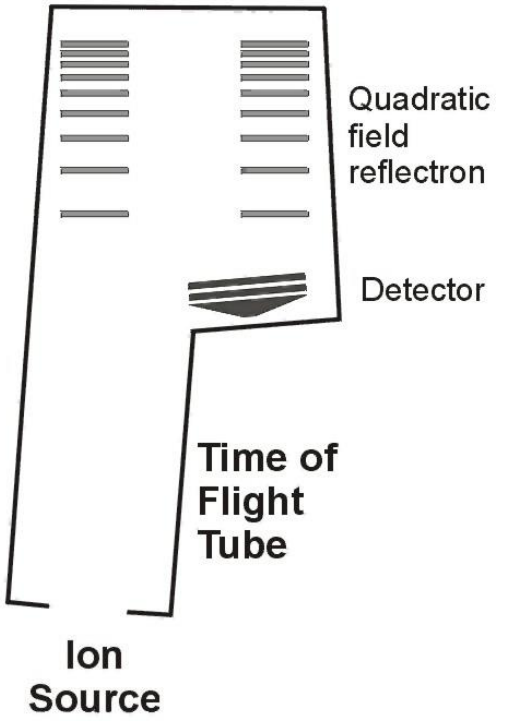
Normal and Reflectron analyzers

The mass spec resolution can be improved by increasing the length of the flight tube or (better) by refocusing the ion beam with a “reflectron”

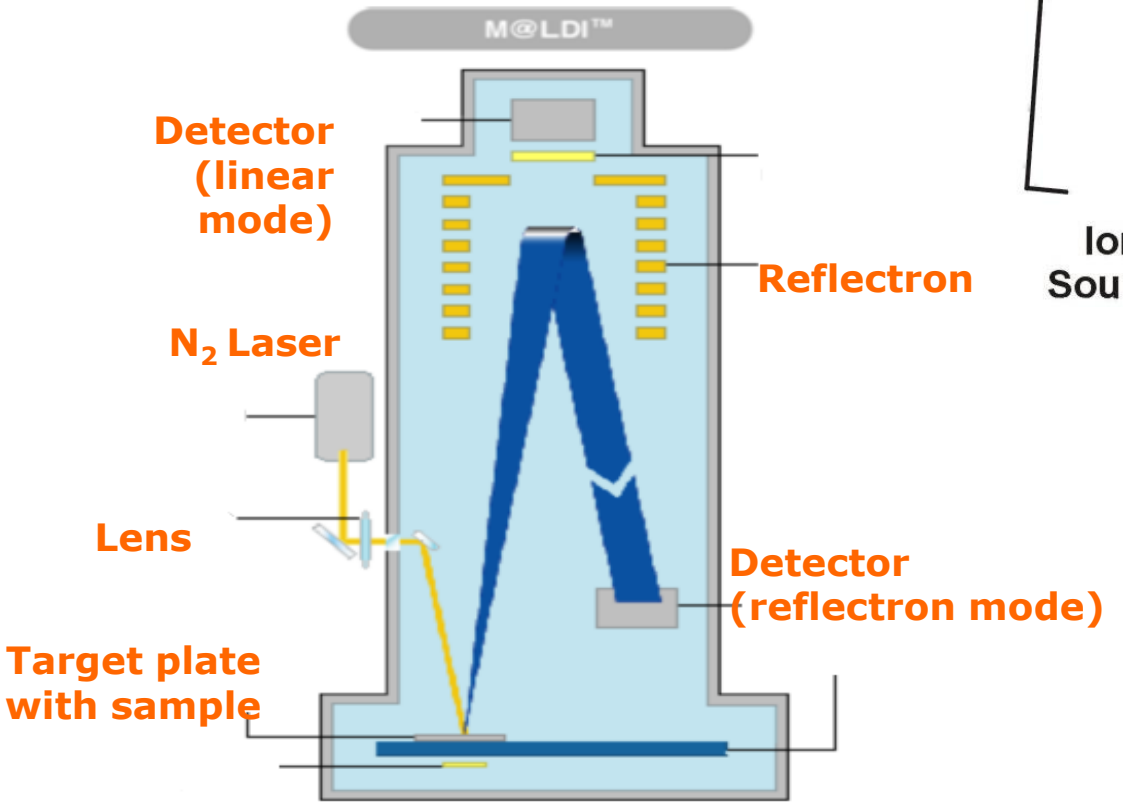


Ions with same m/z ratio but different kinetics energies are refocused by the reflectron therefore reducing differencies in fligh time due to different initial velocities

Reflectron



M@LDI™ LR by Micromass, UK



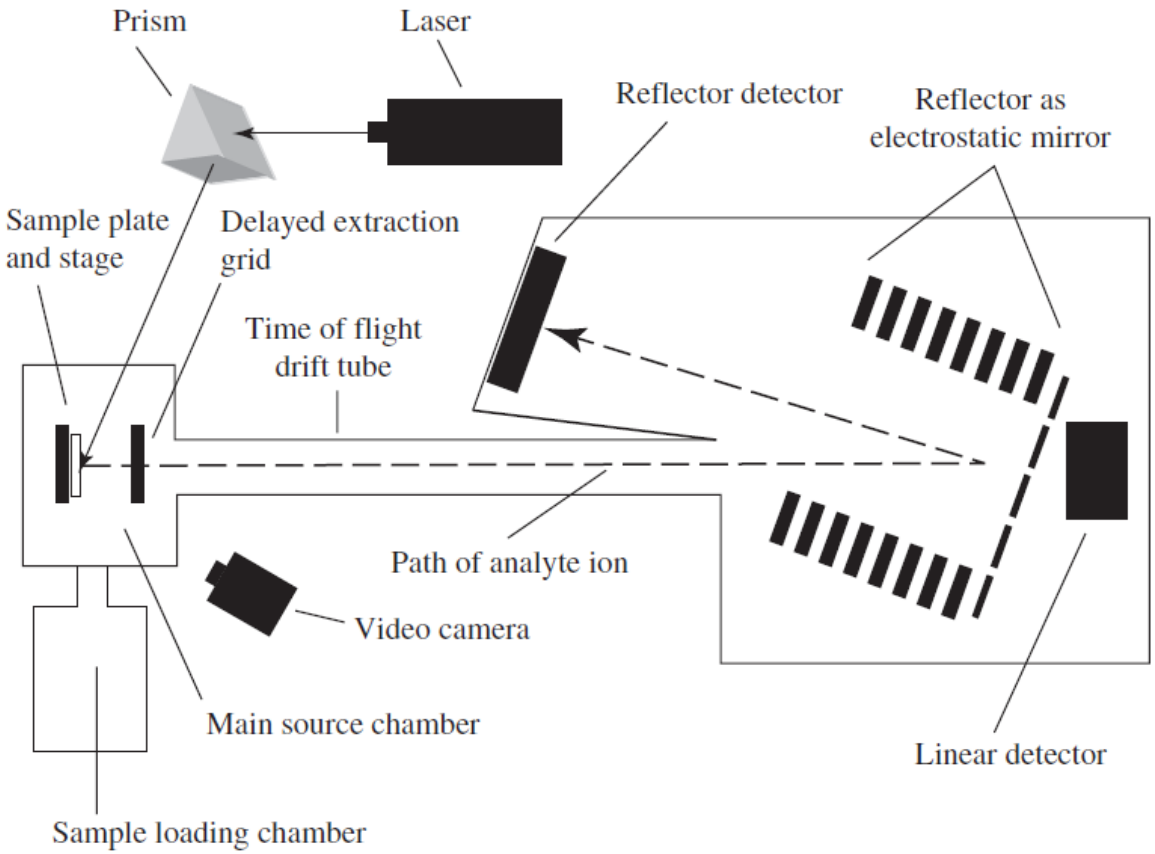


FIGURE 30.7 Components of a time-of-flight mass spectrometer illustrating the major sections including the source, drift tube, reflectron electrostatic mirror, and detector.

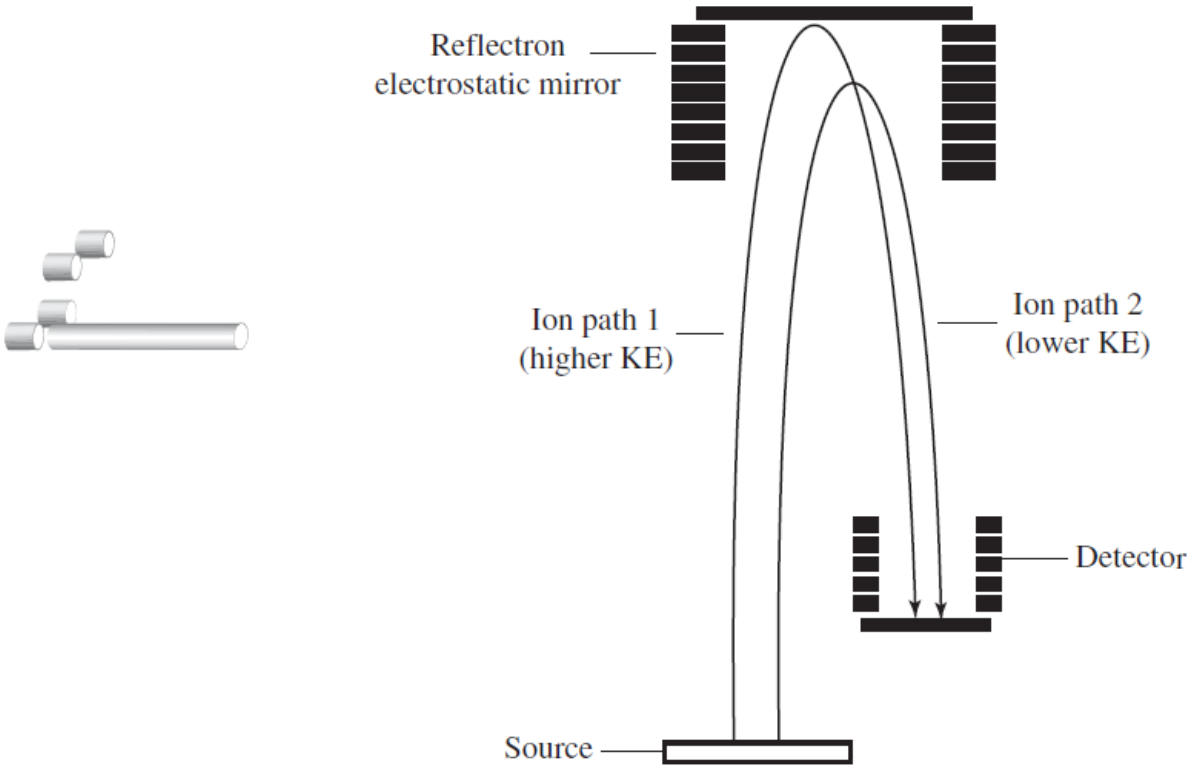
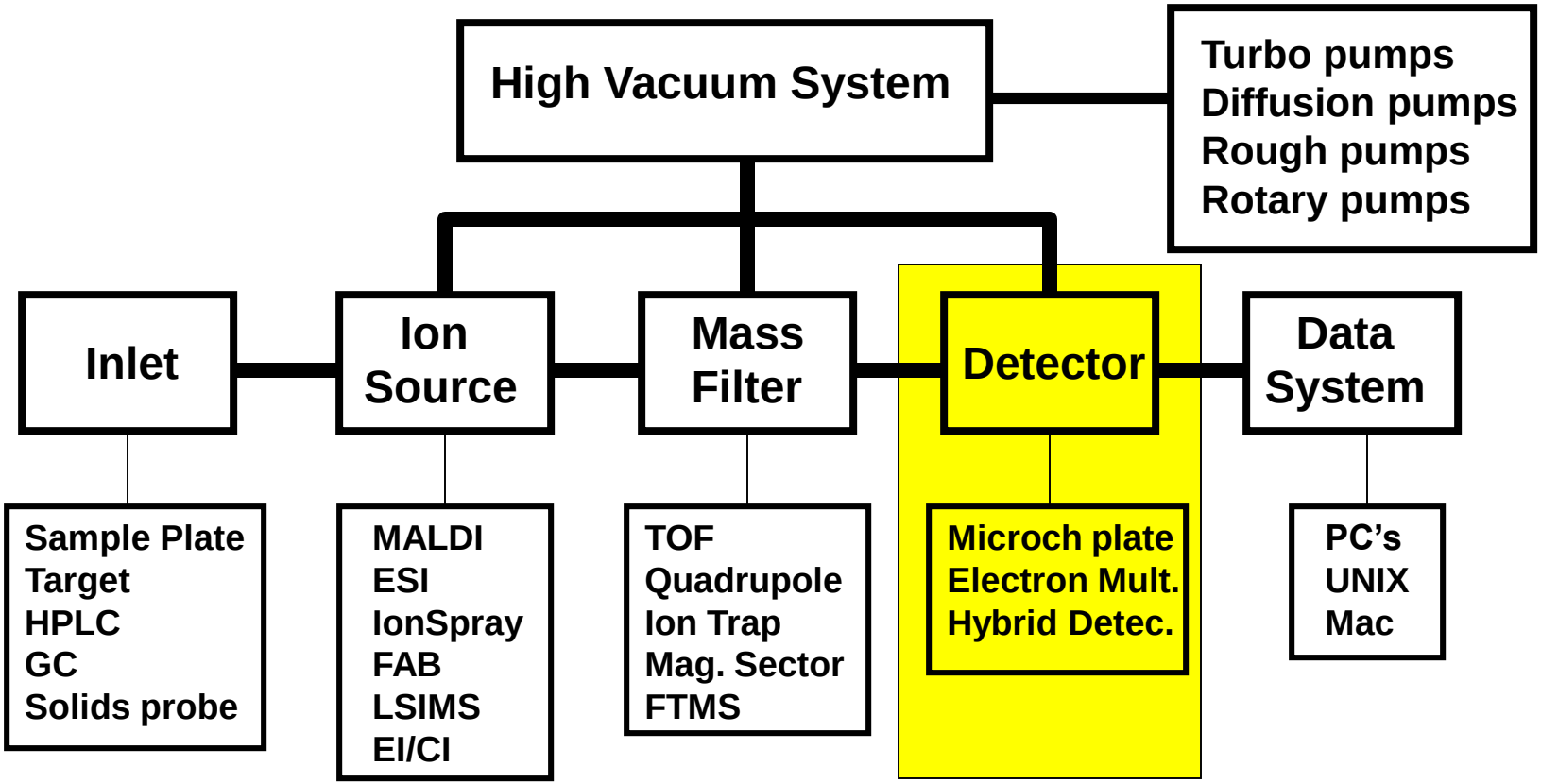


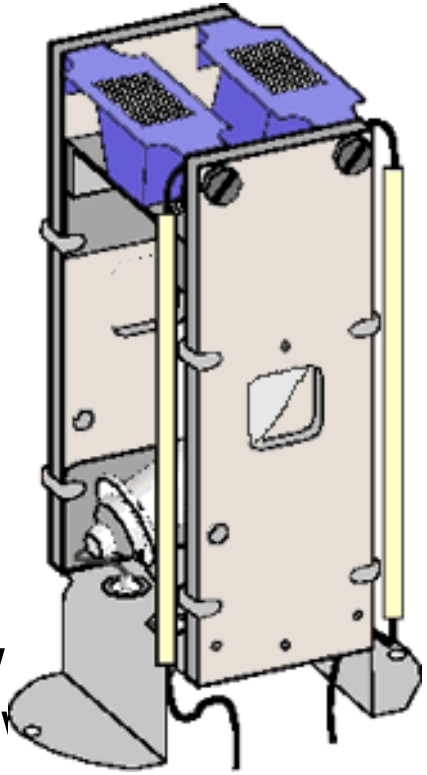
FIGURE 30.9 Electrostatic mirror focusing of two ions that have the same *m/z* value but slightly different kinetic energies. Ion path 1 possesses slightly higher kinetic energy in relation to ion path 2. Ion path 1 travels slightly farther to match that of ion path 2. The two ions are focused and arrive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

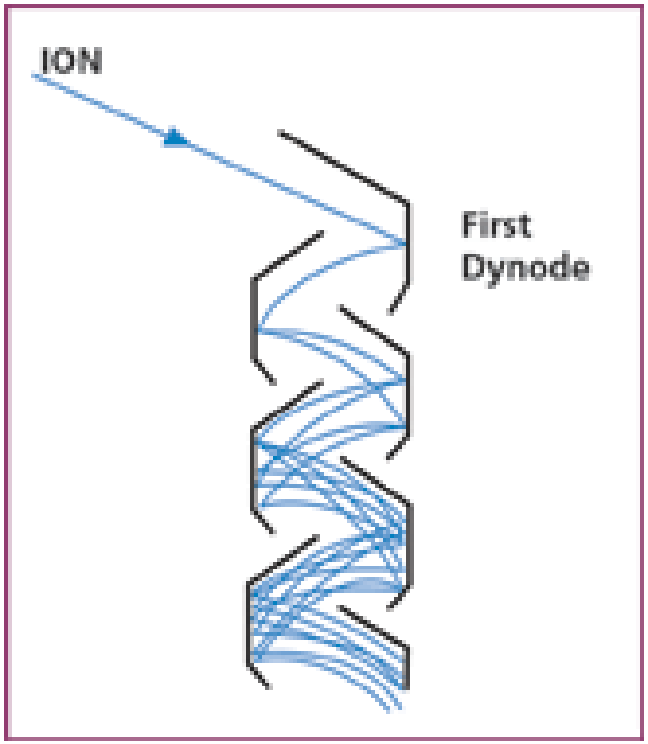
Mass Spectrometer Schematic



MS Detectors

- Early detectors used photographic film
- Today's detectors (**ion channel and electron multipliers**) produce electronic signals via secondary electronic emission when struck by ion
- Timing mechanisms integrate these signals w scanning voltages to allo the instrument to report which m/z has struck the detector
- Need constant and regular calibration





Electron Multiplier (Dynode)

Hybrid instruments /Tandem MS

Combines two or more mass analyzers of the same or different types

First mass analyzer isolates the ion of interest (parent ion)

The ions are then fragmented between the first and second mass analyzer via collisions

The last mass analyzer obtains the mass spectrum of the fragments ions (daughter ions spectrum)

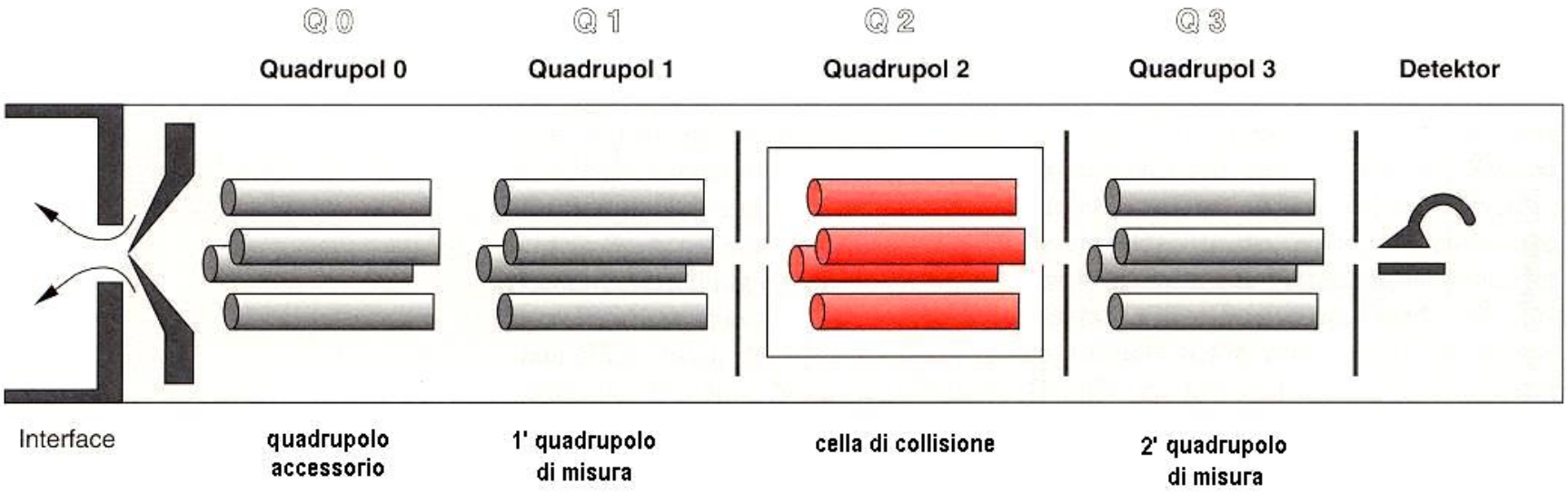
MS-MS spectra reveal fragmentation patterns to provide structural information about a molecule

- Different MS-MS configurations
  - Quadrupole-quadrupole (low energy)
  - Magnetic sector-quadrupole (high)
  - Quadrupole-time-of-flight (low energy)
  - Time-of-flight-time-of-flight (low energy)

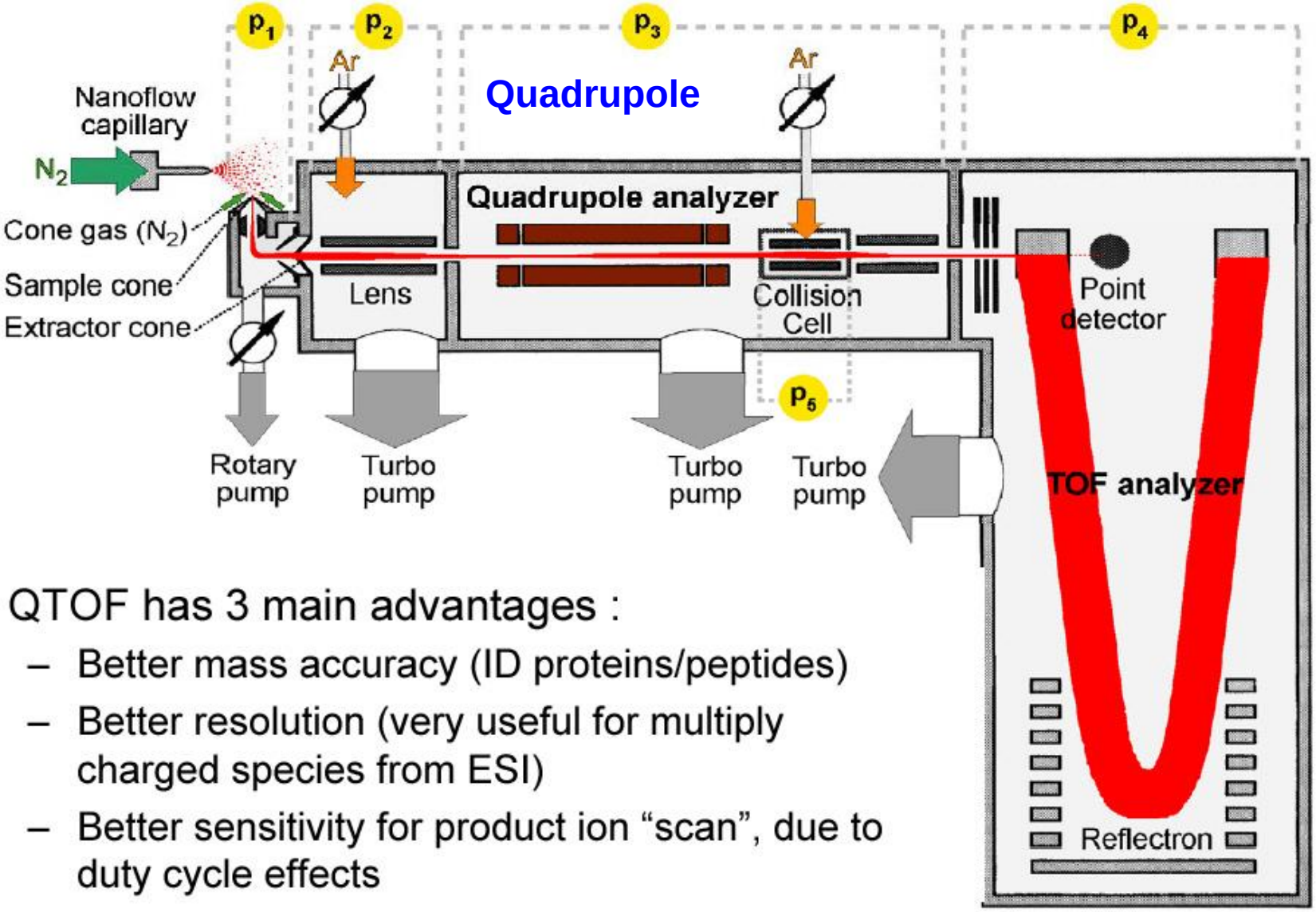
Tandem Mass Spectrometry

- Purpose is to fragment ions from parent ion to provide *structural information* about a molecule
- Also allows separation and identification of compounds in complex mixtures
- Uses two or more mass analyzers/filters separated by a collision cell filled with Argon or Xenon
- Collision cell is where selected ions are sent for further fragmentation

Triple quadrupole mass spectrometer



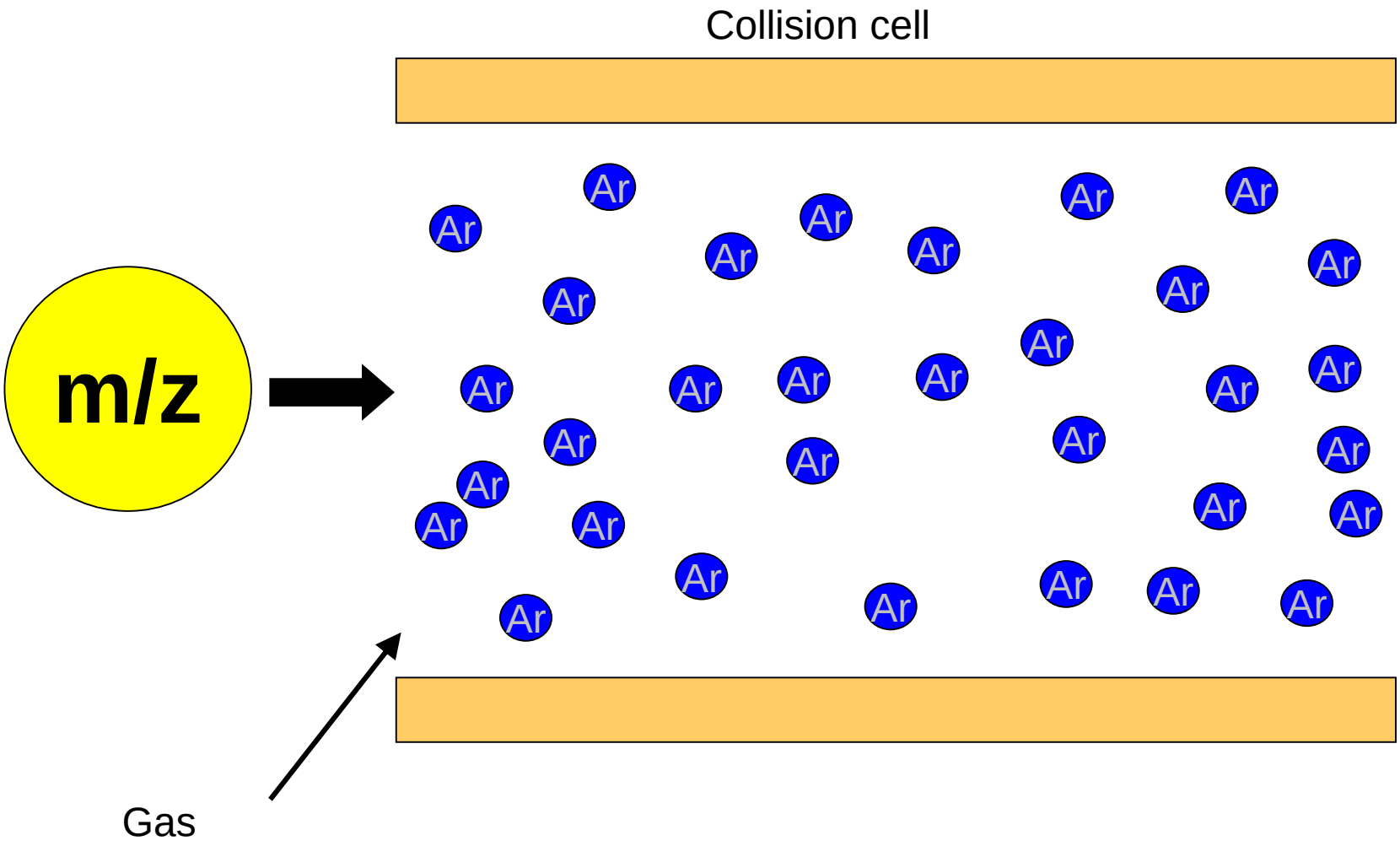
Q-TOF Mass Analyzer



QTOF has 3 main advantages :

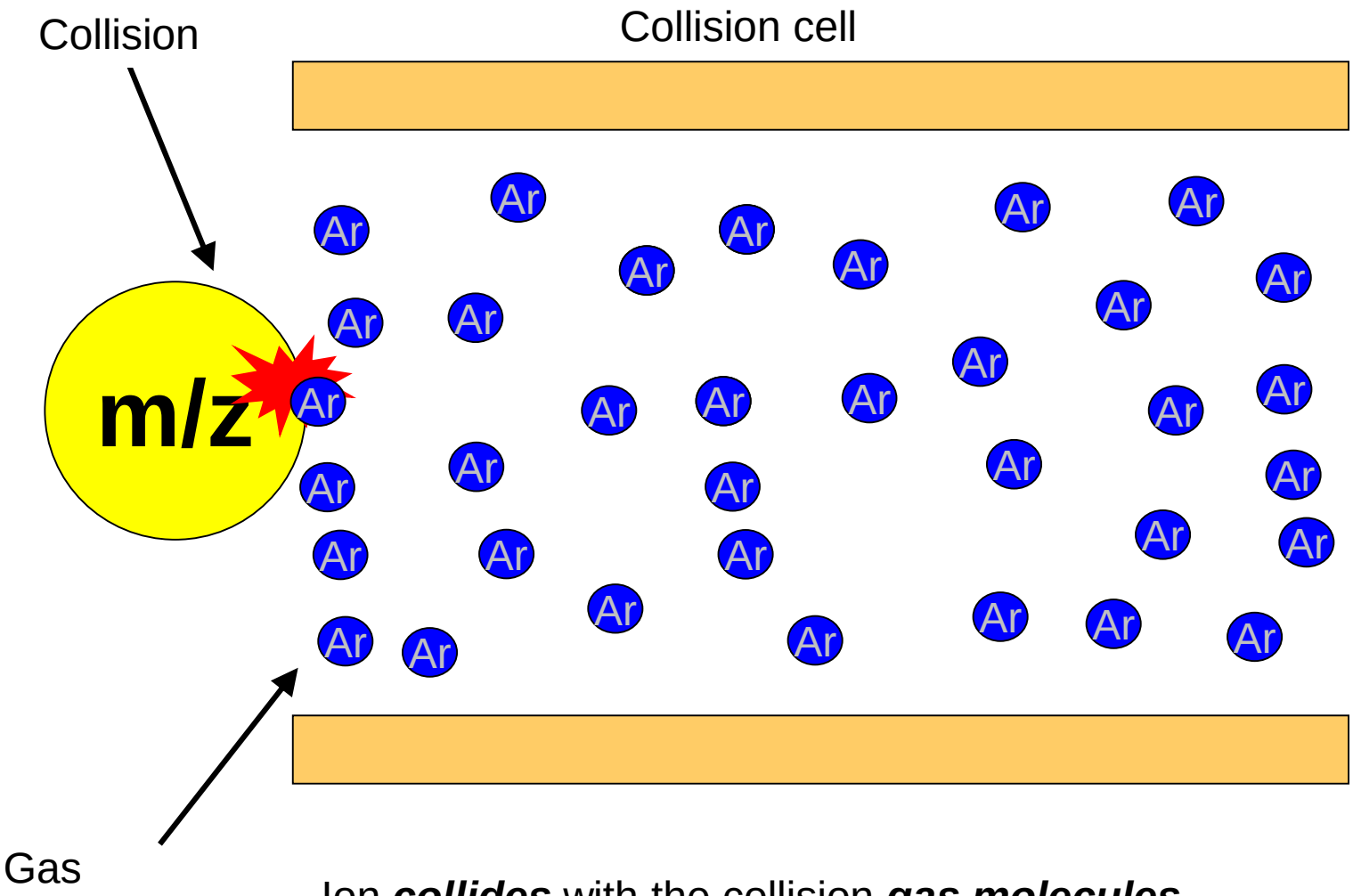
- Better mass accuracy (ID proteins/peptides)
- Better resolution (very useful for multiply charged species from ESI)
- Better sensitivity for product ion “scan”, due to duty cycle effects

CID: Collision-Induced Dissociation
CAD: Collision Activated Dissociation



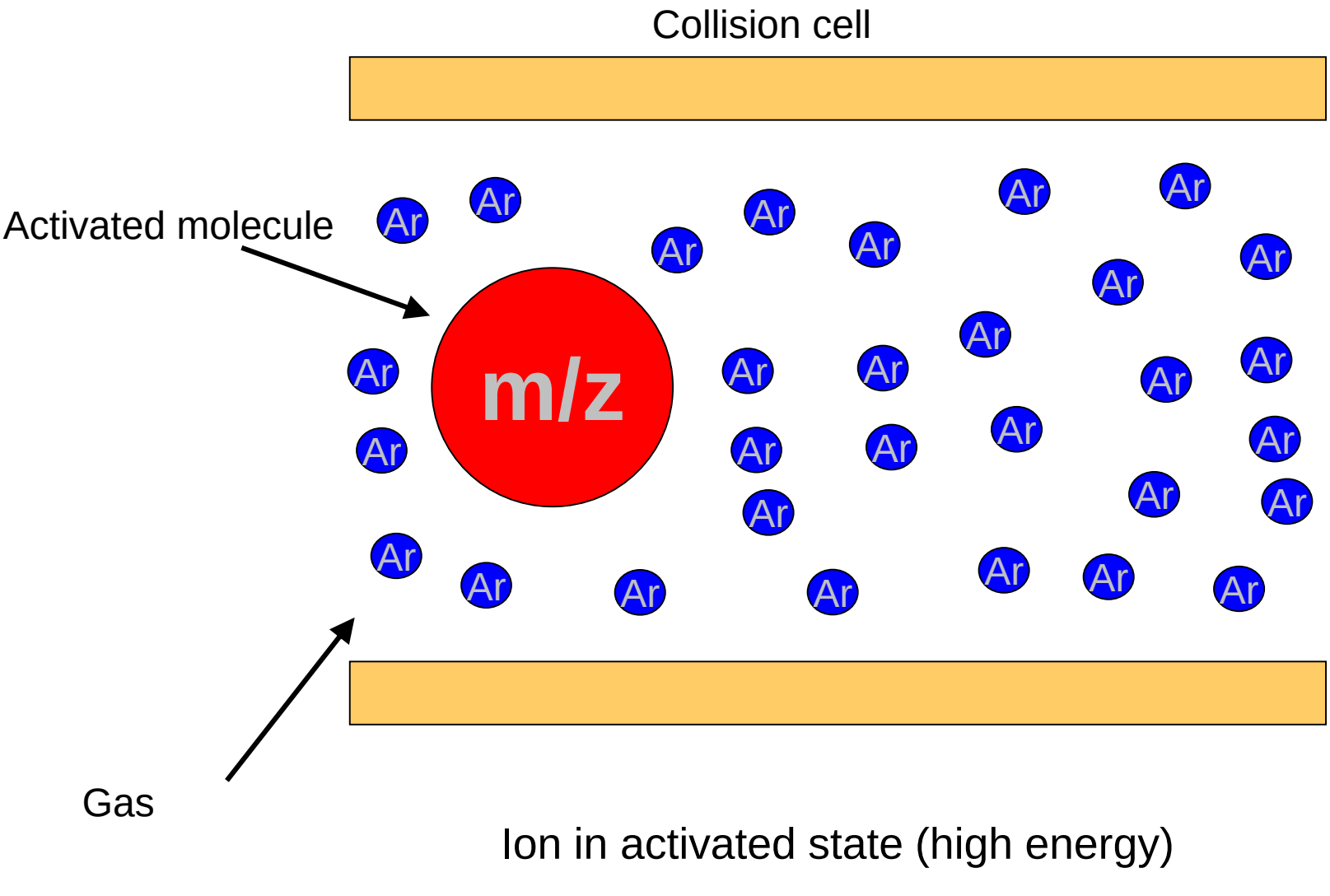
Ion enters into collision cell

CID: Collision-Induced Dissociation
CAD: Collision Activated Dissociation

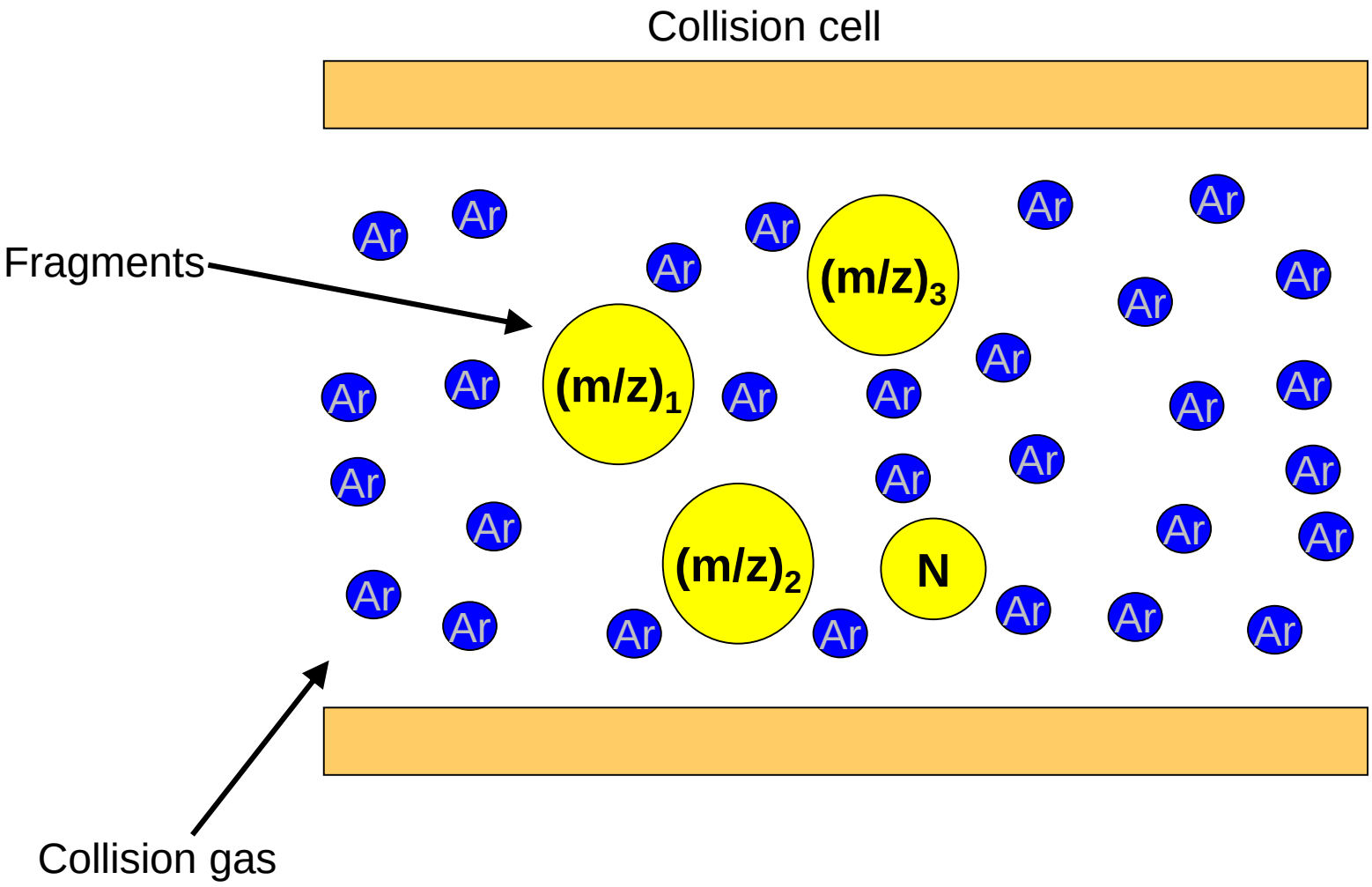


Ion ***collides*** with the collision ***gas molecules***

CID: Collision-Induced Dissociation
CAD: Collision Activated Dissociation

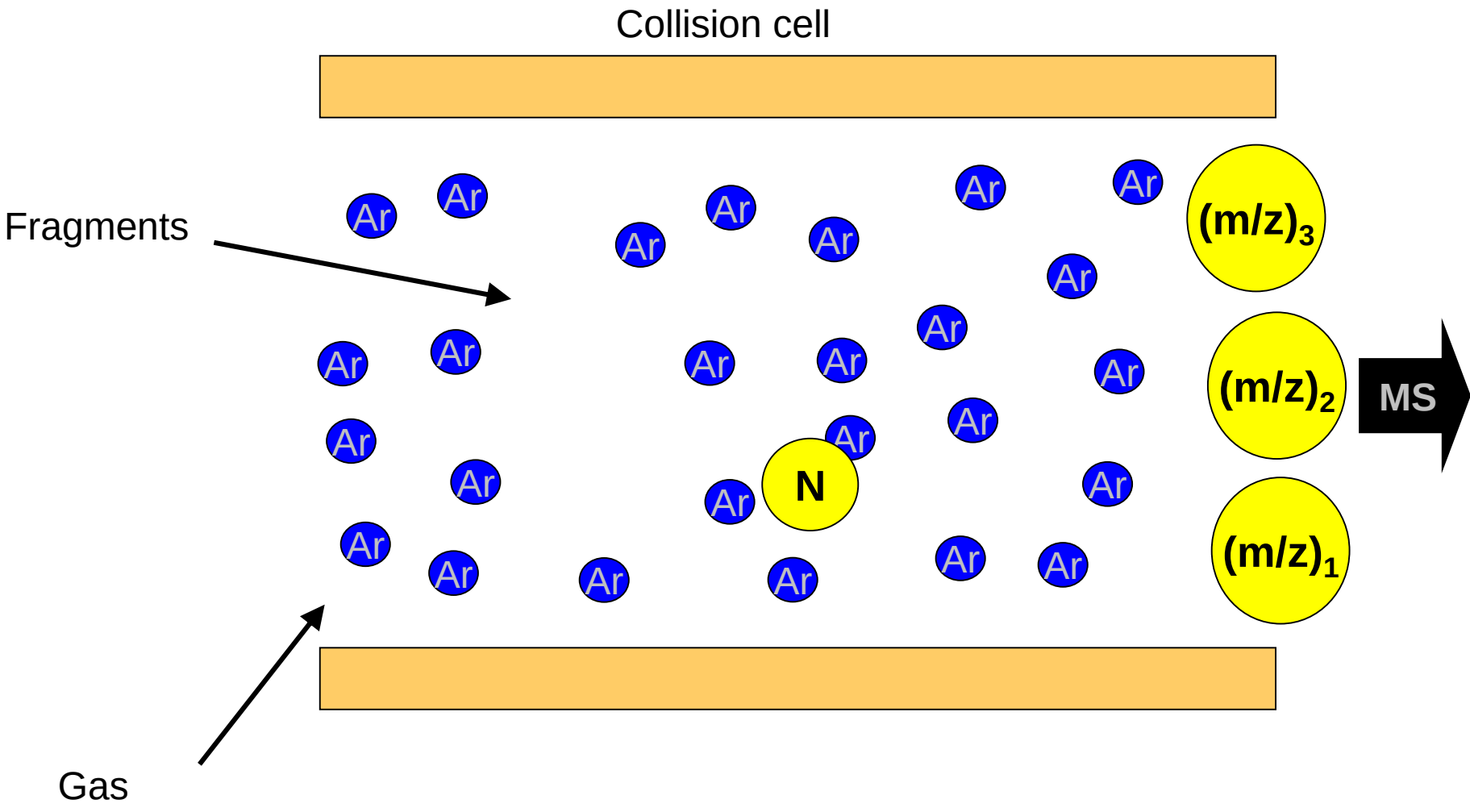


CID: Collision-Induced Dissociation
CAD: Collision Activated Dissociation



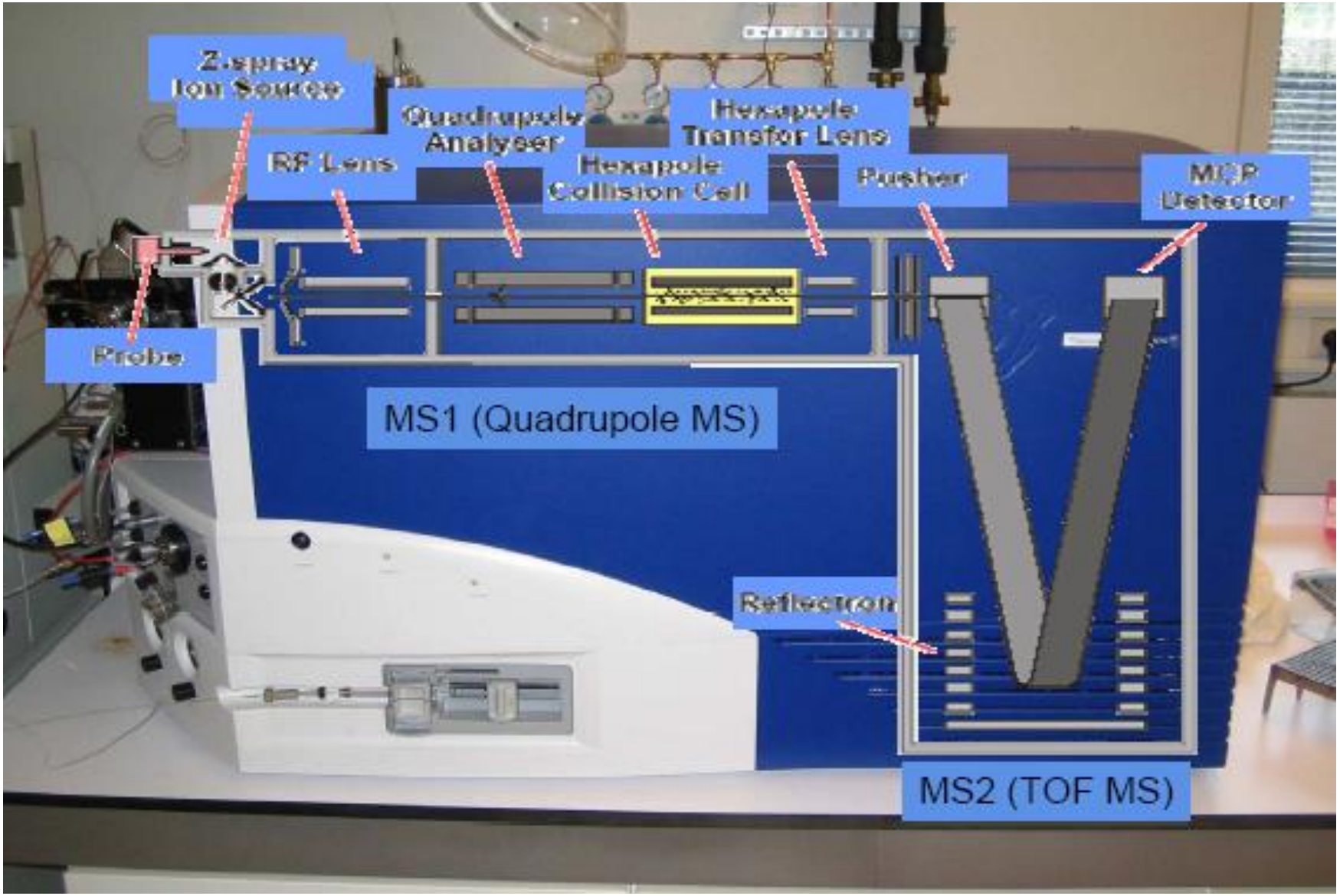
Fragmentation of the ion (to decrease E)

CID: Collision-Induced Dissociation
CAD: Collision Activated Dissociation



Ions produced by fragmentation can be analyzed by second analyzer

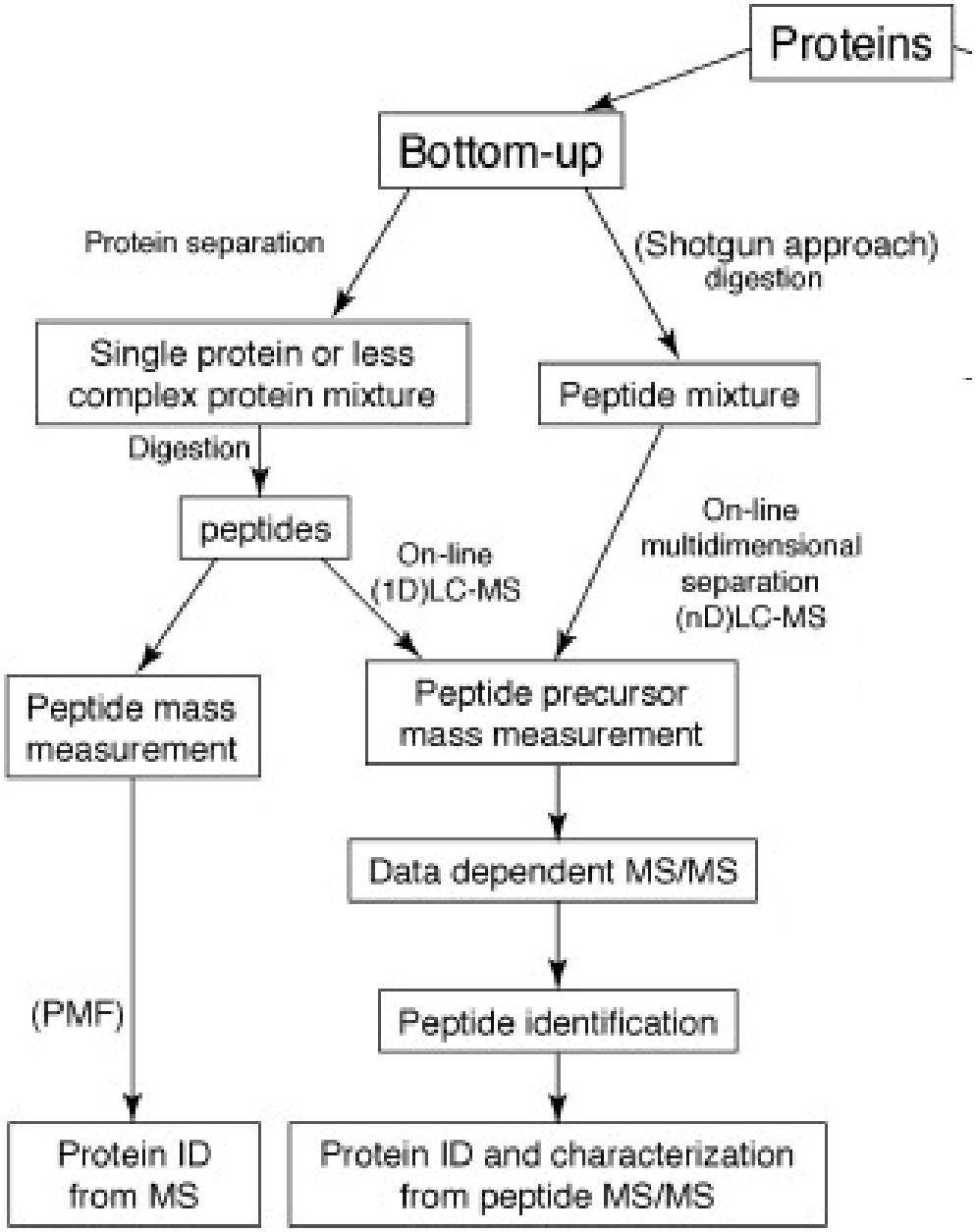
ESI-QTOF MS


Z-spray
Ion Source
Quadrupole
Analyser
Hexapole
Transfer Lens
RF Lens
Hexapole
Collision Cell
Pusher
MCP
Detector
Probe
MS1 (Quadrupole MS)
Reflectron
MS2 (TOF MS)

Proteomics Applications

- Protein sample identification/confirmation
- Protein sample purity determination
- Detection of post-translational modifications
- Detection of amino acid substitutions
- De novo peptide sequencing
- Mass fingerprint identification of proteins
- Monitoring protein folding
- Monitoring protein-ligand complexes/struct.

Bottom-up approach



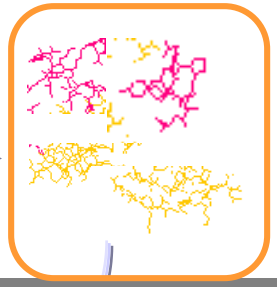
In the **bottom-up approach**, proteins in complex mixtures can be **separated before enzymatic (or chemical) digestion** followed by direct peptide mass fingerprinting-based acquisition or further peptide separation on-line coupled to tandem mass spectrometry.

Alternatively, the protein mixture can be directly digested into a collection of **peptides** (‘shotgun’ approach), which are then separated by multidimensional chromatography on-line coupled to tandem mass spectrometric analysis.

Overview of Shotgun Proteomics:

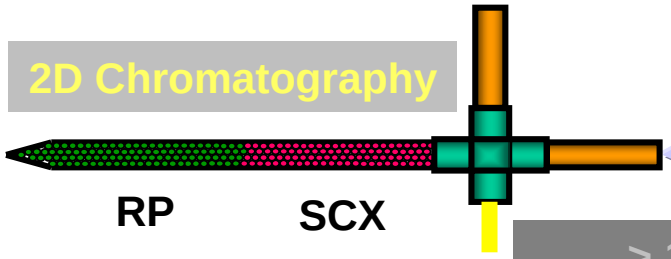
Tandem Mass Spectrometer

Digestion



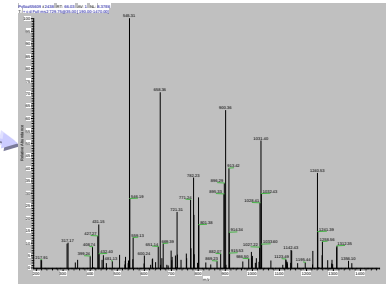
Peptide Mixture

2D Chromatography

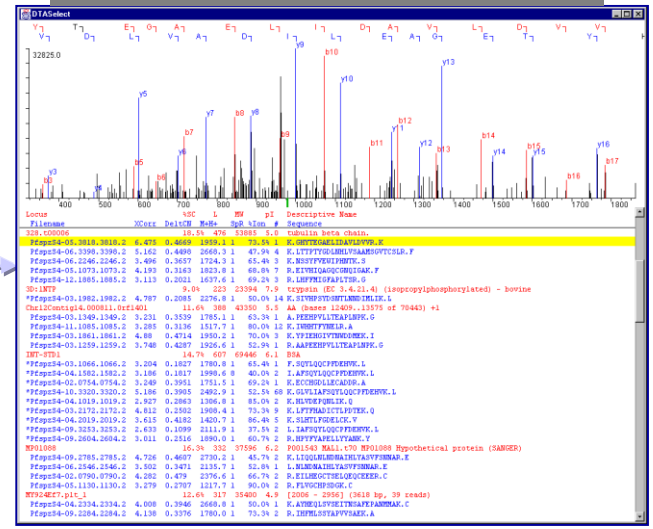


> 1,000 Proteins Identified

MS/MS Spectrum



SEQUEST®
DTASelect &
Contrast



Software for protein identification

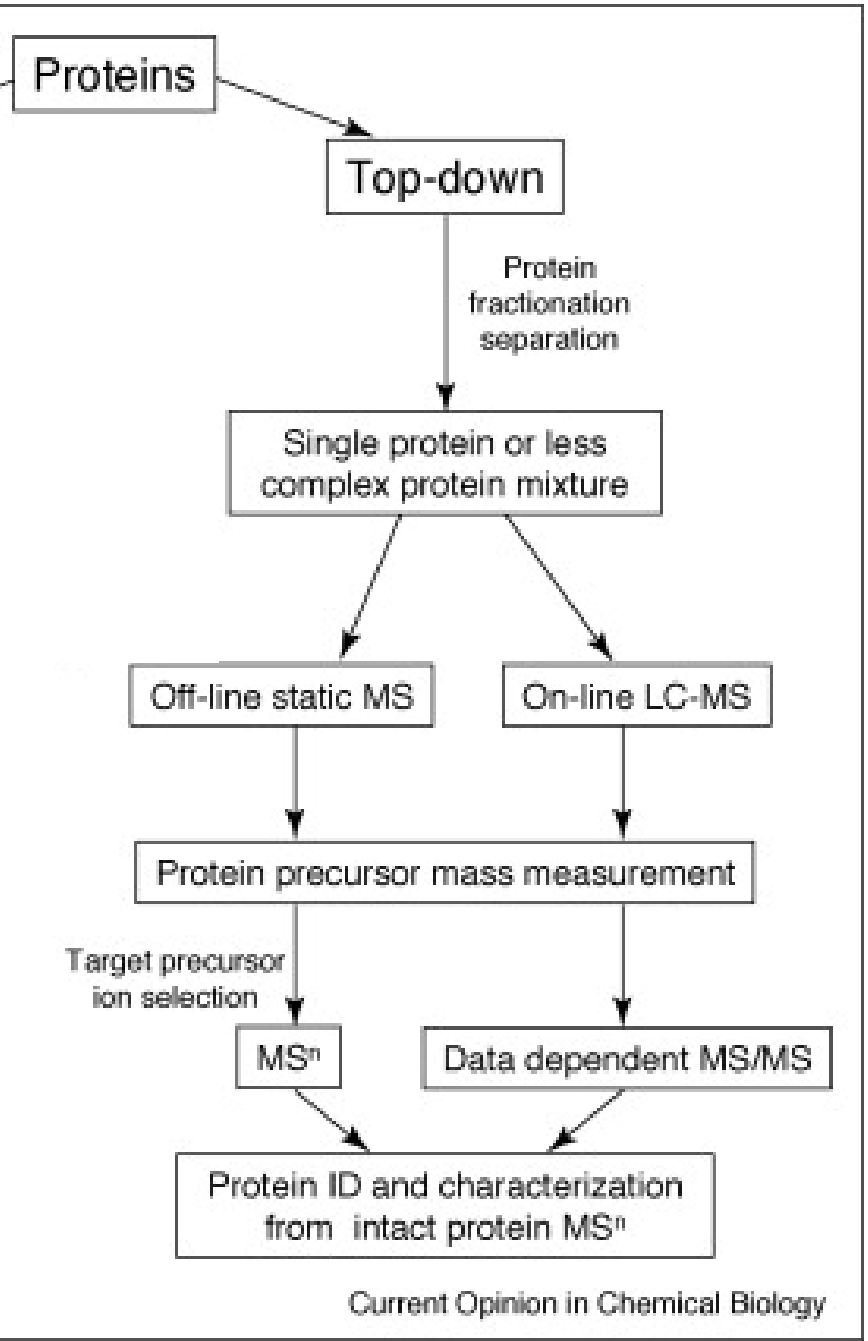
SEQUEST was one of the first tandem mass spectrometry database search program used for protein identification that correlates uninterpreted tandem mass spectra of peptides with amino acid sequences from protein and nucleotide databases.

MASCOT http://www.matrixscience.com/

Proteome Discoverer Software

Multiple database search capability provides the option of applying multiple search algorithms (SEQUEST, Z-Core, Mascot, etc.) and combining their outputs to maximize and cross-validate results

Top-down approach



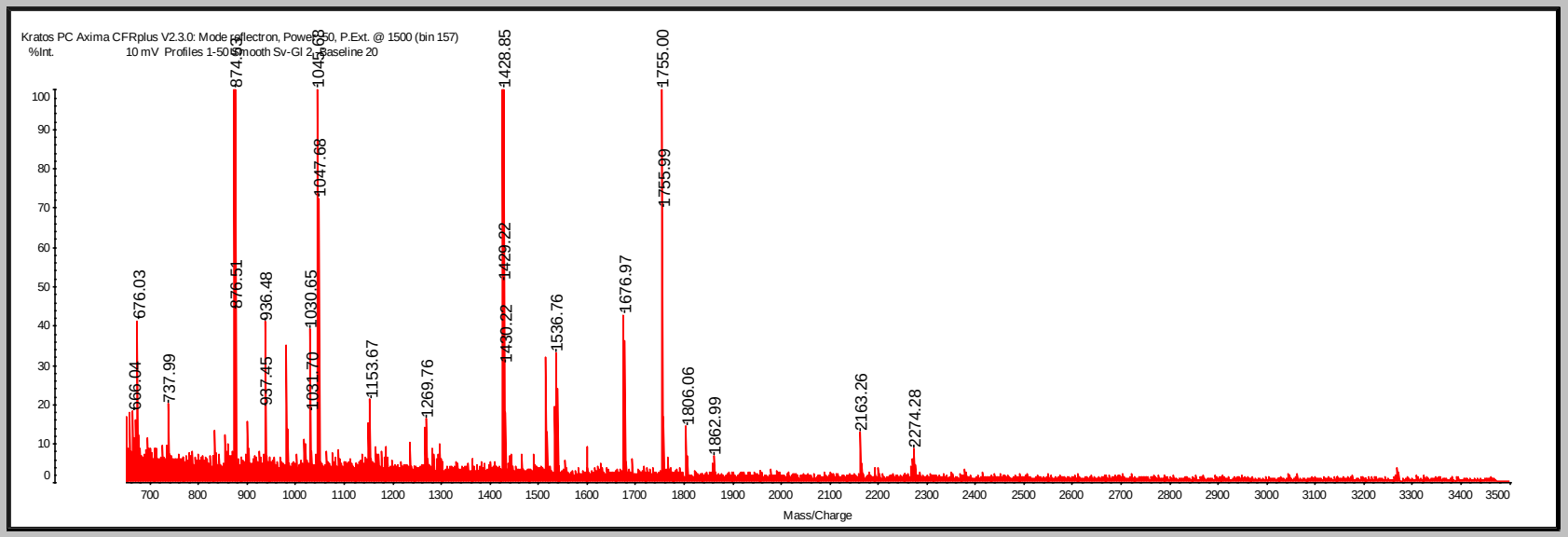
In the **top-down approach**, proteins in complex mixtures are fractionated and separated into pure single proteins or less complex protein mixtures, followed by off-line static infusion of sample into the mass spectrometer for **intact protein mass measurement** and **intact protein fragmentation**. An on-line LC–MS strategy can also be used for large-scale protein interrogation.

Software for protein identification

- Mascot is a powerful search engine which uses mass spectrometry data to identify proteins from primary sequence databases.
- X! Tandem open source is software that can match tandem mass spectra with peptide sequences

Peptide Mass Fingerprinting by MALDI-TOF...

…the most common procedure to identify proteins by MS



MS analysis of proteolytic fragments is a common way to identify a protein.

The following masses are entered for protein identification: 737.99 - 874.44 - 936.48 - 1030.65 - 1047.06 - 1153.67 - 1269.76 - 1428.85 1536.76 - 1676.97 - 1808.06 - 1862.99 - 2163.26 - 2274.28

Low MW peaks (e.g. < 500 Da) are not generally used because of the high interference of the matrix.

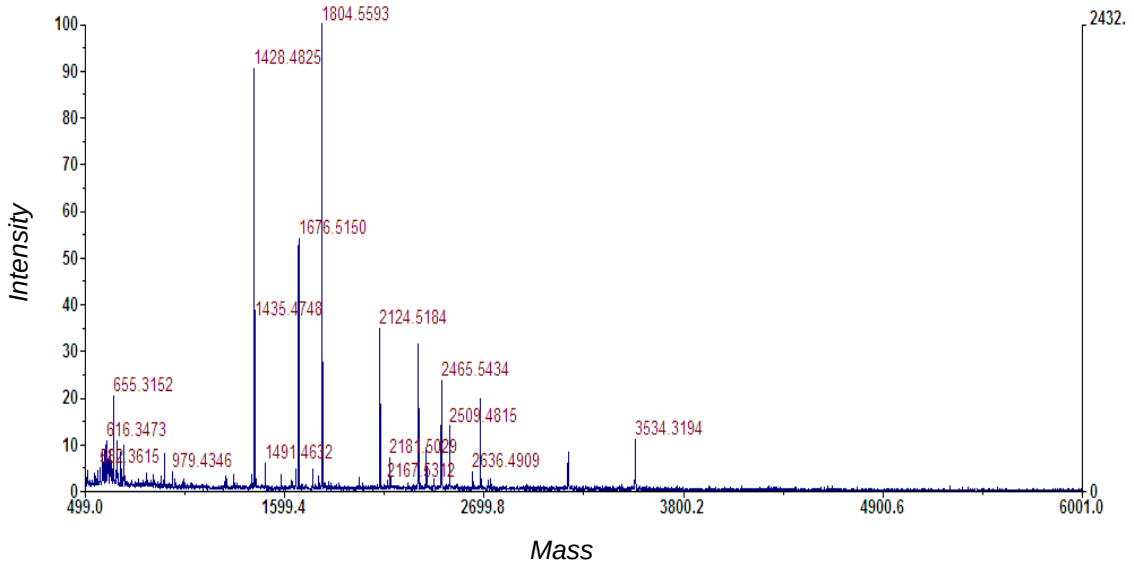
Peptide Mass Fingerprinting

- Used to identify protein spots on gels or protein peaks from an HPLC run
- Depends of 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
- Requires a database of known sequences
- Uses software to compare observed masses with masses calculated from database

Peptide Mass Fingerprint (PMF)

Database search → *Match beween teoretical mass and experimental ones*

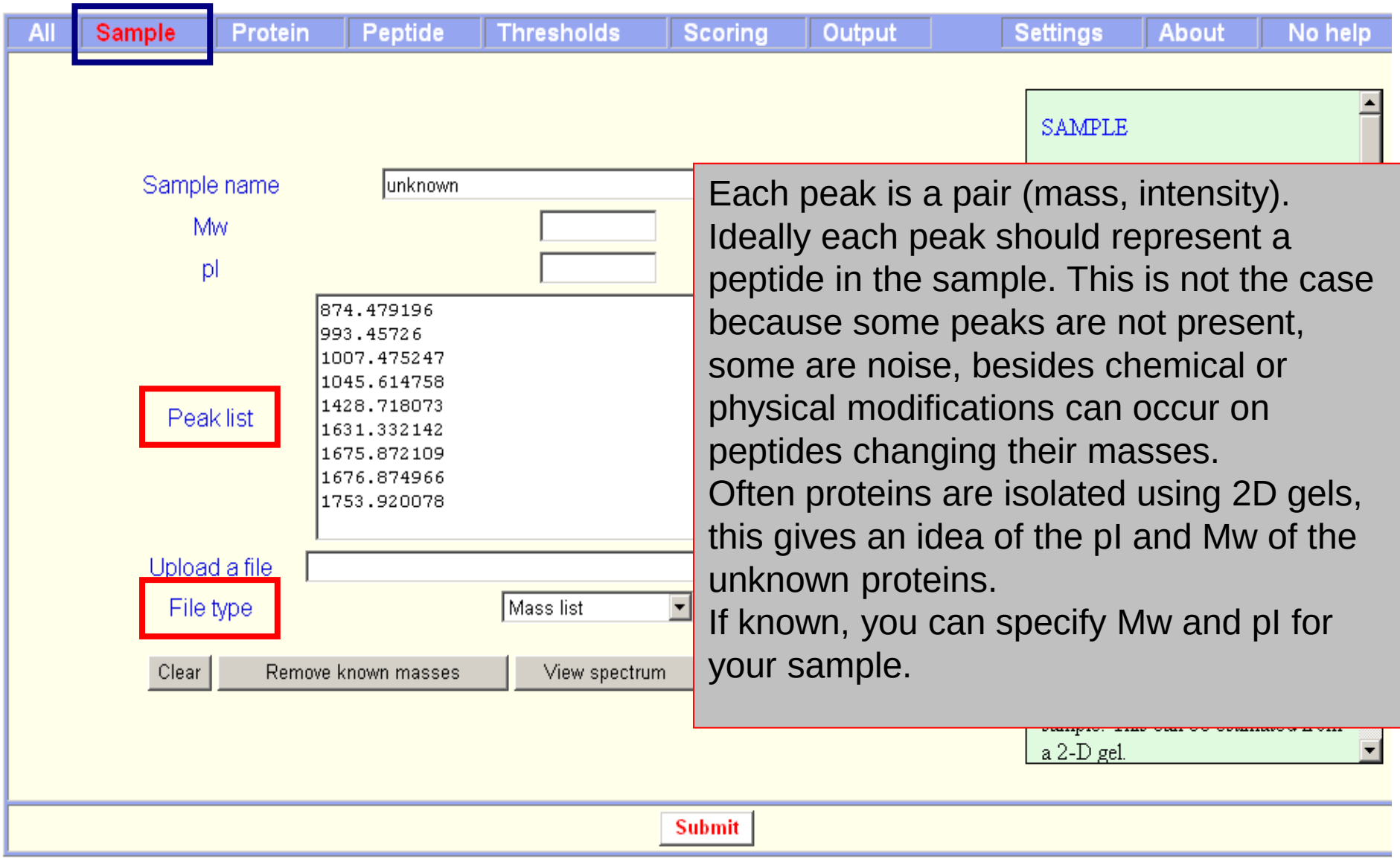
MASCOT	http://www.matrixscience.com
ALDENTE	http://www.expasy.org/tools/aldente/
PROFOUND	http://www.unb.br/cbsp/paginiciais/profound.htm
MS PROSPECTOR	http://prospector.ucsf.edu/prospector/4.0.7/html/msfit.htm
SWISS-PROT	http://www.expasy.org/sprot/



Mass list

1428.482505
1434.487611
1675.51502
1803.562355
2123.519746
2336.502887
2464.526762

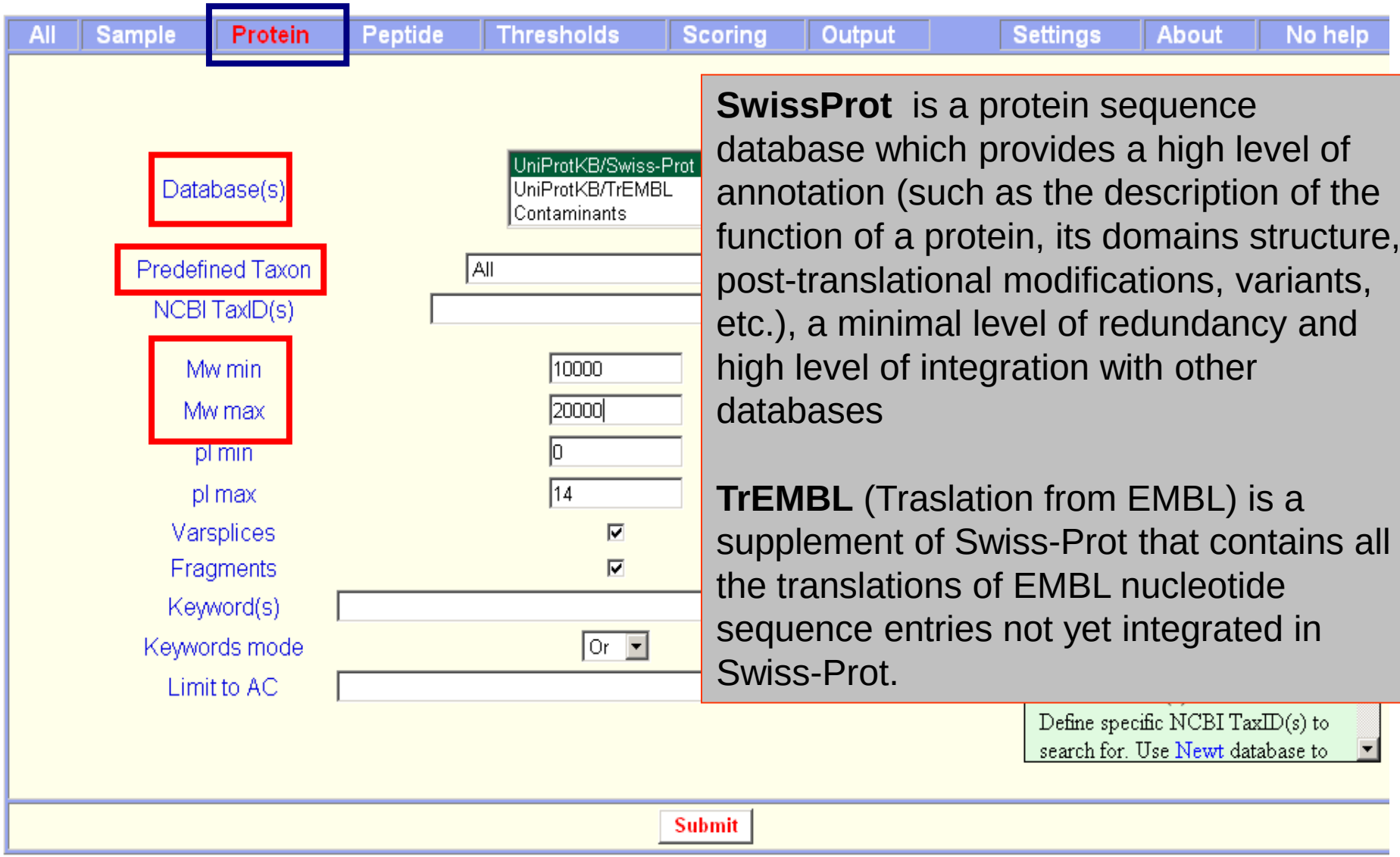
ALDENTE: sample



Each peak is a pair (mass, intensity). Ideally each peak should represent a peptide in the sample. This is not the case because some peaks are not present, some are noise, besides chemical or physical modifications can occur on peptides changing their masses.
Often proteins are isolated using 2D gels, this gives an idea of the pI and Mw of the unknown proteins.
If known, you can specify Mw and pI for your sample.

http://www.genebio.com/products/

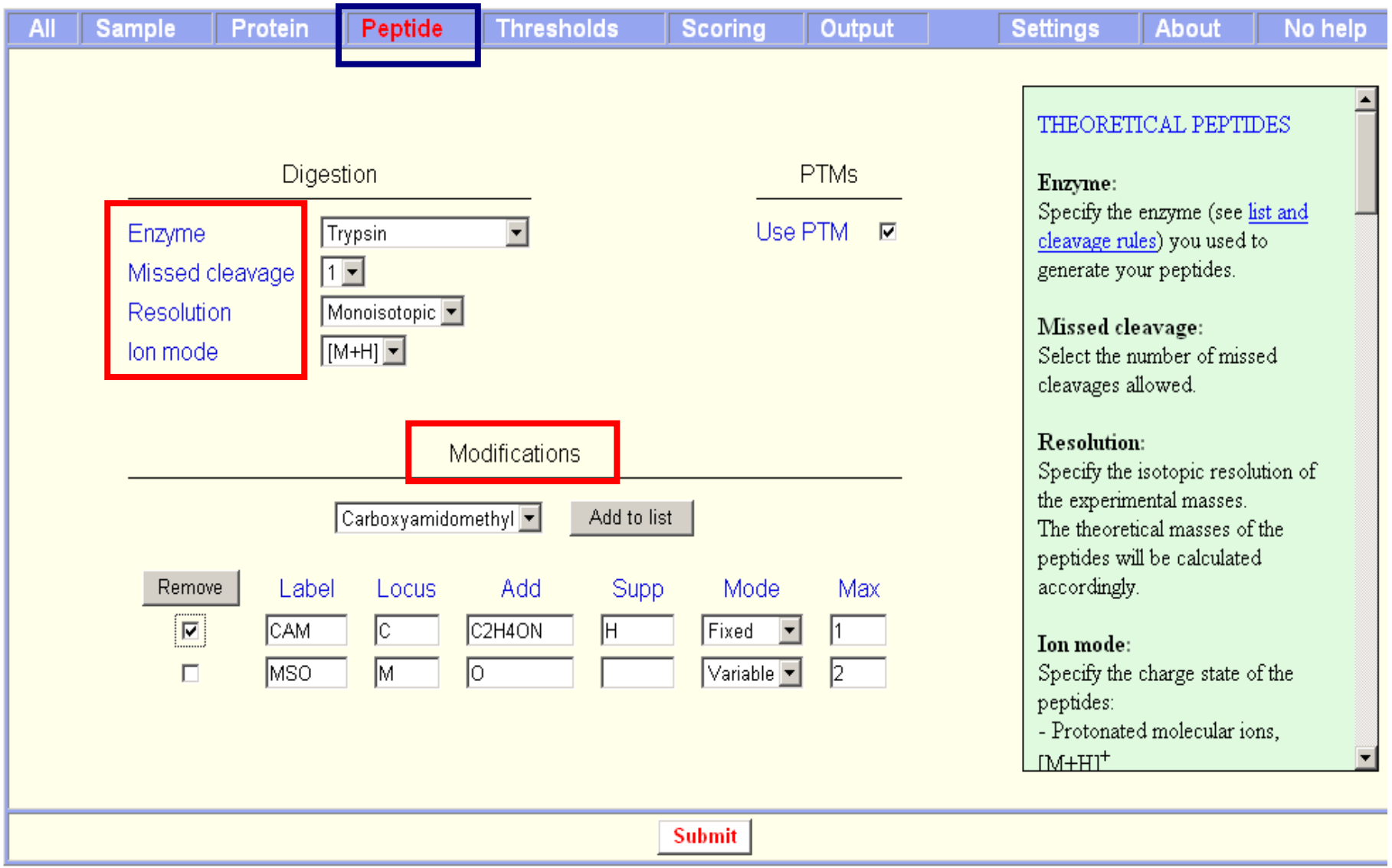
ALDENTE: protein



SwissProt is a protein sequence database which provides a high level of annotation (such as the description of the function of a protein, its domains structure, post-translational modifications, variants, etc.), a minimal level of redundancy and high level of integration with other databases

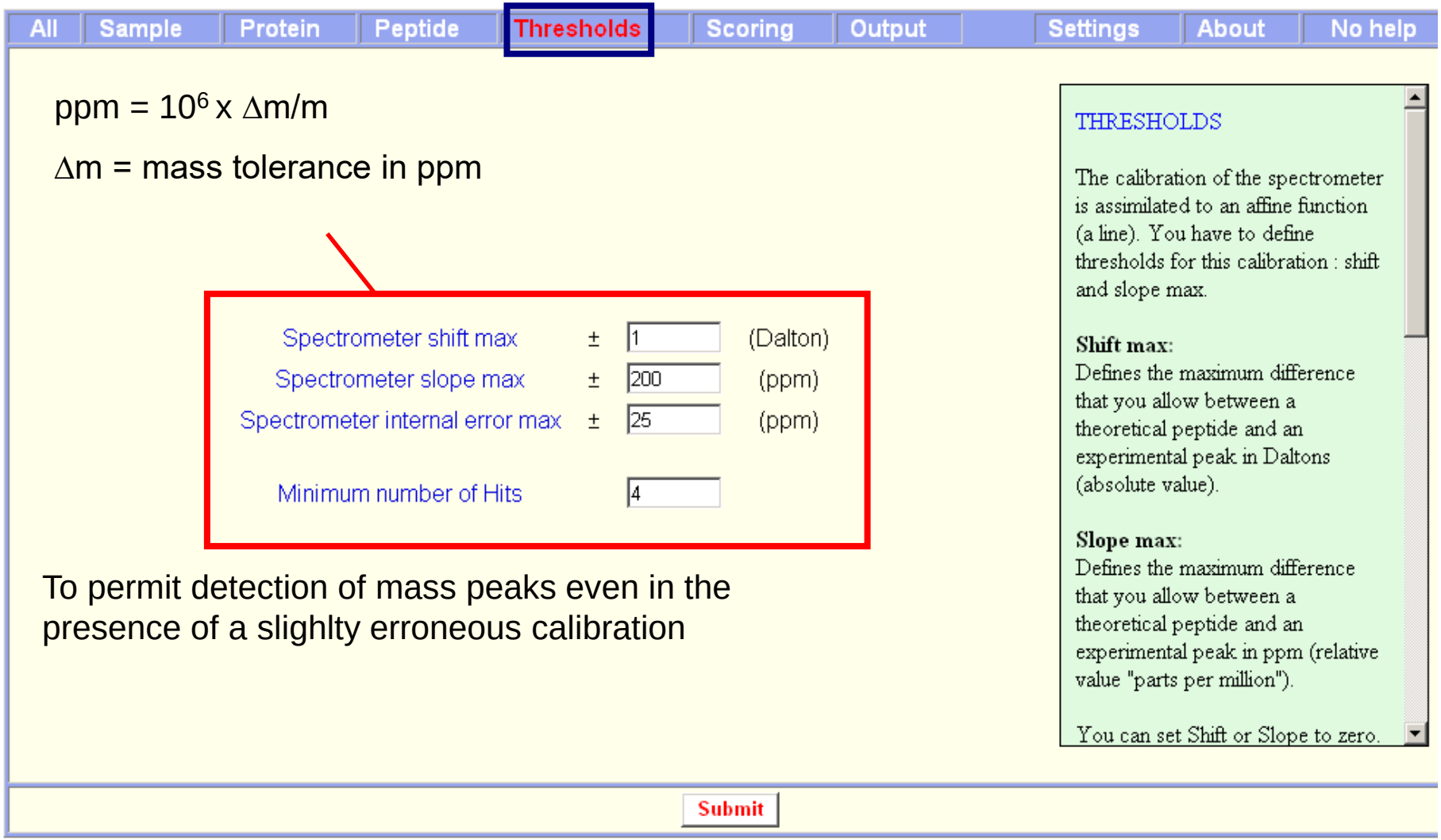
TrEMBL (Traslation from EMBL) is a supplement of Swiss-Prot that contains all the translations of EMBL nucleotide sequence entries not yet integrated in Swiss-Prot.

ALDENTE: peptide



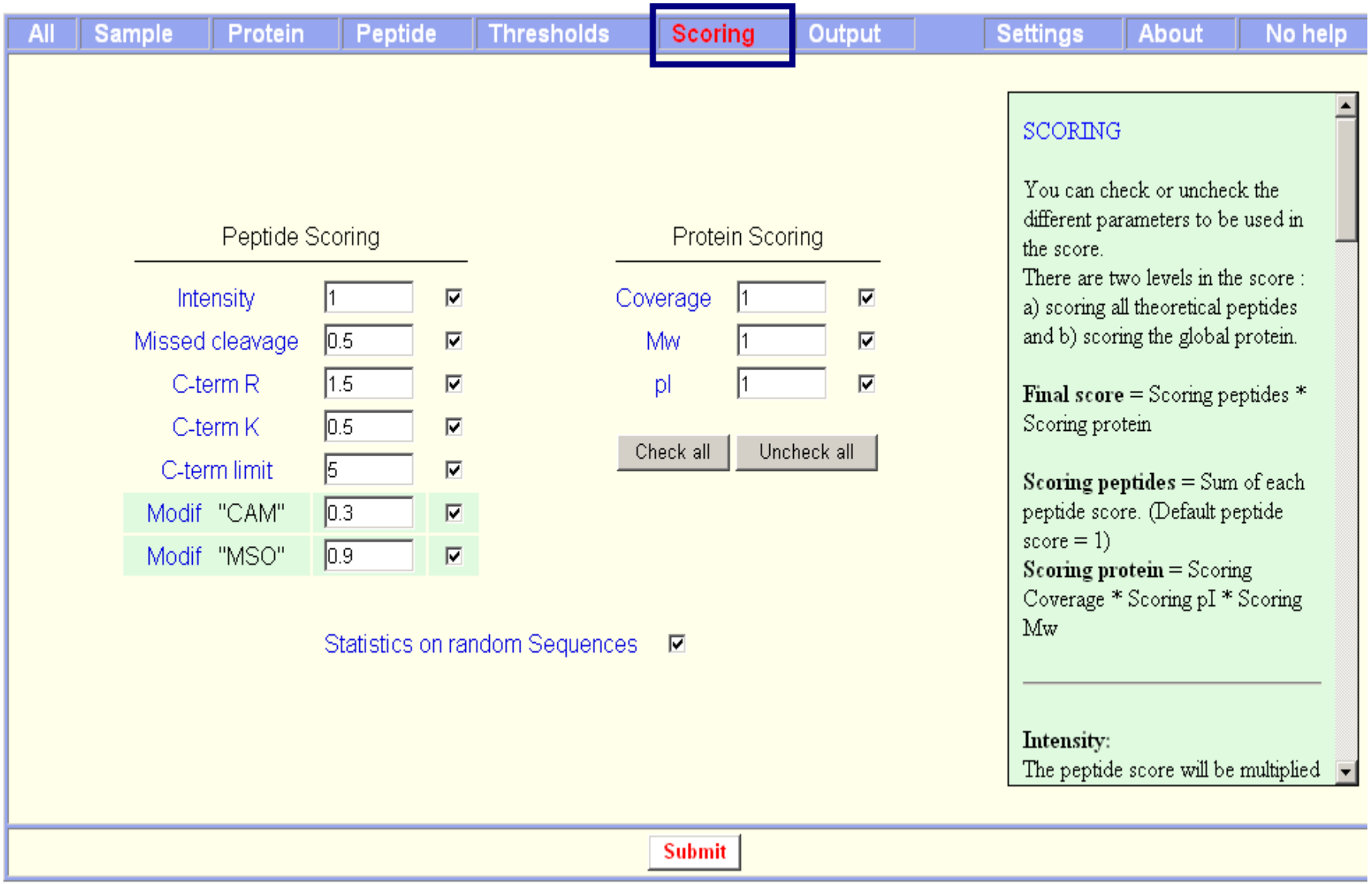
* CAM carboxyamidomethyl cysteine (produced by the reaction of ***iodoacetamide*** with cysteine groups)

ALDENTE: thresholds

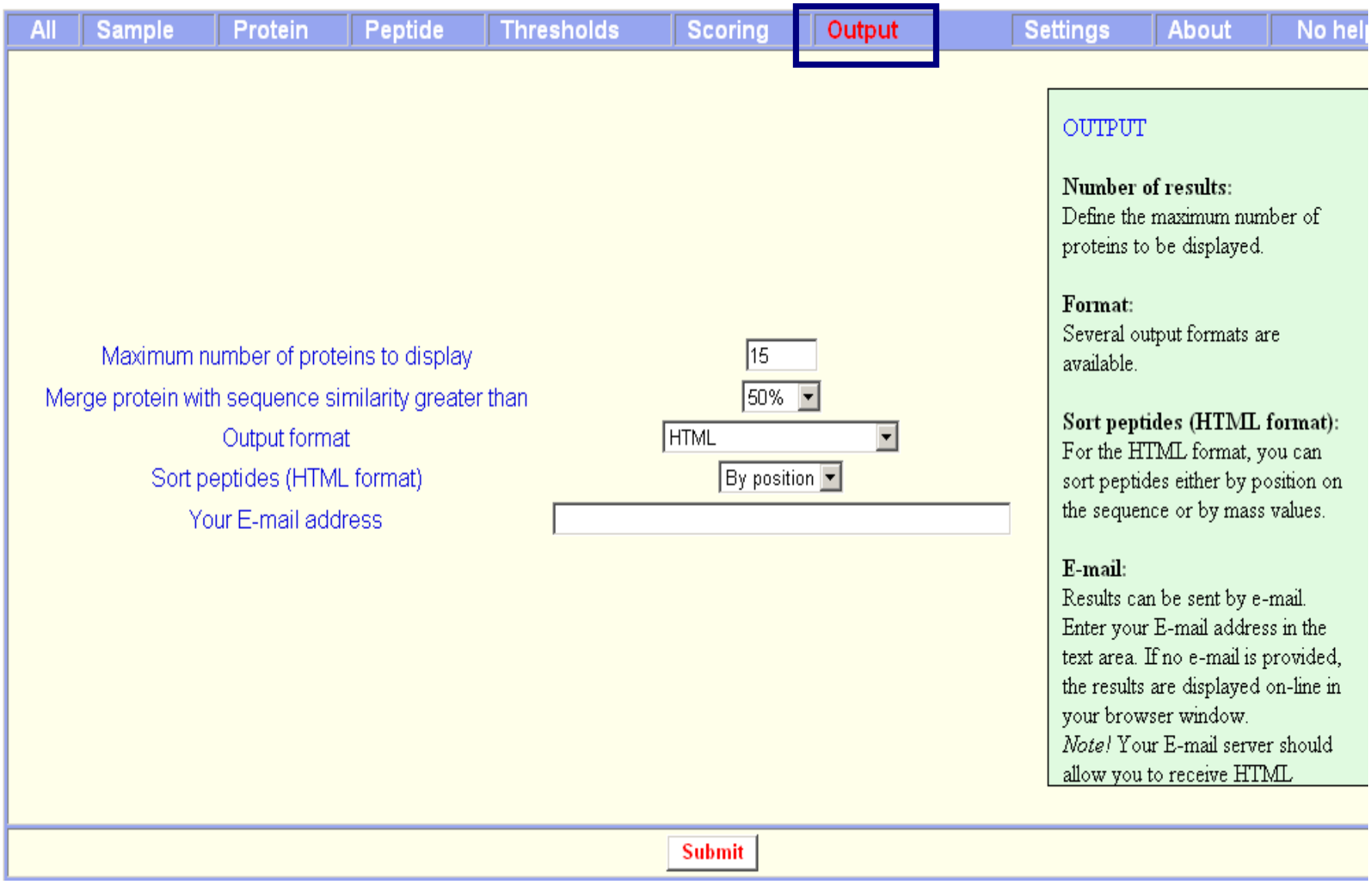


To permit detection of mass peaks even in the presence of a slighlty erroneous calibration

ALDENTE: scoring



ALDENTE: output



ALDENTE: results

The results are composed of three parts:
The Header gives general information on the request
The Summary displays a summary of the best proteins candidates
The details are displayed for each protein candidate

Sample	unknown / Peaks 9 / Mass [874.479196 - 1753.920078] / Intensity [1 - 1]
Date	29/11/2006 09:35:09 UTC
Release	UniProtKB/Swiss-Prot Release 51.1 of 14-Nov-2006: 251512 entries.
Proteins	- In range 46923 / After digestion 40 - First Analysis on 40 sequences : After Alignment 1 - Second Analysis on best 1 of first analysis : After Alignment 1 / Displayed 1
Peptides	Generated 1901717 / Matching a peak 13873 / Average per protein 40
Random	Generated 46923 / Best score 0.40 / Mean score 0.00 / Standard deviation 0.00

In range: Number of sequences in the selected database(s) and taxon(s), in the defined mass range, pI range, keywords, etc... Several sequences can be merged in a single Swiss-Prot or TrEMBL entry, whereas sequence variants for one protein entry are counted as separated sequences.
After digestion: Number of proteins with at least the number of hits after digestion and before the alignment process. This means: Number of proteins with enough theoretical peptides matching an experimental peak in the error space defined by the user to limit the comparison (e.g. 0.2 Da and 200 ppm) and before alignment (e.g. 25 ppm for internal error).
First analysis: The first analysis is run with a low precision for a quick scan of the set of proteins selected according to the previous filters (database, taxon, mass and pI ranges, number of hits, error tolerance). This first run will discared very unlikely proteins.
After alignment: Number of proteins with at least the number of hits aligned (low resolution).
Second analysis: The second analysis is run on the best X proteins of the first analysis (X is the number of proteins to displayed, at least 30). If the number of proteins after the first analysis is less than 30, only these proteins goes into the second analysis.
After alignment: Number of proteins with at least the number of hits aligned (high resolution).
Displayed: Number of proteins displayed defined by the user or less if there are not enough proteins to display.
Note! Because of the higher resolution, some proteins can be discarded after the second analysis. Then the protein number displayed may be less than the max number of proteins defined by the user.

Summary

Identification Summary

Rank	Score	Hits	AC	ID	DE	Mw	pI	Cov	TaxId
					30 first characters	kDa		%	
1	65.75	21	P17844	DDX5_HUMAN (C_1)	Probable ATP-dependent RNA hel...	69	9.1	33	9606
2	32.37	13	Q9Y2X3	NOL5_HUMAN (C_1)	Nucleolar protein 5.	60	9.0	28	9606
3	7.27	6	P00761	TRYP_PIG (C_1)	Trypsin.	23	8.3	31	9823
4	3.74	9	P02769	ALBU_BOVIN (C_1)	Serum albumin.	66	5.6	18	9913
5	2.53	10	Q8K4G5	ABLM1_MOUSE (C_1)	Actin-binding LIM protein 1.	97	8.9	13	10090
6	1.92	6	P55194	3BP1_MOUSE (C_1)	SH3 domain-binding protein 1.	65	5.5	14	10090
7	1.87	10	Q8K449	ABCA9_MOUSE (C_1)	ATP-binding cassette sub-famil...	183	6.5	7	10090
8	1.83	13	P41233	ABCA1_MOUSE (C_1)	ATP-binding cassette sub-famil...	254	6.4	7	10090
9	1.63	6	Q9Z176	2A5R_MOUSE (C_1)	Protein phosphatase 2A, 59 kDa...	56	4.9	13	10090
10	1.61	7	Q6P542	ABCF1_MOUSE (C_1)	ATP-binding cassette sub-famil...	95	6.2	11	10090
11	1.59	6	P12023	A4_MOUSE (C_1)	Amyloid beta A4 protein.	85	4.7	11	10090
12	1.55	9	Q8K442	ABC8A_MOUSE (C_1)	ATP-binding cassette sub-famil...	184	6.9	6	10090
13	1.53	8	Q80XL6	ACD11_MOUSE (C_1)	Acyl-CoA dehydrogenase family ...	87	8.7	11	10090
14	1.39	7	Q99P81	ABCG3_MOUSE (C_1)	ATP-binding cassette sub-famil...	74	6.7	13	10090
15	1.27	6	P48193	41_MOUSE (C_1)	Protein 4.1.	96	5.4	8	10090

Score

Aldente score of the protein. The background color indicates in red or green if the score is lower or greater than the best random score; this threshold should help to eliminate wrong identifications, taking into account the possibility to produce a better score randomly for these proteins.

- Details of a candidate

1) P00698 LYSC_CHICK (C_1) Up

Lysozyme C.
UniProtKB/Swiss-Prot : 9031

Score	Z-score	Mw	pI	Hits	Coverage	Shift (Da)	Slope (ppm)
18.42	9872.50	14313	9.33	6	51%	0.04	17

methionine oxidation (***MSO***)
carbamidomethyl cysteine (**CAM)**

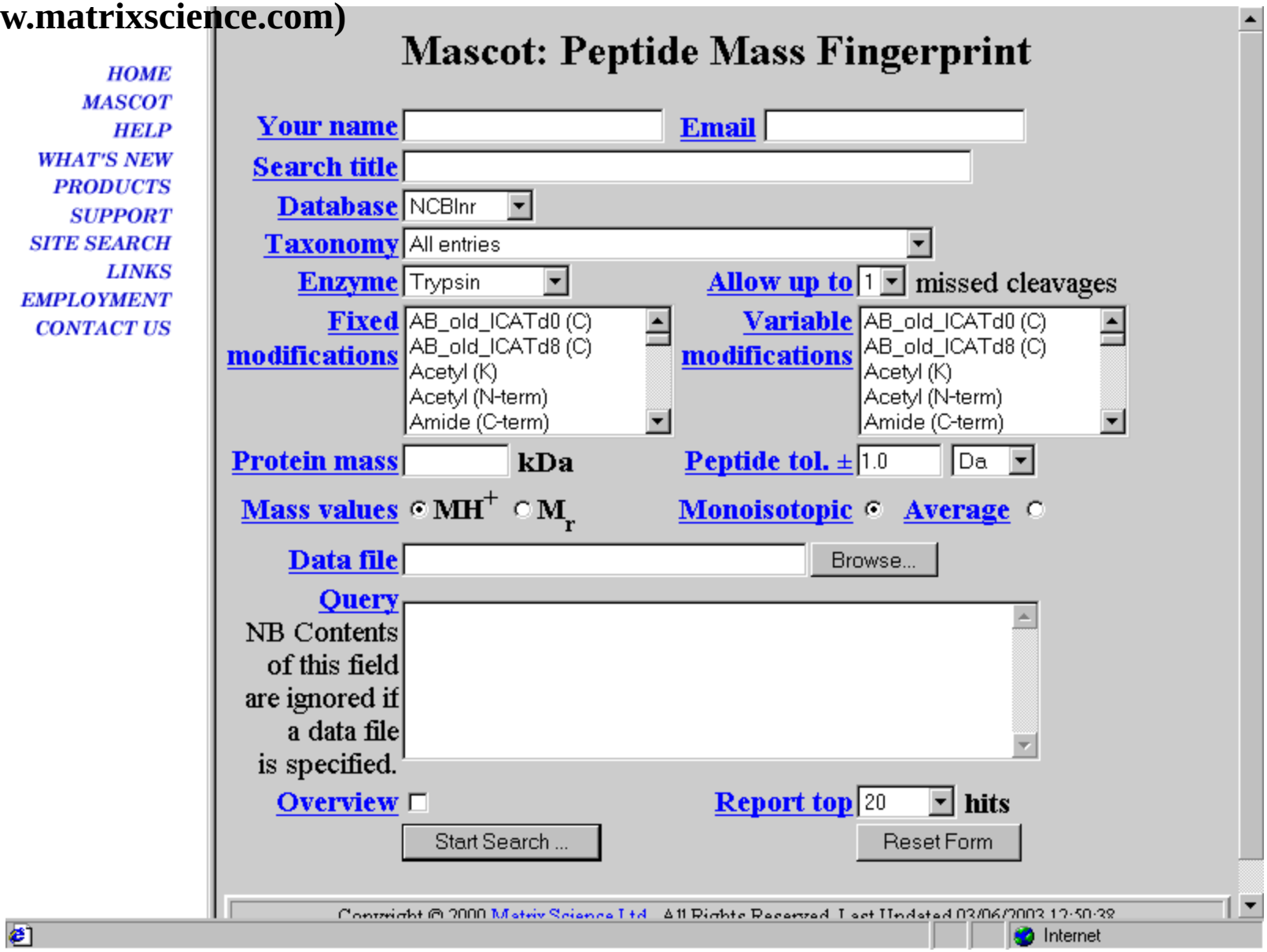
	Exp	Theo	Intensity			Delta		Dev	MC	CAM	MSO	PTM	Position			Sequence
	Da	Da	UI	%	rank	Da	ppm	ppm					start		end	
	874.479196	874.416572	1	100	1	0.06	72	9	-	-	-		33	-	39	HGLDNYR
	1428.718073	1428.6502	1	100	1	0.07	48	3	-	-	-		52	-	63	FESNFNTQATNR
	1753.920078	1753.835092	1	100	1	0.08	48	9	-	-	-		64	-	79	NTDGSTDYGILQINSR
*	993.45726	993.39954	1	100	1	0.06	58	1	-	1/1	-		80	-	86	WWCNDGR
	1675.872109	1675.8009	1	100	1	0.07	42	2	-	-	0/1		116	-	130	IVSDGNGMNAWVAWR
	1045.614758	1045.542498	1	100	1	0.07	69	14	-	-	-		135	-	143	GTDVQAWIR

View all couples (1 outliers)

```
  1  ---------- --------kv fgrcelaaam krHGLDNYRg yslgnwvcaa kFESNFNTQA TNRNTDGSTD YGILQINSRW
 81  WCNDGRtpgs rnlcnipcsa llssditasv ncakkIVSDG NGMNAWVAWR nrckGTDVQA WIRgcrl
```

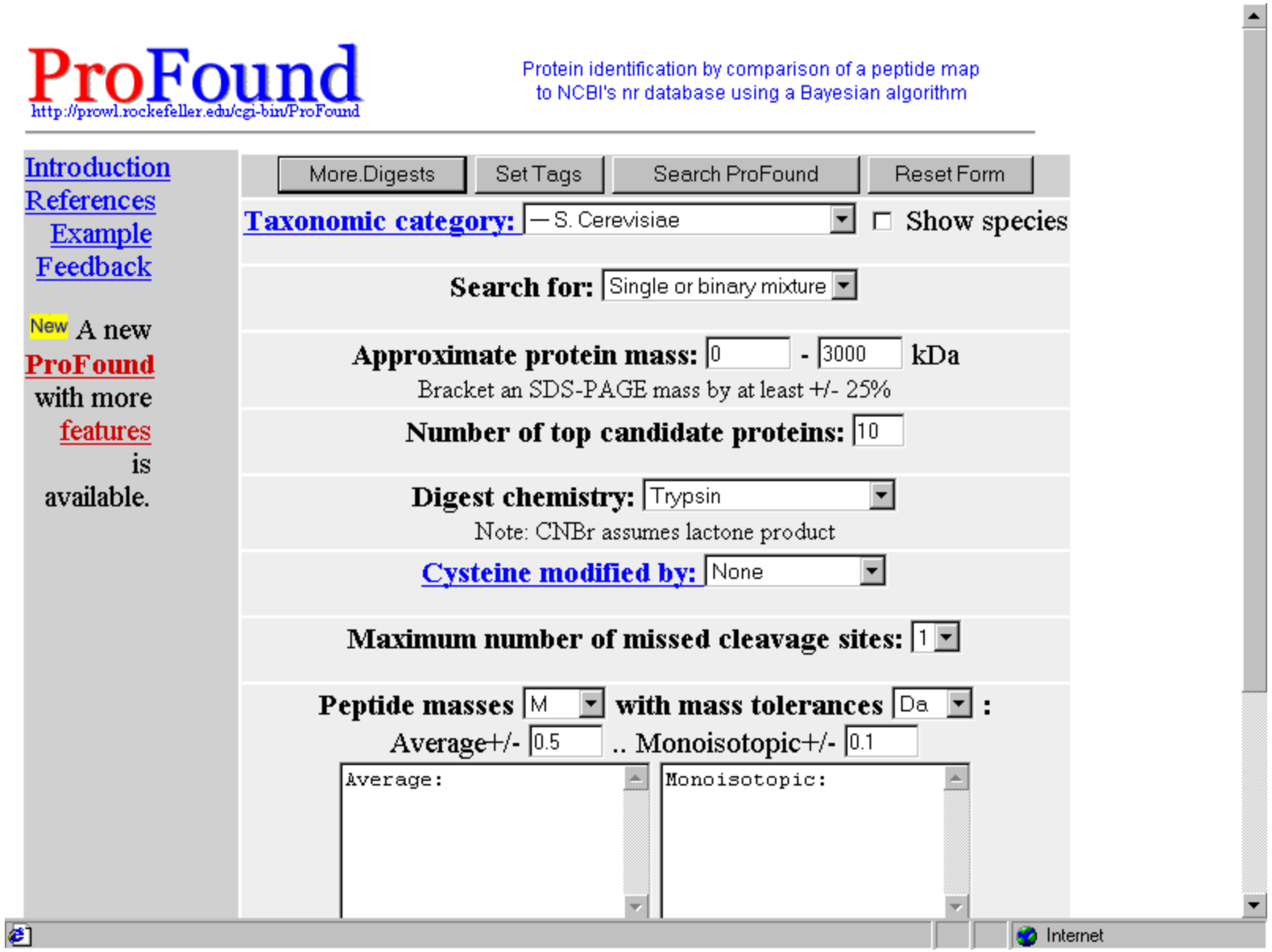
Search Engine for protein identification: MASCOT

(www.matrixscience.com)



Search Engine for protein identification: ProFound

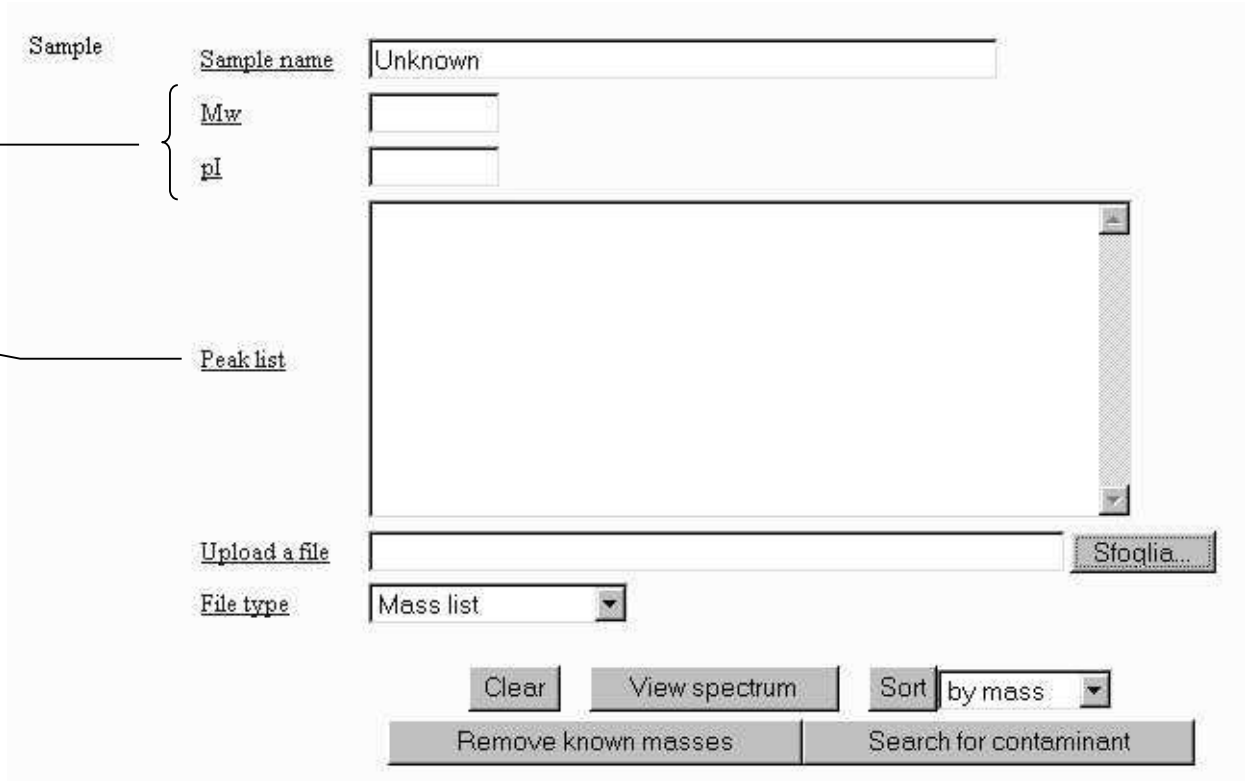
(http://prowl.rockefeller.edu)



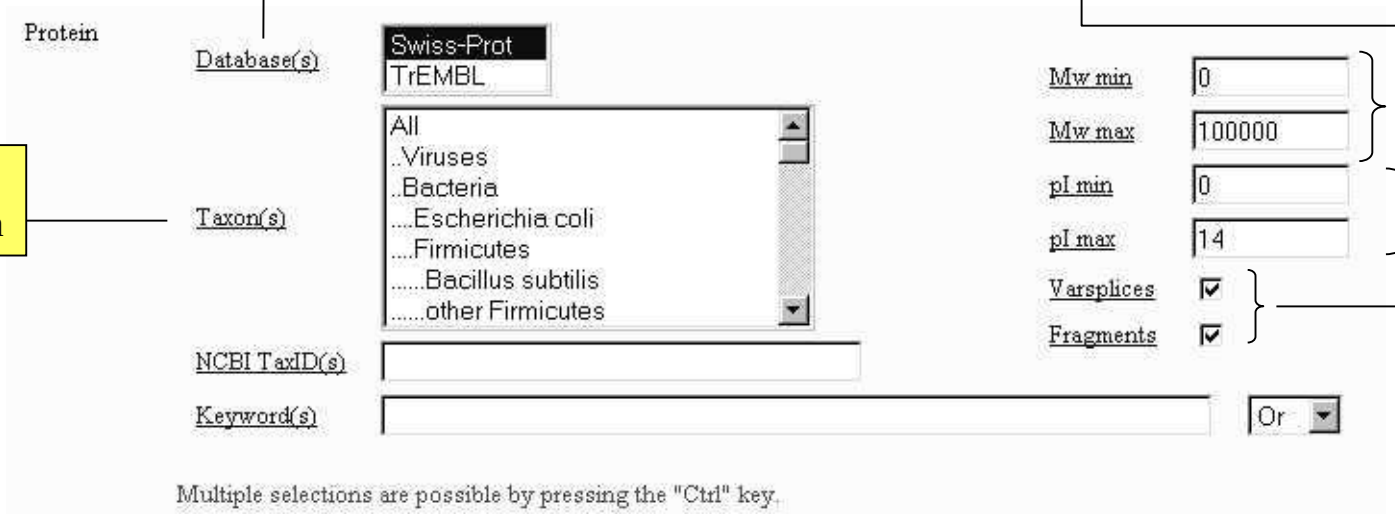
PEPTIDE FINGERPRINTING SOFTWARE (1)

MW & pI: information about 2D el.

Peak list: lits of MW of peptides from in gel digestion



PEPTIDE FINGERPRINTING SOFTWARE (2)



Include in the search only proteins in a MW range

Database to perform the search

Source of the protein

Include in the search also splicing variants or fragments of a protein present in a database

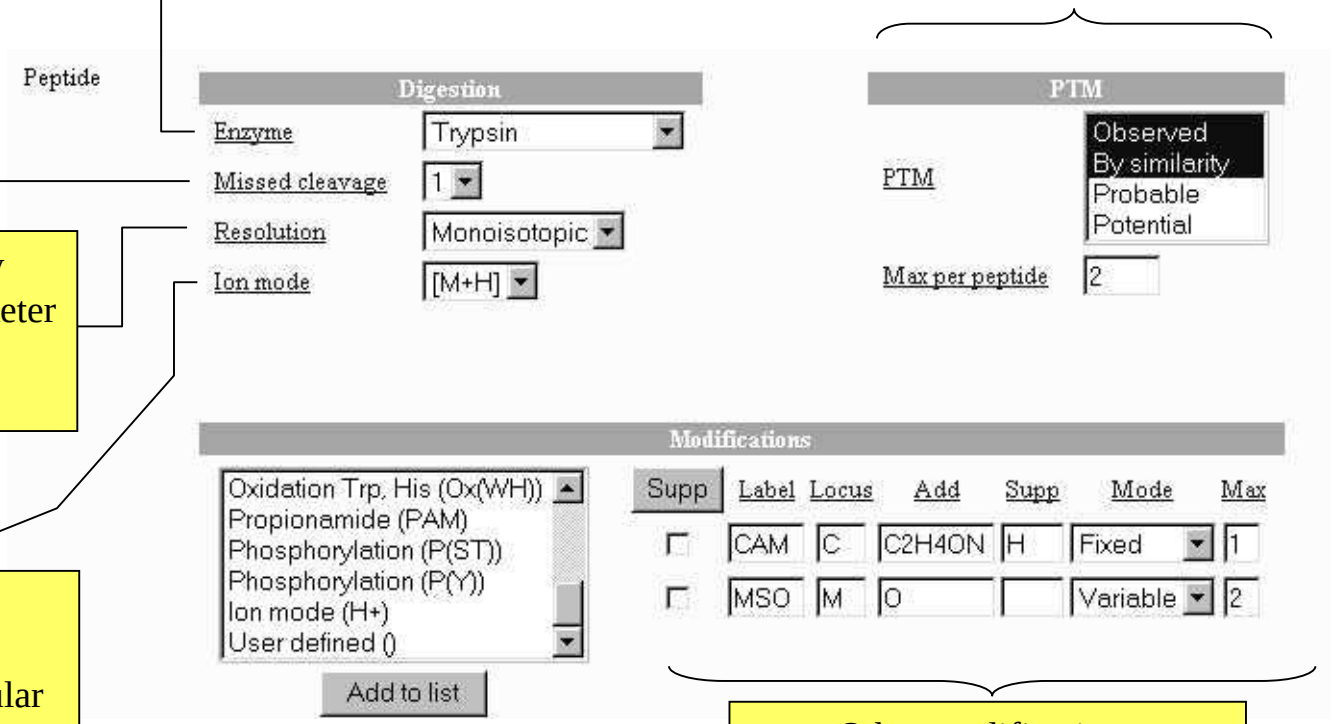
Include in the search only proteins in a pI range

PEPTIDE FINGER PRINTING SOFTWARE (3)

Number of missing clevages allowed for each peptide

Which PTM are kept into consideration

Enzyme used for the digestion



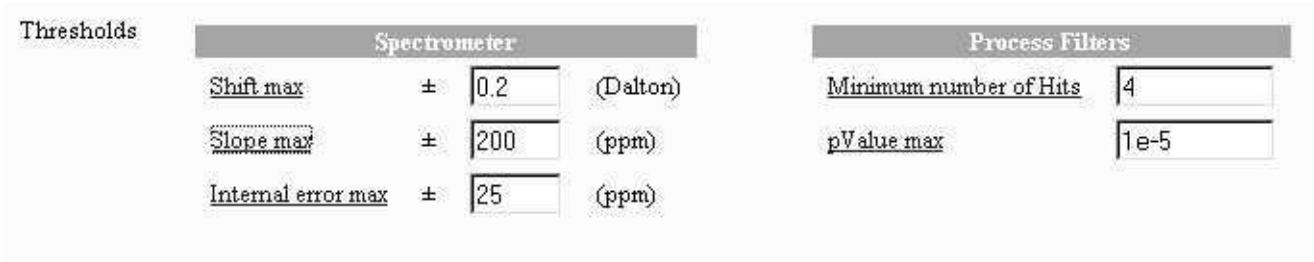
Mass provided by the mass spectometer (average or monoisotopyc)

Ion modality: molecular ion [M+H] or molecular weight [M]

Other modifications

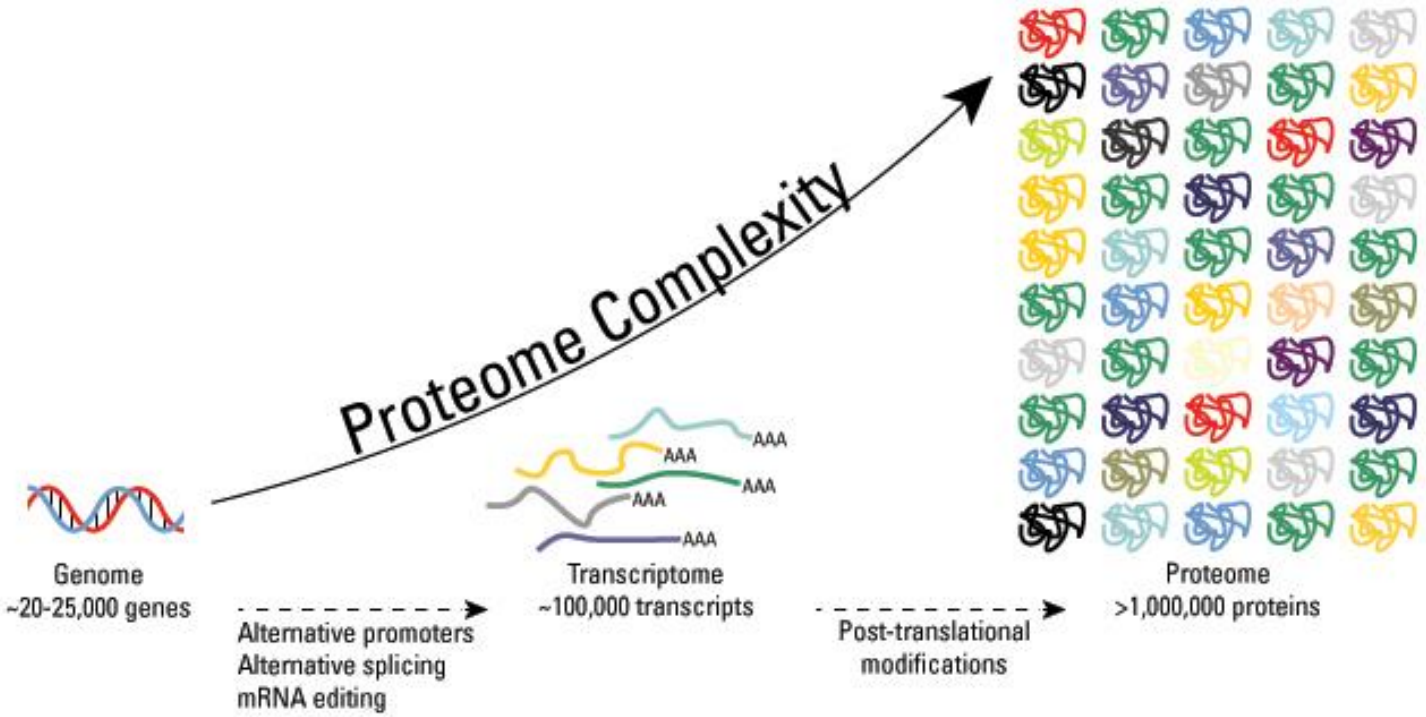
Post-translational modifications allowed to the peptide

PEPTIDE FINGER PRINTING SOFTWARE (4)



Information about precision of the values obtained with mass spectrometer

Information about the identification process. Minimum number of peptides that must be mapped on the sequence of the protein- exclusion from the results of those proteins who do not have a sufficient score


Proteome Complexity
AAA
AAA
AAA
AAA
Genome
~20-25,000 genes
Alternative promoters
Alternative splicing
mRNA editing
Transcriptome
~100,000 transcripts
Post-translational
modifications
Proteome
>1,000,000 proteins

Post Translational Modifications

Post-translational modification	Mass difference (Da)
Methylation	14.03
Propylation	42.08
Sulfation	80.06
Phosphorylation	79.98
Glycosylations by:	
Deoxyhexoses (Fuc)	146.14
Hexosamines (GlcN, GalN)	161.16
Hexoses (Glc, Gal, Man)	162.14
N-Acetylhexosamines (GlcNAc, GalNAc)	203.19
Pentoses (Xyl, Ara)	132.12
Sialic acid (NeuNAc)	291.26
Reduction of a disulfide bridge	2.02
Carbamidomethylation	57.03
Carboxymethylation	58.04
Cysteinylation	119.14
Ethylpyridylation	105.12
Acetylation	42.04
Formylation	28.01
Biotinylation	226.29
Farnesylation	204.36
Myristoylation	210.36
Pyridoxal phosphate Schiff condensation	231.14
Stearoylation	266.47
Palmitoylation	238.41
Lipoylation	188.30
Carboxylation of Asp or Glu	44.01
Deamidation of Asn or Gln	0.98
Hydroxylation	16.00
Met Oxidation	16.00
Proteolysis of a peptide bond	18.02
Deamination from Gln to pyroglutamic	-17.03

PTMs

Arg, Lys

Ser, Thr, Tyr

Cys

Lys

Most common PTM

Tutte queste modificazioni post-traduzionali alterano il peso molecolare della proteina, alcune di esse ne modificano anche il punto isoelettrico: Acetilazione e Fosforilazione
Acetilazione: rimuove una carica positiva
Fosforilazione: introduce due cariche negative

hHMGA1a

[1-106] mass = 11544.8

Cleavage at KR

Small polar:	D(1)	E(14)	N(1)	Q(6)			
Large polar:	K(16)	R(11)	H(0)				
Small non-polar:	S(14)	T(8)	A(4)	G(11)			
Large non-polar:	L(3)	I(1)	V(3)	M(0)	F(0)	Y(0)	W(0)
Special:	C(0)	P(13)					

```
  1  S E S S S K s s q p l a s k Q E K d g t e k R g r G R P R k  30
 31  Q P P V S P G T A L V G S Q K e p s e v p t p k R P R g r p  60
 61  k G S K n k G A A K t r K t t t t p g r K P R g r p k K l e  90
 91  k E E E E G I S Q E S S E E E Q                             106
```

(1)	[1-6] = 623.6	(2)	[7-14] = 816.9	(3)	[15-17] = 403.4
(4)	[18-22] = 548.5	(5)	[23-23] = 174.2	(6)	[24-25] = 231.3
(7)	[26-29] = 484.6	(8)	[30-30] = 146.2	(9)	[31-45] = 1465.7
(10)	[46-54] = 983.1	(11)	[55-57] = 427.5	(12)	[58-61] = 456.5
(13)	[62-64] = 290.3	(14)	[65-66] = 260.3	(15)	[67-70] = 345.4
(16)	[71-72] = 275.3	(17)	[73-73] = 146.2	(18)	[74-80] = 732.8
(19)	[81-83] = 399.5	(20)	[84-87] = 456.5	(21)	[88-88] = 146.2
(22)	[89-91] = 388.5	(23)	[92-106] = 1738.6		

hHMGA1a

[1-106] mass = 11544.8
Cleavage at KR

UNMODIFIED PROTEIN

Small polar:	D(1)	E(14)	N(1)	Q(6)			
Large polar:	K(16)	R(11)	H(0)				
Small non-polar:	S(14)	T(8)	A(4)	G(11)			
Large non-polar:	L(3)	I(1)	V(3)	M(0)	F(0)	Y(0)	W(0)
Special:	C(0)	P(13)					

1	S	E	S	S	S	K	s	s	q	p	l	a	s	k	Q	E	K	d	g	t	e	k	R	g	r	G	R	P	R	k	30
31	Q	P	P	V	S	P	G	T	A	L	V	G	S	Q	K	e	p	s	e	v	p	t	p	k	R	P	R	g	r	p	60
61	k	G	S	K	n	k	G	A	A	K	t	r	K	t	t	t	t	p	g	r	K	P	R	g	r	p	k	K	l	e	90
91	k	E	E	E	E	G	I	S	Q	E	S	S	E	E	E	Q															106

(1)	[1-6] = 623.6	(2)	[7-14] = 816.9	(3)	[15-17] = 403.4
(4)	[18-22] = 548.5	(5)	[23-23] = 174.2	(6)	[24-25] = 231.3
(7)	[26-29] = 484.6	(8)	[30-30] = 146.2	(9)	[31-45] = 1465.7
(10)	[46-54] = 983.1	(11)	[55-57] = 427.5	(12)	[58-61] = 456.5
(13)	[62-64] = 290.3	(14)	[65-66] = 260.3	(15)	[67-70] = 345.4
(16)	[71-72] = 275.3	(17)	[73-73] = 146.2	(18)	[74-80] = 732.8
(19)	[81-83] = 399.5	(20)	[84-87] = 456.5	(21)	[88-88] = 146.2
(22)	[89-91] = 388.5	(23)	[92-106] = 1738.6		

hHMGA1a

[1-106] mass = 11784.7

Cleavage at KR

PROTEIN MODIFIED WITH 3 PHOSPHATE GROUPS

Small polar:	D(1)	E(14)	N(1)	Q(6)			
Large polar:	K(16)	R(11)	H(0)				
Small non-polar:	S(14)	T(8)	A(4)	G(11)			
Large non-polar:	L(3)	I(1)	V(3)	M(0)	F(0)	Y(0)	W(0)
Special:	C(0)	P(13)					

S[98] + 80.0 S[101] + 80.0 S[102] + 80.0

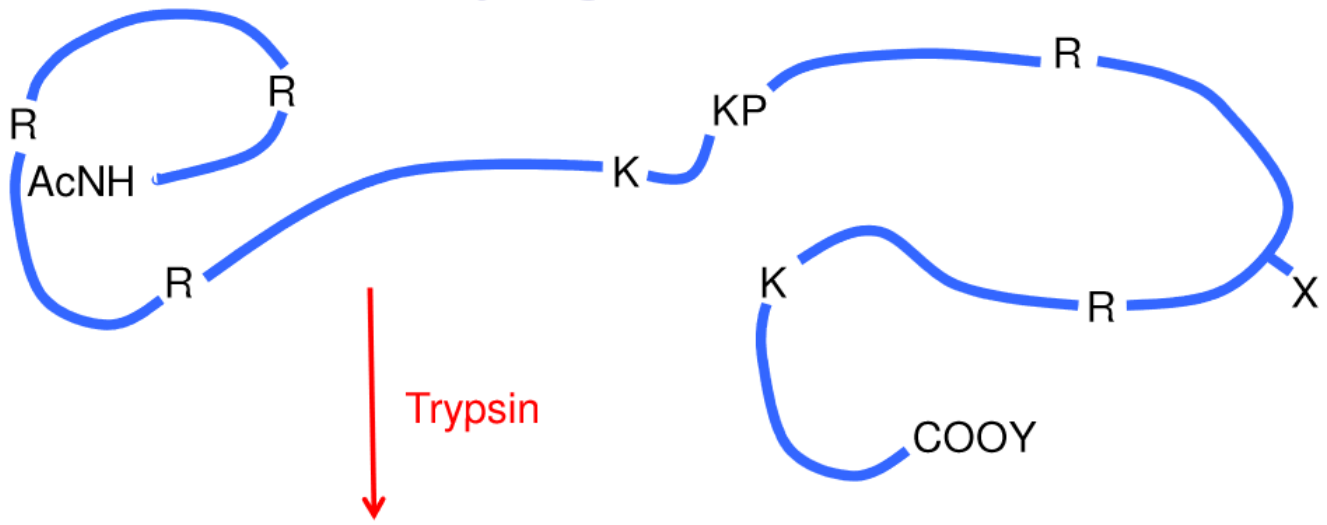
PHOSPHORILATED SITES: + 80 Da

1	S	E	S	S	S	K	s	s	q	p	l	a	s	k	Q	E	K	d	g	t	e	k	R	g	r	G	R	P	R	k	30
31	Q	P	P	V	S	P	G	T	A	L	V	G	S	Q	K	e	p	s	e	v	p	t	p	k	R	P	R	g	r	p	60
61	k	G	S	K	n	k	G	A	A	K	t	r	K	t	t	t	t	p	g	r	K	P	R	g	r	p	k	K	l	e	90
91	k	E	E	E	E	G	I	S	Q	E	S	S	E	E	E	Q															106

(1)	[1-6] = 623.6	(2)	[7-14] = 816.9	(3)	[15-17] = 403.4
(4)	[18-22] = 548.5	(5)	[23-23] = 174.2	(6)	[24-25] = 231.3
(7)	[26-29] = 484.6	(8)	[30-30] = 146.2	(9)	[31-45] = 1465.7
(10)	[46-54] = 983.1	(11)	[55-57] = 427.5	(12)	[58-61] = 456.5
(13)	[62-64] = 290.3	(14)	[65-66] = 260.3	(15)	[67-70] = 345.4
(16)	[71-72] = 275.3	(17)	[73-73] = 146.2	(18)	[74-80] = 732.8
(19)	[81-83] = 399.5	(20)	[84-87] = 456.5	(21)	[88-88] = 146.2
(22)	[89-91] = 388.5	(23)	[92-106] = 1978.6		

MODIFIED PEPTIDE: MW CHANGES

Identifying Modifications



1 AcNH — R — Predicted mass + 42

2 — R — Predicted mass OK

3 — R — Predicted mass OK

4 — K — Predicted mass OK

5 (— K) — Probably no peptide

6 — KP — R — Missed cleavage

7 — X — R — "Wrong" mass

8 — K — Predicted mass OK

9 — COOY — "Wrong" mass