

BE611 Synthetic Biology and Metabolic Engineering

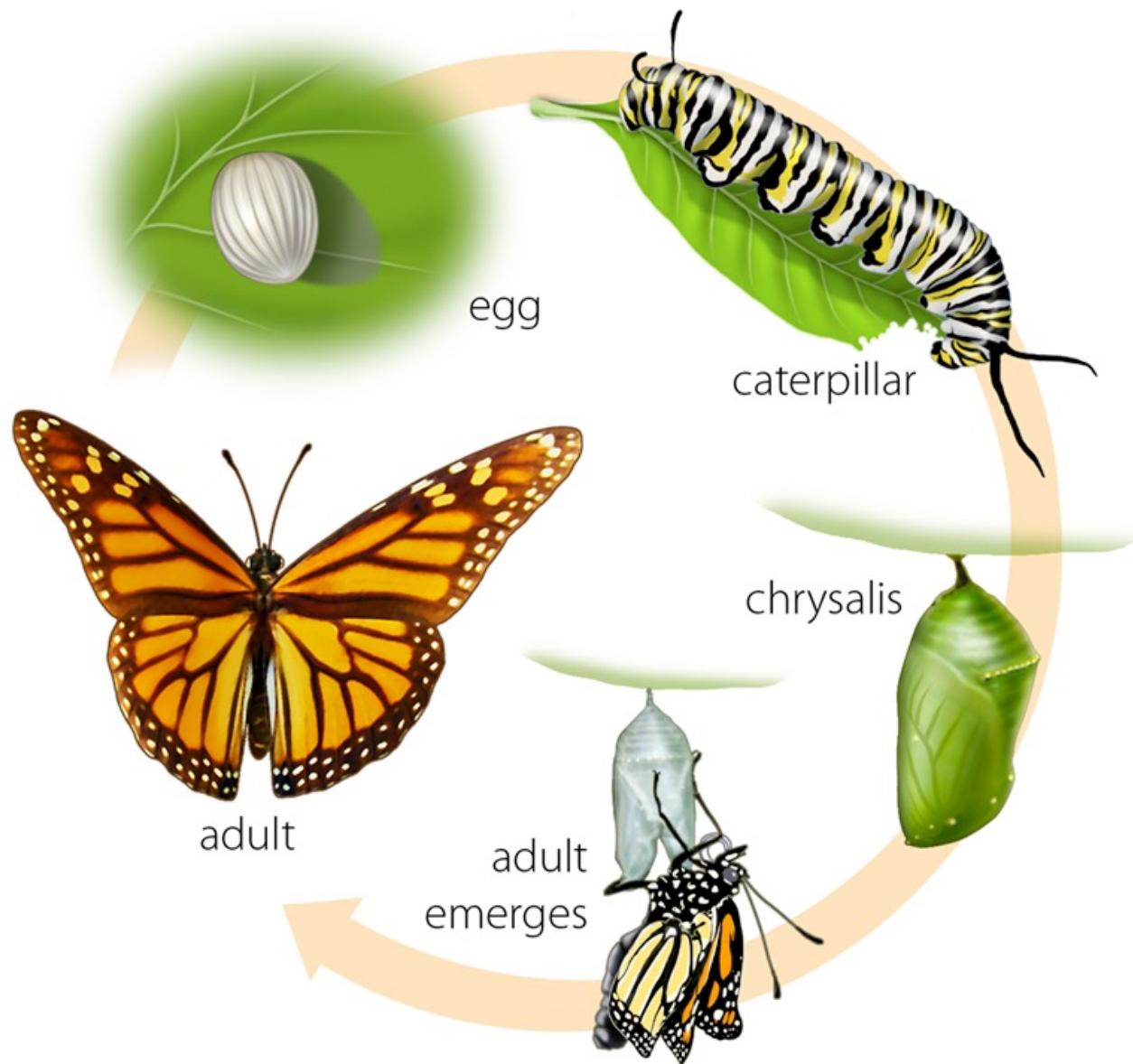
Proteomics and mass spectrometry

May 3rd, 2022

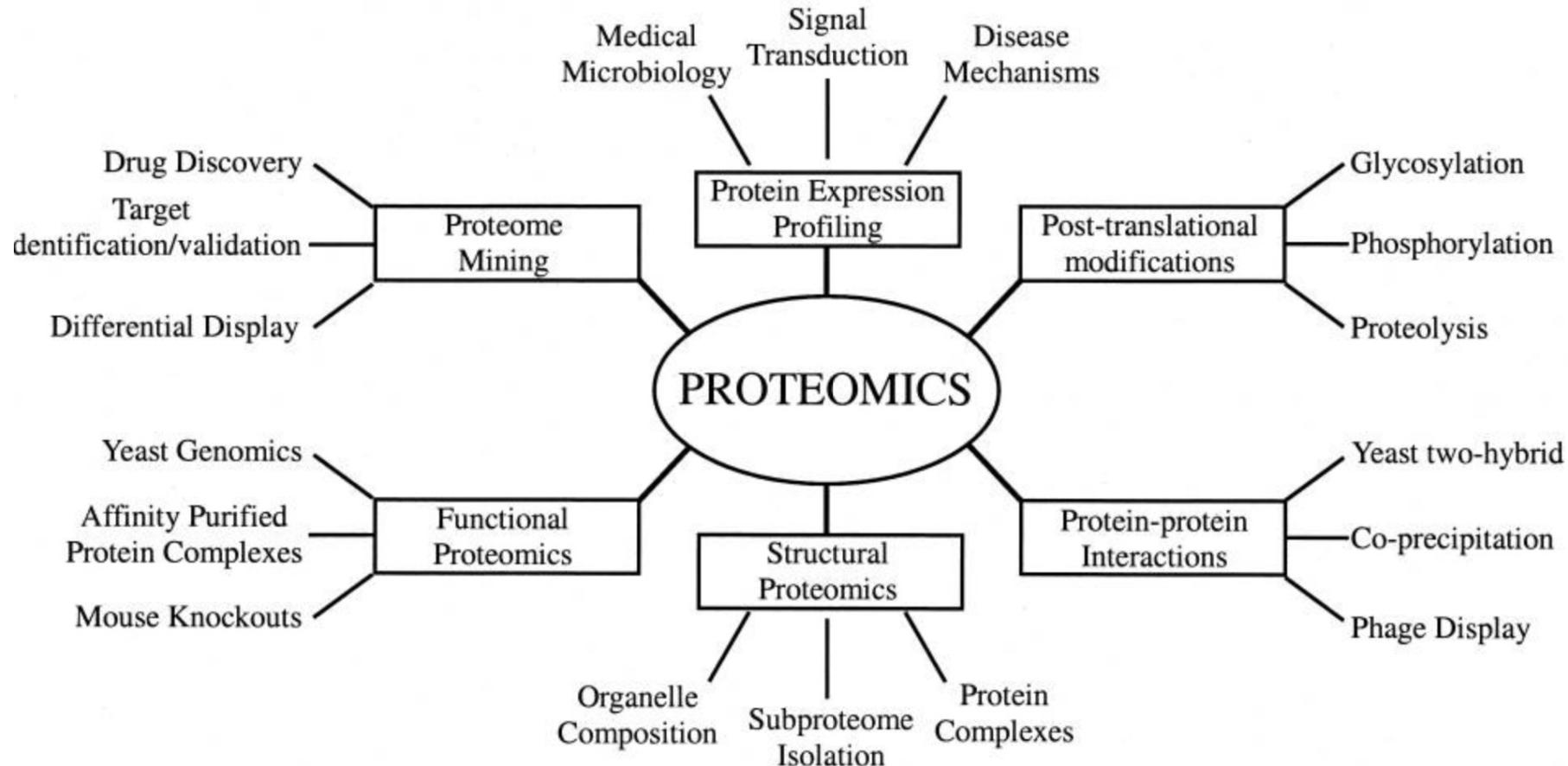


Sira Sriswasdi, Ph.D.
Computational Molecular Biology Group
Faculty of Medicine, Chulalongkorn University

Genome is static, proteome is dynamic



What is proteomics?



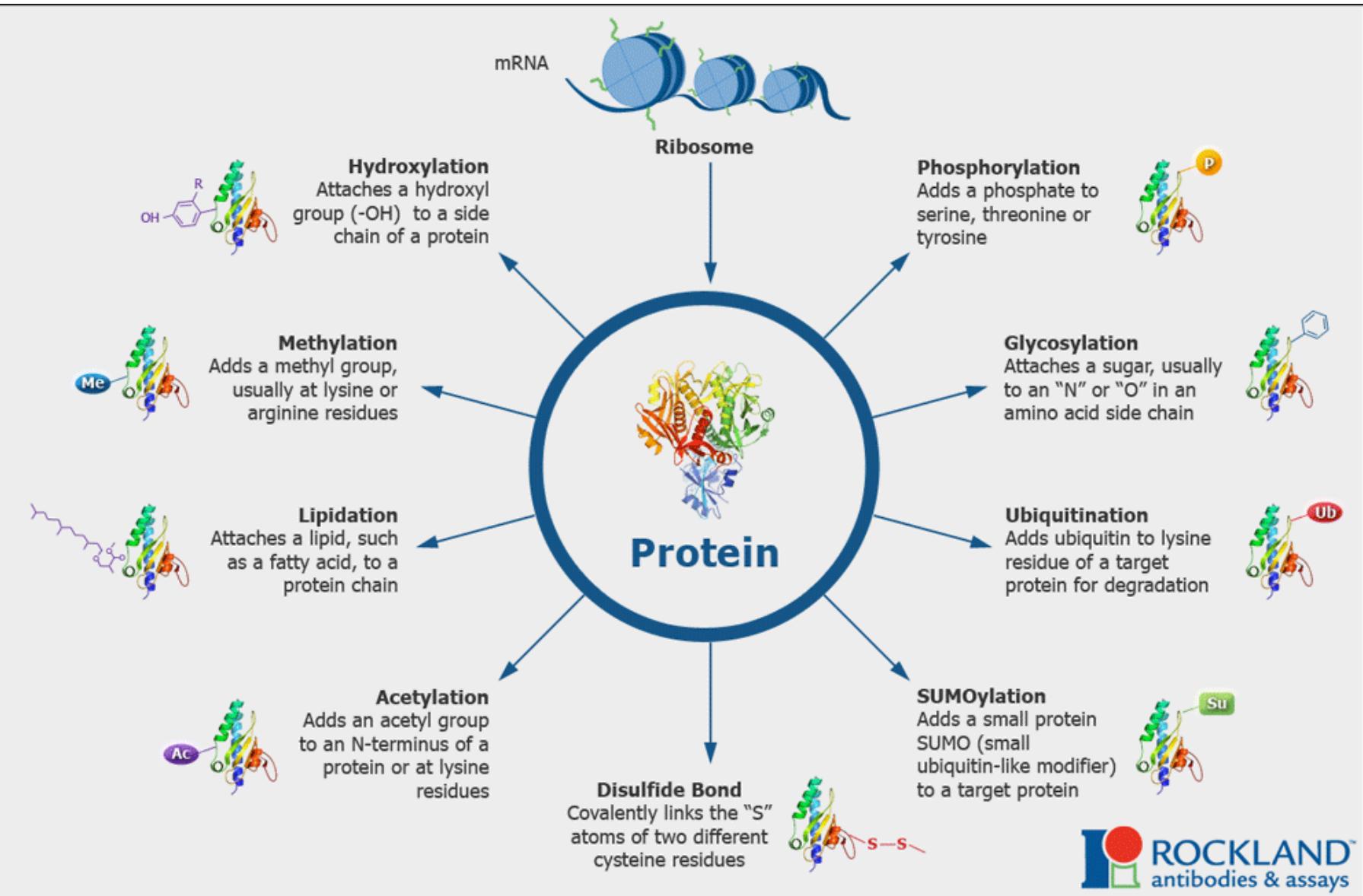
Grave and Haystead. *Microbiol Mol Biol Rev*, 66(1):39-63 (2002)

- Unbiased characterization of proteins and peptides within a biological sample
- Scope can be global or targeted

Key challenges in proteomics

- Proteome is much more complex than genome and transcriptome
 - Post-translational modification
- Proteins cannot be amplified
 - Cannot analyze limited amount of sample
- Amino acids do not have complementary pairing
 - Cannot sequence by synthesis like DNA
 - Instead use Mass Spectrometry technique

Post-translational modification (PTM)



PTM expands the space of proteins

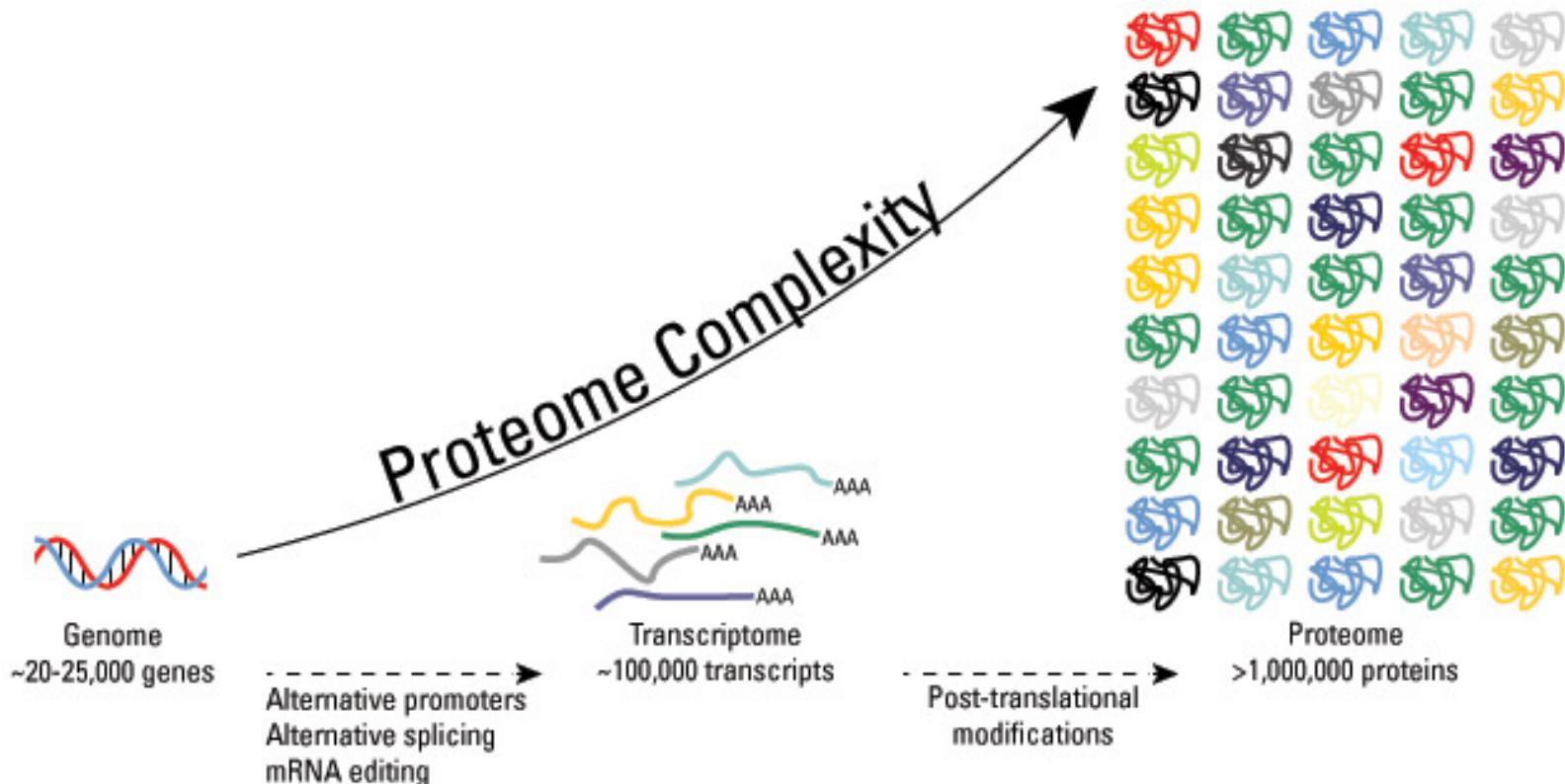


Image from <https://www.thermofisher.com>

- Critical for protein functions
 - Ex: phosphorylation and glycosylation
- Greatly increases the complexity of protein and peptide

Mass spectrometry

Mass spectrometry-based peptide sequencing

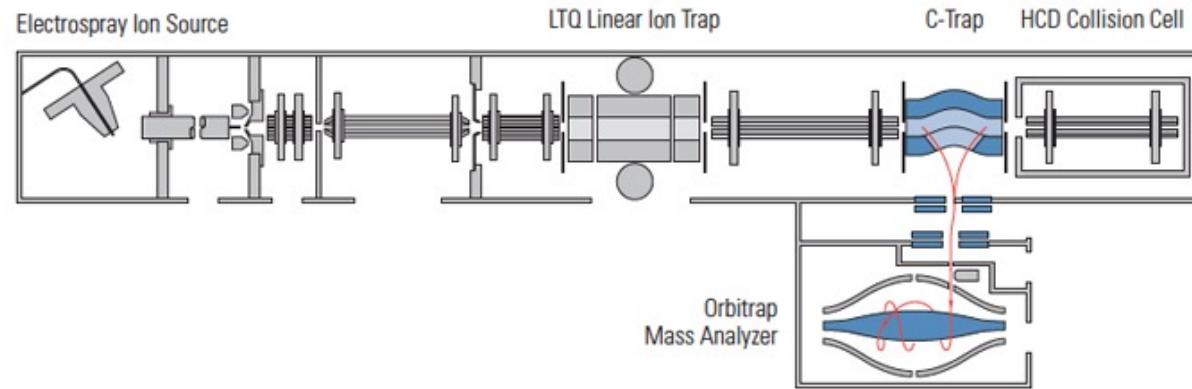


Image from <http://planetorbitrap.com/>

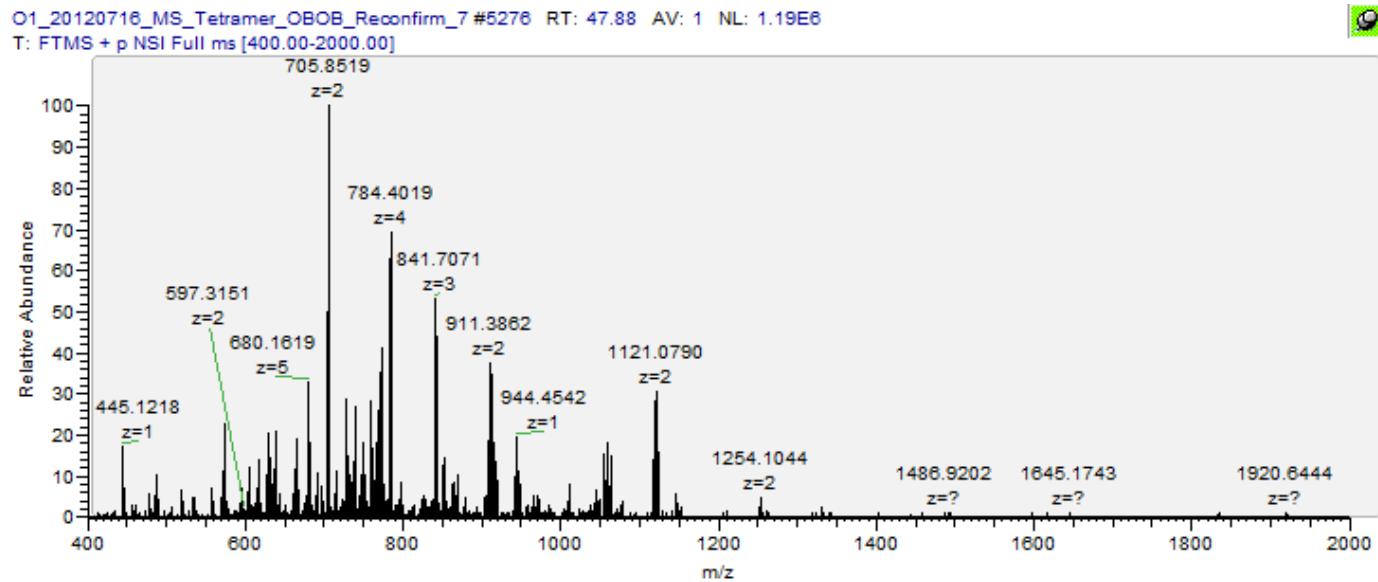
- Ionize peptide molecules (liquid → gas)
- Fragment peptide molecules into prefix/suffix ions
- Measure ions' masses
- Deduce amino acid sequence

What is mass spectrometry?

A really expensive scale!



Abundance profile of all ions



- Measure **mass-to-charge** ratios of ions based on their behaviors in electric-magnetic fields

Mass analyzer varieties

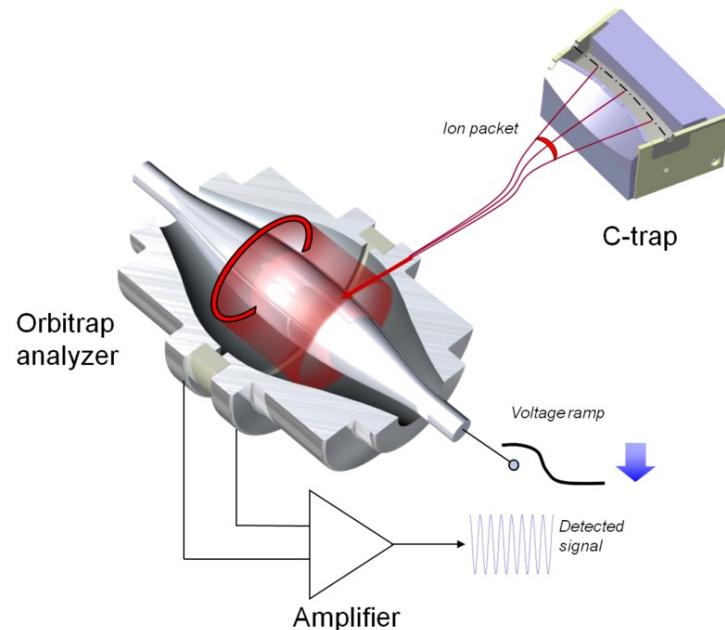
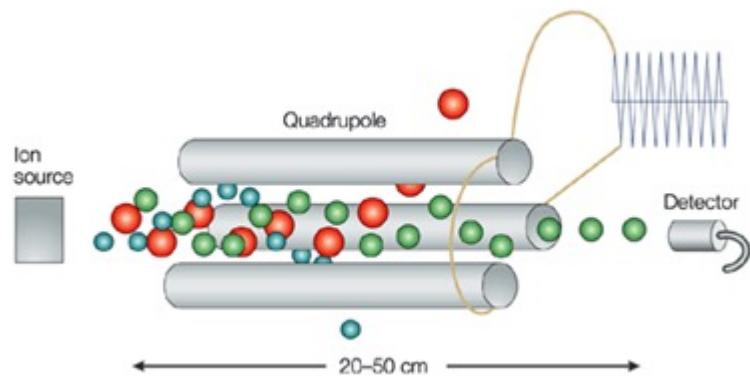
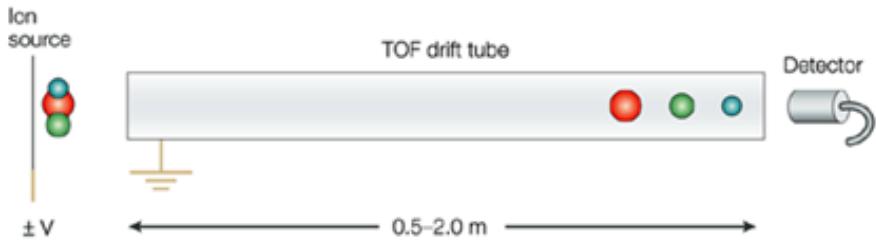
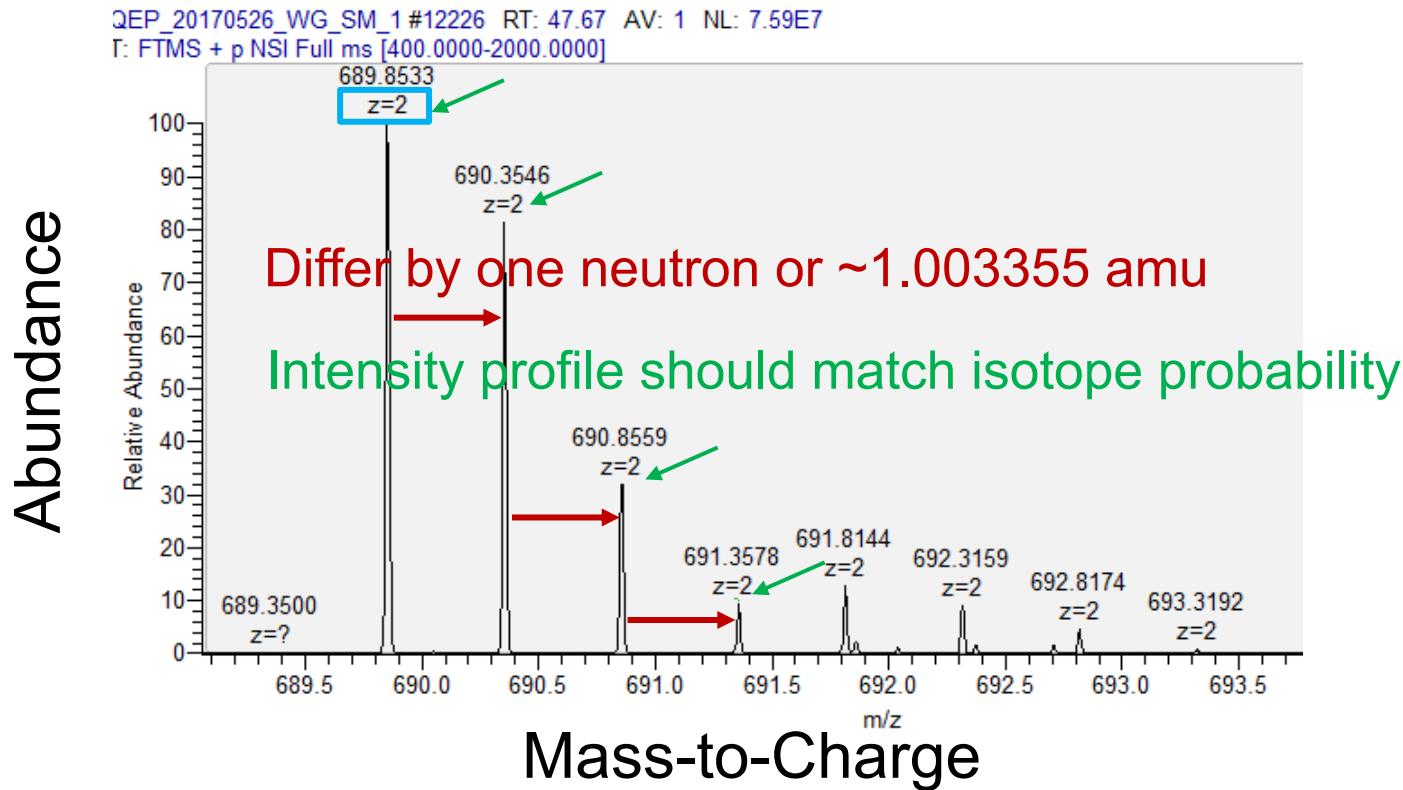


Image from <https://en.wikipedia.org/wiki/Orbitrap>

Glish and Vachet. Nature Reviews Drug Discovery 2, 140-150
(2003)

- **Time-of-Flight (TOF): measure flight time of ions**
 - $qV = \frac{1}{2} mv^2 \rightarrow m / q = 2V / v^2$
- **Quadrupole: select ions using varying electric field**
- **Orbitrap: orbit frequency of ions inside magnetic field**

Identify charge state

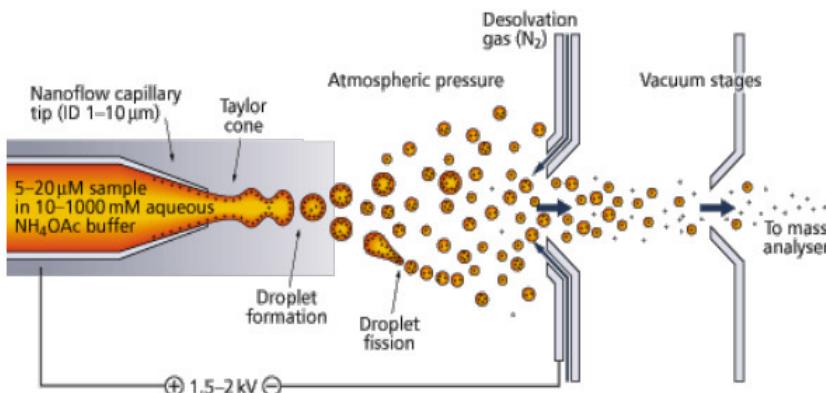


- Carbon, Nitrogen, etc. exist in nature as a mixture of isotopes such as ^{13}C , ^{14}C , ^{15}N , and ^{18}O .
- Differences between adjacent isotopes = 1 neutron

“Soft” ionization for peptides and proteins

- Peptides and proteins are fragile and will break into fragments under regular ionization methods
- Two major techniques:
 - Electrospray Ionization (ESI)
 - Matrix-assisted laser desorption/ionization (MALDI)

ESI



MALDI

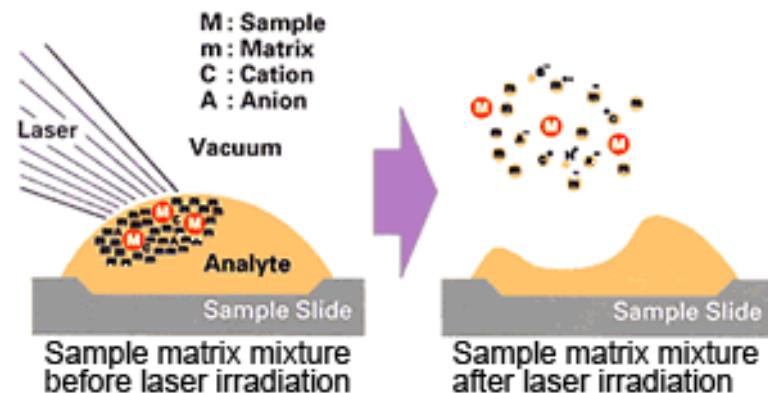


Image from <https://www.thermofisher.com>

Image from <https://www.shimadzu.com>

2002 Nobel Prize in Chemistry



John B. Fenn
Prize share: 1/4



Koichi Tanaka
Prize share: 1/4



Kurt Wüthrich
Prize share: 1/2

Image from <https://www.nobelprize.org>

The Nobel Prize in Chemistry 2002 was awarded "*for the development of methods for identification and structure analyses of biological macromolecules*" with one half jointly to John B. Fenn and Koichi Tanaka "*for their development of soft desorption ionisation methods for mass spectrometric analyses of biological macromolecules*" and the other half to Kurt Wüthrich "*for his development of nuclear magnetic resonance spectroscopy for determining the three-dimensional structure of biological macromolecules in solution*".

An overview of MS experiment (LC-MS/MS)

Protein/Peptide Samples Liquid Chromatography Mass Spectrometry

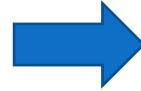
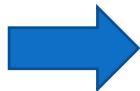


Image from <http://planetorbitrap.com/>

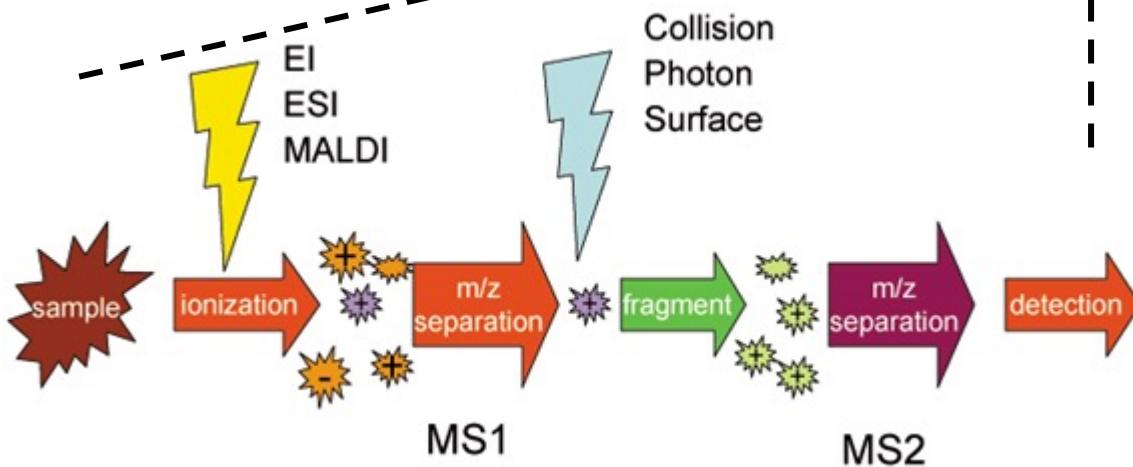


Image from https://en.wikipedia.org/wiki/Tandem_mass_spectrometry

Sample preparation for MS

- Proteins are typically digested into peptides
- Remove salts and small molecule contaminants
- Enrich specific protein/peptide/PTM

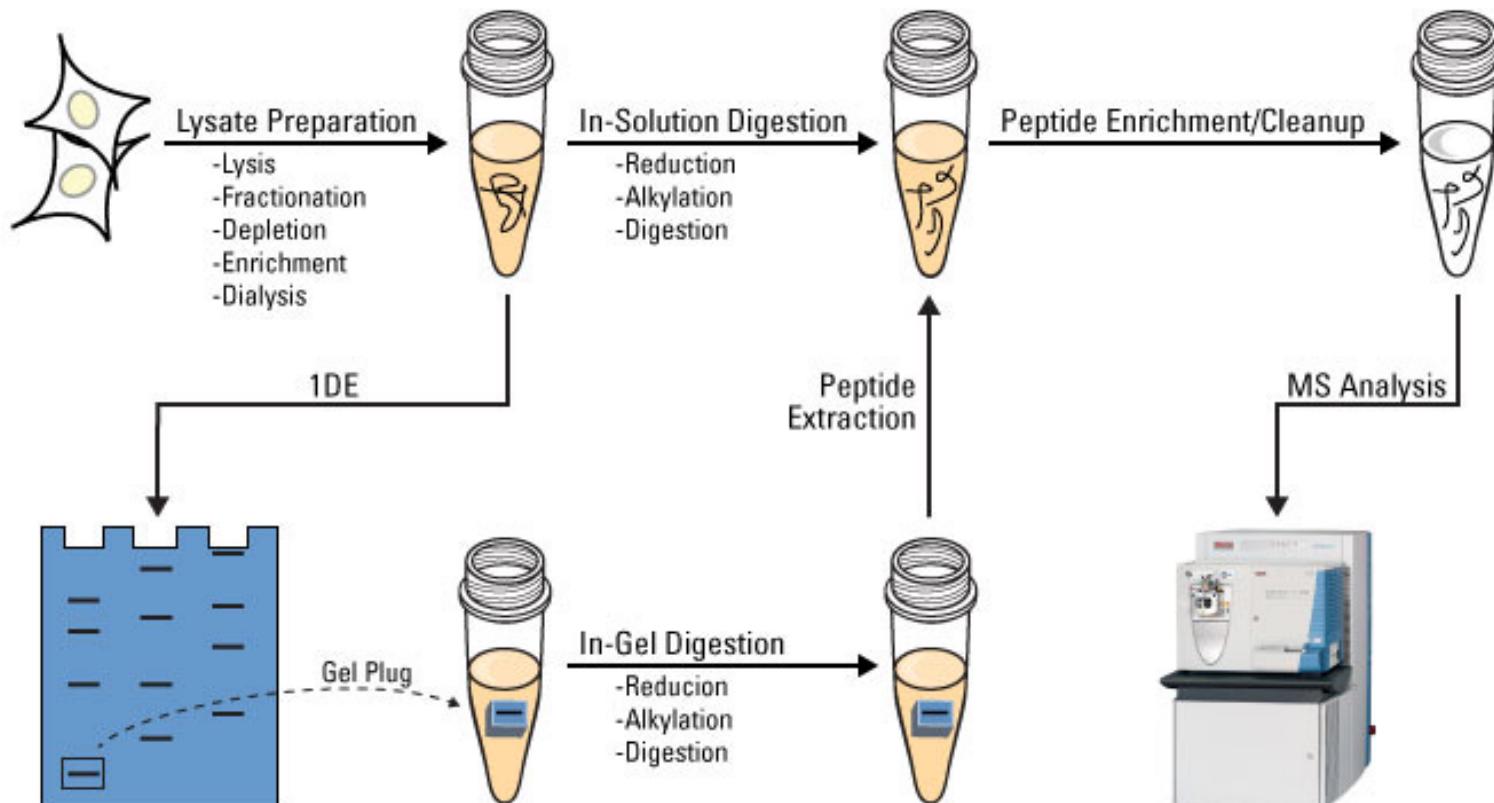
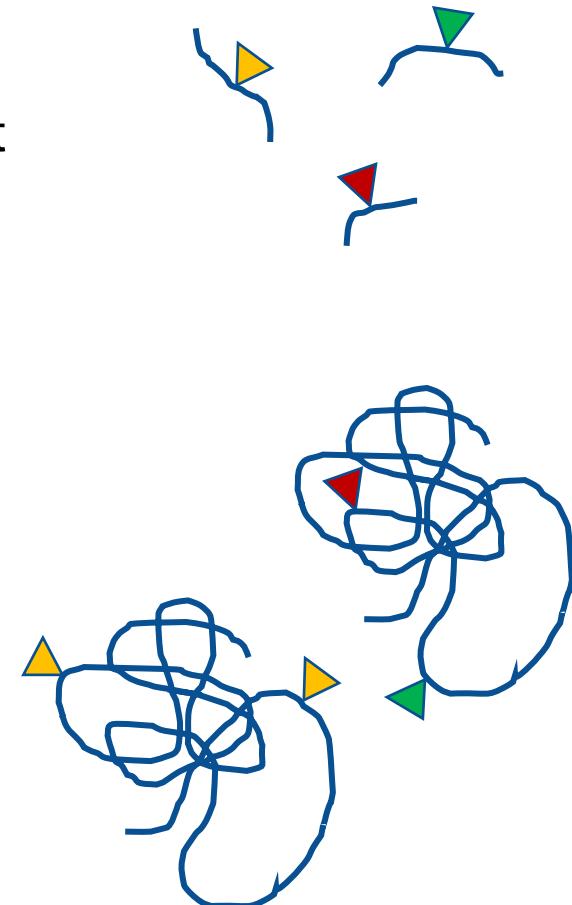


Image from <https://www.thermofisher.com>

Top-down vs bottom-up MS

- Bottom-up
 - Proteins are digested into peptides
 - MS/MS spectra of peptides are easy to interpret
 - Can be applied to any proteome
 - Lose information on some part of proteins

- Top-down
 - Analyze intact proteins
 - MS/MS spectra are difficult to interpret
 - Samples are difficult to prepare
 - Limited to <50 kDa proteins
 - Able to see all modified species of proteins



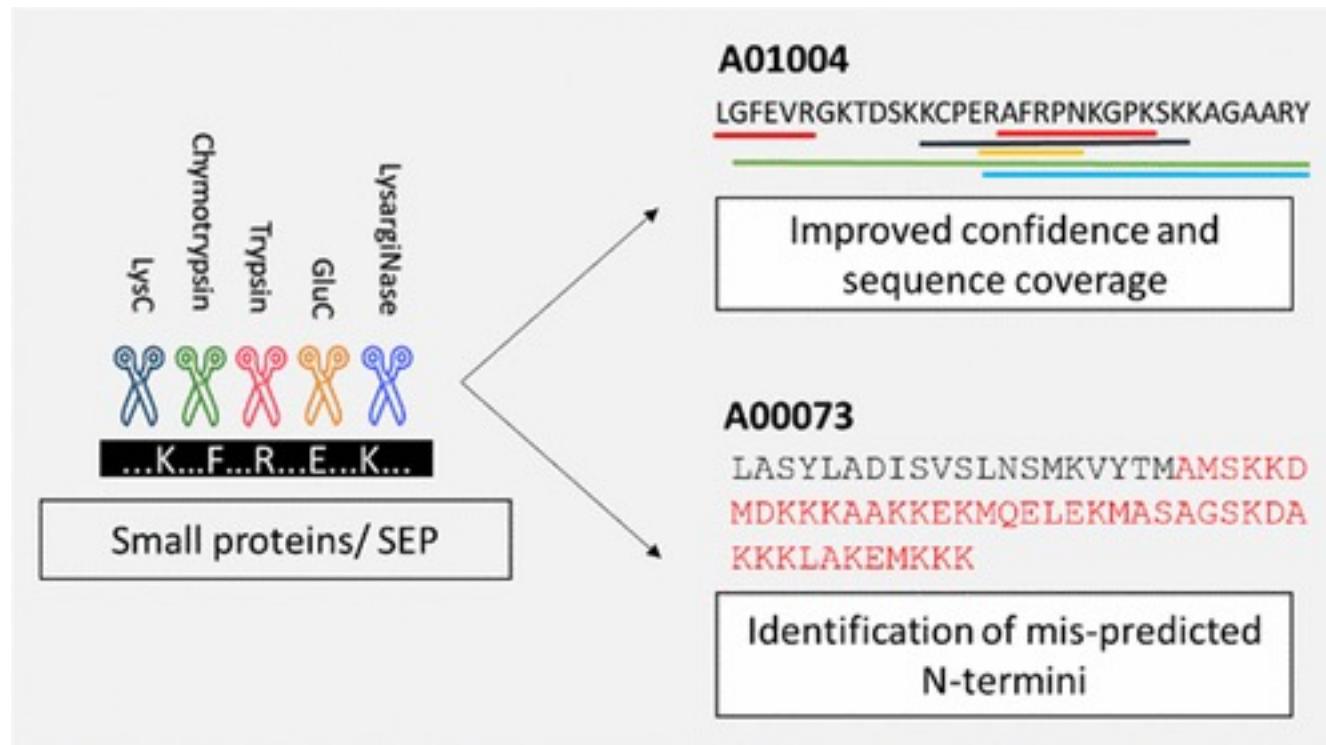
Choices of protease enzyme

| | |
|---------------|--|
| Trypsin | Cleaves after lysine and arginine but not before proline |
| Trypsin/P | Cleaves after lysine and arginine also if a proline follows |
| LysC | Cleaves after lysine but not before proline |
| LysC/P | Cleaves after lysine also if a proline is following |
| D.P | Cleaves D.P pairs. Can be added to other enzymes to include this desired breakage. |
| ArgC | Cleaves after arginine |
| AspC | Cleaves after aspartic acid |
| GluC | Cleaves after glutamic acid |
| GluN | Cleaves before glutamic acid |
| AspN | Cleaves before aspartic acid |
| LysN | Cleaves before lysine |
| Chymotrypsin+ | Cleaves after tyrosine, tryptophane, phenylalanine, leucine, methionine |
| Chymotrypsin | Cleaves after tyrosine, tryptophane, phenylalanine |

Image from MaxQuant software

- Trypsin is typically used, cut at R/K
- What if your protein does not contain many R/K?

Using multiple proteases improve coverage



- Some regions on the protein may not contain any cleavage site for a protease

PTM enrichment

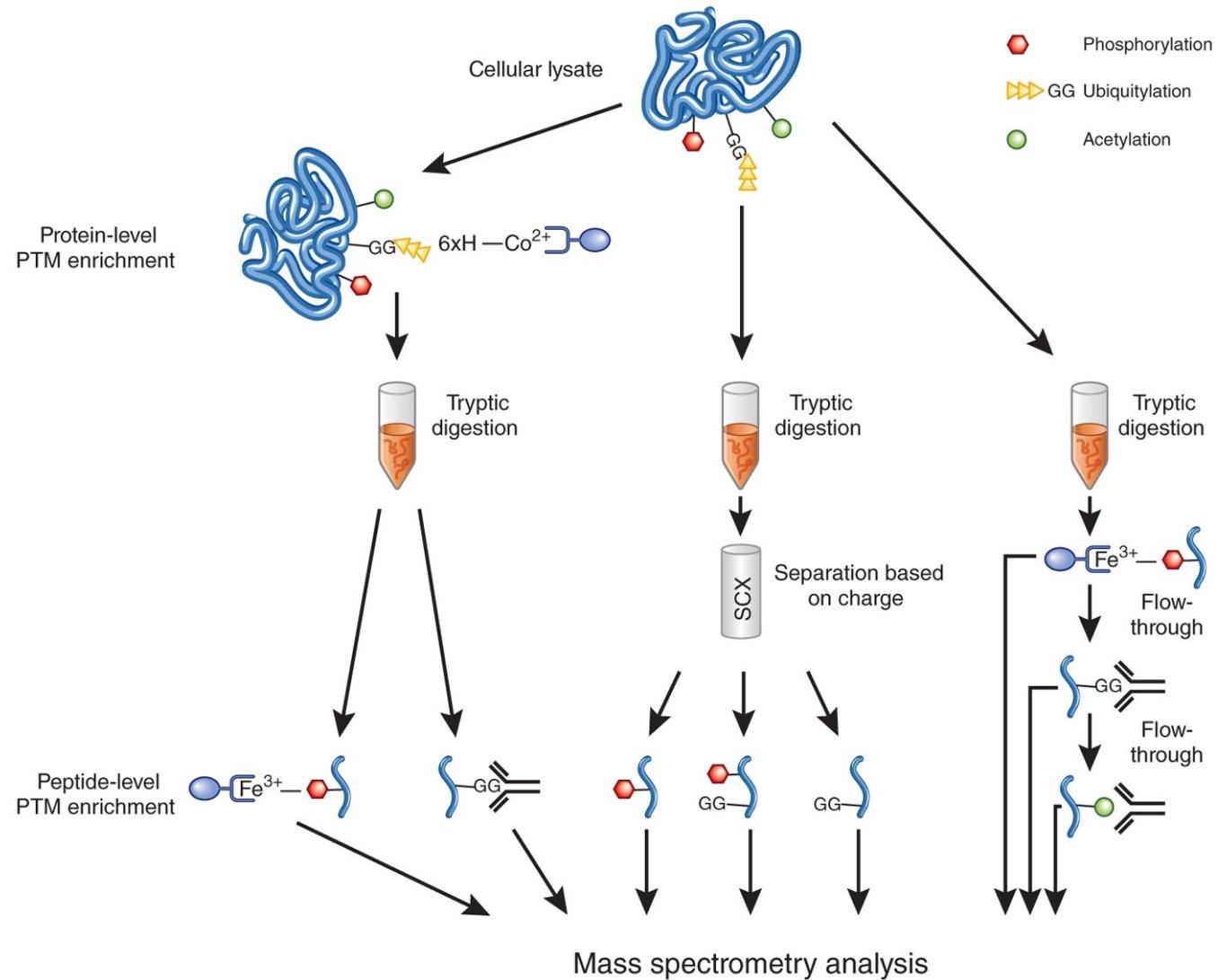


Image from Webb and Bennett. Nature Methods, 10: 620-621 (2013)

Sample fractionation

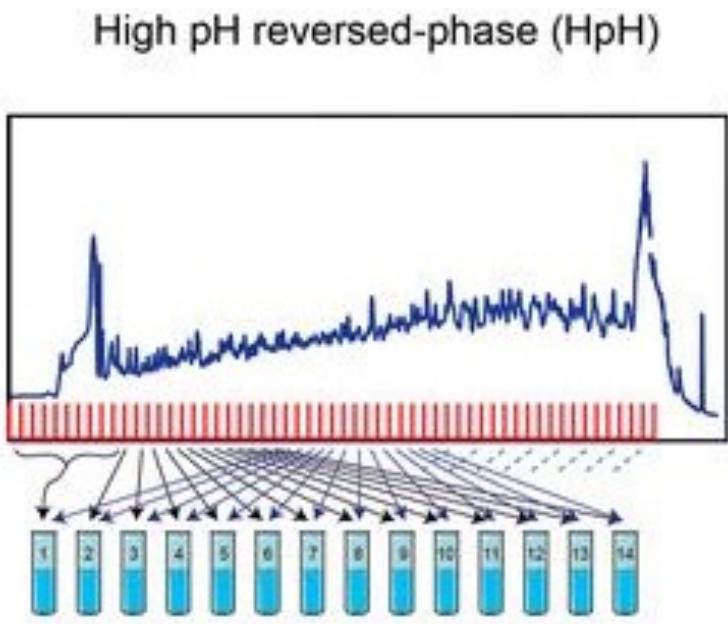


Image from Chen et al. Analyst, 2018

Proteins are eluted with increasing salt (NaCl) gradient

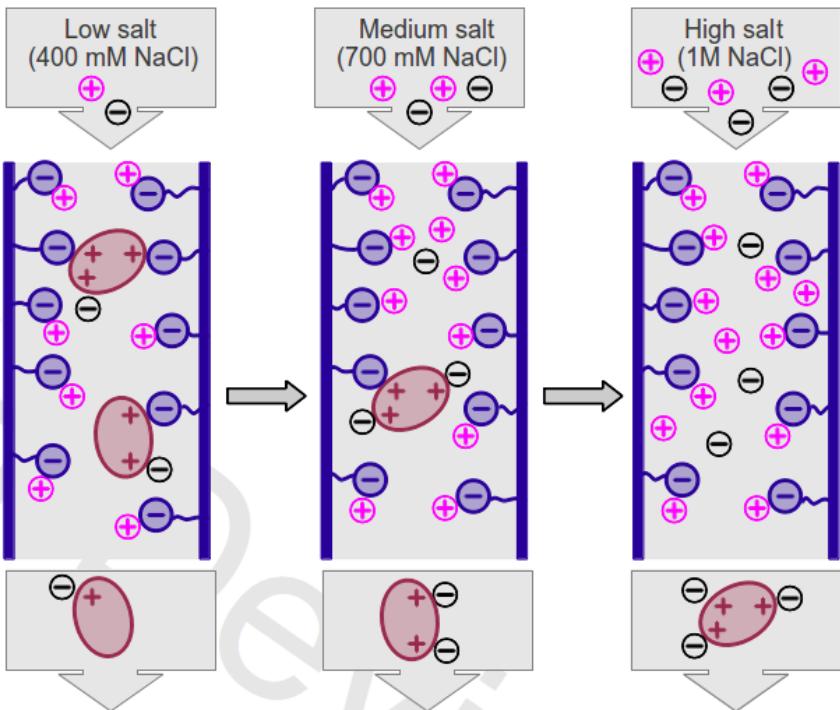


Image from <http://www.reachdevices.com/>

- When a sample is too complex for one LC-MS/MS run, it may be fractionated
- Each fraction = one LC-MS/MS experiment

Ultra-performance liquid chromatography

- UPLC separates peptides by isoelectric potential, hydrophobicity, and size
- Different peptides will elute into the mass spectrometer at different times (called “retention time” of peptide)
- Mass spectrometer receives only a small number of peptides at a time

Injected Sample Band (Appears “Black”) (Blue, Red, Yellow)

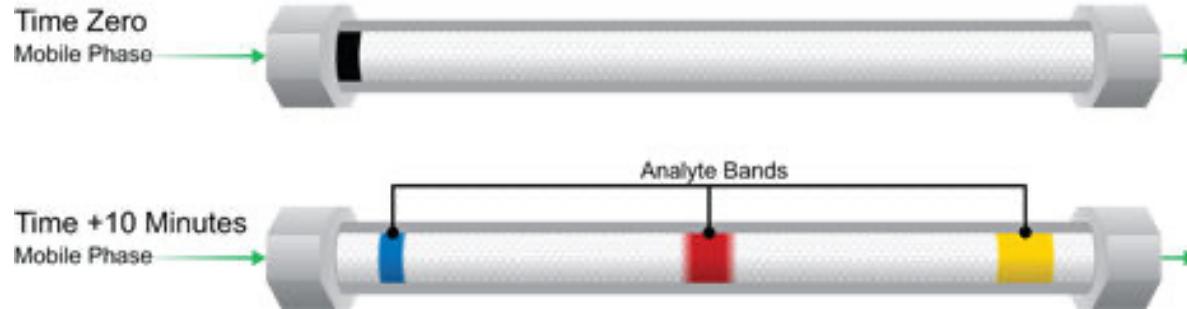
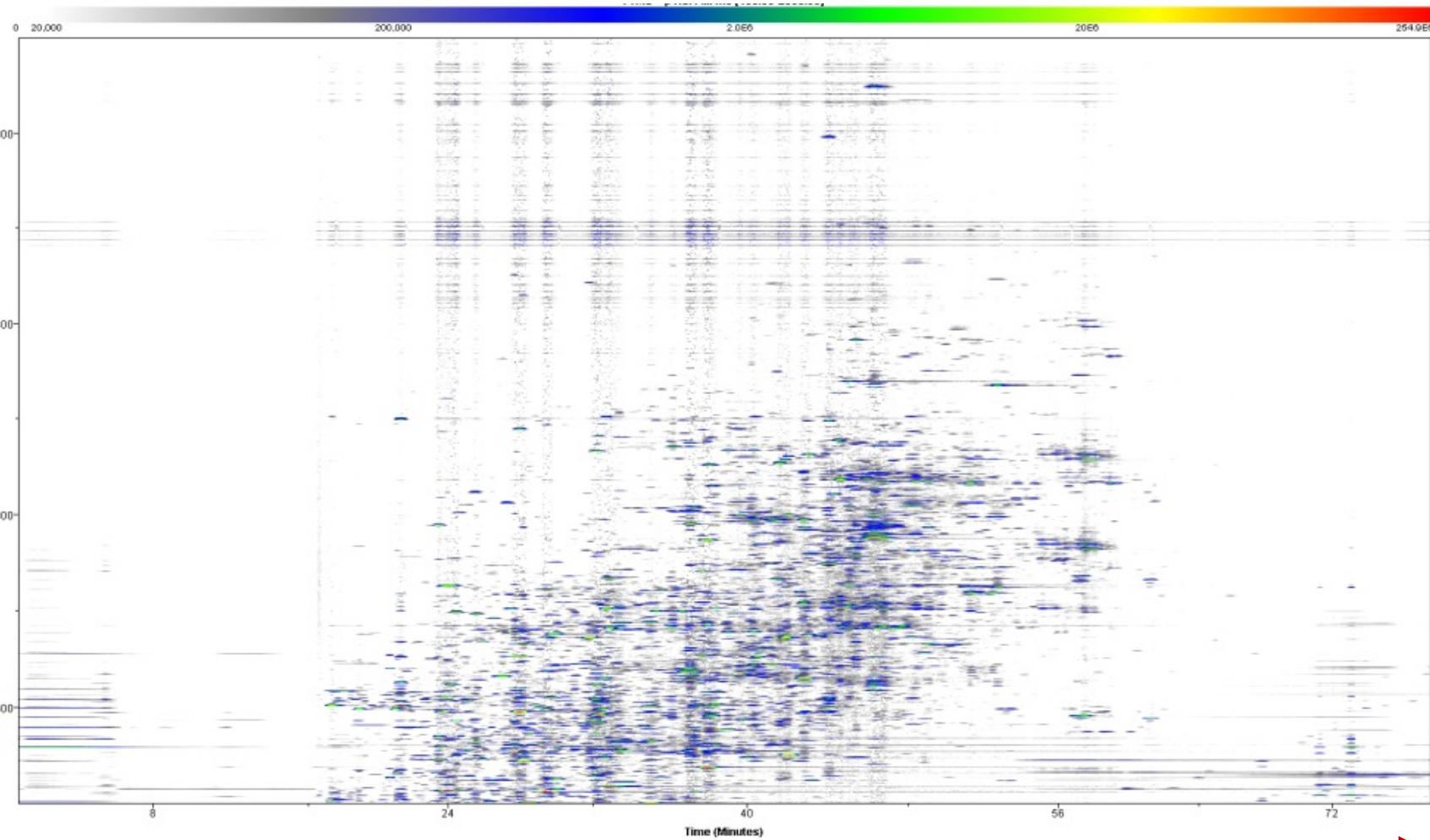


Image from <https://www.waters.com>

Profile of ions seen by MS

Color indicates abundance

Mass-to-Charge

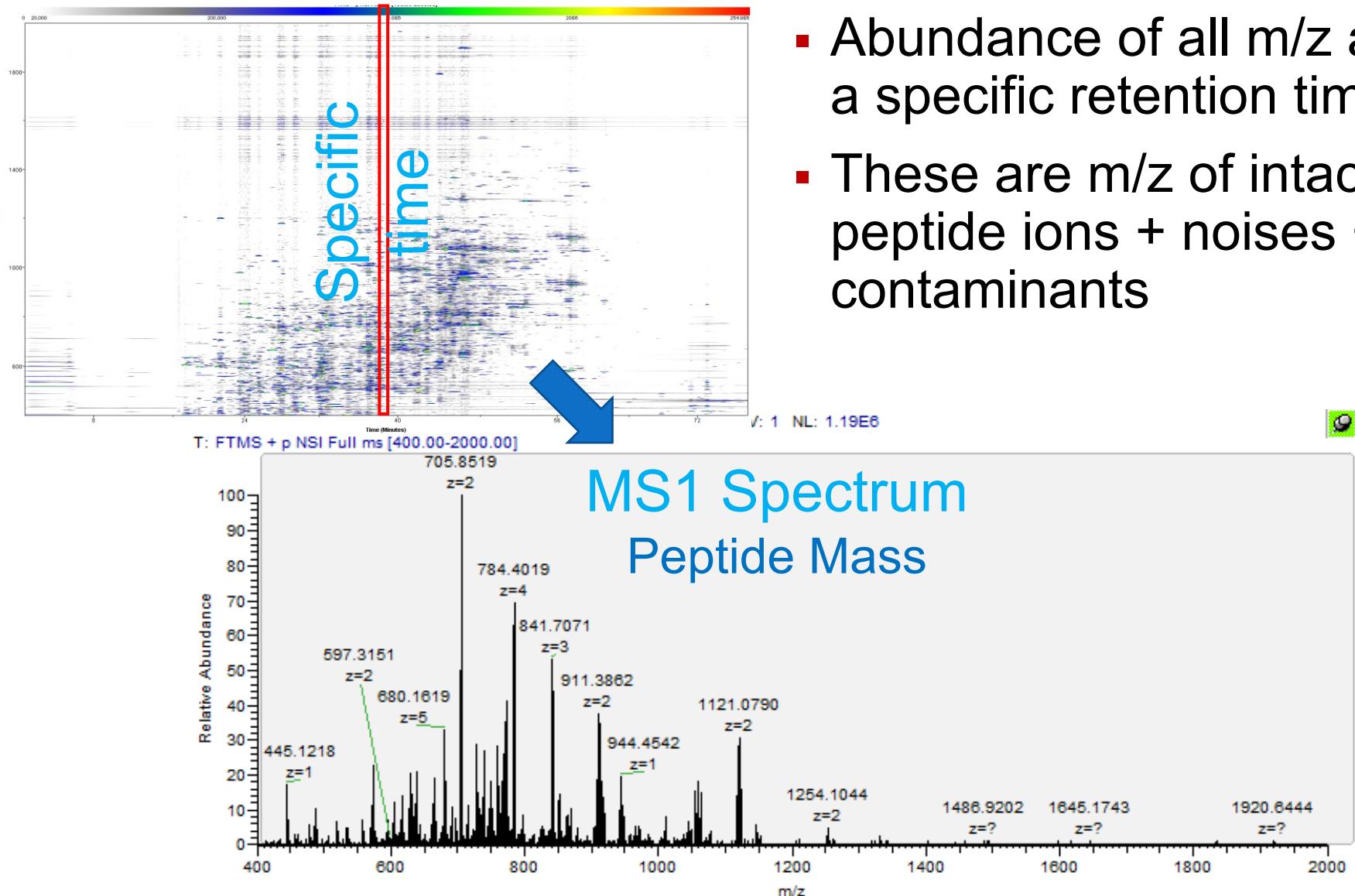


Early ions

Retention Time

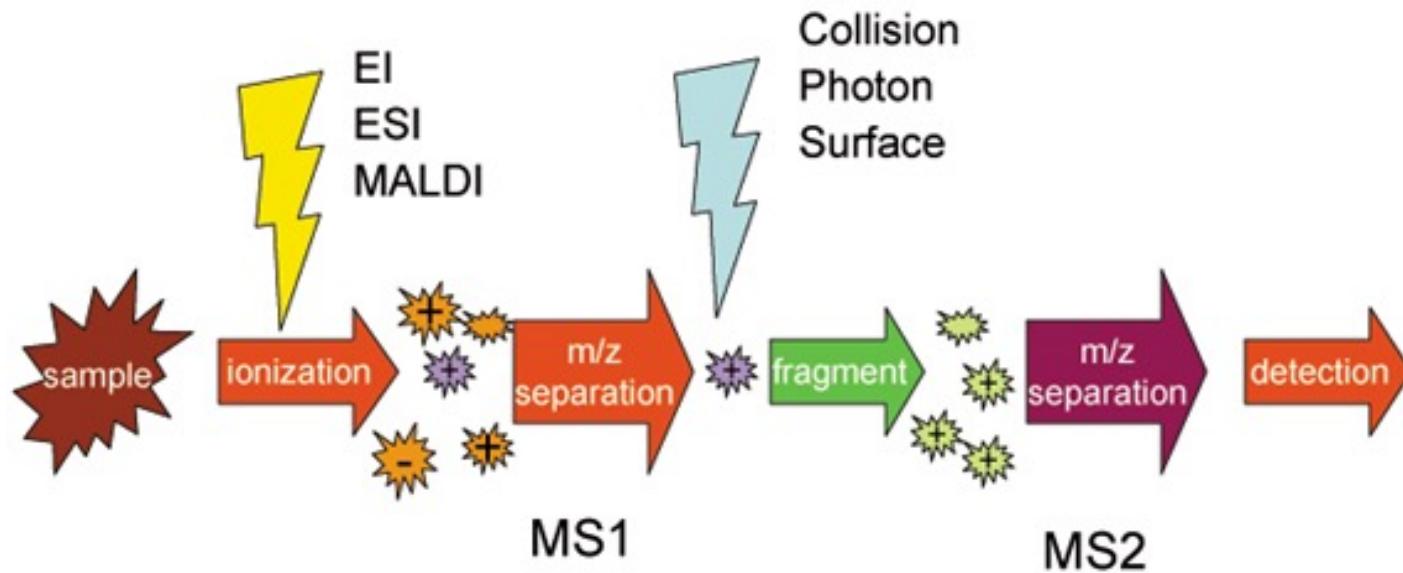
Late ions

Full MS Scan (MS1)



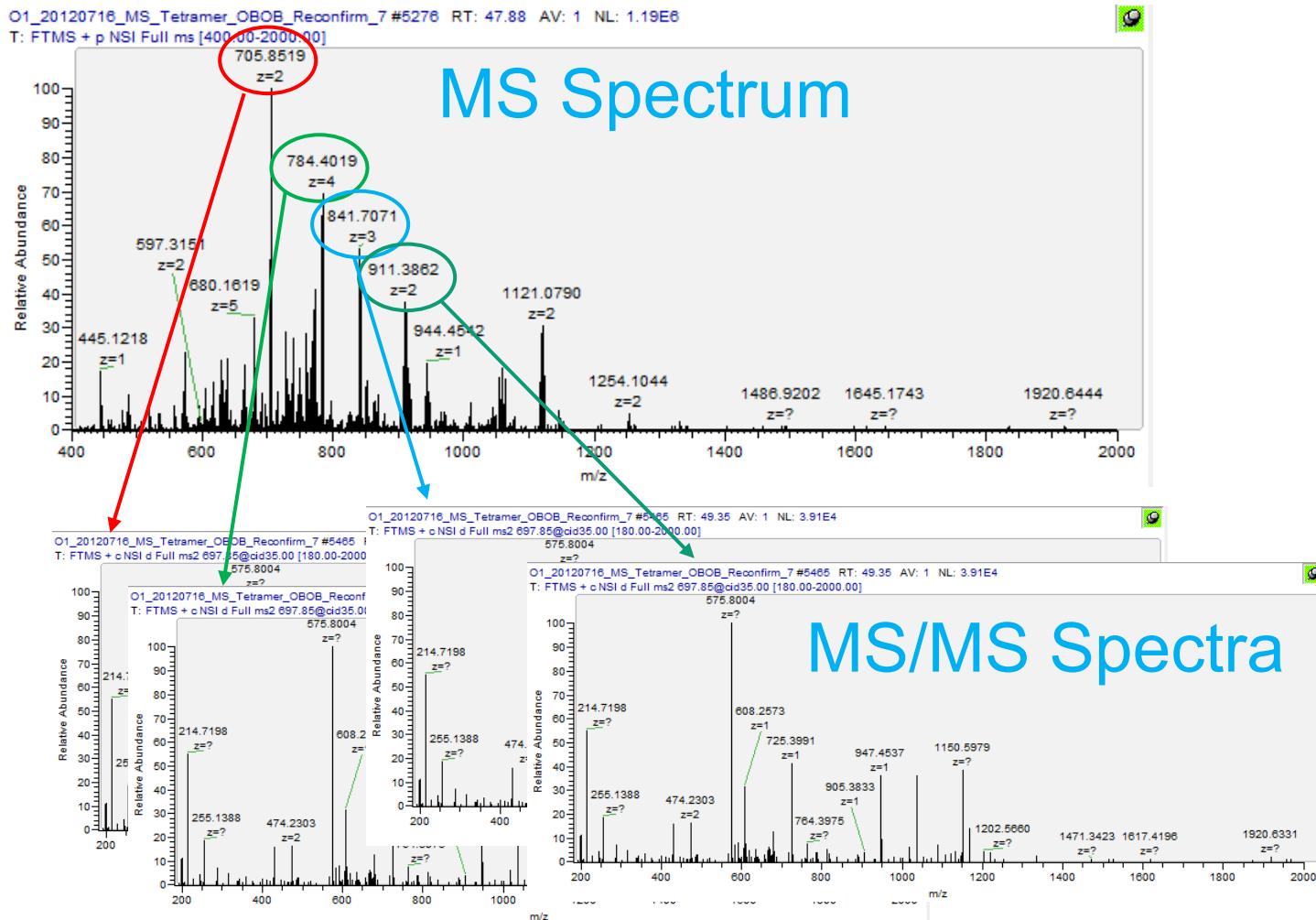
Tandem mass spectrometry

- MS1 spectrum tells us the **mass of intact peptides**
- But mass alone cannot pinpoint the identity of the molecule!
- Ions in MS1 spectrum can be selected for further MS analyses (called MS2 or MS/MS)



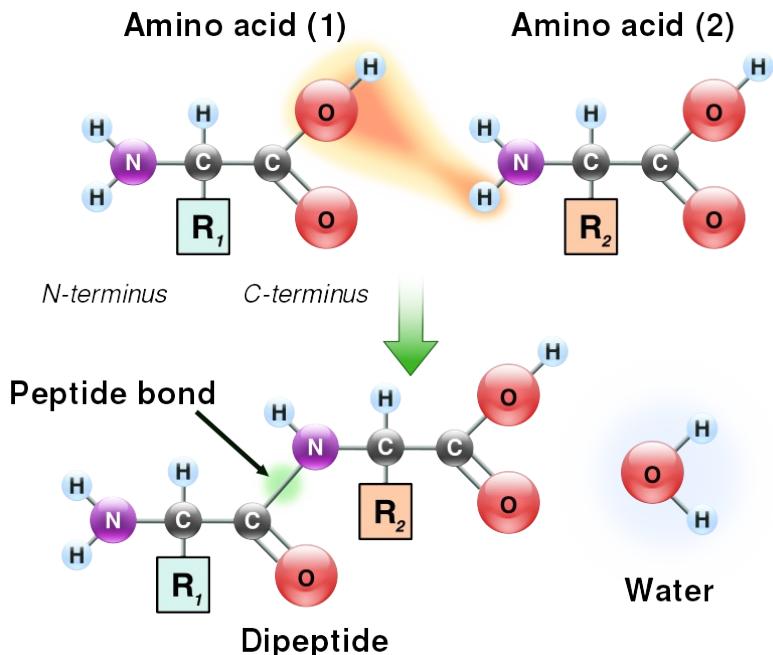
Generation of MS/MS spectra

- Top 5-20 most abundant ions in each MS spectrum will be isolated for individual MS/MS analyses



Peptide fragmentation

- Peptide is a polymer with amino acids as monomers
- Breaking a peptide bond yield prefix (b) and suffix (y) ions
- MS/MS spectra capture the profile of b-ions and y-ions



Fragmentation of peptide yields b- and y-ions

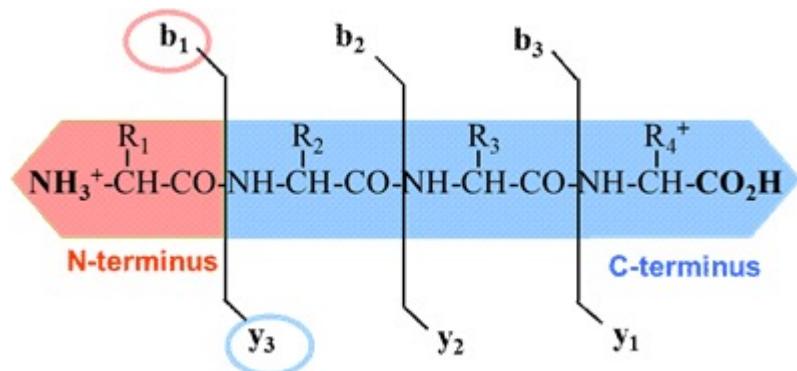
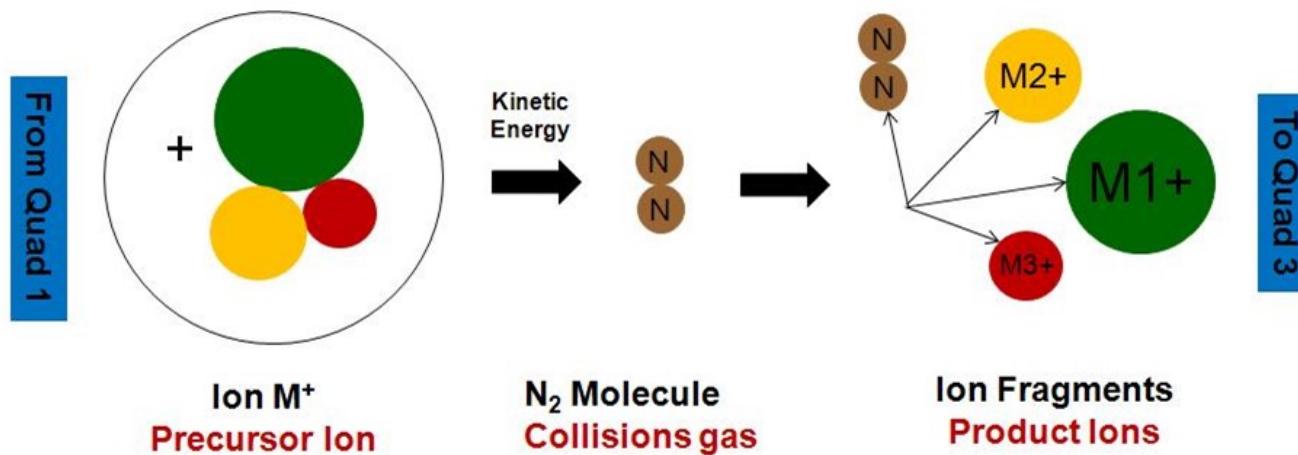


Image from https://en.wikipedia.org/wiki/Peptide_bond

Image from <https://www.molgen.mpg.de>

Collision-based dissociation

- **Collision-induced dissociation (CID)**: peptide ions are accelerated, and then collide into neutral gas such as helium, nitrogen, or argon.
- **Higher-energy collisional dissociation (HCD)**: is a variant of CID specific to Orbitrap mass spectrometer. HCD permits observation of ions with low masses and is therefore useful for techniques that generate low-mass reporter ions.



Electron-transfer dissociation (ETD)

- Radical anions are introduced which causes the transfer of an electron to each peptide ion.
- Peptide ions with extra electron become unstable and break at one of the N-C_α bond, resulting in c- and z-ions.

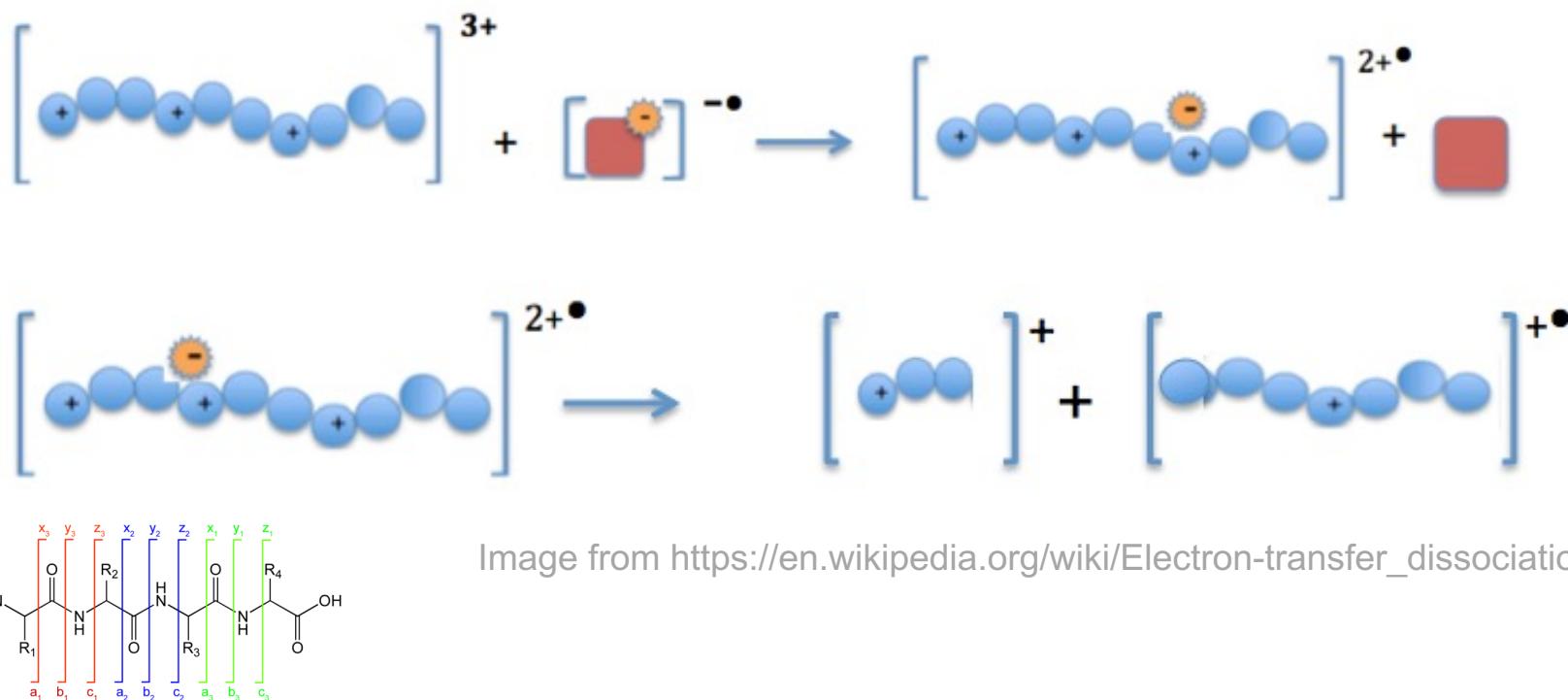
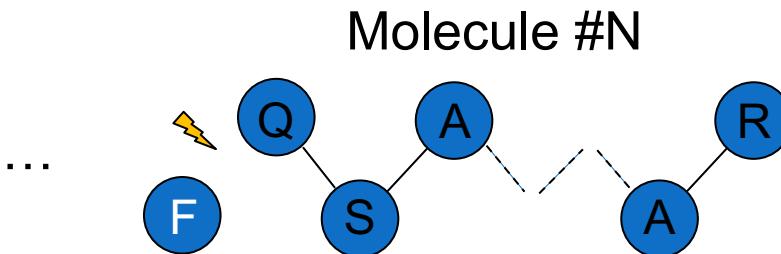
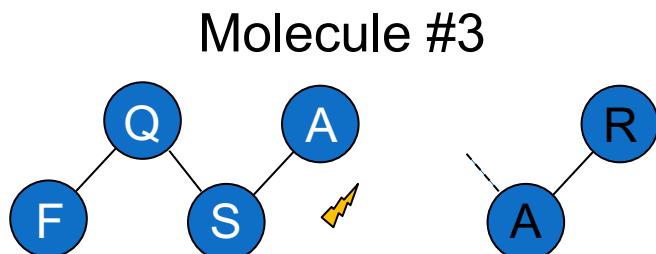
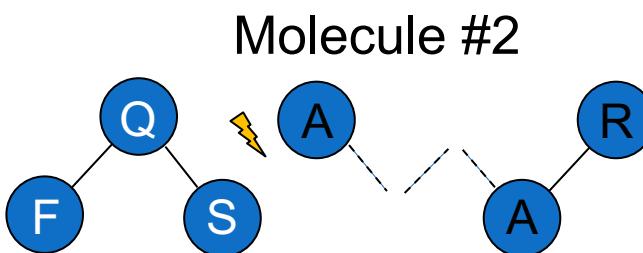
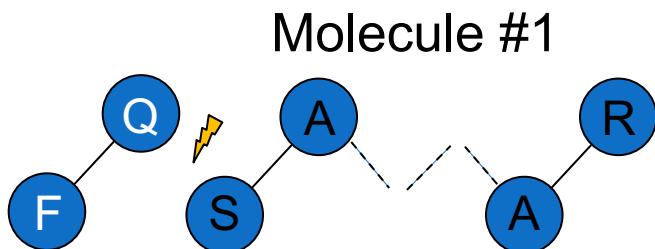
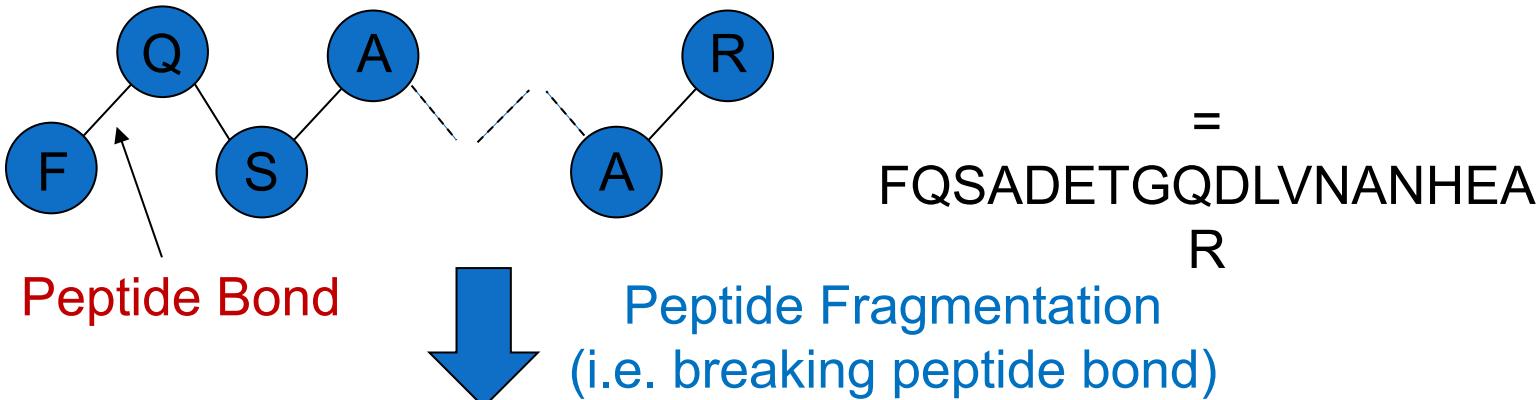


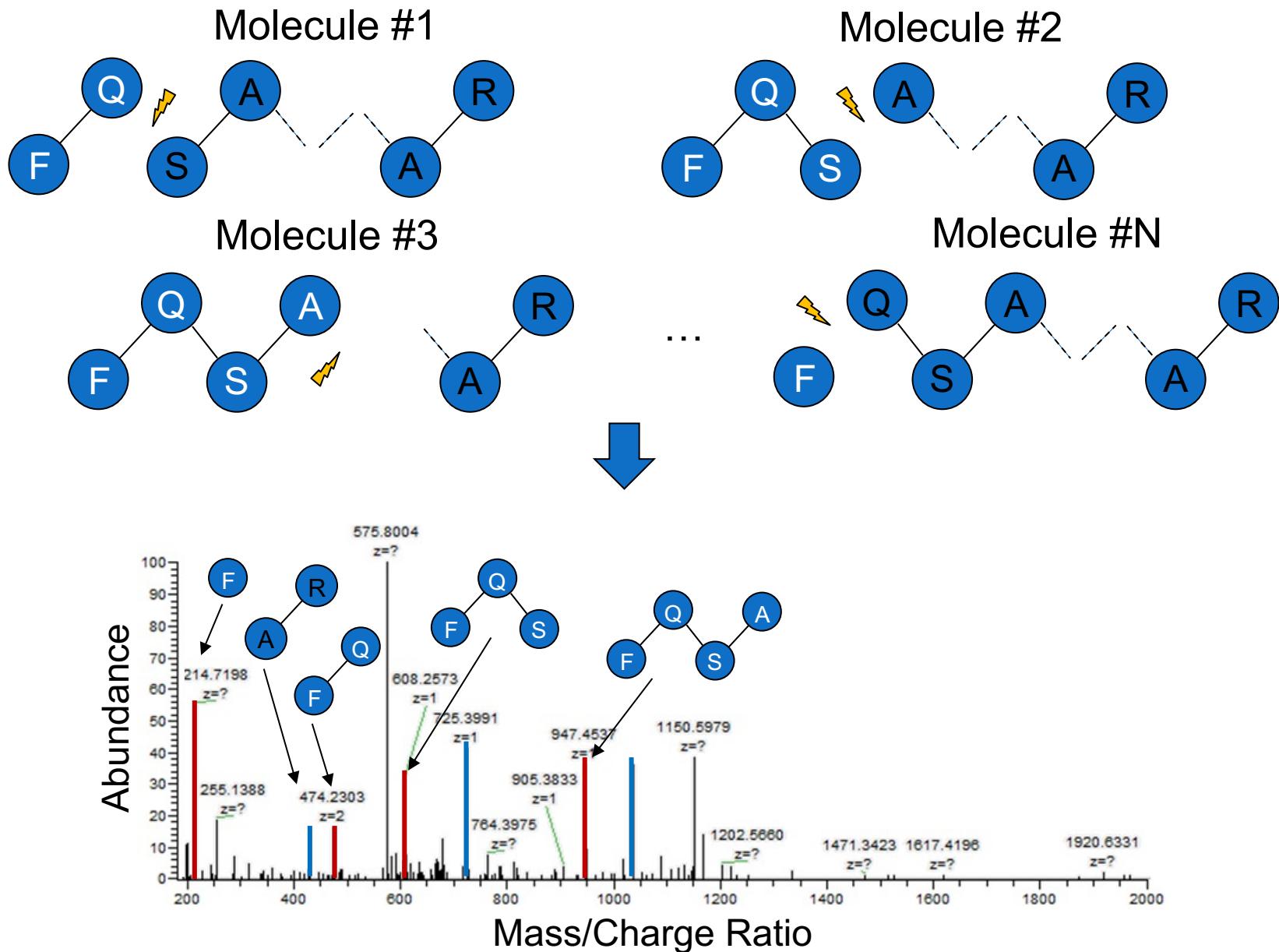
Image from https://en.wikipedia.org/wiki/Electron-transfer_dissociation

Peptide sequencing

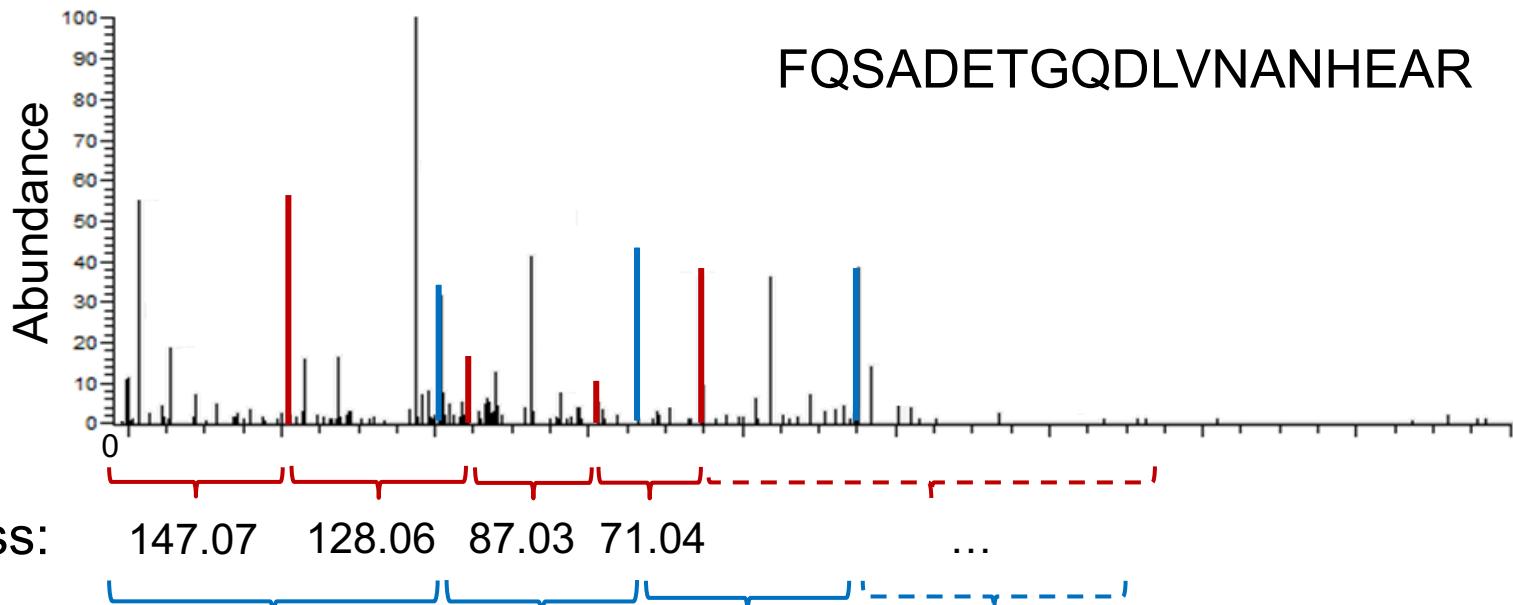
Peptide fragmentation (cont'd)



MS/MS spectrum = profile of b-ions and y-ions



Peptide sequencing from MS/MS data



Amino Acid Mass Table

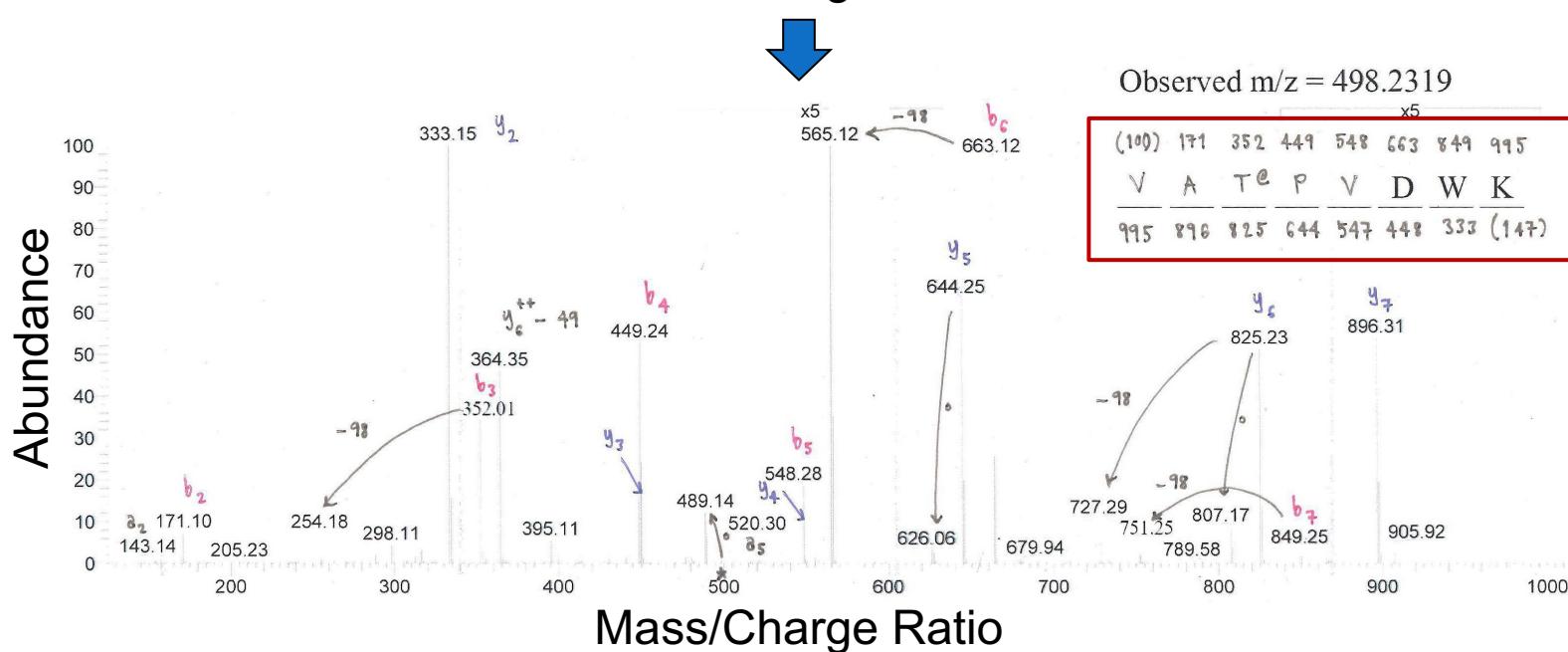
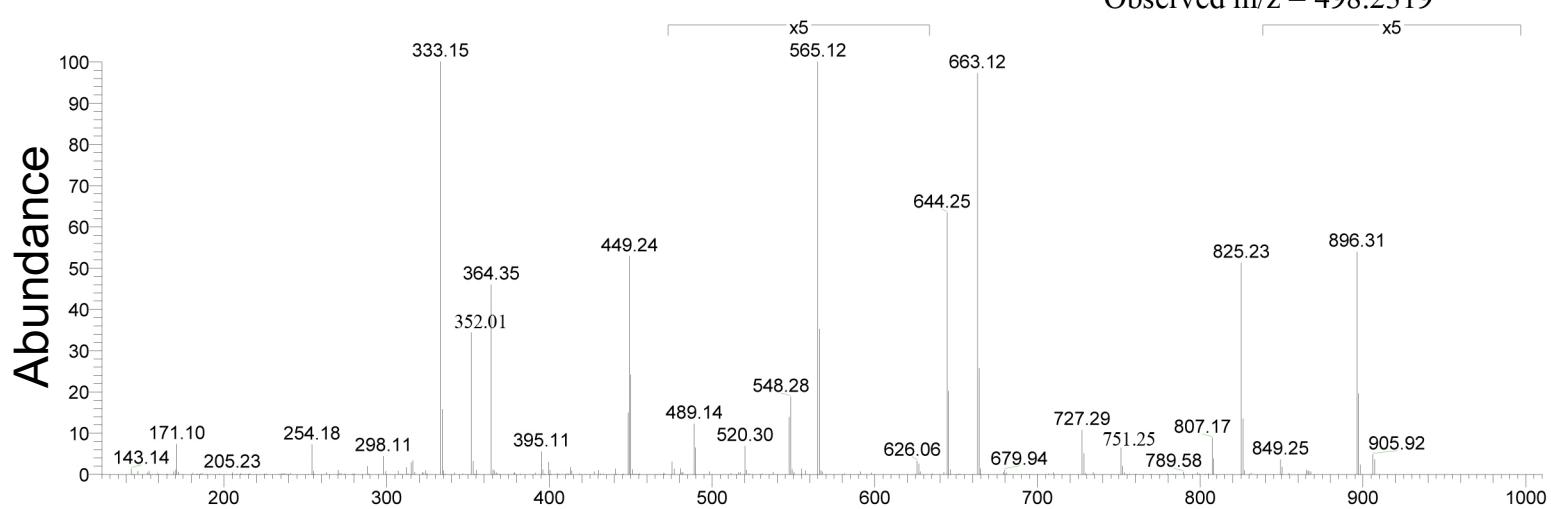
| Amino Acid | Symbol | Residue Mass |
|------------|--------|--------------|
| Glycine | G | 57.021463 |
| Alanine | A | 71.037114 |
| Serine | S | 87.032028 |
| ... | ... | ... |



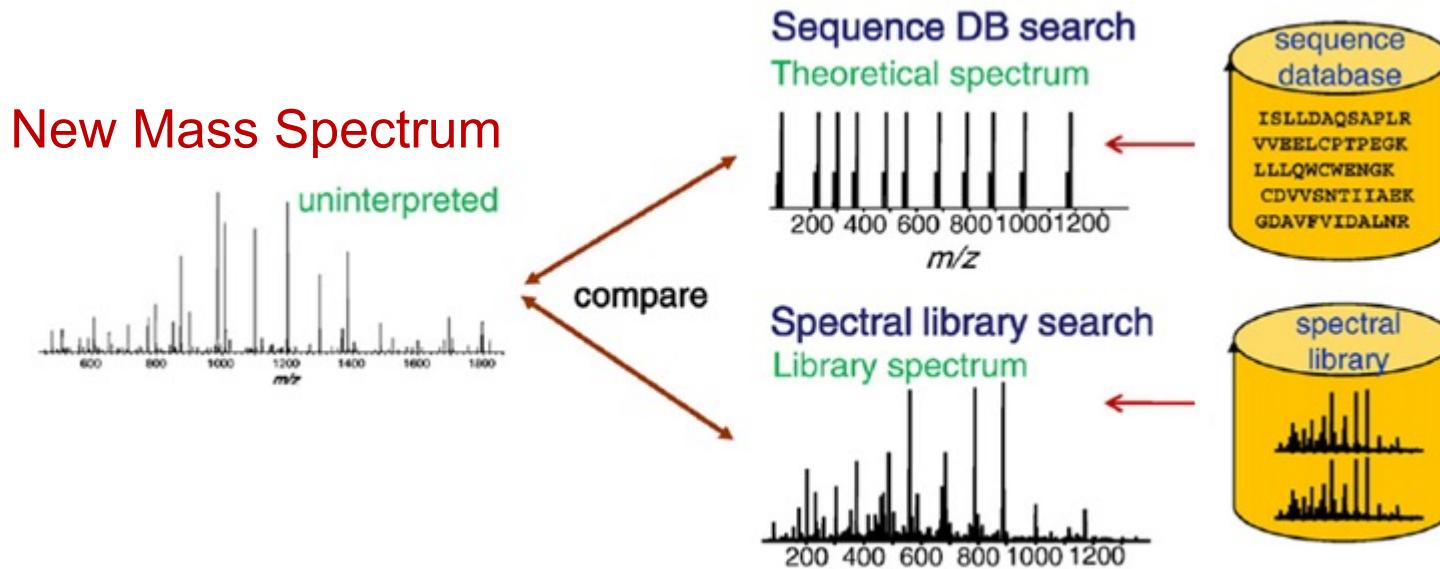
| | |
|--------|-------|
| 147.07 | F |
| 128.06 | Q |
| 87.03 | S |
| ... | ... |
| 227.14 | A + R |
| 129.04 | E |
| ... | ... |

Sequence Direction
→ ↑ ↓ ←

Manual peptide sequencing



Database search approaches



Adapted from Nescizhskii. Journal of Proteomics 73: 2092-2123 (2010)

- Direct peptide sequencing is difficult
- Instead, we compare new MS/MS spectra to databases of known peptides or previously annotated MS/MS spectra

Popular search engines

- **MaxQuant** (free, Max Planck Institute of Biochemistry)
 - **Comet** (free, open source, integrated with other pipelines)
 - **X! TANDEM** (free, open source, The Global Proteome Machine Organization)
-
- **Mascot** (license, Matrix Science)
 - **Proteome Discoverer** (license, Thermo Scientific)
 - **Spectrum Mill** (license, Agilent)

Limitation of database search

- Database search returns **the best known answer**
- Cannot identify sequences not included in the database
 - New species
 - Uncharacterized mutations/polymorphisms
 - Unexpected PTM
- Using large databases results in a lot of false positives
 - MS/MS spectra provide incomplete information
 - Different peptide can produce similar profiles of b-ions and y-ions

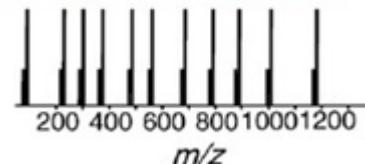
Adding PTM to database-search

Without PTM

FQSADETMAR



Theoretical spectrum

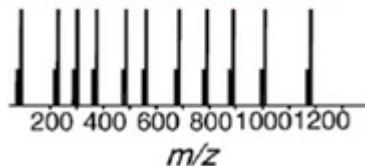


With MetOx

FQSADETMAR



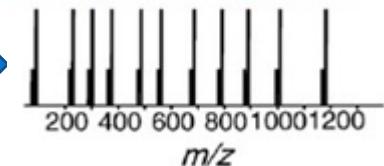
Theoretical spectrum



FQSADETmAR



Theoretical spectrum



With MetOx and STY phosphorylation

FQSADETMAR

FQSADETmAR

FQsADEtMAR

FQsADEtMAR

FQsADETmAR

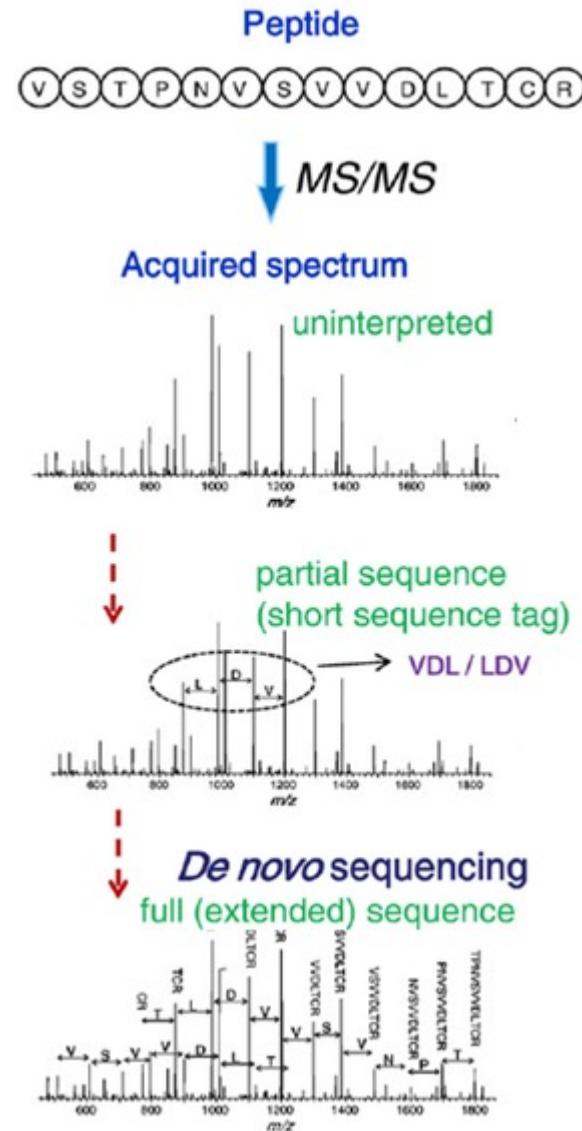
FQsADEtMAR

FQSADeTMAR

FQSADEtMAR

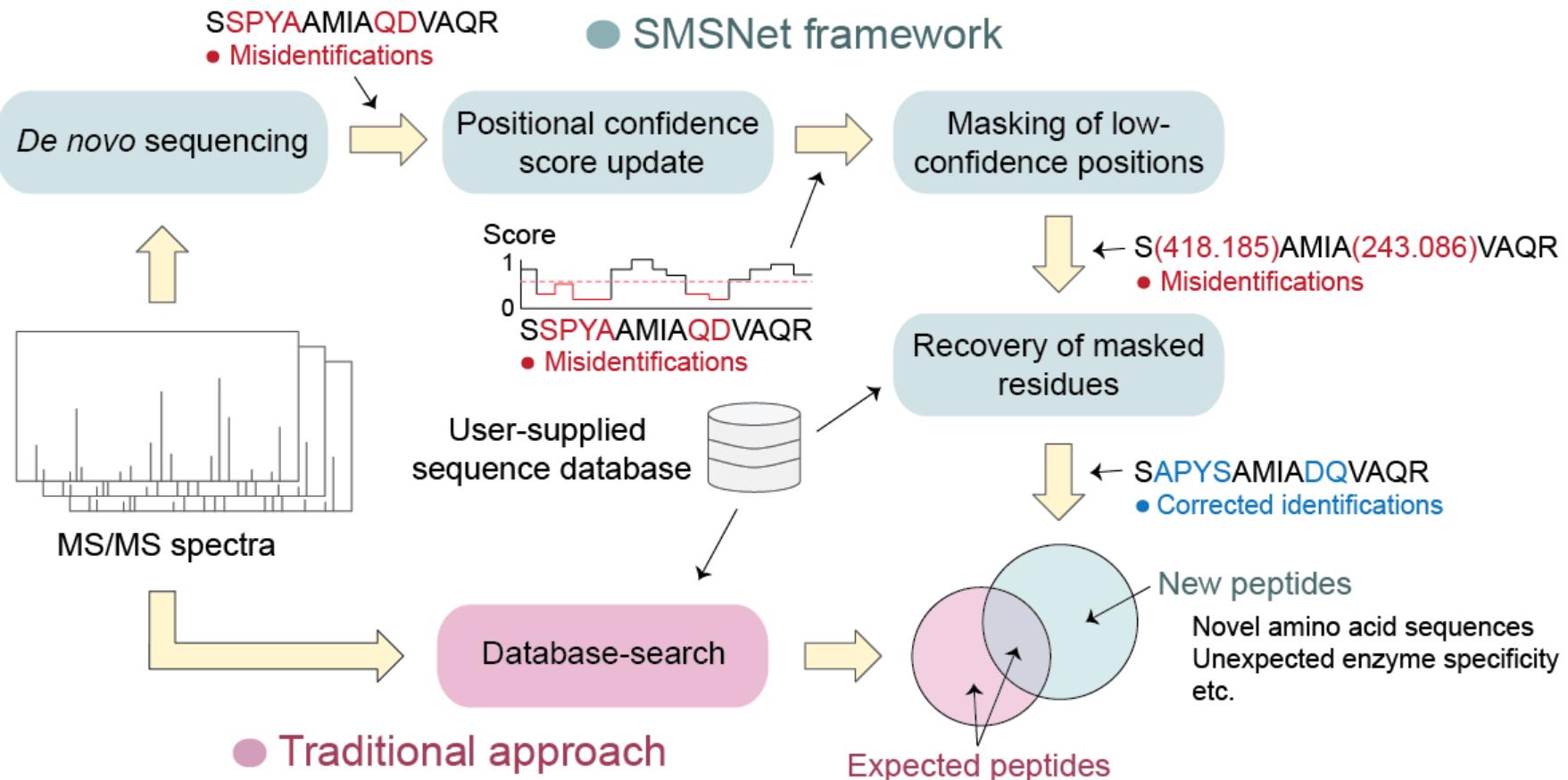
De novo peptide sequencing

- Direct deduction of amino acid sequence from MS/MS spectra
- Can provide partial sequence tag for filtering large database
- **Software:** SMSNet, PEAKS, PepNovo, DeepNovo, Novor



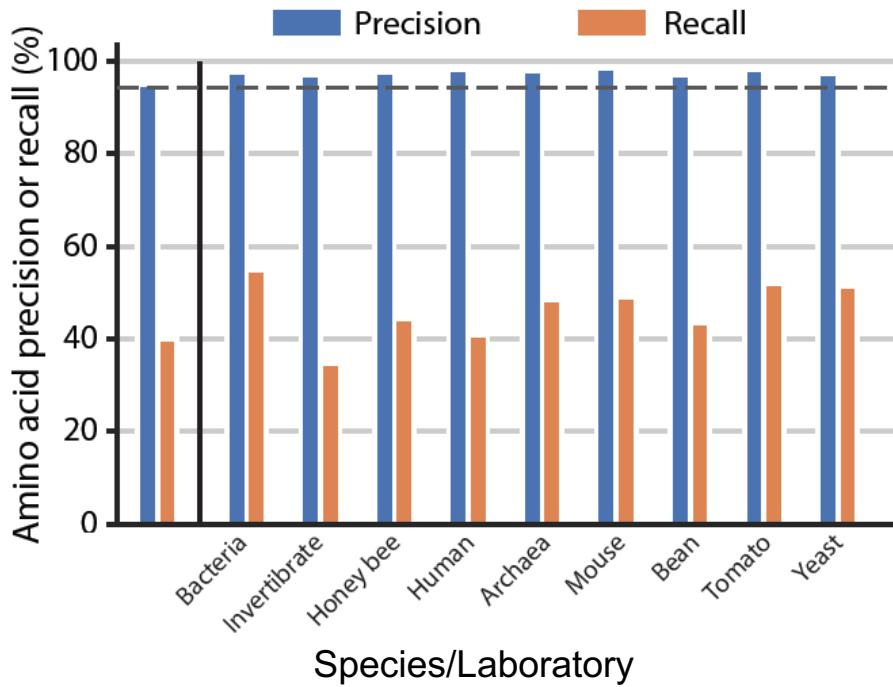
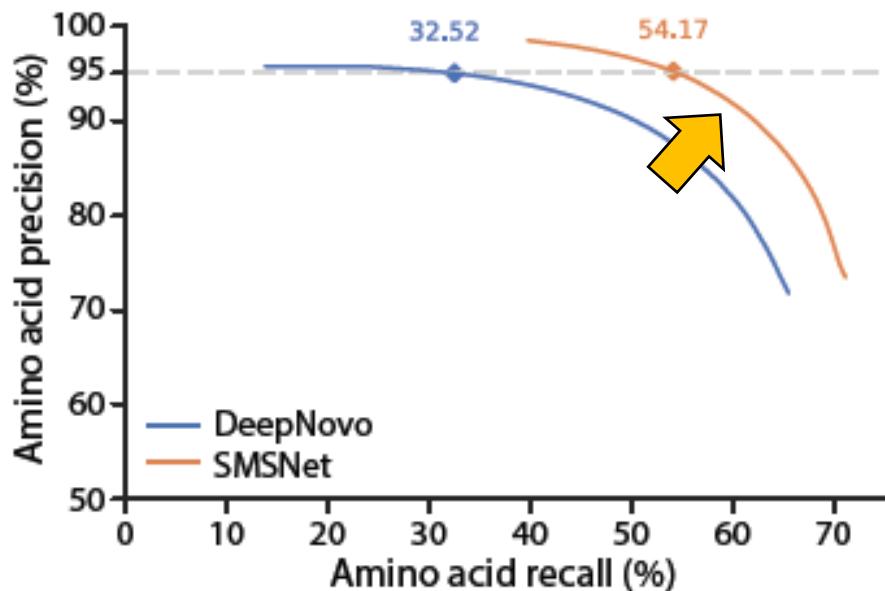
Adapted from Nescizhskii. Journal of Proteomics 73: 2092-2123 (2010)

De novo-assisted database search



Source: Karunratanakul *et al.*, MCP 18:2478-91 (2019)

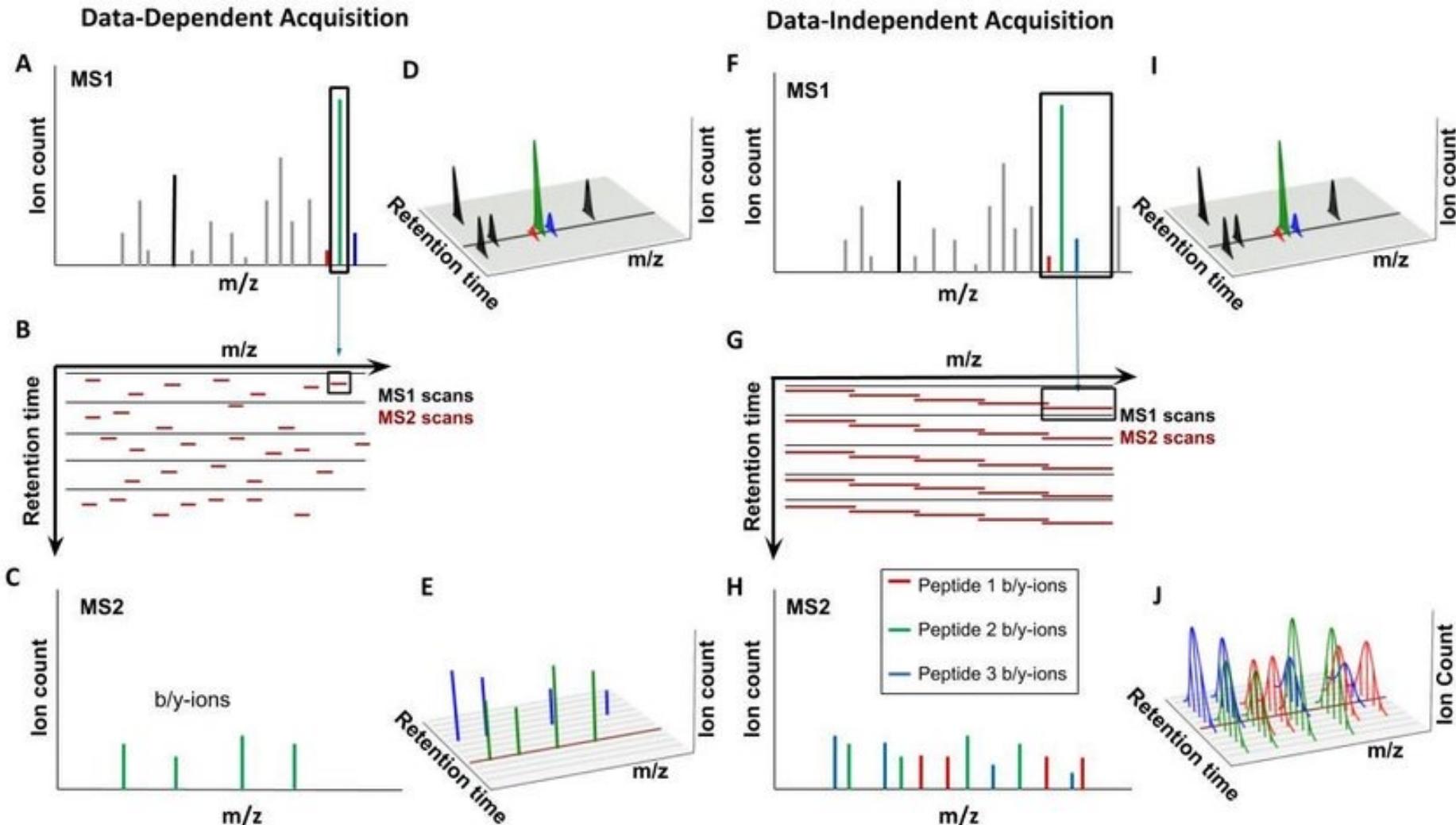
Current state-of-the-art performance



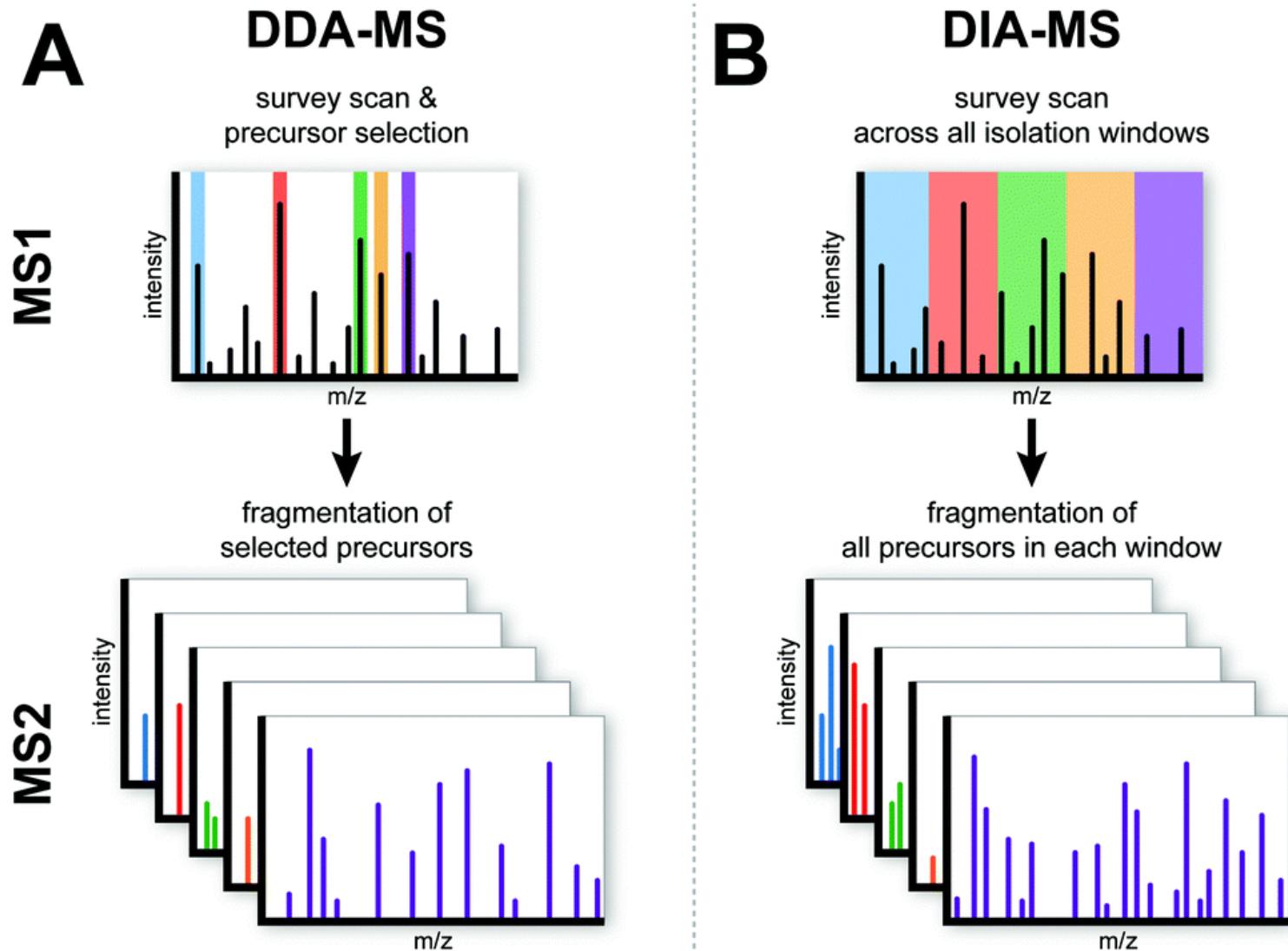
- *De novo* peptide sequencing is far from being solved
- Recall (sensitivity) is still only ~40-60%

Data-independent MS

Data-independent MS/MS acquisition

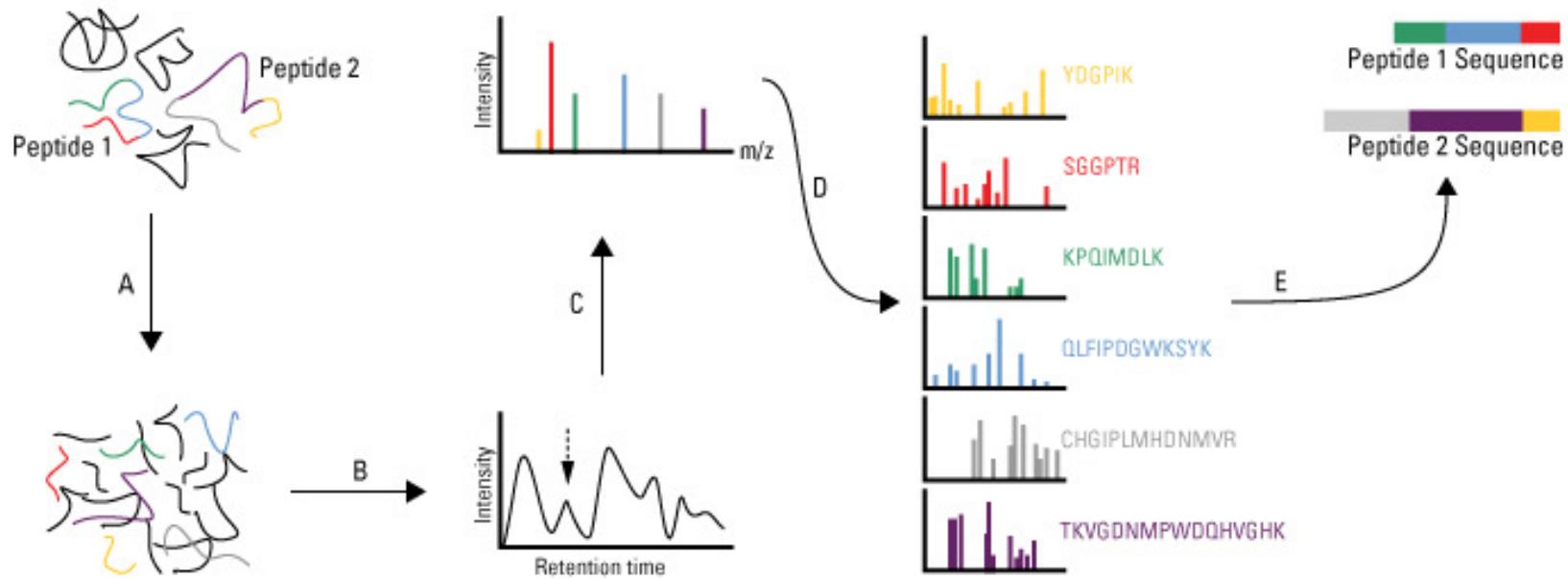


DIA vs DDA mass spectra



Peptide and protein quantification

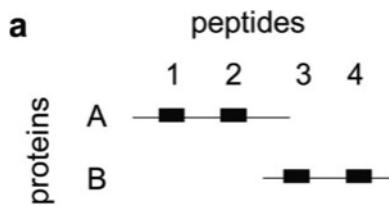
From peptides to proteins



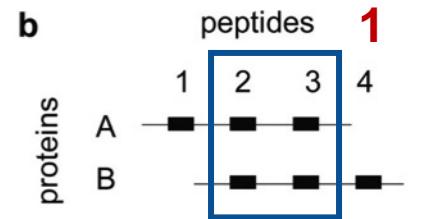
Source: fields.scripps.edu/yates/wp/

- Protein abundance = sum of peptide abundance
- **Problem 1:** How to handle shared peptides?
- **Problem 2:** Mass spectrometry cannot provide absolute abundance quantification
 - Major bias in peptides ionization

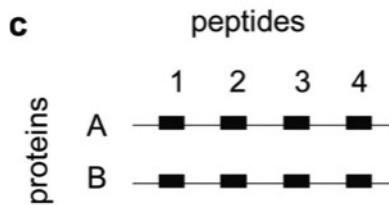
Quantification issues: Isoform



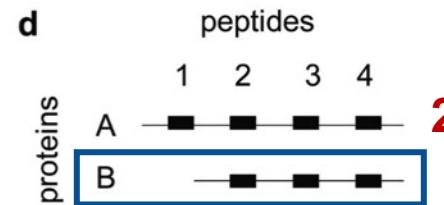
distinct proteins



differentiable proteins

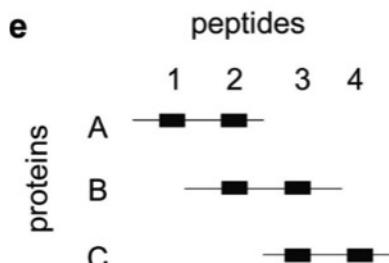


indistinguishable proteins

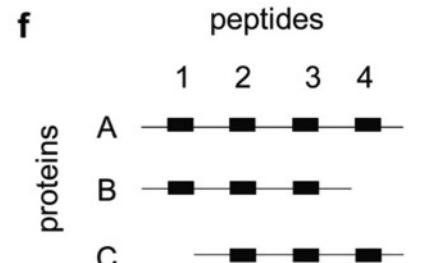


B: subset protein

3



B: subsumable protein

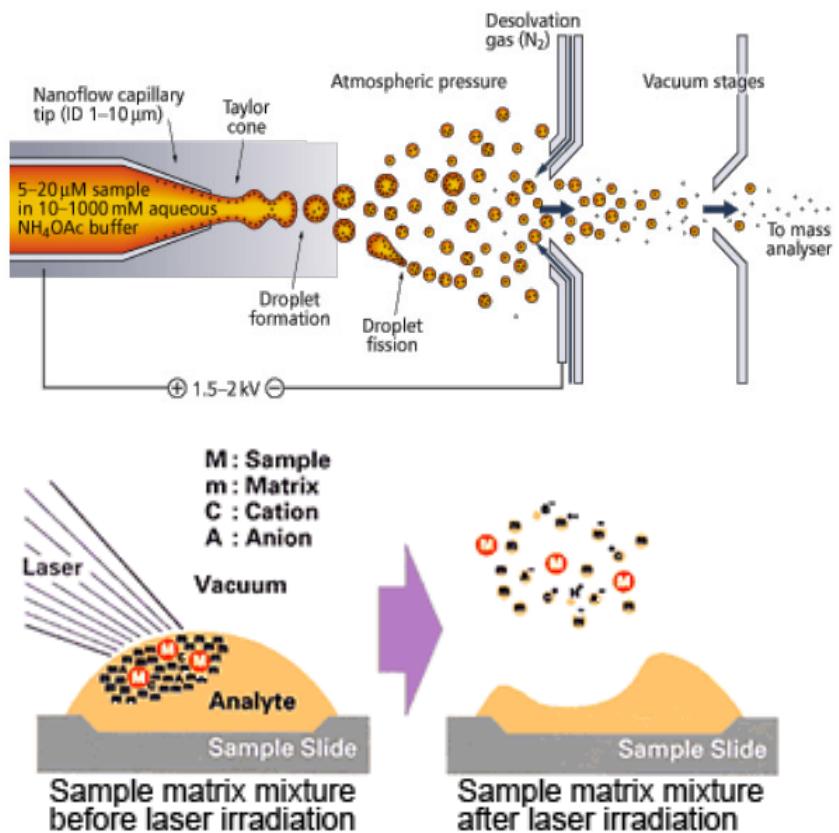


a group of proteins identified
by shared peptides only

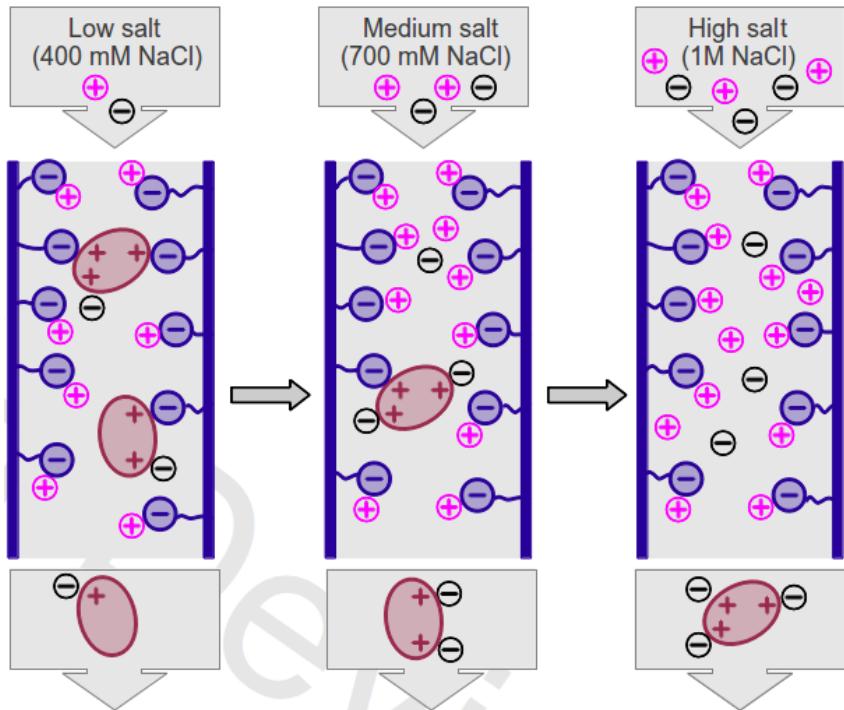
Protein quantification

- Abundance units
 - **Spectral count**
 - **Intensity** (~number of ion molecules)
- Protein quantitation
 - Use all peptides assigned to each protein
 - Use only top-N most abundant peptides
- Occam's razor principle (MaxQuant)
 - If a peptide is shared between protein A and B, **assign it to the one with higher abundance**

Quantification issues: MS pipeline



Proteins are eluted with increasing salt (NaCl) gradient



- Different ionization efficiency for each peptide
- Different elution rate for each peptide

Targeted quantitation

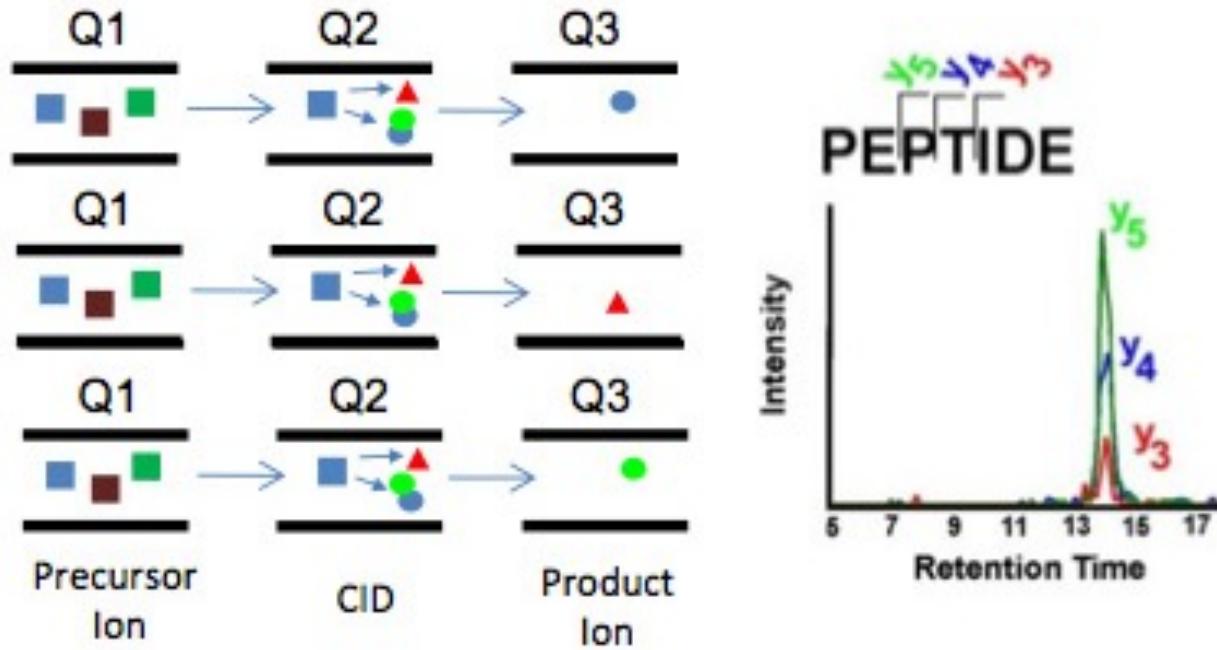


Image from <https://medicine.yale.edu/keck/proteomics>

- **Multiple reaction monitoring (MRM)** is an MS technique that isolates and quantifies specific fragment ions from peptides of interest
- Most commonly done on quadrupole mass analyzer

Comparative proteomics



- The most straightforward approach, called **label-free comparison**, relies on
 - Innate reproducibility of LC over a short period of time
 - High similarity in peptide content between samples being compared

Stable isotope labeling (SILAC)

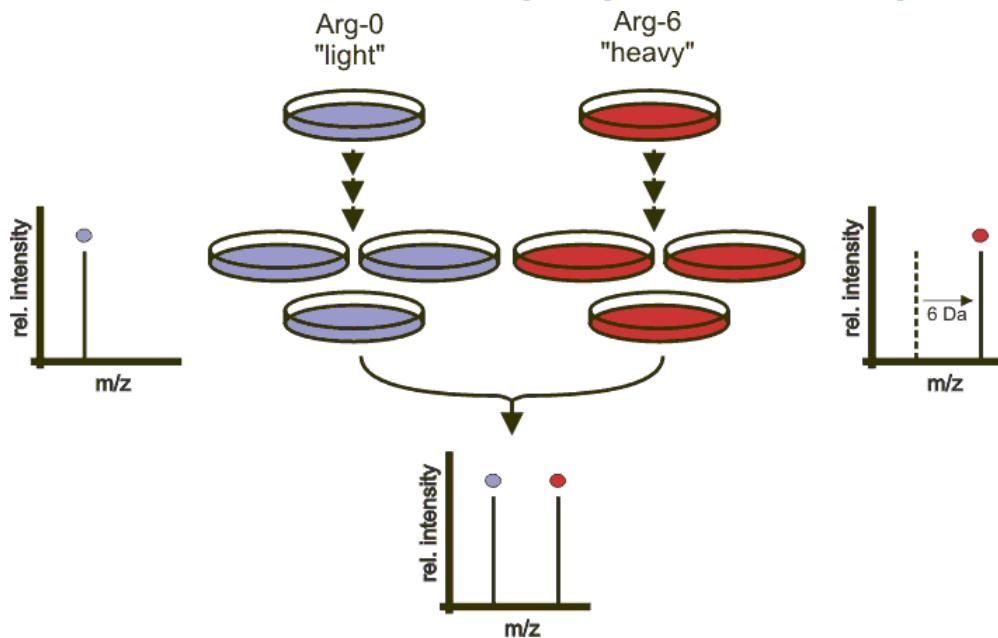
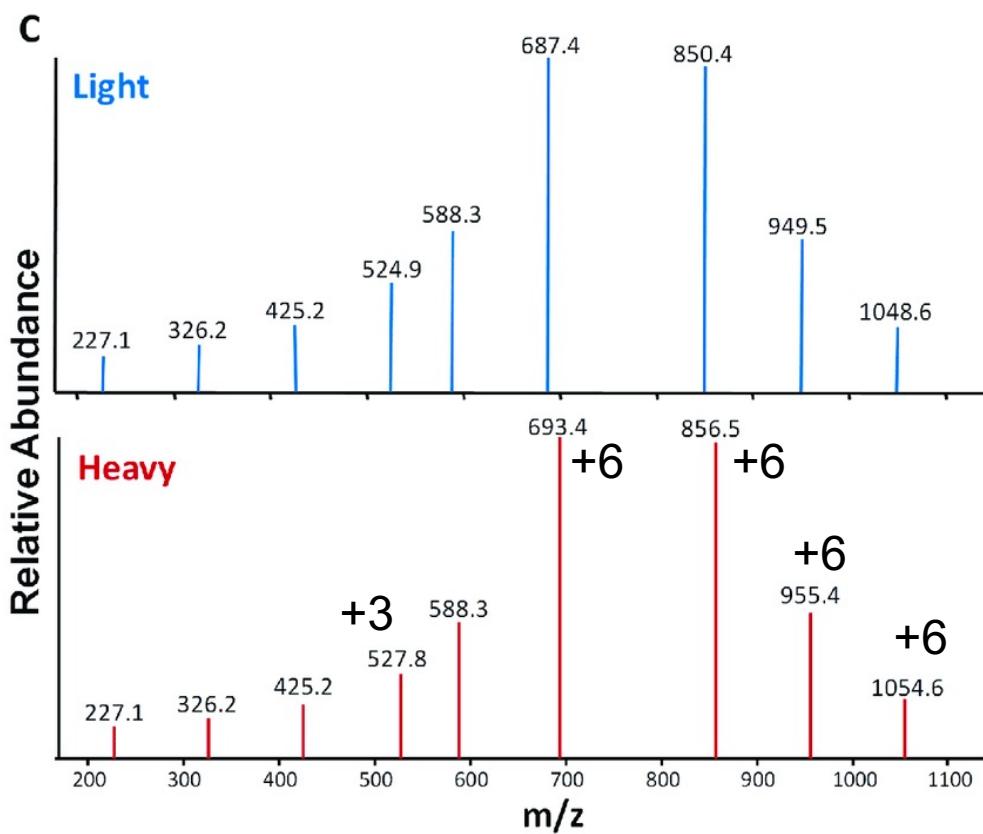
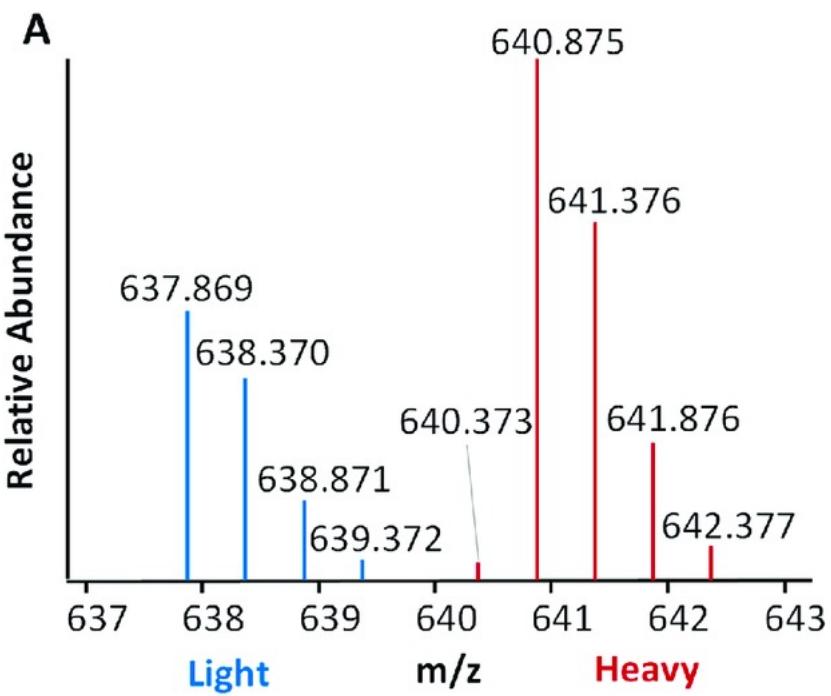


Image from https://en.wikipedia.org/wiki/Stable_isotope_labeling_by_amino_acids_in_cell_culture

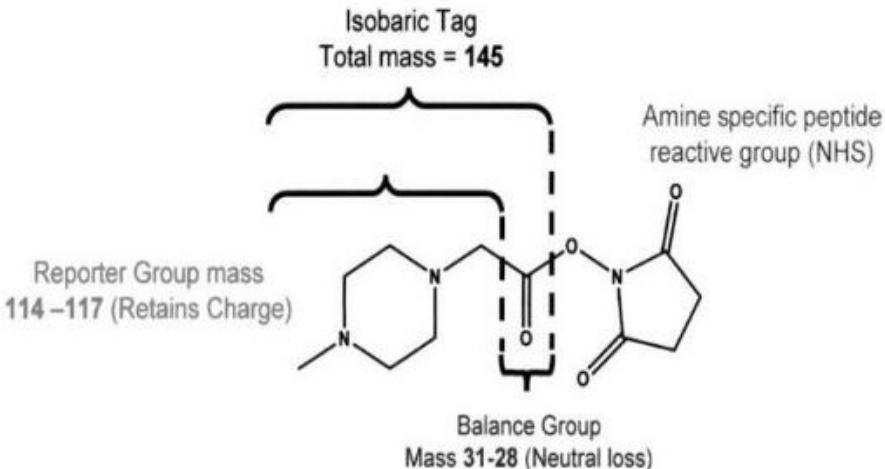
- Stable “light” and “heavy” forms of amino acids are generated with C13 or N15 and fed to each cell culture
- Samples are then combined and run together in a single LC-MS/MS. This minimizes technical variability
- Detect mass shift between light and heavy peptides

SILAC mass spectra

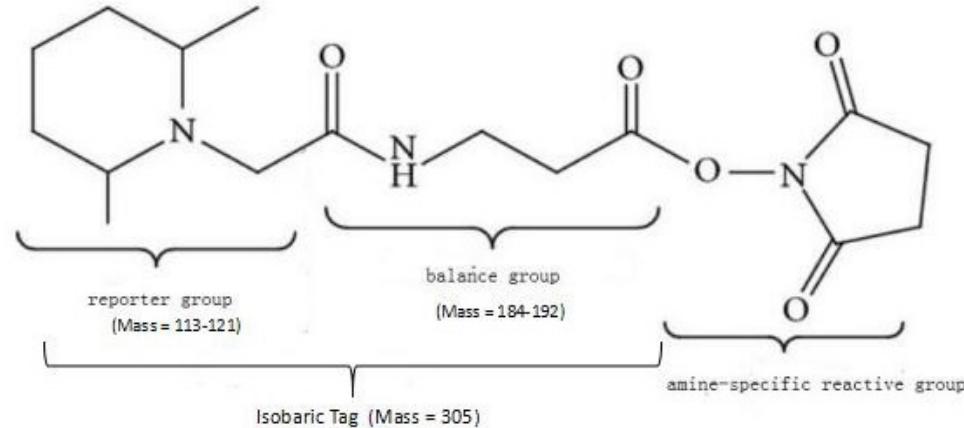
Heavy isotope = +6
Mass shift = +3 at z = 2



Isobaric and tandem mass tag



iTRAQ Tag



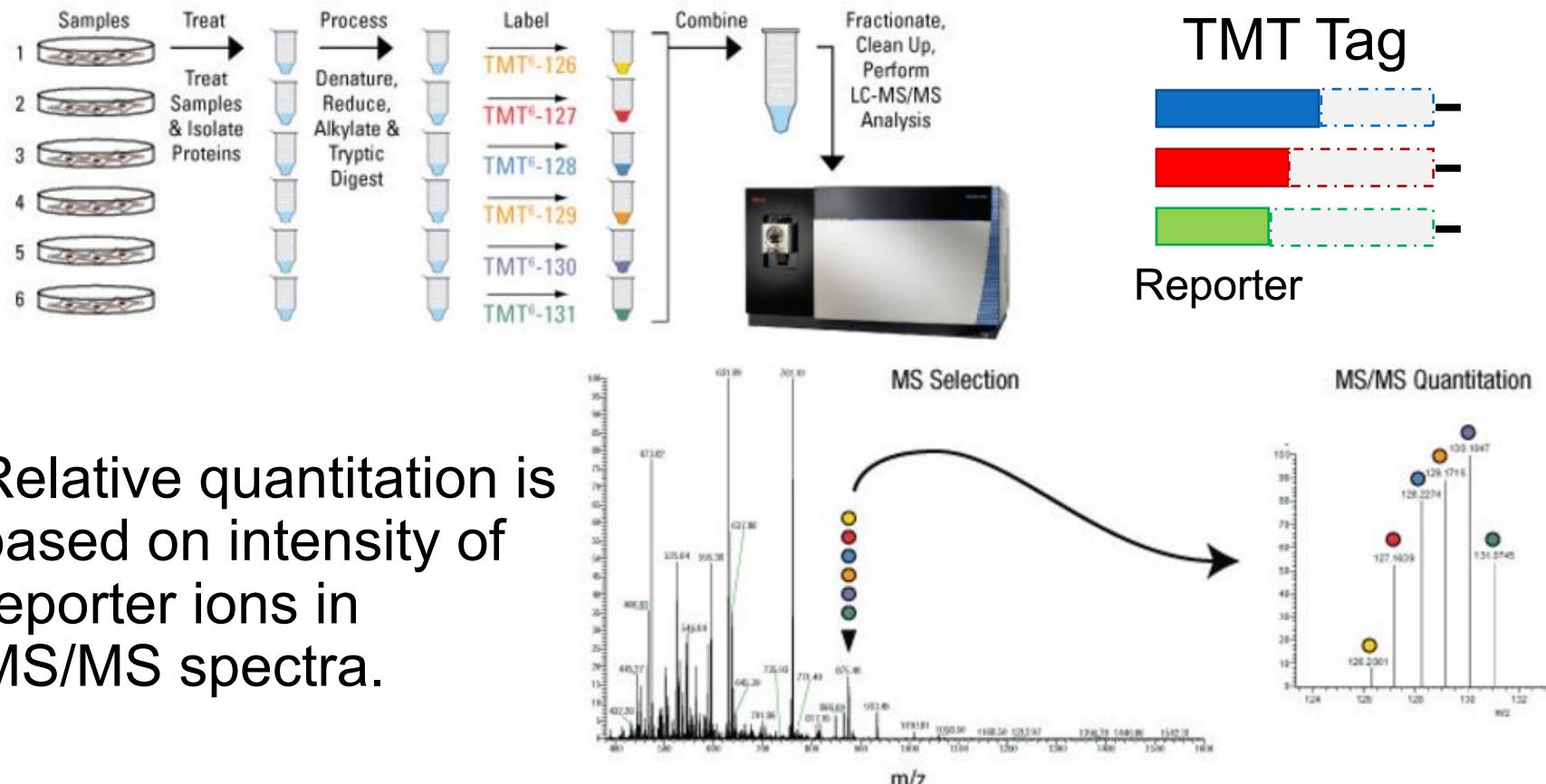
TMT Tag

Source: medium.com

- Group of isobaric chemical compound with:
 - Reactive group for peptide N-terminus
 - Reporter group that can be cleaved inside mass spectrometry
 - Different species yield report group with different masses

Tandem mass tag (TMT)

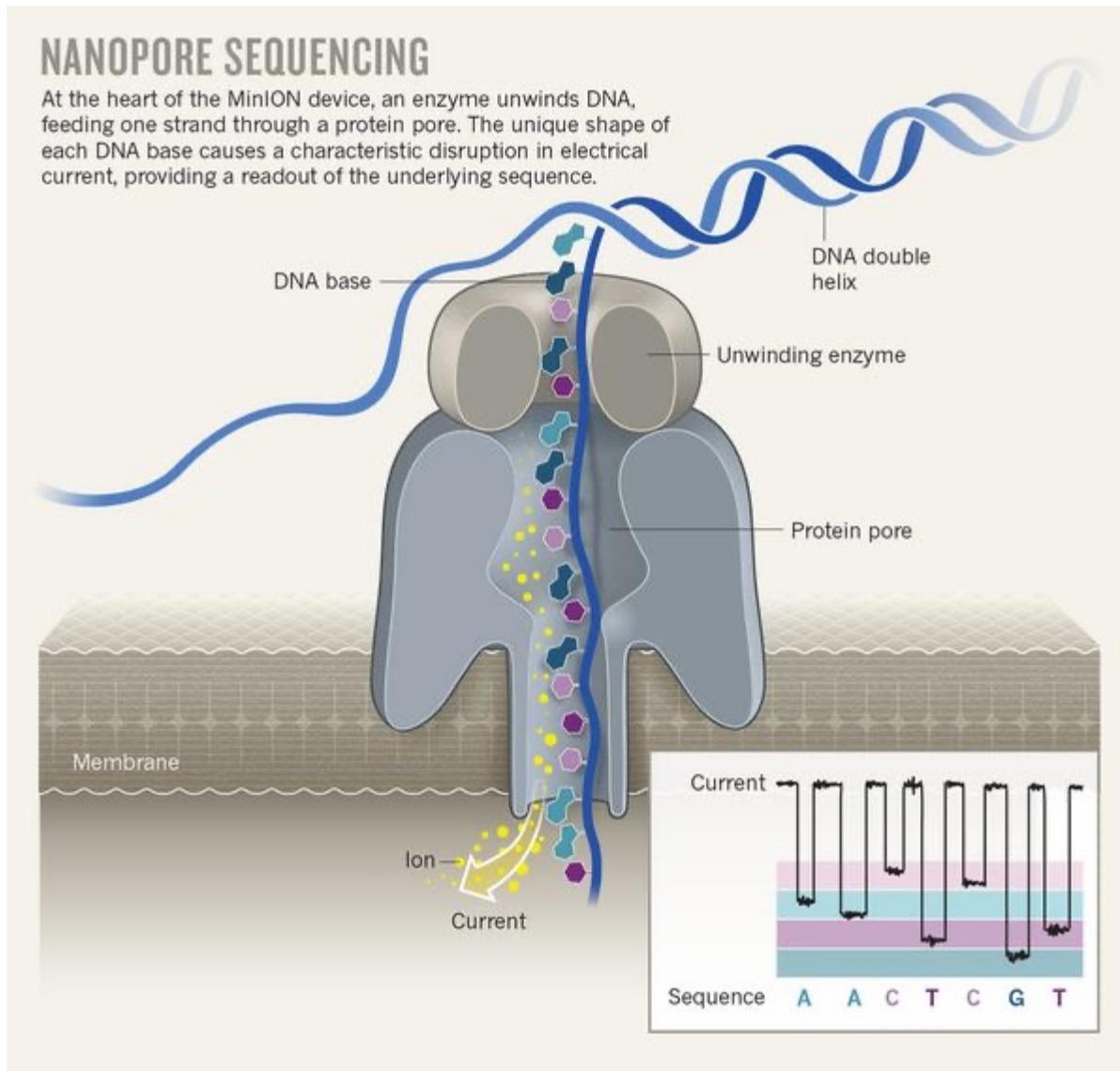
- Isobaric tags which produce distinct reporter ions
- Samples are then mixed and analyzed by a single LC-MS/MS run



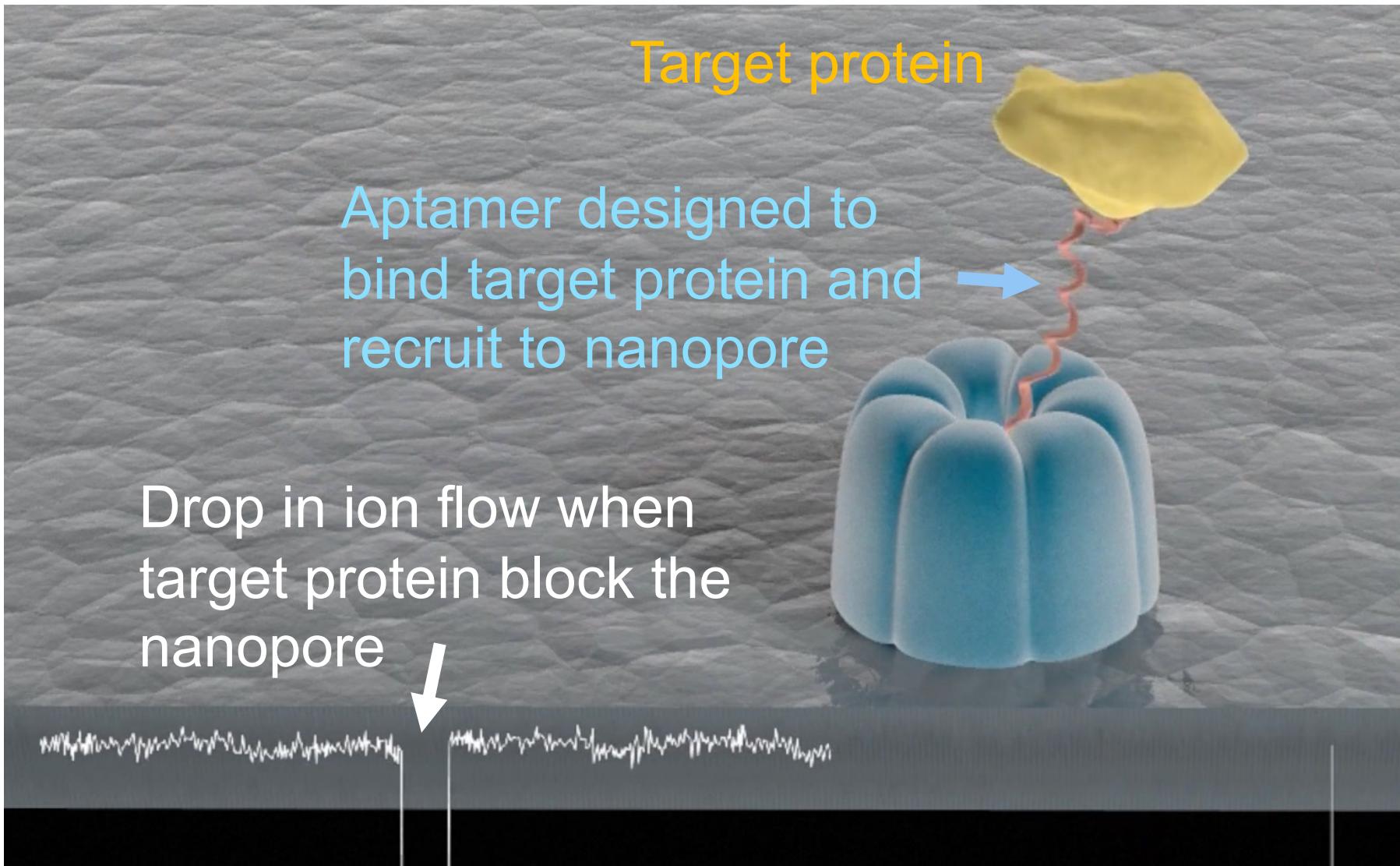
- Relative quantitation is based on intensity of reporter ions in MS/MS spectra.

A possible future of proteomics

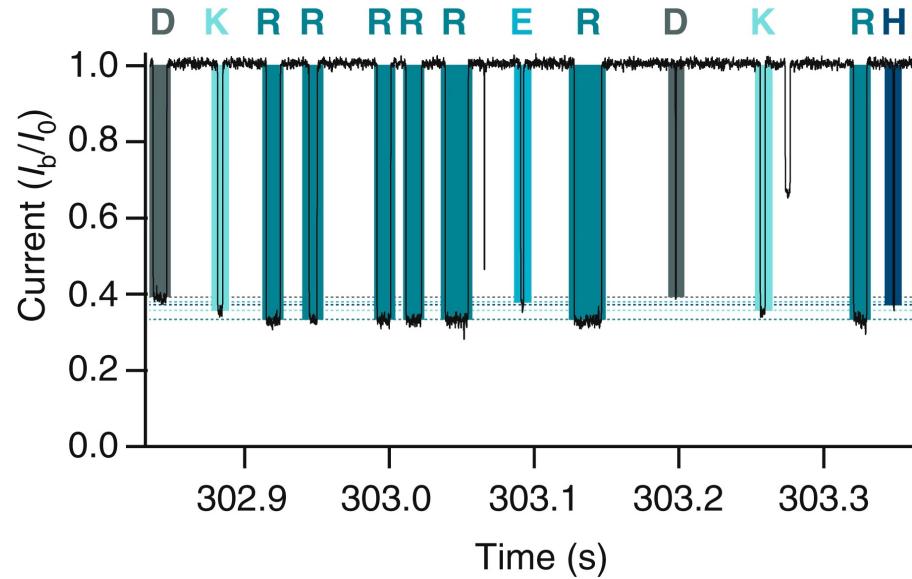
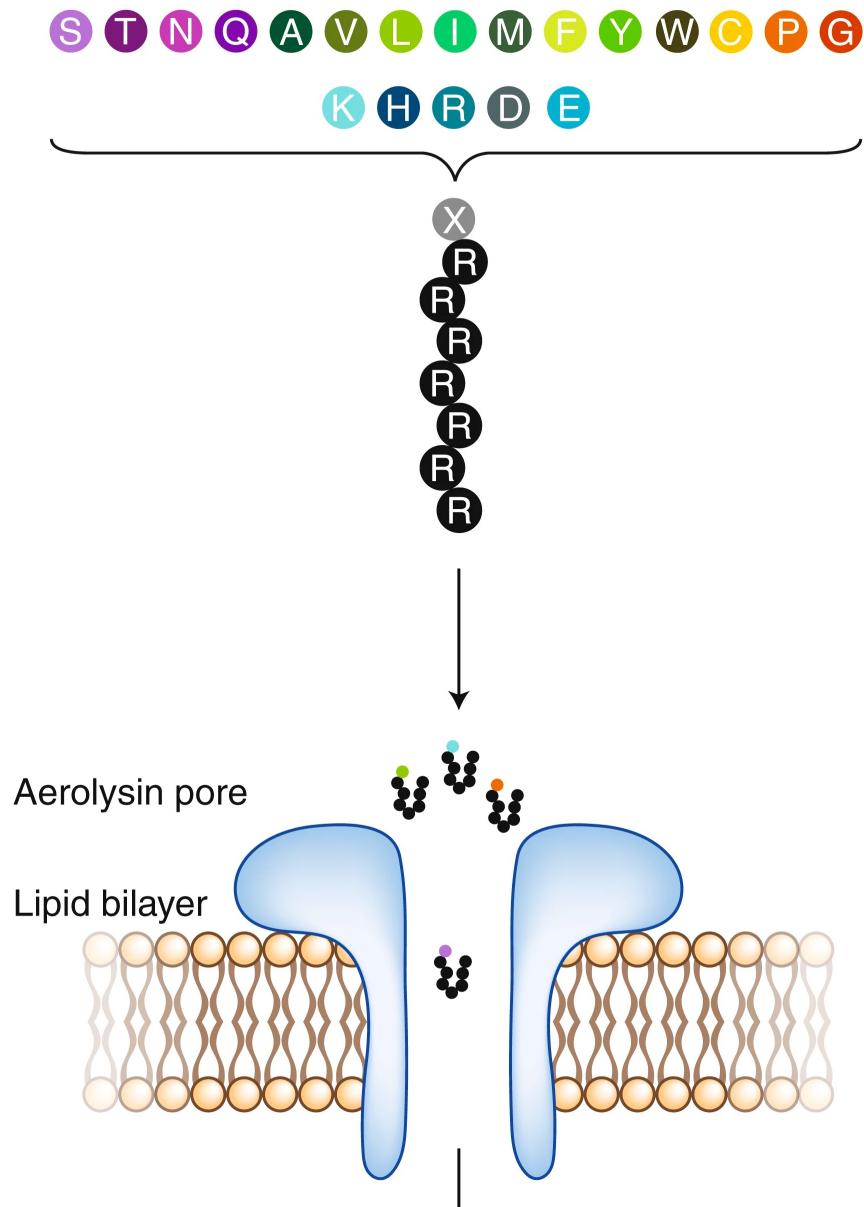
Nanopore sequencing



Protein abundance with Nanopore



Protein sequencing with Nanopore

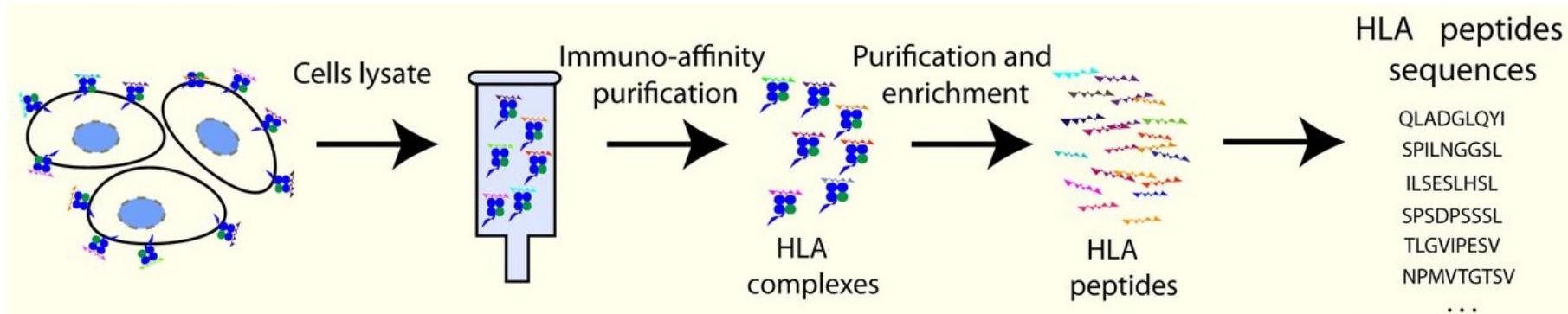


Howorka and Sivý. Nature Biotech (2020)

- Preliminary research to improve nanopore to recognize amino acid

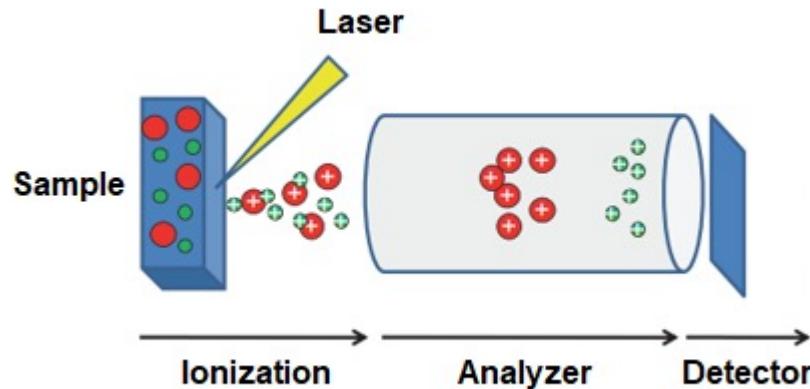
Applications

HLA peptidomics



- Human Leukocyte Antigen (HLA) is a protein in immune system responsible for presenting peptides on cell surface
 - T-cell then can recognize foreign peptides and trigger immune response
- Extract membrane-bound HLA proteins (immuno-purification)
- Extract HLA-bound peptides
- Analyze with mass spectrometry

MALDI-TOF



Source: en.wikipedia.org/wiki/Matrix-assisted_laser_desorption/ionization

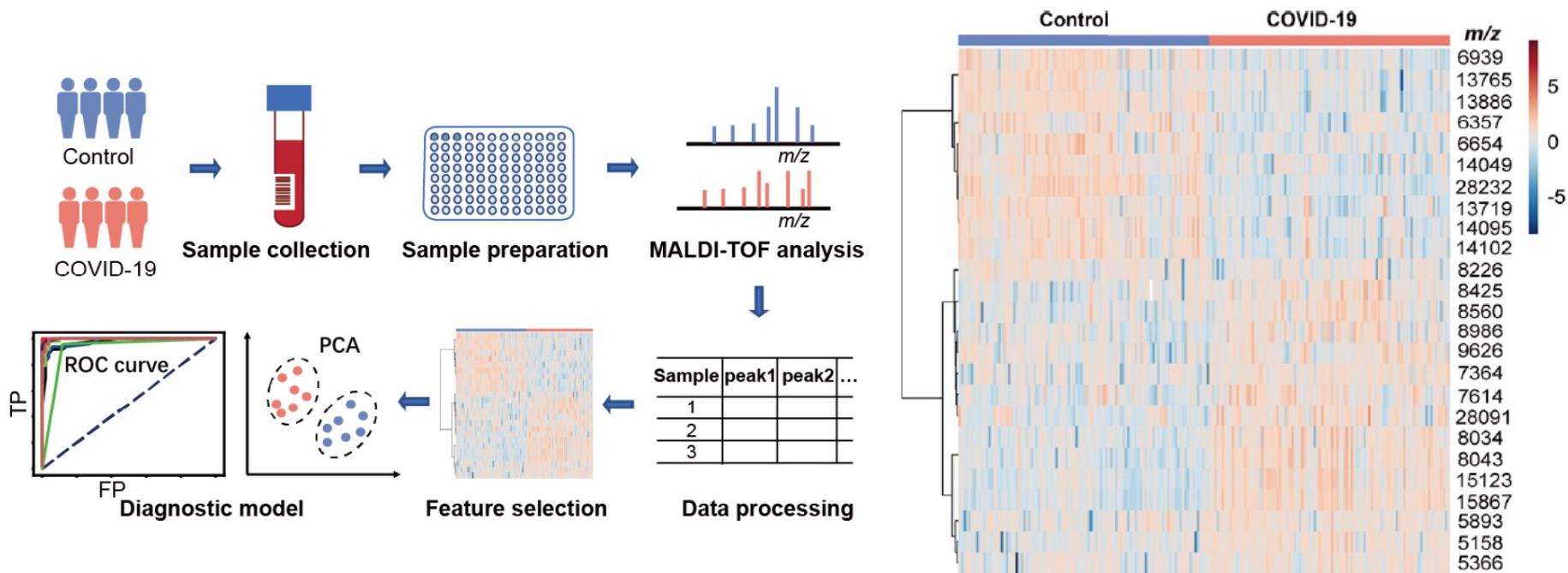


Source: www.shimadzu.com

- Multiple laser activations → multiple MS spectra
- Very fast (0.3-0.5 second per sample)
 - Compared to 2-hr to 4-hr LC time per sample
 - Good for clinical testing
- But provide only MS data (no MS/MS for peptide sequencing)
 - Performed LC-MS/MS in parallel to confirm peptide/protein identity

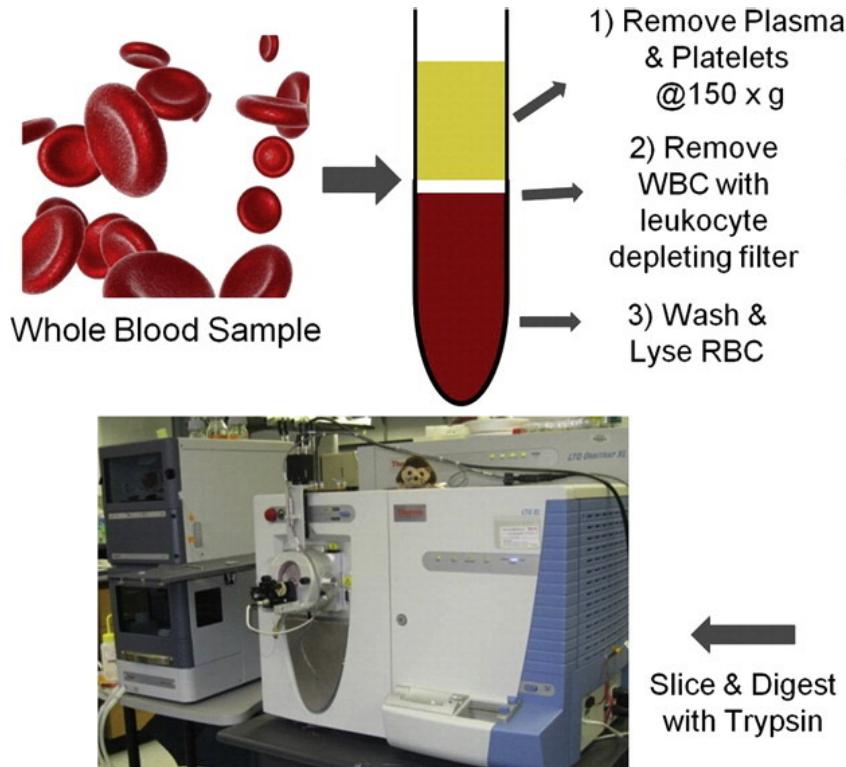
Rapid Detection of COVID-19 Using MALDI-TOF-Based Serum Peptidome Profiling

Ling Yan,[△] Jia Yi,[△] Changwu Huang,[△] Jian Zhang,[△] Shuhui Fu, Zhijie Li, Qian Lyu, Yuan Xu, Kun Wang, Huan Yang, Qingwei Ma, Xiaoping Cui, Liang Qiao,* Wei Sun,* and Pu Liao*



- AUC > 0.99, but can these m/z be mapped to proteins that are relevant to the context of COVID-19 infection?

Blood biomarker identification



- Blood samples from patients and healthy individuals
- Identify differentially expressed proteins
- Validate in new cohort + functional knowledge

Structural proteomics

Hydrogen-Deuterium exchange

HDX-MS: Conformational Changes

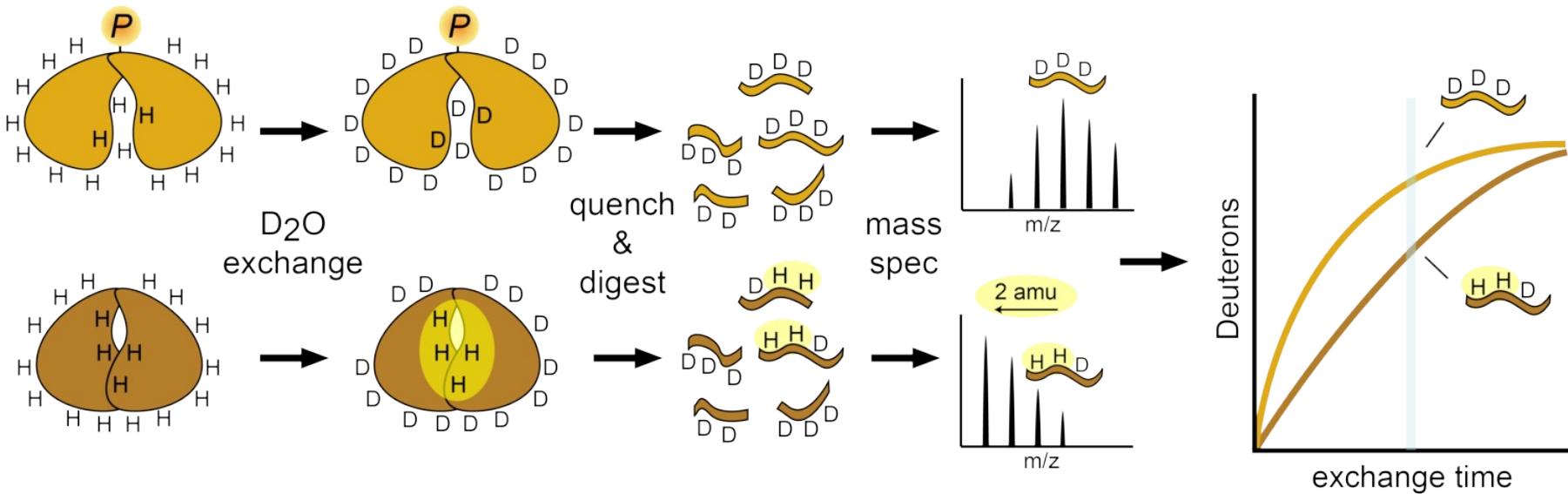
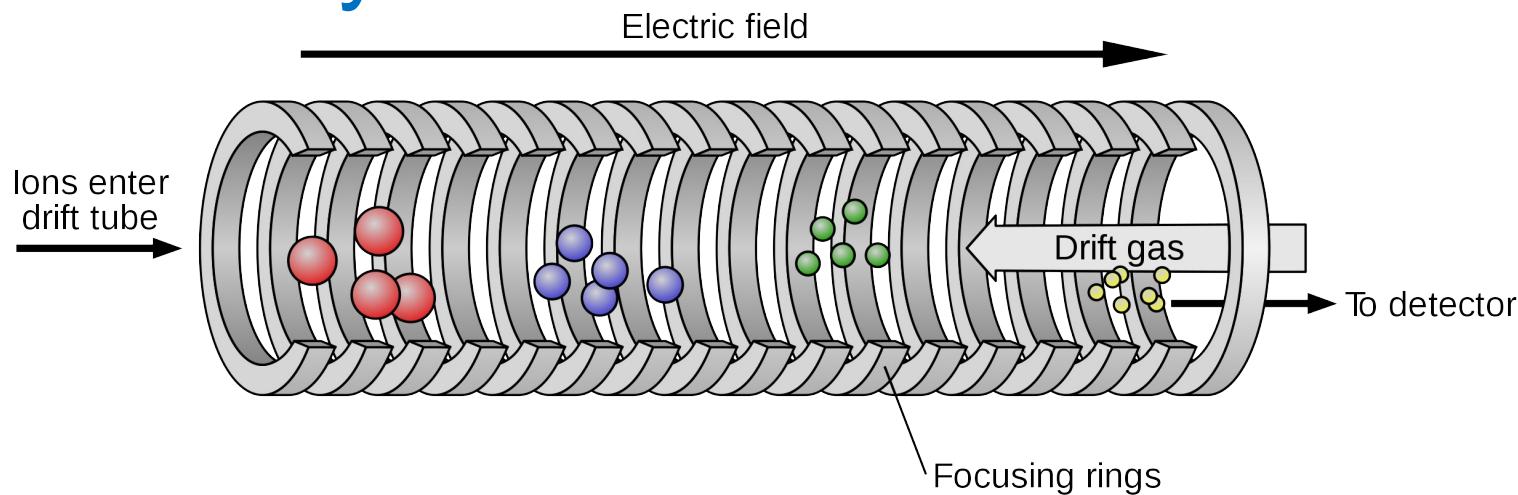


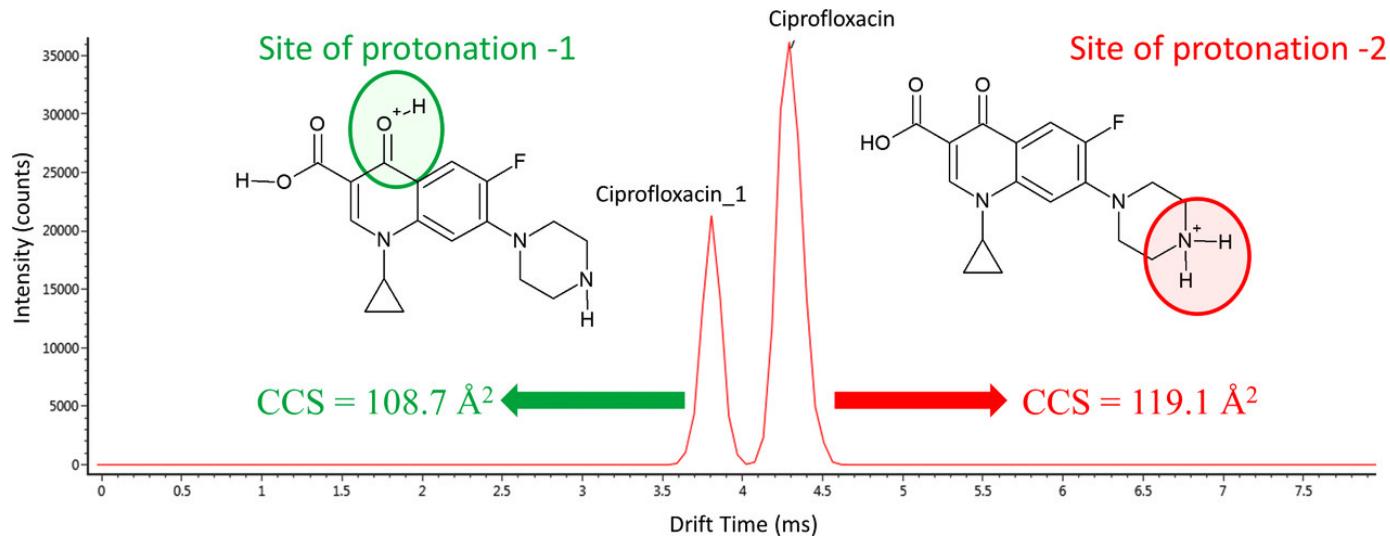
Image from
<https://www.underbakkelab.org/techniques>

- Hydrogen-Deuterium exchange MS (HDX-MS)
- Label solvent-accessible regions of the protein
- Study protein conformational change, protein-ligand binding

Ion mobility



Source: en.wikipedia.org/wiki/Ion-mobility_spectrometry



- Change in conformation → change in collision cross section (CCS) → change in drift time

Chemical crosslinking (CX-MS)

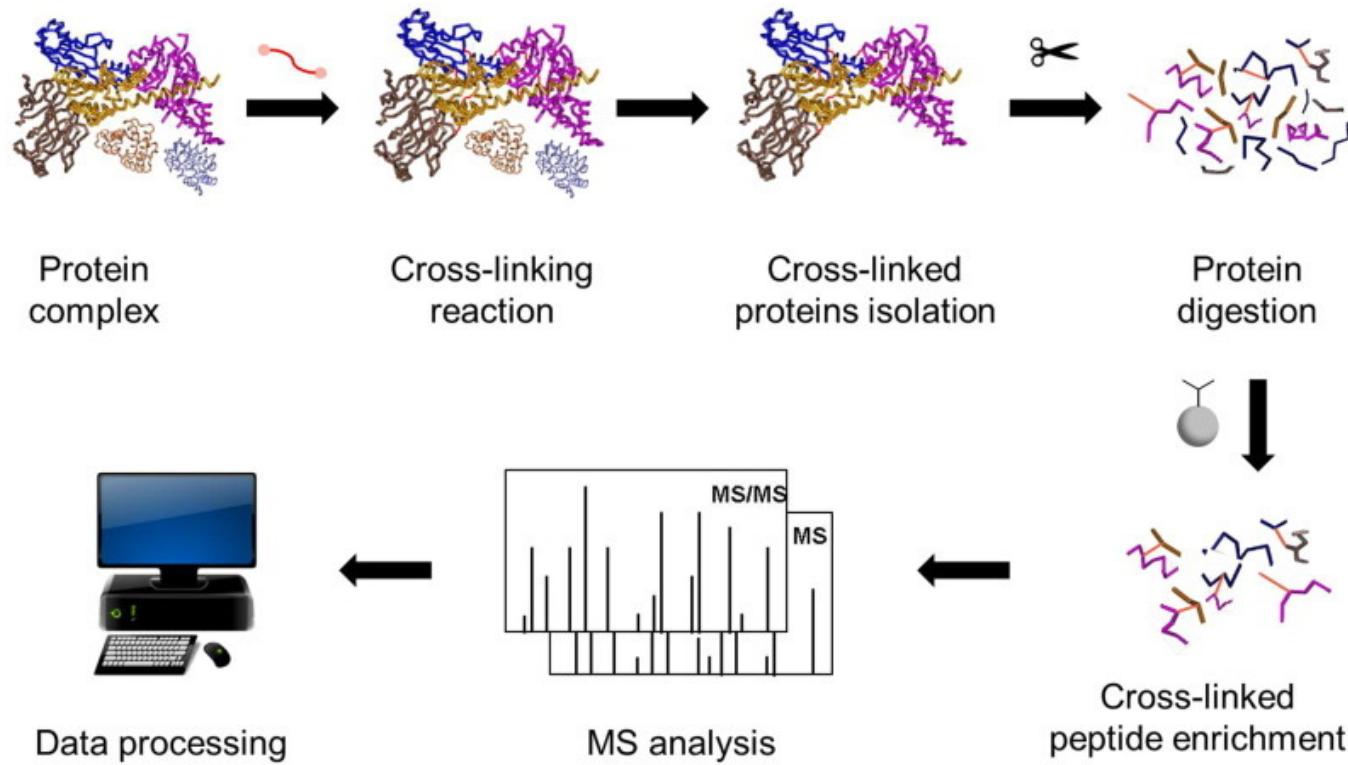
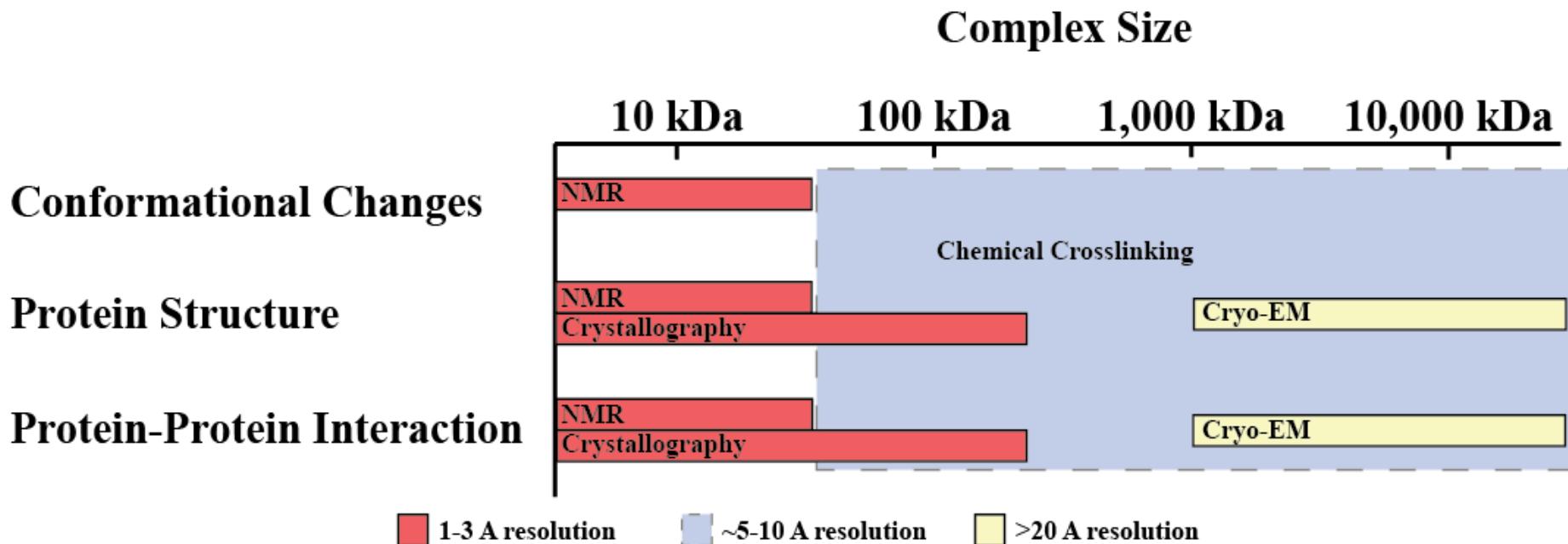


Image from Tran *et al.* BBA Proteins and Proteomics (2016)

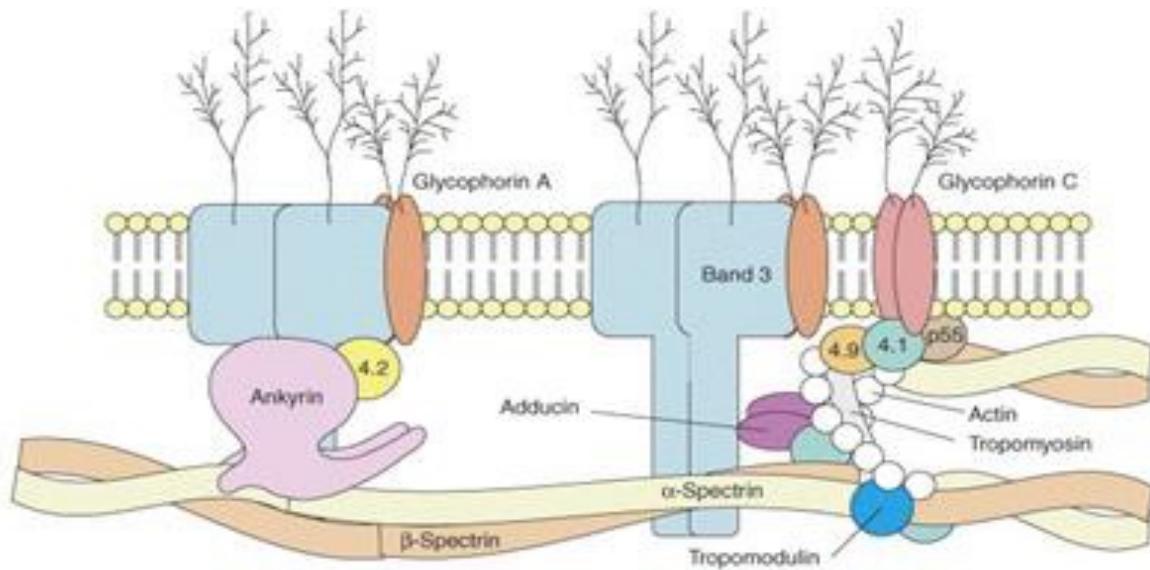
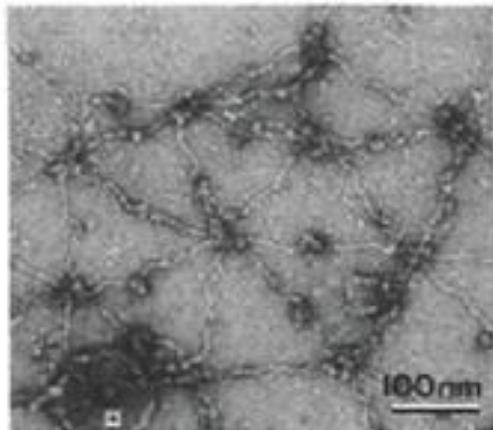
- Induce covalent bonds between proximal amino acids
- Use as distance constraints in *in silico* modeling
- Identify interacting protein complex subunits

CX-MS complements other techniques

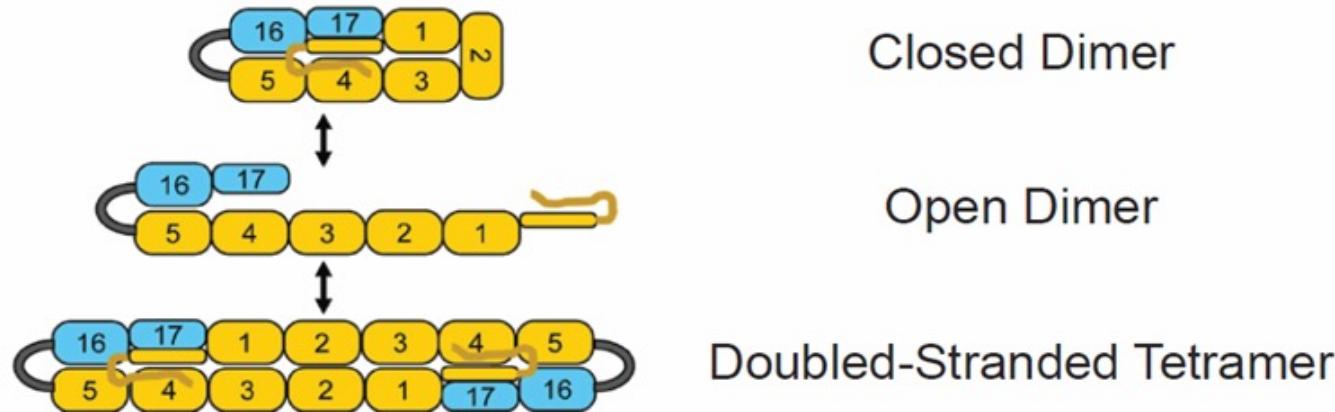


- Can be applied to protein complex of any size
 - Limitation is computation analysis of MS data
- Work on flexible and disordered proteins
- **Caution:** Too much crosslinking can disrupt natural structure

Spectrin



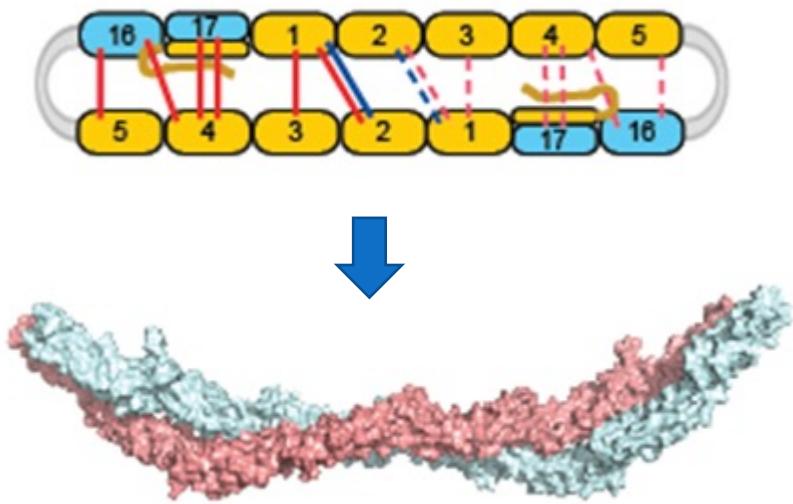
Liem and Gallagher. Drug Discovery Today, 2005 (2)



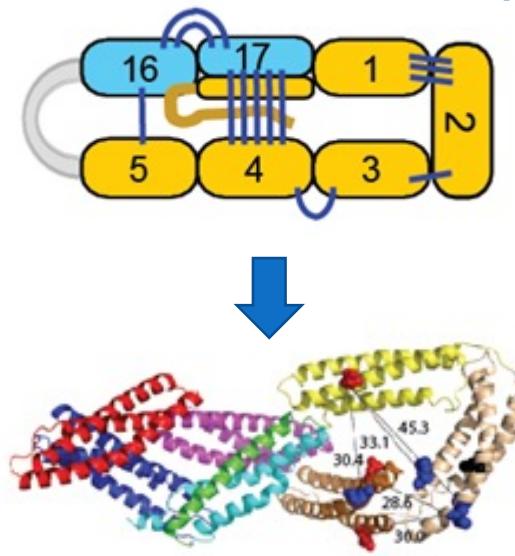
- Spectrin switches between dimer and tetramer to regulate the stiffness of red blood cell's surface

CX-MS for spectrin study

CX-MS of tetramer complex

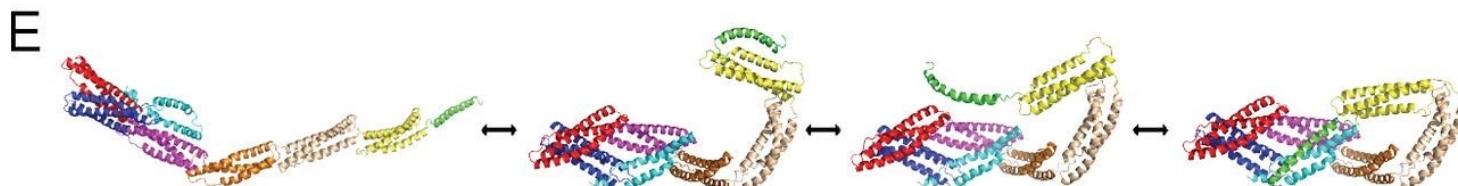
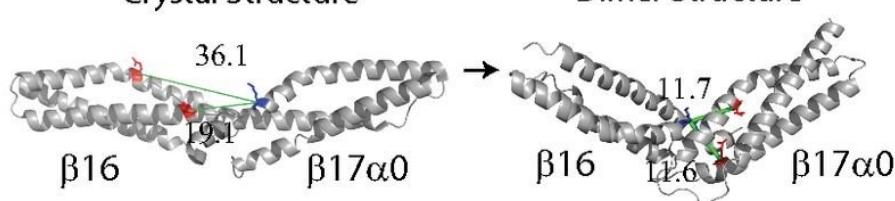
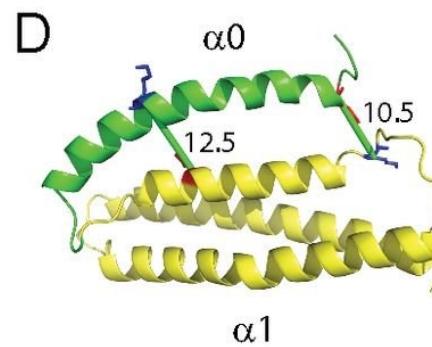
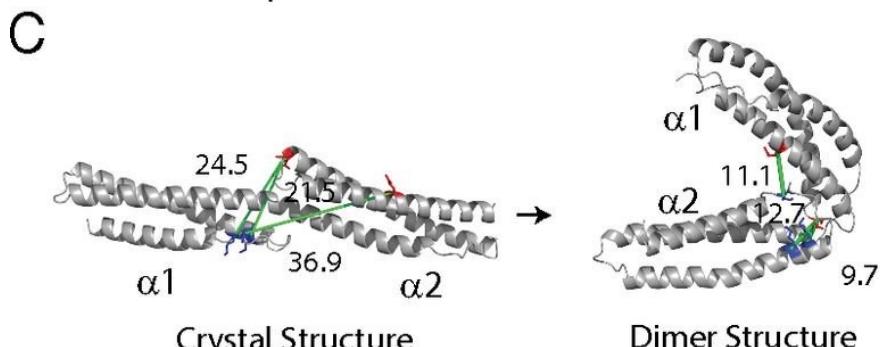
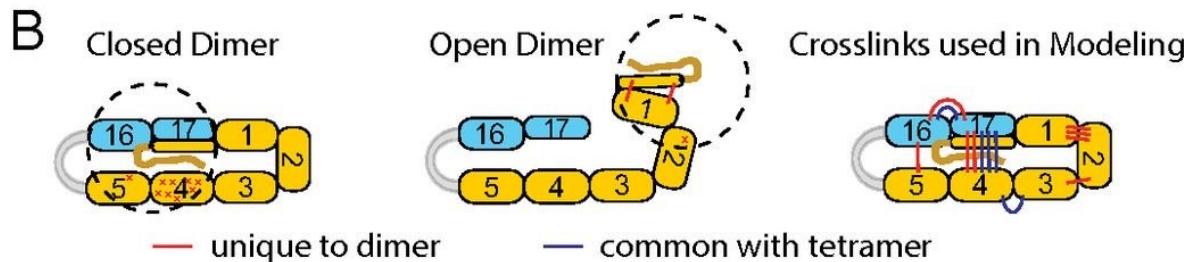
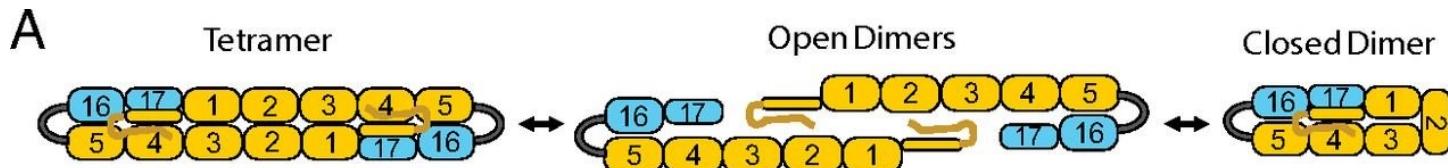


CX-MS of dimer complex

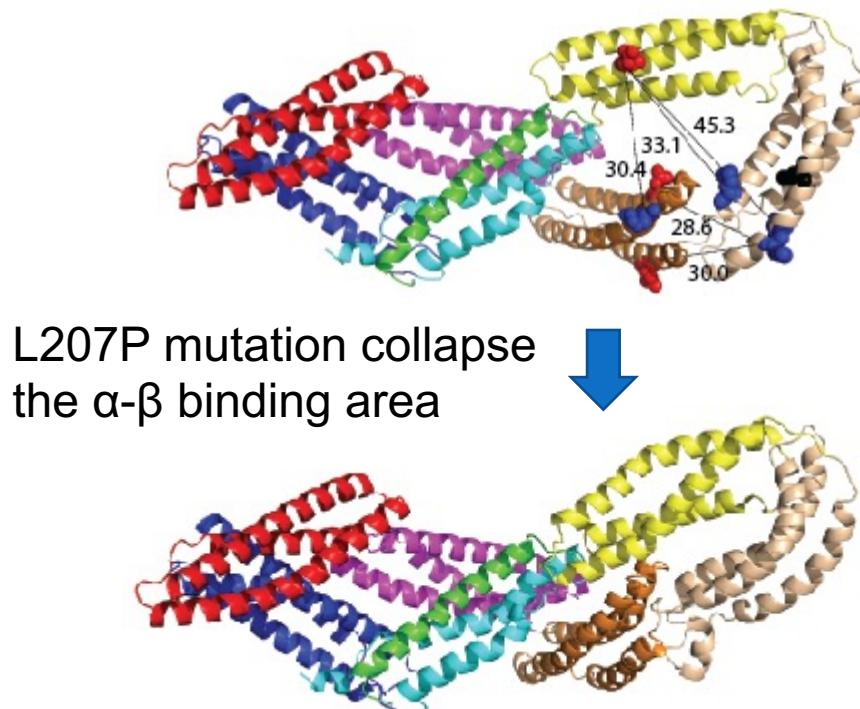
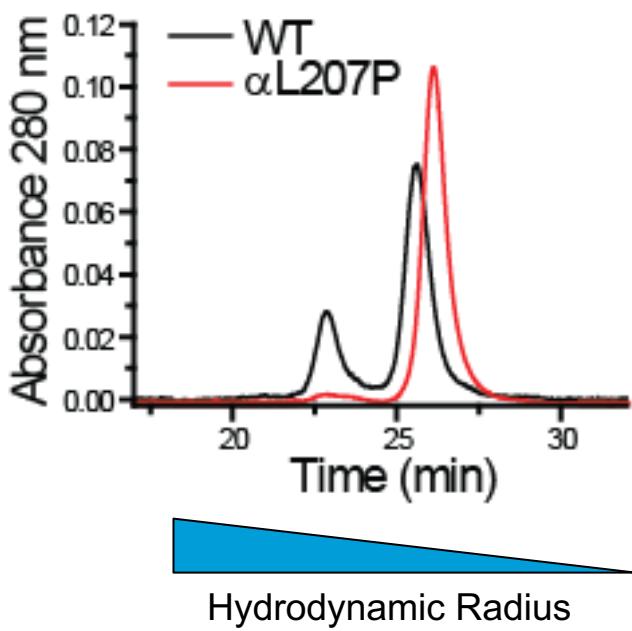
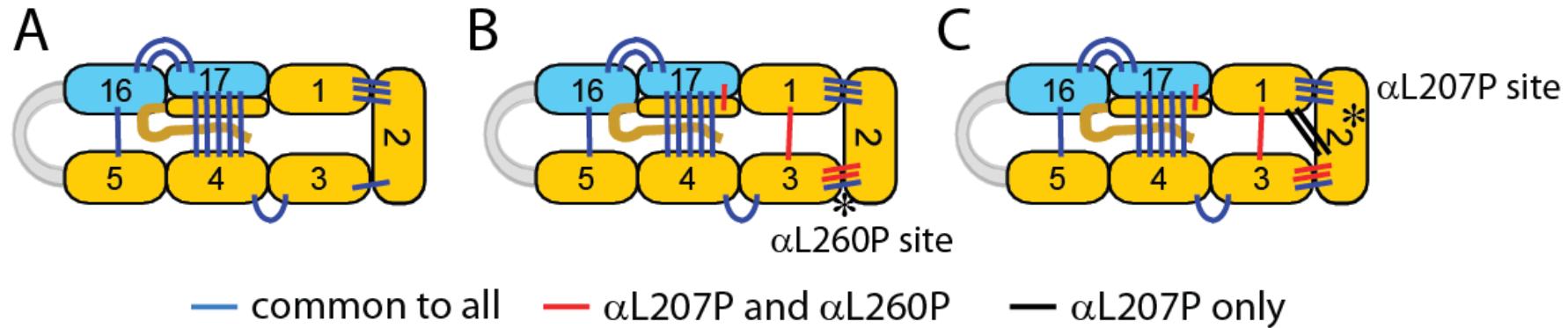


- Distance constraints from chemical crosslinking data combined with 3D structure modeling
 - Crosslinked amino acid residues must be within 12-15 Angstrom

Probing conformational changes

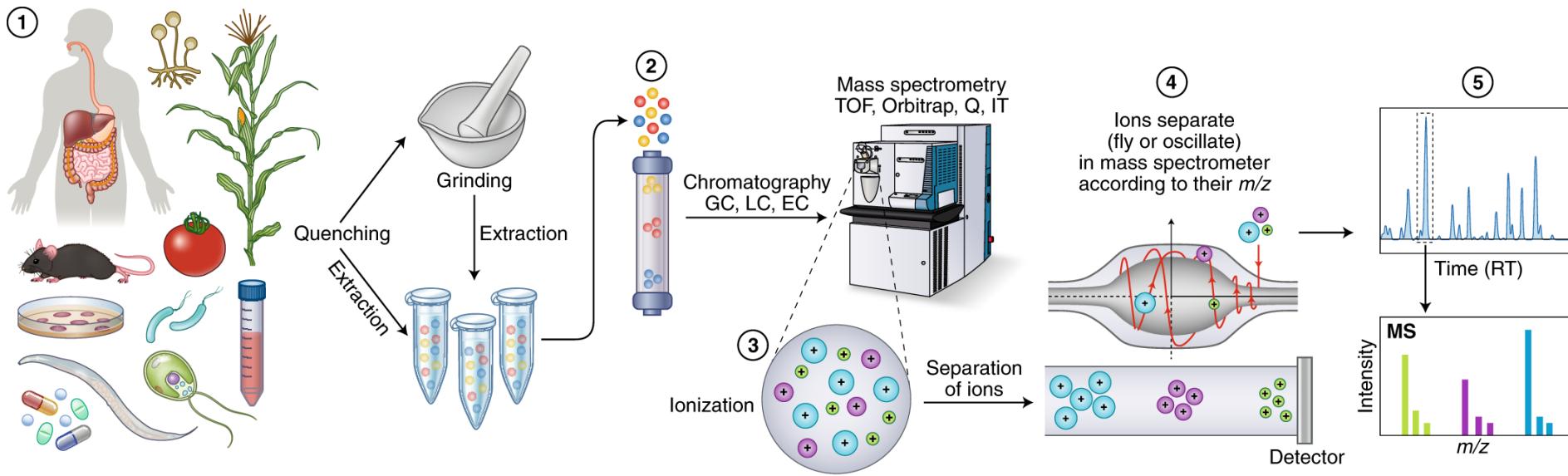


Disease causing mutation on spectrin



Beyond peptides and proteins

Metabolomics



Sample preparation and extraction

- Avoid environmental perturbation during harvesting
- Control environment: harvesting at the same time and under the same conditions
- Snap-freezing in liquid nitrogen
- Enzyme quenching: completely terminate all enzyme activities
- Standards spiked into the quenching solvent
- Grinding, isolation of cells, fast-filtration or aspiration

Sample replication and randomization

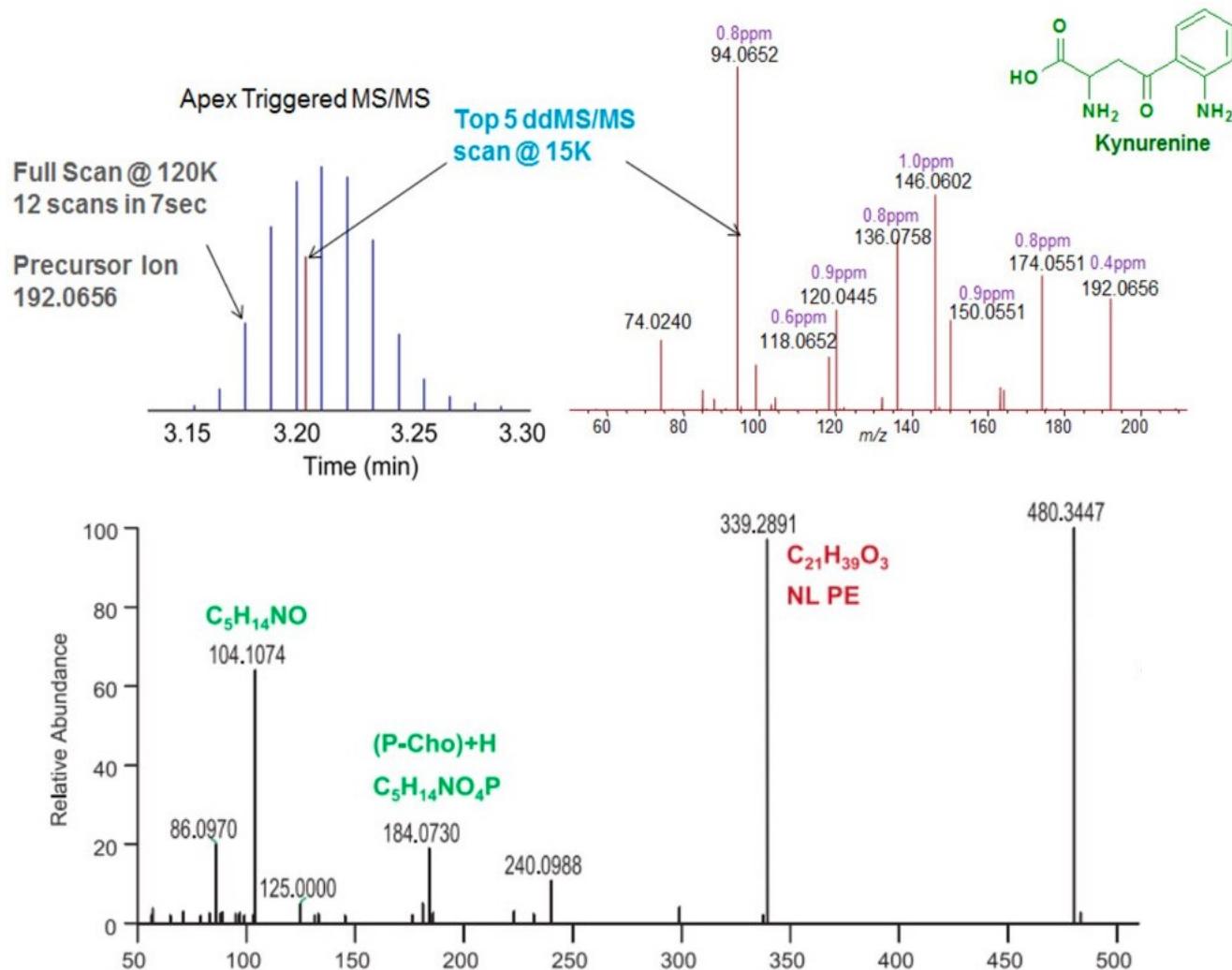
- At least four biological replicates, preferably more
- Technical and analytic replicates are worthy of consideration
- Randomization of samples throughout workflows is essential
- In large-scale studies, quality-control samples and batch correction are essential

Chromatography–mass spectrometry

- Separation methods, composition of the mobile phase, column properties and injection volume
- Metabolites are within their range of detection
- Avoid ion suppression: dilution of extracts, sonication, filtration or centrifugation, recovery test
- Choosing ionization source and type of detection mode, MS method, scan number and speed, MS/MS and energy for fragmentation

- Mass spectrometry can be used to identify molecules other than peptides and proteins
- The difficulty is how to deduce molecule identity from MS/MS

MS/MS of non-peptides



- Primarily rely on database search

Summary

- Mass spectrometry technology can be used to study the composition of biological and environmental samples
 - Proteins and peptides
 - Metabolite and chemical compounds
- Proteomics can be used to study protein-protein interaction and 3D structure, in addition to identification and quantification
- However, proteomics methods still have limitations:
 - Specialized sample preparation and instrument protocols
 - Rely on database of expected amino acid sequences
 - Need some work to obtain absolute quantification