Protein mass spectrometry and proteomics

L. Dayon

Spring Semester 2025

EPFL

Course outline

- 1. Introduction

Introduction to protein analysis and proteomics; Reminders in mass spectrometry; Why proteomics and mass spectrometry?; Ionization sources, analysers, and detectors used in proteomics; Latest generation of mass spectrometers used in proteomics

- 2. Proteomic strategy and workflows

Bottom-up versus top-down strategies; Data-dependent acquisition (DDA) and data-independent acquisition (DIA) approaches; Sample preparation

- 3. Separations techniques in proteomics

Gel electrophoresis; Isoelectric focusing; Liquid chromatography (RP, IEX)

- 4. Quantitative proteomic workflows

Label-free methods; Labelling-based techniques; Other quantitative techniques

- 5. Proteomic bioinformatics

Databases; Identification of protein; Quantification of proteins; Bioinformatics tools; Practical examples

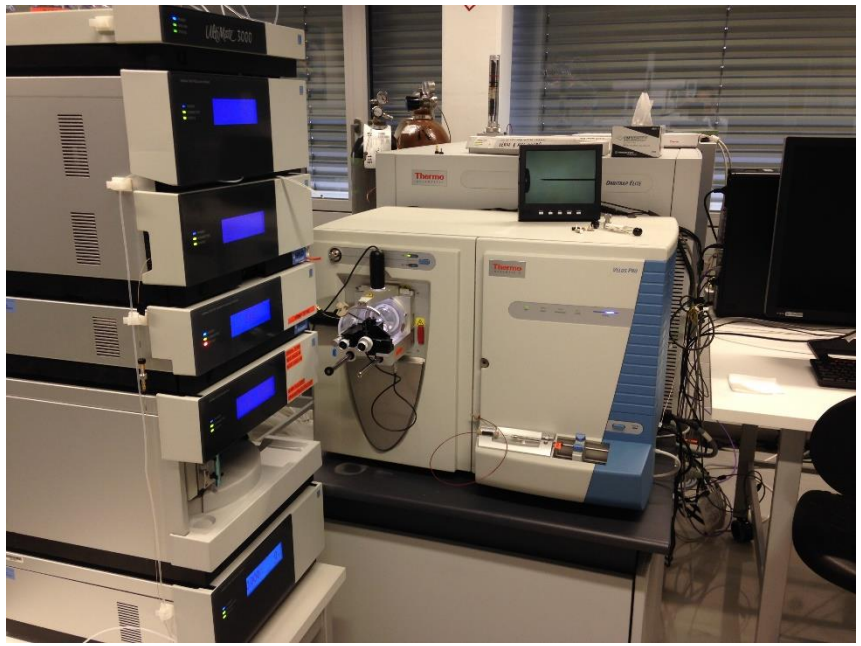
- 6. Applications to biology and clinical research

What strategy?; Experimental design; Biomarker discovery; Industrialized and population proteomics; Forensics; Targeted mass spectrometry-based approaches; Other biological applications of mass spectrometry; Advanced innovations (single-cells, 4D proteomics, multi-omics) and emerging technologies; Limitations and ethical consideration; Lab visit

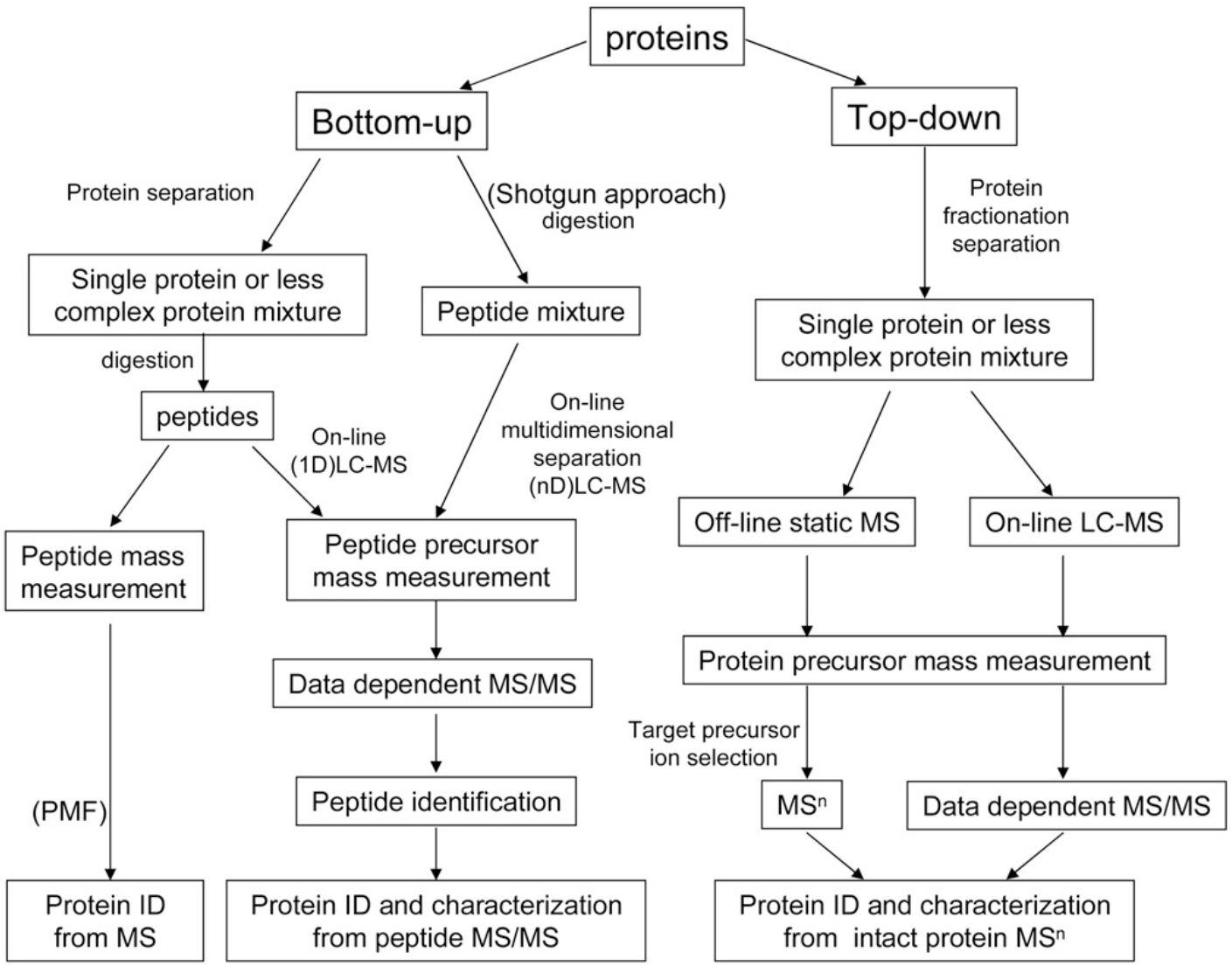
Course outline

- 2. Proteomic strategy and workflows

Bottom-up versus top-down strategies; Sample preparation



Choice of the proteomic strategy



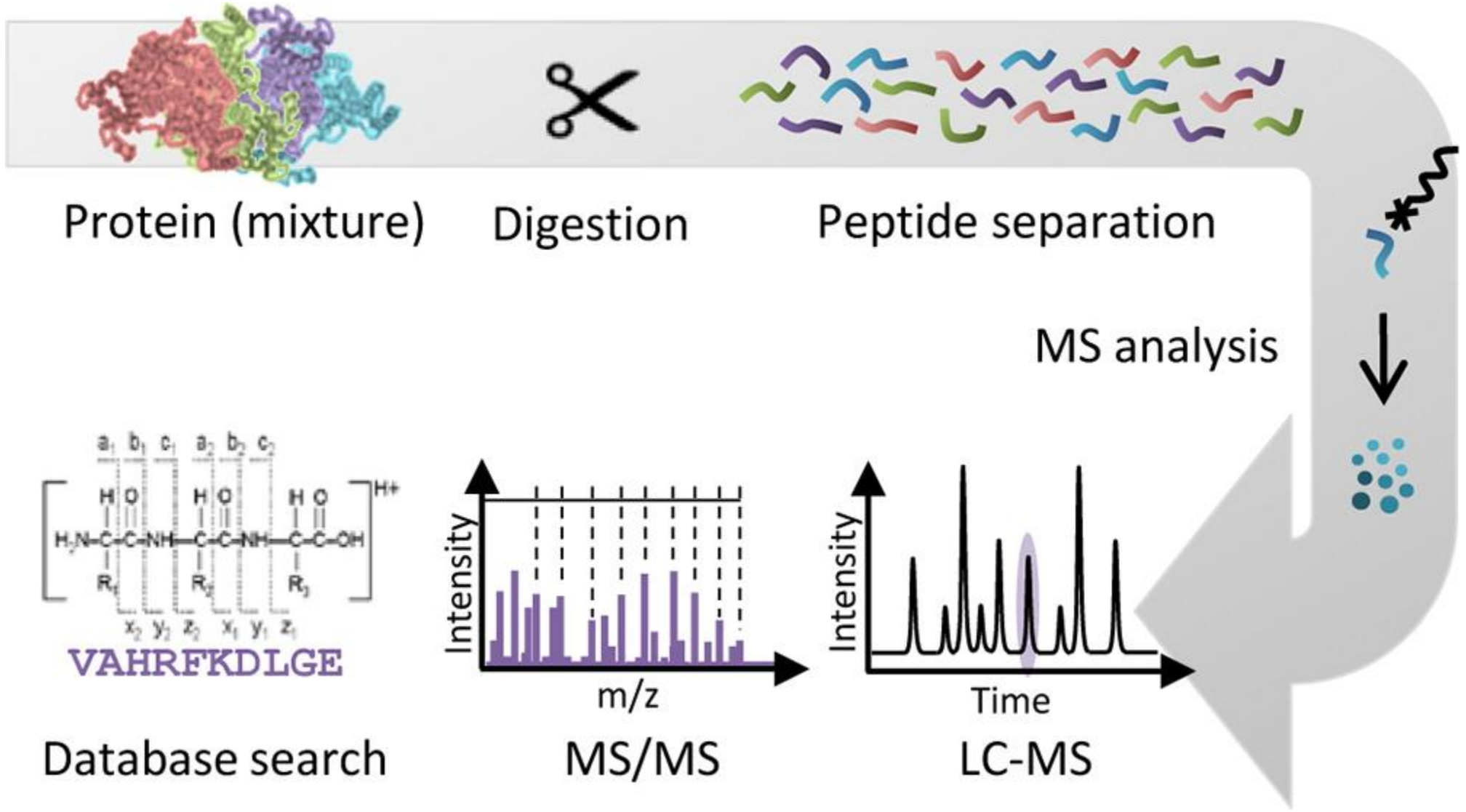
Sample preparation and separation

Mass spectrometry

Data processing and analysis

doi: 10.1016/j.cbpa.2008.07.024

2.1. Bottom-up proteomics



J. Proteome Res., 2013, 12 (3), pp 1067–1077

Why bottom-up proteomics?

- Large proteins are derived into smaller peptides which are easier to separate and analyse with MS
- Fragmentation of tryptic peptides well understood
- Reliable software available for analysis
- It is the most commonly used method to identify proteins
- Several quantitative methods are available
- It is compatible with many workflows and methodologies
- It is robust and sensitive

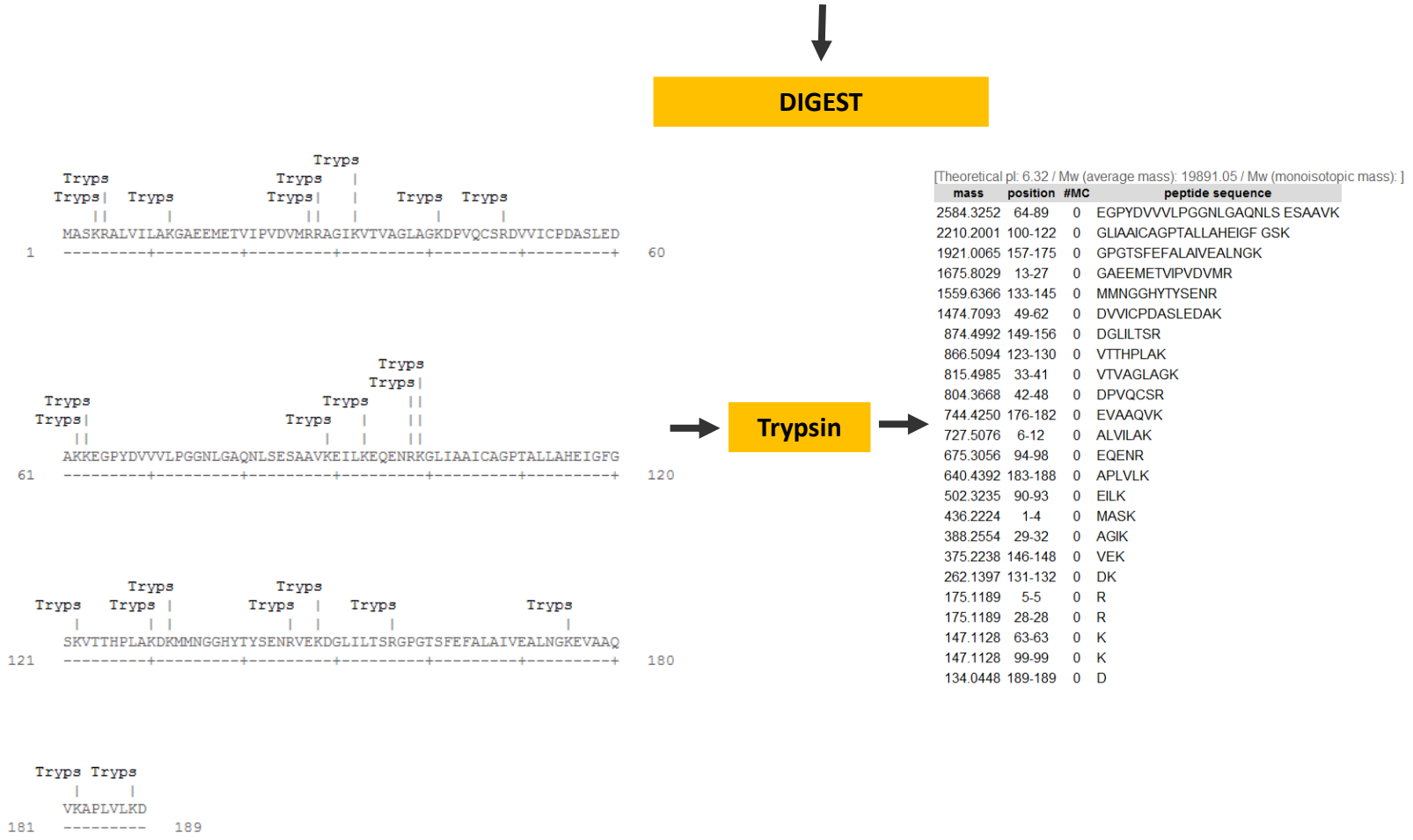
But:

- Protein sequence coverage might be limited
- There might be ambiguity because of non-unique sequences (protein inference problem)
- Post-translational modification (PTM) information may be lost

How to identify proteins from peptides with MS?

MASKRALVILAKGAEEMETVIPVDVMRRAGIKVTVAGLAGKDPVQCSRDVVICPDASLEDAKKEGPYDVVVLPGGNLGAQNLSESA AVKEILKEQENRKGLIAAICAGPTALLAHEIGFGSKVTTHPLAKDKMMNGGHYTYSERVEKDGLILTSRGPGTSFEFALAIVEALNGKEV AAQVKAPLVLKD

Q99497 (PARK7_HUMAN)



[Theoretical pI: 6.32 / Mw (average mass): 19891.05 / Mw (monoisotopic mass):]

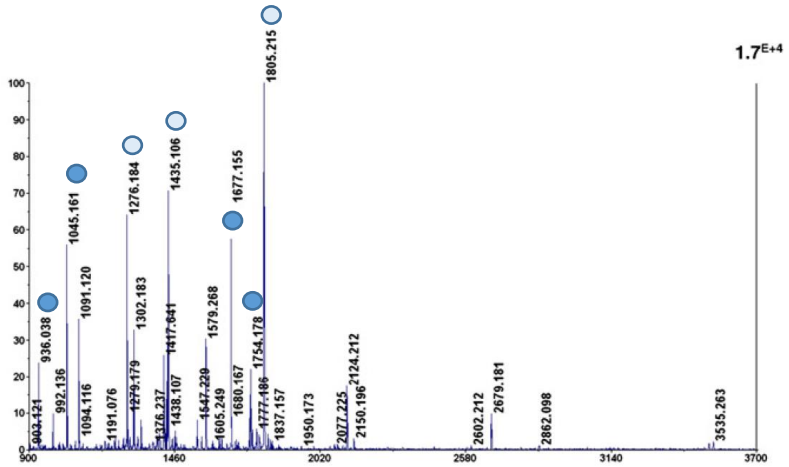
mass	position	#MC	peptide sequence
2584.3252	64-89	0	EGPYDVVVLPGGNLGAQNLS ESAAVK
2210.2001	100-122	0	GLIAAICAGPTALLAHEIGF GSK
1921.0065	157-175	0	GPGTSFEFALAIVEALNGK
1675.8029	13-27	0	GAEEMETVIPVDVMR
1559.6366	133-145	0	MMNGGHYTYSENR
1474.7093	49-62	0	DVVICPDASLEDAK
874.4992	149-156	0	DGLILTSR
866.5094	123-130	0	VTTHPLAK
815.4985	33-41	0	VTVAGLAGK
804.3668	42-48	0	DPVQCSR
744.4250	176-182	0	EVAAQVK
727.5076	6-12	0	ALVILAK
675.3056	94-98	0	EQENR
640.4392	183-188	0	APLVLK
502.3235	90-93	0	EILK
436.2224	1-4	0	MASK
388.2554	29-32	0	AGIK
375.2238	146-148	0	VEK
262.1397	131-132	0	DK
175.1189	5-5	0	R
175.1189	28-28	0	R
147.1128	63-63	0	K
147.1128	99-99	0	K
134.0448	189-189	0	D

Peptide mass fingerprinting (PMF)

P00698 (LYSC_CHICK)

MRSLLILVLC FLPLAALGKV FGRCELAAAM KRHGLDNYRG YSLGNWVCAA KFESNFNTQA TNRNTDGSTD YGILQINSRW WCNDGRTPGS RNLCNIPCSA LLSSDITASV NCAKKIVSDG NGMNAWVAWR NRCKGTDVQA WIRGCRL

DIGEST WITH TRYPSIN



- Chain Lysozyme C at positions 19 - 147 [Theoretical pI: 9.32 / Mw (average mass): 1

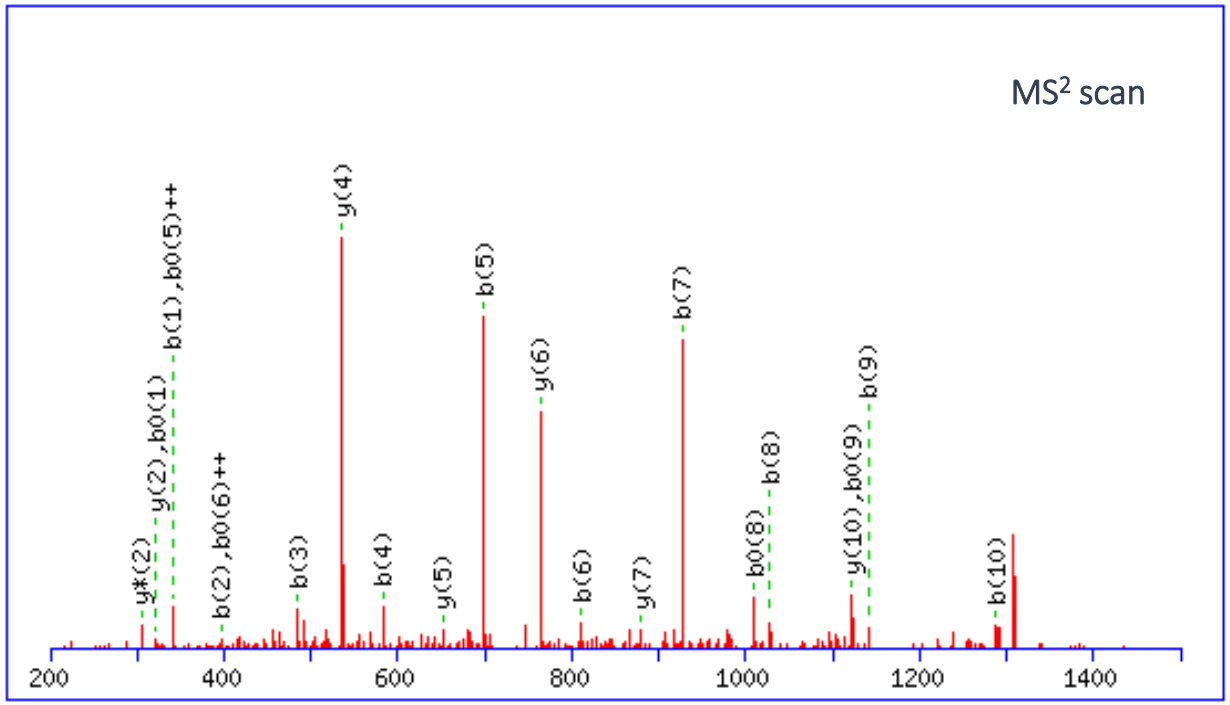
mass	position	#MC	modifications	peptide sequence
3163.4675	52-79	1		FESNFNTQATNRNTDGSTDY GILQINSR
2835.3797	87-114	1		TPGSRNLCNIPCSALLSSDI TASVNCAK
2678.2416	40-63	1		GYSLGNWVCAAKFESNFNTQ ATNR
2671.1954	64-86	1		NTDGSTDYGILQINSRWWCN DGR
2465.2196	92-115	1		NLCNIPCSALLSSDITASVN CAKK
2337.1247	92-114	0		NLCNIPCSALLSSDITASVN CAK
2124.0079	33-51	1		HGLDNYRGYSLGNWVCAAK
1945.9449	116-132	1		IVSDGNGMNAWVAWRNR
1803.8959	115-130	1		KIVSDGNGMNAWVAWR
1753.8351	64-79	0		NTDGSTDYGILQINSR
1675.8009	116-130	0		IVSDGNGMNAWVAWR
1434.6331	80-91	1		WWCNDGRTPGSR
1428.6502	52-63	0		FESNFNTQATNR
1361.6742	135-146	1		GTDVQAWIRGCR
1295.6598	20-31	1		VFGRCELAAAMK
1276.6466	133-143	1		CKGTDVQAWIR
1268.6092	40-51	0		GYSLGNWVCAAK
1045.5425	135-143	0		GTDVQAWIR
1030.5177	32-39	1		RHGLDNYR
992.5016	24-32	1		CELAAAMKR
936.3781	80-86	0		WWCNDGR

Experimental masses

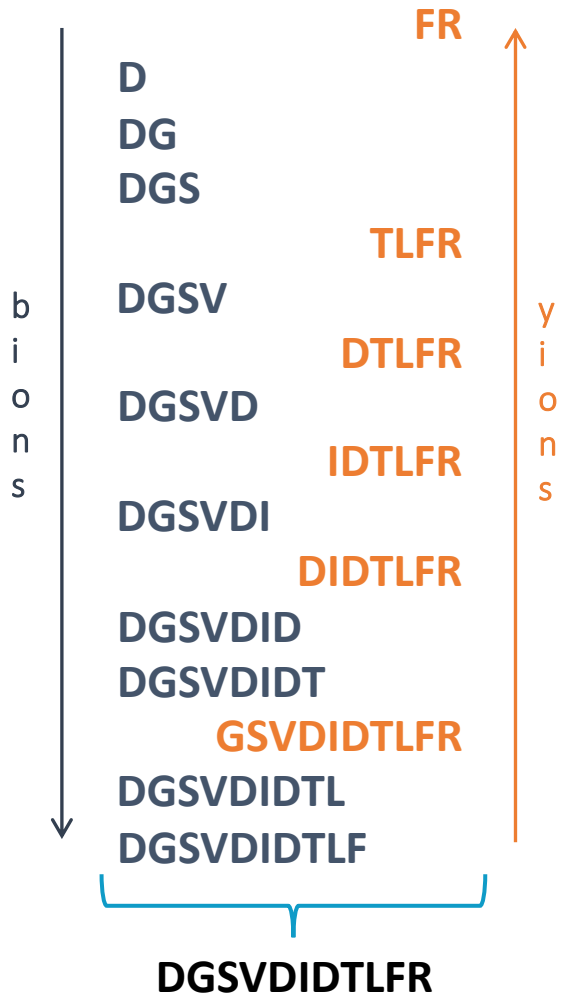
Theoretical masses

Protein identification using MS/MS data

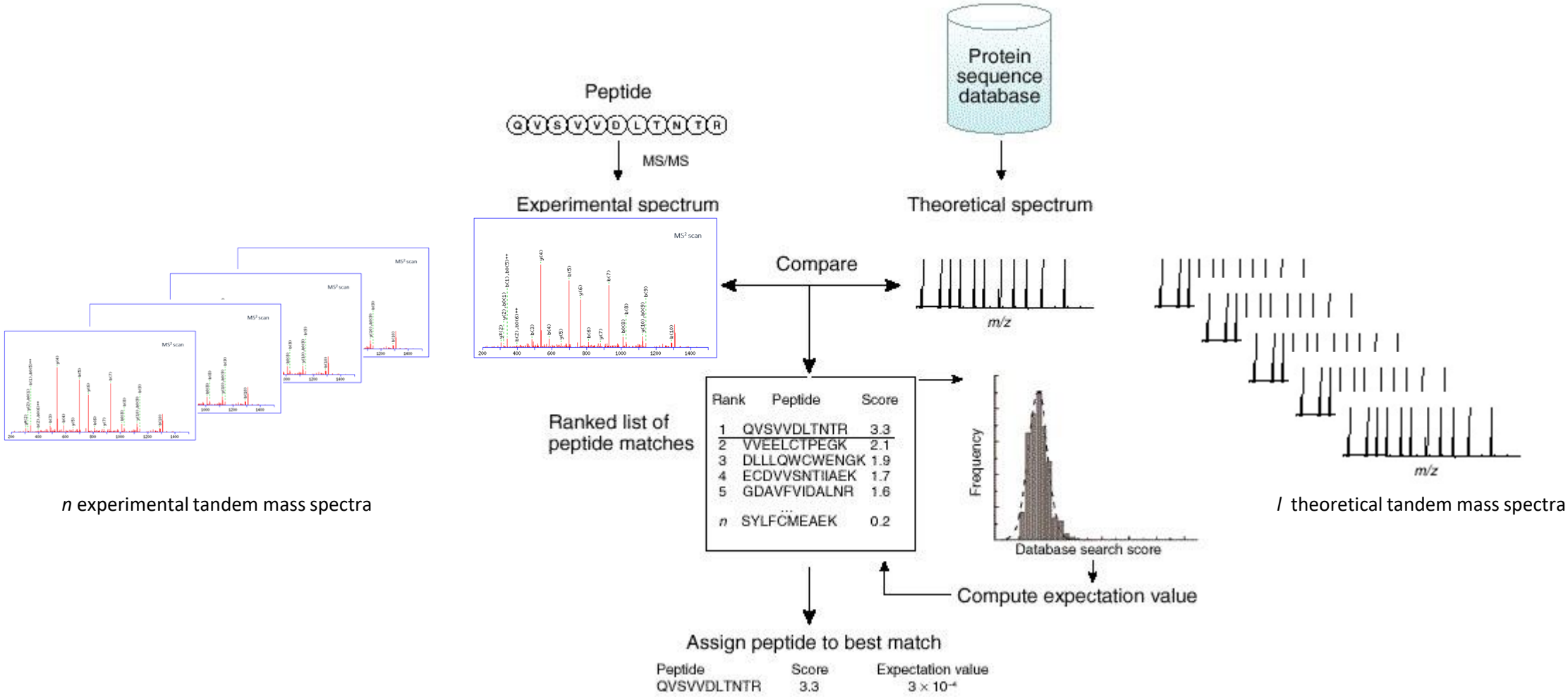
Here, we measure masses of peptide-fragment ions



We match with theoretical fragments



Protein identification using MS/MS data (MS/MS ion search)

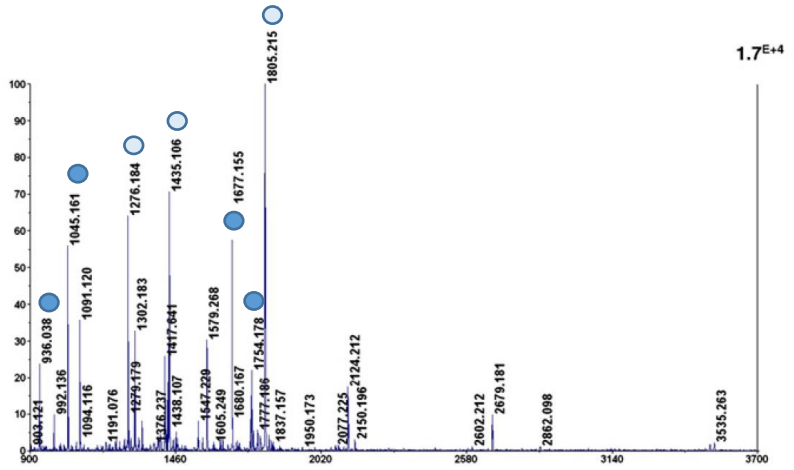


n experimental tandem mass spectra

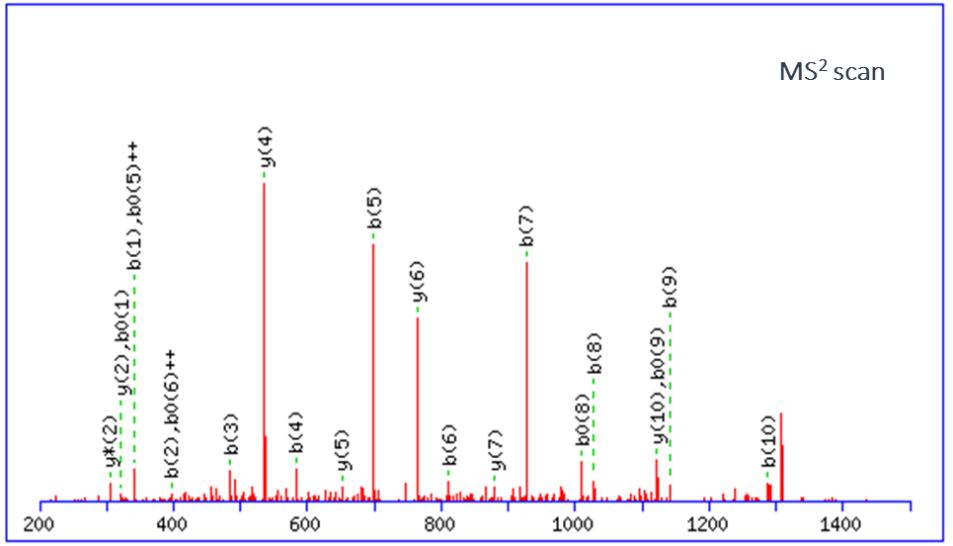
l theoretical tandem mass spectra

DOI: 10.1038/nmeth1088

PMF *versus* MS/MS ion search



versus

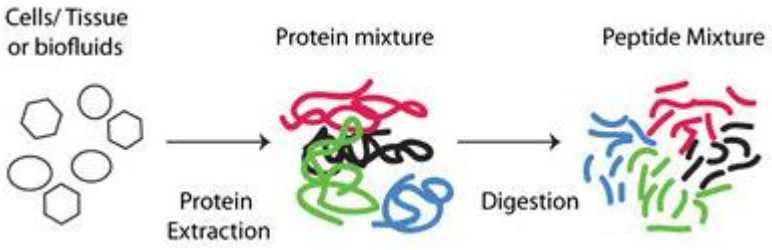


Experimental masses

Experimental fragment masses

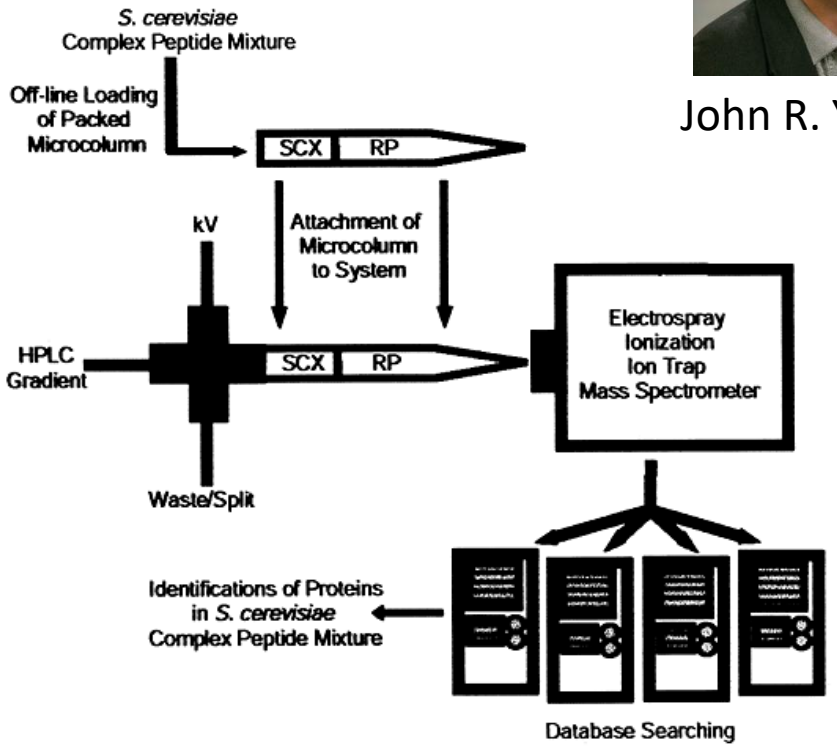
Q1: Which one(s) do(es) work for purified proteins? Which one(s) do(es) work for complex protein mixtures?

Direct analysis of complex protein mixtures



DOI: 10.1186/s40169-014-0034-1

John R. Yates III

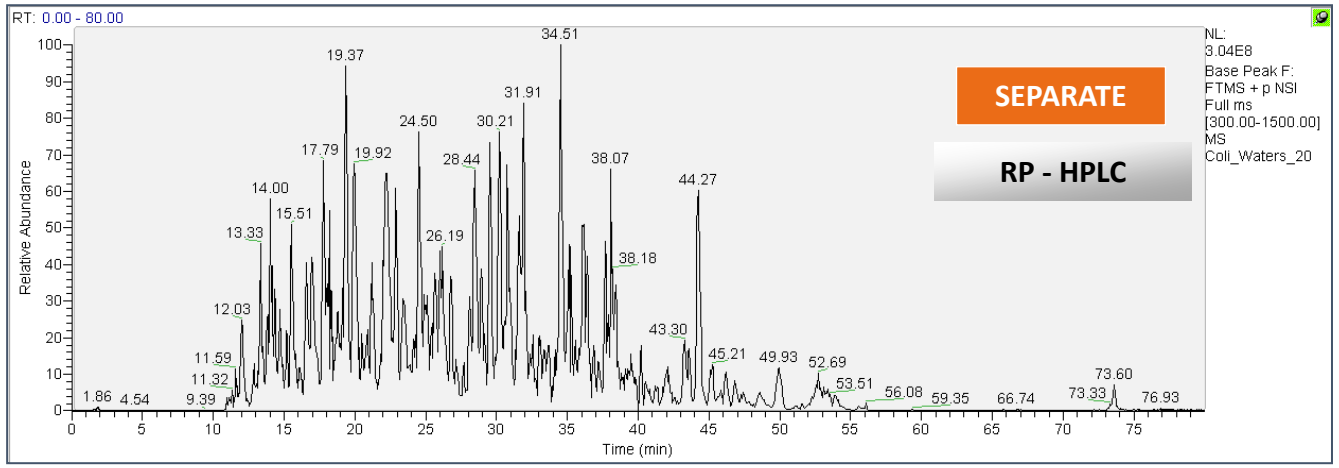


doi:10.1038/85686

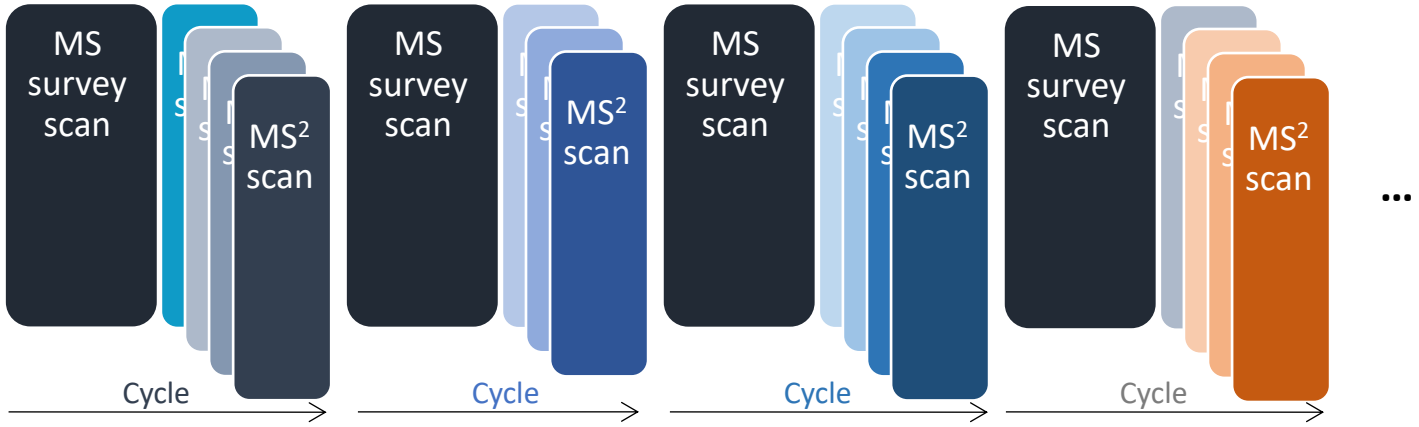
- Thousands of proteins generate hundreds of thousands of peptides
- Mass spectrometry need to be coupled to a separation technique
- A systematic methodology should be applied
- **This is shotgun proteomics!**

Liquid chromatography tandem mass spectrometry (LC-MS/MS)

DIGEST => complex peptide mixture

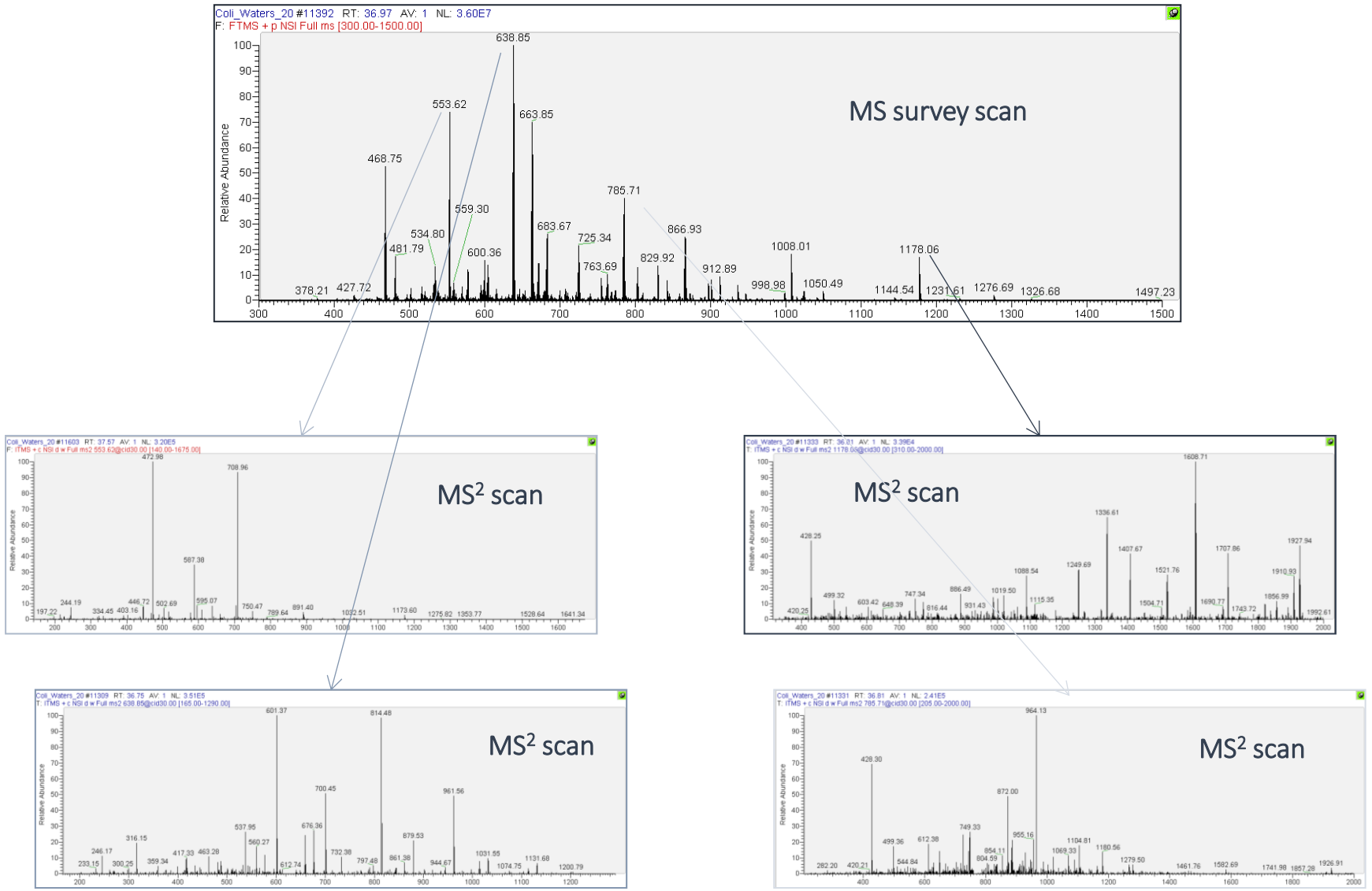


Data-dependent acquisition (**DDA**)



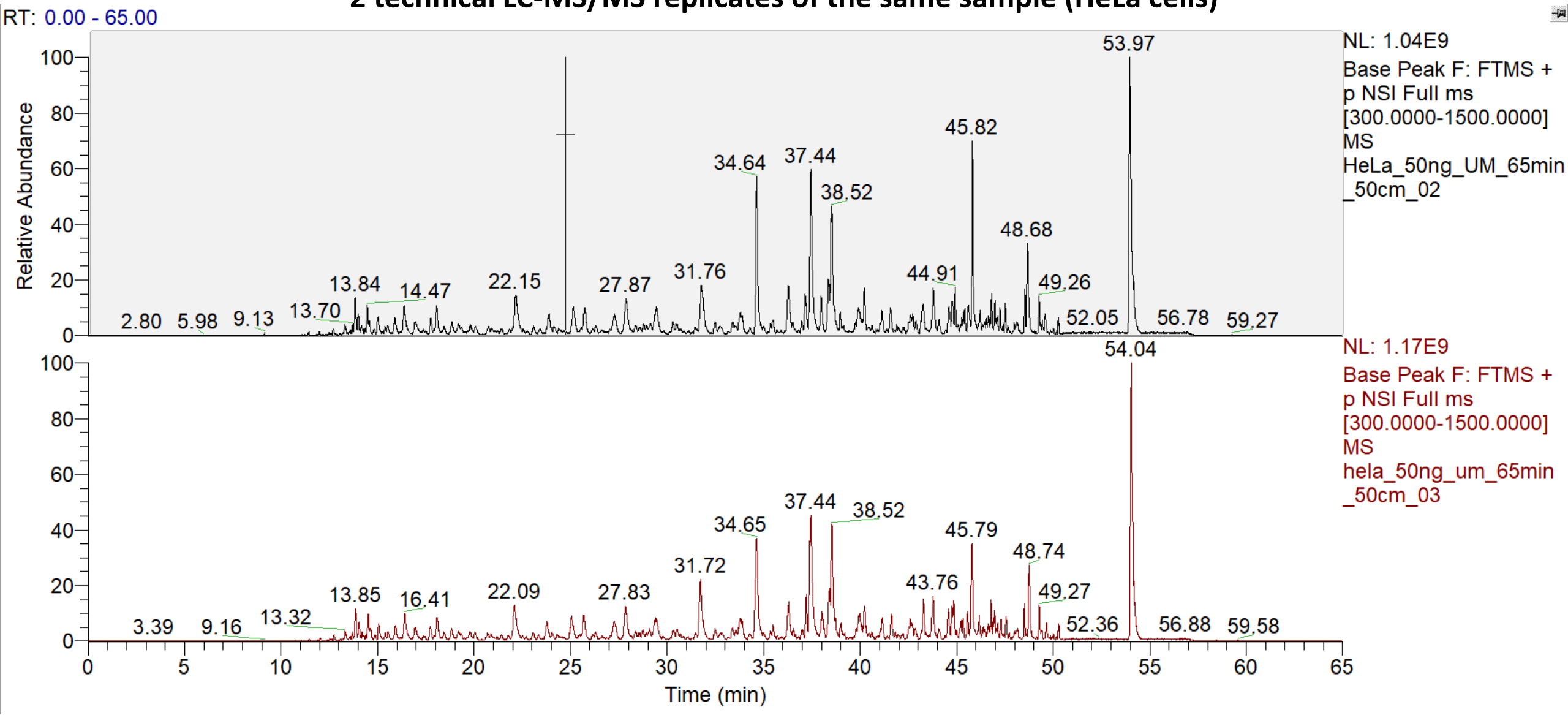
Note: other MS data acquisition types will be seen in next chapters

Data-dependent acquisition (DDA)

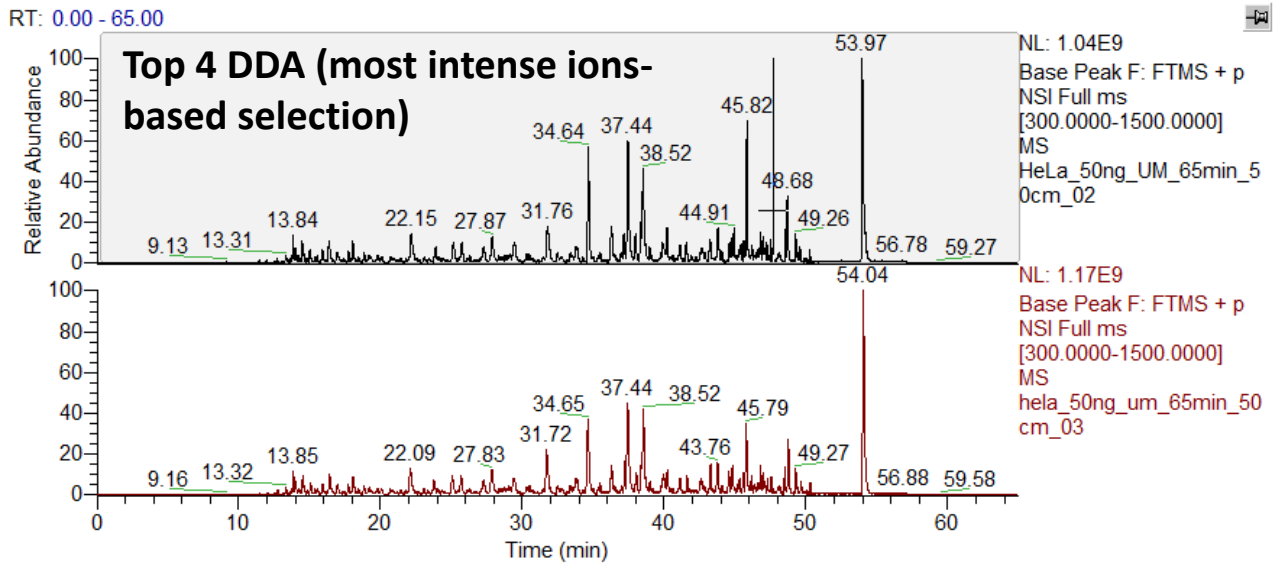


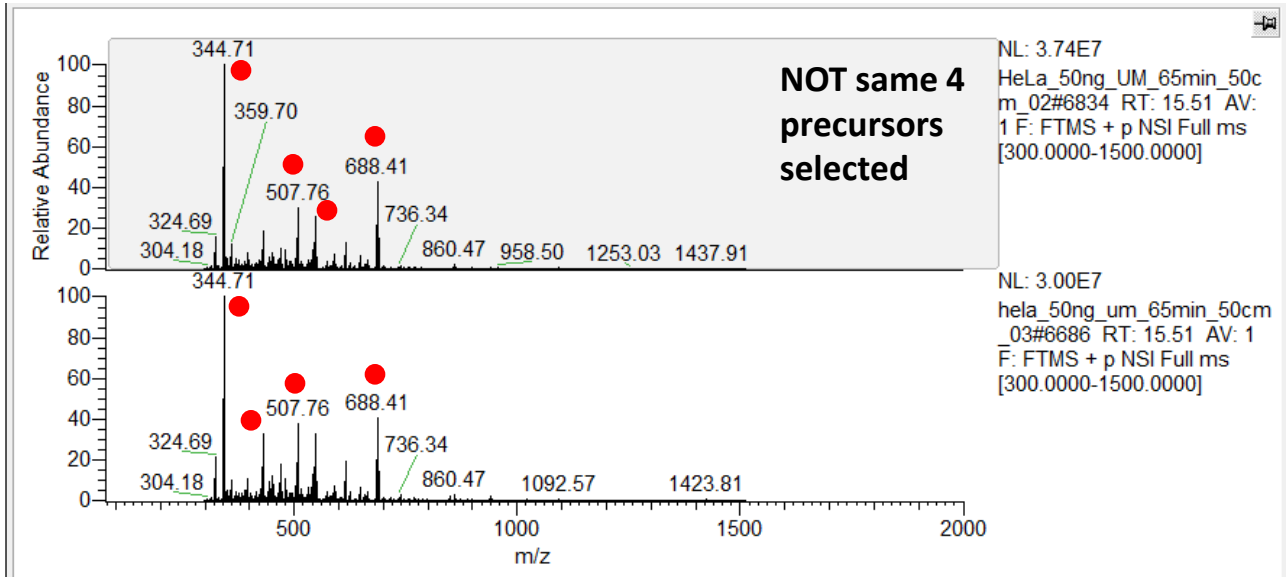
DDA is a stochastic process (1)

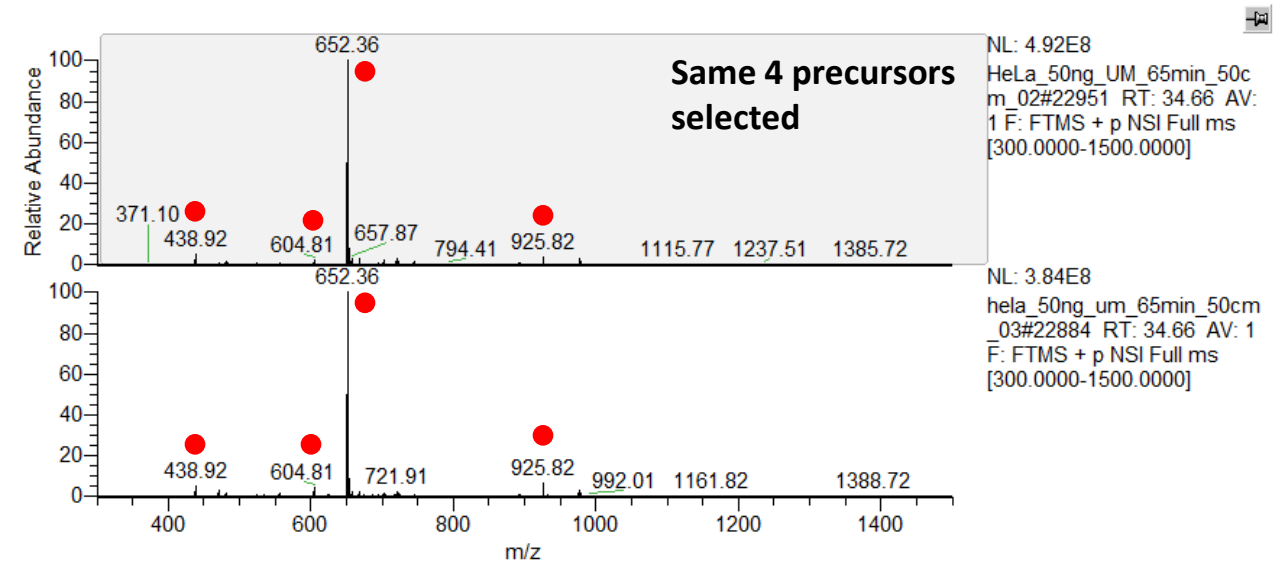
2 technical LC-MS/MS replicates of the same sample (HeLa cells)

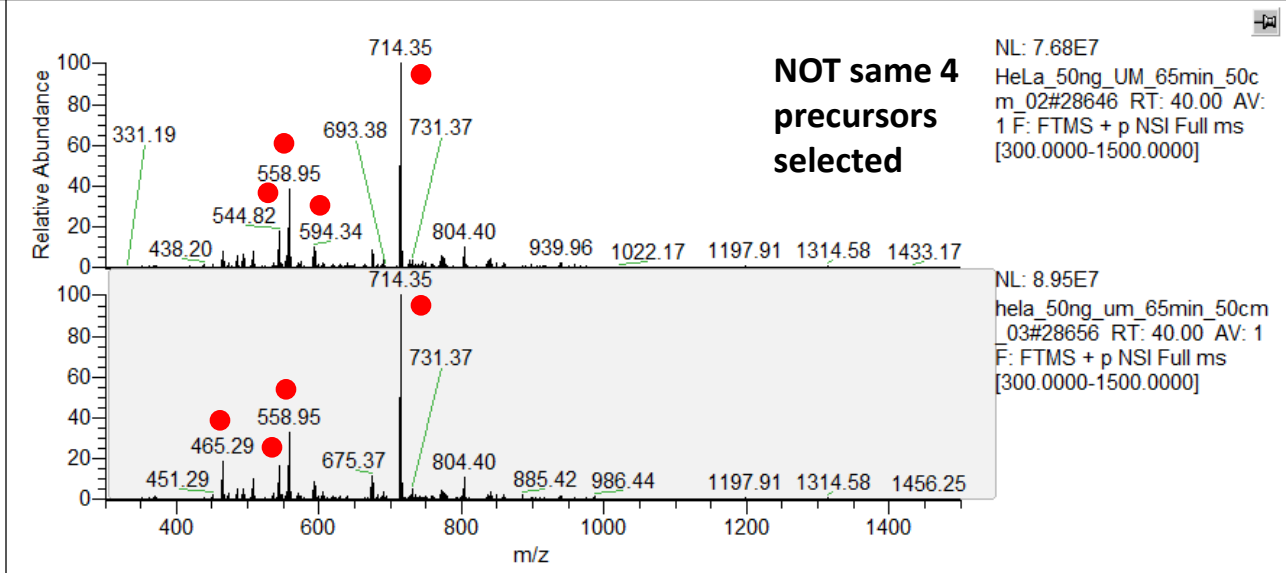


DDA is a stochastic process (2)



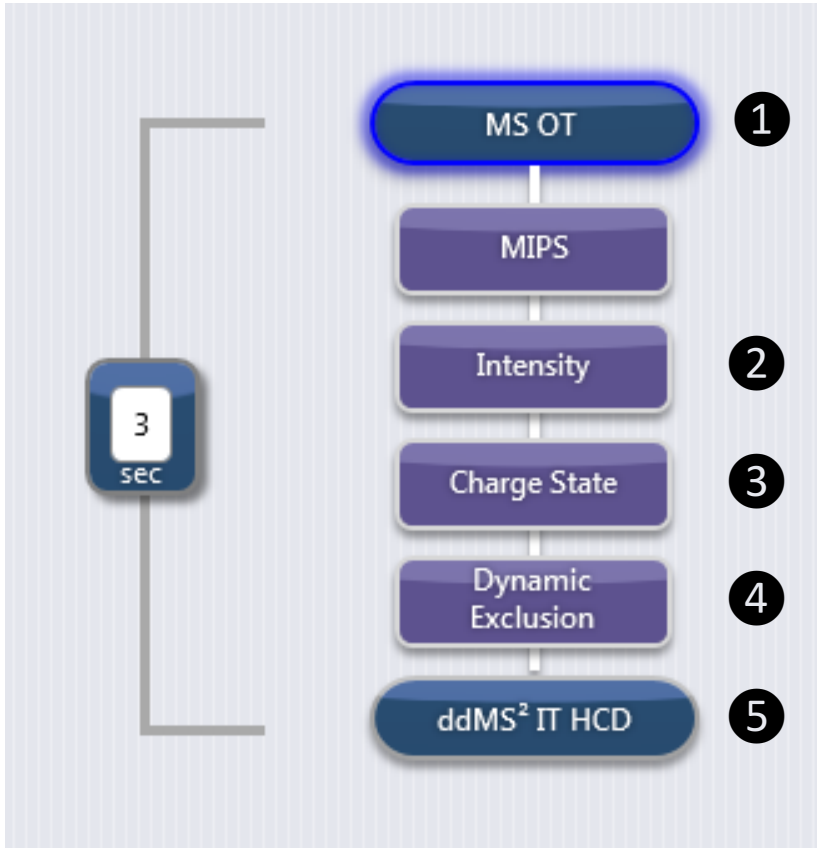






The selection of precursors is not always reproducible -> intrinsic stochastic nature of DDA, partial reproducibility of data collection

The acquisition method is defined by the user in the mass spectrometer software



1

MS Scan Properties — Show All

Orbitrap Resolution	120000
Scan Range (m/z)	300-1500
RF Lens (%)	30
AGC Target	2.0e5
Maximum Injection Time (ms)	100
Use EASY-IC™	☐

2

Intensity Properties

Filter Type	Intensity Threshold
Intensity Threshold	5.0e3

3

Charge State Properties

Include charge state(s)	2-7
Include undetermined charge states	☐
Include charge states 25 and higher	☐

4

Dynamic Exclusion Properties

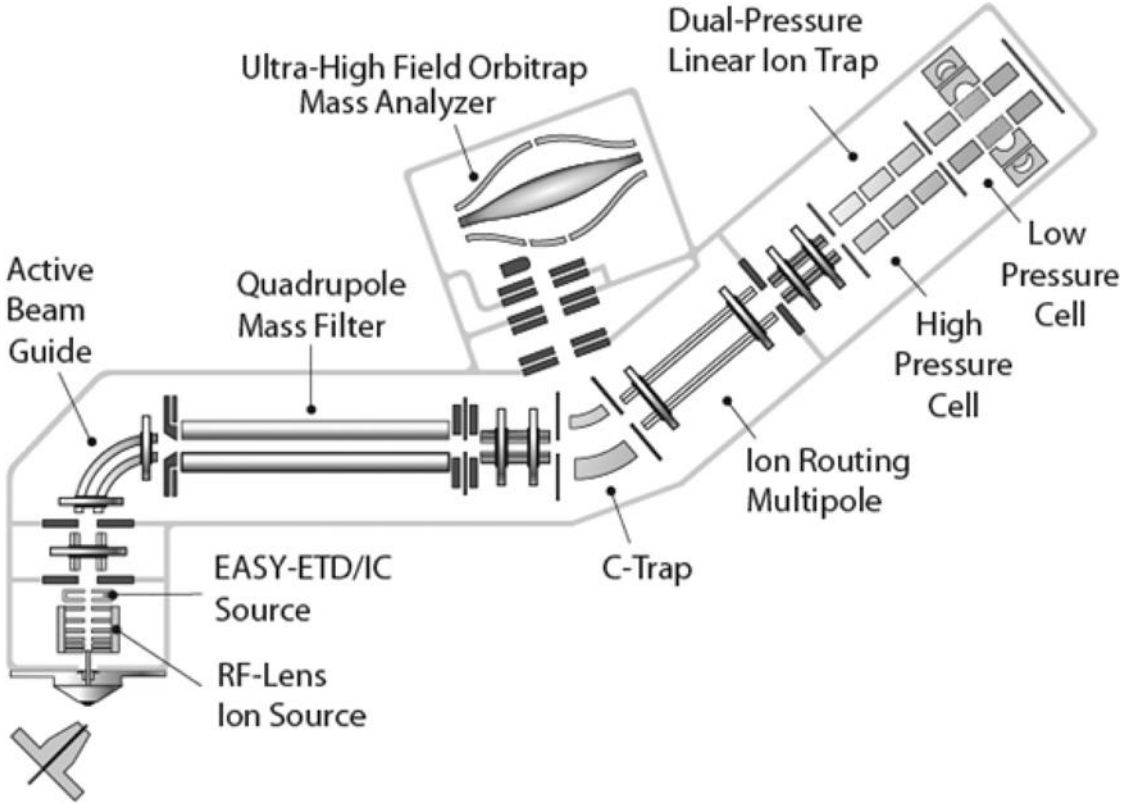
Exclude after n times	1
Exclusion duration (s)	60
Mass Tolerance	ppm
Low	10.00
High	10.00
Exclude Isotopes	☑
Perform dependent scan on single charge state per precursor only	☑

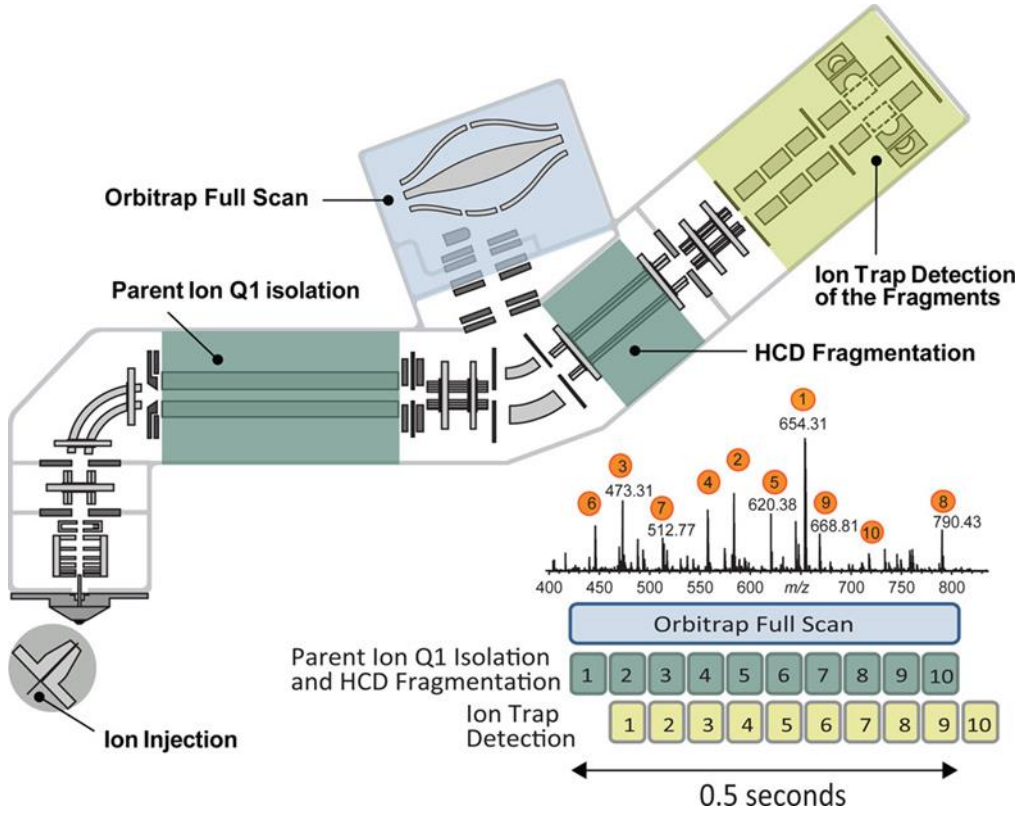
5

Data-Dependent MSⁿ Scan Properties — Show All

Isolation Window (m/z)	1.2
Activation Type	HCD
HCD Collision Energy (%)	30
Detector Type	Ion Trap
Ion Trap Scan Rate	Rapid
First Mass (m/z)	120
AGC Target	2.0e3
Inject Ions for All Available Parallelizable Time	☑
Maximum Injection Time (ms)	300

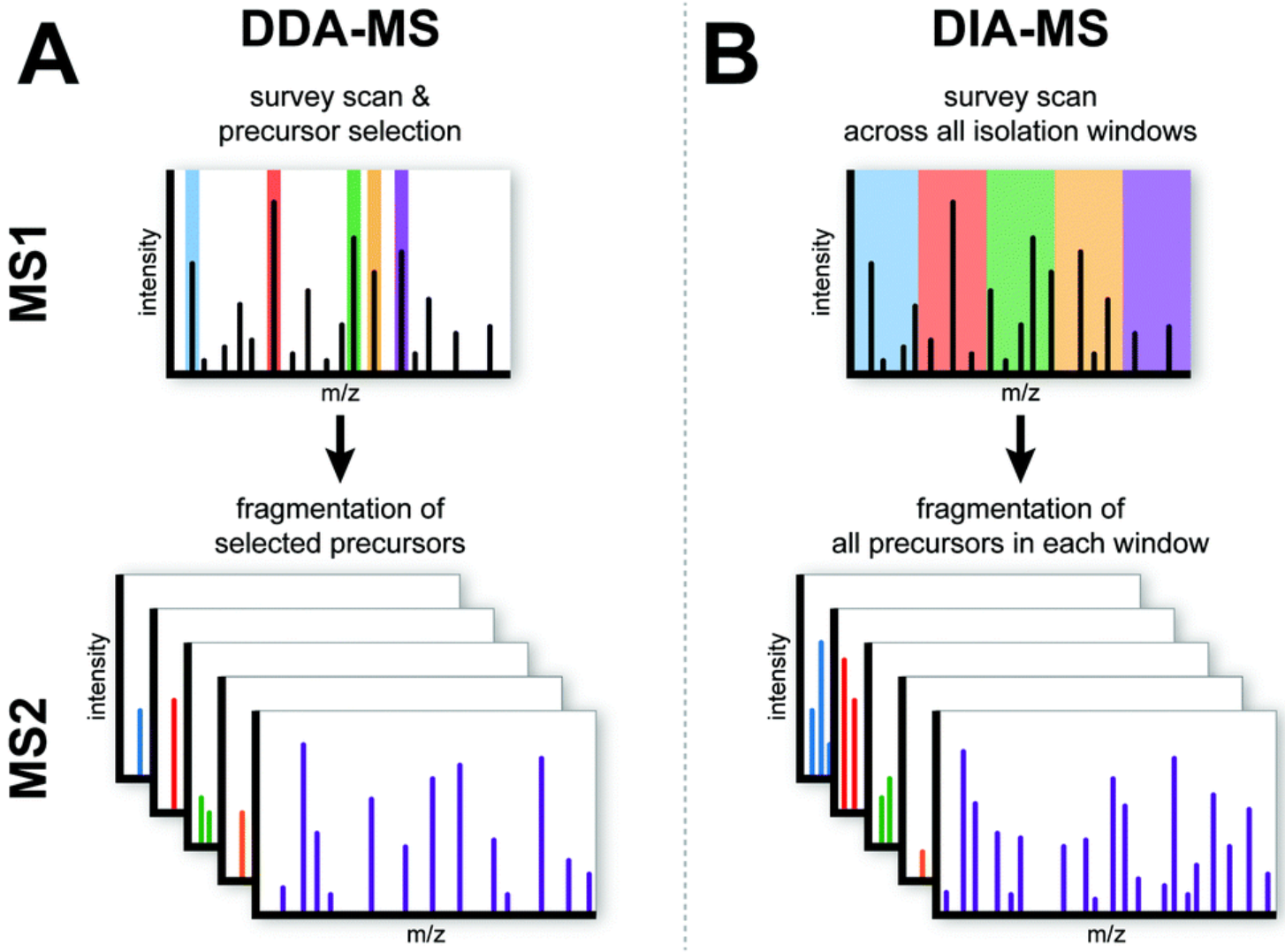
Where does it happen in our mass spectrometer?





DOI: 10.1021/ac403115c

Data-independent acquisition (DIA)



Mol. Omics, 2021, 17, 29-42, doi: 10.1039/D0MO00072H

… And finally, we identify peptide and proteins

545. P07014 **Mass:** 27379 **Score:** 54 **Matches:** 5(3) **Sequences:** 5(3) **emPAI:** 0.41
Succinate dehydrogenase iron-sulfur subunit OS=Escherichia coli (strain K12) GN=sdhB PE=1 SV=1

Query	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	Unique	Peptide
5426	581.2624	1160.5103	1160.5098	0.46	0	21	0.0086	1	U	R.YNPDVDDAPR.M
8989	718.3796	1434.7447	1434.7462	-1.03	0	0	8.6	2	U	R.DMMLLDALIQLK.E
9003	719.3087	1436.6029	1436.6024	0.33	0	30	0.0014	1	U	R.EGVCGSDGLNMNGK.N
9074	481.6569	1441.9488	1441.9497	-0.64	0	30	0.001	1	U	K.IVIRPLPGLPVIR.D
11457	565.6497	1693.9273	1693.9263	0.58	0	5	0.39	1	U	K.IKPYLLNNGQNPPAR.E

546. P13035 **Mass:** 56886 **Score:** 54 **Matches:** 4(2) **Sequences:** 4(2) **emPAI:** 0.12
Aerobic glycerol-3-phosphate dehydrogenase OS=Escherichia coli (strain K12) GN=glpD PE=1 SV=3

Query	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	Unique	Peptide
2909	494.2769	986.5393	986.5396	-0.28	0	1	7.4	2	U	R.LVSEALAER.E
2935	494.7949	987.5753	987.5753	-0.03	0	31	0.0014	1	U	K.APLLSVFGGK.L
5101	571.3206	1140.6266	1140.6291	-2.26	0	38	0.0003	1	U	R.GLVNATGPWVK.Q
13902	664.3203	1989.9391	1989.9392	-0.04	1	3	3	1	U	K.ESVLPGGAIEGDRDDYAAR.L

547. P0AEM0 **Mass:** 16071 **Score:** 54 **Matches:** 2(2) **Sequences:** 2(2) **emPAI:** 0.47
FKBP-type 16 kDa peptidyl-prolyl cis-trans isomerase OS=Escherichia coli (strain K12) GN=fkpB PE=1 SV=2

Query	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	Unique	Peptide
5500	583.2700	1164.5255	1164.5259	-0.32	0	51	2.1e-05	1	U	K.LDDGTTAESTR.N
13131	627.3418	1879.0036	1879.0051	-0.80	0	21	0.037	1	U	R.LGDASLSEGLEQHLLGLK.V

Proteins matching the same set of peptides:
P0AEM1 **Mass:** 16071 **Score:** 54 **Matches:** 2(2) **Sequences:** 2(2)
FKBP-type 16 kDa peptidyl-prolyl cis-trans isomerase OS=Escherichia coli O6:H1 (strain CFT073 / ATCC 700928
P0AEM2 **Mass:** 16071 **Score:** 54 **Matches:** 2(2) **Sequences:** 2(2)
FKBP-type 16 kDa peptidyl-prolyl cis-trans isomerase OS=Escherichia coli O157:H7 GN=fkpB PE=3 SV=2

548. P33368 **Mass:** 27105 **Score:** 54 **Matches:** 2(1) **Sequences:** 2(1) **emPAI:** 0.12
Uncharacterized oxidoreductase YohF OS=Escherichia coli (strain K12) GN=yohF PE=3 SV=2

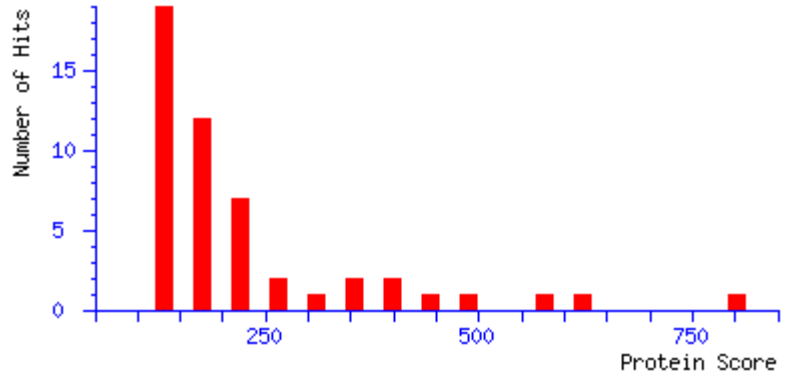
Query	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	Unique	Peptide
8923	715.8959	1429.7772	1429.7776	-0.32	0	54	2.5e-05	1	U	M.AQVAIITASDSGIGK.E
16389	613.0732	2448.2639	2448.2649	-0.41	0	9	0.51	1	U	R.IINITSVHEHTPLPDASAYTAAK.H

549. P0A786 **Mass:** 34463 **Score:** 54 **Matches:** 3(2) **Sequences:** 3(2) **emPAI:** 0.20
Aspartate carbamoyltransferase catalytic chain OS=Escherichia coli (strain K12) GN=pyrB PE=1 SV=2

Query	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	Unique	Peptide
3244	505.2872	1008.5598	1008.5604	-0.57	0	5	1.7	1	U	K.ANPQPELLK.H
4613	552.7739	1103.5333	1103.5346	-1.20	0	25	0.014	1	U	R.VDEIATDVDK.T
6569	622.3483	1242.6821	1242.6819	0.13	0	46	0.00026	1	U	R.DDLNLVLATAAK.L

EFTU_ECOLI **Mass:** 48613 **Score:** 801 **Matches:** 46(35) **Sequences:** 10(9) **emPAI:** 1.35
Elongation factor Tu OS=Escherichia coli (strain K12) GN=tufA PE=1 SV=2

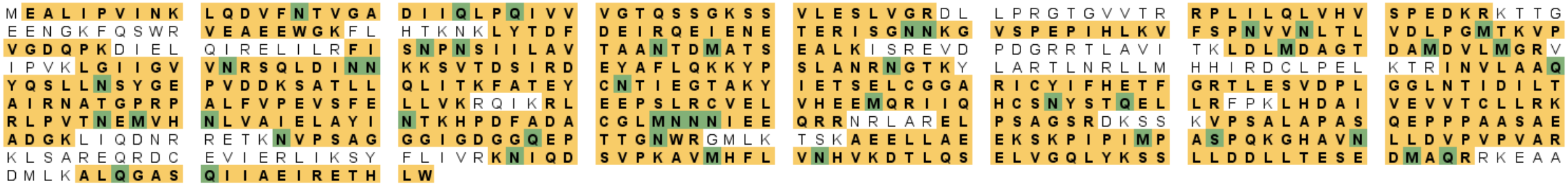
Query	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	Unique	Peptide
1317	460.2534	918.4922	918.4923	-0.15	0	34	0.011	1	U	K.TYGGAAR.A
4389	625.3976	1248.7806	1248.7806	0.02	0	49	0.00044	1	U	R.TVGAGVVAK.V 4392
4416	626.3747	1250.7348	1250.7347	0.13	0	33	0.022	1	U	R.AGENVGVLLR.G
6094	706.3385	1410.6624	1410.6636	-0.80	0	56	3.6e-05	1	U	K.STCTGVEMFR.K
7986	810.4847	1618.9548	1618.9546	0.15	0	61	3.1e-05	1	U	K.VGEEVEIVGIK.E 7969 7974 7979 7984 7989 7991
8409	831.9692	1661.9238	1661.9280	-2.52	0	51	0.00036	1	U	K.FESEVYILSK.D 8410
8457	833.9378	1665.8610	1665.8614	-0.19	0	63	1.9e-05	1	U	K.ALEGDAEWEAK.I 8451 8455 8460 8461 8465 8467 8473
9196	876.5469	1751.0792	1751.0808	-0.94	0	75	8.5e-07	1	U	K.TTLTAAITTVLAK.T 9192 9195 9201 9209 9211 9212
13673	816.7235	2447.1487	2447.1270	8.87	0	67	5.5e-06	1	U	K.CDMVDDEELLELVEMEVR.E 13671 13672
13835	822.0476	2463.1210	2463.1219	-0.37	0	(51)	0.00016	1	U	K.CDMVDDEELLELVEMEVR.E 13841
13836	822.0489	2463.1250	2463.1219	1.27	0	(45)	0.00068	1	U	K.CDMVDDEELLELVEMEVR.E 13834 13837 13838 13839 13840
13956	827.3794	2479.1163	2479.1168	-0.19	0	(38)	0.0019	1	U	K.CDMVDDEELLELVEMEVR.E
14605	642.1268	2564.4780	2564.4618	6.31	0	(31)	0.026	1	U	R.AIDKPFLLPIEDVFSISGR.G
14608	855.8357	2564.4854	2564.4618	9.21	0	49	0.00039	1	U	R.AIDKPFLLPIEDVFSISGR.G 14586 14590 14596



Detailed information on the database search principle will be given in Chapter05

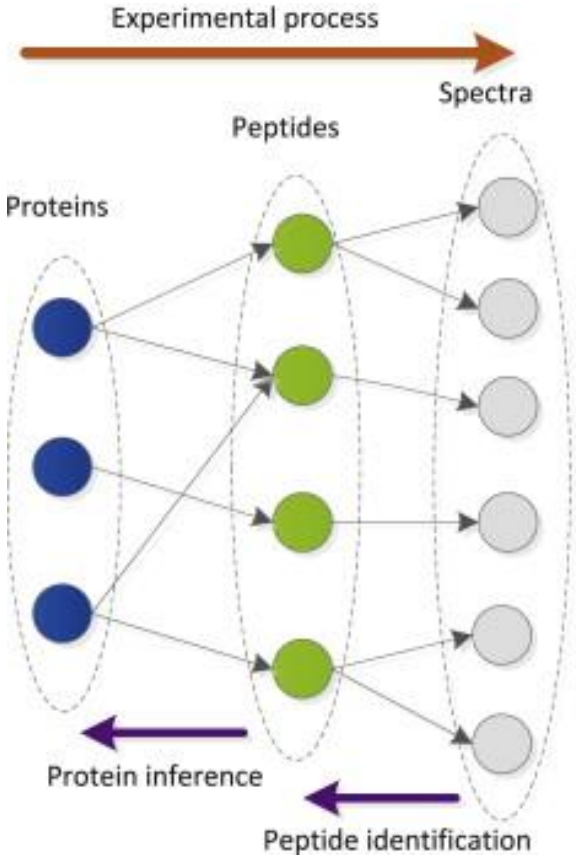
A snapshot of protein characterization with bottom-up proteomics

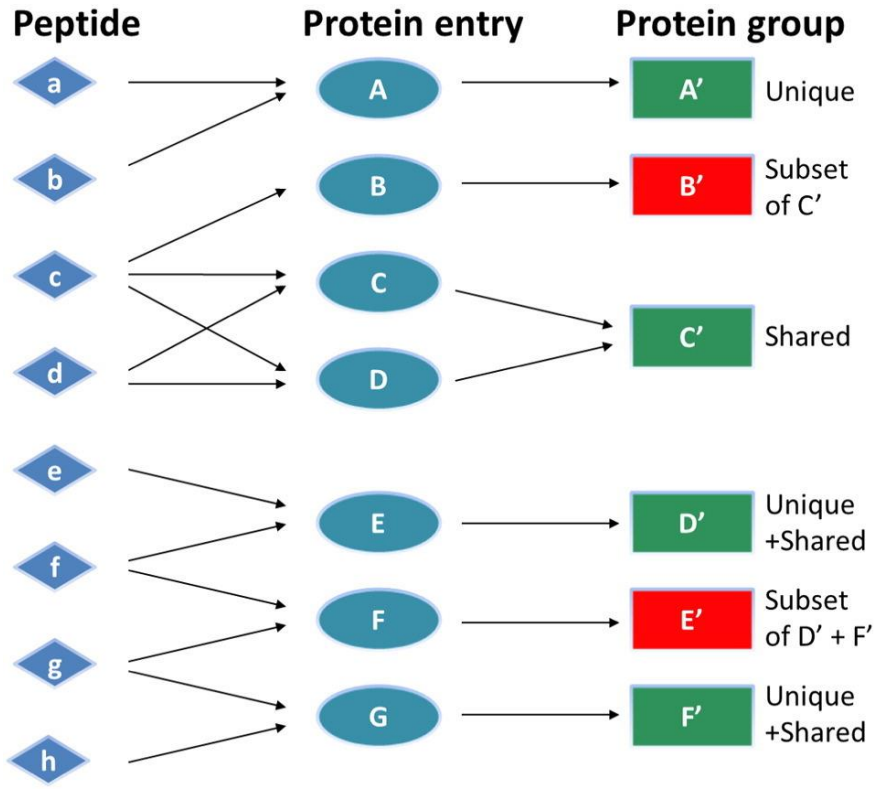
DNM1L_MOUSE (100 %), 82'660.2 Da
Dynamin-1-like protein OS=Mus musculus OX=10090 GN=Dnm1l PE=1 SV=2
52 exclusive unique peptides, 127 exclusive unique spectra, 346 total spectra, 581/742 amino acids (78 % coverage)



GELS_MOUSE (100 %), 85'941.9 Da
Gelsolin OS=Mus musculus OX=10090 GN=Gsn PE=1 SV=3
2 exclusive unique peptides, 2 exclusive unique spectra, 2 total spectra, 13/780 amino acids (2 % coverage)

Protein inference in bottom-up proteomics

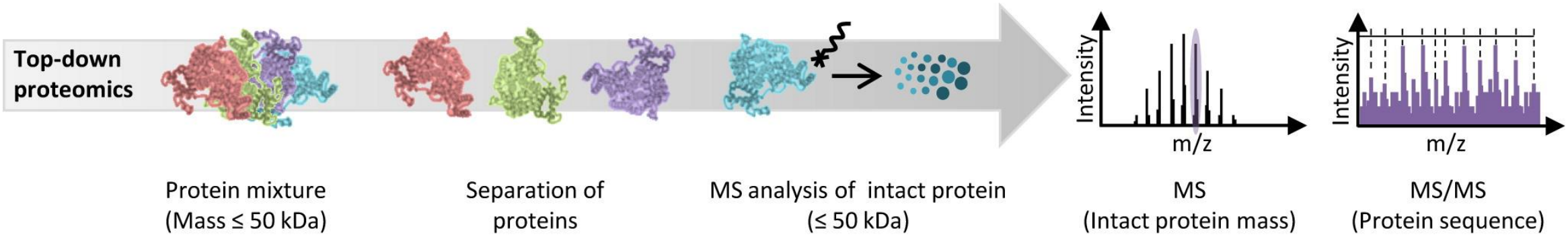




DOI: 10.1016/j.compbiolchem.2015.02.009

DOI: 10.1016/j.jprot.2015.07.006

2.2. Top-down proteomics

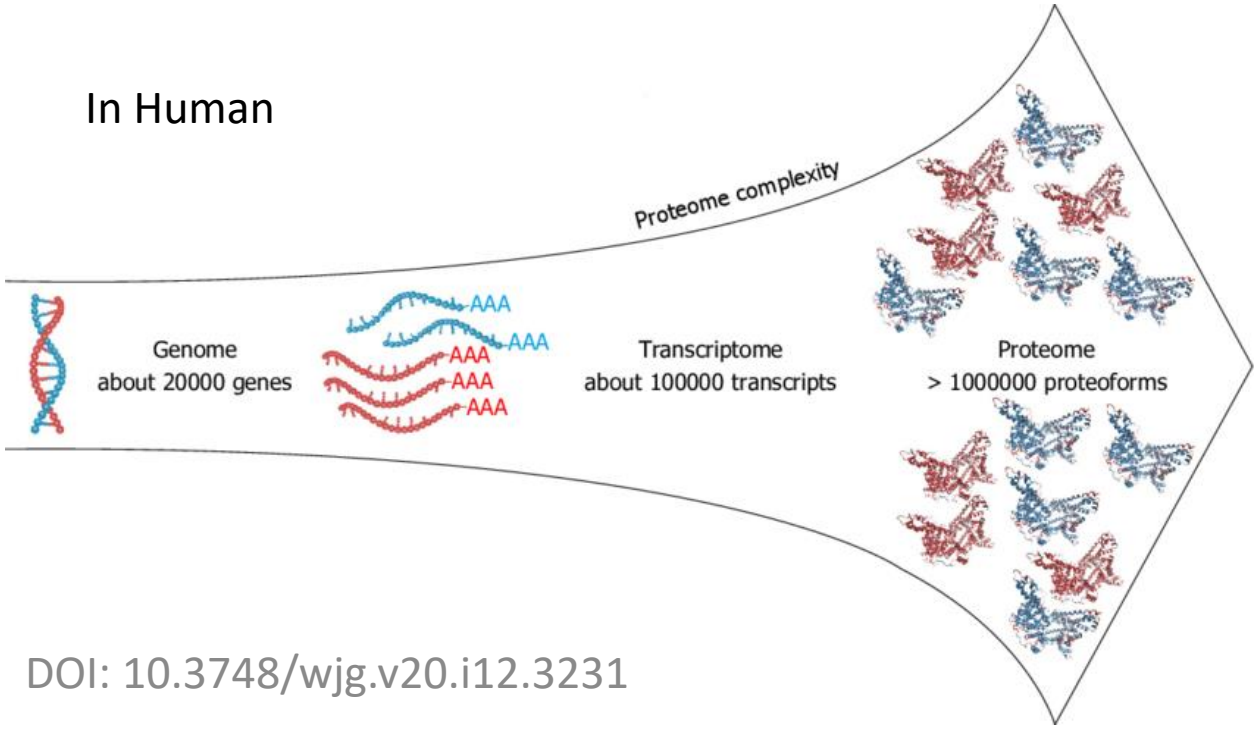


J. Proteome Res., 2013, 12 (3), pp 1067–1077

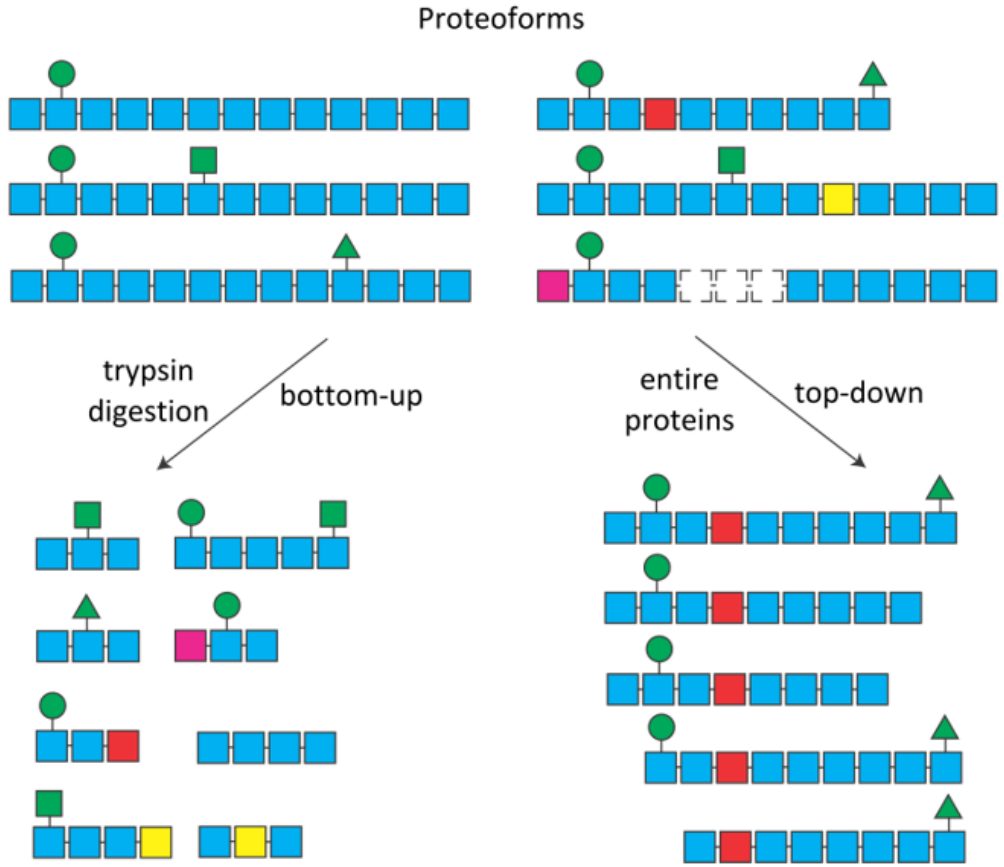
- Proteins are kept intact and directly analyze with MS
- Protein intact and fragment ions masses are measured
- This approach routinely allows for 100% sequence coverage and full characterization of proteoforms

doi: 10.1016/j.bbrc.2014.02.041

Why top-down proteomics?



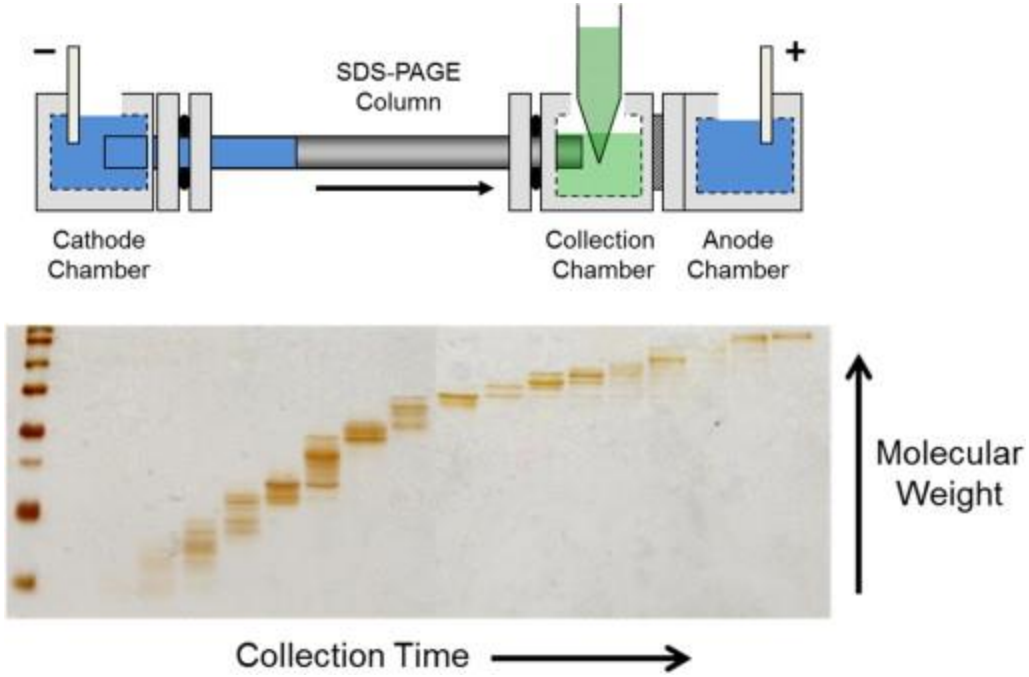
DOI: 10.3748/wjg.v20.i12.3231



http://proteomique.ipbs.fr/front-page/top-down-proteomics/

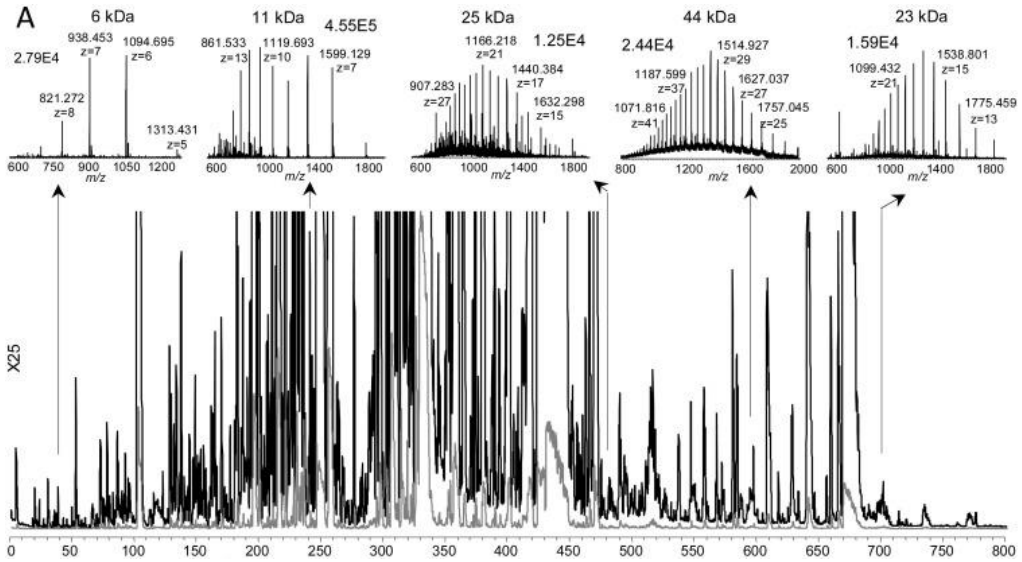
Top-down workflows

- Size-based separation



https://doi.org/10.1016/j.bbrc.2014.02.041

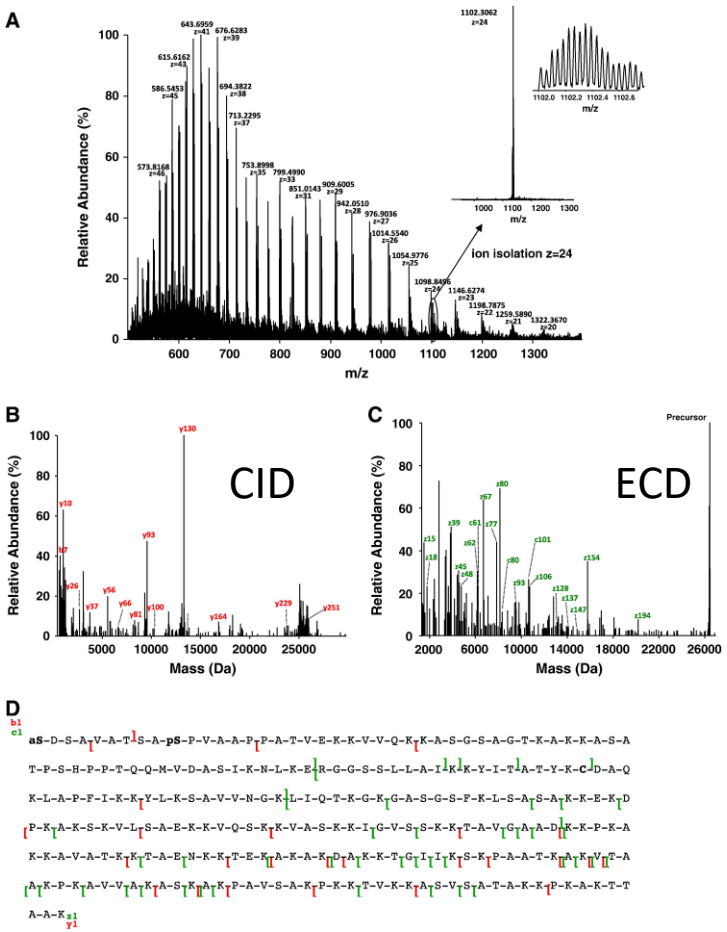
- Reversed-phase liquid chromatography (RPLC), hydrophobic interaction liquid chromatography (HILIC), and ion exchange chromatography (IEX) are three of the most common liquid chromatography approaches applied to intact proteins
- C1-C18-bonded phases had their own limits for eluting various sizes of proteoforms



https://doi.org/10.1016/j.chroma.2017.01.008

MS and fragmentations

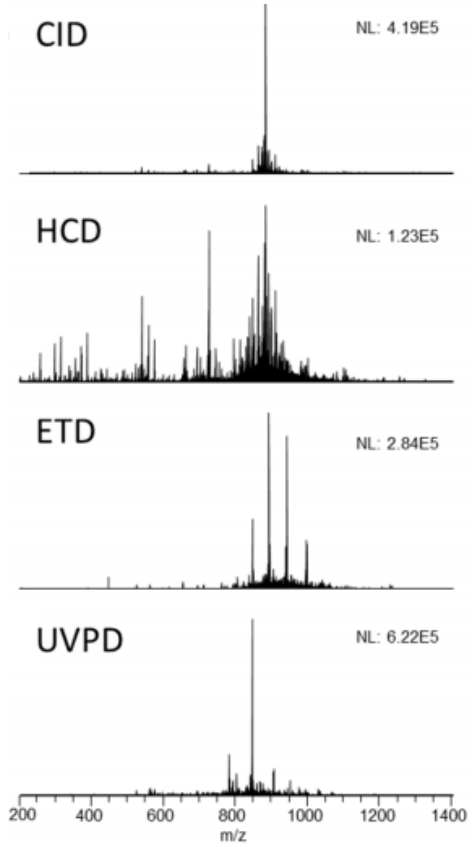
- High resolution mass spectrometry is a must!



https://doi.org/10.1016/j.jprot.2012.05.034

Q2: What mass analyser(s) would you recommend?

- Multiple type of fragmentations (CID often yield selective cleavage of the most labile bonds; electron capture dissociation (ECD) and ETD yield more random and extensive fragmentation; 193 nm ultraviolet photodissociation (UVPD) yields good performances for the characterization of intact proteins)

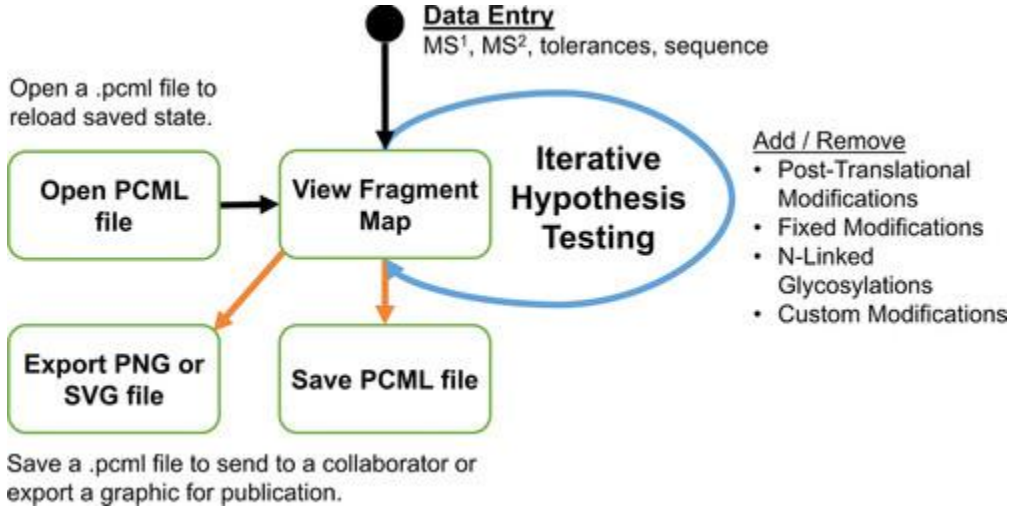


DOI: 10.1021/ja4029654

Data analysis

- Need of dedicated software solution
- The software uses the precursor mass and mass tolerance window to generate a possible list of candidates from a larger annotated database
- The theoretical fragment ions from the candidates are then compared to the experimentally determined fragment ions within a fragment mass tolerance
- A P-score is calculated for each hit, representing the probability that a random sequence could account for the matching ions

https://doi.org/10.1016/j.bbrc.2014.02.041



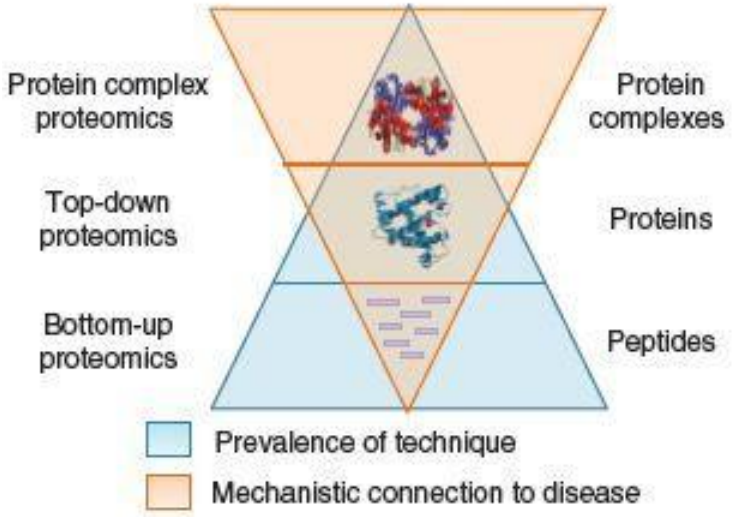
doi: 10.1002/pmic.201570050

Native mass spectrometry

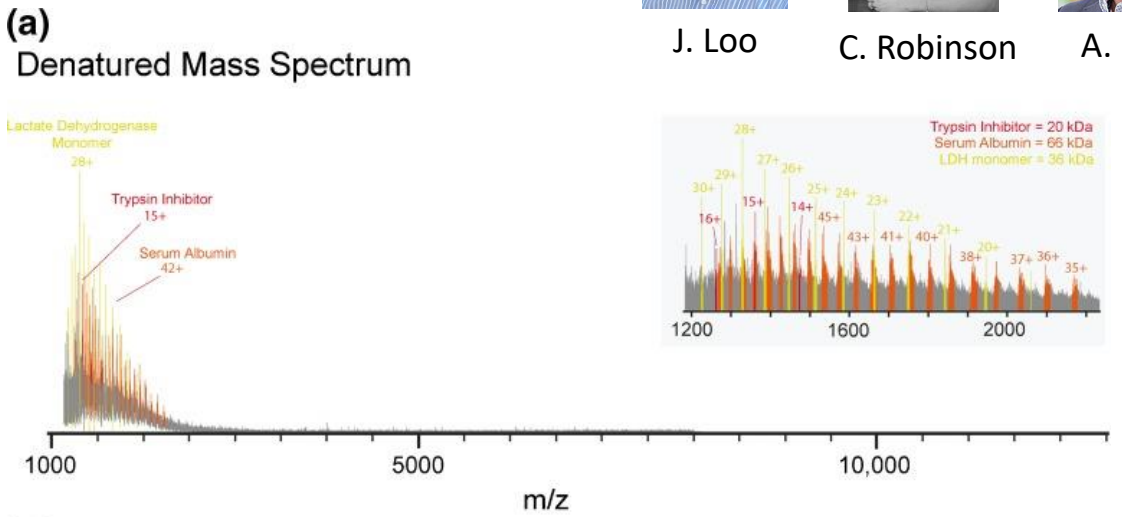

J. Loo

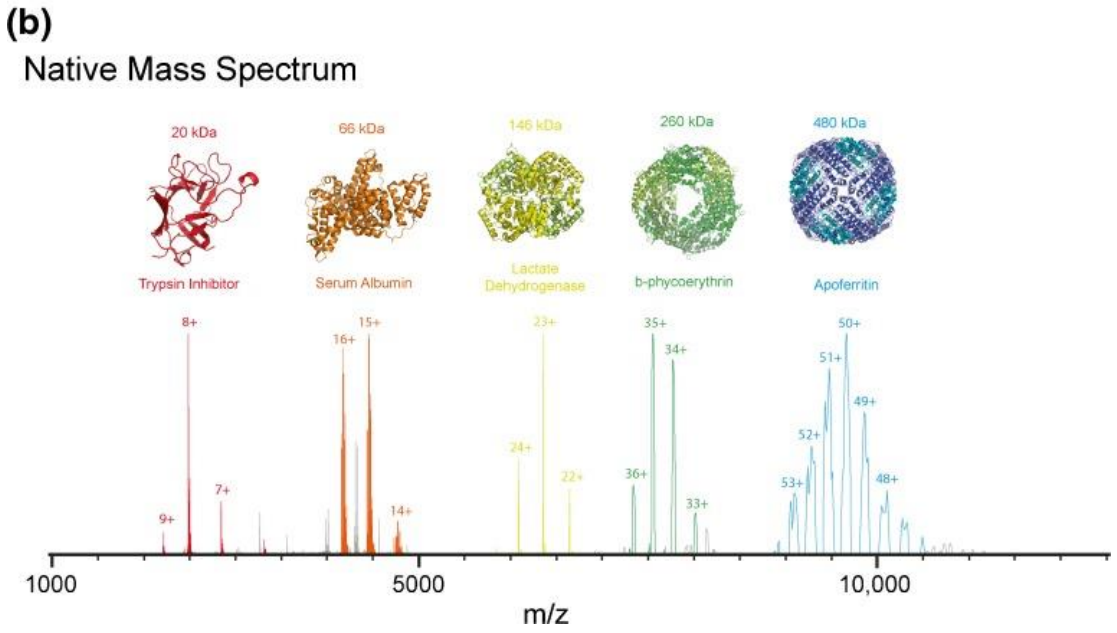

C. Robinson


A. Heck



https://doi.org/10.1186/gm457





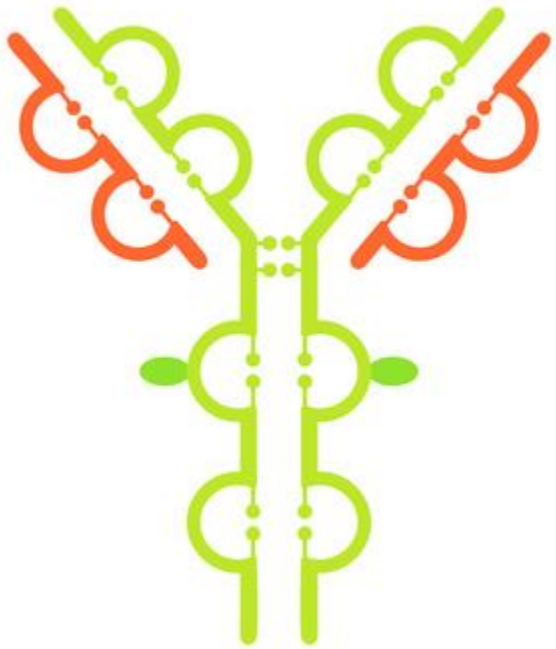
- Many diseases are the result of incorrect protein folding that can hamper the binding of their cofactor and subsequently lead to nonspecific protein aggregation
- Native MS can reveal the composition, stoichiometry, dynamics, stability, and also the spatial arrangement of the subunits of protein assemblies

https://doi.org/10.1016/j.jasms.2009.12.010

doi: 10.1007/s13361-016-1545-3

Applications of top-down proteomics

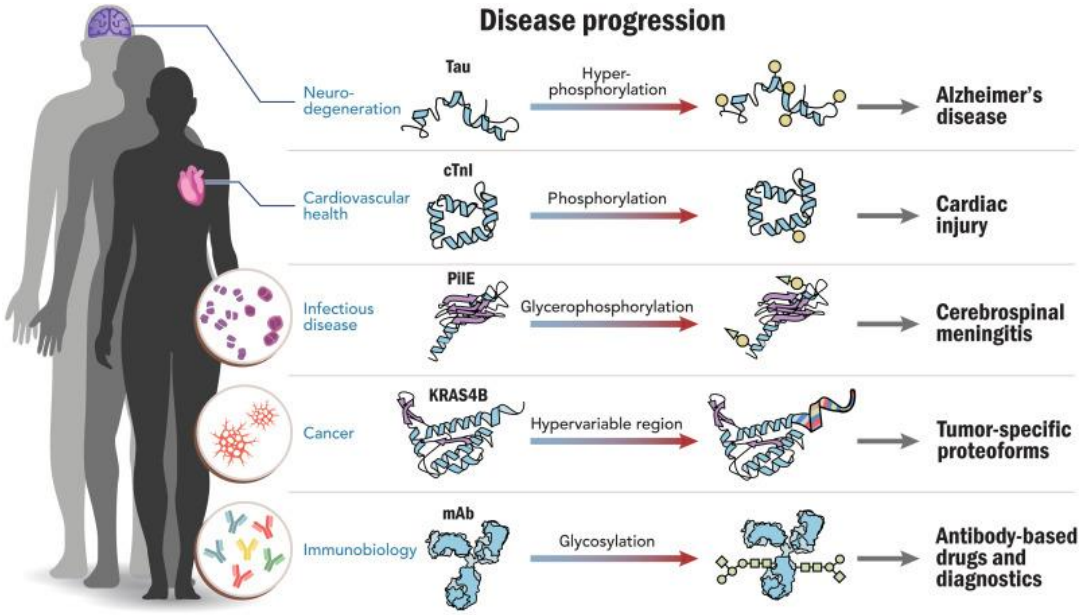
- Assessment of PTMs and sequence variations
- Comprehensive structural characterization of mAbs (therapeutic glycoproteins) to ensure their stringent quality control

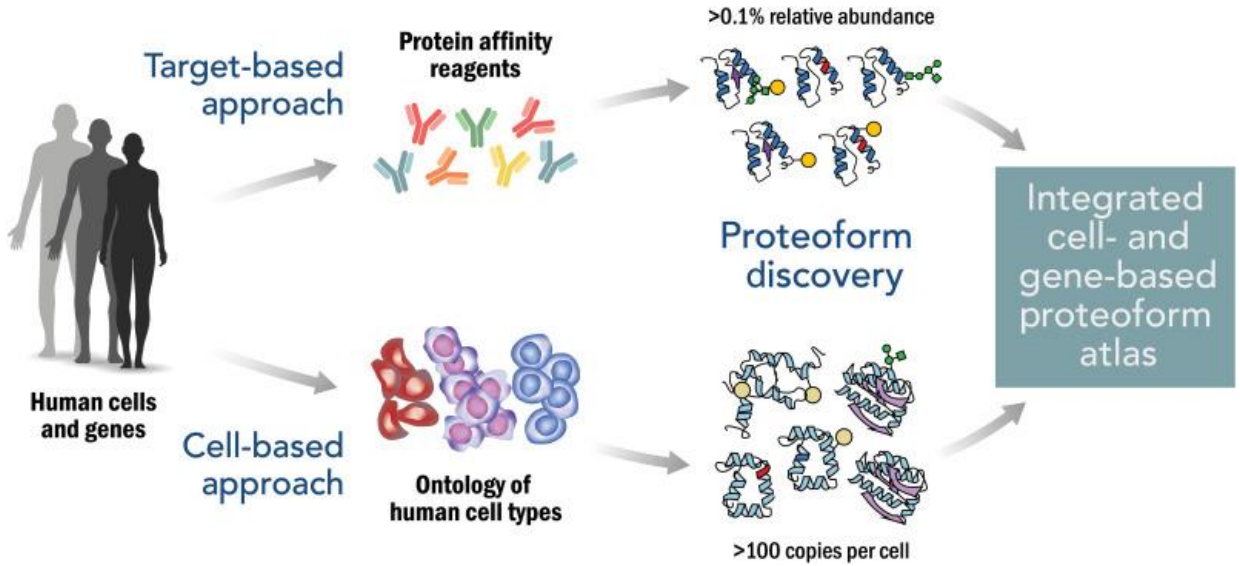


http://www.topdownproteomics.org/

Applications of top-down proteomics

- The Human Proteoform Project





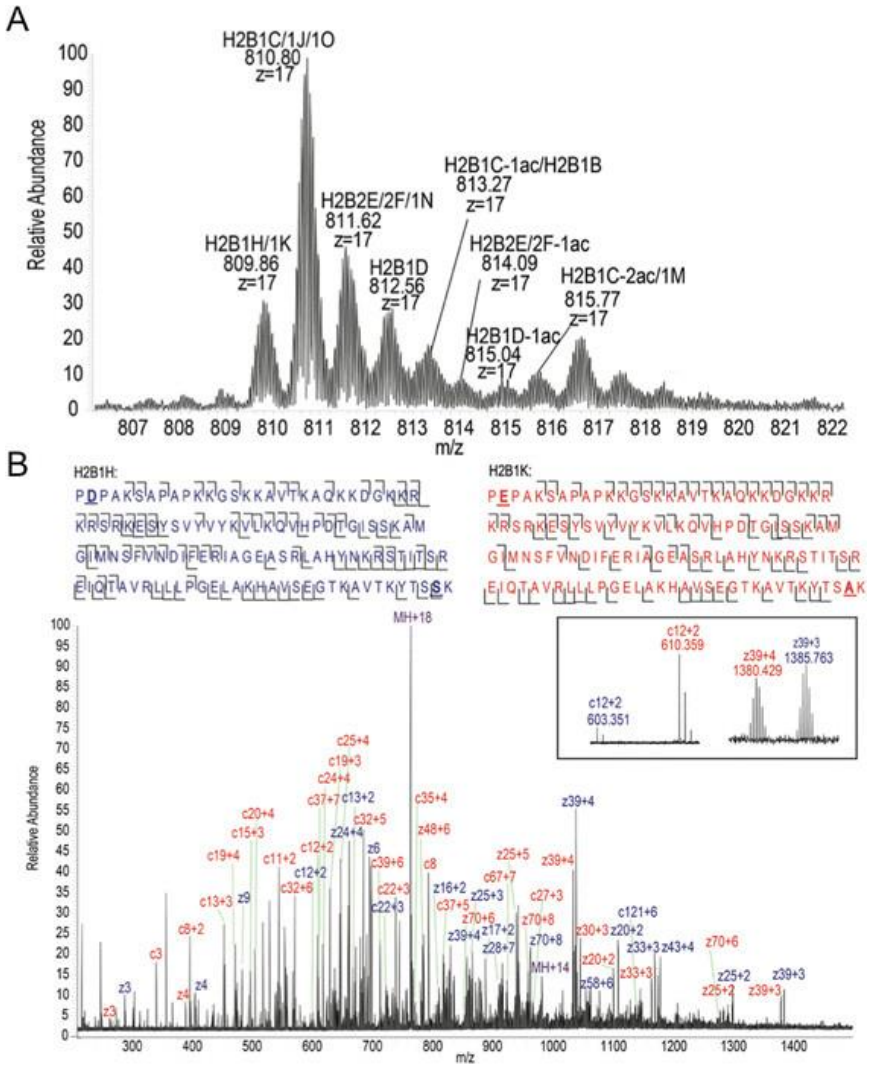
doi: 10.1126/sciadv.abk0734

Challenges moving forward

- Technical difficulty of proteome-wide analysis
- Sample preparation methods
- Protein separation

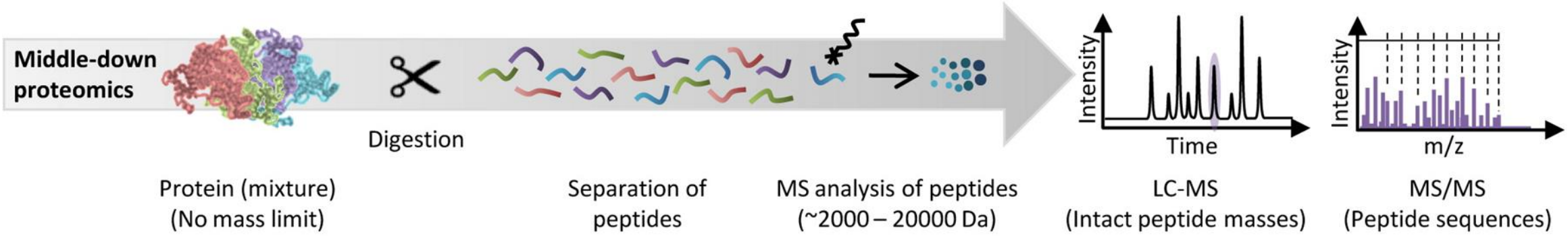
CTDP
Consortium for Top-Down Proteomics

N. Kelleher



DOI: 10.1186/s13072-015-0006-8

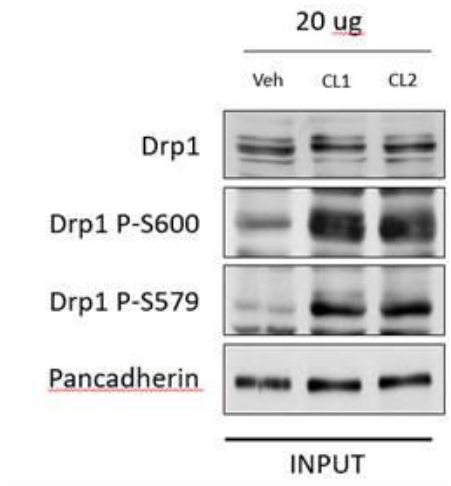
2.3. Middle-down proteomics



J. Proteome Res., 2013, 12 (3), pp 1067–1077

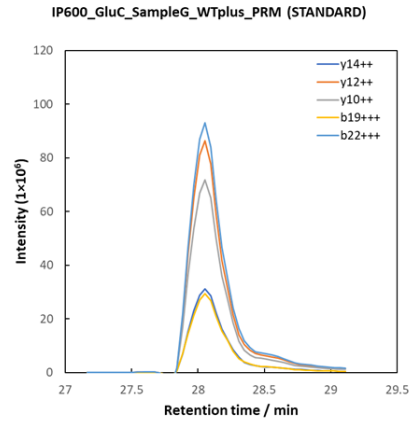
- Middle-range peptides (*i.e.*, 3.0 kDa < *MW* < 10 kDa)
- The middle-down approach (middle) uses a “limited” digest (*e.g.*, Glu-C or Asp-N)
- Good sequence coverage and retention of PTM information

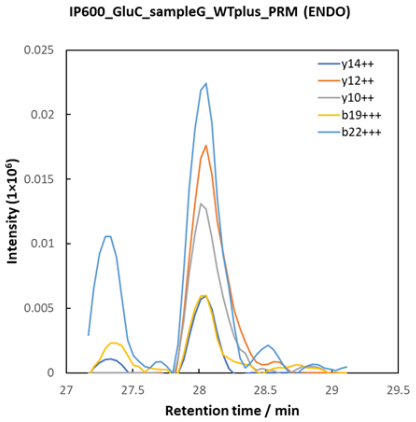
One example in our lab

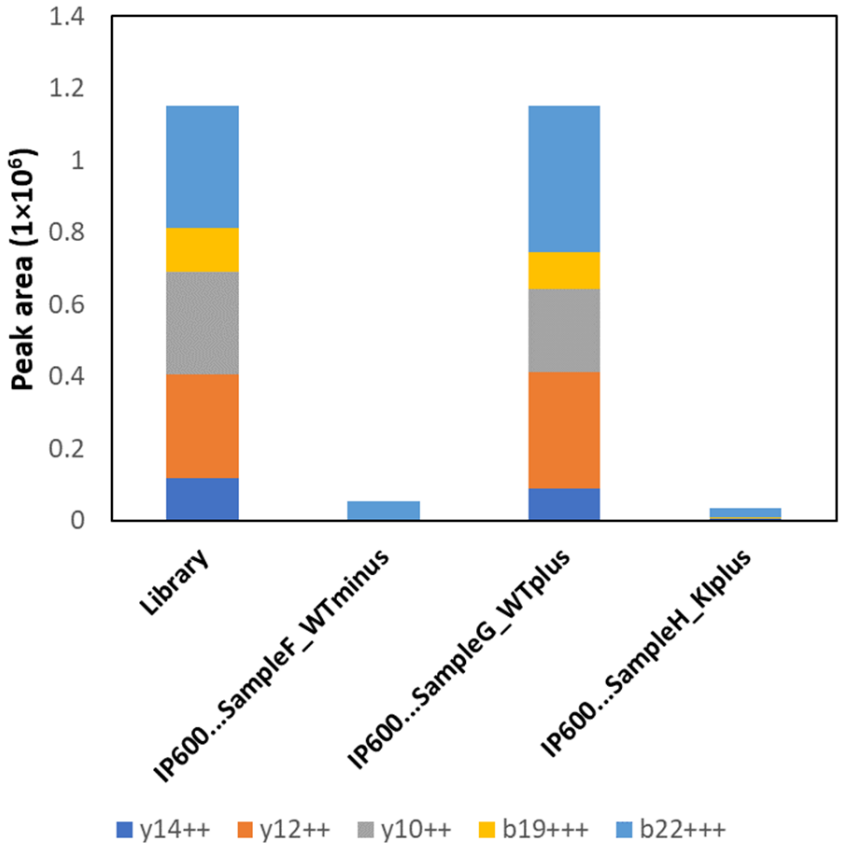


Western blots with phospho-specific antibodies

KSKPIPIMPAS[Phos]PQKGHAVNLLDVPVPVARKLS[Phos]ARE

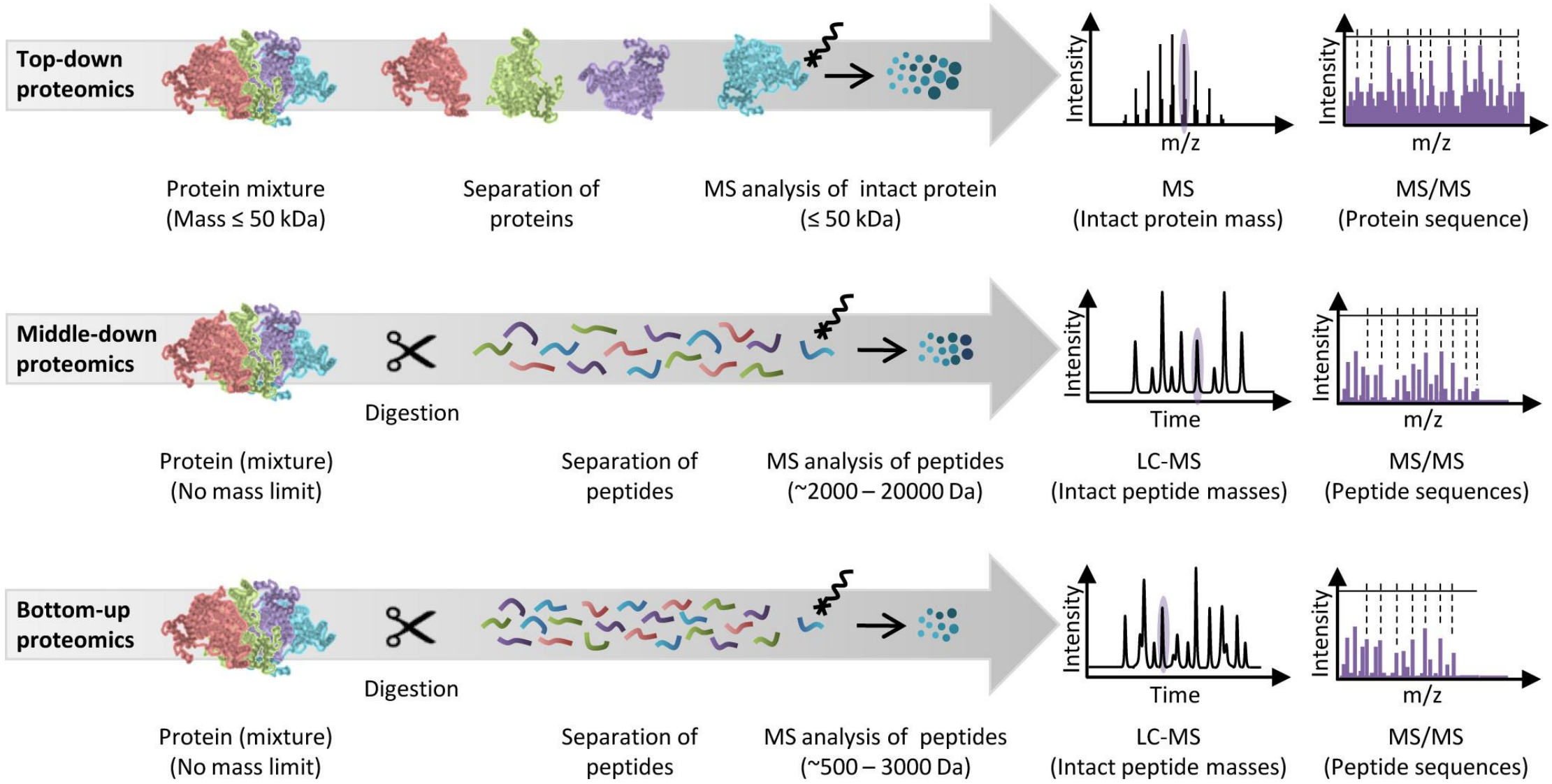






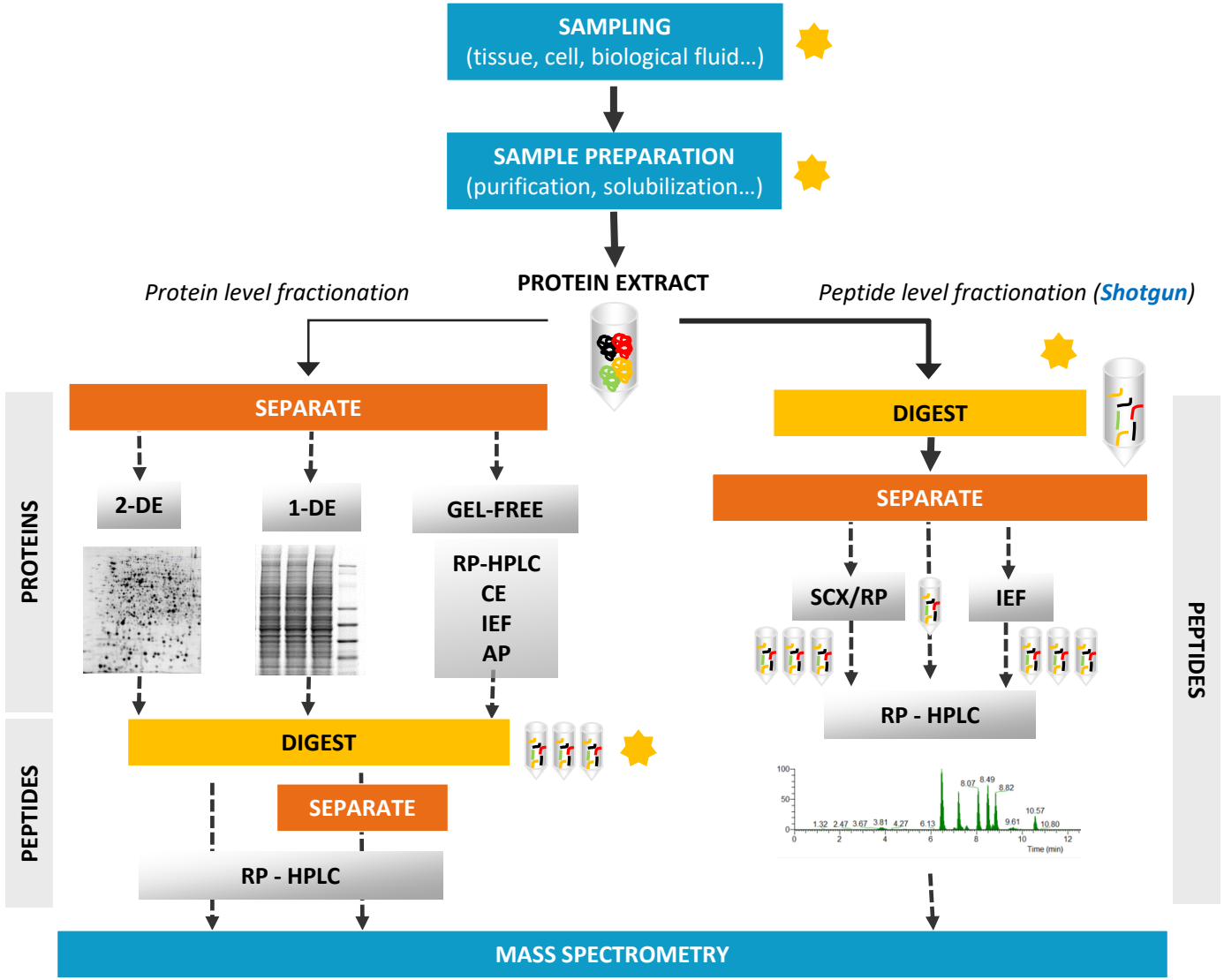
DOI: 10.1016/j.celrep.2021.109565

In summary



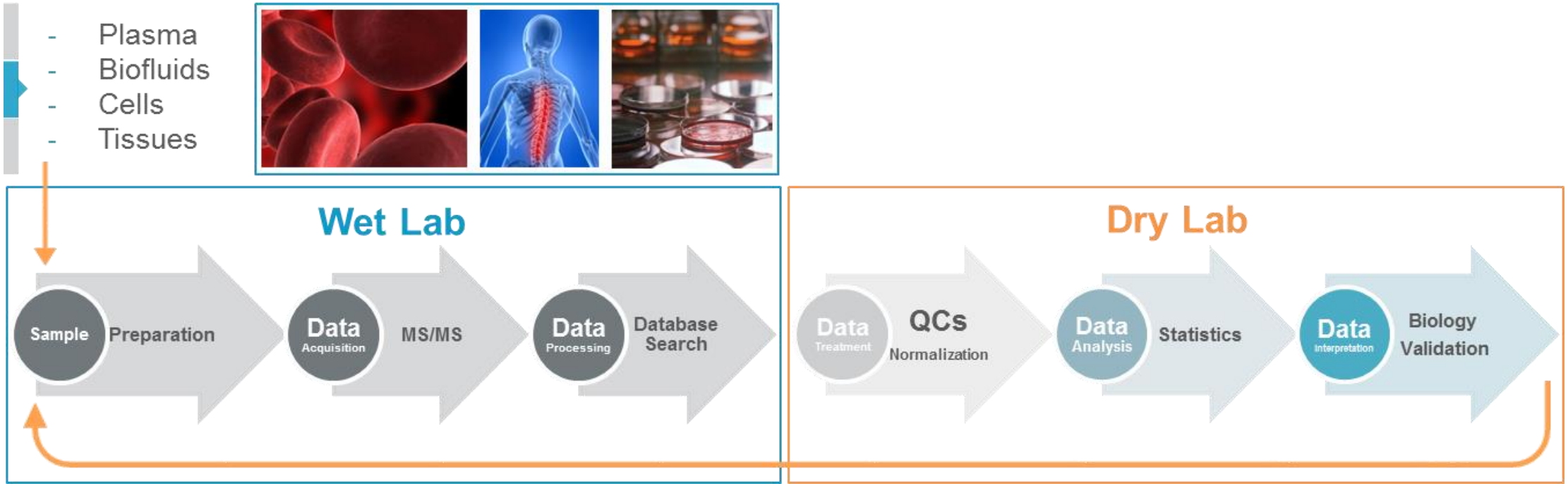
https://doi.org/10.1021/pr301201x

2.4. Sample preparation



M. Abonnenc and M. Mayr, 2012, *Inflammation and Atherosclerosis*, 249-266, DOI: 10.1007/978-3-7091-0338-8_13

The steps of protein and peptide sample preparation

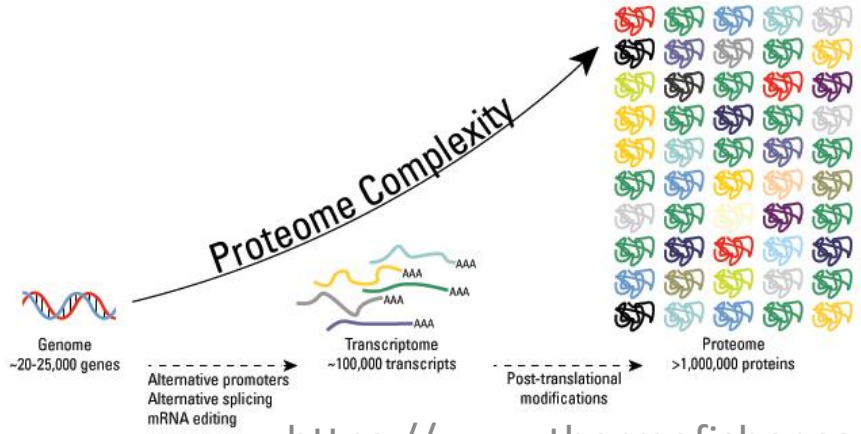


- Sampling
- Extract the proteins
- Protein derivatization
- Digest the proteins
- Peptide derivatization

Sampling and sample complexity

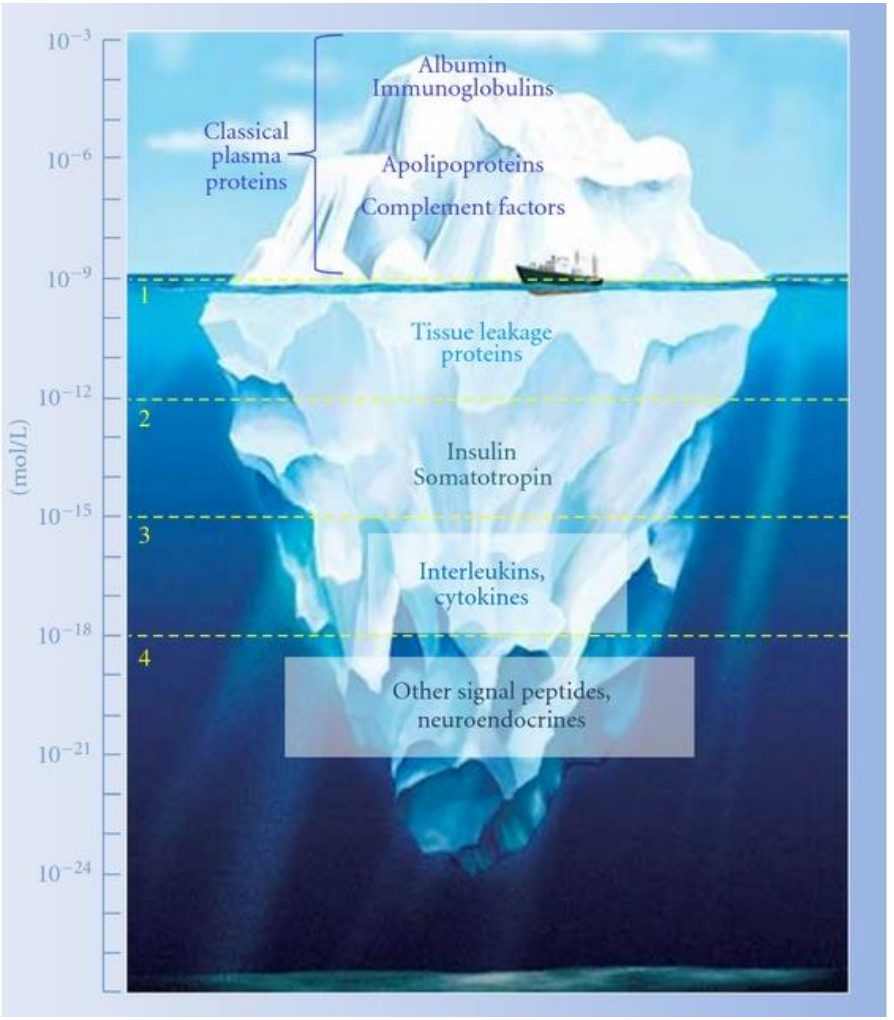
Sampling reproducibility
Sample storage

Samples are complex by number of analytes to measure



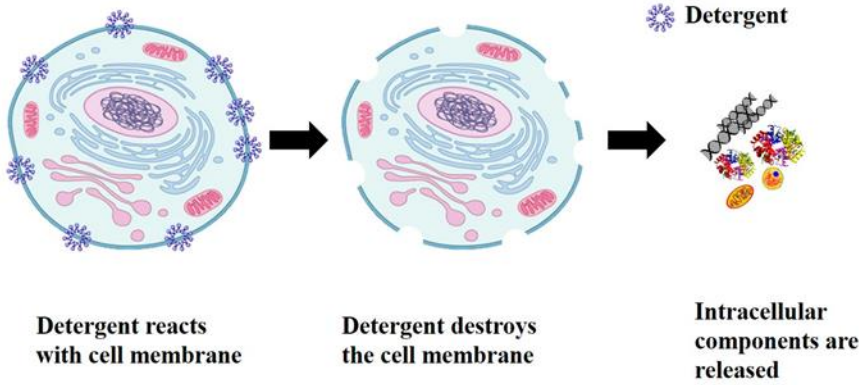
https://www.thermofisher.com

Samples are complex by wide range of abundances

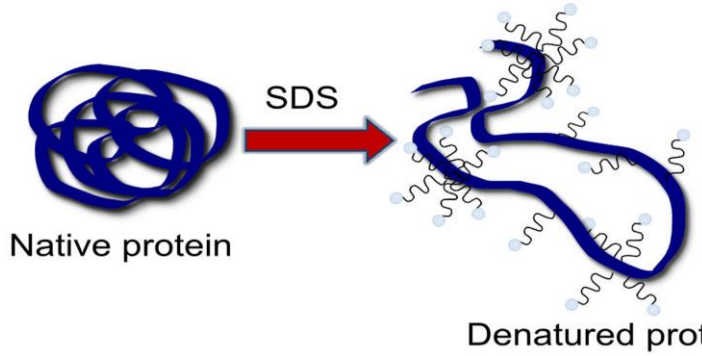


Finoulst *et al., J. Biomed. Biotechnol.*, **2011,** 245291

Lysis, solubilization and denaturation



doi:10.3390/mi8030083



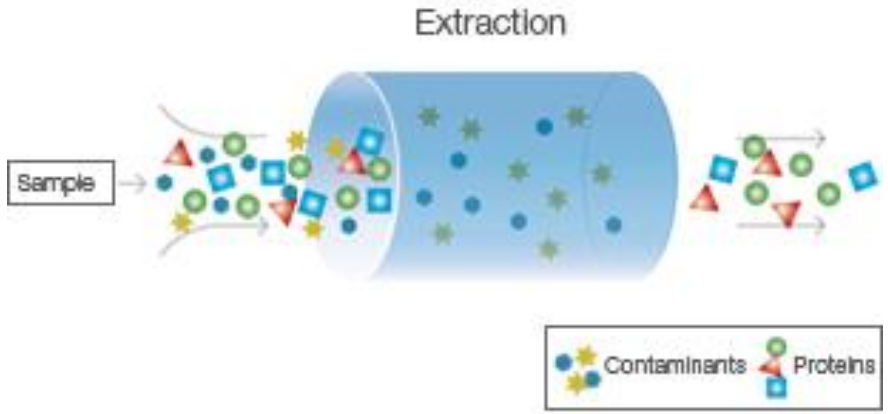
https://doi.org/10.1371/journal.pone.0175838

Detergent Name	Type	Molecular Weight	CMC, mM	Mol. Weight (Micelle)	Suggested Removal
Triton X-100	Nonionic	647	0.24	90,000	TCA/Acetone
NP-40	Nonionic	617	0.29	90,000	Acetone
Tween 20	Nonionic	1228	0.06		Acetone
Tween 80	Nonionic	1310	0.01	76,000	Acetone
Octyl Glucoside	Nonionic	292	23–24	8000	Ethyl acetate
Octyl thioglucoside	Nonionic	308	9		Ethyl Acetate
Big CHAP	Nonionic	878	3–4	8781	Filtration
Deoxycholate	Anionic	415	2–6	2000	Acetone, TCA
Sodium Dodecyl Sulfate	Anionic	288	6–8	17,887	Filtration/FASP
CHAPS	Zwitterionic	615	8–10	6149	Filtration
CHAPSO	Zwitterionic	631	8–10	7000	Filtration

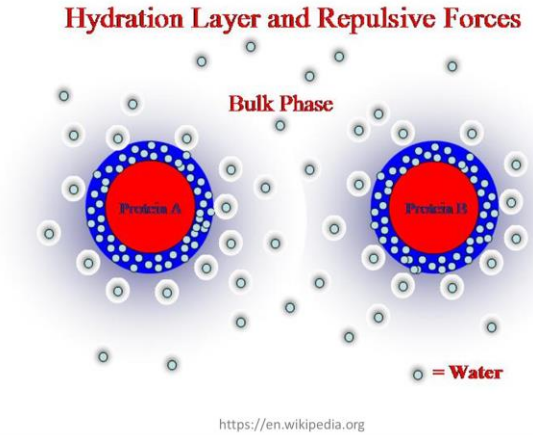
doi: 10.3390/ijms16023537

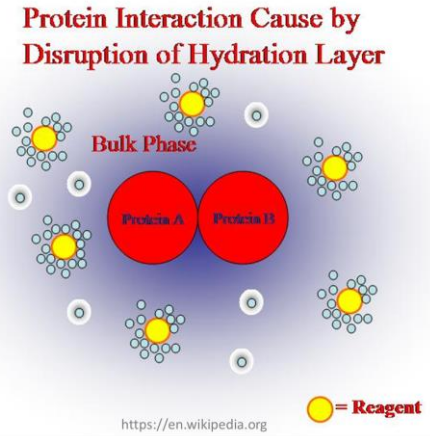
- Cell lysis is frequently the first step and can be accomplished either physically or by using reagents
- Use of different buffers, detergents, salts and reducing agents
- Lysis goes with stabilization to protect extracted proteins from degradation or artifactual modification (*e.g.*, protease and phosphatase inhibitors)

Protein extraction



http://www.bio-rad.com/en-ch/category/protein-extraction



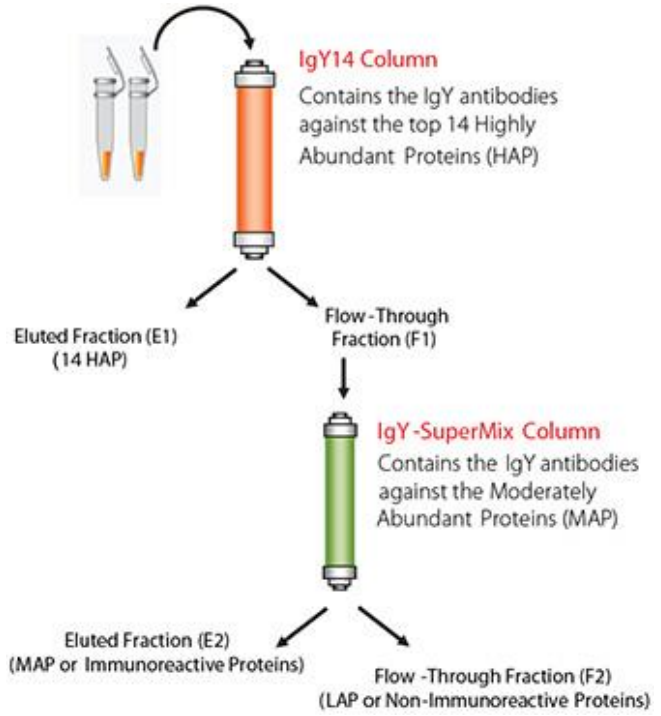


Approach	Description
"Salting out"	Precipitation uses saturation of salt to precipitate protein from solution. Most commonly an ammonium sulfate precipitation, but also uses sodium sulfate.
Ultrafiltration	Centrifugation at high speed using molecular weight cutoff filter to remove contaminants; prominent in Filter-Aided Sample Preparation (FASP).
Polyethyleneimine (PEI)	Cationic polymer precipitates nucleic acids in 1 M NaCl, leaving proteins in the supernatant. PEI must be removed before further analysis.
Isoelectric Point (PI)	The pH of solution is adjusted with mineral acid to the isoelectric point of most proteins (pH 4–6). Neutral proteins will aggregate and precipitate.
Thermal	Cell extracts are denatured using heat; denatured proteins aggregate and precipitate, but stability is enhanced.
Nonionic polymer Polyethylene glycol (PEG)	Concentration of PEG unique to the protein mixture is added. Proteins precipitate based on an excluded volume principle. Centrifugation pellets the precipitated protein. PEG must be removed before mass spectrometry analysis.

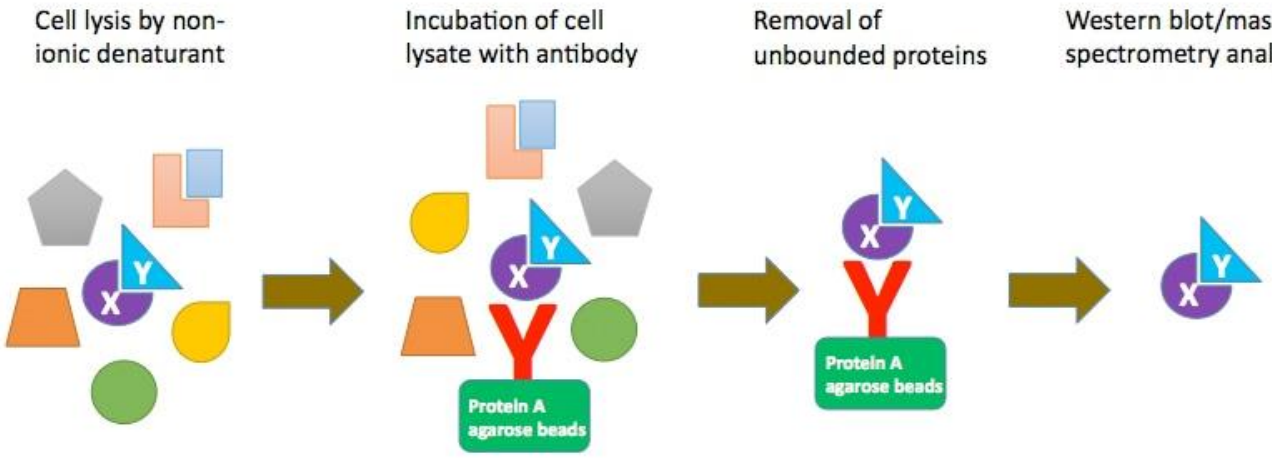
doi: 10.3390/ijms16023537

Depletion and enrichment

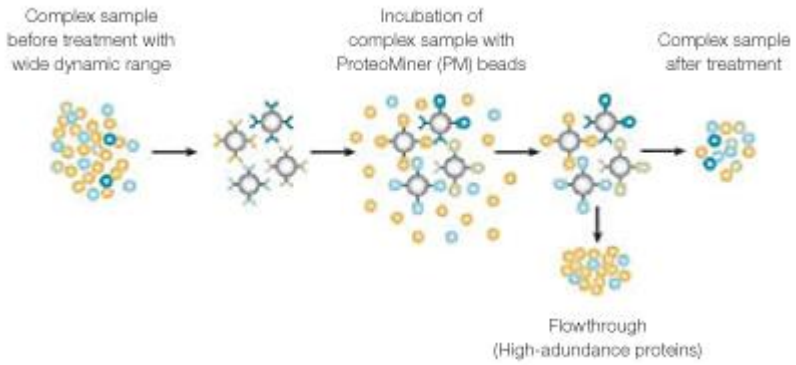
- Depletion and enrichment strategies are often employed to remove high-abundance proteins of no analytical interest and isolate target proteins in the sample



https://www.sigmaaldrich.com/



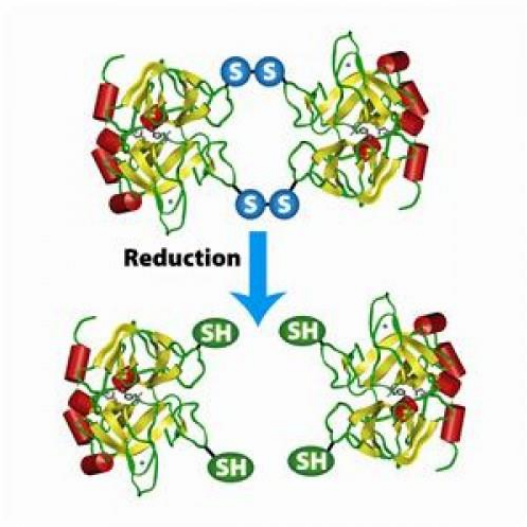
https://www.profacgen.com/Co-Immunoprecipitation-Co-IP.htm

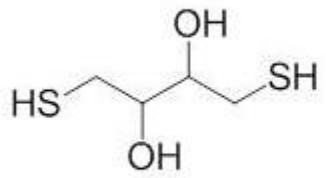


http://www.bioradiations.com/

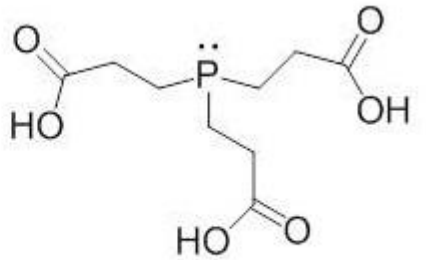
Reduction and alkylation

Reduction step: open the protein





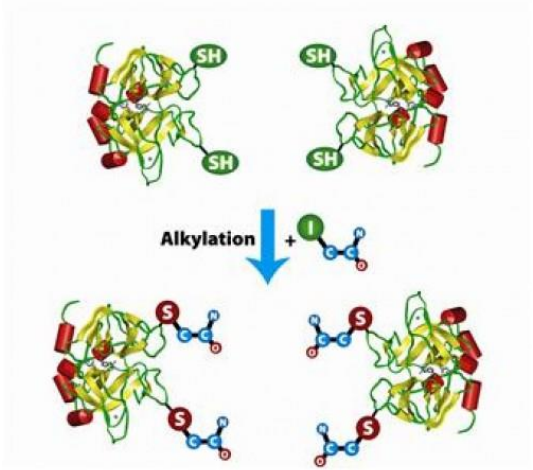
dithiothreitol (DTT)

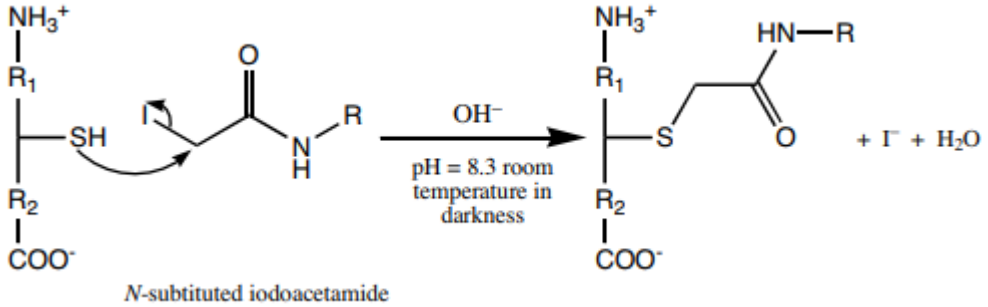


tris(2-carboxyethyl)phosphine (TCEP)

http://sites.psu.edu/msproteomics/2014/05/30/tcep-or-dtt/

Alkylation step: avoid re-oxidation of proteins

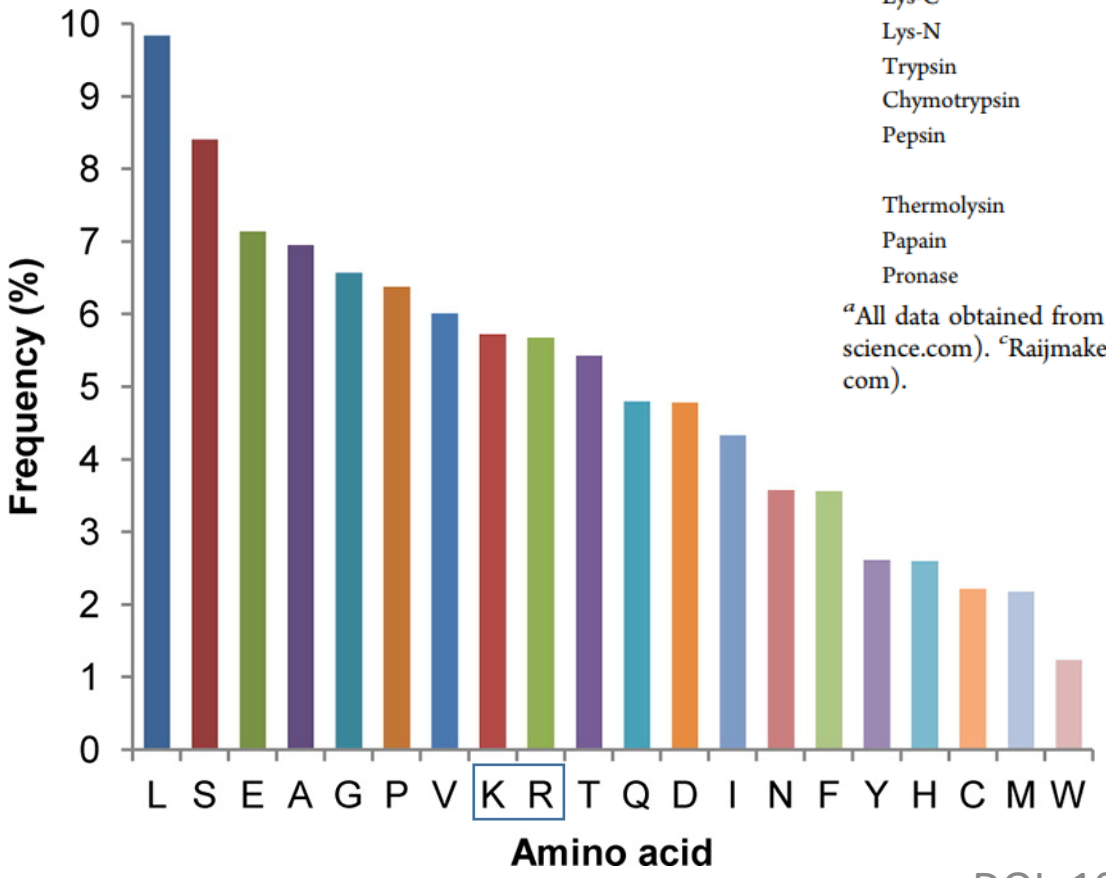




N-subtituted iodoacetamide

https://www.gbiosciences.com

Protein digestion



protease	organism	specificity	pH range	chemical	specificity	pH range
Arg-C	Clostridium histolyticum	R′	7.2–8.0[b]	CNBr	M′	acidic
Asp-N	*Pseudomonas fragi*	‘D	7.0–8.0[b]	HAc	‘D′[d]	acidic
Glu-C	*Staphylococcus aureus*	E′[b]	4.0–7.8[b]	FA	D′	acidic
Lys-C	Lysobacter enzymogenes	K′	8.5–8.8[b]	HCl	D′[e]	2.0[e]
Lys-N	Lysobacter enzymogenes	‘K[c]	8.0[c]	NTCB	‘C[e]	9 −10[f]
Trypsin	*Bos taurus*	K,R′	8.0[b]	Hydroxylamine	N−G	9.0[g]
Chymotrypsin	*Bos taurus*	F,W,Y′	7.0–9.0[b]			
Pepsin	Sus scrofa	‘F,L,W,Y′	1.3			
		‘F,L′	2.0			
Thermolysin	Bacillus thermoproteolyticus	‘A,F,I,L,M,V	8.0[h]			
Papain	Carica papaya	R,K,D,H,G,Y[b]	6.0–7.0[b]			
Pronase	*Streptomyces griseus*	A,E,F,I,L,T,V,W,Y′	6.0–7.5[b]			

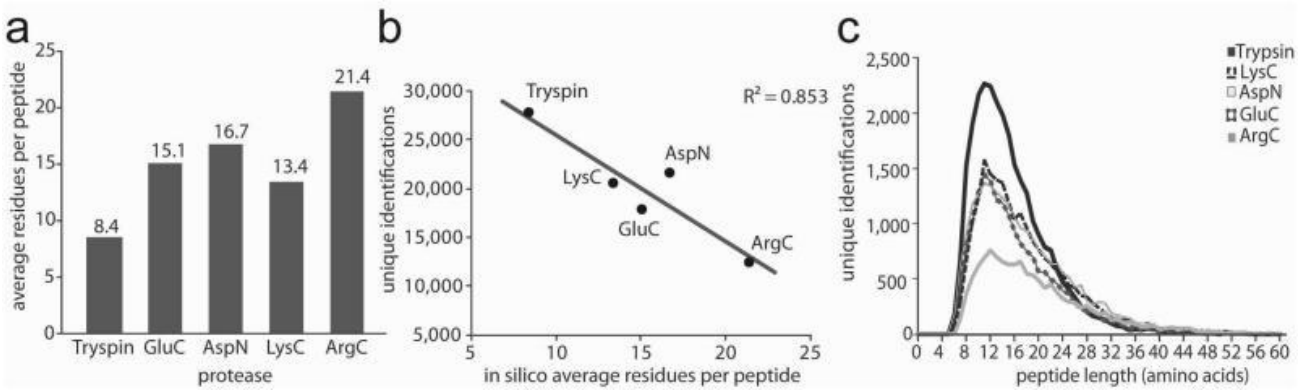
[a]All data obtained from the Expasy bioinformatics resource portal[29] (www.expasy.org), except those noted. [b]Roche Web site (www.roche-applied-science.com). [c]Raijmakers et al.,[30]. [d]Swatkoski et al.,[31]. [e]Smith,[32]. [f]Tang et al.,[33]. [g]Crimmins et al.[34]. [h]Sigma-Aldrich Web site (www.sigma-aldrich.com).

Q3: Why trypsin is so popular in proteomics?

DOI: 10.1021/pr301201x

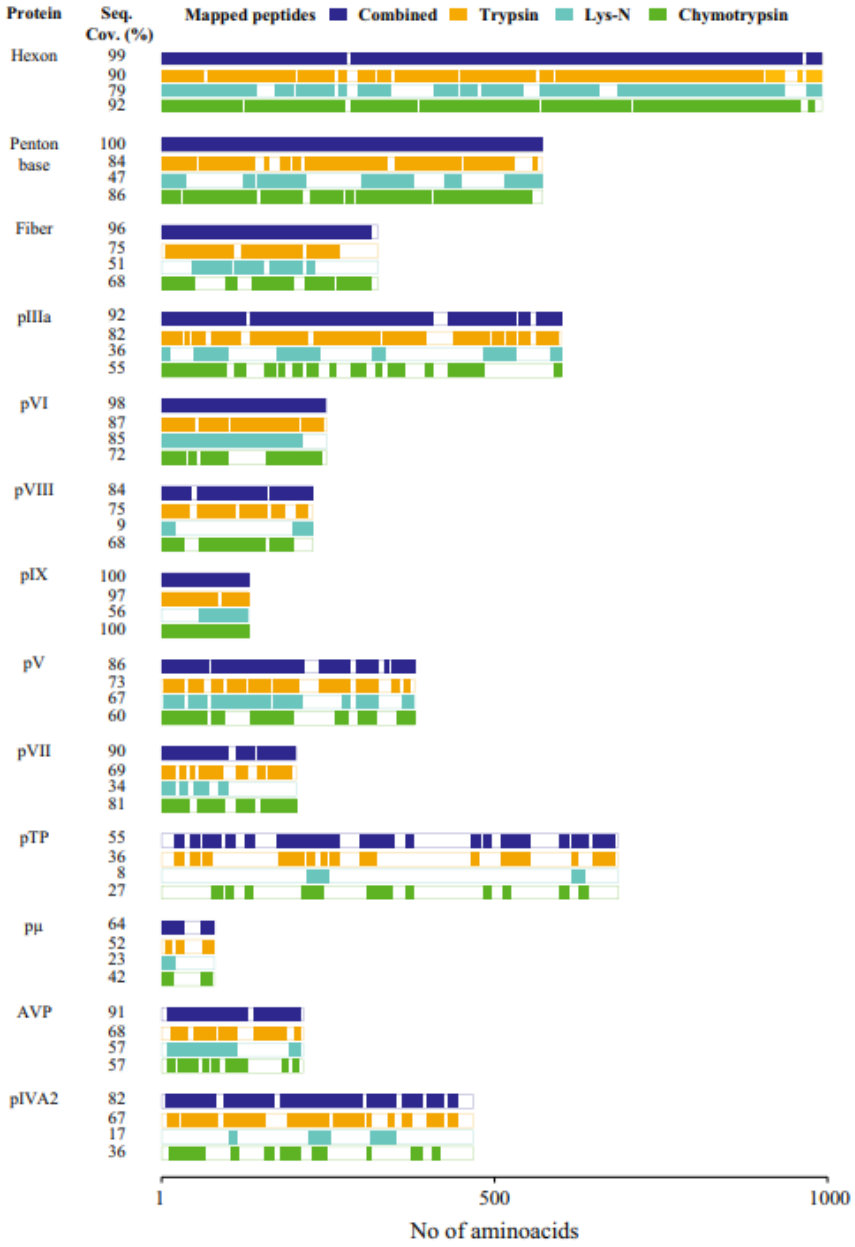
Protein digestion

Q4: Why using different enzymes?



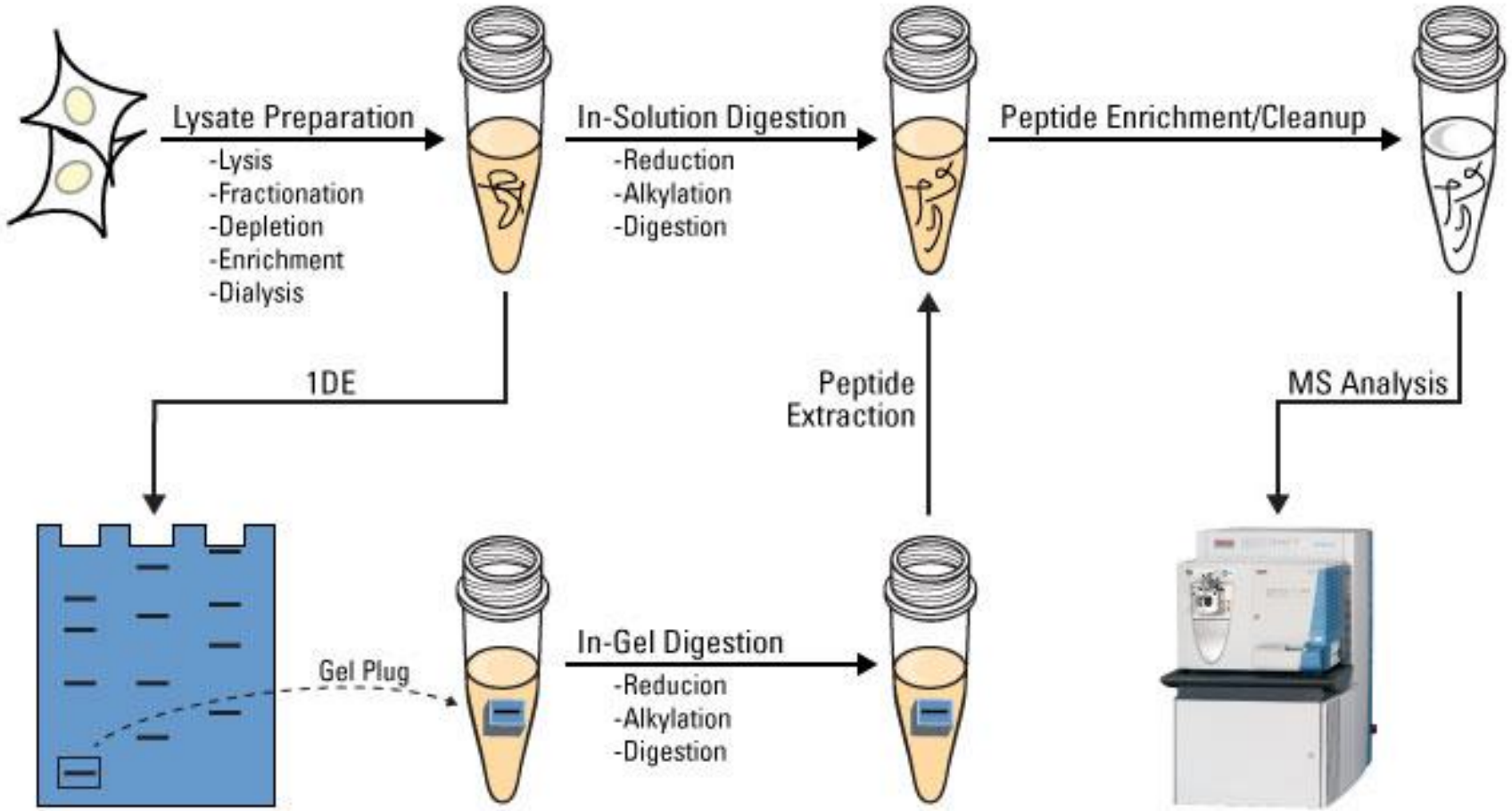
Protease	Trypsin	ArgC	AspN	GluC	LysC	All
Unique peptides	27822	12,452	21,654	17,968	20,619	92,095
CAD	15466	3,518	9,267	7,331	7,807	38,175
ETD	12356	8,934	12,387	10,637	12,812	53,920
Total scans	538,175	540,674	514,607	507,278	524,764	2,625,498
Proteins	3,313	2,708	3,183	2,813	3,030	3,908
Percent of ORFs	56.3	46.0	54.1	47.8	51.5	66.4
Non-redundant amino acids	346,510	191,686	287,188	235,851	304,984	742,312
Non-redundant amino acid proteome coverage (percent)	11.9	6.6	9.8	8.1	10.5	25.5
Average protein sequence coverage (percent)	24.5	18.6	21.5	20.9	24.3	43.4

doi: 10.1021/pr900863u



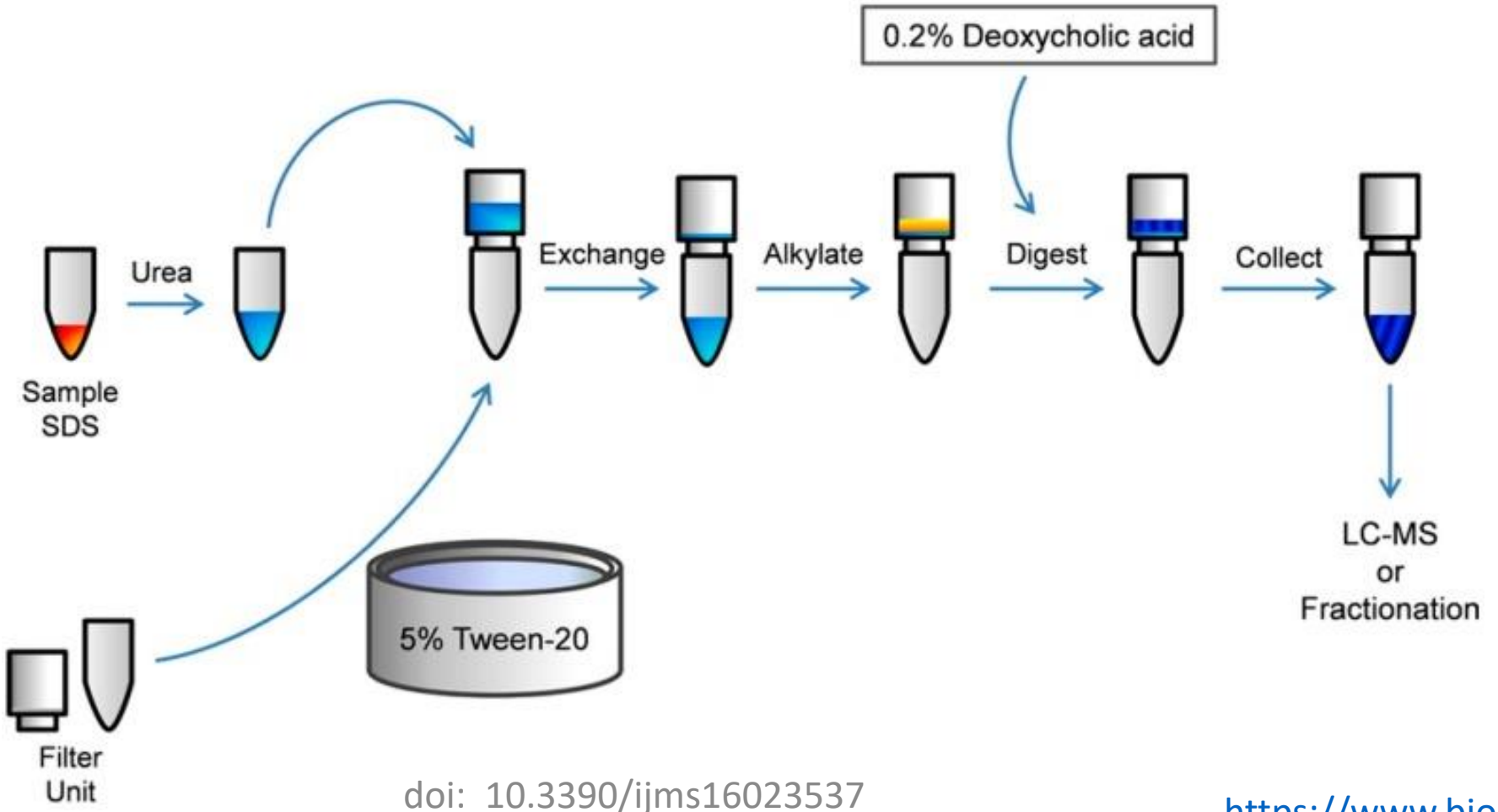
doi:10.1111/febs.13287

In-solution or in-gel digestion



https://www.thermofisher.com

Filter aided sample preparation (FASP)



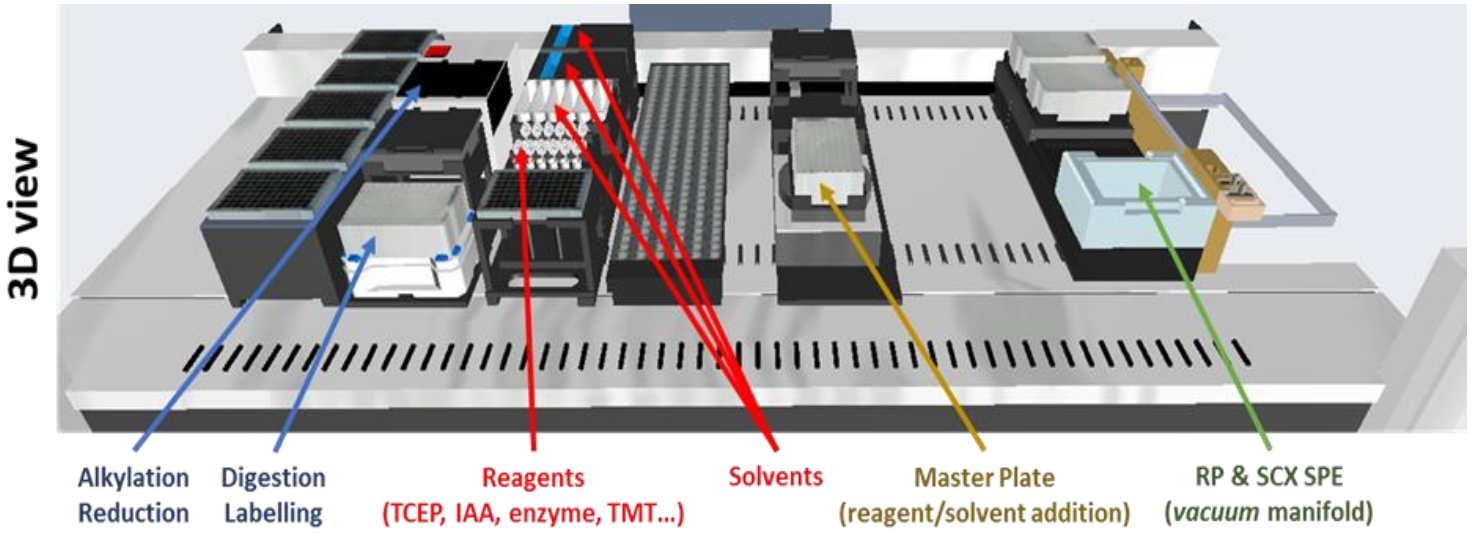
doi: 10.3390/ijms16023537

https://www.biochem.mpg.de/226356/FASP

Automated proteomic sample preparation

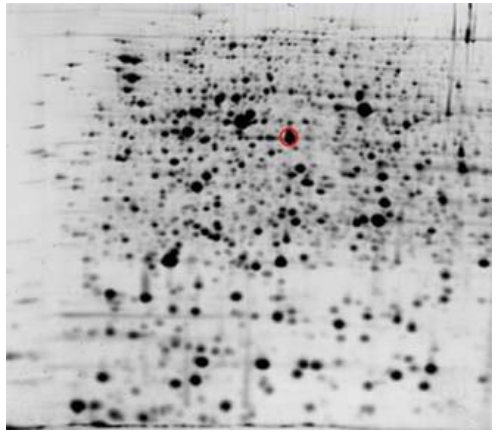
- Reduction/alkylation/digestion
- Labeling
- Purification

Q5: Why automation is valuable?

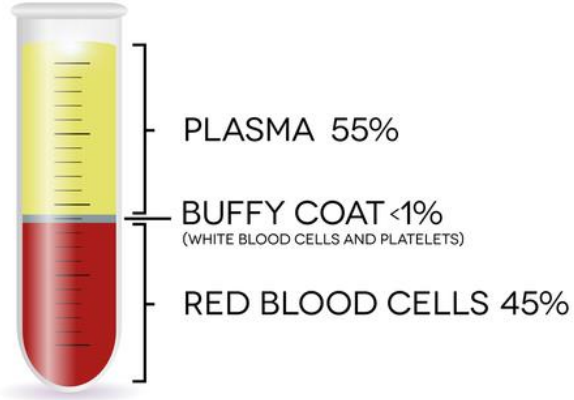


Dayon *et al., Methods in Molecular Biology*

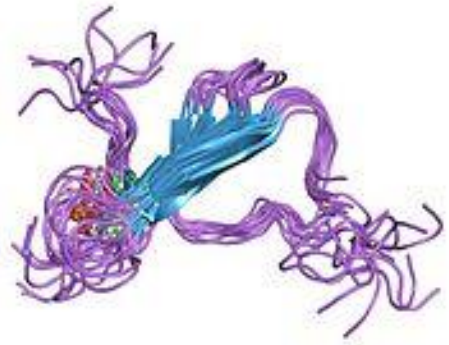
Q6: What procedure(s) would you follow?



http://masse-spec.fr/proteomique



https://www.thermofisher.com



https://en.wikipedia.org/wiki/Tau_protein

Summary of the MS-based proteomic strategies

- Bottom-up and shotgun proteomics
- Top-down proteomics
- Middle-down proteomics

- Procedures to prepare samples for bottom-up, top-down, and middle-down proteomics (as well as peptidomics)

But now:

- You will need to further separate or fractionate your complex samples
- You will need not only to identify but also to quantify your proteins with MS