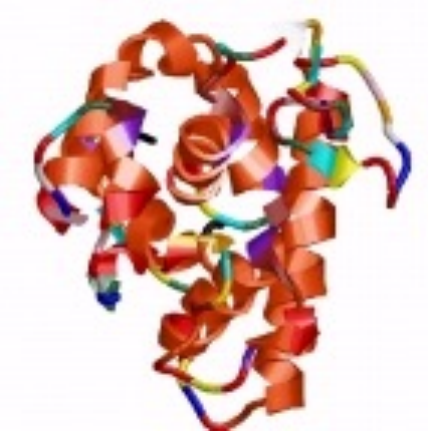
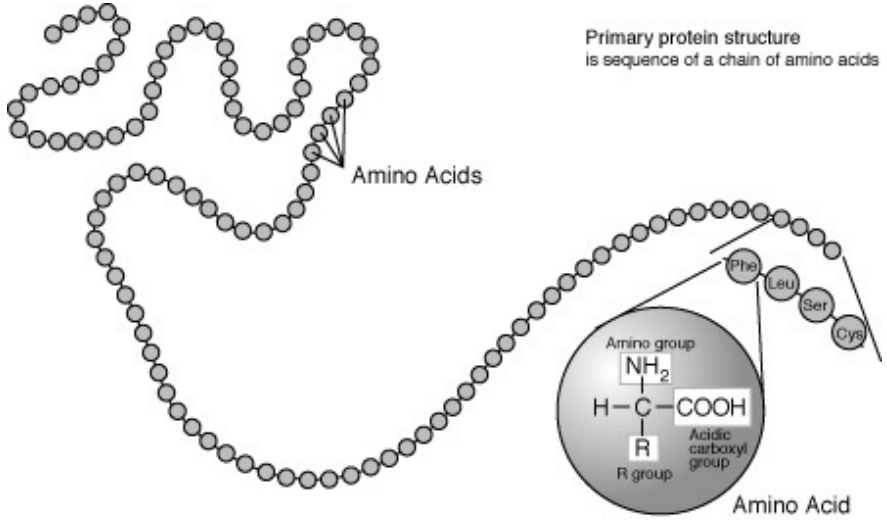
Proteomics and Mass Spectrometry

Outline

- Motivation
- Mass spectrum of a peptide
- Mass spectrometer
- Bottom-up proteomics

Protein





Primary structure is a sequence.
20 frequent amino acids.
Fold into a complex 3D structure.

A Protein Sequence

>P02769|ALBU_BOVIN Serum albumin - Bos taurus (Bovine).

MKWVTFISLLLLFSSAYSRGVFRRDTHKSEIAHRFKDLGEEHFKGLVLIAFS
QYLQQCPFDEHVKLVNELTEFAKTCVADESHAGCEKSLHTLFGDELCKVASL
RETYGDMADCCEKQEPERNECFLSHKDDSPDLPKLKPDPNTLCDEFKADEKK
FWGKYLYEIARRHPYFYAPELLYYANKYNGVFQECCQAEDKGACLLPKIETM
REKVLASSARQRLRCASIQKFGERALKAWSVARLSQKFPKAEFVEVTKLVTD
LTKVHKECCHGDLLECADDRADLAKYICDNQDTISSKLKECCDKPLLEKSHC
IAEVEKDAIPENLPPLTADFAEDKDVCKNYQEAKDAFLGSFLYEYSRRHPEY
AVSVLLRLAKEYEATLEECCAKDDPHACYSTVFDKLKHLVDEPQNLIKQNCD
QFEKLGEYGFQNALIVRYTRKVPQVSTPTLVEVSRSLGKVGTRCCTKPESER
MPCTEDYLSLILNRLCVLHEKTPVSEKVTKCCTESLVNRRPCFSALTPDETY
VPKAFDEKLFTFHADICTLPDTEKQIKKQTALVELLKHKPKATEEQLKTVME
NFVAFVDKCCAADDKEACFAVEGPKLVVSTQTALA

Fundamental Questions

- **Identify, sequence, and quantify** all the proteins in a biological sample.
  - Identification: determine which proteins in a database present in the sample.
  - Sequencing: determine the amino acid sequence without needing a database.
  - Quantification: determine the quantity change of each protein under two different biological conditions.

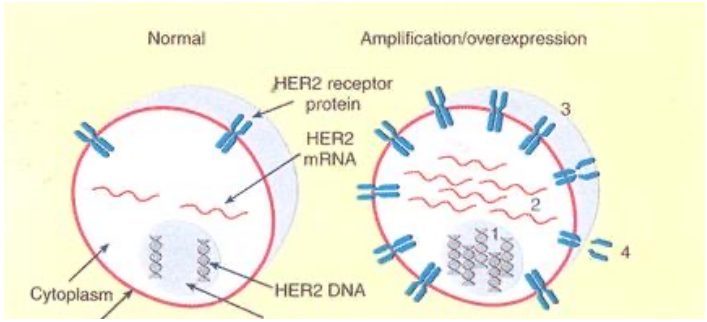
> Human Ig heavy chain variable region

QVQLVQSGAEVKKPGASVKVSCKASGYTFTGYYMH
WVRQAPGQGLEWMGRINPNSGGTNYAQKFQGRVTS
TRDTSISTAYMELSRLRSDDTVVYYCAR

Why Bother? – Example: Biomarker

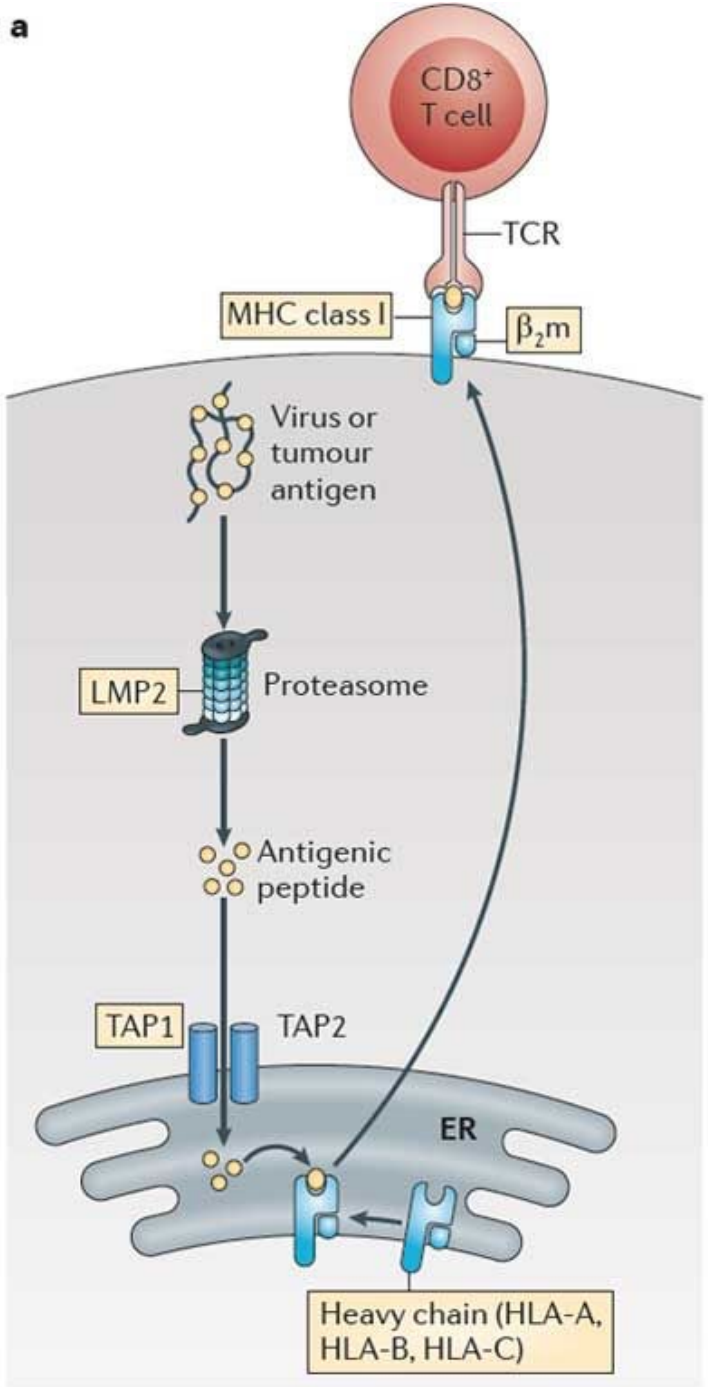
HER2-positive breast cancer is a breast cancer that tests positive for a protein called human epidermal growth factor receptor 2 (HER2), which promotes the growth of cancer cells.

Biomarker



> “HER2-positive breast cancers tend to be **more aggressive** than other types of breast cancer. They're also less responsive to hormone treatment. However, treatments that **specifically target HER2** are very effective.”

MAYO CLINIC

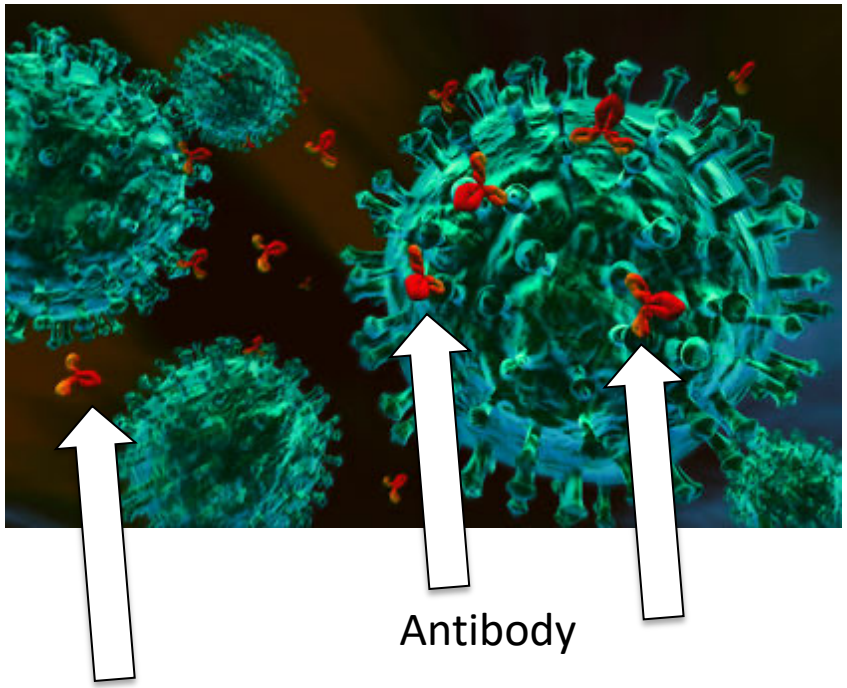


Example: Immunopeptides

- Tumor or infected cells “present” some abnormal peptides at the MHC on cell surface.
- CD8+ T cells (aka T killer cells) recognize the abnormal peptides and kill the cell.
- An actively pursued method in immunotherapy is to identify/predict the peptides presented and train the T cells to target them.
- Mass spec is the best tool to identify these peptides.

https://nanolive.ch/immuno-oncology-t-cells-targeting-and-killing-cancer-cells/

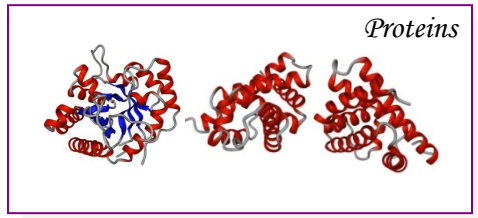
Example: Antibody



- Our immune system produces antibodies to bind to invasive pathogens (virus etc.) and cancer neoantigen.
- Antibody's amino acid sequences determine the binding target (antigen).
- Some people have stronger immune system. E.g. HIV elite controller.
- Sequencing their antibodies provides a new way to discover antibody drugs.

Technology Overview

Biology sample

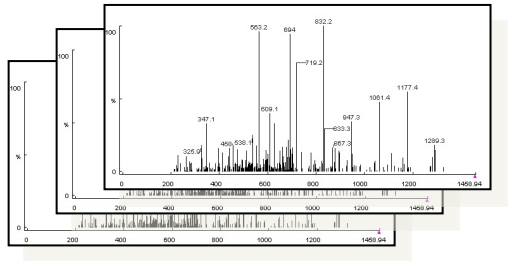


Sample Preparation

Peptides

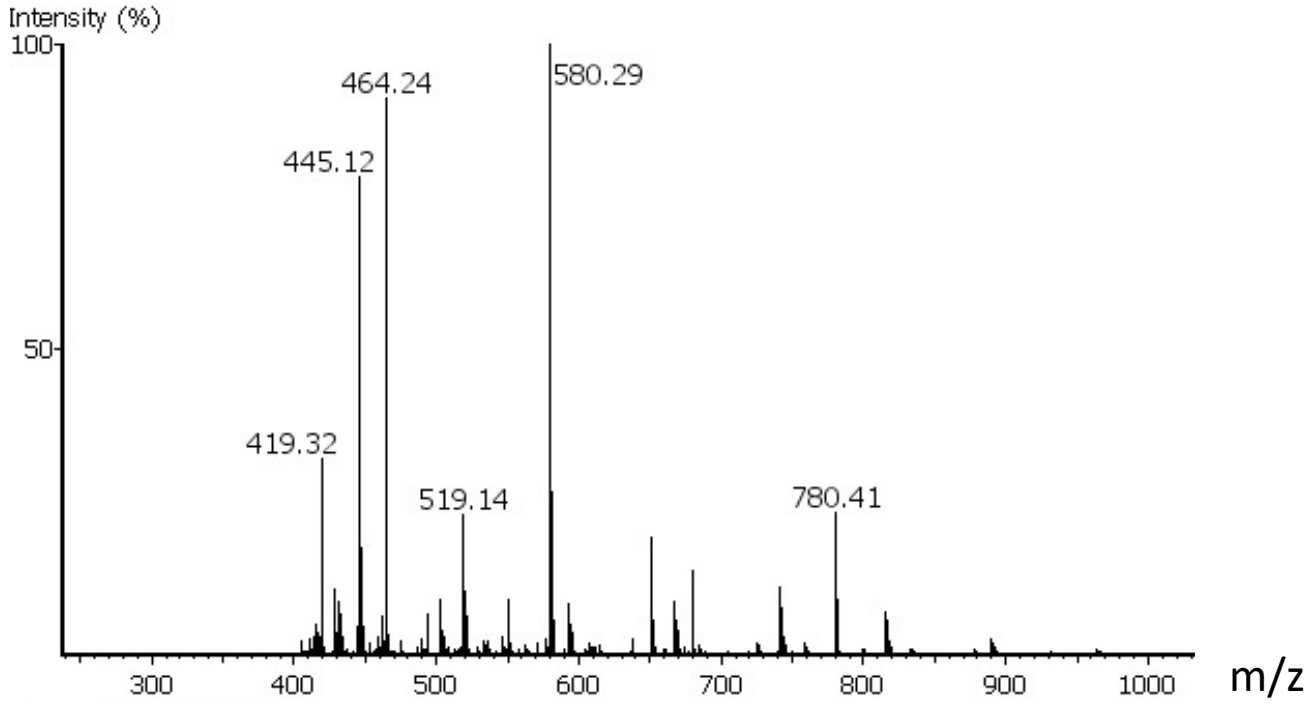
Mass Spectrometry

What're in the sample?



Mass Spectral Data

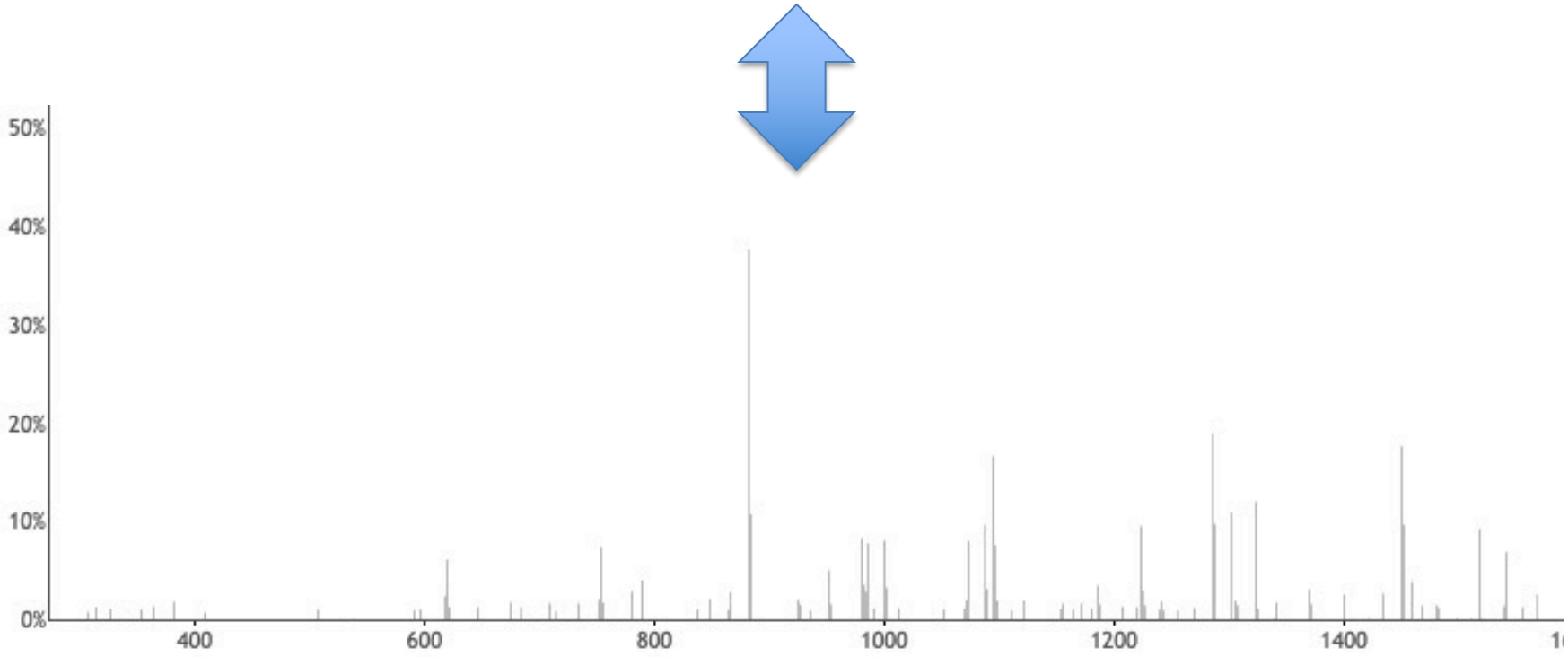
A Mass Spectrum



- Each peak indicates the detection of a particular type of ion (electrically charged molecule) with the corresponding mass to charge ratio.
- Mass of an ion can also be derived if the charge state (z) of an ion is known, which is achievable.
- A typical proteomics experiment produces tens of thousands spectra per hour.

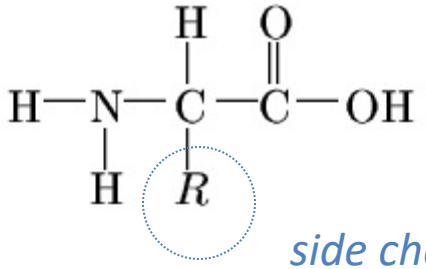
Tandem Mass (MS/MS) Spectrum of a Peptide

VVVDVSHEDPEVKF



Amino Acids

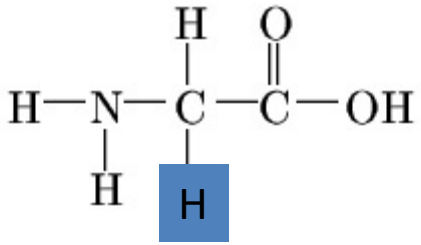
- There are 20 amino acids. All have the same basic structure but with different side chains:



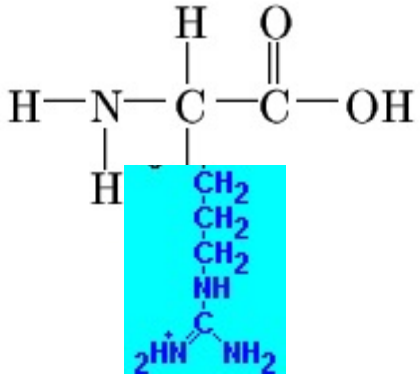
side chain group

C: Carbon
H: Hydrogen
O: Oxigen
N: Nitrogen

- Examples:



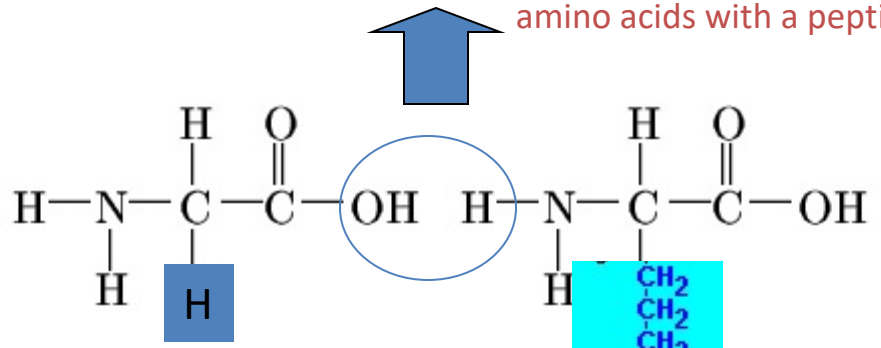
Glycine, or Gly, or G



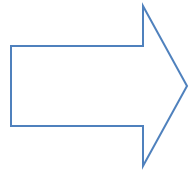
Arginine, or Arg, or R

Peptides and Proteins

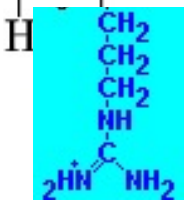
A H2O is lost when connecting two amino acids with a peptide bond.



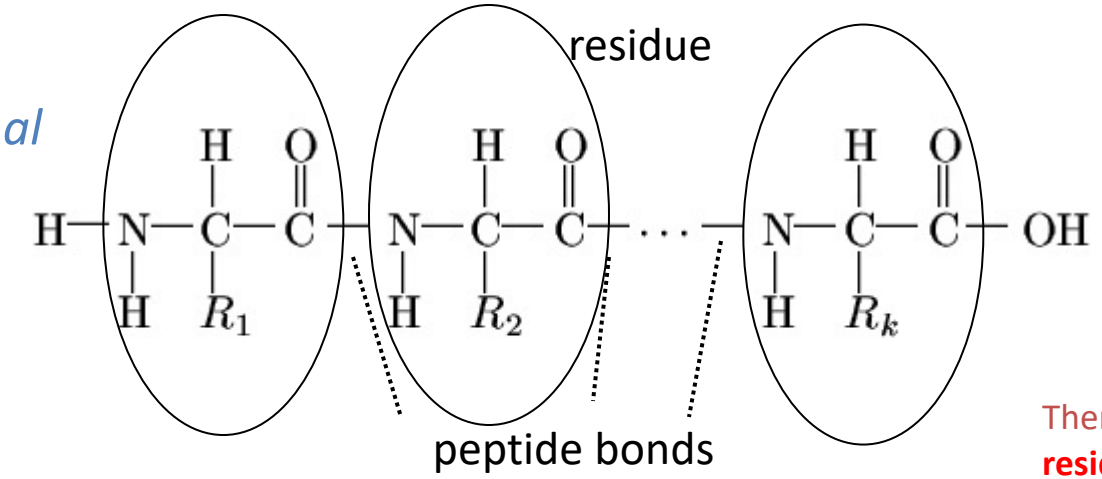
Glycine, or Gly, or G



GR



Arginine, or Arg, or R



Therefore, a peptide is a chain of amino acid **residues** plus H at n-term and OH at c-term.

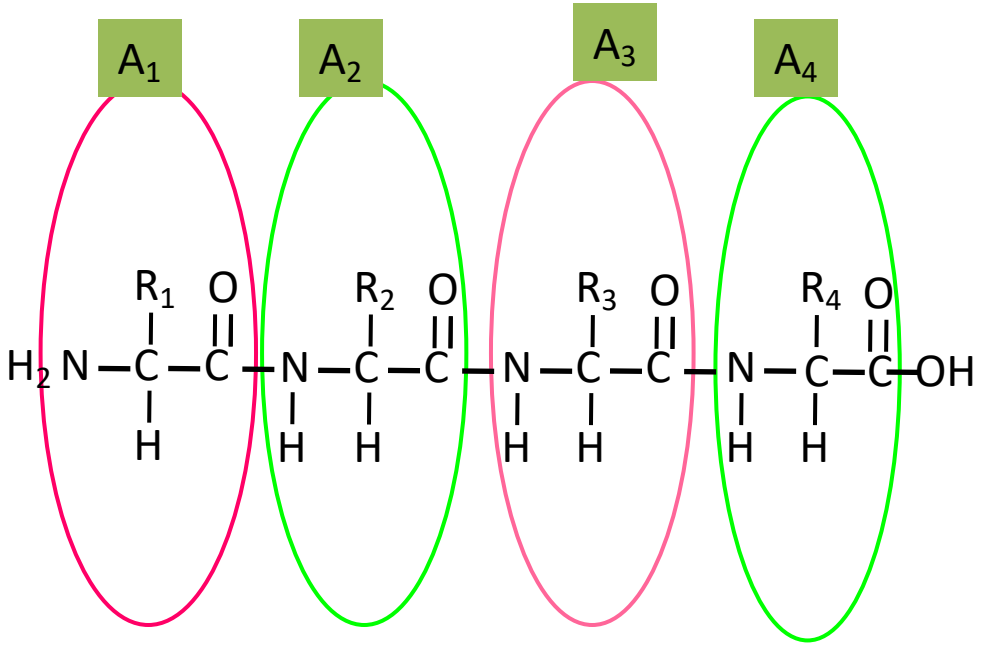
Amino Acid Residue Mass Table

Mass unit: Da (or Dalton)

Name	3-letter code	1-letter code	Monoisotopic Mass	Average Mass	Composition
Alanine	Ala	A	71.03711	71.08	C_3H_5NO
Arginine	Arg	R	156.10111	156.2	$C_6H_{12}N_4O$
Asparagine	Asn	N	114.04293	114.1	$C_4H_6N_2O_2$
Aspartic Acid	Asp	D	115.02694	115.1	$C_4H_5NO_3$
Cysteine	Cys	C	103.00919	103.1	C_3H_5NOS
Glutamic Acid	Glu	E	129.04259	129.1	$C_5H_7NO_3$
Glutamine	Gln	Q	128.05858	128.1	$C_5H_8N_2O_2$
Glycine	Gly	G	57.02146	57.05	C_2H_3NO
Histidine	His	H	137.05891	137.1	$C_6H_7N_3O$
Isoleucine	Ile	I	113.08406	113.2	$C_6H_{11}NO$
Leucine	Leu	L	113.08406	113.2	$C_6H_{11}NO$
Lysine	Lys	K	128.09496	128.2	$C_6H_{12}N_2O$
Methionine	Met	M	131.04049	131.2	C_5H_9NOS
Phenyalanine	Phe	F	147.06841	147.2	C_9H_9NO
Proline	Pro	P	97.05276	97.12	C_5H_7NO
Serine	Ser	S	87.03203	87.08	$C_3H_5NO_2$
Threonine	Thr	T	101.04768	101.1	$C_4H_7NO_2$
Tryptophan	Trp	W	186.07931	186.2	$C_{11}H_{10}N_2O$
Tyrosine	Tyr	Y	163.06333	163.2	$C_9H_9NO_2$
Valine	Val	V	99.06841	99.13	C_5H_9NO

So, what's the approximate (integer) mass of peptide GGK?

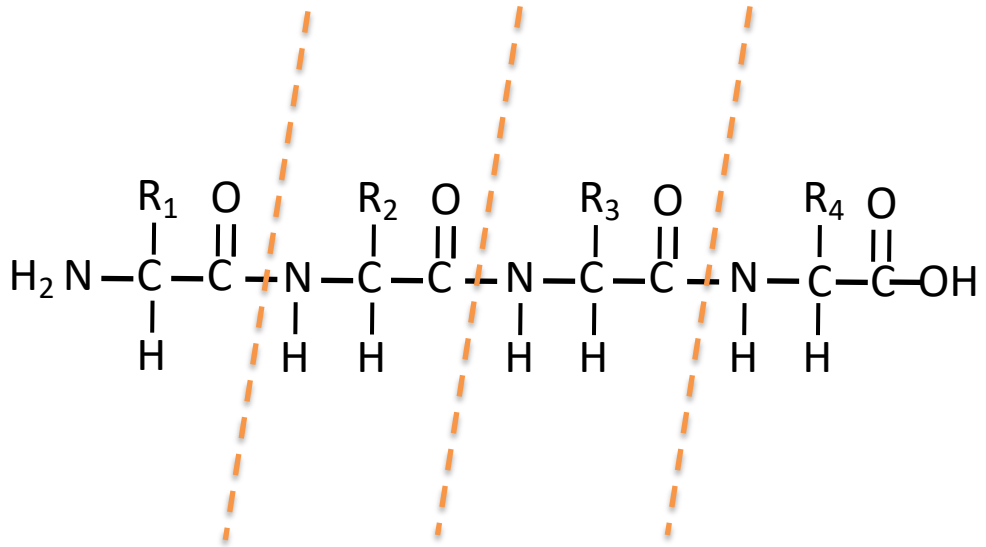
Peptide Mass



$m(A_1)+m(A_2)+m(A_3)+m(A_4)+m(H_2O)$

$m(H_2O) = 18.0105$

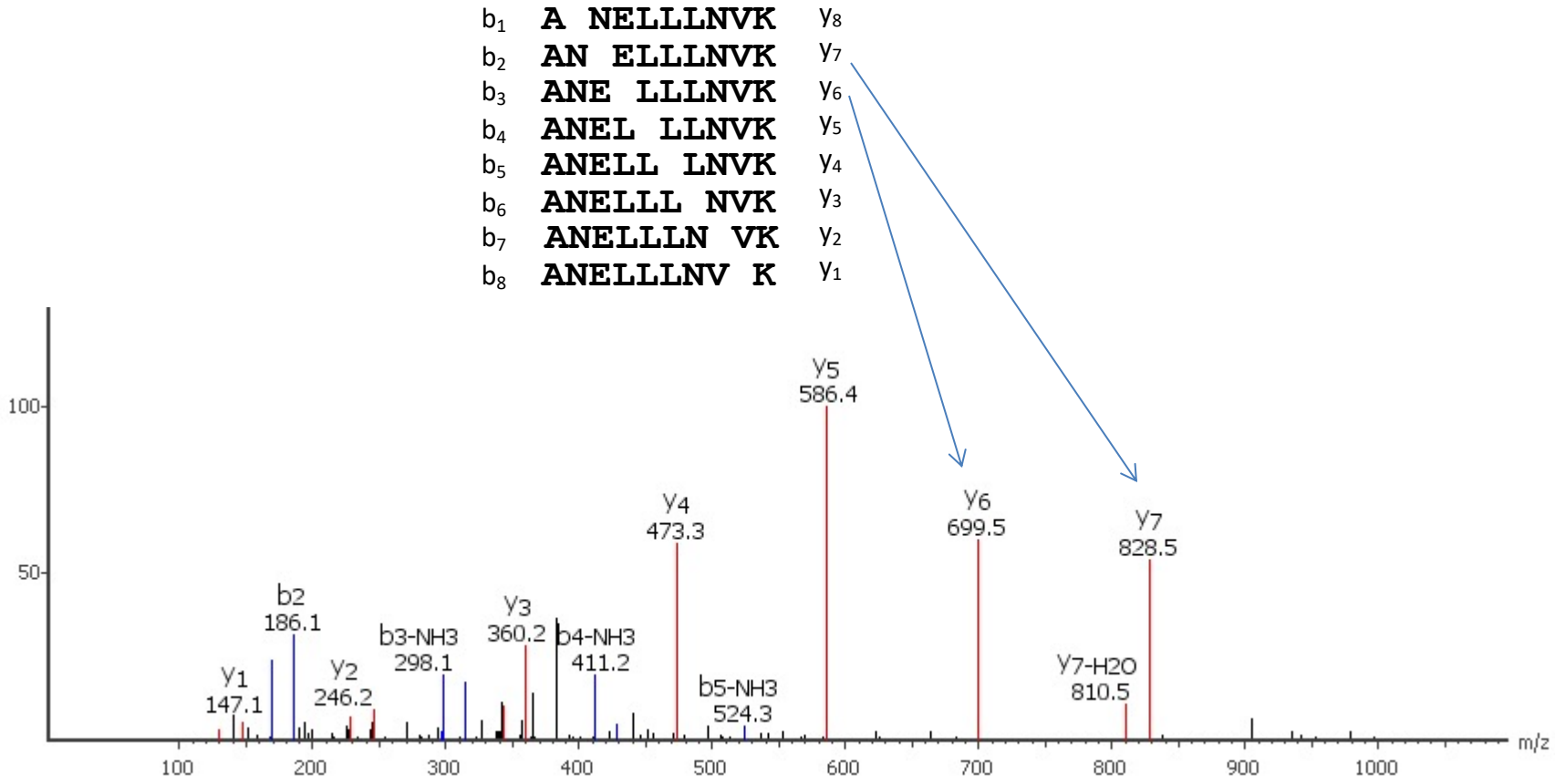
Peptide Fragmentation in MS/MS



$m(A_1)+m(A_2)+m(A_3)+m(A_4)+m(H_2O)$

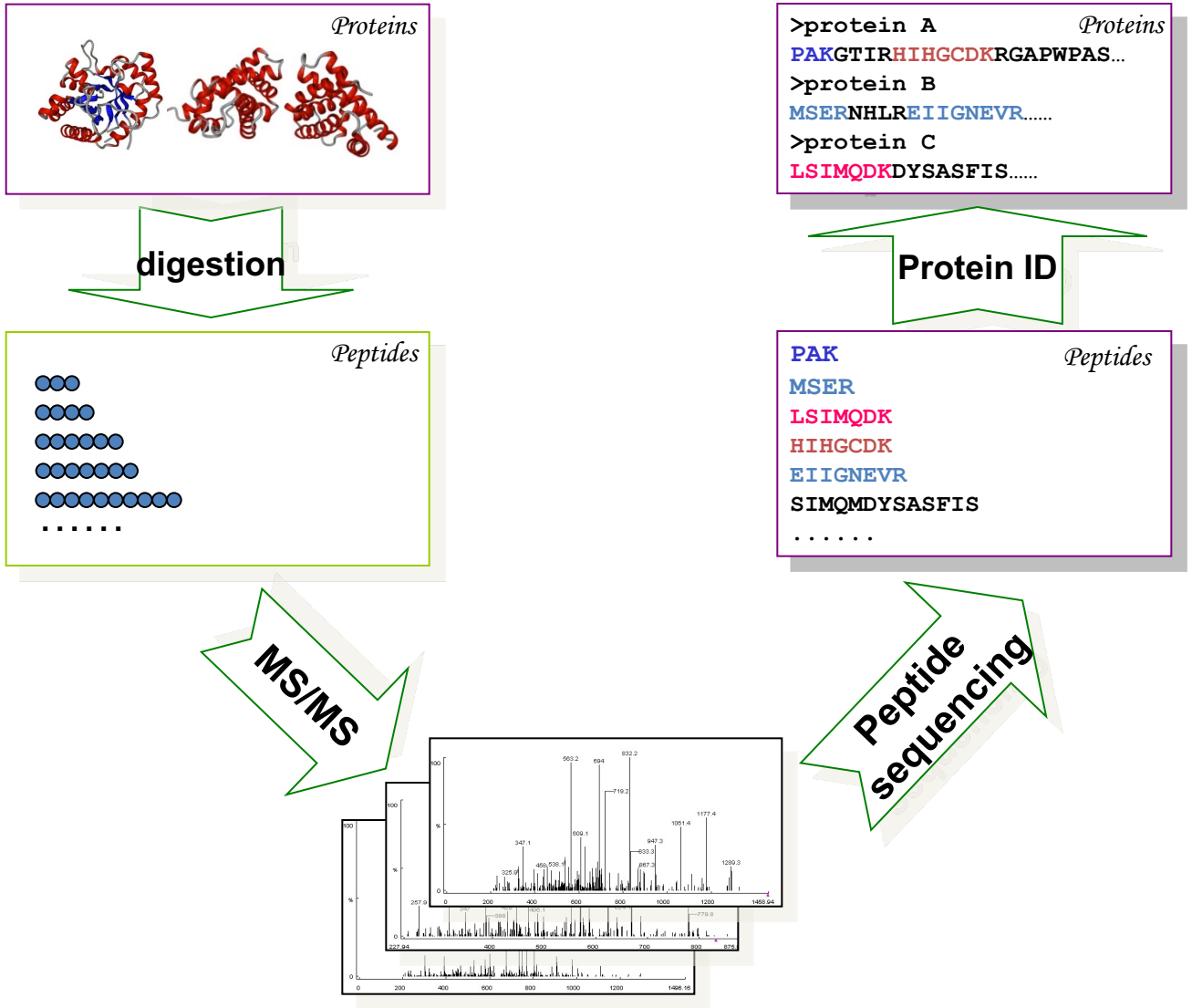
$m(H_2O) = 18.0105$

MS/MS of a Peptide

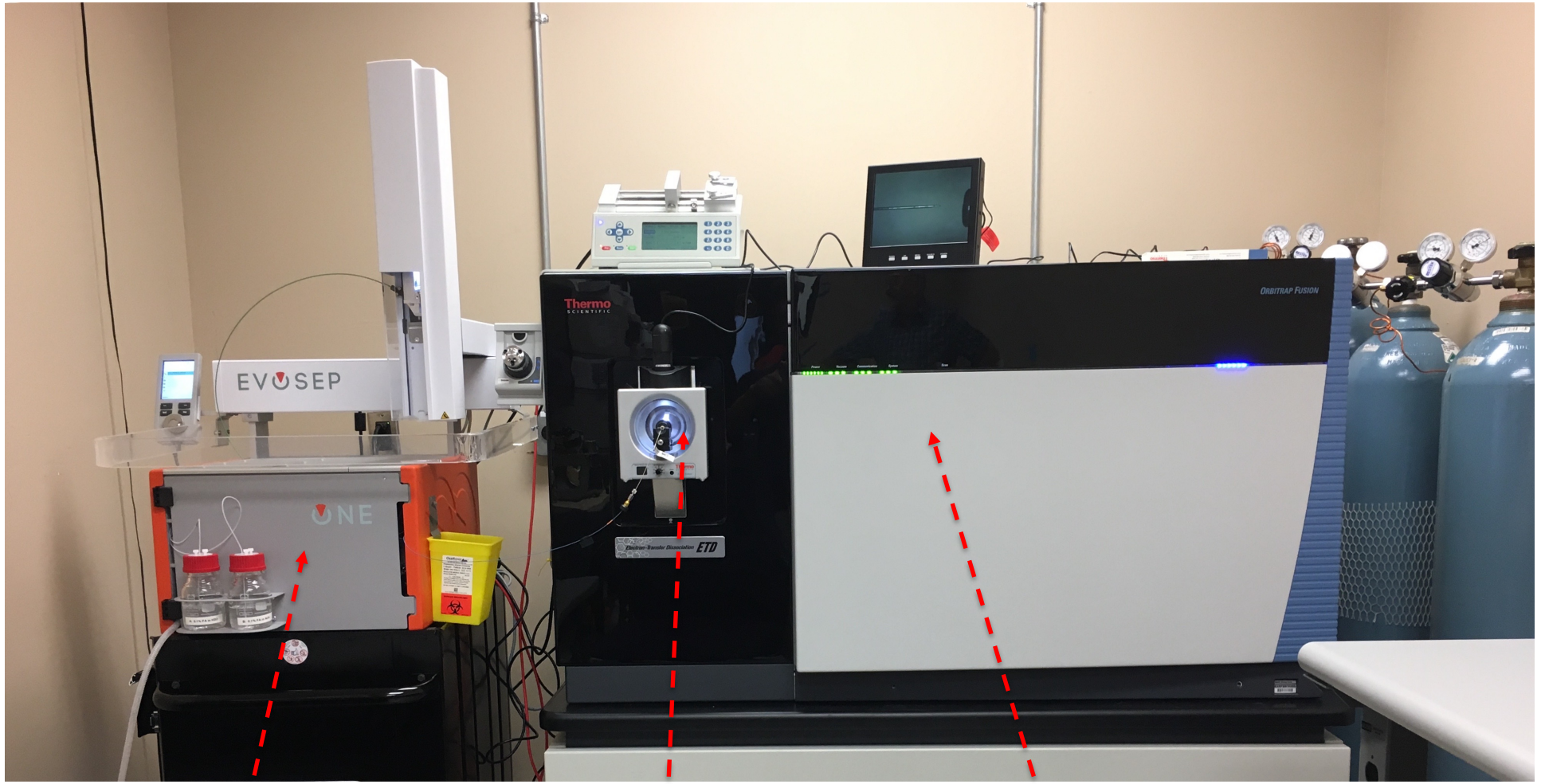


Mass difference between two adjacent “ladder” ions can be used to determine the amino acid (residue).

Bottom Up Proteomics



A Typical LC-MS Setup



LC separation

Ionize and inject

measurement

Three Basic Components

- Ionizer
  - provide electric charges to the molecules
  - MALDI, **ESI**
- Mass Analyzer
  - Separates the ions according to m/z
  - Magnetic Sector, Iontrap, **TOF**, Quadrupole, FT, **Orbitrap**
- Detector
  - Detect the separated ions
  - Electron multiplier, Fourier Transform

Ionization

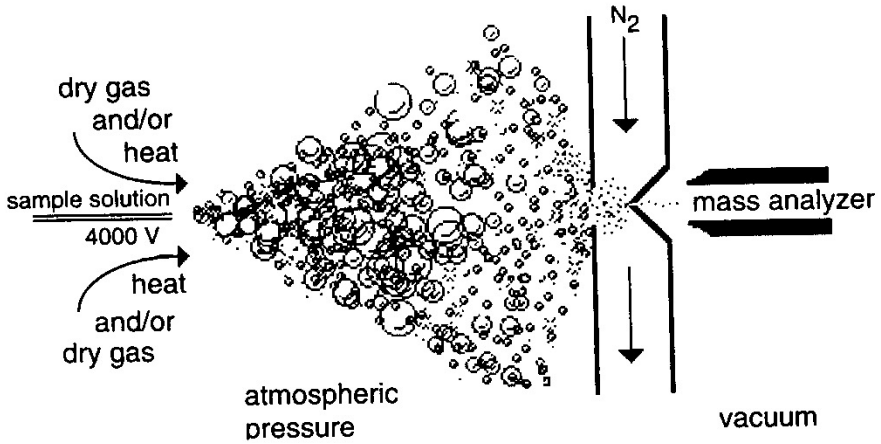
- MALDI (Matrix Assisted Laser Desorption/Ionization)
- ESI (Electrospray Ionization)

Koichi Tanaka

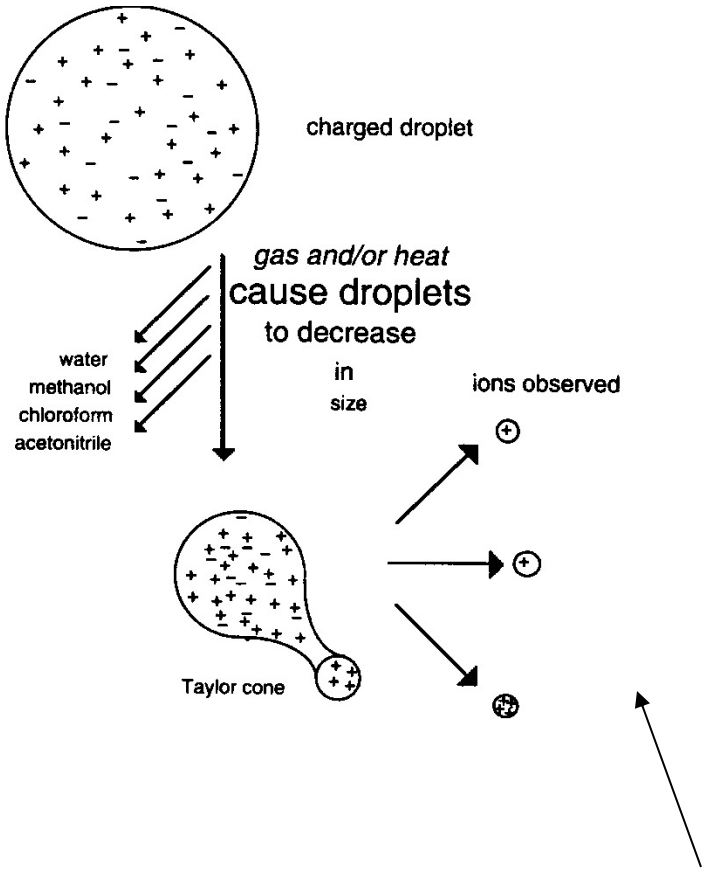
John Fenn

Nobel prize in Chemistry, 2002

Ionization – ESI



Electrospray Ionization: Formation of Charged Droplets



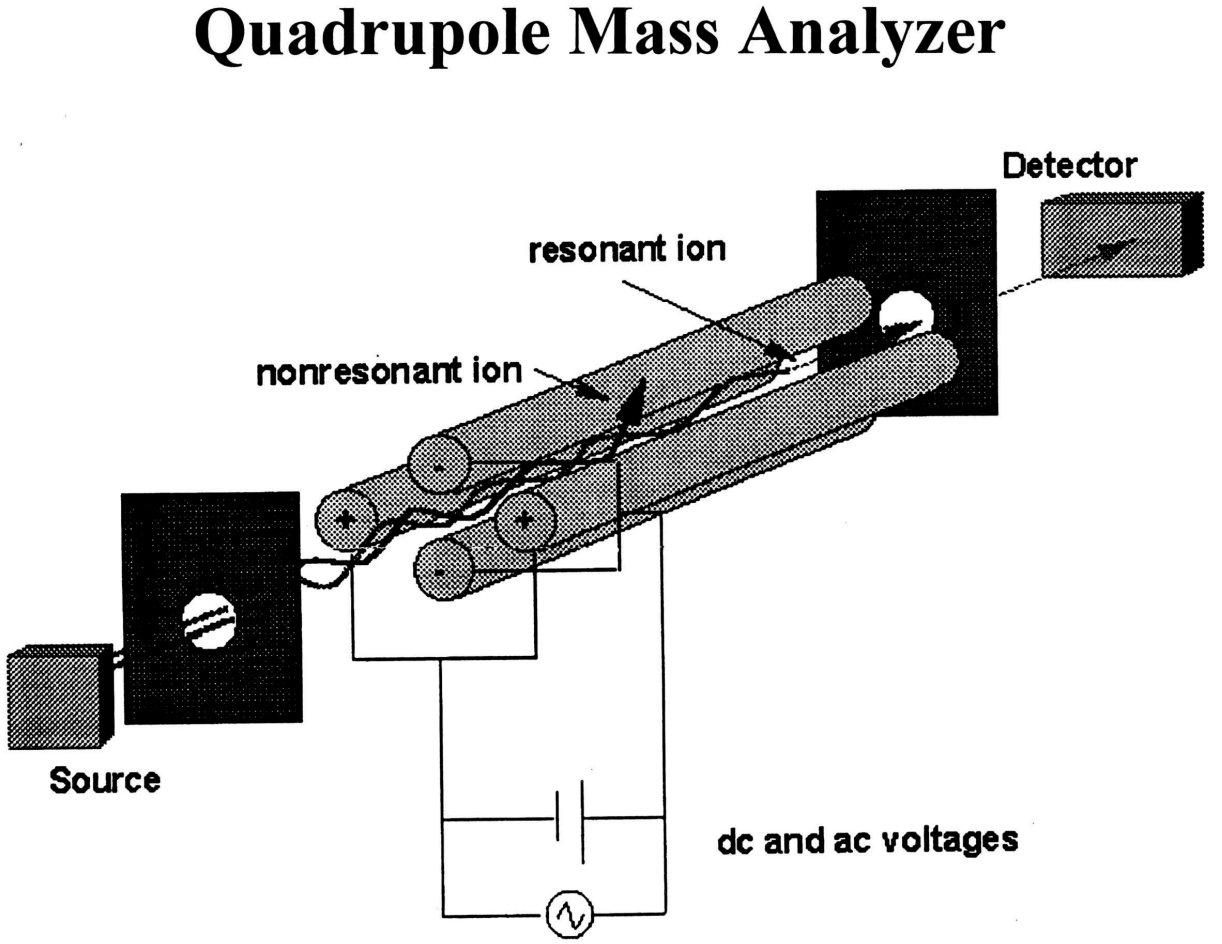
Formation of multiply charged ions

ESI History

- Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. (1989). "Electrospray ionization for mass spectrometry of large biomolecules". Science 246 (4926): 64–71.
- A.B. from Berea College in his new hometown.
- 1940: Ph.D. from Yale University.
- 1962: He joined the Yale University faculty.
- 1987: he reached the mandatory retirement age (70).
- University-mandated move to smaller laboratory space.
- Started to work on ESI.
- 1994: Fenn joined Virginia Commonwealth University.
- The patent rights to ESI became the subject of a legal case between Yale University and Fenn.
- 2005: Yale was awarded over one million dollars and partial patent rights to the technique.

Fenn's first electrospray ionization source (top) coupled to a single quadrupole mass spectrometer

Mass Analyzer (1) – Quadrupole



- **Adjusting DC voltage allows different m/z ions to pass. (Mass filter)**
- **The complete spectrum is obtained by scanning whole range.**

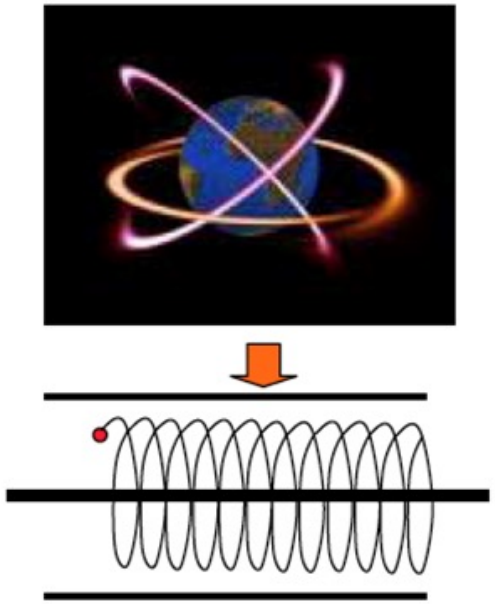
Mass range 10- 4,000 Da

Mass Analyzer (2) – TOF

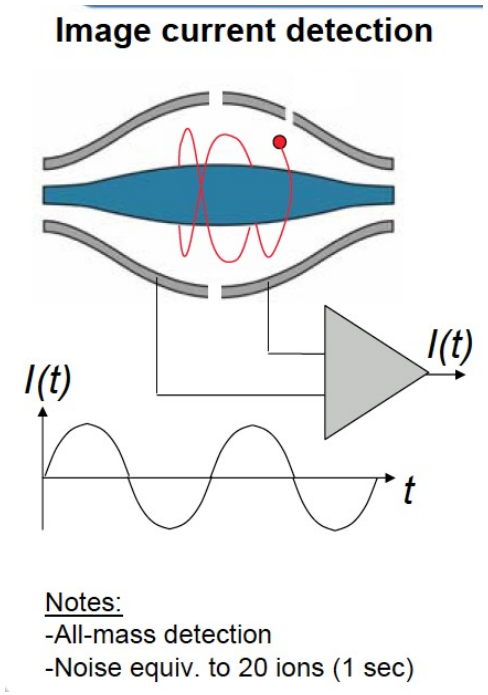
- Time of Flight.

Time of flight is proportional to sqrt(m/z)

Mass Analyzer (3) – Orbitrap



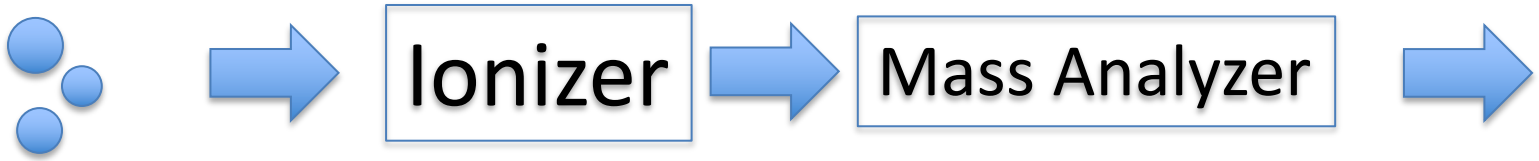
Moving ions are trapped around an electrode.



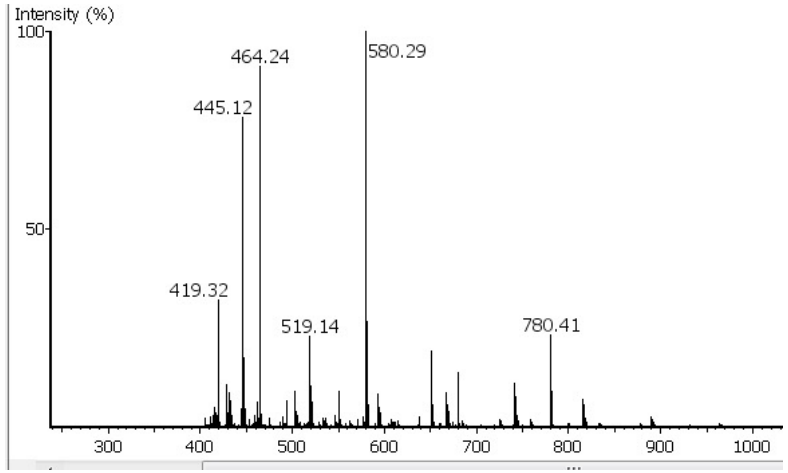
By shaping the electrode appropriately, ions also move left and right. Left-right frequency proportional to sqrt(m/z).

Fourier transform to convert the time domain signal to frequencies.

Mass Spectrometer

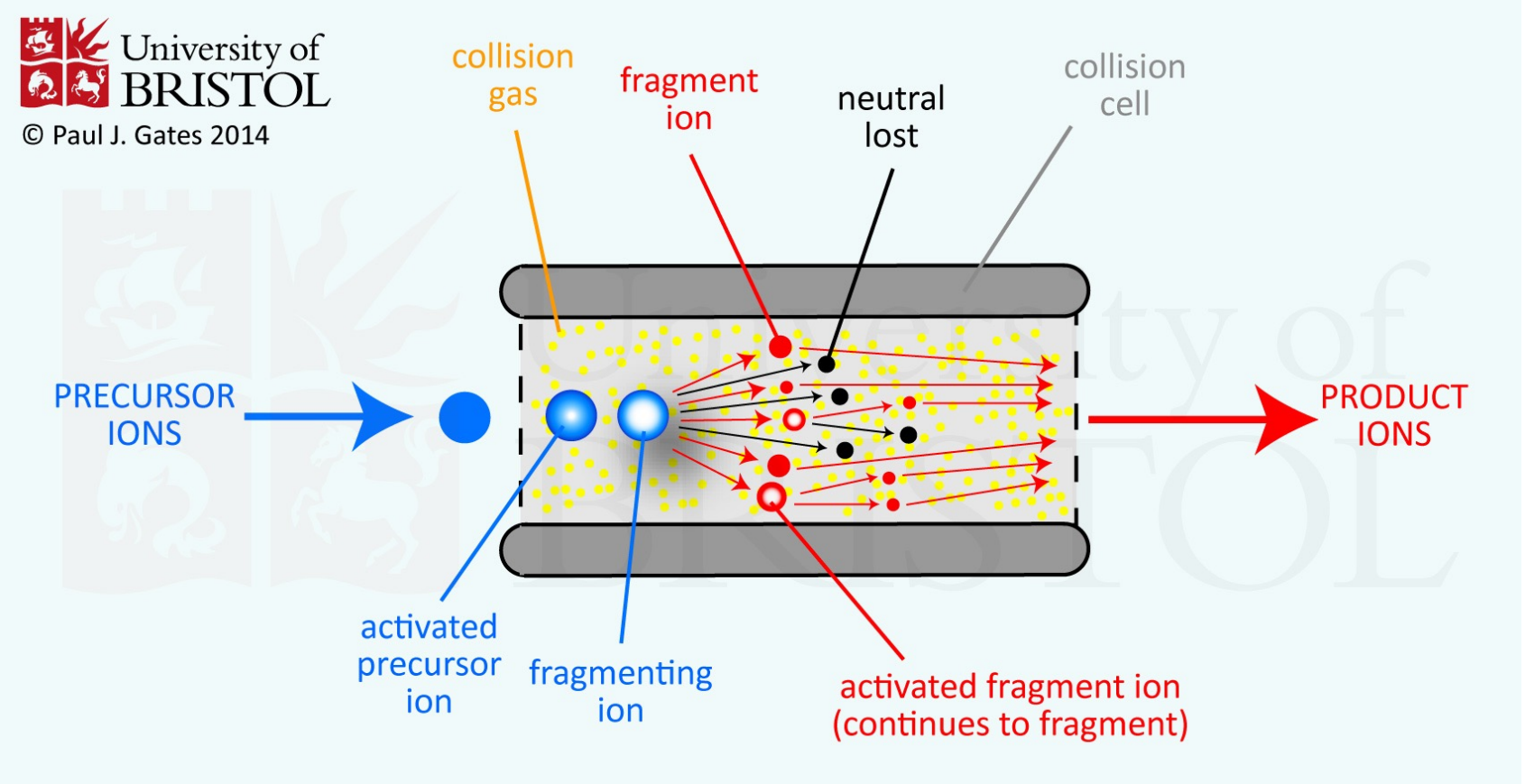


Molecules

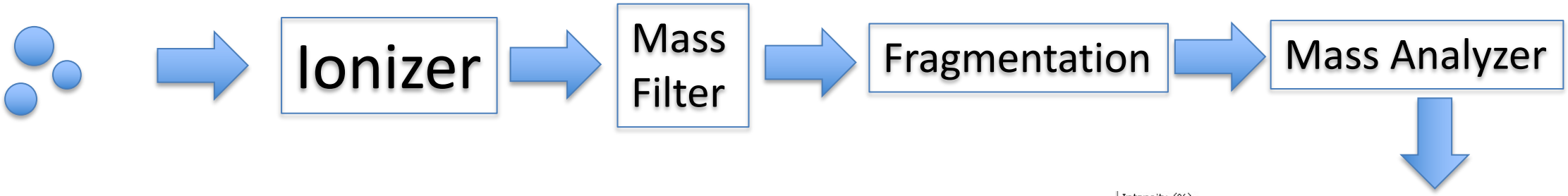


Spectrum

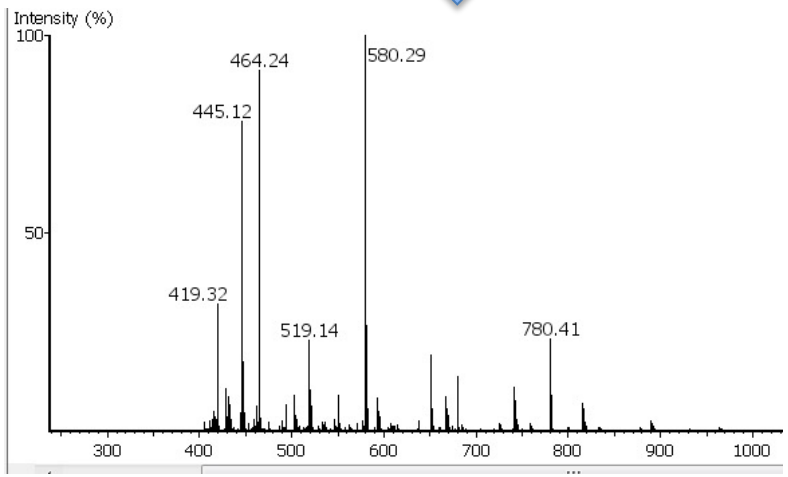
Collision Cell



Tandem Mass Spectrometer



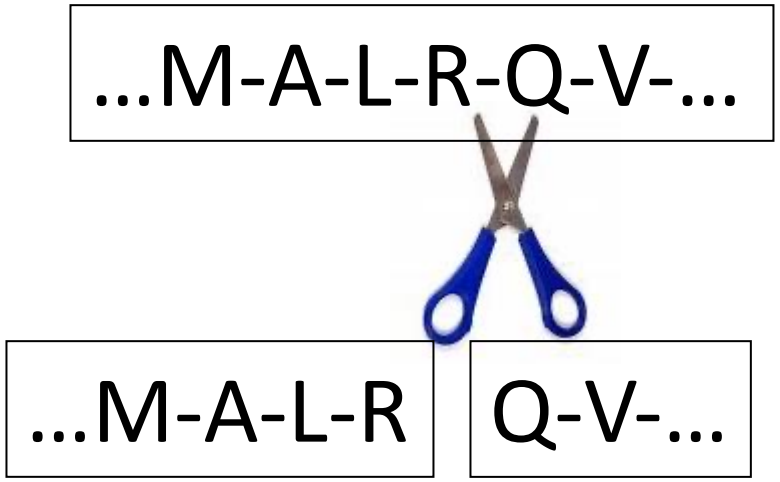
Molecules



MS/MS Spectrum

Protease

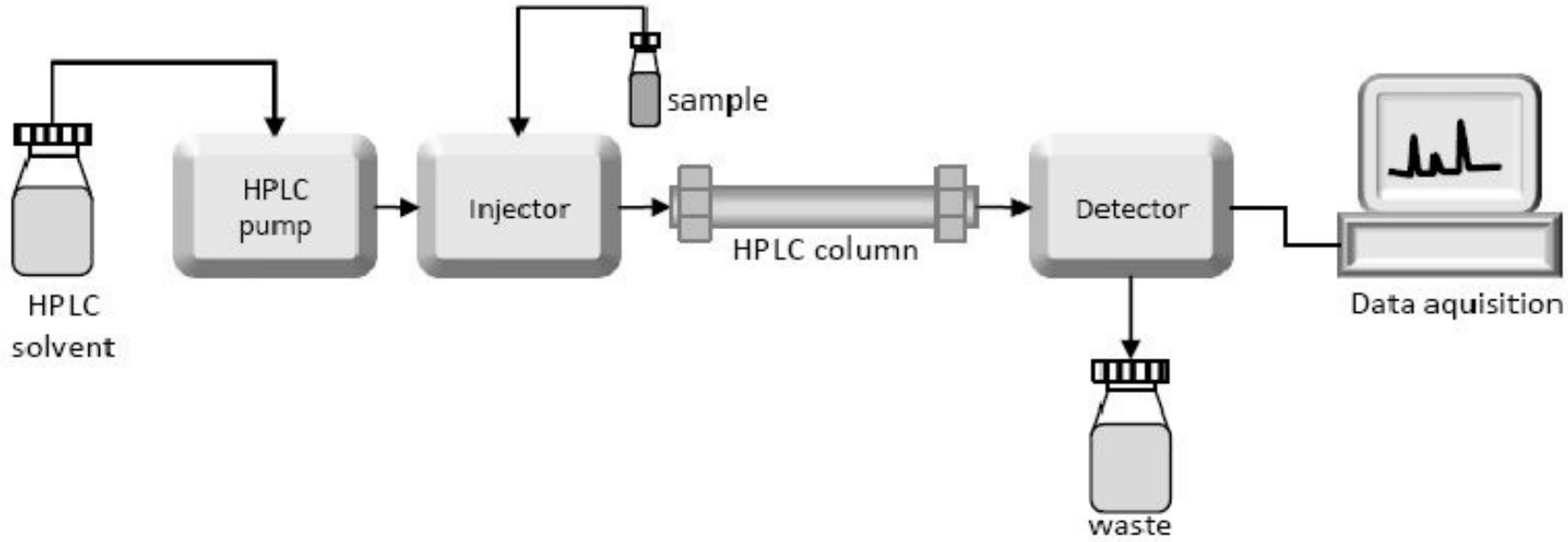
- Proteins are generally too large for mass spec. They need to be cut into short peptides first.
- A **protease** is any enzyme that conducts proteolysis. In another word, a protease breaks protein in water.
- **Trypsin** is the most commonly used enzyme. It digests at site [KR] | [^P]
  - After K or R, but not before P.



Sample Complexity

- Challenge: Suppose there are 10,000 proteins to be analyzed. Each produces 100 peptides. Then there are 1 million peptides.
- If all of them are injected into mass spec simultaneously, we will see a peak everywhere. No useful information at all.
- Solution: separate them and inject a subset of them to mass spec at any given time.

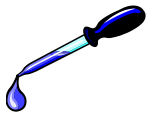
Liquid Chromotography (LC)

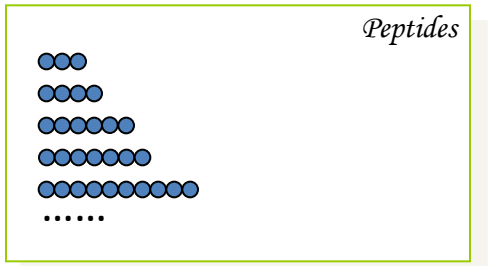


1. Peptides are loaded to the column.
2. Solvent is pumped into the column to wash peptides off.
3. By gradually adjusting the solvent's hydrophobicity, peptides with different hydrophobicity will come off at different time. This is the rentention time (RT) of the peptide.
4. Peptides are therefore separated. (But the separation is not perfect..)

Put All Pieces Together

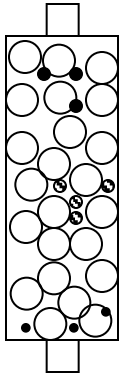
(1) A mixture of protein is digested into peptides with added enzyme (usually trypsin).

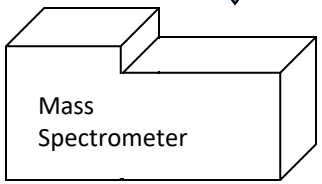




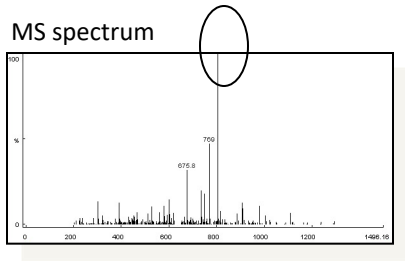
digestion

(2) The resulting peptides are separated with liquid chromatography (LC). Different peptides elute at different time. The separation may not be perfect.



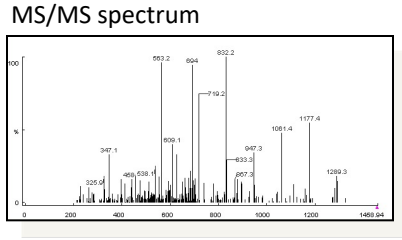


MS



(3) The spectrometer scans the peptide ions at a particular time and obtains a profile MS scan. Each peak in the MS spectrum supposedly corresponds to a peptide.

MS/MS



(4) The spectrometer selects a peak (a peptide ion) with the first mass analyzer, fragments it and produces an MS/MS scan with the second mass analyzer.