3000788 Intro to Comp Molec Biol

Lecture 15: Proteomics and mass spectrometry

Fall 2025

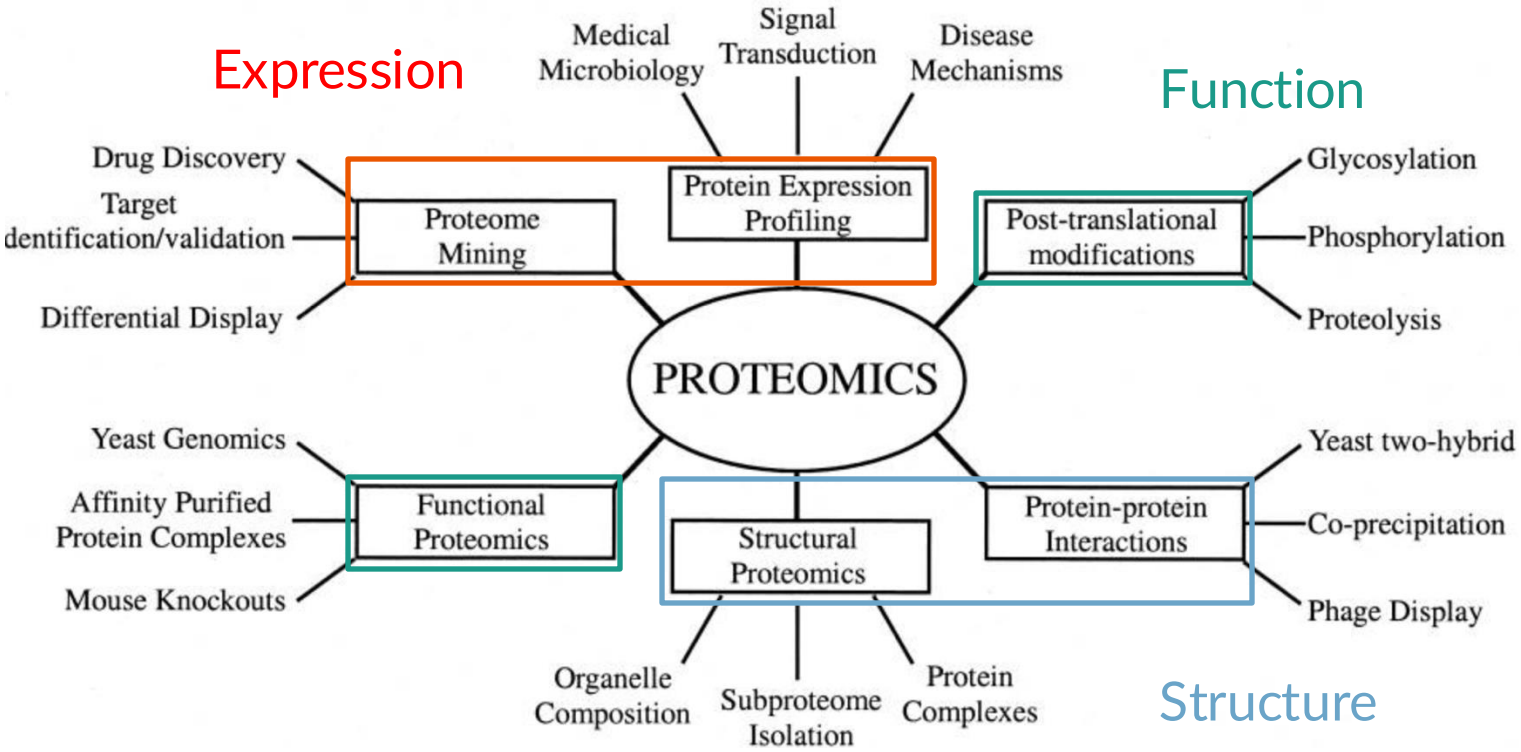
Sira Sriswasdi, PhD

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

Today’s agenda

- What is proteomics?
- Mass spectrometry instrument
- Identification of peptides and proteins
- Extra: Structural proteomics

Proteomics is not just about protein expression



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

Why not proteomics all the time?

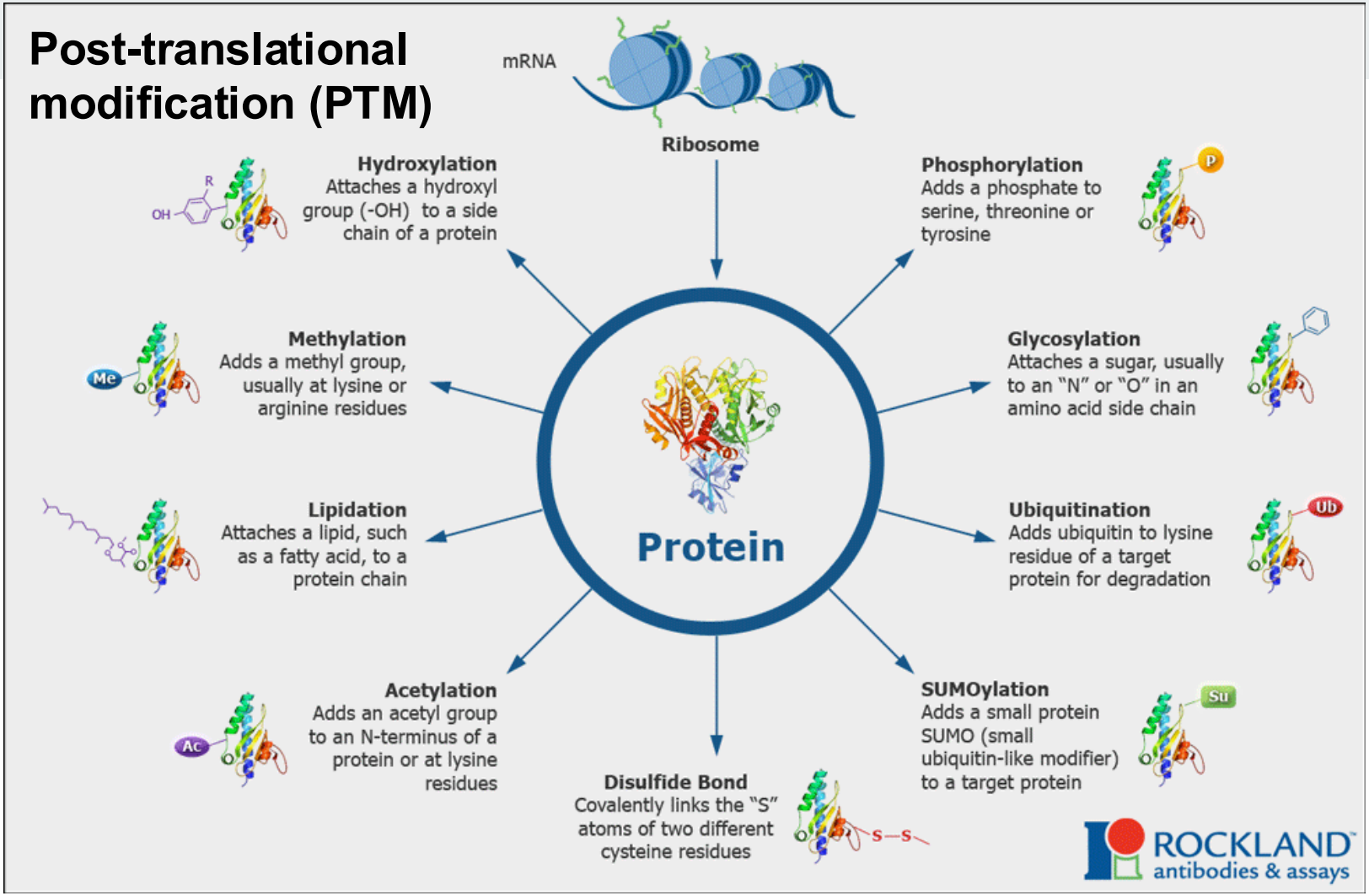
- **Amounts of materials**: Proteins cannot be amplified like DNA

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

- **Difficult to identify**: No direct read out amino acids
  - Similar issue as nanopore (but more costly to synthesize data)

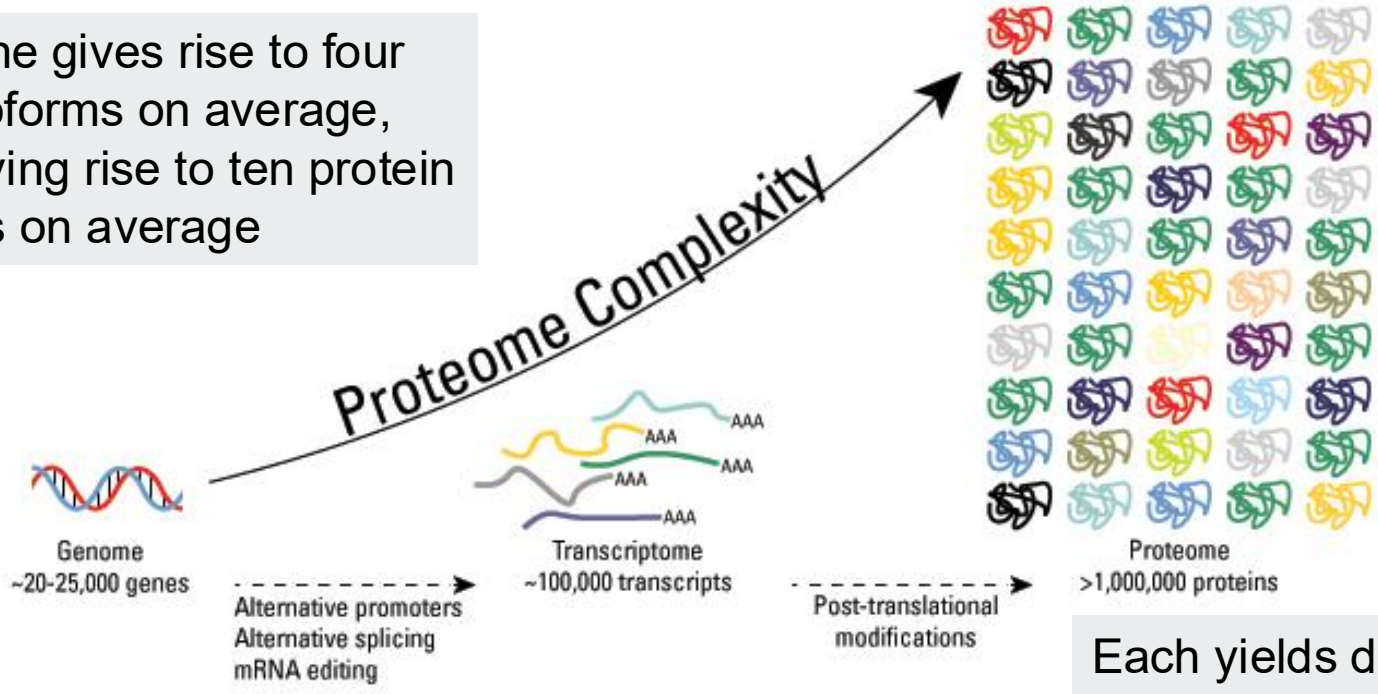
- **Post translational modifications**: Increased molecular diversity

Post-translational modification (PTM)



Explosion of molecular variety of proteins

One gene gives rise to four RNA isoforms on average, each giving rise to ten protein varieties on average



Each yields different signature

Image from https://www.thermofisher.com

Peptide identification sketch

- Peptide: ABCDEFGH
- **Generate fragments: A, AB, ABC, ABCD, …**
- **Measure the weights of all fragments: {10, 15, 22, …}**
- Deduce the original peptide sequence
  - Smallest weight: A = 10
  - Second smallest weight: A + B = 15, B = 5
  - Third smallest weight: A + B + C = 22, C = 7
  - …

In practice, not all possible fragments will be generated.

So, there will be some ambiguity in the deduction.

Also, there will be fragments from other contaminant or background in the measurements

Mass spectrometry

Inside a mass spectrometer

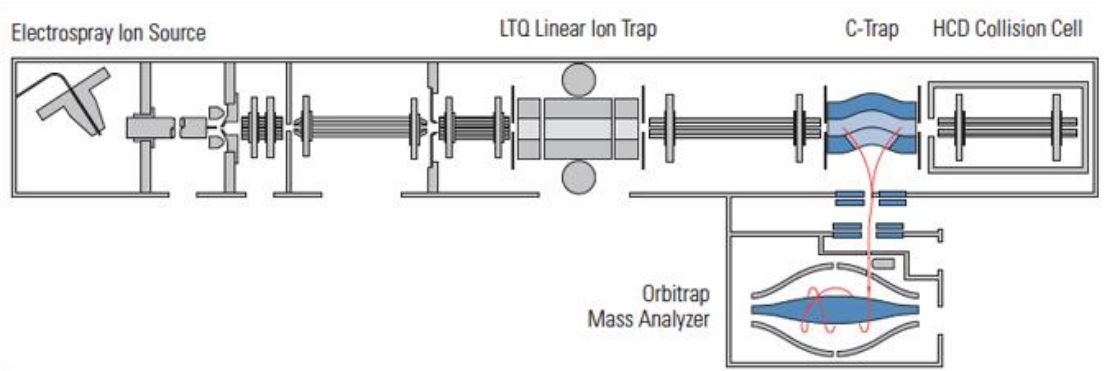
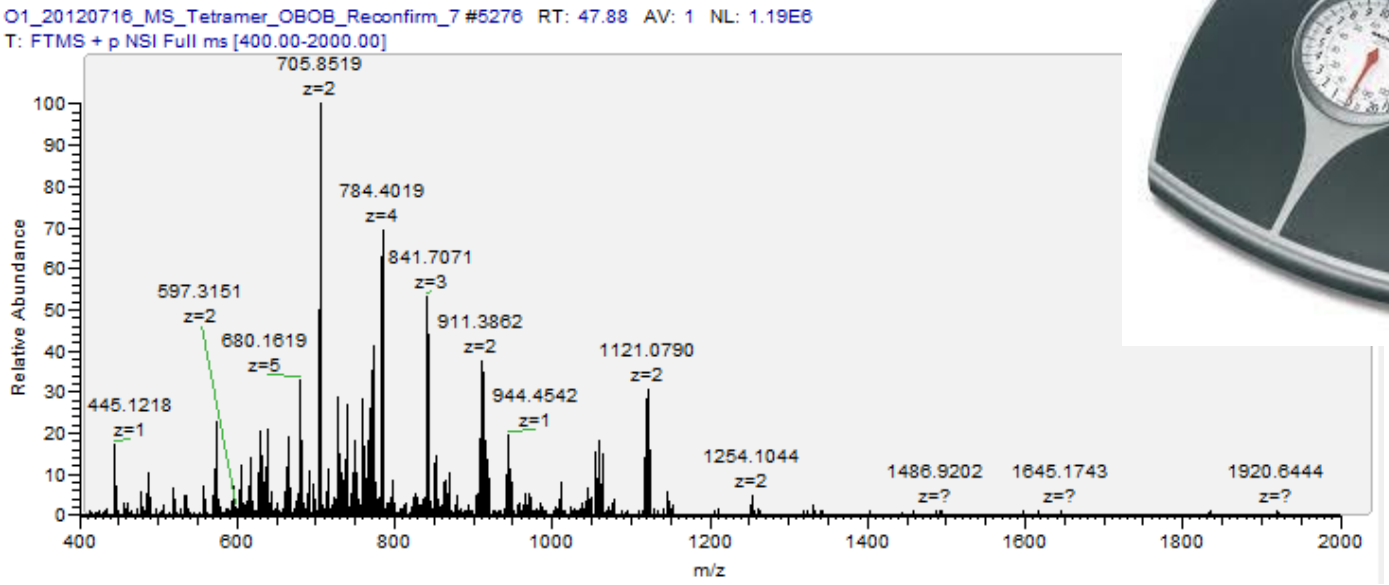


Image from http://planetorbitrap.com/

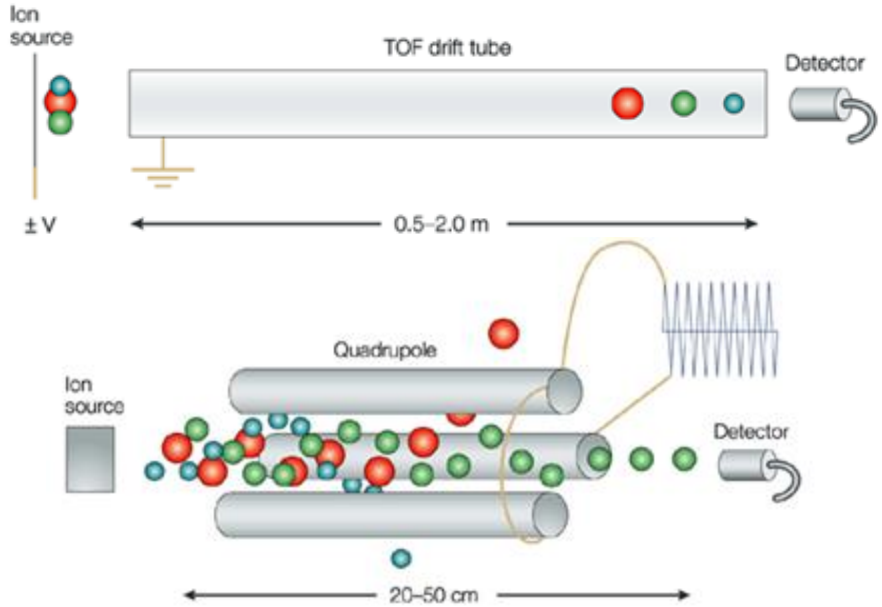
- A series of vacuum chambers for ion trapping and m/z measurement

A mass spectra



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

Physics of m/z measurement



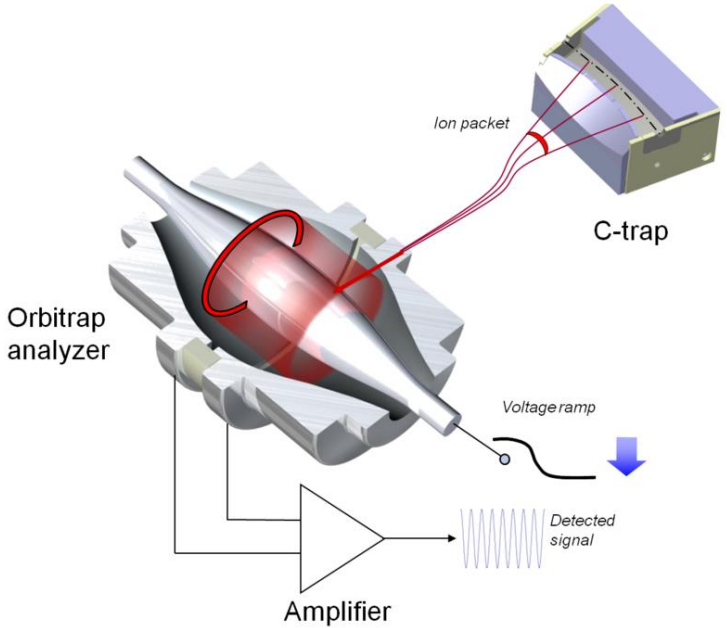


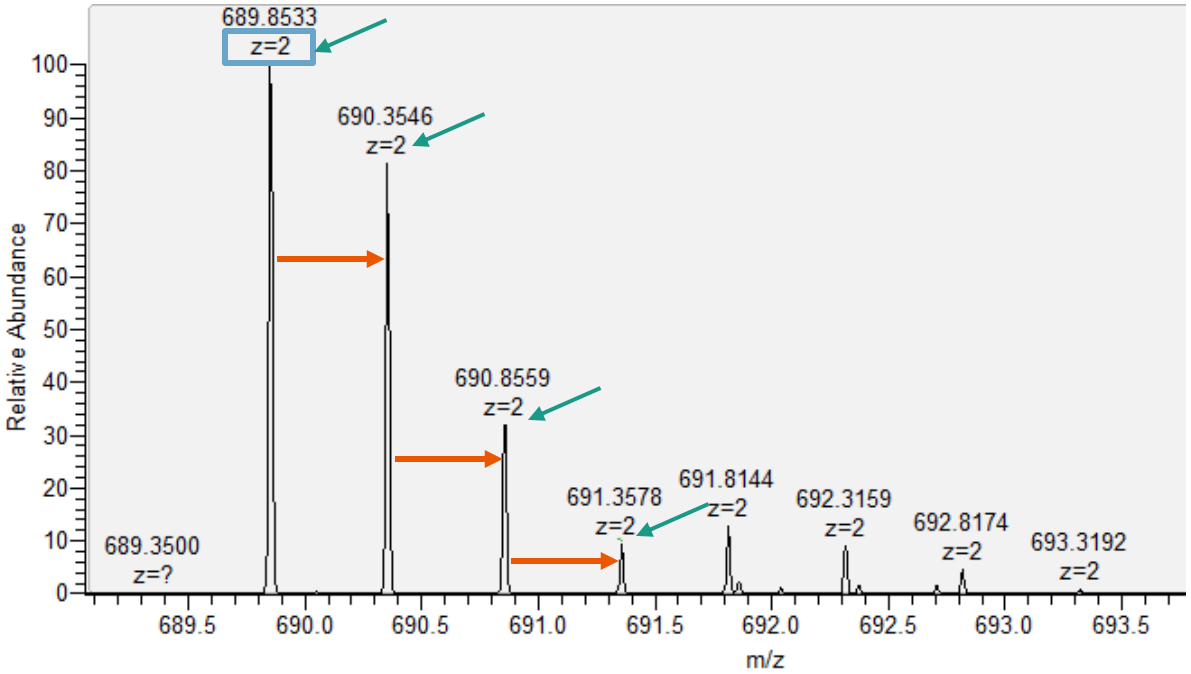
Image from https://en.wikipedia.org/wiki/Orbitrap

- 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 $(m+1)/z = 1/z$

Soft ionization techniques

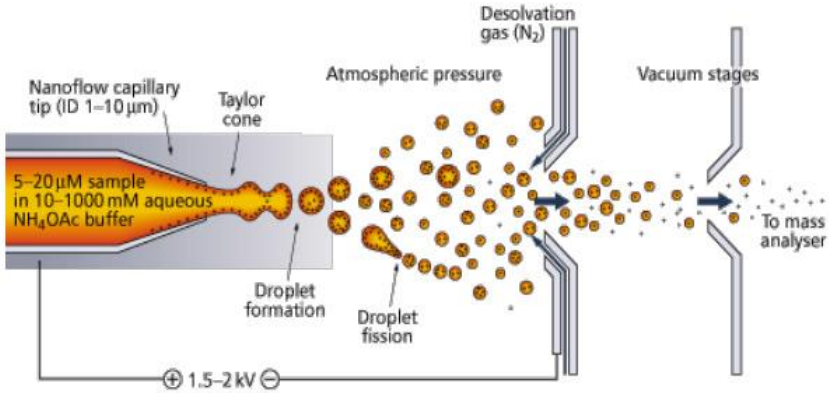


Image from https://www.thermofisher.com

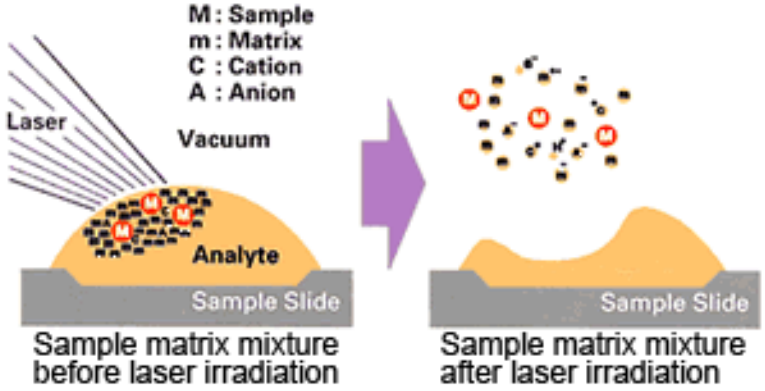


Image from https://www.shimadzu.com

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

Sample preparation for MS

Whole-cell or subcellular proteome

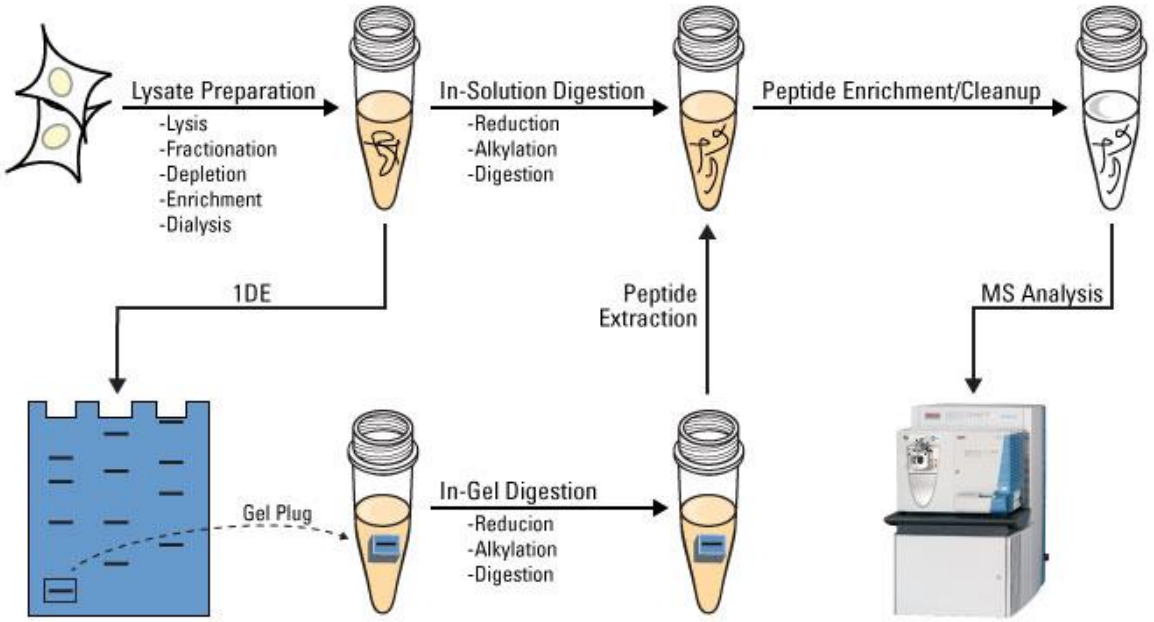


Image from https://www.thermofisher.com

- 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

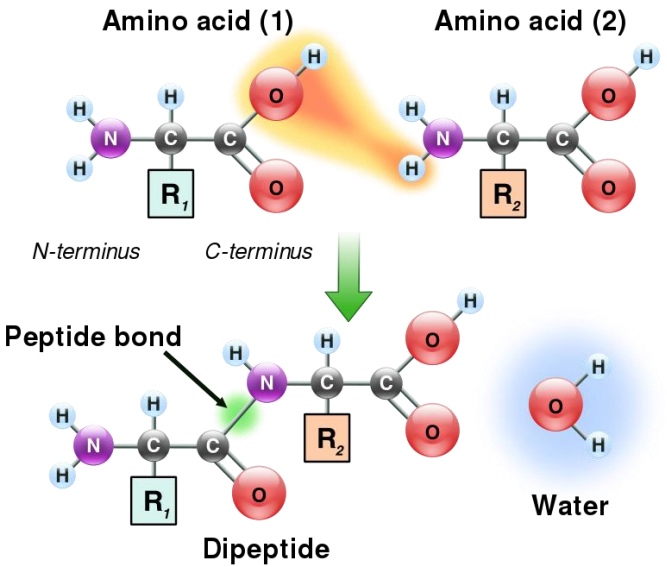
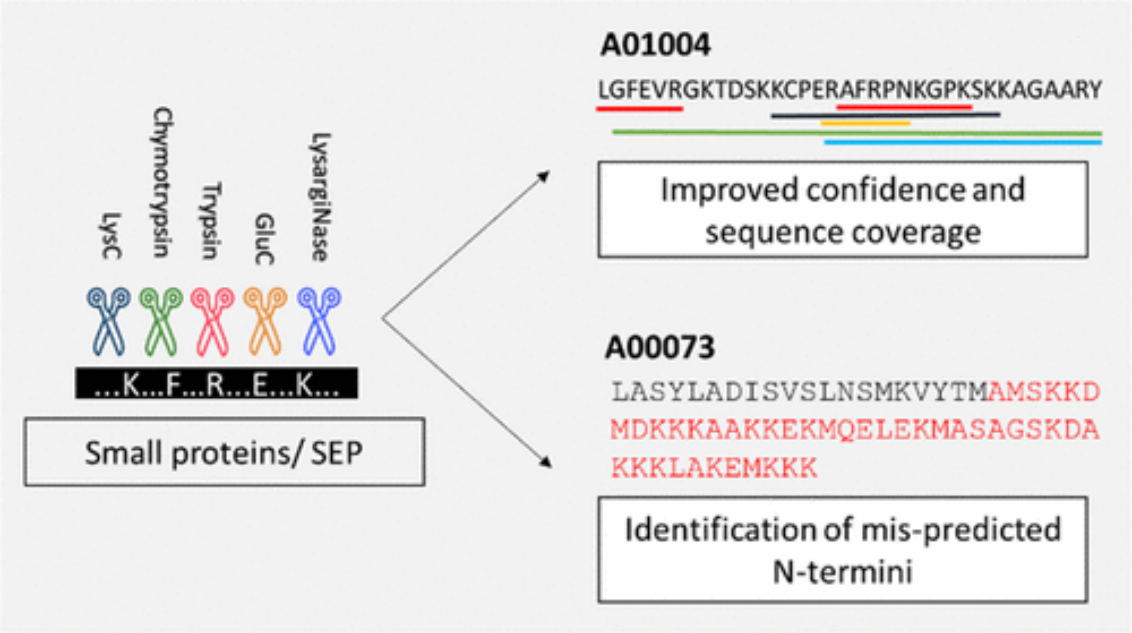


Image from
https://en.wikipedia.org/wiki/Peptide_bond

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

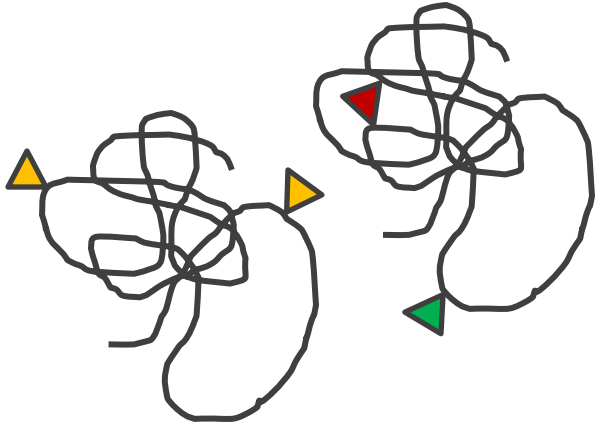
Image from MaxQuant software

Combining multiple proteases to improve coverage



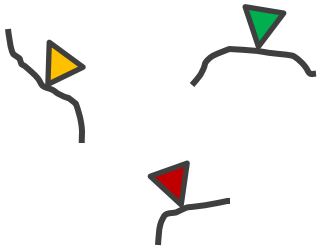
- Different proteins have different cleavage distributions
- Use of multiple enzymes ensure appropriate sizes of peptide fragments

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

Liquid chromatography (LC)

Mass spectrometry takes time to scan molecules

- Protein / peptide molecules can only be stored temporarily in MS
- Not enough time for the MS to analyze every ions
- We have to gradually feed a small number of protein / peptide species into MS at a time

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

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

- Proteins start 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

Fractionation reduces complexity of samples

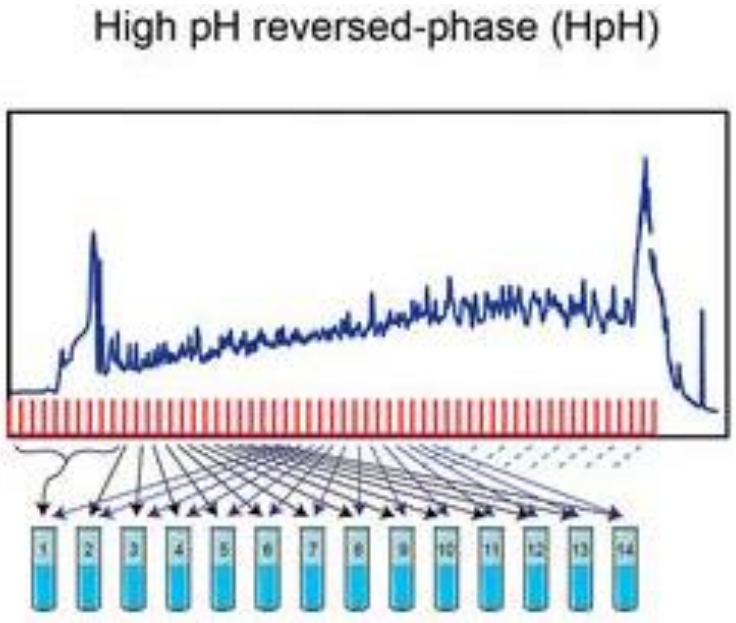


Image from Chen et al. Analyst, 2018

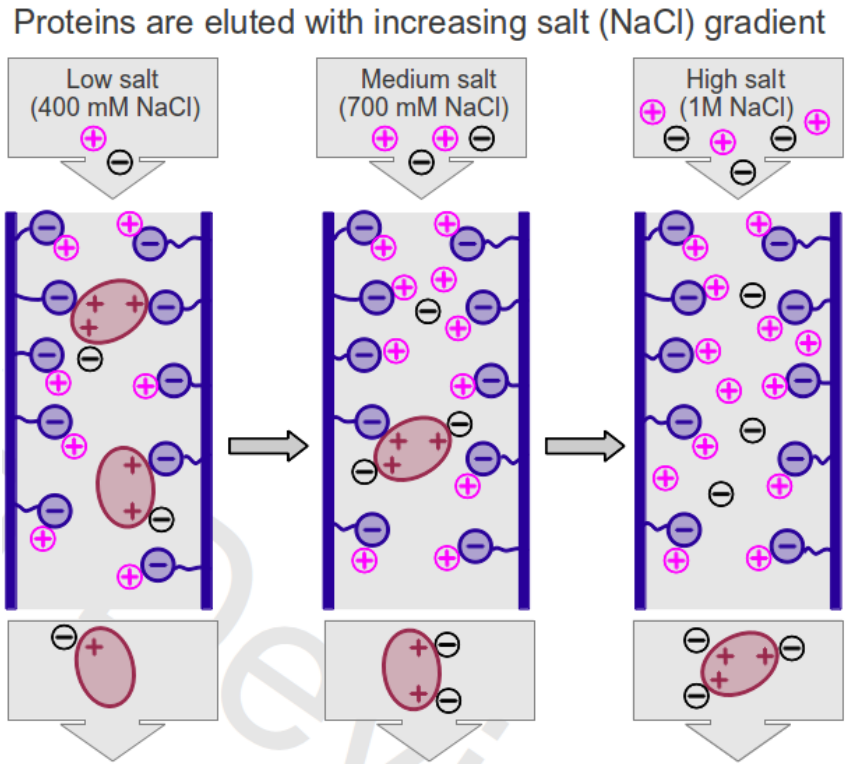


Image from http://www.reachdevices.com/

LC separates out proteins/peptides

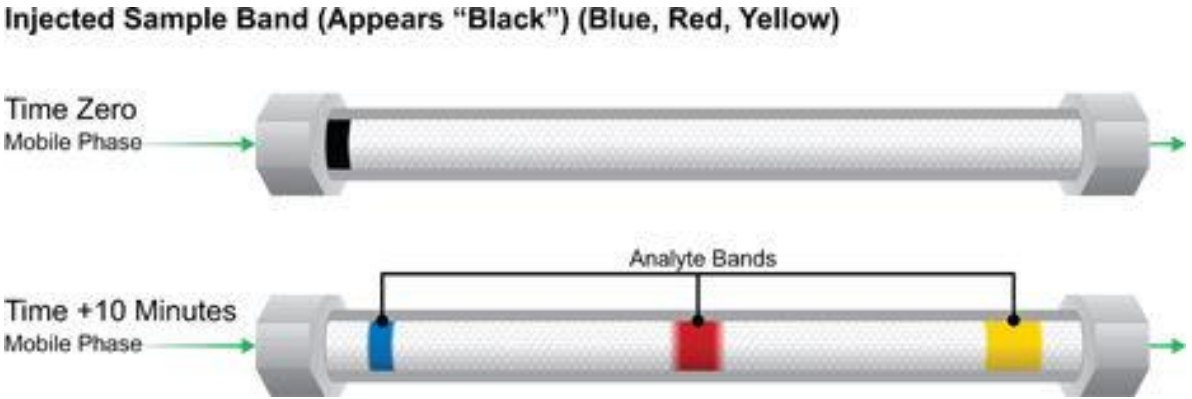
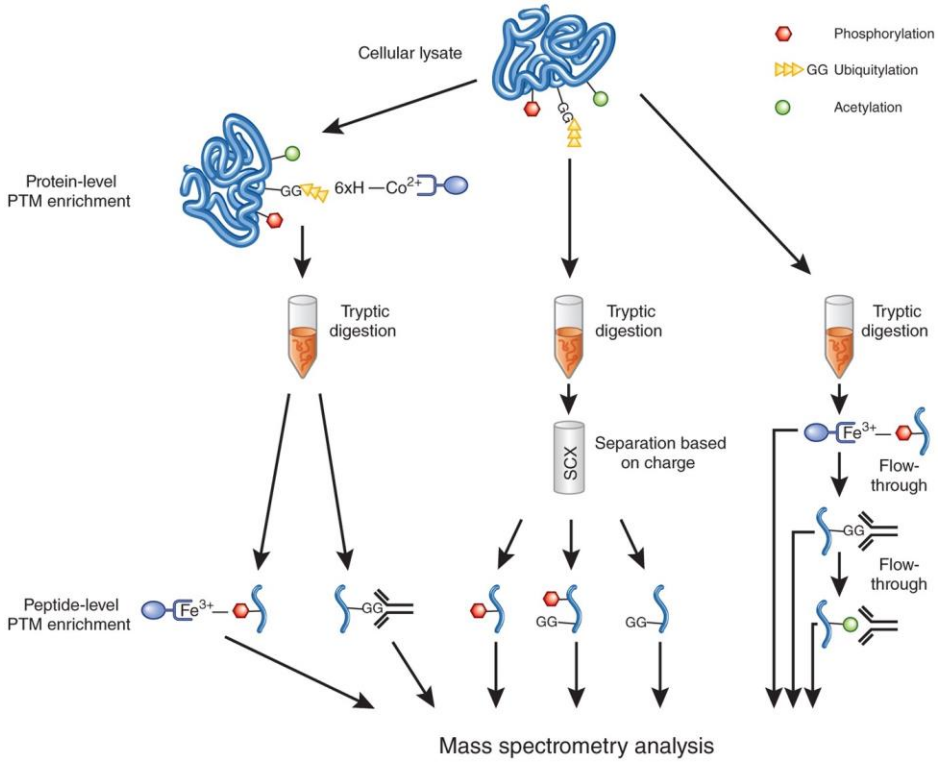


Image from https://www.waters.com

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

Enrichment of target PTM



- Specific PTM can be enriched to focus the MS analysis
  - Phosphorylation
  - Glycosylation
- Enrich phosphorylated peptides to study kinase activity
- Depend on the biology

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

Proteomics can be time-consuming

- Let’s say we have blood samples from 3 patients and 3 controls
- To analyze every protein, each sample is divided into 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

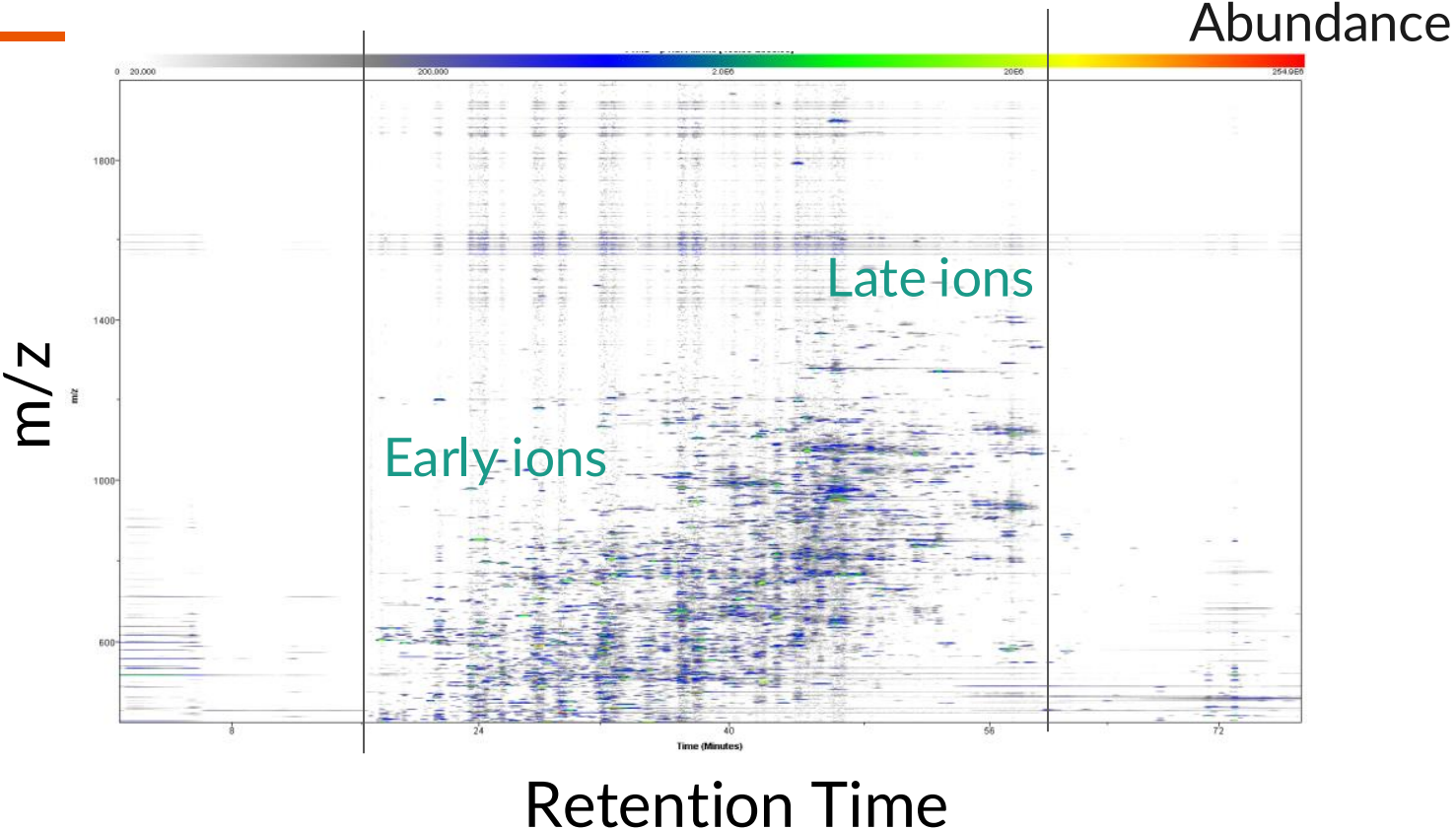
MS/MS process

The journey of a peptide

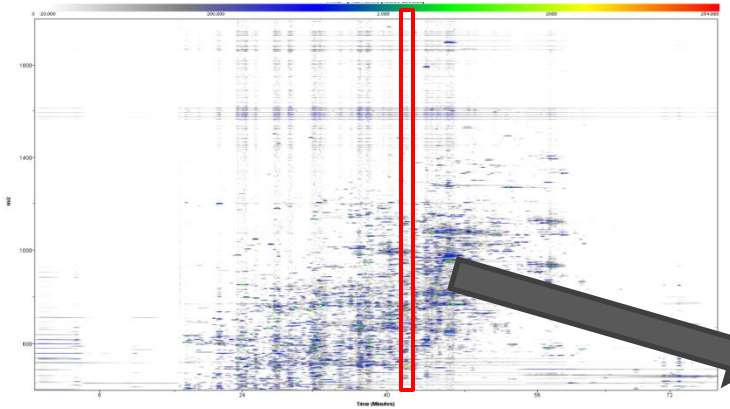
Image from https://en.wikipedia.org/wiki/Tandem_mass_spectrometry

- Masses of intact protein/peptide ions are measured
- Some ions (e.g., high abundances) are isolated for further analysis
- Fragmentation and scanning of selected ions

The profile of peptide ions as they elude from LC

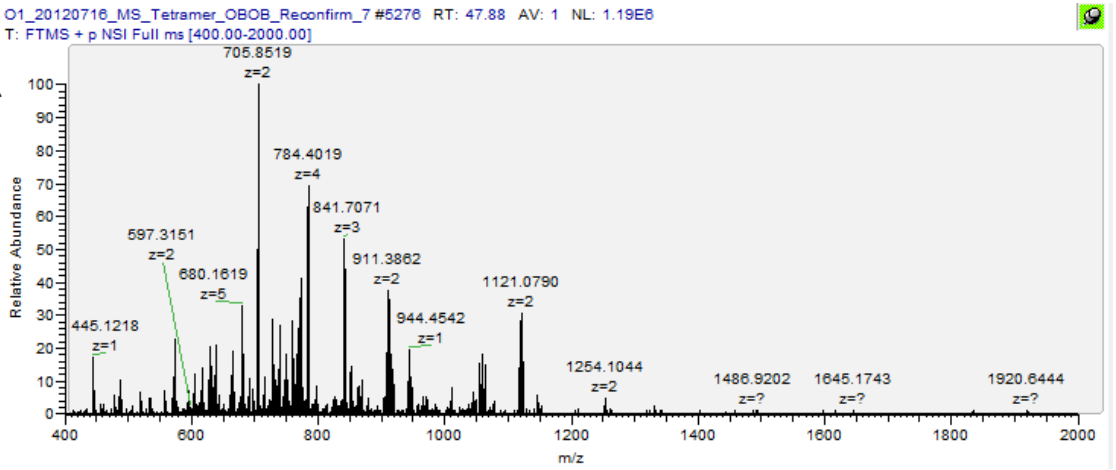


MS1 scans capture proteins/peptides at each time point

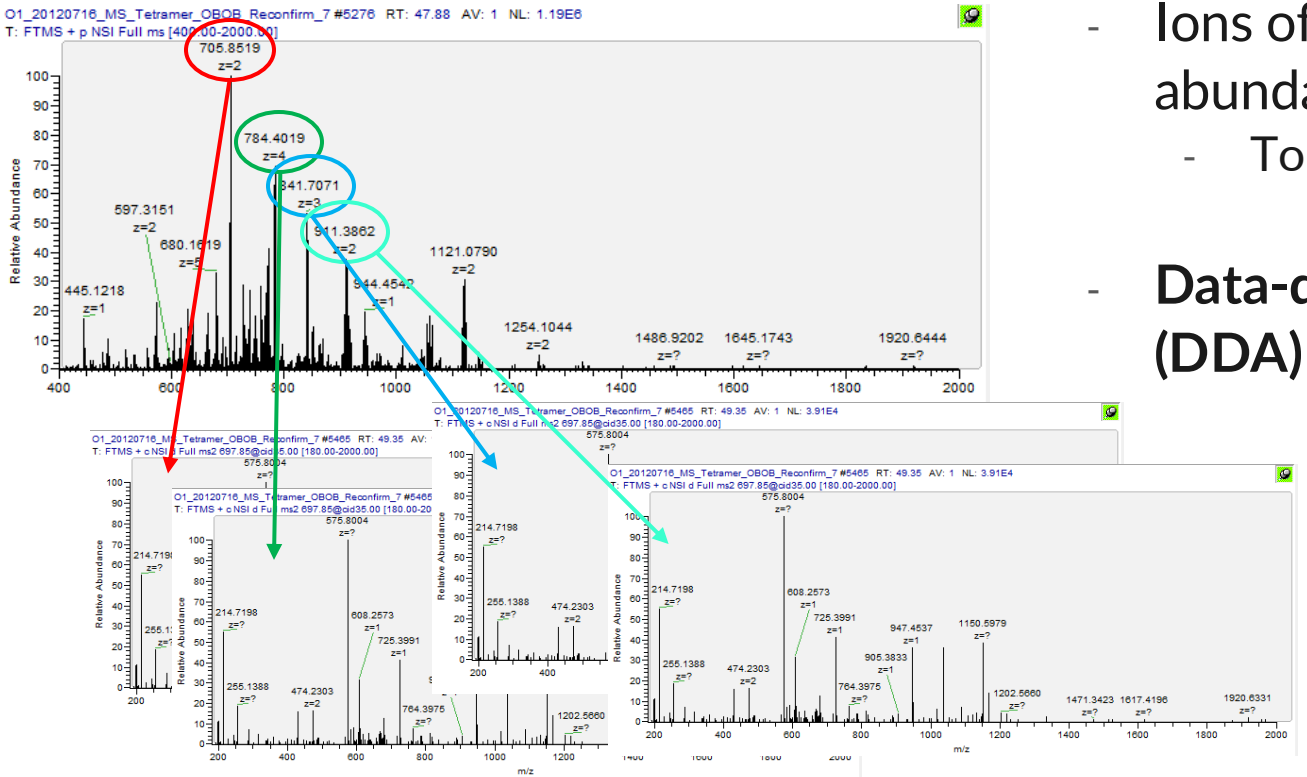


- At a time, a number of peptide ions were injected into MS

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

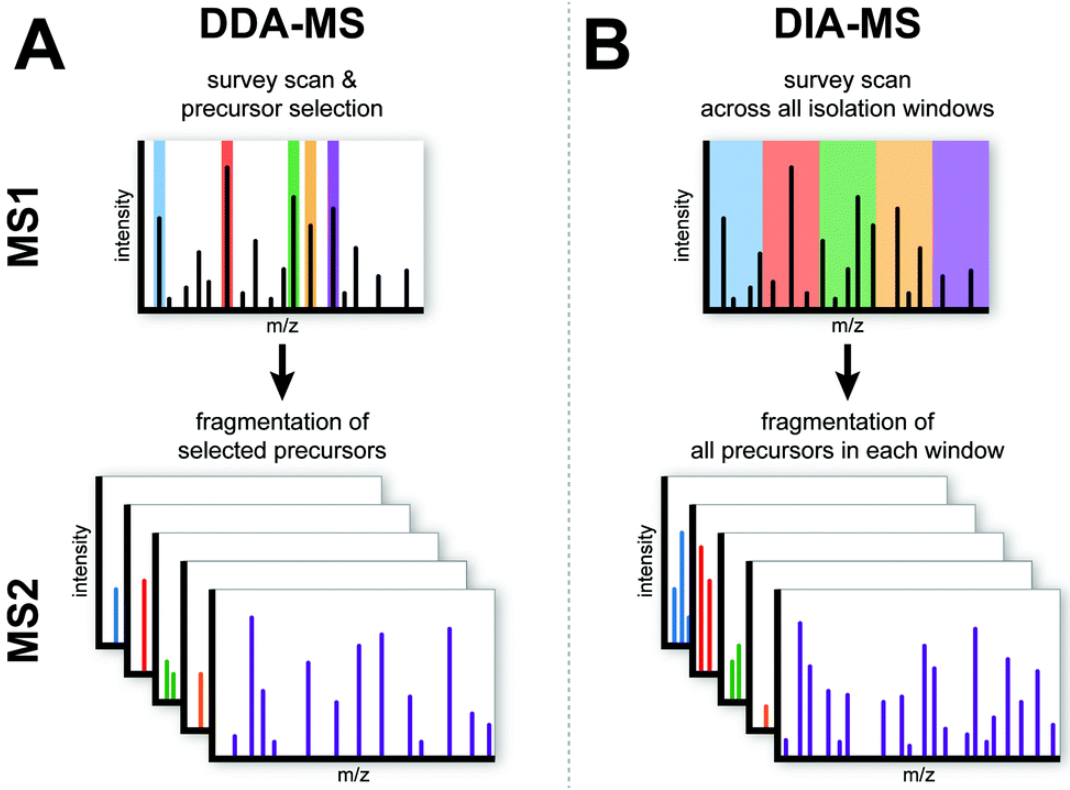


MS/MS analysis of isolated proteins/peptides



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

Data-independent analysis (DIA)



- Scan all ions at once (proteins/peptides and contaminants)
- Good coverage/sensitivity
- Complex MS/MS signals

https://medium.com/@zahraelhamraoui1997/what-are-dia-and-dda-and-how-does-the-data-looks-like-e18a3a532142

MS/MS (tandem MS)

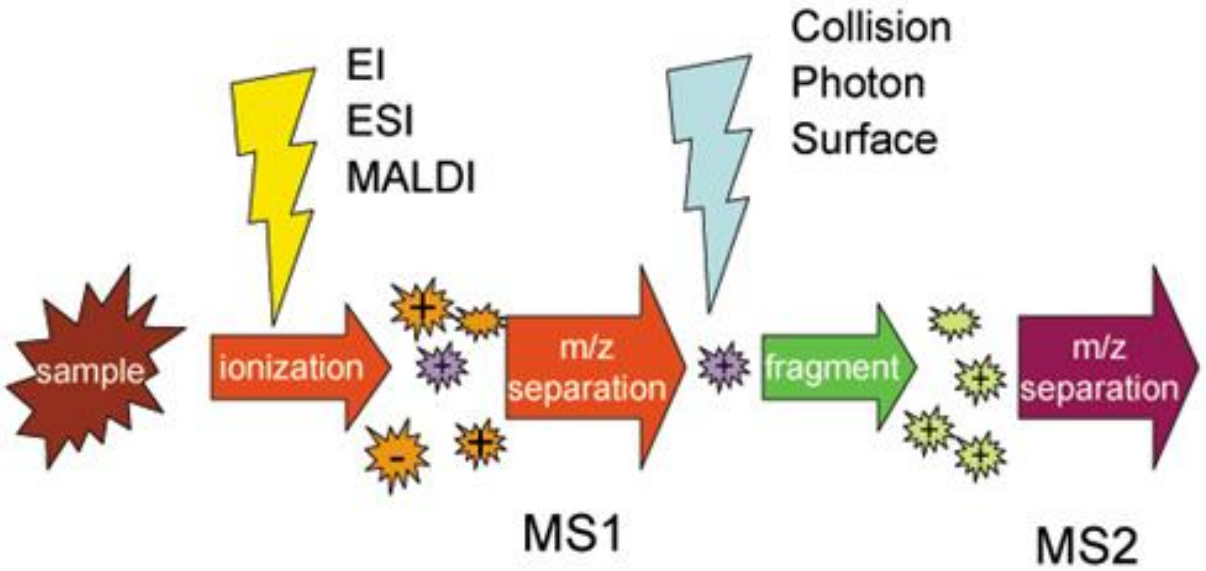


Image from https://en.wikipedia.org/wiki/Tandem_mass_spectrometry

- Fragmentation of each peptide into smaller characteristic ions

Fragmentation of protein/peptide ions

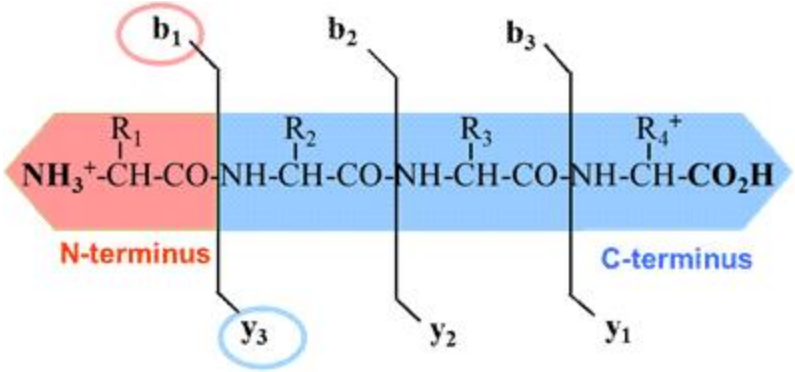


Image from https://www.molgen.mpg.de

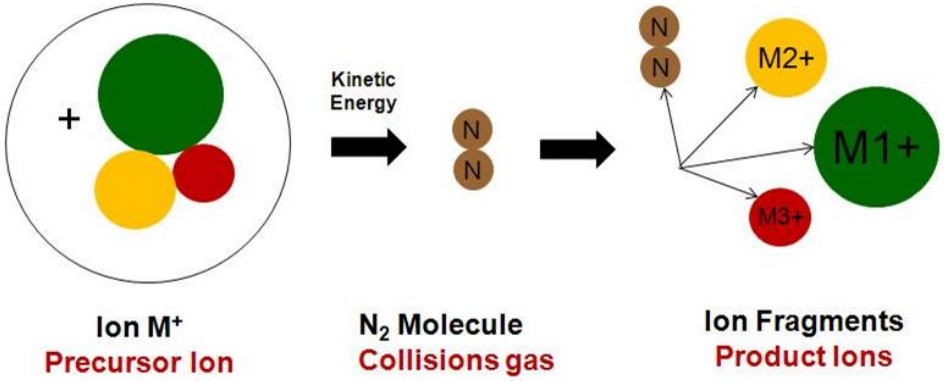
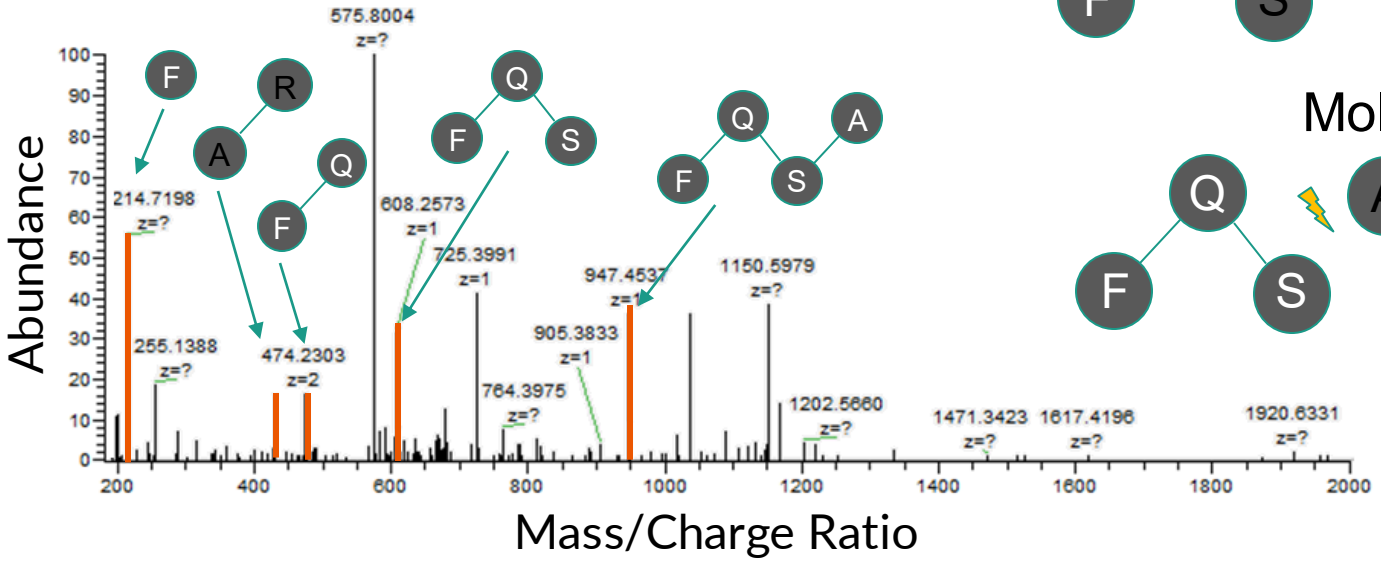


Image from https://www.biologie.hu-berlin.de

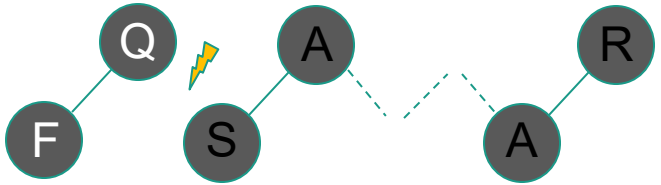
- Accelerate peptide molecules to collide inert gases, such as nitrogen
- Collision at the **right energy level** breaks a peptide bond
  - Generate fragments of peptides: A, AB, ABC, ABCD, ...

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

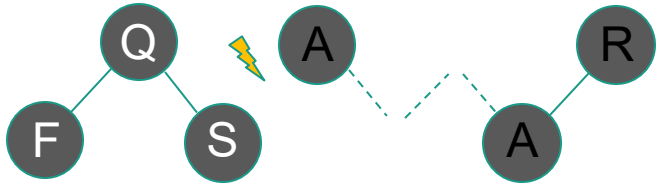
FQSADETGQDLVNANHEAR



Molecule #1

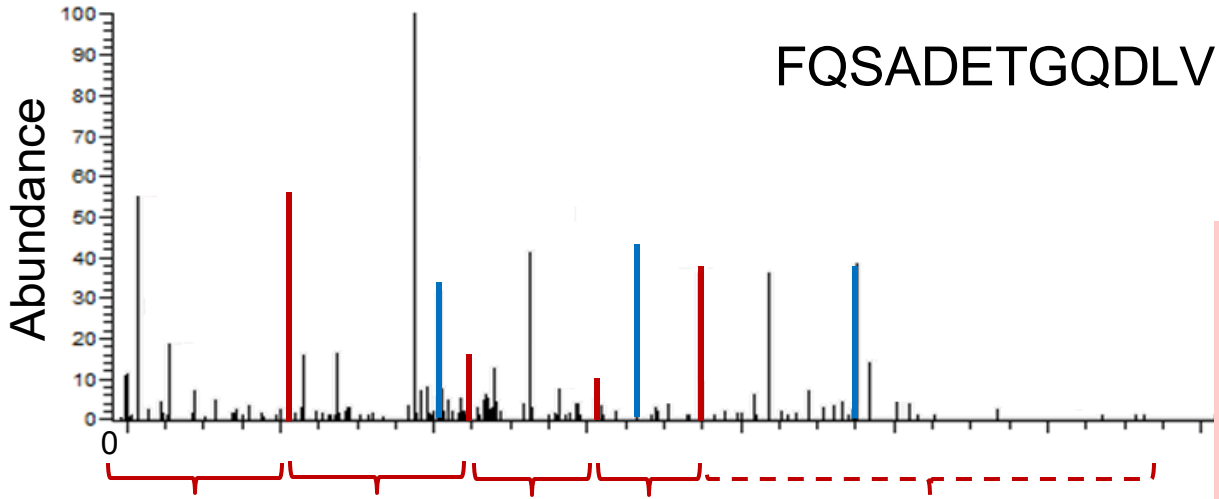


Molecule #2



Peptide sequencing

Decoding MS/MS spectra using known AA masses



FQSADETGQDLVNANHEAR

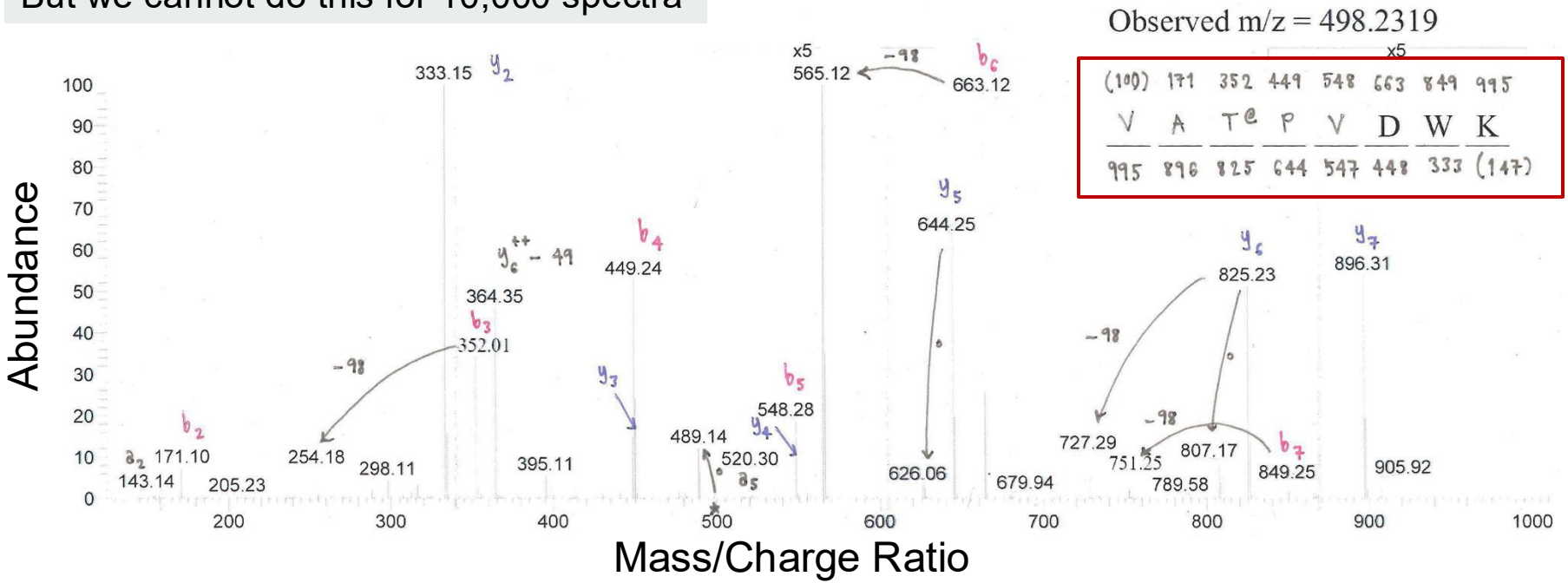
147.07	F
128.06	Q
87.03	S
…	…

Delta Mass: 147.07 128.06 87.03 71.04 …

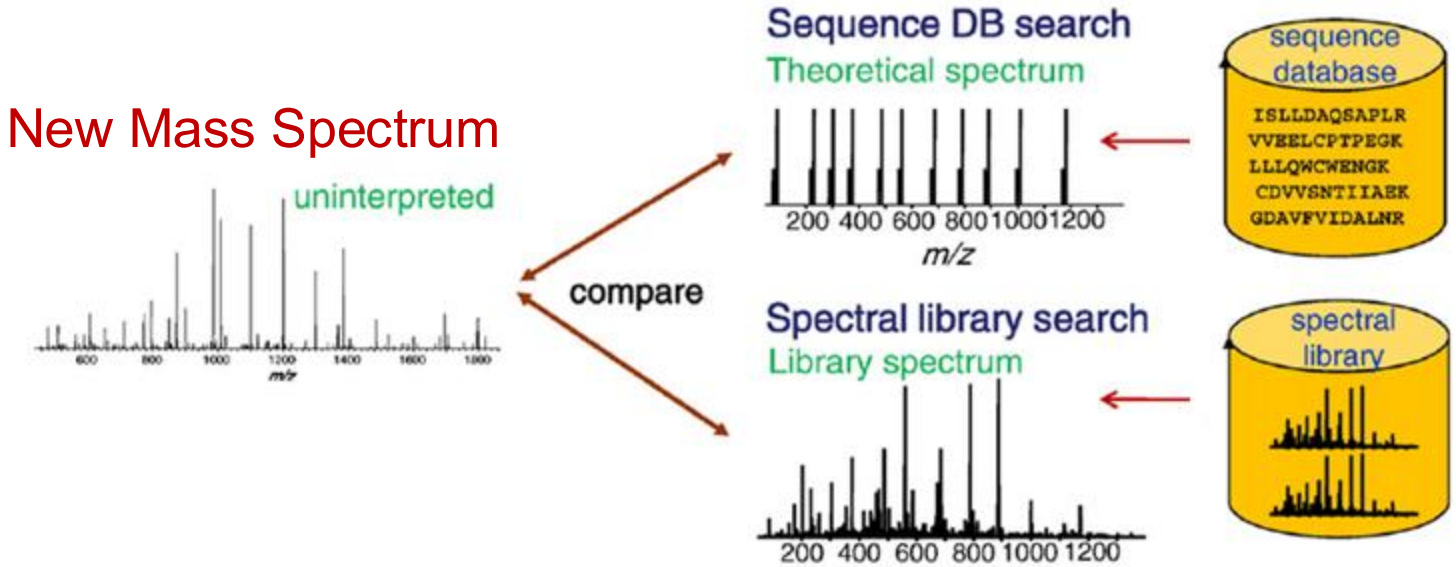
Delta Mass: 227.14 129.04 137.06

Manual peptide sequencing is possible

But we cannot do this for 10,000 spectra



Database search approaches



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

- Compare observed spectra to known spectra or theoretical spectra

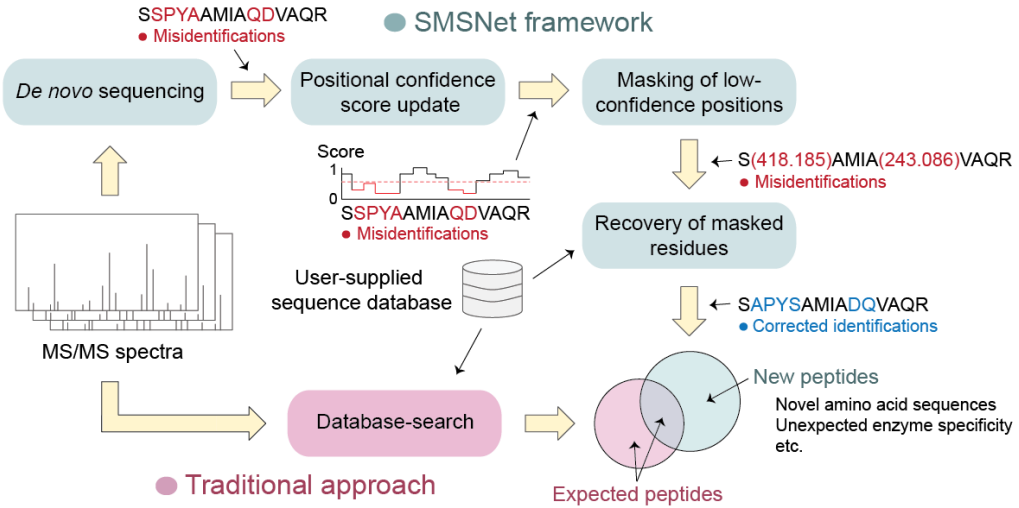
Limitation of database search

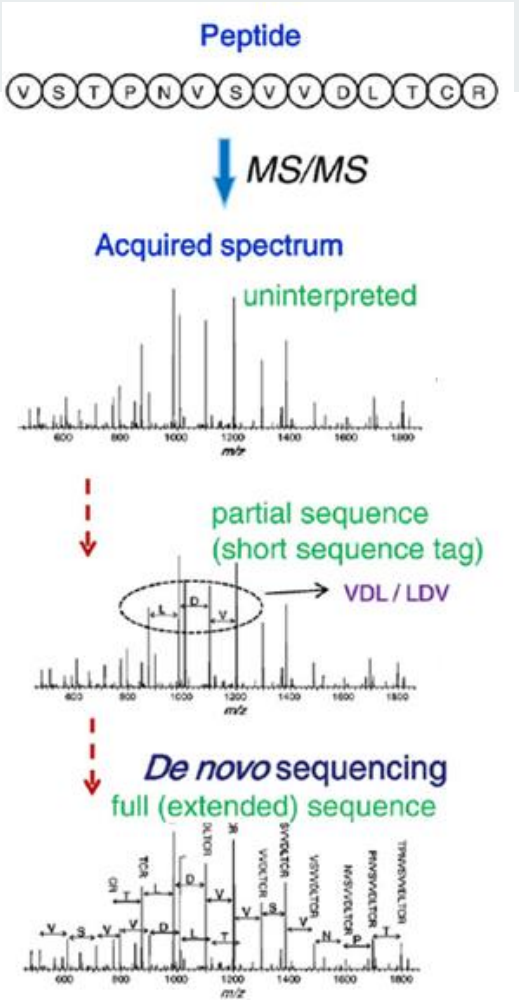
- **Identify exact match only**
  - Unlike BLAST: Peptides are too short to align reliably
  - Doesn't work on data from new species
  - Cannot identify cancer mutations

- **Searching for PTM explodes possibilities**
  - FQSADETMAR with oxidation and phosphorylation = 8 possibilities
  - PTM changes amino acid mass → changes MS/MS spectra

De novo peptide sequencing

- Directly deducing amino acid sequences
- Identify partial sequence
- Rely on deep learning (AI)





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

Protein expression quantification

Peptide abundance from MS1 spectrum

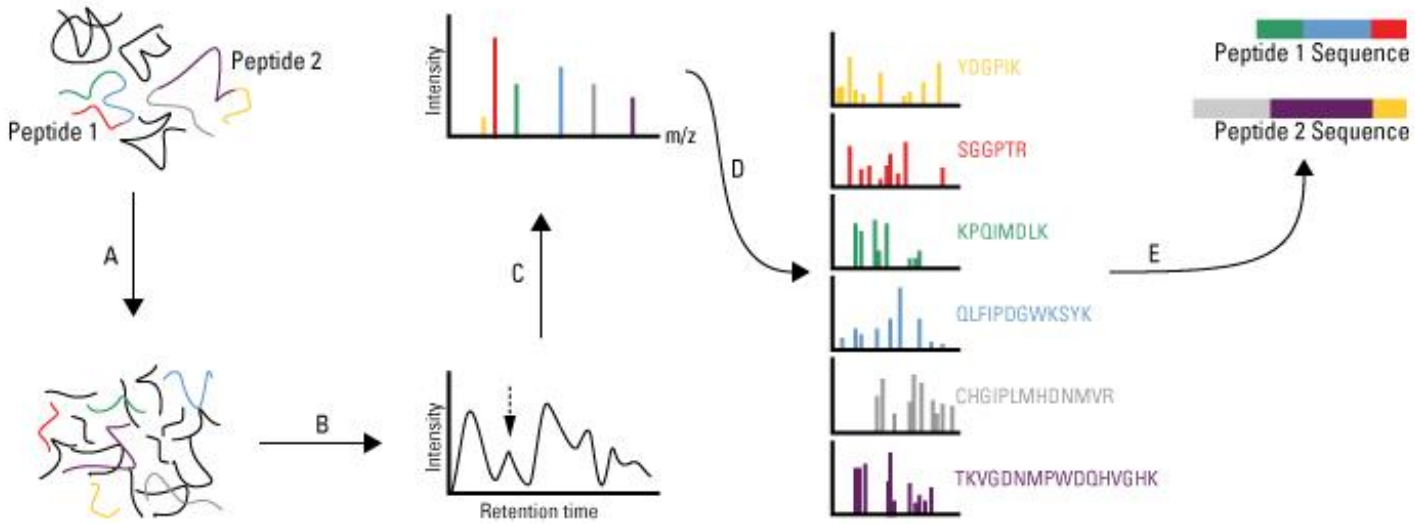


Image from http://fields.scripps.edu/yates/wp/

- Intensity values in MS1 spectrum reflect peptide abundances
- Aggregate abundance for peptides derived from the same protein

Protein quantification

- Spectral count: Number of MS/MS of peptides derived from a protein
- Intensity of the peptide in MS1 spectrum
  - Different ionization efficiencies ← similar to PCR amplification bias
- How to quantify proteins with multiple peptides?
  - Sum of all peptides
  - Sum of top *N* peptides
- Intensity values (ion counts) are log-normally distributed

Isoform issues

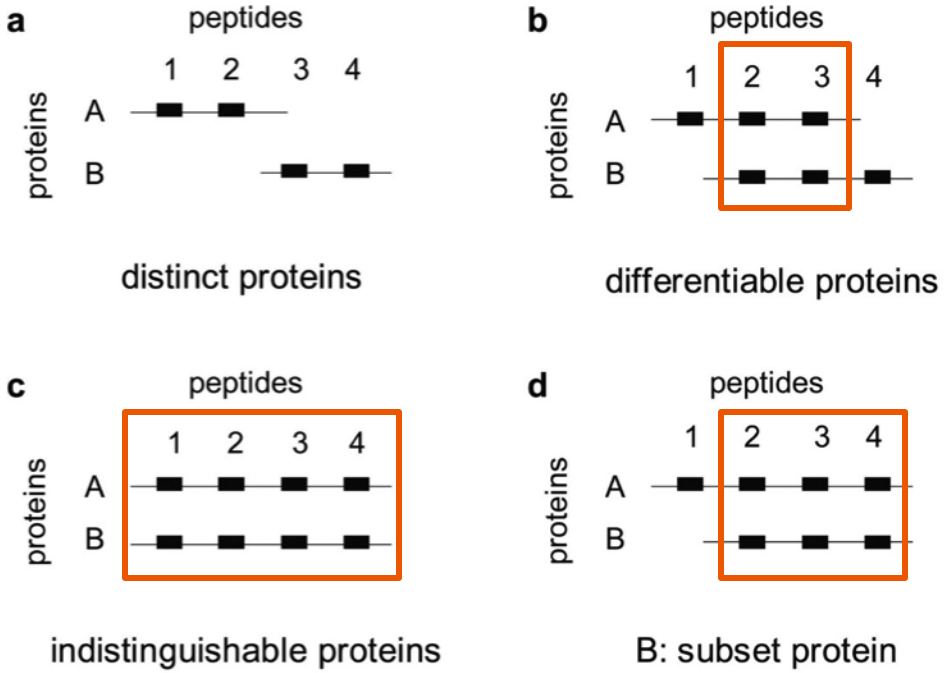
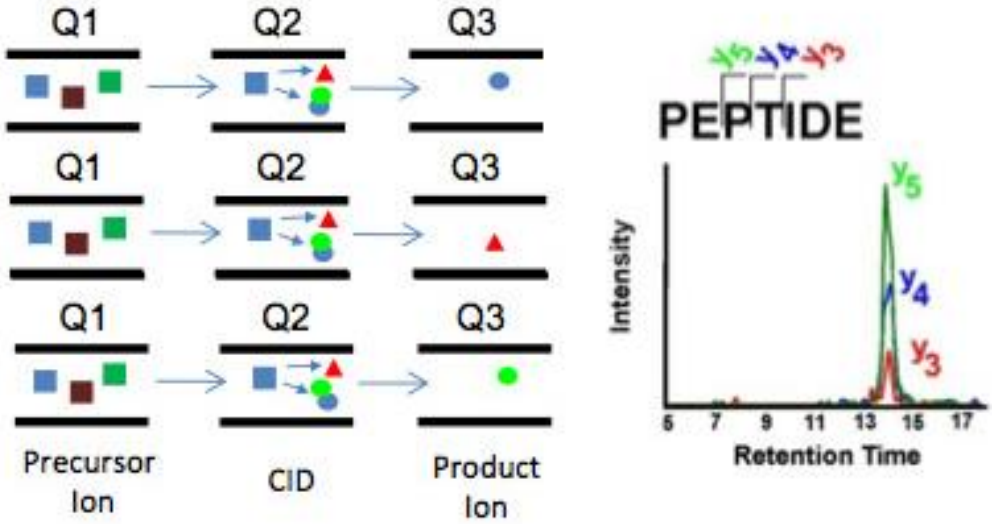


Image from https://www.inf.fu-berlin.de

- **Idea 1**: Assign shared peptides to isoforms with the highest abundance
- **Idea 2**: Distribute shared peptides proportionally
- **Idea 3**: Disregard shared peptides
- Report indistinguishable proteins as a **protein group**

Multiple reaction monitoring (MRM)

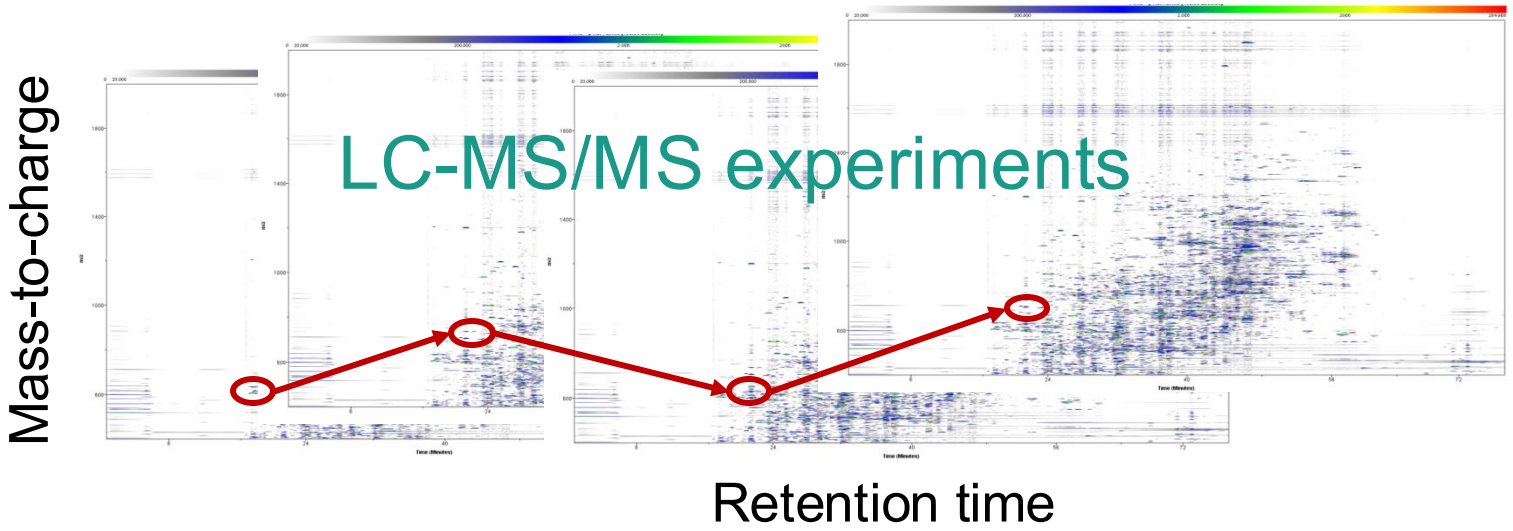


https://medicine.yale.edu/keck/proteomics

- Targeted quantification peptides via specific fragmented ions
- **Peptide mass, retention time, and fragment ions must be known**

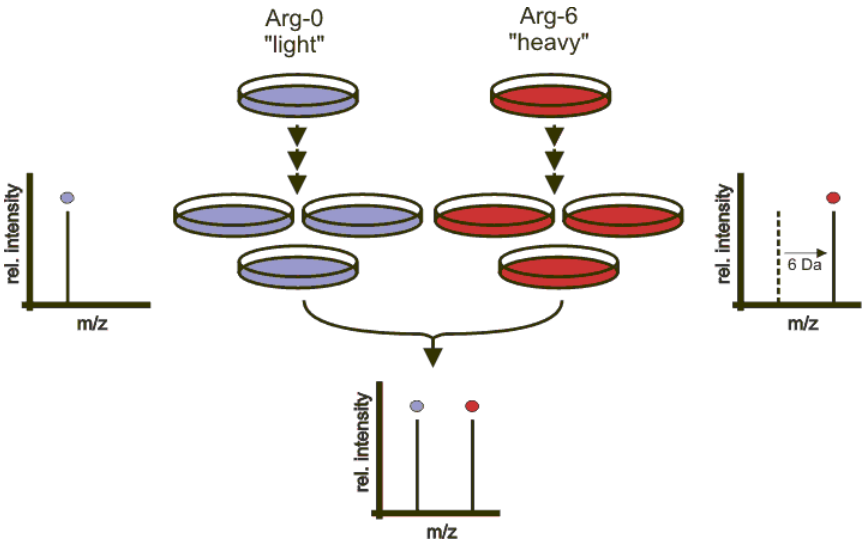
Comparative proteomics

Label-free comparison



- Simply perform multiple LC-MS/MS runs and match the observed m/z and retention time patterns ← Cannot account for all technical biases

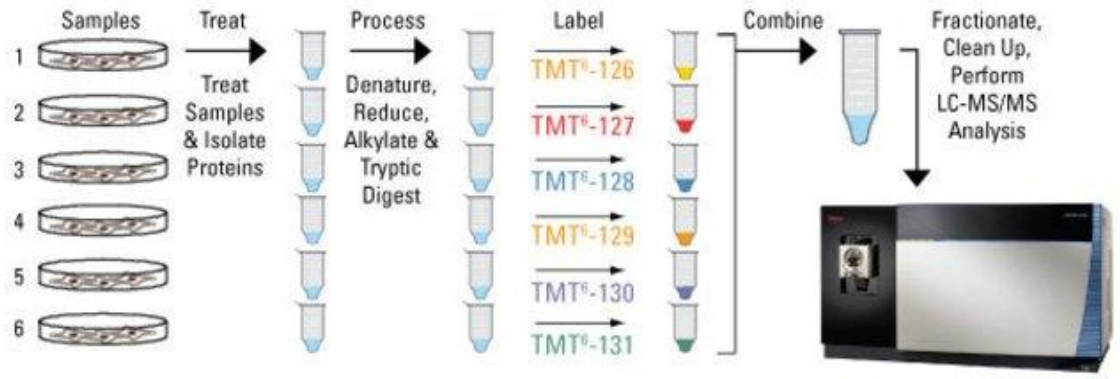
Stable isotope labeling (SILAC)



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

- Feed cell cultures with heavy and light isotopes
- Mix samples and perform a single LC-MS/MS run
  - No technical bias
- Look for a pair of peptide ions with the expected mass shift and similar MS/MS spectra

Tandem mass tag (TMT)



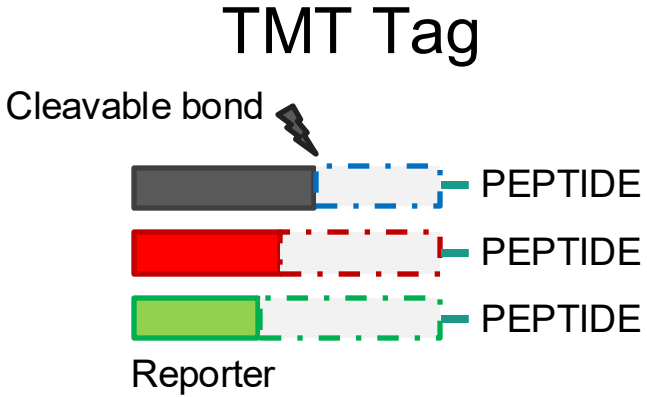


Image from https://www.thermofisher.com

- TMT molecules are added to the *N*-terminus of peptides
- All TMT species have the **same total mass**
- Different TMT species have **different reporter mass** (126 Da, 127 Da, ...)
  - Reporter will be separated from the peptide molecule in tandem MS

Comparative quantification via TMT reporter ions

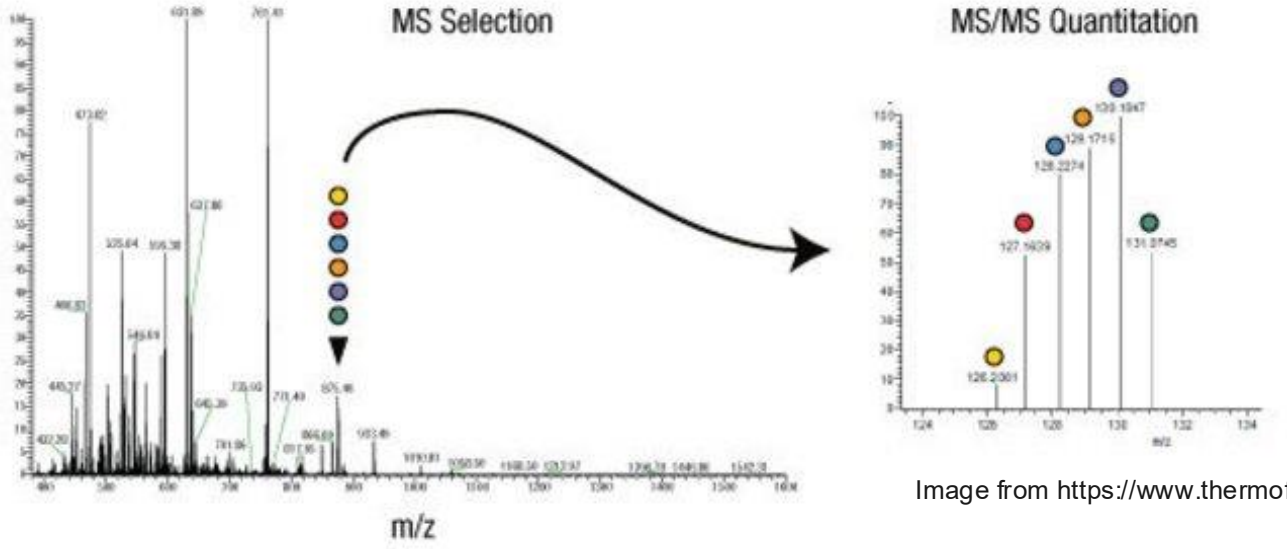


Image from https://www.thermofisher.com

- The same peptides from multiple samples will be isolated together in MS1
- Different sample is associated with different **reporter ion mass**

Structural proteomics

Chemical crosslinking

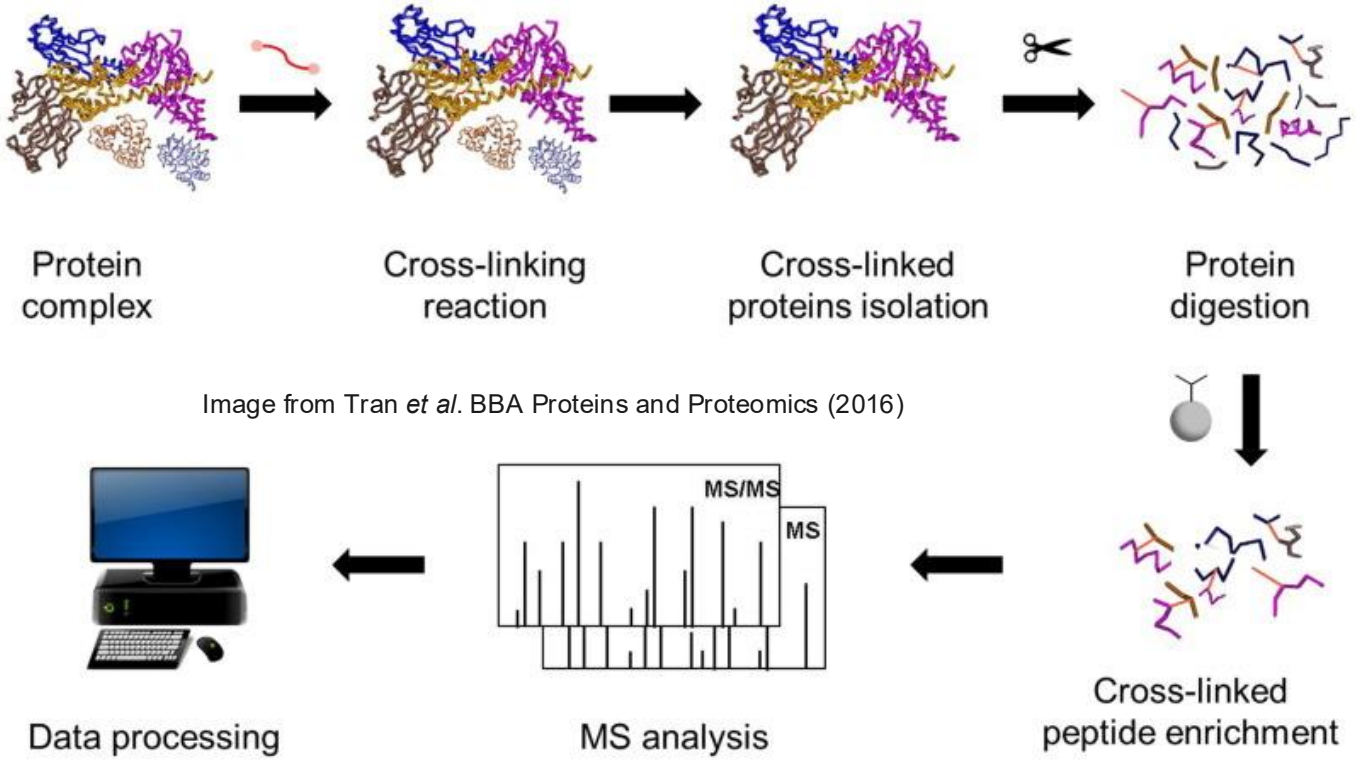


Image from Tran *et al*. BBA Proteins and Proteomics (2016)

- Induce cross-linking between proximal AAs
- Identify linked residues with MS
- Use as distance constraint for modeling

Hydrogen-deuterium exchange

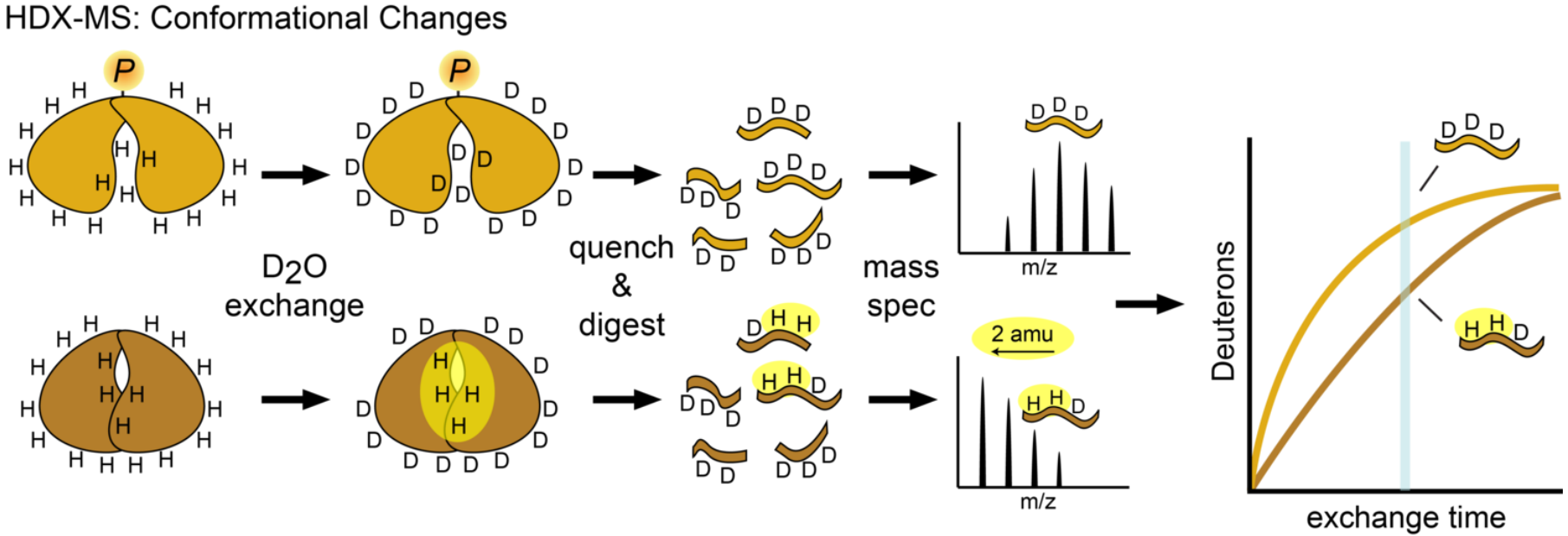
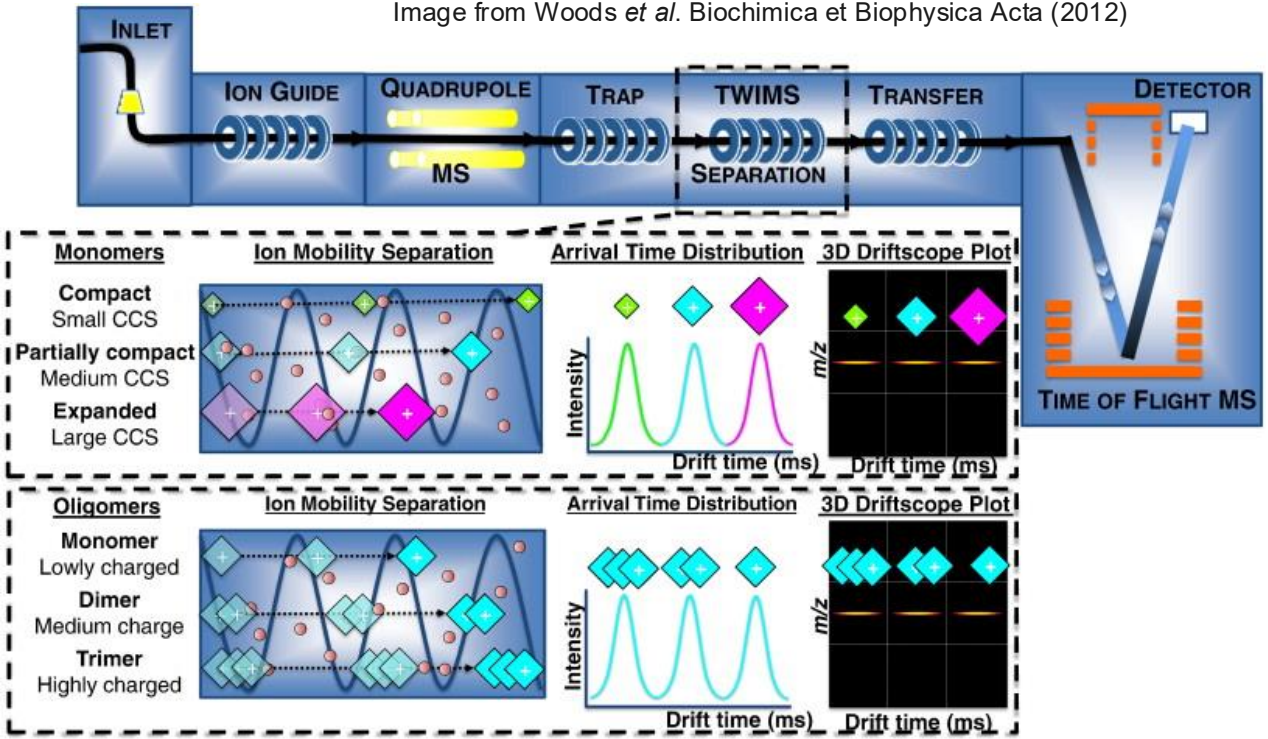


Image from https://www.underbakkelab.org/techniques

- Deuterium incorporation = accessible parts of the structure

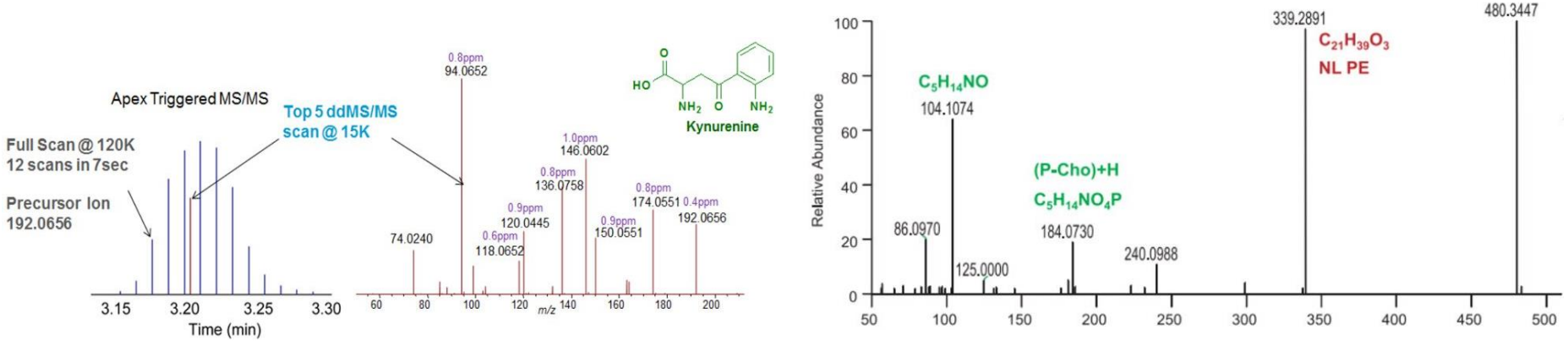
Time-of-flight distinguish protein conformations

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



- Monitor changes in molecular structure, such as cross-sectional surface area
- Result in slower or fast time of flight in the MS

MS of non-peptide molecules: lipids, metabolites



- MS can analyze any molecules
  - Choosing the right collision energy level to break chemical bonds
  - Comparing to a database of MS/MS spectra of known compounds
- **Proteomics is easier because peptides are polymers with limited monomer**

Any question?

- See you next time