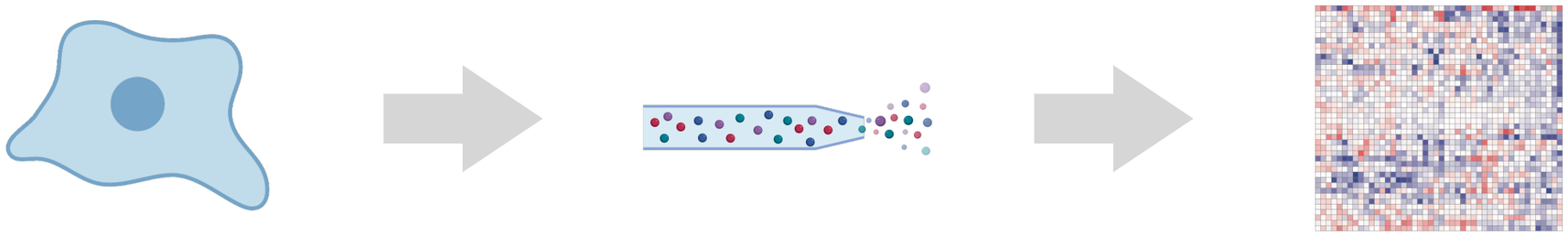
An Introduction to Mass-Spectrometry-Based Proteomics

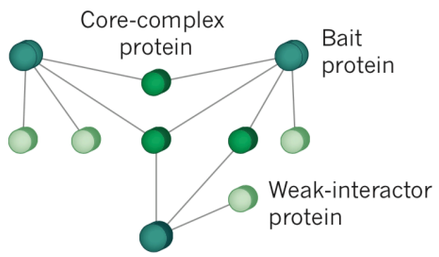


Nick Till

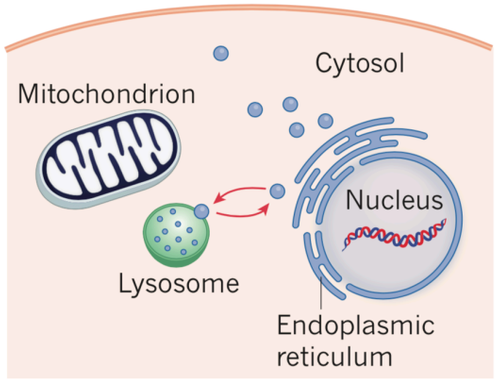
MacMillan Group Meeting 07/29/2020

The Broader Context for MS-Based Proteomics

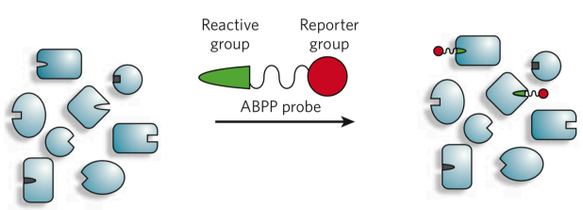
Protein-Interaction Networks



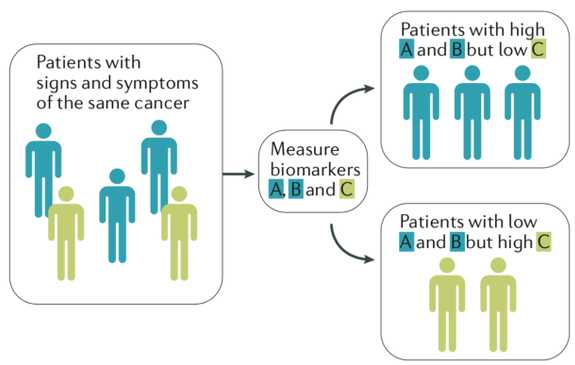
Subcellular Localization



Activity-based Profiling



Biomarker Discovery



The Broader Context for MS-Based Proteomics

DNA → *RNA* → *protein*

genomics (DNA sequencing)	**transcriptomics** (RNA-seq)	**proteomics** (MS analysis)

proteomics: the study of proteomes and their functions (or the large scale study of proteins)

protein measurements can be direct read-outs of biological activity

The Broader Context for MS-Based Proteomics

DNA

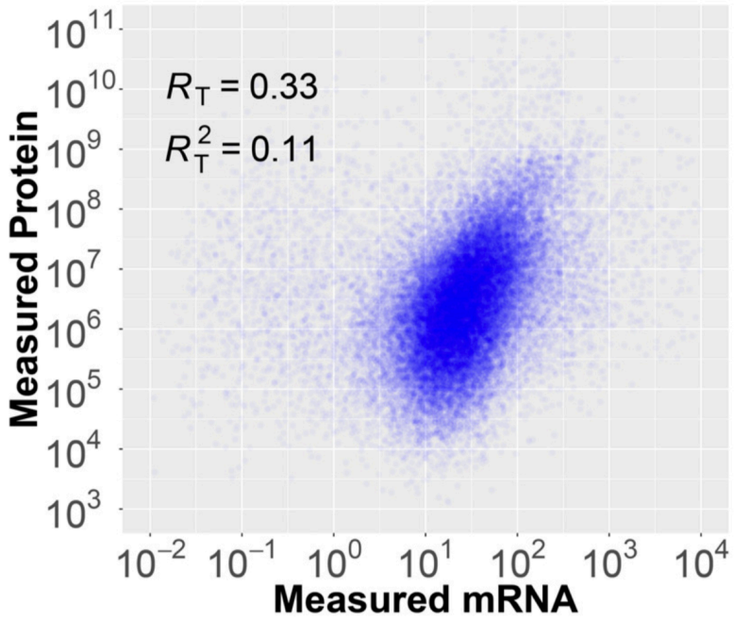
RNA

protein

RNAseq: transcript levels with good coverage

transcript levels can be misleading

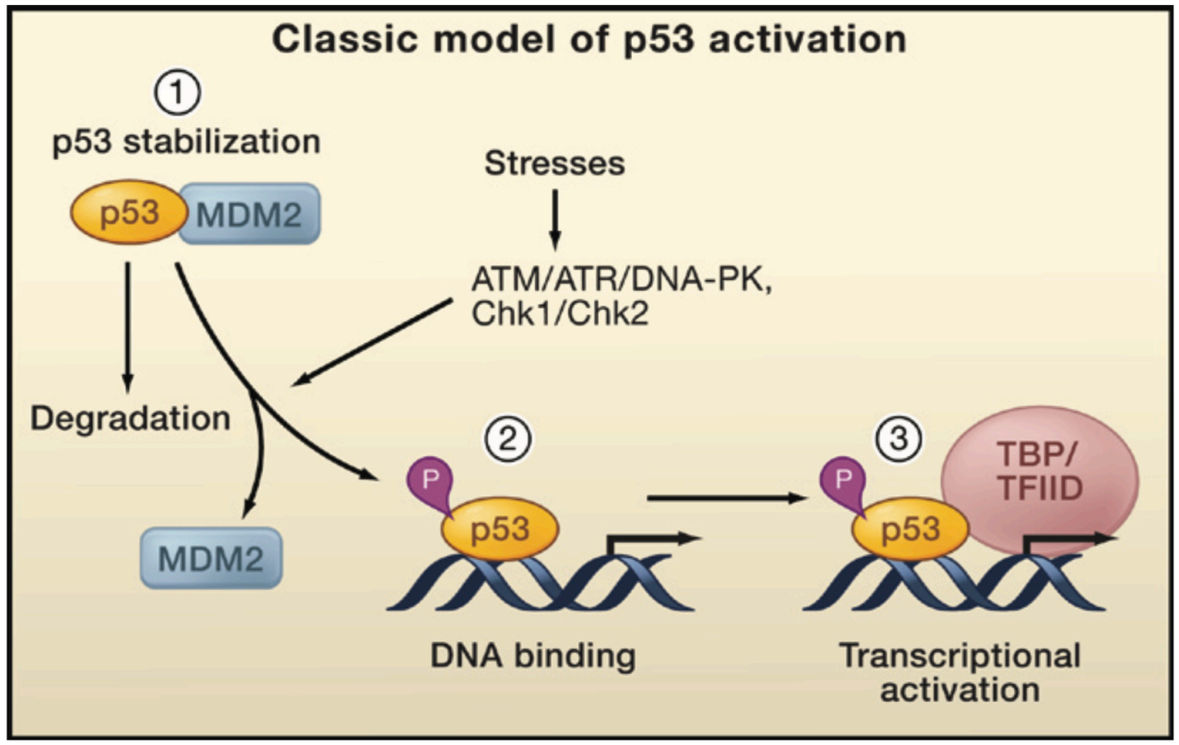
in predicting protein levels



- translational efficiency
- post-translational modifications
- protein degradation kinetics can vary

Franks, A.; Airoldi, E.; Slavov, N. *PLOS Computational Biology* **2017**, *13*, e1005535.

The Broader Context for MS-Based Proteomics



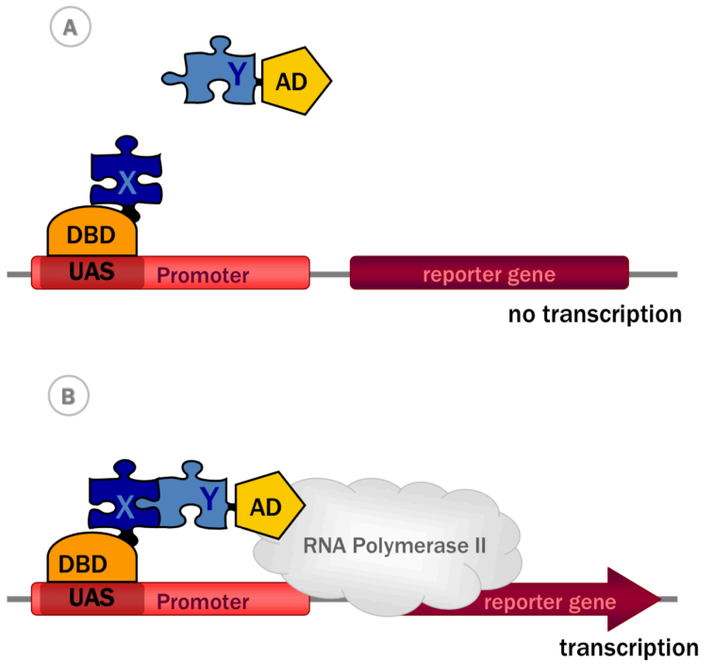
p53 regulation: mediated by phosphorylation and proteasomal degradation

p53 mRNA transcript levels are a poor indicator of p53 levels and activity

Kruse, J.-P.; Gu, W. *Cell* **2009**, *137*, 609–622.

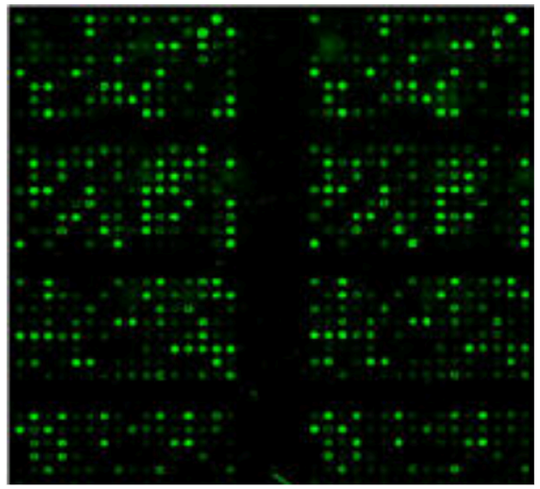
The Broader Context for MS-Based Proteomics

Large-Scale Yeast 2-Hybrid



- large-scale identification of PPIs

Protein Microarrays



- large-scale protein-binding studies

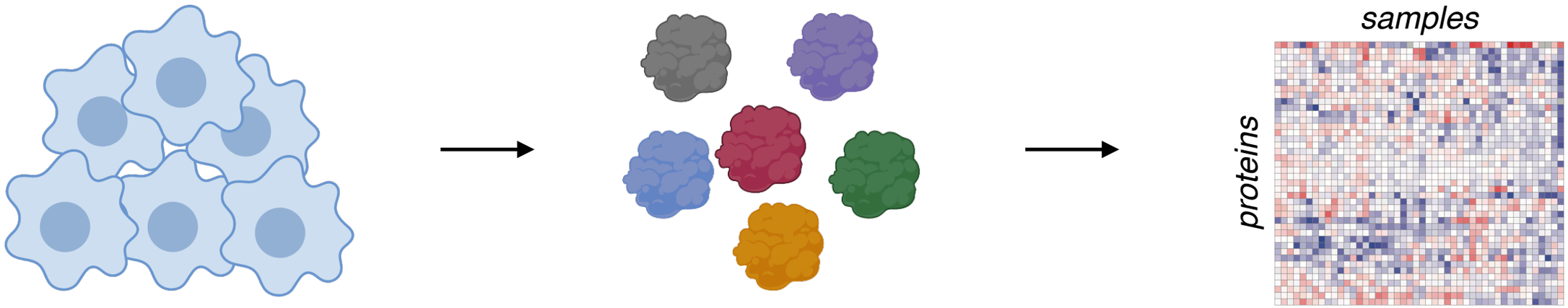
MS-based proteomics: often the tool of choice for large-scale analysis of protein levels and interactions

Sutandy, F. X. R.; Qian, J.; Chen, C.-S.; Zhu, H. *Curr. Protoc. Protein Sci.* **2013**, *72*, 27.1.1-27.1.16.
Brückner, A.; Polge, C.; Lentze, N.; Auerbach, D.; Schlattner, U. Y. *Int. J. Mol. Sci.* **2009**, *10*, 2763–2788.

An Overview of Topics Covered

- Part 1: basic workflow and technology for discovery proteomics
  - How proteins are handled and analyzed
  - Data Peptide assignment and protein inference
- Part 2: methods for (relative) quantitative proteomics
  - Label-free methods
  - Whole-cell isotopic labeling strategies
  - Chemical mass tags
- Part 3: targeted proteomics and its application to biomarker discovery

Technology Development and Workflow for MS Proteomics

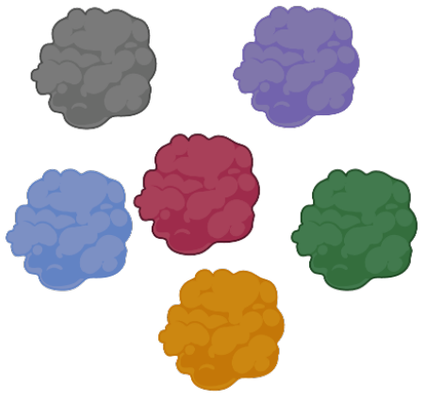


cells or tissue sample

proteins

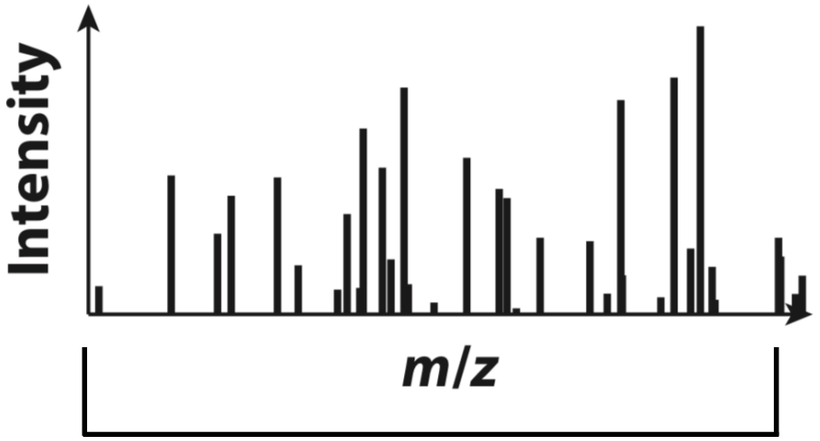
proteome analysis

Technology Development and Workflow for MS Proteomics

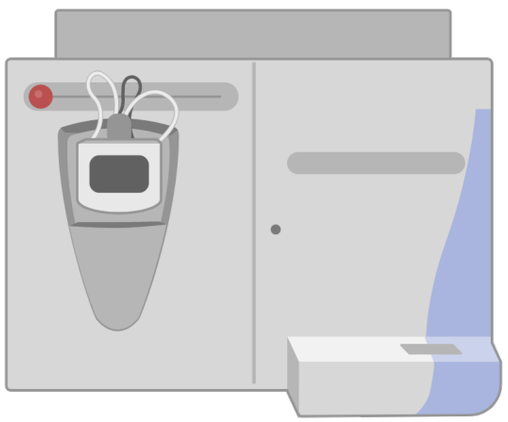


~ 20,000 proteins

(not counting PTMS, alternative splicing)



window ≈ 1500 m/z



direct injection results in high complexity MS

separation step necessary prior to MS analysis

Technology Development and Workflow for MS Proteomics

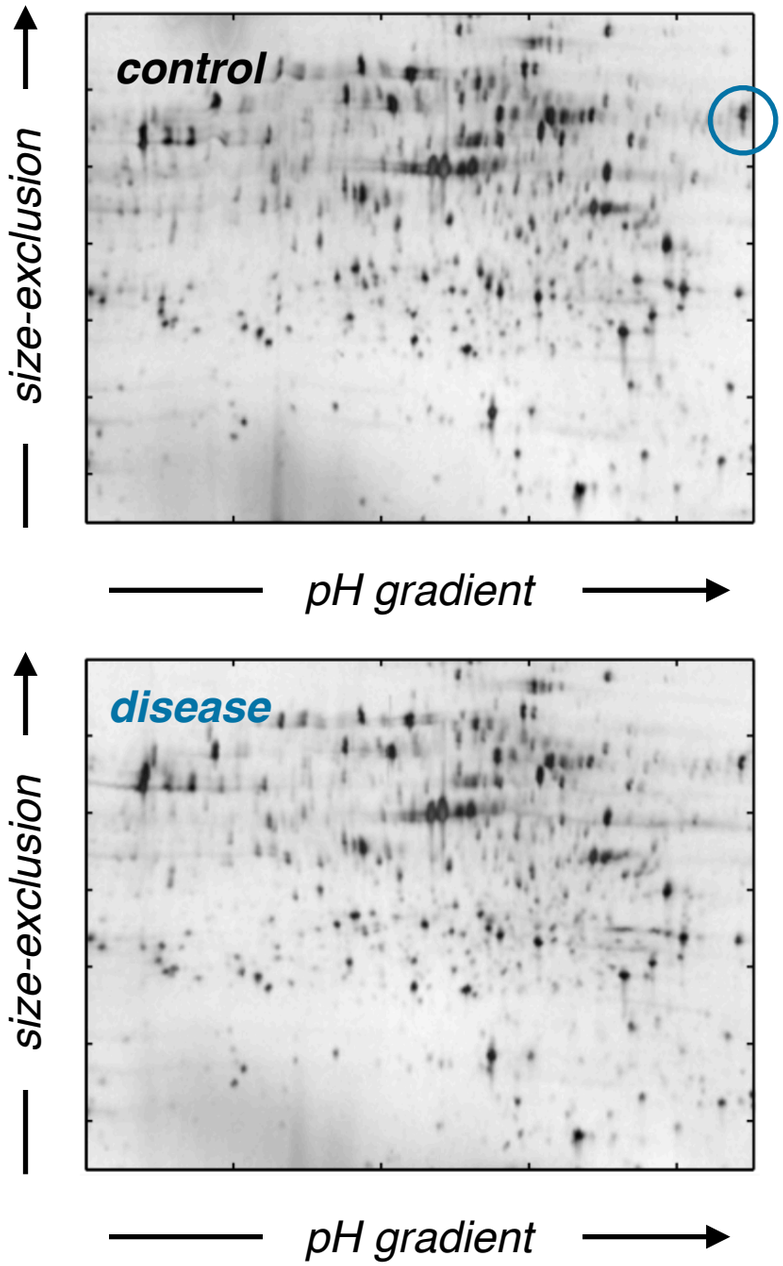
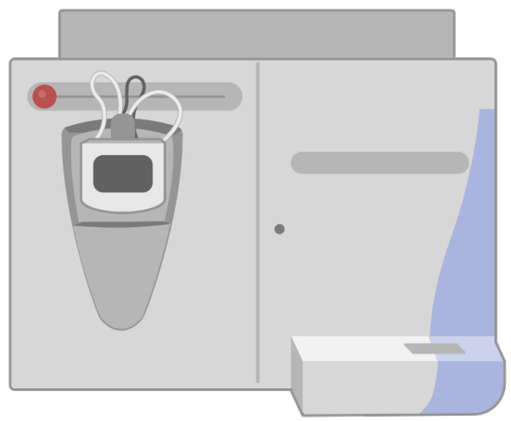


image analysis identifies up/down regulation

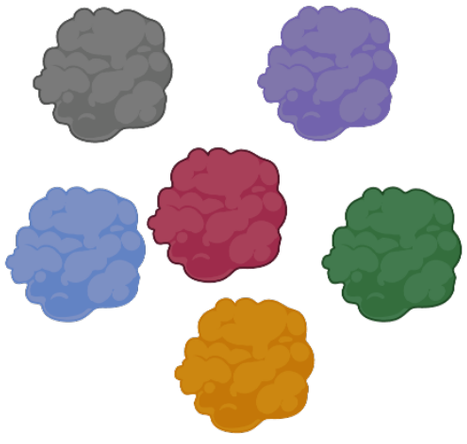
isolate spot from gel

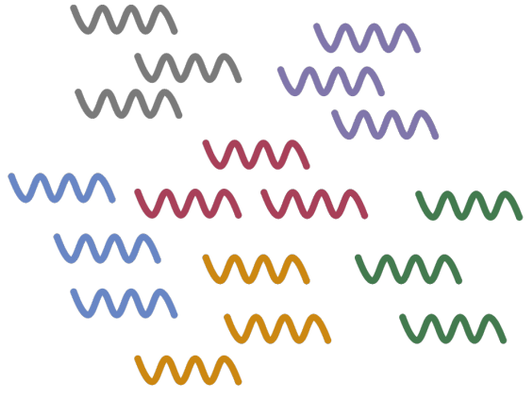


protein ID by MS

Li, F.; Seillier-Moiseiwitsch, F.; Korostyshevskiy, V. R. *Computational Statistics & Data Analysis* **2011**, *55*, 3059–3072.

Technology Development and Workflow for MS Proteomics





MS analysis of peptides

"bottom-up proteomics"

- protein MW > 10,000 Da
- poor recovery by LC

- peptide MWs < 4,000 Da
- good recovery across peptides

peptide fragments

(incomplete)

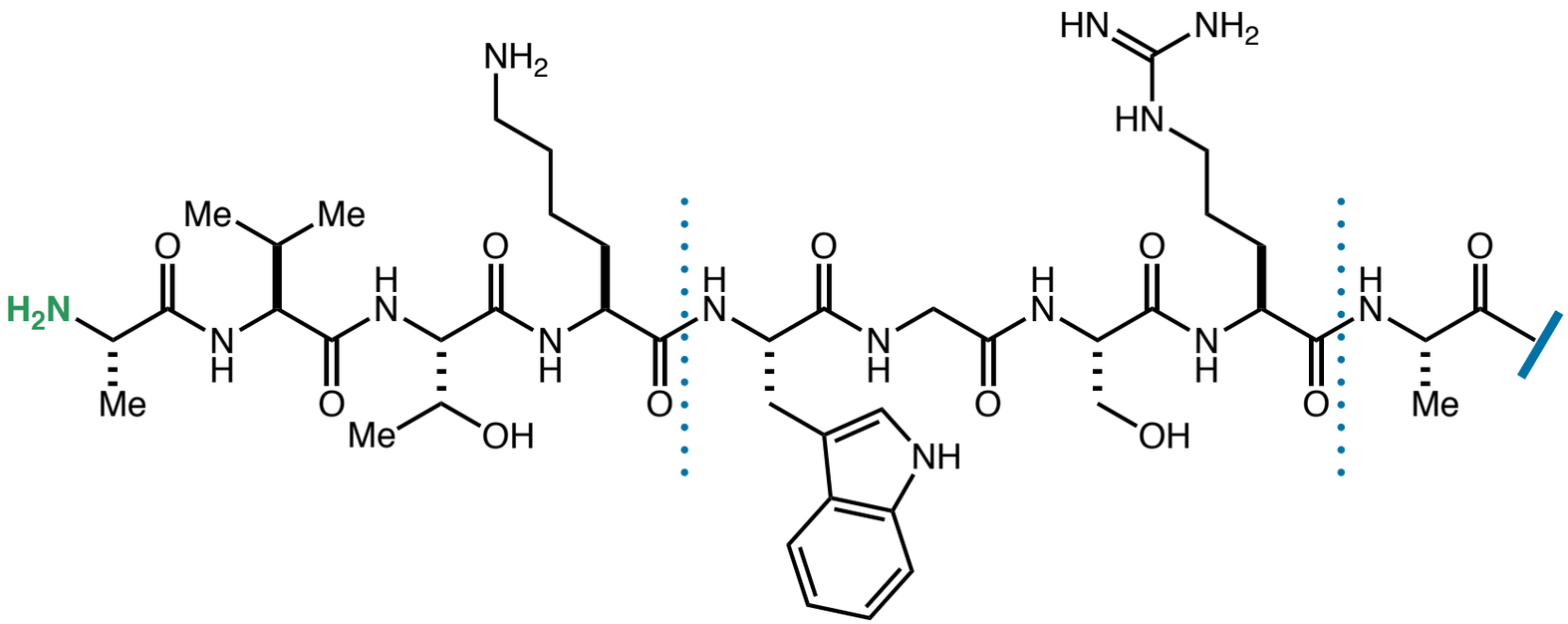
protein identity

data analysis involves

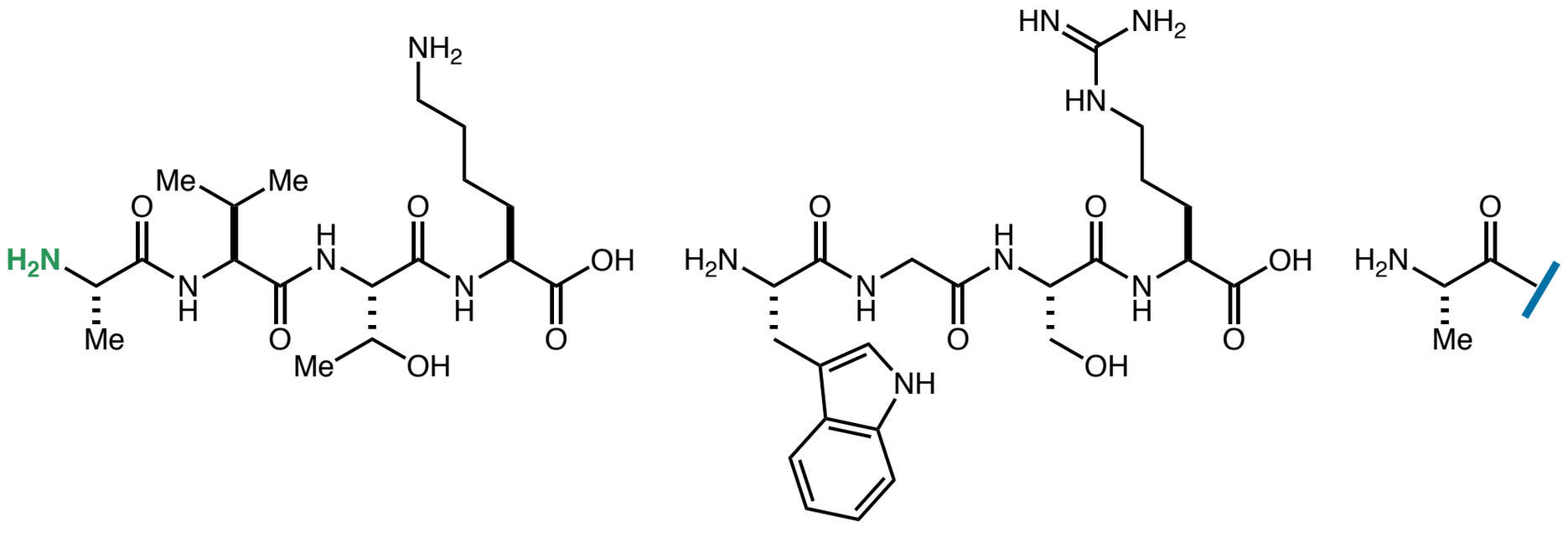
reconstructing protein

(protein inference)

Technology Development and Workflow for MS Proteomics



trypsin cleavage: high-specificity serine protease cleaves after K (lysine) or R (arginine) residues



Gundry, R. L.; White, M. Y.; Murray, C. I.; Kane, L. A.; Fu, Q.; Stanley, B. A.; Eyk, J. E. V. *Current Protocols in Molecular Biology* **2010**, *90*, 10.25.1-10.25.23.

Technology Development and Workflow for MS Proteomics

N-terminus–AVT**K**WGS**R**AGPAVT**K**EIGAASTQV**R**AGDSLQP**K**GTVALE**R**

N-terminus–AVT**K**

WGS**R**

AGPAVT**K**

EIGAASTQV**R**

AGDSLQP**K**

GTVALE**R**

} *tryptic peptides*

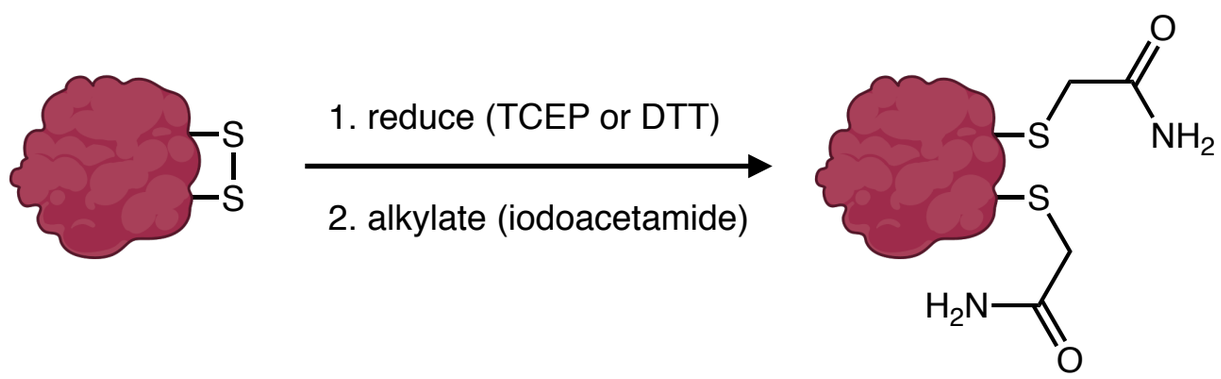
large mixture of tryptic peptides are then subjected to LC/MS² analysis

protease	cleavage specificity[a]
trypsin	-K,R-↑-Z- not -K,R-↑-P-
endoproteinase Lys-C	-K-↑-Z-
chymotrypsin	-W,F,Y-↑-Z- and -L,M,A,D,E-↑-Z- at a slower rate
subtilisin	broad specificity to native and denatured proteins
elastase	-B-↑-Z-
endoproteinase Lys-N	-Z-↑-K-
endoproteinase Glu-C	-E-↑-Z- and 3000 times slower at -D-↑-Z-
endoproteinase Arg-C	-R-↑-Z-
endoproteinase Asp-N	-Z-↑-D- and -Z-↑-cysteic acid- but not -Z-↑-C-
proteinase K	-X-↑-Y-
OmpT	-K,R-↑-K,R-

[a]B – uncharged, nonaromatic amino acids (i.e., A, V, L, I, G, S)

X – aliphatic, aromatic, or hydrophobic amino acids; and Z – any amino acid.

cysteine capping step often included

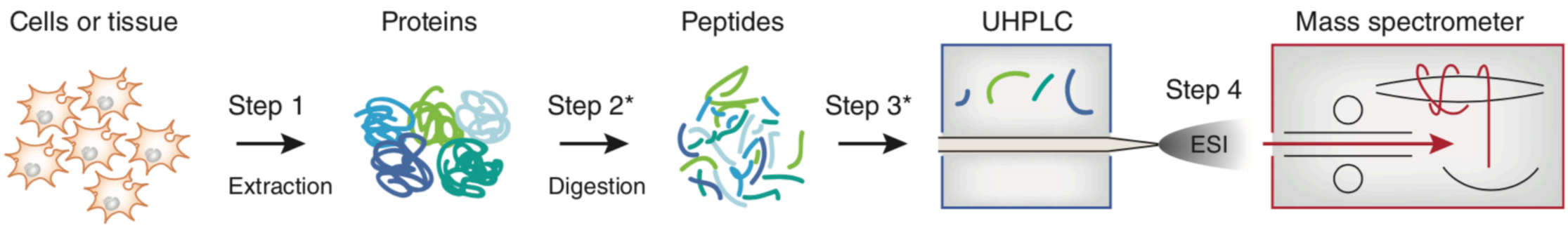


Zhang, Y.; Fonslow, B. R.; Shan, B.; Baek, M.-C.; Yates, J. R. *Chem. Rev.* **2013**, *113*, 2343–2394.

Gundry, R. L.; White, M. Y.; Murray, C. I.; Kane, L. A.; Fu, Q.; Stanley, B. A.; Eyk, J. E. V. *Current Protocols in Molecular Biology* **2010**, *90*, 10.25.1-10.25.23.

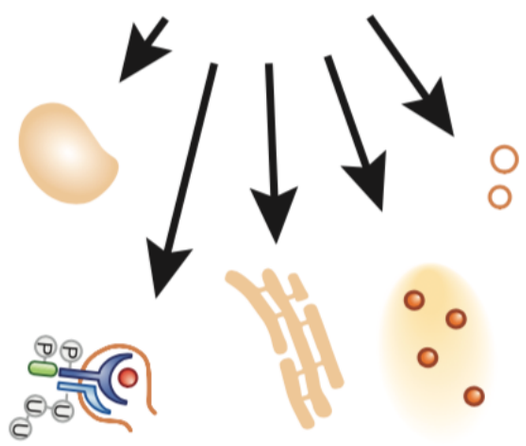
Technology Development and Workflow for MS Proteomics

MS-proteomics sample preparation and workflow

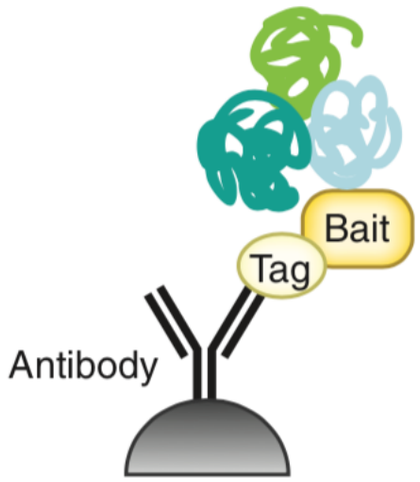


common context-specific modifications

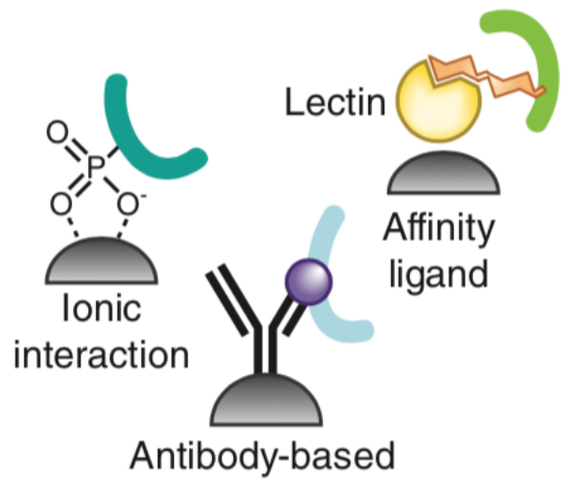
subcellular fractionation



protein interactions



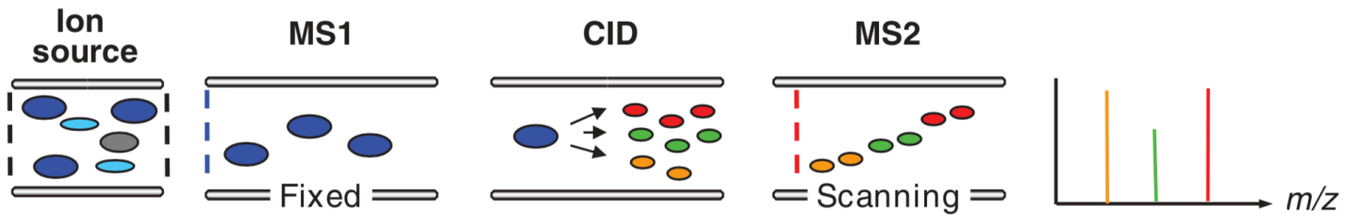
enrichment for PTMs



Meissner, F.; Mann, M. *Nature Immunology* **2014**, *15*, 112–117.

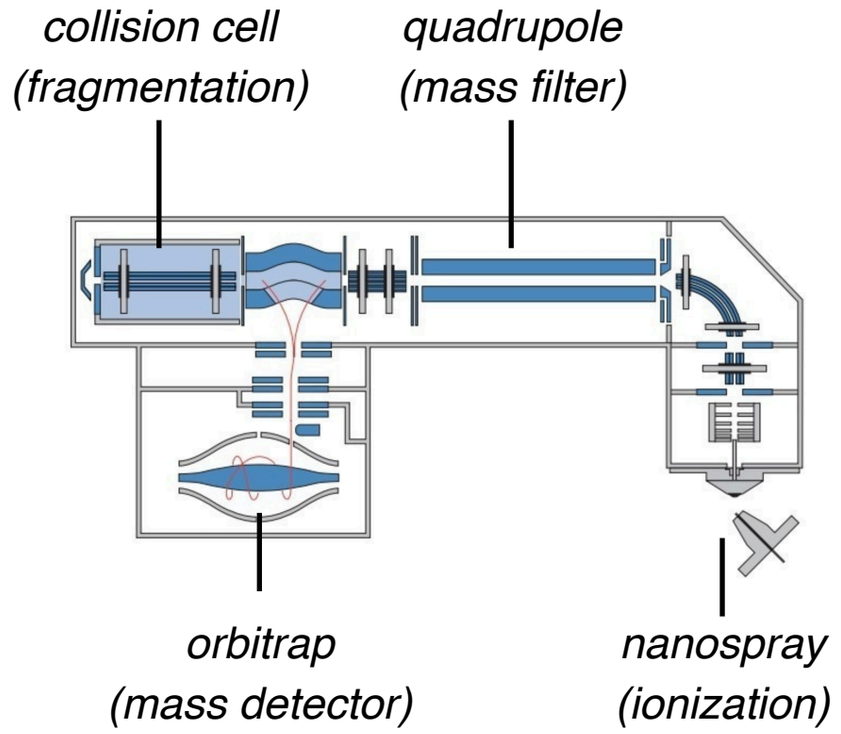
Technology Development and Workflow for MS Proteomics

MS/MS (MS^2) analysis: peptides are further fragmented into ions for sequence identification

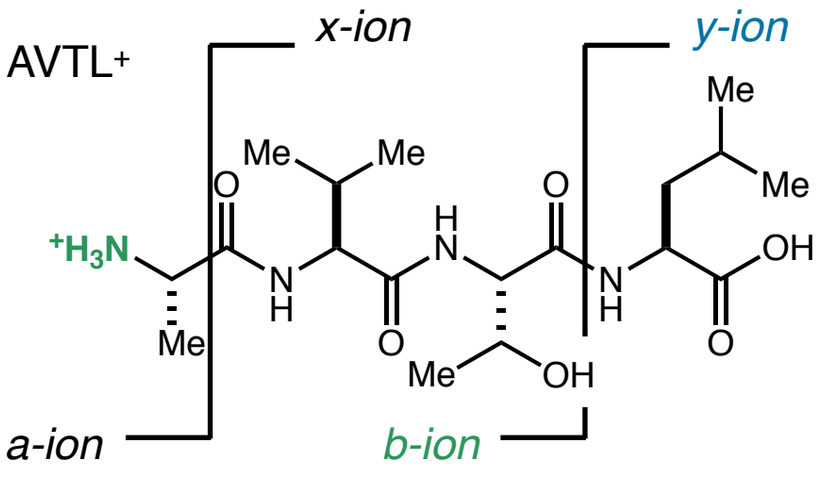


- multiple methods for MS/MS analysis
  - QqQ (triple quadrupole)
  - Q-TOF (quadrupole time-of-flight)
  - Q Exactive (quadrupole orbitrap)
- multiple modes of fragmentation
  - CID (collision-induced dissociation)
  - ECD (electron-capture-dissociation)
  - observed ions depend on method

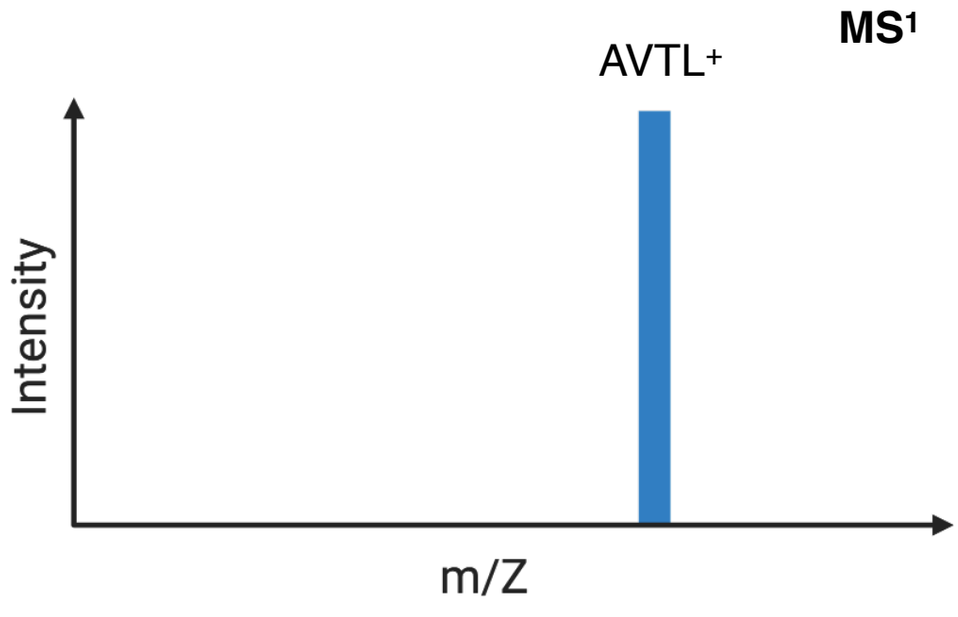
Q Exactive setup (Thermo Fisher)

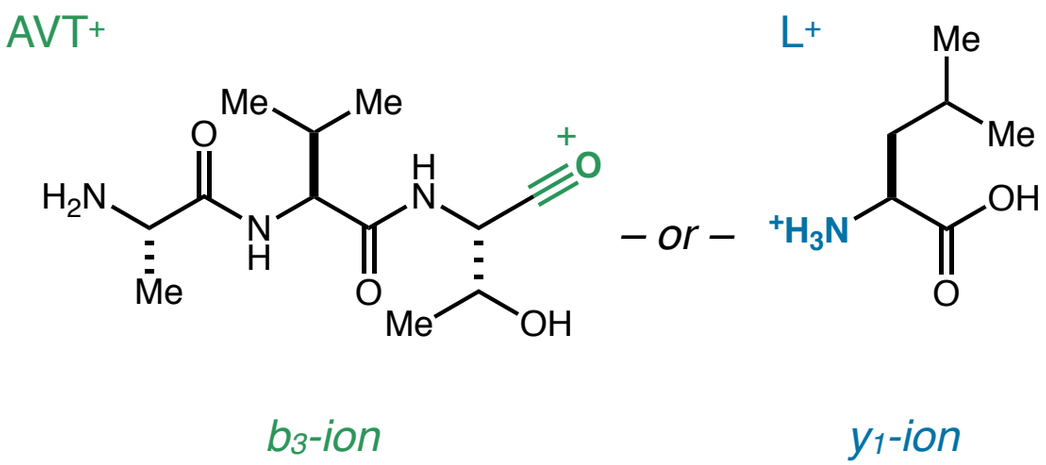


Technology Development and Workflow for MS Proteomics

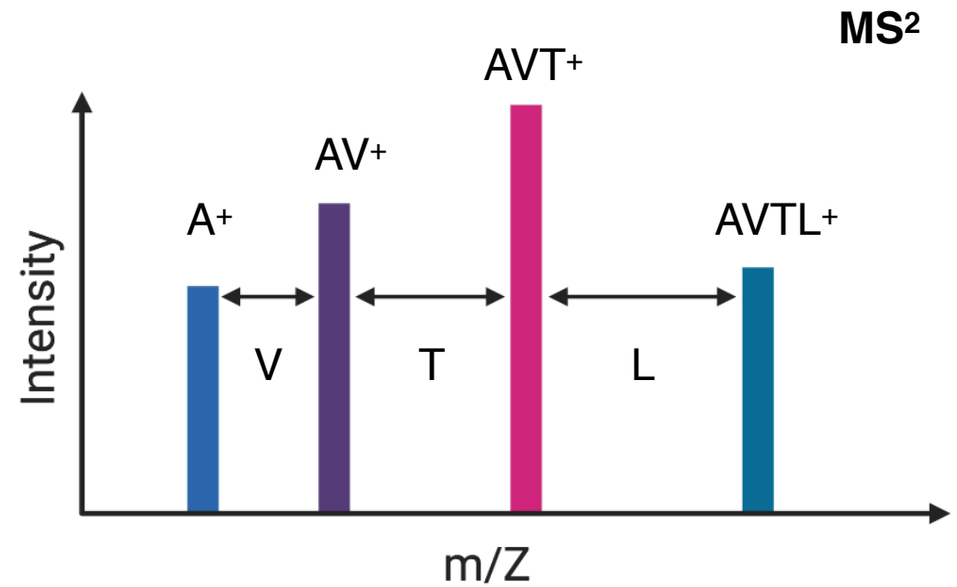


C–N cleavage





b₃-ion *y₁-ion*

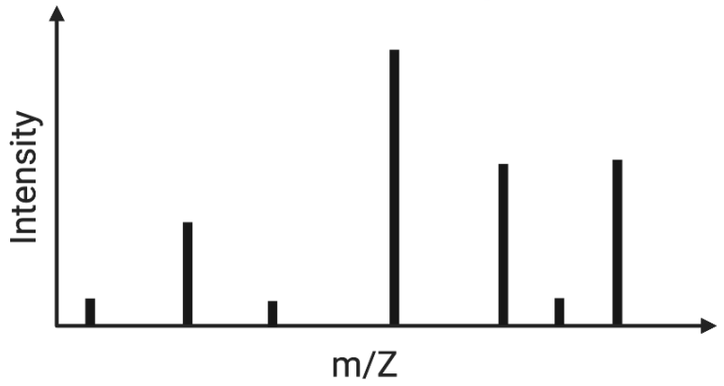


this is an idealized picture, in reality fragmentation is incomplete and requires more analysis for peptide ID

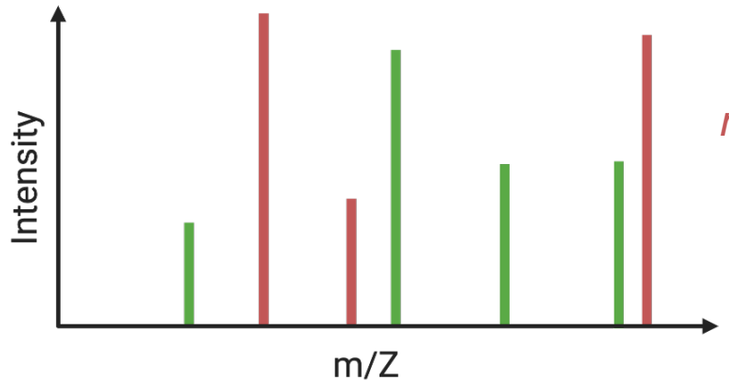
Paizs, B.; Suhai, S. Fragmentation Pathways of Protonated Peptides. *Mass Spectrometry Reviews* **2005**, *24*, 508–548.

Technology Development and Workflow for MS Proteomics

observed MS^2 spectrum



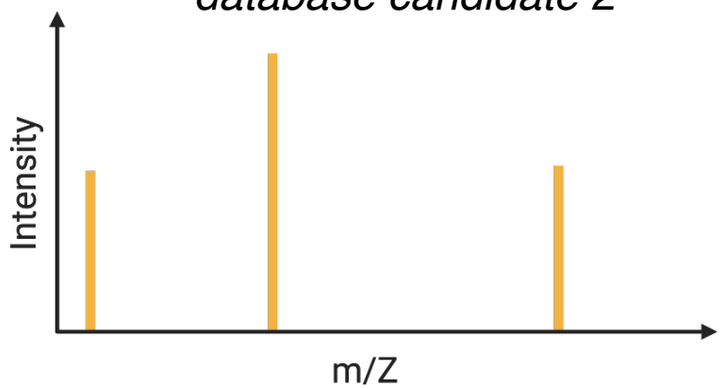
database candidate 1



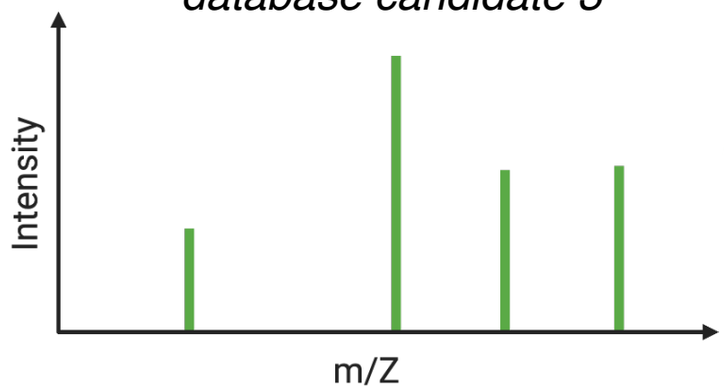
missing peaks penalized

database candidate 2

low intensity
low score



database candidate 3



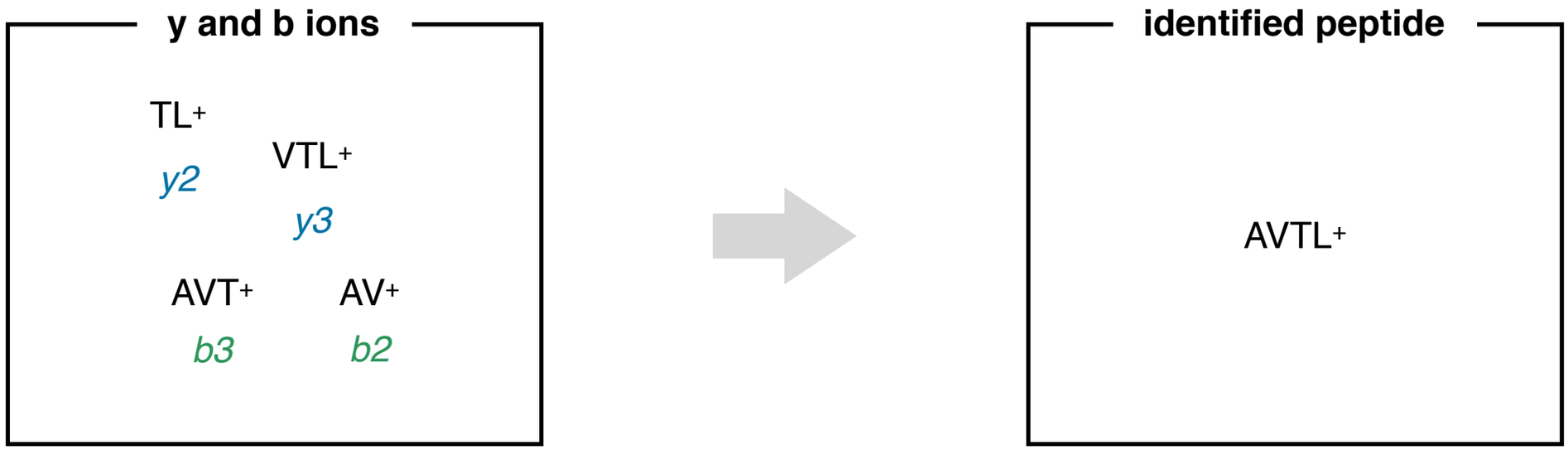
perfect match

observed peptide MS^2 spectra are scored against database MS^2 spectra to identify parent ion

- SEQUEST
- MASCOT
- OMSSA
- X!Tandem
- MaxQuant

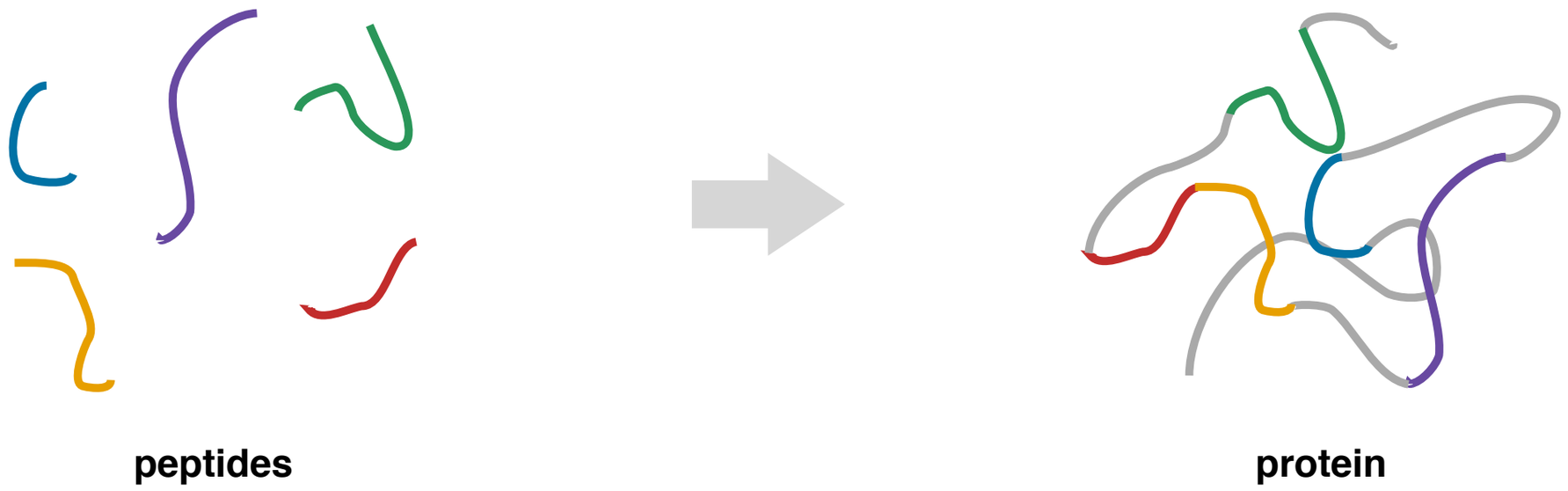
Eng, J. K.; McCormack, A. L.; Yates, J. R. *J. Am. Soc. Mass Spectrom.* **1994**, *5* , 976–989.

Technology Development and Workflow for MS Proteomics



In bottom-up MS-proteomics, peptides (*not proteins*), are directly measured

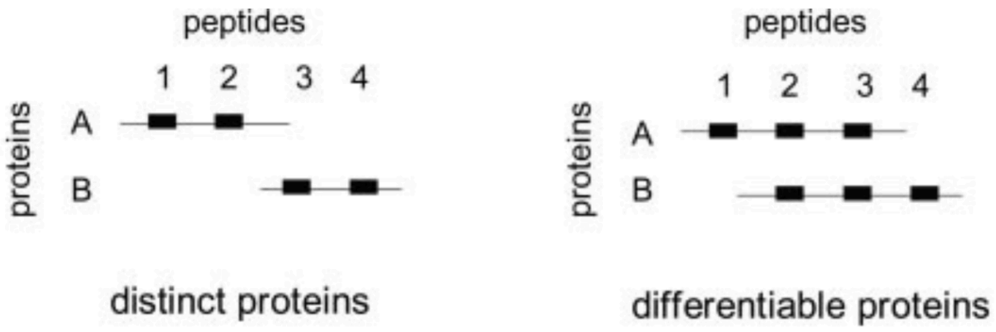
Protein inference is the process of extrapolating protein information from peptide measurements



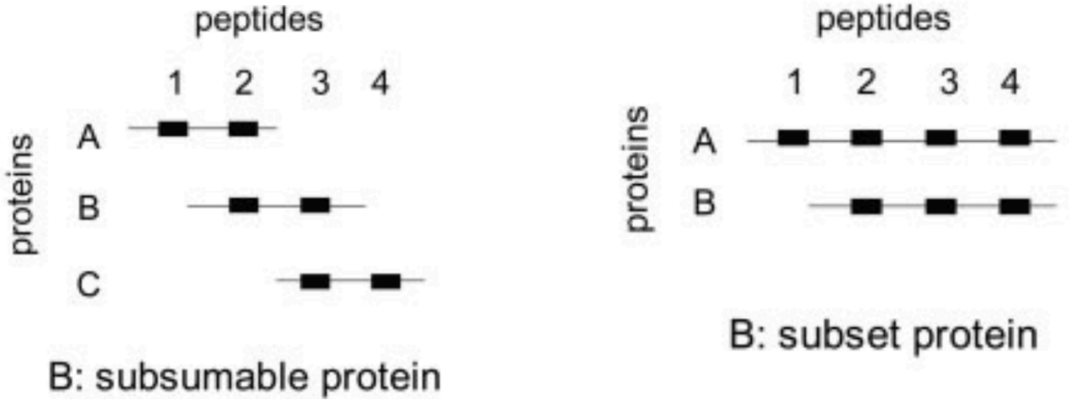
Eng, J. K.; McCormack, A. L.; Yates, J. R. *J. Am. Soc. Mass Spectrom.* **1994**, *5* , 976–989.

Technology Development and Workflow for MS Proteomics

A and B can be identified



B cannot be identified



sequence homology complicates analysis

- protein families
- alternative splicing
- protein isoforms (and point mutations)

peptide signal poses additional challenge

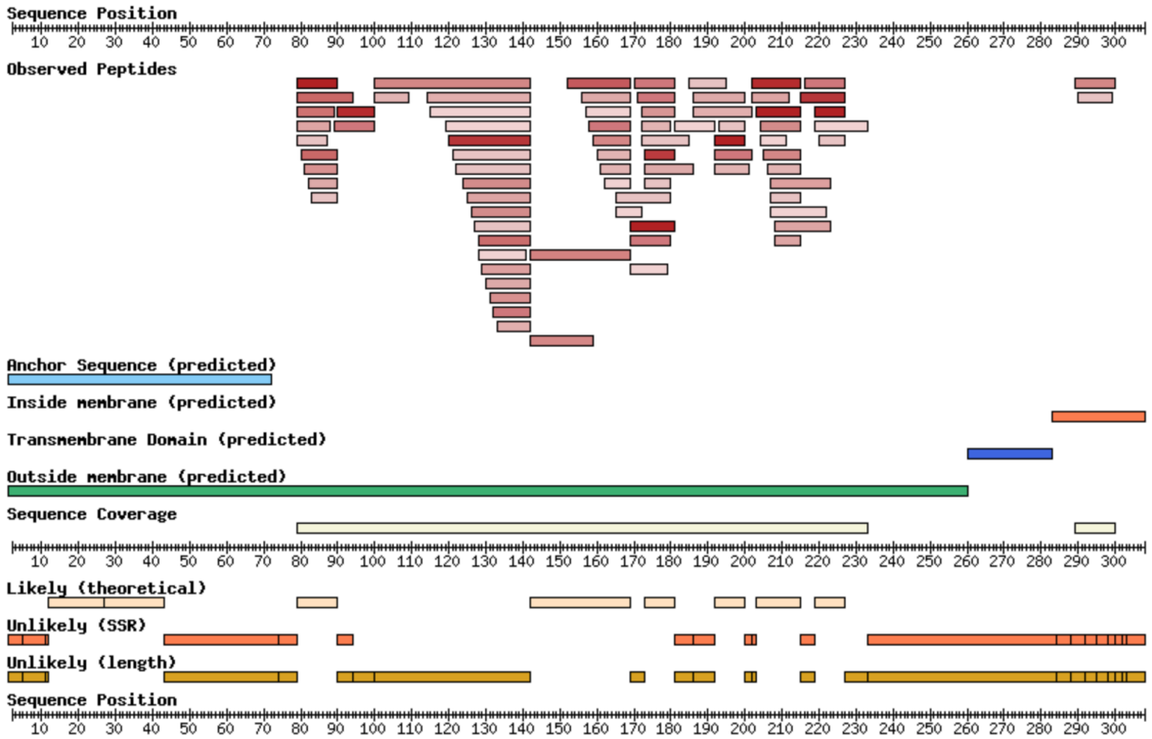
- low intensity ions (especially with DDA)
- small proteins (few tryptic sites)
- PTMs suppress signal

>30% of protein assignments are made based on a single peptide ID

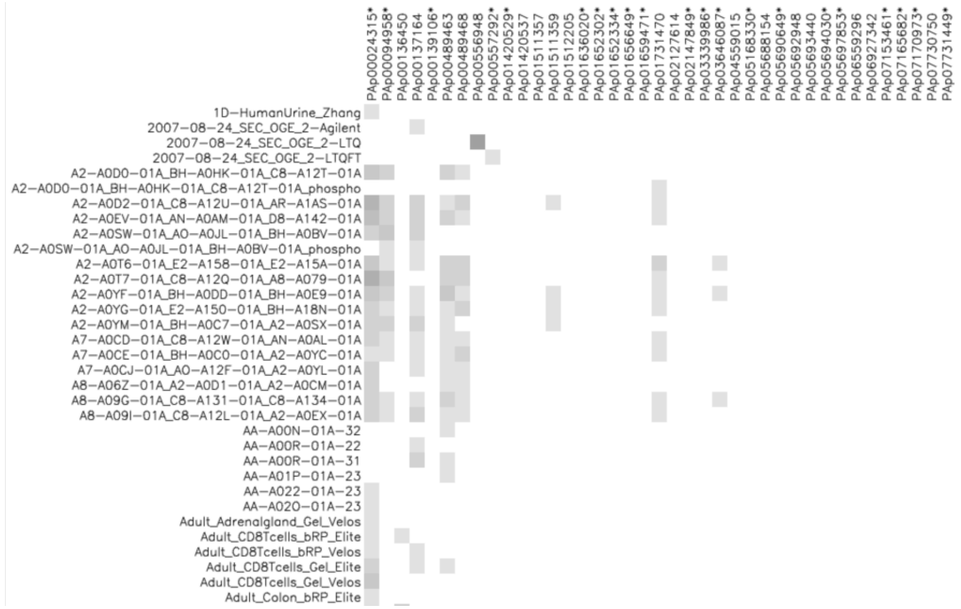
Nesvizhskii, A. I.; Aebersold, R. *Molecular & Cellular Proteomics* **2005**, *4*, 1419–1440.

Technology Development and Workflow for MS Proteomics

protein sequence coverage



observed peptides in experiments



peptide coverage information and much more available on proteomics databases

- jPOST
- MassIVE
- ProteomicsDB
- PeptideAtlas
- MaxQB

peptideatlas.org

An Overview of Topics Covered

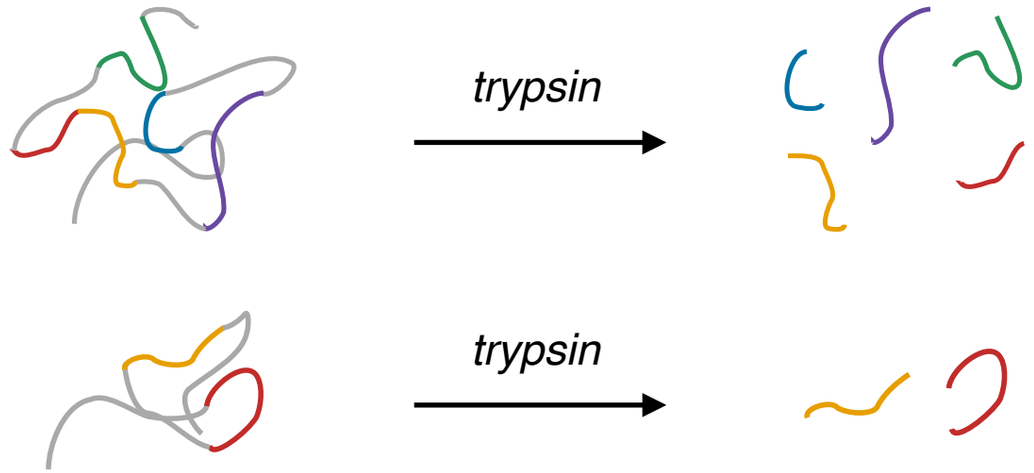
- Part 1: basic workflow and technology for discovery proteomics
  - How proteins are handled and analyzed
  - Data Peptide assignment and protein inference
- Part 2: methods for (relative) quantitative proteomics
  - Label-free methods
  - Whole-cell isotopic labeling strategies
  - Chemical mass tags
- Part 3: targeted proteomics and its application to biomarker discovery

Methods and Applications of Quantitative MS-Proteomics

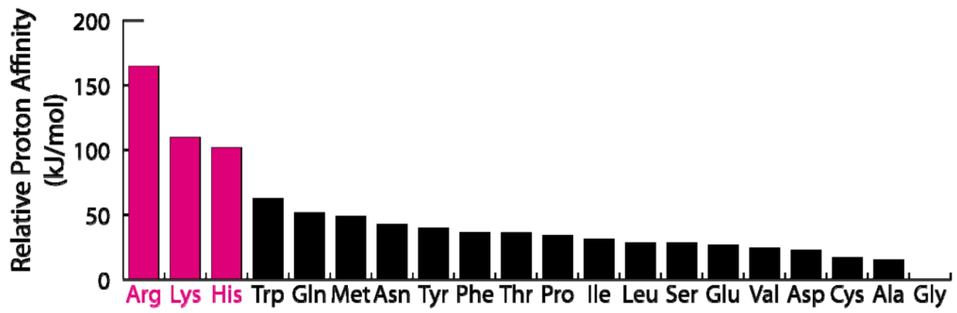
Mass spectrometry is not inherently quantitative…

Digestion efficiency is protein-dependent:

relative amounts of A and B may not be reflected by [peptide]



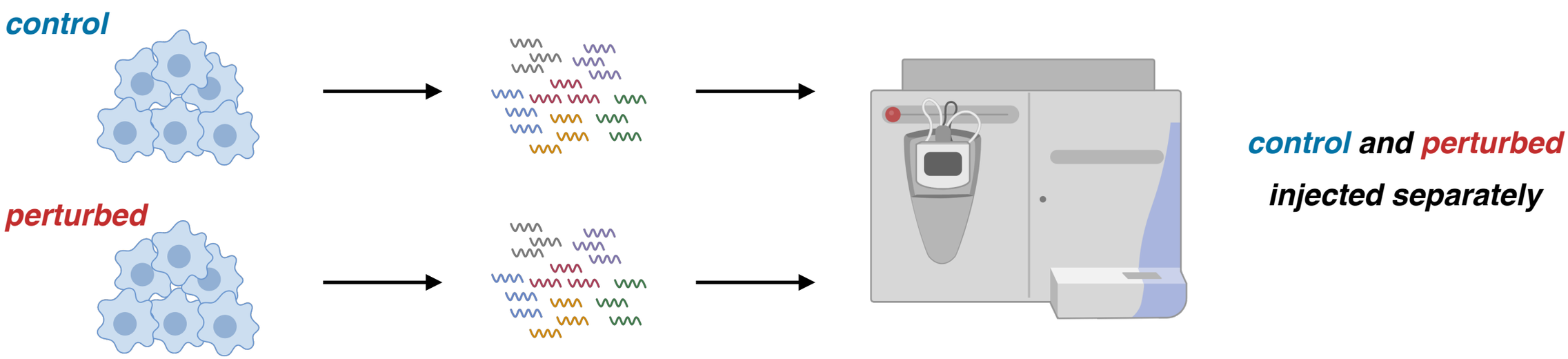
Ionization efficiencies vary by >100-fold



Matrix effects can change signal intensity (as for all MS)

Methods and Applications of Quantitative MS-Proteomics

Method 1: label-free quantitative proteomics

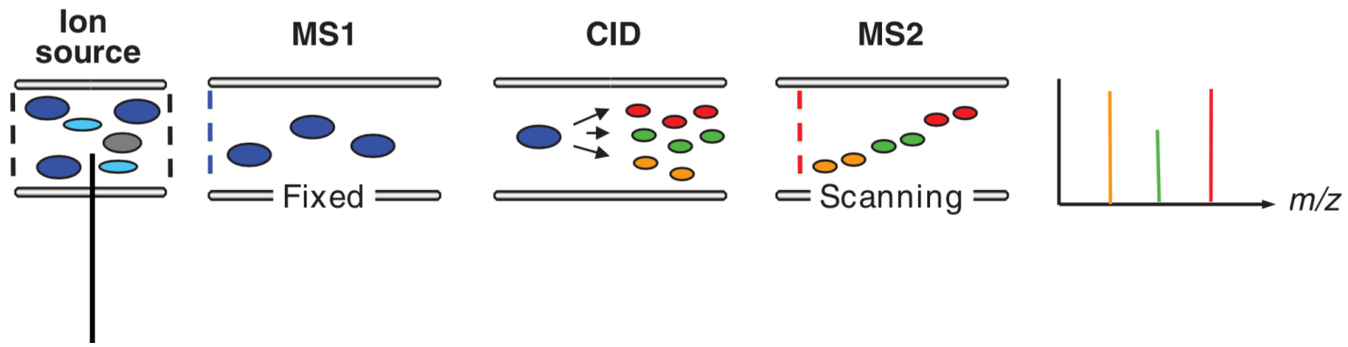


- Spectral counting or XIC used to compare abundances of a protein
- Housekeeping proteins are typically utilized for concentration normalization
- Many experimental arms can be compared, but many replicates (high *n)* often needed

Methods and Applications of Quantitative MS-Proteomics

Method 1: label-free quantitative proteomics

Data-dependent acquisition (DDA): only the highest intensity MS^1 precursor ions are selected for MS^2 analysis

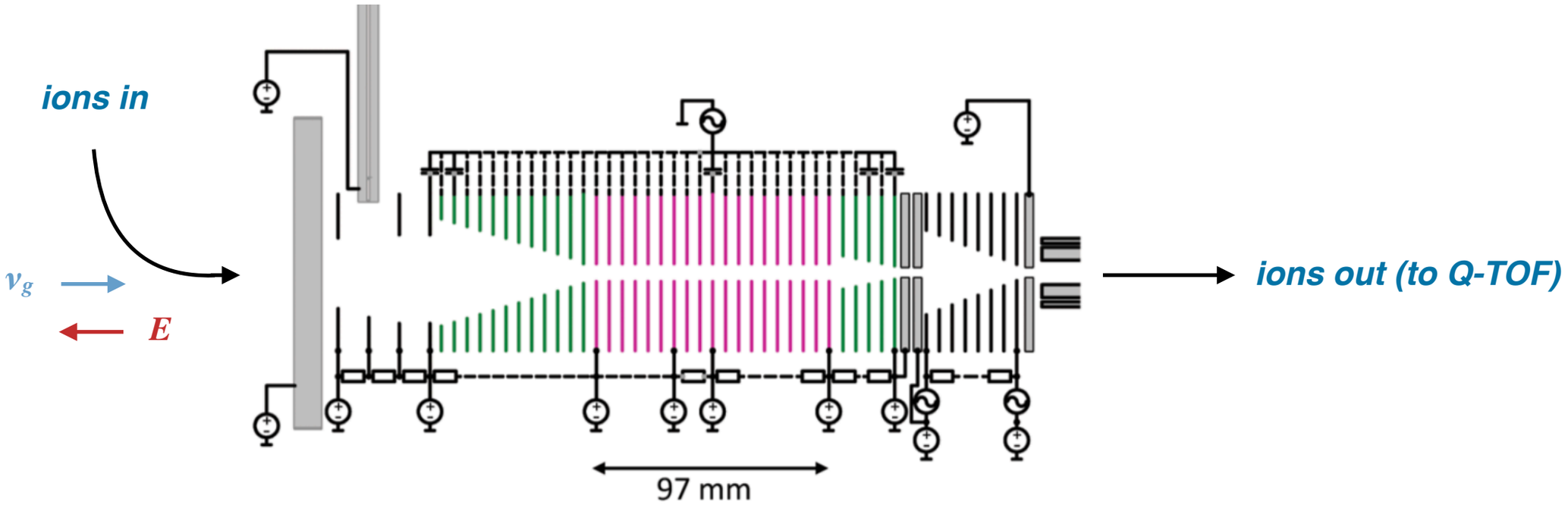


10-20 most intense ions *pass to CID cell per MS^1 scan (previously 3-8)*

Spectral counting: instances of peptide MS^1 ion observation (verified by MS^2) summed up for *relative* quantitation

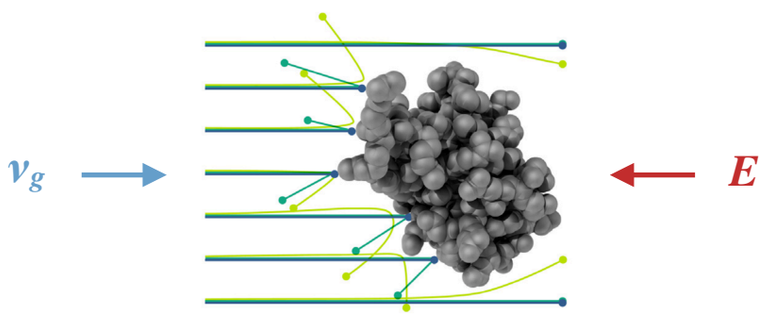
or

Extracted ion chromatogram (XIC): integrate ion intensity vs. time plot to quantitate peptide (MS^1)



v_g: *gas flow directs ions into TIMS (trapped ion mobility spectrometer)*

E: *electric field opposes gas flow, causing ion trapping*

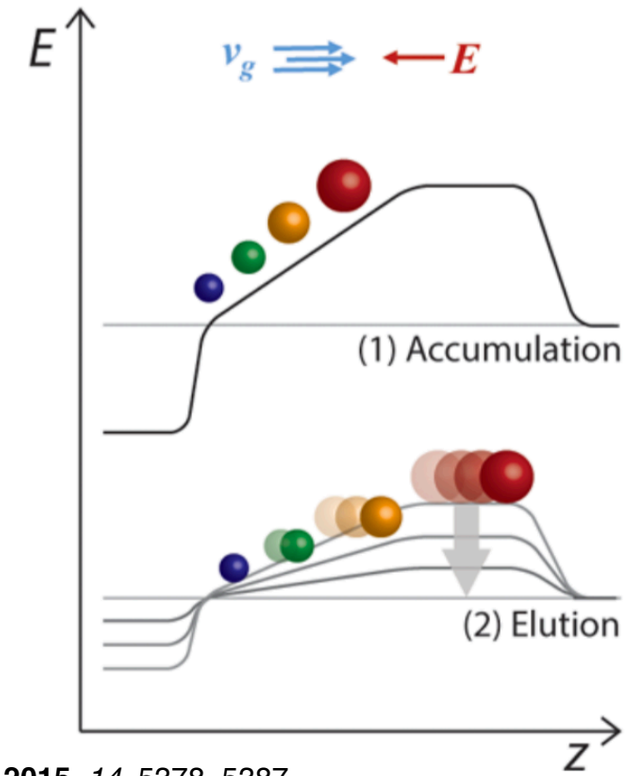


mobility depends on

collisional cross section (CCS)

ions eluted by ramping

down electric field strength



Meier, F.; Beck, S.; Grassl, N.; Lubeck, M.; Park, M. A.; Raether, O.; Mann, M. *J. Proteome Res.* **2015**, *14*, 5378–5387
Gabelica, V.; Marklund, E. *Current Opinion in Chemical Biology* **2018**, *42*, 51–59

Methods and Applications of Quantitative MS-Proteomics

Advantages of TIMS-TOF technology

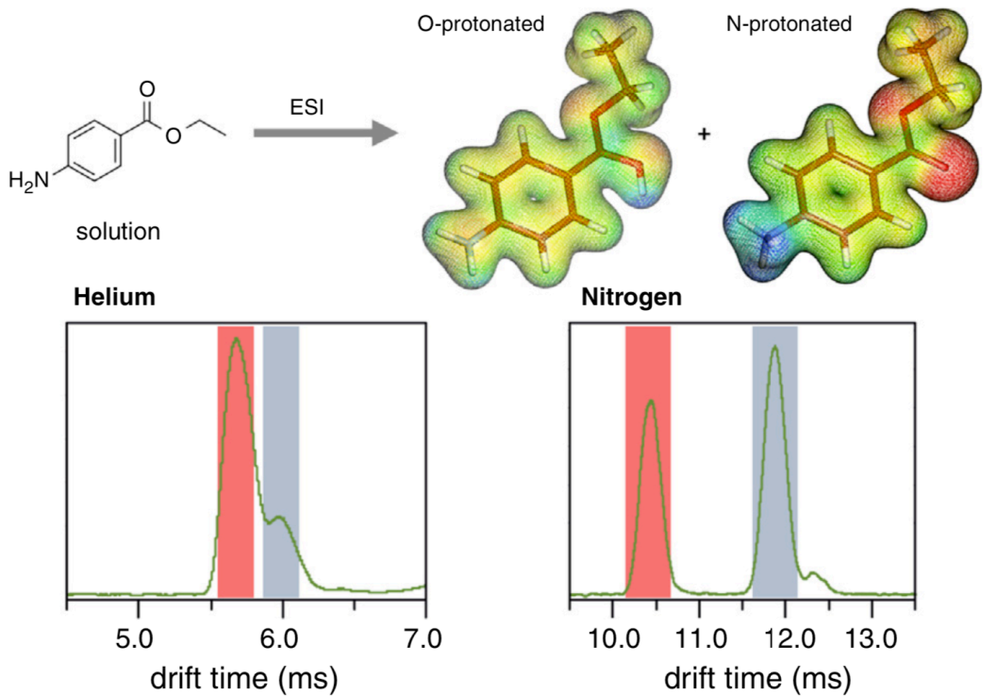
Added dimension of separation

Faster scanning speed

Deeper protein coverage

(can ID 2000 proteins from 10 cells of material)

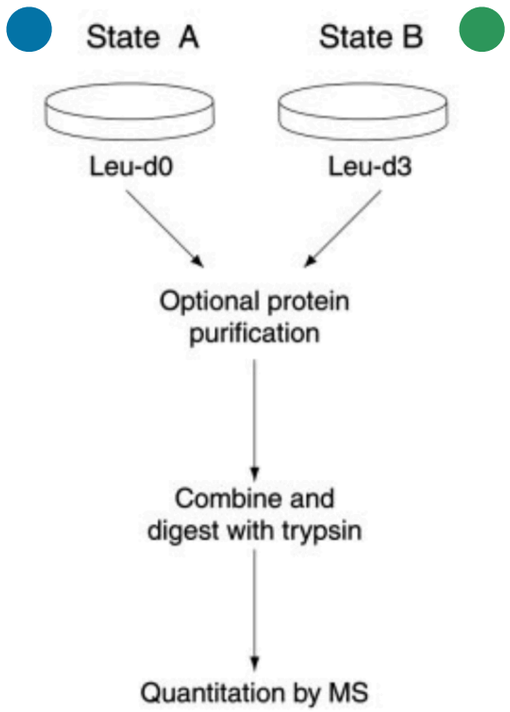
Separation based on CCS is impressive

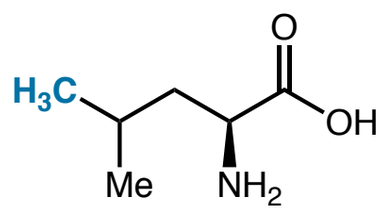


Ridgeway, M. E.; Lubeck, M.; Jordens, J.; Mann, M.; Park, M. A. *International Journal of Mass Spectrometry* **2018**, *425*, 22–35

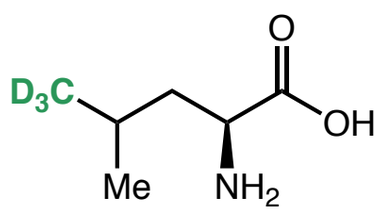
Gabelica, V.; Marklund, E. *Current Opinion in Chemical Biology* **2018**, *42*, 51–59

Method 2: stable isotope labeling with amino acids in cell culture (SILAC)

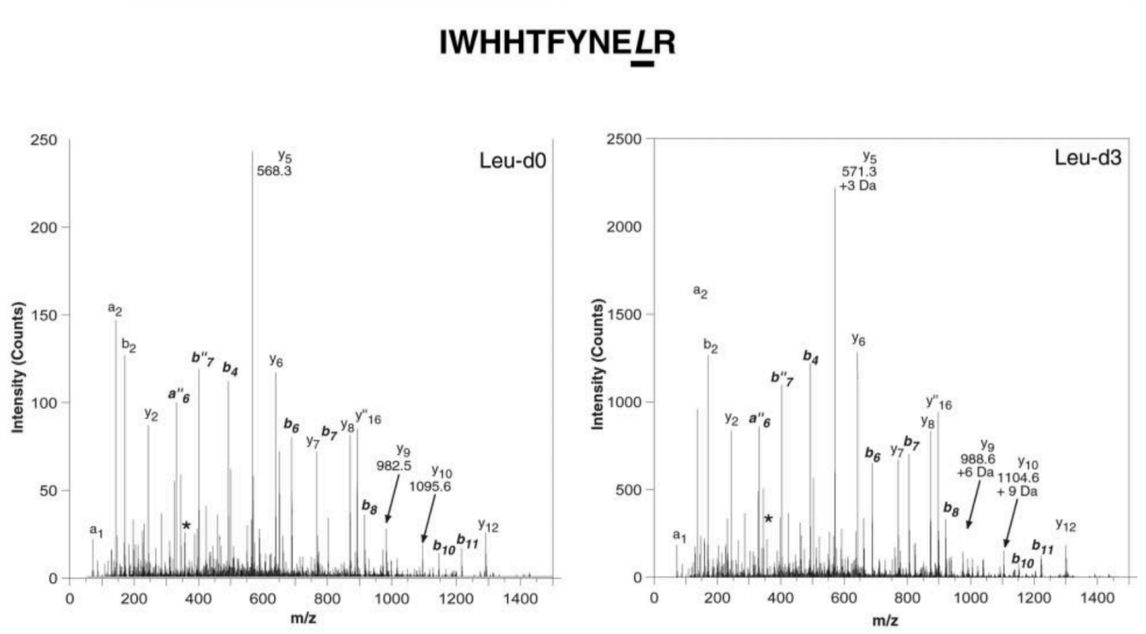




Leu-d0



Leu-d3



controls for ionization efficiency

differences and matrix effects

trypsin cut sites

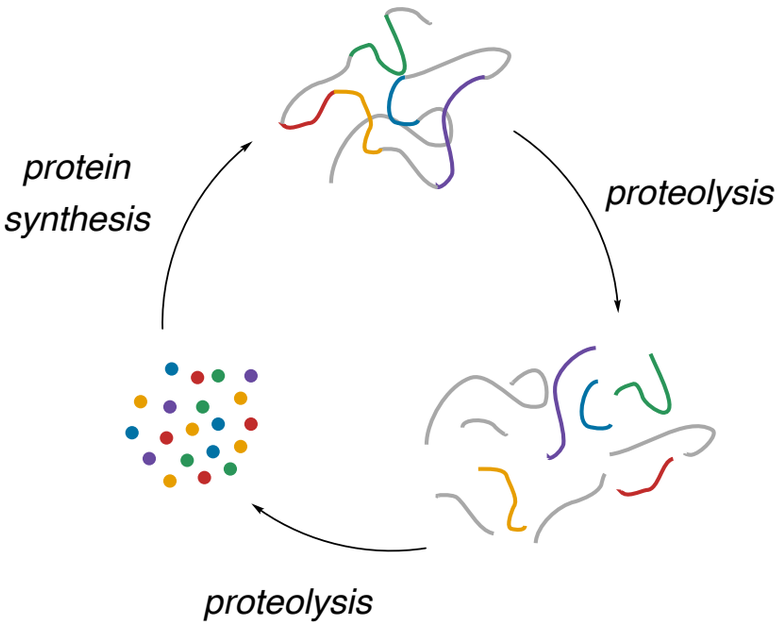
…AVT**K**WGS**R**AGPAVT**K**EIGAASTQV**R**AGDSLQP**K**GTVALE**R**…

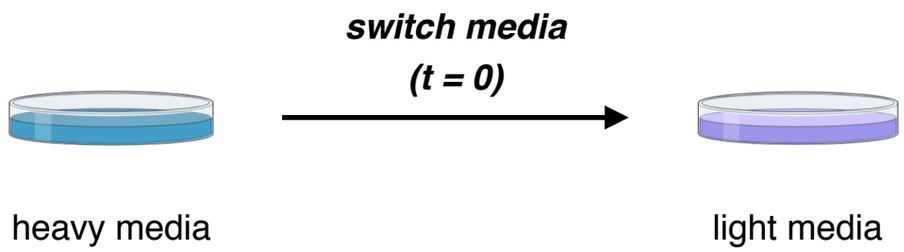
$^{13}C_6$Lys and $^{13}C_6{}^{15}N_2$Arg are optimal

Ong, S.-E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D. B.; Steen, H.; Pandey, A.; Mann, M. *Molecular & Cellular Proteomics* **2002**, *1*, 376–386
Mann, M. *Nature Reviews Molecular Cell Biology* **2006**, *7*, 952–958.

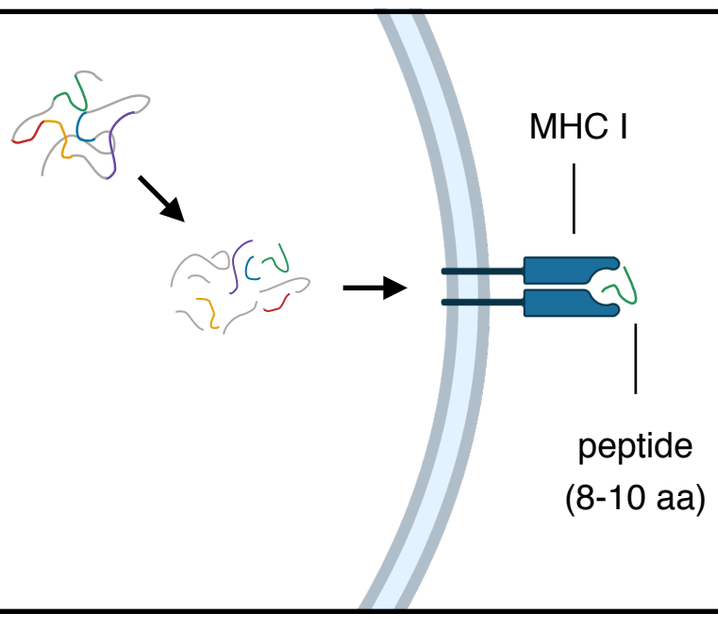
Methods and Applications of Quantitative MS-Proteomics

SILAC is well-suited to measure protein turnover kinetics





- heavy peptide signal decreases with time
- light peptide signal increases with time



MHC class I peptide display

- important for self/non-self recognition, tumor immunity
- peptide display requires proteasomal degradation

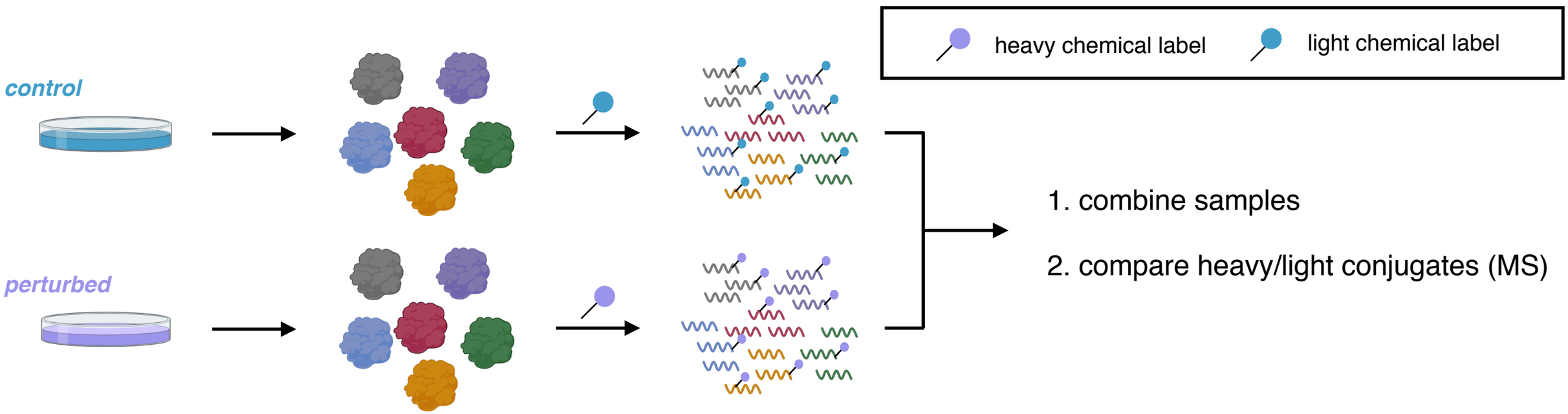
Milner, E.; Barnea, E.; Beer, I.; Admon, A. *Molecular & Cellular Proteomics* **2006**, *5*, 357–365
Mann, M. *Nature Reviews Molecular Cell Biology* **2006**, *7*, 952–958.

SILAC is well-suited to measure protein turnover kinetics

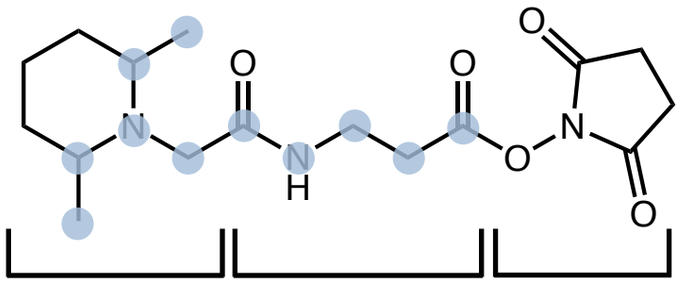
No.	Mass	Sequence	Protein	Complete turnover time	$t_{1/2}$
	amu			*h*	*h*
1	1210.7	LLLDVPTAAVQA	γ-Interferon-inducible protein IP-30 precursor	6	3
2	1011.5	LLLDVPTAAV	γ-Interferon-inducible protein IP-30 precursor	6	3
3	951.6	LLGPRLVLA	TMP21; transmembrane trafficking protein	6	3
4	965.6	ALATLIHQV	COP9 complex subunit 7a	6	5
5	913.6	GLLGTLVQL	Catenin β1	6	5
6	868.5	LLIPGLATA	NADH dehydrogenase (ubiquinone) 1α subcomplex	6	5
7	861.5	ILGPTFTL	3-Hydroxy-3-methylglutaryl-coenzyme A reductase	6	5
8	1037.7	KLLEPVLLL	Similar to 40 S ribosomal protein S16	6	5
9	1034.6	FVFPGELLL	Solute carrier family 1 (neutral amino acid transporter), member 5	6	4
10	910.5	ALPPVLTTV	Unnamed protein product	9	?
11	946.5	SLVEEDALA	Hypothetical protein FLJ30668	9	3
12	1079.7	VLLKARLVPA	NPD019	9	4
13	947.5	ALYVAVVNV	Seven-transmembrane domain protein	9	5
14	1121.5	TLWVDPYEV	B-cell translocation protein 1	9	5
15	959.5	SLFPGQVVI	Polymerase (DNA-directed), α (70 kDa)	9	8
16	974.6	AILPTSIFL	SKB1 homolog	9	4
17	846.5	ALSRITSV	Unknown (protein for MGC:14124)	9	5
18	855.5	ALLGGLVNV	Progestin and adipoQ receptor family member IV	9	6
19	900.5	ALFPGVALL	Protein-disulfide isomerase mER60 precursor	12	8
20	1015.5	FQDPVPLTV	Transcription intermediary factor 1	18	12
21	1032.5	ALPEIFTEL	Similar to eukaryotic translation initiation factor 2, 26 subunit 3γ, 52 kDa	24	4
22	1094.6	SLLPPDALVGL	Sec23 protein	24	11
23	1360.7	ALWDIETGQQTV	Guanine nucleotide-binding protein, β-2 subunit	24	12
24	1115.5	SLFEGTWYL	Hydroxymethylglutaryl-CoA synthase	24	15
25	969.6	VIAEILRGV	Nucleolar protein 5A	24	20
26	**984.5**	**ALMPVLNQV**	**Homolog of yeast mRNA transport regulator 3**	**48**	**6**
27	908.5	NLDTSVFI	Similar to RIKEN cDNA 2610003J06	48	12
28	1258.6	FLFDGSPTYVL	Fatty-acid synthase	48	15
29	989.5	ILGGSLFGLL	ET putative translation product	48	23
30	968.6	SLLDPVPEV	Similar to RIKEN cDNA G431004K08	48	24
31	**1020.6**	**FLSSVIQNL**	**Proteasome 26 S non-ATPase subunit 1**	**96**	**6**
32	**1038.6**	**YLLPAIVHI**	**DEAD box polypeptide 17 isoform p82; probable RNA-dependent helicase p72**	**96**	**11**
33	929.5	SLLDKIIGA	Polymerase I and transcript release factor	96	14
34	1049.6	VLMQDLAFL	Unnamed protein product	96	52
35	786.4	SLAGGILGV	Protein similar to heterogeneous nuclear ribonucleoprotein K	?	24

Milner, E.; Barnea, E.; Beer, I.; Admon, A. *Molecular & Cellular Proteomics* **2006**, *5*, 357–365
Mann, M. *Nature Reviews Molecular Cell Biology* **2006**, *7*, 952–958.

Method 3: chemical labelling with isotopically-labeled tags



Current state-of-the-art: isobaric mass tags (iTRAQ, TMT)

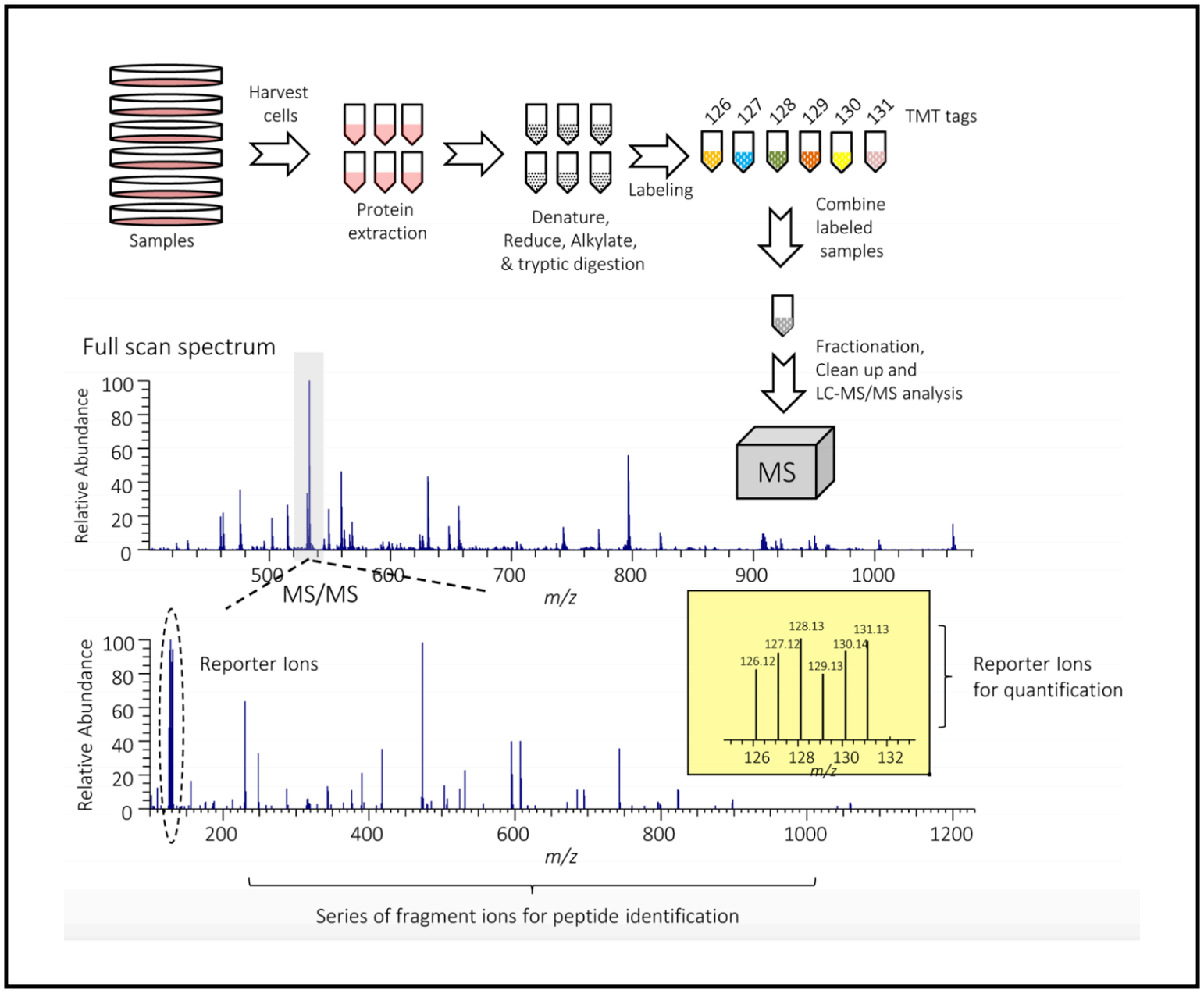


reporter ion + *mass normalizer* = constant

● – sites of ^{15}N and ^{13}C labels

Thompson, A.; Schäfer, J.; Kuhn, K.; Kienle, S.; Schwarz, J.; Schmidt, G.; Neumann, T.; Hamon, C. *Anal. Chem.* **2003**, *75*, 1895–1904

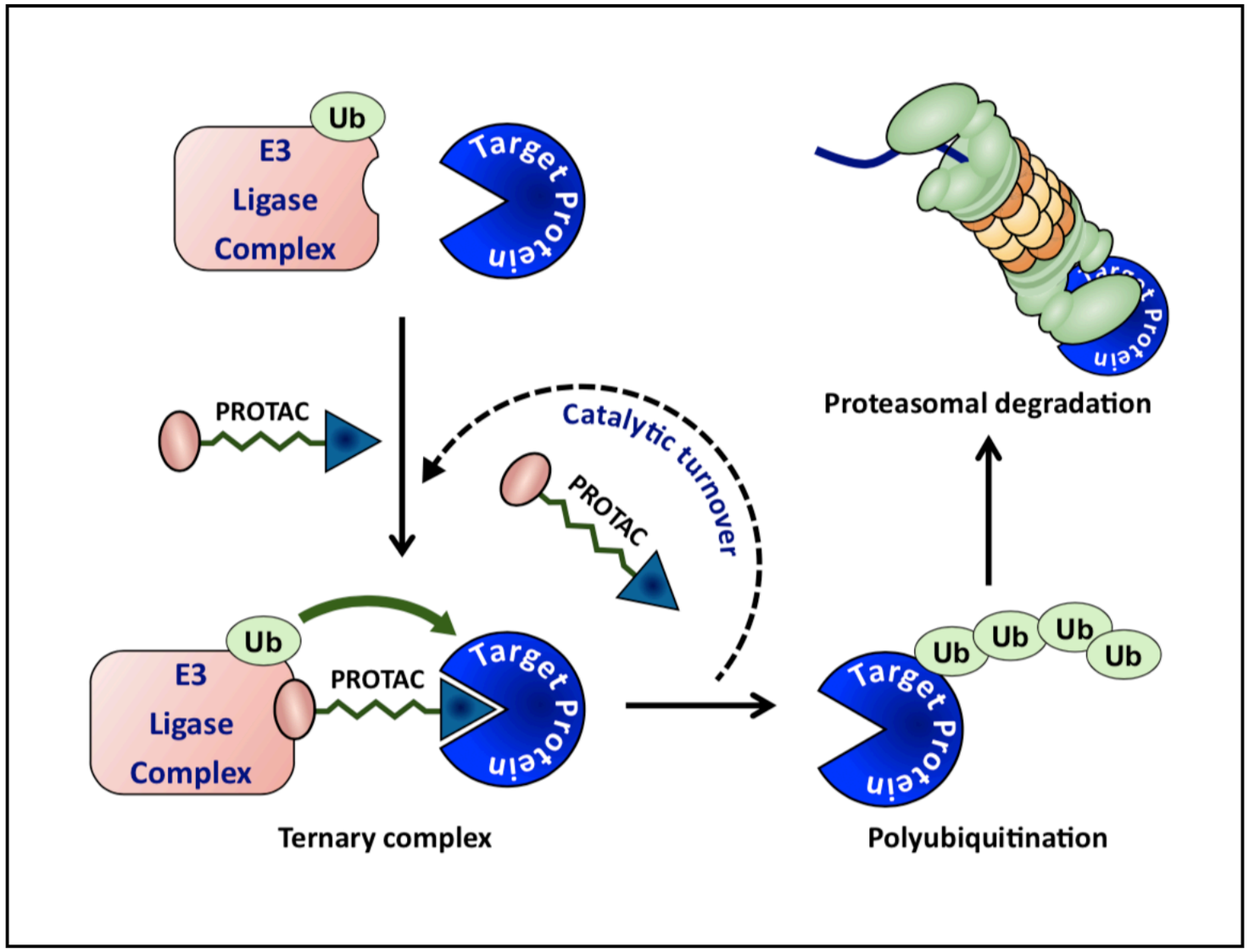
Method 3: chemical labelling with isotopically-labeled tags



- Up to 11-plex possible today
- ^{13}C and ^{15}N, no ^{2}H used
- High resolution MS^2 needed

Rauniyar, N.; Yates, J. R. *J. Proteome Res.* **2014**, *13*, 5293–5309

PROTACs cause changes in expression levels at the proteome level

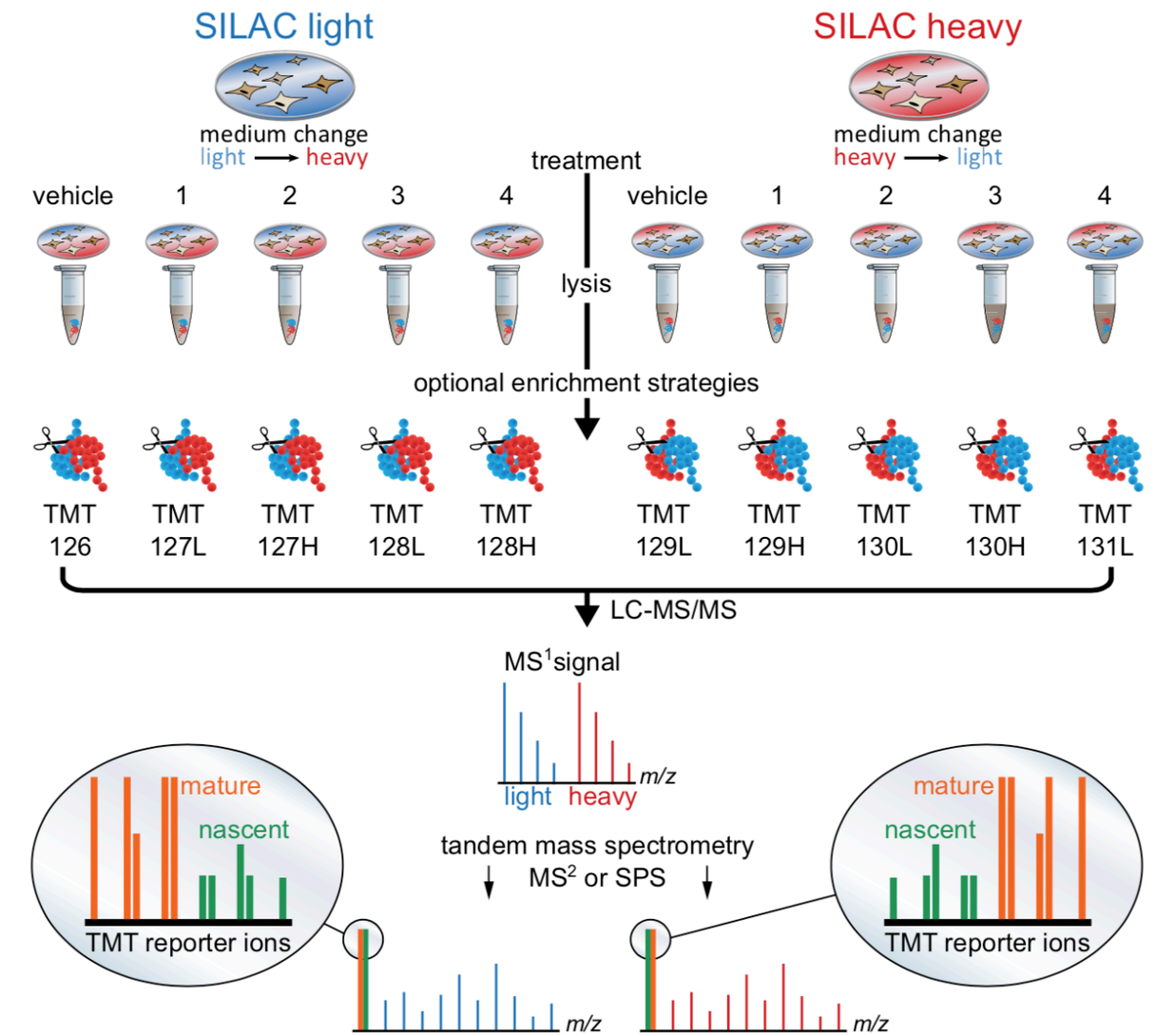


How can these changes be assigned to degradation or transcriptional regulation?

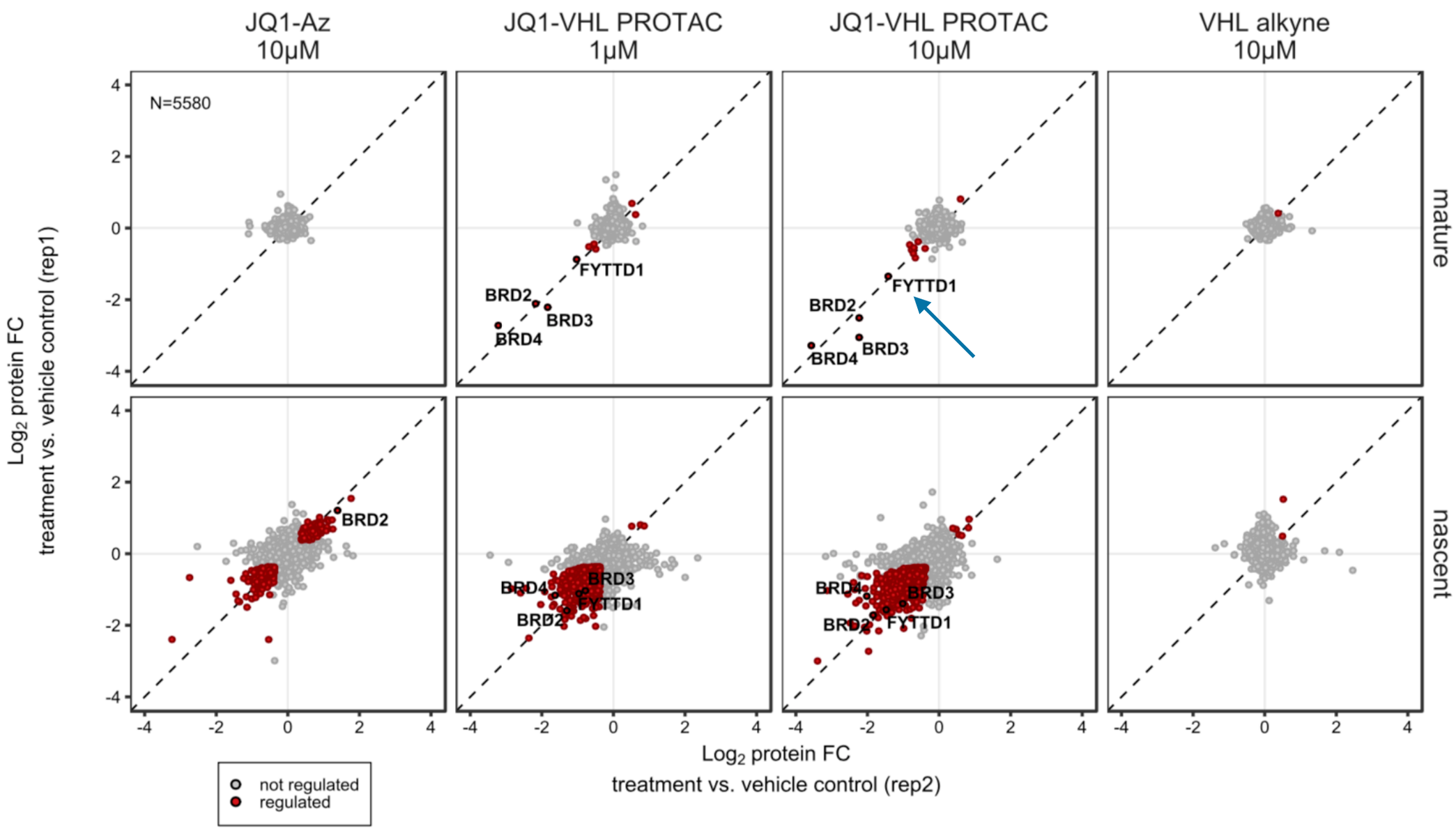
Welle, K. A.; Zhang, T.; Hryhorenko, J. R.; Shen, S.; Qu, J.; Ghaemmaghami, S. *Molecular & Cellular Proteomics* **2016**, *15*, 3551–3563.
Savitski, M. M.; Zinn, N.; Faelth-Savitski, M.;…Bantscheff, M. *Cell* **2018**, *173*, 260-274.e25

Methods and Applications of Quantitative MS-Proteomics

Combining SILAC and TMT to study protein regulation mechanisms



Savitski, M. M.; Zinn, N.; Faelth-Savitski, M.;…Bantscheff, M. *Cell* **2018**, *173*, 260-274.e25



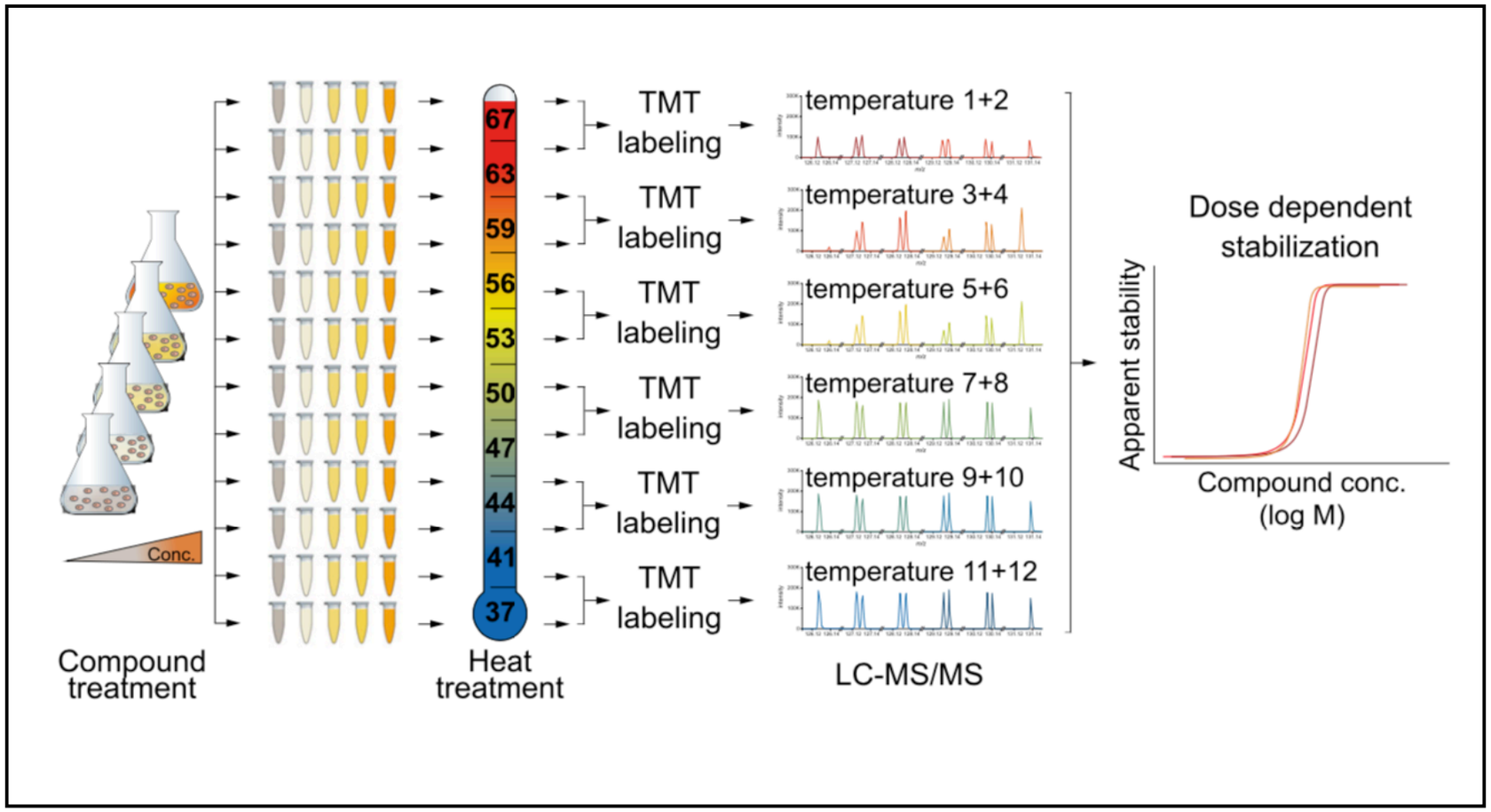
FYTDD1 critical for mRNA nuclear export, hence protein synthesis

off target degradation by JQ1-VHL

Savitski, M. M.; Zinn, N.; Faelth-Savitski, M.;…Bantscheff, M. *Cell* **2018**, *173*, 260-274.e25

Methods and Applications of Quantitative MS-Proteomics

MS-based protein stability/small molecule binding assay



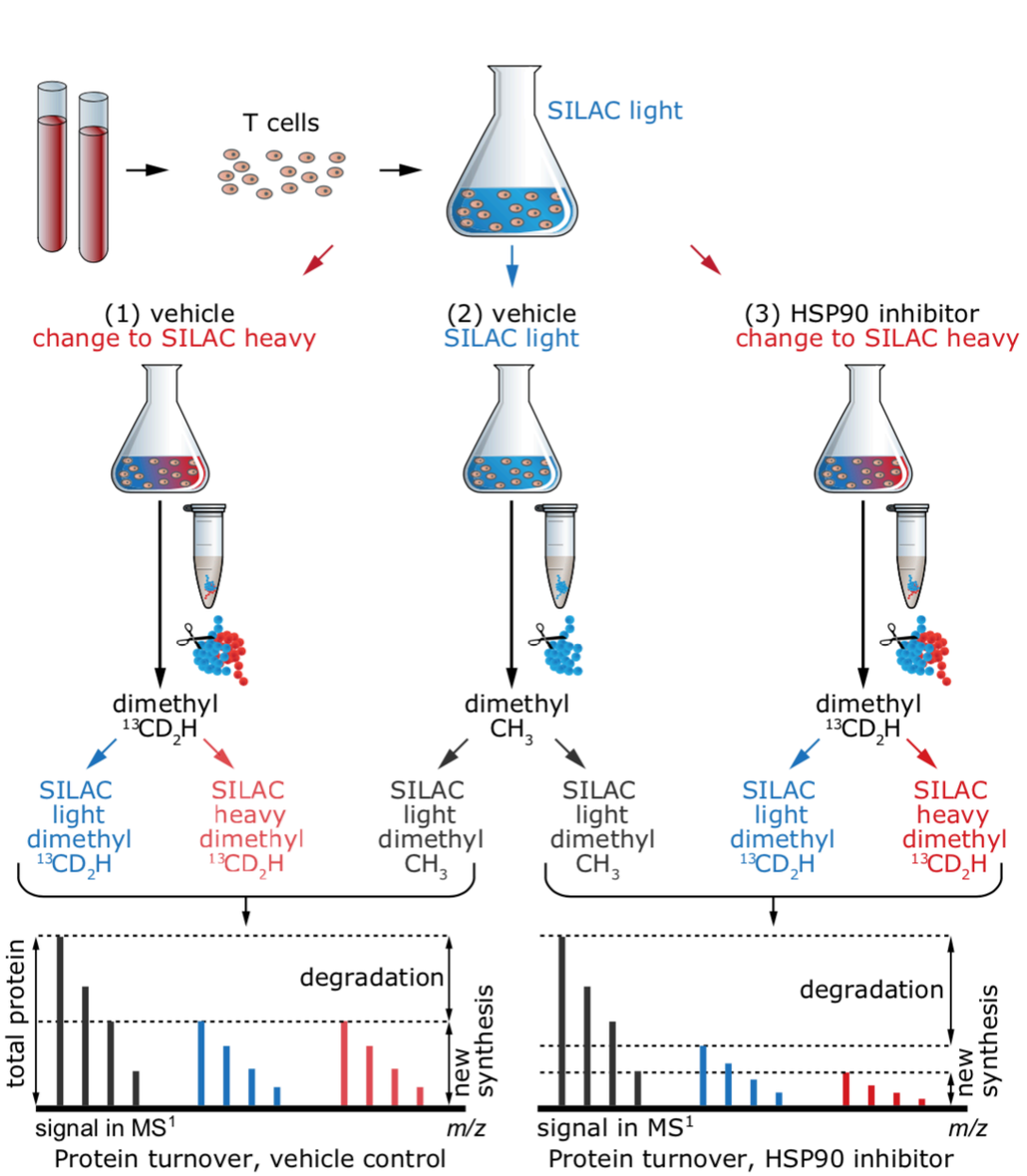
direct interaction between JQ1-VHL and FYTDD1 confirmed,

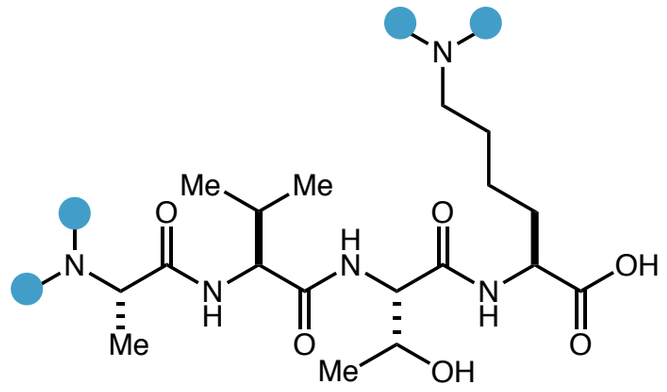
new PROTAC prepared without off-target activity

Savitski, M. M.; Zinn, N.; Faelth-Savitski, M.;…Bantscheff, M. *Cell* **2018**, *173*, 260-274.e25

Methods and Applications of Quantitative MS-Proteomics

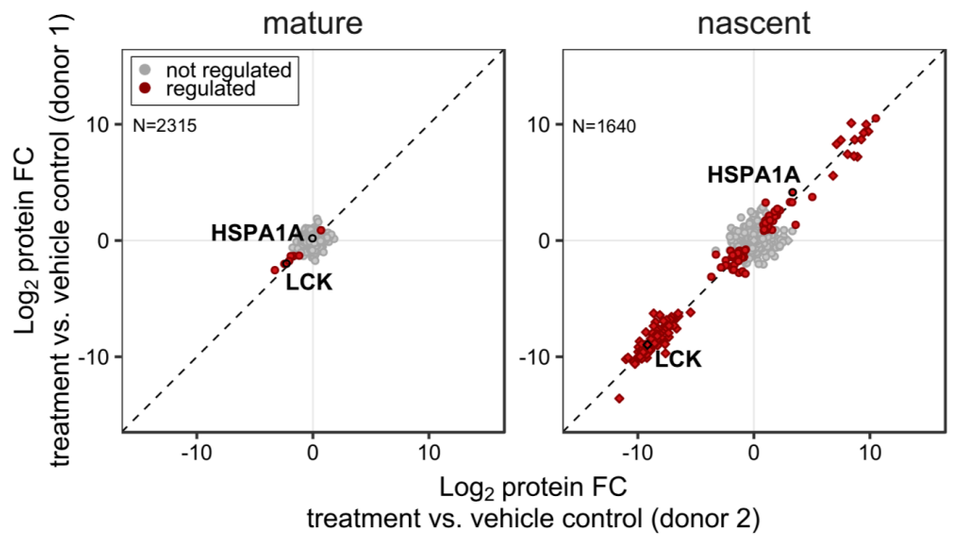
Reductive Dimethylation: alternative chemical labeling strategy with stable isotopes





● – $^{13}CD_2H$ or $^{12}CH_3$

compare in MS^1



Hsu, J.-L.; Huang, S.-Y.; Chow, N.-H.; Chen, S.-H. *Anal. Chem.* **2003**, *75*, 6843–6852

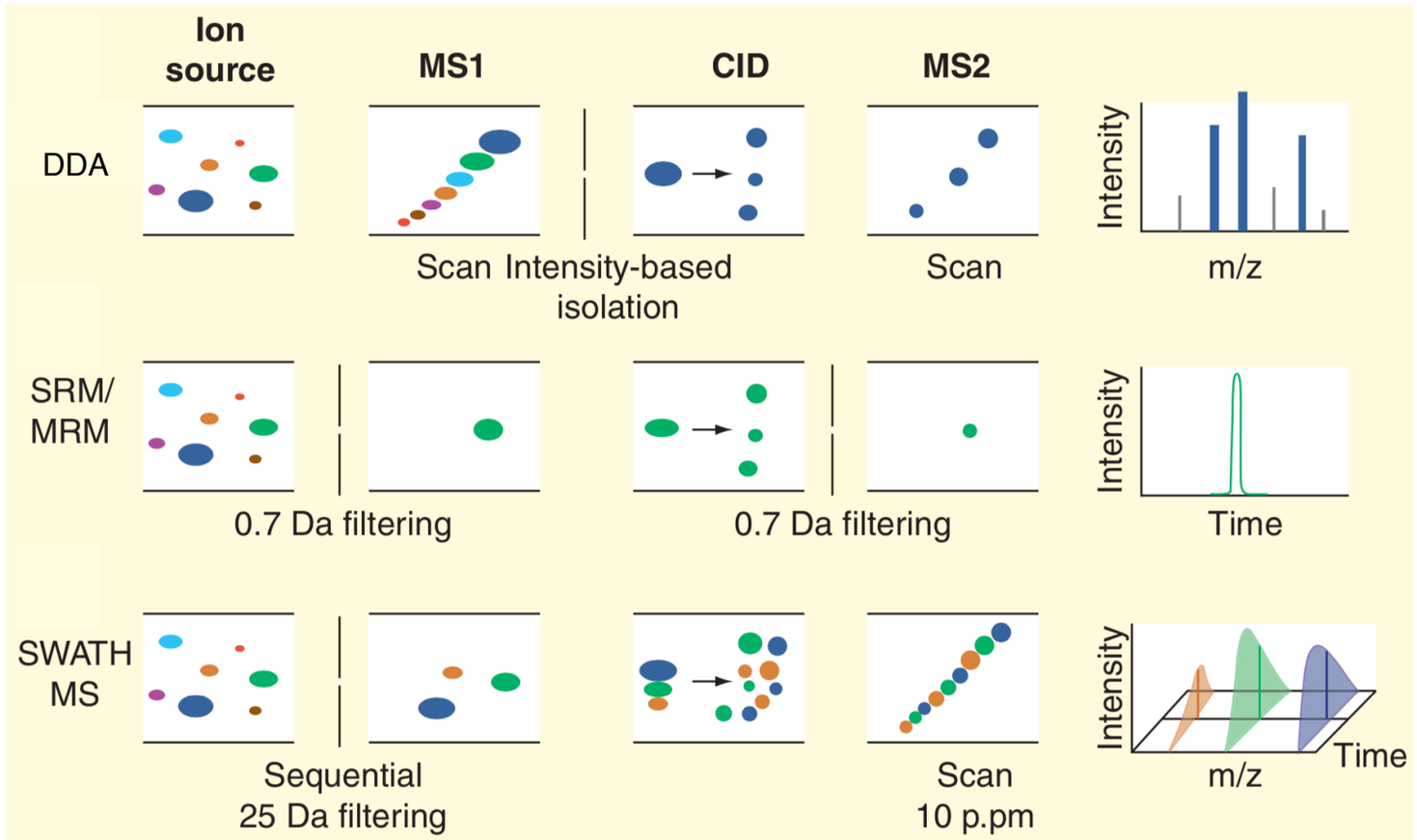
Savitski, M. M.; Zinn, N.; Faelth-Savitski, M.;…Bantscheff, M. *Cell* **2018**, *173*, 260-274.e25

An Overview of Topics Covered

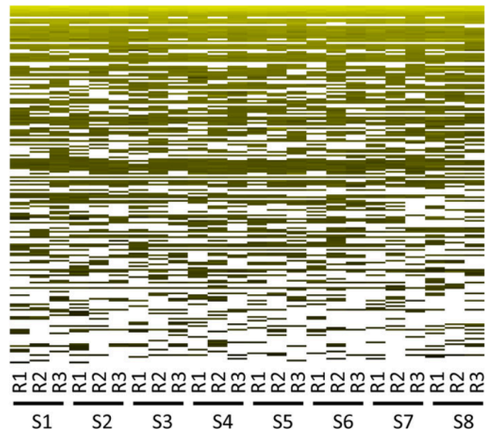
- Part 1: basic workflow and technology for discovery proteomics
  - How proteins are handled and analyzed
  - Data Peptide assignment and protein inference
- Part 2: methods for (relative) quantitative proteomics
  - Label-free methods
  - Whole-cell isotopic labeling strategies
  - Chemical mass tags
- Part 3: targeted proteomics and its application to biomarker discovery

Hsu, J.-L.; Huang, S.-Y.; Chow, N.-H.; Chen, S.-H. *Anal. Chem.* **2003**, *75*, 6843–6852
Savitski, M. M.; Zinn, N.; Faelth-Savitski, M.;…Bantscheff, M. *Cell* **2018**, *173*, 260-274.e25

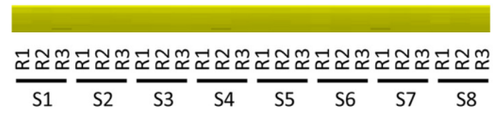
Beyond DDA-based Shotgun Proteomics



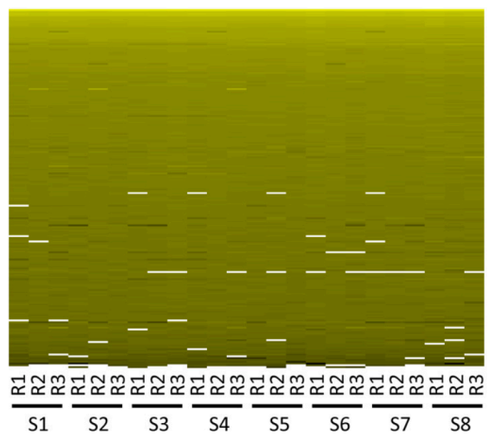
“holey” data with DDA



low coverage with SRM



better coverage with DIA



DDA: most common approach to discovery proteomics, hypothesis-free

SRM: accurate, reproducible quantitation of up to ~500 peptides (chosen)

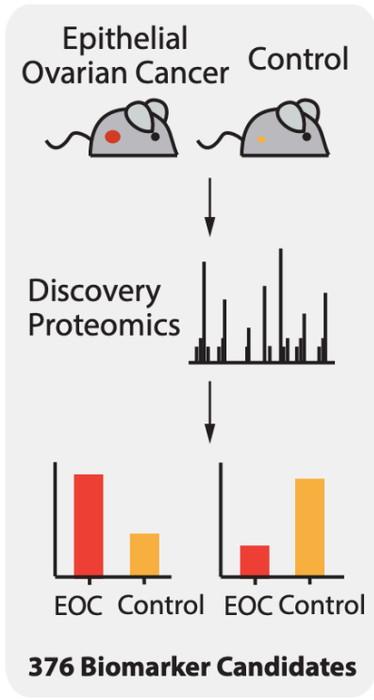
DIA: accurate, reproducible, deep coverage, but requires a DDA measurement first

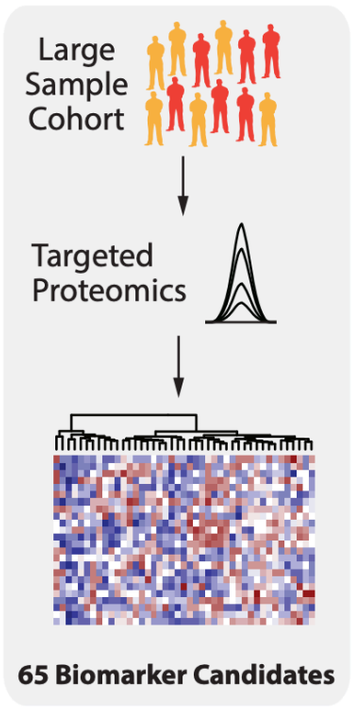
Bruderer, R.; Bernhardt, O. M.; Gandhi, T.;…Reiter, L. *Molecular & Cellular Proteomics* **2015**, *14*, 1400–1410

Liu, Y.; Hüttenhain, R.; Collins, B.; Aebersold, R. *Expert Review of Molecular Diagnostics* **2013**, *13*, 811–825

Technology Development and Workflow for MS Proteomics

Biomarkers "generate clinically useful information that could be used to change the course of the disease for a patient"





biomarker signature development

utilizes targeted proteomics

Carr, S. A.; Abbatiello, S. E.; Ackermann, B. L.;….Weintraub, S. *Molecular & Cellular Proteomics* **2014**, *13*, 907–917