UNIVERSITÀ CATTOLICA del Sacro Cuore

MS-based metabolite profiling in plant

luigi.lucini@unicatt.it

UNIVERSITÀ CATTOLICA del Sacro Cuore

What is a Metabolite?

- Any organic molecule detectable in a system with a MW < 1500 Da
- Includes peptides, oligonucleotides, sugars, nucelosides, organic acids, ketones, aldehydes, amines, amino acids, lipids, steroids, alkaloids, foods, chemical additives, toxins, pollutants,
- Includes human & microbial products
- Concentration > detectable (1 pM)

UNIVERSITÀ CATTOLICA del Sacro Cuore

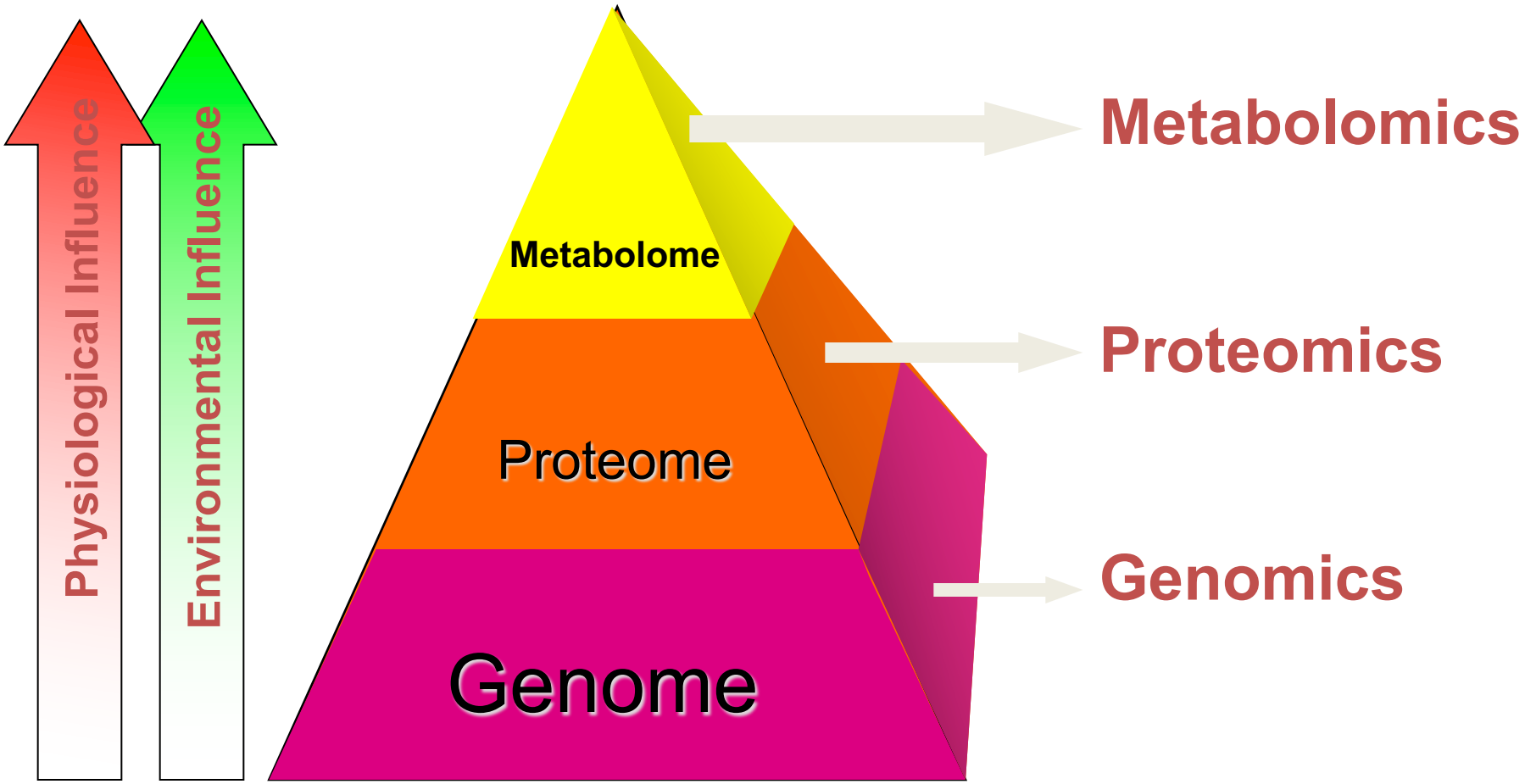
Metabolite ID

- Two scenarios identification of “known unknowns” and “unknown unknowns”
- For “known unknowns” use spectral or metabolite libraries to ID and quantify via spectral deconvolution
- For “unknown unknowns” (truly novel) use computer-aided structure elucidation methods (CASE)

“…there are known unknowns; that is to say we know there are some things we do not know. But there are also unknown unknowns -- the ones we don't know we don't know.”

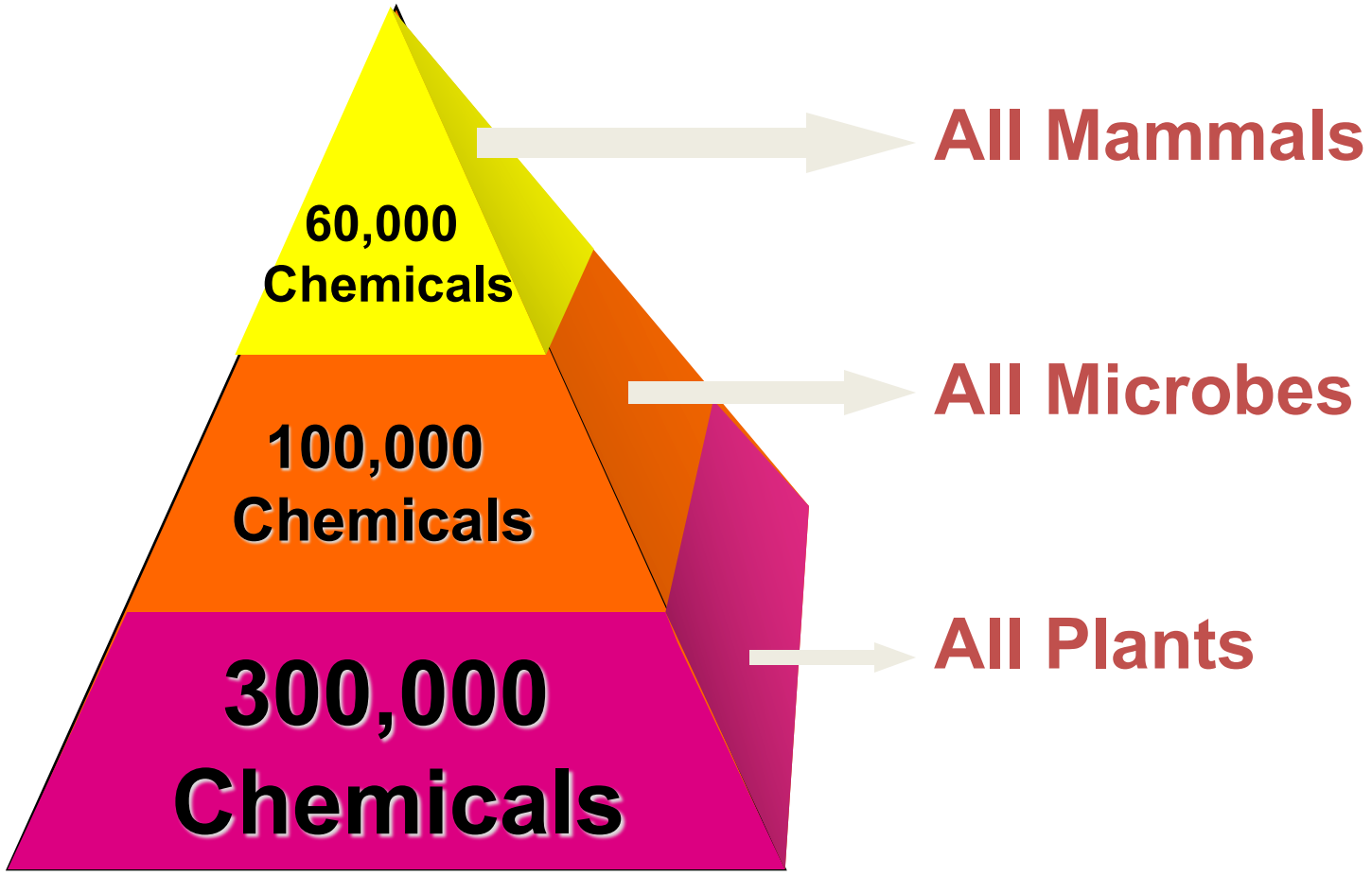
UNIVERSITÀ CATTOLICA del Sacro Cuore

The Pyramid of Life



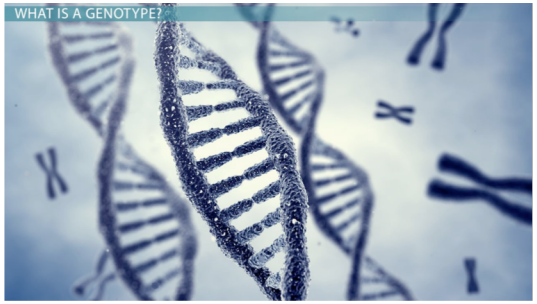
UNIVERSITÀ CATTOLICA del Sacro Cuore

Different Metabolomes



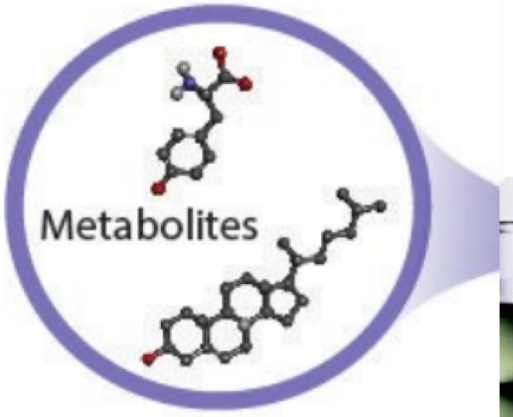
The Pyramid of Life

Chemical fingerprints



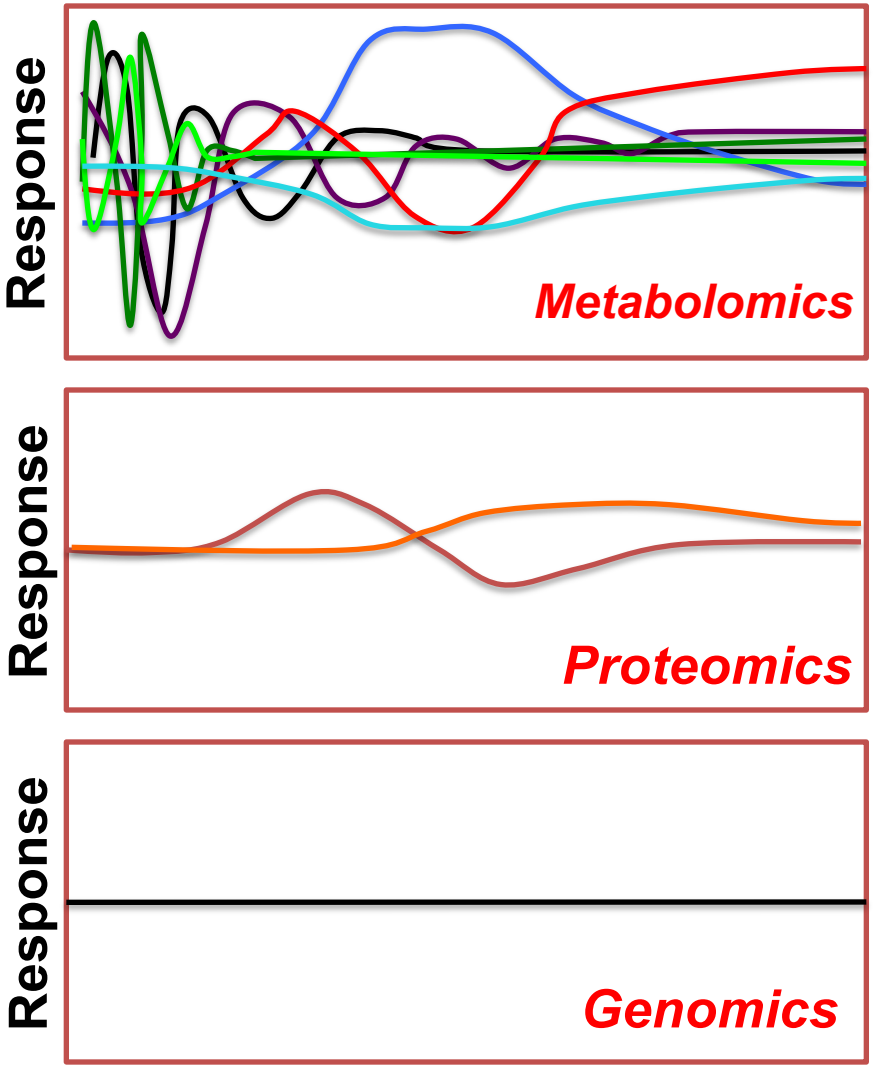
x

=



UNIVERSITÀ CATTOLICA del Sacro Cuore

Metabolomics is More Time Sensitive Than Other "Omics"



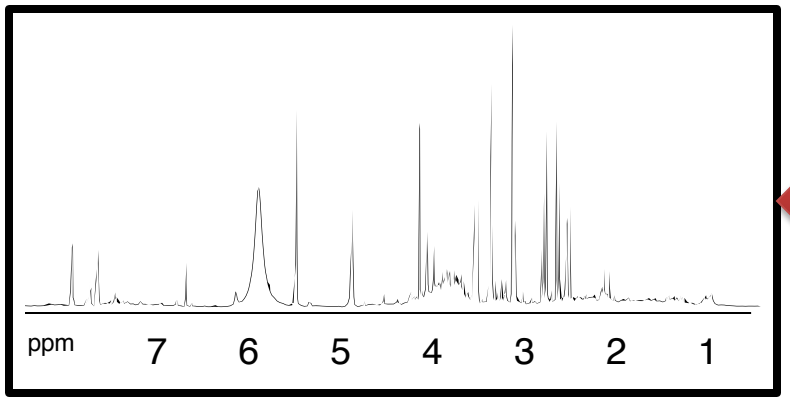
UNIVERSITÀ CATTOLICA del Sacro Cuore

Metabolomics Workflow

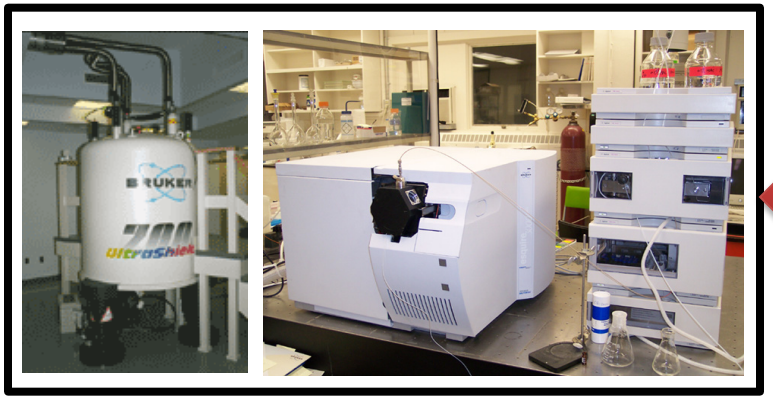
Biological or Tissue Samples

Extraction

Extracts



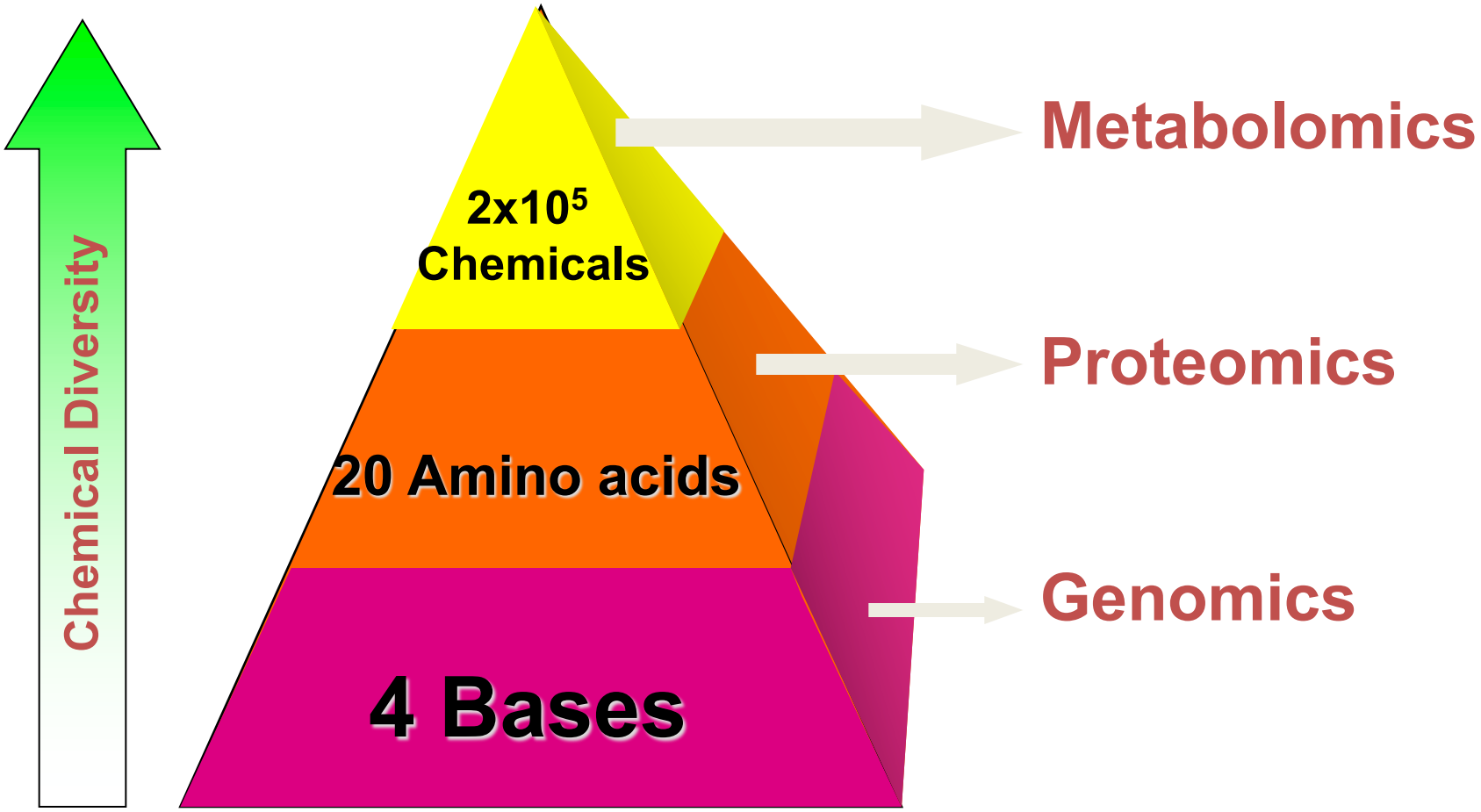
Data Analysis



Chemical Analysis

UNIVERSITÀ CATTOLICA del Sacro Cuore

Chemical diversity



The Pyramid of Life

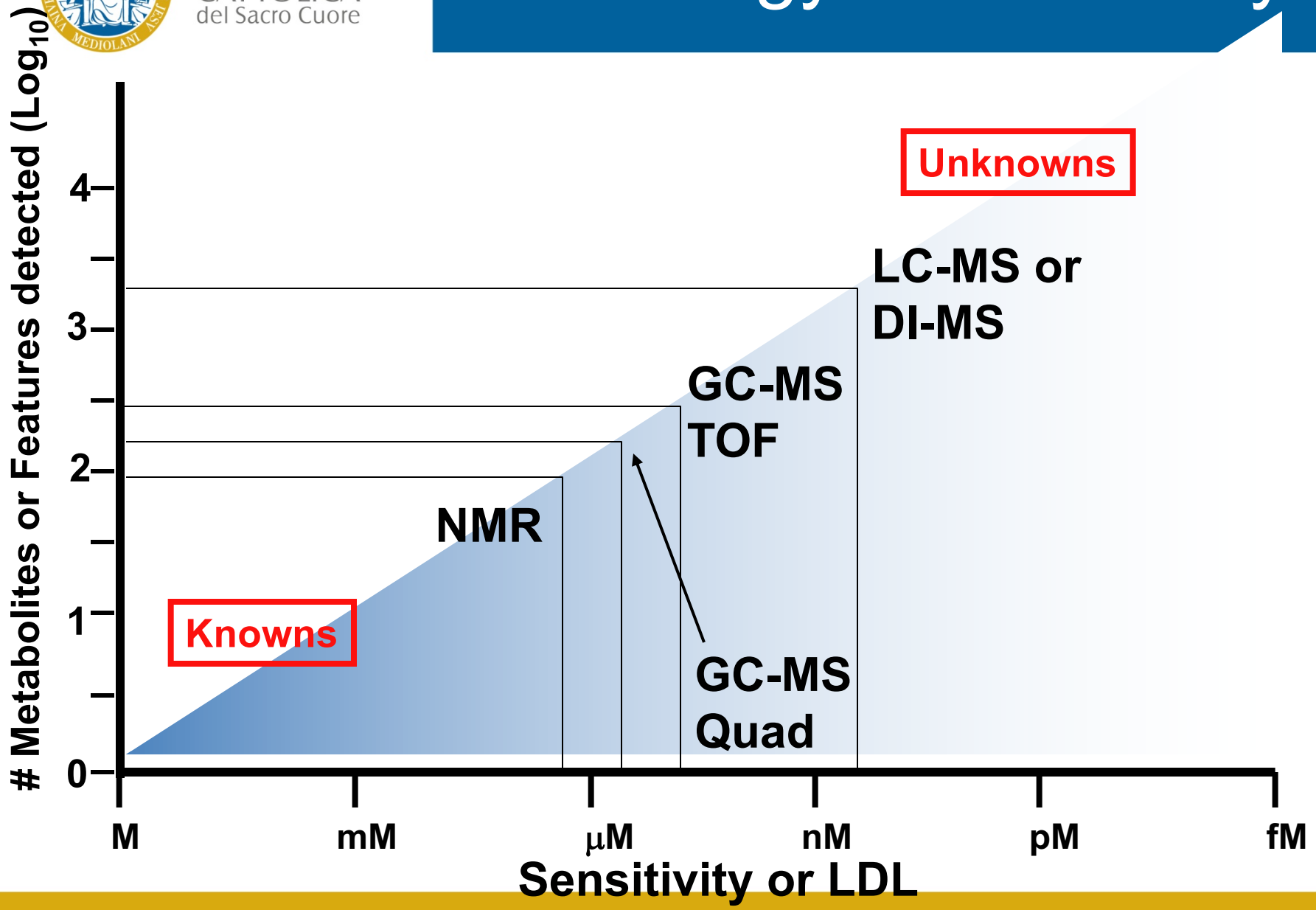
UNIVERSITÀ CATTOLICA del Sacro Cuore

Metabolomics Technologies

- UPLC, HPLC
- CE/microfluidics
- LC-MS
- FT-MS
- QqQ-MS
- NMR spectroscopy
- X-ray crystallography
- GC-MS
- FTIR

UNIVERSITÀ CATTOLICA del Sacro Cuore

Technology & Sensitivity

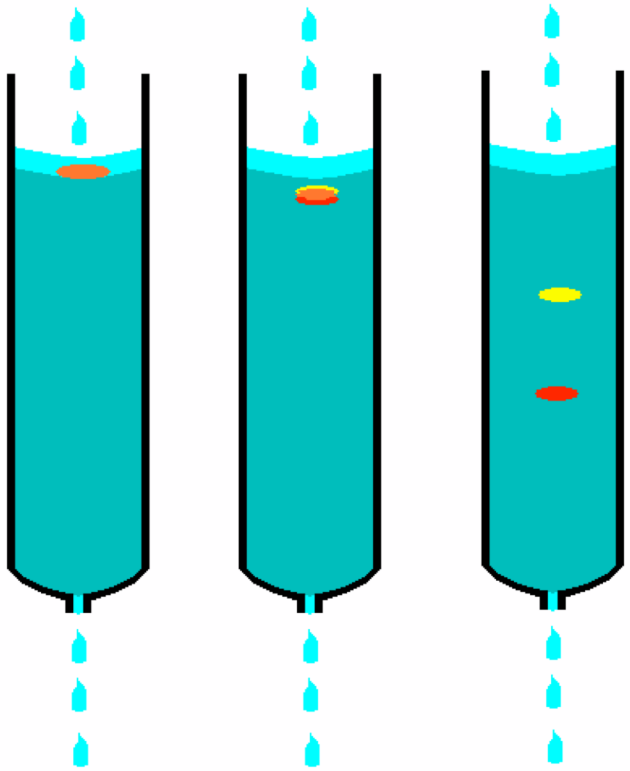


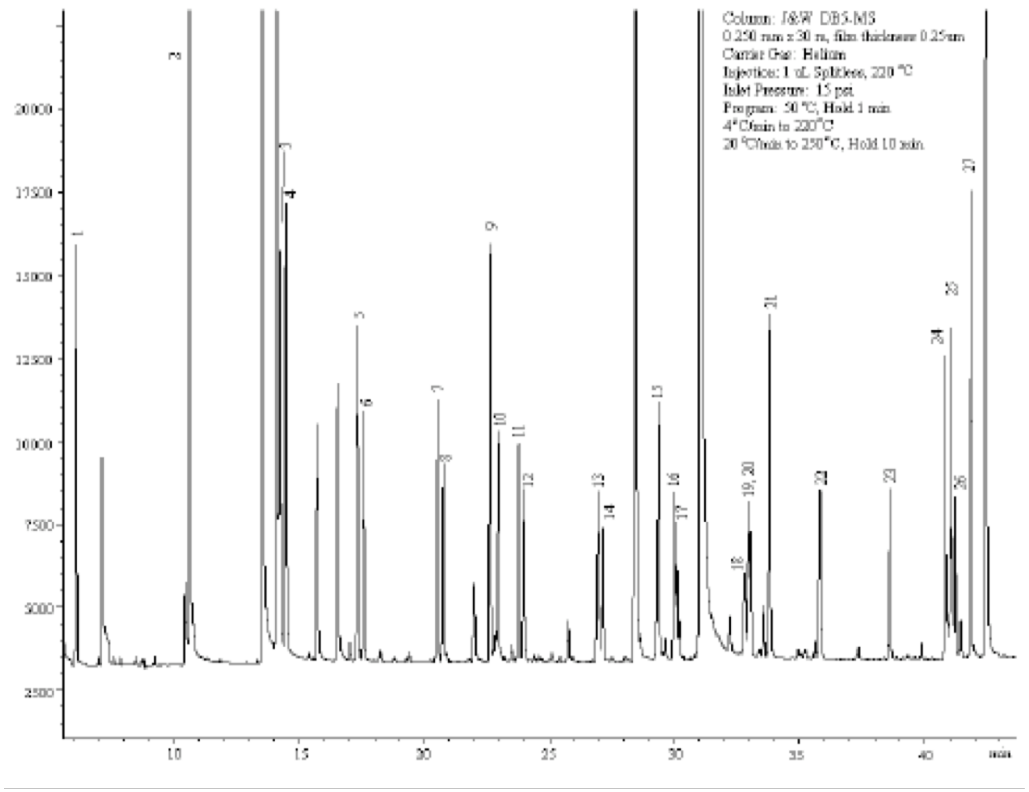
High resolution MS metabolomics

(UHPLC-QTOF)

UNIVERSITÀ CATTOLICA del Sacro Cuore

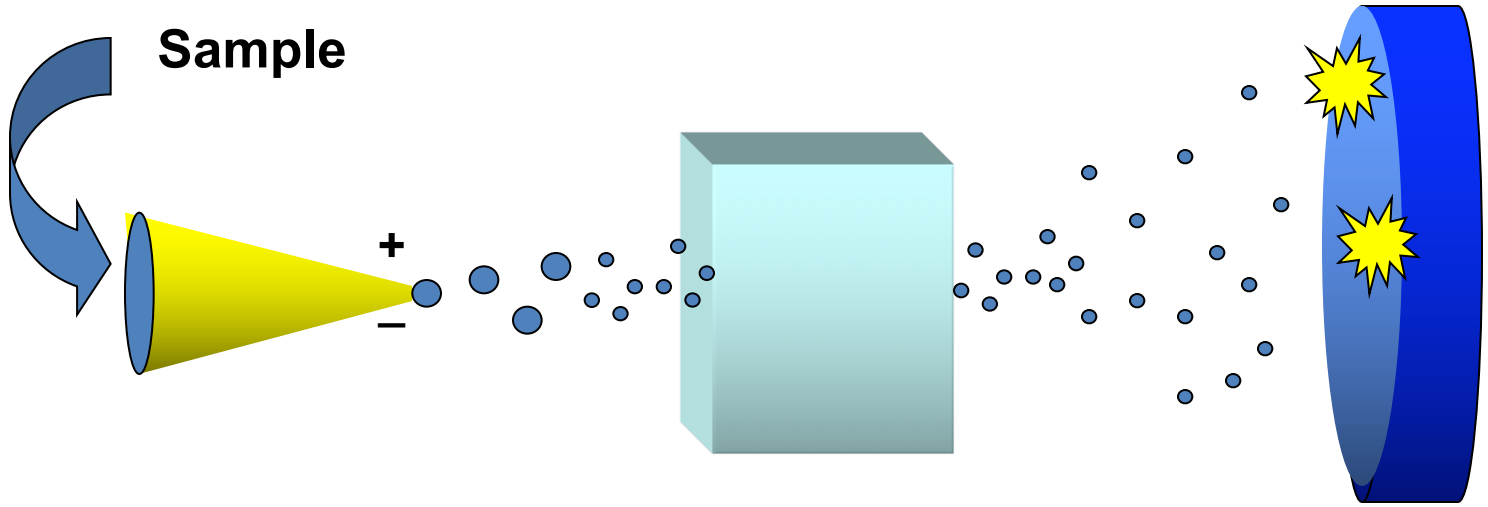
Chromatography





UNIVERSITÀ CATTOLICA del Sacro Cuore

Mass Spec Principles



UNIVERSITÀ CATTOLICA del Sacro Cuore

Mass Analyzers

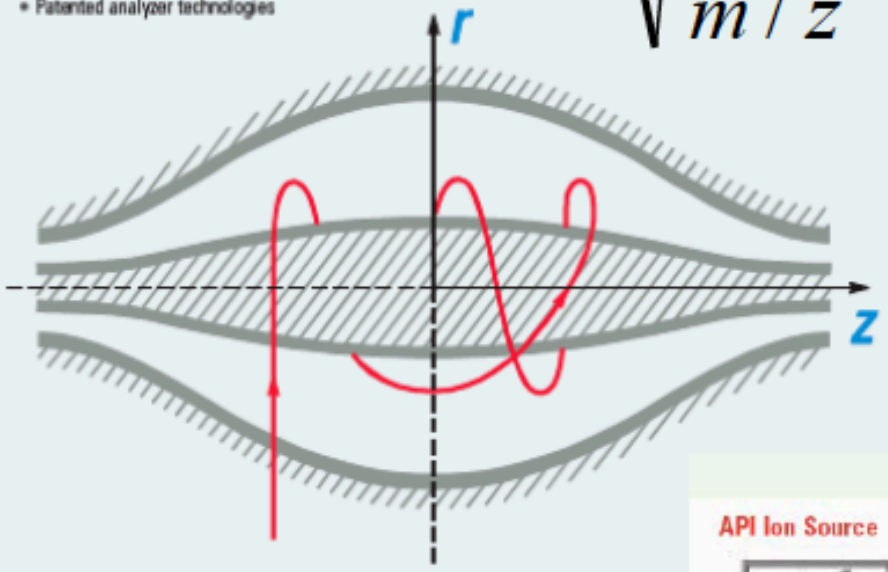
- Magnetic Sector Analyzer (MSA)
  - High resolution, exact mass, original MA
- Quadrupole Analyzer (Q or Q*)
  - Low (1 amu) resolution, fast, cheap
- Time-of-Flight Analyzer (TOF)
  - No upper m/z limit, high throughput
- Ion Cyclotron Resonance (FT-ICR)
  - Highest resolution, exact mass, costly

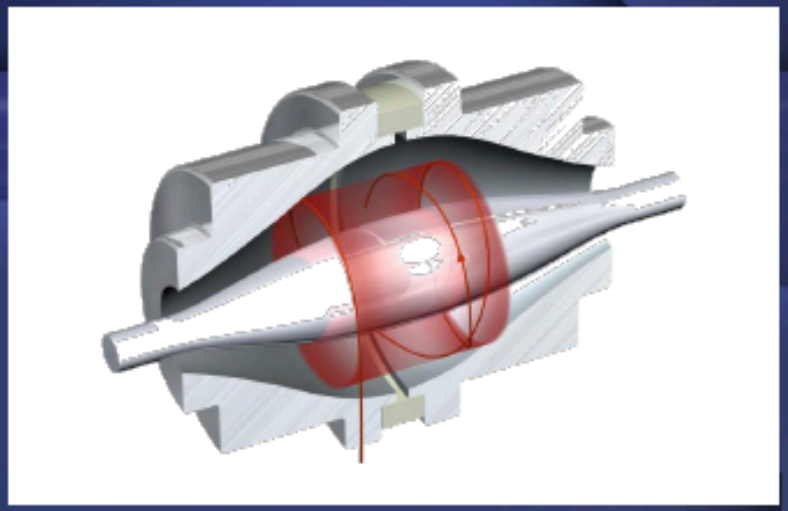
Orbitrap Mass Analyzer

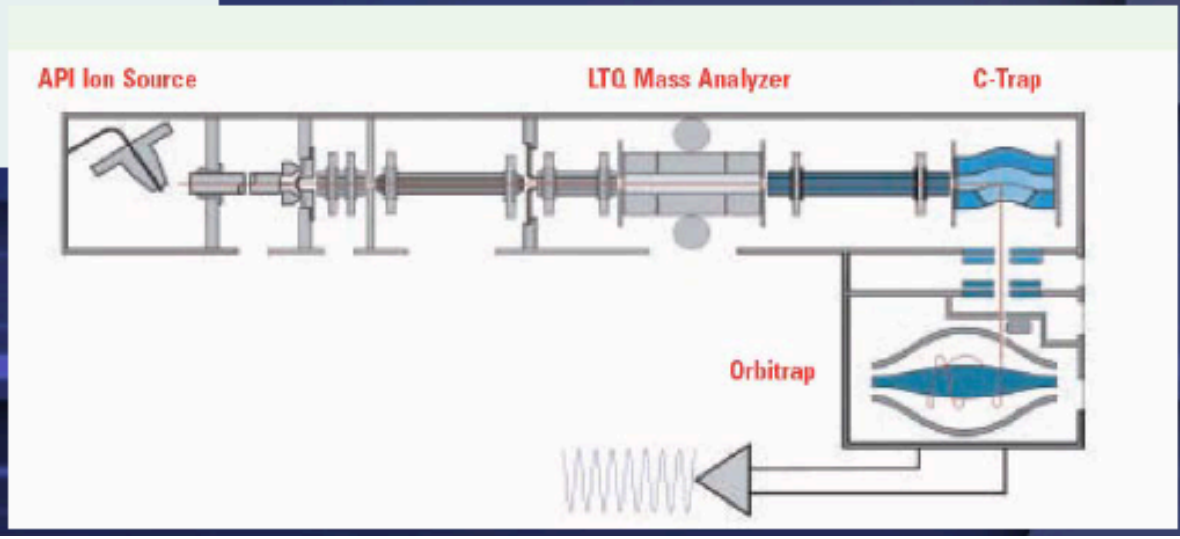
Orbitrap: A Breakthrough Electrostatic Ion Trap

- Highest ion trapping efficiencies
- Large ion capacity
- Stable and robust operation
- Patented analyzer technologies

$$\omega = \sqrt{\frac{k}{m/z}}$$

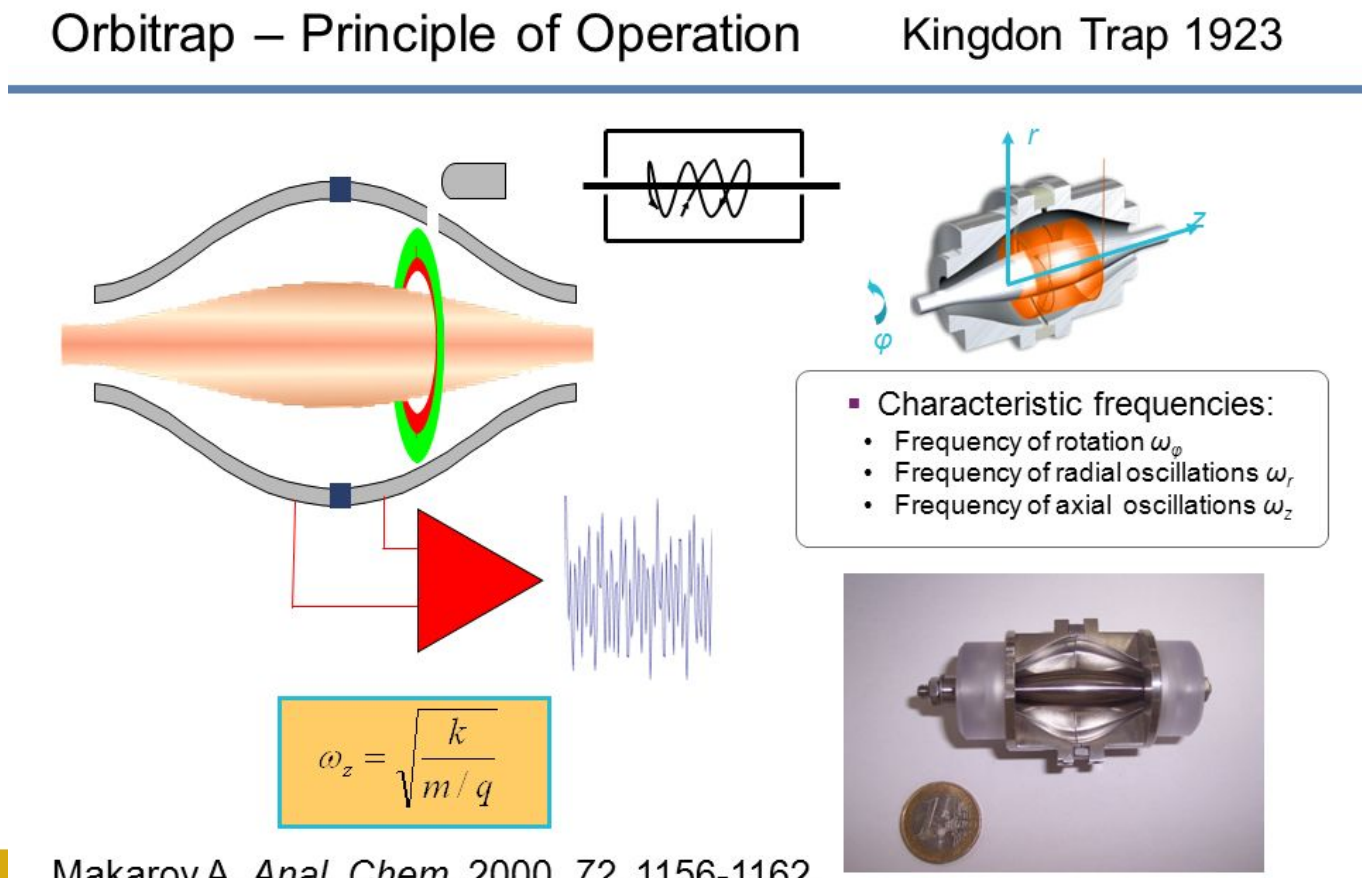






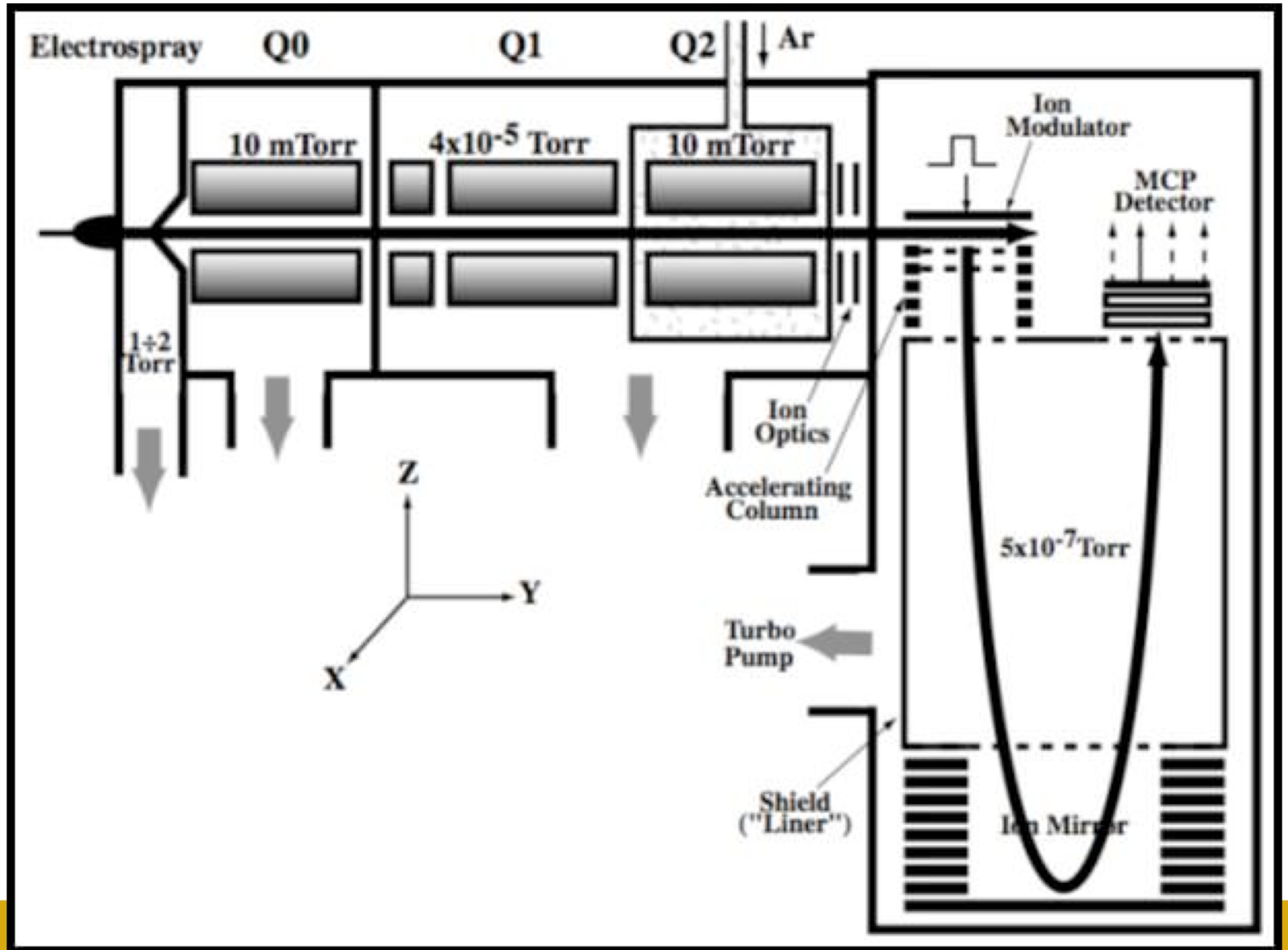
orbitrap

The orbital mass analyser is based on the orbital trapping of ions.
Injected ions cycle around the central electrode and at the same time oscillate along the horizontal axis



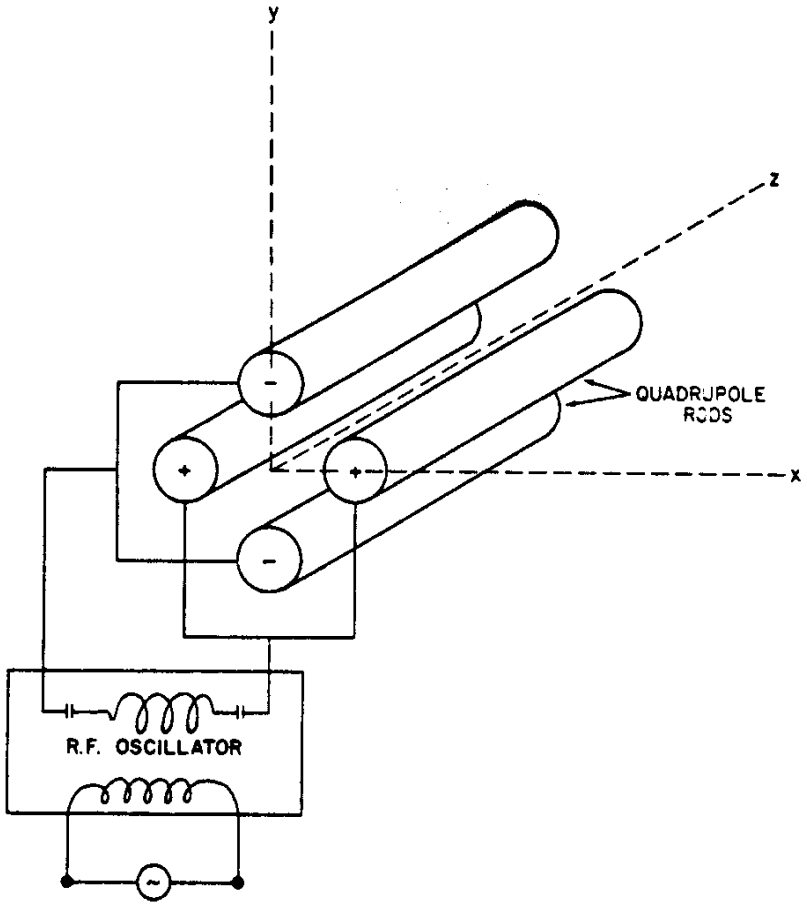
UNIVERSITÀ CATTOLICA del Sacro Cuore

QTOF



UNIVERSITÀ CATTOLICA del Sacro Cuore

Quadrupole Mass Analyzer



Uses a combination of RF and DC voltages to operate as a mass filter.

- Has four parallel metal rods.
- Lets one mass pass through at a time.
- Can scan through all masses or sit at one fixed mass.

UNIVERSITÀ CATTOLICA del Sacro Cuore

Quadrupoles have variable ion transmission modes

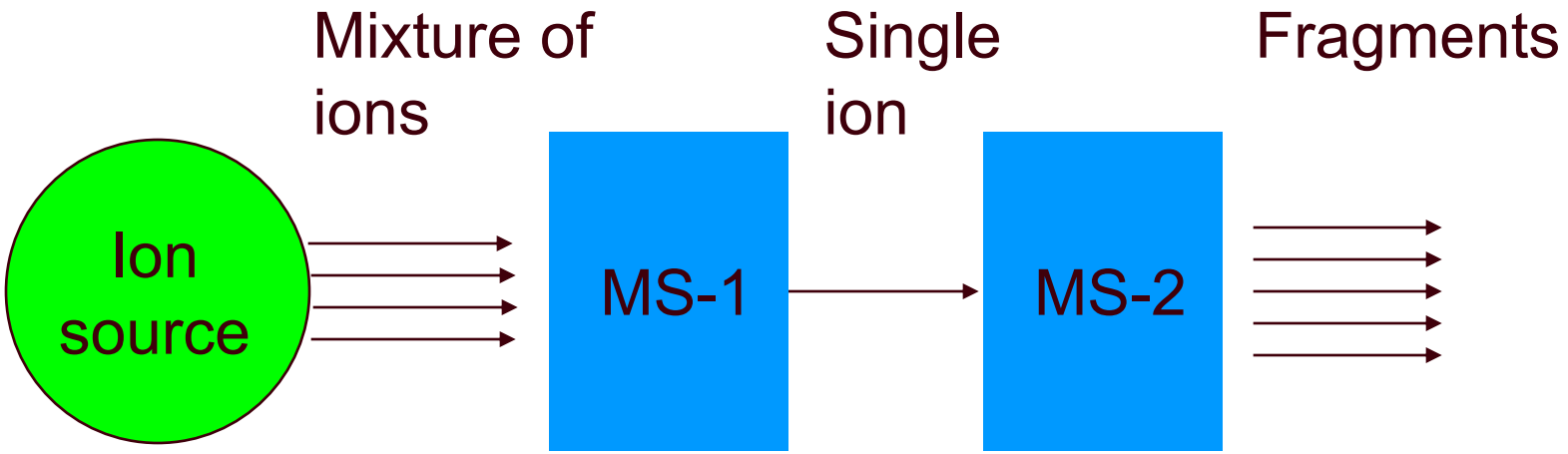
mass scanning mode

single mass transmission mode

UNIVERSITÀ CATTOLICA del Sacro Cuore

What is MSMS?

MS/MS means using two mass analyzers (combined in one instrument) to select an analyte (ion) from a mixture, then generate fragments from it to give structural information.

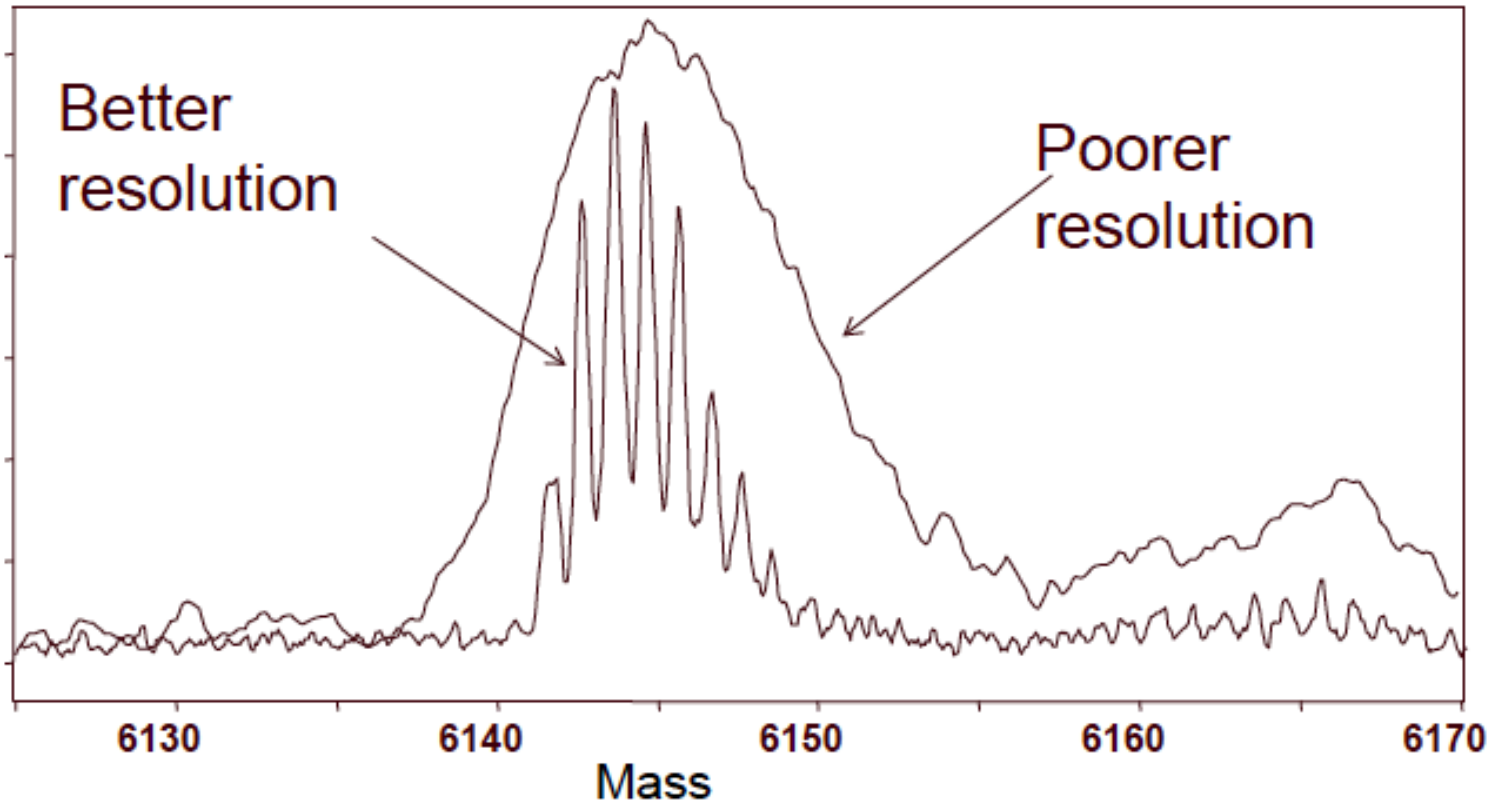


UNIVERSITÀ CATTOLICA del Sacro Cuore

Why HRMS?

What if the resolution is not so good?

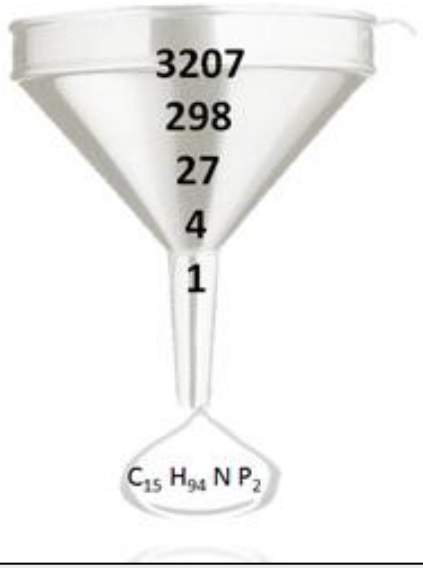
At lower resolution, the mass measured is the average mass.



UNIVERSITÀ CATTOLICA del Sacro Cuore

High Resolution MS

Massa isotopica	n. isomeri possibili	Risoluzione
350	3207	400
350,6	298	3500
350,68	27	7000
350,686	4	11500
350,6861	1	44000



Profilo metabolico

- **Tracciabilità**
- **Adulterazione**
- **Trattamenti tecnologici**
- **Biomarkers**
- **…**

UNIVERSITÀ CATTOLICA del Sacro Cuore

Why HRMS?

Isotopes

+Most elements have more than one stable isotope.

For example, most carbon atoms have a mass of 12 Da, but in nature, 1.1% of C atoms have an extra neutron, making their mass 13 Da.

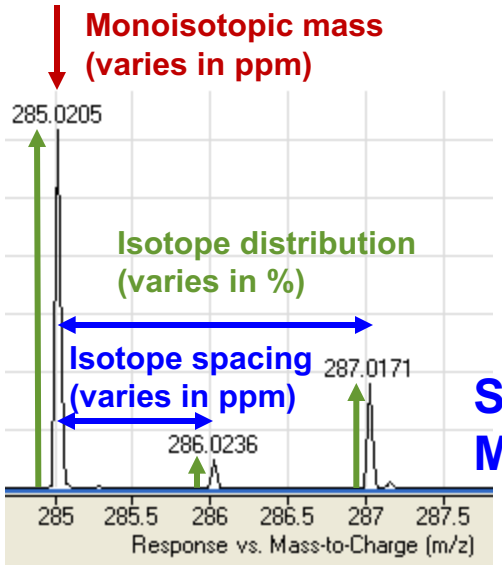
+Why do we care?

Mass spectrometers can “see” isotope peaks if their resolution is high enough.

If an MS instrument has resolution high enough to resolve these isotopes, better mass accuracy is achieved.

Identification / screening

Scoring based on:



Mass Match +

Abund. Match +

Spacing Match =

Overall Score

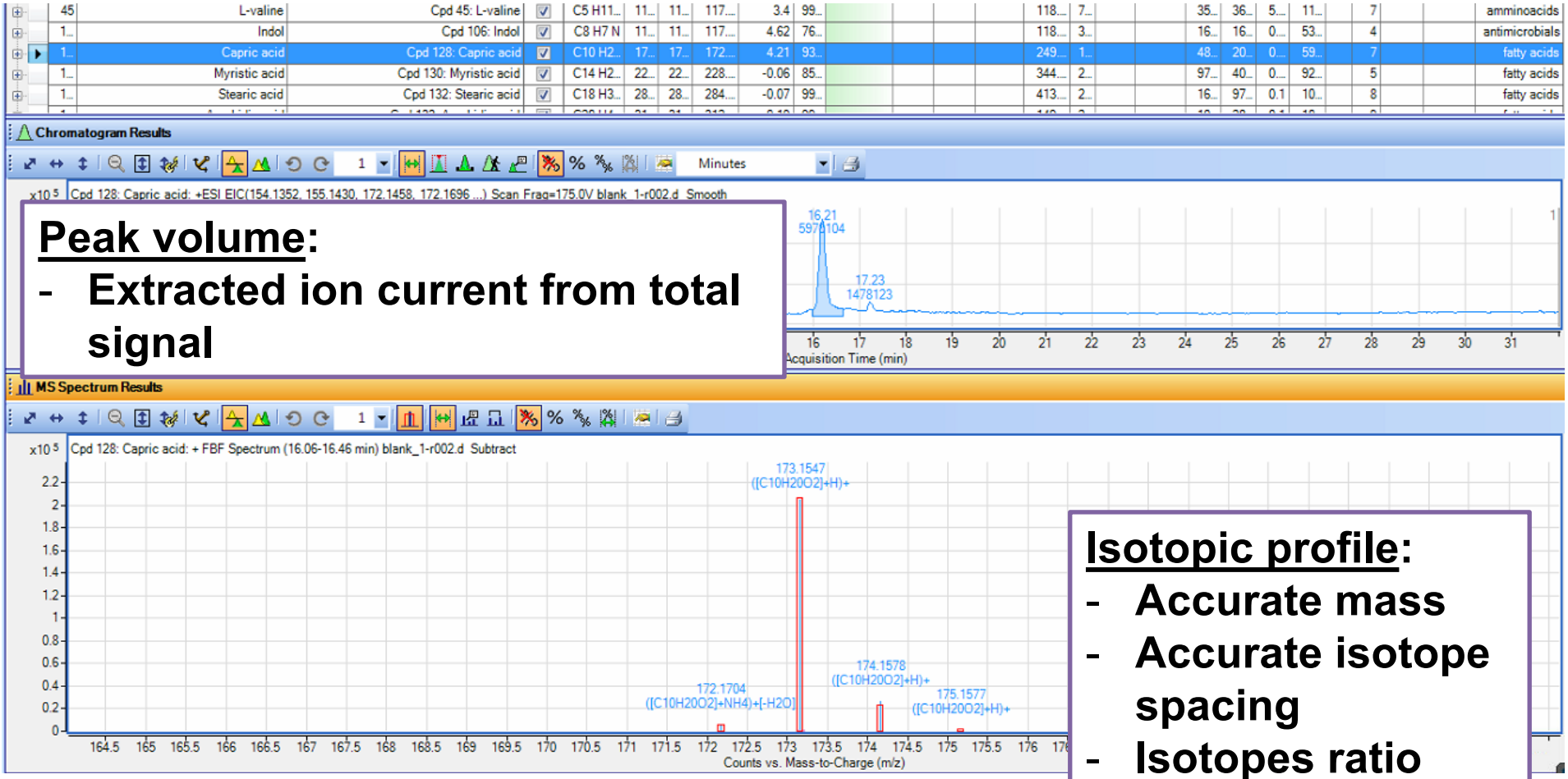
MS Formula Results: Cpd 2: C10H9ClN4O2S

m/z	Ion	Formula	Abundance
285.021	(M+H)+	C10 H10 Cl N4 O2 S	24506.1

Best	Formula (M)	Calc m/z	Score	Cross Score	Mass	Calc Mass	Diff (ppm)	Abs Diff (p	Spacing Matc	Abund Matc	Mass Match	m/z	DBE
☑	C10 H9 Cl N4 O2 S	285.0208	99.55		284.0137	284.0135	-0.71	0.71	99.19	99.26	99.69	285.021	8
☐	C7 H12 N2 O6 S2	285.021	77.28		284.0137	284.0137	0.01	0.01	99.54	1.93	100	285.021	3
☐	C7 H13 Cl N4 O2 S2	285.0241	75.57		284.0137	284.0168	11.12	11.12	99.87	83.87	46.22	285.021	3

UNIVERSITÀ CATTOLICA del Sacro Cuore

Identification / screening



Peak volume:

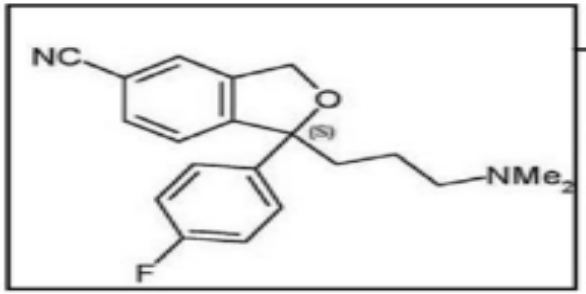
- **Extracted ion current from total signal**

Isotopic profile:

- **Accurate mass**
- **Accurate isotope spacing**
- **Isotopes ratio**

Text/Synonym Search

Structure Search



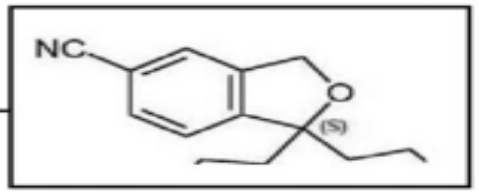
Mol Weight Search

Human Metabolome Database (HMDB)

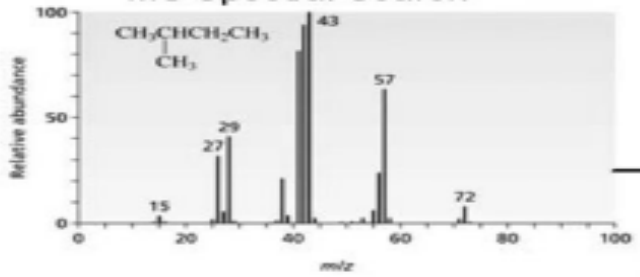
Smiles & Formula Search

CN1C=C(N=C1)

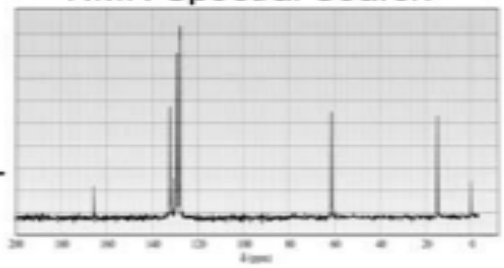
Substructure Search



MS Spectral Search



NMR Spectral Search



METQPRDSKNACDEG
LIGGEDCANKSDRPQ
VTREYQPASFHGFDL
PIYTREWQQNMLKAF

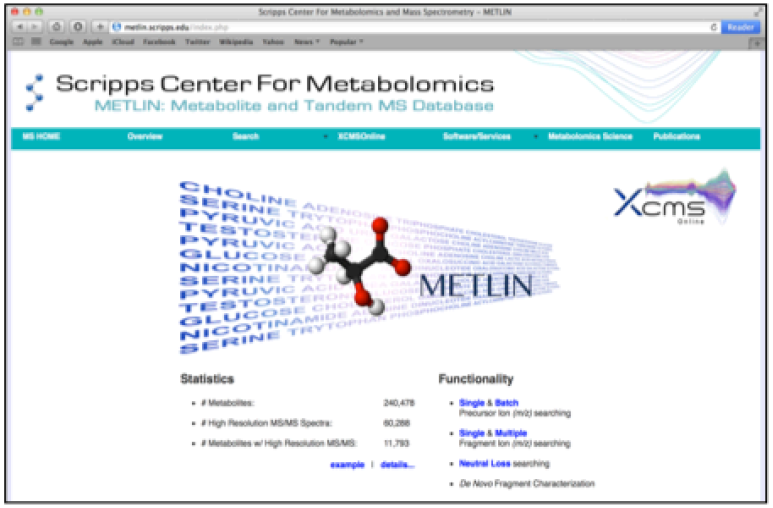
Sequence Search

UNIVERSITÀ CATTOLICA del Sacro Cuore

MS Spectral DBs

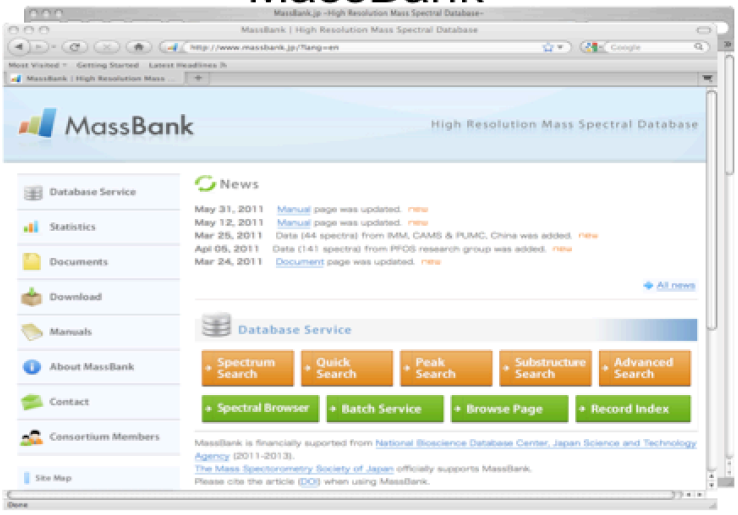
NIST/AMDIS

Metlin



GolmDB

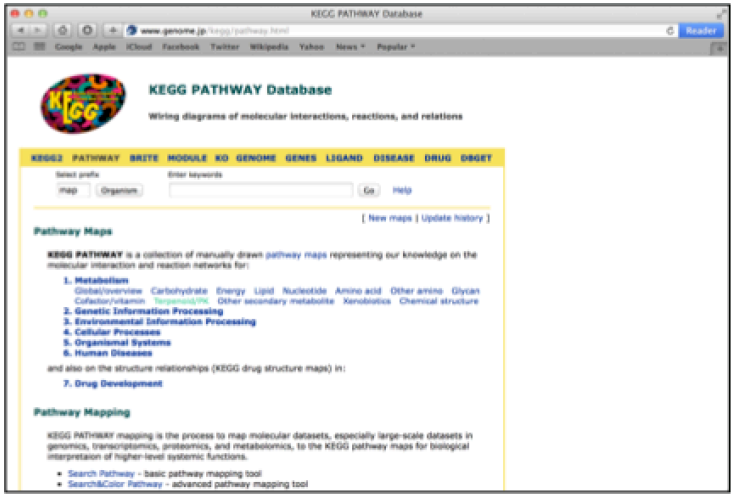
MassBank



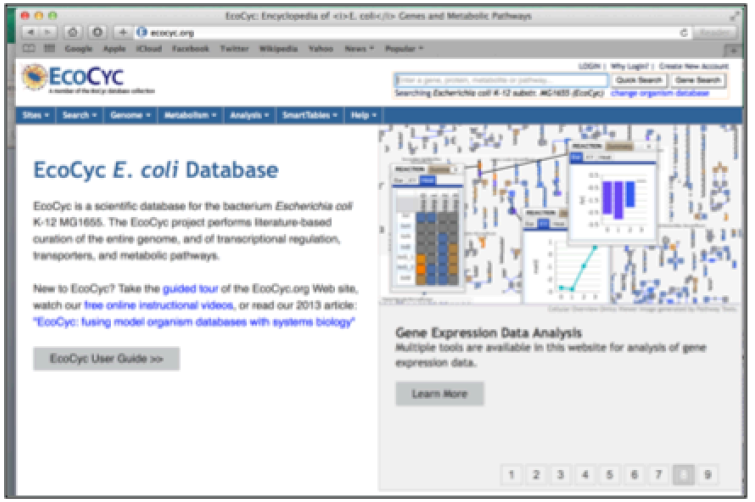
UNIVERSITÀ CATTOLICA del Sacro Cuore

Comprehensive MetDBs

KEGG

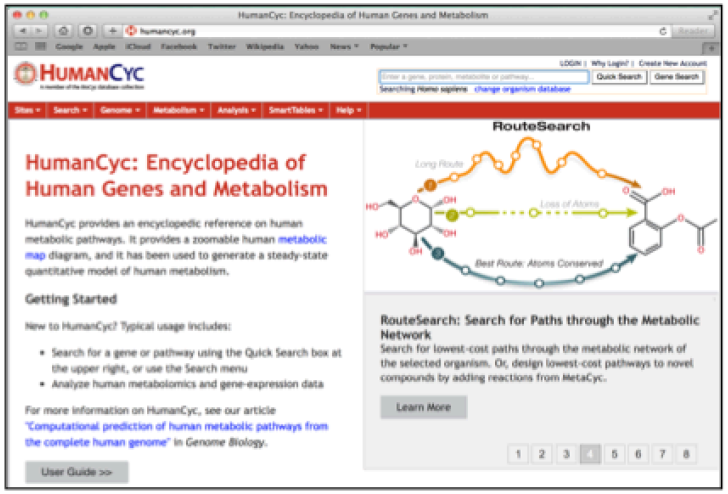


EcoCyc



MetaboLights

HumanCyc



UNIVERSITÀ CATTOLICA del Sacro Cuore

Naive vs forced approach

Forced: Find-by-Formula

- Needs a formula or a source of formula (database)
- Deeper in compounds fishing
- Higher number of false positive

Naive: Find-by-Molecular-Feature

- Needs an isotopic model (e.g. glycans, peptides, common organic molecules)
- Often requires a following identification step
- Lower number of false positives, higher number of false negative

UNIVERSITÀ CATTOLICA del Sacro Cuore

From meaningless to meaningfull data

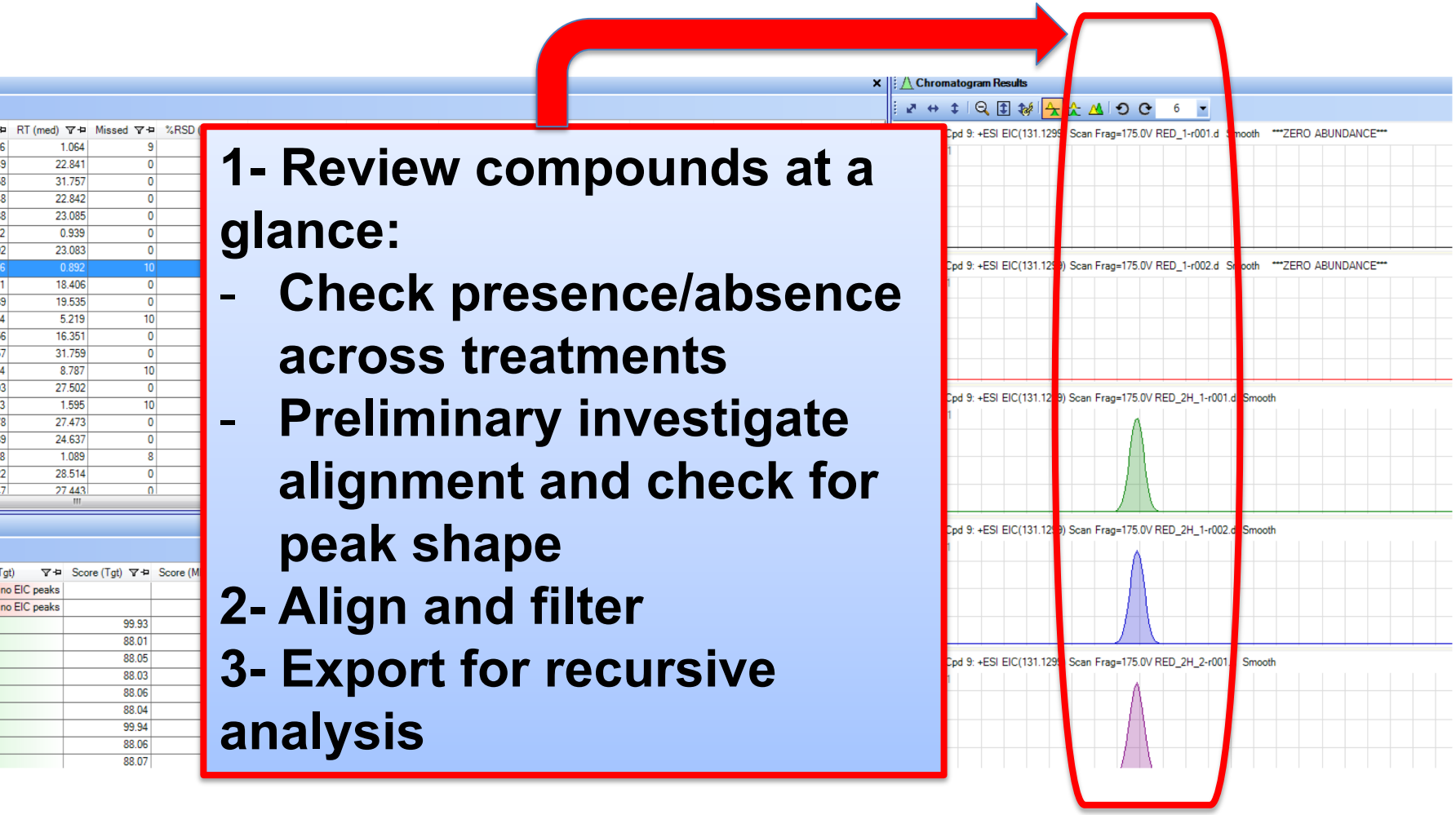
Raw data
- Huge amount of information
- Redundancy? False positive?

dataset
- Data gathering
- Alignment and filtering

Interpret.
- Statistics
- Chemometrics & bioinformatics

Answer(s) to a biological problem?

Alignment and filtering



1- Review compounds at a glance:

- **Check presence/absence across treatments**
- **Preliminary investigate alignment and check for peak shape**

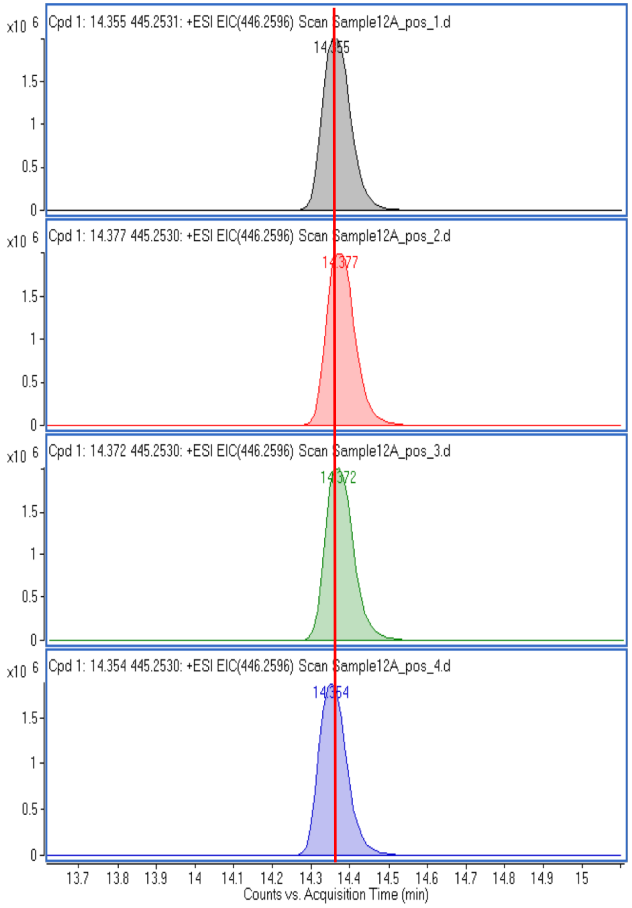
2- Align and filter

3- Export for recursive analysis

UNIVERSITÀ CATTOLICA del Sacro Cuore

Alignment and filtering in Mass Profiler Professional

1. **Alignment (optionally using internal standards) - Retention time tolerance**
2. **Filter by mass tolerance**
3. **Remove irreproducible compounds – Filter by Frequency**
4. **Remove weak compounds – Filter by Abundance**
5. **Remove highly variable compounds – Filter by Sample Variability**



SPECTRAL LIBRARIES BASED METABOLOMICS

(GC-MS)

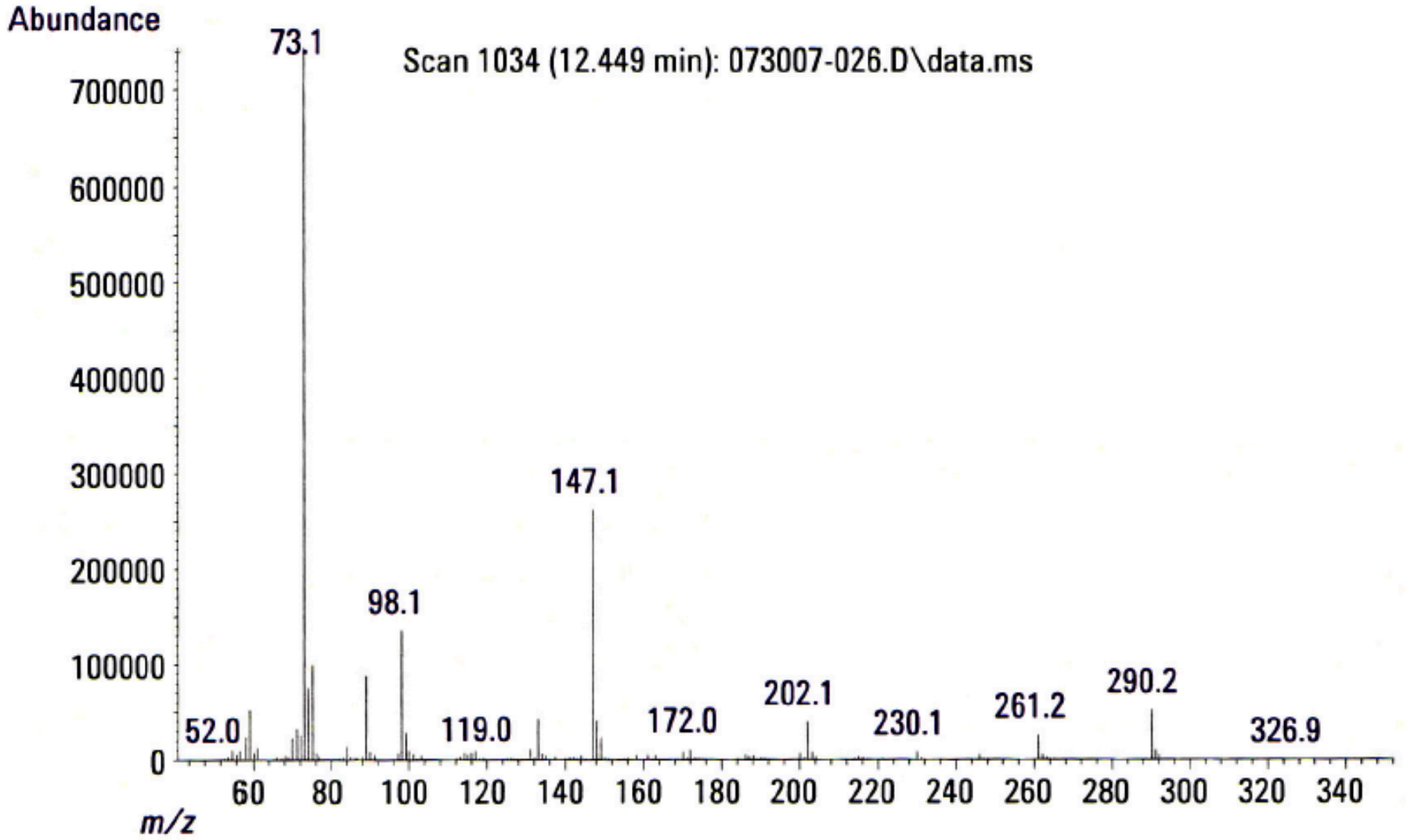
UNIVERSITÀ CATTOLICA del Sacro Cuore

ID by GC-MS

- GC-MS is often best for identification of amino acids, organic acids, sugars, fatty acids and molecules with MW<500
- GC has higher resolution and better reproducibility than LC
- EI-MS is more standardized than soft ionization methods, so EI spectra are more comparable.

UNIVERSITÀ CATTOLICA del Sacro Cuore

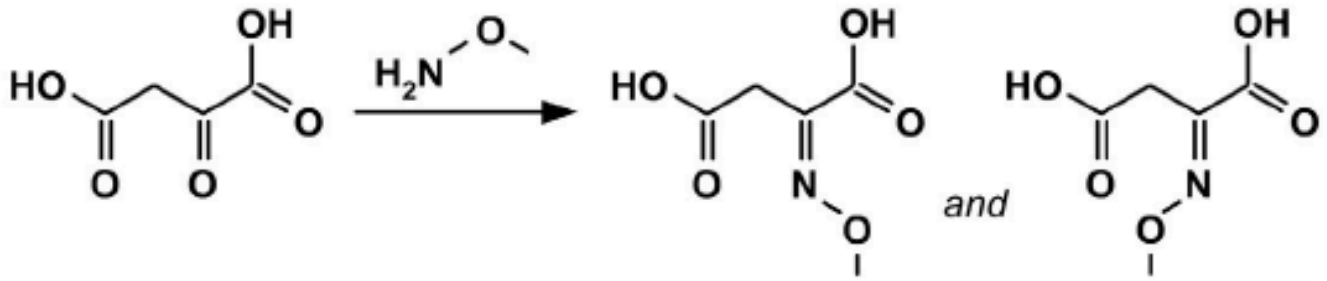
Electron impact MS spectra



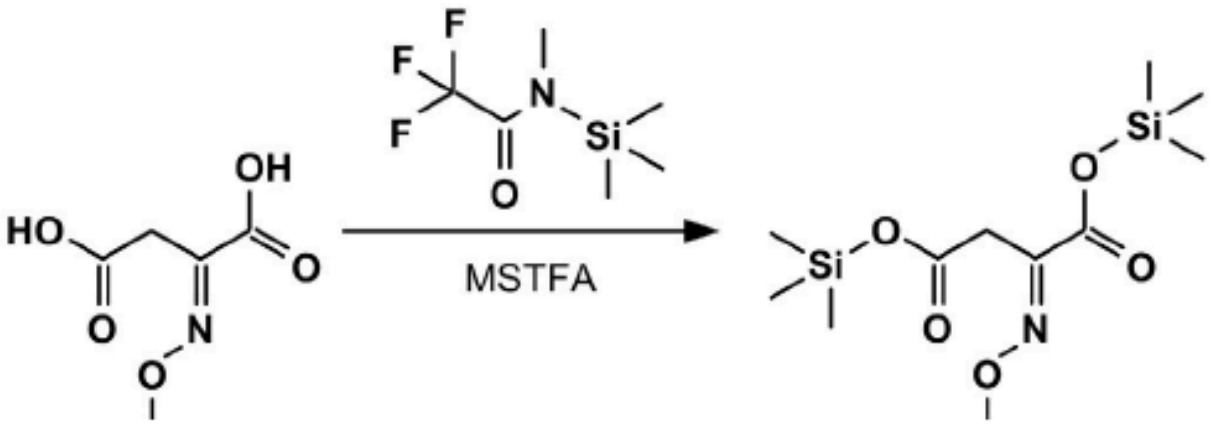
UNIVERSITÀ CATTOLICA del Sacro Cuore

Need for derivatization

Step I. Methoxyamination to protect carbonyl groups

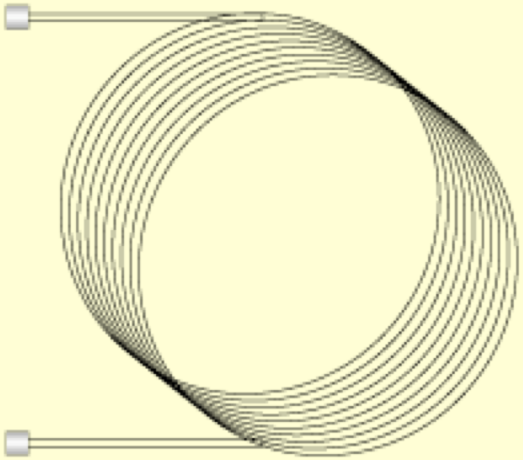


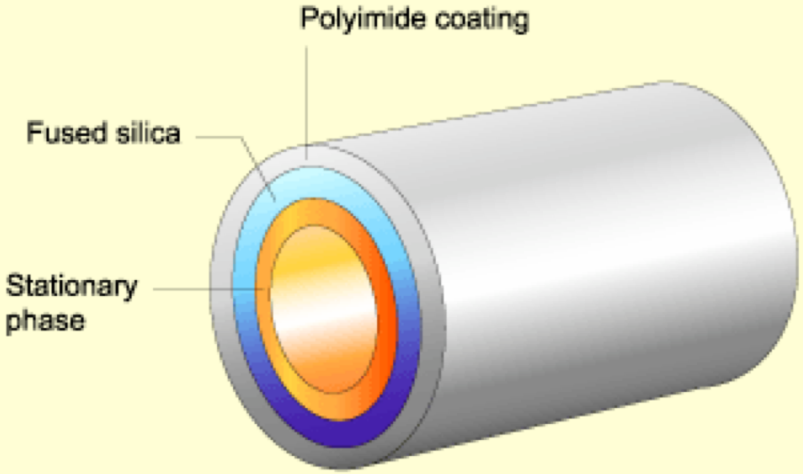
Step II. Trimethylsilylation to decrease the boiling point

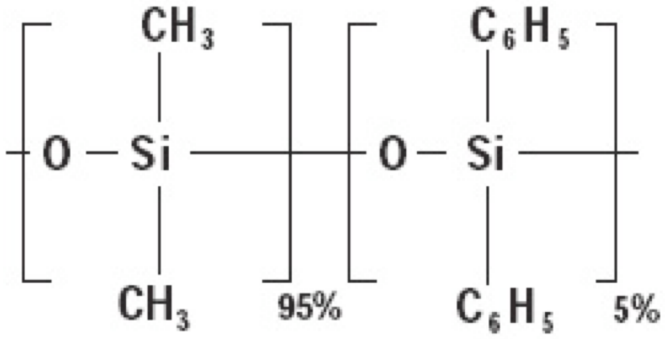


UNIVERSITÀ CATTOLICA del Sacro Cuore

GC-Columns







Polysiloxane

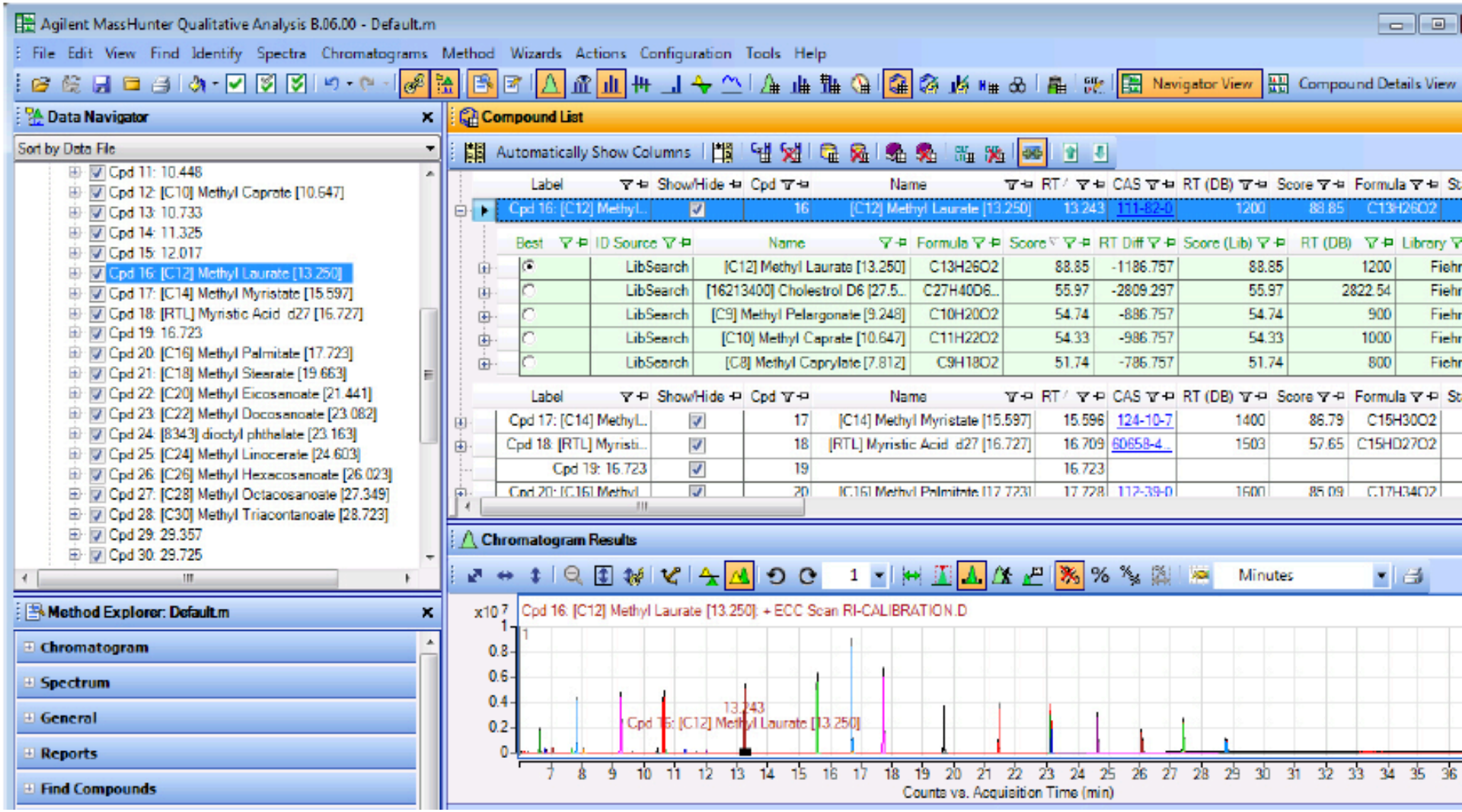
UNIVERSITÀ CATTOLICA del Sacro Cuore

Retention Time/Index

- Retention time (RT) is the time taken by an analyte to pass through a column
- RT is affected by compound, column (dimensions and stationary phase), flow rate, pressure, carrier, temp.
- Comparing RT from a standard sample to an unknown allows compound ID
- Retention index (RI) is the retention time normalized to the retention times of adjacently eluting n-alkanes

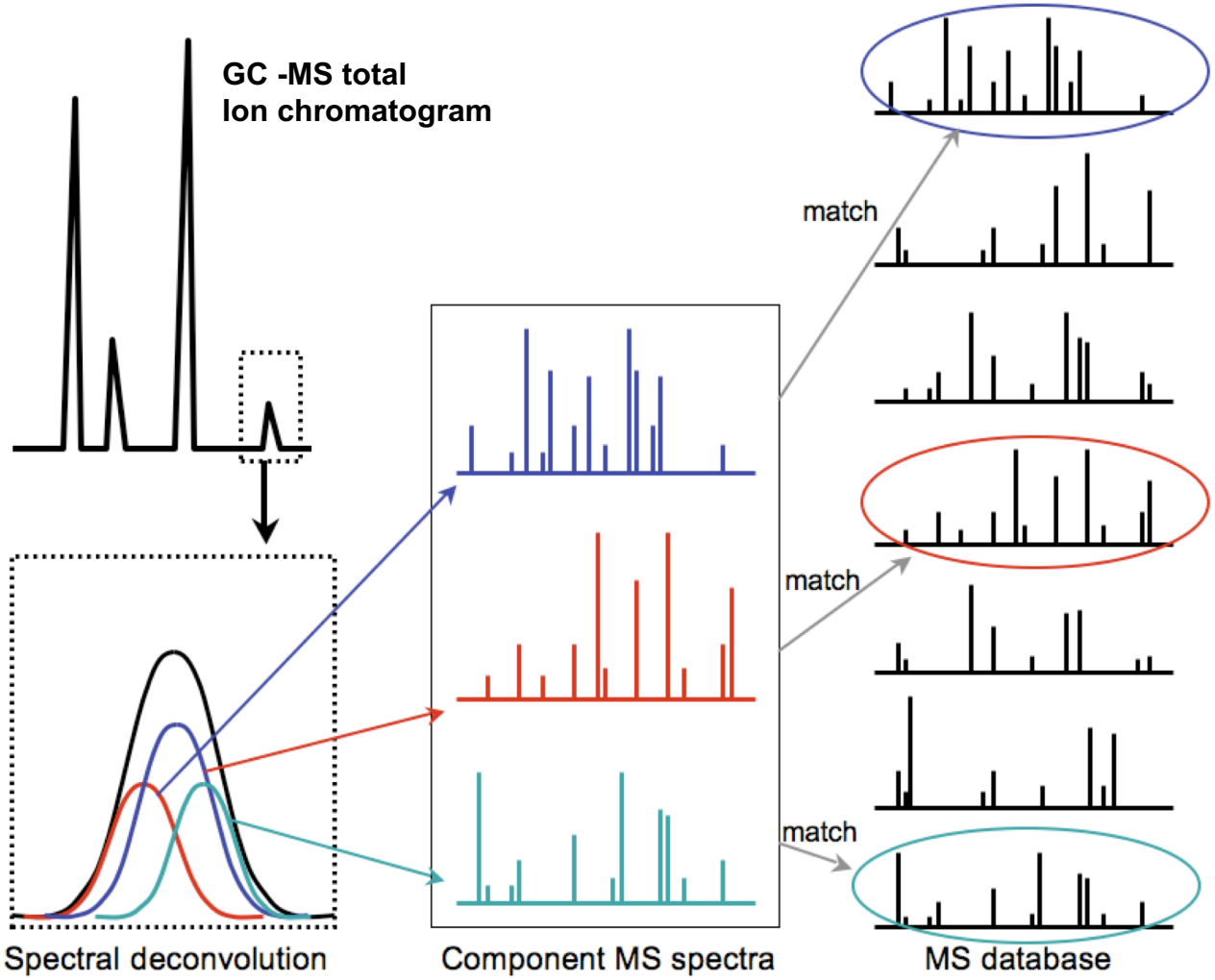
UNIVERSITÀ CATTOLICA del Sacro Cuore

RT locking



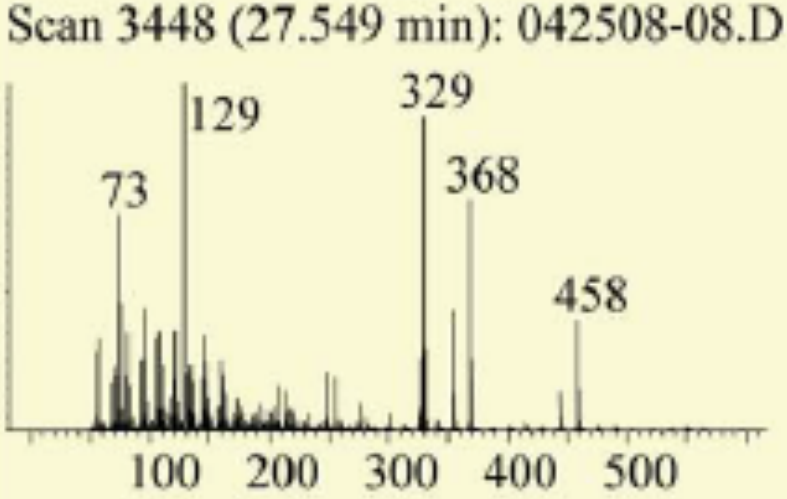
UNIVERSITÀ CATTOLICA del Sacro Cuore

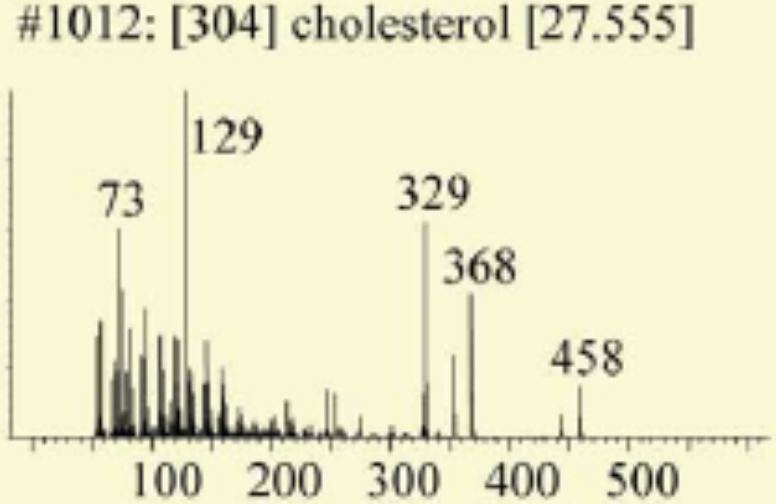
Metabolite ID by GC-MS



UNIVERSITÀ CATTOLICA del Sacro Cuore

Spectral comparison



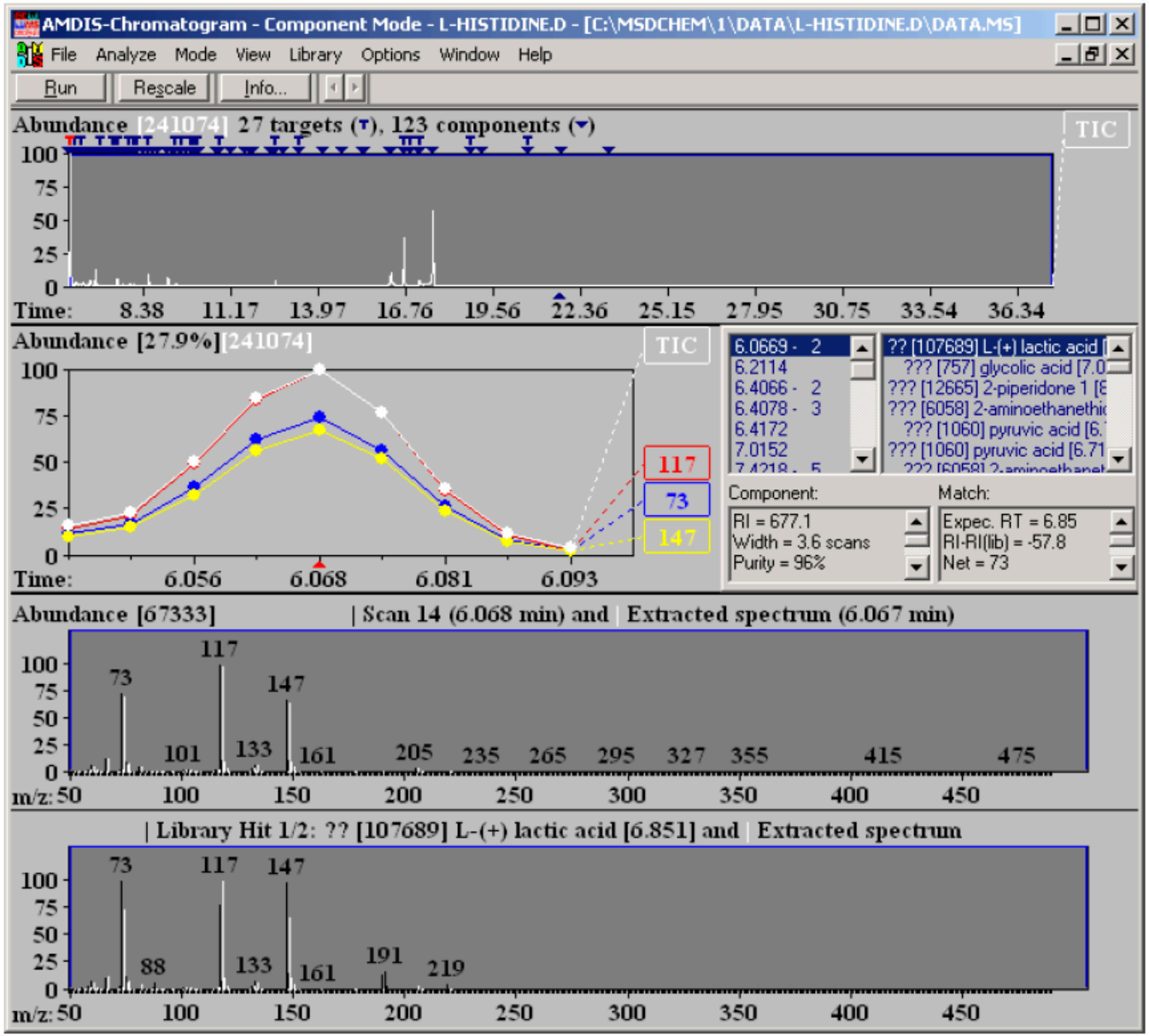


Scan 3448 (27.549 min): 042508-08.D\data.ms blood plasma488755

#	Name		MW	Formula	Qual	CAS#
1	[304] cholesterol	[27.555]	9999	C27H46O	91	000057-88-5
2	[637775] 3,5-dimethoxy-4-hydroxycinnamic acid 2	[20.65]	9999	C11H12O5	7	000530-59-6

UNIVERSITÀ CATTOLICA del Sacro Cuore

Library search



DATA REDUCTION and INTERPRETATION

UNIVERSITÀ CATTOLICA del Sacro Cuore

Data analysis

- **Data collected represented in a matrix**

Variables going across in different columns

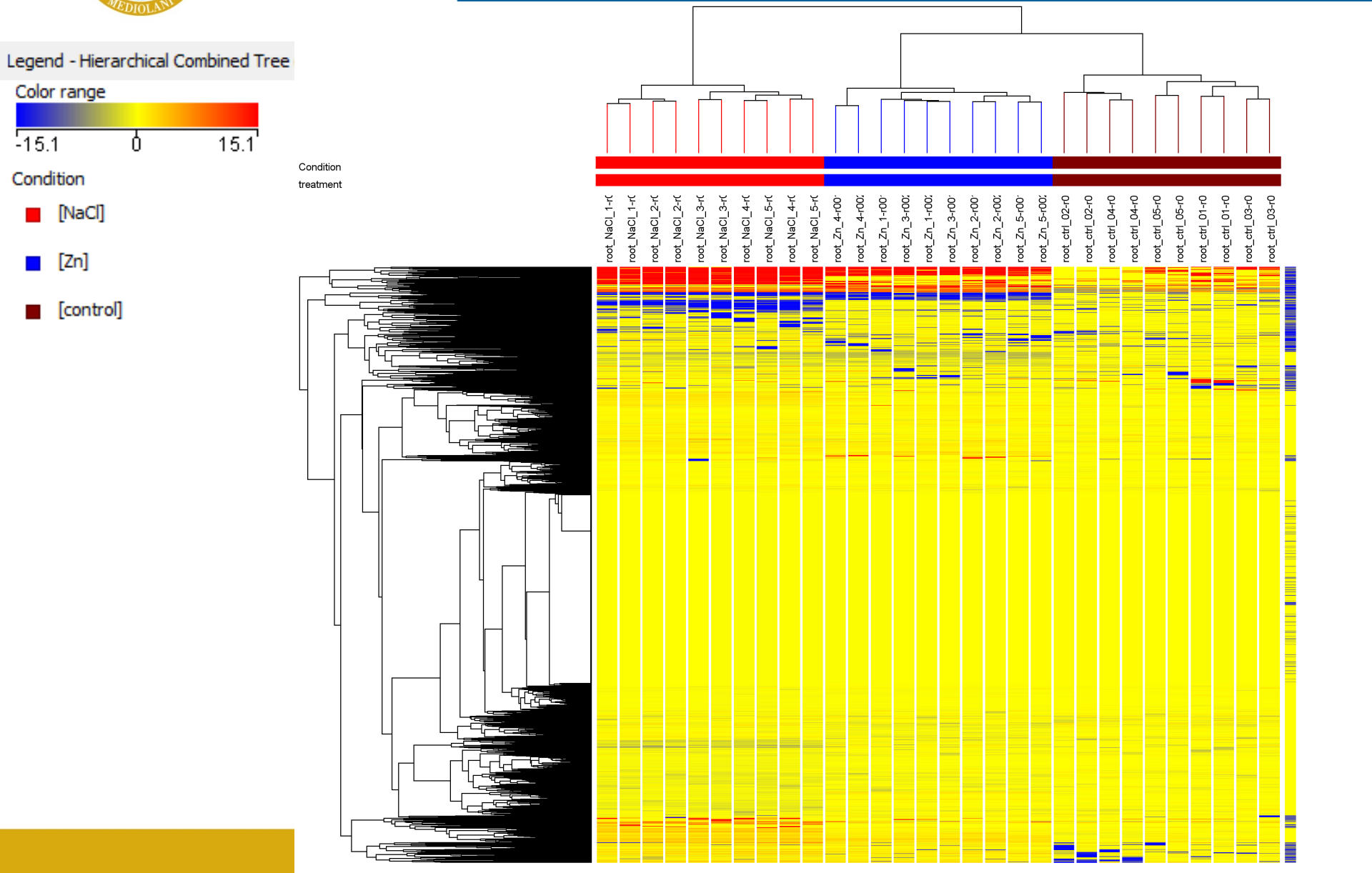
Objects going down in different rows	X-var 1	X-var 2	X-var 3	Y-var 1	Y-var 2
Sample 1					
Sample 2…					

A propositional approach to describing and using metabolomics data (the x-data) for analyzing complex systems. These may have other specific properties (the y-data) which one may also wish to 'explain' in terms of the x-data.

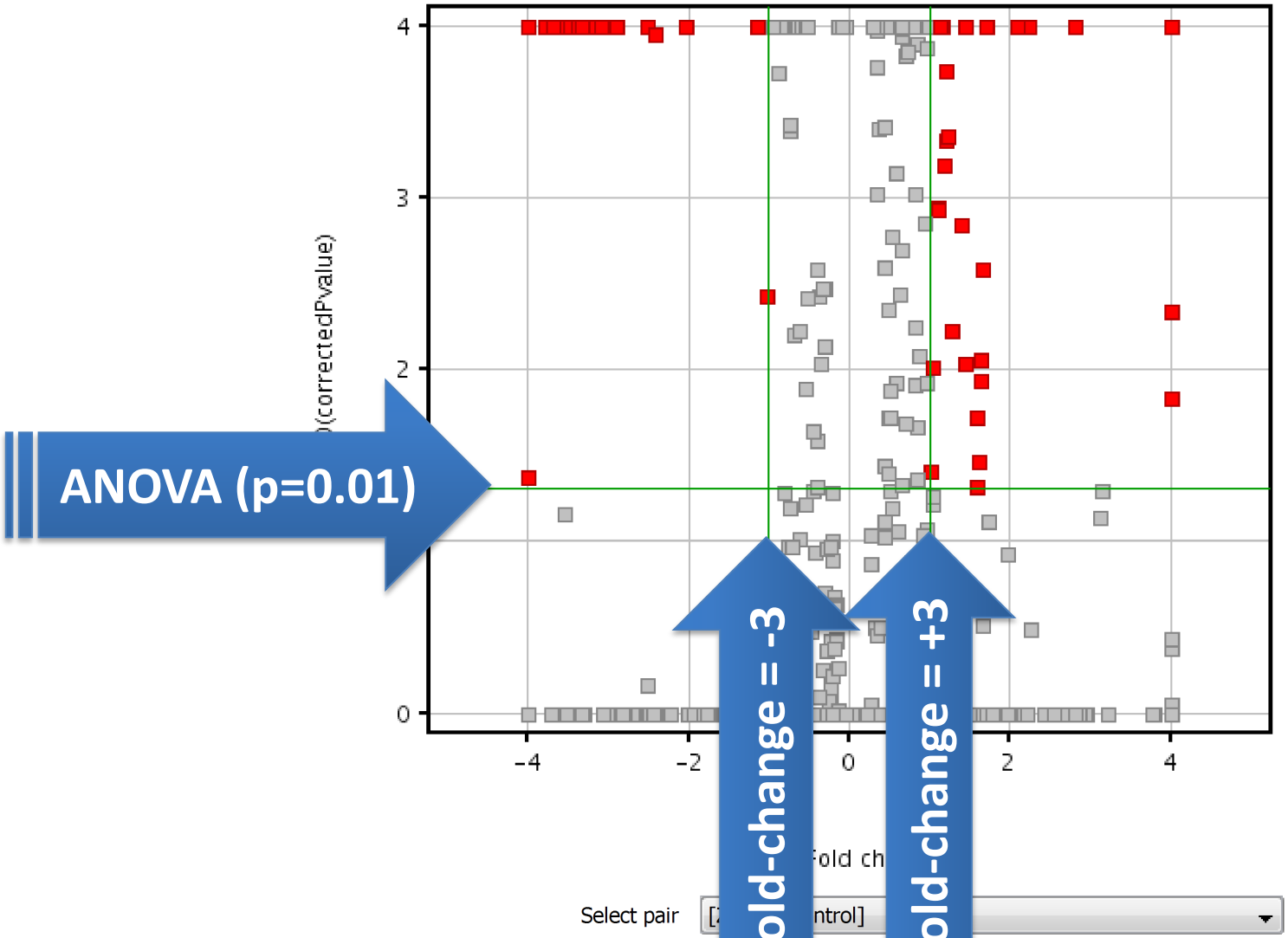
- **Chemometric Approach**
  - **Principle Component Analysis (PCA)**
  - **Partial Least-Squares (PLS) Method**
  - **Orthogonal PLS (OPLS)**
- **Targeted Profiling**

UNIVERSITÀ CATTOLICA del Sacro Cuore

Unsupervised cluster analysis

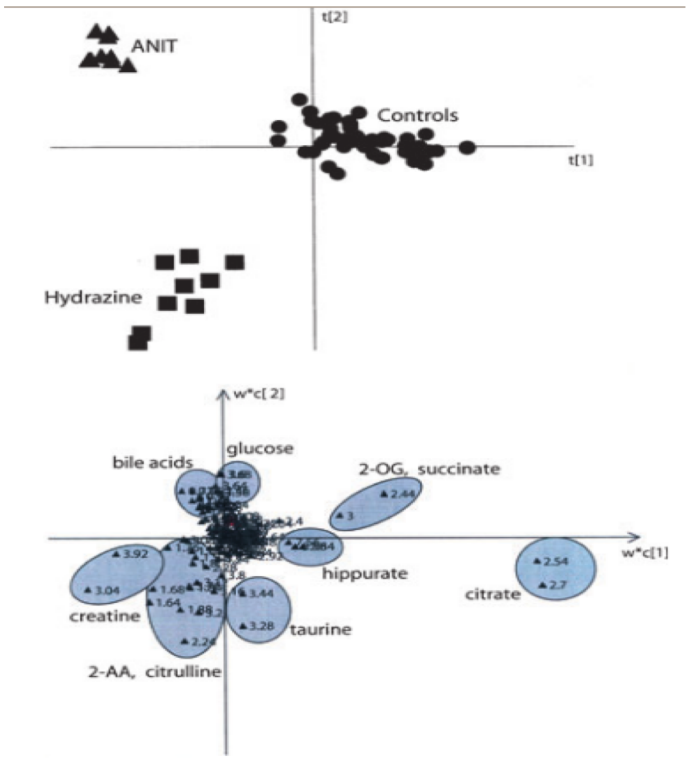


Volcano analysis



PCA

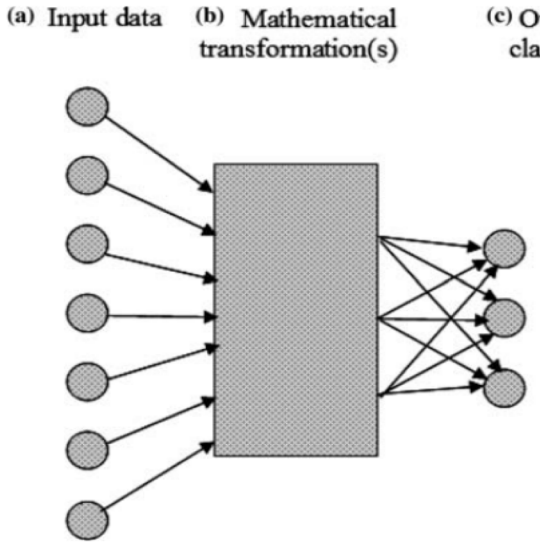
- Unsupervised
- Multivariate analysis based on projection methods
- Extract and display the systematic variation in the data
- Each Principle Component (PC) is a linear combination of the original data parameters
- PCs Orthogonal to each other
- Conversion of original data leads to two matrices, known as scores and loadings
- The scores(T) represent a low-dimensional plane that closely approximates X. Linear combinations of the original variables. Each point represents a single sample spectrum.
- A loading plot/scatter plot(P) shows the influence (weight) of the individual X-variables in the model. Each point represents a different spectral intensity.
- The part of X that is not explained by the model forms the residuals(E)
- $X = TP^T = t_1p_1^T + t_2p_2^T + ... + E$



UNIVERSITÀ CATTOLICA del Sacro Cuore

PLS

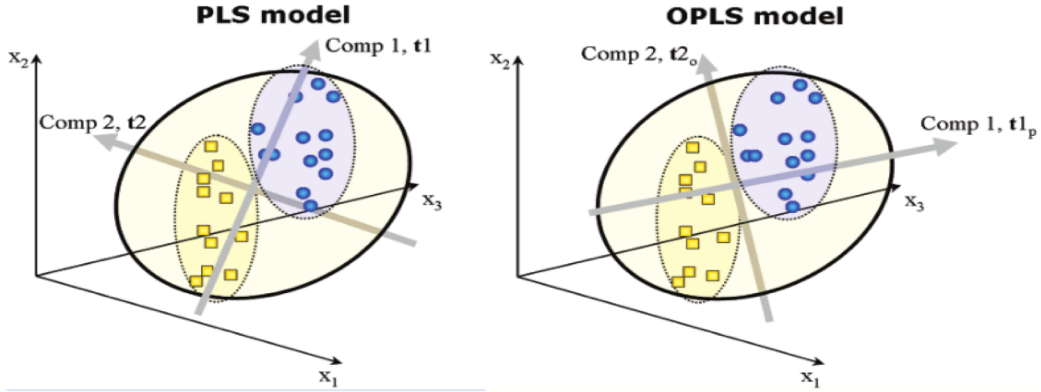
- Supervised learning method.
- Principles that of PCA. But in PLS, a second piece of information is used, namely, the labeled set of class identities.
- Two data tables considered namely X (input data from samples) and Y (containing qualitative values, such as class belonging, treatment of samples)
- The PLS algorithm maximizes the covariance between the X variables and the Y variables



The class assignment problem. The inputs can be considered, and are referred to, as the "explanatory variables" or "x-data" whereas the functional or the other classes of interest, which are still variables associated with the samples, are referred to as "dependent variables" or "y-data" and are to be obtained as the outputs.

UNIVERSITÀ CATTOLICA del Sacro Cuore

OPLS



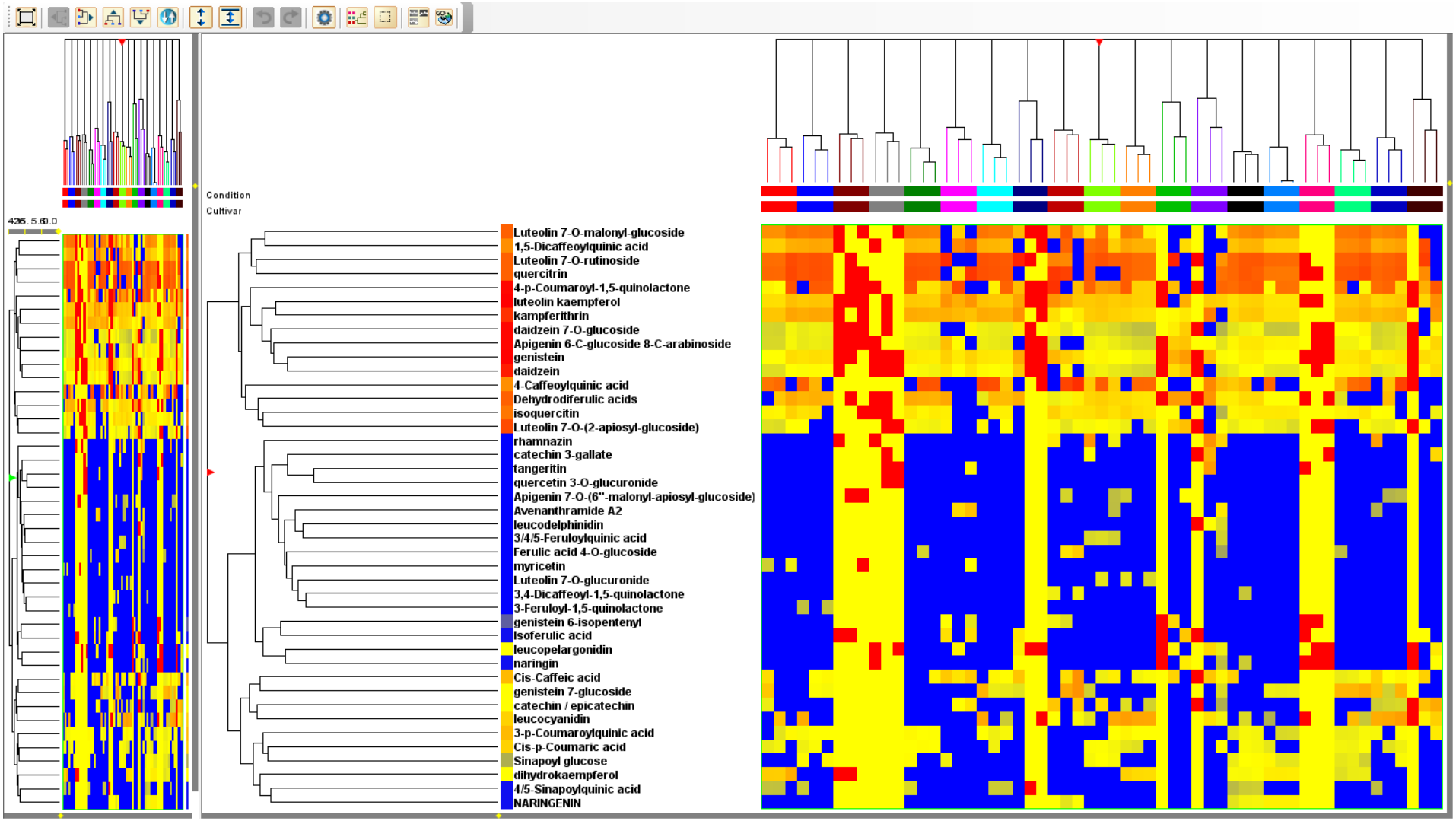
A geometrical illustration of the difference between the PLS-DA and OPLS-DA models. In the left panel, the PLS components cannot separate the between-class variation from the within-class variation, and the resulting PLS component loadings mixes both types of variations. In the right panel, the OPLS components are able to separate these two different variations. Component 1 ($t1_p$) is the predictive component and displays the between-class ([blue circles], [yellow squares]) variation of the samples. The corresponding loading profile can be used for identifying variables important for the class separation. Component 2 ($t2_o$) is the **Y**-orthogonal component and models the within group (within-class) variation.

- **OPLS method is a recent modification of the PLS method to help overcome pitfalls**
- **Main idea to seperate systematic variation in X into two parts, one linearly related to Y and one unrelated (orthogonal).**
- **Comprises two modeled variations, the Y-predictive ($T_pP_p^T$) and the Y-orthogonal ($T_oP_o^T$) compononents.**
- **Only Y-predictive variation used for modeling of Y.**
- **OPLS-DA compared to PLS-DA**

APPLICATIONS

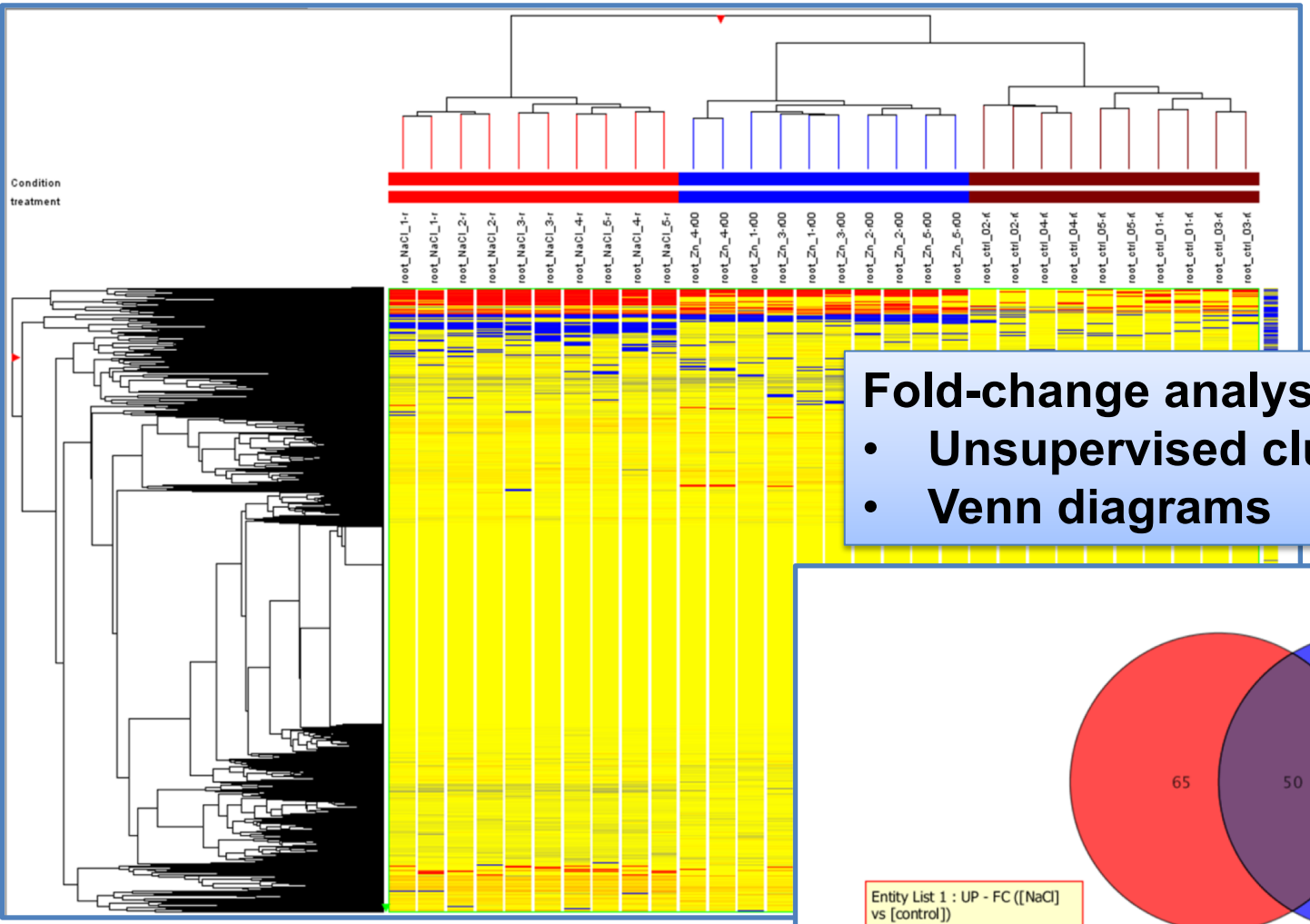
UNIVERSITÀ CATTOLICA del Sacro Cuore

Cluster analysis varietà carciofo (phenolic profiling)



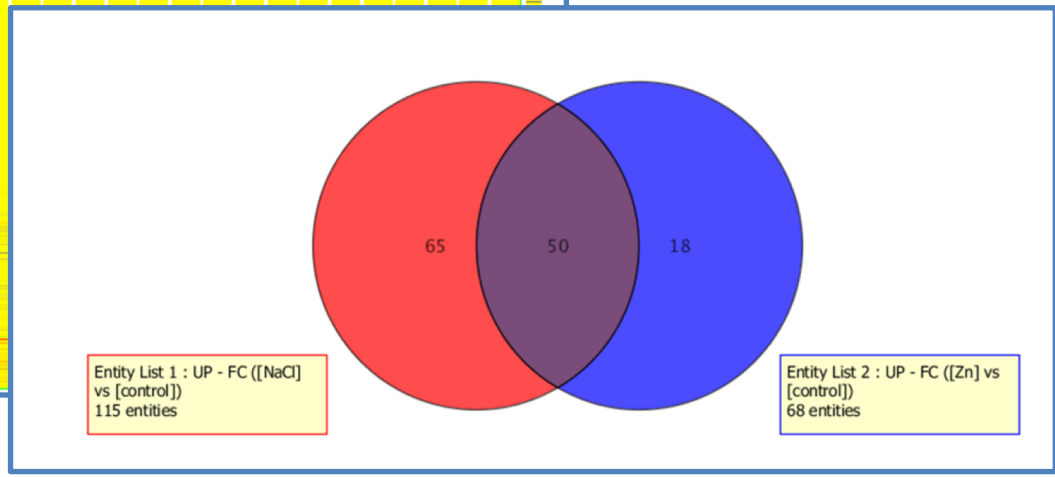
UNIVERSITÀ CATTOLICA del Sacro Cuore

Effect of abiotic stress on plant metabolism (Zn vs salinity vs control)



Fold-change analysis let to perform:

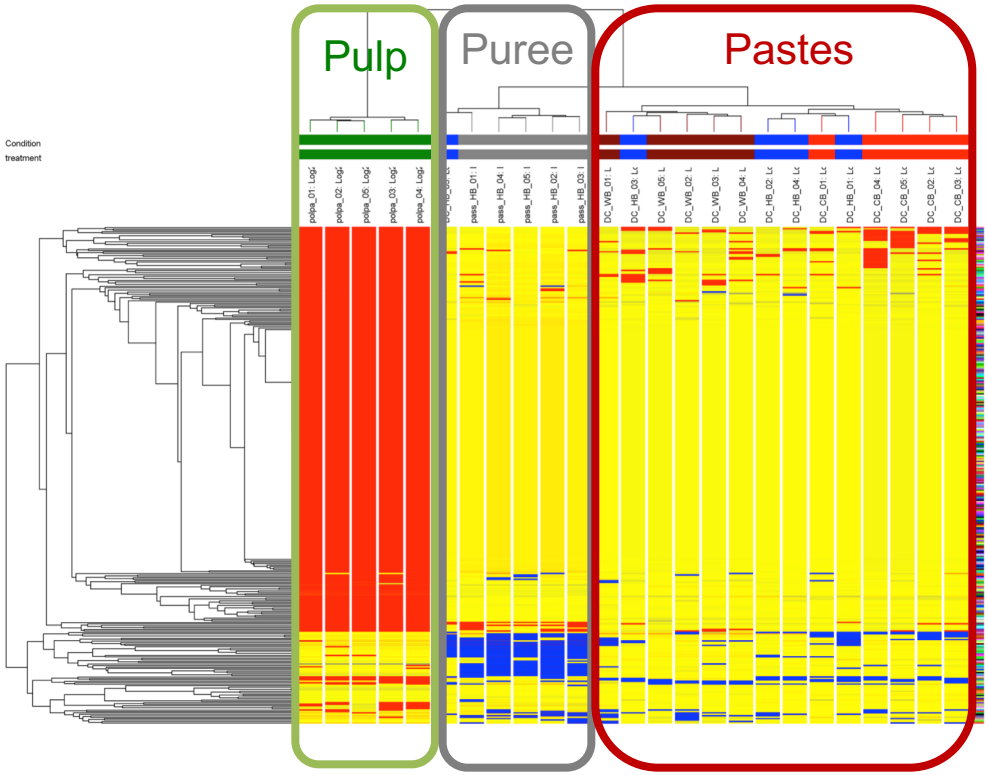
- **Unsupervised cluster analysis**
- **Venn diagrams**



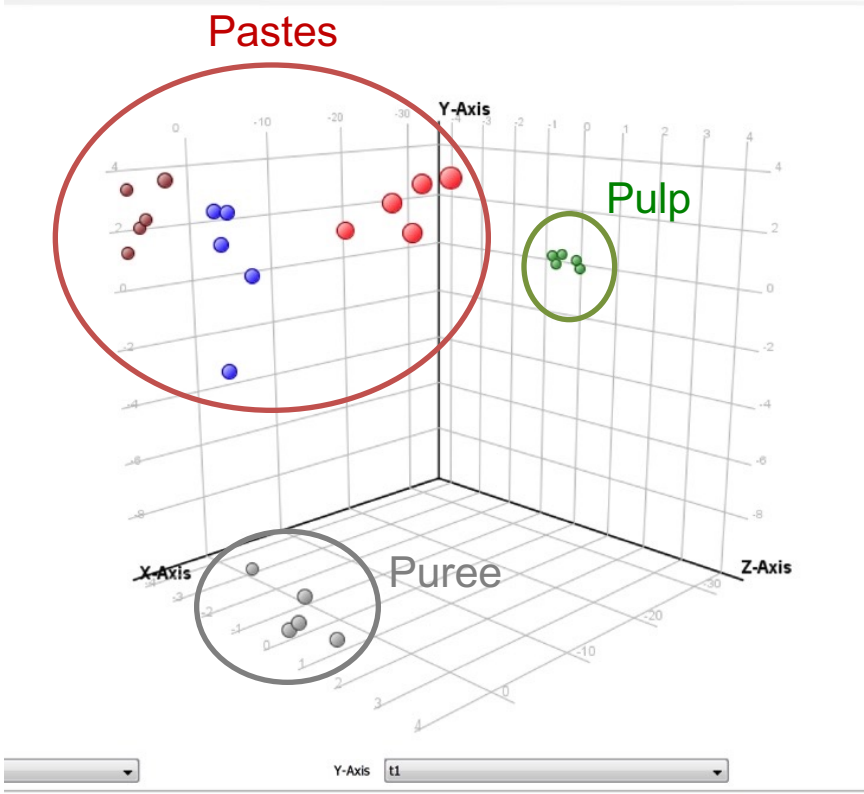
UNIVERSITÀ CATTOLICA del Sacro Cuore

Discrimination of different processing technologies

Unsupervised cluster analysis



Partial Least Square Discriminant Analysis (PLS-DA)

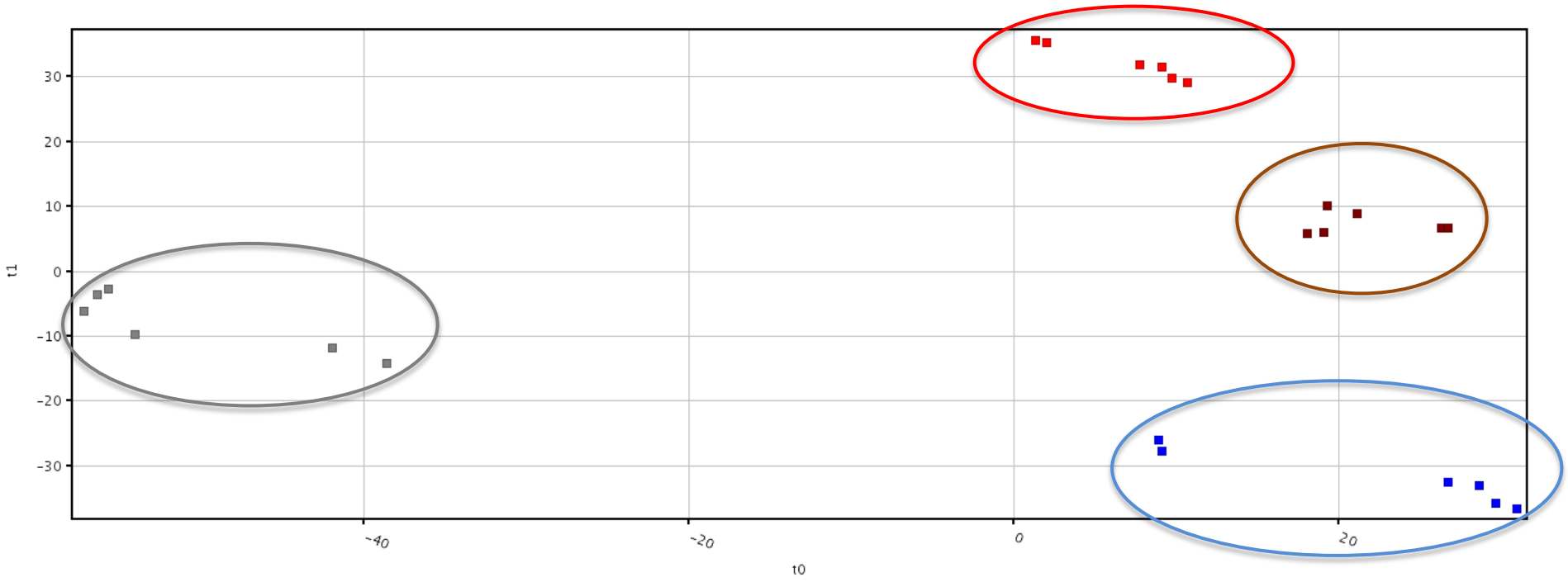


Model accuracy: 100%

Lucini, Rocchetti, Kane, & Trevisan. *Food Control*, 2017

UNIVERSITÀ CATTOLICA del Sacro Cuore

Rintracciabilità dei processi fermentativi (fave cacao)



Legend - Scores

Color by treatment (Non-averaged)

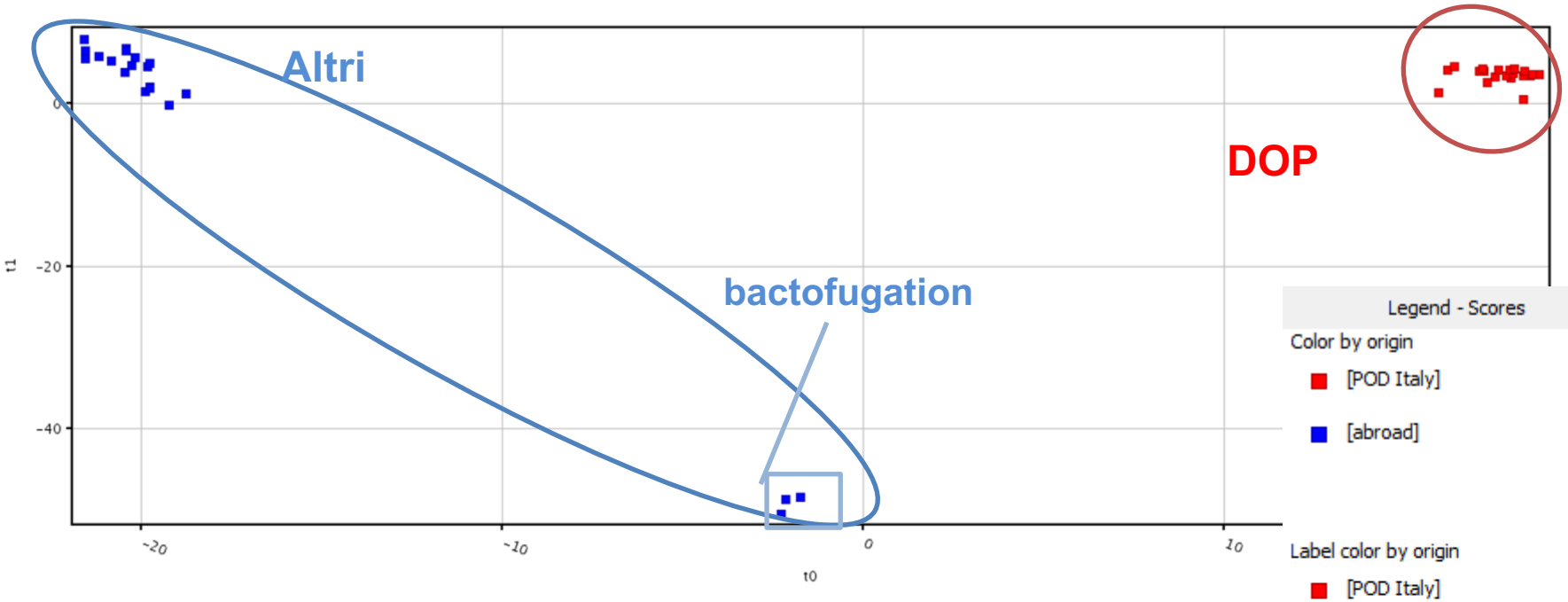
- [T1]
- [T2]
- [T3]
- [control]

X-Axis t0

Y-Axis t1

UNIVERSITÀ CATTOLICA del Sacro Cuore

Tutela produzioni – Grana DOP



Legend - Scores

Color by origin

[POD Italy]

[abroad]

Label color by origin

[POD Italy]

[abroad]

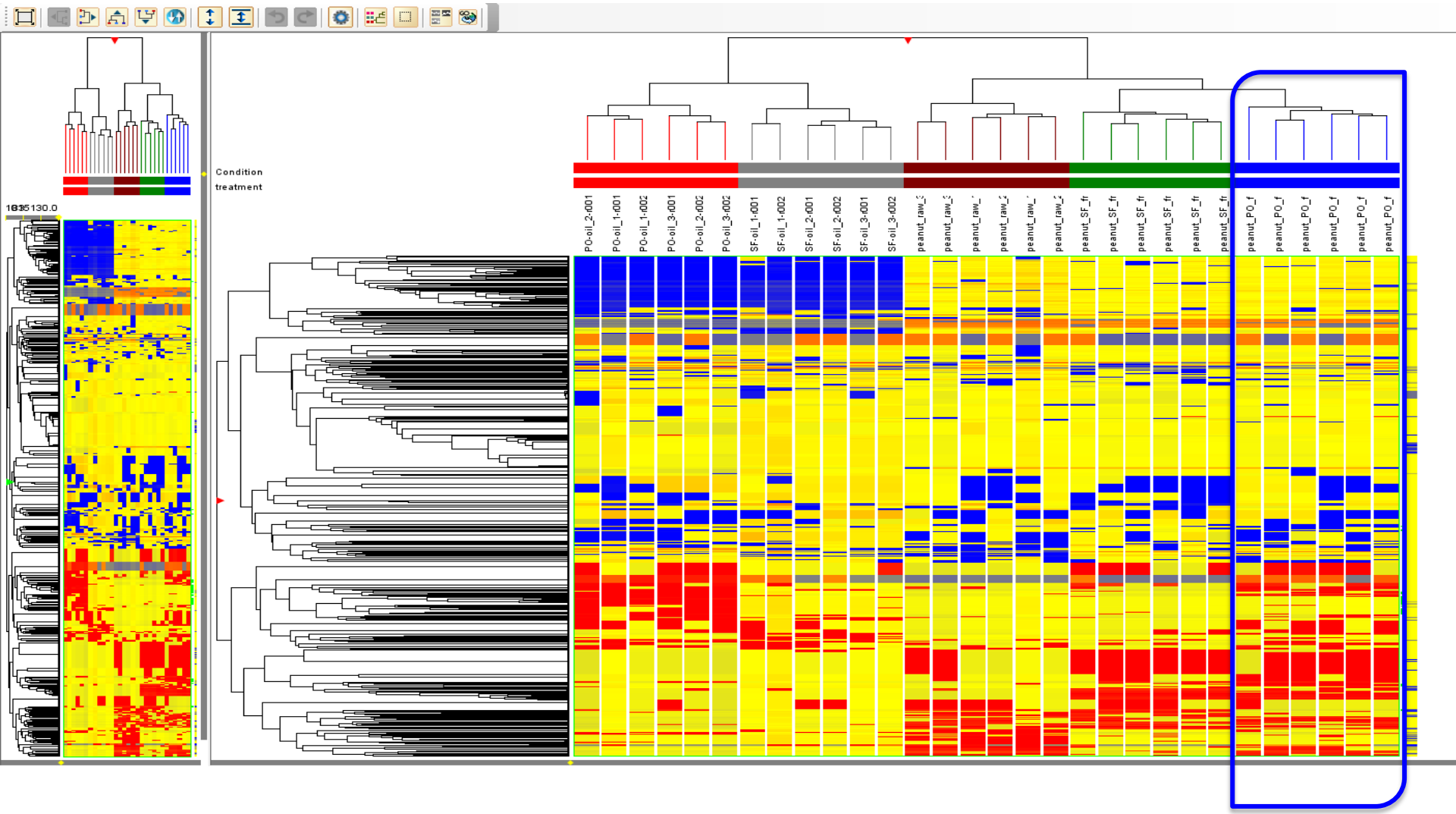
X-Axis t0

Y-Axis t1

	[abroad] (Predicted)	
20	0	
0	20	

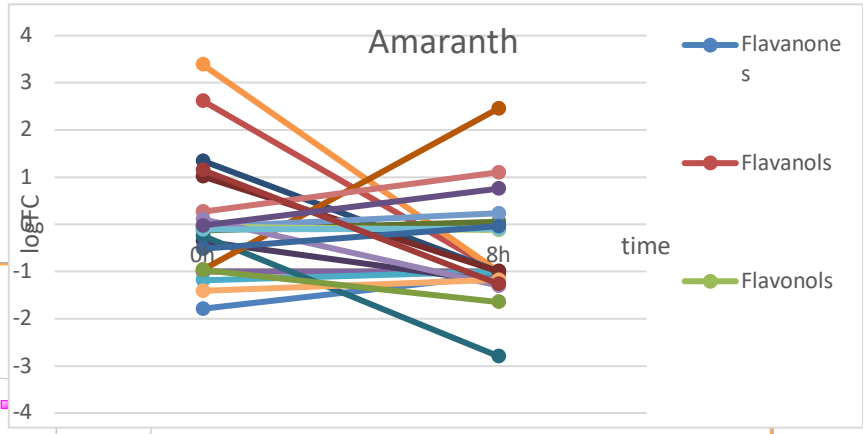
UNIVERSITÀ CATTOLICA del Sacro Cuore

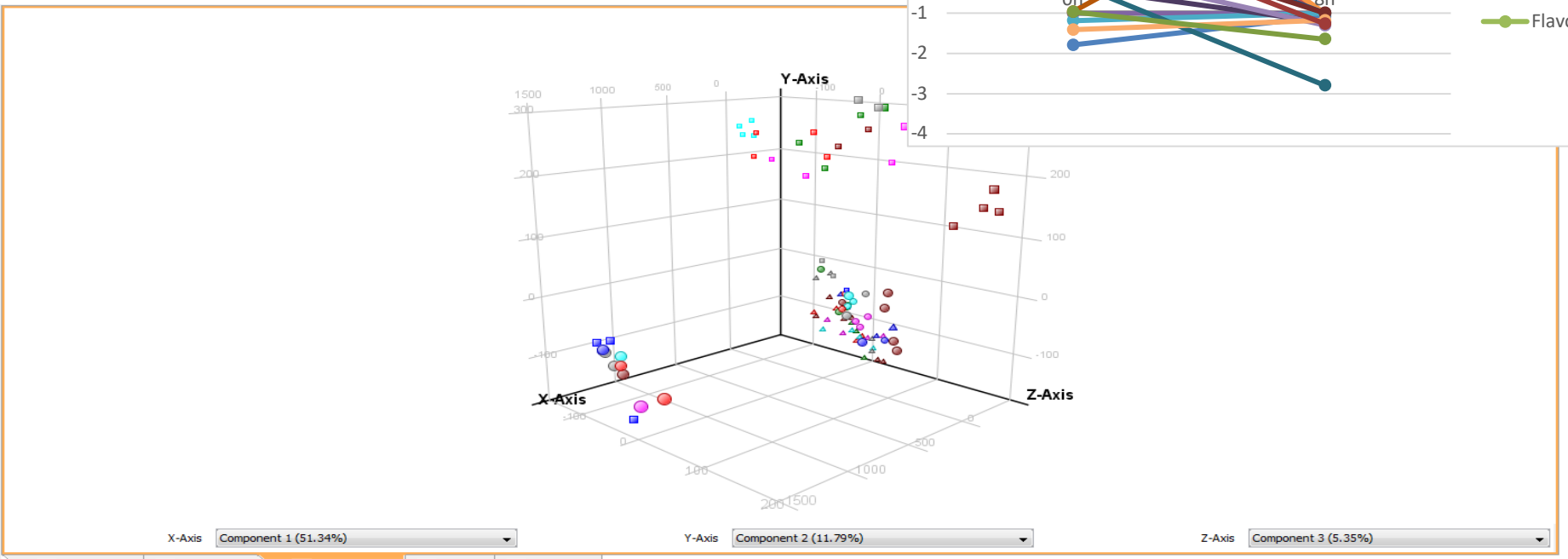
Utilizzo olio di palma

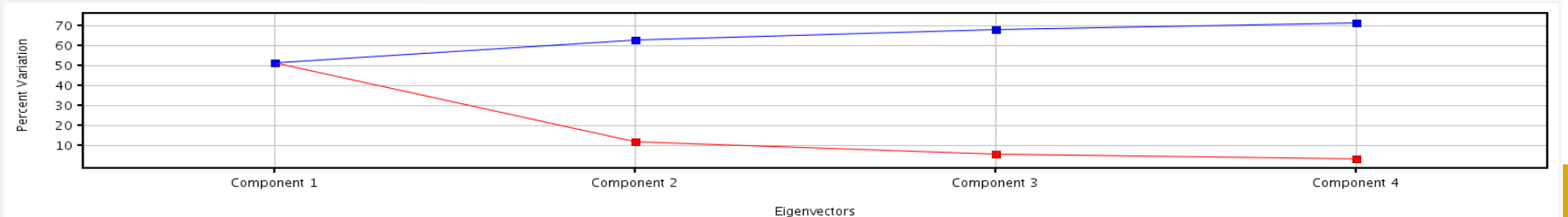


UNIVERSITÀ CATTOLICA del Sacro Cuore

Fermentazione intestinale farine gluten free



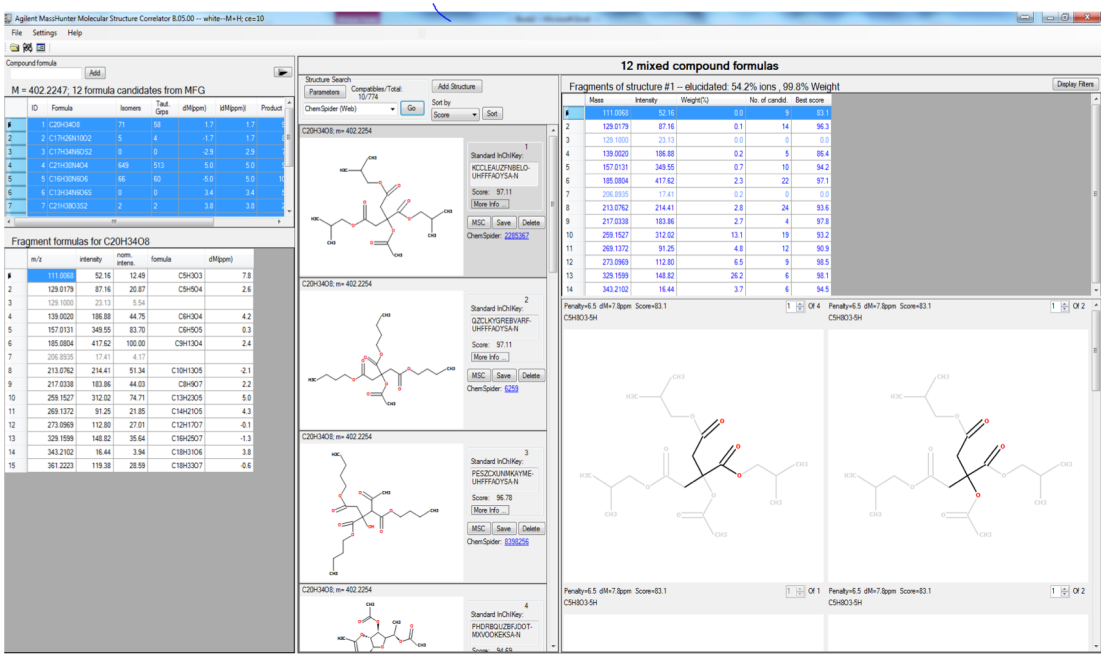




UNIVERSITÀ CATTOLICA del Sacro Cuore

Molecular Structure Correlator

- **"Systematic bond-breaking" approach** *[Hill and Mortishire-Smith, 2005]* **to correlate tandem MS data with chemical structures**

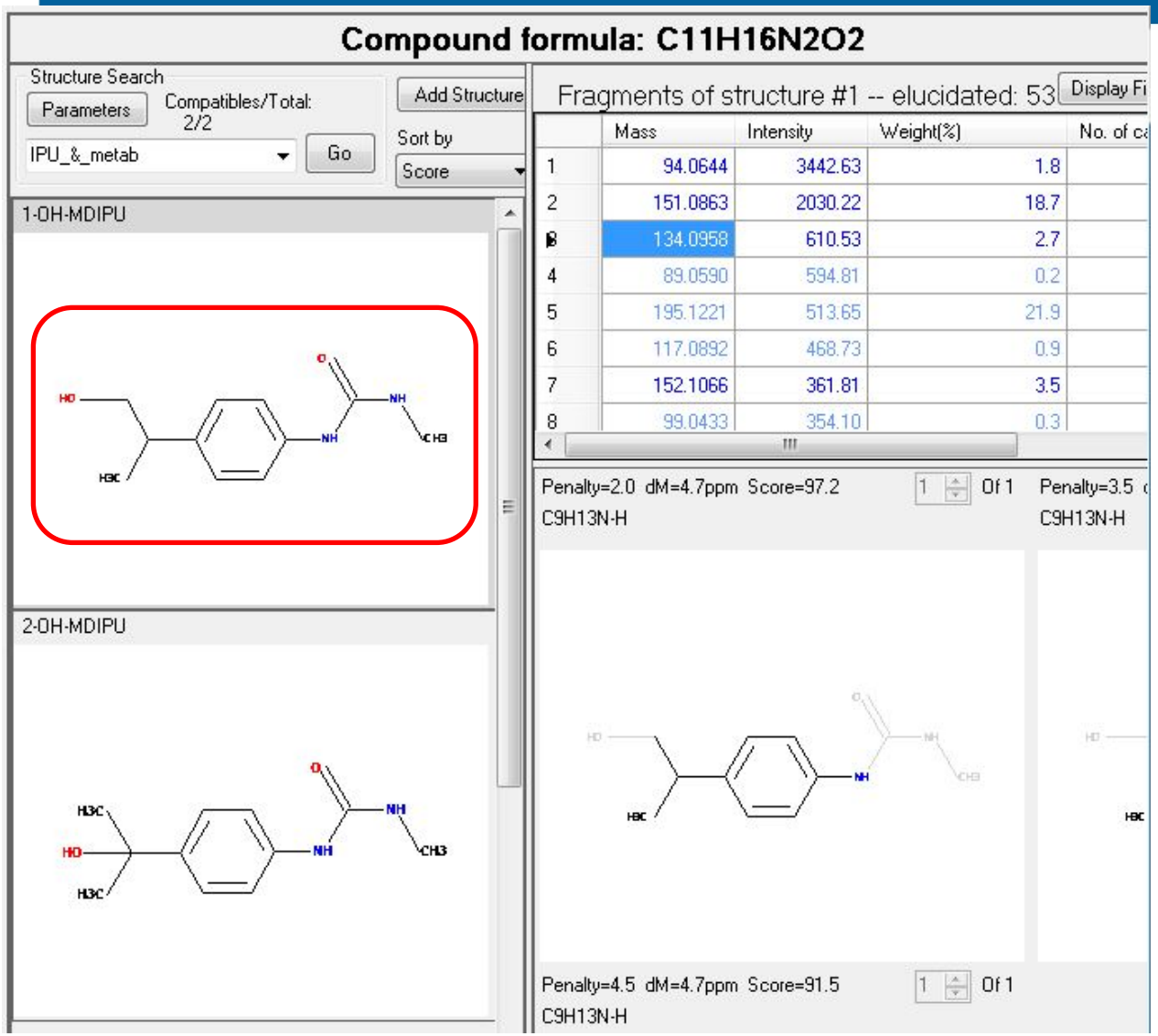


- **Scores are generated from each product ion, the mass accuracy of the fragments, and the overall percentage of ions intensity being plausibly explained with sub-structures**

UNIVERSITÀ CATTOLICA del Sacro Cuore

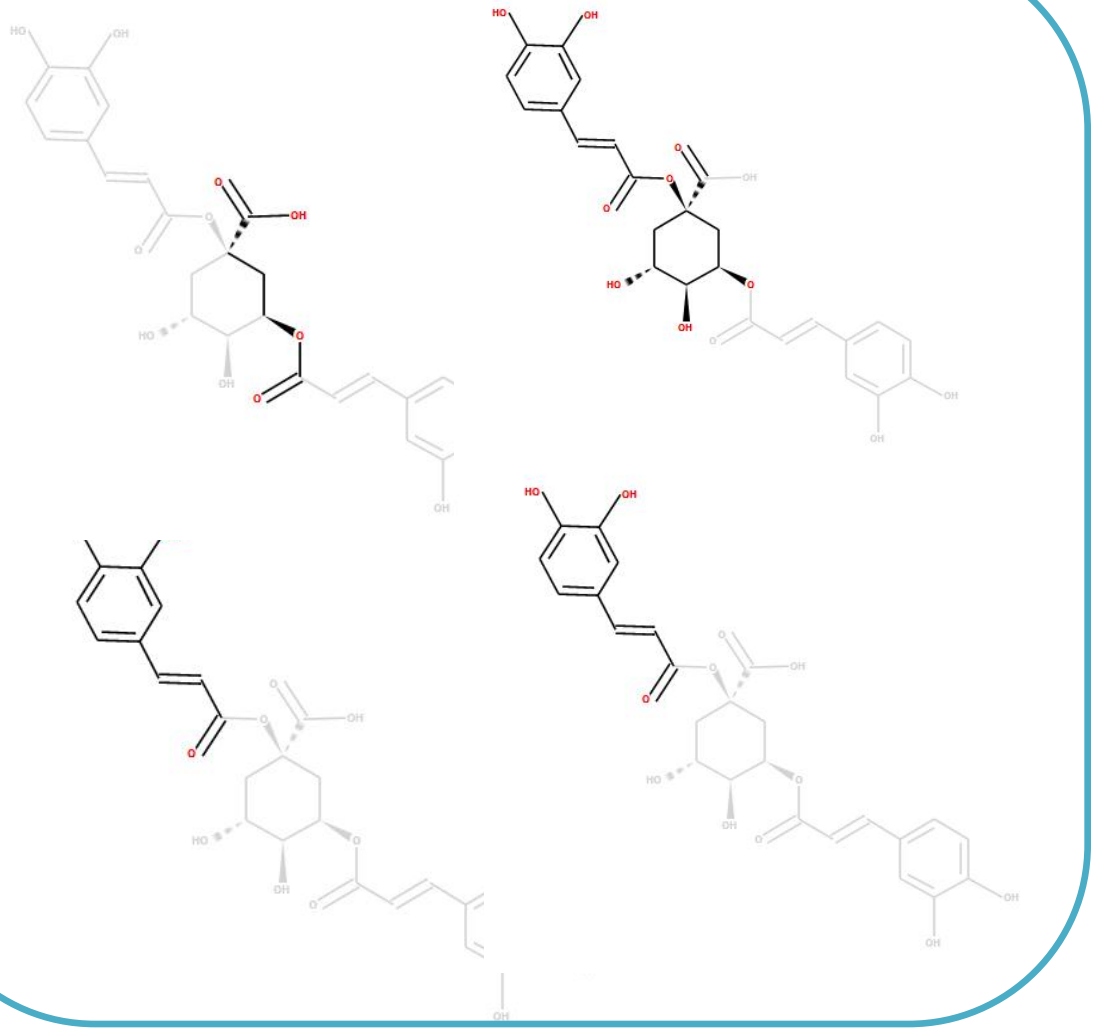
Structure elucidation & metabolite ID

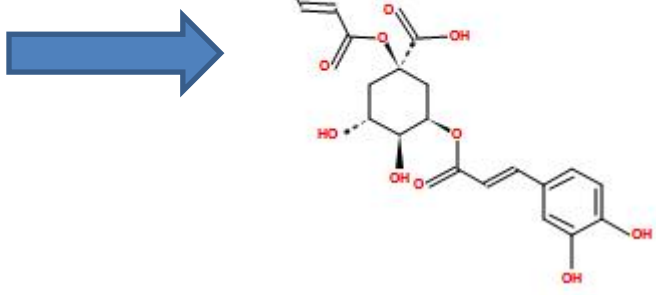
Structural elucidation to discriminate isobaric compounds



UNIVERSITÀ CATTOLICA del Sacro Cuore

Structural confirmation

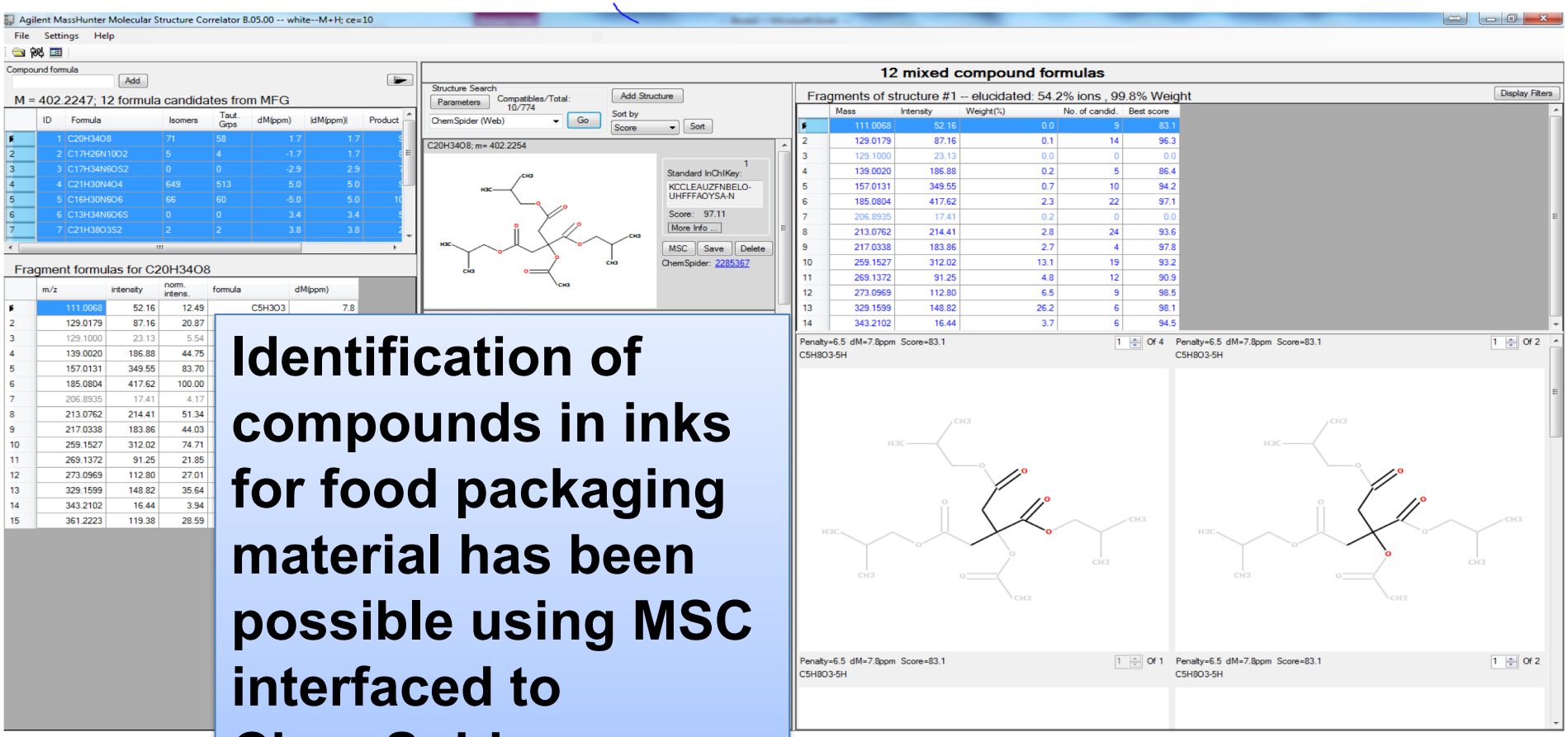




Confirm structure of selected differential metabolites

UNIVERSITÀ CATTOLICA del Sacro Cuore

From totally unknown to structures



Identification of compounds in inks for food packaging material has been possible using MSC interfaced to ChemSpider