

질량분석 기반 단백질체 연구 소개

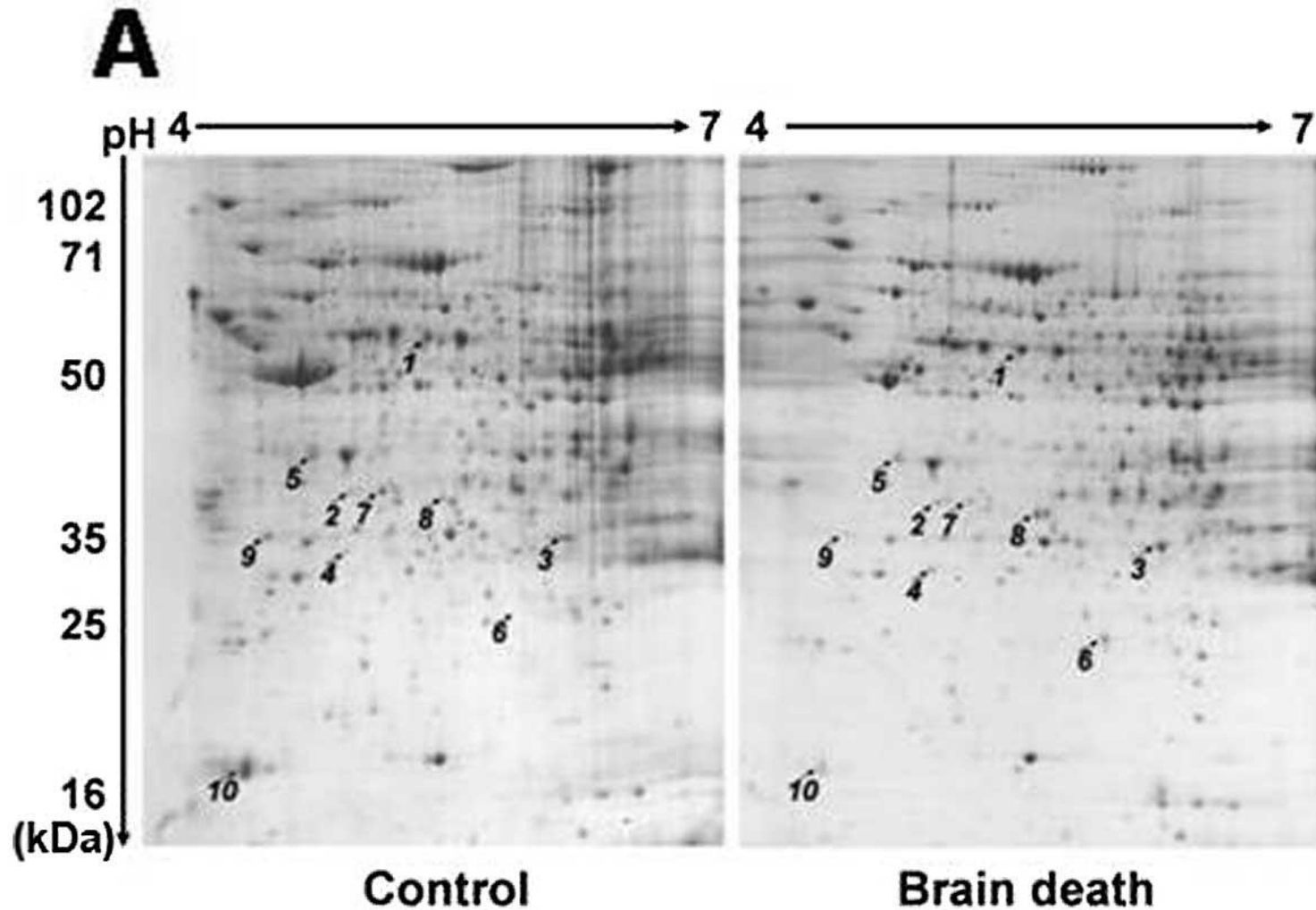
2020. 06

김현우

What is Proteomics?

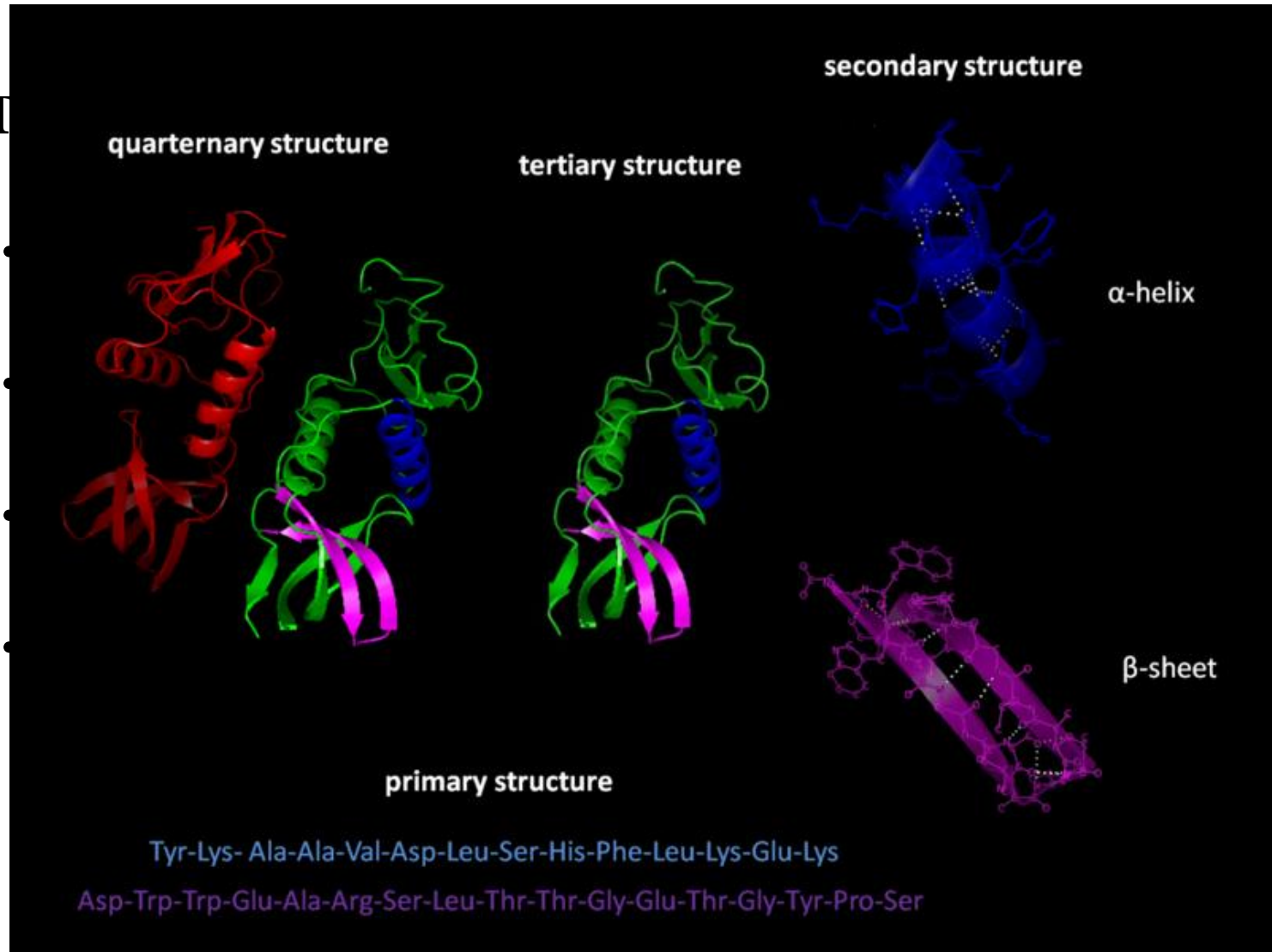
- **The Study of Proteins**
 - Expression
 - Structures
 - Functions
 - Interactions of protein

What is Proteomics?



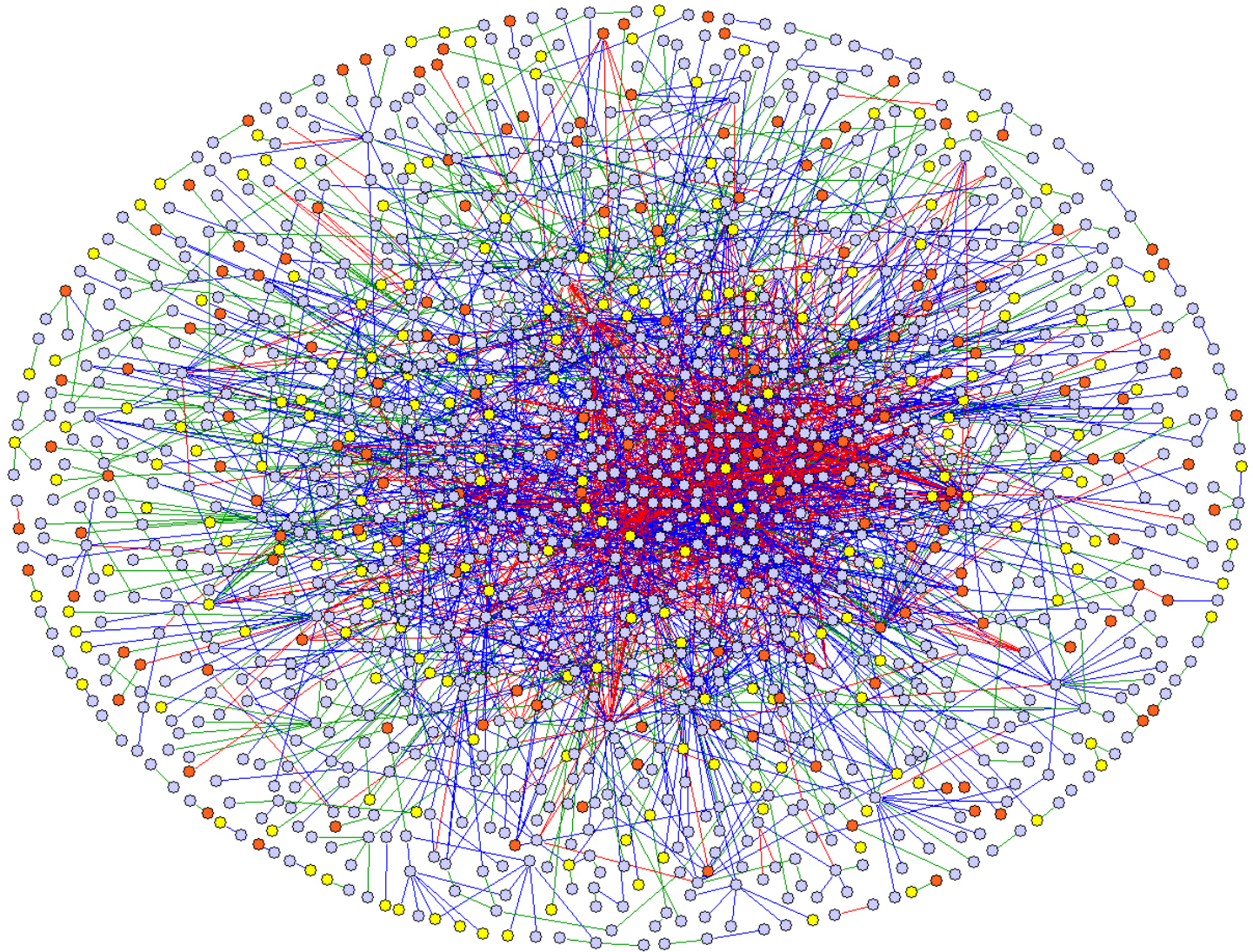
What is Proteomics?

- T



What is Proteomics?

- **T**



What is a Protein?

- **Amino acid**

- 생물의 몸을 구성하는 단백질의 기본 구성단위이다

- **Peptide**

- 20가지의 서로 다른 아미노산들이 화학 결합으로 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

- MEMEKEFEQIDKSGSWAAIYQDIRHEASDFPCRVAKL**LPKNKNRNR**
YRDVSPFDHSRKLHQEDNDYINASLIKMEEAQRSYILTQ....

What is a Protein?

- **Protein**

- An amino acid sequence

- MEMEKEFEQIDKSGSWAAIYQDIRHEASDFPCRVAKL**LPKNKNRNR**
YRDVSPFDHSRKLHQEDNDYINASLIKMEEAQRSYILTQ....

- **Peptide**

- Substring of the protein

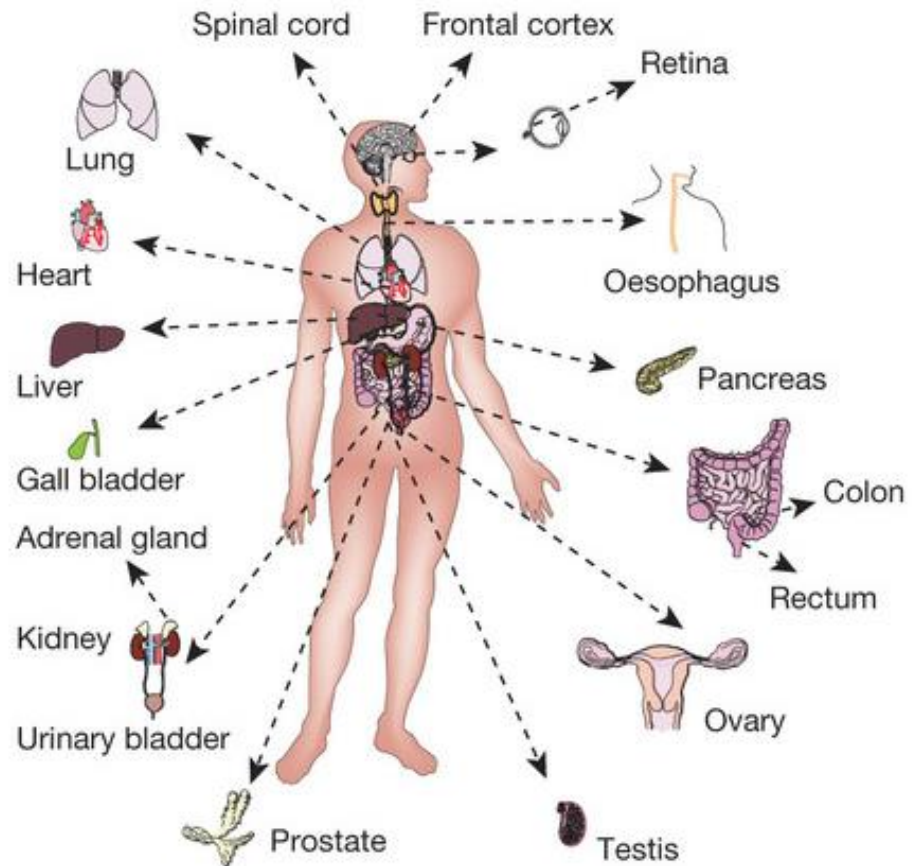
- **LPKNKNRNR****YRDVSPFDHSR**

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?**



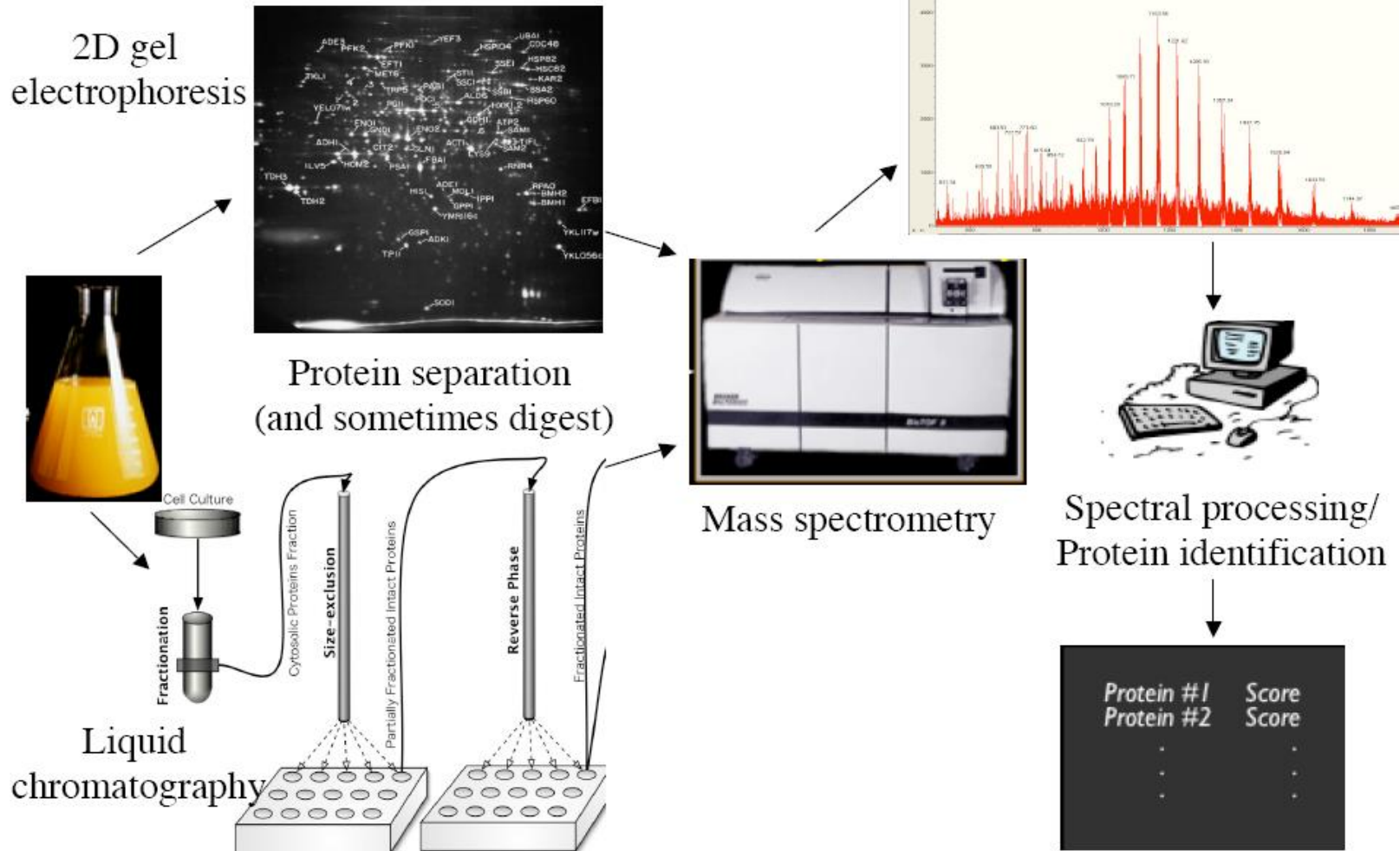
Proteomics

- What is Mass Spectrometry?

	Amino Acids 1						Amino Acids 2					
	AA Codes	AA Residue Composition	Mono.	Avg.	Structure		AA Codes	AA Residue Composition	Mono.	Avg.	Structure	
Glycine	Gly G	C ₂ H ₃ NO	57.021464	57.05	-NH-CH ₂ -CO-		Asp D	C ₄ H ₅ NO ₃	115.02694	115.1		Aspartic acid
Alanine	Ala A	C ₃ H ₅ NO	71.037114	71.08			Gln Q	C ₅ H ₈ N ₂ O ₂	128.05858	128.1		Glutamine
Serine	Ser S	C ₃ H ₅ NO ₂	87.032029	87.08			Lys K	C ₆ H ₁₂ N ₂ O	128.09496	128.2		Lysine
Proline	Pro P	C ₅ H ₇ NO	97.052764	97.12		I O N S O U R C E . C O M	Glu E	C ₅ H ₇ NO ₃	129.04259	129.1		Glutamic acid
Valine	Val V	C ₅ H ₉ NO	99.068414	99.07			Met M	C ₅ H ₉ NOS	131.04048	131.2		Methionine
Threonine	Thr T	C ₄ H ₇ NO ₂	101.04768	101.1			His H	C ₆ H ₇ N ₃ O	137.05891	137.1		Histidine
Cysteine	Cys C	C ₃ H ₅ NOS	103.00919	103.1			Phe F	C ₉ H ₉ NO	147.06841	147.2		Phenylalanine
Leucine	Leu L	C ₆ H ₁₁ NO	113.08406	113.2			Arg R	C ₆ H ₁₂ N ₄ O	156.10111	156.2		Arginine
Isoleucine	Ile I	C ₆ H ₁₁ NO	113.08406	113.2			Tyr Y	C ₉ H ₉ NO ₂	163.06333	163.2		Tyrosine
Asparagine	Asn N	C ₄ H ₆ N ₂ O ₂	114.04293	114.1			Trp W	C ₁₁ H ₁₀ N ₂ O	186.07931	186.2		Tryptophane

Proteomics

- 2D gel electrophoresis



Proteomics

-

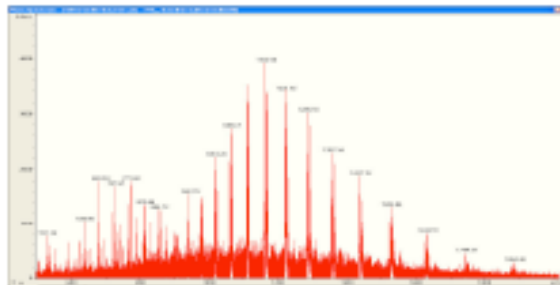
Ion Source

Electrospray Ionization (ESI)
Matrix-assisted laser desorption/ionization (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

- MEMEKEFEQIDKSGSWAAIYQDIRHEASDFPCRVAKL**LPKNKNRNR**Y**RDVSPFDHSR**KLHQEDNDYINASLIKMEEAQRSYILTQ

- Peptide

- **LPKNKNRNRNR**Y**RDVSPFDHSR**

- Amino acid

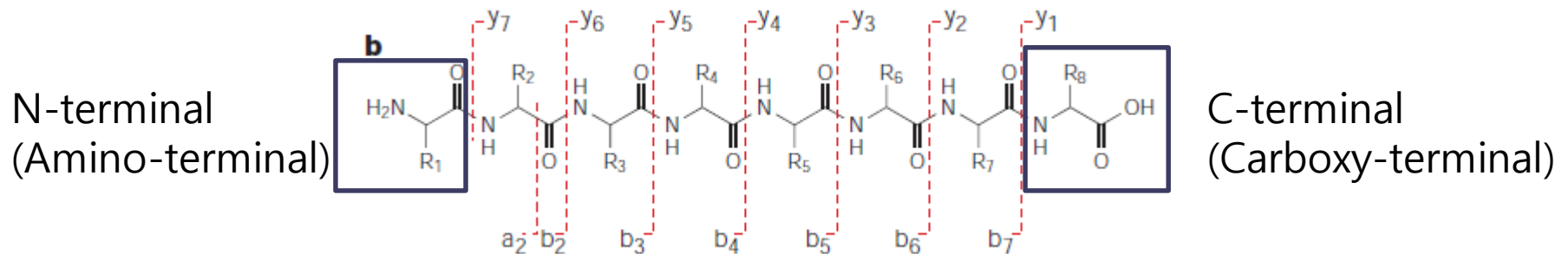
- R

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.



Digesting the Protein

- **Protein digestion(Trypsin)**

- Protein

- MEMEKEFEQIDKSGSWAAIYQDIRHEASDFPCRVAKLPKNKNRNR
RDVSPFDHSRKLHQEDNDYINASLIKMEEAQRSYILTQ

- Peptides

- MEMEKEFEQIDK
 - SGSWAAIYQDIR
 - HEASDFPCRVAK
 - LPK
 - K

Digesting the Protein

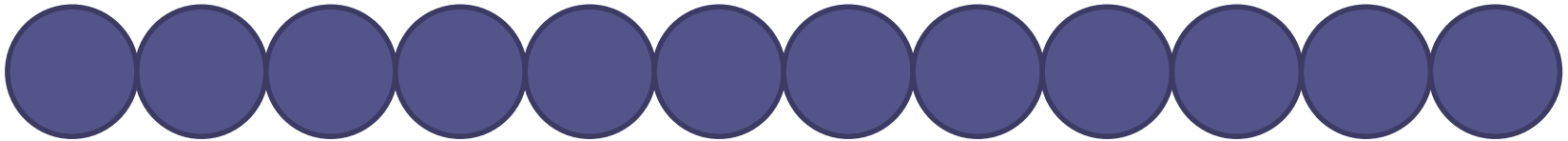
- **Protein digestion(Trypsin)**
 - Protein (Quiz)
 - MTVYAVTGGAEFLGRYIVKLLISADDVQEIRVINVVEDPQPLVSKVK
VINYIQCDINDLIR
 - Peptides ?

Digesting the Protein

- **Protein digestion(Trypsin)**
 - Protein (Quiz)
 - MTVYAVTGGAEFLGRYIVKLLISADDDVQEIRVINVVEDPQPLVSKVK
VINYIQCDINDLIR
 - Peptides ?
 - MTVYAVTGGAEFLGR
 - YIVKLLISADDDVQEIR
 - VINVVEDPQPLVSK
 - VK
 - VINYIQCDINDLIR
- 실습 (?)

Digesting the Protein

- **Protein digestion(Trypsin)**

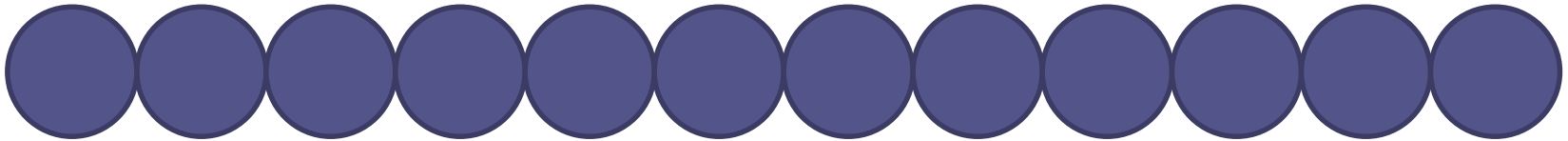


Digesting the Protein

- **Protein digestion(Trypsin)**

N-term

C-term

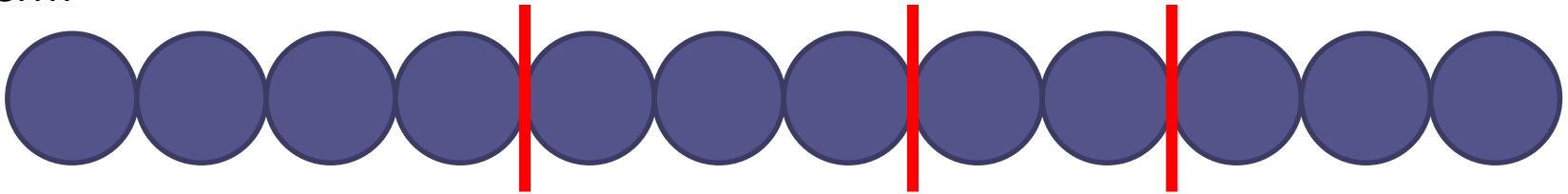


Digesting the Protein

- **Protein digestion(Trypsin)**

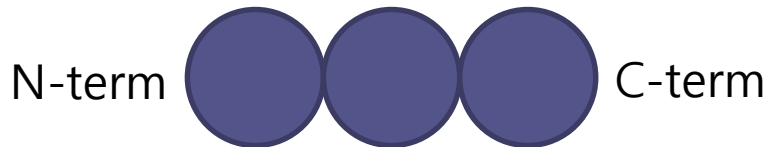
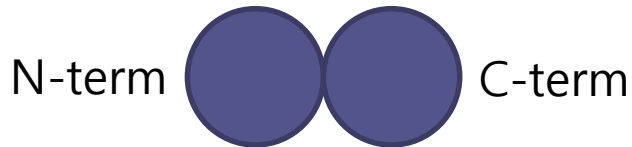
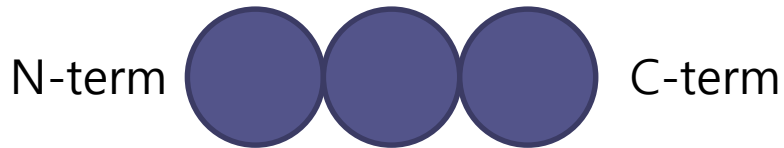
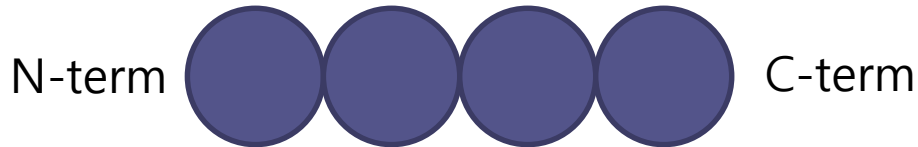
N-term

C-term



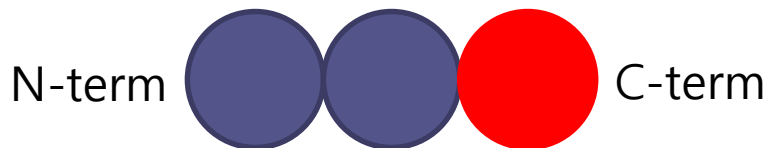
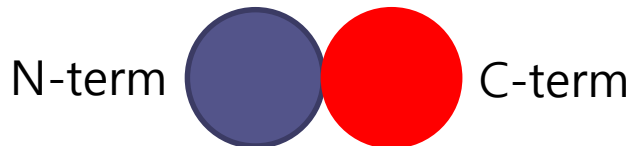
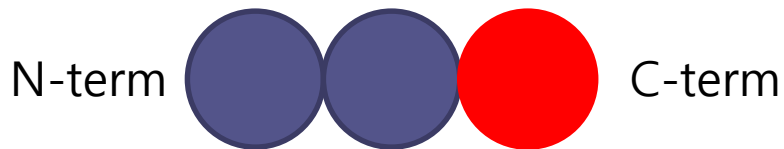
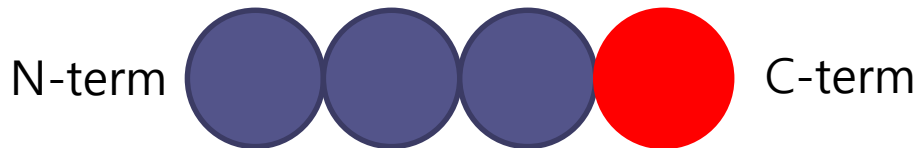
Digesting the Protein

- **Protein digestion(Trypsin)**



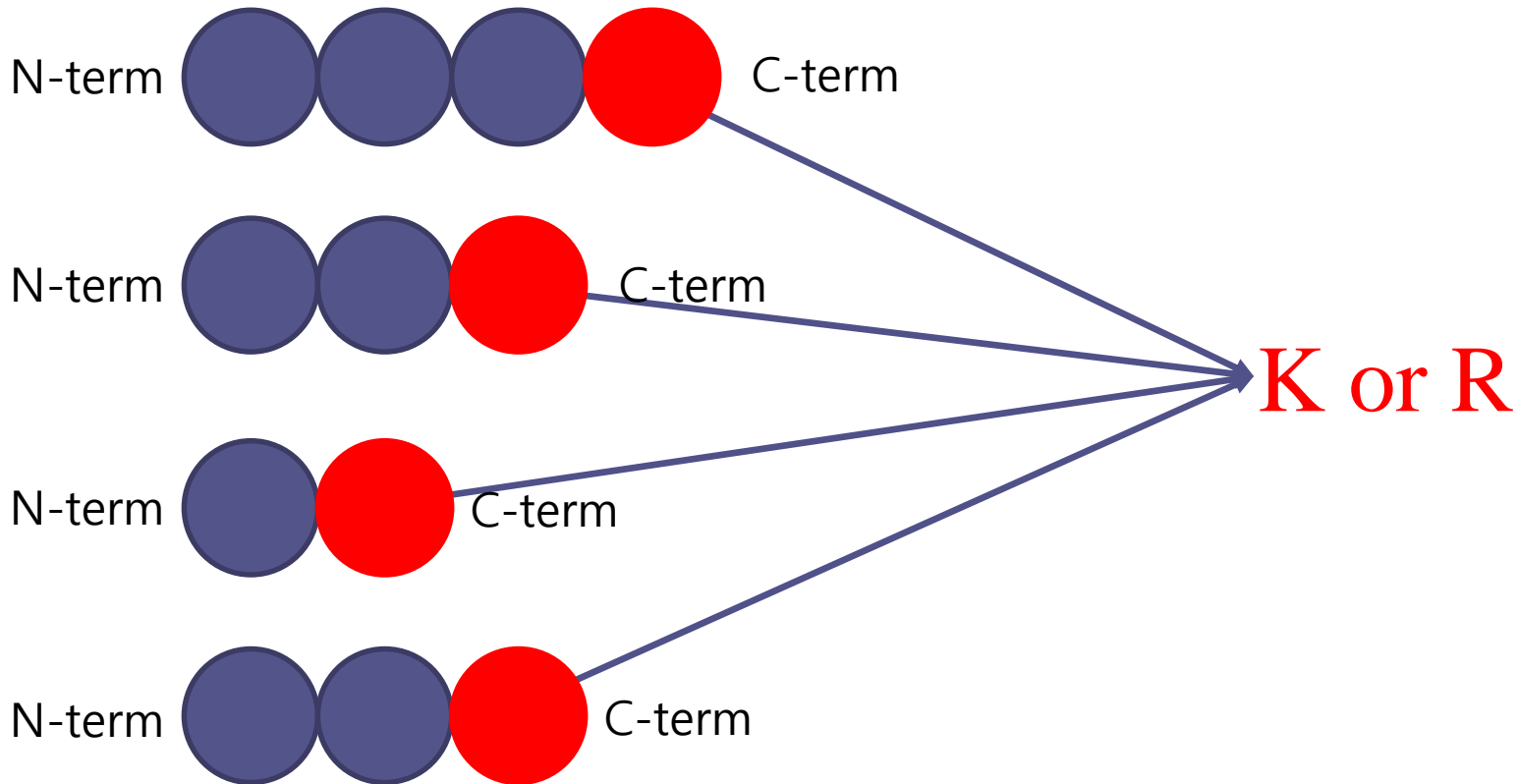
Digesting the Protein

- **Protein digestion(Trypsin)**



Digesting the Protein

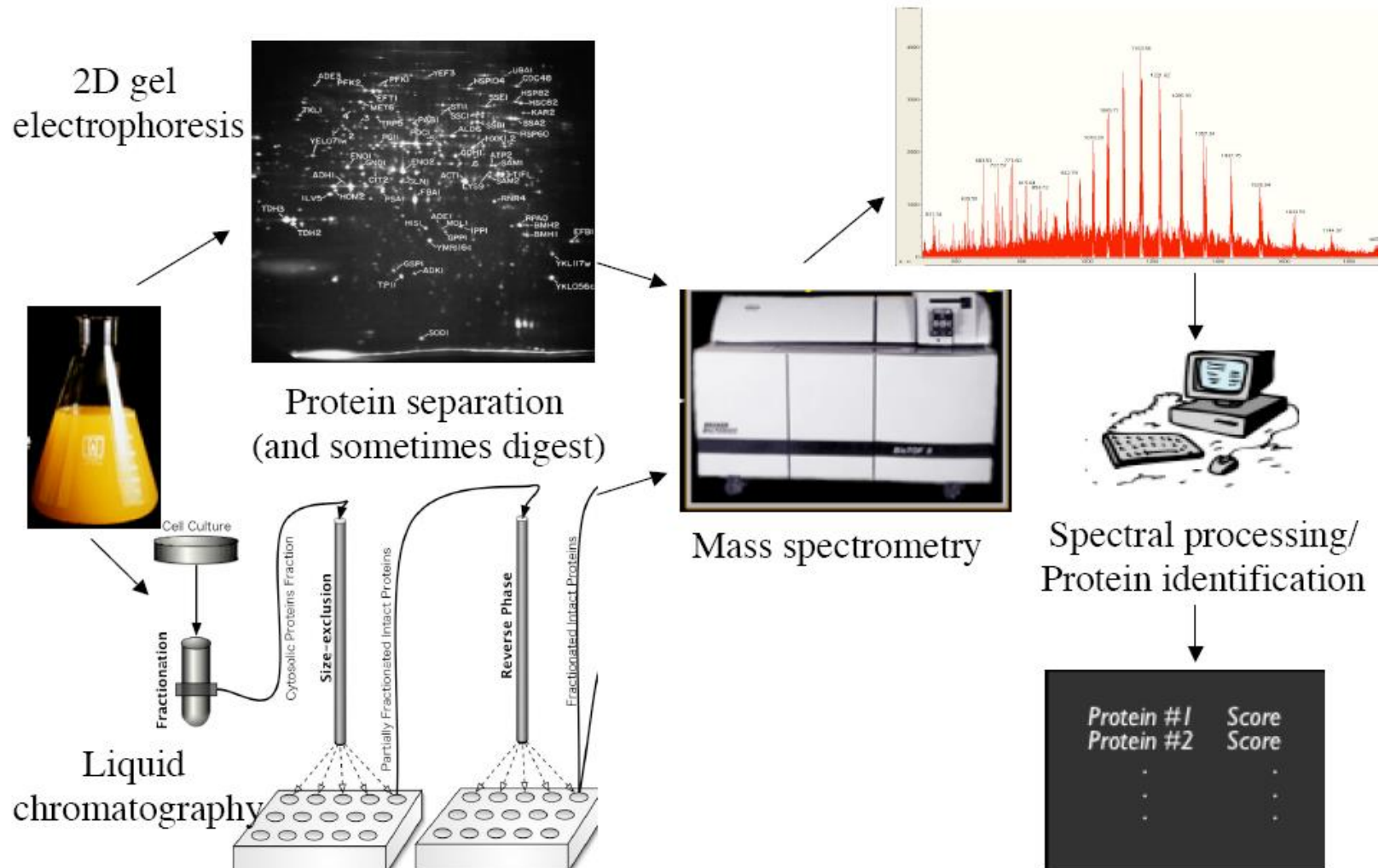
- **Protein digestion(Trypsin)**



Proteomics

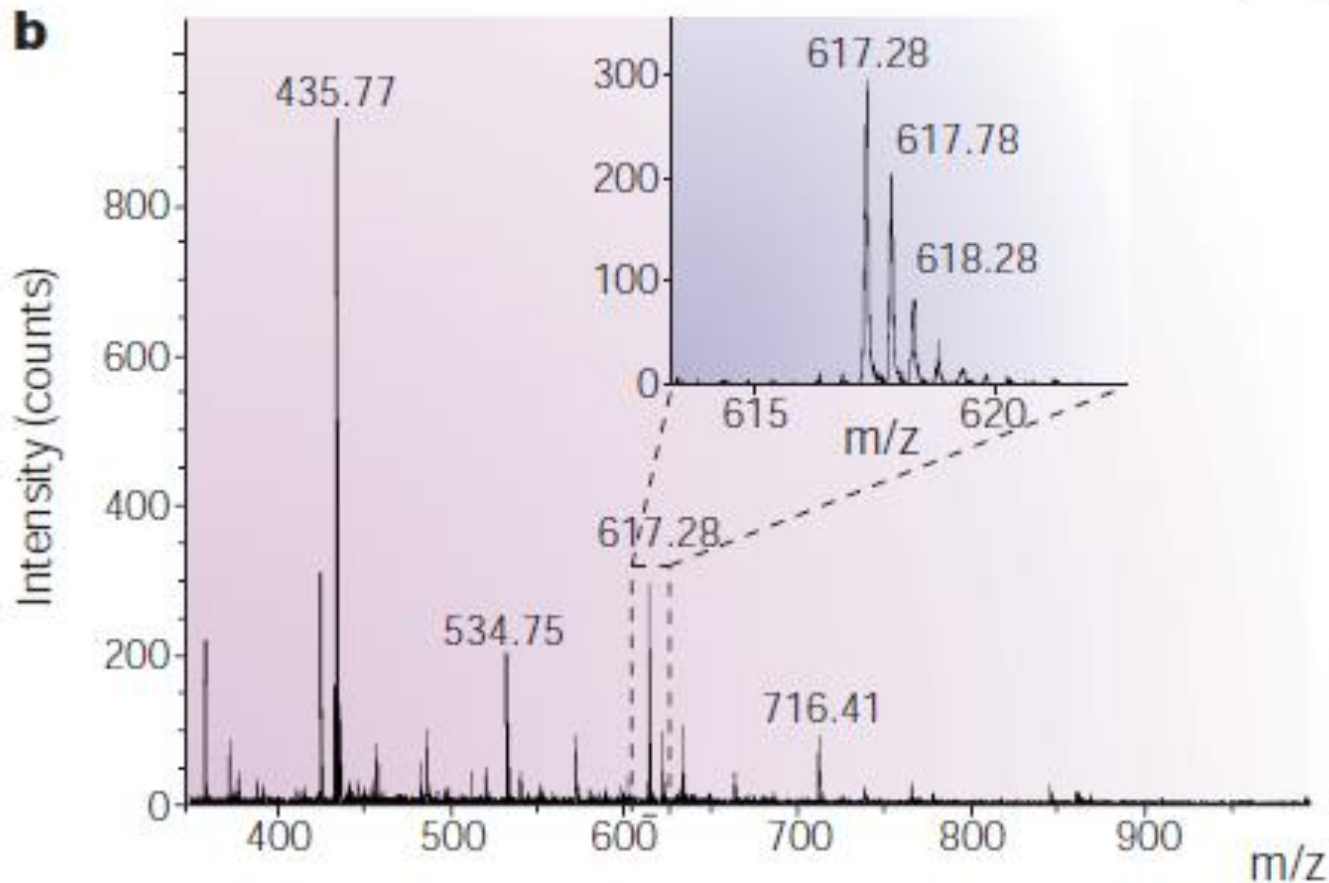
- **Bottom-up method**
 - Protein digestion
 - Mass spectrometry analysis
 - Peptide identification using MS/MS
 - Protein Assembly using identified peptides

Mass Spectrometry



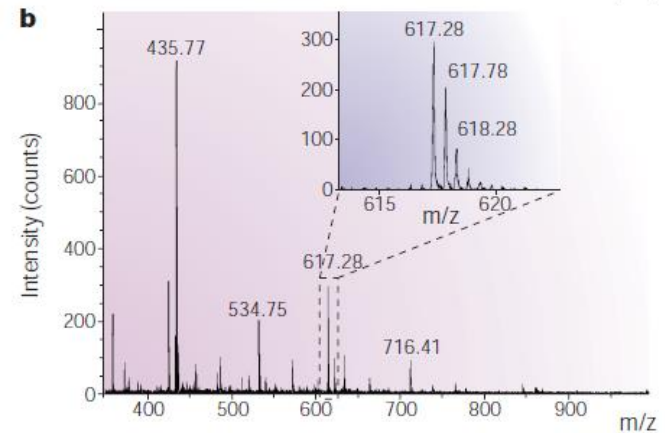
Mass Spectrometry

- 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

- $$\begin{aligned} & (1232.55 + (2 * 1.0073)) / 2 \\ &= (1232.55 / 2) + ((2 * 1.0073) / 2) = (1232.55 / 2) + 1.0073 \\ &= 617.28 \end{aligned}$$

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

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

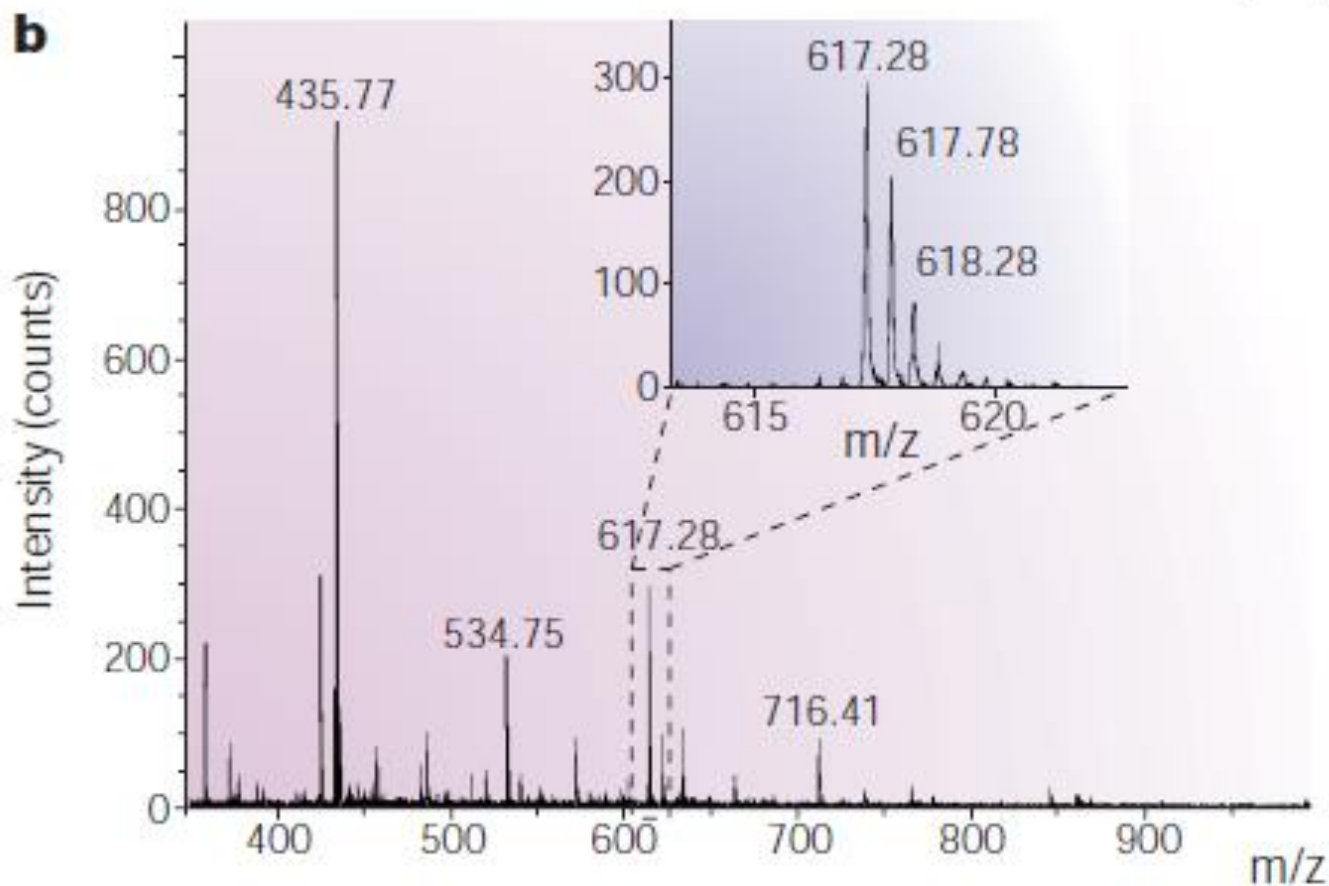
- Mass = 1848.84, Charge state = 3

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

$m/z = 617.28 !!!$

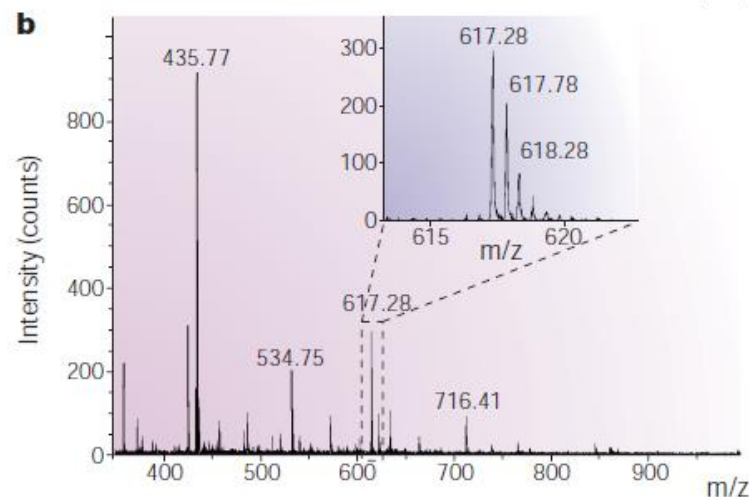
Mass Spectrometry

- How can we decide the charge state?



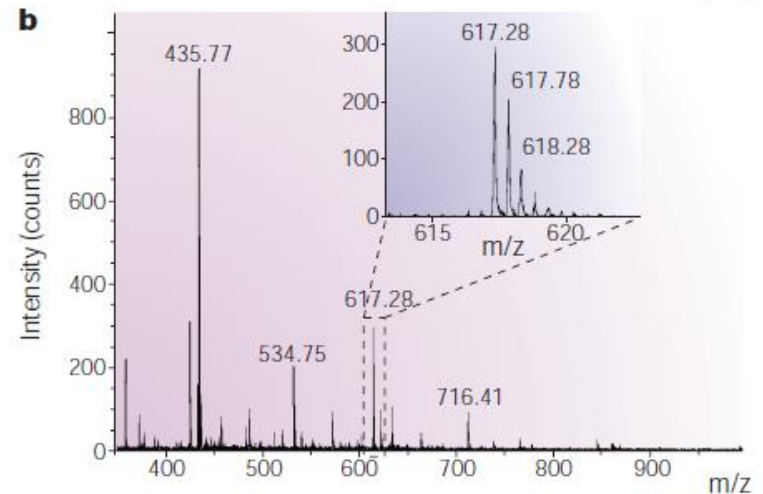
Mass Spectrometry

- **The peptide mass**
 - $m/z = 617.28$, Charge state = 2
 - $(617.28 - 1.0073) * 2 = 1232.55$
 - The peptide mass = 1232.55



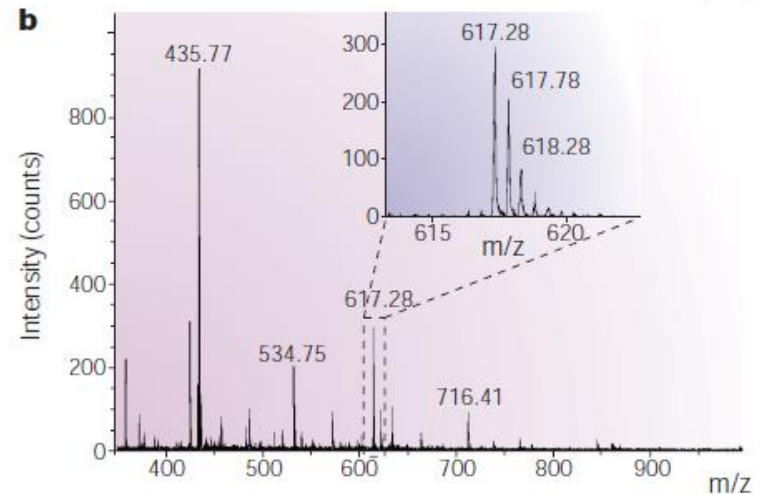
Mass Spectrometry

- **The peptide mass**
 - $m/z = 617.28$, 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**
 - $m/z = 617.28$, Charge state = 2
 - $(617.28 - 1.0073) * 2 = 1232.55$
 - The peptide mass = 1232.55

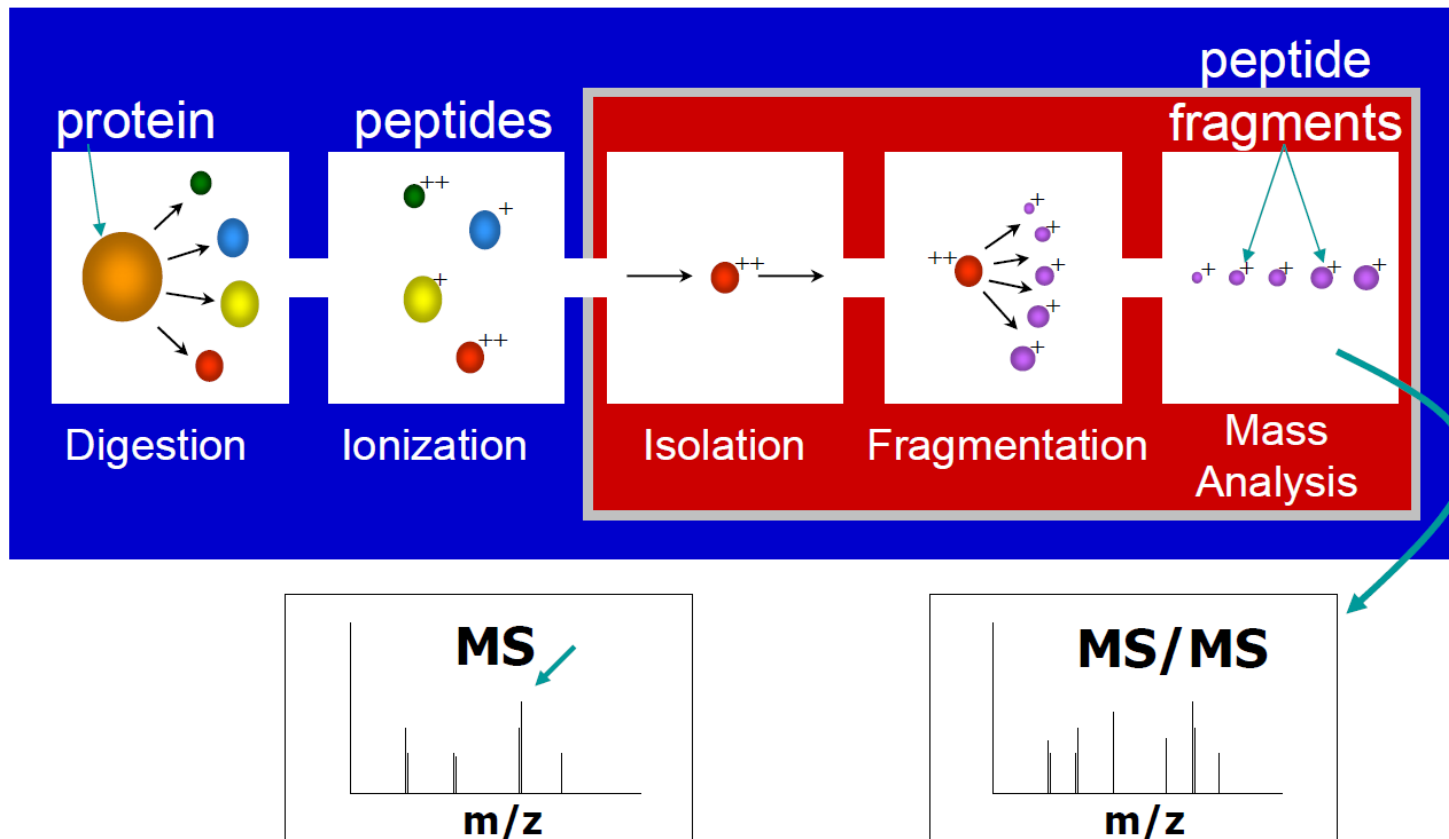


- **How can we know sequences?**

Tandem mass spectrum (MS/MS)

Tandem Mass Spectrum

- Tandem mass spectrum

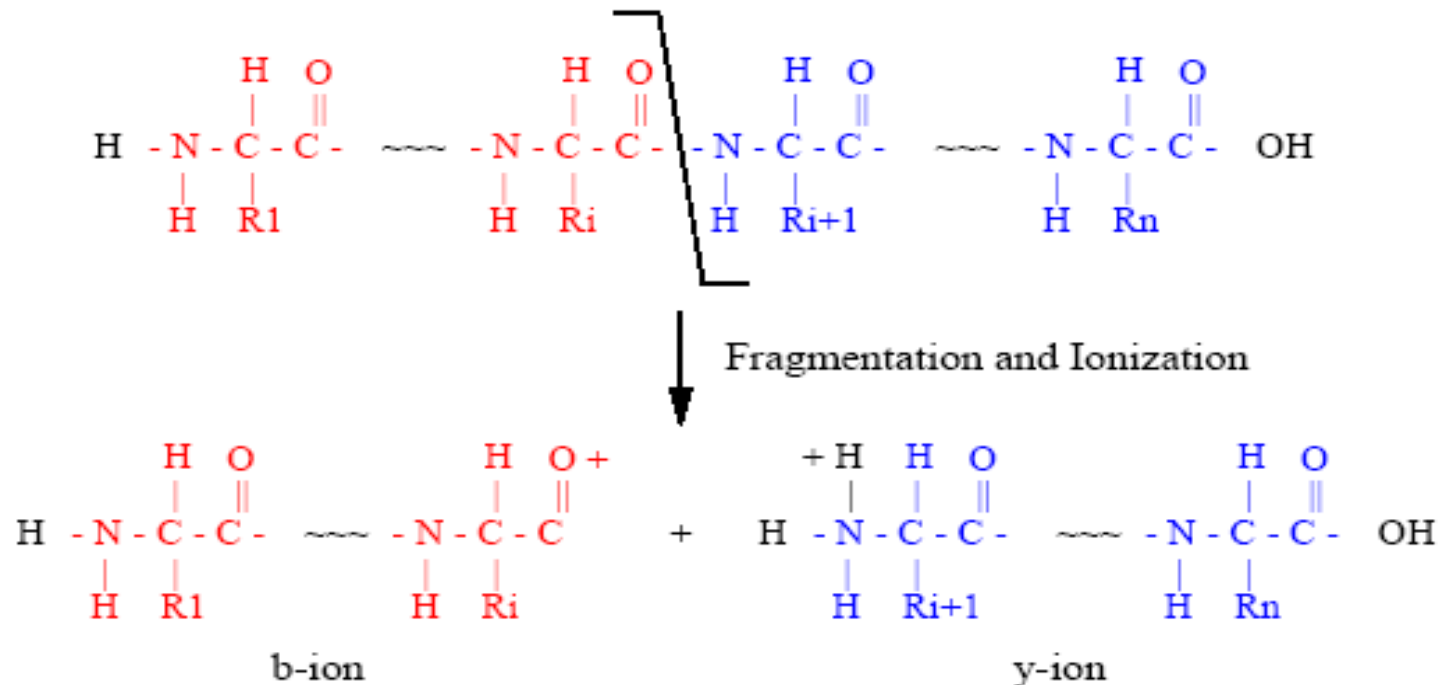


Tandem Mass Spectrum

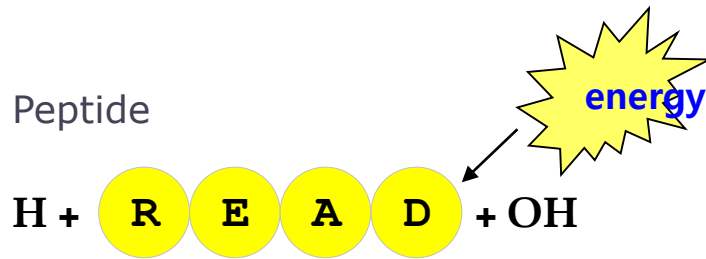
- **Fragmentation**
 - Method: HCD, CID, ETD, ECD
 - LPKKNK
 - LPKKNK
 - L + PKNK
 - LP + KKNK
 - LPK + NK
 - LPKN + K

Tandem Mass Spectrum

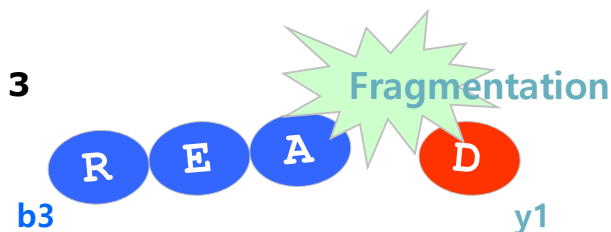
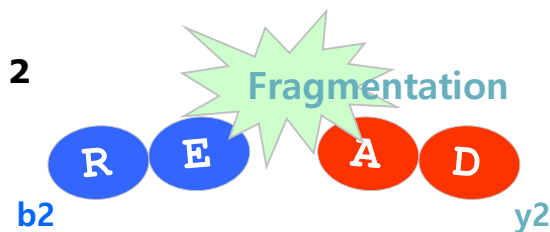
- Fragmentation



Tandem Mass Spectrum

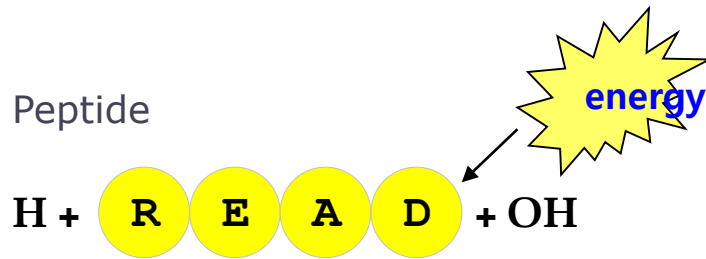


Amino Acid	Mass
A	71
D	115
E	129
R	156

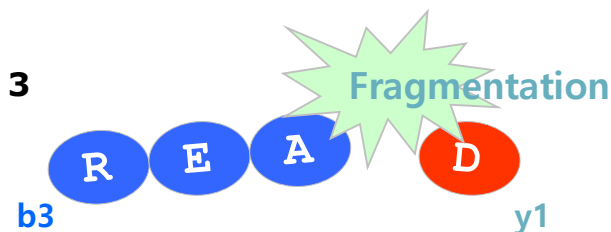
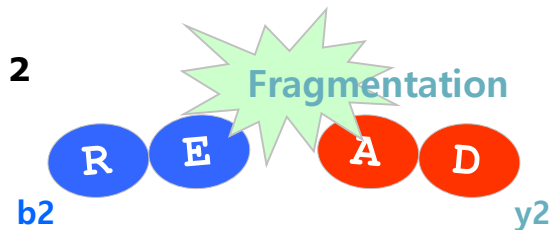


- **b1 mass = R + 1 = 157**
- **y1 mass = D + 19 = 134**

Tandem Mass Spectrum

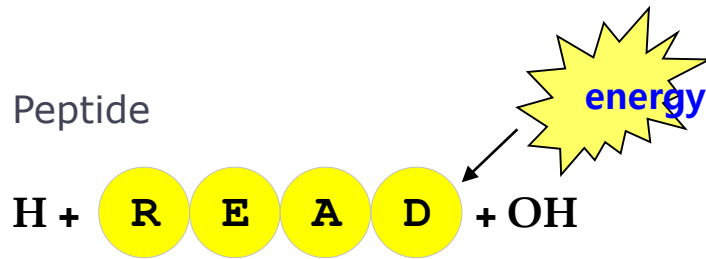


Amino Acid	Mass
A	71
D	115
E	129
R	156

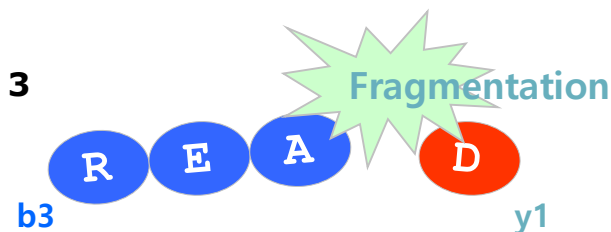
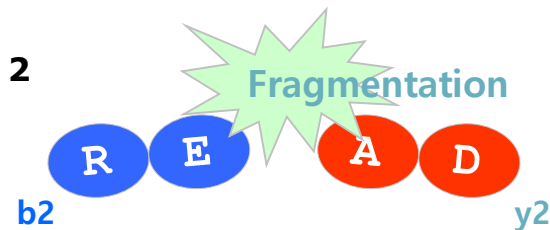


- **b1 mass = R + 1 = 157**
- **y1 mass = D + 19 = 134**
- **Quiz**
- **b2 mass = ?**
- **y3 mass = ?**

Tandem Mass Spectrum

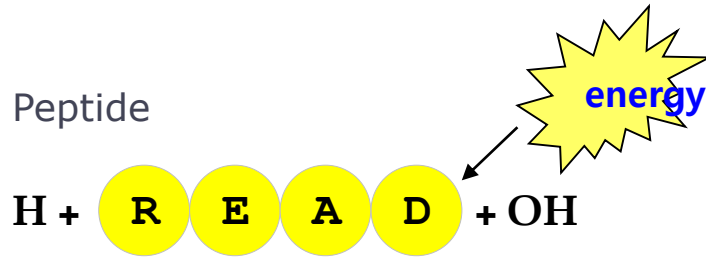


Amino Acid	Mass
A	71
D	115
E	129
R	156

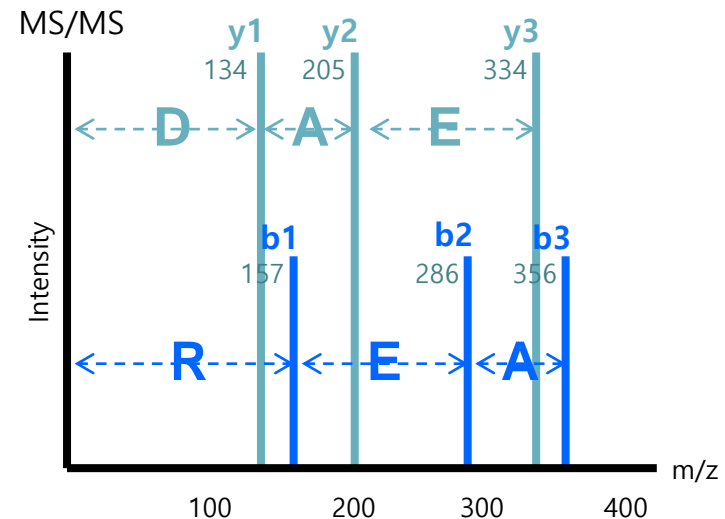
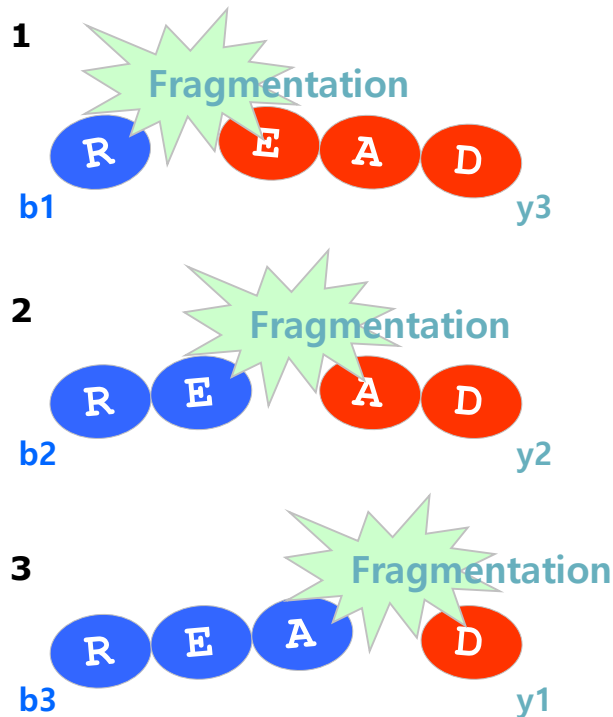


- **b1 mass = R + 1 = 157**
- **y1 mass = D + 19 = 134**
- **Quiz**
- **b2 mass = R + E + 1 = 286**
- **y3 mass = E + A + D + 19 = 334**

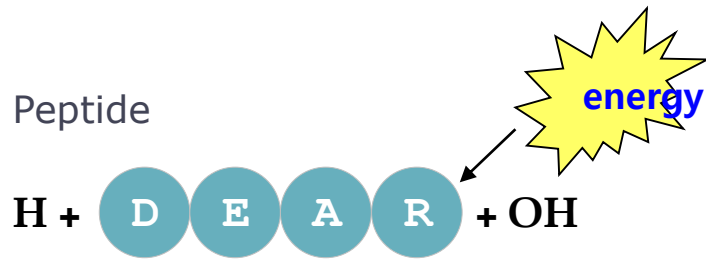
Tandem Mass Spectrum



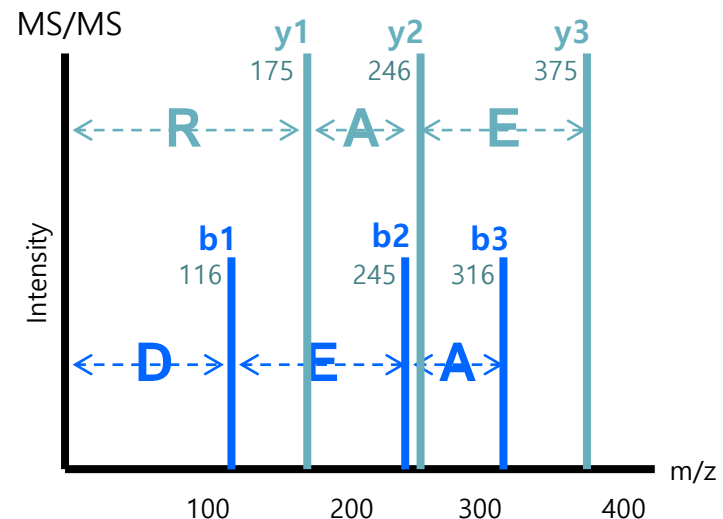
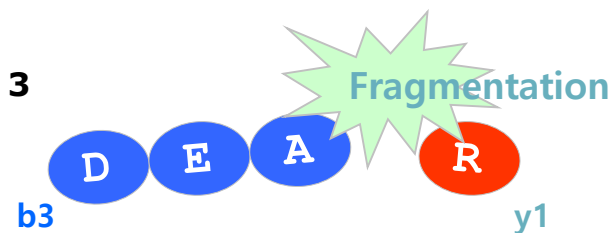
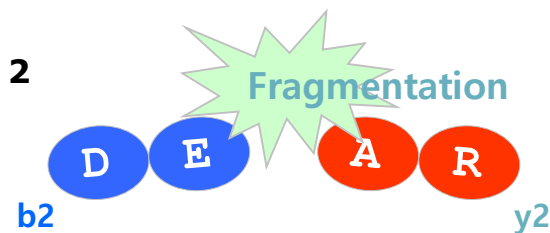
Amino Acid	Mass
A	71
D	115
E	129
R	156



Tandem Mass Spectrum (Quiz)

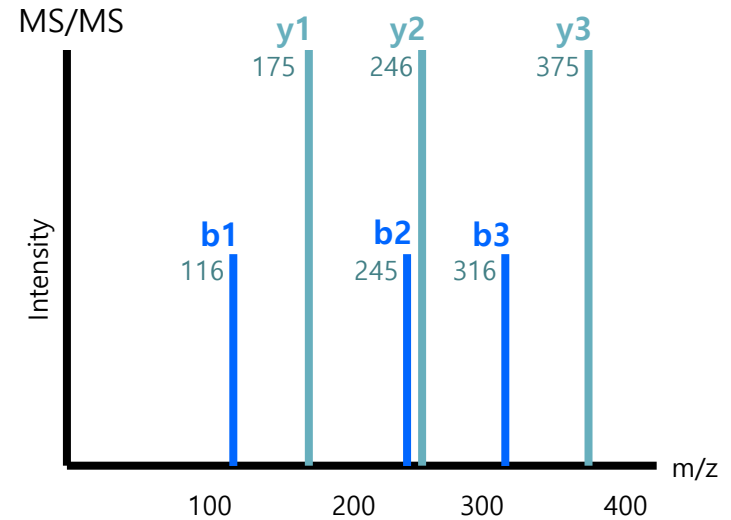
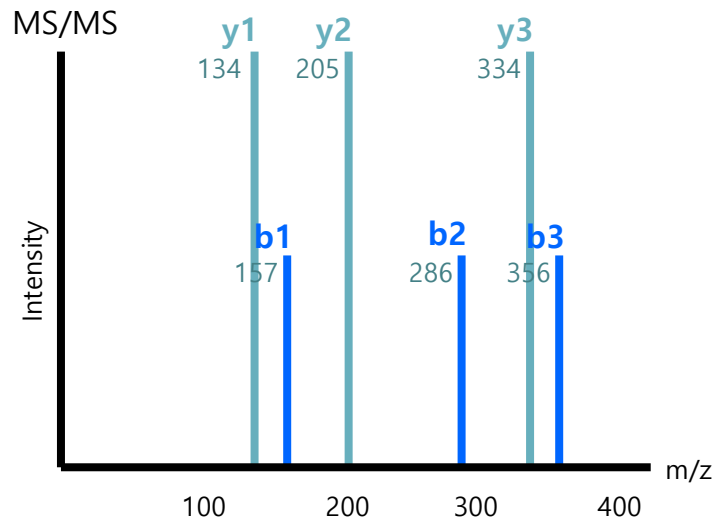


Amino Acid	Mass
A	71
D	115
E	129
R	156



Tandem Mass Spectrum

- **READ vs DEAR**
 - Peptide mass = 471

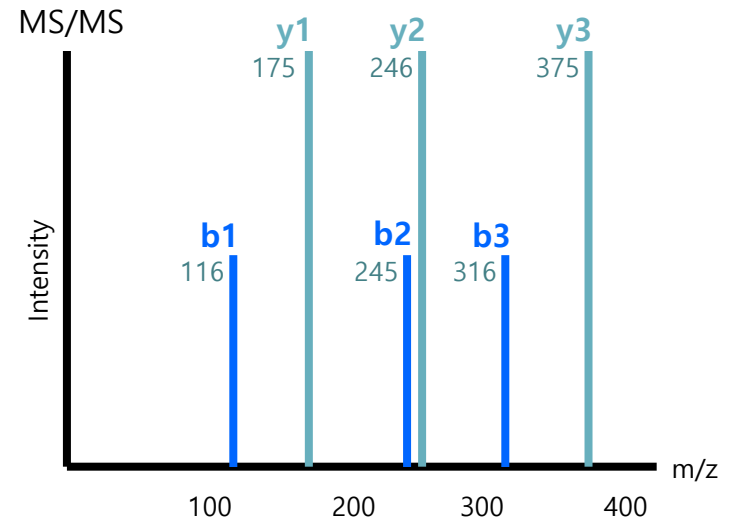
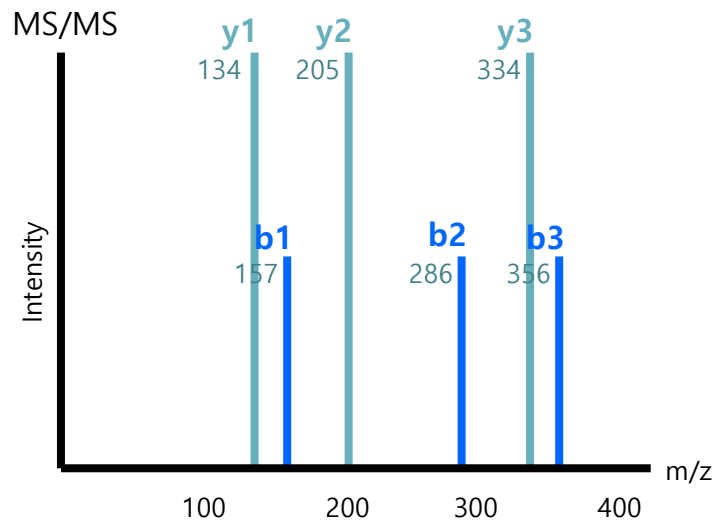


Proteomics

- **Bottom-up method**
 - Protein digestion
 - Mass spectrometry analysis
 - Peptide identification using MS/MS
 - Protein Assembly using identified peptides

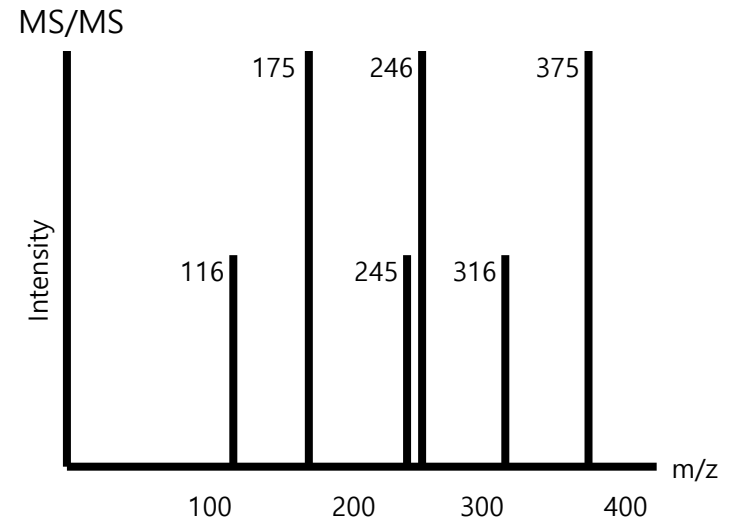
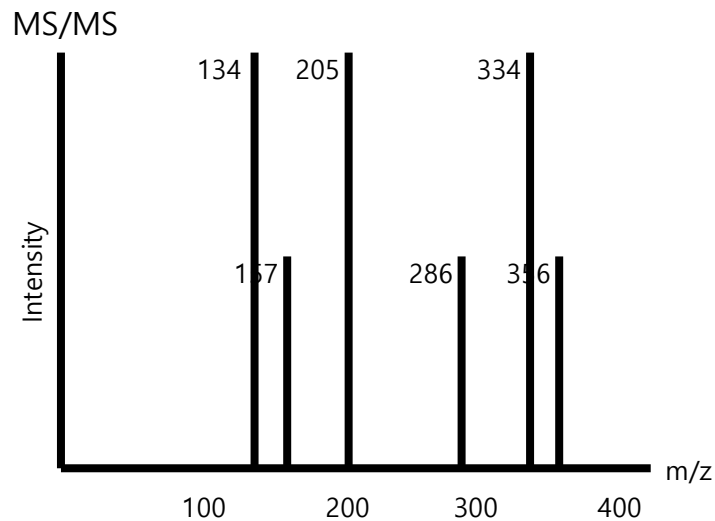
Peptide Identification using MS/MS

- **READ vs DEAR**
 - Peptide mass = 471



Peptide Identification using MS/MS

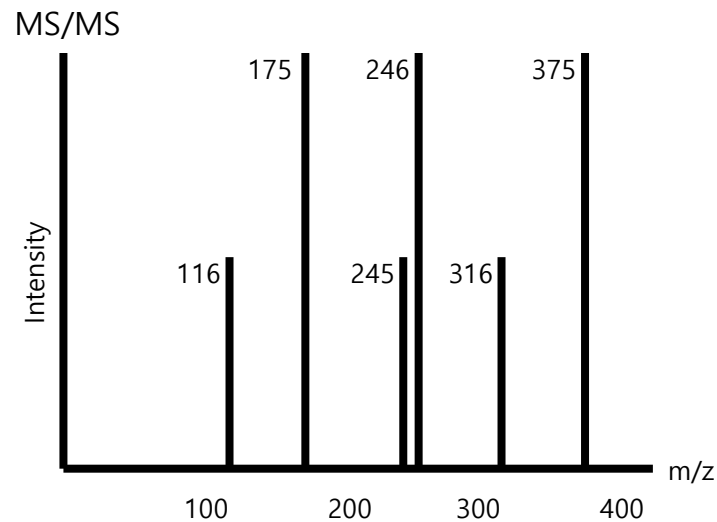
- Peptide mass = 471



Peptide Identification using MS/MS

- **De novo sequencing approach**
 - Peptide mass = 471

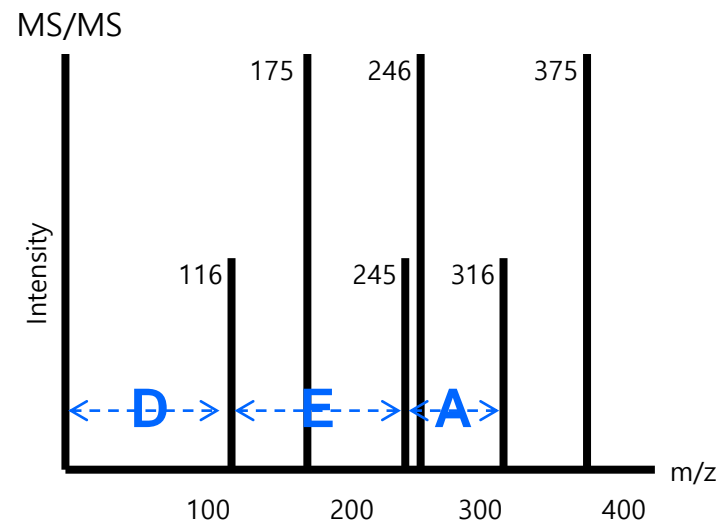
Amino Acid	Mass
A	71
D	115
E	129
R	156



Peptide Identification using MS/MS

- **De novo sequencing approach**
 - Peptide mass = 471

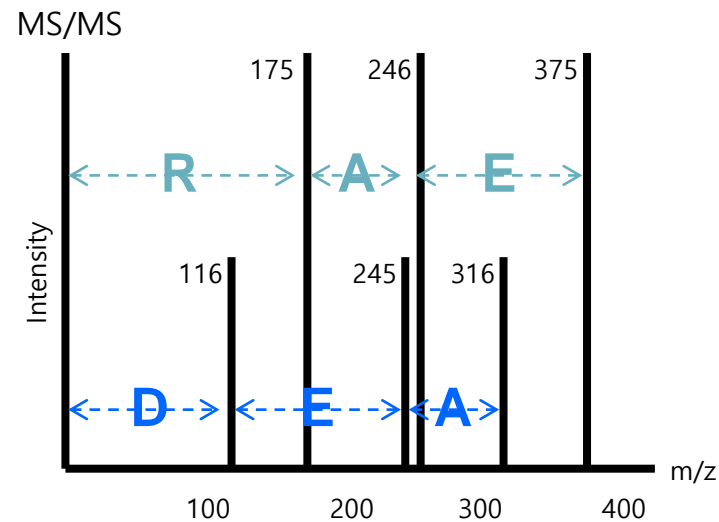
Amino Acid	Mass
A	71
D	115
E	129
R	156



Peptide Identification using MS/MS

- **De novo sequencing approach**
 - Peptide mass = 471

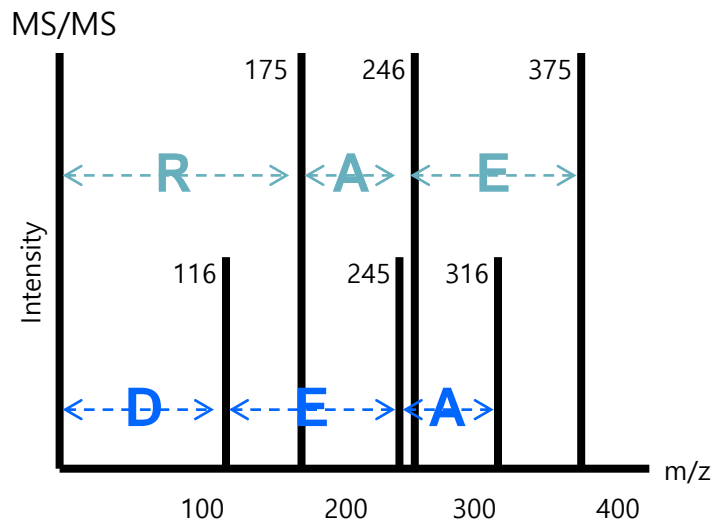
Amino Acid	Mass
A	71
D	115
E	129
R	156



Peptide Identification using MS/MS

- **De novo sequencing approach**
 - Peptide mass = 471

Amino Acid	Mass
A	71
D	115
E	129
R	156

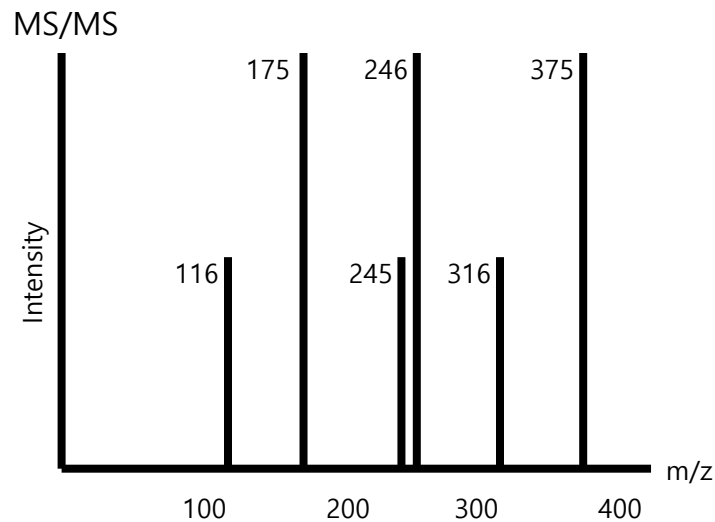


정답: DEAR

Peptide Identification using MS/MS

- **Database search approach**
 - Peptide mass = 471

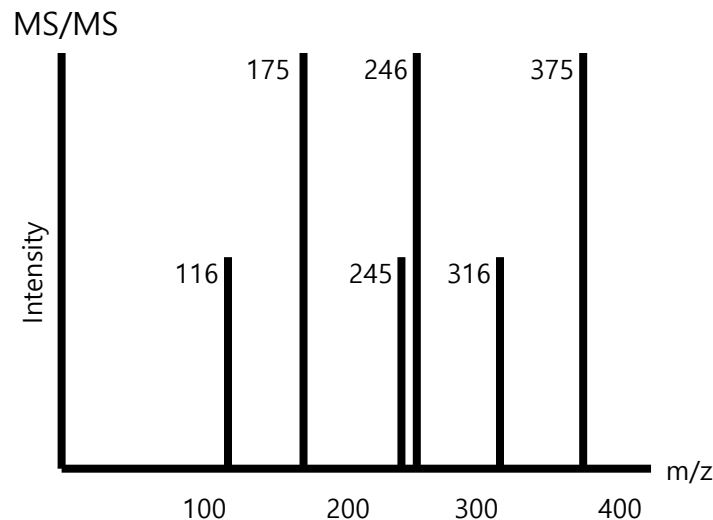
Amino Acid	Mass
A	71
D	115
E	129
R	156



Peptide Identification using MS/MS

- **Database search approach**
 - Peptide mass = 471

Amino Acid	Mass
A	71
D	115
E	129
R	156

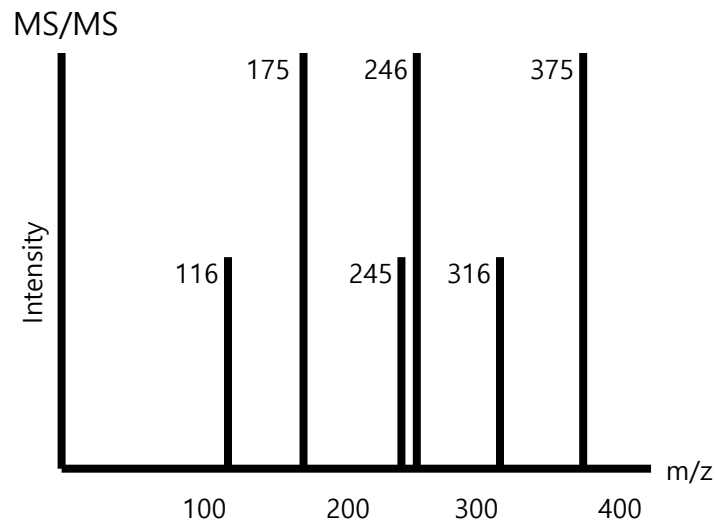


Database
ADER
DEAR
READ
REDA

Peptide Identification using MS/MS

- **Database search approach**
 - Peptide mass = 471

Amino Acid	Mass
A	71
D	115
E	129
R	156

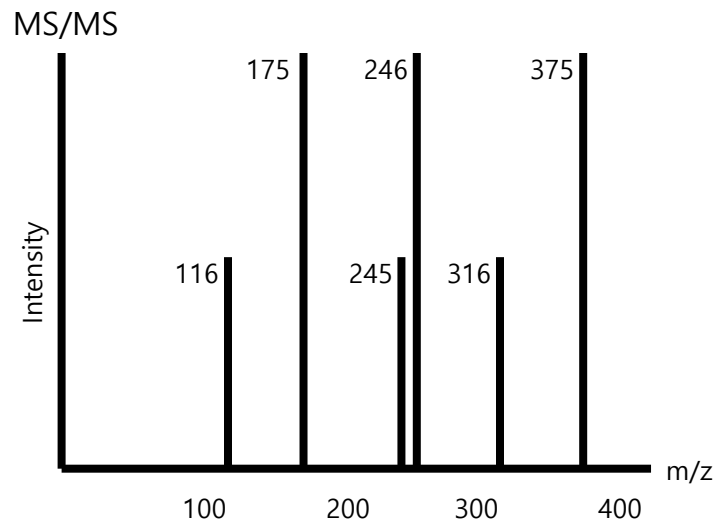


Database
ADER 1.7
DEAR
READ
REDA

Peptide Identification using MS/MS

- **Database search approach**
 - Peptide mass = 471

Amino Acid	Mass
A	71
D	115
E	129
R	156



Database

ADER

DEAR

READ

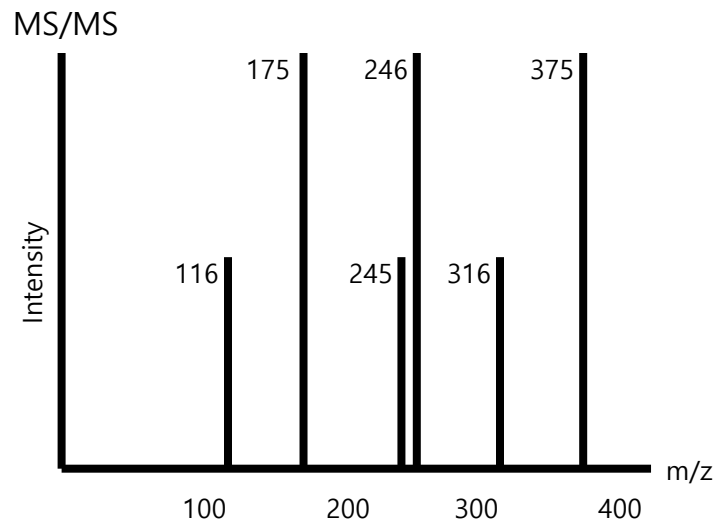
REDA

9.8

Peptide Identification using MS/MS

- **Database search approach**
 - Peptide mass = 471

Amino Acid	Mass
A	71
D	115
E	129
R	156



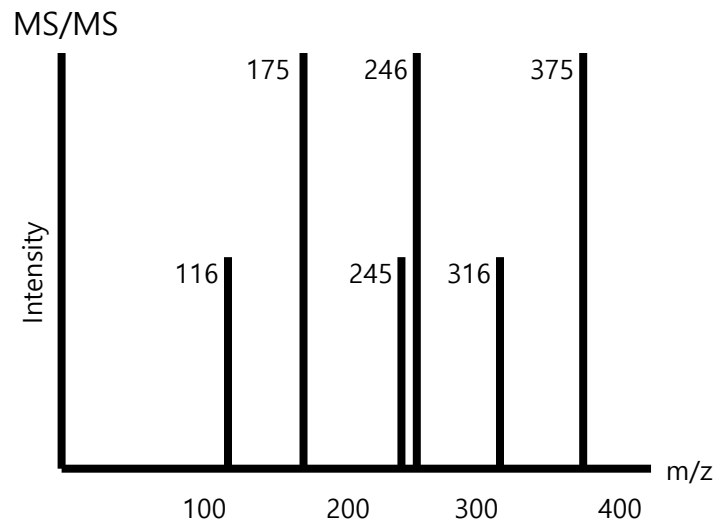
Database
ADER
DEAR
READ
REDA

5.7

Peptide Identification using MS/MS

- **Database search approach**
 - Peptide mass = 471

Amino Acid	Mass
A	71
D	115
E	129
R	156



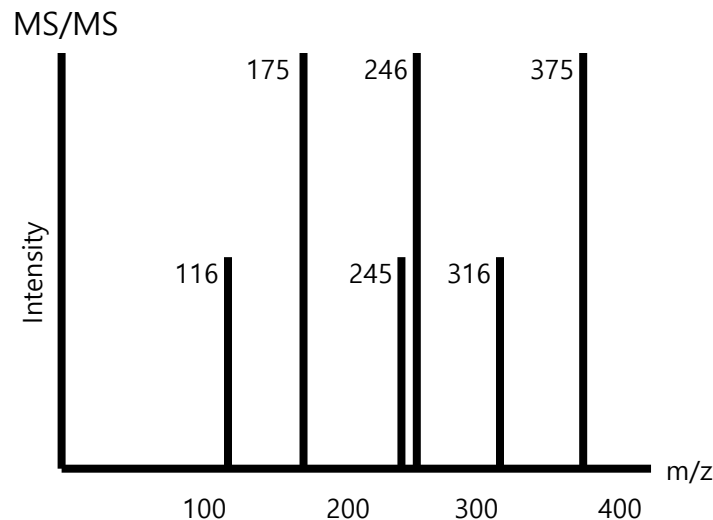
Database
ADER
DEAR
READ
REDA

3.3

Peptide Identification using MS/MS

- **Database search approach**
 - Peptide mass = 471

Amino Acid	Mass
A	71
D	115
E	129
R	156



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
PKNKNRNRYSRDVSPFDHSRKLHQEDNDYINASLIKMEEAQRSYILT
Q

- Protein 2

- MTVYAVTGGAEF**LGRYIVKLLISADD**VQEIRVINVV**EDPQPLVSKVK**
VINYIQCDINDLIR

- Peptide

- LGRYIVKLLISADD
 - DQPPLVSKVK

Protein Inference

- **Protein Inference**

- Protein 1

- MEMEKEFEQIDKSGSWA**GGAEFLGRY**AIYQDIRHEASDFPCRVAKL
PKNKNRNRNRYRDVSPFDHSRKLHQEDNDYINASLIKMEEAQRSYILT
Q

- Protein 2

- MTVYAVT**GGAEFLGRY**IVKLLISADDVQEIRVINVVEDPQPLVSKVK
VINYIQCDINDLIR

- Peptide

- GGAEFLGRY