Exploring the World of Omics: A Multi-Scale Approach to Biology

Basic Metabolomics : Analytical principles and profiling strategies

- Tutor: Dr.ssa Federica De Castro

IR0000032 – ITINERIS, Italian Integrated Environmental Research Infrastructures System
(D.D. n. 130/2022 - CUP B53C22002150006) Funded by EU - Next Generation EU PNRR-Mission 4 “Education and Research” - Component 2: “From research to business” - Investment 3.1: “Fund for the realisation of an integrated system of research and innovation infrastructures”

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani

What is metabolomics?

What are the possible applications?

How does it fit into the context of systems biology?

What is the typical experimental sequence of a metabolomics approach?

Metabolomics:

An approach at the interface between chemistry and biology

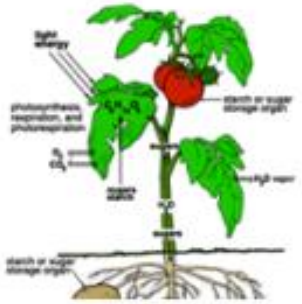
- Each biological phenomenon will produce a measurable chemical response
- Determining the observed changes will provide information on the state of the biological system under analysis.
- Measuring the final products of gene expression, i.e. metabolites, it is possible to obtain essential information on the biological system under examination

Wide range of applications

Natural product chemistry

Stress response

Physiology



Interactions among organisms

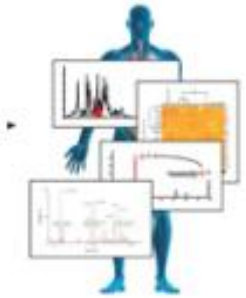
Taxonomy

Food science

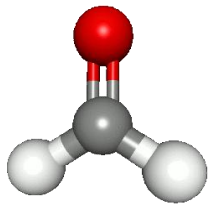
Pharmaceutical research

Microbial biology and chemistry

Human metabolism/ Biomarkers



Environmental science

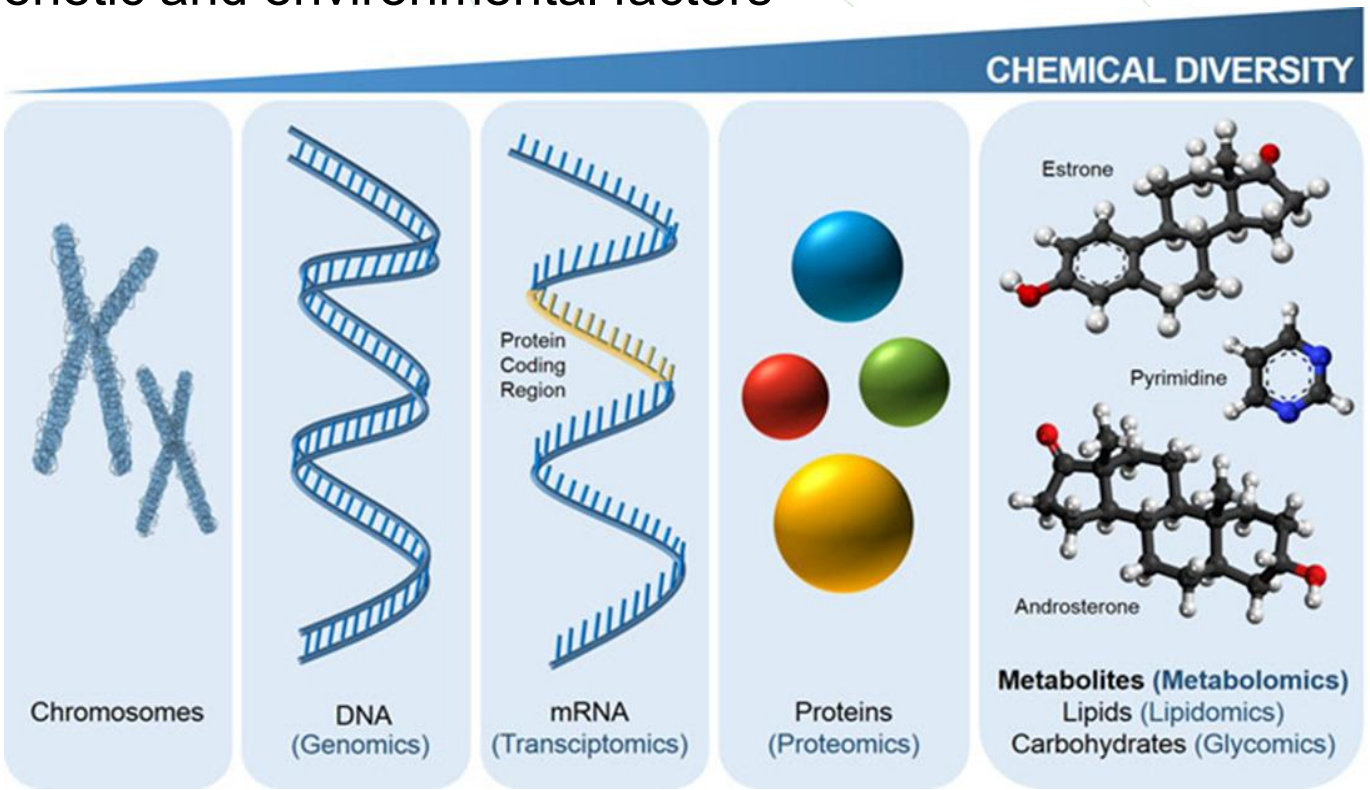


Metabolomics

Metabolomics is the study of chemical processes involving metabolites

Metabolism is the set of chemical reactions that sustain life within cells, biological fluids, tissues, or organisms, which are influenced by both genetic and environmental factors

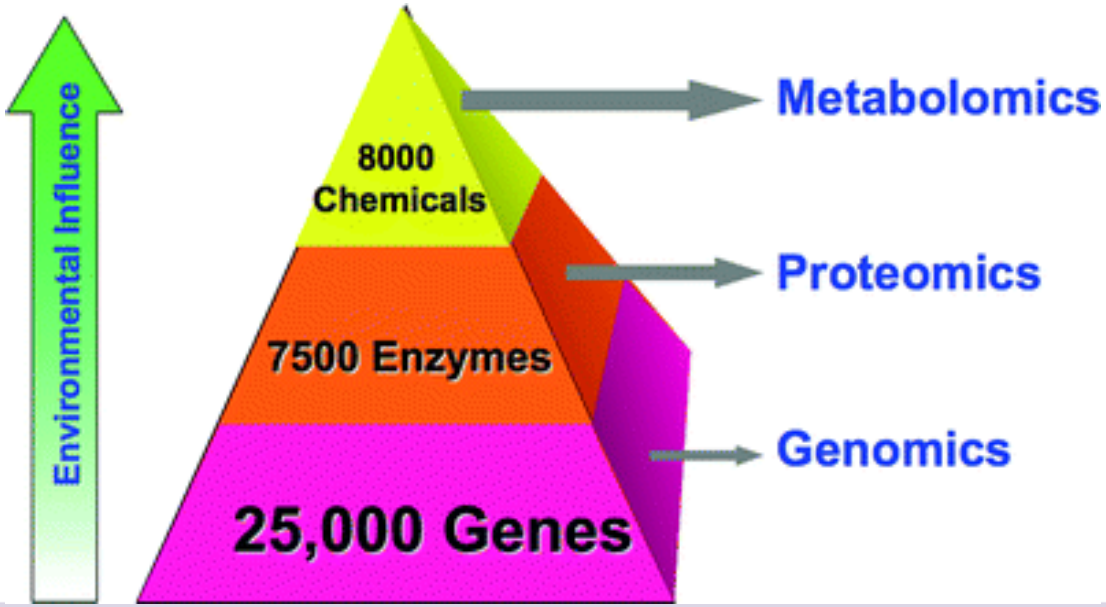
The **metabolome** is the set of small chemical molecules present in biological samples (which are the end products of cellular processes).

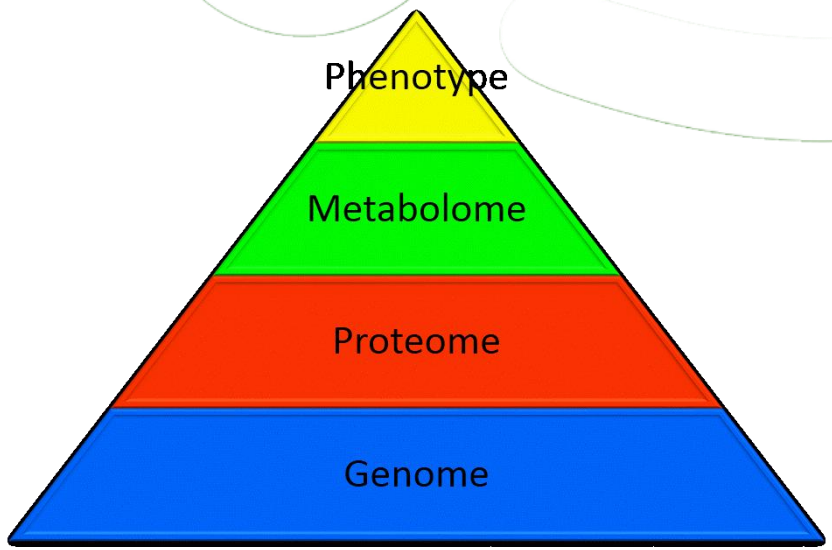


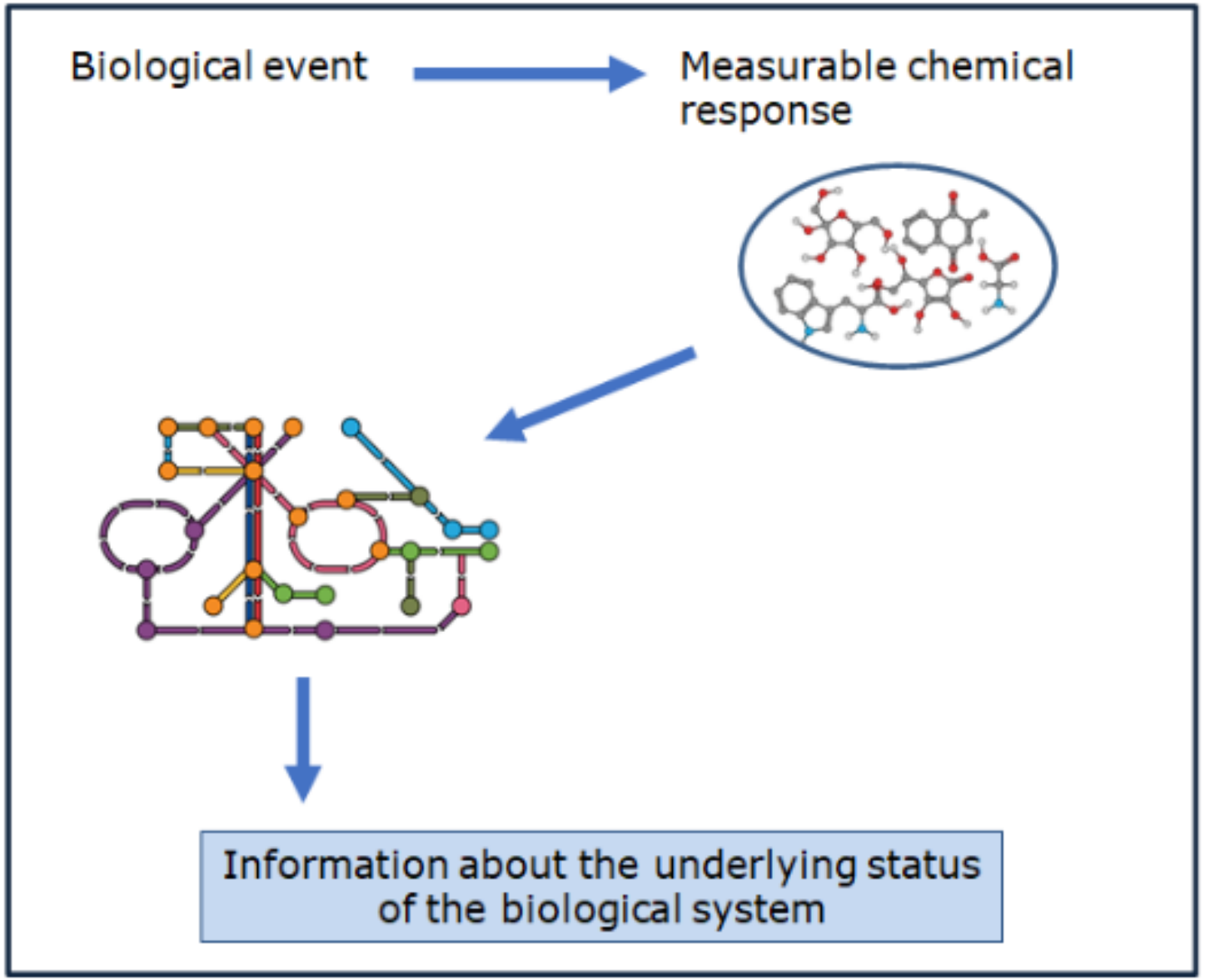
Metabolomics and Bioinformatics

Metabolomics is an extremely important area of bioinformatics research.

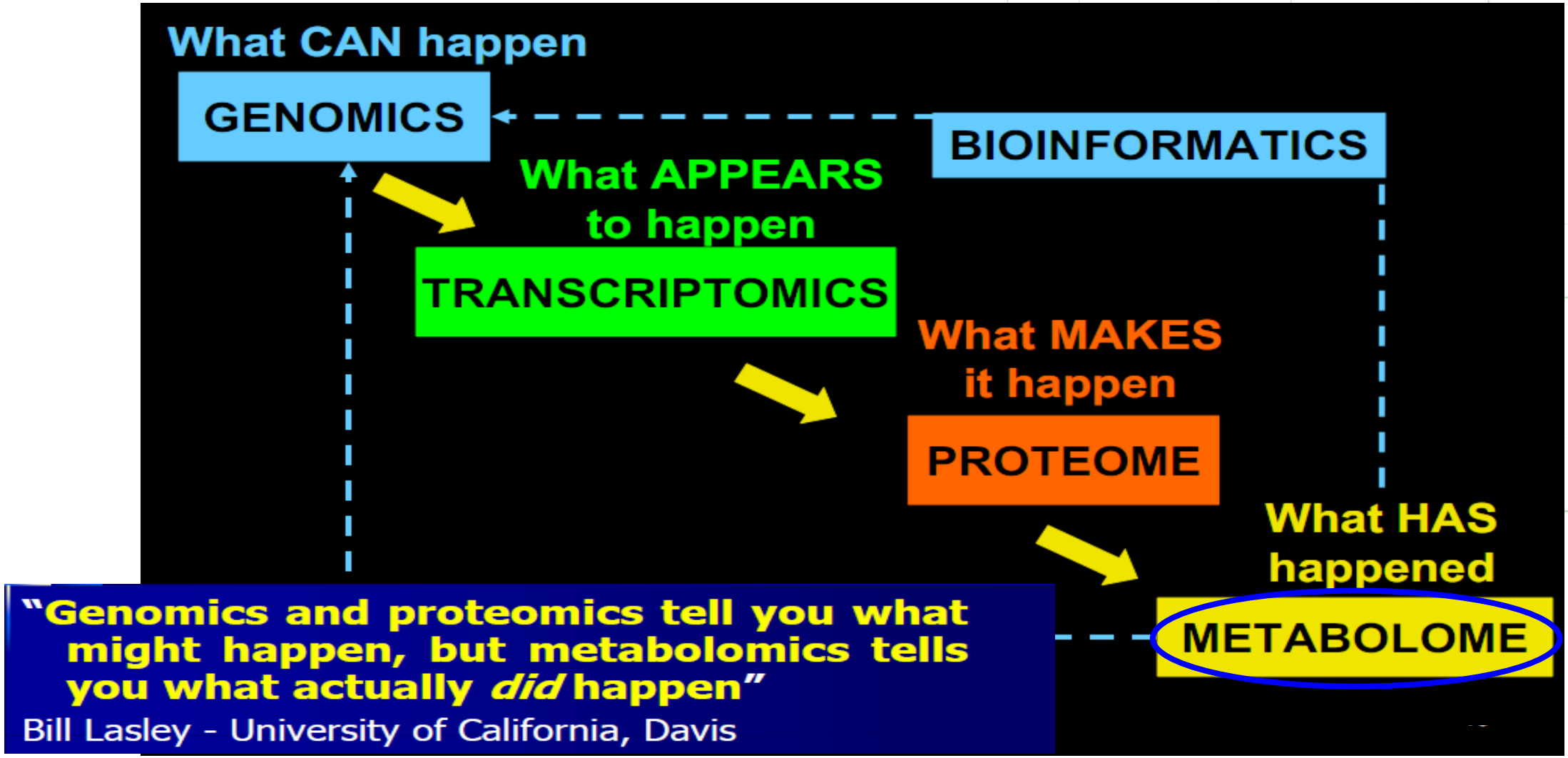
It offers the opportunity to understand how metabolism occurs in the cell, to study and model metabolism, to investigate the compatibility of the functioning of the material elements of the biological system and, consequently, to accelerate the process of creating drugs.






Biological event
Measurable chemical response
Information about the underlying status of the biological system

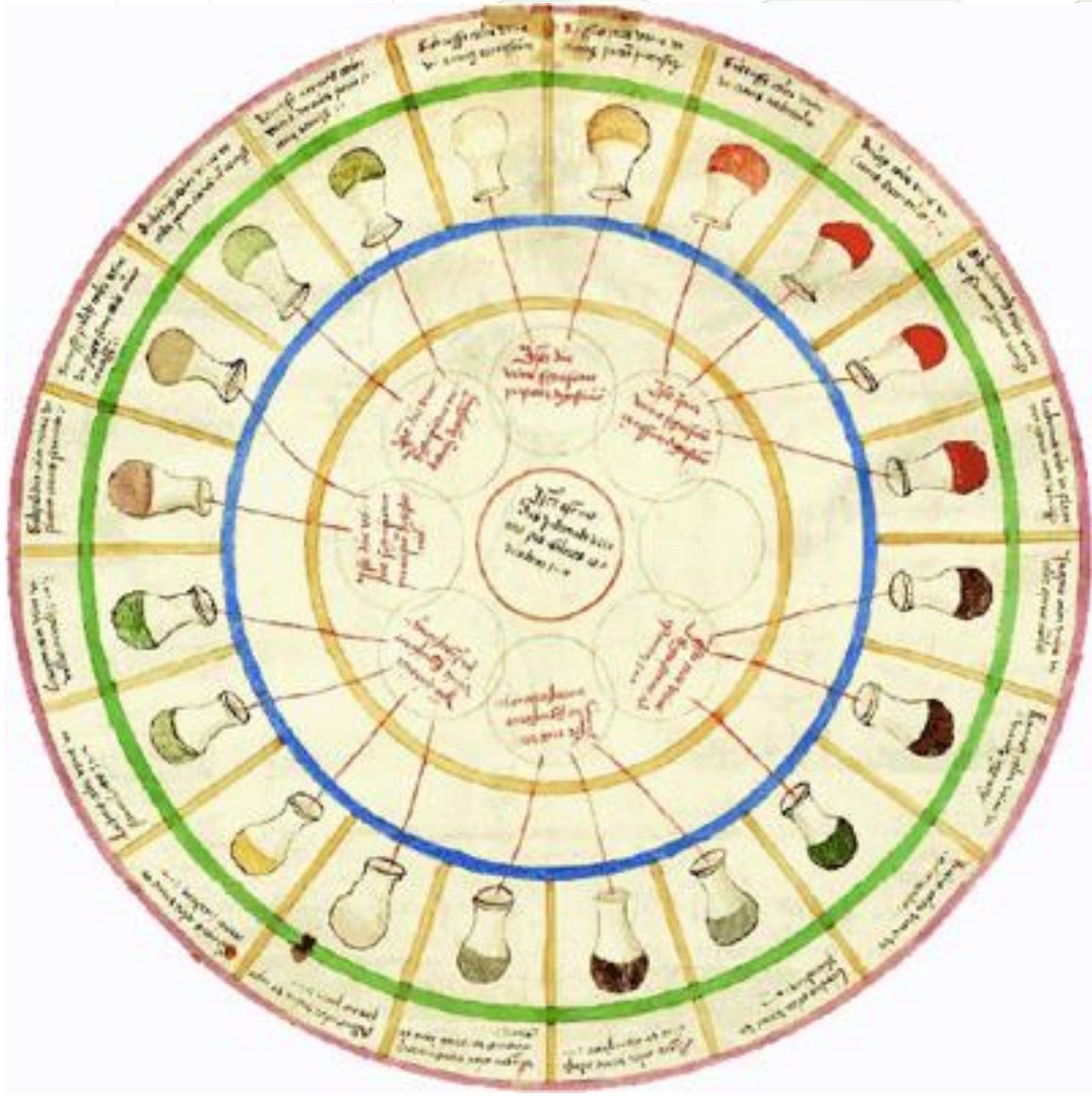
The Omics Cascade



HOW DID IT START?

Metabonomics of yore

This urine wheel was published in 1506 by Ullrich Pinder, in his book Epiphanie Medicorum. It describes the possible colours, smells and tastes of urine, and uses them to diagnose disease.



Ullrich Pinder :physician in Nuremberg and an author of medical writings.

Jeremy K. Nicholson and John C. Lindon, 2008, Nature, 455, 1054-1056

Human Metabolome Project

- In 2005, the first web-based metabolomics database, METLIN, for characterizing human metabolites was developed at Scripps Research Institute and contained over 10,000 metabolites and tandem mass spectrometry spectral data.
- In **2007**, the Human Metabolome Project, led by Dr. David Wishart, Canada, completed the first draft of the human metabolome, consisting of a database of approximately 2500 metabolites, 1200 drugs, and 3500 food components.
- By 2015, METLIN contained over 240,000 metabolites, representing the largest repository of data in metabolomics.Similar projects are ongoing in several plant species.
- In 2015, real-time metabolome profiling was demonstrated for the first time.

Summary of metabolomic databases

Database name	URL or web address	Comments
Human metabolome database	http://www.hmdb.ca	Largest and most complete of its kind. Specific to humans only
BioMagResBank (BMRB – metabolimics)	http://www.bmrb.wisc.edu/metabolomics/	Emphasis on NMR data, no biological or biochemical data Specific to plants (Arabadopsis)
BiGG (database of biochemical, genetic and genomic metabolic network reconstructions)	http://bigg.ucsd.edu/home.pl	Database of human, yeast and bacterial metabolites, pathways and reactions as well as SBML reconstructions for metabolic modeling
Fiehn metabolome database	http://fiehnlab.ucdavis.edu/compounds/	Tabular list of ID'd metabolites with images, synonyms and KEGG links
Golm metabolome database	http://csbdb.mpimp-golm.mpg.de/csbdb/gmd/gmd.html	Emphasis on MS or GC–MS data only No biological data Few data fields Specific to plants
METLIN metabolite database	http://metlin.scripps.edu/	Human specific Mixes drugs, drug metabolites together Name, structure, ID only
NIST spectral database	http://webbook.nist.gov/chemistry/	Spectral database only (NMR, MS, IR) No biological data, little chemical data Not limited to metabolites
Spectral database for organic compounds (SDBS)	http://www.aist.go.jp/RIODB/SDBS/cgi-bin/direct_frame_top.cgi?lang=eng	Spectral database only (NMR, MS, IR) No biological data, little chemical data Not limited to metabolites

Human Metabolome Data Bases

- It is a comprehensive, high-quality, and freely accessible online database of small molecule metabolites present in the human body.
- One of the first databases dedicated to metabolomics, the HMDB facilitates human metabolomics research, including the identification and characterization of human metabolites by NMR spectroscopy, GC-MS spectrometry, and LC/MS spectrometry.

HMDB contains three types of data

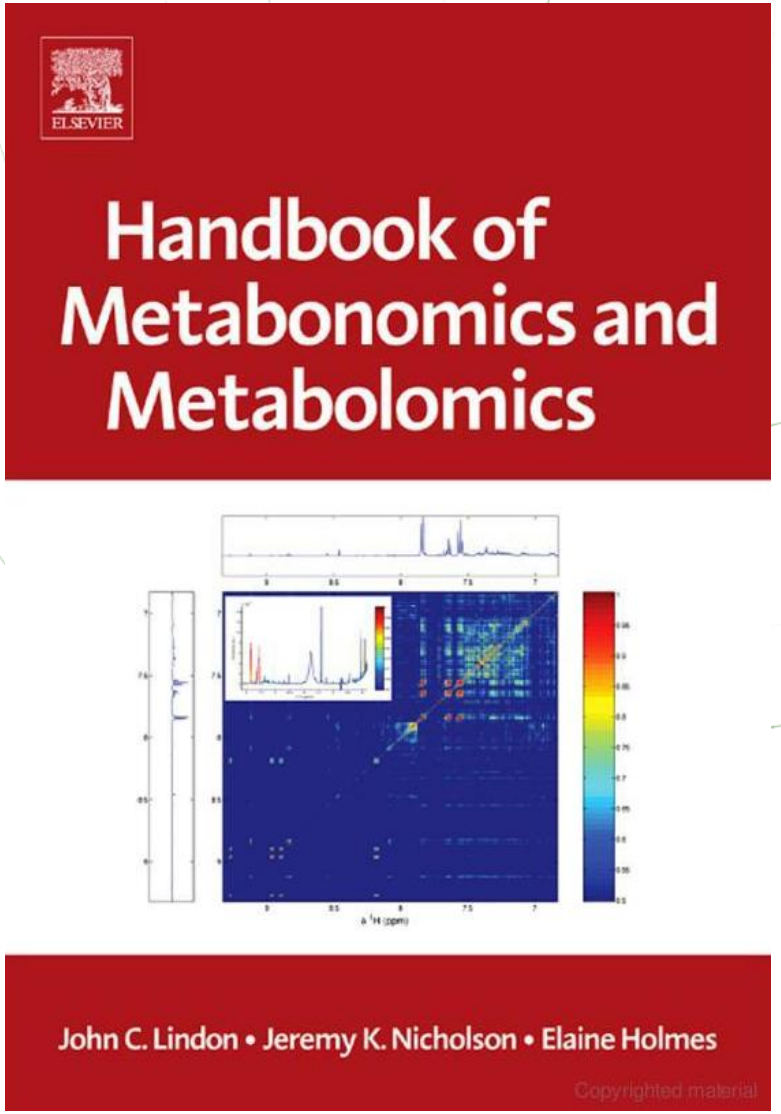
- **Chemical data**
- **Clinical data**
- **Biochemical data / molecular biology**

https://www.hmdb.ca/

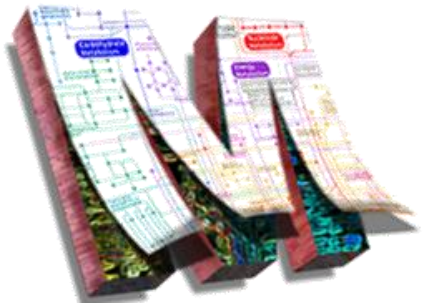
What's in a name?

"**Metabolomics** is the determination of the complete set of context-dependent, low molecular weight metabolites/intermediates that vary with the physiological, developmental, or pathological state of the cell, tissue, organ, or organism."(S. Oliver, 2002)

"**Metabonomics** is the quantitative measurement of the multiparametric, time-related metabolic response of living systems to pathophysiological stimuli and genetic changes."(J.K. Nicholson, 1999)



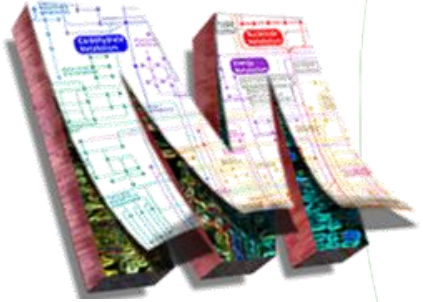
Advantages of metabolomics



- Generate metabolic “signatures”
- Monitor/measure metabolite flux
- Monitor enzyme/pathway kinetics
- Assess/identify phenotypes
- Monitor gene/environment interactions
- Track effects from toxins/drugs/surgery
- Monitor consequences from gene KOs
- Identify functions of unknown genes

from: History of the Omics Cascade Asa Wheelock, Ph.D. Division of Respiratory Medicine & Karolinska Biomics Center

Limitations of metabolomics



- **Inability to** comprehensively **profile all of the metabolome** even with multiple analytical technique available
- **Minor (low concentration) metabolites** are difficult to measure but often of critical importance
- Controlling **analytical variability** is a problem with multi-analyte samples (also reproducibility over time)
- **Biological variance**
- **Dynamic range** of most instrumental approaches is insufficient

General flow for metabolomics

1. Is there a difference between samples?
2. What is the difference between samples?
3. What is the reason of difference?

Main experimental designs used:

Control-Treatment Studies (Unpaired and Paired)

Unpaired Control-Treatment Studies:
In this design, two distinct groups of participants are assigned to a treatment group and a control group. The effect of the treatment is measured by comparing the outcomes between the two groups.
Paired Control-Treatment Studies:
In this design, each participant undergoes both the treatment and the control condition. Outcomes are measured before and after the treatment and compared within the same participant, eliminating inter-individual variability,

Longitudinal Studies

Longitudinal studies follow a group of participants over time, collecting data at regular intervals. This type of design allows researchers to observe the progression of a condition or phenomenon over time

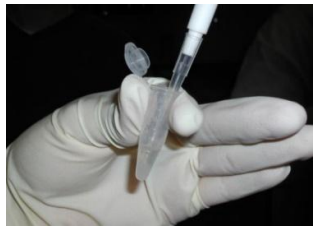
Cross-Over Studies

In a cross-over study, each participant receives multiple treatments in sequence, with a washout period (waiting time) between them. This allows the effects of different treatments to be compared within the same participant

Metabolomics workflow

Design

- *In vivo /in vitro*
- Sample size
- Randomization

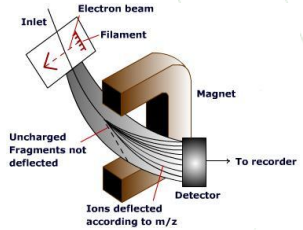


Sampling

- Collection
- Storage
- Extraction

Separation

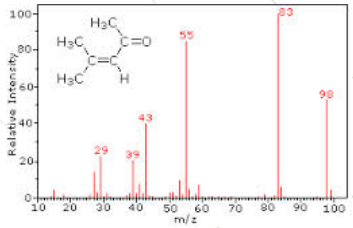
- **Chromatography**
- **Column**
- **Mobile phase**



Detection

Choice of detector

Data acquisition



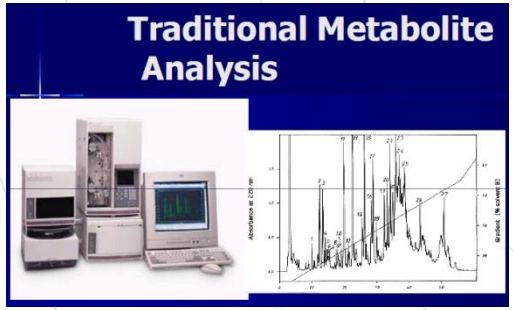
Data Processing

- **Pathway elucidation**
- **Network modelling**

Metabolomics Workflow

- Profiling
  - Finding metabolites with statistically significant variation in abundance within a small set of experimental and control samples
- Identification
- Validation
  - Validating the statistical significance of the metabolites IDs
  - Validate against much larger sample sets to eliminate the effect of natural/biological variation
- Interpretation
  - Evaluation of discovered metabolomic markers in context of relevant biological systems

Metabolomic approaches

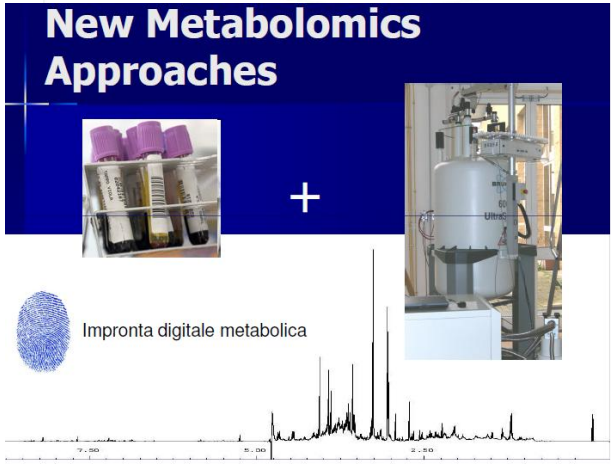


1. Separation Techniques

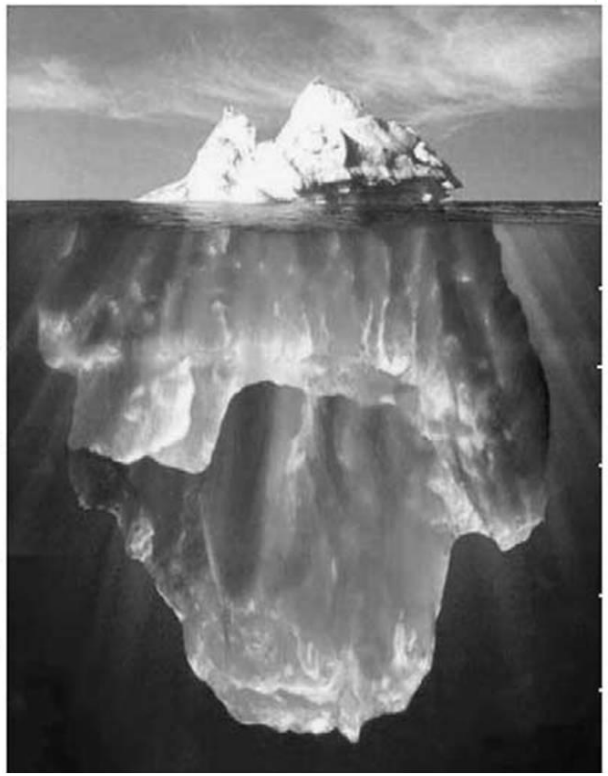
- **Gas Chromatography (GC)**
- **Capillary Electrophoresis (CE)**
- **High Performance Liquid Chromatography (HPLC)**
- **Ultra Performance Liquid Chromatography (UPLC)**

2. Detection Techniques

- **Nuclear Magnetic Resonance Spectroscopy (NMR)**
- **Mass Spectrometry (MS)**



Sensitivity of various metabolomic tools



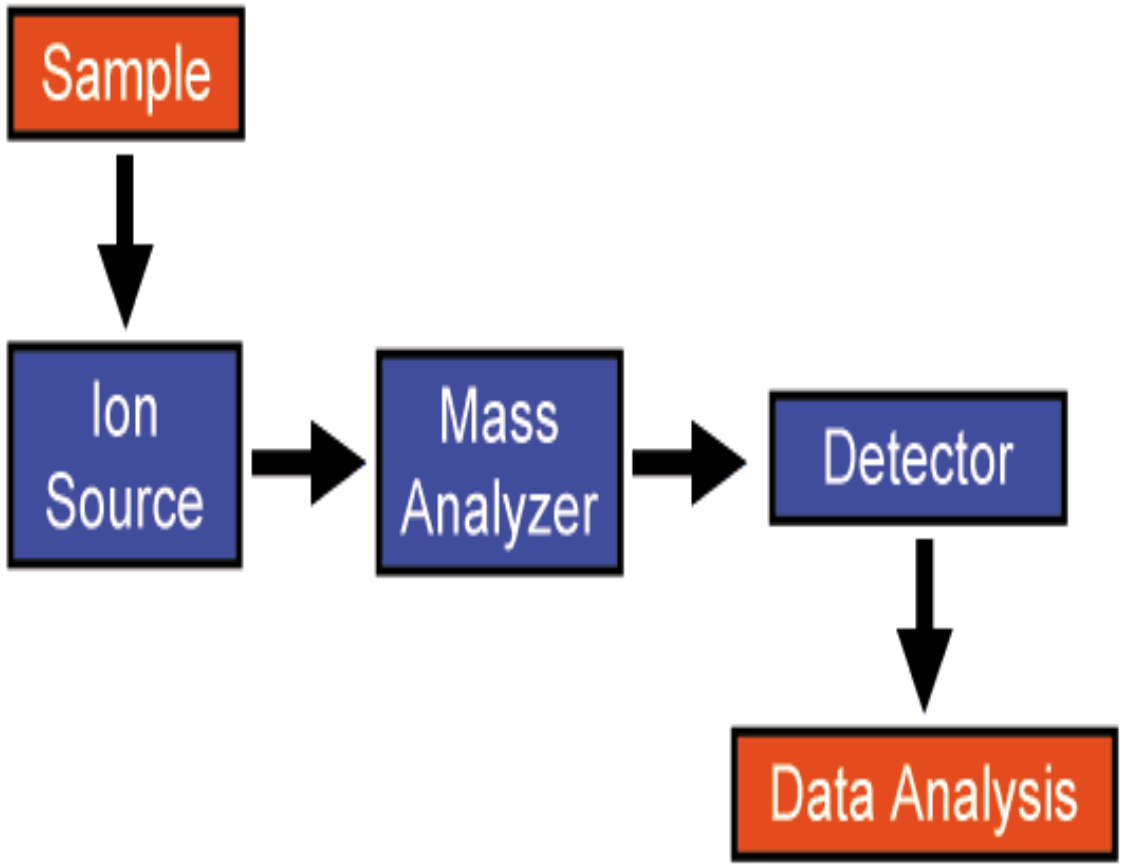
- **NMR has rapid analysis times but has relatively low sensitivity- the tip of the iceberg.**
- **GC-MS and LC-MS provide good selectivity and sensitivity.**

Detection Technique - NMR

- Advantages:
- Non-destructive, non-biased
- Easily quantifiable
- Requires little or no separation
- Permits identification of novel compounds
- Does not require chemical derivatization
- Particularly amenable to cmpds less tractable to GC-MS or LC-MS (sugars, amines, volatile ketones, & relatively non-reactive compounds)

Detection Technique - MS

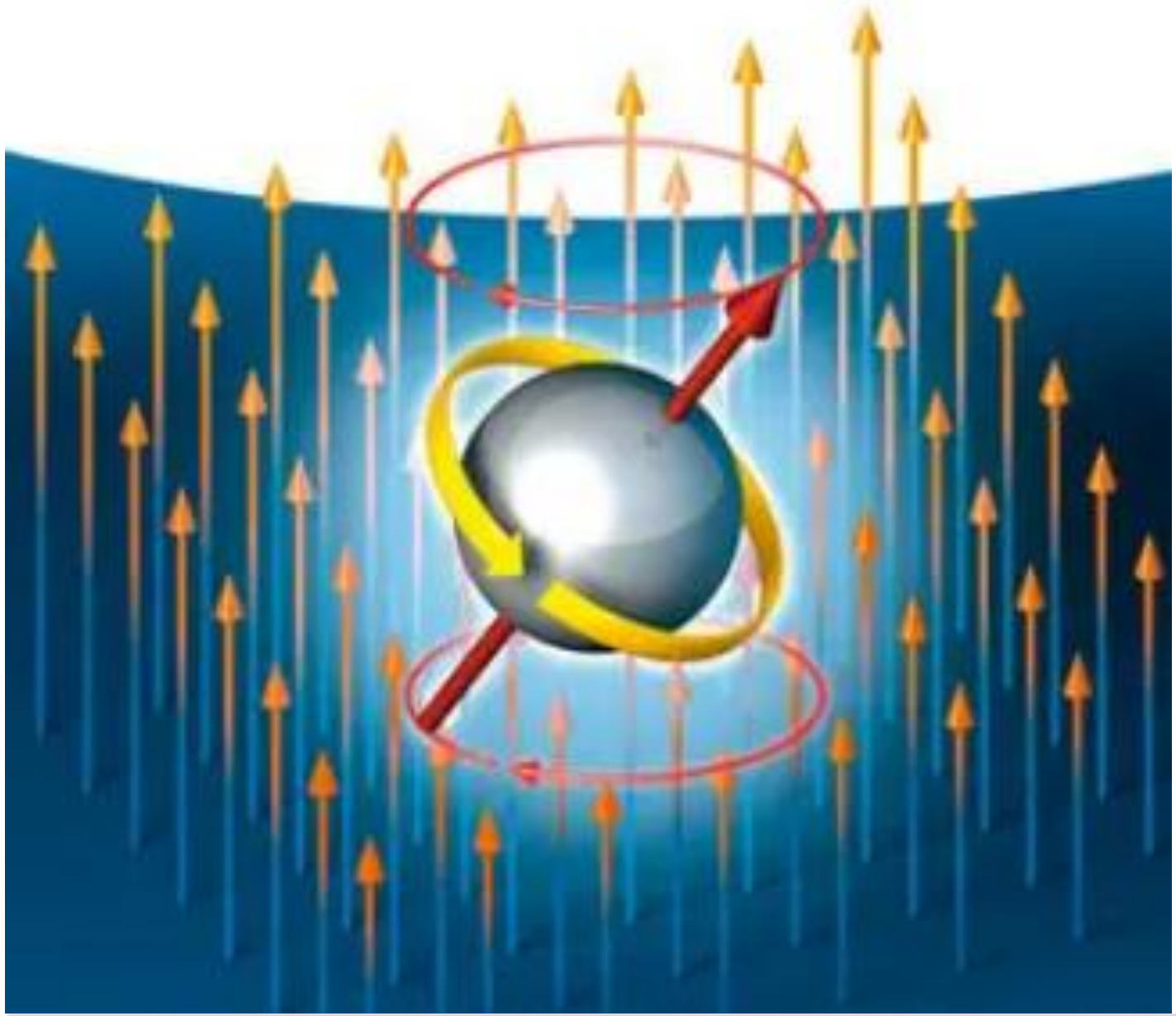
- To identify and to quantify metabolites
- Serves to both separate and to detect
- Mass to charge ratios
- Using electron beam
- Ion source, mass analyzer and detector



NMR versus MS

- Quantitative, fast
- Requires no work up or separation
- Allows ID of 300+ cmpds at once
- Not sensitive
- Needs MS or 2D NMR for positive ID

- Very fast
- Very sensitive
- Allows analysis or ID of 3000+ cmpds at once
- Not quantitative
- Requires work-up



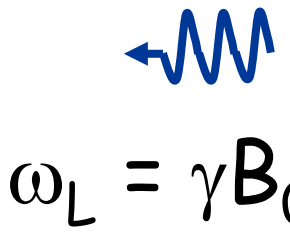
NMR spectroscopy:
Some basic notions

Nuclear Magnetic Resonance

The nuclei of some atoms immersed in a static magnetic field release energy if perturbed by an oscillating electromagnetic field at a suitable frequency. If I can capture this energy (signal) I can deduce structural information on the analyzed sample

Nuclear Magnetic Resonance

All nuclei that contain odd numbers of protons or neutrons have an intrinsic magnetic moment and angular momentum. The most commonly measured nuclei are ^{1}H (the most receptive isotope at natural abundance) and ^{13}C, although nuclei from isotopes of many other elements
(e.g. ^{15}N, ^{14}N ^{19}F, ^{31}P, ^{17}O,^{29}Si, ^{10}B, ^{11}B, ^{23}Na, ^{35}Cl, ^{195}Pt) can also be observed

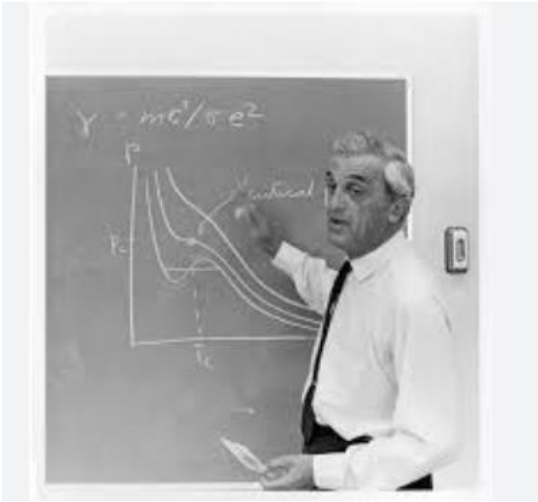


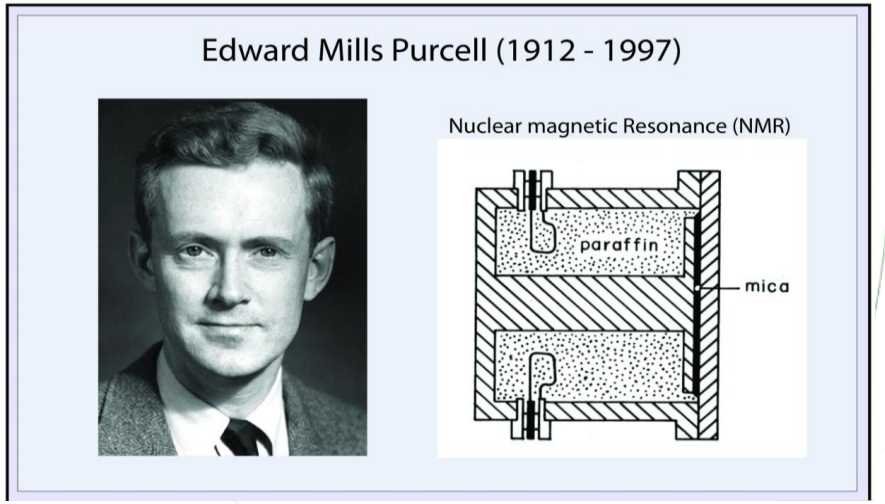
$$\omega_L = \gamma B_0$$

Nucleo	spin *I*	Abb nat. (%)
^{1}H	1/2	99.98
^{2}H	1	0.015
^{13}C	1/2	1.108
^{19}F	1/2	100
^{23}Na	3/2	100
^{31}P	1/2	100

NMR time-line

- **1945** First observation of an NMR signal Bloch et al. Stanford Un. (^1H in H_2O) Purcell et al. Harvard Un. (^1H in paraffin) (Nobel Prize 1952)
- **1950** Discovery of the Chemical Shift
- **1961** First commercial CW spectrometer
- **1970** First commercial FT spectrometer
- **1976** First 2D experiments (Ernst 1991 Nobel Prize)
- **1980** Second generation NMR spectrometers



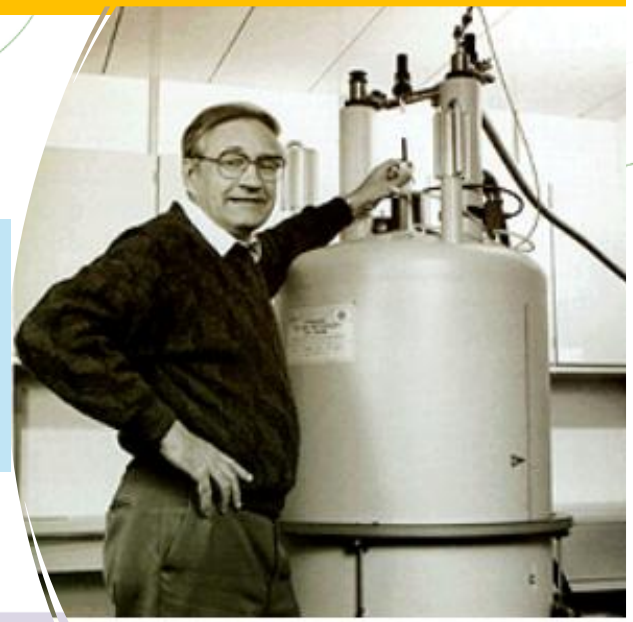


1952 Nobel Prize in Physics : From Radiation to Resonance

"for their development of new methods for nuclear magnetic precision measurements and discoveries in connection therewith"

1976 Nobel Prize in Chemistry:

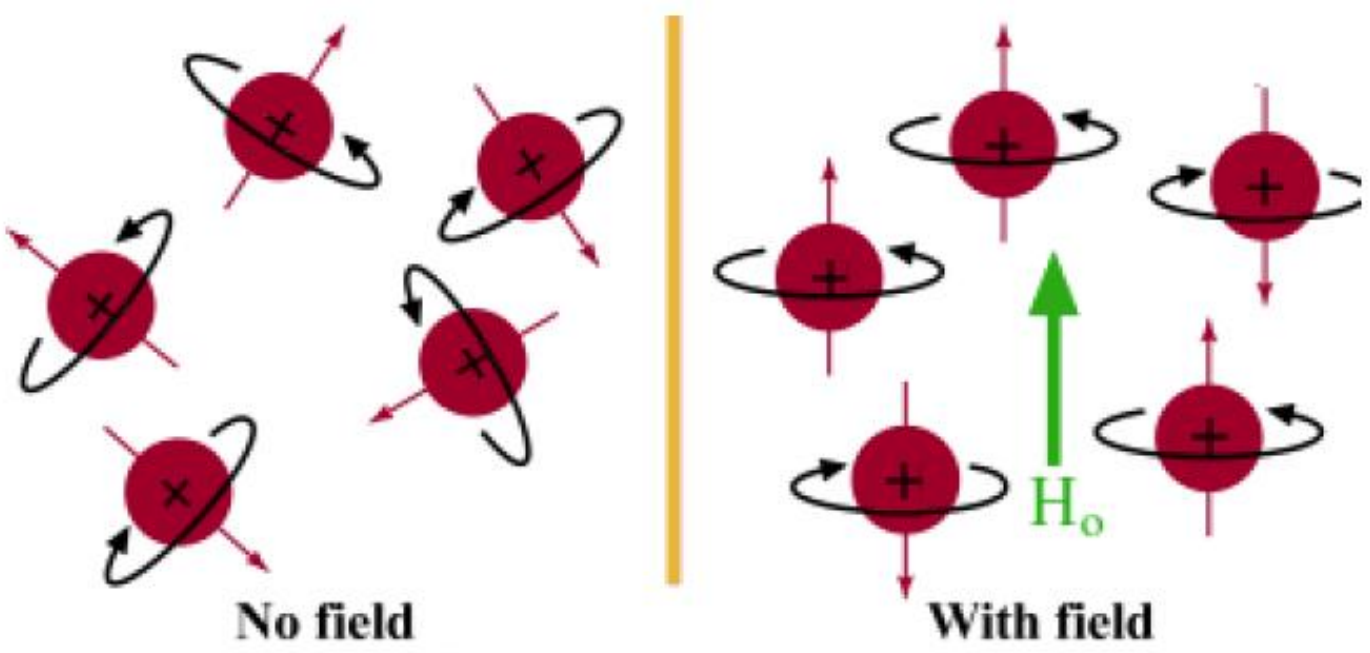
"for his groundbreaking contributions to the development of high-resolution nuclear magnetic resonance (NMR) spectroscopy."



NMR spectroscopy is an absorption spectroscopy. It allows to detect the absorption of electromagnetic radiation by a molecule.

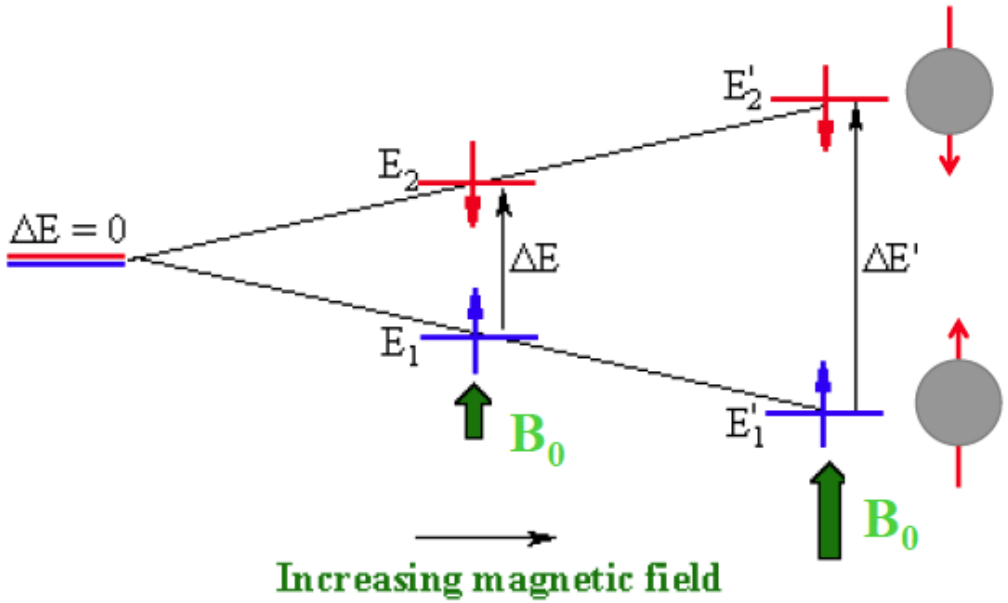
Le regioni dello spettro elettromagnetico				
Regione dello spettro	**Lunghezza d'onda (Å)**	**Lunghezza d'onda (cm)**	**Frequenza (Hz)**	**Energia (eV)**
Radio	$>10^{9}$	>10	$<3\times10^{9}$	$<10^{-5}$
Microonde	10^{9}-10^{6}	10-0.01	3×10^{9} - 3×10^{12}	10^{-5} - 0.01
Infrarosso	10^{6}-7000	0.01-7×10^{-5}	3×10^{12}- 4×10^{14}	0.01 -2
Visibile	7000-4000	7×10^{-5} - 4×10^{-5}	4×10^{14}-7.5×10^{14}	2 -3
Ultravioletto	4000-10	4×10^{-5} - 10^{-7}	7.5×10^{14}-3×10^{17}	3- 10^{3}
Raggi X	10-0.1	10^{-7} - 10^{-9}	3×10^{17} - 3×10^{19}	10^{3} - 10^{5}
Raggi gamma	<0.1	$<10^{-9}$	$>3\times10^{19}$	$>10^{5}$

When a nucleus with spin is immersed in a magnetic field, the nucleus, like the needle of a compass, is subjected to a torque of forces that cause it to rotate to align with the external magnetic field.



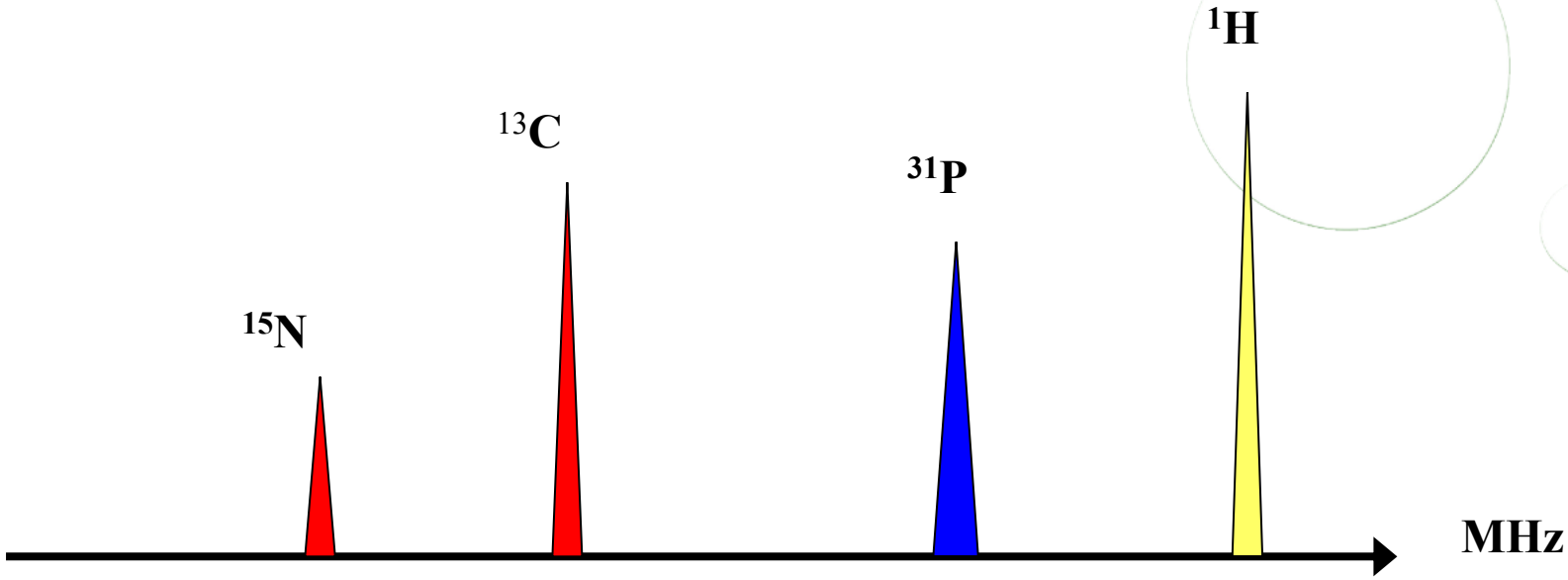
Nuclear Magnetic Resonance

The nucleus immersed in the magnetic field is able to absorb and release electromagnetic radiation with a frequency corresponding to the energy jump between the two levels:



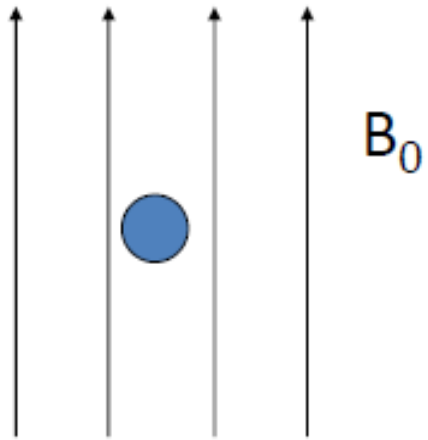
DE (energy jump) or the frequency measured for the excited nucleus **is proportional to the applied magnetic field, to the type of nucleus and also to the chemical environment of the nucleus itself**. If this were not (proportional to the chemical environment) all the protons would give the same signal...

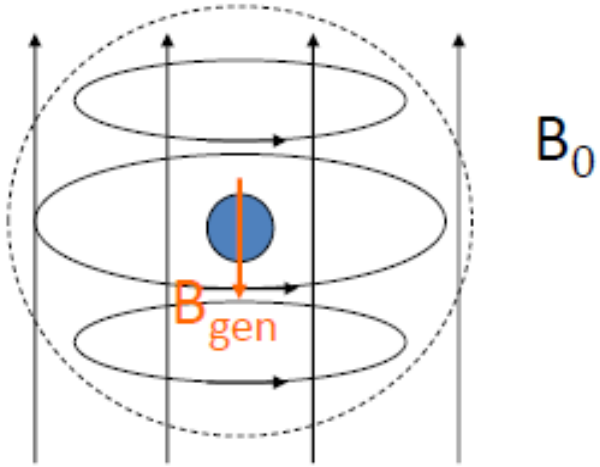
By irradiating the sample with appropriate radio frequencies and measuring its absorption, we could obtain the “spectrum” of the compound under examination.



Chemical shift

If all protons absorbed at the same frequency in an external magnetic field B0, the information would be practically zero.



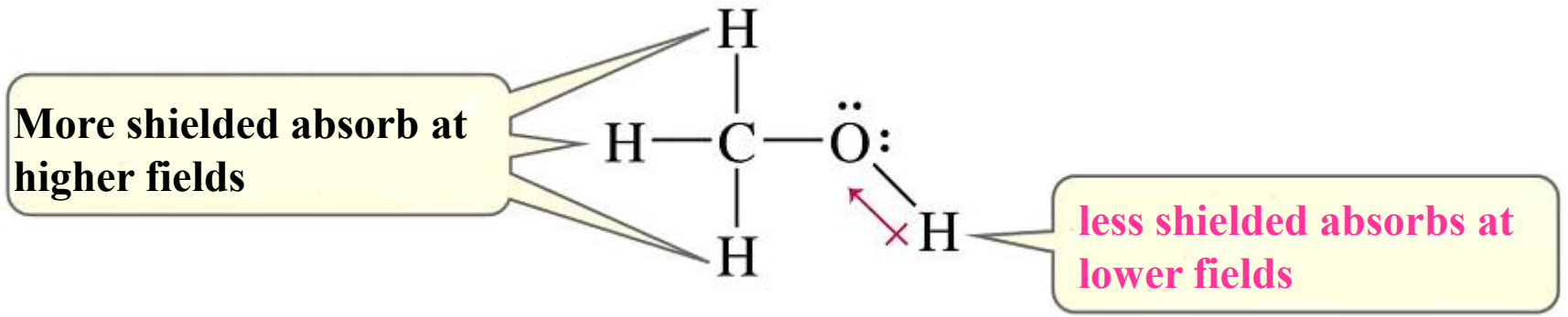


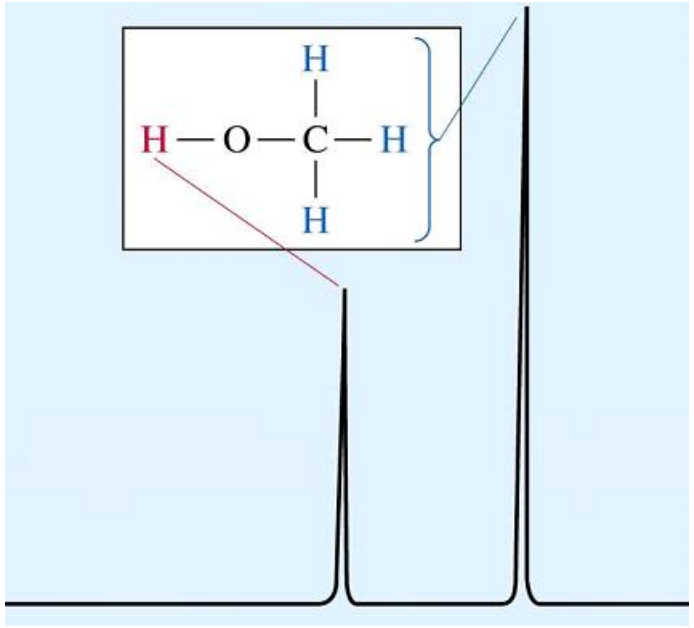
The electrons of an atom induce a small magnetic field that opposes the main one

Nuclei of the same atom absorb slightly different frequencies due to their “chemical surroundings”

$\delta B_0 = -\sigma B_0$ σ screen constant

Nuclei interact with the local magnetic field.

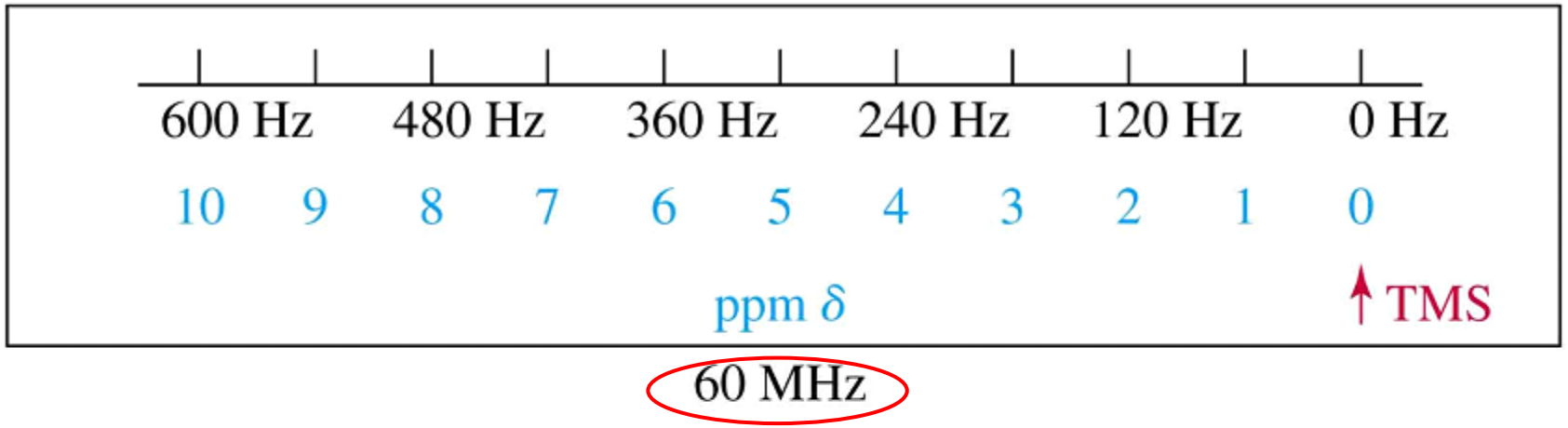




$B_0 \rightarrow$

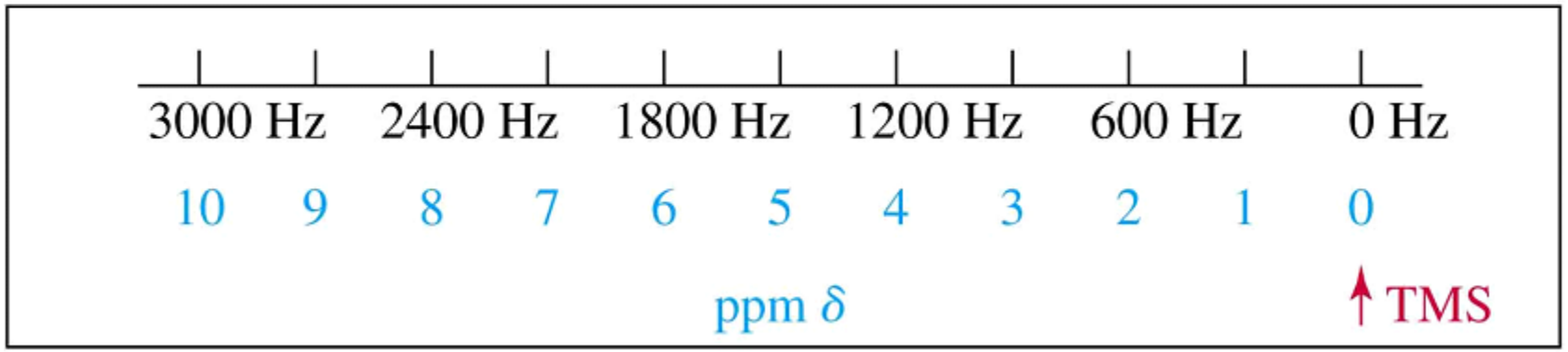
Depending on the chemical environment, the protons in a molecule are shielded in different amounts

δ scale



or TSP

60 MHz



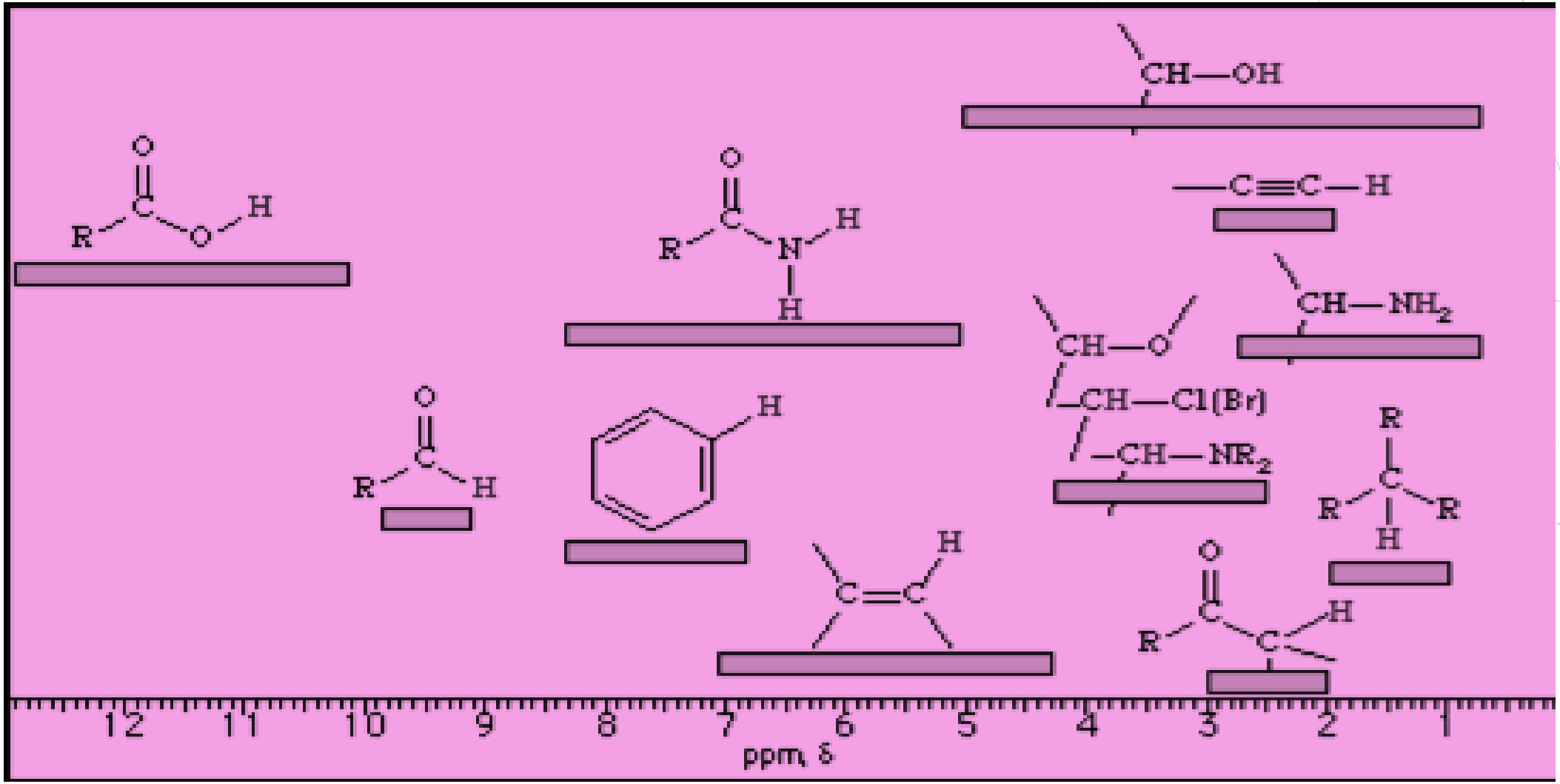
or TSP

300 MHz

standards used respectively in organic or polar solvents whose signals is at 0 in the ^{1}H NMR spectrum

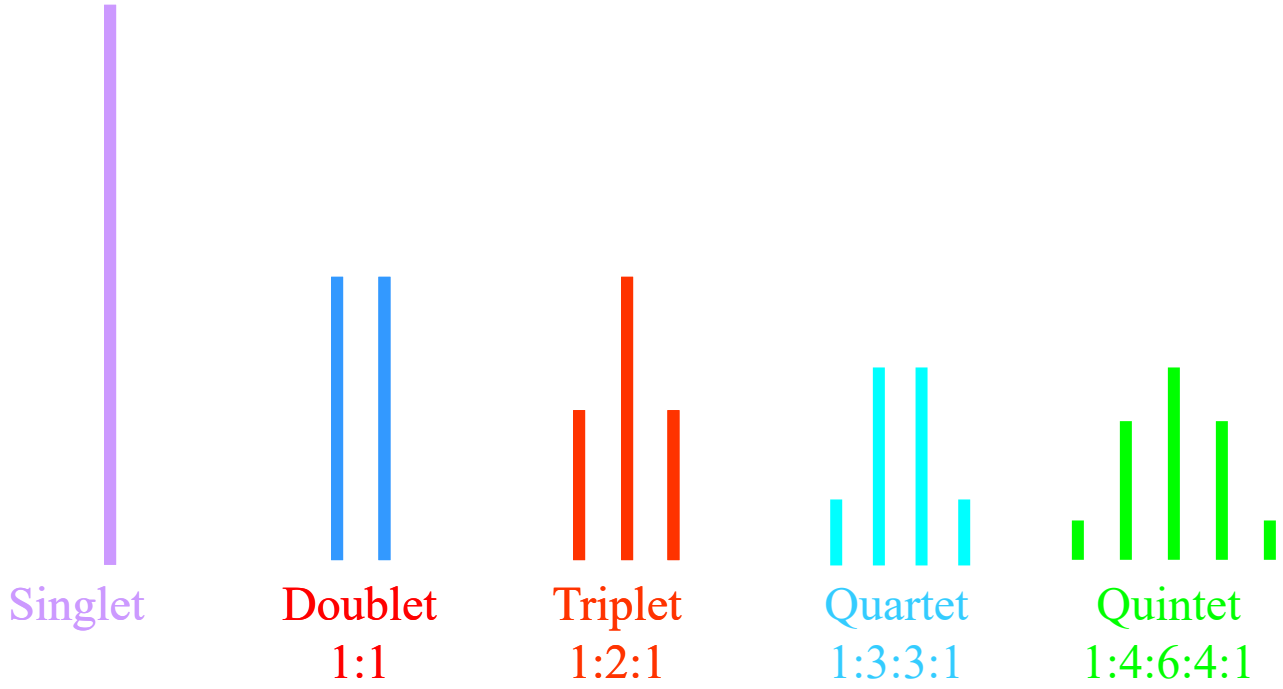
The value of the chemical shift δ does not depend on the power of the magnet (on the equipment used)

^{1}H NMR correlation diagram



“Multiplicity” of NMR signals

- Each chemically different proton (carbon, nitrogen, etc.) corresponds to a signal of different frequency.
- The “shape” of the signal in an NMR spectrum is often not a simple absorption peak but can take on very different aspects.

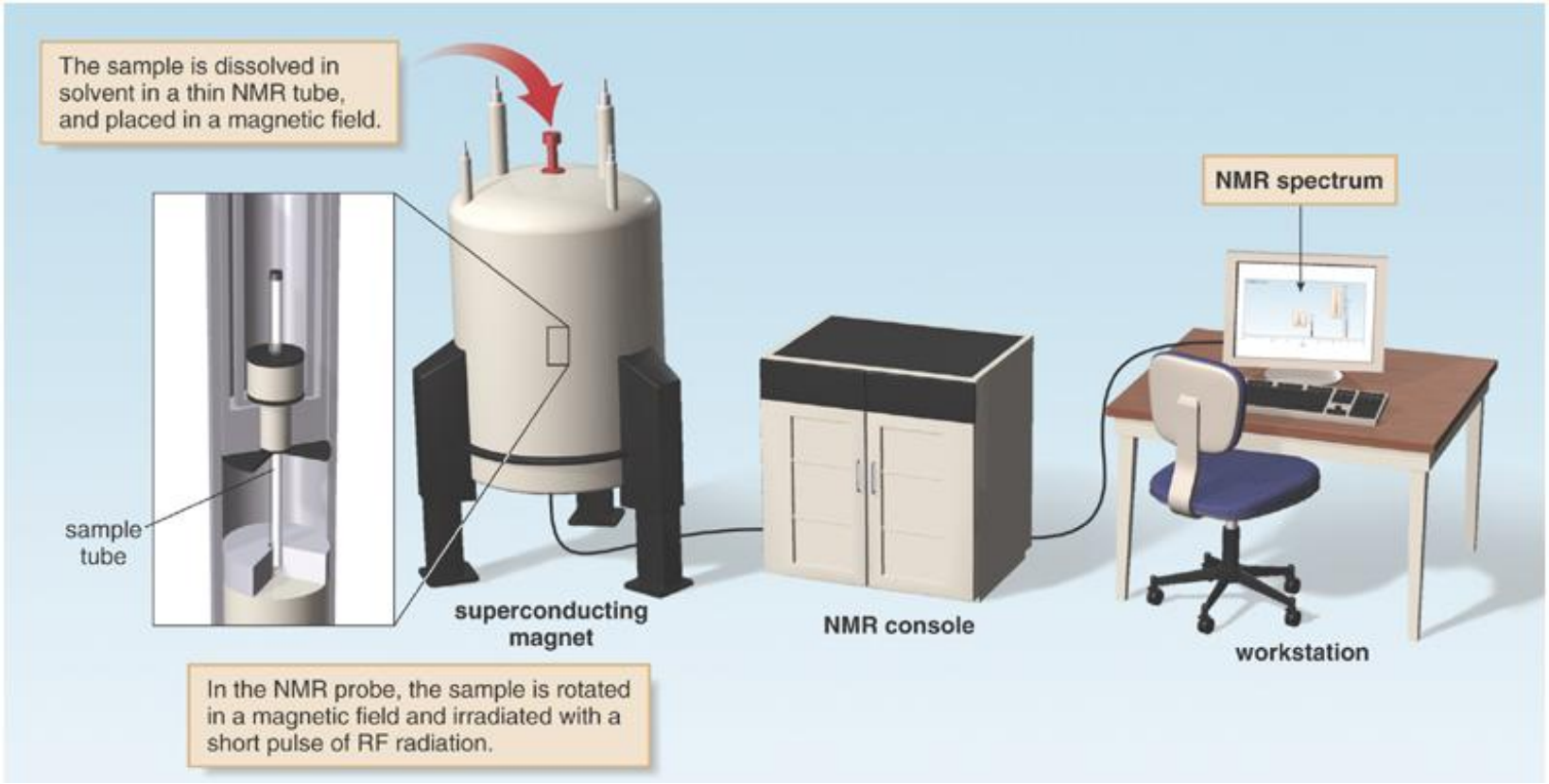


The multiplicity of signals is due to the chemical environment, therefore to the nearby chemical groups

Magnetic Resonance Imaging

- Solution NMR (high resolution) → *Chemists' NMR and for metabolomic*
- Solid state NMR (CP-MAS)
- *In vivo* NMR
- Magnetic resonance imaging

Nuclear Magnetic Resonance Spectroscopy



An NMR spectrometer. The sample is dissolved in a solvent, usually $CDCl_3$ (deuterochloroform), and placed in a magnetic field. A radiofrequency generator then irradiates the sample with a short pulse of radiation, causing resonance. When the nuclei fall back to their lower energy state, the detector measures the energy released, and a spectrum is recorded. The superconducting magnets in modern NMR spectrometers have coils that are cooled in liquid helium and conduct electricity with essentially no resistance.

Nuclear Magnetic Resonance

Inserting the Sample

The tube with the spinner must be placed on the air flow created above the magnet probe

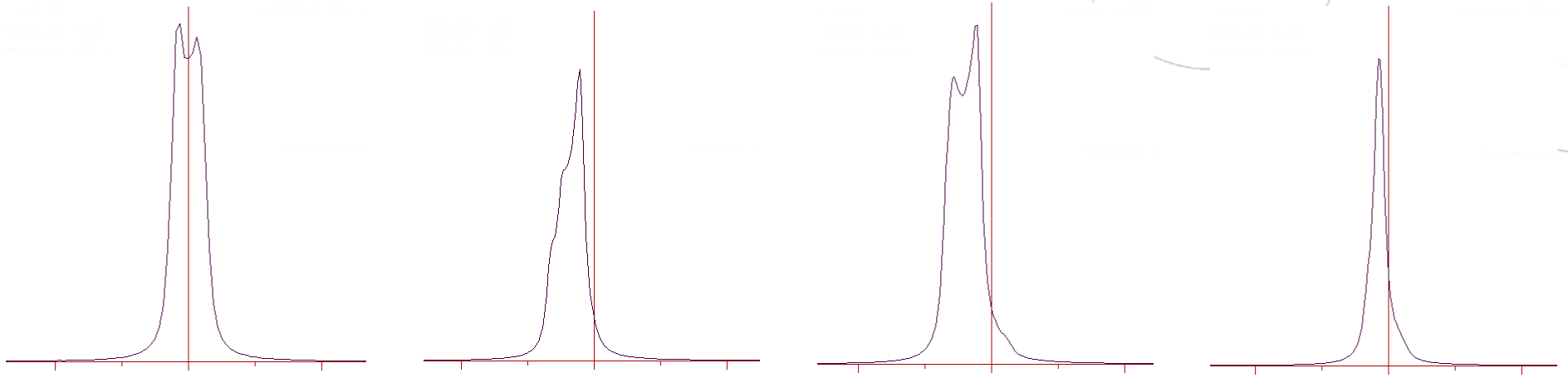
Field-FrequencyLock

The NMR spectrometer has a channel permanently tuned to the resonance frequencies of deuterium (2H)

Nuclear Magnetic Resonance

- After “locking” the lock signal, it is necessary to homogenize the Bo field (make it uniform along the entire sample)

Shimming (field homogeneity)



Examples of the effect of poor homogeneity on a signal

Nuclear Magnetic Resonance

$B(^1H_\alpha) \neq B(^1H_\beta)$

due to microscopic interactions with other nuclei/electrons

Nuclear Magnetic Resonance - analysis of complex mixtures

Nuclear Magnetic Resonance

- Solvent: (about 500 µl)deuterated solvents
- Internal reference: TSP or TMS (a capillary tip)

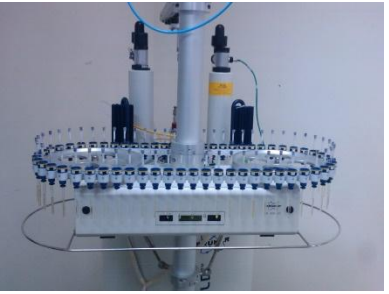
Sample preparation for metabolomic
– *NMR tubes preparation and spectral acquisition*

(A) Always use the same **sample volume** and **sample pH** in NMR tube;

(B) Collect all the1D ^{1}H NMR spectra at one **optimum temperature (typically 298 K);**

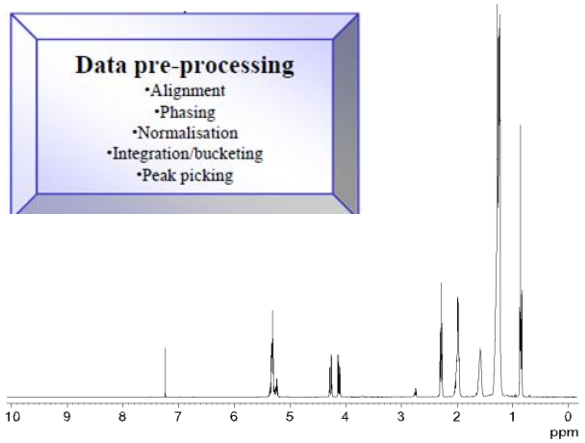
(C) Use the same **internal standard** (TSP or TMS) while recording all the ^{1}H NMR spectra;

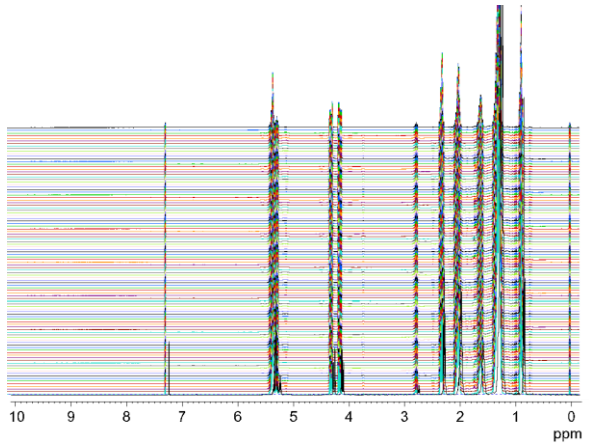
(D) Use the same **acquisition parameters** for recording all the ^{1}H NMR spectra

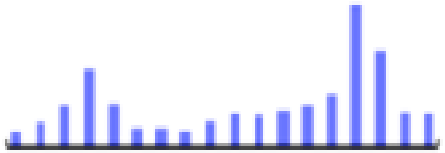


NMR SPECTROSCOPY AND MULTIVARIATE ANALYSIS

Multivariate analysis allows the observations of differences in samples





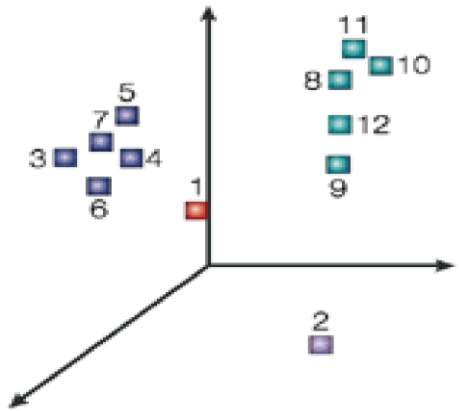


Rectangular bucketing

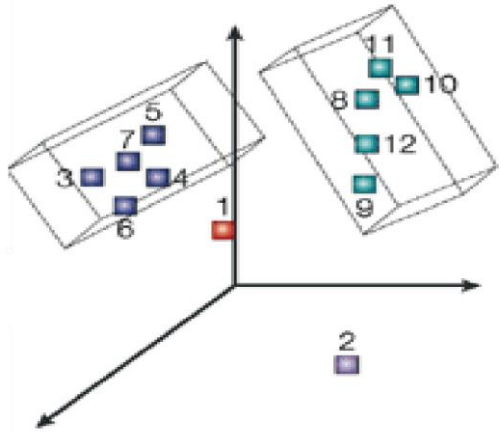
Identification and quantification of metabolites

V1	V2	V3	Vn
1	1.3	0.4	
2	2.3	1.2	
3	2.7	2.1	
4	3.9	4.6	
.....	.....	.....	

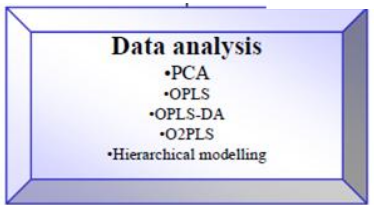
Many variables



Unsupervised mapping of data in three dimensional space

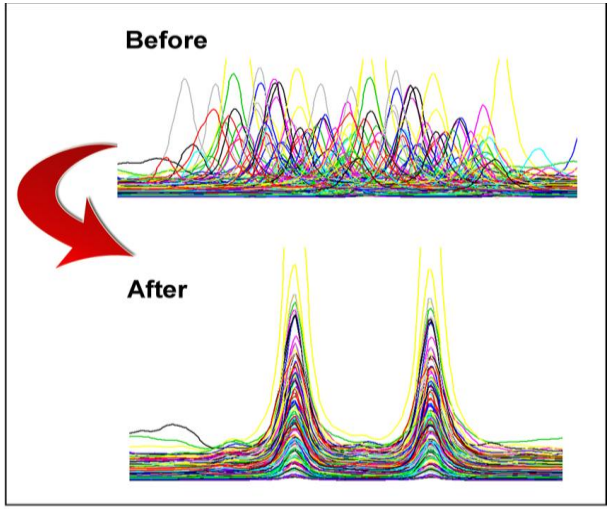


Supervised classification and calculation of confidence limits

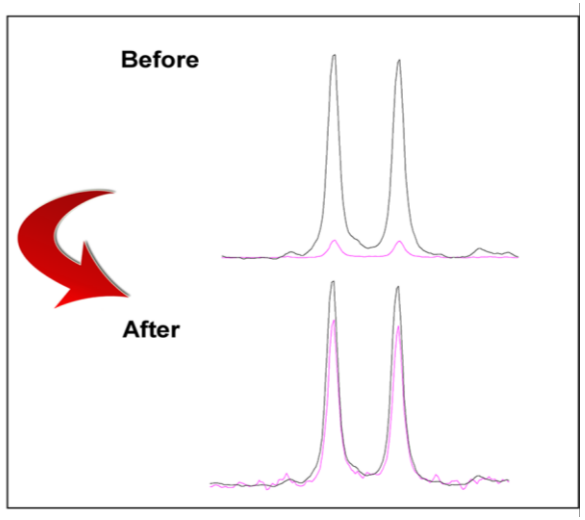


Spectra processing

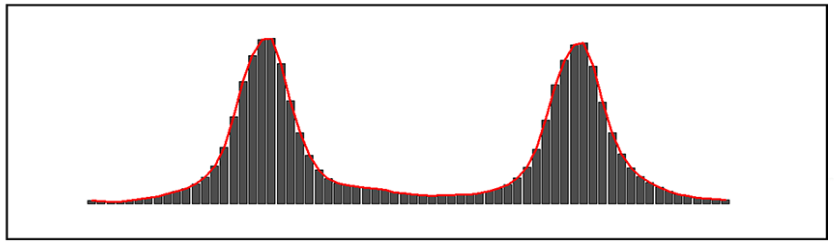
Spectra alignment

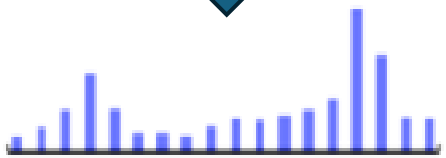


Spectra normalization



Molecule quantification





Rectangular bucketing

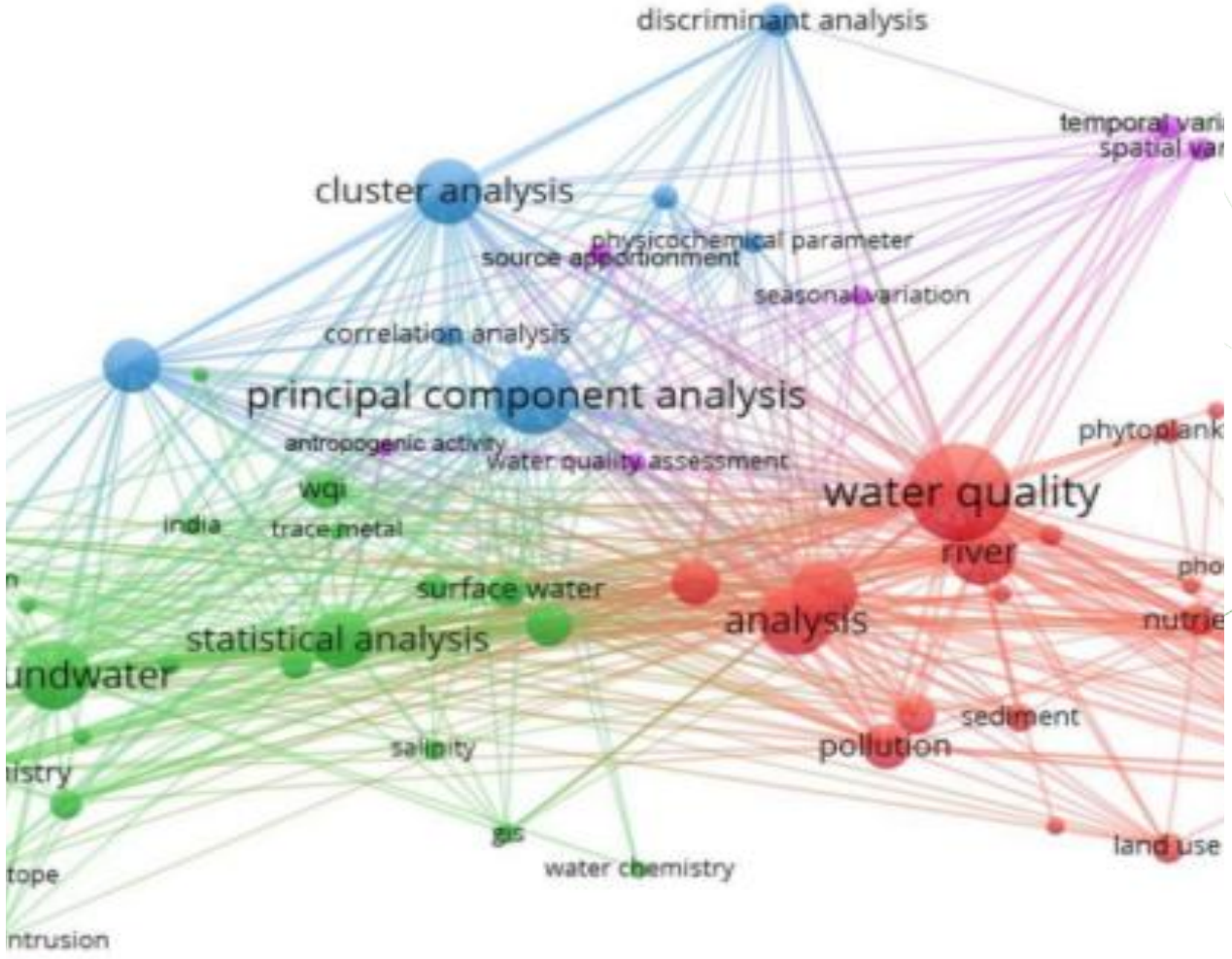
biological matrix

Sample collection

the *quality* of the samples is affected by:

- sample type
- time of collection
- containers used
- preservatives and other additives
- transportation and length of transit time affect

The Handbook of Metabonomics and Metabolomics, *Edited by: John C. Lindon, Jeremy K. Nicholson and Elaine Holmes,* ISBN: 978-0-444-52841-4

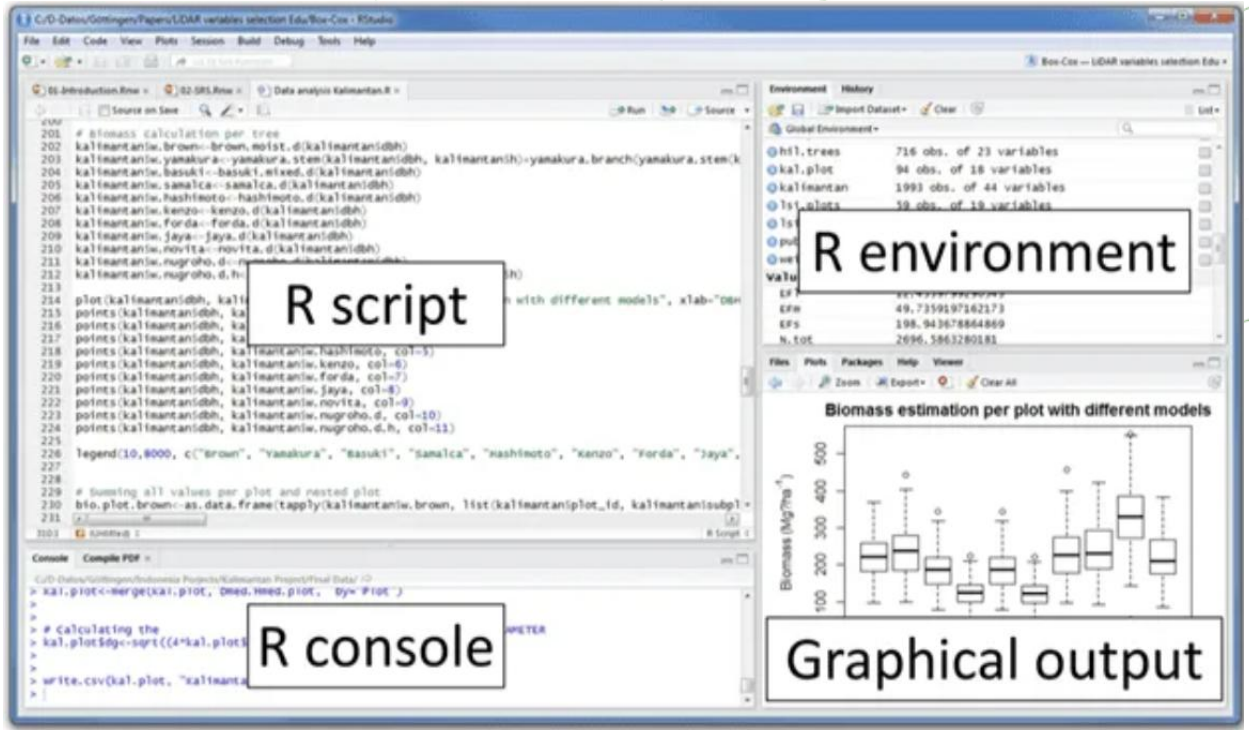


Multivariate data analysis

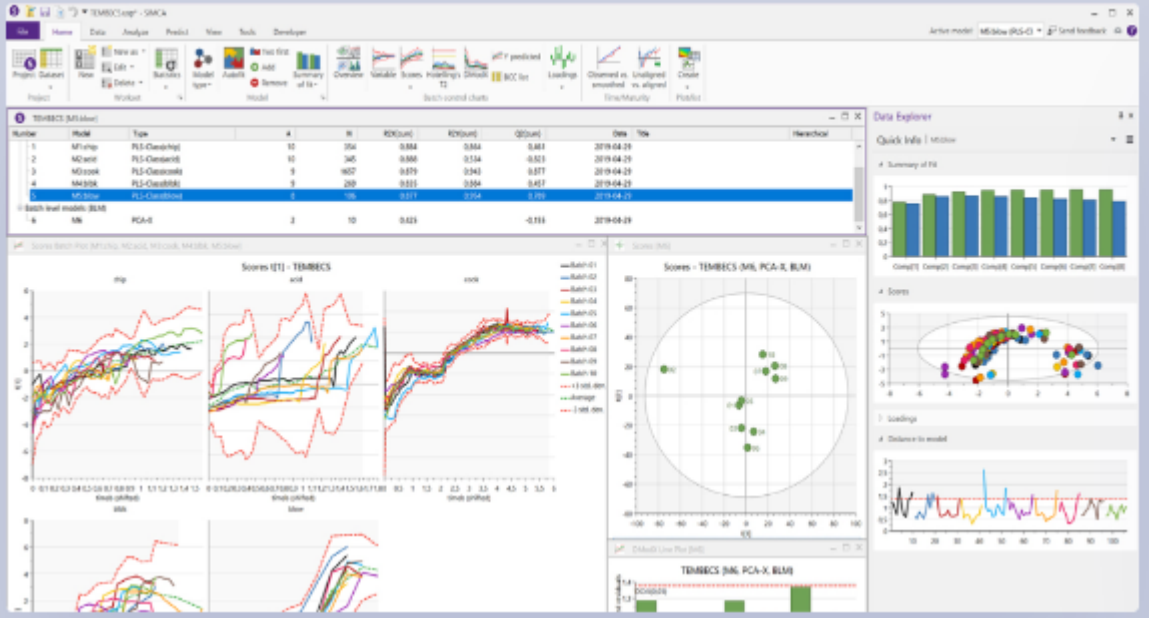
Programming Language for metabolomic

python™

R Programming



Platforms for analyzing metabolomic data



SIMCA
Life Sciences

MetaboAnalyst 6.0 - from raw spectra to biomarkers, patterns, functions and systems biology

Module Overview

Input Data Type	Available Modules (click on a module to proceed, or scroll down to explore a total of 18 modules including utilities)				
LC-MS Spectra (mzML, mzXML or mzData)			Spectra Processing [LC-MS w/wo MS2]		
MS Peaks (peak list or intensity table)		Peak Annotation [MS2-DDA/DIA]	Functional Analysis [LC-MS]	Functional Meta-analysis [LC-MS]	
Generic Format (.csv or .txt table files)	Statistical Analysis [one factor]	Statistical Analysis [metadata table]	Biomarker Analysis	Statistical Meta-analysis	Dose Response Analysis
Annotated Features (metabolite list or table)		Enrichment Analysis	Pathway Analysis	Network Analysis	
Link to Genomics & Phenotypes			Causal Analysis [Mendelian randomization]		

Scaling influences

- UV-scaling, and mean-centering: All variables have had the same chance to influence the model

Centering the data ensures that only the variance relative to the mean is measured, rather than the absolute position of the mean

This is why it is also important to exclude the spectral range of the solvent and other variables

Scaling

- No scaling, but mean-centering: Often, the variable with the highest SD will get too much influence

Scaling: No, Pareto, and UV-scaling

- No scaling (but mean-centering): Useful when all variables are expressed in the same unit, such as with spectroscopic data
- UV-scaling (and mean-centering): Useful when variables are of different kinds and not directly comparable
- Pareto scaling (and mean-centering): Intermediate between the extremes of no scaling and UV-scaling. Gives each variable a variance numerically equal to its initial standard deviation instead of unit variance

UNSUPERVISED ANALYSIS:PCA

PCA creates a visualization of data that **minimizes residual variance** in the least squares sense and **maximizes the variance** of the projection coordinates.

Two PCs form a plane. This plane is a **window into the multidimensional space**. Each observation may be projected onto this plane, giving a score for each.

SUPERVISED ANALYSIS: PLS-DA

Class memberships are explicitly given in PLS-DA; this gives a rotation of the latent variables, such that a maximum separation among the classes is obtained.

- **What is PLS-DA ?**
  - A "dummy" variable is added for each category. PLS is then used to relate X and Y.

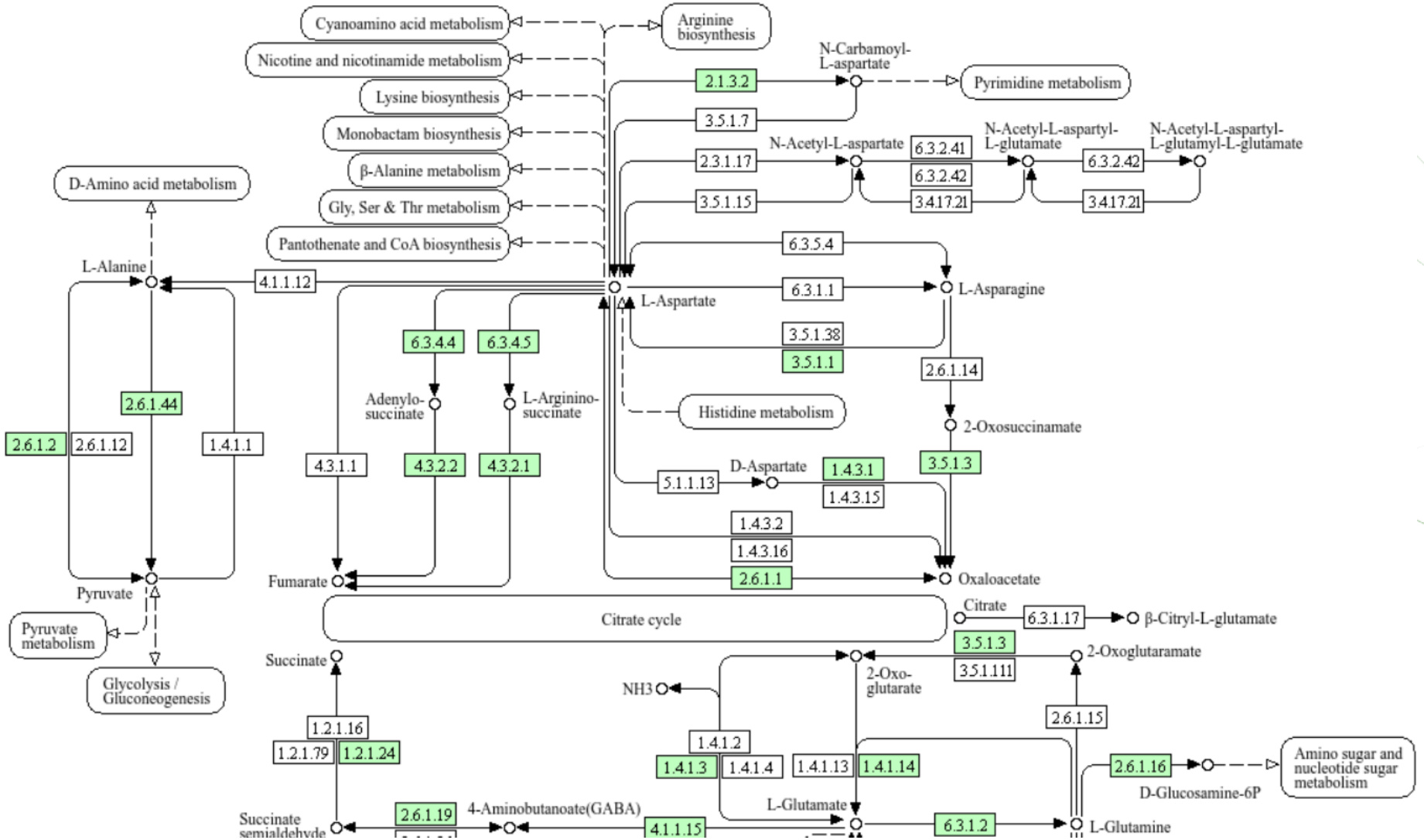
PLS-DA is useful with 2-4 classes; discrimination results may become incomprehensible with too many classes.

SUPERVISED ANALYSIS- OPLS-DA

OPLS-DA relies on a projection of X as does PCA but with **maximum separation**

The **horizontal component** will capture **variation between the groups** and the **vertical dimension** will capture **variation within the groups.**

INTERPRETATION



THANKS!

IR0000032 – ITINERIS, Italian Integrated Environmental Research Infrastructures System
(D.D. n. 130/2022 - CUP B53C22002150006) Funded by EU - Next Generation EU PNRR-Mission 4 “Education and Research” - Component 2: “From research to business” - Investment 3.1: “Fund for the realisation of an integrated system of research and innovation infrastructures”

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani