Lecture (9/26/25)

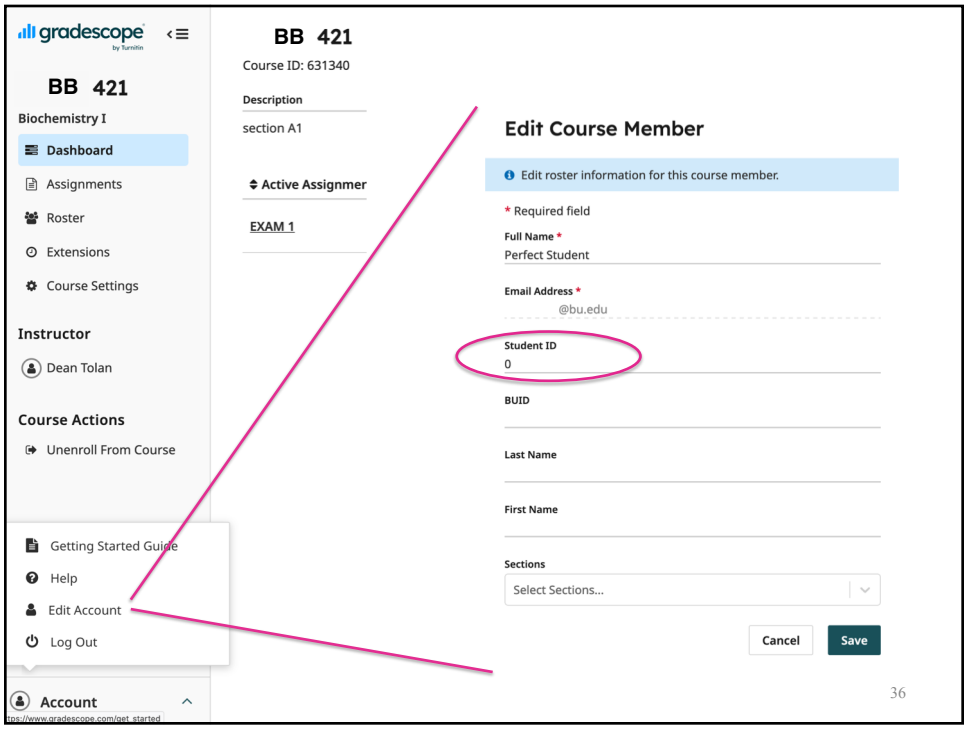
- Reading: Ch3; 90-93, Box 3-2
- Homework: #8

NEXT

- Reading: Ch4; 119-122, 125-127, 131-133;114-115, 120-121, 123-124
- Homework: #9

OUTLINE

I. Protein Structure
  A. Primary
    1. Determination
      a. Sequence determination; CHEMICAL
        i. Cleavage of peptides bonds
        ii. Amino acid composition and stoichiometry
        iii. Disrupt and determine number of chains;
        iv. Divide & Conquer;
        v. Edman Degradation
      b. Sequence determination; PHYSICAL
        i. Mass Spectrometry for proteins
        ii. Use of tandem MS/MS for sequence determination
        iii. Isolation of proteins by 2D PAGE; Isoelectric focusing x SDS-PAGE
      c. Sequence determination; BIOLOGICAL
        i. Genome sequenced
        ii. Bioinformatics to predict protein sequences in predicted genes
        iii. Use of CHEMICAL and/or PHYSICAL methods to get partial sequence



Determination of primary structure

1) Purify protein
2) Determine the amino-acid composition, including stoichiometry
3) Disrupt structure (2°, 3°, 4°, and disulfides)
4) Determine the number of peptide chains by counting number of amino terminal ends
5) Divide into fragments and determine sequence
6) Divide into different set of fragments and determine sequence
7) Determine overlaps and piece original sequence back together

38

Determination of primary structure

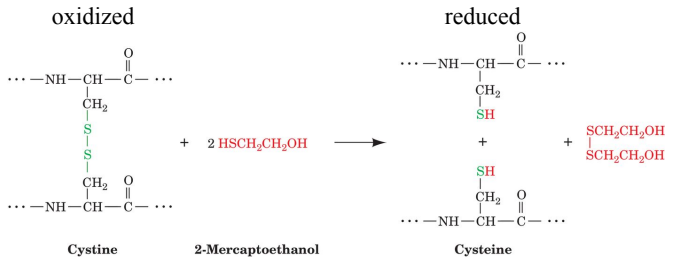
1) Purify protein
2) Determine the amino-acid composition, including stoichiometry
3) Disrupt structure (2°, 3°, 4°, and disulfides)
4) Determine the number of peptide chains by counting number of amino terminal ends
5) Divide into fragments and determine sequence
6) Divide into different set of fragments and determine sequence
7) Determine overlaps and piece original sequence back together

Disrupt structure (2°, 3°, 4°, and disulfides)

What holds these levels of structure together? …….non-covalent bonds (H-bonds, van der Waals, ionic, hydrophobic)

What have you used in the lab that might disrupt non-covalent bonds? …….Urea, **SDS**, pH extremes, heat, etc.

What about the covalent S-S bond? …….2-mercaptoethanol (β-mercaptoethanol, BME) or dithothreitol (DTT).

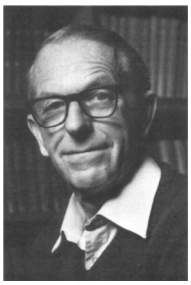


To keep disulfides from reforming…..
1) keep BME at high concentration in buffers
2) alkylate or oxidize the SH groups

Determination of primary structure

1) Purify protein
2) Determine the amino-acid composition, including stoichiometry
3) Disrupt structure (2°, 3°, 4°, and disulfides)
4) Determine the number of peptide chains by counting number of amino terminal ends
5) Divide into fragments and determine sequence
6) Divide into different set of fragments and determine sequence
7) Determine overlaps and piece original sequence back together

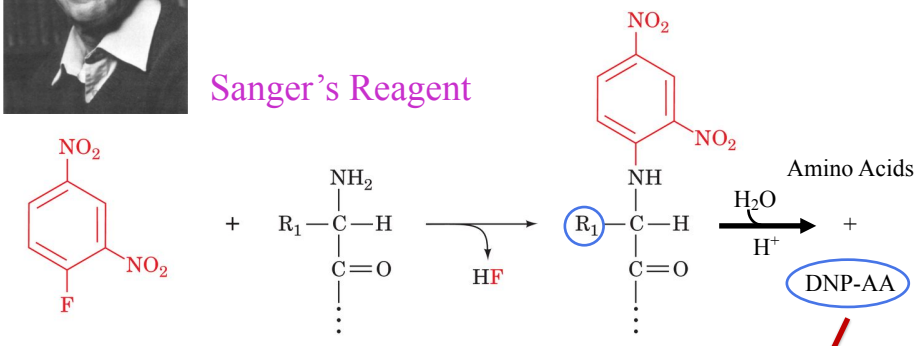
Determine the number of peptide chains by counting number of amino terminal ends



Frederick Sanger
(1918-2013)

Example:
small tripeptide
- AA comp = A,K,L
- C-term = K*
- DNP-A
- Sequence is A-L-K

Sanger's Reagent



2,4-Dinitrofluoro-benzene (DNFB) | **Polypeptide** | **DNP-Polypeptide**

Absorbs at 353 nm (yellow)

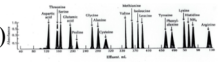
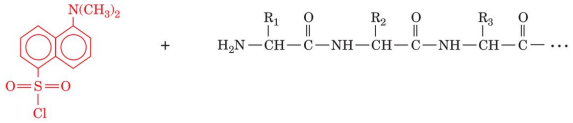


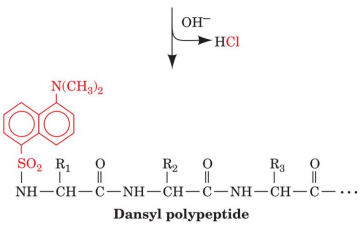
Fig 3-26

**Used carboxylpeptidase (more later)*

Determine the number of peptide chains by counting number of amino terminal ends



5-Dimethylamino-1-naphthalenesulfonyl chloride (dansyl chloride) | **Polypeptide**

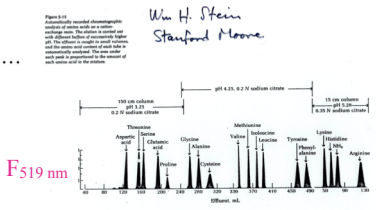


Dansylamino acid (fluorescent) | **Free amino acids**

Dansylation

Fluorescence detection is hundreds of times more sensitive: less protein needed

Wm H. Stein
Stanford Moore



Excitation at 330 nm
Emission at 519 nm

Determination of primary structure

1) Purify protein
2) Determine the amino-acid composition, including stoichiometry
3) Disrupt structure (2°, 3°, 4°, and disulfides)
4) Determine the number of peptide chains by counting number of amino terminal ends
5) Divide into fragments and determine sequence
6) Divide into different set of fragments and determine sequence
7) Determine overlaps and piece original sequence back together

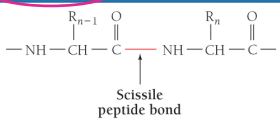
Determination of primary structure

1) Purify protein
2) Determine the amino-acid composition, including stoichiometry
3) Disrupt structure (2°, 3°, 4°, and disulfides)
4) Determine the number of peptide chains by counting number of amino terminal ends
5) Divide into fragments and determine sequence
6) Divide into different set of fragments and determine sequence
7) Determine overlaps and piece original sequence back together

Divide into fragments:

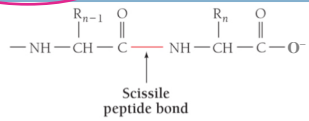
Proteolytic Cleavage

TABLE 5-4 Specificities of Various Endopeptidases



Enzyme	Source	Specificity	Comments
Trypsin	Bovine pancreas	R_{n-1} = positively charged residues: Arg, Lys; $R_n \neq$ Pro	Highly specific
Chymotrypsin	Bovine pancreas	R_{n-1} = bulky hydrophobic residues: Phe, Trp, Tyr; $R_n \neq$ Pro	Cleaves more slowly for R_{n-1} = Asn, His, Met, Leu
Elastase	Bovine pancreas	R_{n-1} = small neutral residues: Ala, Gly Ser, Val; $R_n \neq$ Pro	
Thermolysin	*Bacillus thermoproteolyticus*	R_n = Ile, Met, Phe, Trp, Tyr, Val; $R_{n-1} \neq$ Pro	Occasionally cleaves at R_n = Ala, Asp, His, Thr; heat stable
Pepsin	Bovine gastric mucosa	R_n = Leu, Phe, Trp, Tyr; $R_{n-1} \neq$ Pro	Also others; quite nonspecific; pH optimum = 2
Endopeptidase V8	*Staphylococcus aureus*	R_{n-1} = Glu	

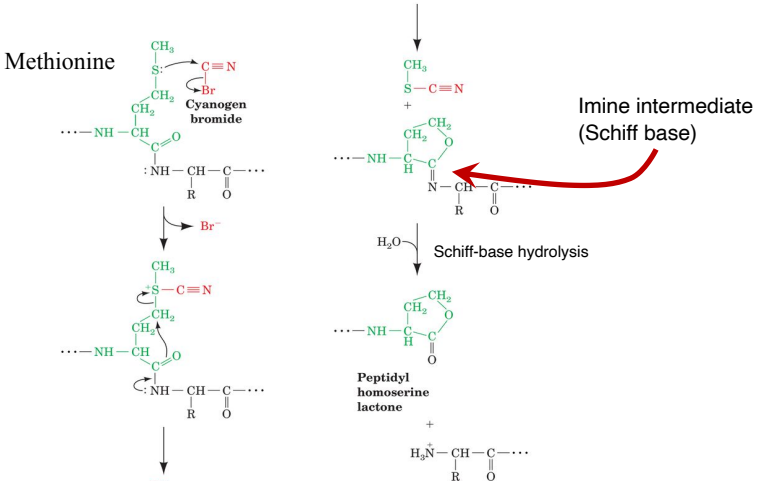
"TABLE 5-5" Specificities of Exopeptidase



Carboxylpeptidase A	Bovine pancreas	R_n = C-terminal; $R_{n-1} \neq$ Pro

Divide into fragments:

Cyanogen Bromide Cleavage

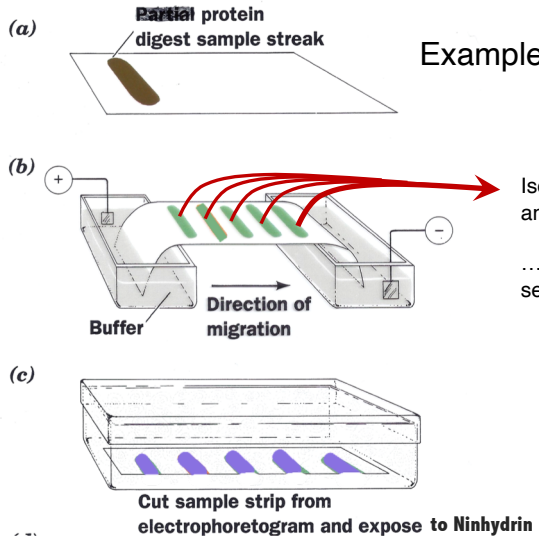


Similar result as an endopeptidase, but leaves a **homoserine lactone** as the C-terminal residue

WILEY

Separation and isolation of peptide fragments

- After fragmentation with either trypsin, chymotrypsin, or cyanogen bromide:



Example: paper electrophoresis

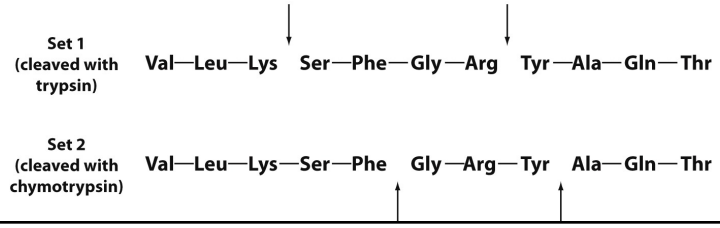
Also, TLC, silica gel, or mass spectrometry, etc.

Isolate each peptide and determine amino-acid composition, N-term, C-term

…..for large peptides, determine the sequence.

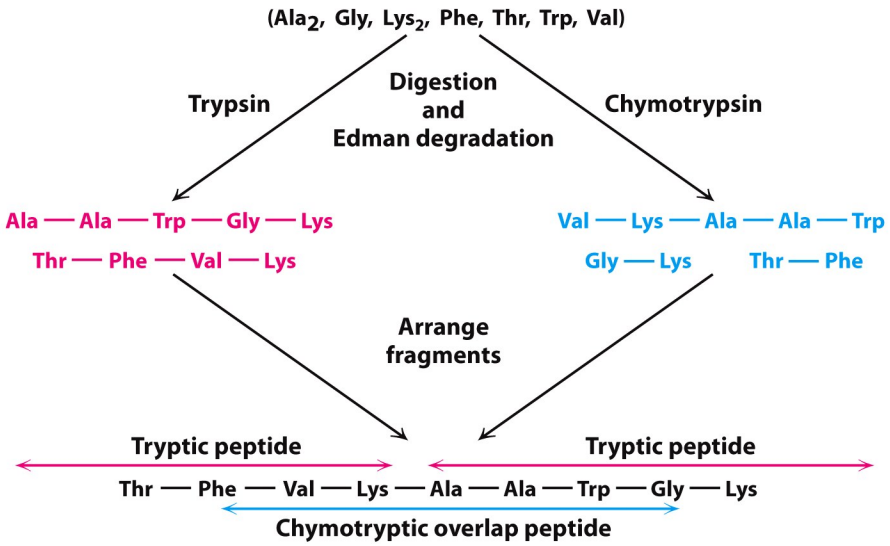
Determination of primary structure

- Determine amino-acid composition (AAC)
- Dansyl chloride or FDNB to determine amino-termini and number
- Proteases: Cleaves peptide bonds only after specific residues.
- Cleave protein with 2 different proteases and/or CNBr.
- Separate and isolate peptides: perform AAC and N-term
- Sequence larger fragments with Edman degradation. Piece together sequence from overlapping fragments.



52

Determination of primary structure: Divide & Conquer



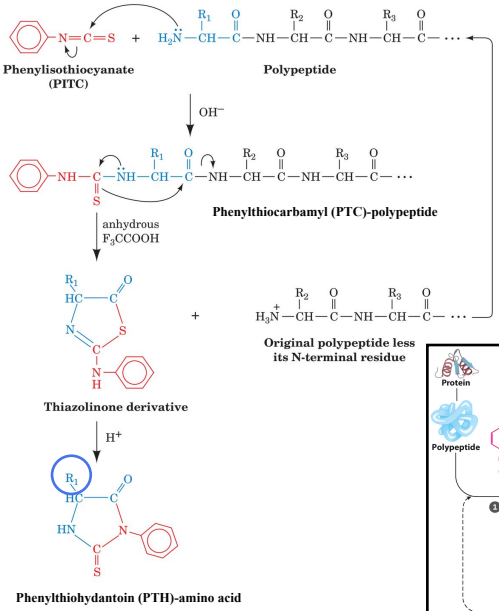
54

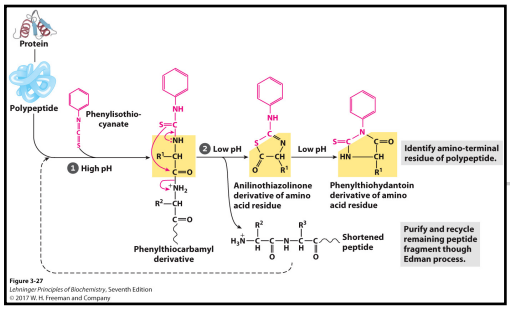
Determination of primary structure

1) Purify protein
2) Determine the amino-acid composition, including stoichiometry
3) Disrupt structure (2°, 3°, 4°, and disulfides)
4) Determine the number of peptide chains by counting number of amino terminal ends
5) Divide into fragments and determine sequence
6) Divide into different set of fragments and determine sequence
7) Determine overlaps and piece original sequence back together

55

Determine the Sequence: Edman degradation

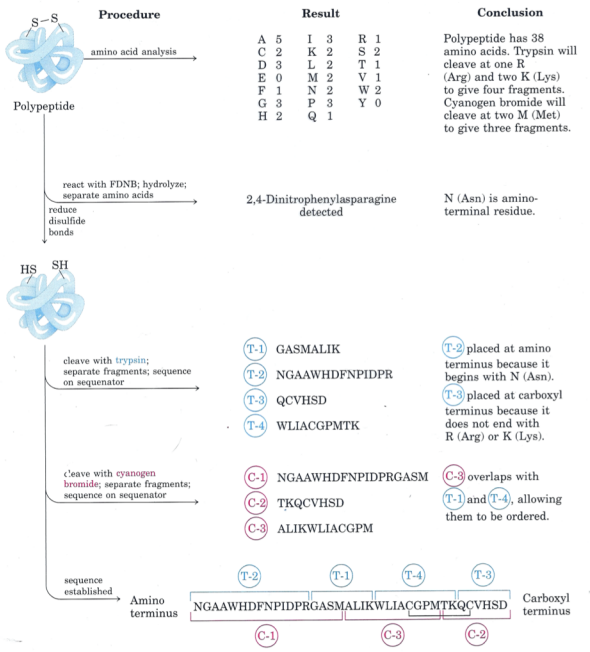




Determination of primary structure

1) Purify protein
2) Determine the amino-acid composition, including stoichiometry
3) Disrupt structure (2°, 3°, 4°, and disulfides)
4) Determine the number of peptide chains by counting number of amino terminal ends
5) Divide into fragments and determine sequence
6) Divide into different set of fragments and determine sequence
7) Determine overlaps and piece original sequence back together

57



Determine overlaps and piece original sequence back together

Determination of primary structure

THREE basic ways to know the primary structure. Only the CHEMICAL method will give the entire covalent structure, including any disulfide bonds. But other methods are more sensitive. One can classify these methods by:

CHEMICAL
PHYSICAL
"BIOINFORMATICAL"

We just went through the CHEMICAL.
The PHYSICAL method still requires the same strategy, including purification, fragmentation, chromatography, and alignment.

But, instead of an Edman degradation the use of tandem Mass Spectrometry (MS) is employed.

Lets look at the use of MS in biochemistry

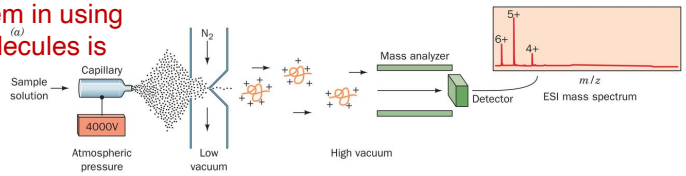
- Ions "fly" in a vacuum toward a target with a velocity ∝ z/m (charge-to-mass ratio)
- Molecules with higher charge and lower mass get detected first.
- Molecules with a lower charge and higher mass get detected last.
- Plotted as m/z to read peaks from left to right
- Instruments can distinguish molecules with same charge by < 1 Da

Determine the Sequence: Tandem MS

The major problem in using MS for macromolecules is getting them to "fly" in a vacuum with a charge.

TWO major methods:

1) Electro-Spray Ionization (ESI)
2) Matrix-Assisted Laser-Desorption Ionization (MALDI)



ESI

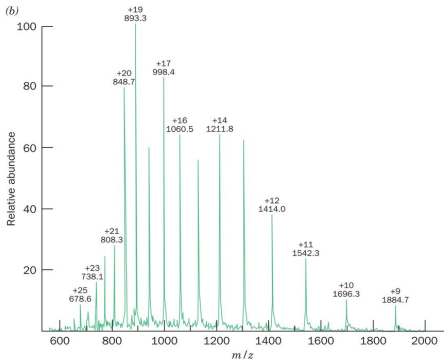


Figure 5-17

- Molecules with higher charge and lower mass get detected first.
- Molecules with a lower charge and higher mass get detected last.
- Plotted as m/z to read peaks from left to right (first detected to last)

What is MALDI?

Mass = (m/z) x z

$(m/z_A) \times z_A = (m/z_B) \times z_B$

$893.3z = 848.7(z+1)$

$893.3z = 848.7z+848.7$

$848.7 = z(893.3–848.7) = z44.6$

$848.7/44.6 = 19 = z$

Mass = 893.3x19 = 16,973

Mass = 848.7x20 = 16,974

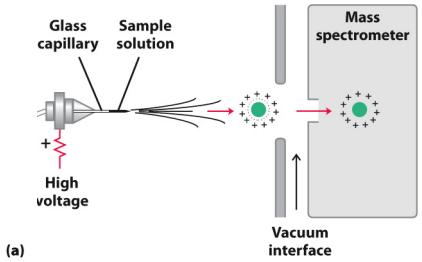
Mass = 1696.3x10 = 16,963

Determine the Sequence: Tandem MS

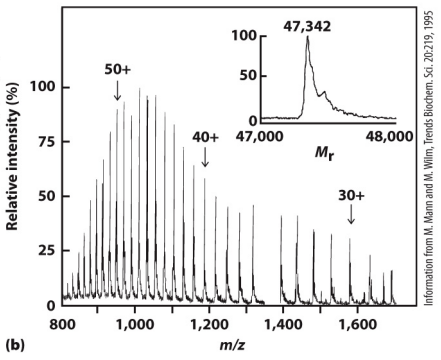
The major problem in using MS for macromolecules is getting them to "fly" in a vacuum with a charge.

TWO major methods:

1) Electro-Spray Ionization (ESI)
2) Matrix-Assisted Laser-Desorption Ionization (MALDI)



ESI



What is MALDI?

Mass = (m/z) x z

$(m/z_A) \times z_A = (m/z_B) \times z_B$

$893.3z = 848.7(z+1)$

$893.3z = 848.7z+848.7$

$848.7 = z(893.3–848.7) = z44.6$

$848.7/44.6 = 19 = z$

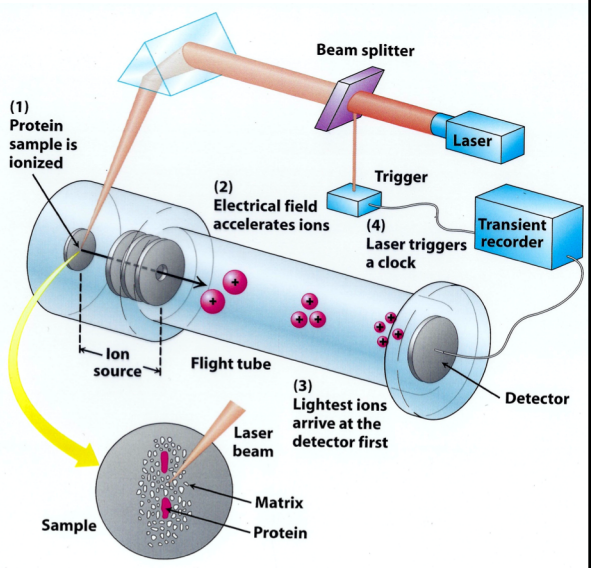
Mass = 893.3x19 = 16,973

Mass = 848.7x20 = 16,974

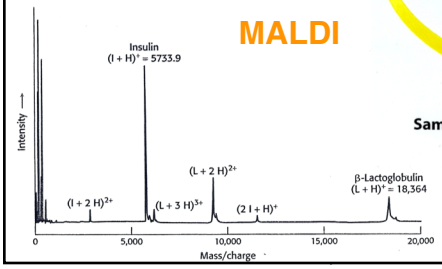
Mass = 1696.3x10 = 16,963

Determine the Sequence: Tandem MS

Matrix-Assisted Laser-Desorption Ionization (MALDI)

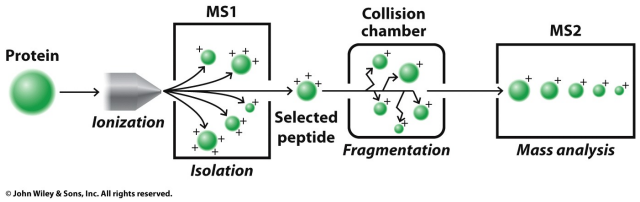


MALDI

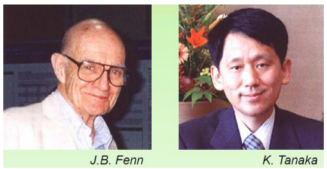


How do you use MS to get protein sequences?

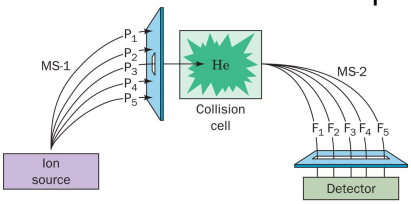
Determine the Sequence: Tandem MS



- Mass spectrometry uses mass-to-charge ratio of different ions to determine mass
- Tandem MS-MS: First selects a peptide, then fragmentation, and second determines mass of fragments
- By comparing results of all fragments, you can find masses that differ by mass of one amino acid to determine sequence



Nobel Prize in Chemistry 2002



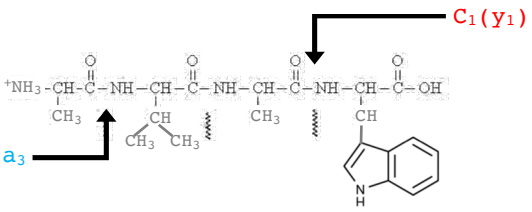
After Biemann, K. and Scoble, H.A., Science 237, 992 (1987)

Determine the Sequence: Tandem MS

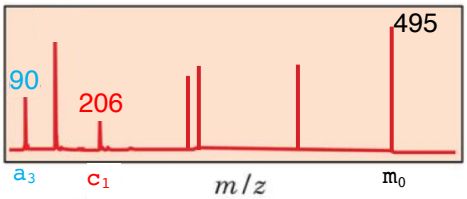
- EXAMPLE

AVAW (m_0=495)

AA	MW
Ala	90
Val	109
Trp	206



AVAW



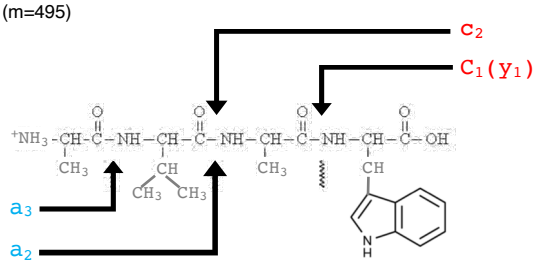
Conclusion:
Trp and Ala are at one end or the other

Determine the Sequence: Tandem MS

- EXAMPLE

AVAW (m=495)

AA	MW
Ala	90
Val	109
Trp	206



What is attached to the Ala?

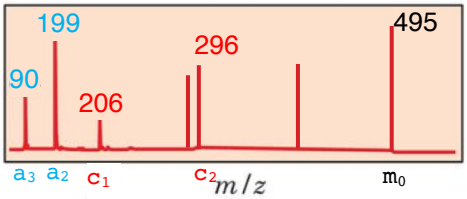
Either Val or Ala:
a_2 is 199 not 180

What is attached to the Trp?

Either Ala or Val:
c_2 is 286 not 315

See 199 (Ala-Val) & 286 (Ala-Trp)

AVAW

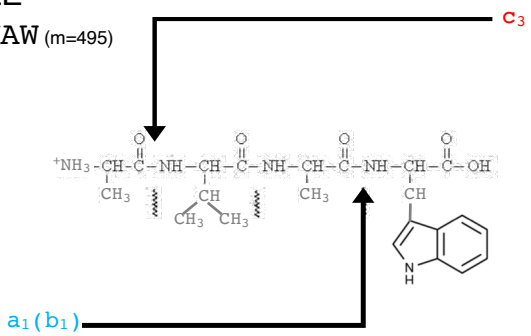


Determine the Sequence: Tandem MS

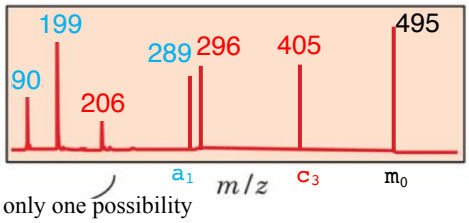
- EXAMPLE

AVAW (m=495)

AA	MW
Ala	90
Val	109
Trp	206



AVAW



CONCLUSION:
Either:
Trp-Ala-Val-Ala
Ala-Val-Ala-Trp

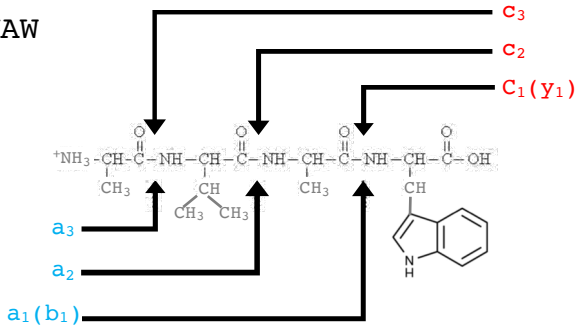
The 405 of c_3 supports Val-Ala-Trp
The 289 of a_1 supports the Ala-Val-Ala;
If you know the 405 is c and 289 is a, there is only one possibility

Determine the Sequence: Tandem MS

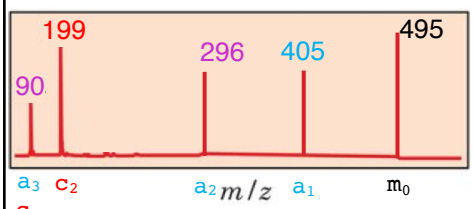
- EXAMPLE

AVAW

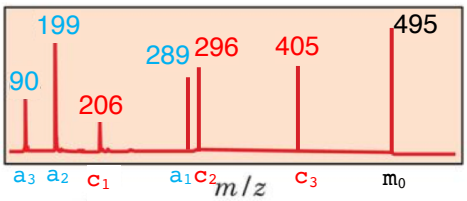
AA	MW
Ala	90
Val	109
Trp	206



AWVA



AVAW

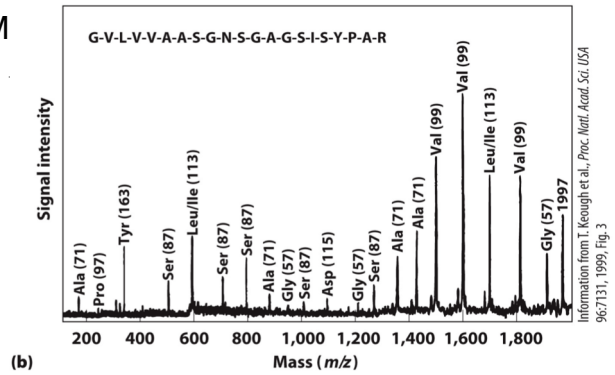


There is an issue with K & Q, which have the same MW!

Determine the Sequence: Tandem MS

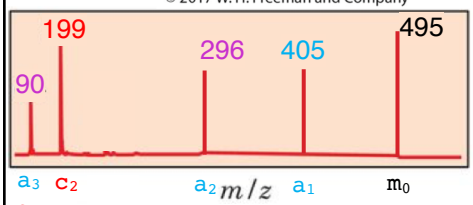
- EXAM

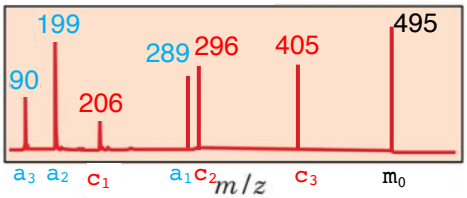
AA	MW
Ala	90
Val	109
Trp	206



Another method first selects for just the C-peptides

Figure 3-31
Lehninger Principles of Biochemistry, Seventh Edition
© 2017 W. H. Freeman and Company





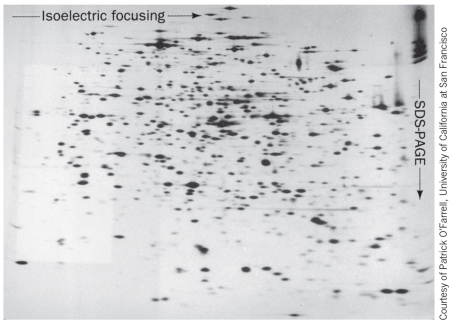
c3 There is an issue with K & Q and L & I, which have the same MW

Determination of primary structure

THREE basic ways to know the primary structure:

CHEMICAL
PHYSICAL
BIOINFOMATICAL

Edman Degradation requires >100 pmole (1-5 µg)
MS/MS requires >1-10 pmole (100-500 ng)



Determination of primary structure

THREE basic ways to know the primary structure:

CHEMICAL Edman Degradation requires >100 pmole (1-5 µg)

PHYSICAL MS/MS requires >1-10 pmole (100-500 ng)

BIOINFOMATICAL

We just went through the CHEMICAL and PHYSICAL.
The BIOINFORMATICAL method requires information from chemical or physical, but only a limited amount of sequence.

- Example: a sequence of 6 AA is only possible as one of 20^6 possible hexa-peptide sequences (1 of 64×10^6).
- There are no more than 50,000 protein-coding genes with ≤400 AA on average. This is ~20×10^6 possible unique sequences.
- So, a hexamer is likely to appear only once; an octomer even rarer.
- Once you have at least 6-8 AA sequence, you can compare that to all possible proteins encoded in the entirety of the gene sequences (genome) for a species for which the genome is known. Then using appropriate bioinformatic tools, you can derive the entire protein sequence.

There is one remaining issue: Where are the Disulfides, if any?

…...This requires chemical and/or physical methods