

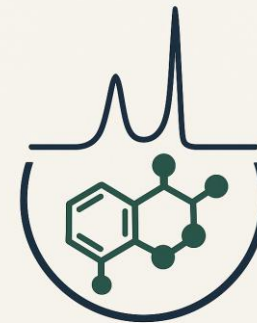


# Exploring the World of Omics: A Multi-Scale Approach to Biology

## 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”



**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

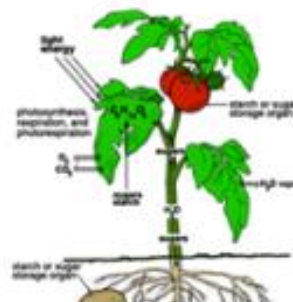
## Natural product chemistry



## Stress response



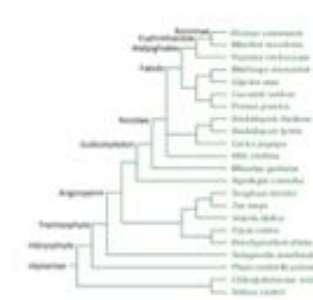
## Physiology



## Interactions among organisms



## Taxonomy



## Food science



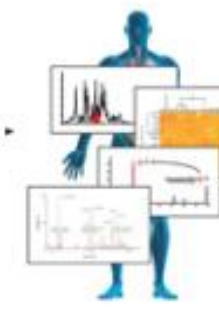
## Pharmaceutical research



## Microbial biology and chemistry

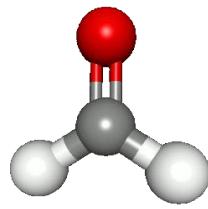


## Human metabolism/ Biomarkers




## Environmental science




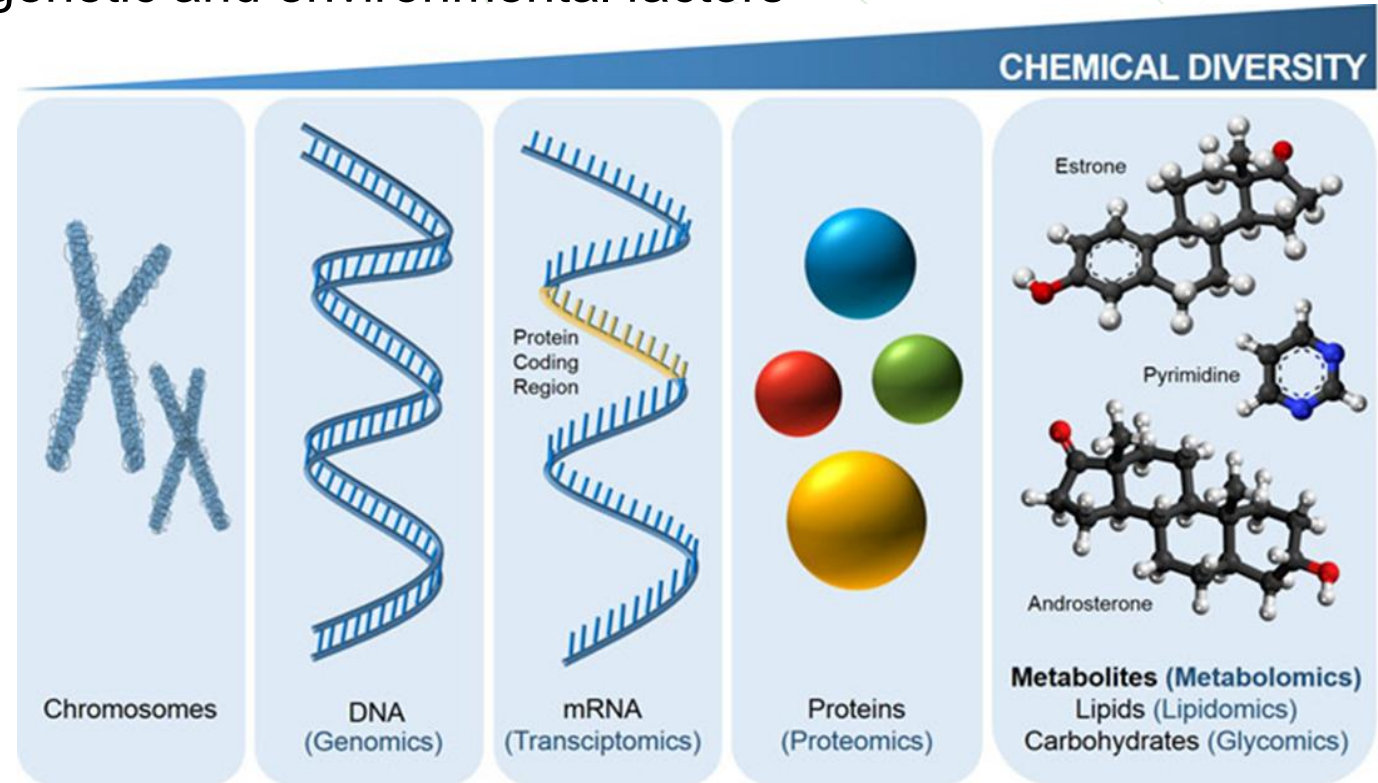


# Metabolomics

 **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

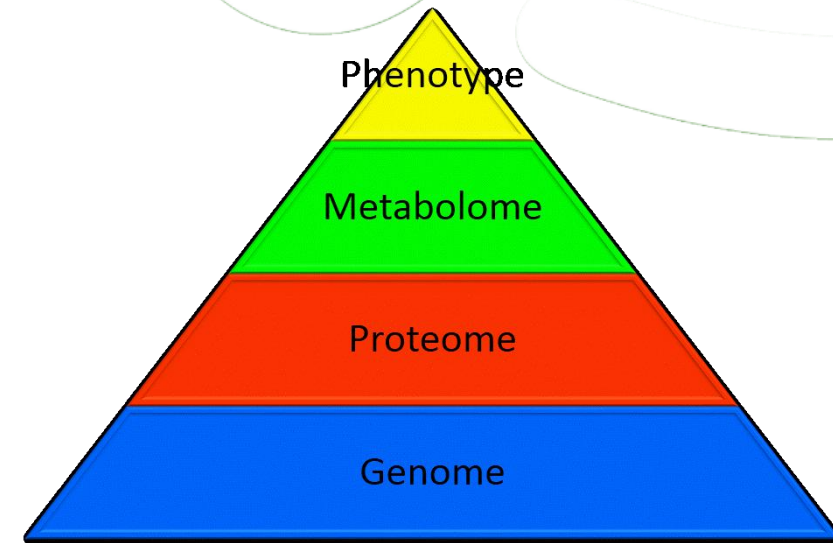
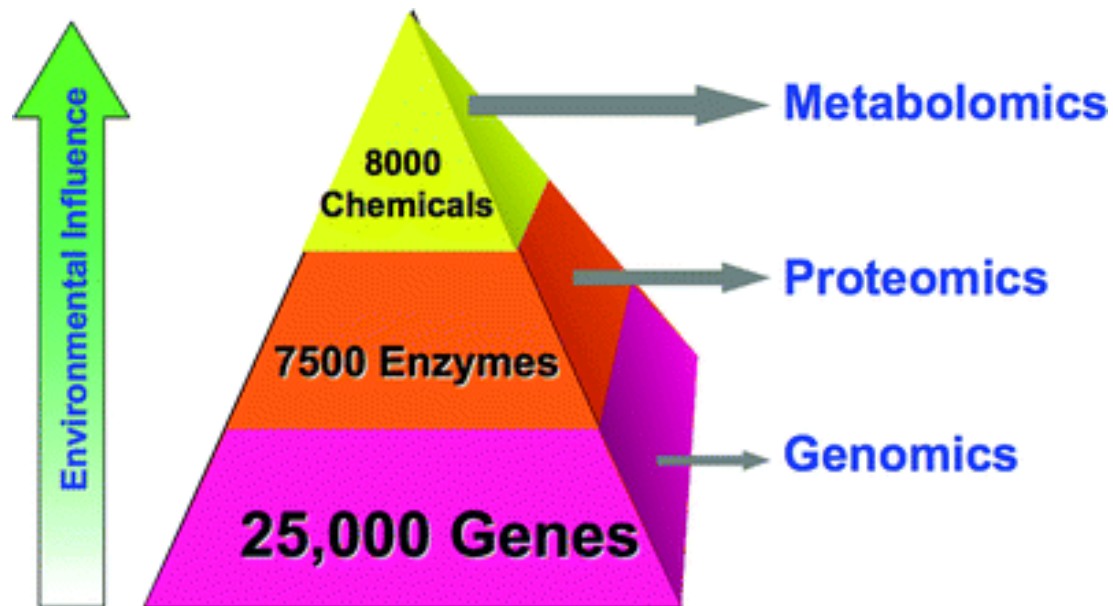
 The **metabolome** is the set of small chemical molecules present in biological samples (which are the end products of cellular processes).

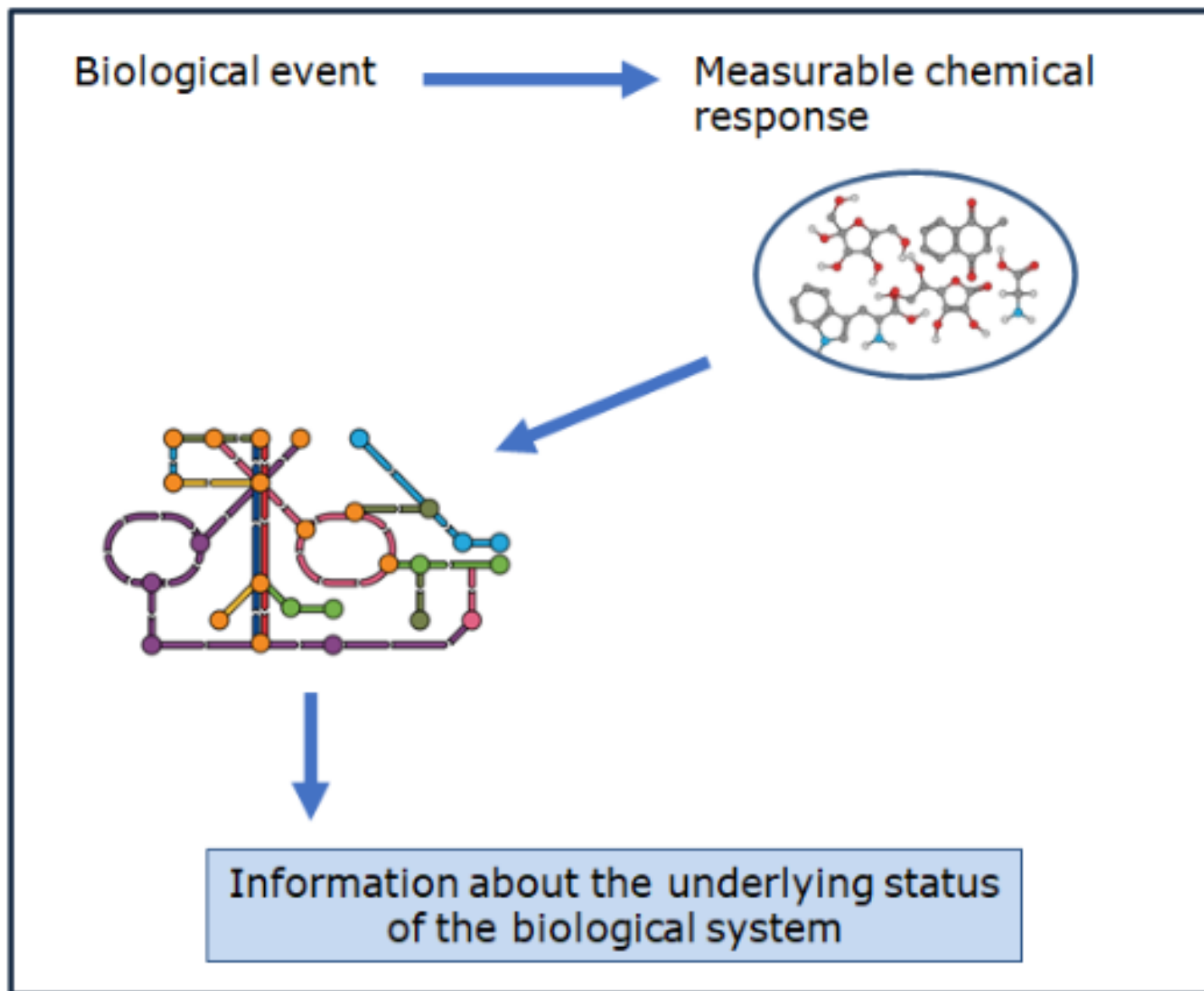


# Metabolomics and Bioinformatics

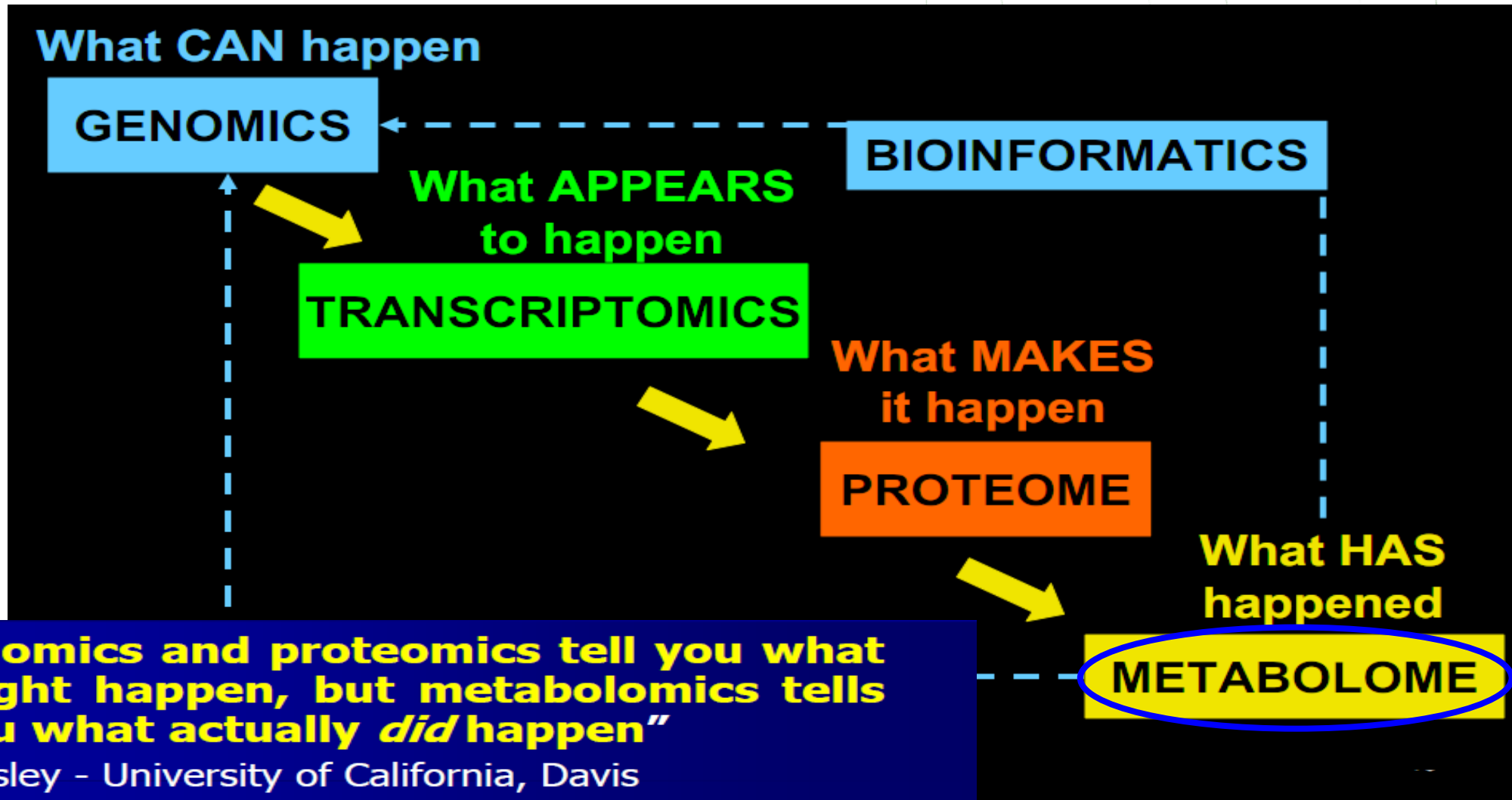
**Metabolomics is an extremely important area of bioinformatics research.**

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.





# 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.

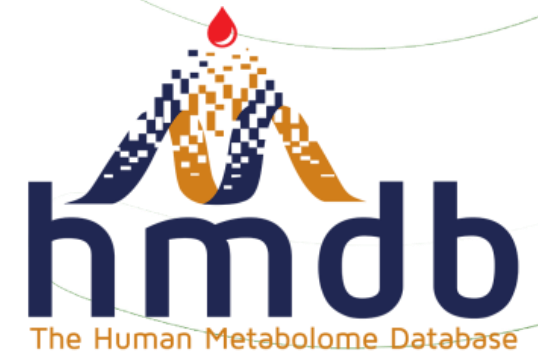


# Summary of metabolomic databases

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

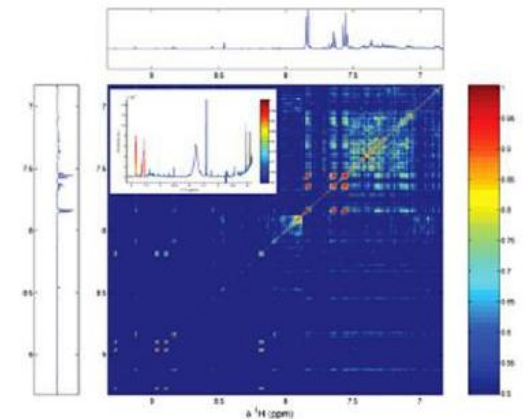
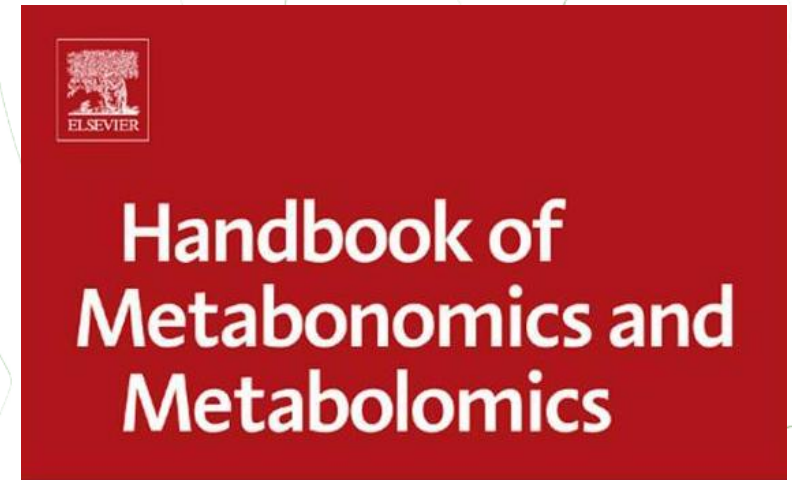


<https://www.hmdb.ca/>

# 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)

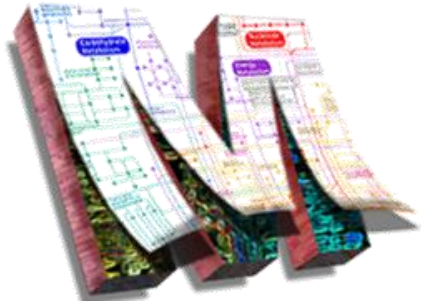
“**Metabonomics** is the quantitative measurement of the multiparametric, time-related metabolic response of living systems to pathophysiological stimuli and genetic changes.”(J.K. Nicholson, 1999)



John C. Lindon • Jeremy K. Nicholson • Elaine Holmes

Copyrighted material

# Advantages of metabolomics

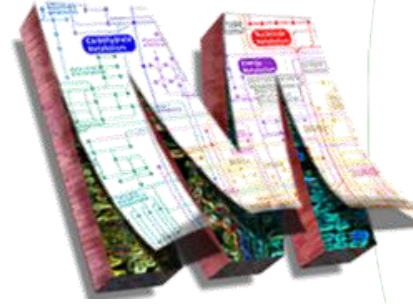


- Generate metabolic “signatures”
- Monitor/measure metabolite flux
- Monitor enzyme/pathway kinetics
- Assess/identify phenotypes
- Monitor gene/environment interactions
- Track effects from toxins/drugs/surgery
- Monitor consequences from gene KOs
- Identify functions of unknown genes



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



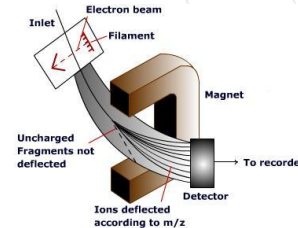
## Sampling

- Collection
- Storage
- Extraction



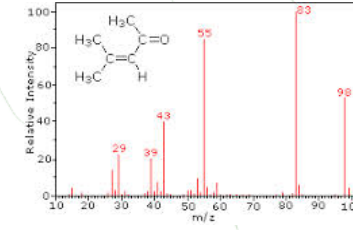
## Separation

- Chromatography
- Column
- Mobile phase



## Detection

- Choice of detector
- Data acquisition



## Data Processing

- Pathway elucidation
- Network modelling



- **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

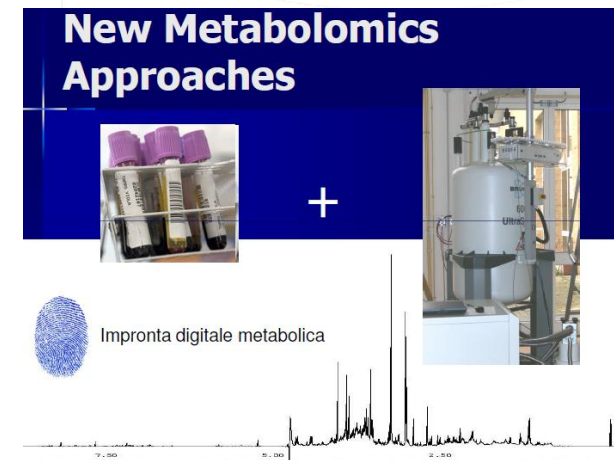
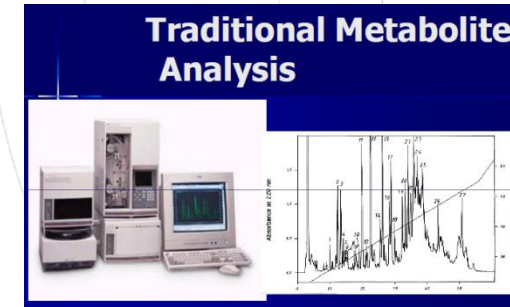
# Metabolomic approaches

## 1. Separation Techniques

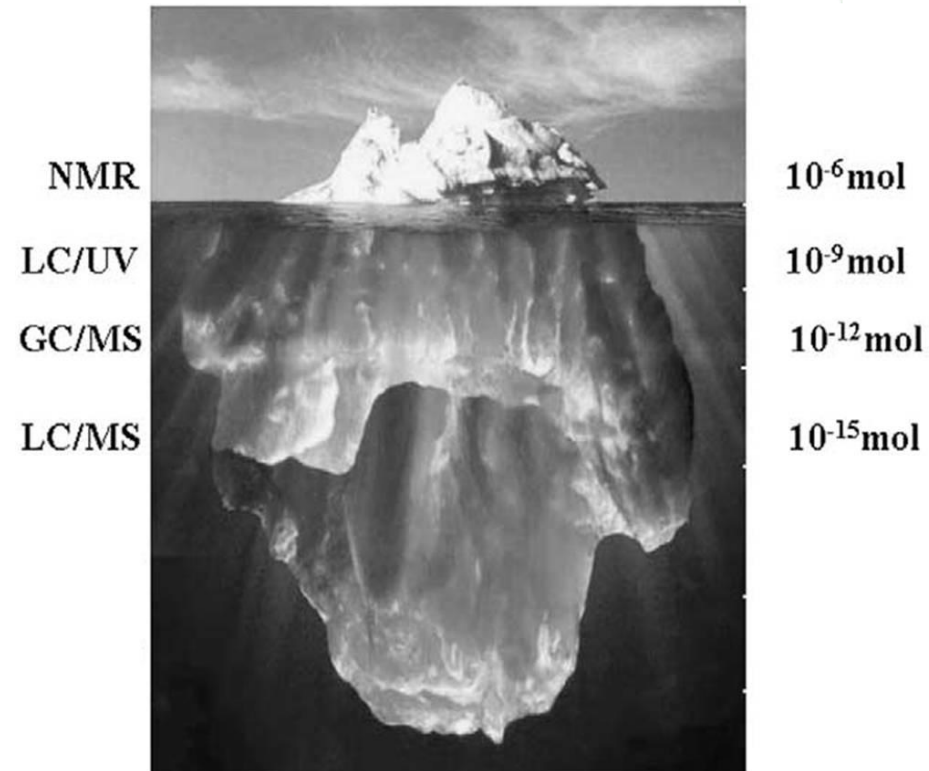
- Gas Chromatography (GC)
- Capillary Electrophoresis (CE)
- High Performance Liquid Chromatography (HPLC)
- Ultra Performance Liquid Chromatography (UPLC)

## 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.**

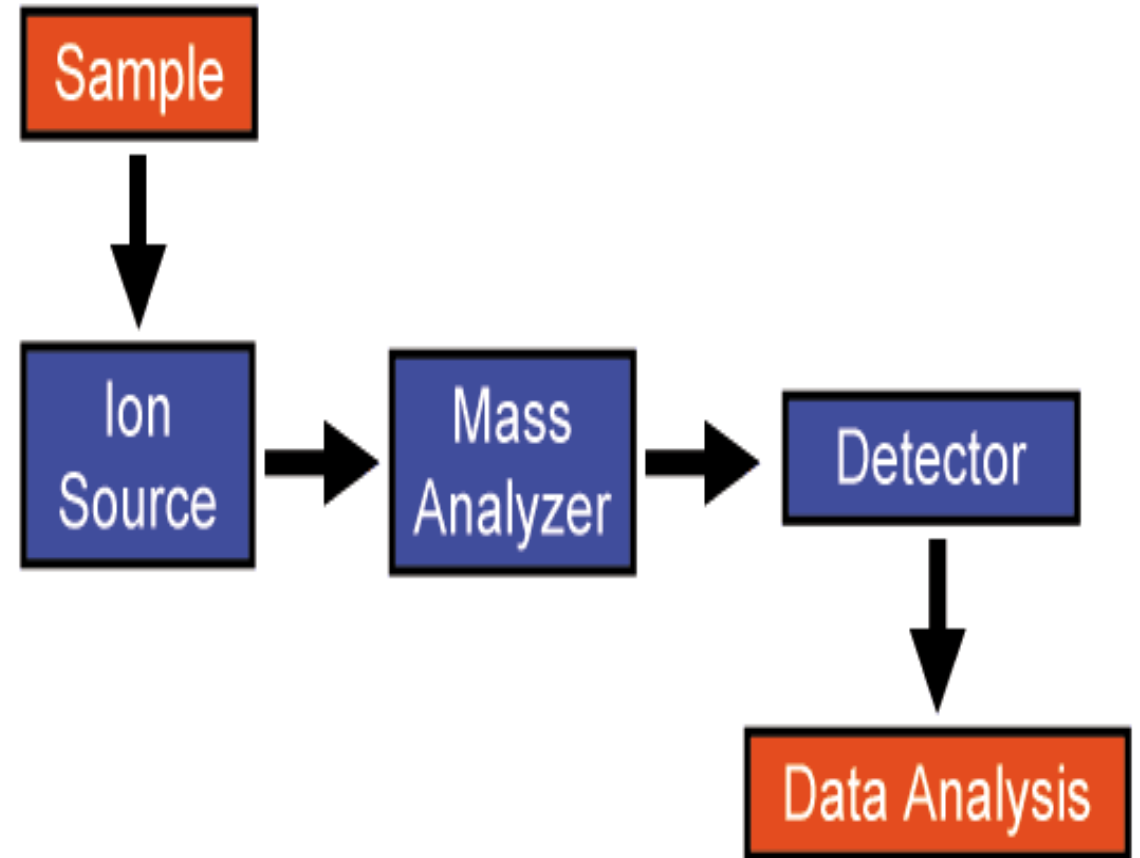
# Detection Technique - NMR

- 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 compounds less tractable to GC-MS or LC-MS (sugars, amines, volatile ketones, & relatively non-reactive compounds)



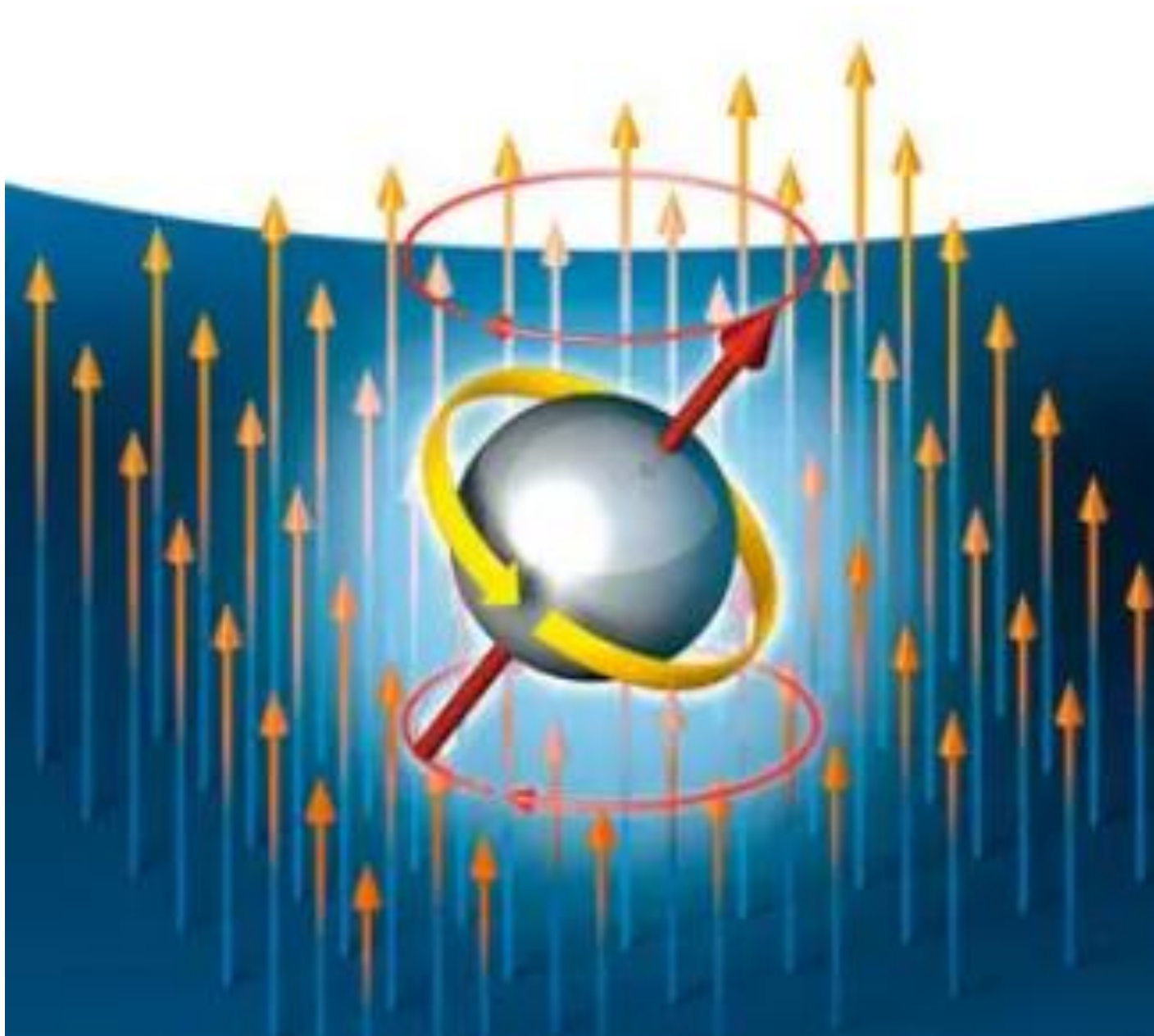
# Detection Technique - MS

- 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  $^1\text{H}$  (the most receptive isotope at natural abundance) and  $^{13}\text{C}$ , although nuclei from isotopes of many other elements (e.g.  $^{15}\text{N}$ ,  $^{14}\text{N}$ ,  $^{19}\text{F}$ ,  $^{31}\text{P}$ ,  $^{17}\text{O}$ ,  $^{29}\text{Si}$ ,  $^{10}\text{B}$ ,  $^{11}\text{B}$ ,  $^{23}\text{Na}$ ,  $^{35}\text{Cl}$ ,  $^{195}\text{Pt}$ ) can also be observed



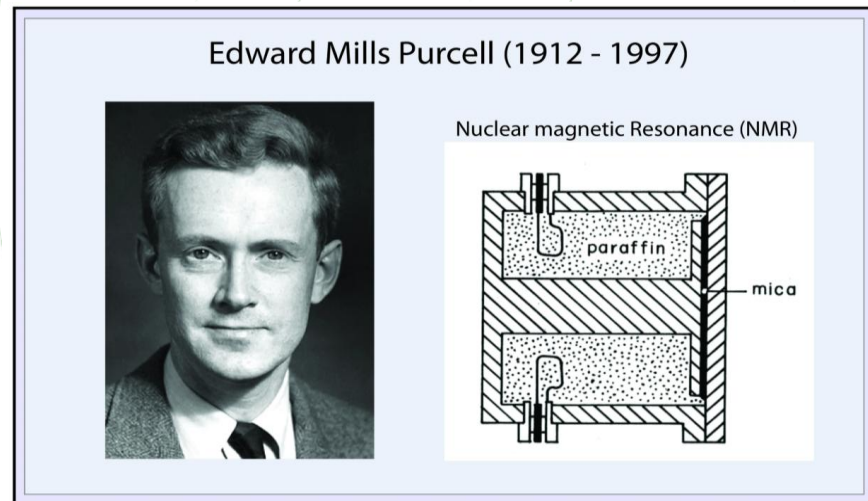
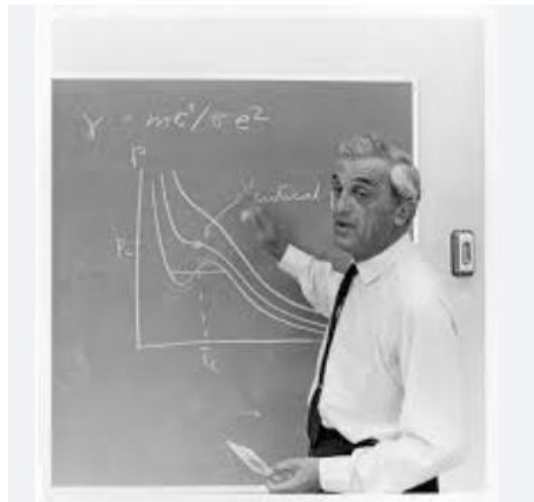
$$\omega_L = \gamma B_0$$



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

# NMR time-line

- 🌐 **1945** First observation of an NMR signal  
Bloch et al. Stanford Un. ( $^1\text{H}$  in  $\text{H}_2\text{O}$ ) Purcell et al. Harvard Un. ( $^1\text{H}$  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:

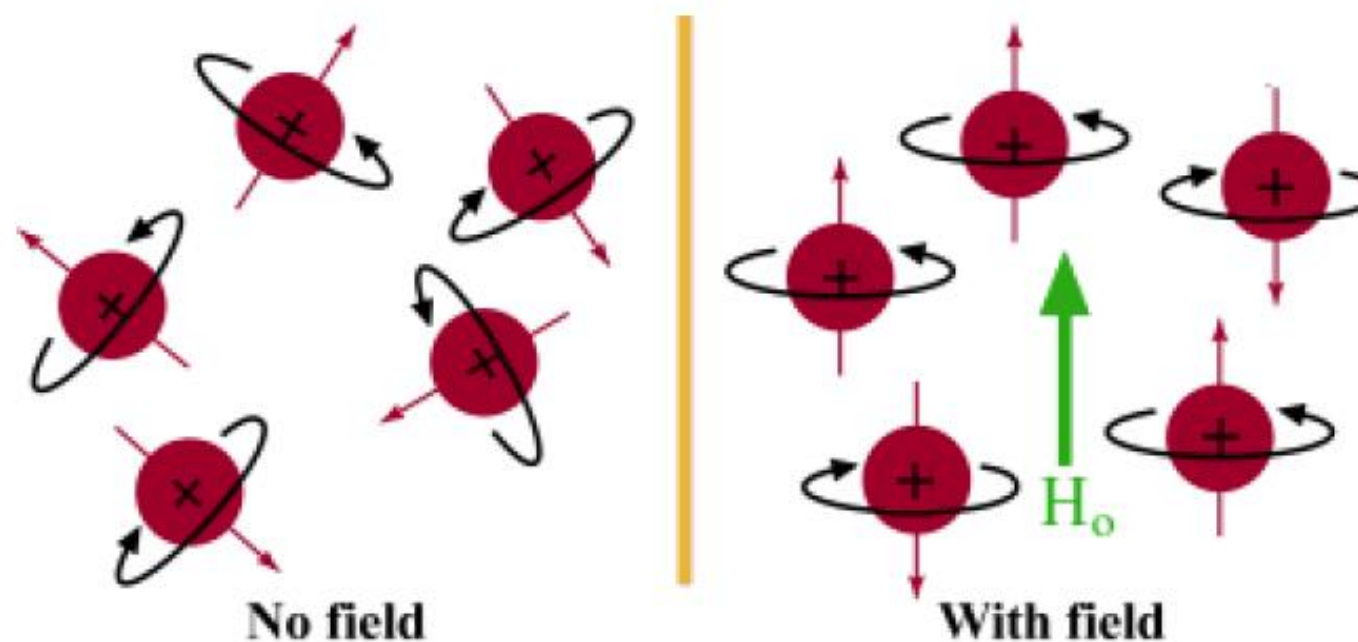
"for his groundbreaking contributions to the development of high-resolution nuclear magnetic resonance (NMR) spectroscopy."



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

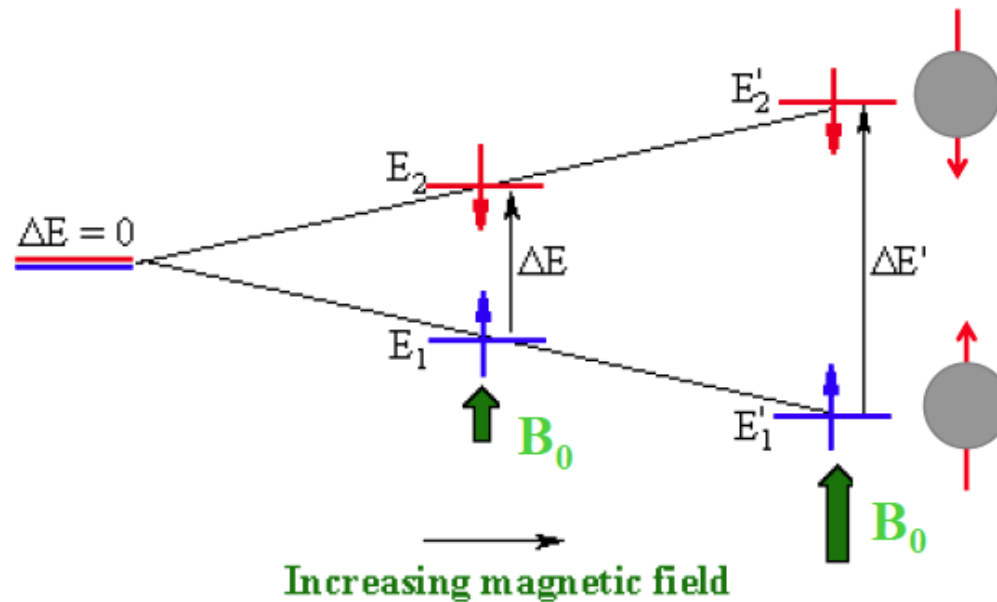
<b>Le regioni dello spettro elettromagnetico</b>				
<b>Regione dello spettro</b>	<b>Lunghezza d'onda (Å)</b>	<b>Lunghezza d'onda (cm)</b>	<b>Frequenza (Hz)</b>	<b>Energia (eV)</b>
Radio	$>10^9$	$>10$	$<3 \times 10^9$	$<10^{-5}$
Microonde	$10^9-10^6$	$10-0.01$	$3 \times 10^9 - 3 \times 10^{12}$	$10^{-5}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

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.



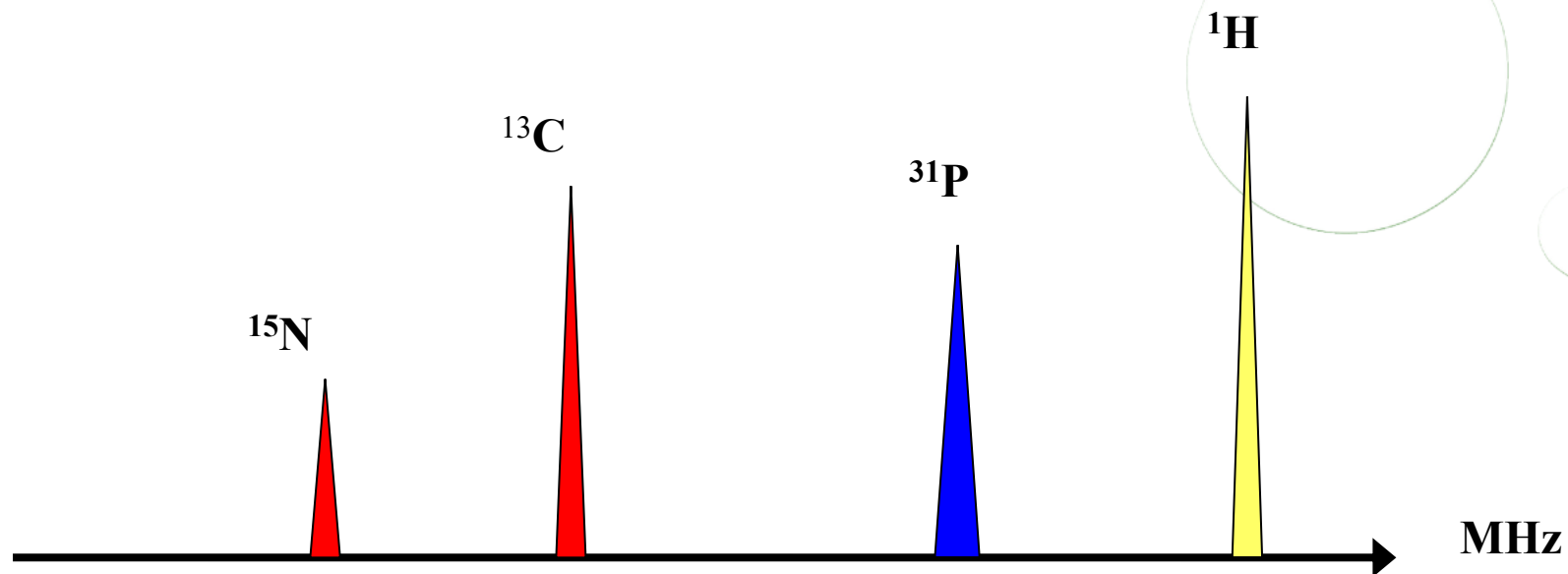
# Nuclear Magnetic Resonance

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:



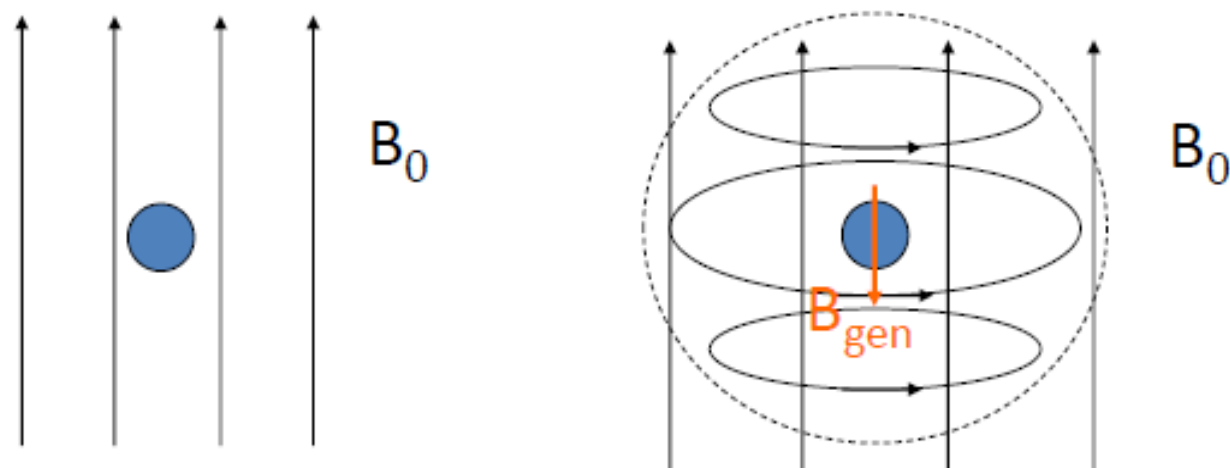
**DE (energy jump)** or the frequency measured for the excited nucleus is **proportional to the applied magnetic field, to the type of nucleus and also to the chemical environment of the nucleus itself**. If this were not (proportional to the chemical environment) all the protons would give the same 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  $B_0$ , the information would be practically zero.

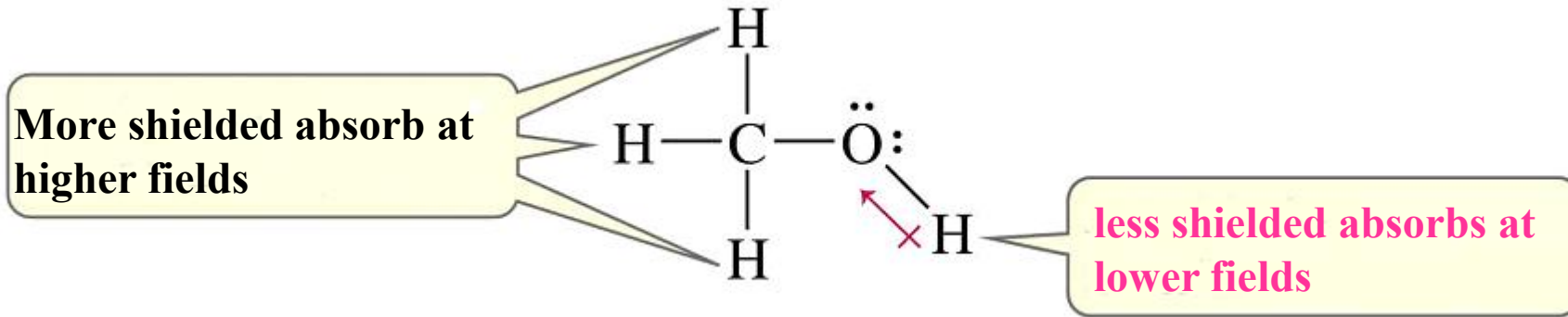


The electrons of an atom induce a small magnetic field that opposes the main one

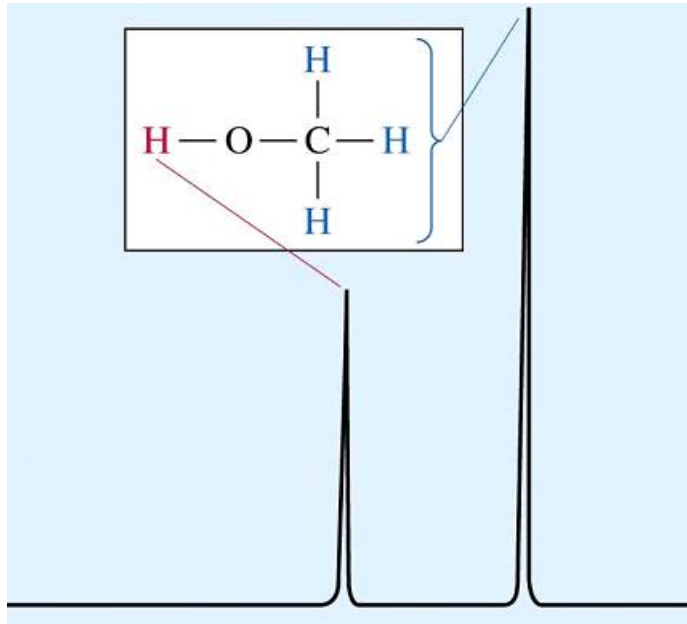
Nuclei of the same atom absorb slightly different frequencies due to their “chemical surroundings”

$$\delta B_0 = -\sigma B_0 \quad \sigma \text{ screen constant}$$

Nuclei interact with the local magnetic field.

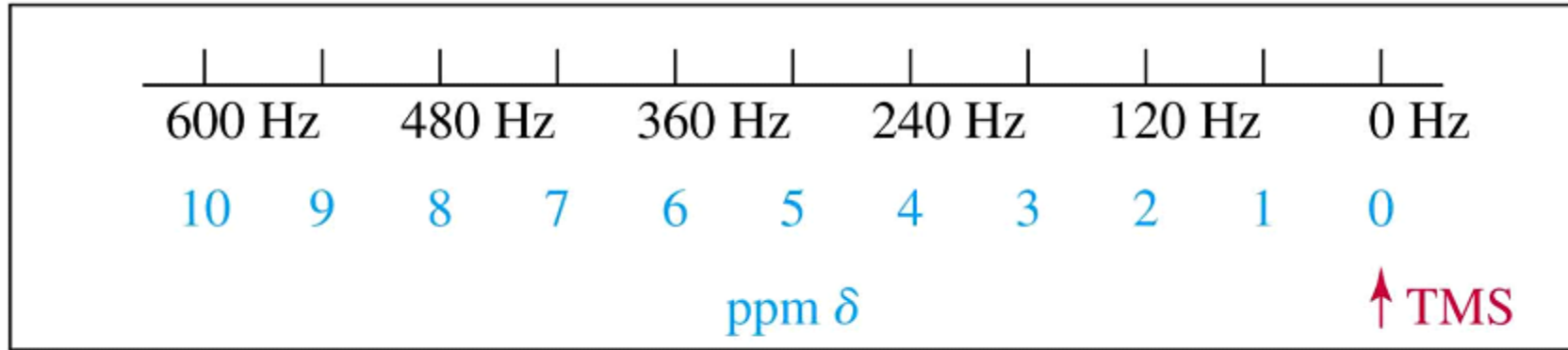


Depending on the chemical environment, the protons in a molecule are shielded in different amounts

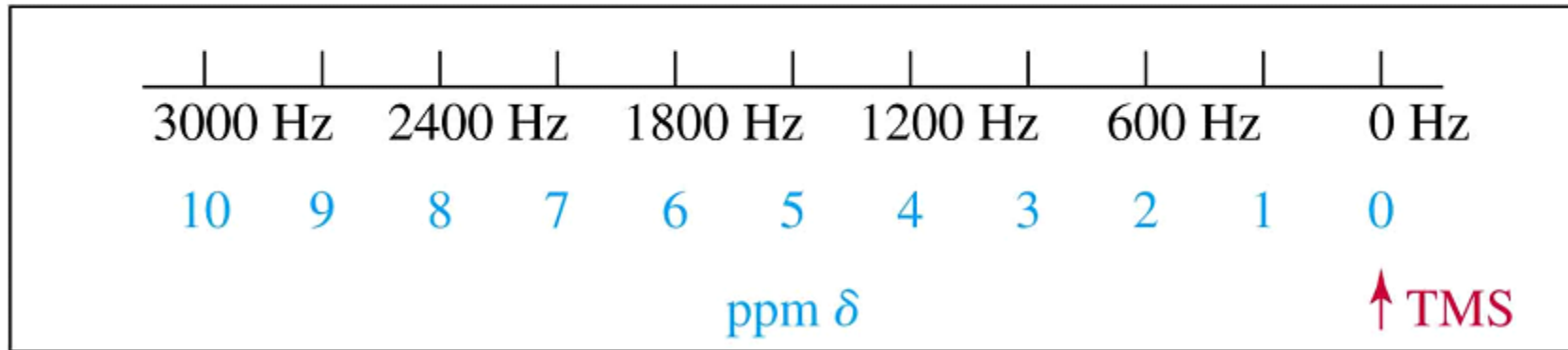


$B_0 \rightarrow$

## $\delta$ scale



or TSP

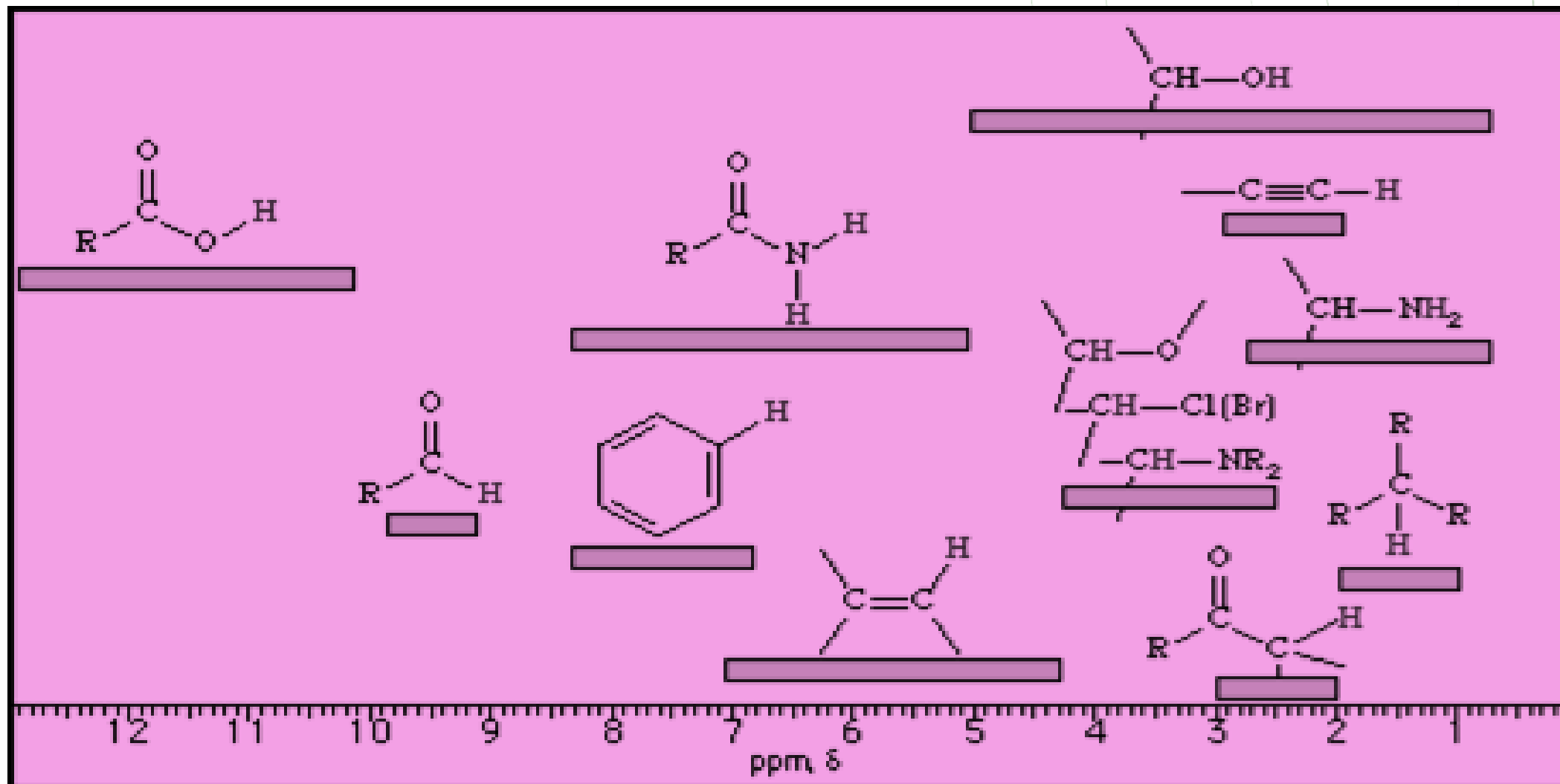


or TSP

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

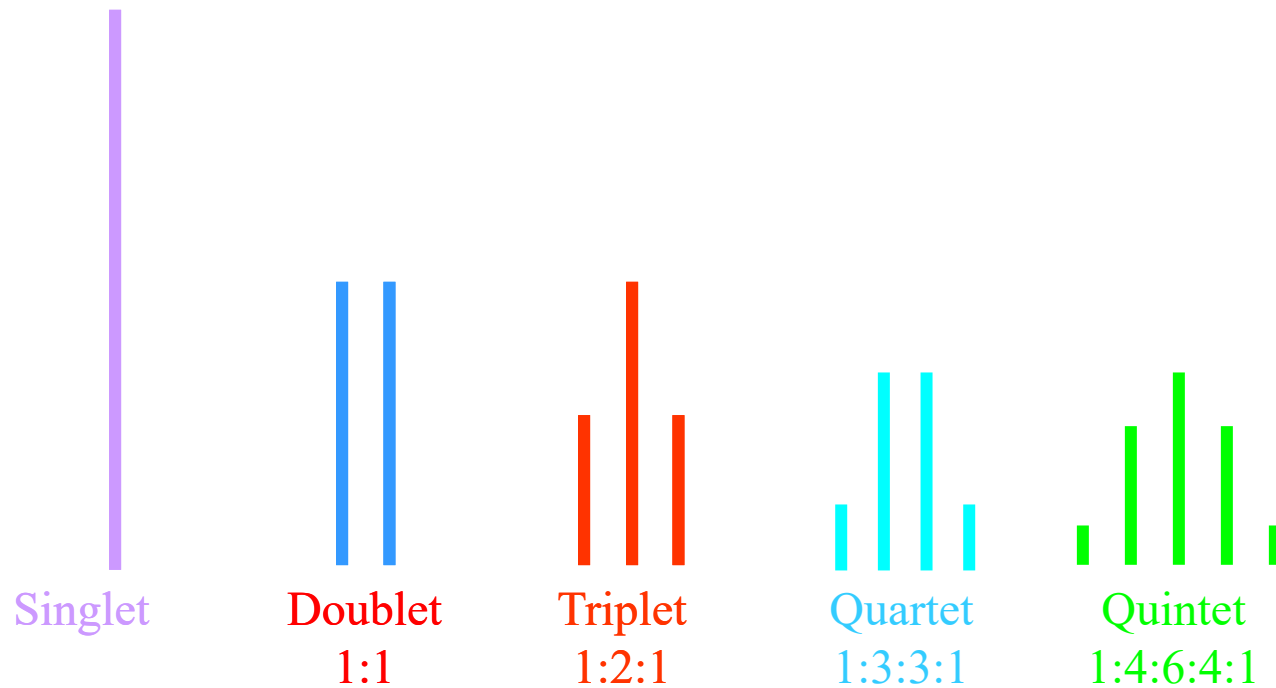
**The value of the chemical shift  $\delta$  does not depend on the power of the magnet (on the equipment used)**

# <sup>1</sup>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.

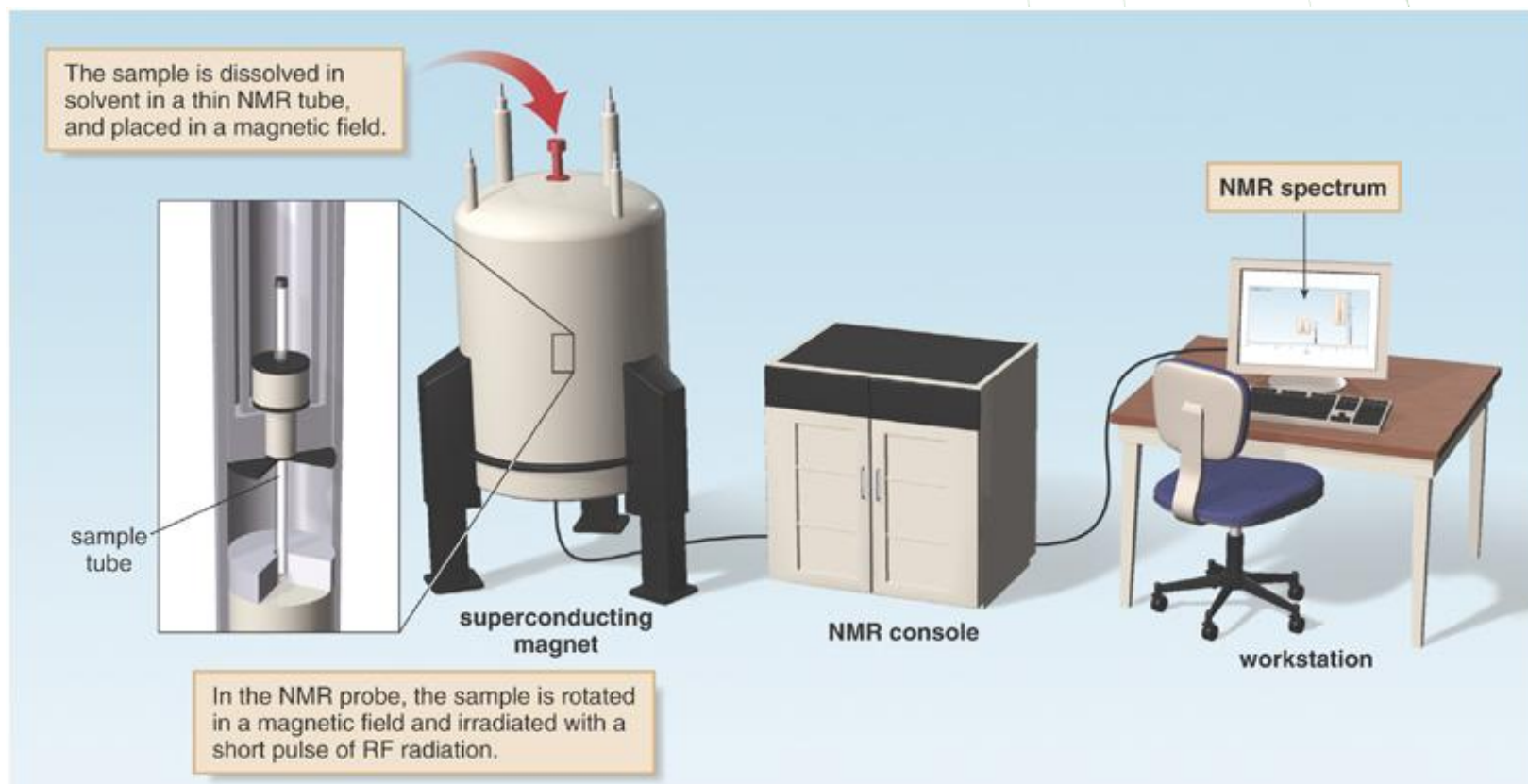


The multiplicity of signals is due to the chemical environment, therefore to the nearby chemical groups

# 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  $\text{CDCl}_3$  (deuteriochloroform), 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

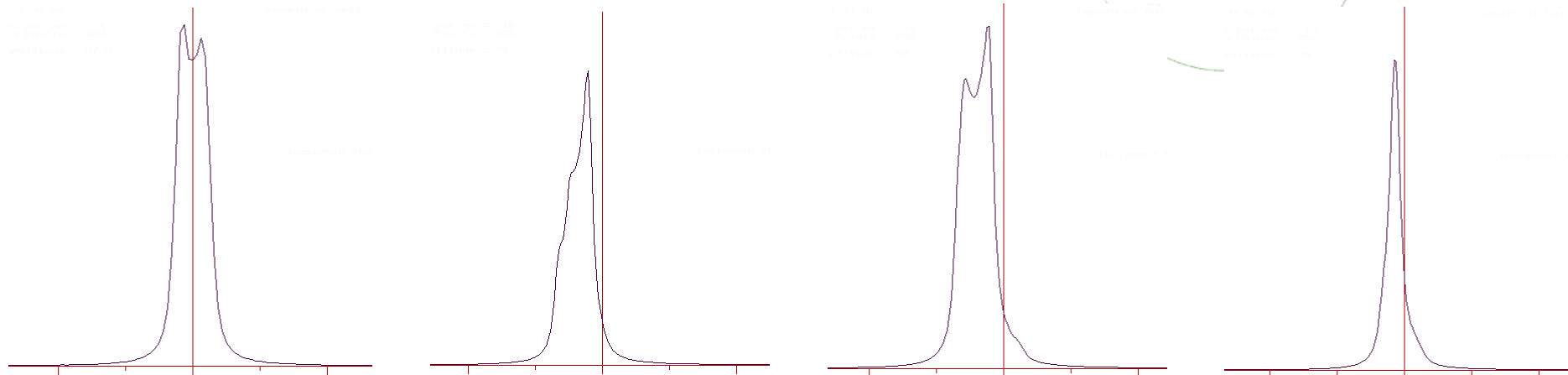
## Field-FrequencyLock

The NMR spectrometer has a channel permanently tuned to the resonance frequencies of deuterium ( $^2\text{H}$ )

# Nuclear Magnetic Resonance

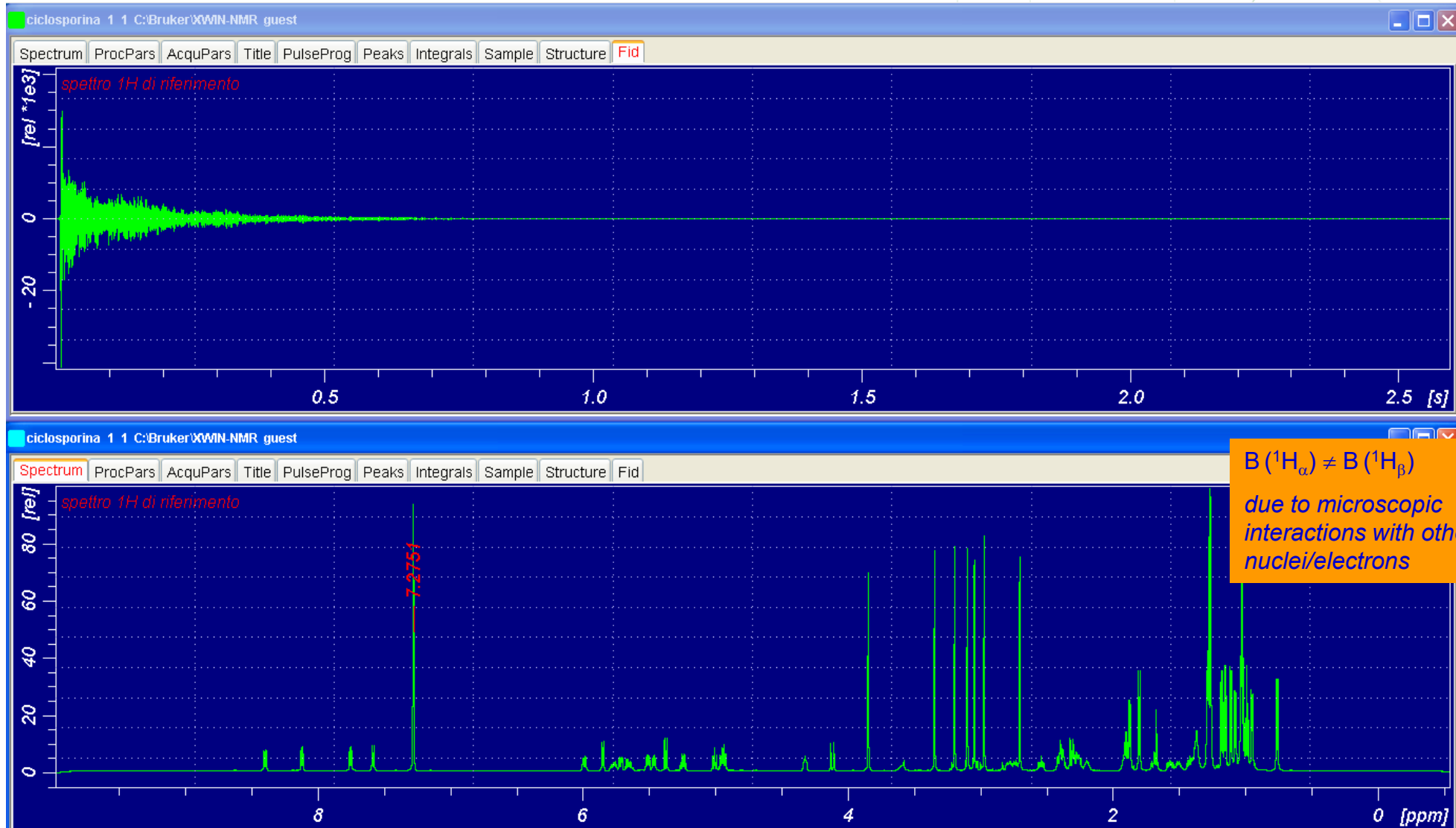
- After “locking” the lock signal, it is necessary to homogenize the  $B_0$  field (make it uniform along the entire sample)

## Shimming (field homogeneity)



Examples of the effect of poor homogeneity on a signal

# Nuclear Magnetic Resonance



## Nuclear Magnetic Resonance - analysis of complex mixtures

# Nuclear Magnetic Resonance

- Solvent: (about 500  $\mu$ l) deuterated solvents
- Internal reference: TSP or TMS (a capillary tip)



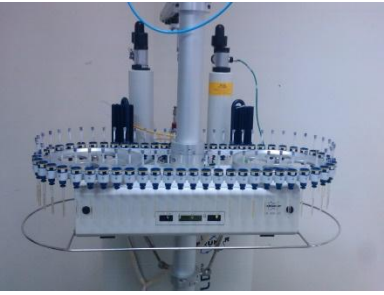
## Sample preparation for metabolomic

### – NMR tubes preparation and spectral acquisition

- (A) Always use the same **sample volume** and **sample pH** in NMR tube;
- (B) Collect all the 1D  $^1\text{H}$  NMR spectra at one **optimum temperature (typically 298 K)**;
- (C) Use the same **internal standard** (TSP or TMS) while recording all the  $^1\text{H}$  NMR spectra;
- (D) Use the same **acquisition parameters** for recording all the  $^1\text{H}$  NMR spectra

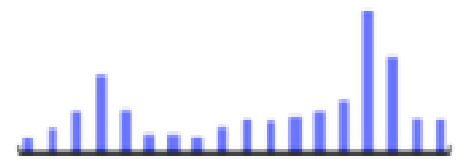
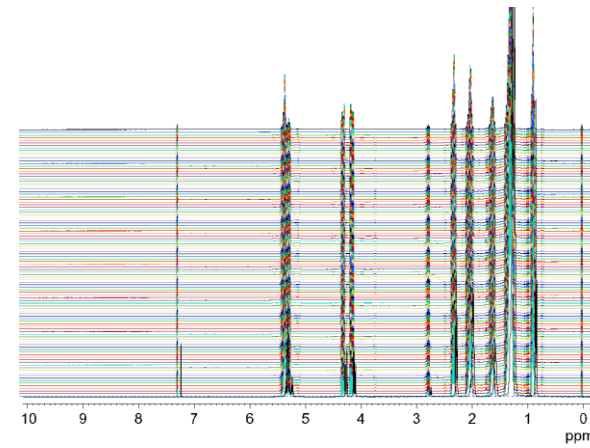
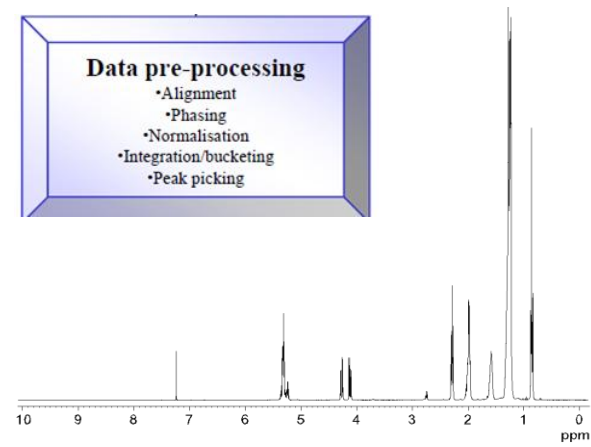
# NMR SPECTROSCOPY AND MULTIVARIATE ANALYSIS

*Multivariate analysis allows the observations of differences in samples*



**Data pre-processing**

- Alignment
- Phasing
- Normalisation
- Integration/bucketing
- Peak picking



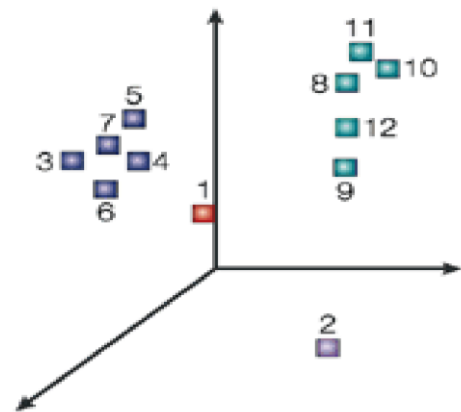
**Rectangular bucketing**

**Identification and quantification of metabolites**

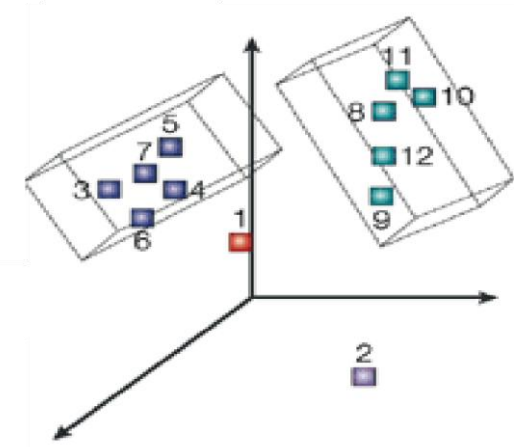
V1	V2	V3	Vn
1	1.3	0.4	
2	2.3	1.2	
3	2.7	2.1	
4	3.9	4.6	
.....	.....	.....	



**Many variables**



**Unsupervised mapping of data in three dimensional space**



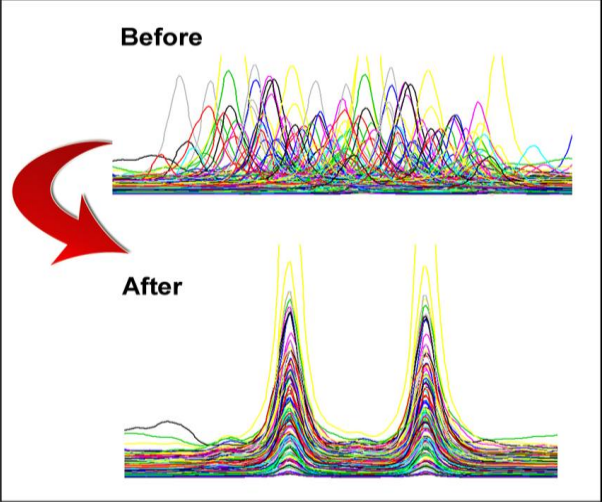
**Supervised classification and calculation of confidence limits**

**Data analysis**

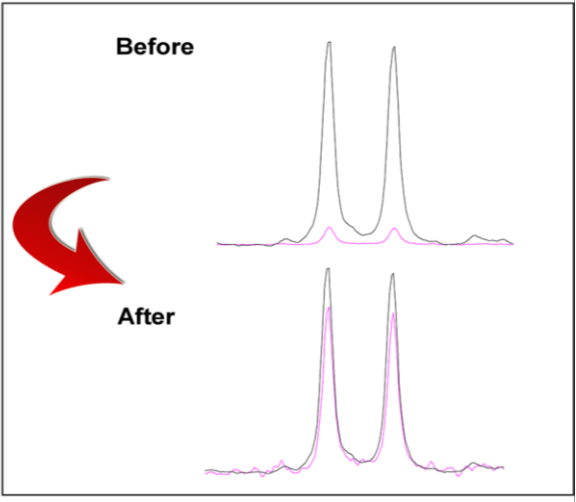
- PCA
- OPLS
- OPLS-DA
- O2PLS
- Hierarchical modelling

# Spectra processing

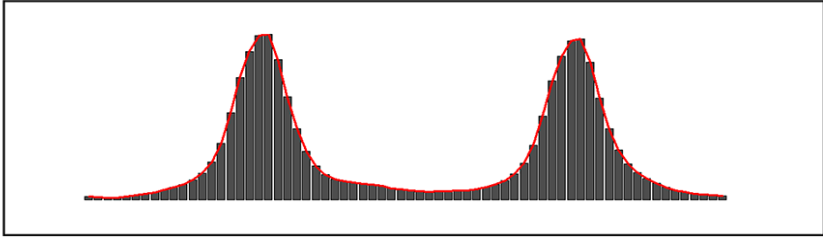
## Spectra alignment



## Spectra normalization



## Molecule quantification





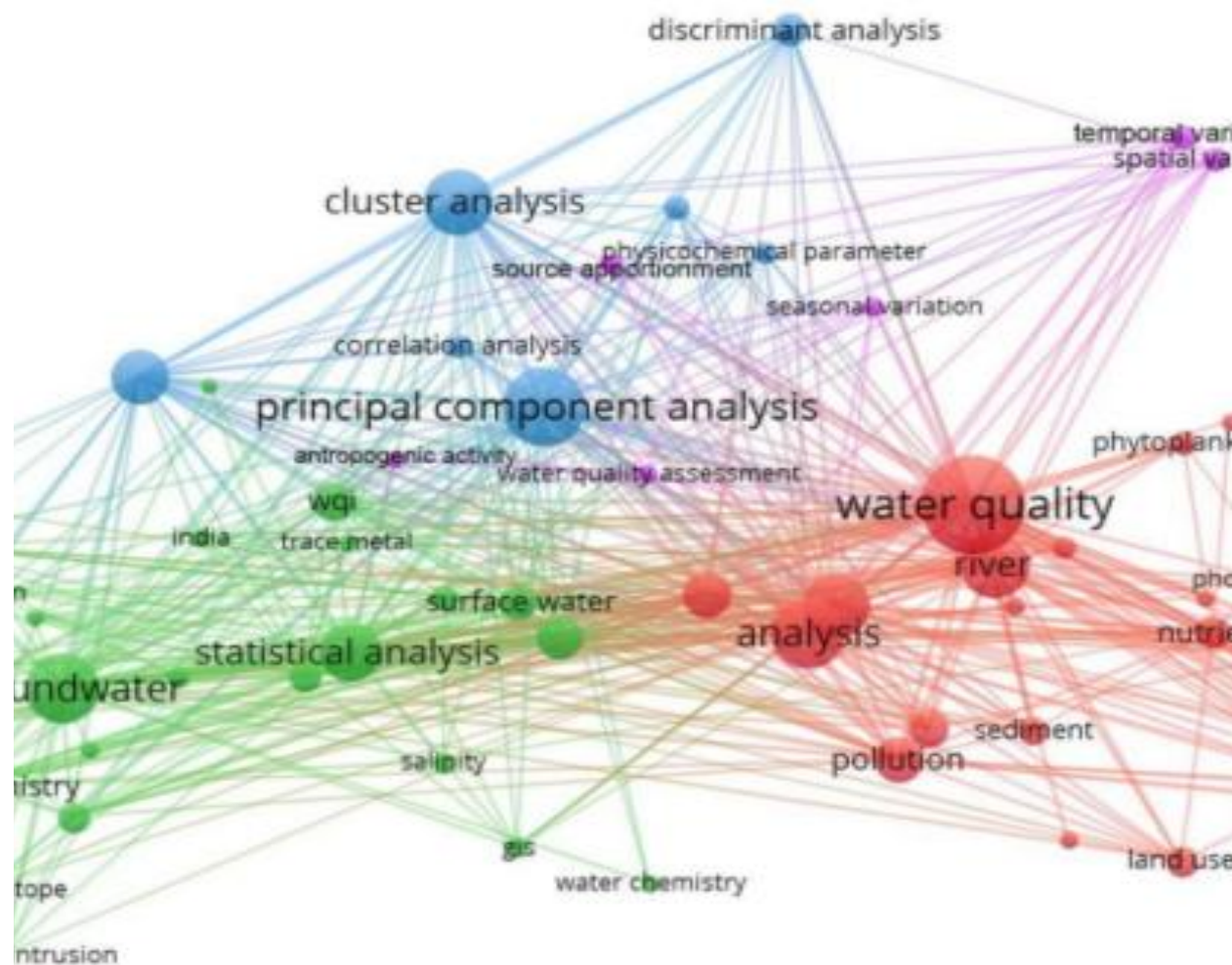
# *biological matrix*

# 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



About Downloads Documentation Community Success Stories News Events

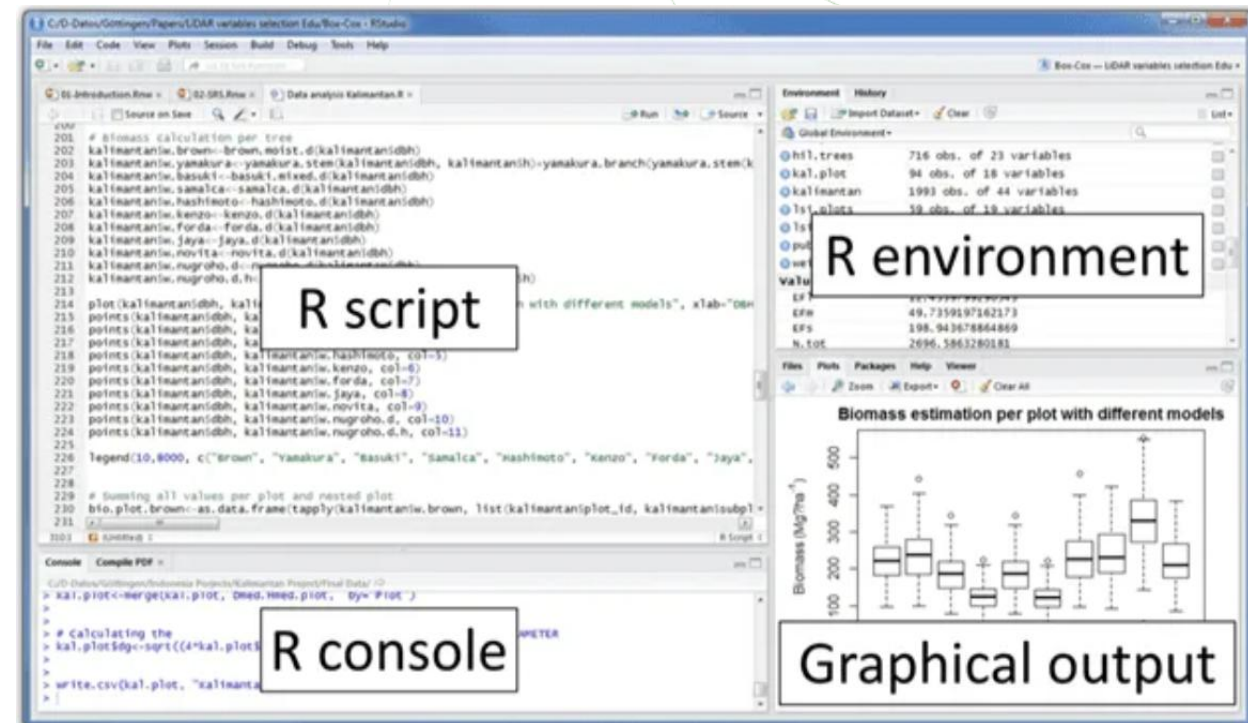
```
# Python 3: Fibonacci series up to n
>>> def fib(n):
>>>     a, b = 0, 1
>>>     while a < n:
>>>         print(a, end=' ')
>>>         a, b = b, a+b
>>>     print()
>>> fib(1000)
0 1 1 2 3 5 8 13 21 34 55 89 144 233 377 610 987
```

### Functions Defined

The core of extensible programming is defining functions. Python allows mandatory and optional arguments, keyword arguments, and even arbitrary argument lists. [More about defining functions in Python 3](#)

1 2 3 4 5

Python is a programming language that lets you work quickly and integrate systems more effectively. [>>> Learn More](#)



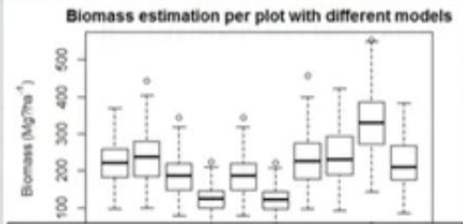
R script

```
## Introduction.Rnw > 62 SRS.Rnw > Data analysis Kalimantan.R >
# biomass calculation per tree
201 kalinantaniw.brown<-brown.mfist.d(kalinantaniwbh)
202 kalinantaniw.yanakura<-yanakura.sten(kalinantaniwbh, kalinantanih<-yanakura.branch<yanakura.sten<
203 kalinantaniw.basuki<-basuki.mfist.d(kalinantaniwbh)
204 kalinantaniw.samalca<-samalca.d(kalinantaniwbh)
205 kalinantaniw.hashimoto<-hashimoto.d(kalinantaniwbh)
206 kalinantaniw.kenzo<-kenzo.d(kalinantaniwbh)
207 kalinantaniw.forda<-forda.d(kalinantaniwbh)
208 kalinantaniw.jaya<-jaya.d(kalinantaniwbh)
209 kalinantaniw.novita<-novita.d(kalinantaniwbh)
210 kalinantaniw.nugroho<-nugroho.d(kalinantaniwbh)
211 kalinantaniw.nugroho.d<-nugroho.d(kalinantaniwbh)
212 kalinantaniw.nugroho.d.h<-nugroho.d(kalinantaniwbh)
213
214 plot(kalinantaniwbh, kalinantaniw.brown, col=1)
215 points(kalinantaniwbh, kalinantaniw.yanakura, col=2)
216 points(kalinantaniwbh, kalinantaniw.basuki, col=3)
217 points(kalinantaniwbh, kalinantaniw.samalca, col=4)
218 points(kalinantaniwbh, kalinantaniw.hashimoto, col=5)
219 points(kalinantaniwbh, kalinantaniw.kenzo, col=6)
220 points(kalinantaniwbh, kalinantaniw.forda, col=7)
221 points(kalinantaniwbh, kalinantaniw.jaya, col=8)
222 points(kalinantaniwbh, kalinantaniw.novita, col=9)
223 points(kalinantaniwbh, kalinantaniw.nugroho.d, col=10)
224 points(kalinantaniwbh, kalinantaniw.nugroho.d.h, col=11)
225
226 legend(10,8000, c("brown", "yanakura", "basuki", "samalca", "hashimoto", "kenzo", "forda", "jaya",
227
228
229 # Summing all values per plot and nested plot
230 bio.plot.brown<-as.data.frame(tapply(kalinantaniw.brown, list(kalinantaniplot_id, kalinantaniwsubp<
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
```

R environment

Environment	History
Global Environment	
h1t.trees	716 obs. of 23 variables
kal.plot	94 obs. of 16 variables
kalinantan	1993 obs. of 44 variables
ts.plots	19 obs. of 19 variables
ts	
plot	
we	
valu	
ES	
EPS	49.7359197162173
EPS	198.943678864869
N.tot	2696.5863280181

Biomass estimation per plot with different models

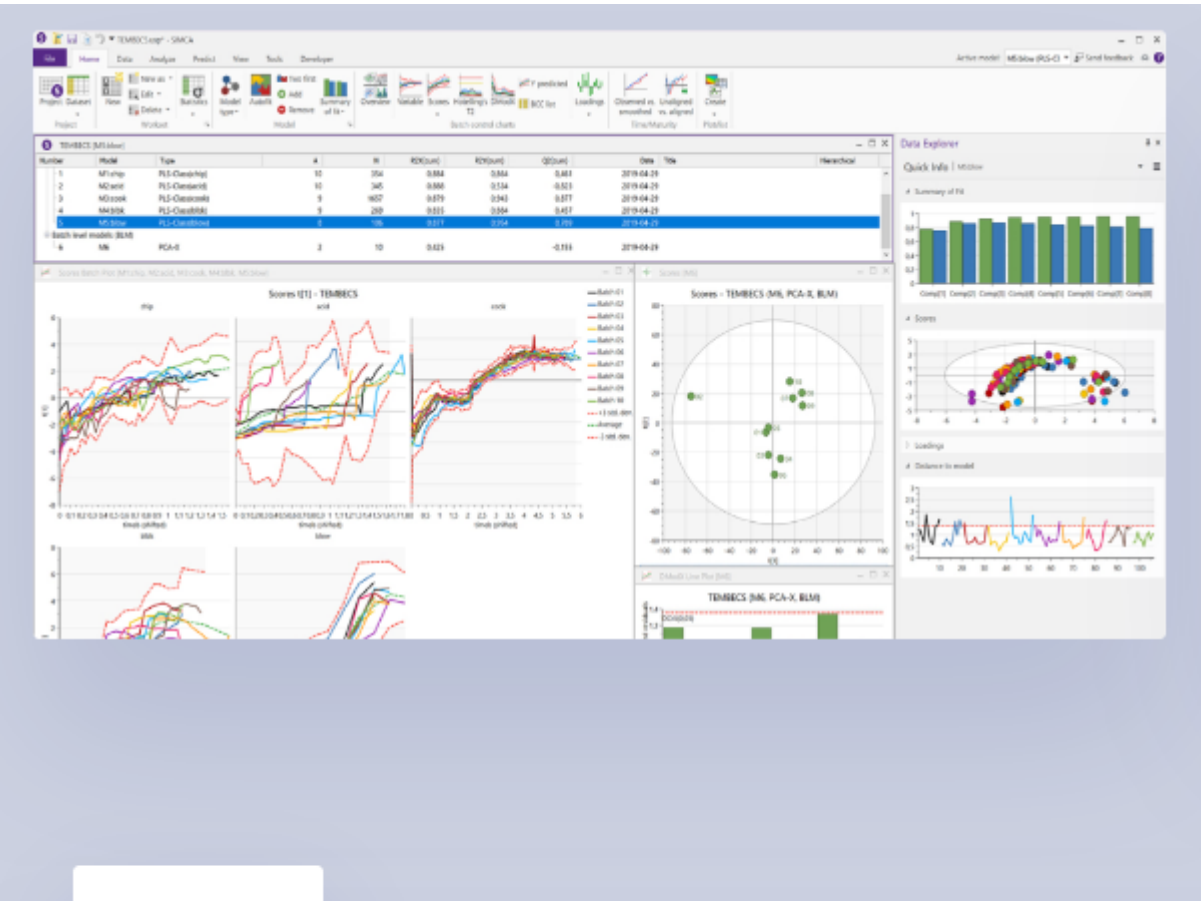


R console

```
> k1.plot<-merge(kal.plot, Dmed, Dmed.plot, by="Plot")
> # calculating the
> kal.plot$dy<-sqrt((4*k1.plot
> write.csv(kal.plot, "kalinanta
```

Graphical output

# Platforms for analyzing metabolomic data



## 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 <a href="#">utilities</a> )				
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 (metabolite list)			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

## Scaling

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

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

# 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.

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 ?

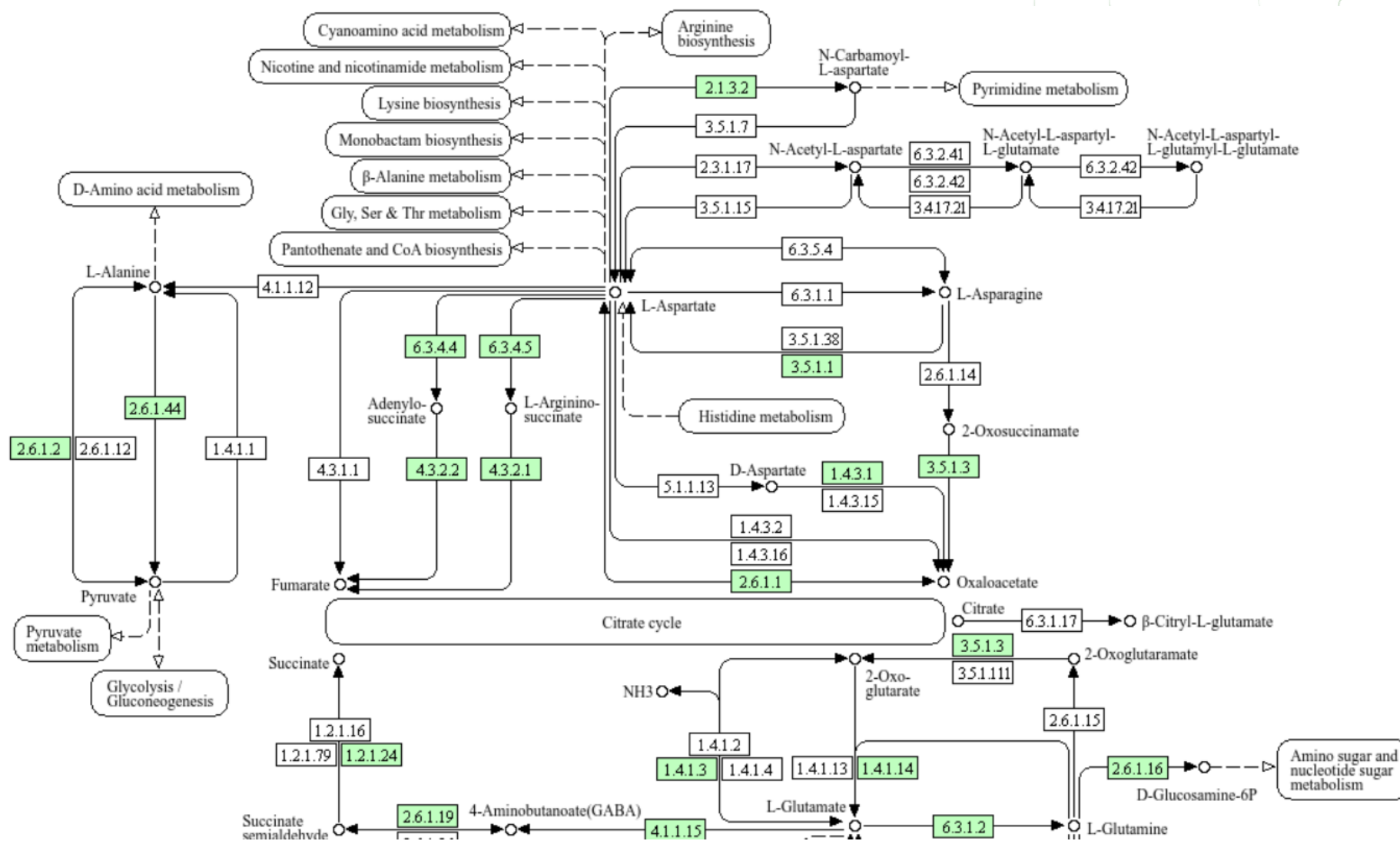
- A "dummy" variable is added for each category. PLS is then used to relate X and Y.

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

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"

