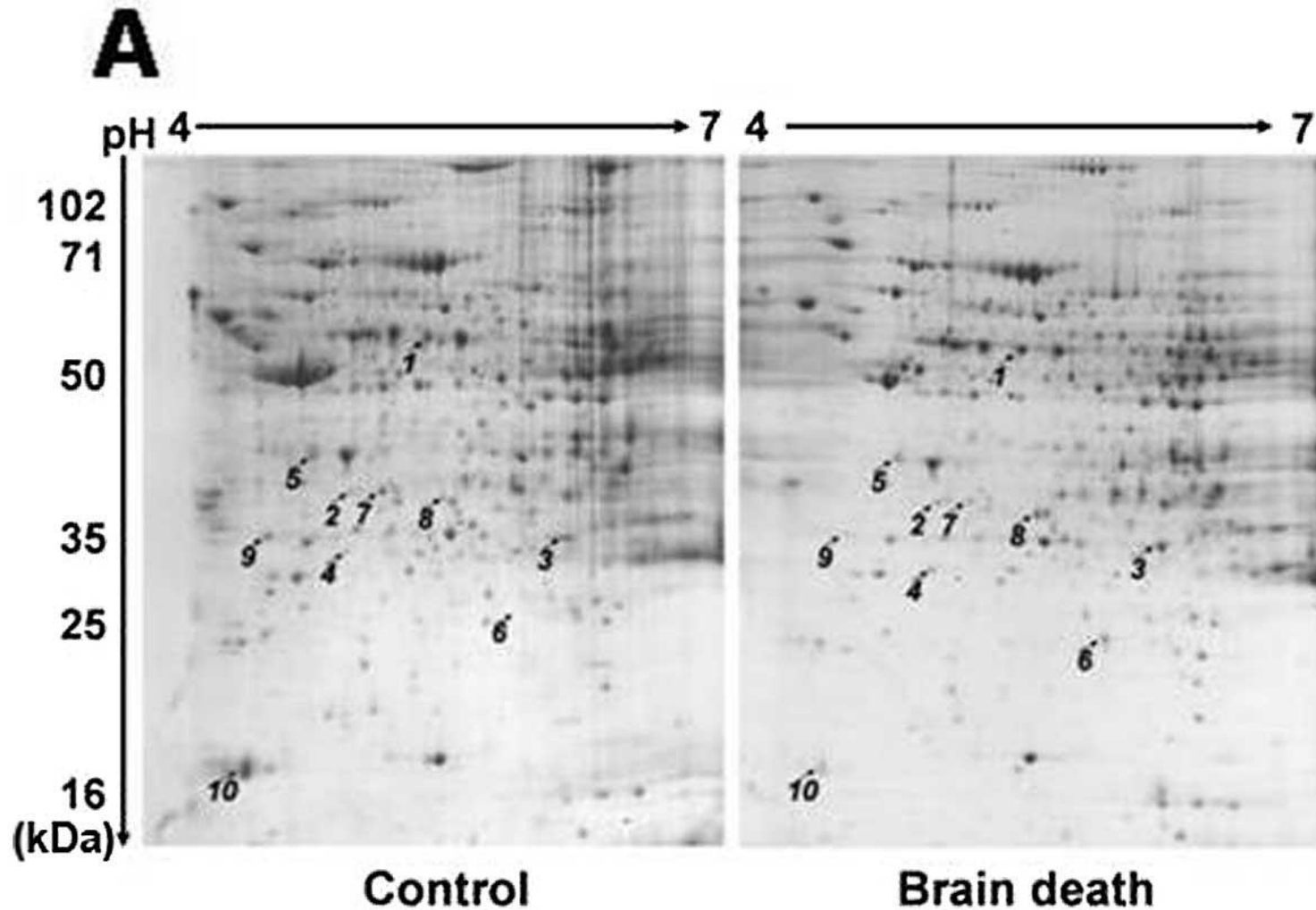# 질량분석 기반 단백체 연구 소개

2020. 06
김현우

What is Proteomics?
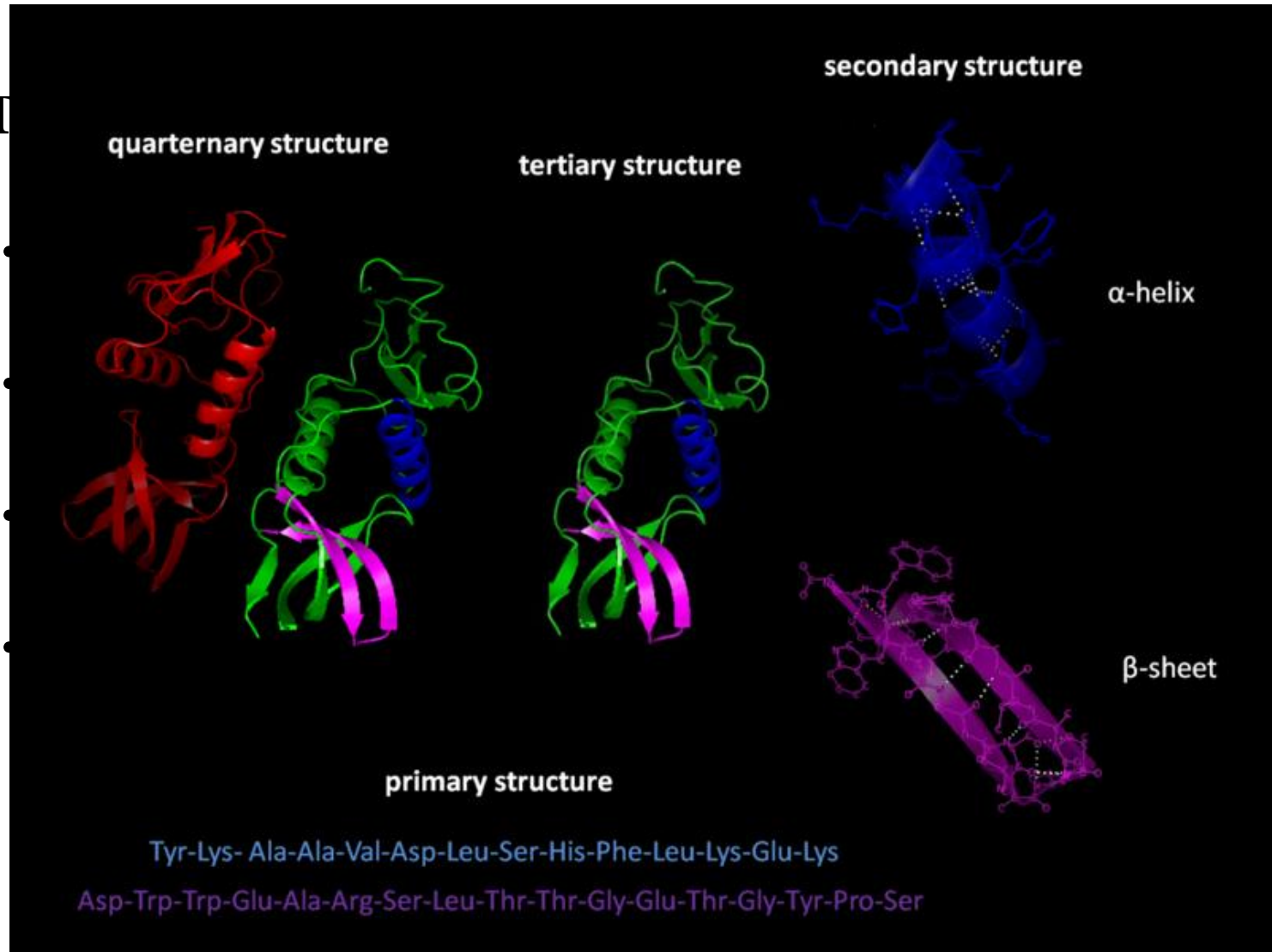
- **The Study of Proteins**

  - Expression

  - Structures

  - Functions

  - Interactions of protein
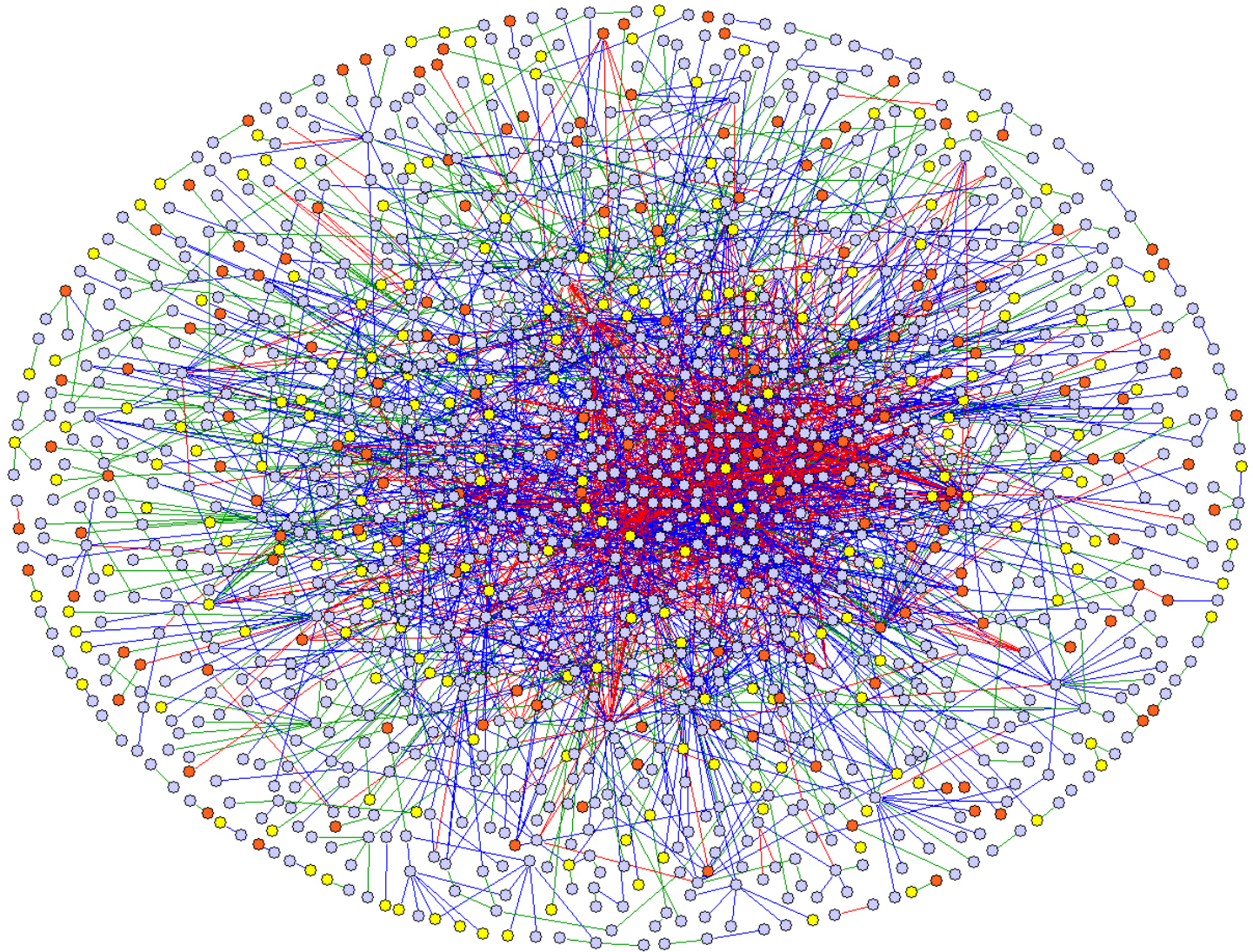
What is Proteomics?

- '

What is Proteomics?

What is Proteomics?

- T

  -

  -

  -

  -

What is a Protein?

- **Amino acid**

  - 생물의 몸을 구성하는 <u>단백질</u>의 기본 구성단위이다

- **Peptide**

  - 20가지의 서로 다른 <u>아미노산들이 화학 결합으로</u> 길게 연결된 것

What is a Protein?

- **Protein**

  - 여러 가지의 펩타이드 사슬이 4차 구조를 이루어 고유한 기능을 갖게 되었을 때 비로소 단백질이라고 불린다.

  - 일반적으로는 분자량이 비교적 작으면 펩타이드라고 하며, 분자량이 매우 크면 단백질이라고 한다.

What is a Protein?

- **Protein**
  - **An amino acid sequence**

- **Peptide**
  - An amino acid sequence
  - **Substring of the protein**

- **Amino acid**
  - **Single Character!**

Amino Acid	3-Letter	1-Letter
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamic acid	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

What is a Protein?

- **Protein**
  - An amino acid sequence
    - MEMEKEFEQIDKSGSWAAIYQDIRHEASDFPCRVAKLPKNKNRNRY RDVSPFDHSRKLHQEDNDYINASLIKMEEAQRSYILTQ….

What is a Protein?

- **Protein**
  - An amino acid sequence
    - MEMEKEFEQIDKSGSWAAIYQDIRHEASDFPCRVAKLPKNKNRNRY RDVSPFDHSRKLHQEDNDYINASLIKMEEAQRSYILTQ….
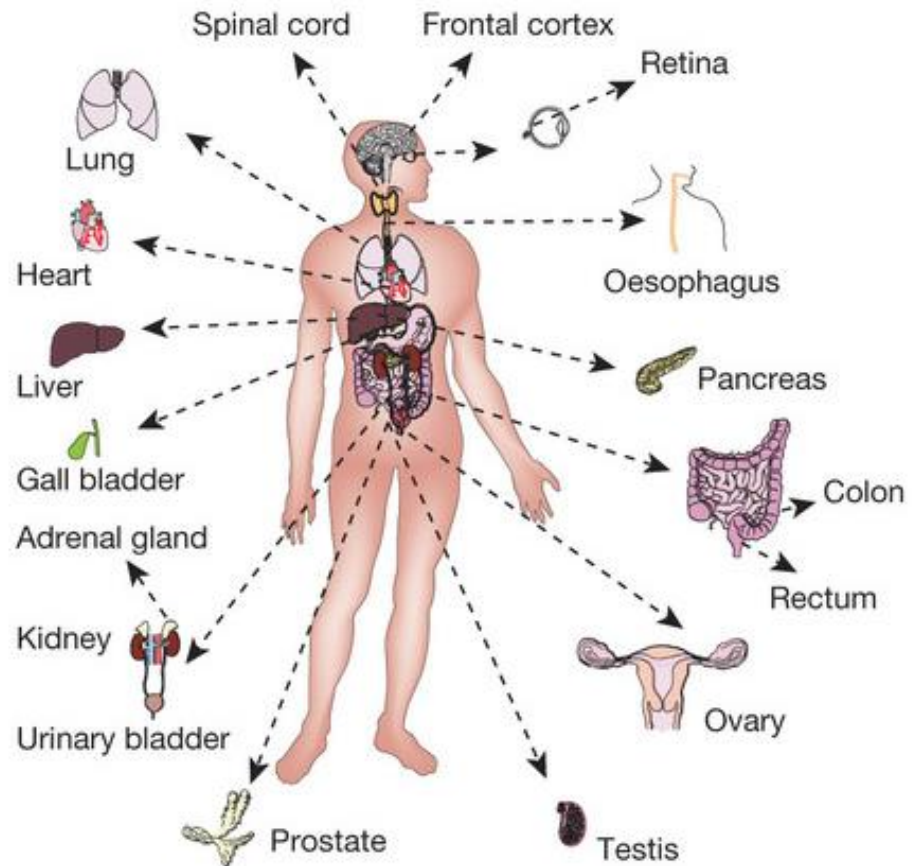
- **Peptide**
  - Substring of the protein
    - LPKNKNRNRYRDVSPFDHSR

What is a Protein?

- **Why are proteins studied?**

Proteomics

- **What is a problem?**

Proteomics

- **How to analyze a protein**

Mass Spectrometry

Proteomics
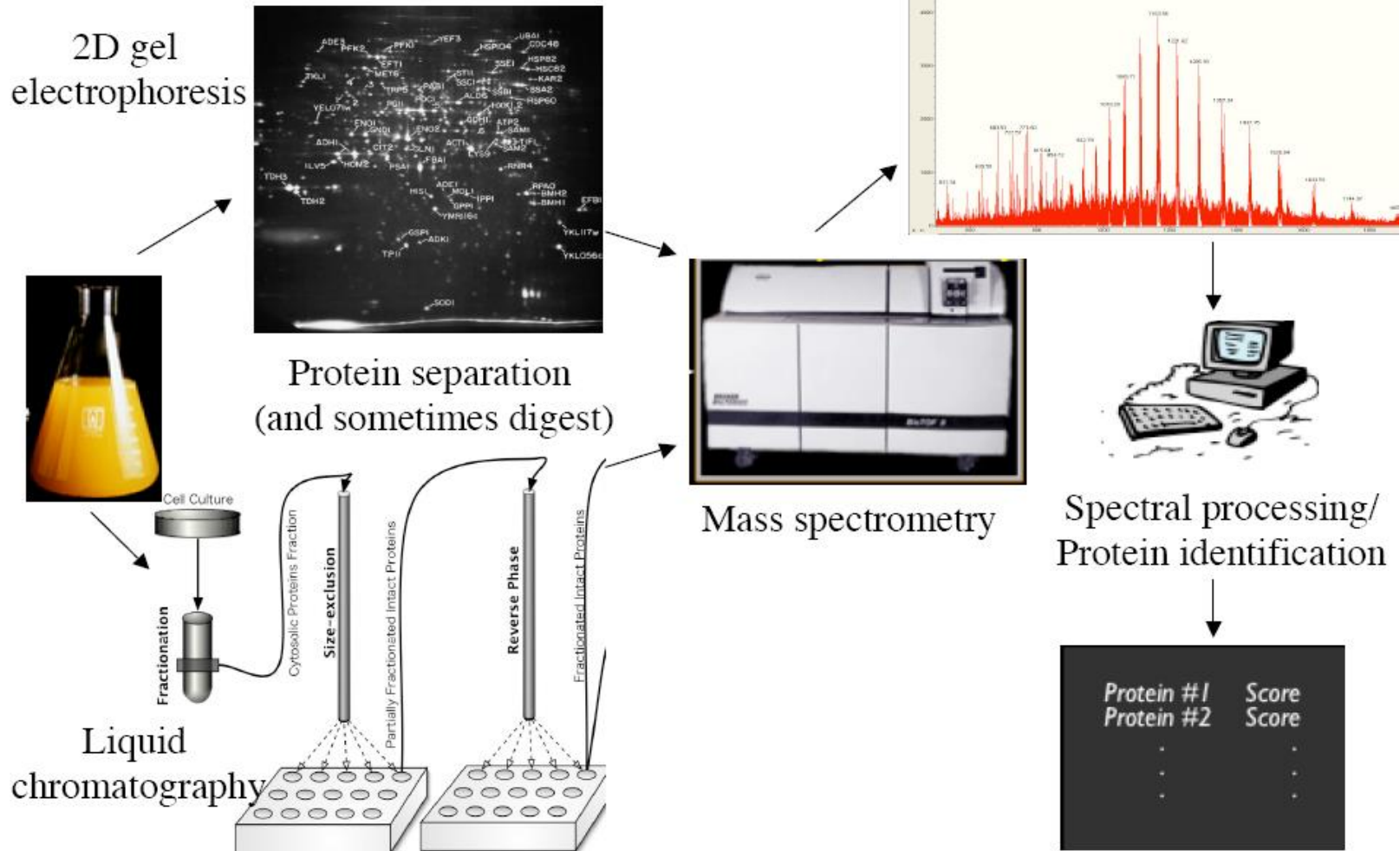
- **What is Mass Spectrometry?**

Proteomics

- **What is Mass Spectrometry?**

	AA Codes		AA Residue Composition	Mono.	Avg.	Structure		AA Codes		AA Residue Composition	Mono.	Avg.	Structure	
Glycine	Gly	G	C_2H_3NO	57.021464	57.05	-NH-CH₂-CO-	Asp	D	$C_4H_5NO_3$	115.02694	115.1		Aspartic acid
Alanine	Ala	A	C_3H_5NO	71.037114	71.08		Gln	Q	$C_5H_8N_2O_2$	128.05858	128.1		Glutamine
Serine	Ser	S	$C_3H_5NO_2$	87.032029	87.08		Lys	K	$C_6H_{12}N_2O$	128.09496	128.2		Lysine
Proline	Pro	P	C_5H_7NO	97.052764	97.12		Glu	E	$C_5H_7NO_3$	129.04259	129.1		Glutamic acid
Valine	Val	V	C_5H_9NO	99.068414	99.07		Met	M	C_5H_9NOS	131.04048	131.2		Methionine
Threonine	Thr	T	$C_4H_7NO_2$	101.04768	101.1		His	H	$C_6H_7N_3O$	137.05891	137.1		Histidine
Cysteine	Cys	C	C_3H_5NOS	103.00919	103.1		Phe	F	C_9H_9NO	147.06841	147.2		Phenylalanine
Leucine	Leu	L	$C_6H_{11}NO$	113.08406	113.2		Arg	R	$C_6H_{12}N_4O$	156.10111	156.2		Arginine
Isoleucine	Ile	I	$C_6H_{11}NO$	113.08406	113.2		Tyr	Y	$C_9H_9NO_2$	163.06333	163.2		Tyrosine
Asparagine	Asn	N	$C_4H_6N_2O_2$	114.04293	114.1		Trp	W	$C_{11}H_{10}N_2O$	186.07931	186.2		Tryptophane

Proteomics

- 2D gel electrophoresis

Protein separation (and sometimes digest)

Liquid chromatography

Cell Culture

Fractionation

Cytosolic Proteins Fraction

Size-exclusion

Partially Fractionated Intact Proteins

Reverse Phase

Fractionated Intact Proteins

Mass spectrometry

Spectral processing/ Protein identification

Protein #1   Score
Protein #2   Score

Proteomics

•

Ion Source

Electrospray Ionization (ESI)
Matrix-assisted laser desorption/ioniation (MALDI)

Mass analyzer

Ion trap
Time of flight (TOF)
Quadrupole
Fourier transform (FT-MS)
(and combinations of these)
Analytes separated by mass (M) and charge (z)
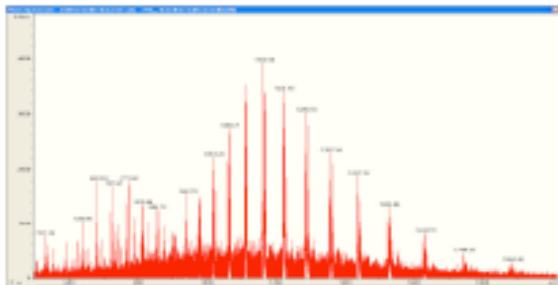
Spectrum of intact proteins,
peptides, or peptide fragments

Proteomics

- **Bottom-up method**

  - Protein digestion

  - Mass spectrometry analysis

  - Peptide identification using MS/MS

  - Protein Assembly using identified peptides

Digesting the Protein

- **Protein digestion**

  - Use proteases

    - Trypsin, Lys-C, Asp-N, Glu-C, etc.

Digesting the Protein

- **Protein digestion**

  - Trypsin

    - Aggressive and stable protease

    - It cleaves proteins very specifically on the carboxy-terminal side (right side) of arginine(R) and lysine(K) residues.

Digesting the Protein

- **N-term & C-term**

  - Protein
    - MEMEKEFEQIDKSGSWAAIYQDIRHEASDFPCRVAKLPKNKNRNRY RDVSPFDHSRKLHQEDNDYINASLIKMEEAQRSYILTQ

  - Peptide
    - LPKNKNRNRYRDVSPFDHSR

  - Amino acid
    - R

Digesting the Protein

- **Protein digestion**

    - Trypsin

        - Aggressive and stable protease

        - It cleaves proteins very specifically on the carboxy-terminal side (right side) of arginine(R) and lysine(K) residues.
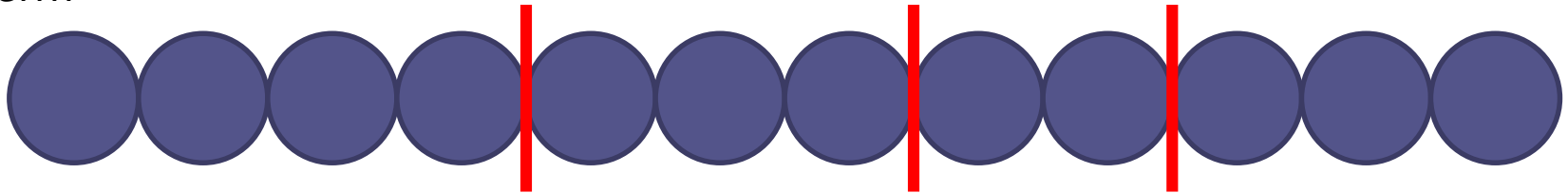

N-terminal
(Amino-terminal)

C-terminal
(Carboxy-terminal)
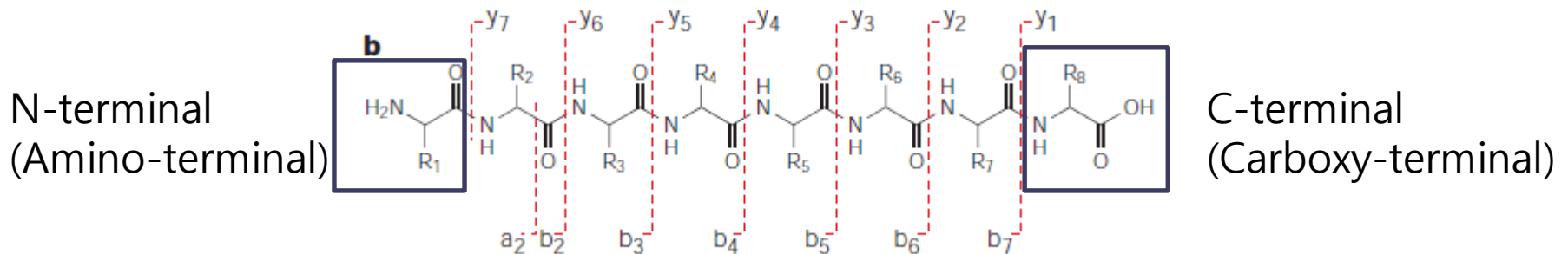
Digesting the Protein

- **Protein digestion(Trypsin)**

  - Protein
    - MEMEKEFEQID**K**SGSWAAIYQDI**R**HEASDFPCRVA**K**LP**K**N**K**N**R**N**R**Y**R**DVSPFDHS**RK**LHQEDNDYINASLI**K**MEEAQ**R**SYILTQ

  - Peptides
    - MEMEKEFEQID**K**
    - SGSWAAIYQDI**R**
    - HEASDFPCRVA**K**
    - LP**K**
    - **K**
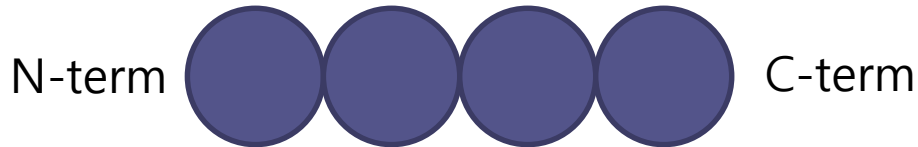
Digesting the Protein

- **Protein digestion(Trypsin)**

  - Protein (Quiz)
    - MTVYAVTGGAEFLGRYIVKLLISADDVQEIRVINVVEDPQPLVSKVK VINYIQCDINDLIR
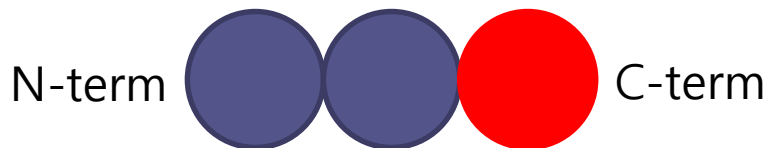
  - Peptides ?

Digesting the Protein

- **Protein digestion(Trypsin)**

  - Protein (Quiz)
    - MTVYAVTGGAEFLGRYIVKLLISADDVQEIRVINVVEDPQPLVSKVK VINYIQCDINDLIR

  - Peptides ?
    - MTVYAVTGGAEFLGR
    - YIVKLLISADDVQEIR
    - VINVVEDPQPLVSK
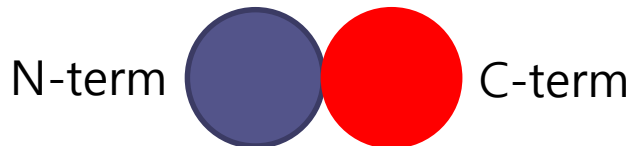   - VK
    - VINYIQCDINDLIR

  - 실습 (?)

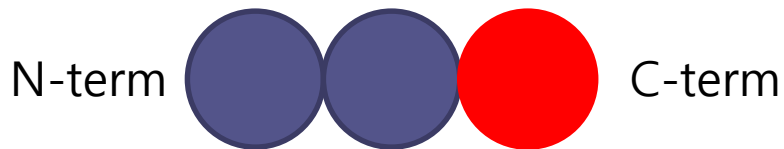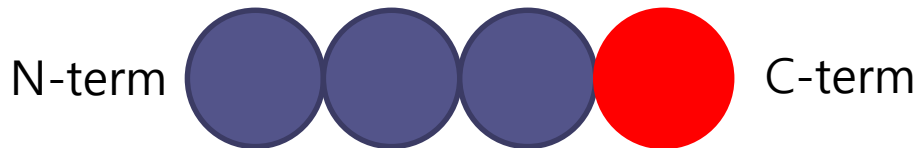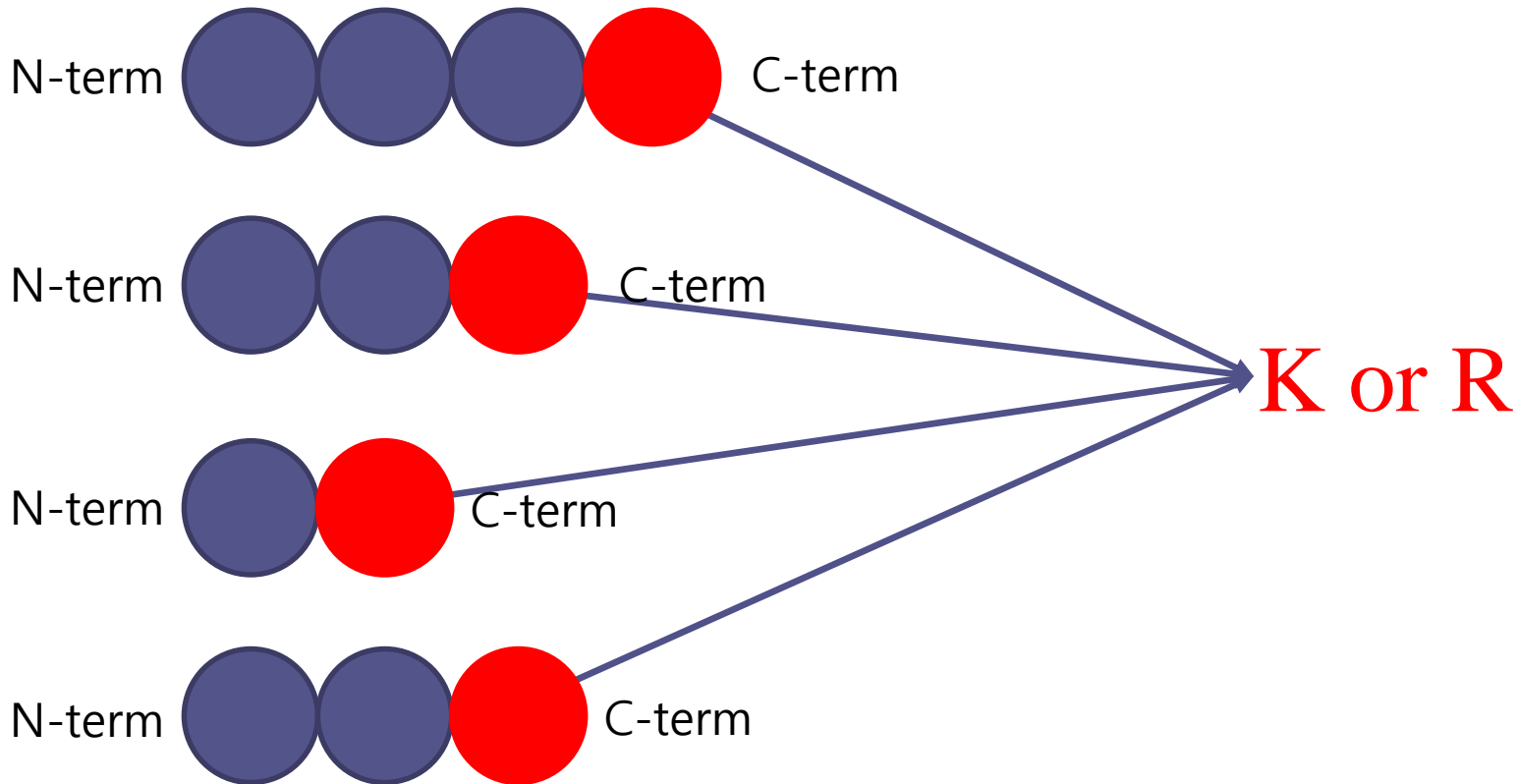# Digesting the Protein
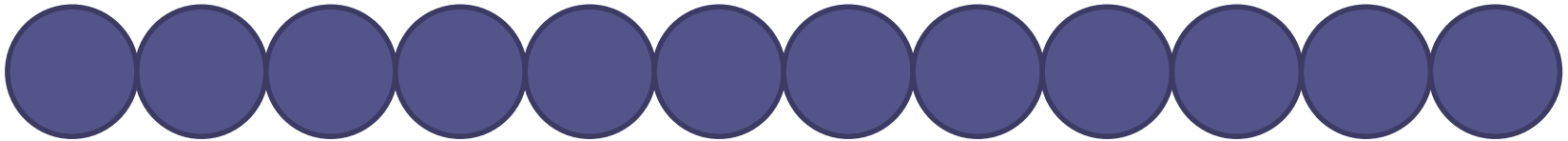
- **Protein digestion(Trypsin)**

Digesting the Protein

- **Protein digestion(Trypsin)**

N-term   C-term
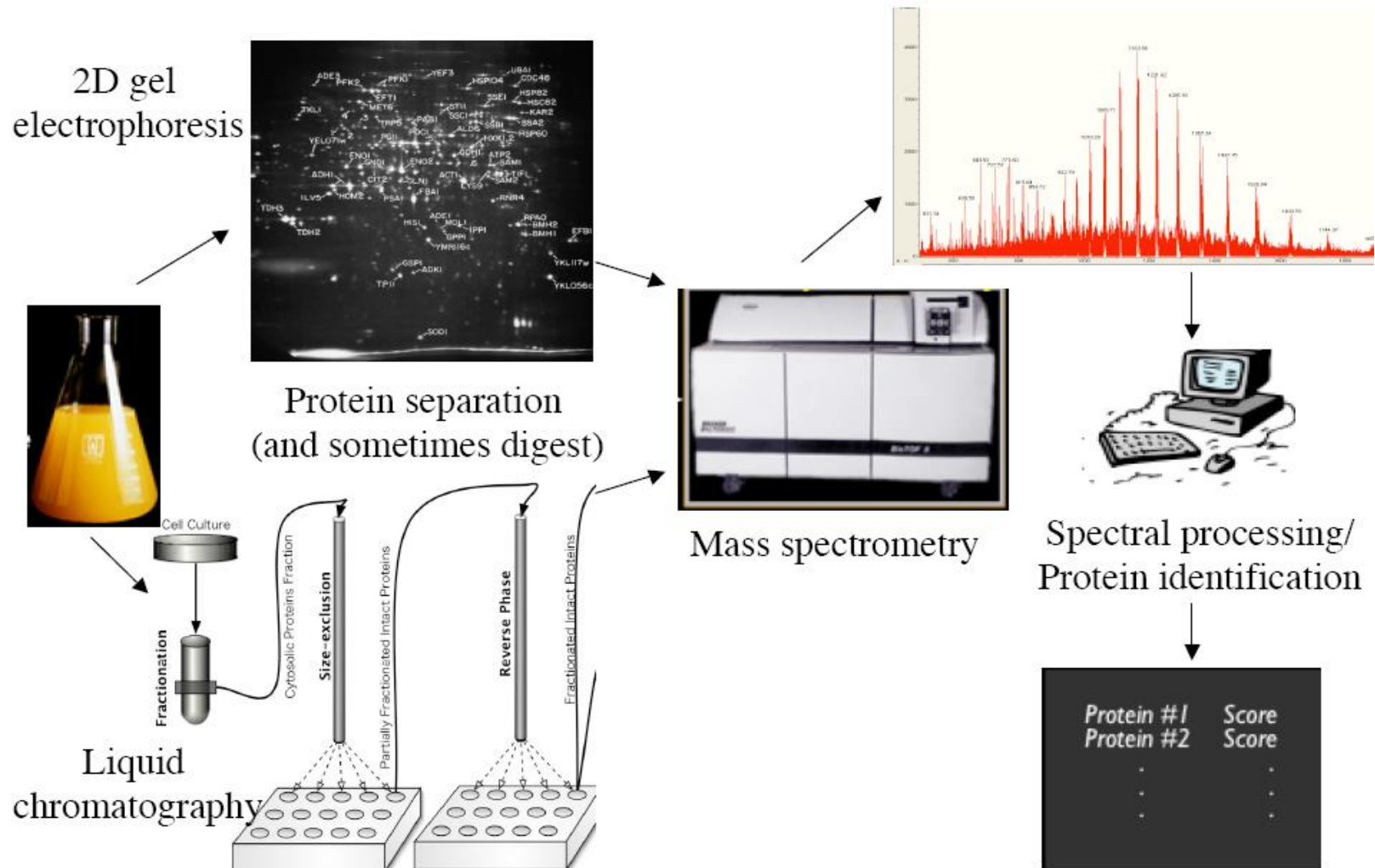
Digesting the Protein

- **Protein digestion(Trypsin)**

N-term

C-term

Digesting the Protein

- **Protein digestion(Trypsin)**

N-term ⬤⬤⬤⬤ C-term

N-term ⬤⬤⬤ C-term

N-term ⬤⬤ C-term

N-term ⬤⬤⬤ C-term

Digesting the Protein

- **Protein digestion(Trypsin)**

N-term ⬤⬤⬤🔴 C-term

N-term ⬤⬤🔴 C-term

N-term ⬤🔴 C-term

N-term ⬤⬤🔴 C-term

Digesting the Protein

- **Protein digestion(Trypsin)**

N-term ⬤⬤⬤🔴 C-term

N-term ⬤⬤🔴 C-term

N-term ⬤🔴 C-term
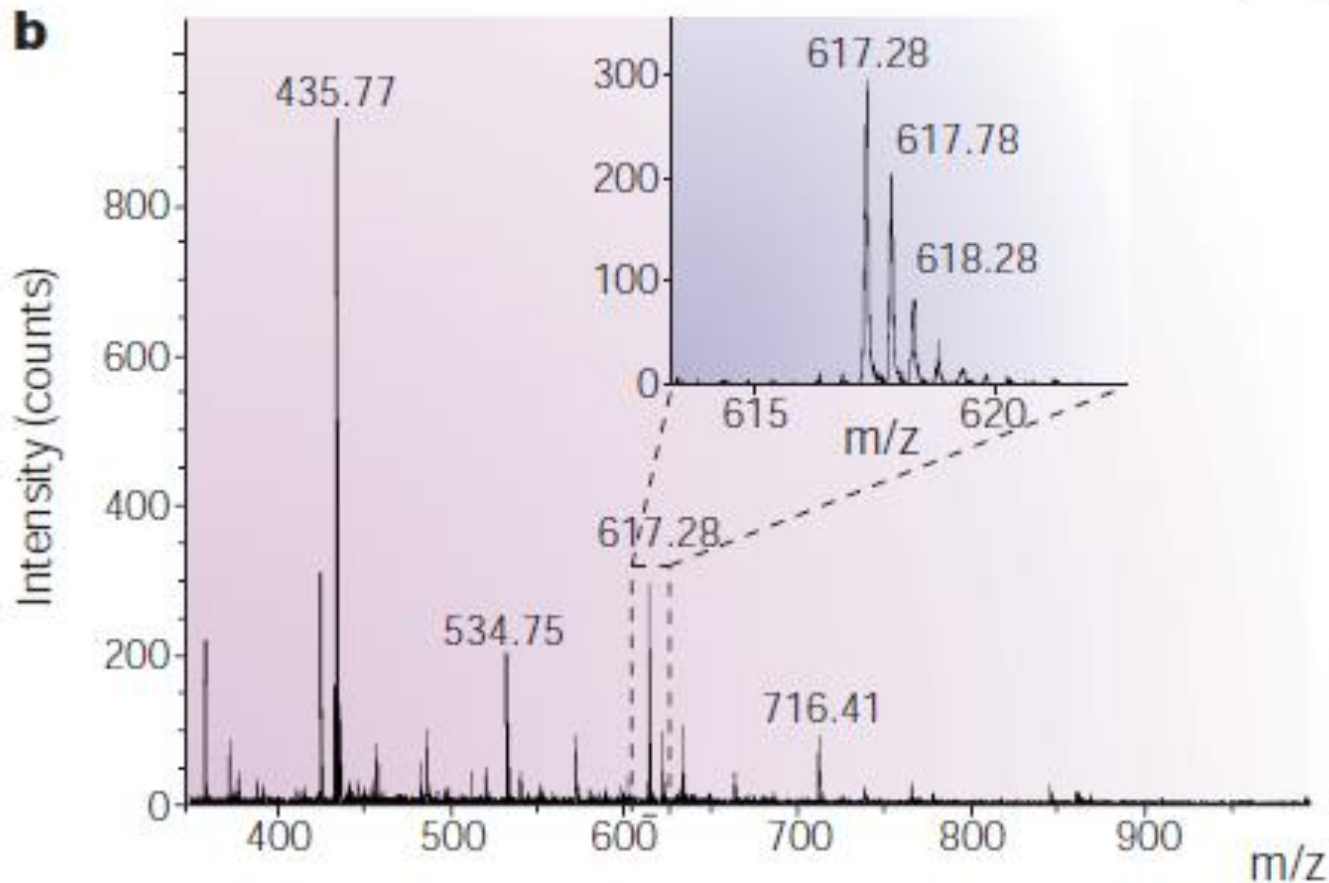
N-term ⬤⬤🔴 C-term

K or R

Proteomics

- **Bottom-up method**

  - Protein digestion

  - Mass spectrometry analysis

  - Peptide identification using MS/MS

 - Protein Assembly using identified peptides

Mass Spectrometry

2D gel electrophoresis

Protein separation (and sometimes digest)

Cell Culture

Fractionation

Cytosolic Proteins Fraction

Size-exclusion

Partially Fractionated Intact Proteins

Reverse Phase

Fractionated Intact Proteins

Liquid chromatography

Mass spectrometry

Spectral processing/ Protein identification

Protein #1   Score
Protein #2   Score

Mass Spectrometry

- **What is the mass spectrum?**

Mass Spectrometry

- **What is the mass spectrum?**

  - Y-axis?
    - Intensity

  - X-axis?
    - m/z (mass to charge ratio)

  - Peak?
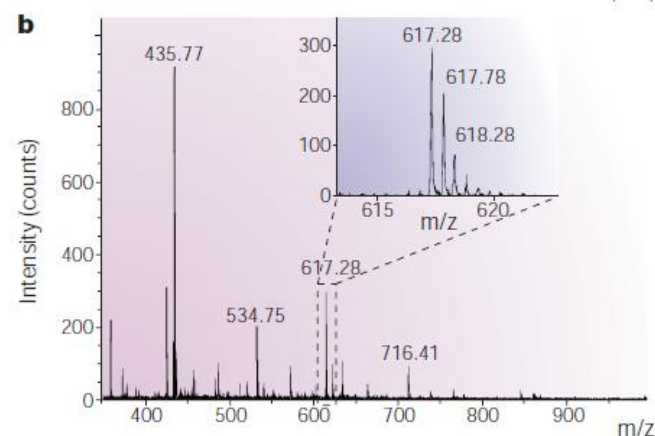
Mass Spectrometry

- **Mass to charge ratio**
  - Peptide with mass of 1232.55
  - The charge state of the peptide ion 2

  - $(M + 2H)^{2+}$
    - M : The mass of the peptide
    - H^+ : The mass of a proton

  - $(1232.55 + (2 * 1.0073)) / 2$
    $= (1232.55 / 2) + ((2 * 1.0073) / 2) = (1232.55 / 2) + 1.0073$
    $= 617.28$

Mass Spectrometry

- **Mass to charge ratio (Quiz)**
  - Peptide with mass of 1848.84
  - The charge state of the peptide ion 3

Mass Spectrometry

- **Mass to charge ratio**
  - Peptide with mass of 1848.84
  - The charge state of the peptide ion 3

  - $(M + 3H)^{3+}$
    - M : The mass of the peptide
    - H^+ : The mass of a proton

  - (1848.84 + (3 * 1.0073)) / 3
    = (1848.84 / 3) + ((3 * 1.0073) / 3) = (1848.84 / 3) + 1.0073
    = 617.28

Mass Spectrometry

- **Mass to charge ratio**
  - Mass = 1232.55, Charge state = 2
    - (1232.55 + (2 * 1.0073)) / 2
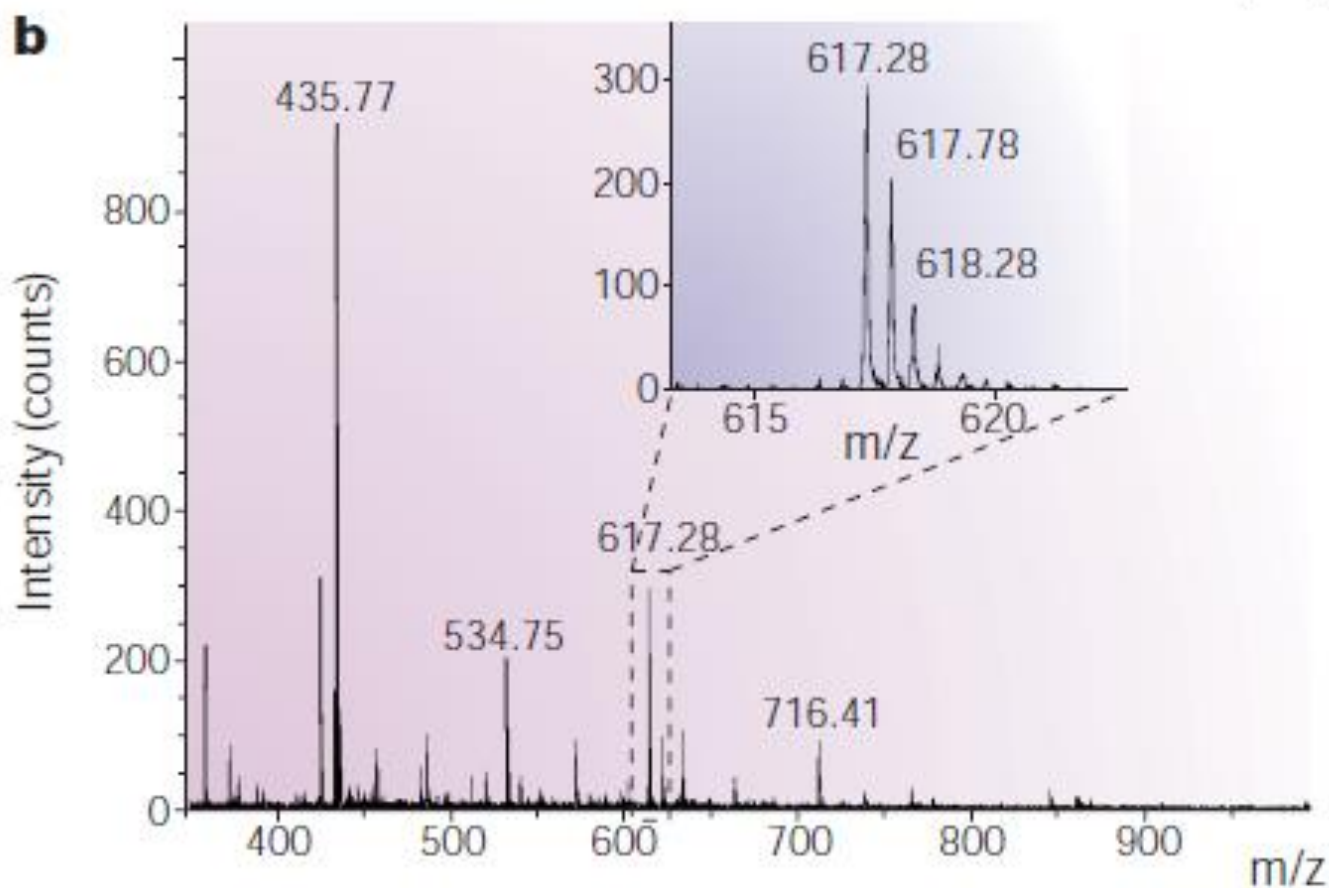   = (1232.55 / 2) + ((2 * 1.0073) / 2) = (1232.55 / 2) + 1.0073
      = 617.28

  - Mass = 1848.84, Charge state = 3
    - (1848.84 + (3 * 1.0073)) / 3
      = (1848.84 / 3) + ((3 * 1.0073) / 3) = (1848.84 / 3) + 1.0073
      = 617.28

Mass Spectrometry

- **Mass to charge ratio**
  - Mass = 1232.55, Charge state = 2
    - (1232.55 + (2 * 1.0073)) / 2
      = (1232.55 / 2) + ((2 * 1.0073) / 2) = (1232.55 / 2) + 1.0073
      = 617.28

  - Mass = 1848.84, Charge state = 3
    - (1848.84 + (3 * 1.0073)) / 3
      = (1848.84 / 3) + ((3 * 1.0073) / 3) = (1848.84 / 3) + 1.0073
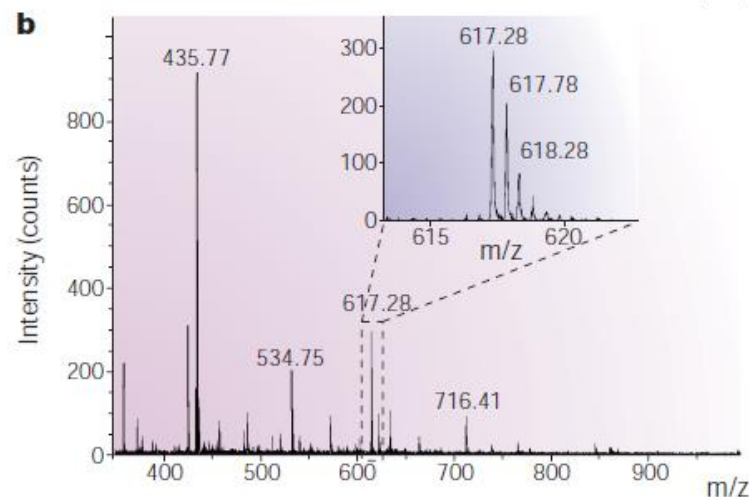   = 617.28

m/z = 617.28 !!!
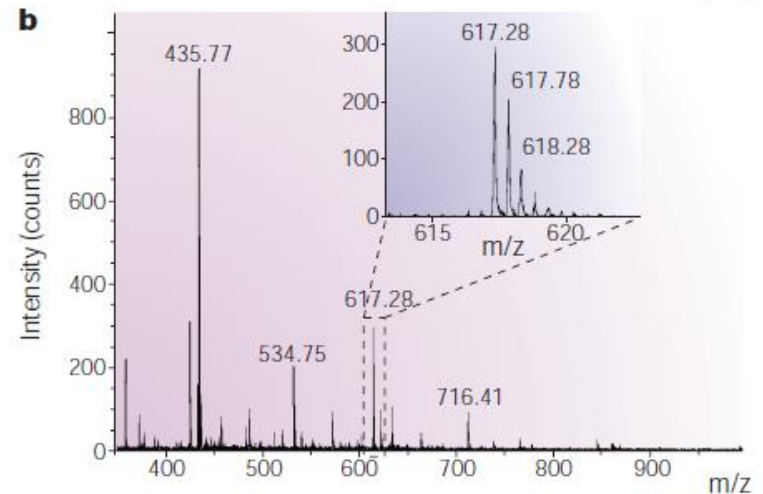
Mass Spectrometry

- **How can we decide the charge state?**

Mass Spectrometry

- **The peptide mass**
  - m/s = 617.28, Charge state = 2

    - (617.28 – 1.0073) * 2 = 1232.55
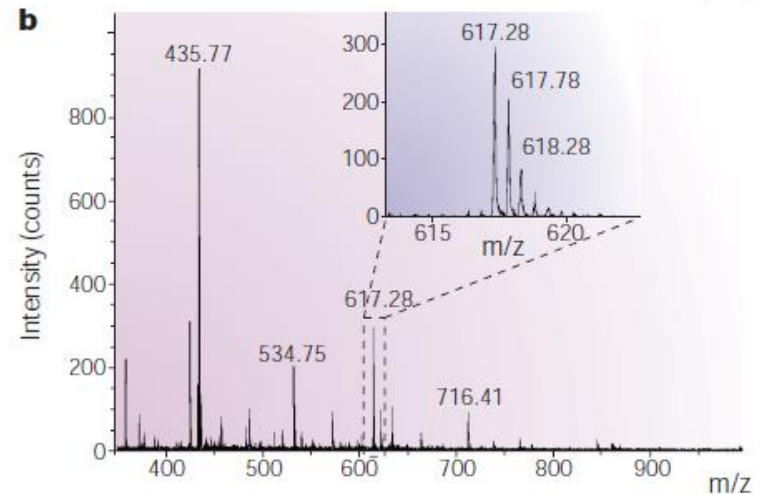
    - The peptide mass = 1232.55
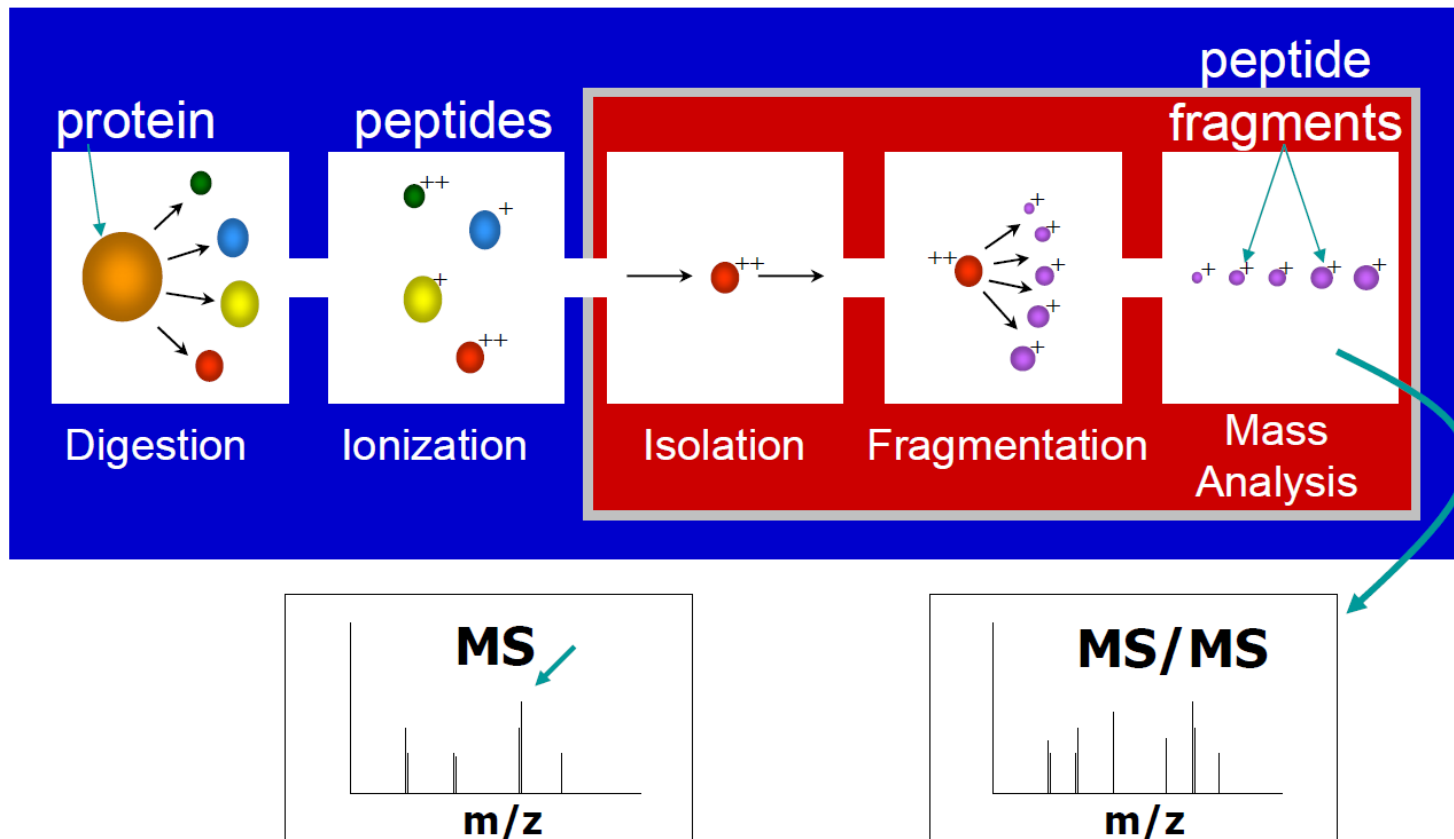
Mass Spectrometry

- **The peptide mass**
  - m/s = 617.28, Charge state = 2

    - (617.28 – 1.0073) * 2 = 1232.55

    - The peptide mass = 1232.55

- **How can we know sequences?**

Mass Spectrometry

- **The peptide mass**
  - m/s = 617.28, Charge state = 2

    - (617.28 – 1.0073) * 2 = 1232.55

    - The peptide mass = 1232.55

- **How can we know sequences?**

  Tandem mass spectrum (MS/MS)

Tandem Mass Spectrum
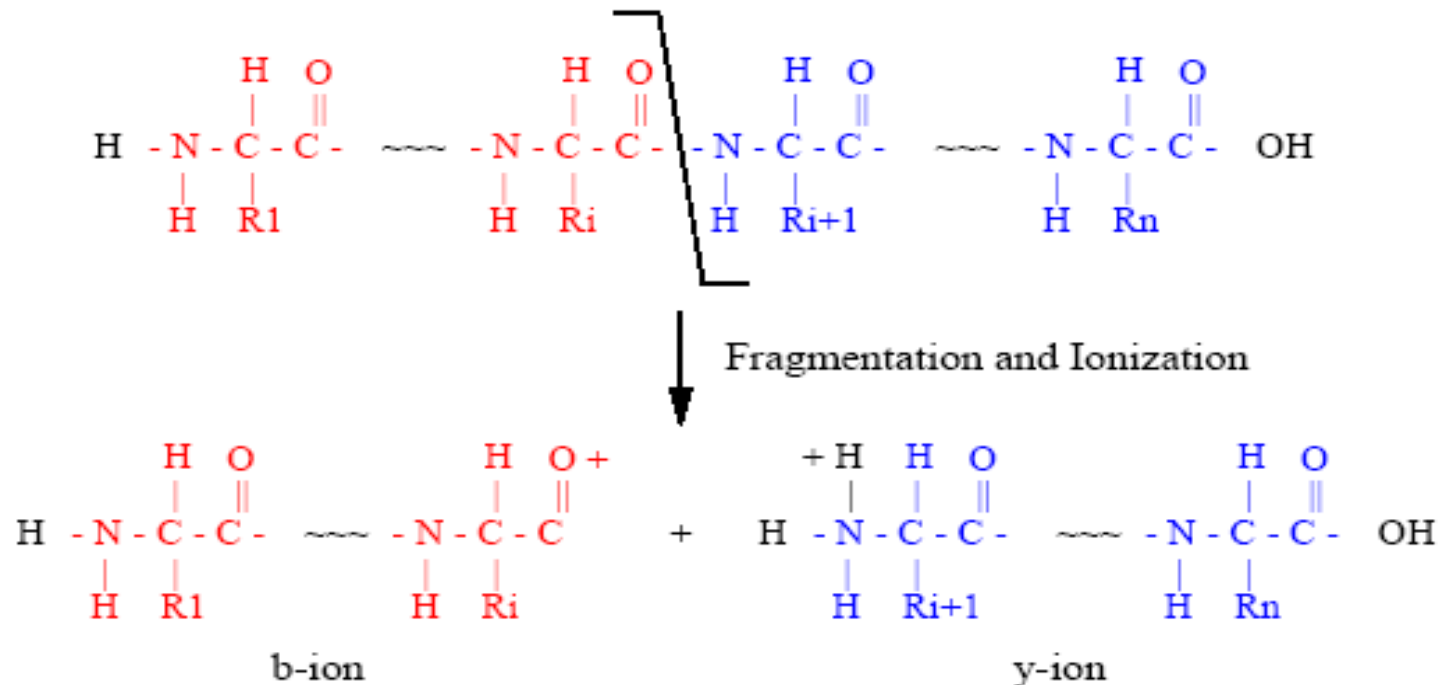
- **Tandem mass spectrum**
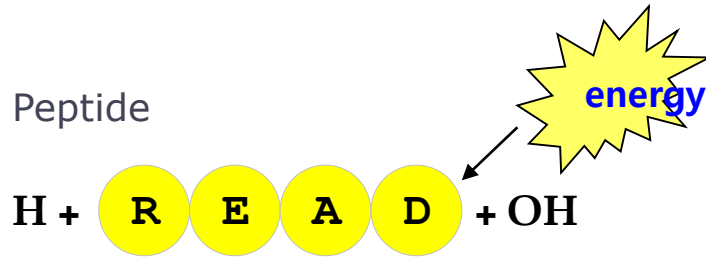
Tandem Mass Spectrum

- **Fragmentation**
  - Method: HCD, CID, ETD, ECD

  - LPKNK
    - LPKNK
    - L + PKNK
    - LP + KNK
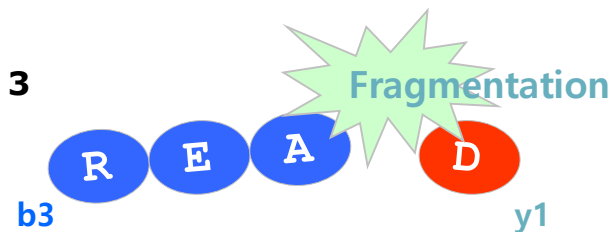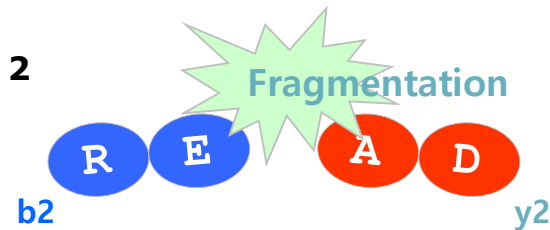   - LPK + NK
    - LPKN + K
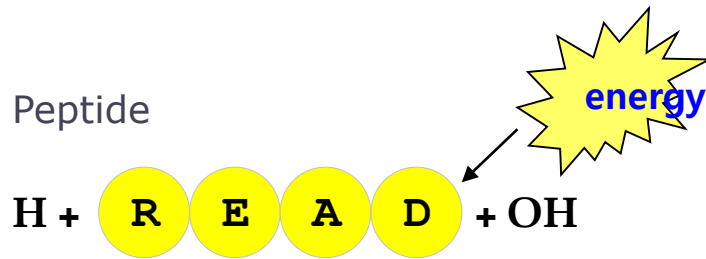
Tandem Mass Spectrum

- **Fragmentation**

Fragmentation and Ionization

b-ion   y-ion

Tandem Mass Spectrum

Peptide

energy

H + R E A D + OH

Amino Acid	Mass
A	71
D	115
E	129
R	156

1

Fragmentation

R   E   A   D

b1   y3

2

Fragmentation

R E   A   D

b2   y2

3

Fragmentation

R   E   A   D

b3   y1

- **b1 mass = R + 1 = 157**
- **y1 mass = D + 19 = 134**

Tandem Mass Spectrum

Peptide

energy

H + R E A D + OH

Amino Acid	Mass
A	71
D	115
E	129
R	156

1

Fragmentation

R   E   A   D

b1   y3

2

Fragmentation

R   E   A   D

b2   y2

3

Fragmentation
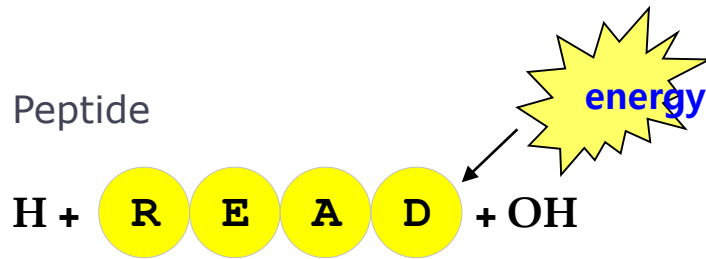
R   E   A   D

b3   y1
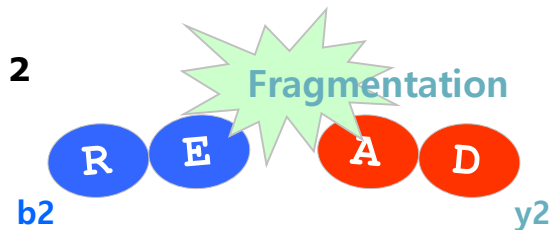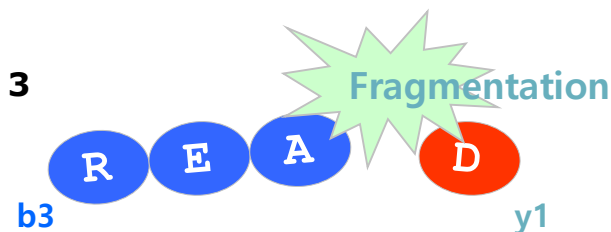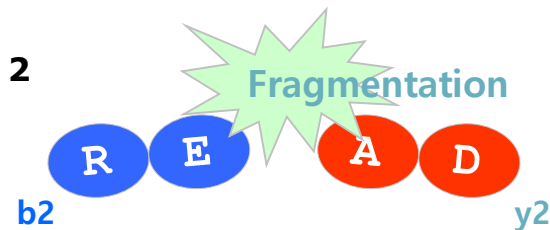
- **b1 mass = R + 1 = 157**
- **y1 mass = D + 19 = 134**

- **Quiz**
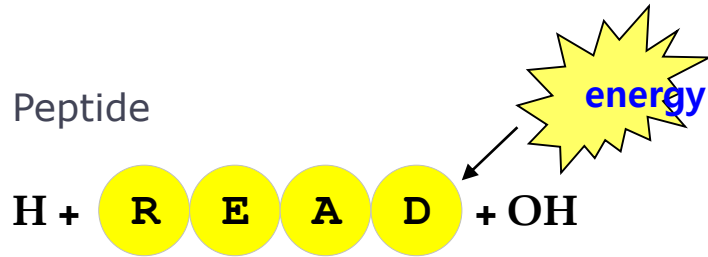- **b2 mass = ?**
- **y3 mass = ?**

Tandem Mass Spectrum

Peptide

H + R E A D + OH

energy

Amino Acid	Mass
A	71
D	115
E	129
R	156

1

Fragmentation

R E A D

b1   y3

2

Fragmentation

R E A D

b2   y2

3

Fragmentation

R E A D

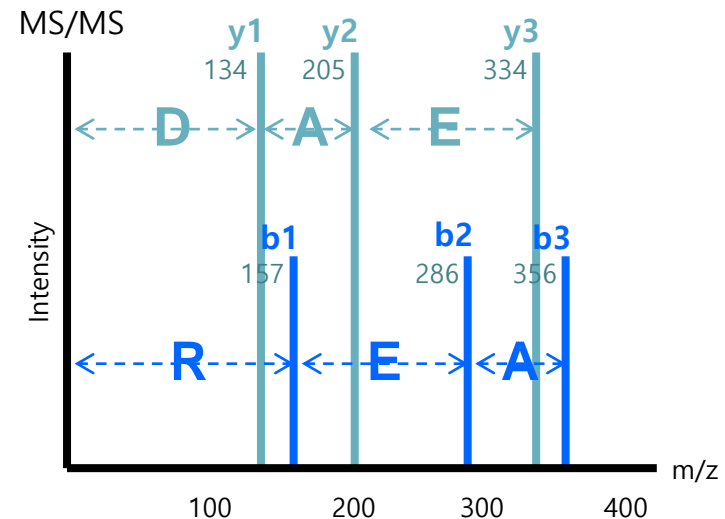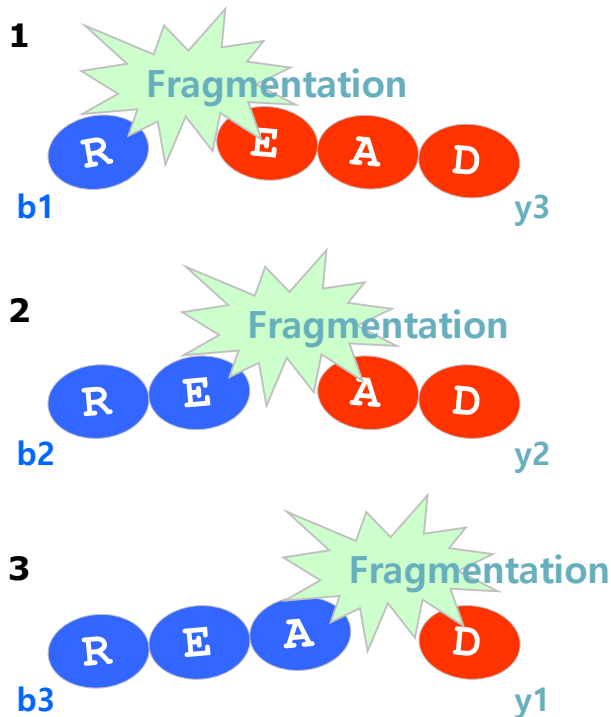b3   y1

- **b1 mass = R + 1 = 157**
- **y1 mass = D + 19 = 134**

- **Quiz**
- **b2 mass = R + E + 1 = 286**
- **y3 mass = E + A + D + 19 = 334**

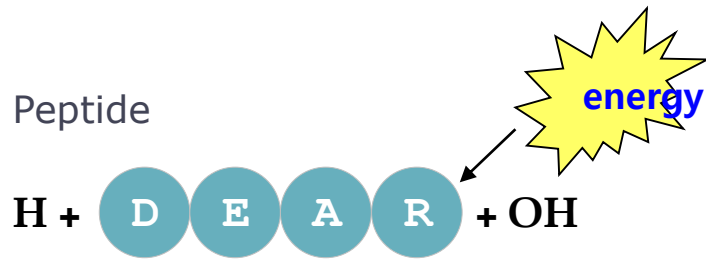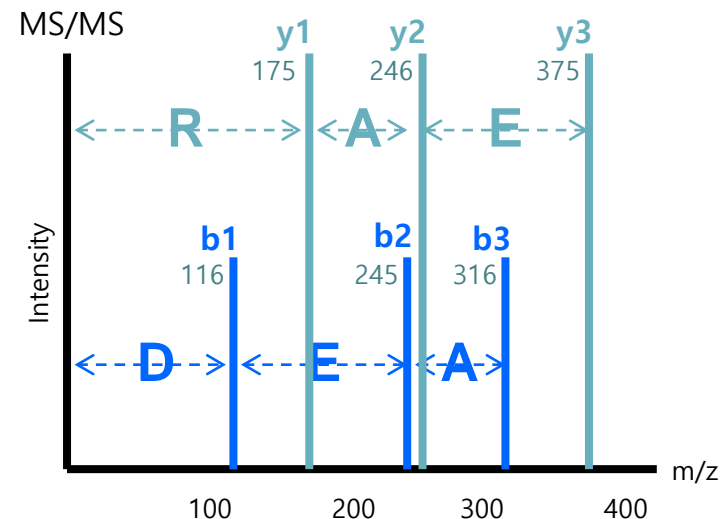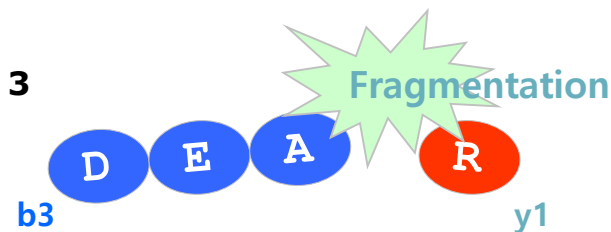# Tandem Mass Spectrum

Peptide

$H +$ R E A D $+ OH$

energy

Amino Acid	Mass
A	71
D	115
E	129
R	156

1

Fragmentation

R   E A D

b1   y3

2

Fragmentation

R E   A D

b2   y2

3

Fragmentation

R E A   D

b3   y1

MS/MS

y1 134   y2 205   y3 334

D → ← A → ← E →

b1 157   b2 286   b3 356

R → ← E → ← A →

m/z

100   200   300   400

Intensity

Tandem Mass Spectrum (Quiz)

Peptide

H + D E A R + OH

energy

Amino Acid	Mass
A	71
D	115
E	129
R	156

1

Fragmentation

D E A R
b1   y3

2

Fragmentation

D E A R
b2   y2

3

Fragmentation

D E A R
b3   y1

MS/MS

y1 175   y2 246   y3 375

R A E

b1 116   b2 245   b3 316

D E A

Intensity

m/z

100   200   300   400
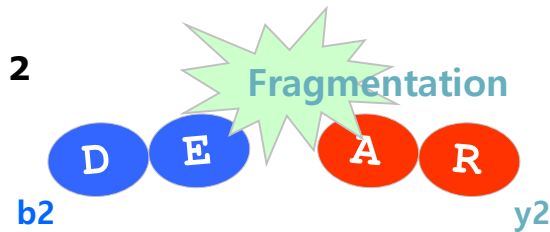
Tandem Mass Spectrum

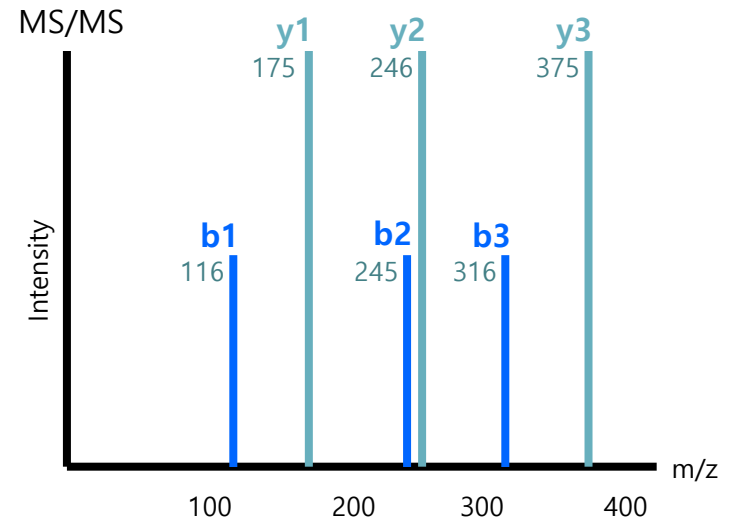- **READ vs DEAR**

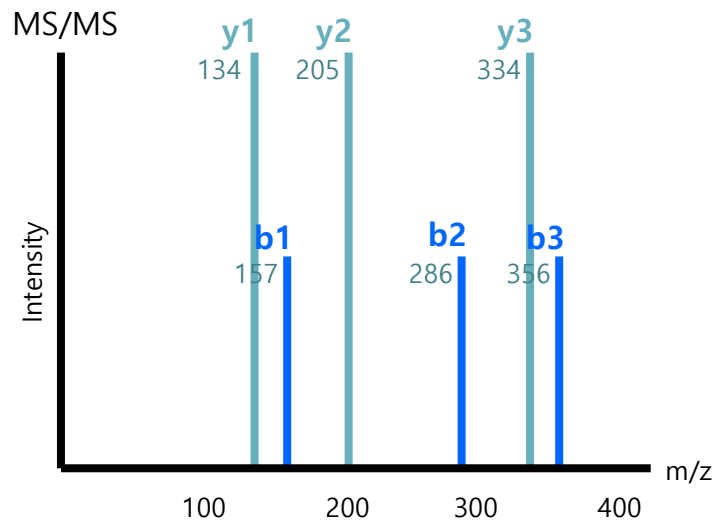 - Peptide mass = 471
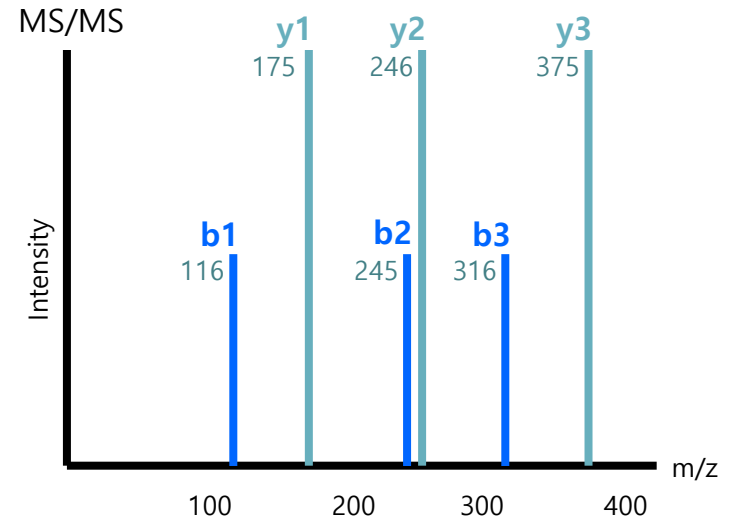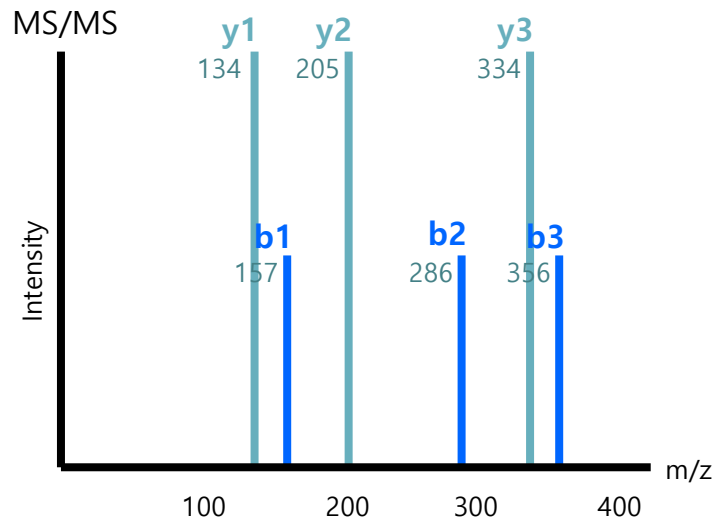
Proteomics

- **Bottom-up method**

  - Protein digestion

  - Mass spectrometry analysis

  - Peptide identification using MS/MS
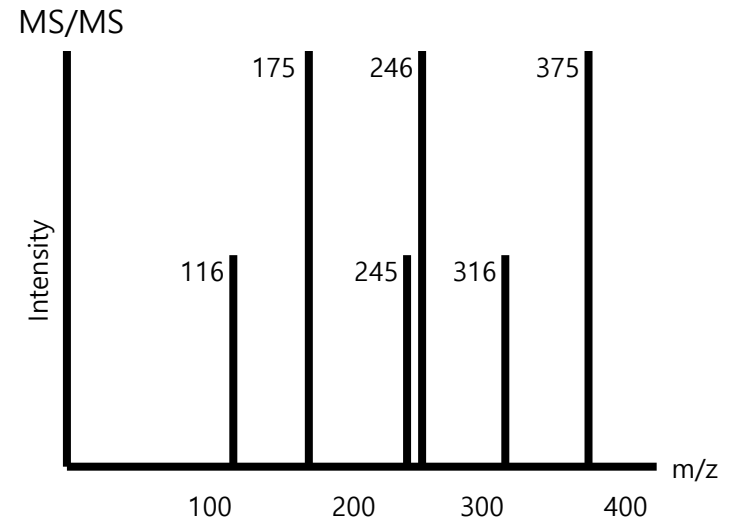
  - Protein Assembly using identified peptides
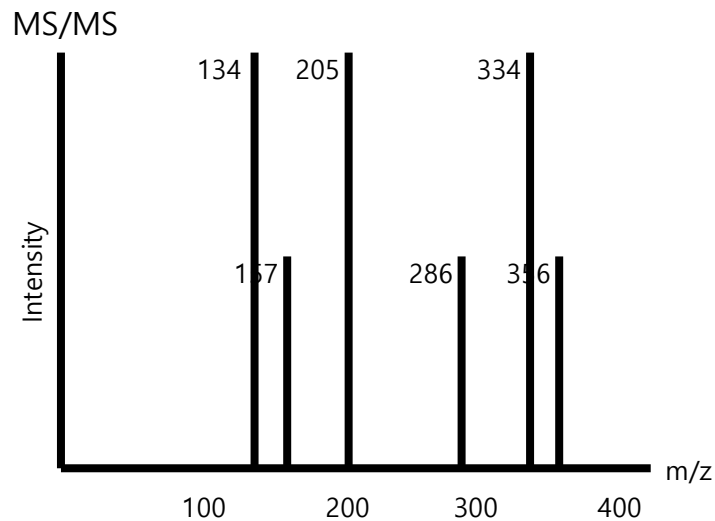
- **READ vs DEAR**

  - Peptide mass = 471

Peptide Identification using MS/MS

- Peptide mass = 471

Peptide Identification using MS/MS

- **De novo sequencing approach**

  - Peptide mass = 471

Amino Acid	Mass
A	71
D	115
E	129
R	156

MS/MS spectrum: peaks at 116, 175, 245, 246, 316, 375 (Intensity vs m/z, axis marked 100, 200, 300, 400)

Peptide Identification using MS/MS

- **De novo sequencing approach**

  - Peptide mass = 471

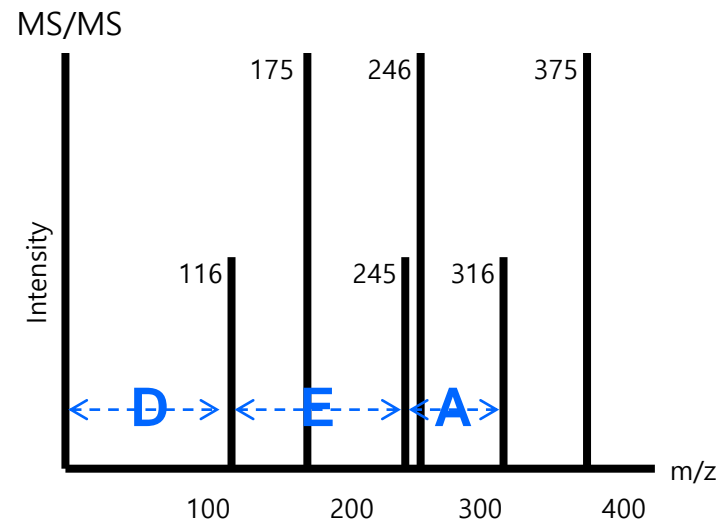| Amino Acid	Mass
A	71
D	115
E	129
R	156

Peptide Identification using MS/MS

- **De novo sequencing approach**

  - Peptide mass = 471

Amino Acid	Mass
A	71
D	115
E	129
R	156

MS/MS

- **De novo sequencing approach**

  - Peptide mass = 471

Amino Acid	Mass
A	71
D	115
E	129
R	156

정답: DEAR

Peptide Identification using MS/MS

- **Database search approach**

  - Peptide mass = 471

Amino Acid	Mass
A	71
D	115
E	129
R	156
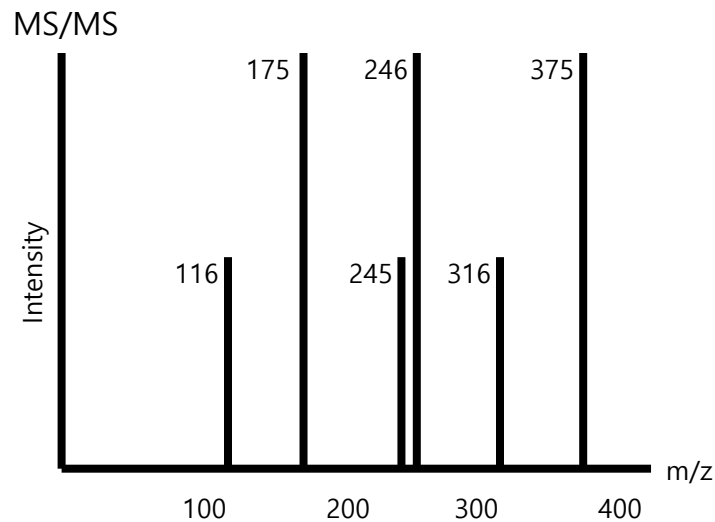
Peptide Identification using MS/MS

- **Database search approach**

  - Peptide mass = 471

Amino Acid	Mass
A	71
D	115
E	129
R	156
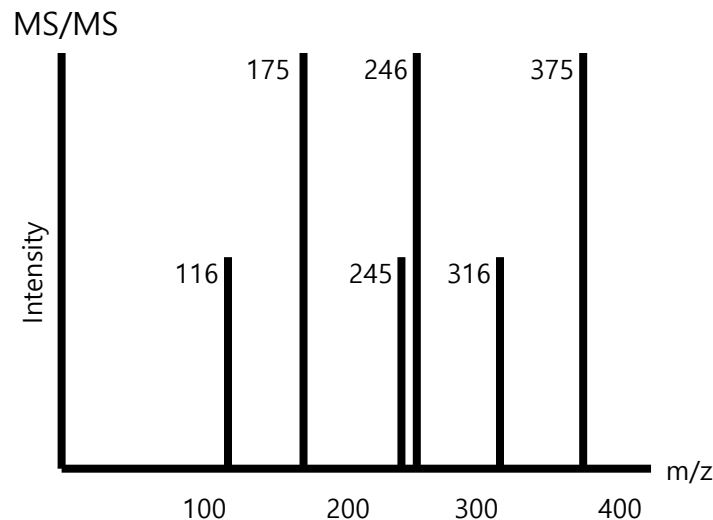
MS/MS

Database
ADER
DEAR
READ
REDA

Peptide Identification using MS/MS

- **Database search approach**

  - Peptide mass = 471

Amino Acid	Mass
A	71
D	115
E	129
R	156

MS/MS

Database

ADER   1.7

DEAR

READ

REDA

Peptide Identification using MS/MS

- **Database search approach**

  - Peptide mass = 471

Amino Acid	Mass
A	71
D	115
E	129
R	156

MS/MS spectrum with peaks at m/z values: 116, 175, 245, 246, 316, 375. Y-axis: Intensity. X-axis: m/z (100, 200, 300, 400).

Database
ADER
DEAR   9.8
READ
REDA

Peptide Identification using MS/MS

- **Database search approach**

  - Peptide mass = 471

Amino Acid	Mass
A	71
D	115
E	129
R	156

MS/MS

Peaks: 116, 175, 245, 246, 316, 375

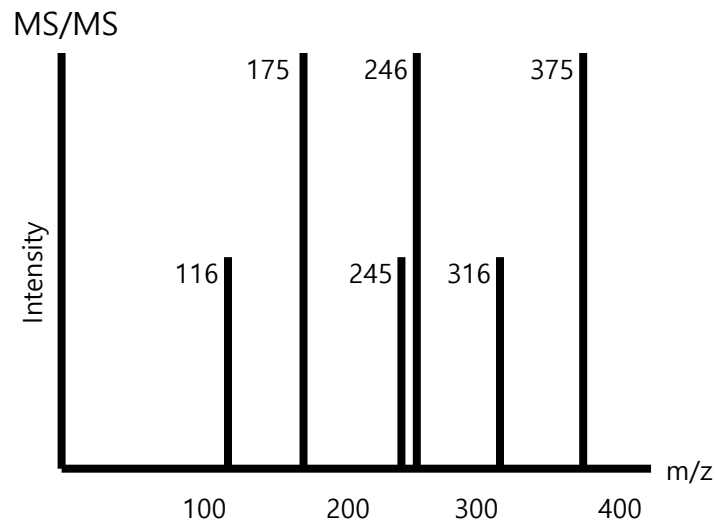Intensity (y-axis), m/z (x-axis): 100, 200, 300, 400

Database
ADER
DEAR   9.8
READ
REDA

Peptide Identification using MS/MS

- **Database search approach**

  - Peptide mass = 471

Amino Acid	Mass
A	71
D	115
E	129
R	156

MS/MS
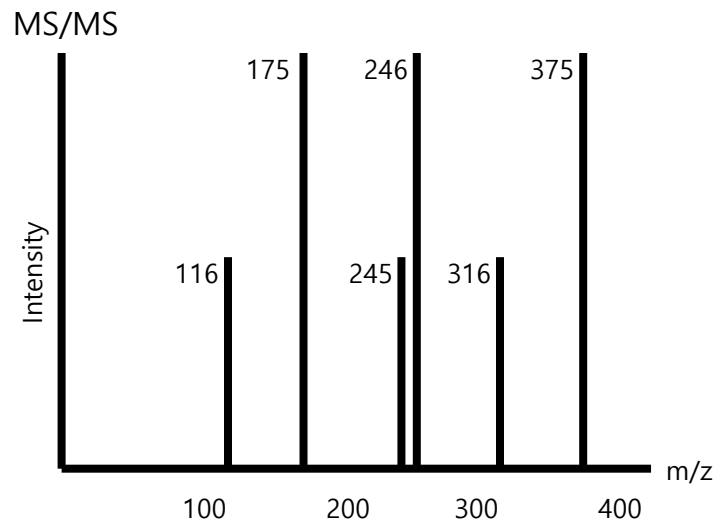

Database
ADER
DEAR
READ   5.7
REDA

Peptide Identification using MS/MS

- **Database search approach**

  - Peptide mass = 471

Amino Acid	Mass
A	71
D	115
E	129
R	156

MS/MS spectrum with peaks at m/z: 116, 175, 245, 246, 316, 375
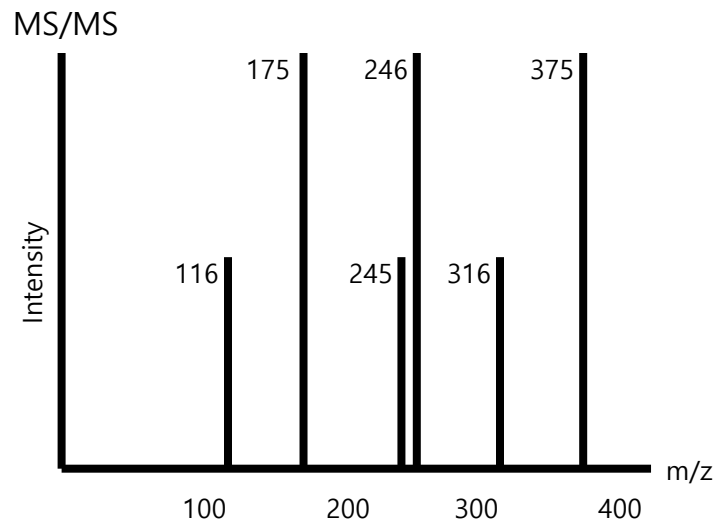
Database
ADER
DEAR
READ
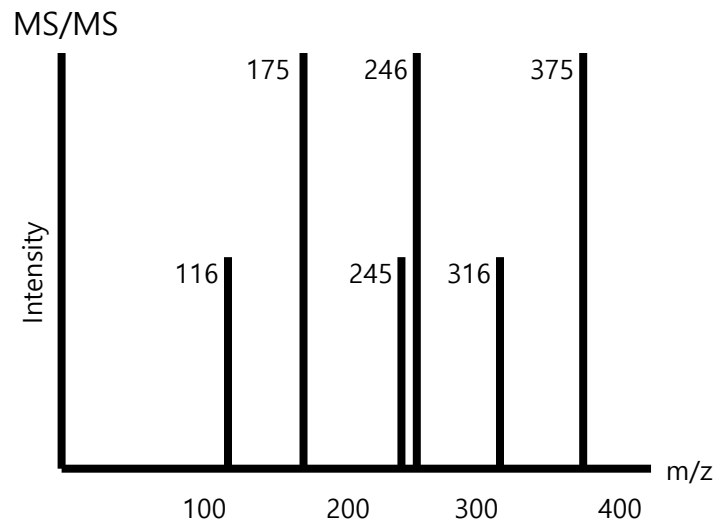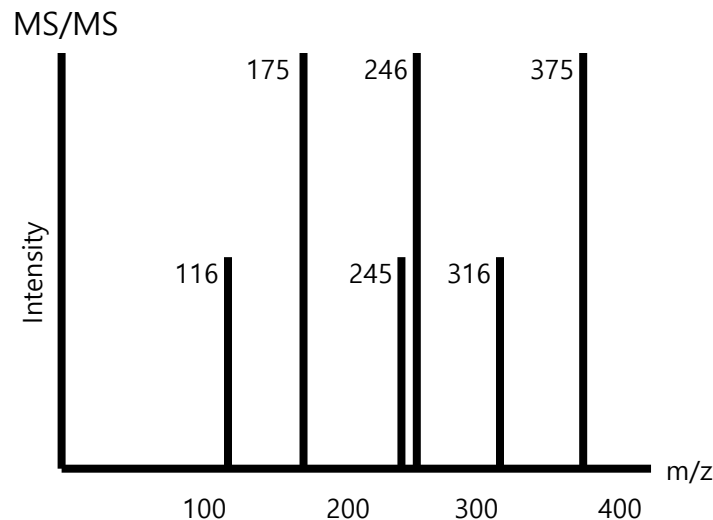REDA   3.3

Peptide Identification using MS/MS

- **Database search approach**

  - Peptide mass = 471

Amino Acid	Mass
A	71
D	115
E	129
R	156

MS/MS spectrum with peaks at m/z: 116, 175, 245, 246, 316, 375

Database

ADER	1.7
DEAR	**9.5**
READ	5.7
REDA	3.3

Peptide Identification using MS/MS

De Novo Approach

VS

Database Search Approach

Proteomics

- **Bottom-up method**

  - Protein digestion

  - Mass spectrometry analysis

  - Peptide identification using MS/MS

  - Protein Assembly using identified peptides

Protein Inference

- **Protein Inference**

  - Protein 1
    - MEMEKEFEQIDKSGSWAGGAEFLGRYAIYQDIRHEASDFPCRVAKL PKNKNRNRYRDVSPFDHSRKLHQEDNDYINASLIKMEEAQRSYILT Q

  - Protein 2
    - MTVYAVTGGAEFLGRYIVKLLISADDVQEIRVINVVEDPQPLVSKVK VINYIQCDINDLIR

  - Peptide
    - LGRYIVKLLISADD
    - DPQPLVSKVK

Protein Inference

- **Protein Inference**

  - Protein 1
    - MEMEKEFEQIDKSGSWAGGAEFLGRYAIYQDIRHEASDFPCRVAKL PKNKNRNRYRDVSPFDHSRKLHQEDNDYINASLIKMEEAQRSYILT Q

  - Protein 2
    - MTVYAVTGGAEFLGRYIVKLLLISADDVQEIRVINVVEDPQPLVSKVK VINYIQCDINDLIR

  - Peptide
    - GGAEFLGRY