Module 1

Separation techniques and Mass Spectrometry for Life Sciences

Wojciech Luczaj, Agnieszka Gegotek, Katarzyna Bielawska,
Elzbieta Skrzydlewska

Department of Analytical Chemistry, Medical University of Białystok, 15-089 Białystok, Poland

e-materials (Intellectual Output 3) produced as part of
Advanced Analytical Chemistry for Life Sciences [AACLifeSci] project

This project has been funded with support from
the European Commission

AACLifeSci

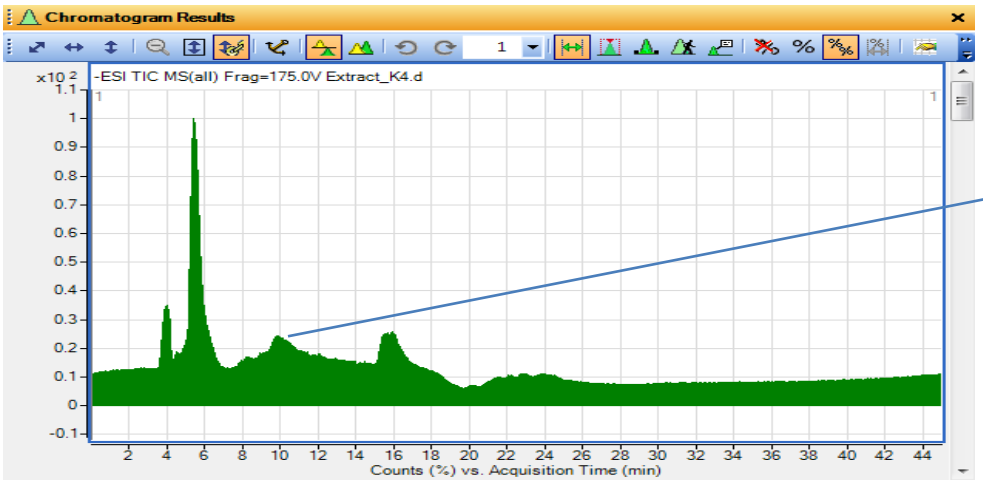
Erasmus+

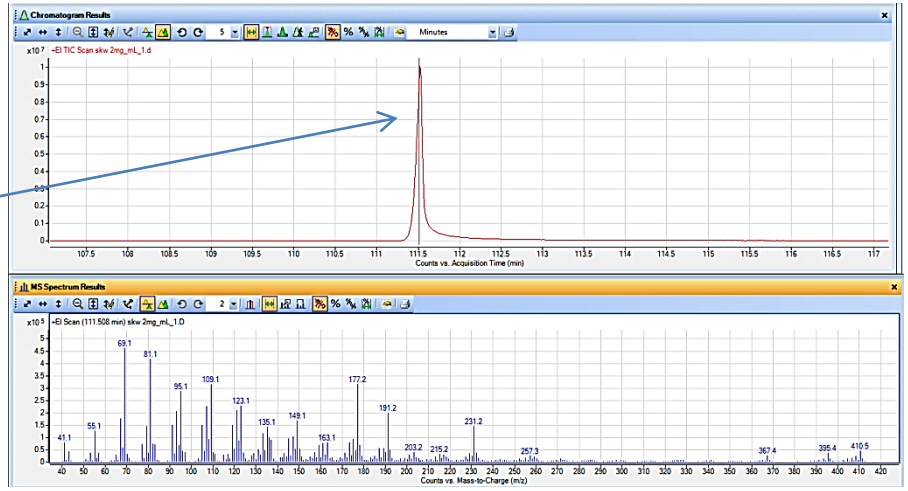
Mass spectrometry

Analytical technique that allows obtaining information on:

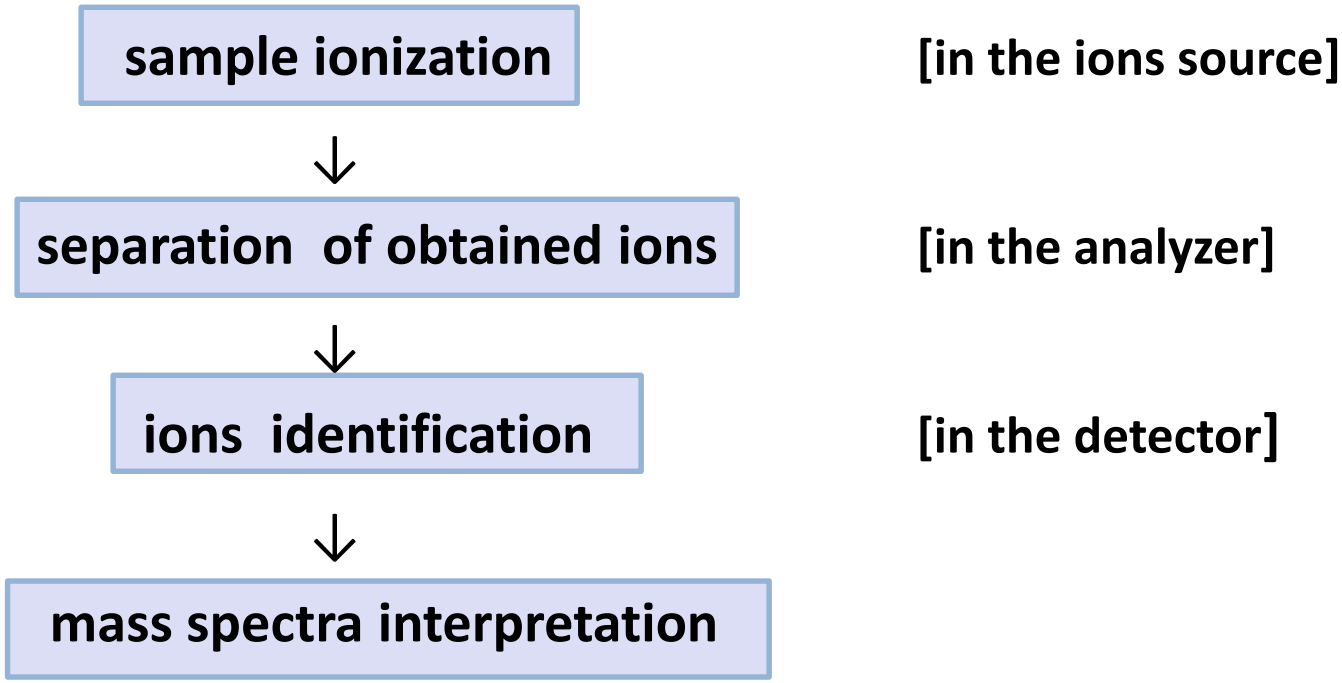
- ✓ molecular weight
- ✓ chemical structure
- ✓ amounts of examined compounds

Analytical technique that allows to separate and identify ions according to their mass-to-charge ratio (m/z)



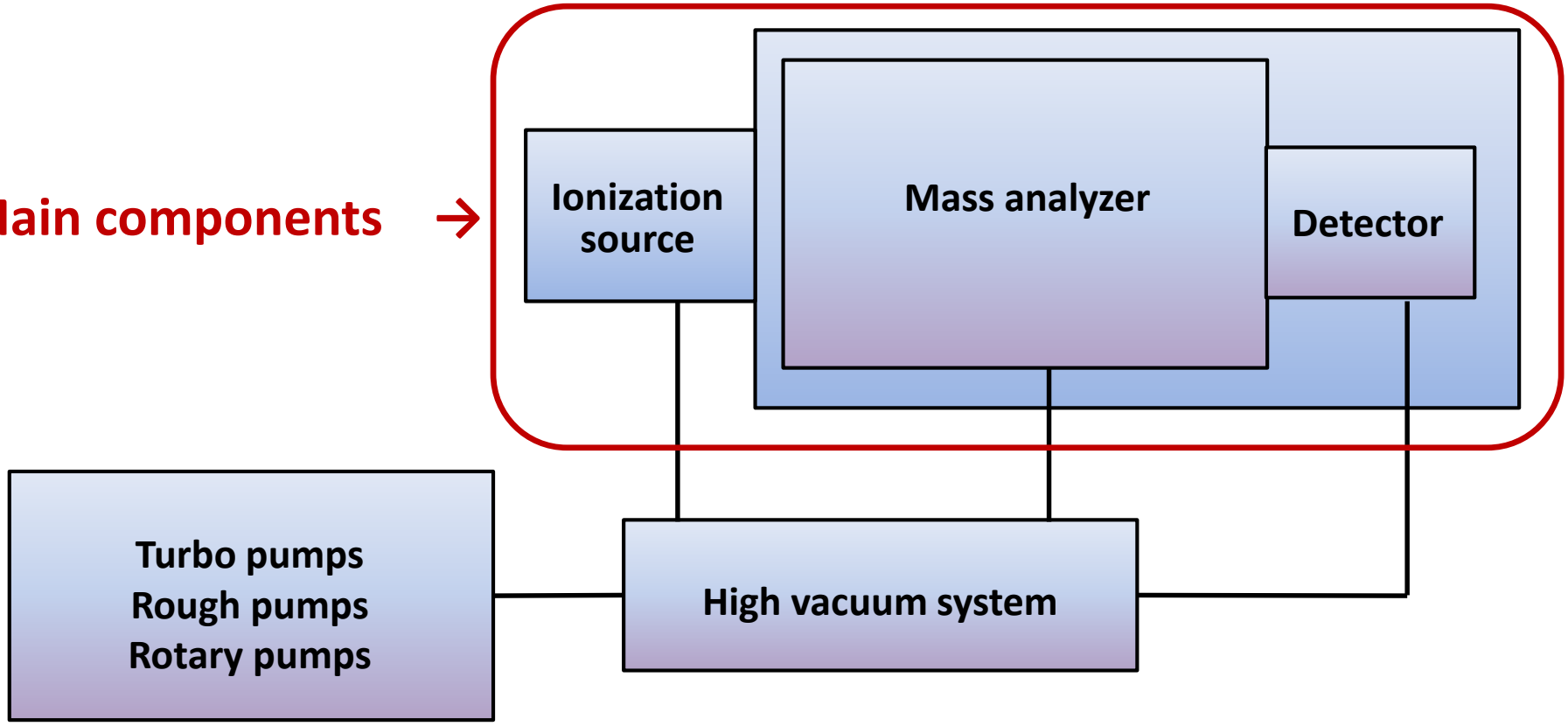


Stages of the analytical procedure based on mass spectrometry



Mass spectrometer

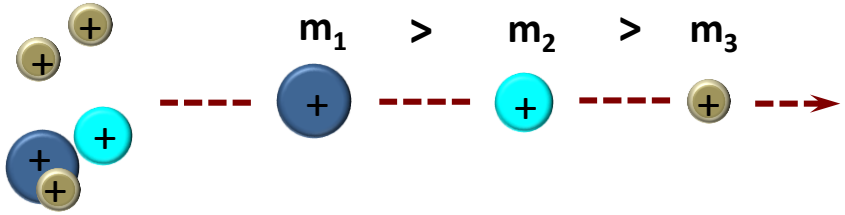
Main components →



MS workflow

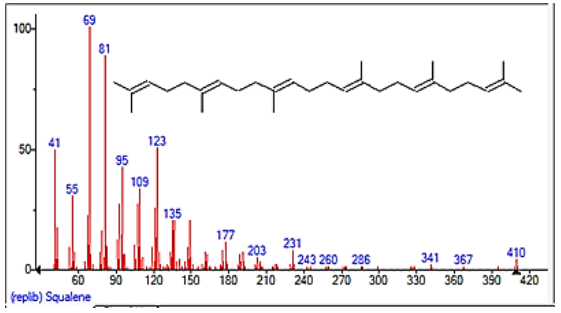
1. **Ionization of the analyte**

2. **Separation of the ionized molecules according to their m/z**



3. **Detection of the ions**

4. **Analysis and interpretation of the mass spectrum**



Ionization methods

- electron impact ionization [EI]
- electrospray ionization [ESI]
- matrix-assisted laser desorption ionization [MALDI]
- atmospheric-pressure chemical ionisation [APCI]

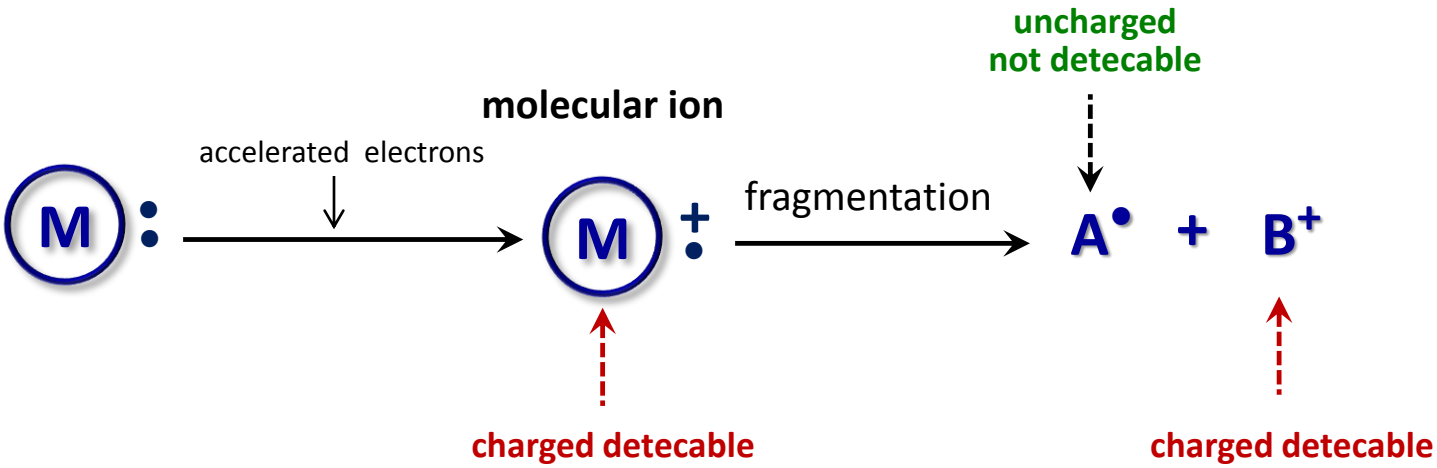
hard ionization – products: molecular ion + fragmented element

✓ **EI**

soft ionization – the main product - molecular ion

✓ **ESI**
✓ **MALDI**
✓ **APCI**

Electron ionization (EI)

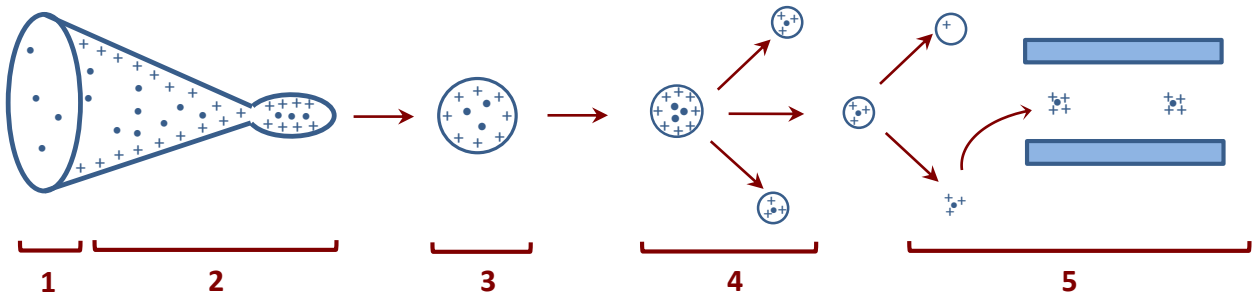


Electron ionization mechanism and subsequent fragmentation

Electron ionization (EI)

- **Typical analytes:**
  - relatively small,
  - non-polar,
  - volatile,
  - thermostable
- **Mass range:**
  - <1 kDa
- **Sample introduction:**
  - GC or liquid/solid
- **Advantages:**
  - non-polar analytes,
  - no ion suppression,
  - easily coupled with GC,
  - spectrum libraries
- **Disadvantages:**
  - analysis
    - ✓ volatile compounds,
    - ✓ thermally stable compounds,
    - ✓ low molecular weight compounds,
  - hard jonization,

Electrospray ionization (ESI)



1- production of ions ,
2- formation of charged droplets spray ,
3- desolvation,
4- „Coulomb fission”,
5- gas phase ions genertion

The mechanism of electrospray ionization

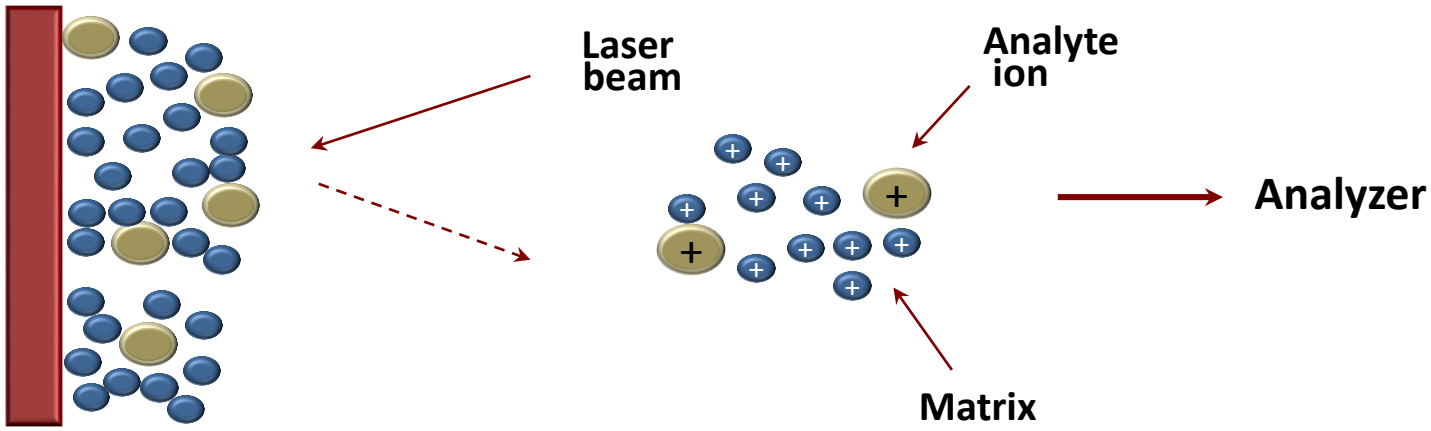
Electrospray ionization (ESI)

- **Typical analytes:**
  - polar compounds
    e.g. peptides, proteins, sugars, nucleotides
- **Mass range:**
  - <200 kDa
- **Sample introduction:**
  - LC or solution

- **Advantages:**
  - thermolabile compounds
  - high MW compounds
  - multi-charged ions
  - sensitivity
  - easy to interface with LC
  - soft ionization method

- **Disadvantages:**
  - ionizable analytes
  - sensitive to salts
  - ion suppression

Matrix-assisted laser desorption ionization (MALDI)



The mechanism of matrix-assisted laser desorption ionization

Matrix-assisted laser desorption ionization (MALDI)

- **Typical analytes:**
  - polar compounds
    e.g. peptides, proteins, sugars, nucleotides
- **Mass range:**
  - <500 kDa
- **Sample introduction:**
  - sample mixed with a solid matrix

- **Advantages:**
  - thermolabile compounds,
  - high MW compounds,
  - sensitivity,
  - less sensitive to salts,
  - soft ionization method

- **Disadvantages:**
  - a wide range of matrices,
  - difficulties in quantitative analysis,
  - ion suppression

Atmospheric-pressure chemical ionization (APCI)

- **Typical analytes:**
  - polar compounds
    e.g. peptides, proteins, sugars, nucleotides
- ✓ **Mass range:**
  - <1 kDa
- ✓ **Sample introduction:**
  - LC or solution

- **Advantages:**
  - thermostabile compounds,
  - sensitivity,
  - allows for large flow rates,
  - easy to interface with LC,
  - soft ionization technique

- **Disadvantages:**
  - needs solubility in polar solvents,
  - sensitive to salts,
  - ion suppression

Mass analyzers

Separate ions according to their mass-to-charge (*m/z*) ratio

- **operate under high vacuum**
- **key specifications are:**
  - resolution
  - mass accuracy
  - sensitivity
  - dynamic range

Resolution

the ability to differentiate between closely related signals

$$R = \Delta m/m$$

where resolving power is defined as:

$$m_1/(m_2-m_1)$$

m_1 is the lighter ion and (m_2-m_1) is the difference between two consecutive ions

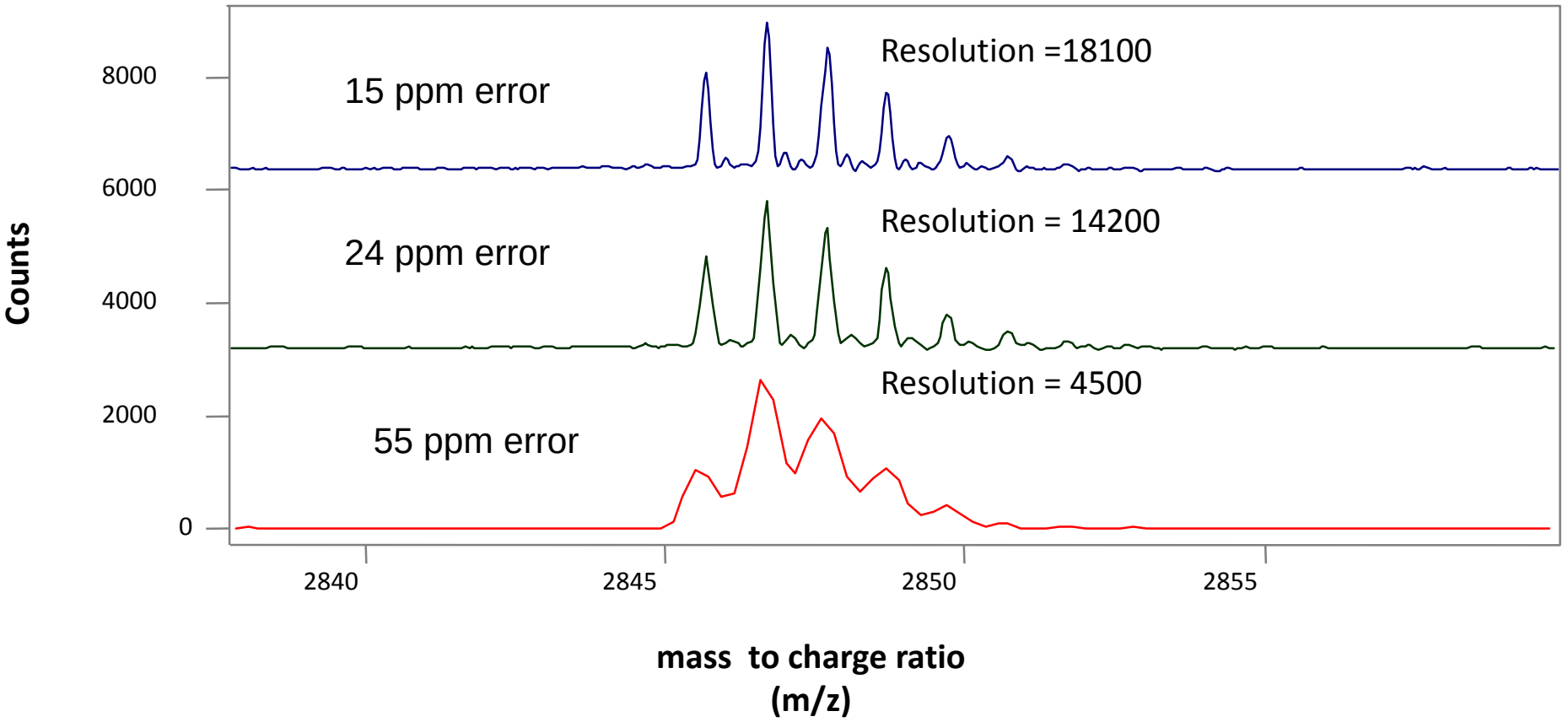
Mass accuracy

the proximity of the experimental mass (accurate mass) to the true value (exact mass)

$$\frac{(\text{monoisotopic exact mass} - \text{measured accurate mass})}{\text{monoisotopic exact mass} \times 10^6}$$

- determined in [ppm]

The higher resolution
the better mass accuracy



Sensitivity

the detector response that is related to the concentration of an analyte which reaches the detector

- determines the limit of detection (LOD)

Dynamic range

the range over which the ion signal is directly proportional to the analyte concentration

- crucial for accurate measurements (quantification analysis)

Mass analyzers

Most frequently used mass analyzers

Quadrupole (Q)

Ion-trap (IT)

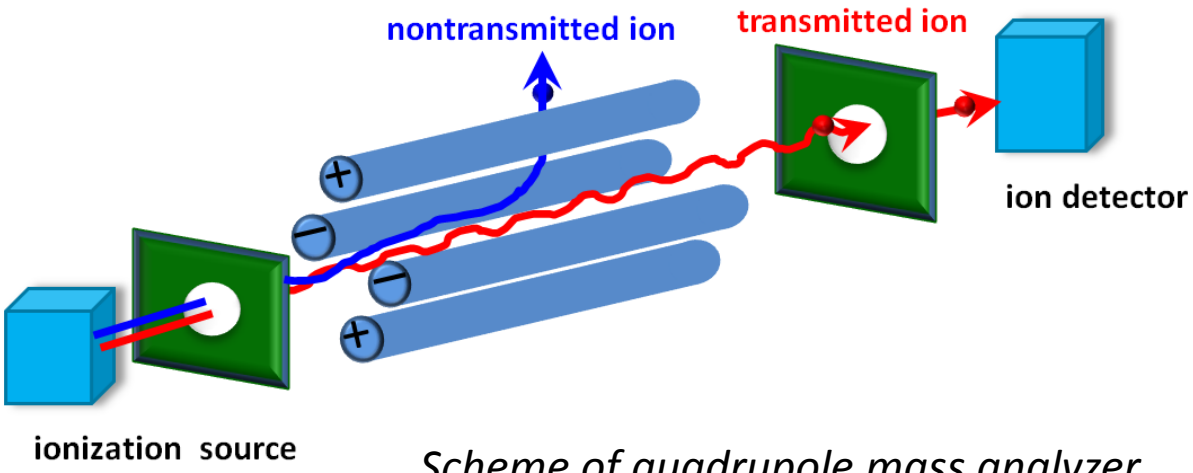
Time-of –flight (TOF)

Orbitrap

Mass analyzers

Quadrupole (Q):

- consists of four parallel rods
- uses combination of RF and DC voltages to operate as mass filter
- has variable ion transmission modes:
  - ✓ ion scanning (SCAN),
  - ✓ single ion monitoring (SIM)
- **low resolution**
- **highest sensitivity (quantitative analysis)**



Scheme of quadrupole mass analyzer

Mass analyzers

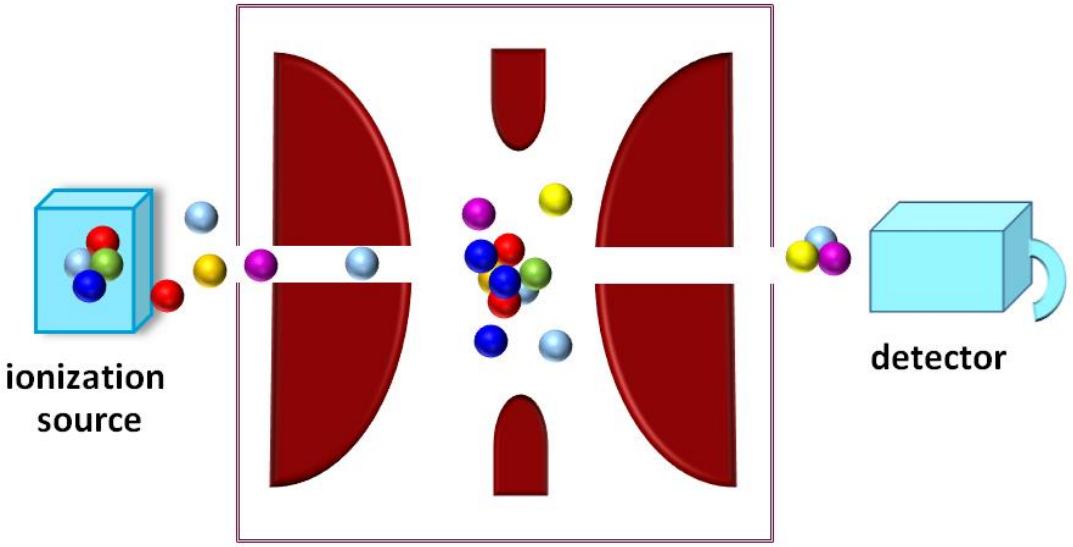
Quadrupole (Q):

Characterization	Quadrupole
Acquisition speed (Hz)	**2-10**
Mass accuracy (ppm)	**low**
Mass range (m/z)	**<3000**
Resolution	**unit**

Mass analyzers

Ion-trap (IT):

- traps ions using quadrupolar fields
- two types:
  - ✓ 2D ion-trap (linear ion-trap)
  - ✓ 3D ion-trap (quadrupole ion trap)
- **low resolution**
- **high scanning rate**



Scheme of 3D ion trap mass analyzer

Mass analyzers

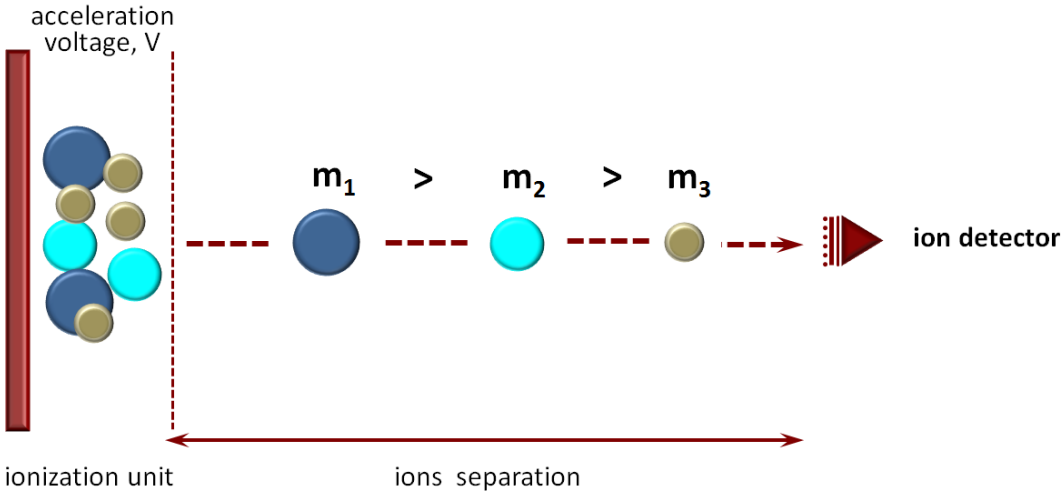
Ion-trap (IT)

Characterization	Ion-trap
Acquisition speed (Hz)	**2-10**
Mass accuracy (ppm)	**low**
Mass range (m/z)	**<6000**
Resolution	**unit**

Mass analyzers

Time-of-flight (TOF):

- ions are formed in pulses
- measures the time for ions to reach the detector
- small ions reach the detector before large ones
- **high resolution**
- **high mass accuracy**
- **high sensitivity**



The mechanism of ions separation in TOF mass analyzer

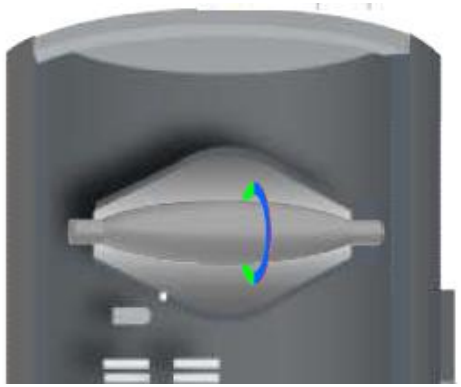
Mass analyzers

Time-of-flight (TOF)

Characterization	TOF
Acquisition speed (Hz)	**10-100**
Mass accuracy (ppm)	**1-10 ppm**
Mass range (m/z)	**<100,000 unlimited**
Resolution	**<50,000**

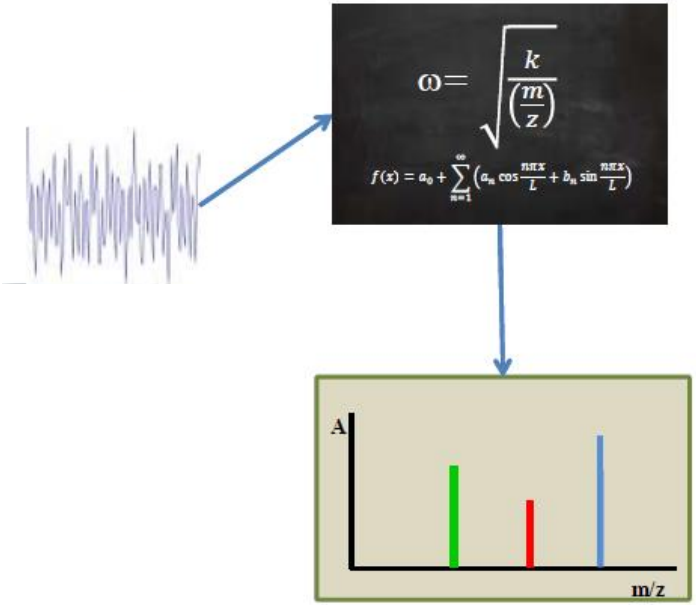
Mass analyzers

Orbitrap:



- consists of barrel-like electrode

- the m/z values are calculated by fast Fourier transform from the oscillation frequencies of the trapped ions



- **high resolution**
- **high mass accuracy**
- **high sensitivity**

Mass analyzers

Orbitrap

Characterization	Orbitrap
Acquisition speed (Hz)	**1-18**
Mass accuracy (ppm)	**1-5 ppm**
Mass range (m/z)	**<6000**
Resolution	**<500,000**

Comparison of different MS analyzers

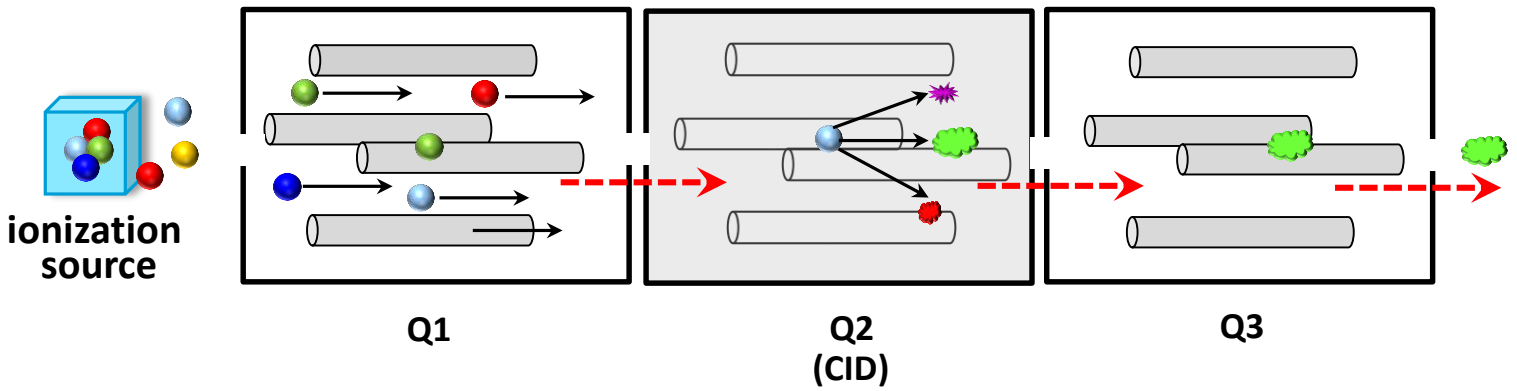
Analyzer	Q	IT	TOF	Orbitrap
Advantages	easily interfaced to various ionization techniques, higher dynamic range, low lost	easily interfaced to various ionization techniques, MS^n, low cost	fast scanning, high mass range, high mass accuracy	high mass accuracy, fast polarity switch
Disadvantages	low resolution, low mass accuracy, low mass range, low scanning speed, MS/MS requires multiple analyzers	low resolution, low mass accuracy, low mass range, low scanning speed	lower dynamic range than Q, high cost	lower scanning rate than QTOF, lower dynamic range than Q, high cost

Tandem mass spectrometry (MS/MS)

Tandem mass spectrometers

Triple quadrupole	[QqQ]
Quadrupole time-of-flight	[QqTOF]
Ion trap	[IT]
Quadrupole orbitrap	[QqOrbitrap]

Tandem mass spectrometry



The tandem mass spectrometry based on triple quadrupole [QqQ] operation

MS/MS analysis:

- **selected/multiply reaction monitoring (SRM/MRM)**
- product ion scan
- precursor ion scan
- neutral loss scan

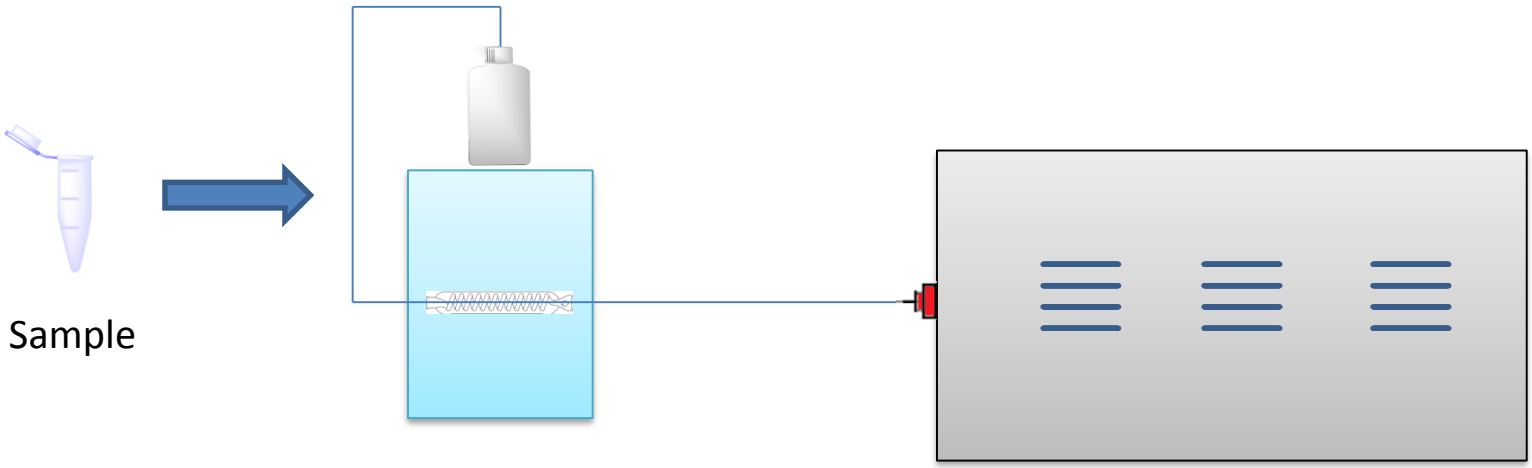
Separation techniques coupled with mass spectrometry

Mass spectrometry is most commonly combined with

- liquid chromatography [LC]
- gas chromatography [GC]
- capillary electrophresis [CE]

Introduction to LC-MS

Combination of the physical separation capabilities of liquid chromatography (LC) with the mass analysis capabilities of mass spectrometry (MS)



liquid chromatograph

LC allows separation of many compounds according their retention time (t_R)

mass spectrometer

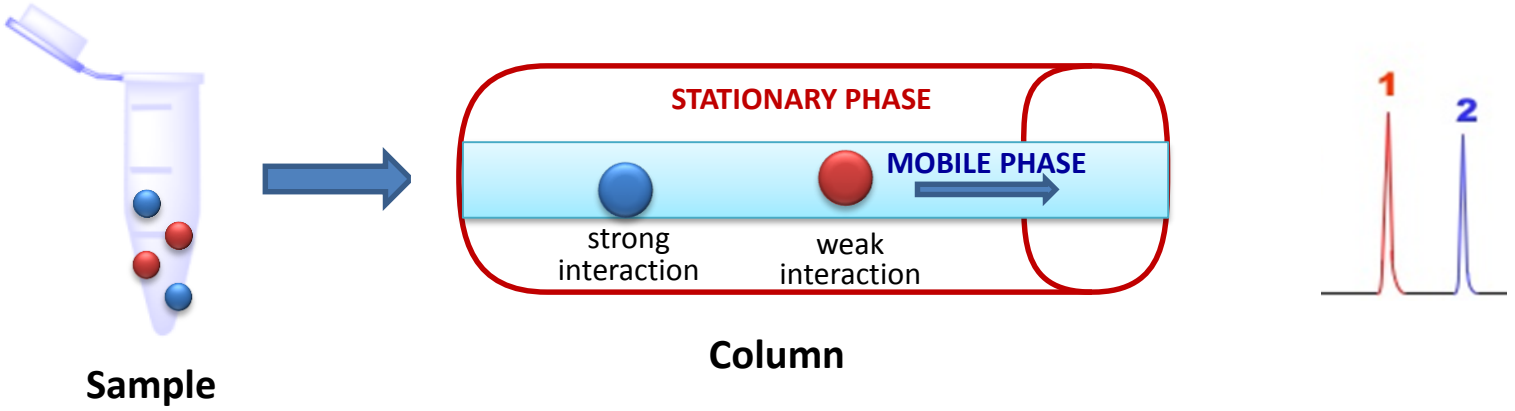
LC-MS allows differentiating many compounds with similar t_R, but with different m/z or fragmentation pattern

Introduction to liquid chromatography [LC]

LC - chromatography in which the mobile phase is a liquid ("eluent")

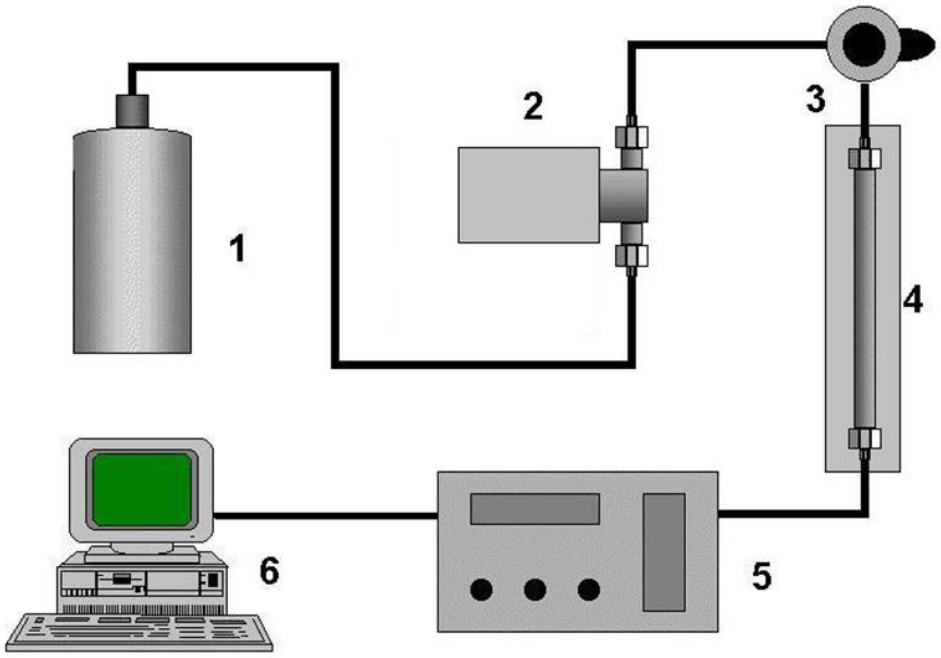
Separation mechanism

Due to different interaction between stationary/mobile phase and polarity of compounds in the sample, their molecules move at different rate and elute from the column in different time.



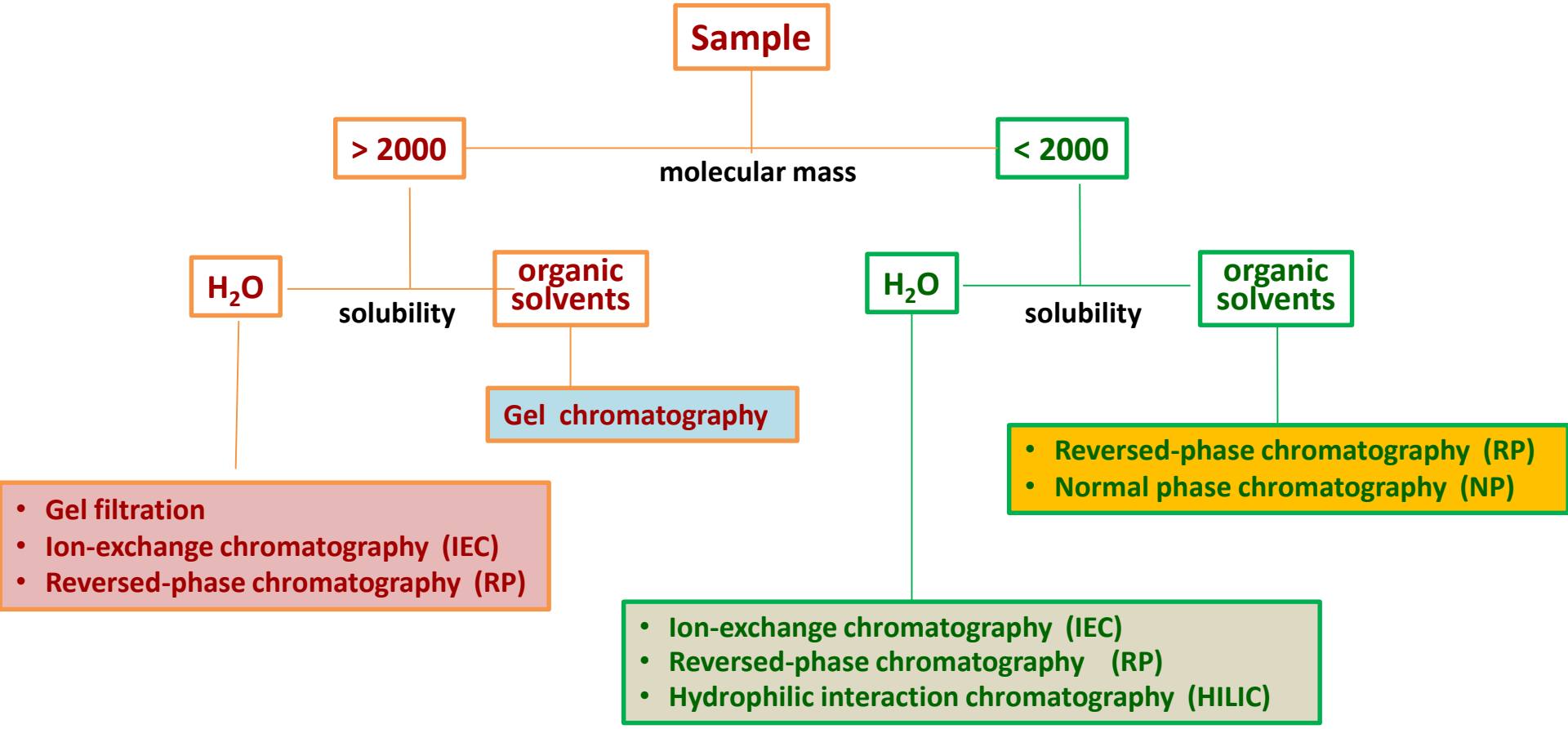
Liquid chromatography (LC)

A typical chromatographic system contains major components:



1 – mobile phase container,
2 - pump,
3 - injector,
4 – chromatographic column,
5 - detector,
6 – computer

Technique selection

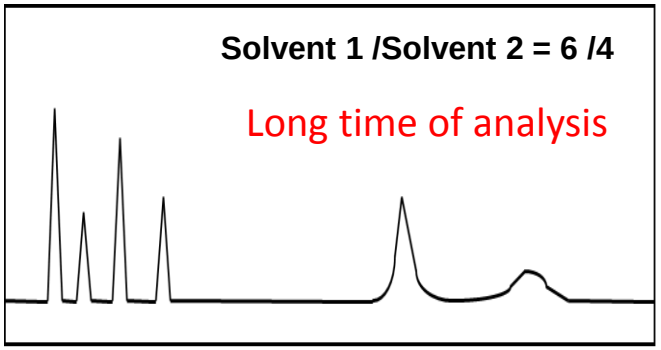


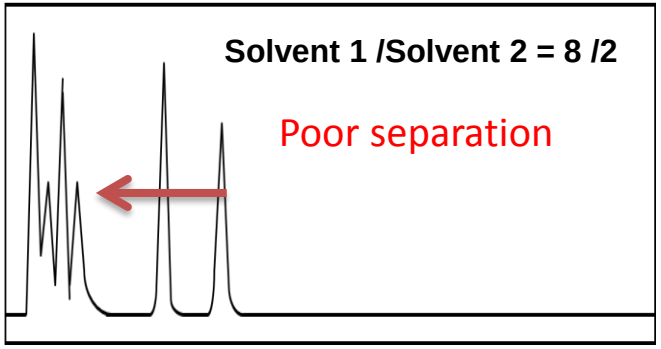
LC pump

The major task of the pump is providing a stable flow, which varies depending on the interface being used in the LC-MS
and
the parameters of the chromatographic column.

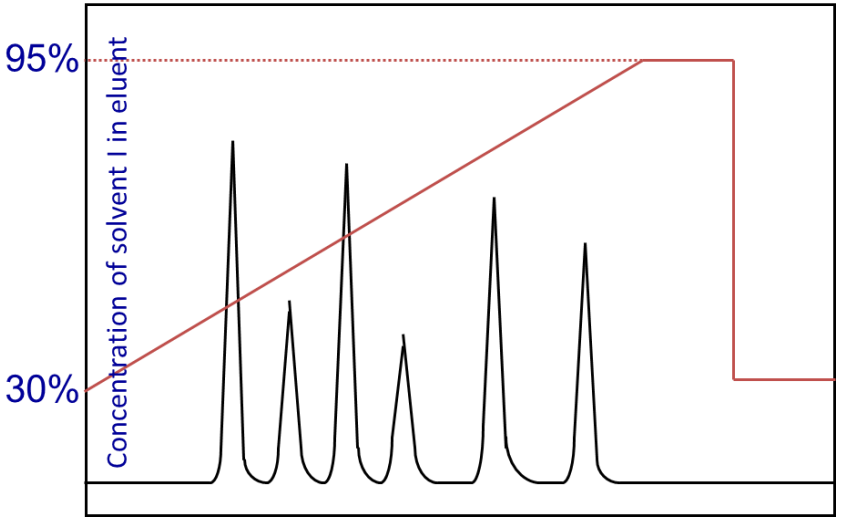
Pump delivers mobile phase to the chromatographic system in:

- **isoctratic mode**
(constant eluent composition)





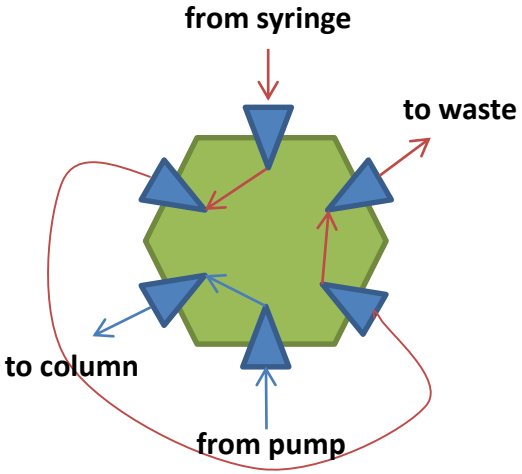
- **gradient mode**
(varying eluent composition)



LC injector

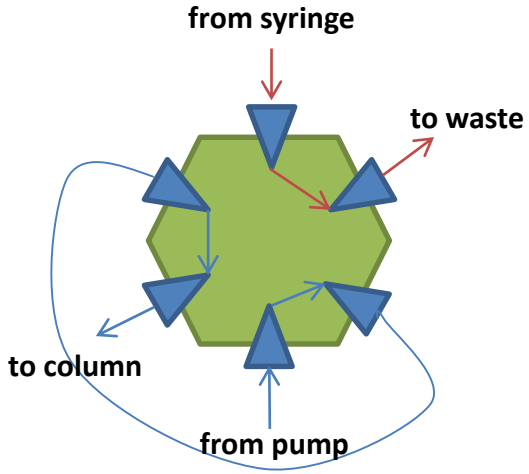
Injector used almost exclusively in LC is known as the *loop injector* (or six-port valve injector).

1. The sample is introduced, using a micro-syringe, into a mobile phase that fills a loop of a nominal volume.



LOADING the sample loop

2. While the loop is filled, the mobile phase is pumped through the valve into the column to keep the column in equilibrium with the mobile phase.



INJECTING the sample

Injector should perform injections with

- high reproducibility,
- accuracy,
- avoiding the presence of air bubbles or pulses.

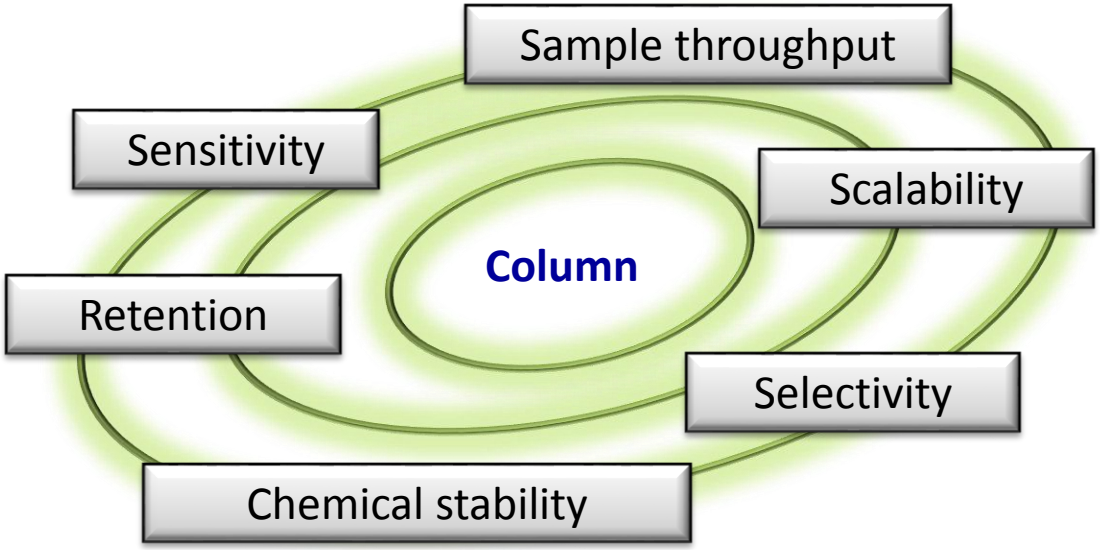
LC column

The format of an LC column refers to:

- the column length,
- column diameter
- particle size of the stationary phase.

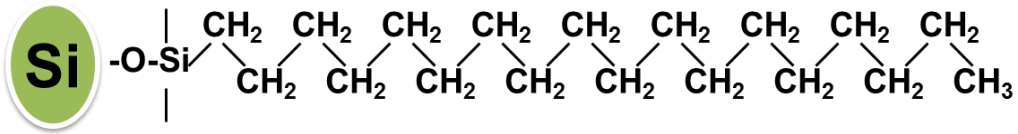
	Column lenght [mm]	Column diamter [mm]	Particle diameter [um]	Optimum flow-rate [µl/min]	Pressure [bars]
Nanobore column	50-1000	0.05-0.1	1-3	0.3	<300
Capillary column	50-1000	0.3	1-5	5	<500
Microbore column	50-1000	0.5-1	1-5	10-50	<800
Narrow(small)-bore column	50-250	2.1	2-5	400	<1200
Normal-bore column	30-250	4.6	2-5	1000	<400

LC column

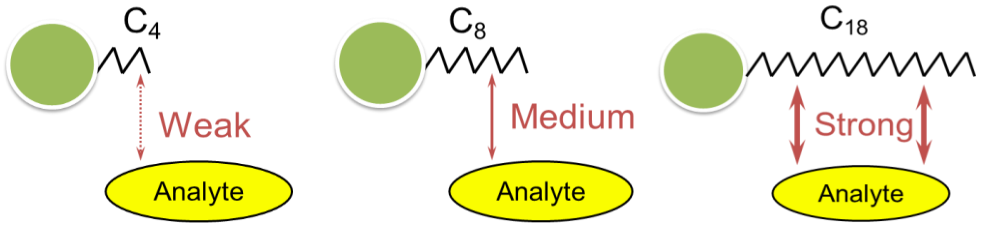


LC column packing materials:

- **C18 (ODS) type**
- C8 (octyl) type
- C4 (butyl) type
- phenyl type
- TMS type
- cyano type



C18 (ODS) type



the interaction between the analyte and column

Mobile and stationary phases

The nature of the analyte /the compounds to be separated/, determines **the stationary phase** and **the mobile phase** selection.

LC mode	Mobile phase	Stationary phase	Type of separated compounds
NP	organics: dichloromethane, ethyl acetate	silica, amino, cyano, diol	organic compounds not soluble in water
RP	water/organic with or without additives	C18, C8, C4, cyano, amino	neutrals, weak acids, weak bases
HILIC	acetonitrile with water, ionic additives	polar, pure silica	polar compounds
IEC	buffered aqueous solutions	anion or cation, exchange resin	ionic, inorganic ions

Normal phase chromatography (NP)

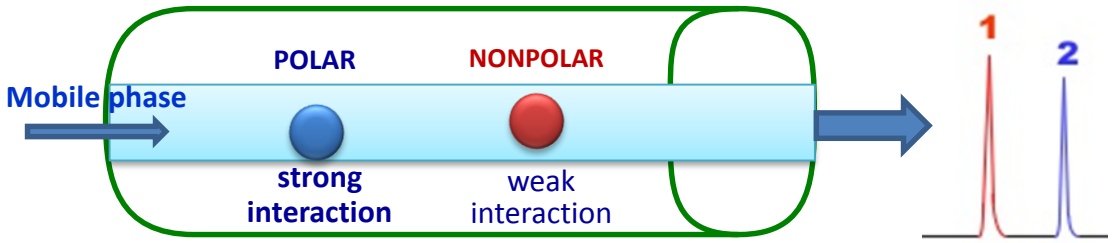
Chromatographic separation in NP results from interactions of separated compounds with **polar stationary phase** and **non-polar mobile phase**

Stationary phase used in NP:

- silica gel: -Si-**OH**
- cyano type: -Si-$CH_2CH_2CH_2$ **CN**
- amino type: -Si-$CH_2CH_2CH_2$ **NH_2**
- diol type: -Si-$CH_2CH_2CH_2OCH$(**OH**)-CH_2 **OH**

Mobile phase used in NP:

- hydrocarbons
- dichloromethane
- ethyl acetate
- other water-immiscible solvent



column

Reversed-phase chromatography (RP)

RP chromatography is the most common of all the methods used in HPLC

Chromatographic separation in RPresults from the interactions of separated compounds with **nonpolar stationary phase** and **polar mobile phase**

Stationary phase used in RP:
(long-chain hydrocarbons covalently bonded to the silica surface)

- **C18**
- **C8**
- **C4**
- **cyano**
- **amino**

The mobile phase in RP:

water or buffer solution, and organic solvents, among which the most frequently used are:

- **methanol**
- **acetonitrile**

NP-LC and RP-LC are used for different purposes in biological samples analysis:

- NP-LC is applied for separation of individual lipid classes **based on the polar head groups**,
- RP-LC is used for separation of lipid species **based on their different hydrophobicities** (fatty acyl chains)
- RP gradient chromatography is also useful for metabolite profiling in metabolomics studies

Ion-exchange chromatography (IEC)

Chromatographic separation in ion-exchange-phase results from the interactions of **ionic and ionizable compounds** with **ionic funcional groups of stationary phase**, usually with opposite charges than that of the analytes

anionic stationary phase → **separation of cations or positively charged molecules**

cationic stationary phase → **separation of anions or negatively charged molecules**

Mobile phase is highly polar, most frequently based on water, with some buffer or salts.

Increasing the ionic strength of the mobile phase is the main factor responsible for the elution of compounds from the IEC column.

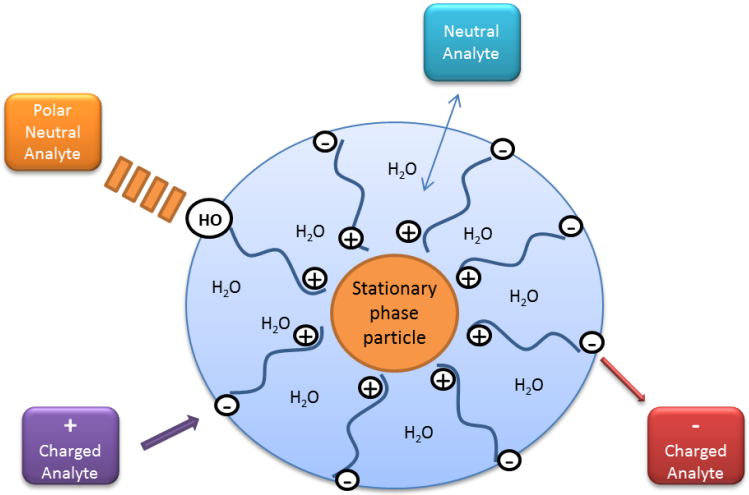
IEC is useful both for **large and small biomolecule** separations, such as of amino acids, carboxylic acids or amines.

Due to the ion suppression phenomenon, IEC, which utilizes high ion strength in the mobile phase, is relatively difficult to be directly coupled with a mass spectrometer.

Hydrophilic interaction chromatography (HILIC)

HILIC for **separation of polar compounds**

HILIHC separation mechanism bases on a **liquid/liquid extraction system** with water layer formation on the surface of the polar stationary phase and organic mobile phase



Schematic of interactions between different types of polar analytes and a stationary phase in HILIC mode

stationary phases

- **hydrophilic**
- HILIC columns typically contain silica polar surfaces or it can be derivatized to amino or amide bonded phases

mobile phase

- **solvent system typical for RP**, most frequently acetonitrile, with a small amount of water
- ammonium acetate or ammonium formate are often added to the mobile phase to increase polarity and ion strength

HILIC is easily adaptable to MS

The use of organic solvents increases MS sensitivity due to a decrease of ion suppression

Detectors

LC detectors:

- UV,
- fluorescence,
- electrochemical,
- conductivity,
- refractive index,
- **MS detectors**

DETECTORS PARAMETERS

selectivity
- ability to determine an analyte of interest without interferences derived from the matrix, the solvents or other substances present in the system

detection limit
- the smallest concentration of an analyte that is sufficient for its detection with a set probability

sensitivity
- the detector response that is related to the concentration of an analyte which reaches the detector.

Mass spectrometers - ideal detectors for both qualitative and quantitative analysis

Introduction to GC-MS

Gas chromatography (GC) is a separation technique capable of separating highly complex mixtures based primarily upon differences of **boiling point/ vapor pressure and of polarity**.

In **GC:**

- the **mobile phase** is a gas (Ar, He, N_2 or H_2)
- the **stationary phase** is either:
  - a solid (adsorbent)- Gas Solid Chromatography (GSC) or
  - an immobilized polymeric liquid - Gas Liquid Chromatography (GLC)

Separation in GC depends on

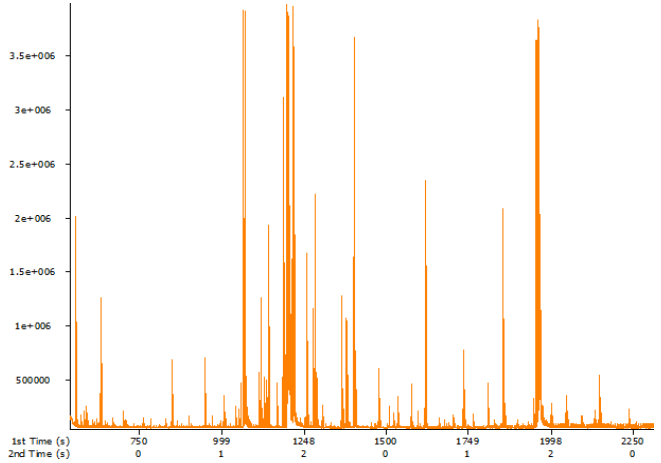
the transfer of a substance (as steam) using carrier gas (mobile phase) through a column.

The rate and degree of compounds partitioning in GC depends upon :

- the **chemical affinity of the analyte for the stationary phase**

and

- the **analyte vapor pressure** – which is governed by the column temperature



Strategies for the GC determination

Sample collection

gaseous
liquid
solid

Pre-treatment

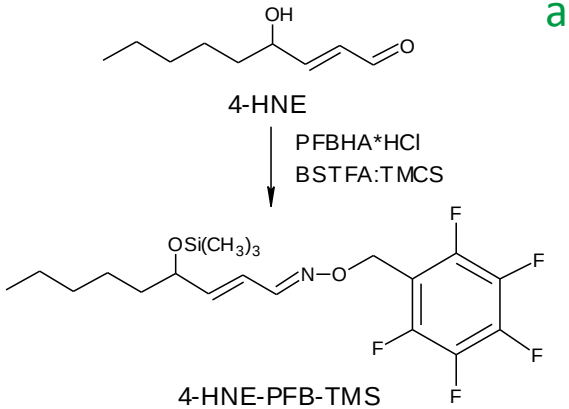
drying
filtration
homogenization
...

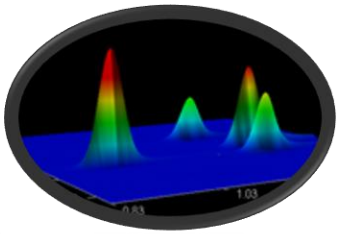
Sample preparation

SPE
SPME
...

Derivatization

silylation
alkylation
...





GC, GCxGC analysis

FID
MS
...

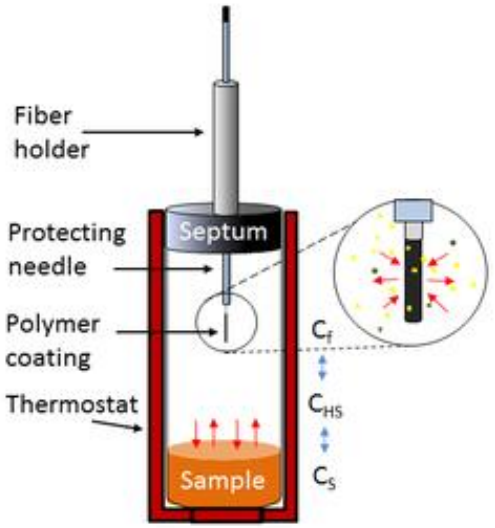
Sample collection

The sample to be analyzed by **GC** could be:

- a gas
- a liquid
- molecules adsorbed on a surface after solid-phase microextraction (SPME)

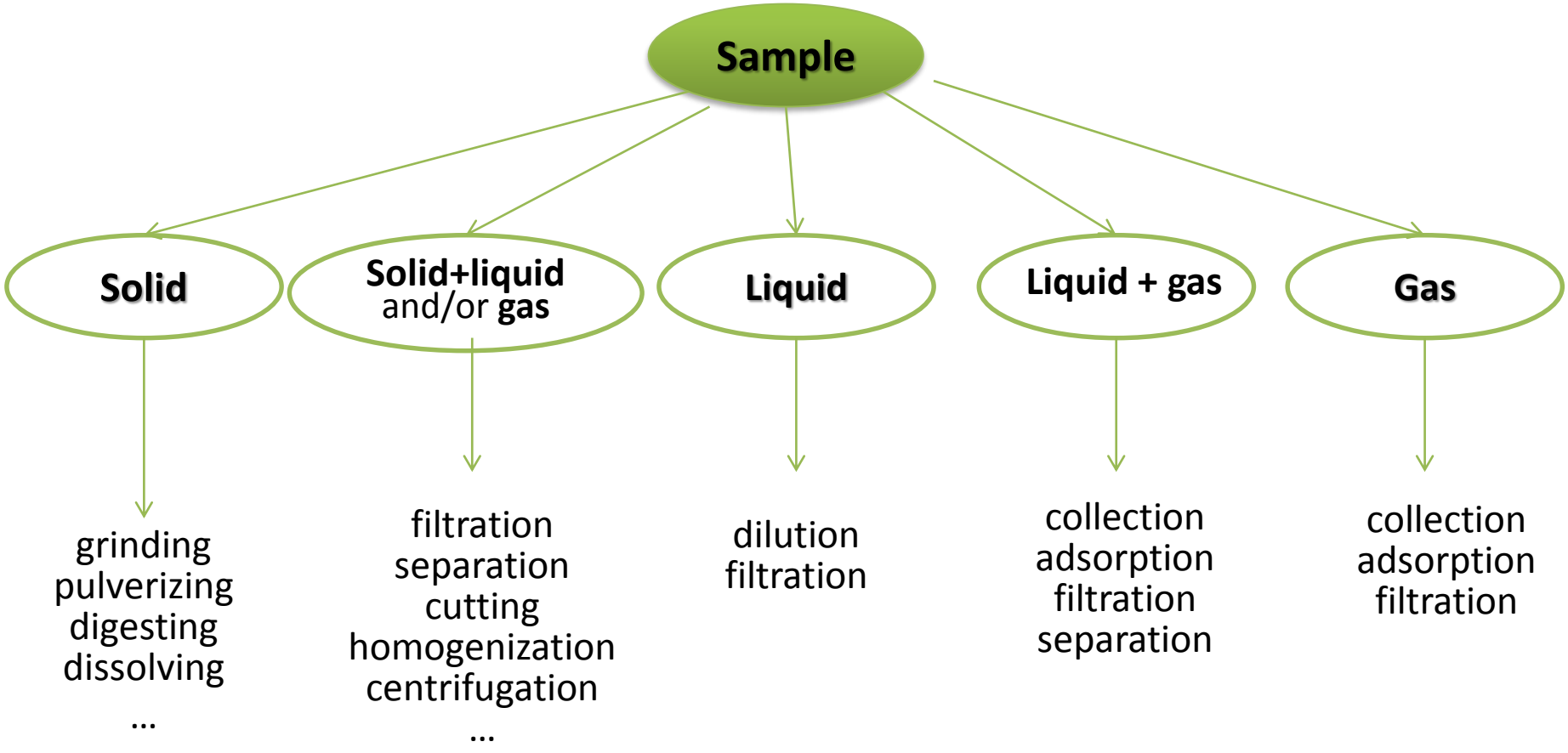
Stabilization of the composition and properties of the samples is often obtained by adding appropriate agents that:

- inhibit biological activity organisms present in the samples
- eliminating the occurence of adsorption of sample components on vascular walls
- eliminating volatilization, thermal decomposition, chemical reactions, etc.

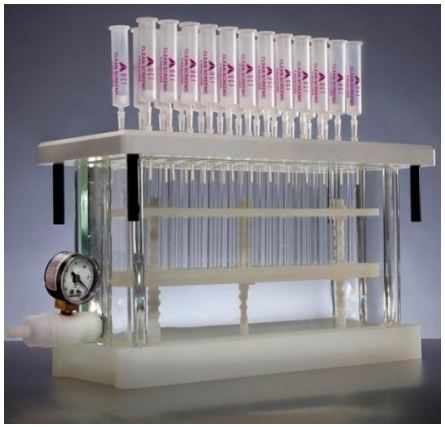


Scheme of SPME

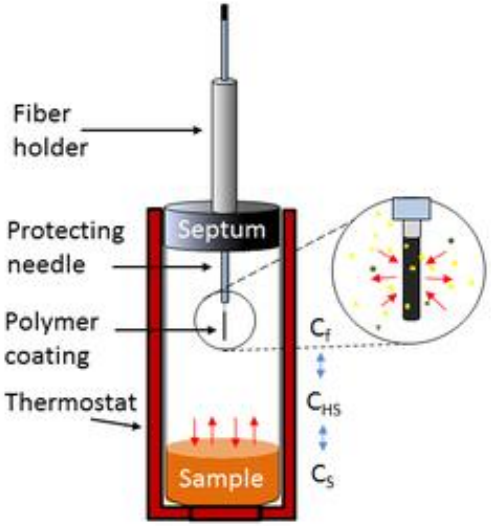
Sample pre-treatment



Sample preparation



SPE equipment



Scheme of SPME

- Extraction techniques of liquid samples
  - Liquid-gas extraction
    - Static headspace analysis (SHSA)
    - Dynamic headspace analysis (DHSA)
    - Purge and trap (PT)
  - Liquid-liquid extraction
    - Classical liquid-liquid extraction (LLE)
    - Continuous liquid-liquid extraction (LLCE)
    - Microextraction (ME)
    - Single drop microextraction (SDME)
    - Stir membrane liquid-liquid microextraction (SM-LLME)
- Extraction techniques of gaseous samples
  - Gas-liquid extraction
    - Extraction to the liquid phase
    - Extraction with using of the semipearmeable membranes (SPMD)
  - Liquid-solid phase extraction
    - Solid phase extraction (SPE)
    - Solid phase microextraction (SPME)
    - Rotating disc extraction (RDE)
    - Membrane Solid phase microextraction (M-SPME)
    - Moleculary imprinted solid-phase extraction (MI-SPE)
    - Dispersive solid-phase extraction (DSPE)
  - Gas-solid extraction
    - Polymeric membrane extraction (PME)

Extraction methods of liquid and gaseous samples

Sample preparation

Microwave extraction system

Thermodesorber

- Extraction techniques of solid samples
  - Solid-gas extraction
    - Static headspace analysis (SHSA)
    - Dynamic headspace analysis (DHSA)
    - Thermal desorption (TD)
  - Solid-liquid extraction
    - Classical methods
      - „Shake extraction”
      - Extraction in a solvent stream
      - Soxlhet extraction
      - Sample homogenization
      - Multi-stage (sequentional) extraction
      - Saponification
    - Modern extraction techniques
      - Classical method - supported
        - Ultrasound assisted extraction (USE)
        - Microwave assisted extraction (MAE)
      - Accelerated solvent extraction (ASE)
      - High- pressure extraction (HPE)
      - Microwave-assisted solvent extraction (MASE)
      - Supercritical fluid extraction (SFE)

Extraction methods of solid samples

Derivatization

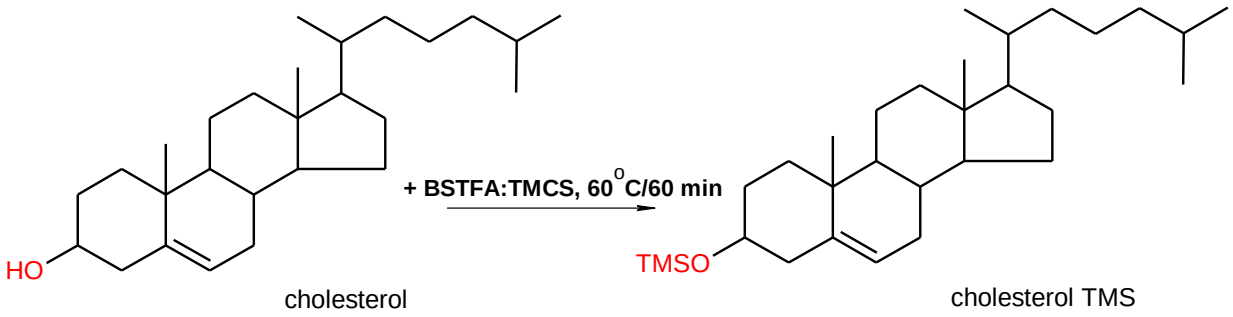
The direct analysis of compounds mixture in GC is difficult due to:

- too low volatility of analysed compounds
- too low thermal stability of analsed compounds
- interactions between the compounds
- interactions between the compounds and the GC column stationary phase
- too low sensitivity or specificity of the assay

Therefore, the main way to prepare a sample for analysis is conversion of analytes to derivatisation products

Derivatization of analytes is performed as a result of

- silylation
- alkylation
- acetylation



Reaction of silylation

Derivatization

Procedure	Functional group - Compound type	Derivative	Reagent
Silylation	-OH -alcohols, phenols -CO -ketones, steroids -COOH -amino acids, fatty acids, steroids -(CH_2OH)n -sugars	Trimethylsilyl ethers	**Bistrimethylsilyltrifluoroacetamide (BSTFA)** N- methyl-N-t-butyldimethylsilyl-trifluoroacetamide (MTBSTFA) N- methyltrimethylsilyltrifluoroacetamide (MSTFA)
	-NH, -NH_2 -amines, urea -CONH, -$CONH_2$ -imides, proteins	Trimethylsilyl amides	**Trimethylsilylimidazole (TMSI)** Halo- methylsilyl reagents
Alkylation	-OH -alcohols, phenols -CO -aldehydes -COOH -amino acids, fatty acids -NH, -NH_2 -amines, amino sugars -CONH –amides -SH -mercaptans	Methyl esters (DMF) Trifluoroacetates (TFAA) Methyl esters (BF_3-methanol) Pentafluorobenzyl ethers (PFBBr) Methyl amides (TMAH) Methyl esters (DMF)	Benzylbromide Boron trifluoride (BF_3) in methanol or butanol Dimethylformamide (DMF) Pentafluorobenzyl- hydroxylamine hydrochloride (PFBHA) Tetrabutylammonium hydroxide (TBH) Trifluoroacetic anhydride (TFAA)
Acylation	-OH -alcohols, phenols -(CH_2OH)n -sugars -NH, -NH_2 -amines -CONH -amides -SH -mercaptans	Pentafluoropropionates (PFPA) Trifluoroacetamides (TFAI) Trifluoroacetamides (MBTFA) Trifluoroacetamides (TFAA) Trimethylsilyl ethers (MBTFA)	Heptafluorobutyric anhydride (HFBA) N-Methyl-bis(trifluoroacetamide) (MBTFA) Pentafluorobenzoyl chloride (PFBCI) Pentafluoropropanol (PFPOH) Trifluoroacetic anhydride (TFAA)

Gas chromatograph

The gas chromatograph consists of

- ***injector***

the sample is injected into the heated injection port where it is volatilized and carried into the column by the carrier gas - inert purity ≥ 99,9995% (Ar, He, N_2 or H_2)

- ***GC column***

the sample is separated inside the column

- ***detector***

responds to some physicochemical property of the analyte and generates an electronic signal measuring the amount of analyte present

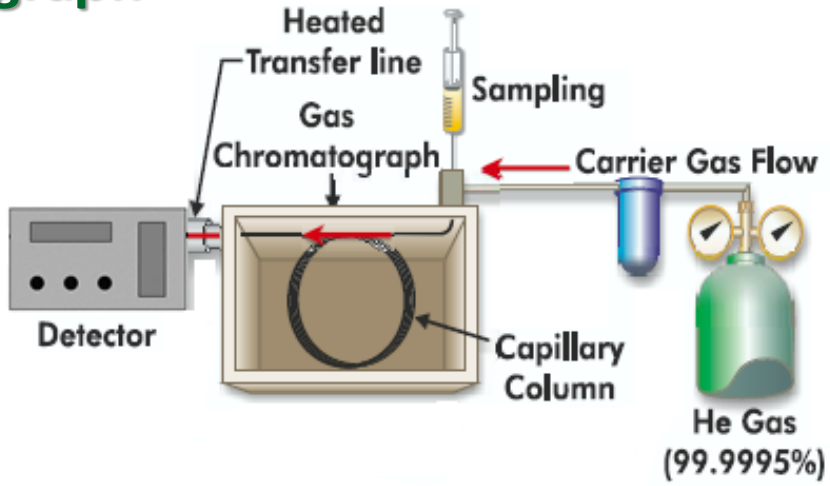
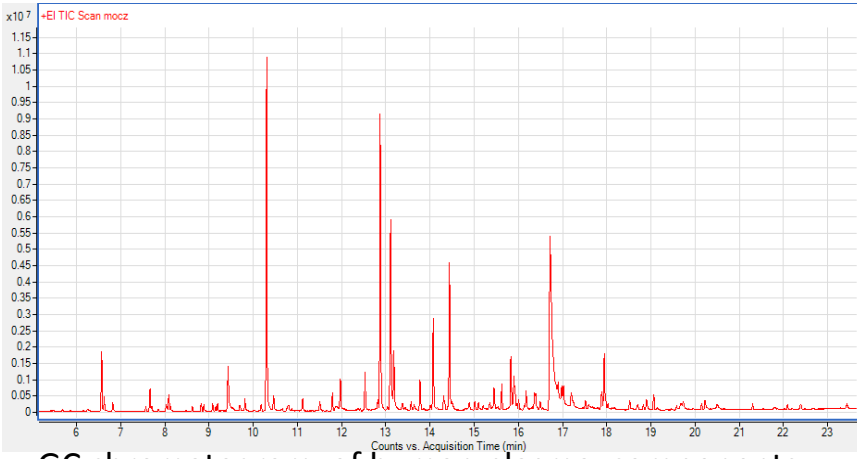


Figure. Diagram of the GC system
(modified from *de.leco-europe.com*)

chromatogram - result of chromatographic separation



GC chromatogram of human plasma components

GC injector

Stages of injection

1. The sample is injected into the heated injection port
2. In the injection port the sample is volatilized
3. The carrier gas entraines volatilized sample into the carrier stream entering the GC column

Types of **GC** injectors are **Split/Splitless (SSI)**

Split injection

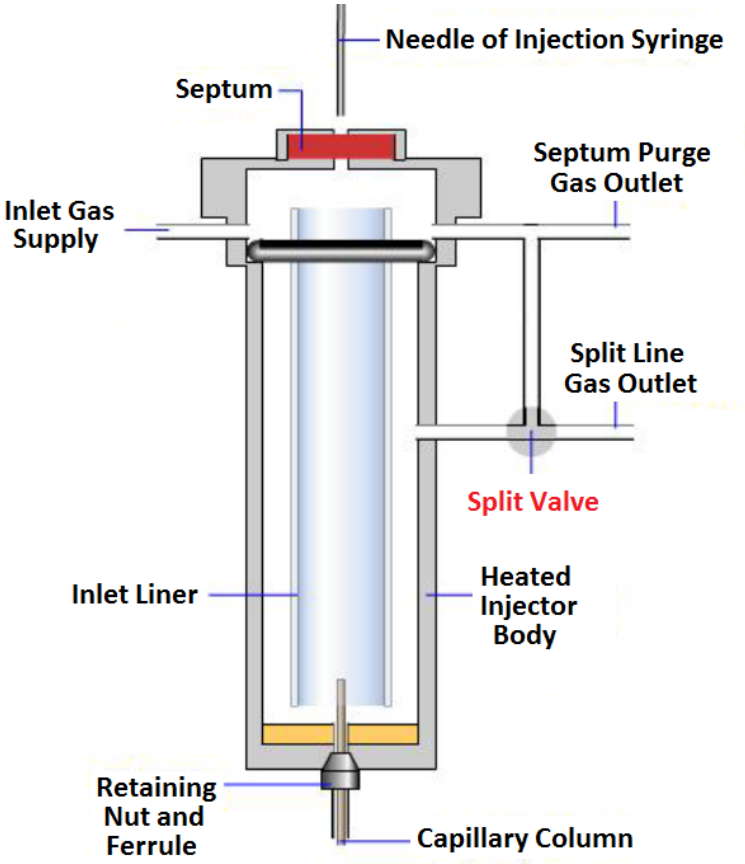
only a small portion of the vaporized samples is applied to the column (usually 1/20 to 1/500)

- primarily used for non-trace analysis of volatile samples

Splitless injection

the whole sample reaches the column

- primarily used for trace and ultra-trace analysis



GC columns

Packed column (analytical, micro packed) filled with solid particles

- carbon adsorbents
- silica
- alumina
- molecular sieves
- porous synthetic polymers

Adsorbents are less common stationary phases due to the:

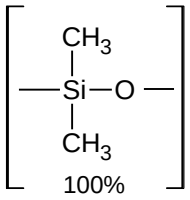
- lower reproducibility of the results
- longer retention time
- appearance of "tails" with much lower separation efficiency

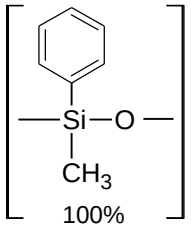
Capillary column (capillary, microcapillary)

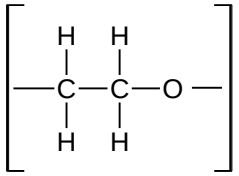
- open tubular capillary tubes are embedded with a liquid
  - silicones
  - squalene
  - polyethylene glycol

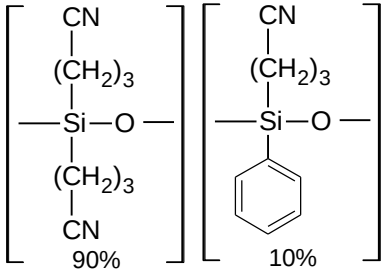
- liquid stationary phase should be:
  - chemically inert
  - capable of dissolving separated components
  - highly selective for the components of the mixture
  - low volatility
  - thermal stability under the operating conditions of the column
  - particularly suitable for the separation of gaseous components with high separation efficiency
- are frequently used

Column phases









Increasing polarity

Methylpolysiloxane

Methylpolysiloxane + 5% phenyl

Methylpolysiloxane + 50% phenyl

Methylpolysiloxane + 7% cyanopropyl + 7% phenyl

Methylpolysiloxane + 25% cyanopropyl + 25% phenyl

Polyethylene Glycol (PEG)

Methylpolysiloxane + 70% cyanopropyl

Polysiloxane + ≥90% cyanopropyl

Increasing selectivity

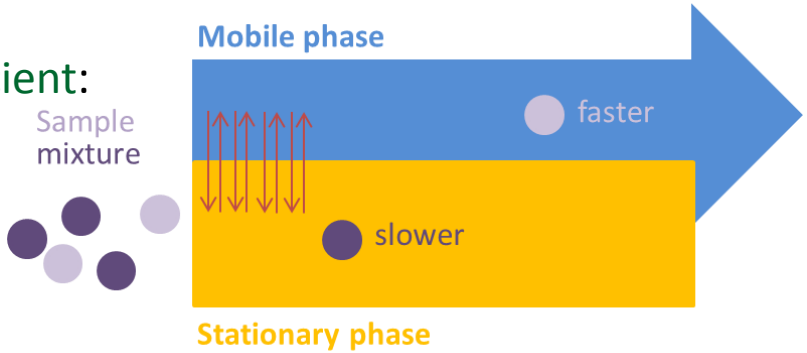
Separation mechanism

The rate and degree of compounds partitioning in the GC column is a function of the distribution of these components in two phases (**mobile and stationary**) remaining in equilibrium

Separation is the result of different migration rates caused by different values of the **partition coefficient (Ks)**

The **Nernst equation** can express the partition coefficient:

$$K_S = C_L/C_G$$



C_L - the concentration of the substances in the stationary phase
C_G - the concentration of the substances in the mobile phases

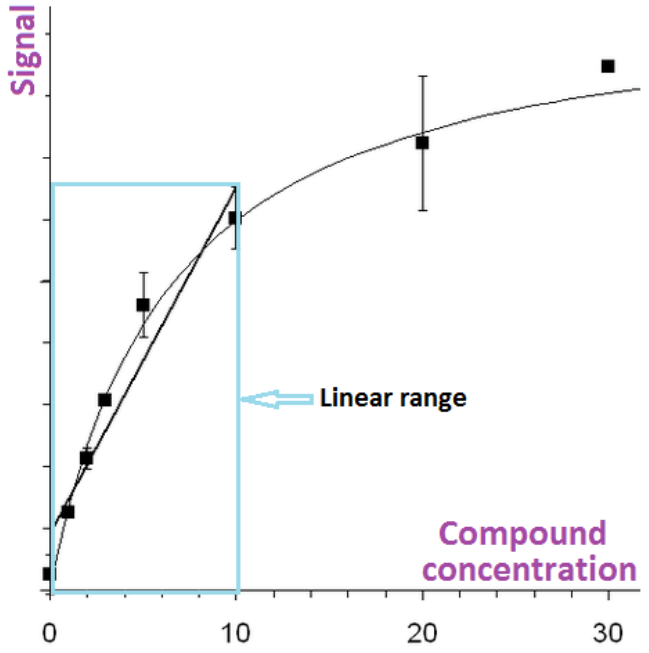
$K_{slower} > K_{faster}$

The higher affinity for the stationary phase material, the higher value of Ks and the higher the value of the retention time (t_R)

GC detectors

Characteristic of ideal GC detector

- high sensitivity
- good stability and reproducibility
- wide range of linear response to solutes that extends over several orders of magnitude (calibration purposes)
- a wide temperature range
- a short response time independent of flow rate
- high reliability and ease of use
- similarity in response toward all solutes
- the detector should be nondestructive

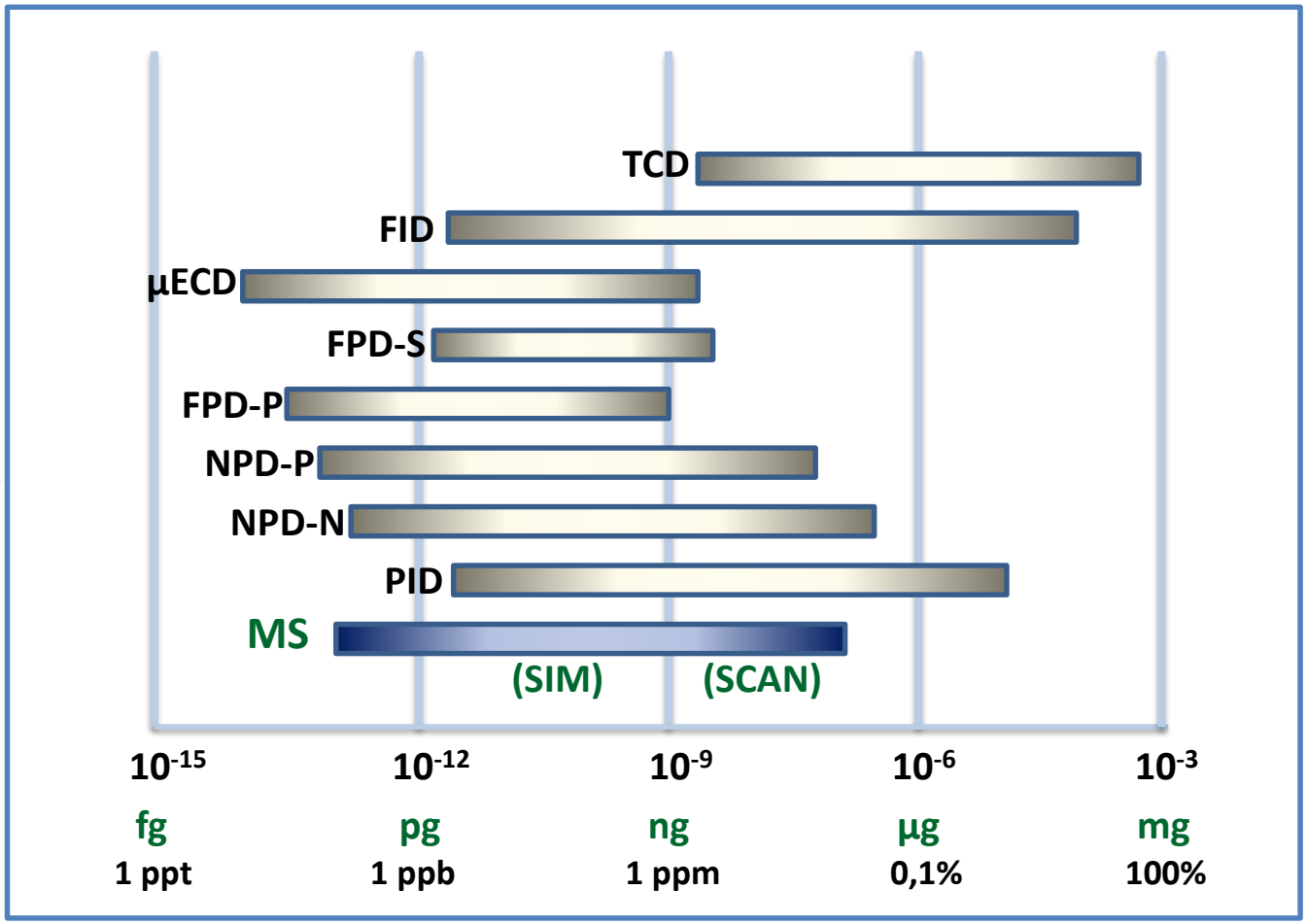


GC detectors

Type	Type of response	LOD [g/sec]	Linear range	Comments
Thermal conductivity (TCD)	**universal**	**10^{-5}-10^{-6}**	**10^{3}-10^{4}**	**measures changes in heat conduction**
Flame ionization (FID)	**universal**	**10^{-12}**	**10^{6}-10^{7}**	**measures ion currents from pyrolysis**
Electron capture (ECD)	selective	10^{-14}	10^{2}-10^{3}	detector for compounds containing atoms with high electron affinities
Flame photometric (FPD)	selective	10^{-13}	10^{2}	detector for compounds containing S, P
Nitrogen-phosphorous (NPD)	selective	10^{-8}- 10^{-14}	10^{5}-10^{7}	selective for compounds containing N, P
Photoionization (PID)	selective	10^{-8}- 10^{-12}	10^{5}	selectivity due to identify of gas in lamp
Fourier-transform infrared (FTIR)	selective	10^{-10}	depends of the functional group	polar molecules
Mass spectrometer (MS)	**universal**	**10^{-12}**	**depends of the type of MS analyzer, operation mode and compound**	**non-destructive detector**, one of the most accurate and efficient tools for analyzing organic samples, **the most powerfull detectors for GC**

GC detectors

Comparison of GC detectors sensitivity and dynamic range

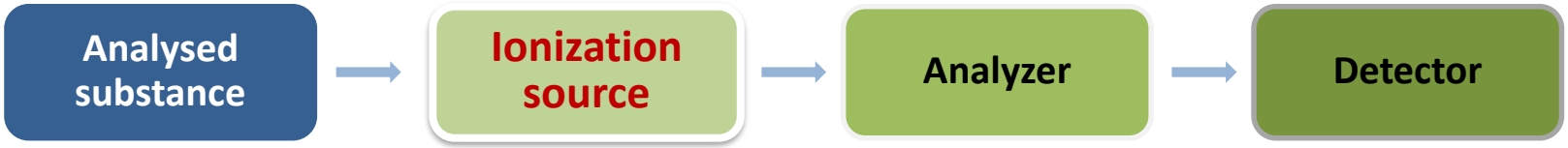


Gas chromatography – mass spectrometry (GC-MS)

GC-MS

- is a versatile tool to:
  - separate
  - identify
  - quantify unknown substances
- is the most effective technique for the analysis of volatile organic compounds in complex matrices in a wide range of concentrations (from ppb to ppm)
- is characterized by high selectivity and sensitivity, providing a **wide range of applications**:
  - medical and pharmaceutical applications
  - biological analysis
  - forensic and criminal applications
  - environmental monitoring
  - security and chemical warfare agent detection
  - food/flavor/fragnance analysis
  - chemical/industrial applications
  - geochemical research
  - petrochemical analysis

GC-MS instrumentation



Ionization techniques

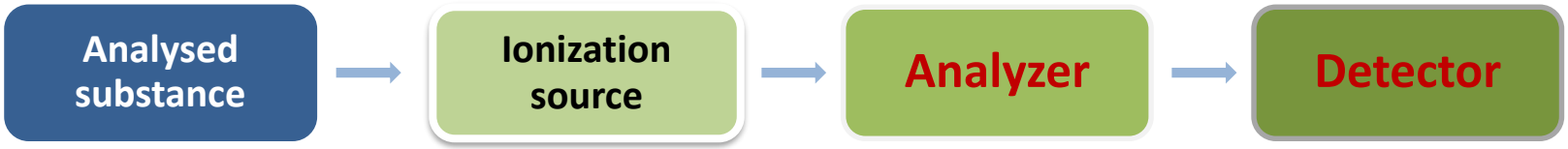
EI - electron ionization - the most commonly used ionization technique in GC-MS, hard ionization

CI - chemical ionization - relatively mild (so-called „mild ionization")

other ionization techniques used in GC-MS:

- photoionization (PI)
- field ionization (FI)
- field desorption (FD)
- laser desorption (LD)
- fast atom bombardment (FAB)
- plasma desorption (PD)
- secondary ion mass spectrometry (SIMS)
- matrix-assisted laser desorption ionization (MALDI)

GC-MS instrumentation



Analyzer - separate charged ions according to their m/z ratio

- **Q** - quadrupole (singleQ, tripleQ)
- **IT** – ion traps (linear, spherical)
- **TOF** – „time-of-flight"
- **hybrid MS analyzers:**
  - QqQ
  - Q/IT
  - Q/TOF
  - Orbitrap
- **other analyzers used in GC-MS:**
  - magnetic (B) sector
  - electric (E) sector
  - electric and magnetic sector
  - ion cyclotron resonance (ICR)
  - Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR MS)

Detector - data registration

- Faraday cup
- electron multiplier
- microchannel plates
- photomultiplier

GC-MS

Advantages

- Fast analysis
- High efficiency – leading to high resolution
- Sensitive detectors (ppb)
- High quantitative accuracy (<1% RSD)
- Non-destructive – (coupling to MS)
- Requires small samples (<1 mL)
- Rugged and reliable techniques
- Well established with extensive literature and applications

Disadvantages

- Limited to volatile samples or derivatisation is required
- Not suitable for thermally labile samples that degrade at elevated temperatures - derivatisation is required
- Not suited to preparative chromatography
- Requires MS detector for analyte structural elucidation (characterization)
- Most non-MS detectors are destructive
- The limited peak capacity in analysis of very complex samples – **GCxGC separation is required**

Two-dimensional gas chromatography (GC×GC; 2D GC)

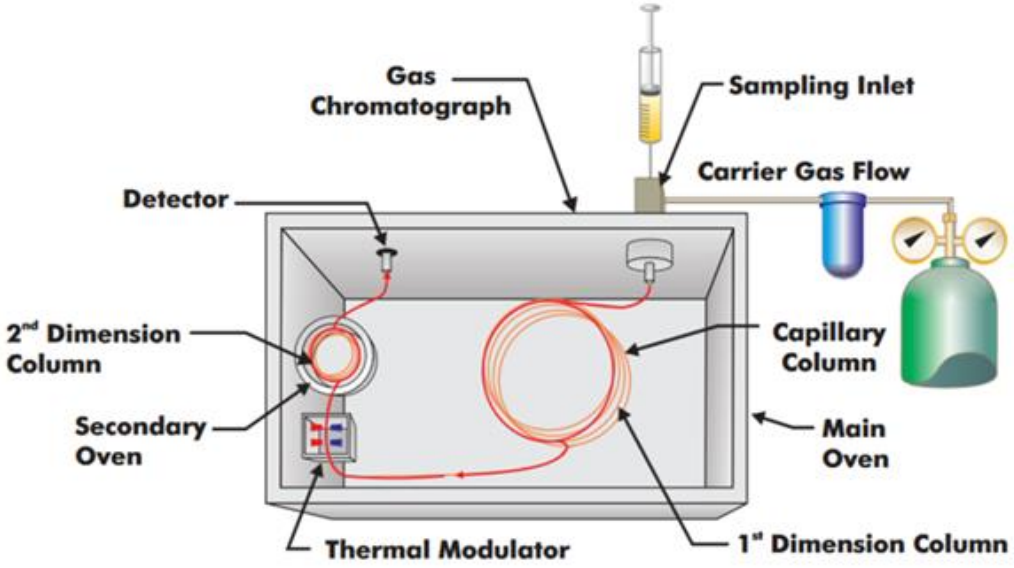
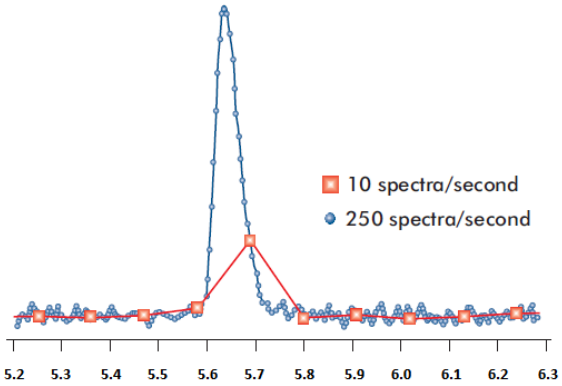


Diagram of the GC×GC system (*de.leco-europe.com*)

GCxGC detectors

- **FID** (max. 300 Hz)
- **TOFMS** (max. 500 Hz)
- **µ-ECD** (50-100 Hz)



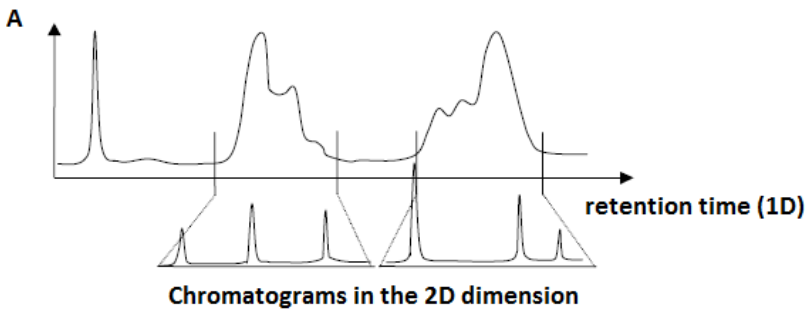
Acquisition rate

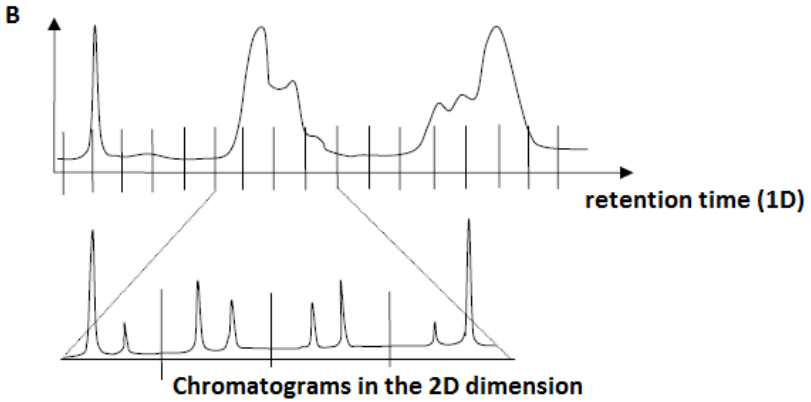
GC×GC columns

Paremeters	1D column	2D column
lenght	15-30 m	0,5-2m
inner diameter	0,25 mm	0,1 mm
film thickness of the stationary phase	0,25-1 µm	0,1-0,25 µm
type of stationary phase	non-polar 100% polydimethylsiloxane or 5% phenyl/95% dimethylsiloxane	polar 50% phenyl/50% dimethylsiloxane or polyethylene glycol (Carbowax)

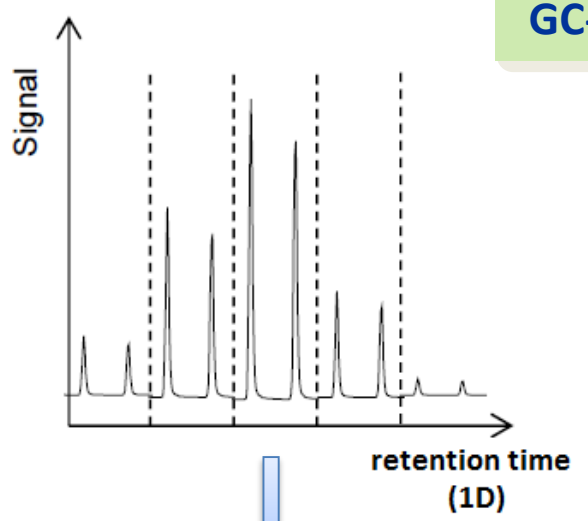
GC×GC

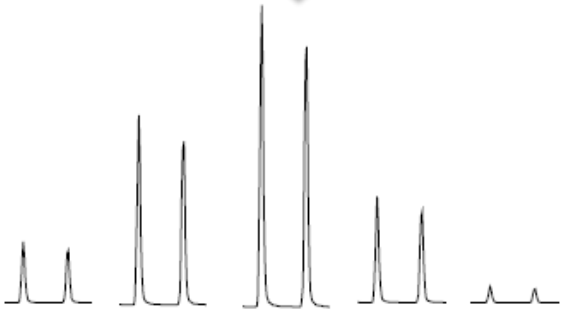
GC-MS

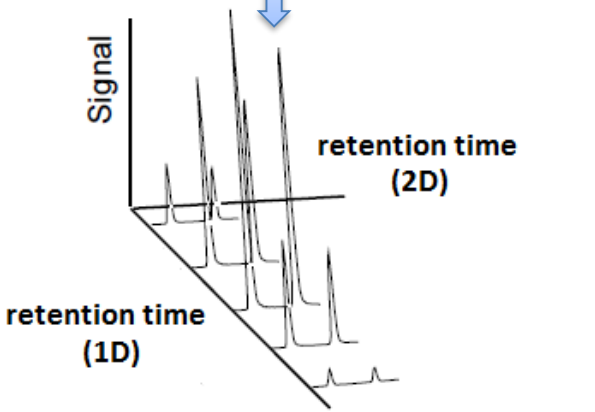




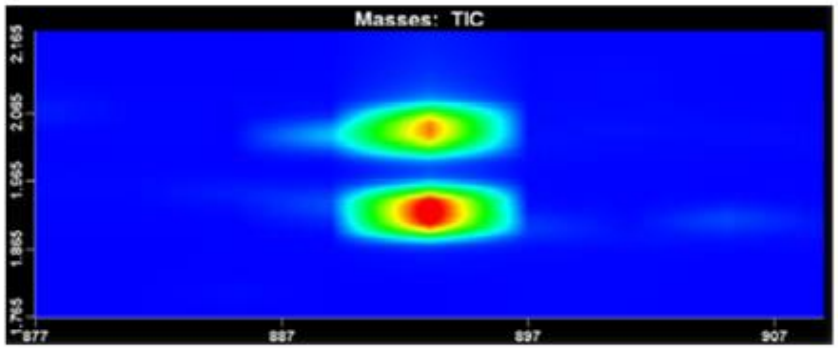
(A) two fractions leaving I column are directed in turn to the II column
(B) comprehensive two-dimensional GC



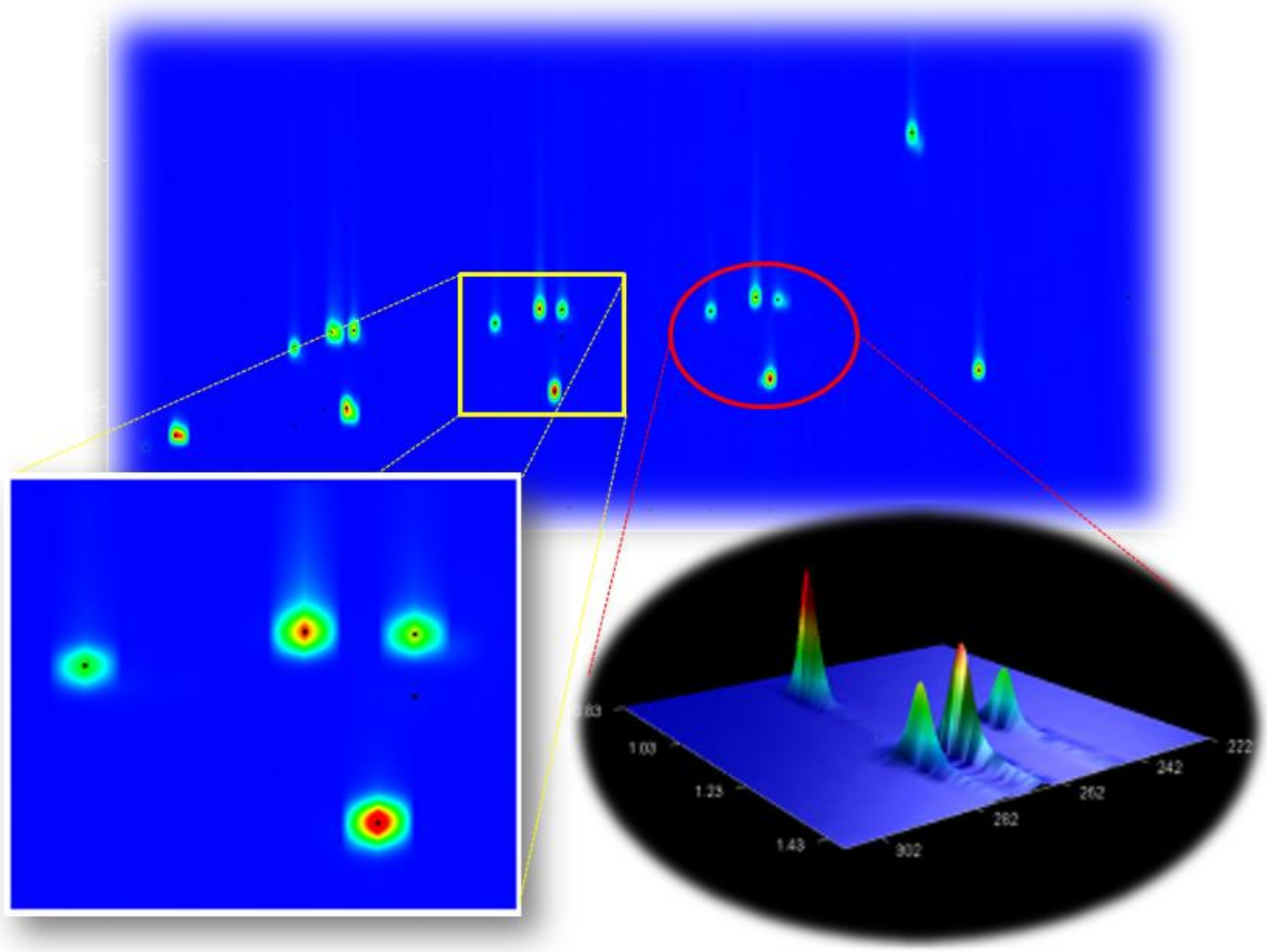




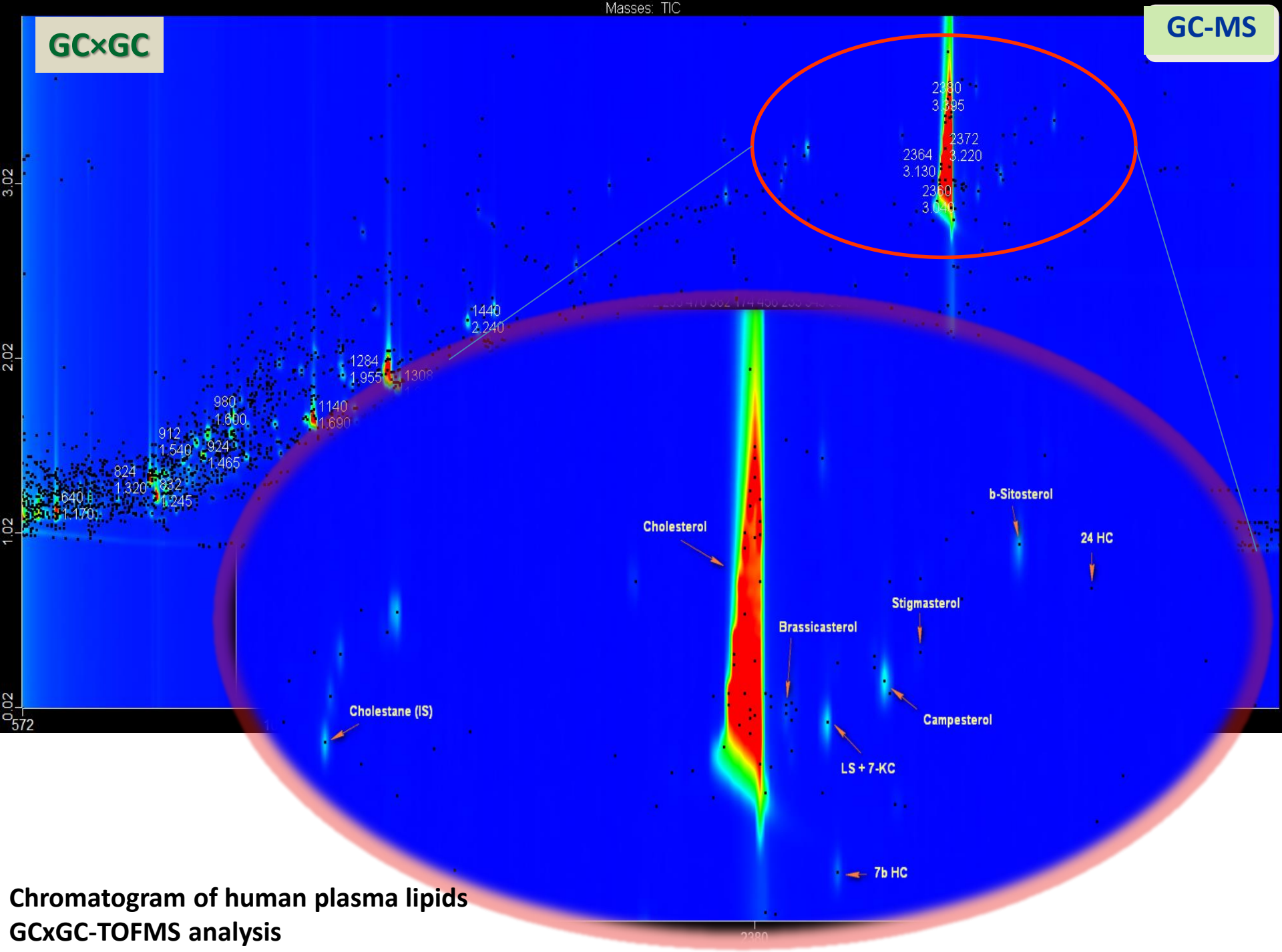
Contour plot



GC×GC



Chromatogram of the mixture of volatile compounds (alkanes, alcohols, aldehydes, ketones, amines) analyzed in 2D GC [Pegasus 4D; Leco].



Chromatogram of human plasma lipids
GCxGC-TOFMS analysis

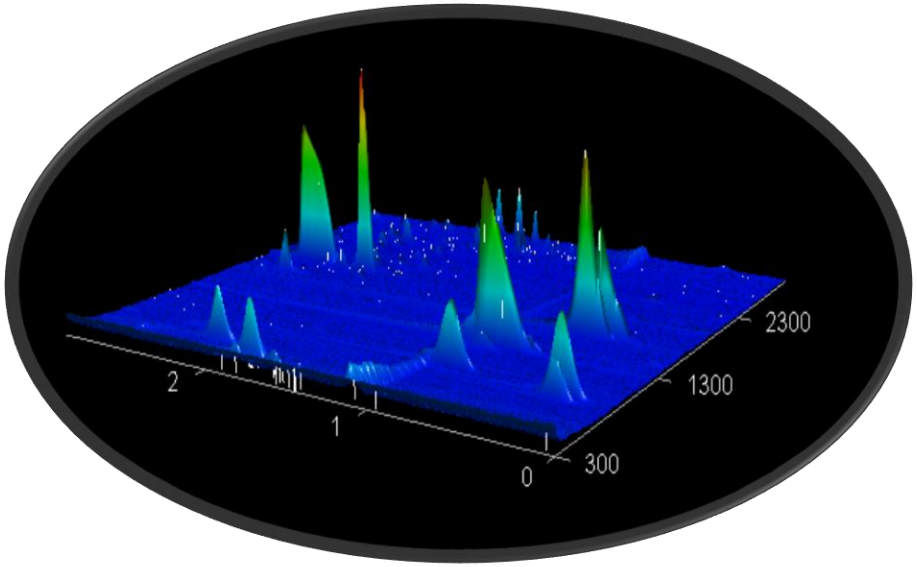
GC×GC

Advantages

- Increased peak capacity - separation of complex mixtures is possible
- Total separation of all components
- Detection limits (signal/noise ratio) is improved (in comparison with 1D GC)
- Structured chromatograms obtained
- 2D GC chromatogram contains much more information than 1D GC - easier and more reliable identification of unknown substances
- The ordered nature of chromatograms makes group analysis much easier

Disadvantages

- Required detectors with a high acquisition rate (TOF, µECD, FID)
- Large capacity of registered data - data processing can be very time–consuming
- Very high cost of equipment



AACLifeSci

Erasmus+

e-materials from

Advanced Analytical Chemistry for Life Sciences [AACLifeSci] project

This project has been funded with support from the European Commission.

Capillary Electrophoresis-Mass Spectrometry (CE-MS)

Coral Barbas
Antonia Garcia

Capillary Electrophoresis

- Separation based on electrophoretic mobility
- Simple instrumentation
- Primary applications in bioanalysis
  - DNA sequencing
  - DNA fragment analysis
- Multiple modes for improved selectivity of neutrals
  - MEKC
  - CEC

Advantages and Disadvantages of CE

Advantages

Offers new selectivity, an alternative to HPLC
Easy and predictable selectivity
High separation efficiency (10^5 to 10^6 theoretical plates)
Small sample sizes (1-10 ul)
Fast separations (1 to 45 min)
Can be automated
Quantitation (linear)
Different "modes" (to be discussed)

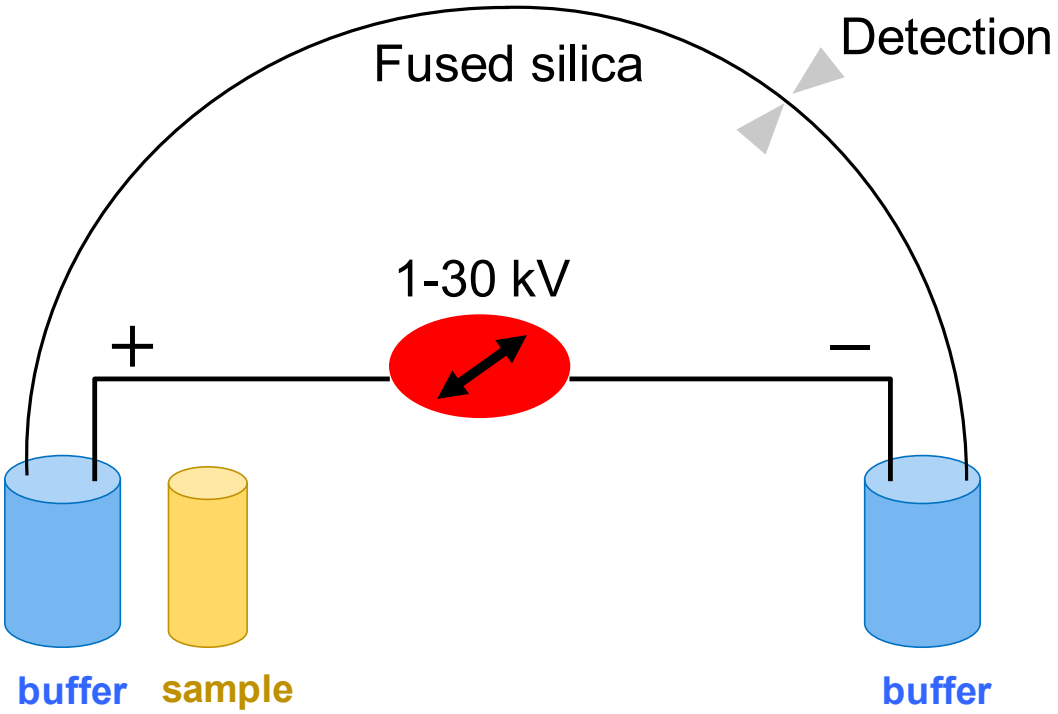
Disadvantages

Cannot do preparative scale separations
Low concentrations and large volumes difficult
"Sticky" compounds
Species that are difficult to dissolve
Reproducibility problems

Applications of CZE

- Wide variety of applications
  - Small molecules
  - Macromolecules (proteins, peptides)
- Limitations
  - Must have different charges
  - Low ionic strength sample
- Advantages
  - Simple
  - Direct analysis of complex systems

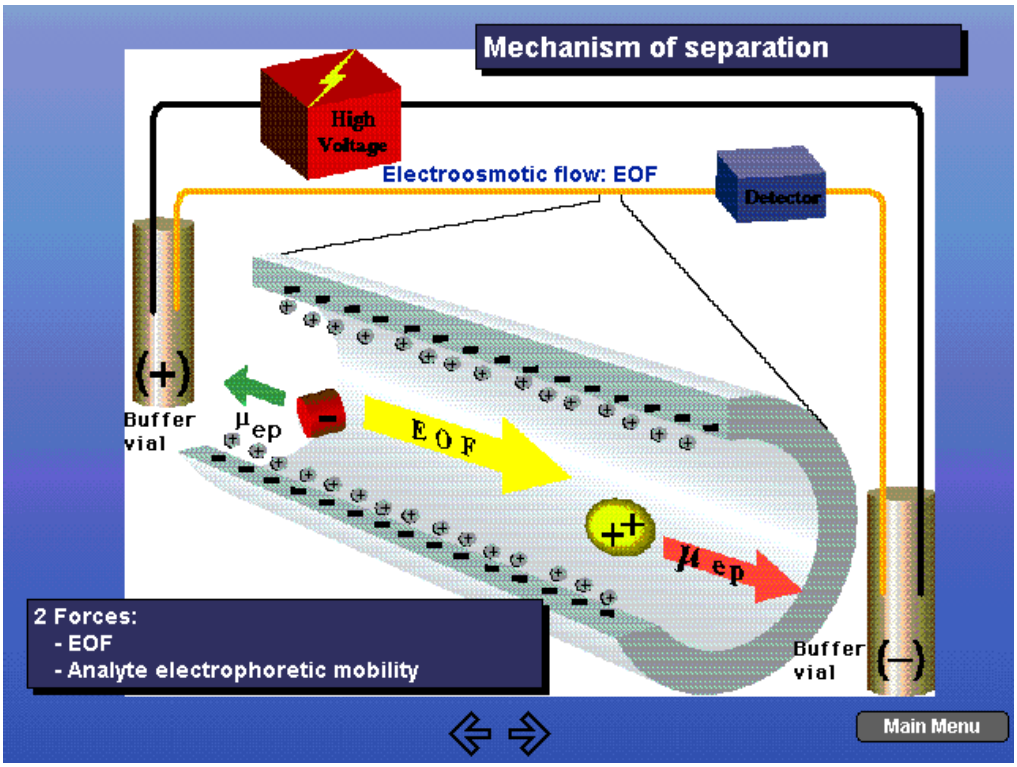
Capillary Electrophoresis (CE)



Modes of CE

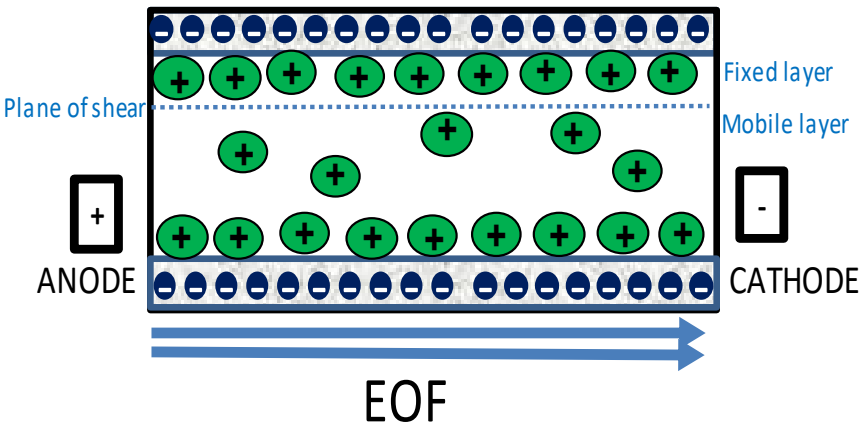
- Capillary Zone Electrophoresis (CZE)
  - Basic mode using open channels
- Micellar Electrokinetic Chromatography (MEKC)
  - Separates compounds with micelles
- Capillary Gel Electrophoresis
  - Size exclusion using sieving gels
- Capillary Electrochromatography
  - Hybrid of CE and HPLC
- Capillary Isoelectric Focusing
- Enantiomeric CE

Capillary Zone Electrophoresis (CZE)



Electroosmotic Flow

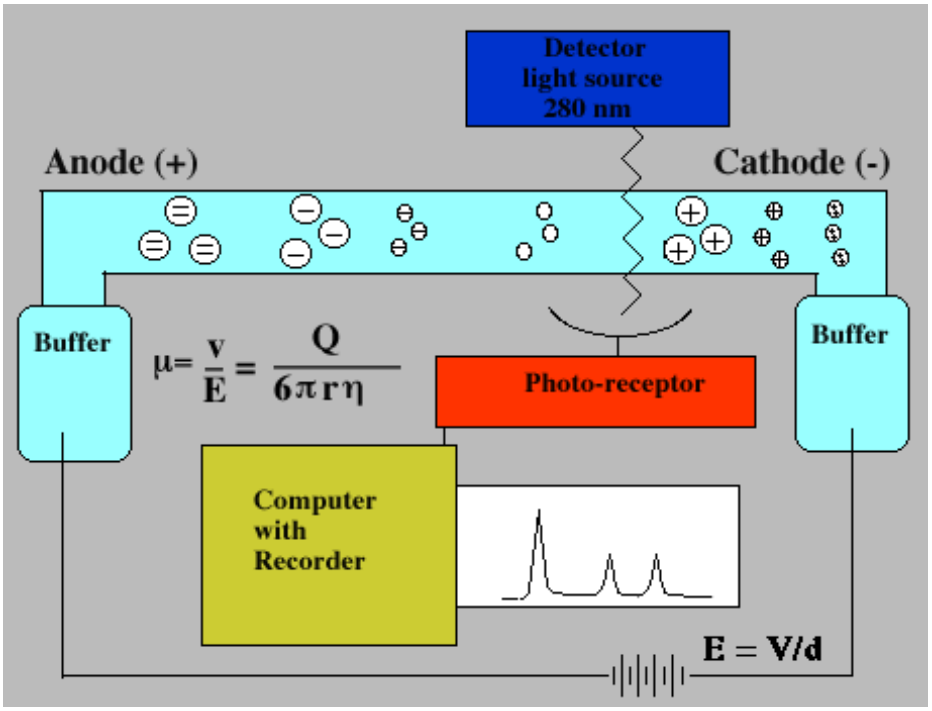
- Capillary flow mechanism based on applied potential and pH
- Provides bulk solution flow in capillaries with moderate to low concentration buffers are used
- Very dependent on solution ionic strength and surface chemistry



Electrophoretic Mobility

$$\mu = \frac{q}{6\pi\eta r}$$

μ= electrophoretic mobility
Q= charge on the particle
η= solution viscosity
r = Stokes radius of the particle

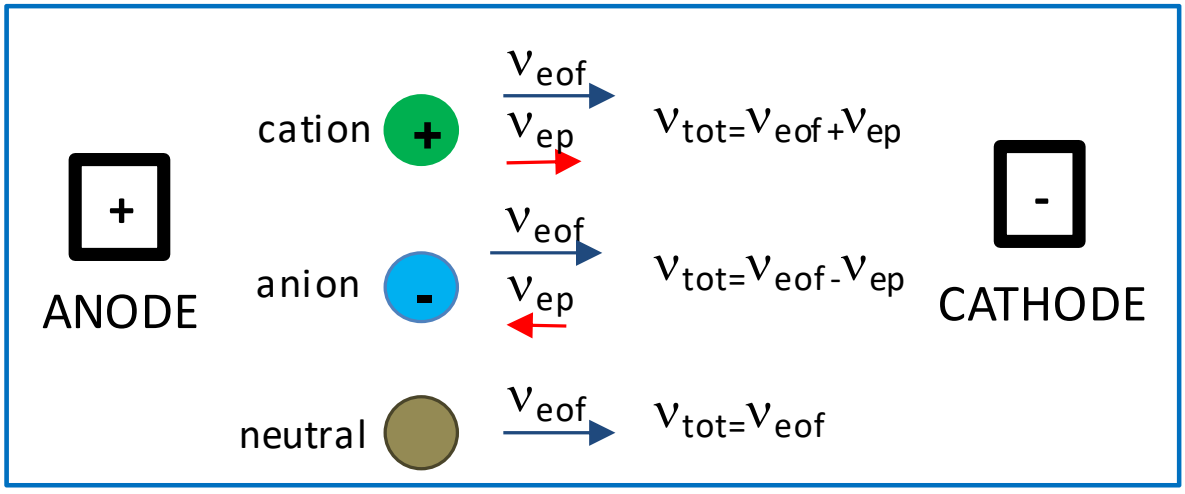


Electrophoresis and Electroosmosis

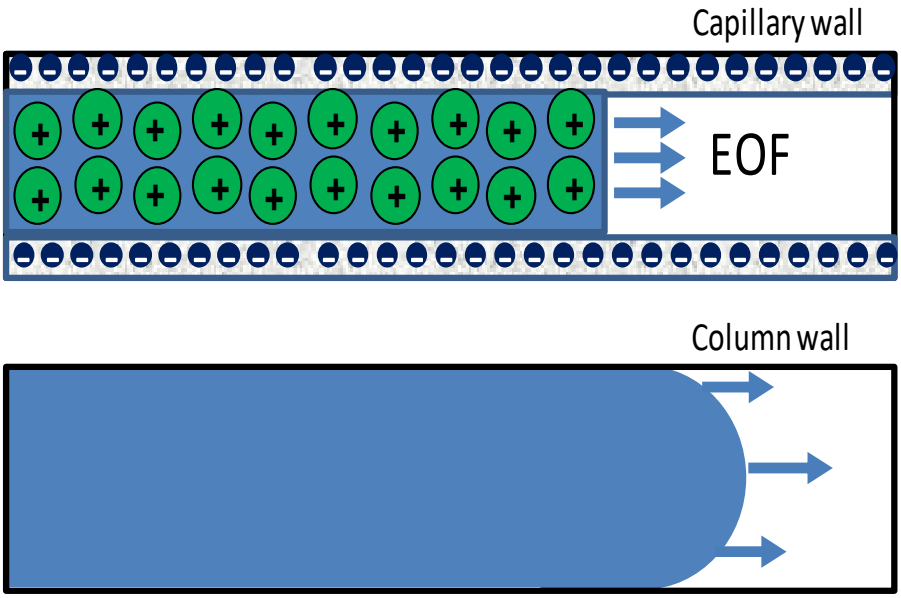
A pictorial representation of the combined effect in a capillary, when EO is faster than EP (the common case):

$$\nu = (\mu_{ep} + \mu_{eo})E = (\mu_{ep} + \mu_{eo})\frac{V}{L}$$

Figure from R. N. Zare, Stanford

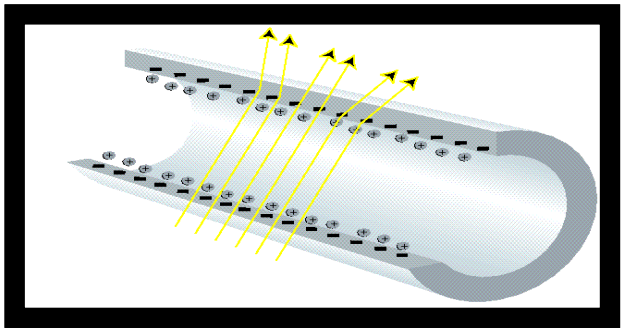


Diagrams of flow in CE and HPLC



Detection Options

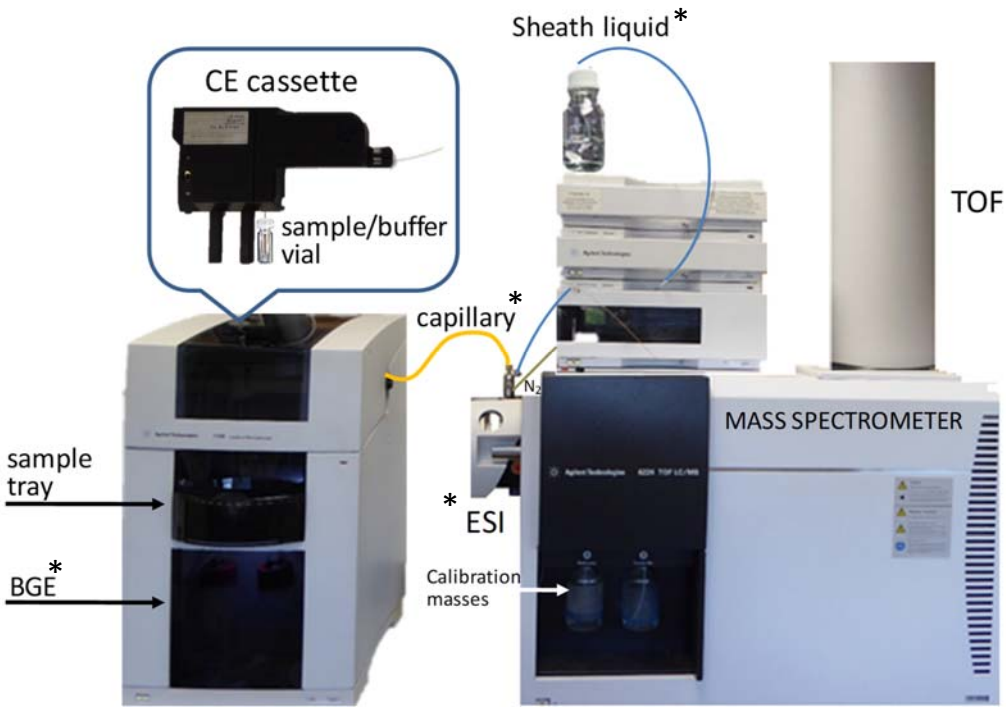
Detector	Advantages	Characteristics	Limit the range of detection (M)
UV/vis absorbance	-The possibility of direct and indirect detection -Very common detector	-Universal	-10^{-3} -10^{-6} for the detection of aromatic compounds
LIF (Laser-induced fluorescence)	-Highly sensitive and highly selective -Used for fluorescent compounds or derivatives	-Selective	-10^{-6} - 10^{-9}
MS	-Qualitative and quantitative information -Highly sensitive and highly selective	-Universal -Selective	-≈10^{-5} (it depends on the type of MS and metabolites)



Optimizing CE Separations PARAMETERS

- pH
  - First paramater to control
  - Effects EOF and mobility (charge)
- Organic Solvent
  - Analyte solvation
- Interacting agent
  - Ion-pairing, solvation, etc.
- Non-aqueous Conditions
  - Solvation and charge
- Temperature
  - Solvation, chemical equilibria

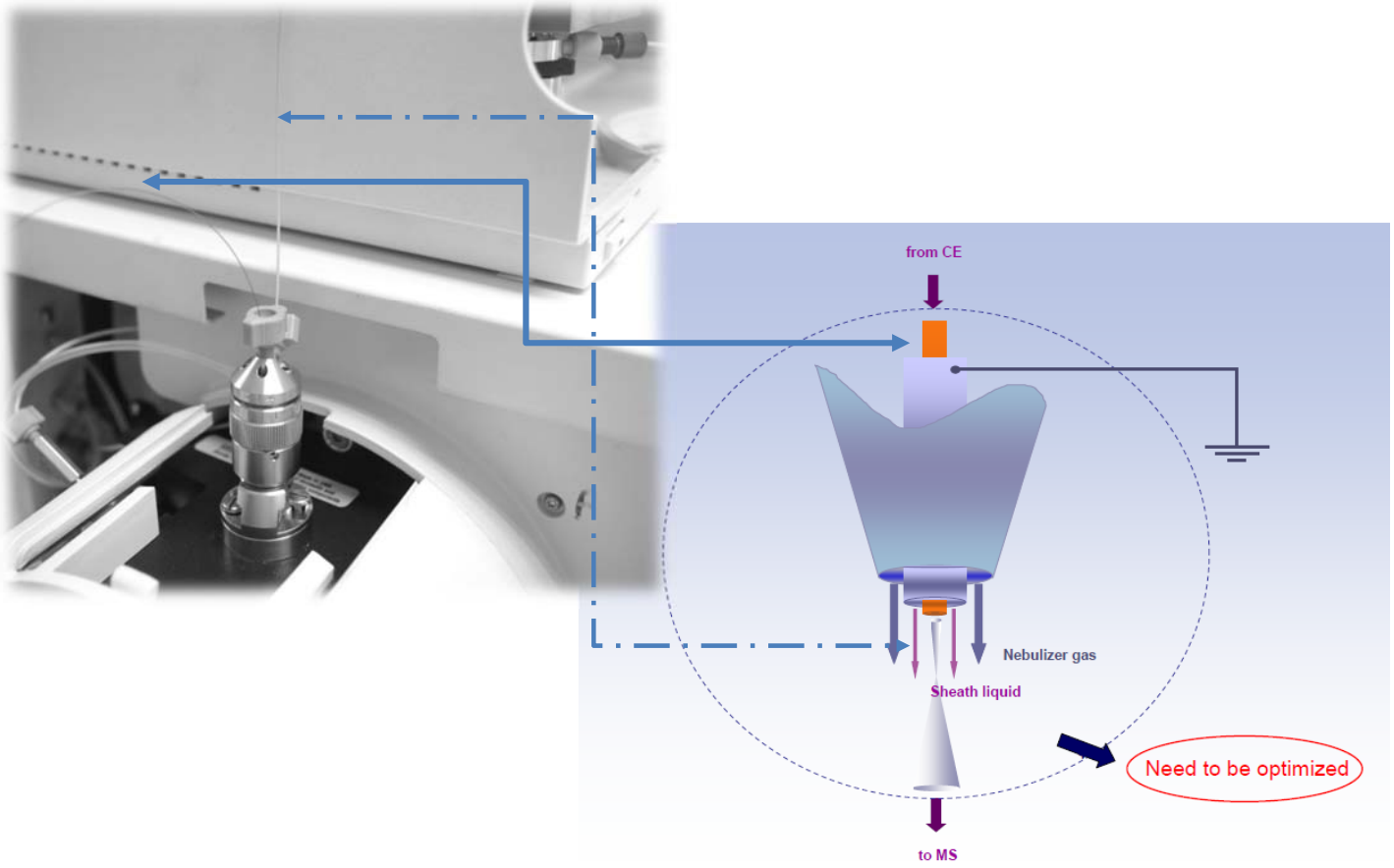
INSTRUMENTATION CE-MS



MS-TOF

* OPTIMIZATION REQUIRED

ESI SOURCE CE-MS



CE-MS: Electrical interfacing

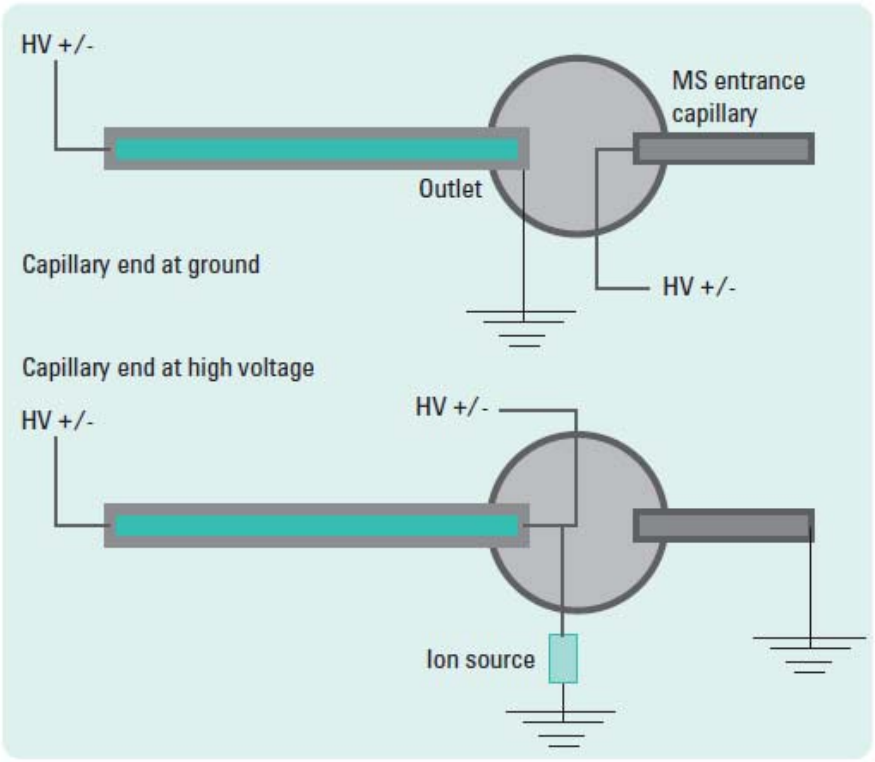


image provided by Agilent Technologies

This project has been funded with support from the European Commission.

This publication reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein

METABOLOMICS

Coral Barbas
Danuta Dudzik
Mª Fernanda Rey-Stolle
Francisco J. Rupérez
Antonia Garcia

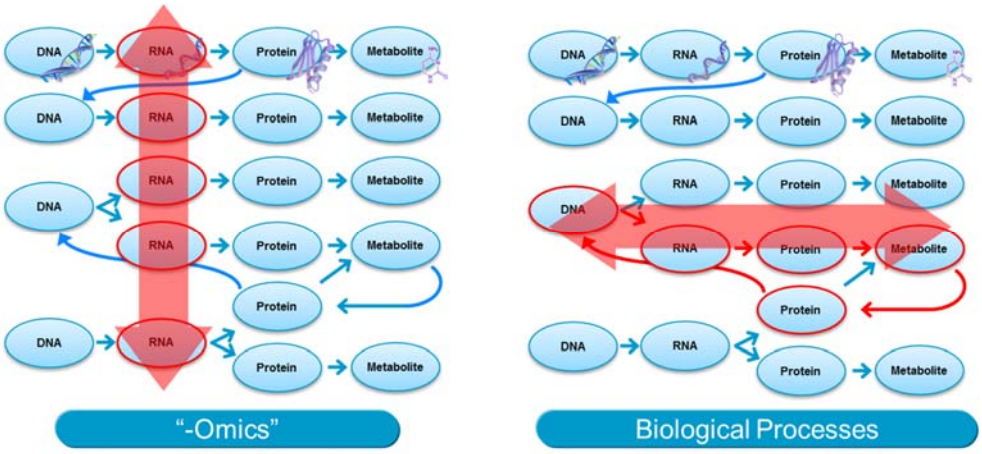
SUMMARY

1. Introduction to metabolomics
2. Analytical approaches in metabolomics
   - Workflow of the metabolomics study
   - Quality Control and Quality Assurance Procedure in Metabolomics
3. Data processing and identification of metabolites
   - Data processing pipeline
   - Non-targeted metabolomics data treatment
   - Metabolite identification
   - Statistical analysis
4. Data analysis
   - From data identification to pathways
   - Biomarker validation
5. Practical sessions
   - Targeted and non-targeted metabolomics
   - Metabolomics with free online tools

Metabolomics

New emerging field of "omics" research (which includes genomics, proteomics and metabolomics) concerned with comprehensive characterization of the small molecule metabolites present in biological systems.

Omics & Systems Biology

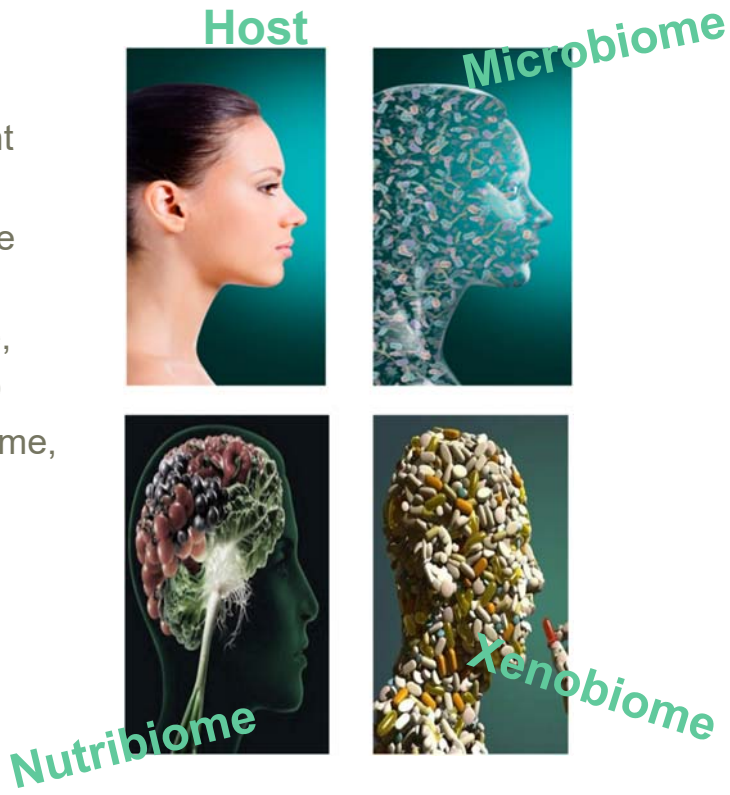


Addendum: Definition of Metabonomics

- Measurement of the dynamic multiparametric metabolic response of living systems to pathophysiological stimuli or genetic modification (Nicholson, 1999)
  - quantitative measurement of the time-related "total" metabolic response to pathophysiological (nutritional, xenobiotic, surgical or toxic) stimuli
- MetaboLomics - the picture, MetaboNomics – the movie
- Nowadays, everything is Metabolomics

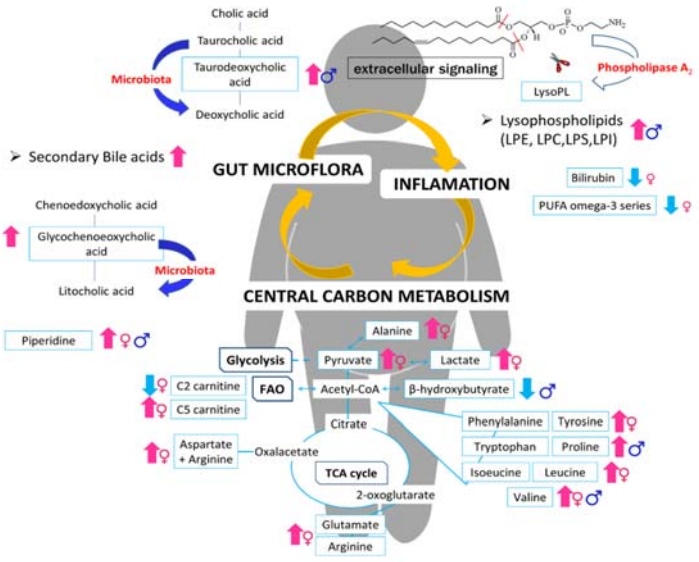
Definition of Metabolome

- "...the **complete set of metabolites**/low-molecular-weight intermediates, which are context dependent, varying according to the physiology, developmental or pathological state of the cell, tissue, organ or organism…" (Oliver 2002)
- Origin: Endometabolome, Microbiome, Xenobiome, Nutribiome…
- Nature: Glycome, lipidome, sphingolipidome, peptidome…
- Metabolome ↔ Phenotype



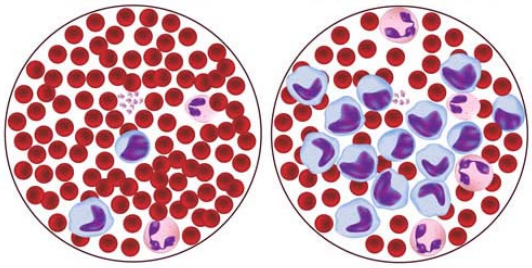
What metabolomics can provide (I)

- Overview of the metabolic status and global biochemical events associated with a cellular or biological system.
  - Pathological situations without known mechanism, i.e. relationship between obesity and insulin resistance



What metabolomics can provide (II)

- Identification (proposal) of new biomarkers, important in the process of new drug discovery or as in vitro diagnostics tools.
  - For instance, new diagnostic biomarkers for aggressiveness in chronic lymphatic leukemia



Utility of validated metabolites as biomarkers of aggressive state of CLL

Metabolite	AUC	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
Acetylcarnitine	0.695	43.2	93.0	86.1	62.1
Butyrylcarnitine	0.548	10.8	98.0	84.4	52.4
Hexanoylcarnitine	0.690	27.0	96.0	87.1	56.8
Octanoylcarnitine	0.651	29.7	95.0	85.6	57.5
Decanoylcarnitine	0.662	27.0	94.0	81.8	56.3
Palmitoylcarnitine	0.719	40.5	94.0	87.1	61.2
Dodecanamide	0.497	8.1	100.0	100.0	52.1
Hexadecanamide	0.516	5.4	100.0	100.0	51.4
Oleamide	0.600	18.9	96.0	82.5	54.2
Linoleamide	0.672	16.2	98.0	89.0	53.9
Acylcarnitines[a]	0.743	32.4	95.0	86.6	58.4
FAA[b]	0.662	13.9	96.0	77.6	52.7
Acylcarnitines and FAA	0.750	54.0	89.0	83.1	65.9

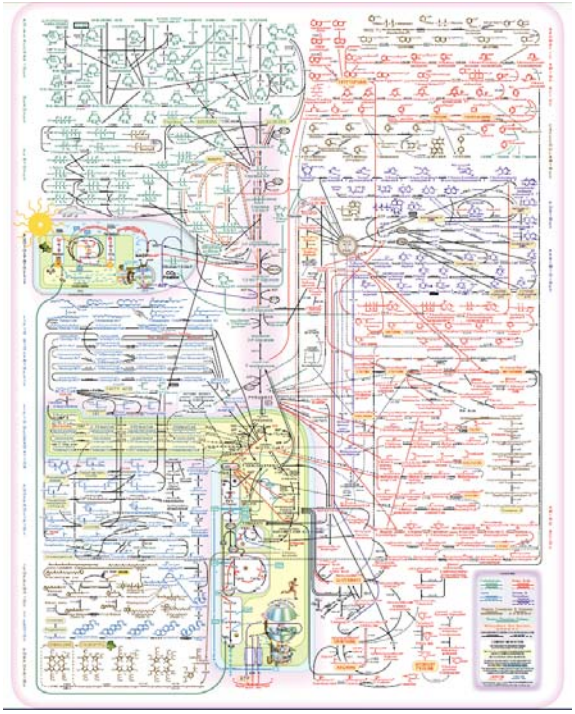
What metabolomics good for…..

- searching for metabolic differences between groups of samples (case vs control; before vs after treatment; One condition vs another)
- identifying compounds that are significant and proposing the mechanisms
- finding out information about the phenotype
- observing the effects of a treatment
- finding new drug targets

What is metabolomics NOT….

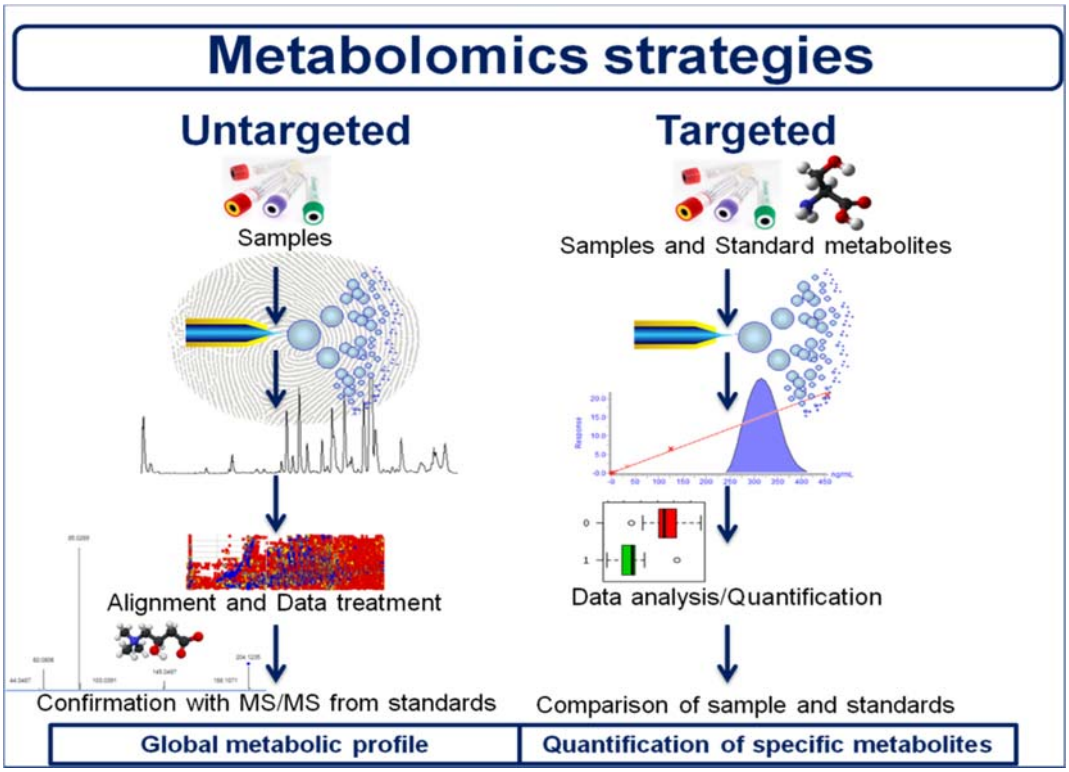
- a method to reveal the fate of a metabolite or drug
- a method for quantification
- the use of a simple kit to quantify a grouop of metabolites (it requires NMR, MS…)
- Possible without simultaneous comparison of samples

Definition of Metabolism

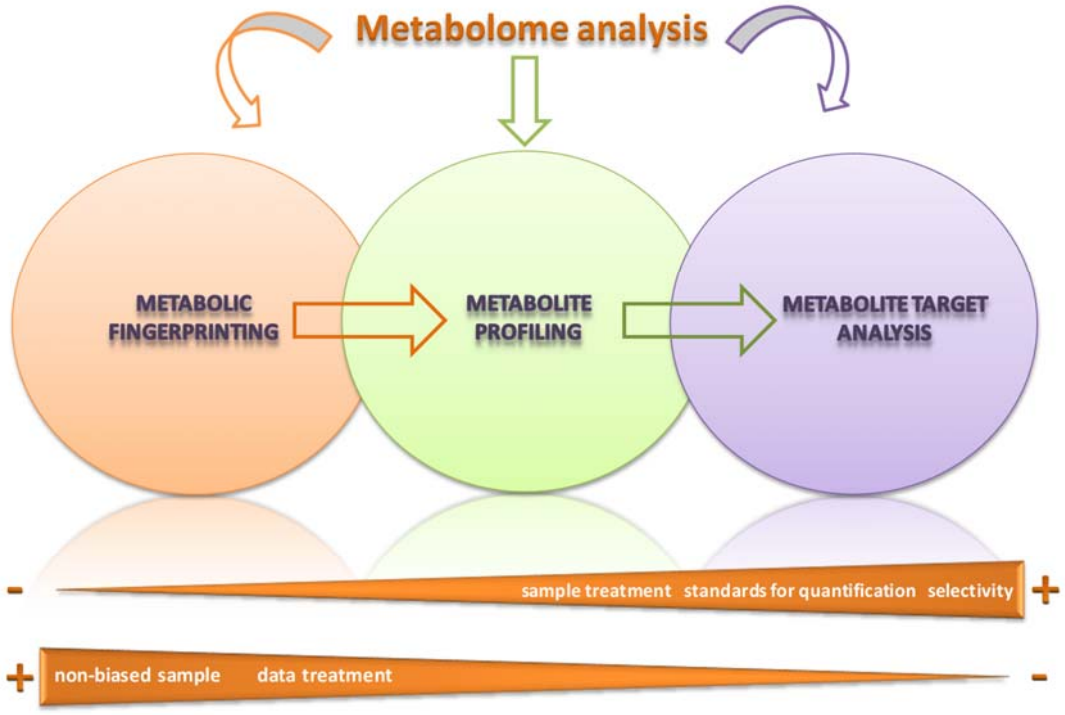


The complete group of (bio)chemical processes within an organelle, cell, tissue, organ or organism, essential for life

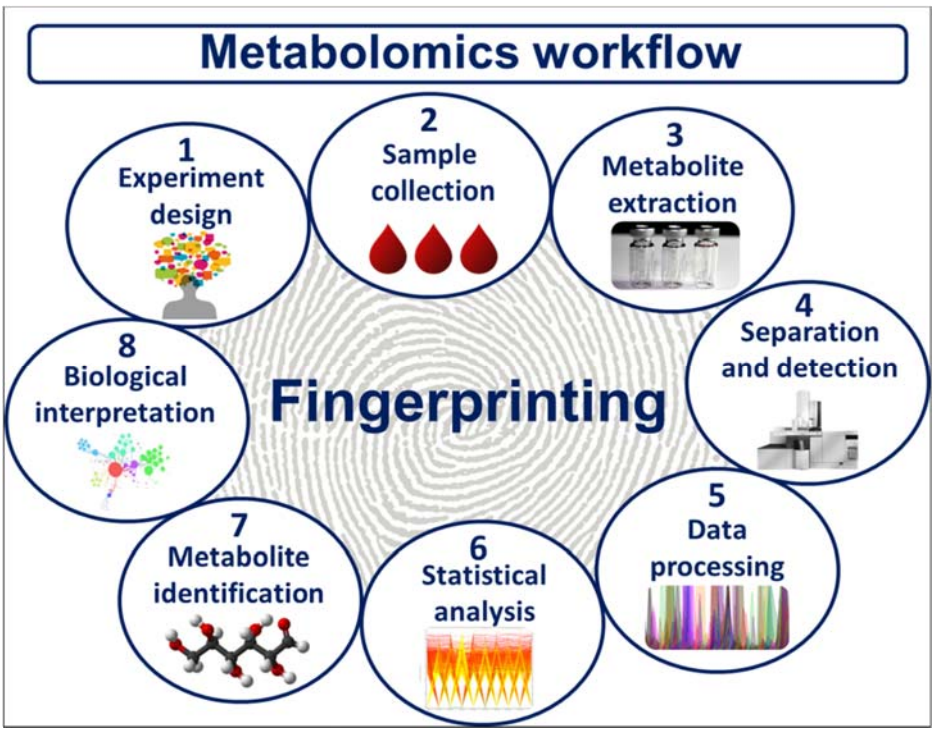
Analytical approaches in metabolomics



Three ways to do metabolomics



WORKFLOW



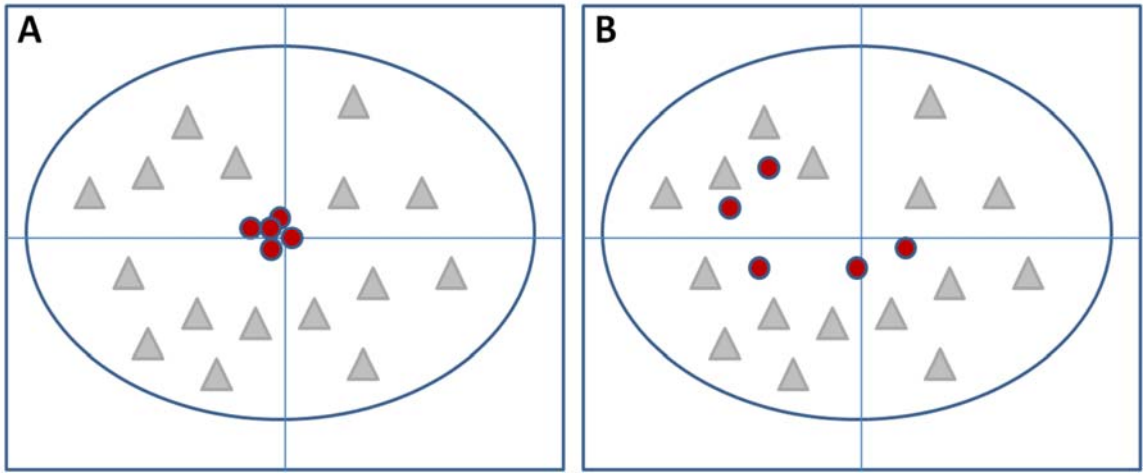
ANALYTICAL TECHNIQUES

- GC/MS: Small polar compounds
  - Mainly water soluble (some hydrophobic)
  - Sample treatment: Derivatization
  - Fragmentation reproducible - databases
- NMR
  - Water-soluble
  - Virtually no sample treatment
  - High LOD
- LC/MS
  - from small to large (<1500 Da) medium to non-polar metabolites
- CE/MS: Small-medium polar compounds
  - Amino acids, acylcarnitines, polyamines, etc.
  - No derivatization

MS Based analytical platforms in metabolomics

Analysis Technique	Application	Advantages	Disadvantages
GC-MS	Separation, identification, and quantification of volatile and thermally stable less polar metabolites.	High chromatographic resolution, availability of large spectrum libraries for metabolites identification.	Inability to analyze thermo-labile and high molecular weight metabolites, the requirement of derivatization for non-volatile metabolites.
LC-MS	Separation, identification, and quantification of very broad groups of metabolites, depending on the type of column and mobile phase.	High sensitivity, large sample capacity, derivatization not required, ability to analyze thermo-labile compounds.	Limited availability of commercial libraries, restriction on LC eluents, matrix effect, limited potential in identification unless an MS-MS technique is used.
CE-MS	Separation, identification, and quantification of polar and ionized metabolites, using reduced sample volumes.	High resolution and rapid analysis, utility for complex biological samples, even if in a small volume.	Limited availability of commercial libraries. Buffer incompatibility, detection limits. Limited potential for identification unless an MS-MS technique is used.

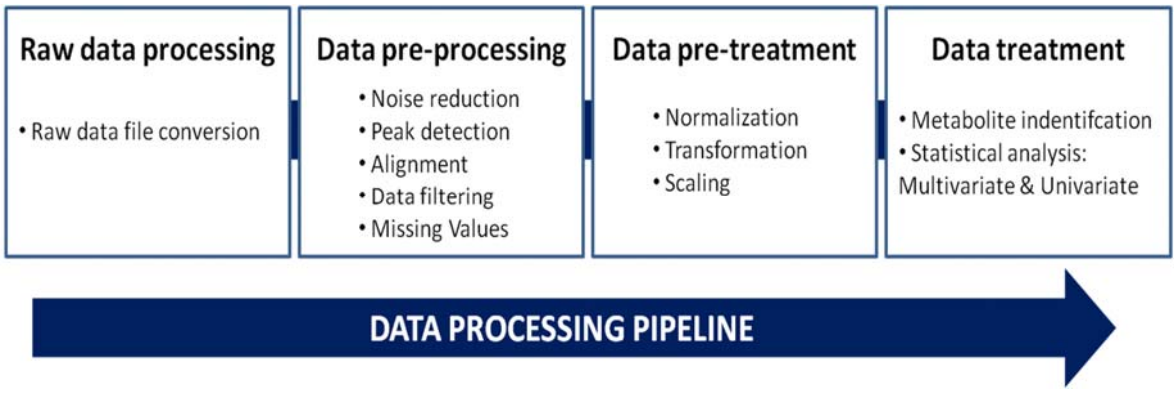
Quality Control and Quality Assurance Procedure in Metabolomics



A: QCs (red dots) clustered together

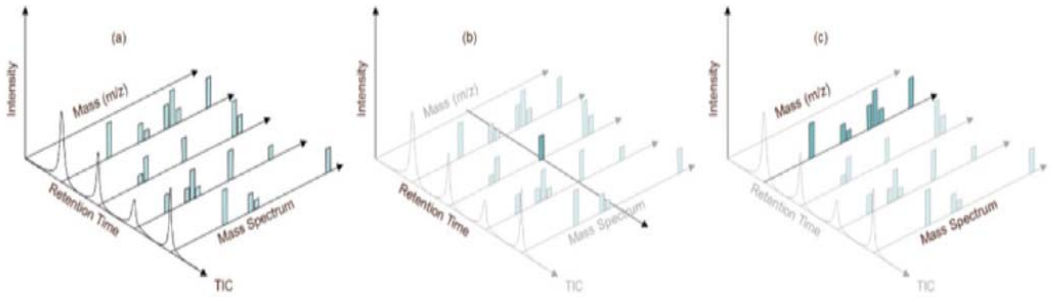
B: QCs spreaded

DATA TREATMENT IN METABOLOMICS: Signal Processing



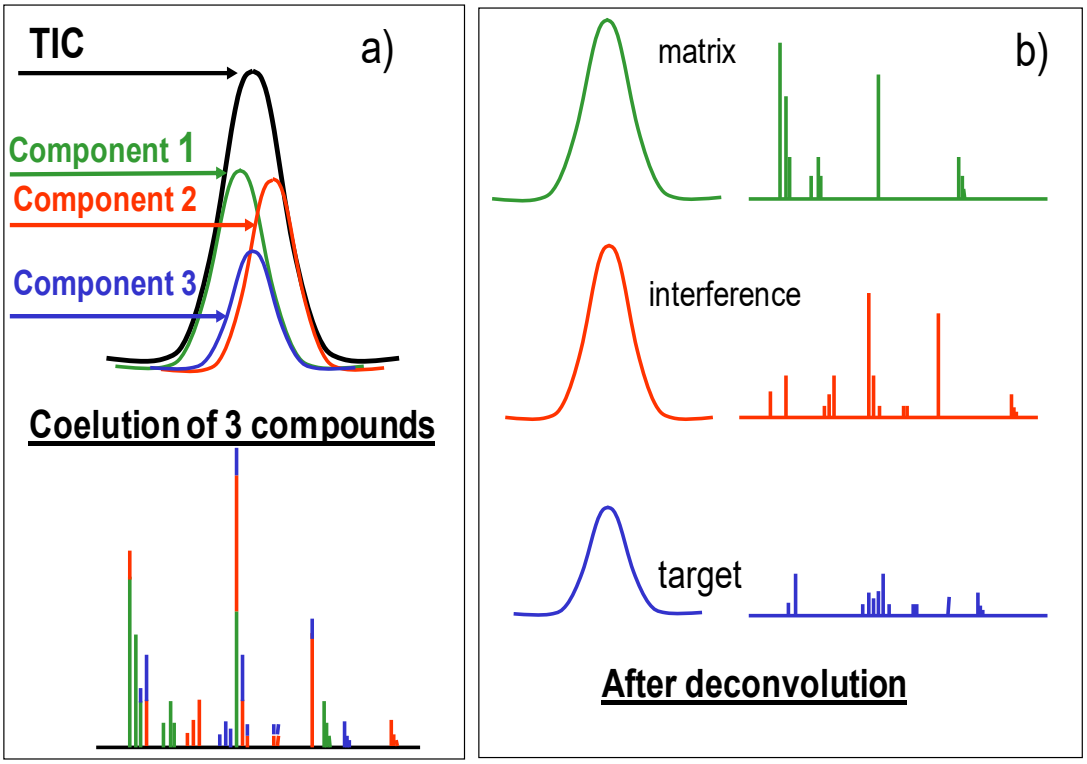
ANALYTICAL TECHNIQUE: GC-MS

- Gas chromatography coupled to mass spectrometry
- Gold standard
  - Highly sensitive and reproducible
  - Information: Quality and Quantity
  - Spectrum libraries for identification purposes
  - 10-20% of the known compounds can be analyzed by GC
    - High metabolic relevance



(a) 3D Data of GC/MS, (b) Extracted Ion chromatogram for the selected ion
(c) A single data point in time gives a single mass spectrum
adapted from Chromatography today

Deconvolution

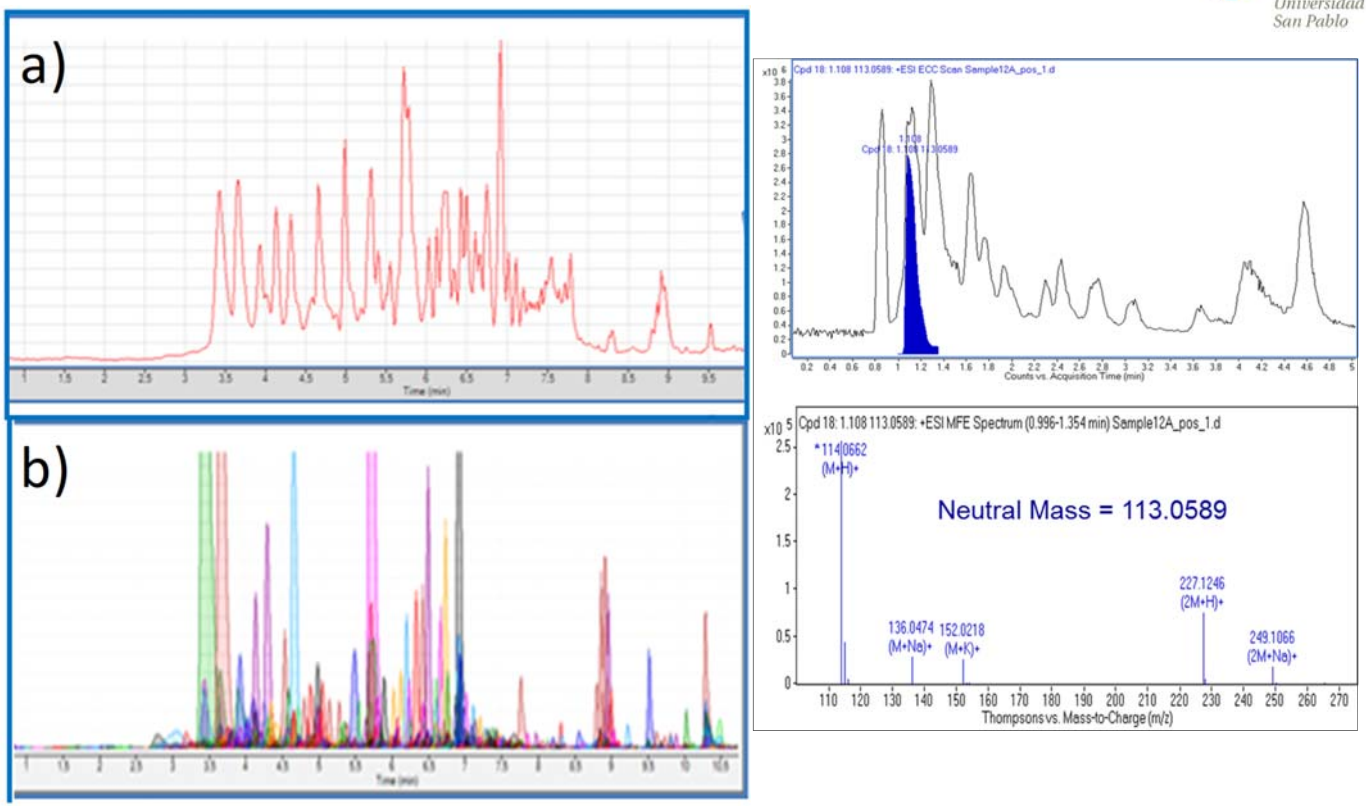


a) Before and b) After the deconvolution process
adapted from https://www.agilent.com/cs/library/Support/Documents/f05017.pdf

Deconvolution in LC-ESI-MS and CE-ESI-MS

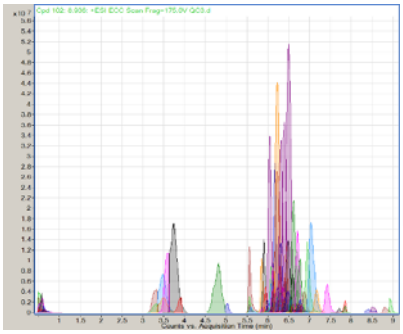
- Peak-based methods
- Molecular Feature Extractor (Agilent) considers the accuracy of the mass measurements to group related ions by charge-state envelope, isotopic distribution, and possible chemical relationships when determining whether different ions are from the same metabolic feature.
- It can consider also related ions like adducts: proton, sodium, potassium and ammonia adducts in positive ionization or loss of a proton, adducts with formate, etc. in negative ionization mode.

After Deconvolution



a) Total Ion Chromatogram
b) Chromatograms from every single compound obtained after deconvolution

Chromatogram or features list?



Automatically Show Columns

Show/Hide	Saturated	RT	m/z	Mass	Polarity	Ions	Height	Area	Vol	Quality Score
☑	S	0.514	280.0923	279.085	Positive	5	2740215		14234373	100
☑	S	0.517	203.0526	202.0454	Positive	6	4045612		20286412	80
☑		0.526	140.0682	139.0609	Positive	3	1530642		7519633	100
☑		0.529	136.0482	135.041	Positive	2	1187260		6016909	100
☑	S	0.57	162.1126	161.1053	Positive	7	2758926		19465836	100
☑	S	0.57	304.2998	303.2926	Positive	4	3021606		14360408	100
☑		0.58	114.0664	113.0591	Positive	4	549599		6985240	80
☑		0.614	175.1192	174.1119	Positive	3	760396		3860452	91.6
☑		0.625	156.0768	155.0696	Positive	7	1055485		6005360	100
☑		0.646	170.0927	169.0854	Positive	4	604901		3355886	100
☑	S	3.29	520.3389	519.3316	Positive	7	3243984		36560884	80.2
☑		3.298	544.3388	543.3315	Positive	4	1339590		12508261	87
☑	S	3.483	520.3389	519.3316	Positive	11	5382675		86088232	85.2
☑	S	3.5	544.3389	543.3316	Positive	7	2216628		36501788	100
☑	S	3.573	496.3389	495.3316	Positive	6	8281163		80624664	86.1
☑	S	3.746	496.3388	495.3316	Positive	13	11742118		200075408	80
☑	S	3.89	522.3545	521.3472	Positive	8	2204035		17763280	87
☑	S	4.801	524.3702	523.3629	Positive	12	6862920		111854136	100
☑		5.027	524.3702	523.3629	Positive	6	1308318		8748089	87
☑		5.559	163.0393	162.032	Positive	4	1349239		5379156	100
☑	S	5.561	391.2839	390.2767	Positive	15	7367381		51796644	100
☑	S	5.561	149.0235	148.0162	Positive	4	2915086		12886771	100
☑		5.597	338.3418	337.3345	Positive	9	787867		4058574	100
☑	S	5.849	675.5425	674.5353	Positive	7	6821314		39496860	100
☑		5.857	627.5339	626.5266	Positive	3	712763		3673685	100
☑	S	5.904	701.5583	700.551	Positive	8	8615328		59468044	100
☑	S	5.941	689.558	688.5507	Positive	8	3178377		18867332	100
☑		5.991	754.5372	753.5299	Positive	4	737180		4127601	100
☑		5.995	730.5369	729.5297	Positive	7	1506513		8484420	100

Chromatogram Results | MS Spectrum

Data preprocesing

- Alignment
  - Peak shifts are observed across the RT axis
  - Two groups:
    - data are aligned before peak detection
    - peak-based alignment methods: detected spectral peaks are aligned across samples.
    - softwares:
      - MetaboAnalyst (metaboanalyst.ca)
      - mzmine and mzmine2 (http://mzmine.sourceforge.net/)
      - metAlign
      - BinBase (fiehnlab.ucdavis.edu)
      - xcms and xcms2 (Scripps)
      - metaXCMS (Scripps)
      - XCMS Online (Scripps)
- Missing values
  - Problems in further analysis
  - Different strategies
    - Replace by the half of the minimum, by mean/median, k-nearest neighbour (KNN), probabilistic PCA (PPCA), Bayesian PCA (BPCA) method, Singular Value Decomposition (SVD) …
- Filtering
  - Variables of very small values - detected using mean or median
  - Variables that are near-constant - detected using standard deviation (SD)
  - Variables that show low repeatability - measured using QC sample

Data pretreatment

- Normalization
  - **Sample-specific** normalization (i.e. weight, volume)
  - Normalization by **sum** or **median**
  - Normalization by reference **sample**
  - Normalization by a **pooled** sample from group control
  - Normalization by reference **feature**
  - **Quantile** normalization
- Data transformation
  - **Log** transformation
  - **Cube root** transformation
- Data scaling
  - **Mean** centering
  - **Auto scaling** (mean-centered and divided by the standard deviation of each variable)
  - **Pareto scaling** (mean-centered and divided by the square root of standard deviation of each variable)
  - **Range scaling** (mean-centered and divided by the range of each variable)

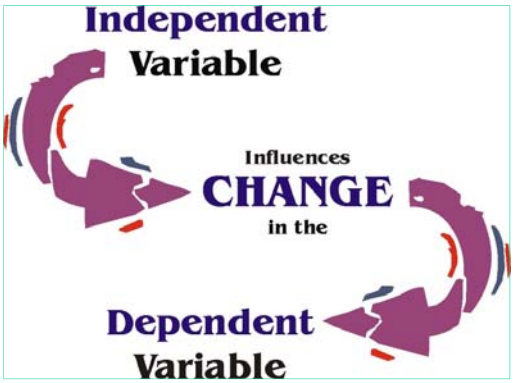
Statistics for Metabolomics

AIMS to:

- detect differences between sample groups at the chemical level
- rank compounds by relative importance for sample differentiation

VARIABLES

- **dependent variable**: represents the output or effect, or is tested to see if there is effect, e.g.: abundance of metabolite
- **independent variable**: represents the inputs or causes, or are tested to see if they are the causes, e.g.: treatment conditions within the experiment

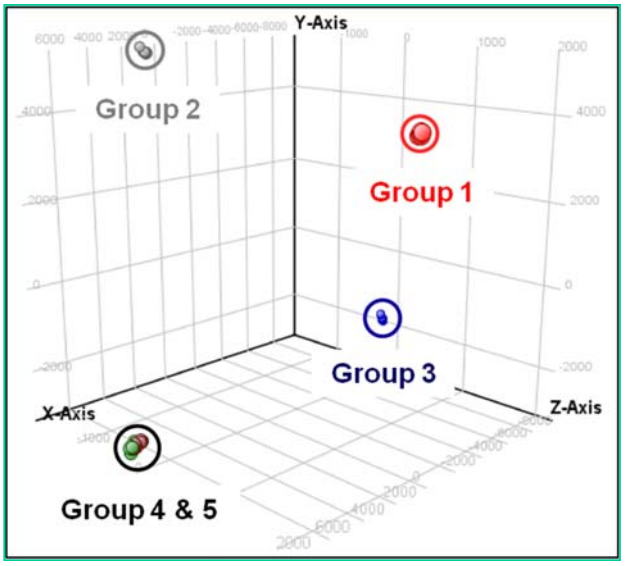


TYPES

- **Univariate analysis UVA:**
  - Normal distribution: Student's *t*-Test, ANOVA,
  - Non-normal distribution: Mann-Whitney U-Test, Kruskal-Wallis
- **Multivariate analysis MVA:** PCA, PLS-DA, OPLSDA

PCA

- used as a tool in exploratory data analysis
- each dot graphically represents each sample measured
- the algorithm has no knowledge of the group associations of the samples – *unsupervised* analysis
- first principal component explains most of the variance
- compound loadings indicate the impact of that compound on the analysis
- each dot is the sum of the compound loadings for a sample
- the tightness of the clustering reflects the variance of the samples



Class prediction

an algorithm using past data to predict the results of future observations

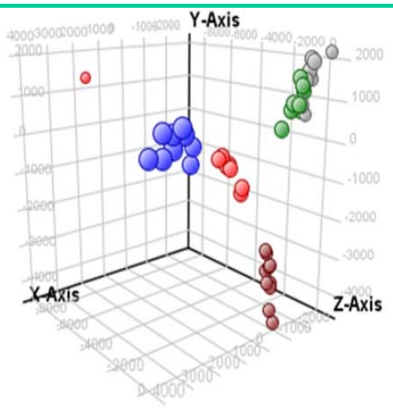
- the algorithm has knowledge of the group associations of the samples – *supervised* analysis
- common algorithms
  - **Partial Least Squares Discriminate Analysis (PLS-DA)**
  - Support Vector Machine
  - Decision Tree
  - Naïve Bayes
  - Neural Network

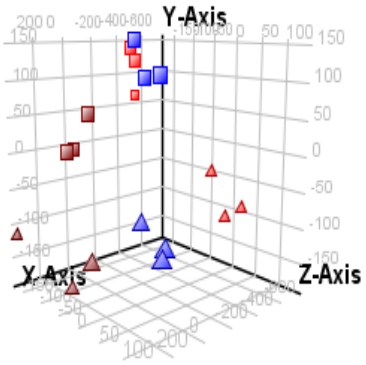
Class prediction: PLS-DA

Partial Least Square - Discriminant Analysis
Projection on Latent Structures - Discriminant Analysis

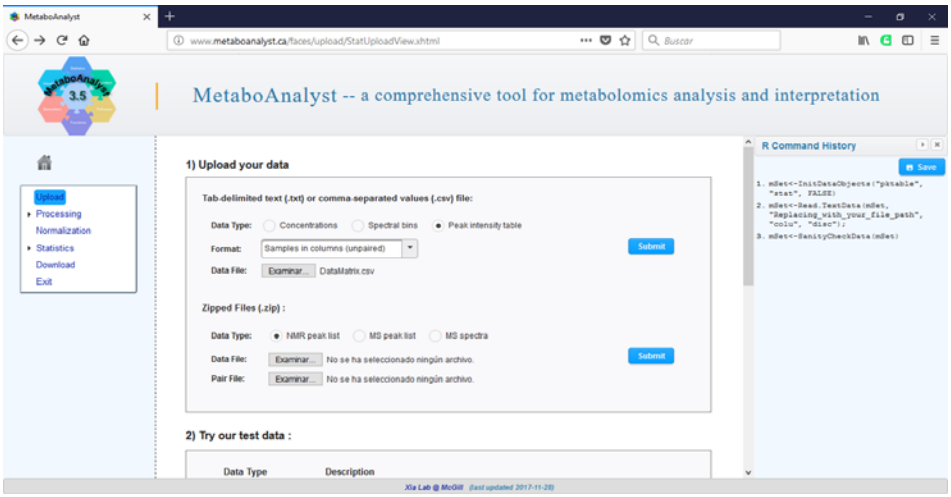
a statistical method that bears some relation to principal components analysis (PCA) but is a *supervised* analysis

- creates a linear regression model by projecting the predicted and observable variables to a new space
- well suited when there are more predictors (compounds) than observations (samples)
- each compound has a t-score that represents its impact on the prediction
- a prediction confidence value is assigned when the model is run





Univariate and Multivariate Statistical Analysis



Class prediction: Validate the model

assesses accuracy of prediction rule that is built and provides an indication of over-fitting models:

leave one out

- all samples in the training set except one is used to build the prediction rule
- using this rule, the class of sample that was left out is predicted
- the sample is returned to the training set while a different sample is left out and the prediction rule is built with remaining samples
- this process is repeated until each sample in training set has been predicted exactly once
- the number of correct and incorrect predictions is then tallied to determine the success rate

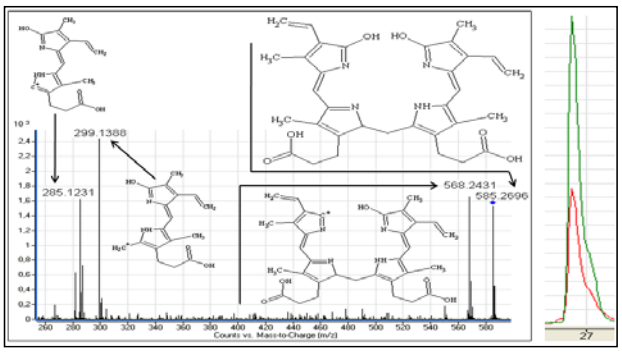
N - fold

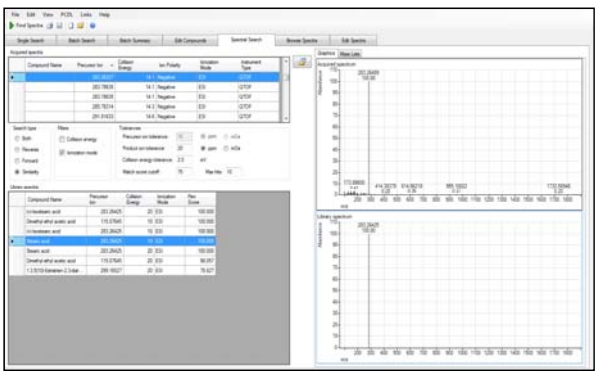
1. samples in the training set are randomly divided into *N* equals subsets, maintaining relative classes frequency
2. *N*-1 subsets are then combined for training and the remaining set is used for testing
3. repeat step 2 step with each group left out in turn
4. repeat step 1, 2, 3 *M* times
5. each sample gets predicted *M* times and majority class predicted over these *M* times is reported in validation results

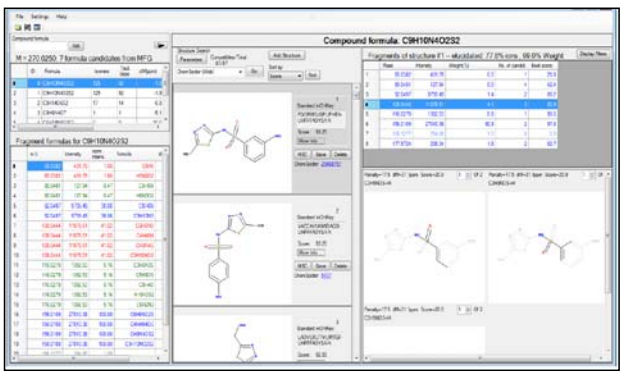
Identification

1. Database matching using accurate mass measurement
2. Database matching with isotope pattern matching
3. Database matching with isotope pattern matching and retention time
4. MS/MS library matching
5. MS/MS library and retention time matching

Confidence







DATABASE CLASIFICATIONS

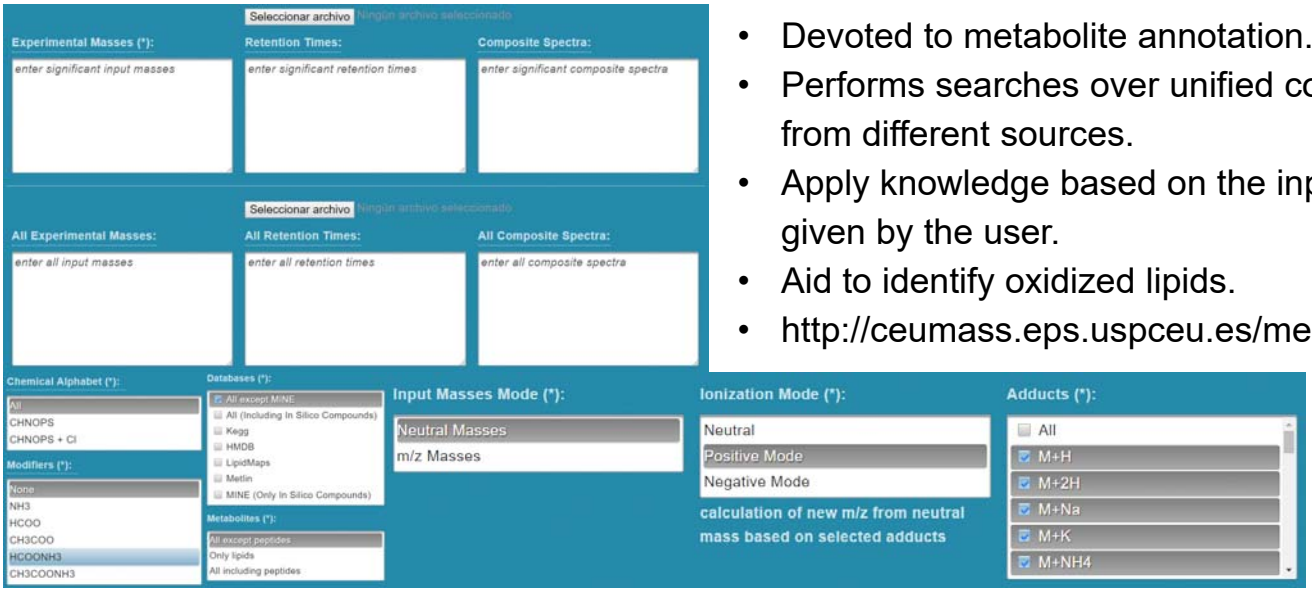
- **Based on Spectral input**
  - Mainly small molecules and not only metabolites
  - NMR
  - MS or MS/MS
- **Based on compound information**
  - Compound name, structures, physical properties, identification
- **Based on Metabolic pathway database**
  - Metabolites, xenobiotics, proteins, signal pathways
- **Complete Metabolomic database**
  - A combination of the previous ones

Database List in 2018

Name	URL	Name	URL
ARALIP	http://aralip.plantbiology.msu.edu/pathways/pathways	KEGG	http://prime.psc.riken.jp/?action=metabolites_index
AtIPD	http://www.atipd.ethz.ch/	KEGG Glycan	http://www.genome.jp/kegg/glycan/
BiGG	http://bigg.ucsd.edu/	KNApSAcK	http://prime.psc.riken.jp/?action=metabolites_index
BioCyc	http://biocyc.org/	LipidMaps	http://www.lipidmaps.org/
BioNumbers	http://bionumbers.hms.harvard.edu/	MarkerDB	http://www.markerdb.ca/users/sign_in
BML-NMR	http://www.bml-nmr.org/	MassBank	http://www.massbank.jp/
BioMagResBank	http://www.bmrb.wisc.edu/metabolomics/	MetaboAnalyst	http://www.metaboanalyst.ca/MetaboAnalyst/
BMDB	http://www.cowmetdb.ca/cgi-bin/browse.cgi	MetaboLights	http://www.ebi.ac.uk/metabolights/index
ChEBI	http://www.ebi.ac.uk/chebi/	MetaCrop	http://metacrop.ipk-gatersleben.de/apex/f?p=269:111:
ChEMBL	https://www.ebi.ac.uk/chembl/about#	MetaCyc	http://metacyc.org/
ChEBI	http://www.ebi.ac.uk/chebi/	METAGENE	http://www.metagene.de/program/a.prg
ChemMine	http://chemminedb.ucr.edu/	METLIN	https://metlin.scripps.edu/index.php
ChemSpider	http://www.chemspider.com/	MMCD	http://mmcd.nmrfam.wisc.edu/
CCD	http://ccd.chemnetbase.com/intro/index.jsp#about	mzCloud	https://mzcloud.org/
CSF Metabolome Database	http://www.csfmetabolome.ca/	OMIM	http://www.ncbi.nlm.nih.gov/omim/
CyberCell Database	http://ccdb.wishartlab.com/CCDB/	OMMBID	http://ommbid.mhmedical.com/
DrugBank	http://www.drugbank.ca/	Oryzabase	http://www.shigen.nig.ac.jp/rice/oryzabase/
ECMDB	http://www.ecmdb.ca/	PepBank	http://pepbank.mgh.harvard.edu/
ExPaSy Pathways	http://web.expasy.org/pathways/	PharmGKB	http://www.pharmgkb.org/
Fiehn GC-MS Database	http://fiehnlab.ucdavis.edu/Metabolite-Library-2007/	PMN	http://www.plantcyc.org/
FooDB	http://www.foodb.ca	PubChem	http://pubchem.ncbi.nlm.nih.gov/
GMDB	http://gmd.mpimp-golm.mpg.de/	Reactome	http://www.reactome.org/
HMDB	http://metabolomics.pharm.uconn.edu/iimdb/	RiceCyc	http://pathway.gramene.org/gramene/ricecyc.shtml
HumanCyc	http://www.genome.jp/kegg/	Serum Metabolome Database	http://www.serummetabolome.ca/
IIDMB	http://www.genome.jp/kegg/glycan/	SetupX & BinBase	http://fiehnlab.ucdavis.edu/projects/binbase_setupx

AACLifeSci

CEU Universidad San Pablo

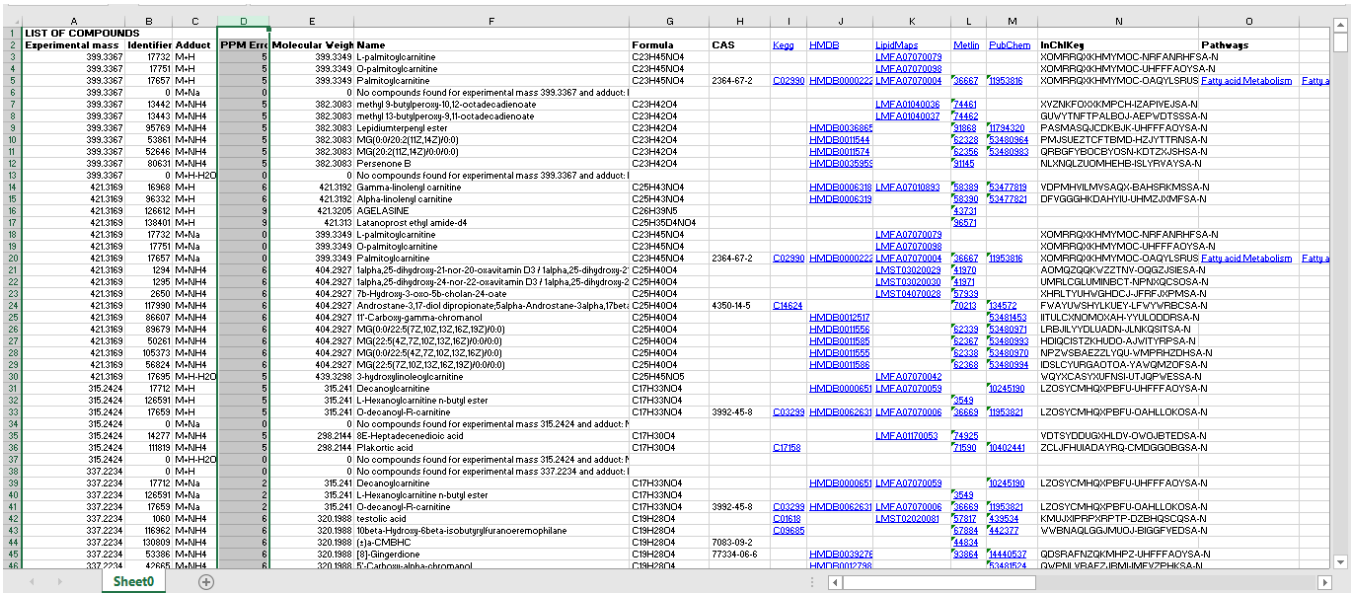


- Devoted to metabolite annotation.
- Performs searches over unified compounds from different sources.
- Apply knowledge based on the input data given by the user.
- Aid to identify oxidized lipids.
- http://ceumass.eps.uspceu.es/mediator

AACLifeSci

CEU Universidad San Pablo

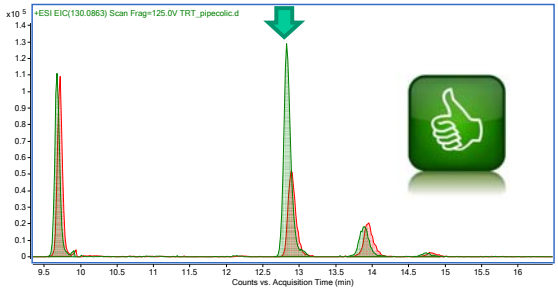
CEU
MASS MEDIATOR



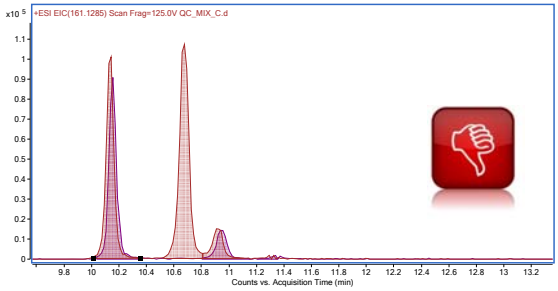
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	LIST OF COMPOUNDS														
2	Experimental mass	Identifier	Adduct	PPM Erro	Molecular Weigh	Name	Formula	CAS	Kegg	HMDB	LipidMaps	Metlin	PubChem	InChIKey	Pathways
3	399.3367	17732	M+H	5	399.3349	L-palmitoylcarnitine	C23H45NO4				LMFA07070079			XOMRRQXKHMYMOC-NRFANRHFSA-N	
4	399.3367	17751	M+H	5	399.3349	O-palmitoylcarnitine	C23H45NO4				LMFA07070098			XOMRRQXKHMYMOC-UHFFFAOYSA-N	
5	399.3367	17657	M+H	5	399.3349	Palmitoylcarnitine	C23H45NO4	2364-67-2	C02990	HMDB0000222	LMFA07070004	36667	11953816	XOMRRQXKHMYMOC-OAQYLSRUS	Fatty acid Metabolism
6	399.3367	0	M+Na	0	0	No compounds found for eseprimental mass 399.3367 and adduct:									
7	399.3367	13442	M+NH4	5	382.3083	methyl 9-butylperoxy-10,12-octadecadienoate	C23H42O4				LMFA01040036	74461		XVZNKFOXXKMPCH-IZAPIVEJSA-N	
8	399.3367	13443	M+NH4	5	382.3083	methyl 13-butylperoxy-9,11-octadecadienoate	C23H42O4				LMFA01040037	74462		GUVYTNFTPALBOJ-AEPVDTSSSA-N	
9	399.3367	95769	M+NH4	5	382.3083	Lepidiumterpengl ester	C23H42O4			HMDB0036865		91868	11794320	PASMASQJCDKBJK-UHFFFAOYSA-N	
10	399.3367	53861	M+NH4	5	382.3083	MG(0:0/20:2(11Z,14Z)/0:0)	C23H42O4			HMDB0011544		52328	53480964	PMJSUEZTCFTBMD-HZJYTTRNSA-N	
11	399.3367	52646	M+NH4	5	382.3083	MG(20:2(11Z,14Z)/0:0/0:0)	C23H42O4			HMDB0011574		52356	53480863	QRBSIFYBOCBYQSN-KDTZXJSHSA-N	
12	399.3367	80631	M+NH4	5	382.3083	Persenone B	C23H42O4			HMDB0025955		91145		NLXNQLZUOMHEHB-ISLYRYAYSA-N	
13	399.3367	0	M+H-H2O	0	0	No compounds found for eseprimental mass 399.3367 and adduct:									
14	421.3169	16968	M+H	6	421.3192	Gamma-linolenyl carnitine	C25H43NO4			HMDB0006318	LMFA07010893	68389	53477819	VDPMHVILMVSAQX-BAHSRKMSSA-N	
15	421.3169	96332	M+H	6	421.3192	Alpha-linolenyl carnitine	C25H43NO4			HMDB0006319		68390	53477821	DFVGGGHKDAHYIU-UHMZJXMFSA-N	
16	421.3169	126612	M+H	9	421.3205	AGELASINE	C26H39N5					43721			
17	421.3169	138401	M+H	9	421.313	Latanoprost ethyl amide-d4	C25H35D4NO4					96571			
18	421.3169	17732	M+Na	0	399.3349	L-palmitoylcarnitine	C23H45NO4				LMFA07070079			XOMRRQXKHMYMOC-NRFANRHFSA-N	
19	421.3169	17751	M+Na	0	399.3349	O-palmitoylcarnitine	C23H45NO4				LMFA07070098			XOMRRQXKHMYMOC-UHFFFAOYSA-N	
20	421.3169	17657	M+Na	0	399.3349	Palmitoylcarnitine	C23H45NO4	2364-67-2	C02990	HMDB0000222	LMFA07070004	36667	11953816	XOMRRQXKHMYMOC-OAQYLSRUS	Fatty acid Metabolism
21	421.3169	1294	M+NH4	6	404.2927	1alpha,25-dihydroxy-21-nor-20-oxavitamin D3 / 1alpha,25-dihydroxy-2	C25H40O4				LMST03020029	41970		AOMQZQQKVZZTNV-OQGZJSIESA-N	
22	421.3169	1295	M+NH4	6	404.2927	1alpha,25-dihydroxy-24-nor-22-oxavitamin D3 / 1alpha,25-dihydroxy-2	C25H40O4				LMST03020030	41971		UMRLCGLUMINBCT-NPNXQCSOSA-N	
23	421.3169	2650	M+NH4	6	404.2927	7b-Hydroxy-3-oxo-5b-cholan-24-oate	C25H40O4				LMST04070028	57939		XHRLTYUHVGHDCJ-JFRFJXPMSA-N	
24	421.3169	117990	M+NH4	6	404.2927	Androstane-3,17-diol dipropionate;5alpha-Androstane-3alpha,17beta	C25H40O4	4350-14-5	C14624			70213	134572	FVAYUVSHYLUEY-LFWYYRBCSA-N	
25	421.3169	86607	M+NH4	6	404.2927	11'-Carboxy-gamma-chromanol	C25H40O4			HMDB0012517			53481453	ITULCXNOMOXAH-YYULODDRSA-N	
26	421.3169	89679	M+NH4	6	404.2927	MG(0:0/22:5(7Z,10Z,13Z,16Z,19Z)/0:0)	C25H40O4			HMDB0011556		52339	53480971	LRBJILYYDLUADN-JLNKQSITSA-N	
27	421.3169	50261	M+NH4	6	404.2927	MG(22:5(4Z,7Z,10Z,13Z,16Z)/0:0/0:0)	C25H40O4			HMDB0011585		52367	53480993	HDIQCISTZKHUDO-AJVITYRPSA-N	
28	421.3169	105373	M+NH4	6	404.2927	MG(0:0/22:5(4Z,7Z,10Z,13Z,16Z)/0:0)	C25H40O4			HMDB0011555		52338	53480970	NPZVSBAEZZLYQU-VMPRHZDHSA-N	
29	421.3169	56824	M+NH4	6	404.2927	MG(22:5(7Z,10Z,13Z,16Z,19Z)/0:0/0:0)	C25H40O4			HMDB0011586		52368	53480994	IDSLCYURGAOTOA-YAWQMZOFSA-N	
30	421.3169	17695	M+H-H2O	5	439.3298	3-hydroxylinoleoylcarnitine	C25H45NO5				LMFA07070042			WQYXCASYXUFNSI-UTJGPVESSA-N	
31	315.2424	17712	M+H	5	315.241	Decanoylcarnitine	C17H33NO4			HMDB0000651	LMFA07070059		10245190	LZOSYCMHQXPBFU-UHFFFAOYSA-N	
32	315.2424	126591	M+H	5	315.241	L-Hexanoylcarnitine n-butyl ester	C17H33NO4					3549			
33	315.2424	17659	M+H	5	315.241	O-decanoyl-R-carnitine	C17H33NO4	3992-45-8	C03299	HMDB0062631	LMFA07070006	36669	11953821	LZOSYCMHQXPBFU-OAHLLOKOSA-N	
34	315.2424	0	M+Na	0	0	No compounds found for eseprimental mass 315.2424 and adduct:									
35	315.2424	14277	M+NH4	5	298.2144	8E-Heptadecenedioic acid	C17H30O4				LMFA01170053	74925		VDTSYDDUGXHLDV-OWOJBTEDSA-N	
36	315.2424	111819	M+NH4	5	298.2144	Plakortic acid	C17H30O4		C17158			71590	10402441	ZCLJFHUIADAYRQ-CMDGGOBGSA-N	
37	315.2424	0	M+H-H2O	0	0	No compounds found for eseprimental mass 315.2424 and adduct:									
38	337.2234	0	M+H	0	0	No compounds found for eseprimental mass 337.2234 and adduct:									
39	337.2234	17712	M+Na	2	315.241	Decanoylcarnitine	C17H33NO4			HMDB0000651	LMFA07070059		10245190	LZOSYCMHQXPBFU-UHFFFAOYSA-N	
40	337.2234	126591	M+Na	2	315.241	L-Hexanoylcarnitine n-butyl ester	C17H33NO4					3549			
41	337.2234	17659	M+Na	2	315.241	O-decanoyl-R-carnitine	C17H33NO4	3992-45-8	C03299	HMDB0062631	LMFA07070006	36669	11953821	LZOSYCMHQXPBFU-OAHLLOKOSA-N	
42	337.2234	1060	M+NH4	6	320.1988	testolic acid	C19H28O4		C09618		LMST02020081	57817	439534	KMUJXIPRPXRPTP-DZBHQSCQSA-N	
43	337.2234	116962	M+NH4	6	320.1988	10beta-Hydroxy-6beta-isobutyrylfuranoeremophilane	C19H28O4		C09685			57884	442377	WWBNAQLGGJMUOJ-BIGGFVEDSA-N	
44	337.2234	130809	M+NH4	6	320.1988	(±)a-CMBHC	C19H28O4	7083-09-2				44824			
45	337.2234	53386	M+NH4	6	320.1988	[8]-Gingerdione	C19H28O4	77334-06-6		HMDB0039276		33864	14440537	QDSRAFNZQKMHPZ-UHFFFAOYSA-N	
46	337.2234	42665	M+NH4	6	320.1988	5'-Carboxy-alpha-chromanol	C19H28O4			HMDB0012798			53481524	QWPNLYBAEZJBMH-MEYZPHKSA-N	

Sheet0

Confirmation by Standard addition



Pipecolic acid

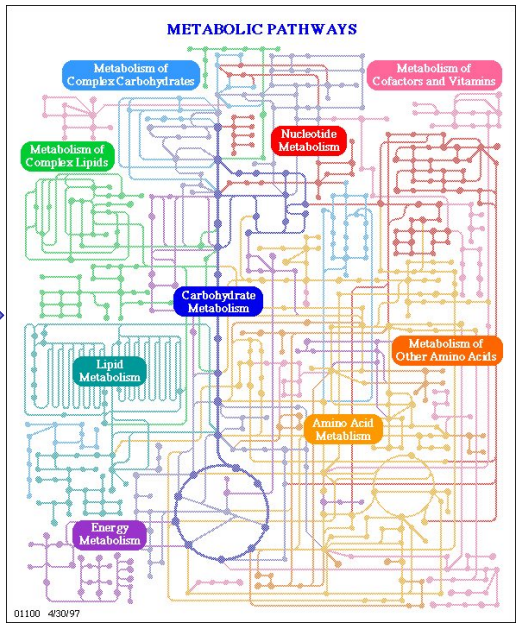


Methyl-lysine

From Lists to Pathways

metabolomics

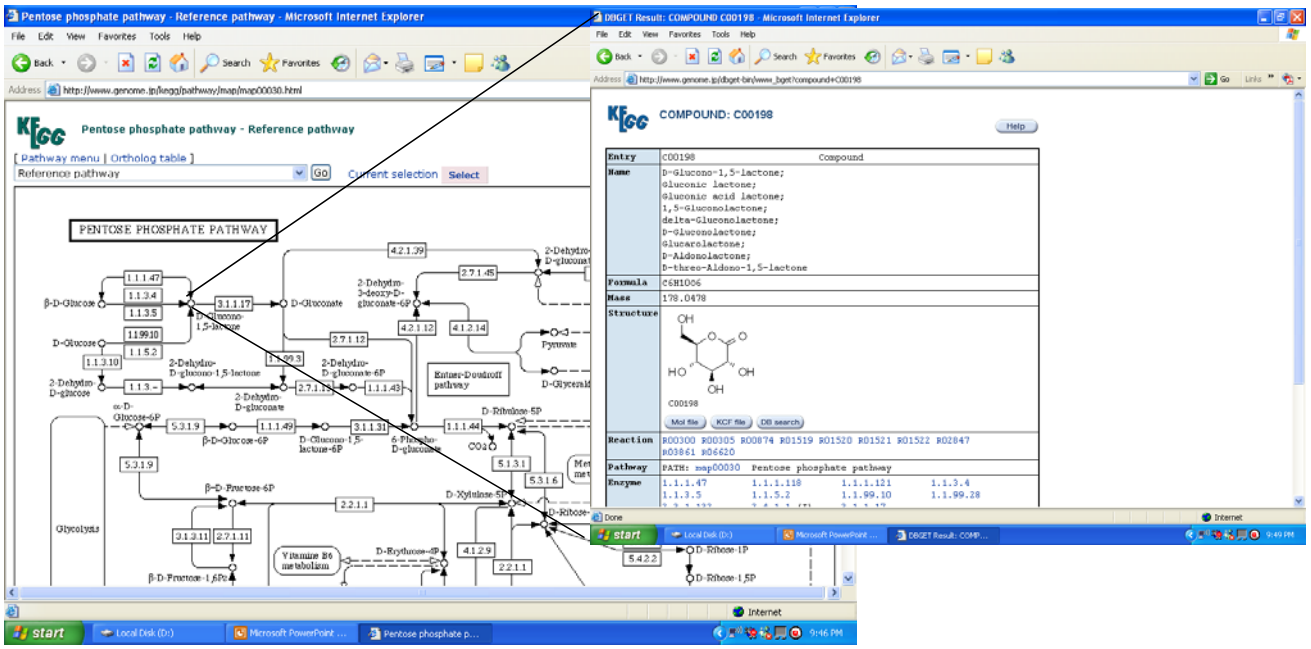
Compound	Retention Time (min)	Conc. in Urine (µM)	Compound	Retention Time (min)	Conc. in Urine (µM)
Dns-o-phospho -L-serine	0.92	<D.L. *	**Dns-Ile**	6.35	25
Dns-o-phospho -L-tyrosine	0.95	<D.L.	**Dns-3-aminosalicylic acid**	6.44	0.5
Dns-adnosine monophosphate	0.99	<D.L.	**Dns-pipecolic acid**	6.50	0.5
Dns-o-phosphoethanolamine	1.06	16	**Dns-Leu**	6.54	54
Dns-glucosamine	1.06	22	**Dns-cystathionine**	6.54	0.3
Dns-o-phospho -L-threonine	1.09	<D.L.	**Dns-Leu-Pro**	6.60	0.4
Dns-6-dimet hylamine purine	1.20	<D.L.	**Dns-5-hydroxylysine**	6.65	1.6
Dns-3-methyl -histidine	1.22	80	**Dns-Cystine**	6.73	160
Dns-taurine	1.25	834	**Dns-N-norleucine**	6.81	0.1
Dns-carnosine	1.34	28	Dns-5-hydroxydopamine	7.17	<D.L.
Dns-Arg	1.53	36	**Dns-dimethylamine**	7.33	293
Dns-Asn	1.55	133	**Dns-5-HIAA**	7.46	18
Dns-hypotaurine	1.58	10	**Dns-umbelliferone**	7.47	1.9
Dns-homocarnosine	1.61	3.9	Dns-2,3 -diaminopropionic acid	7.63	<D.L.
Dns-guanidine	1.62	<D.L.	**Dns-L-ornithine**	7.70	15
Dns-Gln	1.72	633	**Dns-4-acetyamidophenol**	7.73	51
Dns-allantoin	1.83	3.8	**Dns-procaine**	7.73	8.9
Dns-L-citrulline	1.87	2.9	**Dns-homocystine**	7.76	3.3
Dns-1 (or 3 -)-methylhistamine	1.94	1.9	**Dns-acetaminophen**	7.97	82
Dns-adenosine	2.06	2.6	**Dns-Phe-Phe**	8.03	0.4
Dns-methylguanidine	2.20	<D.L.	**Dns-5-methyo xysalicylic acid**	8.04	2.1
Dns-Ser	2.24	511	**Dns-Lys**	8.16	184
Dns-aspartic acid amide	2.44	26	Dns-aniline	8.17	<D.L.
Dns-4-hydroxy -proline	2.56	2.3	**Dns-leu-Phe**	8.22	0.3
Dns-Glu	2.57	21	**Dns-His**	8.35	1550
Dns-Asp	2.60	90	Dns-4-thialysine	8.37	<D.L.
Dns-Thr	3.03	157	Dns-benzylamine	8.38	<D.L.
Dns-epinephrine	3.05	<D.L.	**Dns-1-ephedrine**	8.50	0.6
Dns-ethanolamine	3.11	471	**Dns-tryptamine**	8.63	0.4
Dns-aminoadipic acid	3.17	70	Dns-pyridoxamine	8.94	<D.L.
Dns-Gly	3.43	2510	Dns-2-methyl -benzylamine	9.24	<D.L.
Dns-Ala	3.88	593	**Dns-5-hydroxytrptophan**	9.25	0.12
Dns-aminolevulinic acid	3.97	30	**Dns-1,3 -diaminopropane**	9.44	0.23
Dns-r-amino -butyric acid	3.98	4.6	**Dns-putrescine**	9.60	0.5
Dns-p-amino -hippuric acid	3.98	2.9	**Dns-1,2 -diaminopropane**	9.66	0.1
Dns-5-hydro xymethyluricil	4.58	1.9	**Dns-tyrosinamide**	9.79	29
Dns-tryptophanamide	4.70	5.5	**Dns-dopamine**	10.08	140
Dns-isoguanine	4.75	<D.L.	**Dns-cadaverine**	10.08	0.08
Dns-5-aminopentanoic acid	4.79	1.6	**Dns-histamine**	10.19	0.4
Dns-sarcosine	4.81	7.2	**Dns-3-methoxy -tyramine**	10.19	9.2
Dns-3-amino -isobutyrate	4.81	85	**Dns-Tyr**	10.28	321
Dns-2-aminobutyric acid	4.91	17	Dns-cysteamine	10.44	<D.L.



Pathway Databases

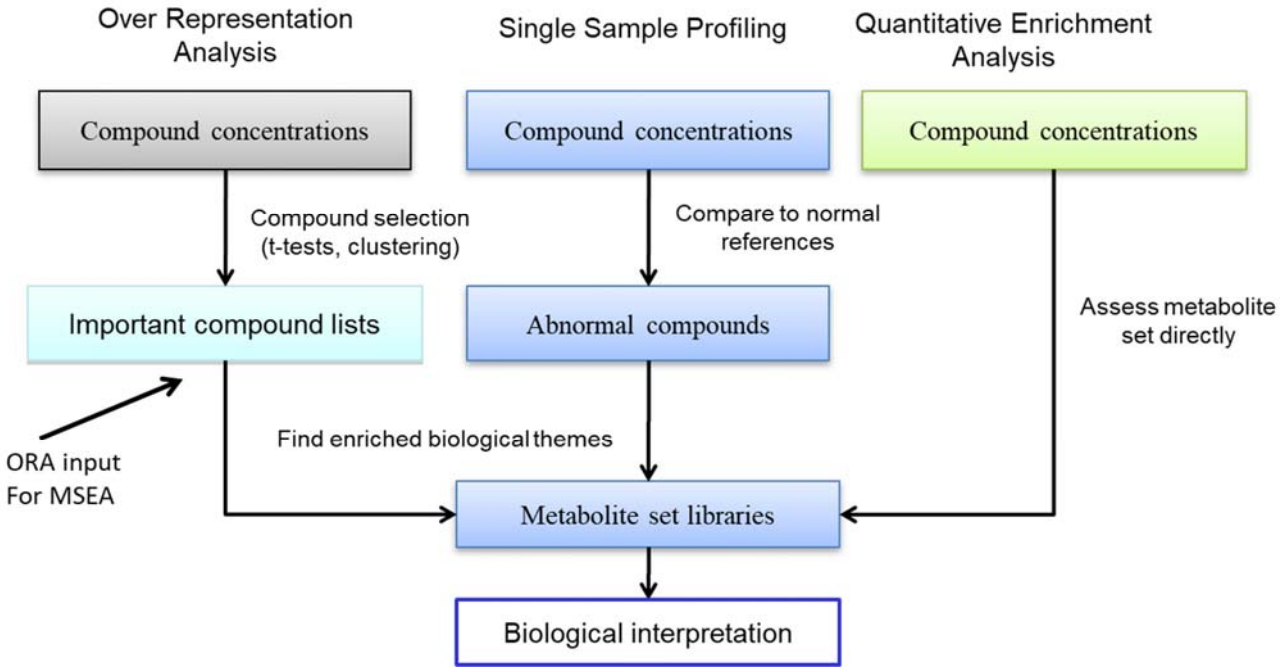
- Rich source of biological data that relates metabolites to genes, proteins, diseases, signaling events and processes
- Provide various tools to permit visualization and gene/metabolite mapping
- Often cover multiple species
- KEGG (www.genome.jp/kegg/), BioCyc/MetaCyc (https://biocyc.org/), SMPDB (www.smpdb.ca), Reactome (www.reactome.org), WikiPathways (http://www.wikipathways.org)...
- "*Strictly speaking, one could argue that pathways don't exist... there are only networks*." (WikiPathways.org)

KEGG – Kyoto Encyclopedia of Genes and Genomes

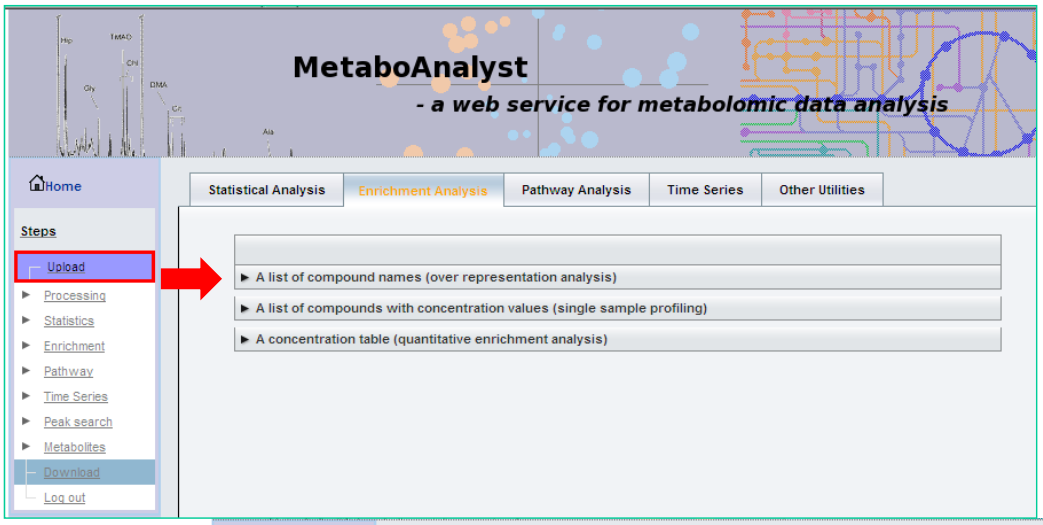


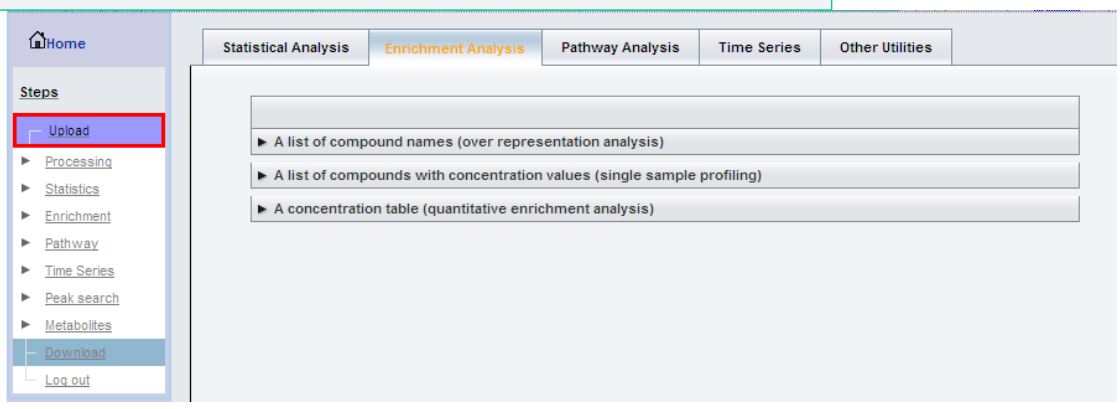
http://www.genome.jp/kegg/

The Metabolite Set Enrichment Analysis MSEA approach

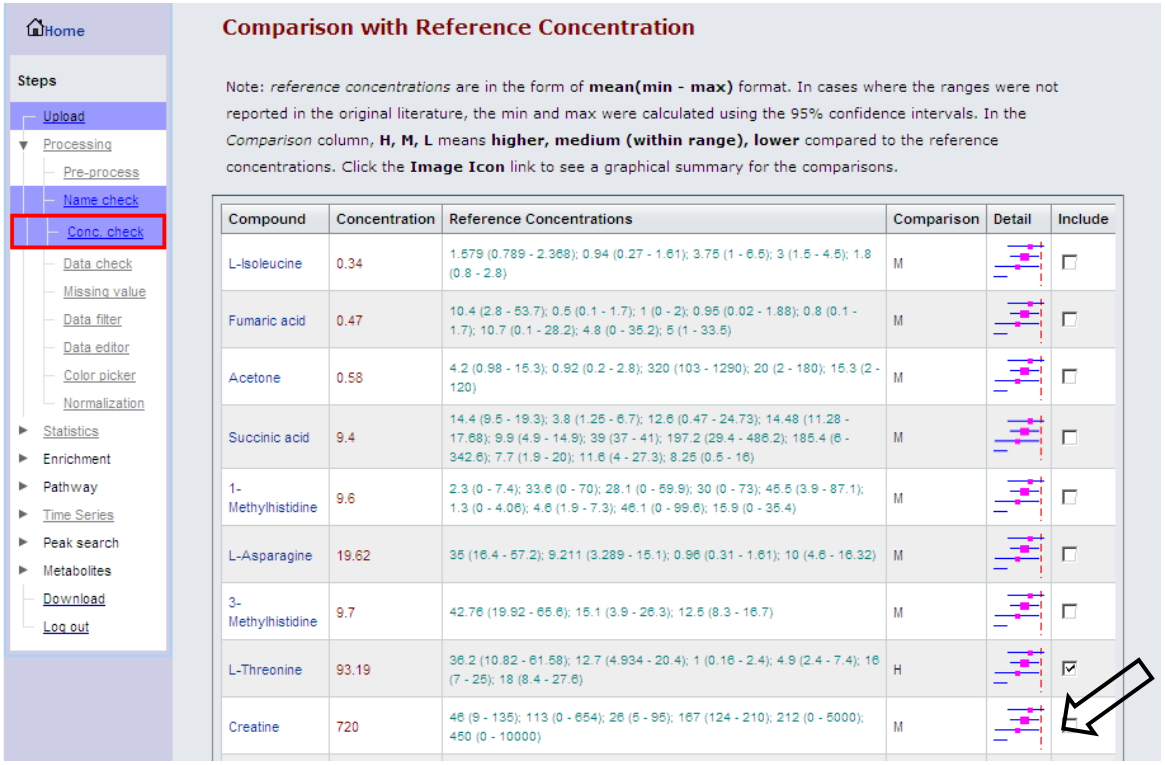


Start with a compound List

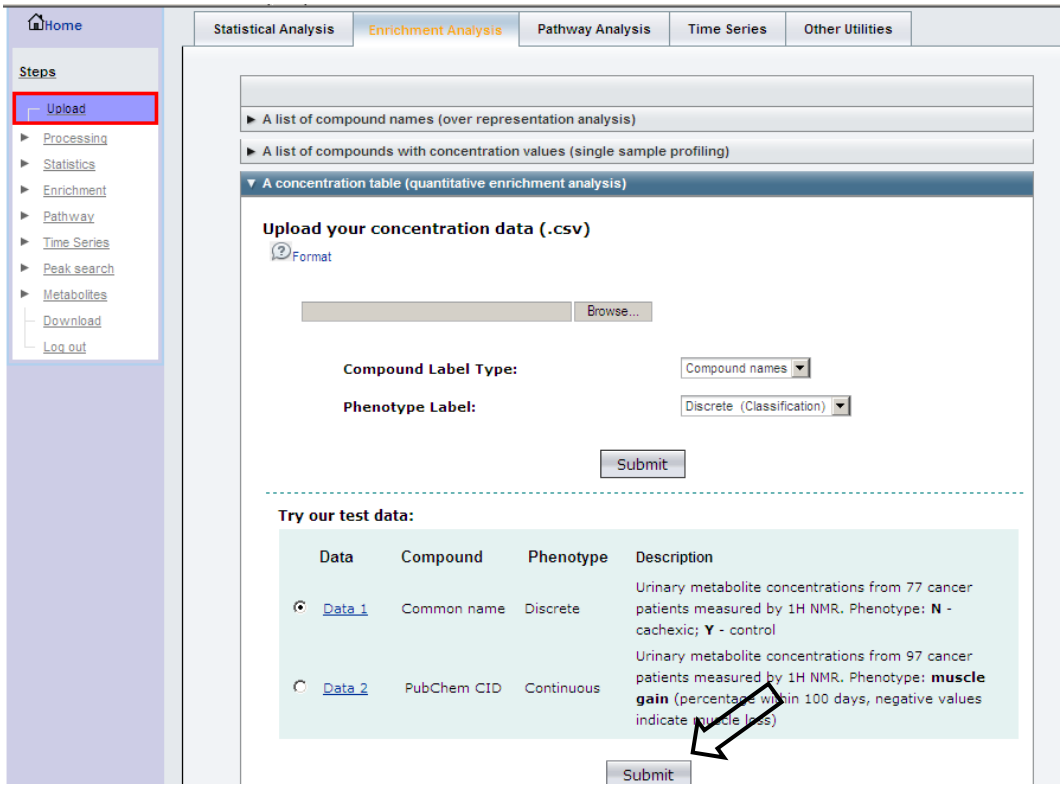




Concentration Comparison



Quantitative Enrichment Analysis

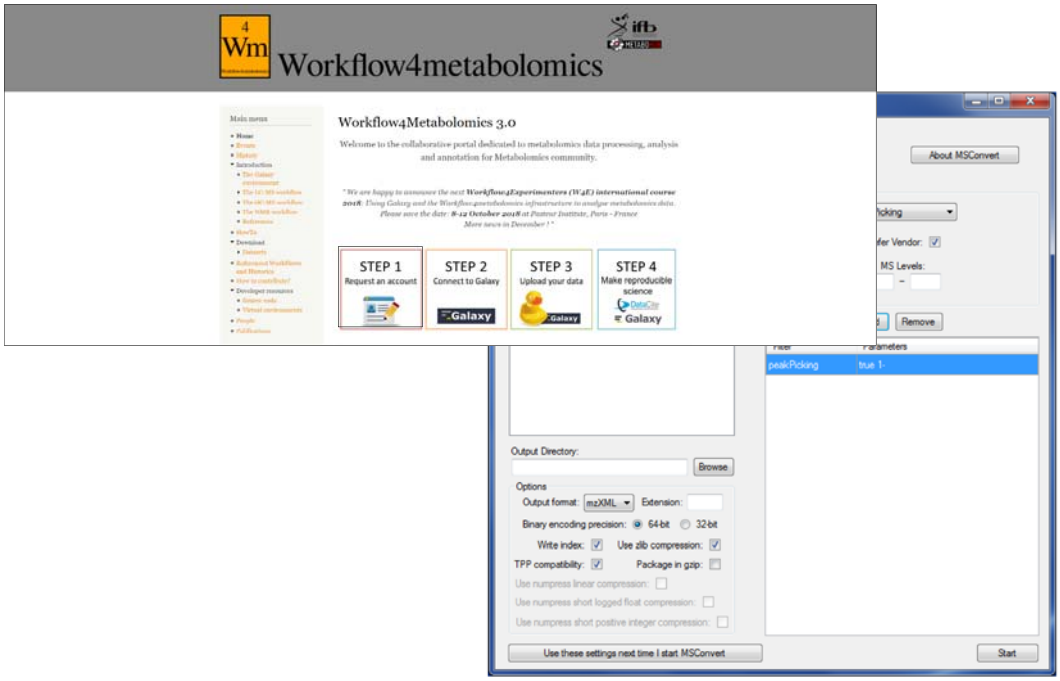


RESULT

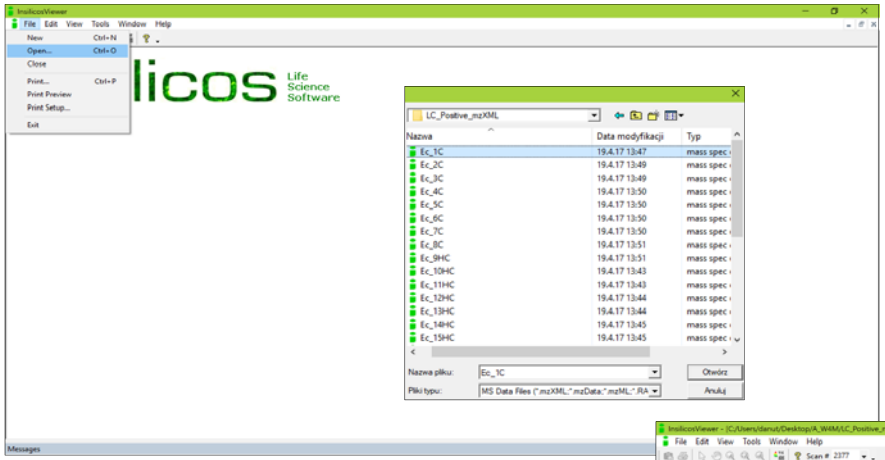
Metaboanalyst
Metabolic Pathway Analysis (MetPA)

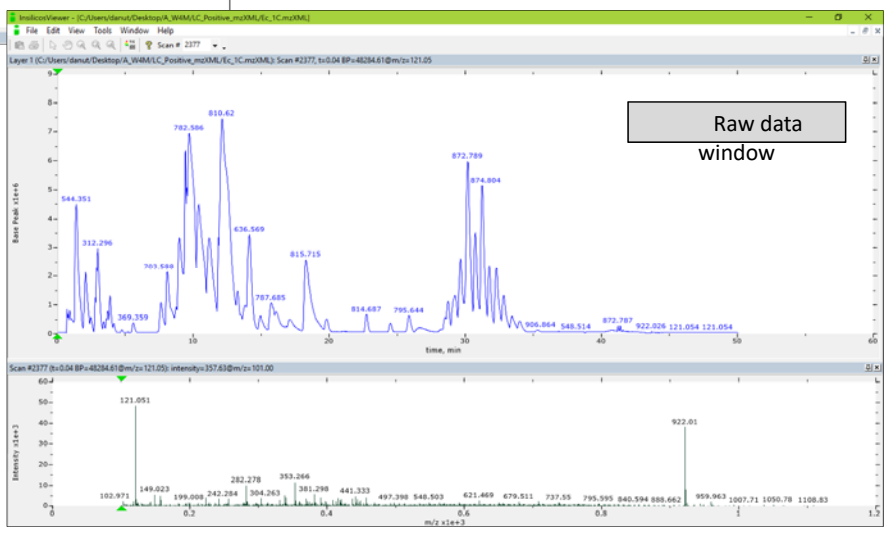
- Purpose: to extend and enhance metabolite set enrichment analysis for pathways by
  - Considering the **structures of pathway**
  - Dynamic pathway visualization
- Currently supports ~1500 pathways covering 17 organisms (based on KEGG)

PRACTICAL SESSION. VISUALS_1

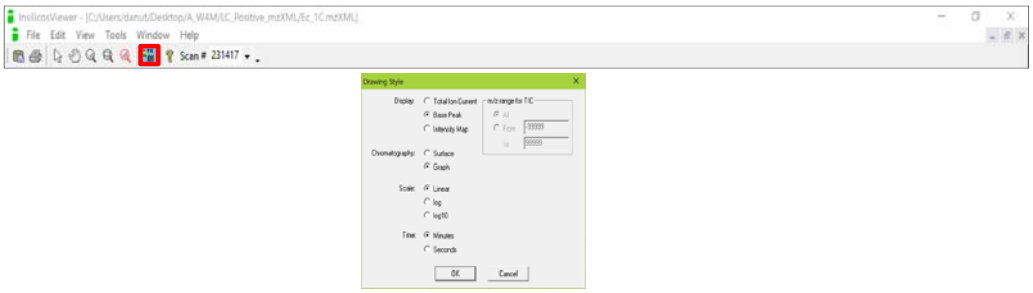


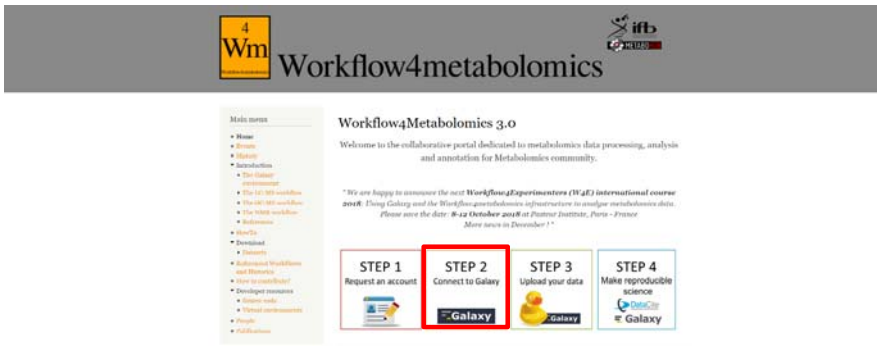
PRACTICAL SESSION. VISUALS_2



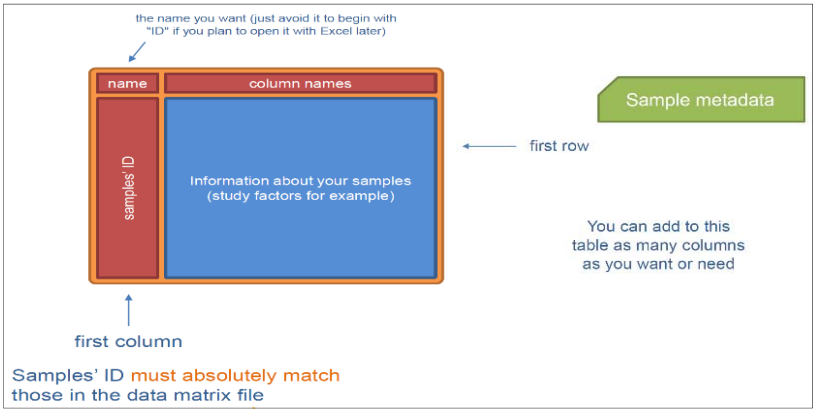


PRACTICAL SESSION. VISUALS_3





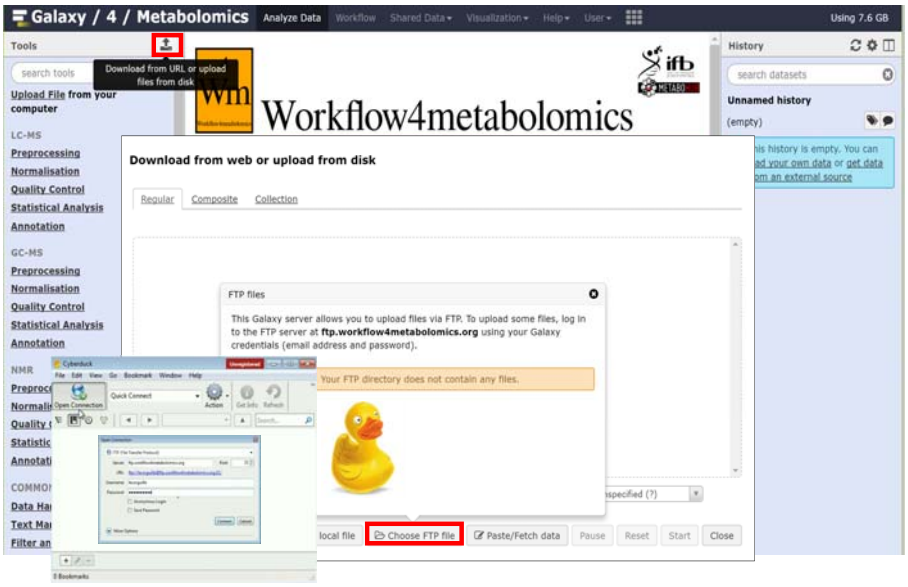
PRACTICAL SESSION. VISUALS_4

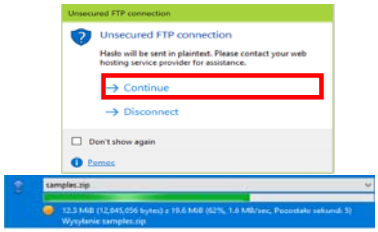


PRACTICAL SESSION. VISUALS_5

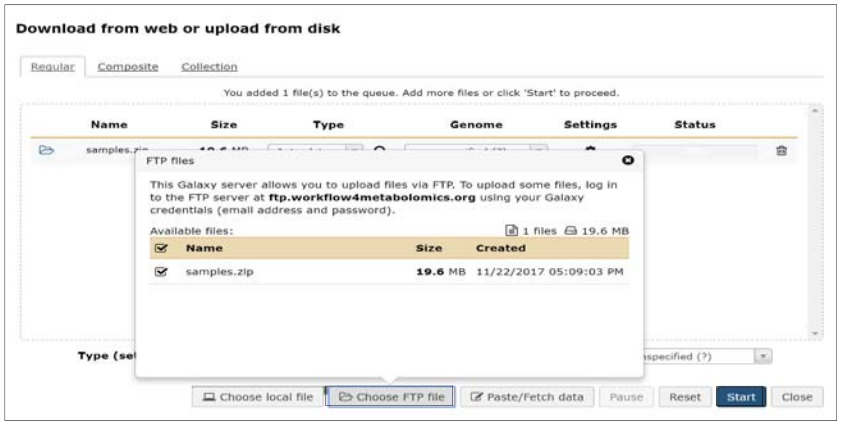
sampleName	class	polarity	sampleType	batch	injectionOrder	diet
QC	one	positive	pool	B1	1	NA
C1	one	positive	sample	B1	7	C
HC3	one	positive	sample	B1	10	HC
BL	one	positive	blank	B1	12	NA
…	…	…	…	…	…	…

PRACTICAL SESSION. VISUALS_6

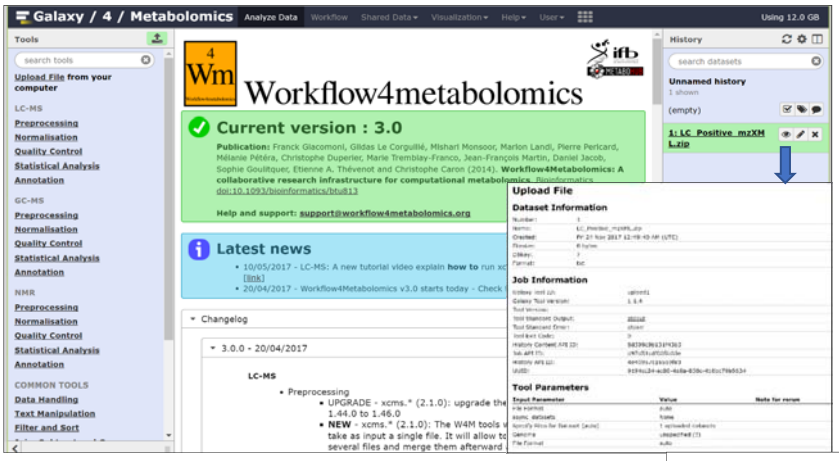


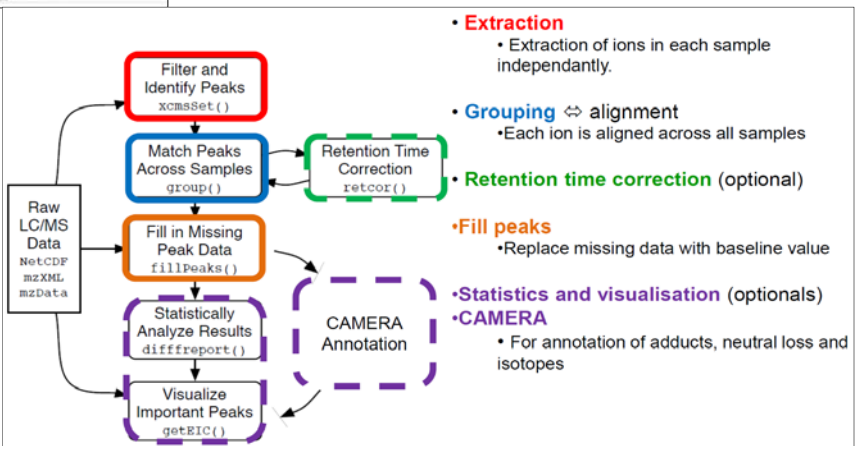


PRACTICAL SESSION. VISUALS_7

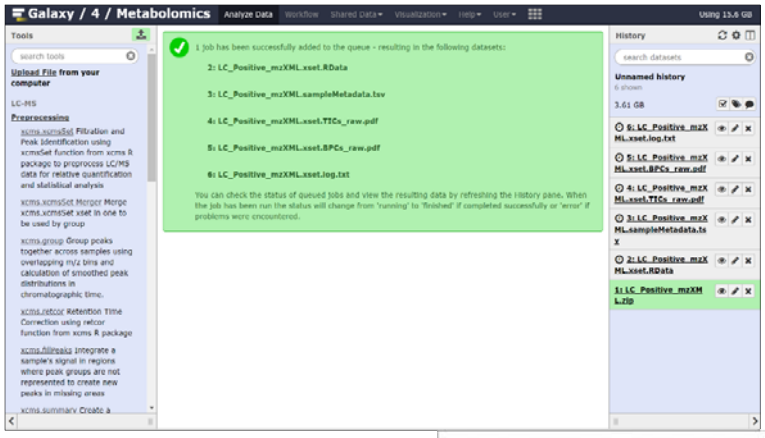


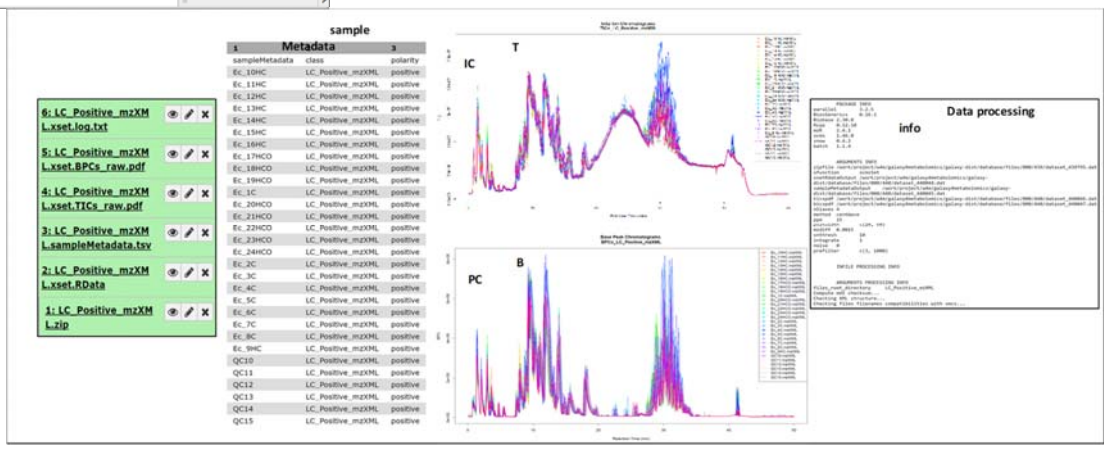
PRACTICAL SESSION. VISUALS_8





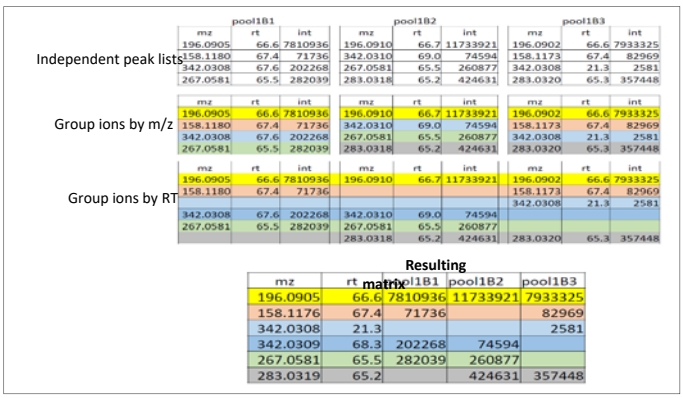
PRACTICAL SESSION. VISUALS_9





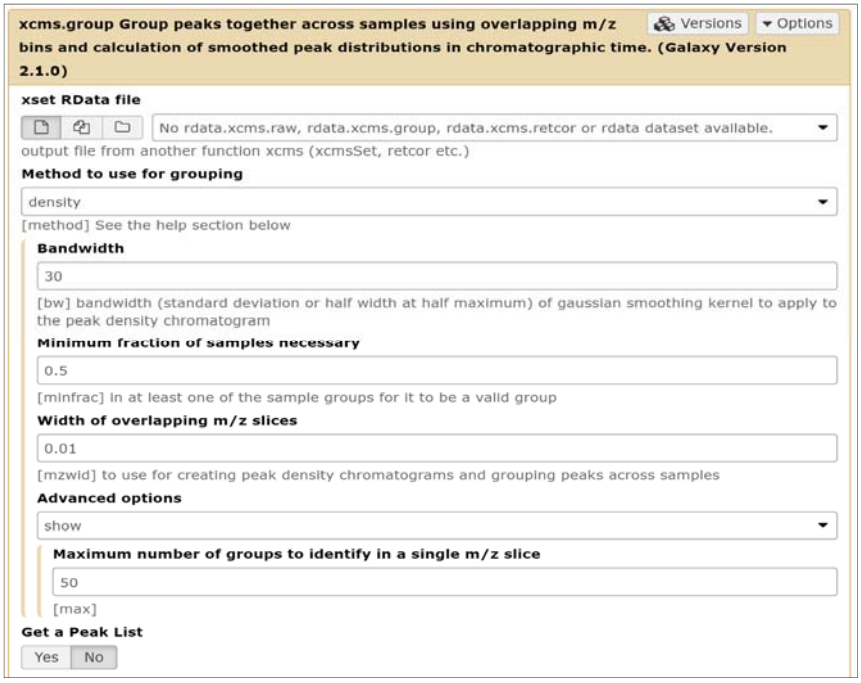
PRACTICAL SESSION. VISUALS_10

xcms.group Group peaks together across samples using overlapping m/z bins and calculation of smoothed peak distributions in chromatographic time.



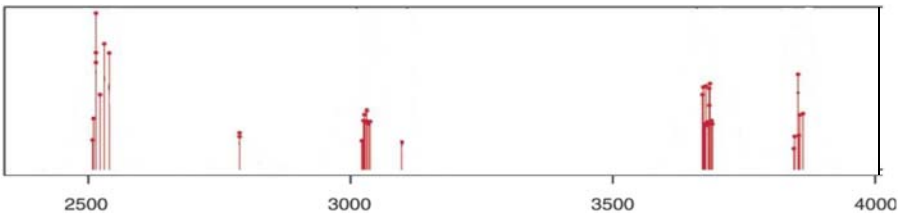
Parameter : num + label	Format
Or : RData file | rdata.xcms.raw	
Or : RData file | rdata.xcms.retcor	

PRACTICAL SESSION. VISUALS_11

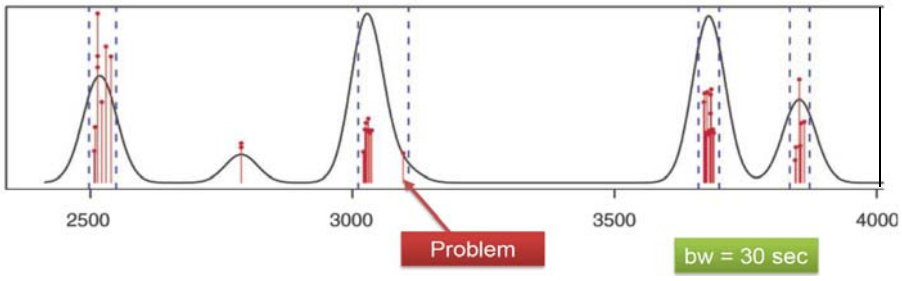


PRACTICAL SESSION. VISUALS_12

Grouping peaks in mass bin: 337.975 – 338.225 m/z (mzwid)

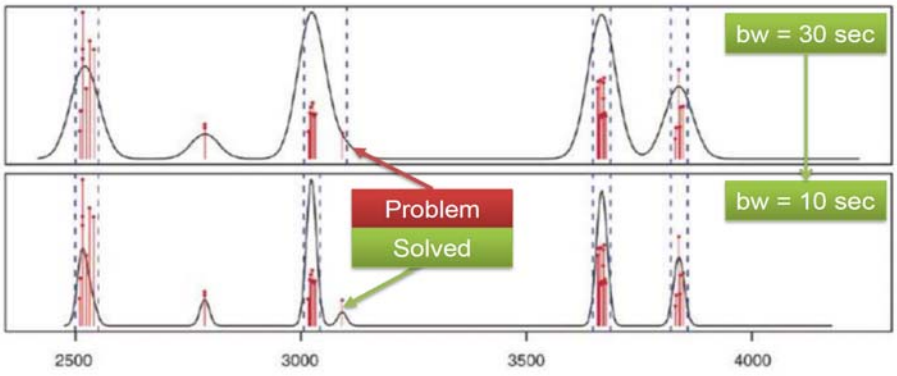


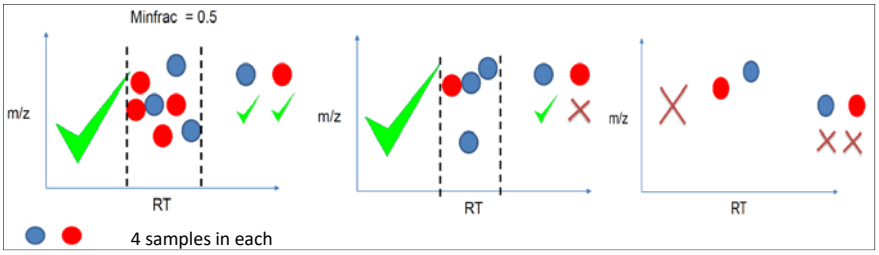
Grouping peaks in mass bin: 337.975 – 338.225 m/z (mzwid)



PRACTICAL SESSION. VISUALS_13

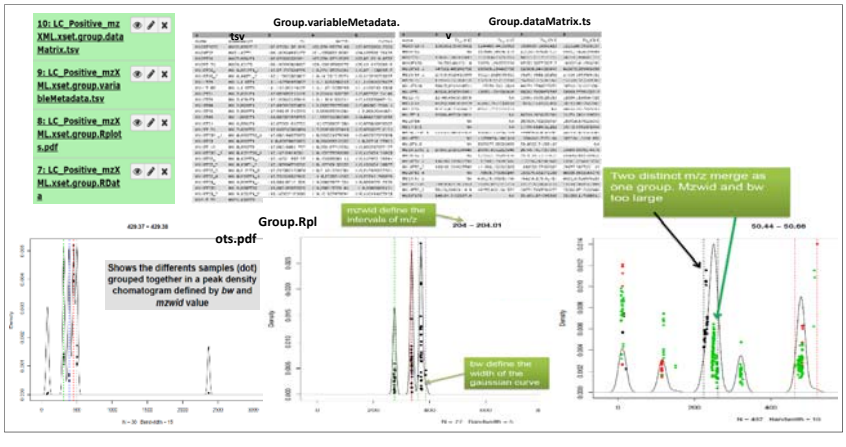
Grouping peaks in mass bin: 337.975 – 338.225 m/z (mzwid)





4 samples in each group

PRACTICAL SESSION. VISUALS_14



PRACTICAL SESSION. VISUALS_15

xcms.retcor Retention Time Correction using retcor function from xcms R package

Parameter : num + label	Format
1 : RData file	rdata.xcms.group

xcms.retcor Retention Time Correction using retcor function from xcms R package (Galaxy Version 2.1.0) Versions Options

xset RData file

No rdata.xcms.raw, rdata.xcms.group or rdata dataset available.

output file from another function xcms (xcmsSet, retcor etc.)

Method to use for retention time correction

peakgroups

[method] See the help section below

Smooth method

loess

[smooth] either 'loess' for non-linear alignment or 'linear' for linear alignment

Number of extra peaks to allow in retention time correction correction groups

1

[extra]

Number of missing samples to allow in retention time correction groups

1

Number of admitted missing well behaved peak in a group.

[missing]

Advanced options

hide

Resubmit your raw dataset or your zip file

Execute

PRACTICAL SESSION. VISUALS_16

Advanced options

show

Degree of smoothing for local polynomial regression fitting

0.2

[span]

Family

gaussian

[family] if gaussian fitting is by least-squares with no outlier removal, and if symmetric a re descending M estimator is used with Tukey's biweight function, allowing outlier removal

plottype

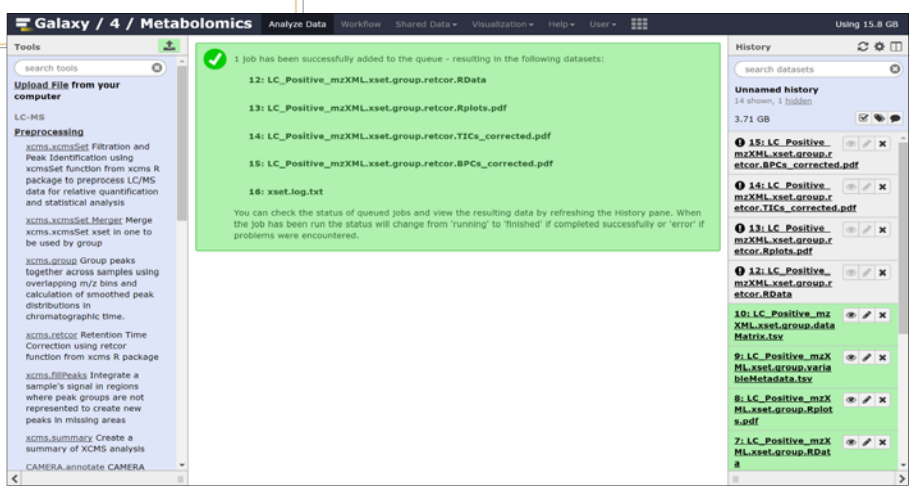
deviation

Plot to visualize the result of the retention time correction.

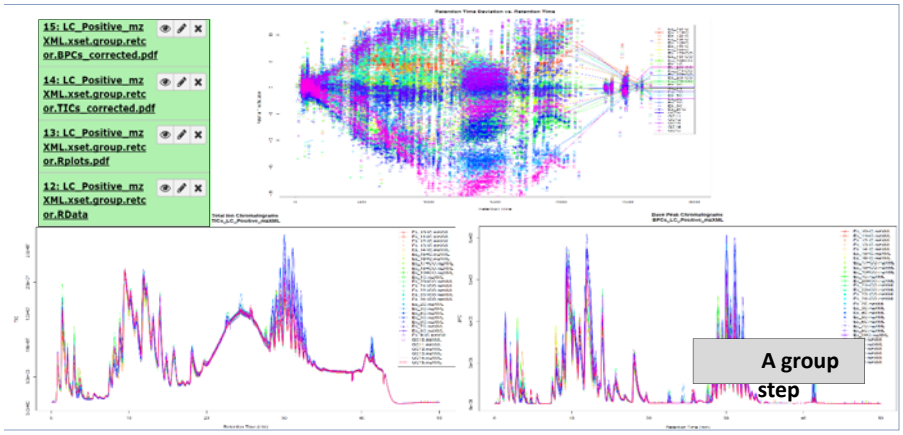
[plottype] if deviation plot retention time deviation points and regression fit, and if mdevden also plot peak overall peak density and retention time correction peak density

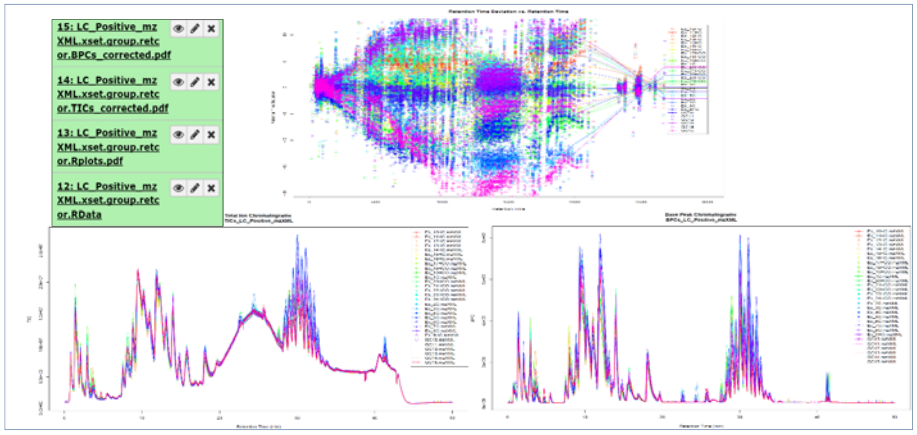
Resubmit your raw dataset or your zip file

Execute

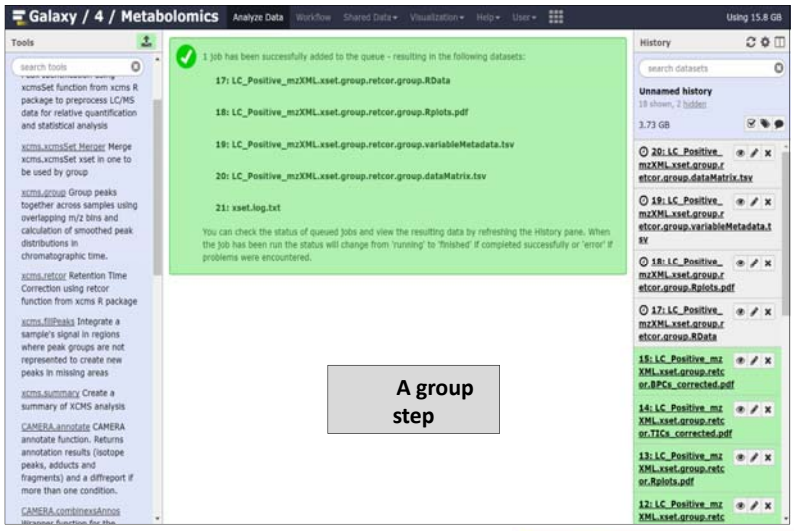


PRACTICAL SESSION VISUALS_17



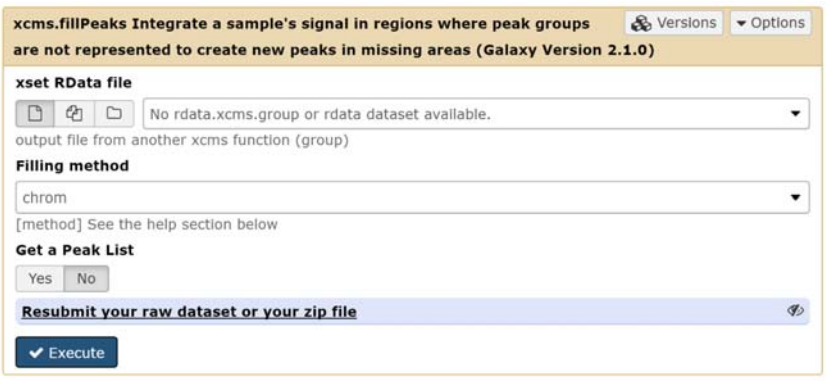


PRACTICAL SESSION. VISUALS_18

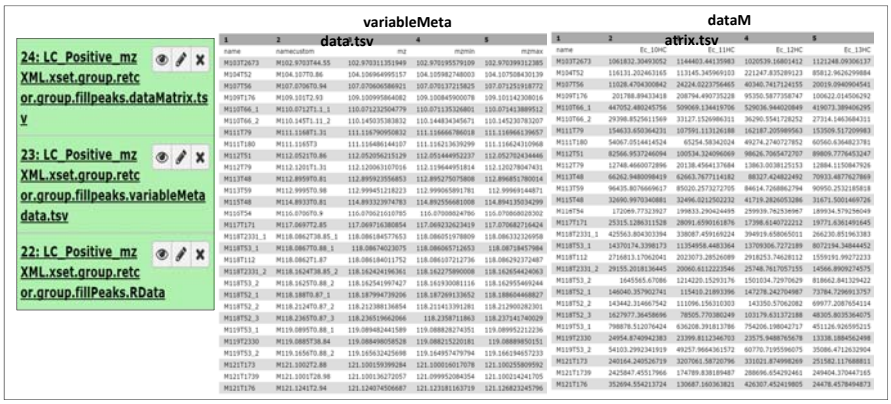


xcms.fillPeaks Integrate a sample's signal in regions where peak groups are not represented to create new peaks in missing areas

Parameter : num + label	Format
1 : RData file	rdata.xcms.group

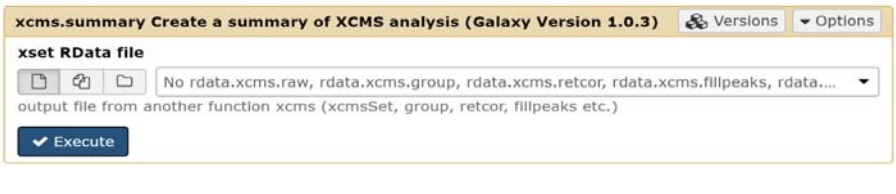


PRACTICAL SESSION. VISUALS_19



PRACTICAL SESSION. VISUALS_20

xcms.summary Create a summary of XCMS analysis



This project has been funded with support from the European Commission.

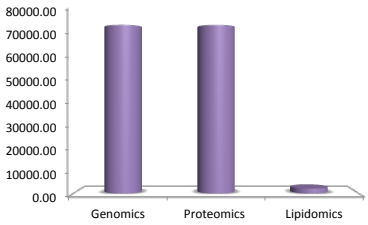
This publication reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein

Mass Spectrometry Based Lipidomics

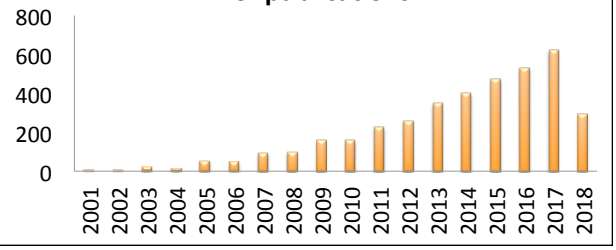
Elisabete Maciel,
Eliana Alves,
Pedro Domingues,
Rosário Domingues

Lipidomics

Nº of published papers by "omics" area(Scopus)



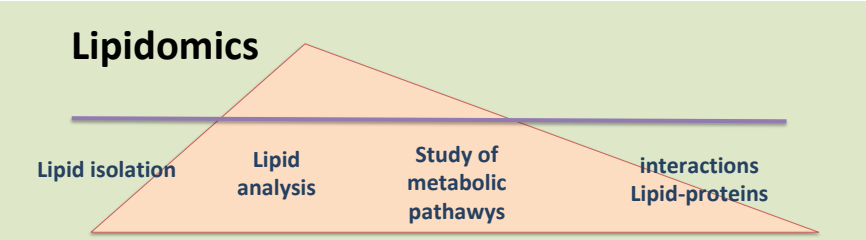
nº of publications



Lipidomics

- The full characterization of lipid molecular species and of their biological roles with respect to expression of proteins involved in lipid metabolism and function, including gene regulation (AOCS Lipids Library)

- Analysis of lipid profile and its relation to cell physiology and pathophysiology

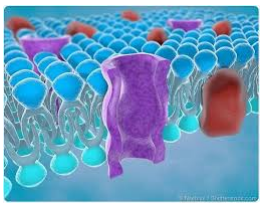


Lipidomics

- Profiling cellular lipidome
- Membrane lipid domains & dynamics
- Regulatory (e.g. signaling) functions of lipids
- Integration of omics & interaction of cellular complement & machinery to form cells/organism

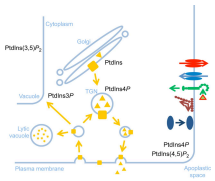
Why lipids are so important?

Membranes



Cellular Regulation

signaling messengers, hormones, …

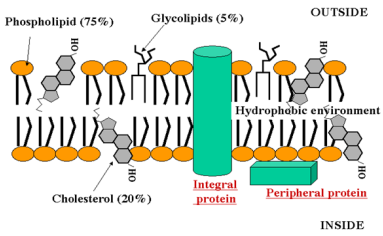


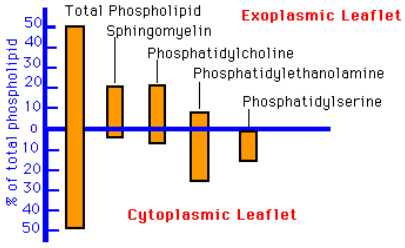
Energy Metabolism/Reserves

Dysfunction of lipid signaling & metabolism plays a central role in ***health & diseases***

Lipids - Cell Membrane

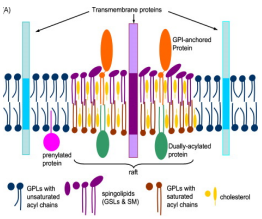
Membrane Assymetry





Membrane domains

Lipid rafts



Lipid profiling in cell, tissues and biofluids

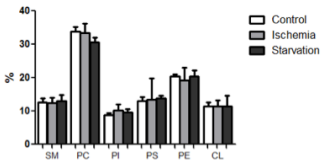
Each type of cell, tissue and body fluid have a characteristic lipid profile with a defined lipid compositions.

Identification of all cellular lipids -

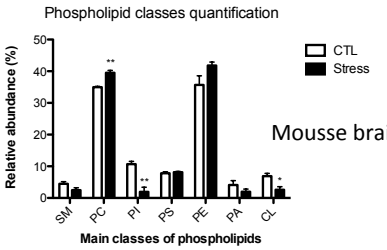
lipidome

	% of total acyl lipid content		
Glycerolipids	Chloroplast thylakoid	inner mitochondrial membrane	plasma membrane
MGDG	51%	0	0
DGDG	26	0	0
SQDG	7	0	0
PC	3	27	32
PS	0	25	0
PG	9	0	0
PE	0	29	46
PI	1	0	19
CL	0	20	0

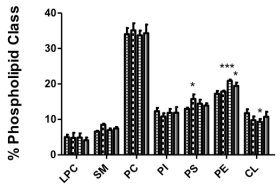
Profile of Phospholipid classes



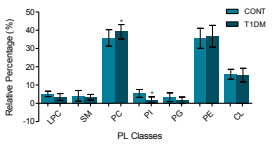
Cardiomyocytes cell line



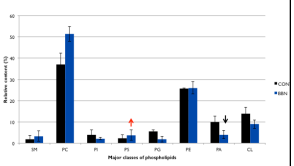
Mousse brain



Breast cancer cell Lines

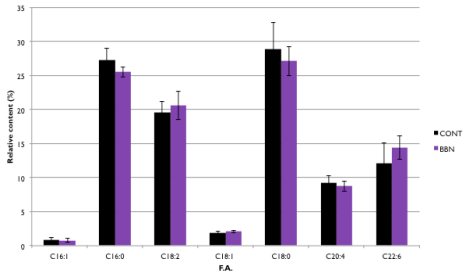


Mitochondria from heart

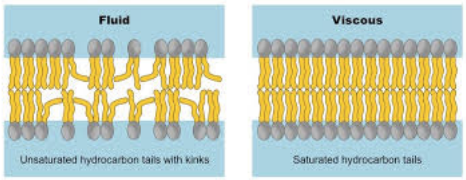


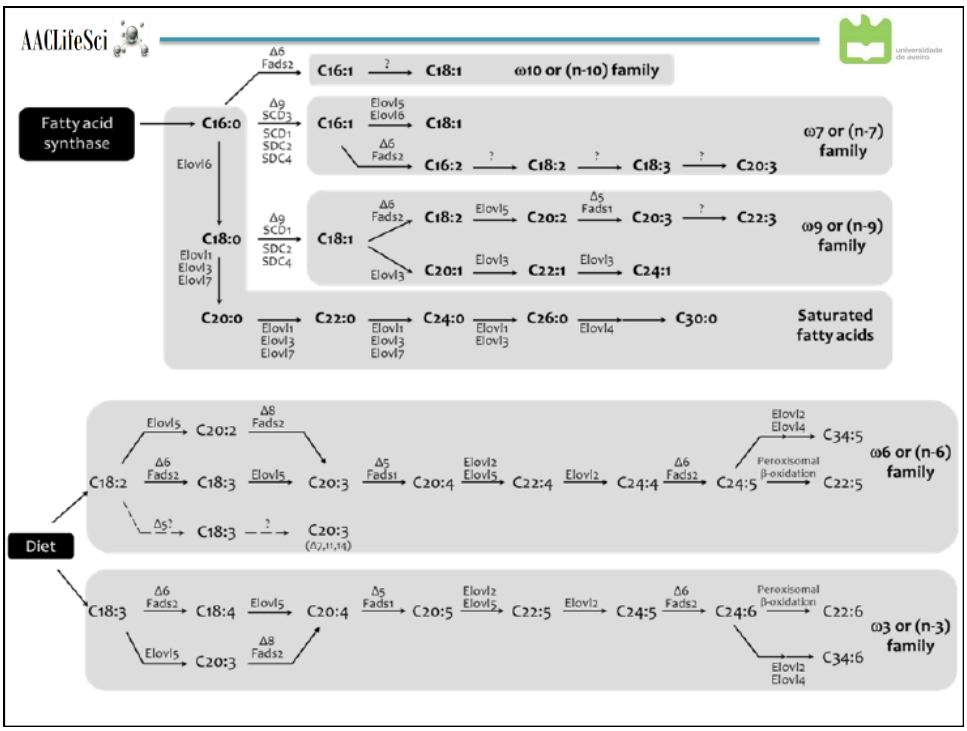
Mitochondria from muscle

Profile of Fatty Acids



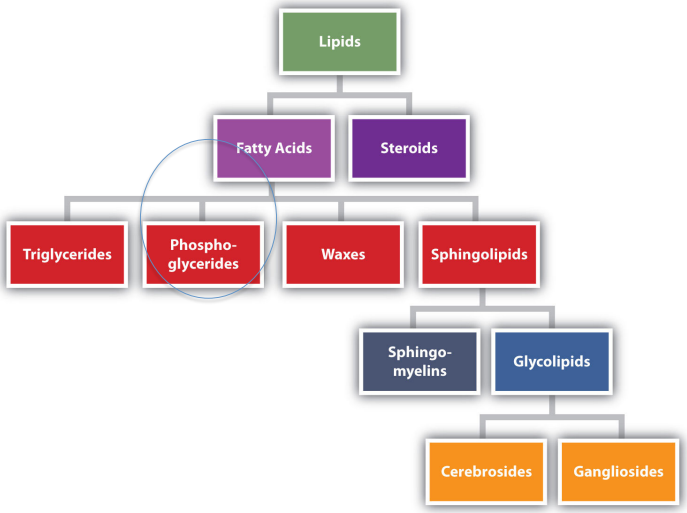
Effects of PL fatty acid composition in membrane properties





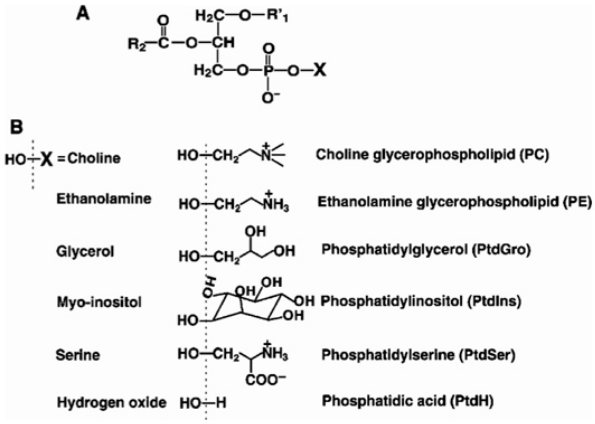
What are the big challenges in lipidomics?

Structural complexity of lipids

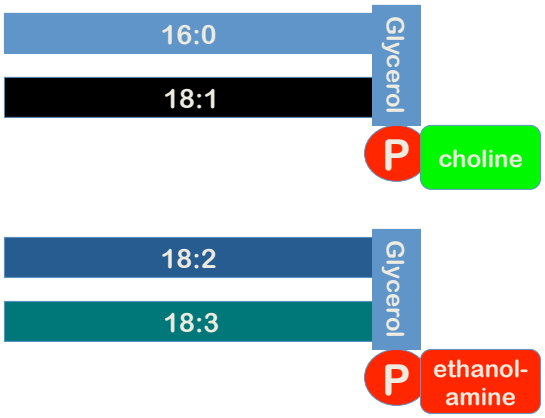


Phospholipids/ Glycerolipids Molecular Species

Also called glycerophospholipids

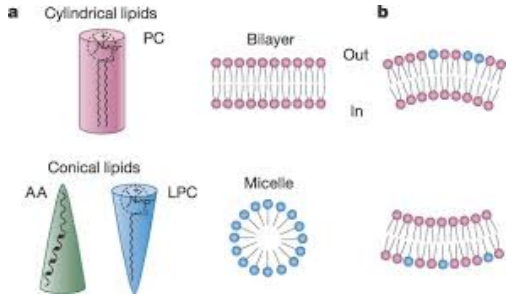


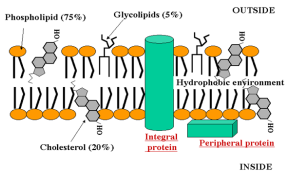
Phospholipids/ Glycerolipids Molecular Species

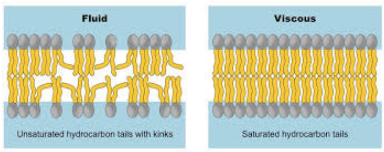


PC, PE, PS, NPE, PG, CL, PI, PIP, PIP_2, PIP_3, LysoPLs……

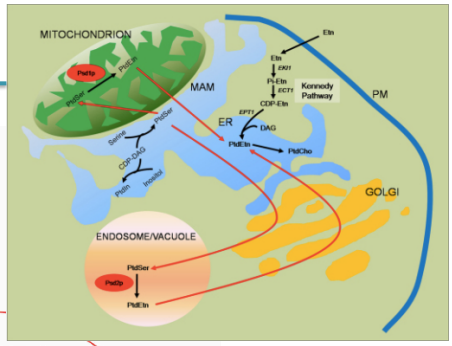
PL & fatty acid composition & membrane properties

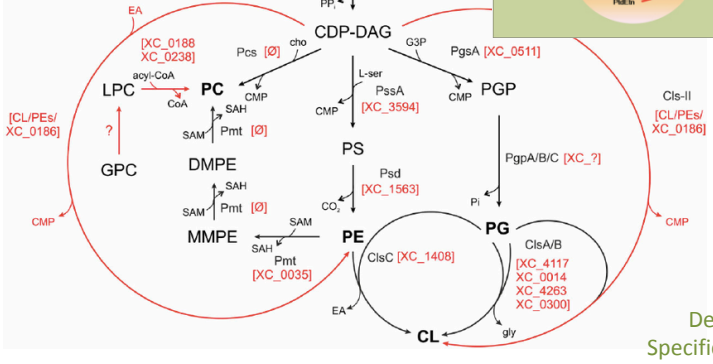






Phospholipid Biosynthesis





Define and regulate
Specific Lipidome signatures

Deviations in the Lipidome

Disease

- Alteration in metabolic pathways
- Oxidative modification of some lipids

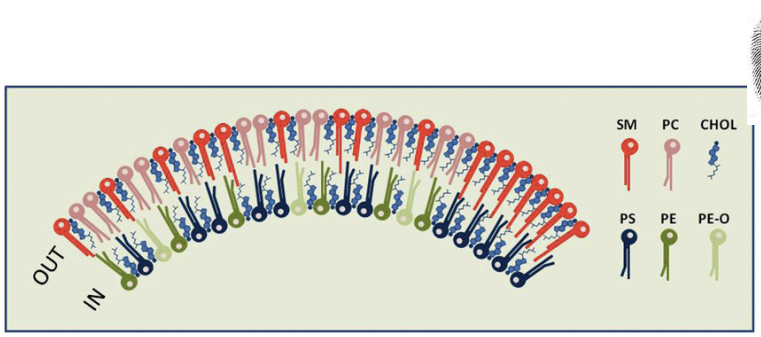
Others:

- Diet –source of different lipids

Importance:

New biomarkers
New therapeutic strategies
New biotechnological applications

Lipidomic analytical strategies to overcome the complexity of the lipidome



Lipidomic approach

Lipossomes | Cells | Tissues

↓

Lipid Extraction

↓

Fractionation/Separation by Chromatography: TLC, HPLC or SPE

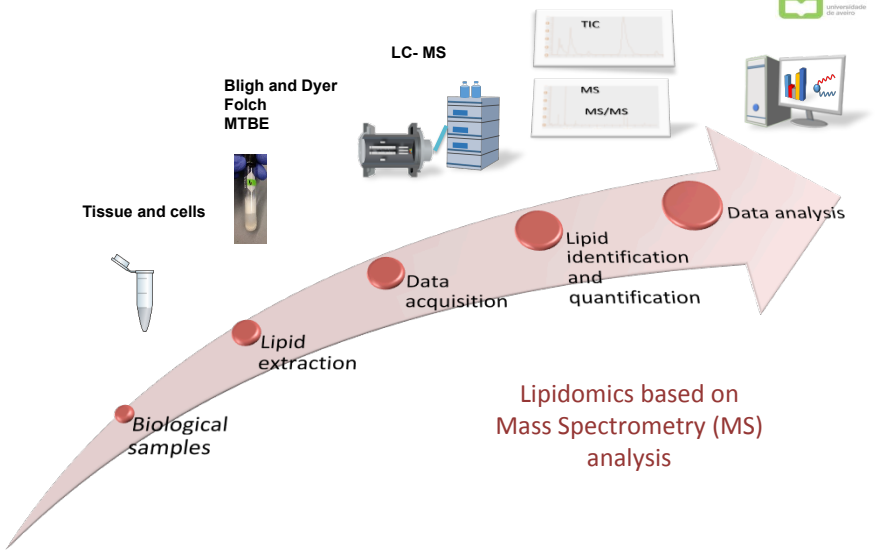
↓

Mass Spectromatry (MS) Analysis

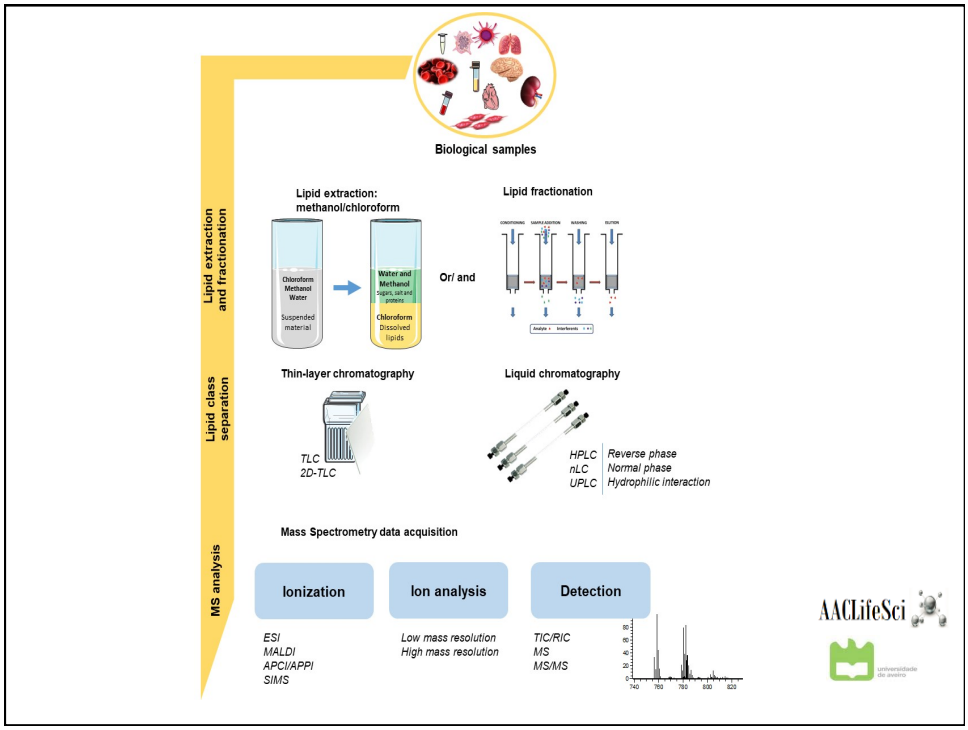
Untarget analysis

Targeret analysis

Lipidomics workflow



Lipidomics based on Mass Spectrometry (MS) analysis

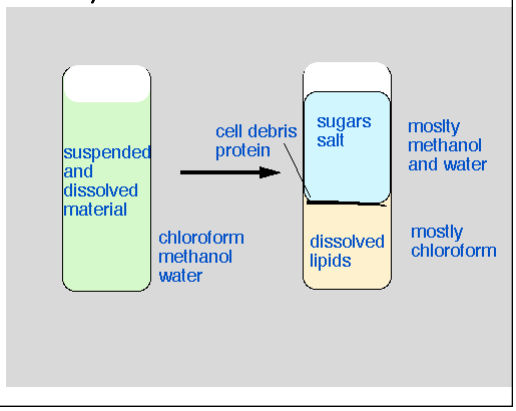


Lipid Extraction

AACLifeSci

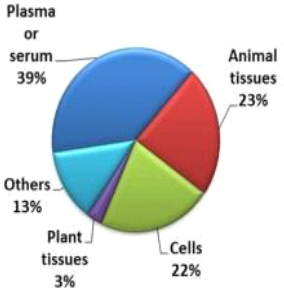
Chemical extraction using organic solvents:

- Folch method ($CHCl_3$:CH_3OH 2:1)
- Bligh and Dyer ($CHCl_3$:CH_3OH 1:2)
- others

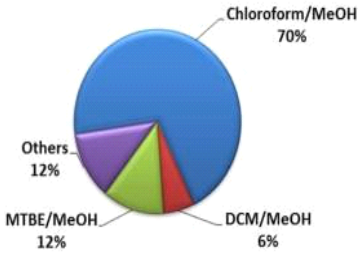


Lipid Extraction

Analyzed Matrices



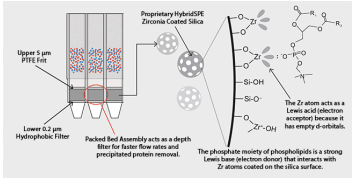
Extration protocols

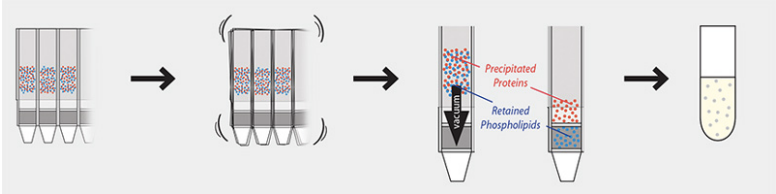


Cajka and Fiehn, Trens in Analytical chemistry, 2014

Lipid Extraction

Selective extraction of phospholipids from plasma using Hybrid SPE



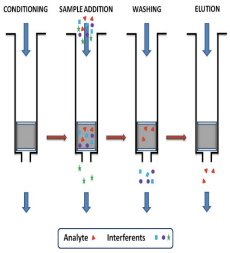


Fractionation of lipid extracts

Solid phase extraction

To separate neutral from polar lipids

Neutral lipids(TG) from PL

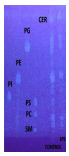


Chromatographic methods

TLC (Thin layer chromatography)

HPLC (High performance liquid chromatography)

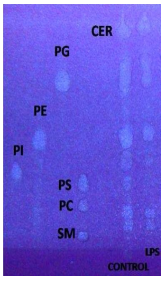
To separate lipid classes/molecular species



Separation of phospholipid classes

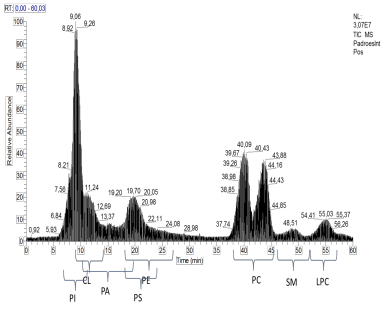
Phospholipid classes can be separated based on their polarity by:

TLC



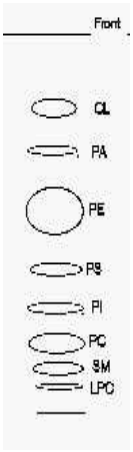
HPLC

I



TLC – Thin Layer Chromatography

Different elution systems- different TLC profiles



CL-Cardiolipin

PA-Phosphatidic Acid

PE-Phosphatidylethanolamine

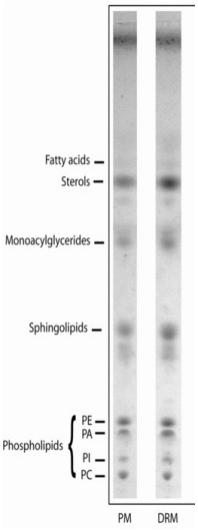
PS-Phosphatidylserine

PI-Phosphatidylinositol

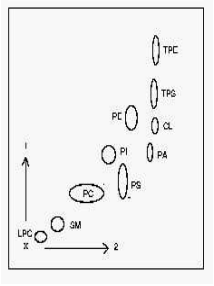
PC-Phosphatidylcholine

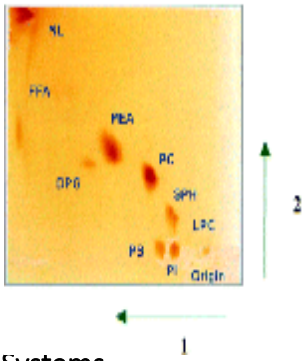
SM-Spingomyelin

LPC-Lyso Phosphatidylcholine



2D-TLC

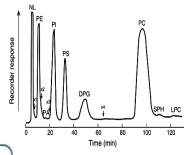


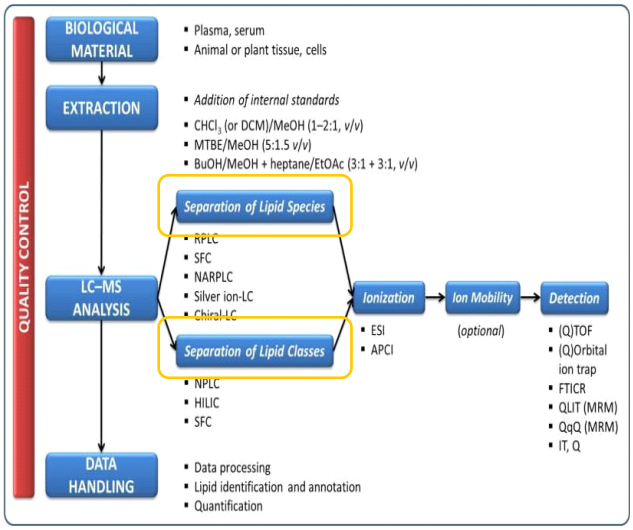


Two Diferent Solvent Systems

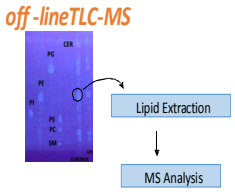
Prof Valerian Kagan lab

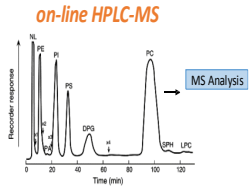
HPLC-MS

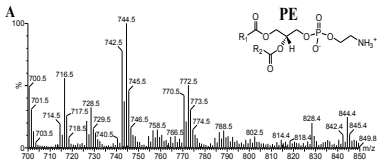


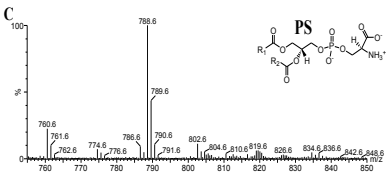


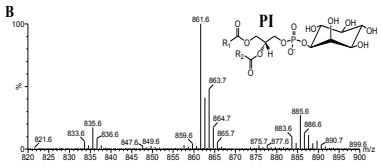
Cajka and Fiehn, Trens in Analytical chemistry, 2014

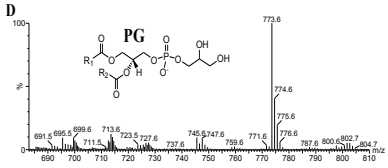












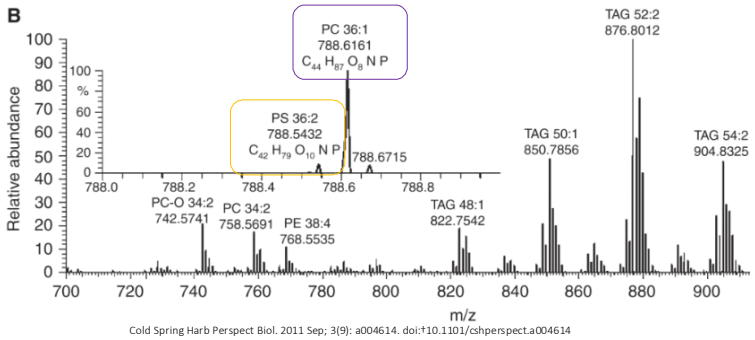
MS DATA ANALYSIS:

Ions formed during ionization of lipids

Lipid class	Positive mode	Negative mode
LPC, PC	$[M + H]^+$, $[M + Na]^+$	$[M–H]^-$, $[M + HCOO]^-$, $[M + CH_3COO]^-$
LPE, PE	**$[M + H]^+$, $[M + Na]^+$**	**$[M–H]^-$**
PG	$[M + H]^+$, $[M + NH_4]^+$, $[M + Na]^+$	$[M–H]^-$
PI	$[M + H]^+$, $[M + NH_4]^+$, $[M + Na]^+$	$[M–H]^-$
PS	**$[M + H]^+$**	**$[M–H]^-$**
PA		$[M–H]^-$
CE	$[M + NH_4]^+$, $[M + Na]^+$	
SM	$[M + H]^+$	$[M + HCOO]^-$, $[M + CH_3COO]^-$
Cholesterol	$[M–H_2O+H]^+$	
MG, DG, TG	$[M + NH_4]^+$, $[M + Na]^+$	
MGDG, DGDG, SQDG	$[M + NH_4]^+$, $[M + Na]^+$	$[M–H]^-$
Fatty acids		$[M–H]^-$
CL	$[M + H]^+$, $[M + NH_4]^+$, $[M + Na]^+$	$[M–H]^-$, $[M–2H]^{2-}$
Cer, GluCer, LacCer	$[M + H]^+$, $[M + NH_4]^+$, $[M + Na]^+$	$[M–H]^-$, $[M + HCOO]^-$, $[M + CH_3COO]^-$

High Resolution Mass Spectrometry (HRMS)

- High mass accuracy:
  - molecular weight calculation
  - elemental composition and molecular formula determination
  - molecular structure
  - Molecular ions of isobaric species (same *m/z* value but different molecular formula and structure) could be distinguished

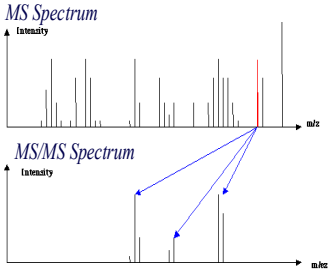


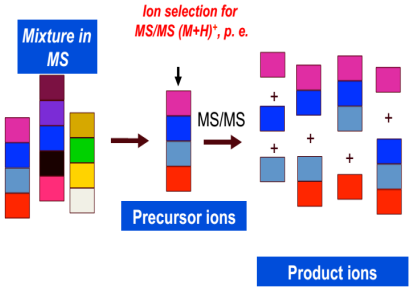
Cold Spring Harb Perspect Biol. 2011 Sep; 3(9): a004614. doi:†10.1101/cshperspect.a004614

Tandem Mass Spectrometry (MS/MS) data analysis

Fragmentation:

- Selection of ion of interest in MS
- Formation of fragment ions in MS/MS
- Structural information





The interpretation of the MS/MS spectrum is like solving a puzzle

↓

Allows us to obtain structural information about the initial compound.

Tandem mass spectrometry (MS/MS)

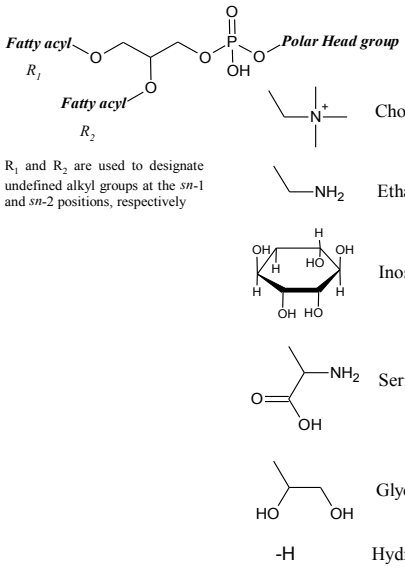
Glycerophospholipids or phospholipids (PL)

Information needed to be confirm:

- Polar head group
- Fatty acids

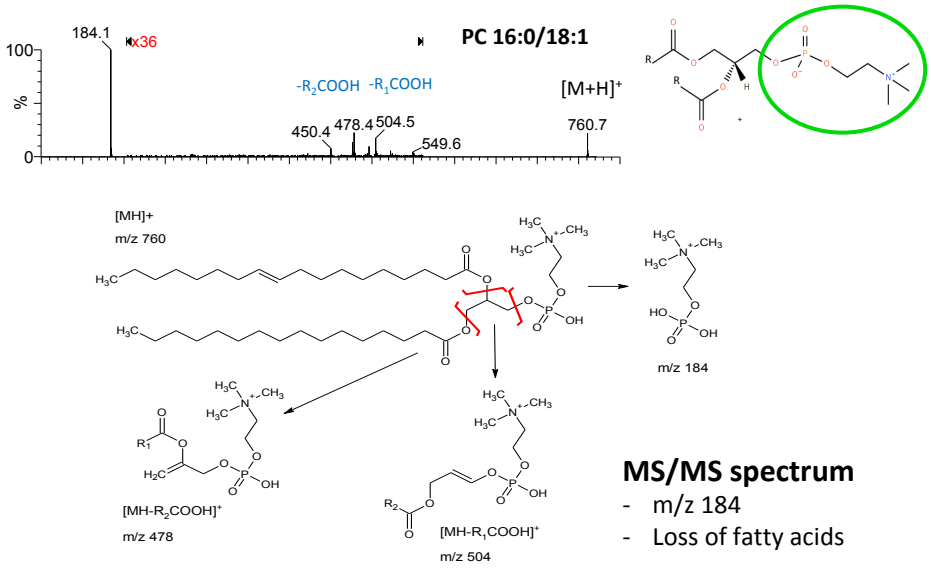
Fragmenation depends on:

- Type of percursor ion
- Colision energy
- others



R_1 and R_2 are used to designate undefined alkyl groups at the *sn*-1 and *sn*-2 positions, respectively

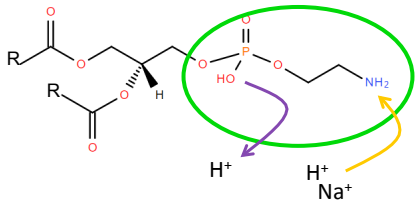
Phosphatidylcholine– MS/MS $[M+H]^+$



MS/MS spectrum

- m/z 184
- Loss of fatty acids

Phosphatidylethanolamine– PE



- Positive Mode $[M+H]^+$
- Negative Mode $[M-H]^-$

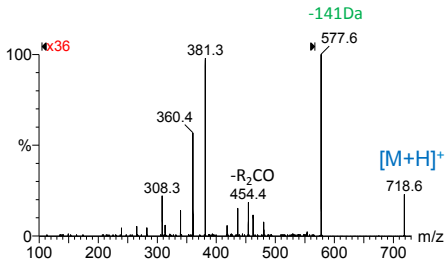
ESI-MS/MS $[M+H]^+$

Characteristic loss of 141 Da

Loss of RCOOH and RC=O

(R_1=CO^+ and R_2=CO^+).

PE 16:0/18:1

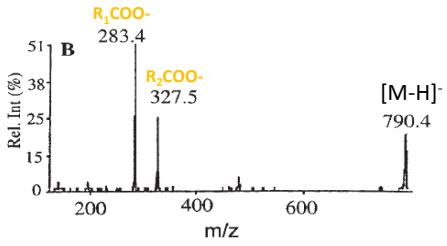


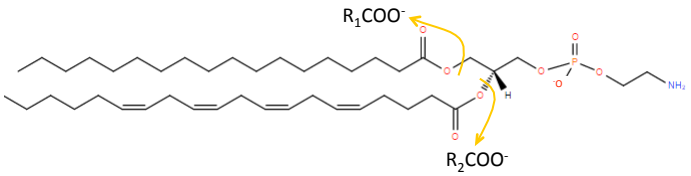
Phosphatidylethanolamine – MS/MS

ESI-MS/MS $[M-H]^-$

MS/MS spectrum in negative mode

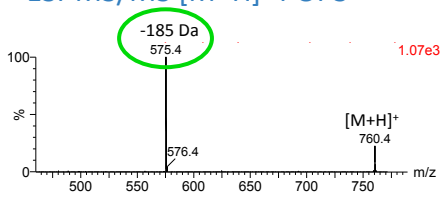
- R_1COO^- ion
- R_2COO^- ion

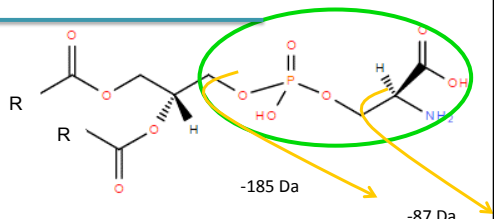




Phosphatidylserines – PS

ESI-MS/MS [M+H]⁺ POPS



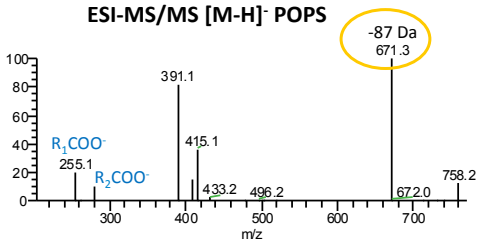


MS/MS [M+H]⁺ **Characteristic loss of 185 Da**

ESI-MS/MS of [M-H]-

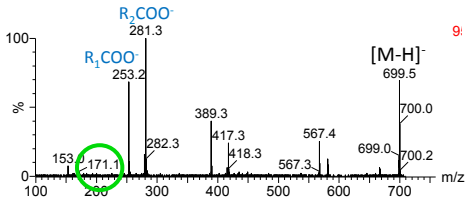
- Loss of 87 Da
- R_1COO^- ion
- R_2COO^- ion

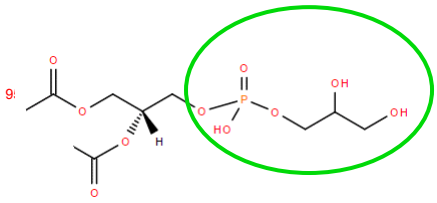
ESI-MS/MS [M-H]⁻ POPS

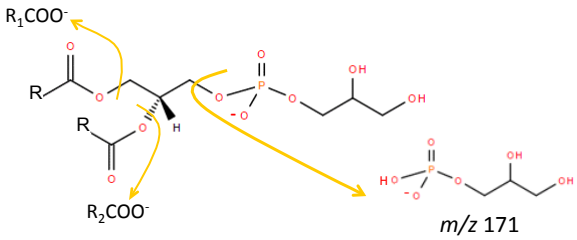


Phosphatidylglycerol- PG

Espectro de ESI-MS/MS in negative mode

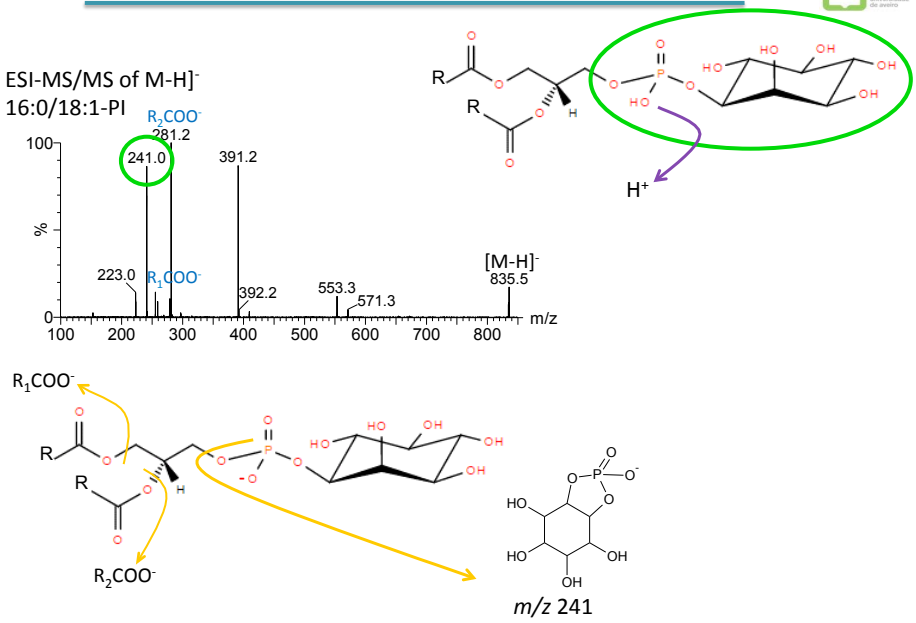






m/z 171

Phosphatidylinositol -PI



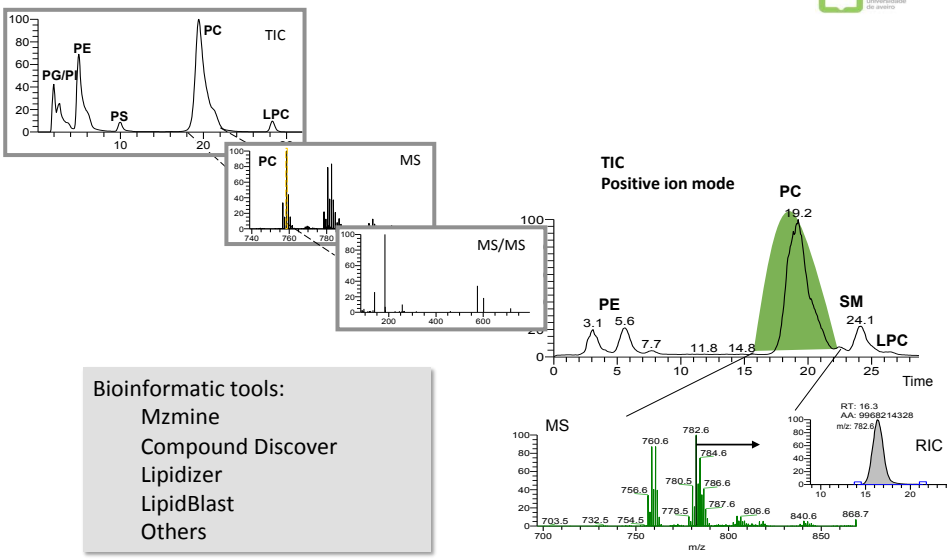
Phospholipid classes and MS/MS

	Positive ion	Negative ion
Headgroup		
Phosphatidylcholine	Precursor ion *m/z* 184	–
Phosphatidylserine	Neutral loss 185 Da	Neutral loss 87 Da
Phosphatidylethanolamine	Neutral loss 141 Da	–
Phosphatidylinositol	–	Precursor ion *m/z* 241
Sphingomyelin	Precursor ion *m/z* 184	

Information about fatty acyl composition

- Positive mode
  - Loss of RCOOH e R=C=O
- Negative mode
  - Formation of carboxylate anions $RCOO^-$

Untarget Lipidomics analysis

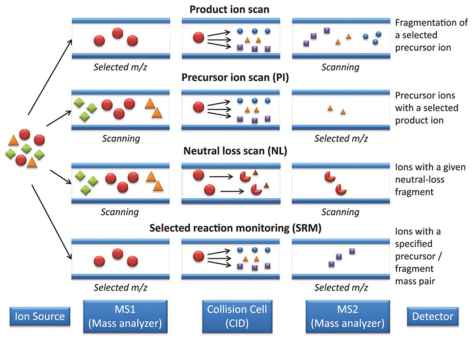


Bioinformatic tools:
- Mzmine
- Compound Discover
- Lipidizer
- LipidBlast
- Others

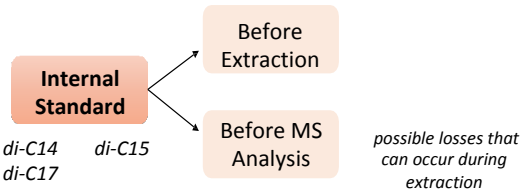
Target Lipidomics analysis

Shotgun Lipidomics

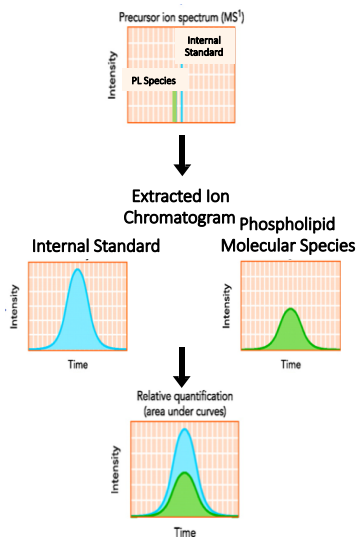
	Positive ion	Negative ion
Headgroup		
Phosphatidylcholine	Precursor ion *m/z* 184	–
Phosphatidylserine	Neutral loss 185 Da	Neutral loss 87 Da
Phosphatidylethanolamine	Neutral loss 141 Da	–
Phosphatidylinositol	–	Precursor ion *m/z* 241
Sphingomyelin	Precursor ion *m/z* 184	



Quantification by LC MS



possible losses that can occur during extraction



- Different lipid species have different quantitative responses
- One internal standard for each one phospholipid class

Normalization of each molecular species to the internal standard of the corresponding class

This project has been funded with support from the European Commission.

This publication reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein

Mass spectrometry-based proteomics

Pedro Domingues
Rosário Domingues
Rita Ferreira
Tânia Melo
Eliana Alves
Elizabete Maciel

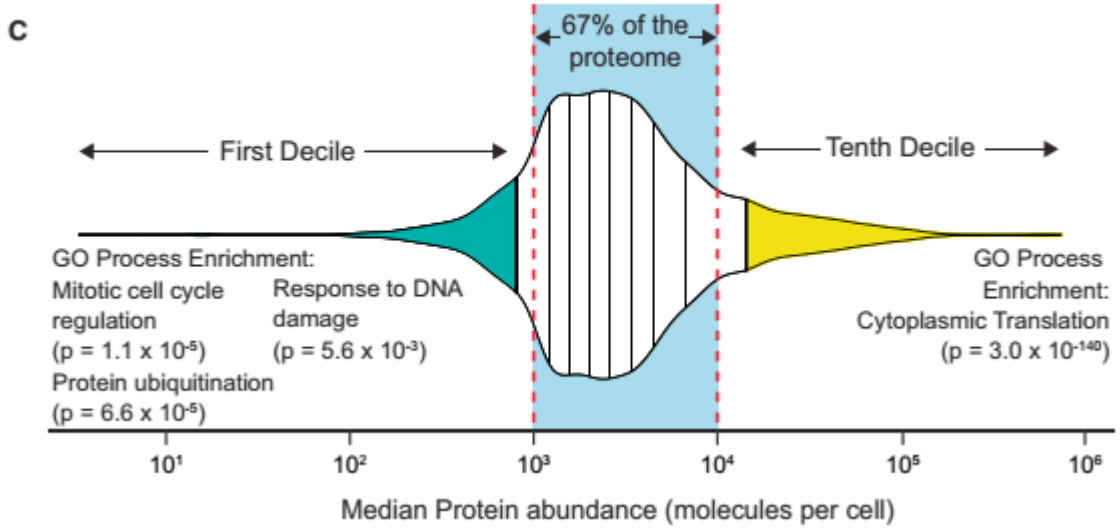
universidade de aveiro

AACLifeSci

Erasmus+

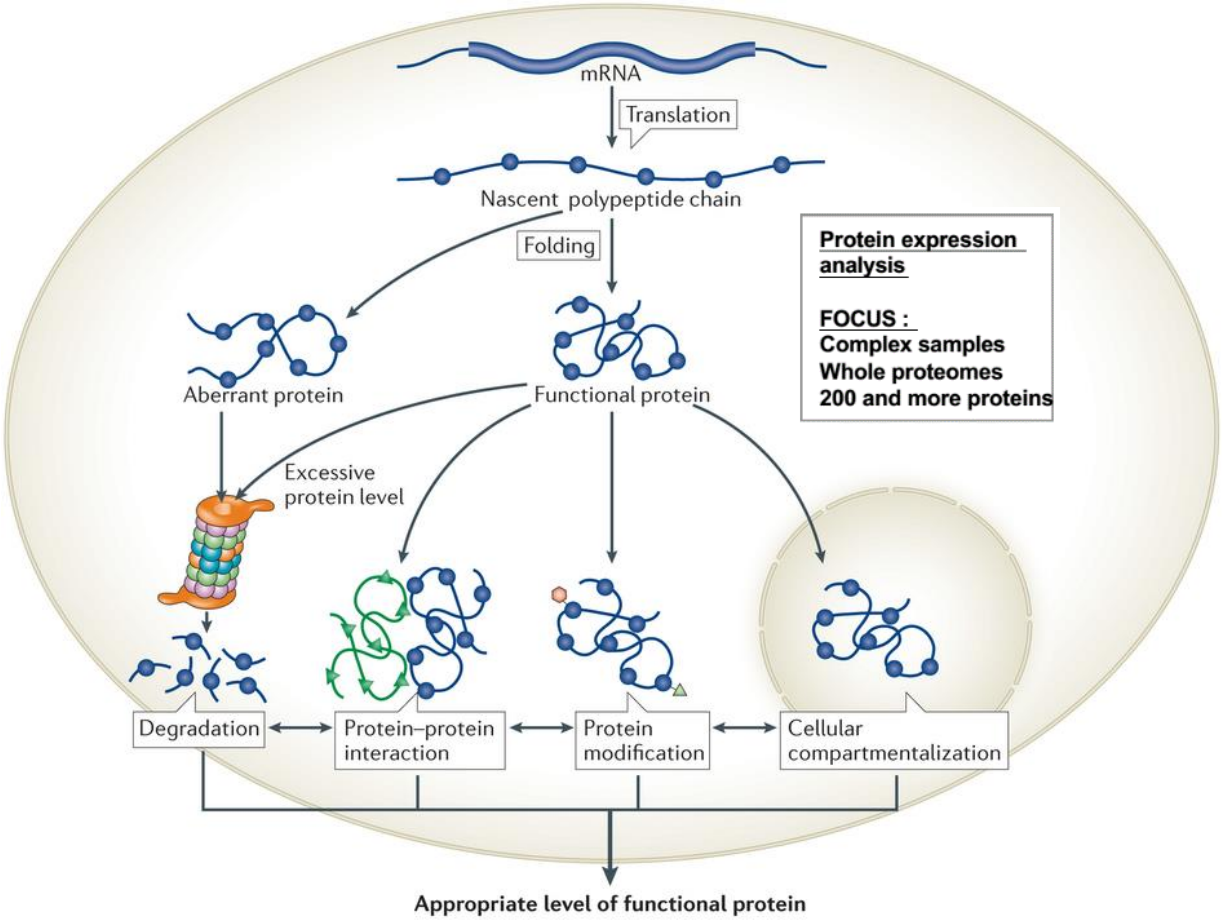
Proteome

S. Serevisiae
Number of proteins (proteome): 5858
Total of proteins/cell: 42 million



Ho et al., 2018, Cell Systems 6, 1–14
February 28, 2018 © 2017 Elsevier Inc.
https://doi.org/10.1016/j.cels.2017.12.004

Proteomics



Protein expression analysis

FOCUS :
Complex samples
Whole proteomes
200 and more proteins

Interaction / Functional Proteomics

FOCUS :
Subcellular fraction
Organelle
Protein Complex

1-200 proteins

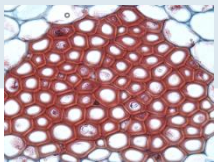
PTM analysis

FOCUS :
Single protein

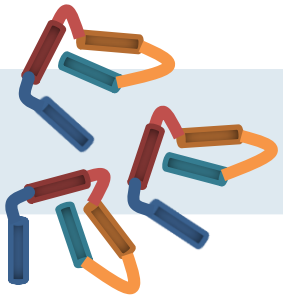
1-20 proteins

Nature Reviews | Molecular Cell Biology

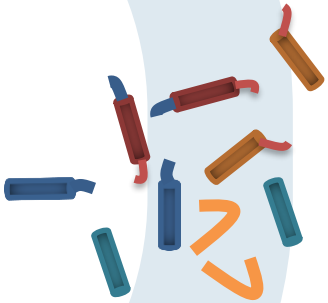
Proteomics



Sample
Tissue or cell culture



Extraction of proteins and purification

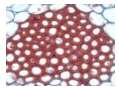


Digestion of protein and purification of peptides

Analysis by LC-MS

Analysis of data and automation

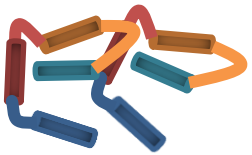
Mass spectrometry-based proteomics



SAMPLE

Cell breakage
- Sonication
- Grinding
- Lysis

Subcellular fractioning
- Density gradient centrifugation



Protein solubilization and stabilization
- buffer
- protease inhibitors

Protein purification
- Ion exchange chromatography
- Hydrophobic interaction
- Affinity chromatography

Protein assay
- Lowry protein assay or RC-DC
- BCA (bicinchoninic acid assay)

1D SDS-PAGE

2D SDS-PAGE

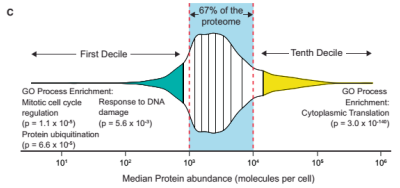
Staining of gel
- Sypro
- Silver
- Coomassie Blue

Digestion of protein
- Trypsin
- Lys-C

Spot quantification
- Melanie
- ImageQuant

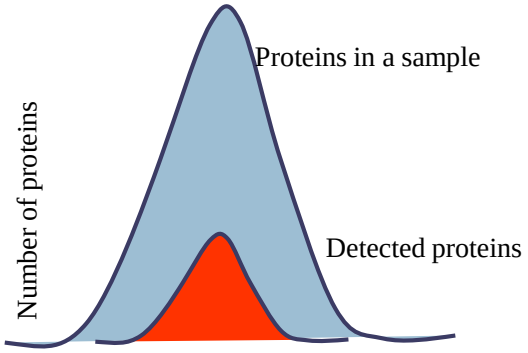
MS

Multistep Liquid Chromatography
- Ion exchange chromatography
- Reversed-phase liquid chromatography



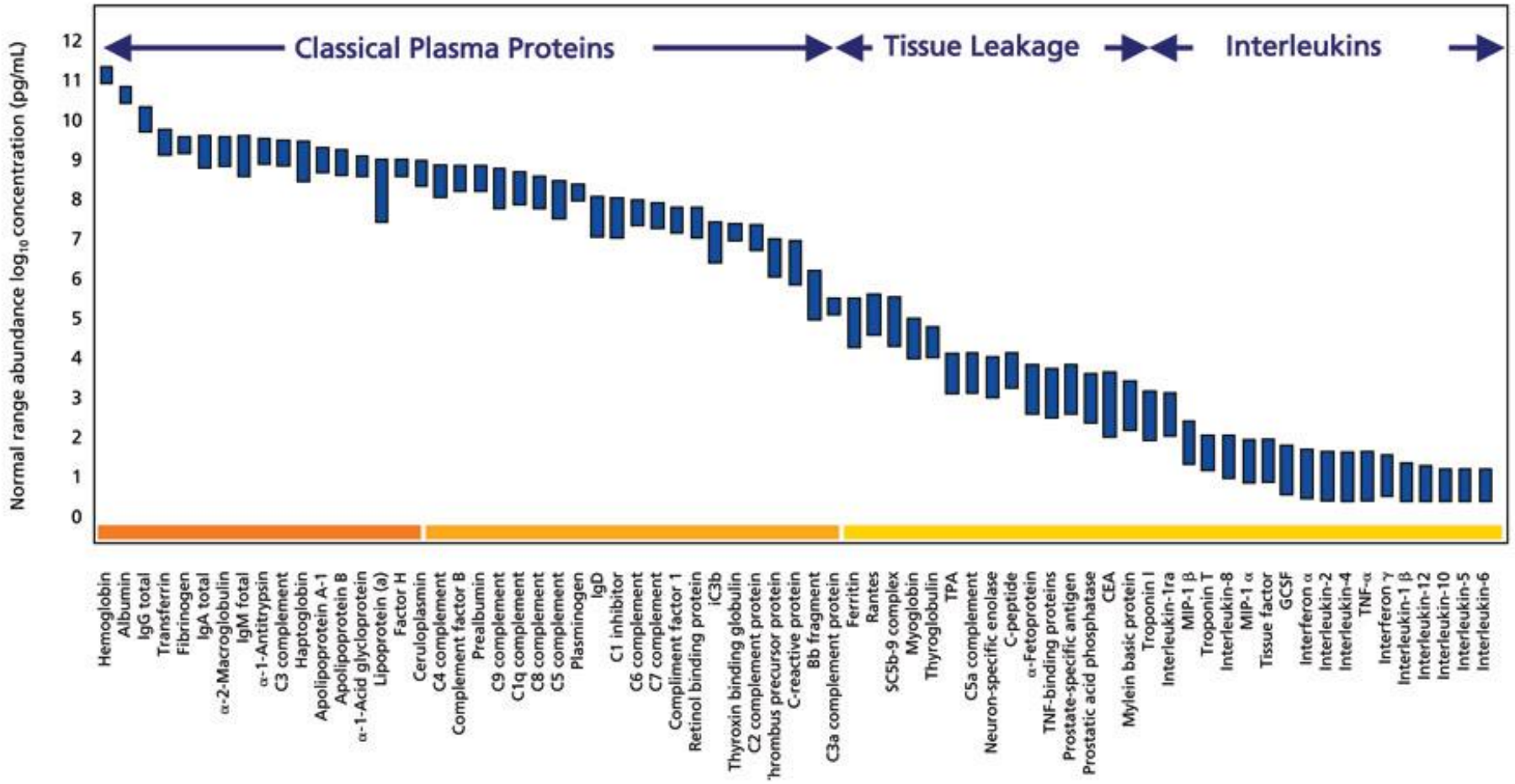
Sensitivity

- **Sensitivity**
  - System response on a calibration curve
- **Detection limit**
  - Signal to noise ratio (S/N)
  - Depends on the matrix
- In Omics:

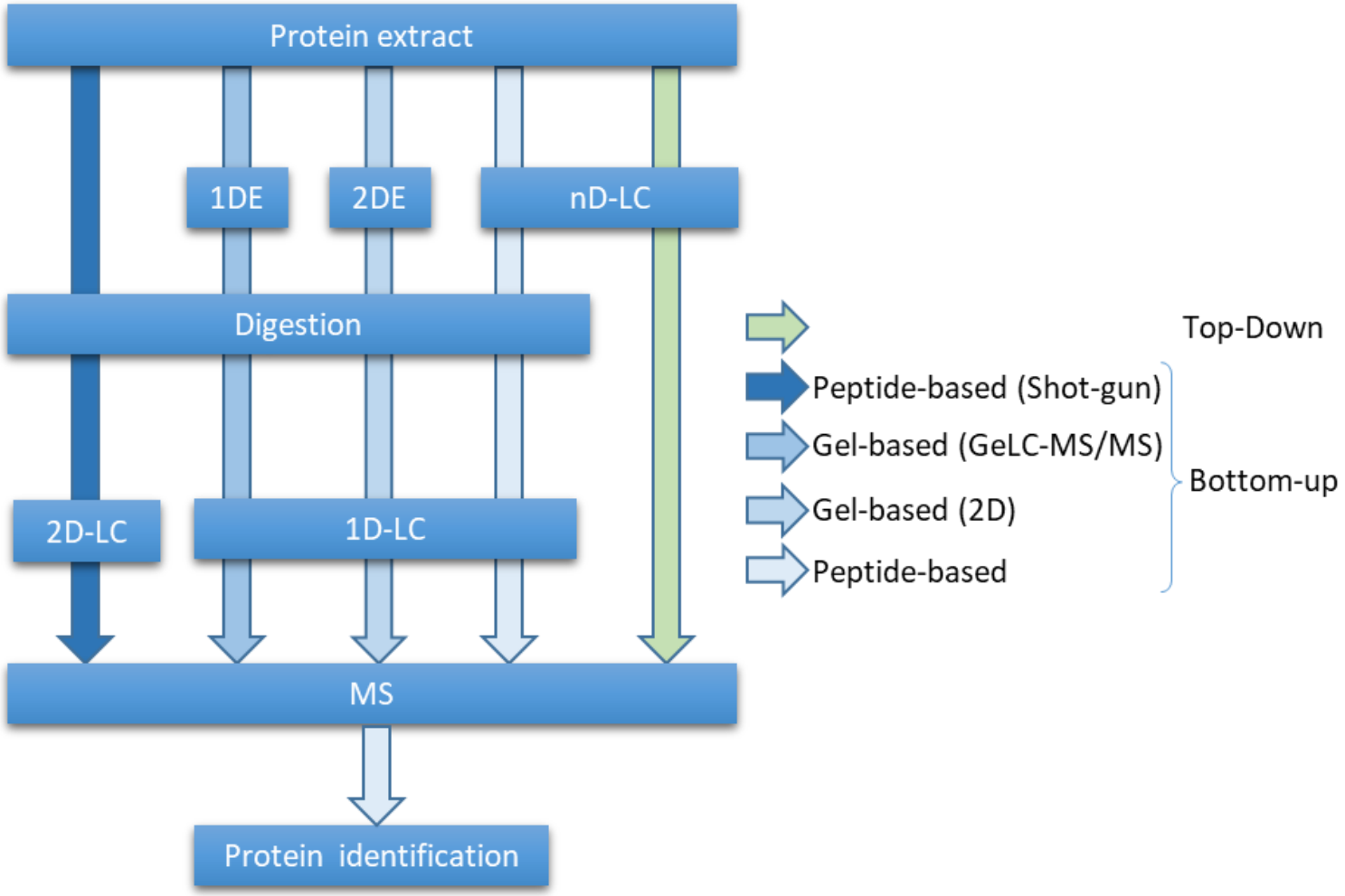


$$\text{Sensitivity} = \frac{\#\text{ of true results not rejected}}{\text{total }\#\text{ of true}}$$

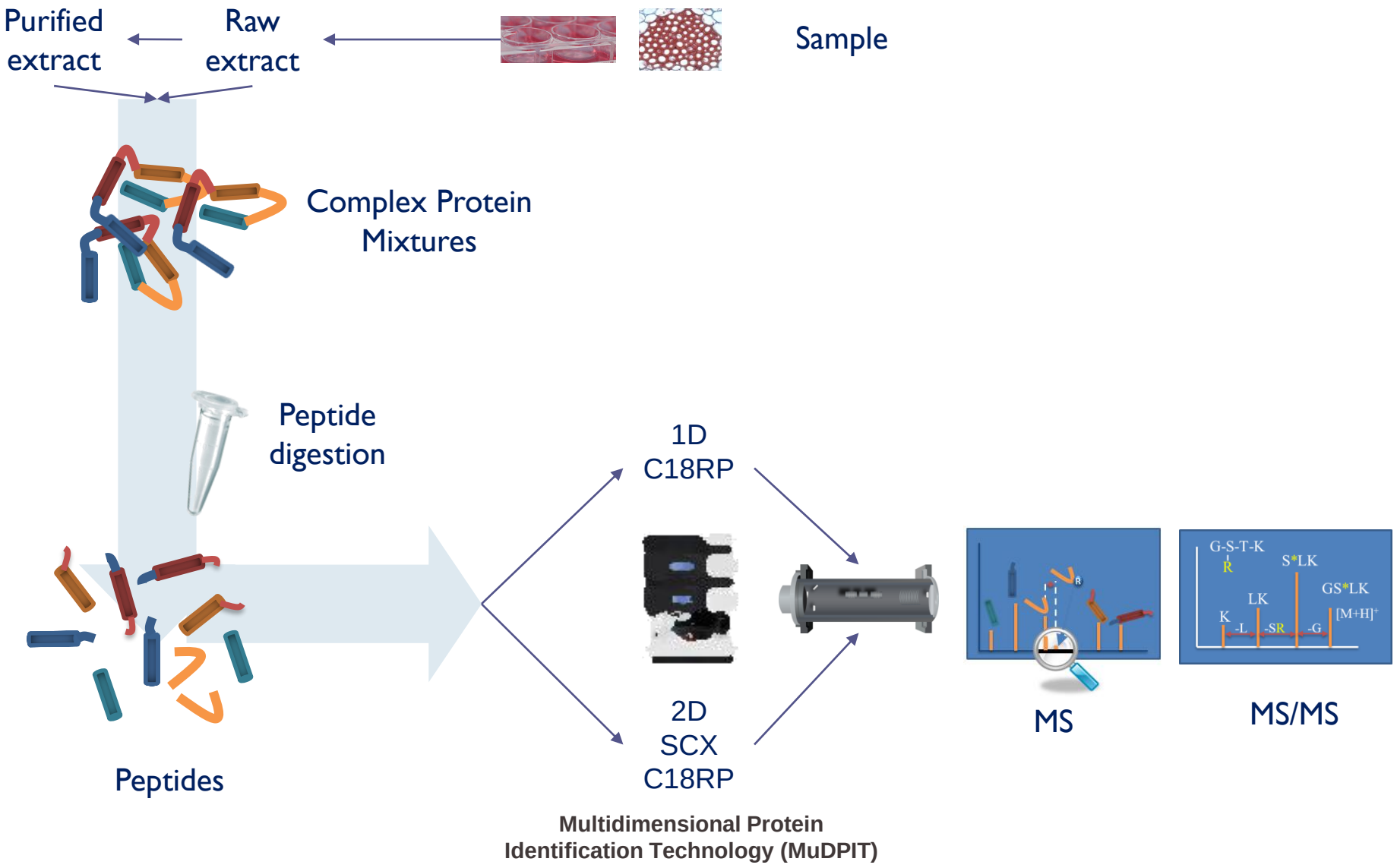
Proteins in plasma



Peptide-based vs protein-based approaches

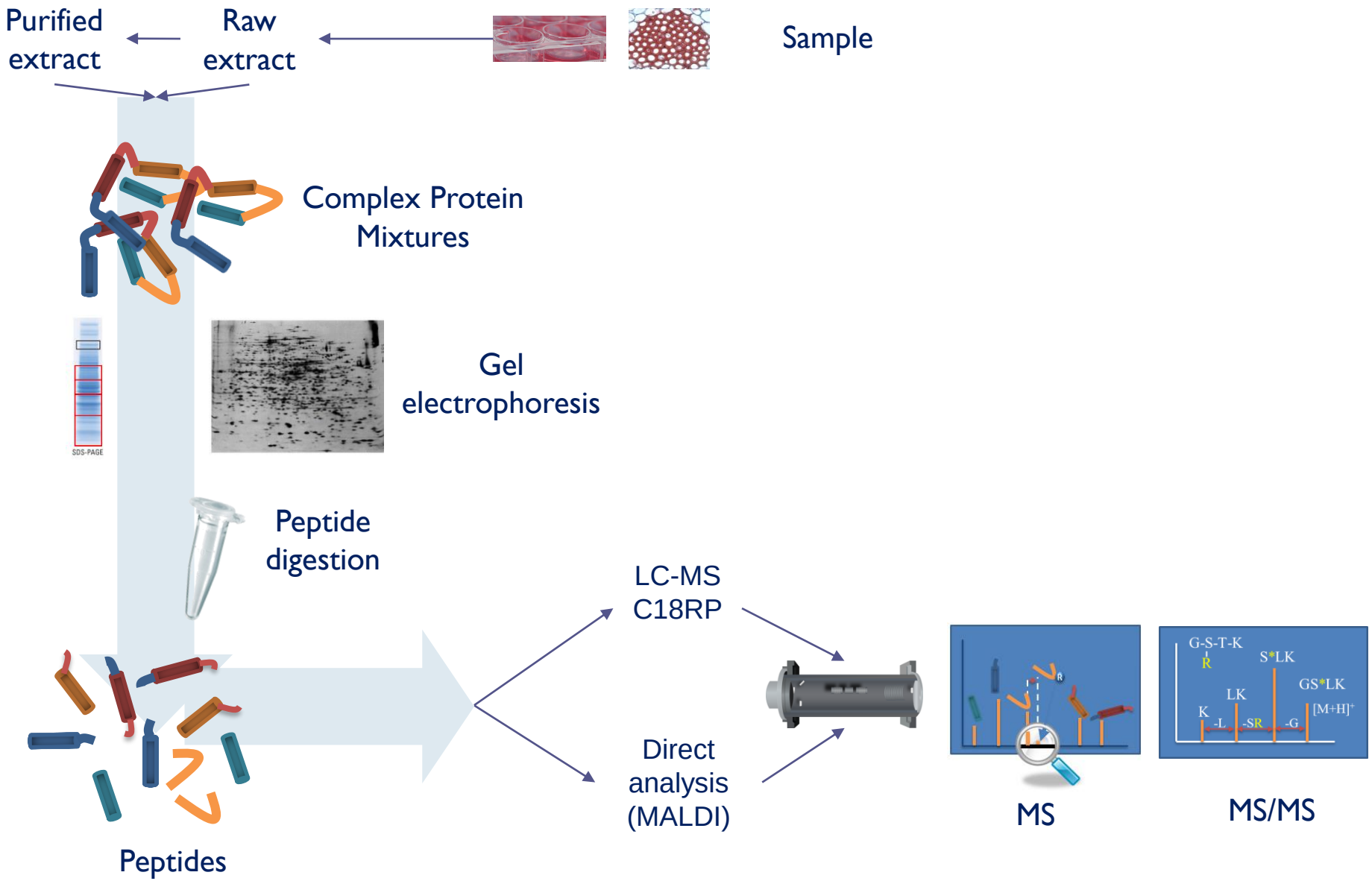


Peptide-based approach (shotgun)

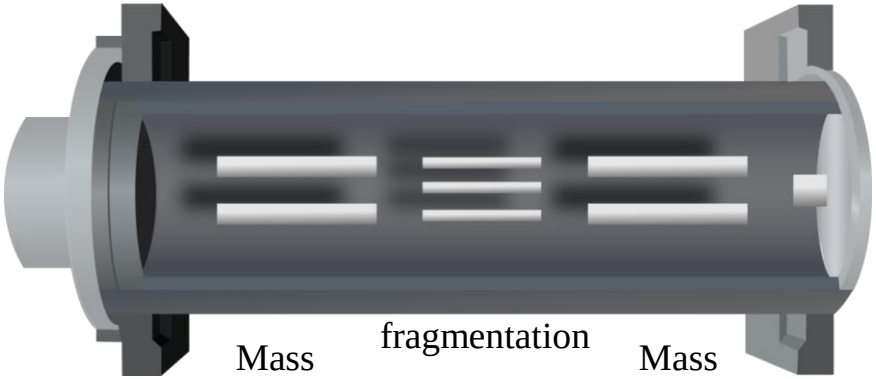


Multidimensional Protein Identification Technology (MuDPIT)

Protein-based approch (Electrophoresis)



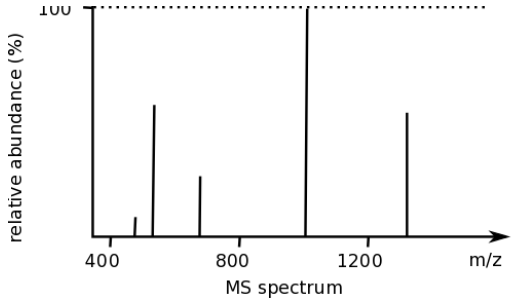
MS-based protein identification : concept

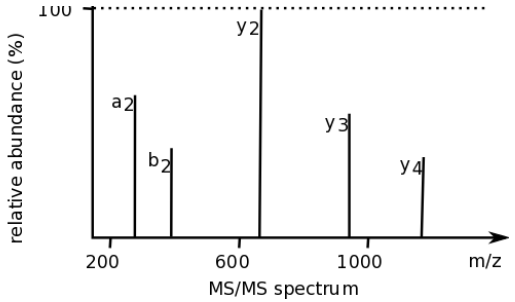


fragmentation

Mass analyser

Mass analyser

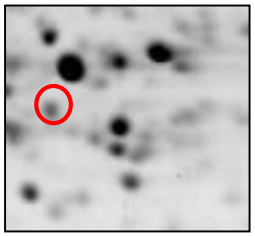




PMF - Peptide Mass Fingerprinting

PFF - Peptide Fragment Fingerprinting

MS-based protein identification : PMF concept



Protein sample

Protein sequence(s)

Sequence database

Protease digestion

In silico digestion

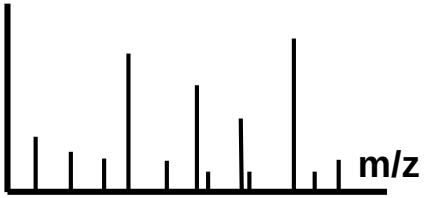
Protein fragments

Protein fragment sequences

MS

software

software

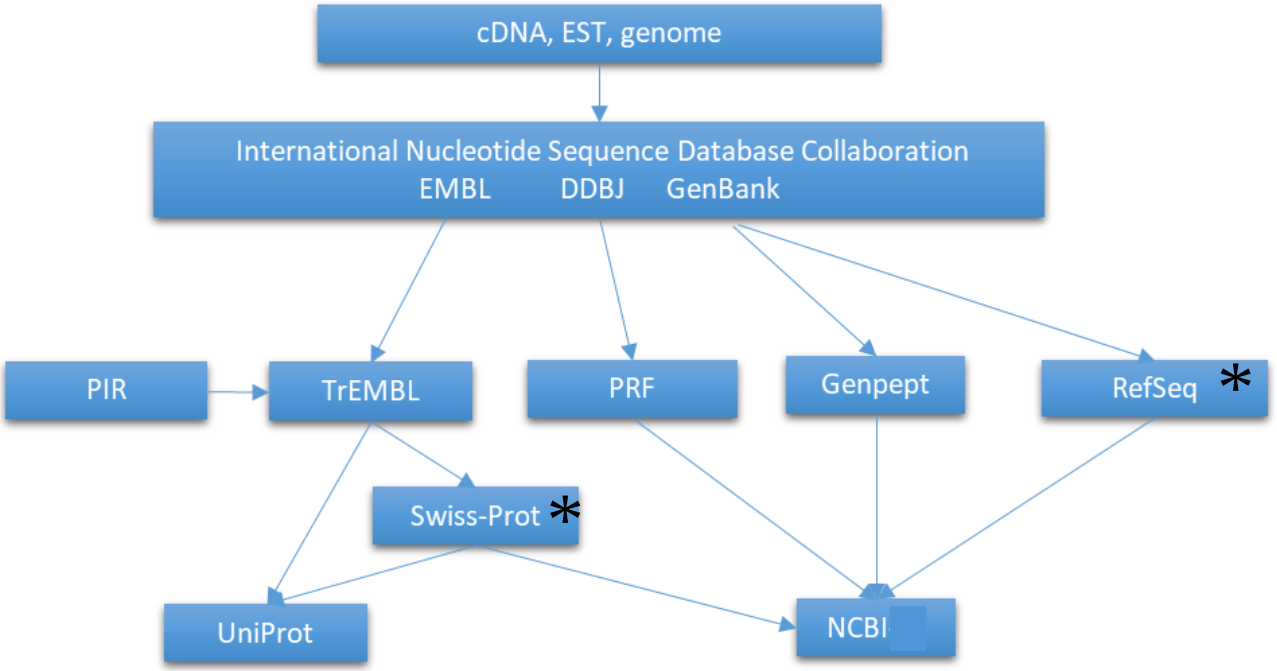


Experimental MS data

Best Match(es)

Theoretical MS data

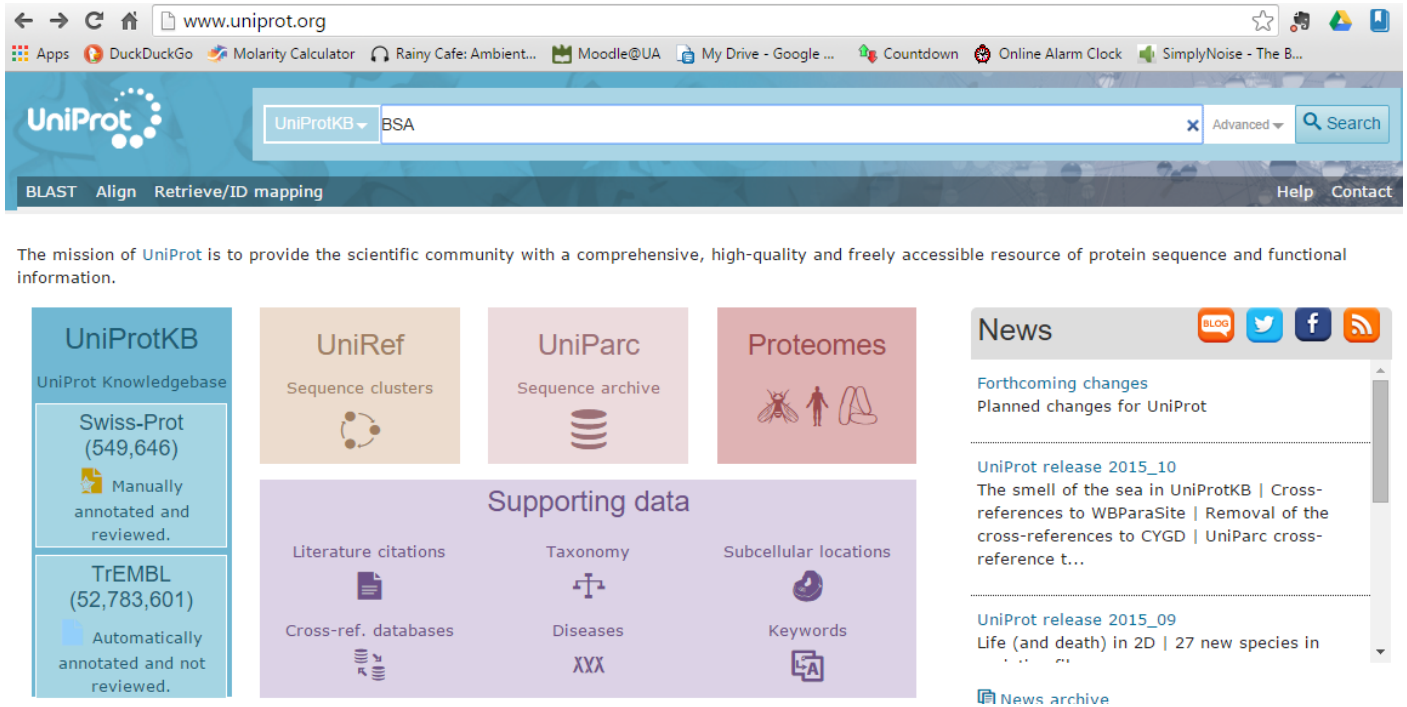
Protein Sequence Databases



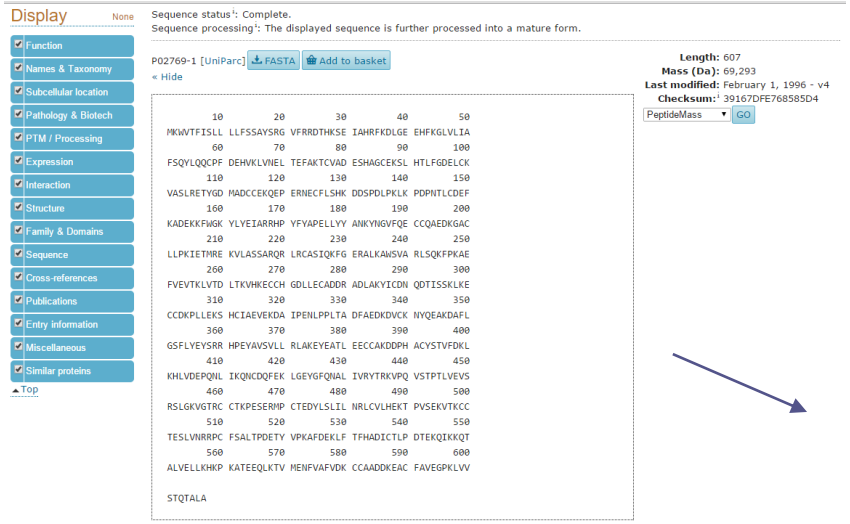
UniProt: Swiss-Prot + TrEMBL + (PIR)
NCBI-nr: Swiss-Prot + GenPept + (PIR) + RefSeq + PDB + PRF

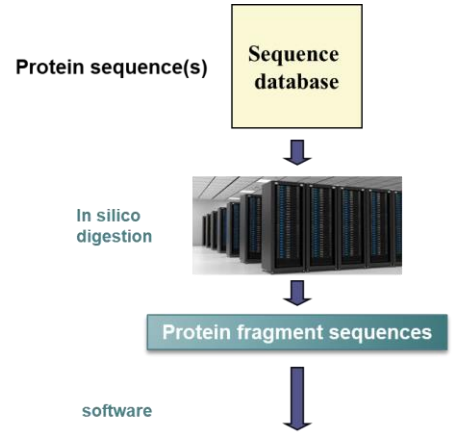
* curated non-redundant protein database

Protein Sequence Databases - UniProt



Protein digestion (in silico)





ExPASy Bioinformatics Resource Portal — PeptideMass

PeptideMass

The entered protein is: P02769

The selected enzyme is: Trypsin

Maximum number of missed cleavages (MC): 2

All cysteines in reduced form.

Methionines have not been oxidized.

Displaying peptides with a mass bigger than 500 Dalton.

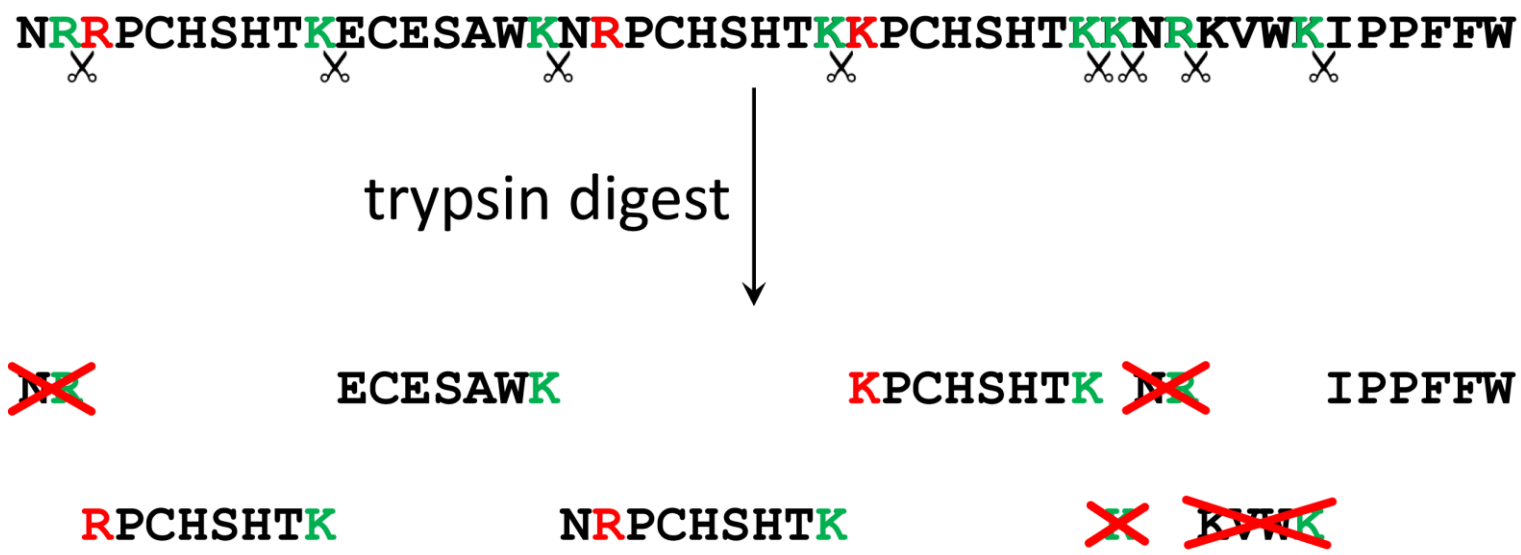
Using monoisotopic masses of the occurring amino acid residues and giving peptide masses as $[M+H]^+$.

You have selected ALBU_BOVIN (P02769) from UniProtKB/Swiss-Prot:

Serum albumin precursor (BSA) (Allergen Bos d 6)
Signal and propep in positions 1-24 have been removed.

- Chain Serum albumin at positions 25 - 607 [Theoretical pI: 5.60 / Mw (average mass): 66432.96 / Mw (monoisotopic mass): 66389.86]

mass	position	#MC	modifications		peptide sequence
4910.3837	45-88	2	PHOS: 82	4990.3500	GLVLIAFSQYLQQCPFDEHV KLVNELTEFAKTCVADESHA GCEK
4535.2954	37-75	2			DLGEEHFKGLVLIAFSQYLQ QCPFDEHVKLVNELTEFAK
4246.0511	508-544	2	PHOS: 512	4326.0174	RPCFSALTPDETYVPKAFDE KLFTFHADICTLPDTEK
4185.9547	169-204	2			HPYFYAPELLYYANKYNGVF QECCQAEDKGACLLPK
4110.9707	300-336	2			ECCDKPLLEKSHCIAEVEKD AIPENLPPLTADFAEDK
3837.8131	66-100	2	PHOS: 82, 89	3997.7458	LVNELTEFAKTCVADESHAG CEKSLHTLFGDELCK
3758.8958	402-433	2			HLVDEPQNLIKQNCDQFEKL GEYGFQNALIVR
3665.8460	35-65	2			FKDLGEEHFKGLVLIAFSQY LQQCPFDEHVK
3659.6722	168-197	2			RHPYFYAPELLYYANKYNGV FQECCQAEDK
3579.8555	45-75	1			GLVLIAFSQYLQQCPFDEHV KLVNELTEFAK
3523.6874	460-489	2			CCTKPESERMPCTEDYLSLI LNRLCVLHEK
3503.5711	169-197	1			HPYFYAPELLYYANKYNGVF QECCQAEDK
3444.4810	267-297	2	PHOS: 296	3524.4473	ECCHGDLLECADDRADLAKY ICDNQDTISSK
3419.6068	499-528	2	PHOS: 512	3499.5731	CCTESLVNRRPCFSALTPDE TYVPKAFDEK
3397.6289	310-340	2			SHCIAEVEKDAIPENLPPLT ADFAEDKDVCK
3390.6826	37-65	1			DLGEEHFKGLVLIAFSQYLQ QCPFDEHVK



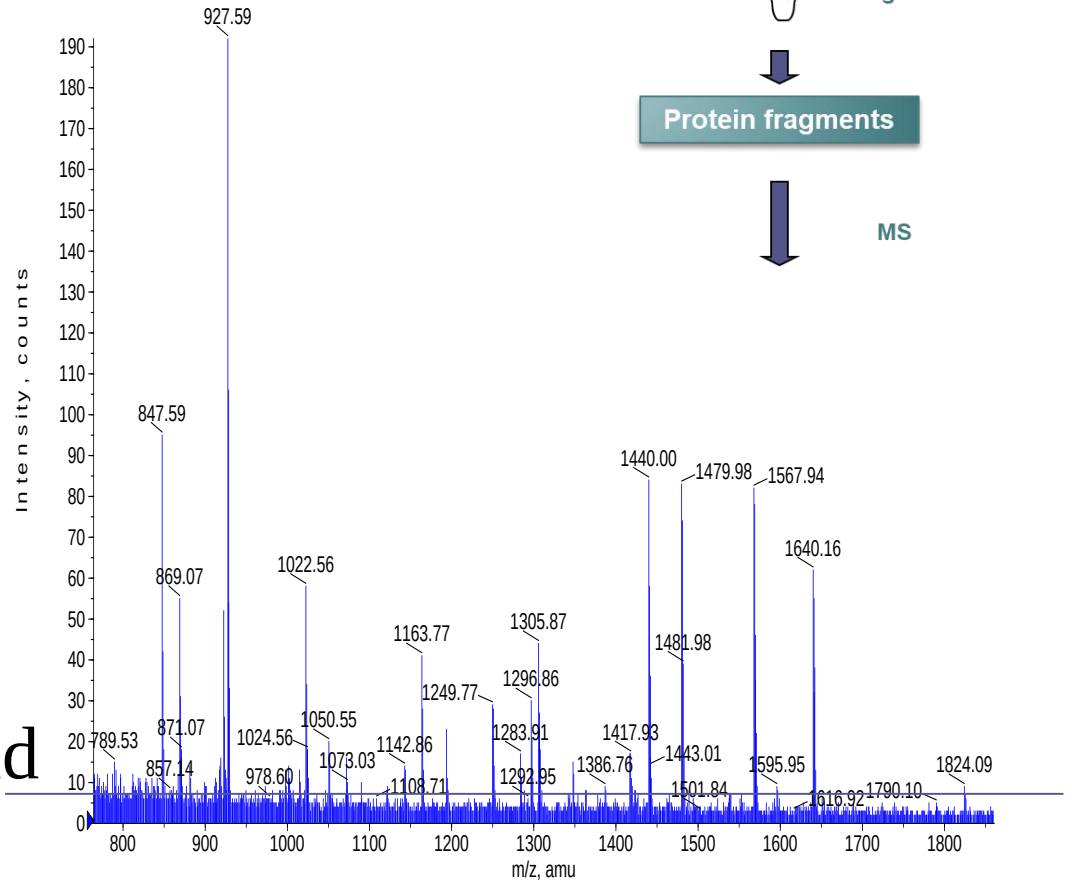
Trypsin predominantly cleaves proteins at the carboxyl side (or "C-terminal side") of the amino acids:
- lysine (K) and
- arginine (R)
- except when either is bound to a C-terminal proline (P).

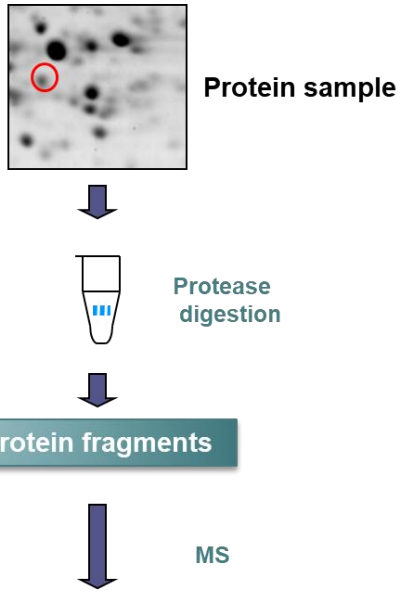
PMF (peptide mass fingerprinting)

Extracted peak list (m/z)

COM=10 pmol digest of Sample BSA
MASS=Monoisotopic
USERNAME=Pedro Domingues
USEREMAIL=p.domingues@ua.pt
TITLE= Cmpd 7, +MS, 16.8 min
847.50413
868.97220
922.46673
923.48150
927.49393
1022.45510
1050.45330
1163.63123
1164.65310
1193.60273
1249.62173
1250.71030
1296.75560
1297.74990
1305.71663
1416.79290
1439.81233
1479.79593
1482.75830
1567.74323
1639.93833
1823.86000

threshold



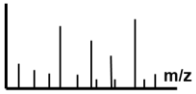


Identification search engines

- Mascot
- X!Tandem,
- MS-GF+,
- MS Amanda,
- MyriMatch,
- Comet,
- Tide,
- Andromeda
- OMSSA
- ProteinProspector
- Sequest
- PEAKS
- PRIDE
- ….

PMF (peptide mass fingerprinting)

Protein Search Engines



Experimental MS data

Best Match(es)

Theoretical MS data

MATRIX SCIENCE

Search this site

Home | Mascot database search | Products | Technical support | Training | News | Blog | Newsletter | Contact

Access Mascot Server | Database search help

Mascot database search > Access Mascot Server

Access Mascot Server

You are welcome to submit searches to this free Mascot Server. Searches of MS/MS data are limited to 1200 spectra and some functions, such as no enzyme searches, are unavailable. Automated searching of batches of files is not permitted. If you want to automate search submission, perform large searches, search additional sequence databases, or customise the modifications, quantitation methods, etc., you'll need to license your own, in-house copy of Mascot Server.

Peptide Mass Fingerprint

The experimental data are a list of peptide mass values from the digestion of a protein by a specific enzyme such as trypsin.

Perform search | Example of results report | Tutorial

Sequence Query

One or more peptide mass values associated with information such as partial or ambiguous sequence strings, amino acid composition information, MS/MS fragment ion masses, etc. A super-set of a sequence tag query.

Perform search | Example of results report | More information

MS/MS Ions Search

Identification based on raw MS/MS data from one or more peptides.

Perform search | Example of results report | Tutorial

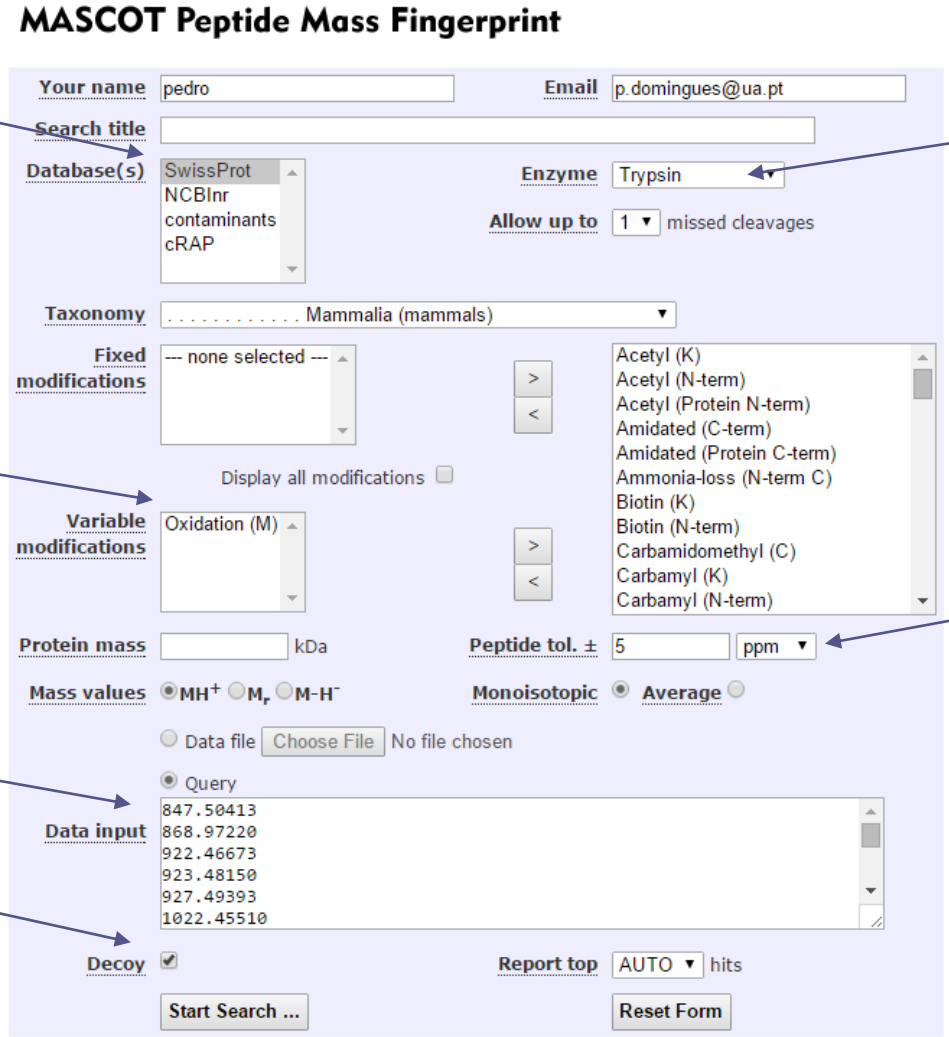
More info

- Mascot overview
- Search parameter reference
- Data file format
- Results report overview

PMF (peptide mass fingerprinting)

Extracted peak list (m/z)

847.50413
868.97220
922.46673
923.48150
927.49393
1022.45510
1050.45330
1163.63123
1164.65310
1193.60273
1249.62173
1250.71030
1296.75560
1297.74990
1305.71663
1416.79290
1439.81233
1479.79593
1482.75830
1567.74323
1639.93833
1823.86000



PMF (Mascot results)

Decoy

1. During the search, every time a protein sequence from the target database is tested, a decoy sequence of the same length is automatically generated and tested.
    1. Reverse sequence
    2. Random sequence
2. The matches and scores for the decoy sequences are recorded separately in the result file.
3. When the search is complete, the numbers of matches and the false discovery rate are reported in the result header.

Contaminants

Sequences for common contaminants, such as keratins, BSA, and trypsin.

Expect is the number of times we would expect to obtain an equal or higher score, purely by chance (smaller is better)

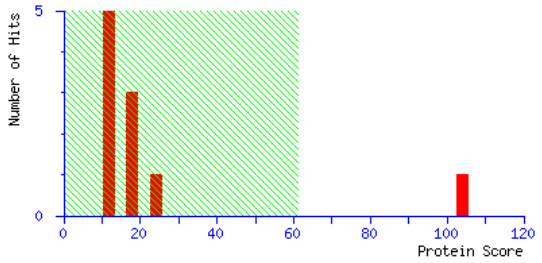
MATRIX SCIENCE Mascot Search Results

```
User            : pedro
Email           : p.domingues@ua.pt
Search title    :
Database        : SwissProt 2015_10 (549646 sequences; 195983064 residues)
Taxonomy        : Mammalia (mammals) (66401 sequences)
Timestamp       : 20 Oct 2015 at 11:43:21 GMT
Top Score       : 104 for ALBU_BOVIN, Serum albumin OS=Bos taurus GN=ALB PE=1 SV=4
```

	SwissProt	Decoy
Protein hits above identity threshold	1	0
Highest scoring protein hit	104	20

Mascot Score Histogram

Protein score is -10*Log(P), where P is the probability that the observed match is a random event.
Protein scores greater than 61 are significant (p<0.05).



Concise Protein Summary Report

Format As: Concise Protein Summary — Help

Significance threshold p< 0.05 — Max. number of hits AUTO

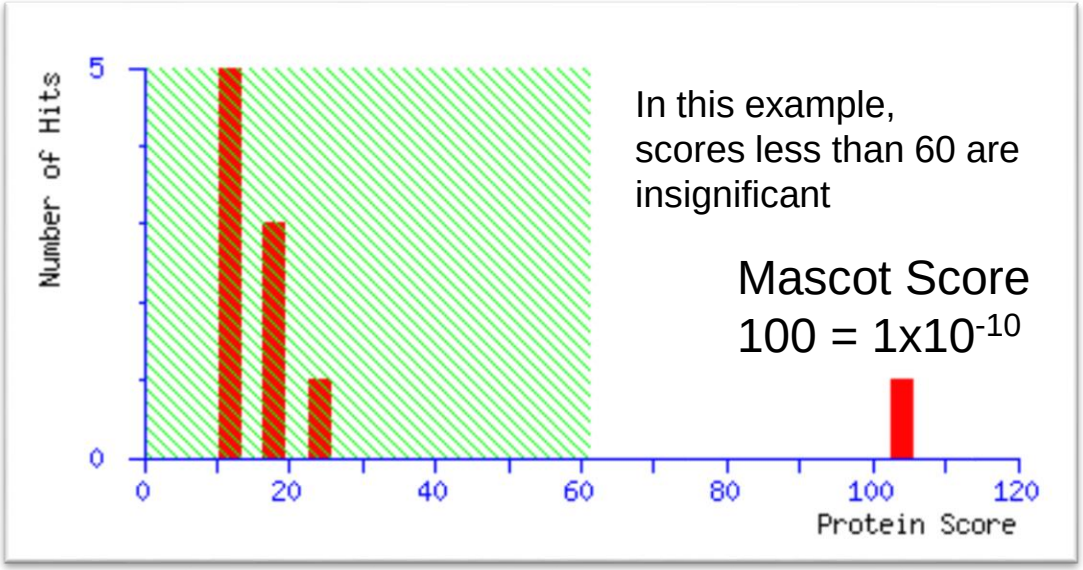
Preferred taxonomy: All entries

Re-Search All | Search Unmatched

```
1.  ALBU_BOVIN   Mass: 69248   Score: 104   Expect: 2.6e-06  Matches: 11
    Serum albumin OS=Bos taurus GN=ALB PE=1 SV=4
    ALBU_FELCA   Mass: 68615   Score: 24    Expect: 2.6e+02  Matches: 4
    Serum albumin OS=Felis catus GN=ALB PE=1 SV=1
    ALBU_CAPHI   Mass: 10048   Score: 24    Expect: 2.8e+02  Matches: 2
    Serum albumin (Fragments) OS=Capra hircus GN=ALB PE=1 SV=2
    ALBU_SHEEP   Mass: 69143   Score: 22    Expect: 4e+02    Matches: 4
```

PMF (Mascot Scoring)

- The Mascot Score is given as S = -10*Log(P), where
  - P is the probability that observed match is a random event
  - P=E*N-1
    - E=expect value
    - N=number of proteins in the database
- The significance of that result depends on the size of the database being searched. Mascot shades in green the insignificant hits using an E=0.05 cutoff



PMF (Mascot results)

MASCOT Search Results

Protein View: ALBU_BOVIN

Serum albumin OS=Bos taurus GN=ALB PE=1 SV=4

Database:	SwissProt
Score:	104
Expect:	2.6e-06
Nominal mass (M_r):	69248
Calculated pI:	5.82
Taxonomy:	Bos taurus

Sequence similarity is available as an NCBI BLAST search of ALBU_BOVIN against nr.

Search parameters

Enzyme:	Trypsin: cuts C-term side of KR unless next residue is P.
Variable modifications:	Oxidation (M)
Mass values searched:	22
Mass values matched:	11

Protein sequence coverage: 18%

Matched peptides shown in ***bold red***.

```
   1 MKWVTFISLL LLFSSAYSRG VFRRDTHKSE IAHRFKDLGE EHFKGLVLIA
  51 FSQYLQQCPF DEHVKLVNEL TEFAKTCVAD ESHAGCEKSL HTLFGDELCK
 101 VASLRETYGD MADCCEKQEP ERNECFLSHK DDSPDLPKLK PDPNTLCDEF
 151 KADEKKFWGK YLYEIARRHP YFYAPELLYY ANKYNGVFQE CCQAEDKGAC
 201 LLPKIETMRE KVLASSARQR LRCASIQKFG ERALKAWSVA RLSQKFPKAE
 251 FVEVTKLVTD LTKVHKECCH GDLLECADDR ADLAKYICDN QDTISSKLKE
 301 CCDKPLLEKS HCIAEVEKDA IPENLPPLTA DFAEDKDVCK NYQEAKDAFL
 351 GSFLYEYSRR HPEYAVSVLL RLAKEYEATL EECCAKDDPH ACYSTVFDKL
 401 KHLVDEPQNL IKQNCDQFEK LGEYGFQNAL IVRYTRKVPQ VSTPTLVEVS
 451 RSLGKVGTRC CTKPESERMP CTEDYLSLIL NRLCVLHEKT PVSEKVTKCC
 501 TESLVNRRPC FSALTPDETY VPKAFDEKLF TFHADICTLP DTEKQIKKQT
 551 ALVELLKHKP KATEEQLKTV MENFVAFVDK CCAADDKEAC FAVEGPKLVV
 601 STQTALA
```

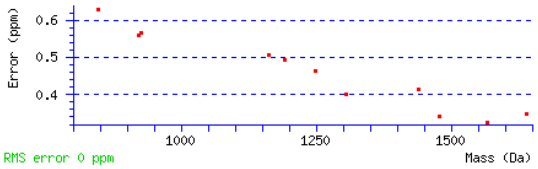
Unformatted sequence string: 607 residues (for pasting into other applications).

Sort peptides by ◉ Residue Number ○ Increasing Mass ○ Decreasing Mass

Show predicted peptides also

Start - End	Observed	Mr(expt)	Mr(calc)	ppm	M	Peptide
25 - 34	1193.6027	1192.5955	1192.5949	0.49	1	R.DTHKSEIAHR.F
35 - 44	1249.6217	1248.6145	1248.6139	0.46	1	R.FKDLGEEHFK.G
66 - 75	1163.6312	1162.6240	1162.6234	0.50	0	K.LVNELTEFAK.T
161 - 167	927.4939	926.4867	926.4861	0.57	0	K.YLYEIAR.R
205 - 211	922.4667	921.4595	921.4589	0.56	1	K.IETMREK.V + Oxidation (M)
242 - 248	847.5041	846.4969	846.4963	0.63	1	R.LSQKFPK.A
347 - 359	1567.7432	1566.7360	1566.7354	0.33	0	K.DAFLGSFLYEYSR.R
360 - 371	1439.8123	1438.8051	1438.8045	0.41	1	R.RHPEYAVSVLLR.L
402 - 412	1305.7166	1304.7094	1304.7088	0.40	0	K.HLVDEPQNLIK.Q
421 - 433	1479.7959	1478.7887	1478.7881	0.34	0	K.LGEYGFQNALIVR.Y
437 - 451	1639.9383	1638.9311	1638.9305	0.35	1	R.KVPQVSTPTLVEVSR.S

No match to: 868.9722, 923.4815, 1022.4551, 1050.4533, 1164.6531, 1250.7103, 1296.7556, 1297




```
AC   P02769; A5PJX3; O02787; P04277; Q3SZR2;
DT   21-JUL-1986, integrated into UniProtKB/Swiss-Prot.
```

PMF (peptide mass fingerprinting)

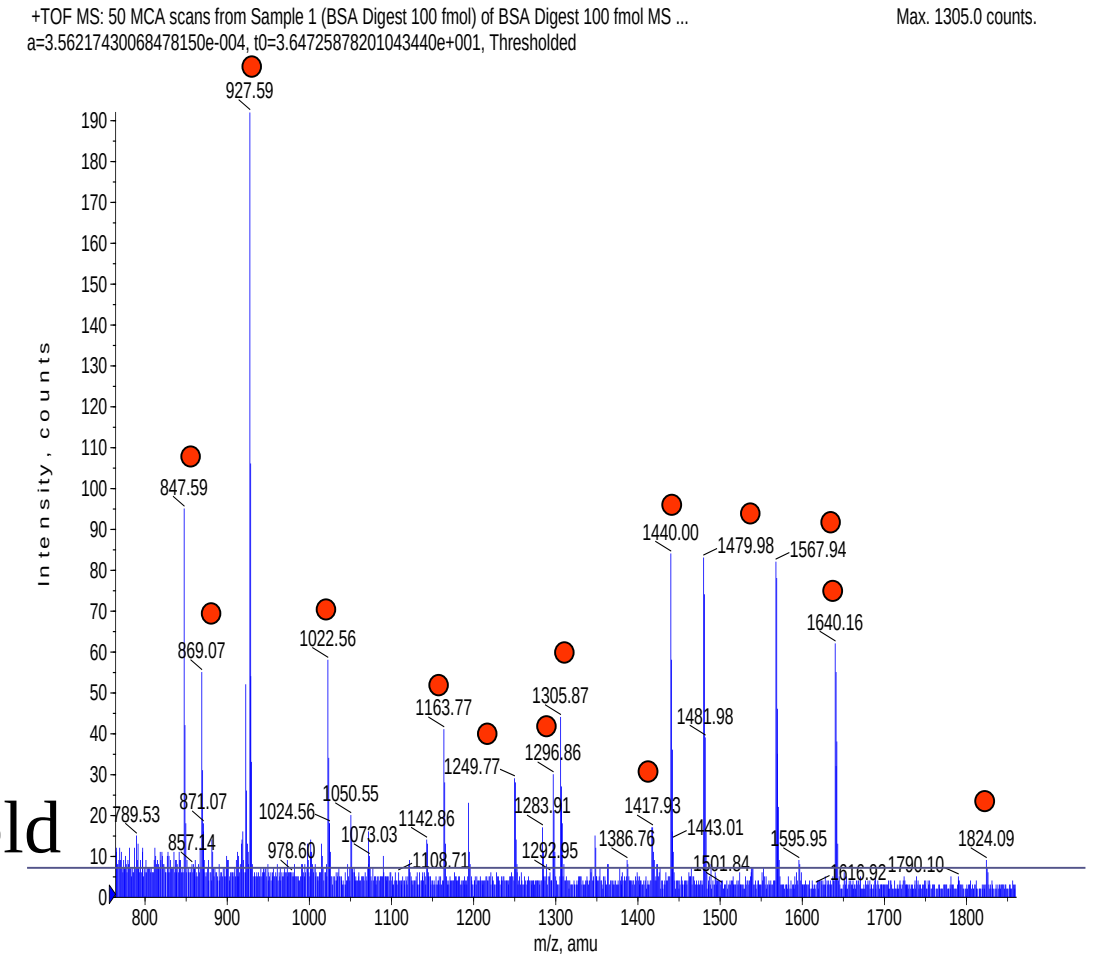
Extracted peak list (m/z)

Mass values searched: 22

Mass values matched: 11

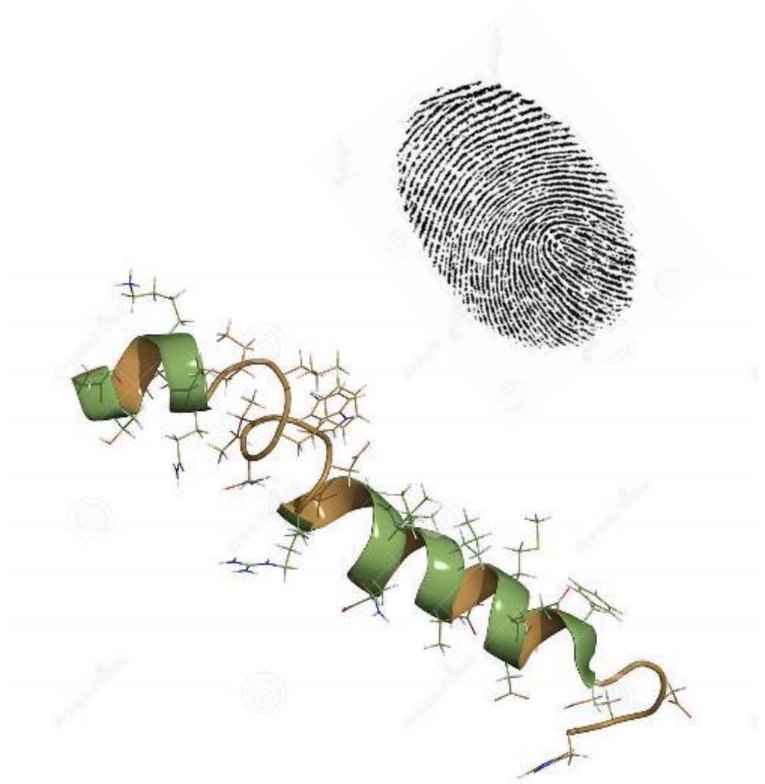
COM=10 pmol digest of Sample BSA
MASS=Monoisotopic
USERNAME=Pedro Domingues
USEREMAIL=p.domingues@ua.pt
TITLE= Cmpd 7, +MS, 16.8 min
847.50413
868.97220
922.46673
923.48150
927.49393
1022.45510
1050.45330
1163.63123
1164.65310
1193.60273
1249.62173
1250.71030
1296.75560
1297.74990
1305.71663
1416.79290
1439.81233
1479.79593
1482.75830
1567.74323
1639.93833
1823.86000

threshold

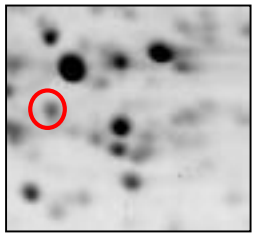


PMF (peptide mass fingerprinting)

- Fast, simple analysis
- High sensitivity
- Not good for mixtures
  - especially a minor component.
- High mass accuracy is necessary.



MS-based protein identification : PFF concept



Protein sample

Protein sequence(s)

Sequence database

Protease digestion (Trypsin)

software

Protein fragments

Protein fragment sequences

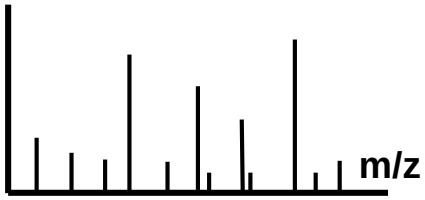
MS
MS/MS

software

Calculated exact masses of peptides

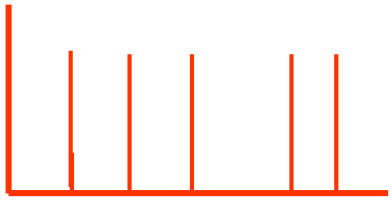
Calculated fragmentation spectrum of each peptide

software



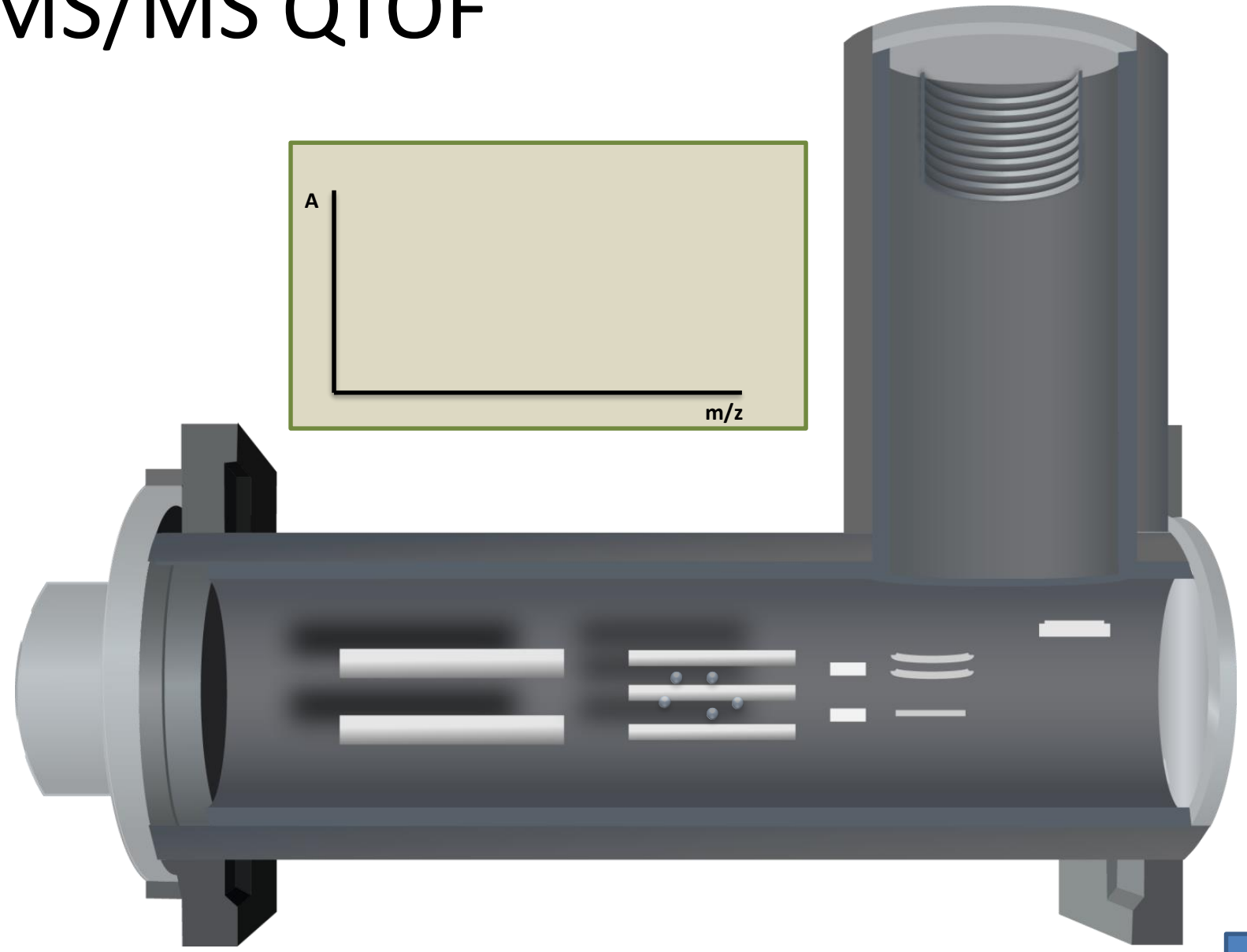
Experimental MS/MS data

Best Match(es)

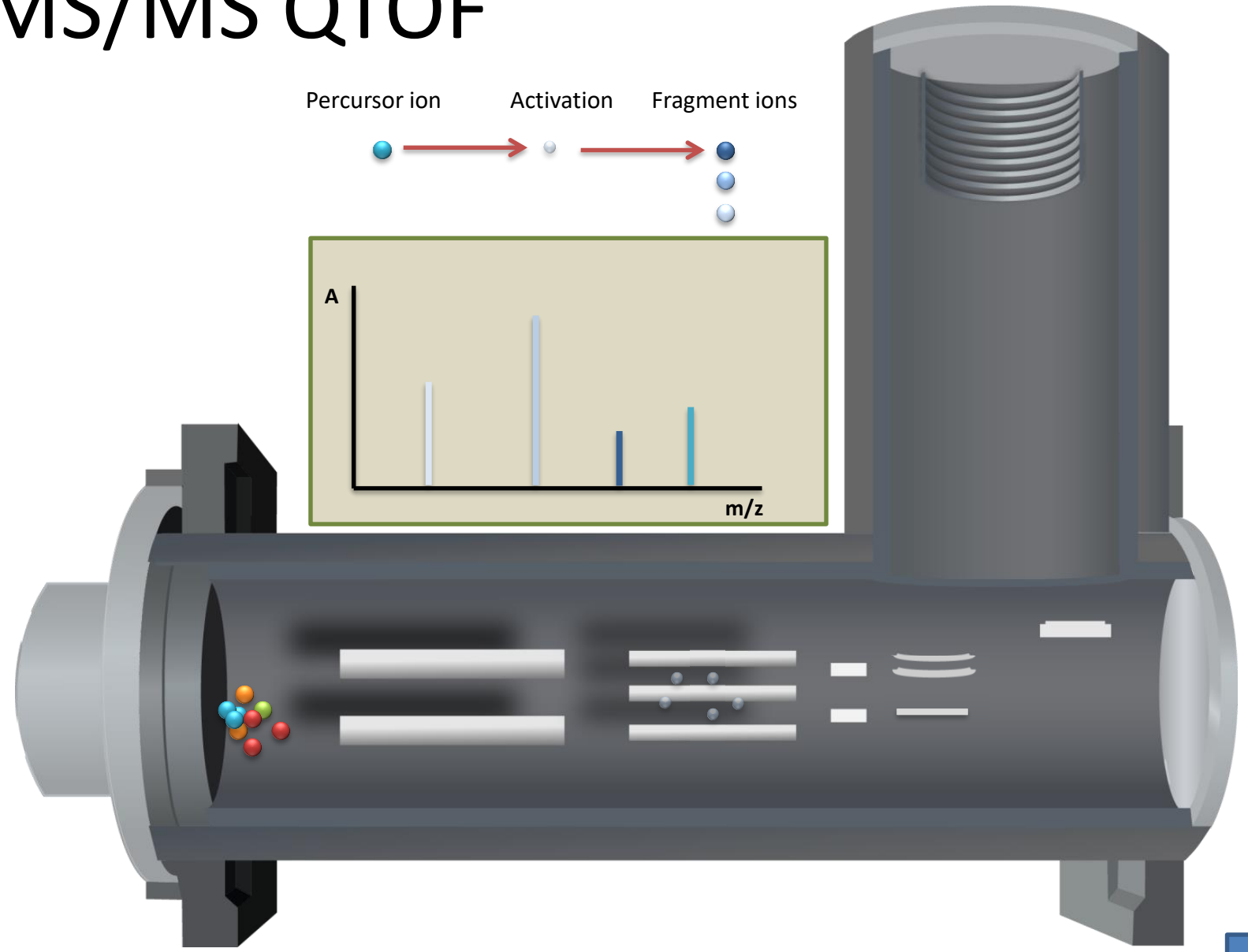


Theoretical MS/MS data

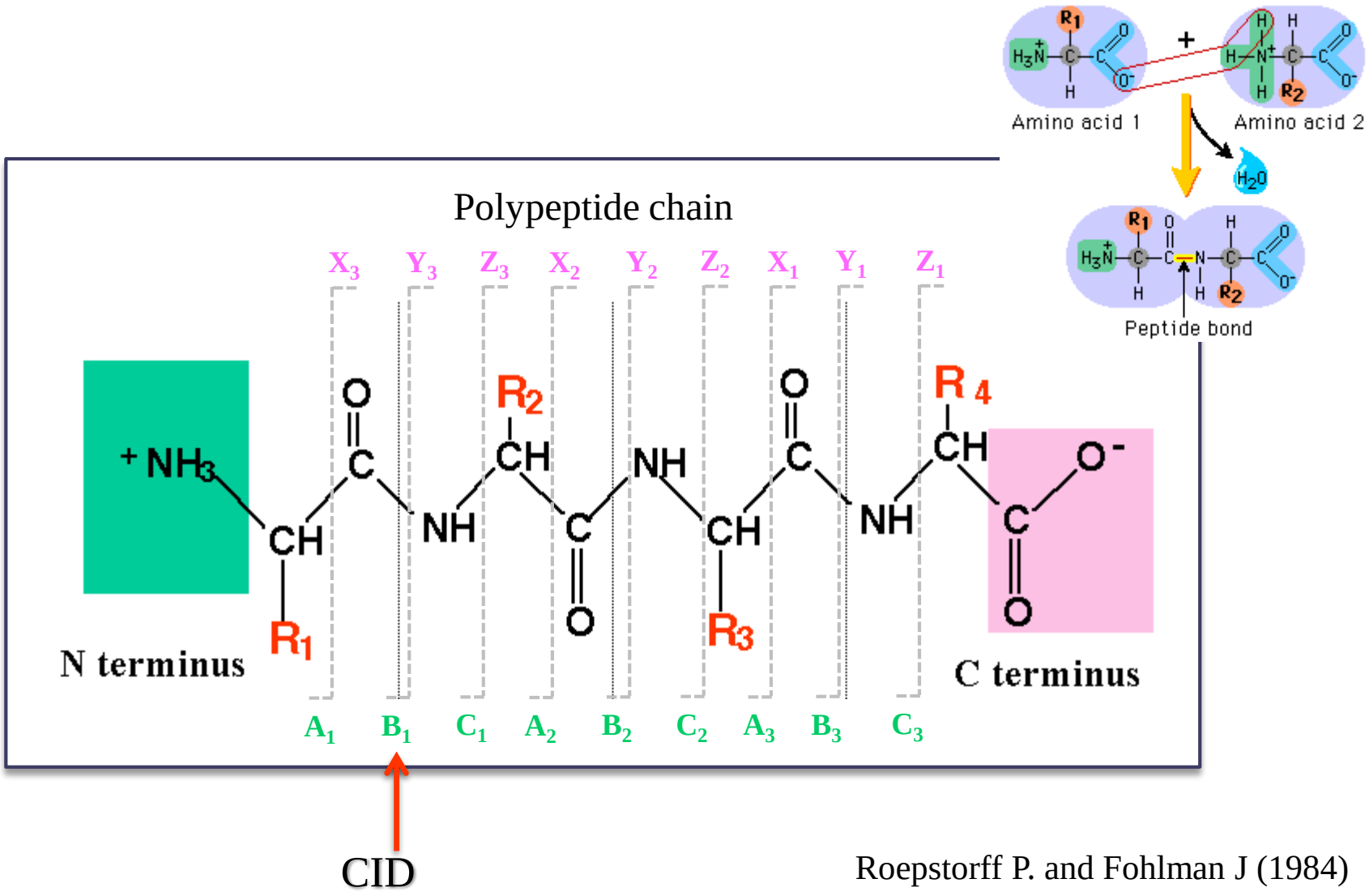
MS/MS QTOF



MS/MS QTOF

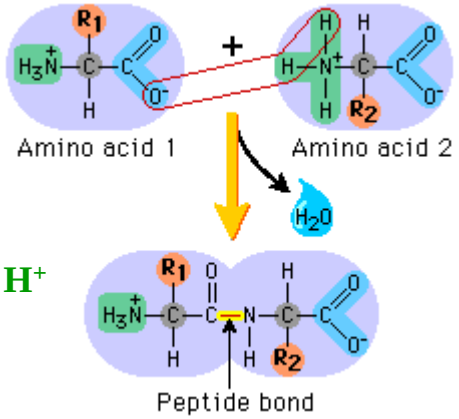


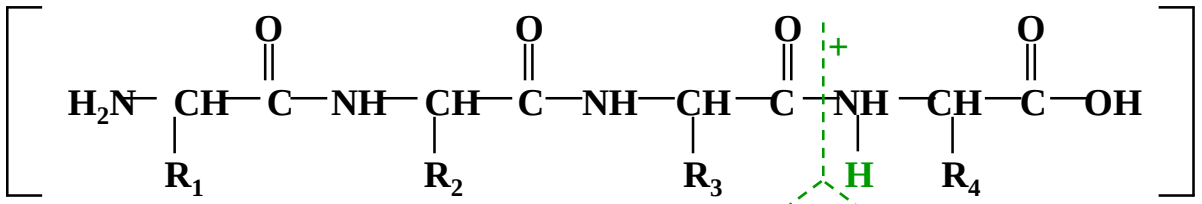
Peptide Fragmentation



Roepstorff P. and Fohlman J (1984)
Biemann, 1988

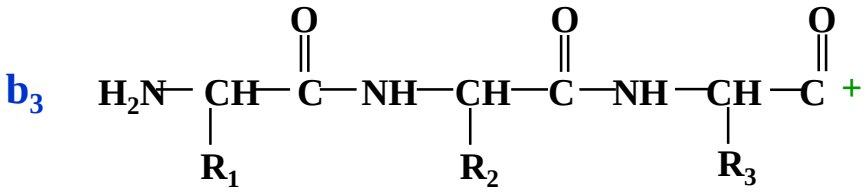
Charge-directed fragmentation

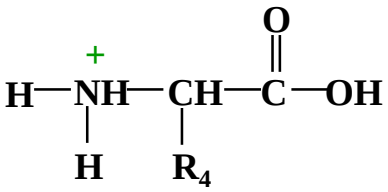




H^+

b ion formation **and/or** **y ion formation**





+ Neutral pumped away by vacuum system

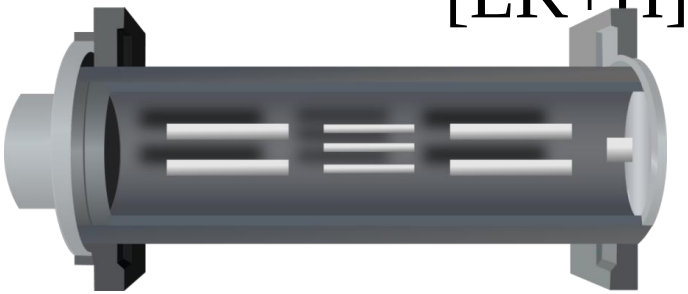
+ Neutral pumped away by vacuum system

Complementary Ions b/y pairs (multiple charged ions)

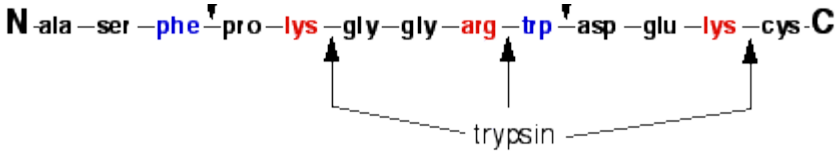
$[ANELK+H]^{2+}$ → $[ANE+H]^{+}$ + $[LK+H]^{+}$
$[ANE+H]^{2+}$ $LK+H$
ANE $[LK+H]^{2+}$

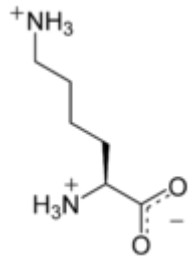
$[ANELK+2H]^{2+}$

$[ANE+H]^{+}$
$[LK+H]^{+}$



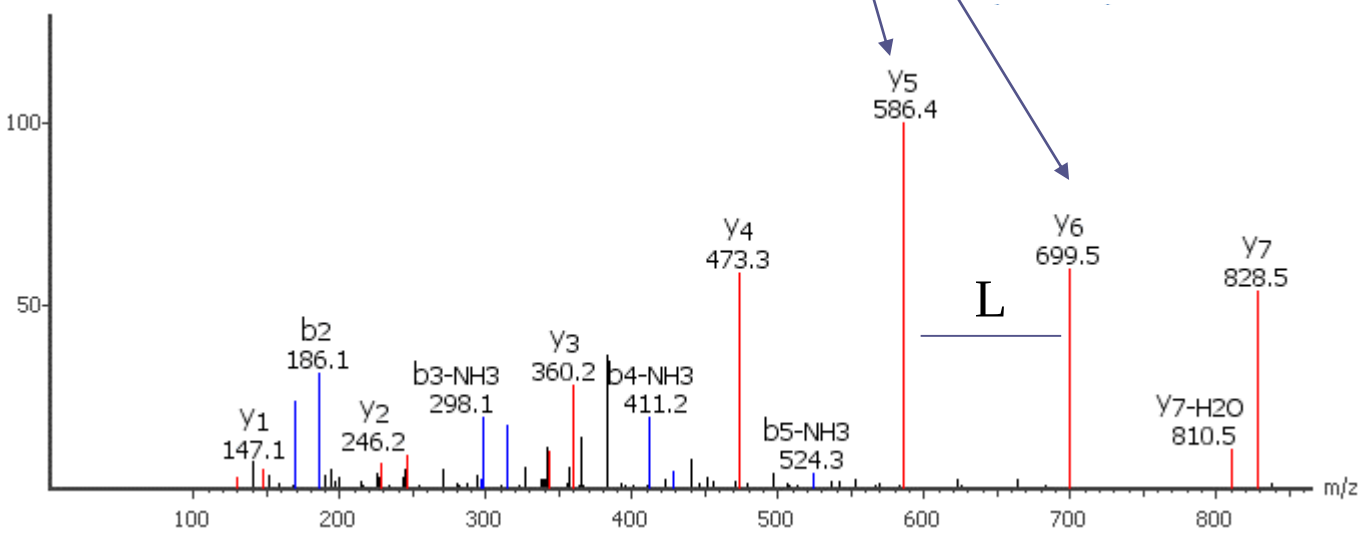
Complementary Ions b/y pairs





Lys (K)

b_1	A NELLLNVK	y_8
b_2	AN ELLLNVK	y_7
b_3	ANE LLLNVK	y_6
b_4	ANEL LLNVK	y_5
b_5	ANELL LNVK	y_4
b_6	ANELLL NVK	y_3
b_7	ANELLLN VK	y_2
b_8	ANELLLNV K	y_1



In silico peptide fragmentation models

Fragment Ion Calculator Results

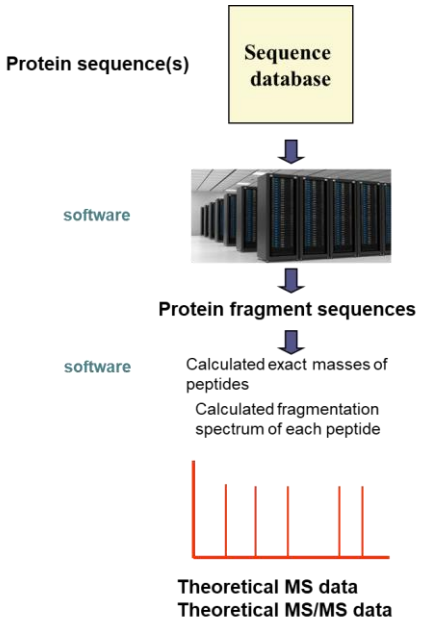
Sequence: **AEFVEVTK**, pI: **4.53158**

Fragment Ion Table, monoisotopic masses

Seq	#	B	Y	# (+1)
A	1	72.04444	922.48807	8
E	2	201.08703	851.45095	7
F	3	348.15544	722.40836	6
V	4	447.22386	575.33995	5
E	5	576.26645	476.27153	4
V	6	675.33486	347.22894	3
T	7	776.38254	248.16053	2
K	8	904.47750	147.11285	1

Mass/Charge Table

	Mass	
	Mono	Avg
(M)	921.48079	922.04593
$(M+H)^{+}$	922.48807	923.05320
$(M+2H)^{2+}$	461.74769	462.03026
$(M+3H)^{3+}$	308.16757	308.35595
$(M+4H)^{4+}$	231.37751	231.51879

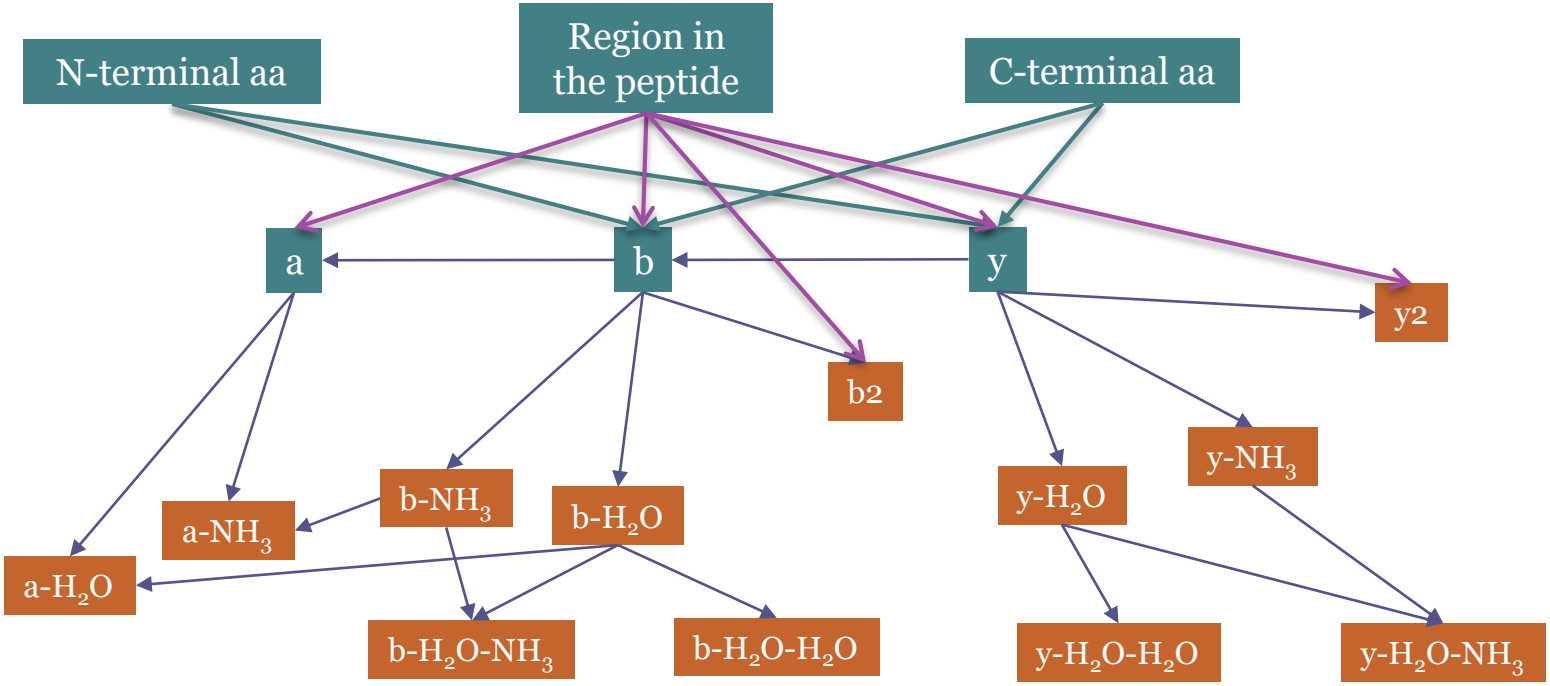


http://db.systemsbiology.net:8080/proteomicsToolkit/FragIonServlet.html

AEFVEVTK

Accurate modeling of peptide fragmentation

Ionization method | Fragmentation method | Fragmentation Energy



(1) correlations between fragment ions;

(2) dependencies due to the relative position of the cleavage site in the peptide;

(3) influence of flanking amino acids to the cleavage site.

Identification search engines

- Mascot
- X!Tandem,
- MS-GF+,
- MS Amanda,
- MyriMatch,
- Comet,
- Tide,
- Andromeda
- OMSSA
- ProteinProspector
- Sequest
- PEAKS
- PRIDE
- ….

Protein identification

MATRIX SCIENCE

Search this site

Home | Mascot database search | Products | Technical support | Training | News | Blog | Newsletter | Contact

Access Mascot Server | Database search help

Mascot database search > Access Mascot Server

Access Mascot Server

You are welcome to submit searches to this free Mascot Server. Searches of MS/MS data are limited to 1200 spectra and some functions, such as no enzyme searches, are unavailable. Automated searching of batches of files is not permitted. If you want to automate search submission, perform large searches, search additional sequence databases, or customise the modifications, quantitation methods, etc., you'll need to license your own, in-house copy of Mascot Server.

Peptide Mass Fingerprint

The experimental data are a list of peptide mass values from the digestion of a protein by a specific enzyme such as trypsin.

Perform search | Example of results report | Tutorial

Sequence Query

One or more peptide mass values associated with information such as partial or ambiguous sequence strings, amino acid composition information, MS/MS fragment ion masses, etc. A super-set of a sequence tag query.

Perform search | Example of results report | More information

MS/MS Ions Search

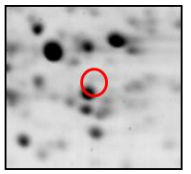
Identification based on raw MS/MS data from one or more peptides.

Perform search | Example of results report | Tutorial

More info

- Mascot overview
- Search parameter reference
- Data file format
- Results report overview

PFF (Peptide Fragment Fingerprinting)



Protein sample

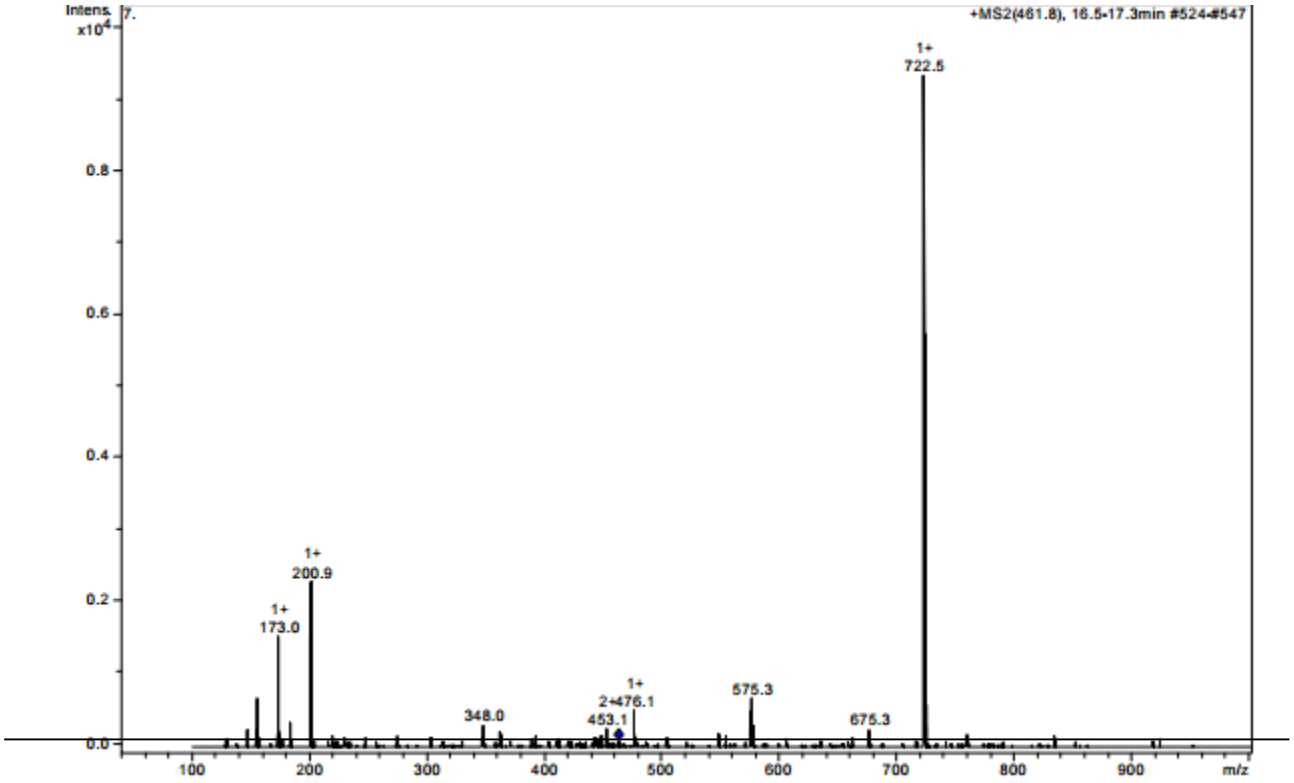
Protease digestion (Trypsin)

Protein fragments

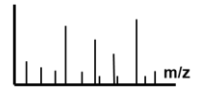
COM=10 pmol digest of Sample BSA
MASS=Monoisotopic
USERNAME=Pedro Domingues
USEREMAIL=p.domingues@ua.pt
BEGIN IONS
TITLE= Cmpd 7, +MSn(461.8), 16.8 min
PEPMASS=461.7491 18565
CHARGE=2+
147.05 229
154.98 787 1+
172.95 1734 1+
183.87 410 1+
200.91 2479 1+
218.92 144
248.16 150
347.22 250
347.96 200
476.17 670 1+
548.13 180
575.33 672
576.37 393 1+
675.33 231
722.45 13125 1+
723.14 357 1+
759.39 159
833.47 141
851.45 149
904.13 432 1+
END IONS

Threshold

MS/MS 461.7



PFF (Mascot)



Experimental MS/MS data

Best Match(es)

Theoretical MS/MS data

MATRIX SCIENCE

Search this site

Home | Mascot database search | Products | Technical support | Training | News | Blog | Newsletter | Contact

Access Mascot Server | Database search help

Mascot database search > Access Mascot Server > MS/MS Ions Search

MASCOT MS/MS Ions Search

Your name: pedro — Email: p.domingues@ua.pt

Search title:

Database(s): SwissProt (AA)

Amino acid (AA)
contaminants
cRAP
NCBIprot
Nucleic acid (NA)
Environmental_EST
Fungi_EST
Human_EST
Invertebrates_EST
Mammals_EST

Taxonomy: . . . . . Mammalia (mammals)

Enzyme: Trypsin — Allow up to: 1 missed cleavages

Quantitation: None

Fixed modifications: --- none selected ---

Display all modifications

Variable modifications: --- none selected ---

Acetyl (K)
Acetyl (N-term)
Acetyl (Protein N-term)
Amidated (C-term)
Amidated (Protein C-term)
Ammonia-loss (N-term C)
Biotin (K)
Biotin (N-term)
Carbamidomethyl (C)
Carbamyl (K)
Carbamyl (N-term)

Peptide tol. ±: 5 ppm — # ^{13}C: 0 — MS/MS tol. ±: 5 ppm

Peptide charge: 2+ — Monoisotopic / Average

Data file: Choose File — Mascot MSMS2.txt

Data format: Mascot generic — Precursor: m/z

Instrument: Default — Error tolerant

Decoy — Report top: AUTO hits

Start Search ... — Reset Form

PFF (Mascot results)

universidade de aveiro

AACLifeSci

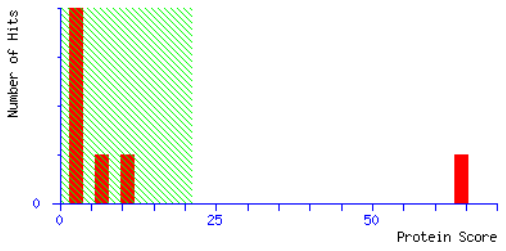
MATRIX SCIENCE Mascot Search Results

```
User            : Pedro Domingues
Email           : p.domingues@ua.pt
Search title    : 10 pmol digest of Sample BSA
MS data file    : new 1.txt
Database        : SwissProt 2015_10 (549646 sequences; 195983064 residues)
Taxonomy        : Mammalia (mammals) (66401 sequences)
Timestamp       : 21 Oct 2015 at 13:31:53 GMT
Protein hits    : ALBU_BOVIN Serum albumin OS=Bos taurus GN=ALB PE=1 SV=4
```

	SwissProt	Decoy	False discovery rate
Peptide matches above identity threshold	1	0	0.00 %
Peptide matches above homology or identity threshold	1	0	0.00 %

Mascot Score Histogram

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 21 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.



Peptide Summary Report

Format As: Peptide Summary ▾ | Help

Significance threshold p< 0.05 | Max. number of hits AUTO

Standard scoring ◉ MudPIT scoring ○ | Ions score or expect cut-off 0 | Show sub-sets 0

Show pop-ups ◉ Suppress pop-ups ○ | Sort unassigned Decreasing Score ▾ | Require bold red ☐

Preferred taxonomy: All entries ▾

Select All | Select None | Search Selected | ☐ Error tolerant

1. ALBU_BOVIN Mass: 69248 Score: 64 Matches: 1(1) Sequences: 1(1)
   Serum albumin OS=Bos taurus GN=ALB PE=1 SV=4
   ☐ Check to include this hit in error tolerant search

	Query	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	Unique	Peptide
☑	1	461.7491	921.4836	921.4807	3.15	0	64	3e-06	1	U	K.AEFVEVTK.L

Search Parameters

```
Type of search         : MS/MS Ion Search
Enzyme                 : Trypsin
Mass values            : Monoisotopic
Protein Mass           : Unrestricted
Peptide Mass Tolerance : ± 5 ppm
Fragment Mass Tolerance: ± 0.5 Da
Max Missed Cleavages   : 1
Instrument type        : Default
Number of queries      : 1
```

Mascot: http://www.matrixscience.com/

PFF (Mascot results)

universidade de aveiro

AACLifeSci

MATRIX SCIENCE **MASCOT Search Results**

Protein View: ALBU_BOVIN

Serum albumin OS=Bos taurus GN=ALB PE=1 SV=4

Database:	SwissProt
Score:	61
Nominal mass (M_r):	69248
Calculated pI:	5.82
Taxonomy:	Bos taurus

Sequence similarity is available as an NCBI BLAST search of ALBU_BOVIN against nr.

Search parameters

MS data file: new 1.txt
Enzyme: Trypsin: cuts C-term side of KR unless next residue is P.

Protein sequence coverage: 1%

Matched peptides shown in ***bold red***.

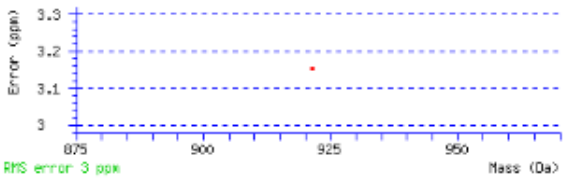
```
  1 MKWVTFISLL LLFSSAYSRG VFRRDTHKSE IAHRFKDLGE EHFKGLVLIA
 51 FSQYLQQCPF DEHVKLVNEL TEFAKTCVAD ESHAGCEKSL HTLFGDELCK
101 VASLRETYGD MADCCEKQEP ERNECFLSHK DDSPDLPKLK PDPNTLCDEF
151 KADEKKFWGK YLYEIARRHP YFYAPELLYY ANKYNGVFQE CCQAEDKGAC
201 LLPKIETMRE KVLASSARQR LRCASIQKFG ERALKAWSVA RLSQKFPKAE
251 FVEVTKLVTD LTKVHKECCH GDLLECADDR ADLAKYICDN QDTISSKLKE
301 CCDKPLLEKS HCIAEVEKDA IPENLPPLTA DFAEDKDVCK NYQEAKDAFL
351 GSFLYEYSRR HPEYAVSVLL RLAKEYEATL EECCAKDDPH ACYSTVFDKL
401 KHLVDEPQNL IKQNCDQFEK LGEYGFQNAL IVRYTRKVPQ VSTPTLVEVS
451 RSLGKVGTRC CTKPESERMP CTEDYLSLIL NRLCVLHEKT PVSEKVTKCC
501 TESLVNRRPC FSALTPDETY VPKAFDEKLF TFHADICTLP DTEKQIKKQT
551 ALVELLKHKP KATEEQLKTV MENFVAFVDK CCAADDKEAC FAVEGPKLVV
601 STQTALA
```

Unformatted sequence string: 607 residues (for pasting into other applications).

Sort peptides by: (•) Residue Number ( ) Increasing Mass ( ) Decreasing Mass

Show predicted peptides also

Query	Start – End	Observed	Mr(expt)	Mr(calc)	ppm	M	Score	Expect	Rank	U	Peptide
1	249 – 256	461.7491	921.4836	921.4807	3.15	0	61	6.1e-06	1	U	K.AEFVEVTK.L




```
AC   P02769; A5PJX3; O02787; P04277; Q3SZR2;
DT   21-JUL-1986, integrated into UniProtKB/Swiss-Prot.
DT   01-FEB-1996, sequence version 4.
DT   14-OCT-2015, entry version 148.
DE   RecName: Full=Serum albumin;
DE   AltName: Full=BSA;
DE   AltName: Allergen=Bos d 6;
DE   Flags: Precursor;
```

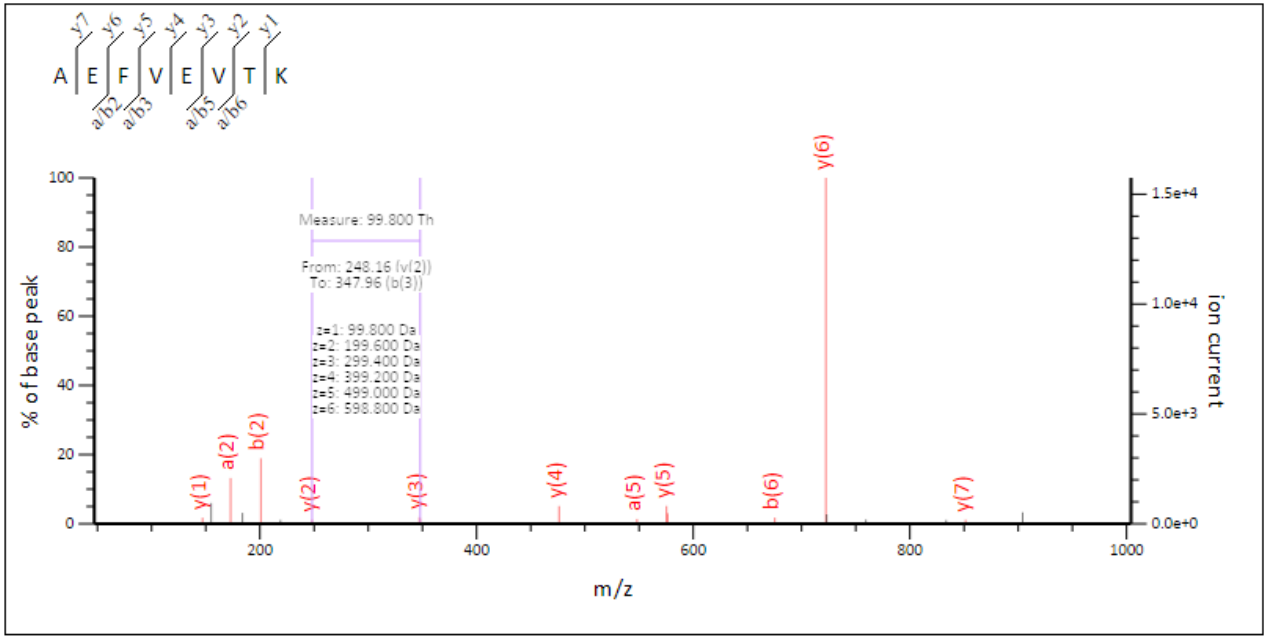
PFF (Mascot results)

MATRIX SCIENCE Mascot Search Results

Peptide View

MS/MS Fragmentation of **AEFVEVTK**
Found in **ALBU_BOVIN** in **SwissProt**, Serum albumin OS=Bos taurus GN=ALB PE=1 SV=4

Match to Query 1: 921.483648 from(461.749100,2+) intensity(18565.0000) index(0)
Title: Cmpd 7, +MSn(461.8), 16.8 min
Data file new 1.txt



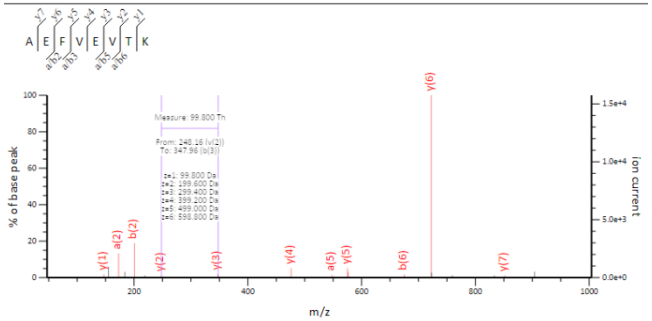
COM=10 pmol digest of Sample BSA
MASS=Monoisotopic
USERNAME=Pedro Domingues
USEREMAIL=p.domingues@ua.pt
BEGIN IONS
TITLE= Cmpd 7, +MSn(461.8), 16.8 min
PEPMASS=461.7491 18565
CHARGE=2+
147.05 229
154.98 787 1+
172.95 1734 1+
183.87 410 1+
200.91 2479 1+
218.92 144
248.16 150
347.22 250
347.96 200
476.17 670 1+
548.13 180
575.33 672
576.37 393 1+
675.33 231
722.45 13125 1+
723.14 357 1+
759.39 159
833.47 141
851.45 149
904.13 432 1+
END IONS

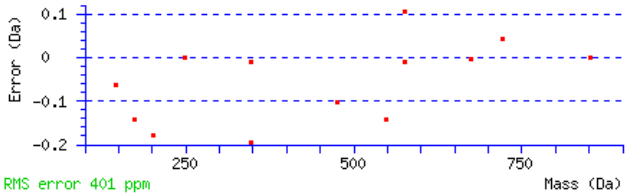
2-PFF (Mascot results)

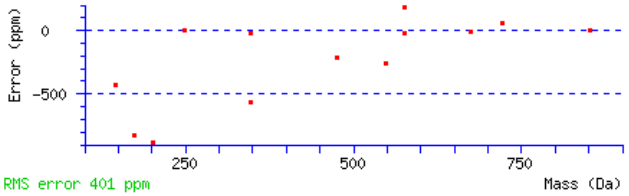
Label all possible matches ○ Label matches used for scoring ◉

```
Monoisotopic mass of neutral peptide Mr(calc): 921.4807
Ions Score: 64  Expect: 2.5e-05
Matches : 13/56 fragment ions using 19 most intense peaks   (help)
```

#	a	a^{++}	b	b^{++}	Seq.	y	y^{++}	y*	y*$^{++}$	#
1	44.0495	22.5284	72.0444	36.5258	A					8
2	173.0921	87.0497	201.0870	101.0471	E	851.4509	426.2291	834.4244	417.7158	7
3	320.1605	160.5839	348.1554	174.5813	F	722.4083	361.7078	705.3818	353.1945	6
4	419.2289	210.1181	447.2238	224.1155	V	575.3399	288.1736	558.3134	279.6603	5
5	548.2715	274.6394	576.2664	288.6368	E	476.2715	238.6394	459.2449	230.1261	4
6	647.3399	324.1736	675.3348	338.1710	V	347.2289	174.1181	330.2023	165.6048	3
7	748.3876	374.6974	776.3825	388.6949	T	248.1605	124.5839	231.1339	116.0706	2
8					K	147.1128	74.0600	130.0863	65.5468	1







NCBI **BLAST** search of AEFVEVTK
(Parameters: blastp, nr protein database, expect=20000, no filter, PAM30)
Other BLAST web gateways

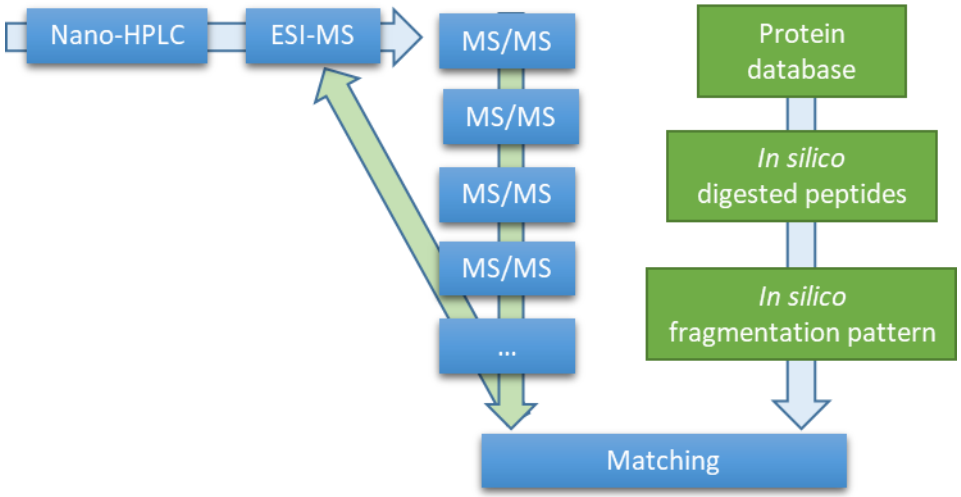
All matches to this query

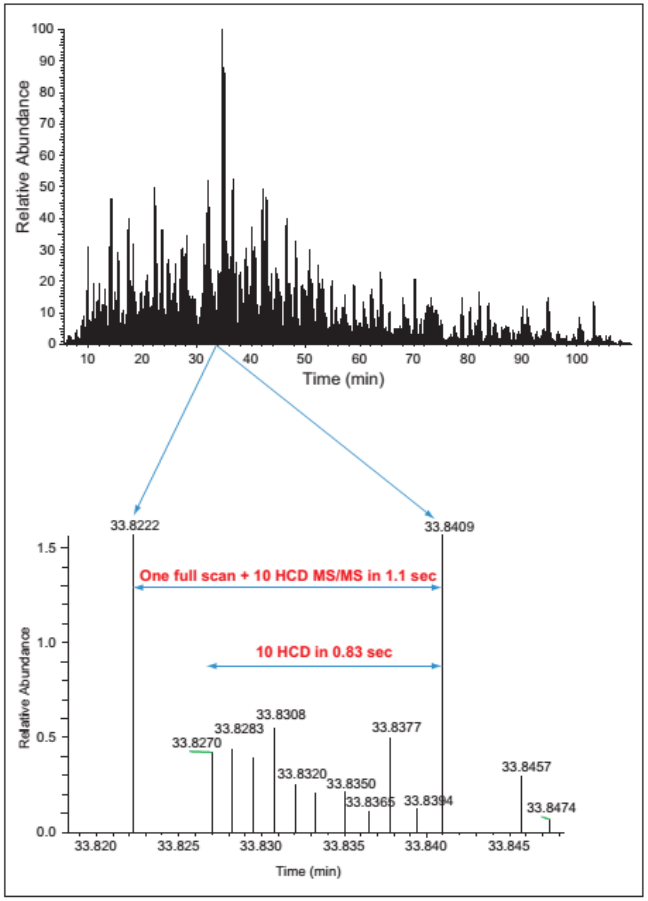
Score	Mr(calc)	Delta	Sequence
64.2	921.4807	0.0029	AEFVEVTK
14.1	921.4807	0.0029	YIADLEAK
12.8	921.4807	0.0029	EADLFISK
12.8	921.4807	0.0029	VTDFLAEK
11.8	921.4841	-0.0005	VTMETLTK
11.5	921.4807	0.0029	AEALYDIK
11.5	921.4807	0.0029	AEAYLLDK
11.5	921.4807	0.0029	AEYIADIK

PFF (Orthogonal datasets and confidence levels)

- Orthogonal datasets and confidence levels
  - Db : 100 000 sequences
  - 500 spectra
- Probability of one (any) spectrum "accidentally" matching a sequence (wrong match) :
  - $1/100\,000 \times 500 = 5.10^{-3}$ (0.005)
- Probability of 2 spectra "accidentally" matching the same sequence (wrong match) :
  - $5.10^{-3} \times 5.10^{-3} = 2.5 \times 10^{-5}$
- Much higher confidence of identification with at least 2 peptides matching the same protein sequence

Data Dependent Tandem Mass Spectrometry



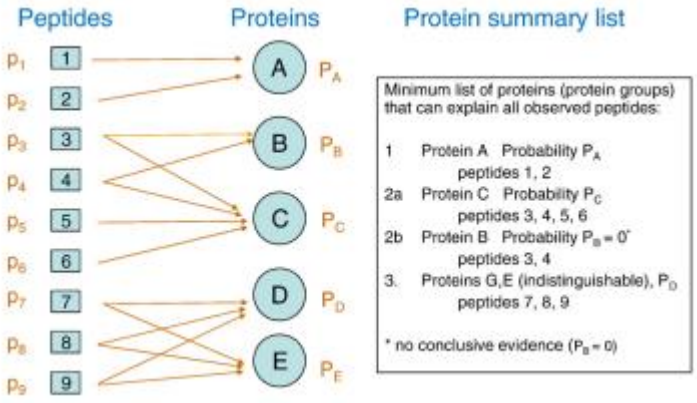


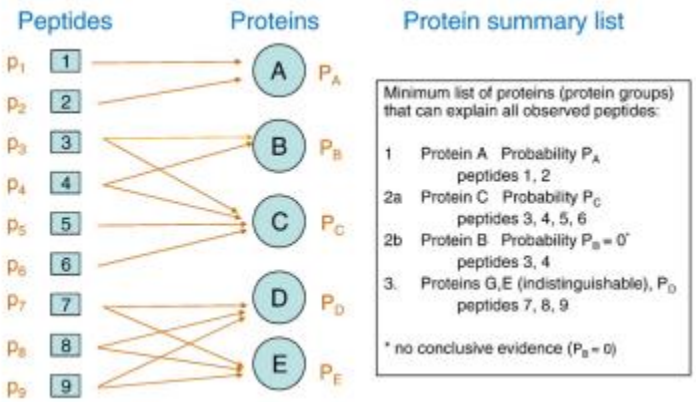
2-PFF

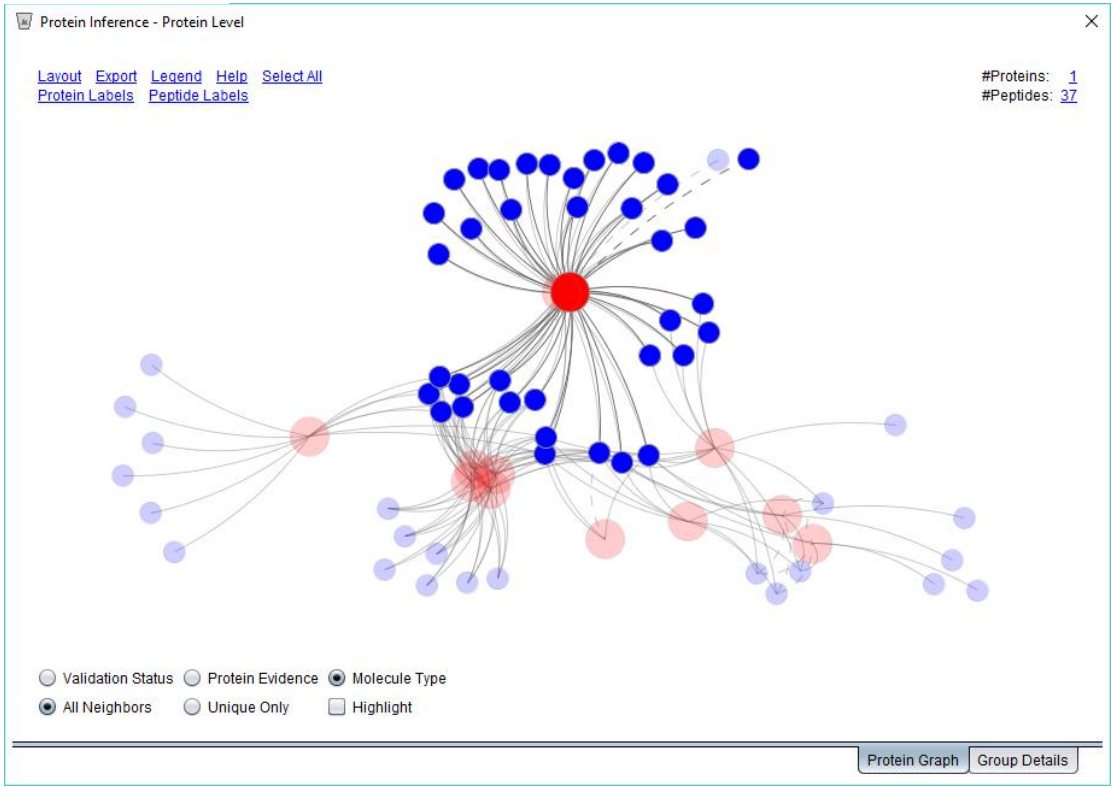
- Easily automated for high throughput
- Can get matches from marginal data

- Can be slow
- Large dataset

- MS/MS is peptide identification
- Proteins by inference.



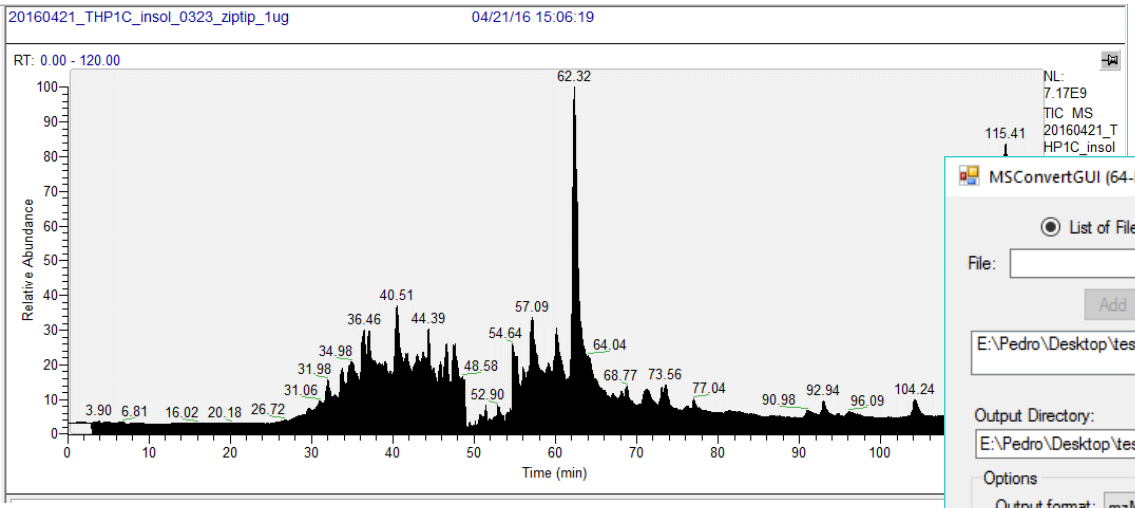

Peptides
Proteins
Protein summary list
p1
p2
p3
p4
p5
p6
p7
p8
p9
1
2
3
4
5
6
7
8
9
A
B
C
D
E
PA
PB
PC
PD
PE
Minimum list of proteins (protein groups) that can explain all observed peptides:
1 Protein A Probability PA
peptides 1, 2
2a Protein C Probability PC
peptides 3, 4, 5, 6
2b Protein B Probability PB = 0*
peptides 3, 4
3. Proteins G,E (indistinguishable), PD
peptides 7, 8, 9
* no conclusive evidence (PB = 0)

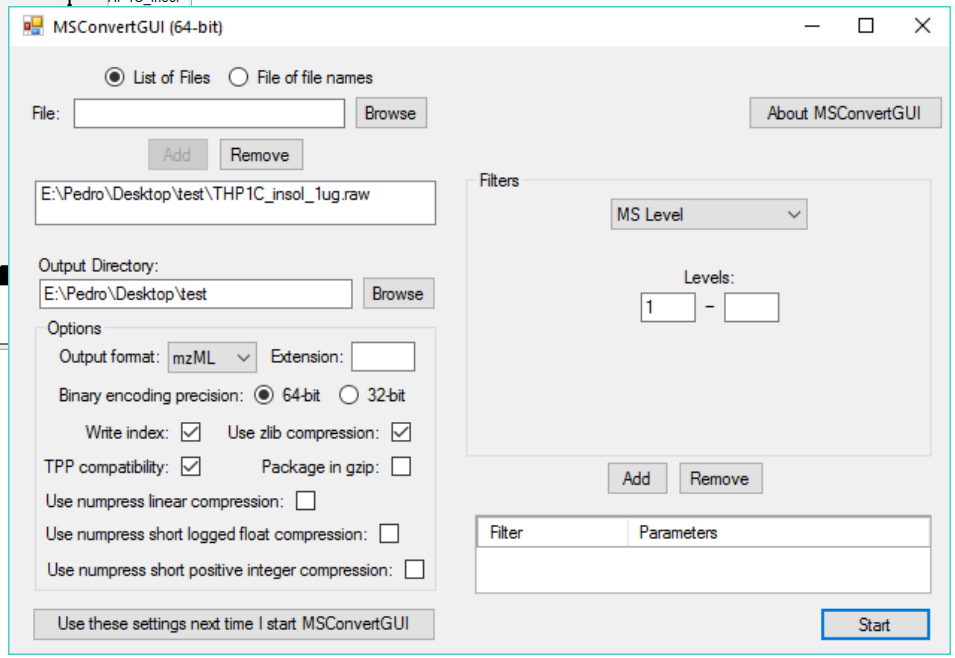

Protein Inference - Protein Level
Layout Export Legend Help Select All
Protein Labels Peptide Labels
#Proteins: 1
#Peptides: 37
Validation Status
Protein Evidence
Molecule Type
All Neighbors
Unique Only
Highlight
Protein Graph
Group Details

Pratical Bioinformatics

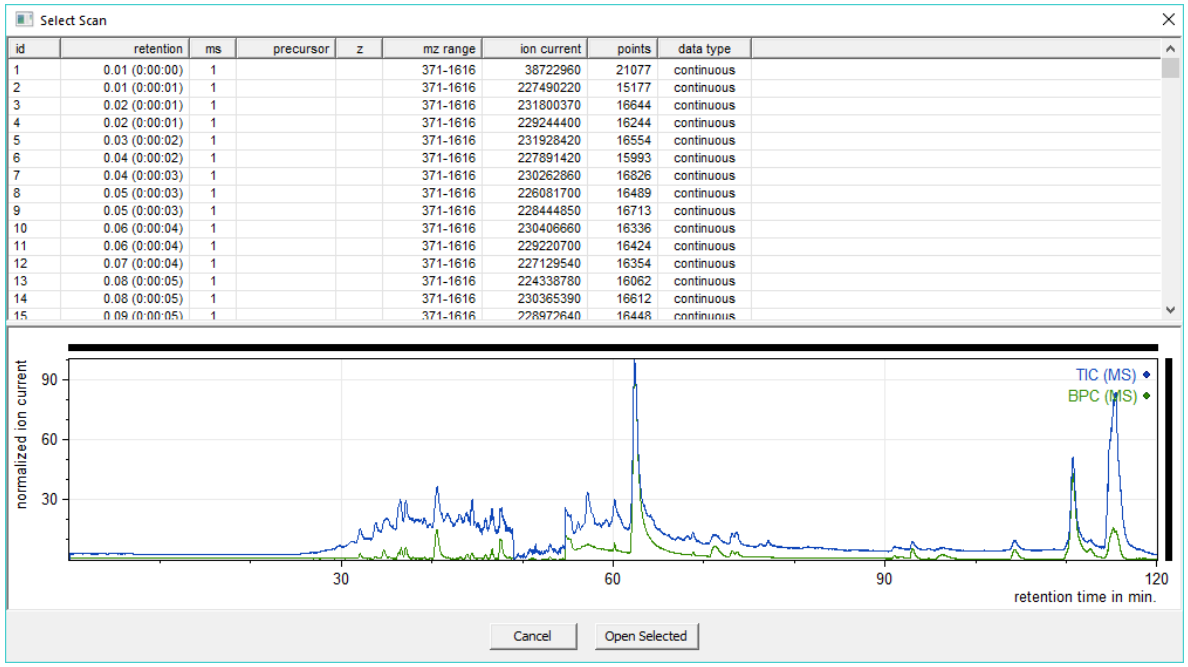
Transformation of data to mzML with MSConverter in ProteoWizard

1) To look at the data in the mMass application, although it can read the mgf file format that is used in SearchGUI, it is best if to convert the RAW data to mzML format.

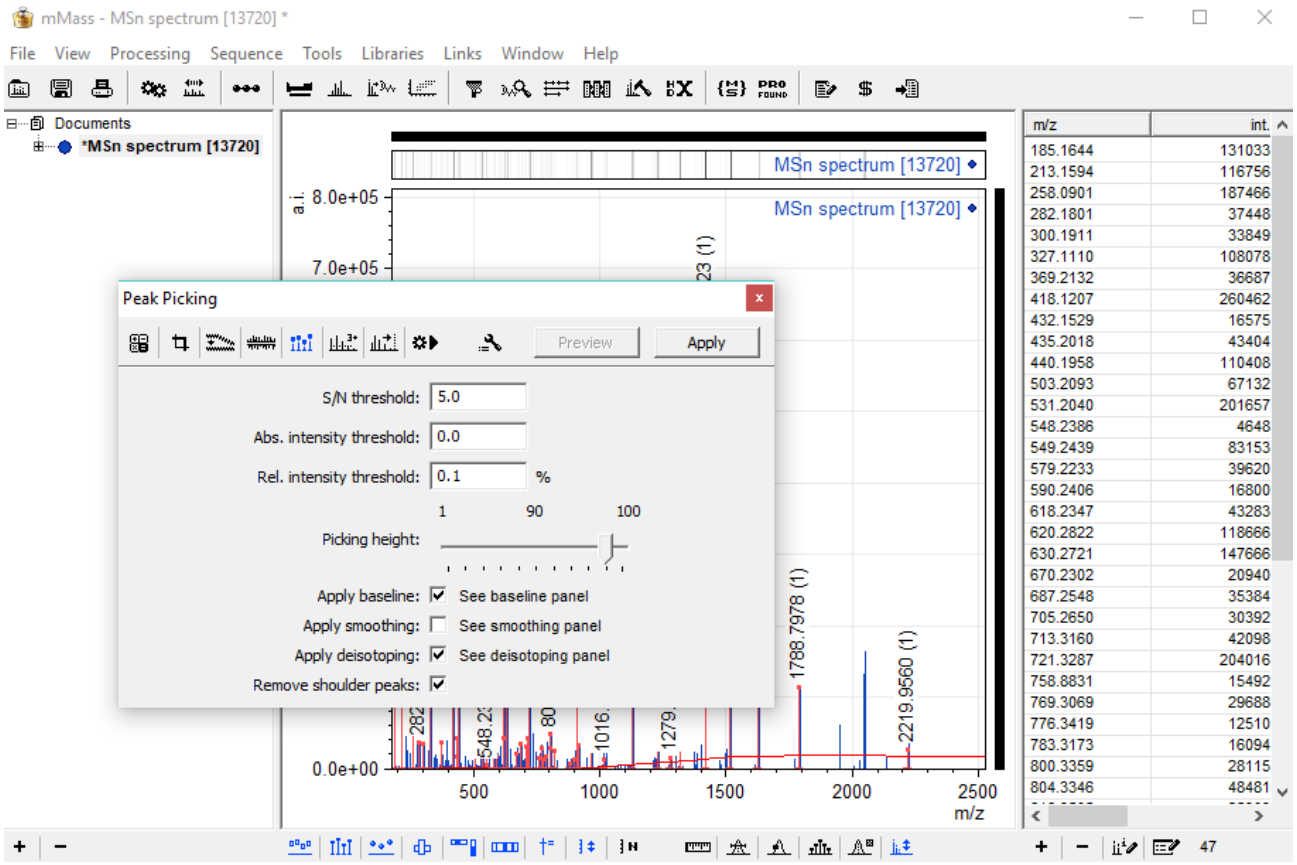




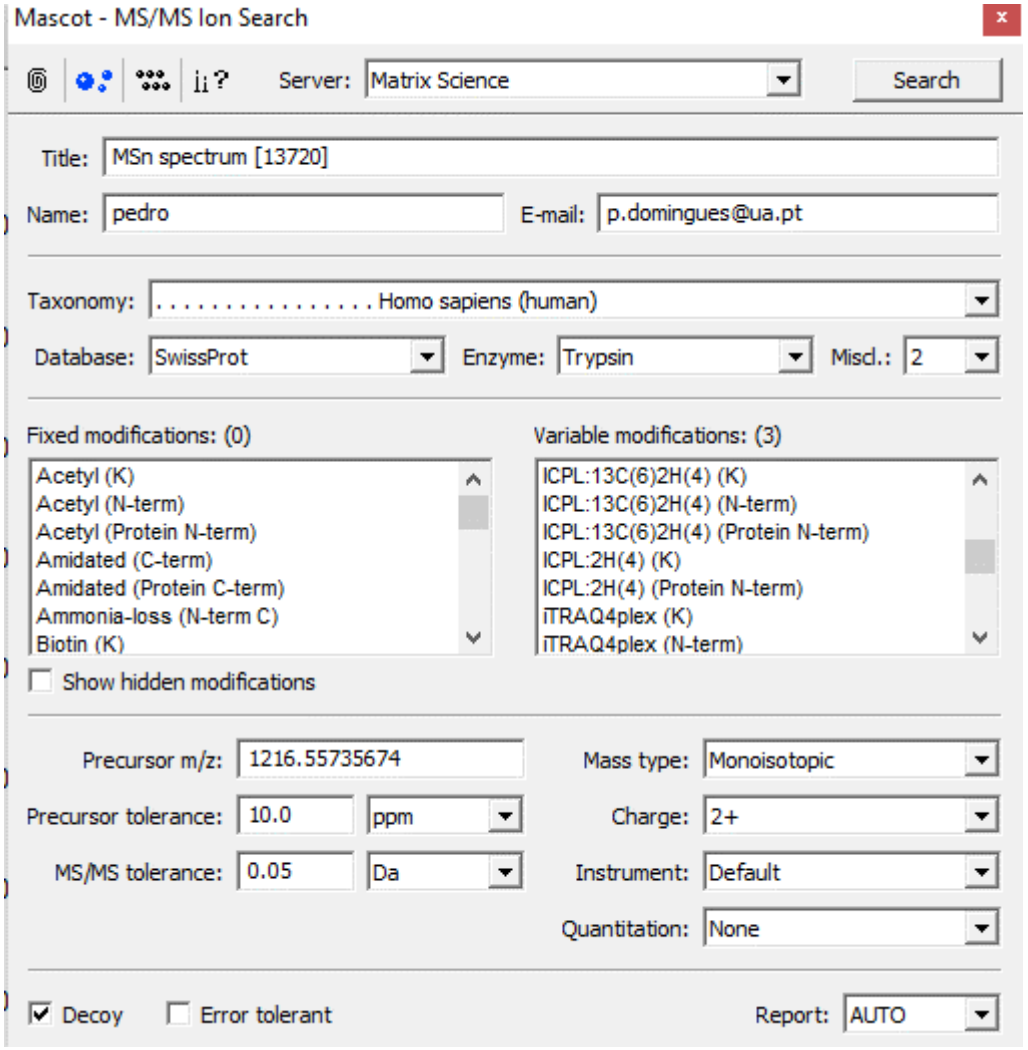
Data observed in the mmass application



1) Open mzML format file
2) Observe the sequence of MS and MS/MS spectra in the acquisition list
3) Open MS at 45.50. Peak piking
4) Observe the MS spectrum at RT 47.50
5) Search MSMS m/z 1216.56(+2) at RT 47.53
6) Search MS/MS in Mascot


mMass - MSn spectrum [13720] *
File View Processing Sequence Tools Libraries Links Window Help
Documents
*MSn spectrum [13720]
MSn spectrum [13720]
MSn spectrum [13720]
a.i.
8.0e+05
7.0e+05
0.0e+00
500
1000
1500
2000
2500
m/z
1788.7978 (1)
2219.9560 (1)
Peak Picking
Preview
Apply
S/N threshold: 5.0
Abs. intensity threshold: 0.0
Rel. intensity threshold: 0.1 %
1 90 100
Picking height:
Apply baseline: See baseline panel
Apply smoothing: See smoothing panel
Apply deisotoping: See deisotoping panel
Remove shoulder peaks:
m/z int.
185.1644 131033
213.1594 116756
258.0901 187466
282.1801 37448
300.1911 33849
327.1110 108078
369.2132 36687
418.1207 260462
432.1529 16575
435.2018 43404
440.1958 110408
503.2093 67132
531.2040 201657
548.2386 4648
549.2439 83153
579.2233 39620
590.2406 16800
618.2347 43283
620.2822 118666
630.2721 147666
670.2302 20940
687.2548 35384
705.2650 30392
713.3160 42098
721.3287 204016
758.8831 15492
769.3069 29688
776.3419 12510
783.3173 16094
800.3359 28115
804.3346 48481
47

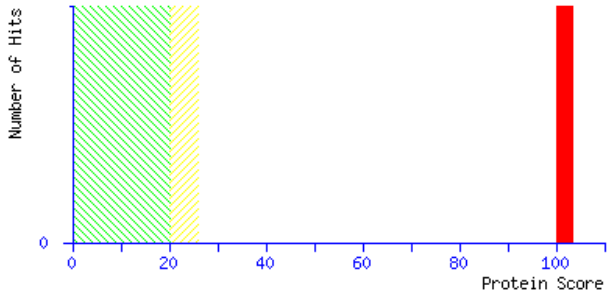
Search MS/MS in Mascot



Modifications: Oxidation M
Carbamidomethyl (C)

Mascot Score Histogram

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 20 indicate peptides with significant homology.
Individual ions scores > 26 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.



Peptide Summary Report

Format As: Peptide Summary | Help

Significance threshold p< 0.05 | Max. number of hits AUTO

Standard scoring | MudPIT scoring | Display non-significant matches | Show sub-sets 0

Show pop-ups | Suppress pop-ups | Sort unassigned: Decreasing Score | Require bold red

Preferred taxonomy: All entries

Select All | Select None | Search Selected | Error tolerant

1. HS90B_HUMAN Mass: 83212 Score: 102 Matches: 1(1) Sequences: 1(1)
Heat shock protein HSP 90-beta OS=Homo sapiens GN=HSP90AB1 PE=1 SV=4
Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	Unique	Peptide
1	1216.5574	2431.1002	2431.0970	1.28	0	102	1.4e-09	1	U	R.LVSSPCCIVTSTYGWTANMER.I + 2 Carbamidomethyl (C)

Search Parameters

```
Type of search         : MS/MS Ion Search
Enzyme                 : Trypsin
Variable modifications : Acetyl (K),Carbamidomethyl (C),Oxidation (M)
Mass values            : Monoisotopic
Protein Mass           : Unrestricted
Peptide Mass Tolerance : ± 10 ppm
Fragment Mass Tolerance: ± 0.05 Da
Max Missed Cleavages   : 2
Instrument type        : Default
Number of queries      : 1
```

universidade de aveiro

AACLifeSci

Sequence Editor

Sequence Modifications

Protein Digest

Mass: Mo Av Max charge: 2 Digest Match Annotate

Enzyme: Trypsin Miscl.: 1 Mass range: 500 - 5000 Ignore mods Coverage: 0/99 %

slice	mis.	m/z	z	sequence	error
[73-82]	0	1194.6477	1	k.IDIIPNPQER.t	
[360-378]					
[482-502]					
[560-568]					
[339-348]					
[584-604]					
[393-402]					
[457-477]					
[584-604]					
[507-526]					
[457-477]					
[507-526]					
[393-402]					
[457-477]					
[338-347]					
[169-180]					
[96-107]					
[429-438]					

Sequence Editor

Sequence type: Regular amino acids Cyclic Pattern

Peptide Fragmentation

Mass: Mo Av Max charge: 2 Fragment Match Annotate

M a b c int-a N-ladder -H2O -CO Defined losses +H2O Allow scrambling

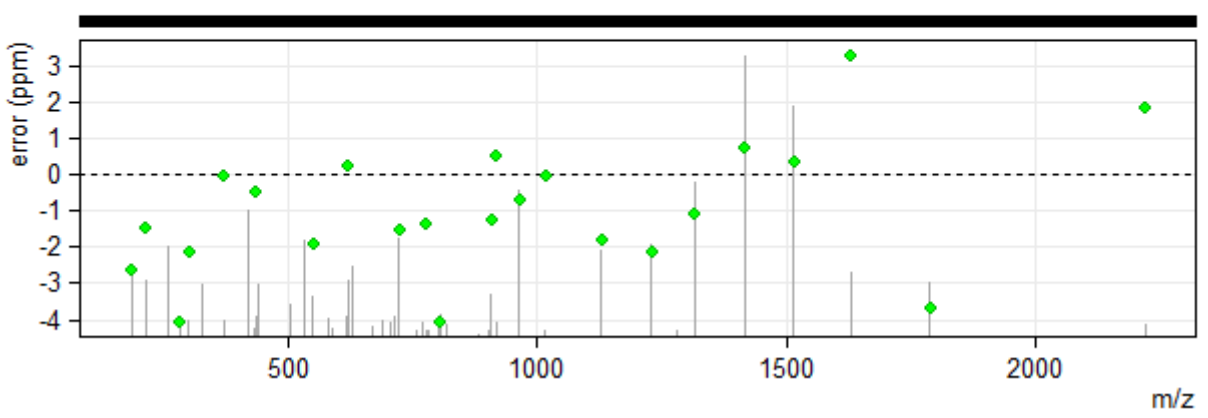
im

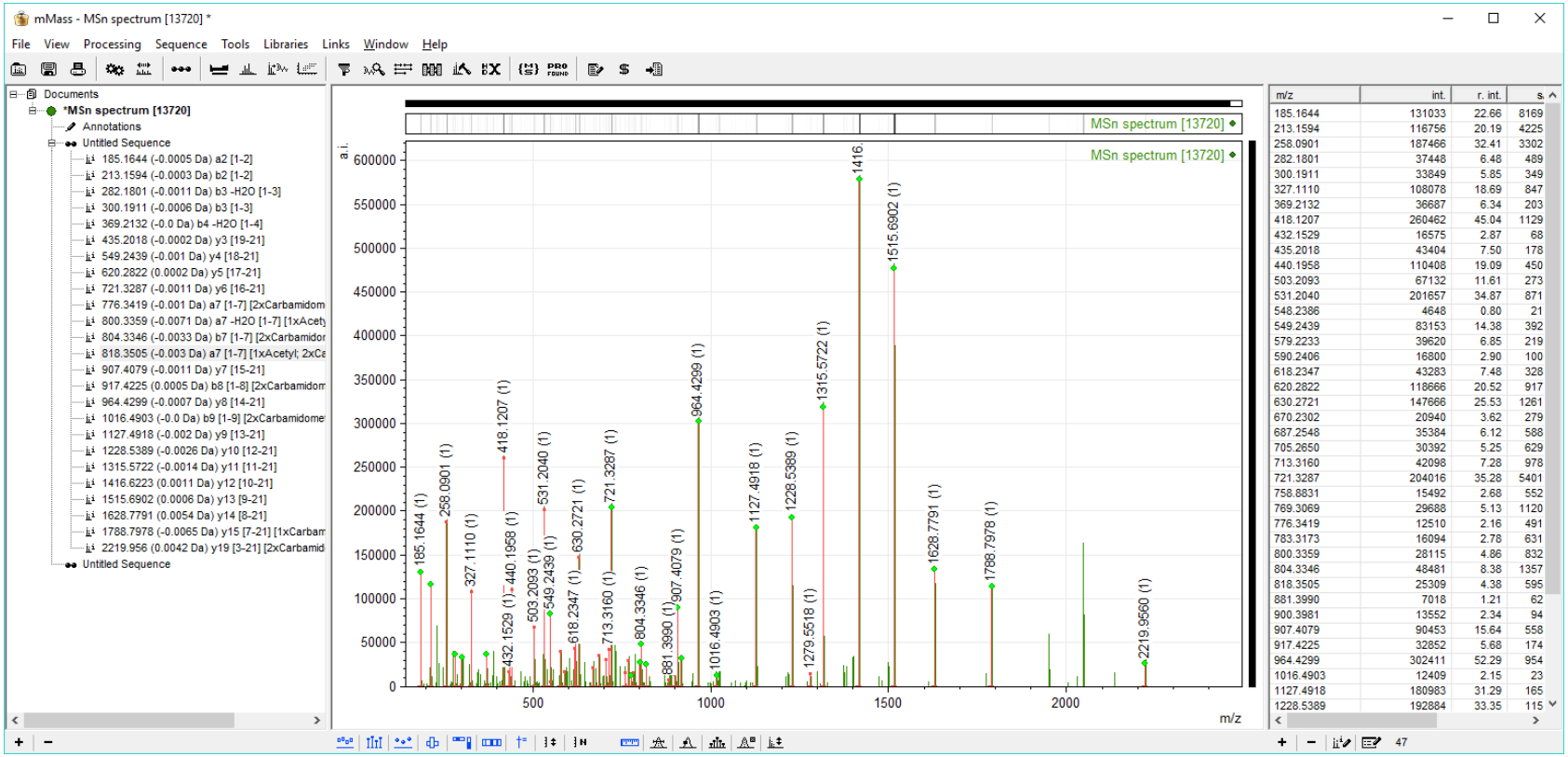
Match Fragments

Tolerance: 10.0 Da ppm Ignore charge Match Calibrate

Remove: Annotated Matched Unselected Isotopes Unknown

ion
M
M
M -H2O
M -H2O
M -NH3
M -NH3
a10
a10
a10 -H2O
a10 -H2O
a11
a11
a11 -H2O
a11 -H2O
a12
a12
a12 -H2O
a12 -H2O




mMass - MSn spectrum [13720] *
File View Processing Sequence Tools Libraries Links Window Help
Documents
*MSn spectrum [13720]
Annotations
Untitled Sequence
185.1644 (-0.0005 Da) a2 [1-2]
213.1594 (-0.0003 Da) b2 [1-2]
282.1801 (-0.0011 Da) b3 -H2O [1-3]
300.1911 (-0.0006 Da) b3 [1-3]
369.2132 (-0.0 Da) b4 -H2O [1-4]
435.2018 (-0.0002 Da) y3 [19-21]
549.2439 (-0.001 Da) y4 [18-21]
620.2822 (0.0002 Da) y5 [17-21]
721.3287 (-0.0011 Da) y6 [16-21]
776.3419 (-0.001 Da) a7 [1-7] [2xCarbamidom
800.3359 (-0.0071 Da) a7 -H2O [1-7] [1xAcet
804.3346 (-0.0033 Da) b7 [1-7] [2xCarbamidor
818.3505 (-0.003 Da) a7 [1-7] [1xAcetyl; 2xCa
907.4079 (-0.0011 Da) y7 [15-21]
917.4225 (0.0005 Da) b8 [1-8] [2xCarbamidom
964.4299 (-0.0007 Da) y8 [14-21]
1016.4903 (-0.0 Da) b9 [1-9] [2xCarbamidome
1127.4918 (-0.002 Da) y9 [13-21]
1228.5389 (-0.0026 Da) y10 [12-21]
1315.5722 (-0.0014 Da) y11 [11-21]
1416.6223 (0.0011 Da) y12 [10-21]
1515.6902 (0.0006 Da) y13 [9-21]
1628.7791 (0.0054 Da) y14 [8-21]
1788.7978 (-0.0065 Da) y15 [7-21] [1xCarbam
2219.956 (0.0042 Da) y19 [3-21] [2xCarbamid
Untitled Sequence
MSn spectrum [13720]
MSn spectrum [13720]
a.i.
m/z
185.1644 (1)
258.0901 (1)
327.1110 (1)
418.1207 (1)
432.1529 (1)
440.1958 (1)
503.2093 (1)
531.2040 (1)
549.2439 (1)
618.2347 (1)
630.2721 (1)
713.3160 (1)
721.3287 (1)
804.3346 (1)
881.3990 (1)
907.4079 (1)
964.4299 (1)
1016.4903 (1)
1127.4918 (1)
1228.5389 (1)
1279.5518 (1)
1315.5722 (1)
1416.
1515.6902 (1)
1628.7791 (1)
1788.7978 (1)
2219.9560 (1)
m/z
int.
r. int.
s.
185.1644 131033 22.66 8169
213.1594 116756 20.19 4225
258.0901 187466 32.41 3302
282.1801 37448 6.48 489
300.1911 33849 5.85 349
327.1110 108078 18.69 847
369.2132 36687 6.34 203
418.1207 260462 45.04 1129
432.1529 16575 2.87 68
435.2018 43404 7.50 178
440.1958 110408 19.09 450
503.2093 67132 11.61 273
531.2040 201657 34.87 871
548.2386 4648 0.80 21
549.2439 83153 14.38 392
579.2233 39620 6.85 219
590.2406 16800 2.90 100
618.2347 43283 7.48 328
620.2822 118666 20.52 917
630.2721 147666 25.53 1261
670.2302 20940 3.62 279
687.2548 35384 6.12 588
705.2650 30392 5.25 629
713.3160 42098 7.28 978
721.3287 204016 35.28 5401
758.8831 15492 2.68 552
769.3069 29688 5.13 1120
776.3419 12510 2.16 491
783.3173 16094 2.78 631
800.3359 28115 4.86 832
804.3346 48481 8.38 1357
818.3505 25309 4.38 595
881.3990 7018 1.21 62
900.3981 13552 2.34 94
907.4079 90453 15.64 558
917.4225 32852 5.68 174
964.4299 302411 52.29 954
1016.4903 12409 2.15 23
1127.4918 180983 31.29 165
1228.5389 192884 33.35 115
47

mmass

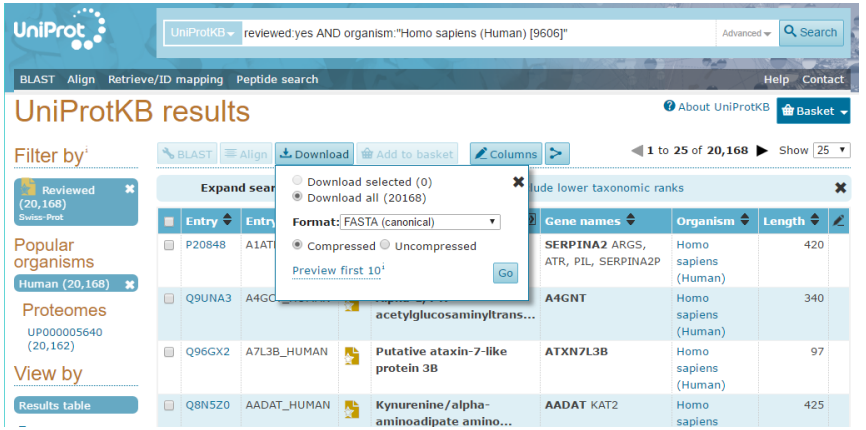
- Tools
  - Mass calculator
  - Mass to formula

Generation of FASTA database from Uniprot (SwissProt)

1) Now, you should download the FASTA file of the proteome of HOMO sapiens.

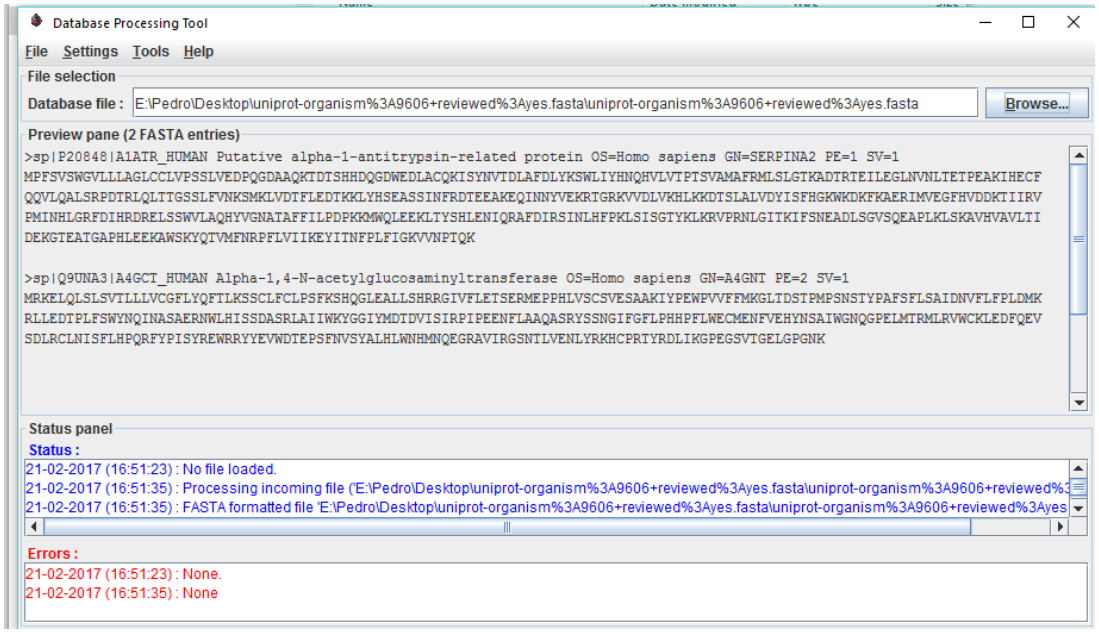
   a) This should be done in the taxonomy page of Uniprot site and search Homo Sapiens

2) Now you should download the Reviewed (Swiss-prot) FASTA file



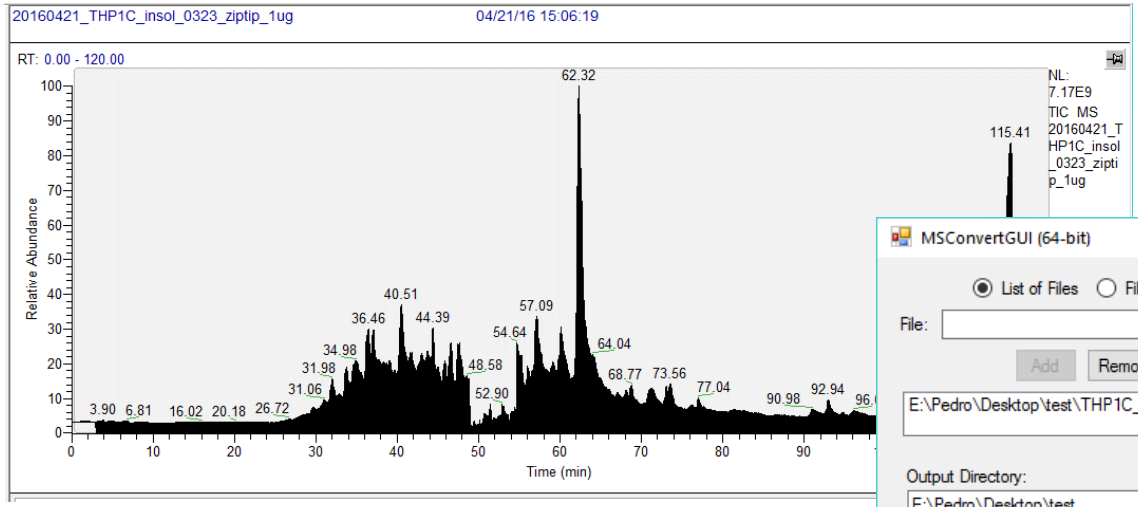
Dbtoolkit dataBase Processing Tool

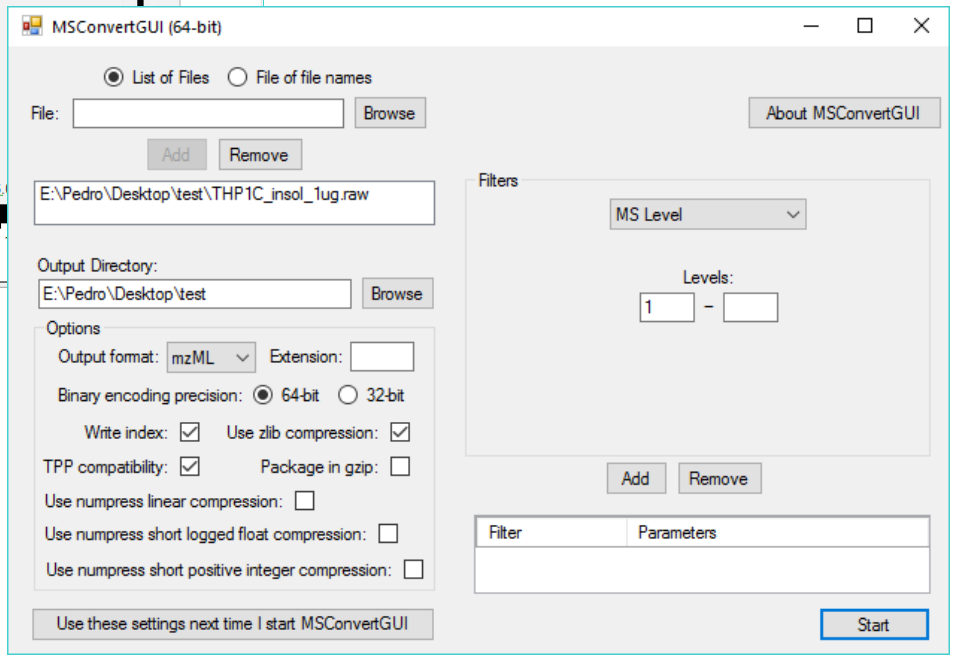
You can look and edit the FASTA file information by using the Dbtoolkit dataBase Processing Tool



Transformation of data to MGF with MSConverter in ProteoWizard

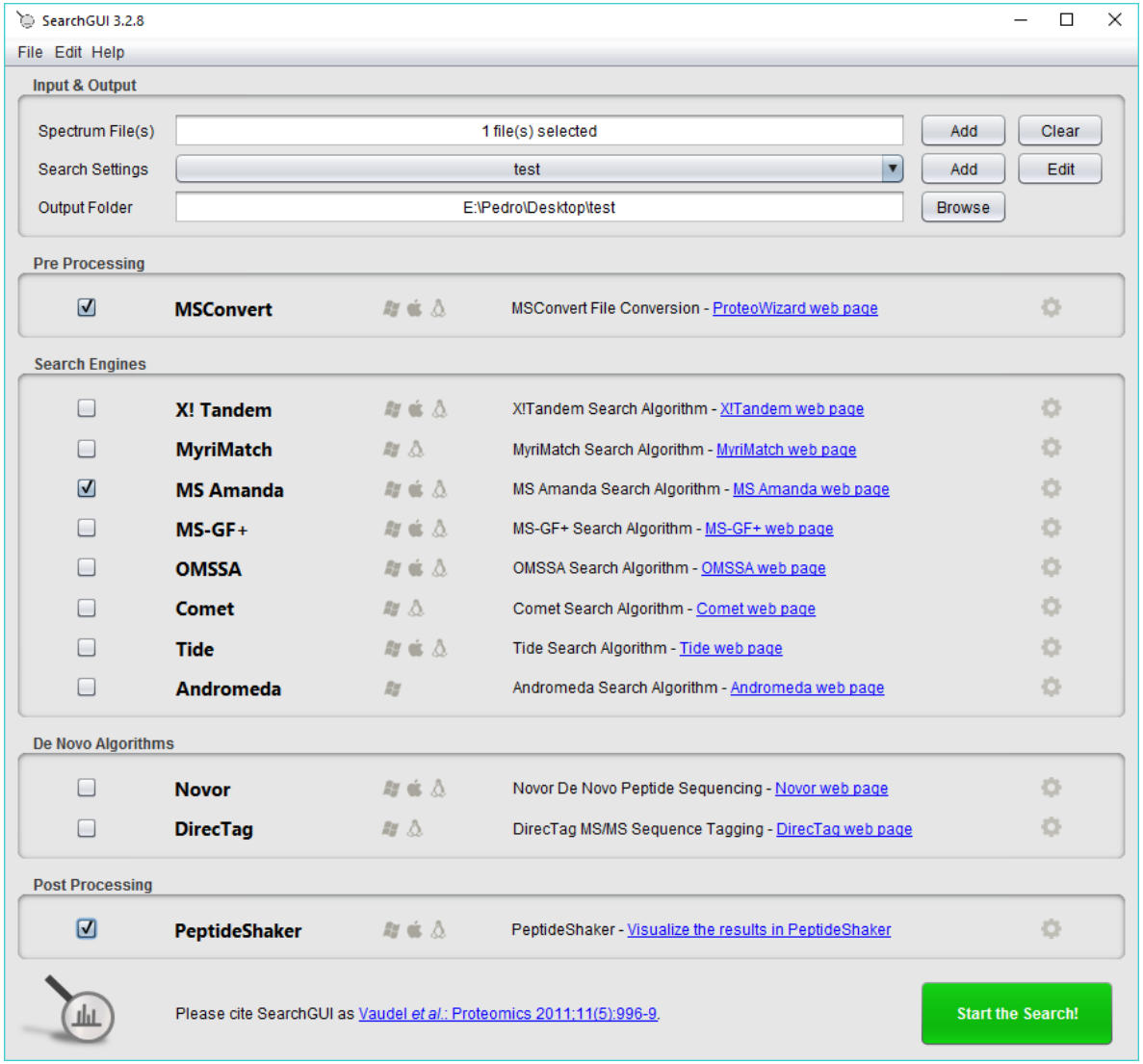
- Use MSConverter in ProteoWizard to convert the HPC-MS data acquired in the orbitrap (RAW data file) to a format that can be read by the SearchGUI (MGF data file).





Search Engines: SearchGUI

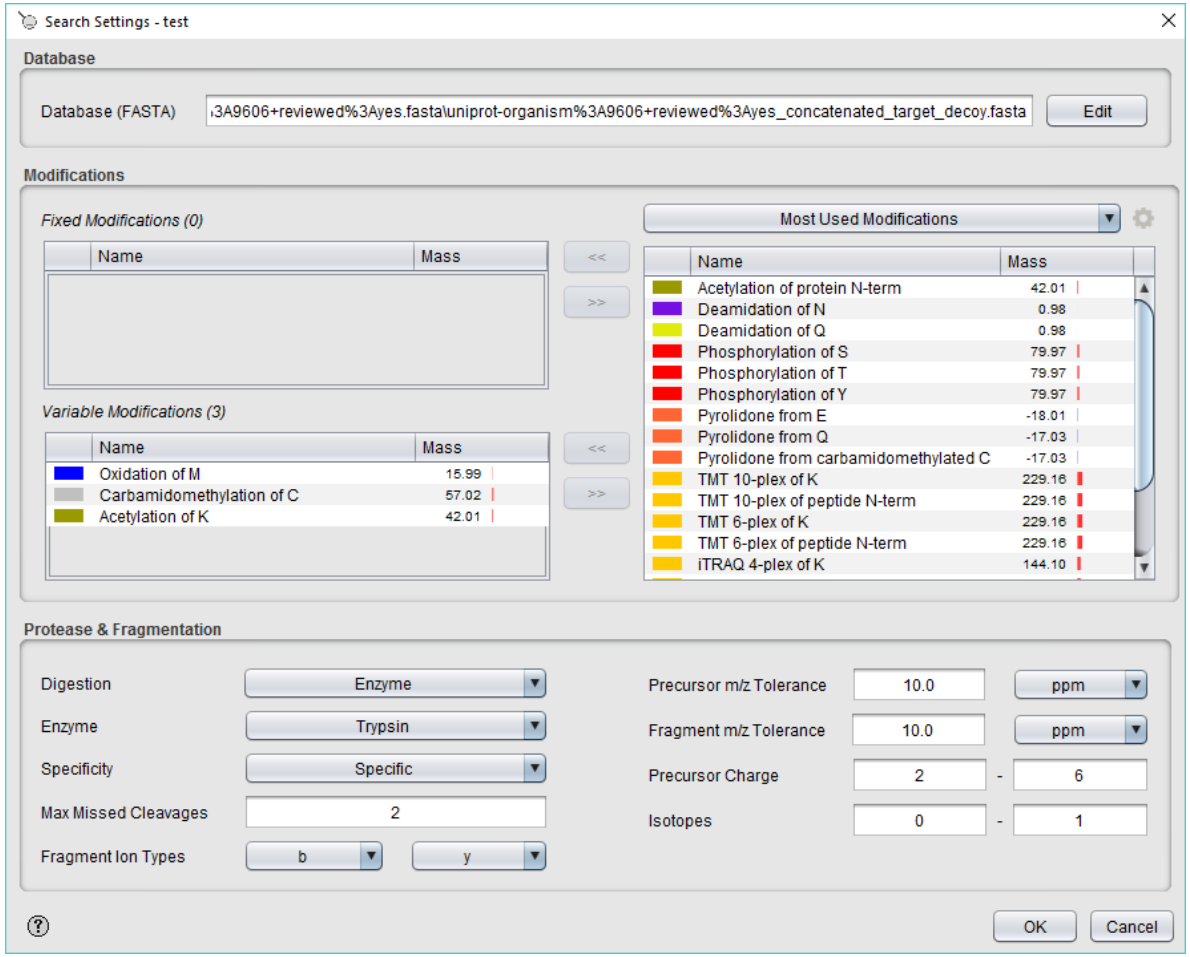
1) Open the searchGUI.
2) In the search setting select edit and fill the form as shown.
3) Variable modification were chosen (why?).
4) You can also configure the peptide shaker to open the results file, as shown bellow.
5) The searchGUI will ask you if you want to create concatenated_target_decoy fasta file. Say yes (Why is this important?)



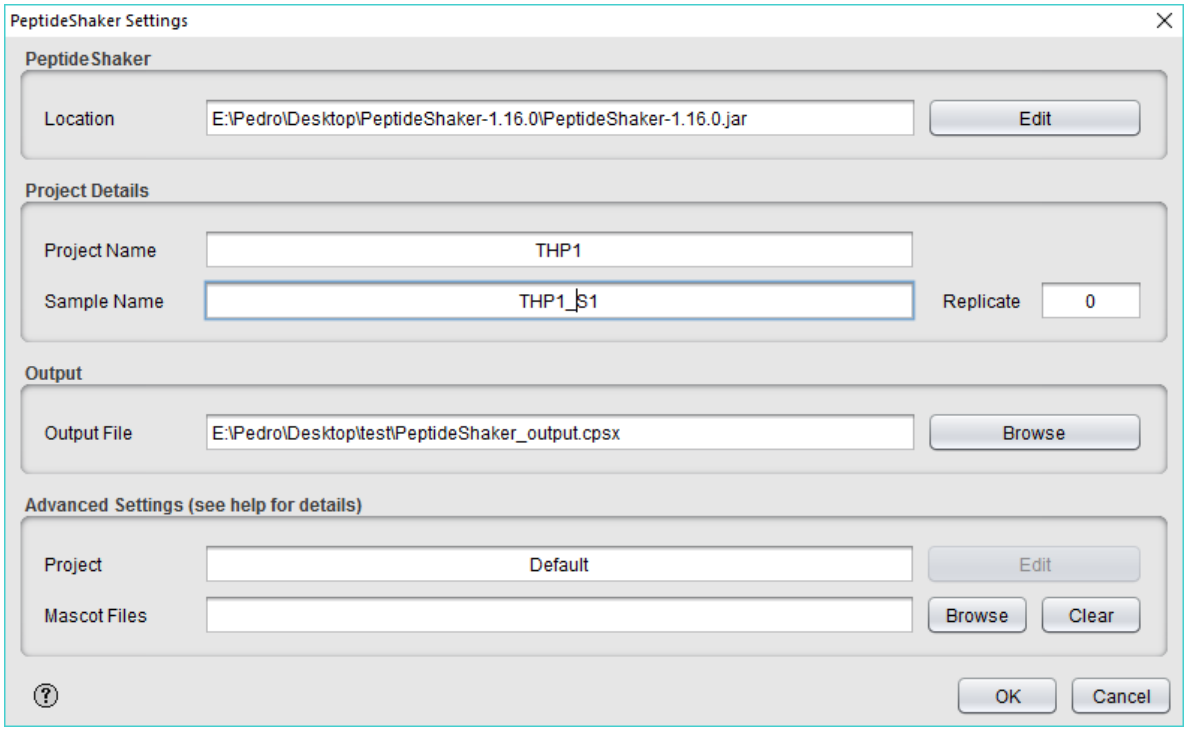
Select the search options and the FASTA database

Variable modification

were chosen (why?).

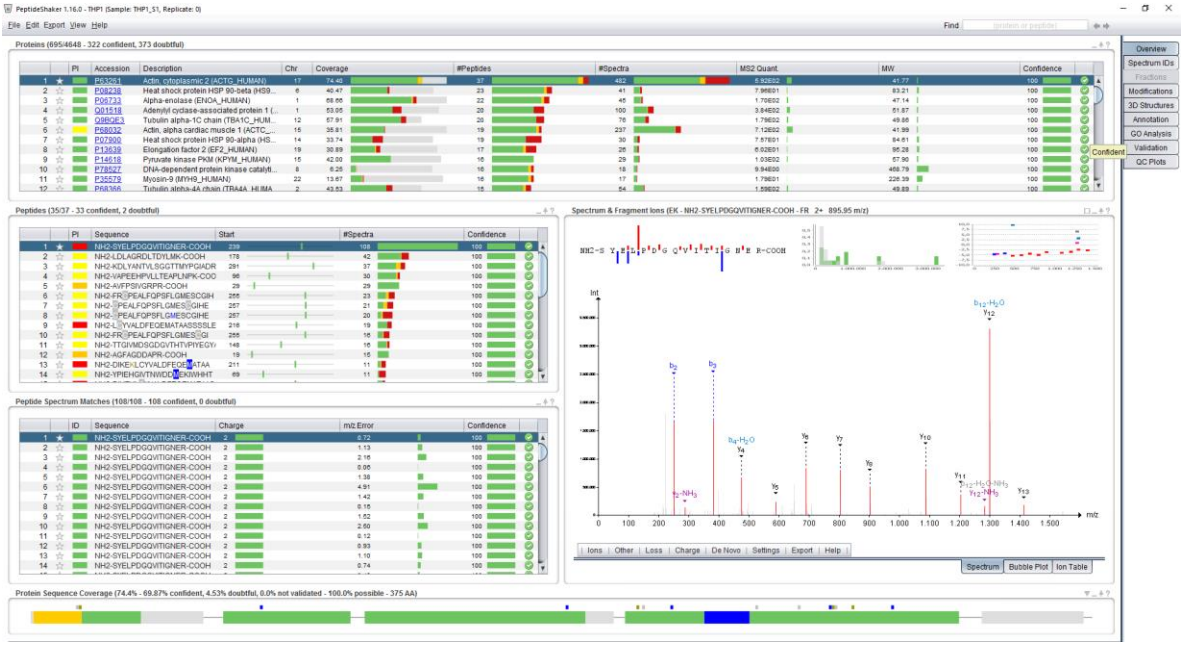


Select the peptideshaker options

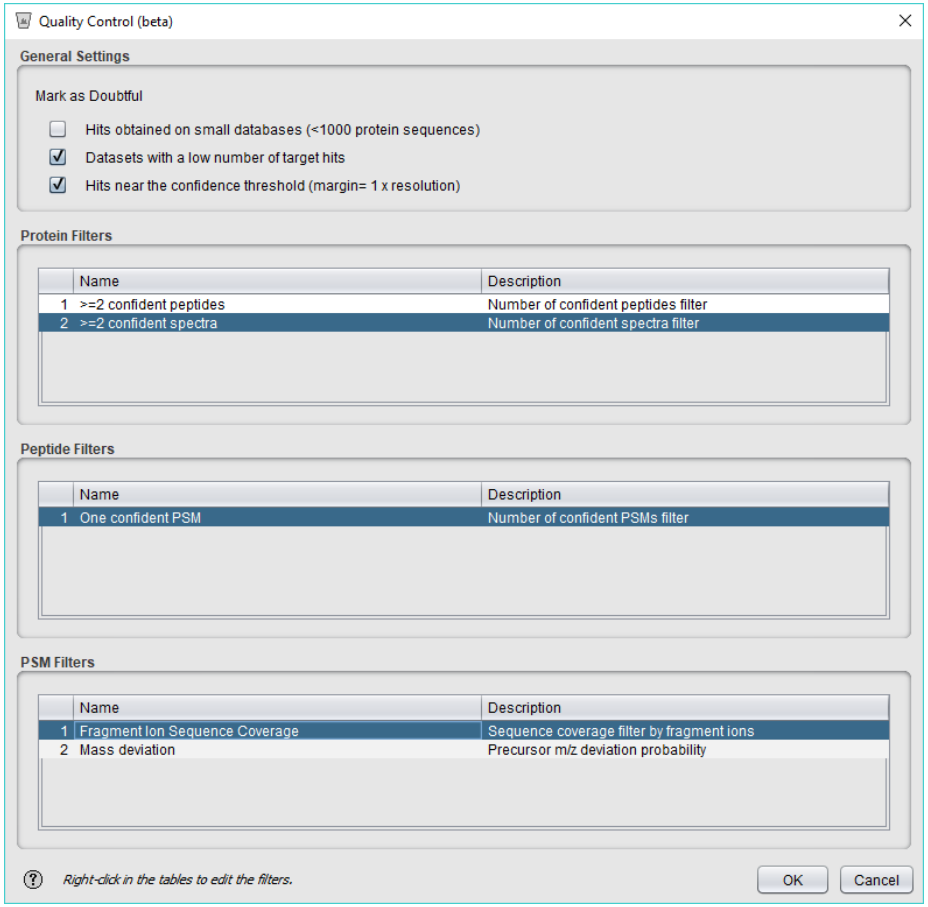


Generation and evaluation of results: PeptideShaker for peptide and protein visualization, and validation. PTM analysis

1) After the searchGUI has performed the search
   - (~3 minutes with an Intel I7-6700K with 16MB of RAM), it will open the results in the peptideshaker platform.
2) Here you will be able to see that ~700 proteins were identified, although 373 have been classified has doubtful (why?).
3) Also, you will be able to see information about the peptides identified for each protein and the mass spectra with the annotated fragmentation pattern.

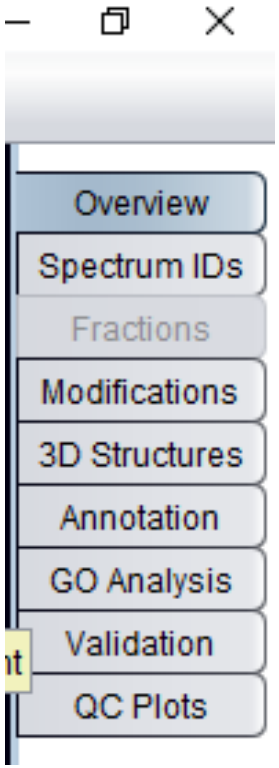


A **peptide-spectrum match (PSM)**



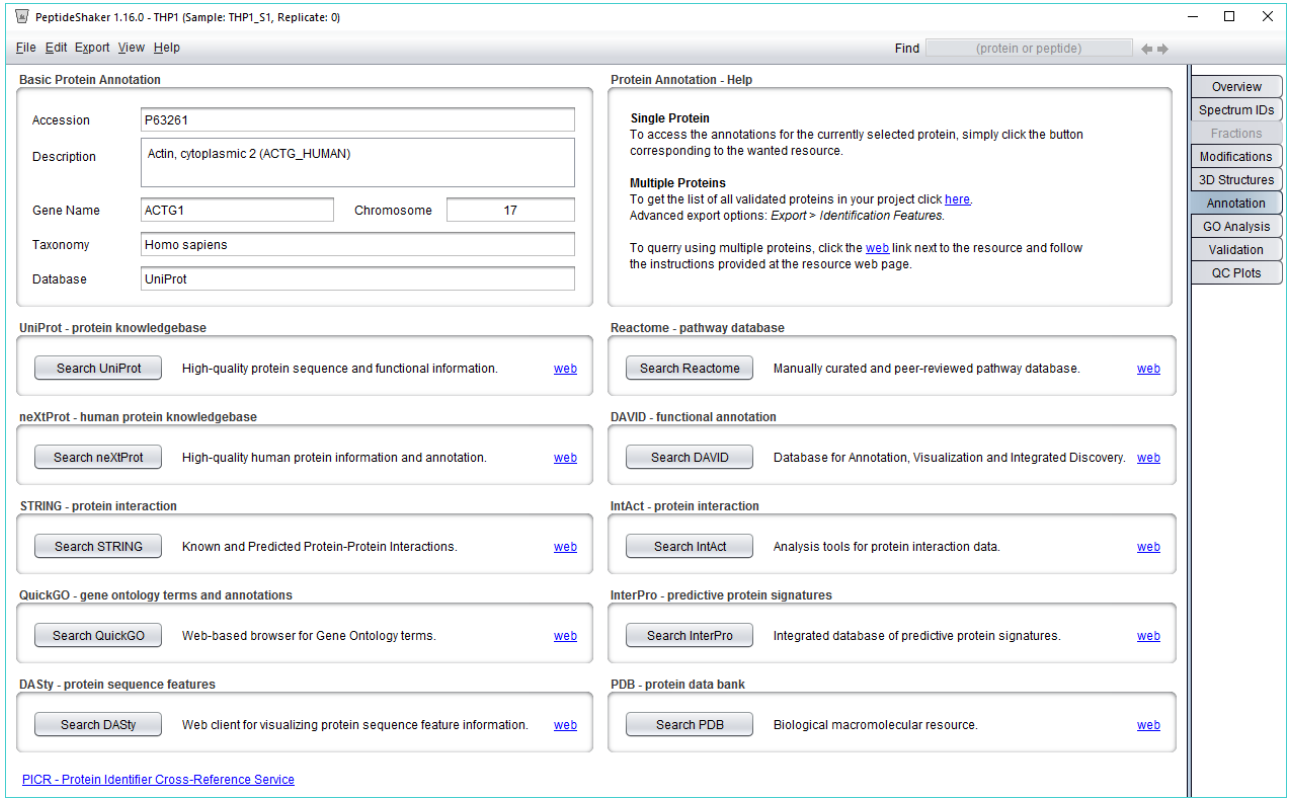
Data analysis: protein information, pathway analysis, and gene ontology

Explore the advanced data analysis options of the peptideshaker by opening the modifications tab, and the GO analysis tab.

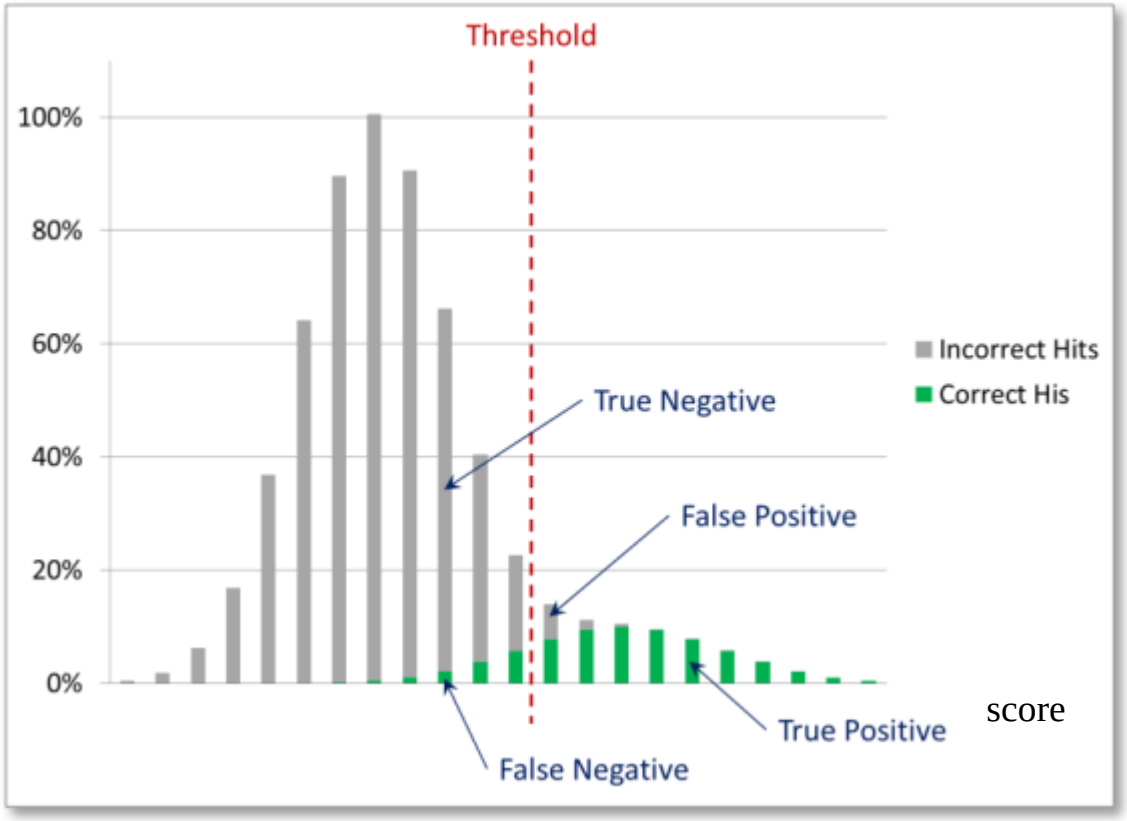


Data analysis: protein information, pathway analysis, and gene ontology

In the annotation tab you will be able to annotate information for each protein. However, if you wish to annotate multiple proteins, you will need to export your results (default protein report) and click the web link next to the resource and follow the instructions provided at the resource web page.

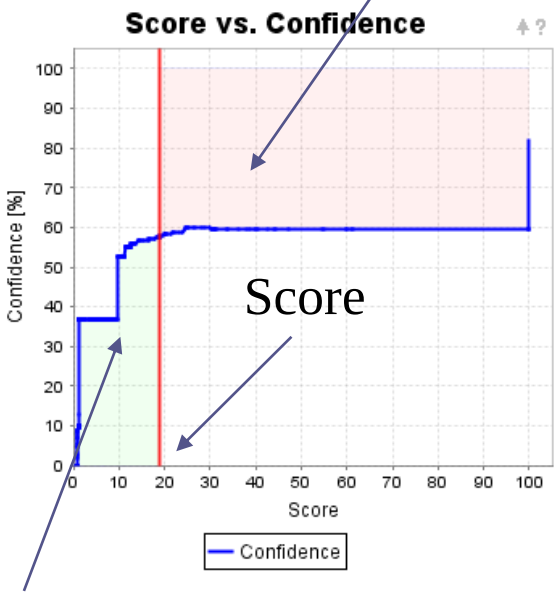


Validation


Threshold
100%
80%
60%
40%
20%
0%
Incorrect Hits
Correct His
True Negative
False Positive
True Positive
False Negative
score

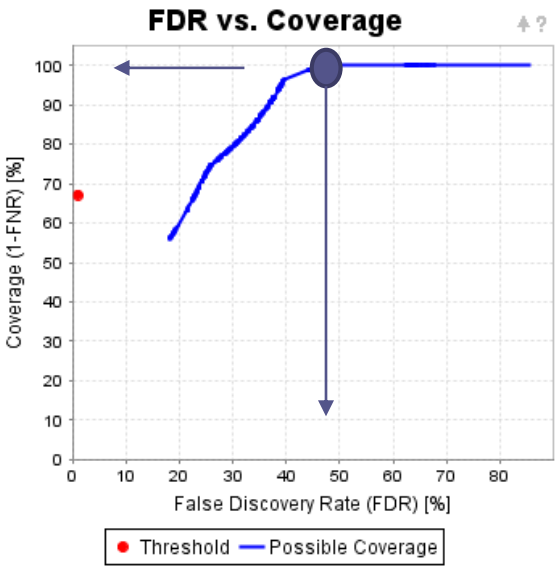
Validation

True Positives



False negatives

Number of target and decoy proteins detected

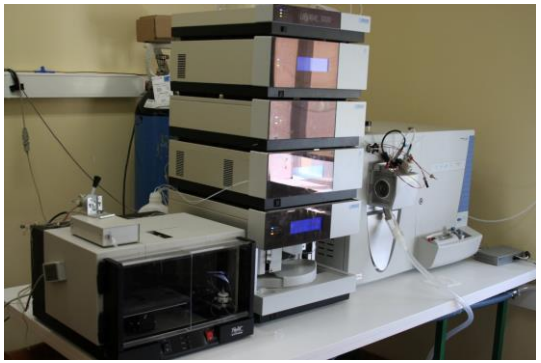


To obtain 100% coverage, one would need a 50% FRD

Report

Show the results of this analysis:

- Number of proteins identified
- Top 3 most abundant proteins
- Select a group based on GO analysis
  - Show the string analysis (functional protein association networks) of this group
  - Annotate one important protein (central node)



This project has been funded with support from the European Commission.

This publication reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

Creative Commons Licence

AACLifeSci

Erasmus+