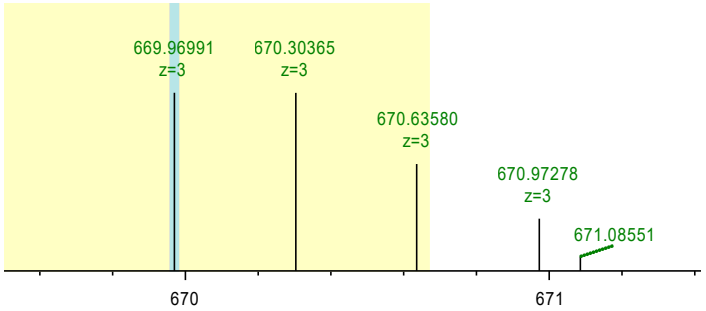
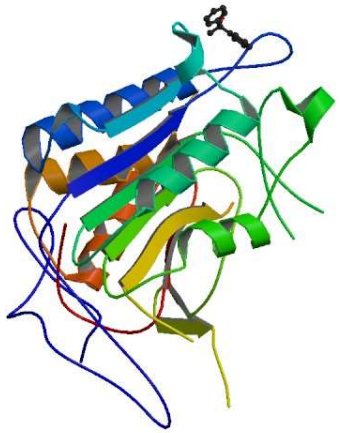
Production of omics data: Proteomics

Thibaut Léger, PhD

LERES, EHESP





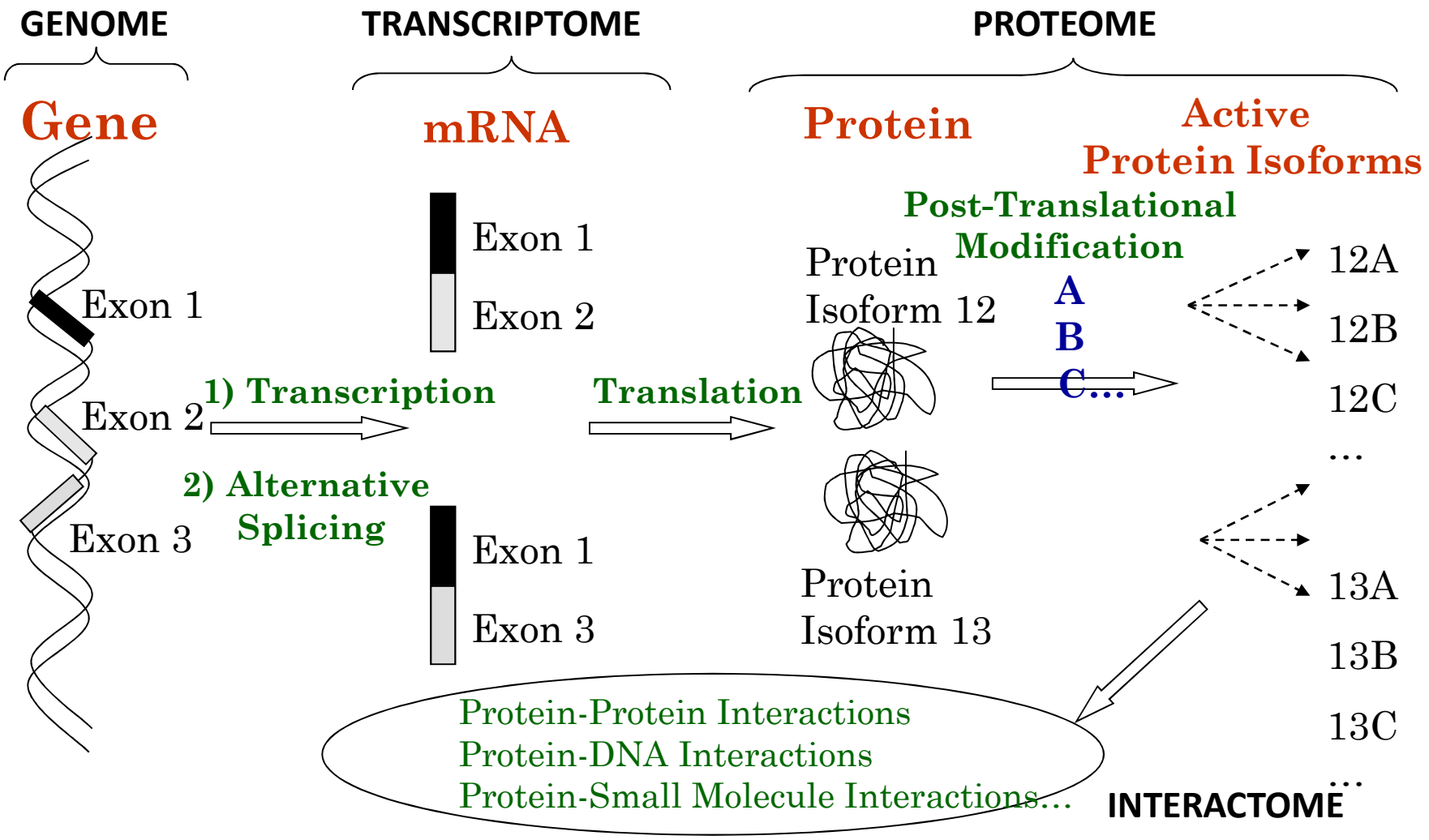
DUBii

25th may 2020

What is behind these data?

													Normalized abundance							
													Sain						Patient	
	Accession	Peptide cou	Unique pep	Confidence	Anova (p)	q Value	Max fold ch	Power	Highest me	Lowest mea	Mass	Description	1845007-F1	1845007-F3	1845007-F5	1845007-F7	1845007-F9	1845007-F1	1845007-F2	1845007-F4
	P40197	12	12	545.23	2.48E-07	4.74E-05	32.042418	1	Patient	Sain	60.921	Platelet glycoprotein V OS	1447.8857	898.47877	2114.249	3517.982	2506.9091	1186.7152	33178.605	73995.78
	P02776	2	2	134.77	2.63E-07	4.74E-05	78.817355	1	Patient	Sain	10.838	Platelet factor 4 OS=Homc	3812.4369	3755.8358	1044.3911	2939.4867	3862.0883	1110.5549	333829.66	68071.51
	Q13201	6	6	221.4	5.33E-07	6.40E-05	72.663027	1	Patient	Sain	138.023	Multimerin-1 OS=Homo sa	160.48528	939.93933	155.7663	732.57482	572.39757	196.60688	75519.303	23299.12
	P04114	336	334	28302.18	7.99E-06	0.0006593	2.1384481	1	Patient	Sain	515.283	Apolipoprotein B-100 OS=	14117253	10601958	8472023.6	9982572.2	10056625	11898407	24415094	1744463
	P07996	42	42	2678.71	9.16E-06	0.0006593	143.08362	1	Patient	Sain	129.3	Thrombospondin-1 OS=Hc	11604.367	10346.75	4059.3393	5652.5533	2121.1343	2367.1543	2303007.2	83175.08
	Q15485	7	7	289.57	1.42E-05	0.0007087	24.427918	0.9999997	Patient	Sain	33.98	Ficolin-2 OS=Homo sapier	8778.176	6132.5206	10104.107	8194.6312	8690.2929	1737.9715	185819.2	105024.9
	P10720	3	3	208.63	1.49E-05	0.0007087	36.552808	0.9999997	Patient	Sain	11.545	Platelet factor 4 variant O	32881.787	11758.269	1451.6395	5448.2604	4187.5362	4196.5766	546497.96	239272.0
	Q12884	2	2	70.47	1.58E-05	0.0007087	Infinity	0.9999996	Patient	Sain	87.657	Prolyl endopeptidase FAP	0	0	0	0	0	0	961.55592	106.477
	Q15061	1	1	3.67	3.71E-05	0.0014832	23.811179	0.9999905	Sain	Patient	74.843	WD repeat-containing prc	567513.12	751475.75	852125.74	1189728.8	625225.18	78135.037	46681.126	36585.06
	P04075	8	8	371.85	4.58E-05	0.0016488	4.2937733	0.9999814	Patient	Sain	39.395	Fructose-bisphosphate al	33219.801	26234.171	28730.699	26696.908	33816.665	32024.353	128602.25	168571.3
	P09486	10	9	463.23	6.25E-05	0.0020451	6.1746973	0.999953	Patient	Sain	34.61	SPARC OS=Homo sapiens (	9522.5101	37758.131	36721.019	20388.641	15117.423	28228.93	299004.24	97804.24
	P10124	1	1	70.16	7.11E-05	0.002134	272.44302	0.9999321	Patient	Sain	17.641	Serglycin OS=Homo sapier	67.740323	0	958.26875	53.15204	160.31231	238.84011	105180.09	42246.24
	P02775	6	6	388.98	8.21E-05	0.0022557	28.012451	0.9998996	Patient	Sain	13.885	Platelet basic protein OS=	61607.9	109352.83	105910.89	67216.894	60865.635	22752.188	3524260.5	1692399.5
	Q9H1K0	2	2	12.94	8.77E-05	0.0022557	1.8919044	0.9998799	Patient	Sain	88.815	Rabenosyn-5 OS=Homo sa	871721.79	1194024.5	1103071.3	1121288.3	1020557.4	1361492.9	1506333.1	1757466.5
	P35542	4	4	338.41	0.0001236	0.0029659	16.656374	0.9997122	Patient	Sain	14.737	Serum amyloid A-4 protein	192222.76	102064.74	9416.0169	8023.3014	46690.651	107524.48	1543185.2	1741696.5
	P80188	6	6	223.58	0.0001847	0.0041561	9.1194461	0.9992657	Patient	Sain	22.574	Neutrophil gelatinase-as	2265.4484	1473.9176	6203.72	3676.0015	7827.3304	4603.3796	66633.817	26485.80
	P02144	4	4	209.23	0.0002392	0.0050651	6.5032553	0.9987201	Patient	Sain	17.173	Myoglobin OS=Homo sapi	2754.3333	2223.1201	1035.6898	1813.5104	1462.164	5472.3528	17977.389	20813.25
	P05067	5	5	179.76	0.000297	0.0059397	18.947351	0.9980148	Patient	Sain	86.888	Amyloid beta A4 protein C	408.20926	53.510036	338.118	494.28658	258.47482	516.31147	8561.0101	2161.775
	Q9NPH3	7	7	223.36	0.0003417	0.0064734	2.9534965	0.9973949	Patient	Sain	65.377	Interleukin-1 receptor acc	19269.326	23550.392	17918.705	19071.463	29165.24	20448.453	74326.28	82743.30
	P22352	7	7	322.73	0.0005632	0.0101356	2.3655577	0.9936247	Patient	Sain	25.537	Glutathione peroxidase 3	356078.09	500176.63	411199.92	354267.54	608490.2	620677.21	1139291.1	1214566.8
	A0A075B6I0	1	1	30.88	0.0005981	0.0102519	15.884721	0.9929515	Sain	Patient	12.806	Immunoglobulin lambda	14017.761	25223.891	6493.2483	70640.84	15600.774	36585.067	4641.3232	1487.298
	P05155	33	33	2513.9	0.0007389	0.0120898	5.4023431	0.9900925	Patient	Sain	55.119	Plasma protease C1 inhib	741421.19	592478.75	377555.54	465885.04	2331779.9	7791080.5	11109177	10511766
	A0A0C4DH2	2	2	100.38	0.0013344	0.0208833	3.8865369	0.9766266	Sain	Patient	12.999	Immunoglobulin heavy va	50500.146	64330.439	38938.111	121804.39	27881.723	32180.678	24057.435	17302.863
	P14780	6	6	207.75	0.0014988	0.0221418	62.066206	0.972746	Patient	Sain	78.408	Matrix metalloproteinase	794.97907	136.39171	262.24899	313.84328	0	555.64619	58928.664	9636.858
	P02649	27	27	1941.9	0.0015378	0.0221418	2.4710215	0.971824	Patient	Sain	36.132	Apolipoprotein E OS=Hom	2552396.1	1536331.8	853676.88	1079240.3	1188651.7	1327575.2	2857655.7	2177433.7
	Q3C1V8	1	1	14.2	0.0019728	0.0268284	5.551993	0.9615339	Patient	Sain	25.917	Brain-specific homeobox	119.76184	56.50442	16.821931	8.5075688	3888.8817	85.190432	6077.0842	3199.583
	P02652	11	11	995.34	0.0020124	0.0268284	2.4014067	0.9606031	Patient	Sain	11.168	Apolipoprotein A-II OS=Hc	34056410	26771849	14632439	15663194	42268641	26890744	89512485	43236635
	P08571	12	12	937.27	0.0025308	0.0320335	1.6971067	0.9485492	Sain	Patient	40.051	Monocyte differentiation	797874.17	434080.26	499432.28	477652.53	733837.28	418725.29	352767.33	318079.37
	P26927	31	31	1222.76	0.0025808	0.0320335	3.3123536	0.9474046	Patient	Sain	80.268	Hepatocyte growth factc								380083.05
	P02655	6	6	819.16	0.0030504	0.0361789	4.7206762	0.9368237	Patient	Sain	11.277	Apolipoprotein C-II OS=I								137623
	P02671	8	8	369.69	0.0031158	0.0361789	191.8384	0.9353762	Patient	Sain	94.914	Fibrinogen alpha chain								107899
	P02763	17	17	1378.36	0.0036234	0.0407586	2.8533872	0.924364	Sain	Patient	23.497	Alpha-1-acid glycoprote								2274357
	Q14831	1	1	23.81	0.0038979	0.0425178	2.9788784	0.9185746	Sain	Patient	102.185	Metabotropic glutamate								31759.88
	P02647	32	32	2367.65	0.0040205	0.0425648	2.2688435	0.9160271	Patient	Sain	30.759	Apolipoprotein A-I OS=H								27802701
	P01876	21	21	1557.26	0.0044316	0.0455765	3.5420267	0.9076475	Sain	Patient	37.631	Ig alpha-1 chain C regio								5212718.5
	Q9UNW1	8	8	227	0.005259	0.0525841	1.7900311	0.8915219	Sain	Patient	55.016	Multiple inositol polypl								30358.52
	Q16853	2	2	81.52	0.0054293	0.0528194	6.1436713	0.8883207	Sain	Patient	84.568	Membrane primary ami								194.7495
	P01833	8	8	219.66	0.006944	0.0656448	9.1555604	0.8614401	Sain	Patient	83.232	Polymeric immunoglobu								2483.330
	P27169	28	27	2064.1	0.0071604	0.0656448	1.8486226	0.8578184	Patient	Sain	39.706	Serum paraoxonase/ary								6280114.6
	P60174	1	1	28.75	0.0074	0.0656448	2.9166777	0.8538651	Patient	Sain	30.772	Triosephosphate isome								4675.589
	P00736	41	39	3067.37	0.0074771	0.0656448	1.4866701	0.8526056	Sain	Patient	80.067	Complement C1r subcon								2638367.5
	P09871	37	37	2588.95	0.0085558	0.0726349	1.93603	0.835606	Sain	Patient	76.635	Complement C1s subcor								2302798.8
	A0A0B4J1V0	3	3	104.52	0.0086769	0.0726349	2.7965659	0.833766	Sain	Patient	12.917	Immunoglobulin heavy va	92355.065	20597.701	46790.32	87259.628	45135.867	81618.911	38454.812	17117.34
	P01619	4	4	262.21	0.0095577	0.0781901	2.30903	0.8207663	Sain	Patient	12.549	Immunoglobulin kappa va	409287.43	332471.99	228180.64	675002.86	220569.52	187876.28	139130.56	148976.38
	Q08380	21	21	1485.84	0.0115602	0.0900161	2.528023	0.7934679	Sain	Patient	65.289	Galectin-3-binding protei	938408.99	614190.09	1091308.6	2622296.8	1676638.8	2590031.7	640186.94	906404.59
	P06732	9	9	314.24	0.0117011	0.0900161	5.8938011	0.7916528	Patient	Sain	43.074	Creatine kinase M-type OS	7288.867	9691.0748	1004.0051	6516.3248	4494.7392	6004.4549	19154.379	89665.94

CONCEPT


GENOME
TRANSCRIPTOME
PROTEOME
Gene
mRNA
Protein
Active Protein Isoforms
Post-Translational Modification
Exon 1
Exon 2
Exon 3
1) Transcription
2) Alternative Splicing
Translation
Exon 1
Exon 2
Exon 1
Exon 3
Protein Isoform 12
Protein Isoform 13
A
B
C…
12A
12B
12C
…
13A
13B
13C
…
Protein-Protein Interactions
Protein-DNA Interactions
Protein-Small Molecule Interactions…
INTERACTOME

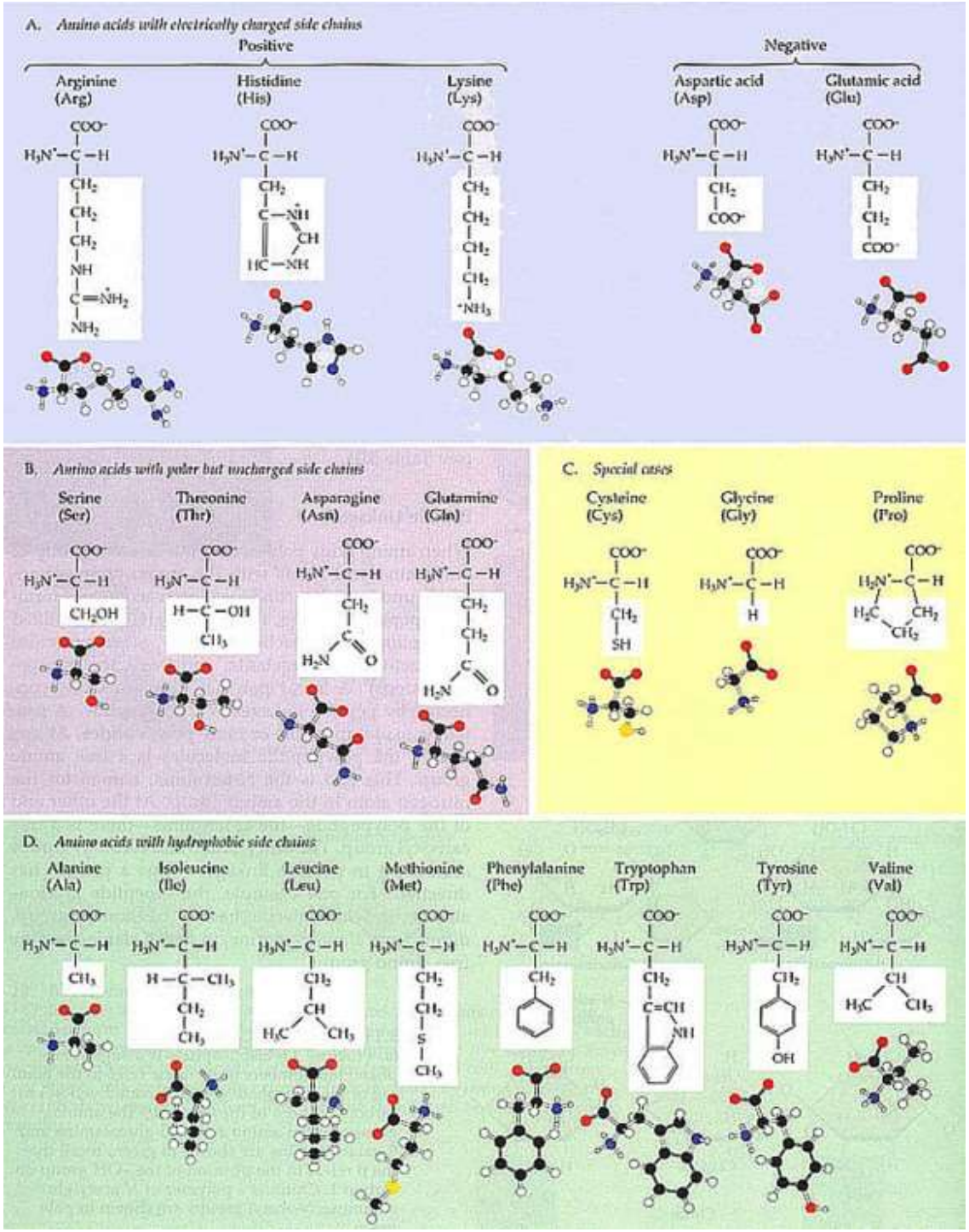
GENOMICS vs PROTEOMICS

Genome (DNA)

- Static (no change with time)
- Can be amplified (PCR)
- Little sample complexity
- *(4 base pairs, very similar, same order of concentration)*
- Good solubility

Proteome (proteins)

- Dynamic
- *(highly variable with time; many proteomes for one genome)*
- Cannot be amplified
- High sample complexity (wide variety of physical and chemical properties; concentrations can differ by 9 orders of magnitude)
- Various solubility; some proteins are insoluble in water



Alanine	A, Ala	71.079
Arginine	R, Arg	156.188
Asparagine	N, Asn	114.104
Aspartic acid	D, Asp	115.089
Cysteine	C, Cys	103.145
Glutamine	Q, Gln	128.131
Glutamic acid	E, Glu	129.116
Glycine	G, Gly	57.052
Histidine	H, His	137.141
Isoleucine	I, Ile	113.160
Leucine	L, Leu	113.160
Lysine	K, Lys	128.17
Methionine	M, Met	131.199
Phenylalanine	F, Phe	147.177
Proline	P, Pro	97.117
Serine	S, Ser	87.078
Threonine	T, Thr	101.105
Tryptophan	W, Trp	186.213
Tyrosine	Y, Tyr	163.176
Valine	V, Val	99.133

Proteomics?

- Proteomics is the large-scale study of proteomes, it means all proteins from a cell, an organelle, a tissue, an organ or from an organism at a one point, under specific conditions.
- Proteomics is at the crossroads of biochemistry, analytical chemistry and bioinformatics.

⇒Proteins can be modified by different biological or chemical processes; The different variants of proteins are called now:

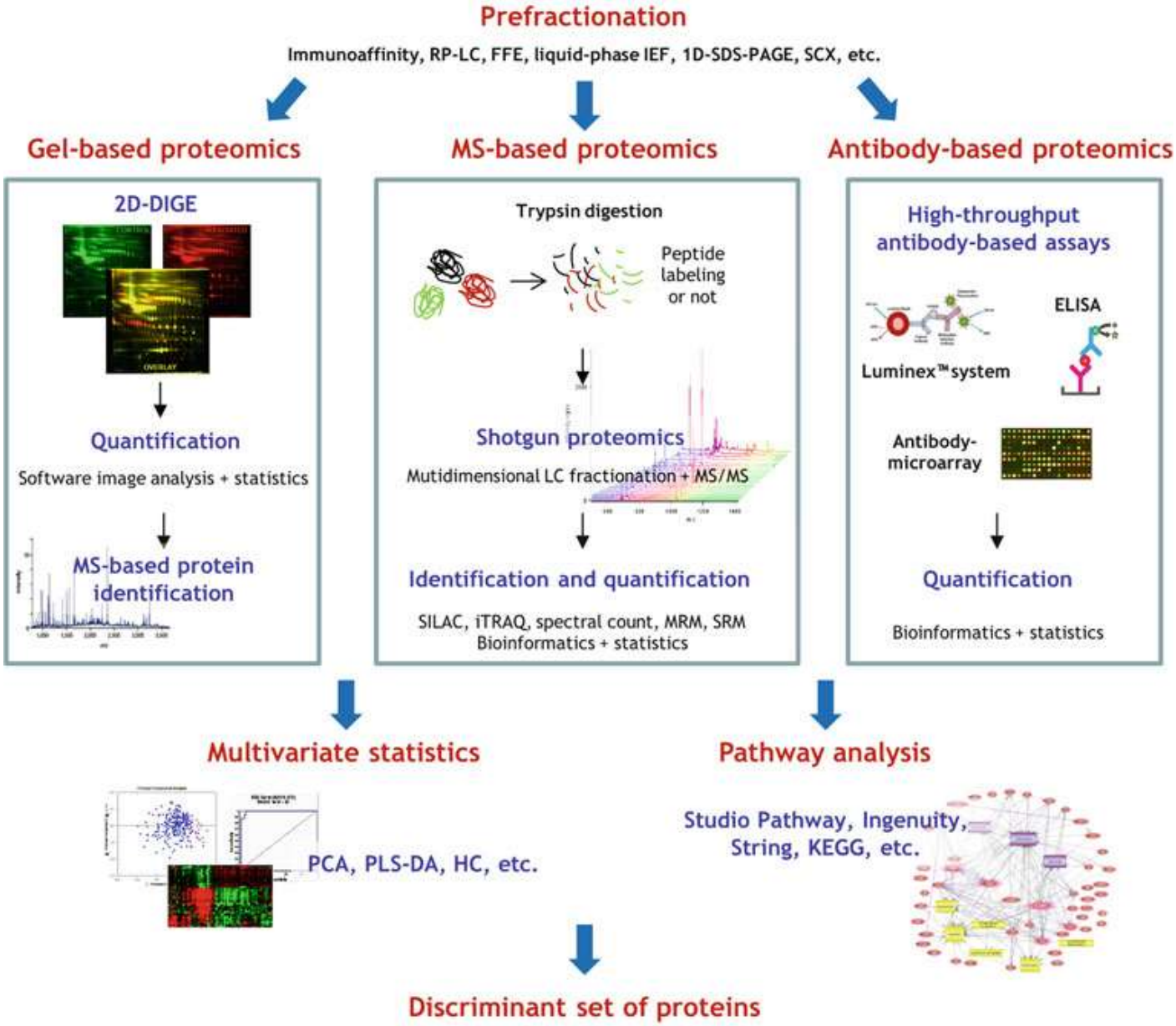
Proteoforms

Nat Methods. 2013 Mar;10(3):186-7. doi: 10.1038/nmeth.2369.
Proteoform: a single term describing protein complexity.
Smith LM, Kelleher NL; Consortium for Top Down Proteomics

PROTEOMICS GOALS

- Identification of all proteins in a proteome
- Search for new, hypothetical or predicted proteins
- Analysis of differential expression between 2,3,... different conditions (protein up- or downregulation)
- Identification of post-translational modifications
- Characterization of proteins by function, pathway, cellular location, etc.
- Study of protein-protein interactions

Proteomics techniques



HUMAN PROTEOME MAP

Statistics	
Organs/cell types	30
Genes identified	17,294
Proteins identified	30,057
Peptide sequences	293,700
N-terminal peptides	4,297
Splice junctional peptides	66,947
Samples	85
Adult tissues	17
Fetal tissues	7
Cell types	6

Welcome to ProteomicsDB!

ProteomicsDB is a joint effort of the Technische Universität München (human proteome and its use across the scientific community.

Browse proteins

Explore the human proteome protein by protein.

Status

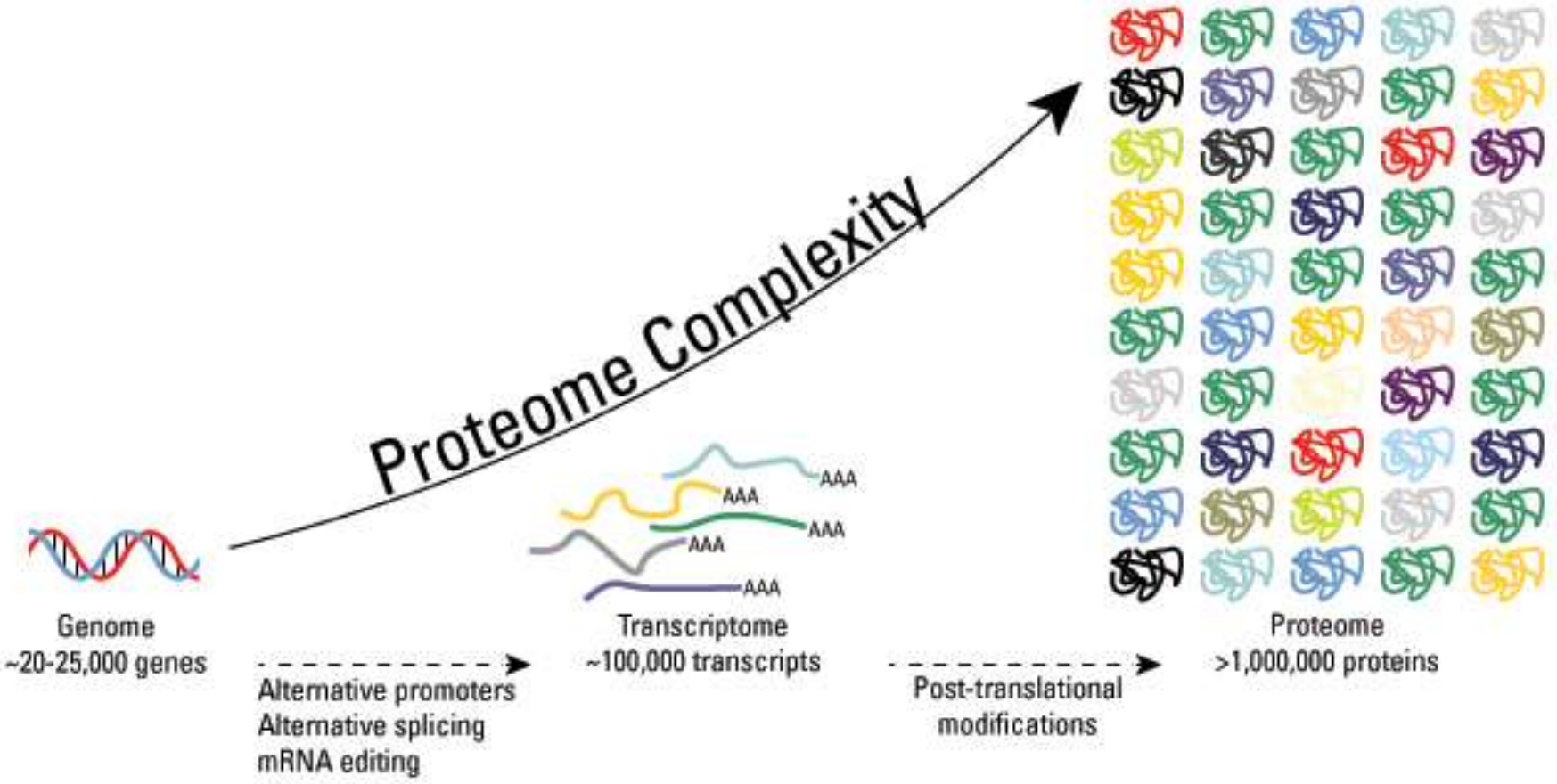
Human Proteome

Coverage:	80%
Proteins:	15721 of 19629
Isoforms:	11353 of 86771
Unique Peptides (Isoform):	113944
Unique Peptides (Gene):	455289
Spectra:	43237800

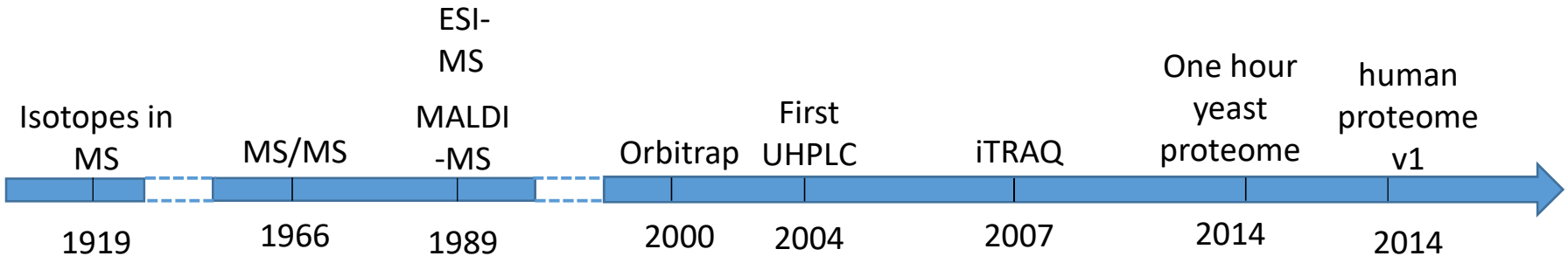
Repository

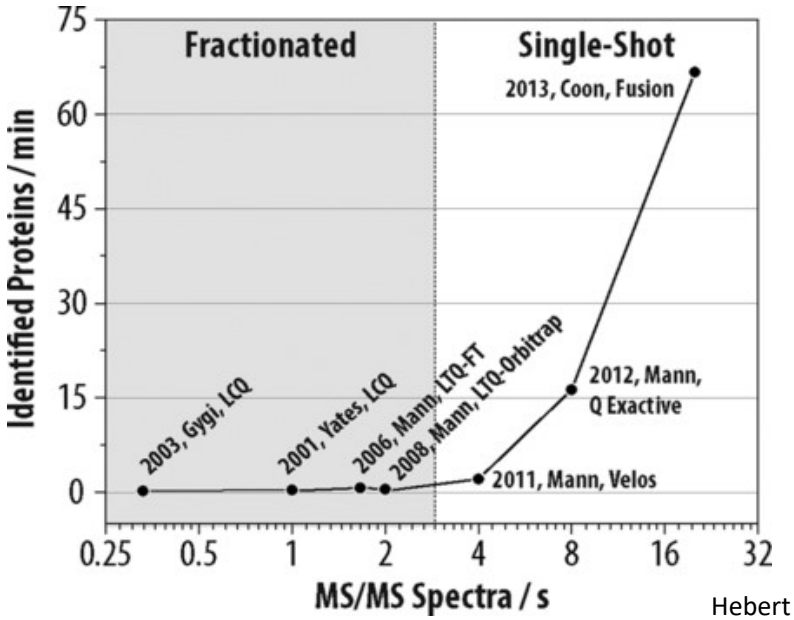
Registered Users:	533
Projects:	75
Experiments:	397
Files:	19459
Data Volume:	7.84 TB

Problem of proteome complexity



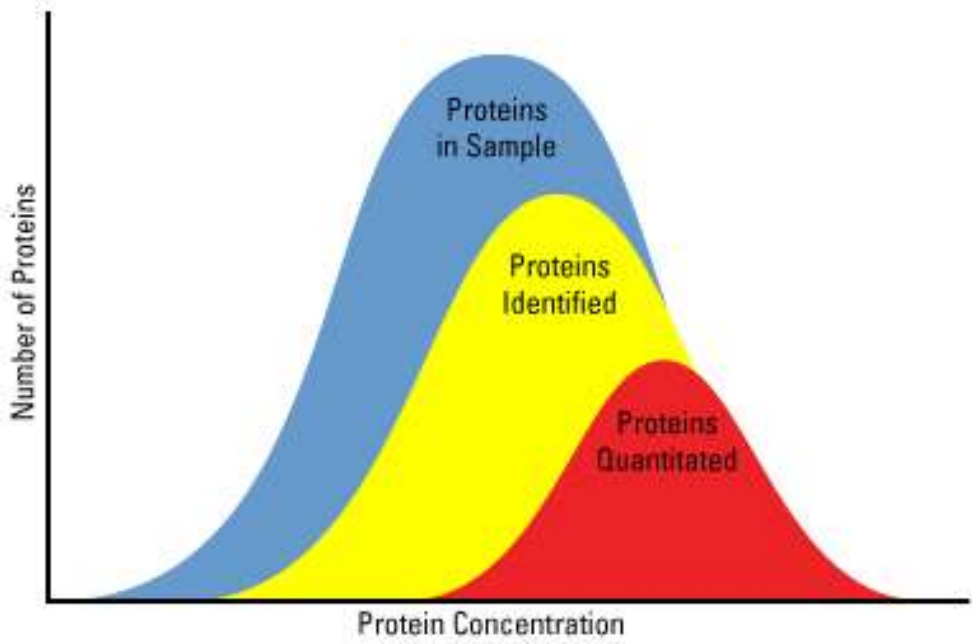
Evolution of proteomics performances

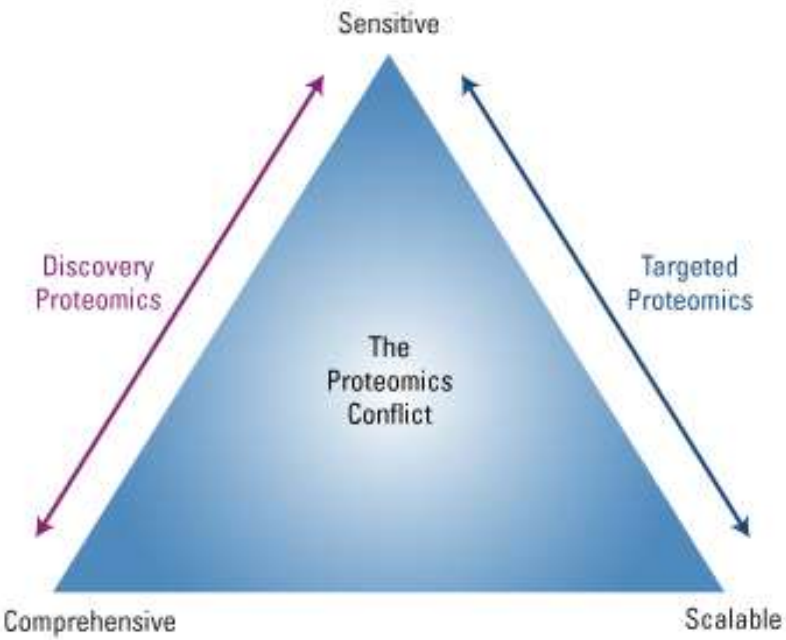




Hebert *et al.* 2009

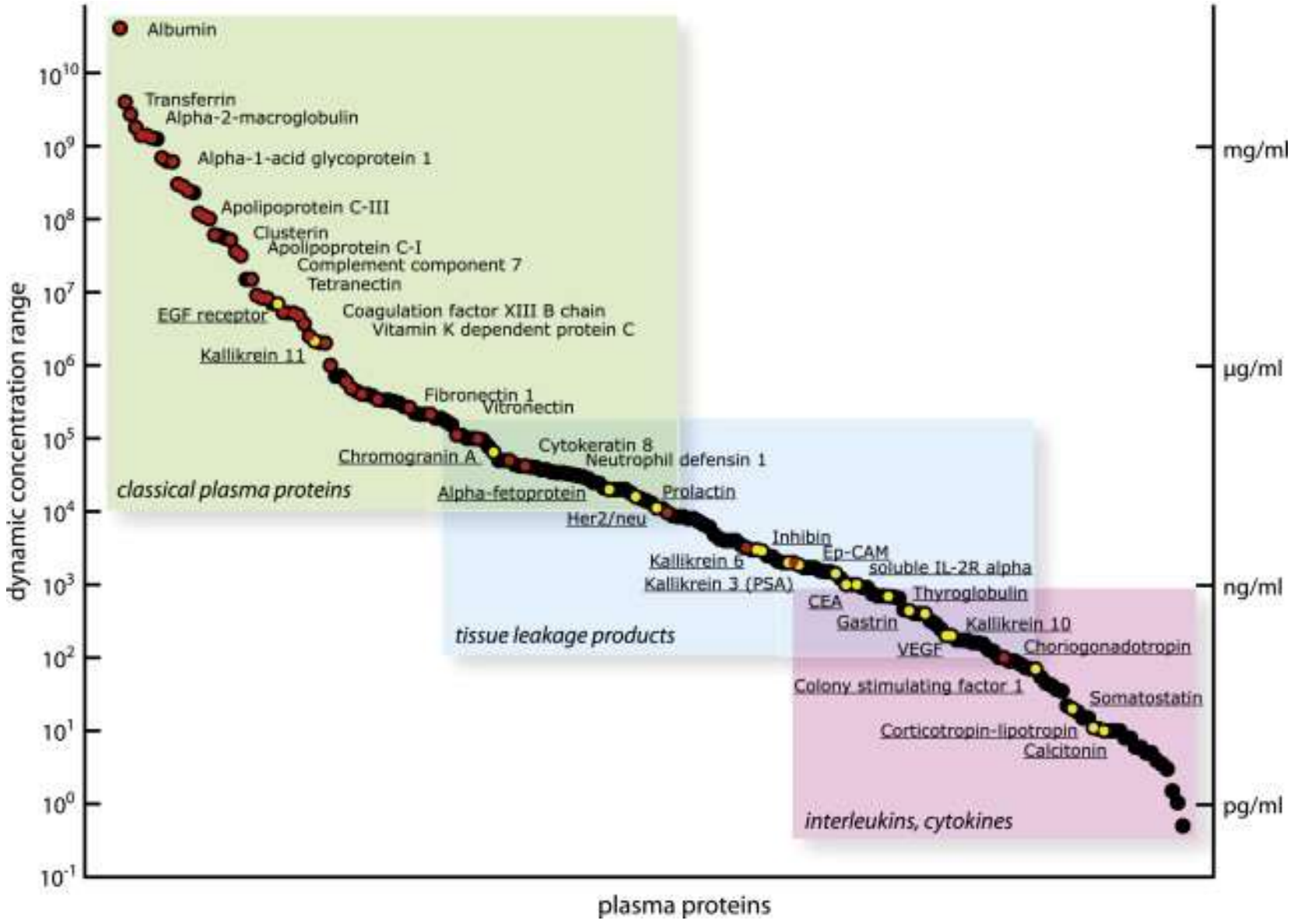
Inherent dilemma linked to proteomics



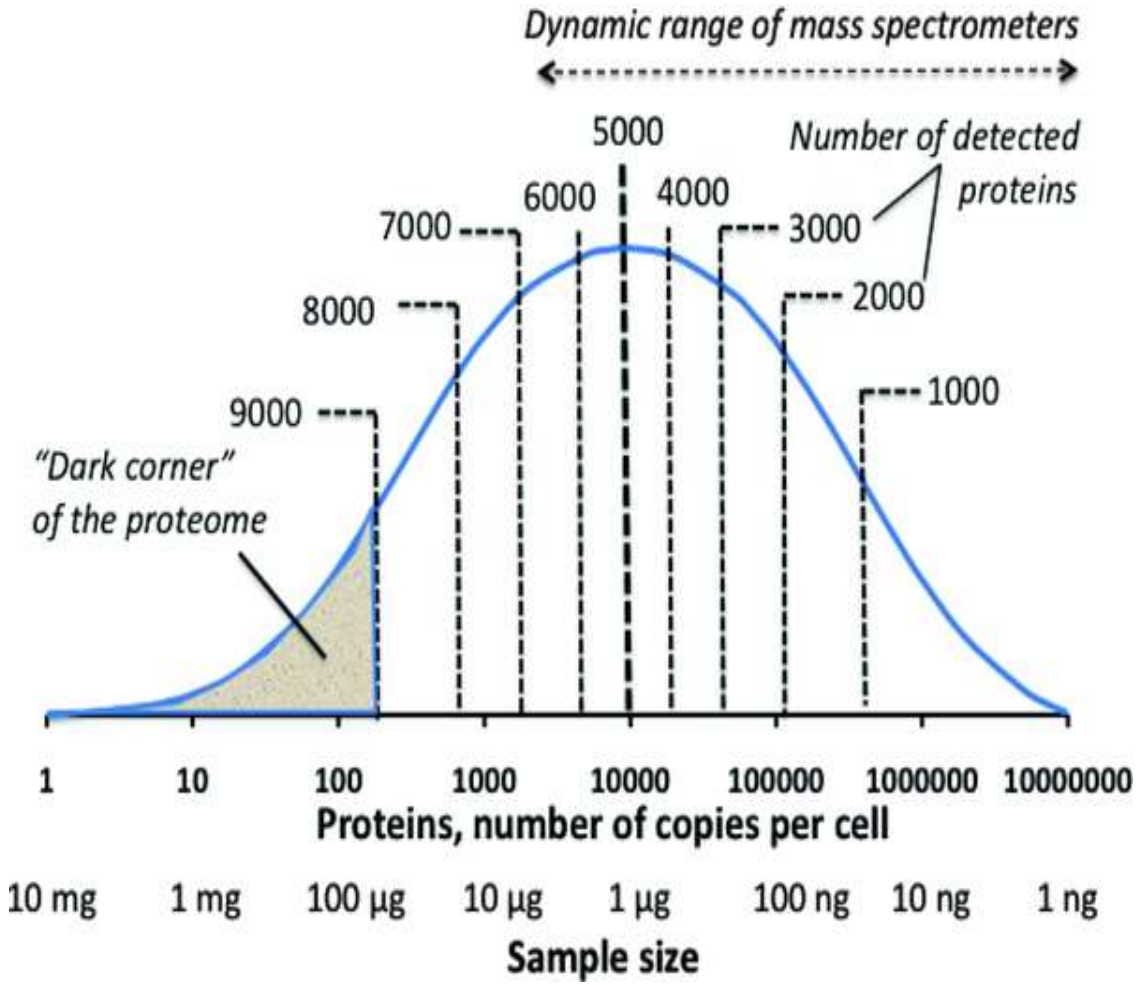


http://www.piercenet.com/method/quantitative-proteomics

Dynamic range in human plasma



Proteomics and proteome coverage



Zubarev *et al.* 2013

Dynamic range of proteomes

Key questions in proteomics

- What is the protein content of my biological sample?

  => problem of identification

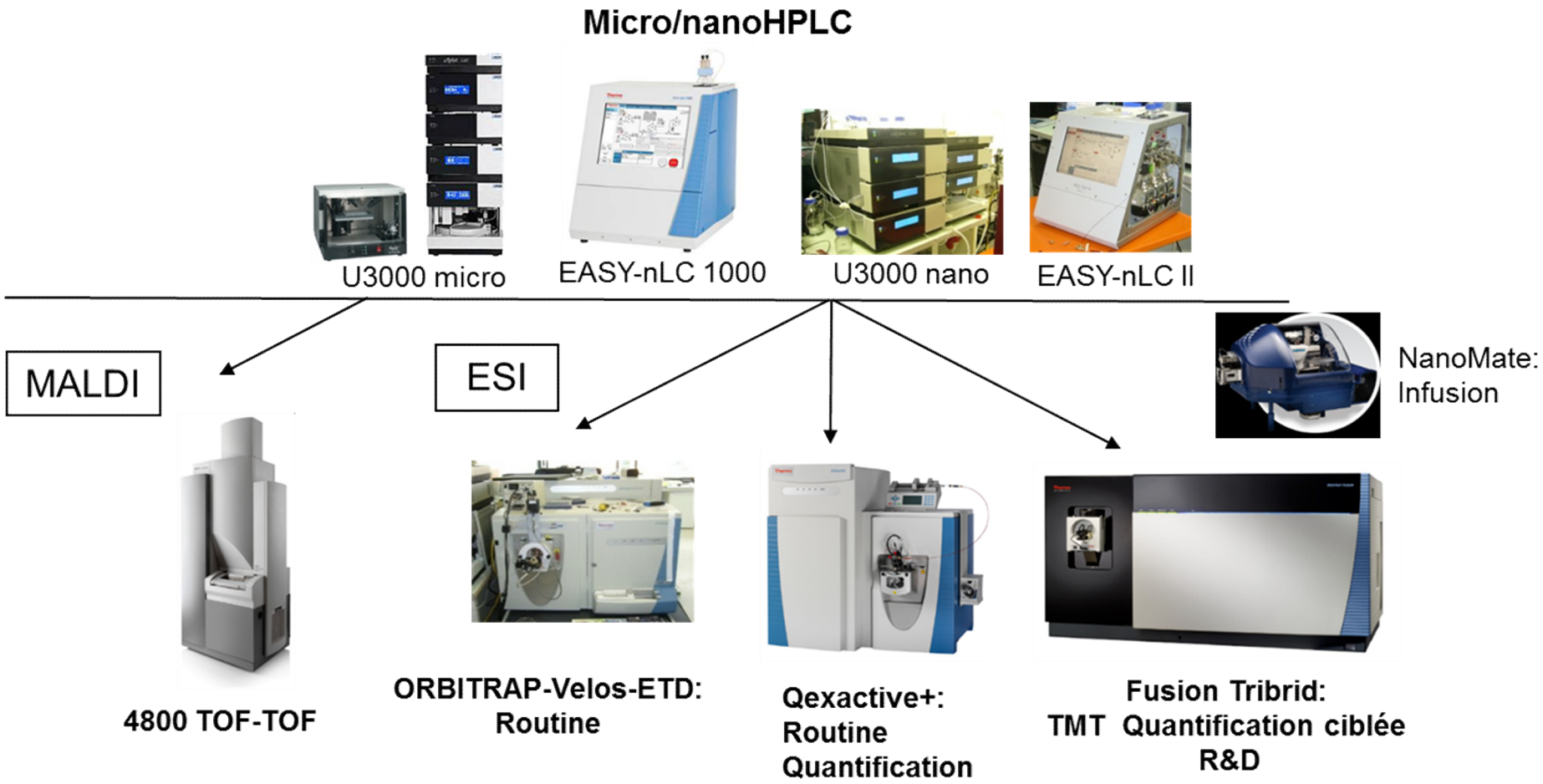
- What is the abundance of my protein of interest?

  => quantification

- Relative question: What are the protein abundance variations of the proteomes studied?
- What are the partners of my protein of interest?
- Are there any signature proteins related to a particular biological process?

  => biomarkers identifications and quantifications

Instrumentations

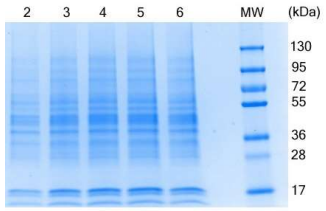


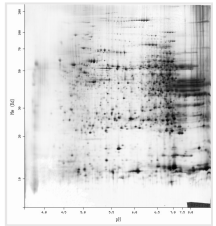
Proteomics workflows

Top – Down

(< 50000 Da in particular conditions)

Proteins





Peptides

HPLC

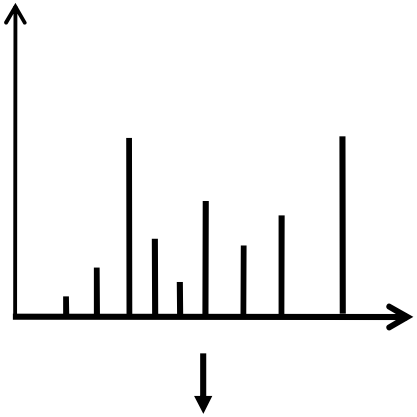
PROTEASES

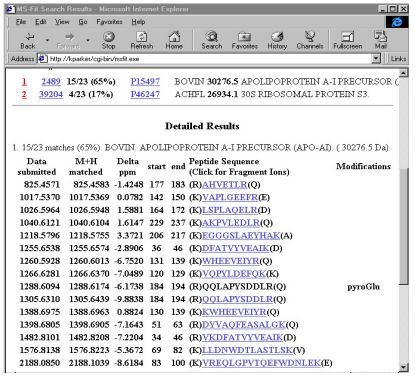
Bottom – Up

(< 4000 Da)

MS acquisition

- MALDI TOF/TOF
- Orbitrap





1 2489 15/23 (65%) P15497 BOVIN 30276.5 APOLIPOPROTEIN A-I PRECURSOR (
2 39204 4/23 (17%) P46247 ACHFL 26934.1 30S RIBOSOMAL PROTEIN S3.

Detailed Results

1. 15/23 matches (65%). BOVIN. APOLIPOPROTEIN A-I PRECURSOR (APO-AI). (30276.5 Da)

Data submitted	M+H matched	Delta ppm	start	end	Peptide Sequence (Click for Fragment Ions)	Modifications
825.4571	825.4583	-1.4248	177	183	(R)AHVETLR(Q)	
1017.5370	1017.5369	0.0782	142	150	(K)VAPLGEEFR(E)	
1026.5964	1026.5948	1.5881	164	172	(K)LSPLAQELR(D)	
1040.6121	1040.6104	1.6147	229	237	(K)AKPVLEDLR(Q)	
1218.5796	1218.5755	3.3721	206	217	(K)EGGGSLAEYHAK(A)	
1255.6538	1255.6574	-2.8906	36	46	(K)DFATVYVEAIK(D)	
1260.5928	1260.6013	-6.7520	131	139	(K)WHEEVEIYR(Q)	
1266.6281	1266.6370	-7.0489	120	129	(K)VQPVLDEFQK(K)	
1288.6094	1288.6174	-6.1738	184	194	(R)QQLAPYSDDLR(Q)	pyroGlu
1305.6310	1305.6439	-9.8838	184	194	(R)QQLAPYSDDLR(Q)	
1388.6975	1388.6963	0.8824	130	139	(K)KWHEEVEIYR(Q)	
1398.6805	1398.6905	-7.1643	51	63	(R)DYVAQFEASALGK(Q)	
1482.8101	1482.8208	-7.2204	34	46	(R)VKDFATVYVEAIK(D)	
1576.8138	1576.8223	-5.3672	69	82	(K)LLDNWDTLASTLSK(V)	
2188.0850	2188.1039	-8.6184	83	100	(K)VREQLGPVTQEFWDNLEK(E)	

Identification
Quantification

The strategy is dependent of the sample complexity

Search engine (Mascot, Sequest, Peaks, Maxquant, OMSSA, Prosight, Byonik)

BOTTOM-UP PROTEOMICS: PRO'S AND CON'S

Advantages

- Less sophisticated instrumentation and expertise
- High throughput
- More info about proteins with "extreme" phys.-chem. properties (hydrophobic, Hi/Low MW, acidic/basic)

Disadvantages

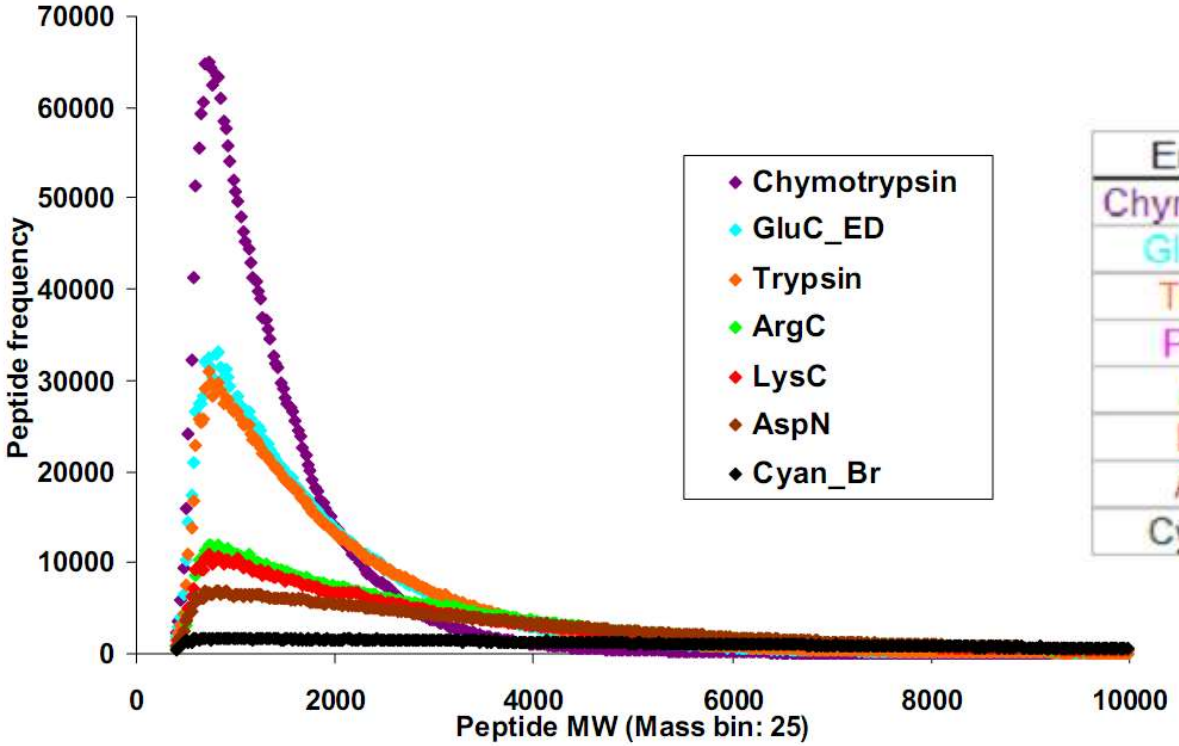
- Confidence in protein ID strongly depends on restriction criteria (subjective; potential bias)
- Since protein ID is often done by 1-2 peptides, PTM and isoform information is often lost

Cleavage rules of proteases

Enzyme or Reagent	Cleaves where?	Exceptions
Trypsin	C-terminal side of K or R	if P is C-term to K or R
Trypsin	(C-term to K/R, even before P)	C-terminal side of K or R
Trypsin (higher specificity)	C-terminal side of K or R	if P is C-term to K or R; after K in CKY, DKD, CKH, CKD, KKR; after R in RRH, RRR, CRK, DRD, RRF, KRR
Lys C	C-terminal side of K	
CNBr	C-terminal side of M	
Arg C	C-terminal side of R	if P is C-term to R
Asp N	N-terminal side of D	
Asp N + N-terminal Glu	N-terminal side of D or E	
Glu C (bicarbonate)	C-terminal side of E	if P is C-term to E, or if E is C-term to E
Glu C (phosphate)	C-terminal side of D or E	if P is C-term to D or E, or if E is C-term to D or E
Chymotrypsin	(C-term to F/Y/W/M/L, not before P, not after Y if P is C-term to Y)	
	C-terminal side of F, L, M, W, Y	if P is C-term to F, L, M, W, Y, if P is N-term to Y
Chymotrypsin (C-term to F/Y/W/, not before P, not after Y if P is C-term to Y)		
	C-terminal side of F, Y, W	if P is C-term to F, Y, W, if P is N-term to Y
Trypsin/Chymotrypsin (C-term to K/R/F/Y/W, not before P, not after Y if P is C-term to Y)		
	C-terminal side of K, R, F, Y, W	if P is C-term to K, R, F, Y, W, if P is N-term to Y
Pepsin (pH 1.3)	C-terminal side of F, L	
Pepsin (pH > 2)	C-terminal side of F, L, W, Y, A, E, Q	
Proteinase K	C-terminal side of A, C, G, M, F, S, Y, W	

PEPTIDE LENGTH AND NUMBER OF PEPTIDES GENERATED DEPENDING ON ENZYME USED FOR DIGESTION

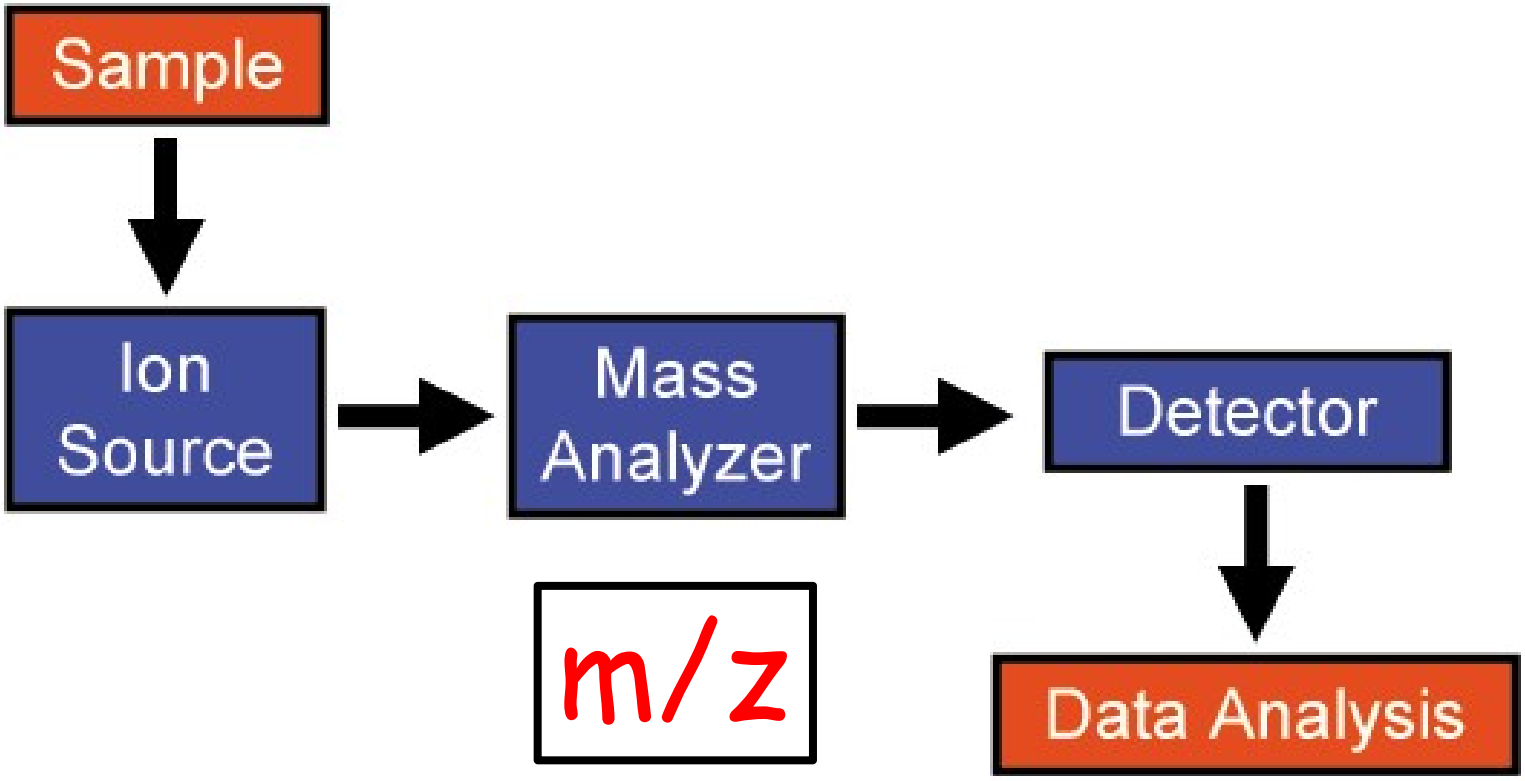
Other enzymes with more or less specific cleavage:



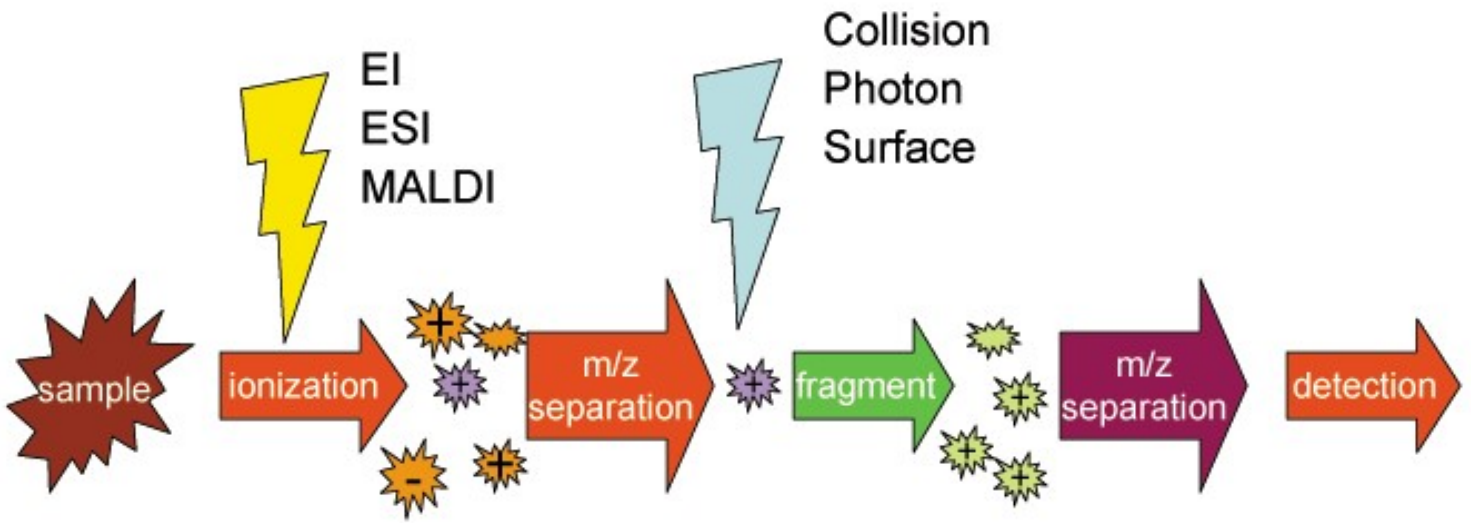
Enzyme	Cleaves at:
Chymotrypsin	FWYL
GluC_ED	ED
Trypsin	KR
Pepsin	FLIWY
ArgC	R
LysC	K
AspN	D
Cyan_Br	M

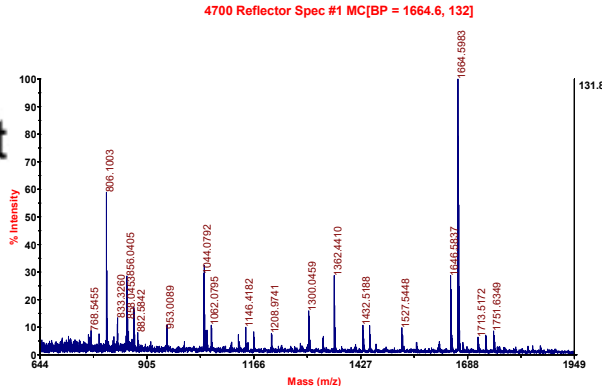
Advantages of a new proteomic approach that uses accurate mass measurements, LC retention time, isoelectric point and dual enzymatic digestion. Petritis K. et. al., Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA 99352; ASMS'2007 poster presentation http://www.chem.agilent.com/Library/posters/Public/Petritis_ASMS_2007.pdf

What is MS?

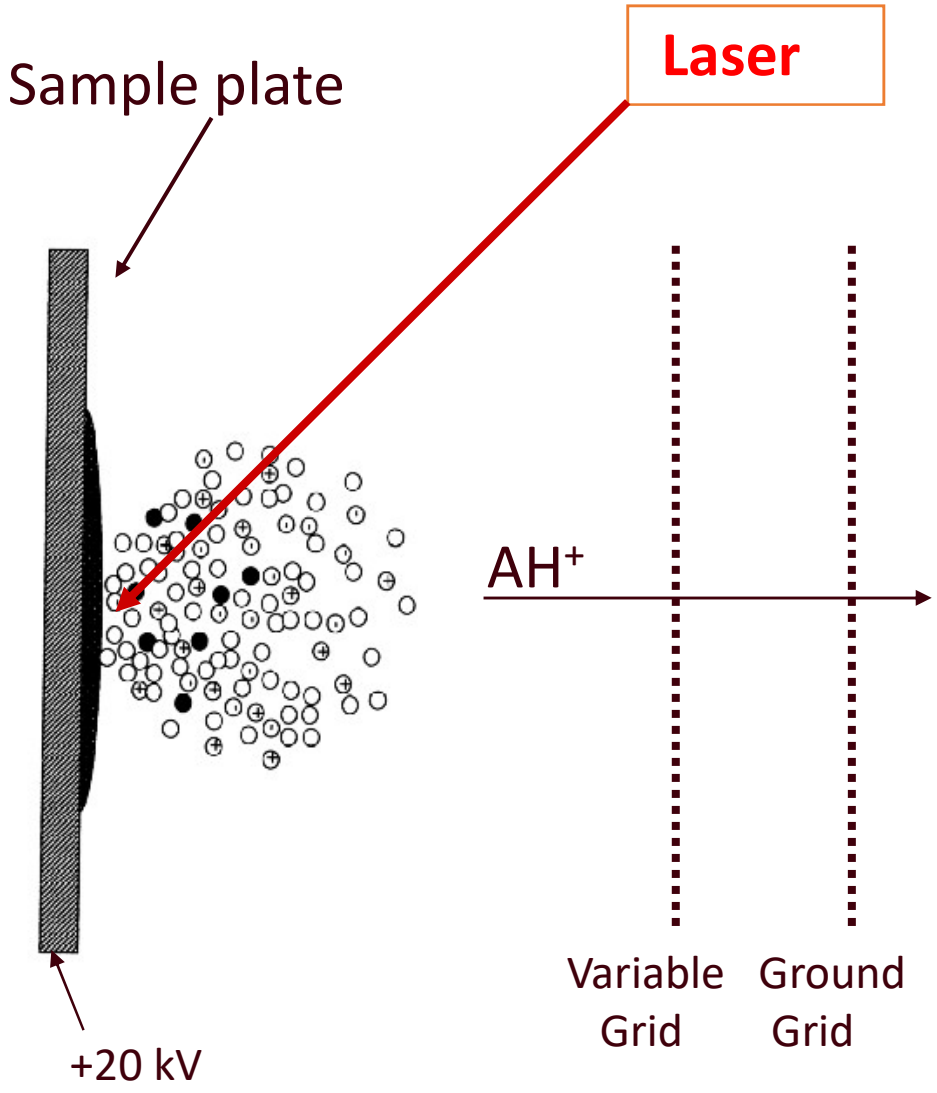


MS and MS/MS


EI
ESI
MALDI
Collision
Photon
Surface
sample
ionization
m/z
separation
fragment
m/z
separation
detection
MS1
MS2
Precursor
ion
Product
ion


4700 Reflector Spec #1 MC[BP = 1664.6, 132]
% Intensity
Mass (m/z)
644
905
1166
1427
1688
1949
806.1003
1664.5983
1044.0792
1362.4410
1062.0795
1300.0459
1432.5188
1527.5448
1146.4182
1208.9741
953.0089
882.5842
768.5455
1713.5172
1751.6349

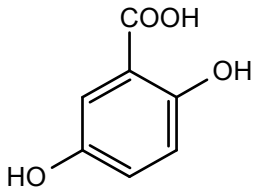
MALDI ionization (Matrix Assisted Laser Desorption Ionization)



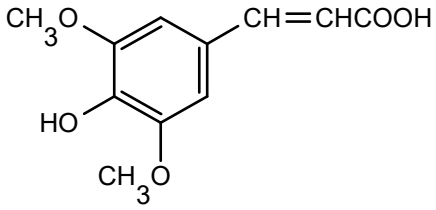
1. L'échantillon (A) est mélangé avec un excès de matrice (M) et séché sur la plaque MALDI
2. Le flash Laser ionise les molécules de matrice
3. Les molécules d'échantillon sont ionisées par transfert de protons de la matrice:

   $MH^+ + A \rightarrow M + AH^+$.

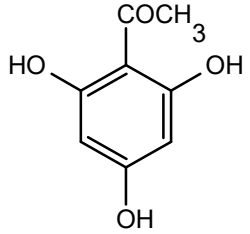
MALDI-TOF Matrix



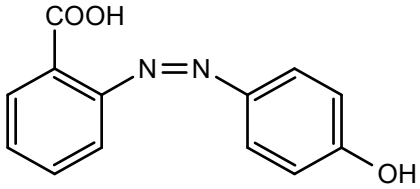
2,5-dihydroxybenzoic acid
(2,5-DHB)



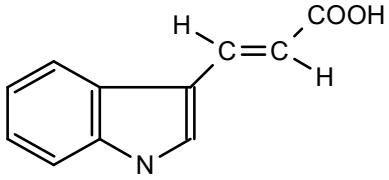
Sinapinic acid (3,5-Dimethoxy-4-hydroxy cinnamic acid)



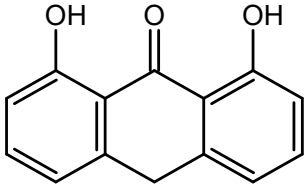
2,4,6-trihydroxy acetophenone (THAP)



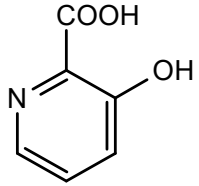
2-(4-hydroxyphenylazo)-benzoic acid
(HABA)



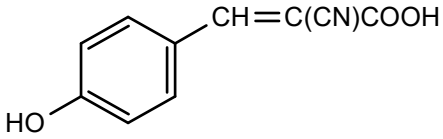
trans-3-indoleacrylic acid



Dithranol



3-hydroxypicolinic acid (3-HPA)

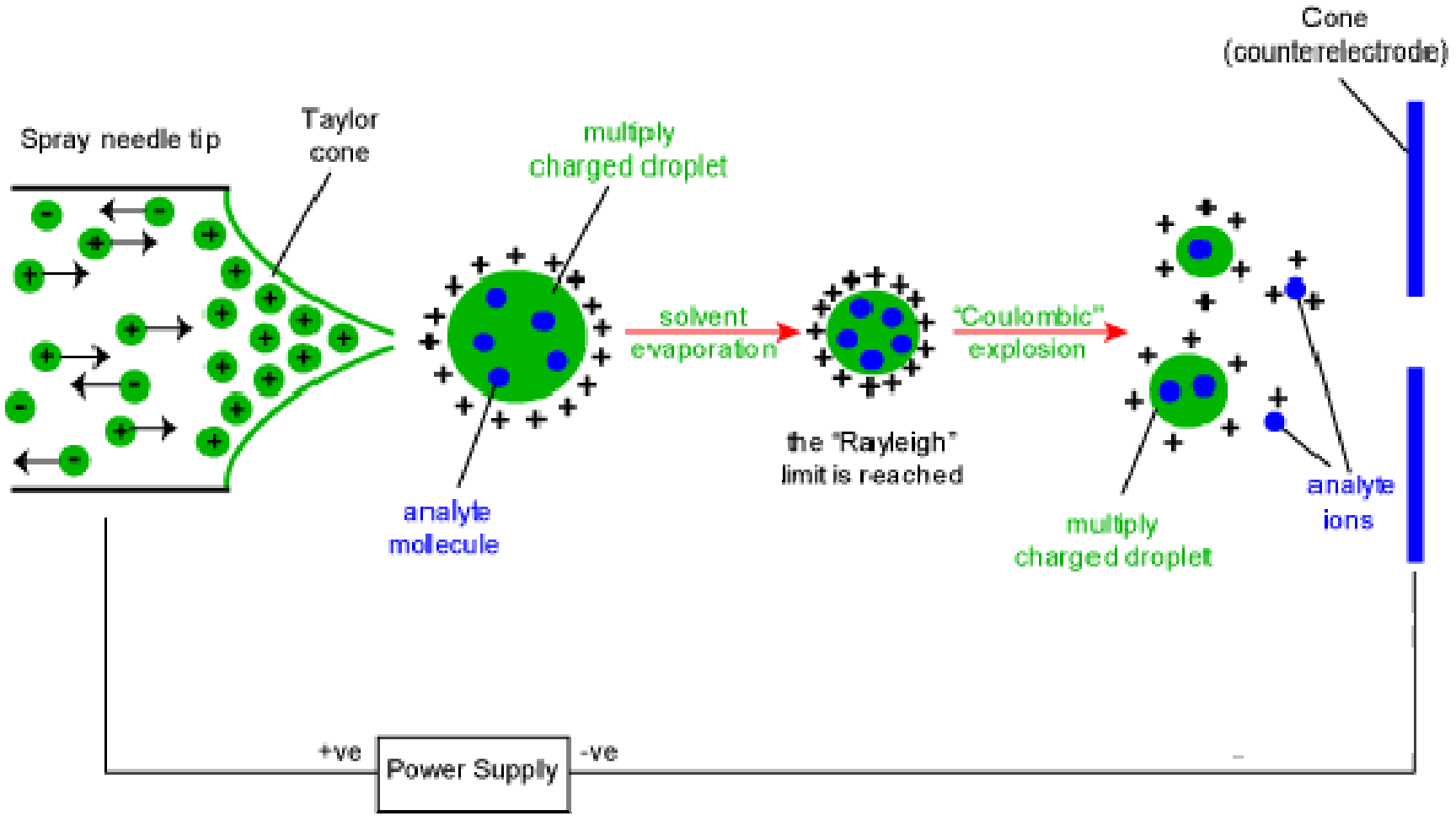


α-cyano-4-hydroxycinnamic acid

Matrix choices

α-Cyano-4-hydroxy-cinnamic acid (CHCA)	Peptides<10kDa
Sinapinic Acid	Proteins >10kDa
2,5-Dihydroxybenzoic acid (DHB)	Neutral carbohydrates, Synthetic Polymers
"Super DHB"	Proteins, Glycosylated proteins
3-Hydroxypicolinic acid	Oligonucleotides
HABA	Proteins, Oligosaccharides

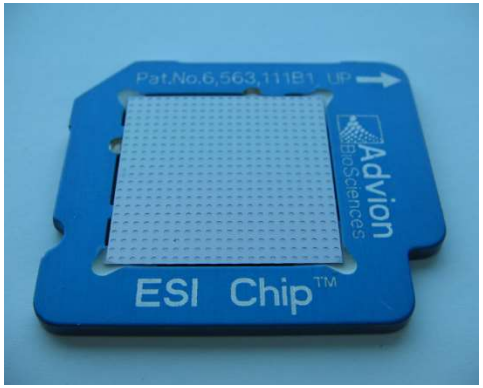
Ionization by electrospray



Electrospray and nanospray sources

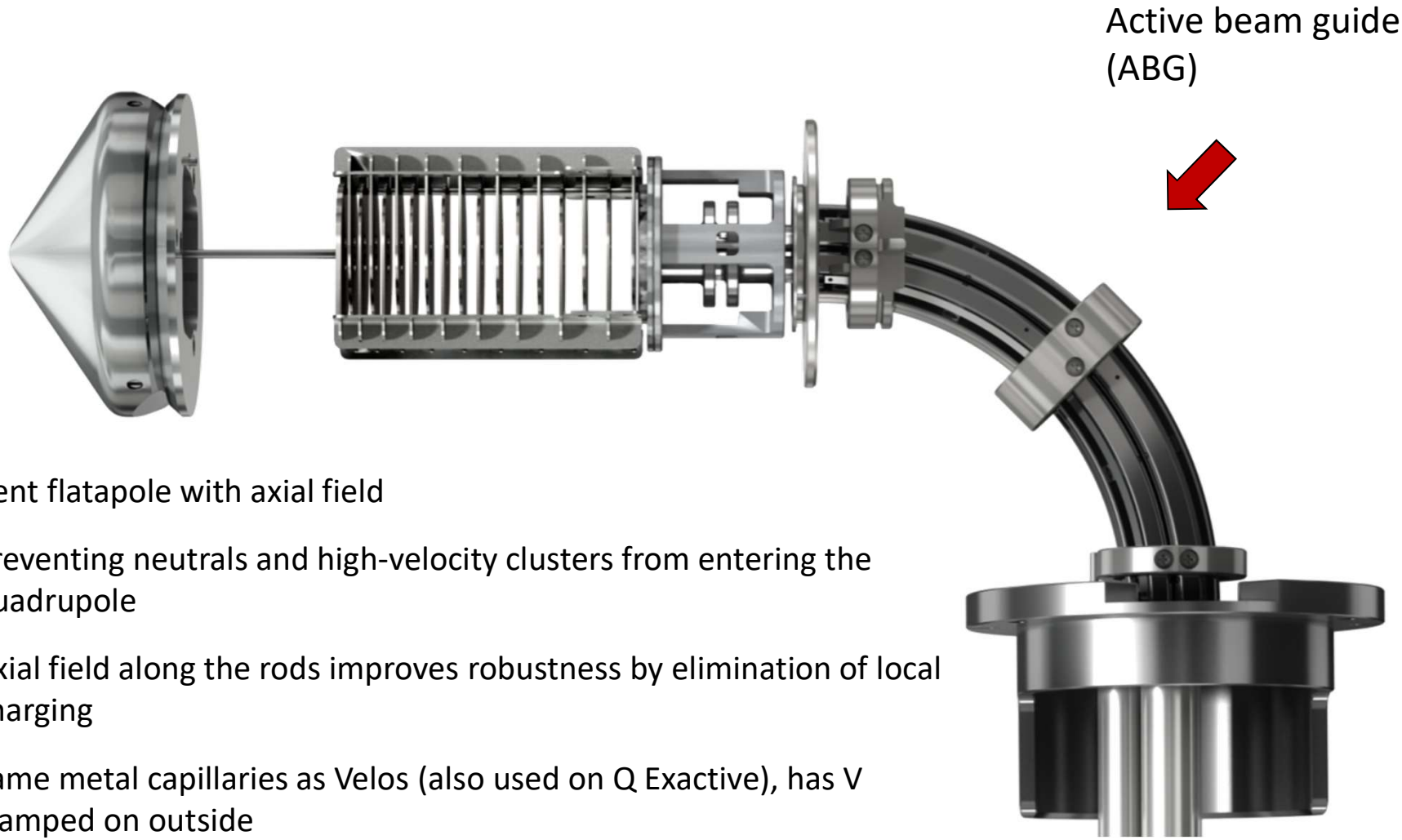
Electrospray

nanospray



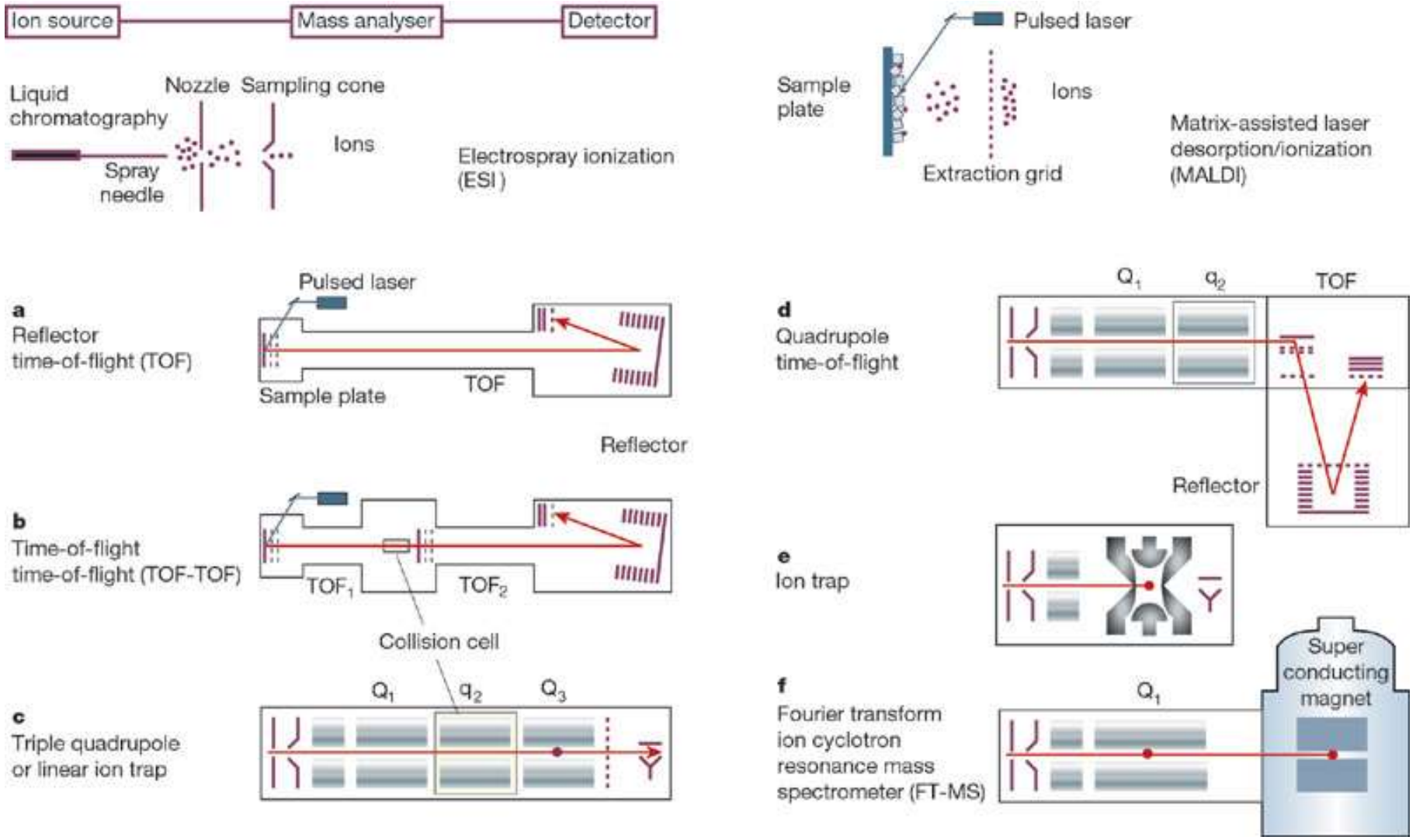
nanospray

“Active Beam Guide” transmission

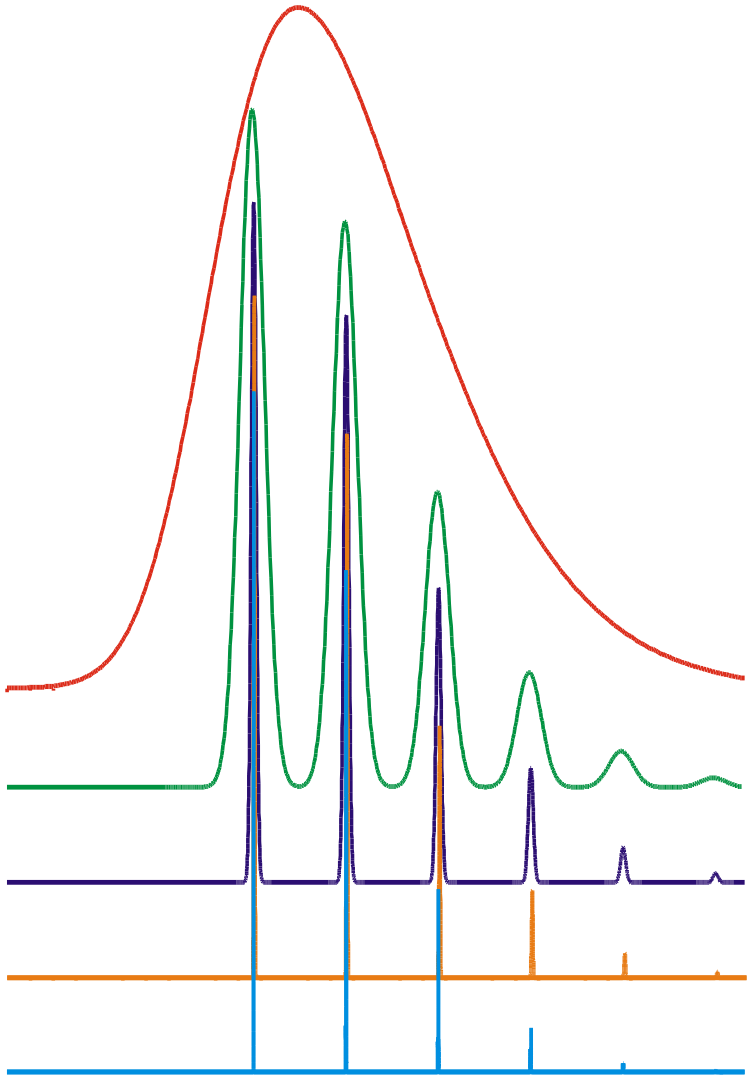


- Bent flatapole with axial field
- Preventing neutrals and high-velocity clusters from entering the quadrupole
- Axial field along the rods improves robustness by elimination of local charging
- Same metal capillaries as Velos (also used on Q Exactive), has V stamped on outside

Different instrumental design



Importance of spectral resolution

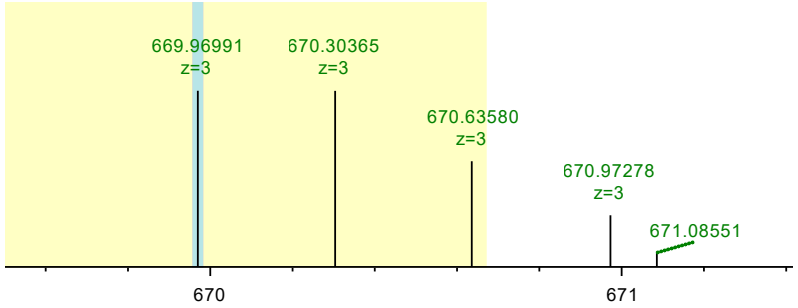


Resolution

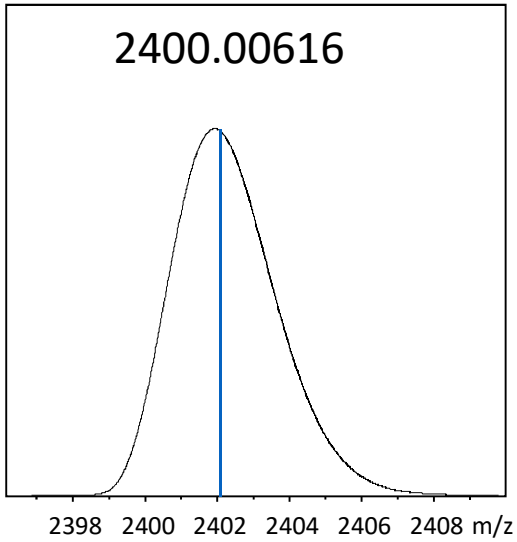
1.000	linear TOF w/o DE
5.000	reflector TOF w/o DE
25.000	reflector TOF with DE
125.000	FTMS wideband mode
1.000.000	FTMS high-res mode

Natural abundance of atoms isotopes in proteins

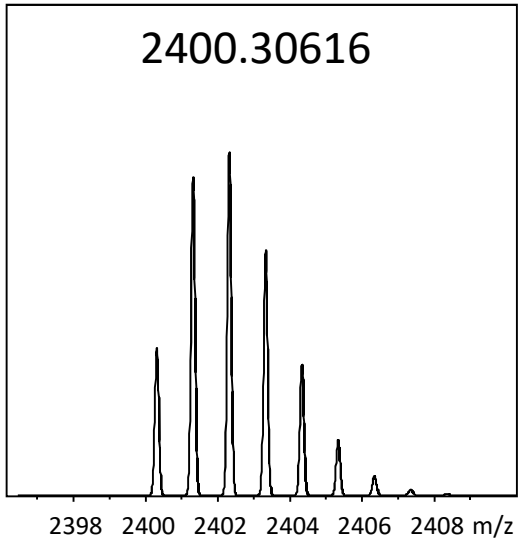
Name	Symbol	Mass (Da)	Abundance (%)
Hydrogen	H	1.007825	99.9885
Deuterium	H	2.014102	0.0115
Carbon	C	12.000000	98.9300
	C	13.003355	1.0700
Nitrogen	N	14.003074	99.6320
	N	15.000109	0.3680
Oxygen	O	15.994915	99.7570
	O	16.999132	0.0380
	O	17.999160	0.2050
Phosphorus	P	30.973762	100.0000
Sulfur	S	31.973762	94.9300
	S	32.971458	0.7600
	S	33.967867	4.2900
	S	35.967081	0.0200



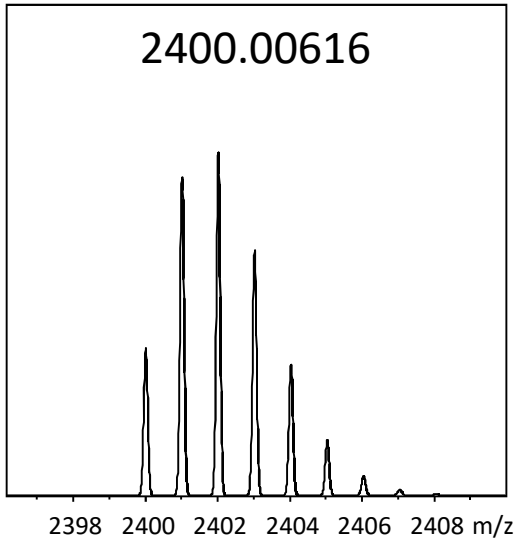
Resolution and mass accuracy



Poor resolution
High mass accuracy



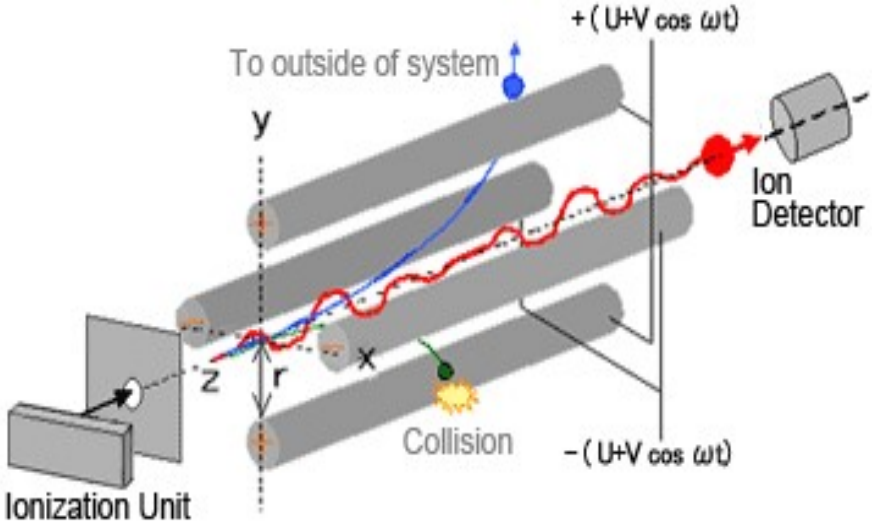
High resolution
Poor mass accuracy

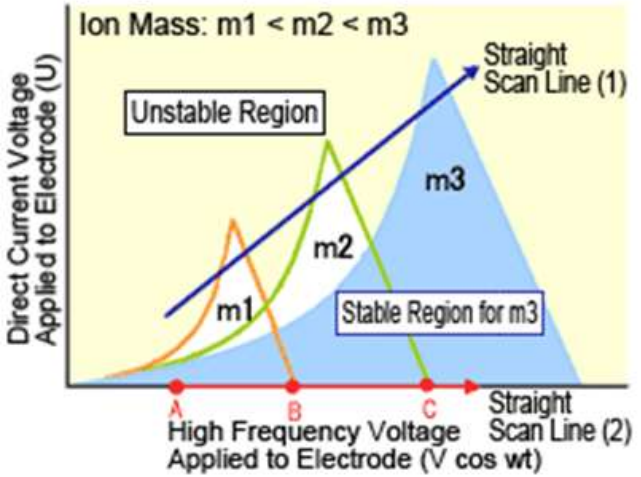


High resolution
High mass accuracy

High resolution makes it easier to achieve high mass accuracy – but high mass accuracy does not necessarily require high resolution! High resolution is only mandatory to avoid overlapping peaks.

Quadrupole analysers

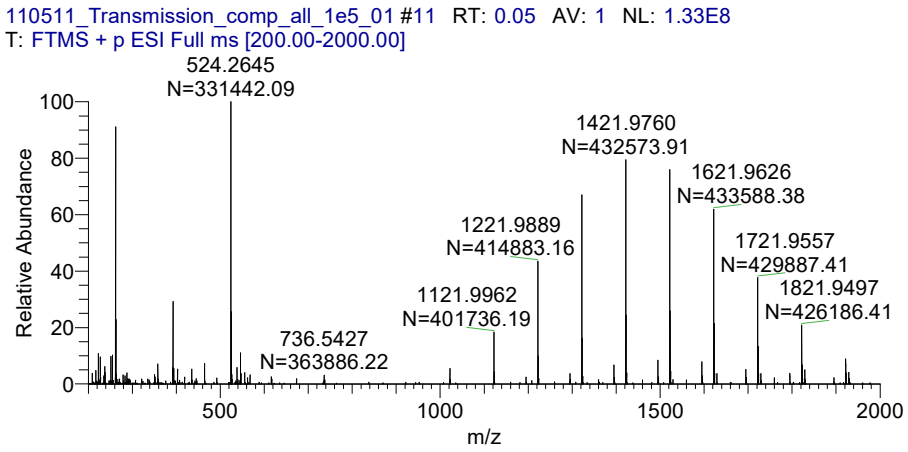




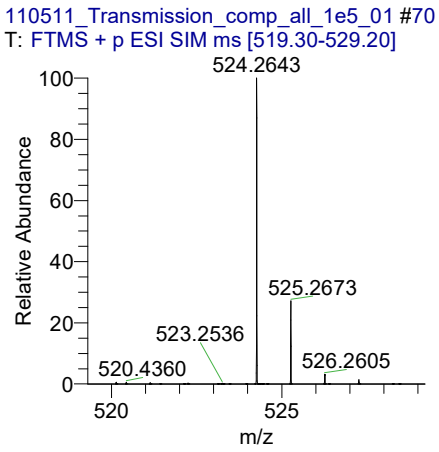
$\varphi 0 = U - V.\cos(2\pi f t)$

$\varphi = \varphi 0.\ (x^2 - y^2)/r0^2$

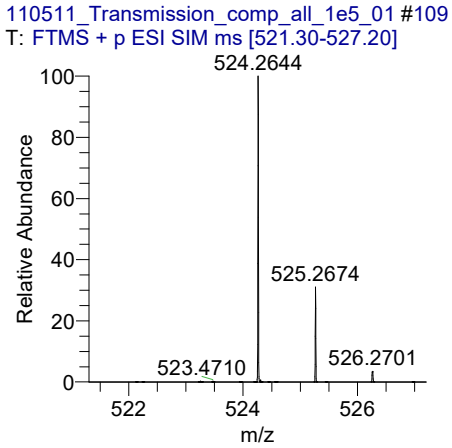
Isolation Width: Full MS to 1 amu (MRFA)



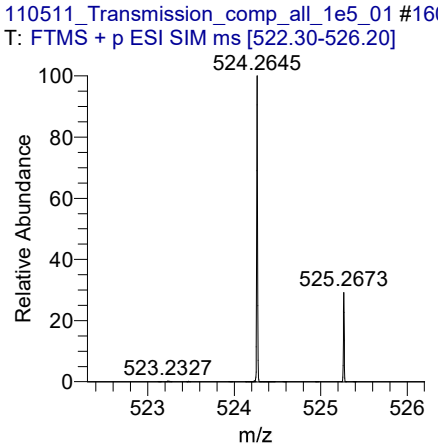
IsoW= 1800



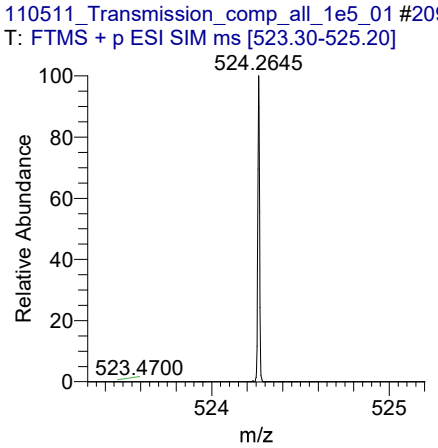
IsoW= 10



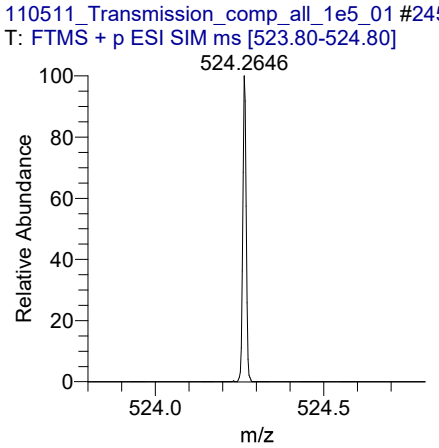
IsoW= 6



IsoW= 4

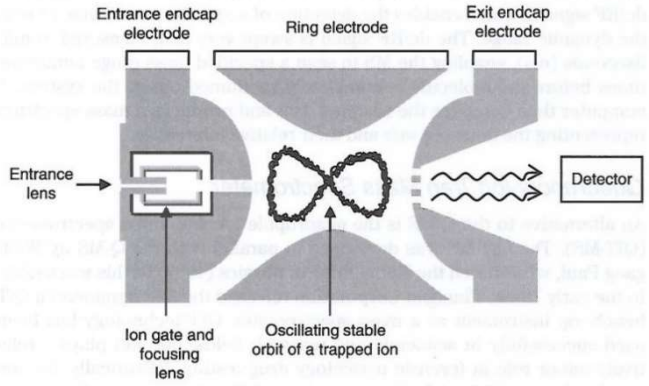


IsoW= 2

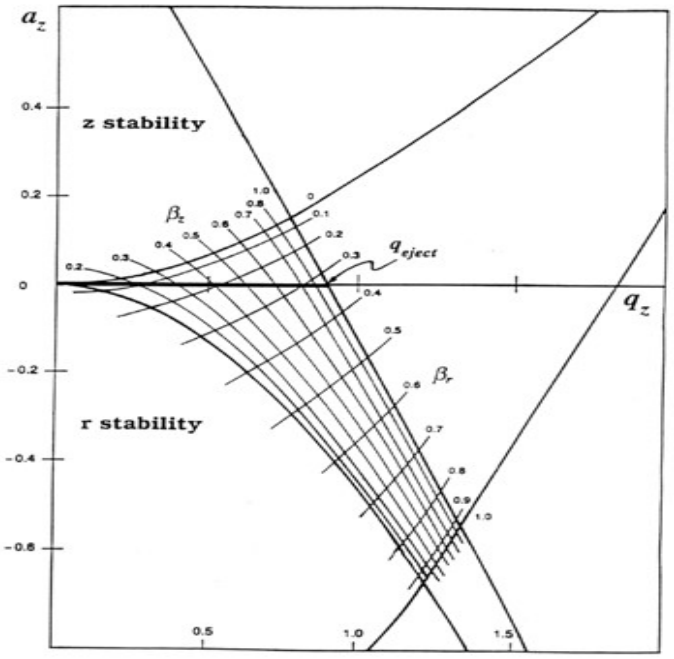


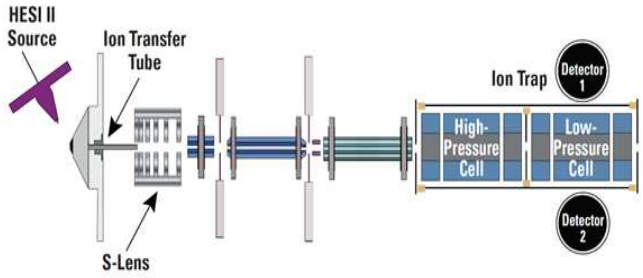
IsoW= 1

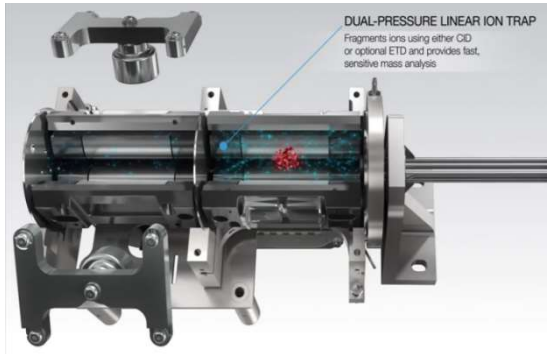
Different instrumental design



Trappe tridimensionnelle (Paul)

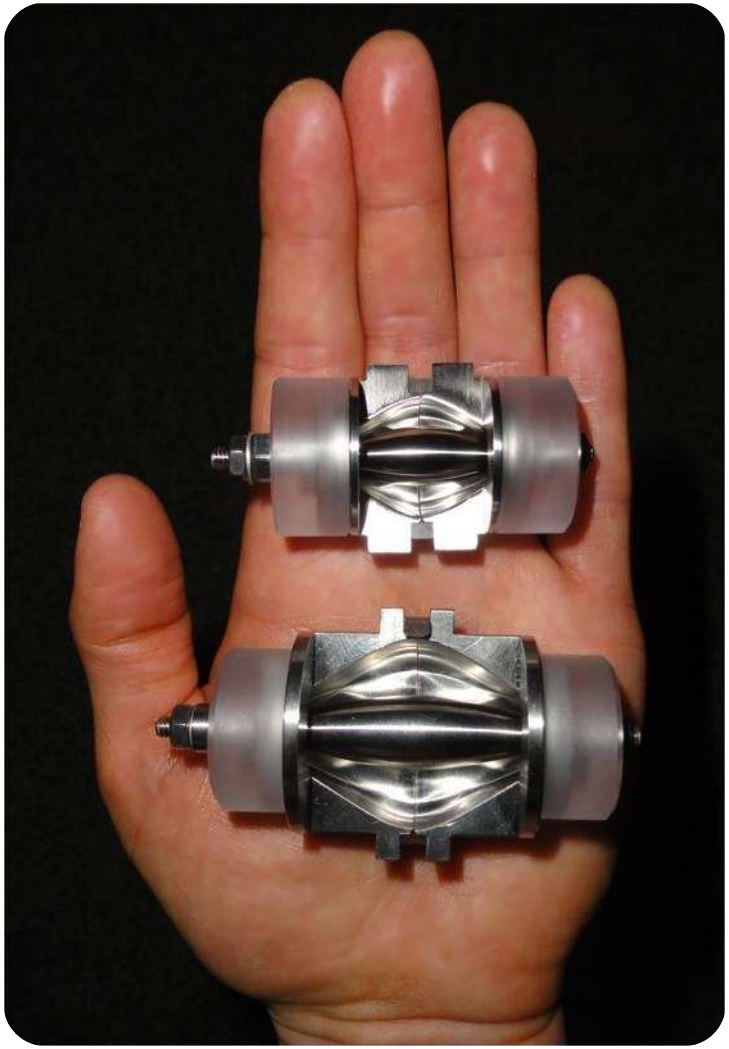


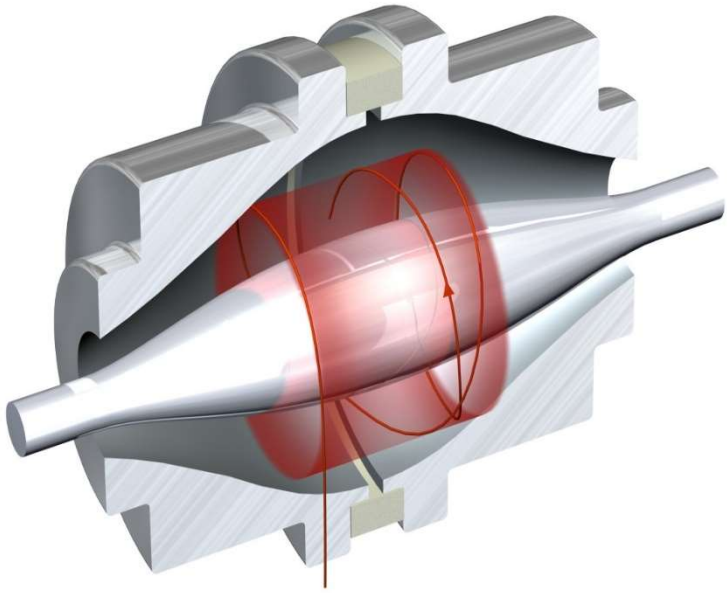




Trappe linéaire (double pression)

the orbitrap cell





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

Developpement of the orbitrap family

2007 LTQ Orbitrap XL and Discovery

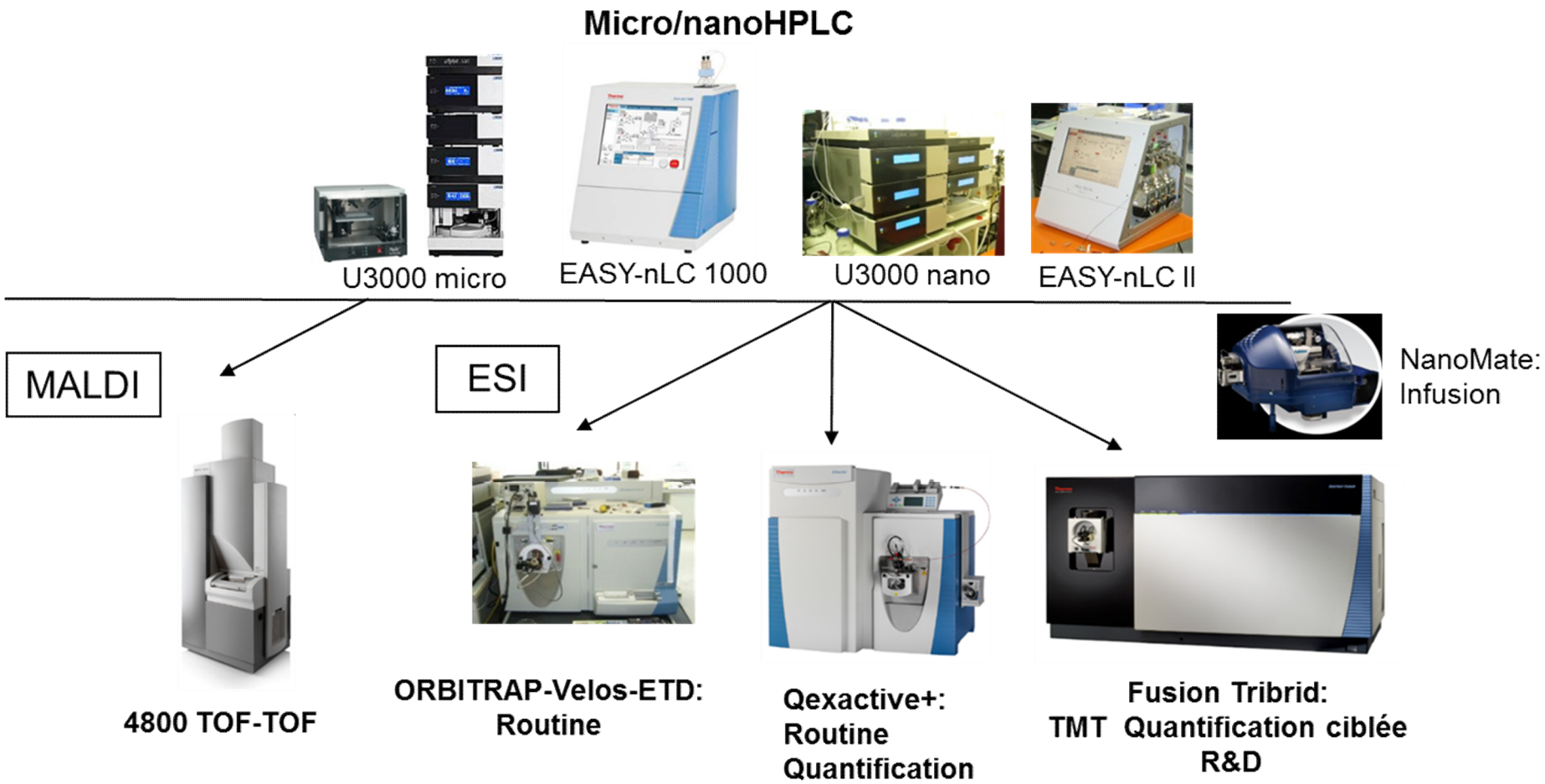
2008 LTQ Orbitrap XL ETD

2009 LTQ Orbitrap Velos

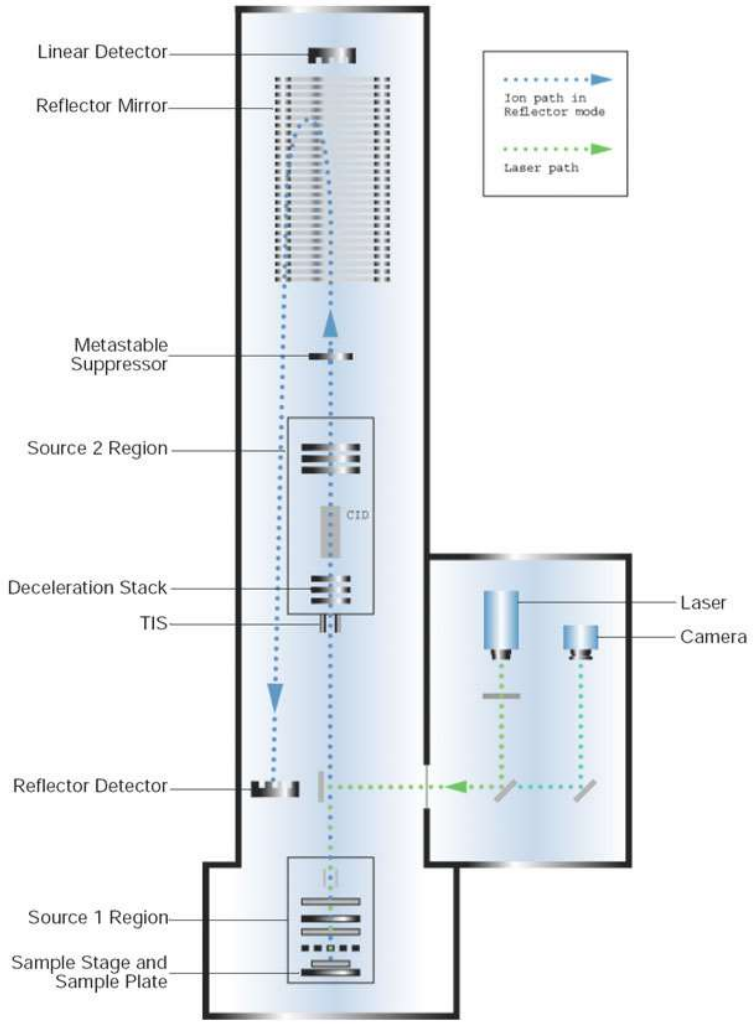
2011 LTQ Orbitrap Velos Pro
Orbitrap Elite

2013 ***Orbitrap Fusion Tribrid***

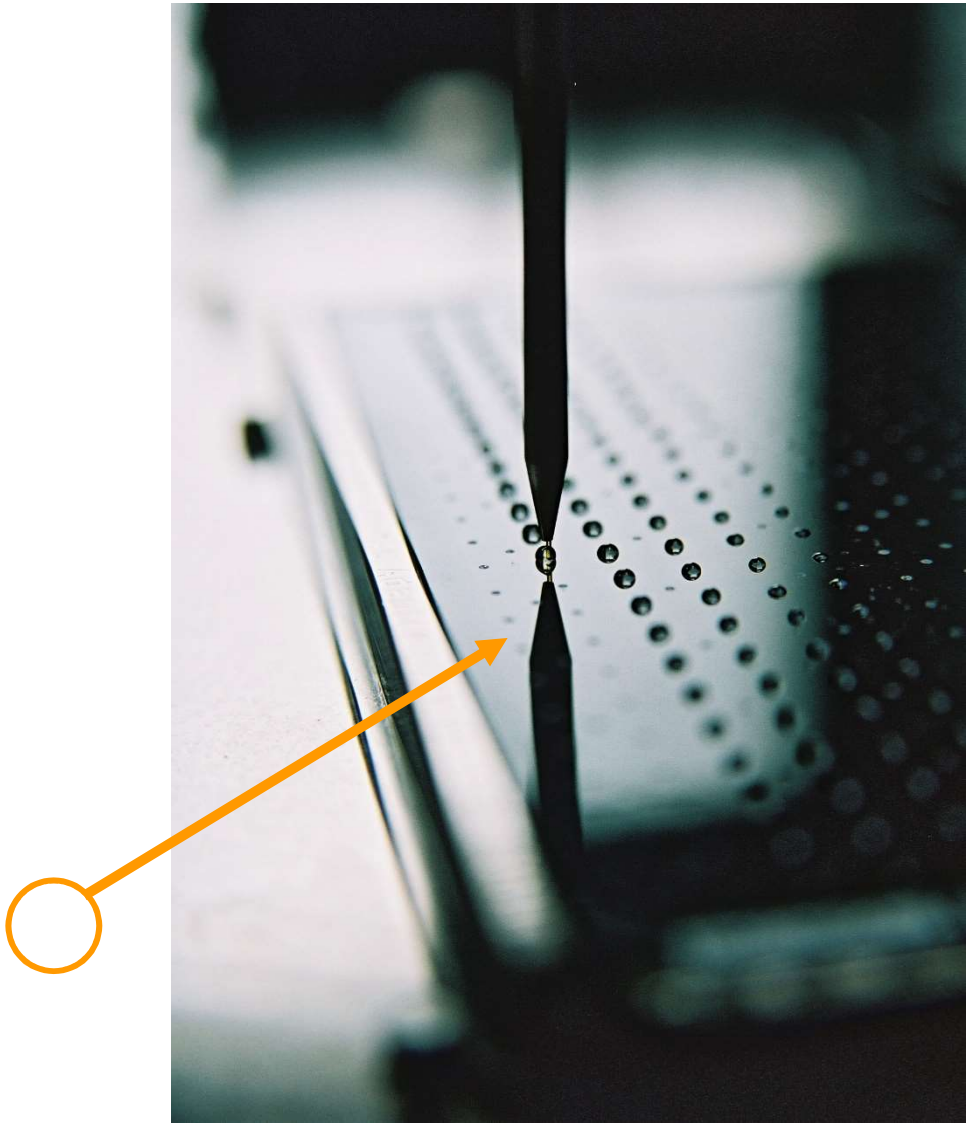
Instrumentations



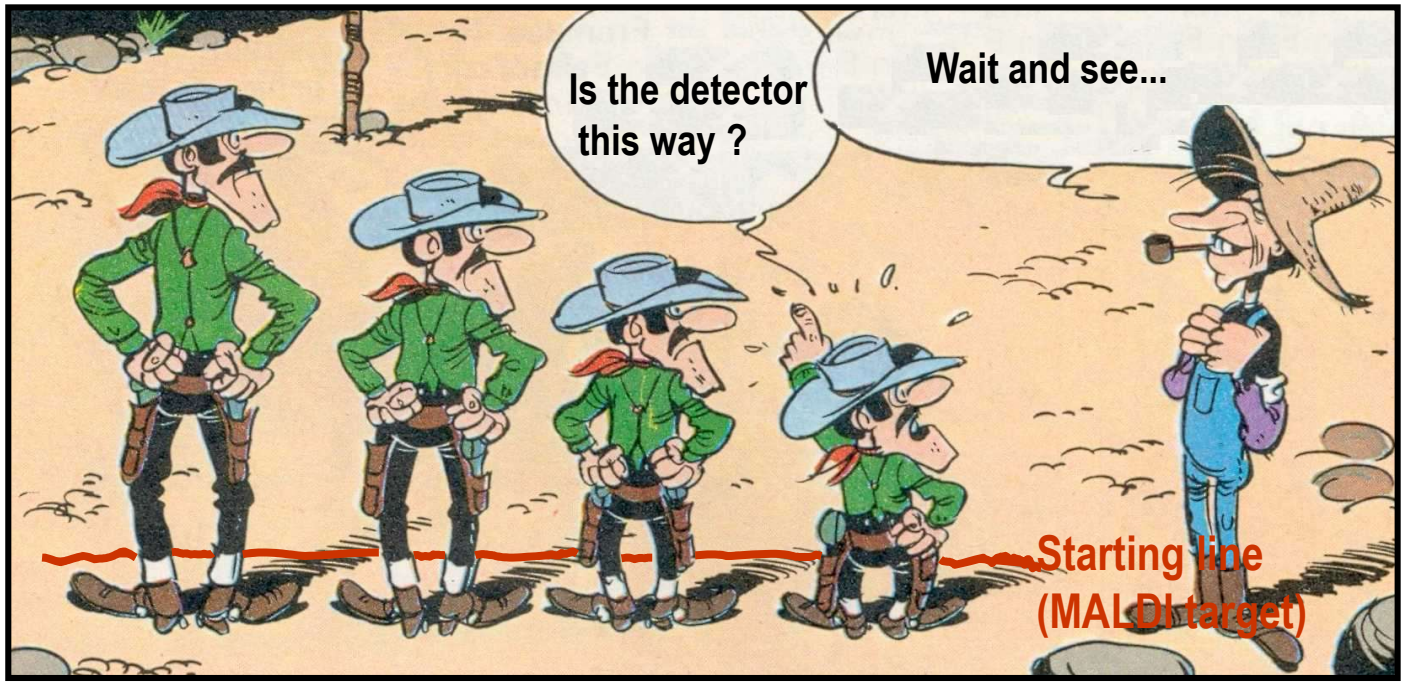
MALDI-TOF/TOF



Samples on MALDI plate



Time of flight – principles (TOF)

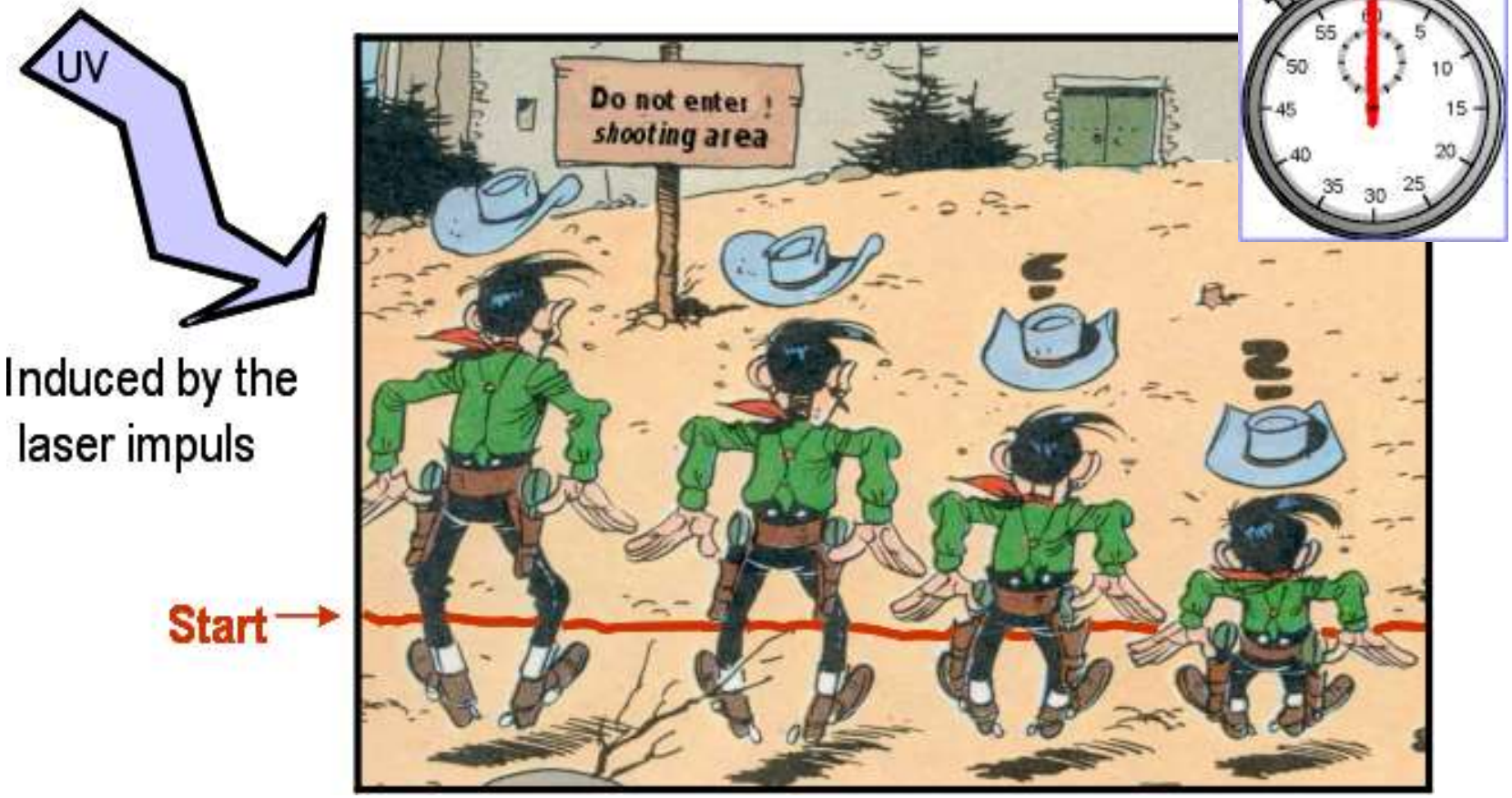


Remember : Mass of an ion is measured in the Dalton units !

Start !

Laser

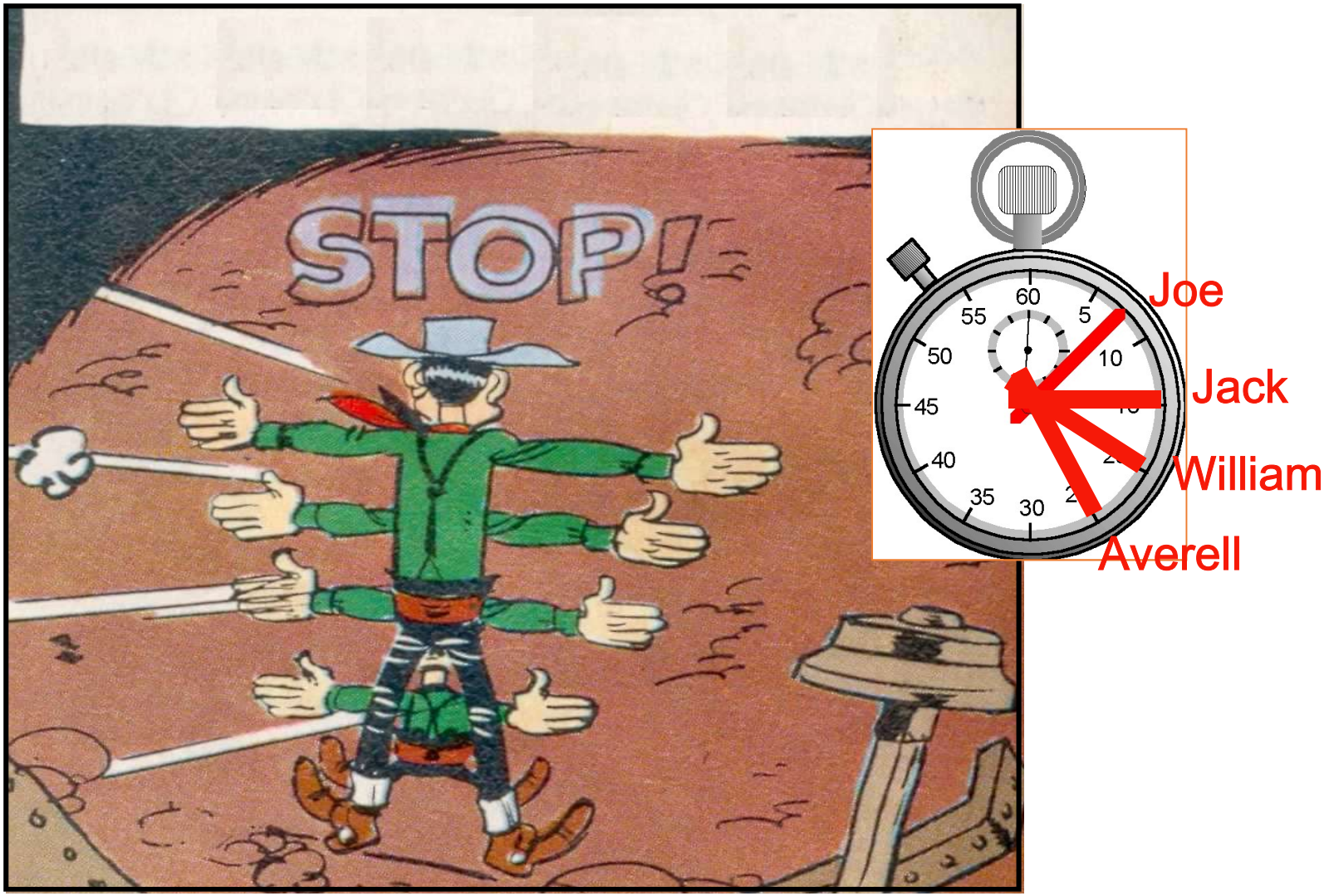
The desorptionevent



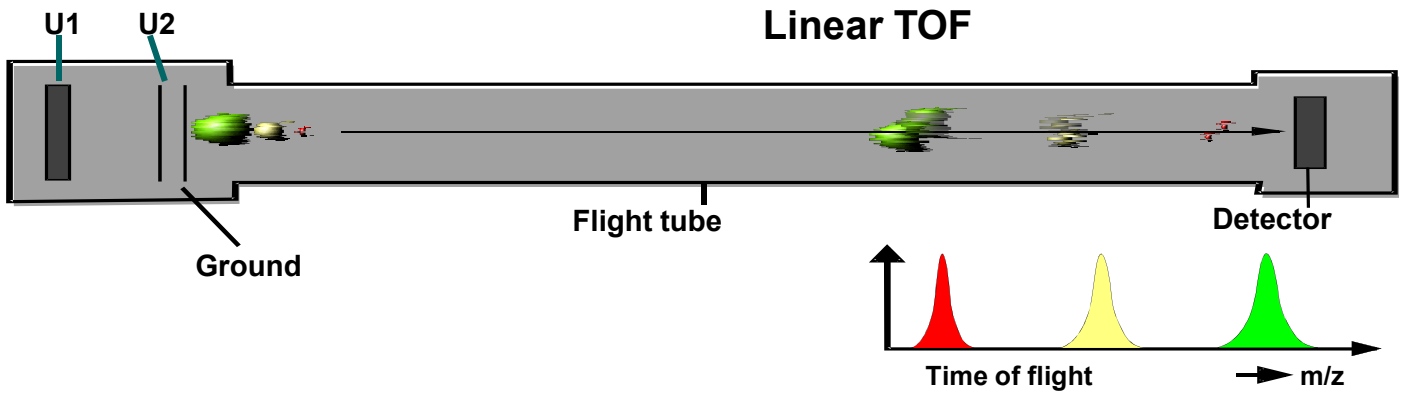
Ions in the time of flight (TOF)

Increasing MW

Ions in the detector



Ions analysis in linear mode



Electric field : $Ec = qU = 1/2\ mV^2$ Identical for all ions

(V=L/t L : tube lenght)

$\Rightarrow$ Simple relation $t^2 = mL^2/2qU$ = Constante x m/z

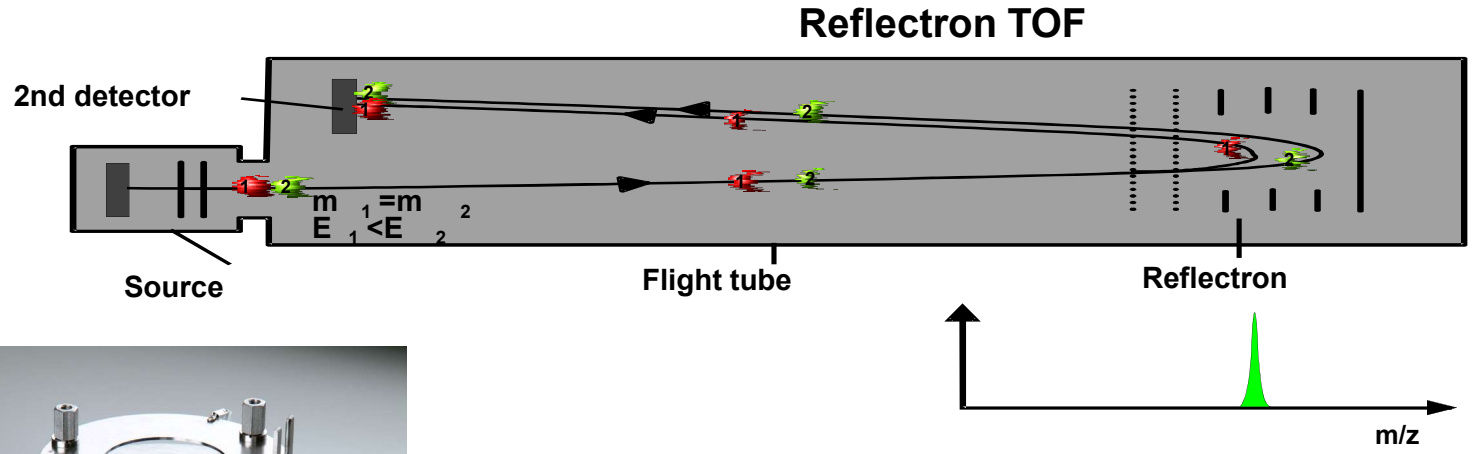
Light Corrections :

$t^2 = Am^2 + Bm + C$ (A : initial desorption Ec

C : Extraction Delay)

=> Simple Quadratic equation

Ions analysis in reflectron mode



Reflectron : 2 effects on resolution :

- Increasing flight path (better separation of particles of different masses) (equivalent 3m flight tube)
- Focusing effet for particles with same mass

Key questions in proteomics

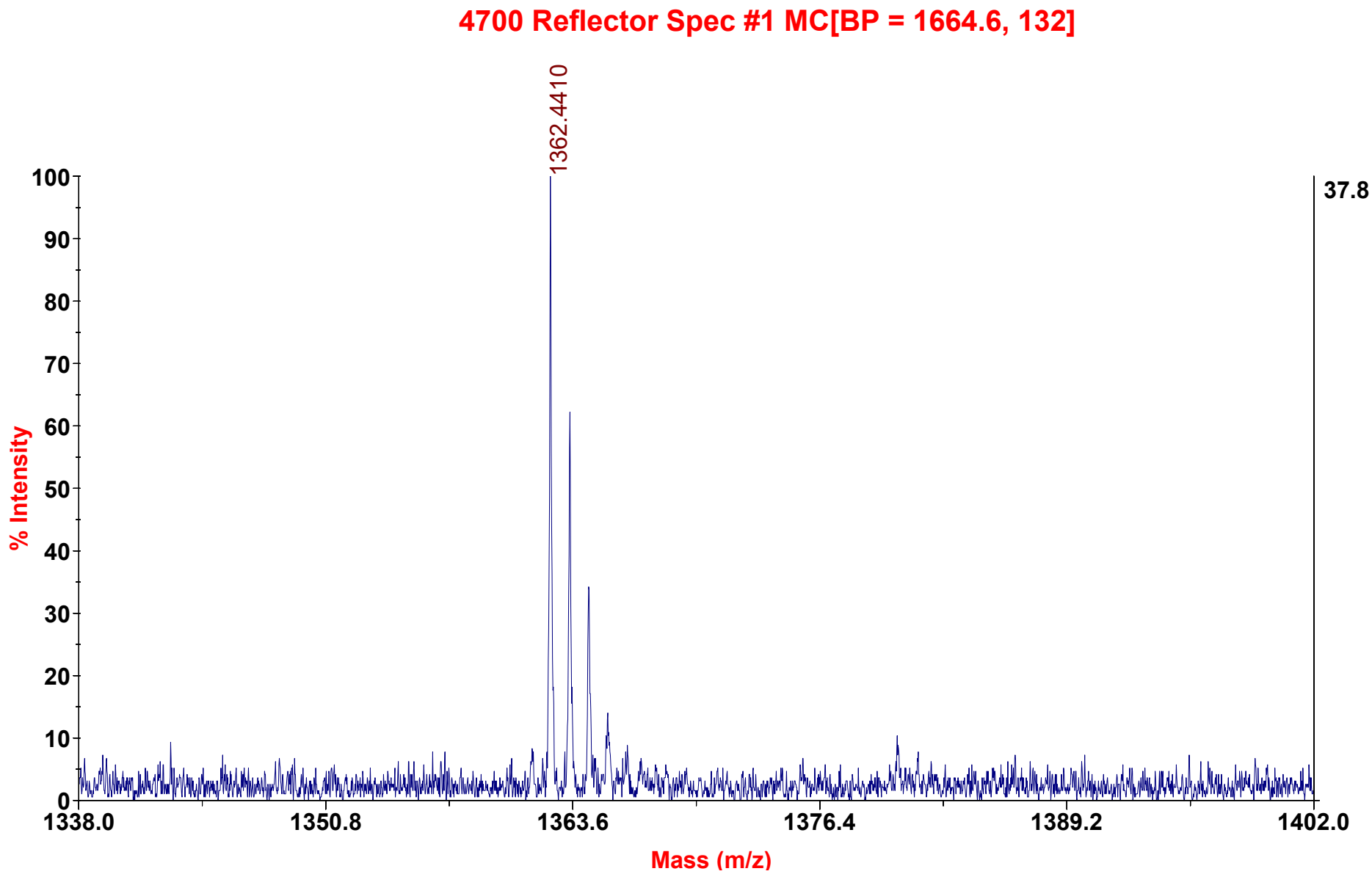
- What is the protein content of my biological sample?

  => problem of identification
- What is the abundance of my protein of interest?

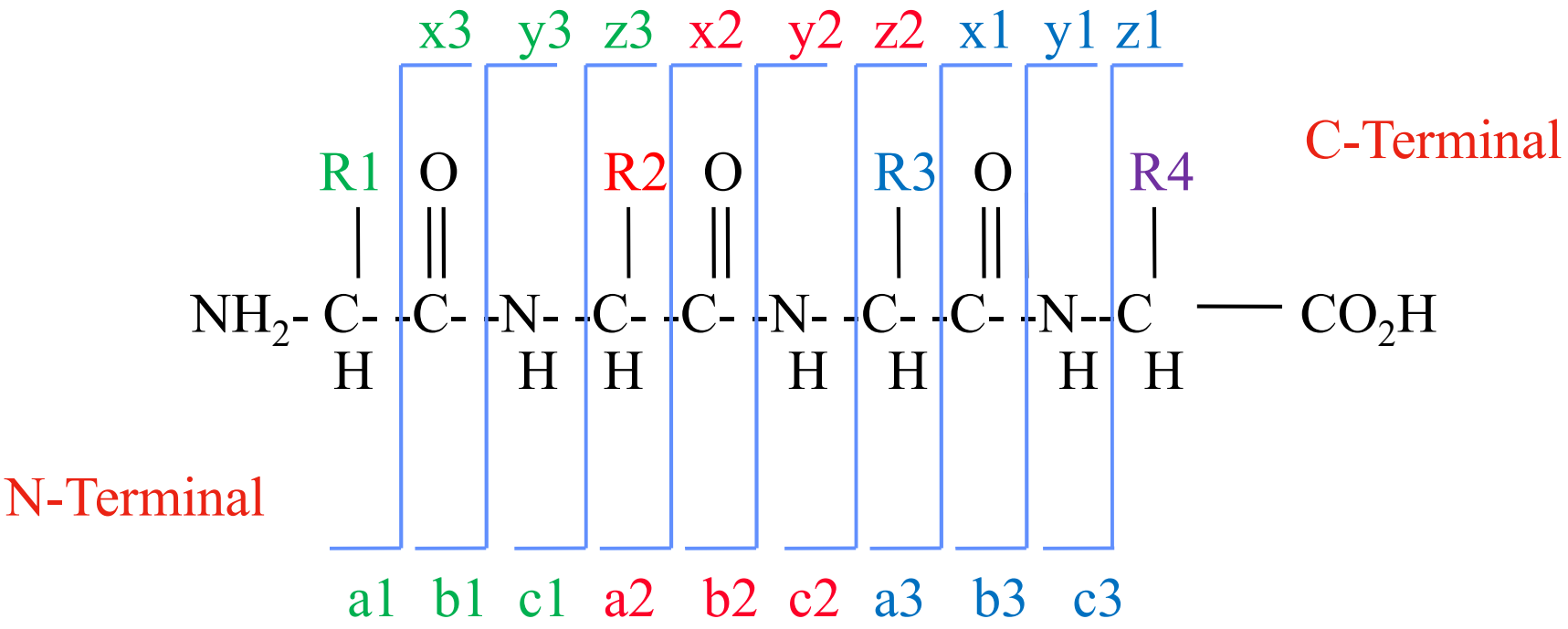
  => quantification problem
- Relative question: What are the protein abundance variations of the proteomes studied?
- What are the partners of my protein of interest?
- Are there any signature proteins related to a particular biological process?

  => biomarkers identifications and quantifications

Ion precursor selection



MS/MS fragmentation for peptides



a $H — (NH\text{-}CHR\text{-}CO)\text{-}\overset{+}{N}H{=}CHR$

b $H — (NH\text{-}CHR\text{-}CO)\text{-}NH{=}CHR\text{-}C{=}\overset{+}{O}$

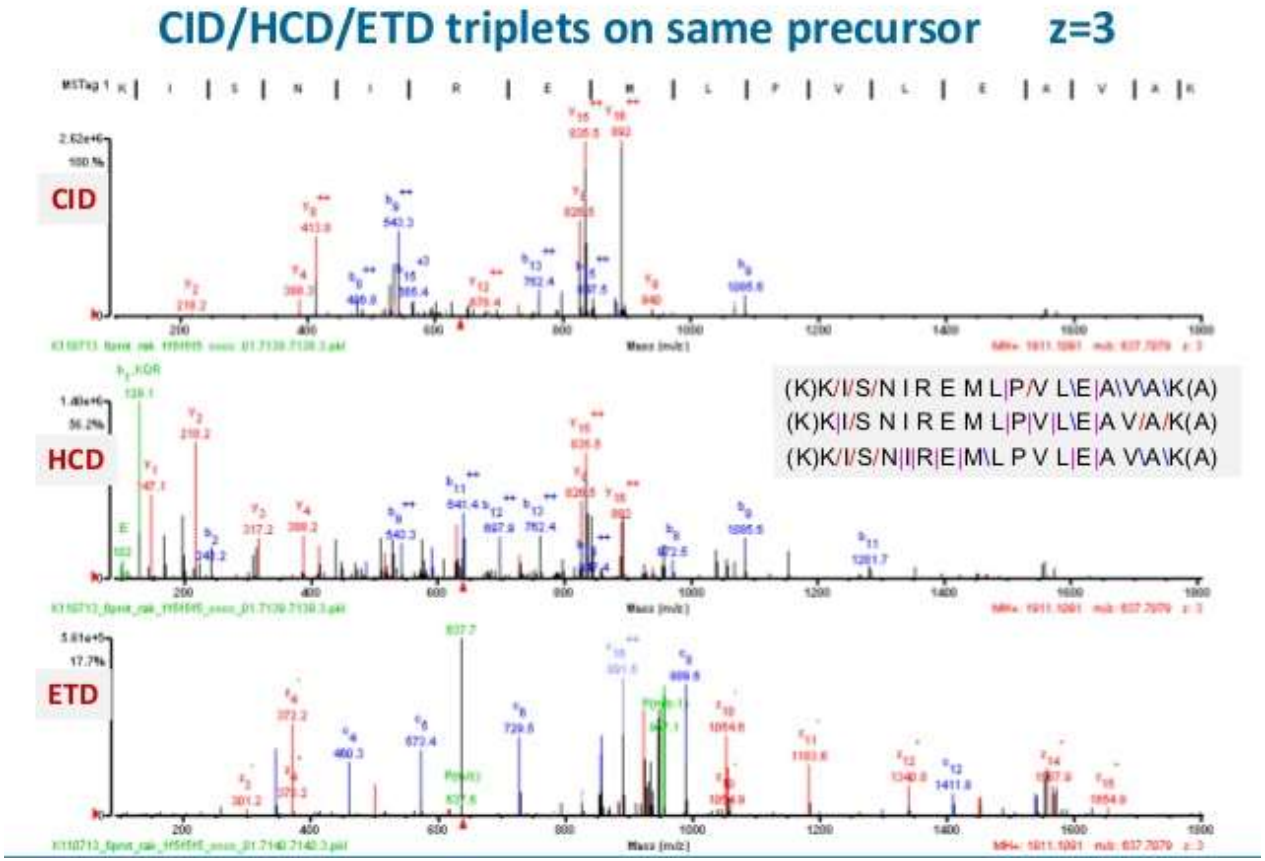
y $H — (\underbrace{NH\text{-}CHR\text{-}CO}_{+H})\text{-}OH$

MS/MS fragmentation for peptides

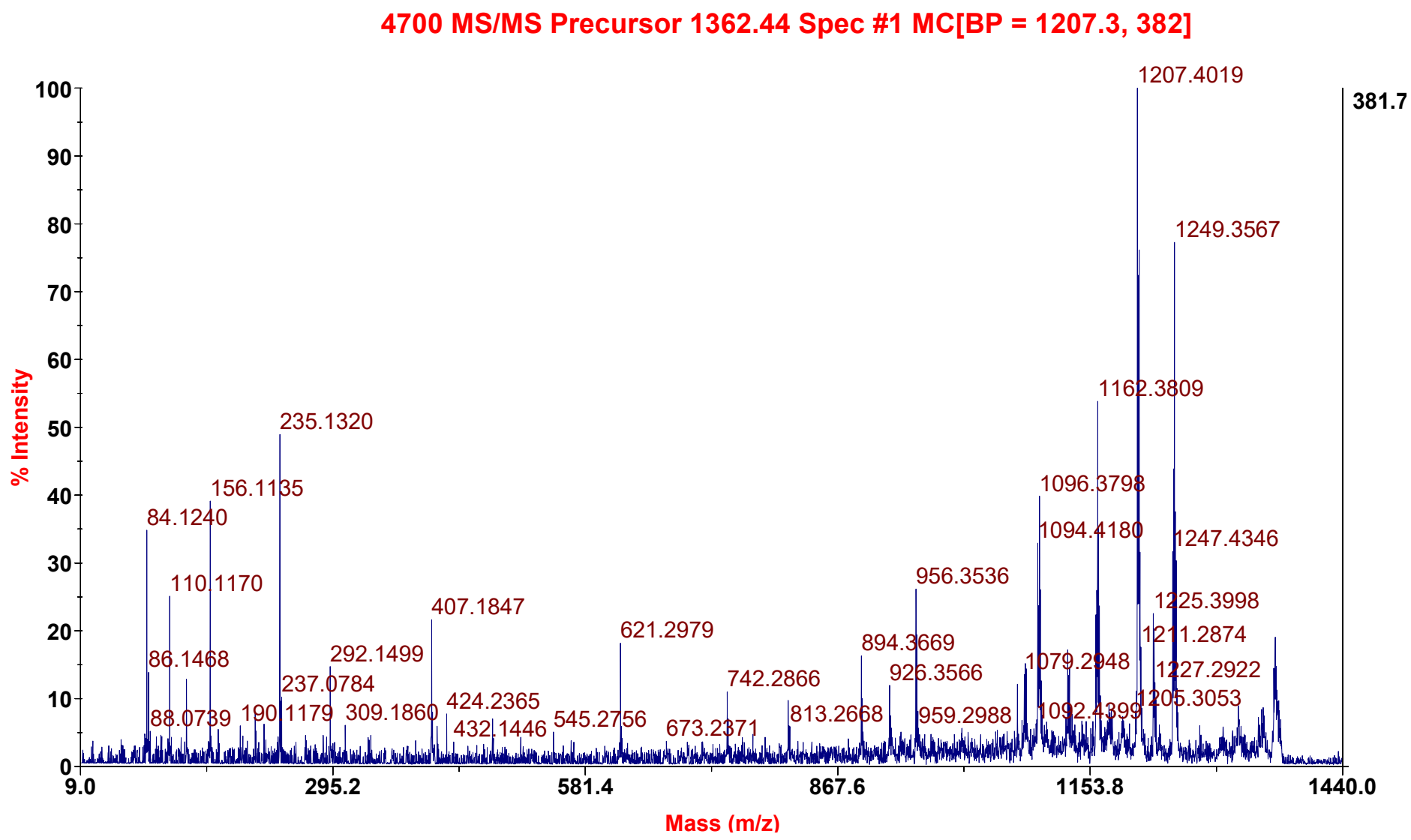
DISSOCIATION INDUITE PAR COLLISION (CID)

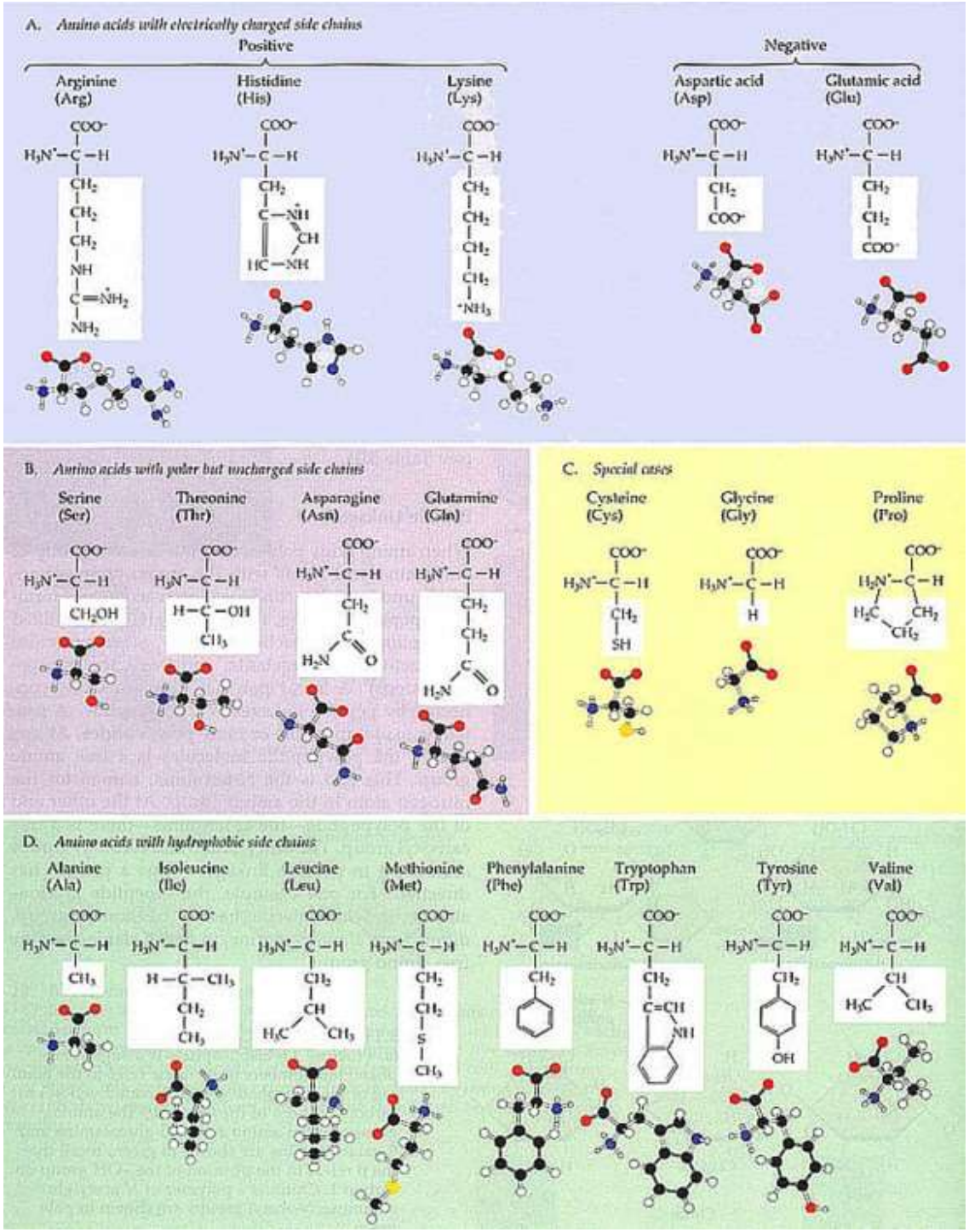
HIGHER ENERGY COLLISIONAL DISSOCIATION (HCD)

ELECTRON TRANSFER DISSOCIATION (ETD)



MS/MS spectrum of the precursor 1362.44 m/z





Alanine	A, Ala	71.079
Arginine	R, Arg	156.188
Asparagine	N, Asn	114.104
Aspartic acid	D, Asp	115.089
Cysteine	C, Cys	103.145
Glutamine	Q, Gln	128.131
Glutamic acid	E, Glu	129.116
Glycine	G, Gly	57.052
Histidine	H, His	137.141
Isoleucine	I, Ile	113.160
Leucine	L, Leu	113.160
Lysine	K, Lys	128.17
Methionine	M, Met	131.199
Phenylalanine	F, Phe	147.177
Proline	P, Pro	97.117
Serine	S, Ser	87.078
Threonine	T, Thr	101.105
Tryptophan	W, Trp	186.213
Tyrosine	Y, Tyr	163.176
Valine	V, Val	99.133

MS/MS spectra interpretation

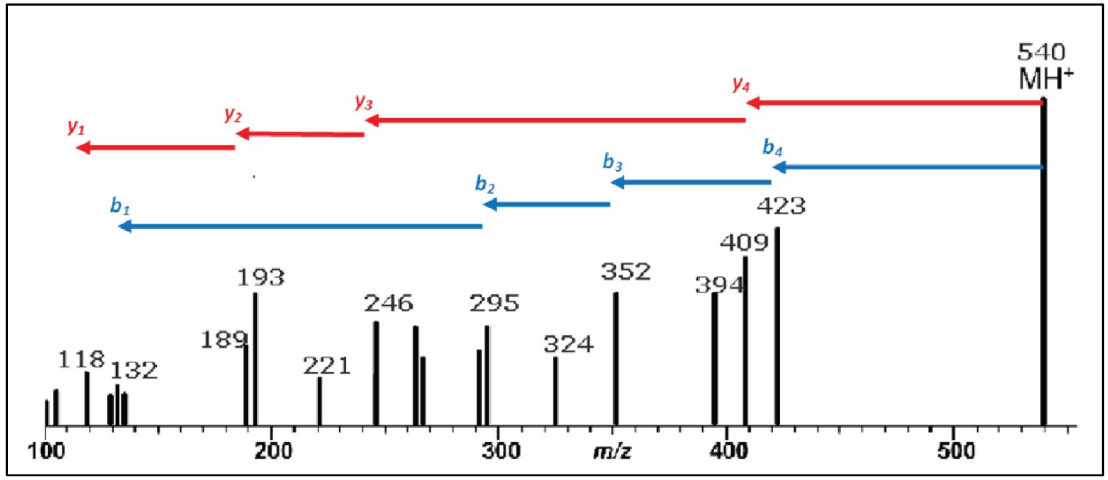


Table 1

Ion	*m/z*	Neutral loss (from previous ion in the series)	Amino Acid Residue
Precursor $[M+H]^+$	540		
y_4	409	131	M
y_3	246	163	Y
y_2	189	57	G
y_1	118	71	A
b_4	423	117 (99+18)	V
b_3	352	71	A
b_2	295	57	G
b_1	132	163	Y
a_4	395?		
a_3	324		
a_2	267		
a_1	104		

MYGAV
User AA Formula 1: C2 H3 N1 O1
Elemental Composition: C24 H38 N5 O7 S1
MH+1(av) MH+1(mono)
540.6627 540.2486
[–] Main Sequence Ions

b				y
---	1	M	5	---
295.1111	2	Y	4	409.2082
352.1326	3	G	3	246.1448
423.1697	4	A	2	189.1234
---	5	V	1	118.0863

http://www.sepscience.com/Information/Archive/MS-Solutions/246-

Current post-translational modifications (PTMs)

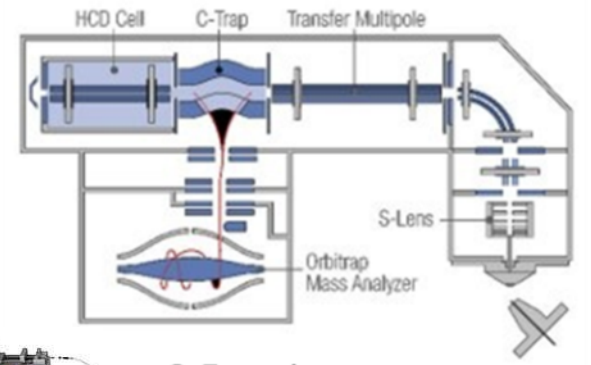
Acids & amides (E/D/Q/N)	Pyroglutamic acid (Q)	-17.0306	Deamidation (Q/N)	+0.9847
	Carboxylation (E/D)	+44.0098		

Hydroxyl groups (S/T/Y)	Phosphorylation	+79.9799	Sulphation	+80.0642

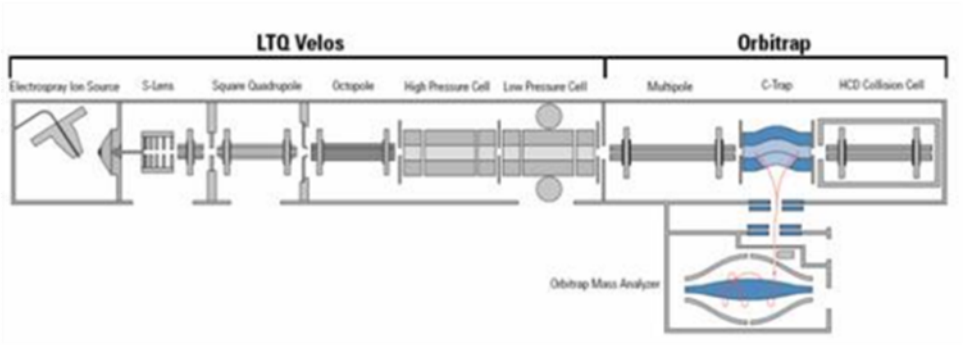
Carbohydrates (S/T/N)	Pentoses	+132.1161	Deoxyhexoses	+146.1430
	Hexosamines	+161.1577	Hexoses	+162.1424
	N-acetylhexosamines	+203.1950	Sialic acid	+291.2579

Sulphydryls (C)	Disulphide bond	-2.0159	Oxidation	+15.9994
	Cysteinylation	+119.1442	Glutathionylation	+305.3117

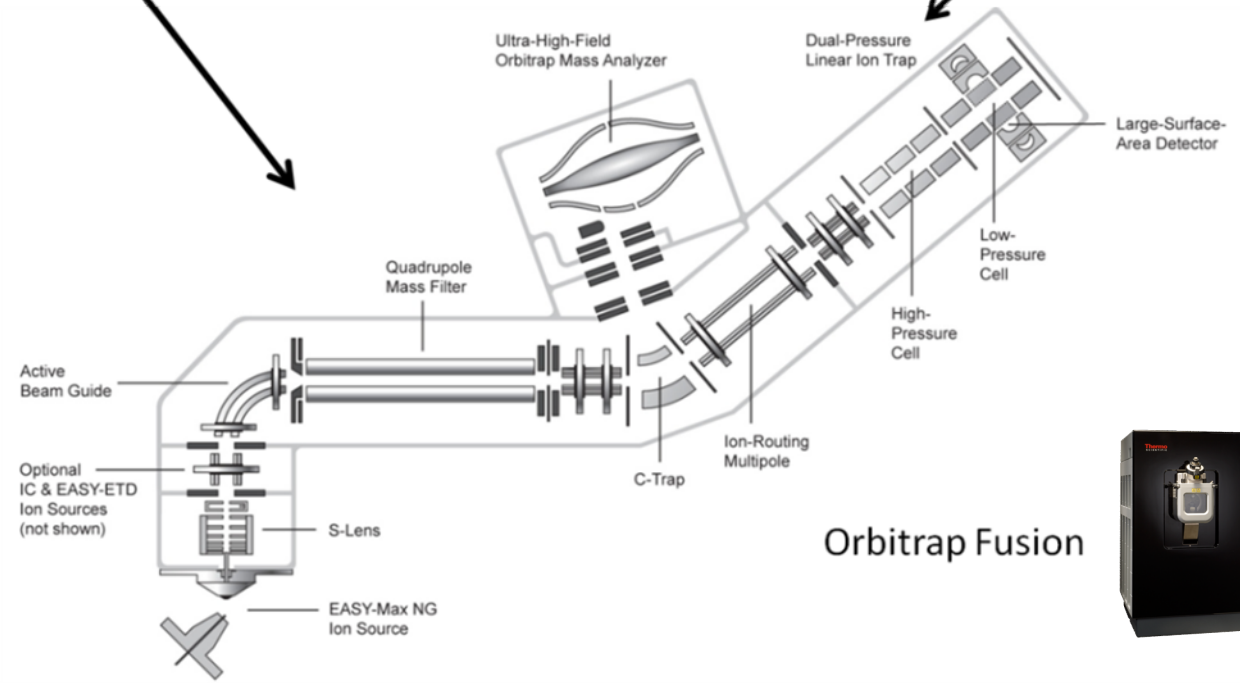
Orbitrap mass spectrometers



Q Exactive



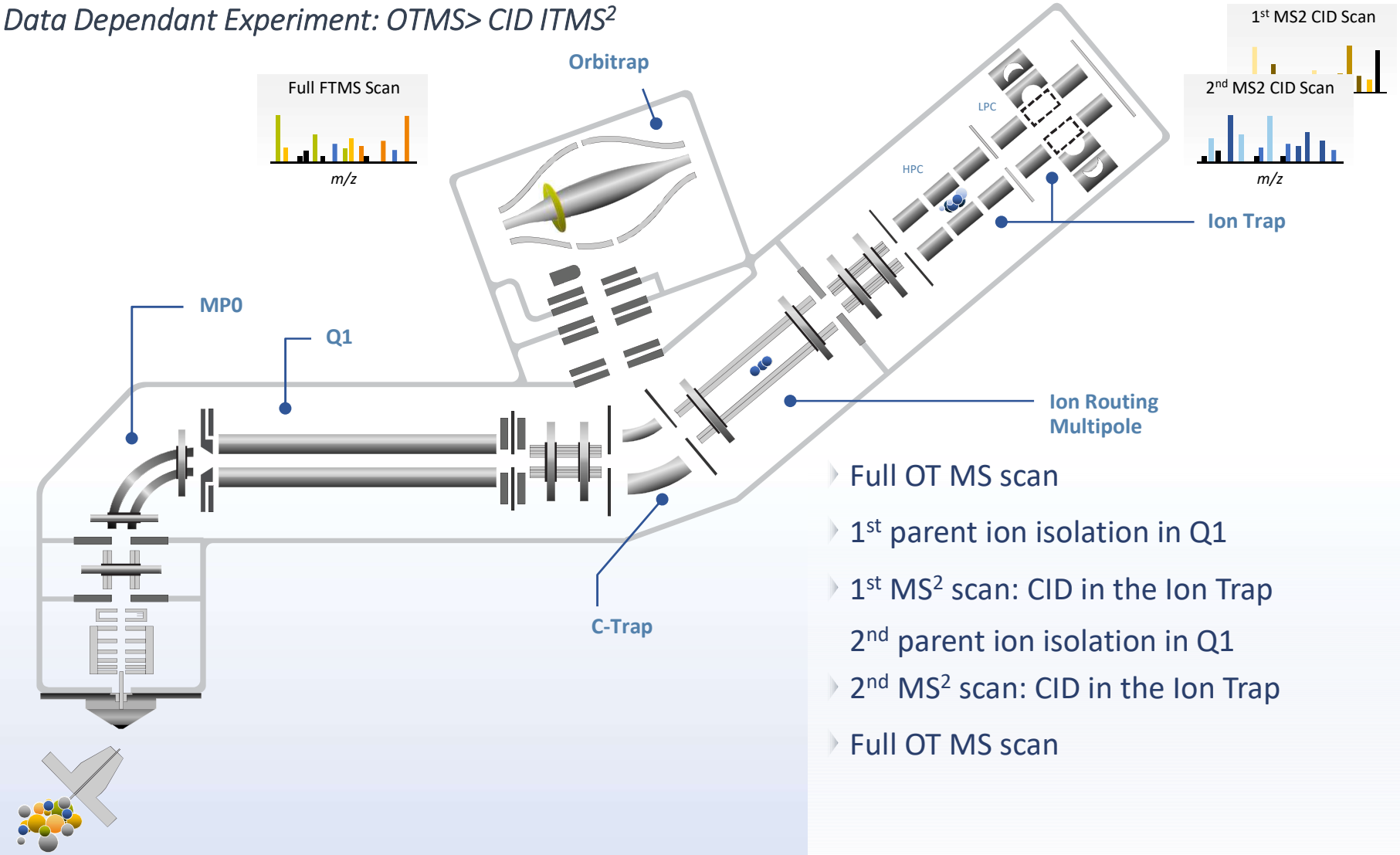
Orbitrap Velos



Orbitrap Fusion

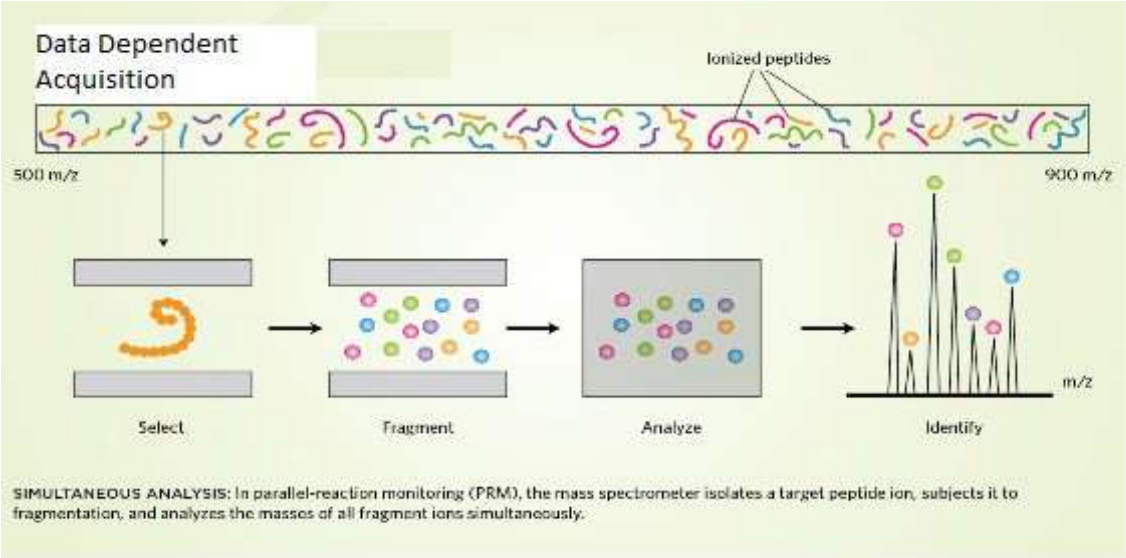
MS and MS/MS spectra generation

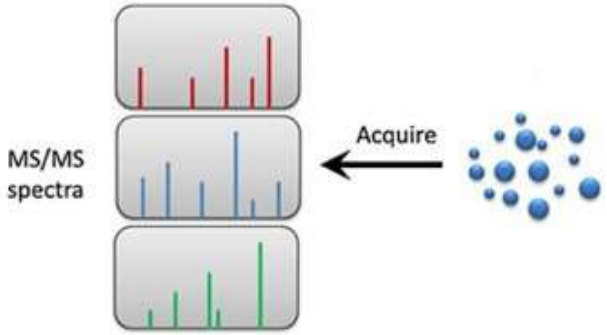
Data Dependant Experiment: OTMS> CID ITMS²

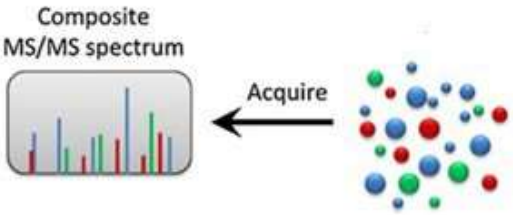


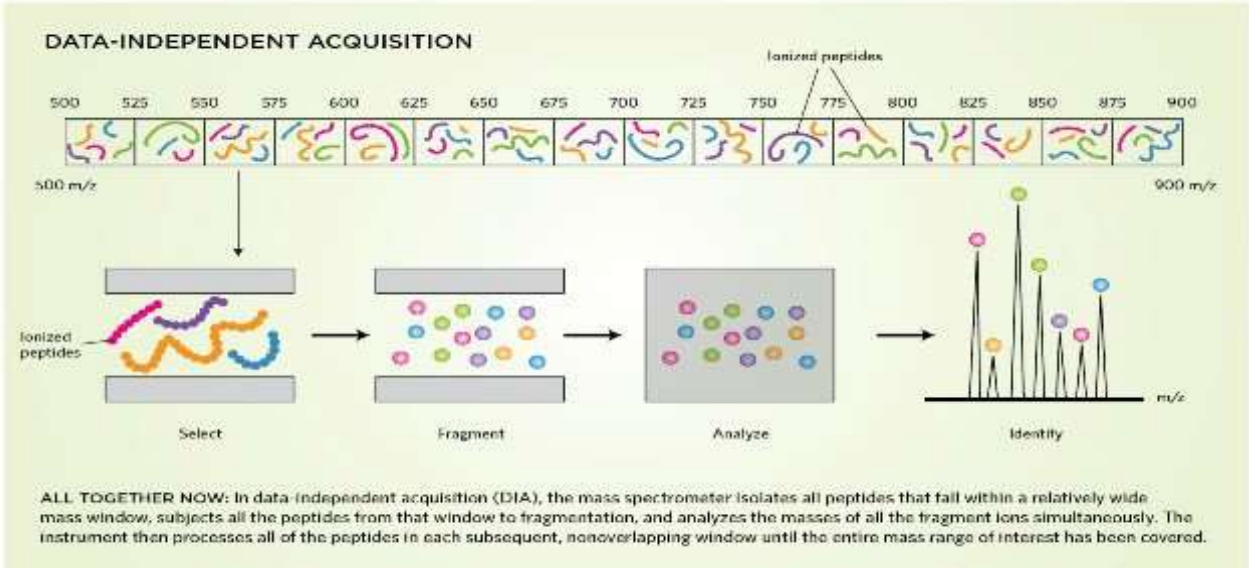
- Full OT MS scan
- 1st parent ion isolation in Q1
- 1st MS^2 scan: CID in the Ion Trap
- 2nd parent ion isolation in Q1
- 2nd MS^2 scan: CID in the Ion Trap
- Full OT MS scan

DDA versus DIA

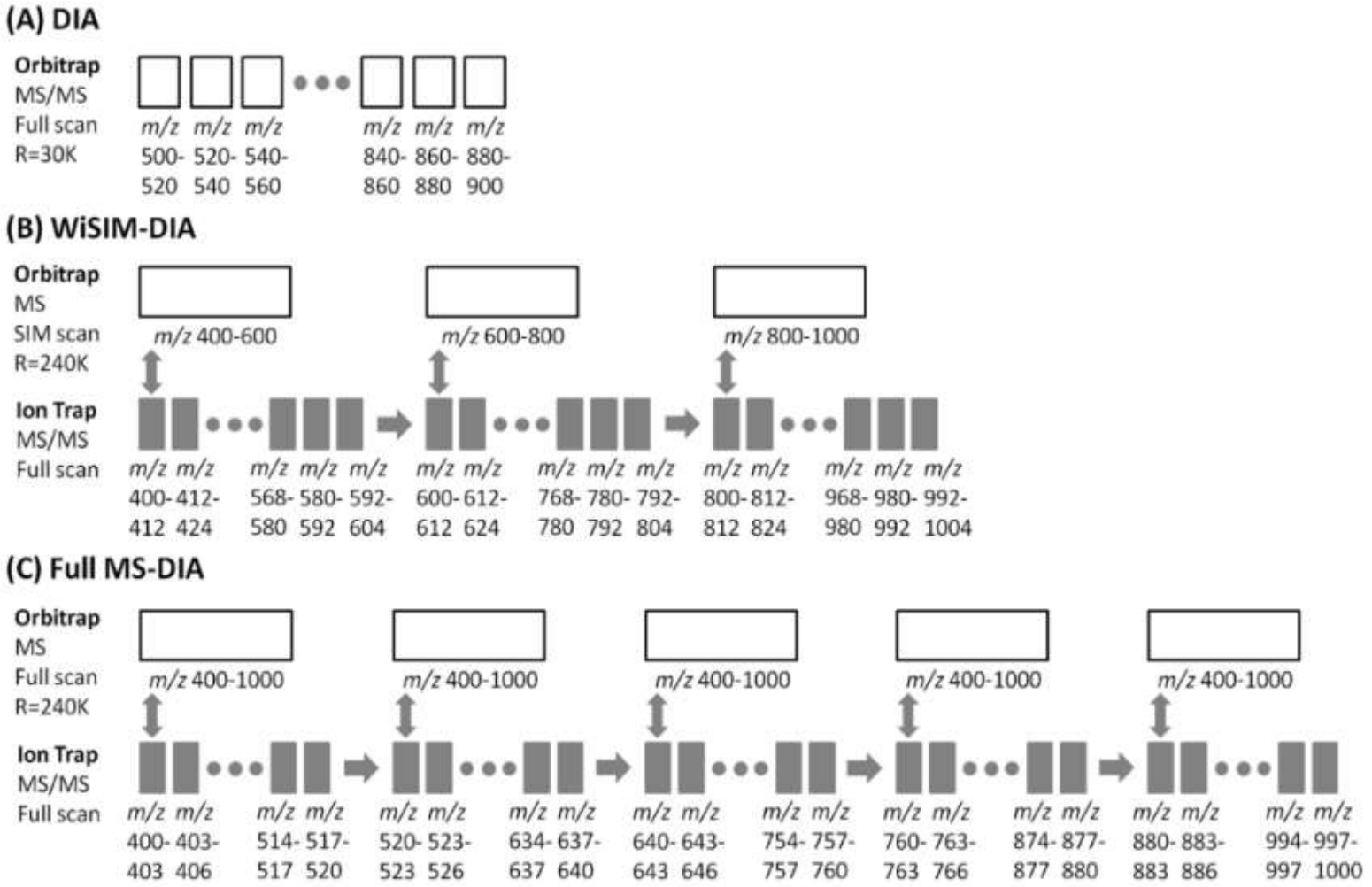




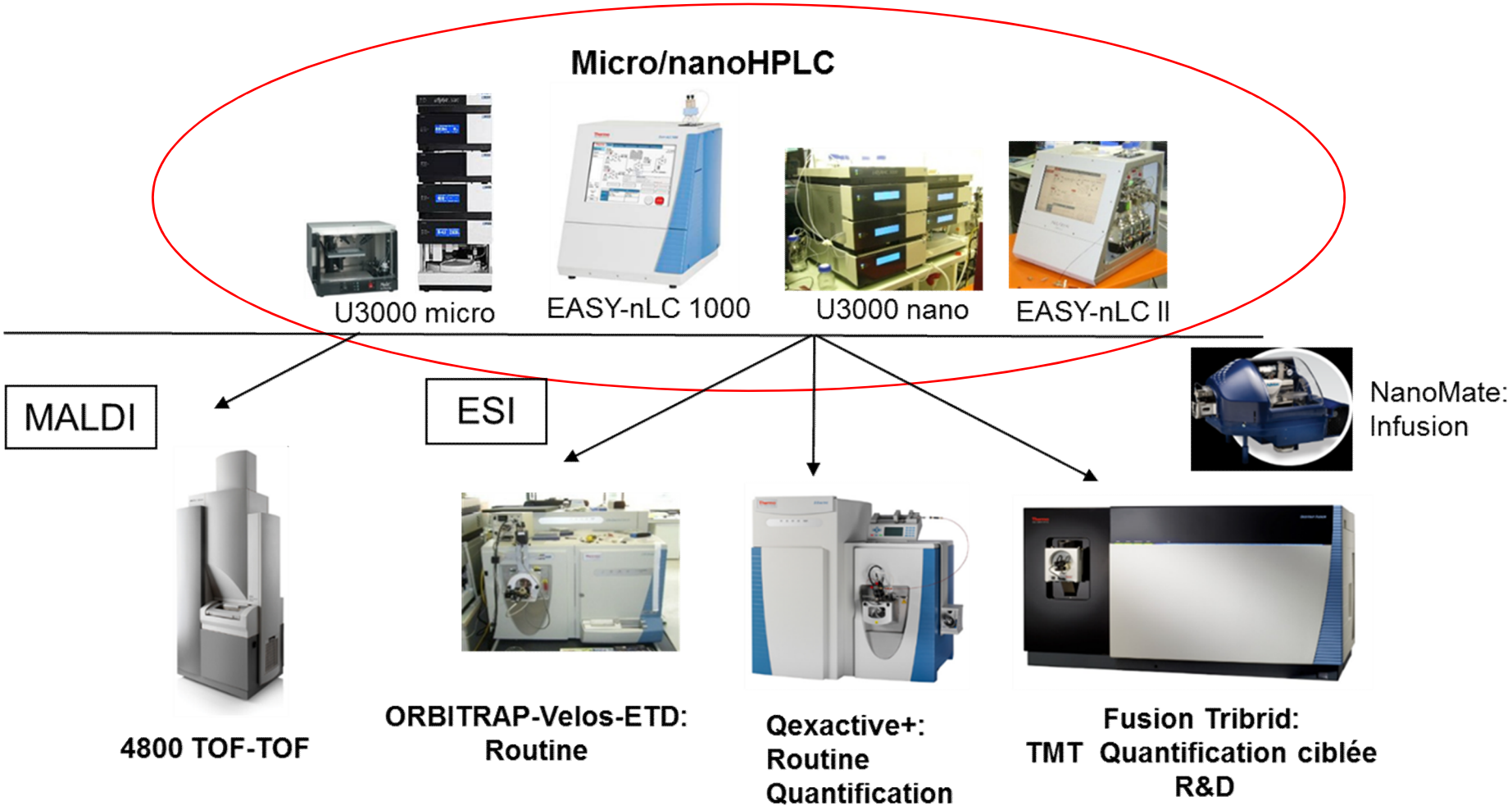




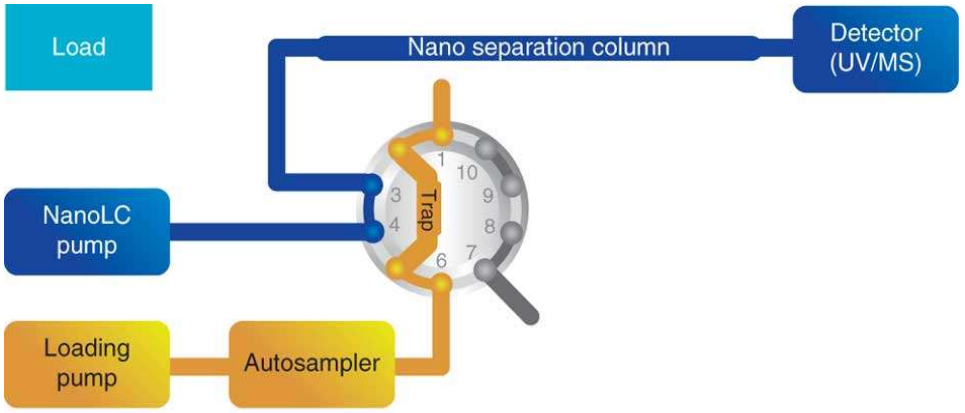
Data Independent Acquisition: DIA

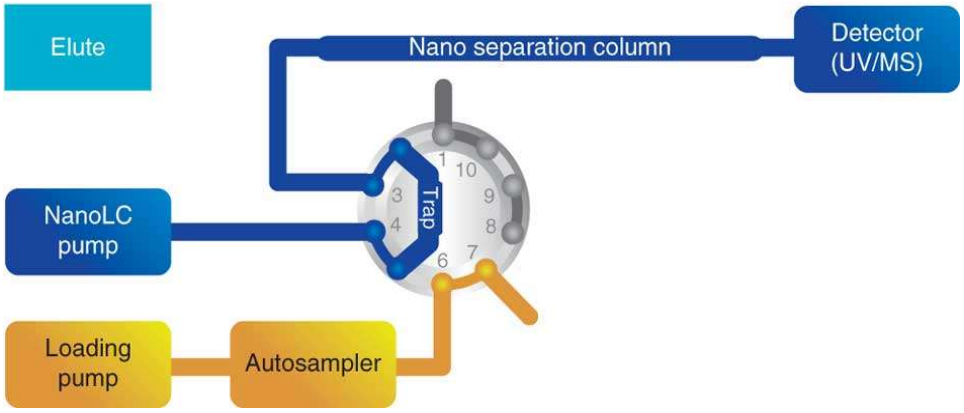


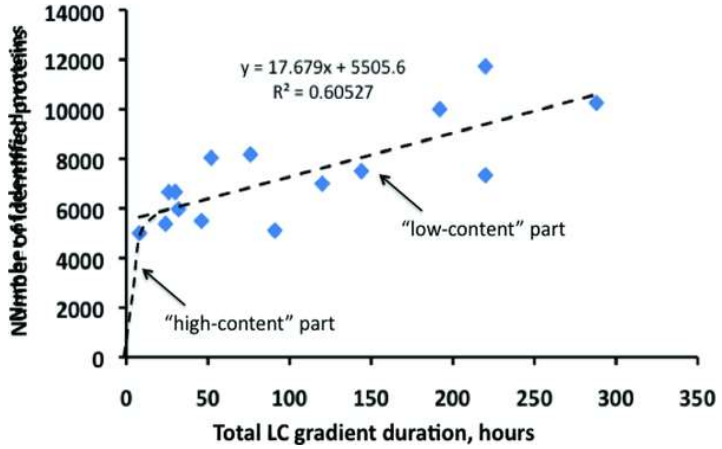
Contribution of nano-HPLC



Peptides separation by nano-LC





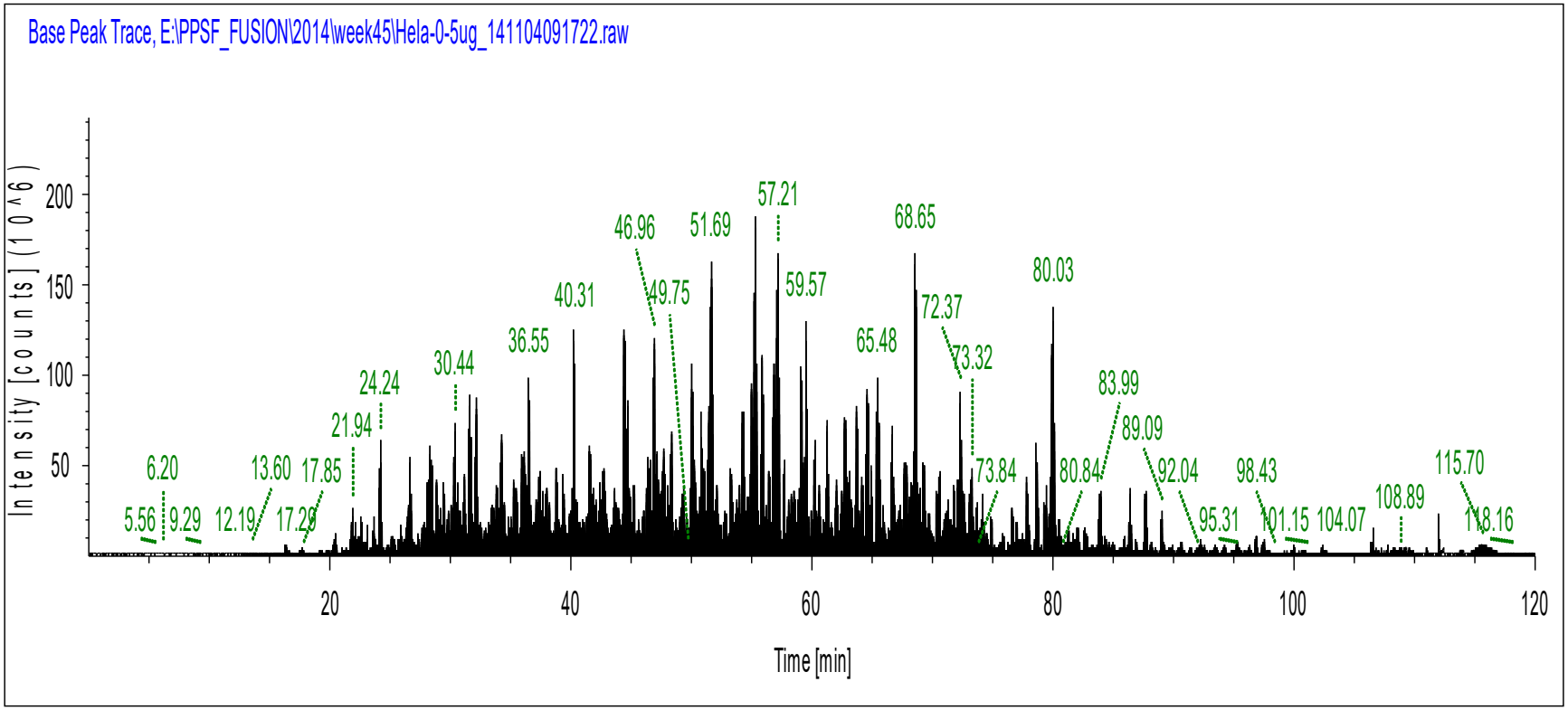


Peptides separation nano-LC

- It is impossible to resolve all species in a proteomics sample using only one separation method
- Multidimensional separation - two or more independent ("orthogonal") separation techniques coupled together for the analysis of a single sample.

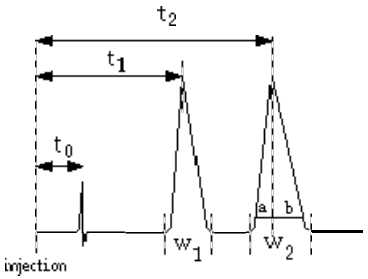
Separation method	Separation by:
Reversed phase	Hydrophobicity
Ion exchange, IsoElectroFocusing (IEF)	Net charge, Isoelectric point
Size exclusion, SDS Gel Electrophoresis	Size, molecular weight
Affinity chromatography	Specific functional groups

Total ion current (TIC) Hela tryptic digest (0.5 µg of total proteins)

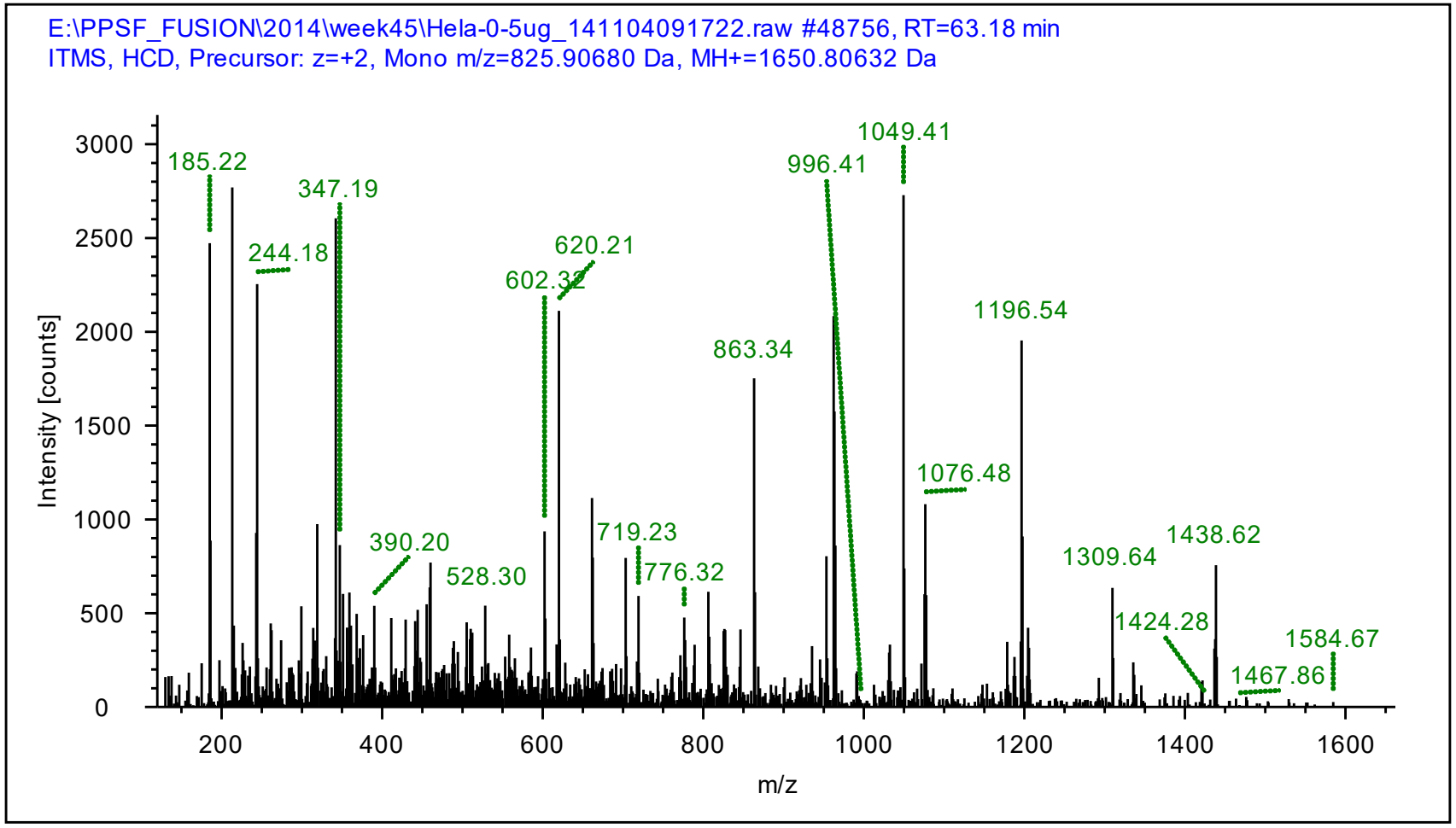


$$R = \frac{2(t2 - t1)}{(w2 - w1)}$$

with t1 and t2 the retention time and w1 and w2 peak widths at mid-height



90000 MSMS in 2h gradient (C18 RPC)



Proprietary MS data formats

Company	Extension	File type
Agilent Bruker	.D (folder)	Agilent MassHunter, Agilent ChemStation, or Bruker BAF/YEP/TDF data format
Agilent/Bruker	.YEP	instrument data format
Bruker	.BAF	instrument data format
Bruker	.FID	instrument data format
Bruker	.TDF	timsTOF instrument data format
ABI/Sciex	.WIFF	instrument data format
ABI/Sciex	.t2d	4700 and 4800 file format
Waters	.PKL	MassLynx peak list format
Thermo PerkinElmer	.RAW*	Thermo Xcalibur PerkinElmer TurboMass
Micromass**/Waters	.RAW* (folder)	Waters MassLynx
Chromtech Finnigan***	.DAT	Finnigan ITDS file format; MAT95 instrument data format
VG		MassLab data format
Finnigan***	.MS	ITS40 instrument data format
Shimadzu	.QGD	GCMSSolution format
Shimadzu	.qgd	instrument data format
Shimadzu	.lcd	QQQ/QTOF instrument data format
Shimadzu	.spc	library data format
Bruker/Varian	.SMS	instrument data format
Bruker/Varian	.XMS	instrument data format
ION-TOF	.itm	raw measurement data
ION-TOF	.ita	analysis data
Physical Electronics/ULVAC-PHI	.raw*	raw measurement data
Physical Electronics/ULVAC-PHI	.tdc	spectrum data

Open MS data formats

JCAMP-DX

This format was one of the earliest attempts to supply a standardized file format for data exchange in mass spectrometry. JCAMP-DX was initially developed for infrared spectrometry. JCAMP was officially released in 1988. JCAMP was found impractical for today's large MS data sets, but it is still used for exchanging moderate numbers of spectra.

ANDI-MS or netCDF

The Analytical Data Interchange Format for Mass Spectrometry is a format for exchanging data. ANDI was initially developed for chromatography-MS data and therefore was not used in the proteomics gold rush where new formats based on XML were developed.

mzData

mzData was the first attempt by the Proteomics Standards Initiative (PSI) from the Human Proteome Organization (HUPO) to create a standardized format for Mass Spectrometry data. This format is now deprecated, and replaced by mzML.

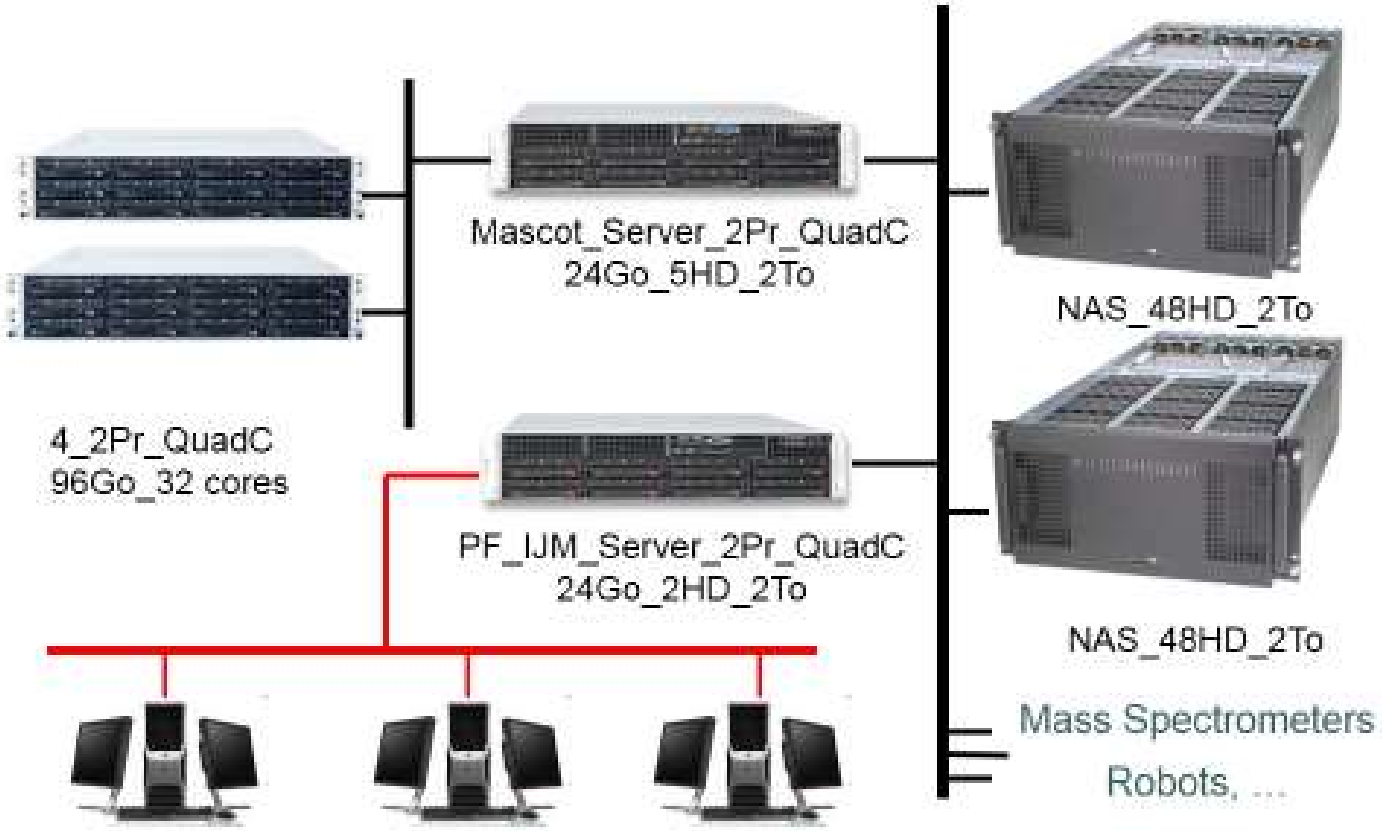
mzXML

mzXML is a XML (eXtensible Markup Language) based common file format for proteomics mass spectrometric data. This format was developed at the Seattle Proteome Center/Institute for Systems Biology while the HUPO-PSI was trying to specify the standardized mzData format, and is still in use in the proteomics community.

mzML

As two formats (mzData and mzXML) for representing the same information is an undesirable state, a joint effort was set by HUPO-PSI, the SPC/ISB and instrument vendors to create a unified standard borrowing the best aspects of both mzData and mzXML, and intended to replace them. The first specification was published in June 2008. This format was officially released at the 2008 American Society for Mass Spectrometry Meeting, and is since then relatively stable with very few updates. On 1 June 2009, mzML 1.1.0 was released. There are no planned further changes as of 2013.

Saving data and servers



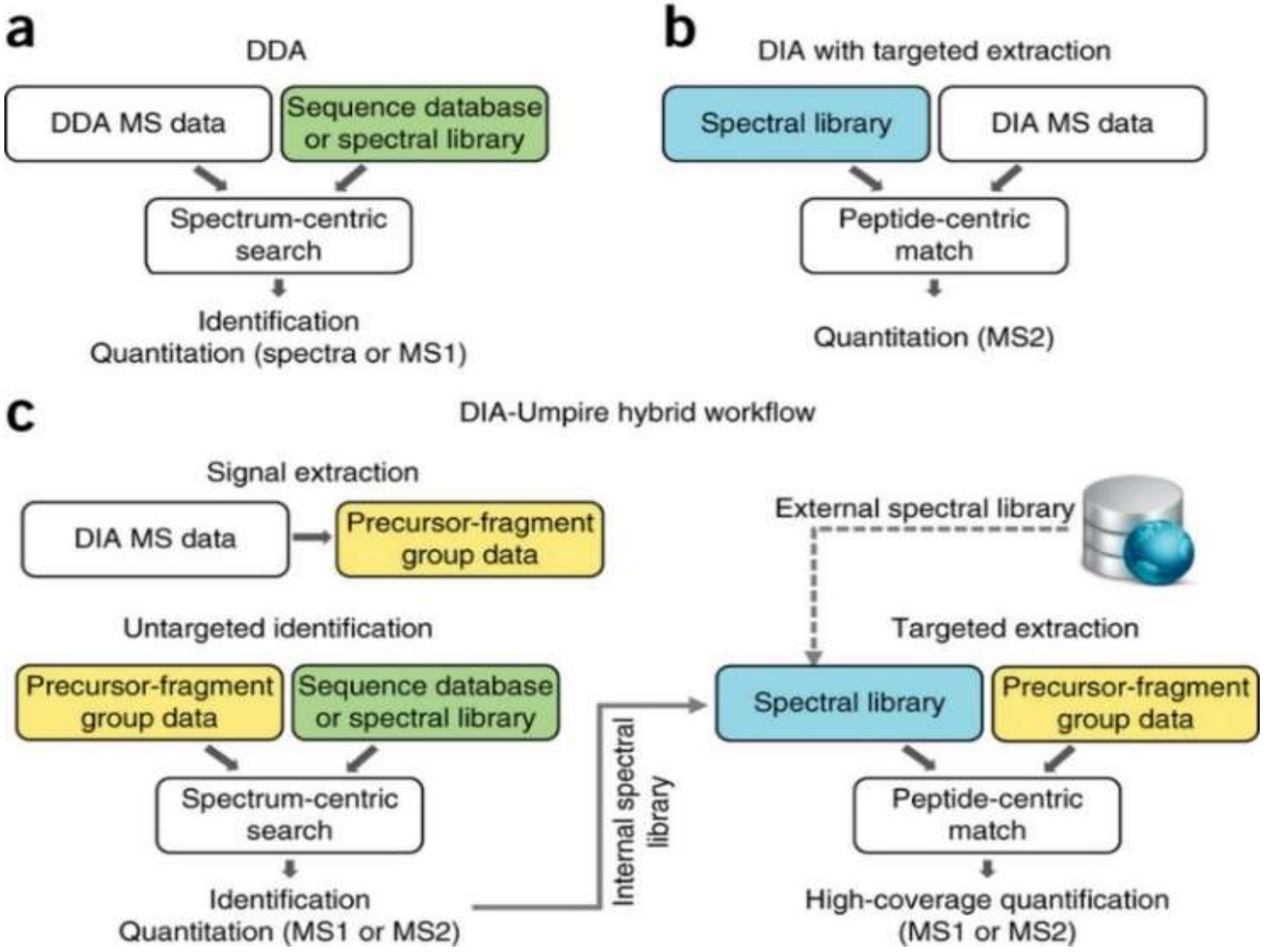
Progenesis QI for proteomics

PROTEOME DISCOVERER 1.4
Mass Informatics Platform for Protein Scientists
Thermo SCIENTIFIC

Byonic™
PROTEIN METRICS INC.

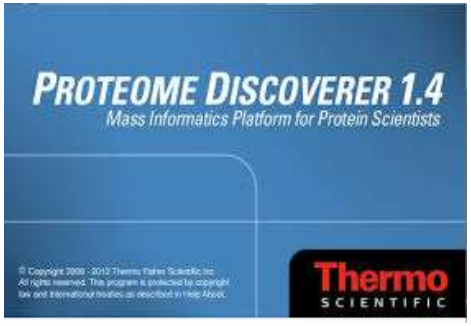
{MATRIX SCIENCE}

Search engine



Search engines and validation of peptides and proteins identifications

$$FDR\ (\%) = \frac{(number\ of\ false\ positive\ peptides)\ X\ 2}{total\ number\ of\ peptides\ (positives + false\ positive)}$$



Critical importance of mass accuracy for database searches

Expressed as Da or as ppm (10 ppm = 0,001%1 ppm = 0,0001%)

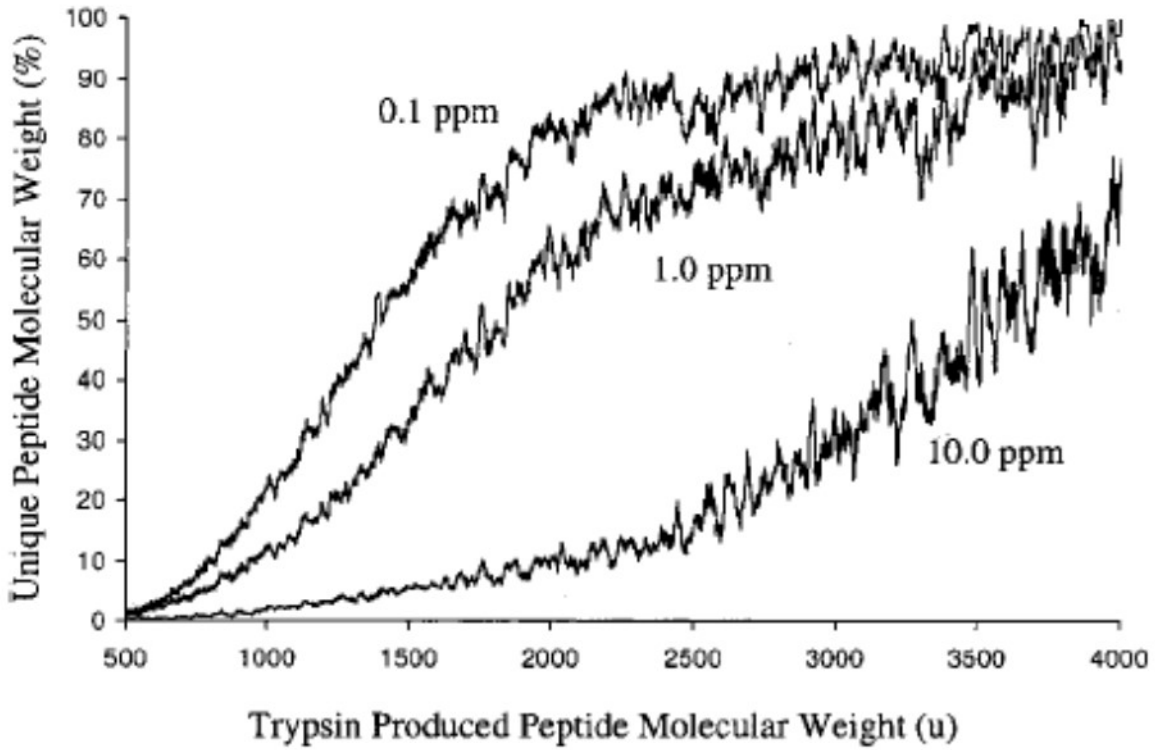
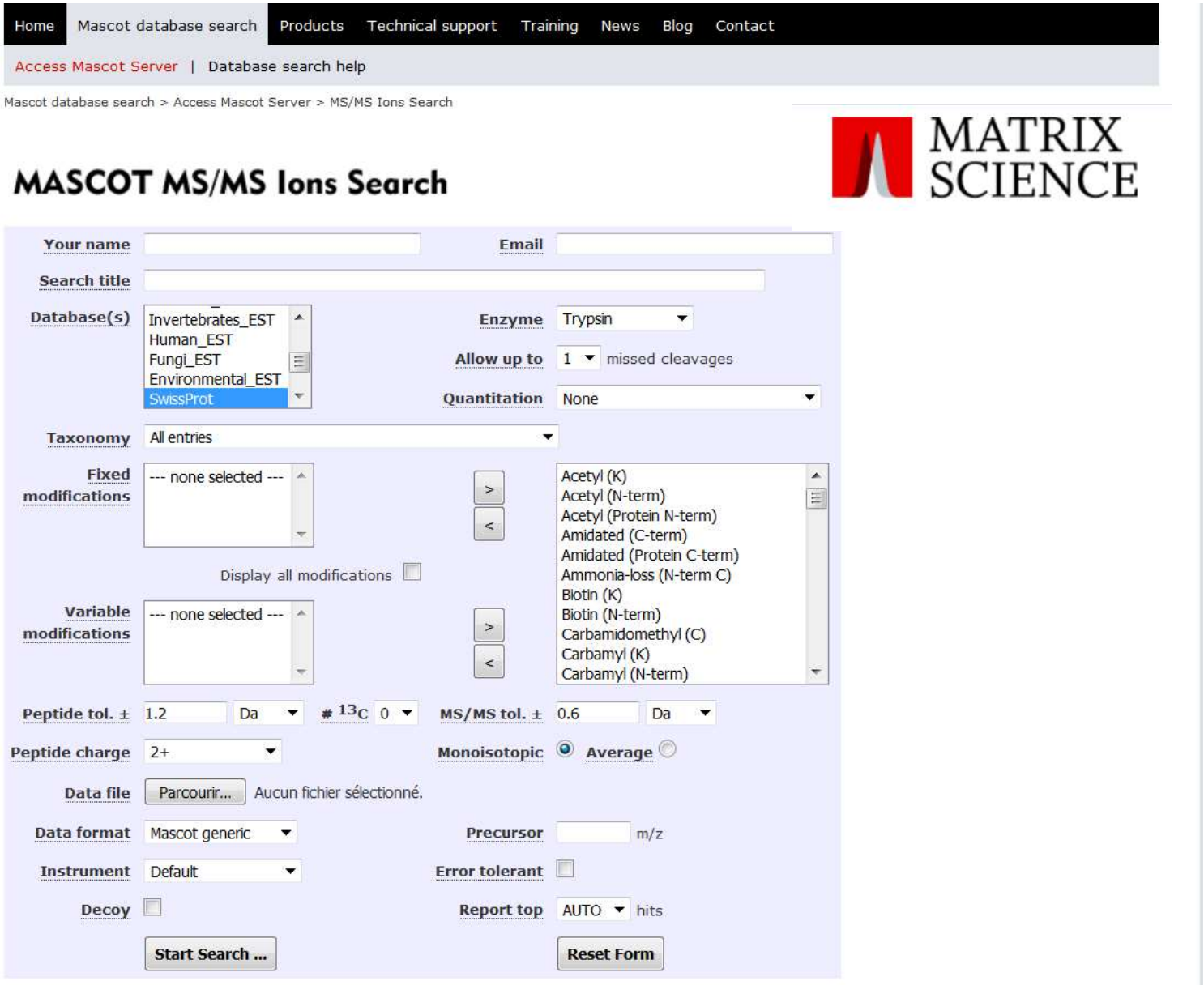
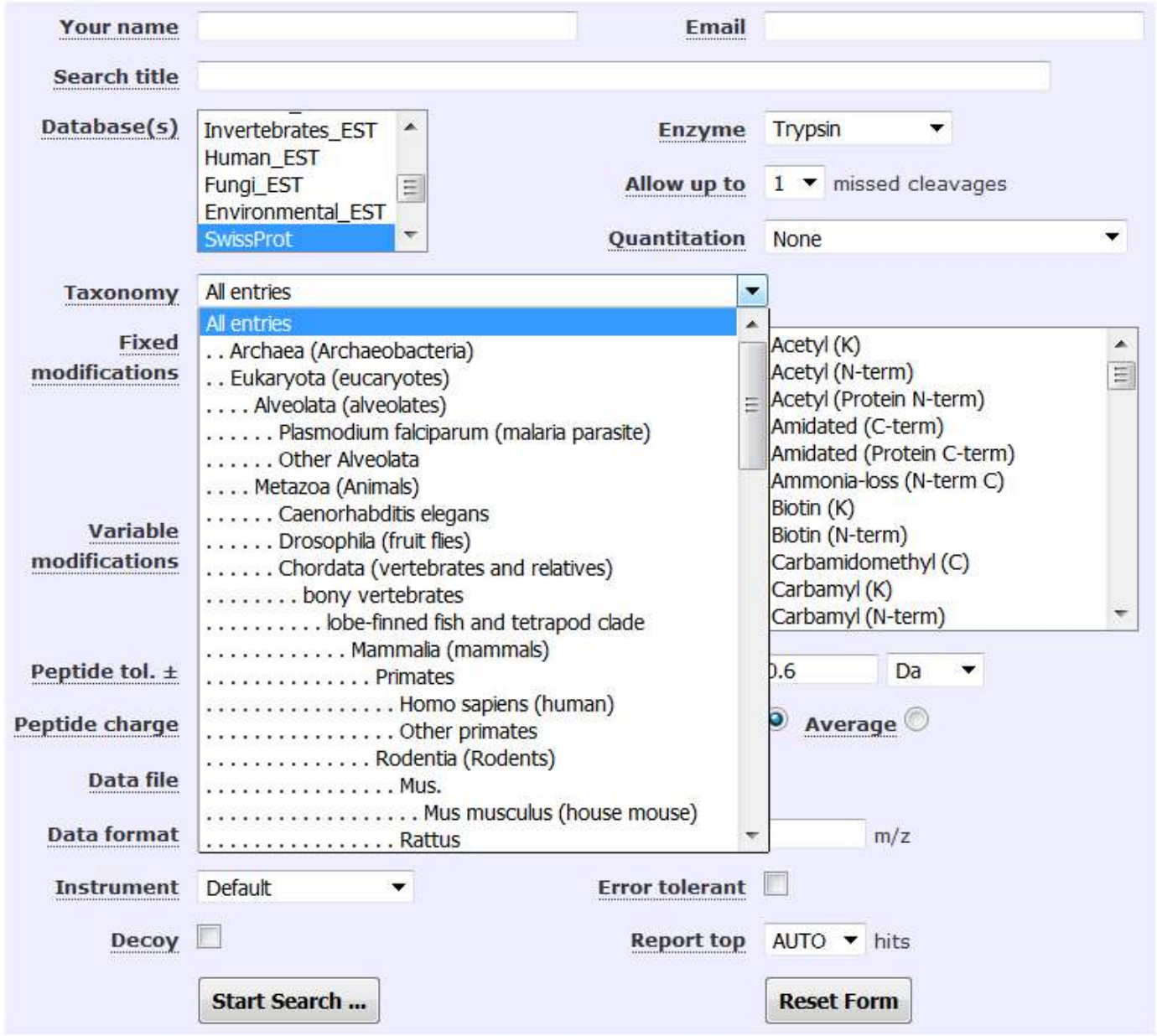


Figure 1. All possible unique peptide molecular weights after digestion of all yeast proteins in the National Center for Biotechnology Information at a mass accuracy of 0.1, 1.0, and 10.0 ppm.

A database search engine : Mascot



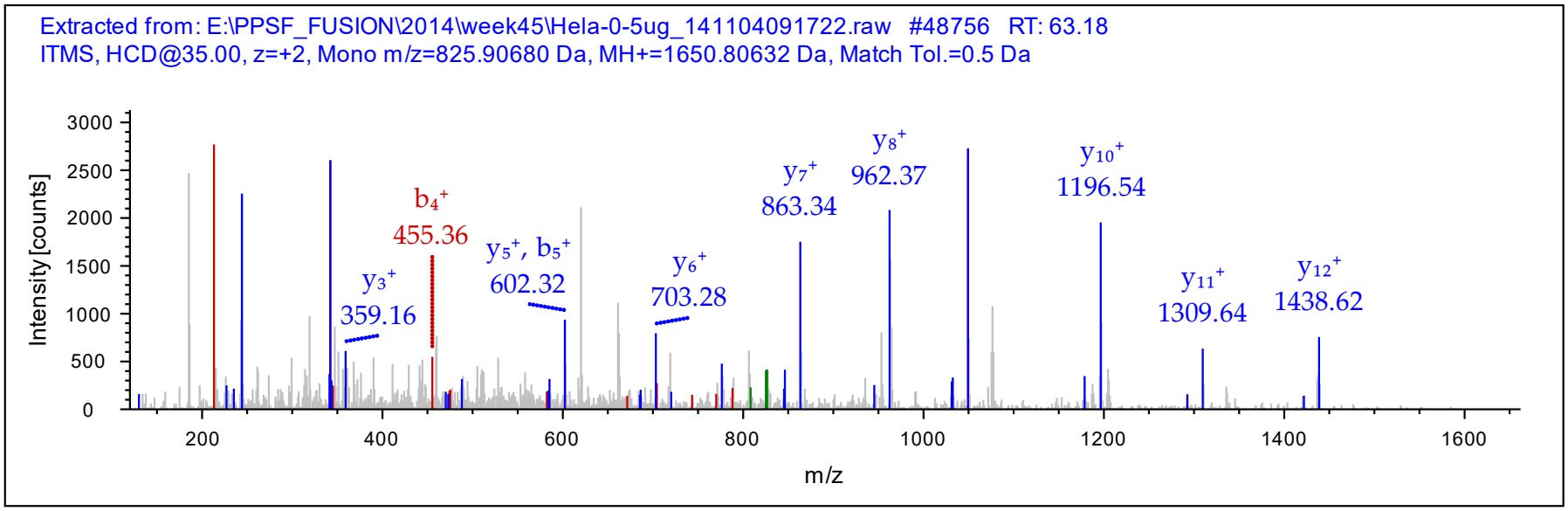
MASCOT MS/MS Ions Search



Search engine output formats

File name	File content
Processed peak lists	Heavily processed form of mass spectrometry data, usually derived from raw data files via various (semi-) automatic steps, e.g.: centroiding, deisotoping and charge deconvolution. These files are formatted in plain text, with typical formats like **dta**, **pkl**, **ms2** or **mgf**.
Search engine output files	These files contain the data and metadata generated by the software (called search engines) used for performing the identification and quantification of peptides and proteins. **Each search engine has its own specific output file format**. The outputs are typically formatted in either plain text or XML. mzIdentML - provides a common format for the export of identification results from any search engine. mzQuantML - provides a common format for the export of quantification results from any search engine. mzTab - represents both identification and basic quantification results. To allow a full representation of the processed results in the PRIDE database and in the PX tool, the search engine output files need to be converted to PRIDE XML. PRIDE Converter and PRIDE Converter 2 are the two tools developed by the PRIDE team to make this conversion possible.
Protein/peptide identifications	Proteomics mass spectra can be matched to peptides or proteins, resulting in identifications for those spectra. Typically a spectrum is considered to have been identified if the score attributed to a peptide or protein match qualifies against an *a priori* or *a posteriori* defined threshold. In the case of fragmentation spectra, the initial identification will consist of a peptide sequence; subsequent steps will derive a list of proteins from the identified peptides. The protein assembly step can be a discernible process with its own input and output files, or it can be implicit in the overall identification software.

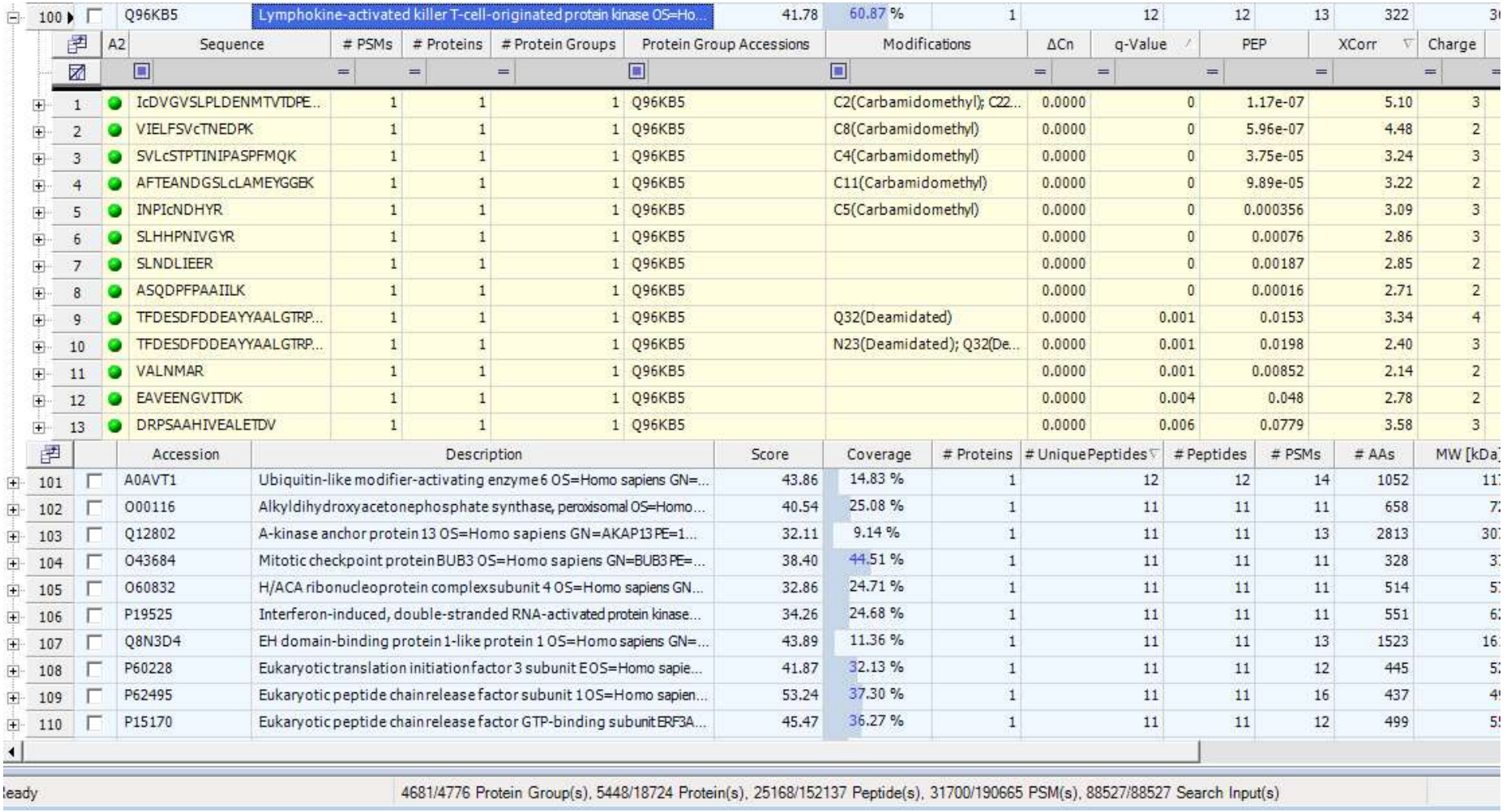
31700 MS/MS spectra interpreted!!!!



Sequence: VIELFSVCTNEDPK, C8-Carbamidomethyl (57.02146 Da)
Charge: +2, Monoisotopic m/z: 825.90680 Da (+0.95 mmu/+1.15 ppm), MH+: 1650.80632 Da, RT: 63.18 min,
Identified with: Sequest HT (v1.3); XCorr:4.48, Ions matched by search engine: 0/0
Fragment match tolerance used for search: 0.5 Da
Fragments used for search: b; b-H_2O; b-NH_3; y; y-H_2O; y-NH_3

Protein references (1):
- Lymphokine-activated killer T-cell-originated protein kinase OS=Homo sapiens GN=PBK PE=1 SV=3 - [TOPK_HUMAN]

5448 identified proteins



		Accession	Description	Score	Coverage	# Proteins	# Unique Peptides	# Peptides	# PSMs	# AAs	MW [kDa]
100		Q96KB5	Lymphokine-activated killer T-cell-originated protein kinase OS=Ho...	41.78	60.87 %	1	12	12	13	322	3

	A2	Sequence	# PSMs	# Proteins	# Protein Groups	Protein Group Accessions	Modifications	ΔCn	q-Value	PEP	XCorr	Charge
1		IcDVGVSLPLDENMTVTDPE...	1	1	1	Q96KB5	C2(Carbamidomethyl); C22...	0.0000	0	1.17e-07	5.10	3
2		VIELFSVcTNEDPK	1	1	1	Q96KB5	C8(Carbamidomethyl)	0.0000	0	5.96e-07	4.48	2
3		SVLcSTPTINIPASPFMQK	1	1	1	Q96KB5	C4(Carbamidomethyl)	0.0000	0	3.75e-05	3.24	3
4		AFTEANDGSLcLAMEYGGEK	1	1	1	Q96KB5	C11(Carbamidomethyl)	0.0000	0	9.89e-05	3.22	2
5		INPIcNDHYR	1	1	1	Q96KB5	C5(Carbamidomethyl)	0.0000	0	0.000356	3.09	3
6		SLHHPNIVGYR	1	1	1	Q96KB5		0.0000	0	0.00076	2.86	3
7		SLNDLIEER	1	1	1	Q96KB5		0.0000	0	0.00187	2.85	2
8		ASQDPFPAAIILK	1	1	1	Q96KB5		0.0000	0	0.00016	2.71	2
9		TFDESDFDDEAYYAALGTRP...	1	1	1	Q96KB5	Q32(Deamidated)	0.0000	0.001	0.0153	3.34	4
10		TFDESDFDDEAYYAALGTRP...	1	1	1	Q96KB5	N23(Deamidated); Q32(De...	0.0000	0.001	0.0198	2.40	3
11		VALNMAR	1	1	1	Q96KB5		0.0000	0.001	0.00852	2.14	2
12		EAVEENGVITDK	1	1	1	Q96KB5		0.0000	0.004	0.048	2.78	2
13		DRPSAAHIVEALETDV	1	1	1	Q96KB5		0.0000	0.006	0.0779	3.58	3

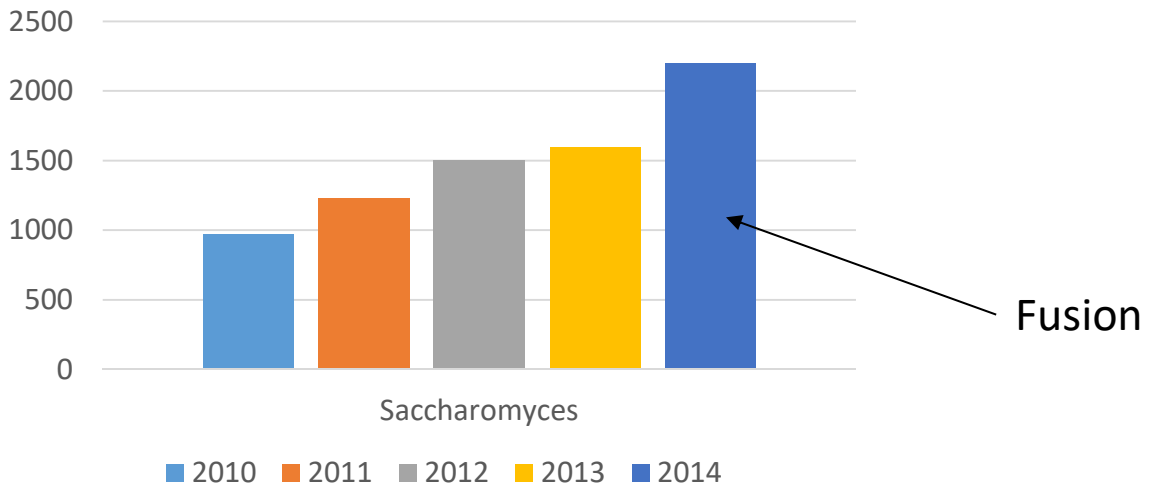
		Accession	Description	Score	Coverage	# Proteins	# Unique Peptides	# Peptides	# PSMs	# AAs	MW [kDa]
101		A0AVT1	Ubiquitin-like modifier-activating enzyme 6 OS=Homo sapiens GN=...	43.86	14.83 %	1	12	12	14	1052	11
102		O00116	Alkyldihydroxyacetonephosphate synthase, peroxisomal OS=Homo...	40.54	25.08 %	1	11	11	11	658	7
103		Q12802	A-kinase anchor protein 13 OS=Homo sapiens GN=AKAP13 PE=1...	32.11	9.14 %	1	11	11	13	2813	30
104		O43684	Mitotic checkpoint protein BUB3 OS=Homo sapiens GN=BUB3 PE=...	38.40	44.51 %	1	11	11	11	328	3
105		O60832	H/ACA ribonucleoprotein complex subunit 4 OS=Homo sapiens GN...	32.86	24.71 %	1	11	11	11	514	5
106		P19525	Interferon-induced, double-stranded RNA-activated protein kinase...	34.26	24.68 %	1	11	11	11	551	6
107		Q8N3D4	EH domain-binding protein 1-like protein 1 OS=Homo sapiens GN=...	43.89	11.36 %	1	11	11	13	1523	16
108		P60228	Eukaryotic translation initiation factor 3 subunit E OS=Homo sapie...	41.87	32.13 %	1	11	11	12	445	5
109		P62495	Eukaryotic peptide chain release factor subunit 1 OS=Homo sapien...	53.24	37.30 %	1	11	11	16	437	4
110		P15170	Eukaryotic peptide chain release factor GTP-binding subunit ERF3A...	45.47	36.27 %	1	11	11	12	499	5

eady

4681/4776 Protein Group(s), 5448/18724 Protein(s), 25168/152137 Peptide(s), 31700/190665 PSM(s), 88527/88527 Search Input(s)

History of standard identifications

Mass Spectometer	HPLC	Gradient Time (min)	Column	Species	Mascot (Protein/Peptide)	Sequest (Protein/Peptide)
Velos	EasynLC Proxeon	75	10	*Saccharomyces cerevisae*	972/3912	1111/4884
	RSLC	120	25	*Saccharomyces cerevisae*	1234/5245	1402/5948
		240		*Saccharomyces cerevisae*	1198/4583	1422/6072
	EasynLC 1000	120	50	*Saccharomyces cerevisae*	1505/8317	1638/8339
				Candida glabrata	1598/7097	
		240		*Saccharomyces cerevisae*	2135/7337	
				Candida albicans	2049/7676	2135/7337
Fusion	EasynLC 1000	120	50	*Saccharomyces cerevisae*	2202/16726	2350/11897
				Candida albicans		



Key questions in proteomics

- What is the protein content of my biological sample?

  => problem of identification

- What is the abundance of my protein of interest?

  => quantification problem

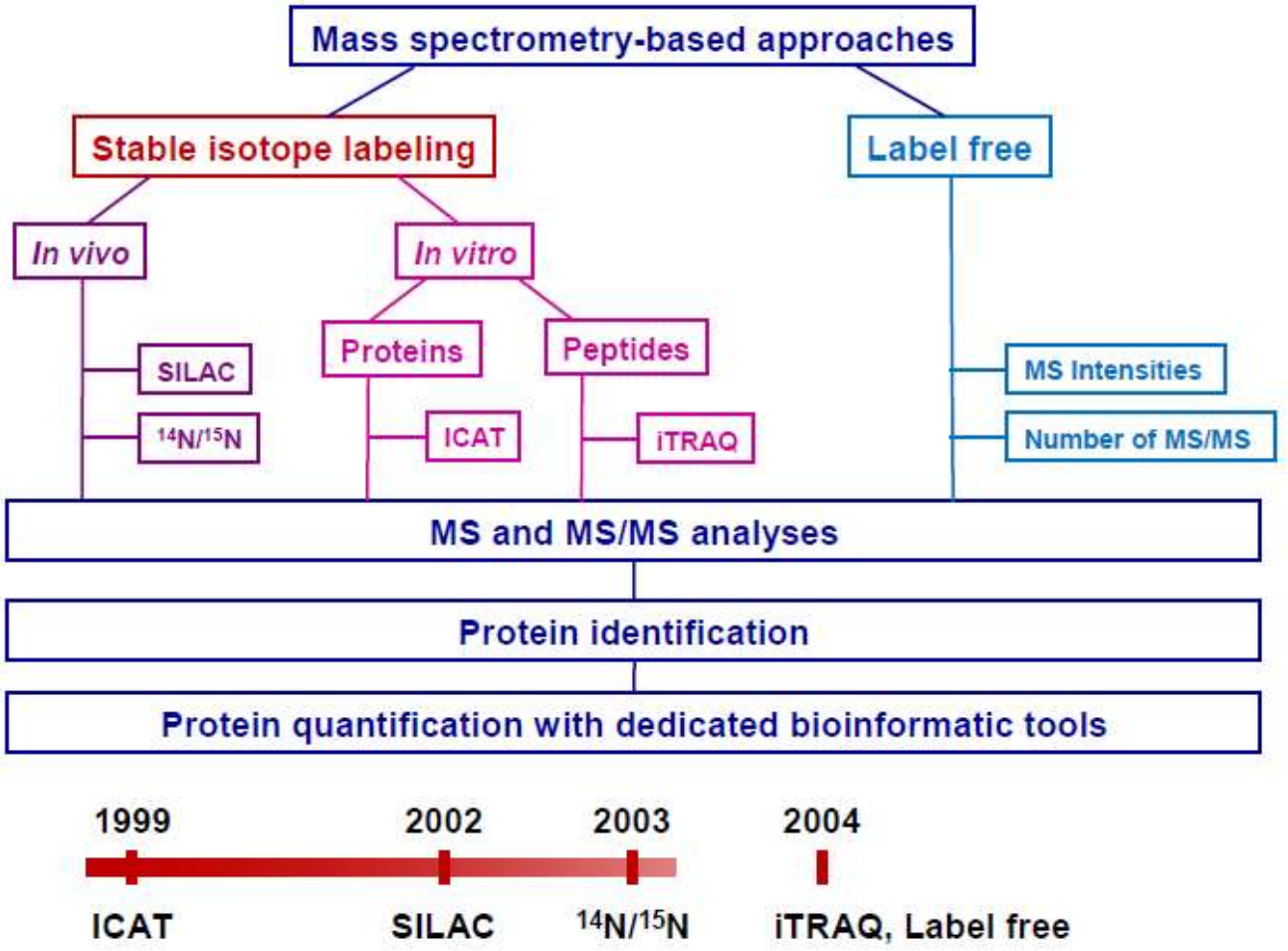
- Relative question: What are the protein abundance variations of the proteomes studied?
- What are the partners of my protein of interest?
- Are there any signature proteins related to a particular biological process?

  => biomarkers identifications and quantifications

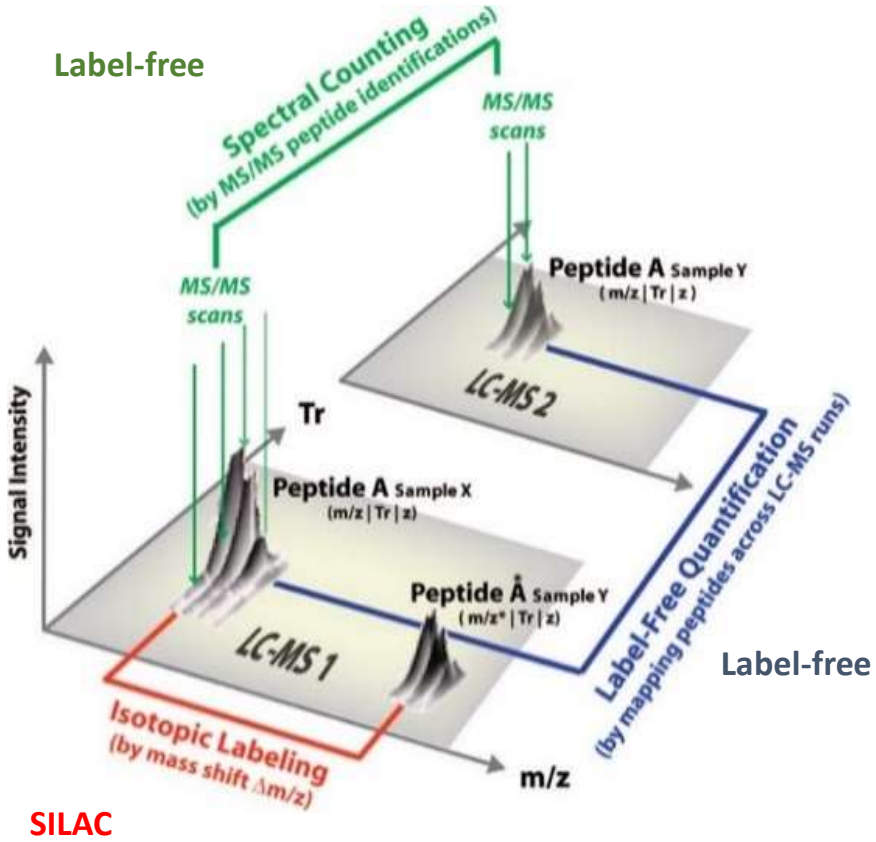
Quantitative proteomics

- Relative quantification
  - Stable isotopes labelling
  - Label-free
  - Metabolic labeling
- Absolute quantification

Quantitative proteomics

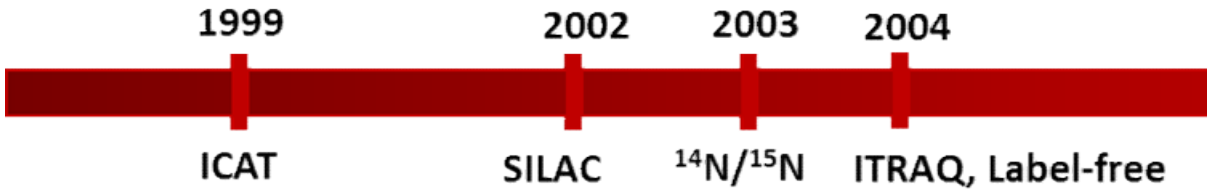


Quantitative proteomics in bottom-up

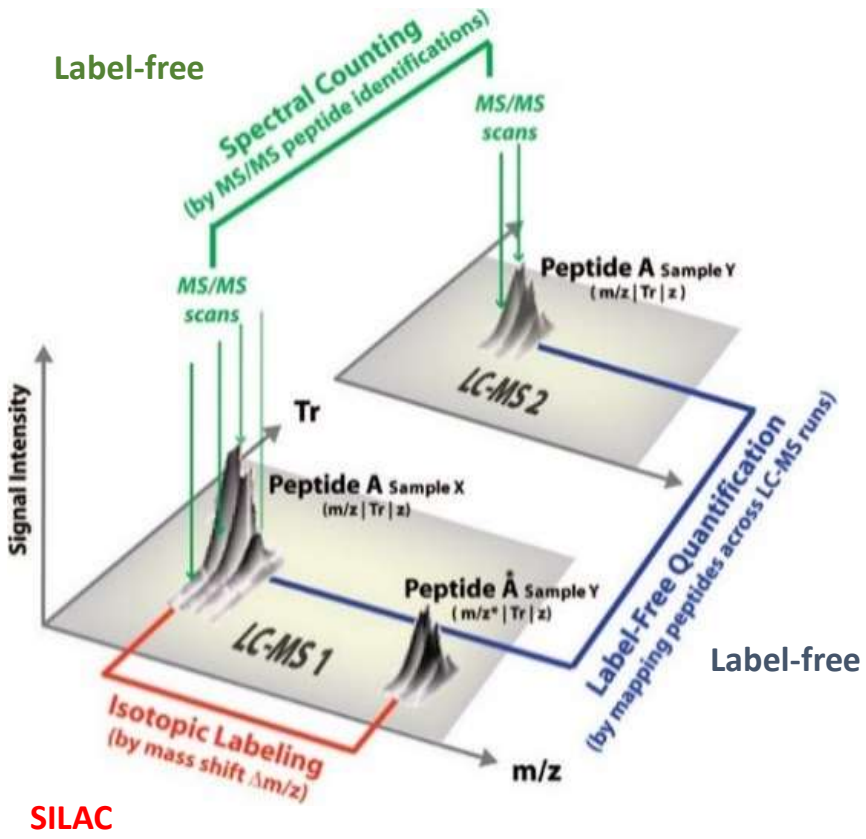


Advantages/Limitations:

- ❑ Label-free:
- ❑ Metabolic Labeling (SILAC, $^{14}N/^{15}N$ – ^{13}C labeling)
- ❑ Chemical labeling (TMT,ITRAQ)



Quantitative proteomics: label-free

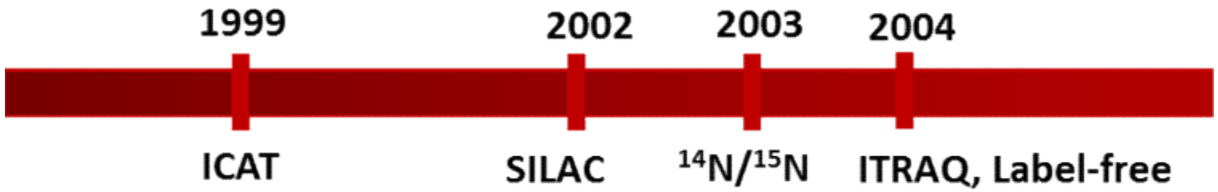


Advantages/Limitations:

❑ Label-free:

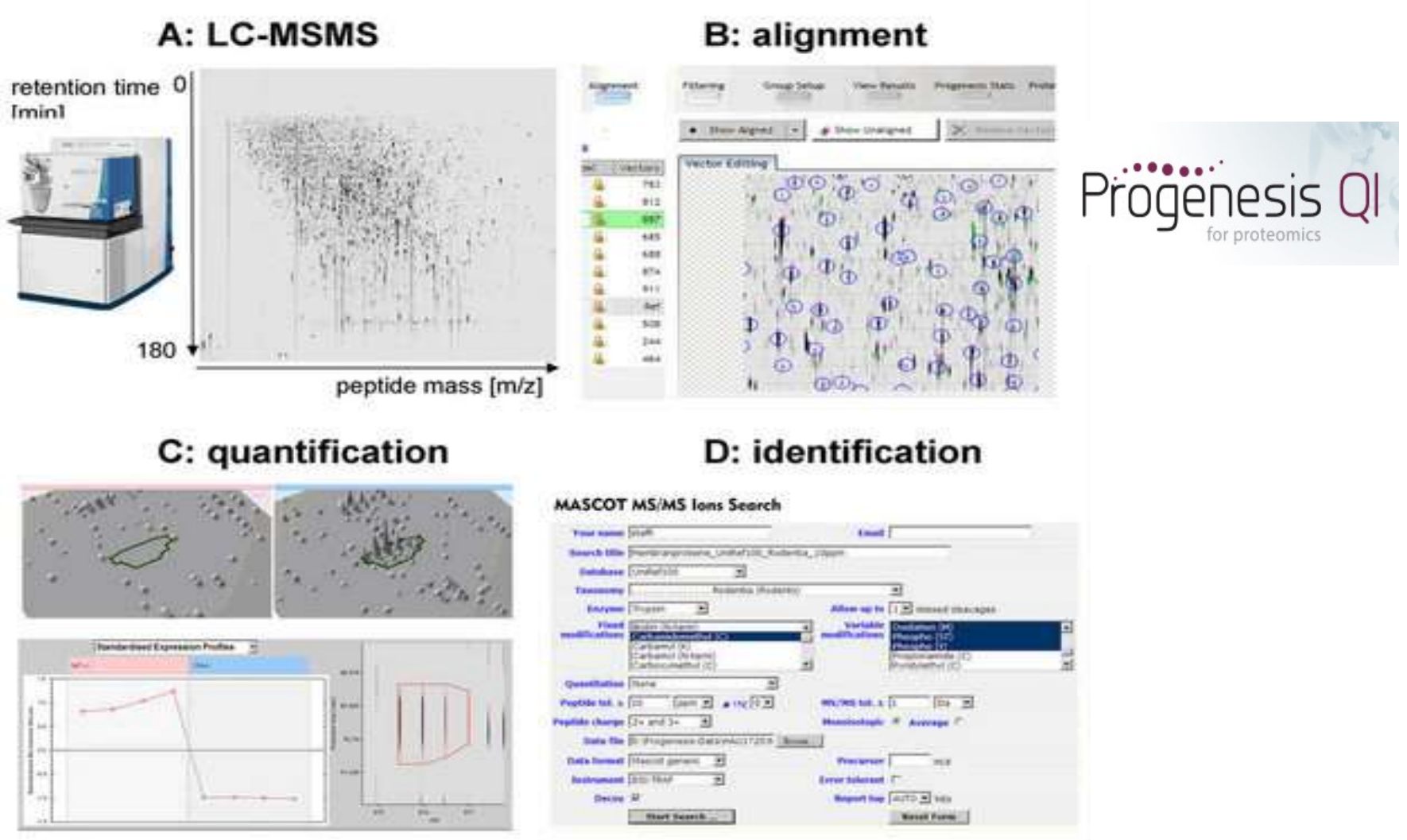
- Simplicity
- Number of identifications
- Reproductibility between runs
- Number of samples to run

Review for Label-free and yeasts:
Leger et al. Methods Mol Biol (2016)



Quantitative proteomics without labeling

Quantification label-free basée sur les intensités MS



Quantitative proteomics without labeling : results

Experiment Design

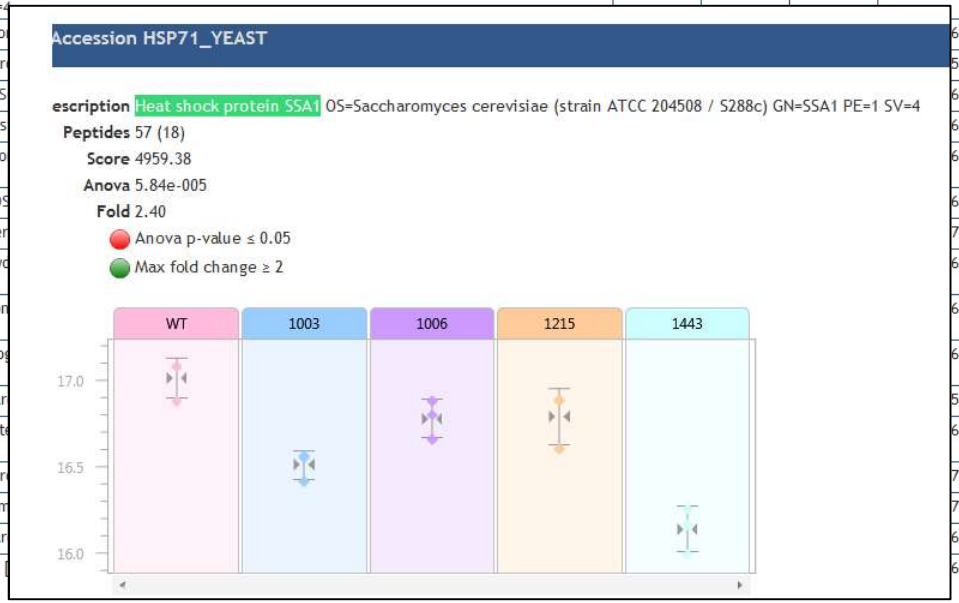
Condition	WT	1003	1006	1215	1443
Replicates	3	3	3	3	3

Proteins

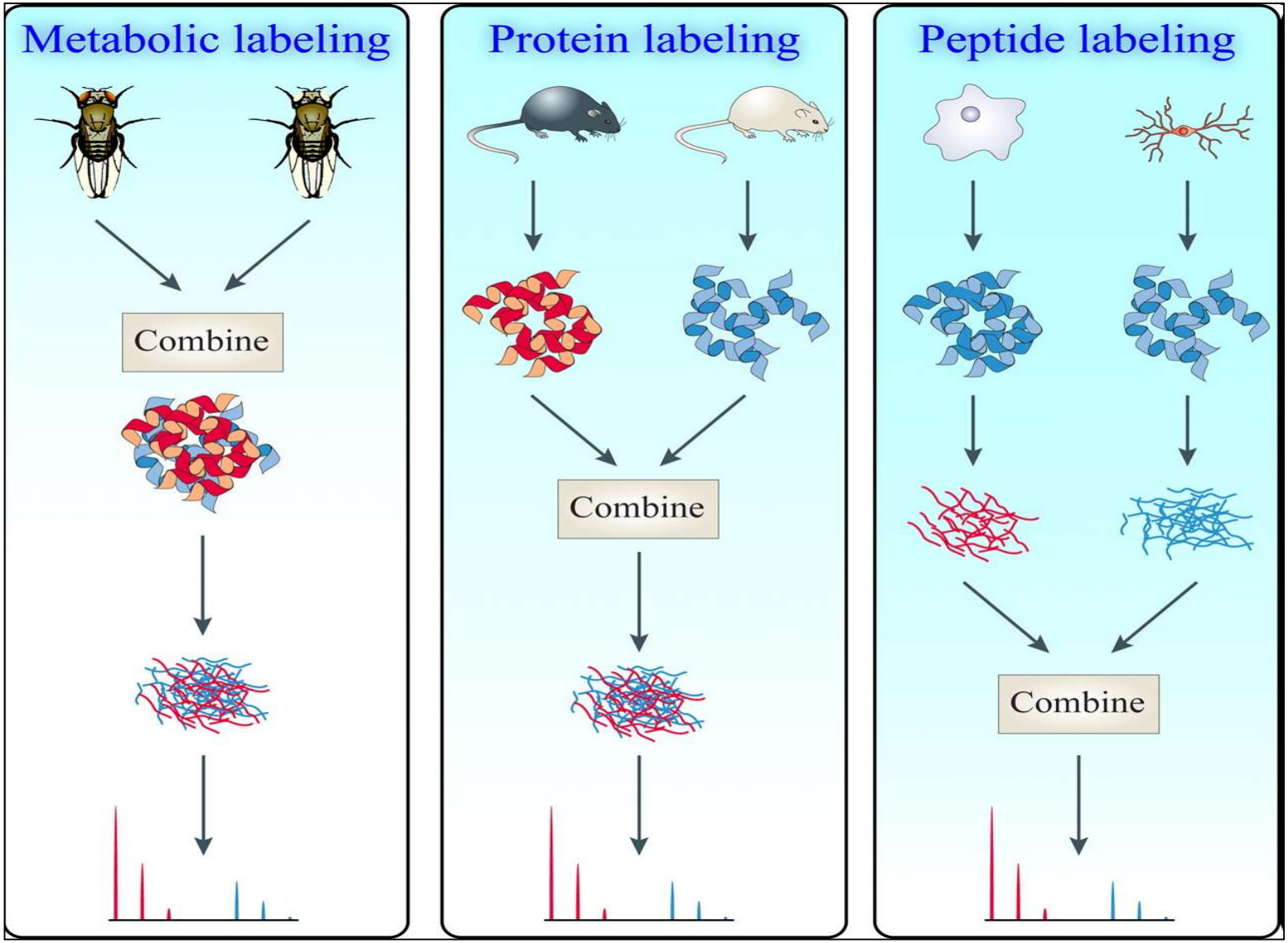
Protein building options

Protein grouping **Group similar proteins**
Protein quantitation **Using only features with no protein conflicts**

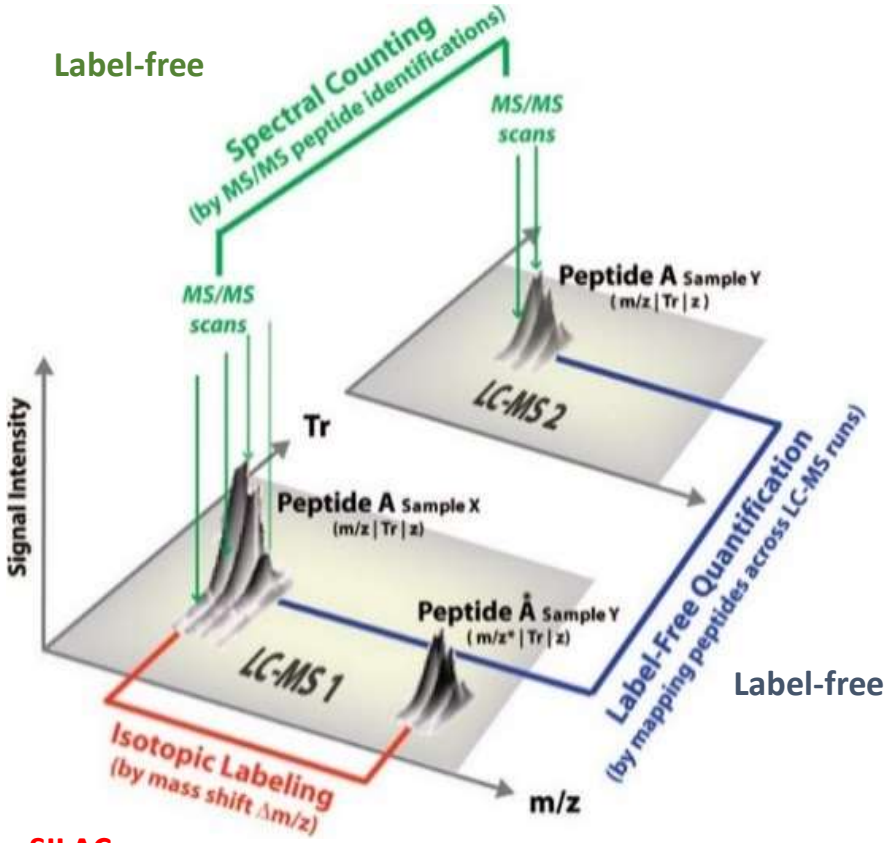
Accession	Peptides	Score	Anova (p)*	Fold	Tags	Description	Average Normalised Abundances				
							WT	1003	1006	1215	1443
HSP71_YEAST	57 (18)	4959.38	5.84e-005	2.40		Heat shock protein SSA1 OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) GN=SSA1 PE=1 SV=4	1.23e+007	7.42e+006	9.72e+006	9.87e+006	5.13e+006
EF2_YEAST	69	4650.11	7.26e-004	2.16		Elongation factor 2 OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) GN=EFT1 PE=1 SV=1	3.40e+007	2.39e+007	2.89e+007	3.28e+007	1.58e+007
FAS1_YEAST	74 (71)	4506.21	9.72e-003	2.33		Fatty acid synthase subunit beta OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) GN=FAS1 PE=1 SV=2	1.03e+007	6.55e+006	8.51e+006	1.01e+007	1.53e+007
EF3A_YEAST	58 (44)	3816.91	7.06e-006	3.32		Elongation factor 3A OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) GN=YEF3 PE=1 SV=4	2.88e+007	1.29e+007	2.33e+007	2.72e+007	8.66e+006
METE_YEAST	46	3373.13	1.85e-006	13.40		5-methyltetrahydropteroyltriglutamate--homocysteine methyltransferase OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) GN=MET6 PE=1 SV=4	4.34e+006	5.55e+006	3.77e+006	5.16e+006	5.05e+007
HS104_YEAST	53	3190.19	6.84e-004	2.29		Heat shock protein 104 OS=Saccharo					3.51e+006
HSP75_YEAST	40 (1)	3062.55	1.21e-006	28.16		Heat shock protein SSB1 OS=Saccharo					1.05e+004
HSP7F_YEAST	39 (32)	2658.69	2.58e-004	2.01		Heat shock protein homolog SSE1 OS					3.50e+006
ENO1_YEAST	31 (15)	2367.12	6.07e-005	2.19		Enolase 1 OS=Saccharomyces cerevis					1.46e+007
ATPA_YEAST	32	2341.09	3.17e-006	2.59		ATP synthase subunit alpha, mitocho SV=5					3.66e+006
SYLC_YEAST	37	2176.12	1.52e-006	2.01		Leucine--tRNA ligase, cytoplasmic OS					1.82e+006
HXKA_YEAST	29 (28)	2162.35	3.17e-004	2.88		Hexokinase-1 OS=Saccharomyces cer					4.61e+006
ALDH6_YEAST	30	2091.58	4.85e-004	2.15		Magnesium-activated aldehyde dehyd GN=ALD6 PE=1 SV=4					2.50e+006
ATPB_YEAST	28	2015.82	4.45e-006	2.39		ATP synthase subunit beta, mitochon SV=2					4.05e+006
G3P1_YEAST	31 (21)	1986.15	8.75e-005	4.16		Glyceraldehyde-3-phosphate dehydro PE=1 SV=3					1.79e+007
HSP74_YEAST	26 (12)	1750.55	0.04	2.68		Heat shock protein SSA4 OS=Sacchar					5.24e+005
PUR92_YEAST	28 (22)	1725.94	6.21e-007	7.07		Bifunctional purine biosynthesis prot PE=1 SV=2					9.63e+006
ADH1_YEAST	24 (17)	1689.13	5.88e-004	2.62		Alcohol dehydrogenase 1 OS=Sacchar					1.03e+007
HSP26_YEAST	18	1538.64	2.83e-006	2.31		Heat shock protein 26 OS=Saccharom					8.43e+006
SAHH_YEAST	27	1535.76	2.79e-006	3.51		Adenosylhomocysteinase OS=Sacchar					1.31e+007
PCKA_YEAST	20	1515.31	3.42e-009	9.67		Phosphoenolpyruvate carboxykinase [SV=2					6.25e+005



Quantitative proteomics with labeling



Quantitative proteomics: metabolic labeling

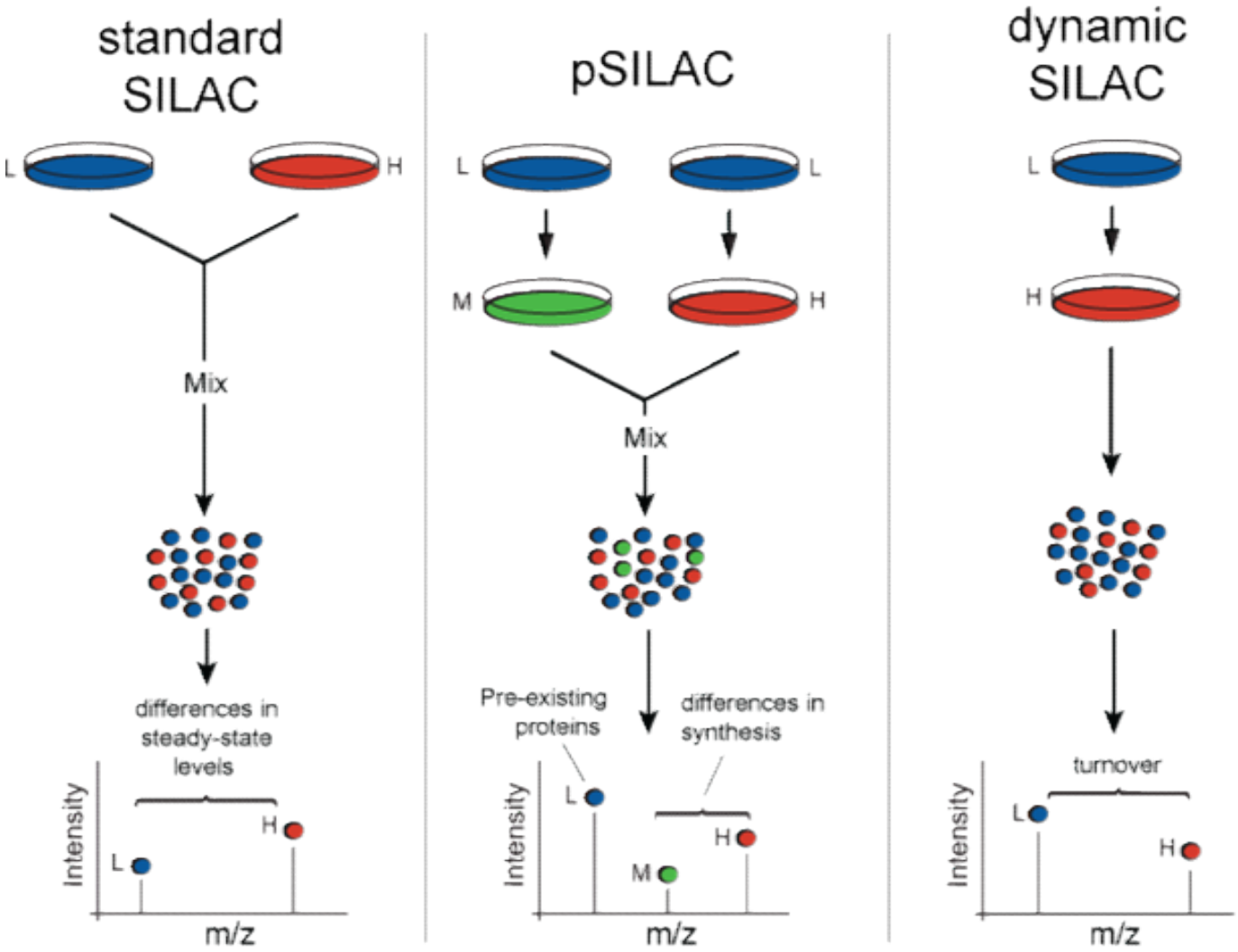


Advantages/Limitations:

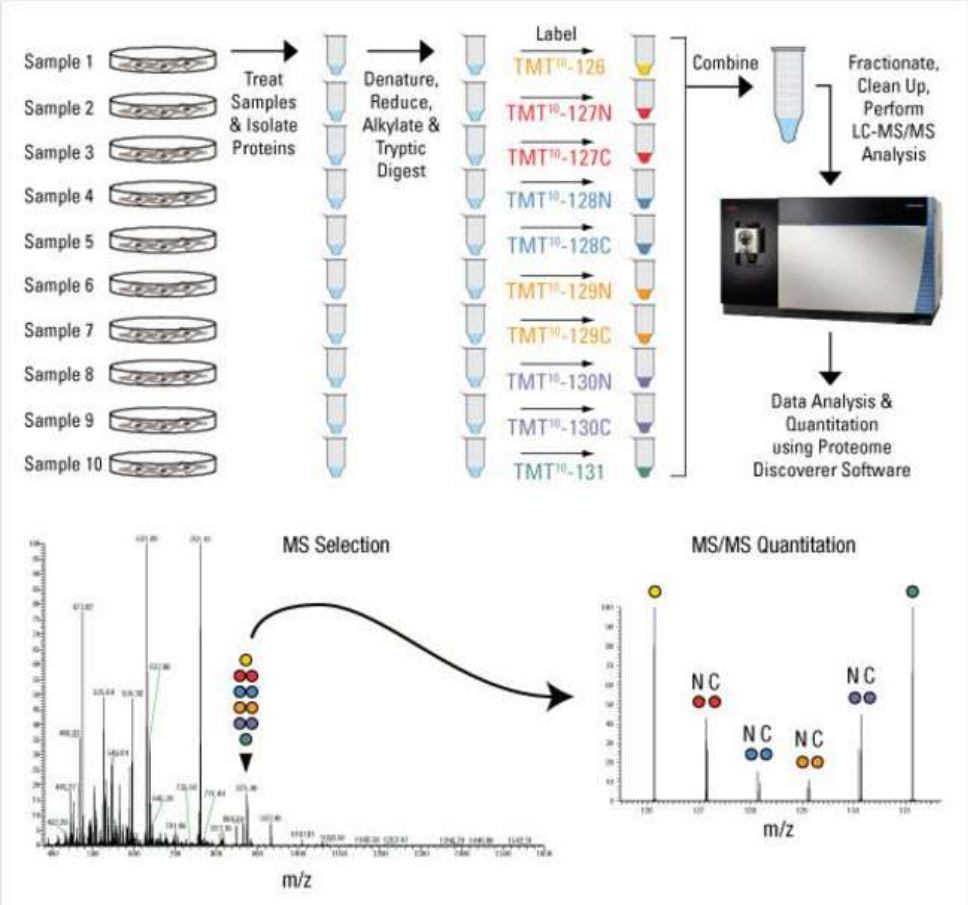
- ❑ SILAC:
  - Multiplexing
  - Reproductibility
  - **2 peaks instead of 1 to analyze by the MS (for 2 samples)**
  - Less identifications
  - Partial labeling
  - Arginine/proline conversion (use of mutants)
  - Trypsin exclusively

- ❑ $^{14}N/^{15}N$ – ^{13}C labeling:
  - Multiplexing
  - Reproductibility
  - **2 peaks instead of 1 to analyze by the MS (for 2 samples)**
  - Less identifications and quantifications
  - Partial labeling
  - **Variable mass shift between heavy and light forms**

SILAC approaches



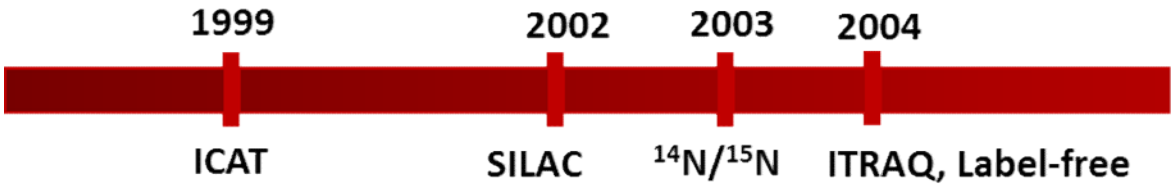
Quantitative proteomics: chemical labeling



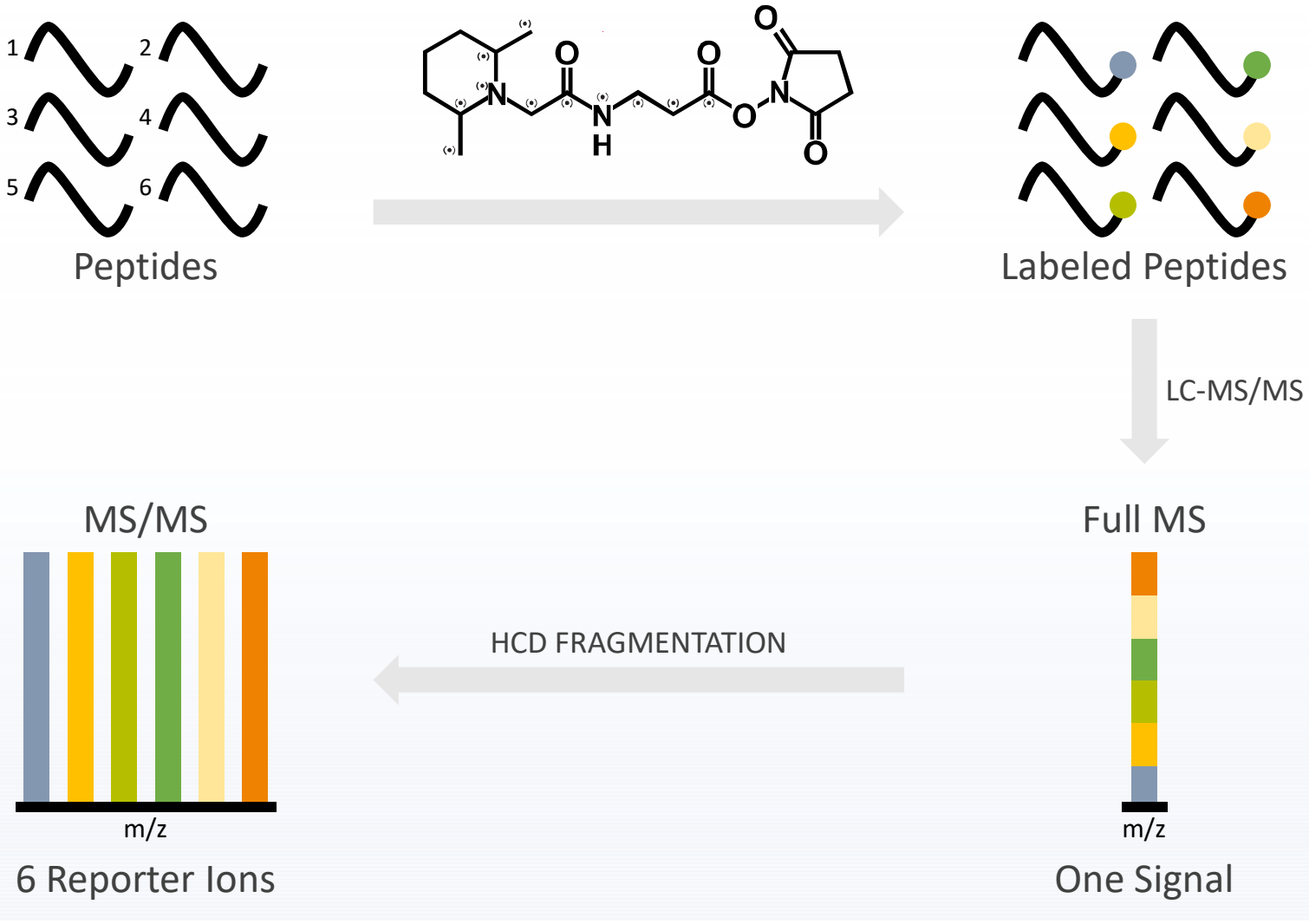
Advantages/Limitations:

❑ Chemical labeling (TMT, ITRAQ)

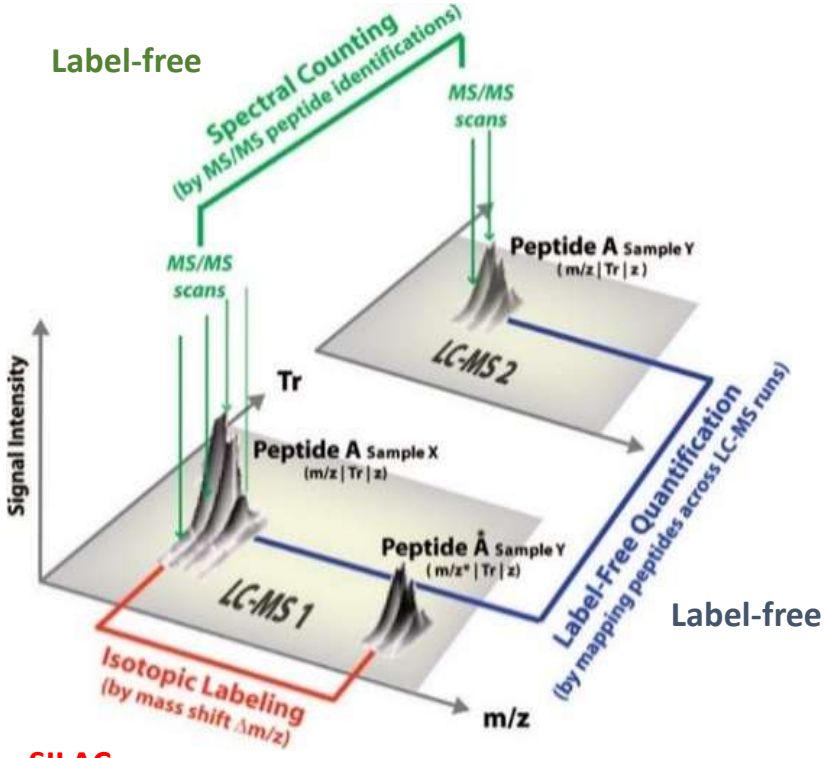
- Multiplexing (until 11plex)
- Reproductibility
- Quantification in MS2 or MS3
- **1 peak instead of N (for N samples) to analyse in MS**
- Amount of materials for the peptide labeling
- Need of resolution in MS2 for quantifications
- Incomplete labeling
- Less identifications and quantifications

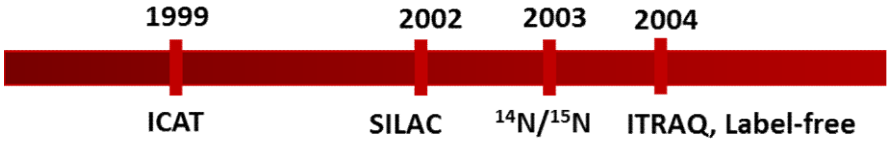


TMT labeling : principles



Quantitative proteomics in bottom-up





Advantages/Limitations:

- ❑ Label-free:
  - Simplicity
  - Number of identifications
  - Reproductibility between runs
  - Number of samples to run

- ❑ SILAC:
  - Multiplexing
  - Reproductibility
  - **2 peaks instead of 1 to analyze by the MS (2 samples)**
  - Less identifications
  - Partial labeling
  - Arginine/proline conversion (use of mutants)
  - Trypsin exclusively

- ❑ $^{14}N/^{15}N$ – ^{13}C labeling:
  - Multiplexing
  - Reproductibility
  - **2 peaks instead of 1 to analyze by the MS (2 samples)**
  - Less identifications and quantifications
  - Partial labeling
  - **Variable mass shift between heavy and light forms**

Metabolic labeling

- ❑ Chemical labeling (TMT,ITRAQ)
  - Multiplexing
  - Reproductibility
  - 1 peak instead of N to analyse in MS (N samples)
  - Amount of materials for the peptide labeling
  - Need of resolution in MS2 for quantifications
  - Incomplete labeling
  - Less identifications and quantifications

Quantitative proteomics without labeling : export

Progenesis QI for proteomics

													Normalized abundance							
													Sain						Patient	
3	Accession	Peptide cou	Unique pep	Confidence	Anova (p)	q Value	Max fold ch	Power	Highest me	Lowest mea	Mass	Description	1845007-F1	1845007-F3	1845007-F5	1845007-F7	1845007-F9	1845007-F1	1845007-F2	1845007-F4
4	P40197	12	12	545.23	2.48E-07	4.74E-05	32.042418	1	Patient	Sain	60.921	Platelet glycoprotein V OS	1447.8857	898.47877	2114.249	3517.982	2506.9091	1186.7152	33178.605	73995.78
5	P02776	2	2	134.77	2.63E-07	4.74E-05	78.817355	1	Patient	Sain	10.838	Platelet factor 4 OS=Homc	3812.4369	3755.8358	1044.3911	2939.4867	3862.0883	1110.5549	333829.66	68071.51
6	Q13201	6	6	221.4	5.33E-07	6.40E-05	72.663027	1	Patient	Sain	138.023	Multimerin-1 OS=Homo sa	160.48528	939.93933	155.7663	732.57482	572.39757	196.60688	75519.303	23299.12
7	P04114	336	334	28302.18	7.99E-06	0.0006593	2.1384481	1	Patient	Sain	515.283	Apolipoprotein B-100 OS=	14117253	10601958	8472023.6	9982572.2	10056625	11898407	24415094	1744463
8	P07996	42	42	2678.71	9.16E-06	0.0006593	143.08362	1	Patient	Sain	129.3	Thrombospondin-1 OS=Hc	11604.367	10346.75	4059.3393	5652.5533	2121.1343	2367.1543	2303007.2	83175.08
9	Q15485	7	7	289.57	1.42E-05	0.0007087	24.427918	0.9999997	Patient	Sain	33.98	Ficolin-2 OS=Homo sapier	8778.176	6132.5206	10104.107	8194.6312	8690.2929	1737.9715	185819.2	105024.9
10	P10720	3	3	208.63	1.49E-05	0.0007087	36.552808	0.9999997	Patient	Sain	11.545	Platelet factor 4 variant O	32881.787	11758.269	1451.6395	5448.2604	4187.5362	4196.5766	546497.96	239272.0
11	Q12884	2	2	70.47	1.58E-05	0.0007087	Infinity	0.9999996	Patient	Sain	87.657	Prolyl endopeptidase FAP	0	0	0	0	0	0	961.55592	106.477
12	Q15061	1	1	3.67	3.71E-05	0.0014832	23.811179	0.9999905	Sain	Patient	74.843	WD repeat-containing prc	567513.12	751475.75	852125.74	1189728.8	625225.18	78135.037	46681.126	36585.06
13	P04075	8	8	371.85	4.58E-05	0.0016488	4.2937733	0.9999814	Patient	Sain	39.395	Fructose-bisphosphate al	33219.801	26234.171	28730.699	26696.908	33816.665	32024.353	128602.25	168571.3
14	P09486	10	9	463.23	6.25E-05	0.0020451	6.1746973	0.999953	Patient	Sain	34.61	SPARC OS=Homo sapiens (	9522.5101	37758.131	36721.019	20388.641	15117.423	28228.93	299004.24	97804.24
15	P10124	1	1	70.16	7.11E-05	0.002134	272.44302	0.9999321	Patient	Sain	17.641	Serglycin OS=Homo sapier	67.740323	0	958.26875	53.15204	160.31231	238.84011	105180.09	42246.24
16	P02775	6	6	388.98	8.21E-05	0.0022557	28.012451	0.9998996	Patient	Sain	13.885	Platelet basic protein OS=	61607.9	109352.83	105910.89	67216.894	60865.635	22752.188	3524260.5	1692399.5
17	Q9H1K0	2	2	12.94	8.77E-05	0.0022557	1.8919044	0.9998799	Patient	Sain	88.815	Rabenosyn-5 OS=Homo sa	871721.79	1194024.5	1103071.3	1121288.3	1020557.4	1361492.9	1506333.1	1757466.5
18	P35542	4	4	338.41	0.0001236	0.0029659	16.656374	0.9997122	Patient	Sain	14.737	Serum amyloid A-4 protein	192222.76	102064.74	9416.0169	8023.3014	46690.651	107524.48	1543185.2	1741696.5
19	P80188	6	6	223.58	0.0001847	0.0041561	9.1194461	0.9992657	Patient	Sain	22.574	Neutrophil gelatinase-as	2265.4484	1473.9176	6203.72	3676.0015	7827.3304	4603.3796	66633.817	26485.80
20	P02144	4	4	209.23	0.0002392	0.0050651	6.5032553	0.9987201	Patient	Sain	17.173	Myoglobin OS=Homo sapi	2754.3333	2223.1201	1035.6898	1813.5104	1462.164	5472.3528	17977.389	20813.25
21	P05067	5	5	179.76	0.000297	0.0059397	18.947351	0.9980148	Patient	Sain	86.888	Amyloid beta A4 protein C	408.20926	53.510036	338.118	494.28658	258.47482	516.31147	8561.0101	2161.775
22	Q9NPH3	7	7	223.36	0.0003417	0.0064734	2.9534965	0.9973949	Patient	Sain	65.377	Interleukin-1 receptor acc	19269.326	23550.392	17918.705	19071.463	29165.24	20448.453	74326.28	82743.30
23	P22352	7	7	322.73	0.0005632	0.0101356	2.3655577	0.9936247	Patient	Sain	25.537	Glutathione peroxidase 3	356078.09	500176.63	411199.92	354267.54	608490.2	620677.21	1139291.1	1214566.8
24	A0A075B6I0	1	1	30.88	0.0005981	0.0102519	15.884721	0.9929515	Sain	Patient	12.806	Immunoglobulin lambda	14017.761	25223.891	6493.2483	70640.84	15600.774	36585.067	4641.3232	1487.298
25	P05155	33	33	2513.9	0.0007389	0.0120898	5.4023431	0.9900925	Patient	Sain	55.119	Plasma protease C1 inhib	741421.19	592478.75	377555.54	465885.04	2331779.9	7791080.5	11109177	10511766
26	A0A0C4DH2	2	2	100.38	0.0013344	0.0208833	3.8865369	0.9766266	Sain	Patient	12.999	Immunoglobulin heavy va	50500.146	64330.439	38938.111	121804.39	27881.723	32180.678	24057.435	17302.863
27	P14780	6	6	207.75	0.0014988	0.0221418	62.066206	0.972746	Patient	Sain	78.408	Matrix metalloproteinase	794.97907	136.39171	262.24899	313.84328	0	555.64619	58928.664	9636.858
28	P02649	27	27	1941.9	0.0015378	0.0221418	2.4710215	0.971824	Patient	Sain	36.132	Apolipoprotein E OS=Hom	2552396.1	1536331.8	853676.88	1079240.3	1188651.7	1327575.2	2857655.7	2177433.7
29	Q3C1V8	1	1	14.2	0.0019728	0.0268284	5.551993	0.9615339	Patient	Sain	25.917	Brain-specific homeobox	119.76184	56.50442	16.821931	8.5075688	3888.8817	85.190432	6077.0842	3199.583
30	P02652	11	11	995.34	0.0020124	0.0268284	2.4014067	0.9606031	Patient	Sain	11.168	Apolipoprotein A-II OS=Hc	34056410	26771849	14632439	15663194	42268641	26890744	89512485	4323663
31	P08571	12	12	937.27	0.0025308	0.0320335	1.6971067	0.9485492	Sain	Patient	40.051	Monocyte differentiation	797874.17	434080.26	499432.28	477652.53	733837.28	418725.29	352767.33	318079.3
32	P26927	31	31	1222.76	0.0025808	0.0320335	3.3123536	0.9474046	Patient	Sain	80.268	Hepatocyte growth factor-	190652.76	43028.557	155598.93	60550.701	75682.634	141408.77	141623.19	380083.05
33	P02655	6	6	819.16	0.0030504	0.0361789	4.7206762	0.9368237	Patient	Sain	11.277	Apolipoprotein C-II OS=Hc	1281811.8	751583.4	312951.49	1013499	603757.76	1691499.6	4859804.1	137623
34	P02671	8	8	369.69	0.0031158	0.0361789	191.8384	0.9353762	Patient	Sain	94.914	Fibrinogen alpha chain O	1792.4021	4843.2609	2616.5938	746.39308	685.49376	8280.2447	1254362.7	107899
35	P02763	17	17	1378.36	0.0036234	0.0407586	2.8533872	0.924364	Sain	Patient	23.497	Alpha-1-acid glycoprotein	67476432	40320887	40796528	53919933	106057189	35895348	25642029	2274357
36	Q14831	1	1	23.81	0.0038979	0.0425178	2.9788784	0.9185746	Sain	Patient	102.185	Metabotropic glutamate r	26519.187	83446.172	72384.114	34456.805	83227.769	36550.192	8908.5356	31759.88
37	P02647	32	32	2367.65	0.0040205	0.0425648	2.2688435	0.9160271	Patient	Sain	30.759	Apolipoprotein A-I OS=Ho	105663566	161836184	96430549	75237801	158864474	104736598	166481100	27802701
38	P01876	21	21	1557.26	0.0044316	0.0455765	3.5420267	0.9076475	Sain	Patient	37.631	Ig alpha-1 chain C region (	10958465	3390608.1	19371147	11129440	8232536	4863588.4	3403586.2	5212718.5
39	Q9UNW1	8	8	227	0.005259	0.0525841	1.7900311	0.8915219	Sain	Patient	55.016	Multiple inositol polypho	62640.486	55052.459	58667.71	30459.856	63293.849	79838.921	34467.455	30358.52
40	Q16853	2	2	81.52	0.0054293	0.0528194	6.1436713	0.8883207	Sain	Patient	84.568	Membrane primary amine	24451.336	3442.0472	3308.1744	4995.8869	4135.7996	9790.8939	236.61601	194.7495
41	P01833	8	8	219.66	0.006944	0.0656448	9.1555604	0.8614401	Sain	Patient	83.232	Polymeric immunoglobuli	14801.731	38882.249	8643.925	36538.175	31852.344	72666.846	1025.7634	2483.330
42	P27169	28	27	2064.1	0.0071604	0.0656448	1.8486226	0.8578184	Patient	Sain	39.706	Serum paraoxonase/aryle	3861076.2	5186138.5	2149700.4	5390878.2	2346566.5	3471134.2	5925120.8	6280114.6
43	P60174	1	1	28.75	0.0074	0.0656448	2.9166777	0.8538651	Patient	Sain	30.772	Triosephosphate isomera	4225.6237	1794.3666	1066.6861	1210.1584	3775.1162	2981.76	10398.334	4675.589
44	P00736	41	39	3067.37	0.0074771	0.0656448	1.4866701	0.8526056	Sain	Patient	80.067	Complement C1r subcomp	5434650.2	5169551.9	3906709.5	4169170.8	3458132.5	4641351.5	4547911.2	2638367.5
45	P09871	37	37	2588.95	0.0085558	0.0726349	1.93603	0.835606	Sain	Patient	76.635	Complement C1s subcomp	5789892.8	5490839.9	3882940.1	5662793.4	5033975.1	2280238.9	1956631.4	2302798.6
46	A0A0B4J1V0	3	3	104.52	0.0086769	0.0726349	2.7965659	0.833766	Sain	Patient	12.917	Immunoglobulin heavy va	92353.063	20397.701	46790.32	87239.628	43135.867	81618.911	38454.812	17117.343
47	P01619	4	4	262.21	0.0095577	0.0781901	2.30903	0.8207663	Sain	Patient	12.549	Immunoglobulin kappa va	409287.43	332471.99	228180.64	675002.86	220569.52	187876.28	139130.56	148976.38
48	Q08380	21	21	1485.84	0.0115602	0.0900161	2.528023	0.7934679	Sain	Patient	65.289	Galectin-3-binding protei	938408.99	614190.09	1091308.6	2622296.8	1676638.8	2590031.7	640186.94	906404.59
49	P06732	9	9	314.24	0.0117011	0.0900161	5.8938011	0.7916528	Patient	Sain	43.074	Creatine kinase M-type OS	7288.867	9691.0748	1004.0051	6516.3248	4494.7392	6004.4549	19154.379	89665.94

Quantitative proteomics without labeling : export

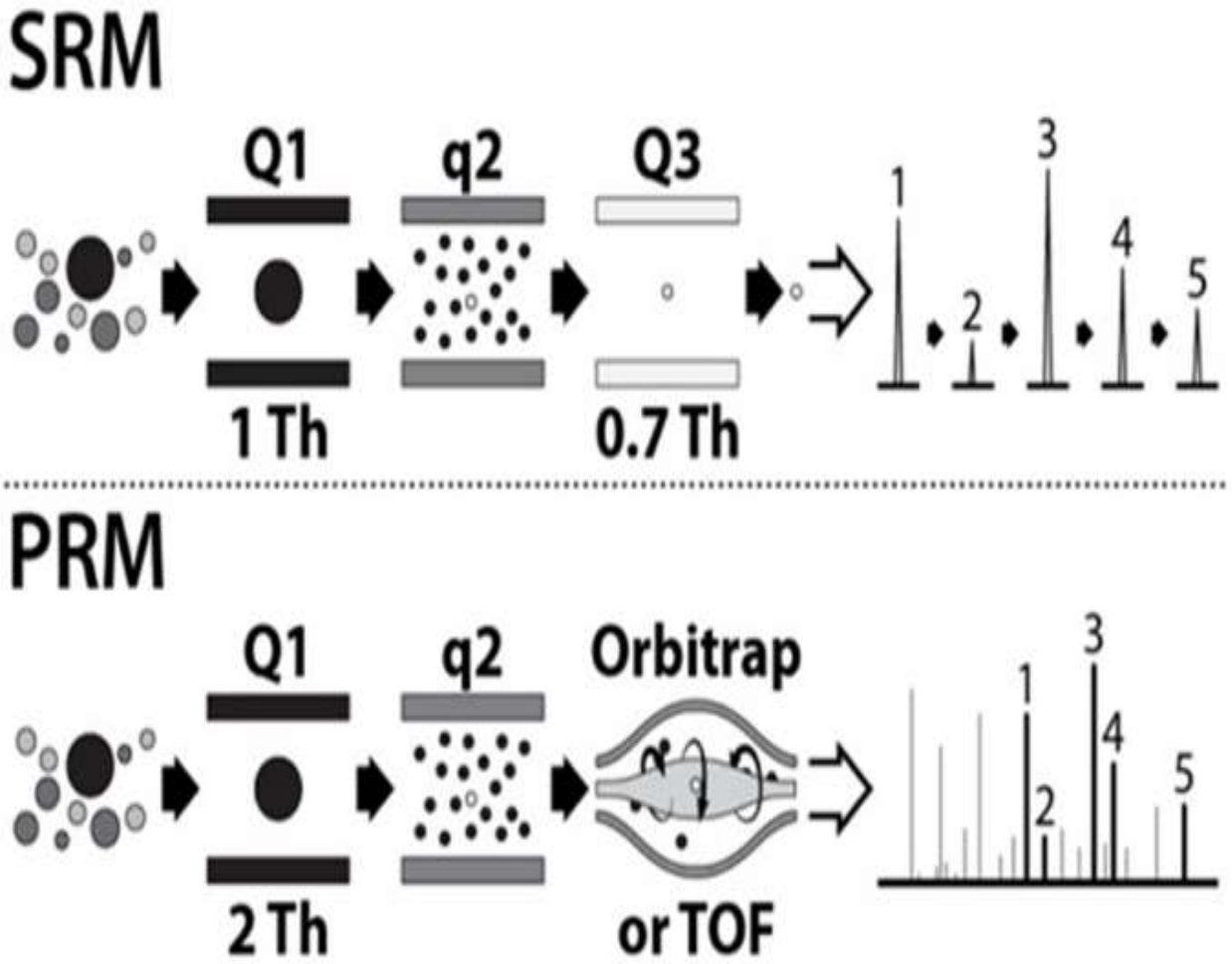
aifMsms.txt	18-Oct-18 6:45 PM	Fichier TXT	0 Ko
allPeptides.txt	18-Oct-18 6:47 PM	Fichier TXT	63,279 Ko
evidence.txt	18-Oct-18 6:44 PM	Fichier TXT	5,765 Ko
matchedFeatures.txt	18-Oct-18 6:45 PM	Fichier TXT	0 Ko
modificationSpecificPeptides.txt	18-Oct-18 6:44 PM	Fichier TXT	2,281 Ko
msms.txt	18-Oct-18 6:44 PM	Fichier TXT	26,108 Ko
msmsScans.txt	18-Oct-18 6:45 PM	Fichier TXT	15,918 Ko
msScans.txt	18-Oct-18 6:47 PM	Fichier TXT	2,211 Ko
mzRange.txt	18-Oct-18 6:47 PM	Fichier TXT	95 Ko
Oxidation (M)Sites.txt	18-Oct-18 6:44 PM	Fichier TXT	125 Ko
parameters.txt	18-Oct-18 6:44 PM	Fichier TXT	2 Ko
peptides.txt	18-Oct-18 6:44 PM	Fichier TXT	2,741 Ko
proteinGroups.txt	18-Oct-18 6:43 PM	Fichier TXT	1,108 Ko
summary.txt	18-Oct-18 6:44 PM	Fichier TXT	2 Ko
tables.pdf	18-Oct-18 6:45 PM	Adobe Acrobat D...	40 Ko

Protein IDs	Majority protein IDs	Peptide counts (all)	Peptide counts (razor+unique)	Peptide counts (unique)	Fasta headers	Number o
C1_00060W_A	C1_00060W_A	10	10	10	>C1_00060W_A translated using codon table 12 (512 amino acids) Verified ORF; (orf19.6109) Transcriptiona	
C1_00070W_A	C1_00070W_A	12	12	12	>C1_00070W_A translated using codon table 12 (362 amino acids) Verified ORF; (orf19.6105) Mevalonate dip	
C1_00110W_A	C1_00110W_A	14	14	14	>C1_00110W_A translated using codon table 12 (540 amino acids) Verified ORF; (orf19.6099) Chaperonin-con	
C1_00140W_A	C1_00140W_A	5	5	5	>C1_00140W_A translated using codon table 12 (1018 amino acids) Verified ORF; (orf19.6092) Kelch repeat	
C1_00150C_A	C1_00150C_A	1	1	1	>C1_00150C_A translated using codon table 12 (622 amino acids) Verified ORF; (orf19.6091) Beta-arrestin-	
C1_00160C_A	C1_00160C_A	4	4	4	>C1_00160C_A translated using codon table 12 (400 amino acids) Verified ORF; (orf19.6090) Putative nucle	
C1_00170W_A	C1_00170W_A	9	9	9	>C1_00170W_A translated using codon table 12 (579 amino acids) Verified ORF; (orf19.6086) Putative 2-iso	
C1_00180W_A	C1_00180W_A	3	3	3	>C1_00180W_A translated using codon table 12 (200 amino acids) Verified ORF; (orf19.6085) Ribosomal prot	
C1_00210C_A	C1_00210C_A	2	2	2	>C1_00210C_A translated using codon table 12 (384 amino acids) Verified ORF; (orf19.6082) Ortholog(s) ha	
C1_00220W_A;C4_04530C_A	C1_00220W_A	5;1	5;1	5;1	>C1_00220W_A translated using codon table 12 (544 amino acids) Verified ORF; (orf19.6081) Gl	
C1_00320W_A	C1_00320W_A	2	2	2	>C1_00320W_A translated using codon table 12 (261 amino acids) Uncharacterized ORF; (orf19.6076) Ortholo	
C1_00330C_A	C1_00330C_A	2	2	2	>C1_00330C_A translated using codon table 12 (182 amino acids) Uncharacterized ORF; (orf19.6075) Putativ	
C1_00340W_A	C1_00340W_A	4	4	4	>C1_00340W_A translated using codon table 12 (248 amino acids) Verified ORF; (orf19.6074) Essential prot	
C1_00380C_A	C1_00380C_A	7	7	7	>C1_00380C_A translated using codon table 12 (745 amino acids) Uncharacterized ORF; (orf19.6071) Ortholo	
C1_00400W_A	C1_00400W_A	7	7	7	>C1_00400W_A translated using codon table 12 (382 amino acids) Uncharacterized ORF; (orf19.6068) Putativ	
C1_00410C_A	C1_00410C_A	13	13	12	>C1_00410C_A translated using codon table 12 (542 amino acids) Uncharacterized ORF; (orf19.6066) Hexadec	
C1_00420W_A	C1_00420W_A	7	7	7	>C1_00420W_A translated using codon table 12 (323 amino acids) Uncharacterized ORF; (orf19.6065) RNA pol	
C1_00440W_A	C1_00440W_A	11	11	11	>C1_00440W_A translated using codon table 12 (478 amino acids) Uncharacterized ORF; (orf19.6063) Putativ	
C1_00450C_A	C1_00450C_A	2	2	2	>C1_00450C_A translated using codon table 12 (150 amino acids) Uncharacterized ORF; (orf19.6062.3) Mitoc	
C1_00460W_A	C1_00460W_A	1	1	1	>C1_00460W_A translated using codon table 12 (106 amino acids) Verified ORF; (orf19.6062) Putative TIM23	
C1_00480C_A	C1_00480C_A	4	4	4	>C1_00480C_A translated using codon table 12 (751 amino acids) Uncharacterized ORF; (orf19.6060) YEF3-su	
C1_00490C_A	C1_00490C_A	2	2	2	>C1_00490C_A translated using codon table 12 (119 amino acids) Verified ORF; (orf19.6059) Putative gluta	
C1_00500C_A	C1_00500C_A	3	3	3	>C1_00500C_A translated using codon table 12 (342 amino acids) Uncharacterized ORF; (orf19.6058) Putativ	
C1_00560W_A	C1_00560W_A	2	2	2	>C1_00560W_A translated using codon table 12 (390 amino acids) Verified ORF; (orf19.6052) Putative co-ch	
C1_00590W_A	C1_00590W_A	14	14	14	>C1_00590W_A translated using codon table 12 (426 amino acids) Uncharacterized ORF; (orf19.6047) Transla	
C1_00610W_A	C1_00610W_A	3	3	3	>C1_00610W_A translated using codon table 12 (590 amino acids) Verified ORF; (orf19.6045) Phosphatidylse	

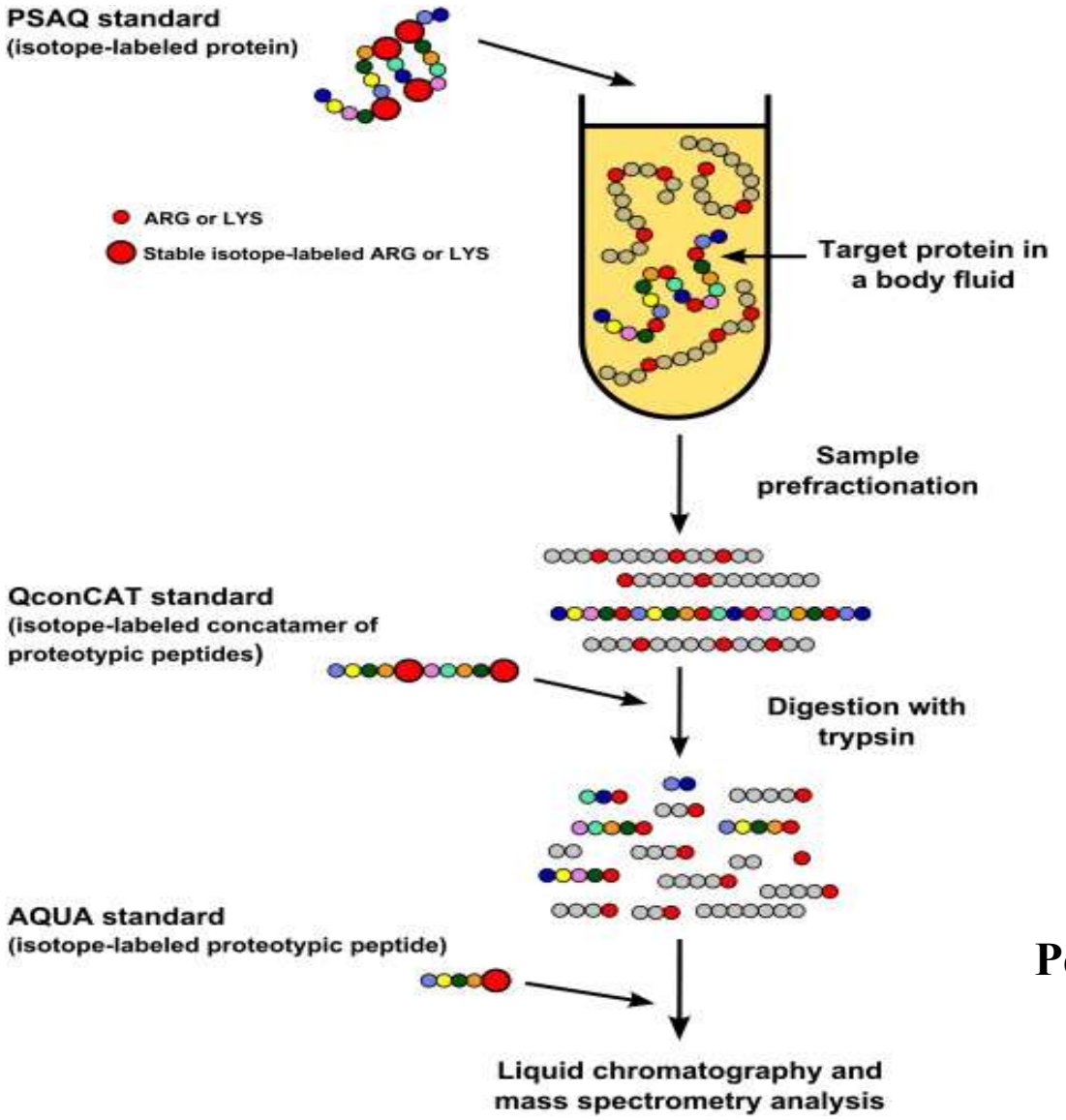
Quantification output formats

File name	File content
Protein/peptide quantification	Protein/peptide expression values can also be obtained from an MS--based proteomics experiment and then this data and metadata is used for performing the quantification analysis of peptides and proteins.
Metadata	A term used to describe data that provides additional information about a particular data set. This information can include how, when and where the data set was generated and what standards were used. In the proteomics context the addition of metadata such as peptide and protein identifications and quantification of their expression values gives meaning to a simple collection of mass spectra output files.

Targeted proteomics : PRM mode



Absolute quantification



PSAQ

QconCAT

Peptides AQUA

Key questions in proteomics

- What is the protein content of my biological sample?

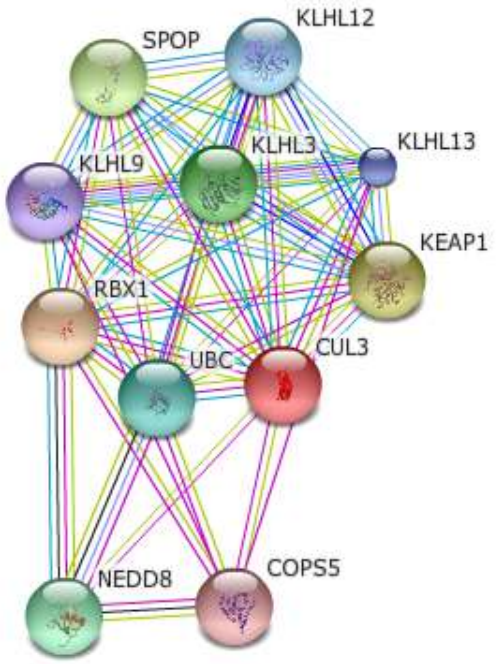
  => problem of identification
- What is the abundance of my protein of interest?

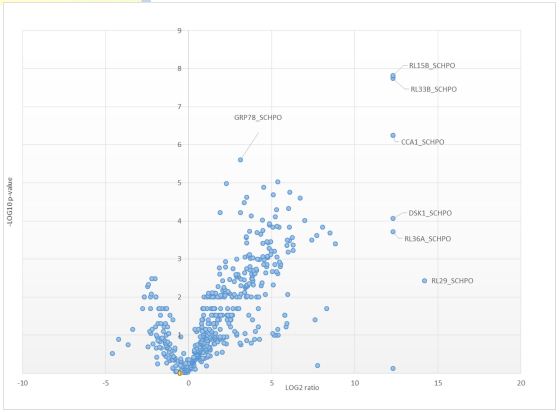
  => quantification problem
- Relative question: What are the protein abundance variations of the proteomes studied?
- What are the partners of my protein of interest?
- Are there any signature proteins related to a particular biological process?

  => biomarkers identifications and quantifications

Co-immunoprecipitation

		Accession	Description	Score A3	Score B3	Score C3
1	☐	Q13618	Cullin-3 OS=Homo sapiens GN=CUL3 PE=1 SV=2 - [CUL3...		1172.08	547.08
2	☐	Q86VP6	Cullin-associated NEDD8-dissociated protein 1 OS=Homo...	0.00	394.42	0.00
3	☐	P62877	E3 ubiquitin-protein ligase RBX1 OS=Homo sapiens GN=R...		251.21	123.68
4	☐	P07437	Tubulin beta chain OS=Homo sapiens GN=TUBB PE=1 SV...		199.36	136.63
5	☐	Q9Y2M5	Kelch-like protein 20 OS=Homo sapiens GN=KLHL20 PE=...		164.62	78.37
6	☐	Q9P2N7	Kelch-like protein 13 OS=Homo sapiens GN=KLHL13 PE=...		158.68	86.54
7	☐	P68371	Tubulin beta-4B chain OS=Homo sapiens GN=TUBB4B PE...		150.00	143.23
8	☐	Q9P2K6	Kelch-like protein 42 OS=Homo sapiens GN=KLHL42 PE=...		149.87	62.55
9	☐	P05141	ADP/ATP translocase 2 OS=Homo sapiens GN=SLC25A5 P...		148.12	57.72
10	☐	Q92905	COP9 signalosome complex subunit 5 OS=Homo sapiens...		142.70	29.32
11	☐	Q99627	COP9 signalosome complex subunit 8 OS=Homo sapiens...		135.68	37.43
12	☐	P68363	Tubulin alpha-1B chain OS=Homo sapiens GN=TUBA1B P...	37.89	135.00	103.98
13	☐	Q9P2J3	Kelch-like protein 9 OS=Homo sapiens GN=KLHL9 PE=1 S...		131.61	110.90
14	☐	P12236	ADP/ATP translocase 3 OS=Homo sapiens GN=SLC25A6 P...		131.31	78.50
15	☐	Q96M94	Kelch-like protein 15 OS=Homo sapiens GN=KLHL15 PE=...		130.71	35.01
16	☐	Q53G59	Kelch-like protein 12 OS=Homo sapiens GN=KLHL12 PE=...		127.87	23.09
17	☐	P49411	Elongation factor Tu, mitochondrial OS=Homo sapiens GN...		112.52	119.93
18	☐	Q9P2G9	Kelch-like protein 8 OS=Homo sapiens GN=KLHL8 PE=2 S...		110.71	28.11
19	☐	Q7L5N1	COP9 signalosome complex subunit 6 OS=Homo sapiens...		106.43	25.42





Ilektra Kouranti (HEGP)

Key questions in proteomics

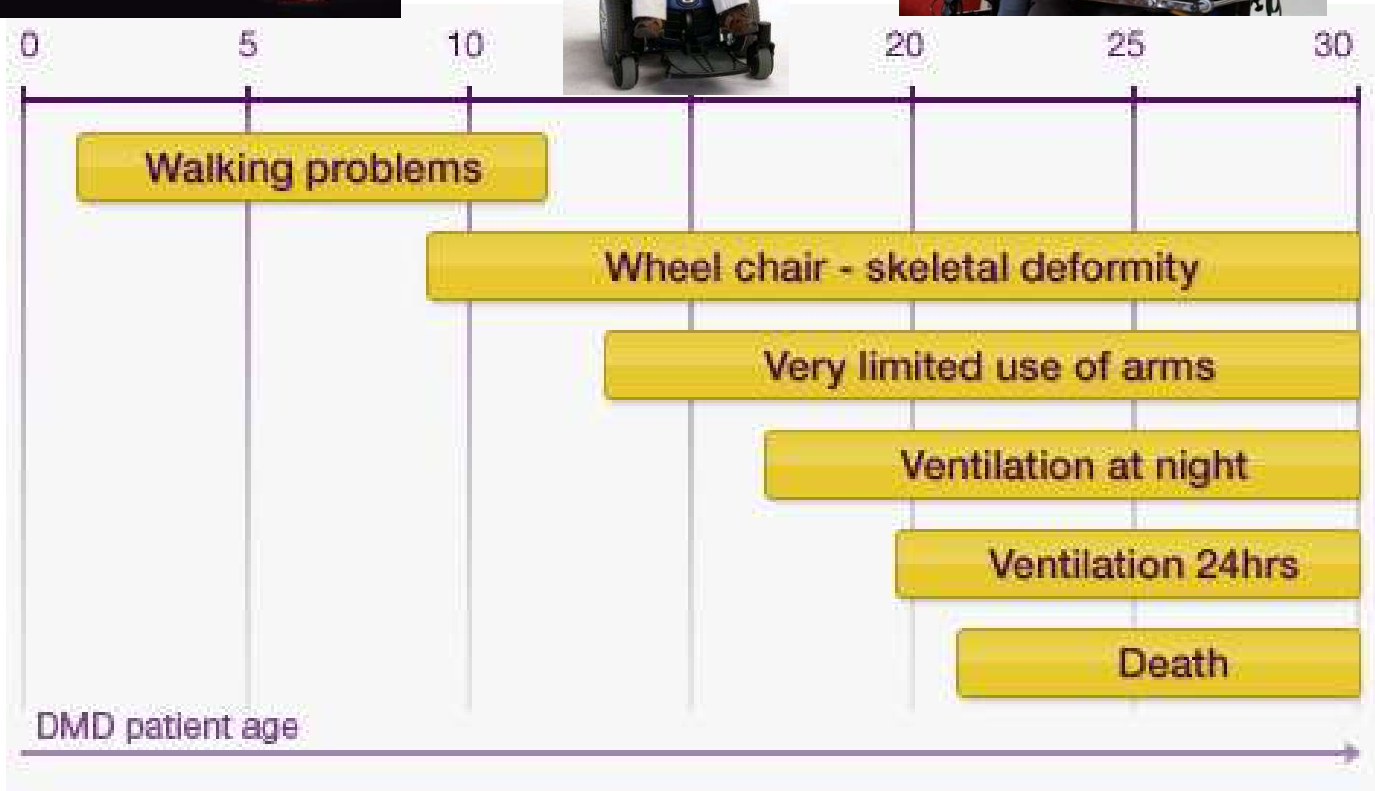
- What is the protein content of my biological sample?

  => problem of identification
- What is the abundance of my protein of interest?

  => quantification problem
- Relative question: What are the protein abundance variations of the proteomes studied?
- What are the partners of my protein of interest?
- Are there any signature proteins related to a particular biological process?

  => biomarkers identifications and quantifications

Biomarkers: applications to Duchenne dystrophy



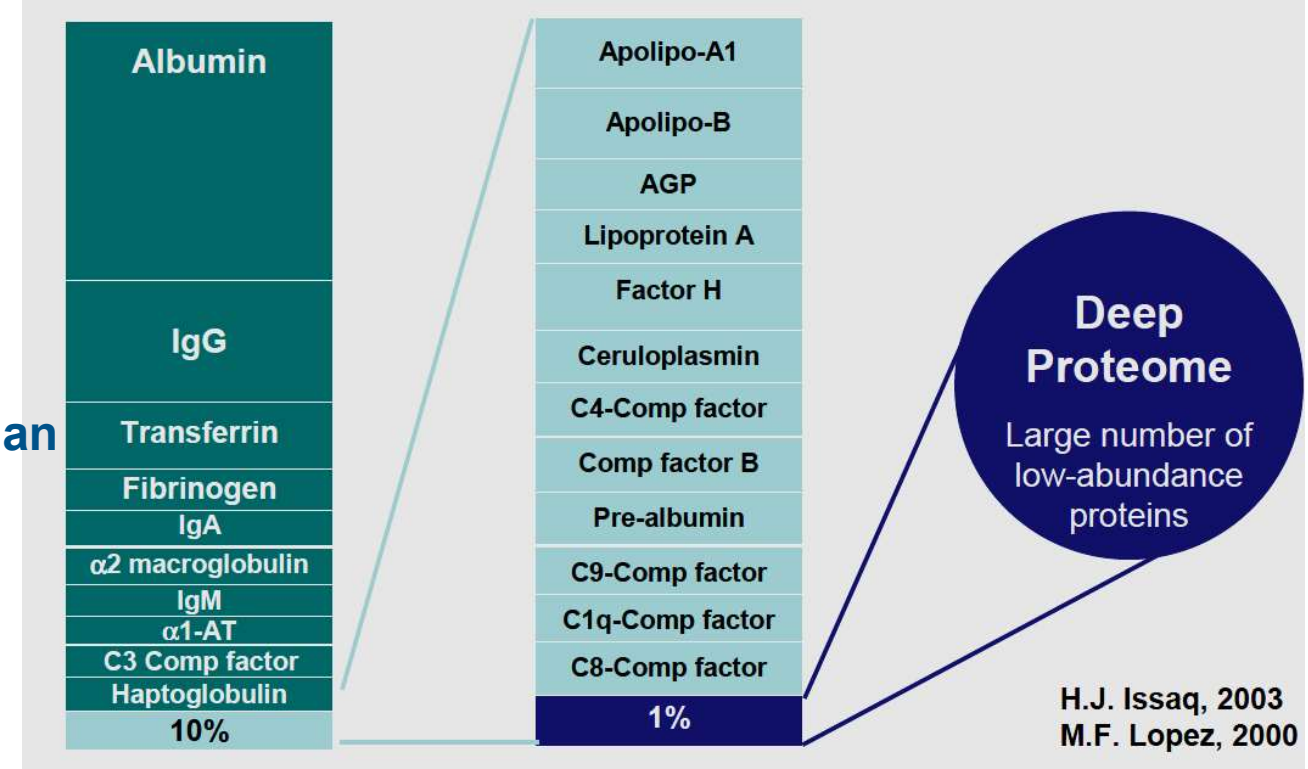
- **Death** of DMD patients usually occurs ~ 30's

Serum: a “tricky” fluid for Mass Spectrometry

- **Serum : Mixture of proteins with different ranges of proteins concentration (from mg/ml to pg/ml)**

- **99% of serum proteome = 20 major proteins**

- **1% remaining = more than one thousand proteins**



- Albumin : ~40 mg/ml (60% of serum proteome)

- C-reactive protein: ~1 µg/ml (40 000 times less than albumin)

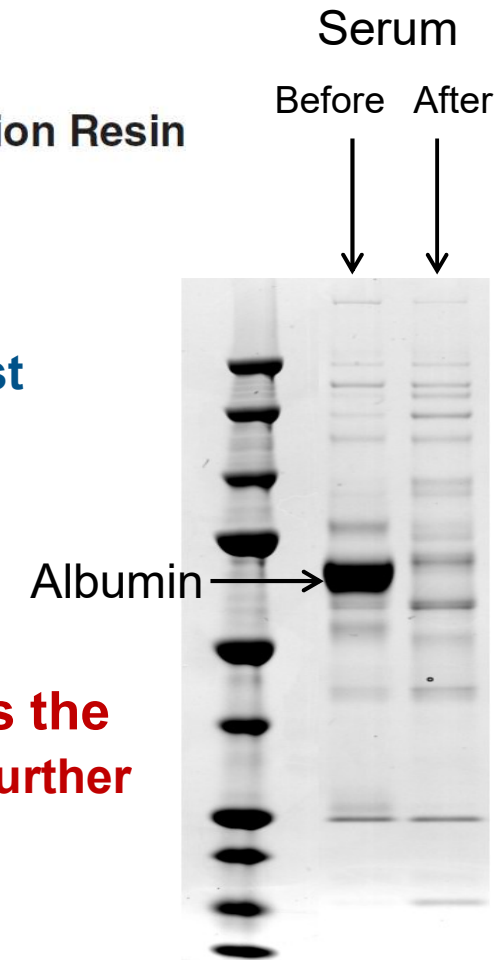
- FGF-9 : ~400 pg/ml (100 000 000 times less than albumin)

Serum: Depletion of high abundance proteins

Proteome Purify™ 12

R&D Systems

Human Serum Protein Immunodepletion Resin



- **Antibody based column raised against the 12 most abundant proteins in serum**
- **Reduction of albumin by > 90%**

Depletion of high abundance proteins gave us the highest number of identifications: selected for further analysis

Biomarkers: applications to Duchenne dystrophy

Description	No. of peptides	Score	ANOVA (*p*-value)	Fold change DMD/Healthy
Titin	23	1469.0	3.88E-06	37.4
Uromodulin	13	777.4	3.96E-03	5.5
Cubilin	10	576.8	2.55E-03	−2.3
Nuclear transport factor 2	5	356.9	1.05E-04	5.8
TNF-receptor superfamily member 16	4	308.7	4.07E-05	3.3
Myosin-1	3	265.3	8.66E-04	39.4
Fibulin-2	3	256.7	1.75E-03	2.9
β-galactosidase	6	253.7	1.23E-03	−2.4
Complement C1r subcomponent-like protein	5	235.8	3.47E-05	2.7
Aminopeptidase	3	213.7	2.71E-03	2.4

From: **"Proteomics profiling of urine reveals specific titin fragments as biomarkers of Duchenne muscular dystrophy", *J. Rouillon, A. Zocevic, T. Leger, C. Garcia, J-M. Camadro, B. Udd, B. Wong, L. Servais, T. Voit, F. Svinartchouk,* 2014 *Neuromuscular Disorders***

Biomarqueurs: applications à la dystrophie de Duchenne

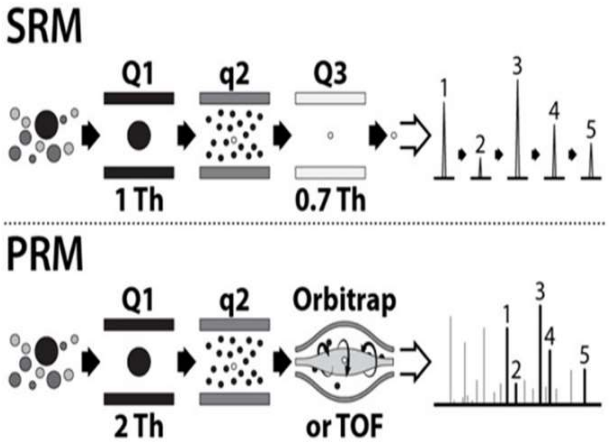
No. accession	Description	Localization	Peptides	Score	ANOVA (P-value)	Fold change
MYG_HUMAN	Myoglobin	Cytoplasm	4	195	2.7e-03	234.8
MYOM2_HUMAN	MYOM2	Myofibril	10	390	9.8e-05	100.1
MYOM3_HUMAN	**MYOM3**	**Myofibril**	**11**	**491**	**1.5e-05**	**49.7**
TPIS_HUMAN	Triosephosphate isomerase	Cytoplasm	3	128	2.3e-03	48.4
AATC_HUMAN	Aspartate aminotransferase	Cytoplasm	3	75	4.7e-04	45.7
KCRM_HUMAN	**CK-M**	**Cytoplasm**	**15**	**849**	**2.9e-05**	**39.8**
MYH7_HUMAN	**Myosin-7**	**Myofibril**	**11**	**520**	**2.2e-05**	**38.3**
ENOB_HUMAN	**β-enolase**	**Cytoplasm**	**4**	**178**	**7.4e-05**	**34.8**
G6PI_HUMAN	Glucose-6-phosphate isomerase	Cytoplasm/Secreted	4	130	1.6e-03	29.5
CAH3_HUMAN	**Carbonic anhydrase 3**	**Cytoplasm**	**5**	**182**	**8.6e-05**	**23.9**
FLNC_HUMAN	Filamin-C	Myofibril	4	145	4.3e-04	19.4
ALAT1_HUMAN	**Alanine aminotransferase 1**	**Cytoplasm**	**4**	**127**	**3.0e-05**	**15.6**
ALDOA_HUMAN	**Fructose-bisphosphate aldolase A**	**Cytoplasm**	**15**	**729**	**9.3e-05**	**14.2**
KPYM_HUMAN	**Pyruvate kinase PKM**	**Cytoplasm**	**16**	**845**	**1.1e-05**	**12.8**
TITIN_HUMAN	Titin	Myofibril	14	495	1.9e-03	10.8
VINC_HUMAN	**Vinculin**	**Cytoplasm/Membrane**	**2**	**74**	**7.2e-05**	**10.3**
PYGM_HUMAN	Glycogen phosphorylase, muscle form	Cytoplasm	8	257	6.1e-04	9.9
LDHA_HUMAN	L-lactate dehydrogenase A chain	Cytoplasm	8	378	9.1e-04	9.5
HPT_HUMAN	Haptoglobin	Secreted	29	1867	1.5e-04	7.6
HBD_HUMAN	Haemoglobin subunit δ	Cytoplasm	3	100	5.1e-03	6.2
LDHB_HUMAN	**L-lactate dehydrogenase B**	**Cytoplasm**	**10**	**598**	**2.4e-05**	**5.4**
HBB_HUMAN	Haemoglobin subunit β	Cytoplasm	7	552	8.0e-03	3.6
HBA_HUMAN	Haemoglobin subunit α	Cytoplasm	7	407	5.3e-03	3.4
TPM2_HUMAN	Tropomyosin βchain	Myofibril	5	170	2.0e-02	2.6

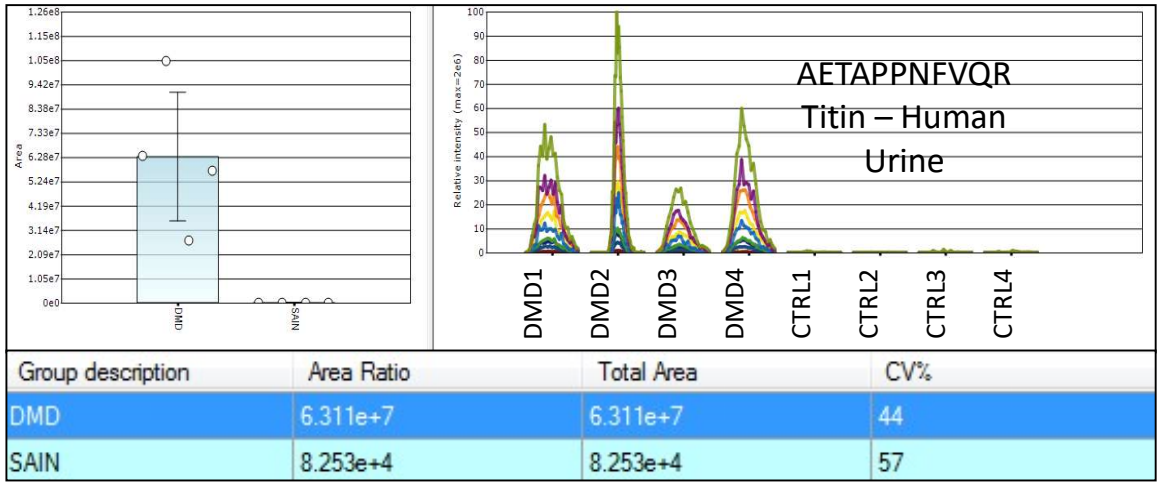
GENETHON
INNOVER POUR GUERIR

Rouillon, J., Zocevic, A., Poupiot, J., Amor, F., **Léger, T., Garcia, C., Camadro, J.M.**, Wong, B., Cosette, J., ML Coenen-Stass, A., McClorey, G., C Roberts, T., JA Wood, M., Servais, L., Voit, T., Richard, I., Svinartchouk, F. (2015). *Serum proteomic profiling reveals specific MYOM3 fragments as biomarkers of Duchenne muscular dystrophy with applications for the follow-up of gene therapy treatment in a mouse model of muscular dystrophies.* – Human Mol. Genetics

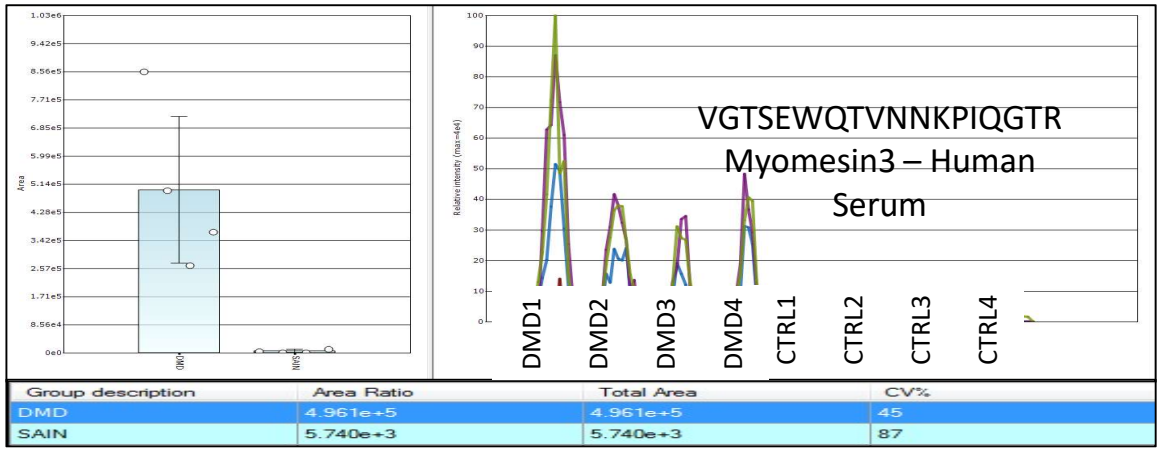
Jeremy Rouillon; Aleksandar Zocevic; **Thibaut Léger**; **Camille Garcia; Jean-Michel Camadro**; Bjarne Udd; Laurent Servais; Thomas Voit; Fedor Svinartchouk. (2014). *Proteomics profiling of urine reveals specific titin fragments as biomarkers of Duchenne muscular dystrophy.* Neuromuscular disorders.

Protéomique ciblée de type PRM





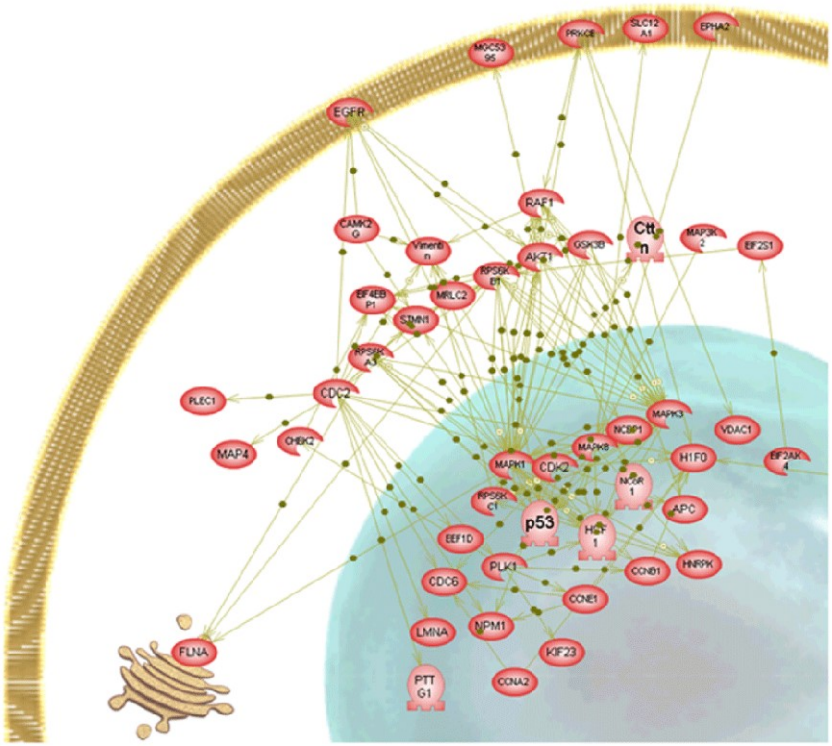
Group description	Area Ratio	Total Area	CV%
DMD	6.311e+7	6.311e+7	44
SAIN	8.253e+4	8.253e+4	57



Group description	Area Ratio	Total Area	CV%
DMD	4.961e+5	4.961e+5	45
SAIN	5.740e+3	5.740e+3	87

Why study PTMS?

- Cells can rapidly respond to stimuli and perturbations
- Important cellular mechanisms are tightly controlled
- Often, diseases (e.g. cancer) are due to aberrantly activated proteins
  - Protein expression is much too slow for quick adaption
  - PTMs are crucial regulator
  - MS-based proteomics allows to analyze complex networks of post-translationally modified proteins



http://www.omicsonline.org/ArchiveJPB/2010/April/03/JPB-03-113.php

PTMs *in vivo*

- **Phosphorylation** (Ser, Thr, Tyr; +80 Da)
  - Phosphorylation is one of the most important PTMs
  - A key event in signaling
  - Catalyzed by kinases/phosphatases
- **Glycosylation** (Asn, Ser, Thr)
  - marks proteins for degradation
  - s for degradation
- **Glycation** (Asn, Ser, Thr)
  - marks proteins for degradation
  - s for degradation
- **Ubiquitination** (Lys; +114 Da)
  - marks proteins for degradation
- **Proteolytic cleavage**
- **Acetylation** (N-termini and Lys +42 Da)
  - often combined with removal of protein initial Met

Others: oxidations, methylations, sumoylations, glutathionylations…

PTMs characterization: techniques

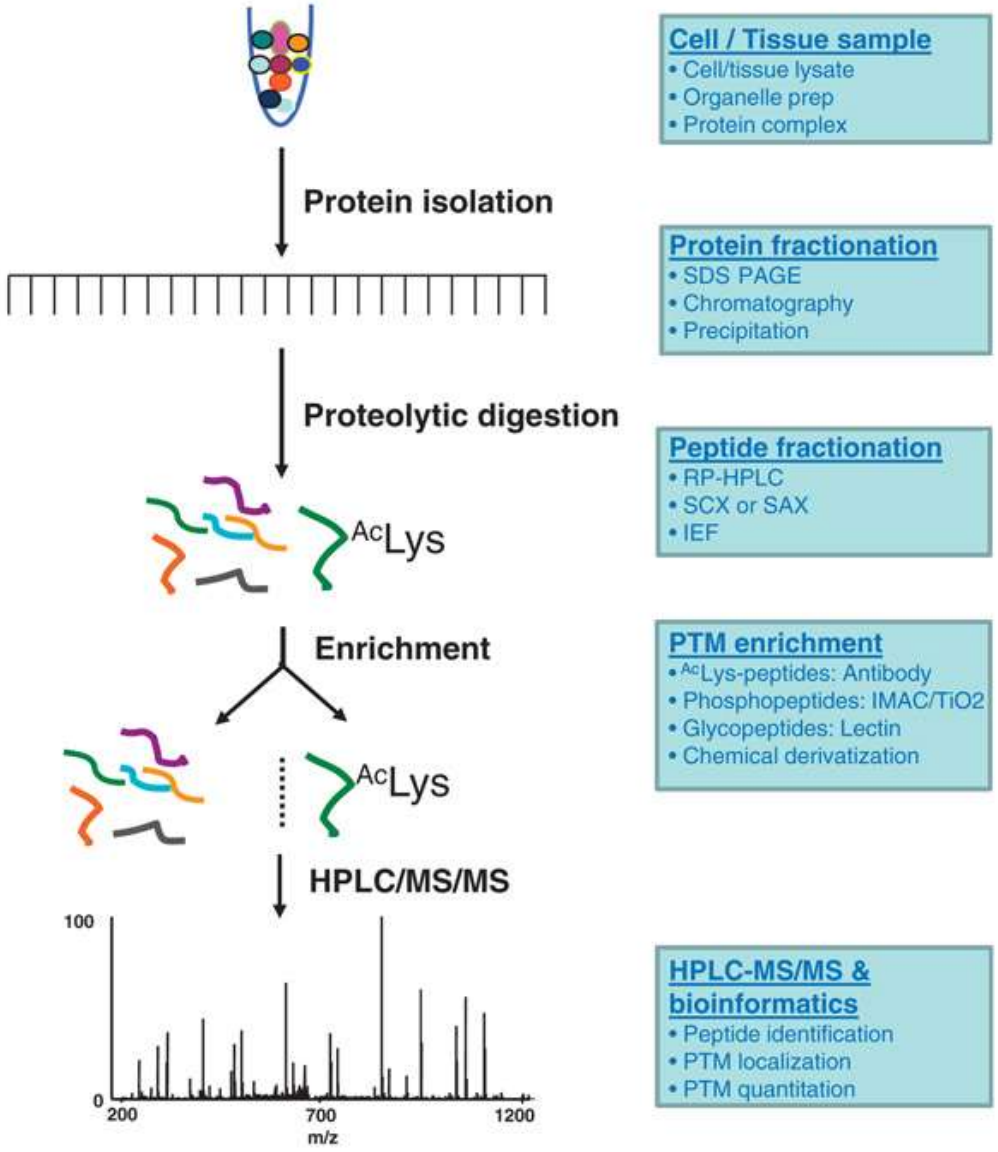
Techniques for detection and identification of PTM substrates

Method	*In vitro* or *in vivo*	Case studies	Advantages	Disadvantages
Radioactive isotope labeling	*In vitro* or *in vivo*	^{32}P (pSer, pThr, pTyr) ^{3}H, or ^{14}C for ^{Ac}Lys or ^{Me}K	Reagents accessible	Inconvenience/hazard low sensitivity
Western blotting	*In vitro* or *in vivo*	pTyr, ^{Ac}Lys or ^{Me}K	Good affinity	Moderate sensitivity
Peptide/protein array	*In vitro*	pSer/Thr/Tyr, ^{Ac}Lys or ^{Me}K	Rapid, global scale	Possibly non-specific, low sensitivity, requires verification
MS-proteomics	*In vitro*	pSer/Thr/Tyr, ^{Ac}R or ^{Me}K	Specific, global scale	Need enrichment methods

$^{Ac}_{R}$, $^{Me}_{K}$, pSer, pThr, and pTyr, represent acetylated arginine, methyllysine, phosphorylated serine, threonine, and tyrosine residues, respectively.

Proteomics. 2009 Oct; 9(20): 4632–4641.

Workflow for PTMs characterization



Cell / Tissue sample
- Cell/tissue lysate
- Organelle prep
- Protein complex

Protein fractionation
- SDS PAGE
- Chromatography
- Precipitation

Peptide fractionation
- RP-HPLC
- SCX or SAX
- IEF

PTM enrichment
- ^{Ac}Lys-peptides: Antibody
- Phosphopeptides: IMAC/TiO2
- Glycopeptides: Lectin
- Chemical derivatization

HPLC-MS/MS & bioinformatics
- Peptide identification
- PTM localization
- PTM quantitation

Proteomics. 2009 Oct; 9(20): 4632–4641.

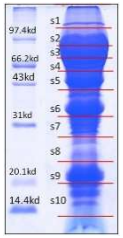
Quantitative proteomics and phosphorylations

CONDITION 1:
Mast cells inactivated
CO-IP PY20

Immunoaffinity

1 Hour

SDS-PAGE

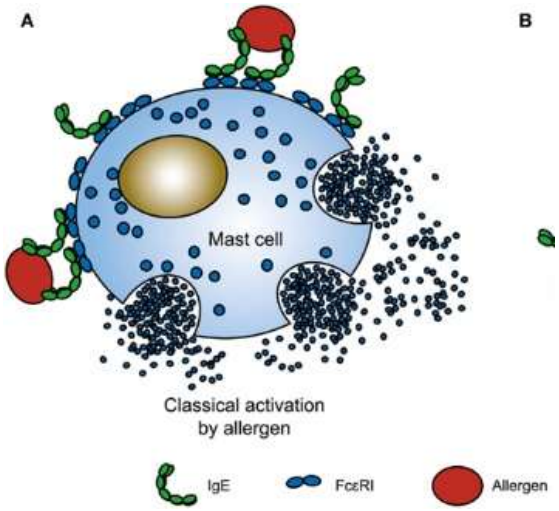


Trypsin digestion overnight
(phosphate buffer)

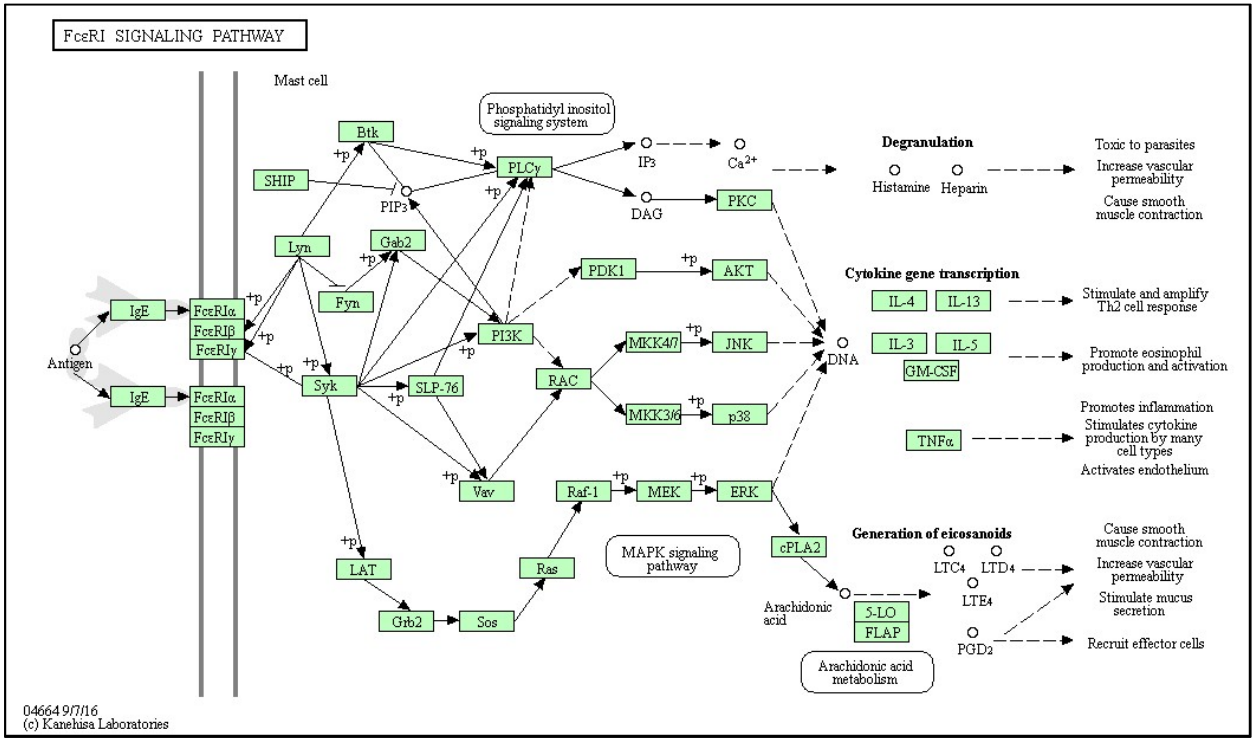
LC-MS/MS

Differential phosphorylated peptide and protein identification

CONDITION 1:
Mast cells activated (BC4)
CO-IP PY20

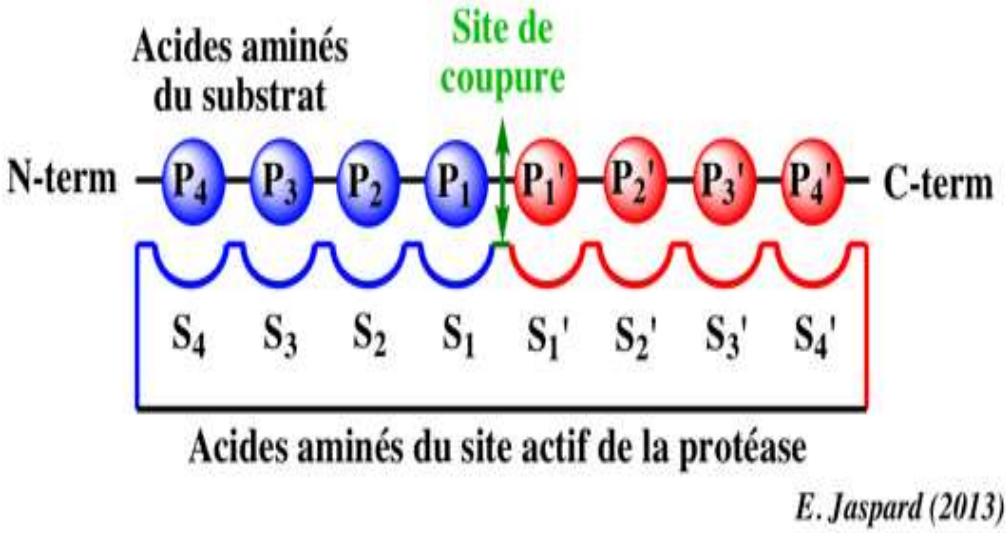


935	Q9JM80	Phosphoprotein associated with glycosphingolipid-enriched microdomains 1		membrane			763.10	646.62	45.9	4.81					
953	P20411	High affinity immunoglobulin epsilon receptor subunit gamma	Signal transducer activity	cell surface; membrane	response to stimulus; transport	Pf11628	573.65	352.21	9.8	6.00					
963	Q64725	Tyrosine-protein kinase SYK OS=Rattus norvegicus GN=Syk PE=1 SV=1	Signal transducer activity	nucleus; organelle lumen	response to stimulus; transport	Pf00017; Pf00069; Pf07714	835.28	323.20	71.5	8.15					
964		Sequence	# PSMs	# Proteins	# Protein Groups	Protein Group Accessions	Modifications	MH+ [Da]	phosphoRS Site Probabilities	A4	IonScore A4	Exp Value A4	B4	IonScore B4	
965		LLTLEDNELGSGNFGTVK	2	1	1	Q64725		1906.97726		High	107	1.50125E-10			
966		DESEQTVLIGSK	4	1	1	Q64725		1305.65435		High	84	2.49757E-08	High	67	
967		ELNGTYAISGGR	7	1	1	Q64725		1237.61919		High	55	2.38967E-05	High	58	
968		ADENYYK	3	1	1	Q64725		902.38894		High	46	0.000100972	High	23	
969		NVLLVTQHYAK	8	1	1	Q64725		1285.72683		High	44	0.000137239	High	31	
970		ISDFGLSK	6	1	1	Q64725		866.46208		High	42	0.000234007	High	40	
971		LIATTAHEK	7	1	1	Q64725		983.55124		High	37	0.000403124	High	37	
972		LRNYYYDVVN	4	1	1	Q64725		1318.64326		High	37	0.001486192	Medium	24	
973		YLEESNFVHR	4	1	1	Q64725		1293.62261		High	36	0.00176677	High	27	
974		MGCPPGCPR	4	1	1	Q64725	M1(Oxidation); C3(Carbamidon	1047.41679		High	35	0.000734469	High	41	
975		GSEVTAMLEK	2	1	1	Q64725		1064.52982		High	34	0.002136328			
976		GSEVTAMLEK	4	1	1	Q64725	M7(Oxidation)	1080.52370		High	34	0.002154235	Medium	20	
977		EVYLDRK	2	1	1	Q64725		922.49926		High	33	0.001493862			
978		ALRADENYYK	2	1	1	Q64725		1242.61182		High	30	0.007003353			
979		VLTVPCQK	5	1	1	Q64725	C6(Carbamidomethyl)	944.52278		Medium	25	0.020099256			
980		GSEVTAMLEKGER	1	1	1	Q64725	M7(Oxidation)	1422.68938		Medium	25	0.027162273			
981		LRNYYYDVVN	6	1	1	Q64725	Y4(Phospho)	1398.61077	Y(4): 93.9; Y(5): 5.7; Y(6): 0.4	Medium	24	0.023646873			
982		TGPFEDLKENLIR	2	1	1	Q64725		1531.81073		Medium	22	0.025122601			
983		GSEVTAMLEKGER	2	1	1	Q64725		1406.69446		Medium	20	0.067444564			
984		KPFNRPPGVQPK	1	1	1	Q64725		1364.77961		Medium	19	0.035807446			
985		LLTLEDNELGSGNFGTVKK	1	1	1	Q64725		2035.07124		Low	19	0.093821803			
986		WYAPECINYFK	1	1	1	Q64725	C6(Carbamidomethyl)	1490.67815		Low	17	0.171057347			
987		NYYYDVVN	2	1	1	Q64725		1049.45671		Low	16	0.121475657			
988		MGCPPGCPR	1	1	1	Q64725	C3(Carbamidomethyl); C7(Carl	1031.42131		Low	15	0.108098357			
989		MPWFHGNISR	1	1	1	Q64725	M1(Oxidation)	1260.59497		Low	13	0.348397318			
990		YLQQNR	1	1	1	Q64725		821.42674		Low	12	0.268819476			
991		NYLGGFALSVAHNR	1	1	1	Q64725	Y2(Phospho)	1598.76055	Y(2): 99.9; S(9): 0.1				Low	11	
992	Q5U2U2	Crk-like protein OS=Rattus norvegicus GN=Crkl PE=1 SV=1	protein binding	m; cytosol; membrane	development	Pf00017; Pf00018; Pf07653	234.55	96.32	33.8	6.74					
1005	P60868	40S ribosomal protein S20 OS=Rattus norvegicus GN=Rps20	tural molecule activity	ism; cytosol; ribosome	metabolic process	Pf00338	172.28	36.35	13.4	9.94					

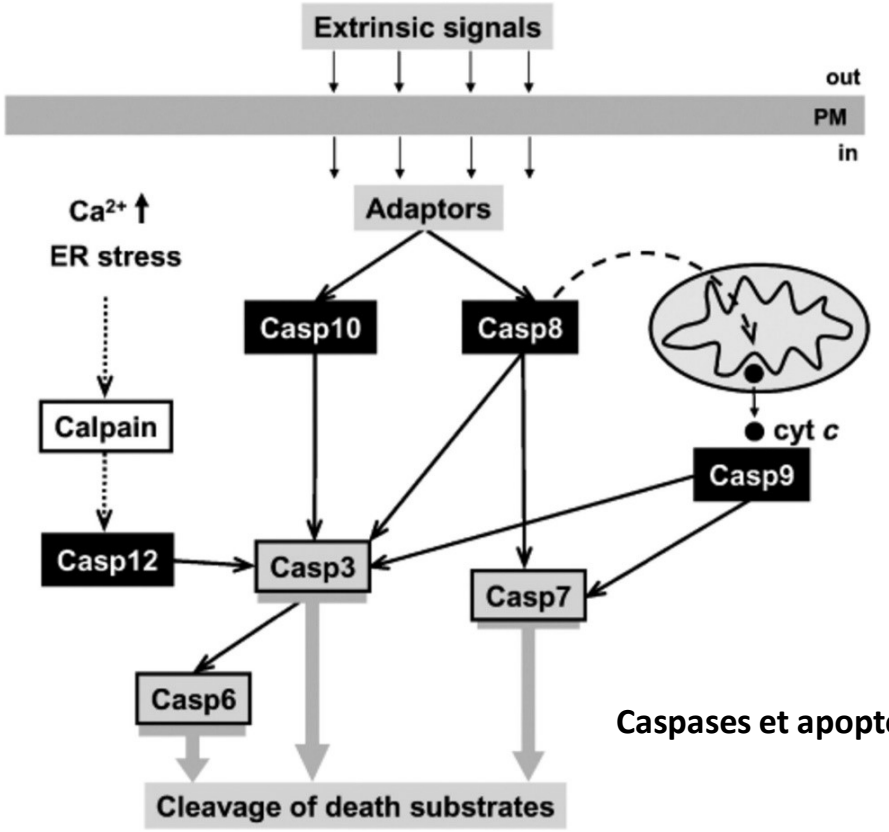


Proteolytic cleavages as PTMs

- Enzymes hydrolysant des liaisons entre acides aminés
- Classification des protéases : aspartate-, cystéine-, glutamate-, métallo-, sérine-, thréonine-, et les asparagine- protéases)
- Autres classifications



Biological functions of proteases

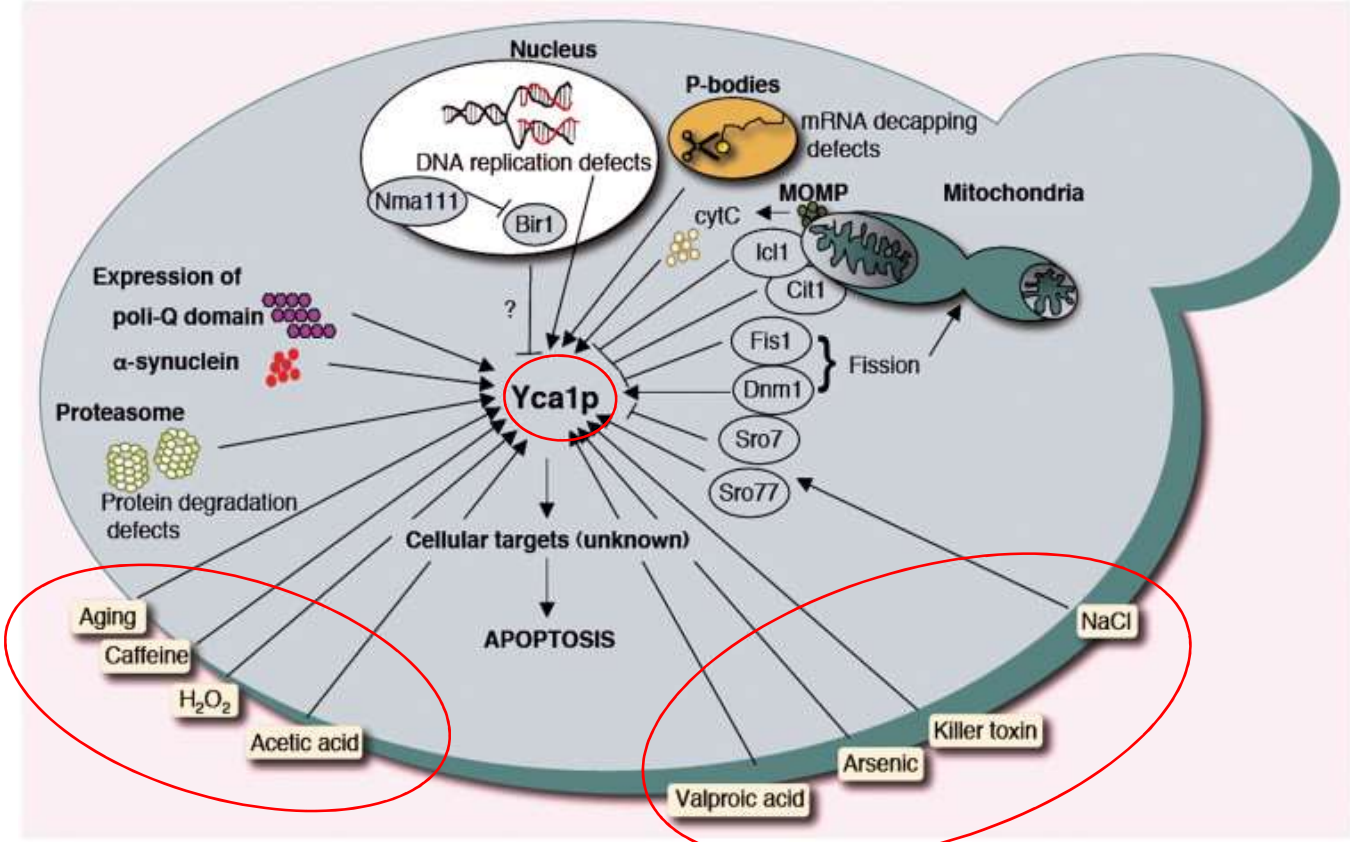


Caspases et apoptose

Vachova *et al.* 2007

- **Various proteases**
- **Biological process regulations**
  - Protein turnover
  - Misfolded proteins degradation
  - Cell addressing
  - Protein activation
- **Deregulations associated to pathologic states**
- **Model organism to study proteases**

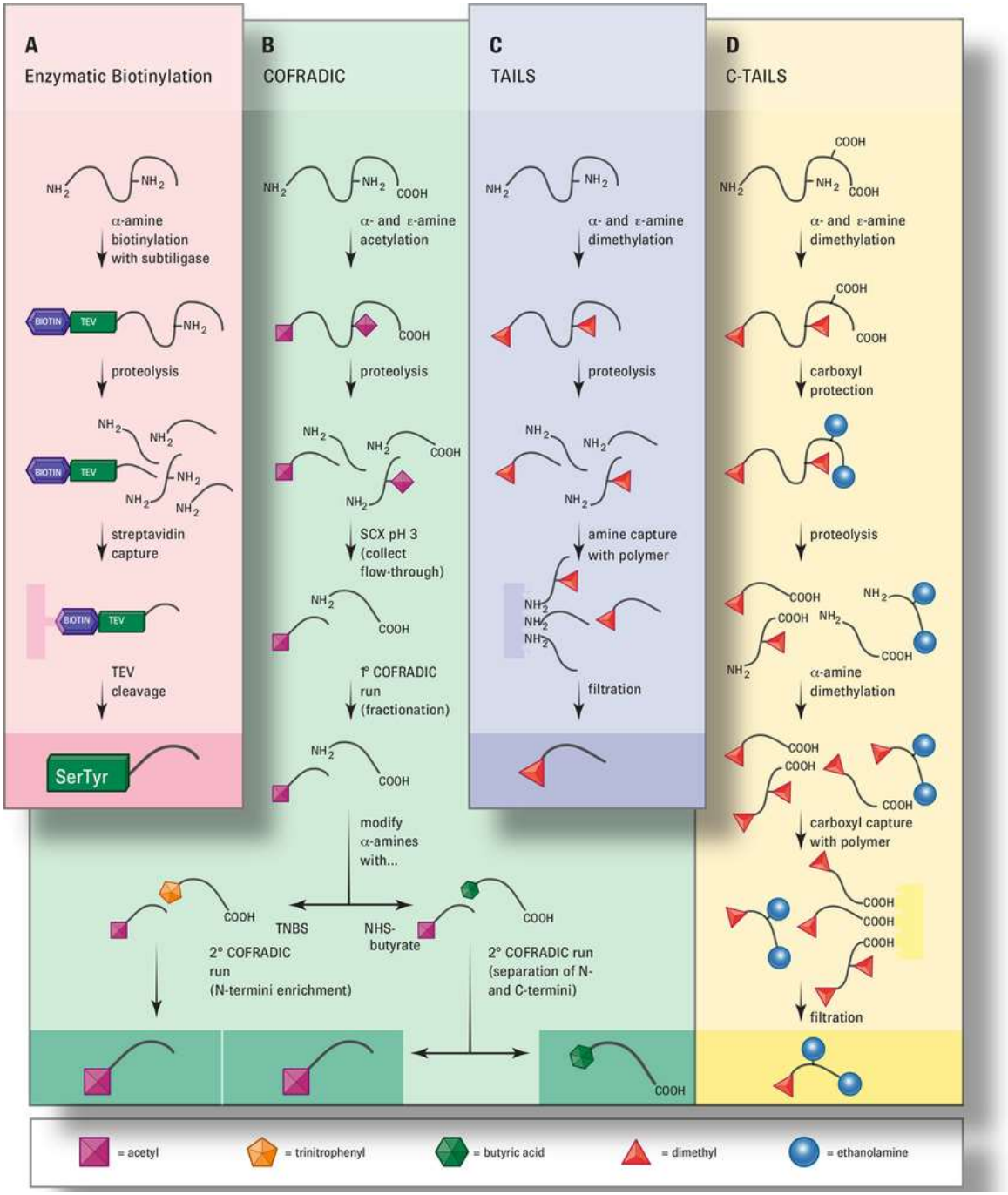
Mca1p activation and apoptosis release



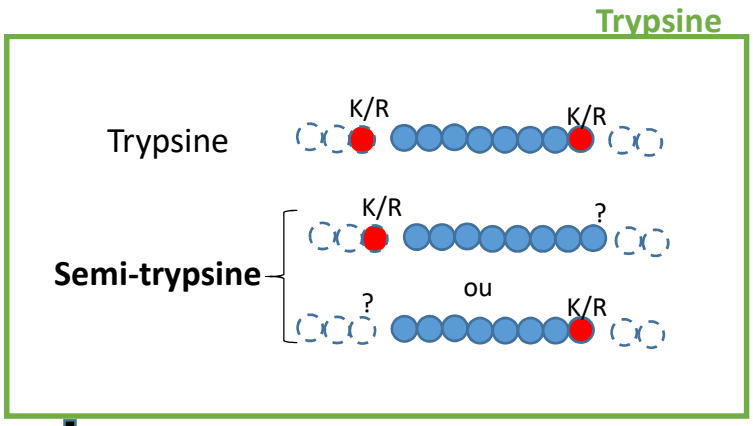
Mazzoni *et al.* 2008

- **Activée par de nombreuses molécules dont la molécule de quorum-sensing farnésol**
- Pas d'informations sur la spécificité de clivage de ces substrats (coupures suspectées au niveau des résidus K et R)
- Un seul substrat caractérisé *in vitro* pour la métacaspase de *S. cerevisiae* (Gapdhp).

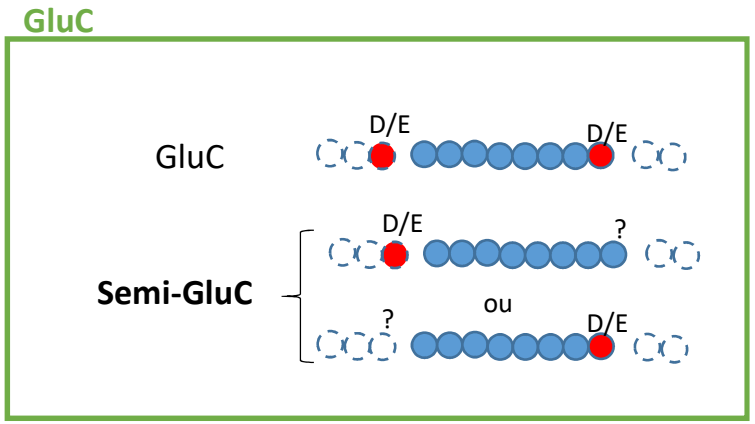
Terminomics



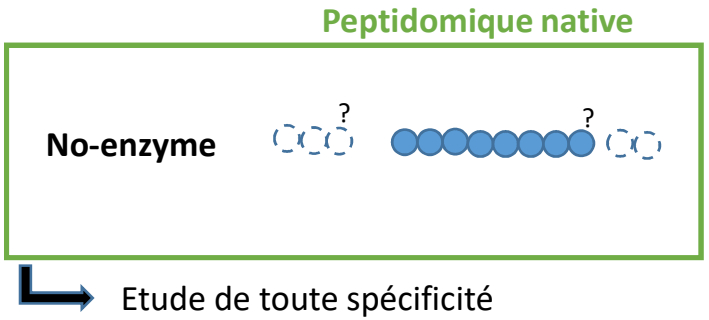
Recherche de la spécificité de clivage de Mca1p



Etude de spécificité autre que K ou R



Etude de spécificité autre que D ou E



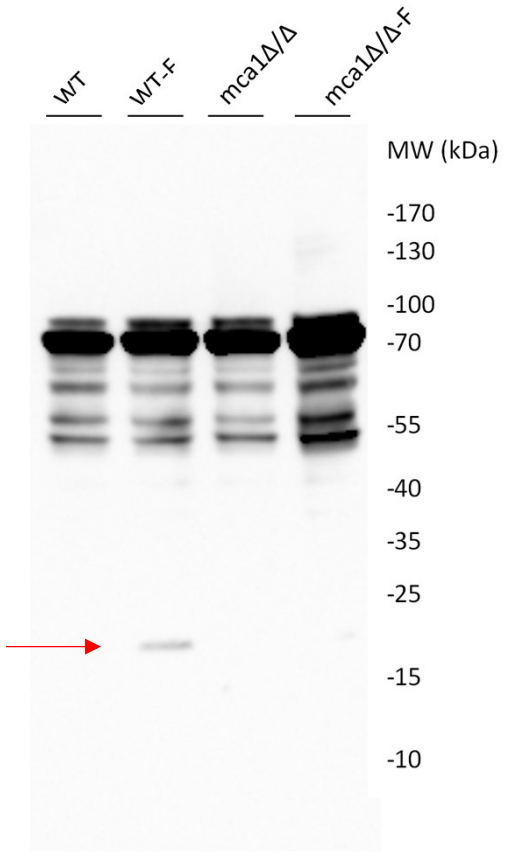
Etude de toute spécificité

Jeux de donnée

Acide aminé

Substrats potentiels de Mca1p et protéines de réponse aux stress

Accession	Description	Experiment	Sequences
orf19.778	PIL1, composant de l'éisosome	Native	WGEDNEDDISDVTDK
orf19.7350	RCT1 Protéine induite par le fluconazole	Native	YDPKRSSNQGSSSSNDEQQDR
Orf19.4309	GRP2, Methylglyoxal réductase	GluC	K.EKPNFTLSVINPVYVFGPQAFE
Orf19.2340	CDC48; ATPase microsomale	GluC	R.FALGNSNPSALRE R.GQFSSFRFNE
orf19.2483	RIM1; protéine liant l'AND simple brin	GluC	K.VGSLVHVD
orf19.2644	QCR2; Ubiquinol-cytochrome-c réductase	GluC	R.GLGNPLFYNE
orf19.1435	TEF1; Facteur d'élongation	GluC	HALLAYTLGVK K.SGKVTGKTLLE
orf19.6515	HSP90; Protéine chaperon essentielle	GluC	K.LVDAPAAIRTGQFGWSANME
Orf19.6367	SSB1; Protéine de choc thermique (HSP70)	GluC	R.LIGRAFDDE
orf19.4980	HSP70; Protéine chaperon de famille HSP70	GluC	K.LVSDFFNGKE K.RTLSSSAQTSIE
orf19.1065	SSA2; Protéine chaperon de famille HSP70	GluC	K.RTLSSSAQTSIE R.LIGDAAKNQAAMNPANTVFD
Orf19.5928	RPP2B; protéine ribosomale acide	GluC	R.LQALLKDLE



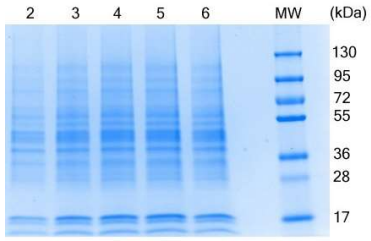
77 substrats potentiels de Mca1p (pour 62 protéines), dont 13 validés dans des conditions de sélection les plus drastiques

Proteomics workflows

Top – Down

(< 50000 Da in particular conditions)

Proteins



Peptides

HPLC

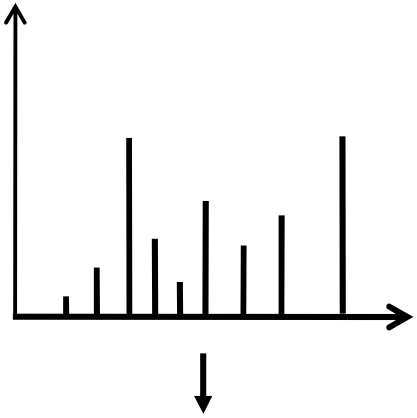
PROTEASES

Bottom – Up

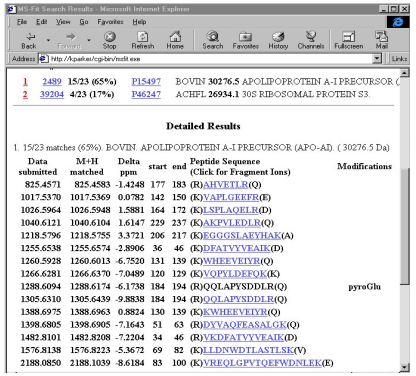
(< 4000 Da)

MS acquisition

- MALDI TOF/TOF
- Orbitrap



Identification
Quantification



1 2489 15/23 (65%) P15497 BOVIN 30276.5 APOLIPOPROTEIN A-I PRECURSOR (
2 39204 4/23 (17%) P46247 ACHFL 26934.1 30S RIBOSOMAL PROTEIN S3.

Detailed Results

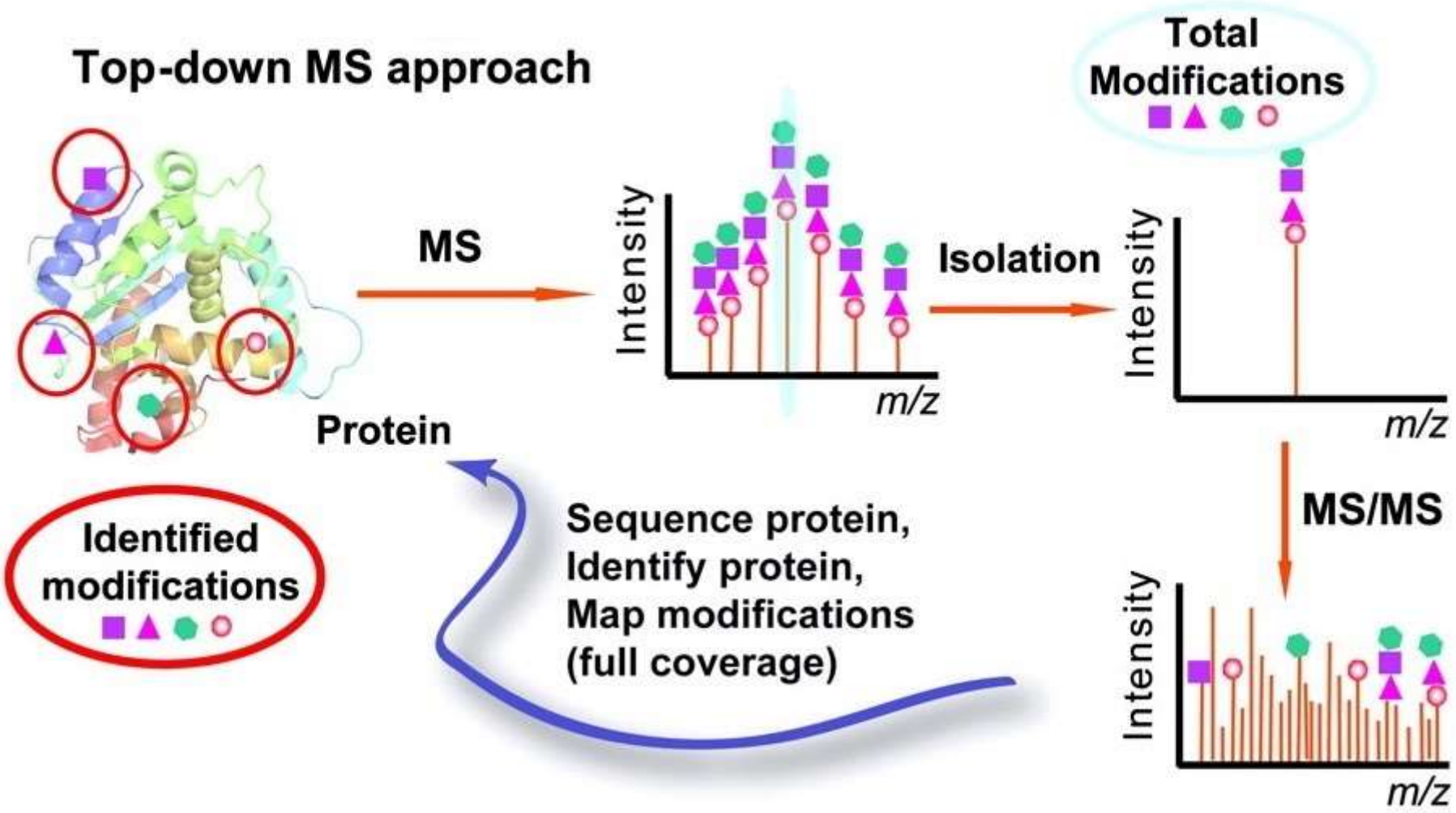
1. 15/23 matches (65%). BOVIN. APOLIPOPROTEIN A-I PRECURSOR (APO-AI). (30276.5 Da)

Data submitted	M+H matched	Delta ppm	start	end	Peptide Sequence (Click for Fragment Ions)	Modifications
825.4571	825.4583	-1.4248	177	183	(R)AHVETLR(Q)	
1017.5370	1017.5369	0.0782	142	150	(K)VAPLGEEFR(E)	
1026.5964	1026.5948	1.5881	164	172	(K)LSPLAQELR(D)	
1040.6121	1040.6104	1.6147	229	237	(K)AKPVLEDLR(Q)	
1218.5796	1218.5755	3.3721	206	217	(K)EGGGSLAEYHAK(A)	
1255.6538	1255.6574	-2.8906	36	46	(K)DFATVYVEAIK(D)	
1260.5928	1260.6013	-6.7520	131	139	(K)WHEEVEIYR(Q)	
1266.6281	1266.6370	-7.0489	120	129	(K)VQPVLDEFQK(K)	
1288.6094	1288.6174	-6.1738	184	194	(R)QQLAPYSDDLR(Q)	pyroGlu
1305.6310	1305.6439	-9.8838	184	194	(R)QQLAPYSDDLR(Q)	
1388.6975	1388.6963	0.8824	130	139	(K)WHEEVEIYR(Q)	
1398.6805	1398.6905	-7.1643	51	63	(R)DYVAQFEASALGK(Q)	
1482.8101	1482.8208	-7.2204	34	46	(R)VKDFATVYVEAIK(D)	
1576.8138	1576.8223	-5.3672	69	82	(K)LLDNWDTLASTLSK(V)	
2188.0850	2188.1039	-8.6184	83	100	(K)VREQLGPVTQEFWDNLEK(E)	

The strategy is dependent of the sample complexity

Search engine (Mascot, Sequest, Peaks, Maxquant, OMSSA, Prosight, Byonik)

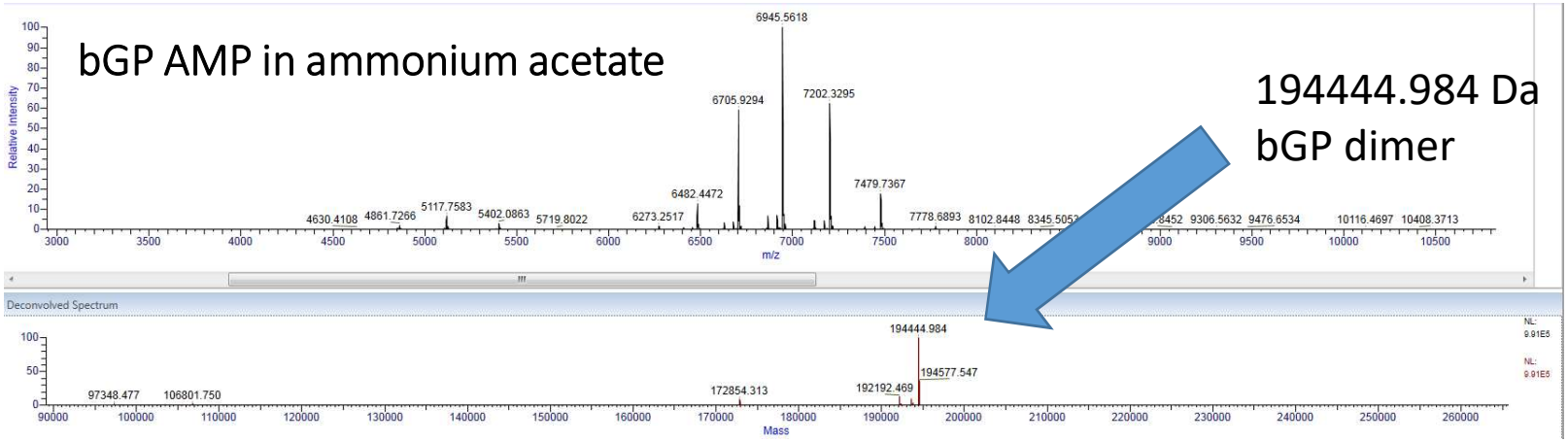
TOP DOWN proteomics for PTMs characterization

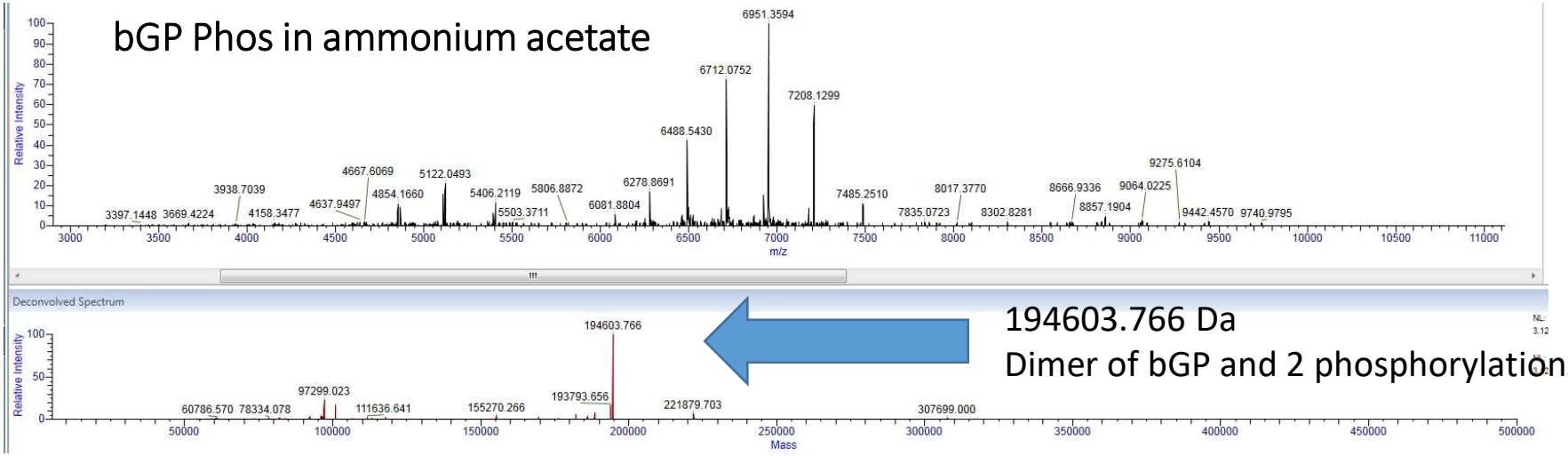


Challenges in TOP-DOWN proteomics

Challenges	Innovations
1. Protein solubility Conventional surfactant (e.g. SDS) not compatible with MS	**Develop new top-down MS compatible surfactant**
2. Proteome complexity Intact protein chromatography under-developed	**Develop novel multi-dimensional chromatography for intact protein separation**
3. Proteome dynamic range Difficulty in detecting low abundant proteins	**Develop novel nanomaterials for enriching low abundant proteins**
4. Protein MS data interpretation Software for top-down proteomics underdeveloped	**Develop user-friendly and versatile software interface**

Analysis in intact protein mode: human brain glycogen phosphorylase

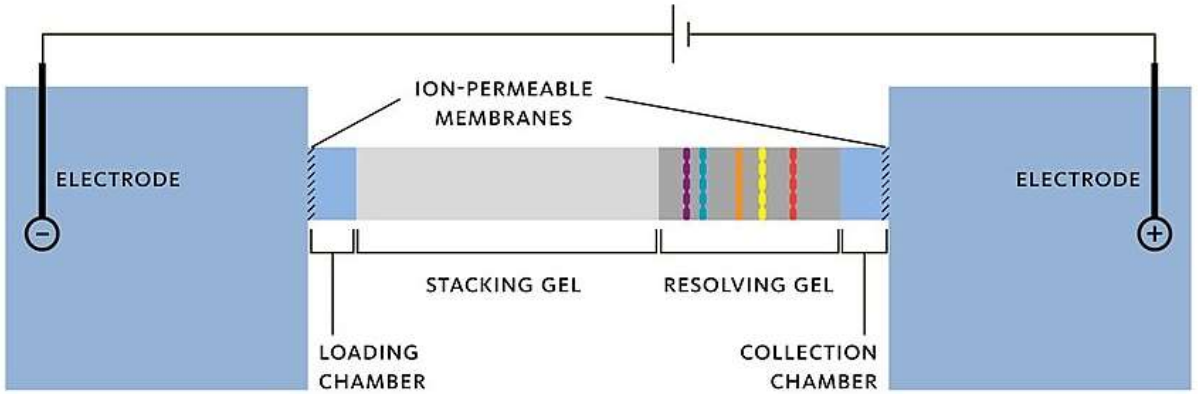




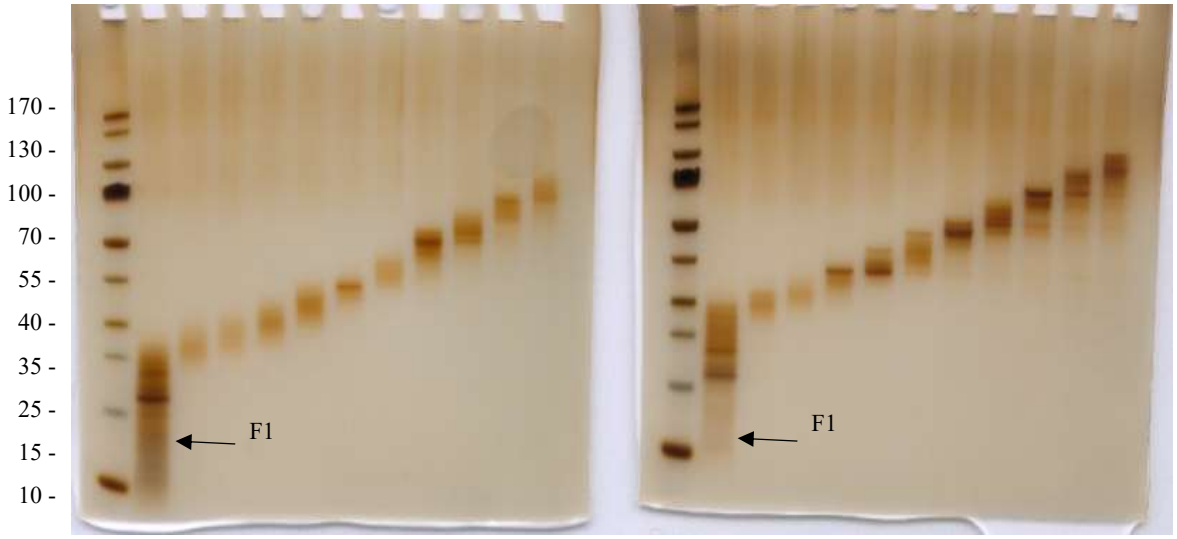
Crystal structure of human brain glycogen phosphorylase. Cécile Mathieu, Ines de la Sierra-Gallay, Romain Duval, Ximing Xu, Angélique Cocaign, Thibault Léger, Jean-Michel Camadro, Catherine Etchebest, Ahmed Haouz, Jean-Marie Dupret, Fernando Rodrigues-Lima. Under review.

Analysis in intact protein mode

Système Gelfree (Expedeon)

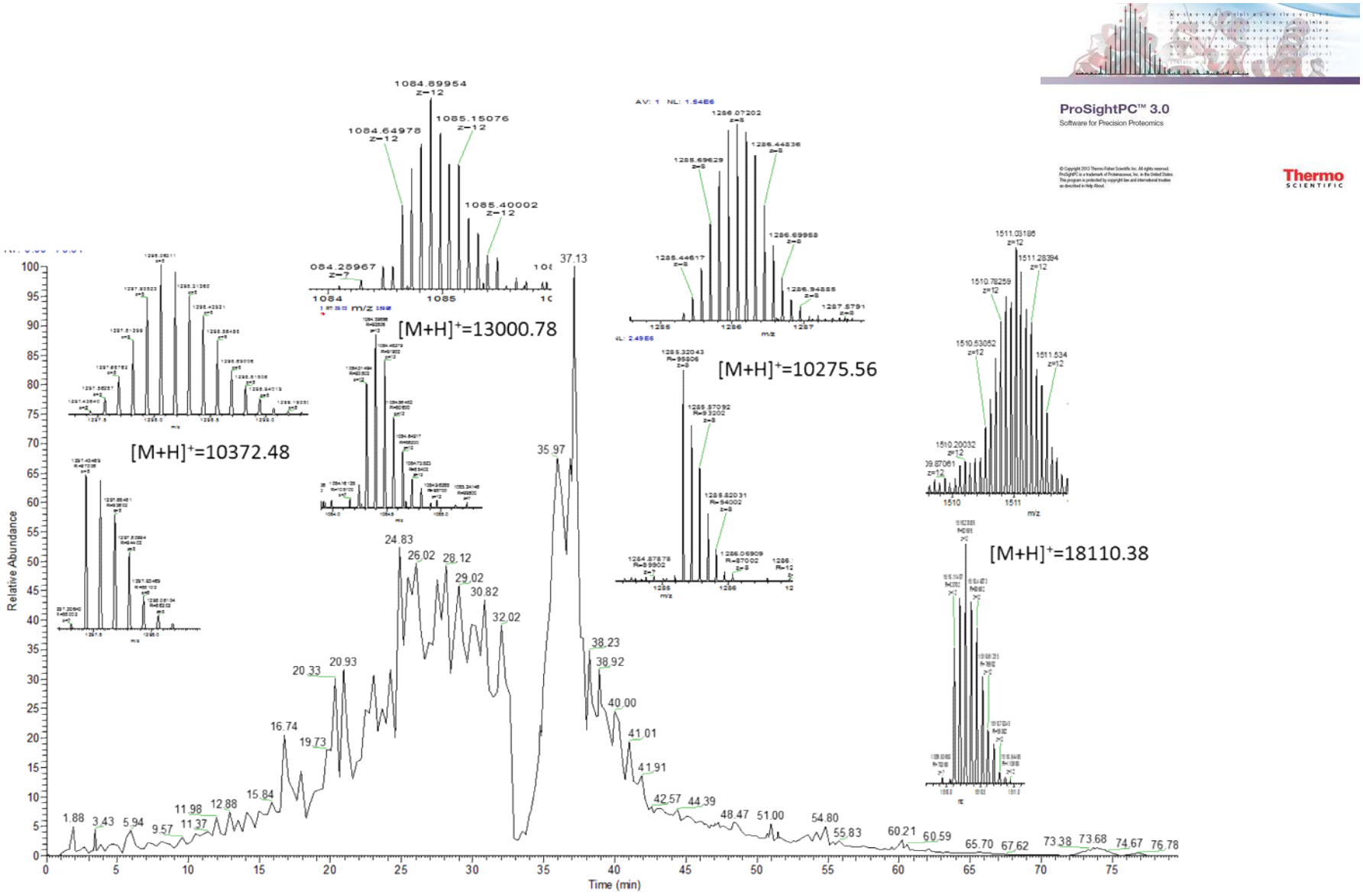


CTRL SDS-PAGE

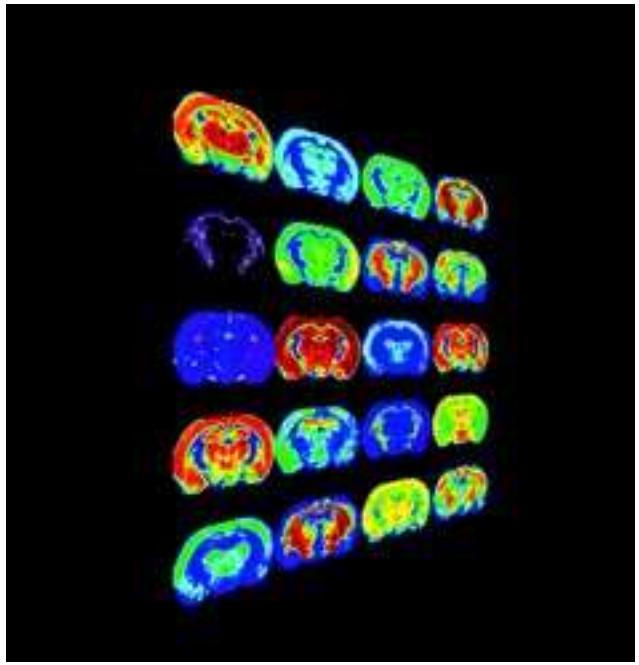


NanoLC-MS/MS
(Colonne pepswift)

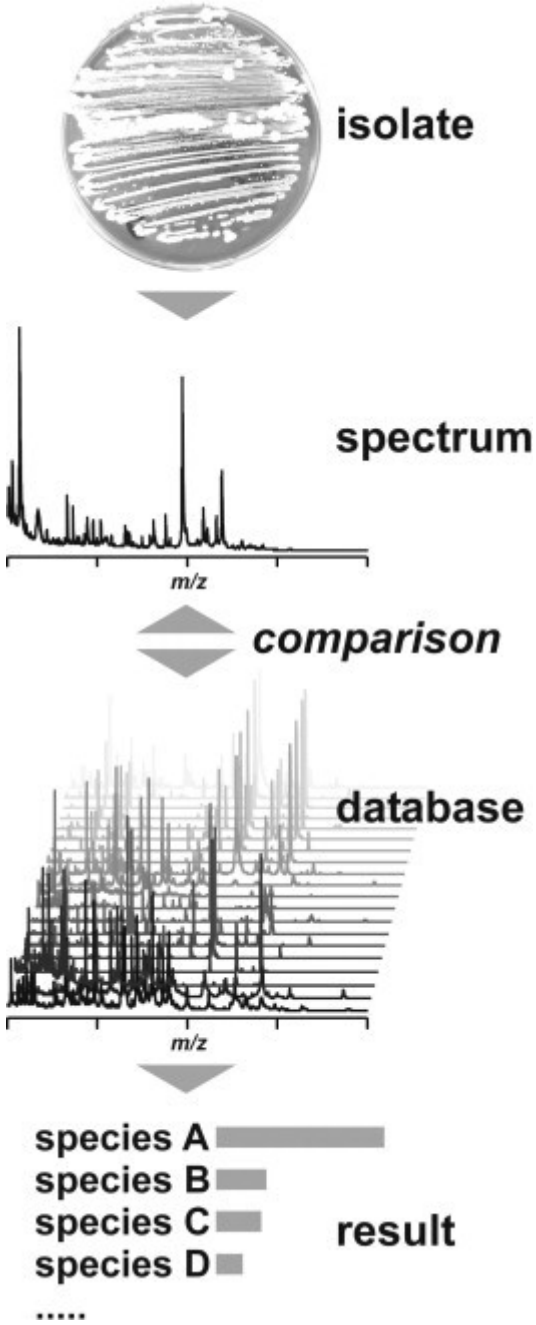
Analysis in intact protein mode

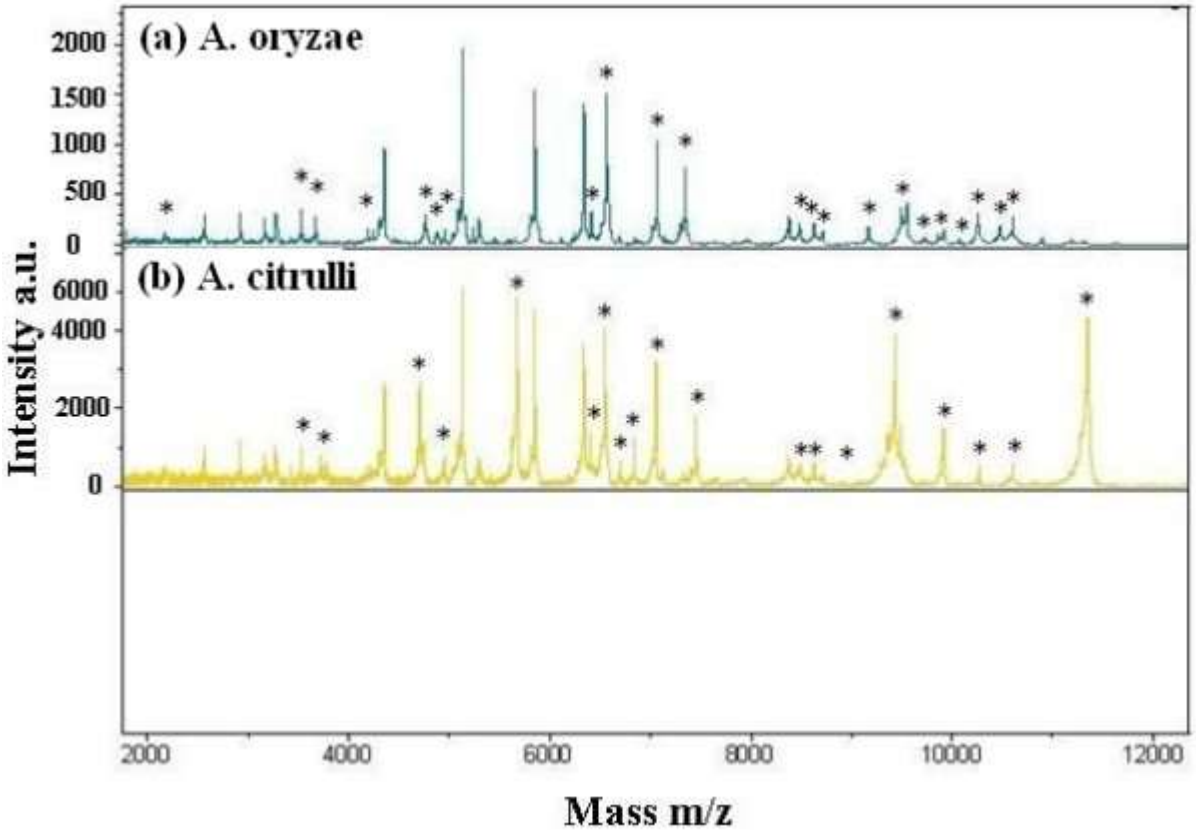


Emerging MS technologies

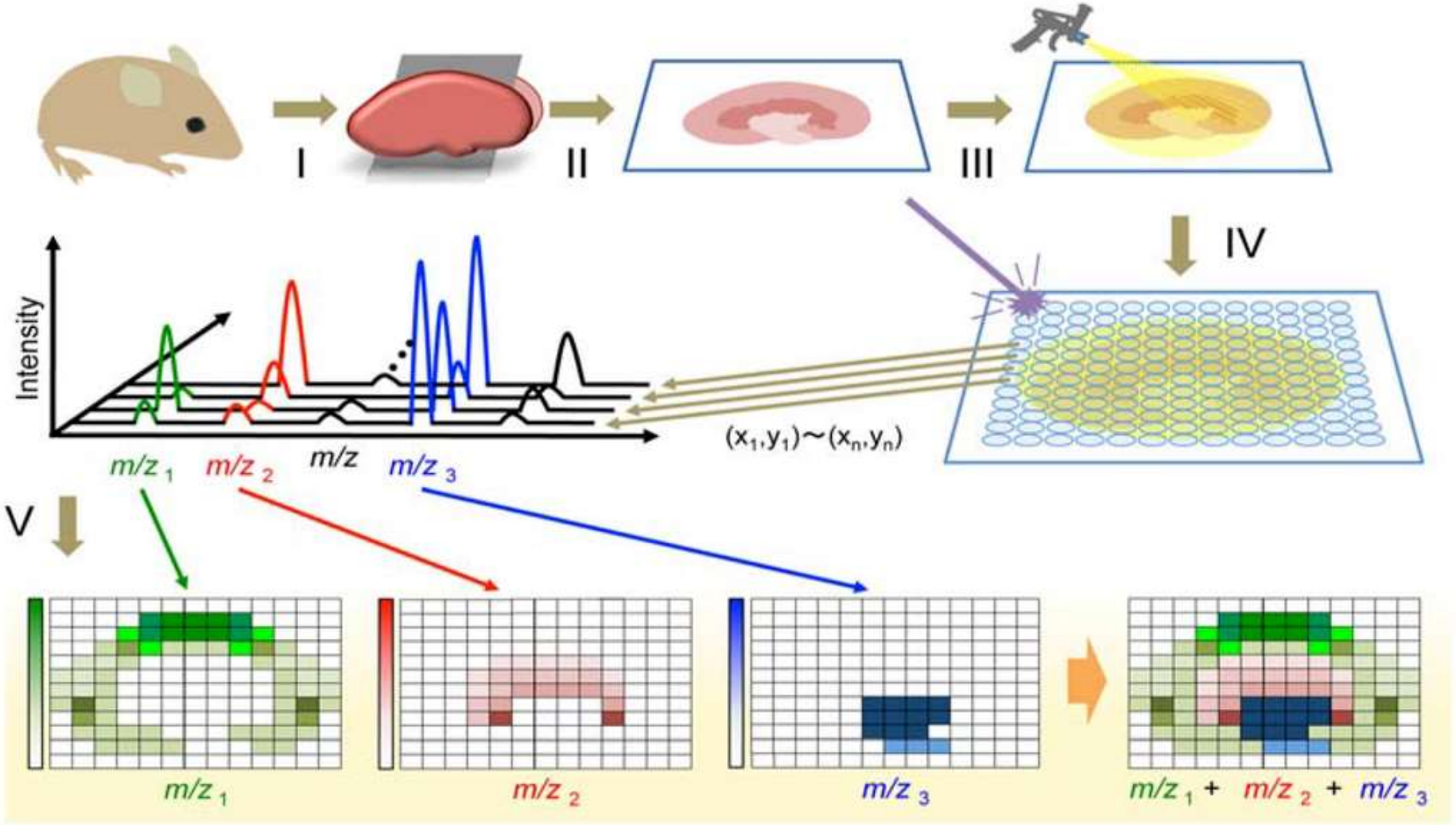


MS identification in Bacteriology (Biotyper)



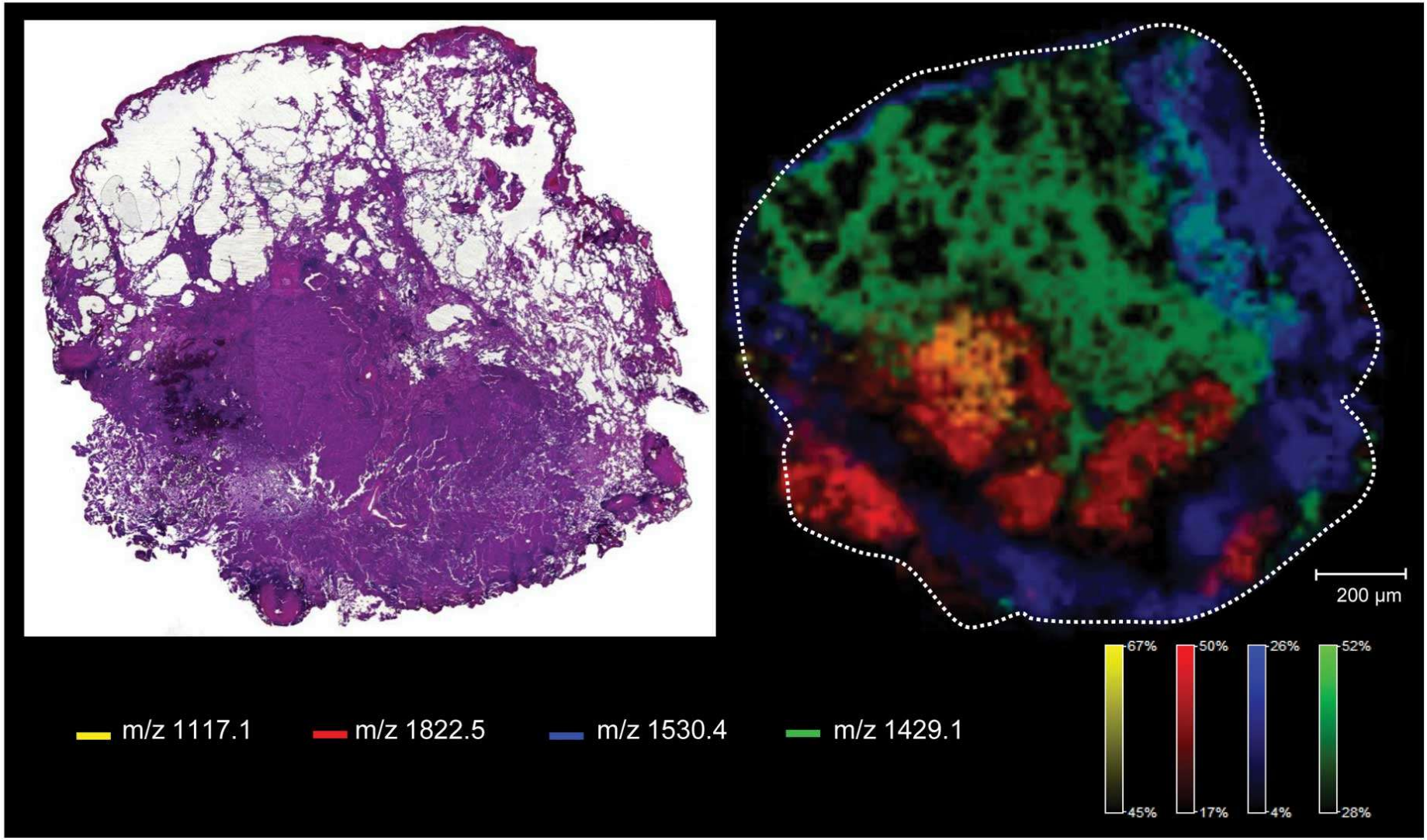


Mass spectrometry imaging

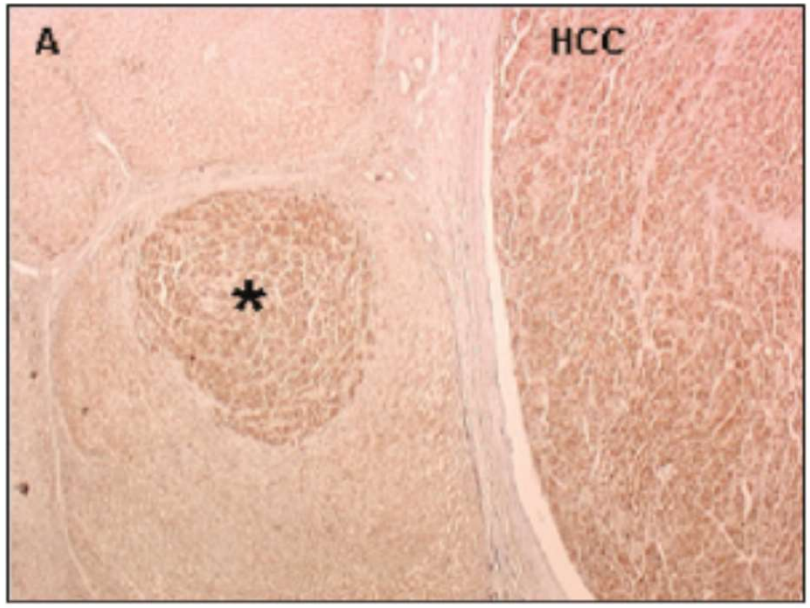


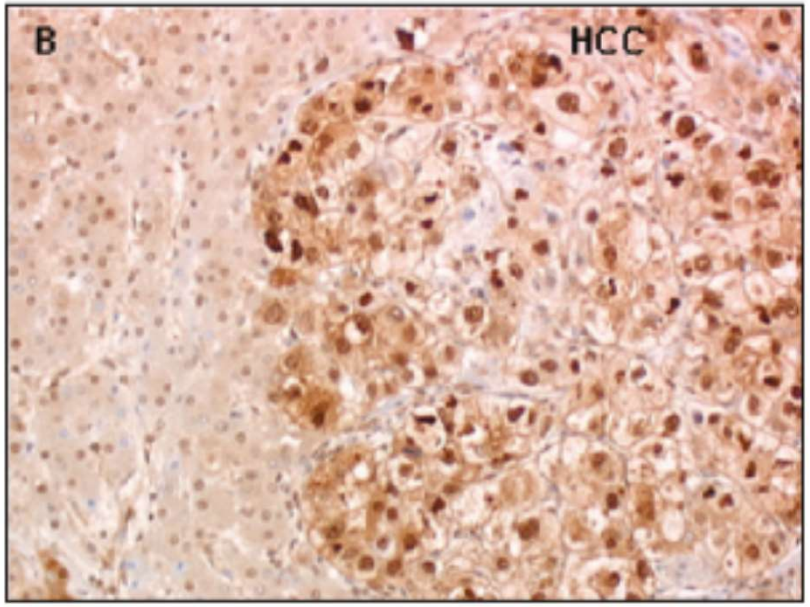
I Sacrifice and organ dissection
II Cryosectioning and moving to ITO glass slide
III Matrix deposition
IV MALDI laser 2D scanning
V Reconstruction of intensity image

Mass spectrometry imaging



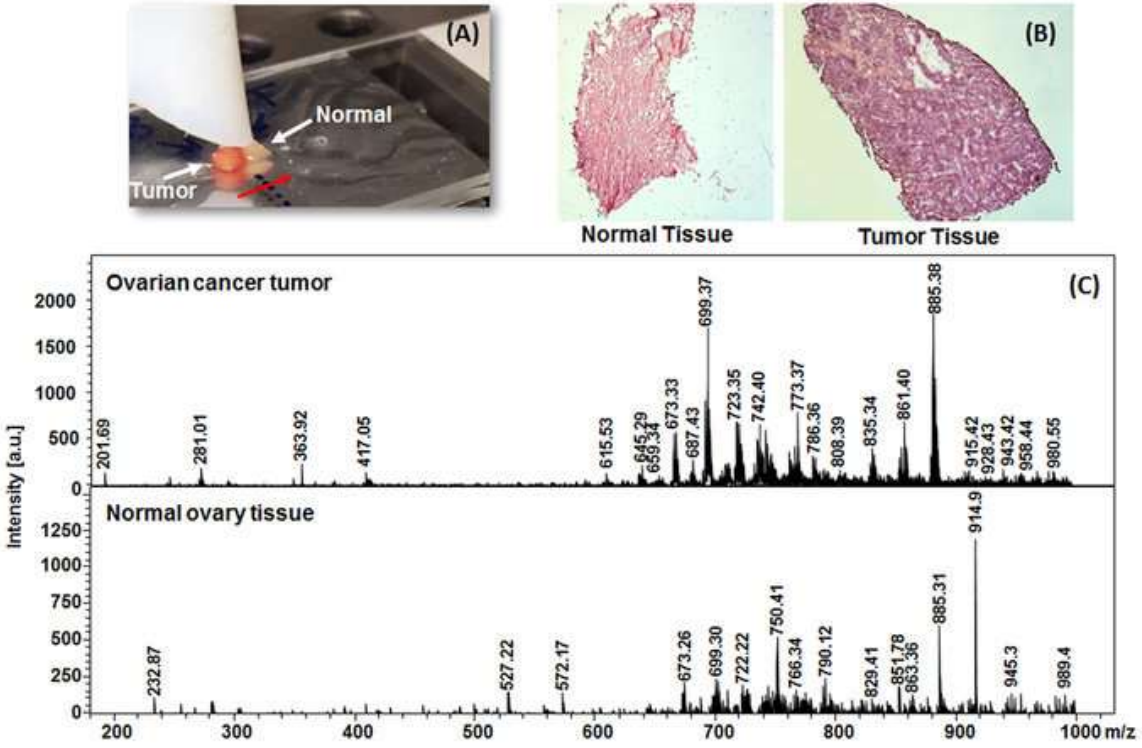
Immunohistochemical validation





SPIDERMASS

PRISM
Protéomique • Réponse Inflammatoire • Spectrométrie de Masse
INSERM U 1192 – Université de Lille



LERES

ANALYSES - RECHERCHE

https://leres.ehesp.fr/

Thanks for your attention!!!