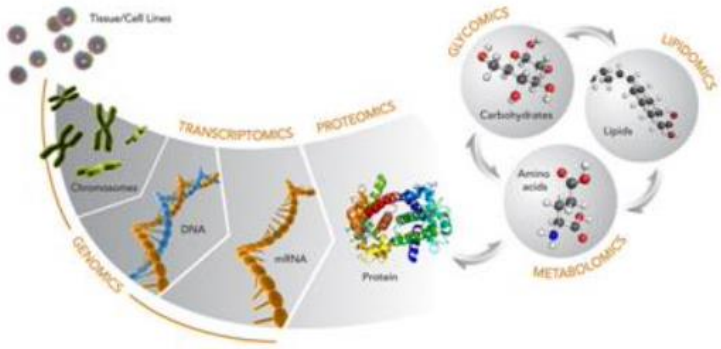
Review:

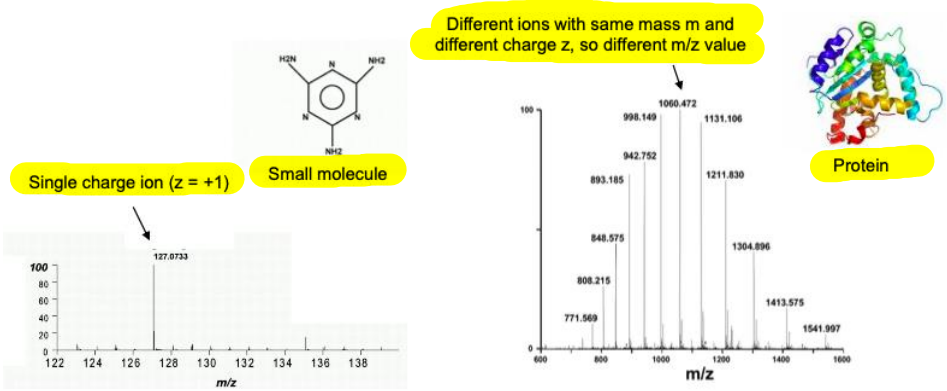
Mass spectrometry techniques measure the mass/charge ratio of analytes. They enable both IDENTIFICATION and QUANTIFICATION of analytes by measuring molecular mass (m/z) or their fragmentation products.

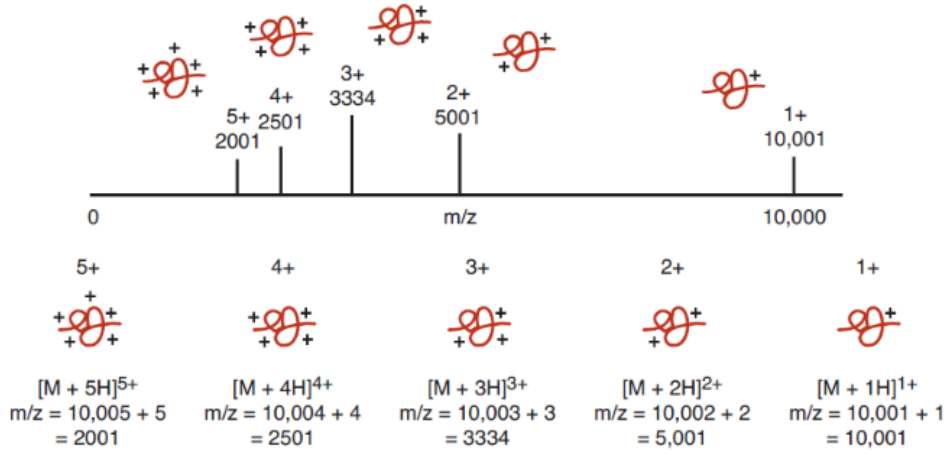
Mass spectrometry (MS), especially combined with chromatography and computer technologies, has been widely used in organic chemistry, biochemistry, drug metabolism, clinical, toxicology, determination of pesticides, environmental protection, petroleum chemistry, geochemistry, food chemistry, chemical plants, cosmic chemistry and chemical defense and other fields.

Different omics take advantage of mass spectrometry, and all of these share the electron magnetics that has effect on ions and the idea is to ionize the molecules, provide the mass/charge value (we have different ions and different charges) and all these signals are same masses but different charges.



A mass spectrometer is based on electromagnetic fields, acting only on ionic species. In fact, electrically charged particles are affected by a magnetic field while electrically neutral ones aren't. A mass spectrometer measures the m/z value and species able to produce ions with different charges will thus produce more signals at different m/z values of the mass spectrum (this is not an issue for small molecules, forming only single charged ions).





The figure shows that a theoretical protein with a molecular weight of 10,000 can be multiply charged, which will generate numerous peaks. A mass spectrometer with a relatively small mass range can detect the multiply charged ions since the m/z is reduced.

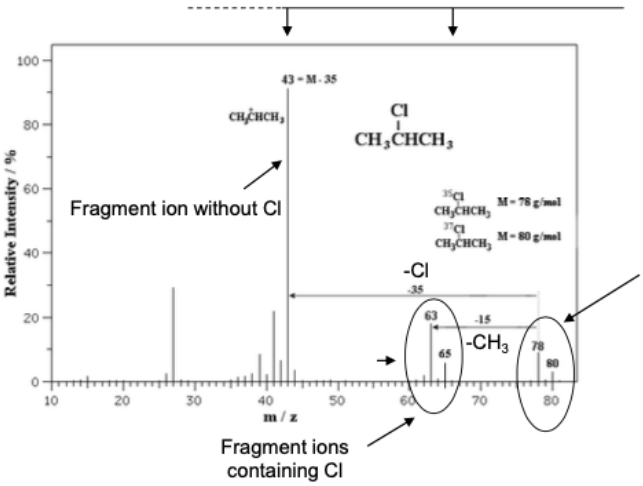
The Mass spectrum

The mass spectrum of a compound graphically represents masses (or better m/z values) of all ions that are obtained analyzing the compound with a mass spectrometer in given experimental conditions.
It represents the ID of a compound and contains all information necessary for its identification and for understanding its molecular structure.

FRAGMENT IONS: molecular ions formed in the initial ionization dissociate to fragment ions (with minor mass) because of excess internal energy remaining after ionization.

MOLECULAR ION: is the ion deriving from the original species in its intact form, but it does not necessarily represent the most intense signal of the mass spectrum.

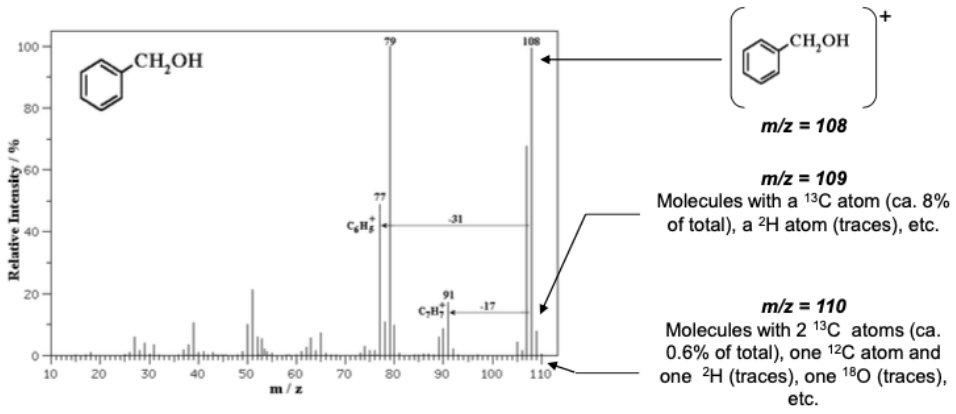
The mass may differ from the nominal one (e.g., H addition or removal produce masses of m + 1 or m - 1).
In this case two molecular ions are present with masses of m and m + 2, corresponding to two molecules with Cl isotopes 35Cl e 37Cl.



Analyte Identification:

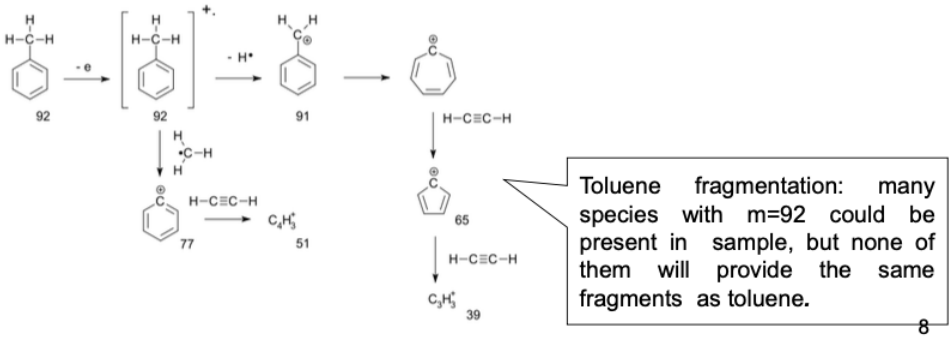
The easiest way to **identify** an analyte from a mass spectrum is using the **m/z value**.

M/z values are dependent on the exact molecular mass, therefore mass spectrometry enables to identify different isotopes of the analyte.
One molecule will produce signals corresponding to all possible isotopic forms (e.g., presence of 13C is responsible for isotopes of most organic molecules).

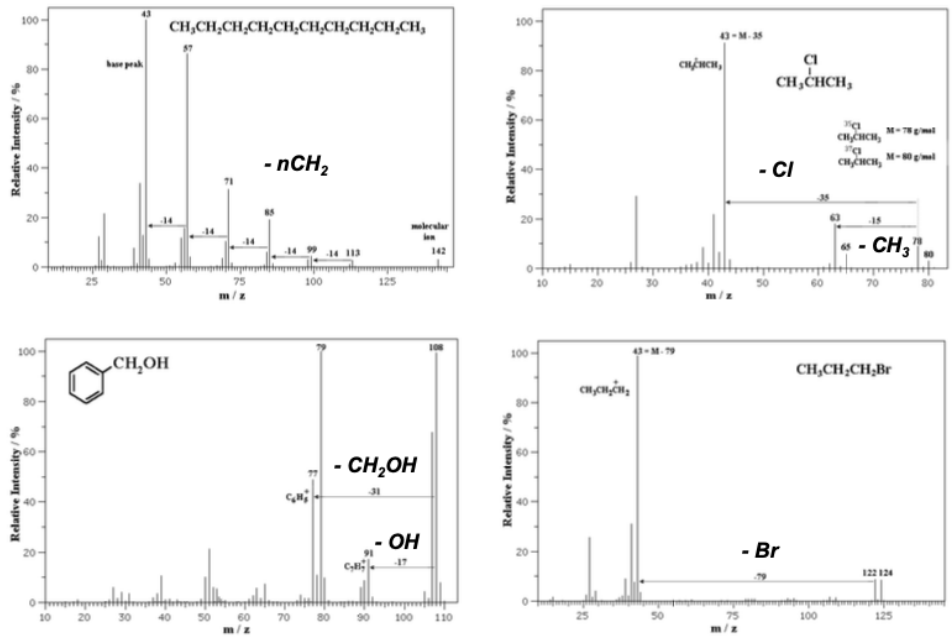


To achieve absolute identification of a target analyte with mass spectrometry measurements performed at low resolution, **fragmentation pattern** is used.

By providing high energy, ions of minor size are produced depending on the original structure of the molecule. These ions are specific for each species and enable its identification.



Also in this case, with toluene, we have a typical fragmentation pattern: it represents the fingerprint of this molecule, for classical metabolomics application and analysis we need fragmentation during ionization. Whereas for proteins they are too big and the ionization/fragmentation would be there to just keep it intact. The easiest tool to apply for peptide mass fingerprinting is the MALDI-TOF.

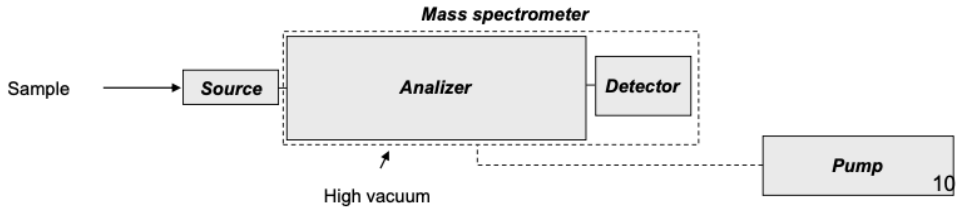


Instrument has three main components:

•**Ionization source**: converts the analyte into an ionized form in the gas phase, thus suitable to be analyzed with the mass spectrometer.
•**Analyzer**: selects ions according to their m/z and sends them to the detector.
•**Detector**: produces a signal proportional to the ions that arrive.

For standard mass spectrometry we have other ion sources that can be used: we have Electron Impact (EI), which is the most used.

Movement of the ions requires high vacuum (10^{-5} - 10^{-6} mBar): coupling with ionization sources working at atmospheric pressures represents a technical challenge.



IONIZATION SOURCES – ELECTRON IMPACT (EI)

The sample for analysis is introduced into the ion source, either through a solid inlet or through a gas chromatography column. A beam of electrons produced by a heated filament of either Tungsten or Rhenium collides with the sample gas molecules, removes an electron and produces a positively charged ion corresponding to the relative molecular mass of the sample being analyzed.

Electron impact is an energetic ionization technique and produces fragment ions which are smaller parts of the original molecule.

$$M + e^- (70\ eV) \rightarrow \underbrace{M^{+\bullet}}_{\text{Molecular ion}} + e^- + e^-$$

$$\text{Molecular ion} \longrightarrow \text{Fragments}$$

Electron impact sources generally produce positive ions. In case target molecule contains halogens (electronegative elements such as Cl, Br, F) negative ions can be produced but the process is less efficient.

$$M + e^- (70\ eV) \rightarrow \underbrace{M^{-\bullet}}_{\text{Molecular ion}}$$

$$\text{Molecular ion} \longrightarrow \text{Fragments}$$

Ionization sources-chemical ionization (CI):

CI--> lower ionization energy and generally quite used for metabolites.

Chemical ionization is a low energy ionization process that avoids excessive fragmentation. CI uses a reagent gas to gently transfer protons to the sample, usually producing (M+H)+ quasimolecular ions (close to the starting ion). Such ions have very little tendency to fragment because little excess of energy is imparted to them. The reagent gas (methane, isobutane, or

ammonia) is present in the ion source at a pressure of 1 torr. It is ionized by an electron beam and the resulting ions undergo a complex series of ion-molecule reactions to produce species such as CH_5^+ and methane.

$$CH_4 + e^- (70\,eV) \rightarrow CH_4^{+\bullet} + 2e^-$$
$$CH_4^{+\bullet} + CH_4 \rightarrow CH_5^+ + CH_3^\bullet$$

Cation CH_5^+ is very acidic and can transfer a proton to every organic molecule. Produced ions have a low energy content and generally do not fragment.

$$CH_5^+ + M \rightarrow MH^+ + CH_4$$

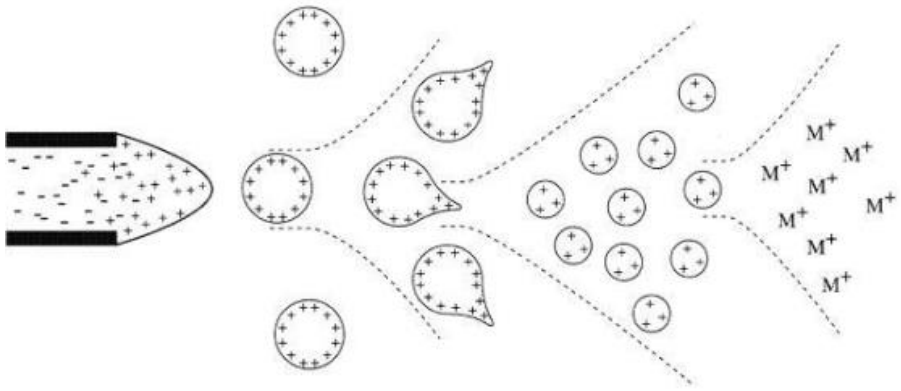
IONIZATION SOURCES - ELECTROSPRAY

ESI uses electrical energy to assist the transfer of ions from solution into the gaseous phase before they are subjected to mass spectrometric analysis.
The transfer of ionic species from solution into the gas phase by ESI involves three steps: (1) dispersal of a fine spray of charge droplets, followed by (2) solvent evaporation and (3) ion ejection from the highly charged droplets.

Sample is nebulized through a needle maintained at a high voltage (e.g. 2.5 – 6.0 kV) negative or positive depending on the type of ions (positive for cations, negative for anions). The application of a nebulizing gas (e.g. nitrogen) at high temperature, which shears around the eluted sample solution, enhances a higher sample flow rate.

The charged droplets are continuously reduced in size by evaporation of the solvent, leading to an increase of surface charge density and a decrease of the droplet radius. Finally, the electric field strength within the charged droplet reaches a critical point at which it is kinetically and energetically possible for ions at the surface of the droplets to be ejected into the gaseous phase.

Lower energy ionization, generally quite used for metabolites.



Mechanism of electrospray ionization

The emitted ions are sampled by a sampling skimmer cone and are then accelerated into the mass analyzer for subsequent analysis of molecular mass and measurement of ion intensity.

We have sample in liquid form, and when sprayed in the magnetic forces it charges, we have multiple charged ions but with MALDI we have singly charged ions, with MALDI we need to manually pipe and spray. The MALDI is more tolerant to the contaminations so it's more used for biological and dirty samples as well.

ANALYZERS

A mass analyzer is the component of the mass spectrometer that takes ionized masses and separates them based on mass-to-charge ratios and outputs them to the detector where they are detected and later converted to a digital output.

- Quadrupole Mass Analyzer
- Time-of-Flight Mass Analyzer (TOF)
- Magnetic Sector Mass Analyzer
- Electrostatic Sector Mass Analyzer
- Quadrupole Ion Trap Mass Analyzers
- Ion Cyclotron Resonance

Applications:

In theory mass spectrometry allows to analyze complex samples (without pre-treatment) but resulting spectra would be too complex for correct interpretation. Therefore, separation techniques are required to simplify the output and remove contaminants when complex samples are analyzed.

• Coupling with gas-chromatography: for the analysis of volatile substances with ionization obtained with electron impact or chemical ionization.

• Coupling with HPLC: for the analysis of non-volatile organic molecules with ESI

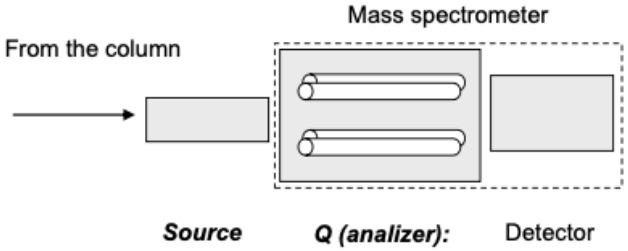
Exception: elemental analysis techniques based on the use of atomization sources such as inductively coupled plasma torch coupled to mass spectrometry (ICP-MS techniques). In this case, the relative simplicity of mass spectra (which only contain atomic species) allows this technique for direct sample analysis without any prior separation.

Mass-spectrometry detection techniques coupled with separation techniques have significant advantages over other conventional detection techniques:

- Mass spectrometry is **universal**: any substance capable of being ionized can be detected, even with variable detection limits.
- Mass spectrometry is **sensitive**: its performance is often comparable or better than that of any other available detector.
- It can be **applied to all separation techniques**: by modifying the ionization source, it can be used for elemental analysis (ICP-MS), gas chromatography (GC-MS) and liquid chromatography (HPLC-MS).
- **Allows both quantification and identification**, based on the molecular mass or fragmentation spectrum. In fact, it is also able to distinguish and quantify co-eluted substances (i.e. not separated from the chromatographic column): the separating capacity of the chromatographic system is no longer a limit to the ability to identify the compounds present in the sample.

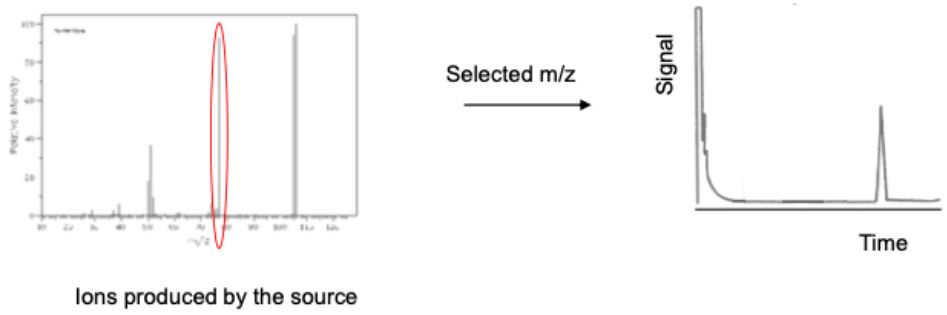
CONFIGURATIONS: SINGLE QUADRUPOLE

Single-analyzer mass spectrometers (in this case single-quadrupole) are the simplest and cheapest tools.



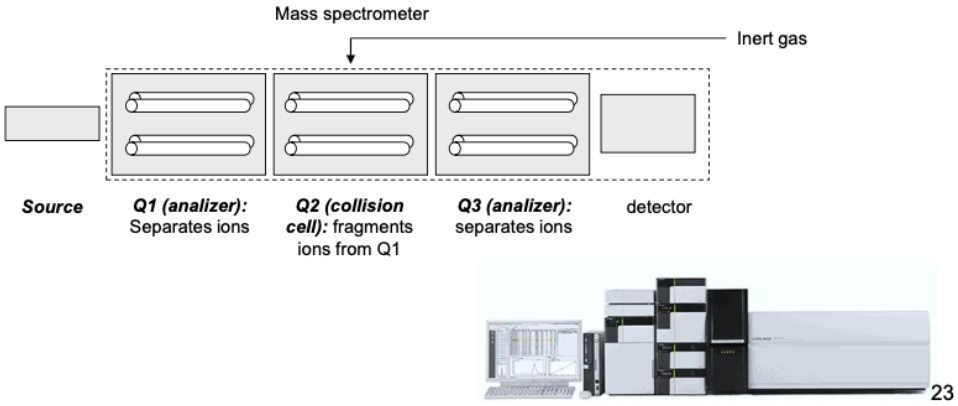
Single-quadrupole mass spectrometer interfaced with an HPLC chromatograph

A single quadrupole mass spectrometer allows you to scan the ions produced in the source in order to identify ions at specific m/z values, but also to monitor the ions corresponding to the target analytes during a chromatographic separation.
In this **Single Ion Monitoring (SIM)** mode, the quadrupole is set to pass only one ion with a given m/z ratio: every time this source is generated in the source, a signal is obtained in the chromatogram. This mode is suitable, for example, to obtain a chromatogram relevant only to the presence of the species under consideration. By rapidly changing different **Multiple Ion Monitoring (MIM)** values, different ions can be monitored simultaneously.

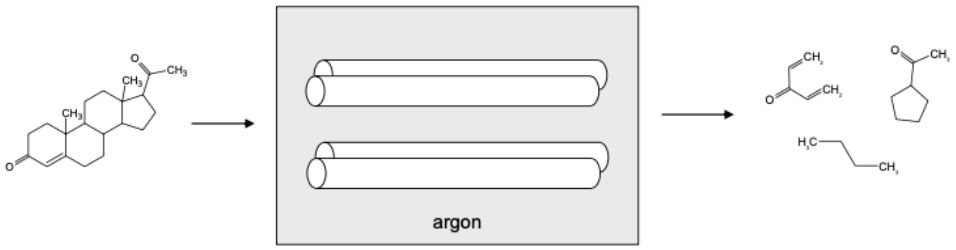


CONFIGURATIONS: TRIPLE QUADRUPOLE

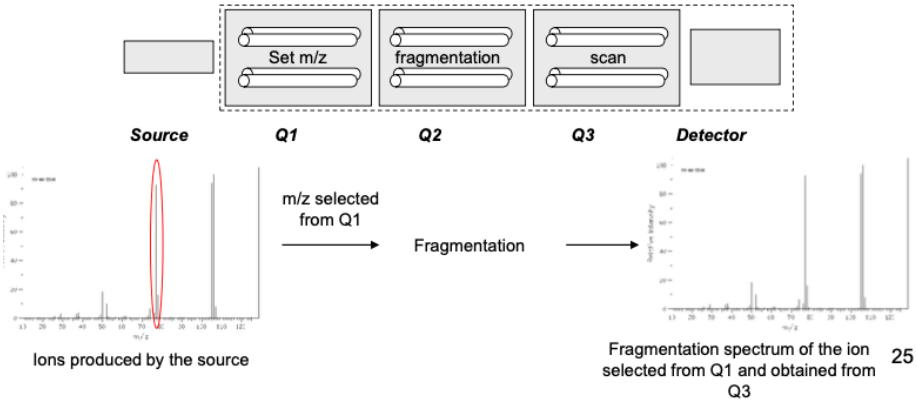
To be able to use mass spectrometry techniques based on the fragmentation of the produced ions, however, instruments with several series analyzers are needed. The most common of these is the **triple quadrupole mass spectrometer**, which is equipped with three independent quadruples. Two of them (Q1 and Q3) act as analyzers, while the intermediate quadrupole (Q2) acts as a collision cell.



The collision cell, which is structurally similar to a quadrupole (but may also be an example or other geometry) has the function of fragmenting the ions by accelerating them by a predetermined potential and colliding with an inert gas (usually argon). The electrodes within the cell have the function of accelerating the ions to a given energy (the magnitude of the fragmentation process depends on the energy of the ions themselves) and not allowing its dispersion during collision processes.

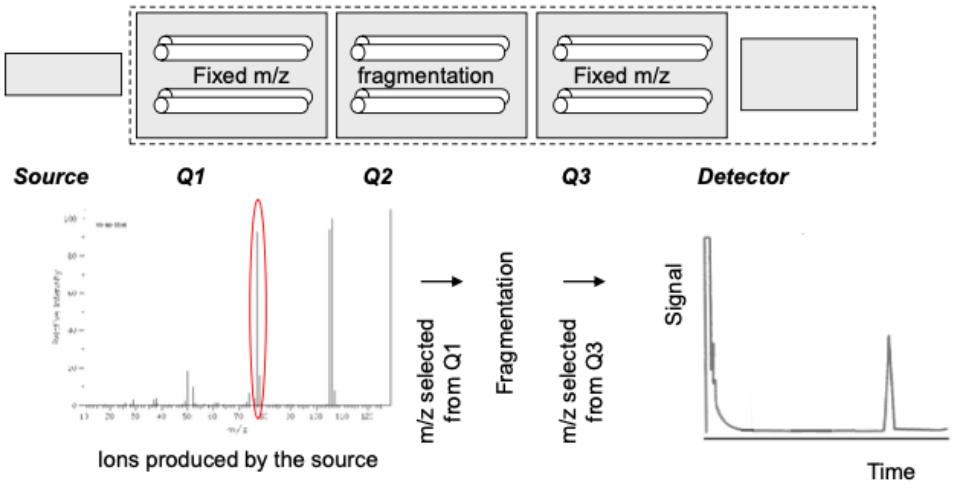


A triple quadrupole tool allows to study and exploit fragmentation phenomena. For example, it is possible to study the fragmentation spectrum of a specific ion (**Product Ion Scan**): the first quadrupole is set to pass only that ion, which is then fragmented into the collision cell. The third quadrupole scans the fragment ions by providing the fragmentation spectrum of the selected ion. This mode is suitable for confirming the **identity of an ion through the study of its fragmentation products.**



25

Maximum selectivity is achieved through **Single Reaction Monitoring (SRM)** and **Multiple Reaction Monitoring (MRM)** techniques, which are based on a dual selection of ions and are used for qualitative and quantitative analysis of the target species in low concentrations and in complex matrices. In these modes, the first quadrupole is set in order to allow passage of target ion, which is then fragmented into the collision cell. The second quadrupole then selects one (or more) fragment ions with specific m/z values.



In general, **techniques based on fragmentation are also called "tandem" or MS/MS mass techniques.**

The Multiple Reaction Monitoring (MRM) mode in which a source ion product is selected from the first quadrupole and multiple fragments are revealed with fixed m/z values is the one that

ensures maximum selectivity to confirm analyte identity. **High selectivity typically also has the lowest detection limits**. In "official" analysis, the detection of molecular ions and their fragments is considered the confirmatory and ultimate analysis.

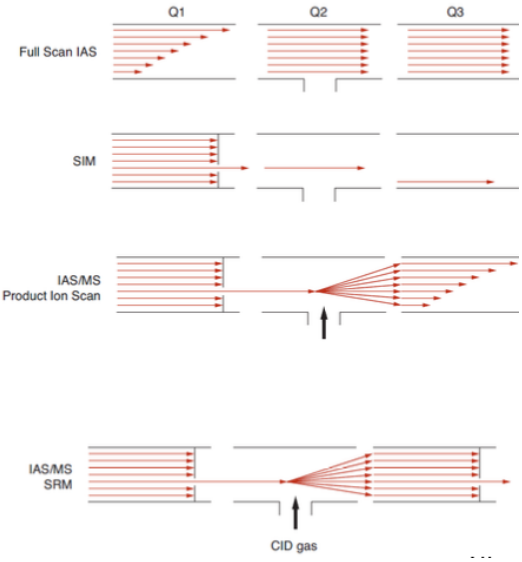
Scanning modes used in a triple quadrupole mass spectrometer:

A. Full scan detects all ions

B. **Selected ion monitoring (SIM)** detects ions of one selected *m/z*

C. **Product ion scans** select ions of one *m/z* in Q1 to pass on to Q2, the collision cell, where the ion is fragmented. All ion fragments are allowed to pass through to the detector

D. **Selected reaction monitoring (SRM)** is similar to the product ion scan, but only fragments of one selected m/z are allowed to pass on to the detector.



Peptide de novo sequencing:

Peptide de novo sequencing is the analytical process that derives a peptide's amino acid sequence from its tandem mass spectrum (MS/MS) without the assistance of a sequence database. ONLY FOR TANDEM MS.

It is in contrast with another popular peptide identification approach, "database search", which searches in each database to find the target peptide.

A clear advantage of de novo sequencing is that it works for both database and novel peptides.

In a tandem mass spectrometer, the peptide is fragmented along the peptide backbone and the resulting fragment ions are measured to produce the MS/MS spectrum. Depending on the fragmentation methods used, different fragment ion types can be produced. In PMF, we already have the peptides deriving from trypsin digestion.

The most widely used fragmentation methods today are Collision-Induced Dissociation (CID) and Electron-Transfer Dissociation (ETD).

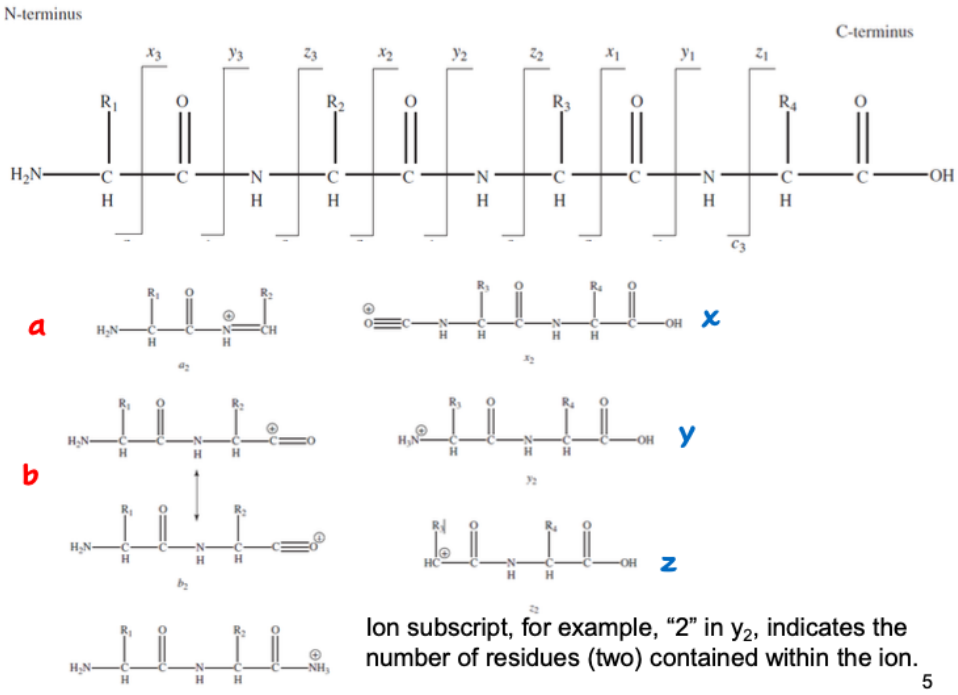
CID produces mostly b- and y-ions.

In most mass spectrometers used in proteomic studies the collision energy is considered low (5–50 eV), and the product ions are generally formed through cleavages of the peptide bonds.

According to the widely accepted nomenclature of Roepstorff and Fohlman:

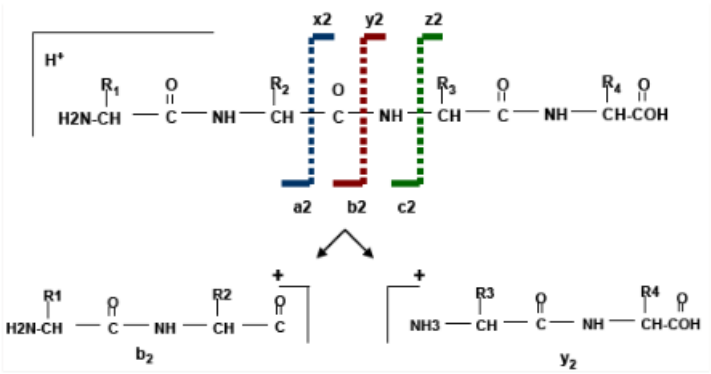
- When the charge is retained on the N-terminal portion of the fragmented peptide, the ions are depicted as a, b, and c.
- When the charge is retained on the C-terminal portion, the ions are denoted as x, y, and z.

Depending on the breaking of molecules we have ions, if the charge is in N or C terminal is A,B,C or X,Y,Z.



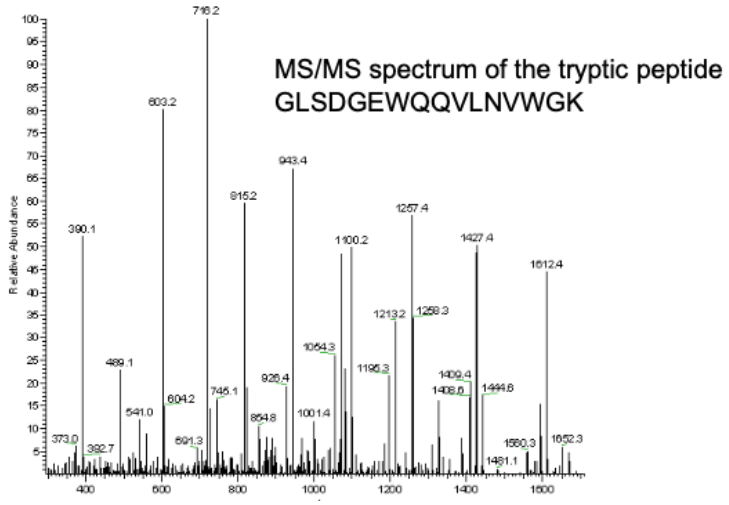
5

b and y ions:



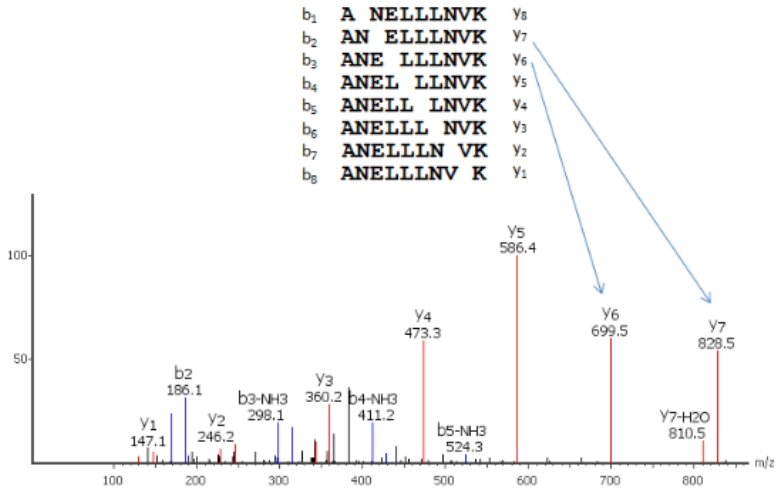
The most common peptide fragments observed in low energy collisions are a, b and y ions.

The b ions appear to extend from the N-terminus and y ions appear to extend from the carboxyl terminus. It is important to distinguish because, with MS/MS, we can read the sequence residue by residue.



In a CID MS/MS, many copies of the same peptide are fragmented at the peptide backbone to form b and y ions. The spectrum consists of peaks at the m/z values of the corresponding fragmented ions. A good quality spectrum often contains many (but not necessarily all) of the theoretical fragment ions.

The main idea of de novo sequencing is to use the mass difference between two fragment ions to calculate the mass of an amino acid residue on the peptide backbone. The mass can usually uniquely determine the residue. For example, the mass difference between the y7 and y6 ions is equal to 129, which is the mass of residue E. Similarly, the next adjacent residue between y6 and y5 can be determined as L by the mass difference. The main limitation of peptide de novo sequencing is that you have to proceed just from one side, otherwise you can cause mess.



We have peptides that we need to assign for each peak. For example, the difference between y7 and y6 is the amount that can be assigned to glutamic acid so this is how we can determine the amino acid sequence.

Glycine	Gly	G	57.02146	57.05	C_2H_3NO
Alanine	Ala	A	71.03711	71.08	C_3H_5NO
Serine	Ser	S	87.03203	87.08	$C_3H_5NO_2$
Proline	Pro	P	97.05276	97.12	C_5H_7NO
Valine	Val	V	99.06841	99.13	C_5H_9NO
Threonine	Thr	T	101.04768	101.1	$C_4H_7NO_2$
Cysteine	Cys	C	103.00919	103.1	C_3H_5NOS
Isoleucine	Ile	I	113.08406	113.2	$C_6H_{11}NO$
Leucine	Leu	L	113.08406	113.2	$C_6H_{11}NO$
Asparagine	Asn	N	114.04293	114.1	$C_4H_6N_2O_2$
Aspartic Acid	Asp	D	115.02694	115.1	$C_4H_5NO_3$
Glutamine	Gln	Q	128.05858	128.1	$C_5H_8N_2O_2$
Lysine	Lys	K	128.09496	128.2	$C_6H_{12}N_2O$
Glutamic Acid	Glu	E	129.04259	129.1	$C_5H_7NO_3$
Methionine	Met	M	131.04049	131.2	C_5H_9NOS
Histidine	His	H	137.05891	137.1	$C_6H_7N_3O$
Phenyalanine	Phe	F	147.06841	147.2	C_9H_9NO
Arginine	Arg	R	156.10111	156.2	$C_6H_{12}N_4O$
Tyrosine	Tyr	Y	163.06333	163.2	$C_9H_9NO_2$
Tryptophan	Trp	W	186.07931	186.2	$C_{11}H_{10}N_2O$

Rules:

- The b ion m/z value is basically the mass of the peptide minus OH, or -17u.
- To calculate the m/z value for the y ions just calculate the (M+H)1+ for the shortened peptide.

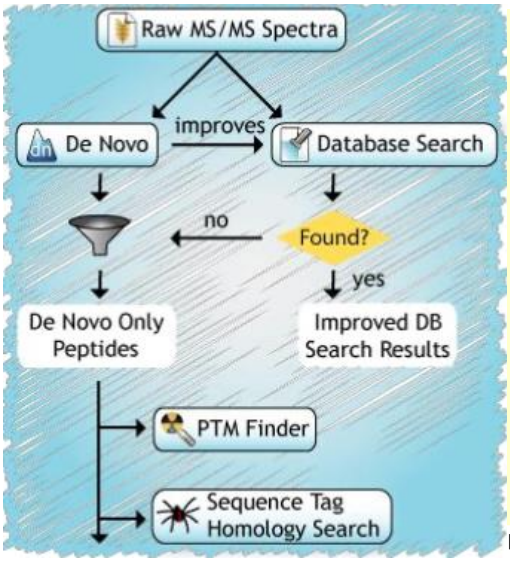
Thus, **if one can identify either the y-ion or b-ion series in the spectrum, the peptide sequence can be determined**. However, the spectrum obtained from the mass spectrometry instrument doesn't tell the ion types of the peaks, which require either a human expert or a computer algorithm to figure out during the process of de novo sequencing.

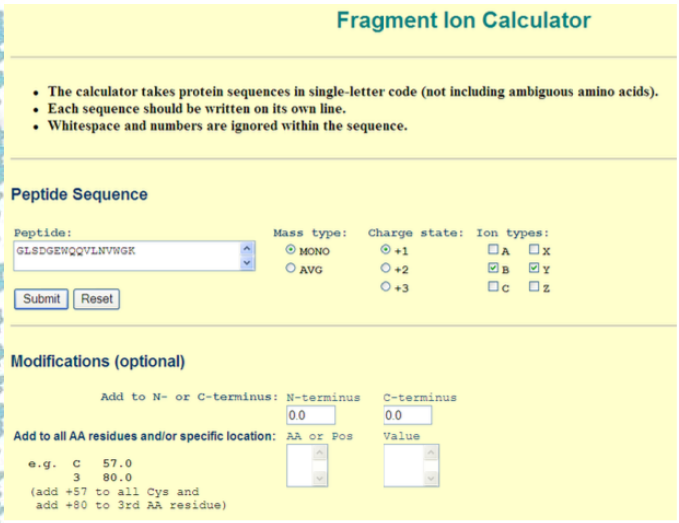
During this process, a few factors can cause difficulties:

- Incorrect assignment of y and b ions.
- Some fragment ions are missing (such as b1 and y8 in slide 8). You carry some replicates to have higher robustness.
- Existence of other fragment ion types (such as the b3-NH_3 ion).
- Existence of noise peaks in the spectrum.
- The same or similar mass of some residues may cause ambiguity (I=L and K=Q).
- The PTMs (post-translational modifications) on the residues may contribute to the mass ambiguity, as well as complicate the peptide fragmentation pattern.

These factors can cause de novo sequencing to figure out only a partially correct sequence tag from the spectrum.

There are some algorithms to assign ions to peptide chains, as sometimes it is not appropriate, and we find that the fragments are not appropriate. We can also find some peptides that are not ionized. We do a lot of replications to obtain more robust data, and correlate.





Free on-line calculator provided by the Institute of Systems Biology.

Fragment Ion Calculator Results

Sequence: **GLSDGEWQQVLNVWGK**, pI: **4.37029**

Fragment Ion Table, monoisotopic masses

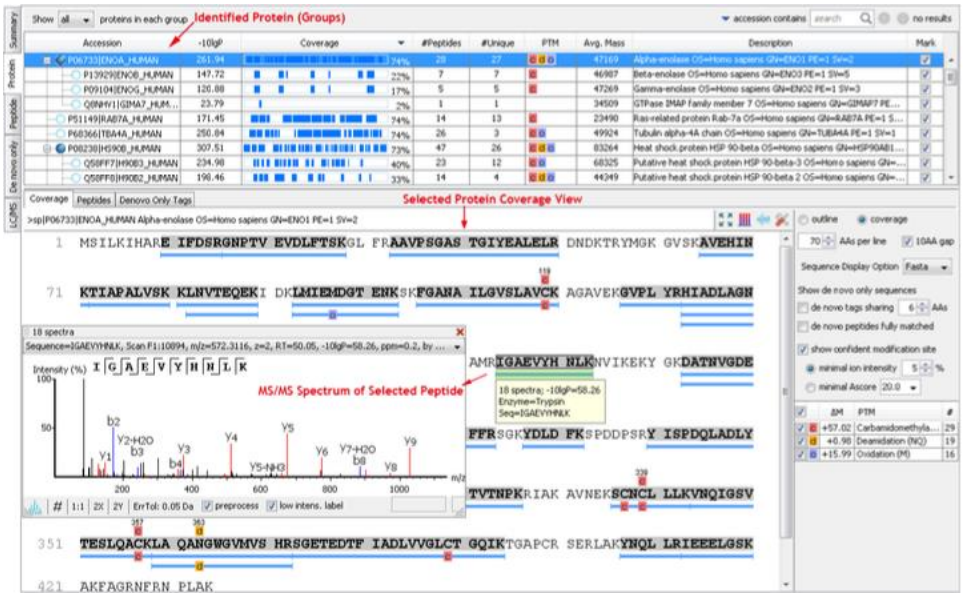
Seq	#	B	Y	# (+1)
G	1	58.02933	1815.90301	16
L	2	171.11340	1758.88155	15
S	3	258.14543	1645.79749	14
D	4	373.17237	1558.76546	13
G	5	430.19383	1443.73851	12
E	6	559.23642	1386.71705	11
W	7	745.31574	1257.67446	10
Q	8	873.37431	1071.59515	9
Q	9	1001.43289	943.53657	8
V	10	1100.50131	815.47799	7
L	11	1213.58537	716.40958	6
N	12	1327.62830	603.32551	5
V	13	1426.69671	489.28259	4
W	14	1612.77602	390.21417	3
G	15	1669.79749	204.13486	2
K	16	1797.89245	147.11340	1

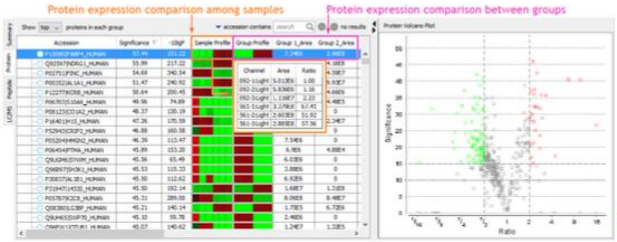
Mass/Charge Table

	Mass	
	Mono	Avg
(M)	1814.89519	1816.00312
$(M+H)^+$	1815.90301	1817.01106
$(M+2H)^{2+}$	908.45544	909.00952

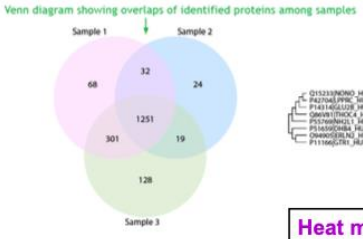
PEAKS is a tool for de novo sequencing in mass spectrometry labs. PEAKS assigns a local confidence score for each amino acid in the de novo sequence. This local confidence ranges from 0% to 99%, indicating how confident the algorithm is about the amino acid. The whole peptide is evaluated by two measures: the **ALC** (Average of Local Confidence) and **TLC** (Total of Local Confidence) scores.

- ALC reflects the average correct ratio for the amino acids in the sequence (or the likelihood of each amino acid assignment in a resultant peptide).
- TLC reflects the expected total number of correct amino acids in the sequence.

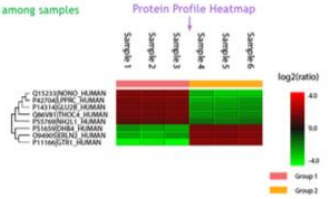




Volcano plot (scatter-plot to quickly identify changes in large data sets composed of replicate data. It plots significance versus fold-change on the y and x axes).



A Venn diagram is a diagram that shows all possible logical relations between a finite collection of different sets



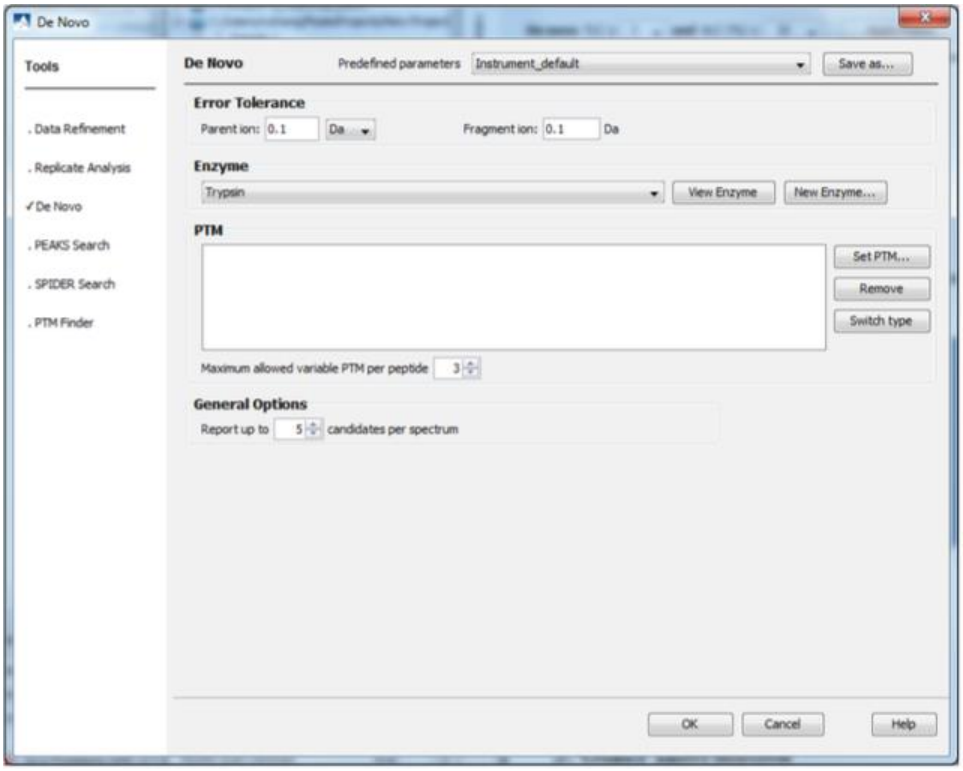
Heat map is a two-dimensional representation of data in which values are represented by colors. A simple heat map provides an immediate visual summary of information.

It can provide elaborate data almost ready for publication, like the volcano plot: these are scattered plots where on the x-axis is the fold change response (+log of the peak value) and on the y-axis is the significance (-log).

Both underexpression and overexpression, those are on the higher part of the graph is higher differentiation change exposure. We are trying to find a high fold change with a high significance.

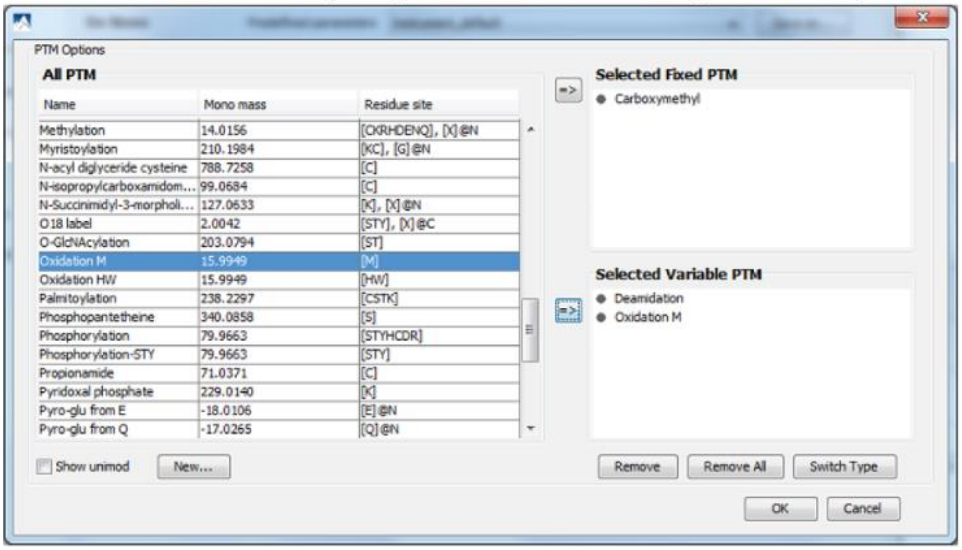
Heat map is a representation of the data, with columns and blocks, overexpression, underexpression with a visual output.

Venn diagram which is we have a different overlapping expression.



Fixed and Variable PTMs

To select the PTMs for the *de novo* sequencing, click the "Set PTM..." button to open the "PTM Setup" window.



The **de novo peptide view** displays the de novo sequencing results in more detail. The table at the top displays all the de novo sequences, and the bottom half of the view provides additional information about the peptide-spectrum match.

Peptide Table

PEAKS displays the peptide sequence candidates at the top of the screen. You can sort the results by clicking on the titles of the columns.

Contents of the columns in the “Peptide Candidates Frame”. The first column is a unique index for the peptides in the list.

• **Scan**: Scan number.

• **Peptide**: The amino acid sequence of the peptide as determined by de novo sequencing. If there is any PTM on an amino acid, the amino acid is followed by a pair of parentheses enclosing the delta mass of the PTM.

• **TLC**: Total local confidence. It is calculated by adding the local confidence for each amino acid in the peptide sequence.

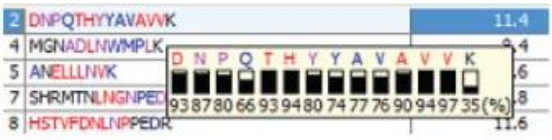
• **ALC**: Average local confidence (TLC divided by the peptide length).
• **m/z**: The measured mass/charge value, in Dalton, for the spectrum.
• **z**: The calculated charge value for the peptide.
• **RT**: Retention time (elution time) for the spectrum as recorded in the data.

• **ppm**: The precursor mass error, calculated as 10^6 × (observed mass - theoretical mass) / theoretical mass.

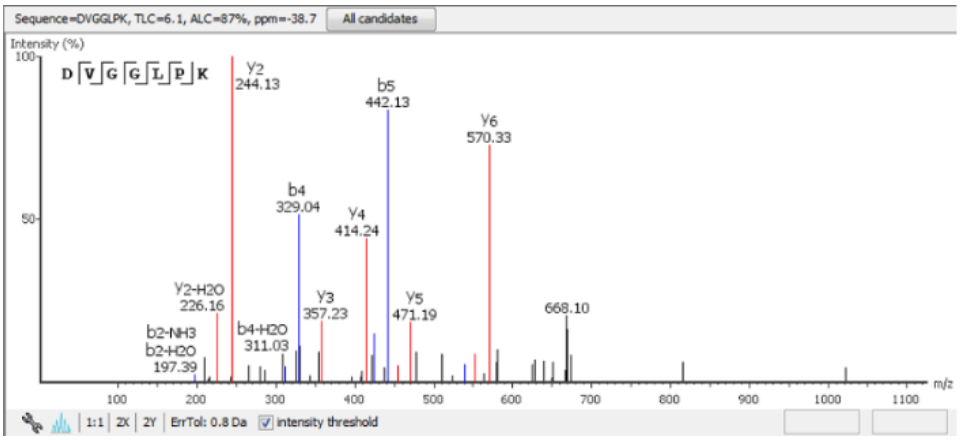
Confidence Scores:

- Next to the proposed sequence candidates, the auto de novo “Total Local Confidence” (TLC) and “Average Local Confidence” (ALC) confidence scores are shown.

- The local confidence scores for each amino acid (that is, confidence that the correct residue in each position has been identified) are represented by color coding.
- Red represents a very high confidence (greater than 90%), purple represents a high confidence (80 to 90%) blue represents a medium confidence (60 to 80%) and black represents a low confidence (less than 60%).



Spectrum Annotation:



Possible questions:

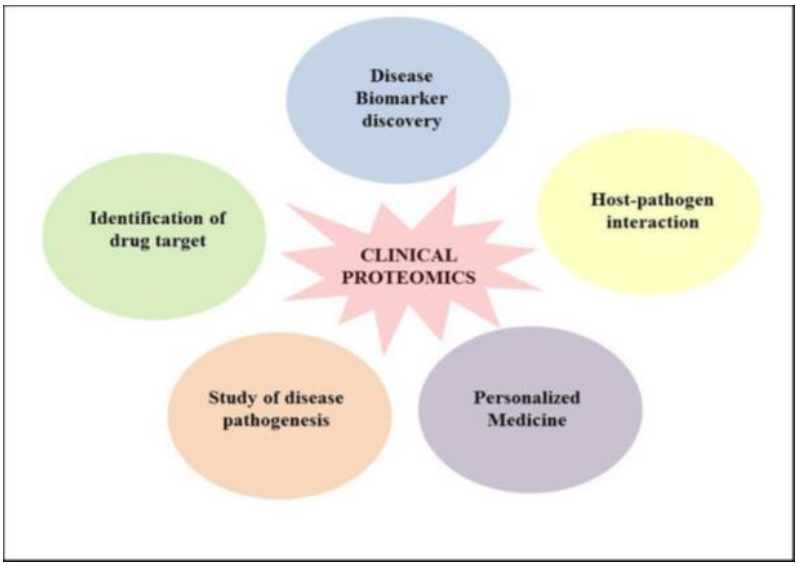
1. Peptide mass fingerprinting

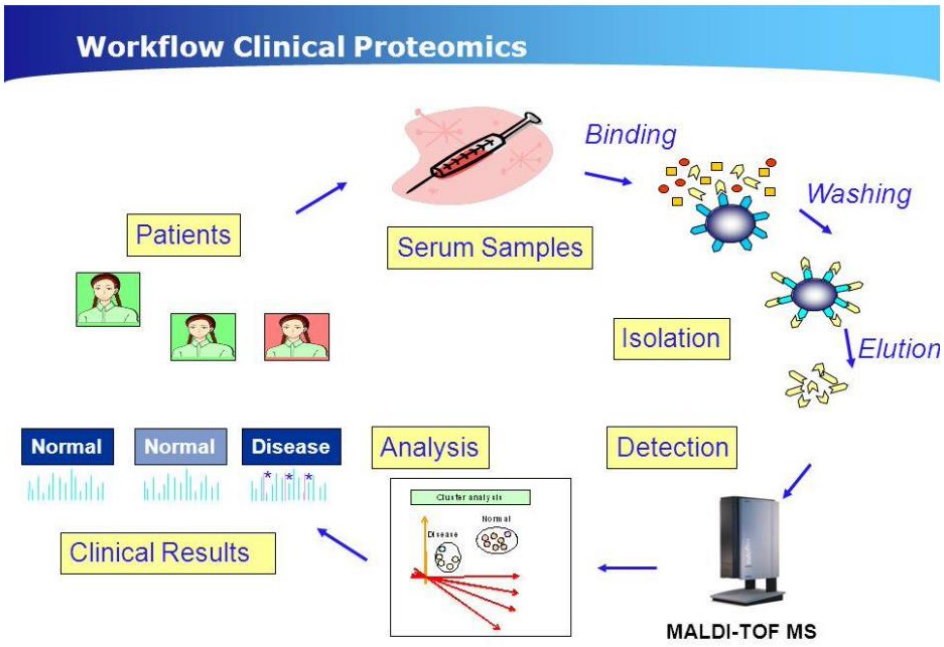
2. MALDI and ESI difference

3. Main issues of 2D Gel

4. Multidimensional HPLC

5. De novo





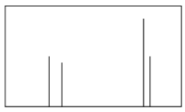
Peptide Mass Fingerprinting:

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

*Considering all aa sequence combinations that are theoretically possible, only a very minor portion of protein sequences is realized in nature, and therefore a short peptide sequence is already highly protein-specific.

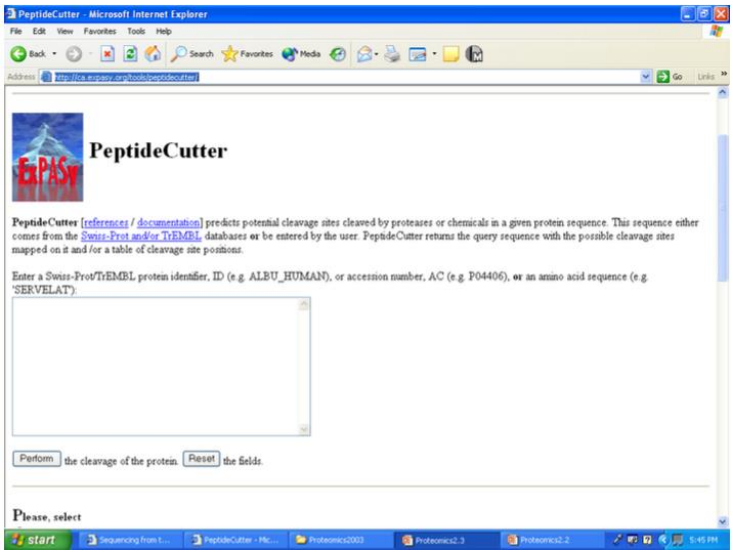
Principles of Fingerprinting:

Sequence	Mass (M+H)	Tryptic Fragments
>Protein 1 acedfhsakdfqea sdfpkivtmeeewe ndadnfekqwfe	4842.05	acedfhsak dfgeasdfpk ivtmeeewendadnfek gwfe
>Protein 2 acekdfhsadfqea sdfpkivtmeeewe nkdadnfeqwfe	4842.05	acek dfhsadfgeasdfpk ivtmeeewenk dadnfeqwfe
>Protein 3 acedfhsadfqeka sdfpkivtmeeewe ndakdnfeqwfe	4842.05	acedfhsadfgek asdfpk ivtmeeewendak dnfegwfe

Sequence	Mass (M+H)	Mass Spectrum
>Protein 1 acedfhsakdfqea sdfpkivtmeeewe ndadnfekqwfe	4842.05	
>Protein 2 acekdfhsadfqea sdfpkivtmeeewe nkdadnfeqwfe	4842.05	
>Protein 3 acedfhsadfqeka sdfpkivtmeeewe ndakdnfeqwfe	4842.05	

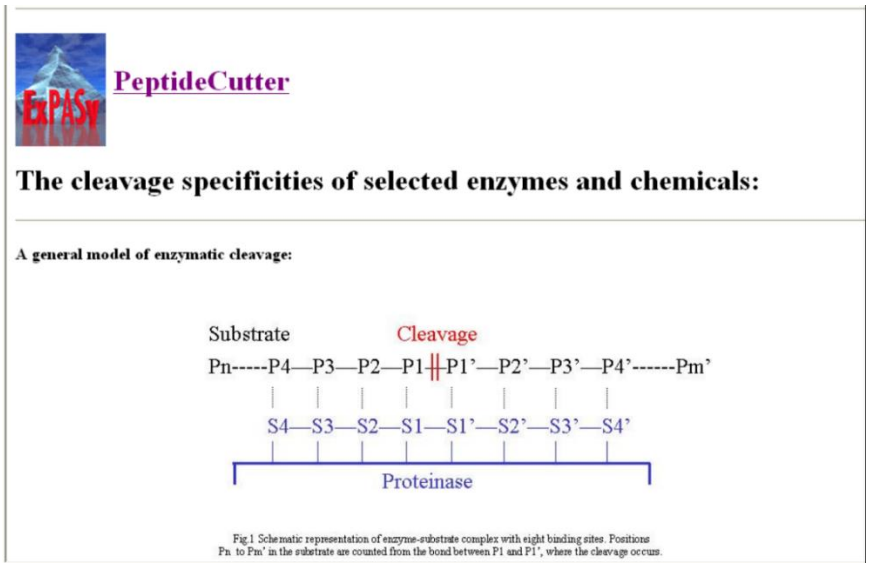
same mass but different mass spectra from maldi tof

Predicting Peptide Cleavages:



https://www.expasy.org/resources/peptidecutter

https://www.expasy.org/resources/peptidecutter#Tryps



Protein Cleavage rules:

Sometimes inhibition occurs

Trypsin	**XXX[KR]--[!P]XXX**
Chymotrypsin	**XX[FYW]--[!P]XXX**
Lys C	**XXXXXK-- XXXXX**
Asp N endo	**XXXXXD-- XXXXX**
CNBr	**XXXXXM--XXXXX**

proline residue impedes the enzyme activity

K-Lysine, R-Arginine, F-Phenylalanine, Y-Tyrosine,
W-Tryptophan,D-Aspartic Acid, M-Methionine, P-Proline

Digest with specific protease:

546 aa 60 kDa; 57 461 Da pI = 4.75

>RBME00320 Contig0311_1089618_1091255 EC-mopA 60 KDa chaperonin GroEL
MAAKDVKFGR TAREKMLRGV DILADAVKVT LGPKGRNVVI EKSFGAPRIT KDGVSVAKEV
ELEDKFENMG AQMLREVASK TNDTAGDGTT TATVLGQAIV QEGAKAVAAG MNPMDLKRGI
DLAVNEVVAE LLKKAKKINT SEEVAQVGTI SANGEAEIGK MIAEAMQKVG NEGVITVEEA
KTAETELEVV EGMQFDRGYL SPYFVTNPEK MVADLEDAYI LLHEKKLSNL QALLPVLEAV
VQTSKPLLII AEDVEGEALA TLVVNKLRGG LKIAAVKAPG FGDCRKAMLE DIAILTGGQV
ISEDLGIKLE SVTLDMLGRA KKVSISKENT TIVDGAGQKA EIDARVGQIK QQIEETTSDY
DREKLQERLA KLAGGVAVIR VGGATEVEVK EKKDRVDDAL NATRAAVEEG IVAGGGTALL
RASTKITAKG VNADQEAGIN IVRRAIQAPA RQITTNAGEE ASVIVGKILE NTSETFGYNT
ANGEYGDLIS LGIVDPVKVV RTALQNAASV AGLLITTEAM IAELPKKDAA PAGMPGGMGG
MGGMDF

Trypsin yields 47 peptides (theoretically)

Peptide masses in Da:

501.3	533.3	544.3	545.3	614.4	634.3
674.3	675.4	701.4	726.4	822.4	855.5
861.4	879.4	921.5	953.4	974.5	988.5
1000.6	1196.6	1217.6	1228.5	1232.6	1233.7
1249.6	1249.6	1344.7	1455.8	1484.6	1514.8
1582.9	1583.9	1616.8	1726.7	1759.9	1775.9
1790.6	1853.9	1869.9	2286.2	2302.2	2317.2
2419.2	2526.4	2542.4	3329.6	4211.4	

http://us.expasy.org/tools/peptide-mass.html

Digest with trypsin

- In practice.....see far fewer by mass spec
- possibly incomplete digestion (we allow 1 missed cleavage site)
- lose peptides during each manipulation
- washes during digestion
- washes during cleanup step some peptides will not ionize well some signals (peaks) are poor low intensity; lack resolution

Missed Cleavage:

Sequence

>Protein 1
acedfhsakdfqea
sdfpkivtmeeewe
ndadnfekqwfe

Tryptic Fragments (no missed cleavage)

acedfhsak (1007.4251)
dfgeasdfpk (1183.5266)
ivtmeeewendadnfek (2098.8909)
gwfe (609.2667)

Tryptic Fragments (1 missed cleavage)

acedfhsak (1007.4251)
dfgeasdfpk (1183.5266)
ivtmeeewendadnfek 2098.8909)
gwfe (609.2667)
acedfhsakdfgeasdfpk (2171.9338)
ivtmeeewendadnfekgwfe (2689.1398)
dfgeasdfpkivtmeeewendadnfek (3263.2997)

Calculating Peptide Masses:

- Sum the monoisotopic residue masses
- Add mass of H_2O (18.01056)
- Add mass of H^+ (1.00785 to get M+H)
- If Met is oxidized add 15.99491
- If Cys has acrylamide adduct add 71.0371
- If Cys is iodoacetylated add 58.0071

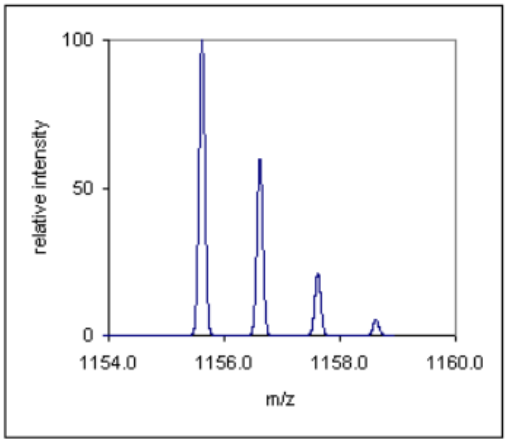
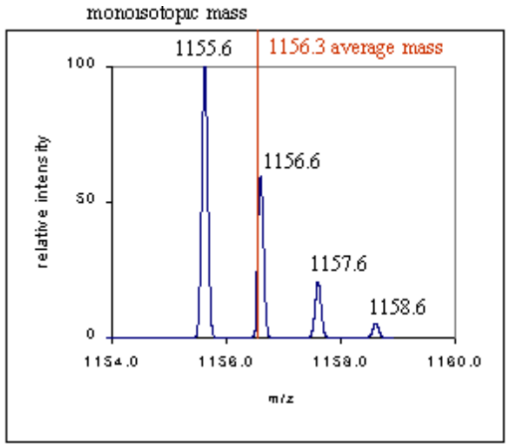
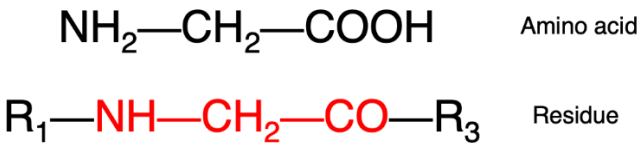


Figure shows a simulated isotopic distribution of the [M+H]+ ion of a compound with the following elemental composition, C48 H82 N16 O17 (Poly- Alanine)



- Monoisotopic mass is the mass determined using the masses of the most abundant isotopes.
- Average mass is the abundance weighted mass of all isotopic components.



Monoisotopic Mass
$^{1}H = 1.007825$
$^{12}C = 12.00000$
$^{14}N = 14.00307$
$^{16}O = 15.99491$

Glycine Amino Acid Mass
5xH + 2xC + 2xO + 1xN
= 75.032015 amu

Glycine Residue Mass
3xH + 2xC + 1xO + 1xN
=57.021455 amu

Preparing a Peptide Mass Fingerprint Database:

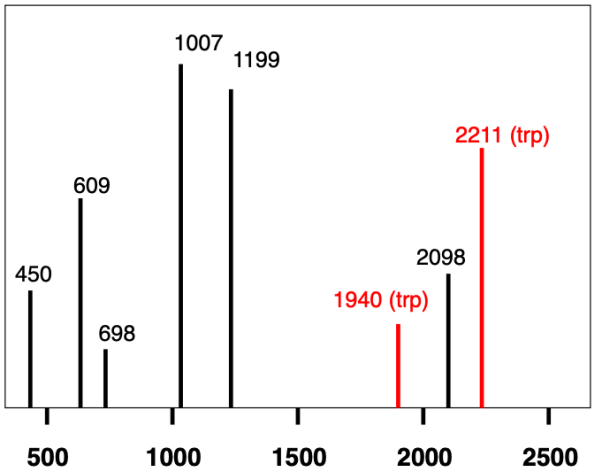
- Take a protein sequence database (SwissProt or nr-GenBank)
- Determine cleavage sites and identify resulting peptides for each protein entry
- Calculate the mass (M+H) for each peptide
- Sort the masses from lowest to highest
- Have a pointer for each calculated mass to each protein accession number in databank

Building A PMF Database

Sequence DB	Calc. Tryptic Frags	Mass List
>P12345 acedfhsakdfqea sdfpkivtmeeewe ndadnfekqwfe	acedfhsak dfgeasdfpk ivtmeeewendadnfek gwfe	450.2017 (P21234) 609.2667 (P12345) 664.3300 (P89212) 1007.4251 (P12345)
>P21234 acekdfhsadfqea sdfpkivtmeeewe nkdadnfeqwfe	acek dfhsadfgeasdfpk ivtmeeewenk dadnfeqwfe	1114.4416 (P89212) 1183.5266 (P12345) 1300.5116 (P21234) 1407.6462 (P21234) 1526.6211 (P89212)
>P89212 acedfhsadfqeka sdfpkivtmeeewe ndakdnfeqwfe	acedfhsadfgek asdfpk ivtmeeewendak dnfegwfe	1593.7101 (P89212) 1740.7501 (P21234) 2098.8909 (P12345)

The Fingerprint (PMF) Algorithm:

- Take a mass spectrum of a trypsin-cleaved protein (from gel or HPLC peak)
- Identify as many masses as possible in spectrum (avoid autolysis peaks of trypsin)
- Compare query masses with database masses and calculate # of matches or matching score (based on length and mass difference)
- Rank hits and return top scoring entry – this is the protein of interest



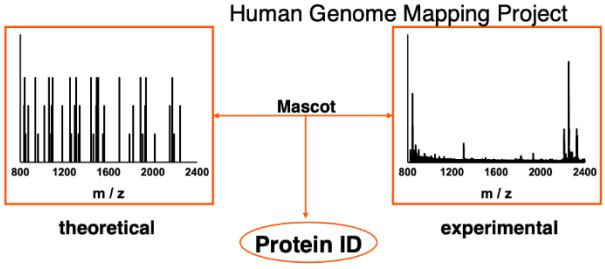
Query vs. Database

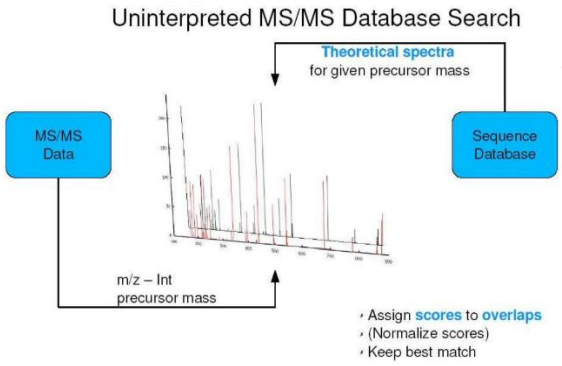
Query Masses	Database Mass List	Results
450.2201	450.2017 (P21234)	2 Unknown masses
609.3667	609.2667 (P12345)	1 hit on P21234
698.3100	664.3300 (P89212)	3 hits on P12345
1007.5391	1007.4251 (P12345)	
1199.4916	1114.4416 (P89212)	Conclude the query
2098.9909	1183.5266 (P12345)	protein is P12345
	1300.5116 (P21234)	
	1407.6462 (P21234)	
	1526.6211 (P89212)	
	1593.7101 (P89212)	
	1740.7501 (P21234)	
	2098.8909 (P12345)	

Database search

PeptIdent (ExPasy)
→ Mascot (Matrix Science)
MS-Fit (Prospector; UCSF)
ProFound (Proteometrics)
MOWSE (HGMP)

Human Genome Mapping Project



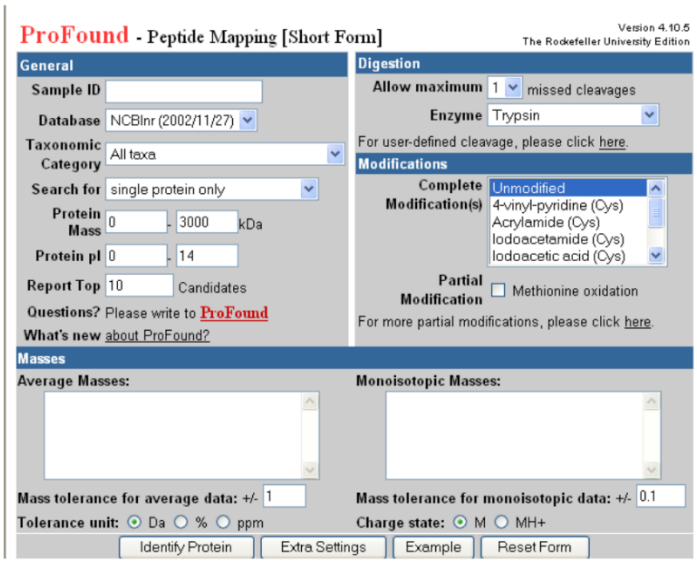


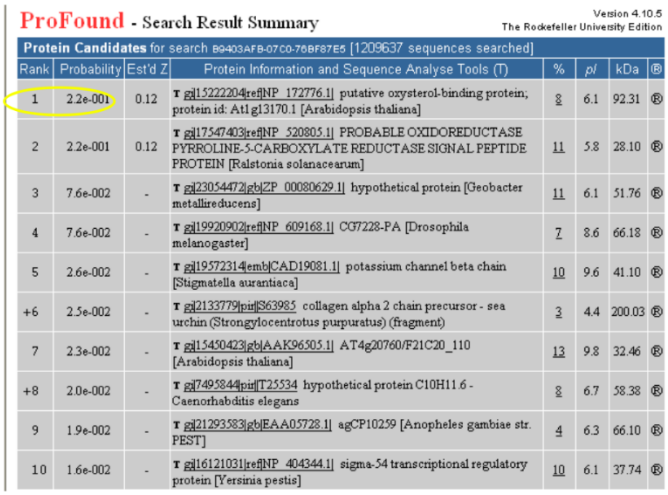
the theoretical spectra doesn't exist, it is just in silico digestion

What do we need to do for PMF:

- A list of query masses (as many as possible)
- Protease(s) used or cleavage reagents
- Databases to search (SWProt, Organism)
- Estimated mass and pI of protein spot
- Cysteine (or other) modifications
- Minimum number of hits for significance
- Mass tolerance (100 ppm = 1000.0 ± 0.1 Da)
- A PMF website (ProFound, Mascot, etc.)

ProFound





ProFound - Search Result Summary

Version 4.10.5
The Rockefeller University Edition

Protein Candidates for search B9403AFB-07C0-76BF87E5 [1209637 sequences searched]

Rank	Probability	Est'd Z	Protein Information and Sequence Analyse Tools (T)	%	pI	kDa	®
1	2.2e-001	0.12	T gi|15222204|ref|NP_172776.1| putative oxysterol-binding protein; protein id: At1g13170.1 [Arabidopsis thaliana]	8	6.1	92.31	®
2	2.2e-001	0.12	T gi|17547403|ref|NP_520805.1| PROBABLE OXIDOREDUCTASE PYRROLINE-5-CARBOXYLATE REDUCTASE SIGNAL PEPTIDE PROTEIN [Ralstonia solanacearum]	11	5.8	28.10	®
3	7.6e-002	-	T gi|23054472|gb|ZP_00080629.1| hypothetical protein [Geobacter metallireducens]	11	6.1	51.76	®
4	7.6e-002	-	T gi|19920902|ref|NP_609168.1| CG7228-PA [Drosophila melanogaster]	7	8.6	66.18	®
5	2.6e-002	-	T gi|19572314|emb|CAD19081.1| potassium channel beta chain [Stigmatella aurantiaca]	10	9.6	41.10	®
+6	2.5e-002	-	T gi|2133779|pir||S63985 collagen alpha 2 chain precursor - sea urchin (Strongylocentrotus purpuratus) (fragment)	3	4.4	200.03	®
7	2.3e-002	-	T gi|15450423|gb|AAK96505.1| AT4g20760/F21C20_110 [Arabidopsis thaliana]	13	9.8	32.46	®
+8	2.0e-002	-	T gi|7495844|pir||T25534 hypothetical protein C10H11.6 - Caenorhabditis elegans	8	6.7	58.38	®
9	1.9e-002	-	T gi|21293583|gb|EAA05728.1| agCP10259 [Anopheles gambiae str. PEST]	4	6.3	66.10	®
10	1.6e-002	-	T gi|16121031|ref|NP_404344.1| sigma-54 transcriptional regulatory protein [Yersinia pestis]	10	6.1	37.74	®

Advantages of PMF: DOMANDA ESAME, IN COSA CONSISTE LA PMF

- Uses a "robust" & inexpensive form of MS (MALDI)
- Doesn't require too much sample optimization
- Can be done by a moderately skilled operator (don't need to be an MS expert)
- Widely supported by web servers
- Improves as databases get larger & instrumentation gets better

Limitations with PMF:

- Requires that the protein of interest already is in a sequence database
- Spurious or missing critical mass peaks always leads to problems
- Mass resolution/accuracy is critical
- Generally found to only be about 40% effective in positively identifying gel spots