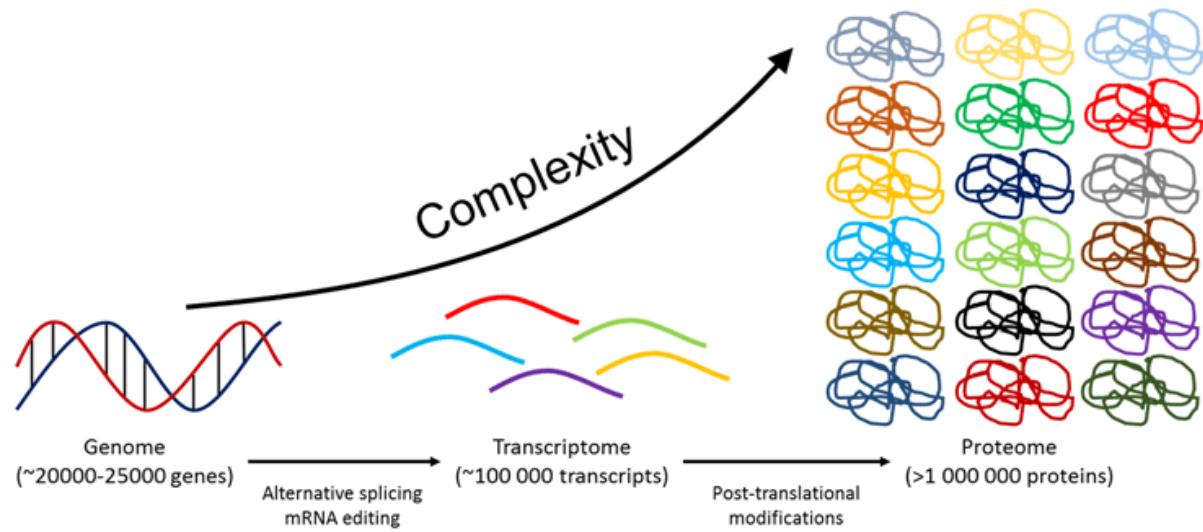


LECTURE 2: IDENTIFICATION OF PROTEINS IN COMPLEX MIXTURES - A MASS SPECTROMETRY APPROACH

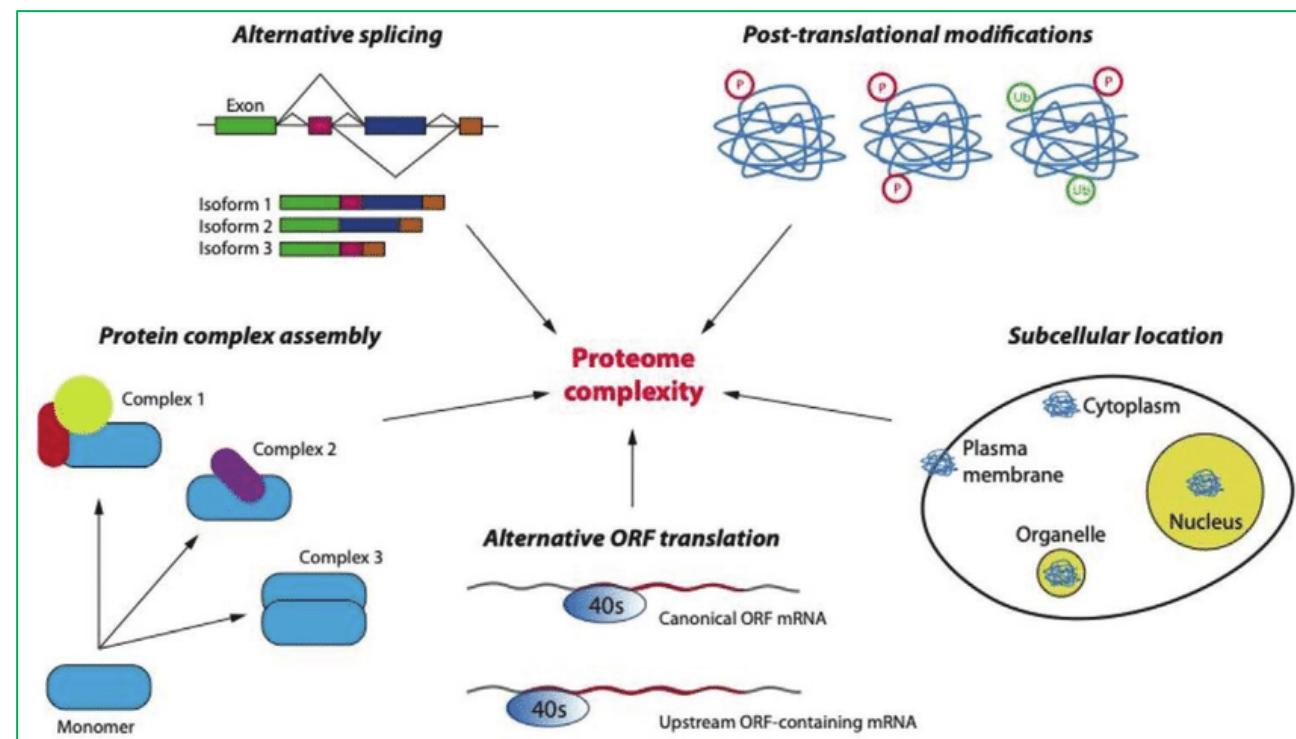
Bio312

Instructor: Dr. Lanlan Han
E-mail: lanlan.han@xjtlu.edu.cn

Complexity of Proteomics



1. Numbers of proteins
2. Diversity of cells or tissues
3. Dynamic changes in protein levels



<https://www.researchgate.net/profile/Juan-Pelta/publication/342275076/figure/fig1/AS:904991556530176@1592778209035/Overview-of-proteome-complexity-Numerous-factors-contribute-to-the-generation-of-complex.ppm>

<https://www.researchgate.net/profile/Floris-Van-Den-Brink/publication/313535513/figure/fig4/AS:460194112118787@1486730234132/The-human-proteome-contains-many-more-species-compared-to-the-human-genome-making-it.png>

General Workflow in Proteomics Analysis

The **charged** groups, **hydrophobic** region, **size** and specific ligand binding **affinity** largely determine the purification behavior of proteins.

Sample Selecting and Storing

2DGE and/or multidimensional LC

Protein Separation (Simplify sample)

Protein Identification

Chemical Sequencing

Mass Spectrometry Sequencing

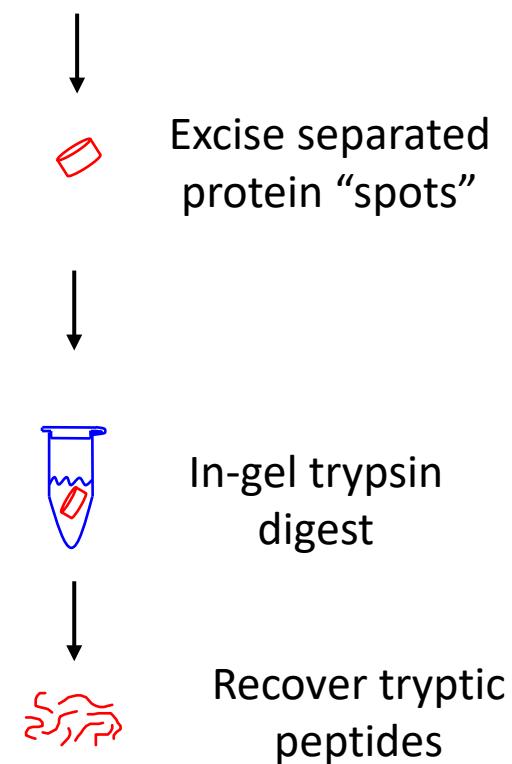
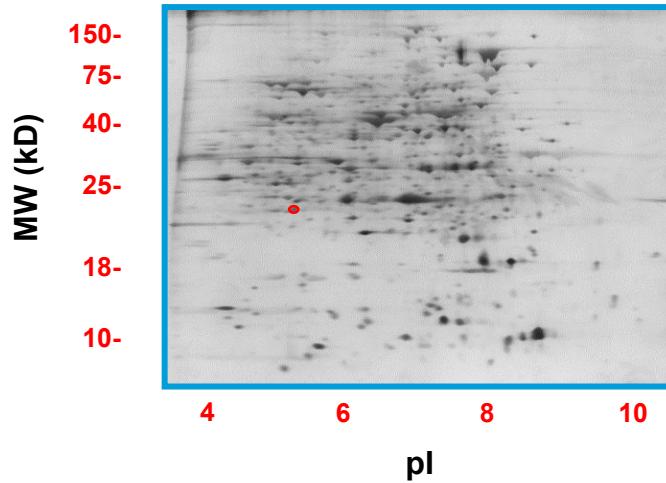
Database enquiry

Protein Structure/Function Screening, etc.

- Edman (N-terminal sequence)
- C-terminal sequence

Strategies for Protein Separation₁

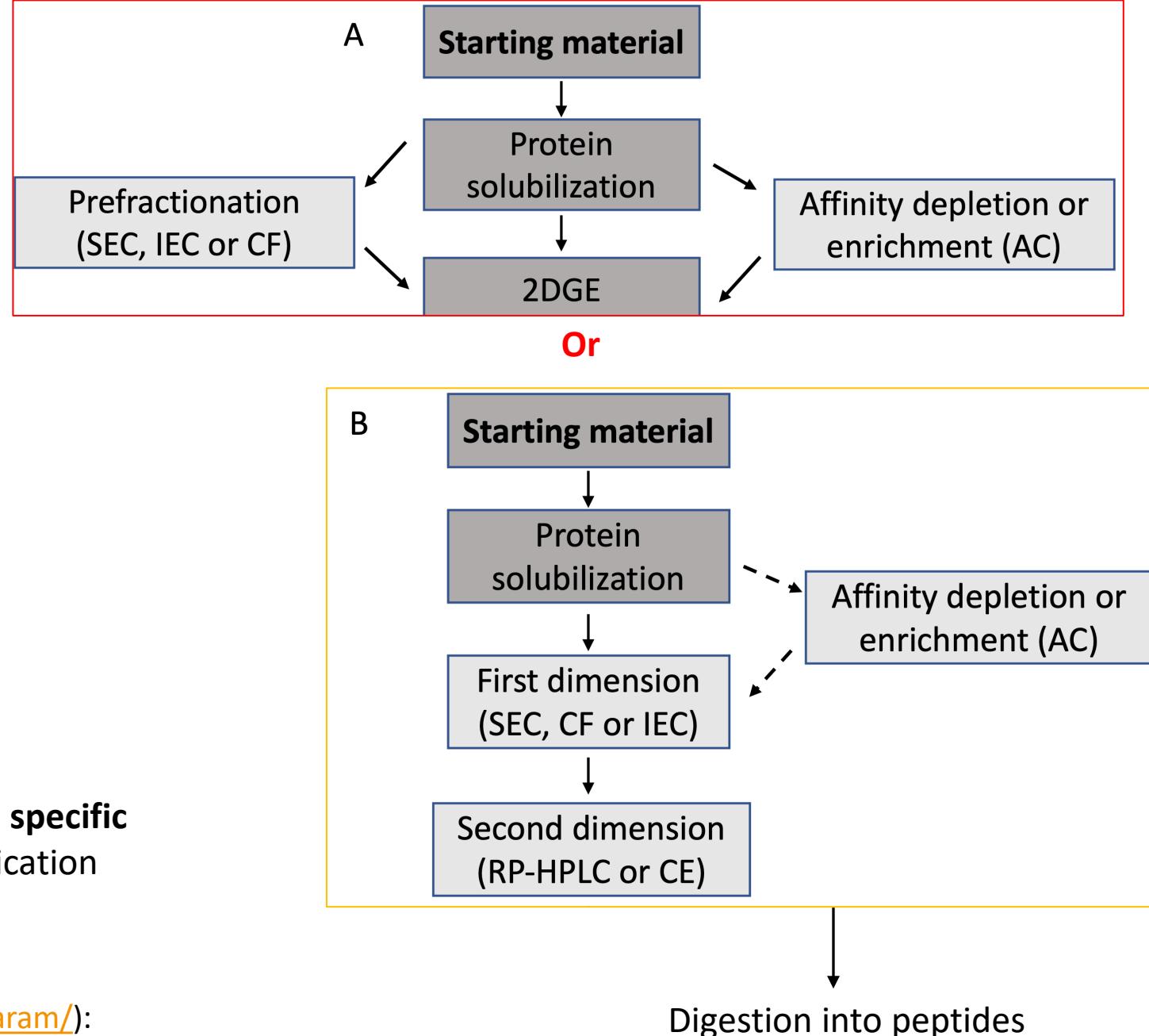
2-D Gel Electrophoresis



Strategies for Protein Separation₂

Multidimensional Liquid chromatography

- The **charged groups, hydrophobic region, size and specific ligand binding affinity** largely determine the purification behavior of proteins.
- **ExPASy - ProtParam tool** (<https://web.expasy.org/protparam/>):
A computation tool to calculate various physical and chemical parameters for a given protein



ExPASy - ProtParam tool

Please note that you may only fill out **one** of the following fields at a time.

Enter a Swiss-Prot/TrEMBL accession number (AC) (for example **P05130**) or a sequence identifier (ID)

Or you can paste your own amino acid sequence (in one-letter code) in the box below:

```
MGKRILLLEKERNLAHFLSLELQKEQYRVDLVEEGQKALSMALQTDYDLILLNVNLGDMMAQDFAEKLRSRTKPASVIMILDHWEDLQEELEVQRFAVSYIYKPVLIENLVARISAIFRGDRFIDQHCSLMKVPRPTYRNLRIDVEHHTVYRGEEMIALTRREYDLLATLMGSKKVLTREQLLESVWKYESATETNIVDVYIRYLRSKLDVKQQKSYIKTVRGVGYTMQE
```

Number of amino acids: 229

Molecular weight: 26880.19

Theoretical pI: 6.55

Amino acid composition: CSV format

Ala (A)	12	5.2%
Arg (R)	18	7.9%
Asn (N)	6	2.6%
Asp (D)	13	5.7%
Cys (C)	1	0.4%
Gln (Q)	11	4.8%
Glu (E)	21	9.2%
Gly (G)	9	3.9%
His (H)	5	2.2%
Ile (I)	14	6.1%
Leu (L)	31	13.5%
Lys (K)	15	6.6%
Met (M)	9	3.9%
Phe (F)	5	2.2%
Pro (P)	3	1.3%
Ser (S)	12	5.2%
Thr (T)	11	4.8%
Trp (W)	2	0.9%
Tyr (Y)	12	5.2%
Val (V)	19	8.3%
Pyl (O)	0	0.0%
Sec (U)	0	0.0%
(B)	0	0.0%
(Z)	0	0.0%
(X)	0	0.0%

Total number of negatively charged residues (Asp + Glu): 34
Total number of positively charged residues (Arg + Lys): 33

Atomic composition:

Carbon	C	1201
Hydrogen	H	1939
Nitrogen	N	327
Oxygen	O	350
Sulfur	S	10

Formula: C₁₂₀₁H₁₉₃₉N₃₂₇O₃₅₀S₁₀
Total number of atoms: 3827

Extinction coefficients:

Extinction coefficients are in units of M⁻¹ cm⁻¹, at 280 nm measured in water.

$$\text{Beer's Law (determine } A = \epsilon bc\text{)}$$

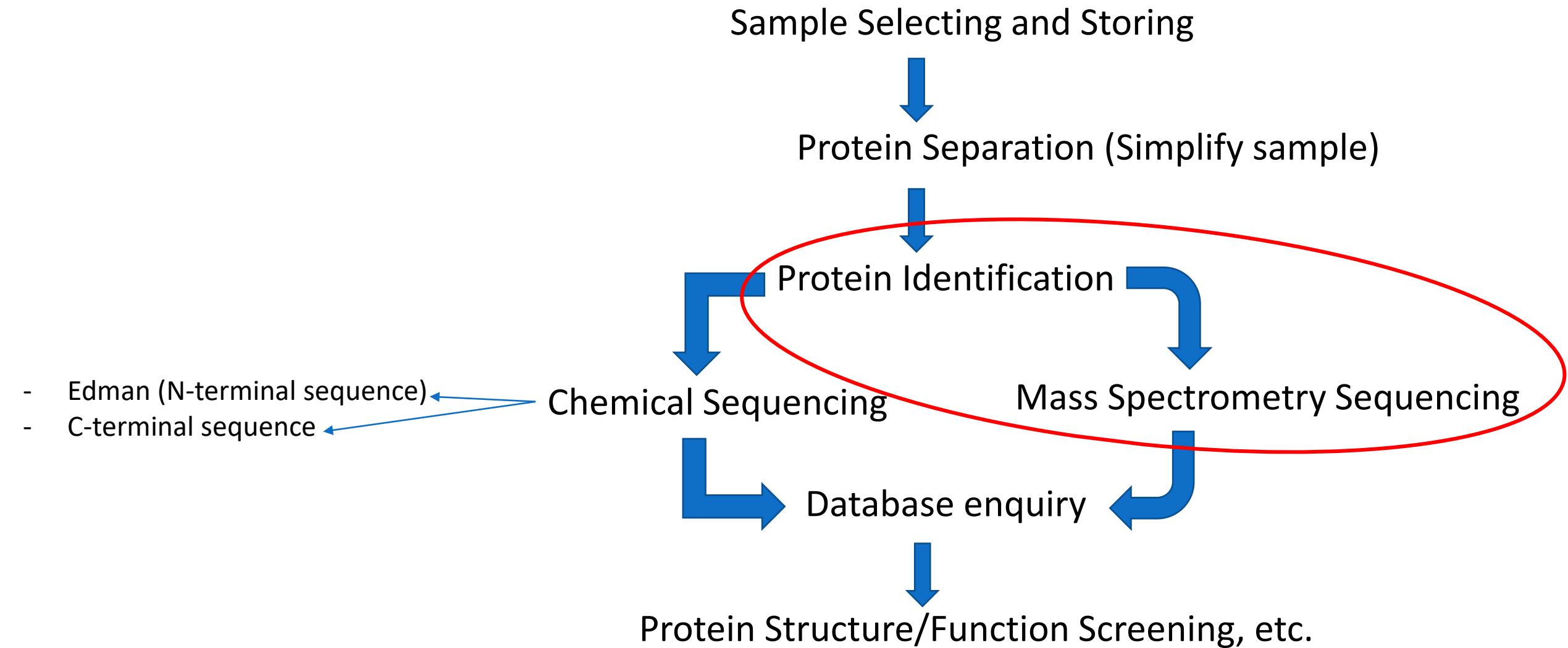
Ext. coefficient 28880
Abs 0.1% (=1 g/l) 1.074, assuming all pairs of Cys residues form cystines

Ext. coefficient 28880
Abs 0.1% (=1 g/l) 1.074, assuming all Cys residues are reduced

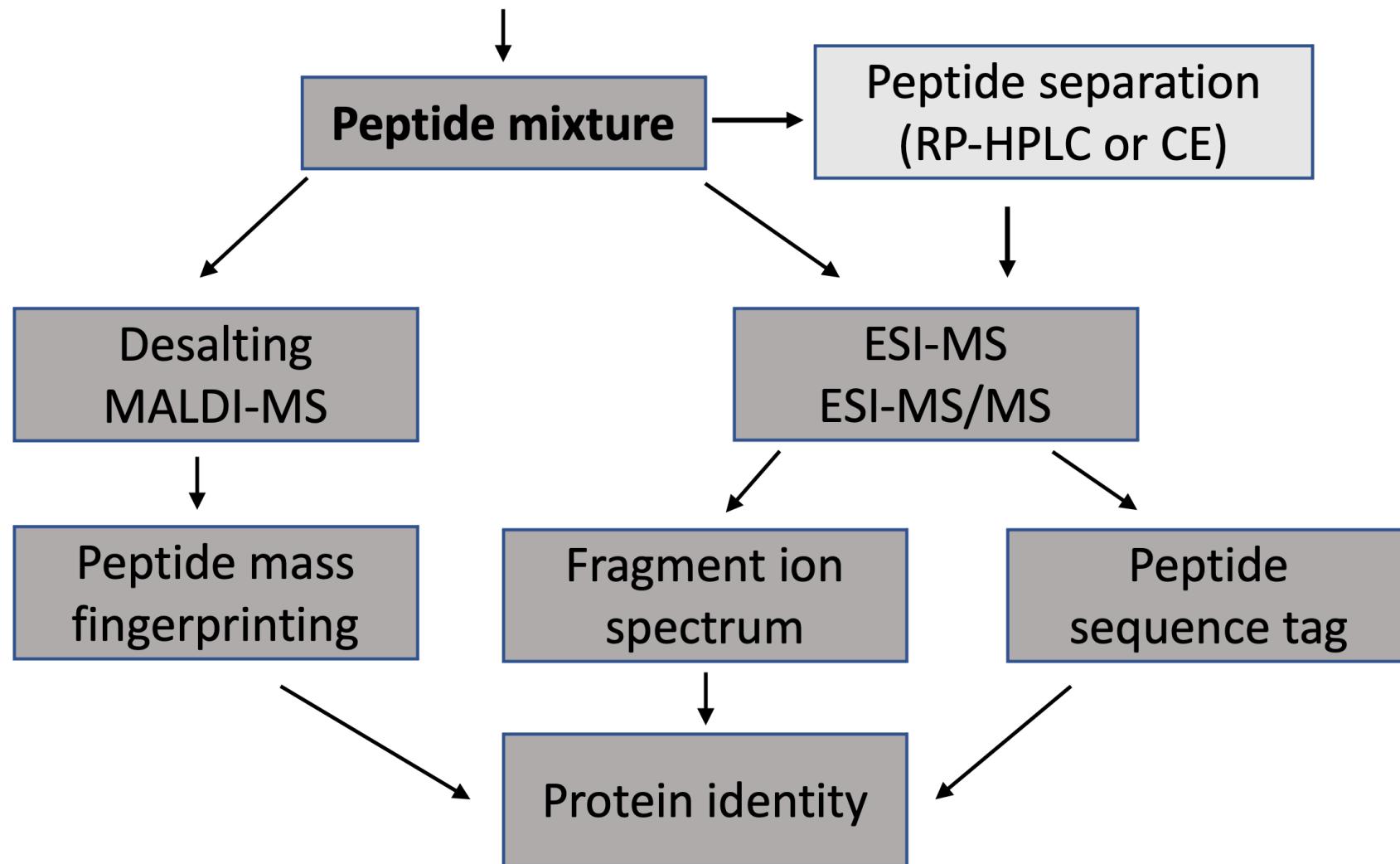
Multi-Dimensional Liquid Chromatography

- Pros:
 - Large sample volume
 - Detect low abundancy protein
 - Good for separating membrane proteins or very basic proteins
 - Separate both proteins and peptides
 - Connect directly to MS
- Cons:
 - Can not know the pI and MW of proteins
 - Need to consider the compatibility of buffers or solvents in different steps

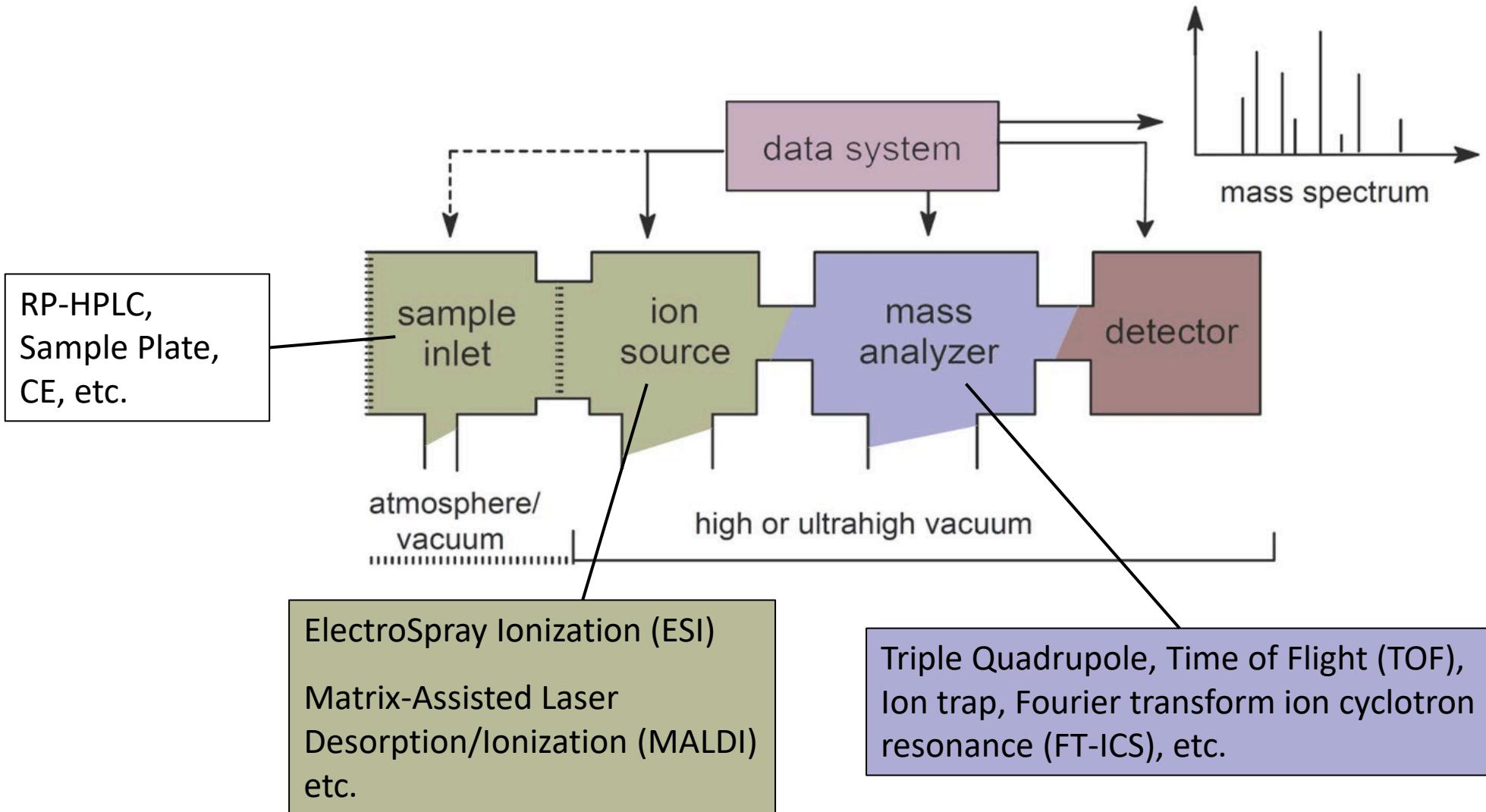
General Workflow in Proteomics Analysis



Mass Spectrometry-based Protein Identification

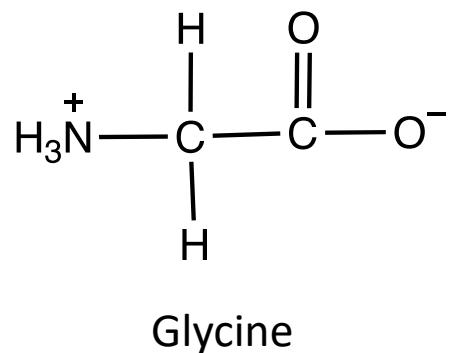
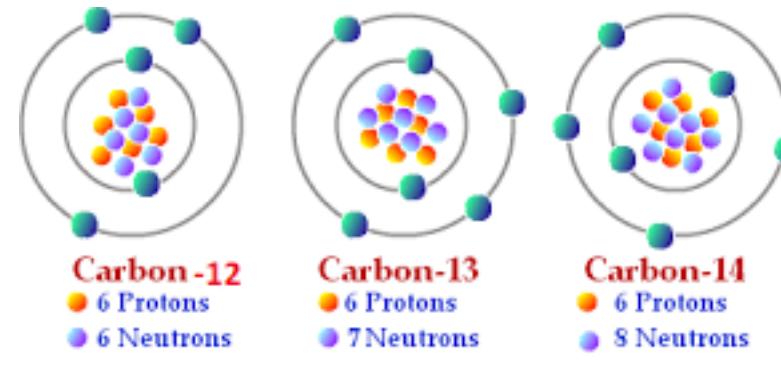
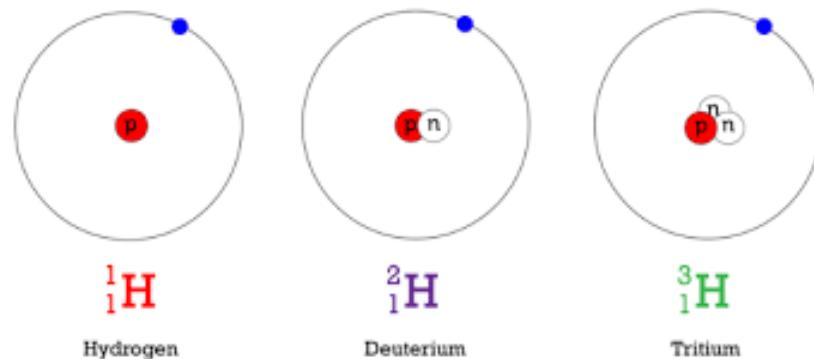


General Scheme of a Mass Spectrometer



How to Calculate Mass?

- Elements and their isotopes have unique masses.

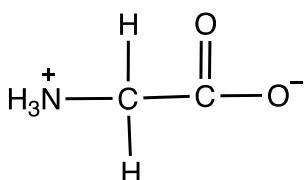


For Glycine ($\text{C}_2\text{H}_5\text{NO}_2$), the nominal mass is $\underline{12 \times 2 + 1 \times 5 + 14 \times 1} = 75$

If one of the carbons is ^{13}C , the nominal mass of Gly is $\underline{12 \times 1 + 13 \times 1 + 1 \times 5 + 14 \times 1} = 76$

- Elements and their isotopes have unique masses.
- **Monoisotopic Mass:** *Exact mass of an ion or molecule calculated using the mass of the most abundant isotope of each element.*
- **Accurate Mass:** *Experimentally determined* mass of an ion of known charge

For Glycine,
 $(\text{C}_2\text{H}_5\text{NO}_2)$,



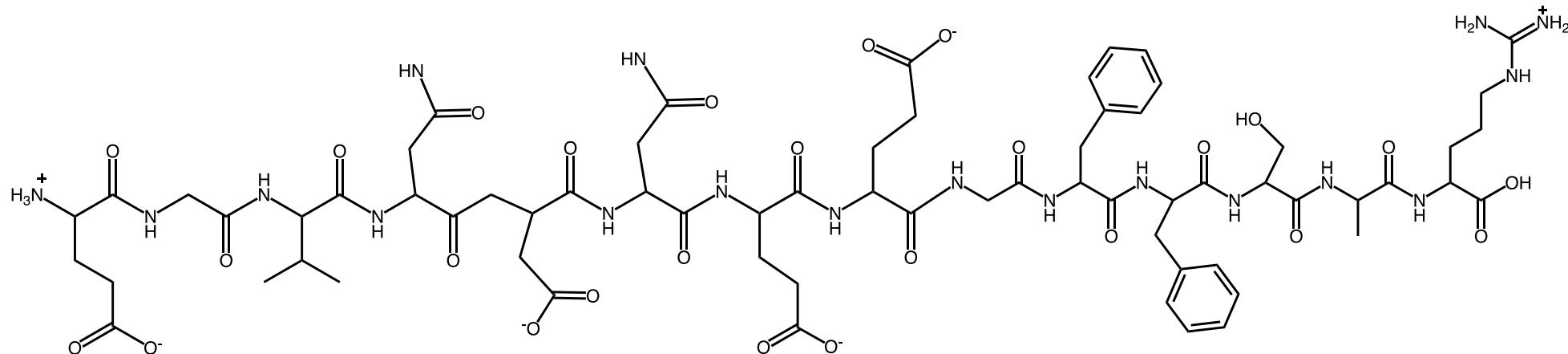
ATOM	Nominal Mass	Monoisotopic Mass
C	$12 \times 2 = 24$	$12.0000 \times 2 = 24.0000$
H	$1 \times 5 = 5$	$1.0078 \times 5 = 5.039$
O	$16 \times 2 = 32$	$15.9949 \times 2 = 31.9898$
N	$14 \times 1 = 14$	$14.0031 \times 1 = 14.0031$
S	$32 \times 0 = 0$	$31.9720 \times 0 = 0$
	75	75.0319
	Low Resolution	High Resolution

Recognizing Multiple Charged Ions

- Mass spectrometers operate on the basis of mass-to-charge ratio (**m/z**, unit: Thomson or Th). In other words, the analytes (peptides) need to be charged (ions).

$$m/z = \frac{m+z}{z}$$

- Single charge (MALDI) $m/z = (M+H^+)$
 - Double charge (ESI) $m/z = (M+2H^+)/2$
 - n charge (ESI) $m/z = (M+nH^+)/n$

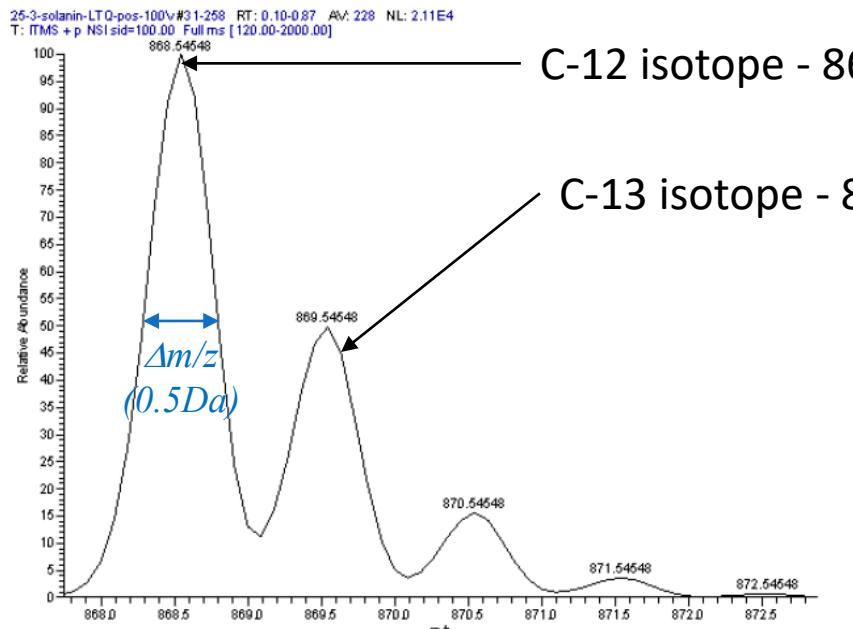


Mass Resolution

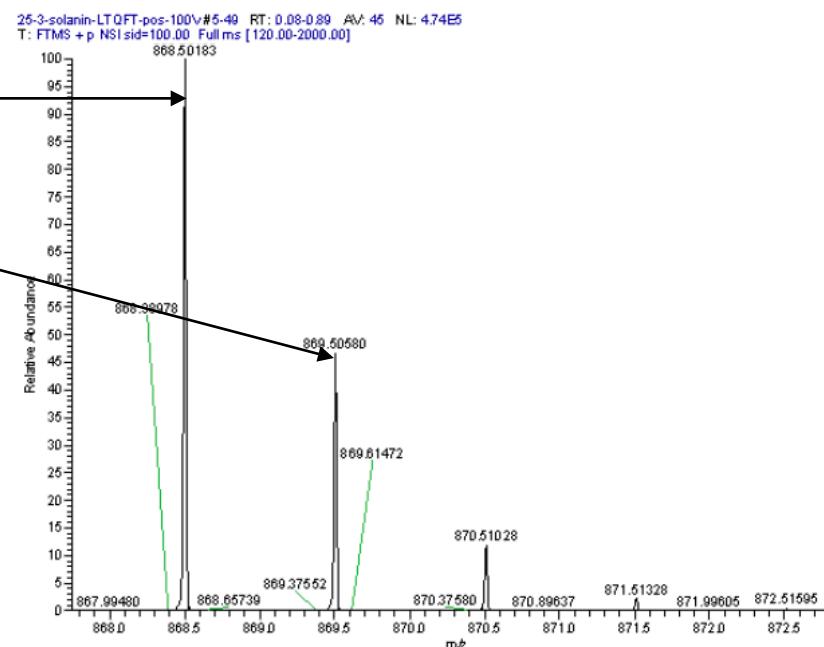
$$\text{Resolution} = \frac{m/z}{\Delta m/z}$$

full width at half maximum

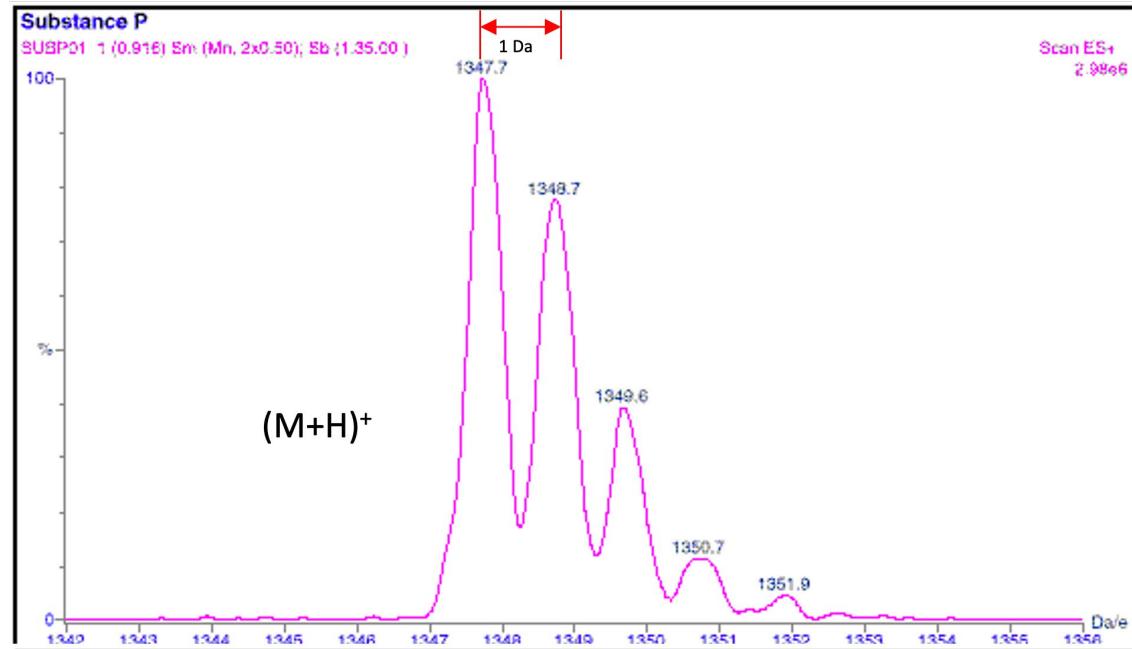
- The ability of the instrument to resolve two closely placed peaks.



low resolution, Resolution = 1737 (868.5/0.5)

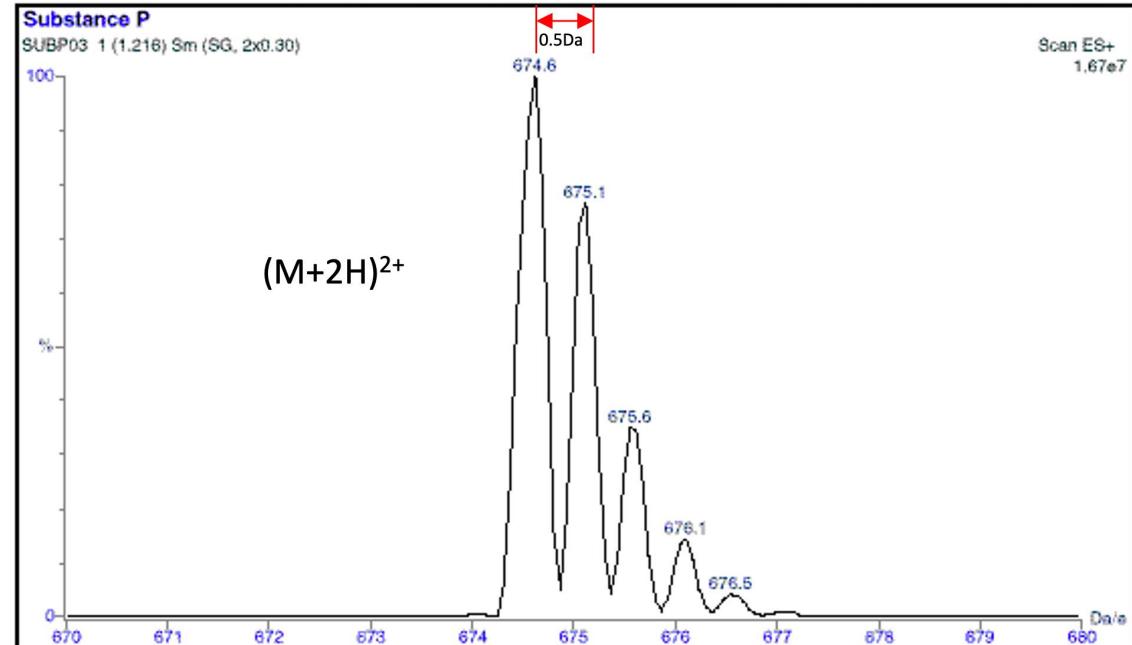


high resolution, resolution = 48,250 (868.5/0.018)
m/z 868.5, peak width ~0.018



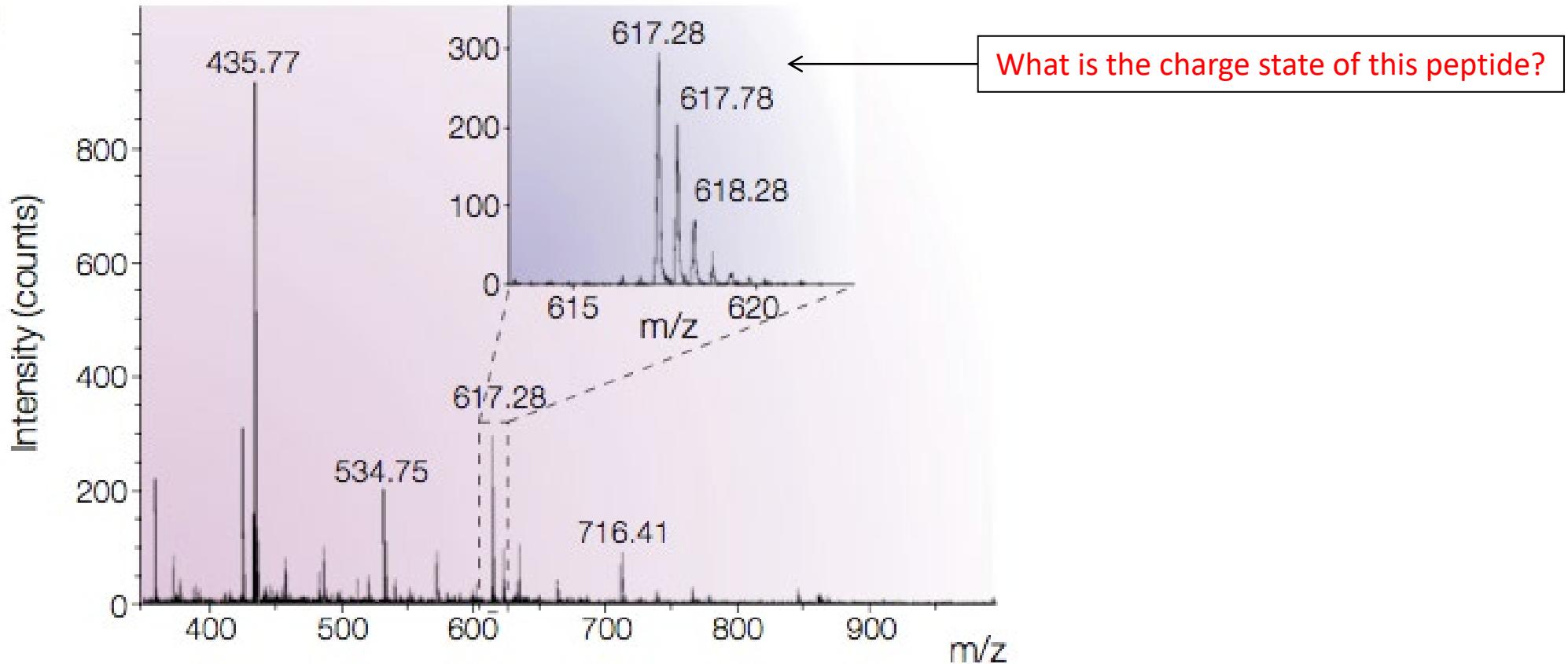
$$m/z = \frac{m+z}{z}$$

Isotopes of singly charged ions are separated by 1 Da



Isotopes of doubly charged ions are separated by 0.5 Da

MS Spectrum (i.e., peptide ions)

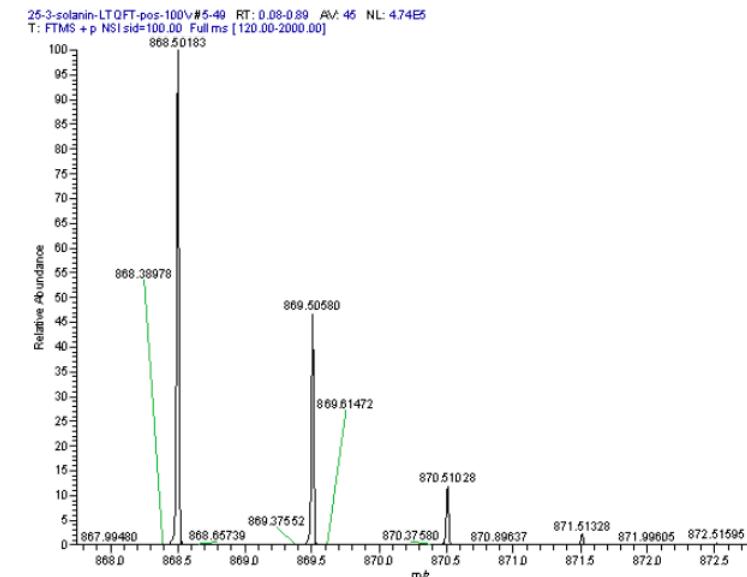
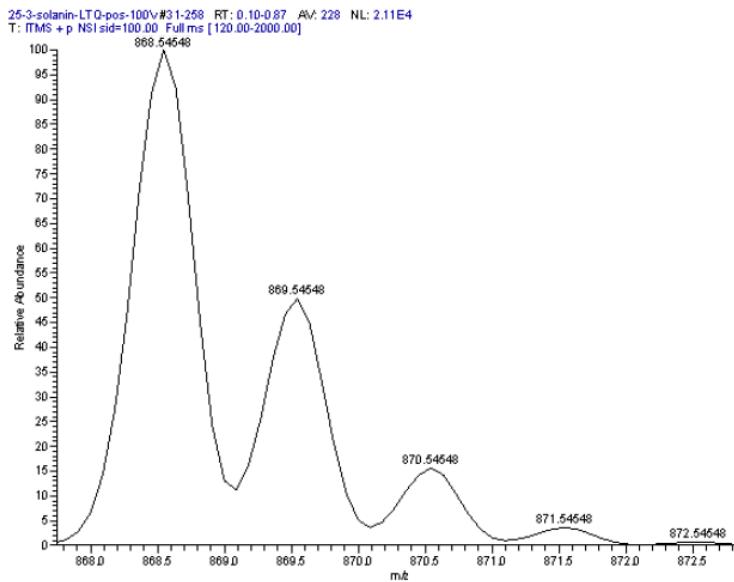


Each peak is a different peptide, separated based on m/z
A single peptide is selected by the instrument for the second MS

Mass Accuracy

$$\frac{M_{\text{exp}} - M_{\text{true}}}{M_{\text{true}}} \times 10^6$$

- The relative percent difference between the measured mass and the true mass (usually represented in ppm).
- The lower the number the better the mass accuracy



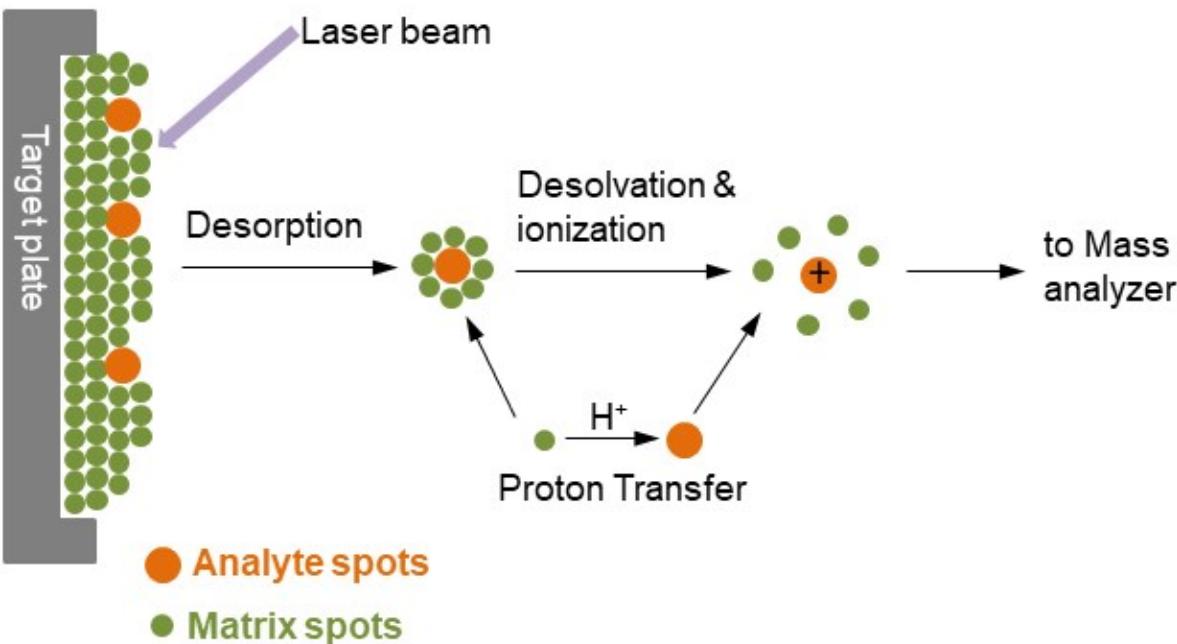
MS Basics

1. Soft Ionization

Goal: ionize (i.e., charge) peptide fragments without destroying molecule

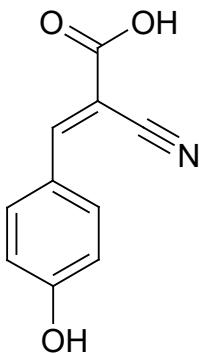
- Positive ionization (protonate amine groups)
- Negative ionization (deprotonate carboxylics and alcohols)
- Two common ionization methods used in proteomics:
 - MALDI (Matrix-Assisted Laser Desorption/Ionization)
 - ESI (ElectroSpray Ionization)

MALDI

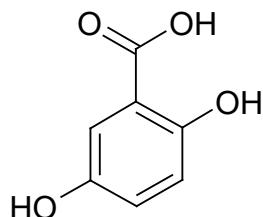


- Peptide/protein analytes of interest are mixed with an excess of **an aromatic matrix molecule** and are co-crystallized on the MALDI target plate
- The crystals are targeted by a shorter laser pulse (UV pulse)
- **Matrix molecules** absorb energy and the heat released results in the desorption (sublimation)
- **Analyte** is ionized by gas-phase proton transfer (perhaps from ionized **matrix molecules**)
 - Protonation with one charge (H^+)

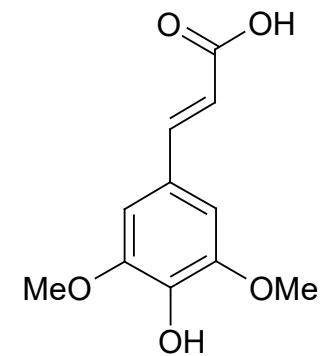
MALDI Matrices



4-hydroxy- α -cyanocinnamic acid
("alpha-cyano" or HCCA)
peptides



2,5-dihydroxybenzoic acid (DHB)
peptides and **proteins**

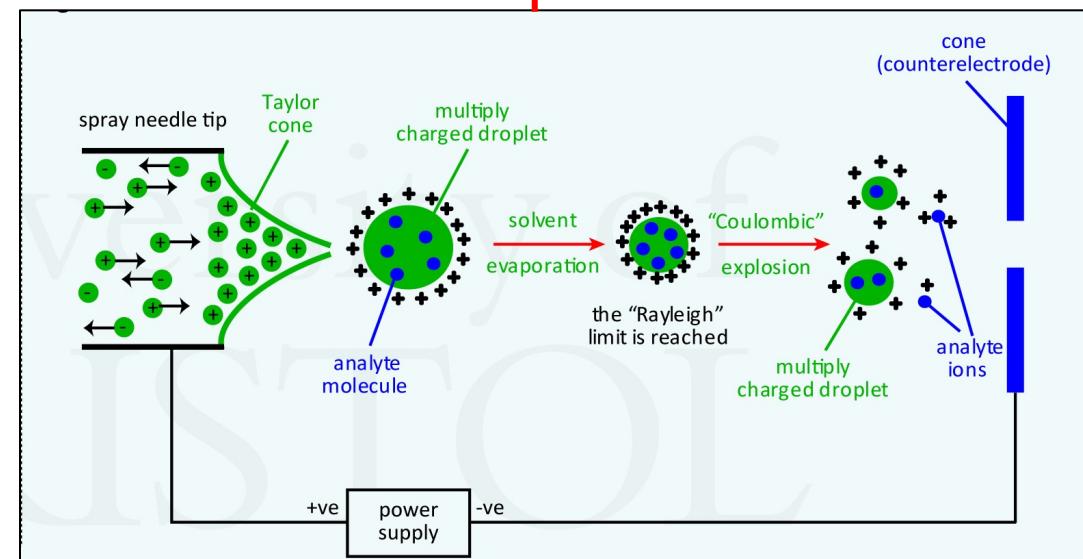
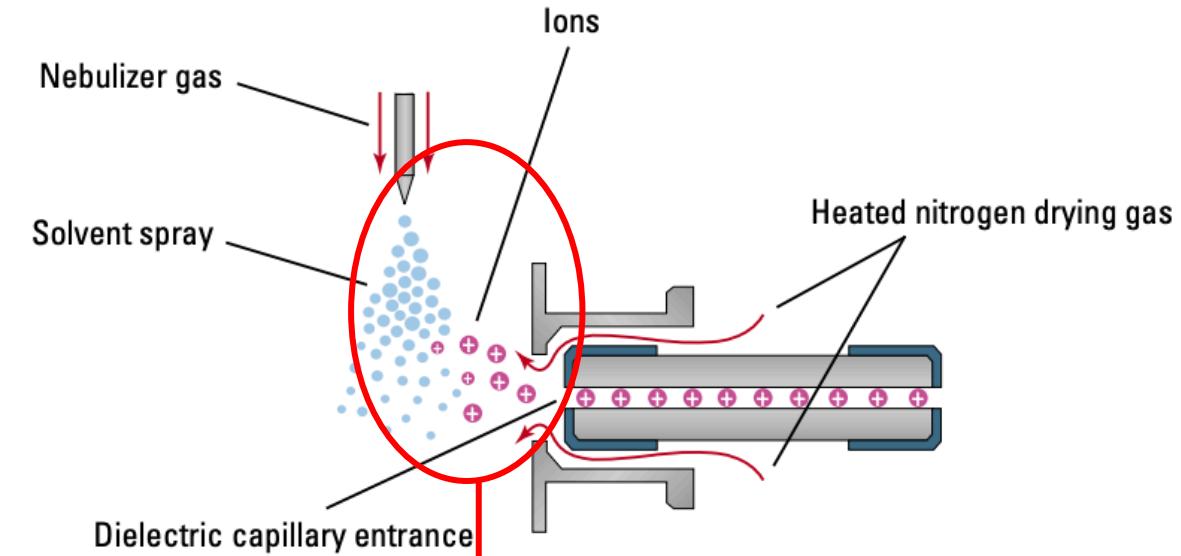


3,5-dimethoxy-4-hydroxycinnamic acid
(sinapinic acid)
proteins

matrices for 337 nm irradiation

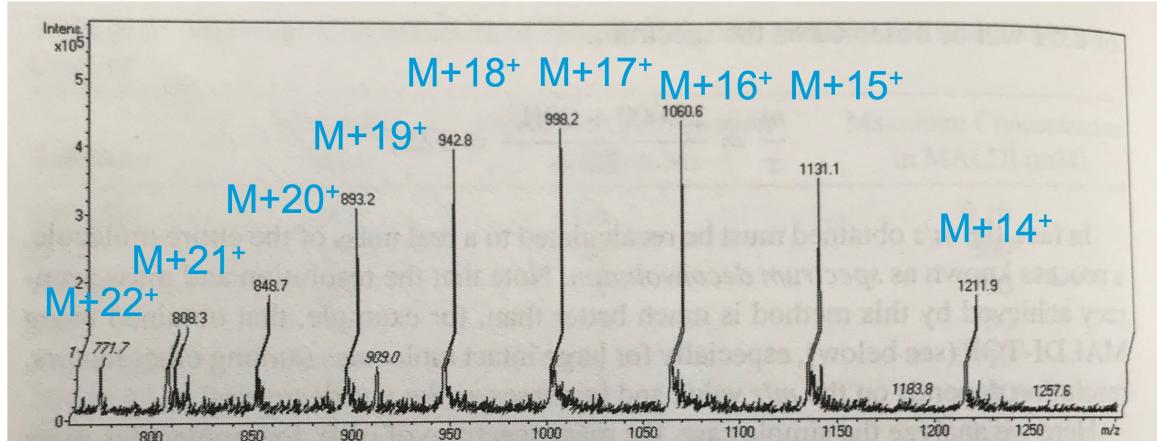
ESI

- Analyte dissolved in a suitable solvent flows through a small diameter capillary tube
 - ESI is compatible with liquid chromatography (LC)
- Liquid in the presence of a high electric field generates a fine “mist” or aerosol spray of highly charged droplets
 - Multiple charged ions are possible (1^+ , 2^+ , 3^+ , 4^+ , etc.)

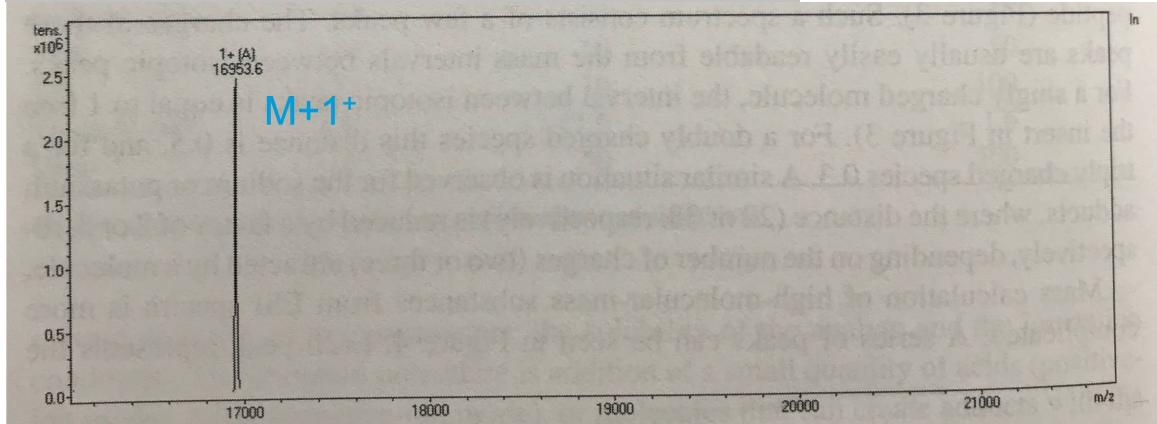


Mw Calculation

ESI-MS of horse heart myoglobin Multiple charged ions



MALDI-MS of horse heart myoglobin Singly charged ions



To deconvolute the myoglobin mass,

1. Calculate the charge

$$z_n = \frac{m_{n+1} - 1.0078}{m_n - m_{n+1}}$$

2. Calculate the mass of each ion
3. Average them with deviation

$$m/z = \frac{m+z}{z}$$

$$\text{myoglobin MW} = 16953.6 - 1 = 16952.6 \text{ Da}$$

MS Basics

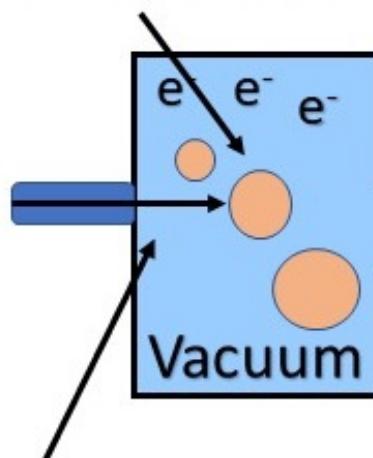
2. MS analyzers

Goal: Separate ions based on m/z (mass/charge) ratio

- Time-of-flight (TOF)
- Triple quadrupole
- Ion trap
- Fourier Transform Ion Cyclotron Resonance (FT-ICR)

TOF

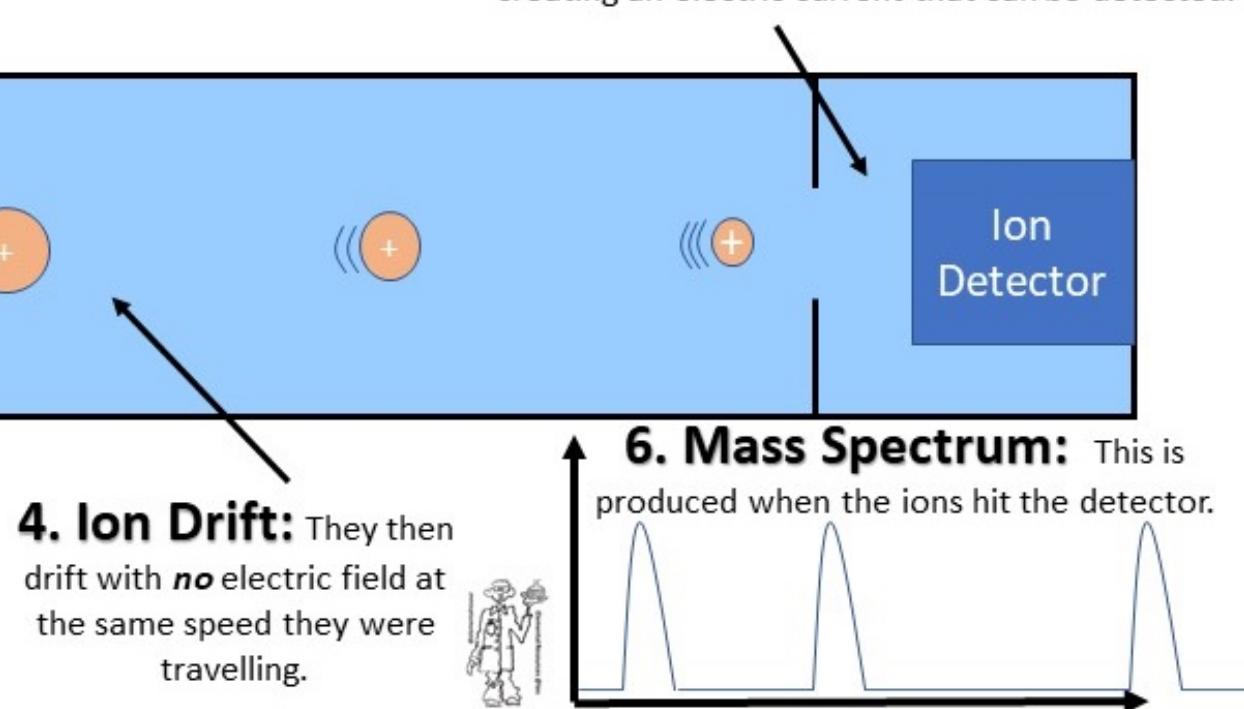
2. Ionisation: They are then ionised into positive ions by firing electrons at them (knocking electrons from the outer energy level) This is called *Electron Impact Ionisation*.



1. Vapourisation:

Samples of unknown substances (elements or compounds) are injected into a vacuum.

3. Electric Field: The positive ions are *accelerated* by an electric field. Lighter ions will accelerate greater than heavier ones.

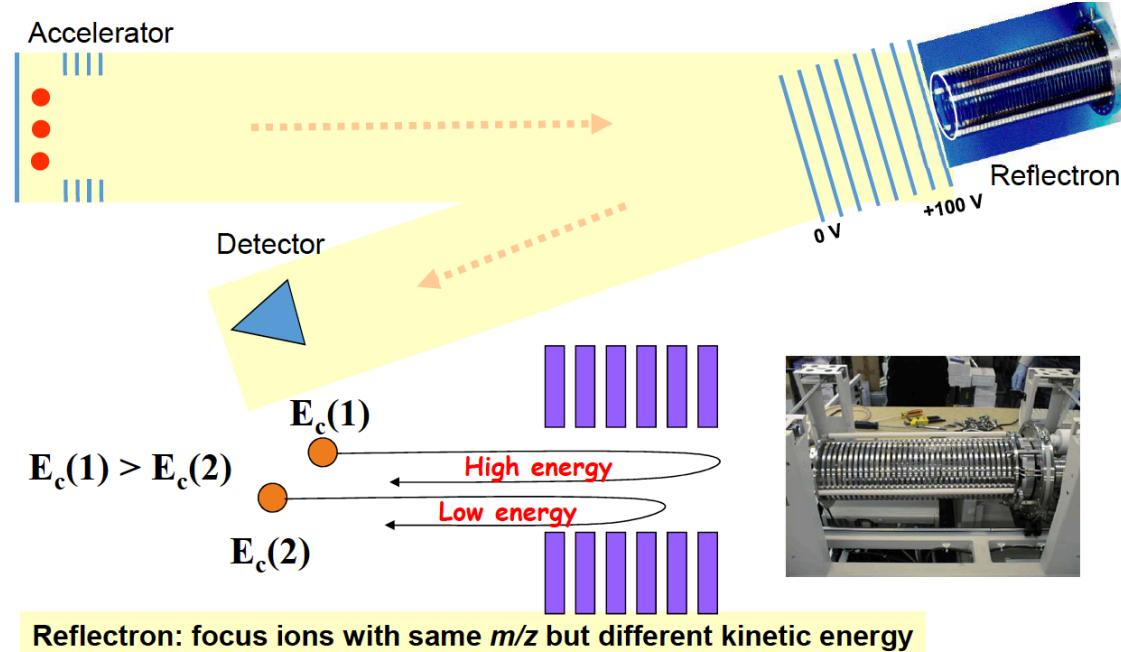


How does a Time of Flight Mass Spectrometer work?

5. Detector: Finally they hit the detector. Lighter ions will hit first, heavier ions later. The ions gain an electron when they hit the detector creating an electric current that can be detected.

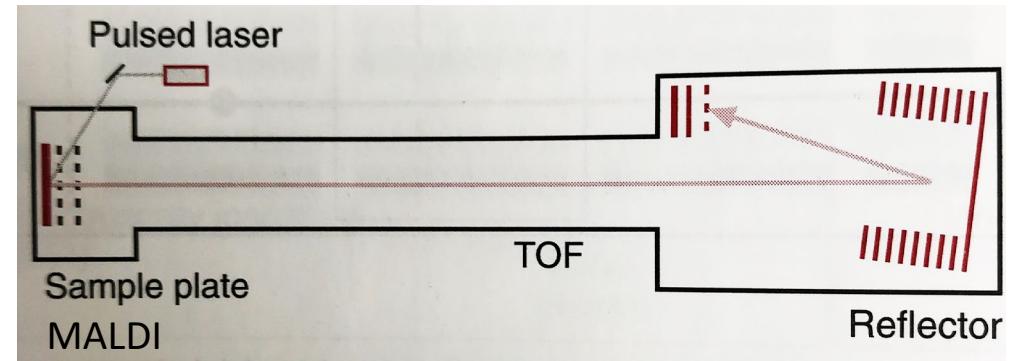
TOF

- Increase of resolution



• MALDI-TOF

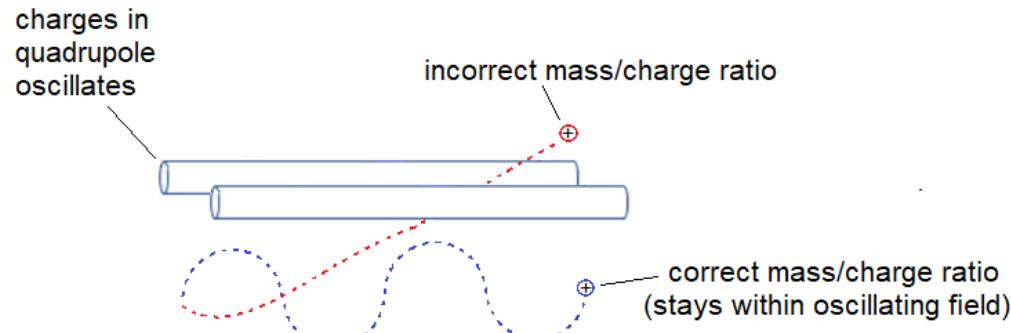
- MALDI produces singly charged ions
- The time of flight of any ions in TOF is inversely proportional to the square root of the molecular mass.
- MALDI-tandem TOF or MALDI-hybrid Quadrupole-TOF analyzers are more sensitive for high-throughput analysis.



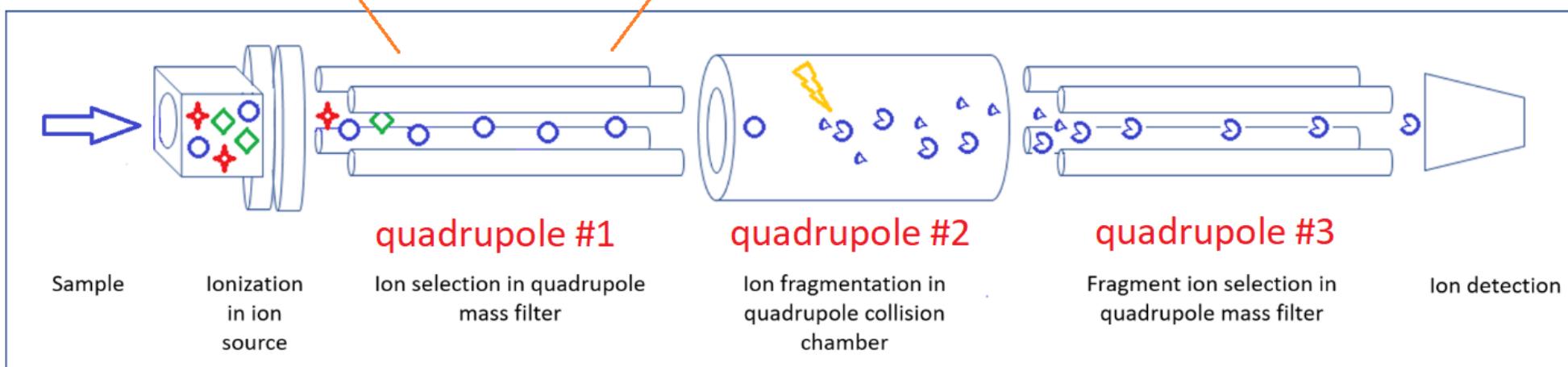
Quadrupole

Video on how triple quadrupole works:

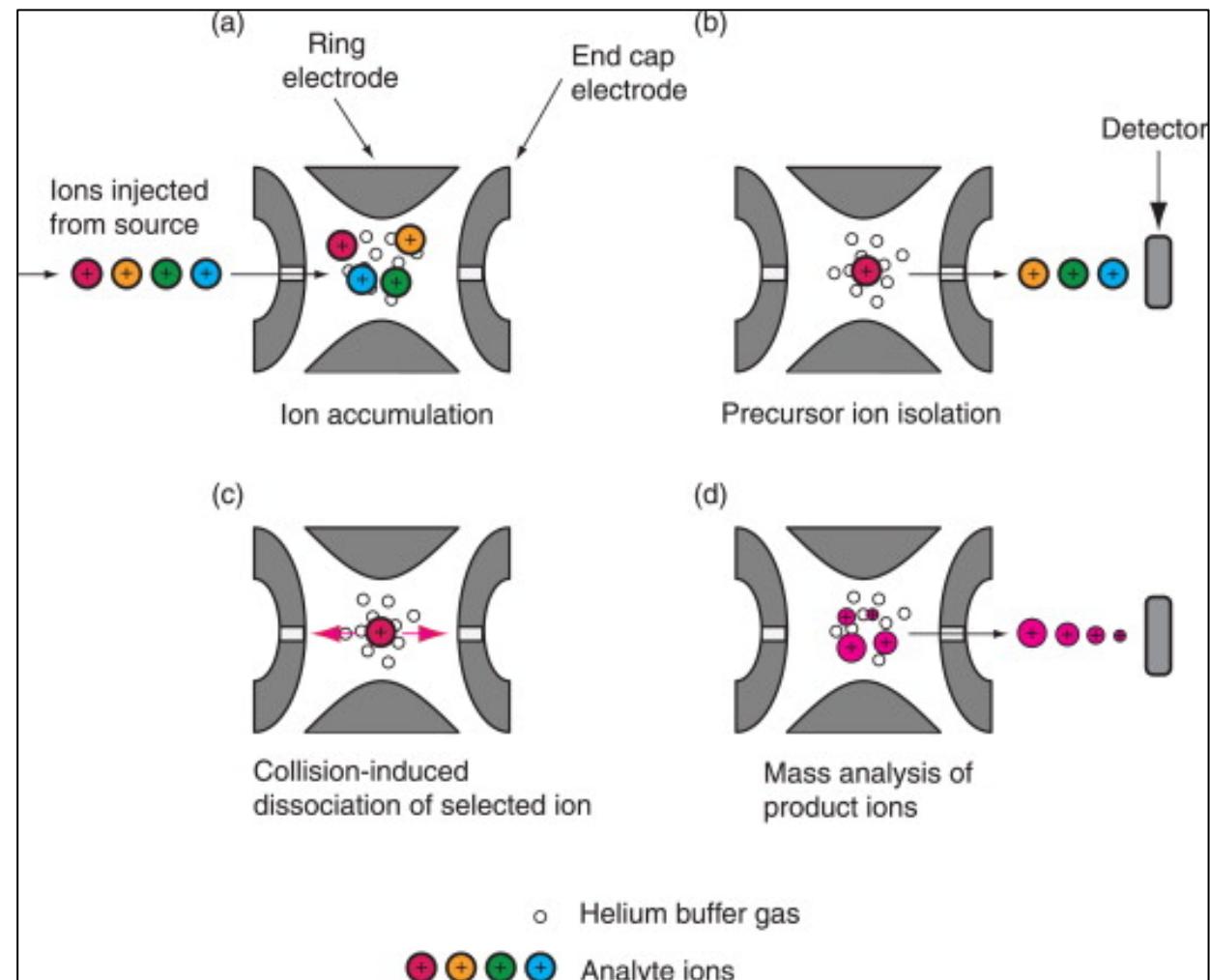
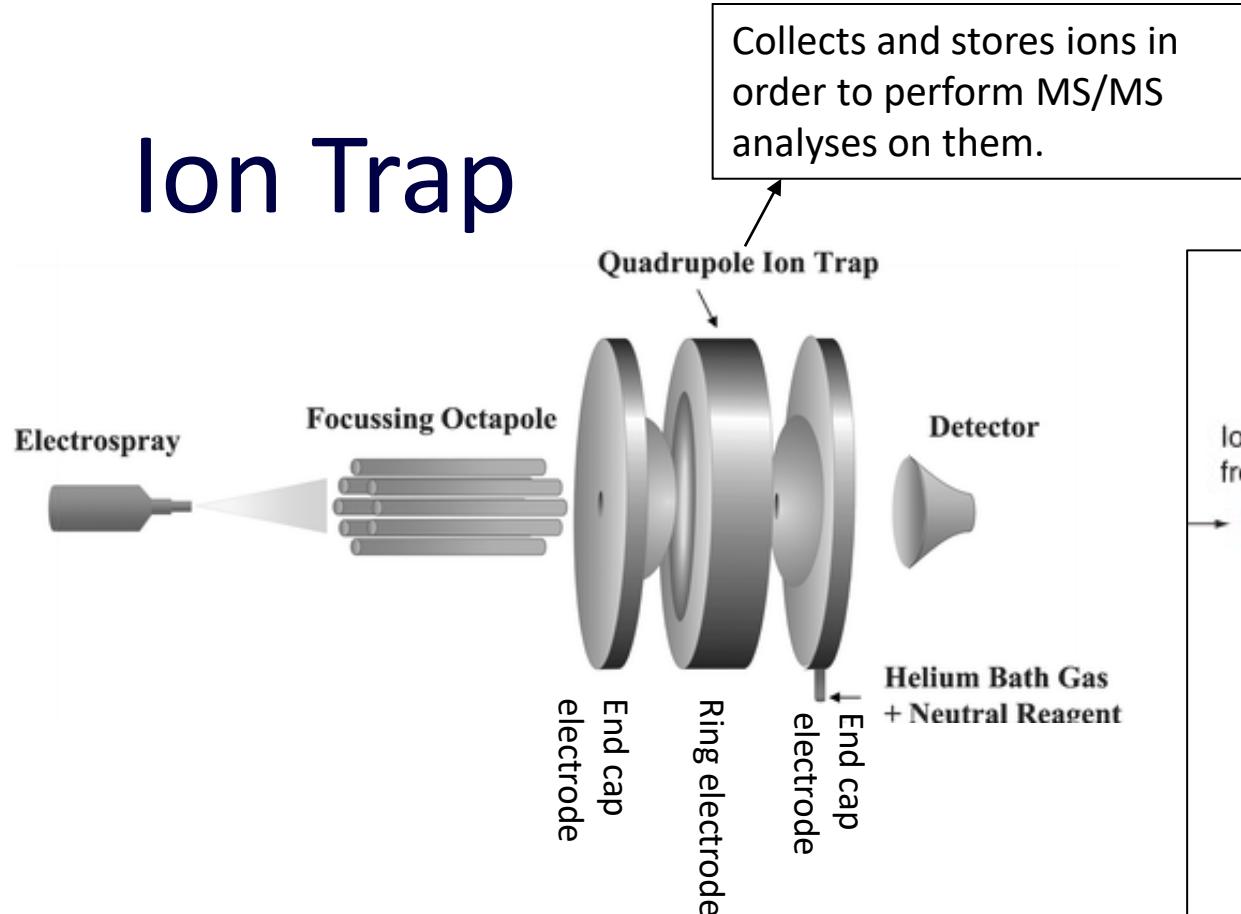
<https://www.youtube.com/watch?v=og2DUvF29zI>



Triple quadrupole



Ion Trap

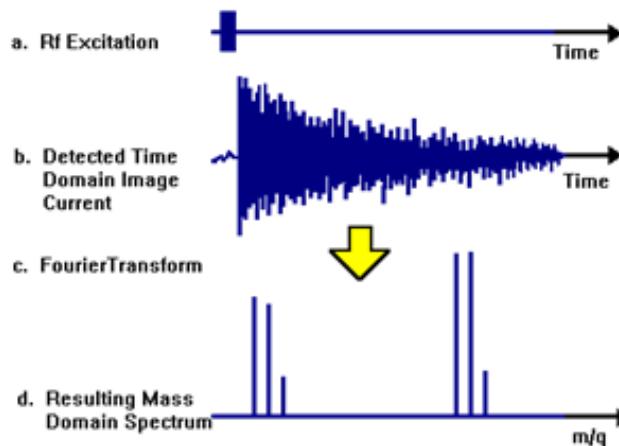
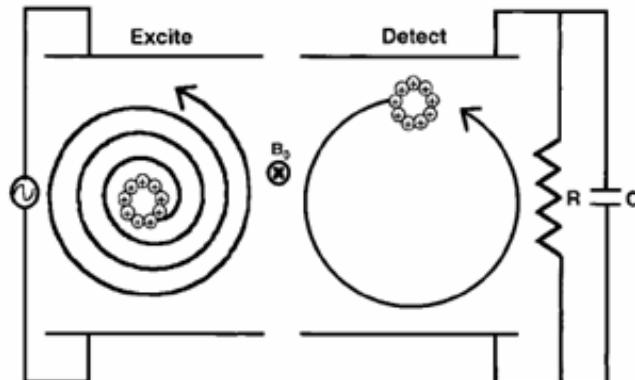


Fourier Transform Ion Cyclotron Resonance

- A mass analyzer for determining the mass-to-charge ratio (m/z) of ions based on the cyclotron frequency of the ions in a fixed magnetic field.

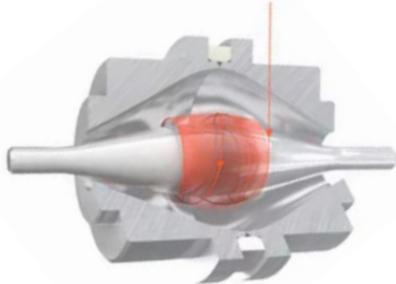
Ions are injected into a magnetic field , that causes them to travel in circular paths.

Excitation with oscillating electrical field increases the radius and enables a frequency measurement

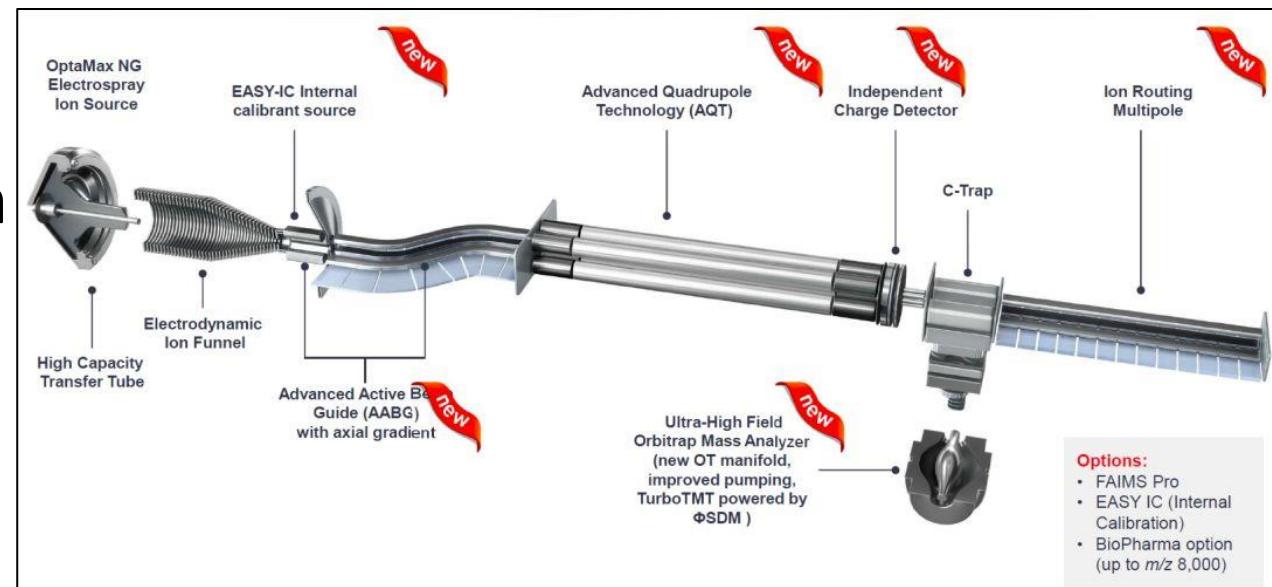


- High resolution
- High accuracy
- Very sensitive (the minimal quantity for detection is in the order of several hundred ions)
- Non destructive – the ions don't hit the detection plate so they can be selected for further fragmentation

Orbitrap



- Ions oscillate around the center electrode and along the z axis at m/z dependent frequencies and are non-destructively detected via z oscillations
- High mass accuracy and resolution
- High stability and speed
- High dynamic range
- Always part of hybrid mass spectrometers



<https://www.youtube.com/watch?v=zJagpUbnv-Y&t=5s>

MS/MS (Tandem MS) Terminology

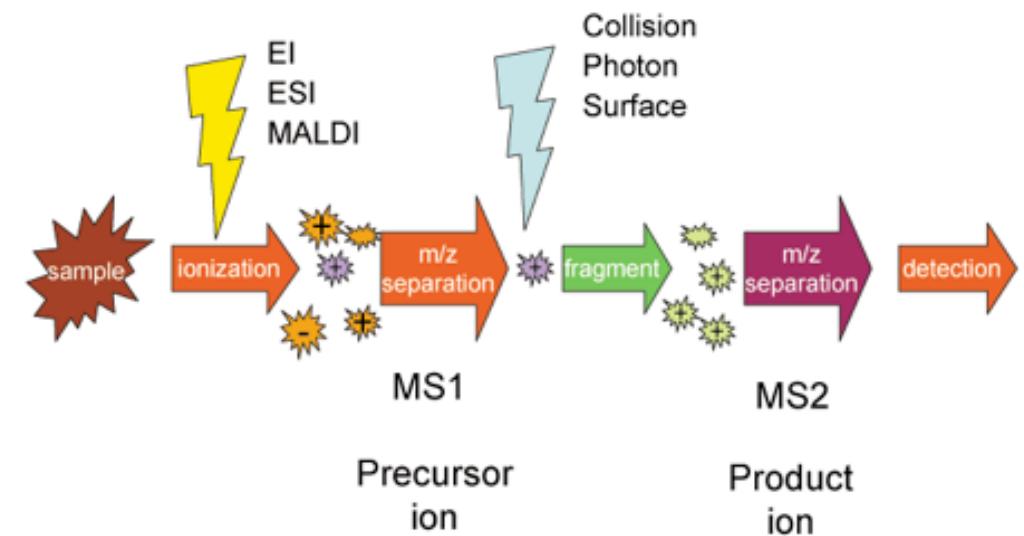
- Tandem mass spectrometry, also known as MS/MS or MS^2 , is a technique in instrumental analysis where two or more mass analyzers are coupled together using an additional reaction step to increase their abilities to analyze samples.

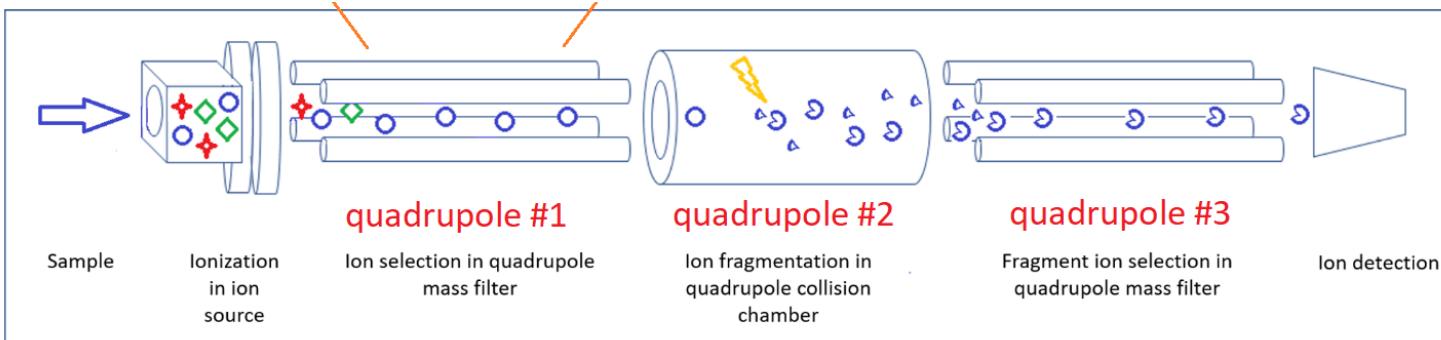
- Molecular ion / precursor ion (parent ion)

Ion formed by ionization of the analyte species

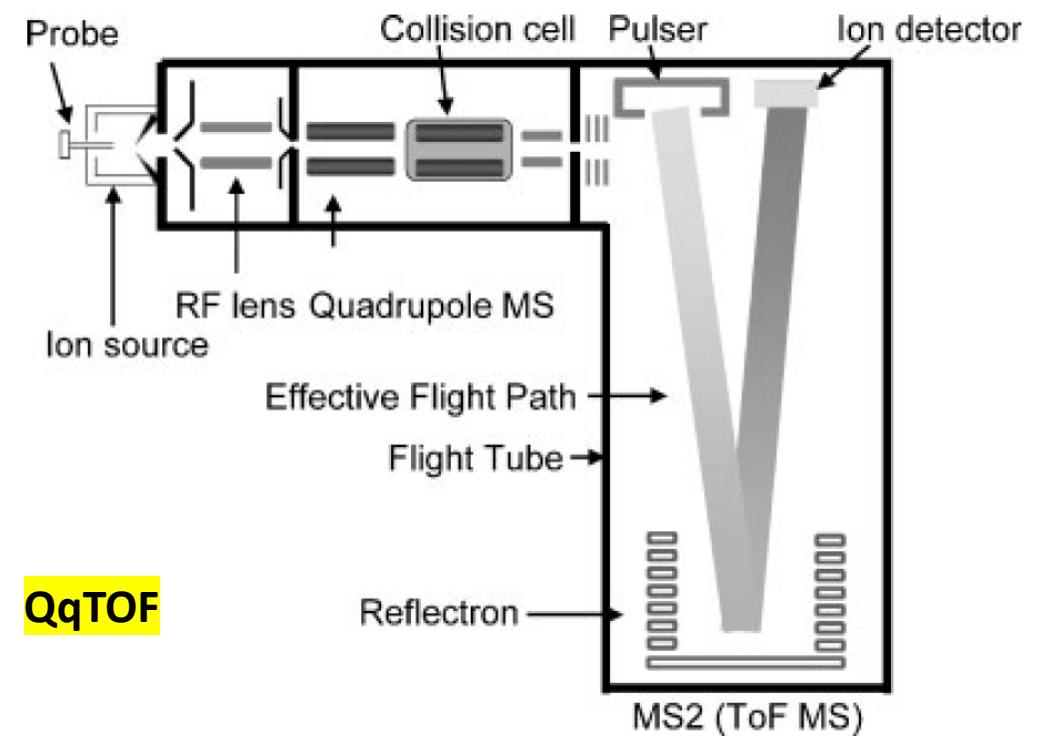
- Fragment ions / product ions (daughter ions)

Ions formed by the gas-phase fragmentation of the molecular ion



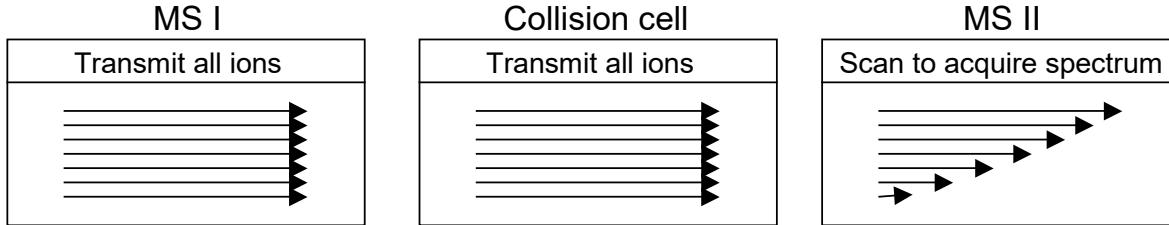


Triple Quadrupole (QqQ)



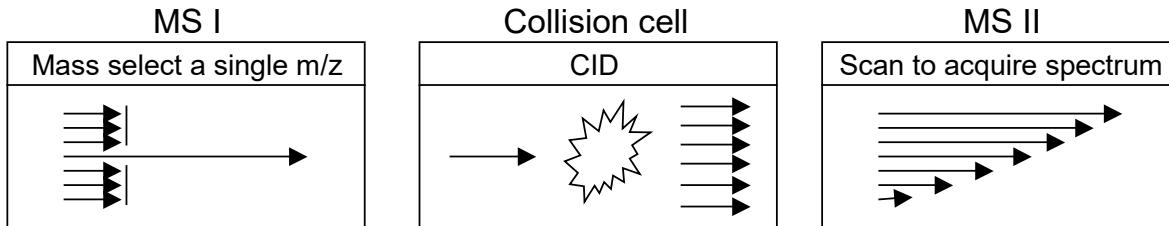
Four Different Strategies

A. Mass spectrum scan



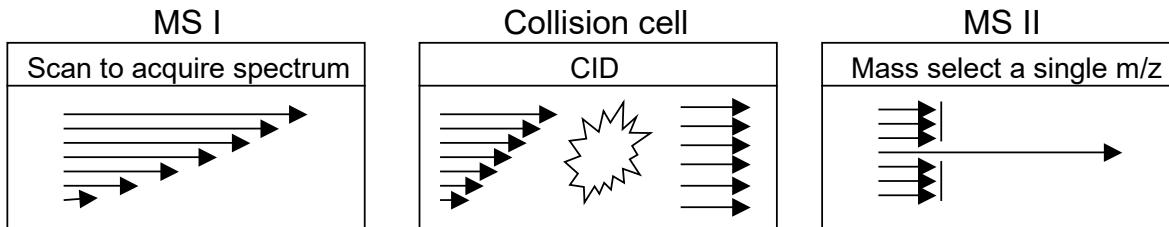
DDA-Data Dependent Analysis.

B. Product ion scan



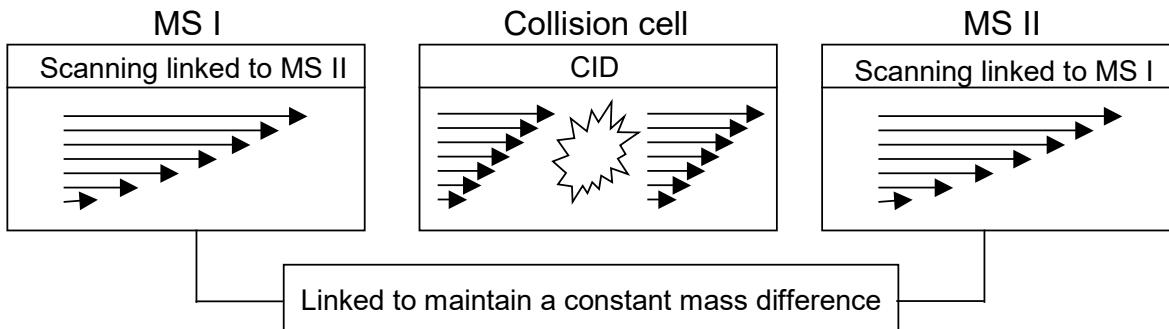
A scan which determines, in a single experiment, all the product ion m/z that are produced by the reaction of a selected precursor ion.

C. Precursor ion scan



A scan which determines, in a single experiment, all the precursor ion m/z that react to produce a selected product ion m/z (sometimes called a 'precursor ion scan').

D. Neutral loss scan



Protein post-translational modification analysis.

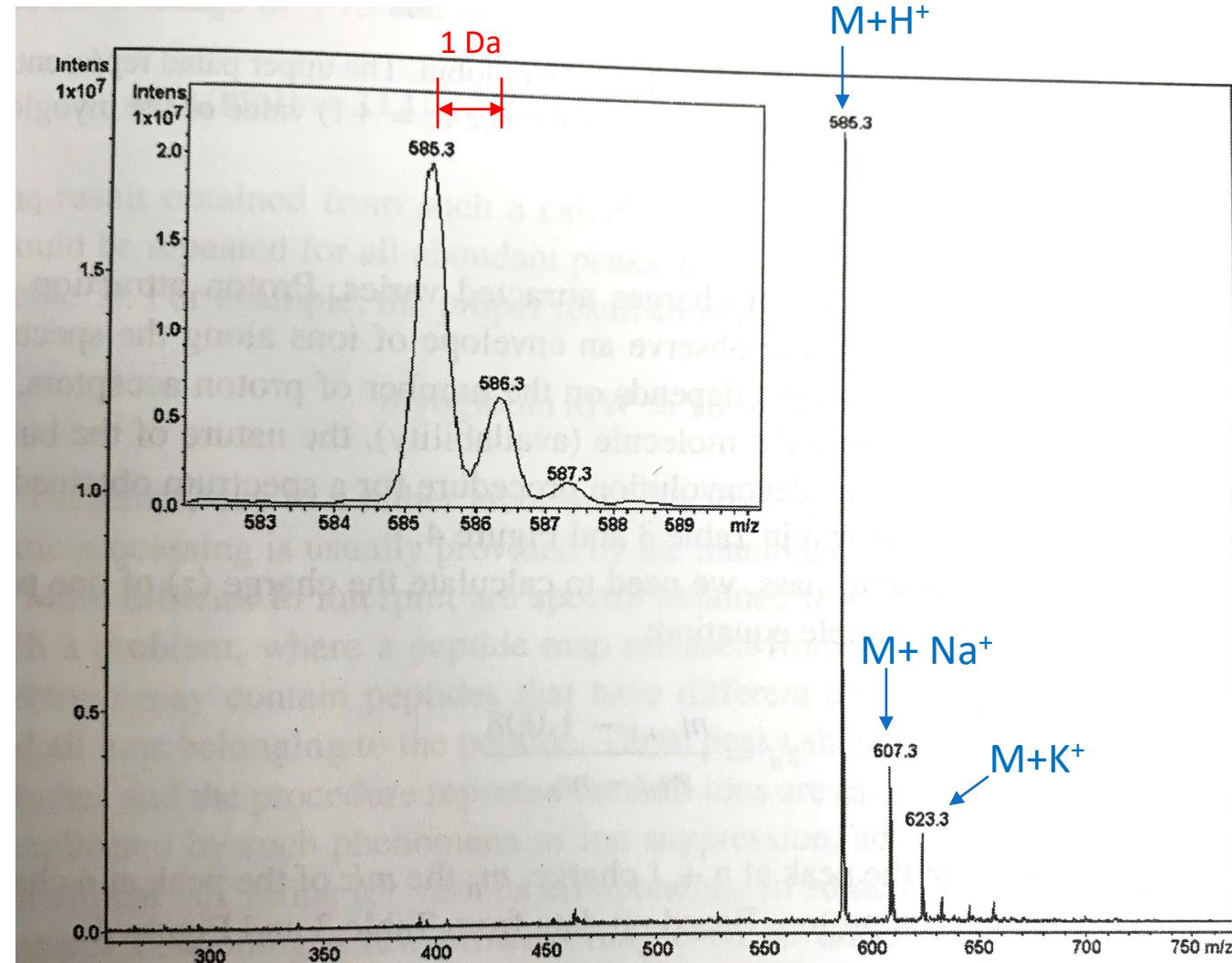
Protein Identification Using Data from MS

1. Peptide mass fingerprinting (aka. Peptide mapping)
2. Fragment ions analysis

Determine protein sequence *de novo* by MS

1. Peptide Mass Fingerprinting (PMF)

- Principle: *each protein can be uniquely identified by the masses of its constituent peptides.*
- A single protein or a simple mixture
 - e.g., a spot on 2D gel or a single LC fraction
- The sample is digested by 1 or more specific cleavage reagent (e.g., trypsin)
- The masses of the peptides are determined, usually by **MALDI-TOF**.
- Search databases for correlative searching
- The algorithm carries out a virtual digest of each protein in the databases with the same specific cleavage and calculate the theoretical peptide masses
- The algorithm attempts to correlate the theoretical masses with the experimentally determined ones. Then Rank proteins from the database in order of best correlation (number of peptides matched).

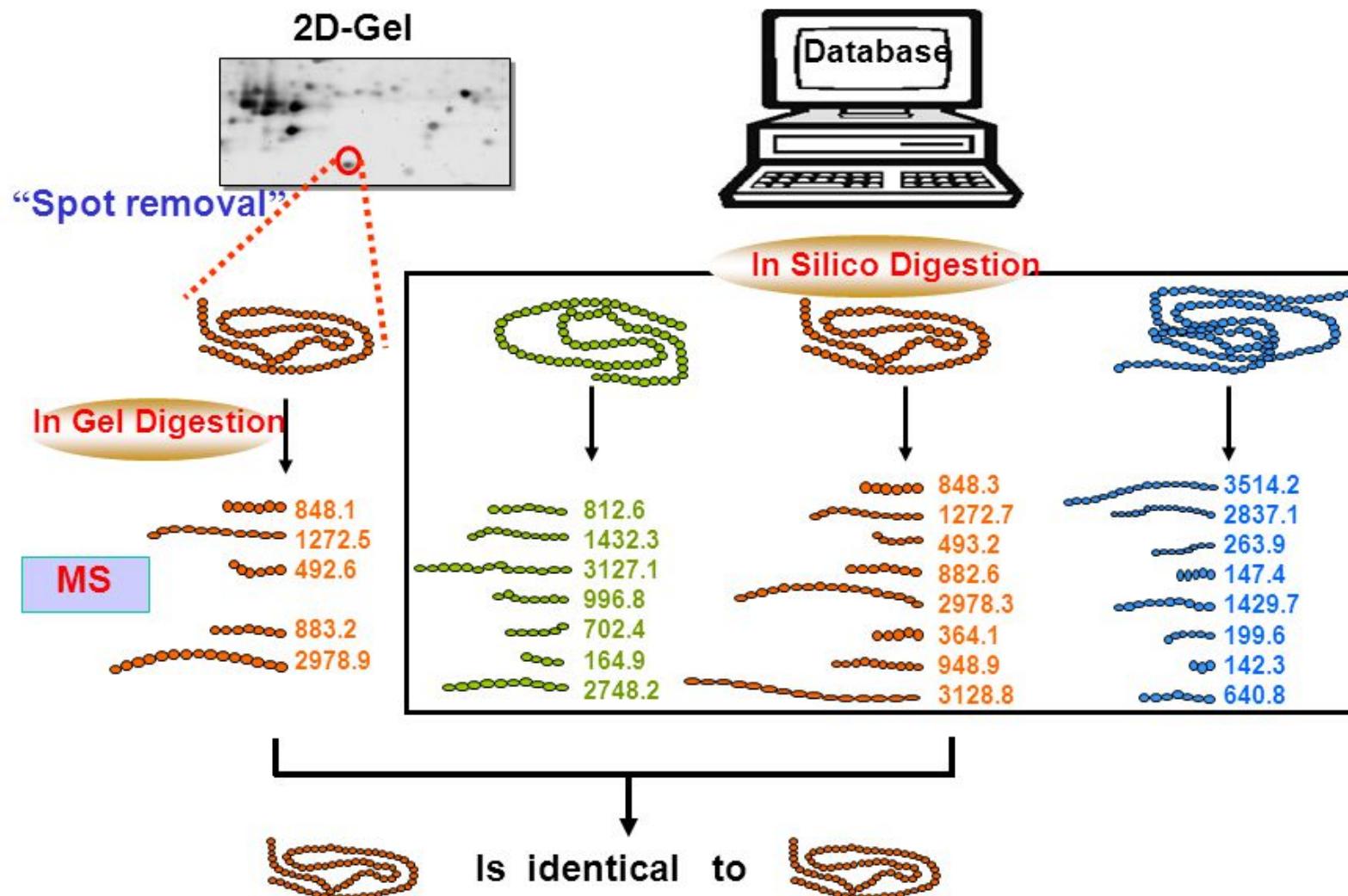


Mass spectrum of the peptide FGGFTG.
The MW of the peptide is 584.3.

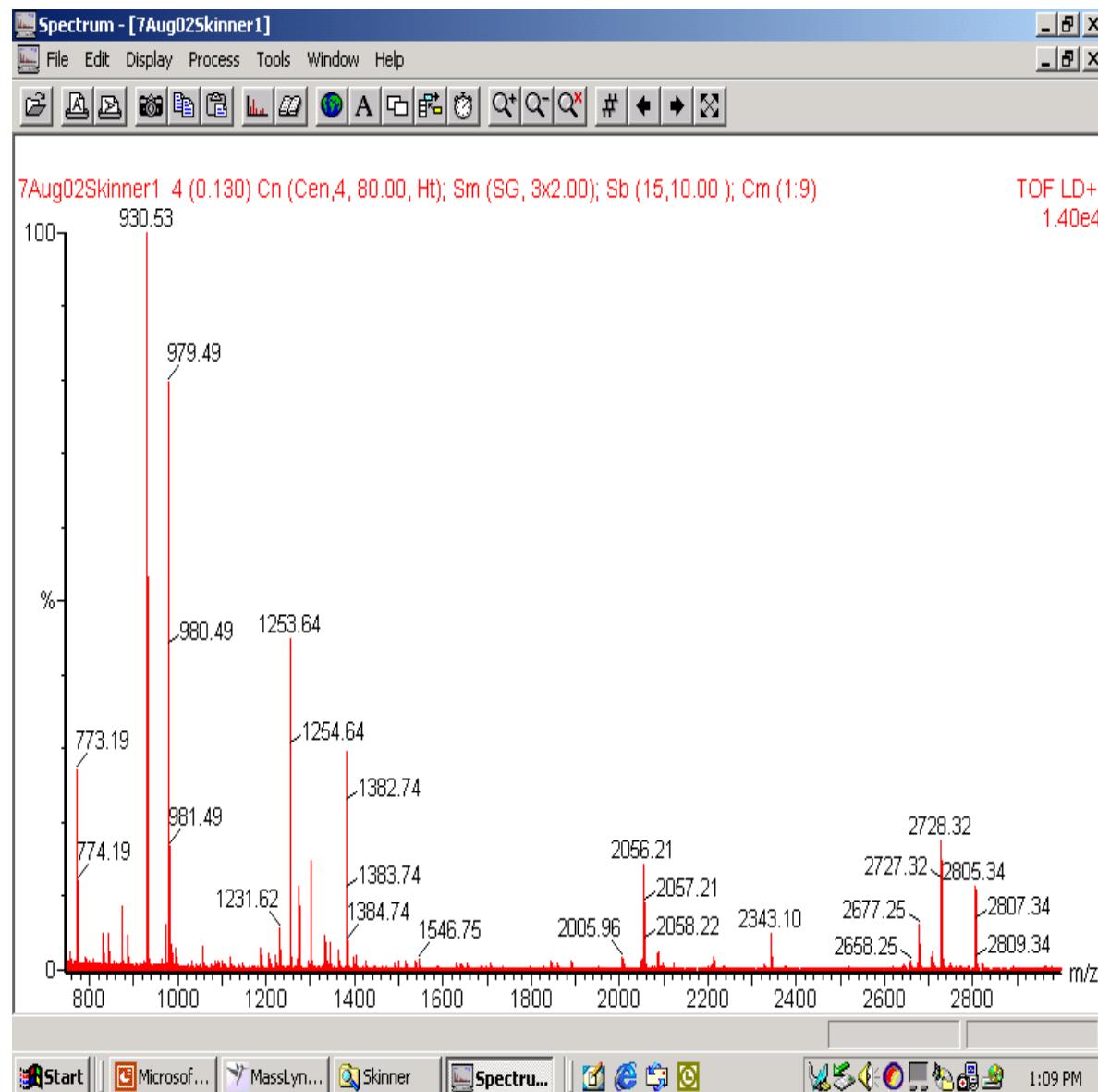
- Proteins and peptides are generally analyzed in the positively mode, which favors production of protonated ($M+H^+$) ions
- In the presence of some buffer ions, sodium ($M+ Na^+$), potassium ($M+K^+$), and ammonium ($M+ NH_4^+$) adducts may be formed.

Website for peptide mass calculation:
https://web.expasy.org/peptide_mass/

Peptide Mass Fingerprinting



MALDI-MS Database Searching



Peaklist

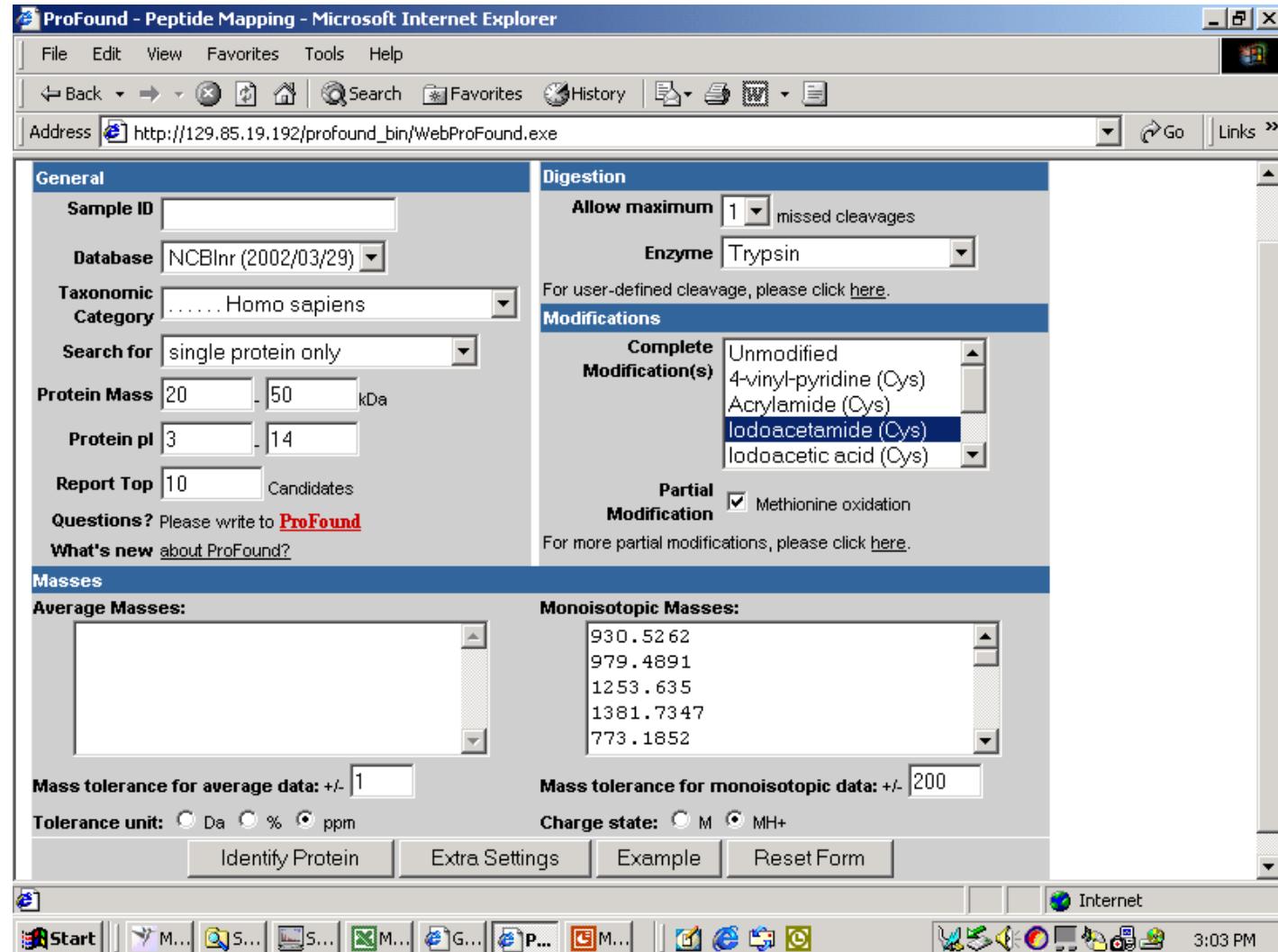
Microsoft Excel - 7Aug02Skinner1

A1		
A	B	C
1 755.0297	164.5261	
2 757.2428	346.1629	
3 762.2136	92.4608	
4 763.0082	119.2401	
5 765.0125	182.3788	
6 771.0146	116.7579	
7 773.1852	3785.743	
8 779.0749	104.9448	
9 781.0849	152.6307	
10 787.0667	106.6553	
11 789.3287	208.7667	
12 791.1122	230.9857	
13 797.1398	107.4736	
14 799.1335	189.2312	
15 805.1334	74.6656	
16 806.1569	121.7381	
17 809.1818	208.7286	
18 815.204	138.5067	
19 817.199	148.6337	
20 823.2642	97.0401	
21 825.2269	131.0409	
22 831.4469	683.3651	
23 837.2399	116.6338	
24 842.5099	224.8589	
25 844.4847	683.1419	

7Aug02Skinner1_mowse Ready

Start Ma... Skinner Sp...

Web-based ProFound Search Engine



The majority of the available search engines allow one to define certain experimental parameters to optimize a particular search.

- Minimum number of peptides to be matched
- Allowable mass error
- Monoisotopic versus average mass data
- Mass range of starting protein
- Type of protease used for digestion
- Information about potential protein modification, such as N- and C-terminal modification, carboxymethylation, oxidized methionine, etc.

ProFound Search Result

SBC9D70B9-05F8-73C2FFD8 - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites History Print Copy Paste Find Links

Address http://129.85.19.192/profound_results/SBC9D70B9-05F8-73C2FFD8.html Go Links

Reprints of our posters at the 48th ASMS Conference are available in PDF format ([ThPB096](#), [ThPB097](#)).

ProFound - Search Result Summary

Version 4.10.5
The Rockefeller University Edition

Protein Candidates for search SBC9D70B9-05F8-73C2FFD8 [88967 sequences searched]

Rank	Probability	Est'd Z	Protein Information and Sequence Analyse Tools (T)	%	pI	kDa	⊕
+1	1.0e+000	2.29	T e 2781338 pdb 1HWGA Chain A, 1:2 Complex Of Human Growth Hormone With Its Soluble Binding Protein	63	5.3	22.34	⊕
+2	2.7e-026	-	T e 10047120 ref NP_061330.1 (NM_018842) insulin receptor tyrosine kinase substrate [Homo sapiens]	18	8.9	45.90	⊕
+3	3.6e-028	-	T e 14760678 ref XP_043920.1 (XM_043920) hypothetical protein FLJ10630 [Homo sapiens] Link to NCBI	14	6.0	31.51	⊕
4	4.0e-030	-	T e 10434501 dbj BAB14278.1 (AN00ZZZ002) unnamed protein product [Homo sapiens]	15	11.9	49.80	⊕
+5	1.6e-030	-	T e 14745041 ref XP_051082.1 (XM_051082) hypothetical protein FLJ12820 [Homo sapiens]	10	9.2	49.53	⊕
+6	5.1e-031	-	T e 364011 prf 1506383A apolipoprotein E mutant E3K [Homo sapiens]	20	6.0	36.25	⊕
7	4.7e-031	-	T e 14767250 ref XP_030233.1 (XM_030233) hypothetical protein XP_030233 [Homo sapiens]	17	8.9	47.73	⊕
8	2.2e-031	-	T e 1688260 gb AAB36943.1 (U78045) metalloelastase [Homo sapiens]	23	9.3	28.39	⊕
9	2.1e-031	-	T e 386971 gb AAA60387.1 (M2S142) myosin heavy chain alpha-subunit [Homo sapiens]	16	5.7	45.25	⊕
10	1.5e-031	-	T e 13646923 ref XP_011188.2 (XM_011188) actin filament associated protein [Homo sapiens]	18	9.6	33.10	⊕

NOTE:

1. To search again using ~~selected masses~~, click the symbol ⊕.
2. Highly similar protein sequences were given the same rank (E.g.: click '+' to expand/collapse).

http://www.ncbi.nlm.nih.gov/htbin-post/Entrez/query?form=6&db=p&Dopt=g&uid=14760678

Start M... S... S... M... G... S... M... Internet

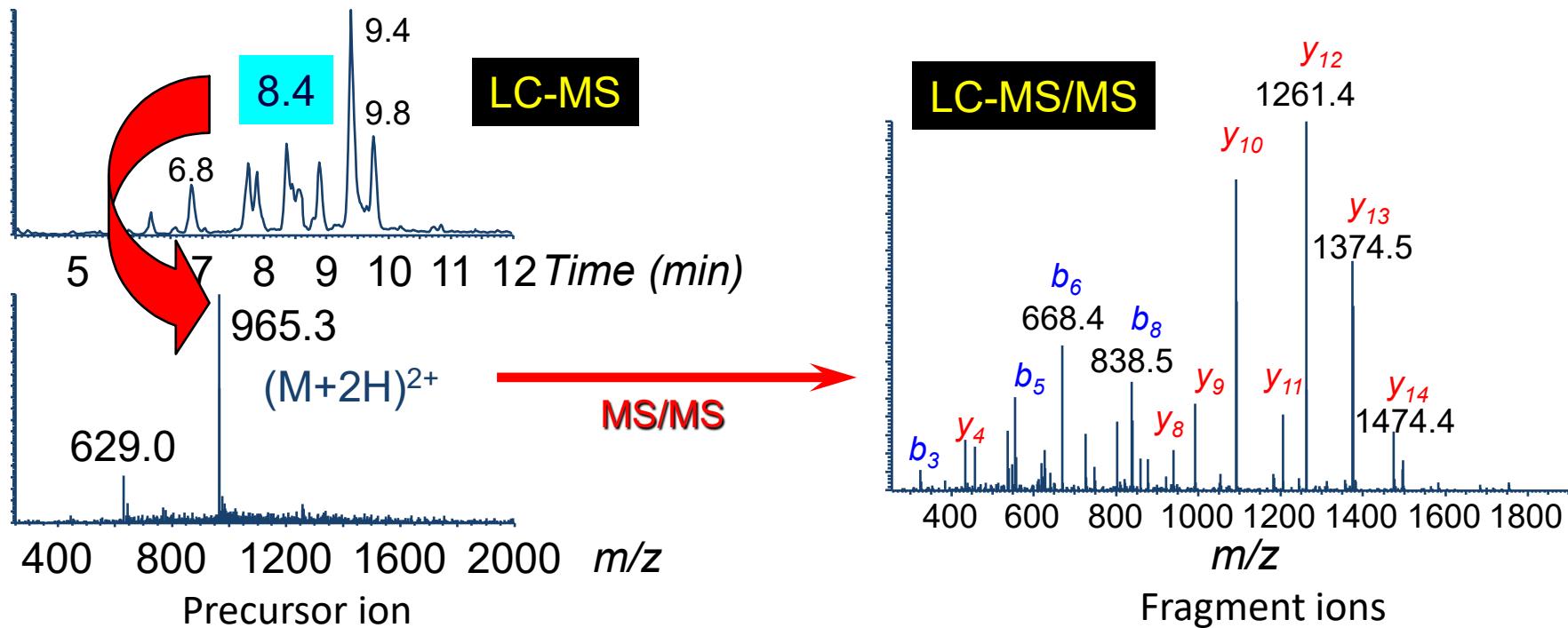
3:04 PM

Searching with Peptide Mass Fingerprints (PMF): Limitations

- Most protein databases contain **primary sequence information only**
 - Any shift in mass incorporated into the primary sequence as a result of **post-translational modification, amino acids substitution** will result in an experimental mass that is in *disagreement with the theoretical mass*, even a protein with a great deal of homology in the database can not be identified.
- **Non-specific cleavage, isobaric peptide** (same mass but different order), etc.

LC-MS/MS for Protein Identification

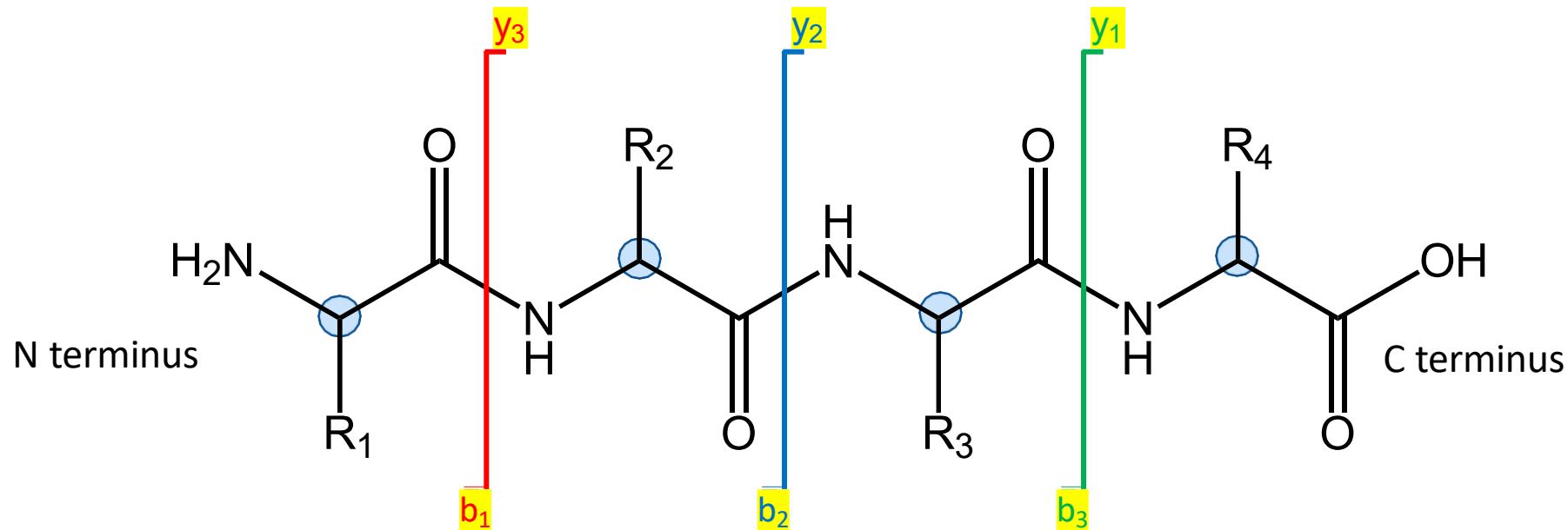
- An improvement in throughput of the overall method can be obtained by performing LC-MS/MS in the data-dependent mode.
 - As full scan mass spectra are acquired continuously in LC-MS mode, any ion detected with a signal intensity above a pre-defined threshold will trigger the mass spectrometer to switch over to MS/MS mode. Thus, the mass spectrometer switches back and forth between MS (molecular mass information) and MS/MS mode (sequence information) in a single LC run.



The data-dependent scanning capability can dramatically increase the capacity and throughput for protein identification.

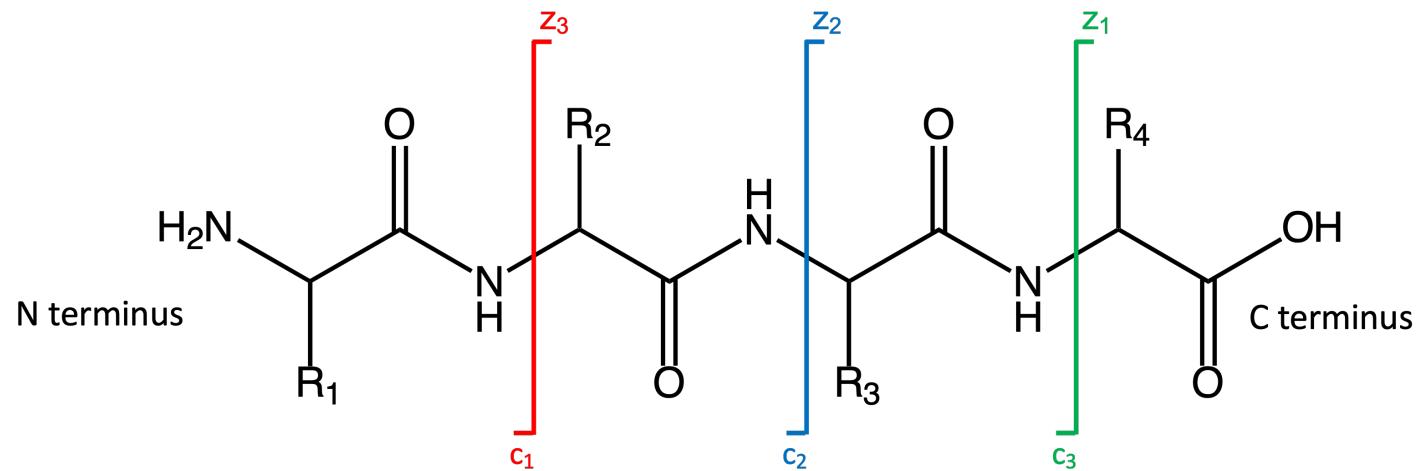
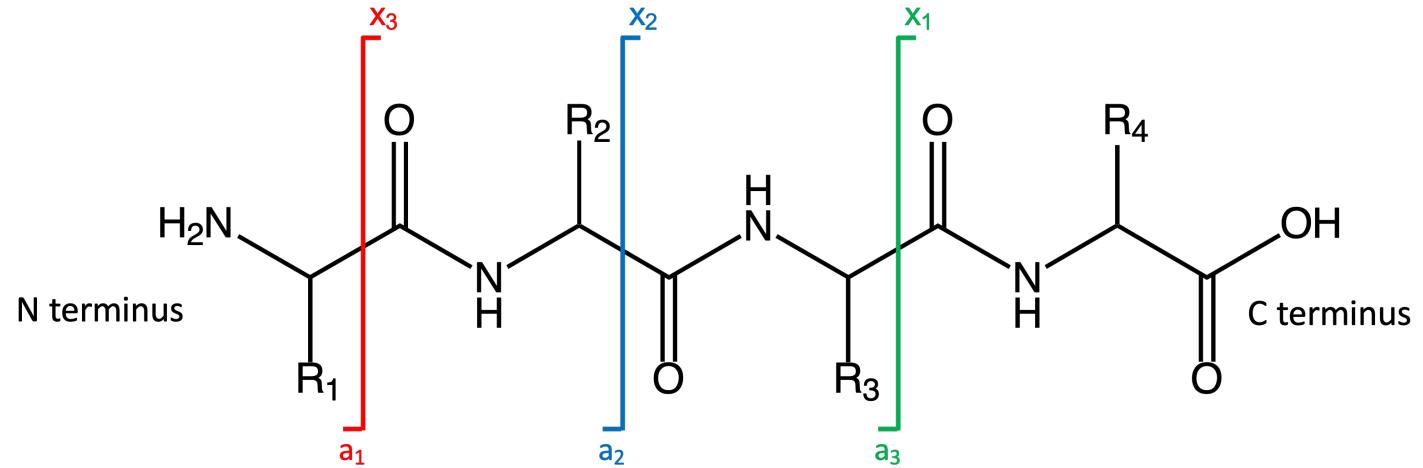
2. Fragment Ion Analysis₁

- Peptide can be fragmented by collision-induced dissociation (CID) (and other methods)
 - Collisions with neutral inert gas molecules (nitrogen, argon, etc.)
 - Charge stays on *either* the ‘left’ (a, b, or c) or ‘right’ (x, y, or z) side of cleavage
 - *Cleavage along the CO-NH bond is most common, generating ‘b’ and ‘y’ ions*



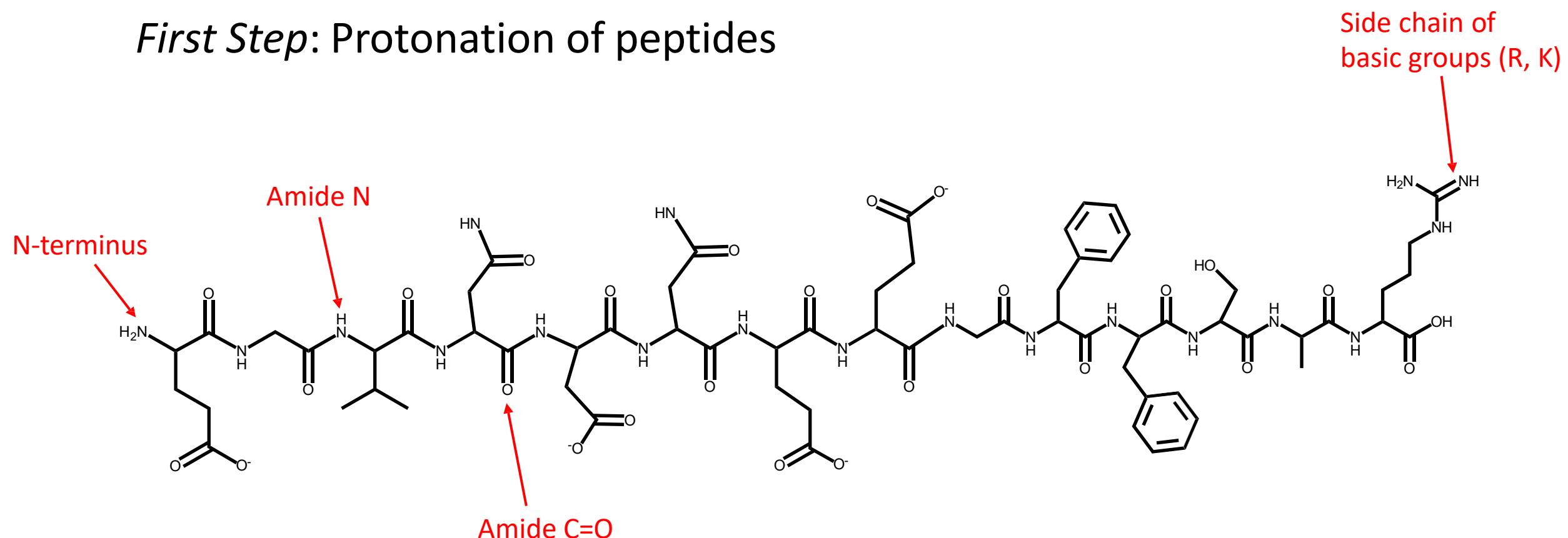
- Letter: Indicates the bond broken and the terminus contained in the fragment
- Number: Indicates the number of C α in the fragment

2. Fragment Ion Analysis₂



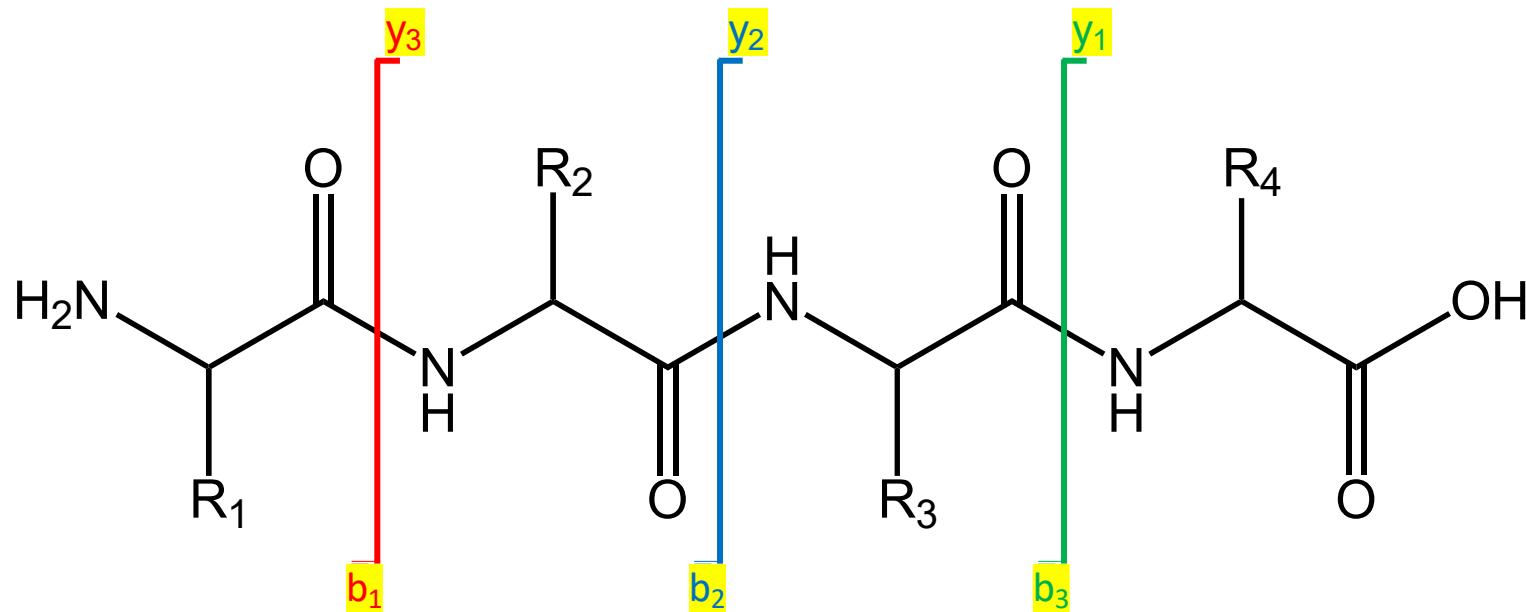
Peptides Fragment by CID

First Step: Protonation of peptides

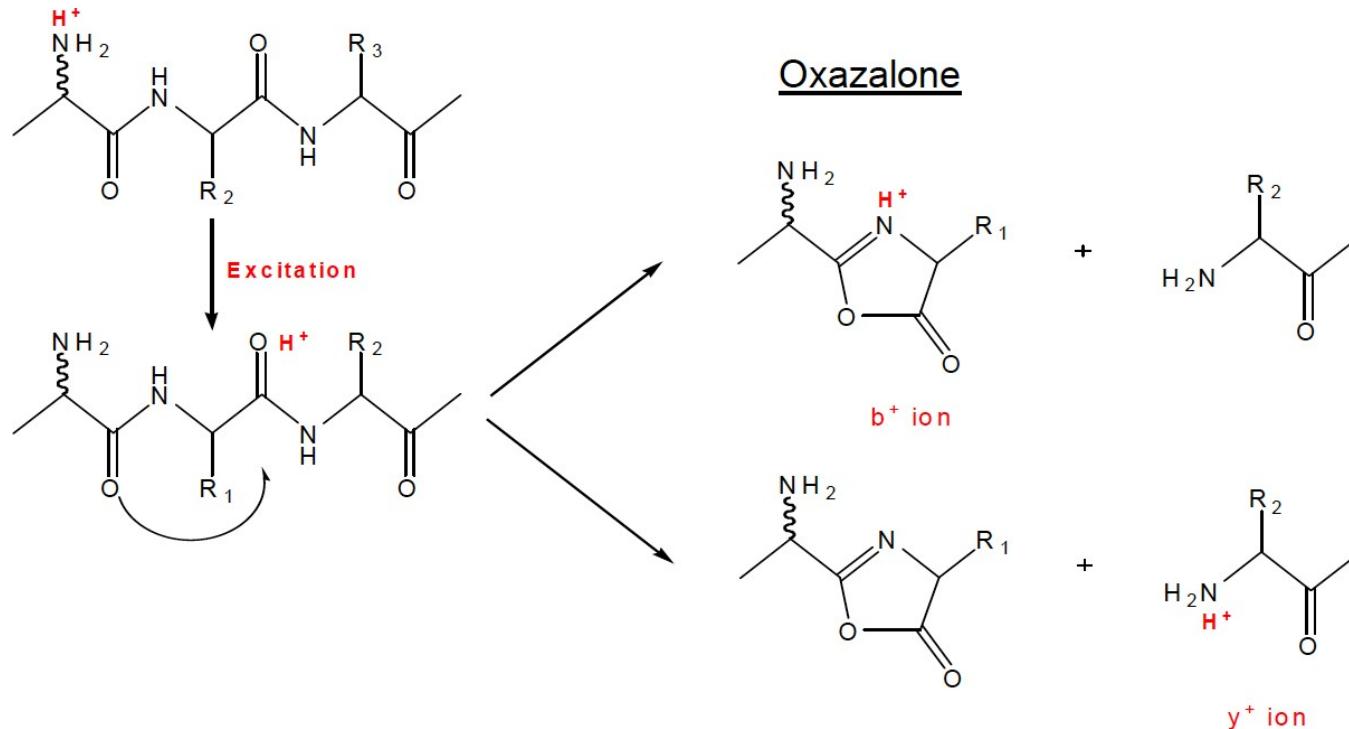


Peptides Fragment by CID

Second step: Cleavage along the CO-NH bond is most common, generating **b** and **y** ions



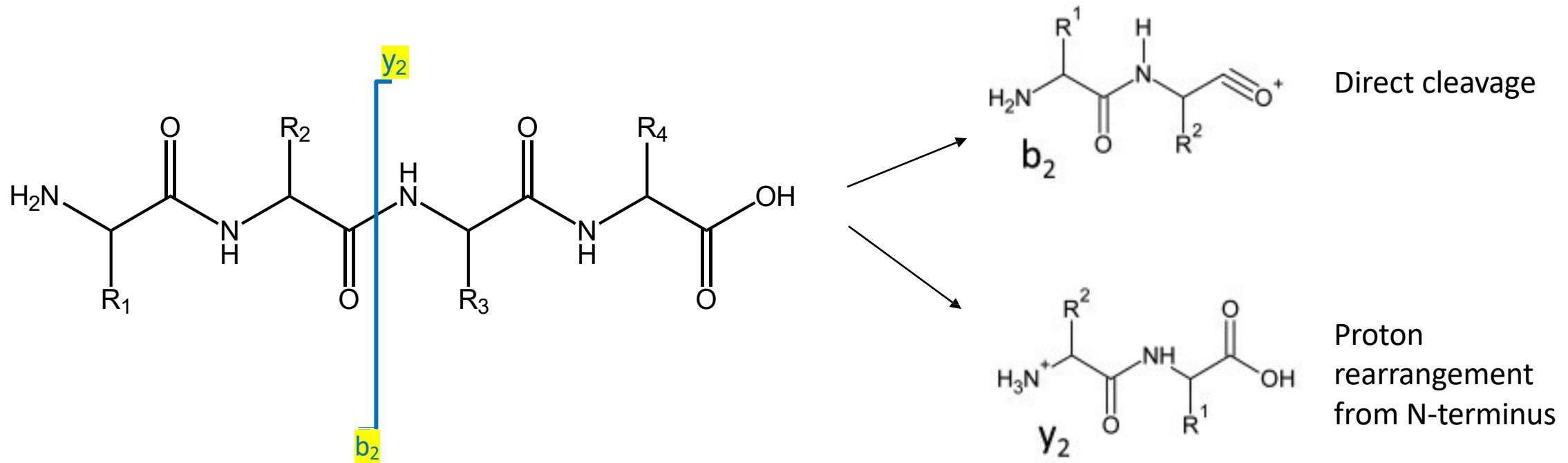
Peptides Fragment by CID



For a singly protonated peptide,

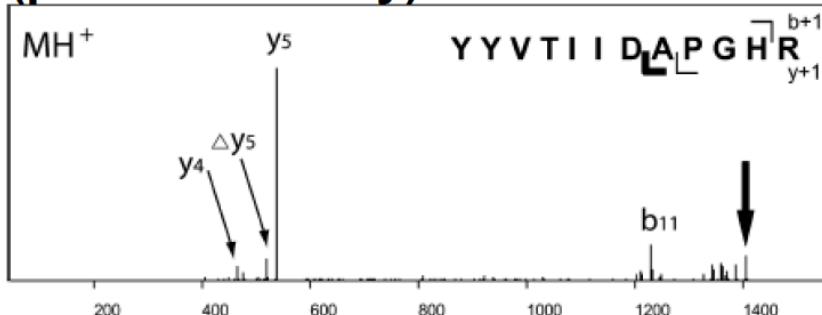
Singly charged N term ion ($+H^+$) and neutral C-term
OR
Neutral N term and Singly charged C-term ion ($+H^+$)

Peptides Fragment by CID

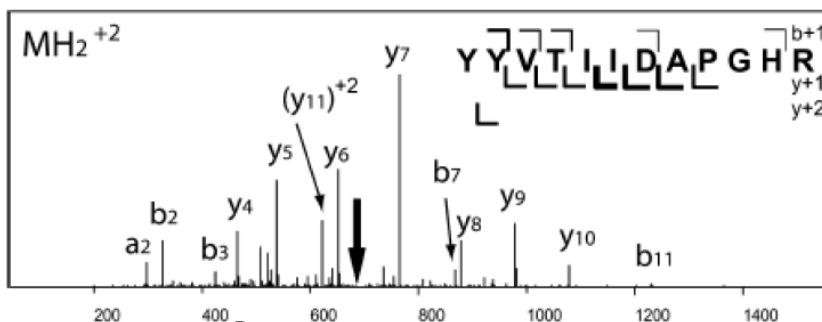


For a doubly protonated peptide, both N- and C-terminal fragments can be generated from a single dissociate event.

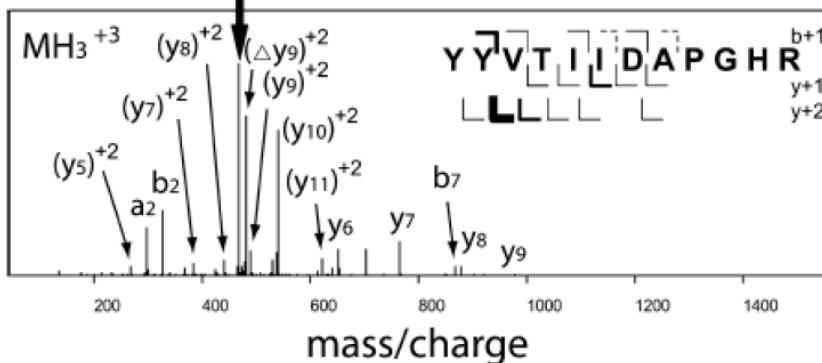
Different Precursor Ion Charge States Have Different Cleavage Patterns



Localized proton, selective fragmentation



Free proton, non- or less selective fragmentation



Free proton, non-selective fragmentation and multiply charged fragments

The Proline Effect in Fragmentation – Cleavage Favored N-terminal to Pro

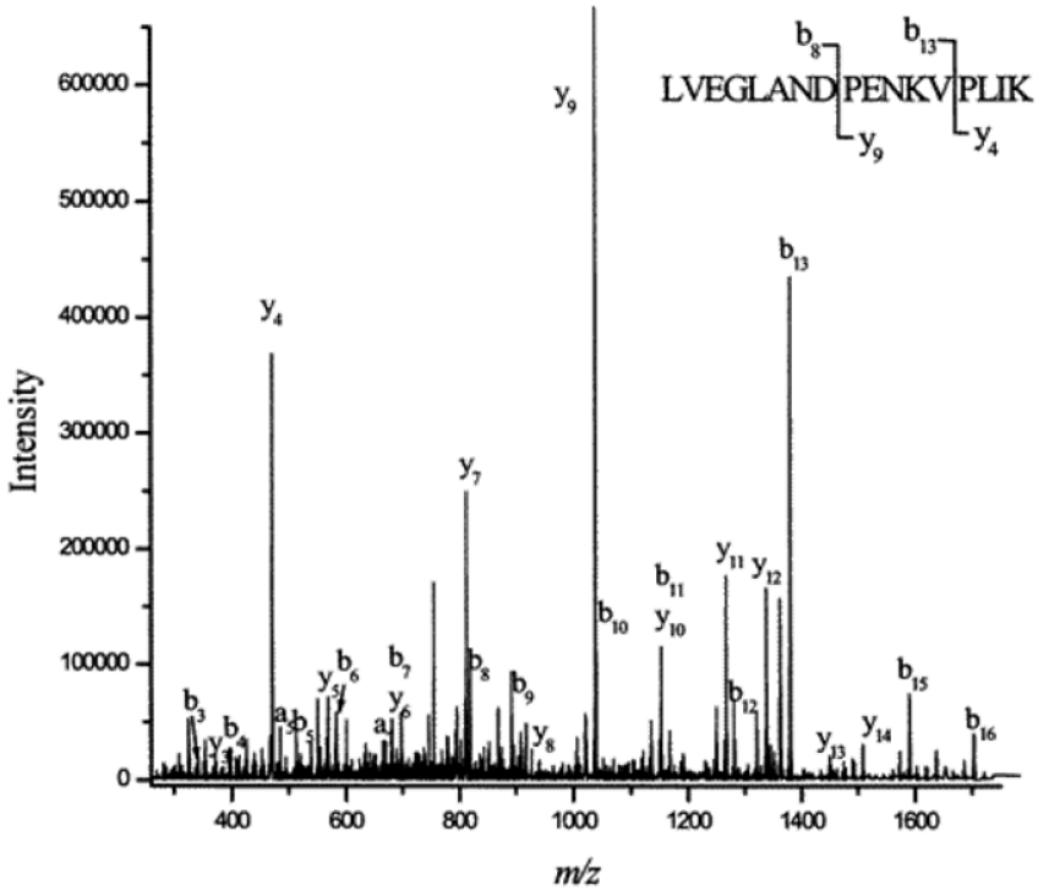
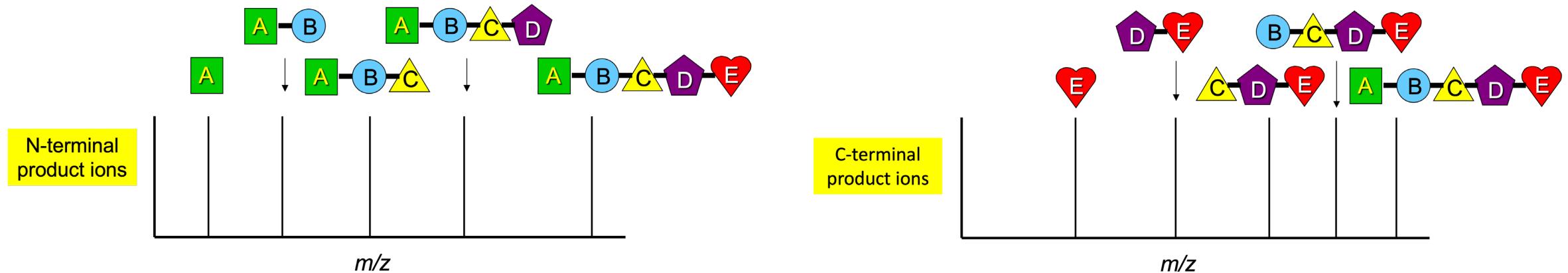


Figure 1 MS/MS spectrum of the peptide [LVEGLANDPENKVPLIK + 2H] $_2$ + acquired by CID in an ion trap. Although many peaks are a-, b-, and y-sequence ions, many other peaks are unidentified.

Peptide Sequencing



- Ideally, one can measure the spacings between product ion peaks to deduce the sequence
 - if each amide bond dissociates with equal probability
 - if only a single amide bond fragments for each molecule
 - if only C-terminal or N-terminal products ions are formed
- In reality, this is not the case...

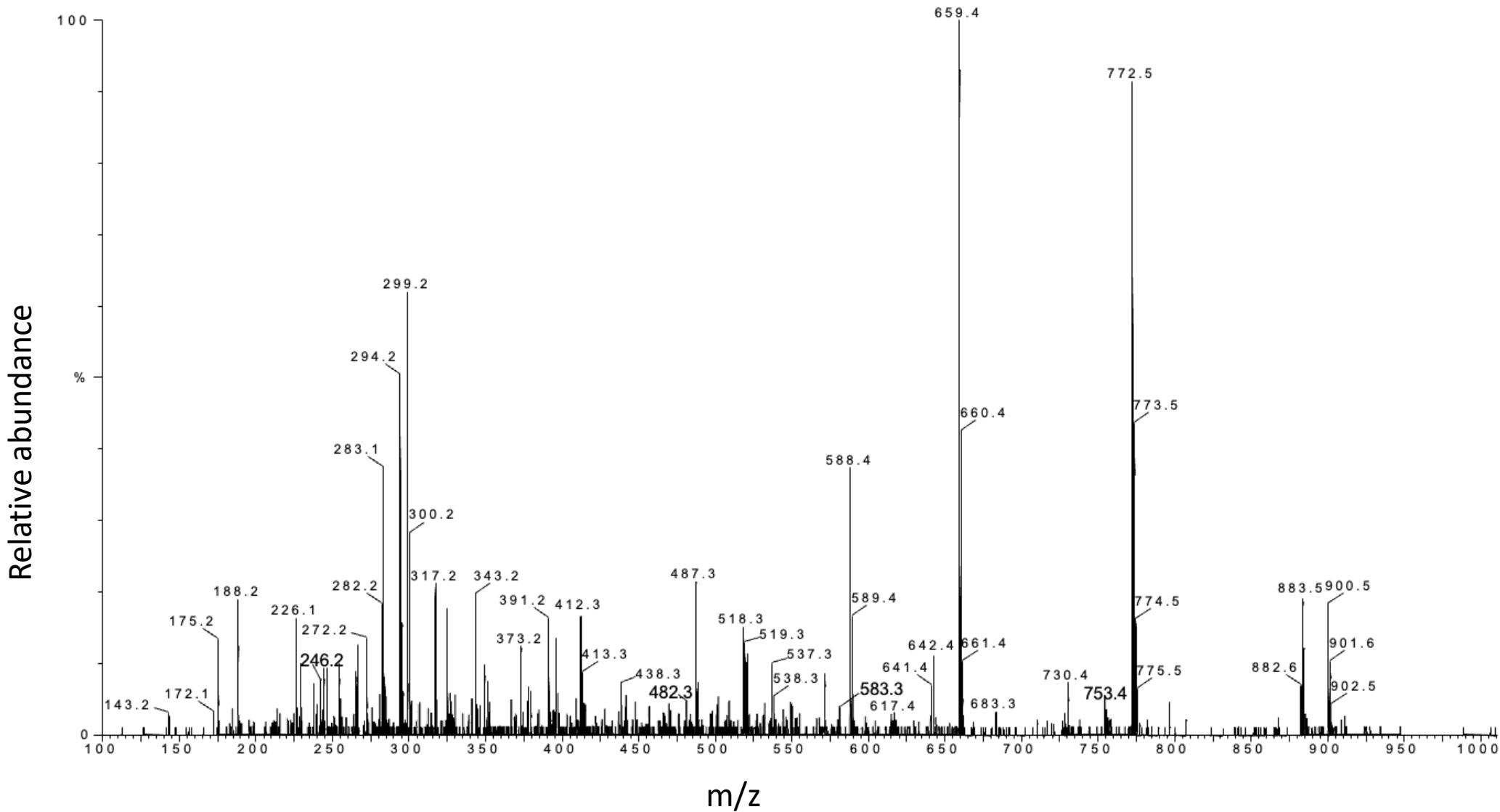


Fragmentation Results in a Peptide “Ladder”

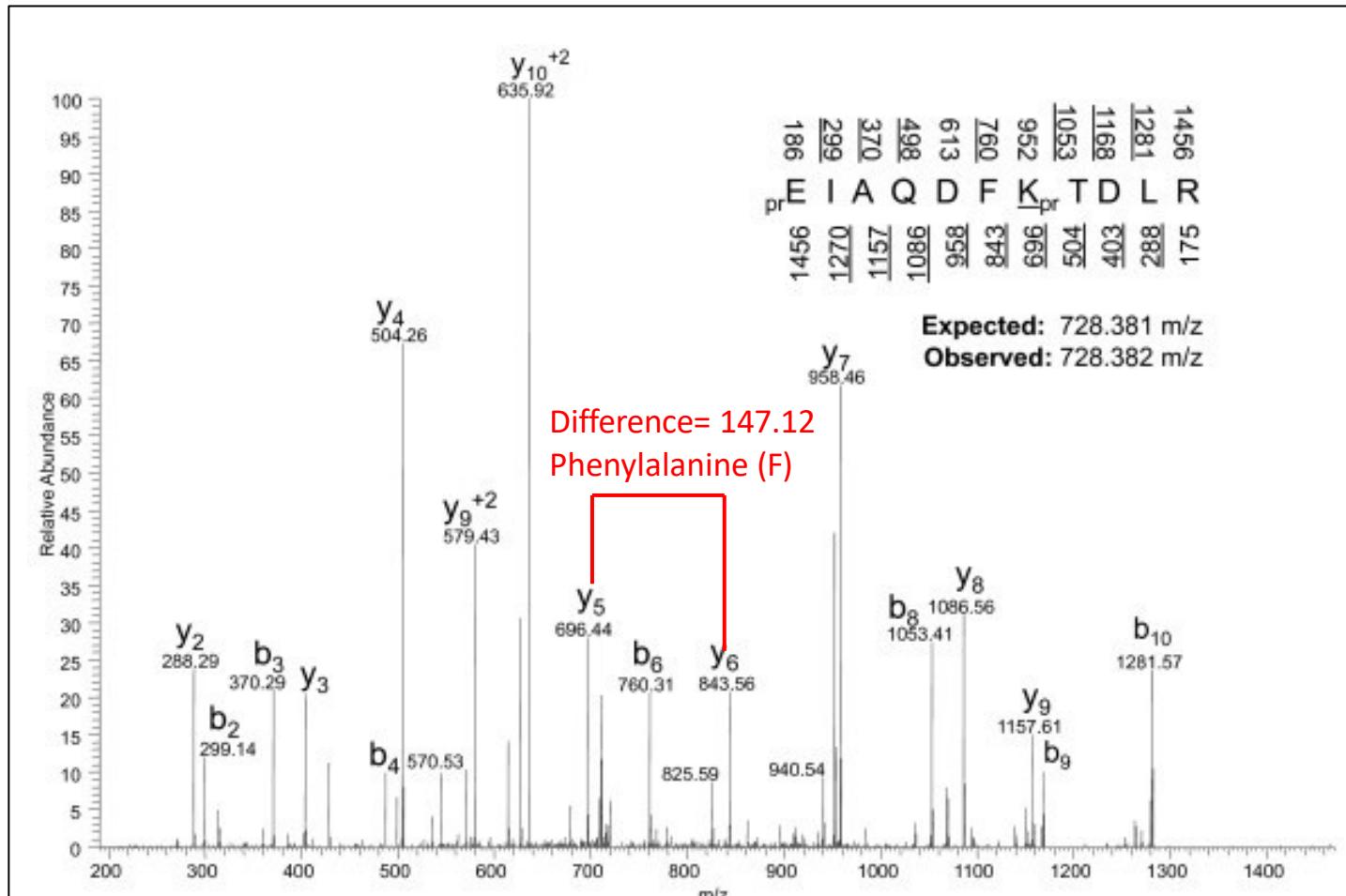
Peptide: A-B-C-D-E

	<u>b-ions</u>		<u>y-ions</u>
b_1^+	A	-----	BCDE y_4^+
b_2^+	AB	-----	CDE y_3^+
b_3^+	ABC	-----	DE y_2^+
b_4^+	ABCD	-----	E y_1^+

MS² Spectrum



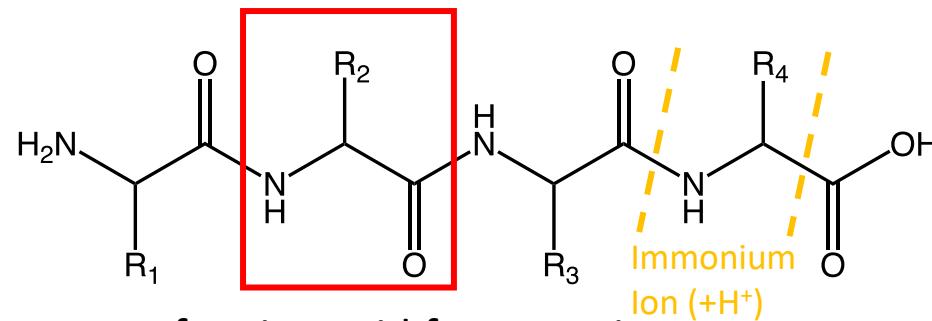
Mass Spectrum (Assignment of *b*- and *y*-ions)



- 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 observed m/z of full peptide is 728.382 Th (It is doubly charged):

$$728.382 \times 2 - 2 = 1454.764 \text{ Da}$$



Mass of amino acid fragment ion

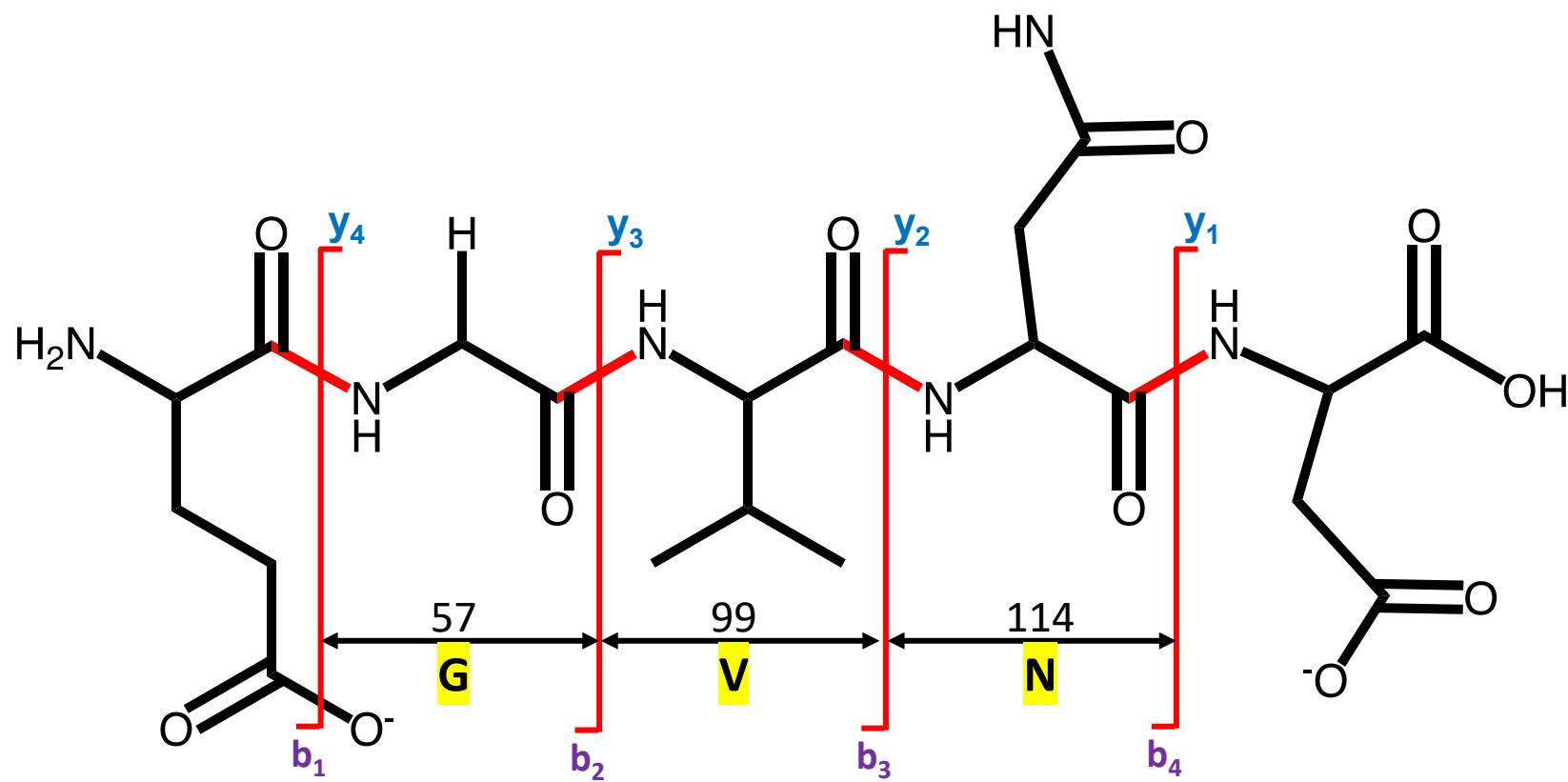
Code (1 letter)	Monoisotopic mass
G	57.021 47
A	71.037 12
S	87.032 03
P	97.052 77
V	99.068 42
T	101.047 68
C	103.009 19
I	113.084 07
L	113.084 07
N	114.042 93
D	115.026 95
Q	128.058 58
K	128.094 97
E	129.042 60
M	131.040 49
H	137.058 91
F	147.068 42
R	156.101 12
Y	163.063 33
W	186.079 32

Mass of **b₂** ions (mass of the two AAs + 1) in peptide fragmentation

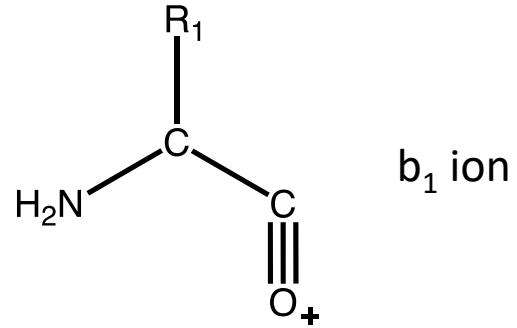
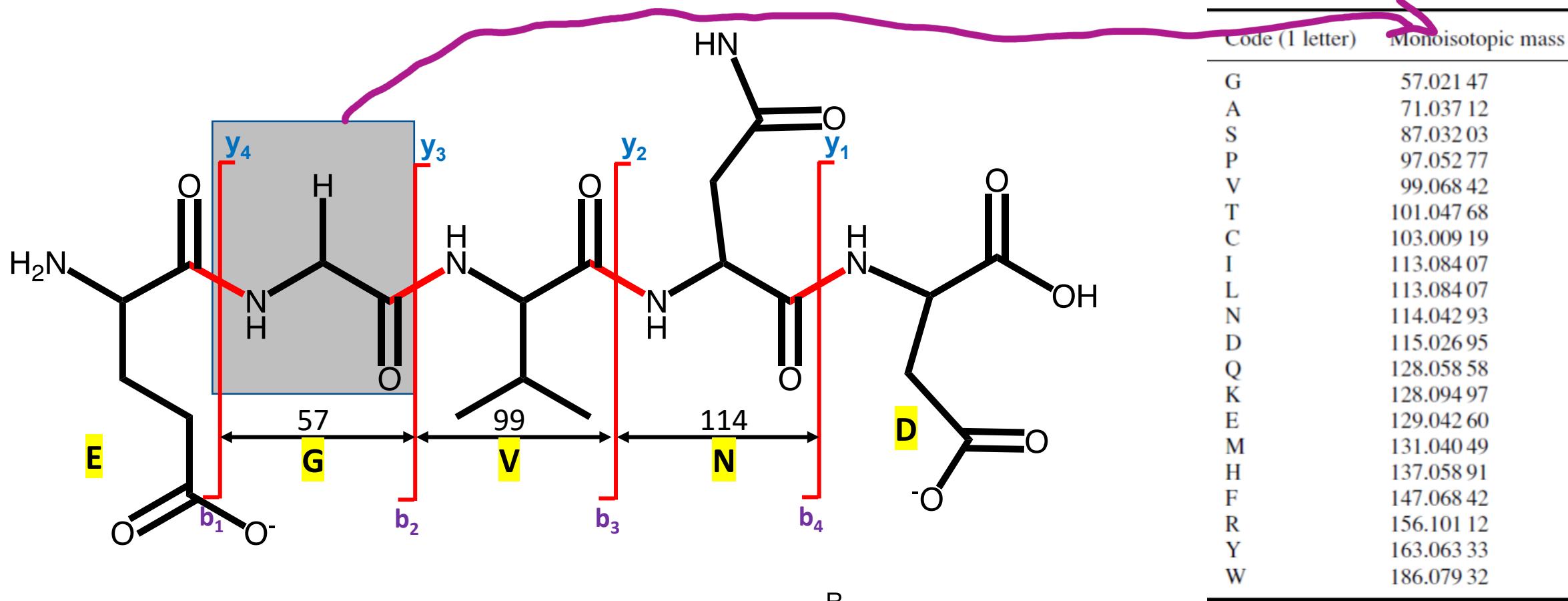
	G	A	S	P	V	T	C	I/L	N	D	K/Q	E	M	H	F	R	Y	W
G	115																	
A	129	143																
S	145	159	175															
P	155	169	185	195														
V	157	171	187	197	199													
T	159	173	189	199	201	203												
C	161	175	191	201	203	205	207											
I/L	171	185	201	211	213	215	217	227										
N	172	186	202	212	214	216	218	228	229									
D	173	187	203	213	215	217	219	229	230	231								
K/Q	186	200	216	226	228	230	232	242	243	244	257							
E	187	201	217	227	229	231	233	243	244	245	258	259						
M	189	203	219	229	231	233	235	245	246	247	260	261	263					
H	195	209	225	235	237	239	241	251	252	253	266	267	269	275				
F ^b	205	219	235	245	247	249	251	261	262	263	276	277	279	285	295			
R	214	228	244	254	256	258	260	270	271	272	285	286	288	294	304	313		
Y	221	235	251	261	263	265	267	277	278	279	292	293	295	301	311	320	327	
W	244	258	274	284	286	288	290	300	301	302	315	316	318	324	334	343	350	
																	373	

GG=N=114; GA=K/Q=128; GV=R=156; GE=AD=SV=W=186.

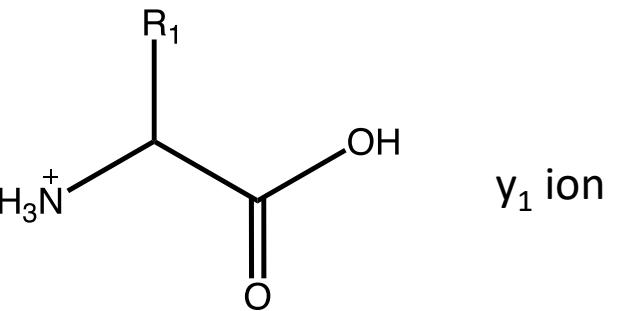
- Amino acid sequence can be deduced by the Δmass between adjacent y ion peaks **or** adjacent b ion peaks



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

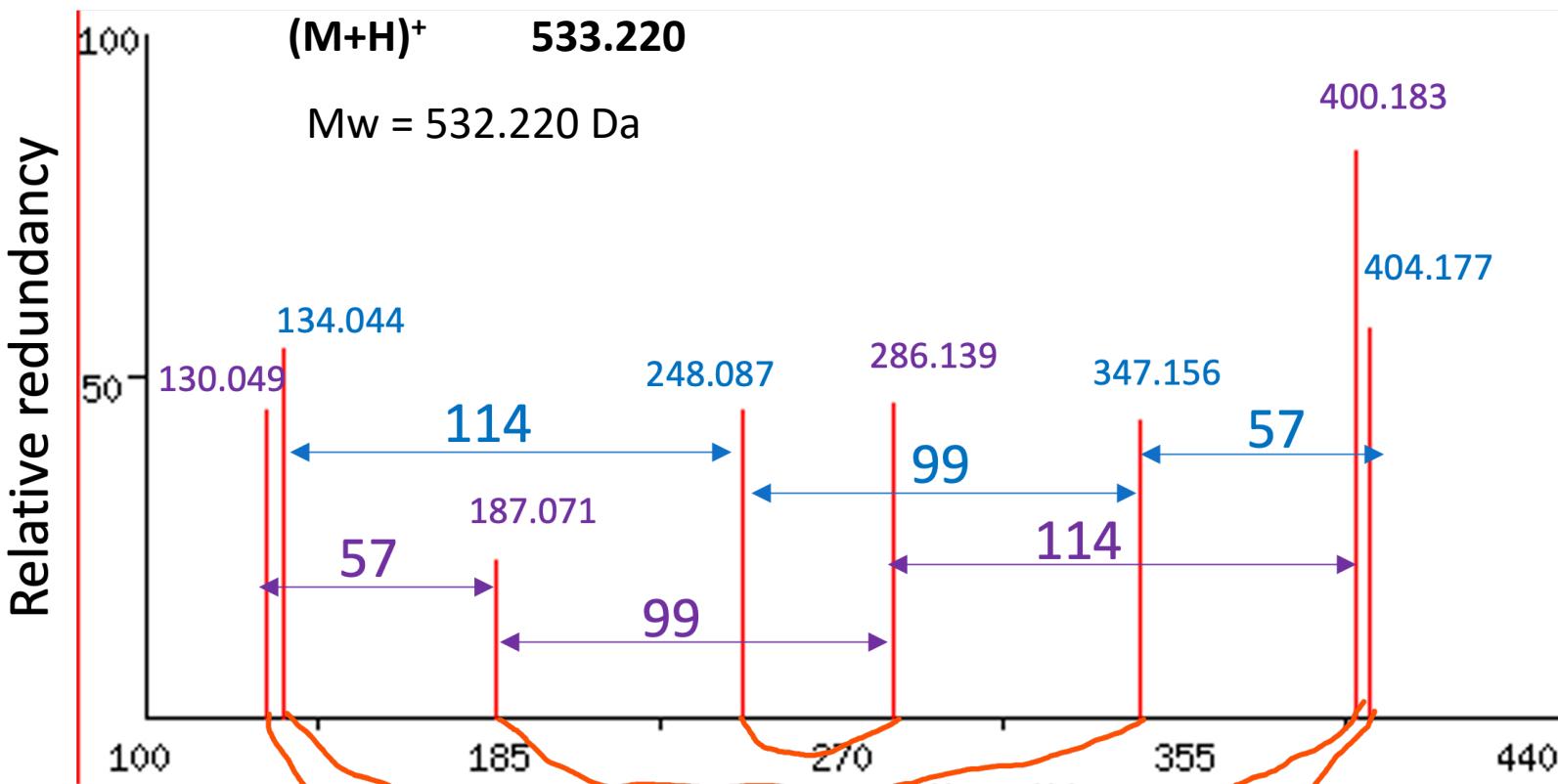


Mass of b-ions = Σ (residue masses) + 1 (H)



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

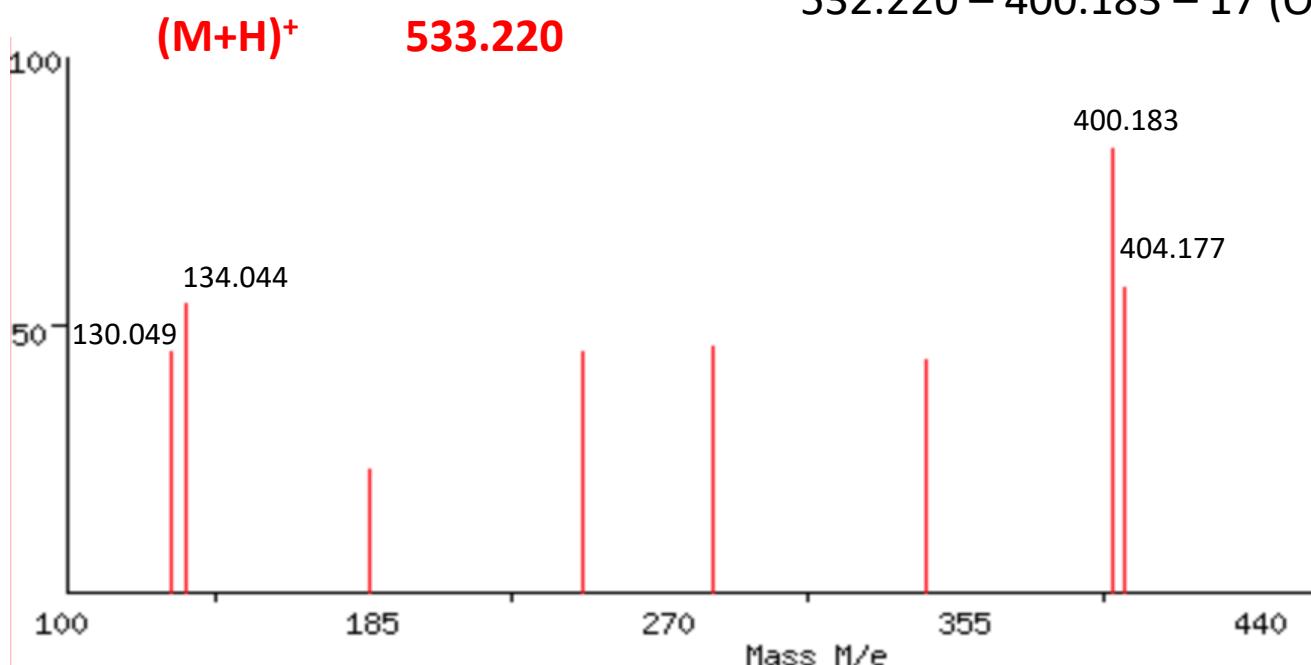
$b_1, y_1, b_{n-1}, y_{n-1}$

$$130.049 - 1 = 129.049$$

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

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

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

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

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

Calculate the Terminal Residues

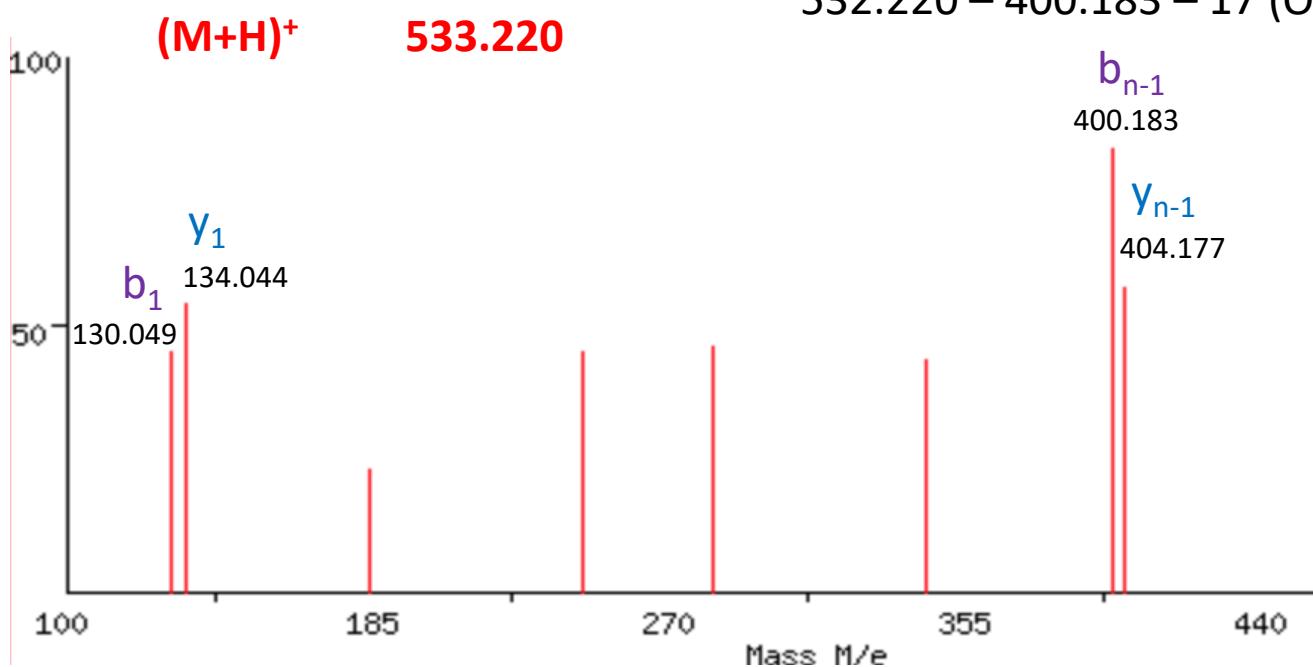
$b_1, y_1, b_{n-1}, y_{n-1}$

$$130.049 - 1 = 129.049$$

$$134.044 - 19 = 115.044$$

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

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

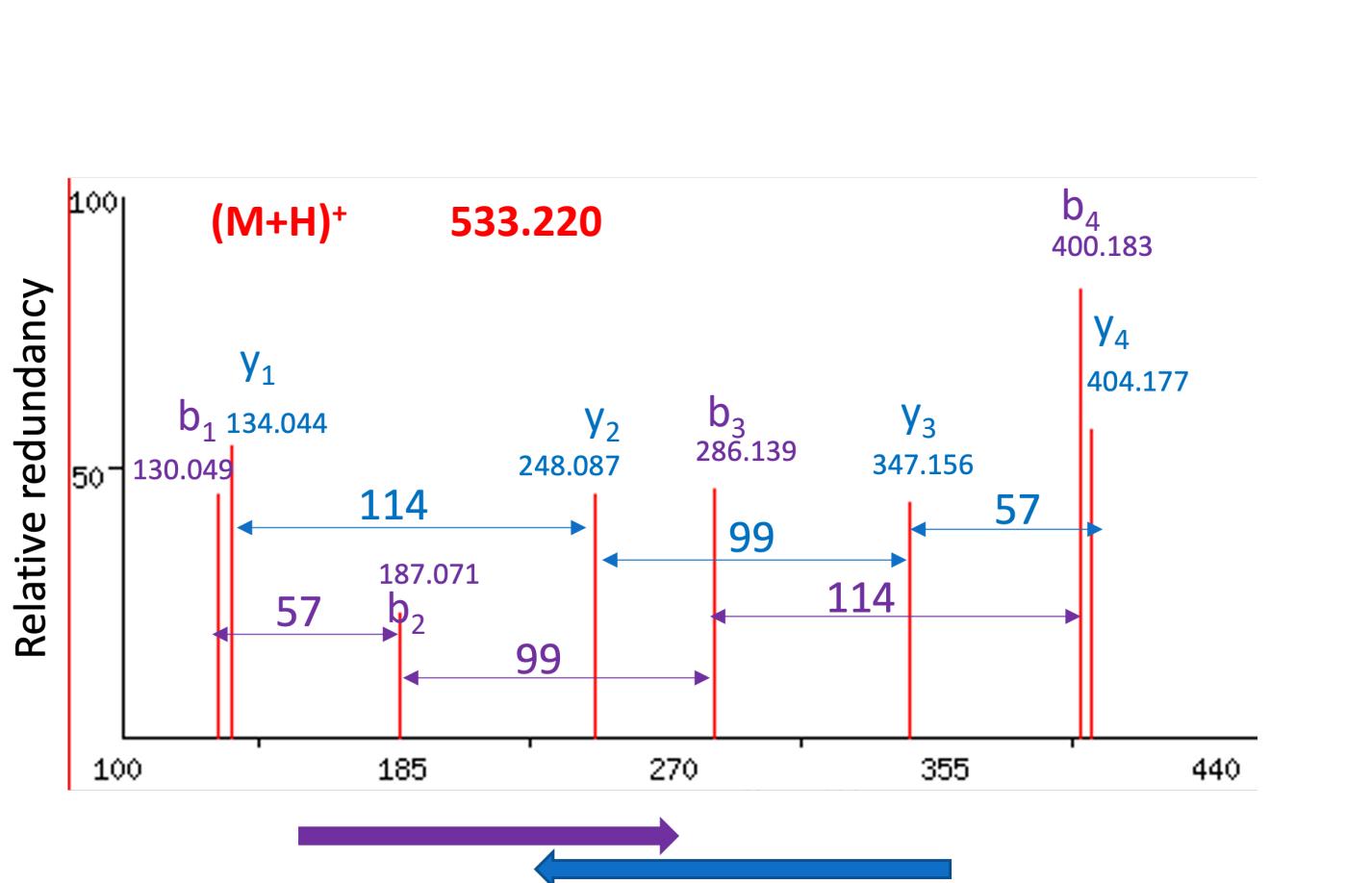
$$532.220 - 400.183 - 17 (\text{OH}) = 115.037 \quad D \text{ on C terminus}$$

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

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

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.

$[M+H]^+ = 1464.7693$

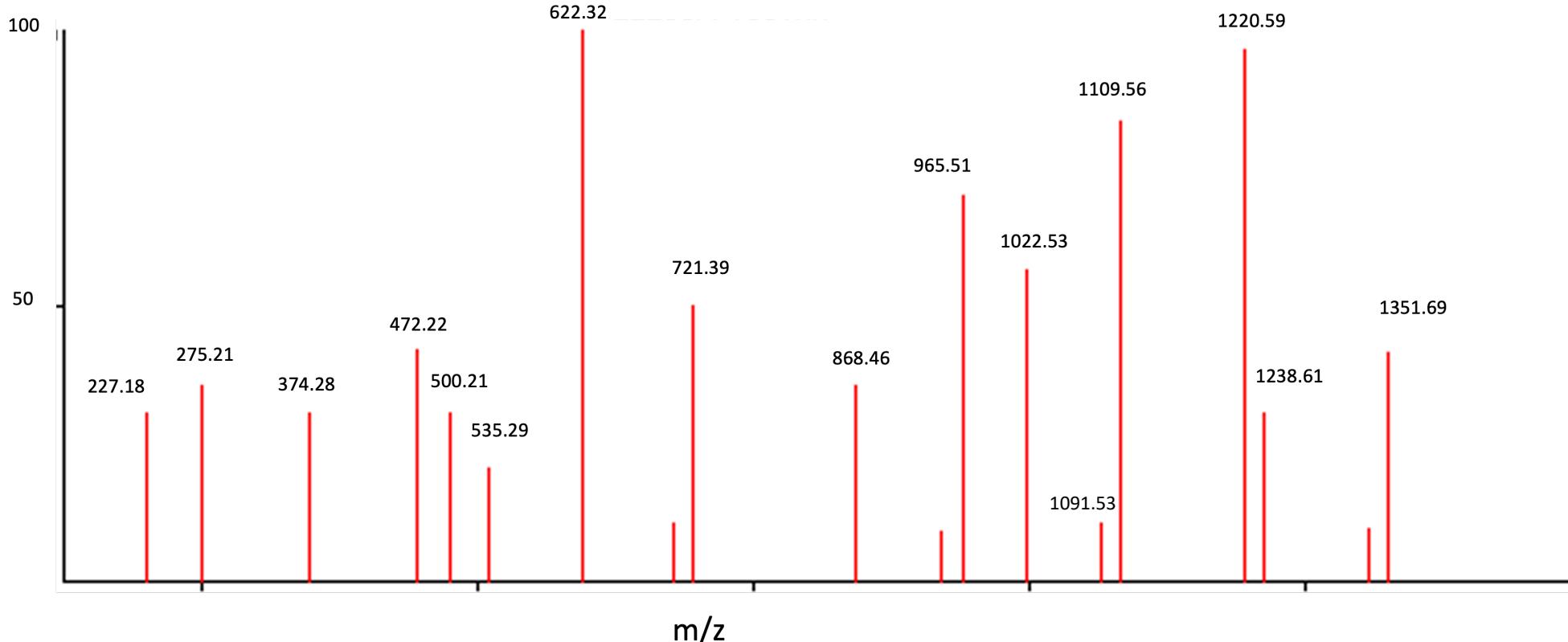
So, Mw = 1463.7693 Da

- 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)

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

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

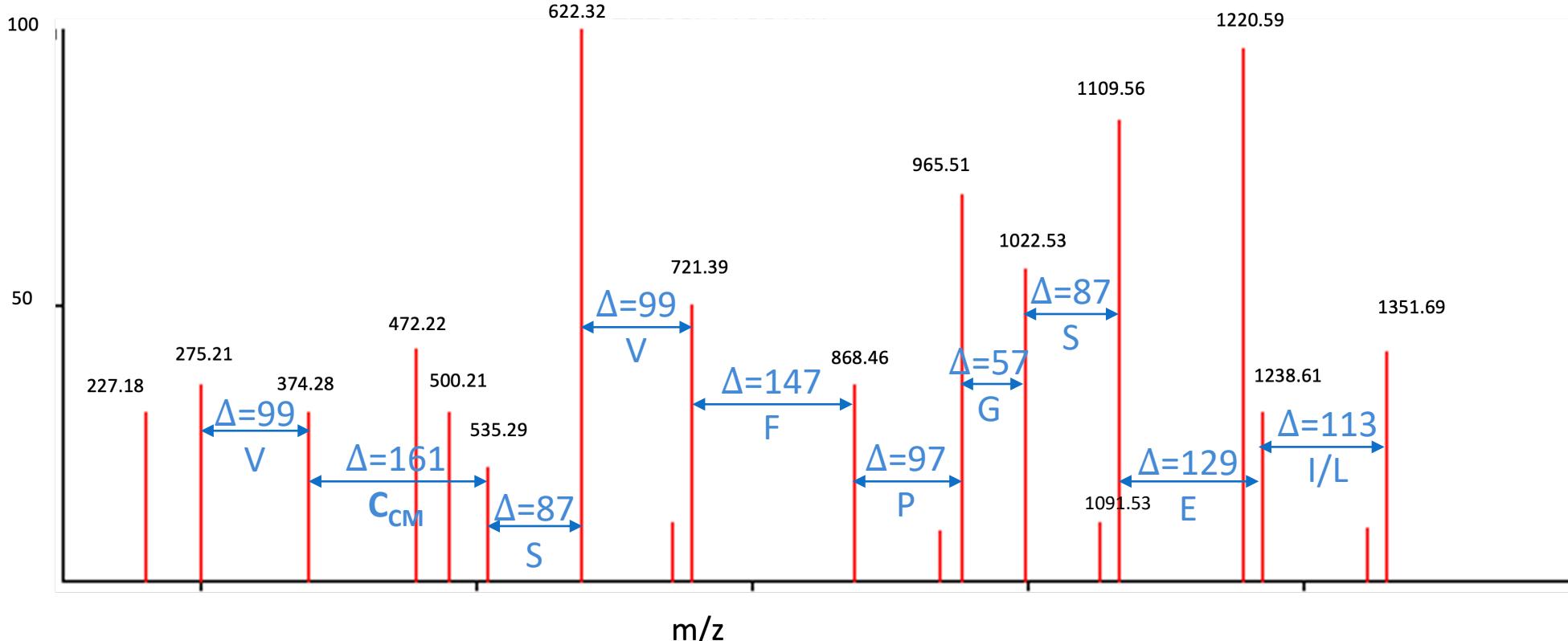
- Amino acid sequence can be deduced by the Δ mass between adjacent y ion peaks or adjacent b ion peaks

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

I/L-I/L

3) See below.....

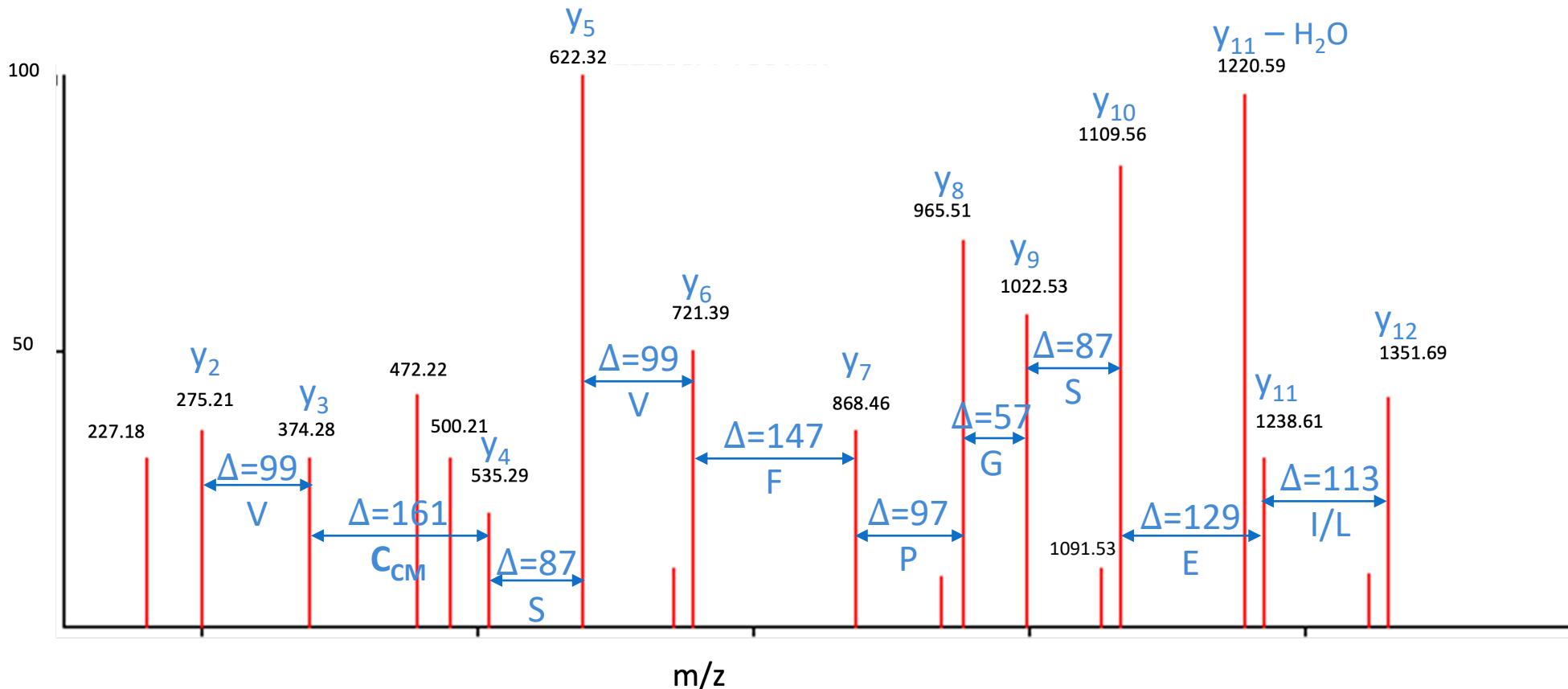
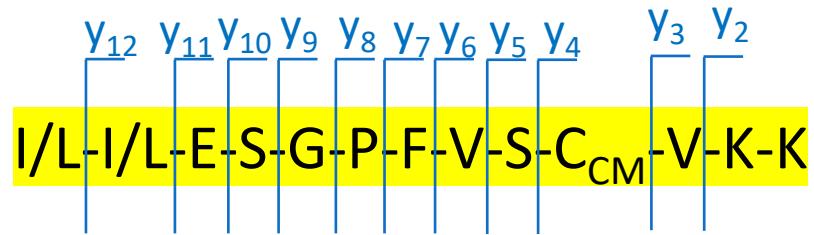
I/L-I/L-E-S-G-P-F-V-S-C_{CM}-V-...



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.



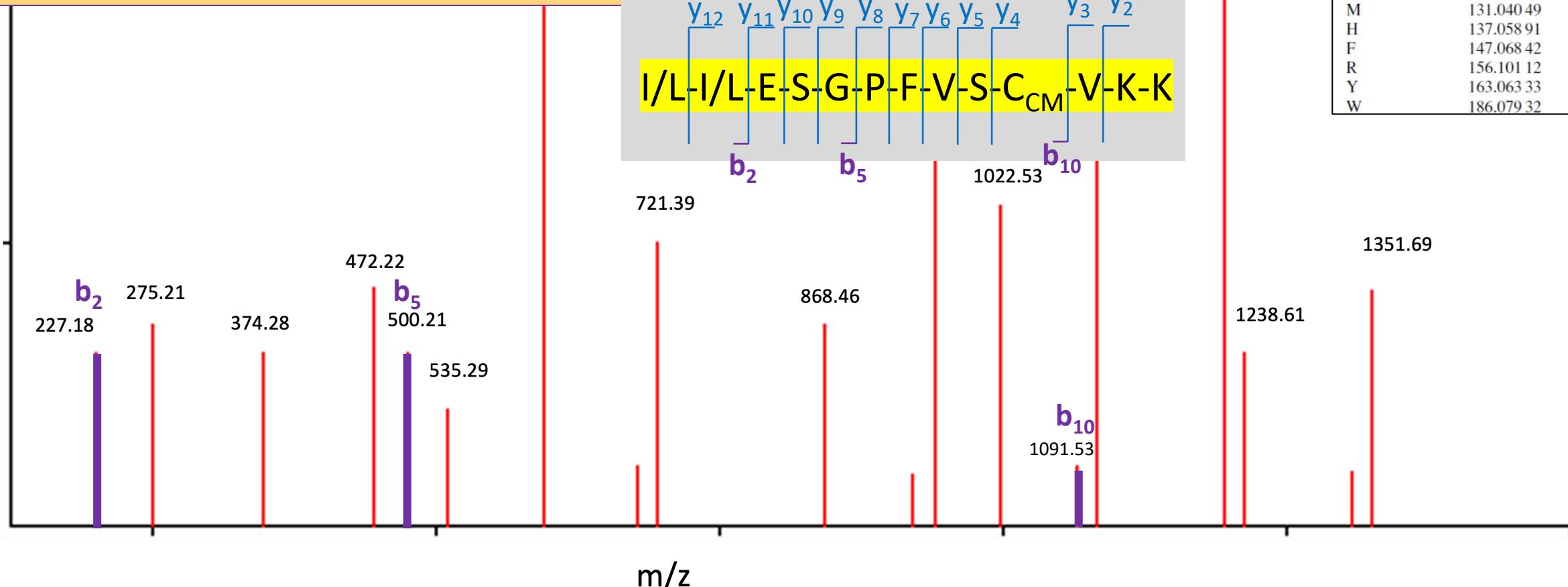
C_{CM} : Cysteine with Carboxymethyl (58.01)

- 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.

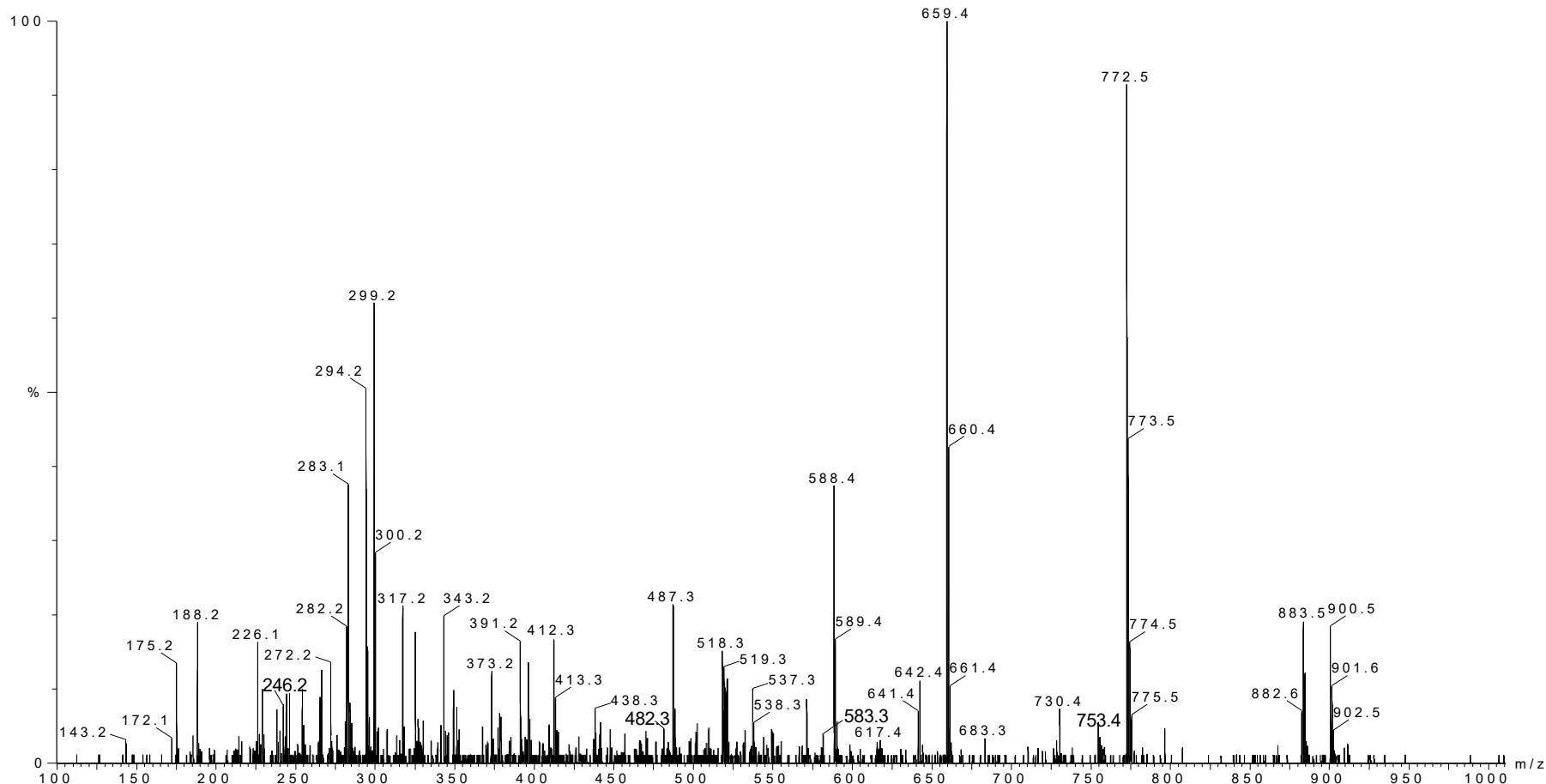
$$500.21 - 227.18 = 273.03 = E+S+G$$

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



Code (1 letter)	Monoisotopic mass
G	57.021 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?

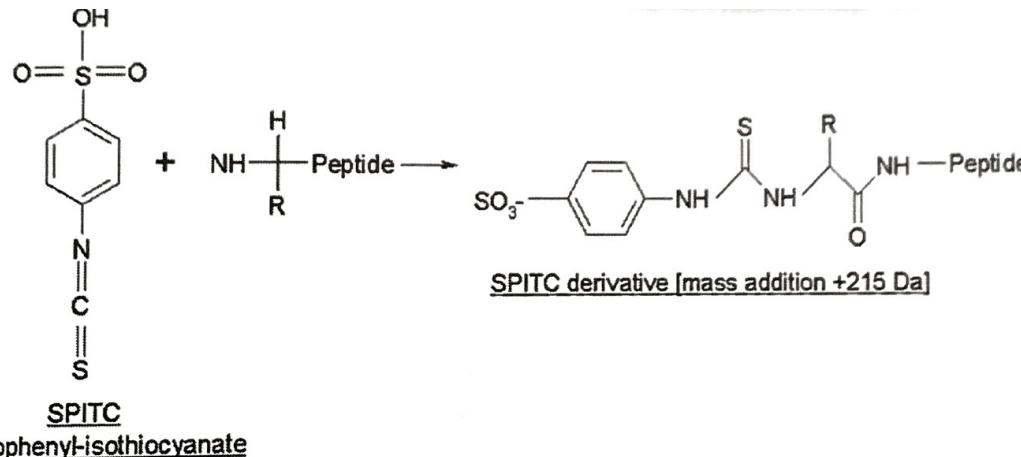


Chemical Derivatization For Sequence Analysis

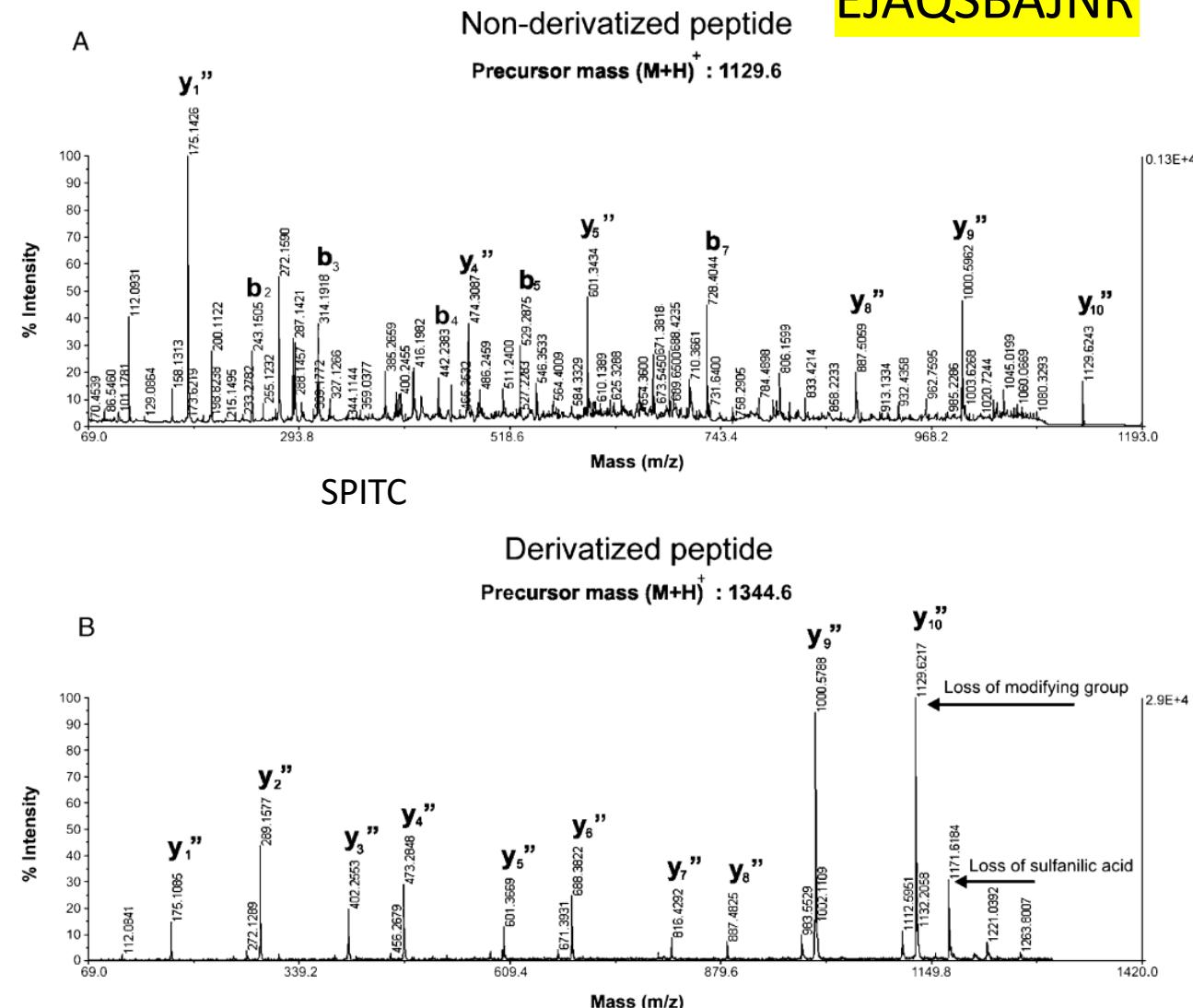
- *Derivatization* is a procedure that utilizes chemical reactions with covalent bond formation
- The aims of derivatization are:
 - To enhance detection of one ion series
 - To improve fragmentation yield
 - To simplify data interpretation (for de novo sequencing)

1. SPITC (4-sulfophenylisothiocyanate)

- SPITC can generate sulfonation of peptides at the amino terminus



- SPITC selectively enhances detection of the y ion series

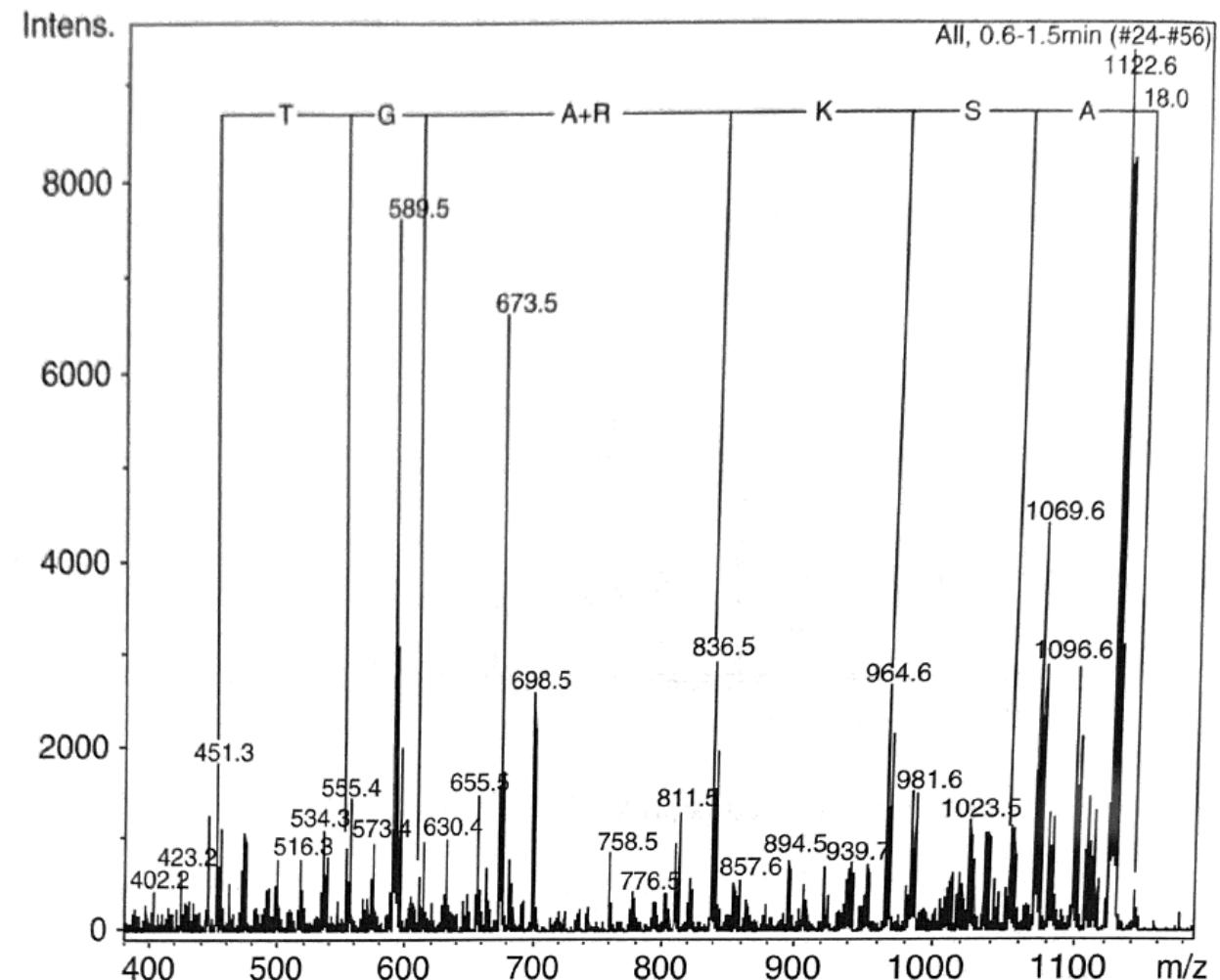


2. Acetylation and DeuteroAcetylation

FGGFTGARKSA

[M+H]⁺ = 1140.6

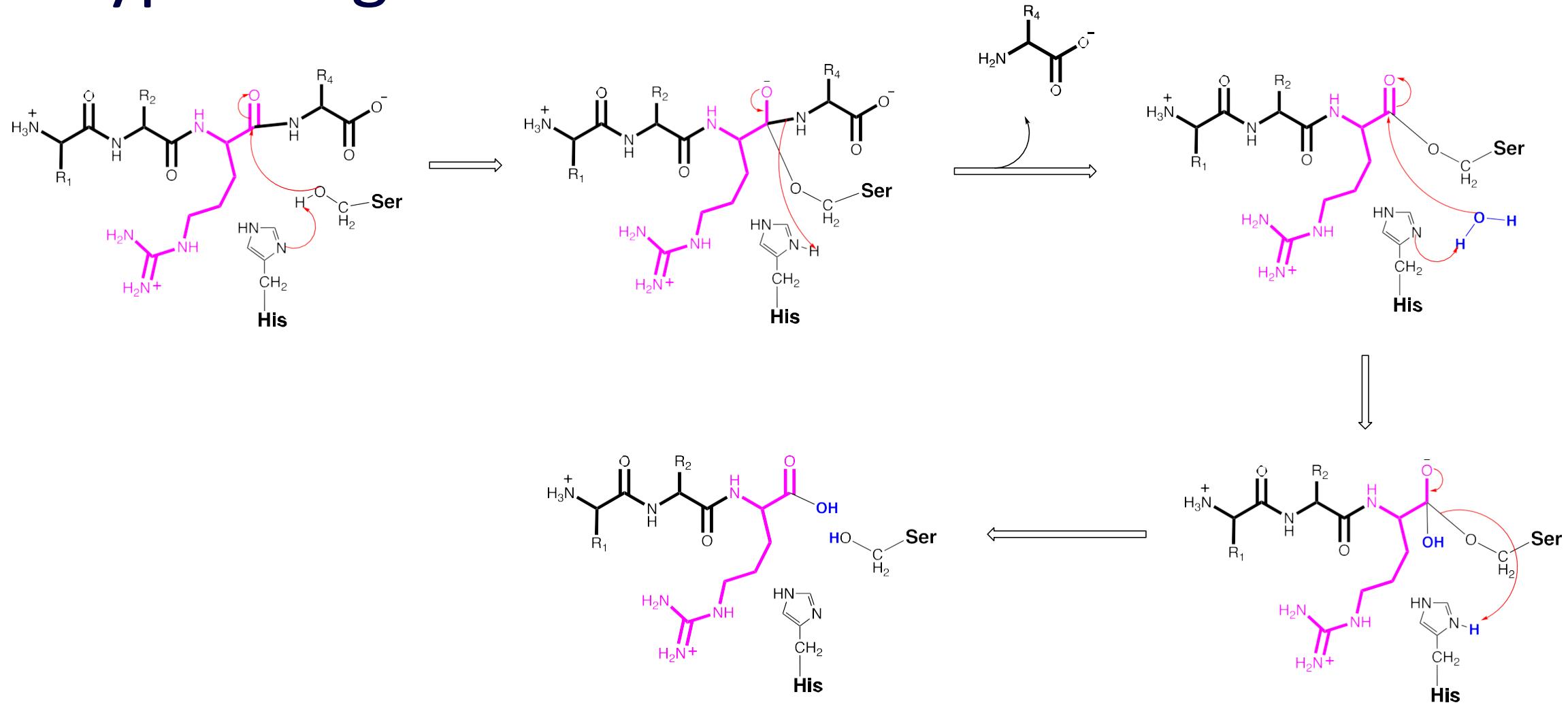
- It utilizes a mixture of acetic anhydride and deuterated acetic anhydride (1:1 v/v) in methanoic solution, which labels N terminal amino groups only.



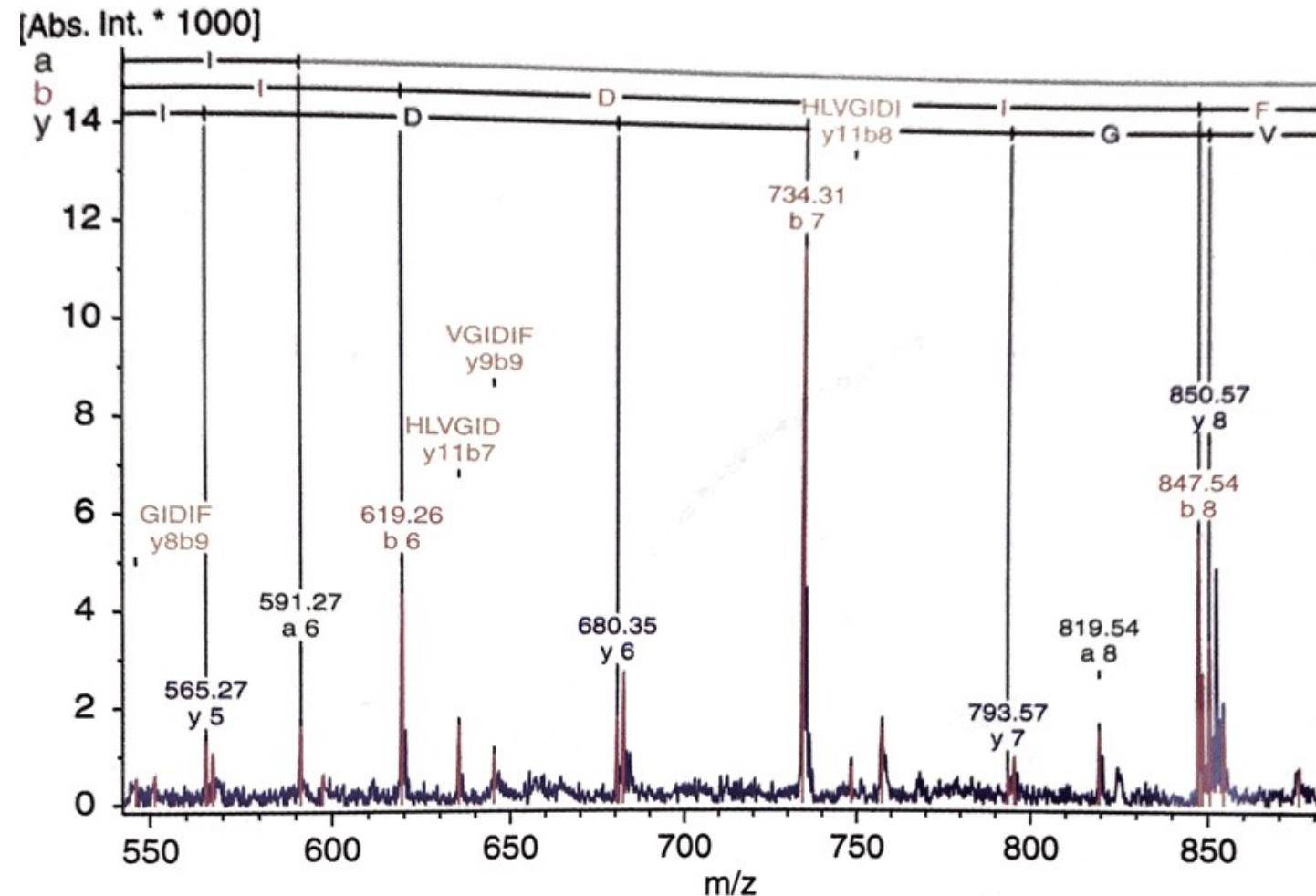
3. ^{18}O Labeling

- 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

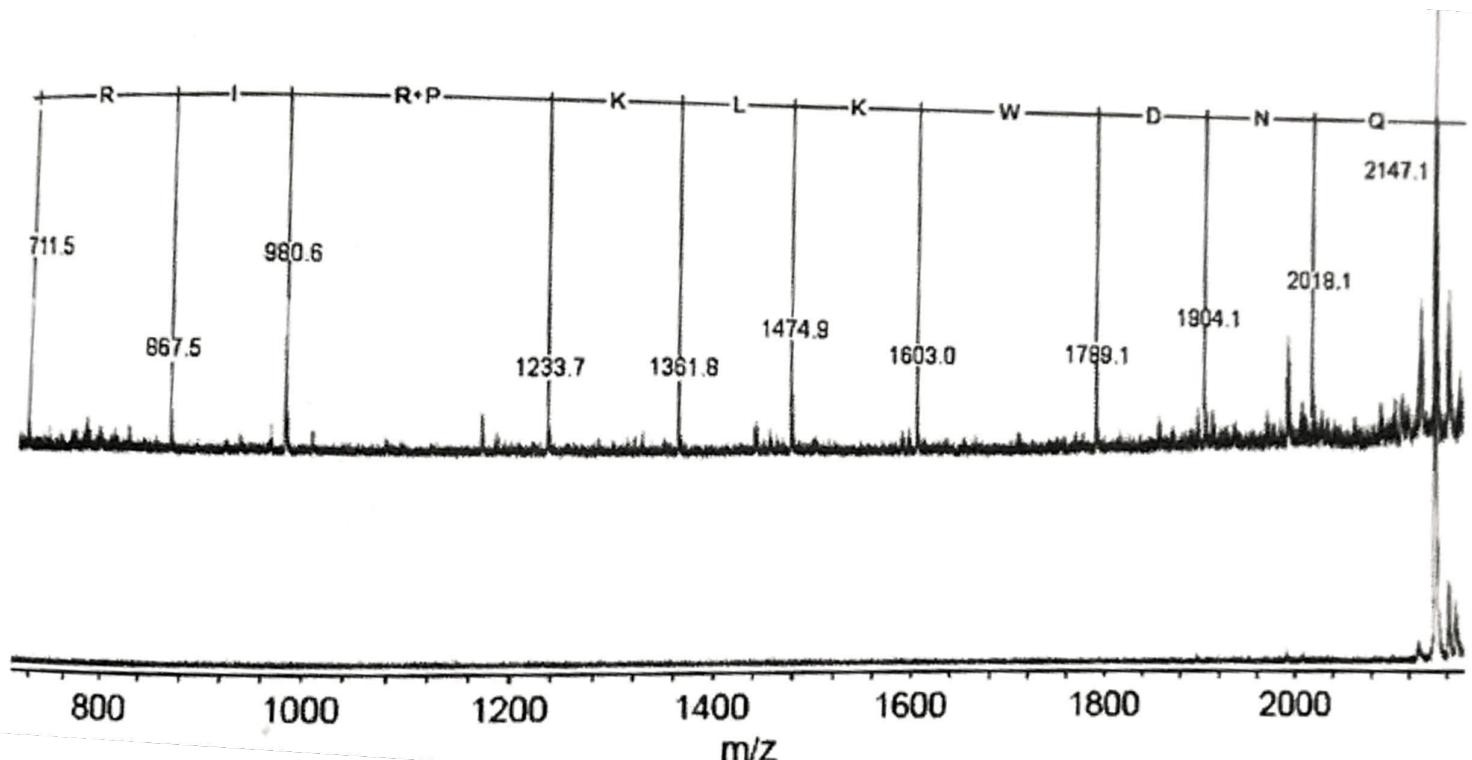
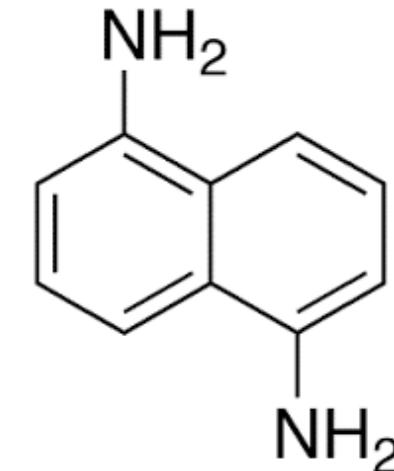


¹⁸O Labeling

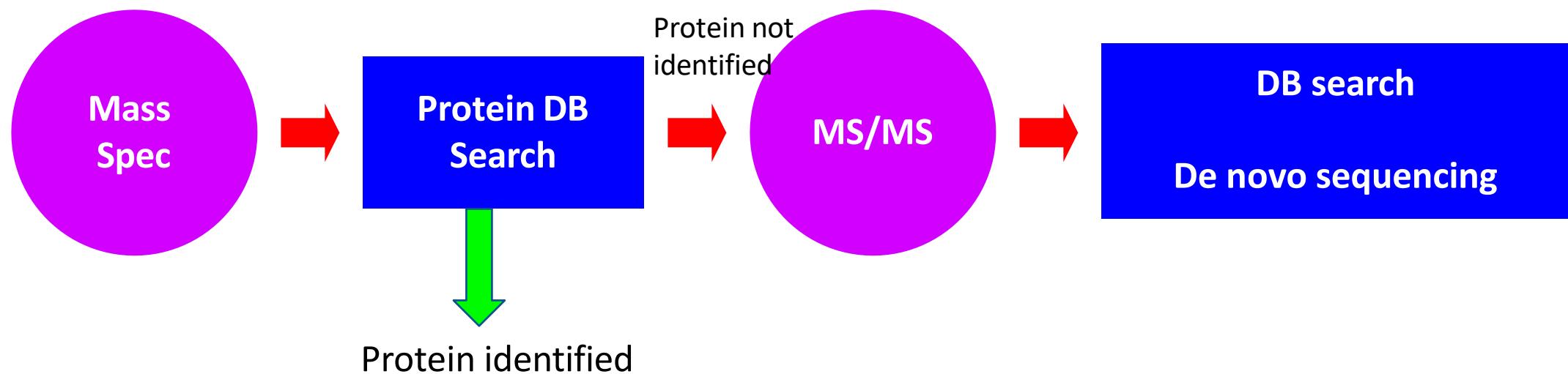


4. DAN (1,5-diaminonaphthalene)

- 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

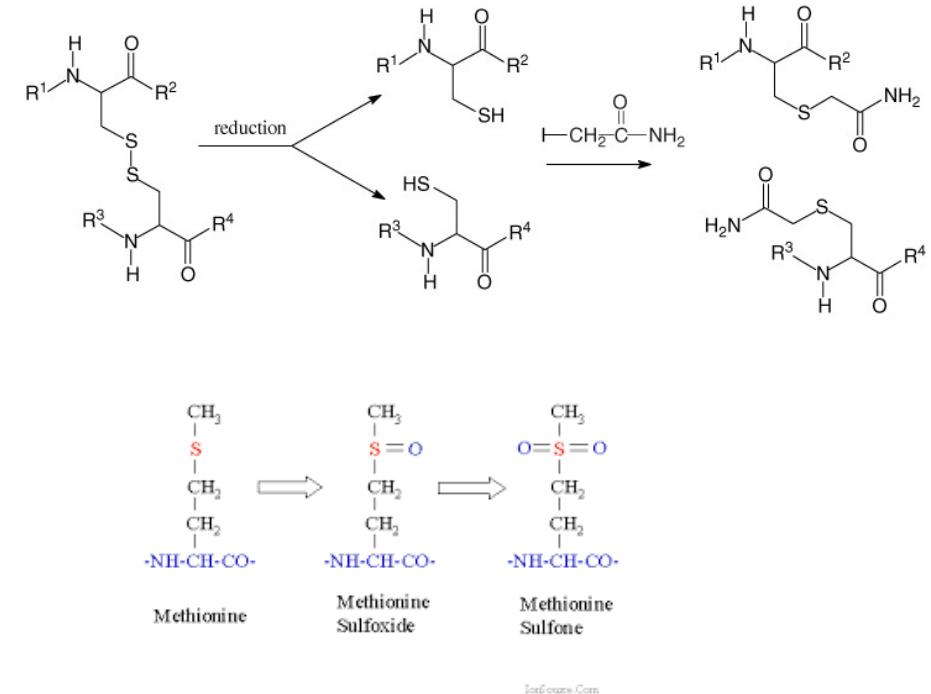


Protein Identification and Characterization Map

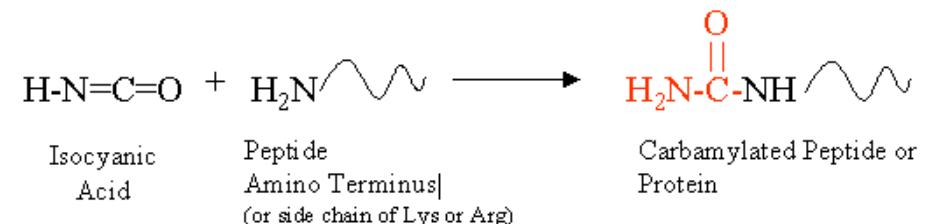


Specific Amino Acids Modification During Sample Handling

- 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.

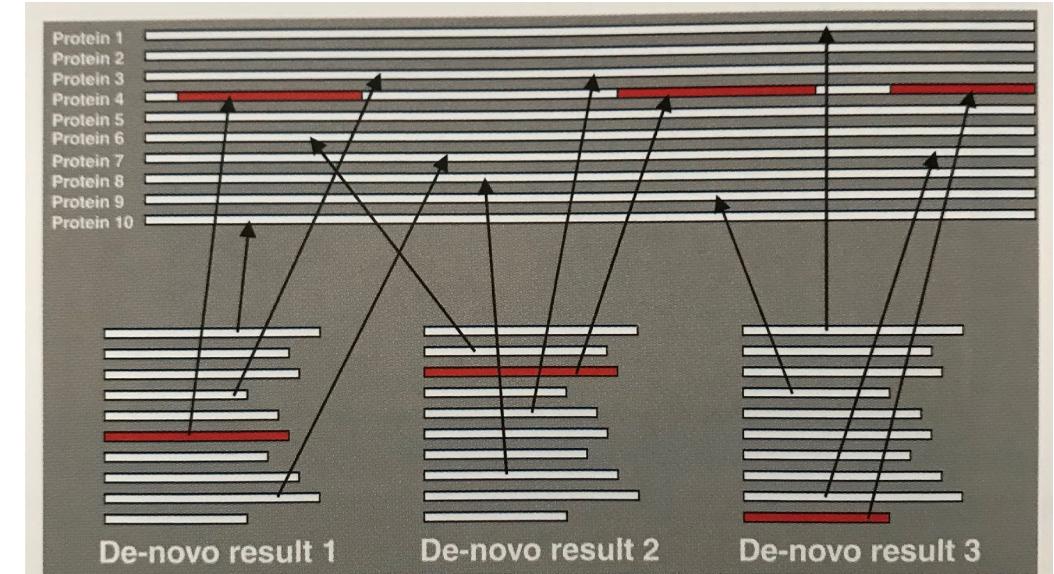


Carbamylation of Proteins
(amino terminus of a peptide used as an example)



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