PSL RESEARCH UNIVERSITY PARIS

ESPCI PARIS

EDUCATION SCIENCE INNOVATION

cnrs dépasser les frontières

LC-MS AND BOTTOM UP PROTEOMICS

JOELLE VINH

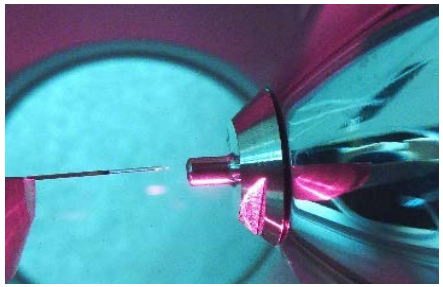
JOELLE.VINH@ESPCI.FR

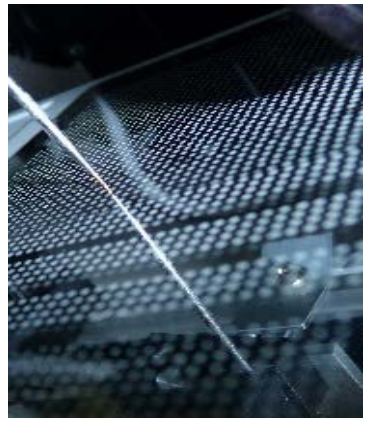
SMBP ESPCI, CNRS USR 3149, PARIS

SPECTROMÉTRIE DE MASSE BIOLOGIQUE ET PROTÉOMIQUE

Summary

- Omics and Mass spectrometry
- Why study the proteome?
- Bottom-up proteomics: an LC-MS story
- DIA and HRM
- Quantification





The interface between Chemistry / Biology is essential

Two of the three most cited papers of all time report analytical chemistry techniques to study biological systems

R. Van Noorden, B. Maher and R. Nuzzo, Nature, 2014, 514, 550–553

1. **Protein measurement with the folin phenol reagent**. Lowry, O. H., Rosebrough, N. J., Farr, A. L. & Randall, R. J. J. Biol. Chem. 193, 265–275 (1951). (305 148 citations)

2. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Laemmli, U. K. Nature 227, 680–685 (1970). (213 005 citations)

3. **A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding**. Bradford, M. M. Anal. Biochem. 72, 248–254 (1976). (155 530 citations)

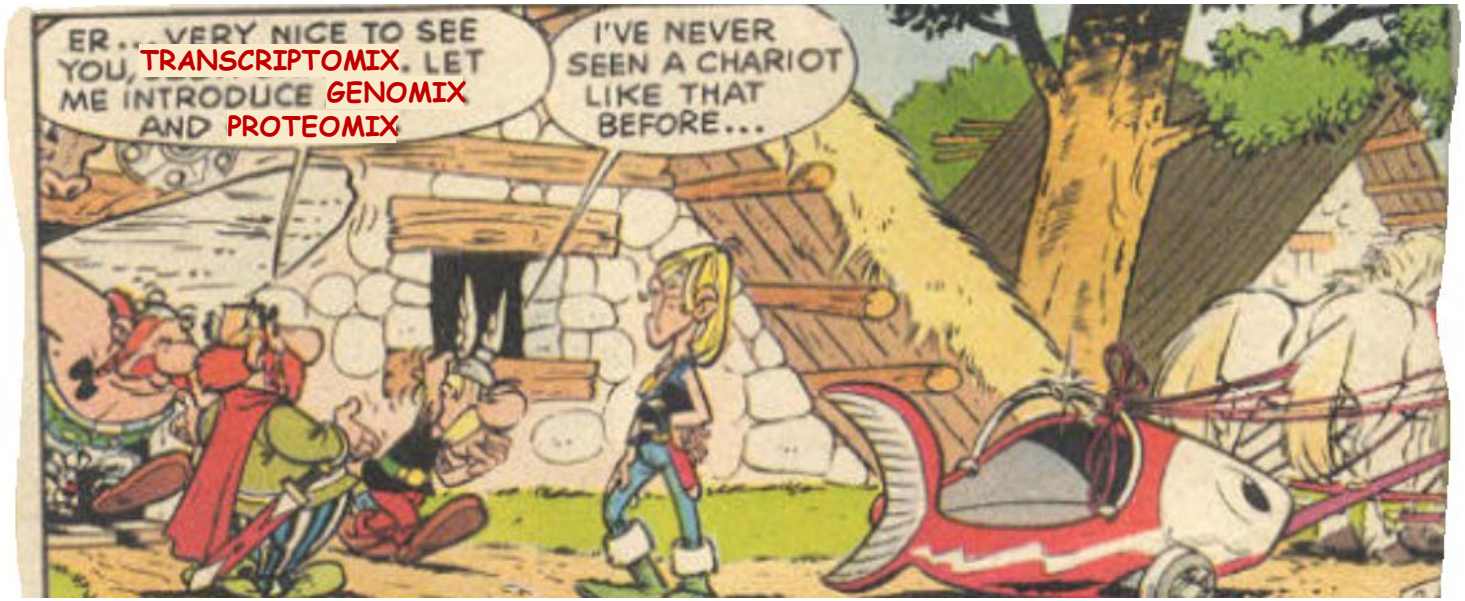
To understand a biological process we often need a better comprehension of the associated chemical environment

Omics or Omix

'Omic' sciences are perhaps the best example of a successful integration of chemistry and biology.

D.J. Hare and E.J. New, *Chem. Commun.*, 2016, 52, 8918

The 'omic' revolution has been driven by advances in analytical chemistry, from DNA microarray technology to mass spectrometry.

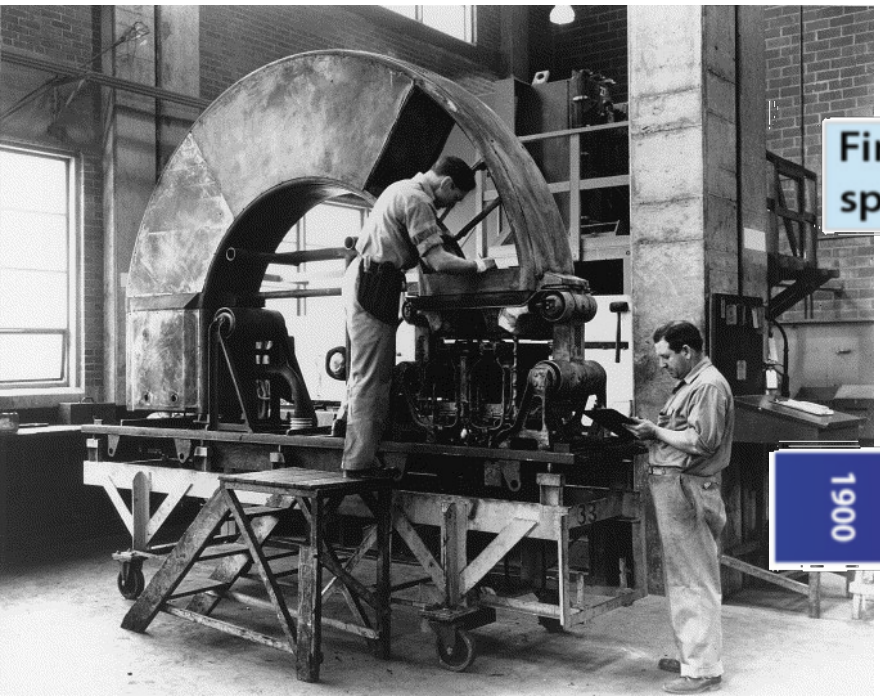


2 divergent lines of enquiry with regard to 'omic' sciences and systems biology :

(1) how can analytical chemistry be improved to better answer key biological questions?

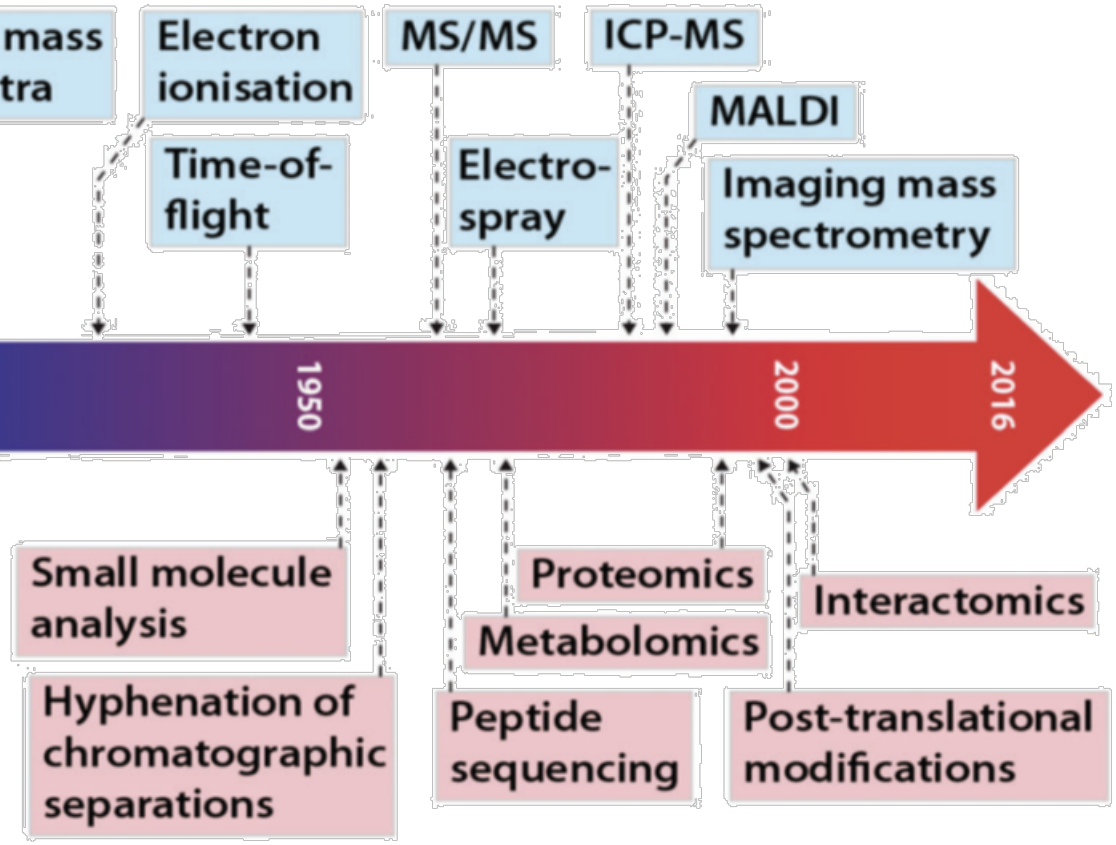
(2) are the right questions being asked that take advantage of the tools at our disposal?

The example of Mass Spec



Part of the Calutron mass spectrometer first used for preparative MS

Yergey, A.L. & Yergey, A.K. Preparative scale mass spectrometry: A brief history of the calutron J Am Soc Mass Spectrom (1997) 8: 943.



A. Doerr, J. Finkelstein, I. Jarchum, C. Goodman and B. Dekker, Nature Milestones: Mass Spectrometry, Nature Publishing Group, 2015.

1. How much ?
2. How deep?
3. How fast?
4. What flexibility?

Summary

- Omics and Mass spectrometry
- **Why study the proteome?**
- Bottom-up proteomics: an LC-MS story
- DIA and HRM
- Quantification

Proteomics

Proteome : The pool of PROteins isoforms expressed by one genOME of a cell or a tissue at one given time in one given environment.

Wilkins et al., Biotechnol Gene Eng Rev, 1995

Objective 1 : Exhaustivity ????

Proteomic Analysis : Dynamic and quantitative analysis of the regulation of expression of **genes products** for a given biological process in order to decipher molecular interaction mechanisms.

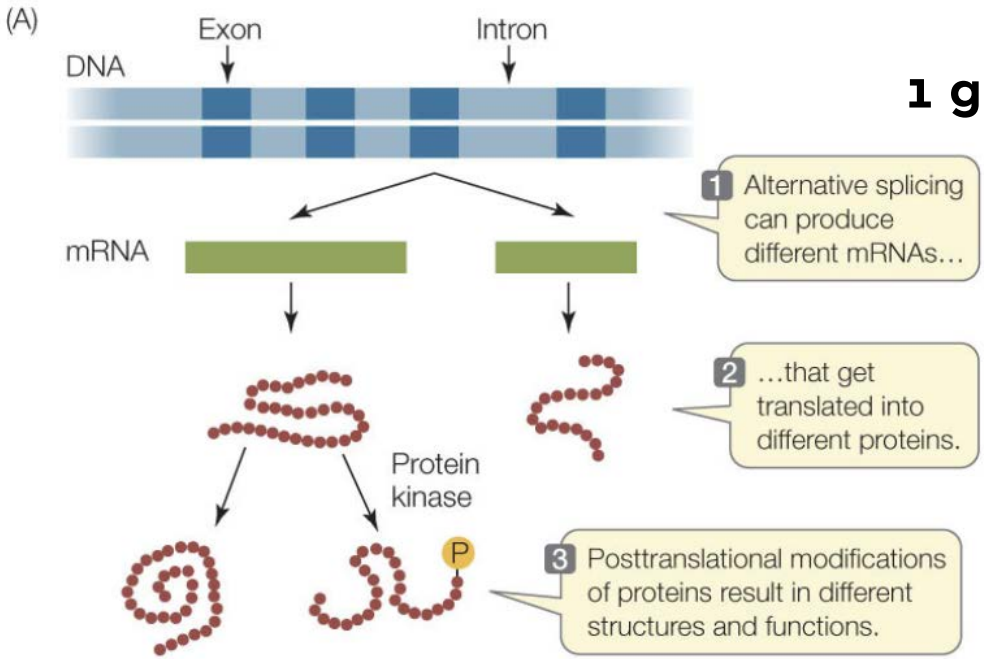
Anderson et Anderson, Electrophoresis, 1998

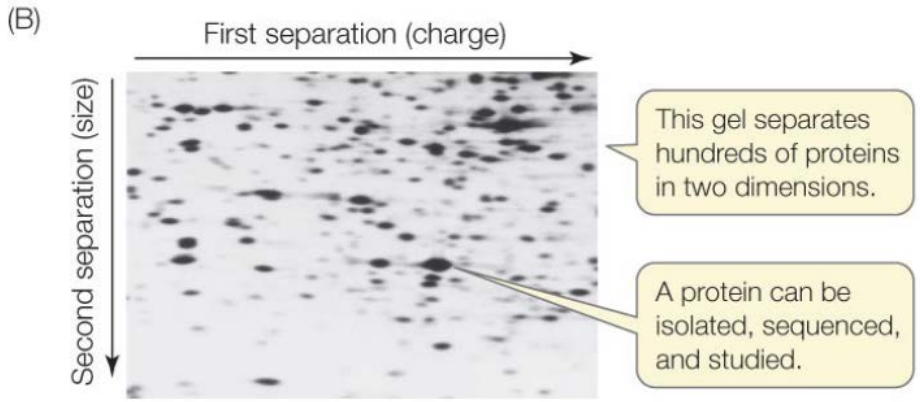
Objective 2 : Quantitative Analysis !!!!

From DNA to proteins

~~1 gene = 1 protein ??~~

1 gene = several proteins

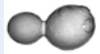




12.4: From P. H. O'Farrell, 1975. High resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem.* 250: 4007. Courtesy of Patrick H. O'Farrell.

phosphorylation, acetylation,oxidation, glycosylation, lipoylation

No apparent correlation between the size of the genome and the complexity of the organism

Organism	# Chromosom	Genome size (pb)
Human	**2x23** 	**3.2 E9**
Mus musculus (mouse)	2x20 	2.7 E9
Drosophila melanogaster (Fly)	2x4 	1.4 E8
Tobacco	2x24 	4.8 E9
Zea mays (corn)	**2x10** 	**3.9 E9**
Saccharomyces cerevisiae (yeast)	17 	1.5 E7
E Coli (bacteria)	1 	4.64 E5
Phage λ	1 	4.85 E4
Allium cepa (onion)	2x8 	1.5 E10
Amoeba dubia (amibae)	**2x6** 	**6.7 E11**
Lysandra atlantica (blue lycaenid butterfly)	250 	N/A
Ophioglossum reticulatum	2x630 	1.6 E11

Protein–coding Human genes

2001: 30000

2004: 25000

2012: 21065

2013: 20774

2014: 19000

Total Number of Protein-Coding Genes

Drosophila melanogaster (fruitfly)	13,917
Pan troglodytes (chimpanzee)	18,746
Canis familiaris (dog)	19,856
Bos taurus (cow)	19,994
Caenorhabditis elegans (nematode)	20.517
***Homo sapiens* (human)**	**20,774**
Arabidopsis thaliana (mustard weed)	27,416
Physcomitrella patens (moss)	35,938
Oryza sativa (rice)	40,577
Populus trichocarpa (poplar)	41,377
Manihot esculenta (cassava)	47,164
Malus domestica (apple)	57,386
Triticum aestivum (bread wheat)	>94,000

Kelly Rae Chi, The dark side of the human genome Nature 538, 275–277 (2016)

Correlation with mRNA

- Lower stability of mRNA (lower half-life)
- Larger distribution of abundance of protein

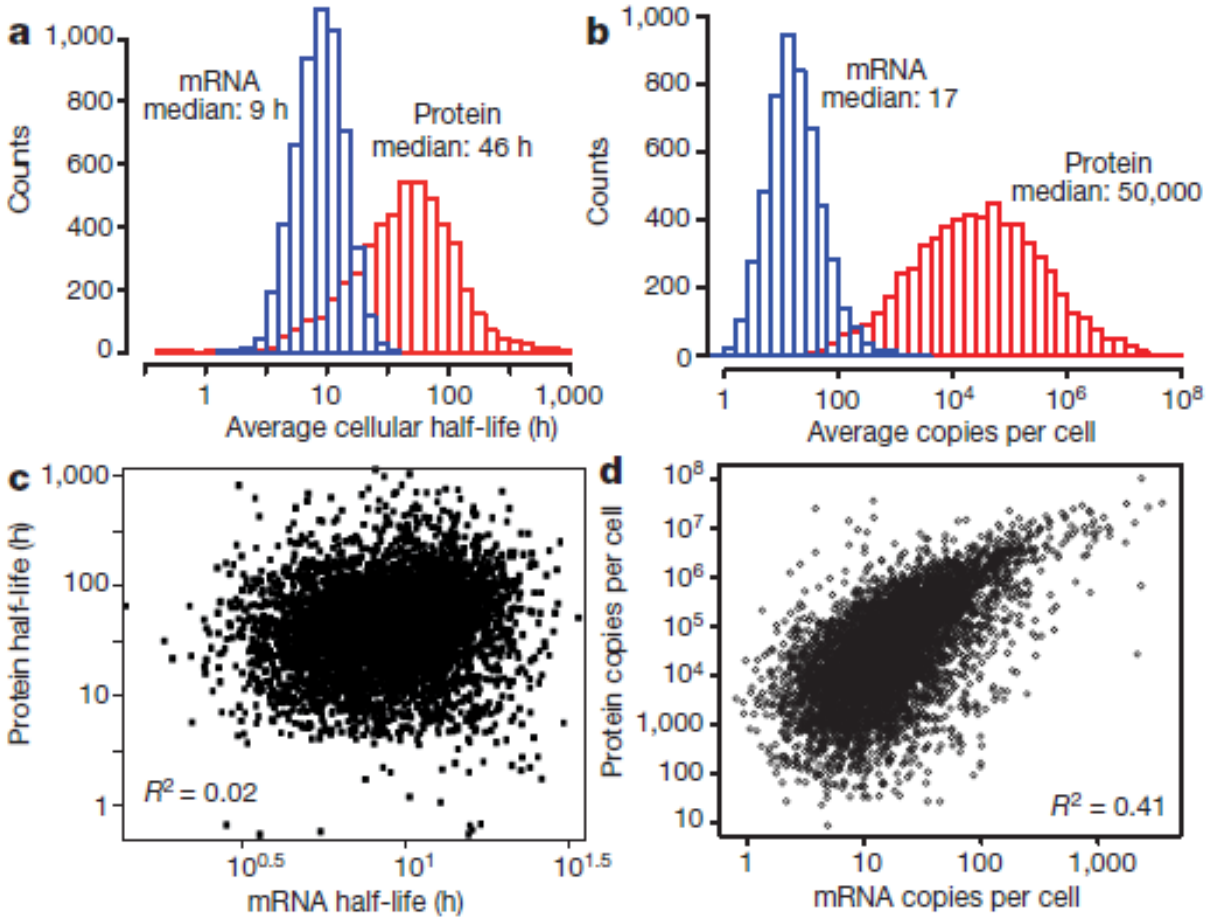


Figure 2 | mRNA and protein levels and half-lives. a, b, Histograms of mRNA (blue) and protein (red) half-lives (a) and levels (b). Proteins were on average 5 times more stable and 2,800 times more abundant than mRNAs and spanned a higher dynamic range. c, d, Although mRNA and protein levels correlated significantly, correlation of half-lives was virtually absent.

Schwanhäusser et al, Nature, 2011

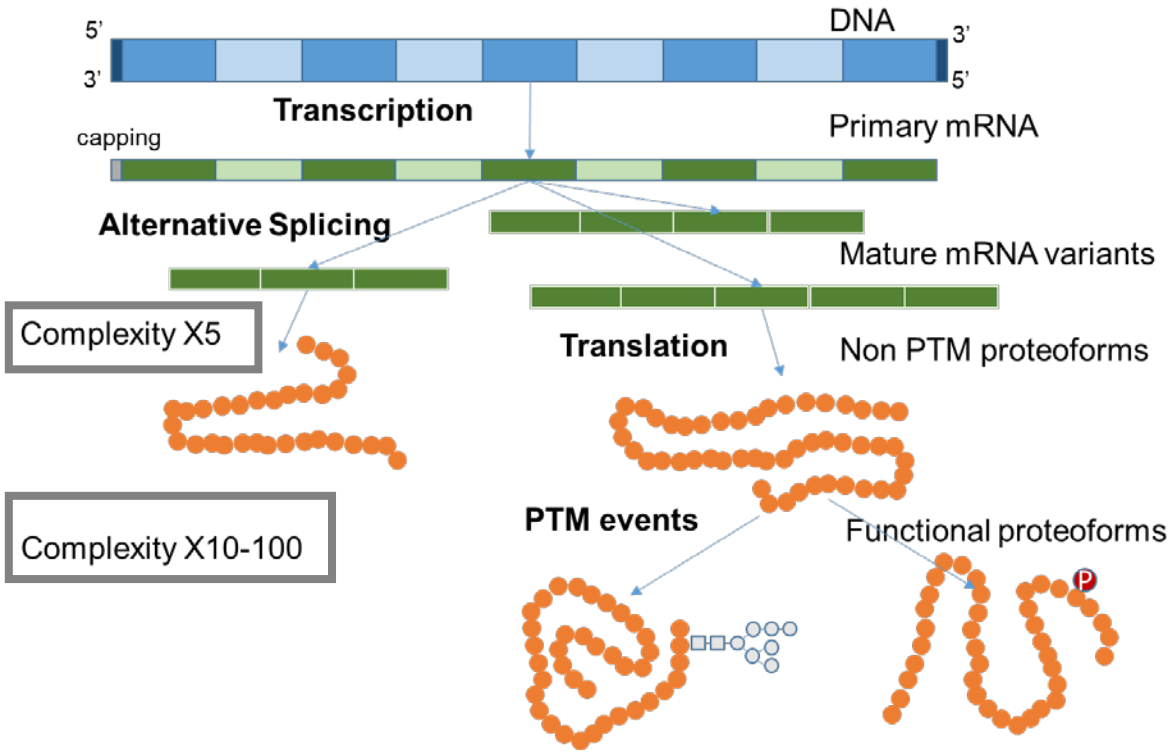
Why study the proteome ?

Genome is static but
Structures change just as the caterpillar develops into a butterfly

Taxonomy of Omics

Let's focus

Genomics	Transcriptomics	Proteomics	Metabolomics
Gene sequences	mRNA sequences & expression level	Proteins isoforms sequences & expression levels	Metabolites as products or substrate of enzymes



PTMomics

Proteoforms sequences & expression levels

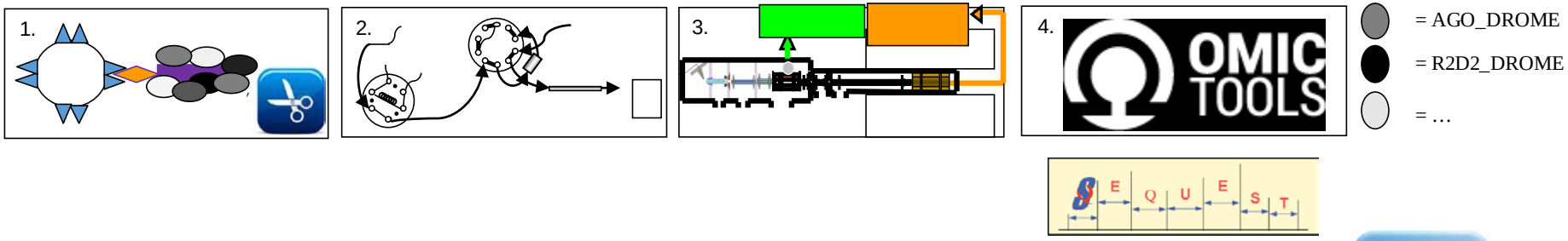
Proteins and Proteoforms: New Separation Challenges Regnier FE and Kim JH, Anal Chem. 2018

How many human proteoforms are there? Aebersold R et al Nat Chem Biol. 2018

Summary

- Omics and Mass spectrometry
- Why study the proteome?
- **Bottom-up proteomics: an LC-MS story**
- DIA and HRM
- Quantification

Bottom-up proteomics workflow



1: Proteins

Extraction, purification of proteins

Proteolysis to generate peptides

2: Peptides

Complex mixtures

Separation and purification

3: Mass spectrometry

Peptide masses

Peptide sequences

4: Data processing

Raw data analysis

MaxQuant

PEAKS

Top-down

Byonic

Bottom-up

MS-based untargeted proteomics

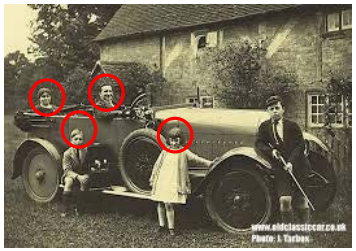
802 tools

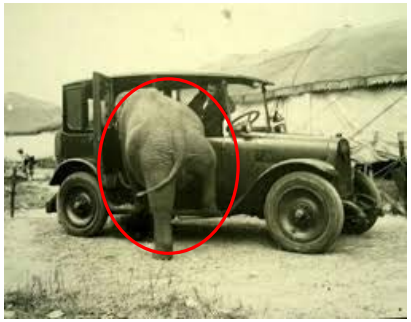
Shotgun Proteomics

Unknown organization

Spare parts inventory

Identification of original composition





Potential modifications

Artifacts

False positive identification

Endoproteolysis

Example: trypsin, which cleaves after the carboxyl moiety

. of lysine (K)

. of arginine (R) (except when followed by proline)

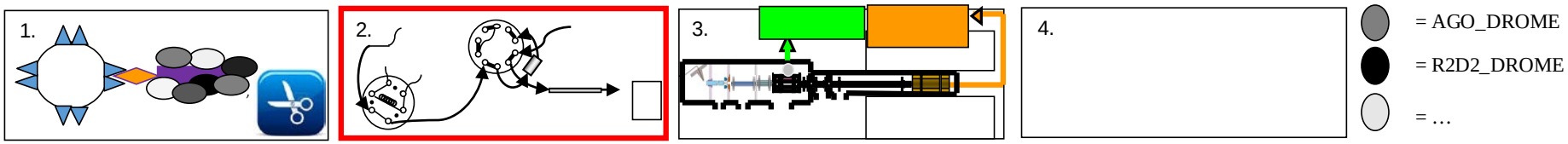
Generate tryptic peptides whose sequencing in tandem mass spectrometry is favored by the basic residue at C-ter

R EMINDER ATTEACTIVITYK FTR YPSINK PEPTIDEBNDS

1 missed cleavage:

R,
EMINDER,
ATTEACTIVITYK,
FTR,
YPSINK,
PEPTIDEBNDS,
REMINDER,
EMINDER**ATTEACTIVITYK**,
ATTEACTIVITYKFTR,
FTRYPSINK,
YPSINK**PEPTIDEBNDS**

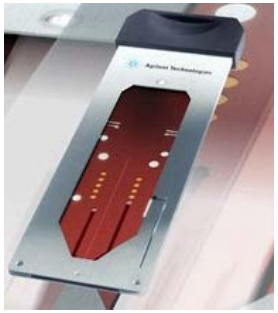
Separation and purification



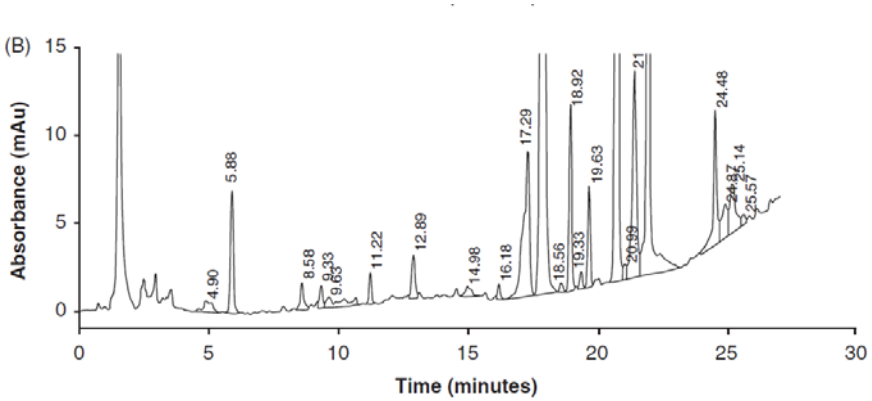
- Charge → ion exchange
- Polarity → HILIC
- **Hydrophobicity → reverse phase chromatography**

Conventional

Nano



Chip-LC/MS



Elution : organic solvent gradient (acetonitrile, methanol, …)

Compatible with mass spectrometry

LC-MS is a Gold standard for bottom-up proteomics

• Liquid chromatography aims at the separation of molecules according to their physico-chemical properties using various stationary and mobile phases

• From people in the separation sciences we need a detector and

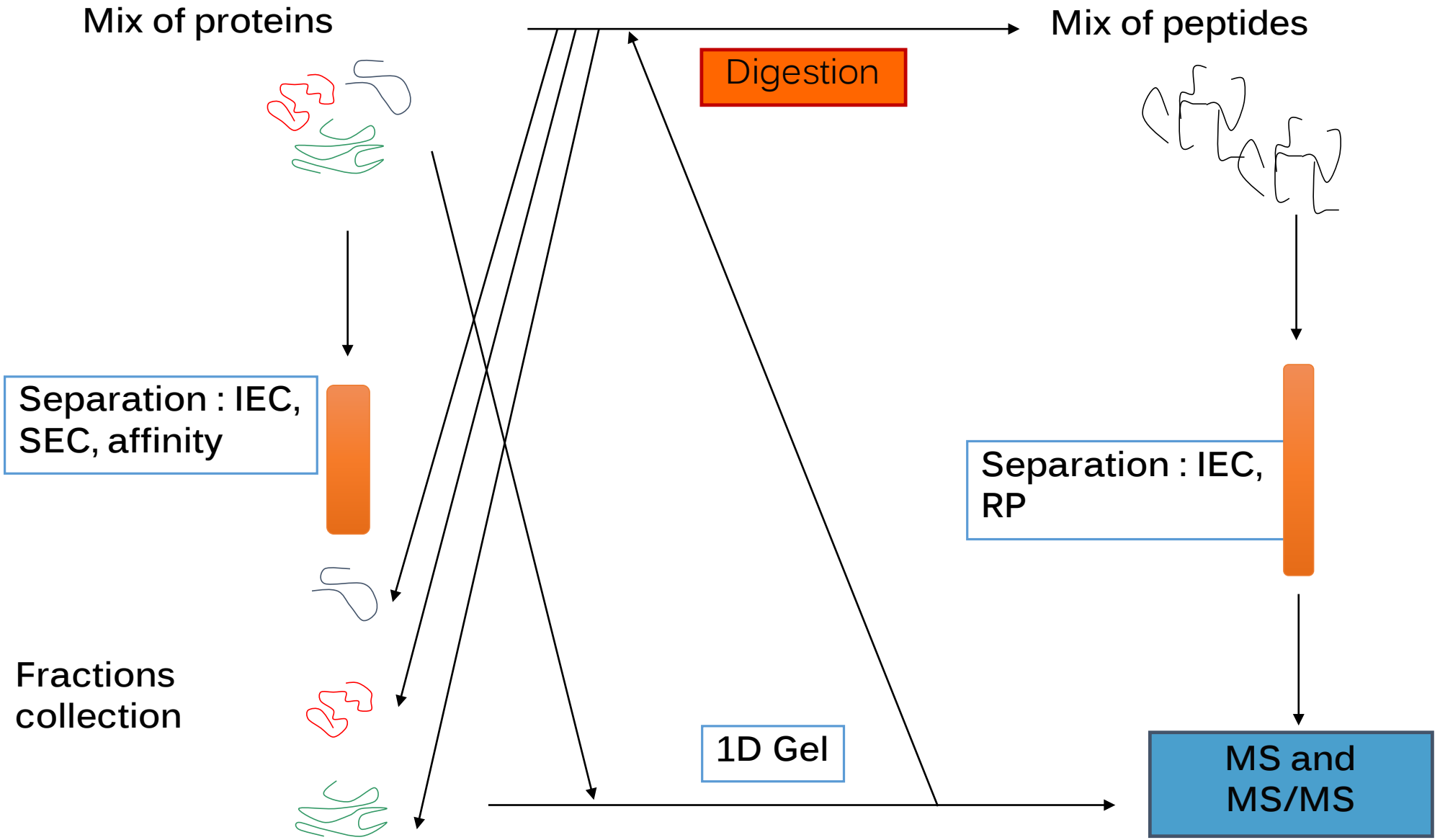
> A mass spectrometer is one of them, that allows the m/z measurement of eluted species.

• Separation of proteins or peptides ?

• Which phase, which columns?

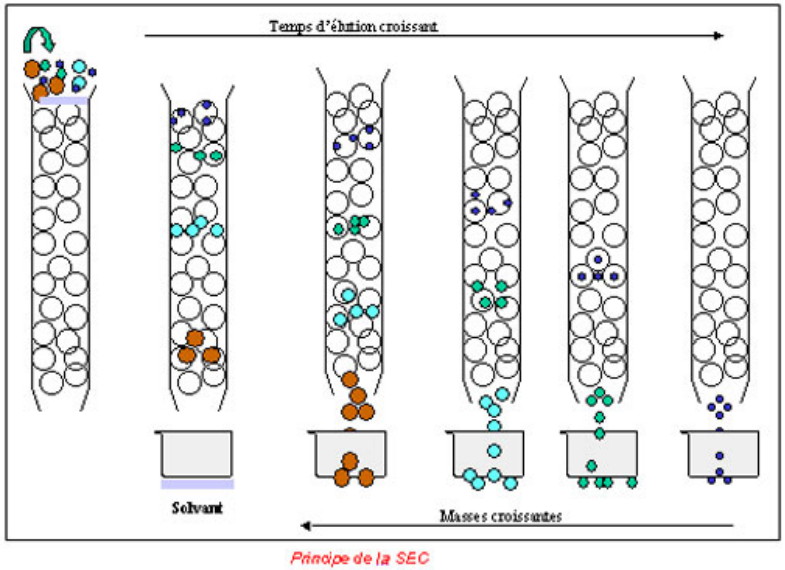
• Which ion source?

Chromatographic tools



Chromatographic separation mode

Properties of proteins		Chromatographic methods
• Size (MW)	⟶	Size/steric exclusion chromatography

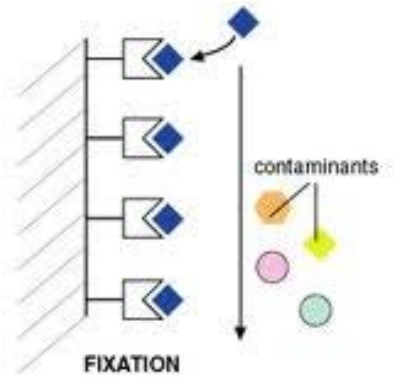


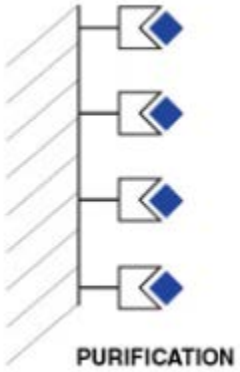
Porous material?

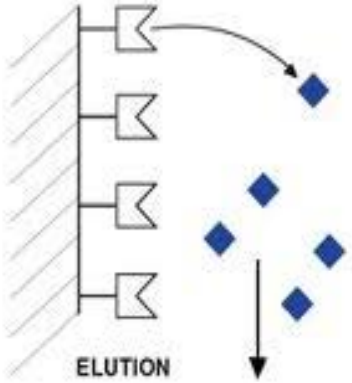
Elution Order: large molecules first

Chromatographic separation mode

Properties of proteins		Chromatographic methods
• Size (MW)	⟶	Size/steric exclusion chromatography
• Activity (affinity)	⟶	Affinity chromatography



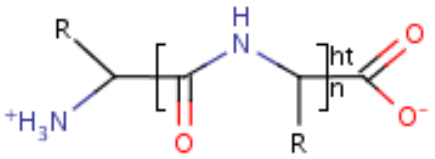




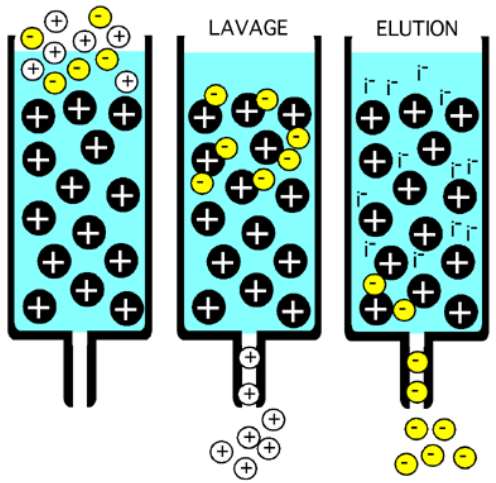
Elution: Ionic strenght gradient
pH gradient
competitive elution

Chromatographic separation mode

Properties of proteins		Chromatographic methods
• Size (MW)	→	Size/steric exclusion chromatography
• Activity (affinity)	→	Affinity chromatography
• Charge (pI)	→	Ion exchange chromatography



pH > pI : -
pH < pI : +

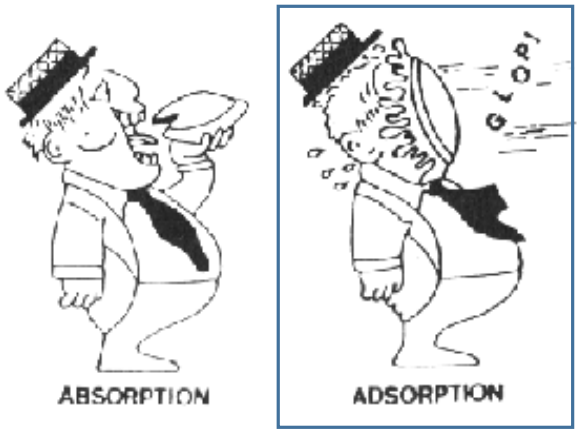


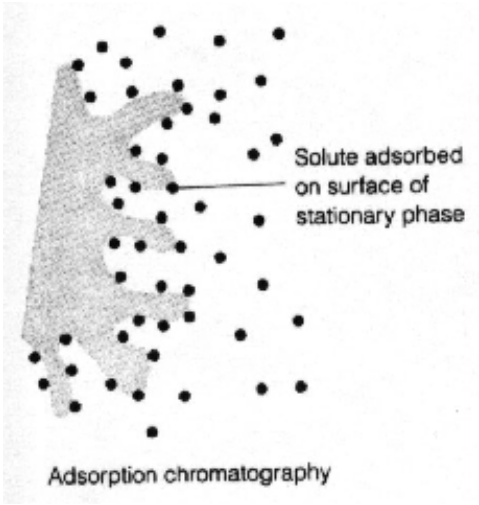
Anions exchange: amino moiety
Cations exchange : sulfonic/phospho/carboxylic moiety

Elution: Ionic strenght gradient
pH gradient

Chromatographic separation mode

Properties of proteins		Chromatographic methods
• Size (MW)	→	Size/steric exclusion chromatography
• Activity (affinity)	→	Affinity chromatography
• Charge (pI)	→	Ion exchange chromatography
• Polarity	→	Normal phase chromatography

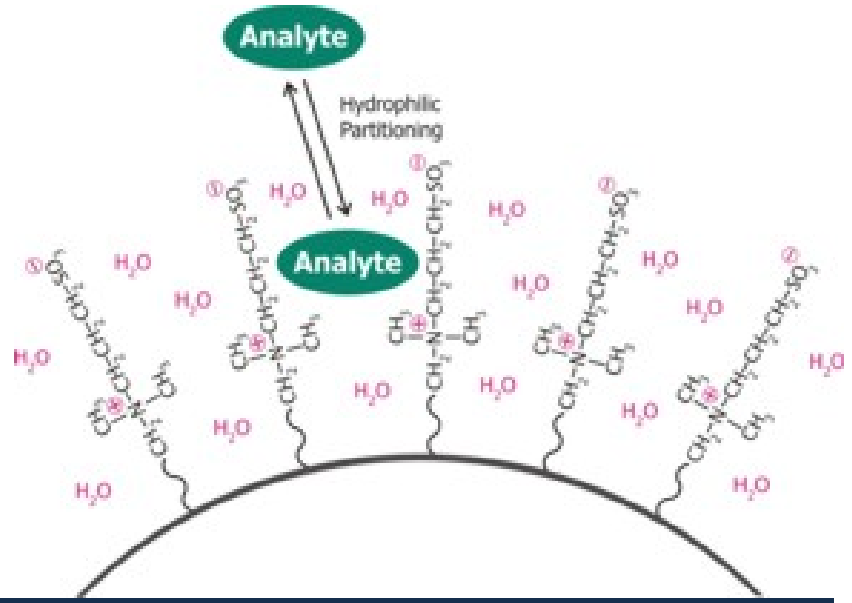




Elution : organic solvent gradient (Hexane, toluene,...)

Chromatographic separation mode

Properties of proteins	Chromatographic methods
• Size (MW)	Size/steric exclusion chromatography
• Activity (affinity)	Affinity chromatography
• Charge (pI)	Ion exchange chromatography
• Polarity	Normal phase chromatography
	Hydrophilic interaction chromatography (HILIC)

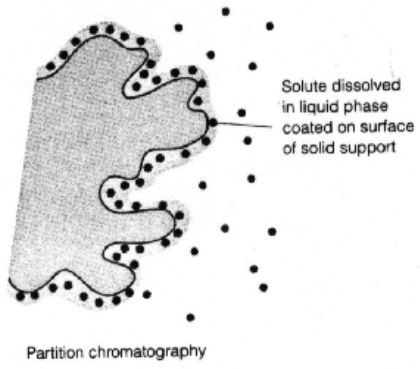


Elution : aqueous phase gradient

Easy LC-MS interface

Chromatographic separation mode

Properties of proteins	Chromatographic methods
• Size (MW)	Size/steric exclusion chromatography
• Activity (affinity)	Affinity chromatography
• Charge (pI)	Ion exchange chromatography
• Polarity	Normal phase chromatography
	Hydrophilic interaction chromatography (HILIC)
• Hydrophobicity	Reversed phase chromatography (RP)



Partition chromatography

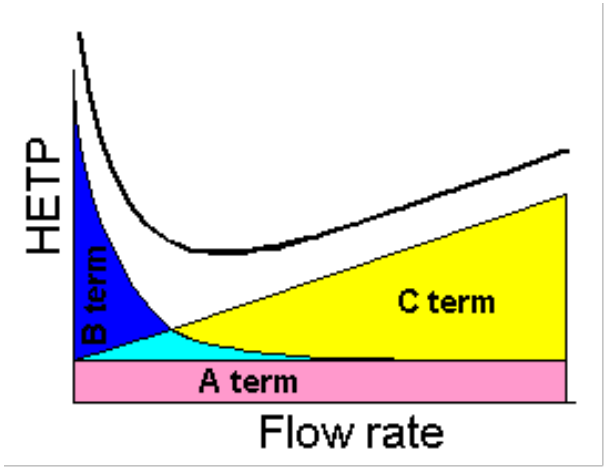
Elution : organic solvent gradient (acetonitrile, methanol, …)

Easy LC-MS interface

Selection of column dimension

- Column selection is a function of the analytes of interest, **nature**, **complexity and abundance**.

➢ Size and nature of particules

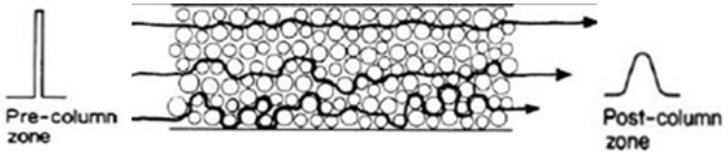


Van Deemter equation

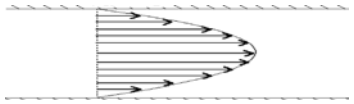
$$H = 2\lambda d_p + 2\frac{\gamma D_m}{v} + \frac{\omega d_p^2}{D_m} v$$

$$H = A + \frac{B}{u} + C * u$$

A: Convection diffusion ($\varphi_{stationnaire}$)



B: Longitudinal diffusion (φ_{mobile})

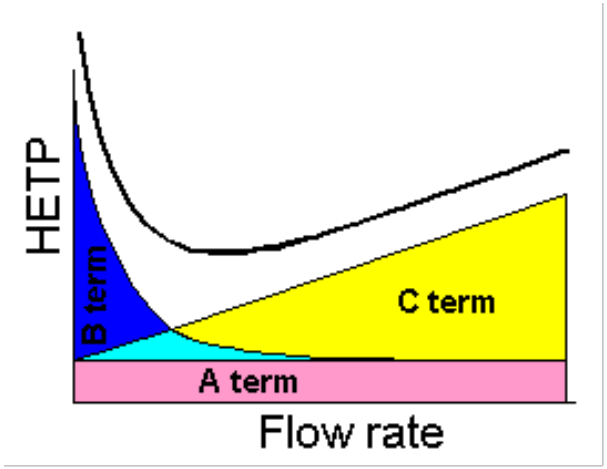


C: Mass transfer resistance ($\varphi_{stationnaire}$ /φ_{mobile})

Selection of column dimension

- Column selection is a function of the analytes of interest, **nature**, **complexity and abundance**.

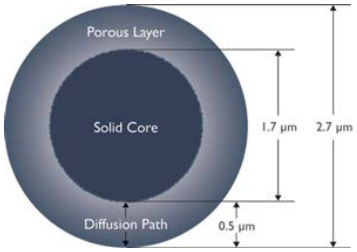
➢ Size and nature of particules

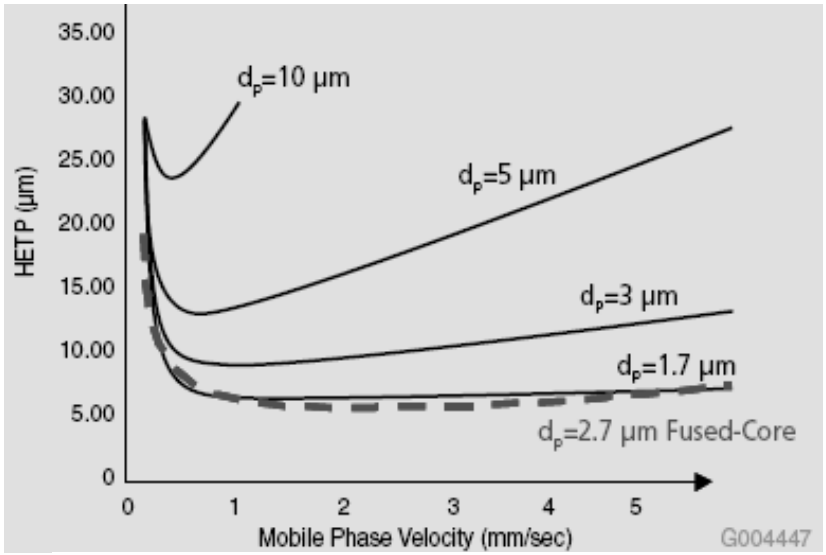


Van Deemter equation

$$H = 2\lambda d_p + 2\frac{\gamma D_m}{v} + \frac{\omega d_p^2}{D_m} v$$

$$H = A + \frac{B}{u} + C * u$$



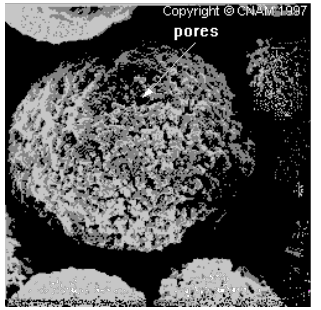


Pressure (Darcy): $\Delta P = \frac{\phi . \eta . L . u}{d_p^2}$

Selection of column dimension

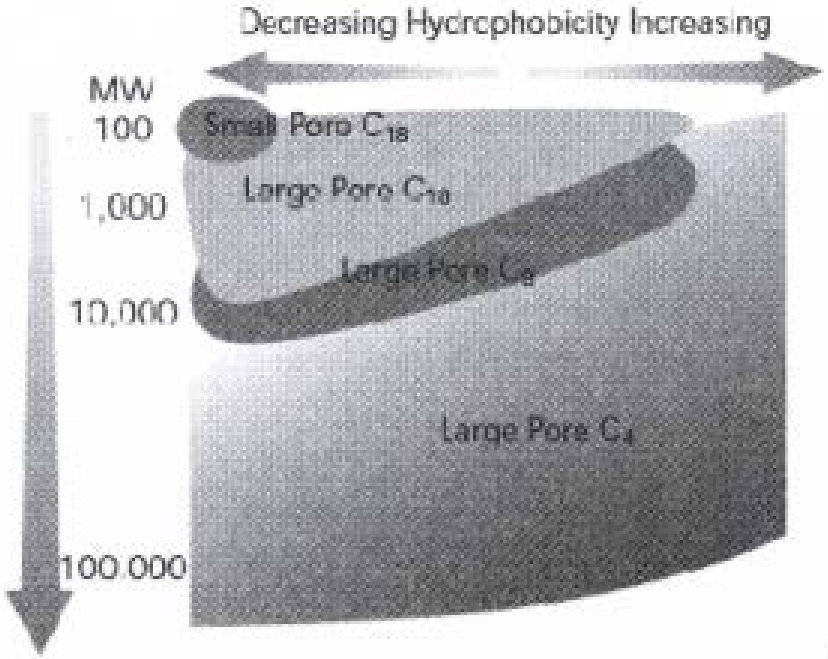
- Column selection is a function of the analytes of interest, **nature**, **complexity and abundance**.

  - Pore size/ grafted alkyle chain



peptides

proteins



d_{pore} > 500 Å macroporous

20 Å < d_{pore} < 500 Å mesoporous

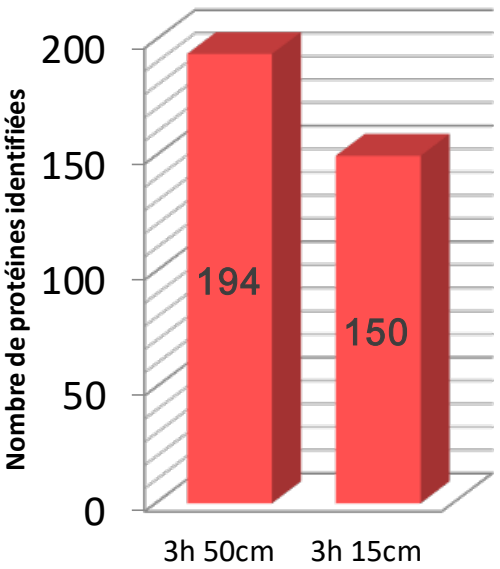
d_{pore} < 20 Å microporous

Selection of column dimension

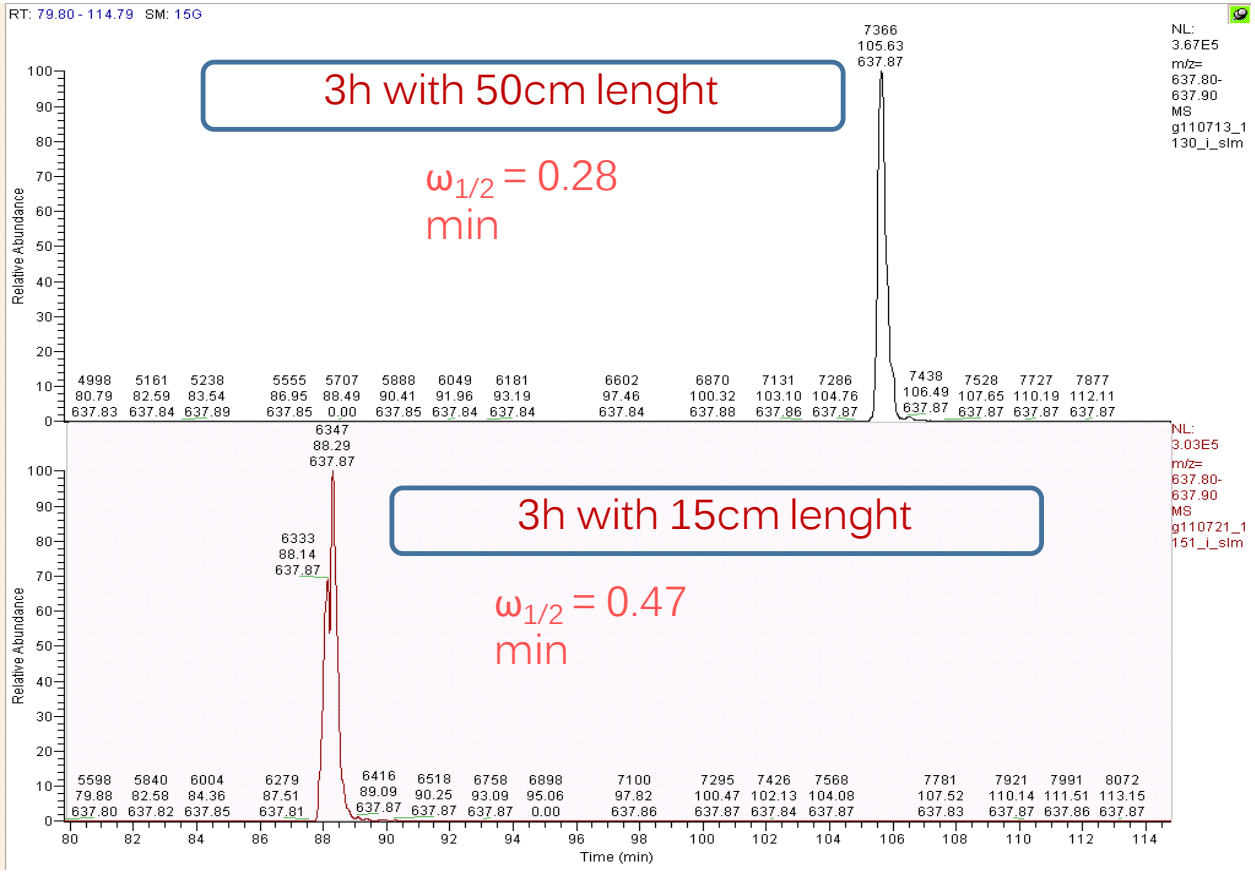
- Column selection is a function of the analytes of interest, **nature**, **complexity** and **abundance**.

➢ Column lenght

$$R = 2\frac{(t_{r2} - t_{r1})}{(\omega_1 + \omega_2)} = \sqrt{N_2}\frac{\alpha - 1}{\alpha}\frac{k_2}{1 + k_2} \qquad N = \frac{L}{H} \quad (eq.2)$$



Sample : Kidney Biopsy (S. Liuu et al)

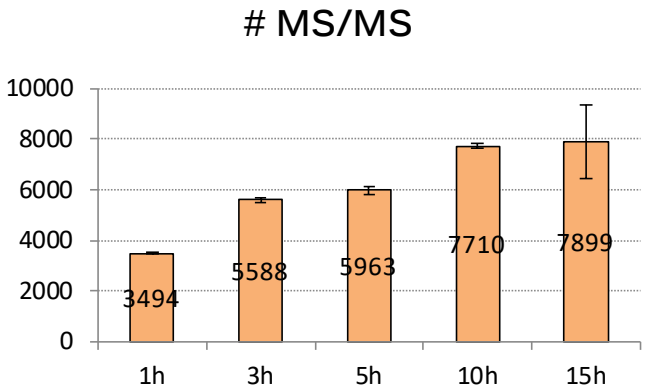


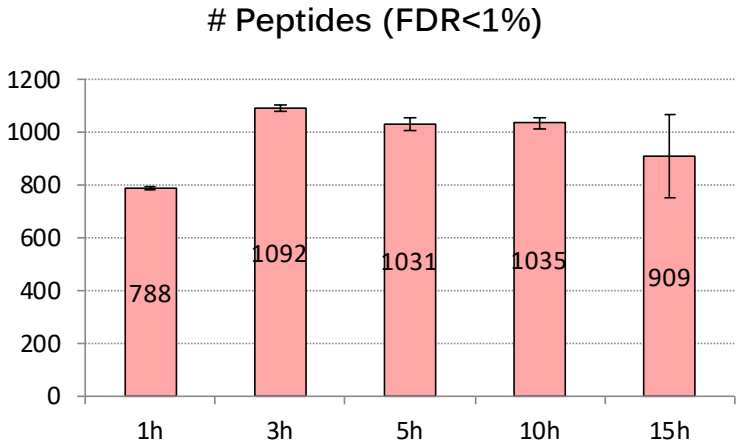
LTQ - FT Ultra (ThermoFisher Scientific)

Selection of column dimension

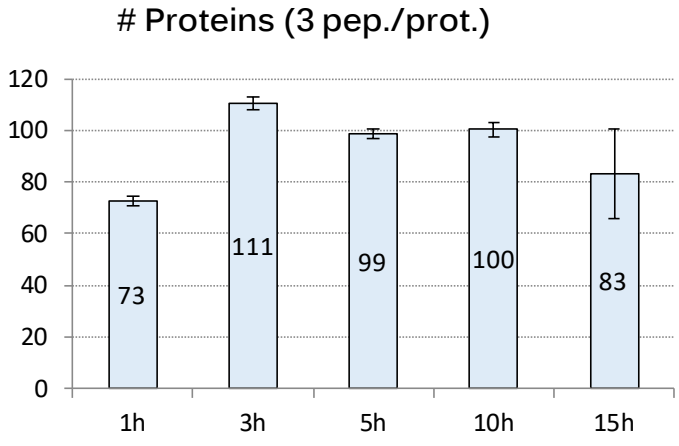
- Column selection is a function of the analytes of interest, **nature**, **complexity** and **abundance**.

➢ Gradient duration/slope





Sample : Kidney Biopsy (S. Liuu et al)



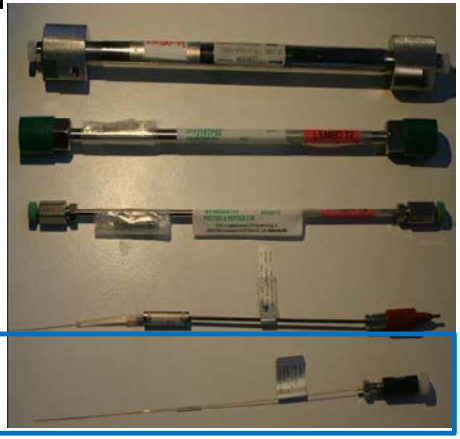
LTQ - FT Ultra (ThermoFisher Scientific)

Selection of column dimension

- Column selection is a function of the analytes of interest, **nature**, **complexity and abundance**.

  ➢ Column internal diameter and flow rate

Diamètre interne	Nomenclature	Débit	Quantité
4 mm	Conventionnel LC	1ml/min	1-200 µg (10pmol)
2 mm	Narrowbore LC	200µl/min	2-50 µg (5pmol)
1 mm	Micro LC	40µl/min	0,05-10 µg (1pmol)
300 µm	Capillaire LC	4µl/min	1 ng-1µg (500 fmol)
75 µm	NanoLC	300nl/min	0,02-0,05 ng (1 fmol)



Increased sensitivity (ESI signal concentration dependant)

Advantage of low flow rates

- The conventional ESI interfaces operate at high flow rates: typically, 1-1000µL.min-1 is a typical concentration-sensitive technique. Within this range, increasing the flow rate does not normally increase the signal.
- NanoESI-MS, at low flow rates (typically, 300-50 nL.min-1), exhibits superior mass sensitivity, with a high ionization efficiency.

$$C_{max} = \frac{mN^{1/2}}{(2\pi)^{1/2}\mathbf{V_0}(1+k)}$$

- Cmax : eluted analyte concentration
- m : absolute abundance
- N : column efficiency
- V_0 : column volume
- k : retention factor

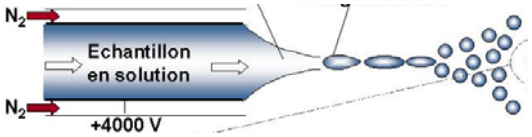
- Max Concentration = K/(Volume colonne)

→ decrease the column diameter

→ optimization of the flow rate for nanoESI mode

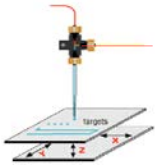
Which ion source ?

- ESI : Compatibility with liquid phase



- MALDI : Fraction collection

Output
nanoLC



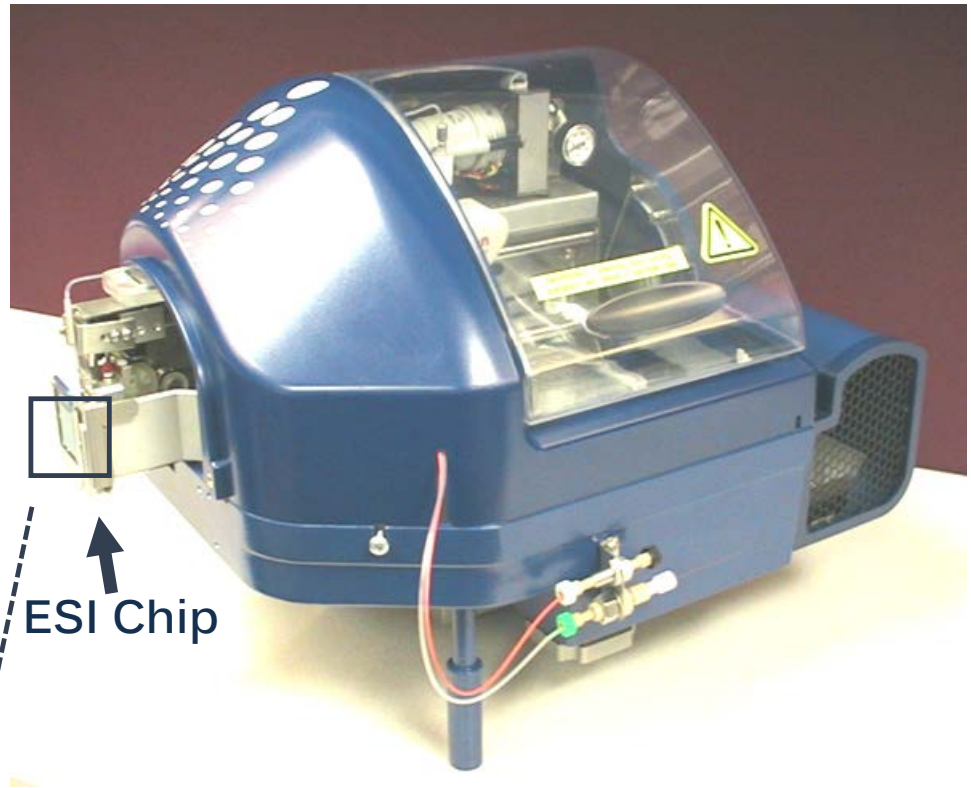
Addition of matrix

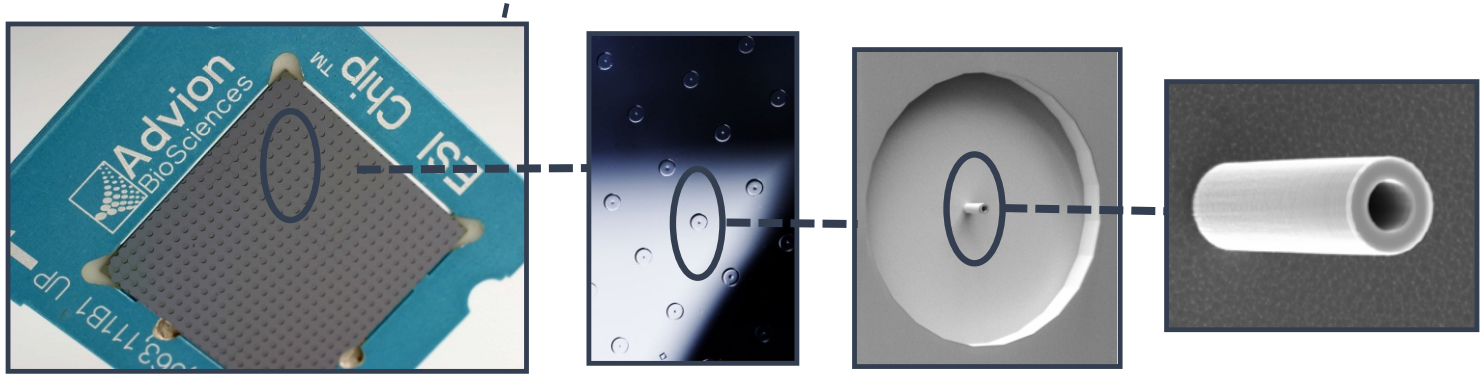
- Complementarity:
  - Ion mode => different ion nature
  - Advantages MALDI: Off-line analysis
  - Advantages ESI: Speed

TriVersa NanoMate (Advion)

Combination Autosampler/ion source
- Fraction collector
- Sample cooling
- “Spray sensing”

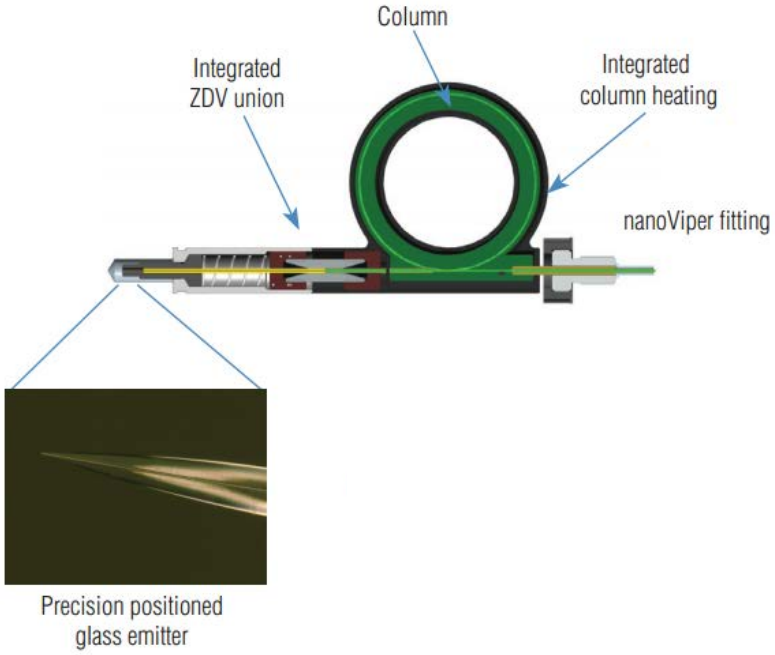
4 operation modes
- direct infusion
- LC Coupling
- Fraction collector
- LESA (Liquid Extraction Surface Analysis)



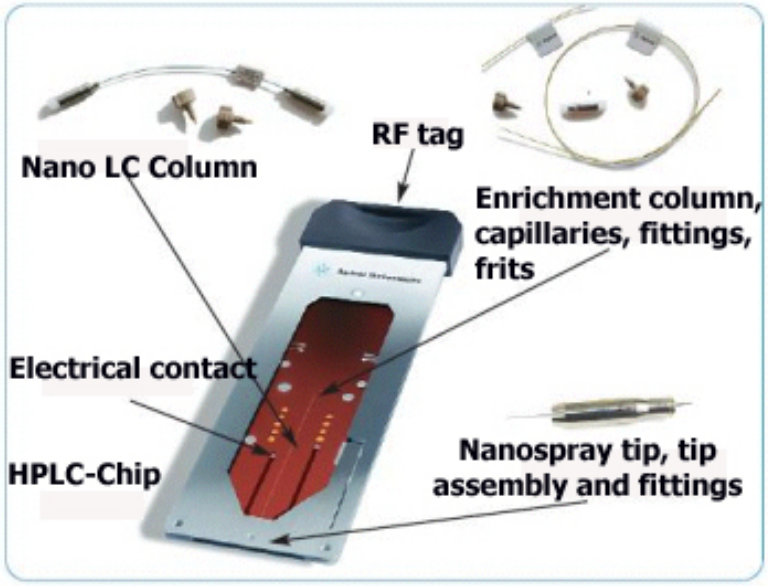


Other interfaces nano-ESI

EASY-Spray

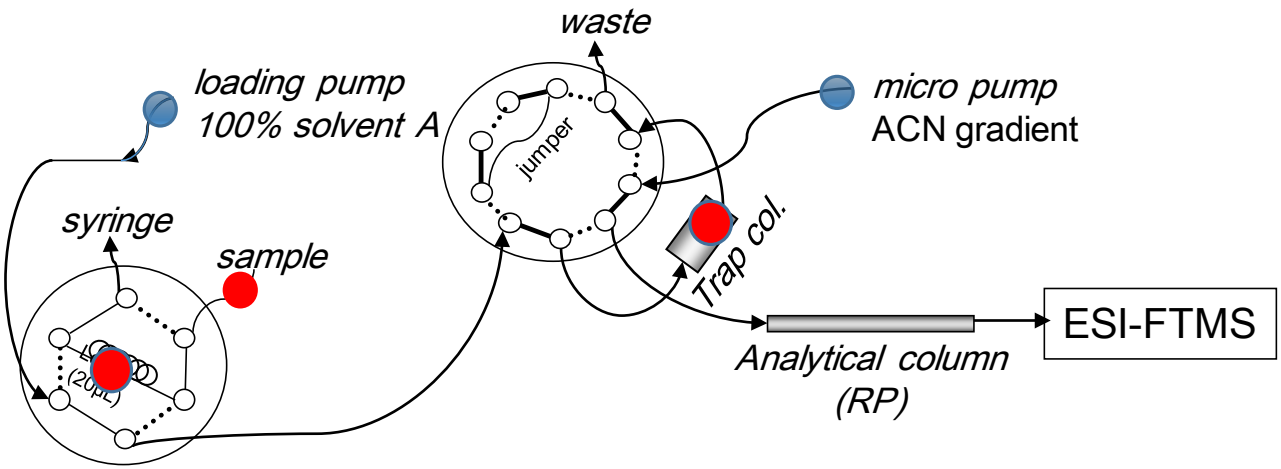


Chip Cube Interface

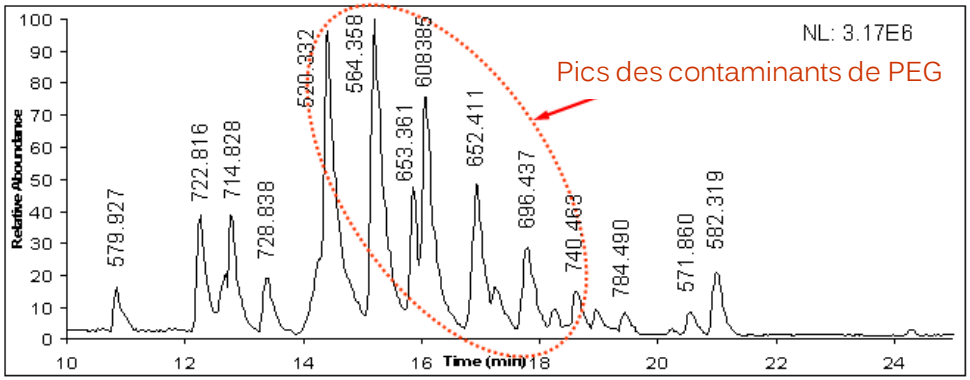


- Reduction of dead volumes (extra-column, leaks, etc.)
- Robustness ease of use
- Costs

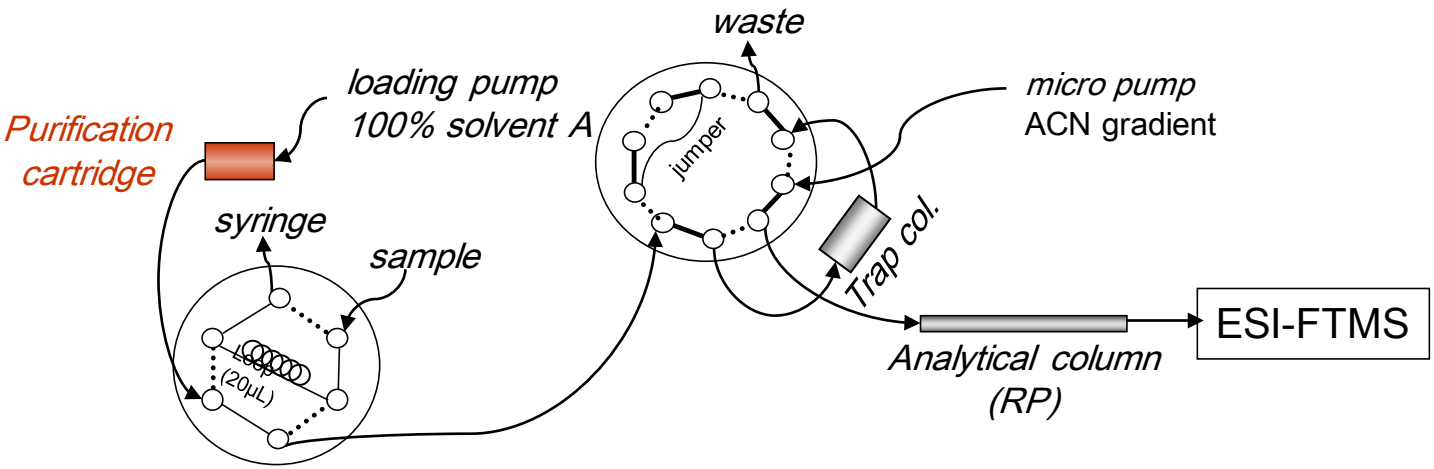
Following the samples … and others



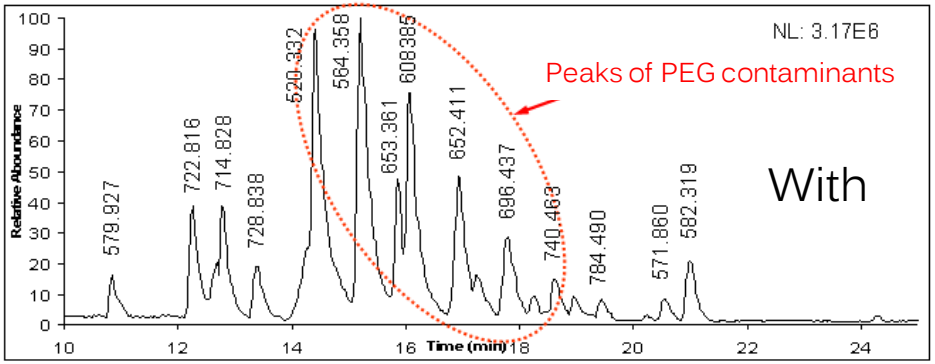
A



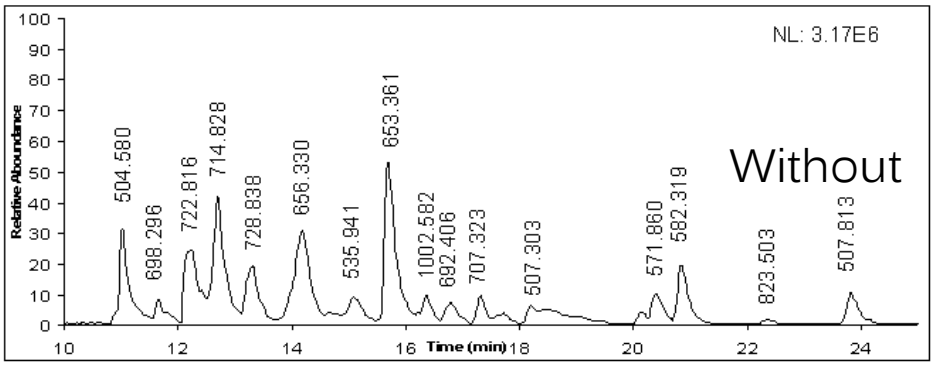
Following the samples … and others



A



B



IDENTIFICATION		CARTOUCHE DE PURIFICATION			
		SANS		AVEC	
	PROTEINE	Score	# peptides	Score	# peptides
ADH1_YEAST	Alcool déshydrogénase 1	139 +/- 19	2 +/- 0	298 +/- 10	5 +/- 0
ALBU_BOVIN	Sérum albumine	401 +/- 20	7 +/- 1	631 +/- 4	14 +/- 1
BGAL_ECOLI	β-galactosidase	149 +/- 21	3 +/- 1	194 +/- 27	4 +/- 1
CYC_BOVIN	Cytochrome C	49 +/- 43	1 +/- 1	196 +/- 46	3 +/- 1
LYSC_CHICK	Lysozyme C	74 +/- 3	1 +/- 1	75 +/- 3	1 +/- 0
TRFE_BOVIN	Sérotransferrine	236 +/- 148	4 +/- 2	632 +/- 46	12 +/- 1

Gain in sensitivité

Hesse *et al.*, J. Chrom. A, 1189 (2008)

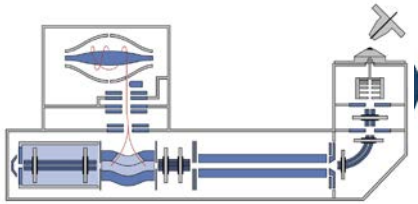
Optimization of peptides analysis in NanoLC-MS/MS

- **In all cases**:
  - ✓ **Duration of gradient** according to sample complexity
  - ✓ Duration of dynamic **exclusion according to peak width**

- **ESI** :
  - ✓ **duty cycle**
  - ✓ Scan time (best **ratio quality/time**)
  - ✓ **# of MS/MS per MS**

- **MALDI** : time is not (?) an issue …
  - ✓ Fraction c**ollection**
  - ✓ **MS/MS** (energy, # laser shots)
  - ✓ **# de MS/MS per spot**

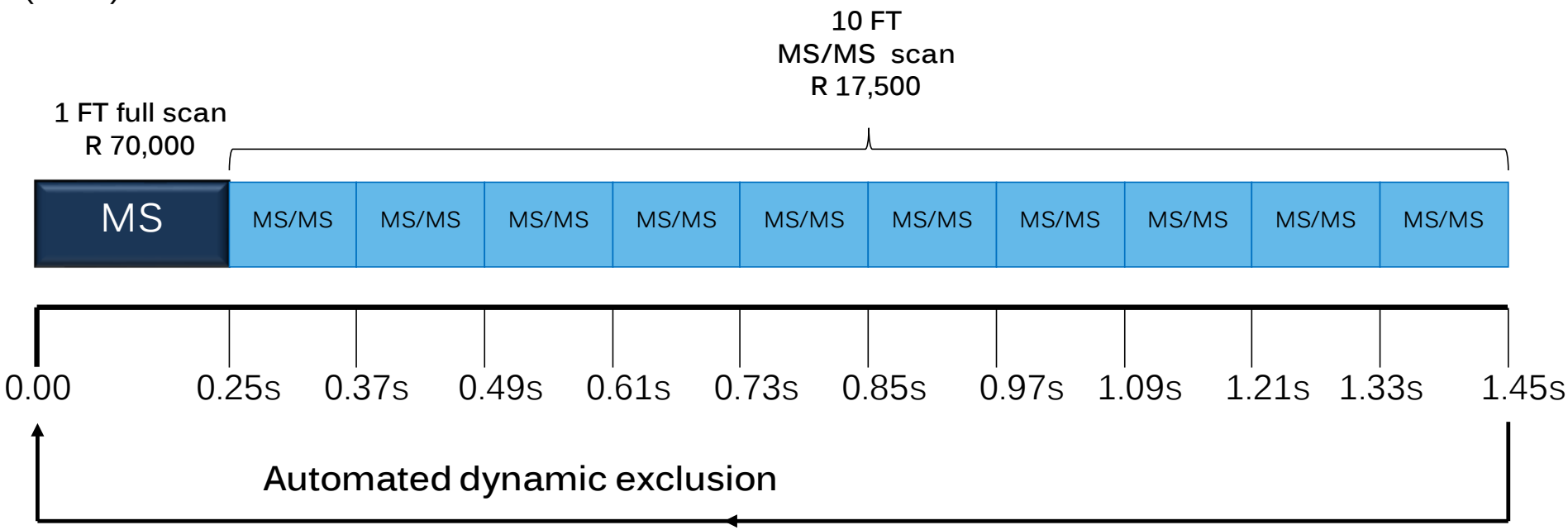
NanoLC-MS with sequential MS/MS and MS



Elution gradient

0 => 50 % B in 180 min then 90% B for 15 min @300nl/min

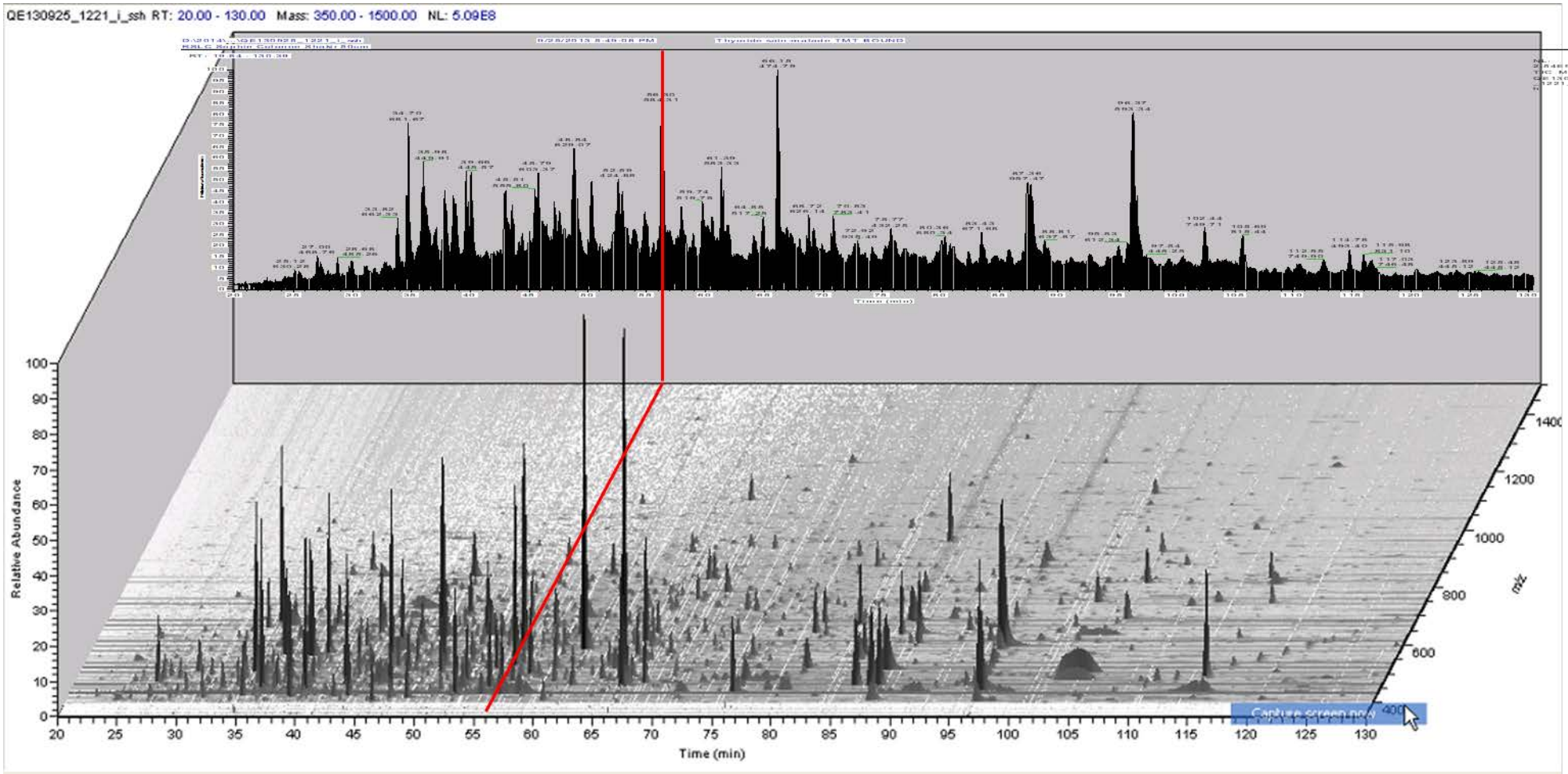
A = 98% water/ 2% ACN/ 0,1% AF (v:v:v) ; B = 10% water/ 90% ACN/ 0,1% AF (v:v:v)



Automatic edition of a precursor list with their fragments

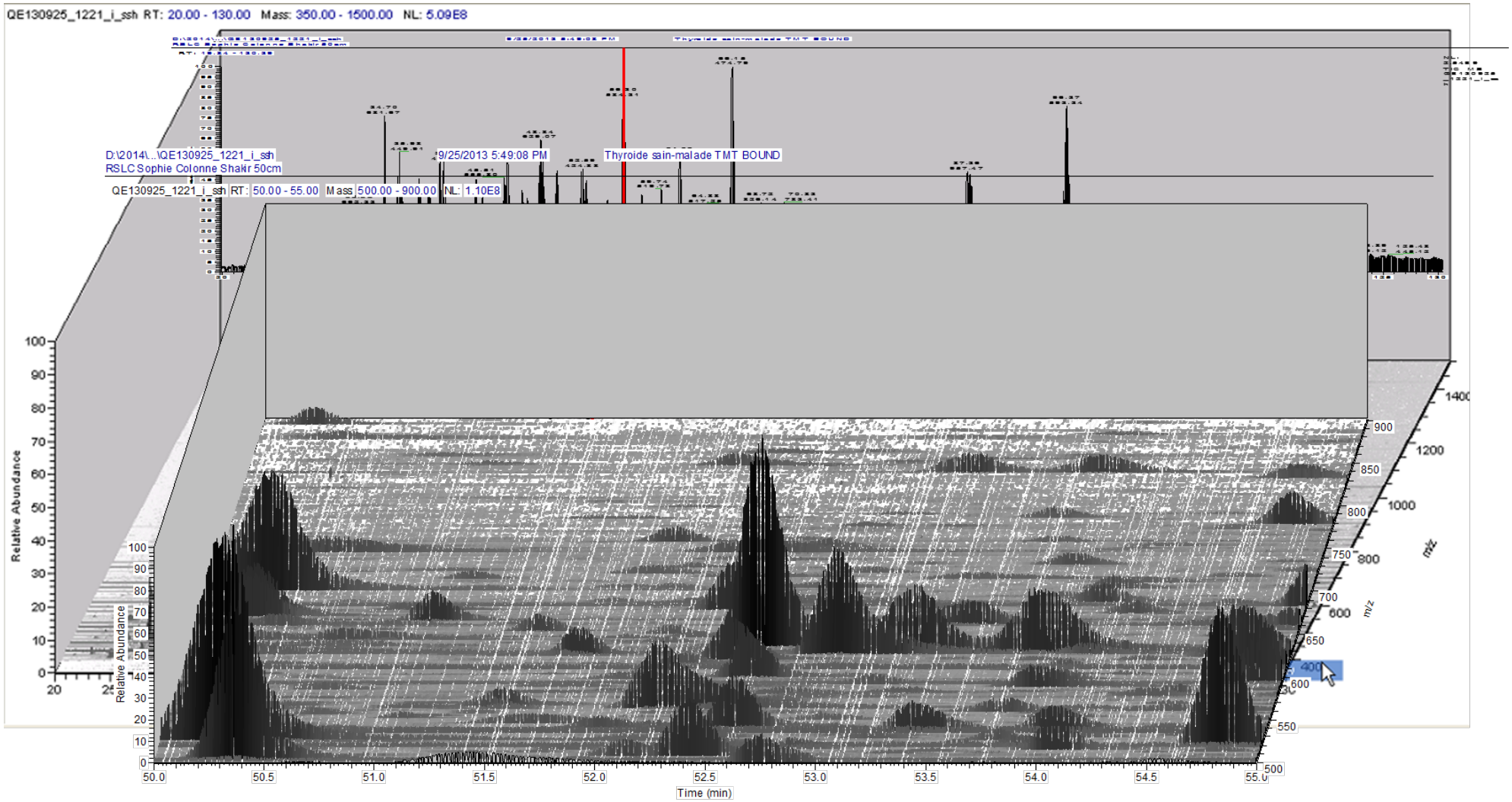
Instruments : Ultimate 3000 RSLC (Dionex), Qexactive (ThermoFisher Scientific)

Example of LC MS profile
Thyroid biopsy



Condition LC: RSLC, column C18, 5µ, 300A, 75µm id, 50 cm, 35° C, 300nL/min, 200bar
Condition MS: Qexactive, DDA scan top 10, $res_{m/z300}$ MS 50000, $res_{m/z300}$ MS/MS 17500

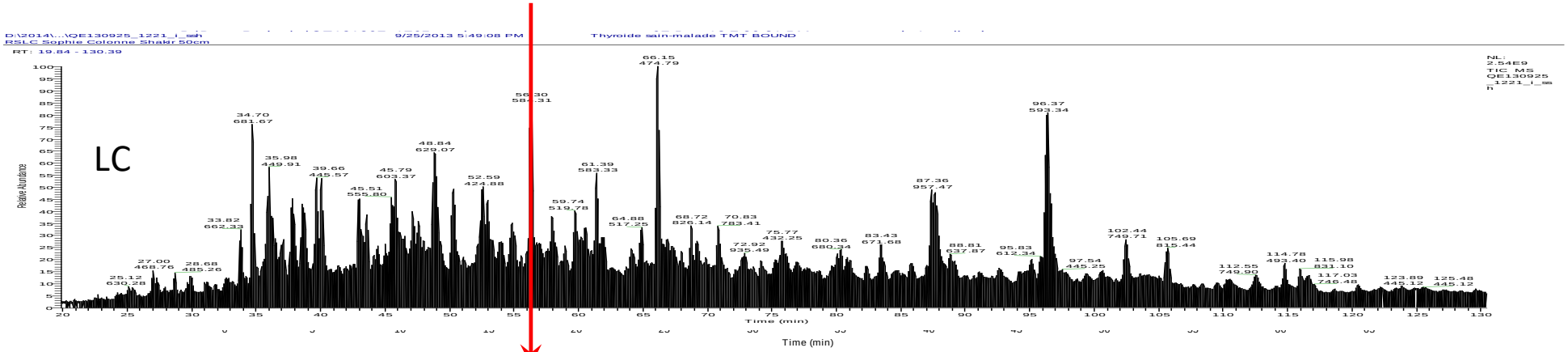
Zoom on 2D map: dynamic range challenge



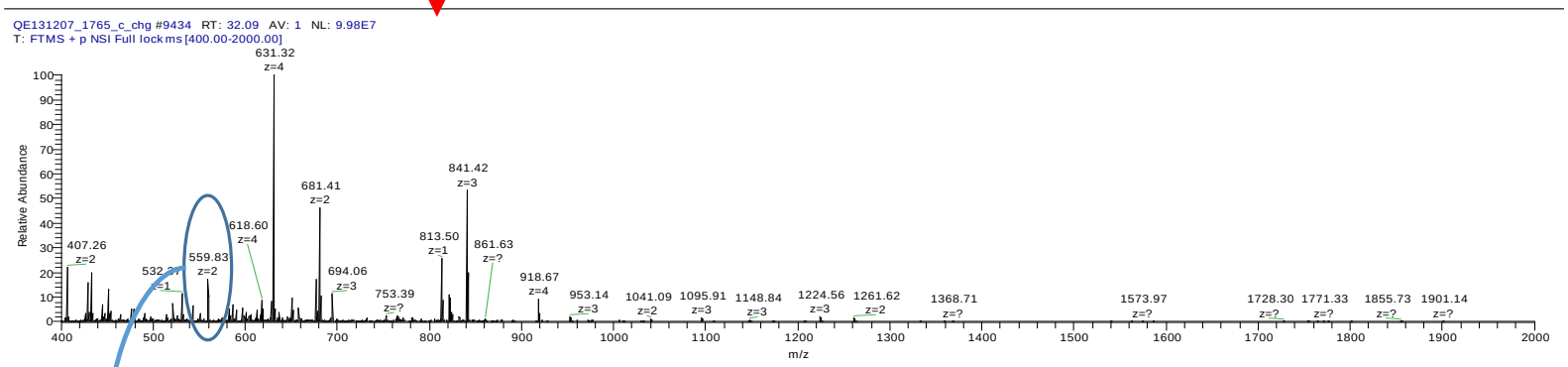
Zoom 50-55min, m/z 500-900, peak width=30s, sampling time MS 0,6s and MS/MS < 120ms

Raw data processing

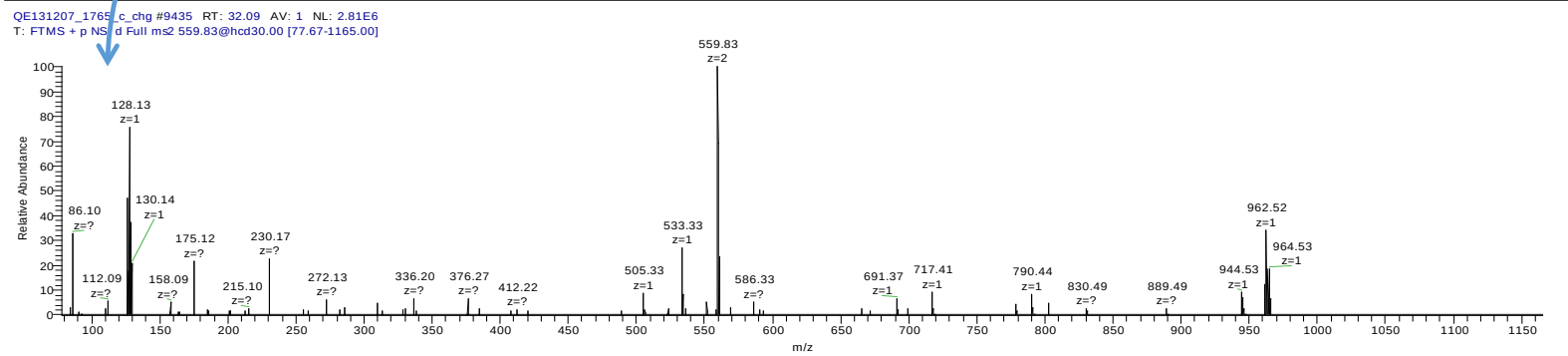
LC



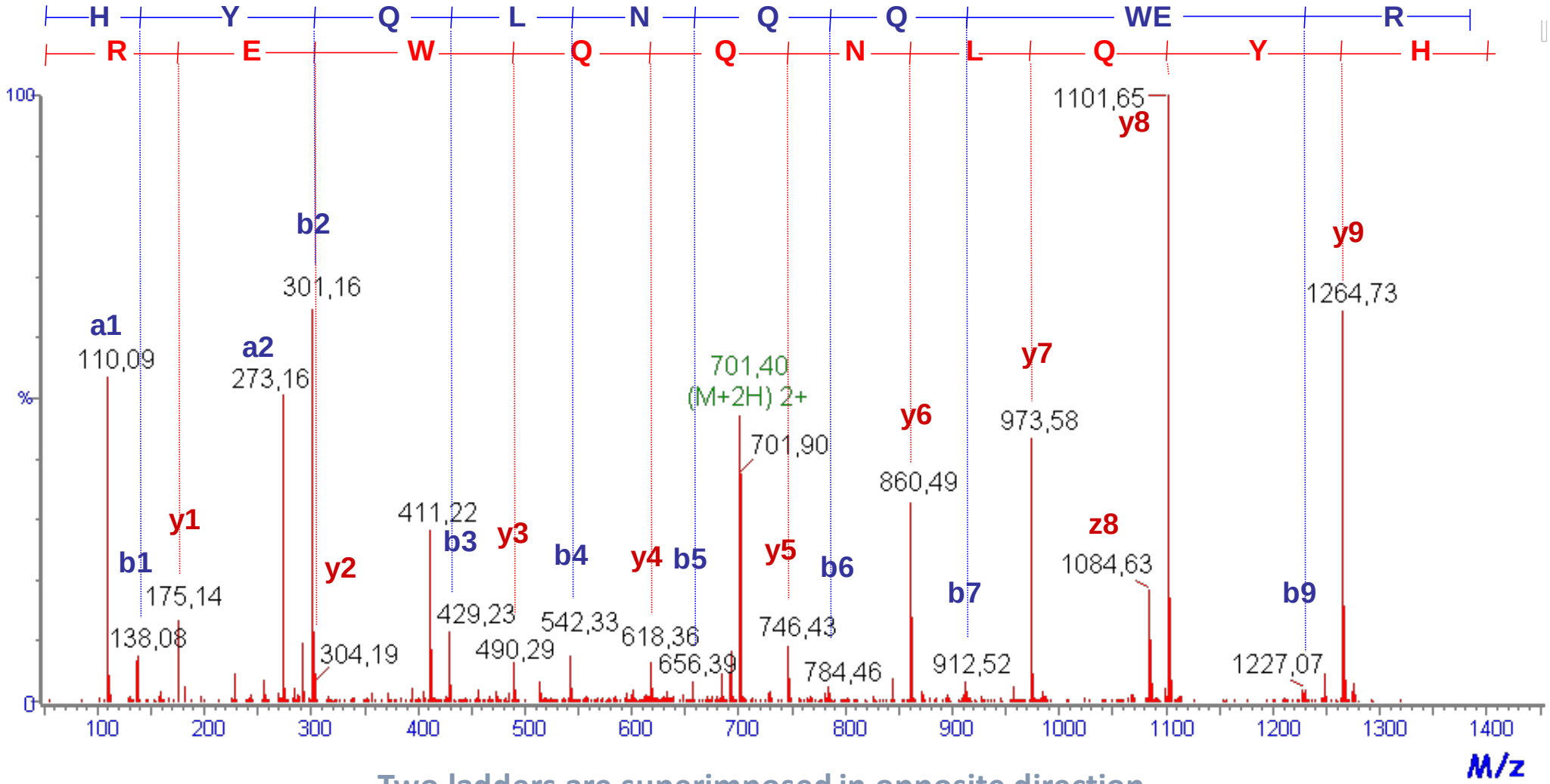
Full scan MS



MS/MS



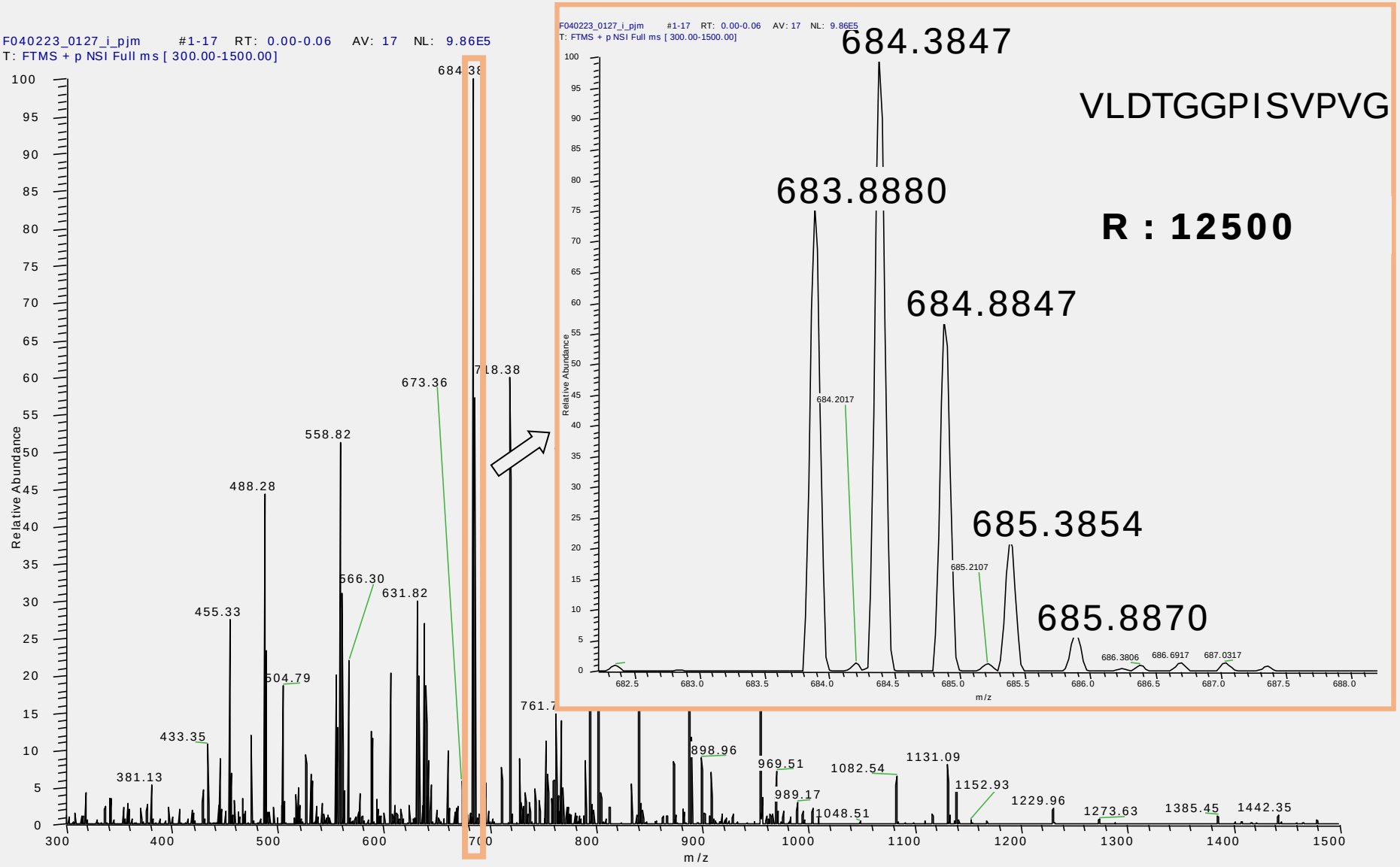
Peptide sequencing: Example of one MS/MS spectra



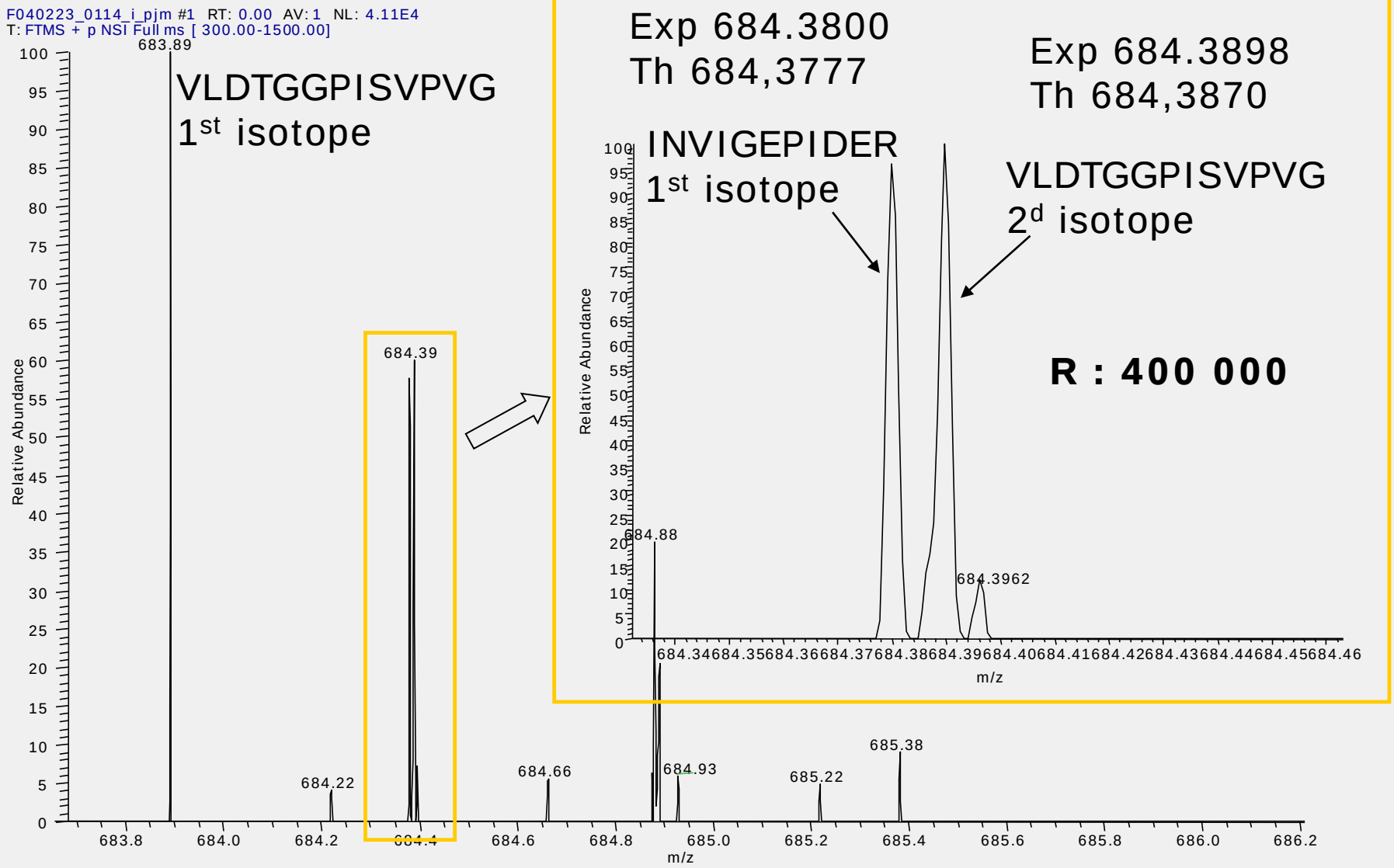
Two ladders are superimposed in opposite direction

Identification of sequence HYQLNQQWER

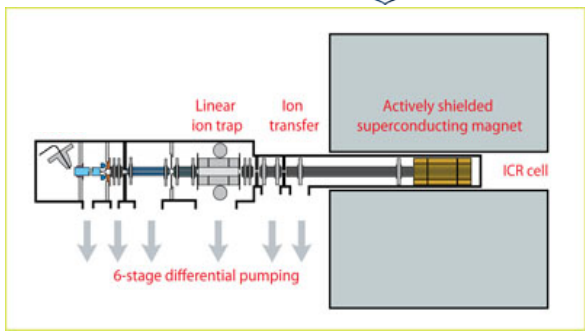
High accuracy / high resolution effect on a peptidic mix



High accuracy / high resolution effect on a peptidic mix

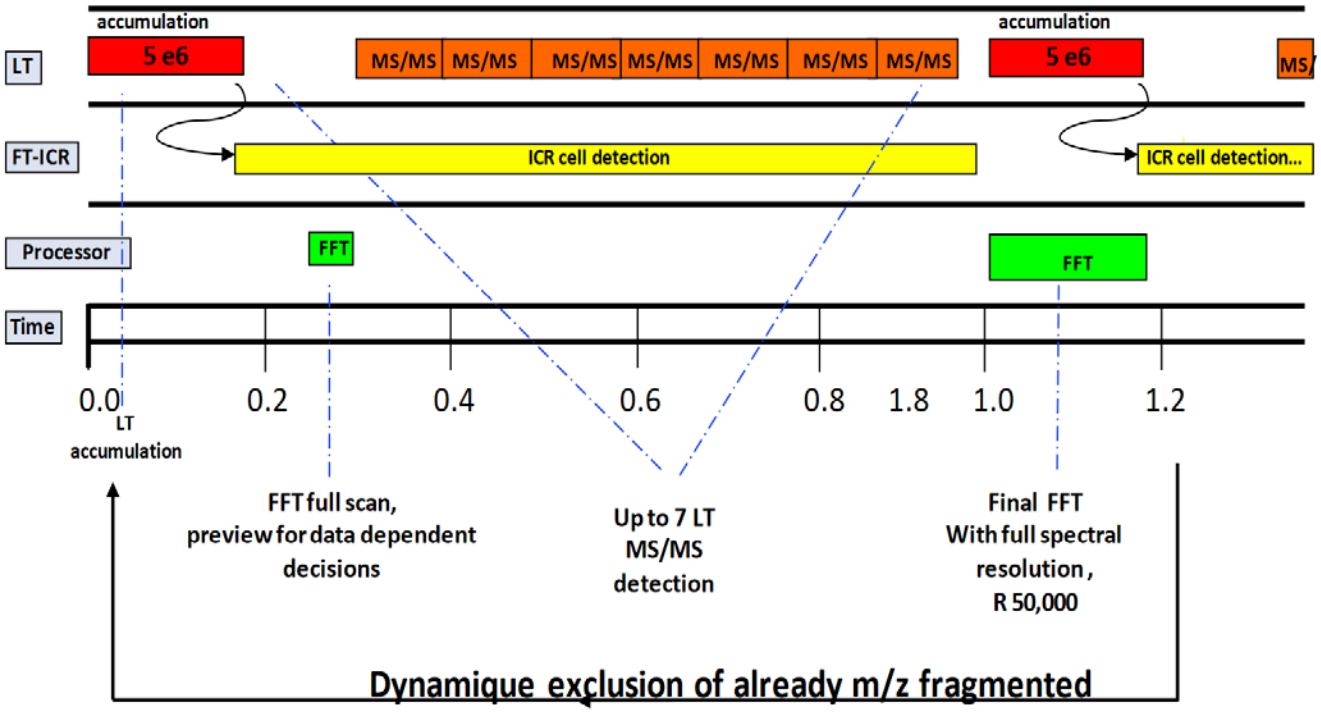


NanoLC-MS with parallel IT MS/MS and FT MS



Elution Gradient

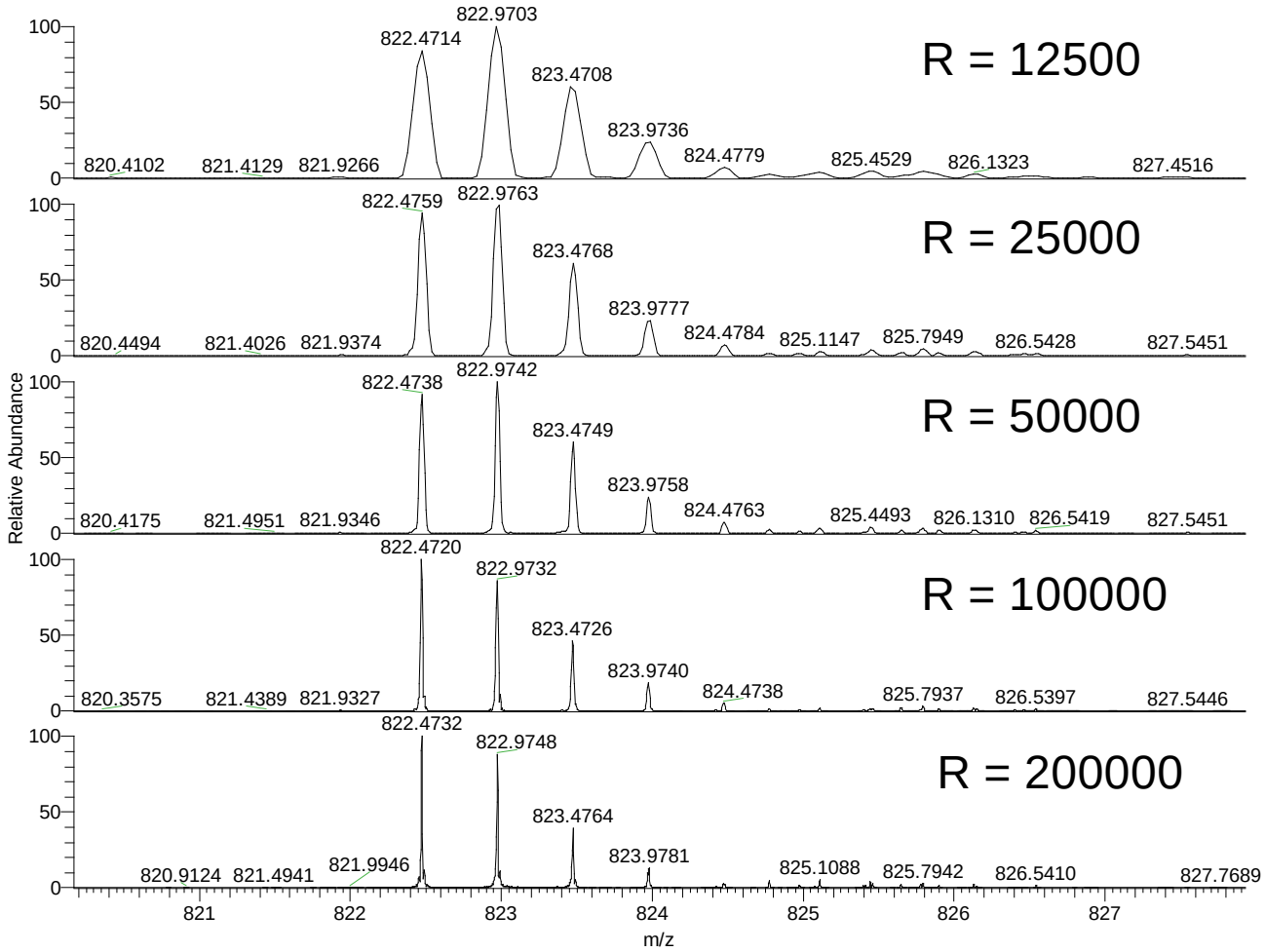
0 => 50 % B in 35 min then 100% B for 10 min
A = 98% water/ 2% ACN/ 0,1% AF (v:v) ; B = 10% water/ 90% ACN/ 0,085% AF (v:v)



Automatic edition of a precursor list with their fragments

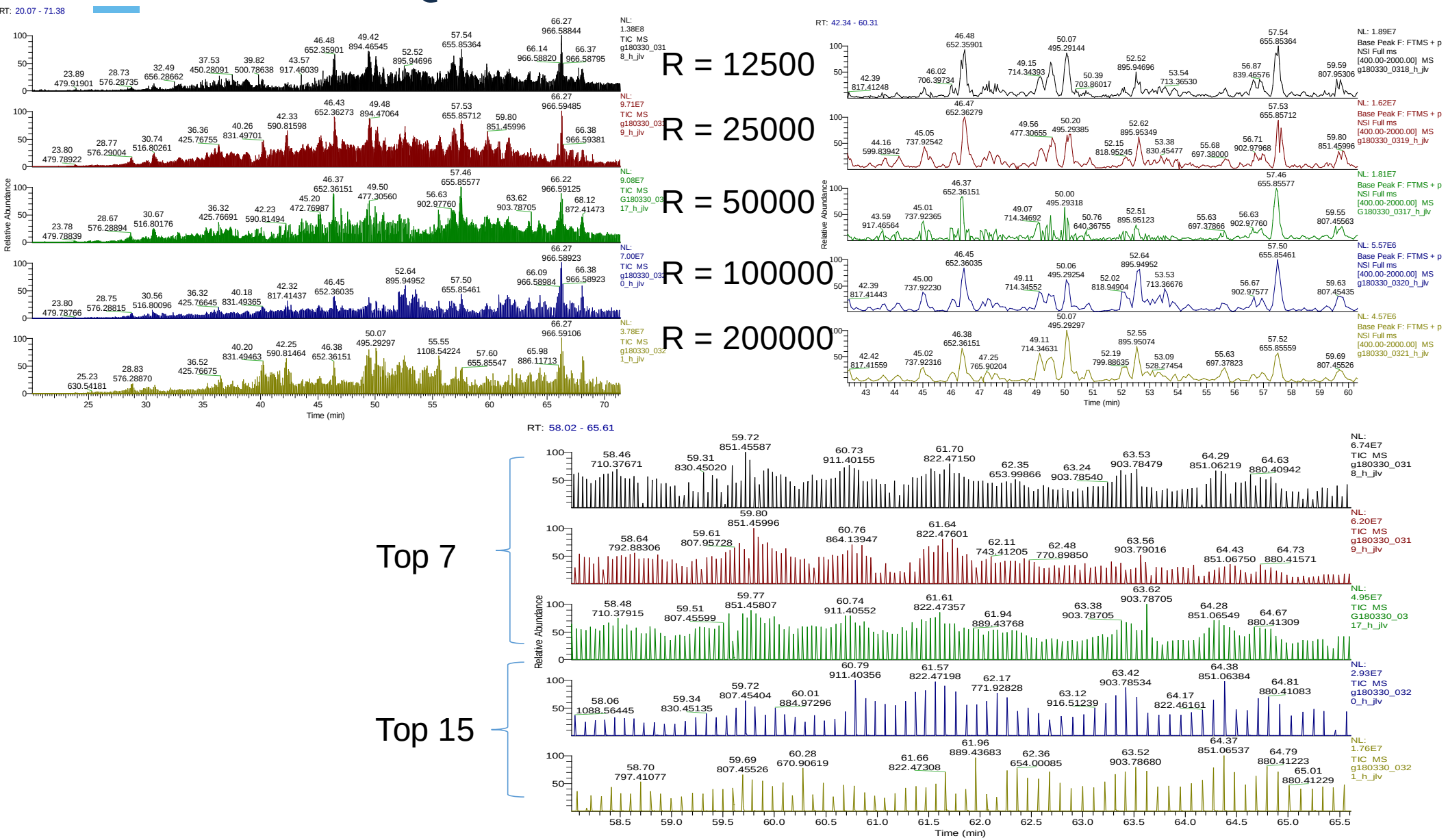
Instruments : Ultimate 3000 (Dionex), LTQ FT (ThermoFisher Scientific)

Hela digest in 30 min gradient by nanoLC LTQ MS/MS FTICR MS

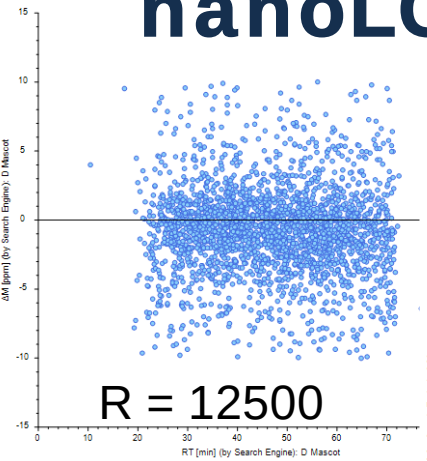


At m/z 400

Hela digest in 30 min gradient by nanoLC LTQ MS/MS FTICR MS

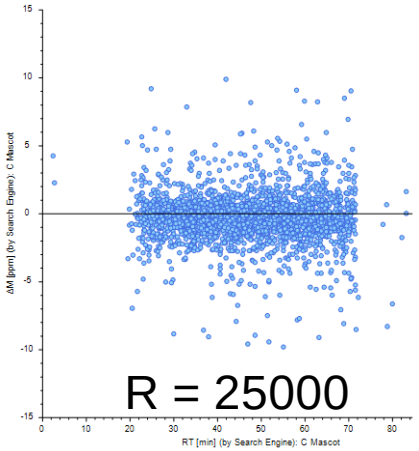


Hela digest in 30 min gradient by nanoLC LTQ MS/MS FTICR MS



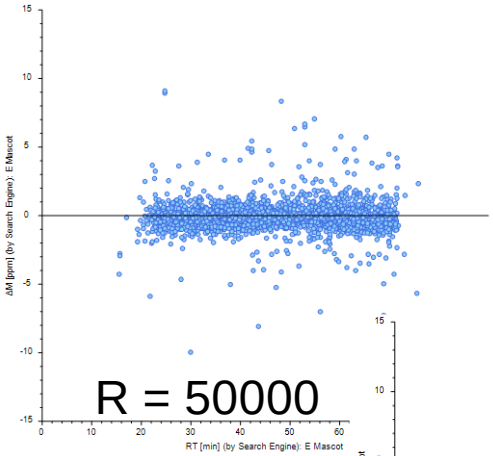
R = 12500

10ppm



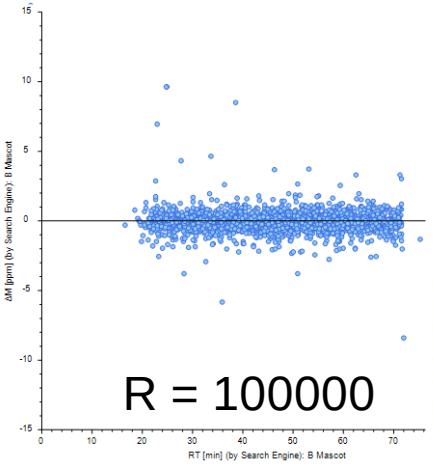
R = 25000

5ppm



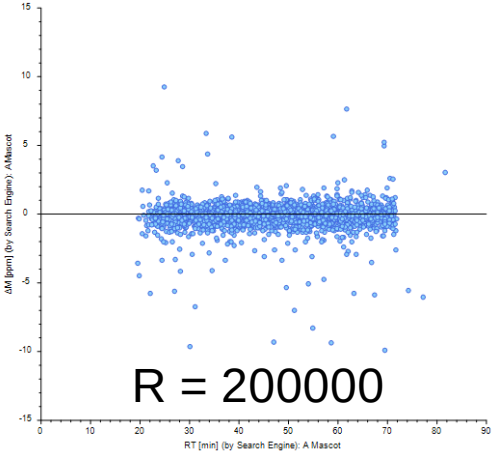
R = 50000

4ppm



R = 100000

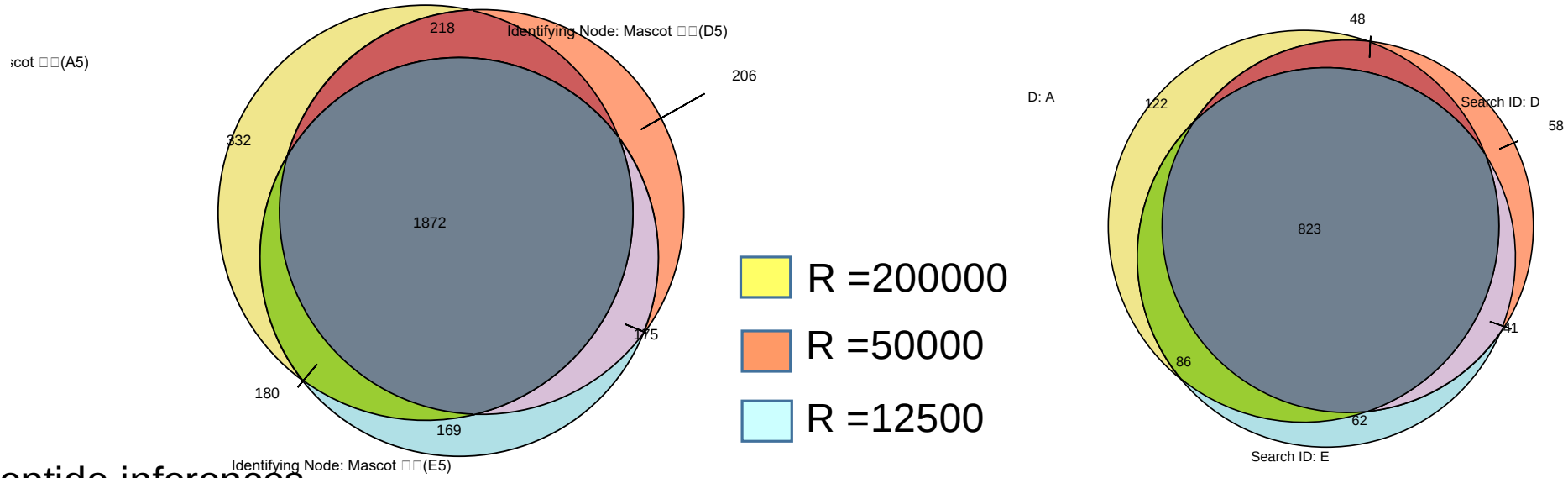
2ppm



R = 200000

Accuracy increases with resolution

Hela digest in 30 min gradient by nanoLC LTQ MS/MS FTICR MS



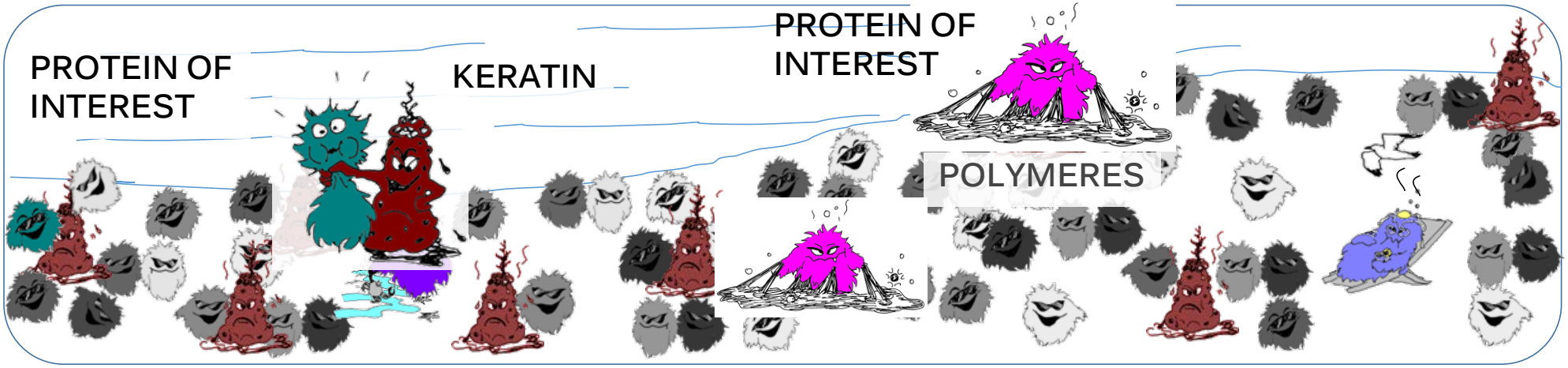
Peptide inferences

Protein inferences

target<5%	10ppm				5ppm			
Reso	Peptides	decoy	FDR	Proteins	Peptides	decoy	FDR	Proteins
12500	4014	87	2.17	825	3636	115	3.16	530
25000	4168	95	2.28	821	4143	108	2.61	531
50000	3856	88	2.28	852	3887	123	3.16	556
100000	3863	110	2.85	844	3918	138	3.52	572
200000	3801	105	2.76	870	3822	125	3.27	600

For the records: 3740 proteins, 19587 peptides in 120 min gradient on a Q Exactive HF

Challenges in LC-MS



© E. Demey, 2009

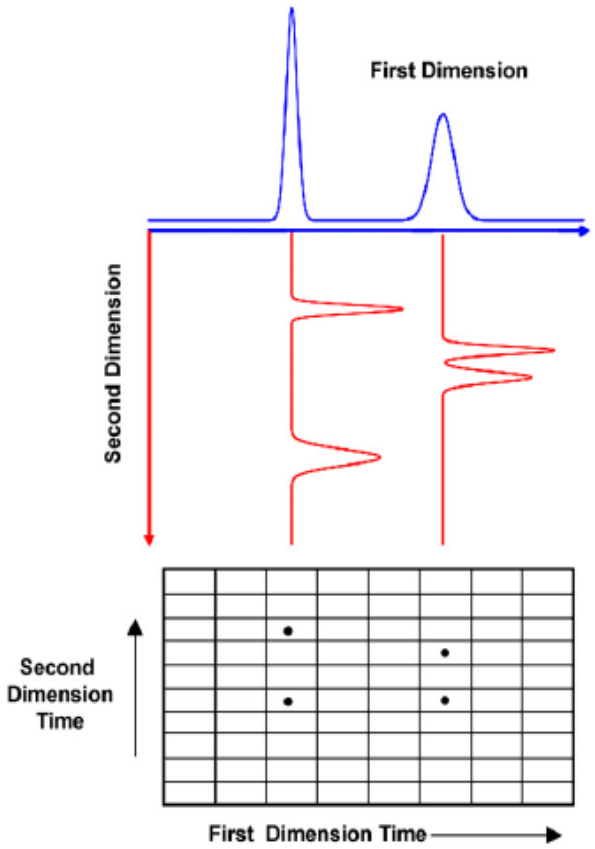
- **A biological sample is always complex !!**
  - Thousands of analytes to identify within a large dynamic range….
- **A biological sample is always contaminated !!**
  - Constitutive engogenous contaminations
  - Exogenous contaminations
- **MS : simple mix**

Solution : Increase peak capacity in MS and in LC

Peak capacity

Max number of resolved compounds using a given analytical method in given analytical conditions.

Giddings, Anal. Chem. 39 (1967)



Stoll *et al.,* J. Chrom. A, 1168 (2007)

Multidimensionnal separation

If 2 orthogonal dimensions

Golden rule: $^{2D}P_c = {}^{1}P_c \times {}^{2}P_c$

↗ peak capacity

↗ dynamic range

Increased separation of peptides

↓

Increased sequenced peptides

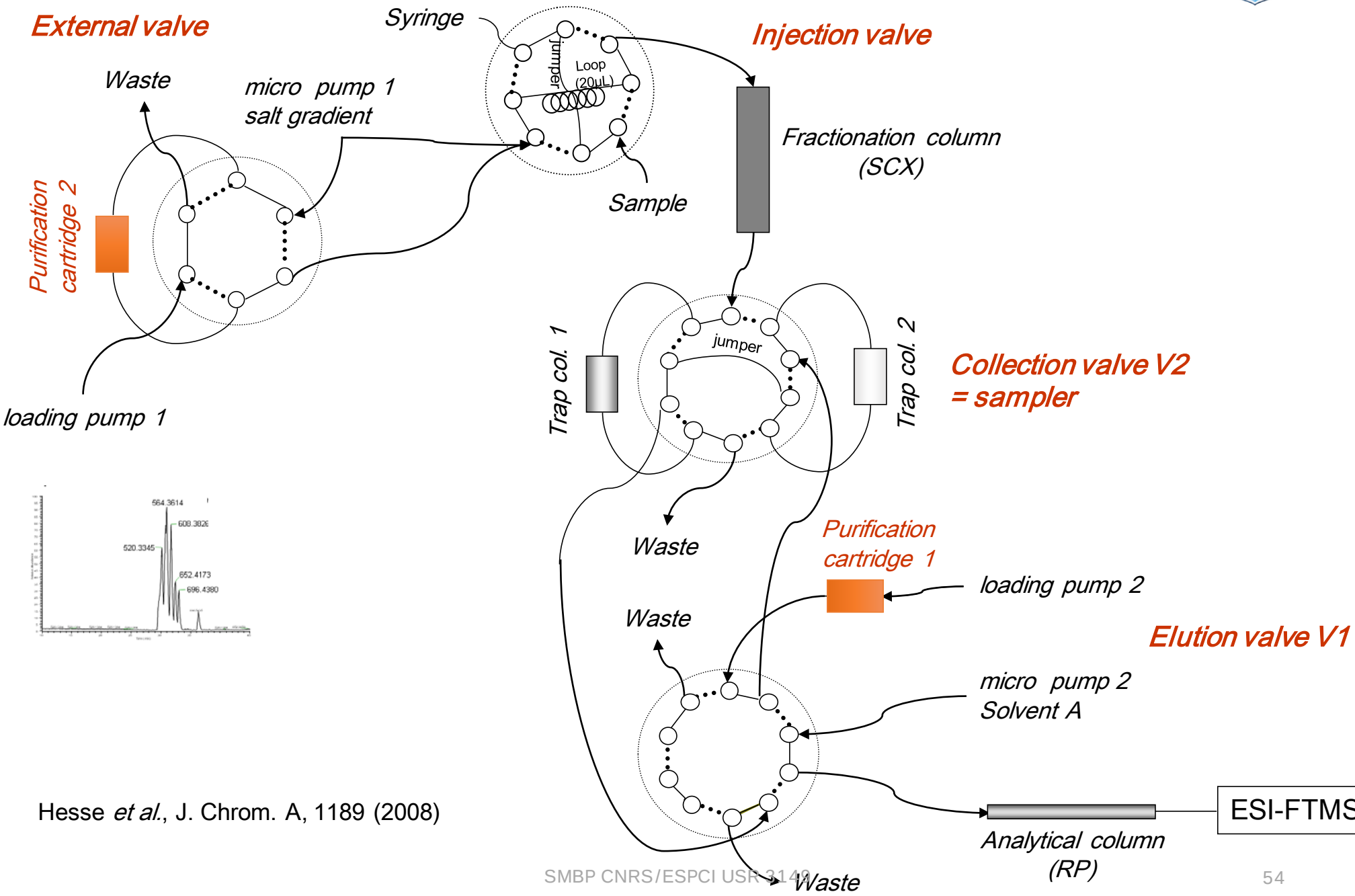
↓

Increased sequence coverage

↓

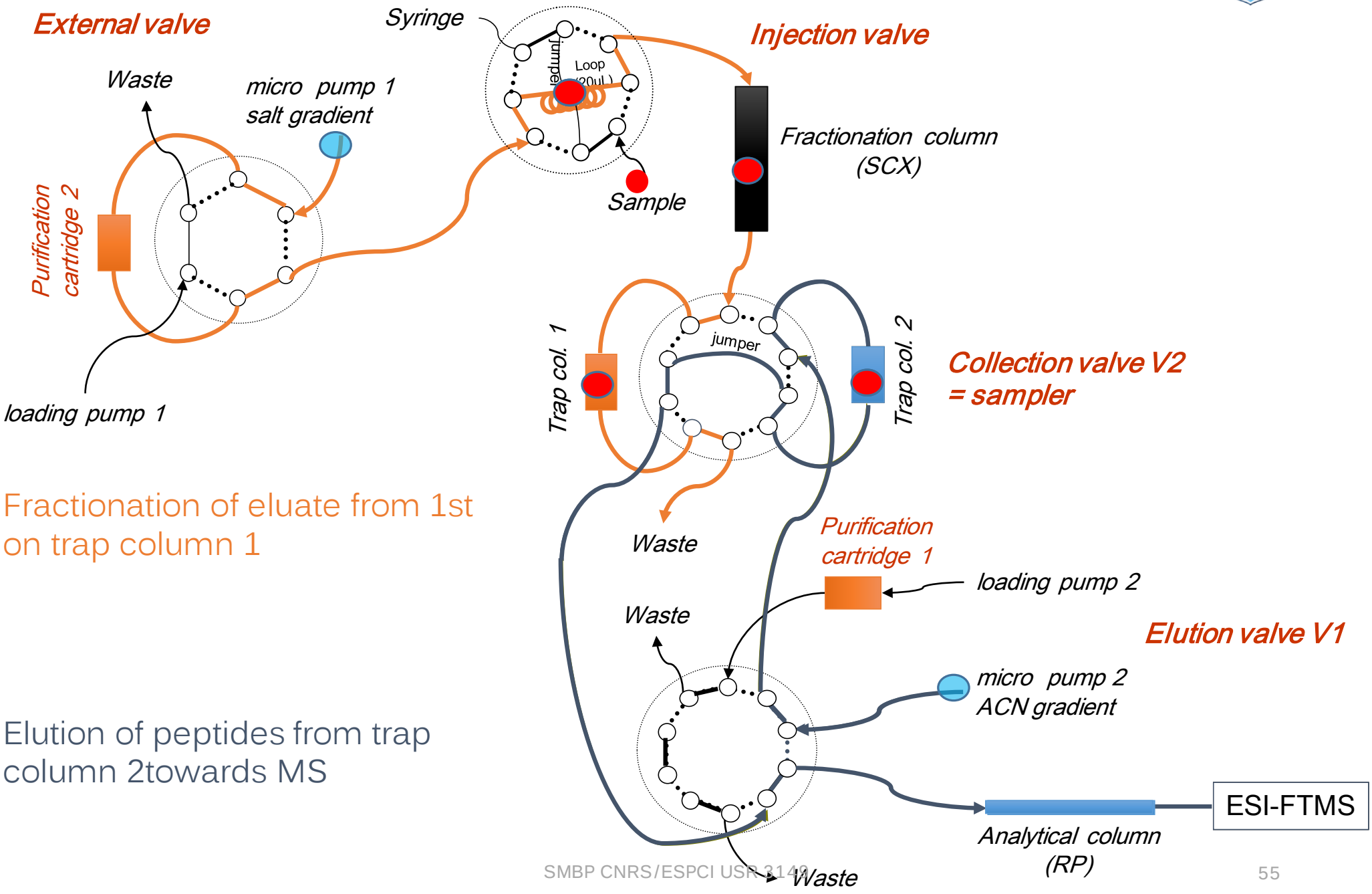
Increased proteome coverage

on-line 2DLC



Hesse *et al.*, J. Chrom. A, 1189 (2008)

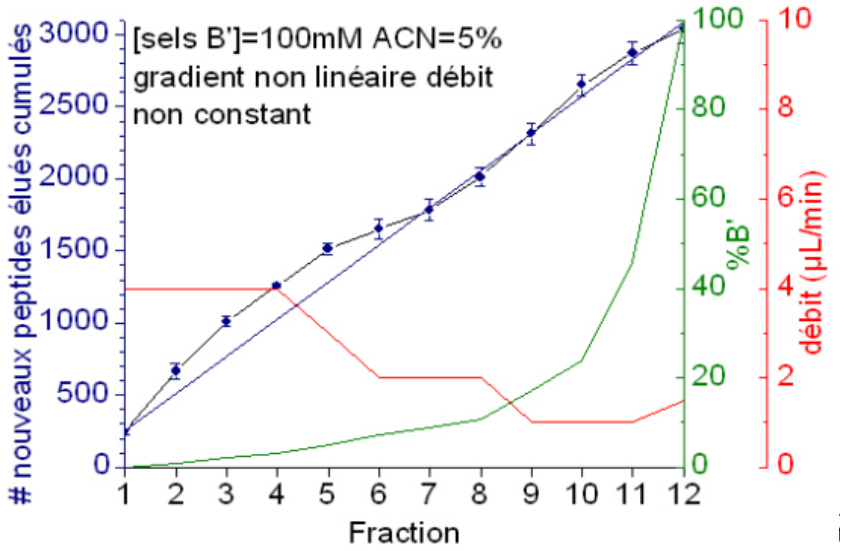
on-line 2DLC

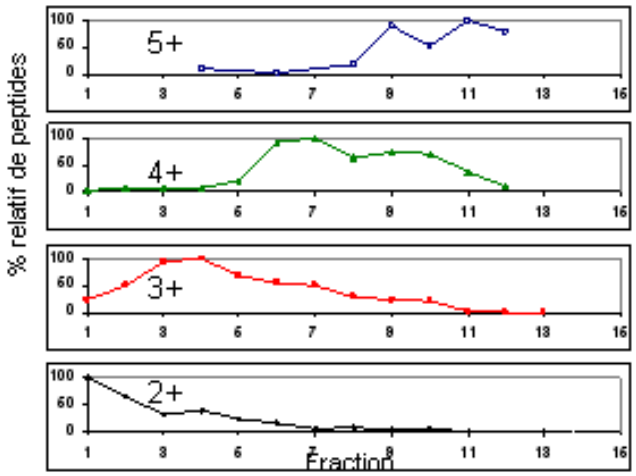


Fractionation of eluate from 1st on trap column 1

Elution of peptides from trap column 2towards MS

on-line 2DLC

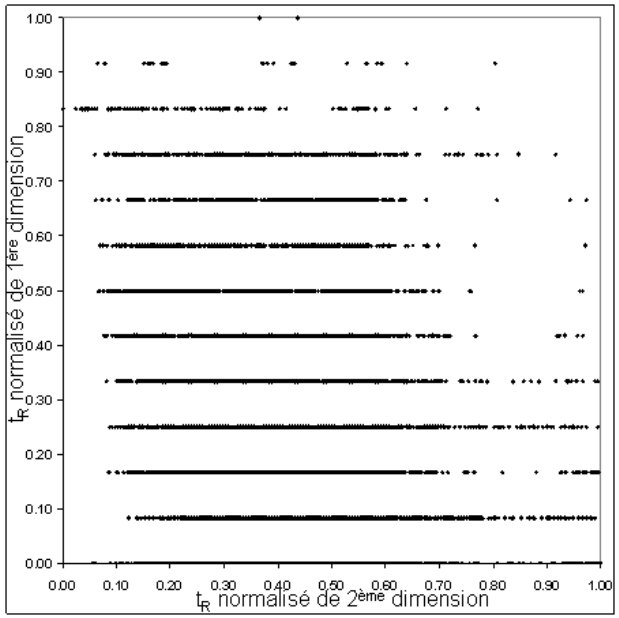




No relationship between retention times of 1st and 2d dimension

↓

Orthogonality



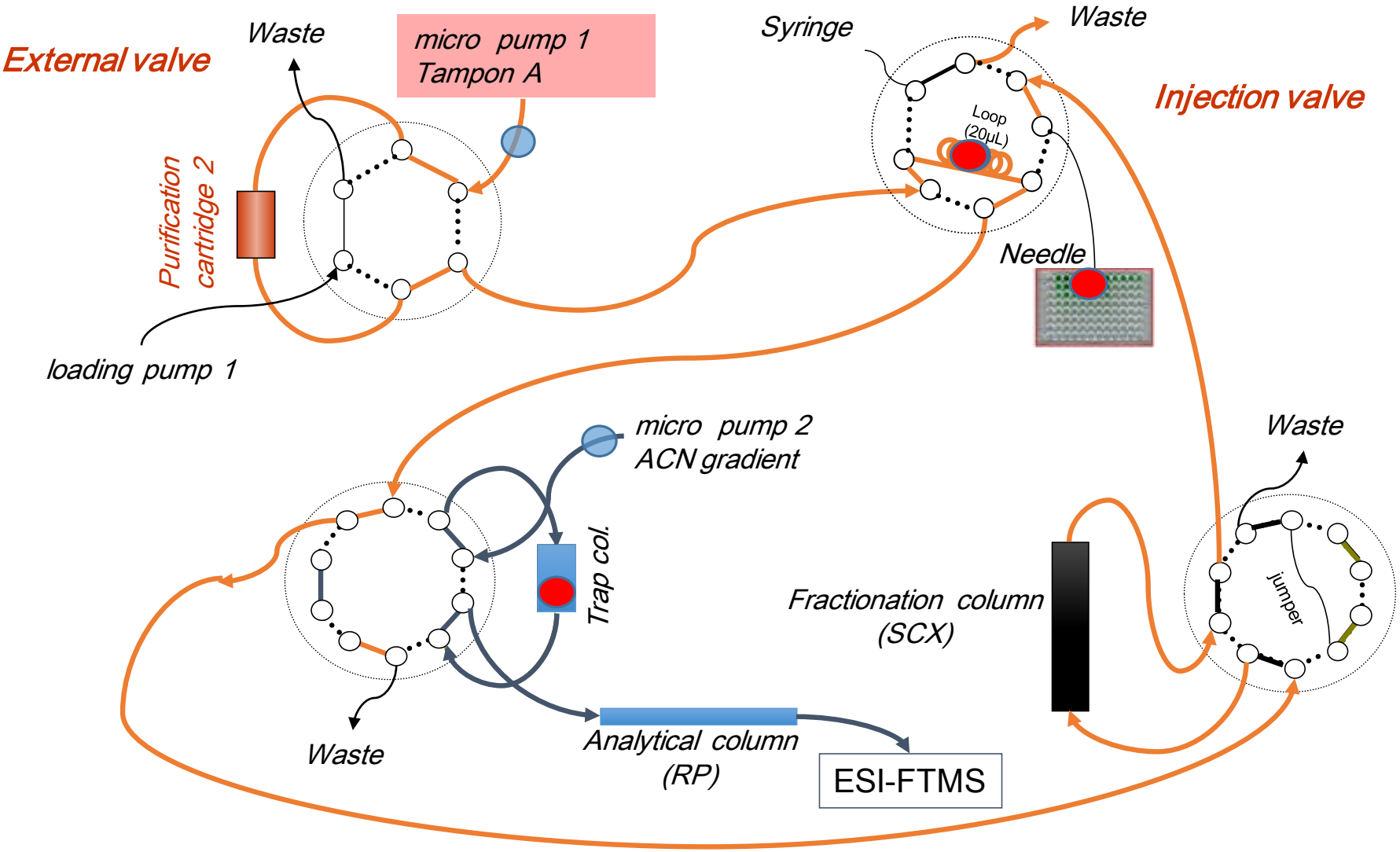
Peak capacity 2DLC x MS/MS

$^{2DMS/MS}P_c = {}^{2D}P_c \times {}^{MS/MS}P_c$

$^{2DMS/MS}P_c = {}^{1}P_c \times {}^{2}P_c \times {}^{MS/MS}P_c$

$^{2DMS/MS}P_c = 12 \times 80 \times 20 = 18000$

off-line 2DLC

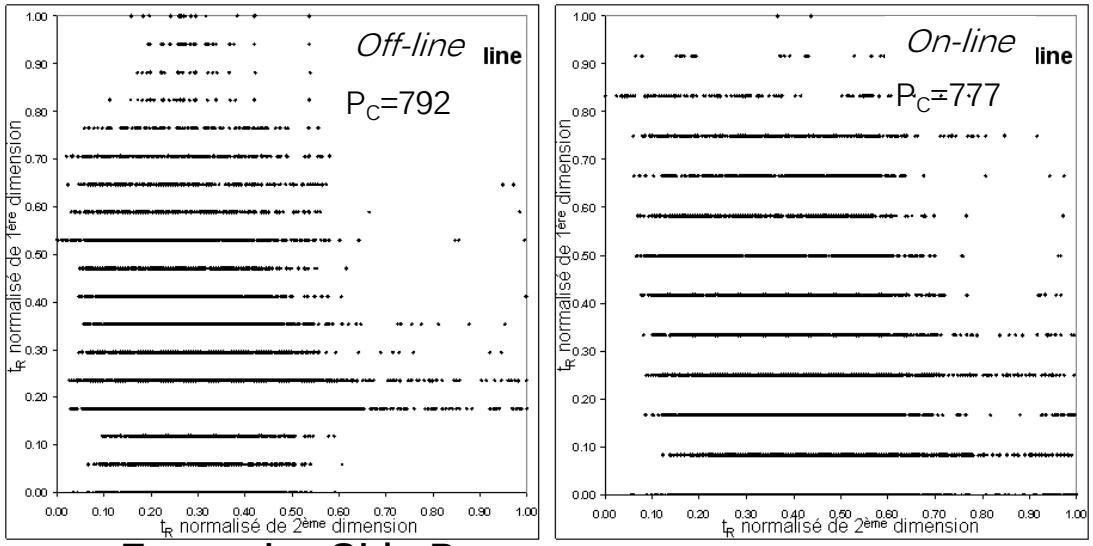


Fractionation of 1ère dimension eluate

Elution of peptides
From fractions towards MS

2DLC : off-line *vs* on-line

Bias against hydrophobic peptides



Bias against hydrophilic peptides

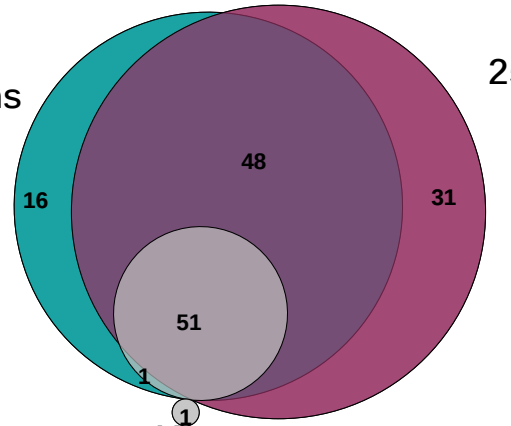
Example : Skin Proteome

2D Off-line

33% of Keratins

2D On-line

25% of Keratins



1D

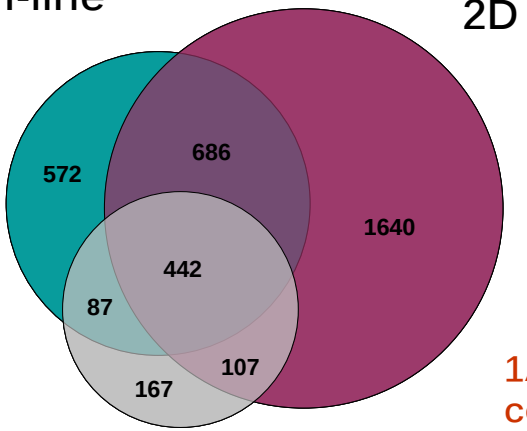
60% of Keratins

Proteins inferences

2/3 de protéines communes

2D Off-line

2D On-line



1D

Peptides inferences

1/3 de peptides communs

Alternatives

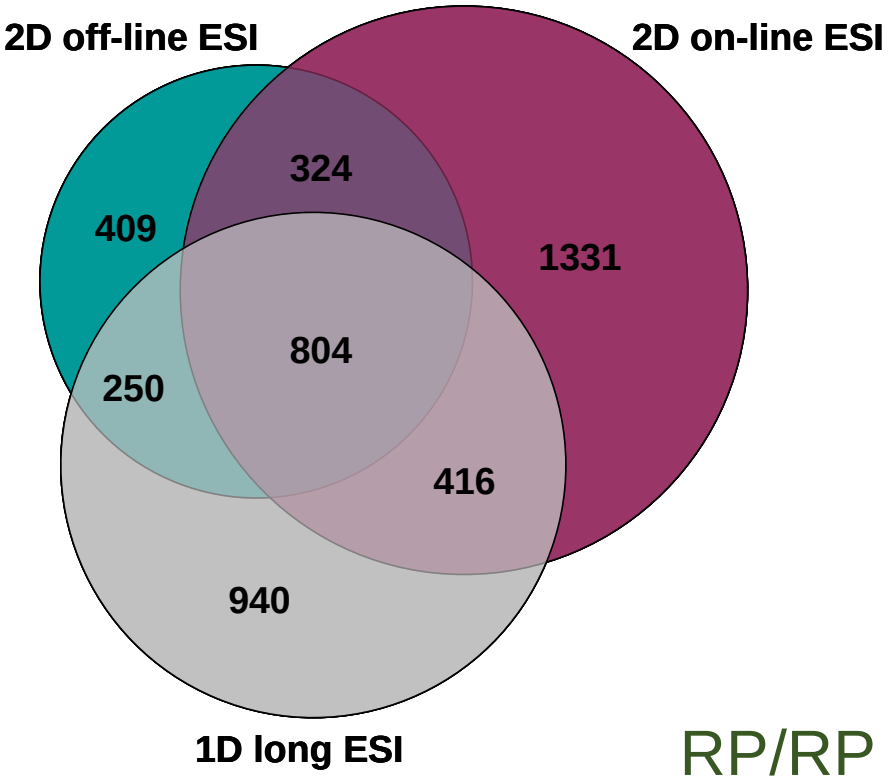
Gradient 1DLC plus long

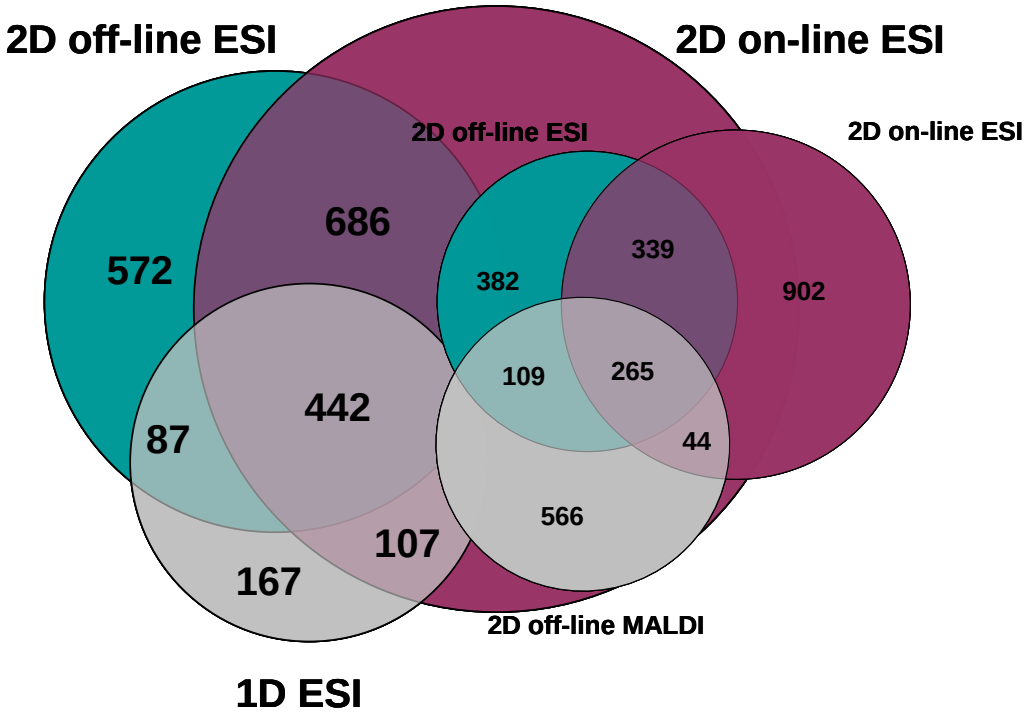
☺ simple setup
☺ Robust Separation
☹ Large peaks

Detection MALDI-MS

☺ Different Ionization Mode
☺ Separation time stopped
☹ Longer acquisition time

Complementarity of analytical strategies





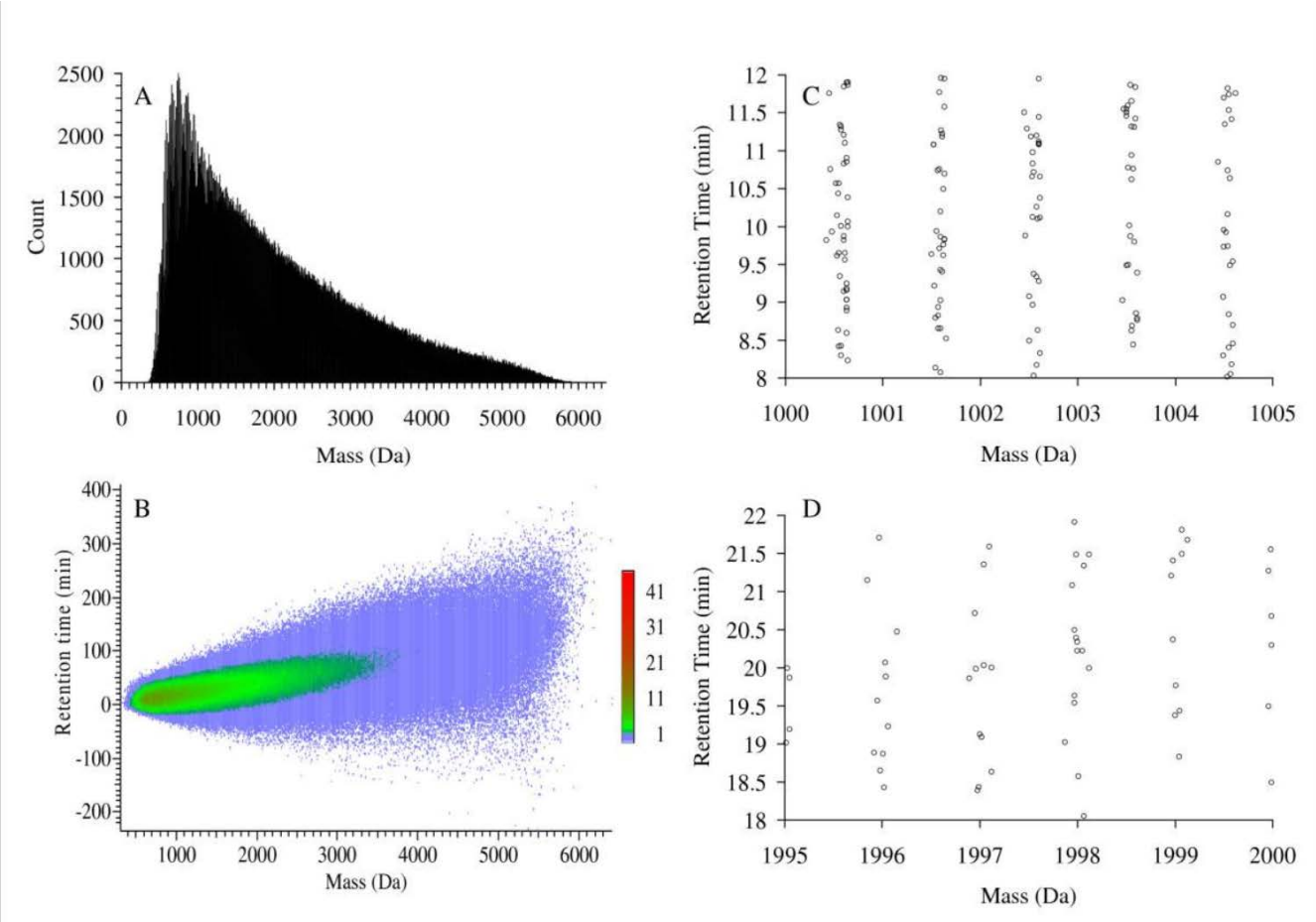
RP/RP MS/MS is also a good solution

Summary

- Omics and Mass spectrometry
- Why study the proteome?
- Bottom-up proteomics: an LC-MS story
- **DIA and HRM**
- Quantification

Co-eluted peptides

Alves G, Ogurtsov A, Kwok S *et al.* Detection of co-eluted peptides using database search methods. *Biol. Direct*3(1), 1–16 (2008).



In most cases the most intense peptides are selected in DDA MS/MS

Minor co-eluted are ignored and information is lost

Example of 9393 proteins from Saccharomyces Cerevisiae after tryptic digestion (2 miscleav. 771753 peptides)

Other analytical strategies

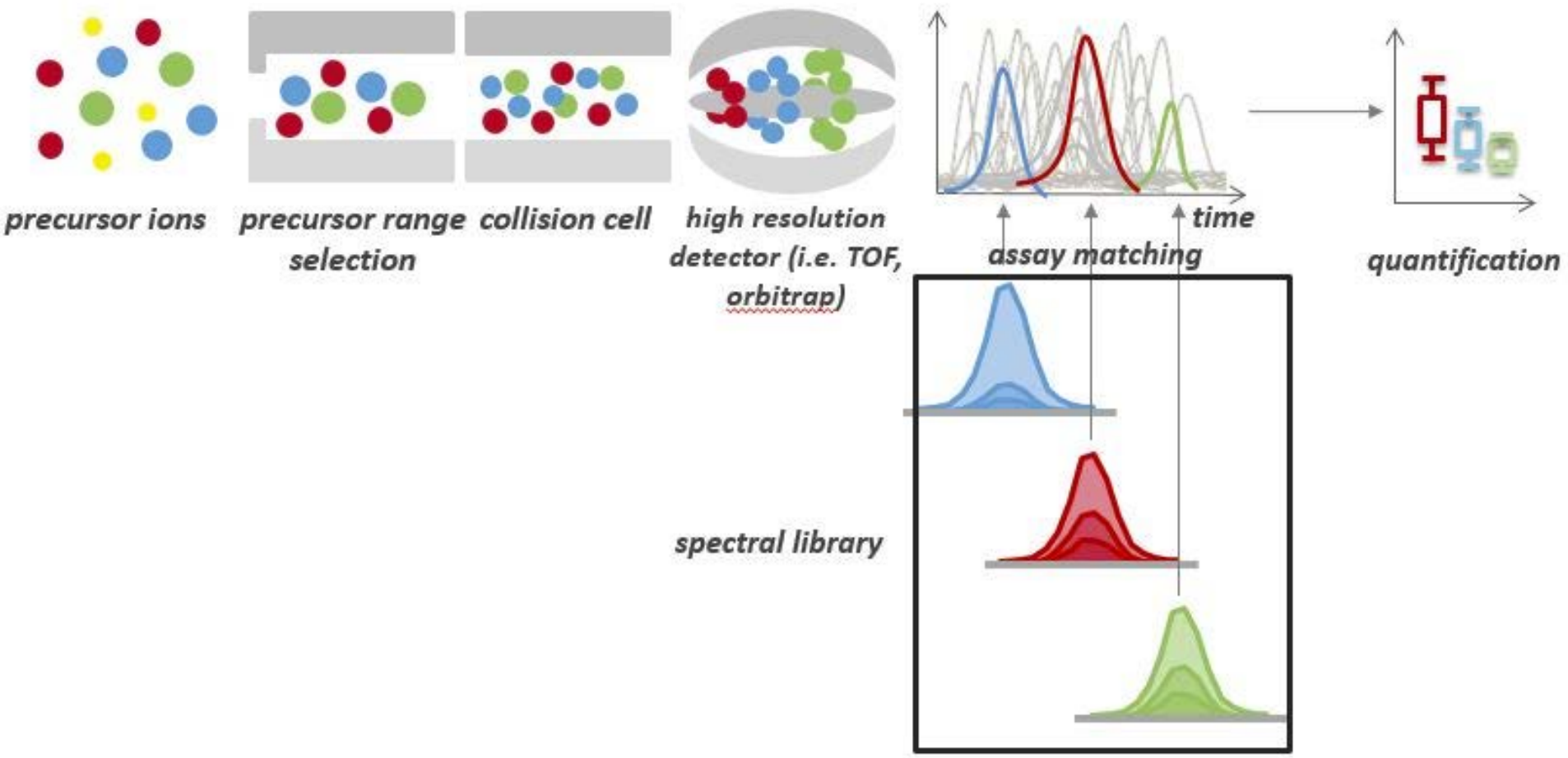
DIA: Data Independant Acquisition

- SWATH™
- MS^E
- AIF (All ion Fragmentation)

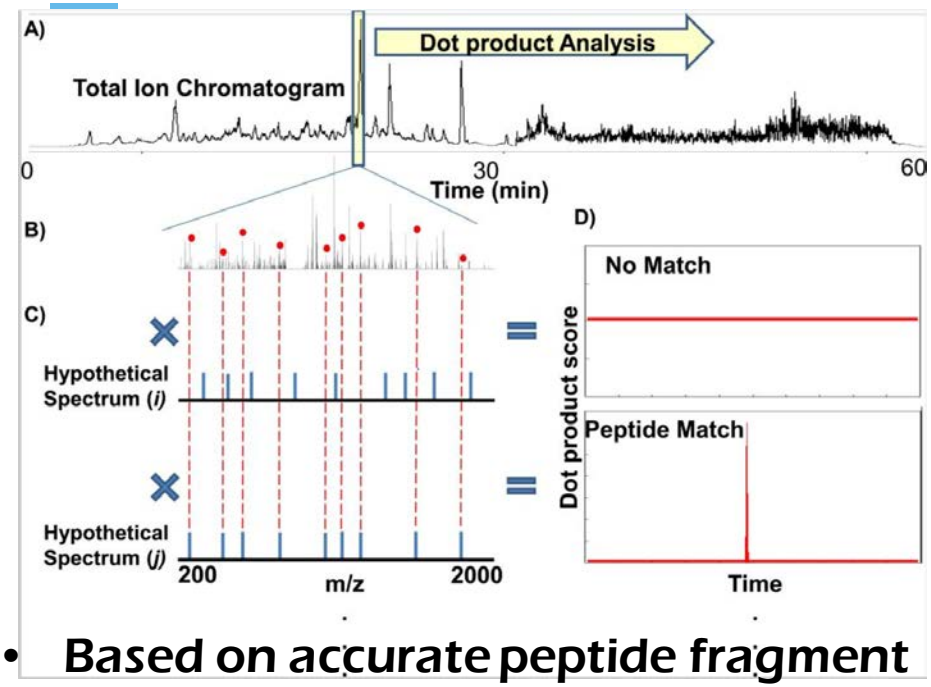
HRM: Hyper Reaction Monitoring

- Multiplex SRM (Selected Reaction Monitoring)
- Pseudo SRM
- PRM (Parallel reaction monitoring)

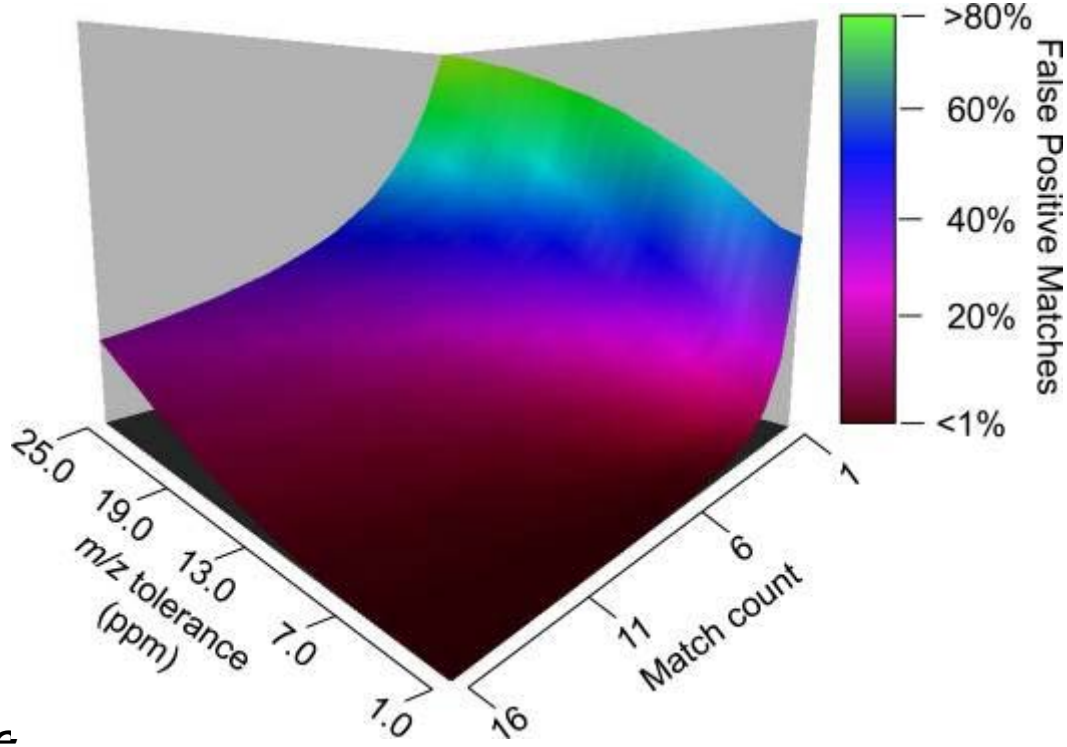
Data Independent Acquisition (DIA)



Fourier transform-all ion reaction monitoring FT ARM



Weisbrod CR, Eng JK, Hoopmann MR, Baker T, Bruce JE. J Proteome Res. Mar 2, 2012; 11(3): 1621–1632.



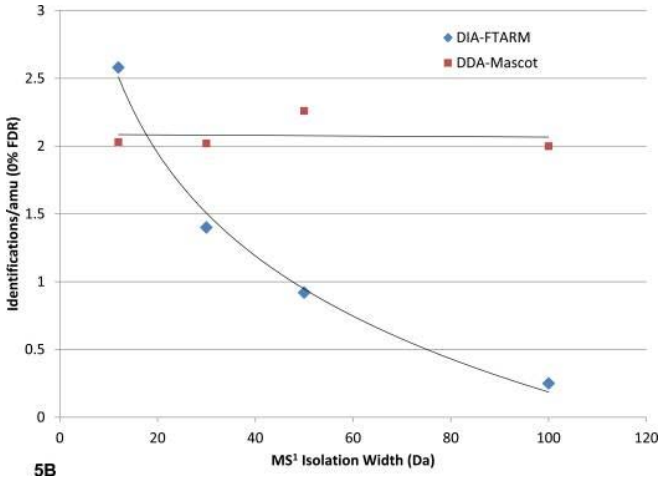
- Based on accurate peptide fragment mass measurements
- All ions fragmented in every scan

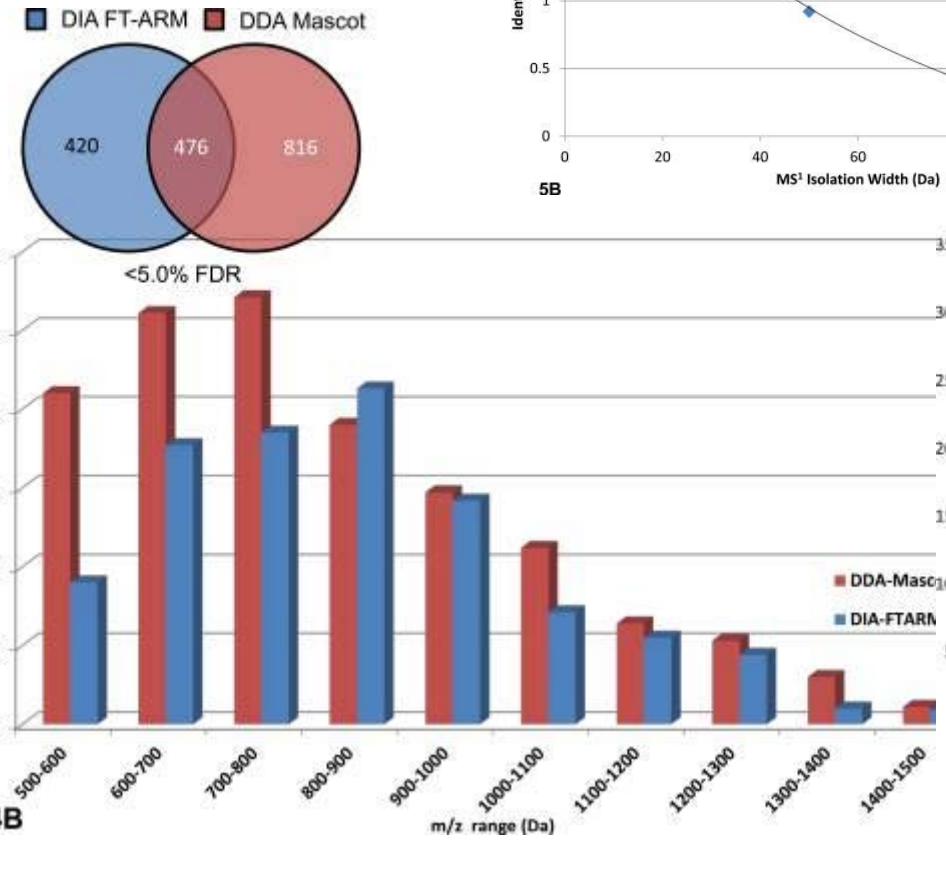
A) TIC B) Complex fragmentation spectrum of all ions. C) Hypothetical fragmentation spectrum of one peptide D) Dot product of the 2 spectra, and *Score Chromatogram*.

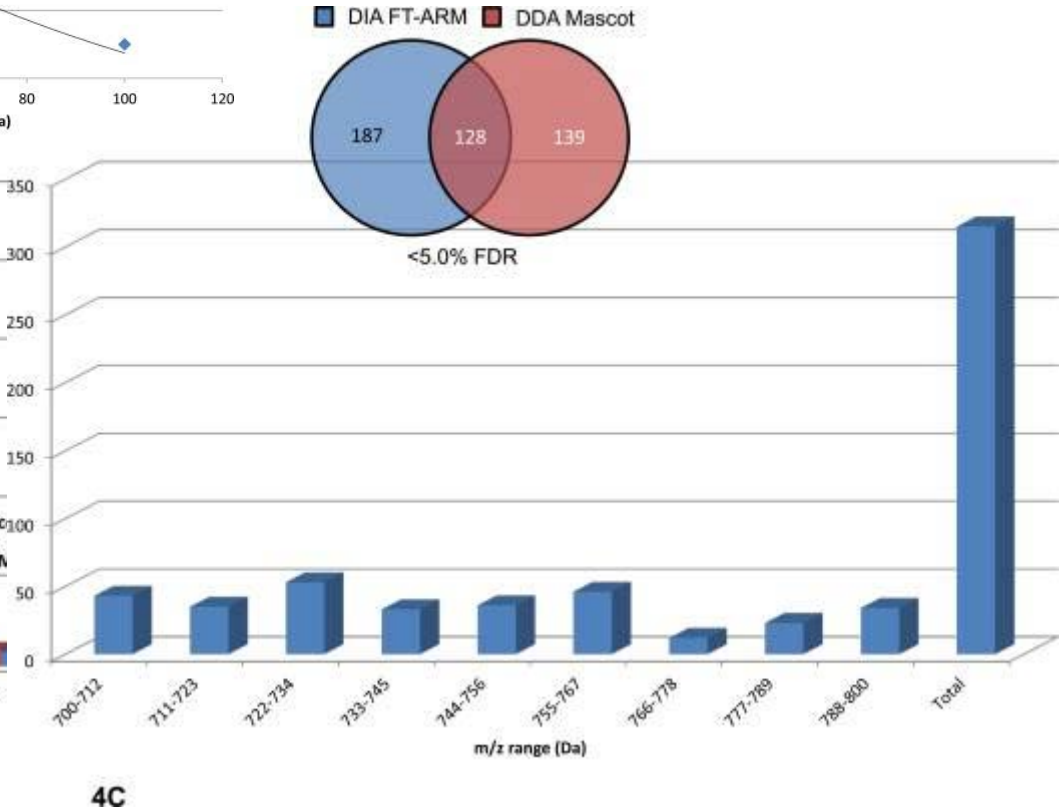
Fourier transform-all ion reaction monitoring FT ARM

Weisbrod CR, Eng JK, Hoopmann MR, Baker T, Bruce JE. J Proteome Res. Mar 2, 2012; 11(3): 1621–1632.

http://brucelab.gs.washington.edu/

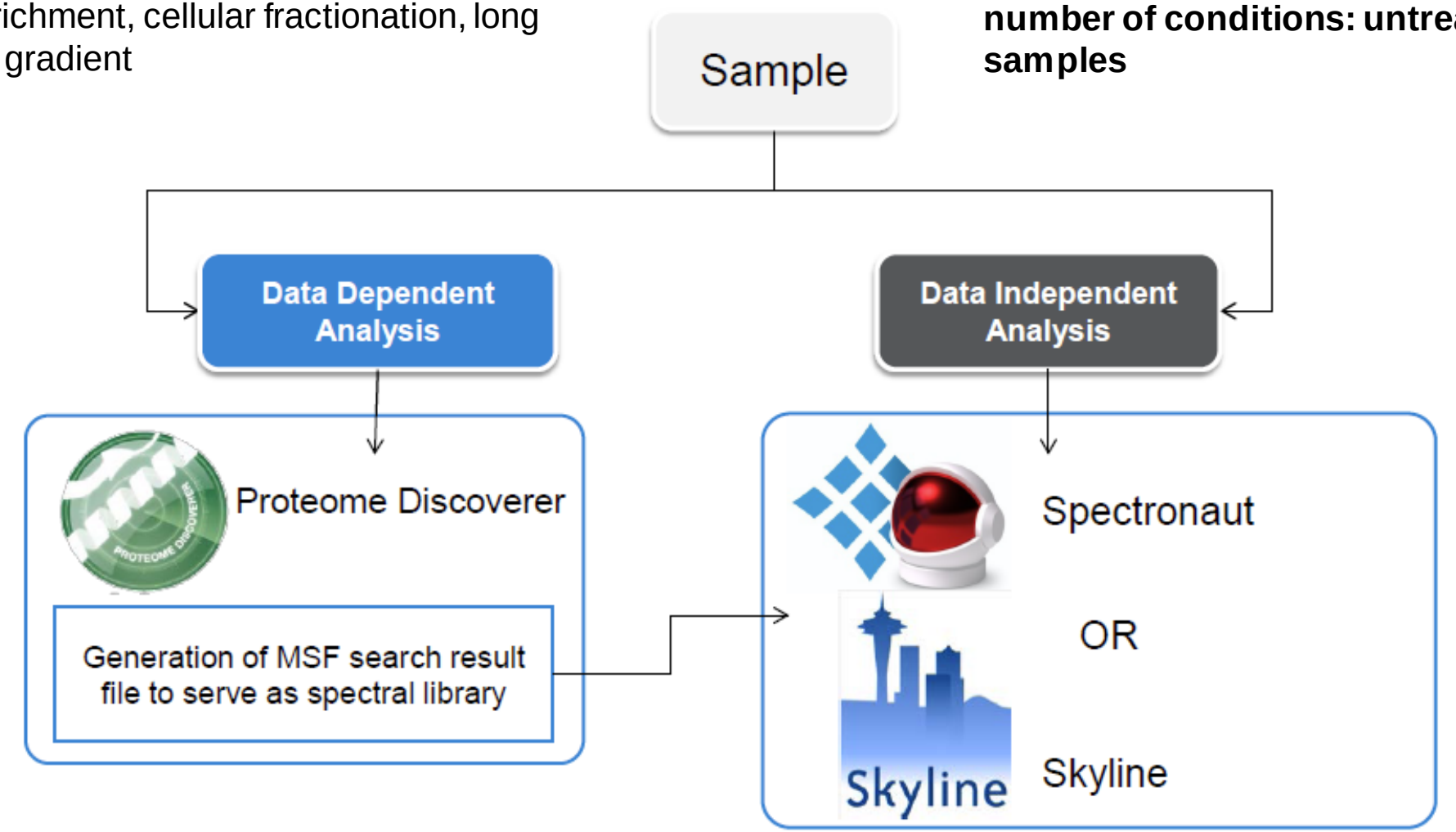




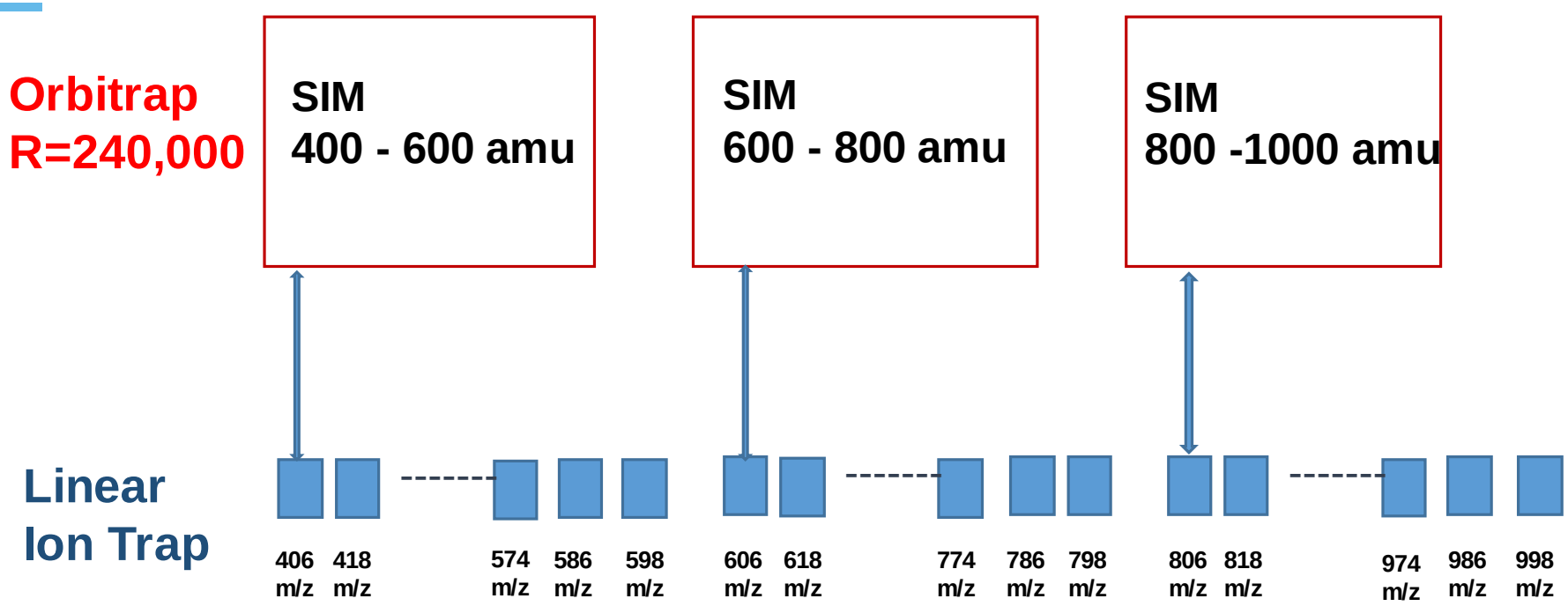


Spectral library: 2D-LC-MSMS, enrichment, cellular fractionation, long LC gradient

Quantitative analysis with a large number of conditions: untreated samples



SIM/tMS² for Data Independent Protein Quantification



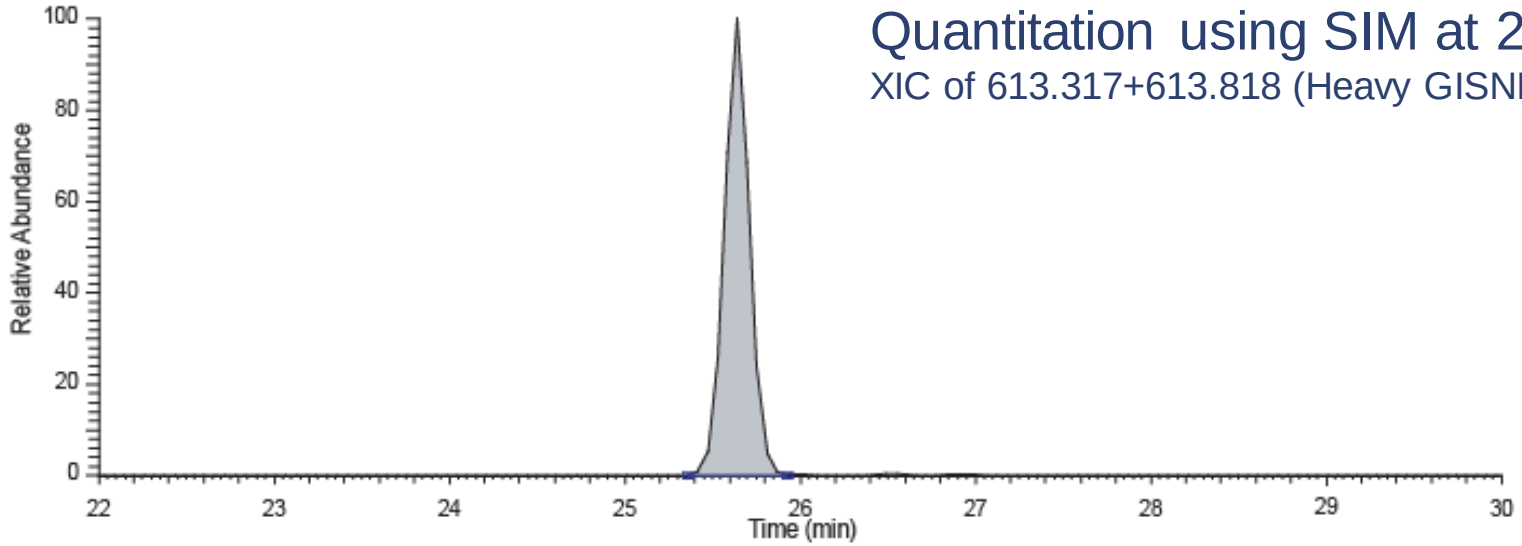
17 sequential CID MS/MS scans with 12 Da isolation Windows

17 sequential CID MS/MS scans with 12 Da isolation Windows

17 sequential CID MS/MS scans with 12 Da isolation Windows

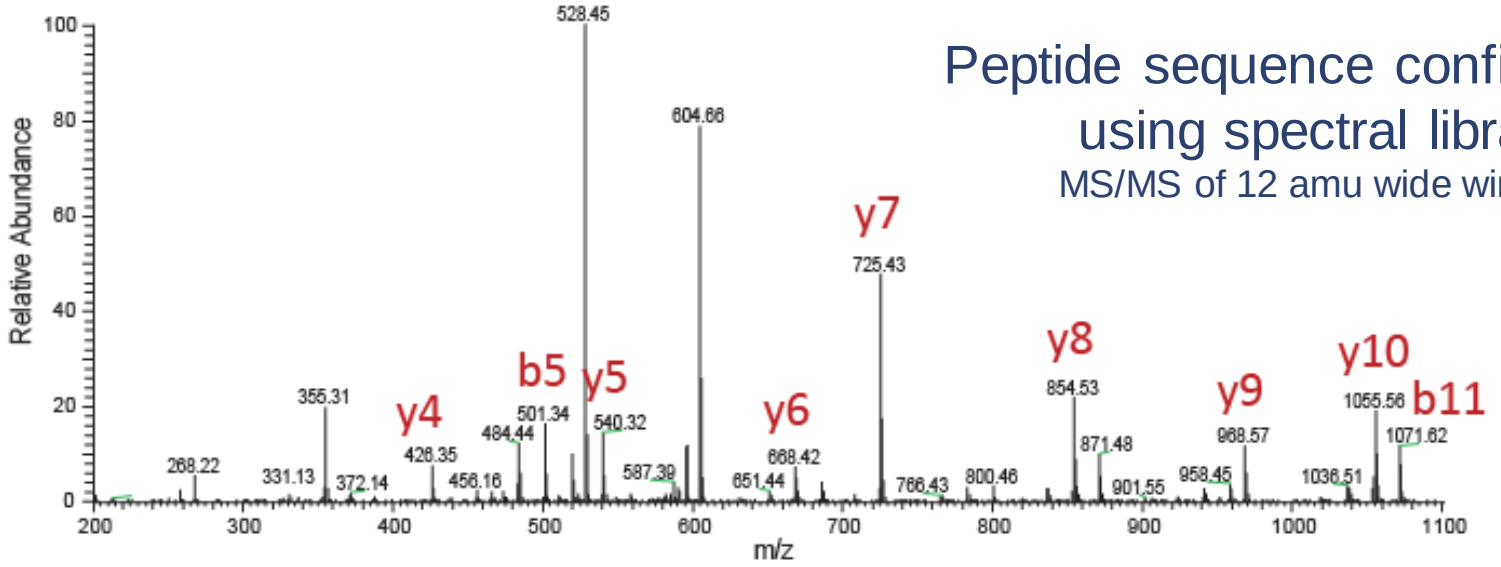
3.6 s cycles, covering 1000 m/z

DIA Workflow on Orbitrap Fusion Tribrid



Quantitation using SIM at 240K

XIC of 613.317+613.818 (Heavy GISNEGQNASIK)

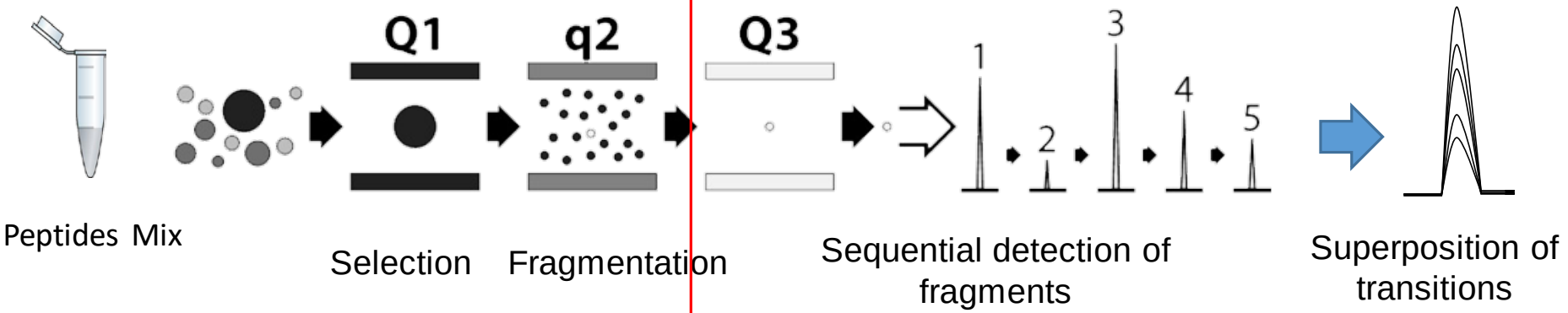


Peptide sequence confirmation using spectral library

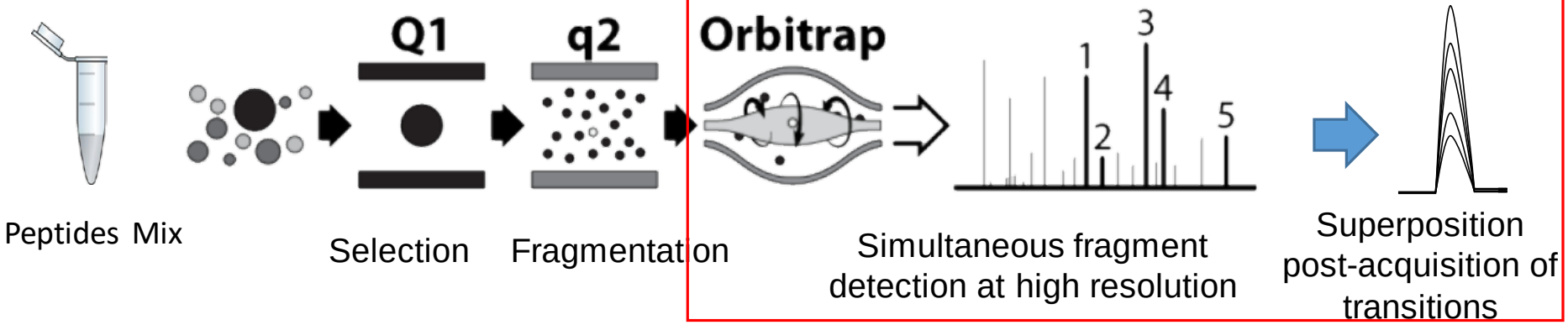
MS/MS of 12 amu wide window

PRM vs SRM

SRM « *Selected Reaction Monitoring* » (QqQ)



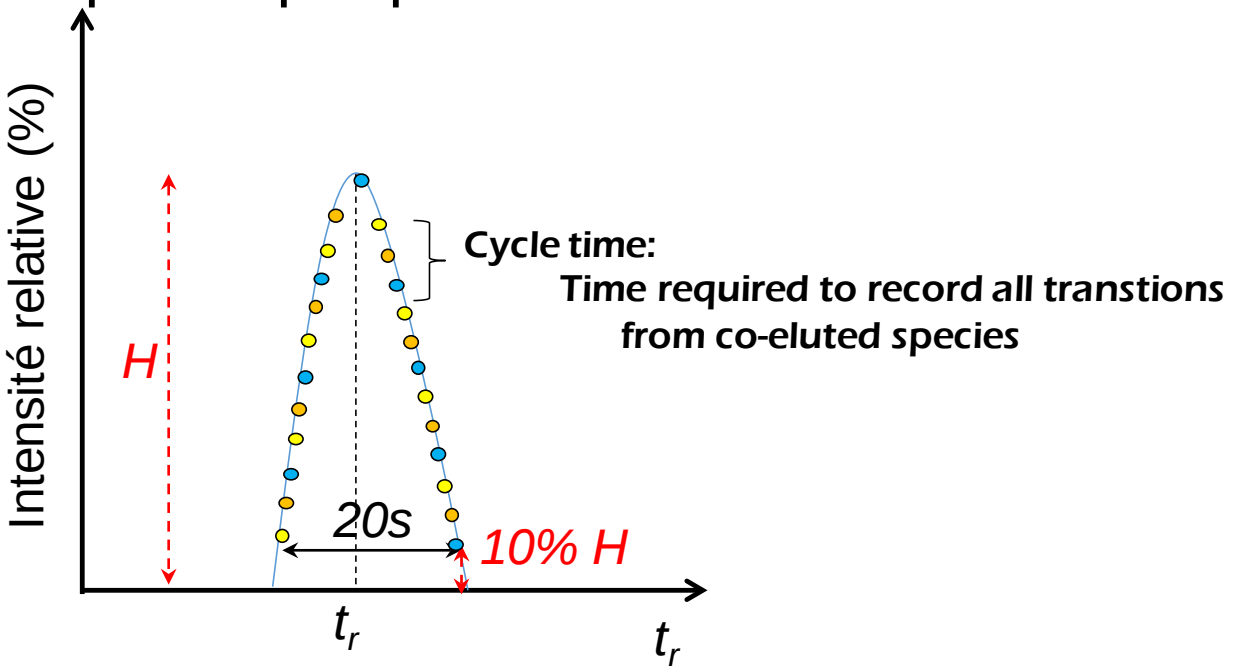
PRM « *Parallel Reaction Monitoring* » (QqOrbitrap)



Peterson A.C. et al., Mol Cell Proteomics. 2012 Nov;11(11):1475-88

Targeted analysis: sample sampling

minimum 8 - 10 points per peak



$$\#peptides = \frac{temps\ de\ cycle}{temps\ d'acquisition}$$

➔Up to 31 peptides / cycle
(R=17 500 at m/z 200, 64ms/acquisition)

Q Exactive (Thermo Scientific)

The case of amyloidoses

Sophie Liuu, Emmanuelle Demey
Gilles Grateau, David Buob (Hop Tenon)

Group of diseases

Extracellular deposit (aggregate) of insoluble misfolded proteins

Hereditary (genetic mutation) or senile (aging)

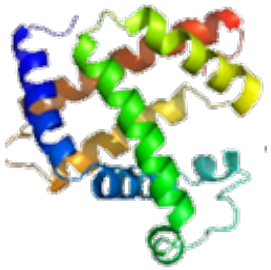
Life-threatening organ failure

ex: myocardial wall thickening

Complex mechanisms

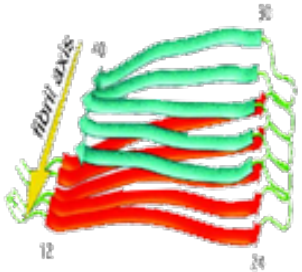
Common components:

- Glycosaminoglycans
- Serum amyloid P-component (SAMP)
- Apolipoprotein E (APOE)

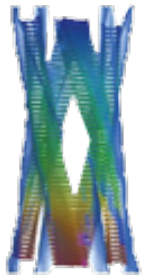


Native precursor protein

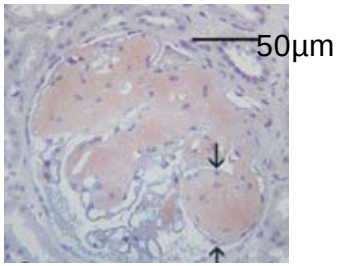
ex: Serum amyloïd A (SAA)



β-sheets rich structure



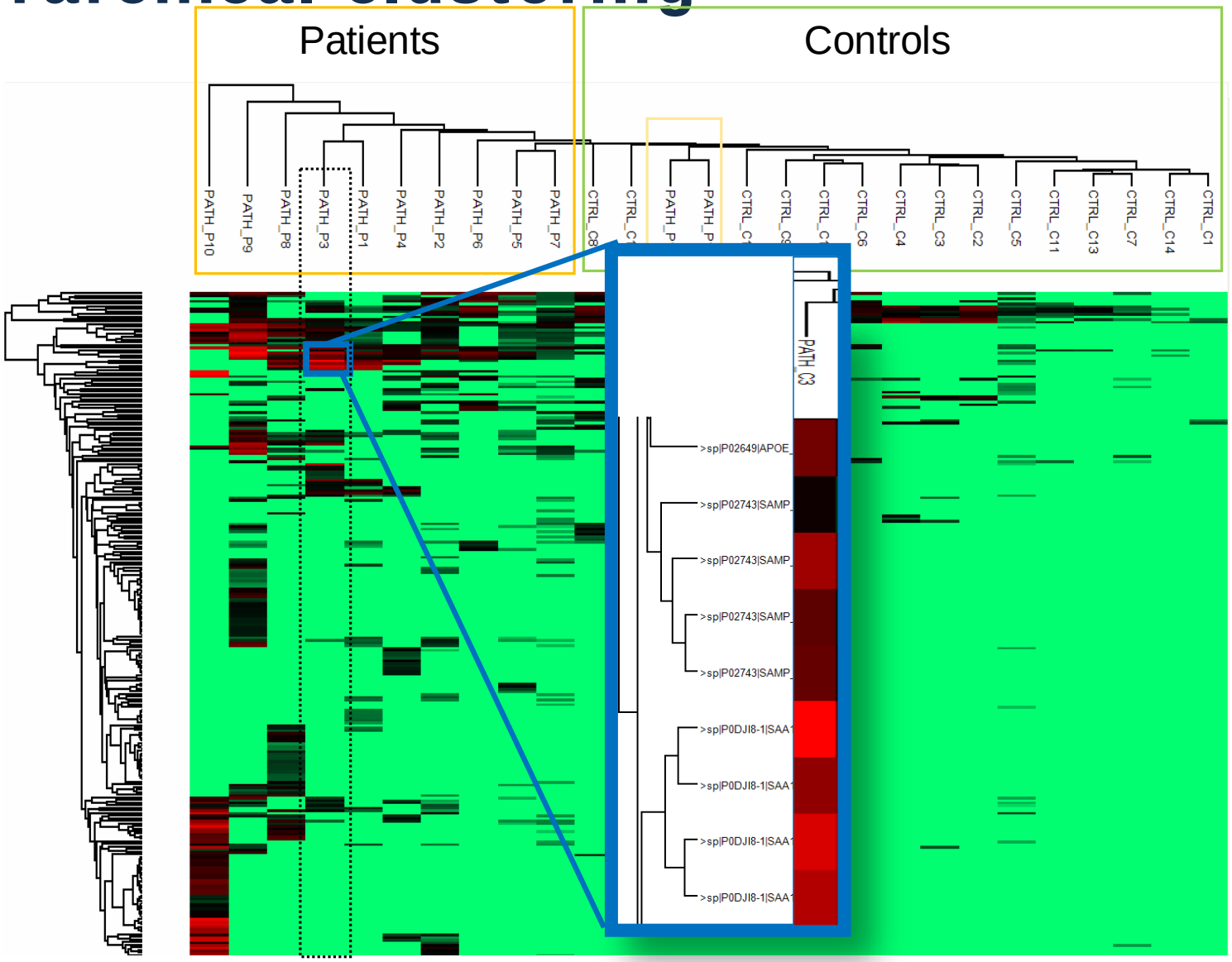
Amyloïd fibrills



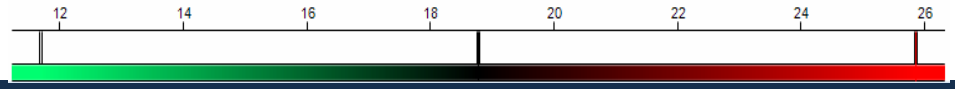
Amyloid deposits

Kisilevsky R., Amyloid. 2000 ,7(1):23-5
Grateau G., Médecine/Sciences 2005 ; 21 : 627-33
Ren R. et al. J Bio Chem, 2010, 285 (4) 37672-82

DDA MS/MS mode : automated hierarchical clustering



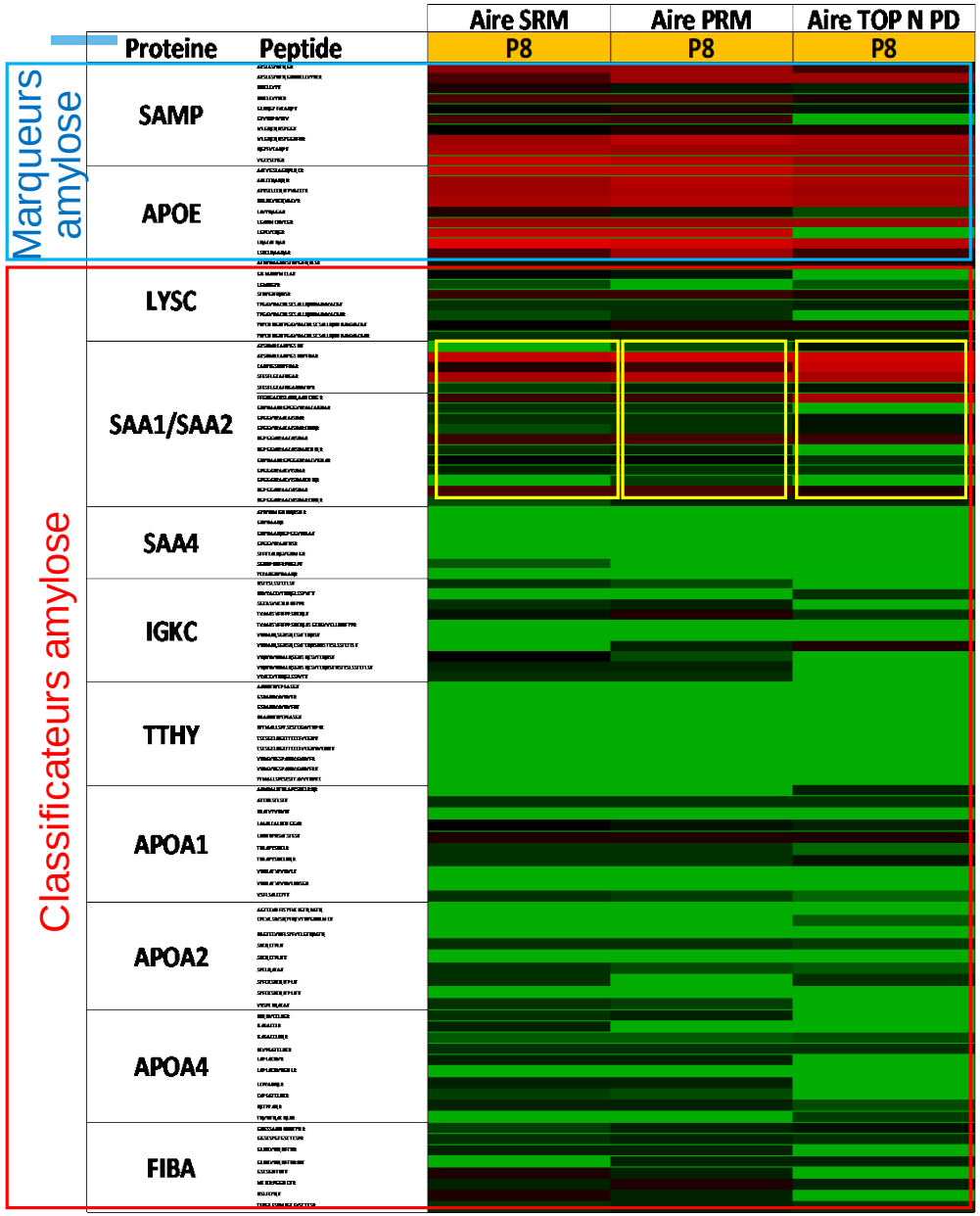
Intensity scale of peptides (area)

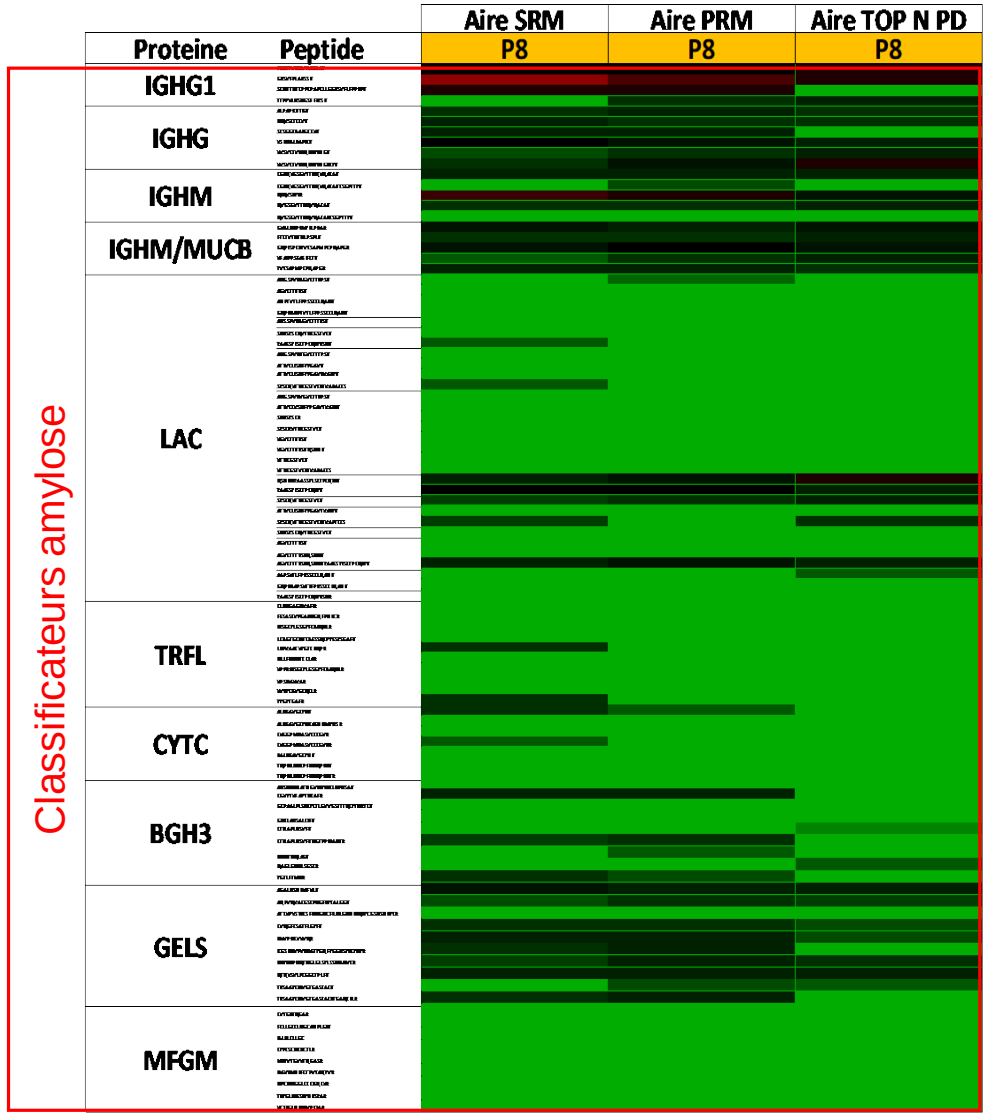


Data processing Maxquant/Perseus

3 technical replicates

Targeted vs non targeted analysis





Intensity scale of peptides (area)

1 E - 6 — 1

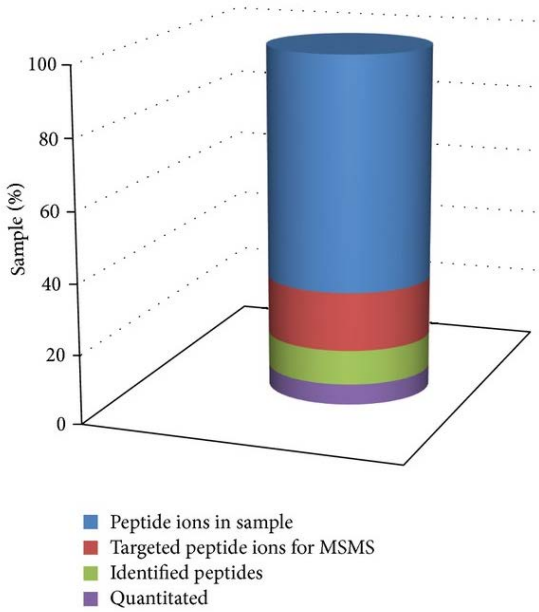
Summary

- Omics and Mass spectrometry
- Why study the proteome?
- Bottom-up proteomics: an LC-MS story
- DIA and HRM
- **Quantification**

Why Quantitative Proteomics ?

Most biological issues cannot be solved by the identification of one protein.

In most cases the variation of the abundance of a protein together with its modification state is required.



Wasinger VC, Zeng M, Yau Y. Int J Proteomics. 2013;2013:180605.

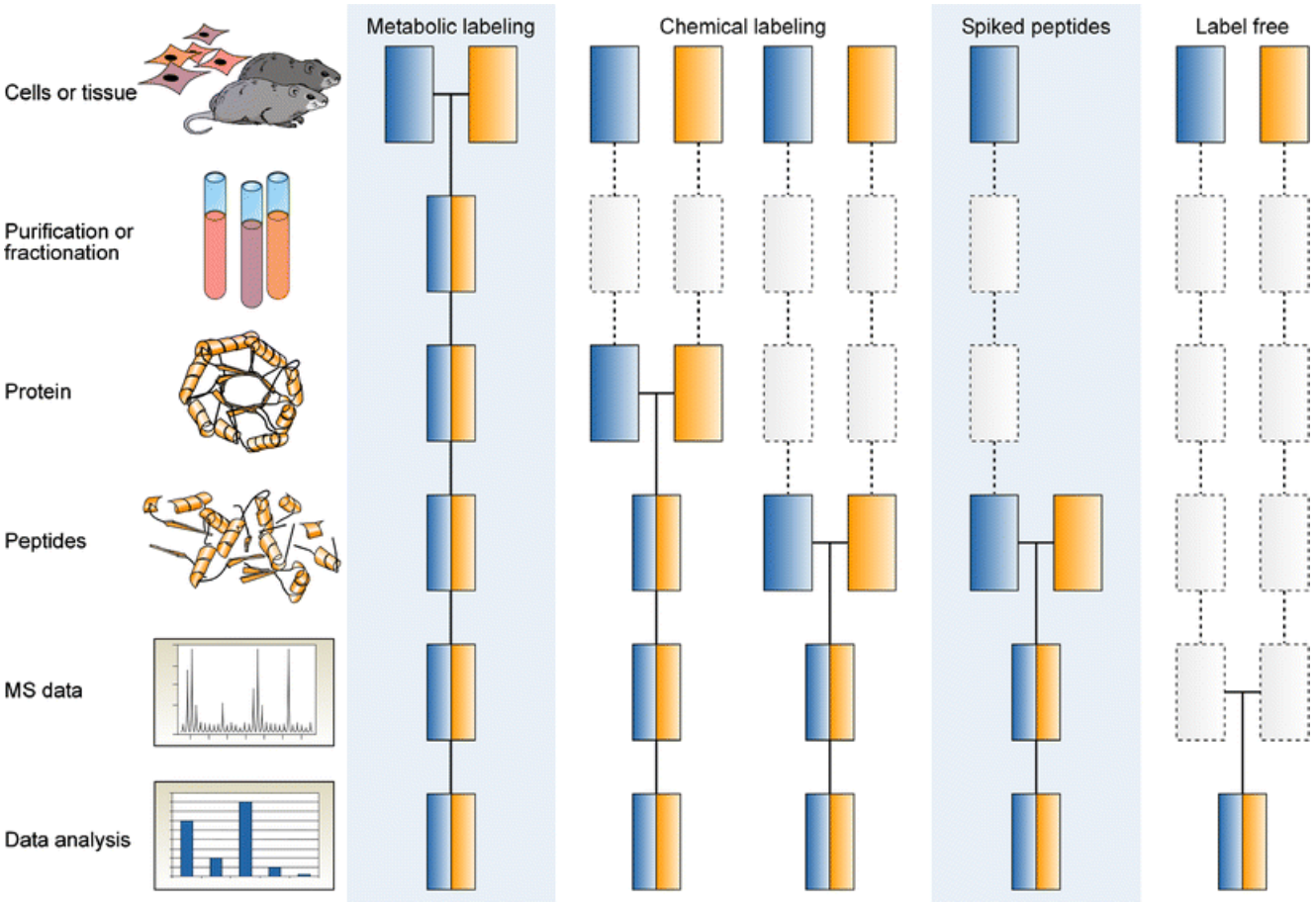
Mass spectrometry is not quantitative

- Due to non uniform instrumental response mass spectrometry cannot be used directly for quantitation of macrobiomolecules
- The intensity of a peptide is not a direct mesure of its abundance
- Comparison of chemically identical species :
    - Two peptides with a different isotopic composition within one single run:

      Isotopic labelling
    - The same peptide detected in 2 different runs:

      Label free quantification

Quantification bottom-up



- Technological limitations to detect and differentiate the isotopic labels
- Bioinformatics and statistics limitation to process the data with a correct normalization

S-E Ong & M Mann; Nature Chemical Biology 1, 252 - 262 (2005)
M Bantscheff, M Schirle, G Sweetman, J Rick & B Kuster; Anal Bioanal Chem 389(4), 1017-1031 (2007)

In vivo Metabolic Labeling Stable Isotope Labeling by Amino Acid in Cell Culture (SILAC)

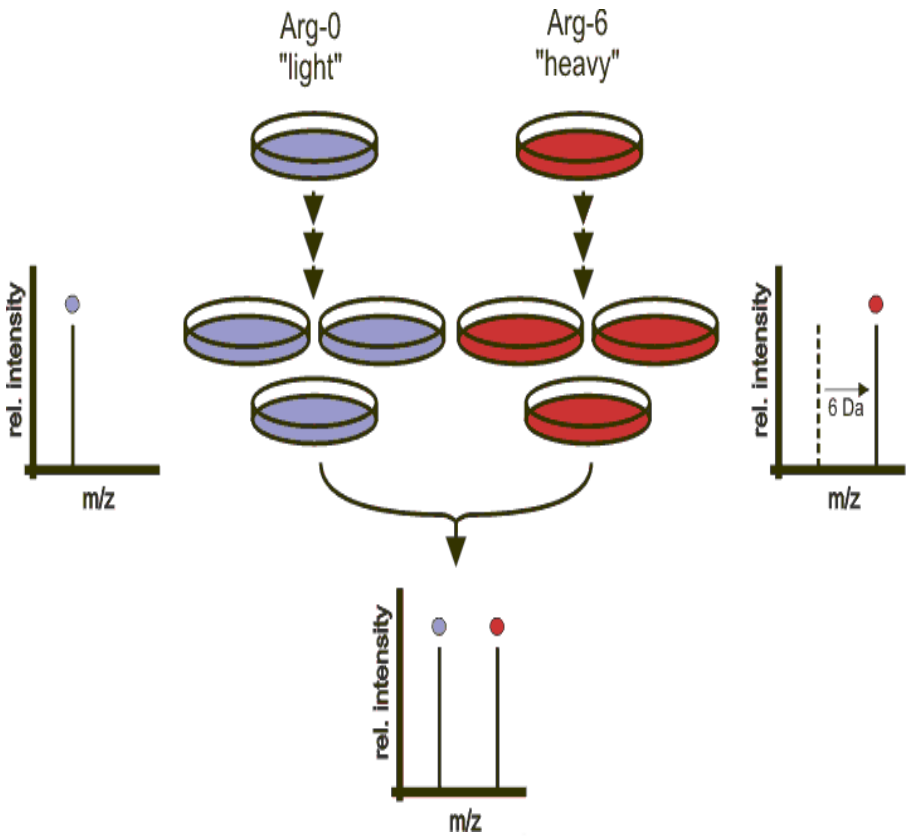
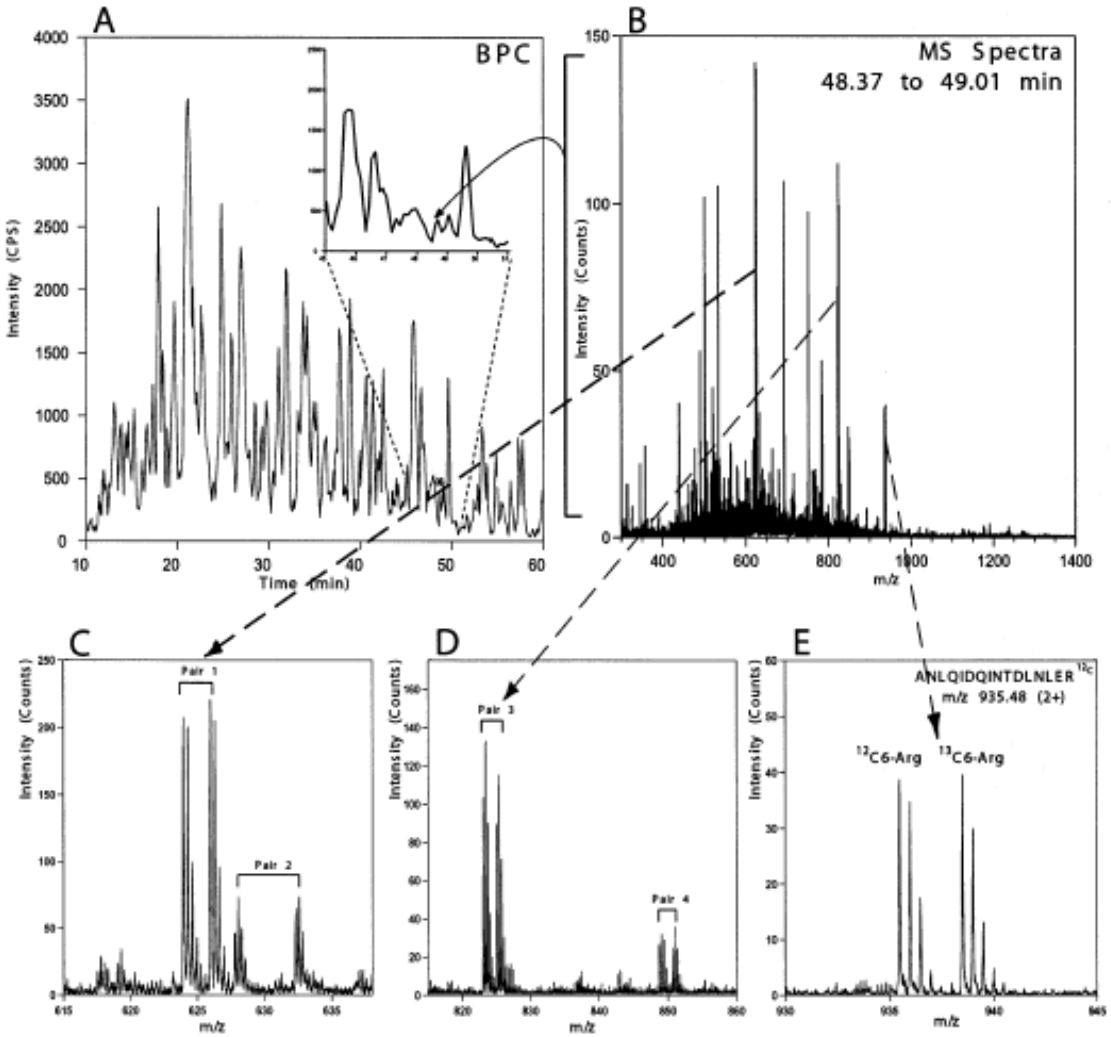
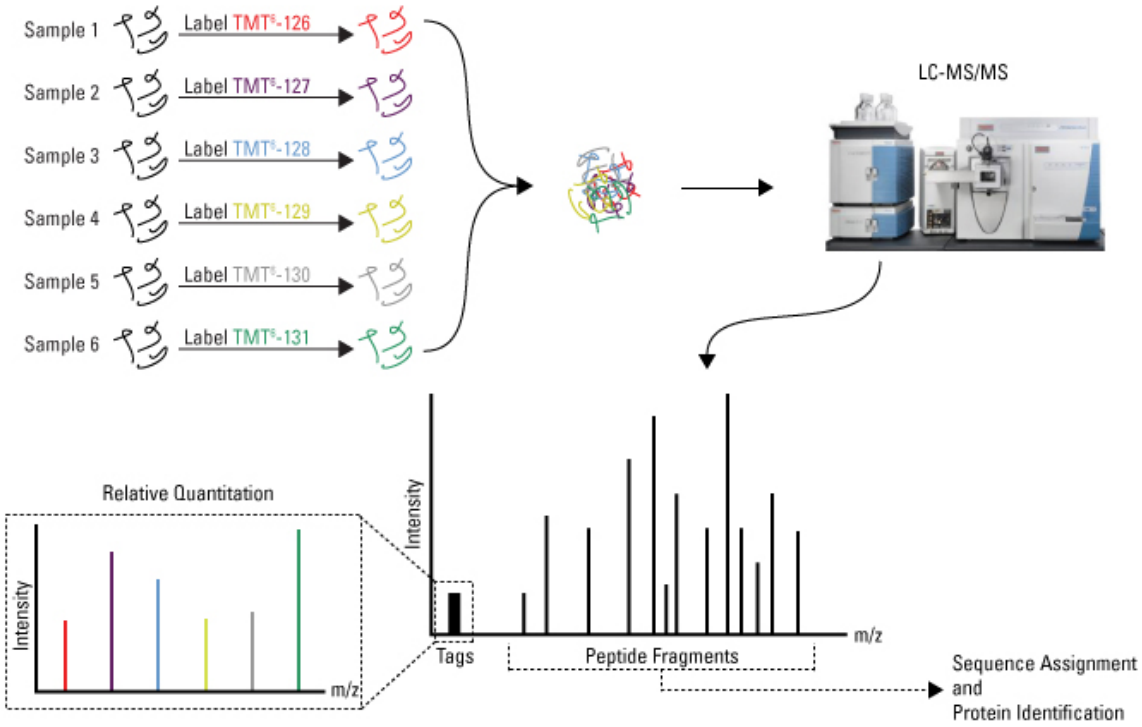


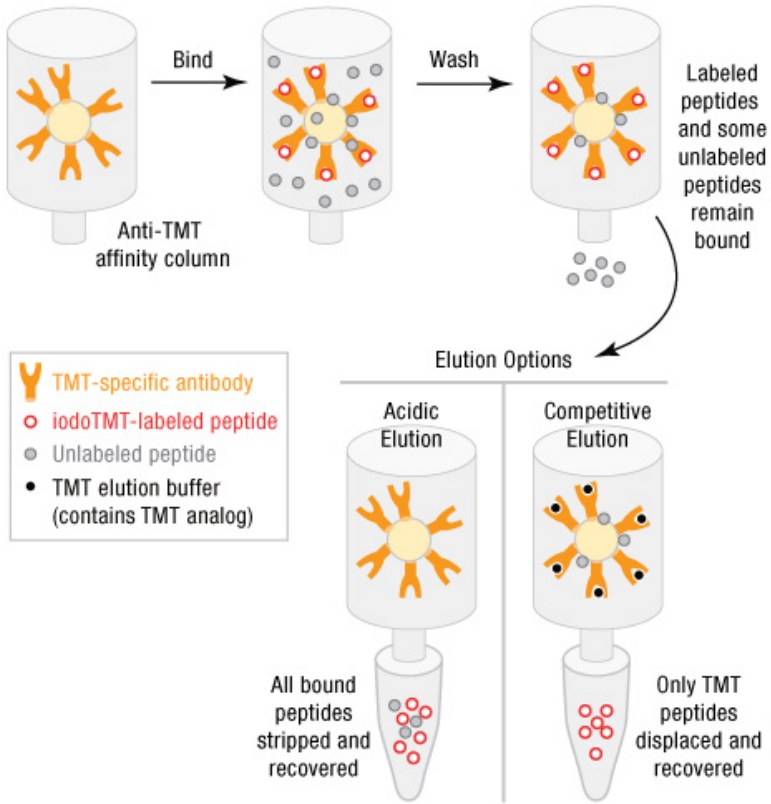
Figure 1. SILAC with ^{13}C6-Arg. One cell population is cultured in ^{13}C6-Arg labeling medium, whereas the other population is grown in normal arginine. In a typical SILAC strategy, the cells are combined and then lysed; however, they can also be lysed separately and mixed at a later stage of the experiment. Proteins are digested with trypsin and analyzed by LC–MS/MS.



Reporter ion relative quantification

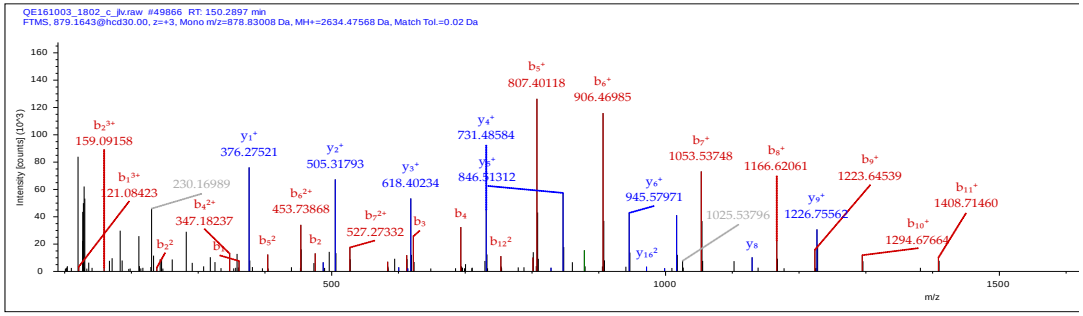


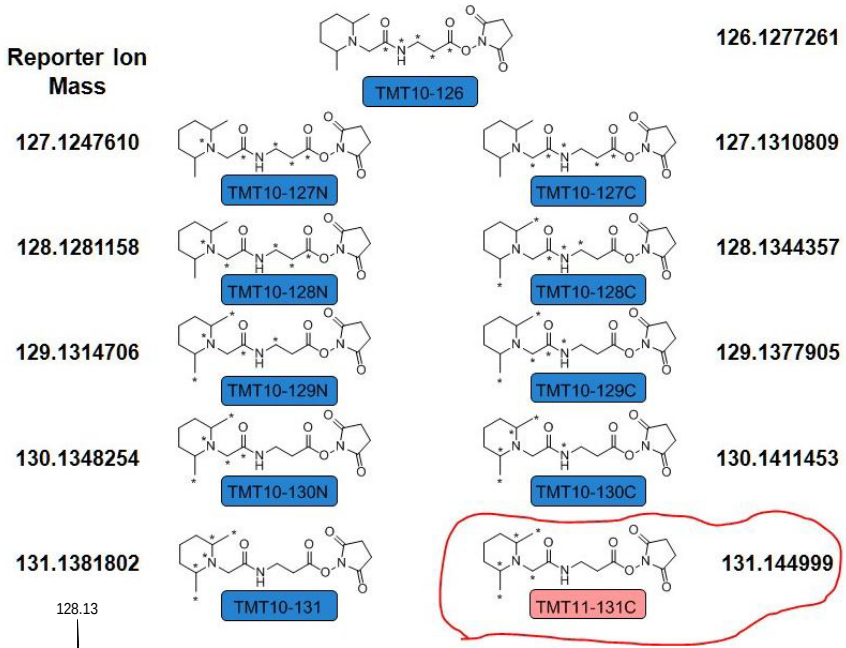
Possibilité d'enrichissement des peptides marqués

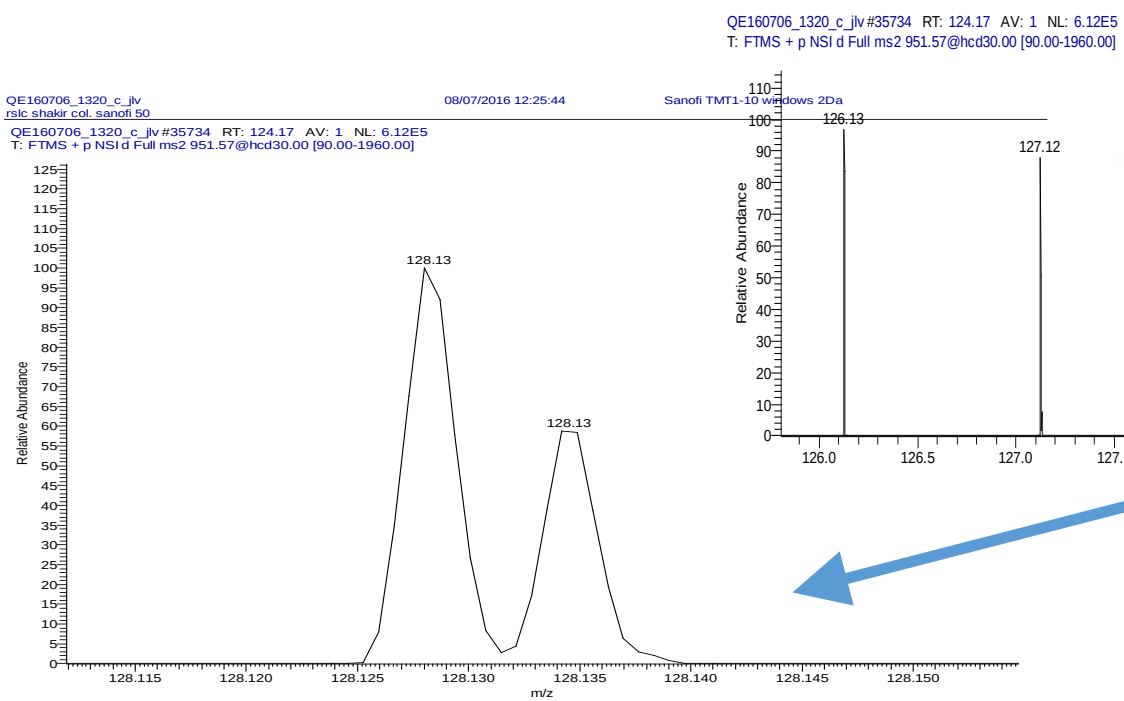


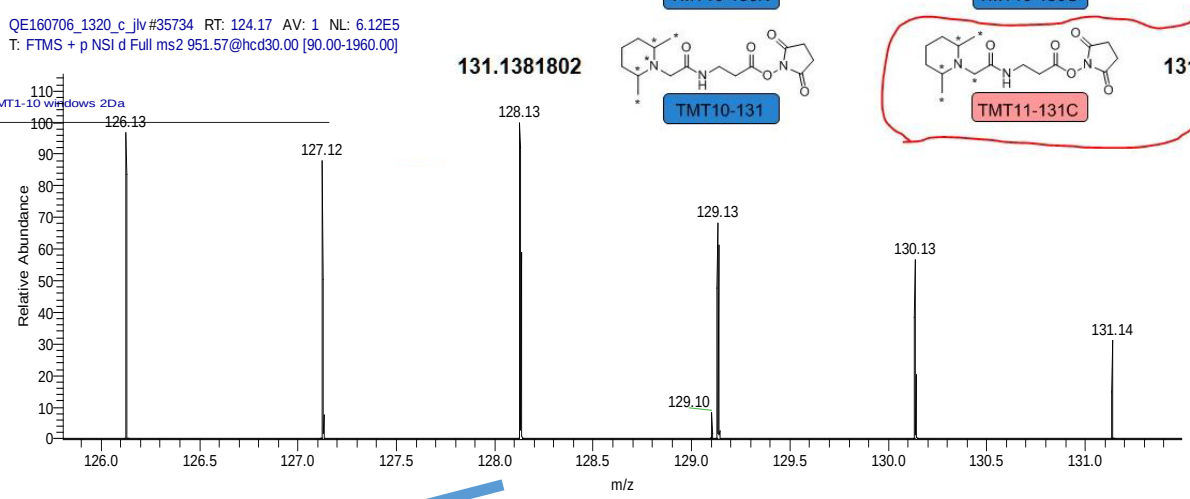
TMT (Tandem Mass Tag)
iTraq

10 and 11-plex TMT: high resolution MS/MS required

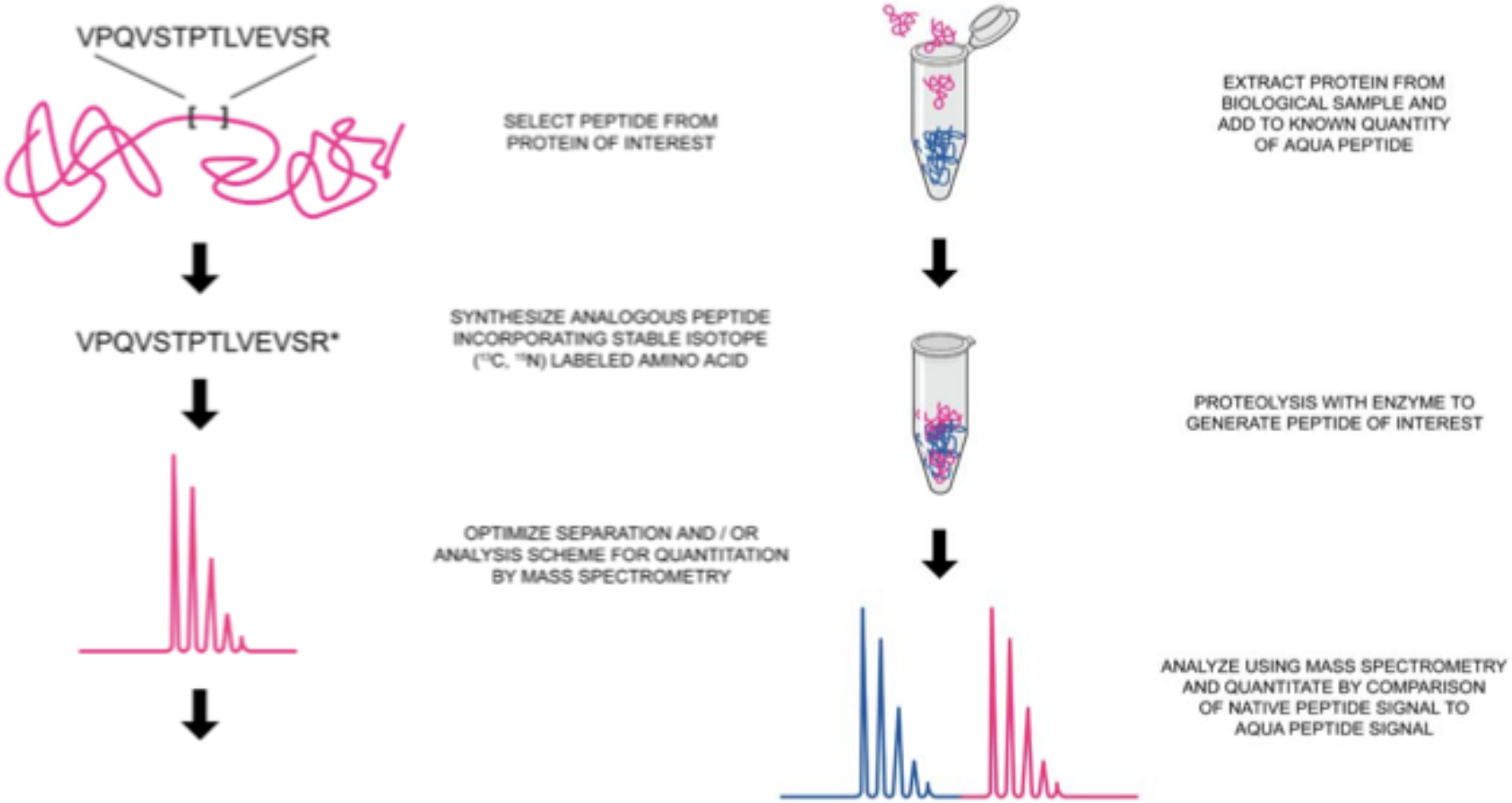








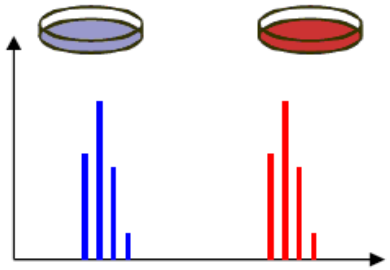
Overview of AQUA Strategy (Developed by Gygi *et al*)



Intensity-measurement incertainty correlation

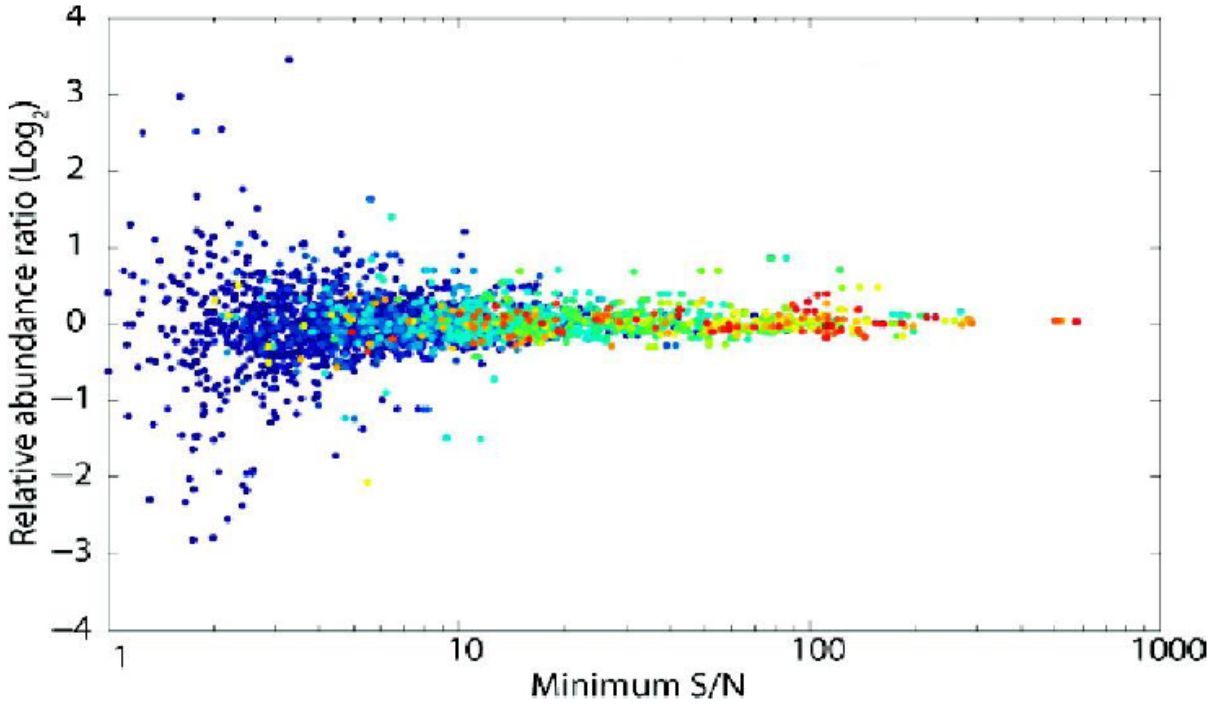
Corrélation avec le rapport signal sur bruit (S/B)

Codage des Arg et Lys



Mélange 1:1

~1/3 des ions à S/B < 5



Targeted analysis improves the S/N

J Proteome Res 2008, *7*, 4756-65, Bakalarski *et al.*

PTMomics and relative quantification

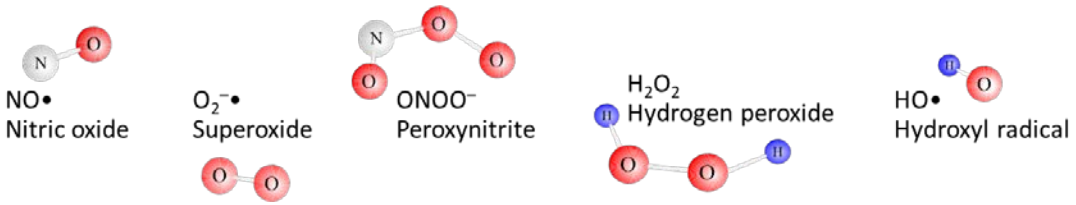
What do we want to know?

- **Sample 1 *vs.* Sample 2**
- **Modified *vs.* unmodified**
- **PTM1 *vs.* PTM2**
- **Absolute titration**

Cysteine redoxome: a case study for quantification

Michel Toledano, CEA
Jean-Marc Ghigo, Nicolas Barraud Inst. Pasteur
Clotilde Policar, Nicolas Delsuc ENS
Sophie Vriz, Alain Jolliot, Collège de France

THE CASE OF REVERSIBLE OXIDATION OF CYS



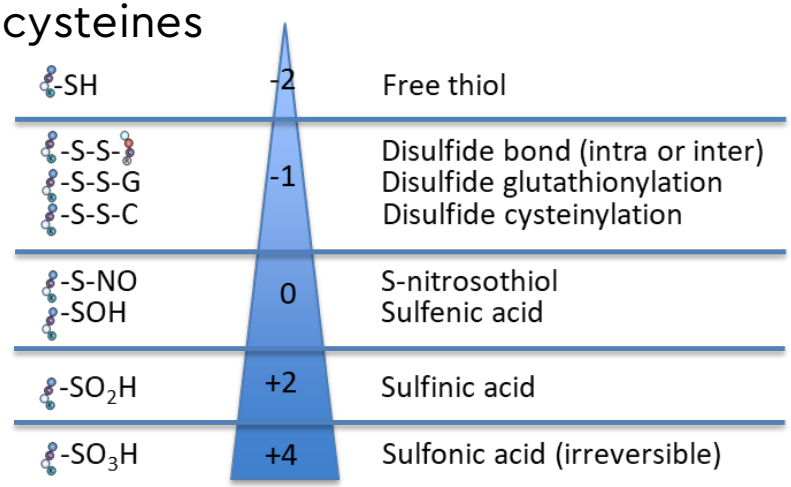
Reactive oxygen and nitrogen species (ROS/RNS)

- Low concentrations
  - regulation physiological processes,
  - reversible

Cell. 2002 Nov 15;111(4):471-81

- High concentrations:
  - Deleterious process (~burning, rusting); e.g. ageing
  - irreversible damage to DNA, proteins and lipids

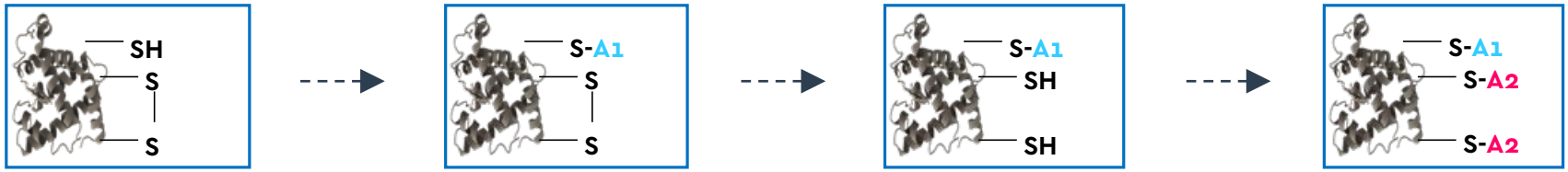
Different redox states of protein cysteines



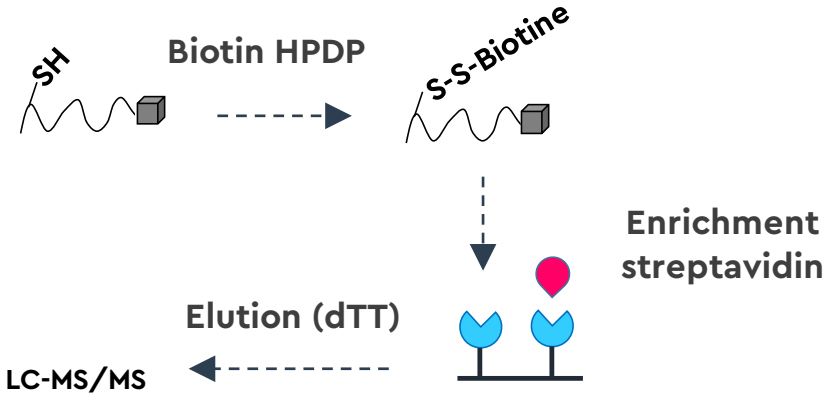
From Antiox. & Redox Signal., 26(7), 2017

Redox proteomics strategies

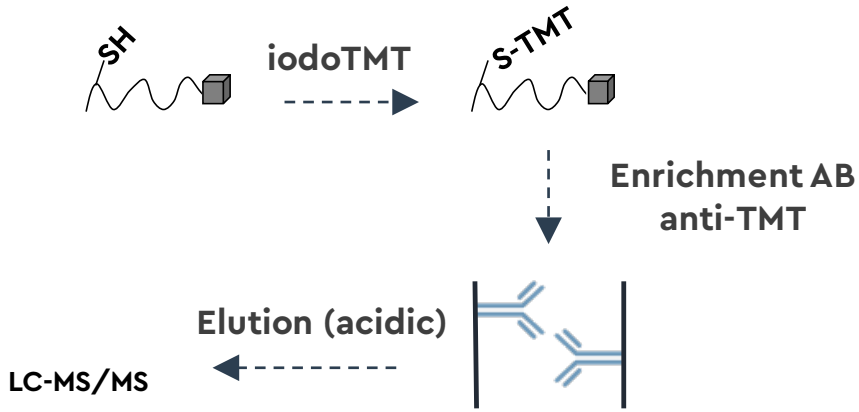
Differential labelling of Cys residues



Enrichment of oxidized Cys residues

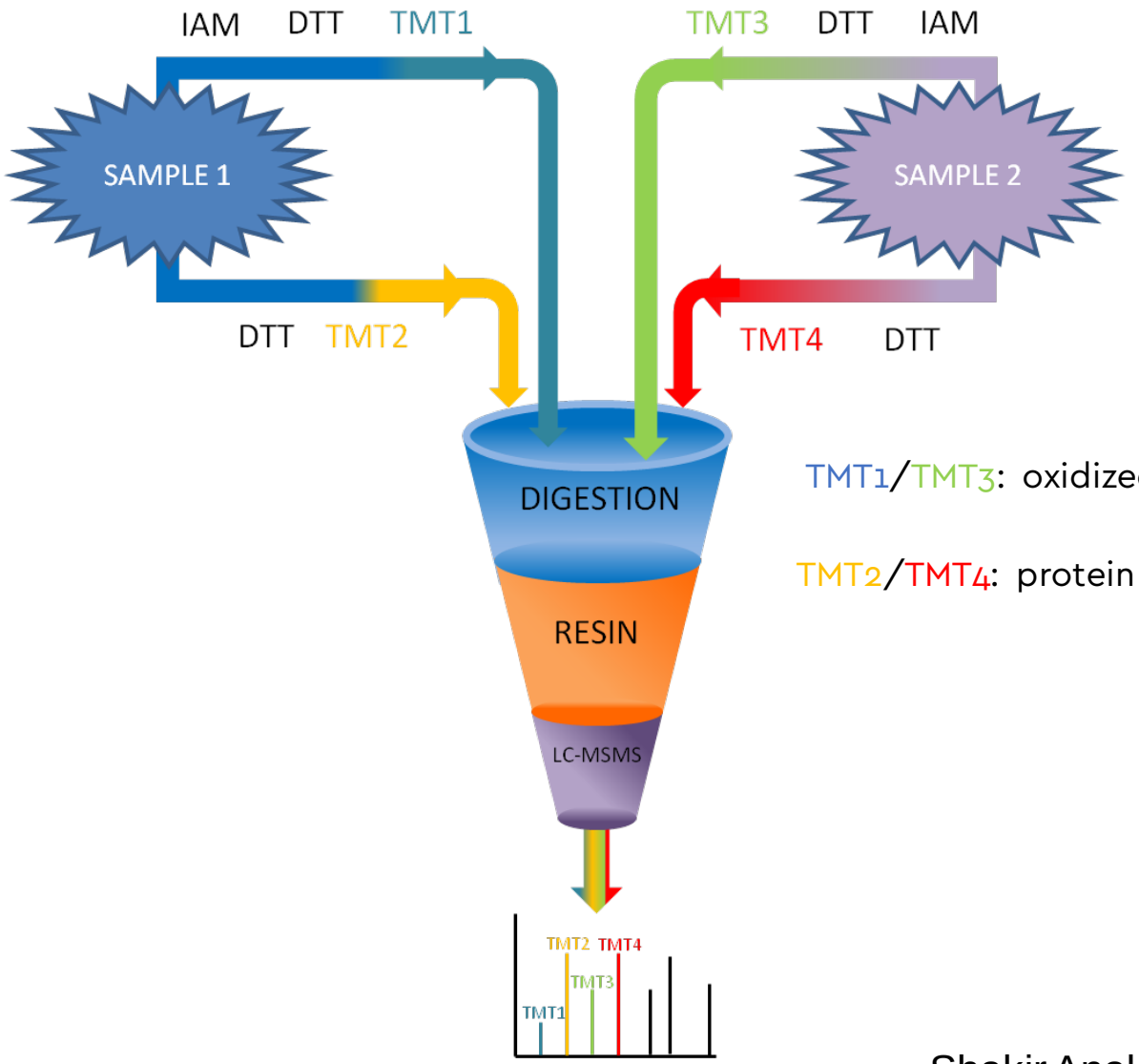


McDonagh B et al. J Proteomics. 2009



OxICAT and ICAT based strategies (Leichert PNAS 2008, Fu MCP 2009, Garcia-Santamarina Nat Protoc 2014)
CysTMT, iodoTMT and OxiTMT (Behring FASEB 2014, Murray MCP 2012, Shakir Anal Bio Chem 2017)
OcSILAC (Chiappetta HUPO 2011, Shakir submitted)

OxiTMT: Label-based relative quantification



TMT1/TMT3: oxidized Cys fold change

TMT2/TMT4: protein fold change

Shakir Anal Bio Chem 2017

PTMs quantification specificities

1 single peptide $\frac{I_{mod\,peptide}2}{I_{mod\,peptide}1}$ TMT1/TMT3: oxidized Cys fold change

All proteolytic peptides $\frac{I_{protein}2}{I_{protein}1}$ TMT2/TMT4: protein fold change

Modification or expression level?

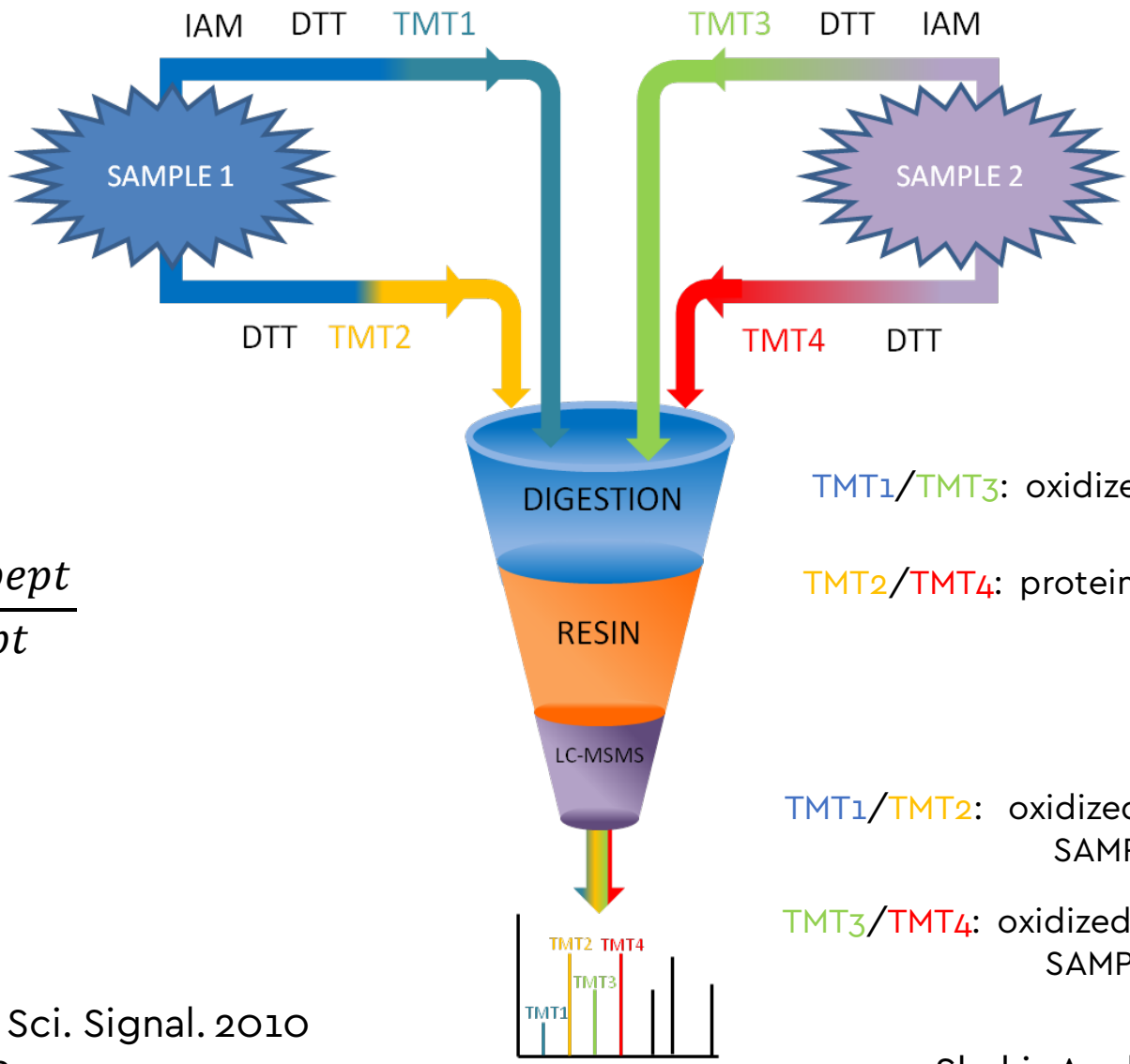
$$\begin{cases} \frac{I_{mod\,peptide}2}{I_{mod\,peptide}1} = 2 \quad \frac{I_{protein}2}{I_{protein}1} = 2 \\ \frac{I_{mod\,peptide}2}{I_{mod\,peptide}1} X \frac{I_{protein}1}{I_{protein}2} = 1 \end{cases}$$

$$\begin{cases} \frac{I_{mod\,peptide}2}{I_{mod\,peptide}1} = 3 \quad \frac{I_{protein}2}{I_{protein}1} = 6 \\ \frac{I_{mod\,peptide}2}{I_{mod\,peptide}1} X \frac{I_{protein}1}{I_{protein}2} = 0.5 \end{cases}$$

$$\begin{cases} \frac{I_{mod\,peptide}2}{I_{mod\,peptide}1} = 3 \quad \frac{I_{protein}2}{I_{protein}1} = 1 \\ \frac{I_{mod\,peptide}2}{I_{mod\,peptide}1} X \frac{I_{protein}1}{I_{protein}2} = 3 \end{cases}$$

1% to 3% or 30% to 90% ????

Label based site occupancy



Site Occupancy

$$\frac{N\ modified\ pept}{N\ total\ pept}$$

TMT1/TMT3: oxidized Cys fold change

TMT2/TMT4: protein fold change

TMT1/TMT2: oxidized CYS site occupancy SAMPLE 1

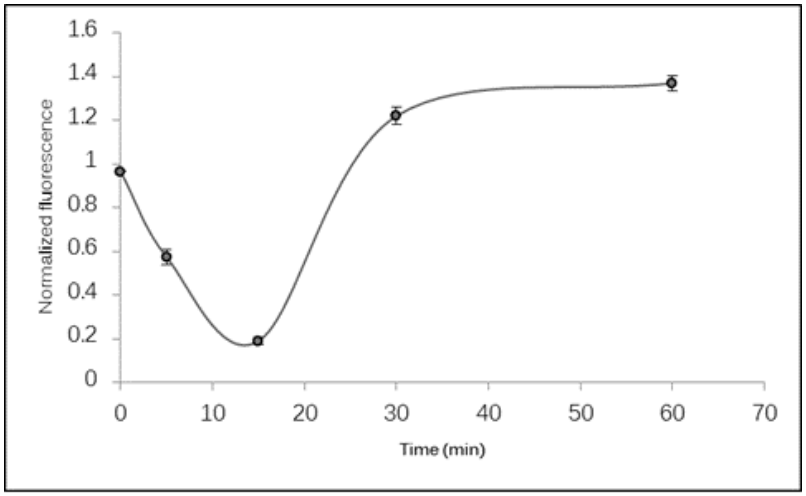
TMT3/TMT4: oxidized CYS site occupancy SAMPLE 2

Olsen J.V.& Mann M., Sci. Signal. 2010
Sharma K et al. Cell Rep. 2014

Shakir Anal Bio Chem 2017

E. Coli redox proteomics analysis by OxiTMT

E.Coli model before (1) and after (2) 30min 1mM H2O2



Araki K. J. Proteome Res 2016
Dardalhon M Free Radical Bio Med. 2012

Quantification: 1229 iodoTMT Cys, 580 proteins

- bound fraction: 886 pept (1019 Cys), 487 proteins (172 specific).
- unbound fraction: 834 peptides (893 Cys), 408 proteins (93 specific)

25 up-regulated proteins in H_2O_2 treated cells

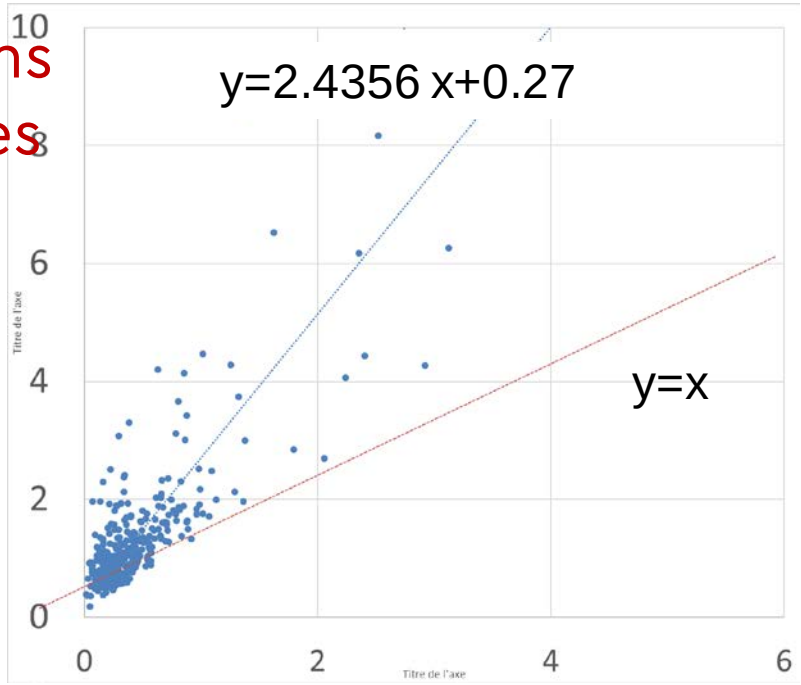
- oxidoreduction and generation of precursor
- metabolites and energy pathways

18 down-regulated oxidized fraction

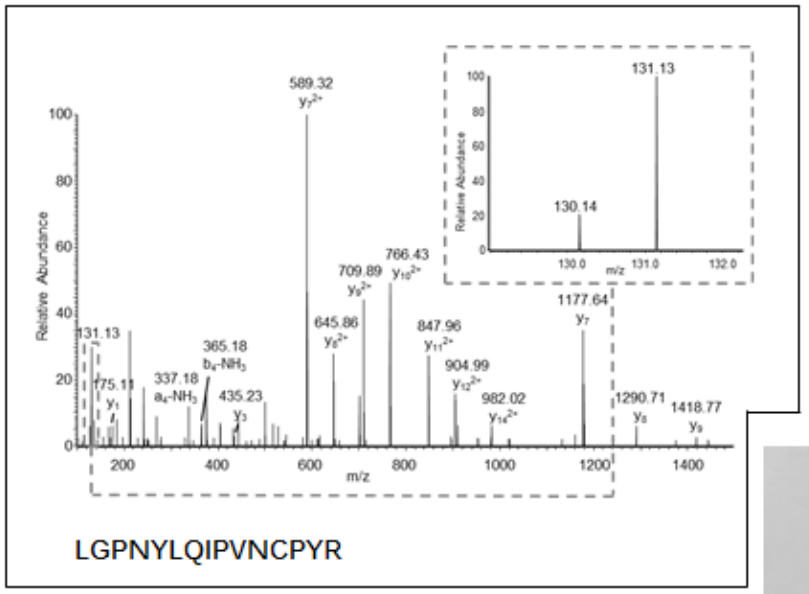
Considering some risks…

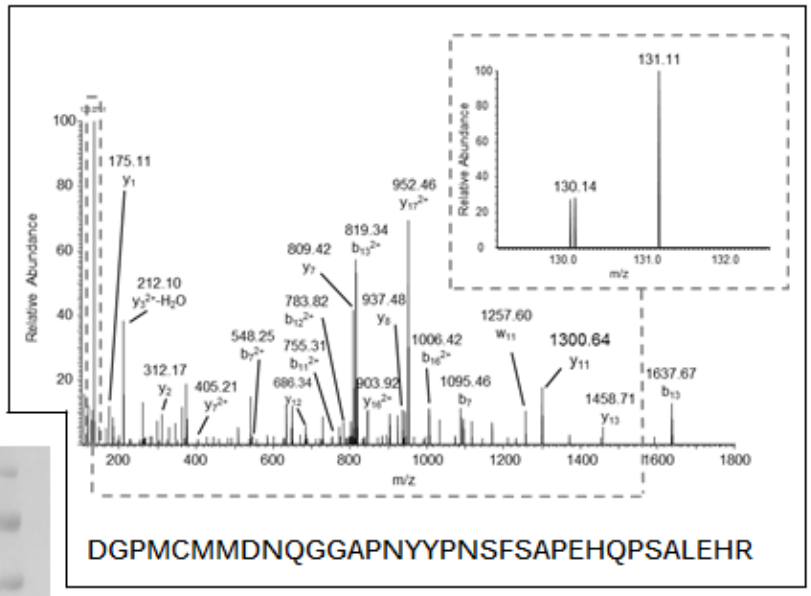
Where do the biases come from?

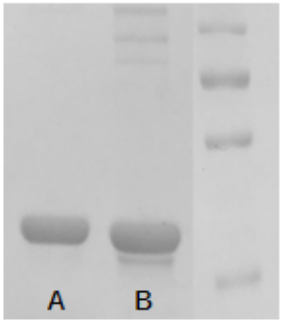
- Sample preparation (biological replicates)
- Sampling
- LC-MS (separation, ionization, matrix effects)
- MS sampling and scan events
- Quantification of each peptide from runs
- Quantification of proteins from peptides

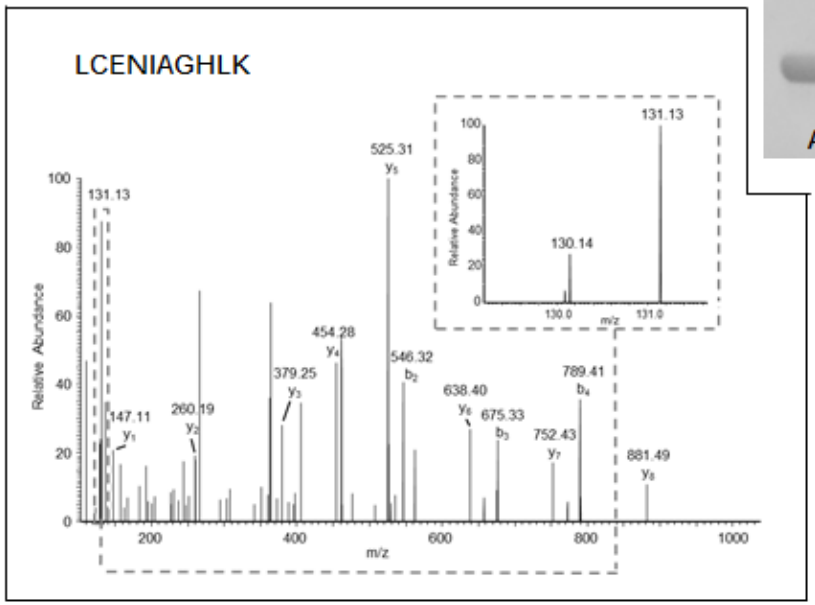


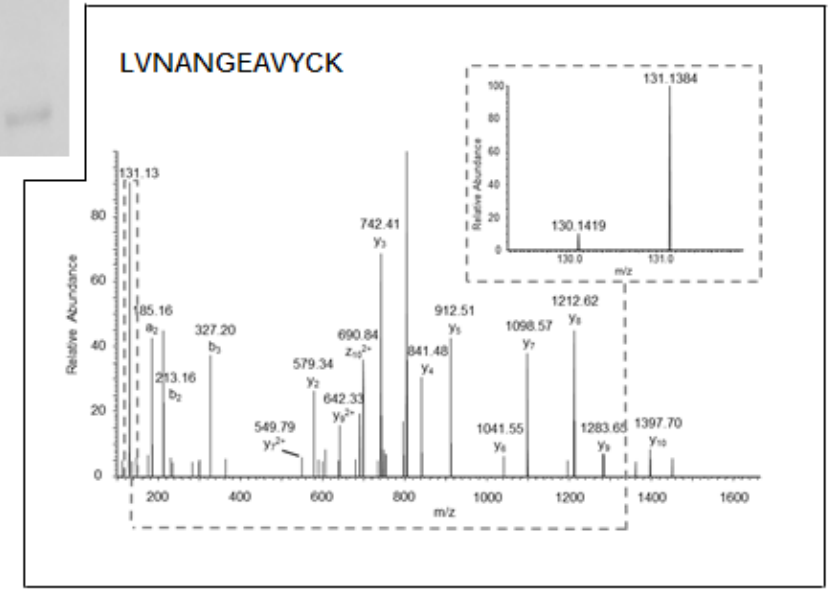
Bias from peptide to peptide







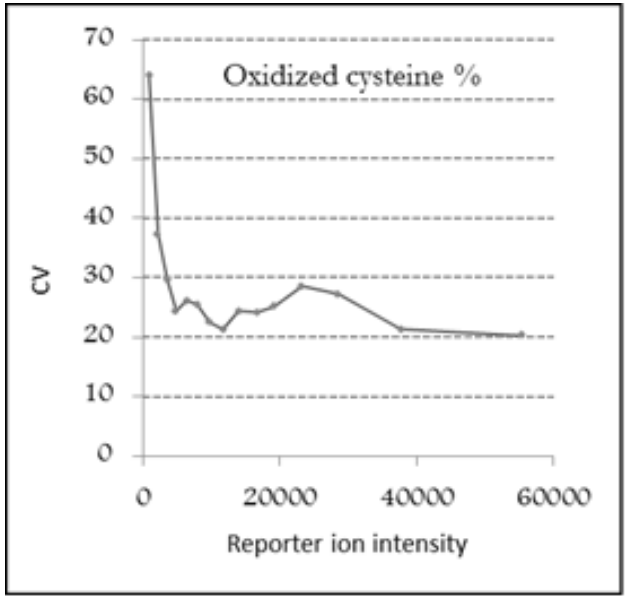


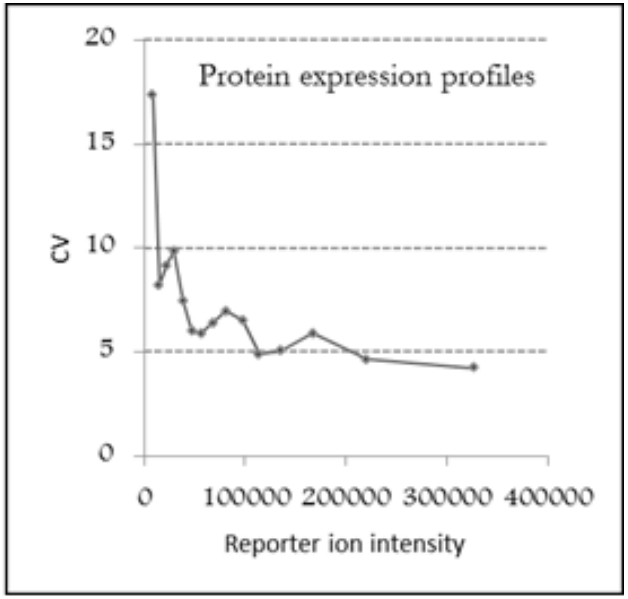


Mini signals, mini results

Co-isolation of peptides is less frequent

Specific challenge of ion reporter PTM quantification compared to protein quantification





Error increase for low abundance
Shift of experimental ratios towards higher values

Biased quantification of low abundance species.

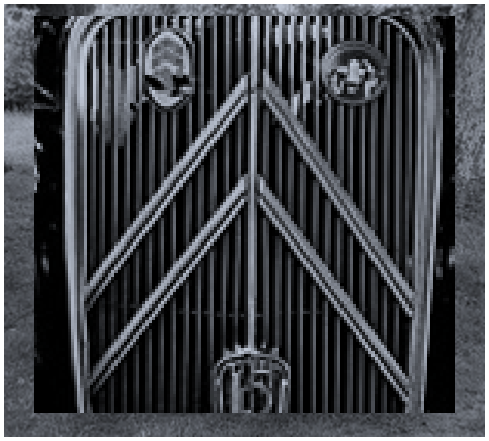
Oxidized unabundant cysteine reporter ions are more affected than total cysteine reporter ions

Protein coverage: Sampling and specificity

Bottom up partial characterization of the species of interest might lead to misinterpretation…

protein1

protein2



protein2

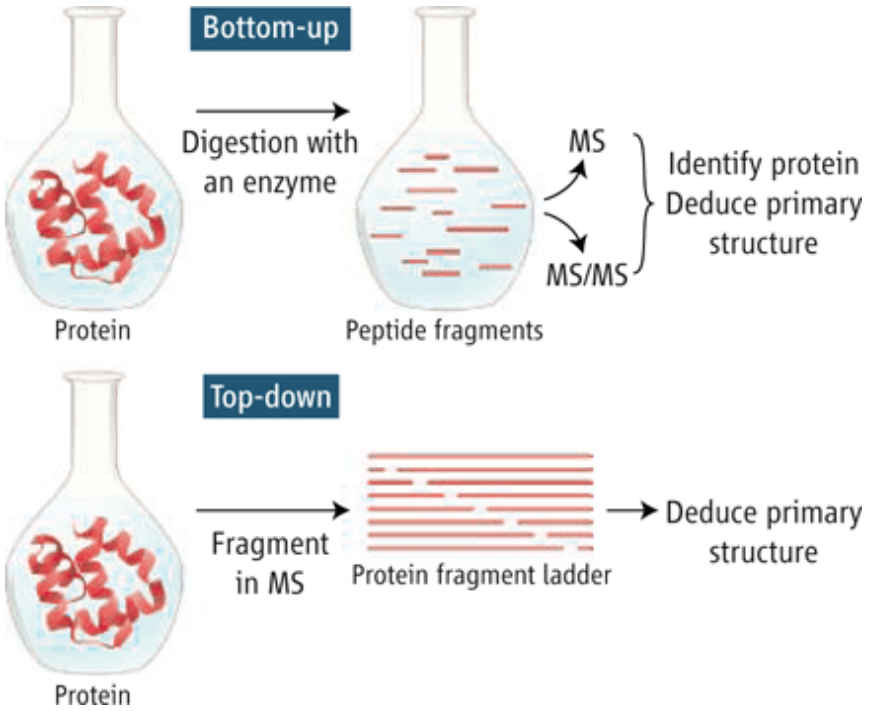
protein1

protein2

protein1

Top-down or Bottom-up?

We known how to purify, fragment and sequence peptides with high sensitivity and high throughput (like « pros »!)



B.T. Chait, Science (2006) 314:65-66

- **Bottom-up approach** : sequential processes each providing necessary information to describe a whole.
  - identification of proteins, sequencing and PTM characterization from proteolytic peptides
- **Top-down approach** : processes starting from the raw material aiming to transform and modify it to simplify its description.
  - identification of proteins and global structural analysis (sequence, PTMs) without proteolysis

For intact proteins, we are still beginners…

Acknowledgments

Oxidative Stress & Cancer Lab CEA
Michel TOLEDANO
Alise PONSERO

Applied Statistics ESPCI
Isabelle RIVALS

- SMBP / ESPCI

- Giovanni Chiappetta
- Shakir Shakir
- Massamba Ndiaye
- Martha Zoupoulaki
- Nicolas Eskenazi

- Emmanuelle Demey-Thomas
- Yann Verdier
- Iman Haddad

- *Sophie Liuu*
- *Anne Marie Hesse*
- *Sega Ndiaye*

- €