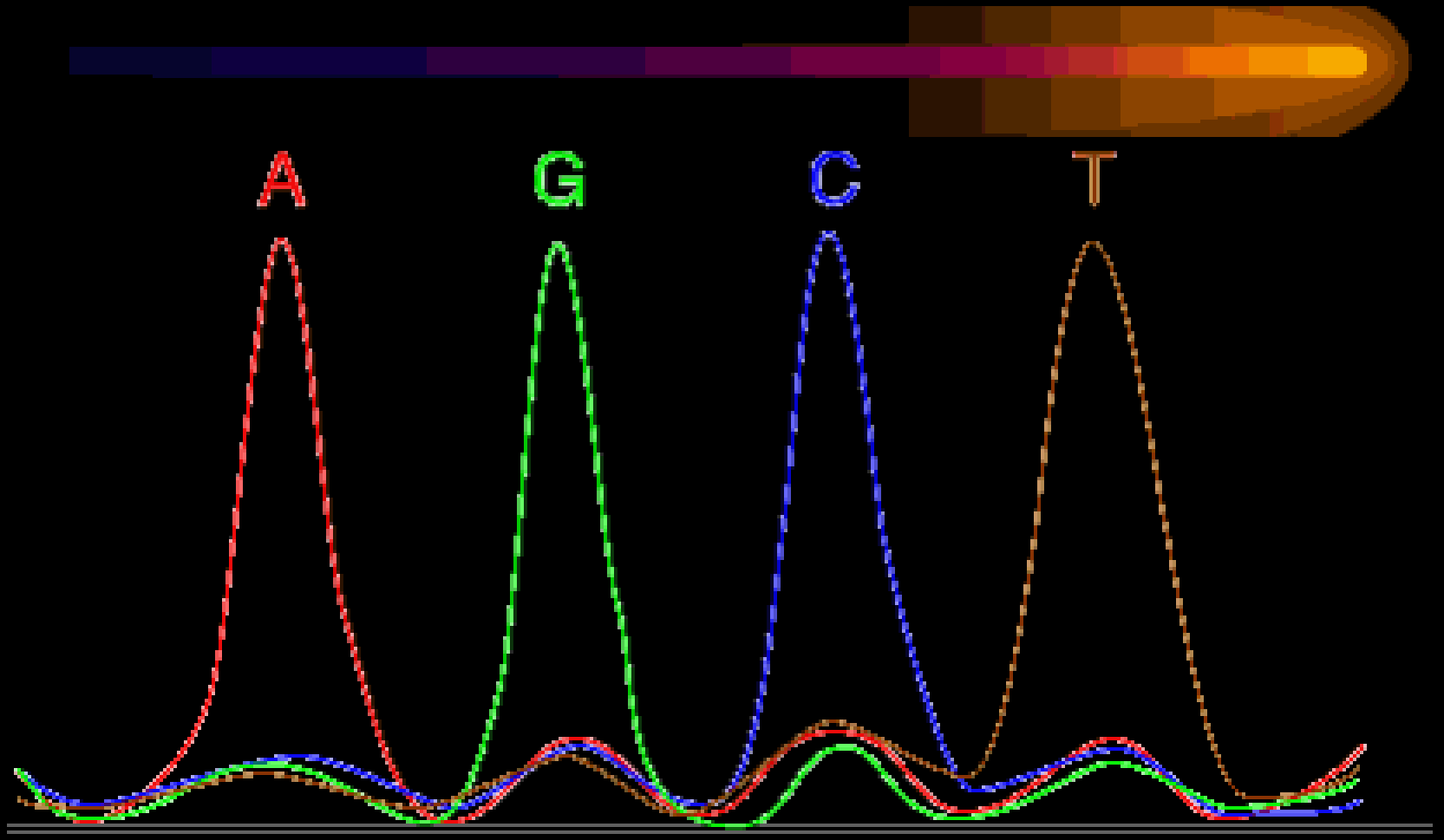
BENG 183

Trey Ideker

Protein Sequencing



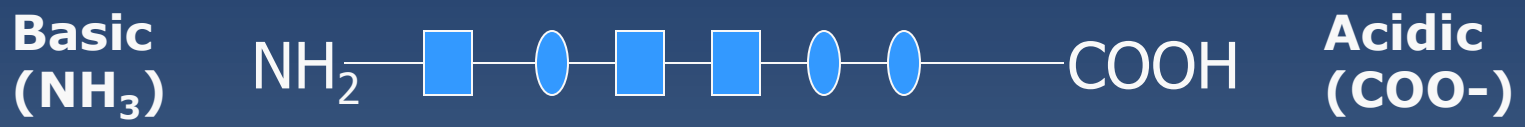
Introduction to Proteins

Proteins are of vital importance to biological systems because of

- their ability to interact with other molecules based on 3D structures;
- their ability to catalyze chemical reactions.

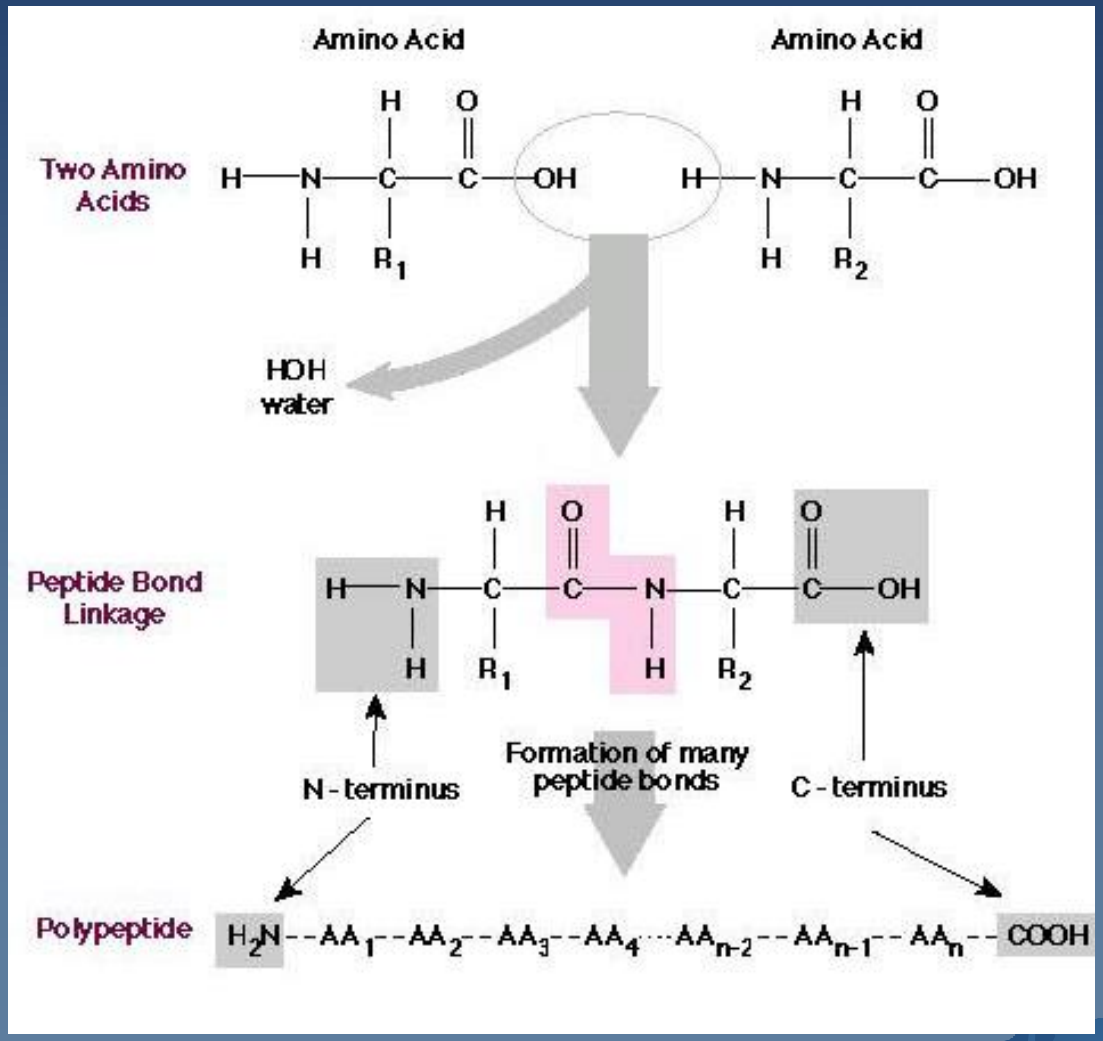
The following slides borrowed from Hong Li's Biochemistry Course: www.sb.fsu.edu/~hongli/4053NOTES

Proteins are linear polymers of amino acids

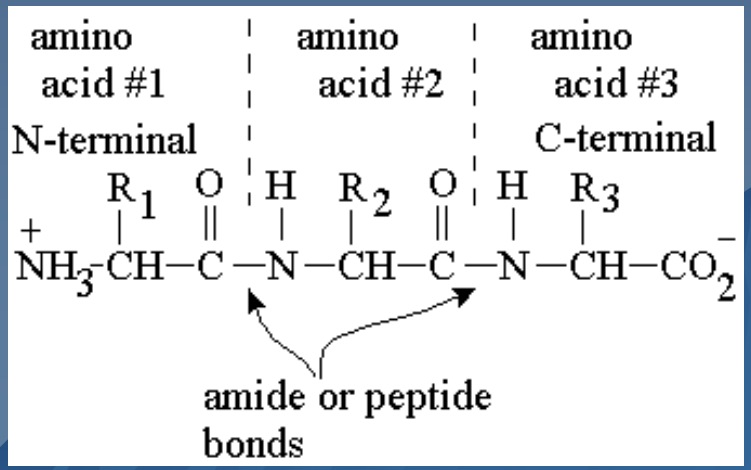


- Terminology:
  - Residues----amino acids
  - Dipeptide---a peptide containing two amino acids
  - Polypeptide---a peptide containing more than a dozen amino acids
  - Monomer -----a protein containing a single polypeptide
  - Multimers------a protein containing more than one polypeptide
  - Homodimer-----a protein contain two identical polypeptide
  - Heterodimers----a protein contain two different polypeptide
  - Tetramers-------a protein containing four polypeptides (can be homo- or hetero-)
- Although amino acids are weak acids overall, they are actually ***zwitterions*** with both acidic and basic groups

Structure of the peptide bond



2 amino acids are joined by the peptide bond, a reaction catalyzed by the ribosome in all cells:



Crash course in protein chemistry

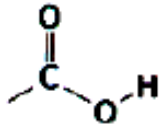
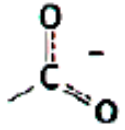
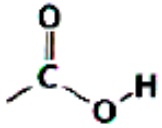
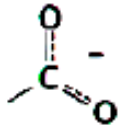
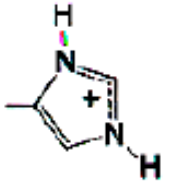
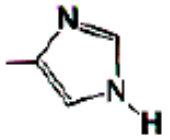
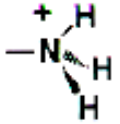
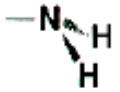
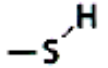
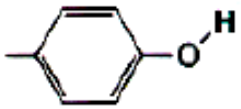
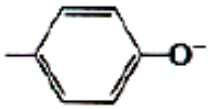
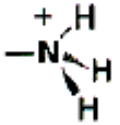
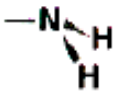
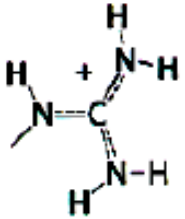
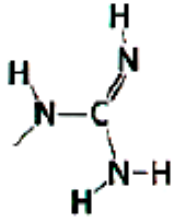
pH [-log(H)], cation (+), anion (-)

pKa (dissocation constant):The pH at which half the molecules of an acid are neutral and half are dissociated/charged: $pH = pKa + \log([A^-]/[AH])$

pI (isoelectric point): The pH at which a protein has an equal number of pos. and negative charges: $pI = \frac{1}{2} \cdot (pK_i + pK_j)$

Type of AA	pK_i	pK_j
Non-polar	α-Carboxyl Group	α-Amino Group
Polar Neg	α-Carboxyl Group	Side Chain
Polar Pos	Side Chain	α-Amino Group

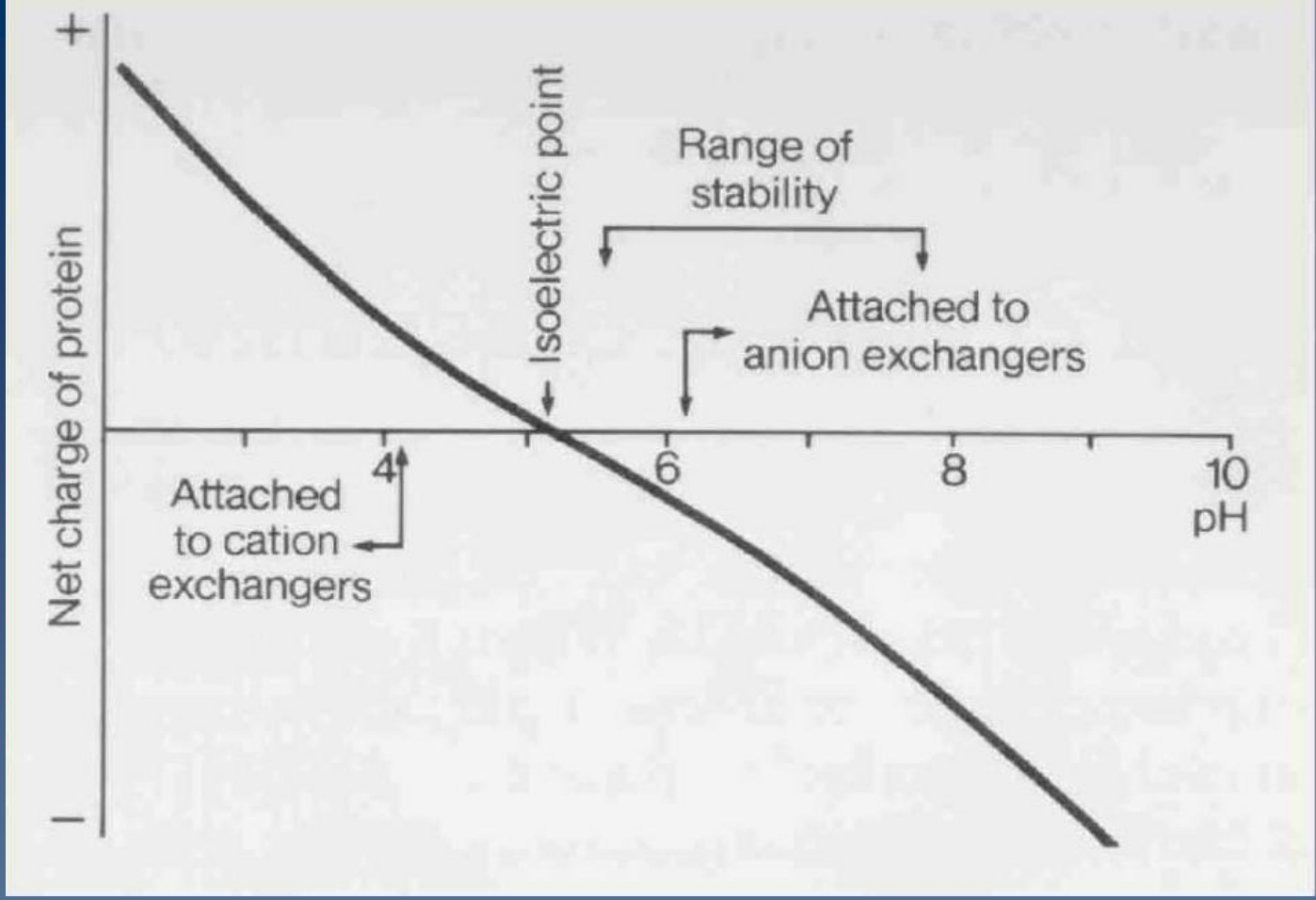
Typical pK_a values of ionizable groups in proteins

Group	Acid	⇌	Base	Typical pK_a*
Terminal α-carboxyl group		⇌		3.1
Aspartic acid Glutamic acid		⇌		4.1
Histidine		⇌		6.0
Terminal α-amino group		⇌		8.0
Cysteine		⇌		8.3
Tyrosine		⇌		10.9
Lysine		⇌		10.8
Arginine		⇌		12.5

pKa chart

Amino Acid	a-carboxylic acid	a-amino	Side chain
Alanine	2.35	9.87	
Arginine	2.01	9.04	12.48
Asparagine	2.02	8.80	
Aspartic Acid	2.10	9.82	3.86
Cysteine	2.05	10.25	8.00
Glutamic Acid	2.10	9.47	4.07
Glutamine	2.17	9.13	
Glycine	2.35	9.78	
Histidine	1.77	9.18	6.10
Isoleucine	2.32	9.76	
Leucine	2.33	9.74	
Lysine	2.18	8.95	10.53
Methionine	2.28	9.21	
Phenylalanine	2.58	9.24	
Proline	2.00	10.60	
Serine	2.21	9.15	
Threonine	2.09	9.10	
Tryptophan	2.38	9.39	
Tyrosine	2.20	9.11	10.07
Valine	2.29	9.72	

Net charge of a protein as a function of pH



Acidic pH = proteins present as cations -- carboxy group does not dissociate and amino group is protonated.

Basic pH = proteins present as anions -- amino group is a free base and the carboxy group is dissociated.

Under either state, it is easy to bind proteins to an oppositely charged stationary phase as long as the salt concentration is kept low.

Ion Exchange Column

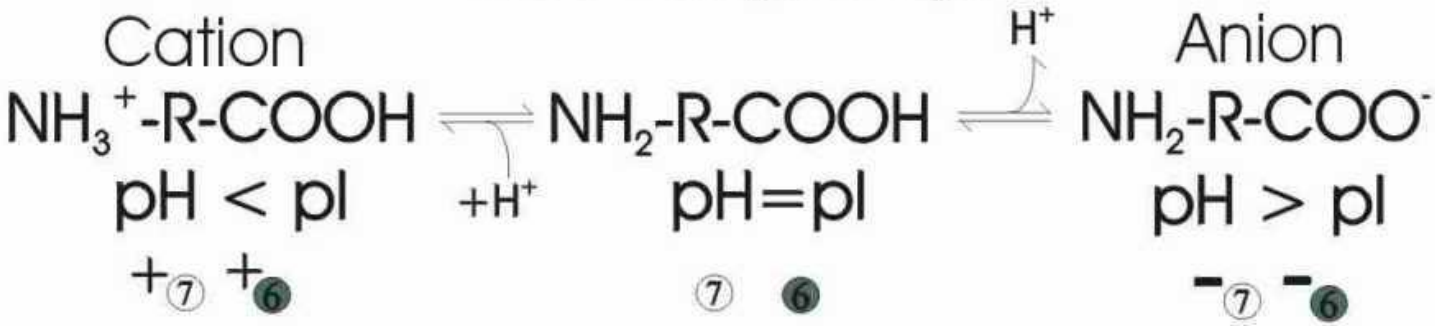
- Separates molecules by their net surface charge
- Uses covalently-bound column materials with opposite charge
  Diethyl-amino-ethyl (DEAE) for +
  Carboxy-methyl (CM) for – charge
- The ionic groups of the column are compensated by small concentrations of counter ions in the buffer (anions – DEAE; cations + CM).
- When a sample is added to the column, an exchange with the weakly bound counter ions takes place.

Elution from Ion Exchange Column

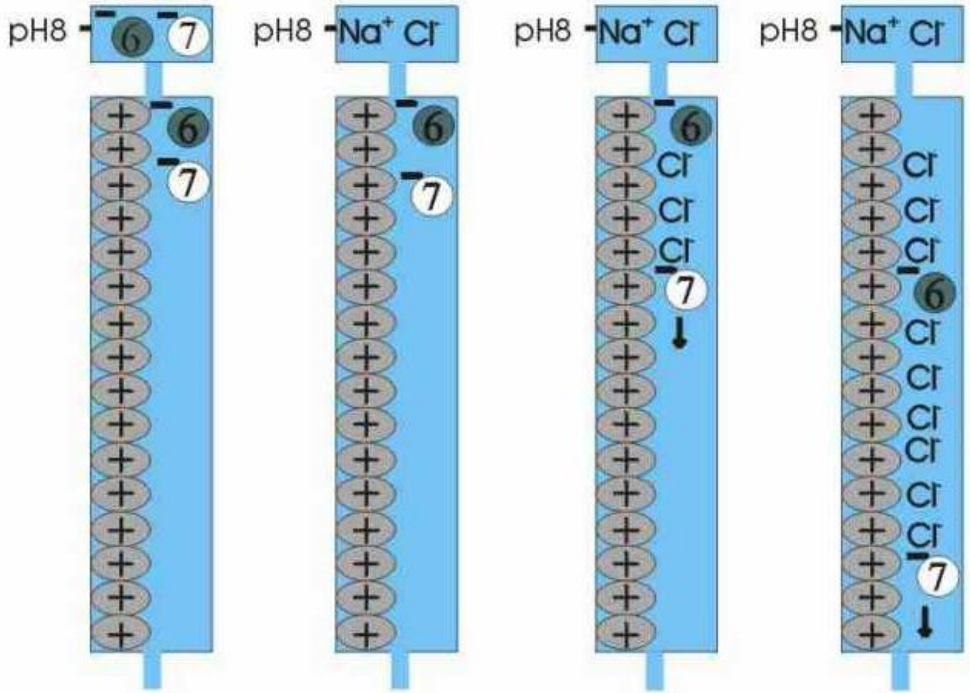
- At low salt, all components with affinity for the ion exchanger will be tightly adsorbed at the top of the column and nothing will remain in the mobile phase.
- When the ionic strength of the mobile phase is increased by adding a neutral salt, the salt ions will compete with the protein. More of the sample components will be partially desorbed and start moving down the column.
- The higher the net charge of the protein, the higher the ionic strength needed to bring about desorption.

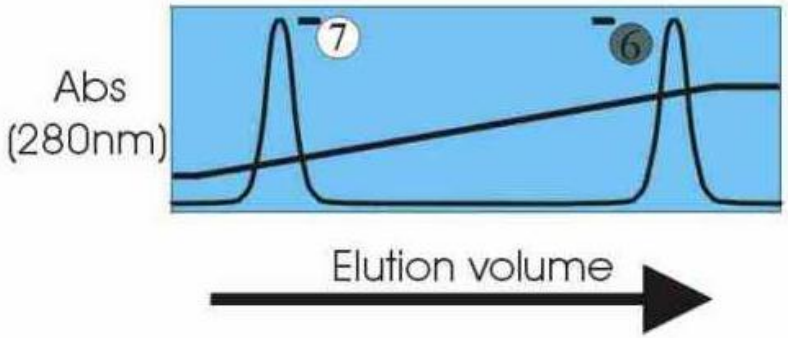
Ion Exchange Chromatography

Separation based on Surface Charges on the Protein-detergent complex



Anion Exchange





Elution

1) Increase salt concentration
2) Change pH towards pI

Analysis of Protein Sequences

Proteins can be sequenced in one of two ways:

(1) Real amino acid sequencing

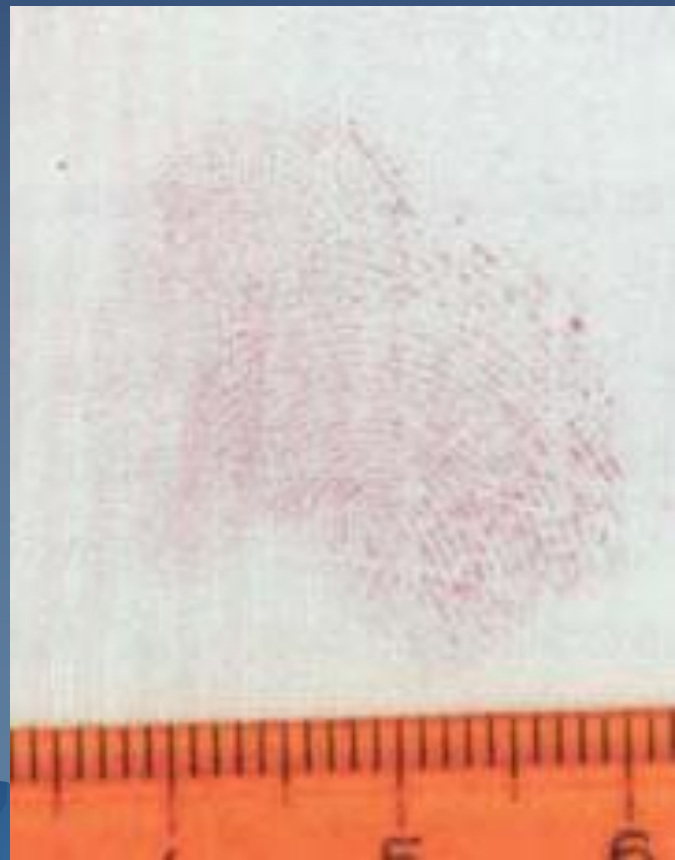
- Amino-acid analyzer
- Edman sequencer
- Mass spectrometry

(2) Sequencing the corresponding DNA from the gene

Amino Acid Analysis of Proteins

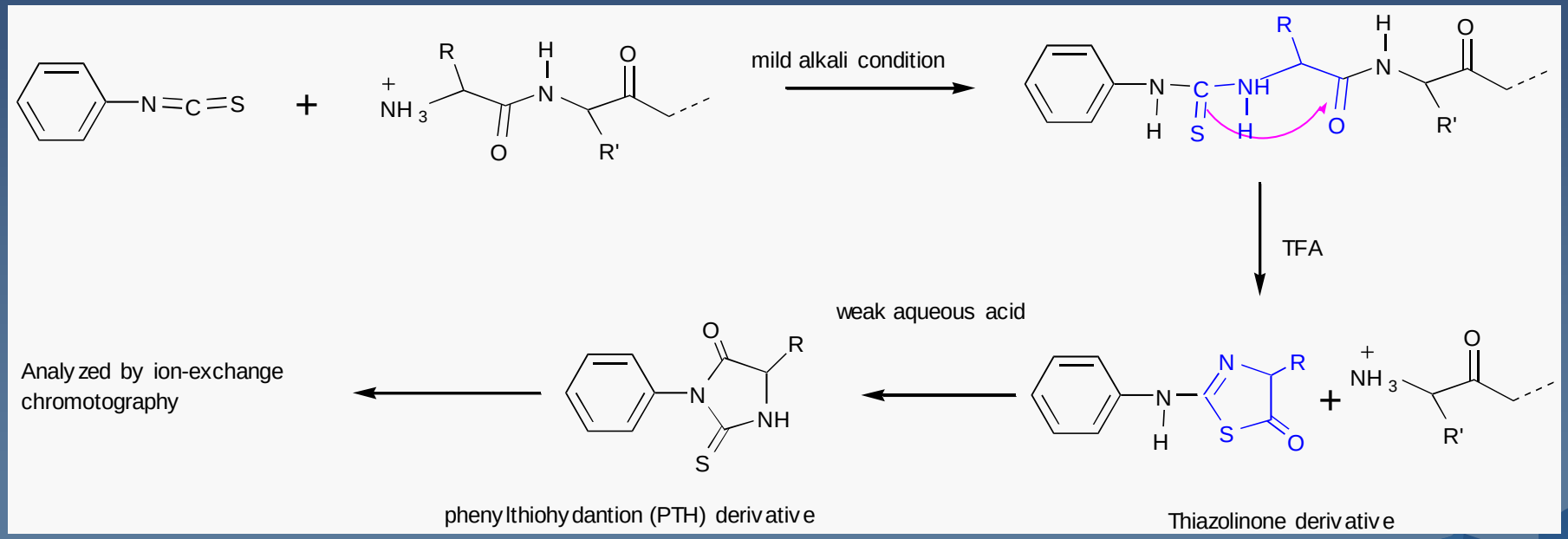
- Analyzes the amount of each type of amino acid, but gives no information on the order of amino acids
- 6N HCl @ 110° C for 24, 48, 72 hours in sealed glass vials (Trp is destroyed, must be analyzed by other means)
- Each reaction mixture is loaded on an ion-exchange column where each amino acid can be eluted according to its charge
- Eluted amino acids are quantified by reacting to ninhydrin (postcolumn derivatization using amino-acid "fingerprinting" reagent as for forensics)

  This method has been automated (amino acid analyzer)



Edman Degradation Method

- Edman degradation uses Edman reagent to determine the order of amino acids from the N-terminal end of a protein.



Enzymatic Methods

- C-terminal analysis by carboxypeptidase A, B, C, Y:
  - Carboxypeptidase A cleaves all except for P, R, K;
  - Carboxypeptidase B cleaves only R, and K;
  - Carboxypeptidase C cleaves all;
  - Carboxypeptidase Y cleaves all;
- Trypsin - cleaves on the C-side of Lys or Arg;
- Chymotrypsin - C-side of Phe, Tyr, Trp
- Clostripain - like trypsin, but attacks Arg more than Lys;
- Endopeptidase Lys-C cleaves at C-side of Lys;
- Staphylococcal protease
  - C-side of Glu or Asp in phosphate buffer
  - specific for Glu in acetate or bicarbonate buffer
- Cyanogen Bromide (CNBr) – cuts after methionine

An example (left to problem set):

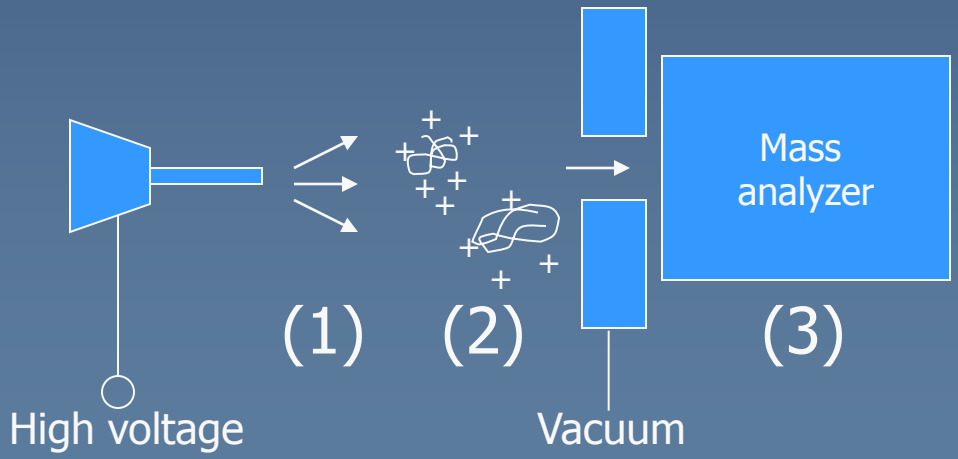
- Amino Acid Analysis yielded: Asn, Gln, Leu, Lys, Met, Tyr, Trp
- Trypsin had no effect;
- Edman degradation yielded PTH-Tyr;
- CNBr treatment yielded a tetrapeptide of positive charge and a tripeptide of zero charge at pH 7.0;
- Brief chymotrypsin (cleaved at Trp) yielded a dipeptide and a tetrapeptide which contains Gln, Leu, Lys, and Met;
- What is the sequence of the peptide?

Mass Spectrometry

- Because previous approaches are so labor intensive and non-systematic, mass spectrometry is increasingly used in protein identification
- MS determines the mass-to-charge (m/z) ratio of protein mixtures or peptides (if hydrolyzed by a protease):
  - Evaporate and ionize molecules in a vacuum
  - Separate the ions in space and/or time based on their m/z ratios;
  - Measure different m/z ratios

Mass Spectrometer

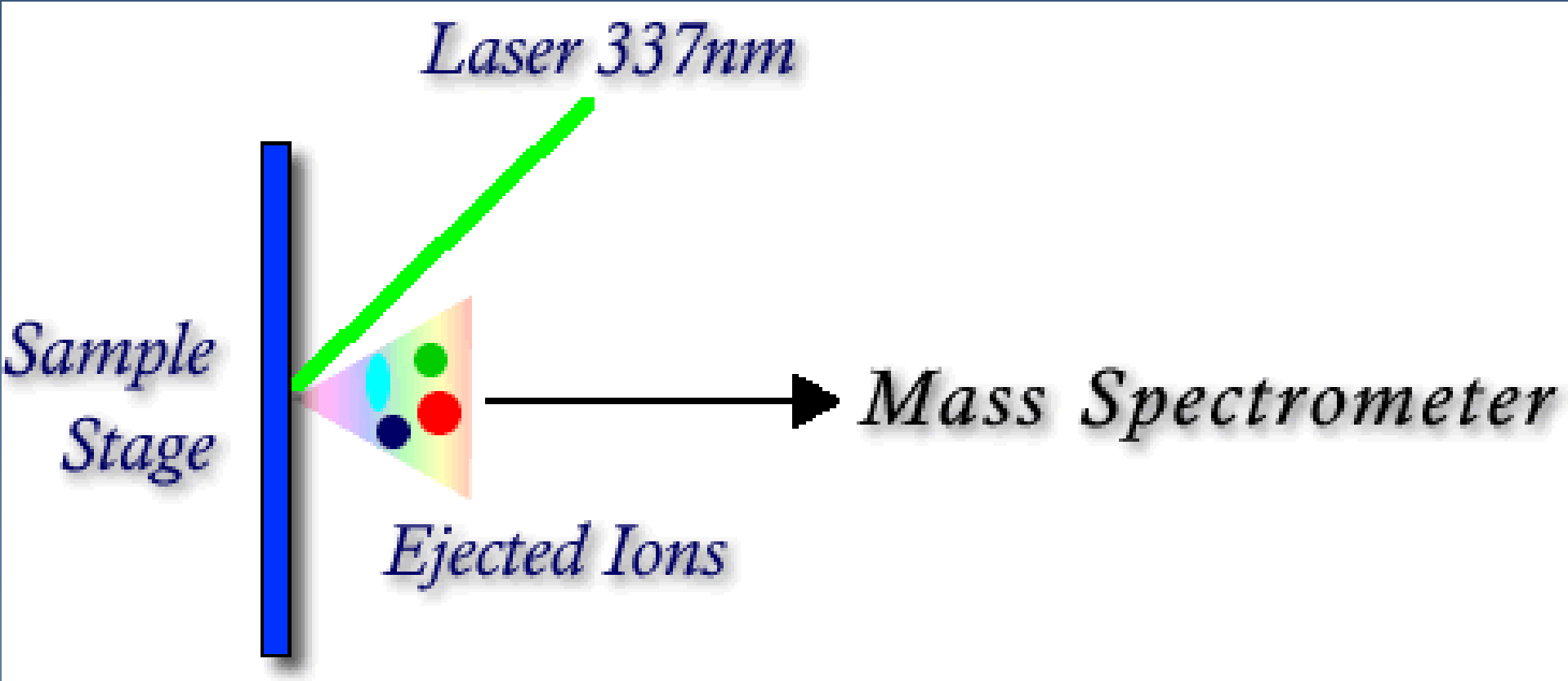
Schematic drawing of a mass spectrometer. (1) Highly charged protein droplets are formed by electrostatic dispersion of a protein solution through a glass capillary subjected to a high electric field (**electrospray**); This step can also be replaced by **m**atrix-**a**ssisted **d**esporption **i**onization (**MALDI**) (2) Protein droplets are accelerated by electrostatic forces in vacuum; (3) the time it takes for a protein to arrive at the mass spectrometer depends on its m/z. Thus an unknown protein mass can be obtained by comparing its m/z with that of a standard protein.



Matrix assisted laser disorbtion ionization (MALDI) of protein sample

Creates ions by excitation of molecules isolated from the energy of the laser by an energy absorbing matrix.

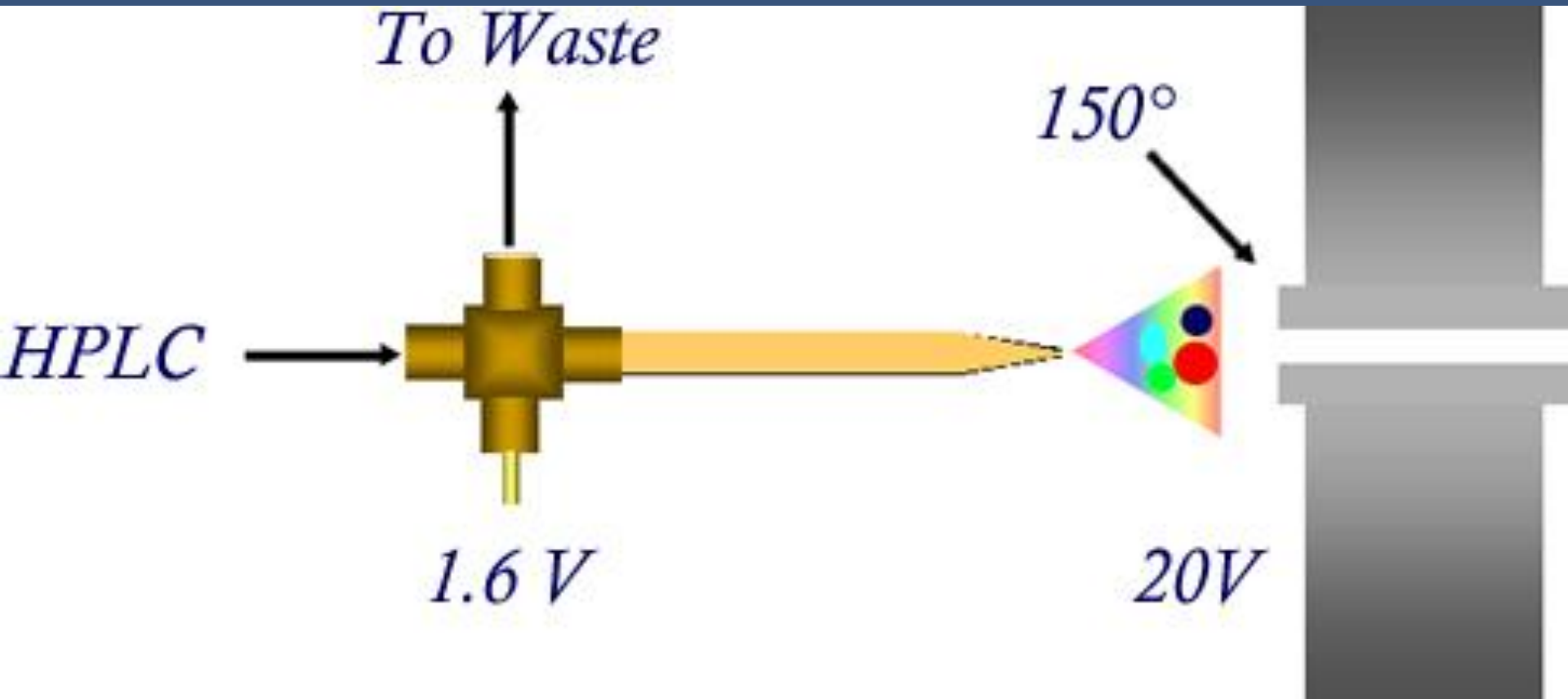
Ionized typically results in addition of one or several protons. A peptide of weight 1000 daltons will have a m/z value of 1001 after ionization by the addition of one proton and 501 with the addition of two (M+2H)/2.



Electrospray ionization (ESI)

Creates ions by application of a potential to a flowing liquid causing the liquid to charge and subsequently spray.

Creates very small droplets of solvent-containing analyte. Solvent is removed as the droplets enter the mass spectrometer by heat.

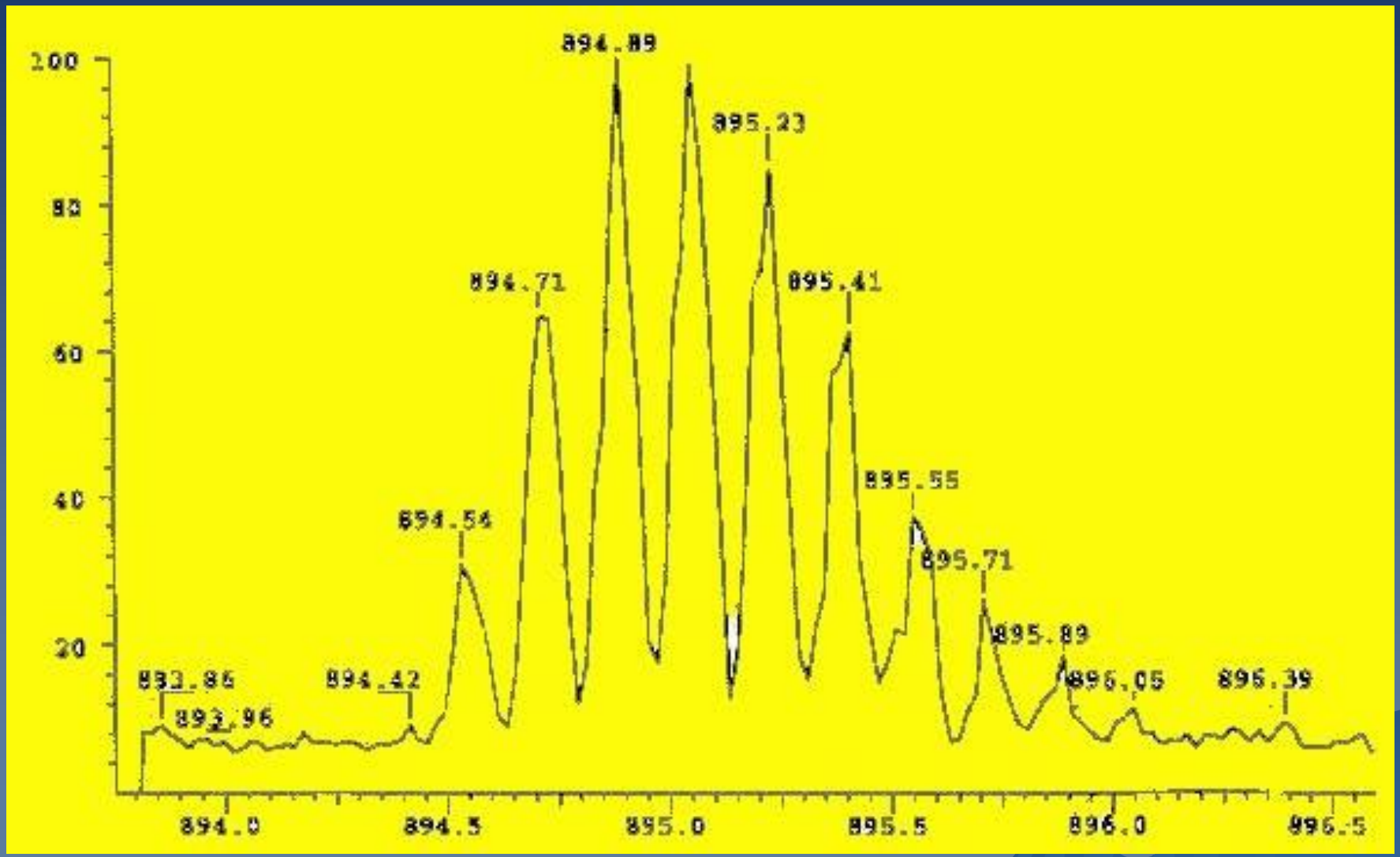


Typical nanoelectrospray source

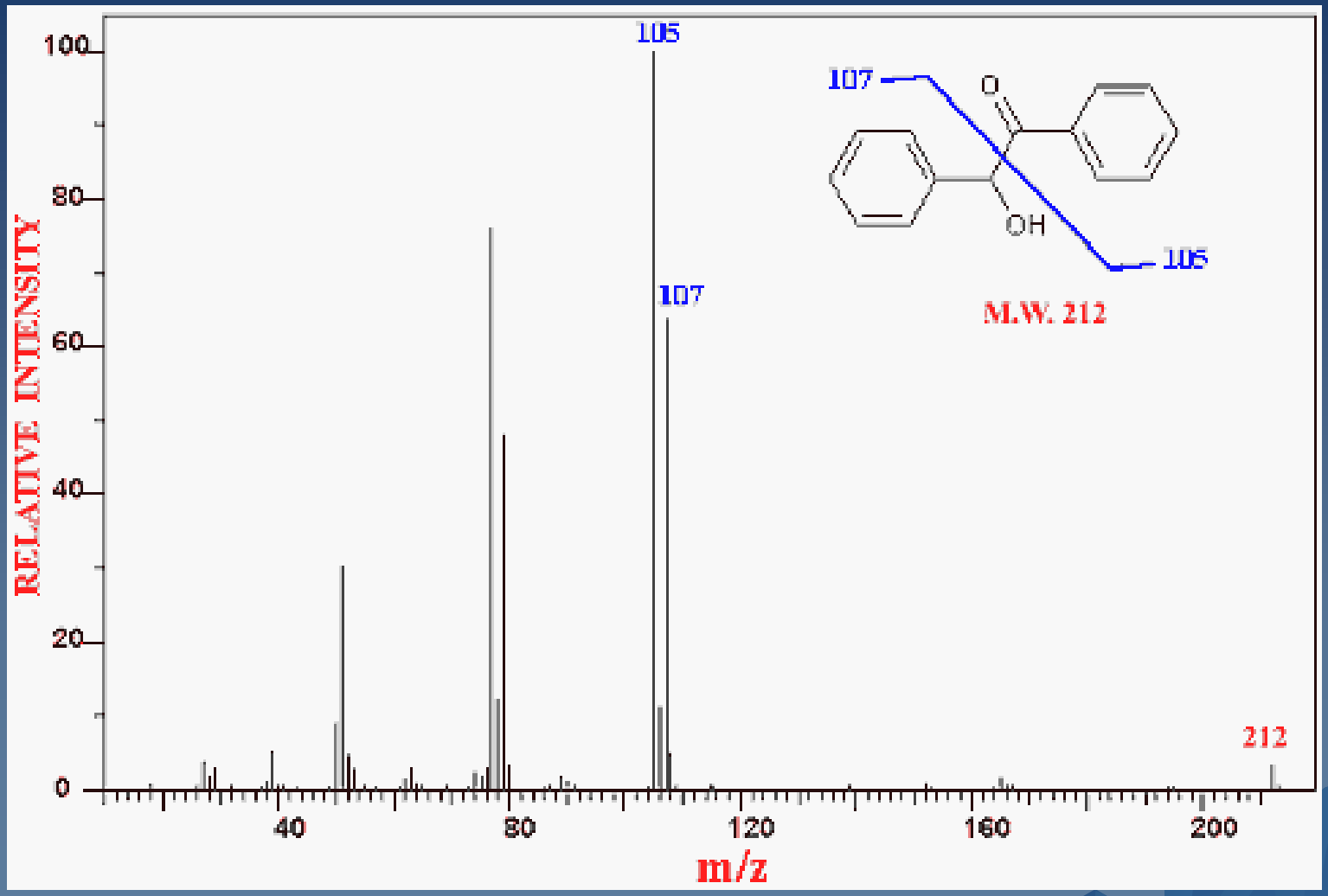
The mass analyzer & ion detector

- Once ions are created, individual mass-to-charge ratios (m/z) are separated by a second device, a mass analyzer, and transferred to the third device, an ion detector.
- A mass analyzer uses some physical property (e.g., electric or magnetic fields or time of flight) to separate ions of a particular m/z value, which then strike the ion detector.
- The magnitude of the current produced at the detector as a function of time (e.g., the physical field in the mass analyzer is changed as a function of time or the time it takes the ion to move a certain distance) is used to determine the m/z value of the ion.

Example mass spectrum

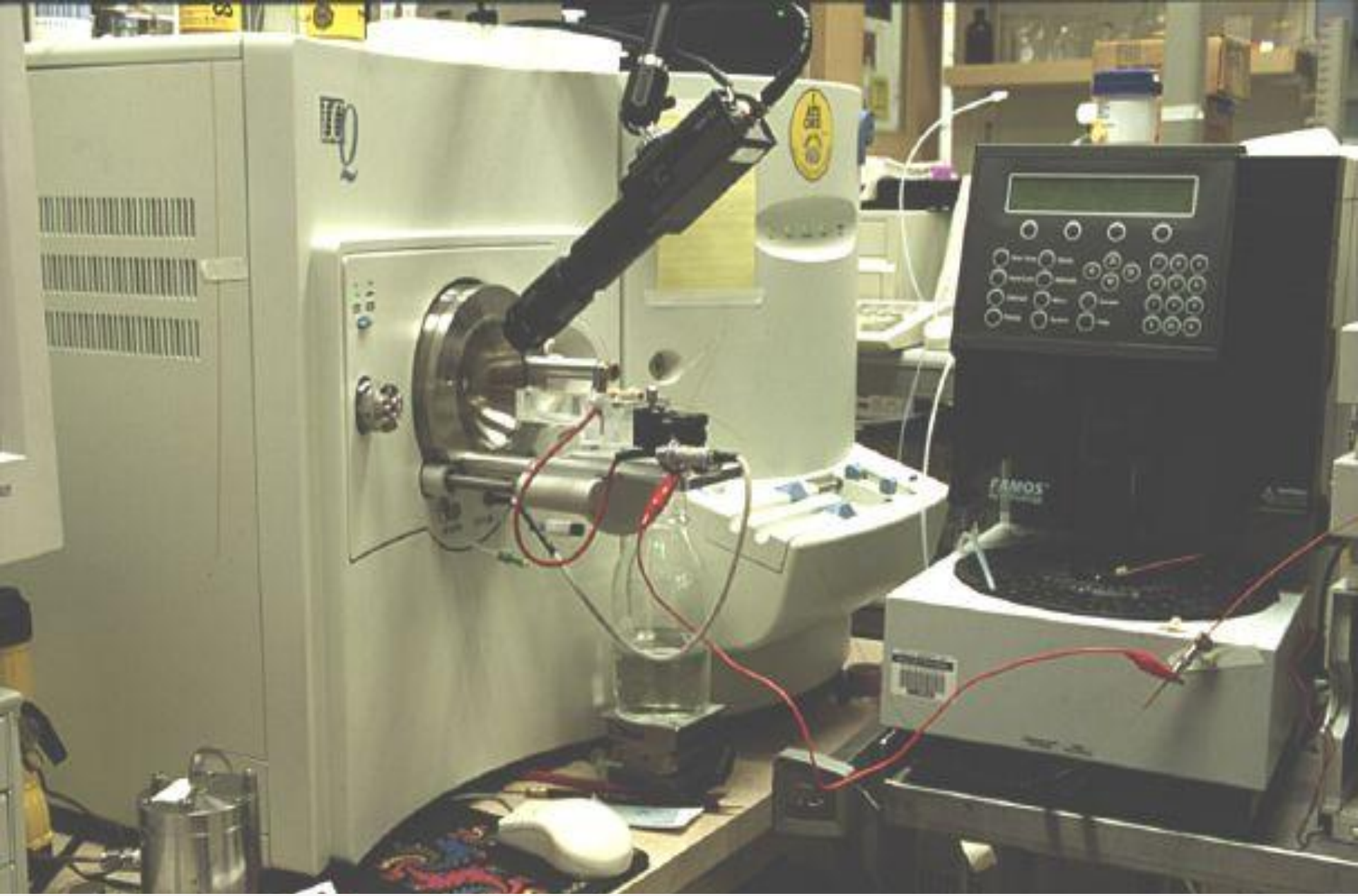


Not just proteins: benzoin (benzoic acid or balsam)



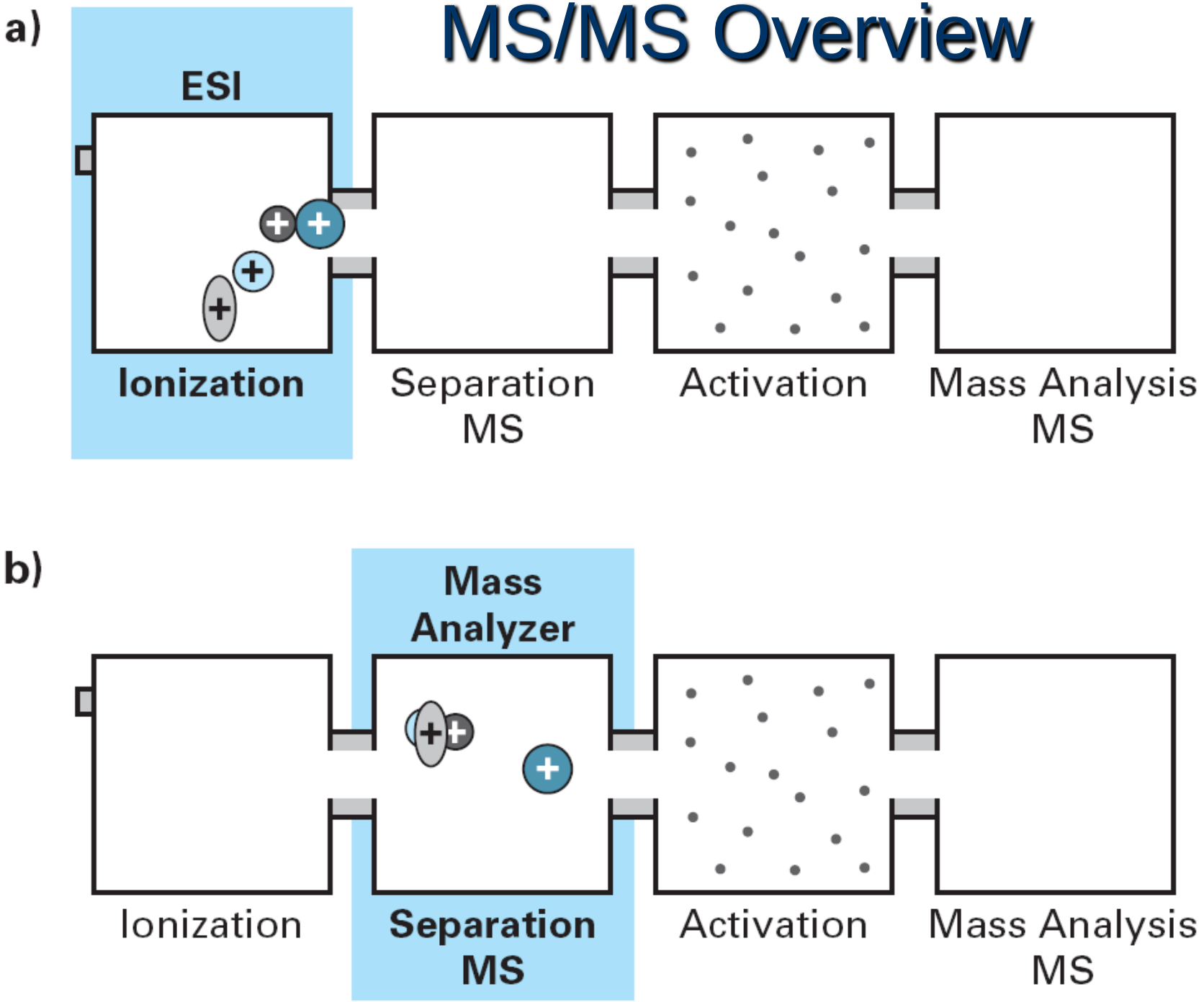
Tandem Mass Spectrometry (MS/MS)

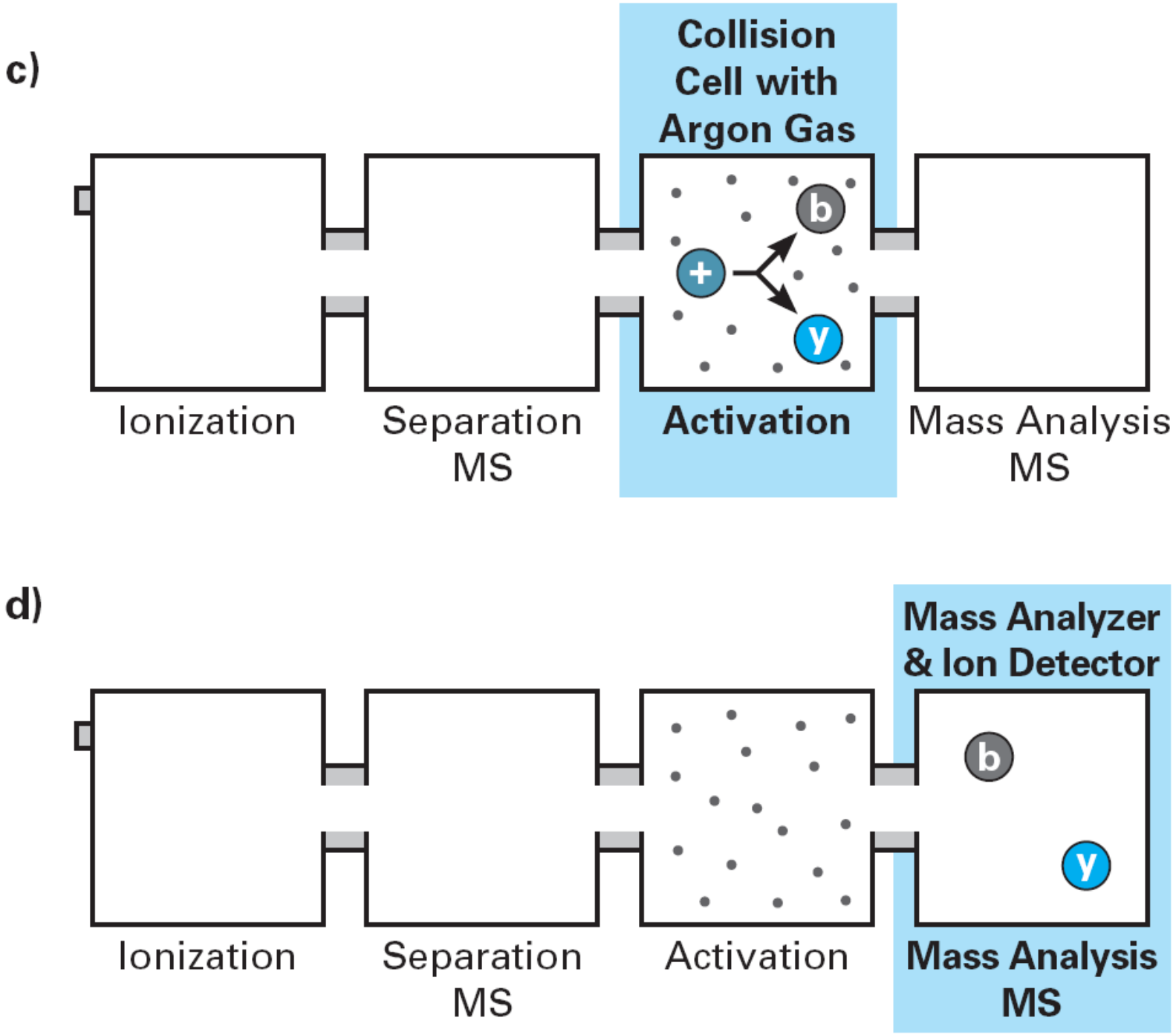
- Two mass specs separated by a collision cell.
- The first mass spectrometer is set to pass just one m/z value
- This ion enters the collision cell and collides with argon.
- The kinetic energy of ions is converted to vibrational energy and the ions fragment.
- The m/z values of fragment ions are then determined in the second mass spectrometer.
- Computer control: Typically, a scan of the mass range reveals several ions above a preset ion-abundance threshold. The computer signals MS/MS on each of the ions



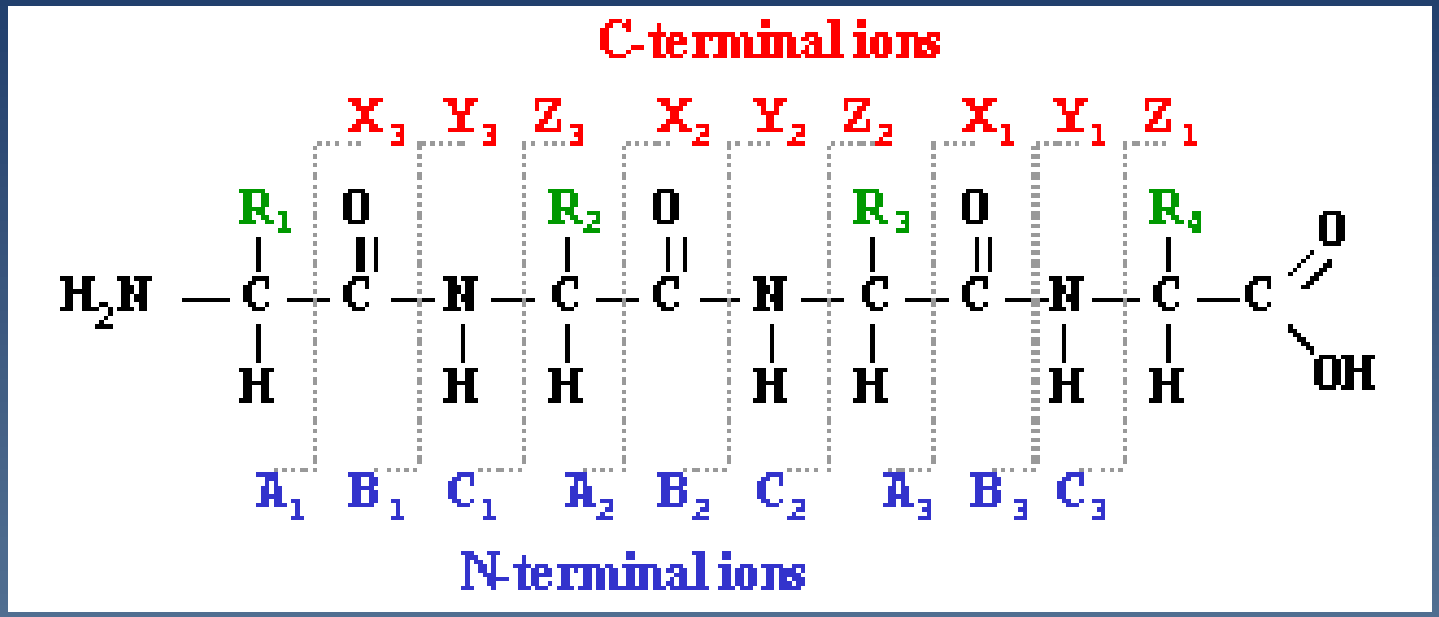
Tandem Mass Spec (MS/MS)

MS/MS Overview



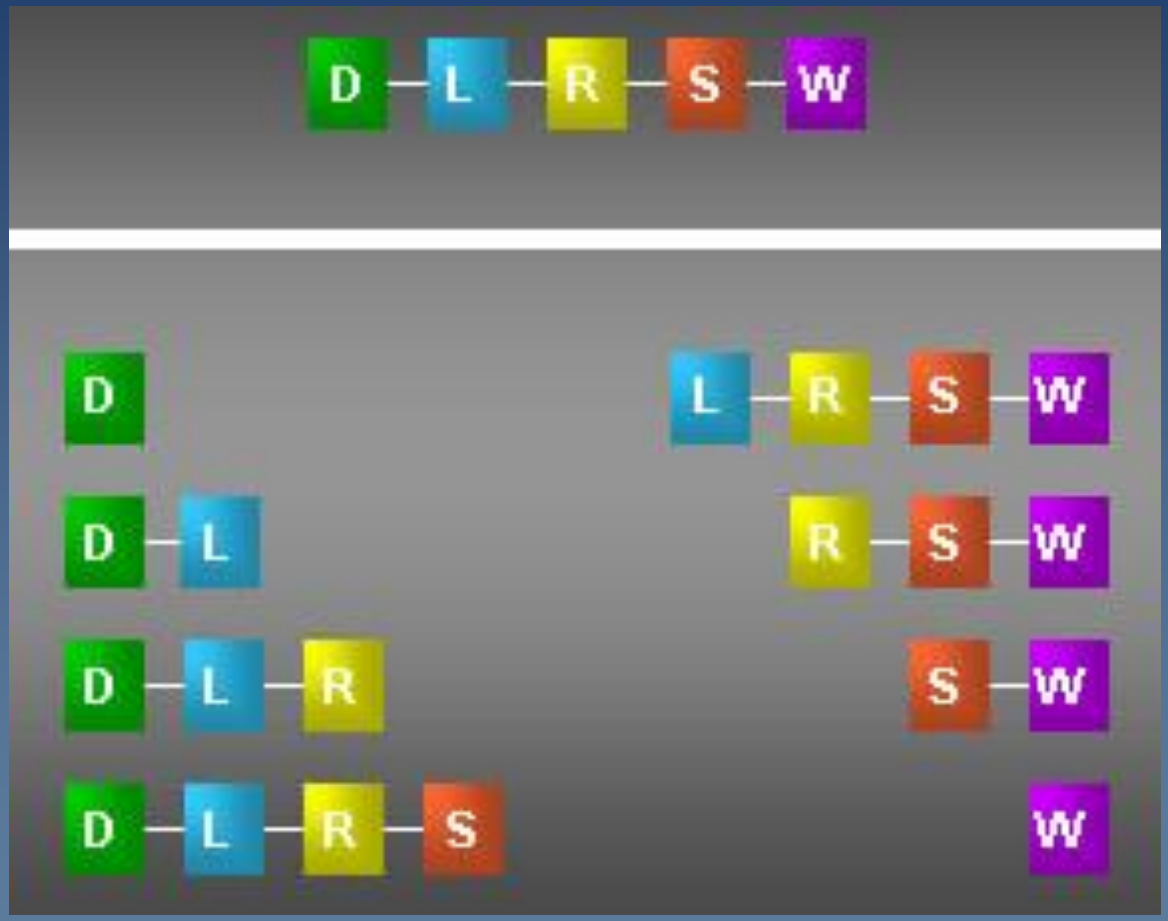

c)
Collision Cell with Argon Gas
b
+
y
Ionization
Separation MS
Activation
Mass Analysis MS
d)
Mass Analyzer & Ion Detector
b
y
Ionization
Separation MS
Activation
Mass Analysis MS

Peptide sequencing via MS/MS



- By using tandem mass spectrometry, fragmentation info determines the amino acid sequence of a peptide.
- Different bond cleavage positions are possible, known as (A,B,C) and (X,Y,Z) depending on whether the fragment contains the amino (N_2H) or carboxy (COOH) terminus
- The most common are b and y ions. These fragments are used to infer the original peptide sequence

B and Y ion fragmentation



b ions <----------> y ions

A raw fragmentation spectrum

By calculating the molecular weight difference between ions of the same type the sequence can be determined.

SEQUEST uses the fragmentation pattern to search through a complete protein database to identify the sequence which best fits the pattern.

a)

S–P–A–F–D–S–I–M–A–E–T–L–K
(protonated mass 1410.6)

Mass$^+$	b-ions	y-ions	Mass$^+$
81.1	S	PAFDSIMAETLK	1323.6
185.2	SP	AFDSIMAETLK	1226.4
256.3	SPA	FDSIMAETLK	1155.4
403.5	SPAF	DSIMAETLK	1008.2
518.5	SPAFD	SIMAETLK	893.1
605.6	SPAFDS	IMAETLK	806.0
718.8	SPAFDSI	MAETLK	692.3
850.0	SPAFDSIM	AETLK	561.7
921.1	SPAFDSIMA	ETLK	490.6
1050.2	SPAFDSIMAE	TLK	361.5
1151.3	SPAFDSIMAET	LK	260.4
1264.4	SPAFDSIMAETL	K	147.2

b)