

Protein Identification: Review

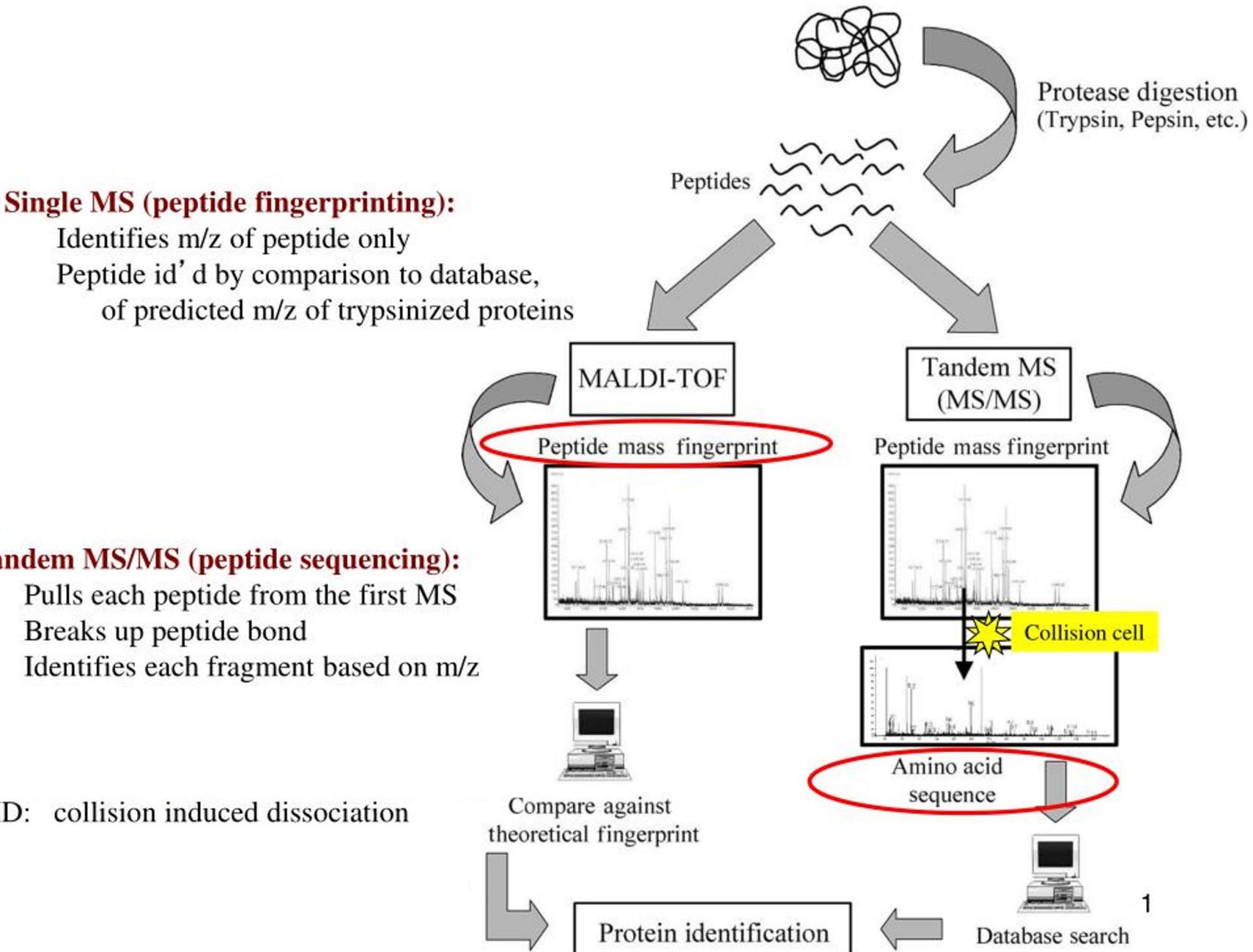
Single MS (peptide fingerprinting):

Identifies m/z of peptide only
Peptide id'd by comparison to database,
of predicted m/z of trypsinized proteins

Tandem MS/MS (peptide sequencing):

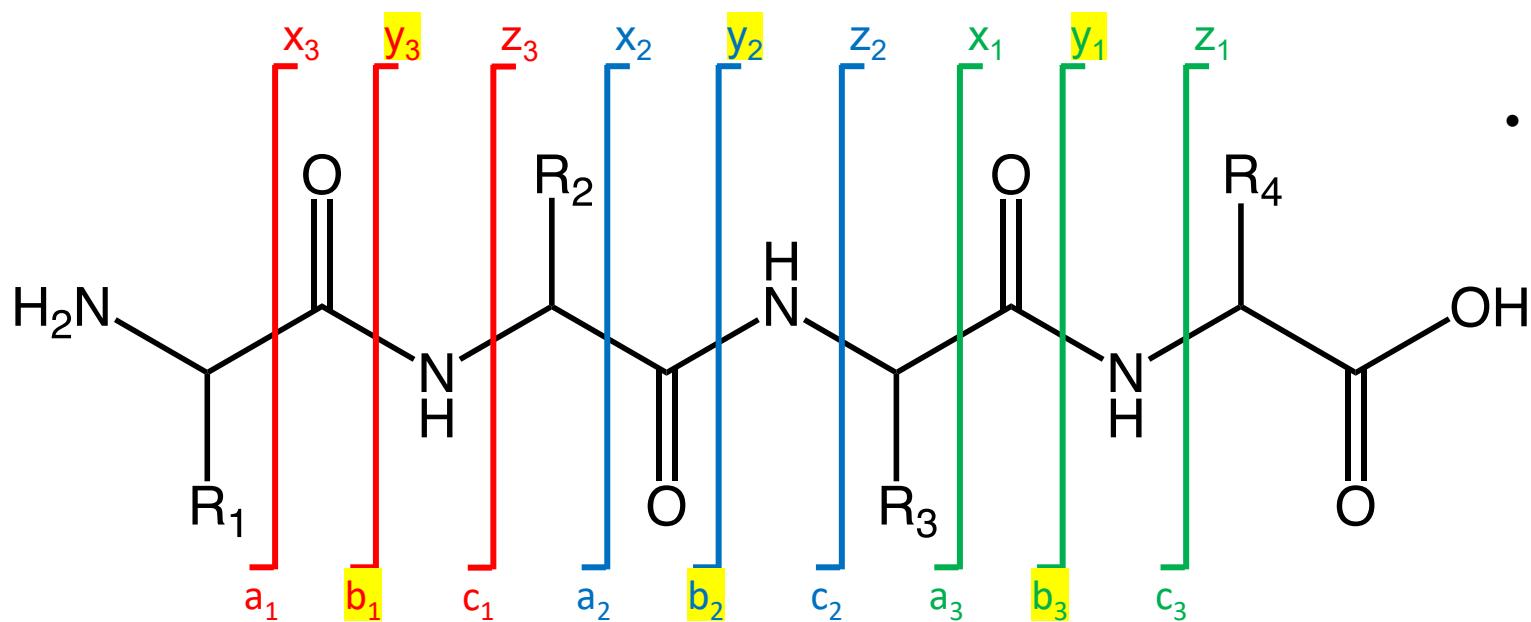
Pulls each peptide from the first MS
Breaks up peptide bond
Identifies each fragment based on m/z

CID: collision induced dissociation



Nomenclature of Fragment Ions: Review

Charge stays on *either* the 'left' (a, b, or c) or 'right' (x, y, or z) side of cleavage



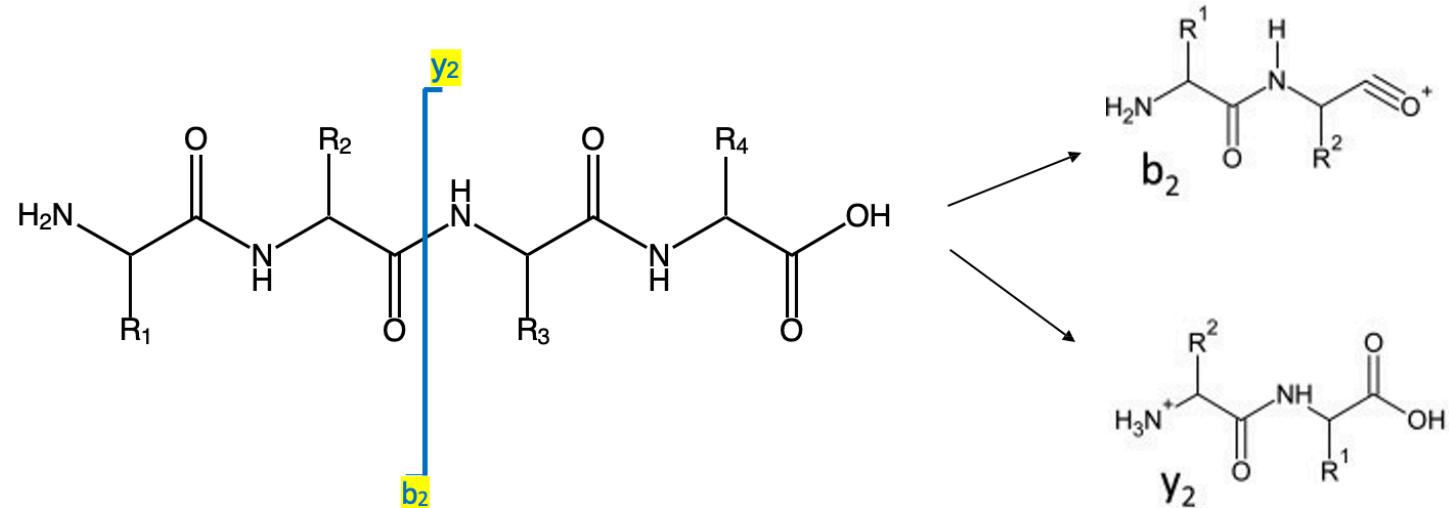
- Letter: Indicates the bond broken and the terminus contained in the fragment
- Number: Indicates the number of $C\alpha$ in the fragment

Peptides Fragment by CID: Review

For a singly protonated peptide,

Singly charged N term ion ($+H^+$) and neutral C-term
Or
Neutral N term and Singly charged C-term ion ($+H^+$)

For a doubly protonated peptide, **both** N- and C-terminal fragments can be generated from a single dissociate event.



Fragmentation Results in a Peptide “Ladder”: Review

S-P-A-F-D-S-I-M-A-E-T-L-K

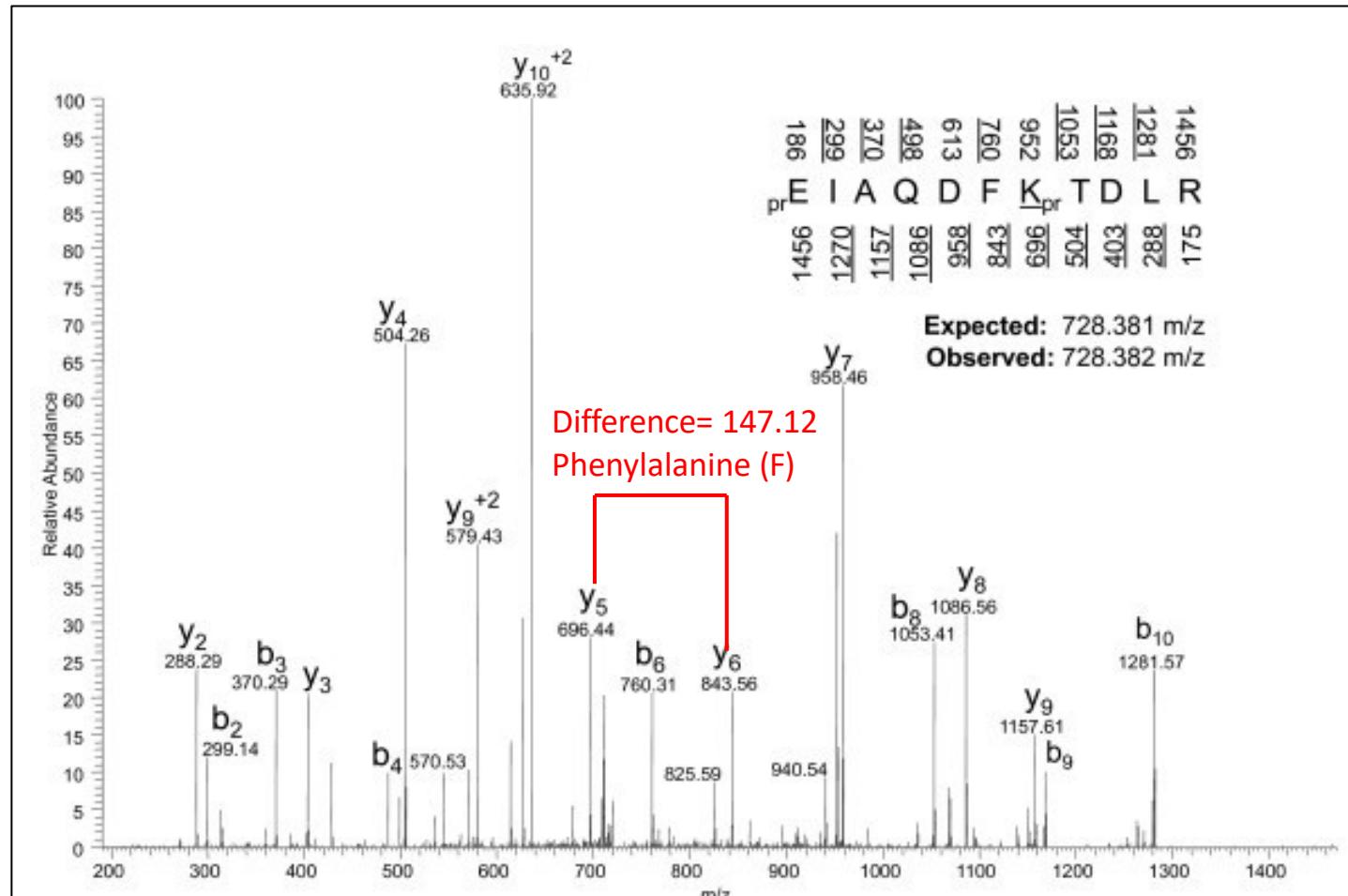
(Charge +1 = m/z of 1410.6)

(Charge +2 = m/z of 705.3)

(Charge +3 = m/z of 470.2)

<u>mass¹⁺</u>	<u>b-ions</u>	<u>y-ions</u>	<u>mass¹⁺</u>
b_1^+ 88.1	S	PAFD SIMAE TLK	1323.6 y_{12}^+
b_2^+ 185.2	SP	AFDS IMAET LK	1226.4 y_{11}^+
b_3^+ 256.3	SPA	FDS IMAET LK	1155.4 y_{10}^+
b_4^+ 403.5	SPAF	DS IMAET LK	1008.2 y_9^+
b_5^+ 518.5	SPAFD	S IMAET LK	893.1 y_8^+
b_6^+ 605.6	SPAFDS	IMA E TLK	806.0 y_7^+
b_7^+ 718.8	SPAFDSI	MA E TLK	692.3 y_6^+
b_8^+ 850.0	SPAFDSIM	A E TLK	561.7 y_5^+
b_9^+ 921.1	SPAFDSIMA	ETLK	490.6 y_4^+
b_{10}^+ 1050.2	SPAFDSIMAE	TLK	361.5 y_3^+
b_{11}^+ 1151.3	SPAFDSIMAET	LK	260.4 y_2^+
b_{12}^+ 1264.4	SPAFDSIMAE T L	K	147.2 y_1^+

Mass Spectrum (Assignment of *b*- and *y*-ions): Review

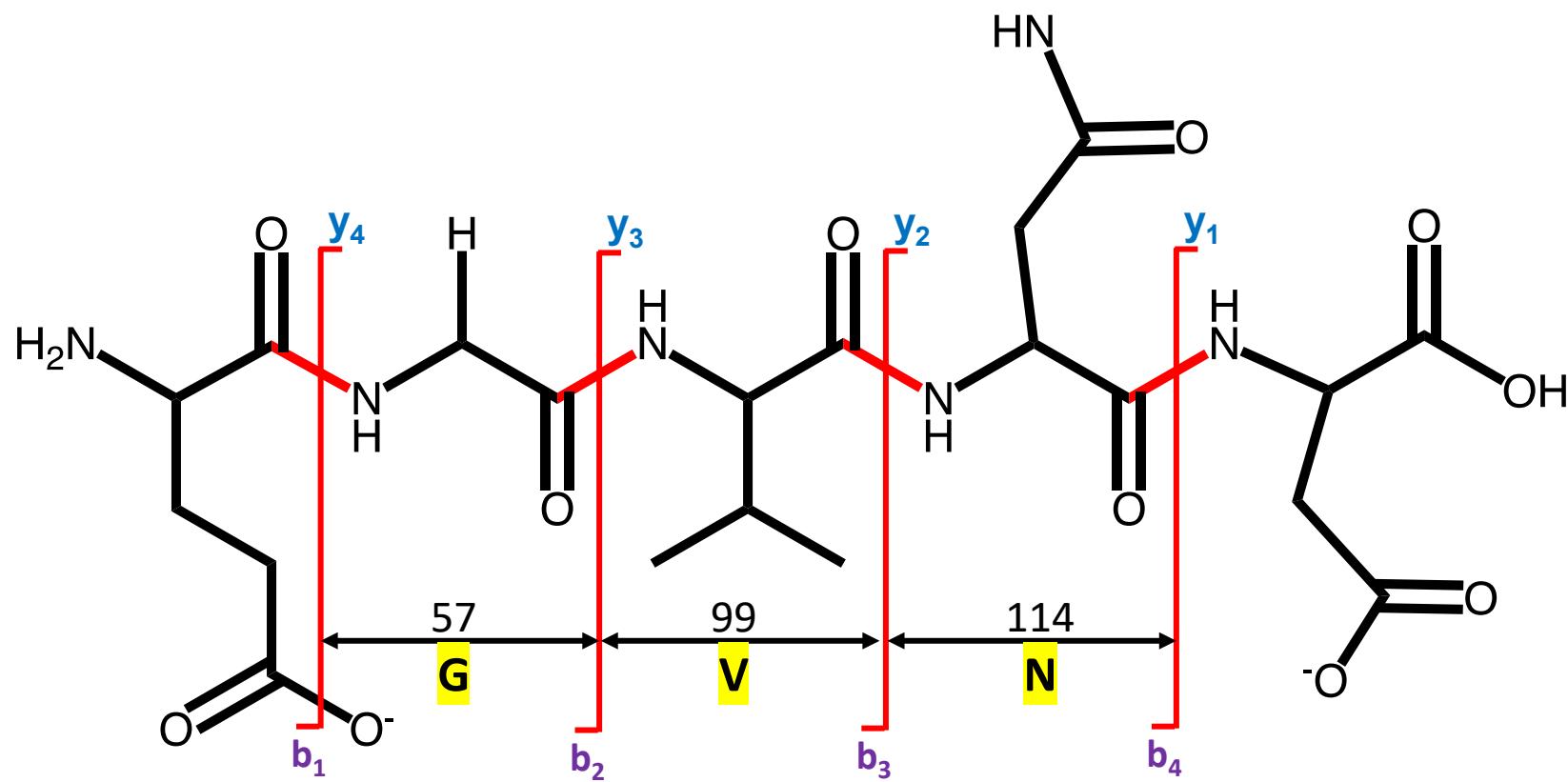


- Mixture of *b* ions and *y* ions
- MS/MS of 2⁺ charged tryptic peptides yield (often) 1⁺ charged product ions (but 2⁺ charged products can be observed as well)
- Not all *b* ions or *y* ions are visible

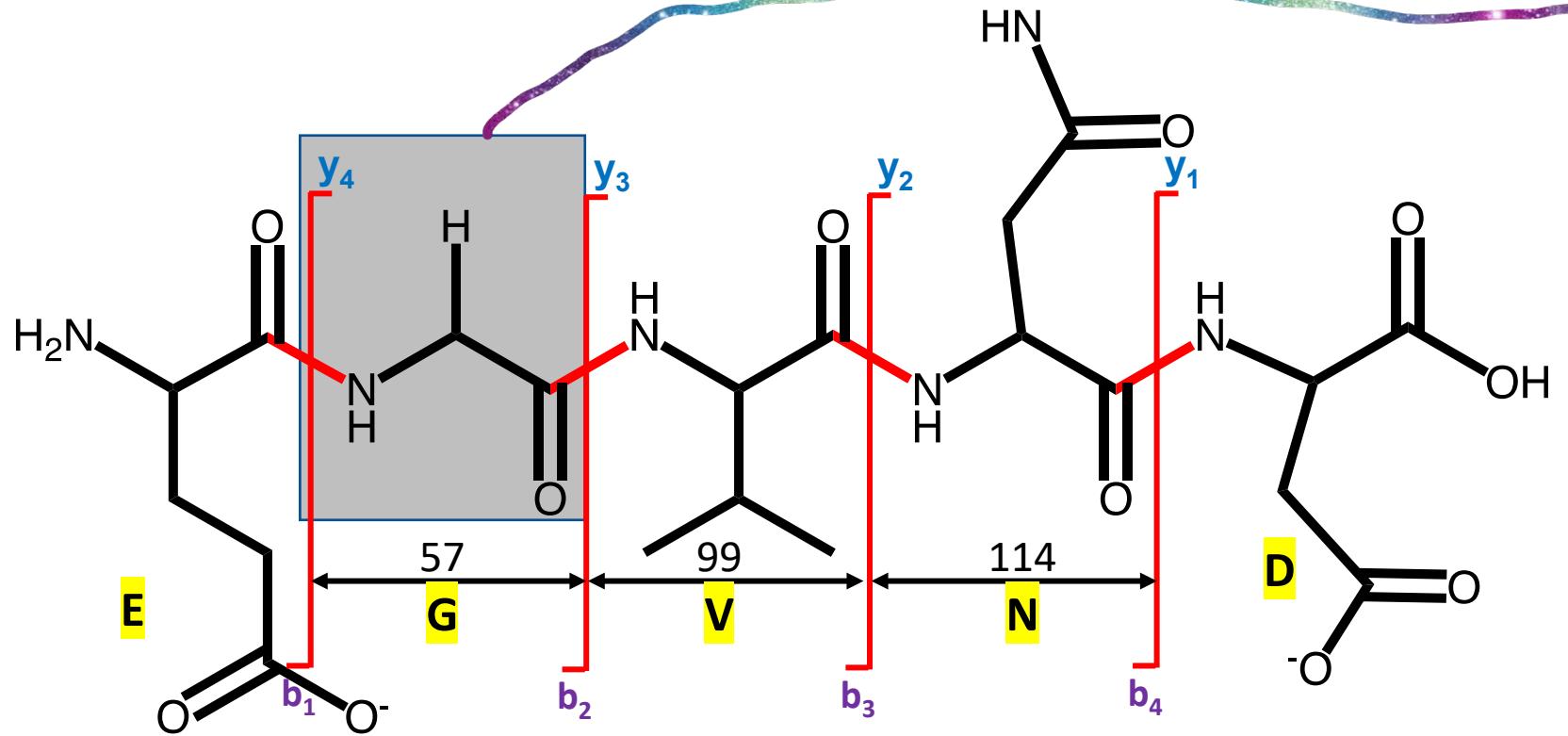
The mass of the precursor is 1454.764

the observed ion was doubly charged
 $728.382 \times 2 - 2 = 1454.764$ Da

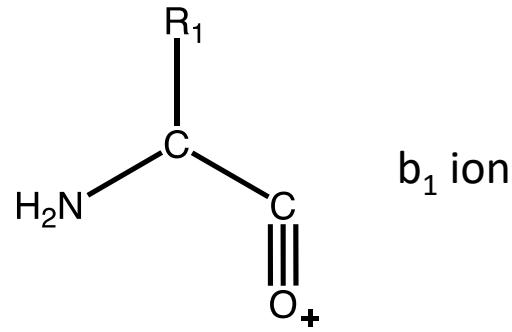
- Amino acid sequence can be deduced by the Δmass between adjacent y ion peaks **or** adjacent b ion peaks



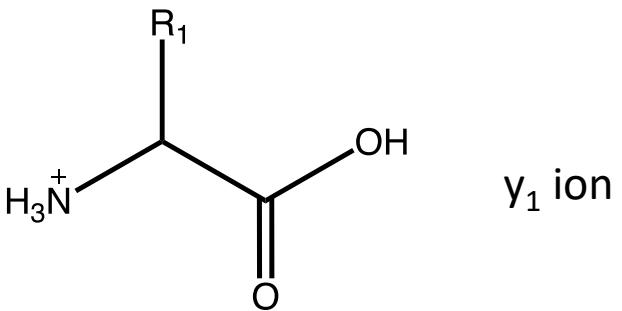
Code (1 letter)	Monoisotopic mass
G	57.021 47
A	71.037 12
S	87.032 03
P	97.052 77
V	99.068 42
T	101.047 68
C	103.009 19
I	113.084 07
L	113.084 07
N	114.042 93
D	115.026 95
Q	128.058 58
K	128.094 97
E	129.042 60
M	131.040 49
H	137.058 91
F	147.068 42
R	156.101 12
Y	163.063 33
W	186.079 32



Code (1 letter)	Monoisotopic mass
G	57.021 47
A	71.037 12
S	87.032 03
P	97.052 77
V	99.068 42
T	101.047 68
C	103.009 19
I	113.084 07
L	113.084 07
N	114.042 93
D	115.026 95
Q	128.058 58
K	128.094 97
E	129.042 60
M	131.040 49
H	137.058 91
F	147.068 42
R	156.101 12
Y	163.063 33
W	186.079 32

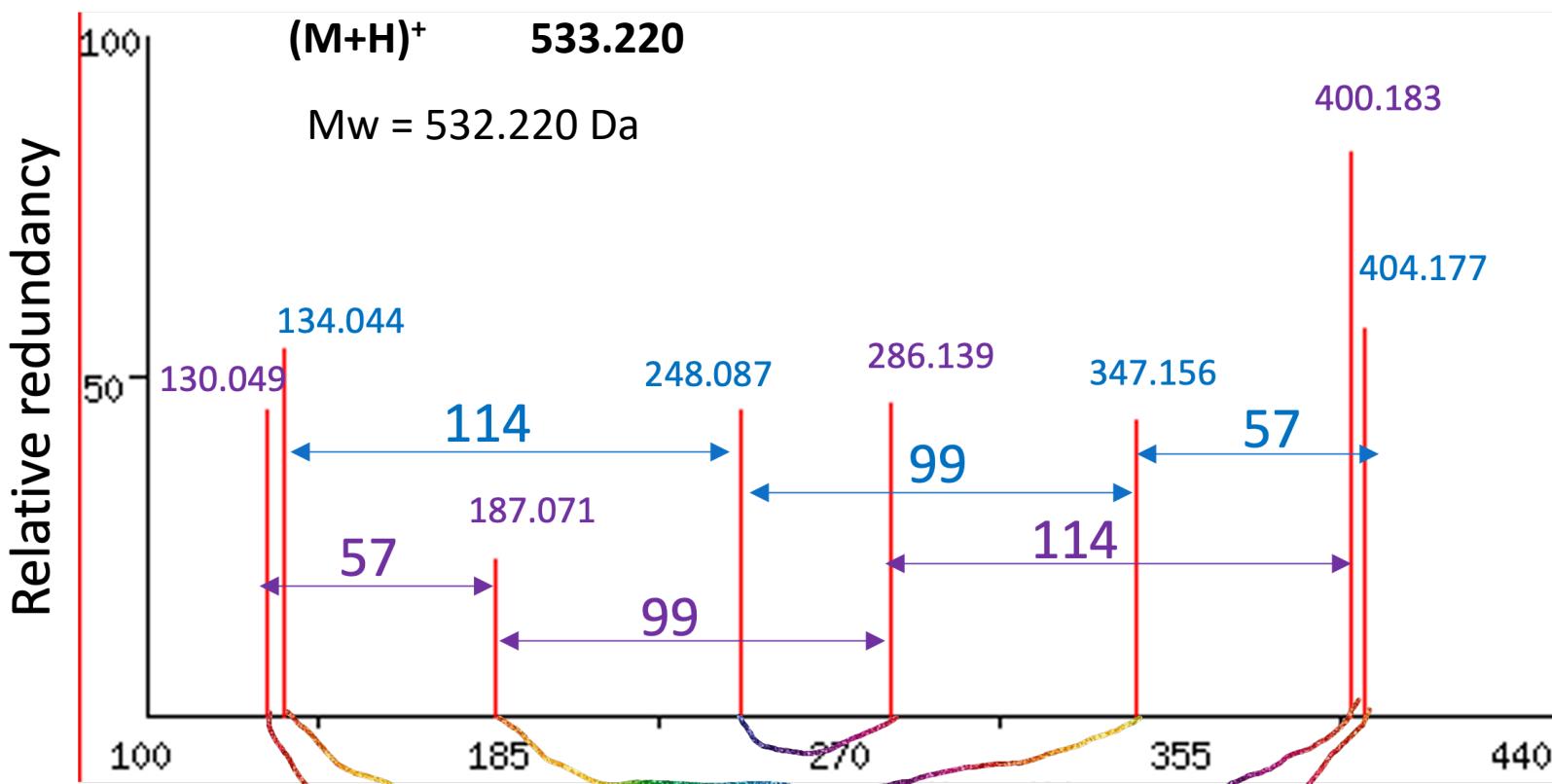


Mass of b-ions = Σ (residue masses) + 1 (H)



Mass of y-ions = Σ (residue masses) + 19 (OH + H + H⁺)

Complementary b/y Ion Pairs



Code (1 letter)	Monoisotopic mass
G	57.021 47
A	71.037 12
S	87.032 03
P	97.052 77
V	99.068 42
T	101.047 68
C	103.009 19
I	113.084 07
L	113.084 07
N	114.042 93
D	115.026 95
Q	128.058 58
K	128.094 97
E	129.042 60
M	131.040 49
H	137.058 91
F	147.068 42
R	156.101 12
Y	163.063 33
W	186.079 32

GVN or NVG

Calculate the Terminal Residues

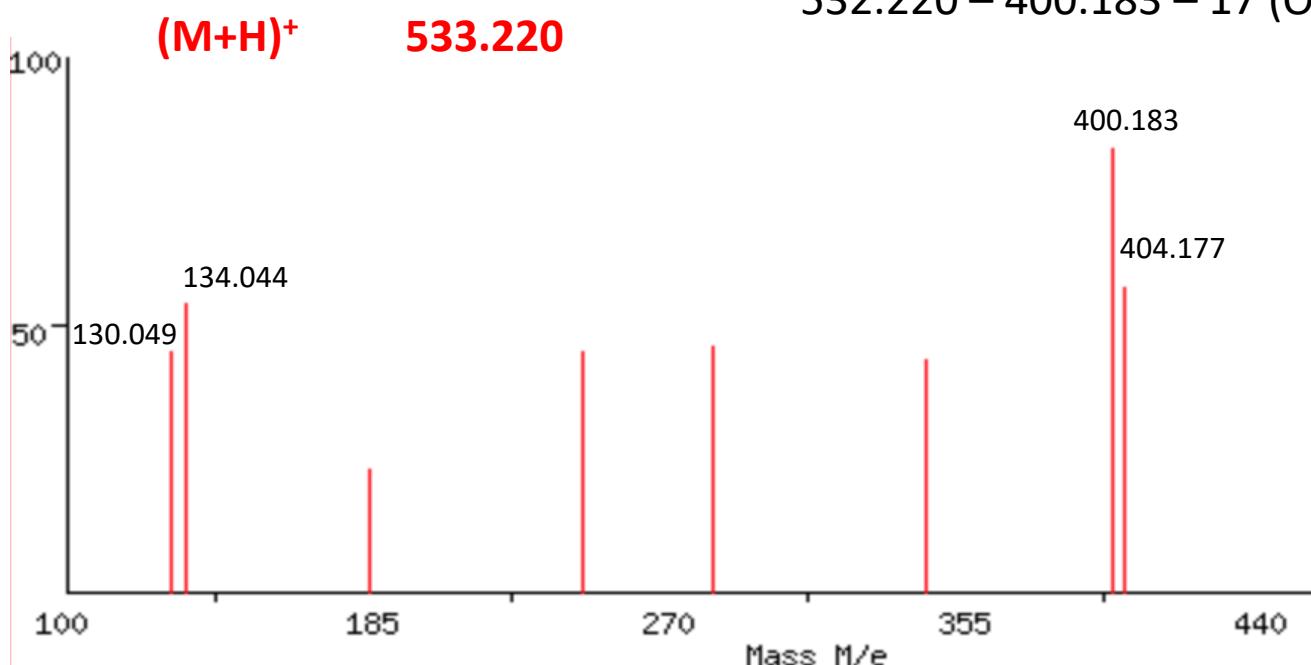
$b_1, y_1, b_{n-1}, y_{n-1}$

$$130.049 - 1 = 129.049$$

E on N terminus

$$134.044 - 19 = 115.044$$

D on C terminus



Code (1 letter)	Monoisotopic mass
G	57.02147
A	71.03712
S	87.03203
P	97.05277
V	99.06842
T	101.04768
C	103.00919
I	113.08407
L	113.08407
N	114.04293
D	115.02695
Q	128.05858
K	128.09497
E	129.04260
M	131.04049
H	137.05891
F	147.06842
R	156.10112
Y	163.06333
W	186.07932

$M - y_{n-1} \text{ ion} + 1 = \text{mass of 1}^{\text{st}} \text{ residue on N terminus}$

$M - b_{n-1} \text{ ion} - 17 = \text{mass of 1}^{\text{st}} \text{ residue on C terminus}$

Calculate the Terminal Residues

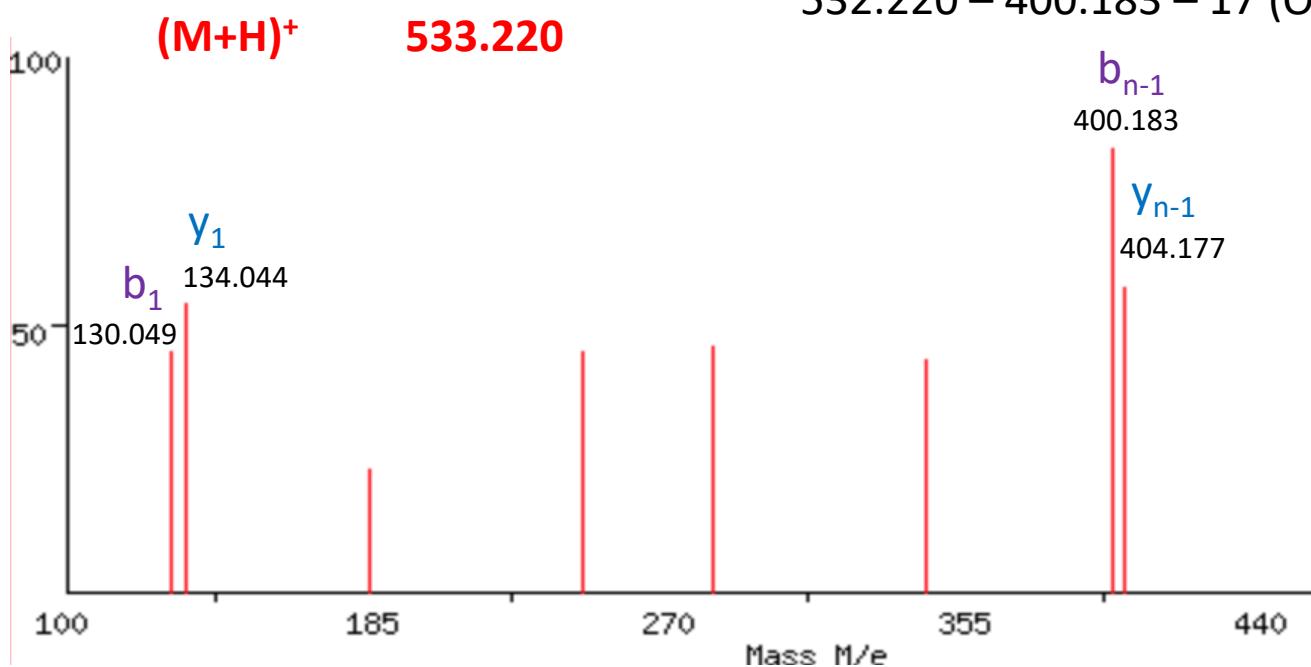
$b_1, y_1, b_{n-1}, y_{n-1}$

$$130.049 - 1 = 129.049$$

$$134.044 - 19 = 115.044$$

E on N terminus

D on C terminus



Code (1 letter)	Monoisotopic mass
G	57.02147
A	71.03712
S	87.03203
P	97.05277
V	99.06842
T	101.04768
C	103.00919
I	113.08407
L	113.08407
N	114.04293
D	115.02695
Q	128.05858
K	128.09497
E	129.04260
M	131.04049
H	137.05891
F	147.06842
R	156.10112
Y	163.06333
W	186.07932

$$532.220 - (404.177 - 2) - 1 = 129.043 \quad E \text{ on N terminus}$$

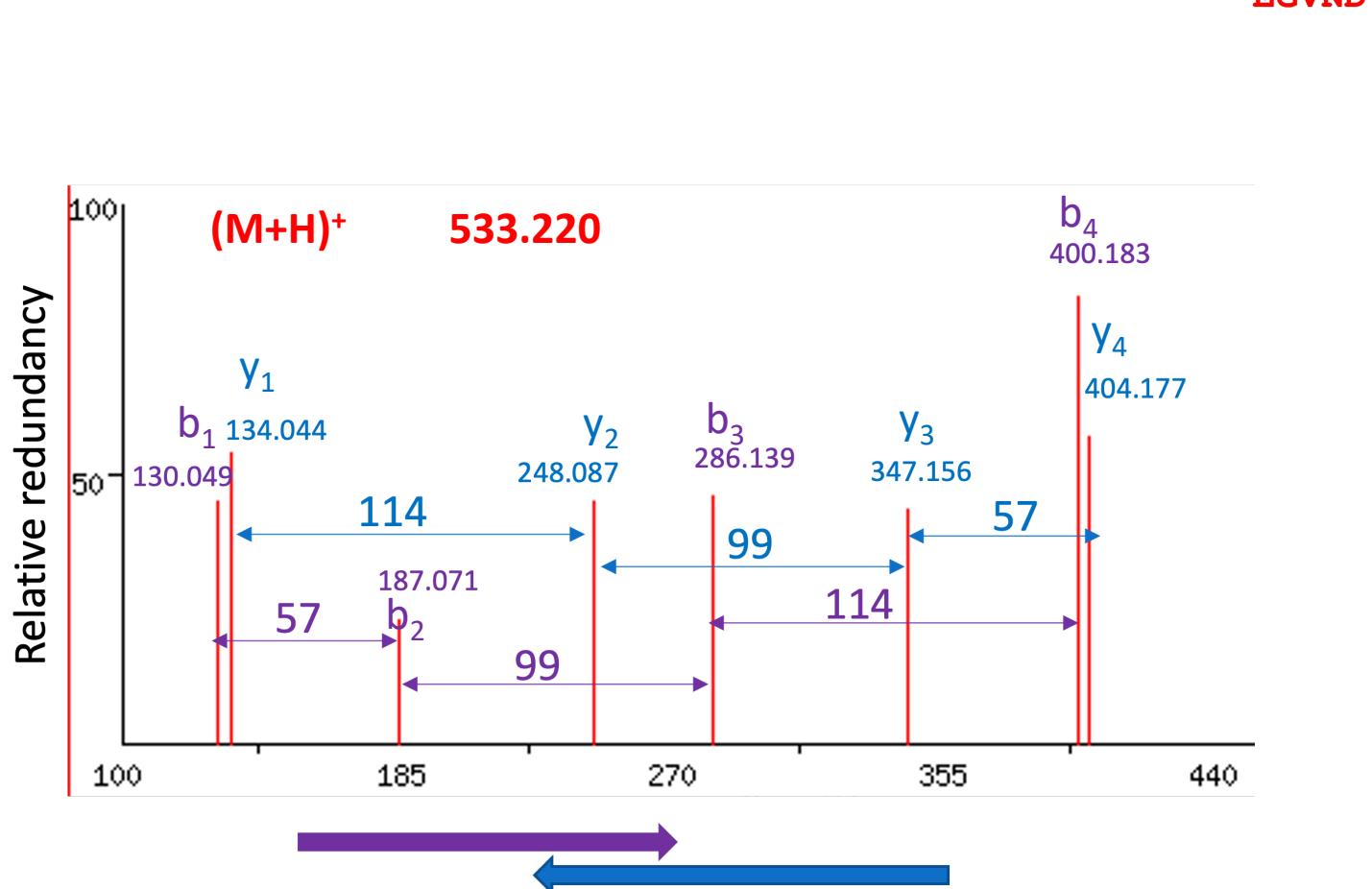
$$532.220 - 400.183 - 17 (\text{OH}) = 115.037 \quad D \text{ on C terminus}$$

$M - y_{n-1} \text{ ion} + 1 = \text{mass of 1}^{\text{st}} \text{ residue on N terminus}$

$M - b_{n-1} \text{ ion} - 17 = \text{mass of 1}^{\text{st}} \text{ residue on C terminus}$

Δ mass and Complementary b/y Ion Pairs

	<u>mass¹⁺</u>	<u>b-ions</u>	<u>y-ions</u>	<u>mass¹⁺</u>	Code (1 letter)	Monoisotopic mass
b_1^+	130.049	E	GVND	404.177	J	57.021 47
b_2^+	187.071	EG	VND	347.156	J	71.037 12
b_3^+	286.139	EGV	VND	248.087	J	87.032 03
b_4^+	400.182	EGVN	D	134.044	J	97.052 77
					V	99.068 42
					T	101.047 68
					C	103.009 19
					I	113.084 07
					L	113.084 07
					N	114.042 93
					D	115.026 95
					Q	128.058 58
					K	128.094 97
					E	129.042 60
					M	131.040 49
					H	137.058 91
					F	147.068 42
					R	156.101 12
					Y	163.063 33
					W	186.079 32



Summary of Peptide Mass Calculation

- Mass of b-ions = Σ (residue masses) + 1 (H^+)
- Mass of y-ions = Σ (residue masses) + 19 ($OH + H + H^+$)
- $M - y_{n-1}$ ion + 1 = mass of 1st residue on N terminus
- $M - b_{n-1}$ ion - 17 = mass of 1st residue on C terminus
- Mass of a-ions = mass of b-ions – 28 (CO)
- Ser-, Thr-, Asp- and Glu-containing ions generate neutral molecular loss of water (-18).
- Asn-, Gln-, Lys-, Arg-containing ions generate neutral molecular loss of ammonia (-17).
- A complementary b-y ion pair can be observed in multiply charged ions spectra.
 - For this b-y ion pair, the sum of their subscripts is equal to the total number of amino acid residues in the unknown peptide.

Code (1 letter)	Monoisotopic mass
G	57.021 47
A	71.037 12
S	87.032 03
P	97.052 77
V	99.068 42
T	101.047 68
C	103.009 19
I	113.084 07
L	113.084 07
N	114.042 93
D	115.026 95
Q	128.058 58
K	128.094 97
E	129.042 60
M	131.040 49
H	137.058 91
F	147.068 42
R	156.101 12
Y	163.063 33
W	186.079 32

$[M+H]^+ = 1464.7693$

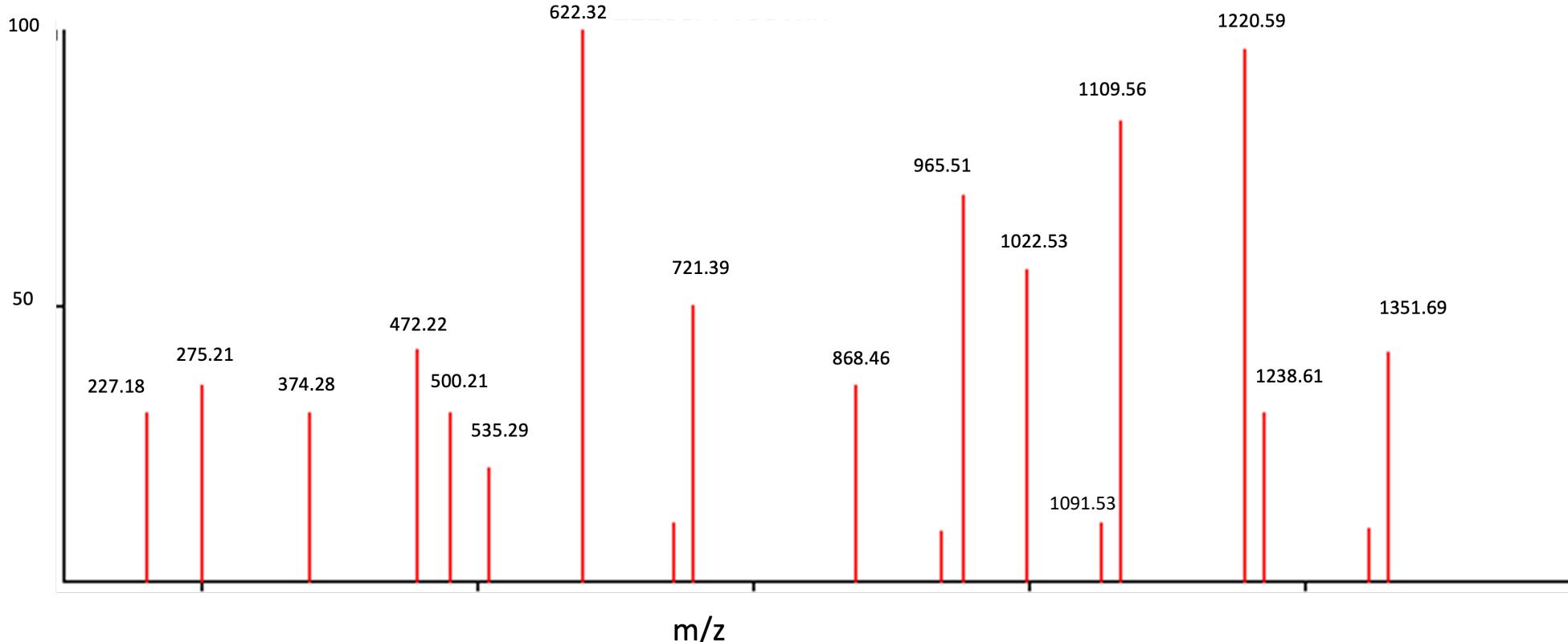
So, Mw = 1463.7693 Da

- First look at the dominant peak that below the mass.
- $M - y_{n-1}$ ion + 1 = mass of 1st residue on N terminus
- $M - b_{n-1}$ ion - 17 = mass of 1st residue on C terminus

1) $1463.7693 - 1351.69 + 1 = 113.0793$, which is the mass of I/L. SO 1351.69

I/L-

m/z represents an y_{n-1} ion and I/L is the N terminus residue.



Code (1 letter)	Monoisotopic mass
G	57.021 47
A	71.037 12
S	87.032 03
P	97.052 77
V	99.068 42
T	101.047 68
C	103.009 19
I	113.084 07
L	113.084 07
N	114.042 93
D	115.026 95
Q	128.058 58
K	128.094 97
E	129.042 60
M	131.040 49
H	137.058 91
F	147.068 42
R	156.101 12
Y	163.063 33
W	186.079 32

C_{CM} : Cysteine with Carboxymethyl (58.01)

$$[M+H]^+ = 1464.7693$$

$$\text{So, Mw} = 1463.7693 \text{ Da}$$

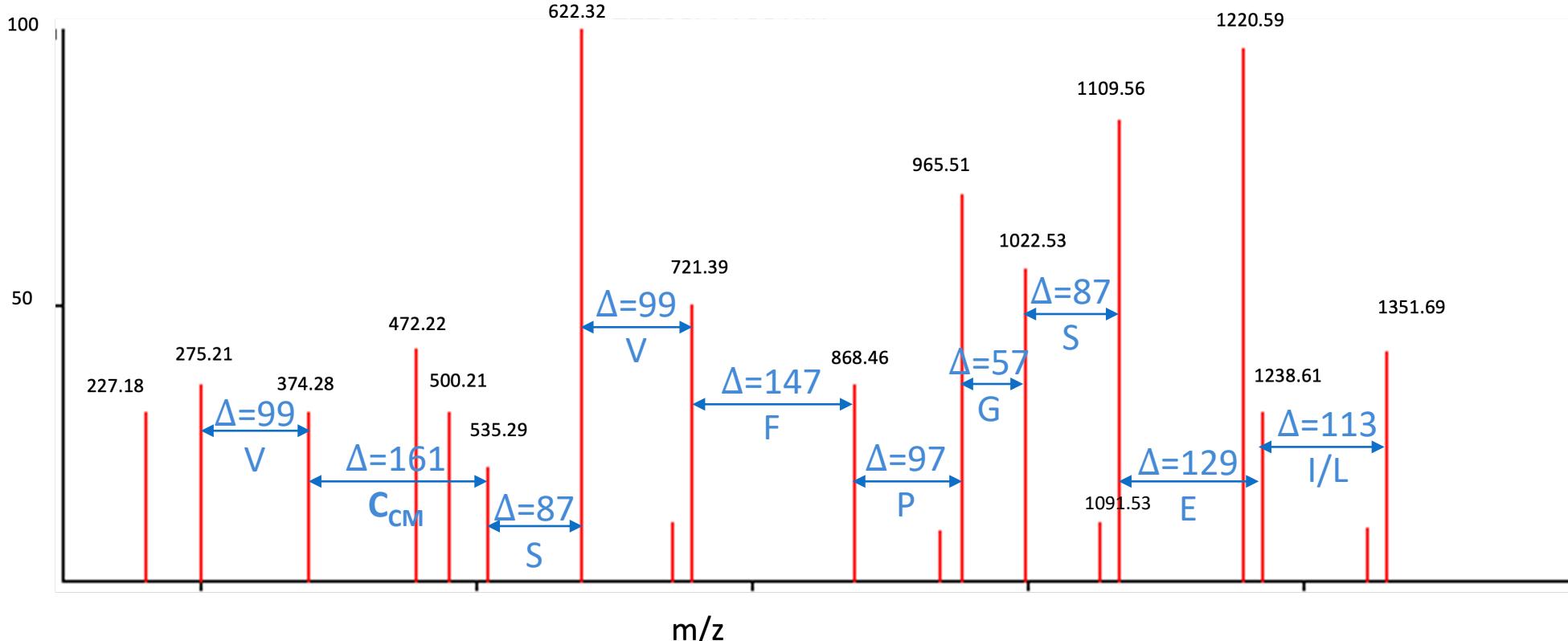
- Amino acid sequence can be deduced by the Δ mass between adjacent y ion peaks or adjacent b ion peaks

2) $\Delta m/z = 1351.69 - 1238.61 = 113.08$, which is the mass of I/L.

I/L-I/L

3) See below.....

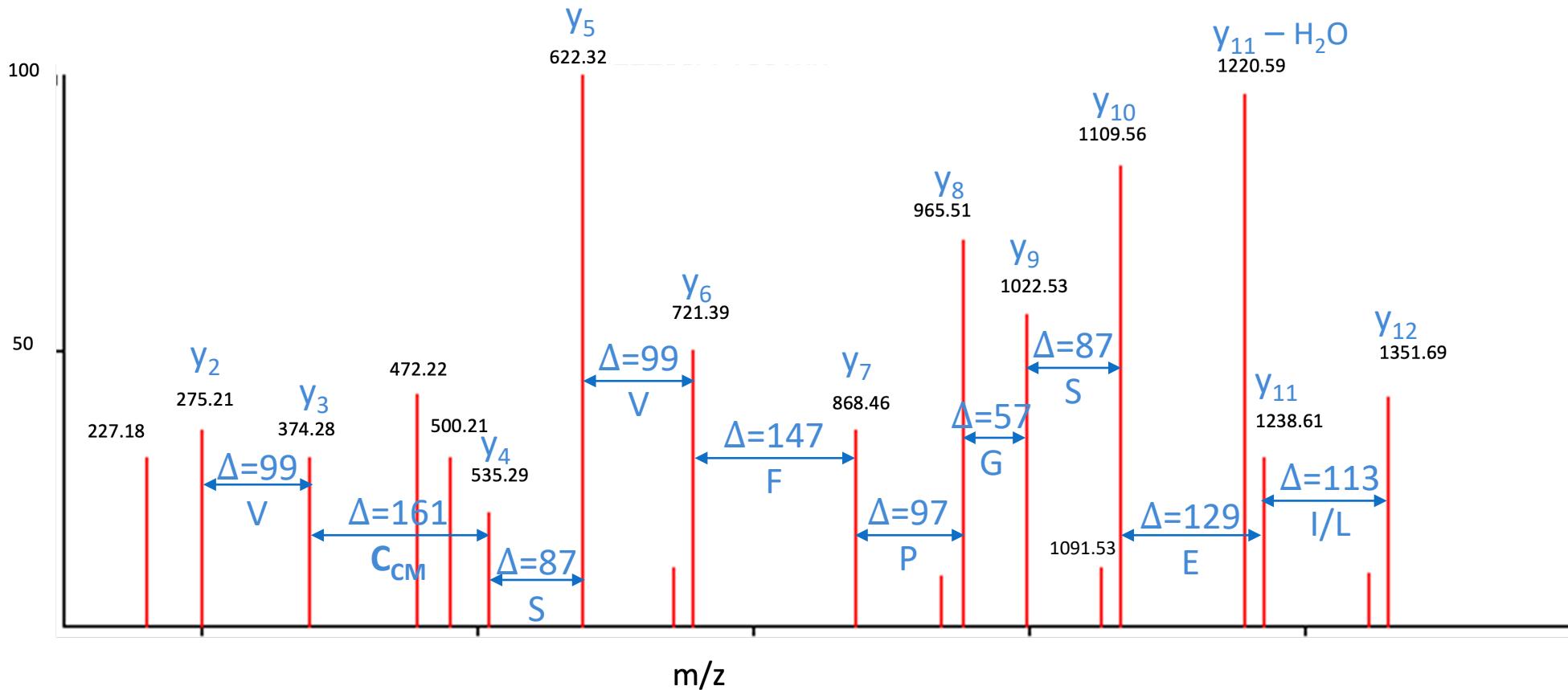
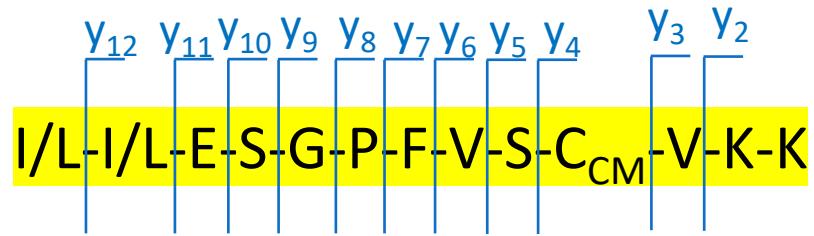
I/L-I/L-E-S-G-P-F-V-S-C_{CM}-V-...



Code (1 letter)	Monoisotopic mass
G	57.021 47
A	71.037 12
S	87.032 03
P	97.052 77
V	99.068 42
T	101.047 68
C	103.009 19
I	113.084 07
L	113.084 07
N	114.042 93
D	115.026 95
Q	128.058 58
K	128.094 97
E	129.042 60
M	131.040 49
H	137.058 91
F	147.068 42
R	156.101 12
Y	163.063 33
W	186.079 32

C_{CM}: Cysteine with Carboxymethyl (58.01)

- 4) 275.21 m/z is probably the y2 ion with 2 residues. Because it is an y ion, so the mass of two residues = $y_2 - 19 = 256.21$, which are the sum of K and K.



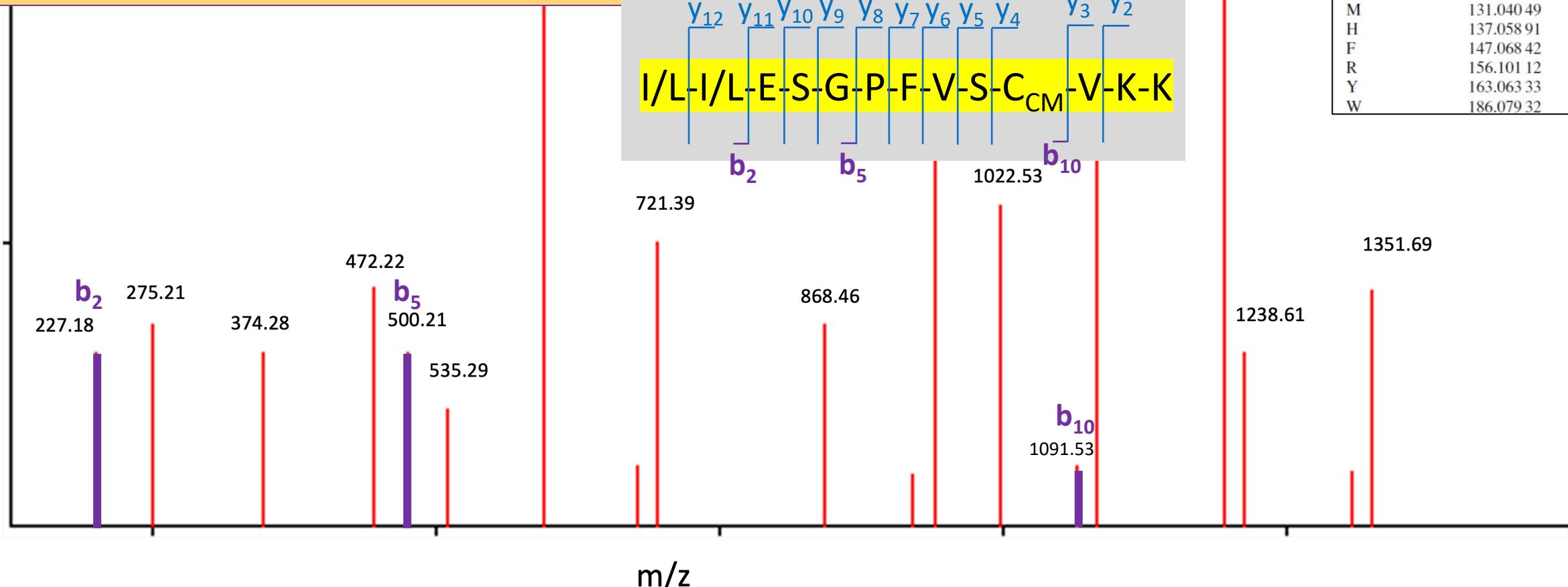
C_{CM} : Cysteine with Carboxymethyl (58.01)

- Then to verify the high mass y ion assignments, we look for the complimentary low mass b ions.
- We may not be able to see b1. Usually, we will start by looking for b2.

1) $227.18 > 186.07932$ (W), SO the first ion on the left is a b2 ion. So the first two residues are I/L-I/L.

$$500.21 - 227.18 = 273.03 = E+S+G$$

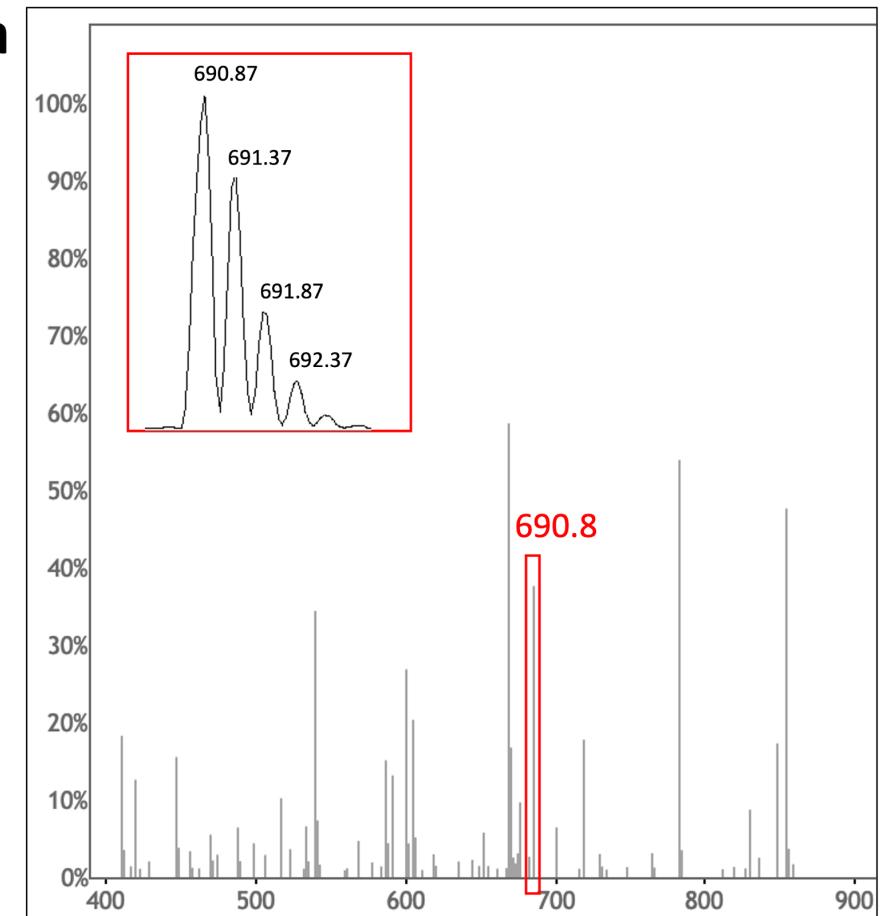
$$1091.53 - 500.21 = 591.32 = P+F+V+S+C_{CM}$$



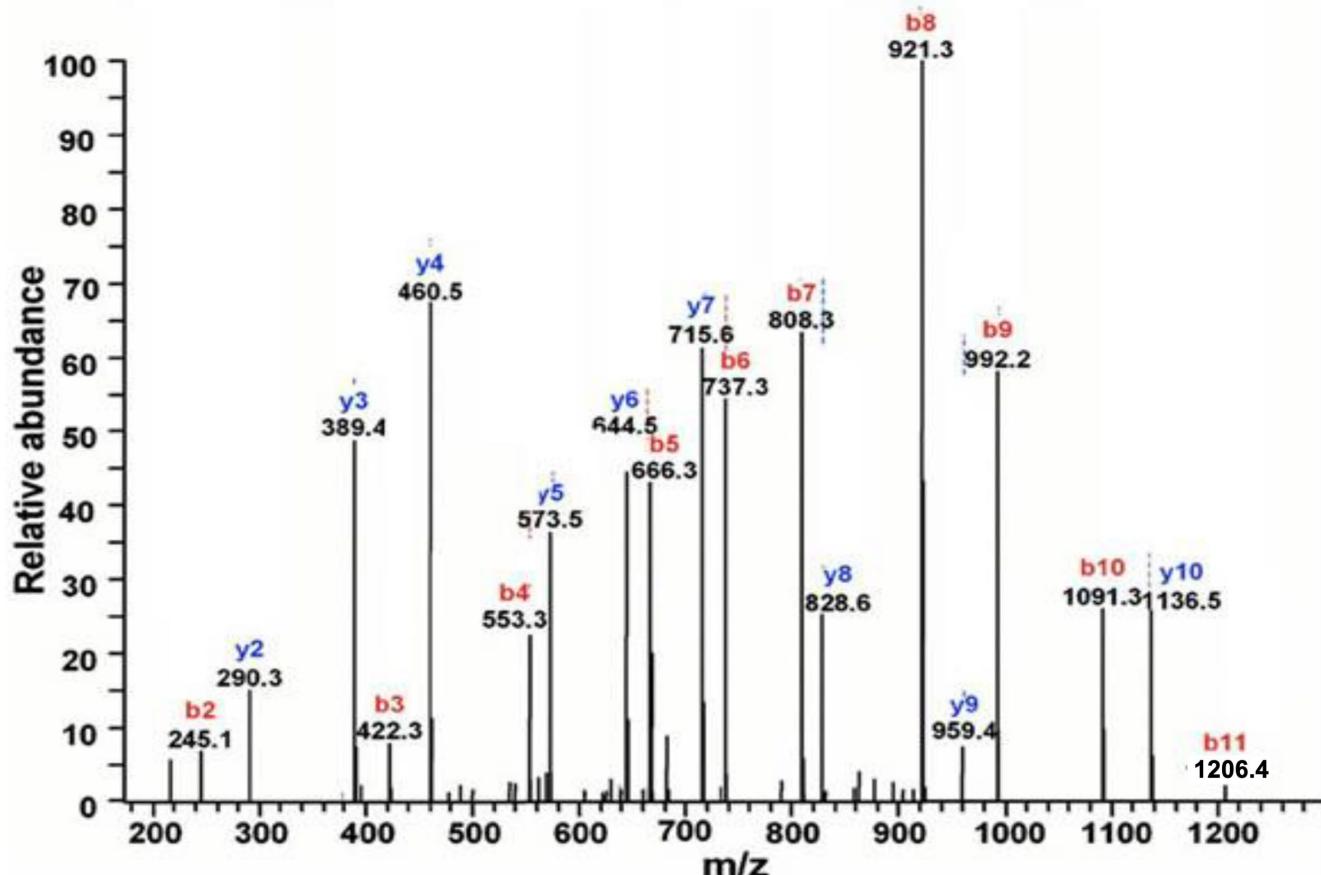
Code (1 letter)	Monoisotopic mass
G	57.02147
A	71.03712
S	87.03203
P	97.05277
V	99.06842
T	101.04768
C	103.00919
I	113.08407
L	113.08407
N	114.04293
D	115.02695
Q	128.05858
K	128.09497
E	129.04260
M	131.04049
H	137.05891
F	147.06842
R	156.10112
Y	163.06333
W	186.07932

Practice 1

- Figure **a** is the full scan MS1 spectrum of a mixture **a** of tryptic peptides. The inset is the isotopic peak of a modified 12-residue peptide. The modification is either methylation or acetylation. To identify the amino acid sequence and the type of modification, the precursor ions of the modified 12-residue peptide are selected for CID fragmentation and analyzed in the second MS analyzer. The tandem mass spectrum is shown in Figure **b**. Using the spectra and the information provided, answer the following questions:
 - What is the mass of the modified peptide? What is the charge of this peptide?

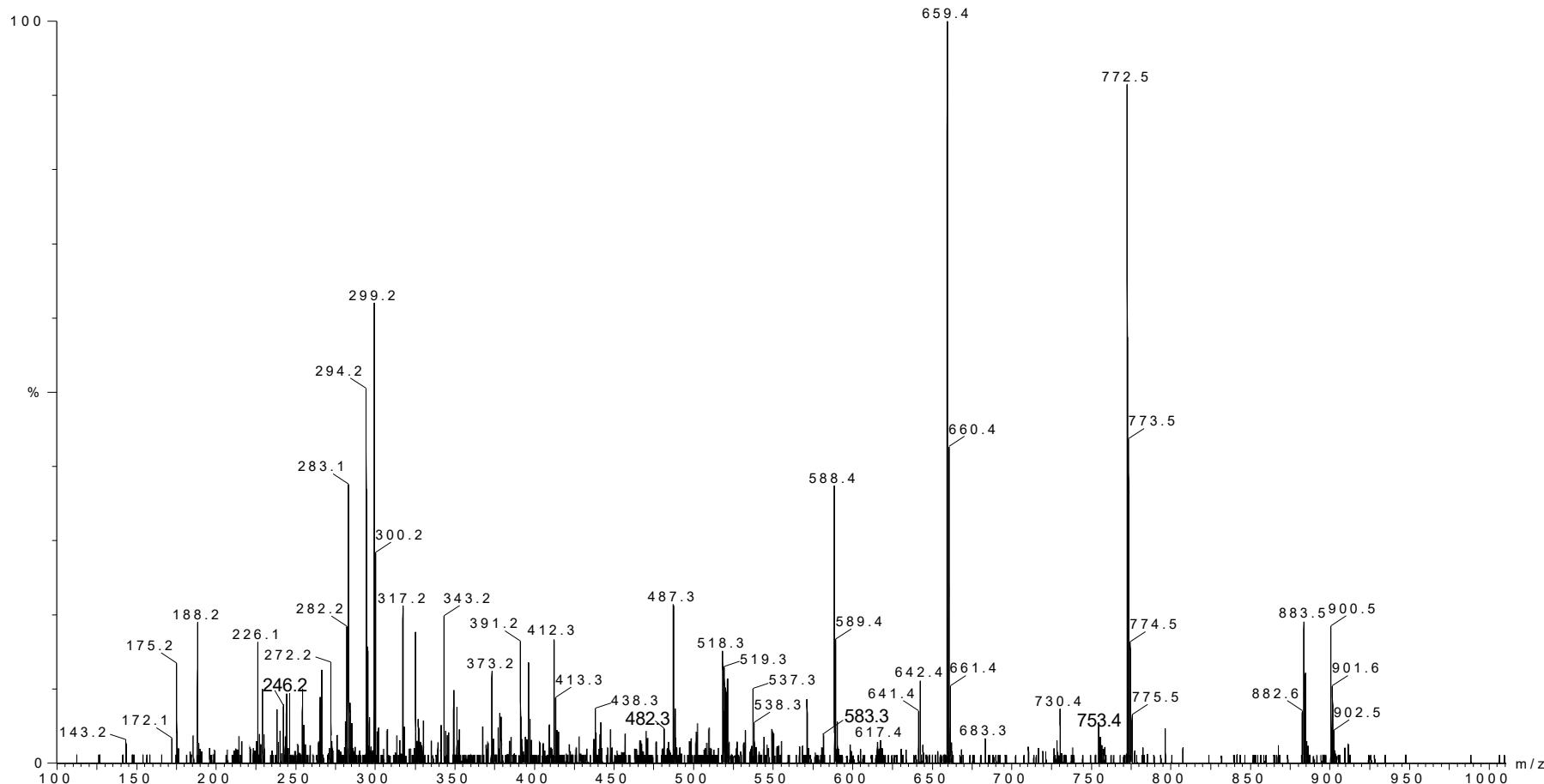


- ii) Determine the type and location of the modification.
- iii) Determine the full sequence of the modified sequence. Can you assign the sequence with this data alone? If not, what technique can be used to help you identify the full sequence?



Code (1 letter)	Monoisotopic mass
G	57.021 47
A	71.037 12
S	87.032 03
P	97.052 77
V	99.068 42
T	101.047 68
C	103.009 19
I	113.084 07
L	113.084 07
N	114.042 93
D	115.026 95
Q	128.058 58
K	128.094 97
E	129.042 60
M	131.040 49
H	137.058 91
F	147.068 42
R	156.101 12
Y	163.063 33
W	186.079 32

Are there easy ways to differentiate between fragment ions originating from the N terminus and C terminus of a peptide?

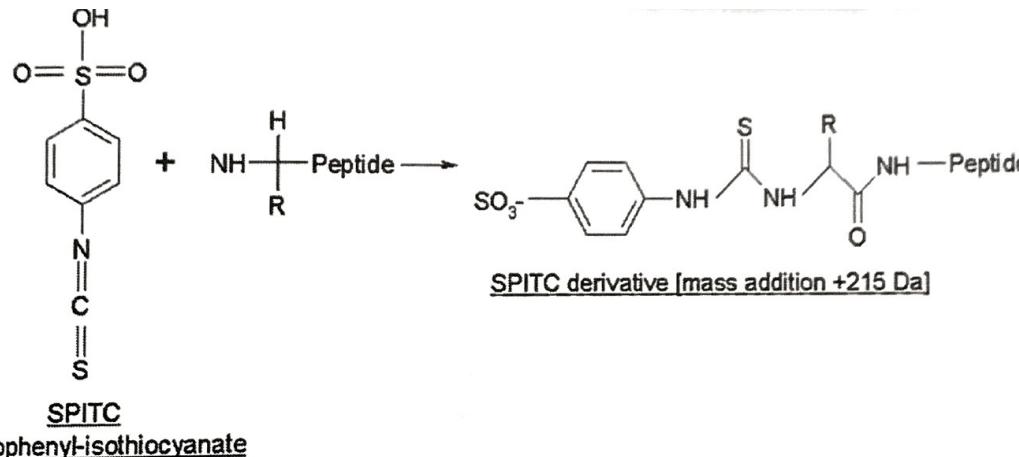


Chemical Derivatization For Sequence Analysis

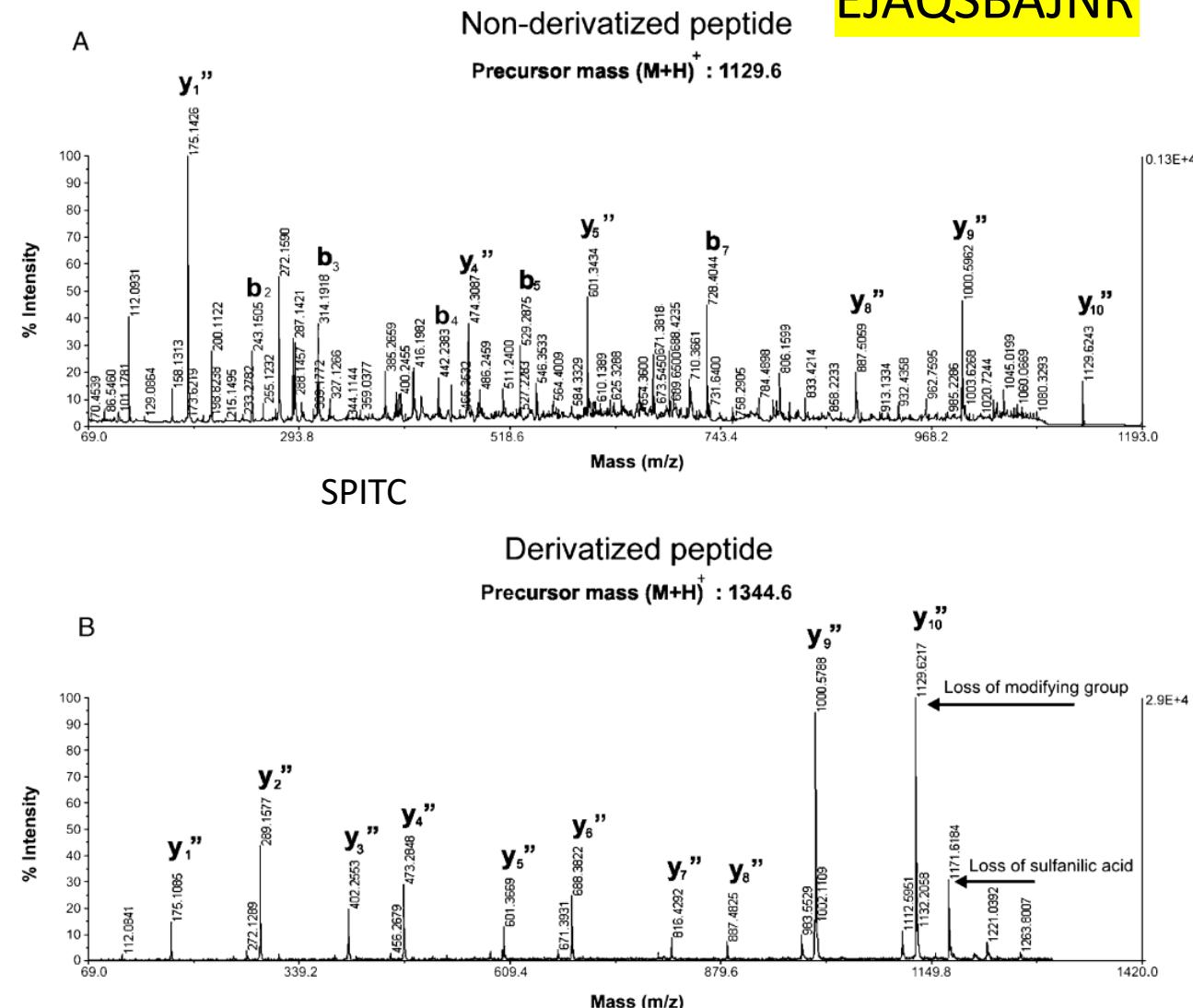
- *Derivatization* is a procedure that utilizes chemical reactions with covalent bond formation
- The aims of derivatization are:
 - To enhance detection of one ion series
 - To improve fragmentation yield
 - To simplify data interpretation (for de novo sequencing)

1. SPITC (4-sulfophenylisothiocyanate)

- SPITC can generate sulfonation of peptides at the amino terminus



- SPITC selectively enhances detection of the y ion series

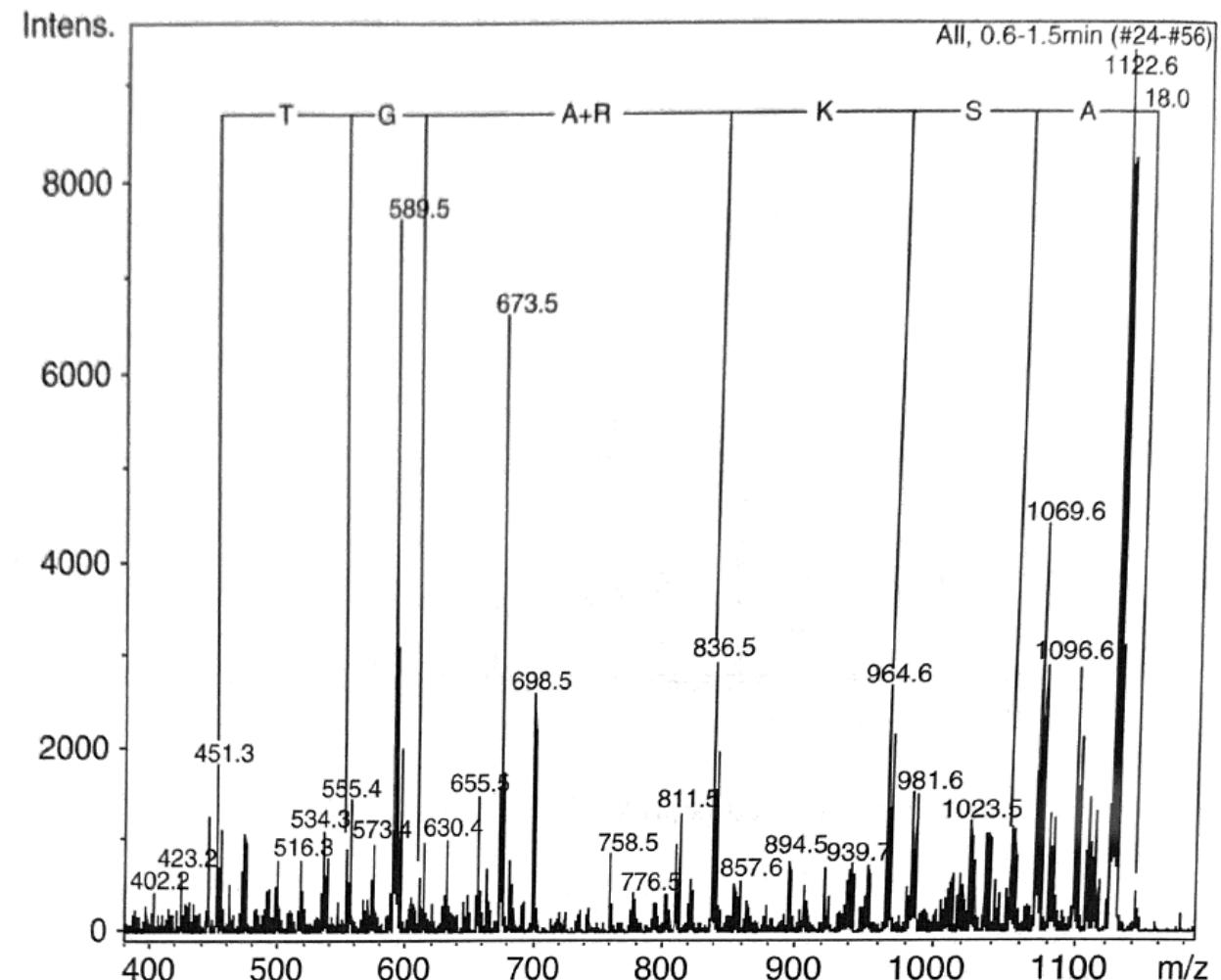


2. Acetylation and DeuteroAcetylation

FGGFTGARKSA

[M+H]⁺ = 1140.6

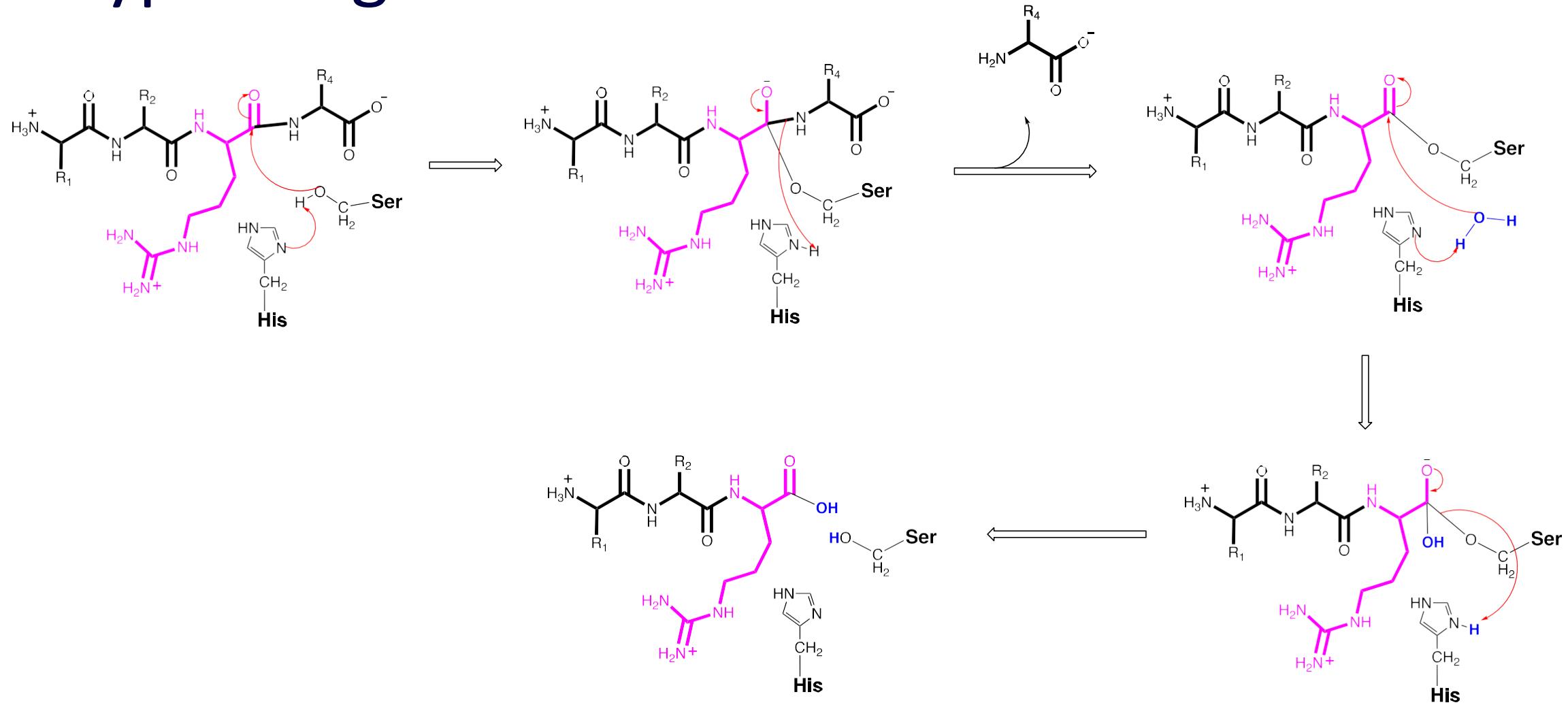
- It utilizes a mixture of acetic anhydride and deuterated acetic anhydride (1:1 v/v) in methanoic solution, which labels N terminal amino groups only.



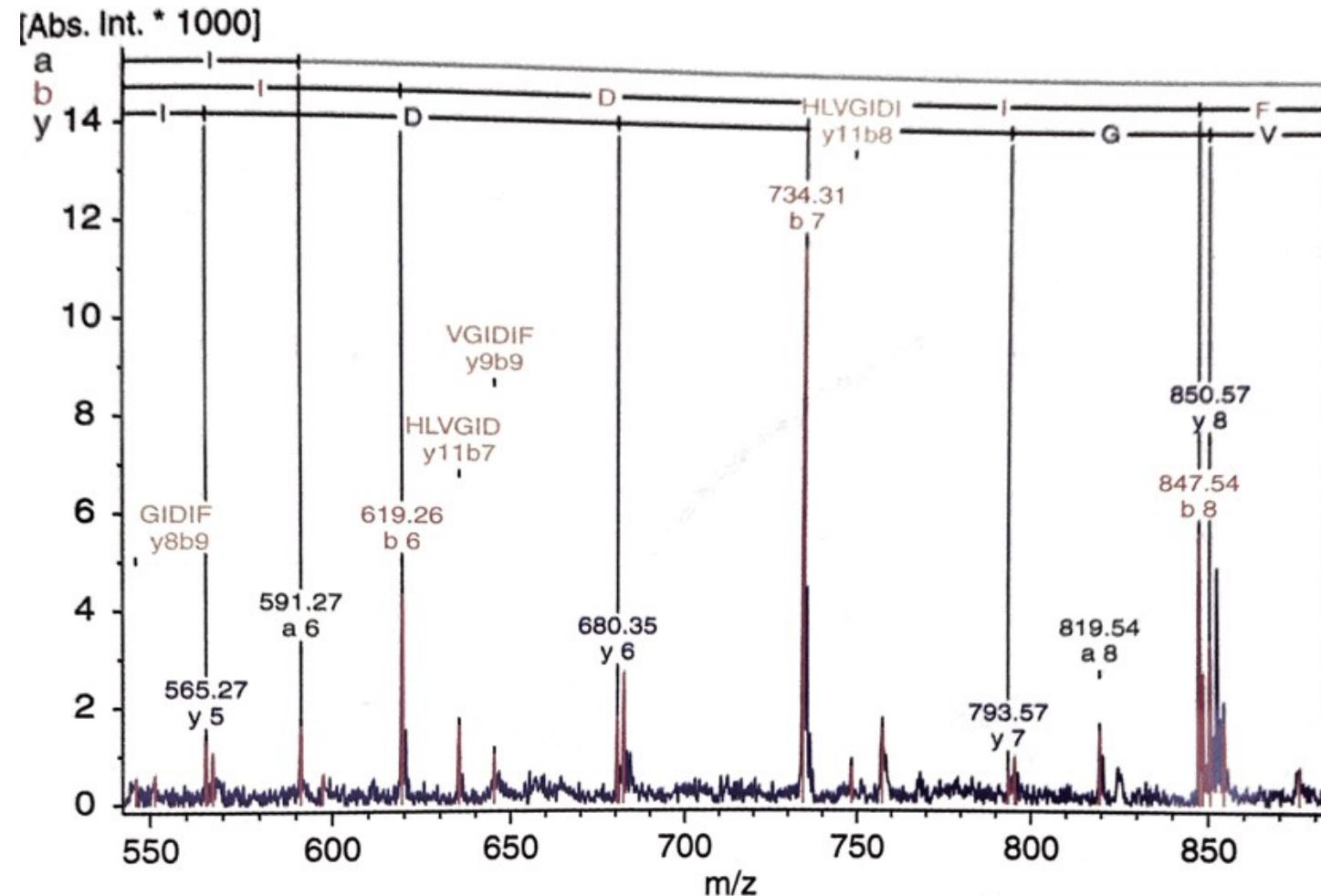
3. ^{18}O Labeling

- When the peptide was digested by some protease (e.g., trypsin), one H_2O was added into the newly formed peptide termini. This water is taken from the solvent.
- When the solvent consists of a mixture of water containing ^{16}O and ^{18}O . This results in an incorporation of both ^{16}O and ^{18}O into the peptide.
- Thus, every peptide appears as a pair of peaks separated by 2 Th in precursor mass spectrum.
- After fragmentation, the MS/MS spectrum will show both single peaks (usually b ions) and single **peak pairs (usually y ions)**.

Trypsin Digestion mechanism

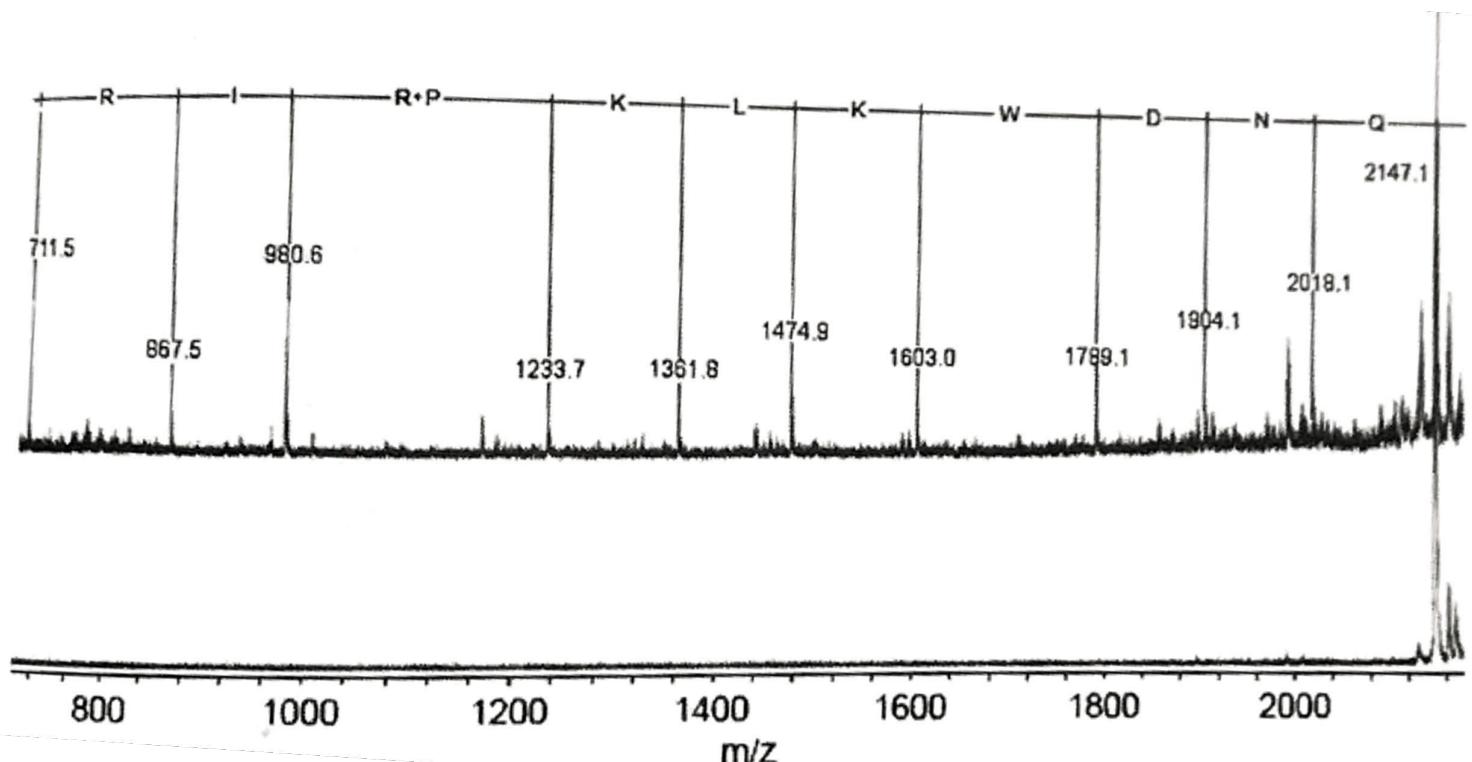
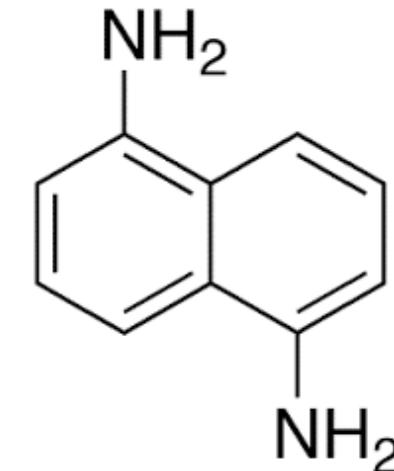


¹⁸O Labeling

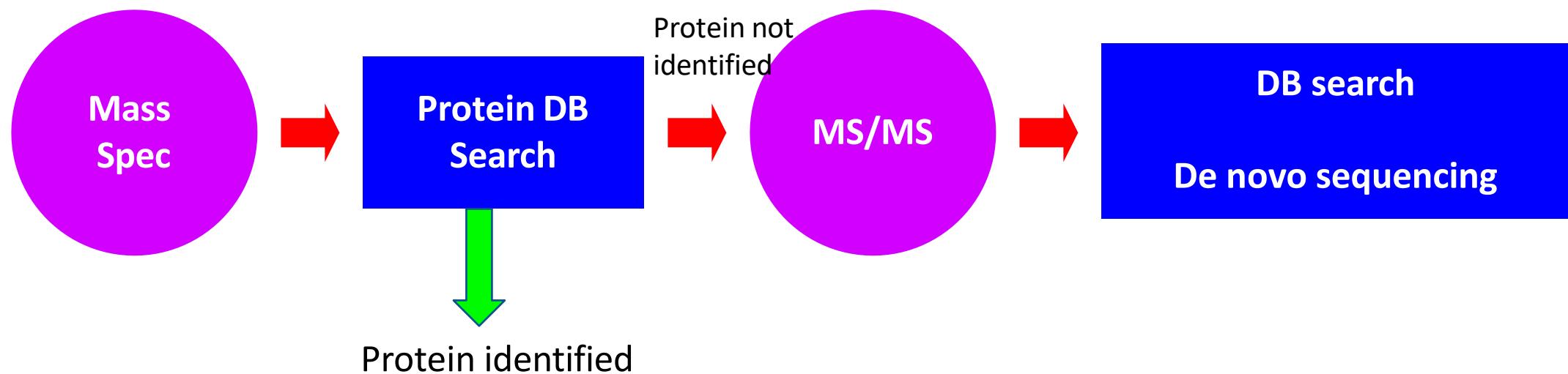


4. DAN (1,5-diaminonaphthalene)

- Serves as both a matrix in MALDI method and a reducing reagent for S-S bond.
- It is suitable for top-down proteomics
- c ion series is usually most abundant in the spectrum

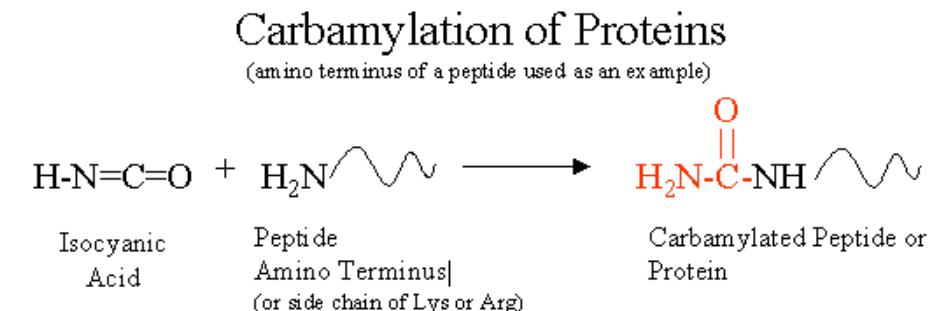
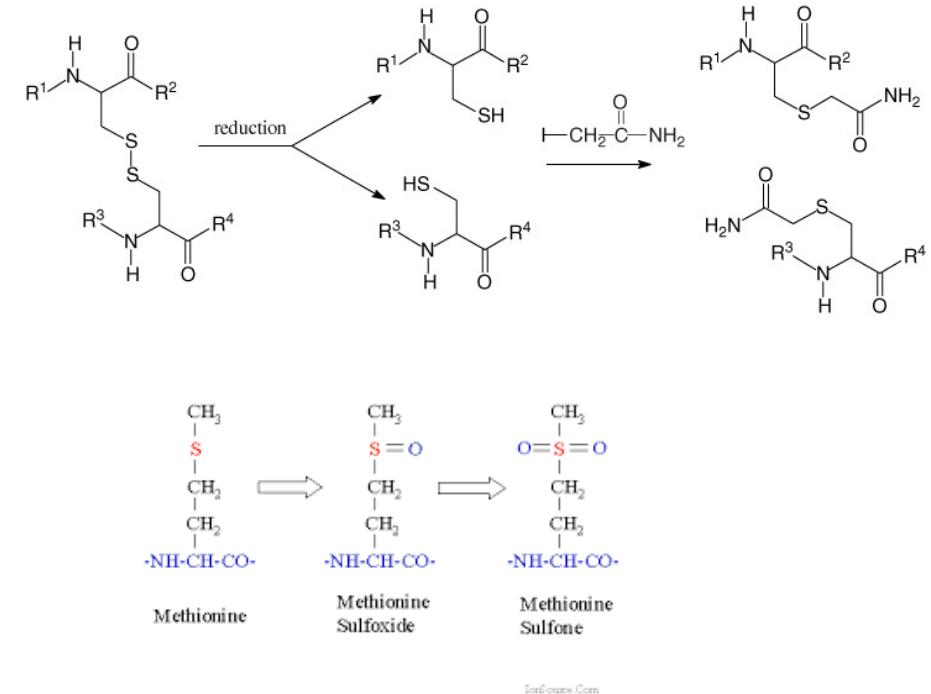


Protein Identification and Characterization Map



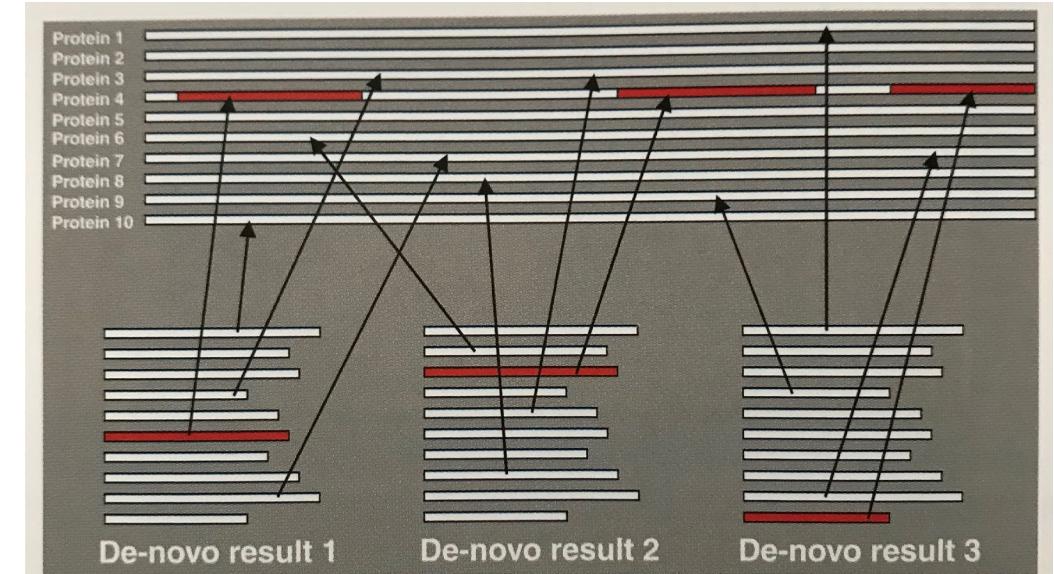
Specific Amino Acids Modification During Sample Handling

- Reduction and Alkylation on Cys
 - Routinely done prior to enzymatic digestion to break disulfide bonds, unfolding proteins to make them more susceptible to enzymatic cleavage
- Methionine is easily mono-oxidized (Met sulfoxide)
- Cyclization of N-terminal Glutamine (Q) and carboxamidomethyl-Cys
- Urea exposure can carbamylate N termini of protein/peptide and side chains of Lys
- etc.



Physiochemical Complications to Spectrum Interpretation

- Incomplete fragmentation
- Inconsistent intensity of fragment ion types
- Chemical or posttranslational modifications
- Isobaric AAs
 - I = L
 - K = Q
- Isobaric AA combinations
 - GG = N
 - GA = K = Q
 - W = DA = VS



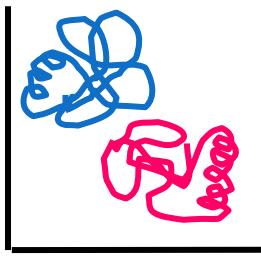
Schematic view of the function of MS-BLAST

LECTURE 3: BOTTOM-UP PROTEOMICS AND TOP-DOWN PROTEOMICS

Protein Analysis by Mass Spectrometry

- **Bottom up:** sequence fragments of larger proteins
 - Proteins are cut into smaller pieces with enzymes (proteases).
 - Ion traps, triple quadrupoles, and hybrid instruments are ideal
- **Top Down:** extract and identify intact proteins in images
 - Intact protein mass measurement
 - Use high resolution MS (FTICR)

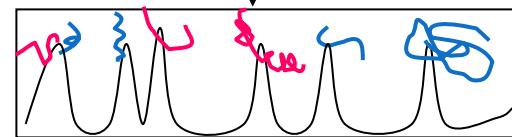
Bottom-up Proteomics



Reduction/
Alkylation
→
Digest - enzyme
(e.g. trypsin)



→
Multi-dimensional
HPLC



Separation of peptides

↓
Determine masses of intact peptides
Perform MS/MS fragmentation
on each peptide as it elutes

↓
Data Analysis

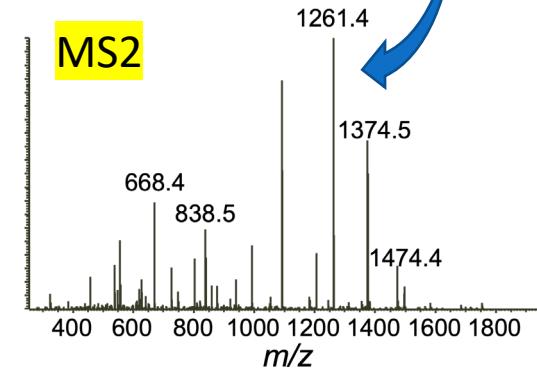
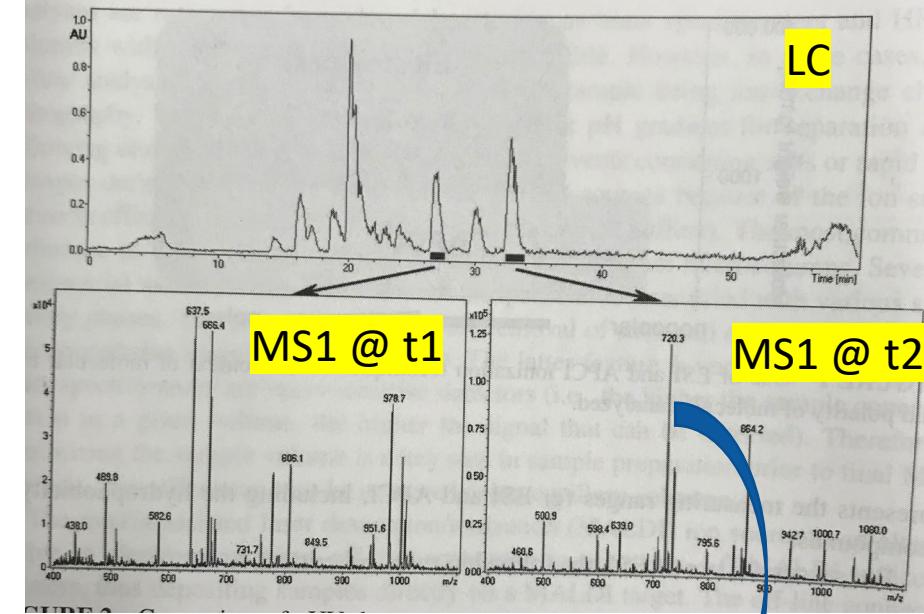
>ipi|IPI0024789|IPI0024789.6
MDQHQHLNK**TAESASSEKEKKTRR**
CNGFK**AMFLAALSFSYIAK**ALGGIIM
KISITQIERRFD...

Protein sequences
↑

TAESASSEKE
AMFLAALSFSYIAK

...

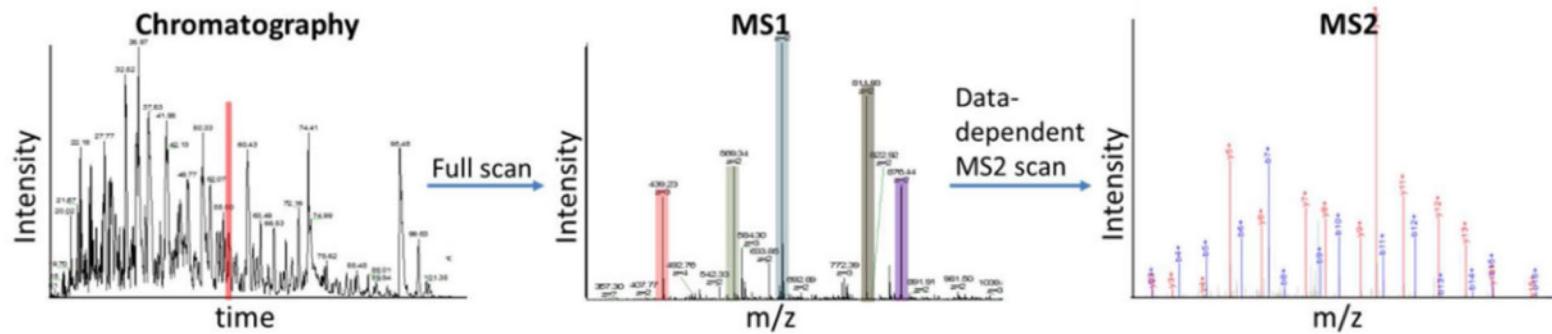
Peptide sequences



Shot Gun Proteomics

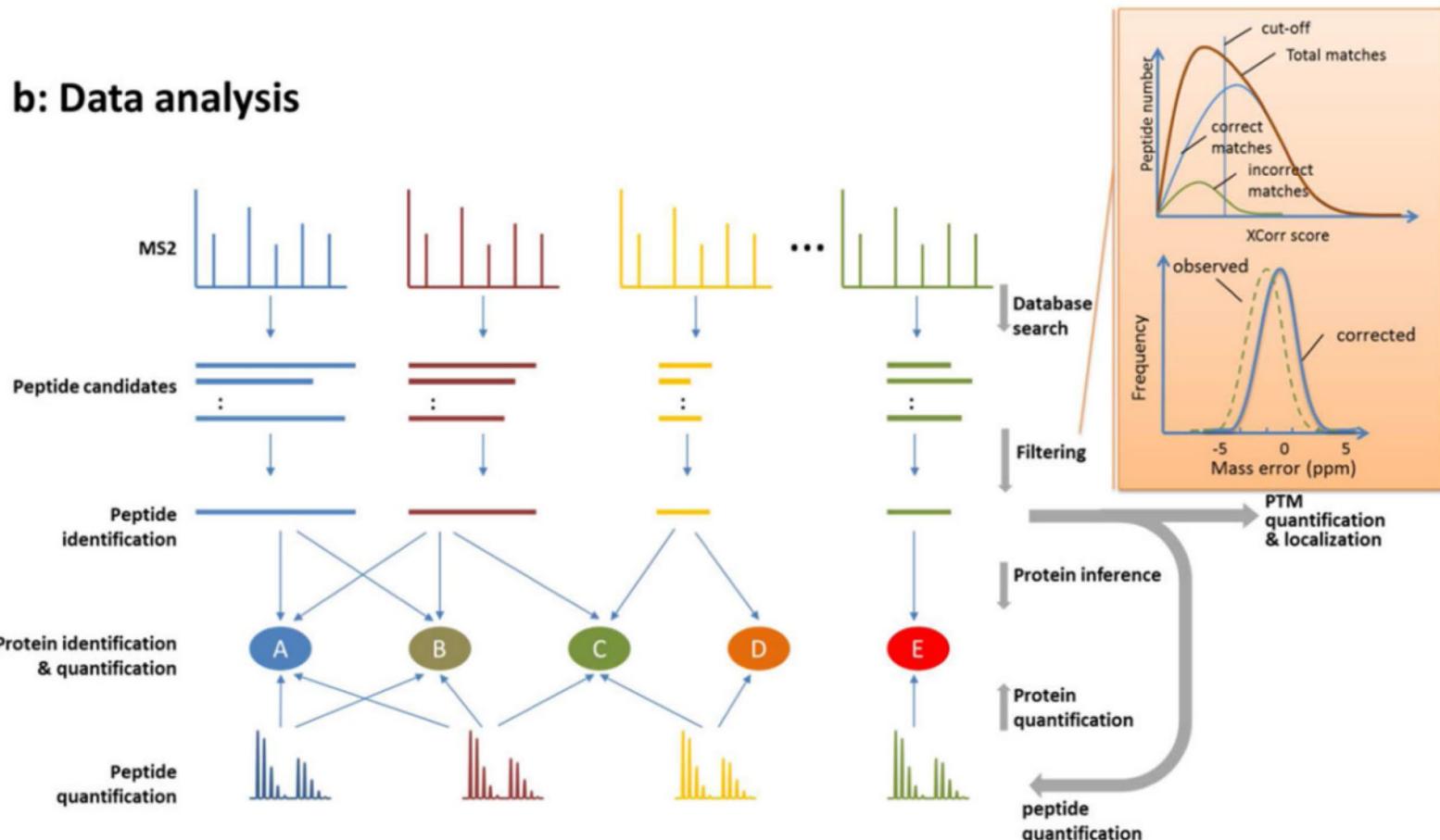
- “Bottom-up” protein analysis refers to the characterization of proteins by analysis of peptides released from the protein through proteolysis.
- When bottom-up is performed on a mixture of proteins it is called **shotgun proteomics**.
 - In a typical shotgun proteomics experiment, the peptide mixture is fractionated and subjected to LC-MS/MS analysis.
 - Peptide identification is achieved by comparing the tandem mass spectra derived from peptide fragmentation with theoretical tandem mass spectra generated from *in silico* digestion of a protein database.
 - Protein inference is accomplished by assigning peptide sequences to proteins.

a: Data acquisition



Representative LC-MS/MS data and a generalized bioinformatic analysis pipeline for protein identification and quantification in shotgun proteomics.

b: Data analysis



Bottom-up Proteomics

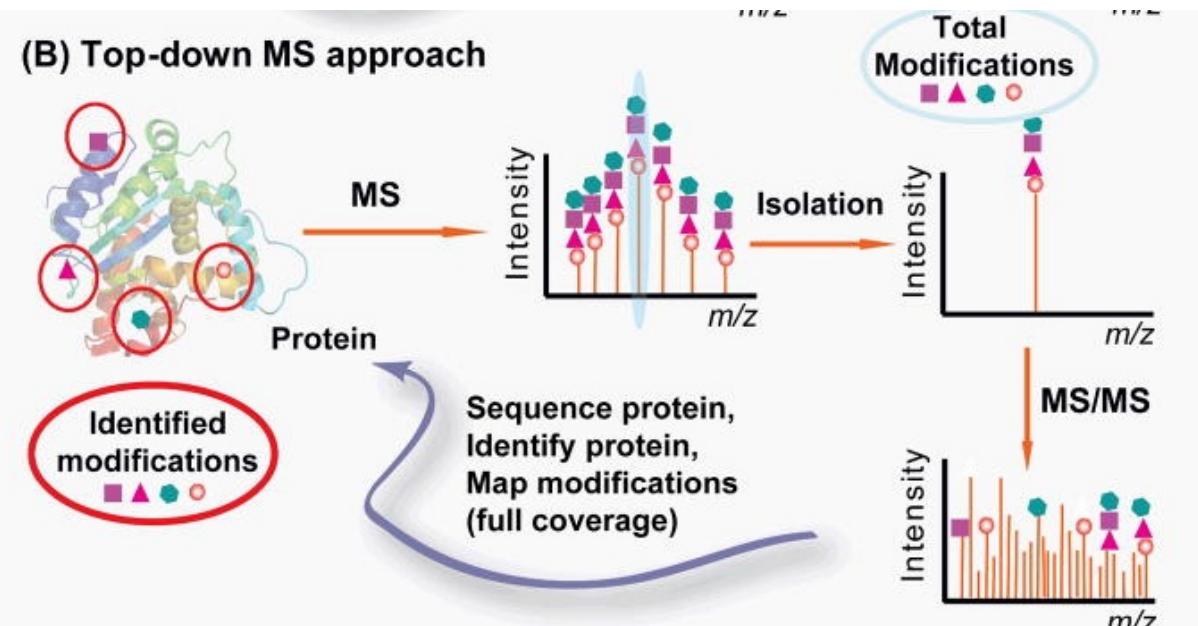
Advantages

- The most mature and most widely used approach for protein identification and characterization
- Easier to analyze larger or hydrophobic proteins
- peptides have more uniform physico-chemical properties, i.e., solubility, hydrophobicity.
- Less sophisticated instrumentation and expertise

Disadvantages

- A low percentage coverage of the protein sequence
- A significant amount of information about PTMs and alternative splice variants is lost

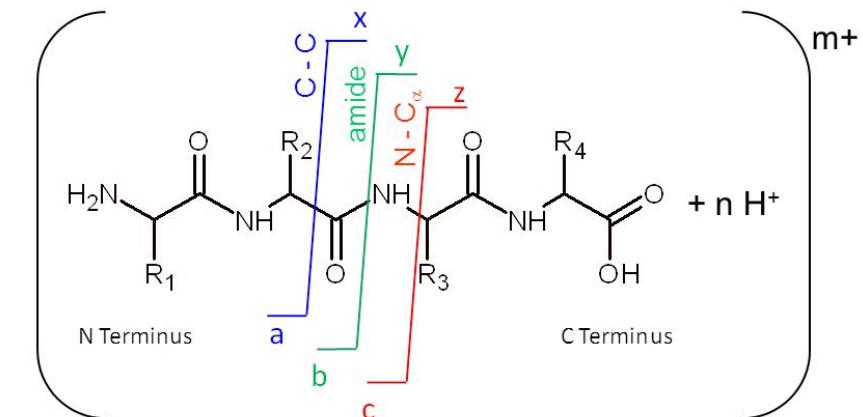
Top-down Proteomics



- Characterize **intact proteins** from complex biological systems
- Most useful for single proteins or relatively simple mixture
- Proteins are typically ionized by ESI and trapped in FT-ICR or orbit trap mass spectrometer
- Fragmentation for tandem MS is accomplished by electron-capture dissociation (ECD) or electron-transfer dissociation (ETD)
 - ECD and ETD typically provide more uniform dissociation than conventional CID (collision induced dissociation), while preserving the labile modifications

Electron Capture Dissociation (ECD) Electron Transfer Dissociation (ETD)

- Different mechanisms for fragmentation than CAD
- Free radical cleavage chemistries
- Ions can be fragmented more efficiently.
- Favors high charge states (+3 and higher)
- Efficient sequencing of peptides with PTMs



There are three different bond types in the peptide backbone.

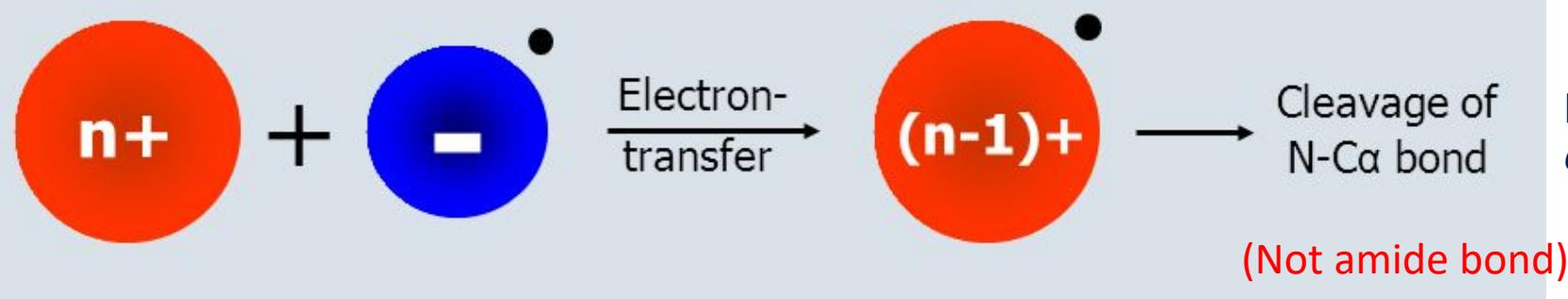
- **CID** produces **b** and **y** type anions by heterolytic amide bond cleavage.
- **ETD** and **ECD** produce **c** and **z** type ions by homolytic bond cleavage.

ETD Reaction Scheme

Multiply charged analyte ($n \geq 2$)

Reagent radical anion

odd-electron protonated peptide



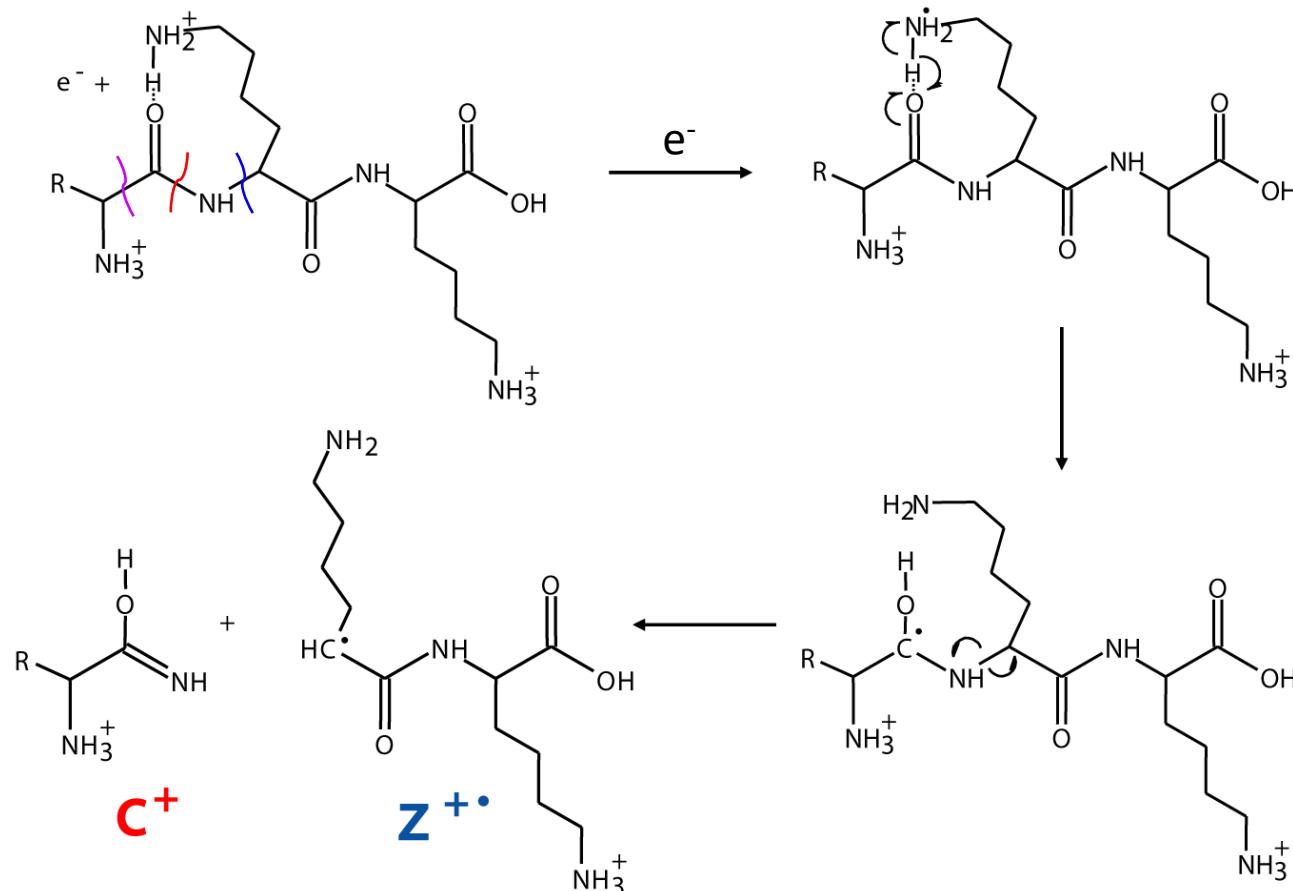
Prerequisite: multiply charged precursor ions, $n \geq 2$!

ETD is not applicable to 1+ or negatively charged ions

The multiply charged ions interact with a *low-energy electron beam* in ECD.

Electron Transfer Dissociation

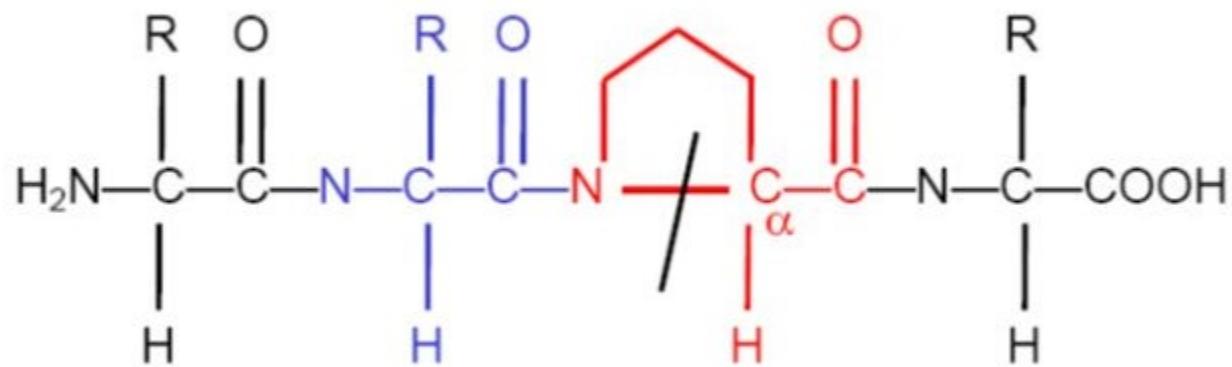
(produces primarily c and z ions)



Anions are used as vehicles for delivery to multiply-protonated peptides in ion-trap mass spectrometry

ETD: No Cleavage at Proline

Even though the N-C_a bond is cleaved, no respective c and z fragments are formed since they stay connected via the Proline ring system.



Phosphate

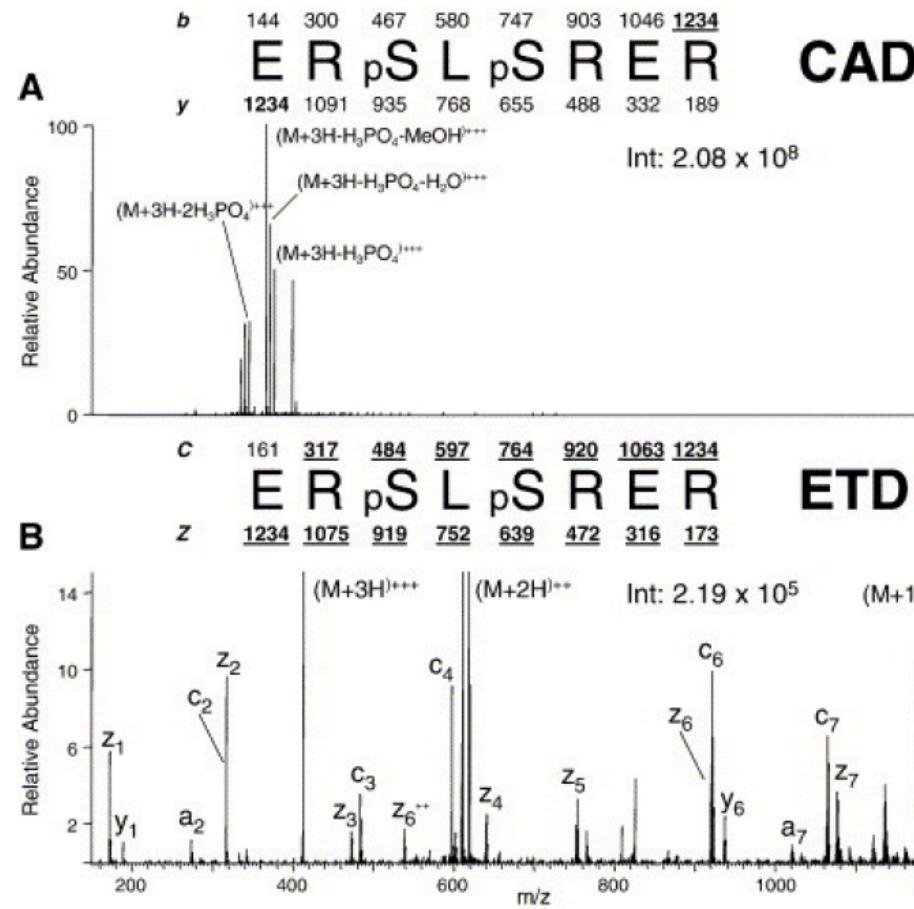


Fig. 3. Comparison of CAD vs. ETD spectrum of a phosphorylated peptide. Consecutive single-scan CAD vs. ETD mass spectrum comparison of phosphorylated peptides generated from a tryptic digest of human nuclear proteins recorded during a data-dependent analysis (nHPLC- μ ESI-MS/MS). All peptides were converted to methyl esters and enriched for phosphorylated peptides by immobilized metal affinity chromatography before analysis. (A) CAD spectrum dominated by fragment ions corresponding to the loss of phosphoric acid and either methanol or water. (B) ETD spectrum containing a near complete series of c- and z-type product ions. Note that the spectrum is devoid of fragment ions corresponding to the loss of phosphoric acid

Top-Down Proteomics

Advantages

- Complete protein sequence
- The ability to locate and characterize PTMs
- Protein isoforms determination
- Elimination of the time-consuming protein digestion

Disadvantages

- Expensive instruments
- Many proteins not soluble, and activation/fragmentation methods (ECT, ETD) not efficient.
- Not applicable to a large scale due to a lack of intact protein fractionation methods that are integrated with tandem MS