3000788 Intro to Comp Molec Biol

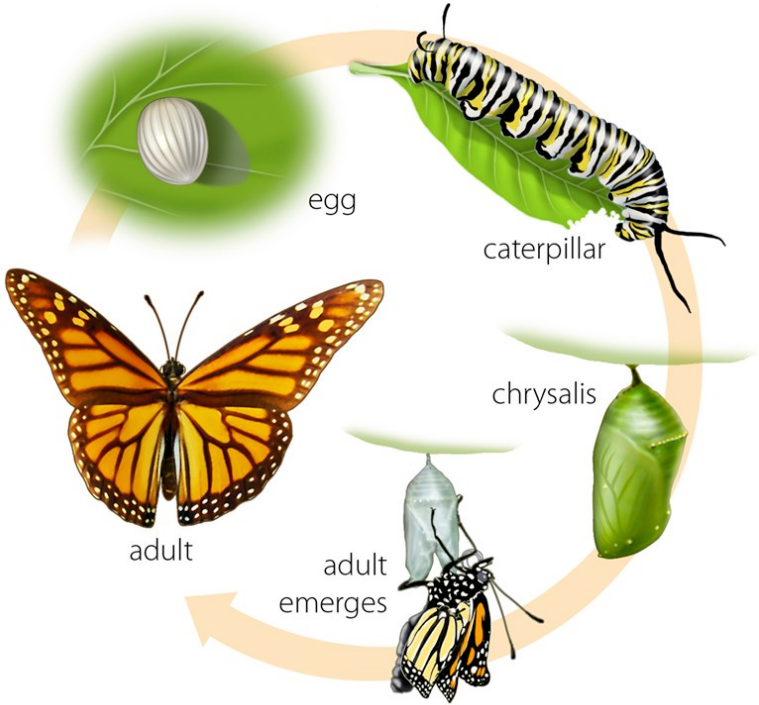
Lecture 16: Proteomics and mass spectrometry

October 9, 2023

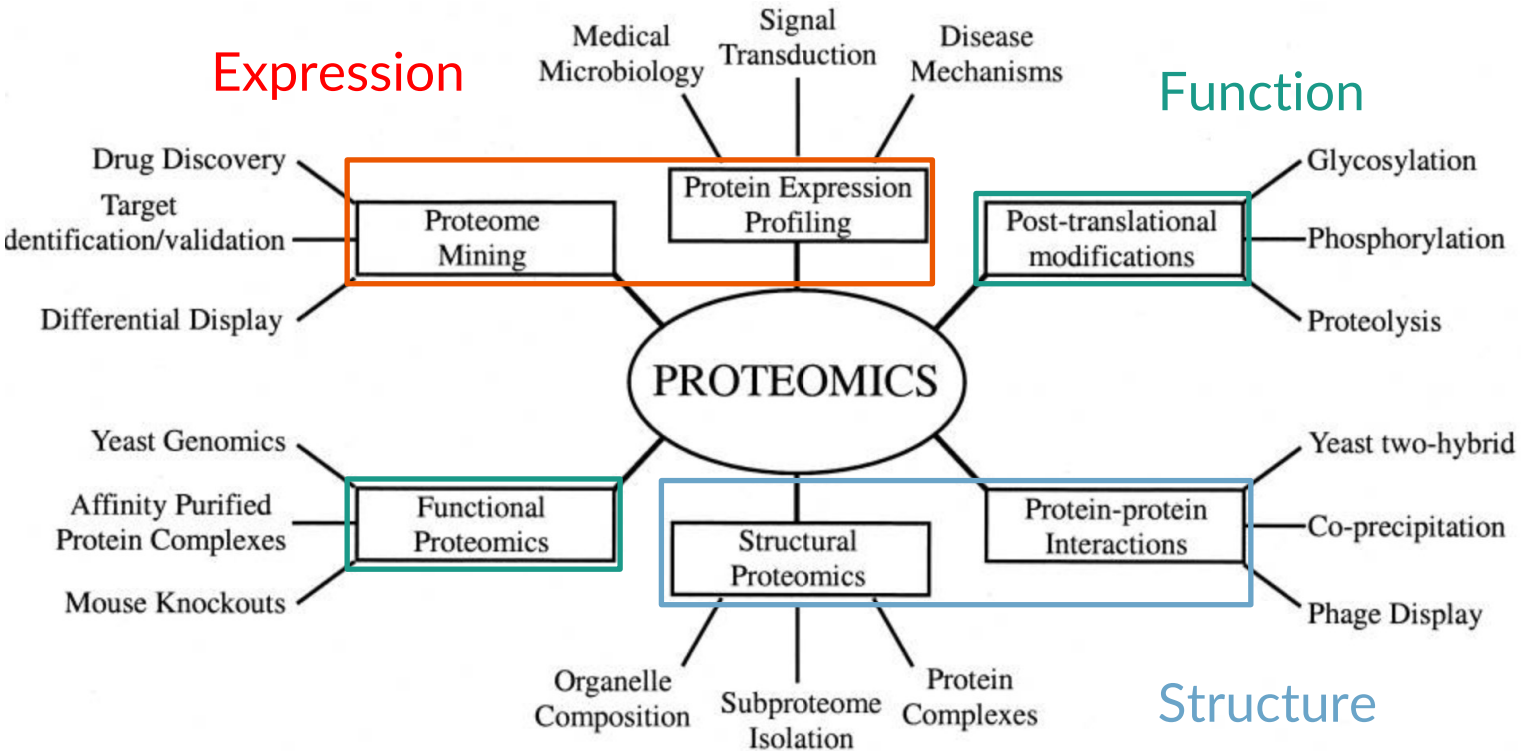
Sira Sriswasdi, PhD

- Research Affairs
- Center of Excellence in Computational Molecular Biology (CMB)
- Center for Artificial Intelligence in Medicine (CU-AIM)

Genome is static, proteome is dynamic



Focuses in proteomics



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

Why not proteomics all the time?

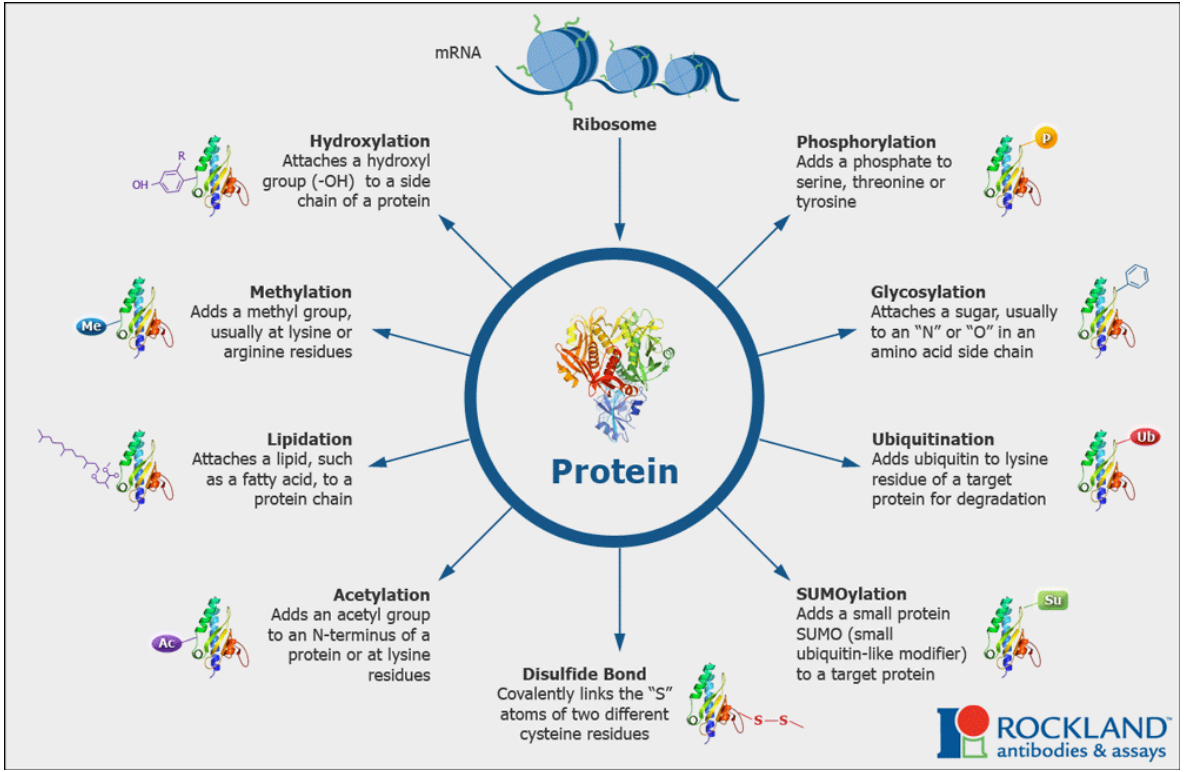
- **Amounts of materials**

- **Difficult to extract**
  - Proteins are chemically heterogeneous
  - Integrated into cellular structure

- **Difficult to identify**
  - Cannot rely on complementary pairing to read out amino acids
  - Similar issue as nanopore (but more costly to synthesize data)

- **Post translational modifications**

Post translational modification (PTM)



PTM greatly increases the variety of proteins

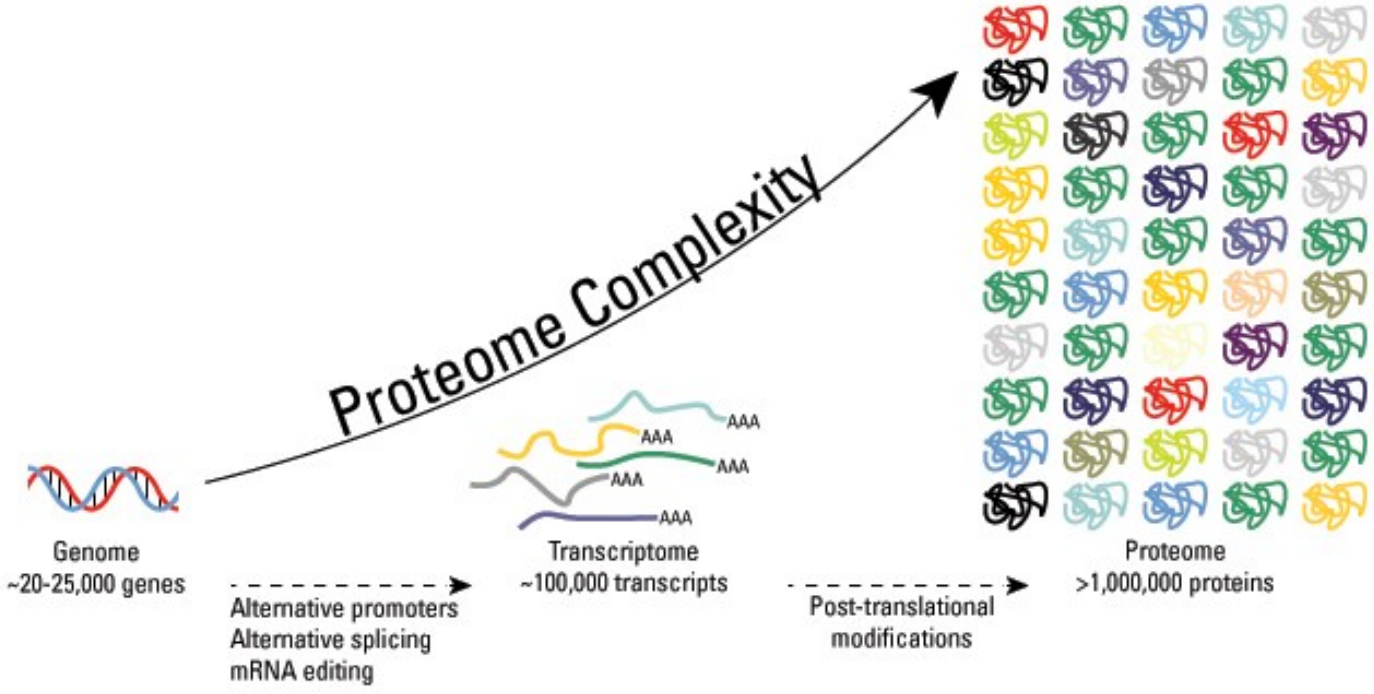


Image from https://www.thermofisher.com

Mass spectrometry (MS)

Inside a mass spectrometer

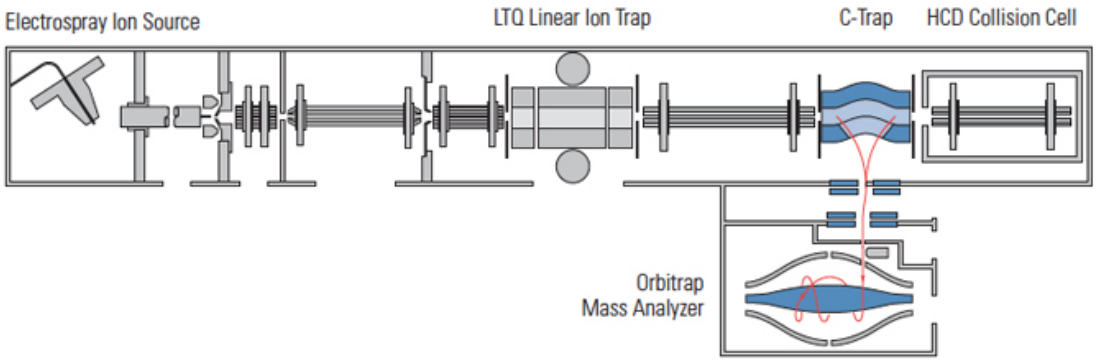
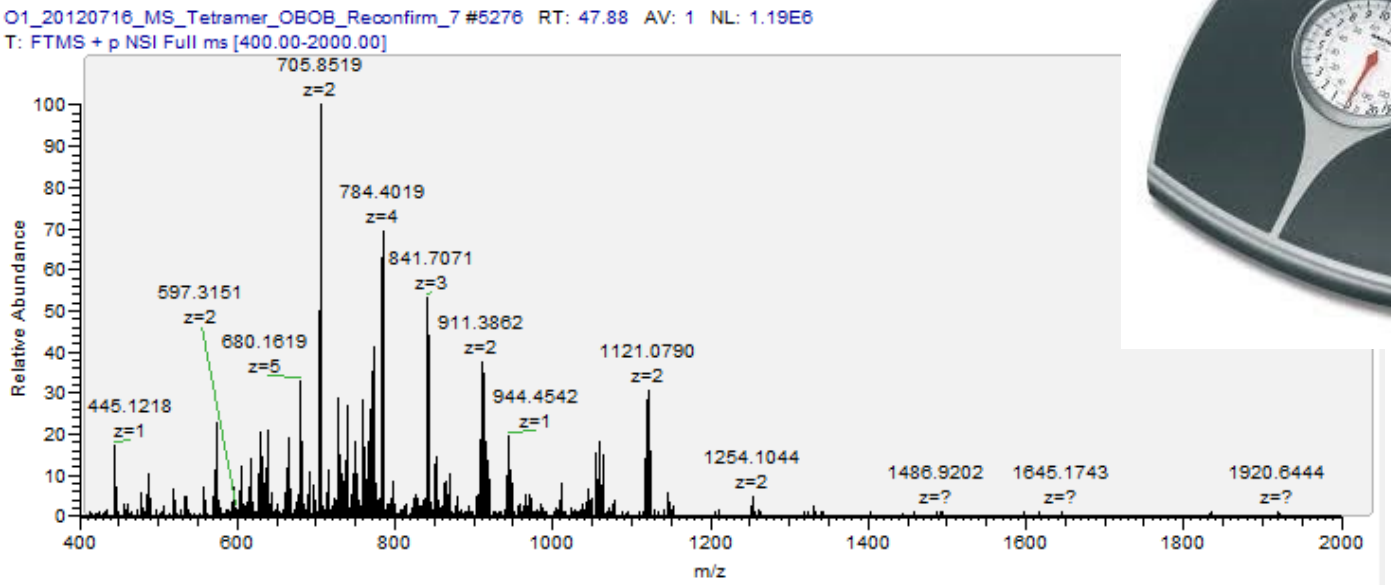


Image from http://planetorbitrap.com/

- A series of vacuum chambers for ion trapping and m/z measurement

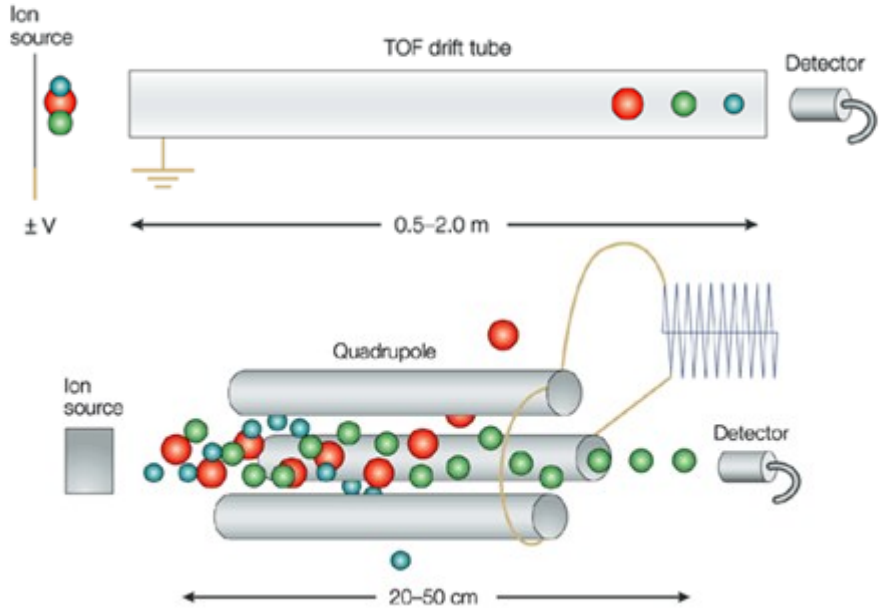
A mass spectra



Shalek and Benson. Science Trans Med. 9:eaan4730 (2017)

- Abundances + mass-to-charge ratio (m/z) of all detected ions

Physics of m/z measurement



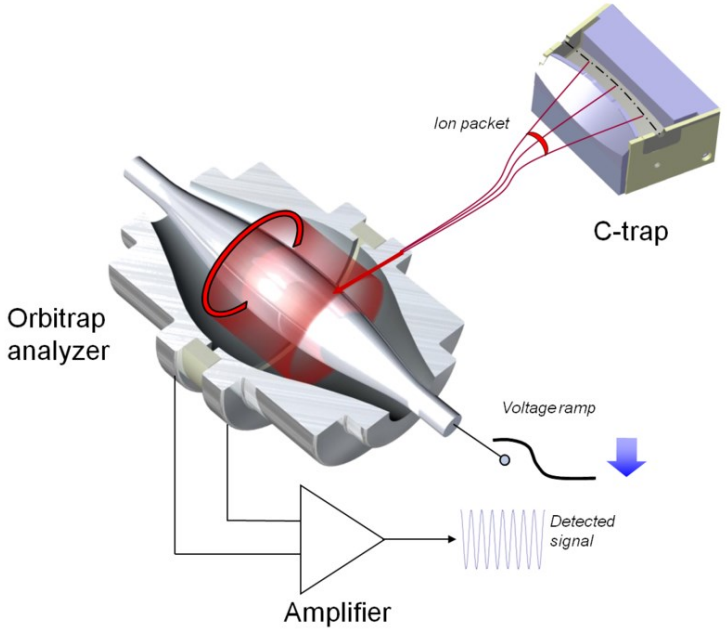


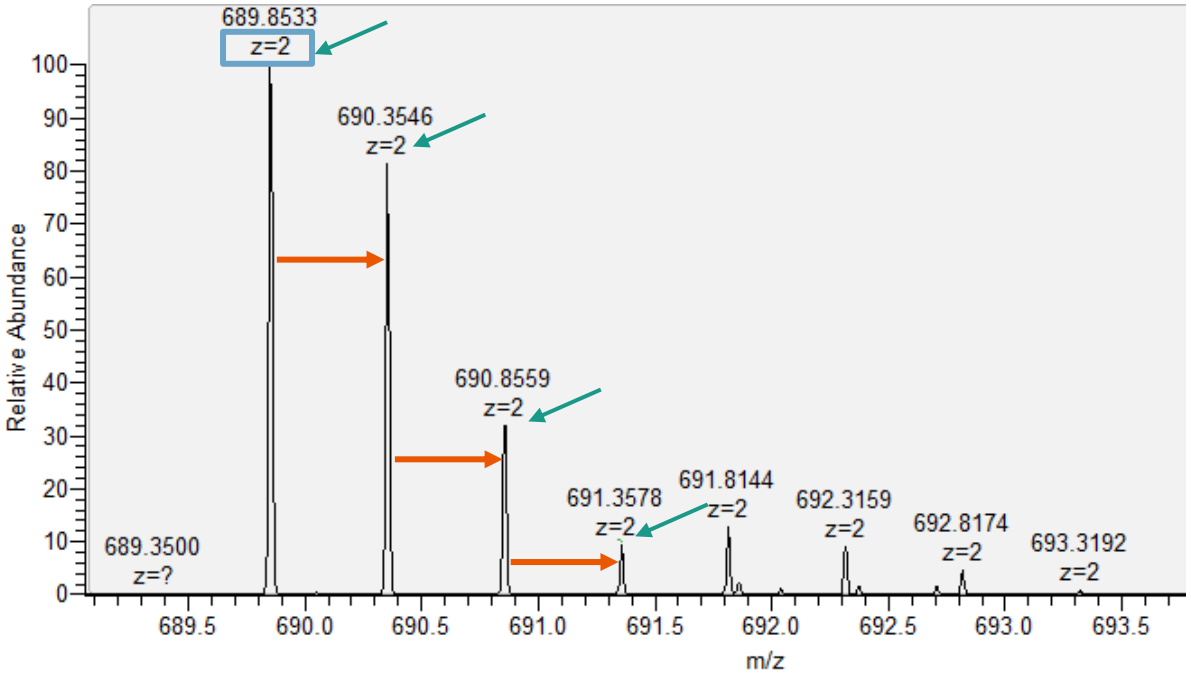
Image from https://en.wikipedia.org/wiki/Orbitrap

- Ion in electric field → time of flight
- Ion in magnetic field → orbital frequency

$zV = \frac{1}{2} mv^2$

$m/z = 2V/v^2$

Solving mass from mass-to-charge ratio



- Ions exist as isotopes
- ^{13}C, ^{14}C, ^{15}N, and ^{18}O
- Adjacent isotopes differ by a neutron
- Difference between m/z and $(m+1)/z = 1/z$

Making MS works for biology

Soft ionization techniques

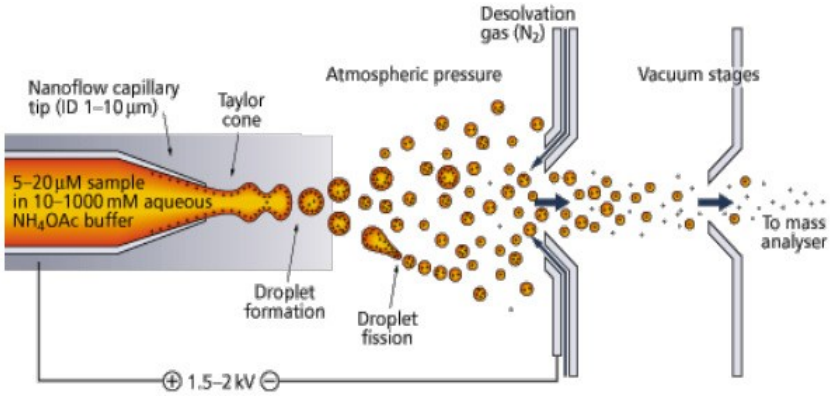


Image from https://www.thermofisher.com

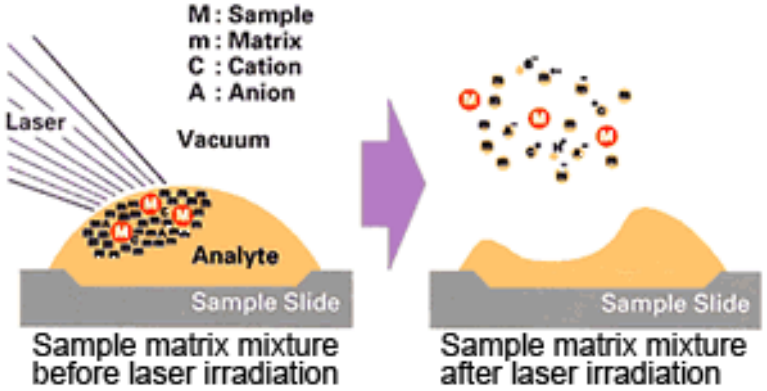


Image from https://www.shimadzu.com

- MS analyzes ions in **gas phases**
- Biomolecules break apart under regular ionization techniques
- Electrospray (ESI) and Matrix-assisted LASER desorption (MALDI)

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"*.

LC-MS/MS (tandem MS) analysis of proteins

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

- Proteins in aqueous phase
- Passed through liquid chromatograph (LC)
- Ionized into gas phase and injected into MS
- Several rounds of m/z measurements and ion fragmentation

Preparation of protein/peptide samples for MS

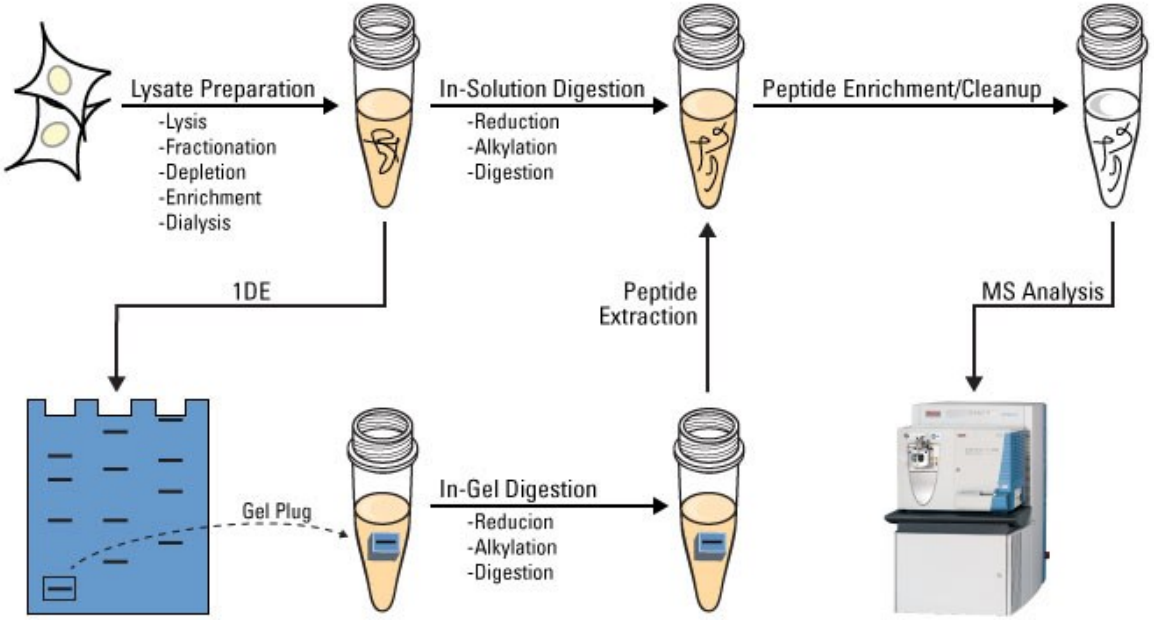


Image from https://www.thermofisher.com

- Whole-cell lysate or sub-cellular
  - Scope of study
- In-solution vs in-gel
  - Selection of specific protein complexes
- Intact proteins or digested into peptides

Protein digestion by protease enzymes

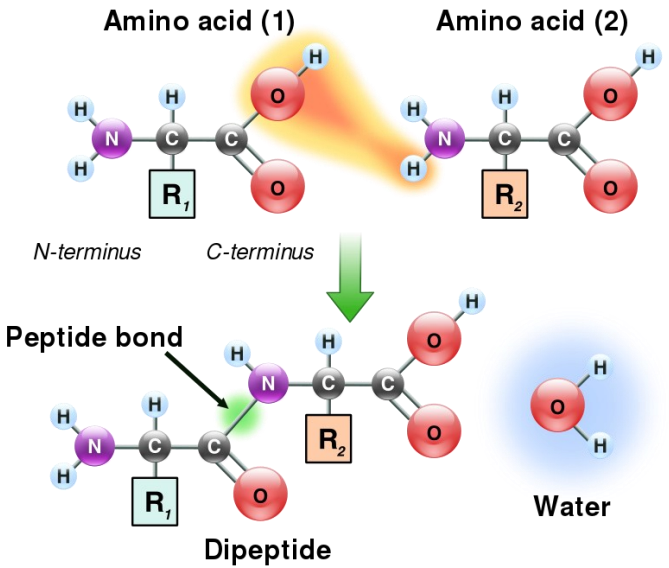
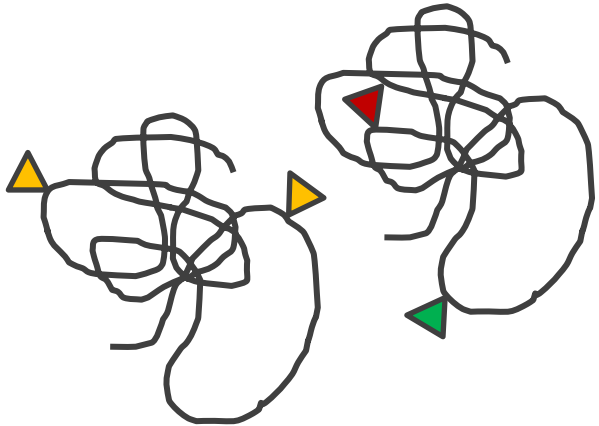


Image from
https://en.wikipedia.org/wiki/Peptide_bond

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

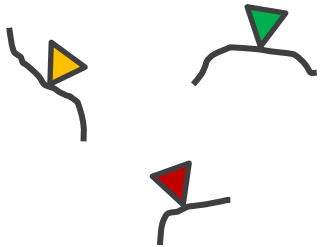
Image from MaxQuant software

Top-down vs bottom-up proteomics



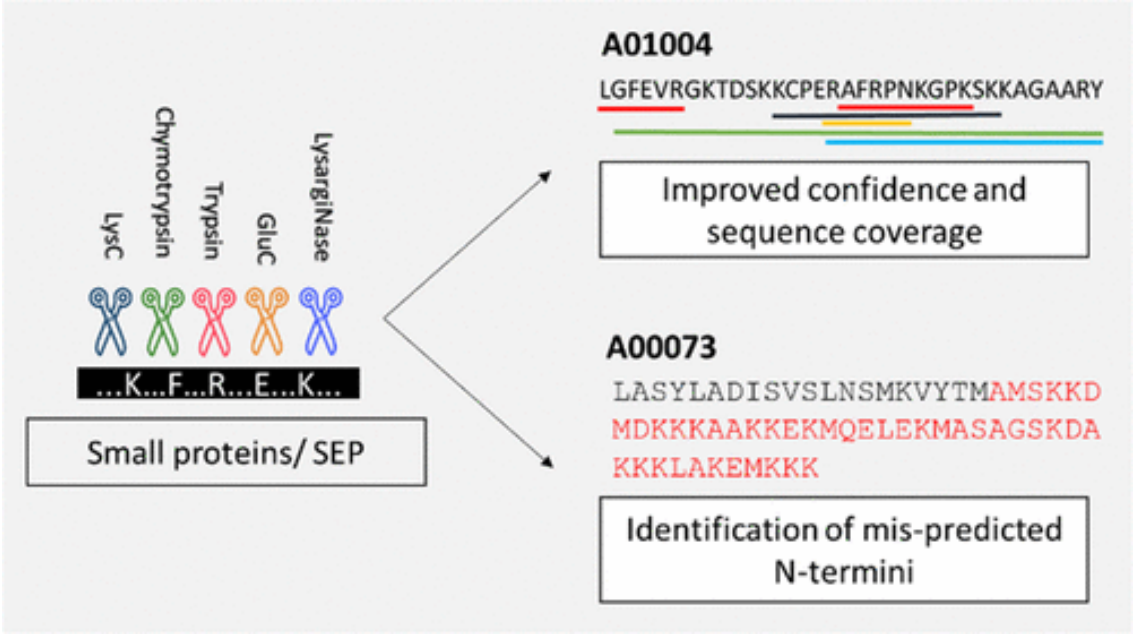
- Top-down = analysis of in-tact proteins
  - Complex MS data
  - Limited to ~50-100 kDa proteins
  - Can identify co-occurring PTMs
  - Can identify multiple species of a proteins

- Bottom-up = analysis of peptides from digested proteins
  - Easier to analyze
  - Applicable to all protein samples
  - PTMs on multiple peptides cannot be linked



- Similar to short-read vs long-read benefits

Combining multiple proteases



- Ensure well-sized peptides throughout the protein

Handling the complexity of proteme

The need for chromatography

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

- It takes time for MS to collect ions and analyze their m/z profiles
- Cannot dump the entire proteome in the MS at once
  - Ions that cannot be analyzed in time will be lost

High/Ultra-performance LC (HPLC / UPLC)

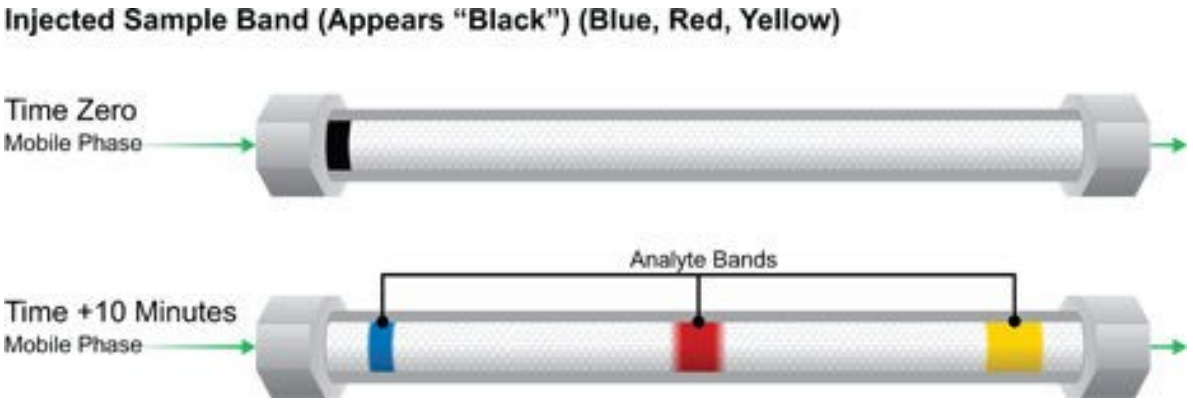


Image from https://www.waters.com

- Separate molecules by their isoelectric potential, hydrophobicity, and size
- Different molecules take different amounts of time to pass through (retention time)

Sample fractionation

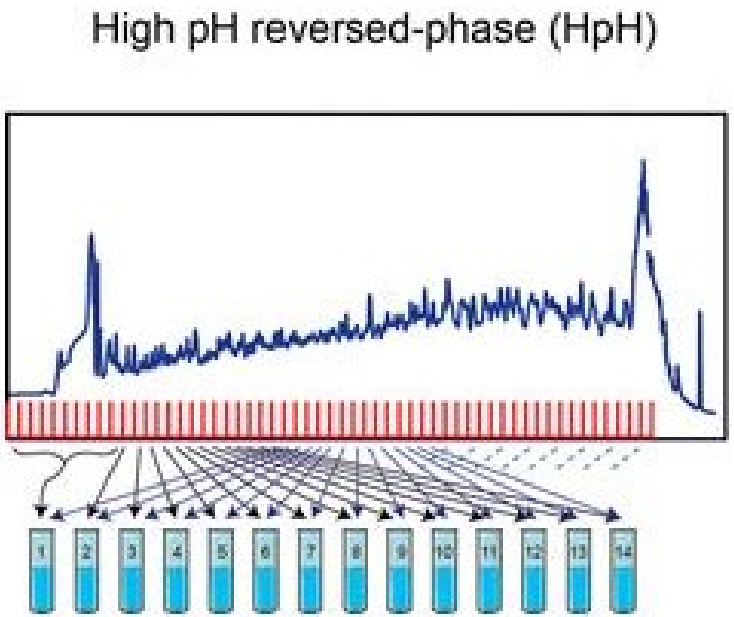


Image from Chen et al. Analyst, 2018

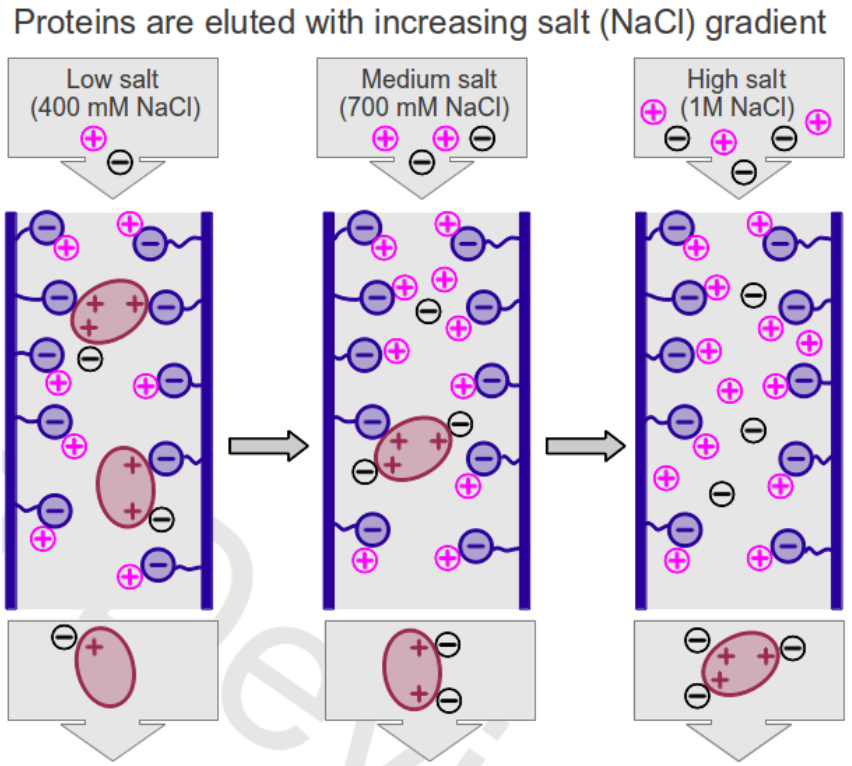
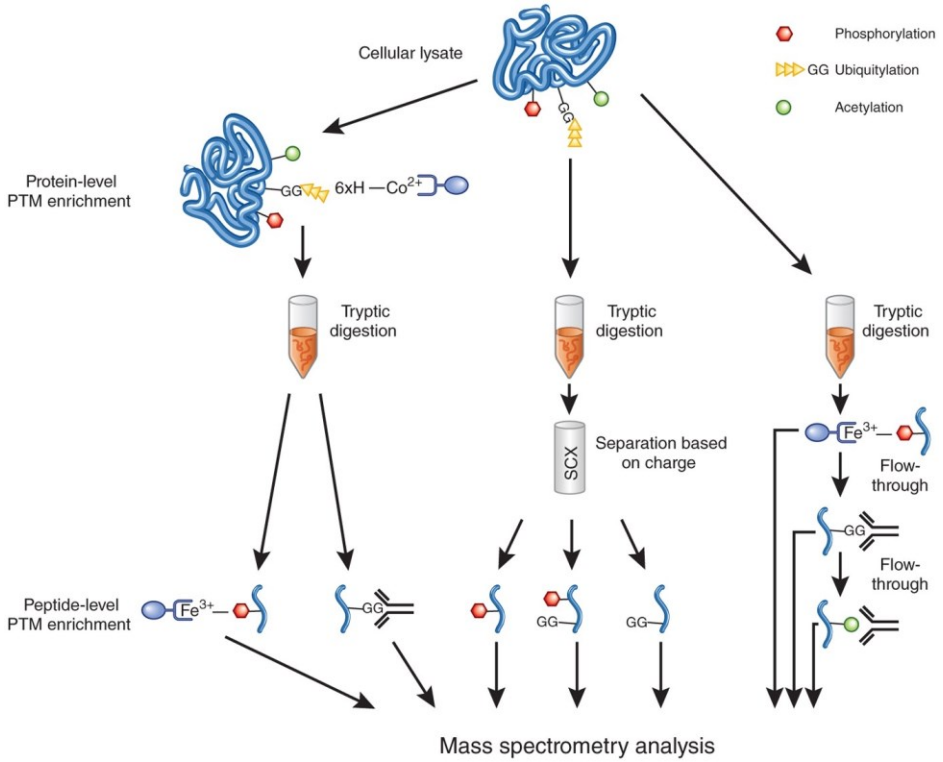


Image from http://www.reachdevices.com/

Enrichment of target proteins / peptides



- Same idea as the scope of DNA sequencing
- Analysis of kinases may focus on phosphorylated peptides
- Depend on the biological mechanism of interest

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

Another reason why proteomics is not that popular

- Let’s say we have blood samples from 3 patients and 3 controls
- To analyze every protein, each sample is divided in to 10 fractions
- LC-MS/MS of each fraction takes 2-4 hours to complete

- How much time do we need ?
  - 6 x 10 x 4 = 240 hours = 10 days

Peptide identification

The journey of a peptide

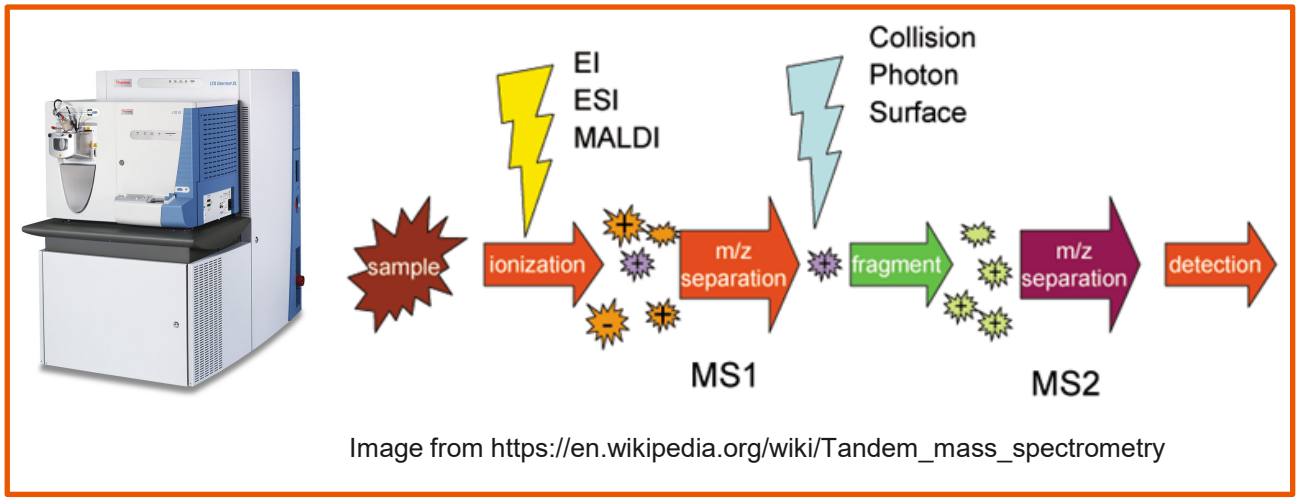
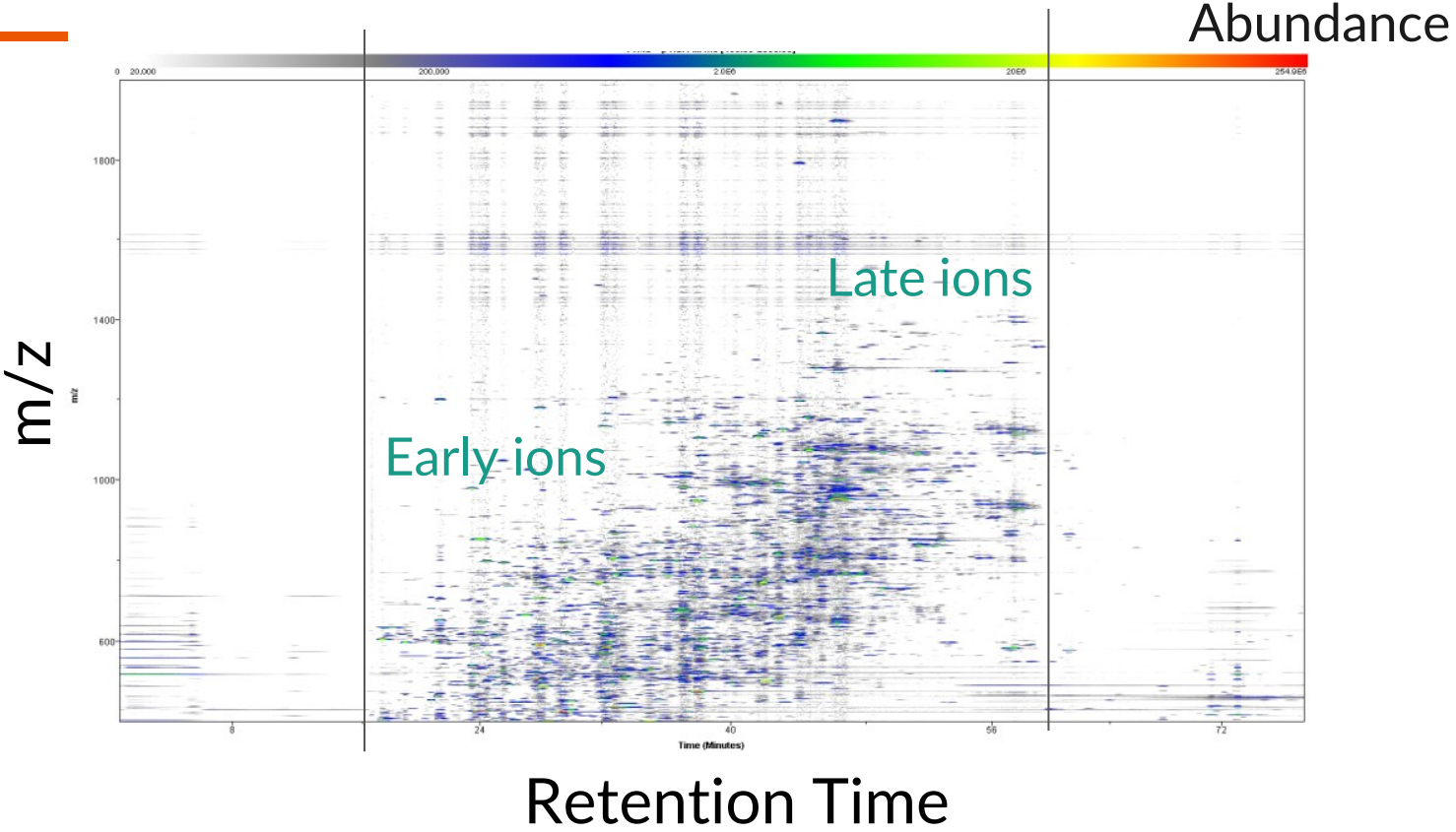
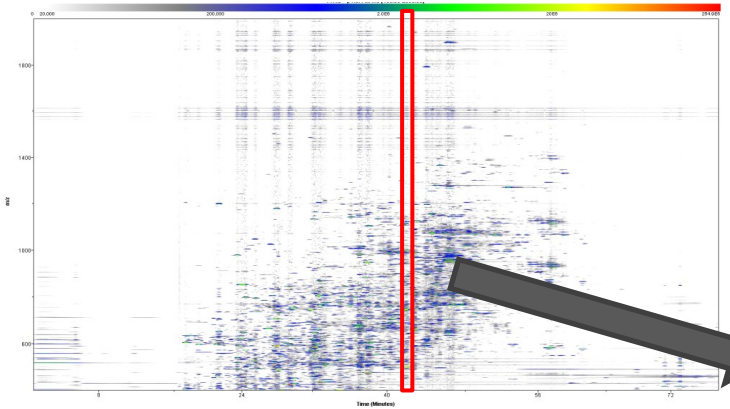


Image from https://en.wikipedia.org/wiki/Tandem_mass_spectrometry

The profile of all peptide ions

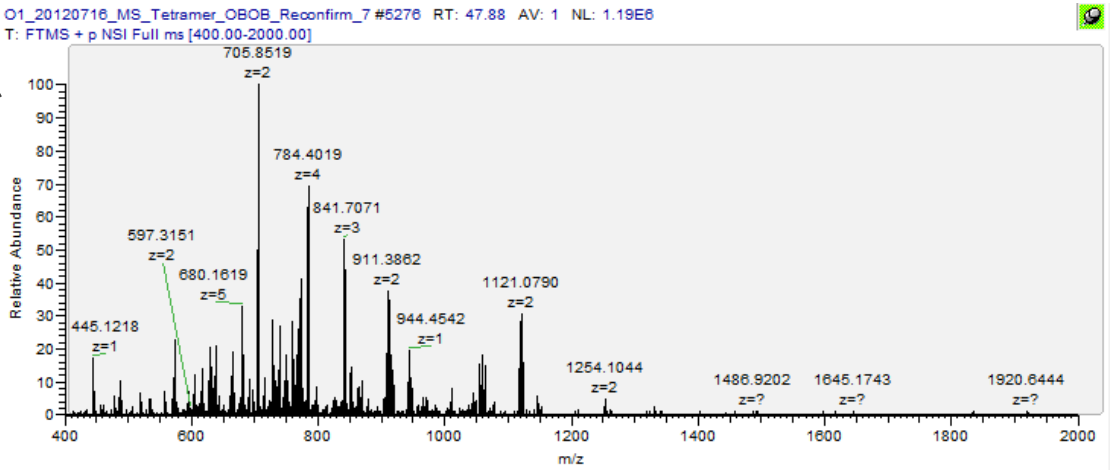


Profile of peptides at a specific retention time

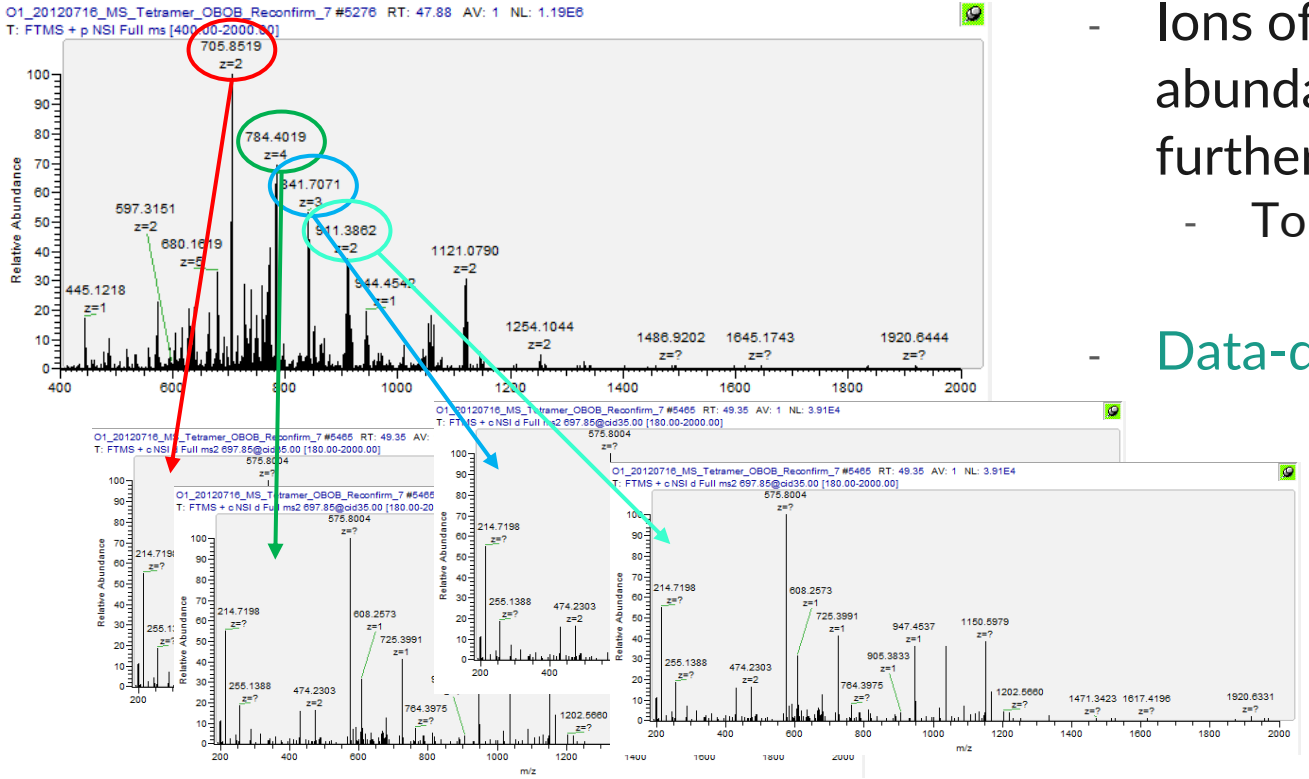


- At a time, a number of peptide ions were injected into MS

- MS first measures the m/z values of all peptide ions (MS1 spectrum)



MS analysis of individual peptides



- Ions of interest (with high abundances) are isolated and further analyzed
  - Top 5 up to top 20 ions
- Data-dependent mode

MS/MS (tandem MS)

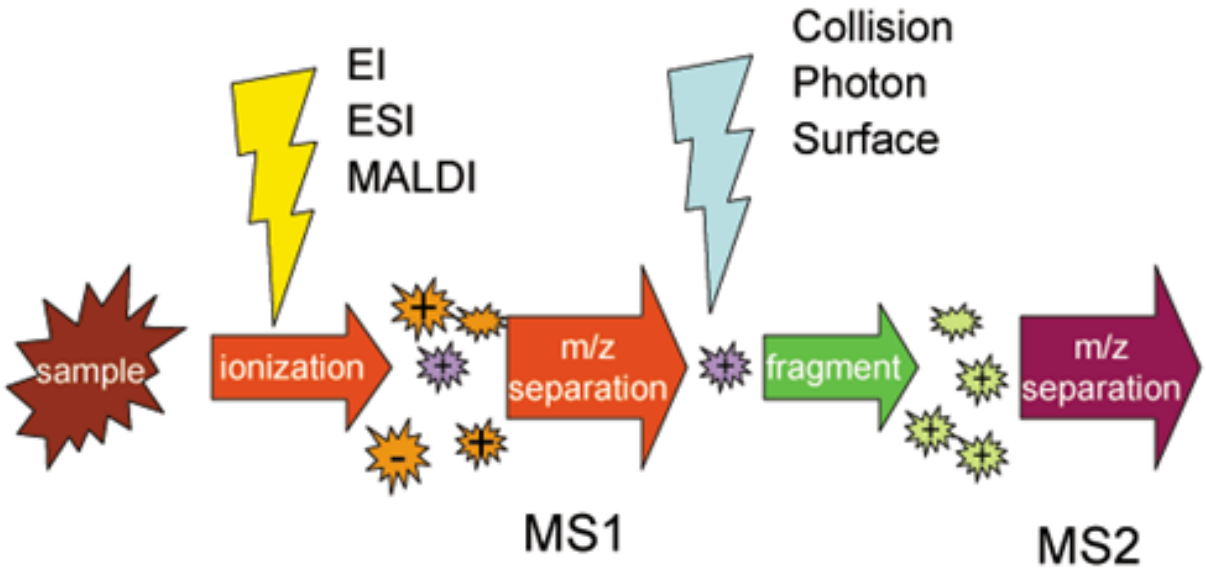


Image from https://en.wikipedia.org/wiki/Tandem_mass_spectrometry

- Fragmentation of each peptide into smaller characteristic ions

Collision-induced dissociation

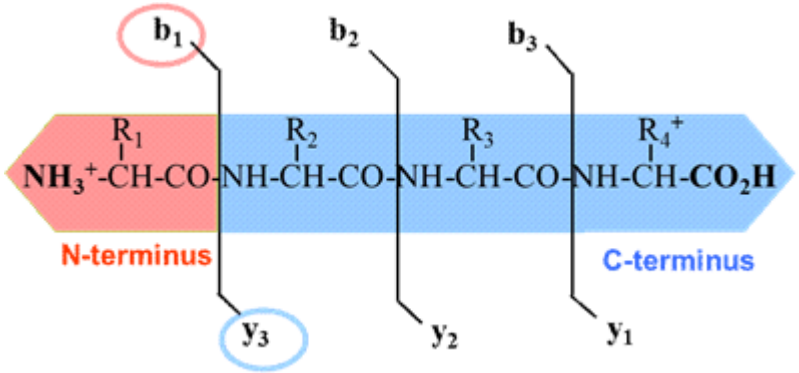


Image from https://www.molgen.mpg.de

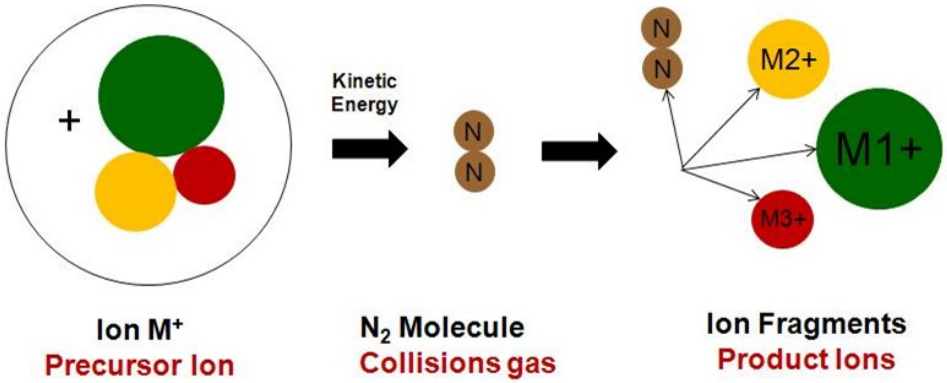


Image from https://www.biologie.hu-berlin.de

- Accelerate peptide molecules to run into inert gases, such as helium, nitrogen, or argon
- Collision at the **right energy level** breaks a peptide bond at random

Electron transfer dissociation

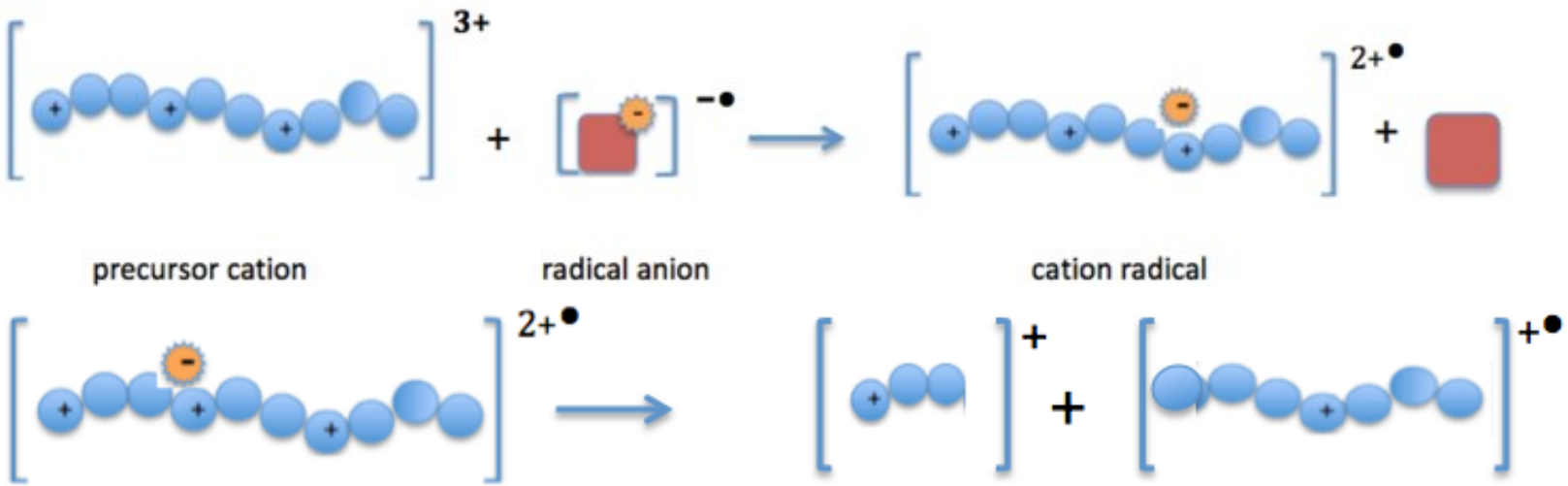
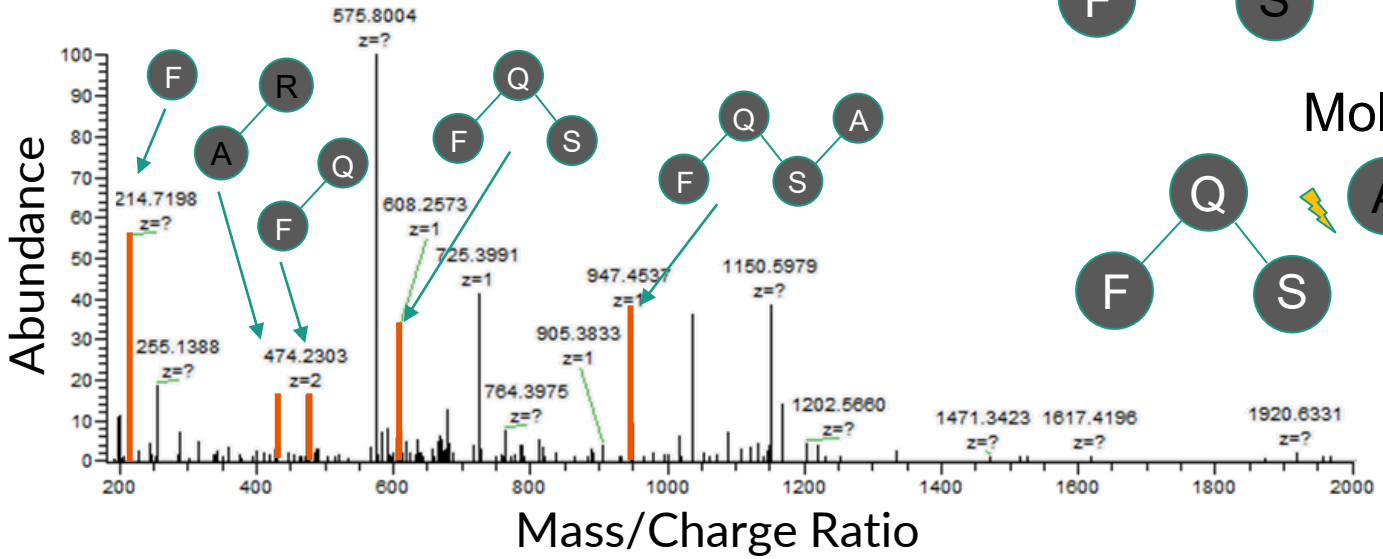


Image from https://en.wikipedia.org/wiki/Electron-transfer_dissociation

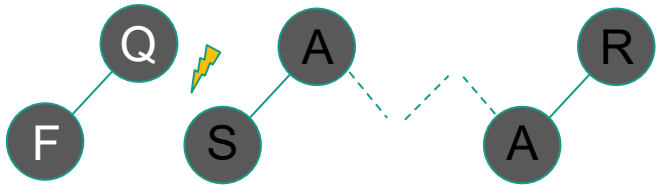
- Transfer of electron onto a positively-charged peptide breaks it
- Fragmentation occurs on the R-CH-NH bond, not the peptide bond

MS/MS spectrum = profile of fragment ions of a peptide

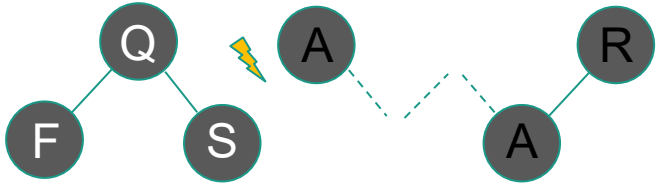
FQSADETGQDLVNANHEAR



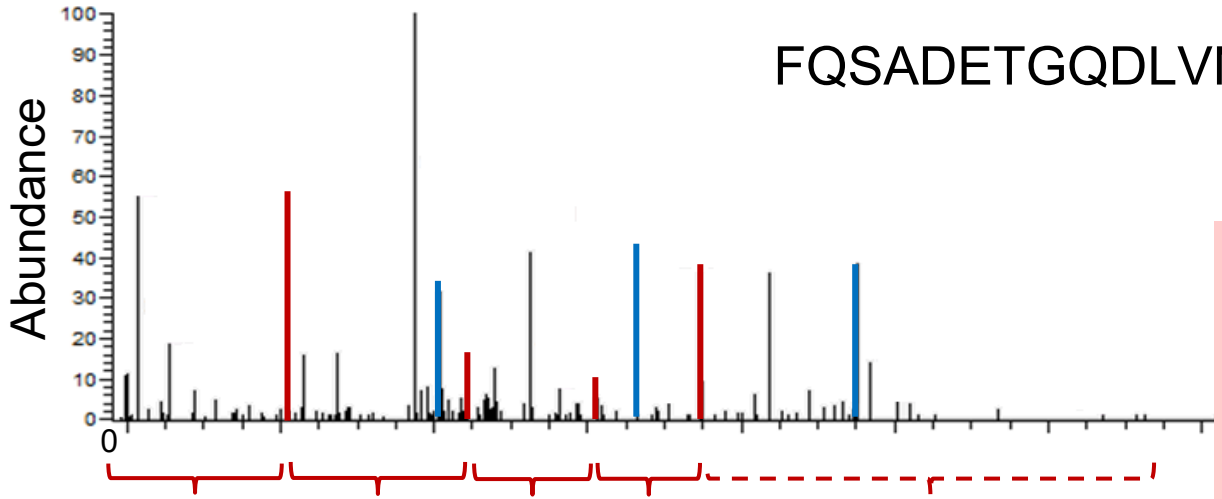
Molecule #1



Molecule #2



Peptide sequencing = decoding MS/MS spectra



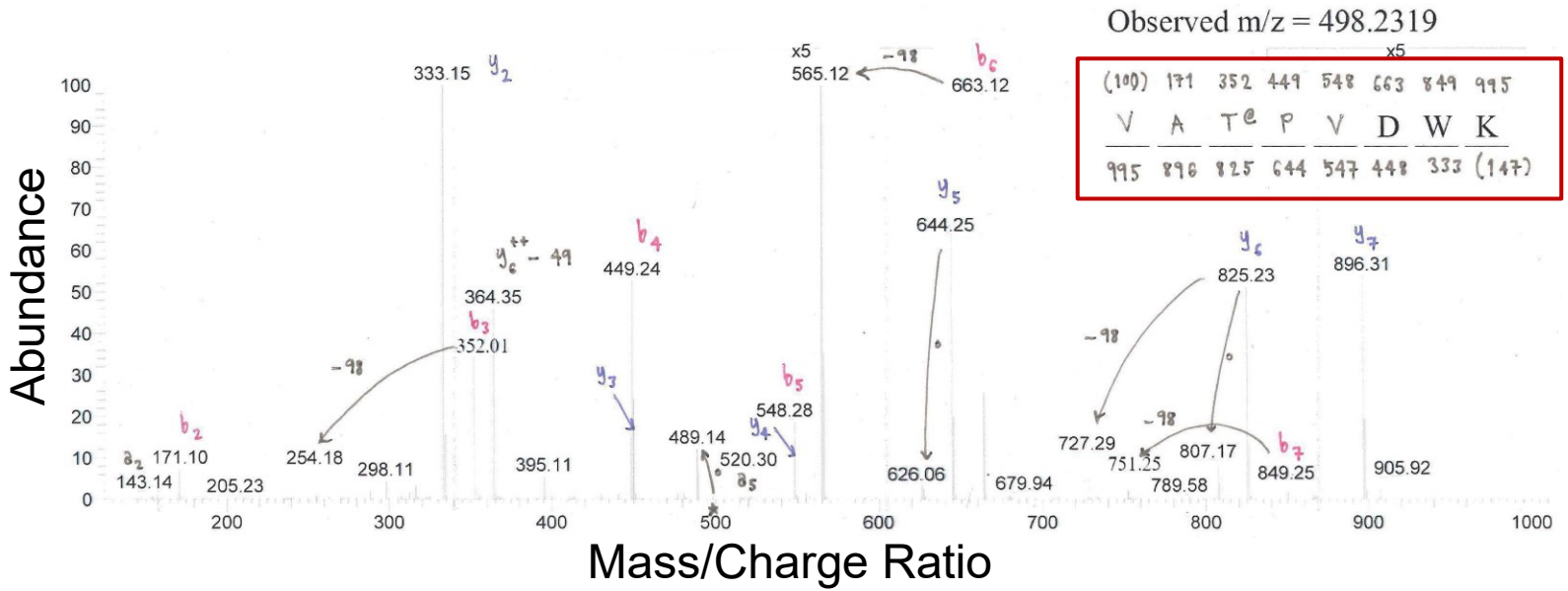
FQSADETGQDLVNANHEAR

147.07	F
128.06	Q
87.03	S
…	…

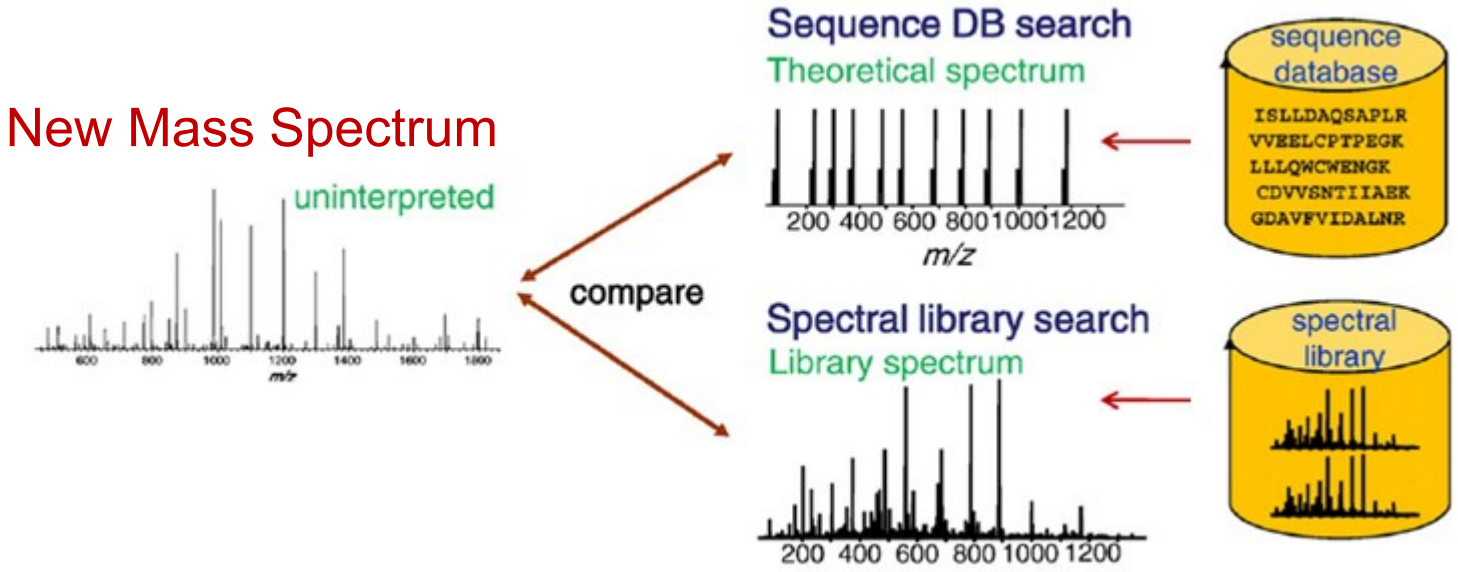
Delta Mass: 147.07 128.06 87.03 71.04 …

Delta Mass: 227.14 129.04 137.06

Manual peptide sequencing



Database search for peptide sequencing



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

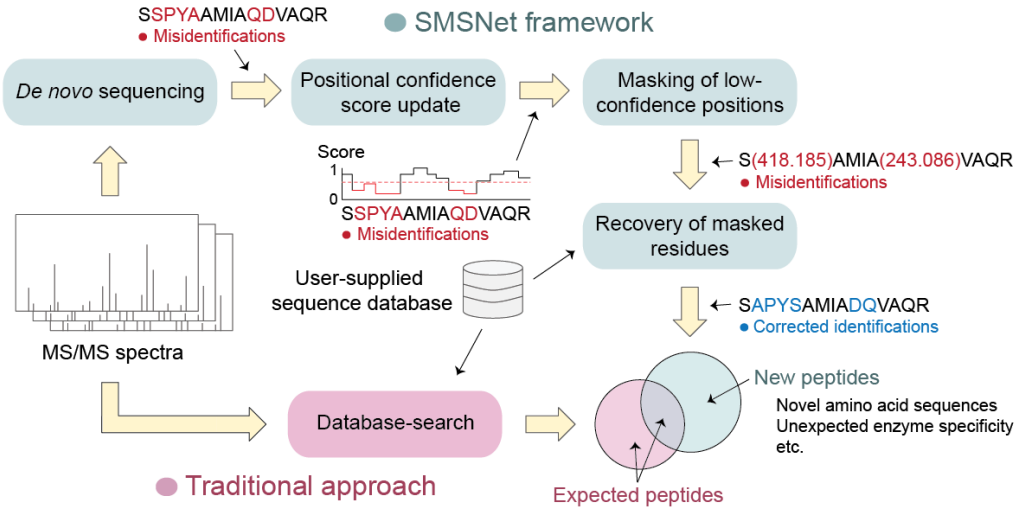
- We will do a demo with MaxQuant in the next session

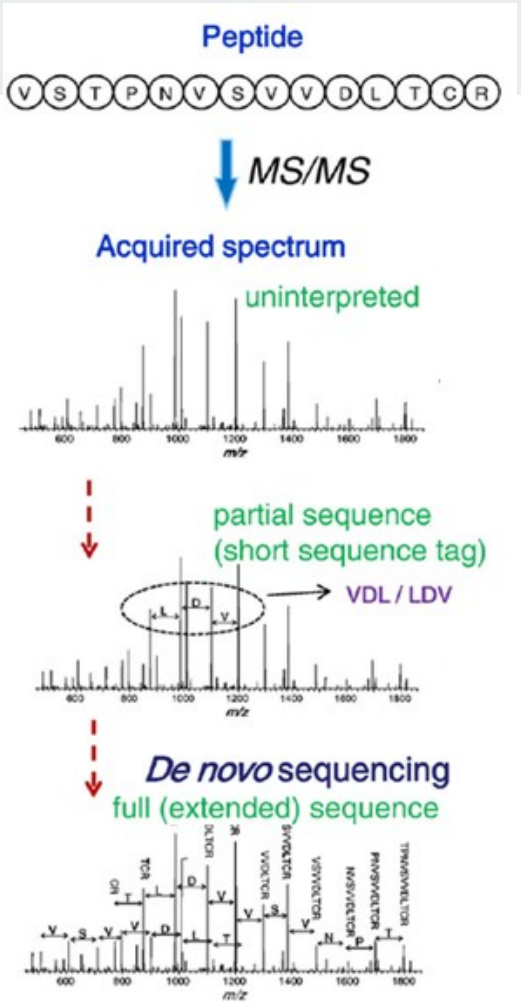
Limitation of database search

- Only identify the **best known answer**
  - Doesn't work on data from new species
  - Cannot identify cancer mutations
- Including too many proteins → False hits
- Not practical to search many PTMs
  - FQSADETMAR with oxidation and phosphorylation = 8 possibilities
  - PTM changes amino acid mass → changes MS/MS m/z values

De novo peptide sequencing

- Directly deducing amino acid sequences
- Can identify sequence tags (partial sequence)
- Rely on deep learning (AI)





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

Peptide quantification

Peptide abundance from MS1 spectrum

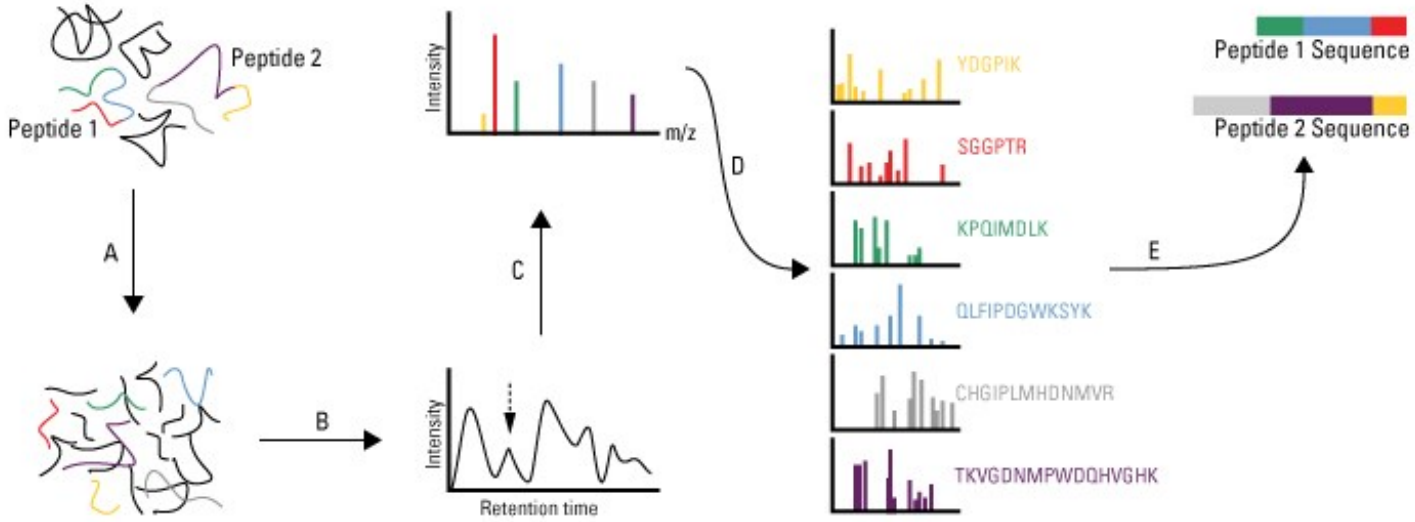


Image from http://fields.scripps.edu/yates/wp/

- Intensity values in MS1 spectrum reflect peptide abundances

Peptide and protein quantification

- Number of MS/MS of a peptide sequence
- Intensity of the peptide in MS1 spectrum
  - Not all peptides can be ionized
  - Different ionization efficiencies

- How to quantify proteins with multiple peptides?
  - Sum of all peptides
  - Sum of top *N* peptides
  - Consider only high-quality peptides
  - Linear effect model

- Intensity values are log-normal → Log-transformed + *t*-test

Handling of shared peptides

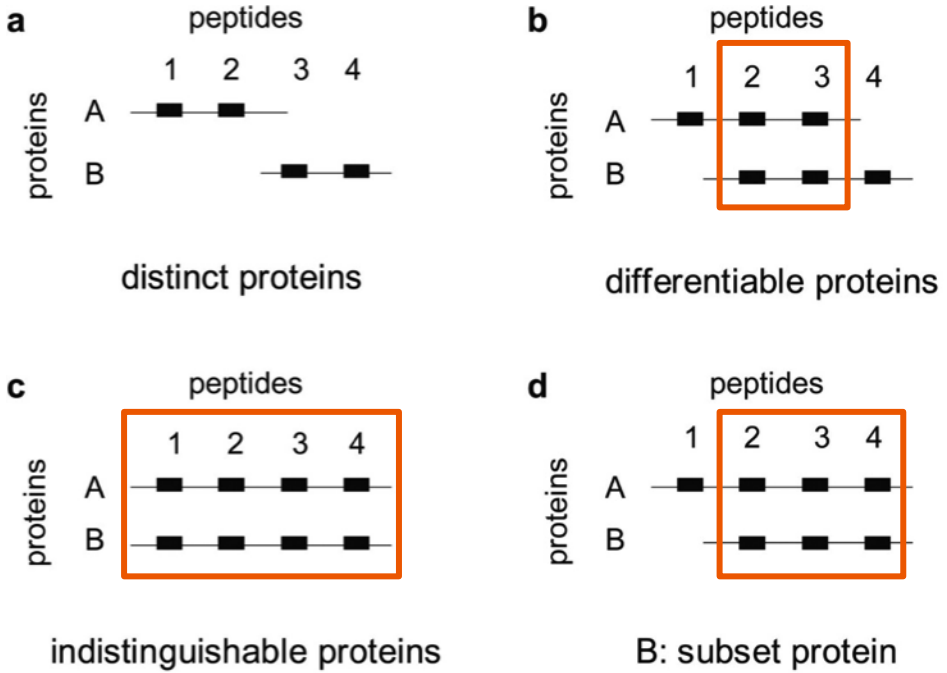
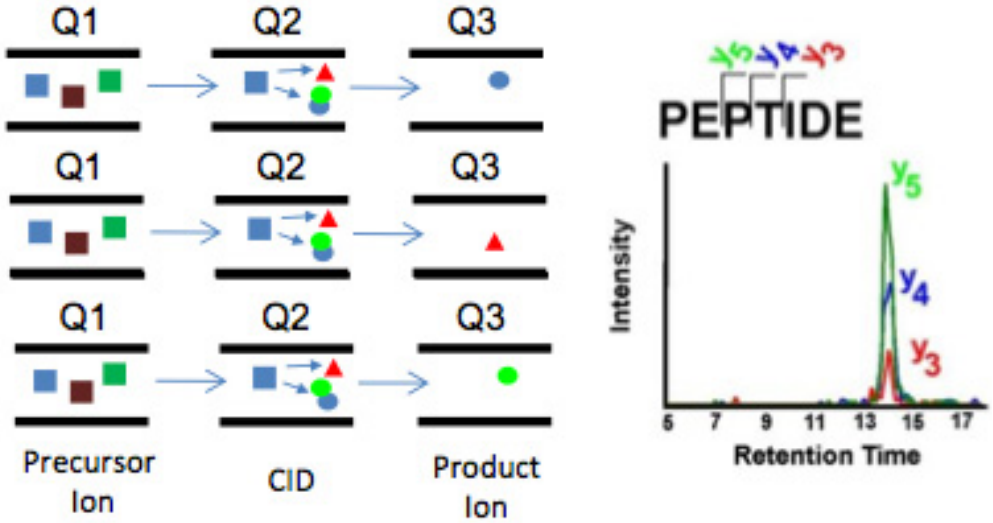


Image from https://www.inf.fu-berlin.de

- Assign shared peptides to isoforms with the highest abundance
- Distribute shared peptides proportionally
- Disregard shared peptides
- Report indistinguishable proteins as a group

Multiple reaction monitoring (MRM)

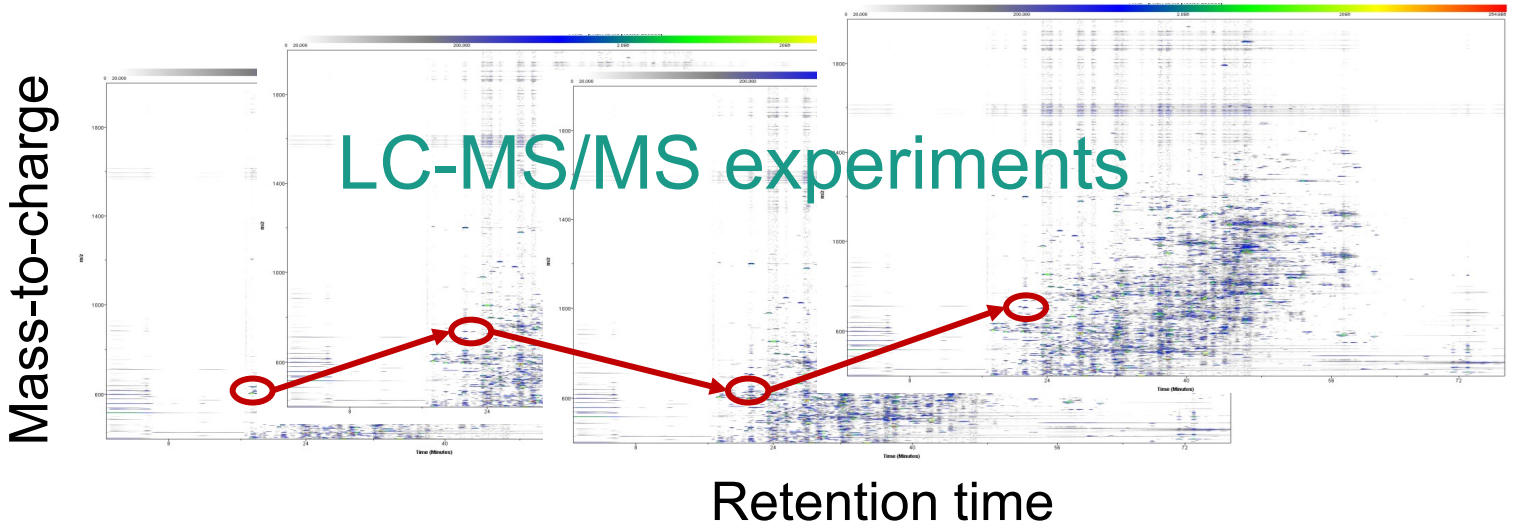


https://medicine.yale.edu/keck/proteomics

- Targeted MS analysis of specific peptides and their fragmented ions
- Commonly performed on quadrupole MS (different from proteomics)

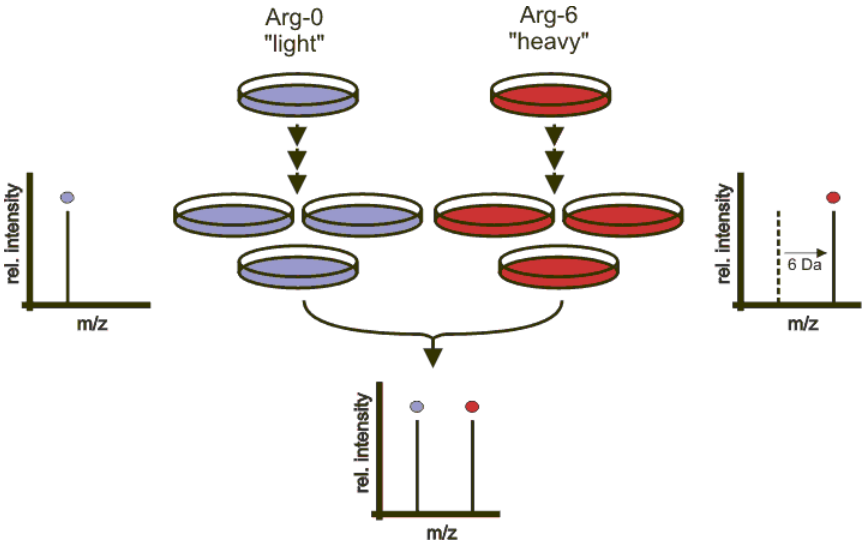
Comparative proteomics

Label-free comparison



- Simply perform multiple LC-MS/MS runs and match the observed m/z and retention time patterns

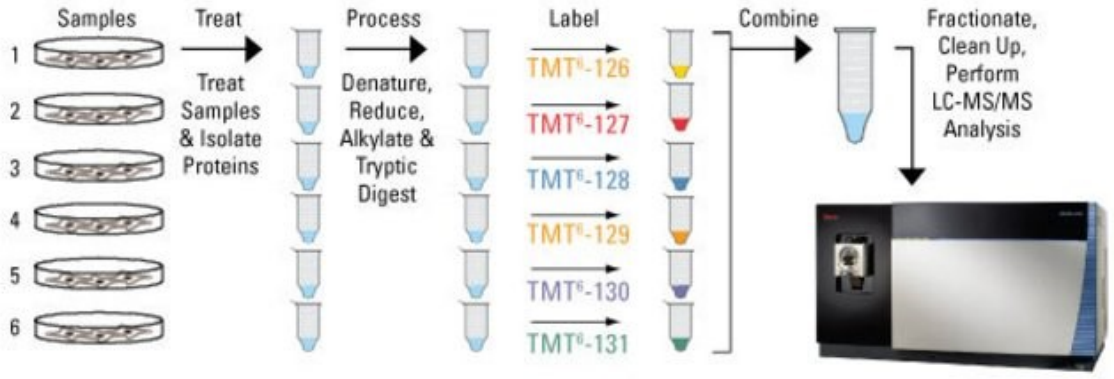
Stable isotope labeling (SILAC)



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

- Feed cell cultures with heavy and light isotopes
  - Integrated into all proteins
- Mix samples and perform a single LC-MS/MS run
  - No technical bias
- Look for a pair of peptides
  - With mass shift
  - With similar MS/MS

Tandem mass tag (TMT)



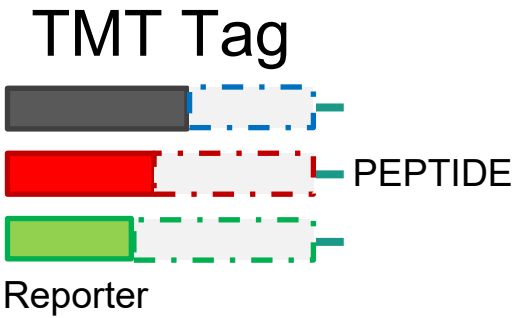


Image from https://www.thermofisher.com

- TMTs are added to N-terminus of peptides
- All TMTs have the same total mass
- Different TMTs have different reporter mass (126 Da, 127 Da, etc.)
  - Will be separated from the peptide molecule in tandem MS

Comparative quantification via TMT

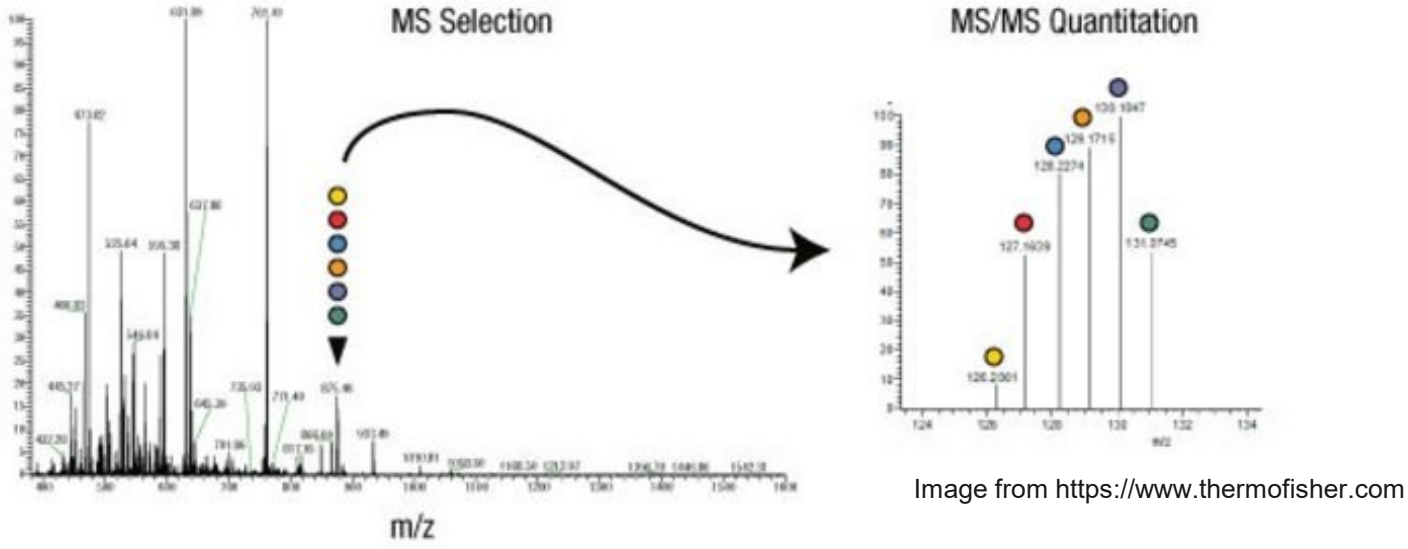
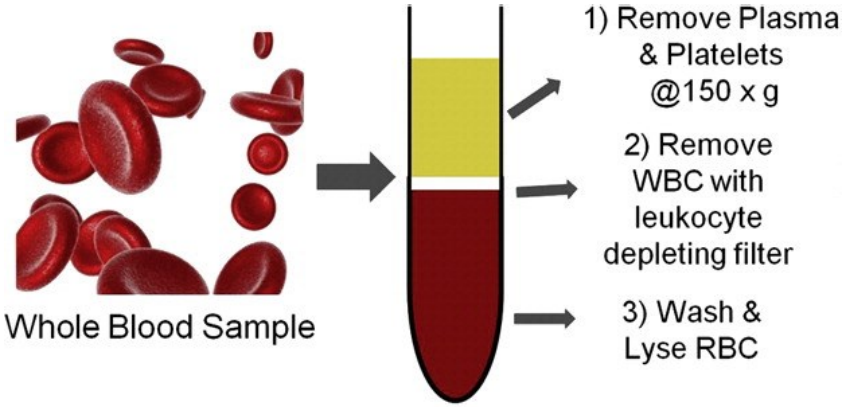


Image from https://www.thermofisher.com

- The same peptides attached with TMT will be isolated together
- Peptide with different TMT will produce different reporter ions in MS/MS

Some clinical applications

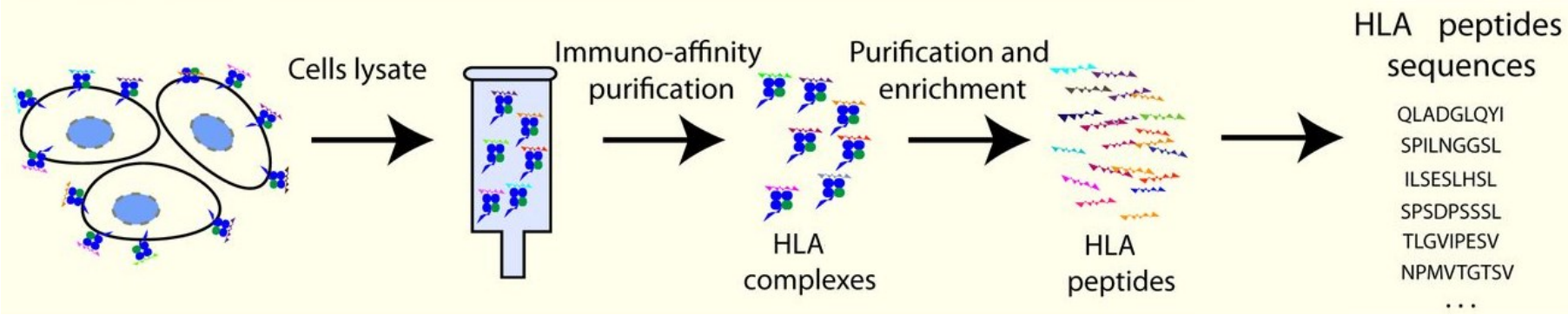
Tissue consists of multiple cell types



Slice & Digest with Trypsin

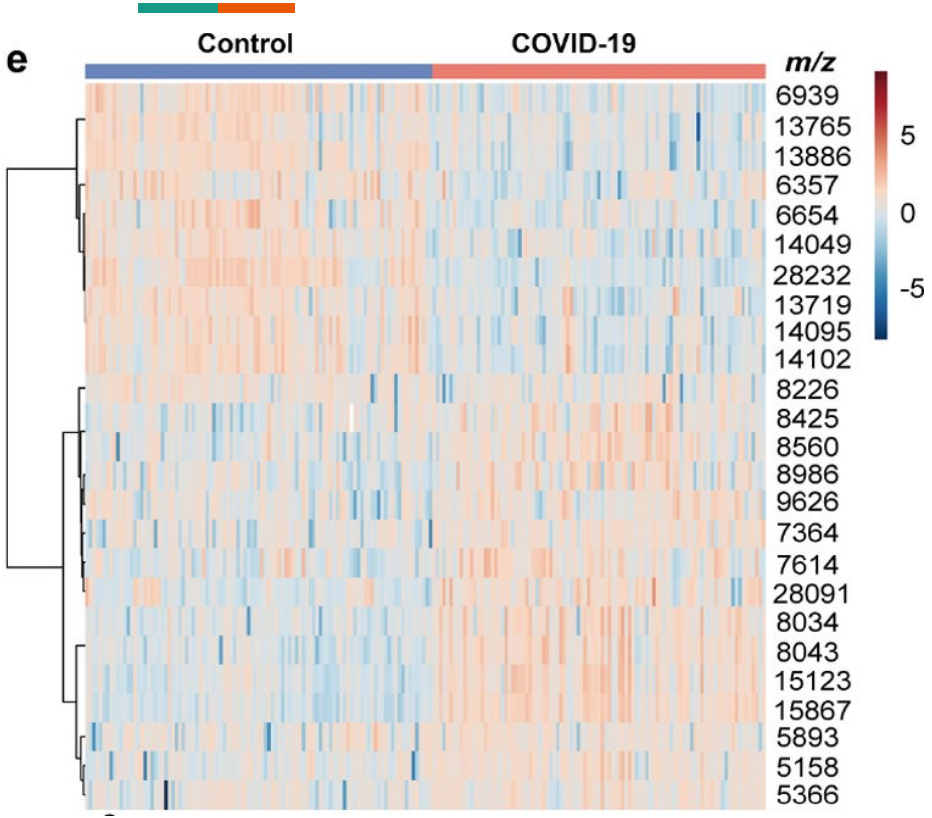
- Finding plasma biomarkers for various diseases
- Protocols for depleting abundant blood cell-specific proteins (uninteresting for disease study)

HLA peptidomics



- Targeted MS analysis of HLA-bound antigen peptides
- Identify cancer-specific neo-antigens → develop into cancer vaccine

Saliva biomarkers for COVID-19 diagnosis



- Attempted during PCR reagent shortage
- Compare whole proteins in saliva from patients to controls
- No MS/MS = only m/z signatures

Structural proteomics

Chemical crosslinking

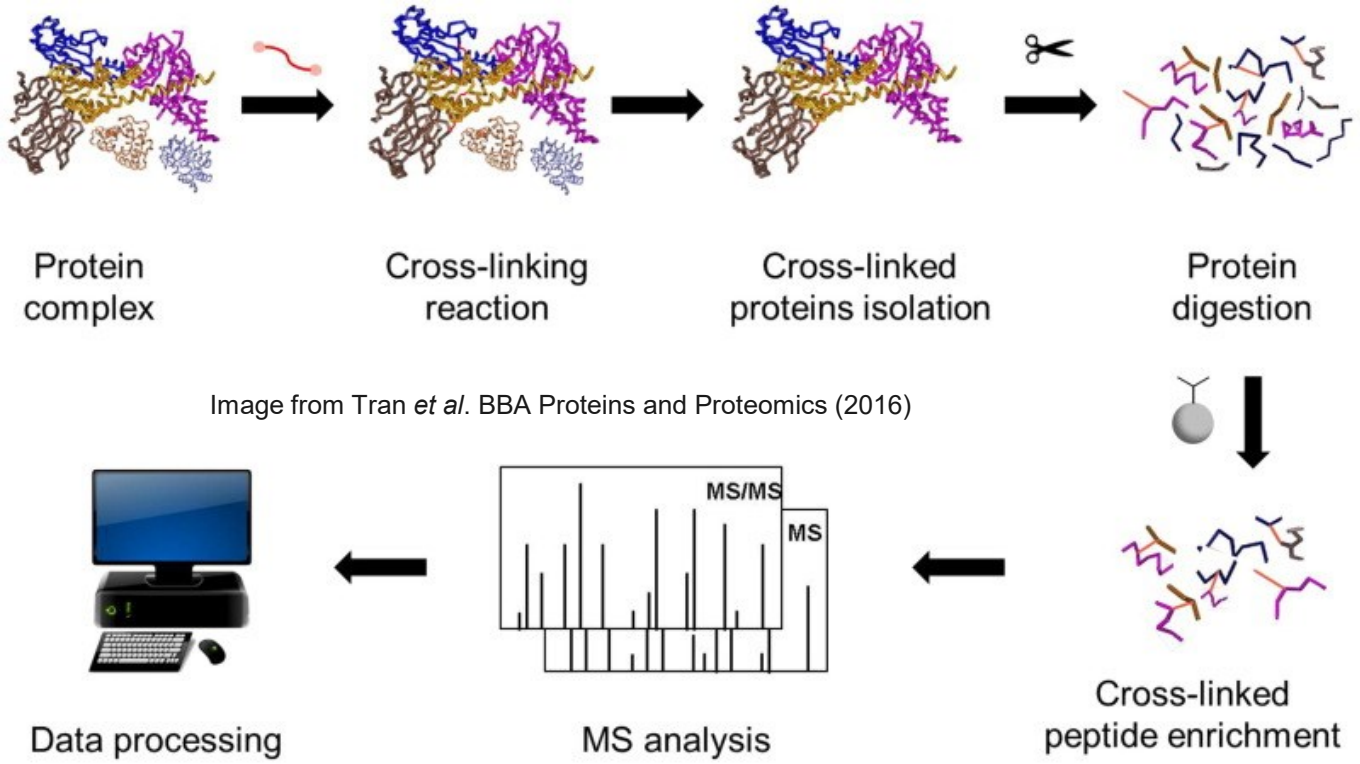


Image from Tran *et al*. BBA Proteins and Proteomics (2016)

Hydrogen-deuterium exchange

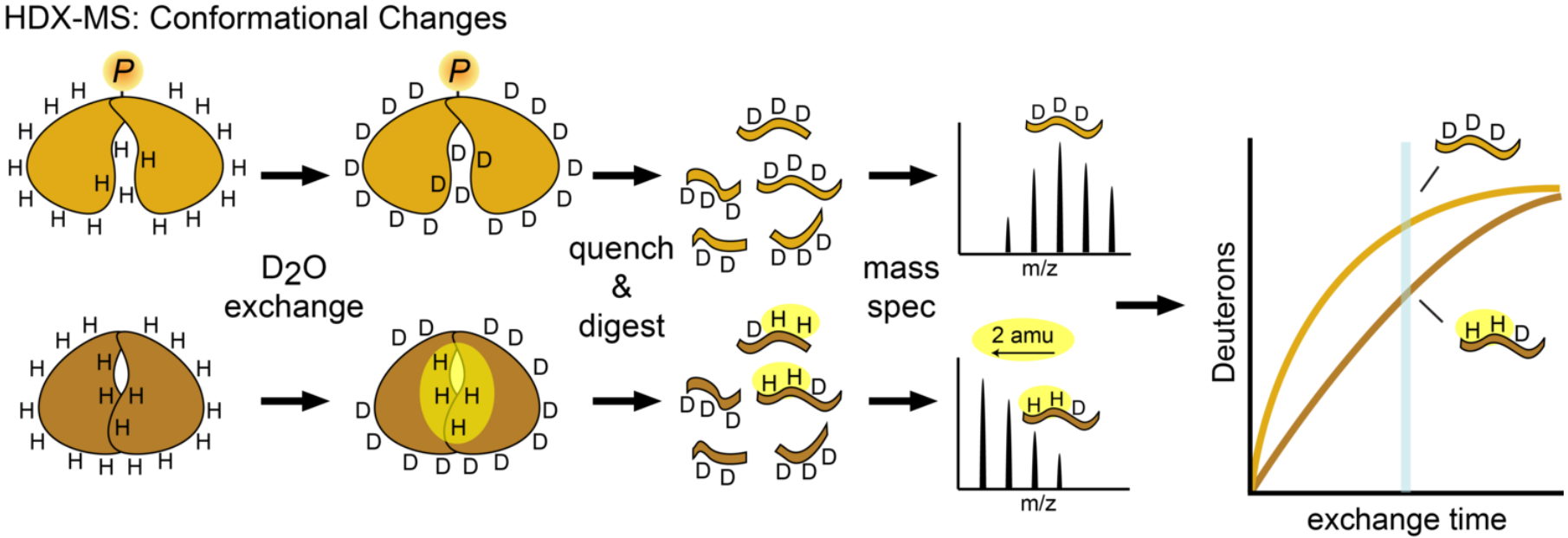
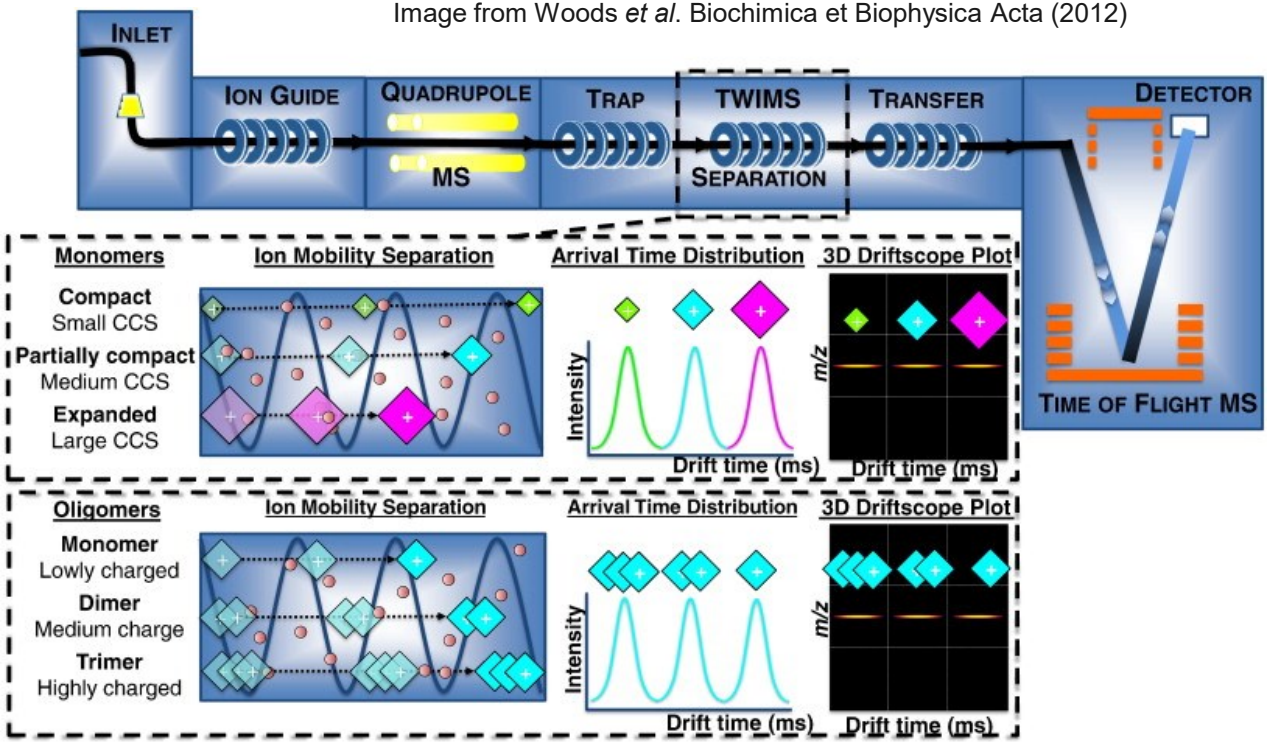


Image from https://www.underbakkelab.org/techniques

- Deuterium readily replaces hydrogen wherever accessible

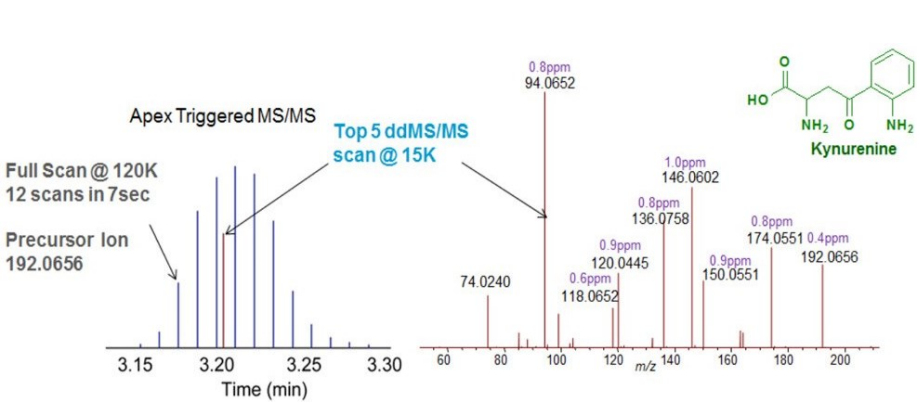
Time-of-flight distinguish protein conformations

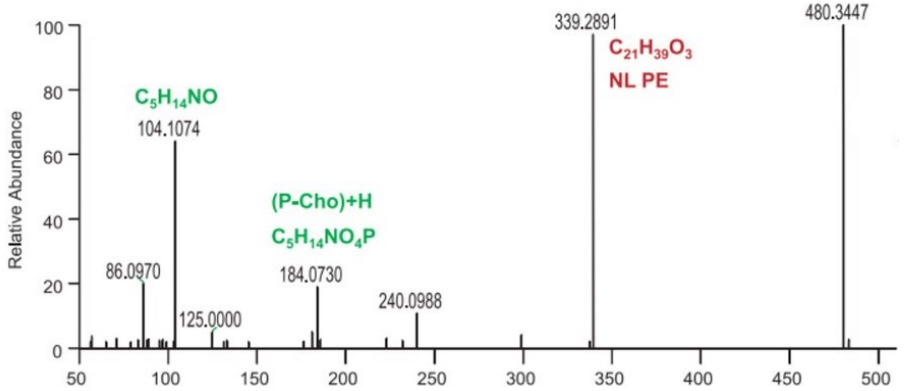
Image from Woods *et al*. Biochimica et Biophysica Acta (2012)



MS for other molecules

MS/MS of non-peptides





- Choosing the right collision energy level to break chemical bonds
- Comparing to a database of known MS/MS spectra of known compounds

Summary

- Key points and issues in proteomics
- Mass spectrometry
- Peptide sequencing
- Some applications

Any question?

- Oct 12 session will be online (find zoom link on Announcement)