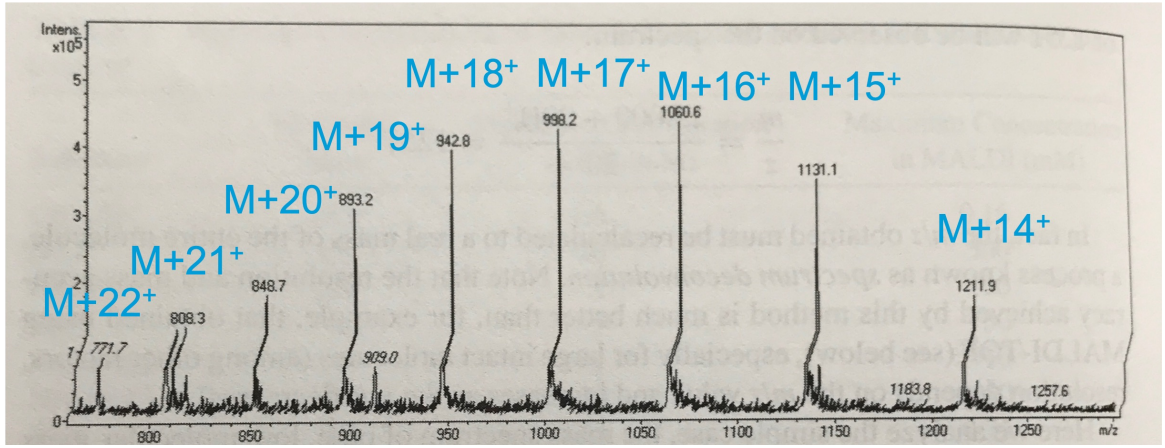
LECTURE 2-2: IDENTIFICATION OF PROTEINS IN COMPLEX MIXTURES - A MASS SPECTROMETRY APPROACH

Bio312

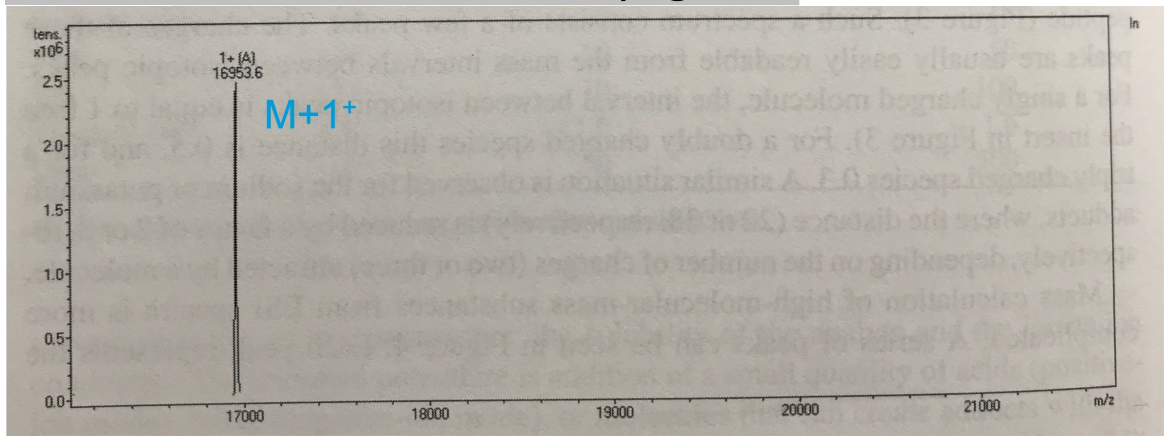
Instructor: Dr. Lanlan Han
E-mail: lanlan.han@xjtlu.edu.cn

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

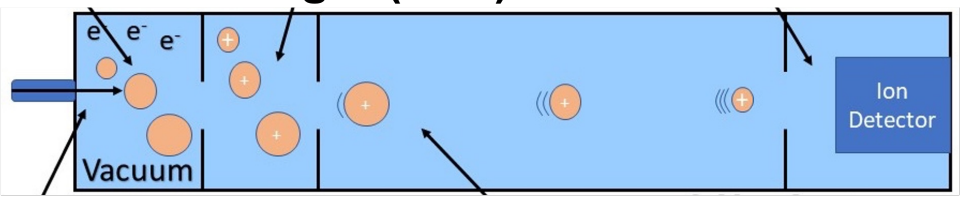
$$\mathbf{m/z = \frac{m+z}{z}}$$

myoglobin MW = 16953.6 – 1 = 16952.6 Da

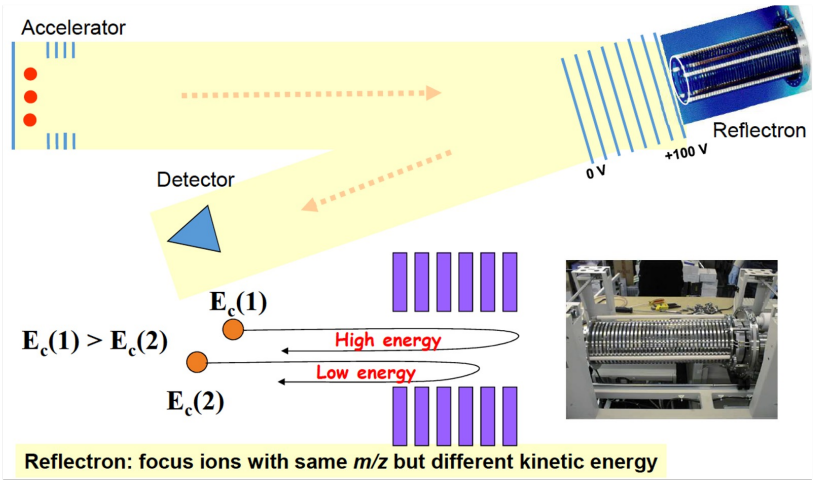
MS basics: Review

2. MS analyzers: Separate ions based on m/z (mass/charge) ratio

Time-of-flight (TOF)



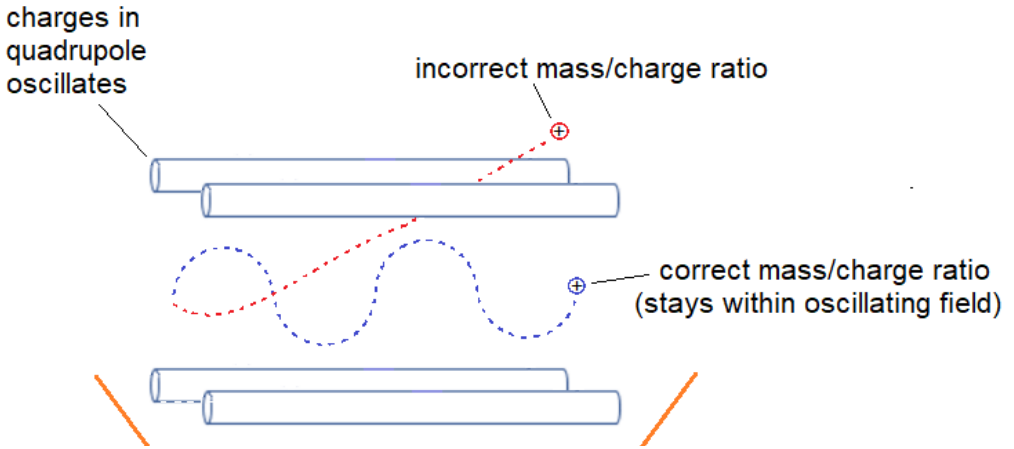
Low resolution,
Large molecules



High resolution.
Small molecules, <10 kDa

Large molecules can have metastable decay, which results in broad peak.

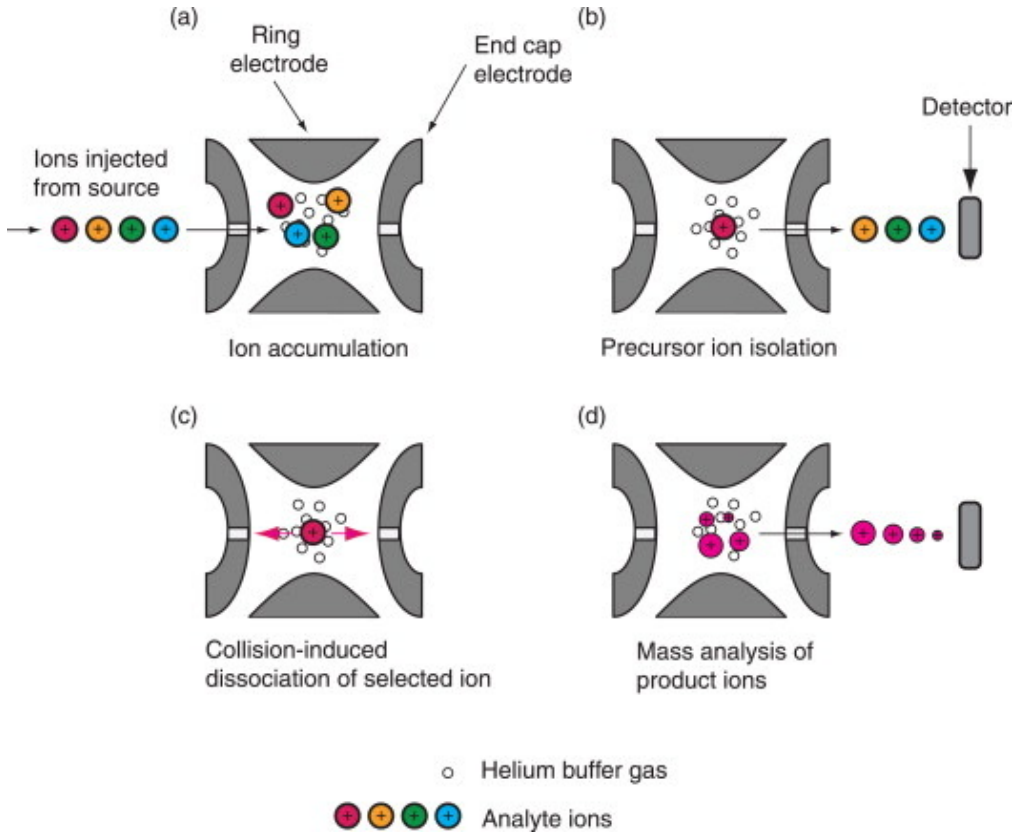
Quadrupole



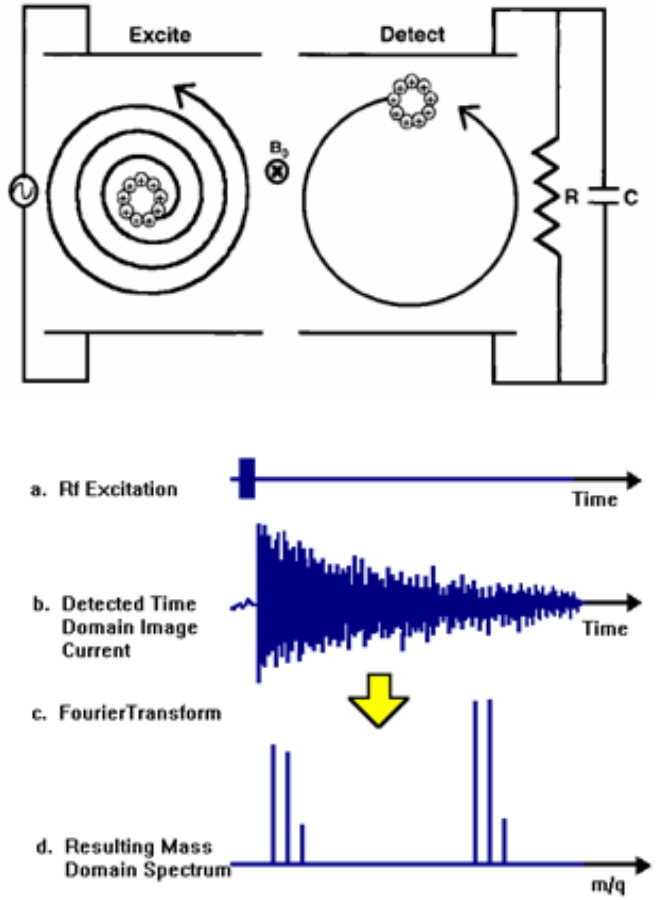
MS basics: Review

2. MS analyzers: Separate ions based on m/z (mass/charge) ratio

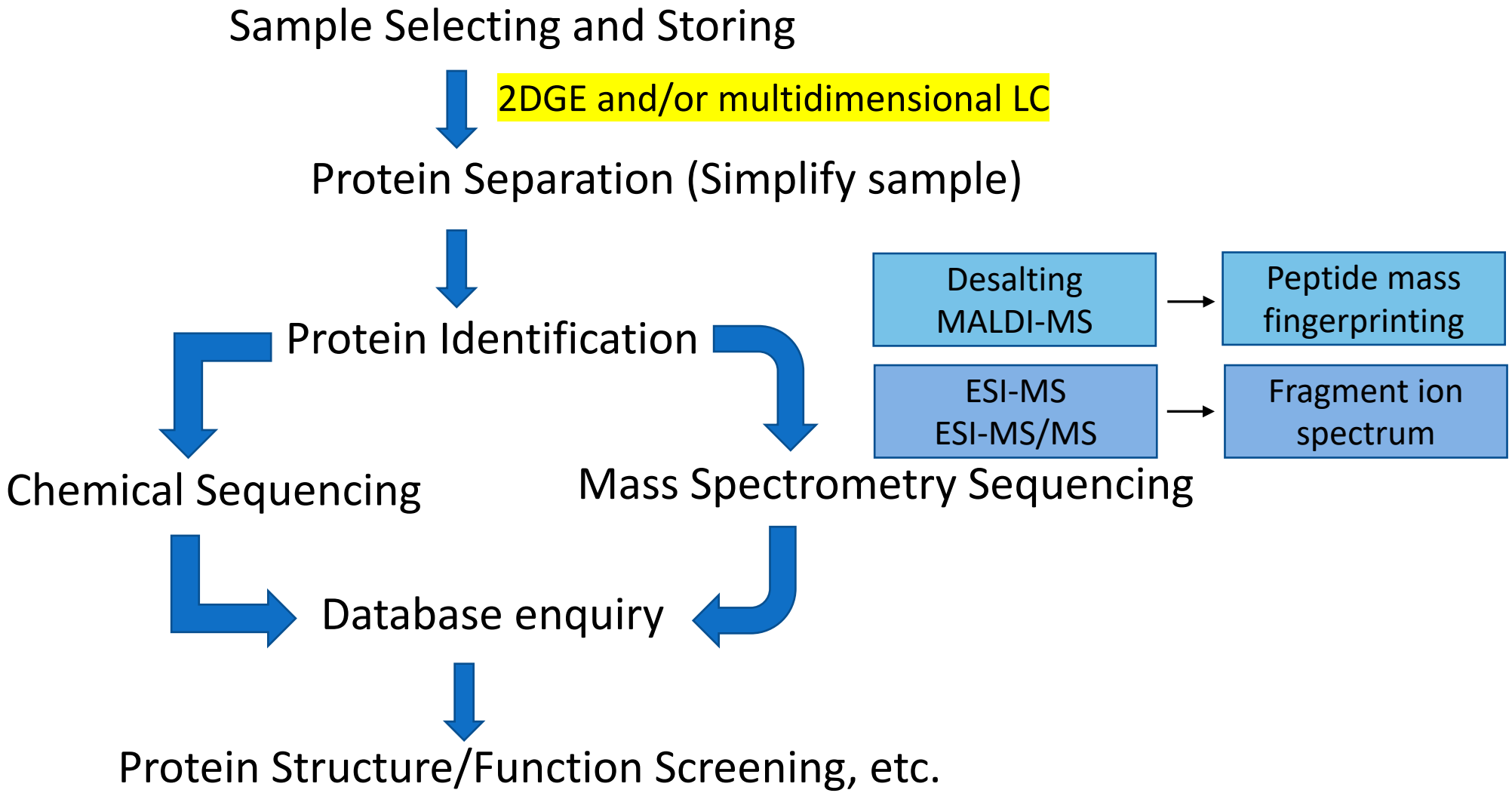
Iontrap



FT-ICR



General Workflow in Proteomics Analysis： Review

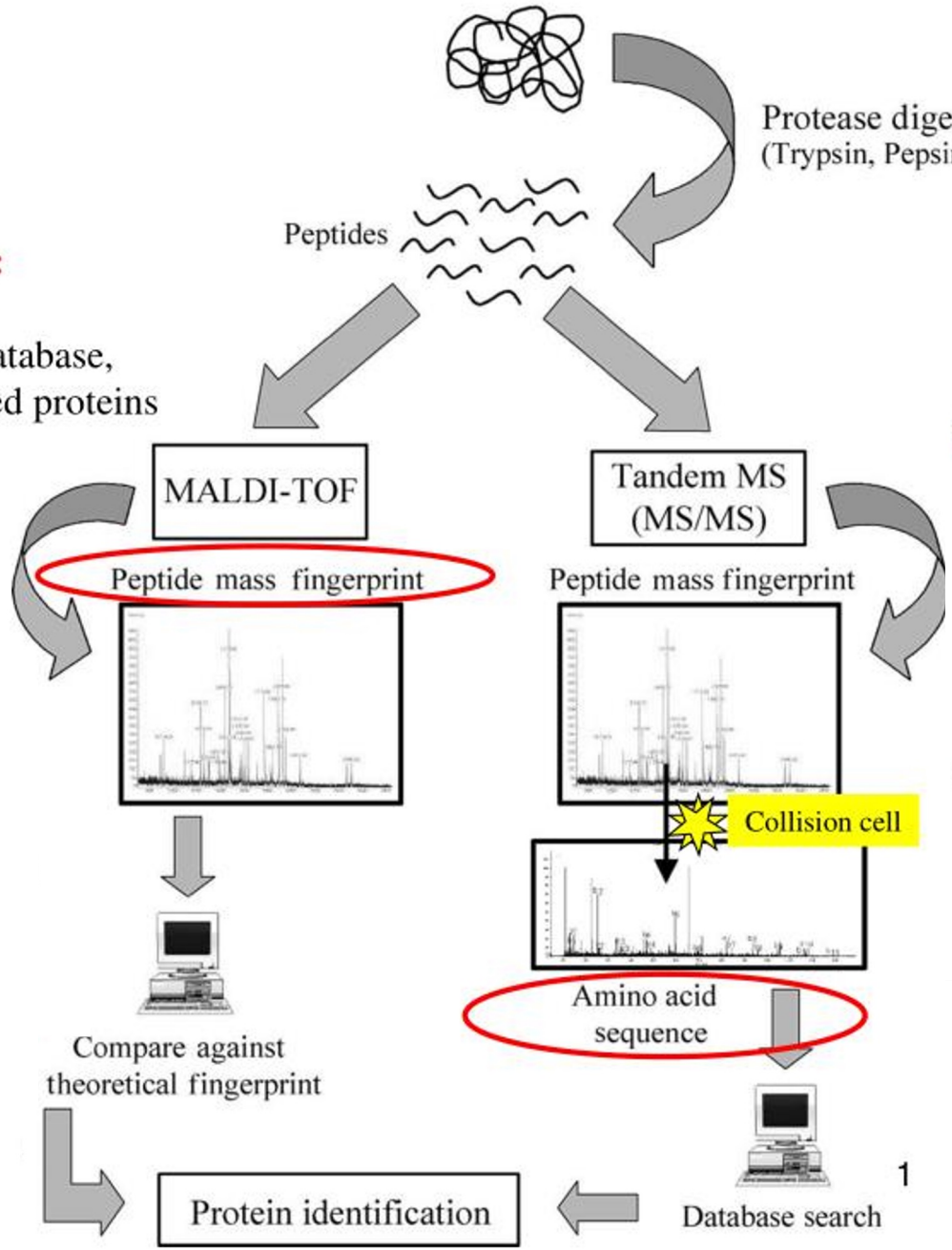


Single MS (peptide fingerprinting):

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

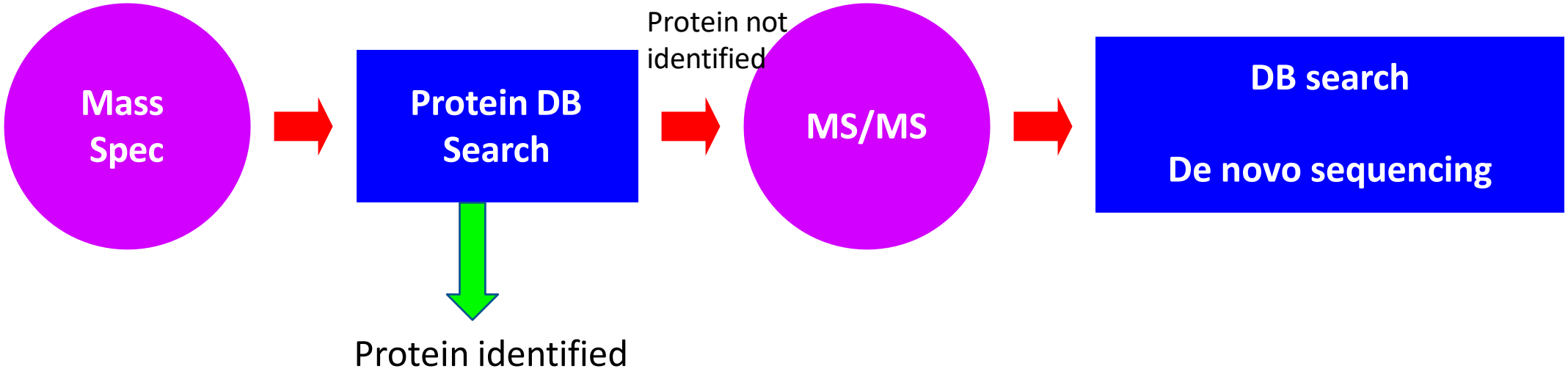
Breaks up peptide bond

Identifies each fragment based on m/z

CID: collision induced dissociation

each protein in the databases with the same specific cleavage and calculate the theoretical peptide masses.

Protein Identification and Characterization Map



Databases

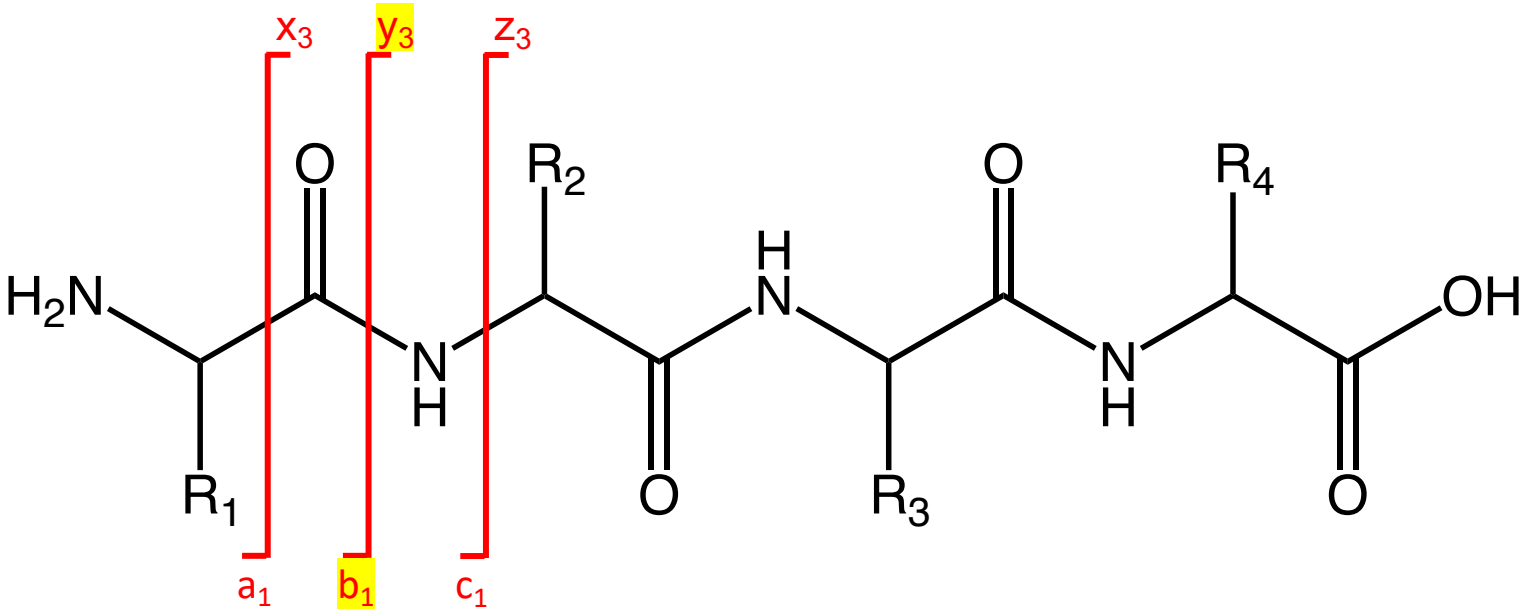
- Three components are required for database searching support of proteomics: MALDI or MS/MS data, the algorithms used to search protein databases, and the protein databases.
- A reality for database searching is that these protein databases are constantly changing, making database search results potentially obsolete as new entries are added that better fit the MALDI or MS data.
    - Even as genomes are completed, there is still flux as new coding regions are identified and novel mechanisms of increased translational complexity are better understood, such as alternative splice products, RNA editing, and ribosome slippage leading to novel, unexpected translation products.

Some Representative Internet Sources for Protein Identification from Mass Spectrometric Data

Program	Web Address
BLAST	http://www.ebi.ac.uk/blastall/
Mascot	http://www.matrixscience.com/cgi/index.pl?page=/home.html
MassSearch	http://cbrg.inf.ethz.ch/Server/ServerBooklet/MassSearchEx.html
MOWSE	http://srs.hgmp.mrc.ac.uk/cgi-bin/mowse
PeptideSearch	http://www.narrador.embl-heidelberg.de/GroupPages/PageLink/peptidesearchpage.html
Protein Prospector	http://prospector.ucsf.edu/
Prowl	http://prowl.rockefeller.edu/
SEQUEST	http://fields.scripps.edu/sequest/

2. Fragment Ion Analysis: Review

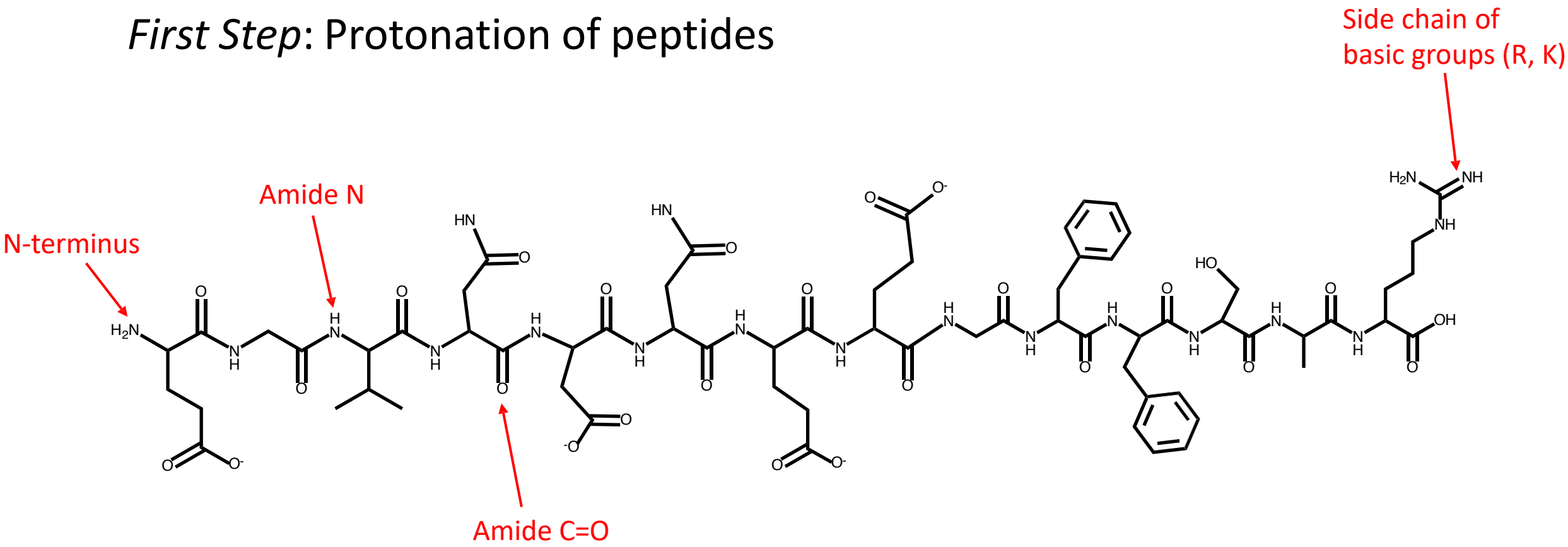
- Peptide can be fragmented by collision-induced dissociation (CID) (and other methods)
  - Collisions with neutral inert gas molecules (nitrogen, argon, etc.)
  - Charge stays on *either* the 'left' (a, b, or c) or 'right' (x, y, or z) side of cleavage
  - *Cleavage along the **CO-NH bond** is most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



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

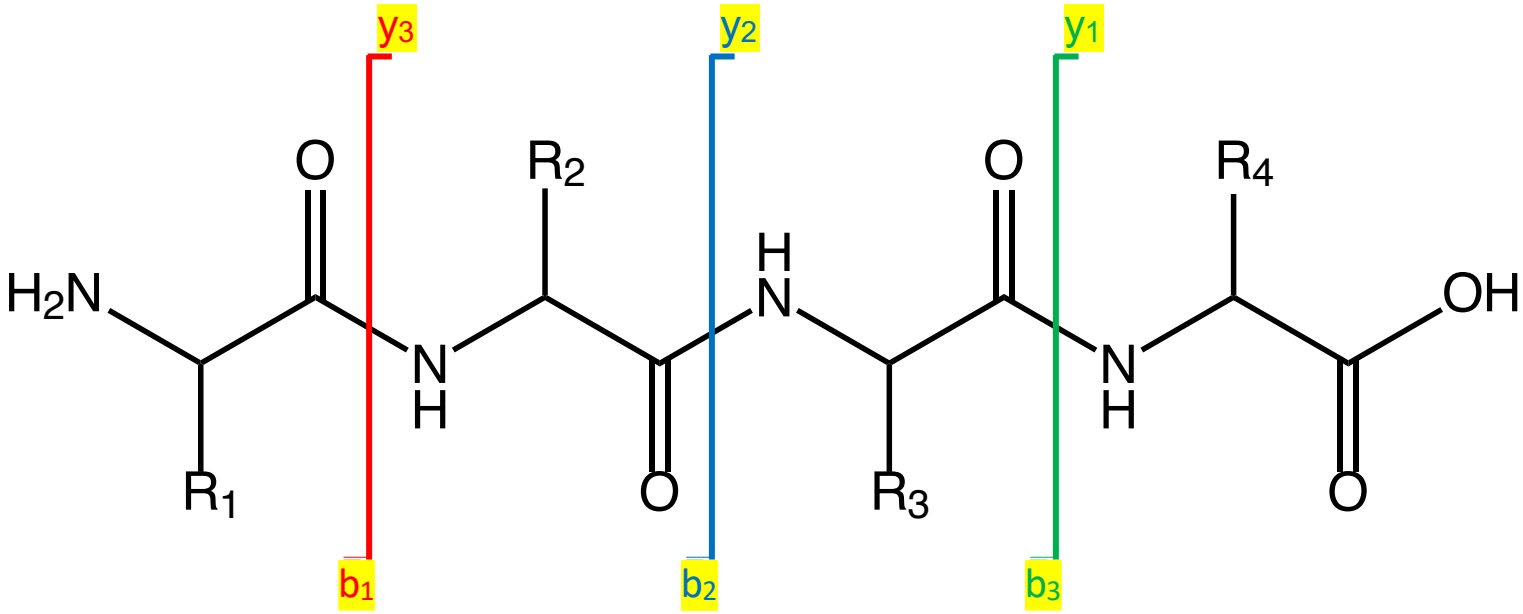
Peptides Fragment by CID

First Step: Protonation of peptides

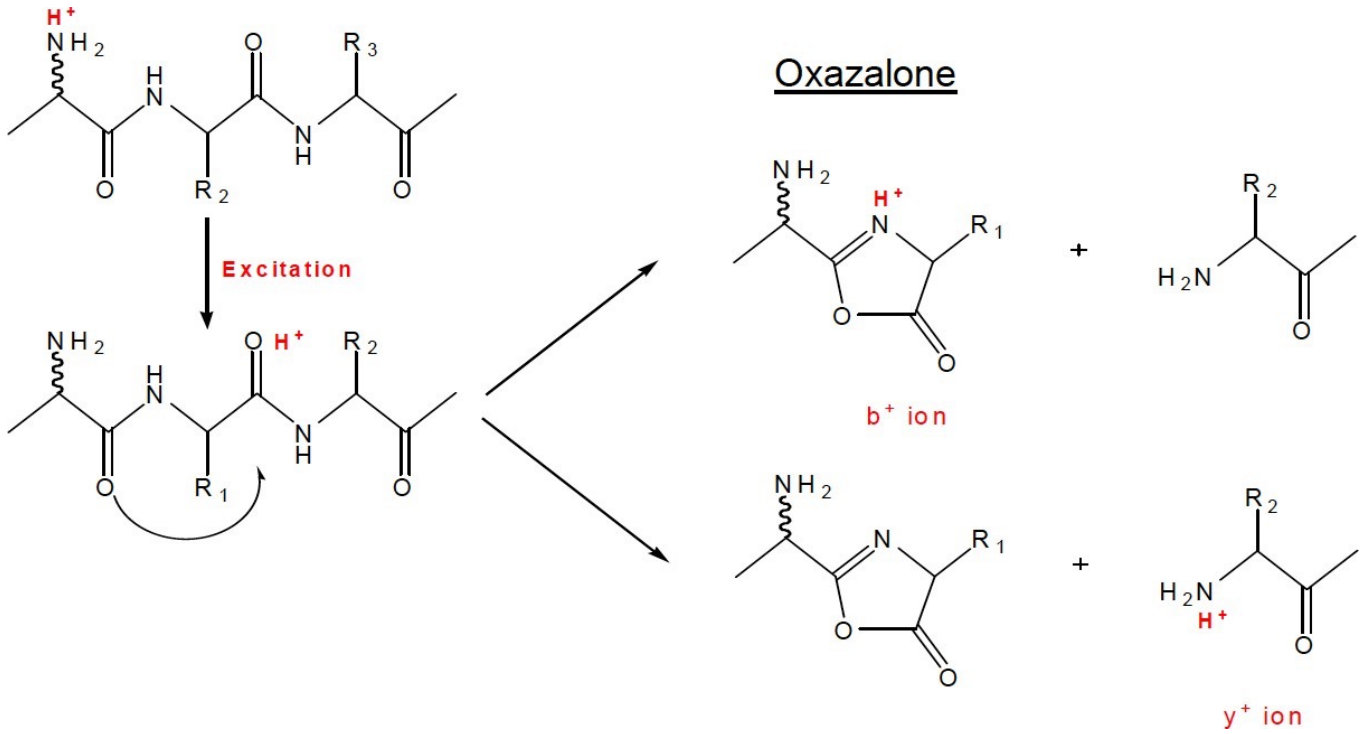


Peptides Fragment by CID

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,

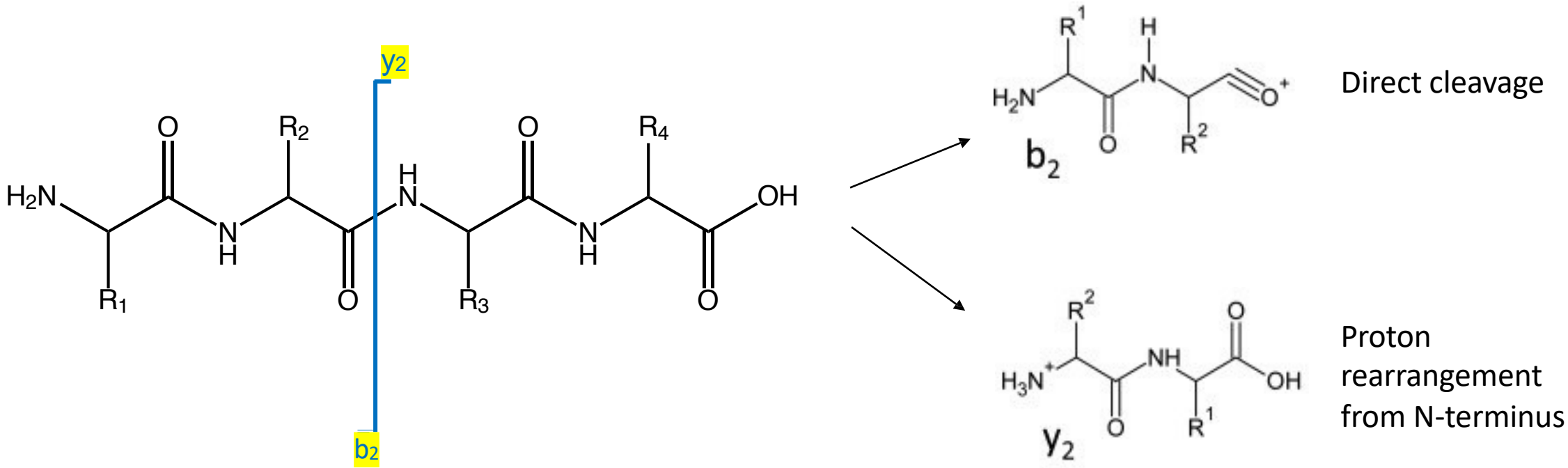
Singly charged N term ion ($+H^+$) and neutral C-term

OR

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

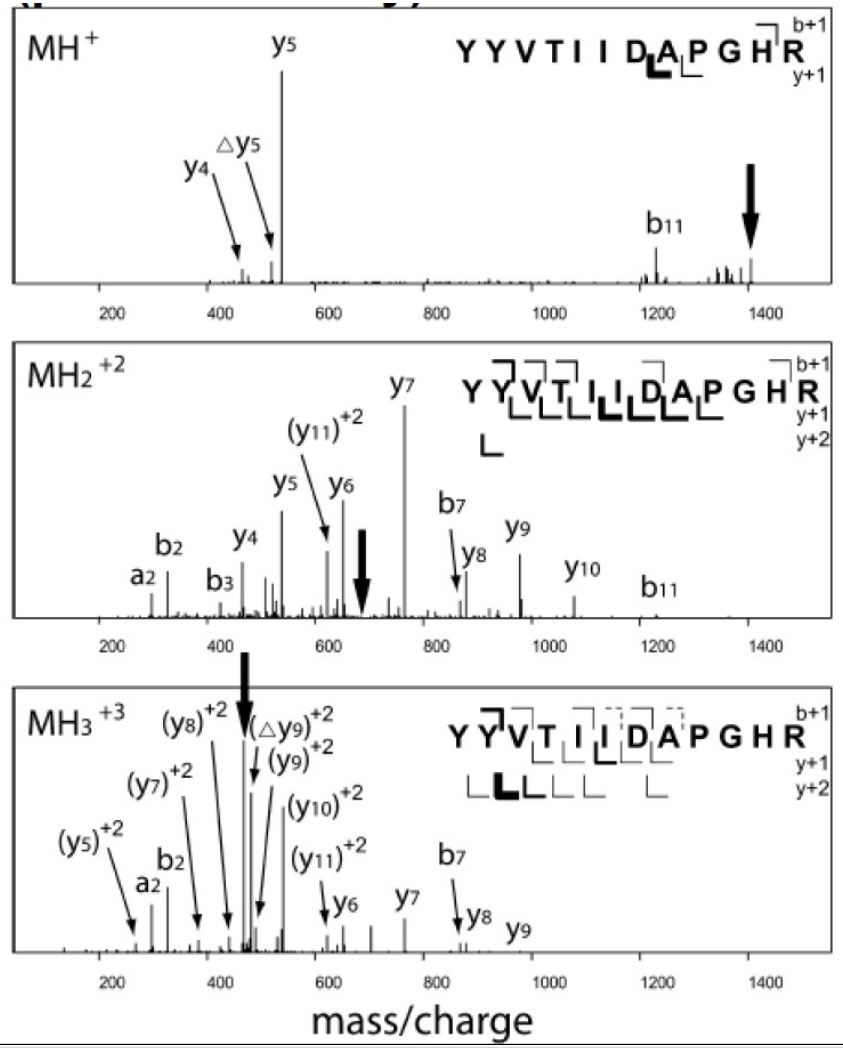
From Wysocki et al., J. Mass Spectrom. (2000) 35:1399

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

From Prof. William Old in University of Colorado - Boulder

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

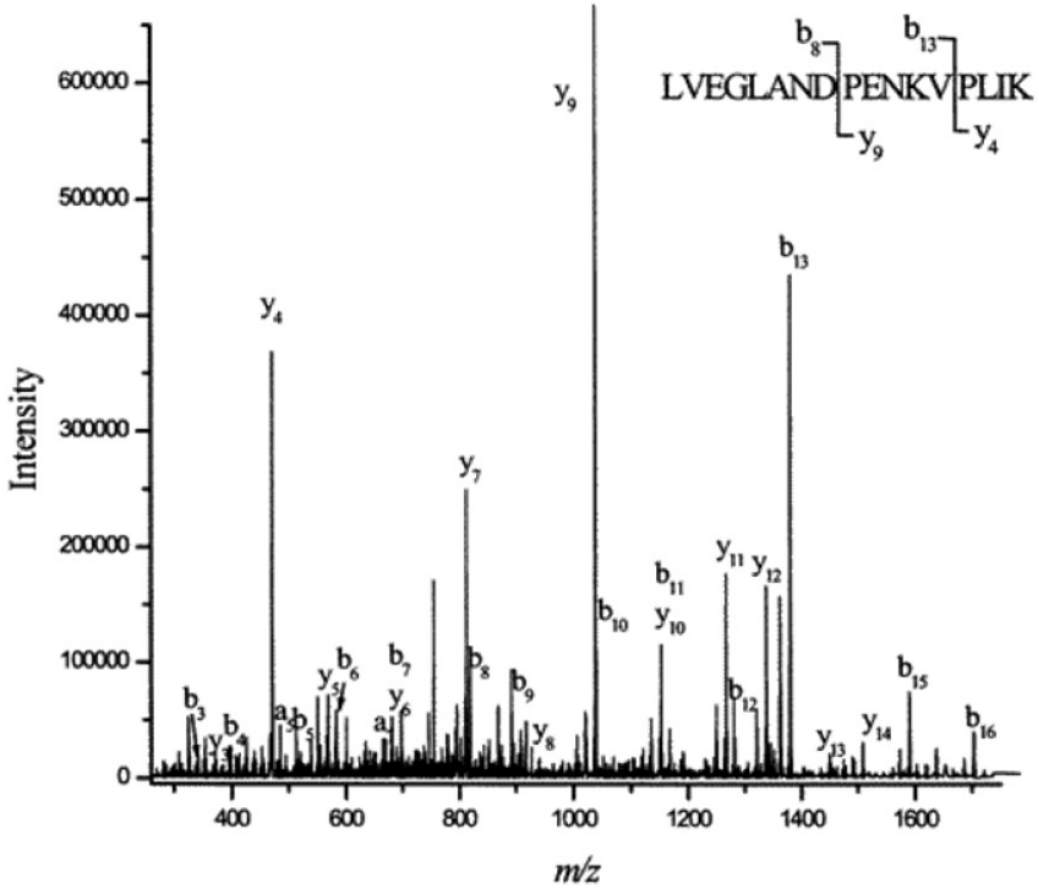
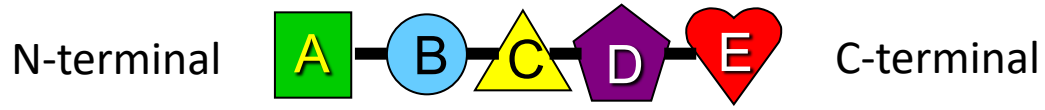


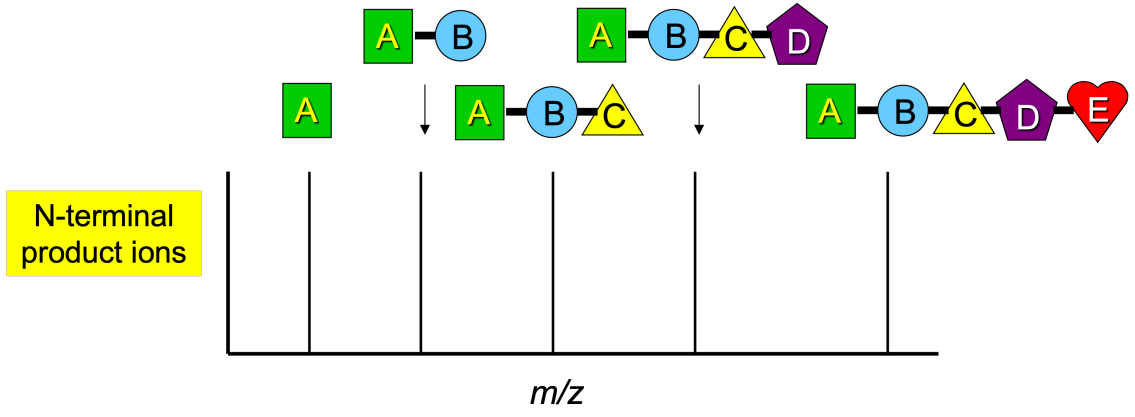
Figure 1 MS/MS spectrum of the peptide [LVEGLANDPENKVPLIK + 2H]2+ acquired by CID in an ion trap. Although many peaks are a-, b-, and y-sequence ions, many other peaks are unidentified.

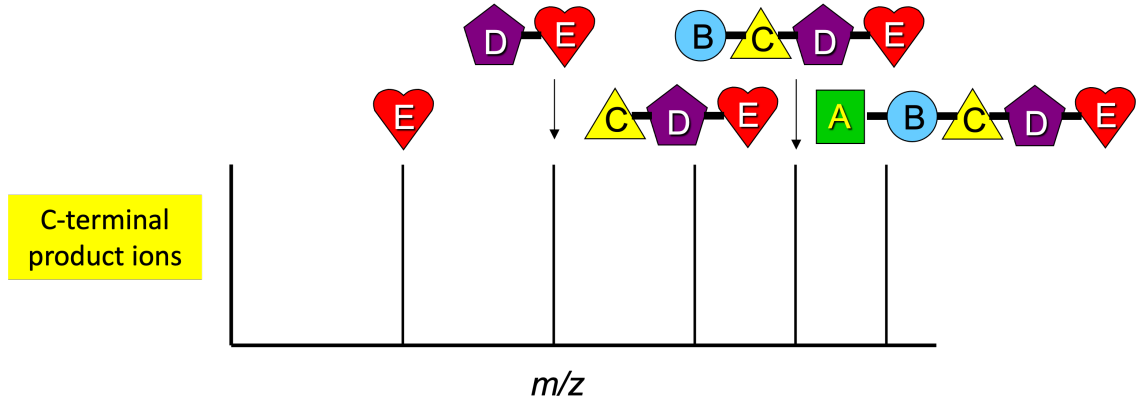
Breci et al 2003 Anal. Chem., 75 (9), 1963 -1971

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…



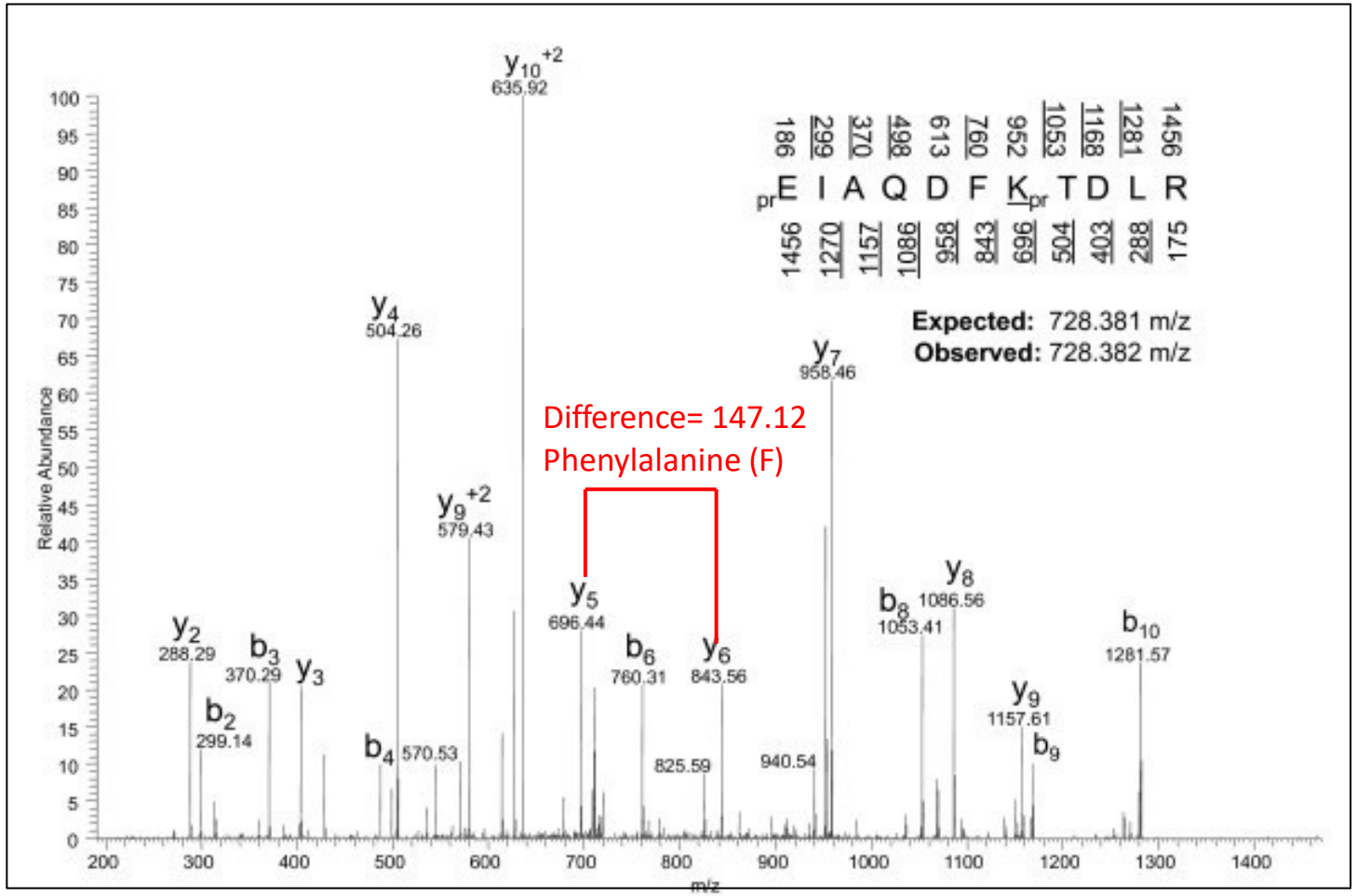


Fragmentation Results in a Peptide “Ladder”

Peptide: A-B-C-D-E

	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)

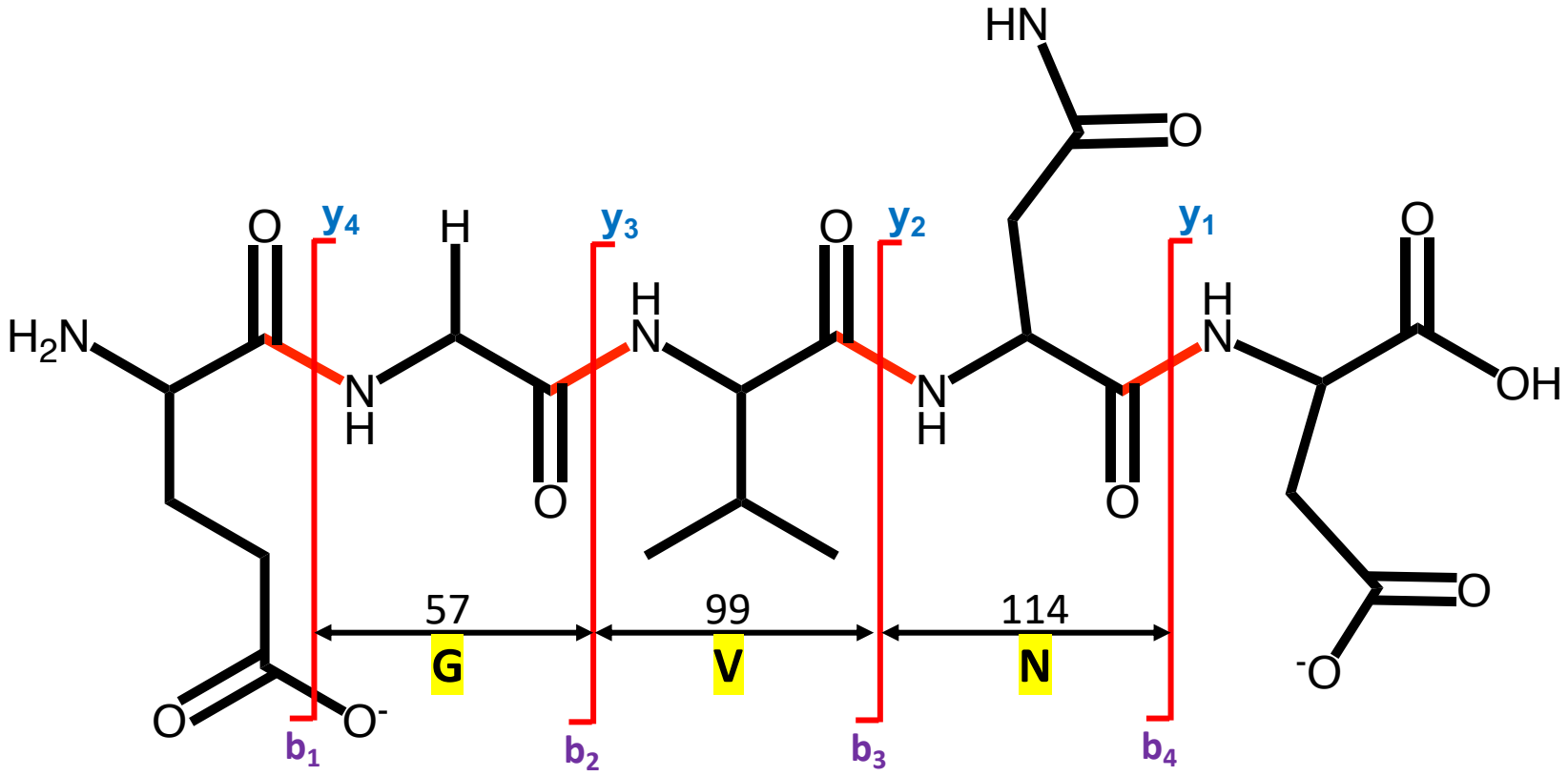


- Mixture of b ions and y ions
- MS/MS of 2^+ charged tryptic peptides yield (often) 1^+ charged product ions (but 2^+ charged products can be observed as well)
- Not all b ions or y ions are visible

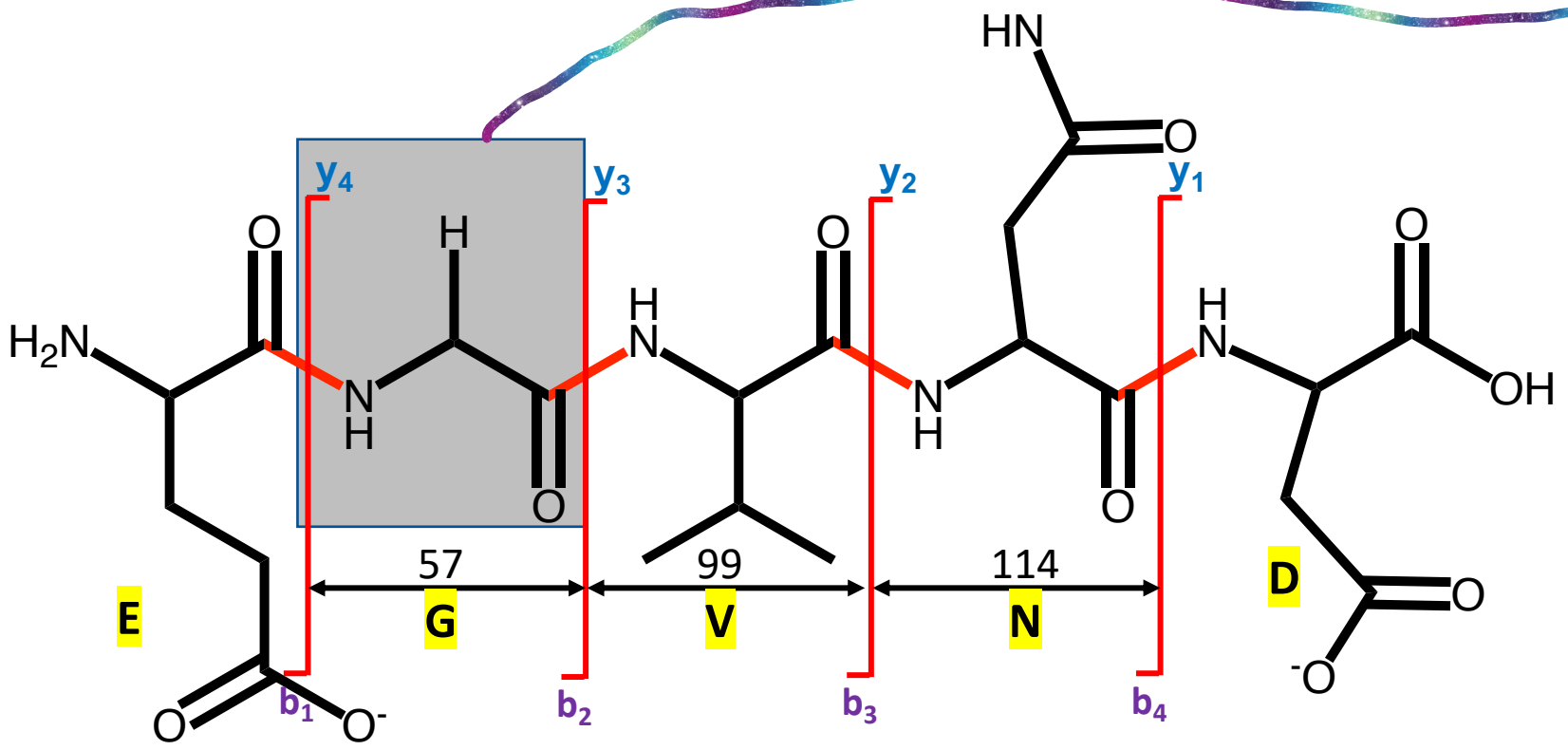
The mass of the precursor is 1454 (the observed ion was doubly charged)
728.382 x 2 - 2 = 1454.764 Da
Precursor ion ($M+2H^+$) is 1456.764 Da.

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

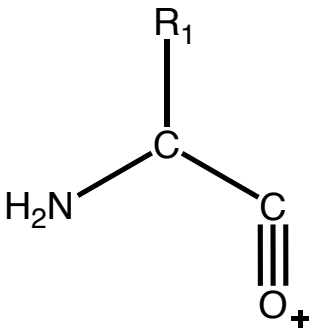
- Amino acid sequence can be deduced by the **Δmass** between adjacent y ion peaks **or** adjacent b ion peaks



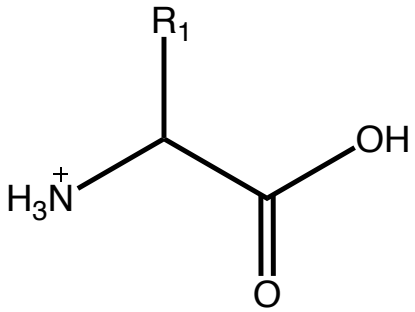
Code (1 letter)	Monoisotopic mass
G	57.021 47
A	71.037 12
S	87.032 03
P	97.052 77
V	99.068 42
T	101.047 68
C	103.009 19
I	113.084 07
L	113.084 07
N	114.042 93
D	115.026 95
Q	128.058 58
K	128.094 97
E	129.042 60
M	131.040 49
H	137.058 91
F	147.068 42
R	156.101 12
Y	163.063 33
W	186.079 32



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



b_1 ion

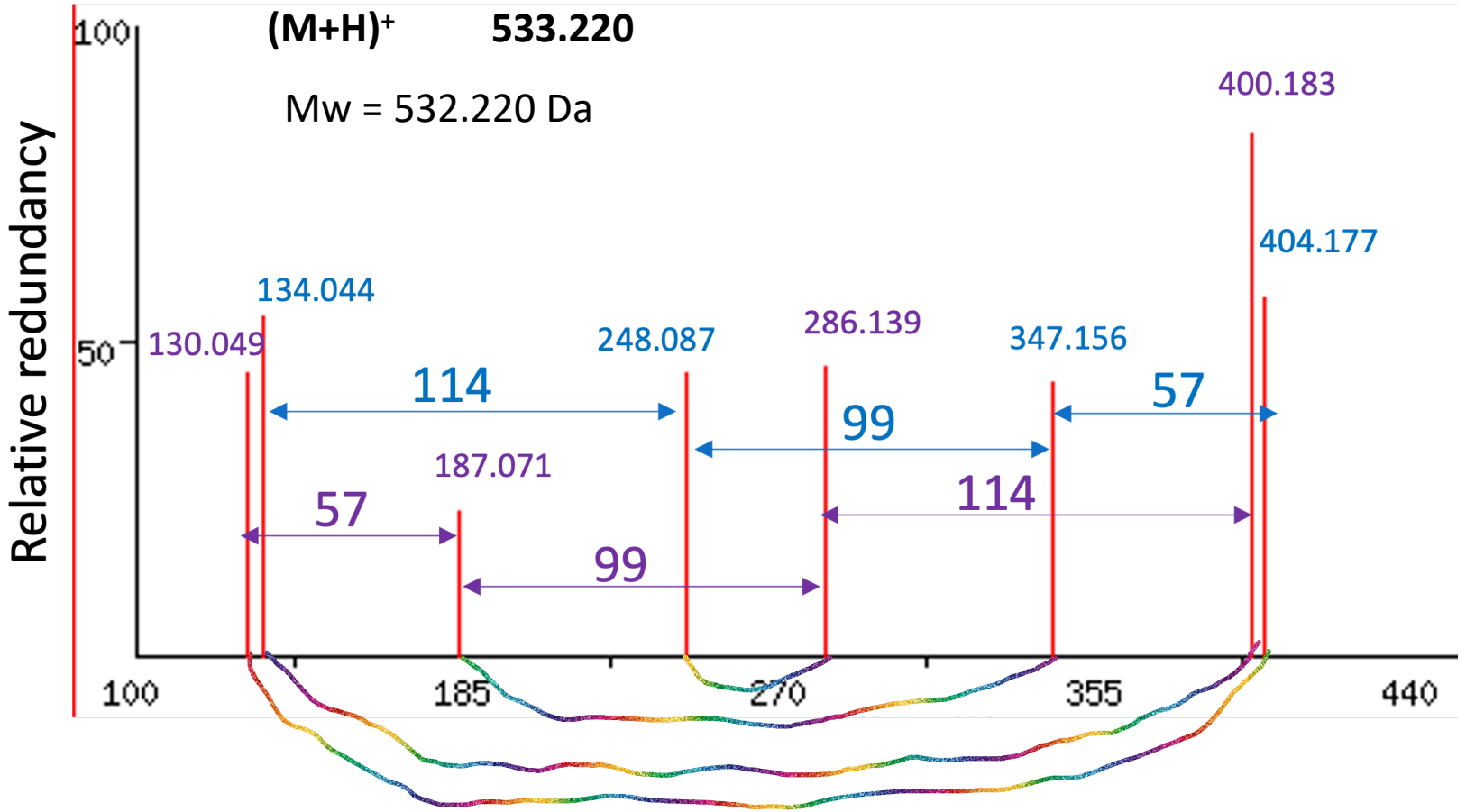


y_1 ion

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

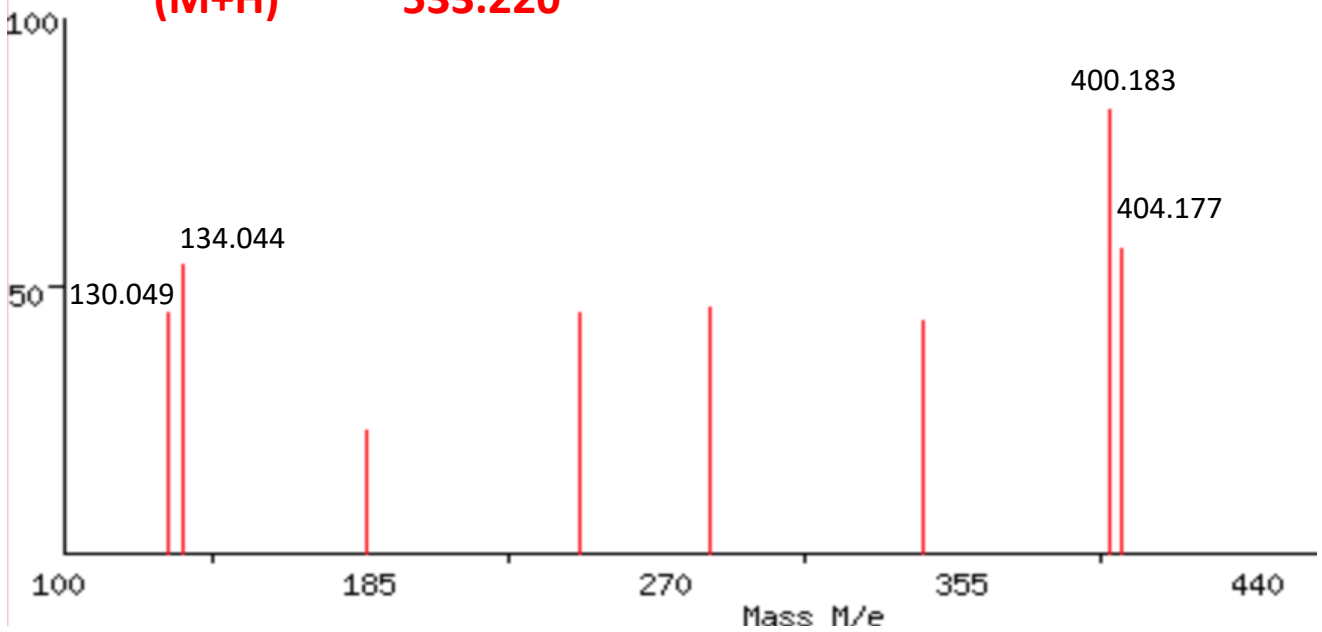
130.049 – 1 = 129.049 **E** on N terminus

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

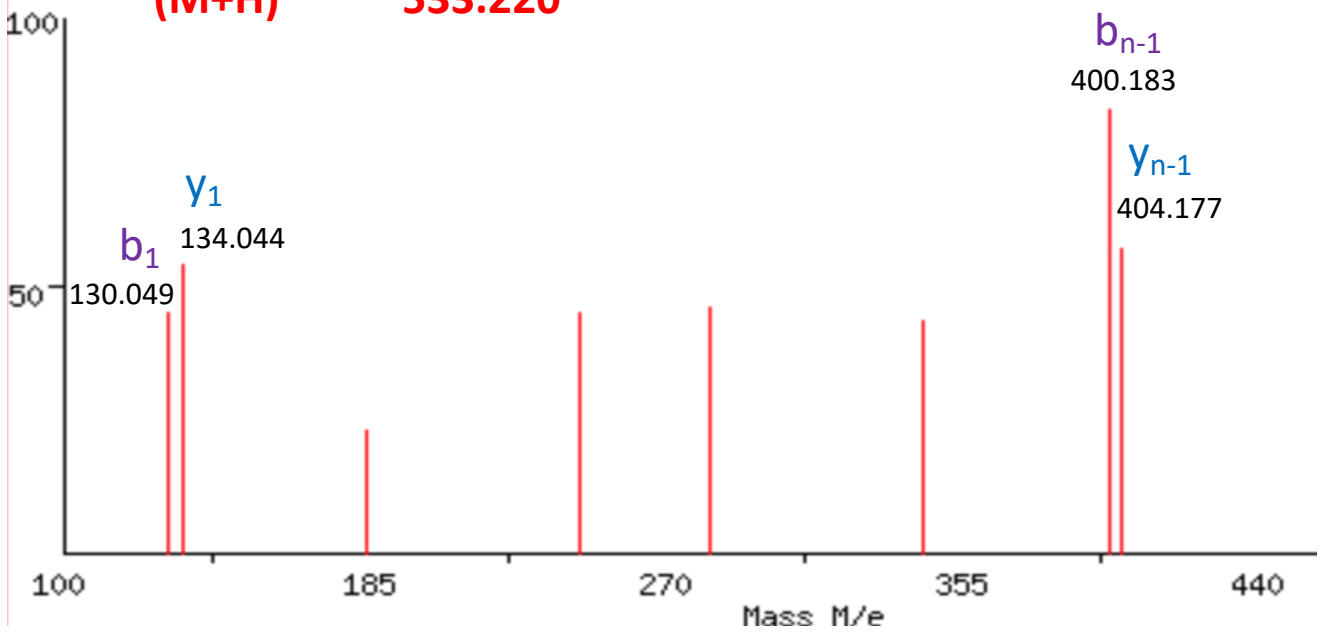
130.049 – 1 = 129.049 **E** on N terminus

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



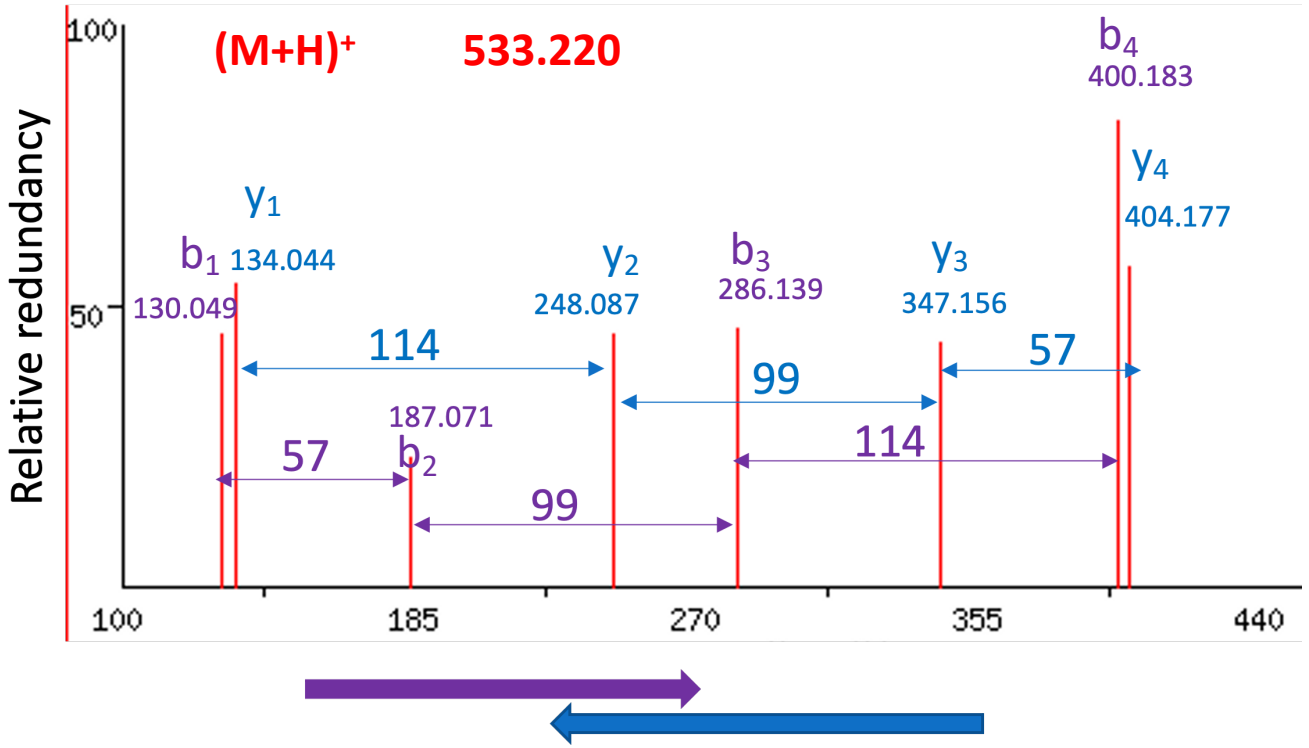
M - y_{n-1} ion + 1 = mass of 1st residue on N terminus

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

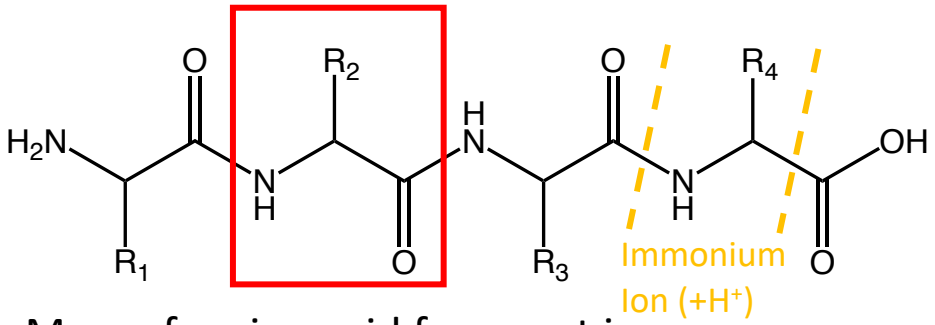
EGVND



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.

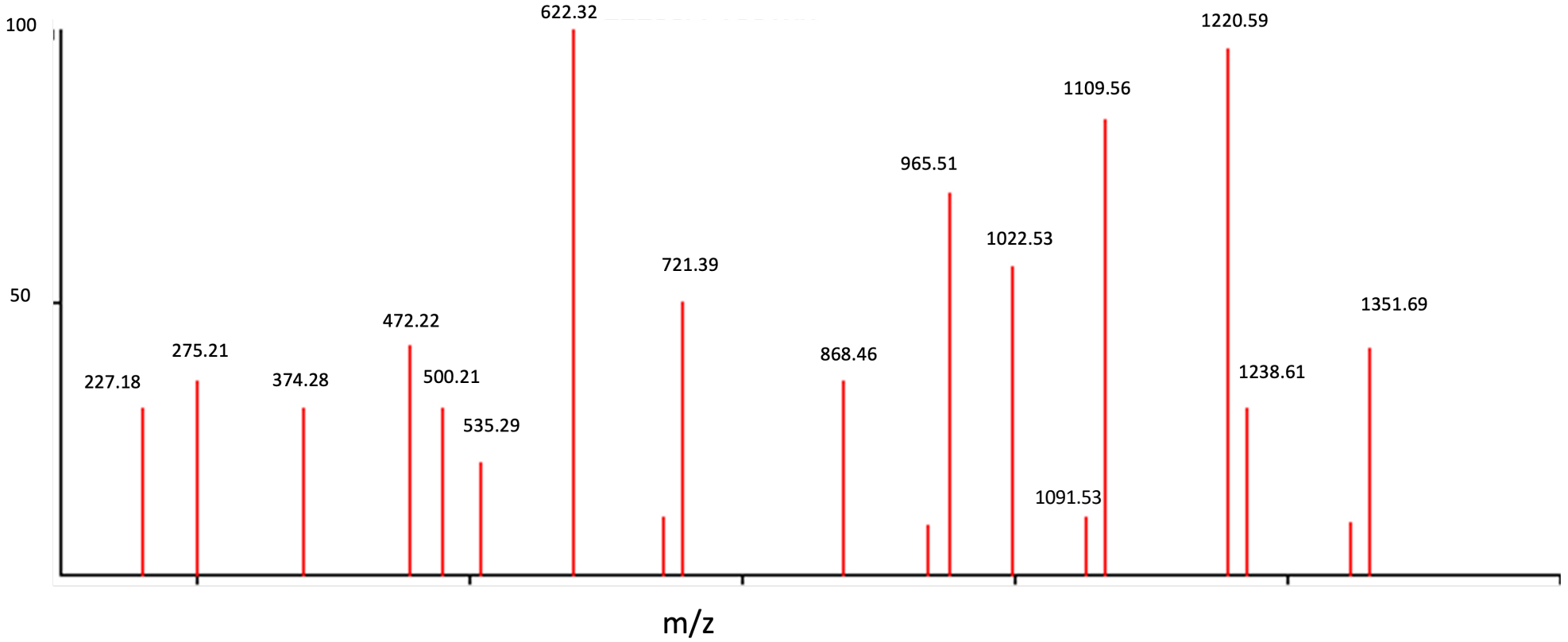
$[M+H]^+$ = 1464.7693

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)

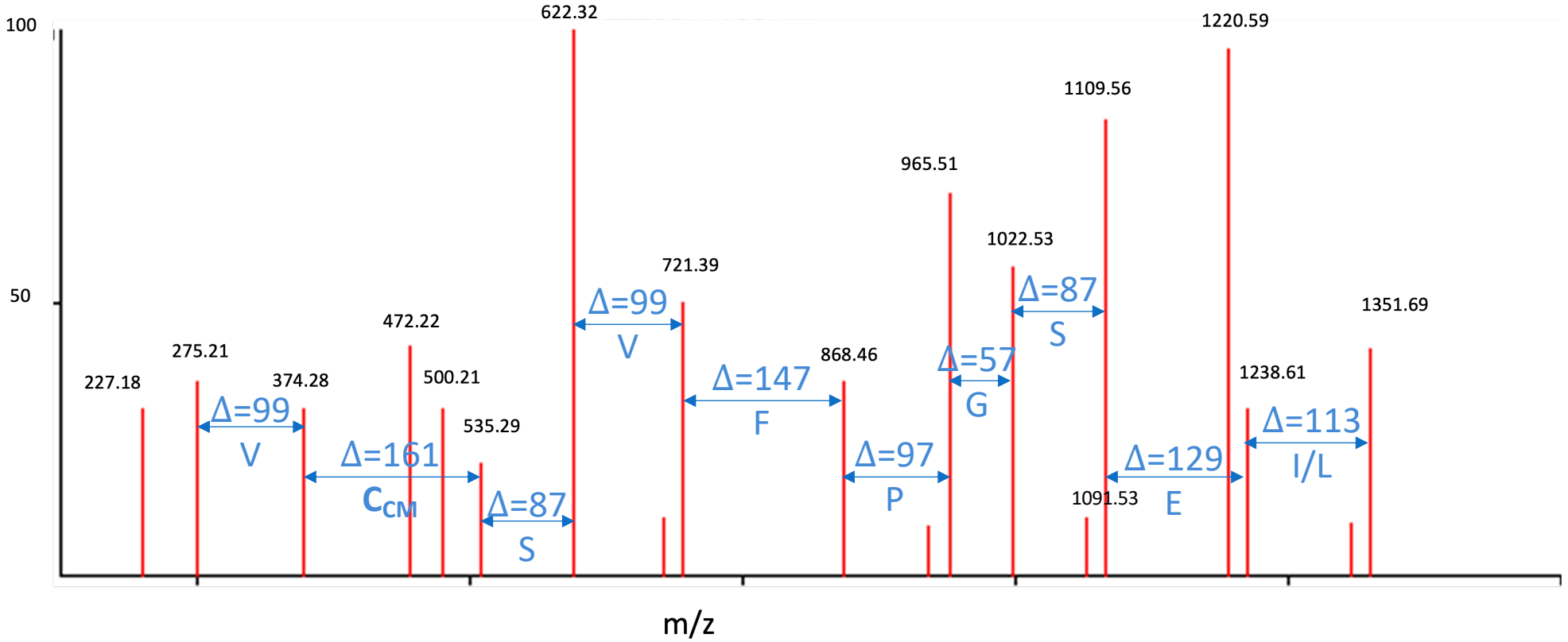
$[M+H]^+$ = 1464.7693
So, Mw = 1463.7693 Da

- 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

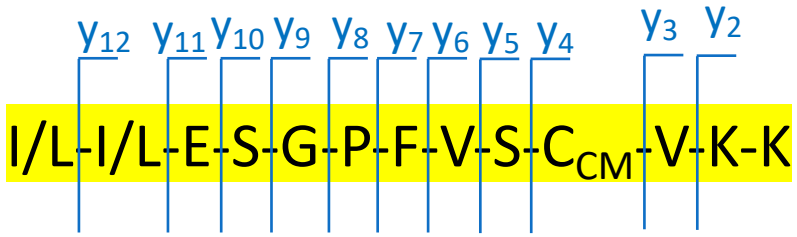
I/L-I/L-E-S-G-P-F-V-S-C_{CM}-V-…

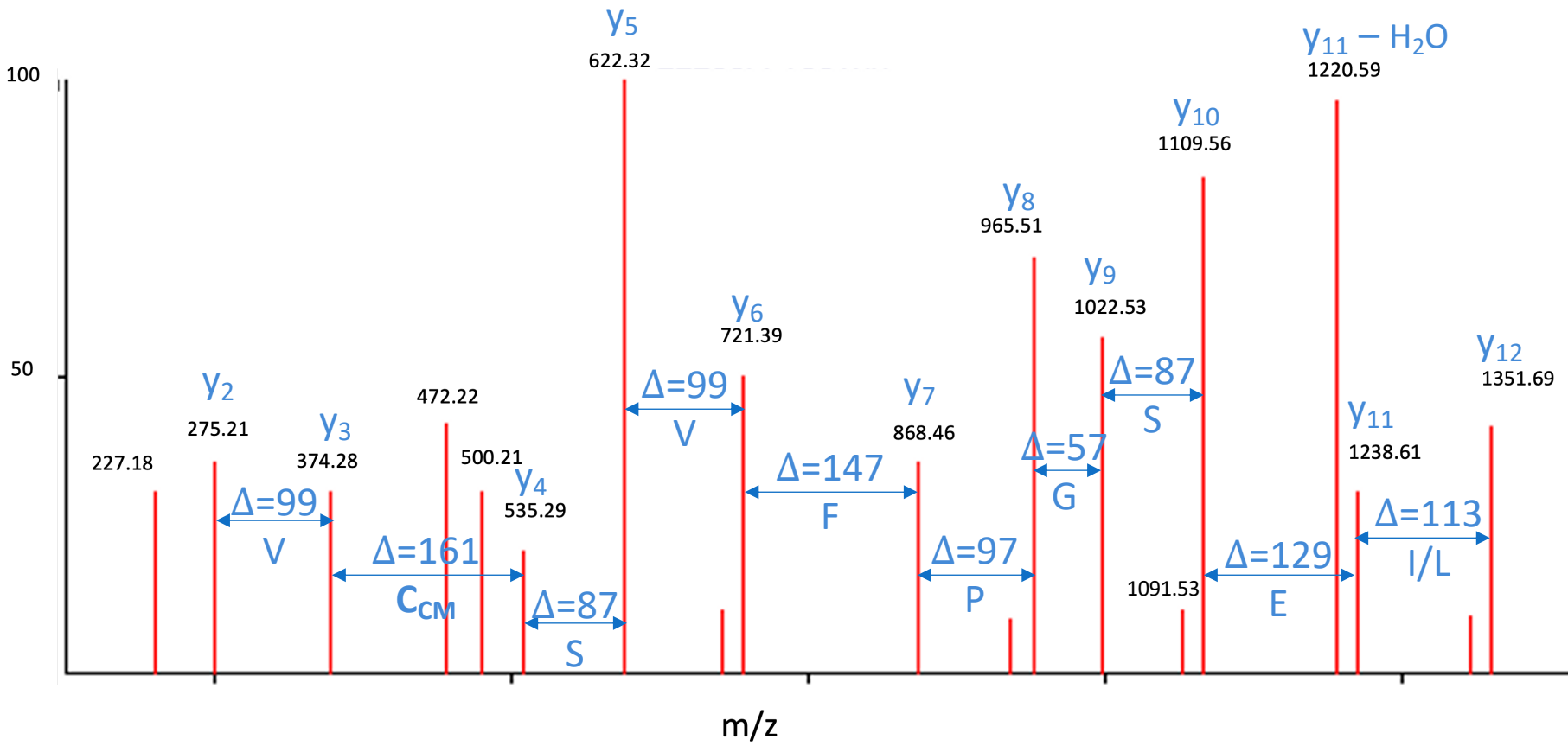


Code (1 letter)	Monoisotopic mass
G	57.021 47
A	71.037 12
S	87.032 03
P	97.052 77
V	99.068 42
T	101.047 68
C	103.009 19
I	113.084 07
L	113.084 07
N	114.042 93
D	115.026 95
Q	128.058 58
K	128.094 97
E	129.042 60
M	131.040 49
H	137.058 91
F	147.068 42
R	156.101 12
Y	163.063 33
W	186.079 32

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.





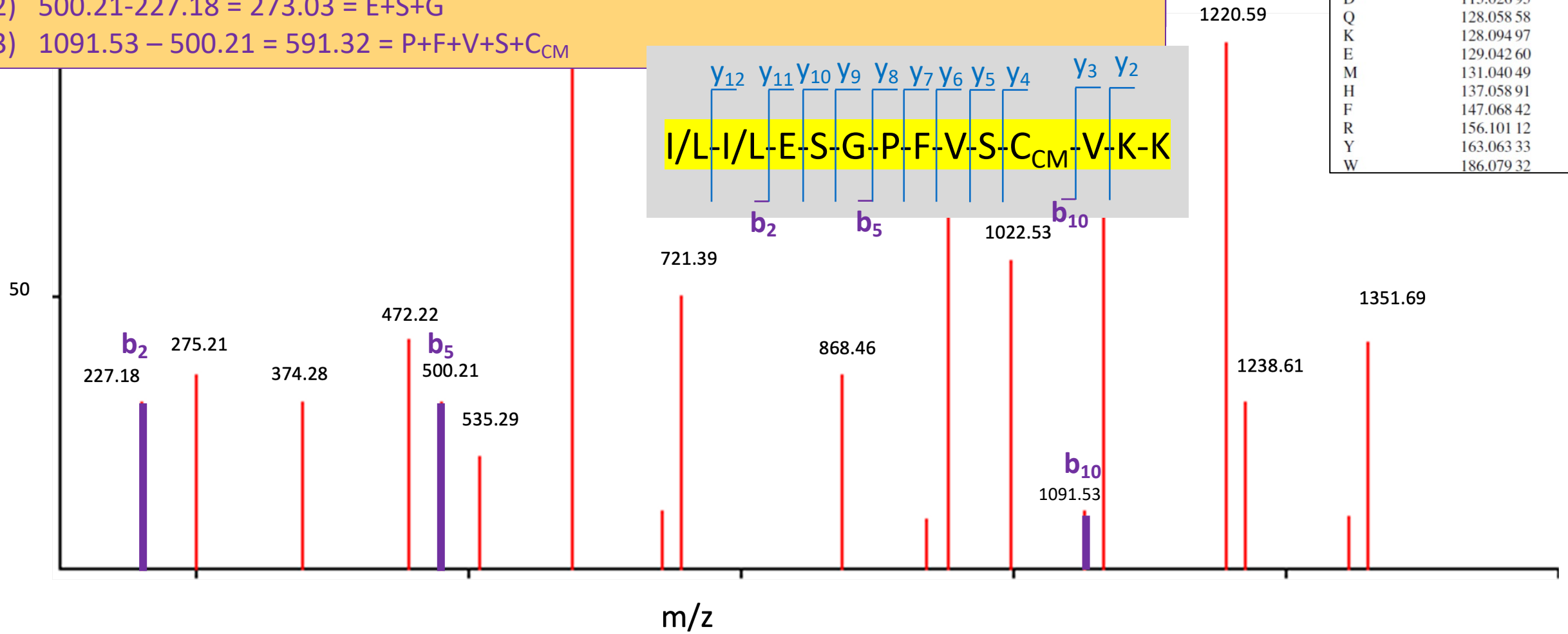
Code (1 letter)	Monoisotopic mass
G	57.021 47
A	71.037 12
S	87.032 03
P	97.052 77
V	99.068 42
T	101.047 68
C	103.009 19
I	113.084 07
L	113.084 07
N	114.042 93
D	115.026 95
Q	128.058 58
K	128.094 97
E	129.042 60
M	131.040 49
H	137.058 91
F	147.068 42
R	156.101 12
Y	163.063 33
W	186.079 32

C_{CM} : Cysteine with Carboxymethyl (58.01)

- 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.

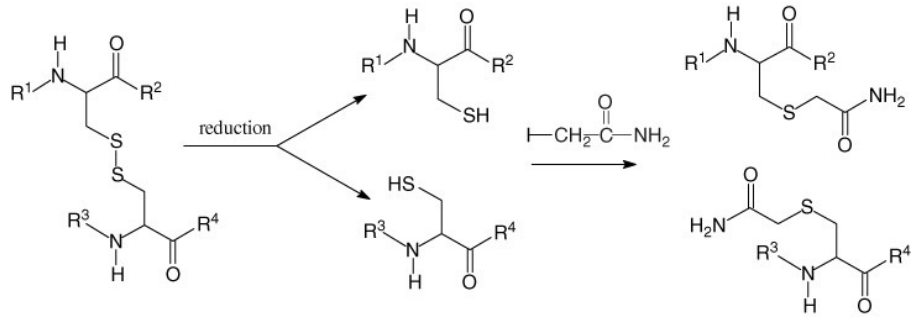
Code (1 letter)	Monoisotopic mass
G	57.021 47
A	71.037 12
S	87.032 03
P	97.052 77
V	99.068 42
T	101.047 68
C	103.009 19
I	113.084 07
L	113.084 07
N	114.042 93
D	115.026 95
Q	128.058 58
K	128.094 97
E	129.042 60
M	131.040 49
H	137.058 91
F	147.068 42
R	156.101 12
Y	163.063 33
W	186.079 32

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}

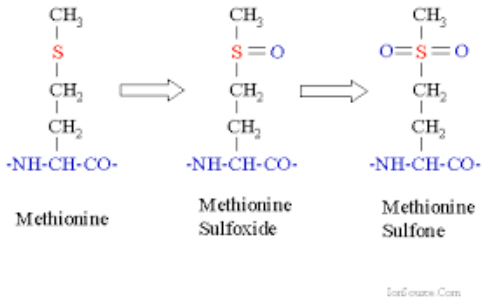


Specific Amino Acids Modification During Sample Handling

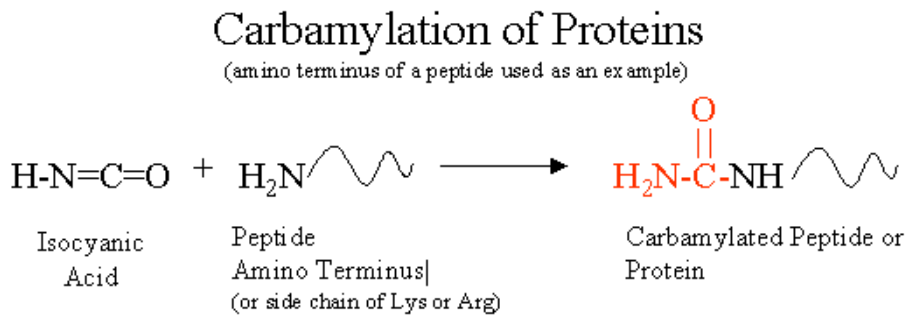
- Reduction and Alkylation on Cys
  - Routinely done prior to enzymatic digestion to break disulfide bonds, unfolding proteins to make them more susceptible to enzymatic cleavage



- Methionine is easily mono-oxidized (Met sulfoxide)



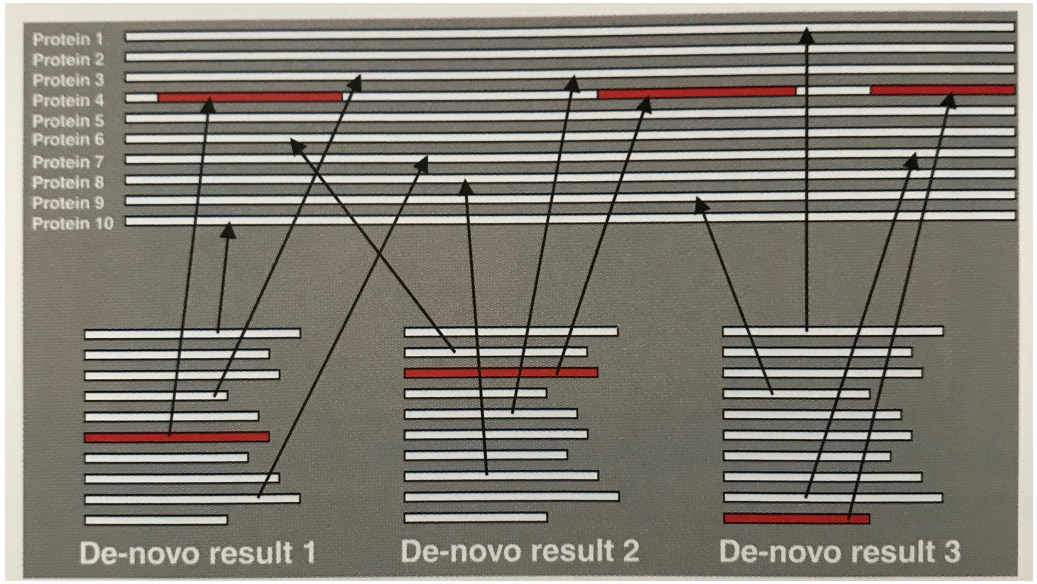
- Cyclization of N-terminal Glutamine (Q) and carboxamidomethyl-Cys
- Urea exposure can carbamylate N termini of protein/peptide and side chains of Lys



- 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