Proteomics

- Introduction -

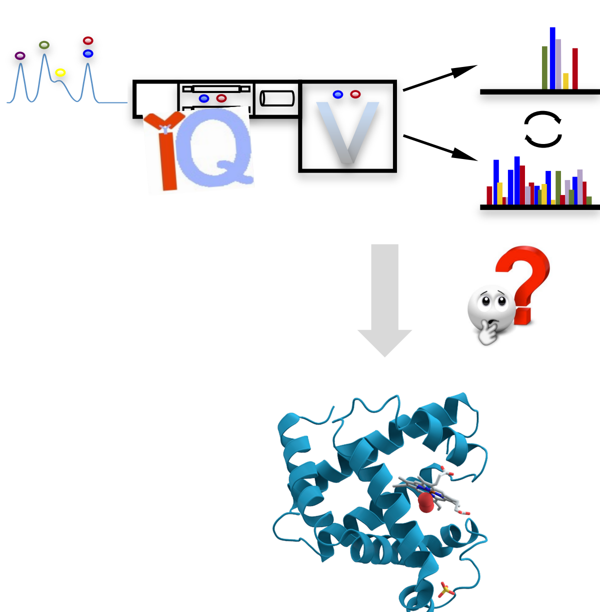
Ute Distler

2022/03/10

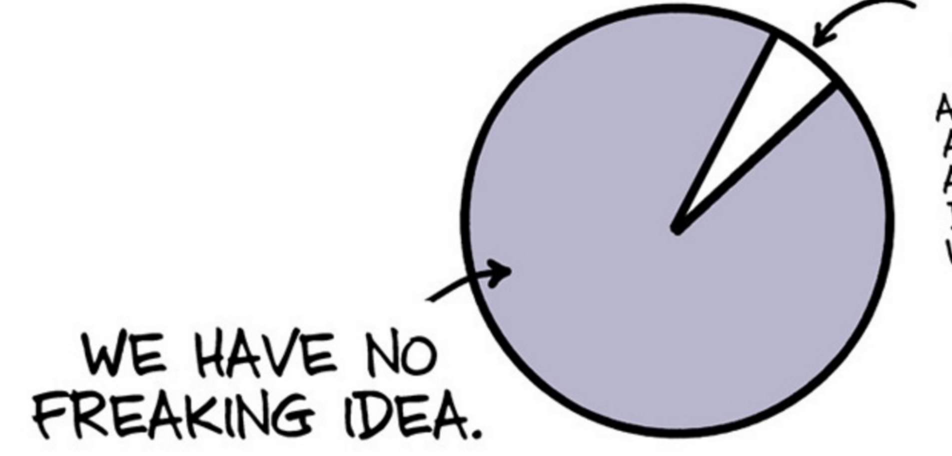
Introduction to Proteomics

- What is proteomics? And why do we do this?

- Mass spectrometry-based proteomics
  - Sample preparation
  - Protein Identification – „from mass spectrum to protein“
  - Acquisition modes



Current state-of-the-art!!!



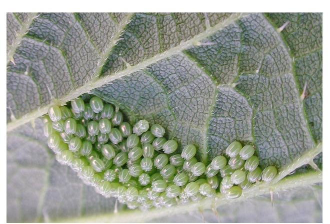
http://phdcomics.com/noidea/

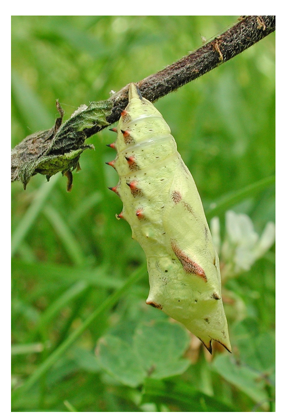
Why proteomics?

DNA: what could be

RNA: what it is trying to be

Protein: what it is



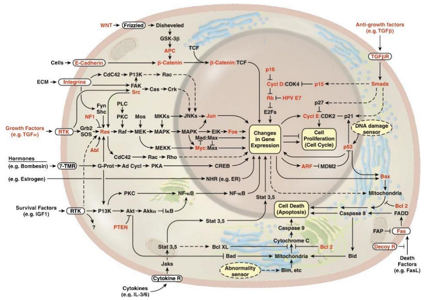


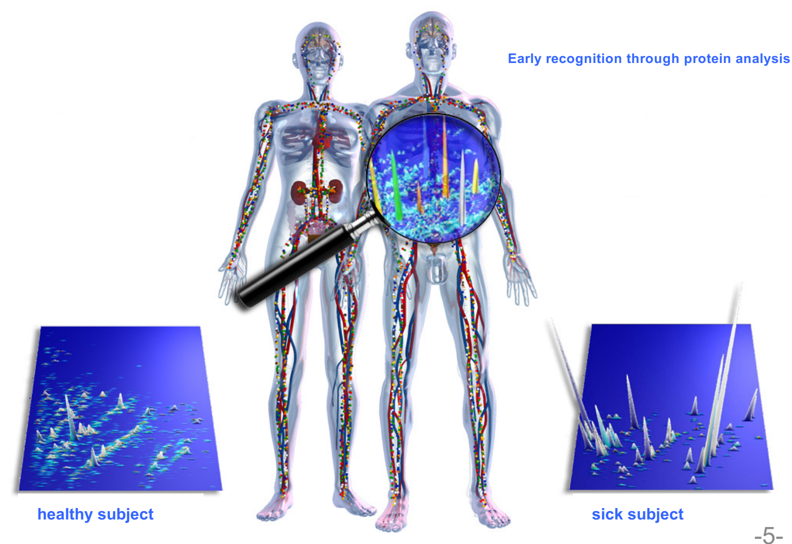
→ larva and adult butterfly: same genome … different proteome

Picture source - http://www.bund-nrw-naturschutzstiftung.de

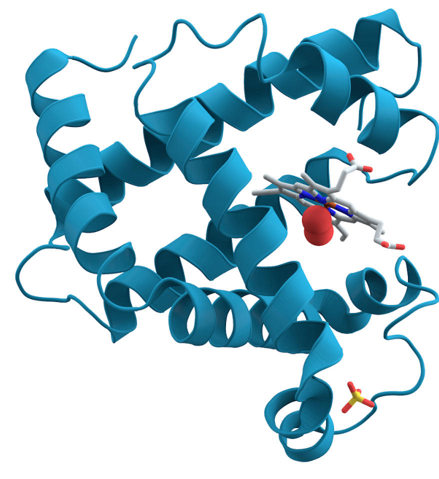
Main Applications of Proteomics

- Identification of proteins whose presence or absence correlates with disease (e.g., cancer)
- Identification of proteins as diagnostic markers or targets for the development of therapeutics
- Elucidation of biological mechanisms of action
- Identification of proteins in signaling pathways
- Detection of drug side effects





How do we define "proteome“/„proteomics“?



>sp|P02144|MYG_HUMAN Myoglobin
OS=Homo sapiens GN=MB PE=1 SV=2
MGLSDGEWQLVLNVWGKVEADIPGHG
QEVLIRLFKGHPETLEKFDKFKHLKSED
EMKASEDLKKHGATVLTALGGILKKKG
HHEAEIKPLAQSHATKHKIPVKYLEFISE
CIIQVLQSKHPGDFGADAQGAMNKALE
LFRKDMASNYKELGFQG

Proteome:

- The **PROTE**in complement of the gen**OME**
  (Mark Wilkins, 1994)
- The entirety of all proteins in a cell (compartment), tissue, or organism (under defined conditions and at a specific time point)

Proteomics:

- The study of proteomes
  (Mark Wilkins, Denis Hochstrasser, Ron Appel, 1996)
- The large-scale study of the structure and functions of proteins (including protein modifications, protein expression, the influence of proteins on metabolic processes, protein-protein interactions,…)

How do we define "proteome"/„proteomics"?

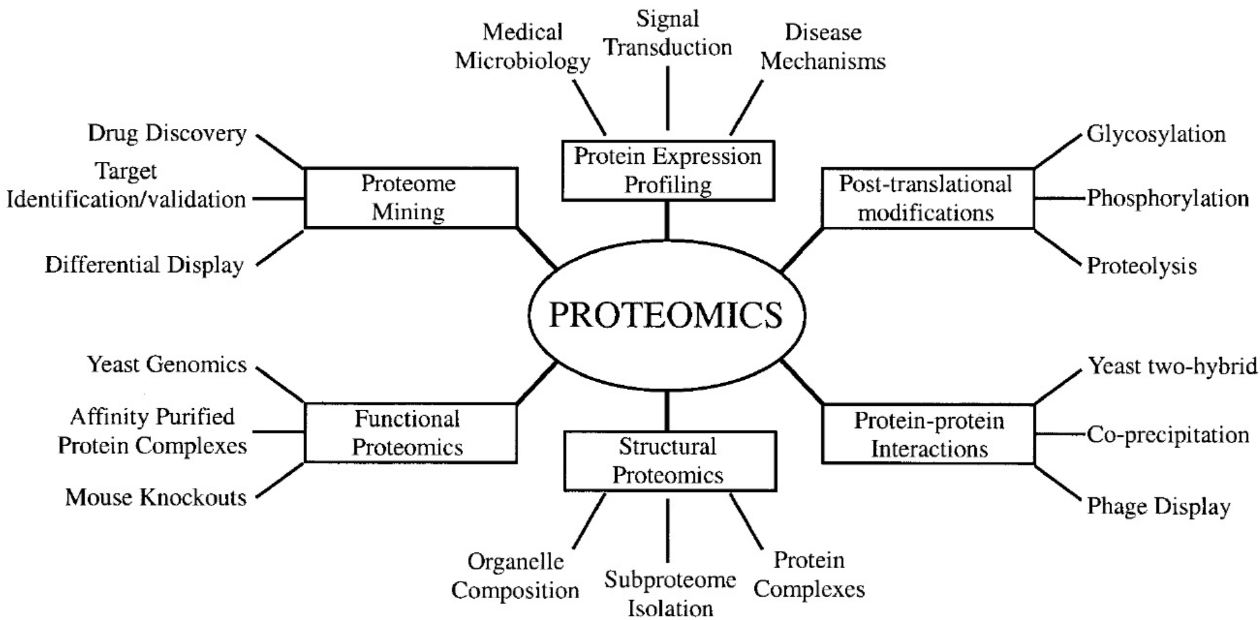


FIG. 1. Types of proteomics and their applications to biology.

Graves and Haystead, Microbiology and Molecular Biology Reviews, **66**, 39–63 (2002); doi: 10.1128/MMBR.66.1.39–63.2002

Deciphering the human proteome

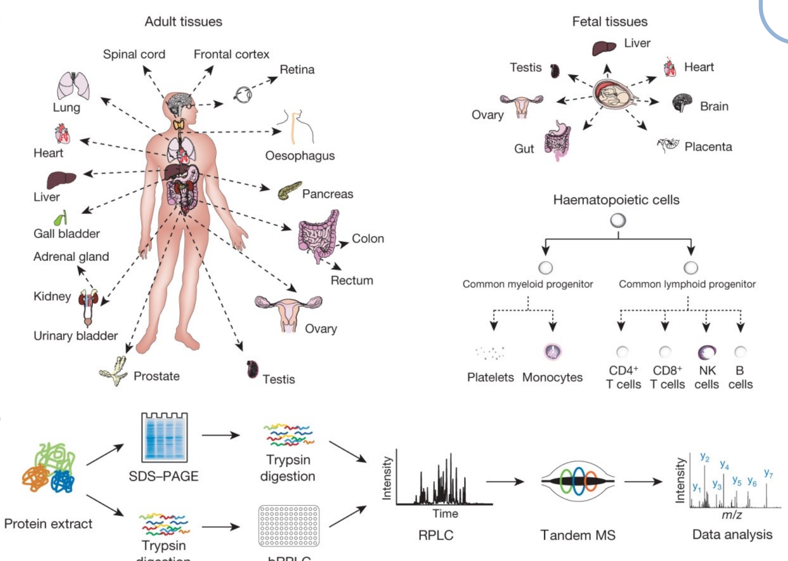
- a milestone for mass spectrometry-based protein analysis-

ARTICLE

doi:10.1038/nature13302

A draft map of the human proteome

Min-Sik Kim[1,2], Sneha M. Pinto[3], Derese Getnet[1,4], Raja Sekhar Nirujogi[3], Srikanth S. Manda[3], Raghothama Chaerkady[1,2], Anil K. Madugundu[3], Dhanashree S. Kelkar[3], Ruth Isserlin[5], Shobhit Jain[5], Joji K. Thomas[3], Babylakshmi Muthusamy[3], Pamela Leal-Rojas[1,6], Praveen Kumar[3], Nandini A. Sahasrabuddhe[3], Lavanya Balakrishnan[3], Jayshree Advani[3], Bijesh George[3], Santosh Renuse[3], Lakshmi Dhevi N. Selvan[3], Arun H. Patil[3], Vishalakshi Nanjappa[3], Aneesha Radhakrishnan[3], Samarjeet Prasad[1], Tejaswini Subbannayya[3], Rajesh Raju[3], Manish Kumar[3], Sreelakshmi K. Sreenivasamurthy[3], Arivusudar Marimuthu[3], Gajanan J. Sathe[3], Sandip Chavan[3], Keshava K. Datta[3], Yashwanth Subbannayya[3], Apeksha Sahu[3], Soujanya D. Yelamanchi[3], Savita Jayaram[3], Pavithra Rajagopalan[3], Jyoti Sharma[3], Krishna R. Murthy[3], Nazia Syed[3], Renu Goel[3], Aafaque A. Khan[3], Sartaj Ahmad[3], Gourav Dey[3], Keshav Mudgal[7], Aditi Chatterjee[3], Tai-Chung Huang[1], Jun Zhong[1], Xinyan Wu[1,2], Patrick G. Shaw[1], Donald Freed[1], Muhammad S. Zahari[2], Kanchan K. Mukherjee[8], Subramanian Shankar[9], Anita Mahadevan[10,11], Henry Lam[12], Christopher J. Mitchell[1], Susarla Krishna Shankar[10,11], Parthasarathy Satishchandra[13], John T. Schroeder[14], Ravi Sirdeshmukh[3], Anirban Maitra[15,16], Steven D. Leach[1,17], Charles G. Drake[16,18], Marc K. Halushka[15], T. S. Keshava Prasad[3], Ralph H. Hruban[15,16], Candace L. Kerr[19]†, Gary D. Bader[5], Christine A. Iacobuzio-Donahue[15,16,17], Harsha Gowda[3] & Akhilesh Pandey[1,2,3,4,15,16,20]



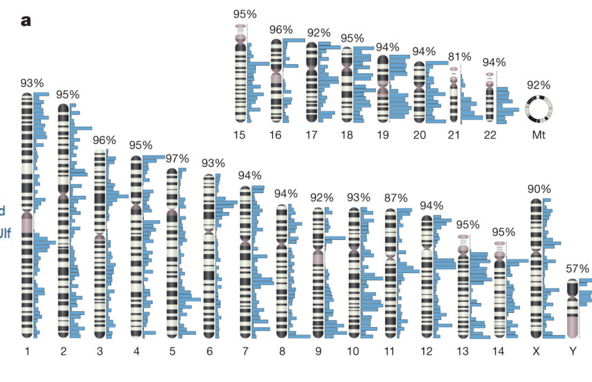
M-S Kim *et al. Nature* **509**, 575-581 (2014) doi:10.1038/nature13302

- First mass spectrometry-based drafts of the human proteome in 2014
- First mass spectrometric detection of gene products for a total of 84% (*Kim et al.*) and 92% (*Wilhelm et al.*) of the annotated protein-coding genes in the human genome

Published: 28 May 2014

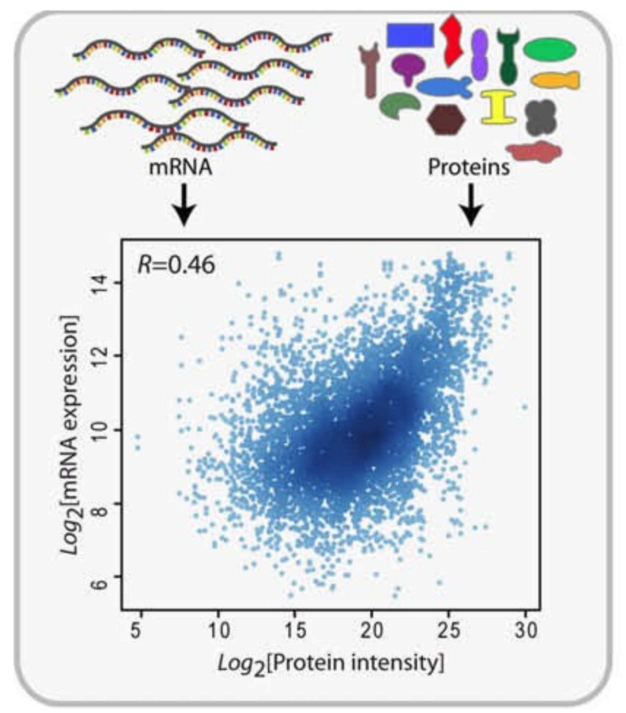
Mass-spectrometry-based draft of the human proteome

Mathias Wilhelm, Judith Schlegl, Hannes Hahne, Amin Moghaddas Gholami, Marcus Lieberenz, Mikhail M. Savitski, Emanuel Ziegler, Lars Butzmann, Siegfried Gessulat, Harald Marx, Toby Mathieson, Simone Lemeer, Karsten Schnatbaum, Ulf Reimer, Holger Wenschuh, Martin Mollenhauer, Julia Slotta-Huspenina, Joos-Hendrik Boese, Marcus Bantscheff, Anja Gerstmair, Franz Faerber & Bernhard Kuster



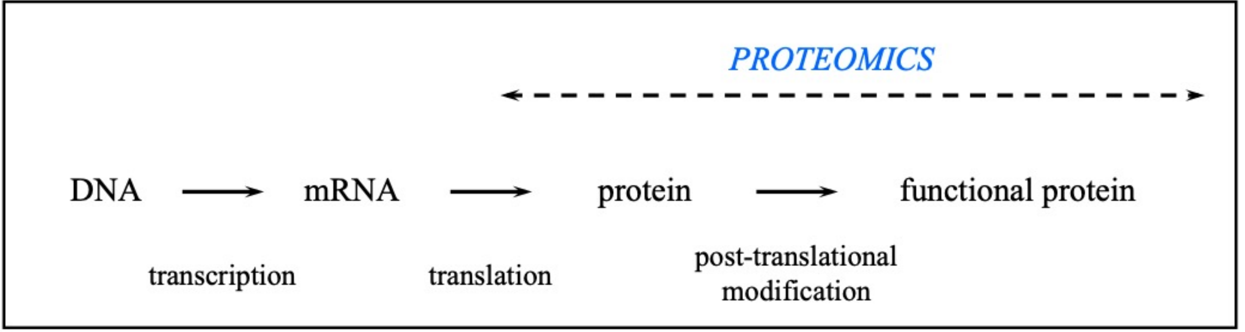
M Wilhelm *et al. Nature* **509**, 582-587 (2014) doi:10.1038/nature13319

Why not simply analyze the transcriptome?



[Tenzer]

The route from genome to proteome



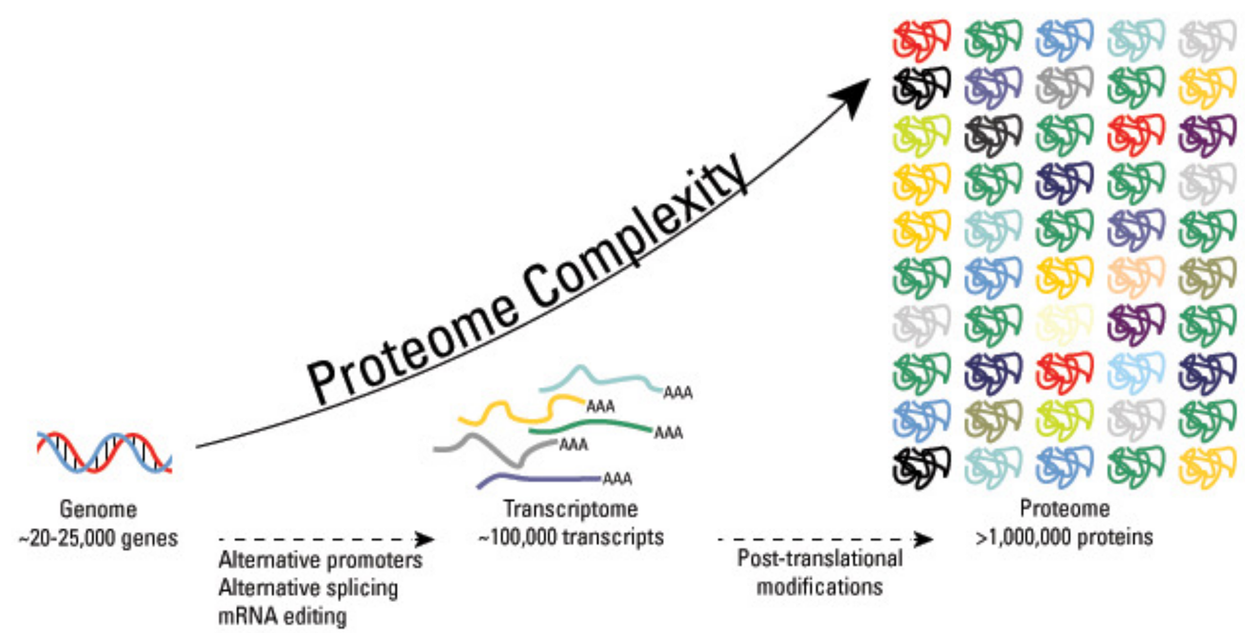
Cell metabolism is driven by active proteins, thus functional proteins

- post-translational modifications,
  e.g. phosphorylation, glycosylation, S-S bond formation
- many proteins are only active in complexes

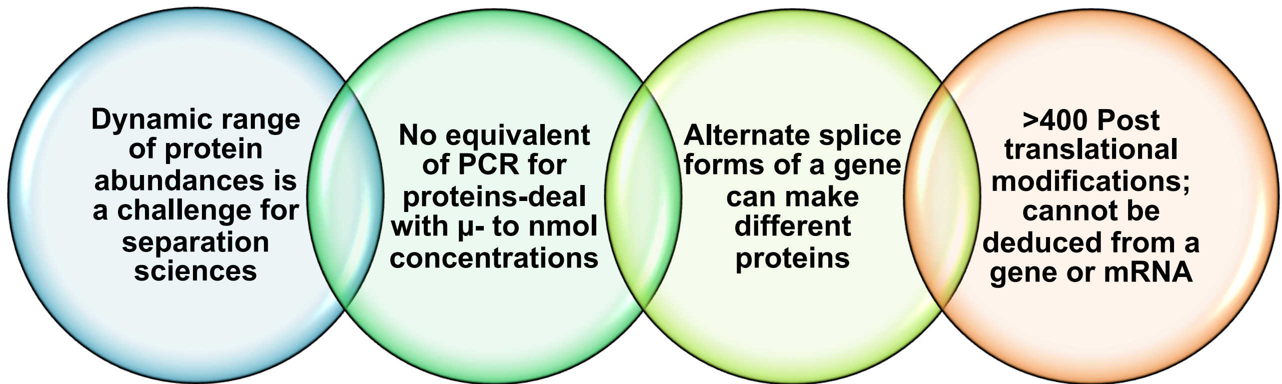
[Tenzer]

The route from genome to proteome

Central dogma „one gene - one enzyme (protein)“ obsolete
Human: approx. 19,773 protein-coding genes, 1 million potential proteoforms?



Characterizing a proteome: An Analytical Challenge



Genome
- Essentially static over time
- Non subcellular location specific
- Human genome mapped (2000)
  - ~20,000 genes

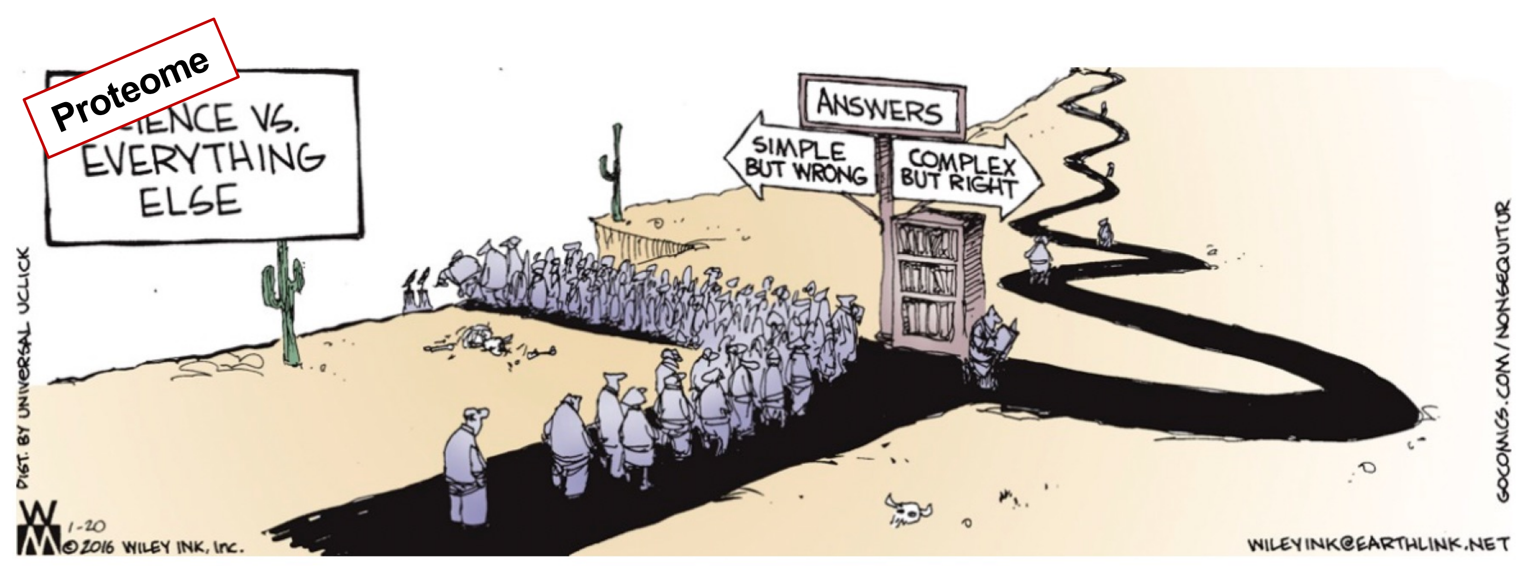
- PCR is available to amplify DNA

Proteome
- Dynamic over time
- Subcellular location specific
- Human proteome non-mapped:
  - ~400,000 proteoforms ???

- No equivalent of PCR for proteins

[Tenzer]

Characterizing a proteome: An analytical challenge



→ We must embrace the complexity…

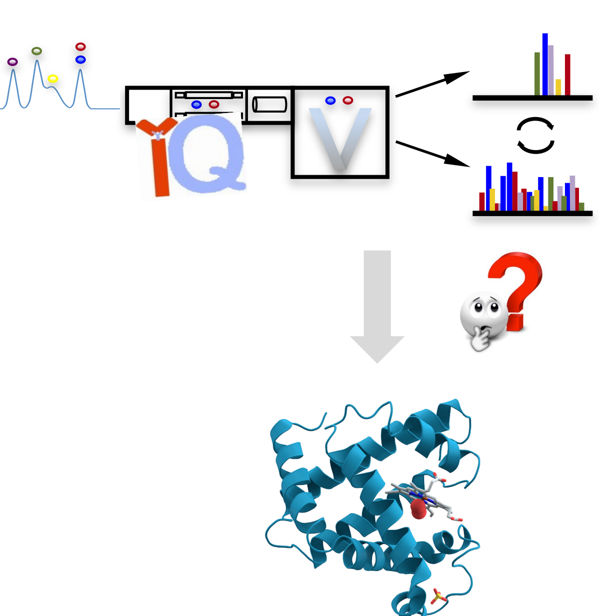
→ No "gold standard" for protein analysis

K Carbonara *et al. Proteomes,* **9**, 38 (2021)

Introduction to Proteomics

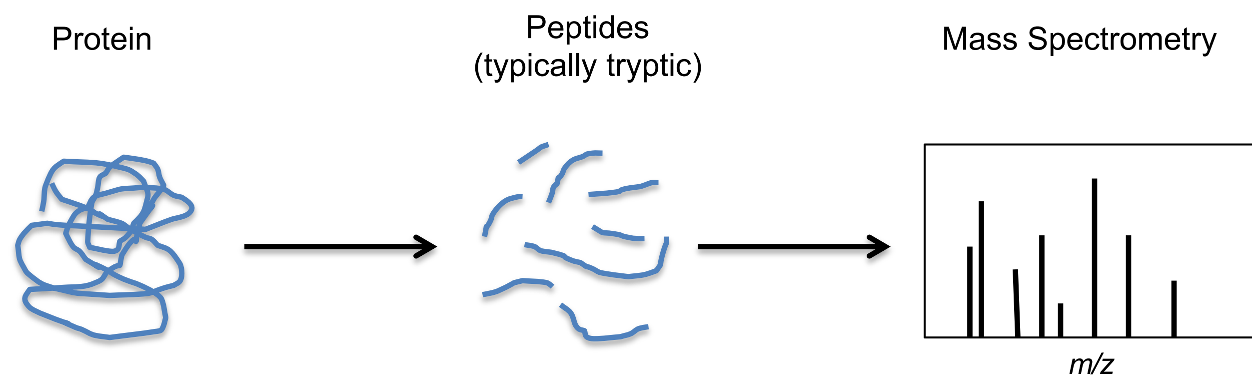
- What is proteomics? And why do we do this?

- Mass spectrometry-based proteomics
  - Sample preparation
  - Protein Identification – „from mass spectrum to protein“
  - Acquisition modes



Mass spectrometry-based proteomics

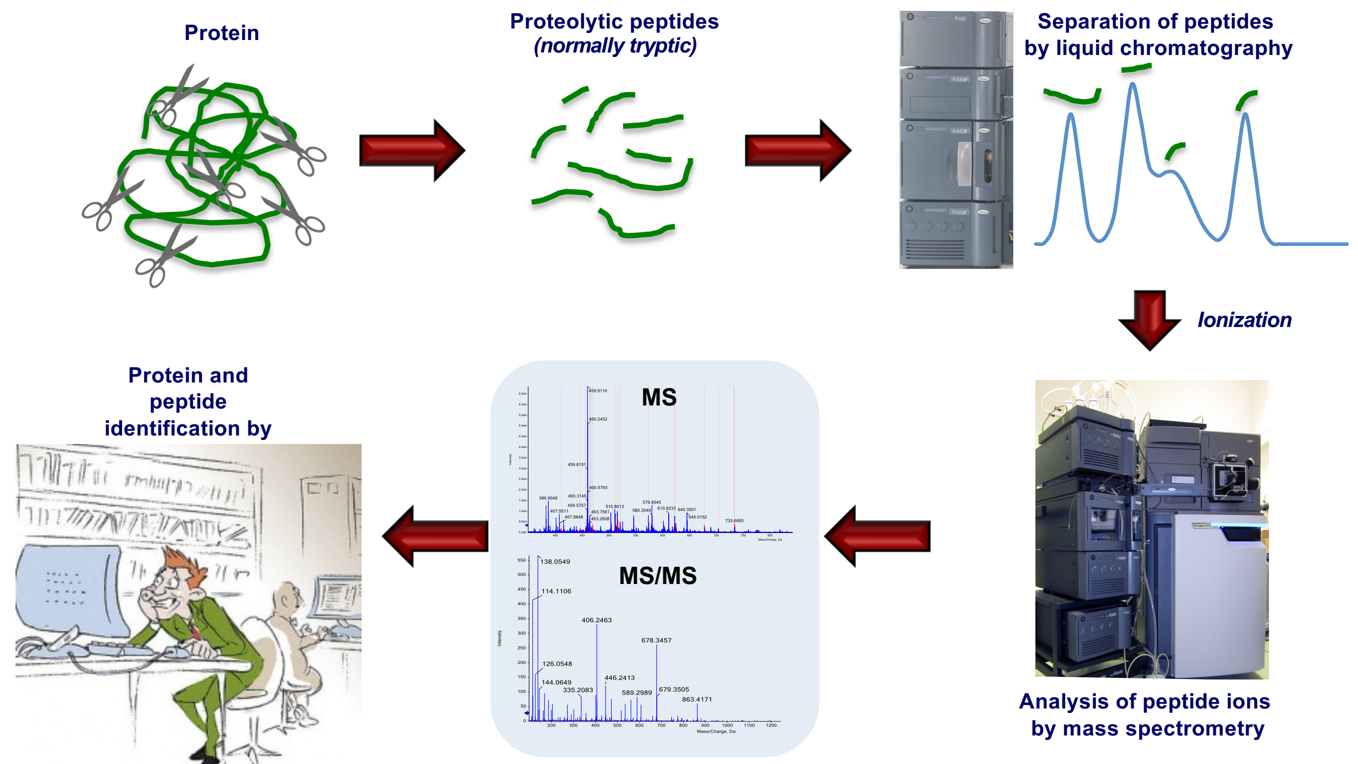
1) *Bottom-up proteomics* (short peptides: 0.7 kDa < M_w < 3.0 kDa)



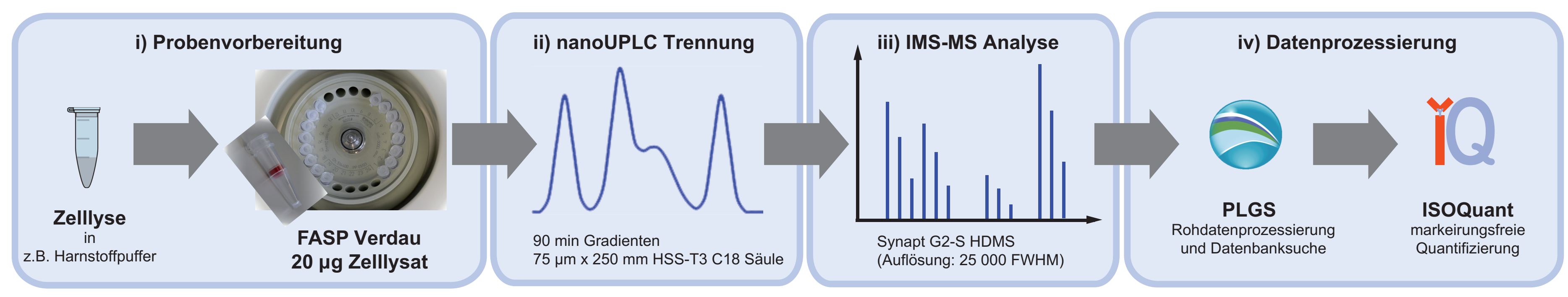
2) *Middle-down*: Analysis of „medium"-length peptides (3.0 kDa < M_w < 10 kDa)

3) *Top-down:* Analysis of intact proteins including fragmentation (e.g. 10 kDa < M_w < 50 kDa/~200 kDa)

Typical *bottom-up* proteomics workflow



Bottom-up proteomics workflow



(1) Sample preparation (Lysis of cells or tissue, proteolytic digest, purification)

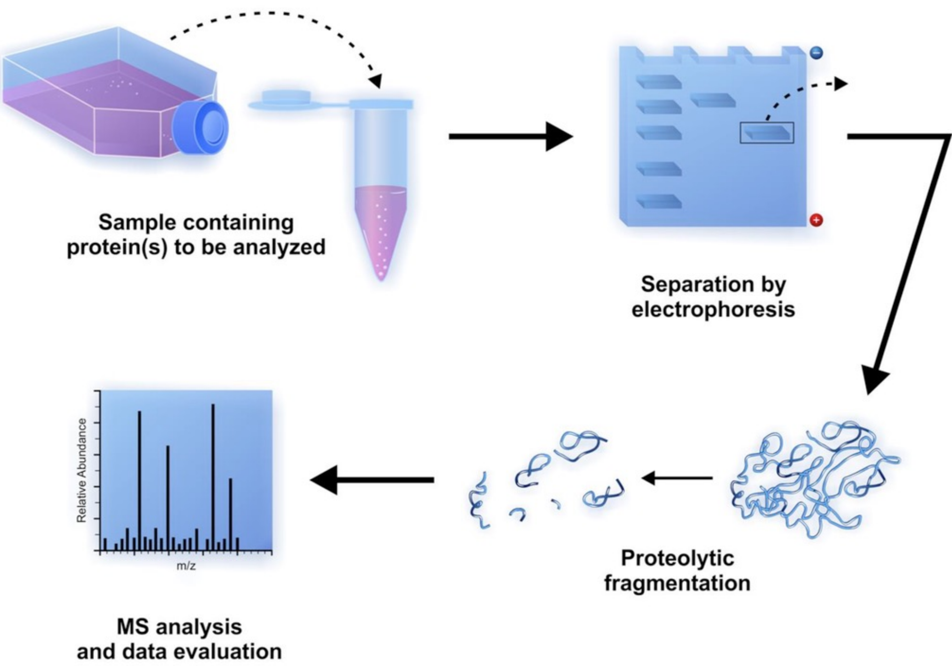
(2) Liquid Chromatography

(3) Mass spectrometry (precursor and fragment level)

(4) Data processing (database search/quantification)

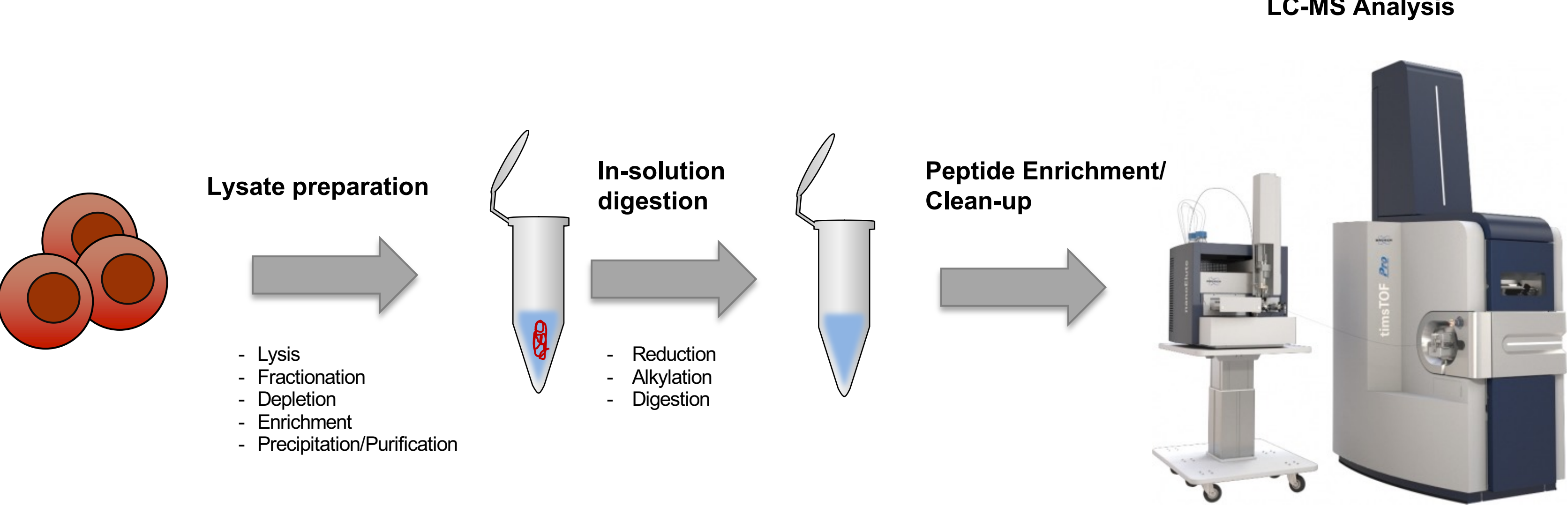
(1) Sample Preparation: Generating peptides from proteins

- In-gel digest



(1) Sample Preparation: Generating peptides from proteins

- In-solution digest



Cell lysis
- Choice of lysis buffer -

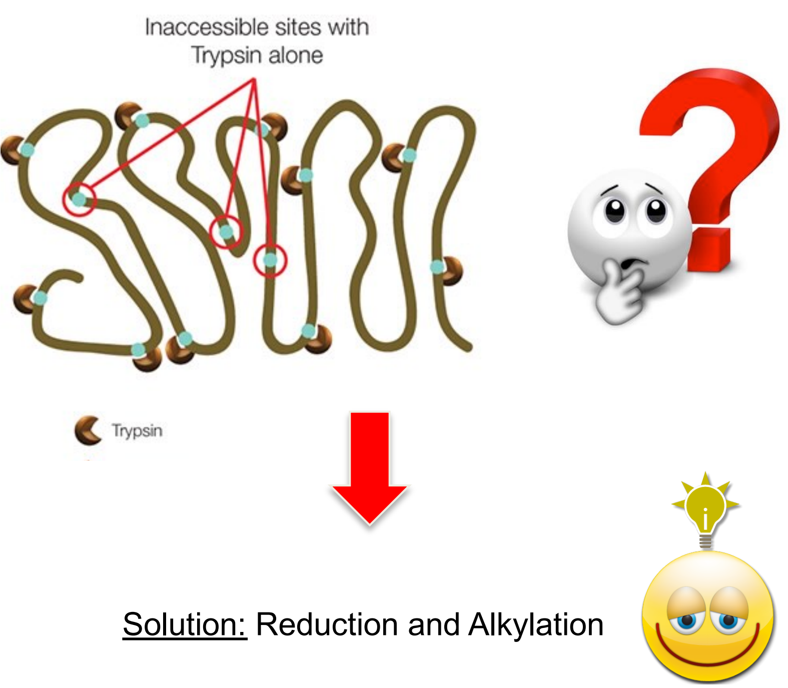
- Which **pH**?
- E.g. Tris, phosphate and HEPES → good buffering capacity around physiological pH conditions

Additive	Example	Goal
Salt	NaCl, KCl, $(NH_4)_2SO_4$, …	Provide ionic strength
Glycerol		Protein stabilization
Reducing agents	DTT, DTE, TCEP, β-mercaptoethanol	Reduce oxidation damage
Detergents	Tween20, Triton-X100, octylglucoside, dodecylmaltoside, CHAPS, …	For poorly soluble and membrane(-associated) proteins
Co-factors	Zn^{2+}, Mg^{2+}, GTP, ATP, NAD, …	Protein stabilization
Chelating agents	EDTA, EGTA	Reduce oxidation damage

…plus (phospho)protease inhibitors, benzonase or Dnase…

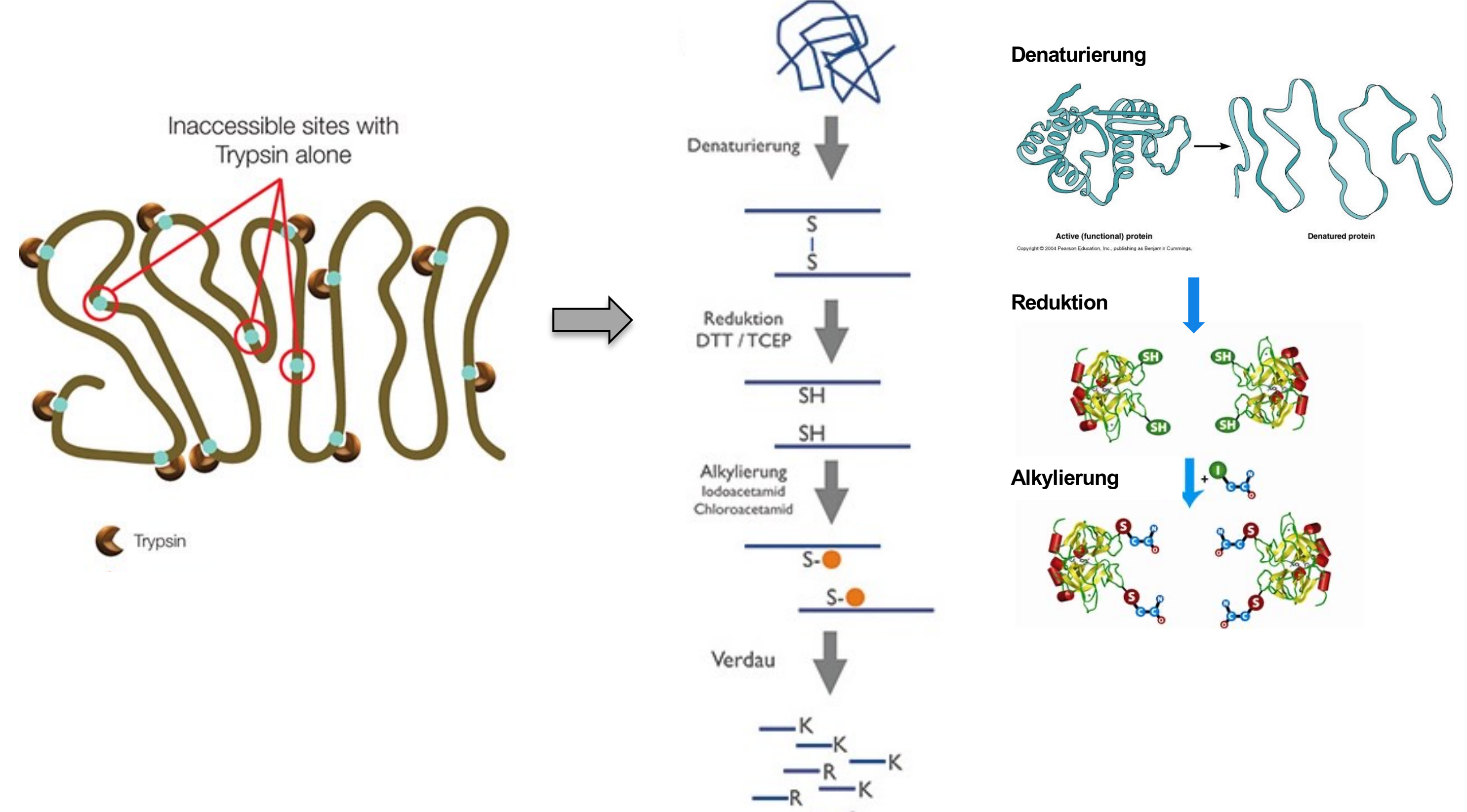
Generating peptides from proteins

Problem: Many cleavage site are not accessible for proteases within the native protein…



Solution: Reduction and Alkylation

Reduction and alkylation facilitates access to cleavage sites



Specificities of different endoproteinases

Trypsin R K → R K

Lys-C K → K

Asp-N D → D

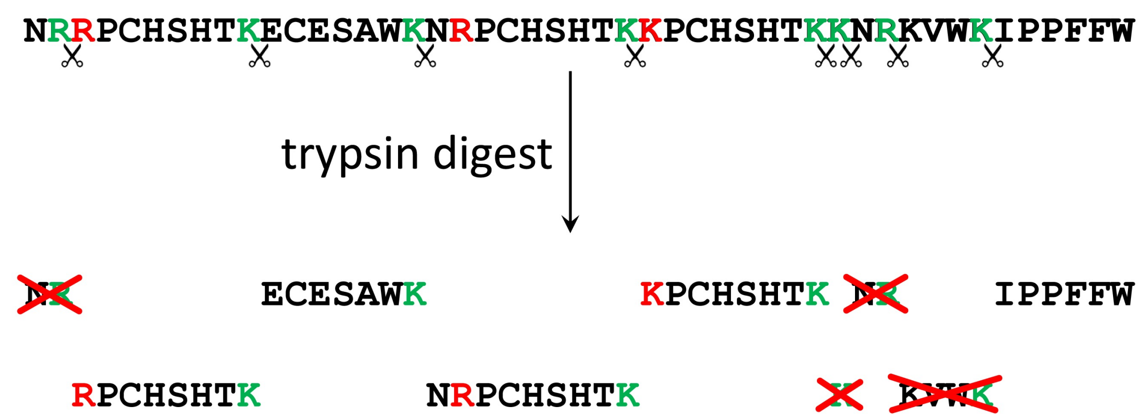
V8/Glu-C E (D) → E (D)

Chymotrypsin F Y W (L, I, V, M) → F Y W (L, I, V, M)

Arg-C R → R

Workflow: Digestion with trypsin – why??

- Relatively known and precise lysis behavior: C-ter from K and L
- Generates mostly di-charged peptides (N-ter and R/K)
- Generate small peptides (0.5–3 kDa) = good for LC-MS
- But sometimes too small: 56% are≤6 residues = not specific enough for protein ID
- Can't result on 100% protein coverage (no single protease can)



[David Gómez-Zepeda]

https://unipept.ugent.be/clidocs/casestudies/tpa

Peptide clean-up prior to MS analysis

Desalting using C18 material

Peptides

60% ACN
0.1% TFA

C_{18}

filter resin

Salt Contaminats

Desalted Peptides

LC-MS analysis

mAU

min

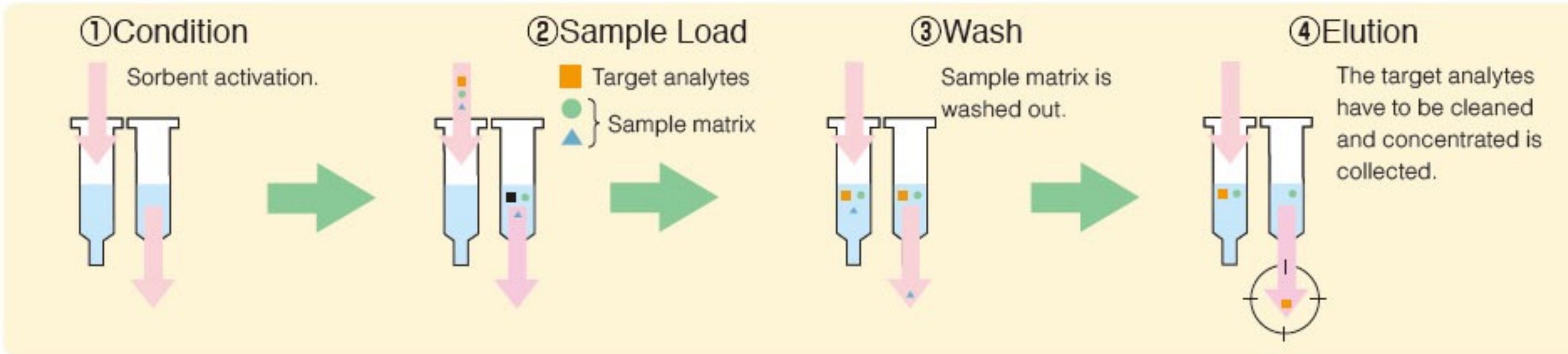
MS

MS/MS

www.waters.com
www.globalspec.com

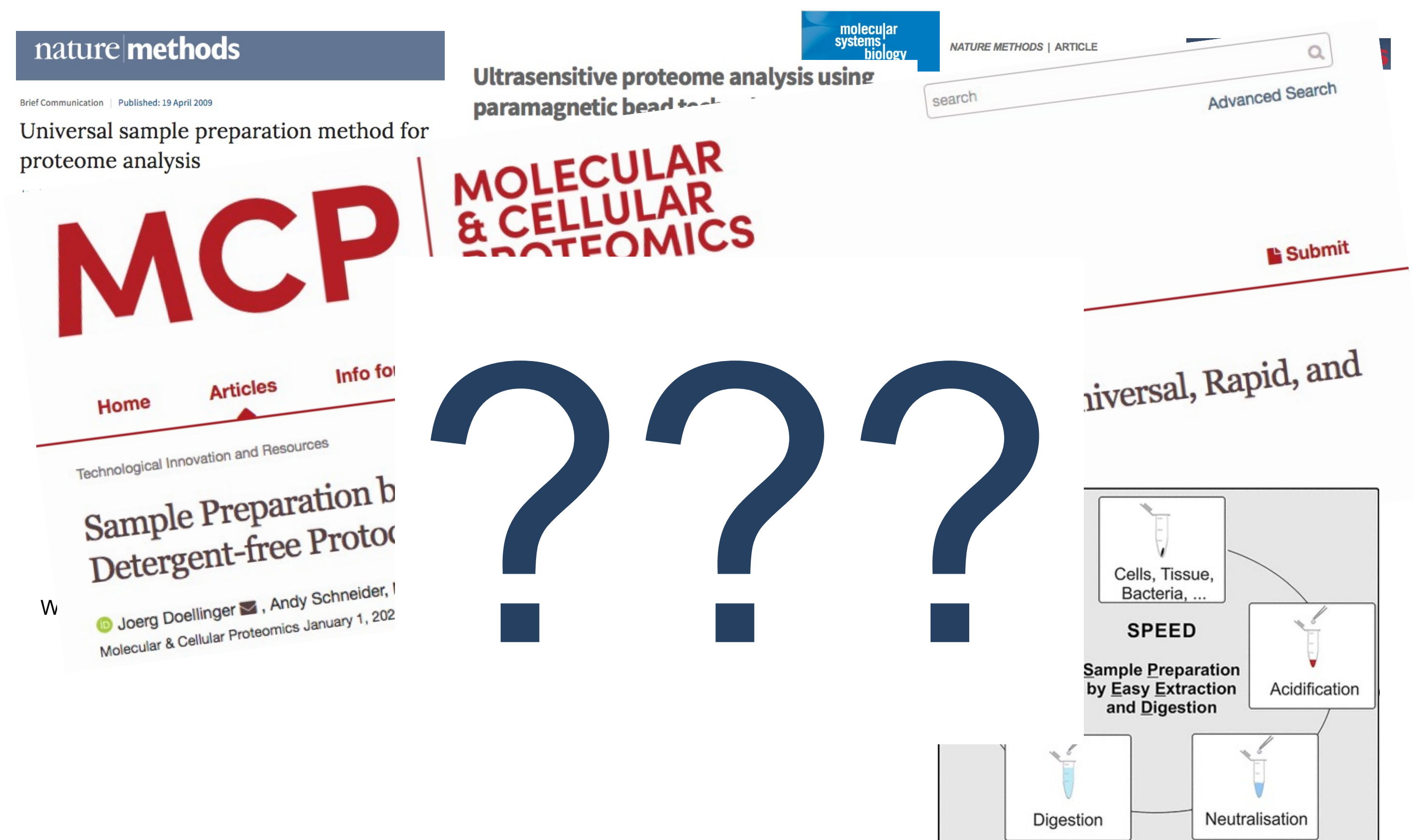
Peptide clean-up prior to MS analysis

Remaining buffer salts (or other impurities) from proteolytic digestion might impair subsequent MS analysis



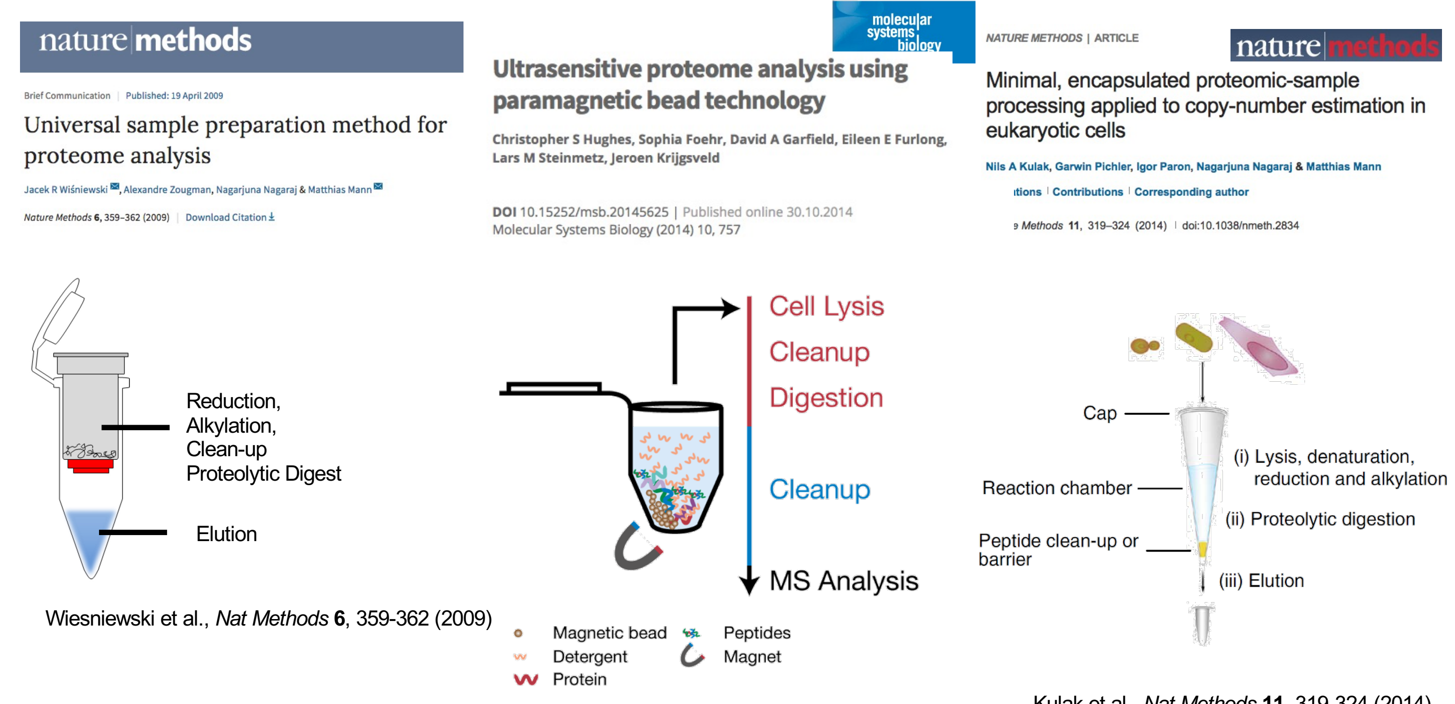
https://www.glsciences.com/product/spe_columns/about_a_spe/01042.html

Digestion protocols for *bottom-up* proteomic analyses



Digestion protocols for *bottom-up* proteomic analyses

Pubmed search “sample preparation + proteomics + mass spectrometry”: 2,365 hits

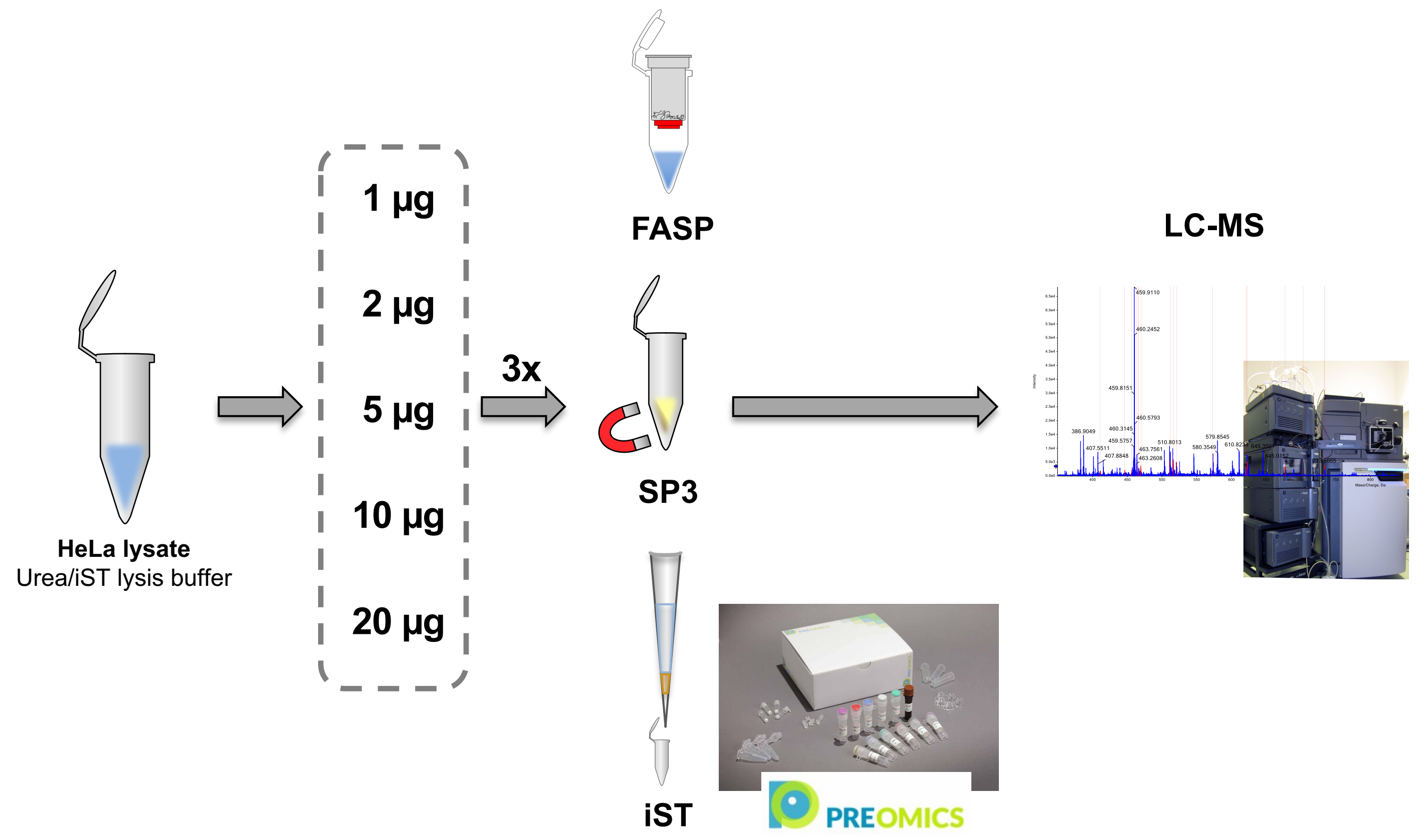


Wiesniewski et al., *Nat Methods* **6**, 359-362 (2009)

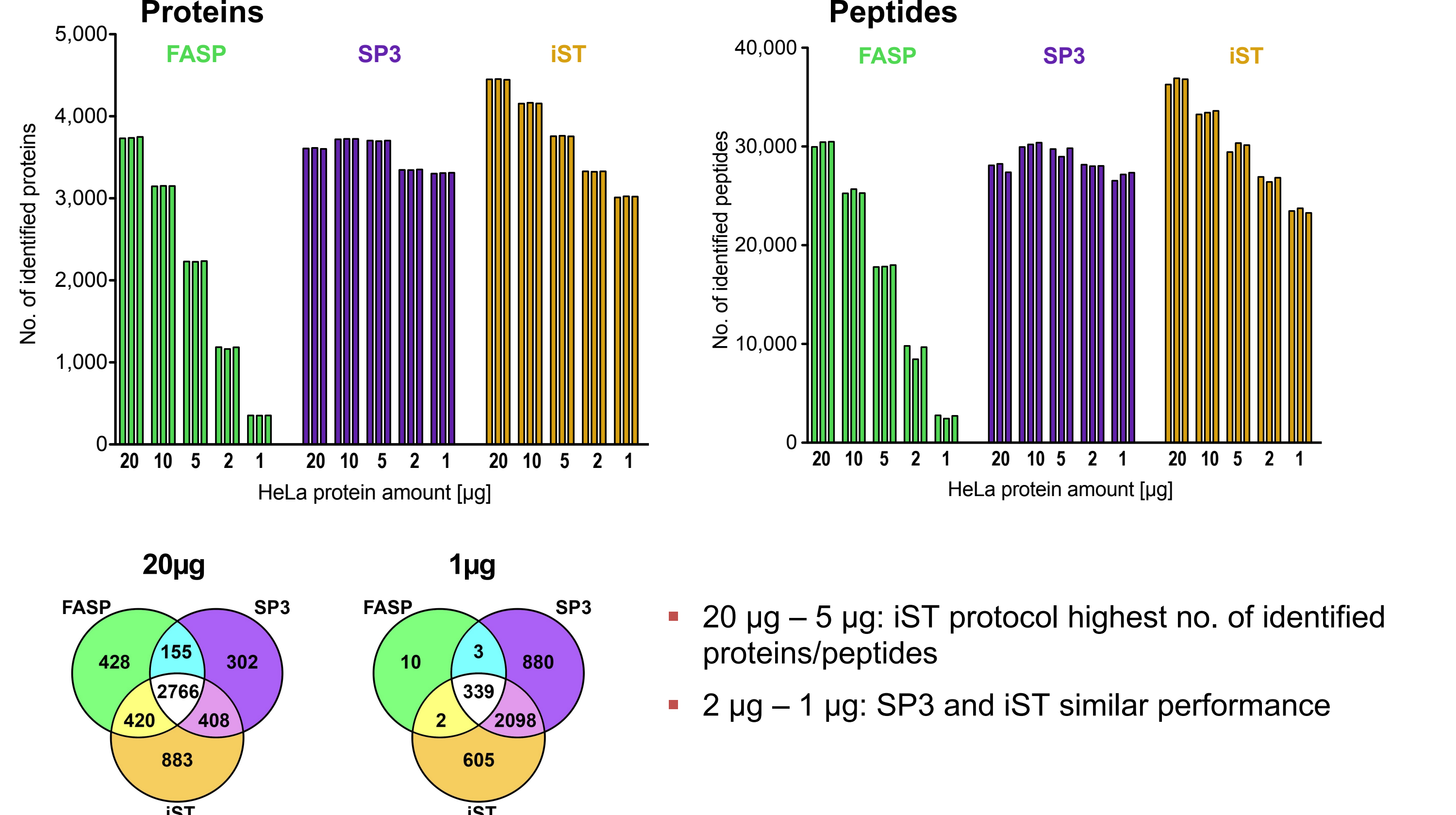
Hughes et al., *Mol Sys Biol* **10**, 757 (2014)

Kulak et al., *Nat Methods* **11**, 319-324 (2014)

Digestion protocols for *bottom-up* proteomic analyses
- Comparison of three popular protocols for low input material -

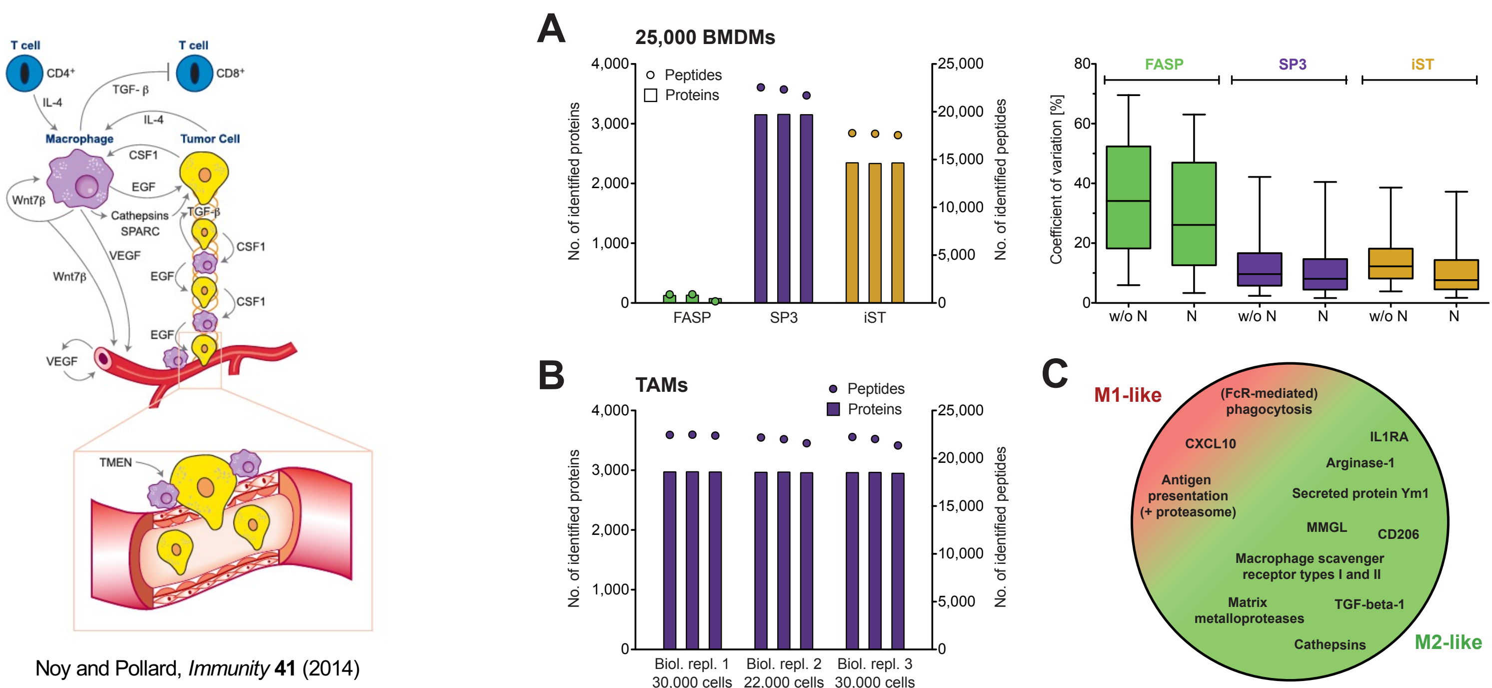


Digestion protocols for *bottom-up* proteomic analyses
- Comparison of three popular protocols for low input material -



- 20 µg – 5 µg: iST protocol highest no. of identified proteins/peptides
- 2 µg – 1 µg: SP3 and iST similar performance

Comparison of three popular protocols for low input material
- FACS sorted cells -

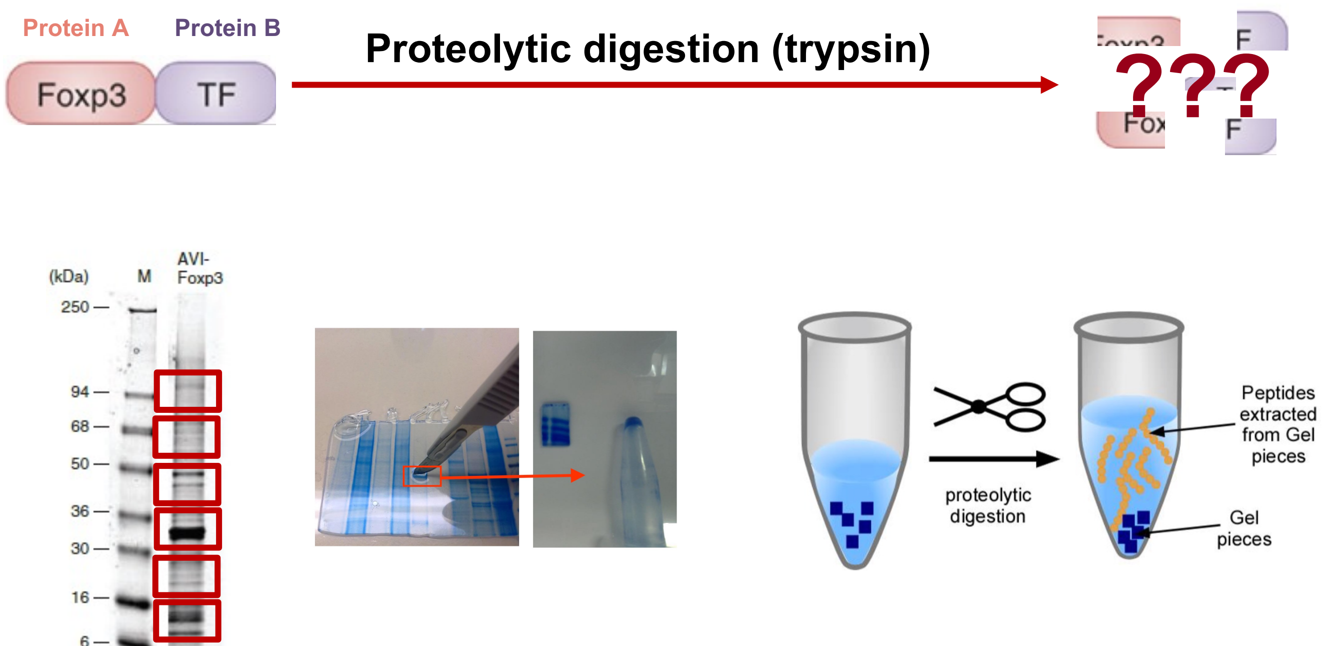


Noy and Pollard, *Immunity* **41** (2014)

- SP3 combined with LC-IMS-MS allows the identification and quantification of around 3,000 proteins from TAMs isolated from murine tumors by FACS

Sielaff et al., *J Proteome Res* **16**, 4060-4072 (2017)

Protein identification via mass spectrometry



Tryptic digestion

Asp-Ala-Gly-Arg-His-Cys-Lys-Trp-Lys-Ser-Glu-Asn-Leu-Ile-Arg-Thr-Tyr

↓ Trypsin, H_2O

Asp-Ala-Gly-Arg

His-Cys-Lys

Trp-Lys

Ser-Glu-Asn-Leu-Ile-Arg

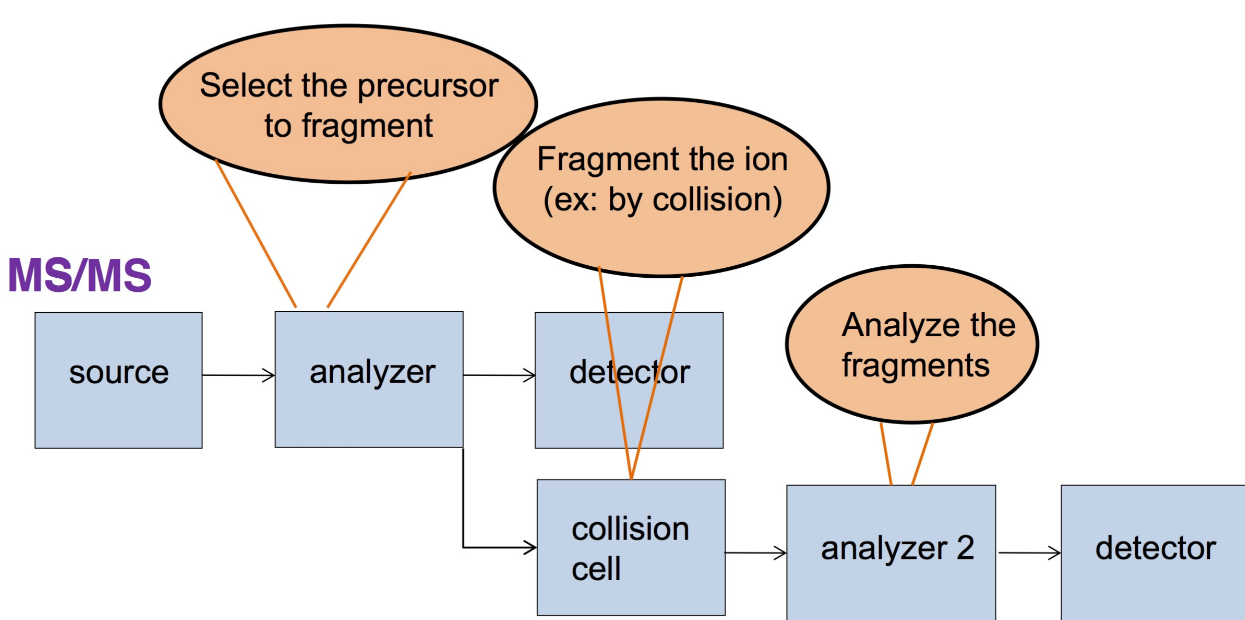
Thr-Tyr

485 Da   378 Da   312 Da   536 Da   257 Da

Tandem mass spectrometry (MS/MS)

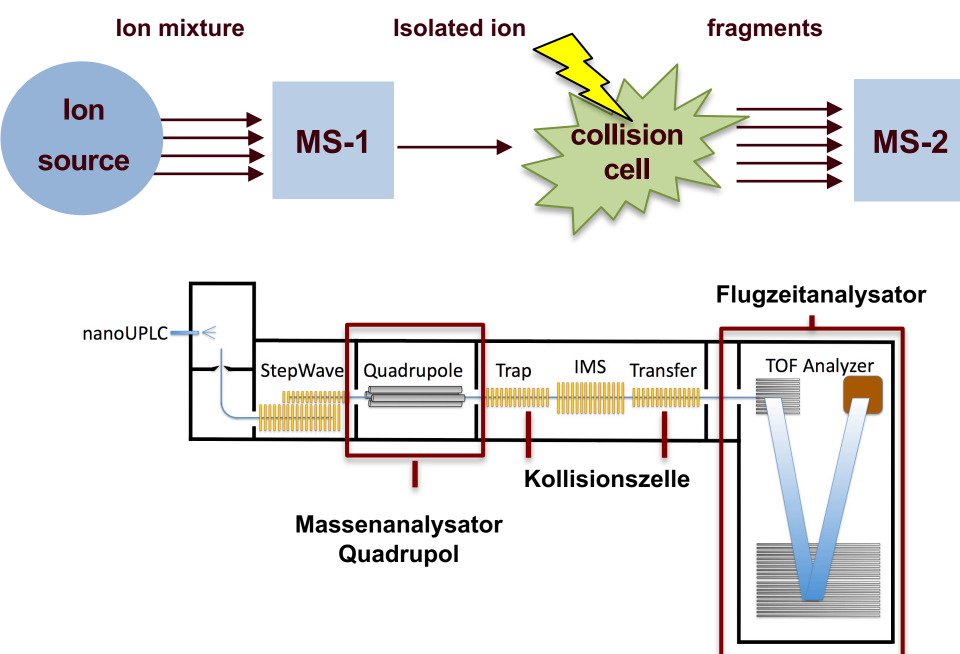
Goal: Sequence information of the peptide

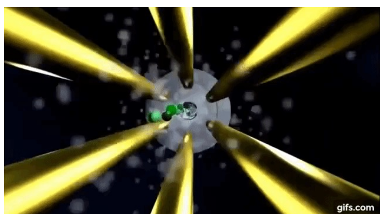
- MS (MS1): *m/z* of the intact peptide
- MS/MS: fragmentation of the peptide into smaller „pieces“ to determine its primary structure



Tandem mass spectrometry (MS/MS)

MS/MS:
The use of two mass analyzers (combined in one instrument, i.e.tandem mass spectrometer) for ion isolation, fragmentation and fragment ion detection



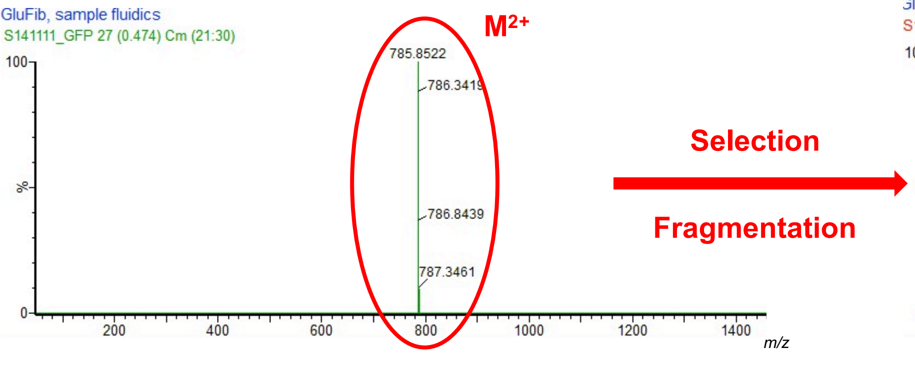


What is MS/MS?

[Glu1] - Fibrinopeptide B
EGVNDNEEGFFSAR

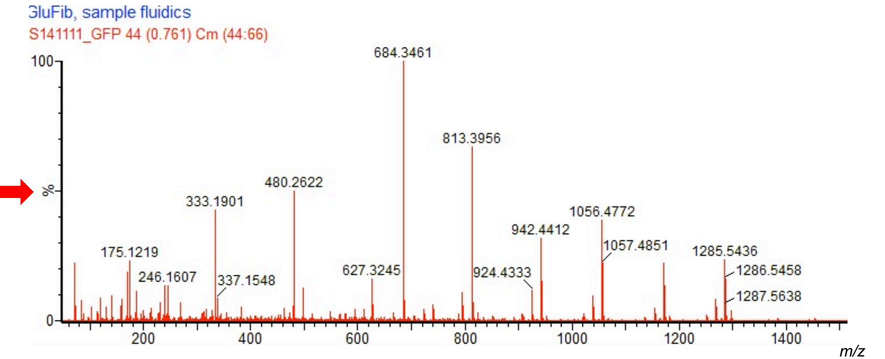
1. Analysis of the precursor ion(s) (*m/z)*

2. Analysis of the fragment ions (*m/z)*



Selection

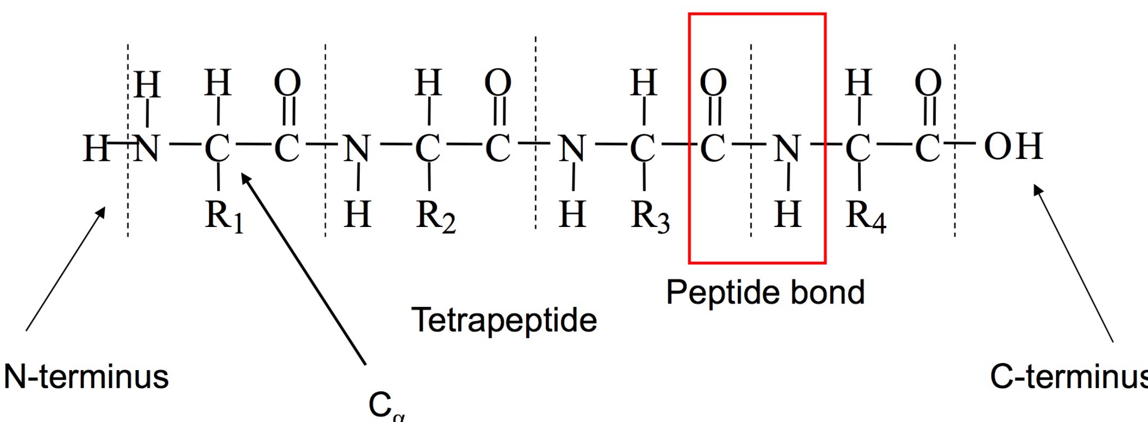
Fragmentation



Tryptic Digest
- Terminology -

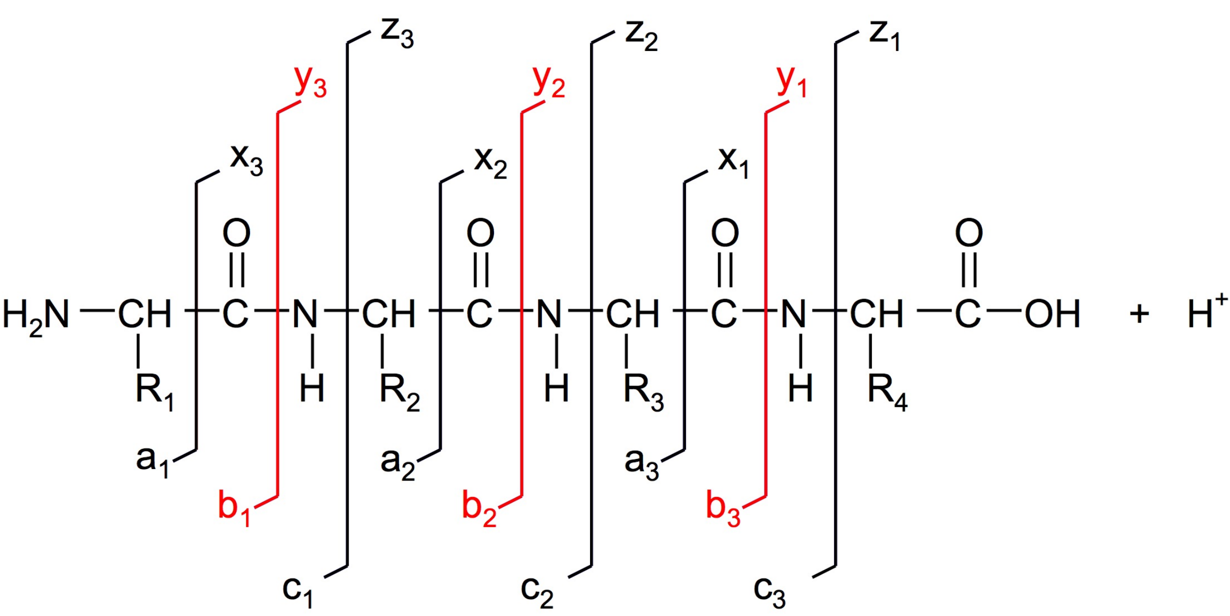
Amino acid

Residue (amino acid minus H_2O)



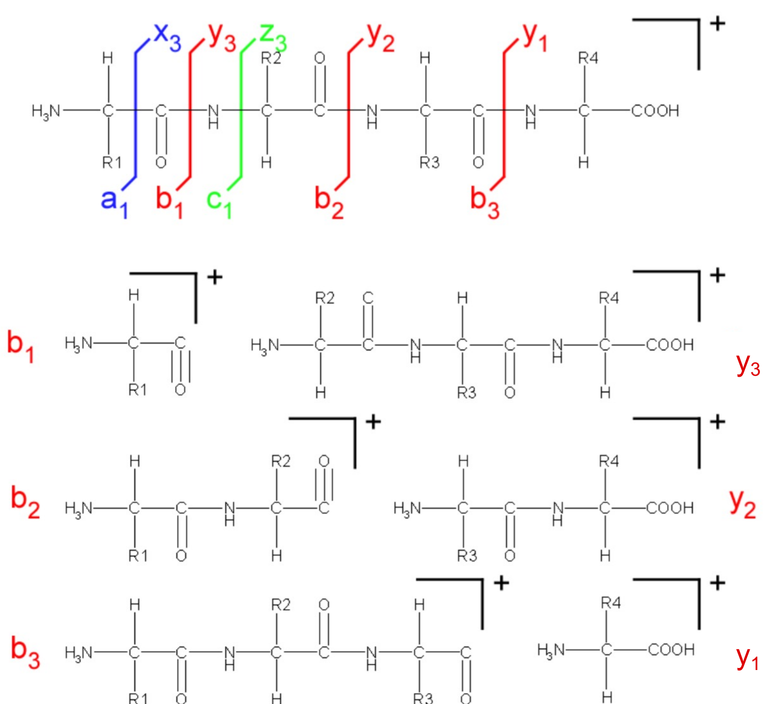
Identification of peptides by fragmentation

Fortunately, peptides fragment at predictable locations along the peptide backbone



Identification of peptides by fragmentation

- Scheme of peptide fragmentation (nomenclature according to Roepstorff and Fohlmann)
- After separation by LC, peptide ions are selected and fragmented in the mass spectrometer



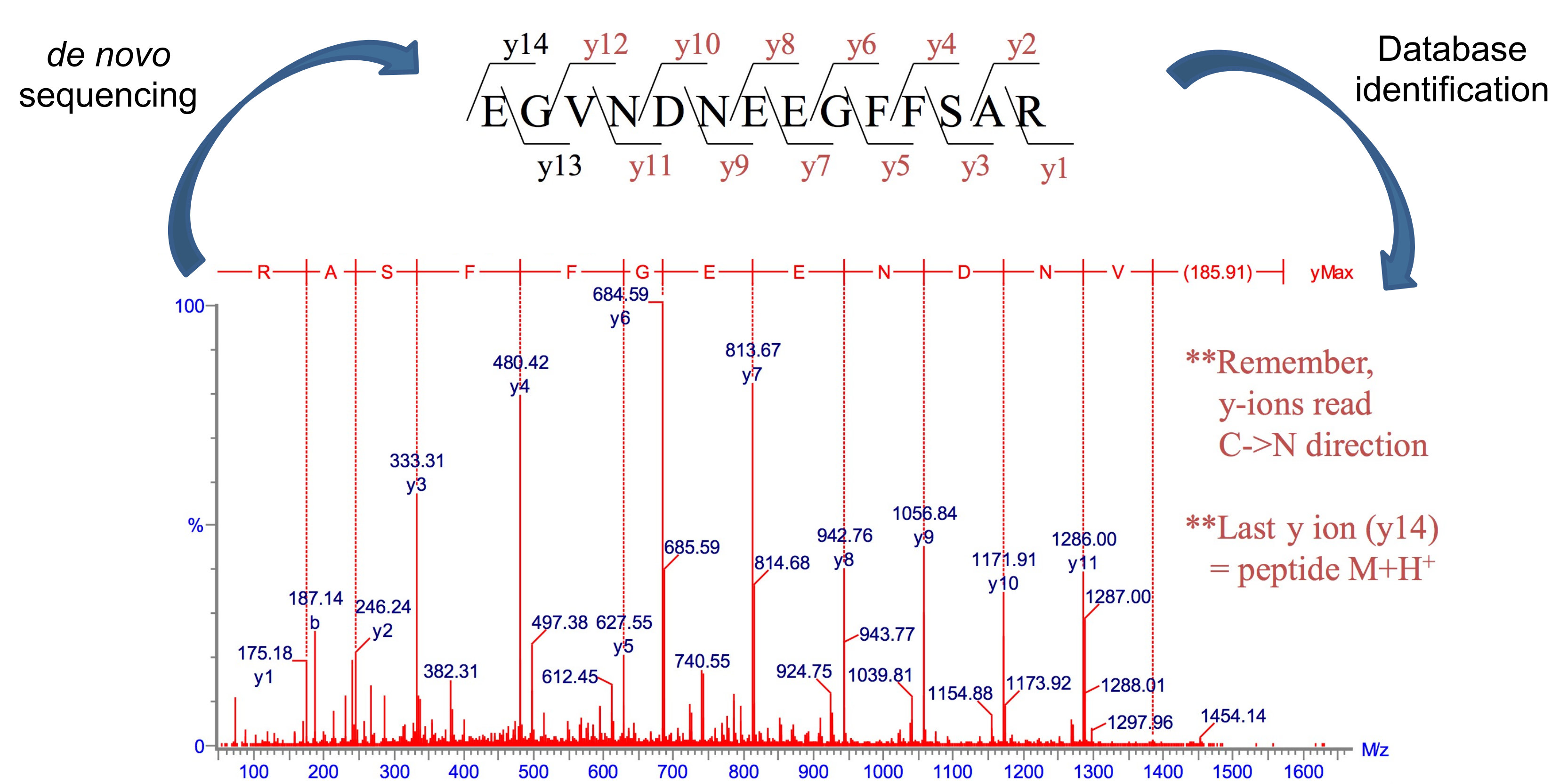
How can we calculate the mass of a peptide?

$$H-N(H)-C(H)(R_1)-C(=O)-N(H)-C(H)(R_2)-C(=O)-/\!/-N(H)-C(H)(R_{n-1})-C(=O)-N(H)-C(H)(R_n)-C(=O)-OH \quad + \quad H^+$$

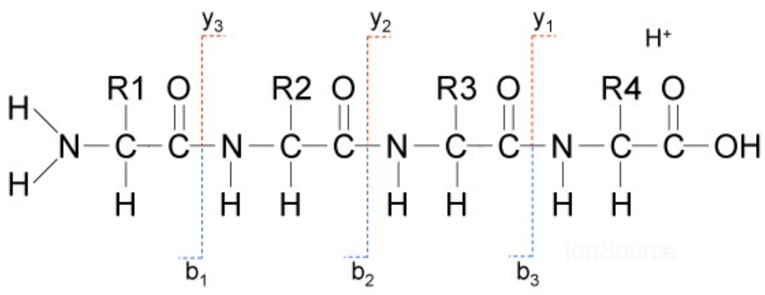
1	+	mass residue 1	+	mass residue 2	+	mass residue n-1	+	mass residue n	+ 16+1	+ 1

- **Mass of all AS residues plus water and proton(s)**

Identification of peptides by fragmentation

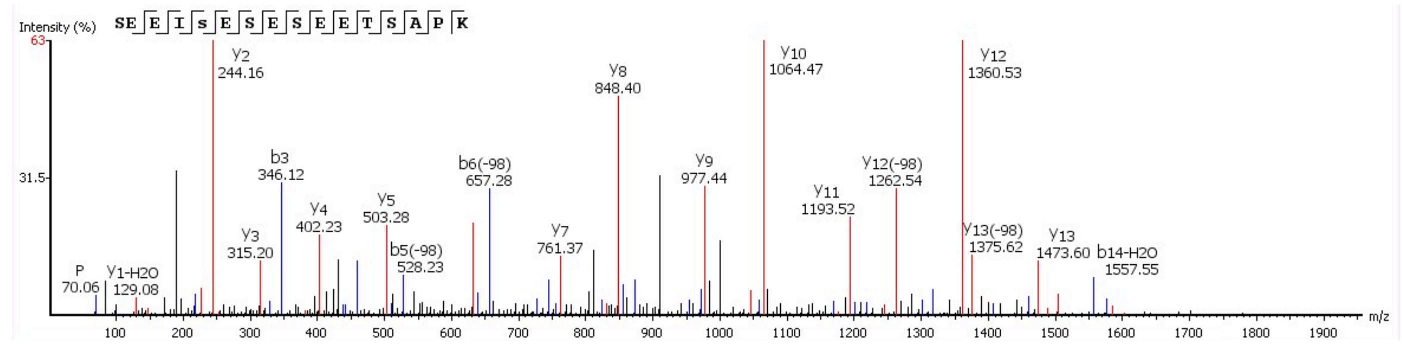


Identification of peptides by fragmentation



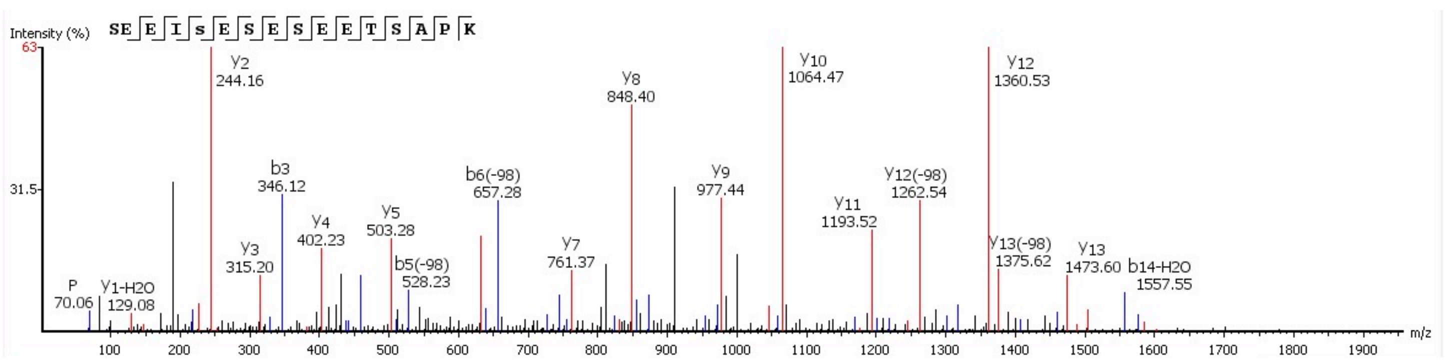
according to Roepstorff P, Fohlman J., *Biomed Mass Spectrom* **11**, 601 (1984)

Atrophin-1



Identification of peptides by fragmentation

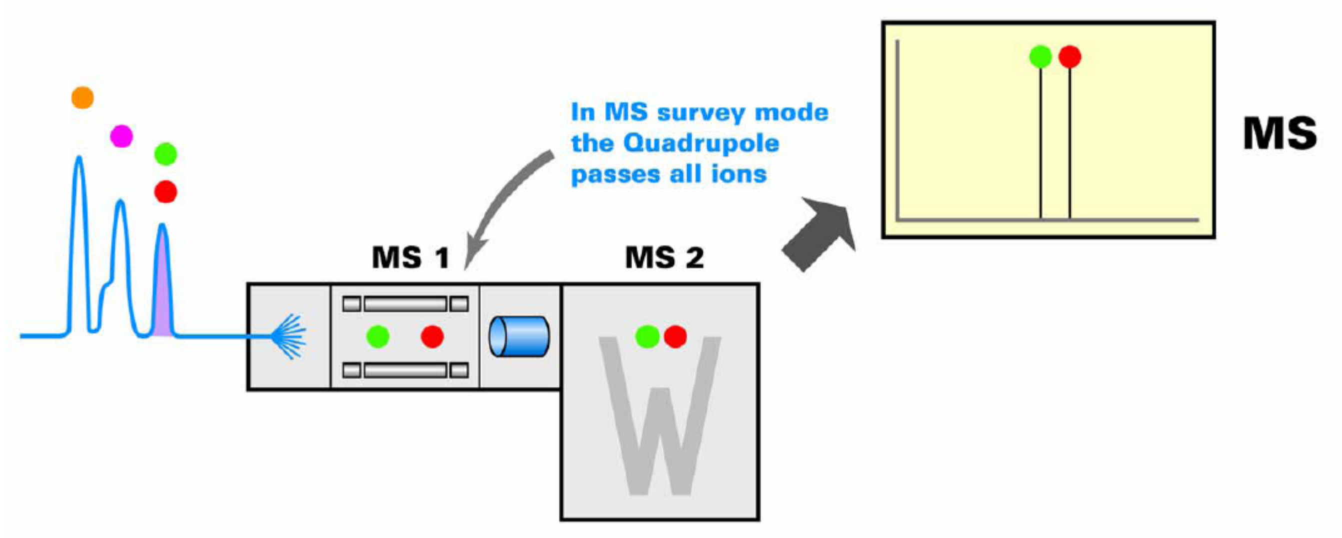
Atrophin-1



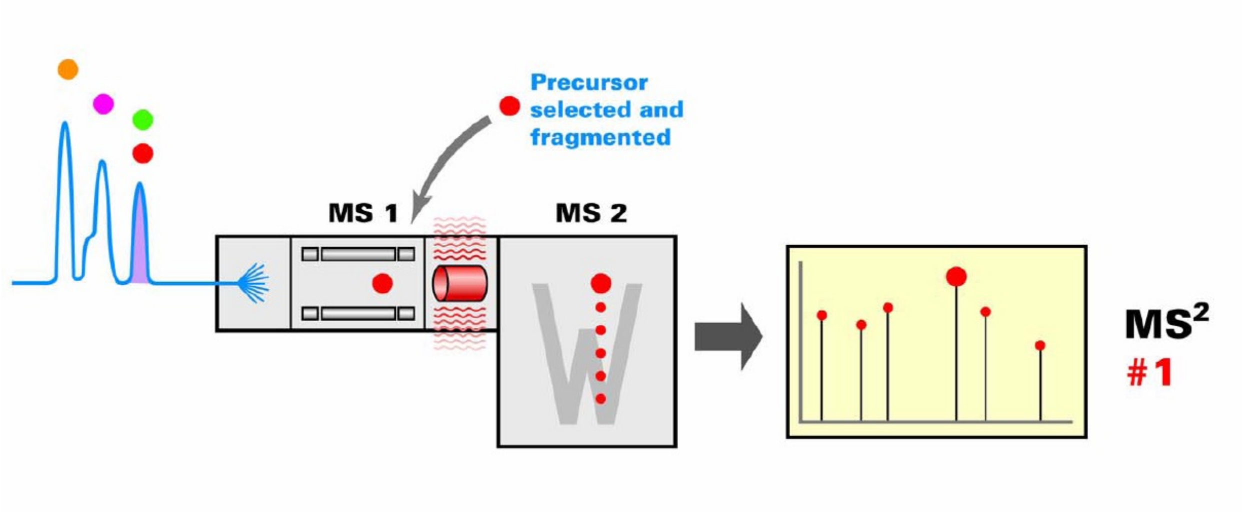
Ion Match

#	Immonium	b	b-H2O	b-NH3	b (2+)	Seq	y	y-H2O	y-NH3	y (2+)	#
1	60.04	88.04	70.03	71.01	44.52	S					16
2	102.05	217.08	199.07	200.06	109.04	E	1731.68	1713.67	1714.66	866.34	15
3	102.06	346.12	328.12	329.10	173.56	E	1602.65	1584.63	1585.63	801.82	14
4	86.10	459.20	441.20	442.20	230.10	I	1473.60	1455.59	1456.57	737.30	13
5	140.01	626.21	608.20	609.18	313.60	S(+79.97)	1360.53	1342.50	1343.49	680.76	12
6	102.06	755.25	737.24	738.22	378.12	E	1193.52	1175.51	1176.52	597.26	11
7	60.04	842.28	824.27	825.26	421.64	S	1064.47	1046.46	1047.45	532.74	10
8	102.06	971.34	953.33	954.30	486.16	E	977.44	959.43	960.42	489.22	9
9	60.04	1058.36	1040.35	1041.33	529.68	S	848.40	830.41	831.37	424.70	8
10	102.06	1187.40	1169.42	1170.37	594.20	E	761.37	743.36	744.34	381.17	7
11	102.06	1316.43	1298.43	1299.41	658.72	E	632.32	614.31	615.30	316.66	6
12	74.06	1417.49	1399.48	1400.46	709.24	T	503.28	485.27	486.26	252.13	5
13	60.04	1504.52	1486.51	1487.49	752.76	S	402.23	384.22	385.21	201.62	4
14	44.05	1575.57	1557.55	1558.53	788.28	A	315.20	297.19	298.18	158.10	3
15	70.06	1672.61	1654.60	1655.58	836.81	P	244.16	226.15	227.14	122.58	2
16	101.11					K	147.12	129.08	130.09	74.06	1

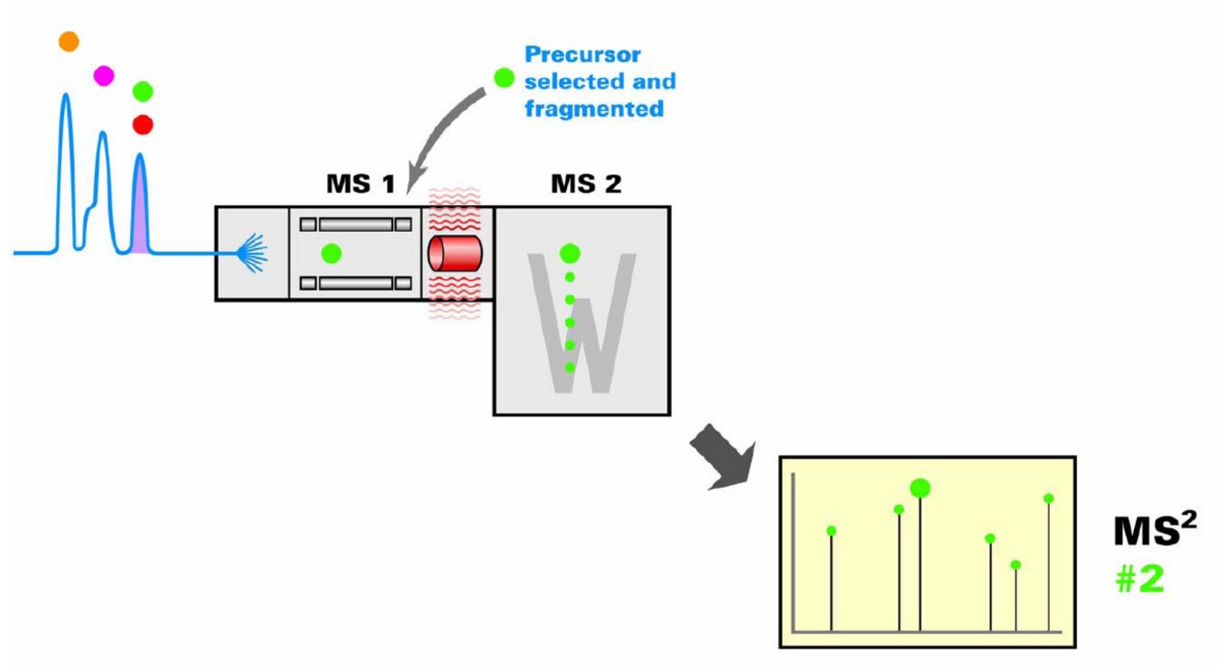
Data dependent acquisition (DDA) via MS/MS



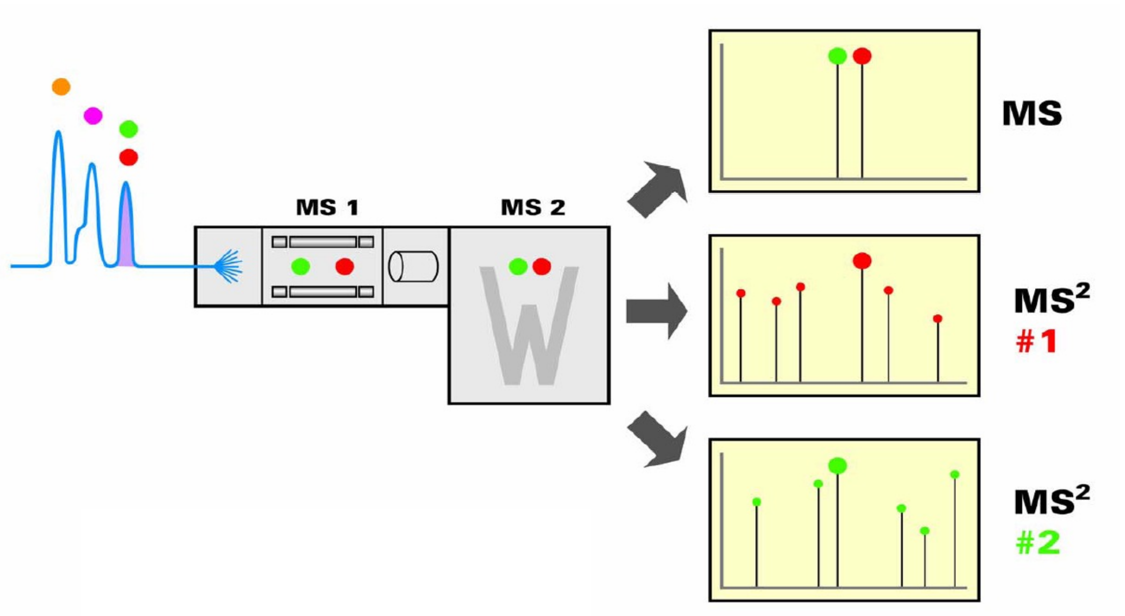
Data dependent acquisition (DDA) via MS/MS



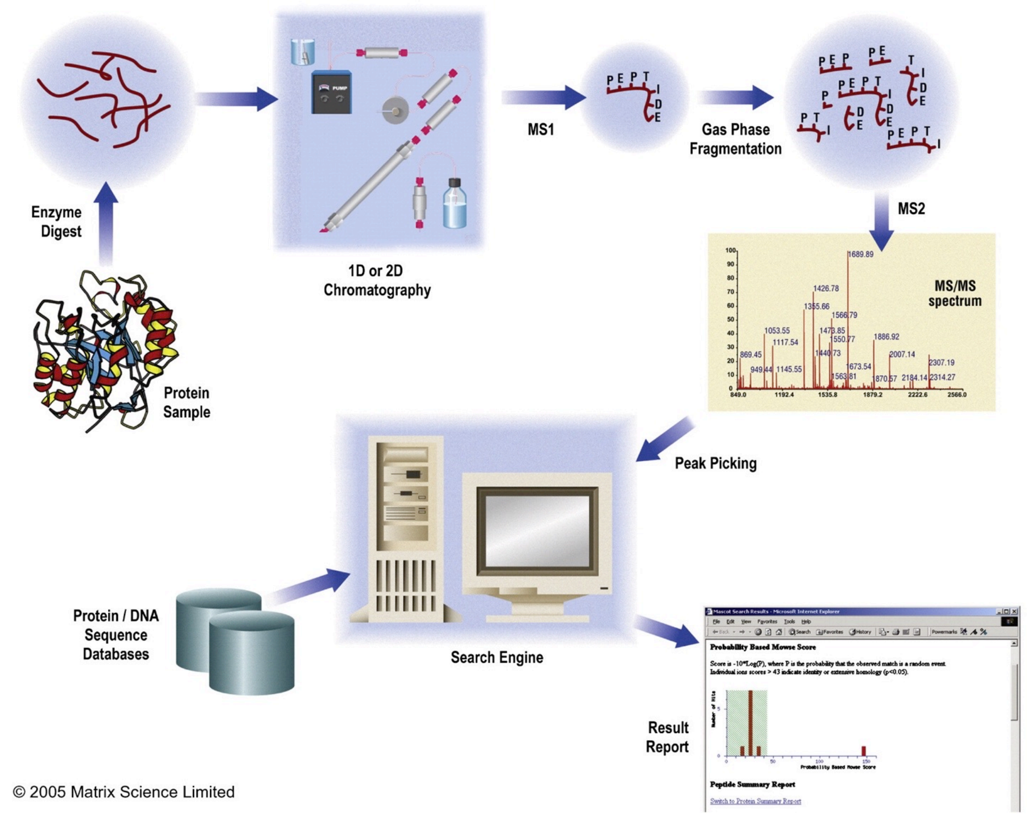
Data dependent acquisition (DDA) via MS/MS



Data dependent acquisition (DDA) via MS/MS

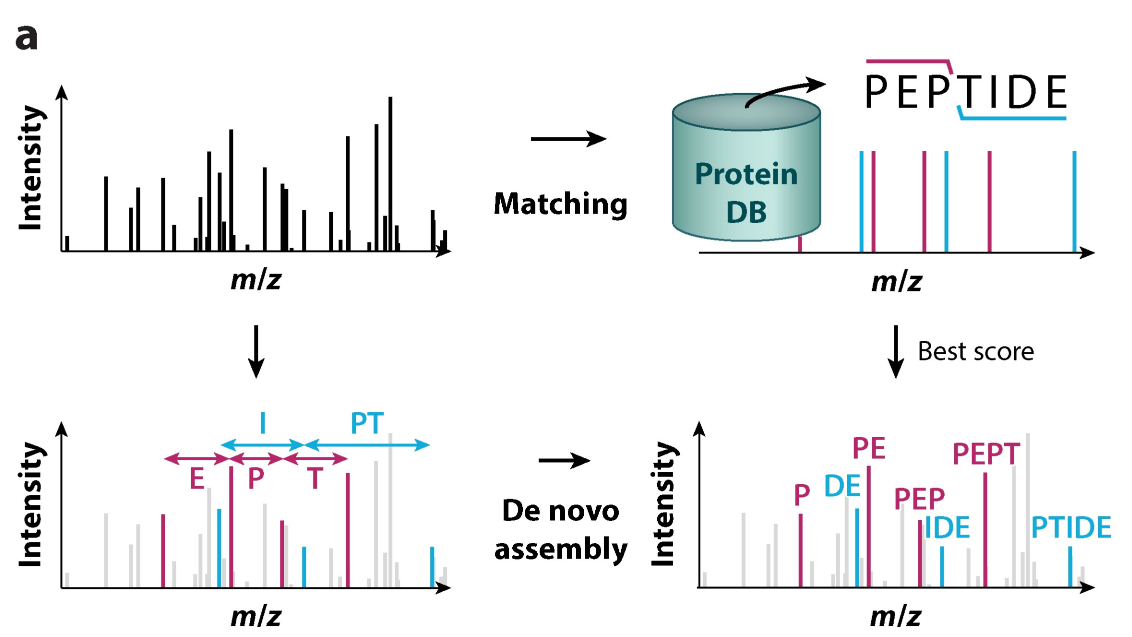


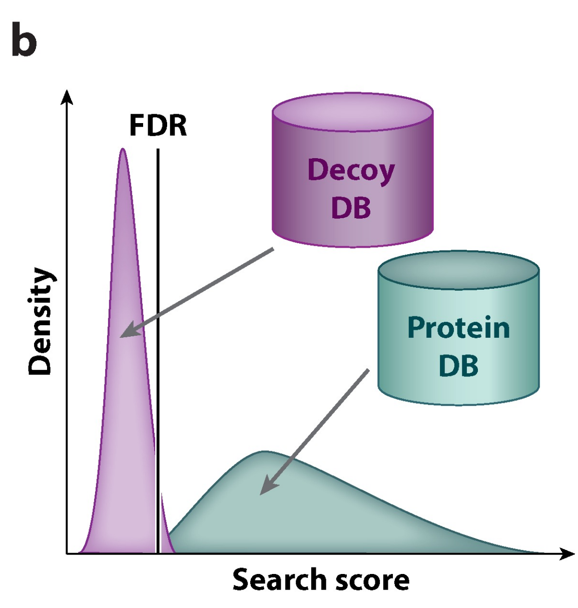
Identification of peptides by MS/MS and database search



© 2005 Matrix Science Limited

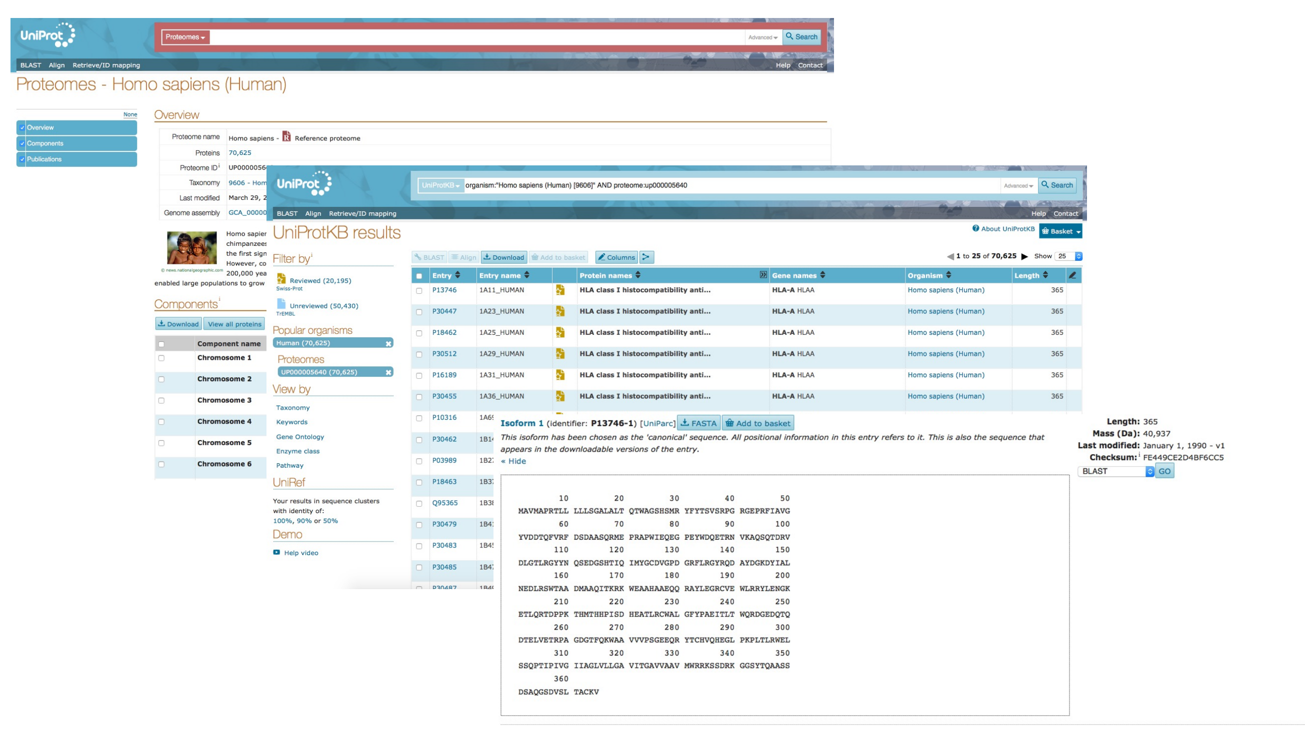
Identification of peptides by MS/MS and database search



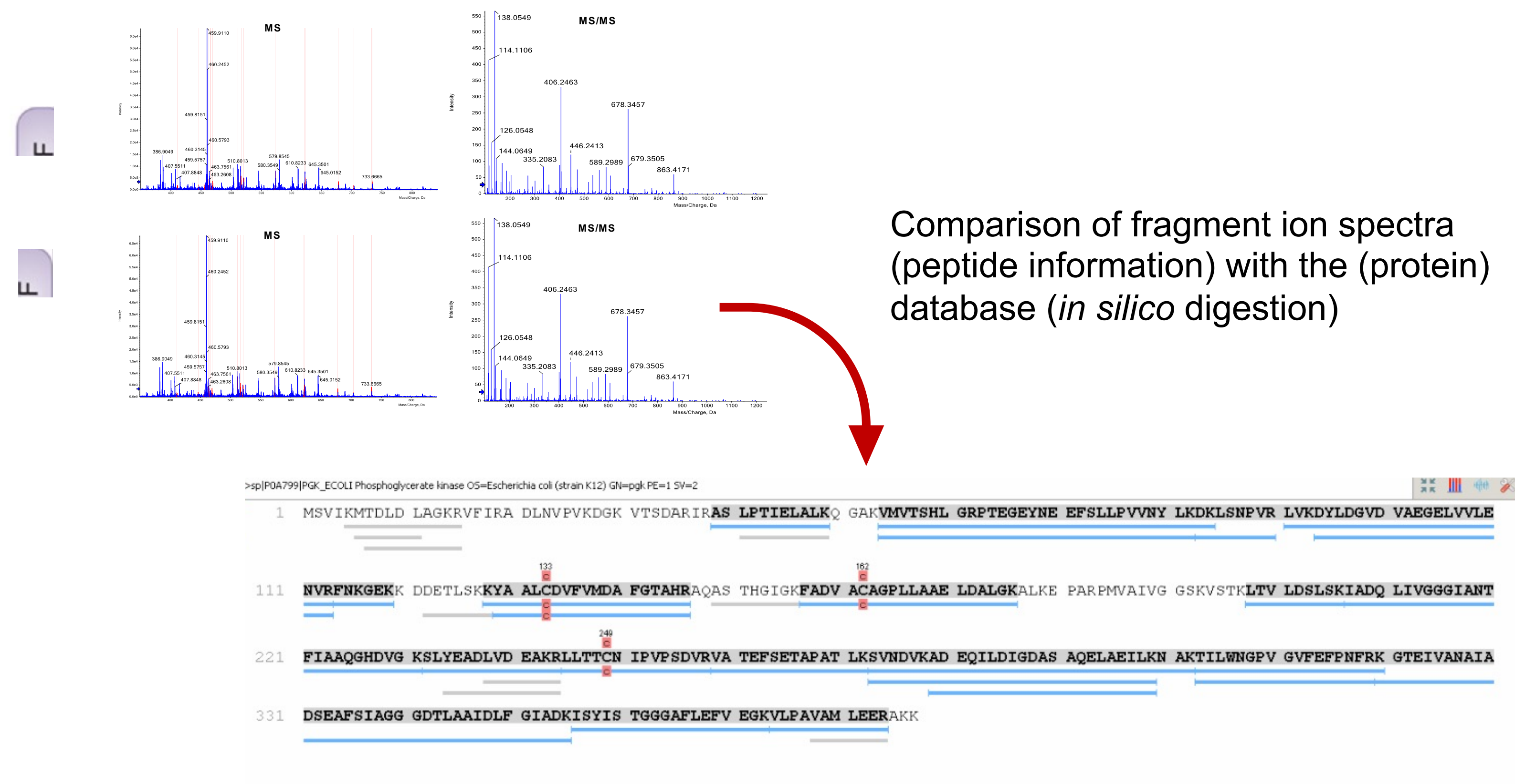


Sinitcyn P, et al. 2018.
Annu. Rev. Biomed. Data Sci. 1:207–34

Identification of peptides by MS/MS and database search



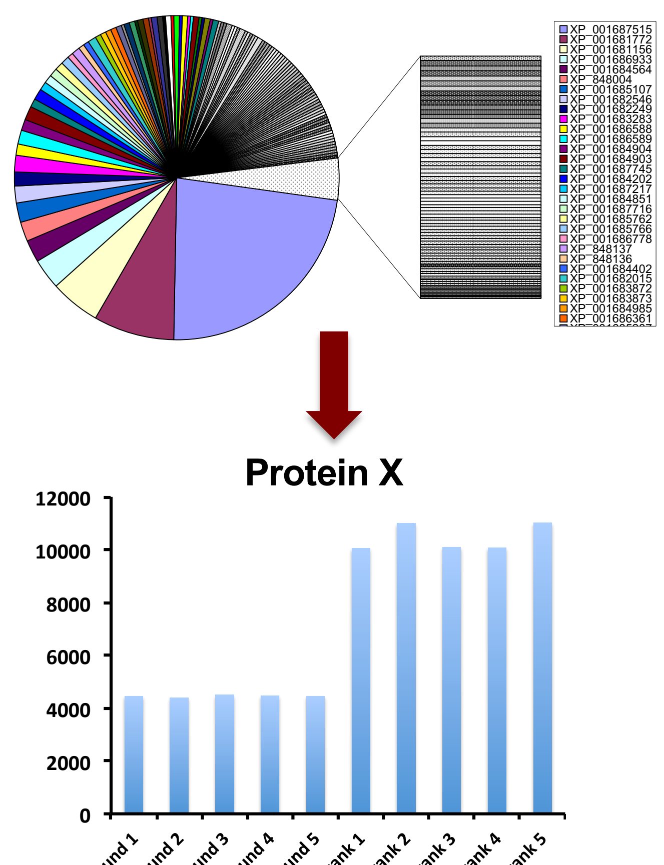
Identification of proteins by means of detected peptides



Comparison of fragment ion spectra (peptide information) with the (protein) database (*in silico* digestion)

What information does *bottom-up* proteome analysis provide?

ppm

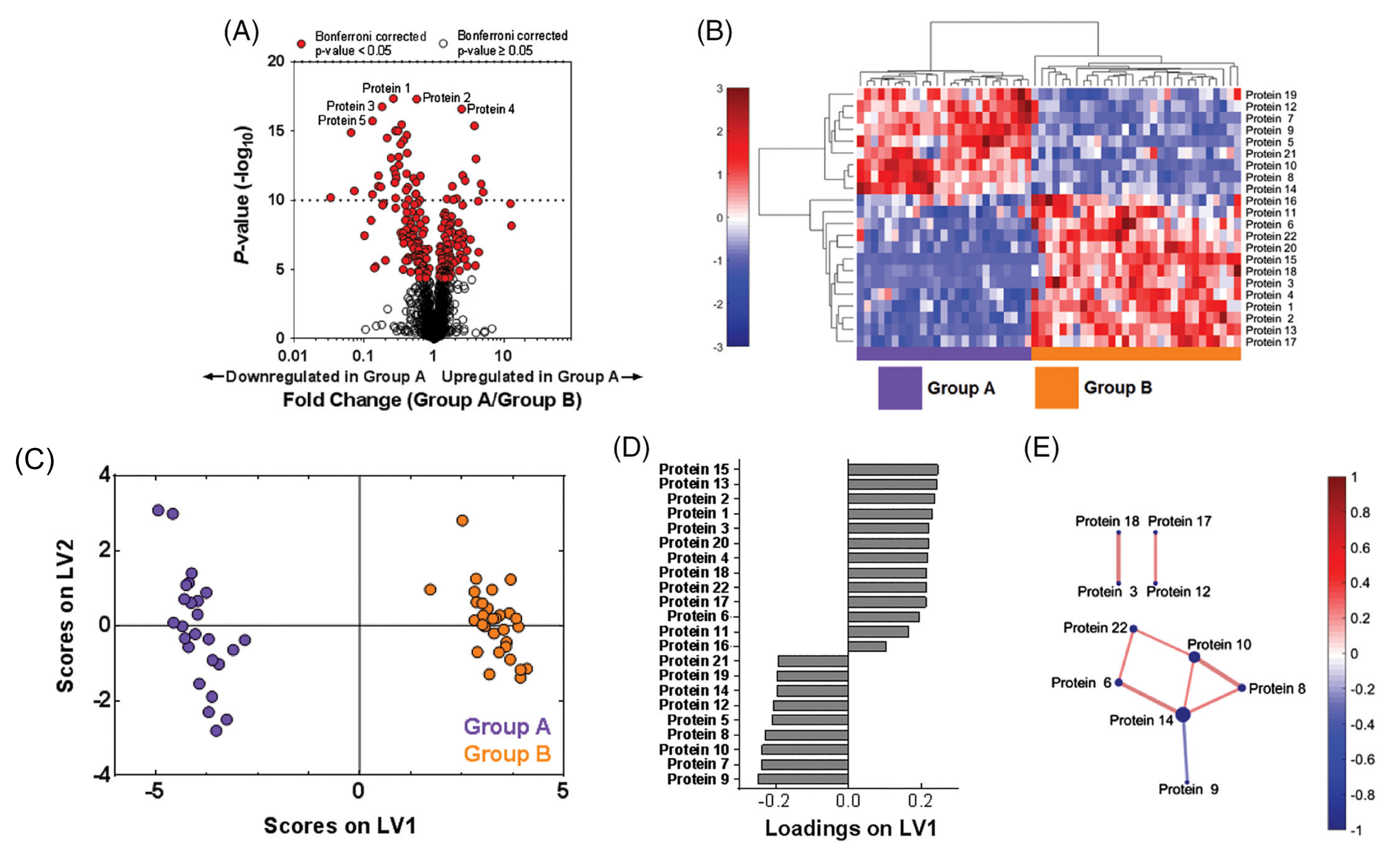


→ Type and quantity of proteins contained in a sample

→ Up to 5,000-10,000 proteins are identified per measurement

→ Differentiation between different samples

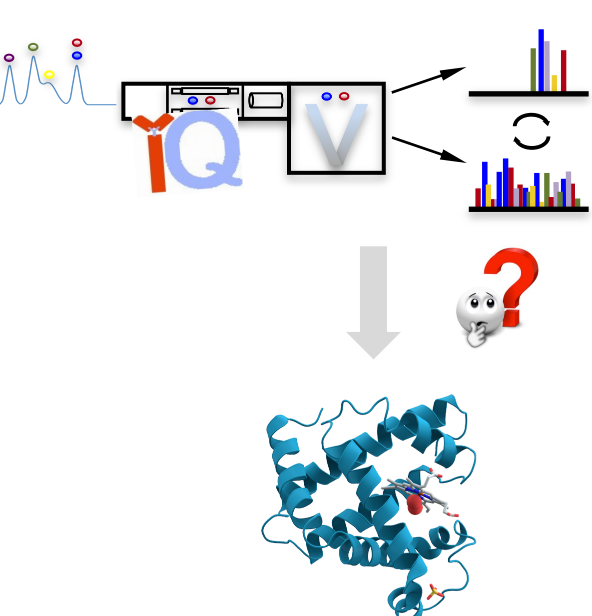
What information does *bottom-up* proteome analysis provide?



Respirology, Volume: 23, Issue: 11, Pages: 993-1003, First published: 13 August 2018, DOI: (10.1111/resp.13383)

Introduction to Proteomics

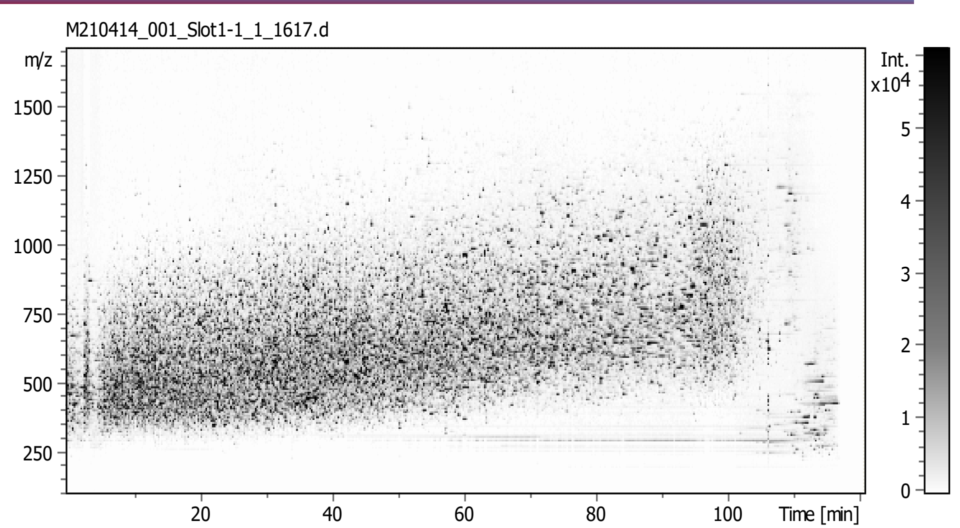
- What is proteomics? And why do we do this?
- Mass spectrometry-based proteomics
  - Sample preparation
  - Protein Identification – „from mass spectrum to protein“
  - Acquisition modes



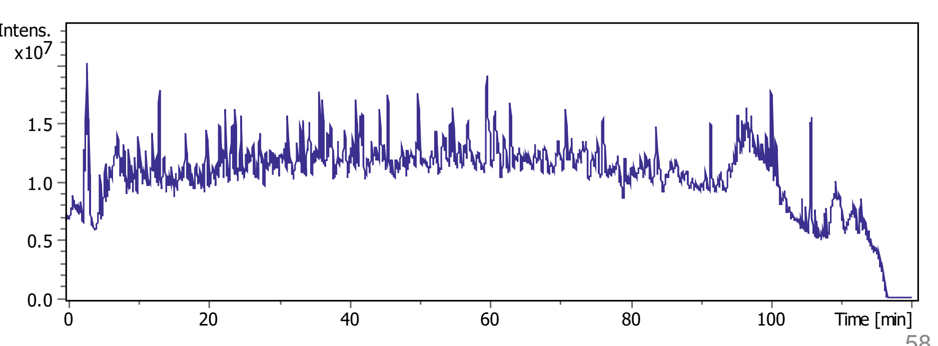
LC-MS analysis of complex proteomic samples

200 ng HeLa Lysate
2h nanoLC-MS/MS
(TIMS-TOF-Pro)

Heatmap or "gel-like"



Total Ion chromatogram (TIC)



Data dependent acquistion, DDA

→ *Selection of analytics based on your intensity*

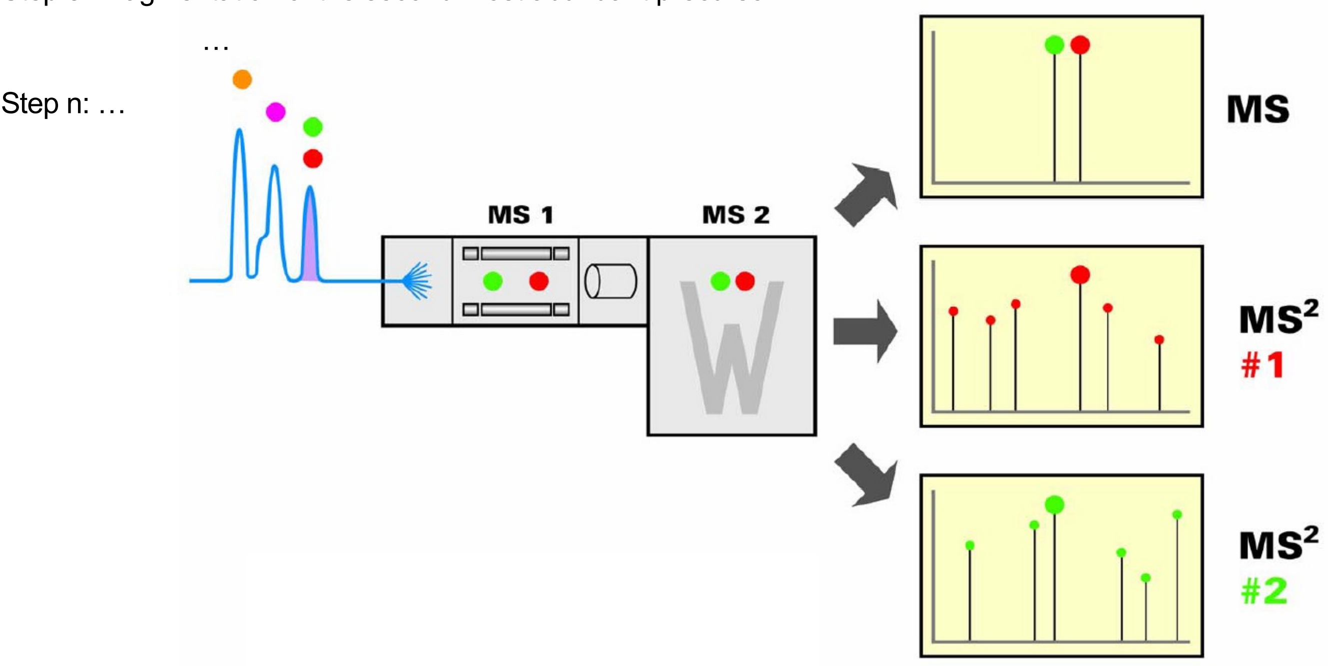
Step 1: Detection of precursor ions (intact peptides)

Step 2: Fragmentation of the most abundant precursor

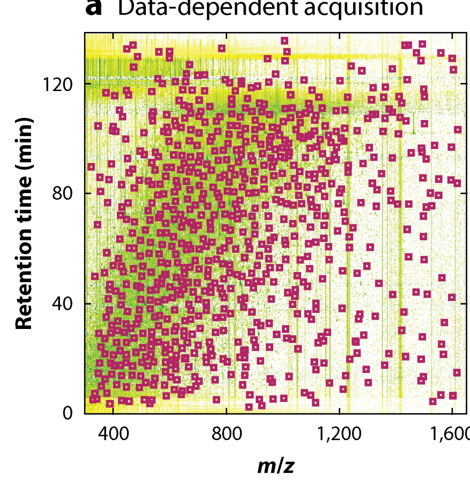
Step 3: Fragmentation of the second most abundant precursor

…

Step n: …



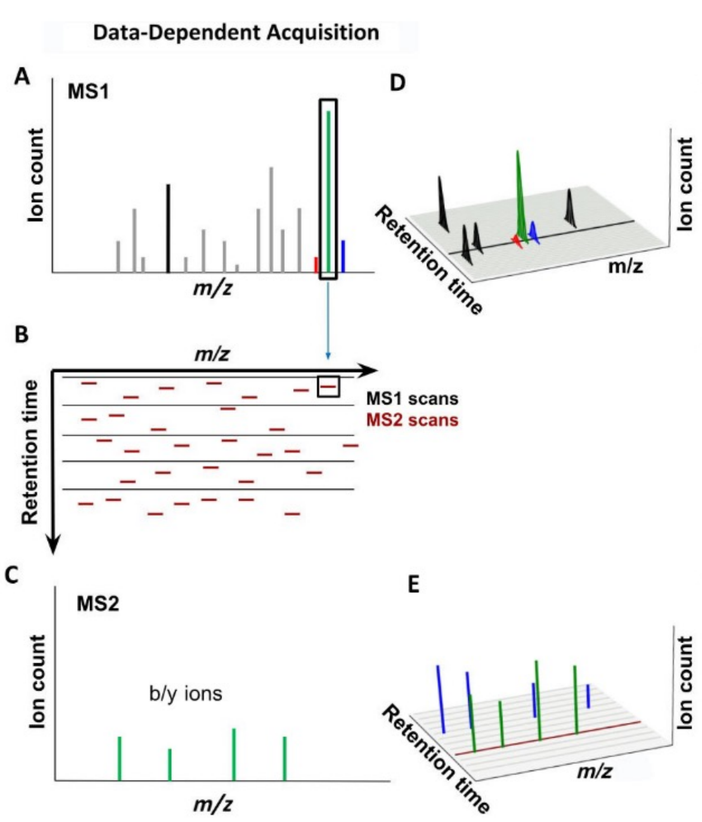
Data dependent acquistion
- DDA -



Sinitcyn P, et al. 2018.
Annu. Rev. Biomed. Data Sci. 1:207–34

stochastic/biased
Easy mapping of
precursor and fragment ions

Data dependent acquistion
- DDA -



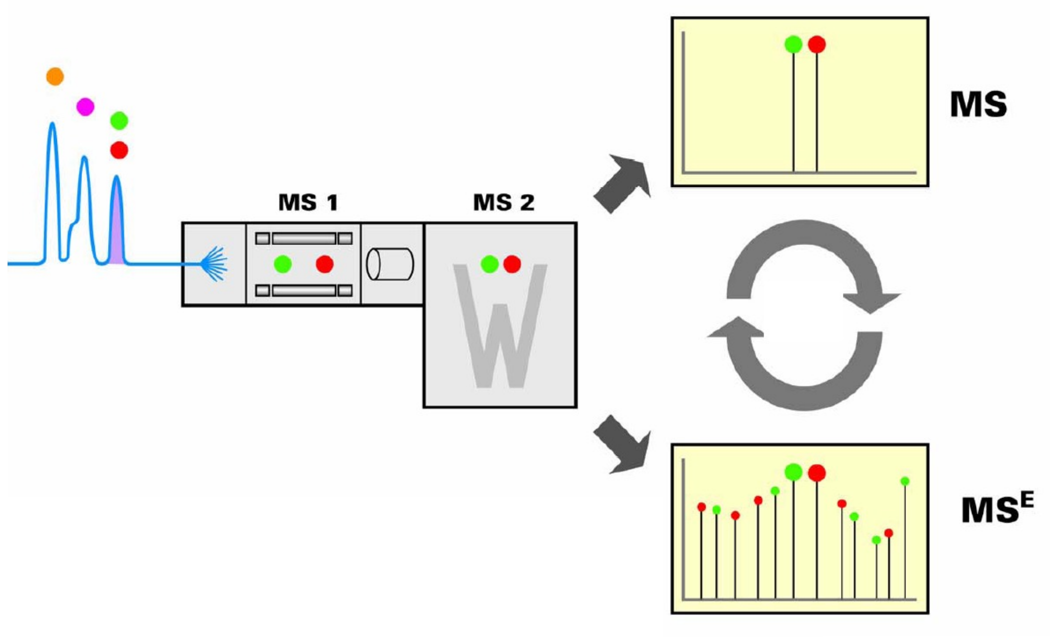
- Fragmentation of top N ions = high quality spectra
- Narrow selection window = high specificity
- But still some co-fragmentation
- Serial, biased and discontinuous process

Pappireddi *et al*. *ChemBioChem*. 2019;20(10):1210-1224.

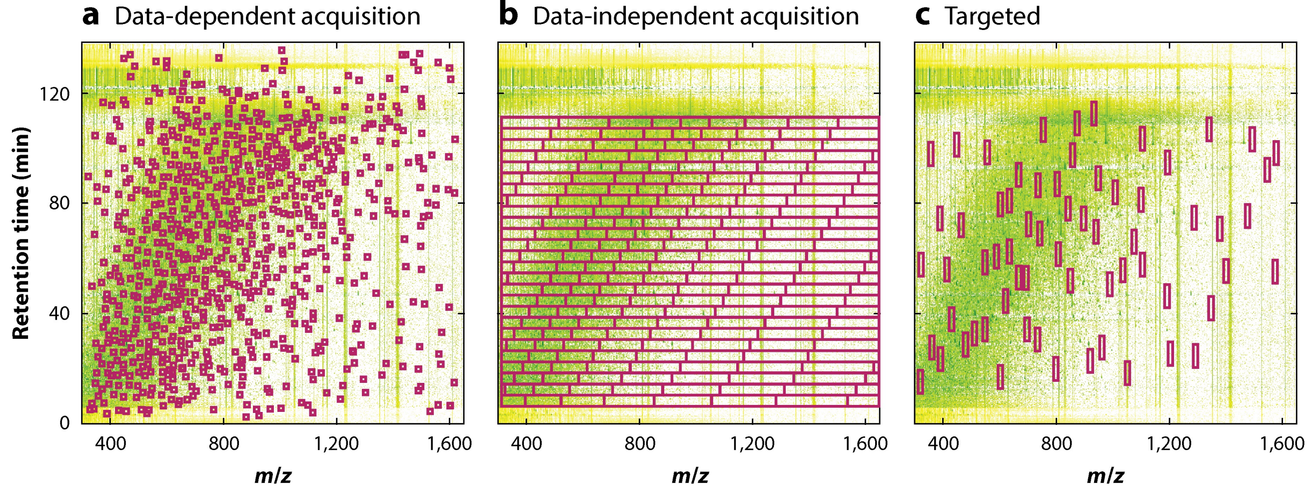
Data-independent acquistion, DIA

Step 1: Detection of precursor ions (intact peptides)

Step 2: Parallel fragmentation of multiple or all precursors



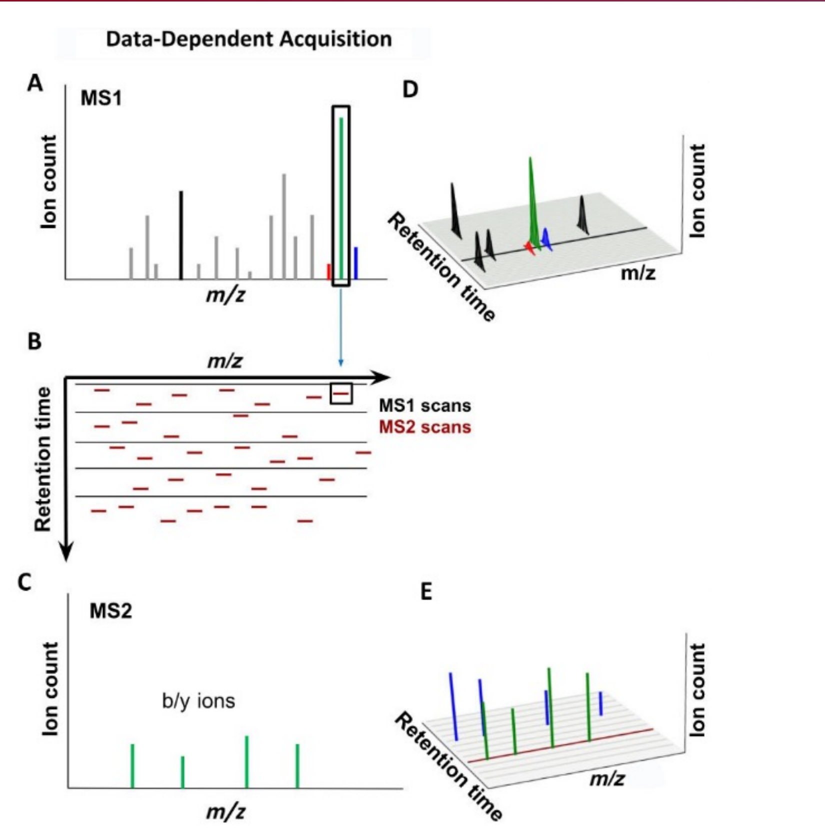
Different modes of acquisition
- Summary -

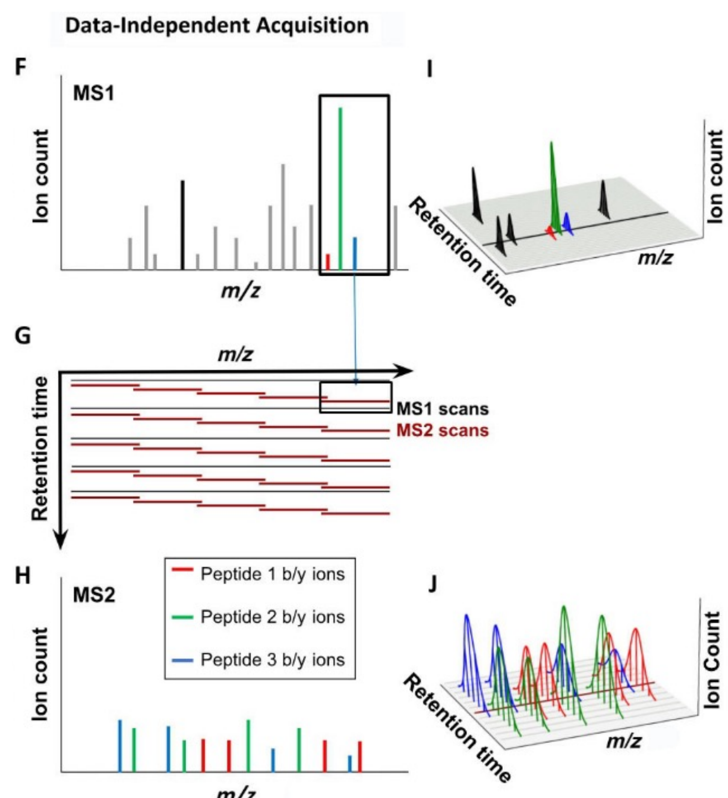


stochastic/biased
Easy mapping of
precursor and fragment ions

unbiased
complex fragment ion spectra
data processing challenging

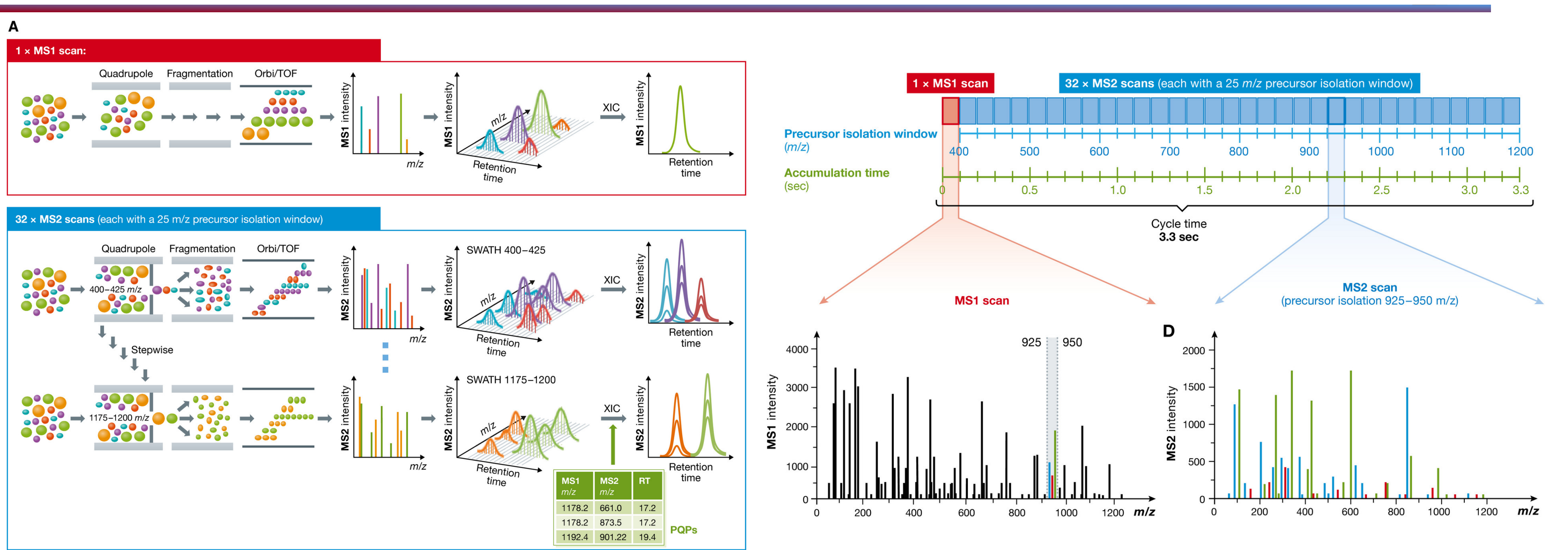
Data Dependent Acquisition (DDA) vs Data Independent Acquisition (DIA)





Pappireddi *et al*. *ChemBioChem*. 2019;20(10):1210-1224.

Data Independent Acquisition (DIA)



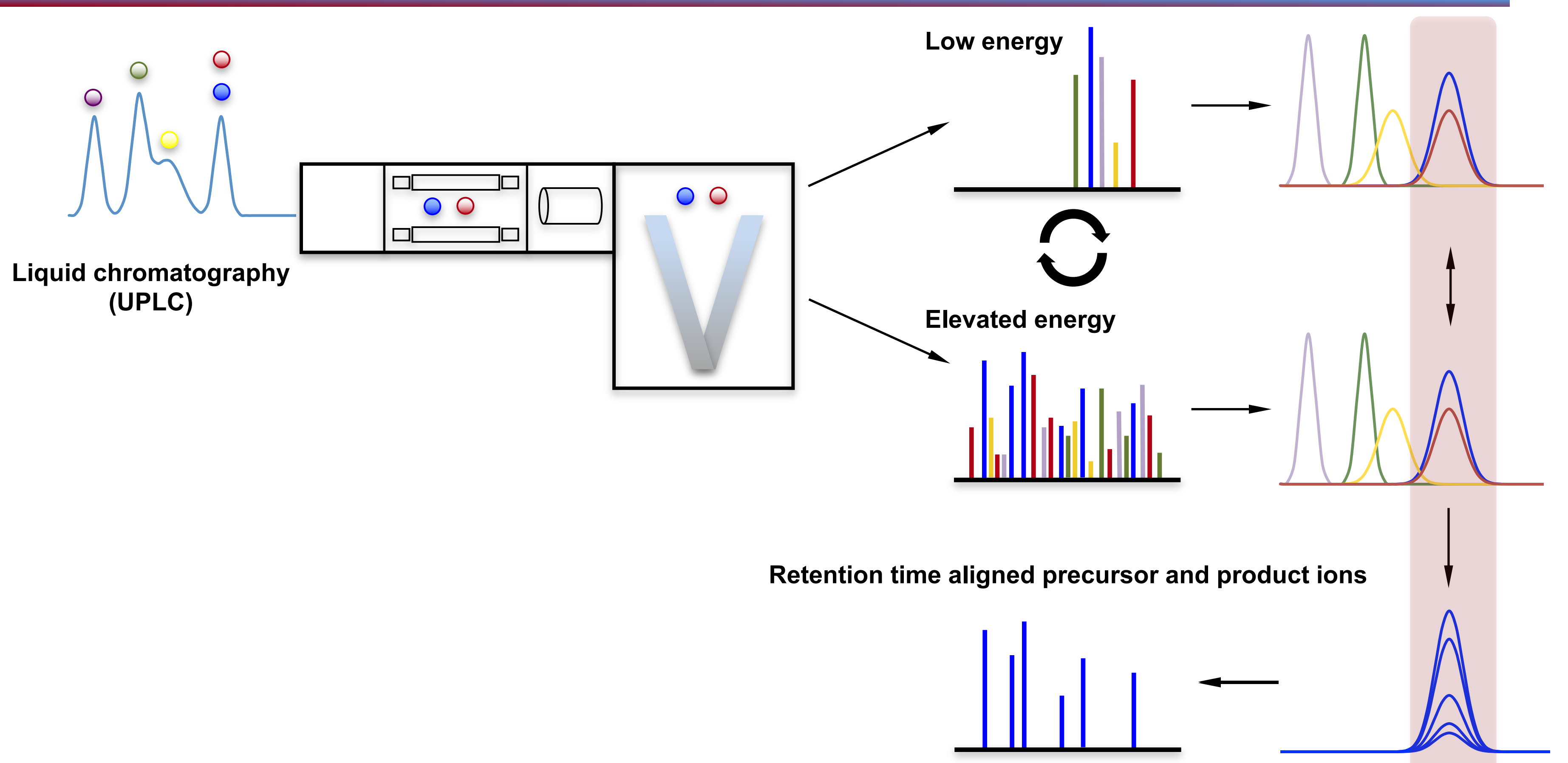
MS1 m/z	MS2 m/z	RT
1178.2	661.0	17.2
1178.2	873.5	17.2
1192.4	901.22	19.4

© EMBO

- Fragment everything within the wide window (e.g. 25 *m/z* vs <= 1 *m/z* in DDA)
- Non-biased by precursor selection
- Generates chimeric MS2 spectra, harder to deconvolute than DDA fragment ion spectra

Mol Syst Biol, Volume: 14, Issue: 8, First published: 13 August 2018, DOI: (10.15252/msb.20178126)

Deconvolution of DIA fragment ion spectra



Ion mobility as additional dimension of separation

Heatmap - G210417_001_Slot1-5_1_1450.d: +MS, 0.0-120.0 min #1-67463

1^+ → **noise**

2-4^+,... → ions of interest

Orthogonal Accelerator

Funnel 2

Funnel 1

Collision Cell

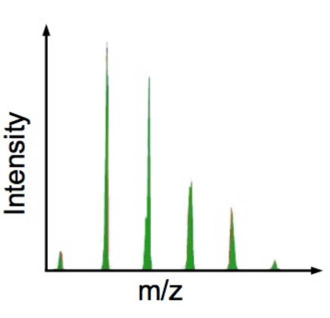
Quadrupole Filter

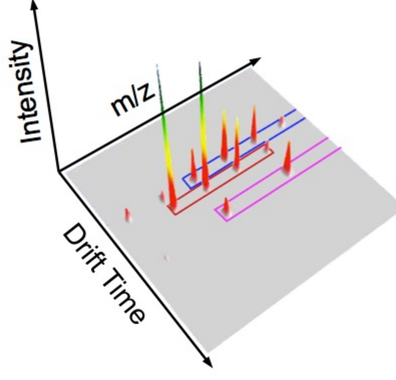
Hexapole

Capillary

TIMS Analyzer

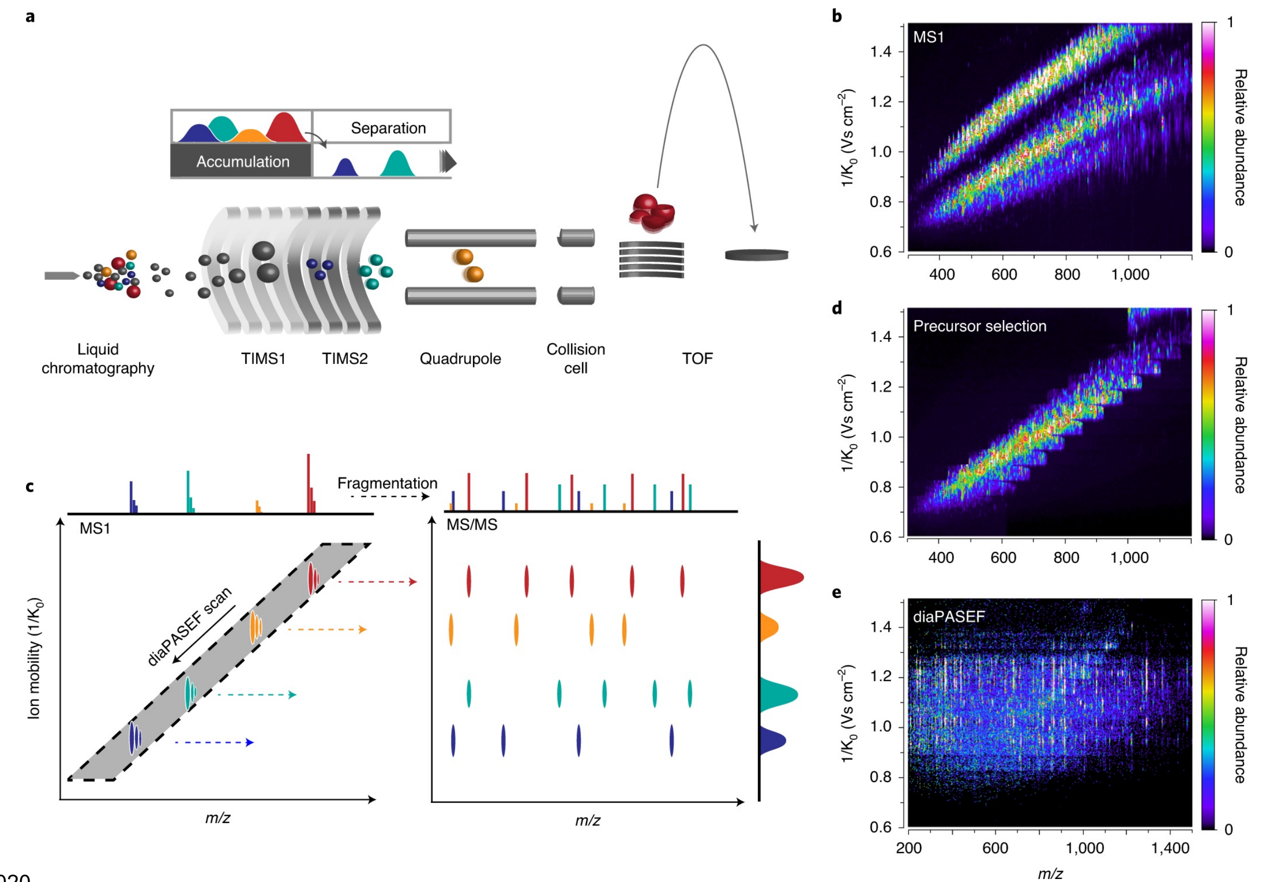
timsTOF Pro
(source: www.bruker.com)





Distler U, Kuharev J, Tenzer S. ***Expert Rev Proteomics.*** 2014

Ion mobility as additional dimension of separation - DIA – PASEF -

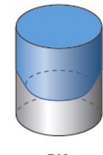


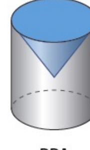
Meier F. *et al.*. ***Nature Methods.*** 2020

Different modes of acquisition
- Summary DDA vs DIA -

	Data independent acquisition-based SWATH-MS	Data-dependent acquisition (DDA)
Ease of data acquisition	** Easy, requires definition of mass range to cover, precursor isolation window width and number of MS2 scans per cycle	*** Easiest, default setup on most mass spectrometers, requires definition of TopN method, MS2 trigger threshold and dynamic exclusion time
Ease of data analysis	Getting easier, several pipelines available, becoming the gold standard for high-throughput proteome analysis	*** Currently easiest, multitude of pipelines available
Breadth of protein and peptide detection/multiplexing	*** 10,000s of peptides per MS injection quantifiable	*** 10,000s of peptides per MS injection quantifiable
Selectivity/sensitivity/dynamic quantification range	** 4 orders of magnitude per MS injection	** 4 orders of magnitude per MS injection
Reproducibility/data consistency	*** High, due to peptide-centric scoring analysis	* Low, due to stochastic sampling in DDA
Retrospective targeting (using chromatogram extraction)	*** Possible on MS1 and MS2 level	** Possible on MS1 level only

*Least optimal performance.
**Medium performance.
***Best performance.





[Gomez-Zepeda]

(Ludwig *et al.* 2018)

What should I know now?

- What is a proteome/proteomics?
- How can we analyse proteins using mass spectrometry?
- What are the challenges when analysing a proteome?
- What are the steps when preparing a sample for *bottom-up* proteomics analysis?
- How do we derive peptide sequence information from a mass spectrum?

Questions??

The proteome

It's a Fan!

It's a Spear!

It's a Wall!

It's a Rope!

It's a Snake!

It's a Tree!

The blind men and the elephant

PM Daigneault, *Methodological Innovations Online* 28(2):82-89; DOI:10.4256/mio.2013.015