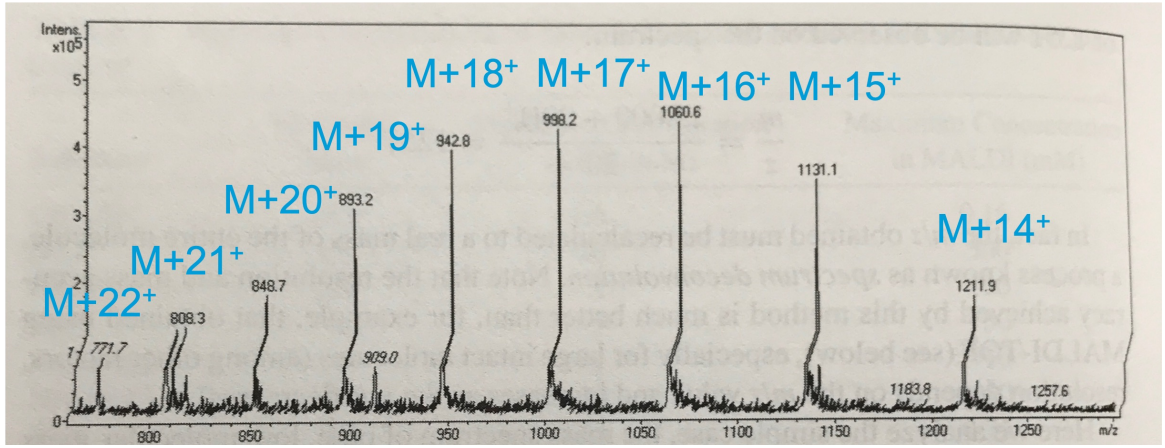
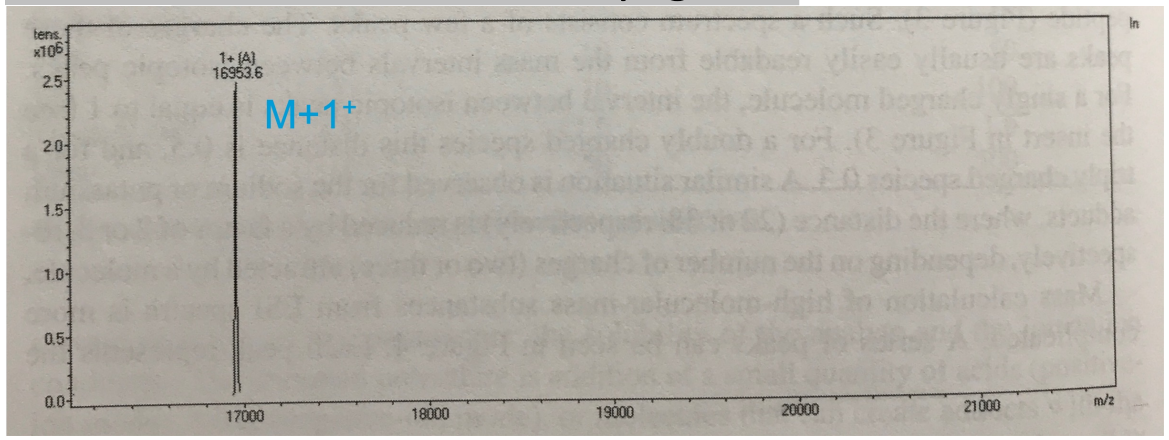
MS basics: Review

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

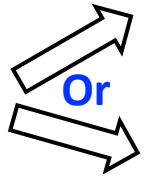
myoglobin MW = 16952.7 $\pm$ 0.87 Da

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

myoglobin MW = 16953.6 – 1 = 16952.6 Da

Peptides Fragment by CID: Review

For a singly protonated peptide,



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

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.

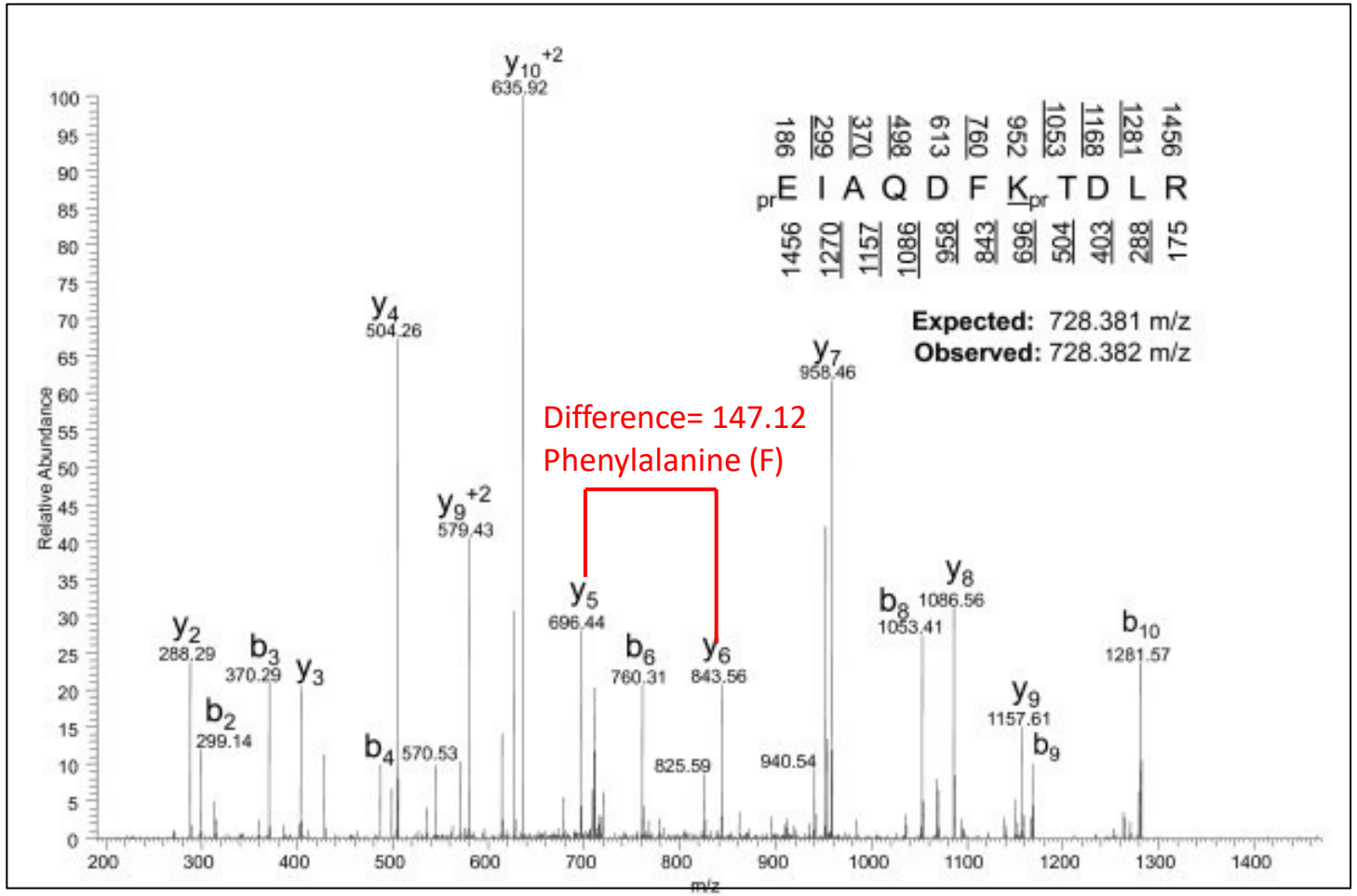
From Wysocki et al., J. Mass Spectrom. (2000) 35:1399

Fragmentation Results in a Peptide “Ladder”

Peptide: A-B-C-D-E

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

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

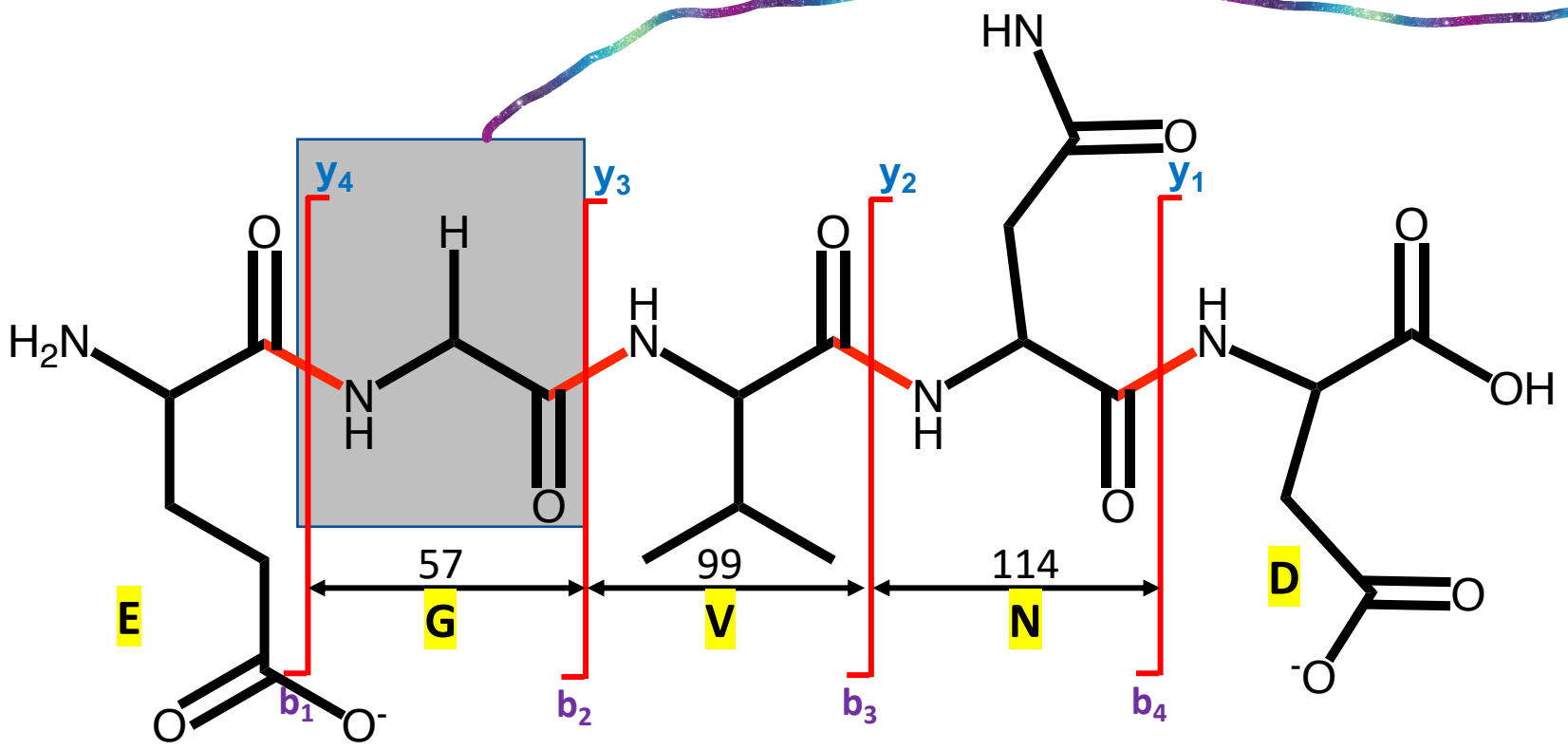


- Mixture of b ions and y ions
- MS/MS of 2^+ charged tryptic peptides yield (often) 1^+ charged product ions (but 2^+ charged products can be observed as well)
- Not all b ions or y ions are visible

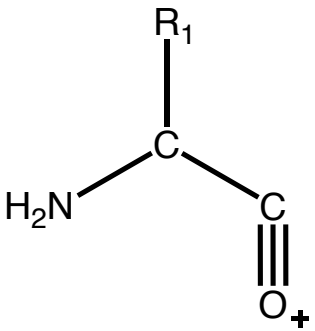
The mass of the precursor is 1454.764

the observed ion was doubly charged
728.382 x 2 - 2 = 1454.764 Da

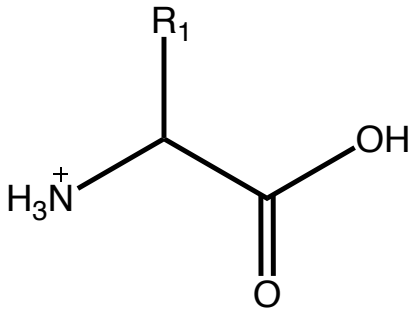
https://www.researchgate.net/profile/Rebecca-Levin-2/publication/49660823/figure/fig2/AS:214315966701573@1428108316420/Tandem-mass-spectrometry-MS-MS-spectrum-of-labeled-lysine-peptide-MS-MS-of-the-H3.png



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



b_1 ion

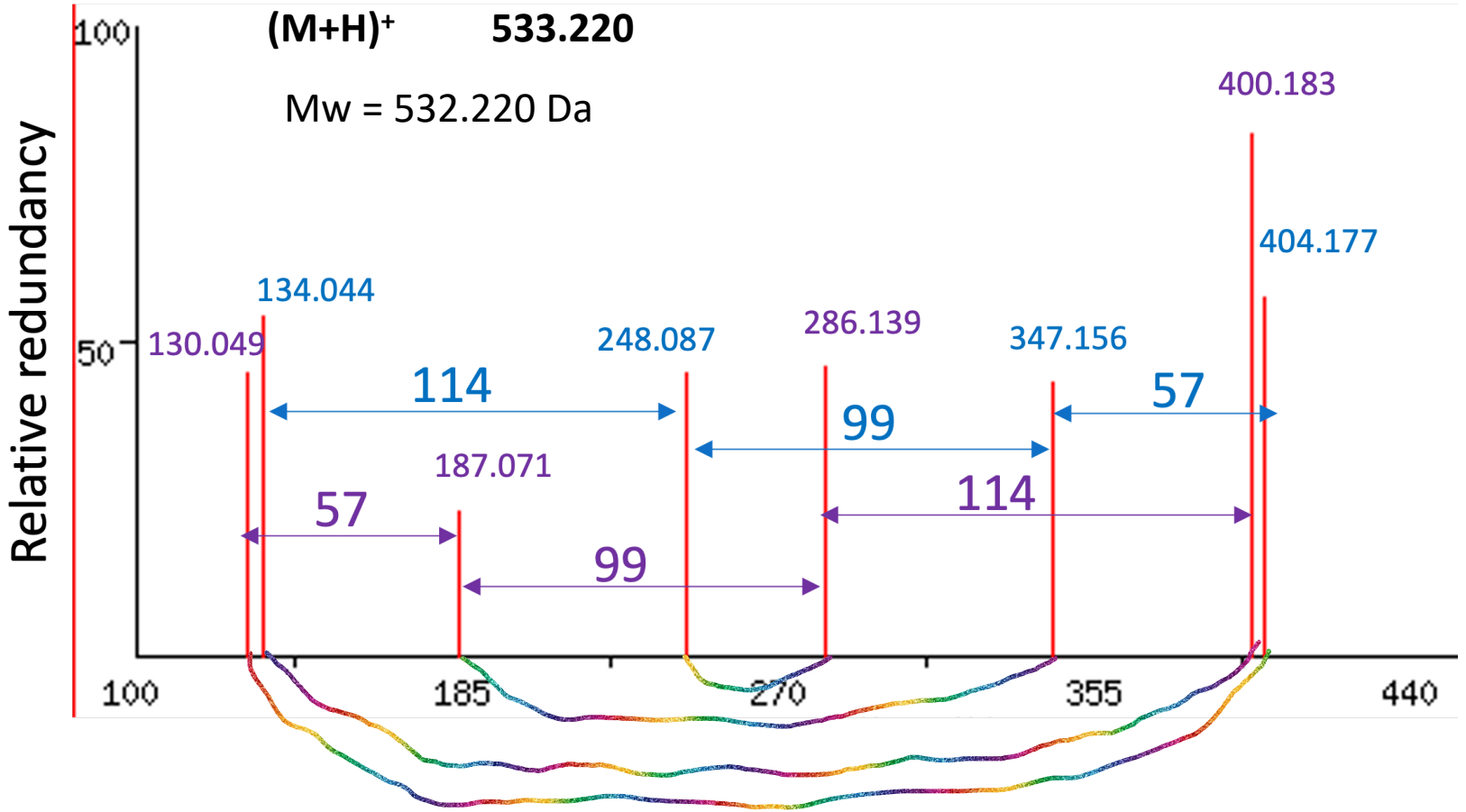


y_1 ion

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}

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

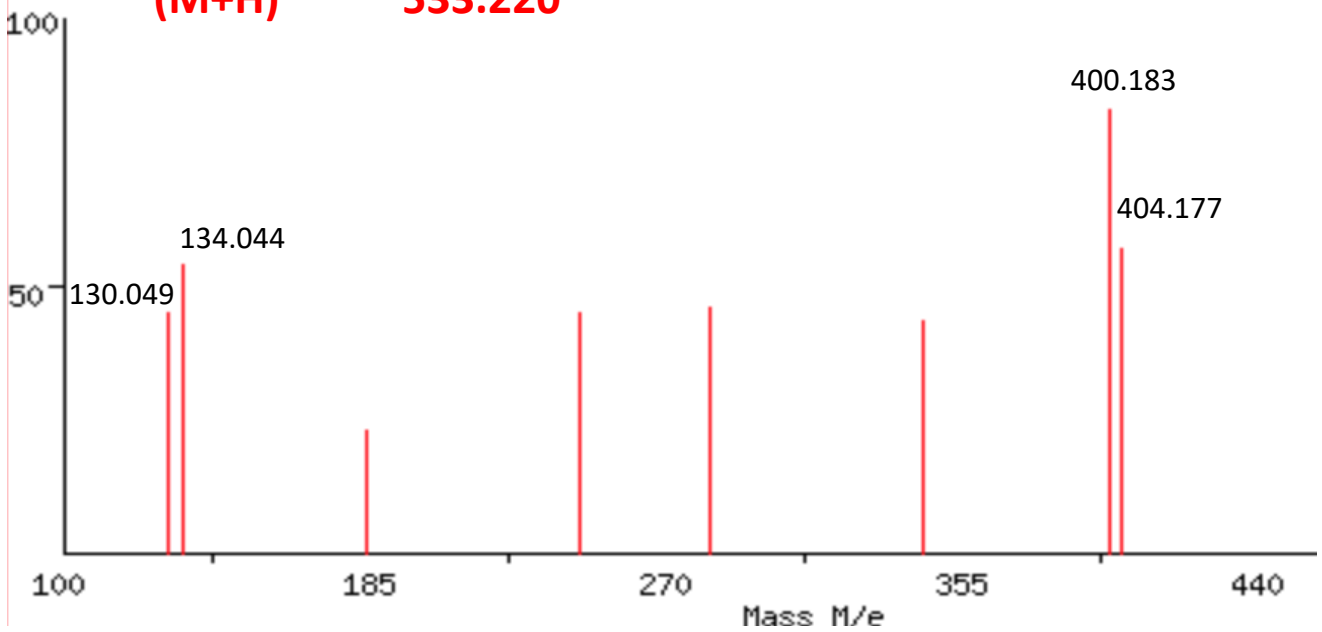
130.049 – 1 = 129.049 **E** on N terminus

134.044 – 19 = 115.044 **D** on C terminus

532.220 – (404.177-2) -1 = 129.043 **E** on N terminus

532.220 – 400.183 – 17 (OH) = 115.037 **D** on C terminus

$(M+H)^+$ 533.220



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

Calculate the Terminal Residues

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

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

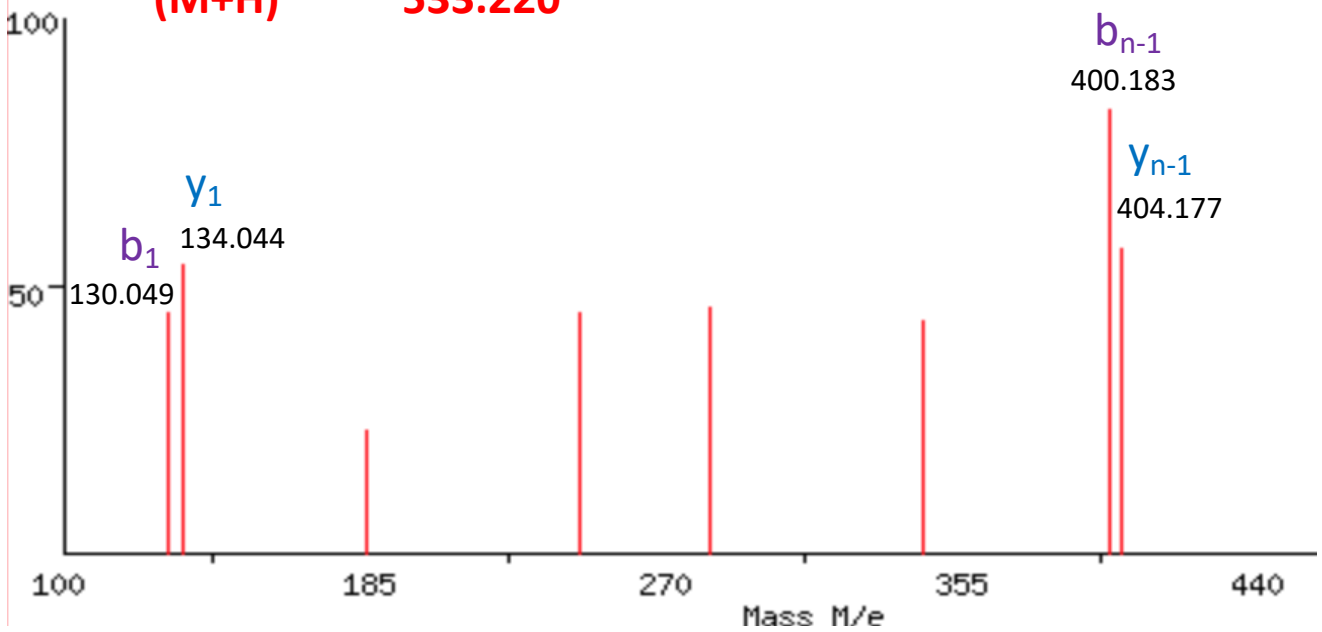
130.049 – 1 = 129.049 **E** on N terminus

134.044 – 19 = 115.044 **D** on C terminus

532.220 – (404.177-2) -1 = 129.043 **E** on N terminus

532.220 – 400.183 – 17 (OH) = 115.037 **D** on C terminus

$(M+H)^+$ 533.220



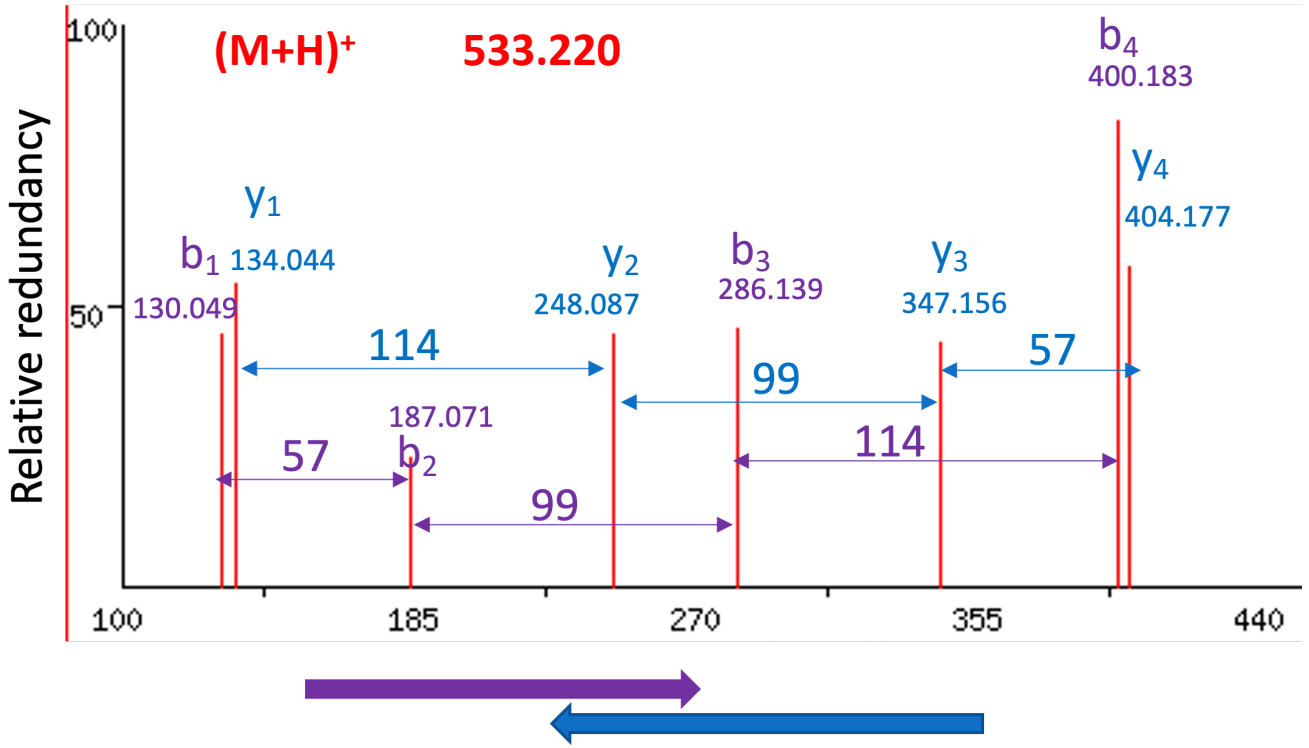
M - y_{n-1} ion + 1 = mass of 1st residue on N terminus

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

Δmass and Complementary b/y Ion Pairs

	mass^{1+}	b-ions	y-ions	mass^{1+}
b_1^+	130.049	E	GVND	404.177
b_2^+	187.071	EG	VND	347.156
b_3^+	286.139	EGV	VND	248.087
b_4^+	400.182	EGVN	D	134.044

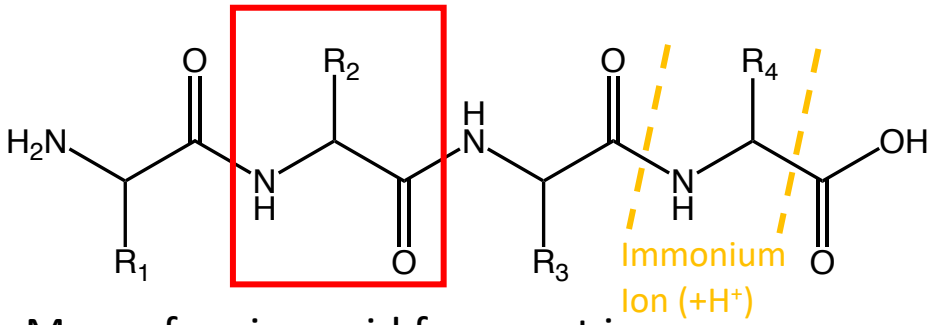
EGVND



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

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 1^{st} residue on N terminus
- M- b_{n-1} ion - 17 = mass of 1^{st} 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.



Mass of amino acid fragment ion

Name	3-letter code	1-letter code	Residue Mass	Immonium ion	Related ions	Composition
Alanine	Ala	A	71.03711	44		C_3H_5NO
Arginine	Arg	R	156.10111	129	59,70,73,87,100,112	$C_6H_{12}N_4O$
Asparagine	Asn	N	114.04293	87	70	$C_4H_6N_2O_2$
Aspartic Acid	Asp	D	115.02694	88	70	$C_4H_5NO_3$
Cysteine	Cys	C	103.00919	76		C_3H_5NOS
Glutamic Acid	Glu	E	129.04259	102		$C_5H_7NO_3$
Glutamine	Gln	Q	128.05858	101	56,84,129	$C_5H_8N_2O_2$
Glycine	Gly	G	57.02146	30		C_2H_3NO
Histidine	His	H	137.05891	110	82,121,123,138,166	$C_6H_7N_3O$
Isoleucine	Ile	I	113.08406	86	44,72	$C_6H_{11}NO$
Leucine	Leu	L	113.08406	86	44,72	$C_6H_{11}NO$
Lysine	Lys	K	128.09496	101	70,84,112,129	$C_6H_{12}N_2O$
Methionine	Met	M	131.04049	104	61	C_5H_9NOS
Phenyalanine	Phe	F	147.06841	120	91	C_9H_9NO
Proline	Pro	P	97.05276	70		C_5H_7NO
Serine	Ser	S	87.03203	60		$C_3H_5NO_2$
Threonine	Thr	T	101.04768	74		$C_4H_7NO_2$
Tryptophan	Trp	W	186.07931	159	11,117,130,132,170,100	$C_{11}H_{10}N_2O$
Tyrosine	Tyr	Y	163.06333	136	91,107	$C_9H_9NO_2$
Valine	Val	V	99.06841	72	44,55,69	C_5H_9NO

Mass of b_2 ions (+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.

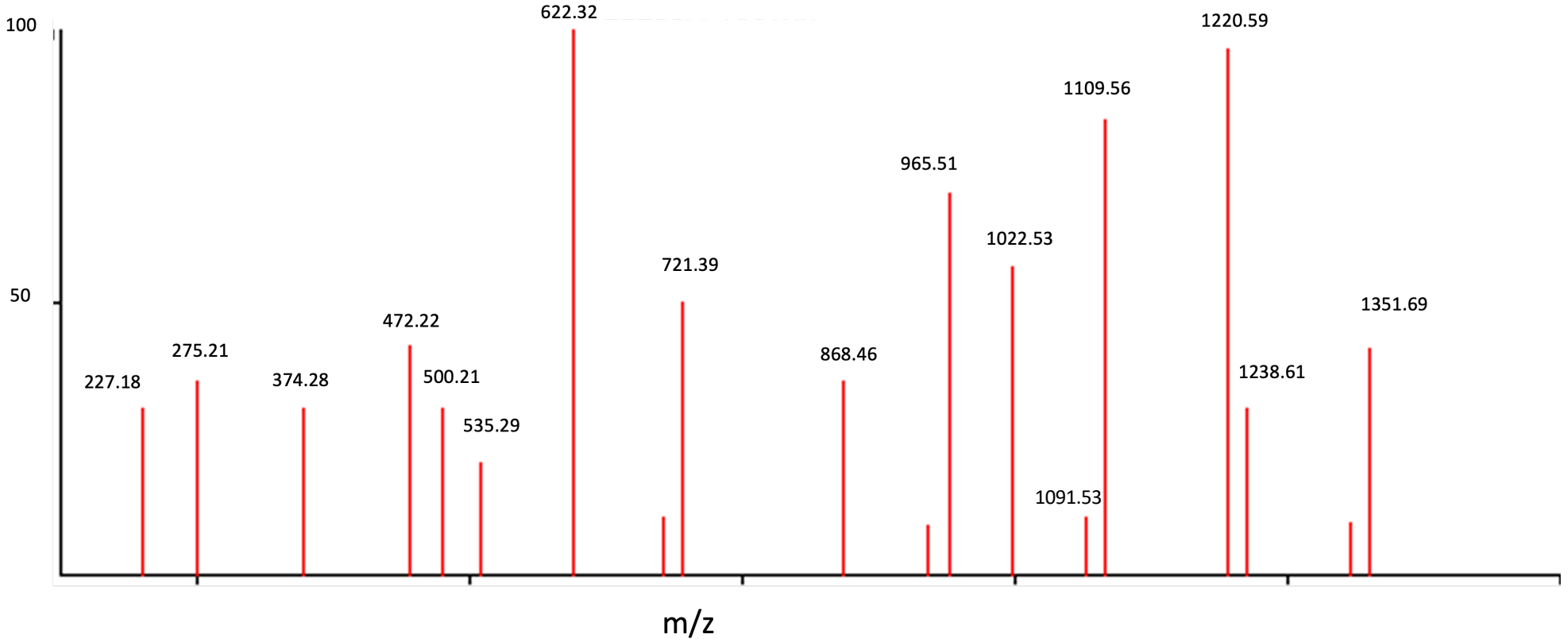
$[M+H]^+$ = 1464.7693

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

I/L-



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)

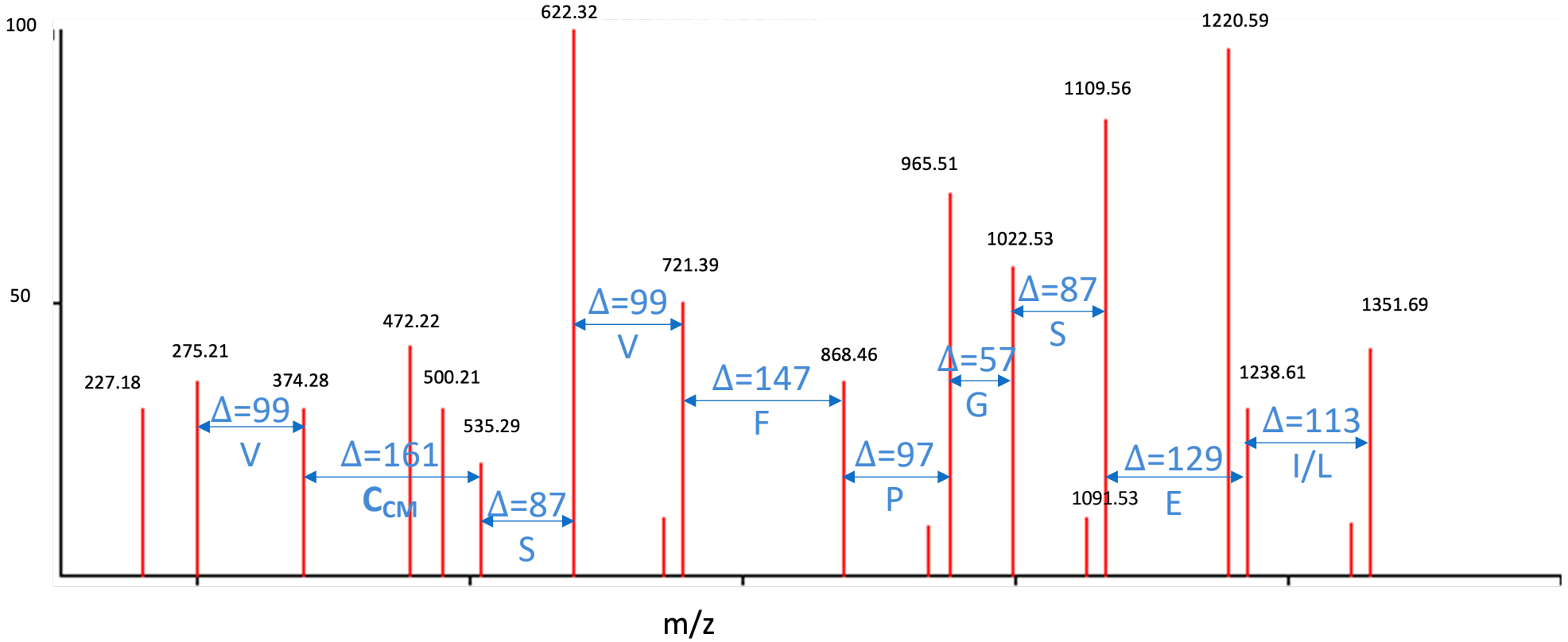
$[M+H]^+$ = 1464.7693
So, Mw = 1463.7693 Da

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

2) Δm/z = 1351.69– 1238.61 = 113.08, which is the mass of **I/L**.
3) See below……

I/L-I/L

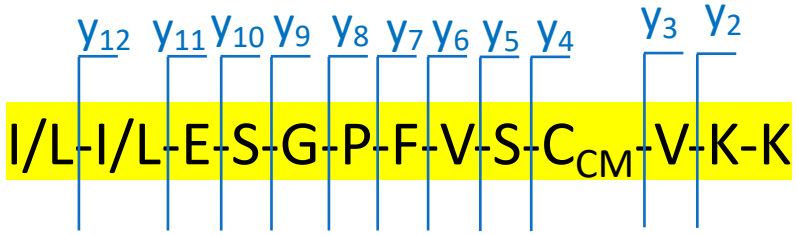
I/L-I/L-E-S-G-P-F-V-S-C_{CM}-V-…

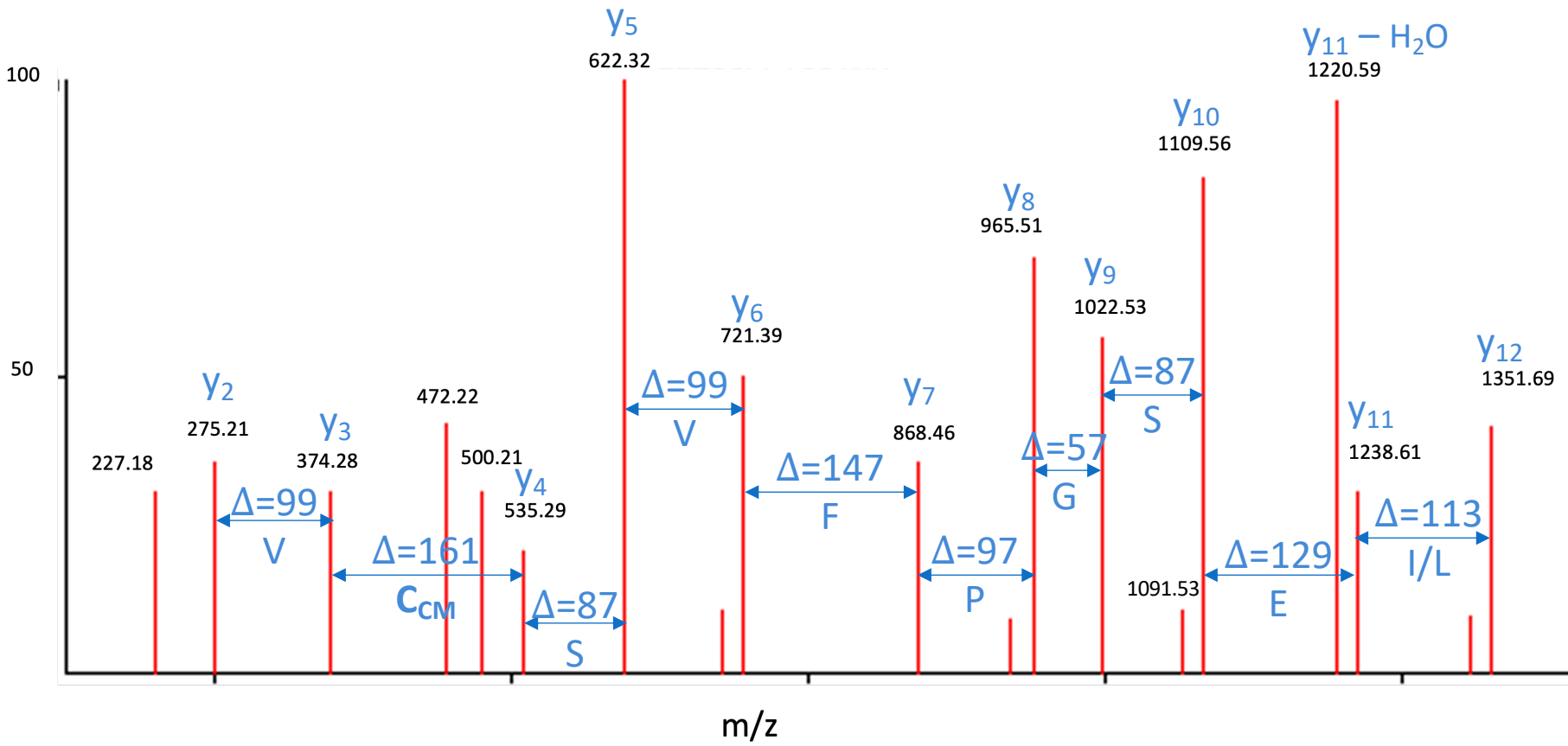


Code (1 letter)	Monoisotopic mass
G	57.021 47
A	71.037 12
S	87.032 03
P	97.052 77
V	99.068 42
T	101.047 68
C	103.009 19
I	113.084 07
L	113.084 07
N	114.042 93
D	115.026 95
Q	128.058 58
K	128.094 97
E	129.042 60
M	131.040 49
H	137.058 91
F	147.068 42
R	156.101 12
Y	163.063 33
W	186.079 32

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 = y2 – 19 = 256.21, which are the sum of K and K.



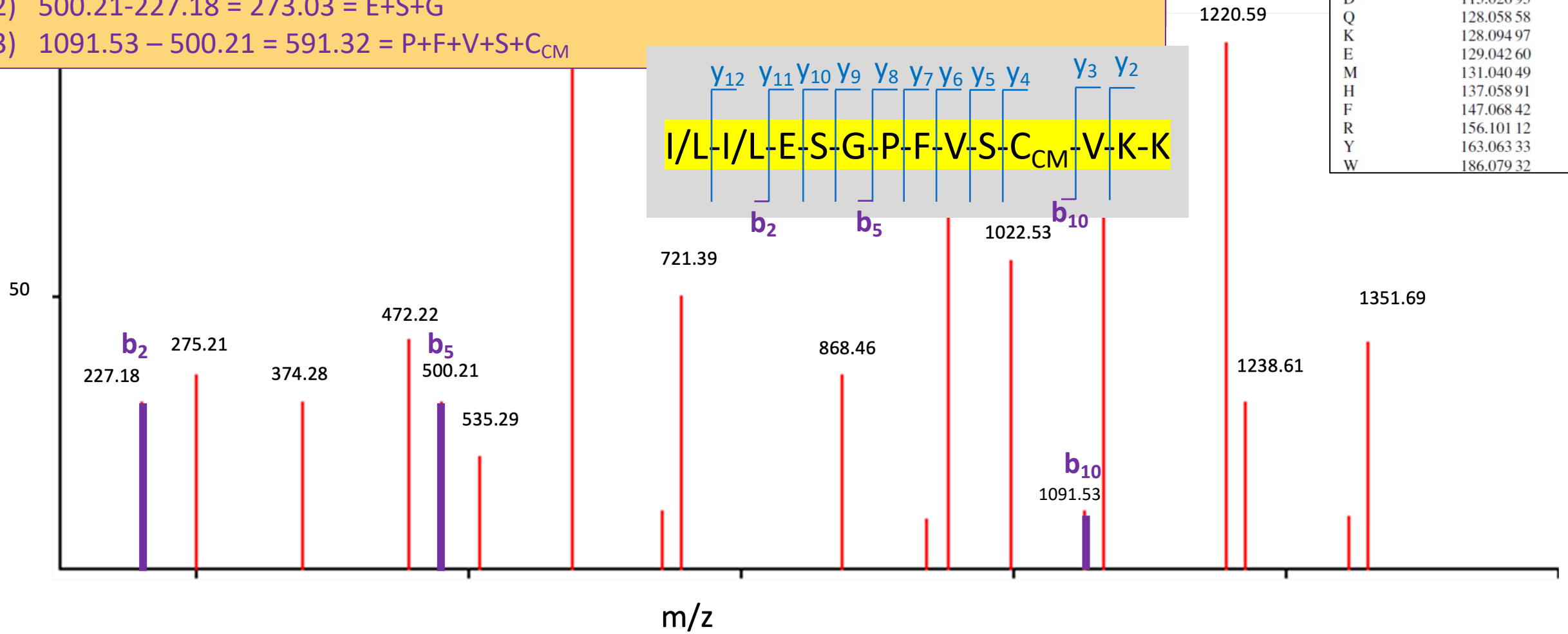


C_{CM} : Cysteine with Carboxymethyl (58.01)

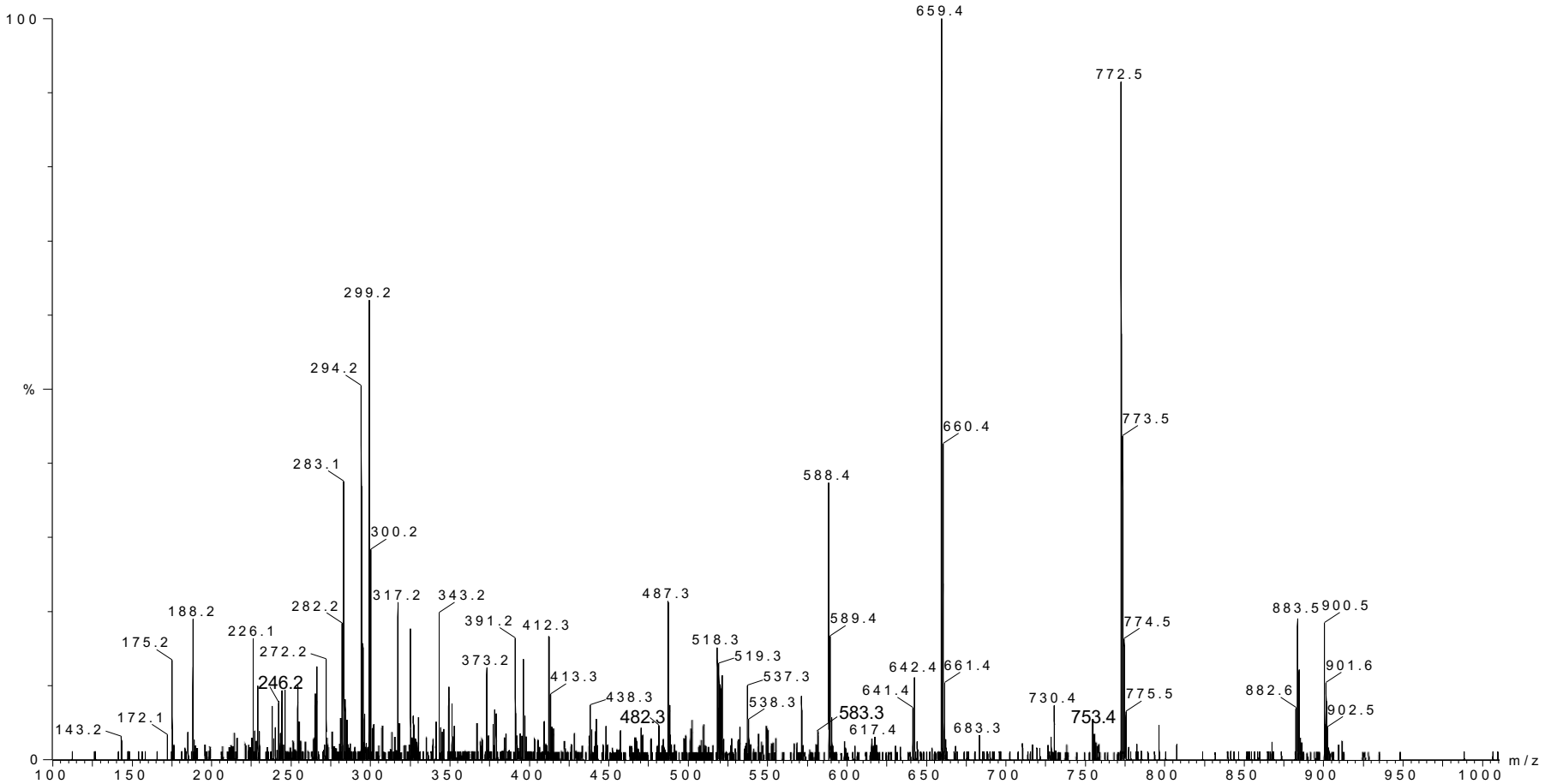
- Then to verify the high mass y ion assignments, we look for the complimentary low mass b ions.
- We may not be able to see b1. Usually, we will start by looking for b2.

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

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.
2) 500.21-227.18 = 273.03 = E+S+G
3) 1091.53 – 500.21 = 591.32 = P+F+V+S+C_{CM}



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

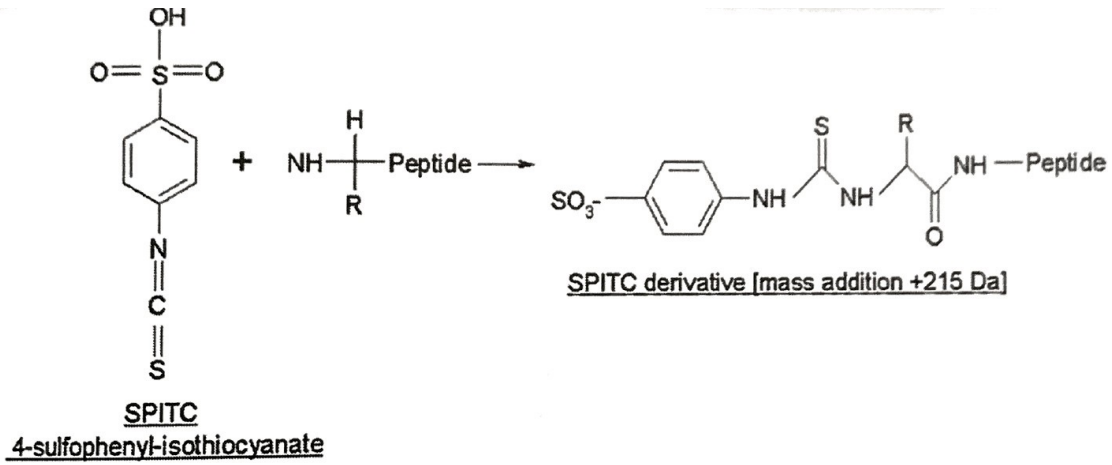
- ***Derivatization*** is a procedure that utilizes chemical reactions with covalent bond formation

- 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)

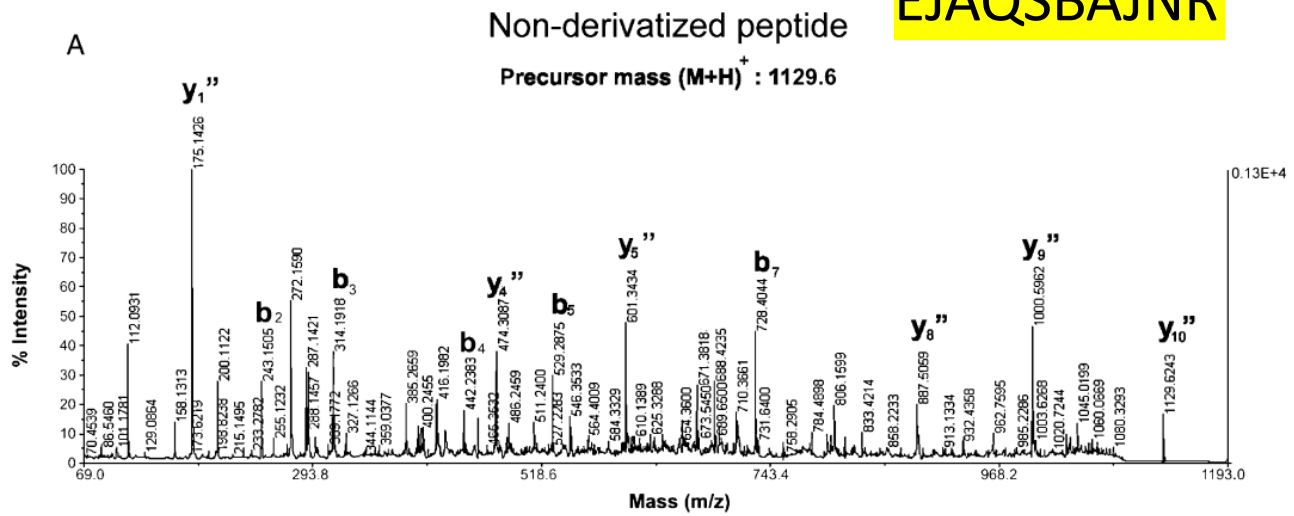
EJAQSBAJNR

- SPITC can generate sulfonation of peptides at the amino terminus

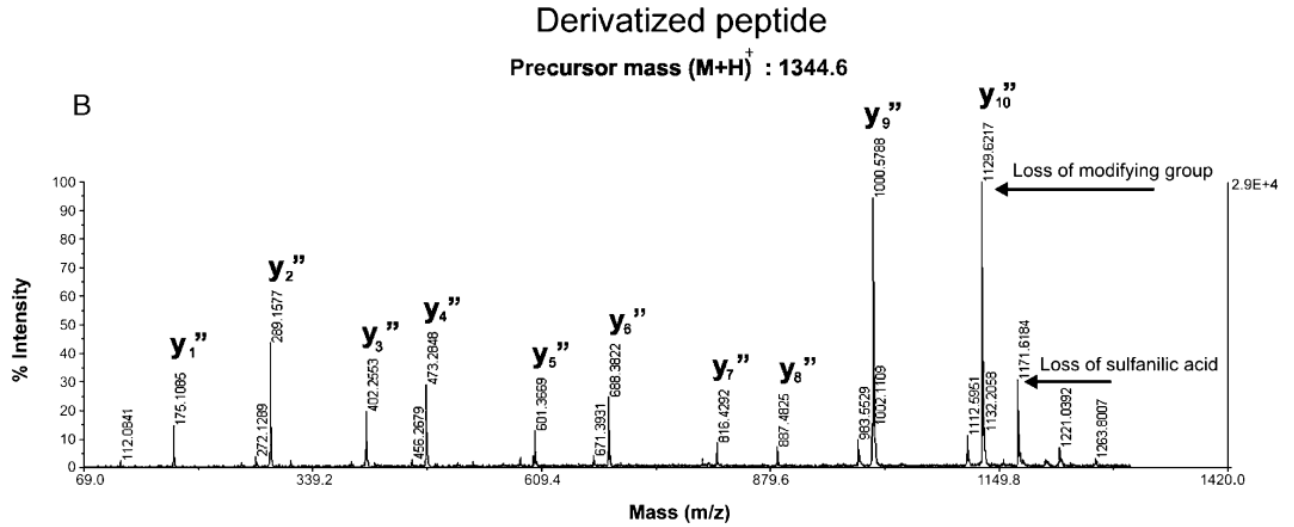


SPITC
4-sulfophenyl-isothiocyanate

SPITC derivative [mass addition +215 Da]



SPITC



- SPITC selectively enhances detection of the y ion series

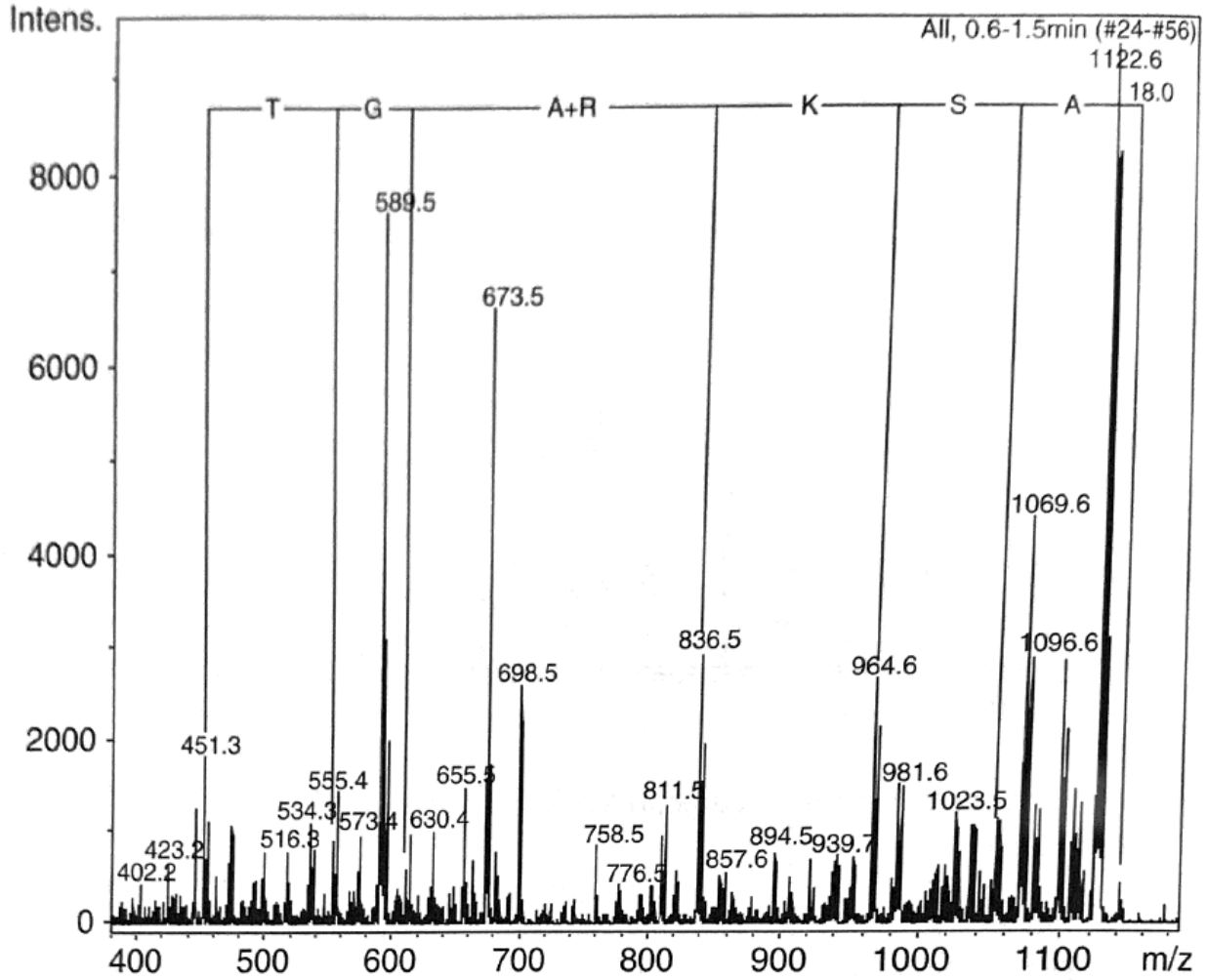
2. Acetylation and DeuteroAcetylation

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.

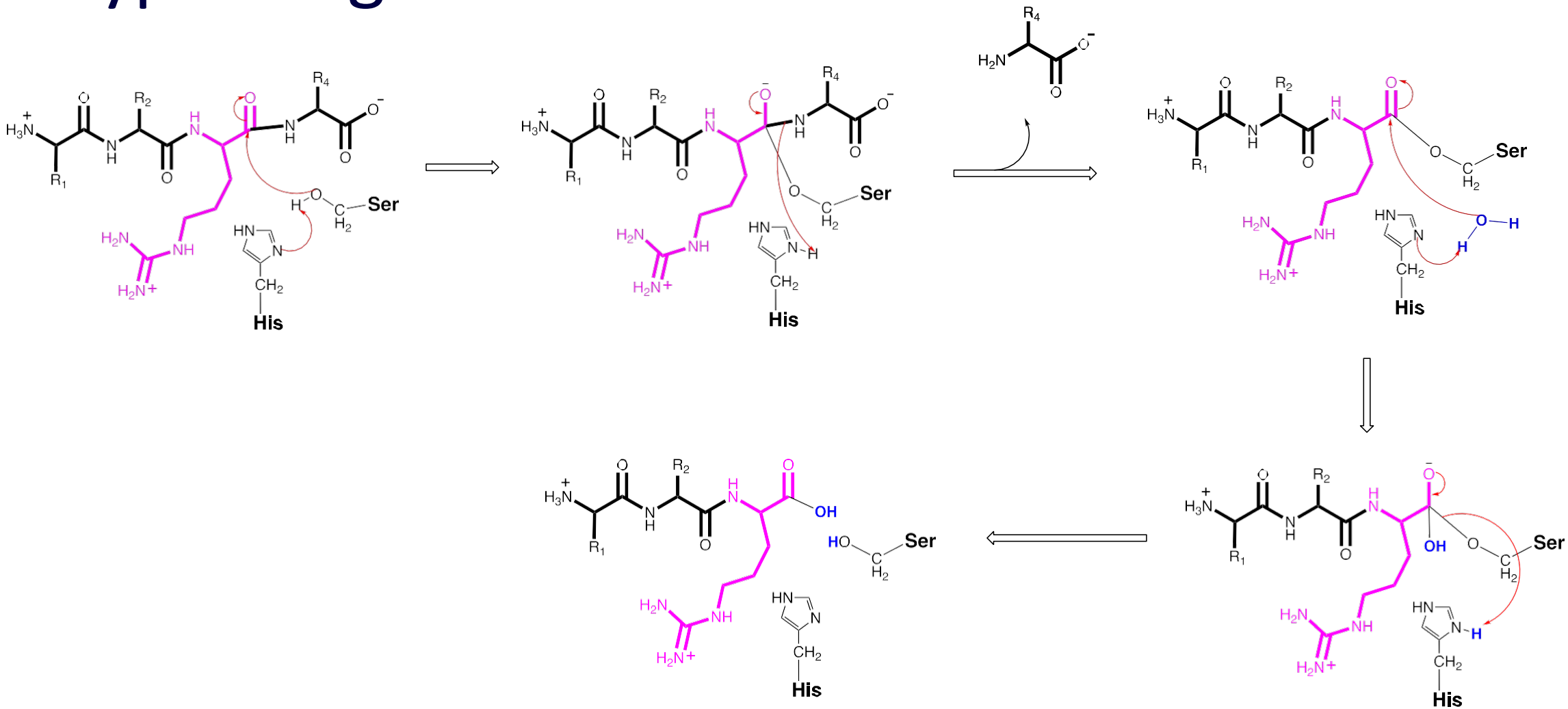
$(CH_3CO)_2O \xleftrightarrow{3\ Th} (CD_3CO)_2O$



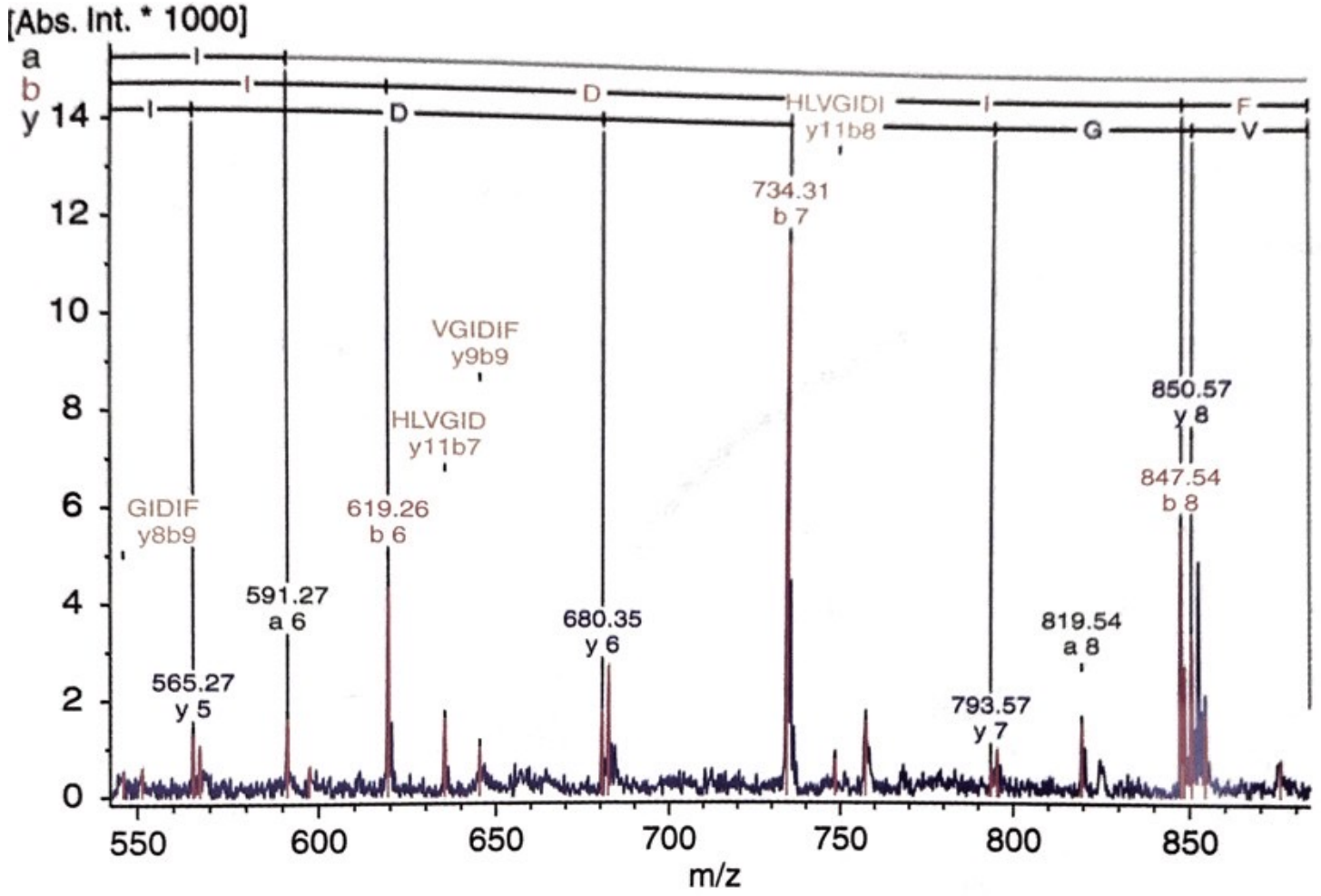
3. ^{18}O Labeling

- 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.
- After fragmentation, the MS/MS spectrum will show both single peaks (usually b ions) and single peak pairs (usually y ions).

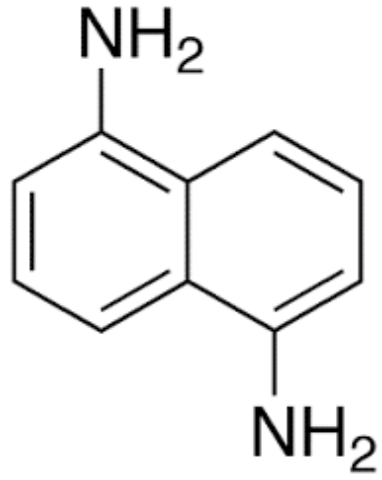
Trypsin Digestion mechanism



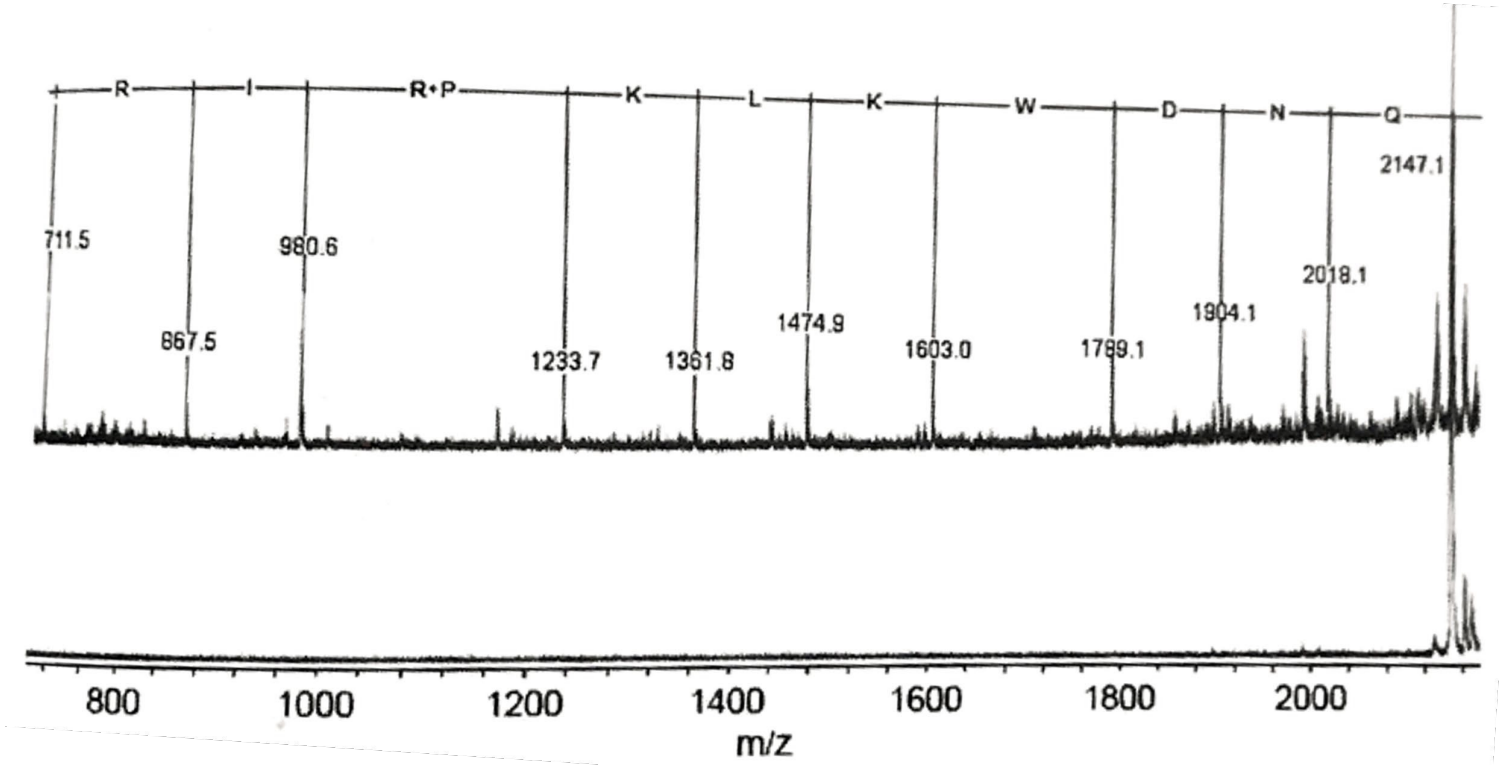
^{18}O Labeling



4. DAN (1,5-diaminonaphtalene)

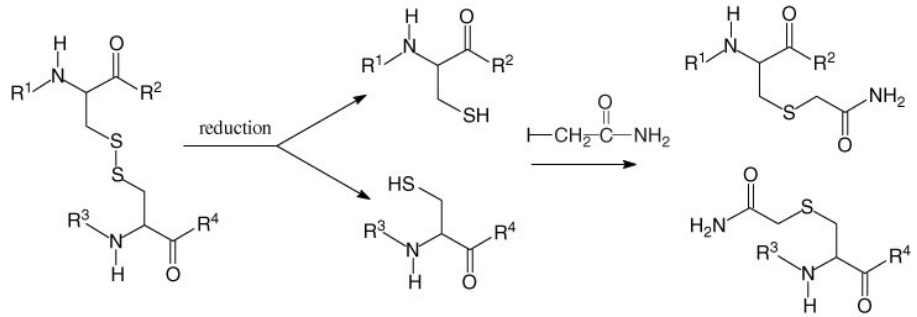


- Serves as both a matrix in MALDI method and a reducing reagent for S-S bond.
- It is suitable for top-down proteomics
- c ion series is usually most abundant in the spectrum

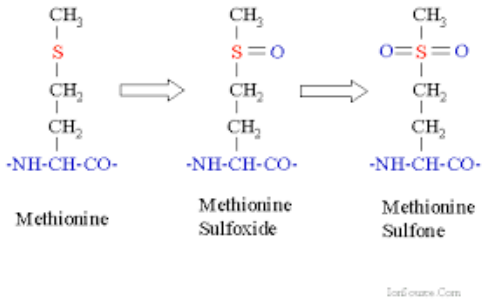


Specific Amino Acids Modification During Sample Handling

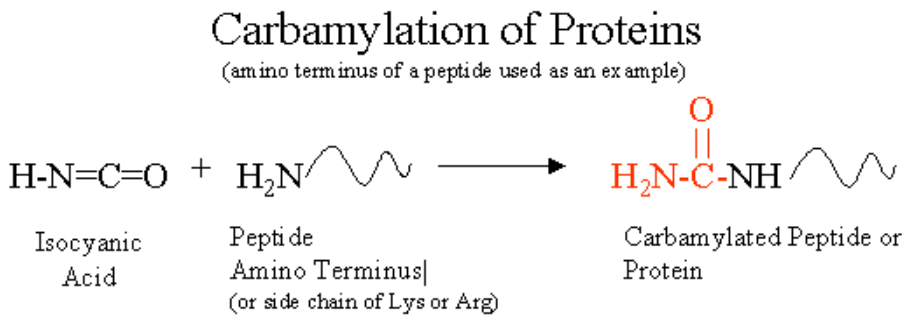
- Reduction and Alkylation on Cys
  - Routinely done prior to enzymatic digestion to break disulfide bonds, unfolding proteins to make them more susceptible to enzymatic cleavage



- Methionine is easily mono-oxidized (Met sulfoxide)



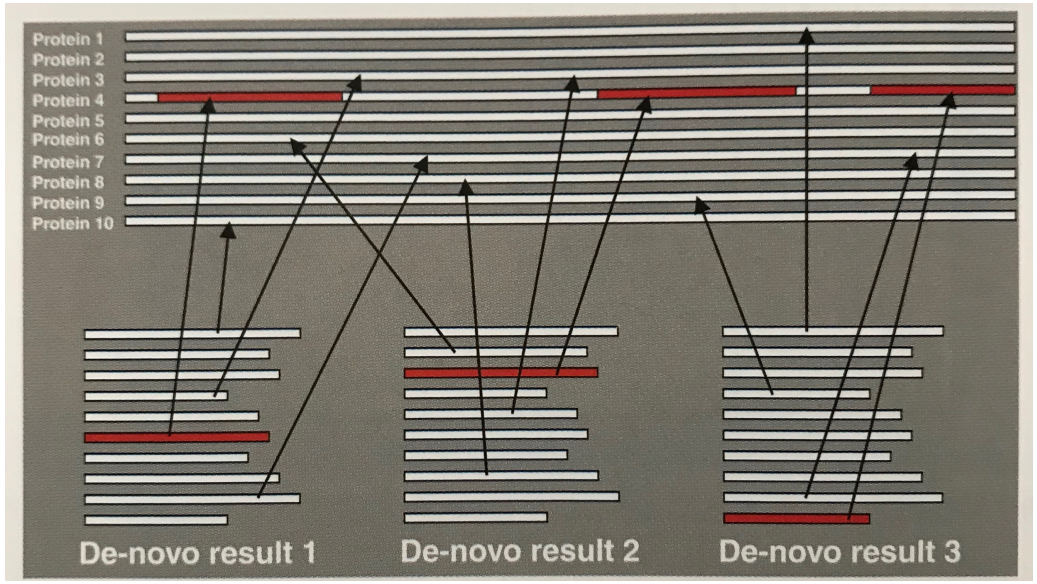
- Cyclization of N-terminal Glutamine (Q) and carboxamidomethyl-Cys
- Urea exposure can carbamylate N termini of protein/peptide and side chains of Lys



- 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

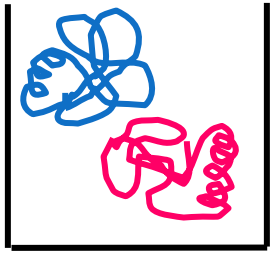
LECTURE 3: BOTTOM-UP PROTEOMICS AND TOP-DOWN PROTEOMICS

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)

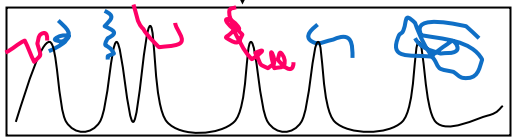
Bottom-up Proteomics



Reduction/
Alkylation

Digest - enzyme
(e.g. trypsin)

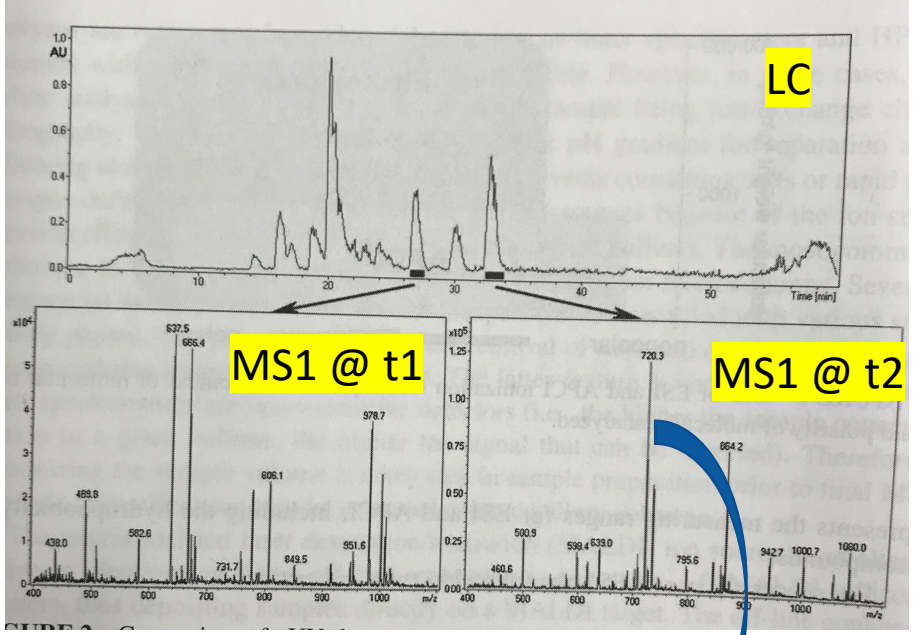
Multi-dimensional
HPLC

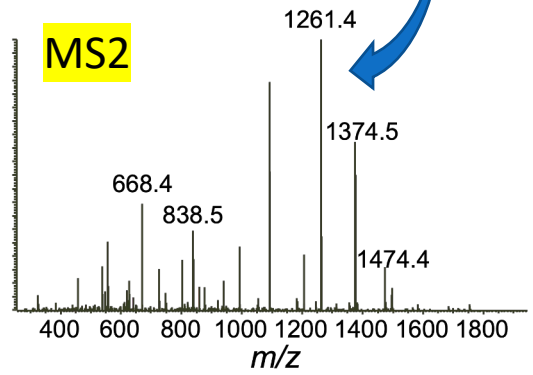


Separation of peptides

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

Data Analysis





>ipi|IPI0024789|IPI0024789.6
MDQHQHLNKTAESASSEKEKKTRR
CNGFKAMFLAALSFSYIAKALGGIIM
KISITQIERRFD…

Protein sequences

TAESASSEKE
AMFLAALSFSYIAK
…

Peptide sequences

Bottom-up Proteomics

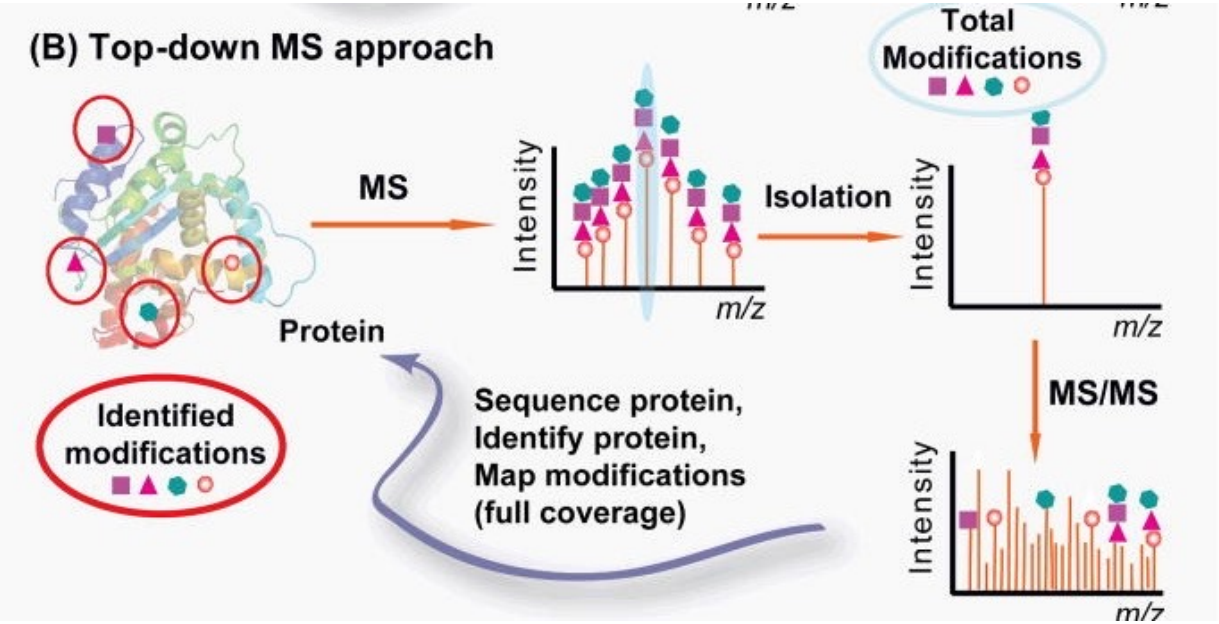
Advantages

- 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

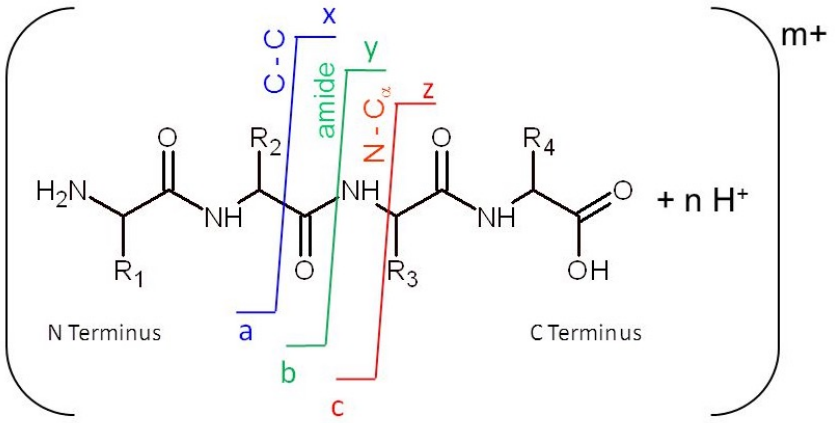
Top-down Proteomics



- 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)

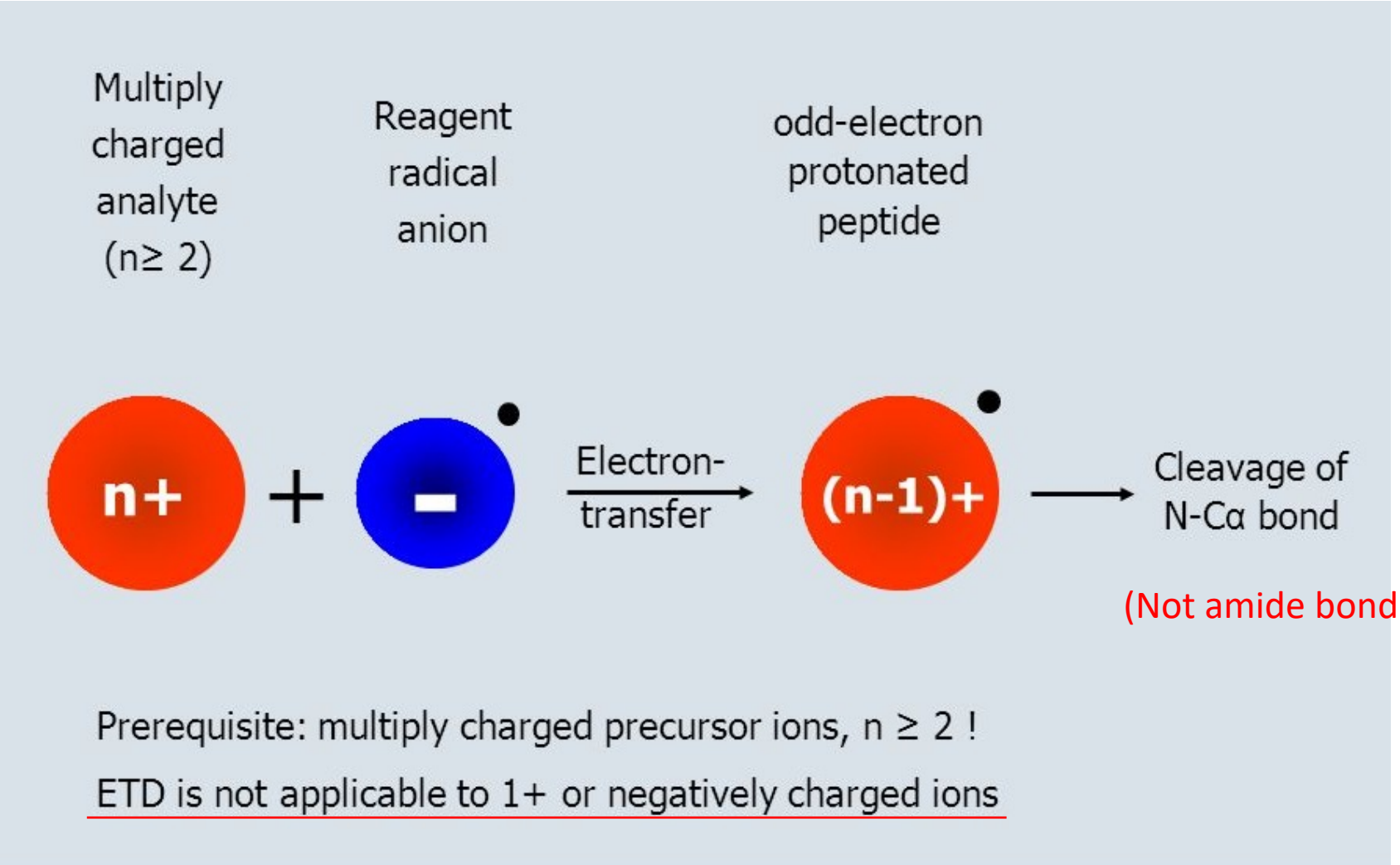
- Different mechanisms for fragmentation than CAD
- Free radical cleavage chemistries
- Ions can be fragmented more efficiently.
- Favors high charge states (+3 and higher)
- Efficient sequencing of peptides with PTMs



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

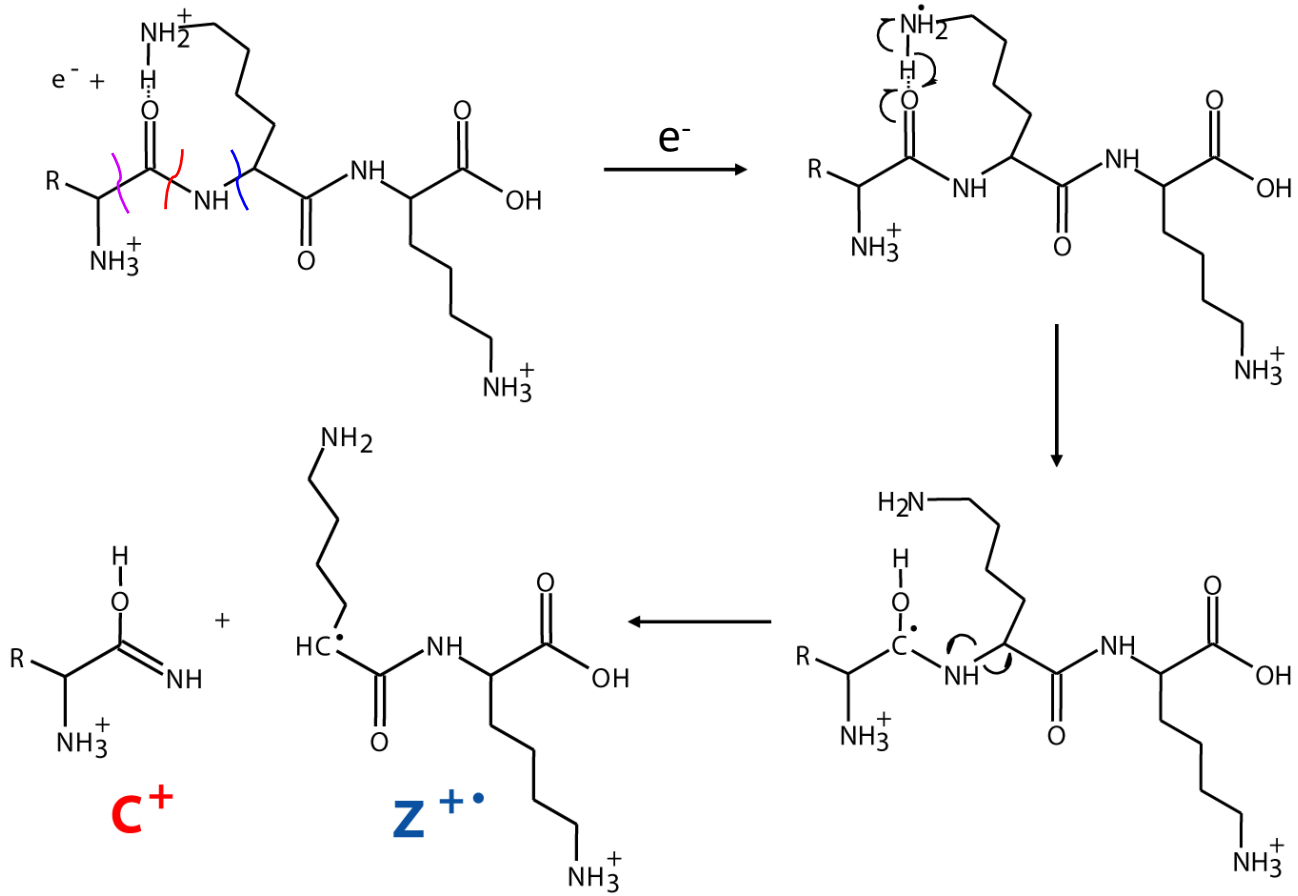


produces primarily *c* and *z* ions

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

Electron Transfer Dissociation

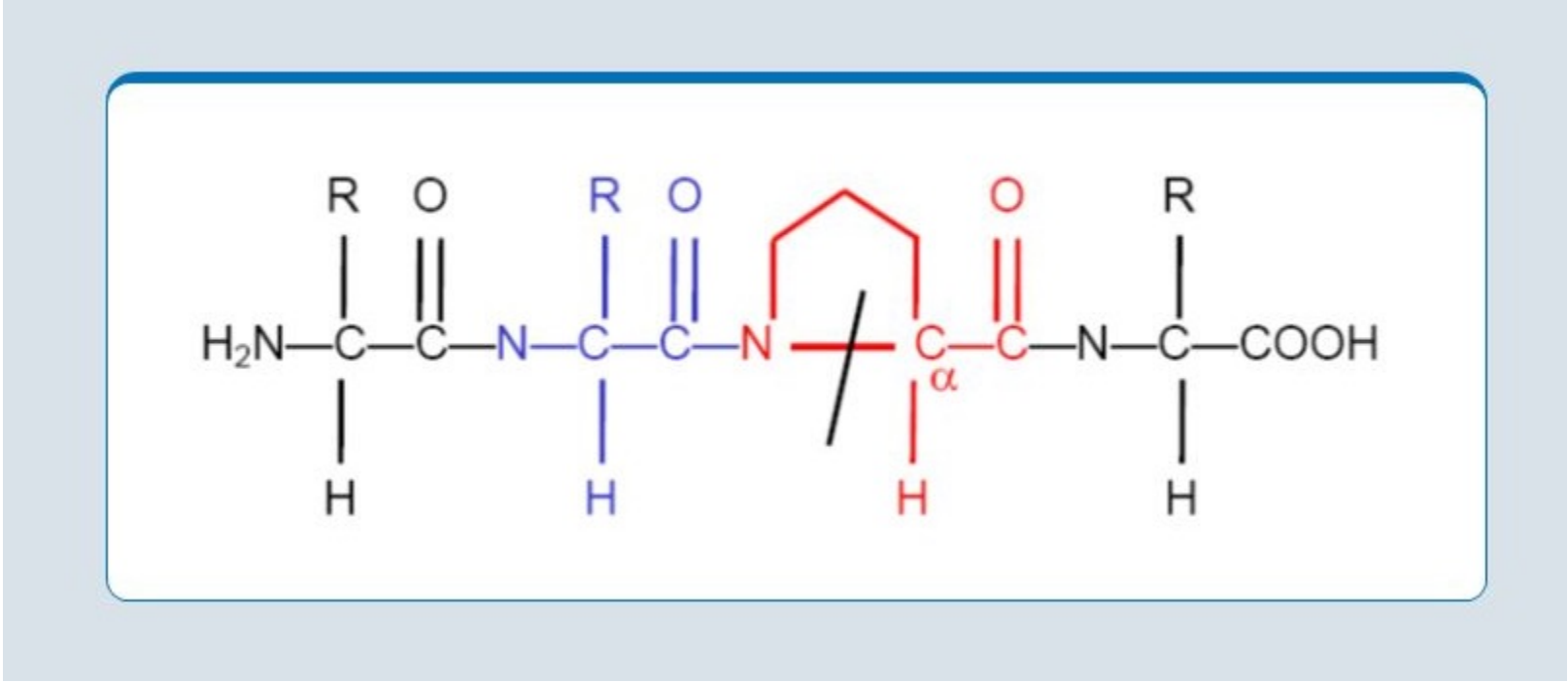
(produces primarily *c* and *z* ions)



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

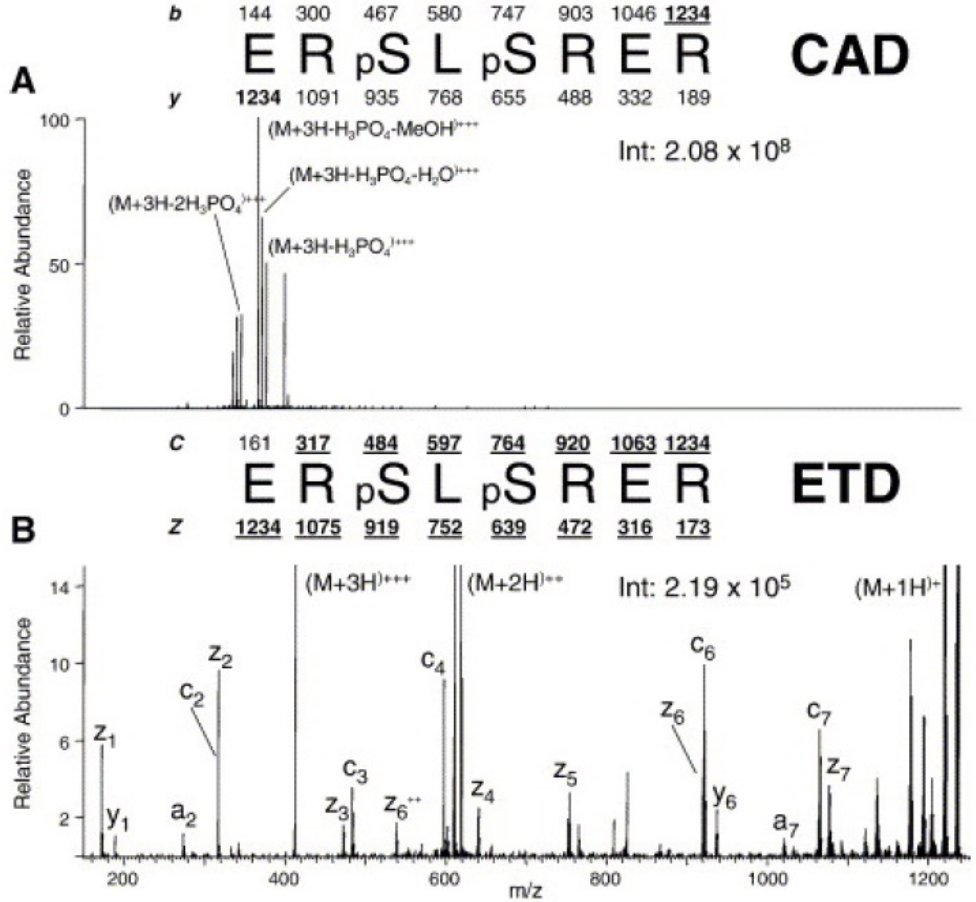


Fig. 3. Comparison of CAD vs. ETD spectrum of a phosphorylated peptide. Consecutive single-scan CAD vs. ETD mass spectrum comparison of phosphorylated peptides generated from a tryptic digest of human nuclear proteins recorded during a data-dependent analysis (nHPLC-μESI-MS/MS). All peptides were converted 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

Mikesh et al (2006) Biochem Biophys Acta 1764:1811-1822 – Review on ETD

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

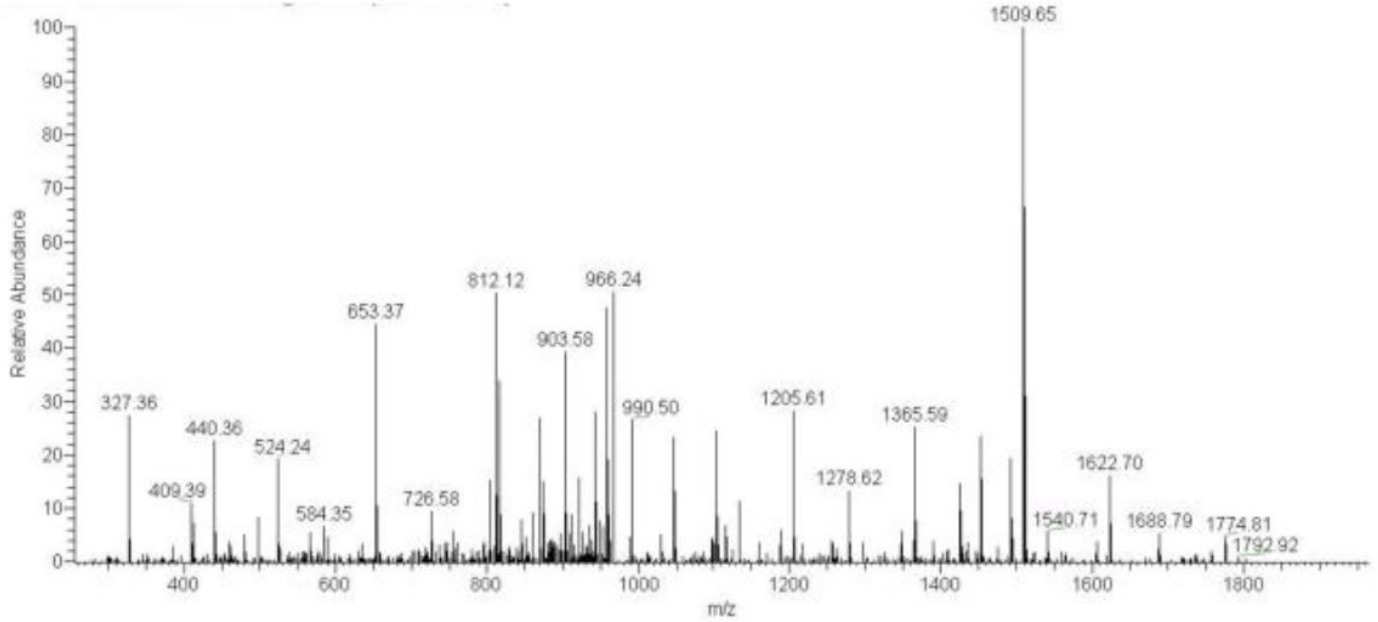
LECTURE 4: QUANTITATIVE MASS SPECTROMETRY-BASED PROTEOMICS

Quantitative Proteomics

- Quantitative proteomics aims at simultaneously quantitation of level differences between many proteins in different samples, not at measurement of their absolute concentrations.

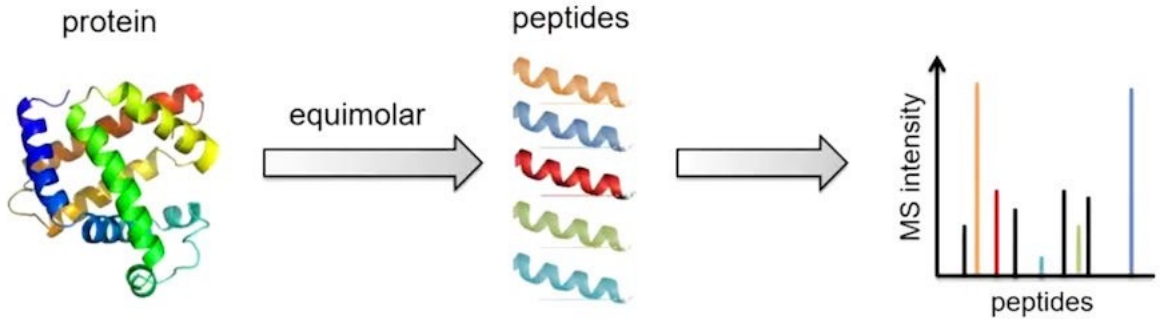
- Mass spectrum records a whole bunch of m/z

- BUT MS intensity does **NOT** tell us peptide abundance directly.



Why does MS^1 intensity not tell us peptide abundance directly?

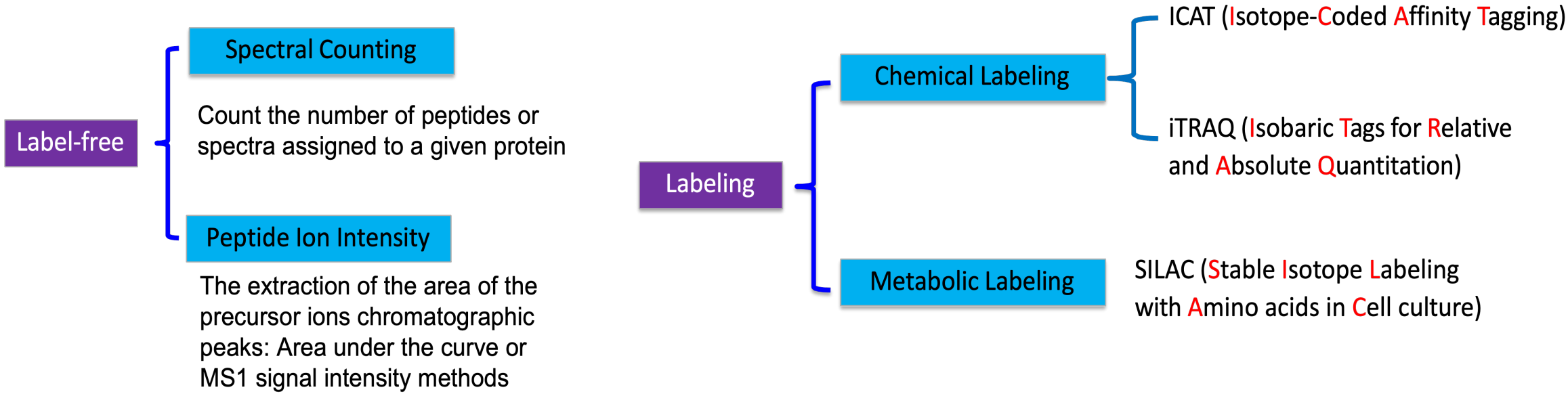
- There is a poor correlation between the amount of a peptide and the MS intensity in a single MS spectrum



- Many factors affect the MS intensity:
  - Peptide concentration
  - Day to day and long-term instrument reproducibility
  - Digestion efficiency
  - Recovery during sample preparation
  - Ionization efficiency
  - Instantaneous matrix effects

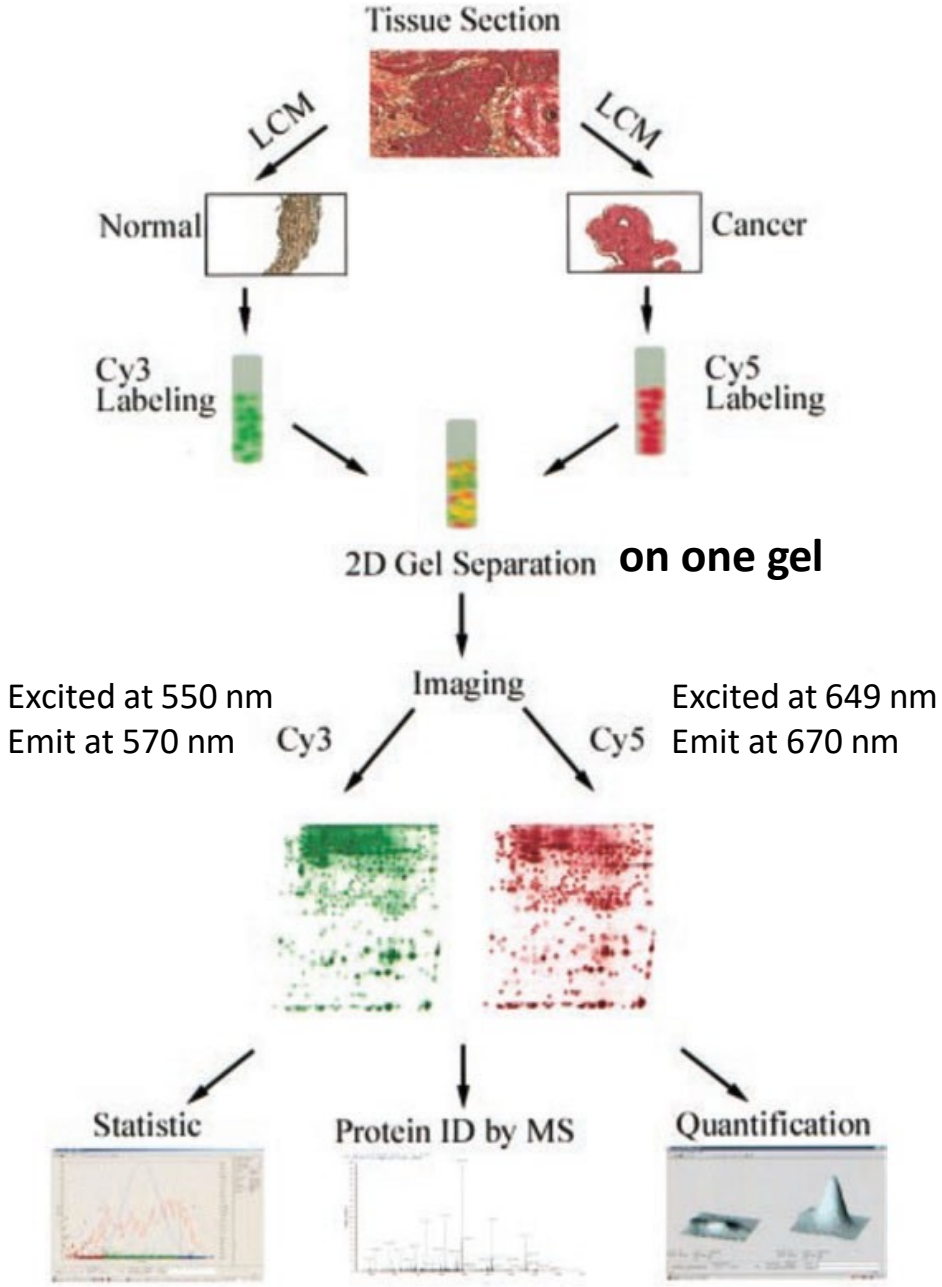
Quantitation Techniques in Proteomics

- Electrophoretic techniques
- Mass spectrometric techniques



Electrophoretic Technique

- It is often difficult to obtain reproducible separations by standard 2D PAGE.
- Quantitative proteomics utilizes fluorescent dyes
  - Ease of use
  - High sensitivity (1 ng)
  - Low background (not staining the gel)
  - Compatibility with MS
  - Broad range of linearity
- Differential gel electrophoresis (DIGE) is used in quantitative proteomics.



Zhou, Ge et al. *Molecular & Cellular Proteomics* 1 (2002): 117 - 123.