Proteomics Informatics –
Protein identification III:
de novo sequencing (Week 6)

De Novo Sequencing of MS Spectra

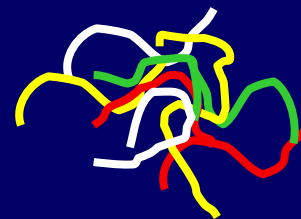## Only a manually confirmed spectrum is a correct spectrum

Beatrix Ueberheide

March 12th 2013

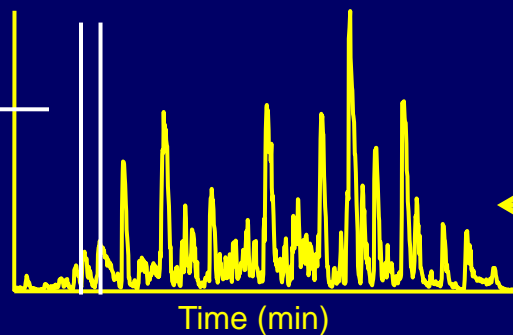Peptide Sequencing using Mass Spectrometry

Peptide Sequencing using Mass Spectrometry

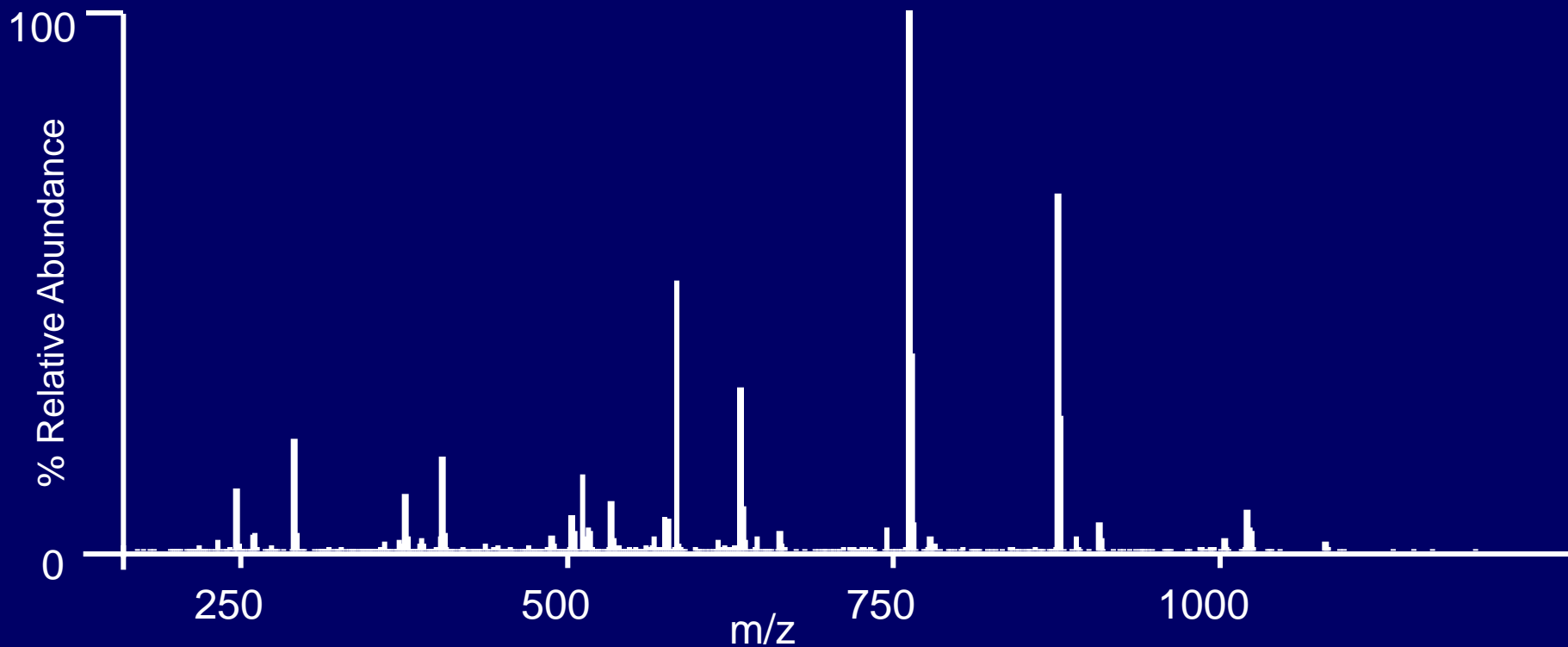| 88	145	292	405	534	663	778	907	1020	1166	b ions
S	G	F	L	E	E	D	E	L	K	

% Relative Abundance

100

0

250   500   750   1000

m/z
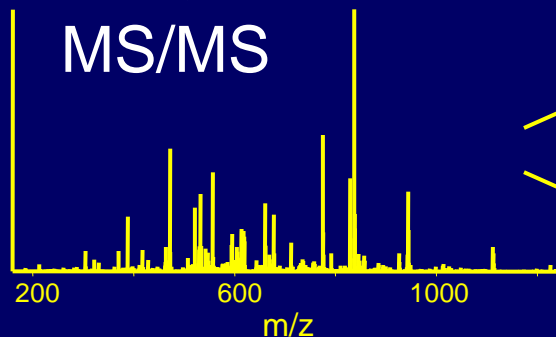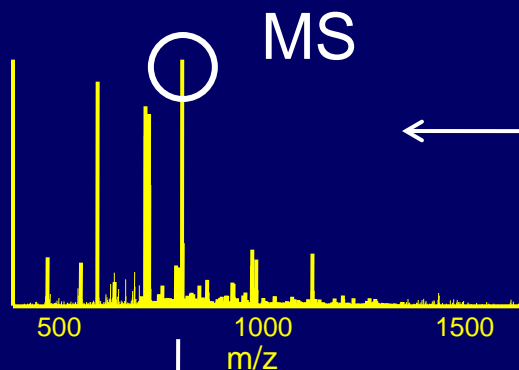
Peptide Sequencing using Mass Spectrometry

S	G	F	L	E	E	D	E	L	K	
1166	1080	1022	875	762	633	504	389	260	147	y ions

Peptide Sequencing using Mass Spectrometry

88	145	292	405	534	663	778	907	1020	1166	b ions
S	G	F	L	E	E	D	E	L	K	
1166	1080	1022	875	762	633	504	389	260	147	y ions

Peptide Sequencing using Mass Spectrometry

Peptide Sequencing using Mass Spectrometry

88	145	292	405	534	663	778	907	1020	1166	b ions
S	G	F	L	E	E	D	E	L	K	
1166	1080	1022	875	762	633	504	389	260	147	y ions

Peptide Sequencing using Mass Spectrometry

88	145	292	405	534	663	778	907	1020	1166	b ions
S	G	F	L	E	(E)	D	E	L	K	
1166	1080	1022	875	762	633	504	389	260	147	y ions

Peptide Sequencing using Mass Spectrometry

Peptide Sequencing using Mass Spectrometry

How to Sequence: CAD

Residue Mass (RM)

The very first N- and C-terminal fragment ions are not just their corresponding residue masses. The peptides N or C-terminus has to be taken into account.

b ion

y ion

b1 = RM + 1
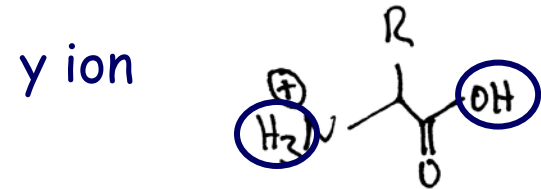
y1 = RM + 19

Example of how to calculate theoretical fragment ions

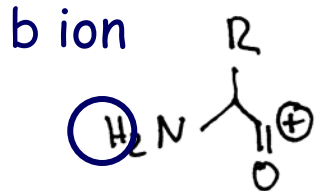| 88	159	290	387	500	629	803
S	A	M	P	L	E	R
803	716	645	514	417	304	175
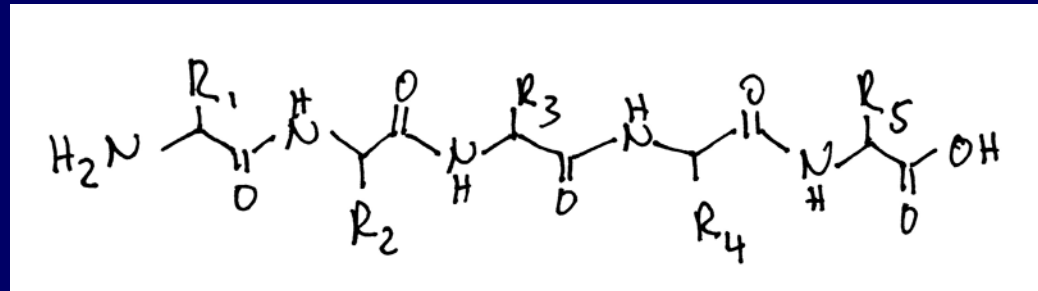
The first b ion
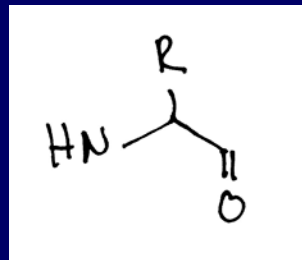
Residue Mass
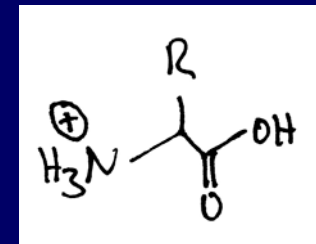
The first y ion

How to calculate theoretical fragment ions

RM+1 + RM   + RM + RM + RM + RM +RM+18
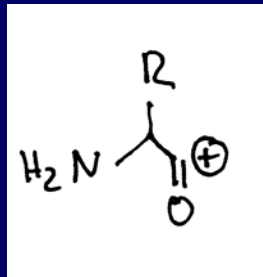
88	159	290	387	500	629	803
S	A	M	P	L	E	R
803	716	645	514	417	304	175

+ RM + RM + RM + RM   + RM + RM   RM+19

The first b ion

Residue Mass

The first y ion

Finding 'pairs' and 'biggest' ions

largest b ion = $(n+H)^+ - (H + HN \overset{R}{\underset{O}{\bigwedge}} + OH)$

$\qquad = (n+H)^+ - (Rn + 18)$

largest y ion = $(n+H)^+ - (HN \overset{R}{\underset{O}{\bigwedge}} y)$

$\qquad = (n+H)^+ - Rn$
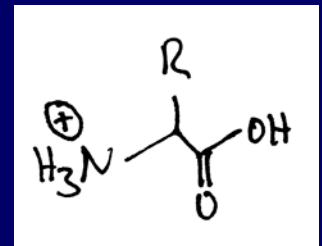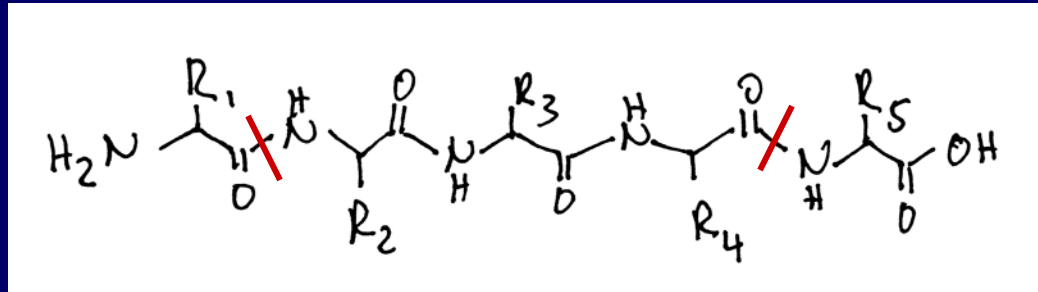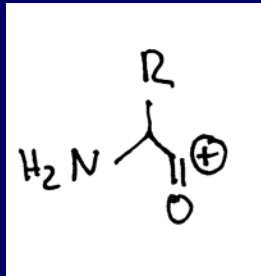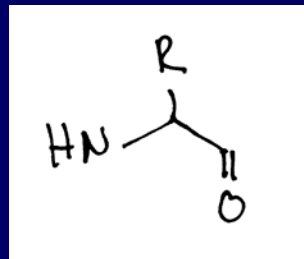
If trypsin was used for digestion, one can assume that the peptide terminates in K or R. Therefore the biggest observable b ion should be:
Mass of peptide [M+H] $^{+1}$ -128 (K) -18
Mass of peptide [M+H] $^{+1}$ -156 (K) -18

y ions are truncated peptides. Therefore subtract a residue mass from the parent ion [M+H] $^{+1}$. The highest possible ion could be at [M+H] $^{+1}$ -57 (G) and the lowest possible ion at [M+H] $^{+1}$ -186 (W)

b and y ion pairs:
Complementary b and y ions should add up and result in the mass of the intact peptide, except since both b and y ion carry $1H^+$ the peptide mass will be by $1H^+$ too high therefore: b (m/z) + y (m/z)-1 = [M+H] $^{+1}$

Check the SAMPLER example

How to start sequencing

- Know the charge of the peptide
- Know the sample treatment (i.e. alkylation, other derivatizations that could change the mass of amino acids)
- Know what enzyme was used for digestion
- Calculate the $[M+1H]+1$ charge state of the peptide
- Find and exclude non sequence type ions (i.e. unreacted precursor, neutral loss from the parent ion, neutral loss from fragment ions
- Try to see if you can find the biggest y or b ion in the spectrum. Note, if you used trypsin your C-terminal ion should end in lysine or arginine
- Try to find sequence ions by finding b/y pairs
- You usually can conclude you found the correct sequence if you can explain the major ions in a spectrum

Table 1. Mass Values for Peptide Fragment Ions

One Let Code	Amino Acid	Three Let Code	Mono Isotopic Integer Residue Mass	Mass Change	Neutral Mass Lost	Methyl Ester Mass	Mono Isotopic Accurate Residue Mass	Average Residue Mass
G	Glycine	Gly	57 Da	0 Da			57.02	57.05
A	Alanine	Ala	71 Da	14 Da			71.04	71.08
S	Serine	Ser	87 Da	16 Da	18 Da		87.03	87.08
P	Proline	Pro	97 Da	10 Da			97.05	97.12
V	Valine	Val	99 Da	2 Da			99.07	99.13
T	Threonine	Thr	101 Da	2 Da	18 Da		101.05	101.11
C	Cysteine	Cys	103 Da	2 Da	34 Da		103.01	103.14
L	Leucine	Leu	113 Da	10 Da			113.08	113.16
I	Isoleucine	Ile	113 Da				113.08	113.16
N	Asparagine	Asn	114 Da	1 Da	17 Da		114.04	114.1
D	Aspartic Acid	Asp	115 Da	1 Da	18 Da	129 Da	115.03	115.09
K	Lysine	Lys	128 Da	13 Da	17 Da		128.09	128.17
Q	Glutamine	Gln	128 Da		17 Da		128.06	128.13
E	Glutamic Acid	Glu	129 Da	1 Da	18 Da	143 Da	129.04	129.12
M	Methionine	Met	131 Da	2 Da	48 Da		131.04	131.19
H	Histidine	His	137 Da	6 Da			137.06	137.14
F	Phenylalanine	Phe	147 Da	10 Da			147.07	147.18
R	Arginine	Arg	156 Da	9 Da	17 Da		156.1	156.19
Y	Tyrosine	Tyr	163 Da	7 Da			163.06	163.18
W	Tryptophan	Trp	186 Da	23 Da			186.08	186.21

Hydrogen	H	1		1.008	1.008
Carbon	C	12		12	12.011
Nitrogen	N	14		14.003	14.007
Oxygen	O	16		15.995	15.999
Sulfur	S	32		31.972	32.06
Methyl		15		15.02	15.04
Water		18		18.01	18.02
Acetyl		43		43	43.05

Residue Mass (RM) ≡ NH—...—R, O

# Common observed neutral losses and mass additions:
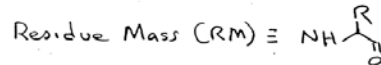
- Ammonia -17

- Water -18

- Carbon Monoxide from b ions -28

- Phosphoric acid from phosphorylated serine and threonine -98

- Carbamidomethyl modification on cysteines upon alkylation with iodoacetamide +57

- Oxidation of methionine +18

Calculate with nominal mass during sequencing, but use the monoisotopic masses to check if the parent mass fits. For high res. MS/MS check that the residue mass difference is correct.

E:\for lecture\Stat3_cytosolic_a   2/28/2011 12:57:52 PM
loaded 3 ul (total ~12ul) of gel digested Stat3
Stat3_cytosolic_a #8004   RT: 63.24   AV: 1   NL: 1.15E3
T: ITMS + c ESI Full ms2 1196.20@cid35.00 [325.00-2000.00]
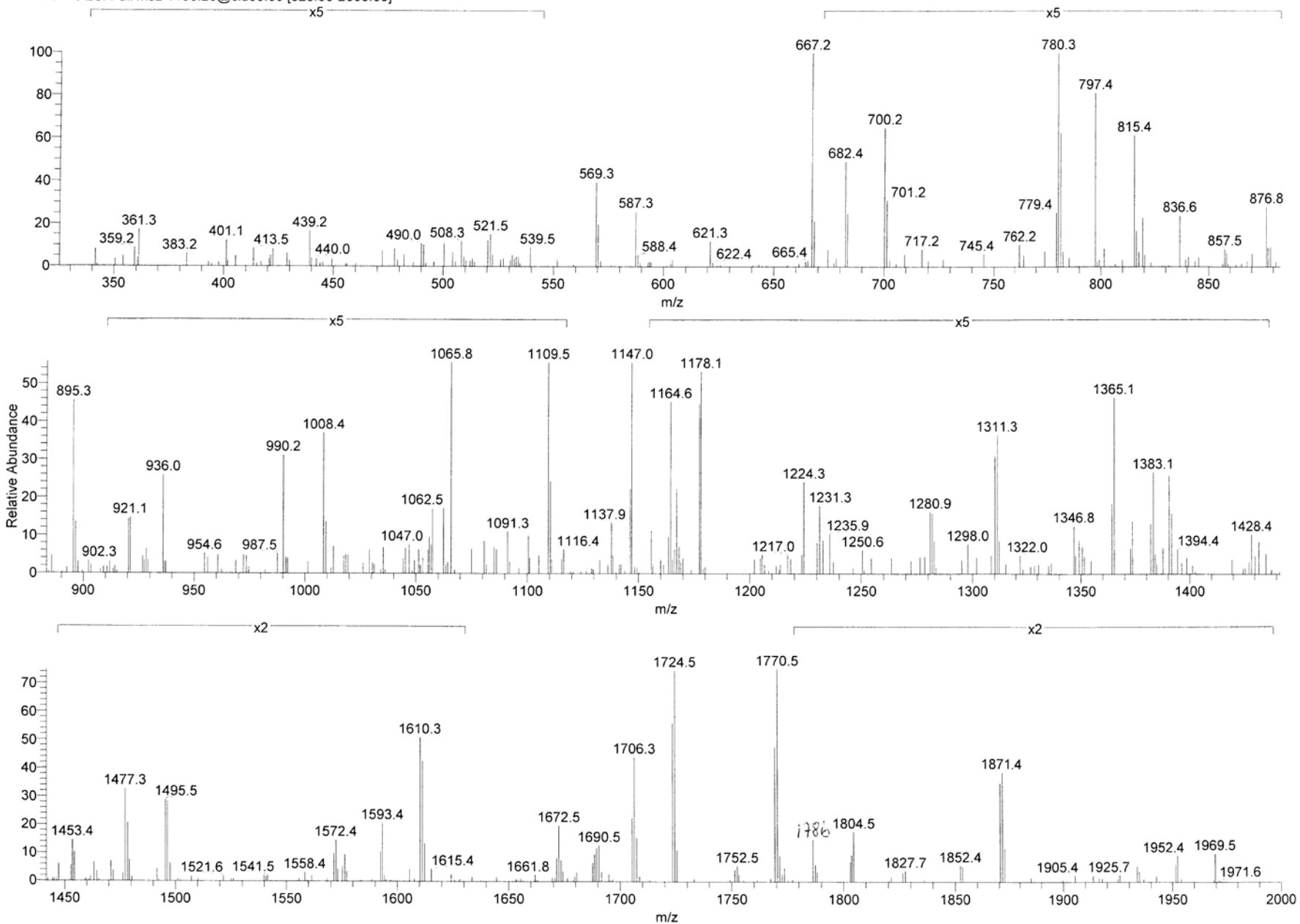
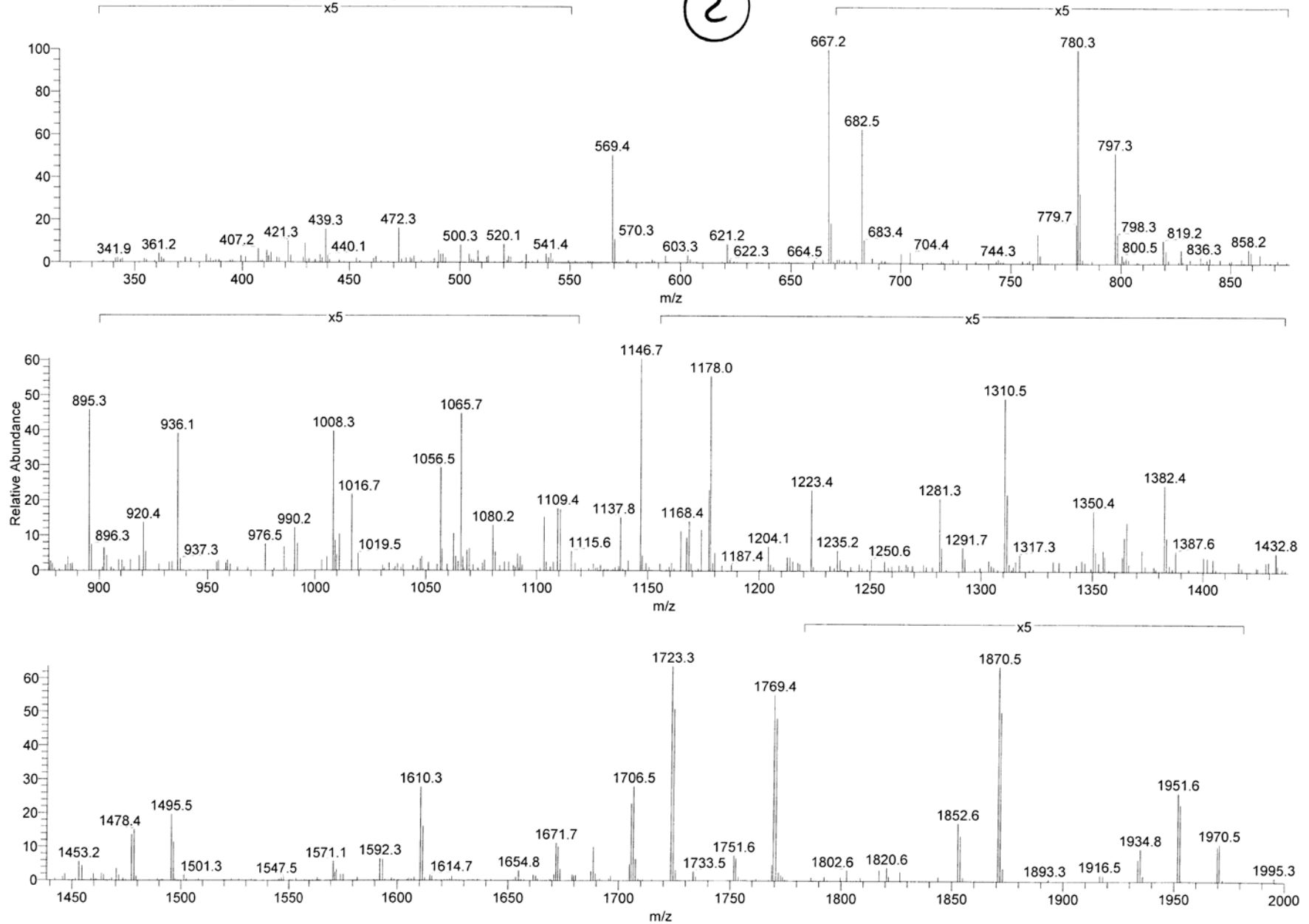E:\for lecture\Stat3_cytosolic_a   2/28/2011 12:57:52 PM
loaded 3 ul (total ~12ul) of gel digested Stat3
Stat3_cytosolic_a #7952   RT: 62.88   AV: 1   NL: 6.00E3
T: ITMS + c ESI d Full ms2 1196.04@cid35.00 [315.00-2000.00]

Mixed Phospho spectra

unmodified

1 Phospho site

1 Phospho site

First 'on your own example'

Remember what you need to know first!

E:\for lecture\030713_jrc_std4
30 fmole AV, BSA, AIIP
030713_jrc_std4 #1725 RT: 18.01 AV: 1 NL: 7.45E4
T: ITMS + c NSI d Full ms2 722.33@cid35.00 [185.00-1455.00]

3/8/2013 11:09:33 AM

What is the charge state?

Neutral loss of water
Water = 18; 18/z; 9
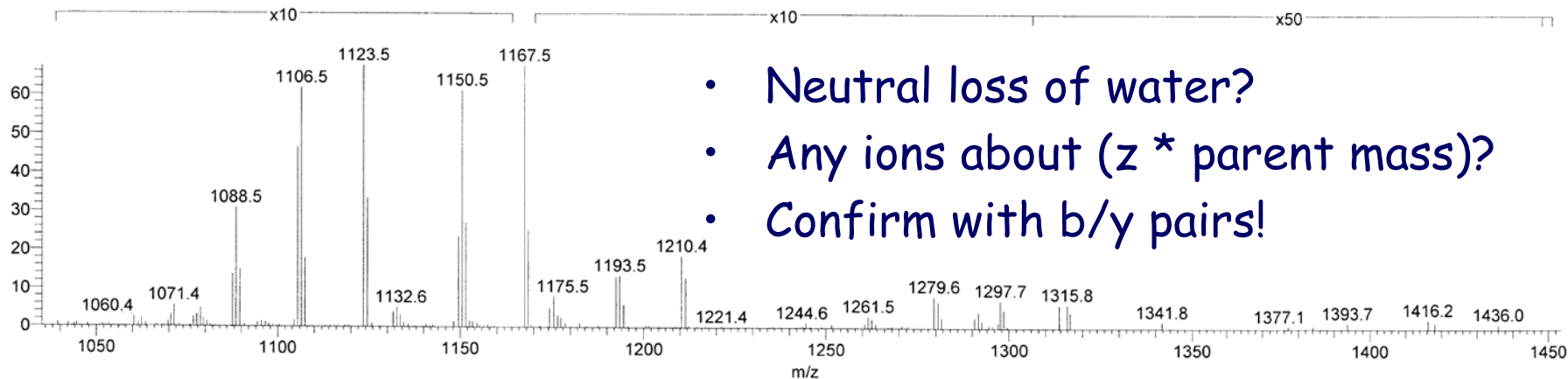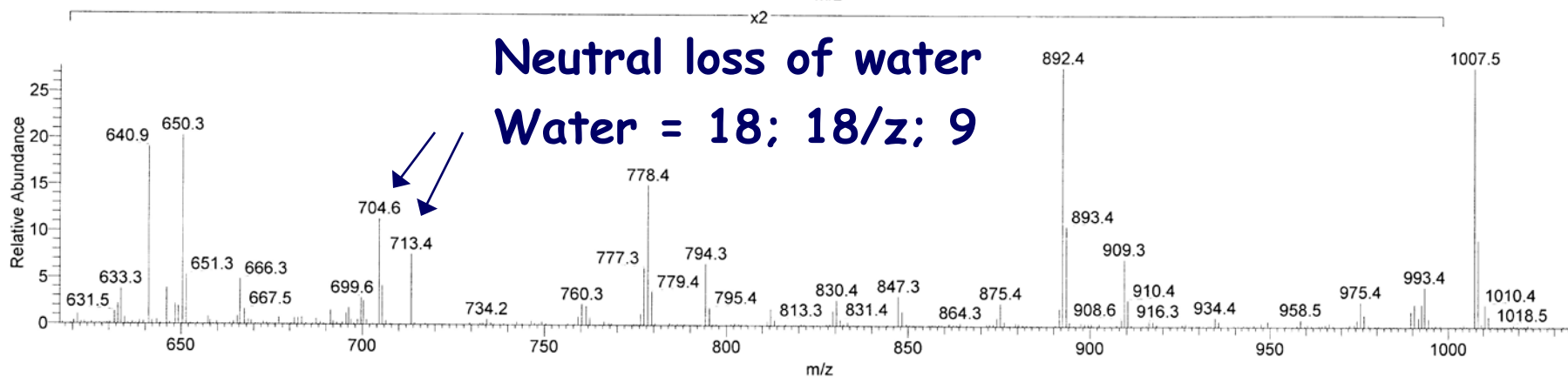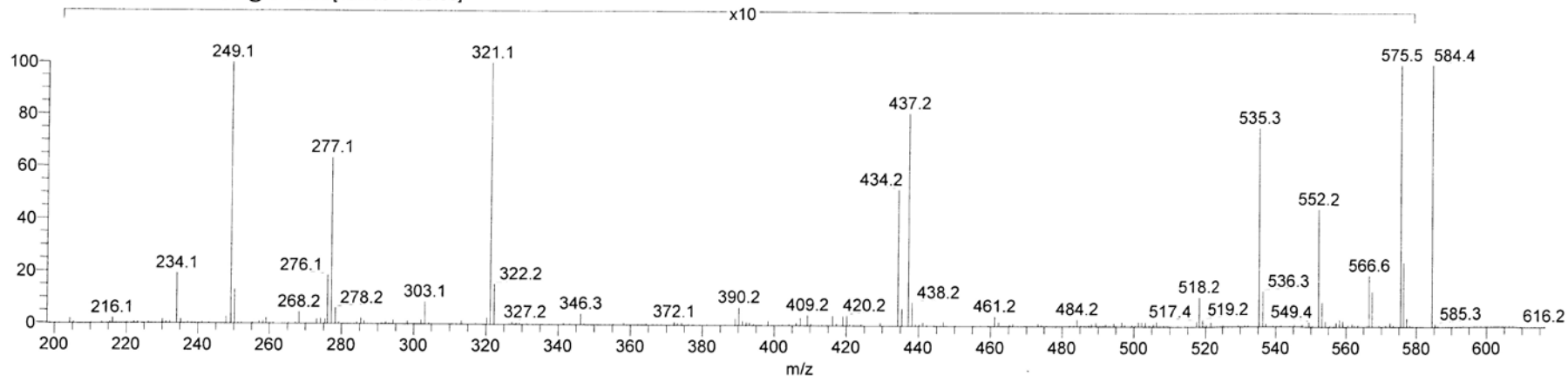
- Neutral loss of water?
- Any ions about (z * parent mass)?
- Confirm with b/y pairs!

Search for 'biggest ion'

E:\for lecture\030713_jrc_std4
30 fmole AV, BSA, AllP
030713_jrc_std4 #1725 RT: 18.01 AV: 1 NL: 7.45E4
T: ITMS + c NSI d Full ms2 722.33@cid35.00 [185.00-1455.00]

3/8/2013 11:09:33 AM
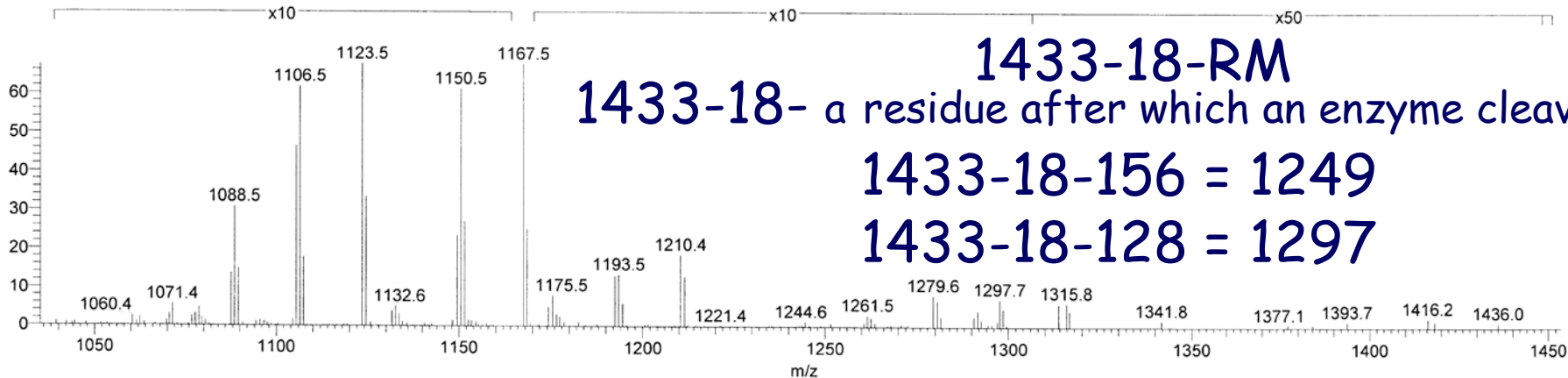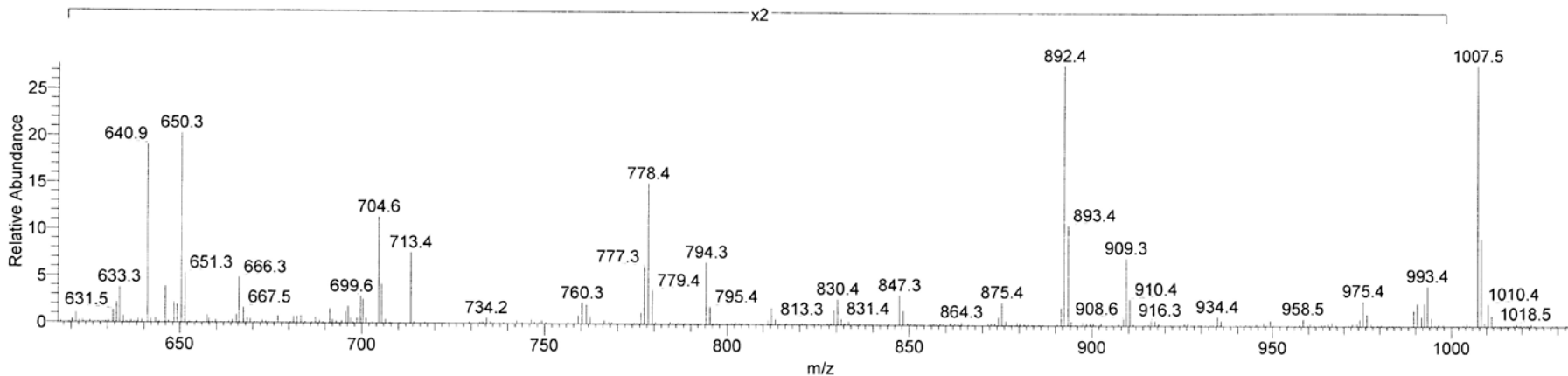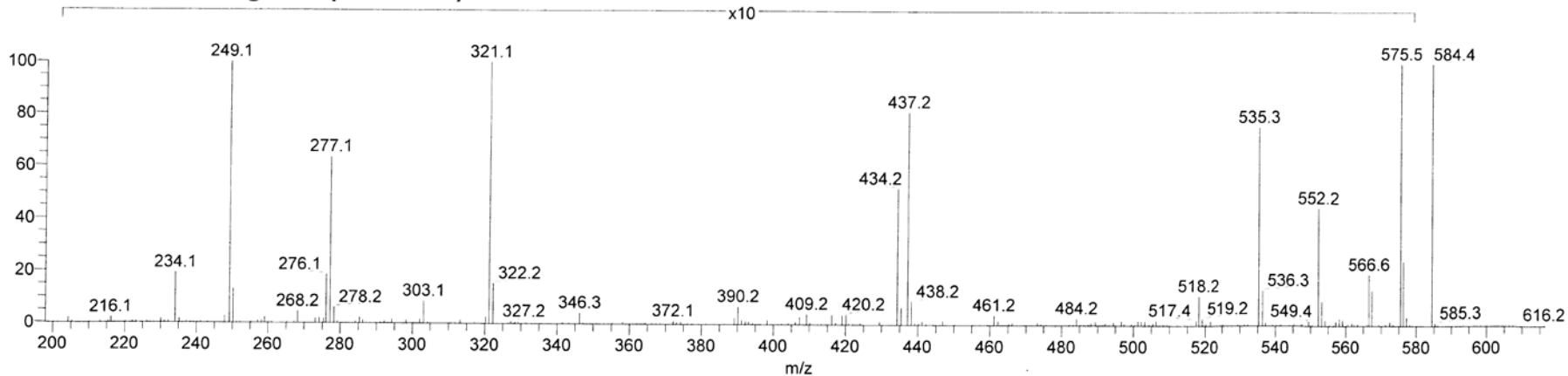
1433-18-RM
1433-18- a residue after which an enzyme cleaves
1433-18-156 = 1249
1433-18-128 = 1297

E:\for lecture\030713_jrc_std4
3/8/2013 11:09:33 AM
30 fmole AV, BSA, AIIP
030713_jrc_std4 #1725 RT: 18.01 AV: 1 NL: 7.45E4
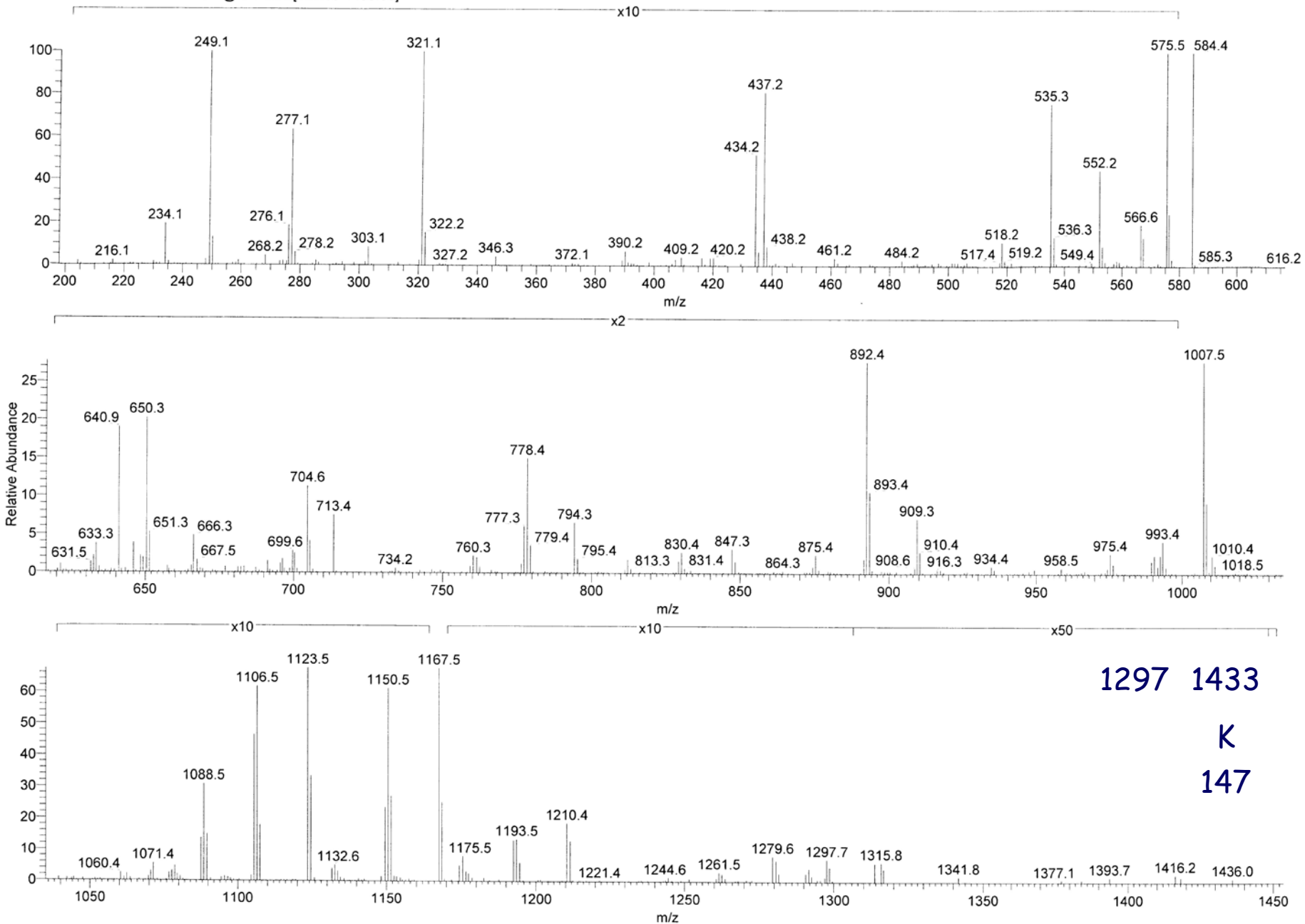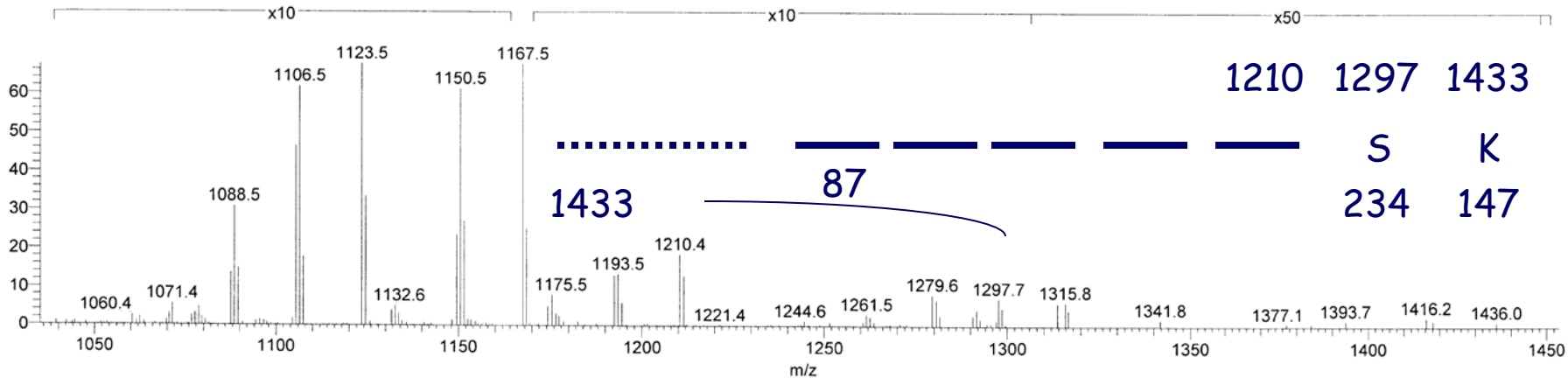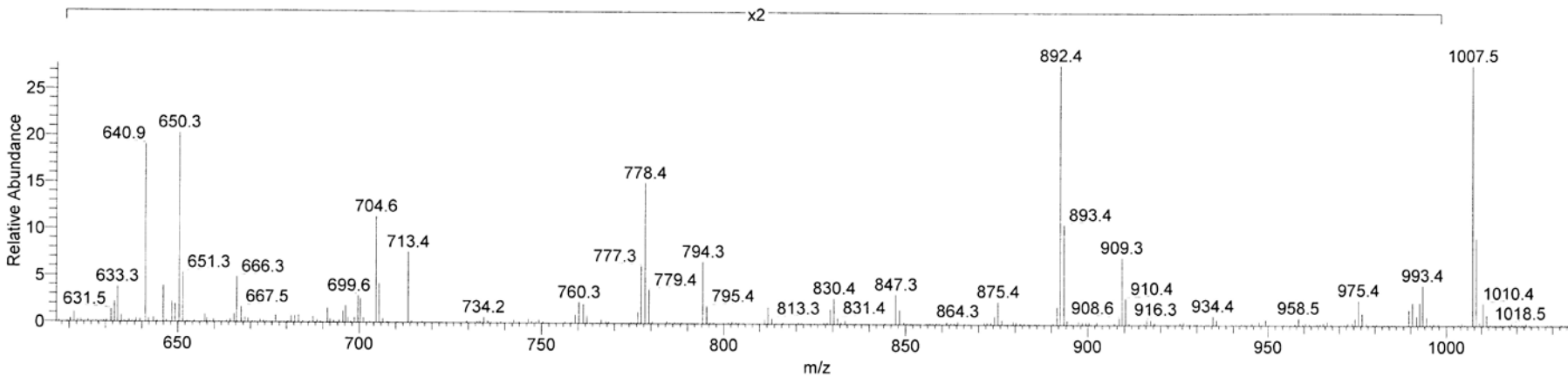T: ITMS + c NSI d Full ms2 722.33@cid35.00 [185.00-1455.00]

x10

100
80
60
40
20
0

216.1 234.1 249.1 268.2 276.1 277.1 278.2 303.1 321.1 322.2 327.2 346.3 372.1 390.2 409.2 420.2 434.2 437.2 438.2 461.2 484.2 517.4 518.2 519.2 535.3 536.3 549.4 552.2 566.6 575.5 584.4 585.3 616.2

200 220 240 260 280 300 320 340 360 380 400 420 440 460 480 500 520 540 560 580 600

m/z

x2

25
20
15
10
5
0

Relative Abundance

631.5 633.3 640.9 650.3 651.3 666.3 667.5 699.6 704.6 713.4 734.2 760.3 777.3 778.4 779.4 794.3 795.4 813.3 830.4 831.4 847.3 864.3 875.4 892.4 893.4 908.6 909.3 910.4 916.3 934.4 958.5 975.4 993.4 1007.5 1010.4 1018.5

650 700 750 800 850 900 950 1000

m/z

x10   x10   x50

60
50
40
30
20
10
0

1060.4 1071.4 1088.5 1106.5 1123.5 1132.6 1150.5 1167.5 1175.5 1193.5 1210.4 1221.4 1244.6 1261.5 1279.6 1297.7 1315.8 1341.8 1377.1 1393.7 1416.2 1436.0

1050 1100 1150 1200 1250 1300 1350 1400 1450

m/z

1297 1433
K
147

E:\for lecture\030713_jrc_std4   3/8/2013 11:09:33 AM
30 fmole AV, BSA, AIIP
030713_jrc_std4 #1725   RT: 18.01   AV: 1   NL: 7.45E4
T: ITMS + c NSI d Full ms2 722.33@cid35.00 [185.00-1455.00]

E:\for lecture\030713_jrc_std4   3/8/2013 11:09:33 AM
30 fmole AV, BSA, AIIP
030713_jrc_std4 #1725   RT: 18.01   AV: 1   NL: 7.45E4
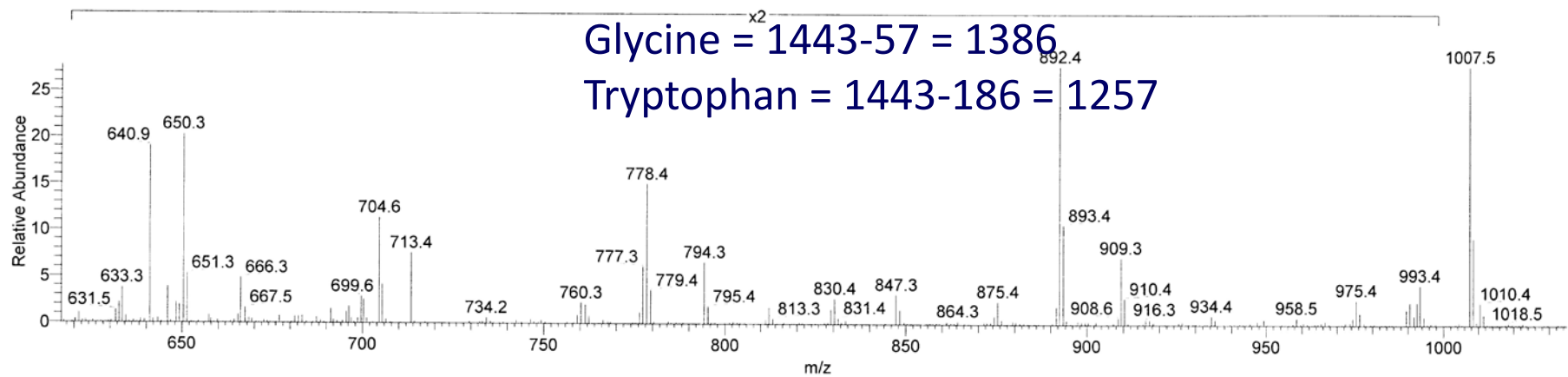T: ITMS + c NSI d Full ms2 722.33@cid35.00 [185.00-1455.00]

Find the biggest y ion!

Peptide Mass – RM

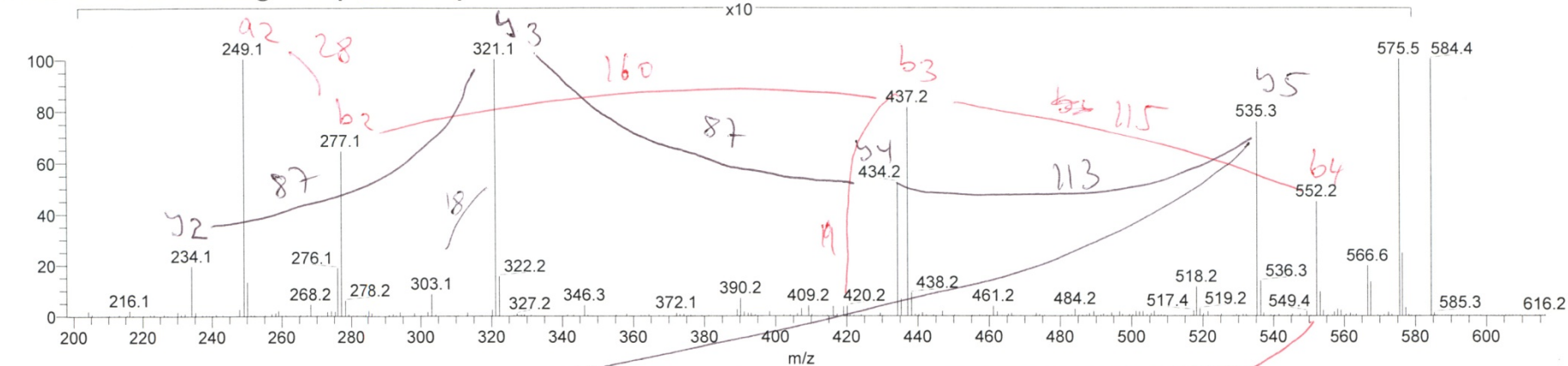Lowest possible ion = Glycine
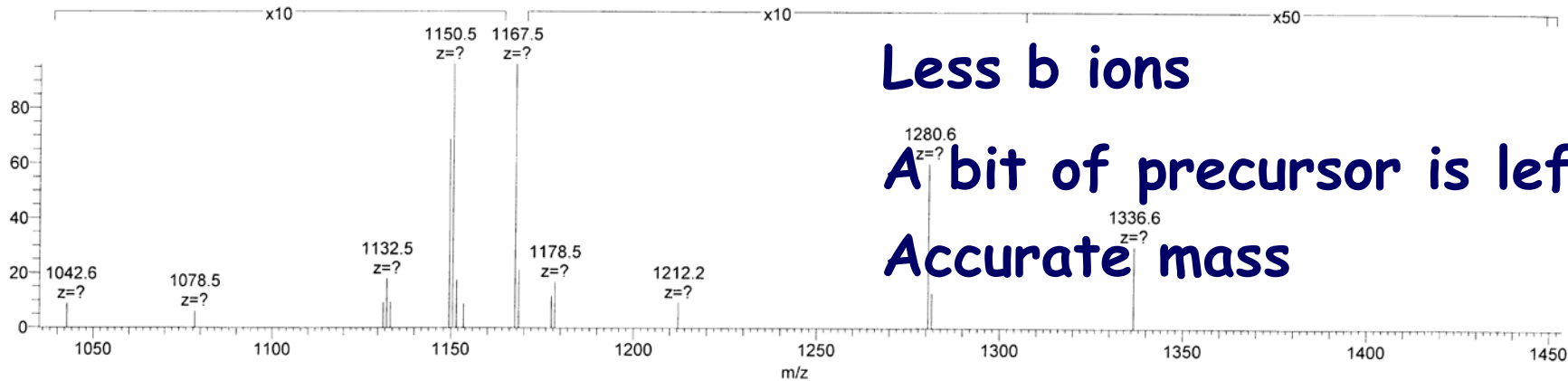
Highest possible ion = Tryptophan

Glycine = 1443-57 = 1386

Tryptophan = 1443-186 = 1257

x10

100
249.1   321.1   575.5 584.4
80
277.1   437.2   535.3
60
434.2
40
552.2
234.1   276.1   566.6
20   268.2 278.2 303.1   322.2   518.2   536.3
216.1   327.2 346.3 372.1 390.2 409.2 420.2 438.2 461.2 484.2 517.4 519.2 549.4 585.3 616.2

200 220 240 260 280 300 320 340 360 380 400 420 440 460 480 500 520 540 560 580 600
m/z

x2

Relative Abundance

25
640.9 650.3   892.4   1007.5
20
778.4
15   893.4
704.6
10   909.3
651.3 666.3 713.4 777.3 794.3
5 633.3   667.5 699.6   779.4 795.4 830.4 847.3 875.4 910.4 975.4 993.4 1010.4
631.5   734.2 760.3   813.3 831.4 864.3 908.6 916.3 934.4 958.5 1018.5

650   700   750   800   850   900   950   1000
m/z

x10   **x10**   **x50**

60   1123.5   1167.5   164   1210 1297 1433
1106.5 1150.5
50   y..............   — — — — — —   S   K
40
1088.5   87
30   1433 1280   234 147
1210.4
20   1193.5   **1443-163 = 1280**
1060.4 1071.4 1132.6 1175.5   1279.6 1297.7 1315.8
10   1341.8 1377.1 1393.7 1416.2 1436.0
1221.4 1244.6 1261.5

1050 1100 1150 1200 1250 1300 1350 1400 1450
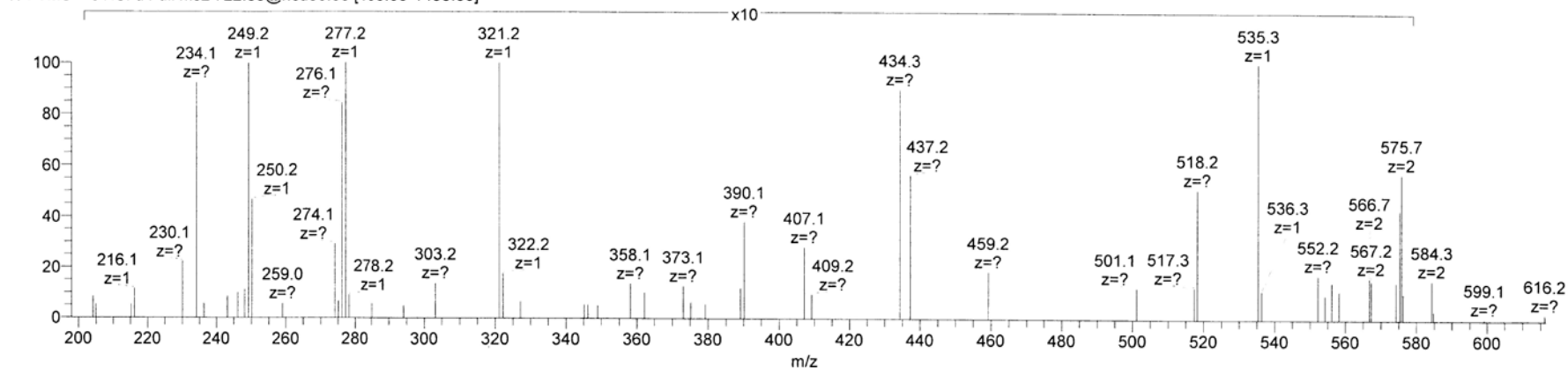m/z

And the sequence is........

What is the difference ?
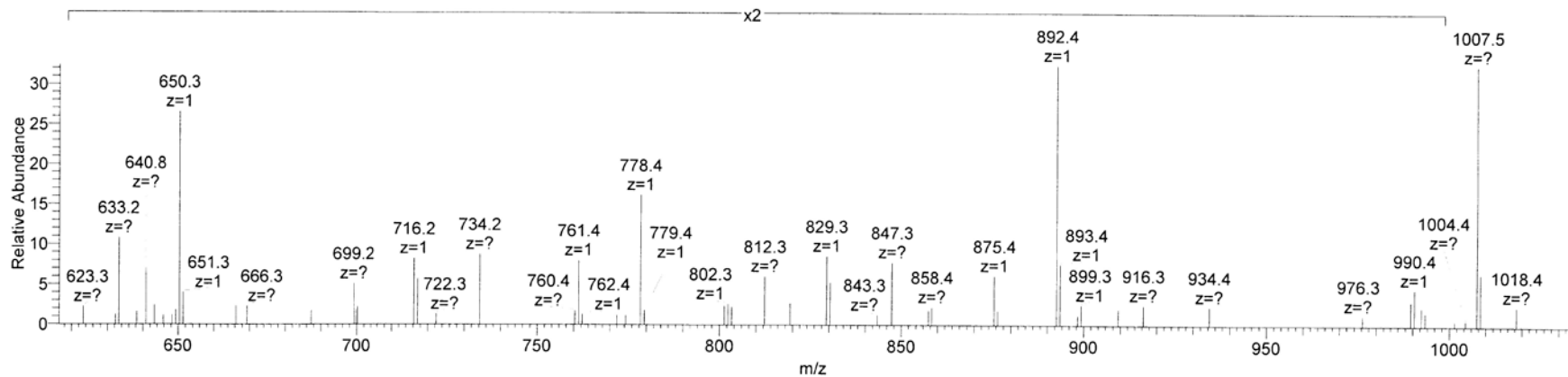
Less b ions
A bit of precursor is left
Accurate mass

What if we do not get good fragmentation?

[M+4H - P]$^{++++}$

Intensity: 2.2 x 10^6

1.

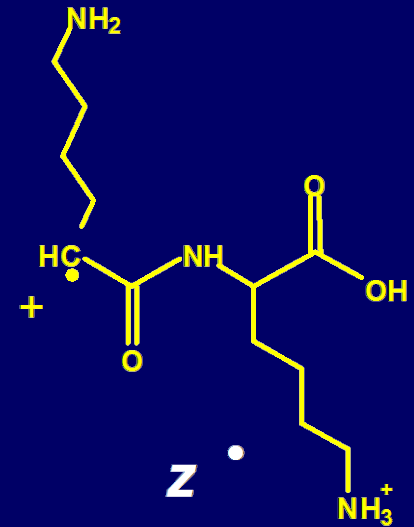CAD
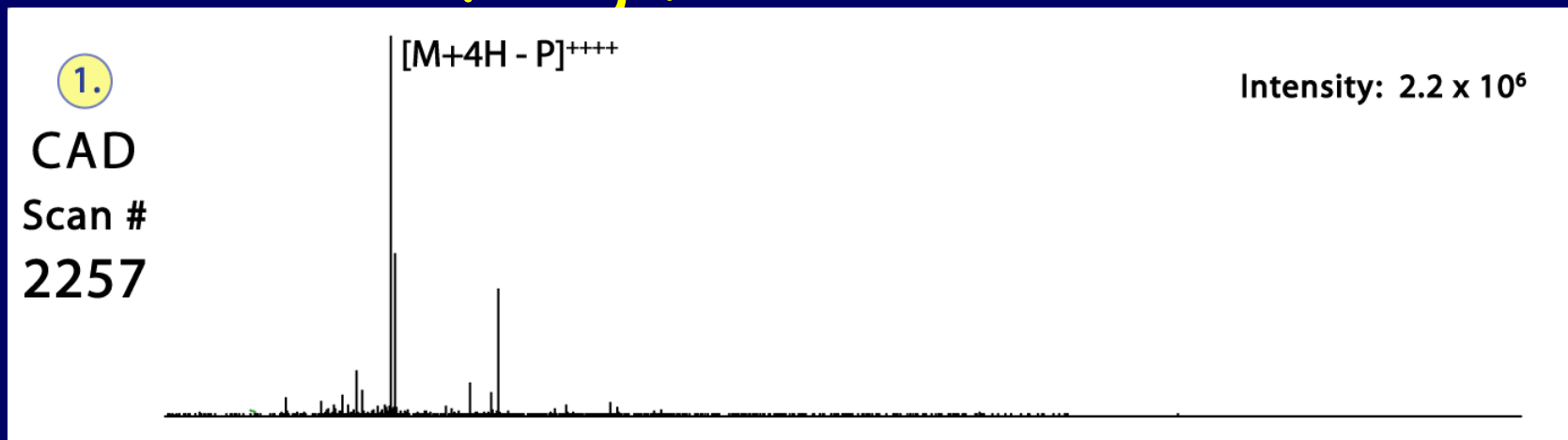
Scan #

2257

Try a different mode of dissociation
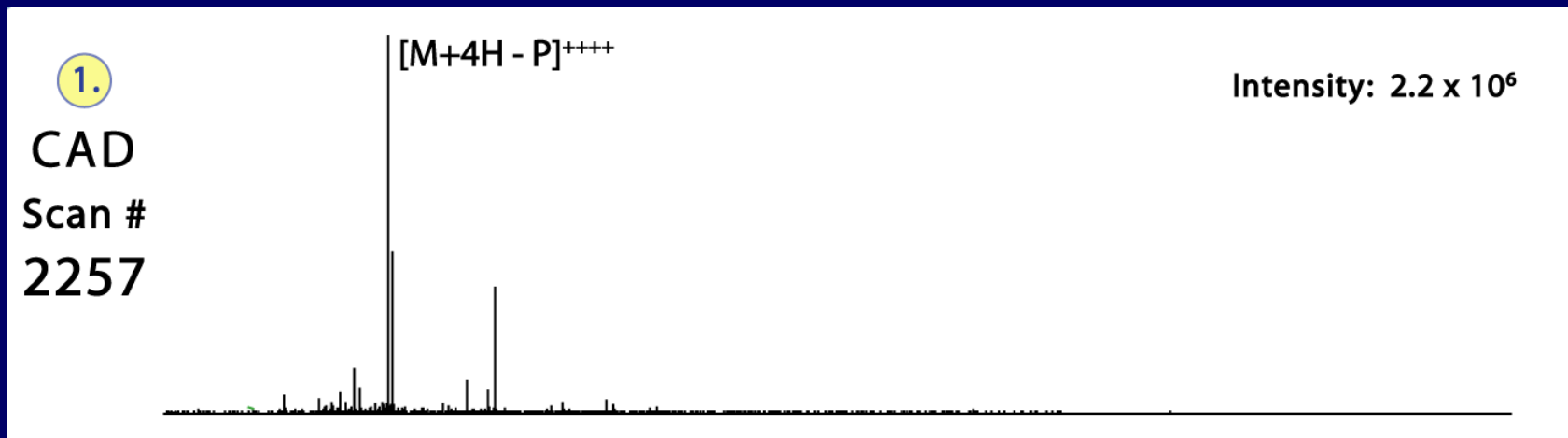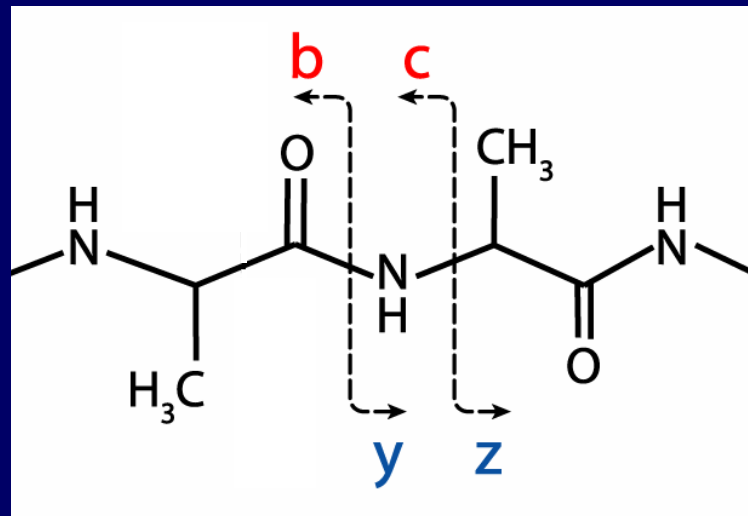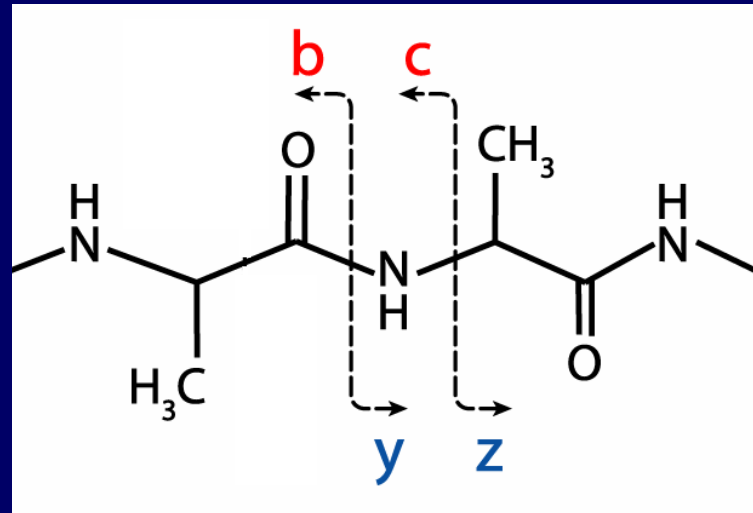
ETD
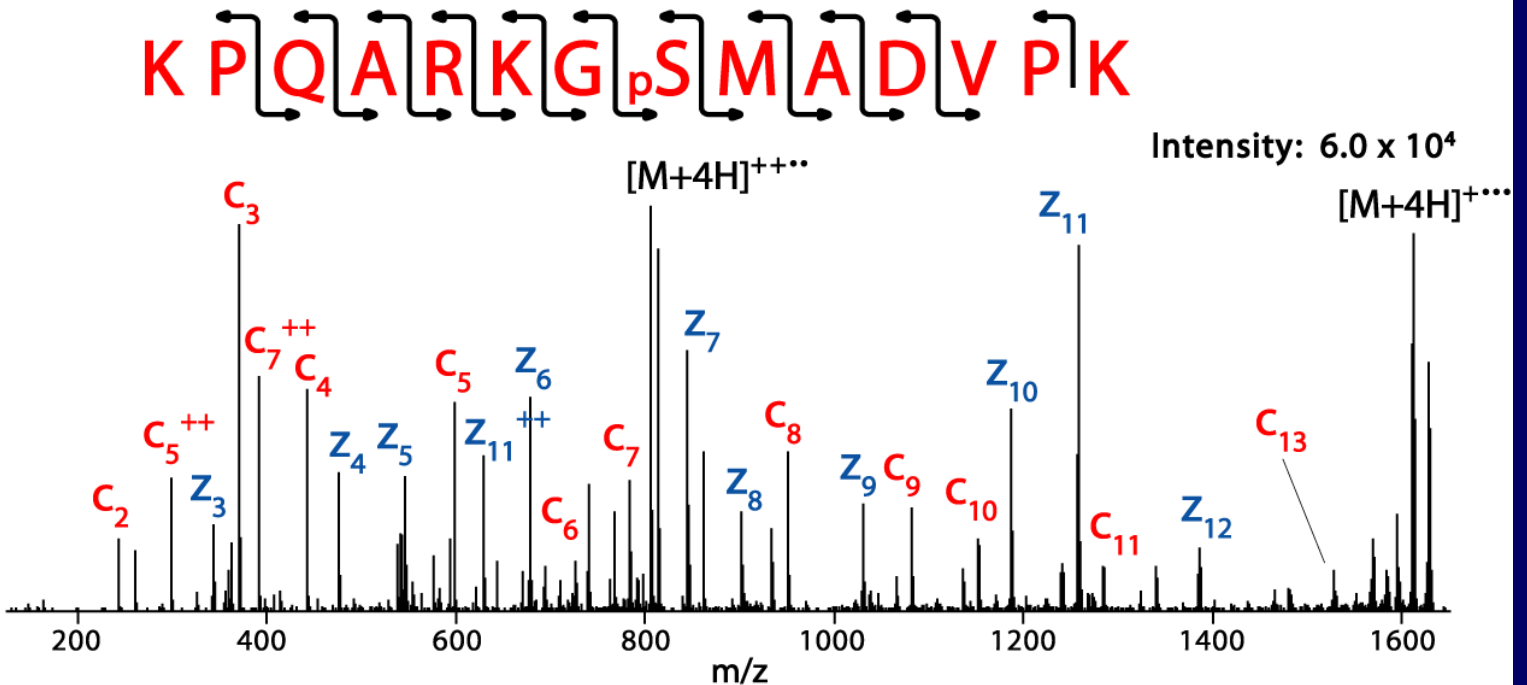
Electron Transfer Dissociation
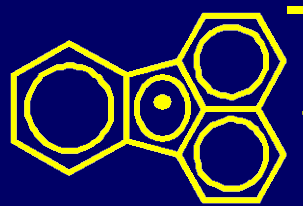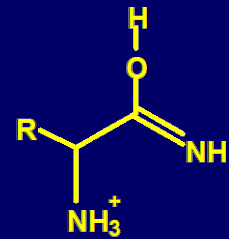
Fluoranthene

Tandem MS - Dissociation Techniques
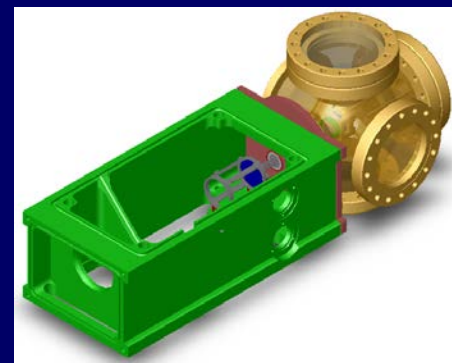
CAD: Collision Activated Dissociation (b, y ions)

⇒increase of internal energy through collisions

ETD: Electron Transfer Dissociation (c, z ions)

⇒bombardment of peptides with electrons
     (radical driven fragmentation)

The Prototype Instrument

Modified rear / CI source

HPLC

LTQ front
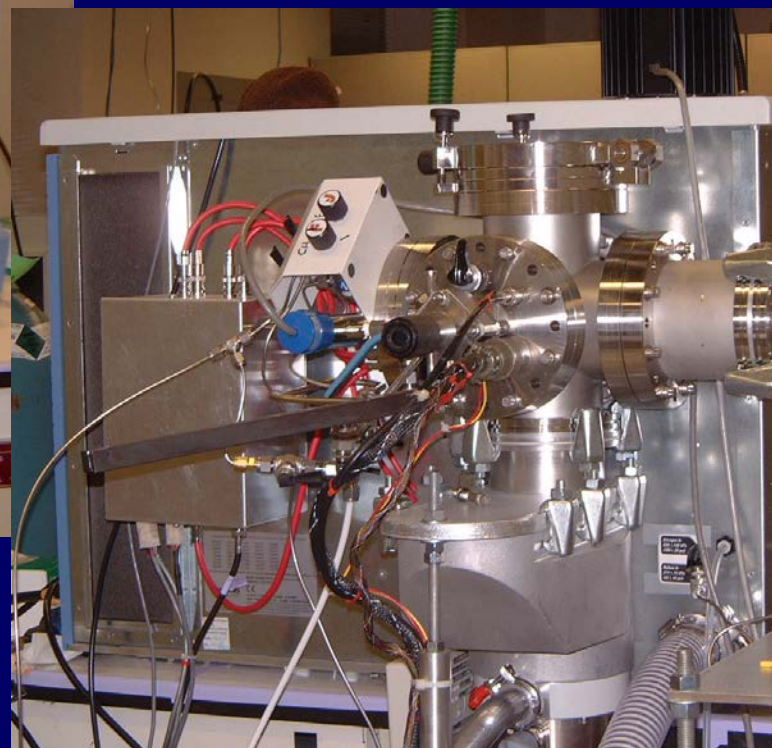
Modifications For Ion/Ion Experiments
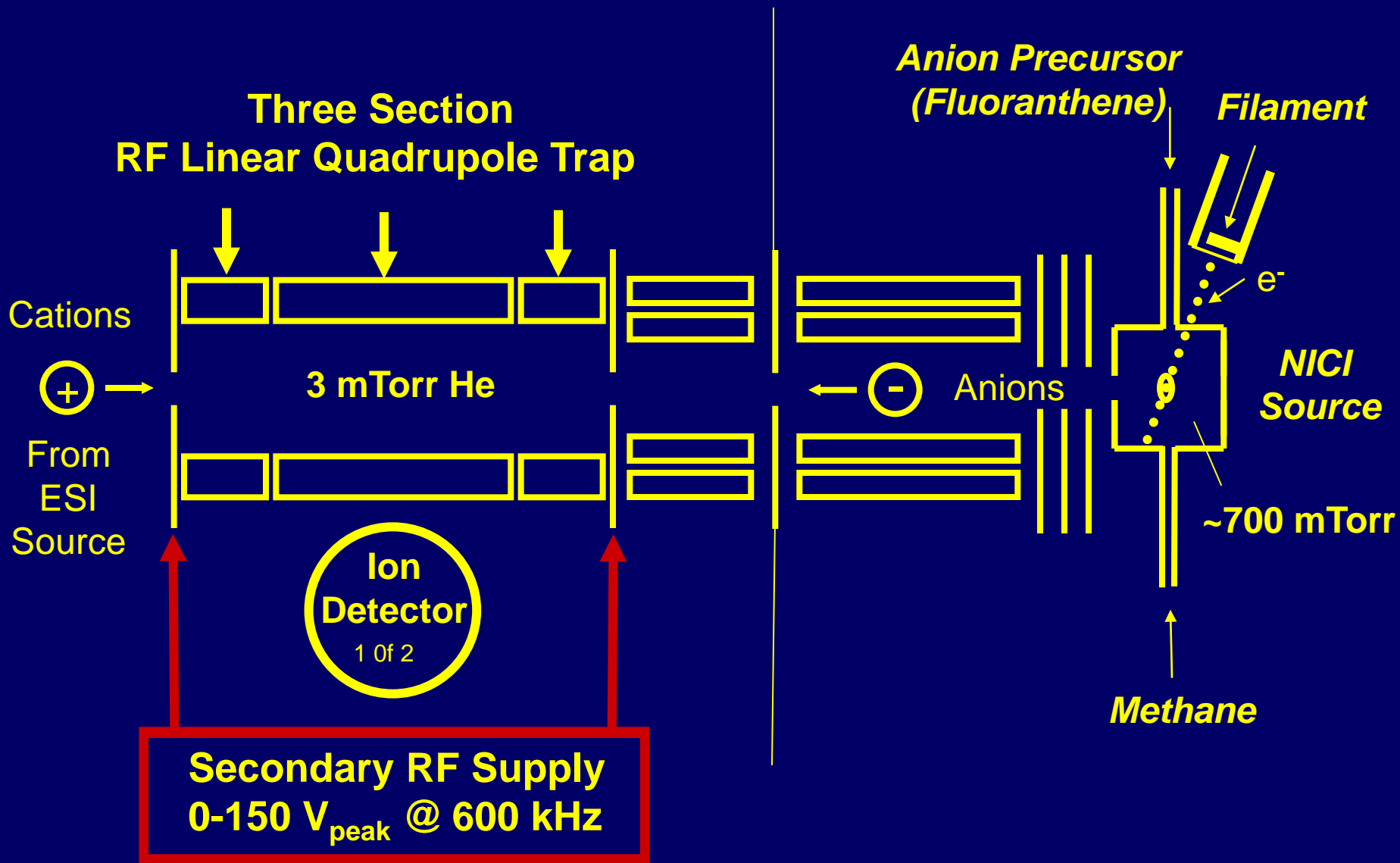
Injection of Positive Ions (ESI)

Front
Section

Center
Section

Back
Section

Front
Lens

Back
Lens

Peptide
Cations

Ions accumulate

0 V

-10 V
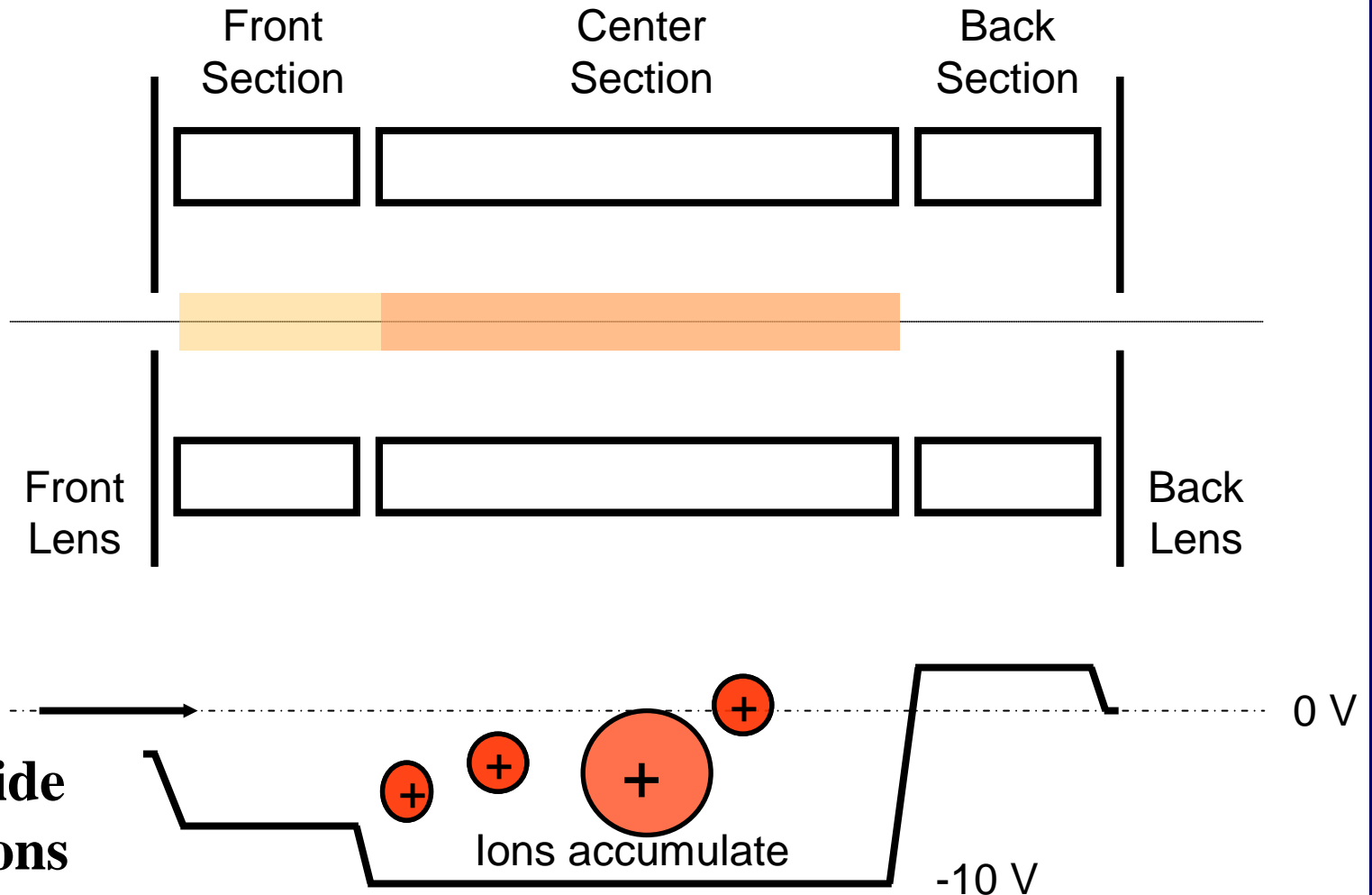
Precursor Storage in Front Section

Front Section

Center Section

Back Section

Front Lens

Back Lens

Precursor ions moved to front section

+

0 V

-10 V
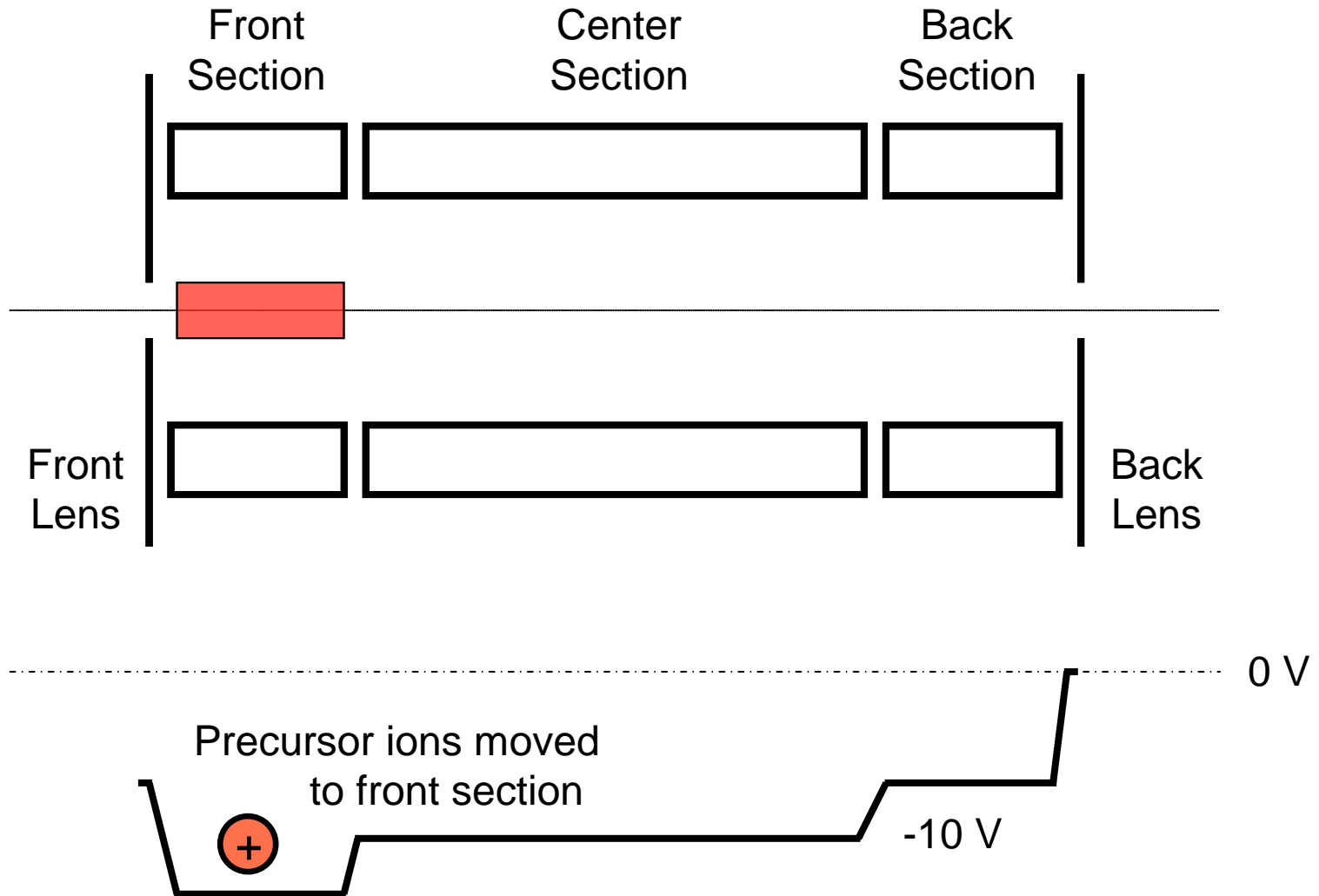
Injection of Negative Ions (CI)

Front Section

Center Section

Back Section

Front Lens

Back Lens

Negative reagent ions accumulate in the center section

+5 V

0 V

Precursor ions held in front section
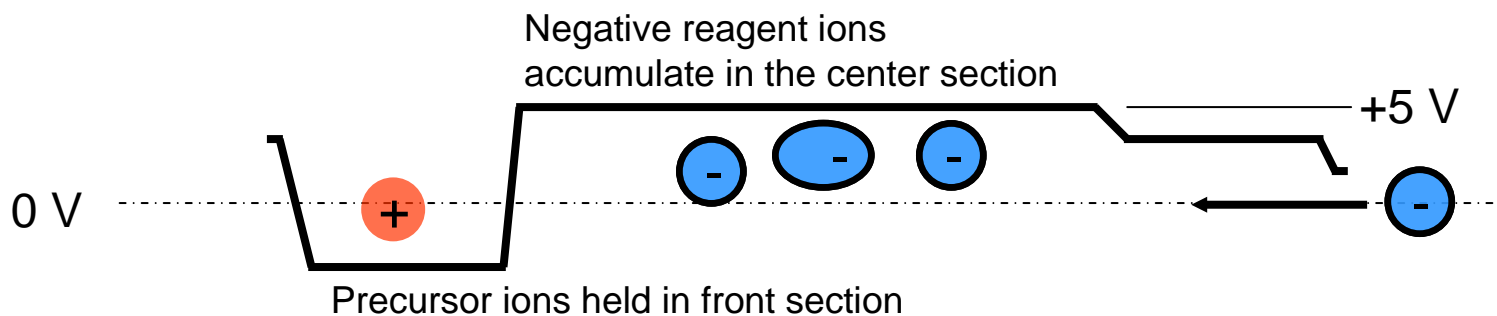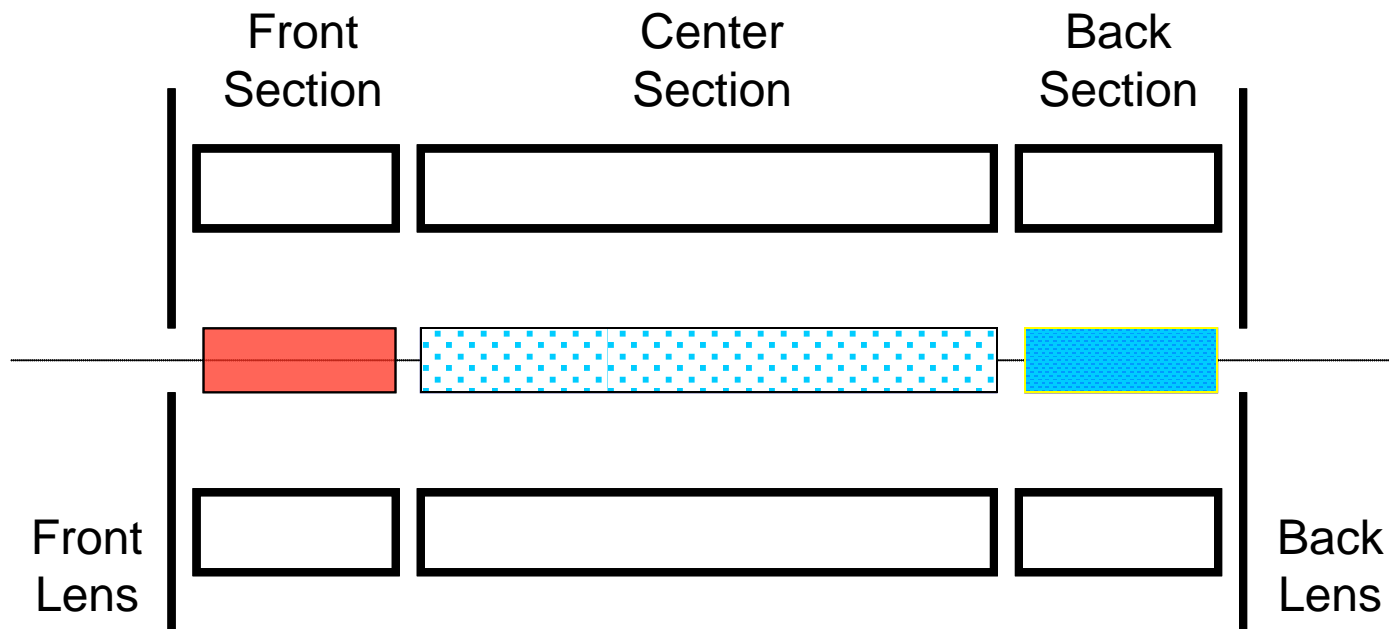
Charge-Sign Independent Trapping
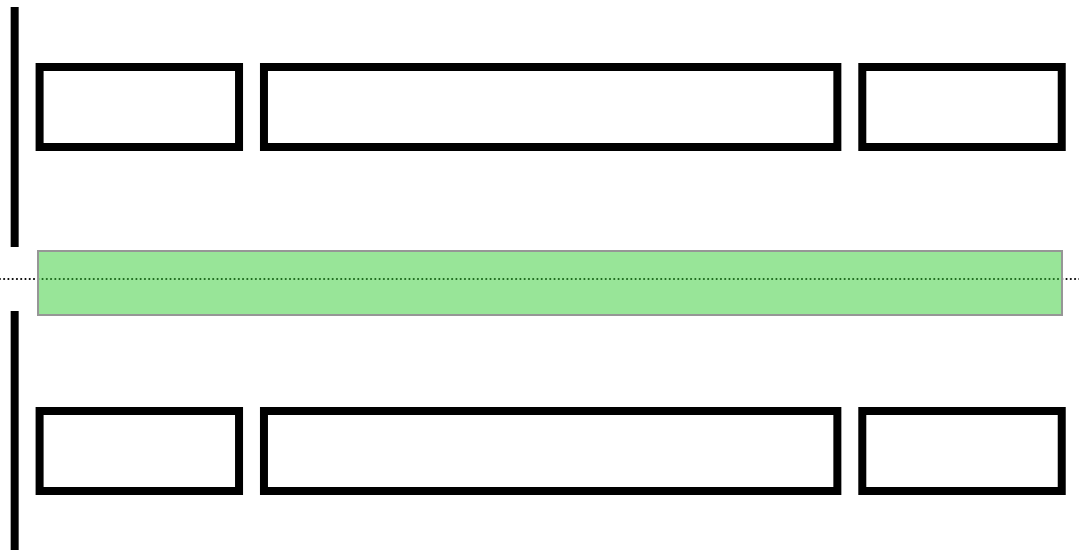
Positive and negative ions react while trapped in axial pseudo-potential

Pseudo-potential created by +150 V_p 600 kHz applied to lenses

+

-

0 V
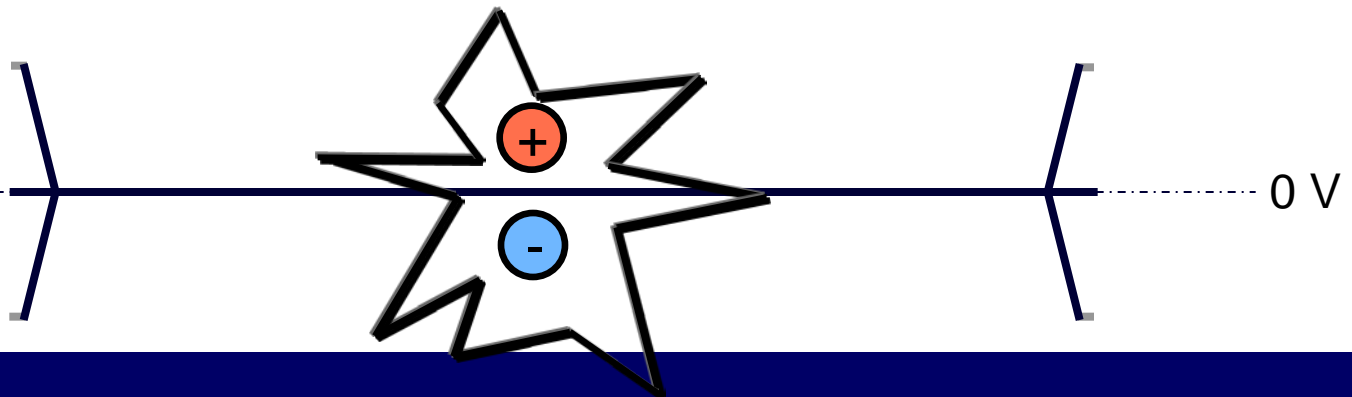
Charge sign independent radial confinement
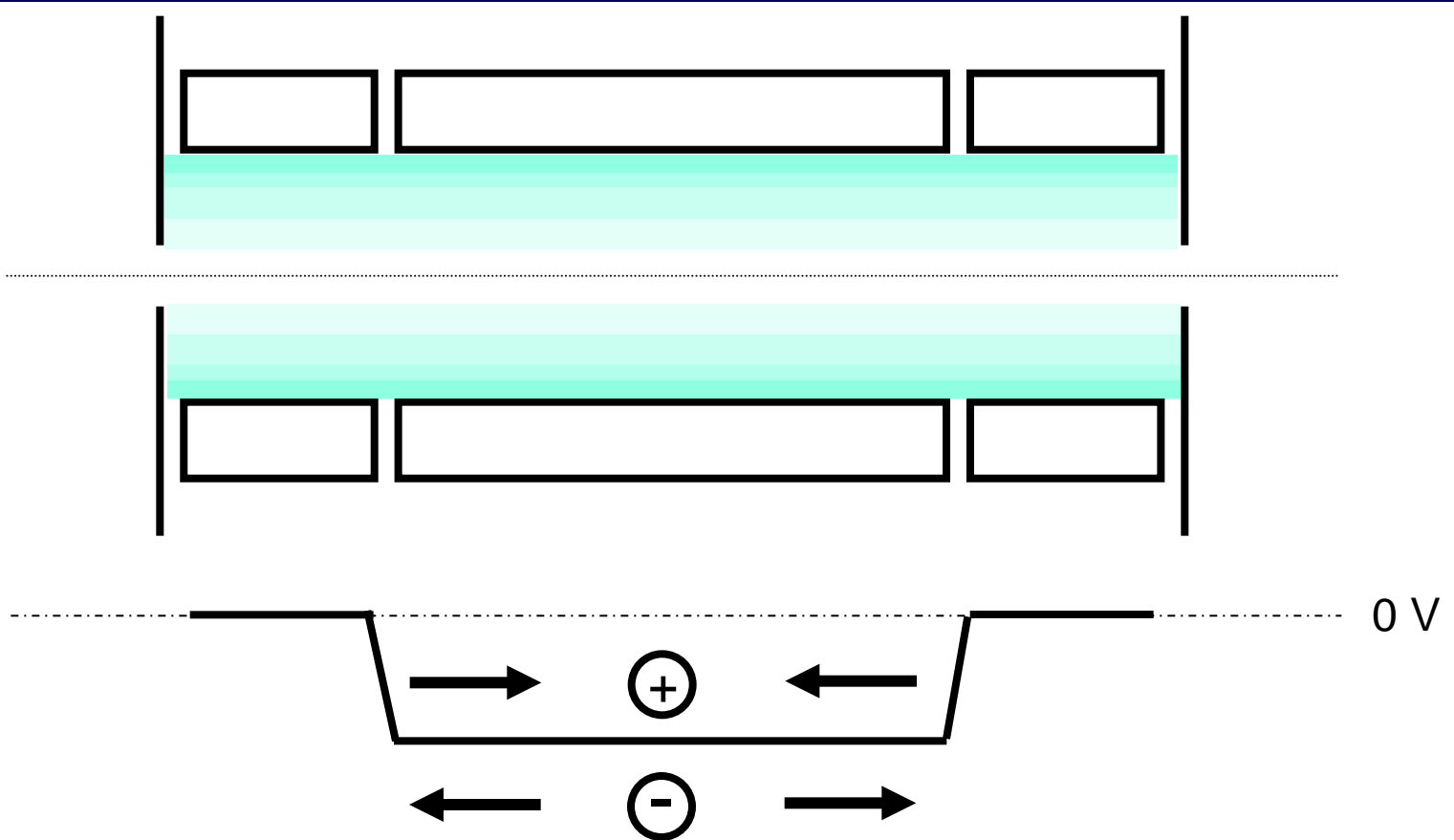
Axial Confinement With DC Potentials
Trapping is Charge Sign Dependent

Charge sign independent axial confinement with combined RF Quadrupole and end lens RF pseudo-potentials
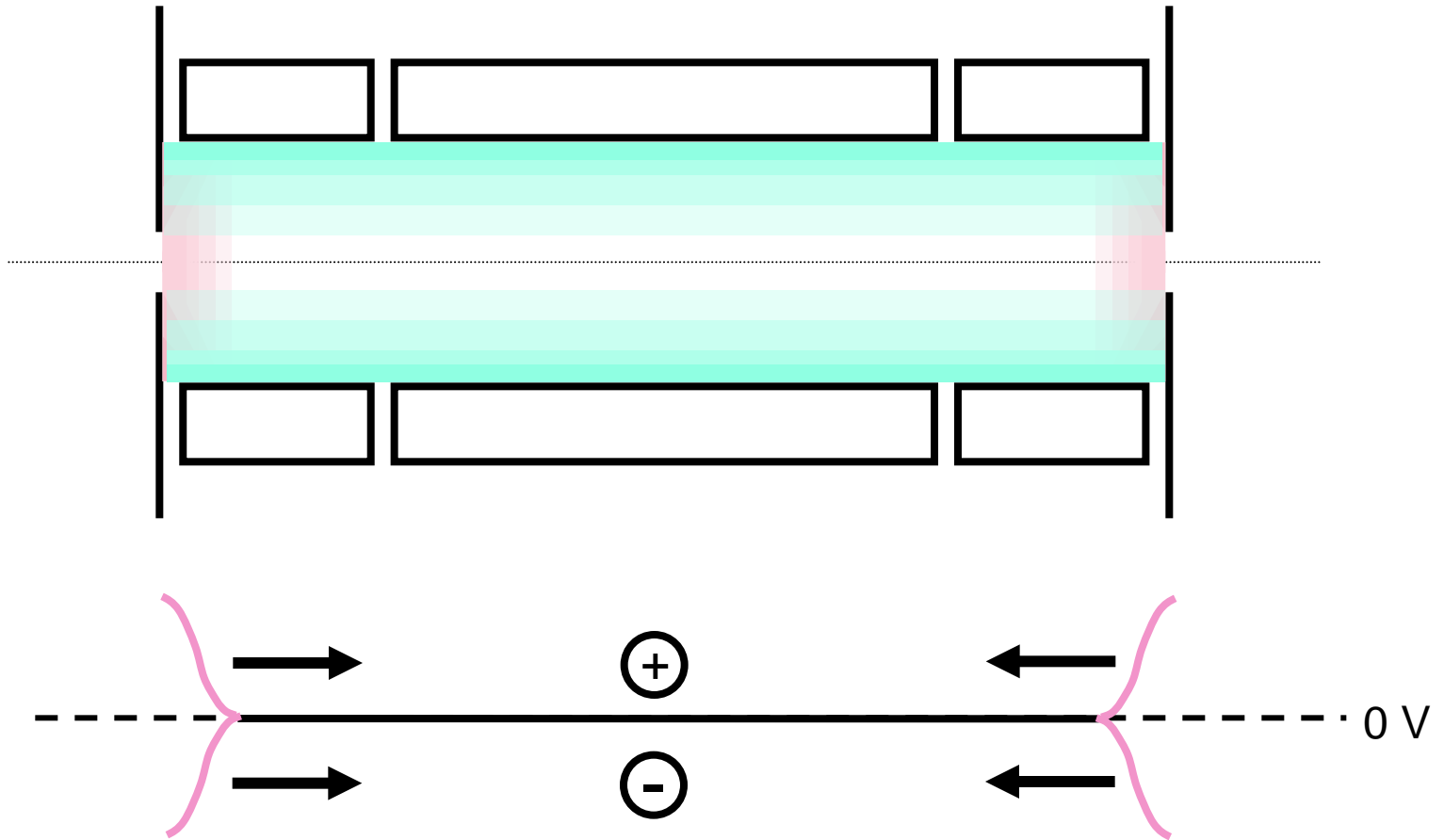
End ion/ion reactions
prepare for product ion analysis
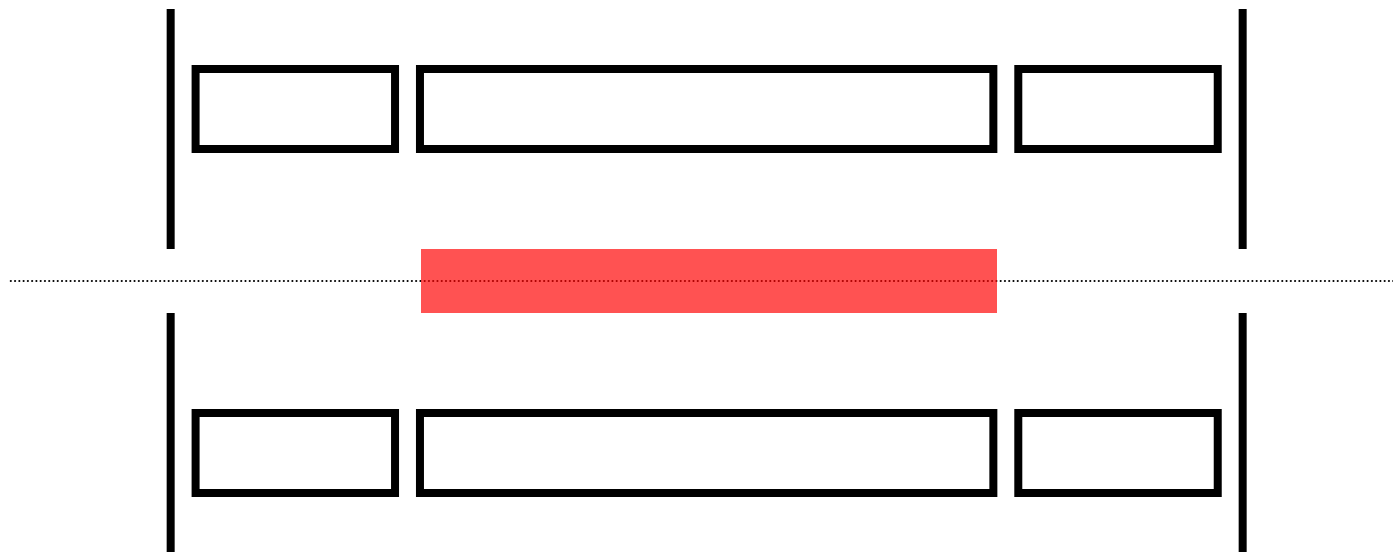
**Product Cations Trapped in
Center Section For Scan Out**

0 V

-12 V

**Reagent Anions
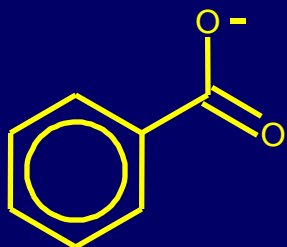Removed Axially**
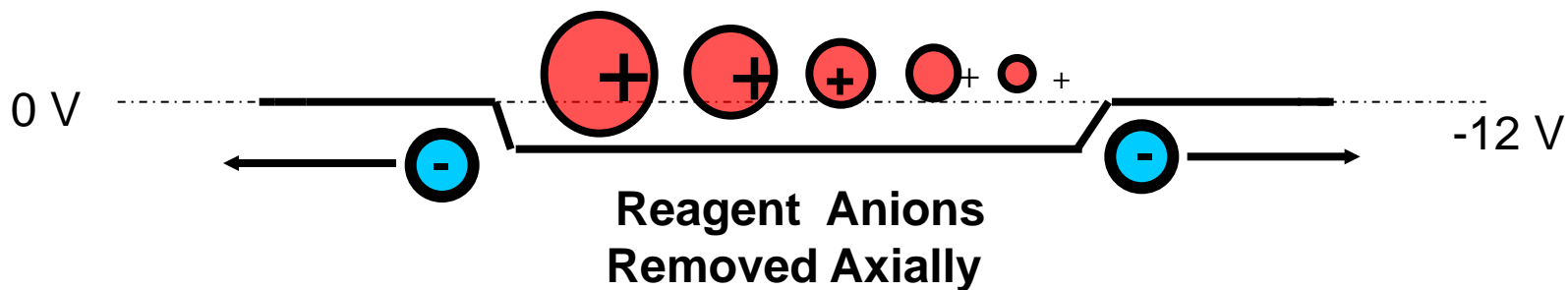
Electron Transfer - Proton Transfer
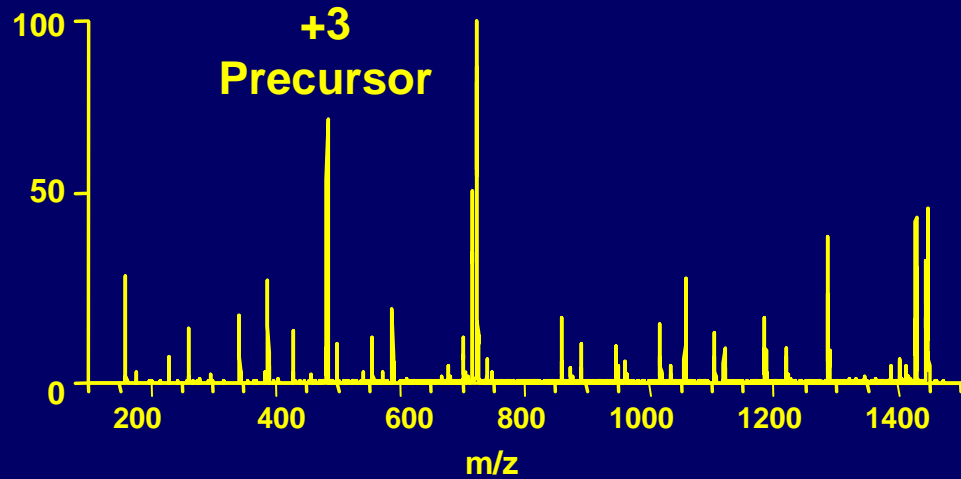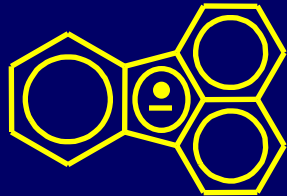
Fragmentation (ETD)
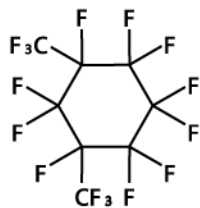
Electron Transfer - Proton Transfer
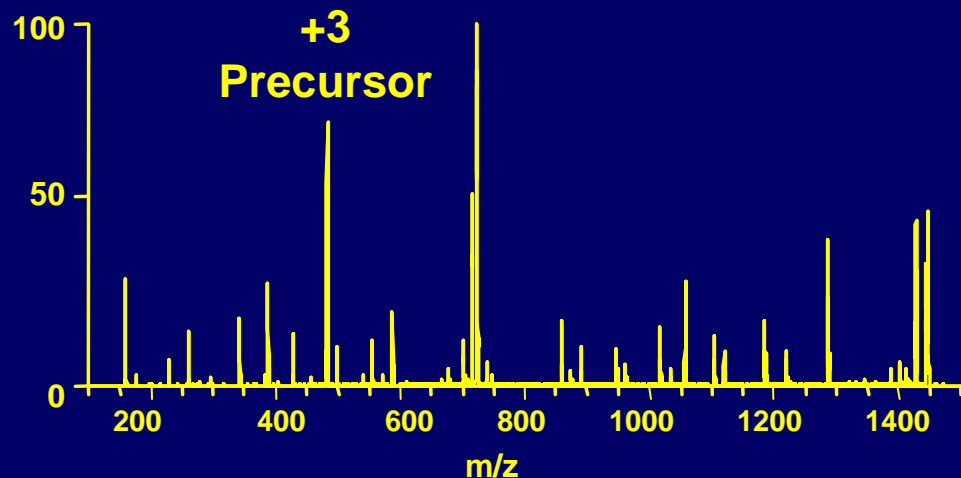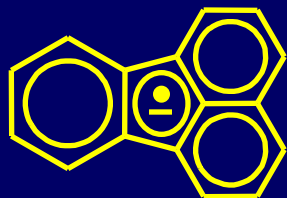
Fragmentation (ETD)

Charge Reduction (PTR)
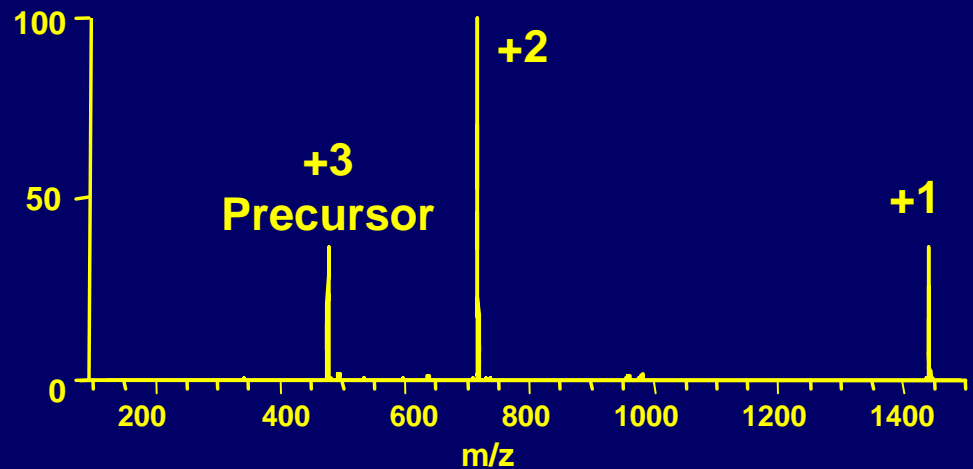
The two types of ion reactions

ET or ETD

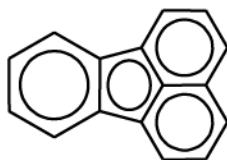$$[M + 3H]^{2+\bullet}$$

Mass ?
m/z ?
Charge (z)
Sequence ?
Temperature ?
Anion ?
He Pressure?

Intact Charge-Reduced Products

$$[M + 3H]^{2+\bullet}$$

Fragmentation Products

c, z, etc.

ET or ETD

$[M + 3H]^{2+\bullet}$

Mass ?
m/z ?
Charge (z)
Sequence ?
Temperature ?
Anion ?
He Pressure?

Intact Charge-Reduced Products

$[M + 3H]^{2+\bullet}$

Fragmentation Products

c, z, etc.
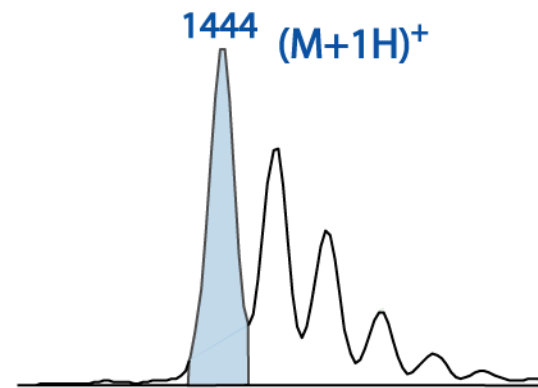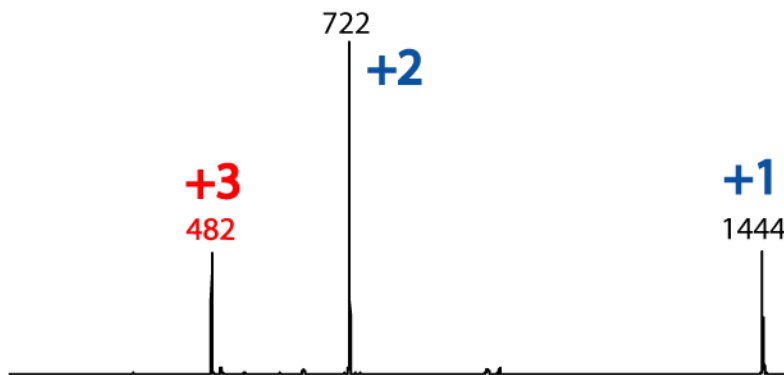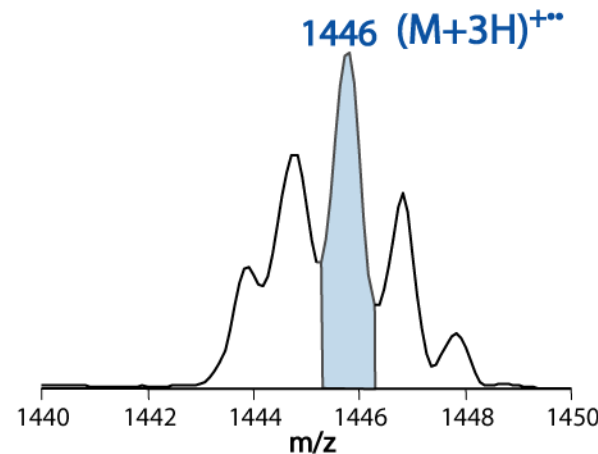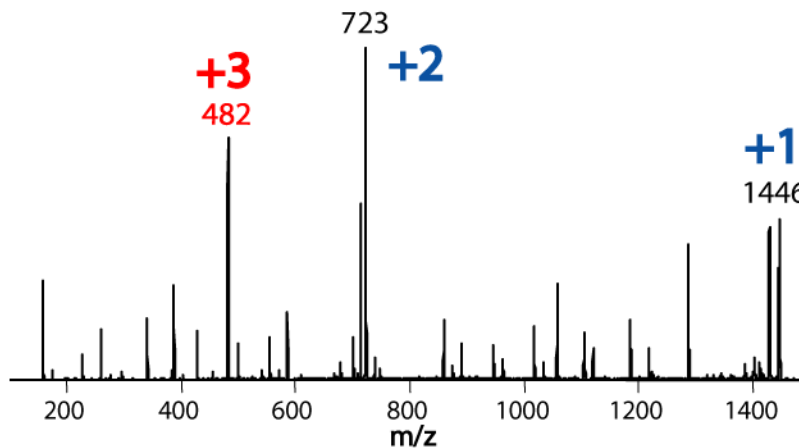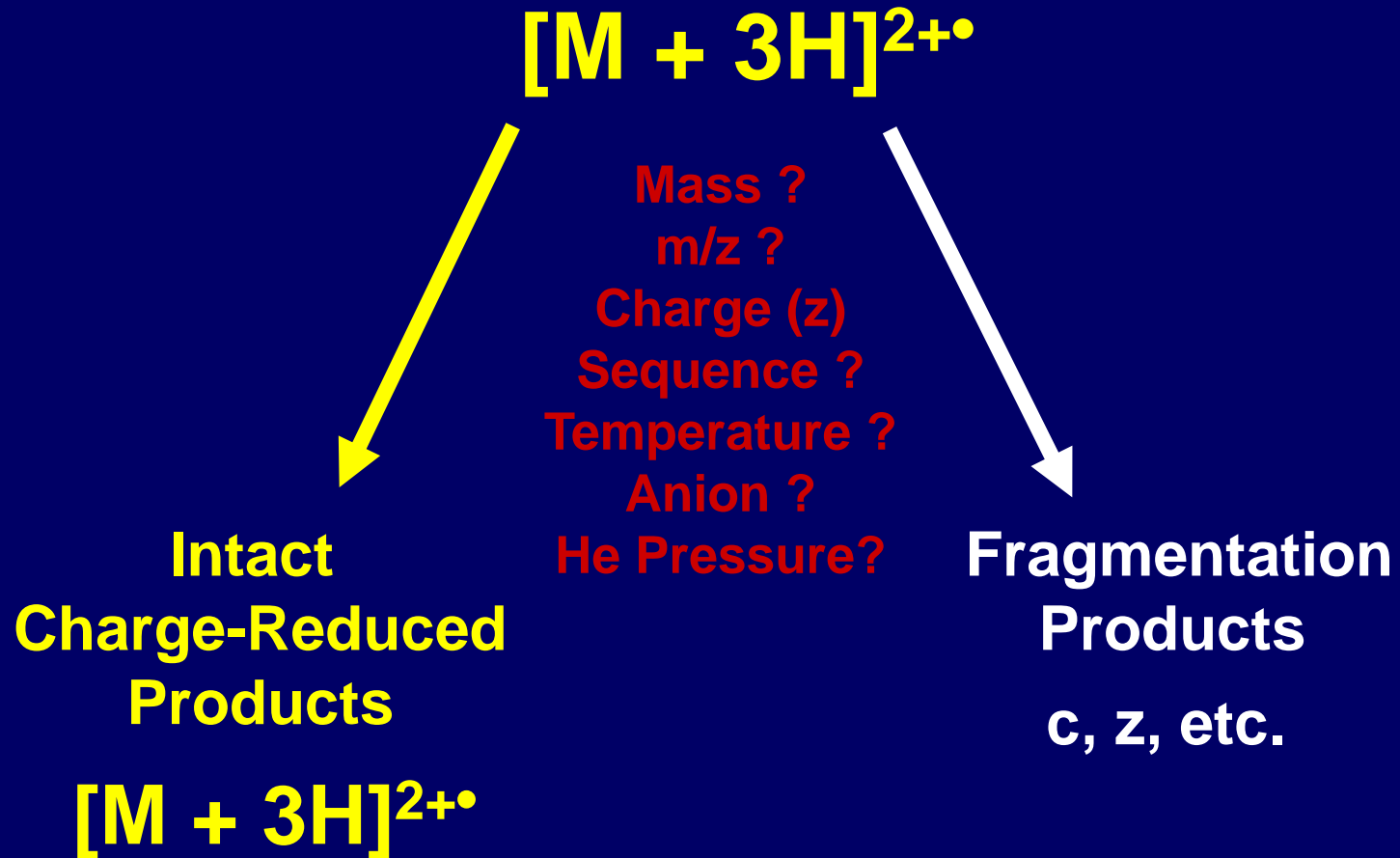
CAD
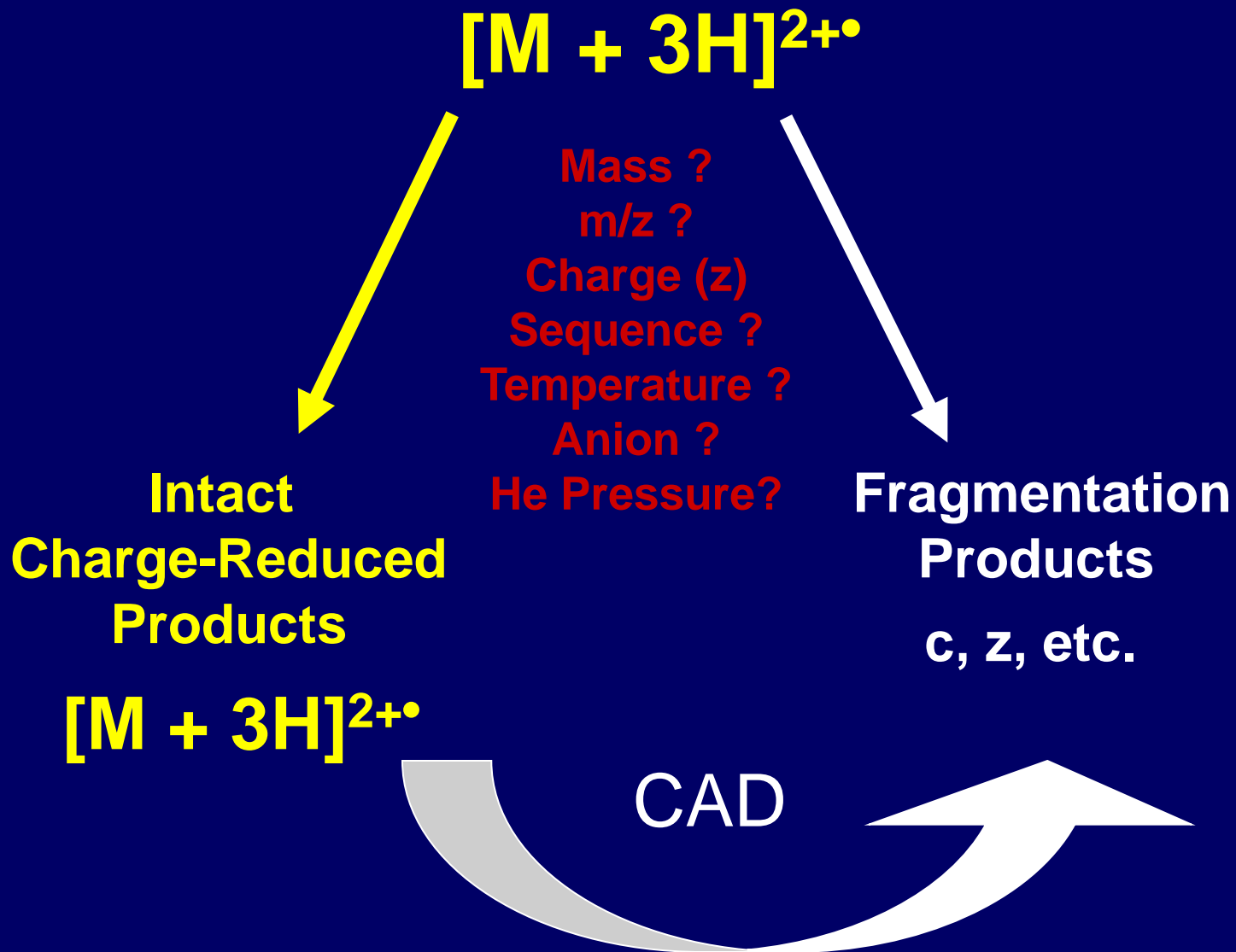
Charge dependence in fragmentation

+6 (m/z 490)

$[M+6H]^{+++\bullet\bullet\bullet}$

$[M+6H]^{++\bullet\bullet\bullet\bullet}$

+5 (m/z 587)

$[M+5H]^{+++\bullet\bullet}$

$[M+5H]^{++\bullet\bullet\bullet}$

+4 (m/z 733)
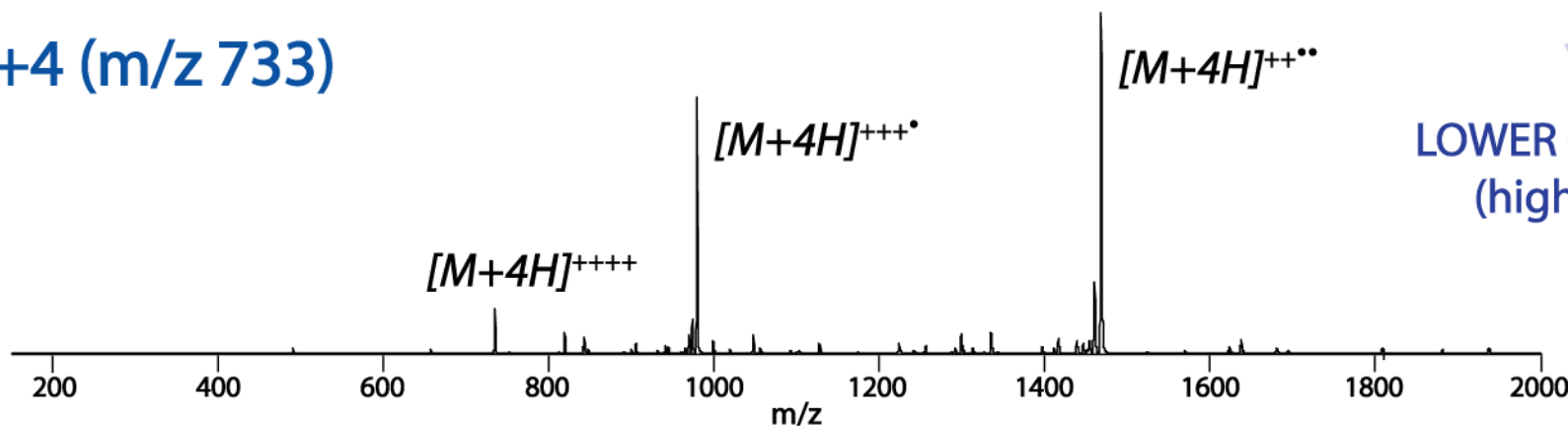
$[M+4H]^{++\bullet\bullet}$

$[M+4H]^{+++\bullet}$

$[M+4H]^{++++}$
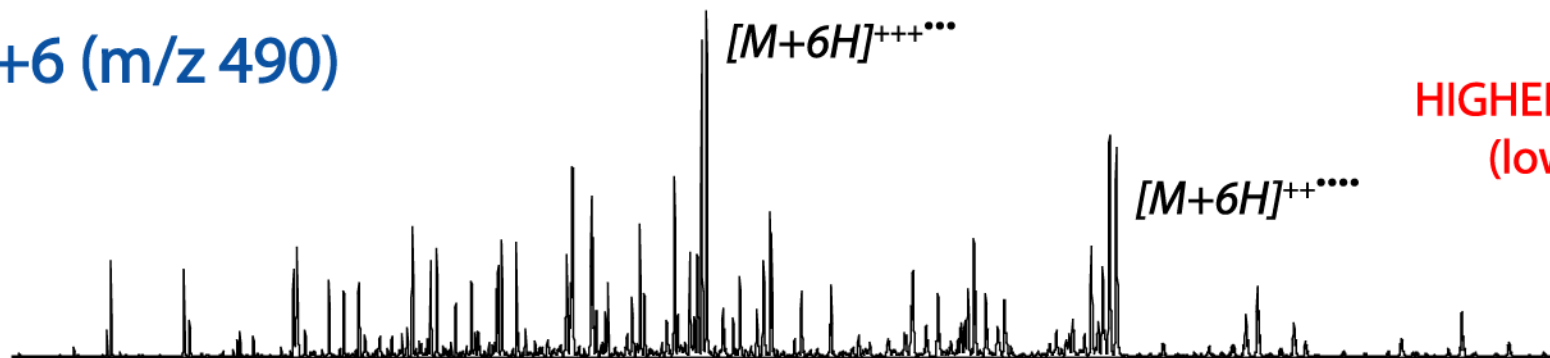
HIGHER CHARGE
(low m/z)
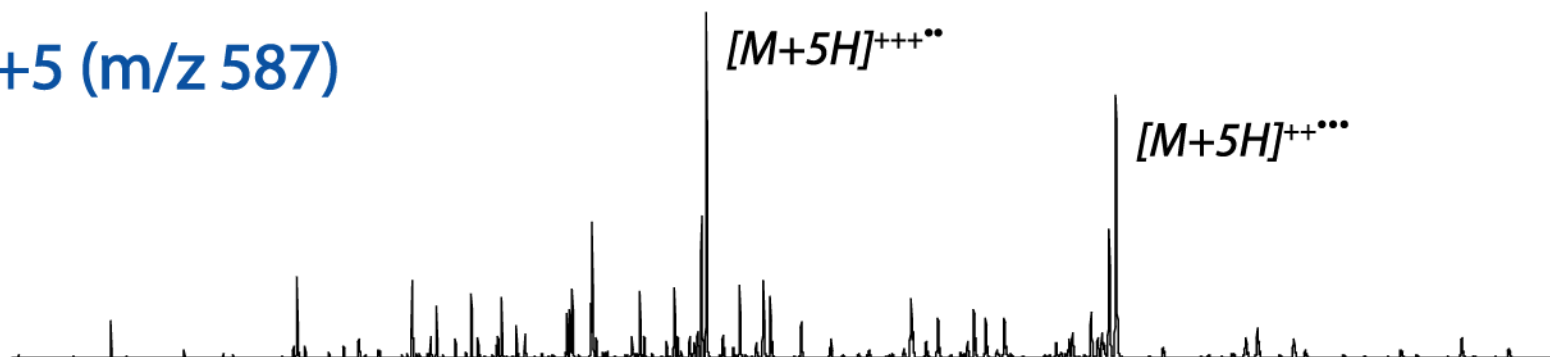
decreasing c/z-product yield
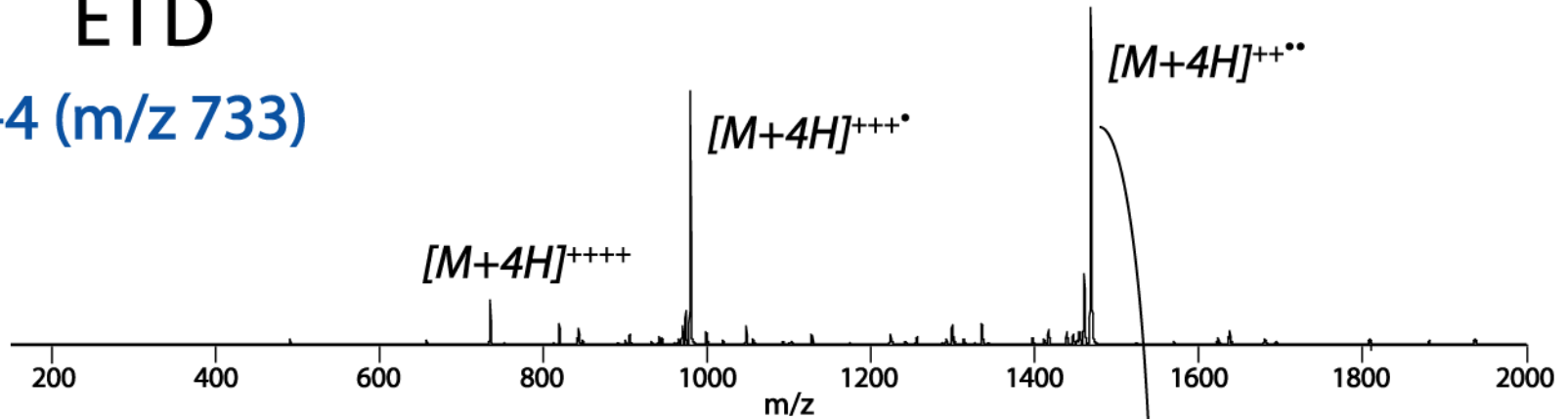
LOWER CHARGE
(high m/z)

m/z

Gentle off resonance activation

ETD

+4 (m/z 733)

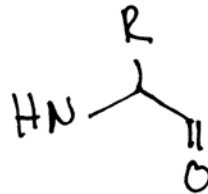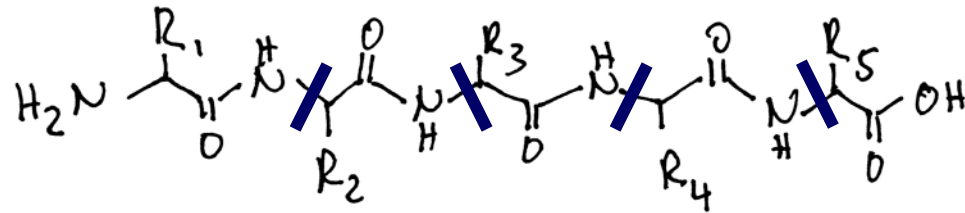$[M+4H]^{++••}$

$[M+4H]^{+++•}$

$[M+4H]^{++++}$

S Y S M E H F R W G K P V G K K R R P V R V Y P

gentle activation (q = 0.15)

x5 x3 x3 x8

z_{20}^{++}

c_{21}^{++} z_{21}^{++}

c_{20}^{++}

c_7

z_7 z_{17}^{++} z_{18}^{++} z_{19}^{++}

c/z_{22}^{++}

side-chain losses from ETD

z_{14}

c_{10} z_{10} z_{23}^{++}

c_6 z_8 c_8 z_9 z_{11}

c_{13} c_{15}

c_{19}^{++} c_9 z_{12} c_{12} c_{14} z_{16} c_{16}
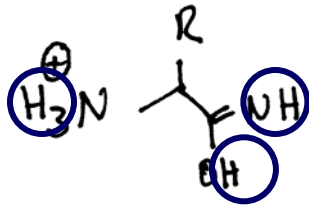
z_{15}
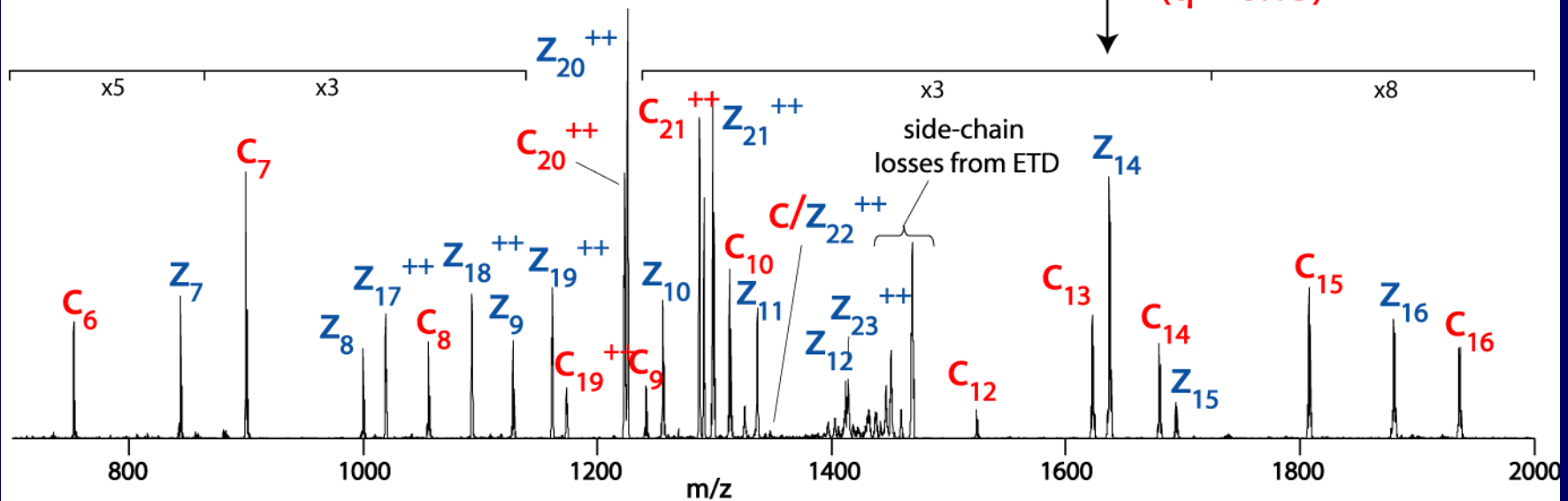
How to Sequence ETD

Residue Mass (RM)

c1 ion

z1 ion

+ NH

c1 = RM + 18
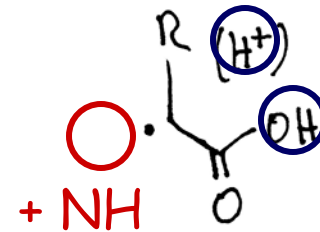
z1 = RM + 3

Example of how to calculate theoretical fragment ions

105	176	307	404	517	646	803
S	A	M	P	L	E	R
803	700	629	498	401	288	159
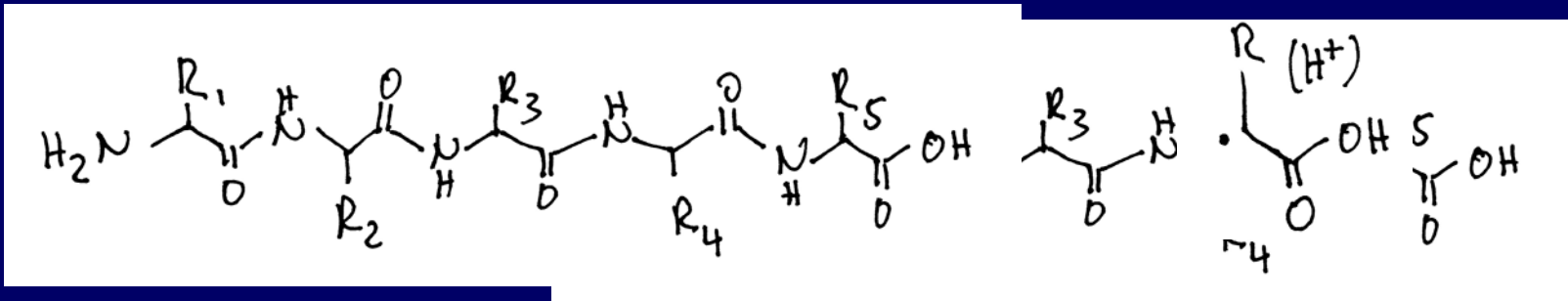
The first c ion

Residue Mass

The first z ion

largest z ion = $(n+H)^{+} - (H + HN$ $+ NH)$
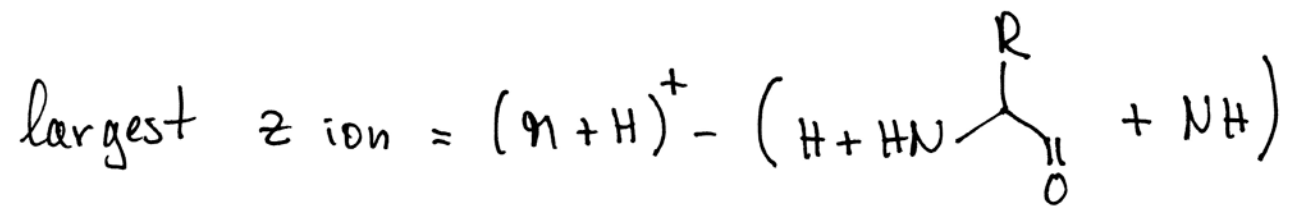
Background necessary to know
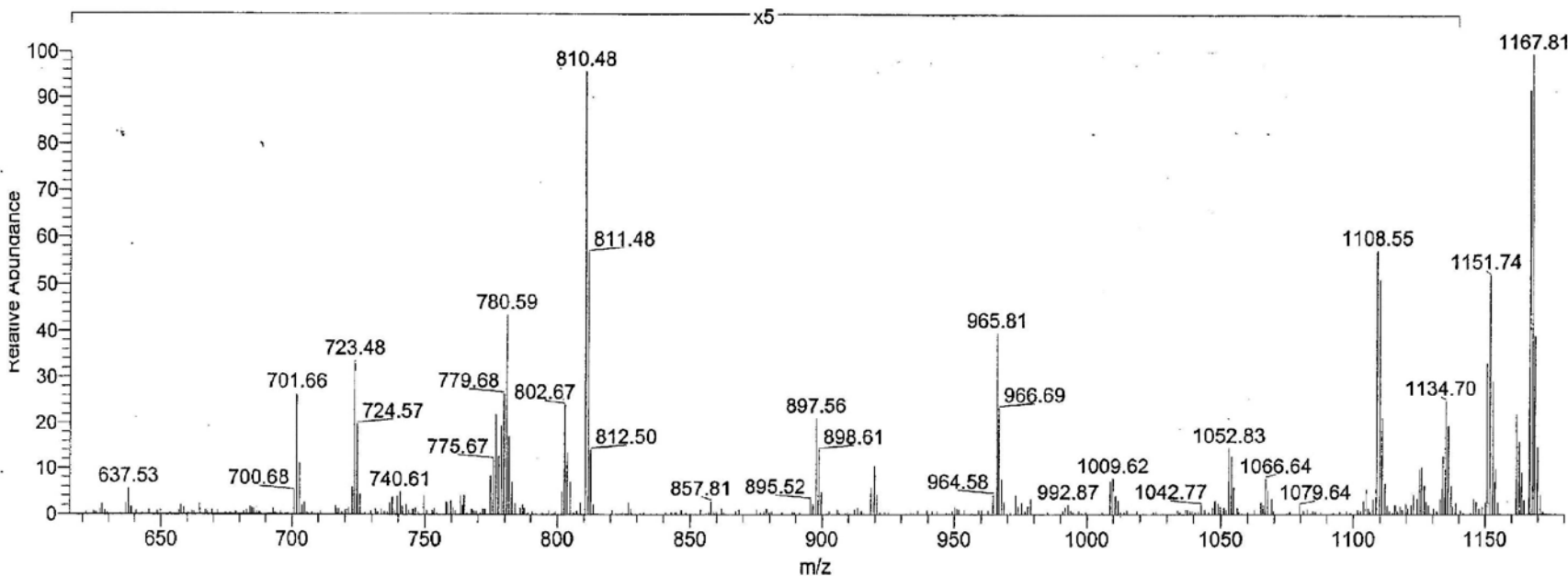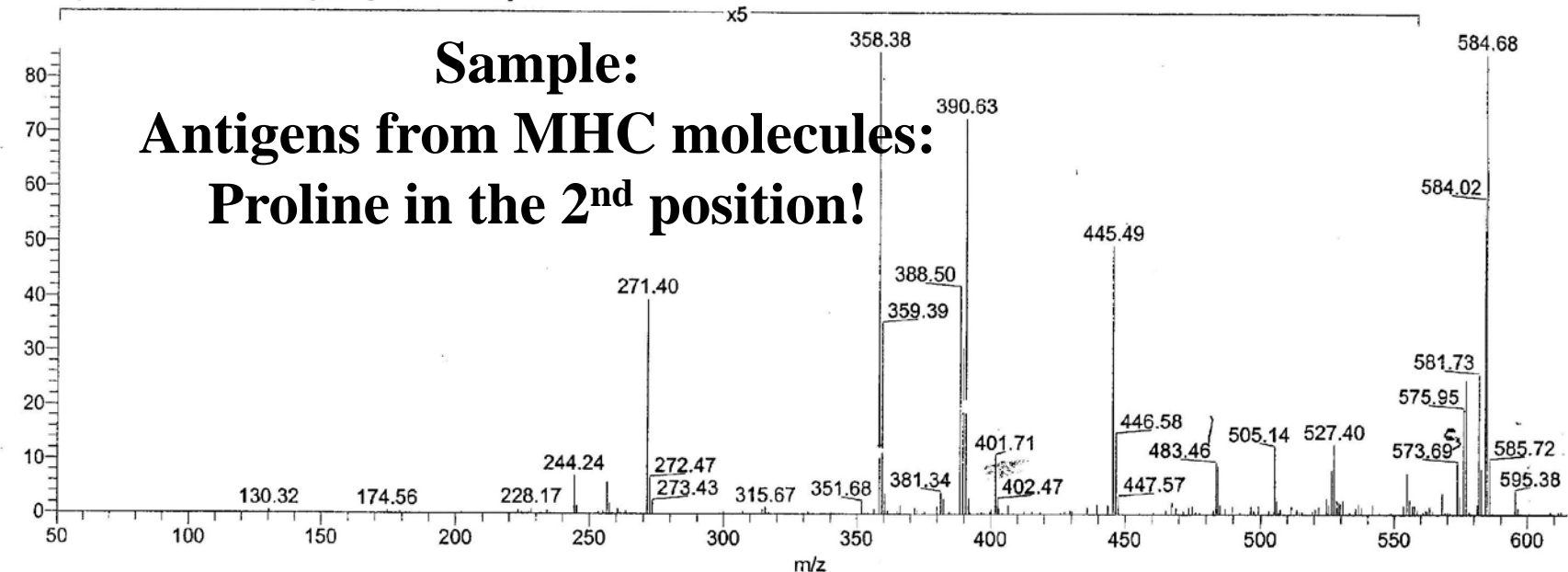
Show a hand annotated spectra from
one of my toxin talks
Show how to calculate a charge state,
also in a spectrum, give a spectrum
where they should find out the charge
state for CAD and then for ETD and
how they confirm that by finding pairs
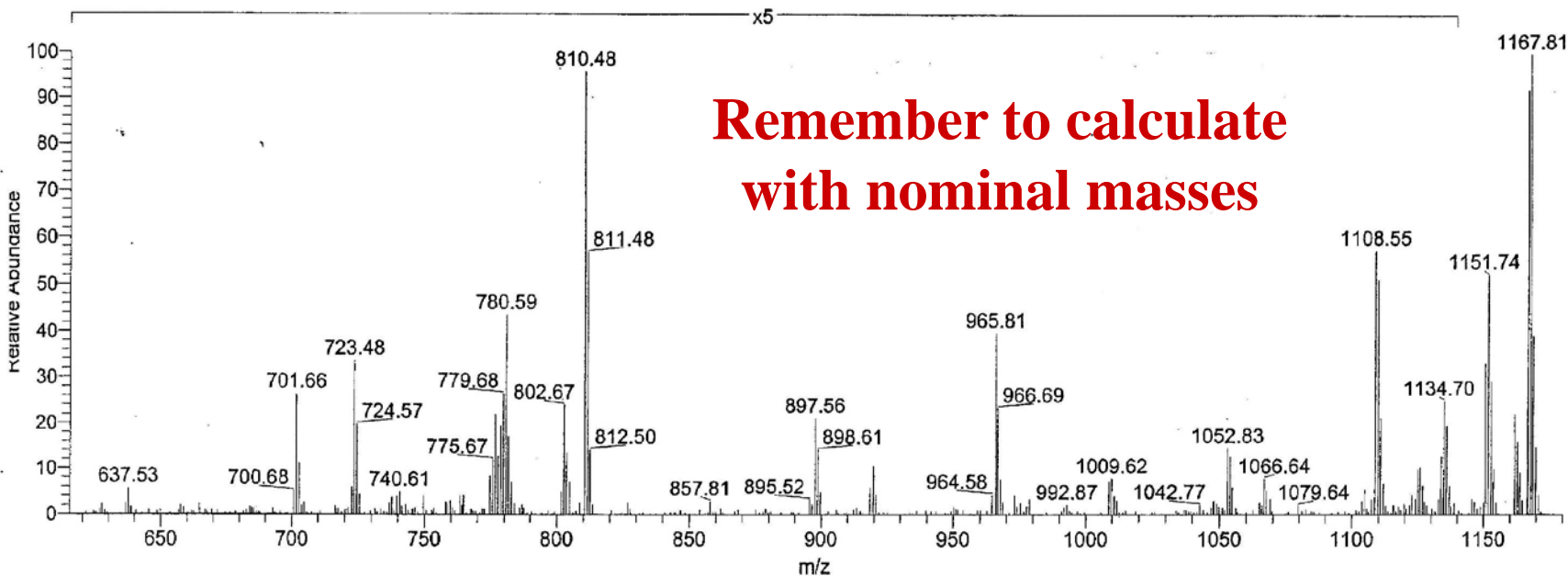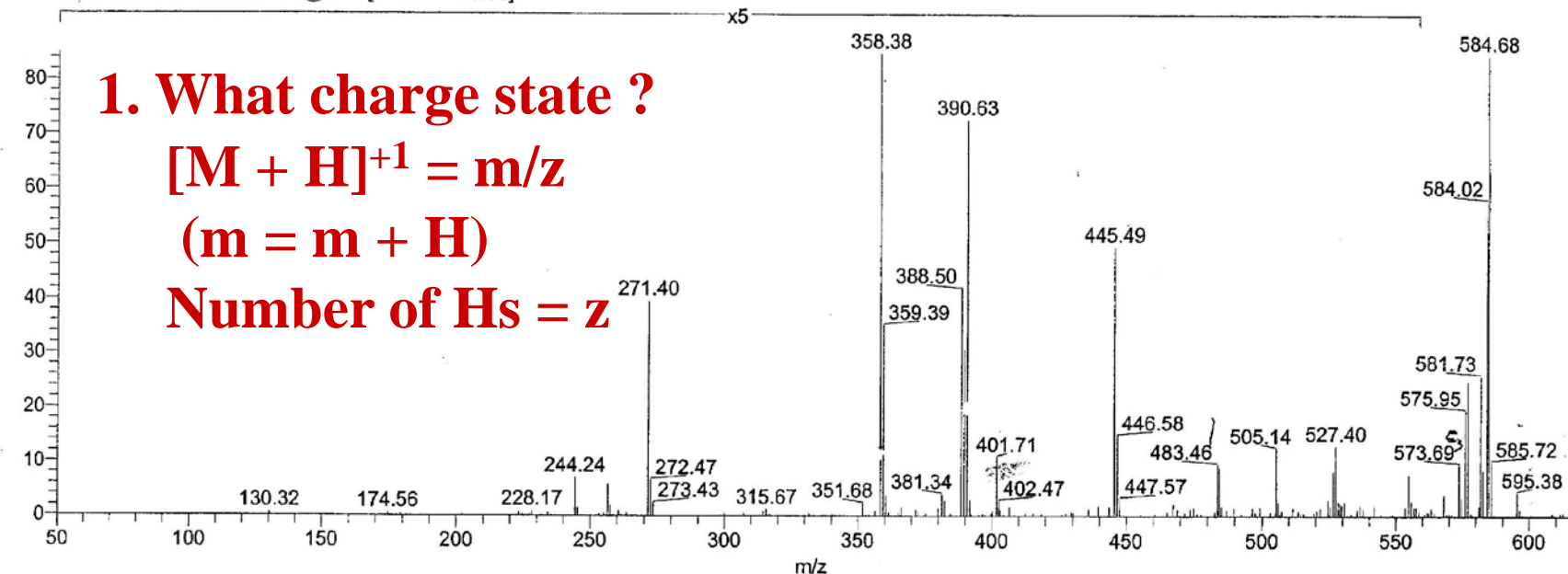(must have mentioned that in the
previous slide for the example)
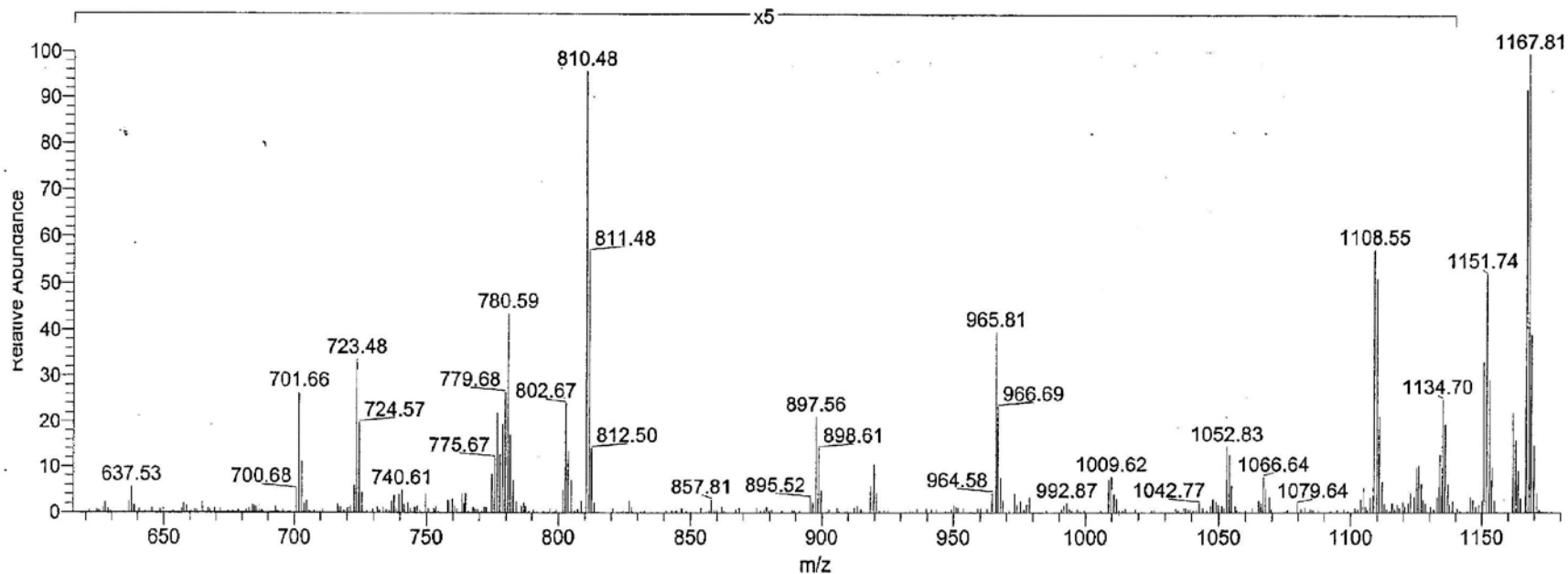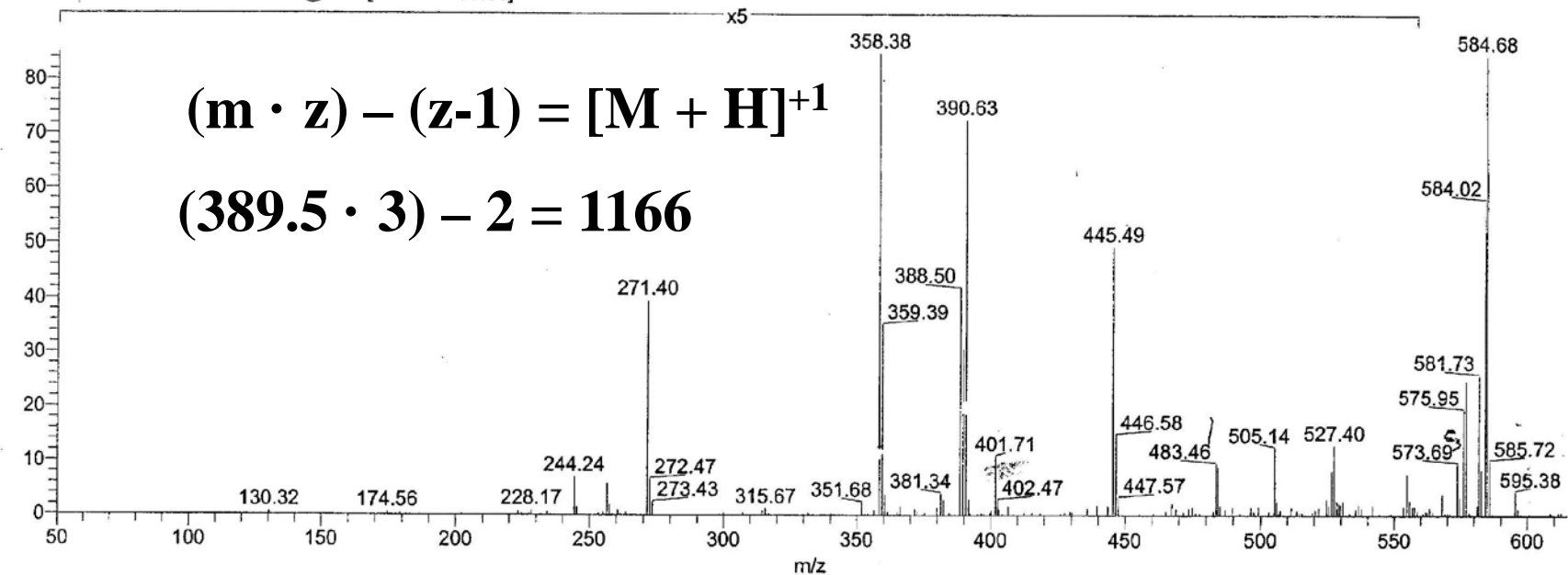Ask them if they know how to calculate
a charge state etc.
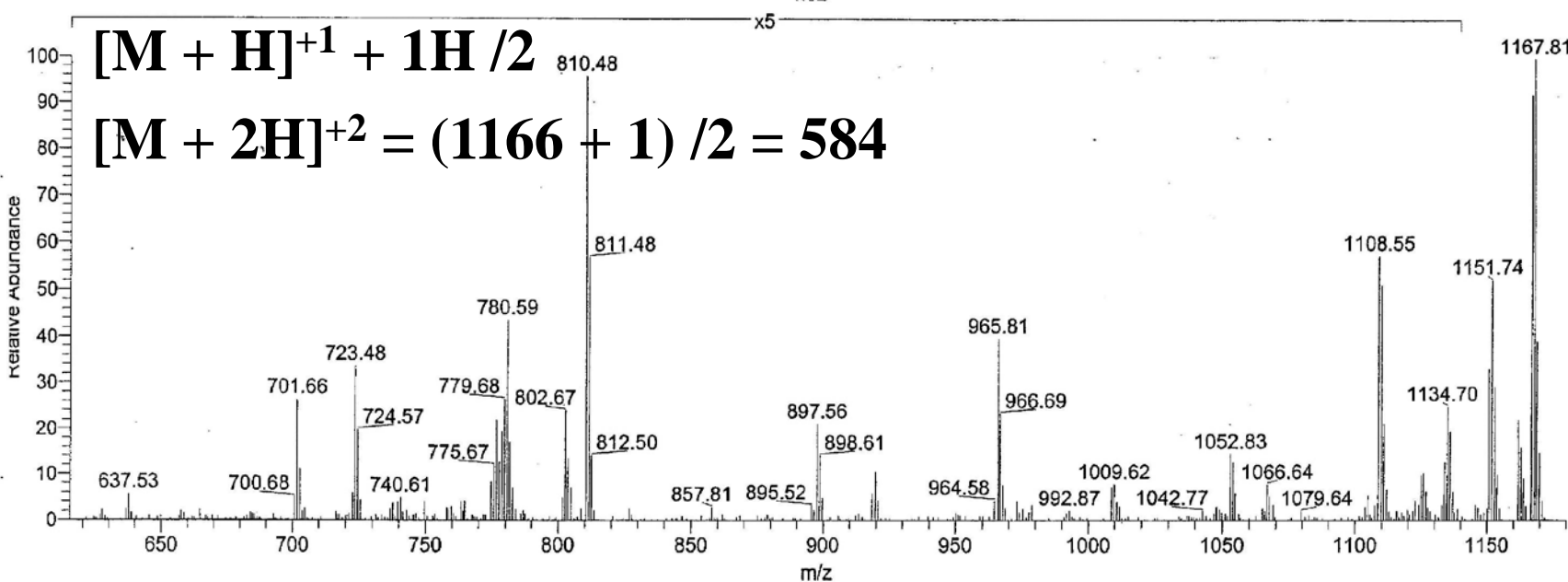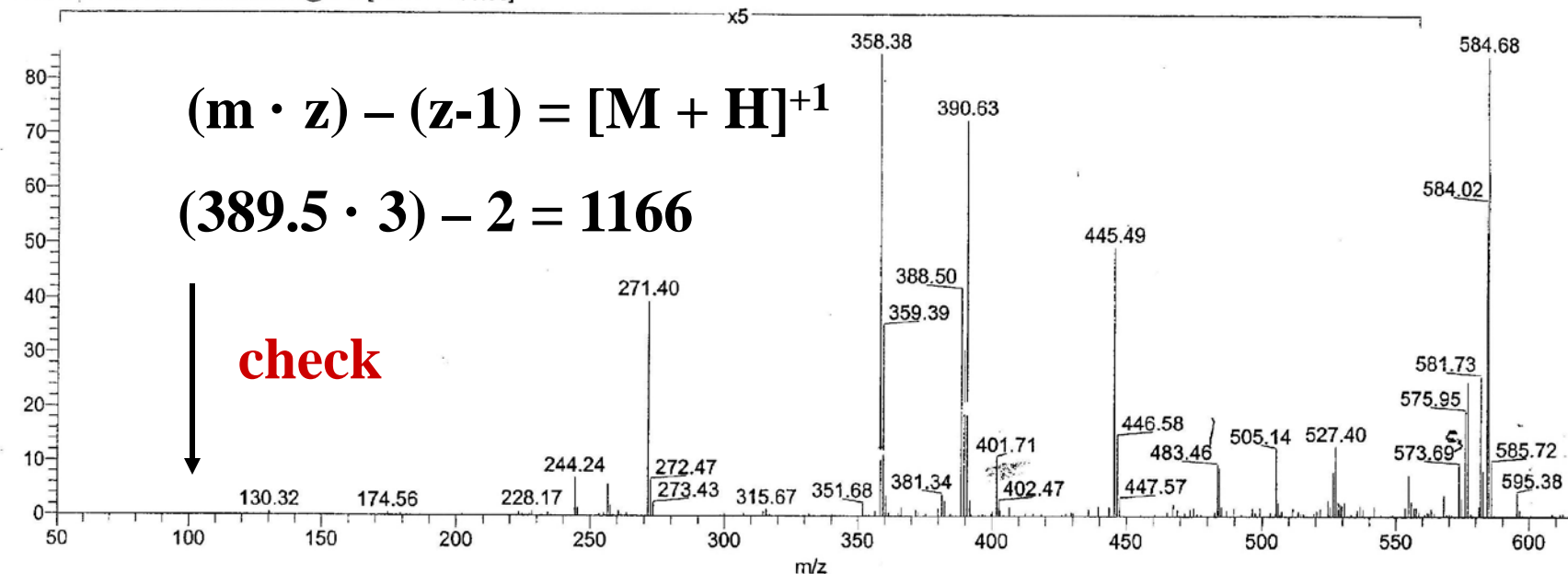
Characteristics of CAD and ETD spectra

Show the precursor issue, show the neutral loss from the parent, show the neutral loss from peptides

Sample:
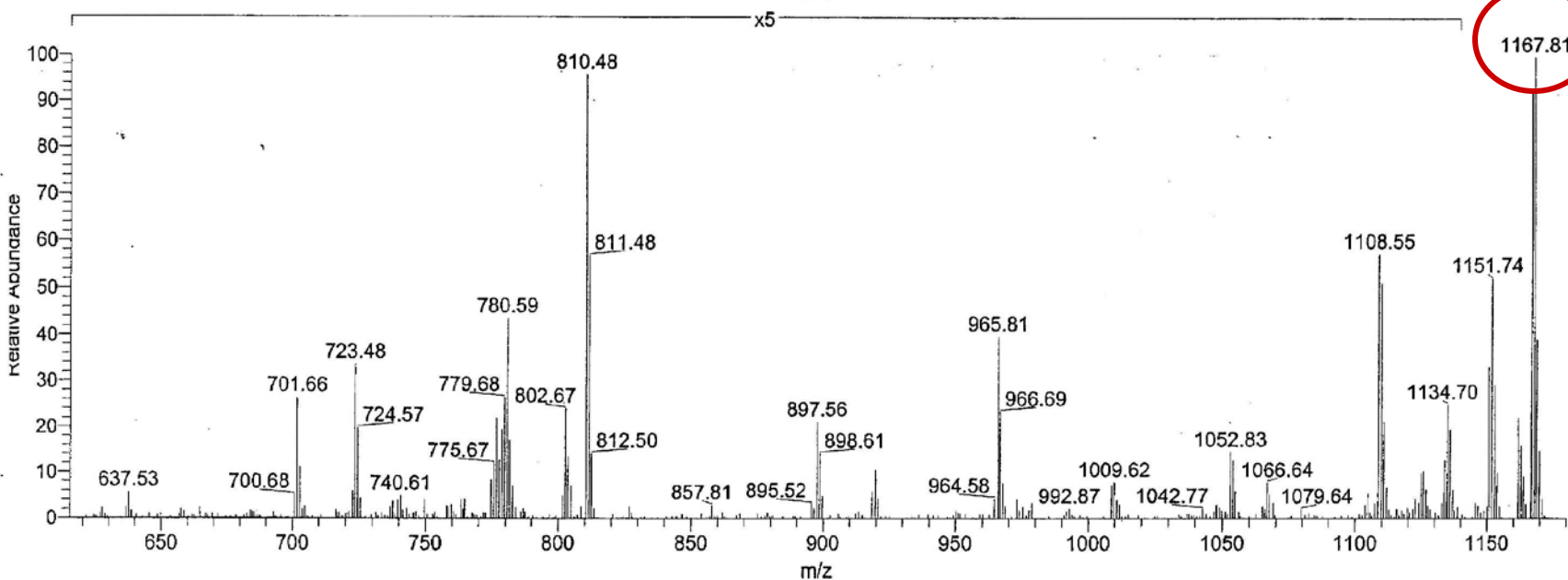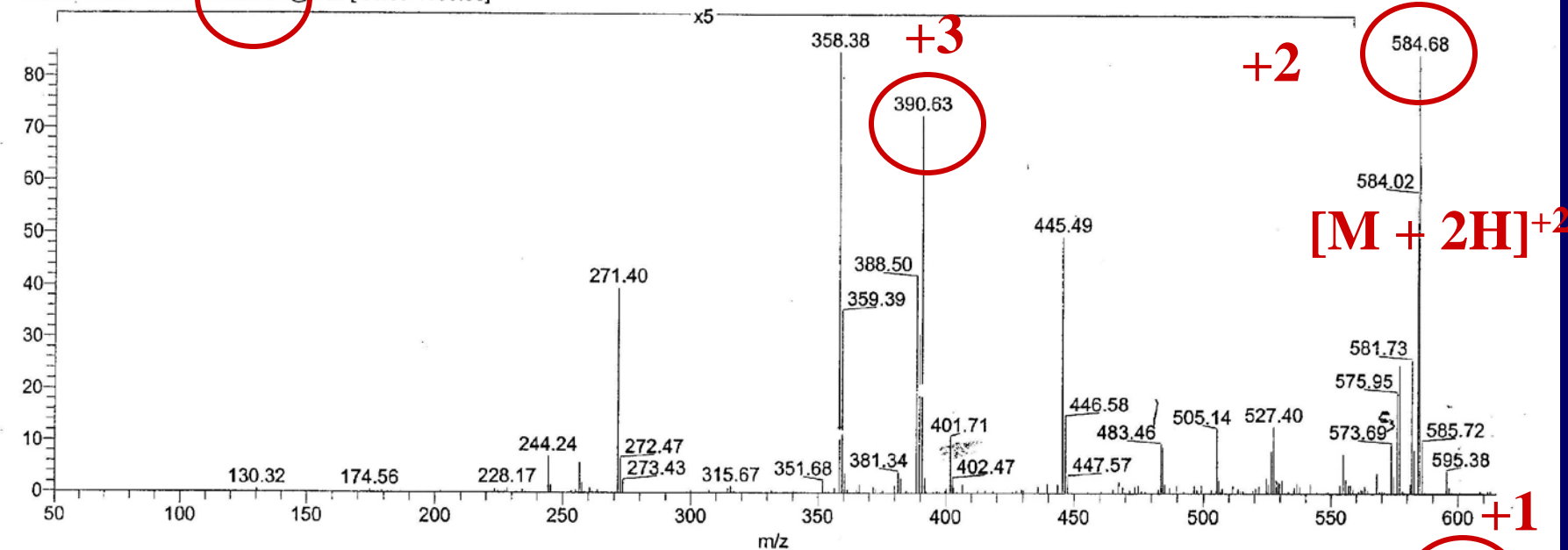Antigens from MHC molecules:
Proline in the 2nd position!

1. What charge state ?
$[M + H]^{+1}$ = m/z
(m = m + H)
Number of Hs = z

Remember to calculate
with nominal masses

J:\Melanie\ETD\JYB7_031104_040311132712   03/11/2004 04:15:45 PM   JYB7
20 microliters, ~ 8 E 7 cell equivalents
YB7_031104_040311132712 #1172   RT: 15.11   AV: 1   NL: 3.38E4
: ITMS + c NSI d Full ms2 389.94@0.00 [50.00-1180.00]

$$(m \cdot z) - (z\text{-}1) = [M + H]^{+1}$$

$$(389.5 \cdot 3) - 2 = 1166$$

check

$$[M + H]^{+1} + 1H\,/2$$

$$[M + 2H]^{+2} = (1166 + 1)\,/2 = 584$$

2. **Eliminate non-sequencing relevant ions!**

3. **biggest c ion ?**

4. **biggest z ion ?**

J:\Melanie\ETD\JYB7_031104_040311132712   03/11/2004 04:15:45 PM   JYB7
0 microliters, ~ 8 E 7 cell equivalents
YB7_031104_040311132712 #1172 RT: 15.11 AV: 1 NL: 3.38E4
: ITMS + c NSI d Full ms2 389.94@0.00 [50.00-1180.00]

Now find the first c ion (c_1), look for c and z pairs, and sequence ladders

Let's find c and z pairs!

$$z_i + c = [M + H]^{+1} + 2$$

$$[M + H]^{+1} - c + 2 = z$$

x5

358.38
390.63
584.68
584.02
271.40
388.50
359.39
581.73
575.95
446.58
505.14 527.40
401.71
483.46
573.69
585.72
244.24
272.47
130.32   174.56   228.17   273.43   315.67 351.68 381.34 402.47 447.57   595.38

m/z

1052 1166

x5

__ **P** __ __ __ __ __ __ **I/L**

1166   **130**

810.48
1167.81
811.48
1108.55
1151.74
780.59
965.81
723.48
1134.70
701.66
779.68 802.67   966.69
724.57
775.67 812.50   897.56   1052.83
637.53   700.68   740.61   857.81 895.52 964.58 898.61 1009.62 1066.64
992.87 1042.77 1079.64

m/z

J:\Melanie\ETD\JYB7_031104_040311132712   03/11/2004 04:15:45 PM   JYB7
20 microliters, ~ 8 E 7 cell equivalents
YB7_031104_040311132712 #1172 RT: 15.11 AV: 1 NL: 3.38E4
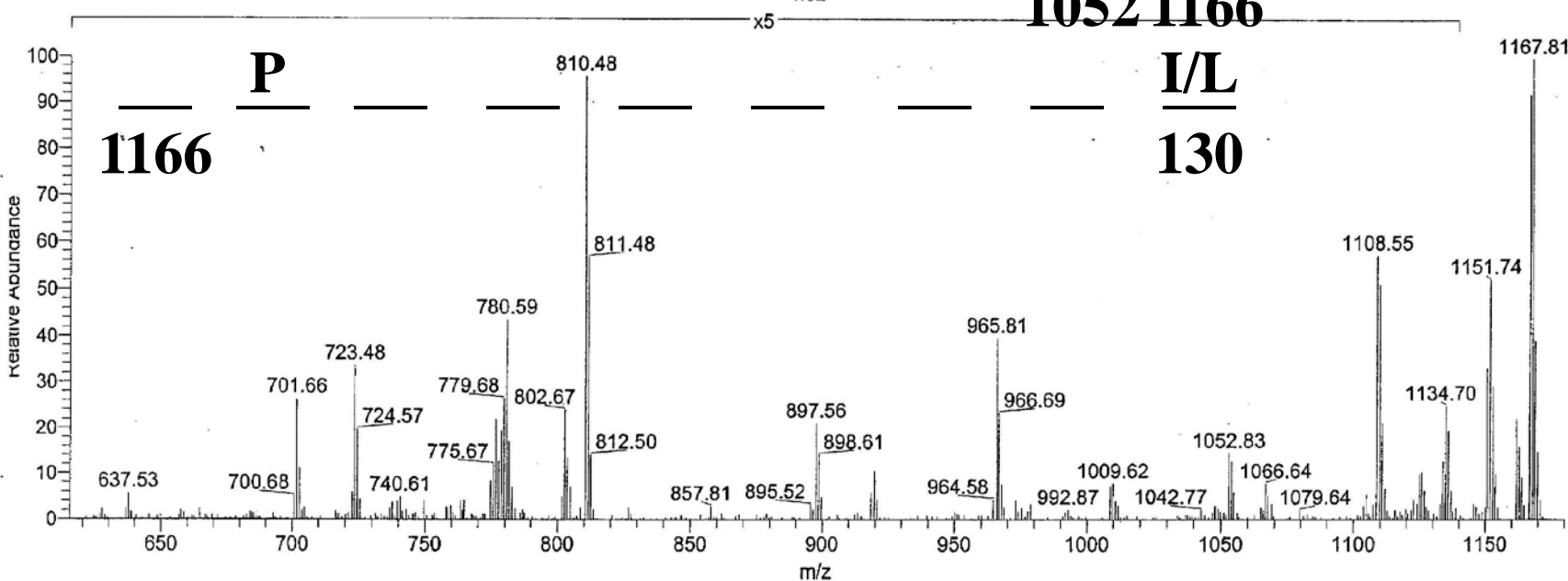: ITMS + c NSI d Full ms2 389.94@0.00 [50.00-1180.00]
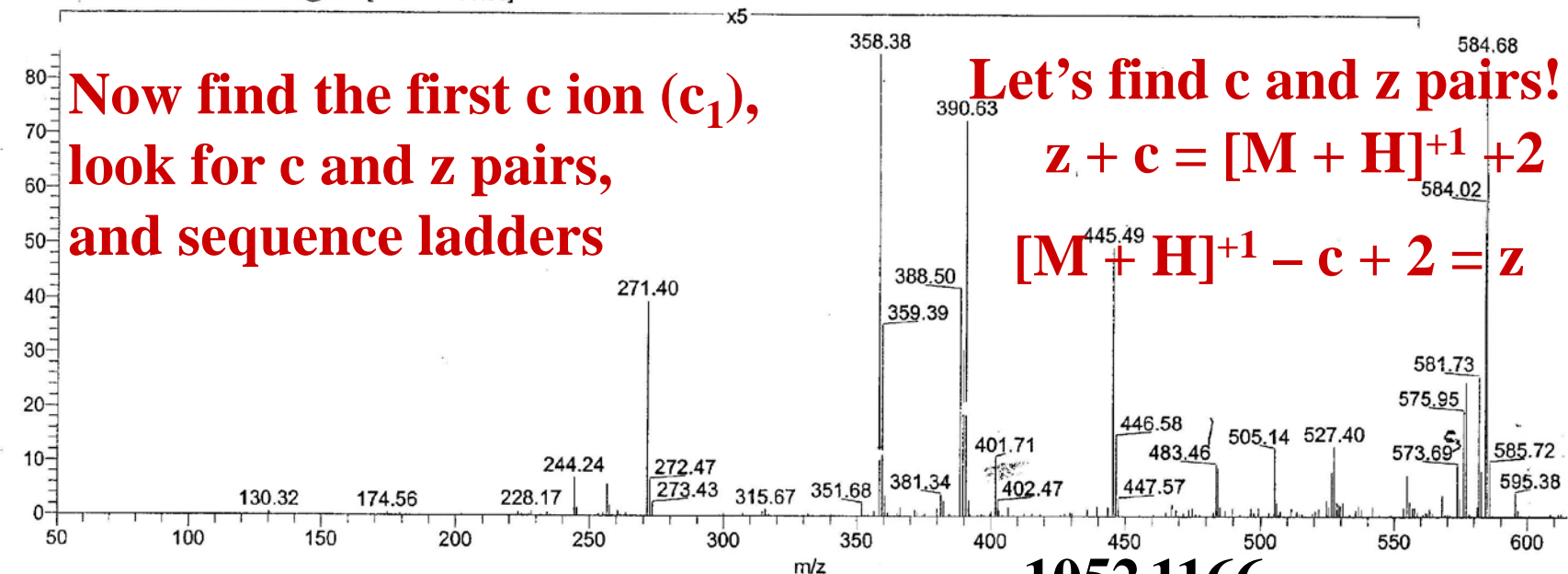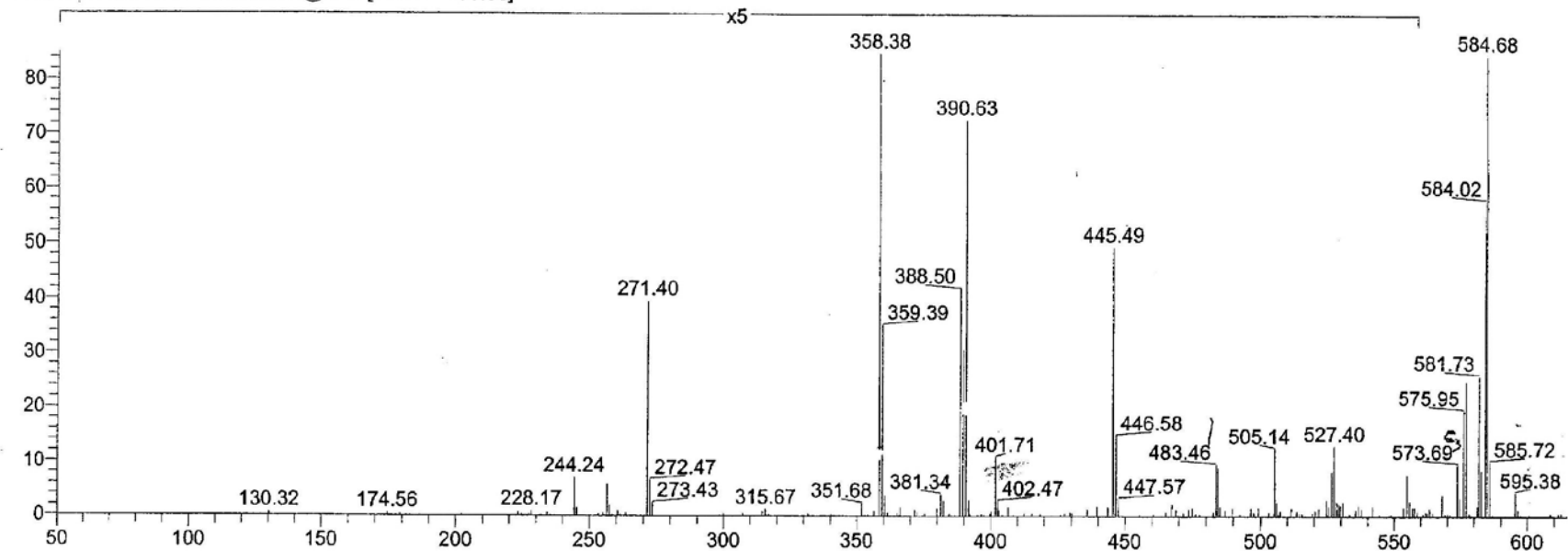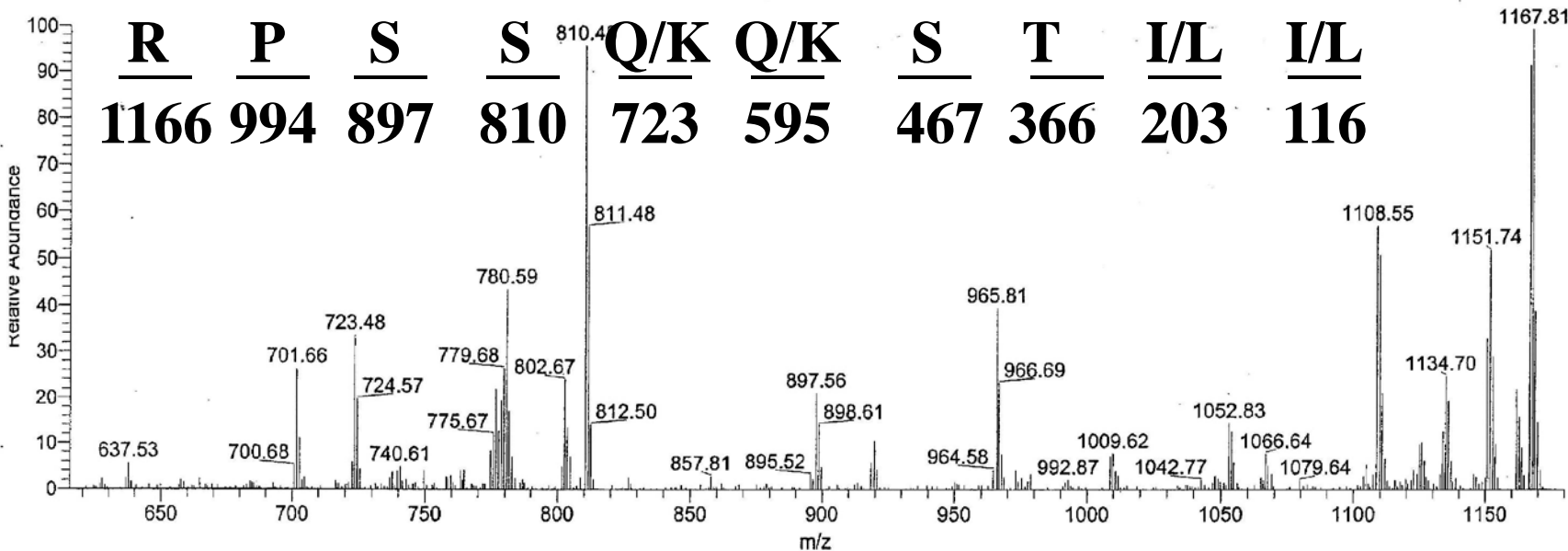
174 271 358 445 573   701   802   965 1052 1166

R   P   S   S   Q/K Q/K   S   T   I/L   I/L
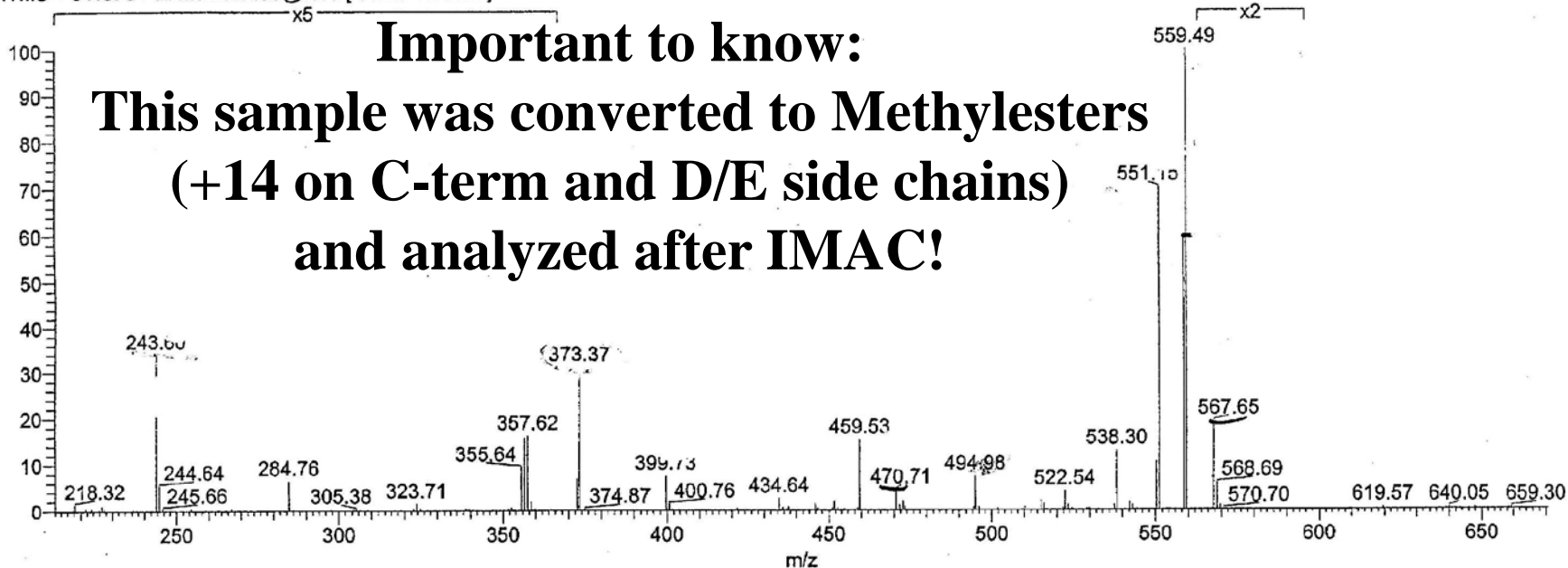
1166 994 897 810 723 595   467 366   203   116

2nd Example

Important to know:
This sample was converted to Methylesters
(+14 on C-term and D/E side chains)
and analyzed after IMAC!

Sample:
Antigens from MHC molecules: 9mer
with Proline in the 2nd position!
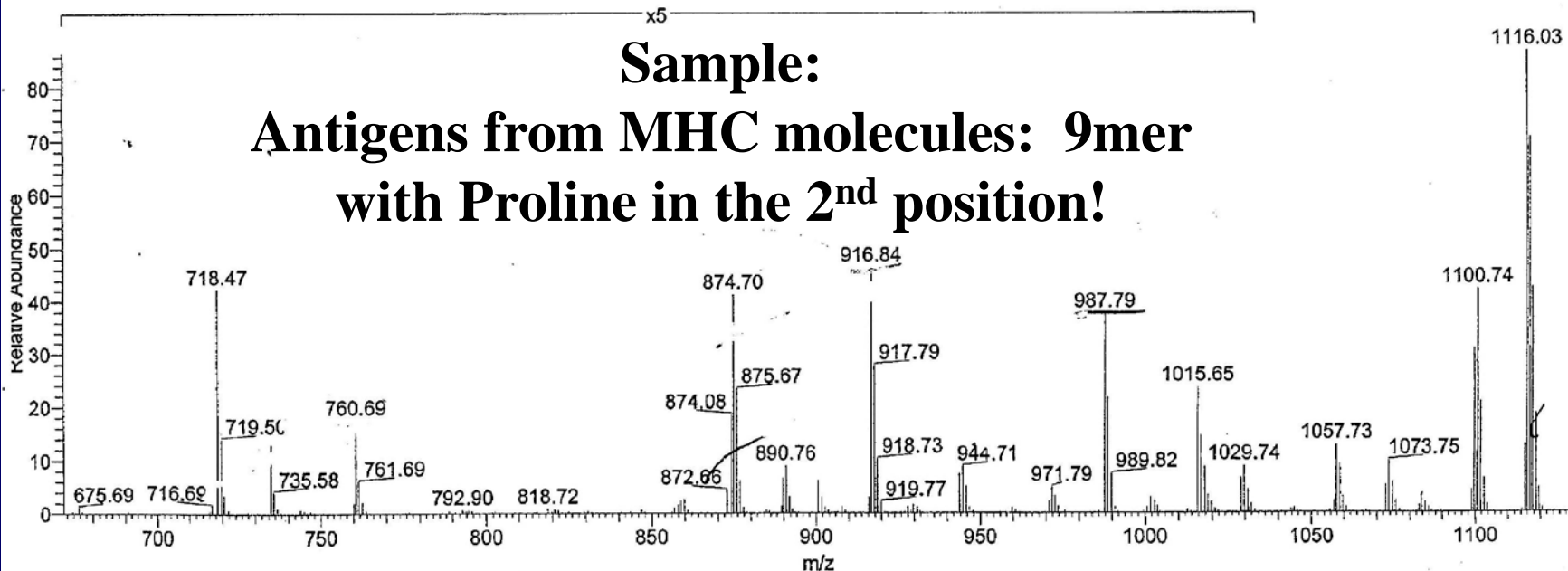
I:\Melanie\ETD\JYB7\mjsJYB7IMAC_030304   03/03/2004 12:25:02 PM   JYB7 IMAC   *methylesters*
td test
mjsJYB7IMAC_030304 #2144 RT: 26.21 AV: 1 NL: 2.12E6
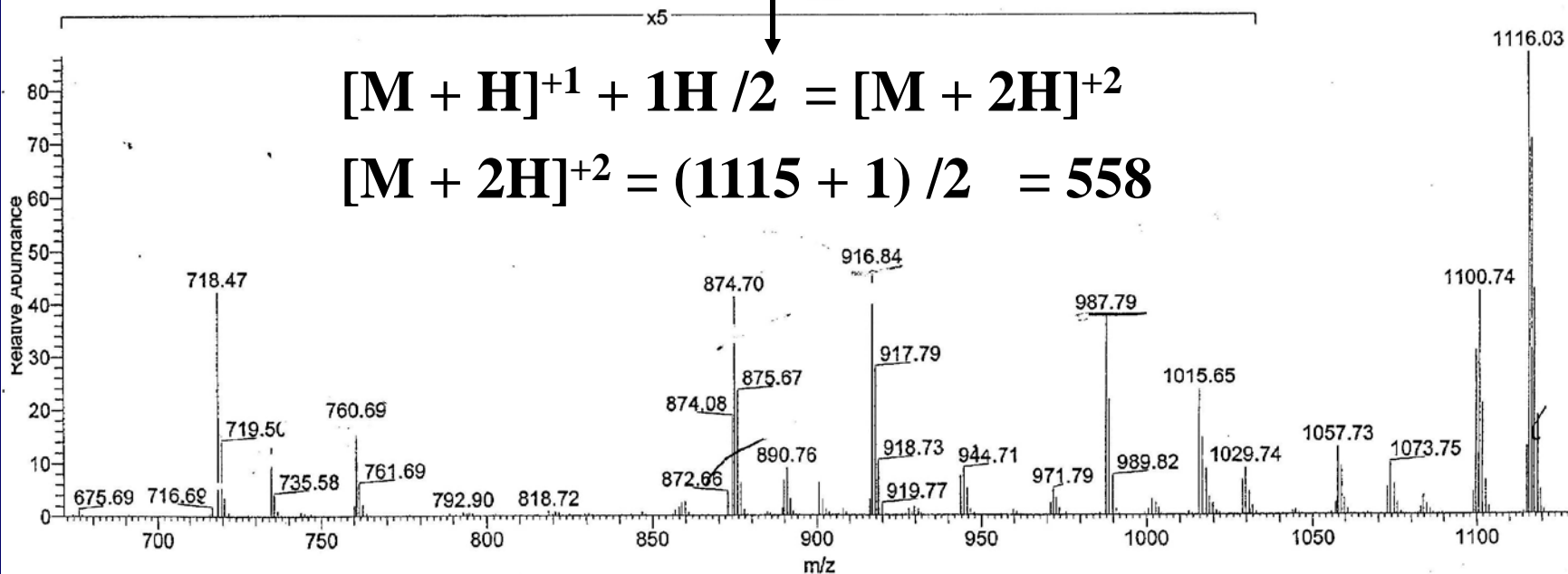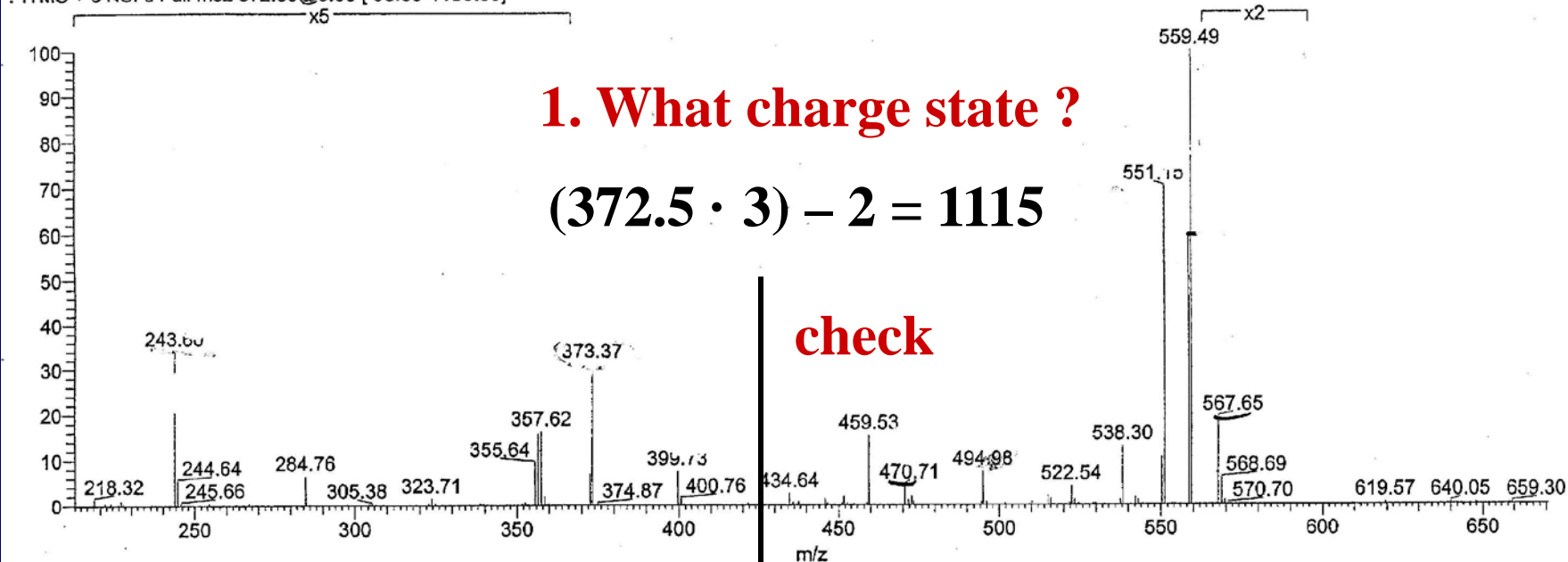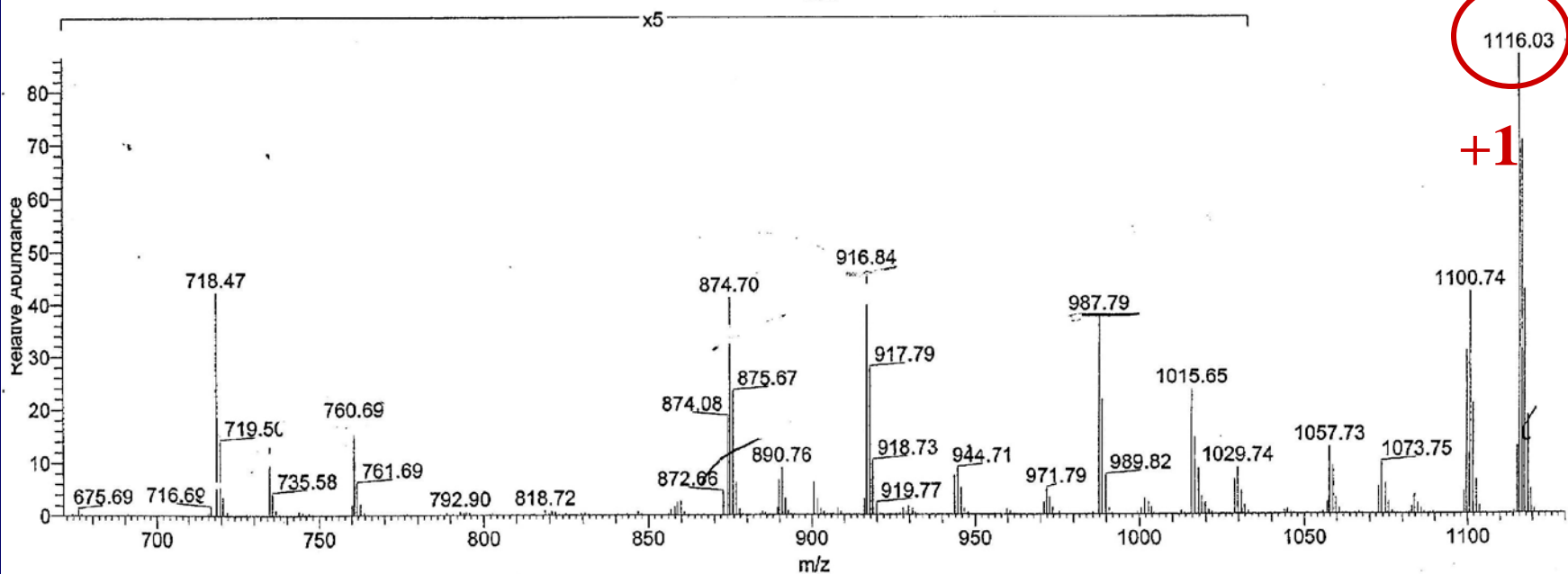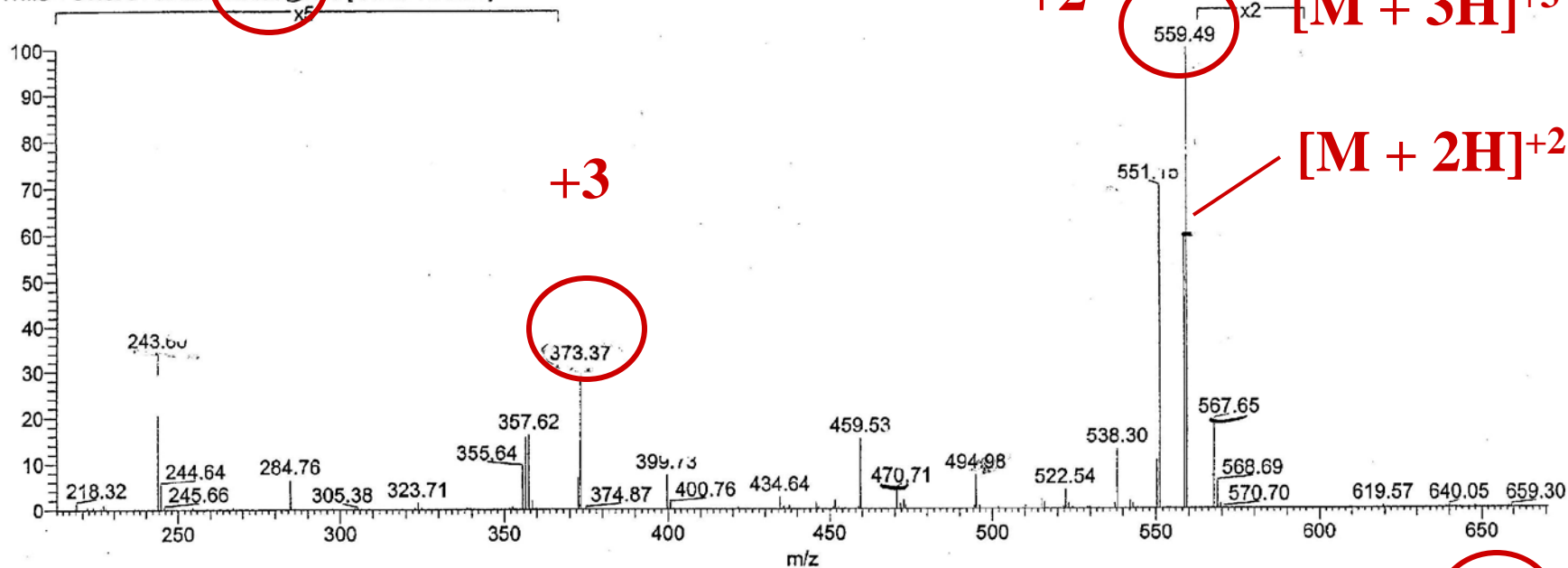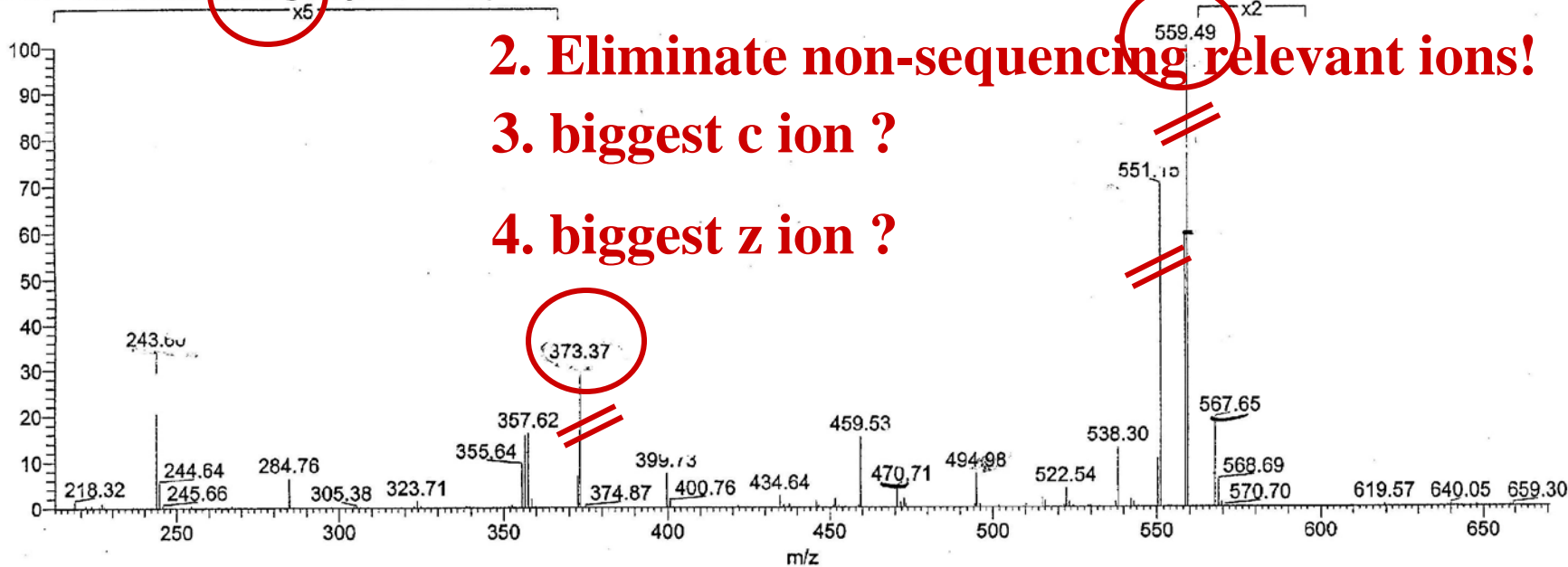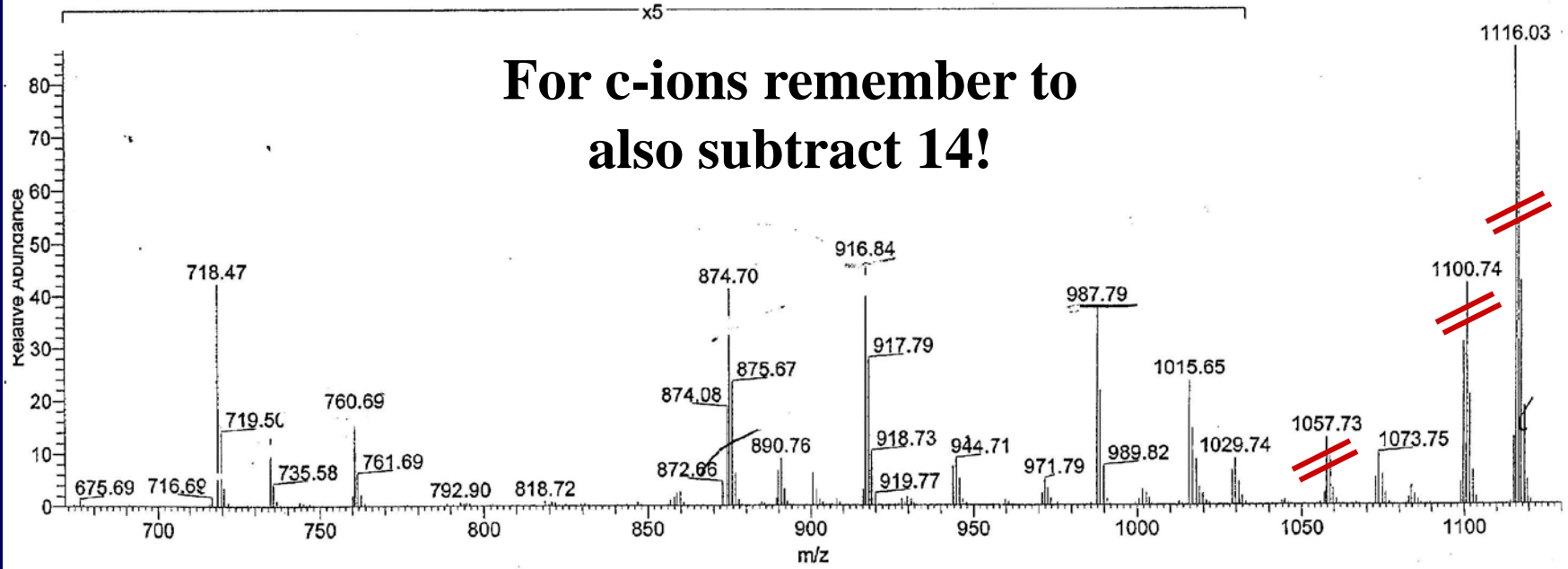: ITMS + c NSI d Full ms2 372.85@0.00 [90.00-1130.00]

2. Eliminate non-sequencing relevant ions!

3. biggest c ion ?

4. biggest z ion ?

**For c-ions remember to
also subtract 14!**

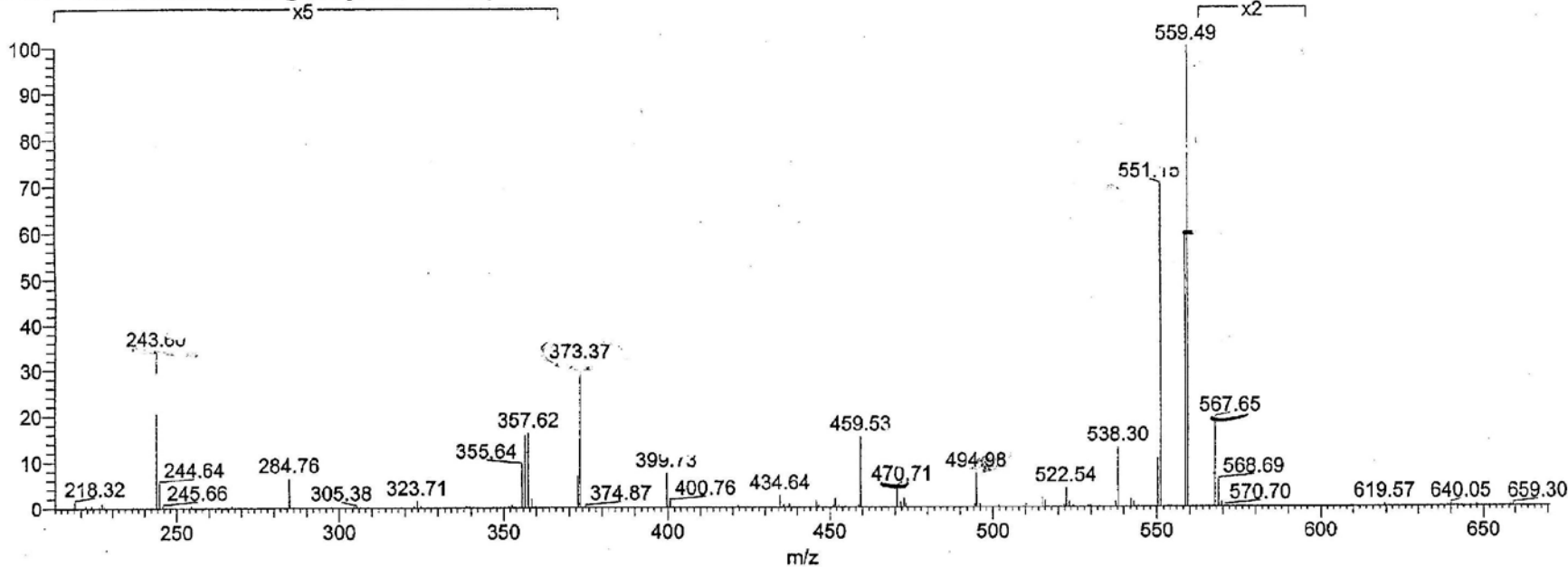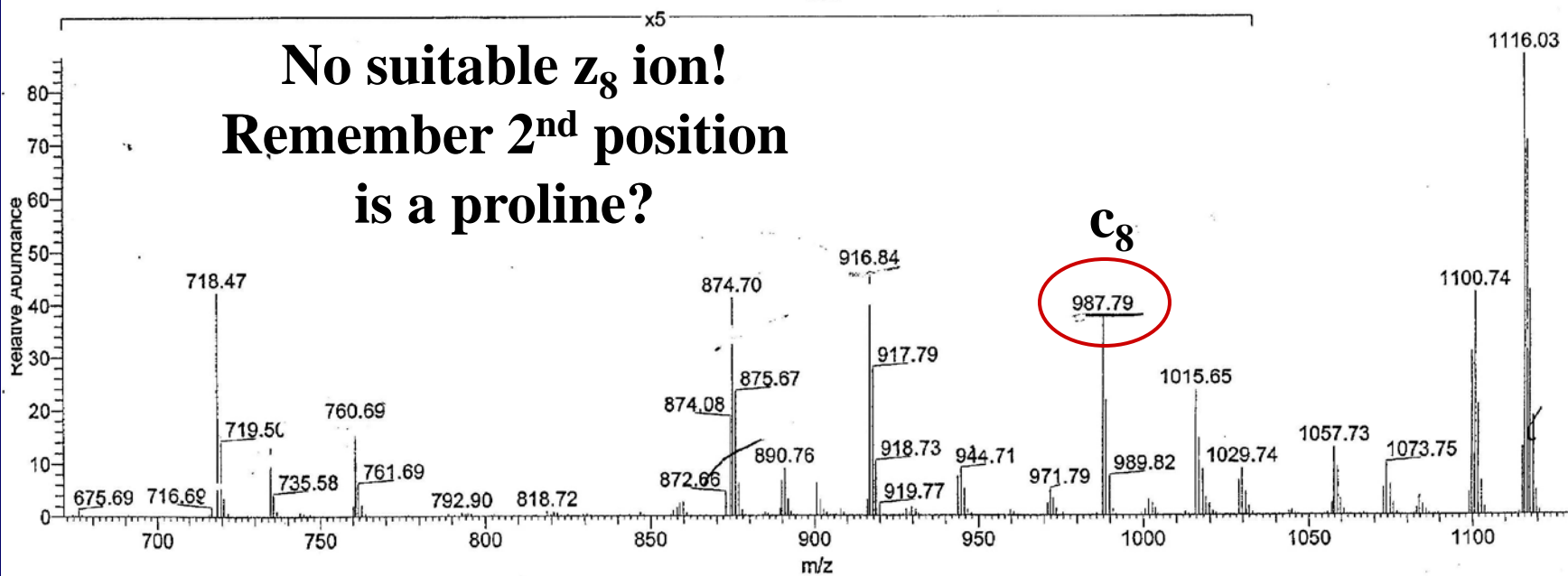mjsJYB7IMAC_030304 #2144 RT: 26.21 AV: 1 NL: 2.12E6
: ITMS + c NSI d Full ms2 372.85@0.00 [90.00-1130.00]
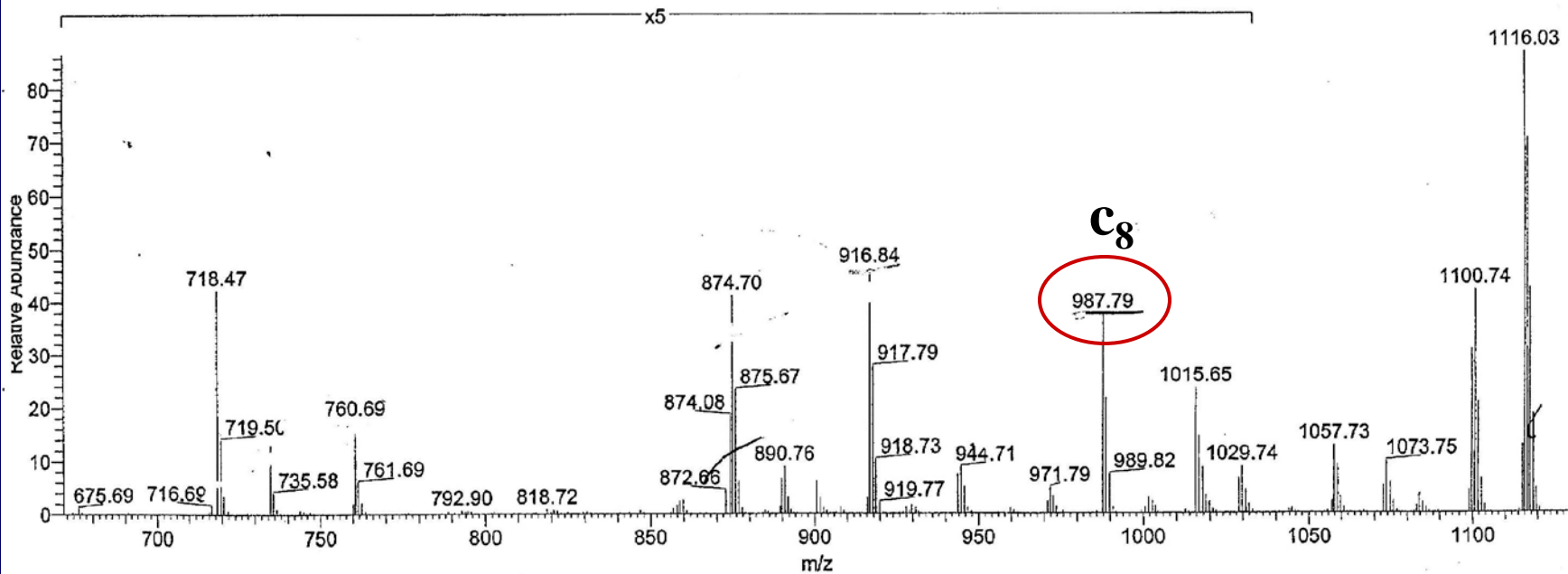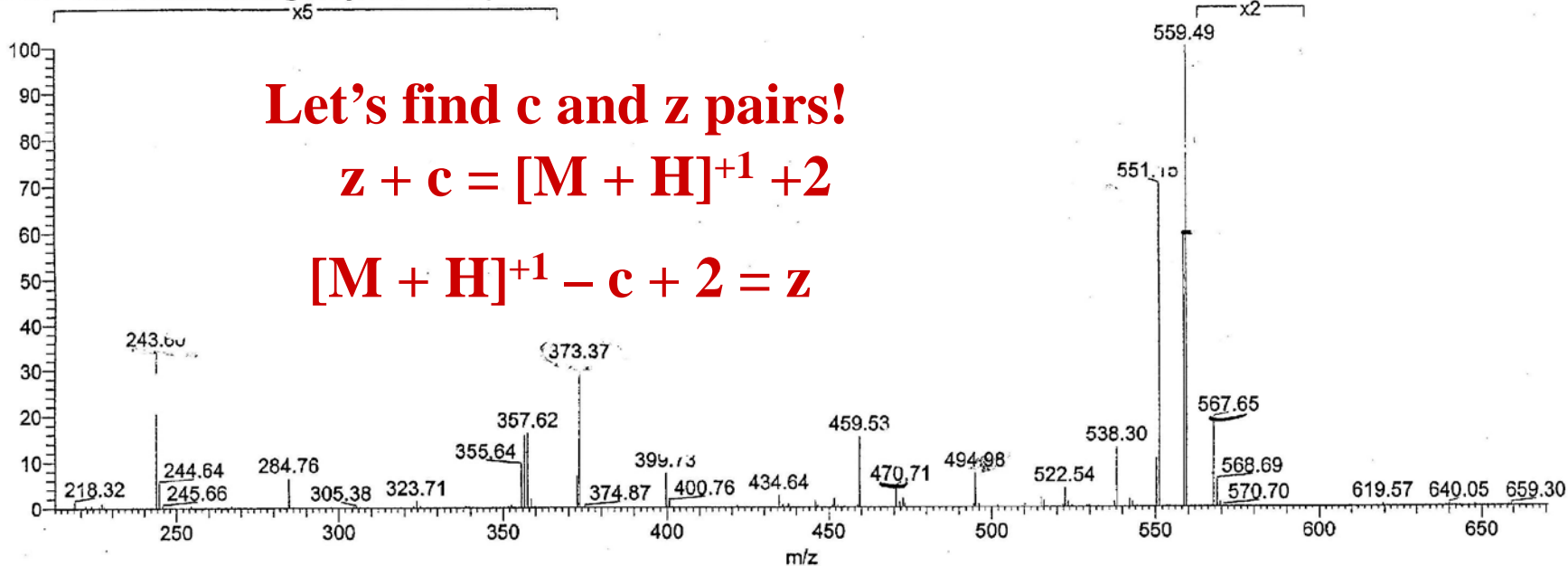

No suitable z_8 ion!
Remember 2nd position
is a proline?

c_8

Let's find c and z pairs!
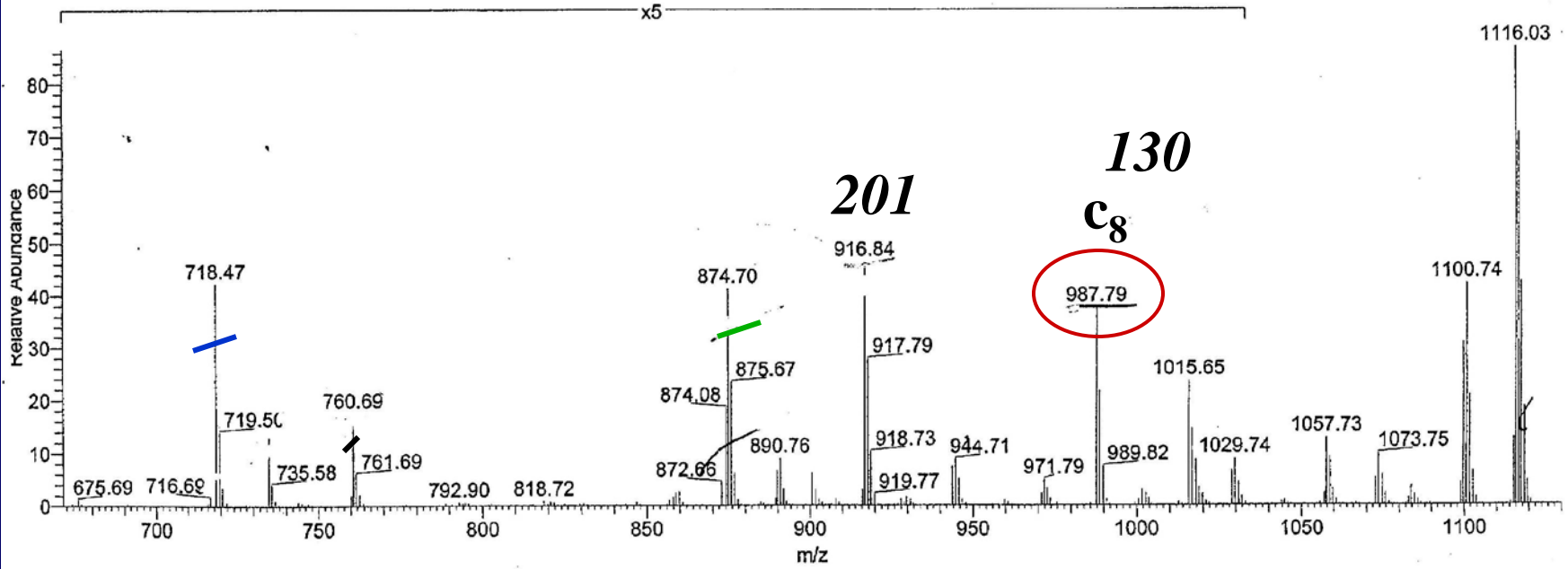
$$z + c = [M + H]^{+1} + 2$$

$$[M + H]^{+1} - c + 2 = z$$

mjsJYB7IMAC_030304 #2144 RT: 26.21 AV: 1 NL: 2.12E6
: ITMS + c NSI d Full ms2 372.85@0.00 [90.00-1130.00]

x5

100
90
80
70
60
50
40
30
20
10
0

243.60
244.64
245.66
218.32
284.76
305.38
323.71
355.64
357.62
373.37
374.87
399.73
400.76
434.64
459.53
470.71
494.98
522.54
538.30
551.15
559.49
567.65
568.69
570.70
619.57
640.05
659.30

x2

m/z
250   300   350   400   450   500   550   600   650

146   243   399   760   916   987 1115
Q/K   P   R   _   _   _   R   A   I/L
1115 971   874   718   357   201   130

x5

Relative Abundance

80
70
60
50
40
30
20
10
0

675.69
716.69
718.47
719.50
735.58
760.69
761.69
792.90
818.72
872.66
874.08
874.70
875.67
890.76
916.84
917.79
918.73
919.77
944.71
971.79
987.79
989.82
1015.65
1029.74
1057.73
1073.75
1100.74
1116.03

m/z
700   750   800   850   900   950   1000   1050   1100
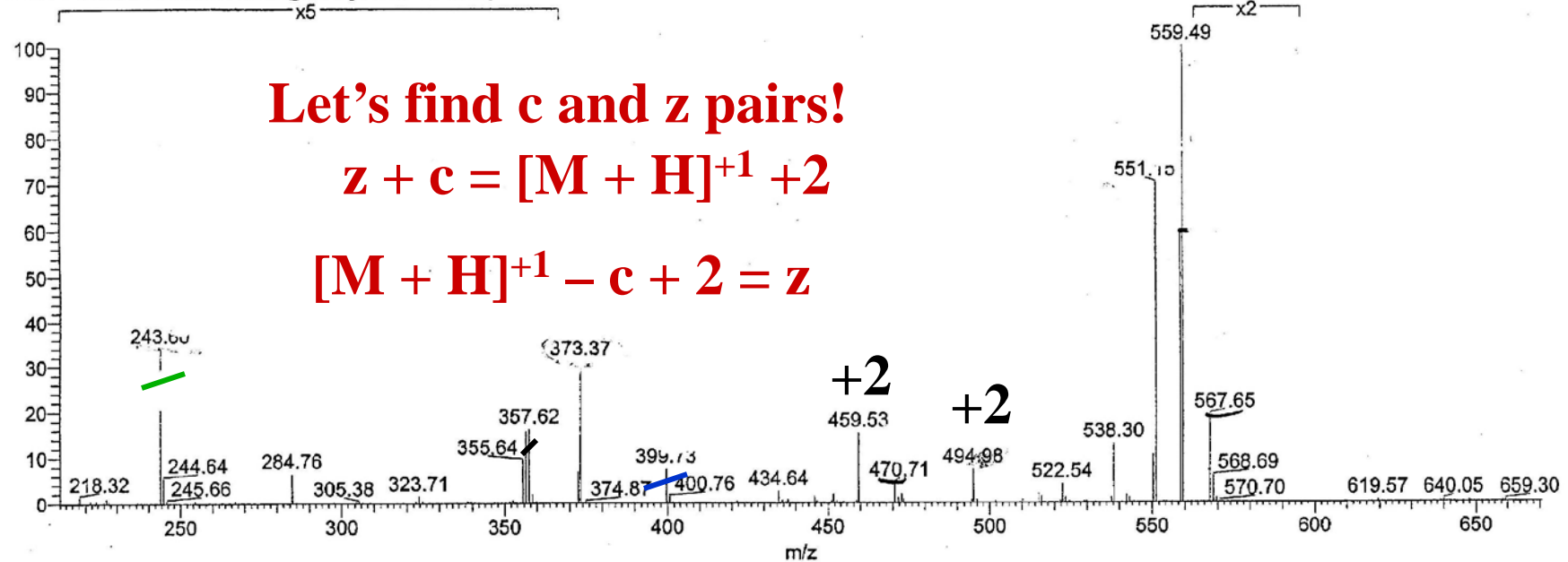
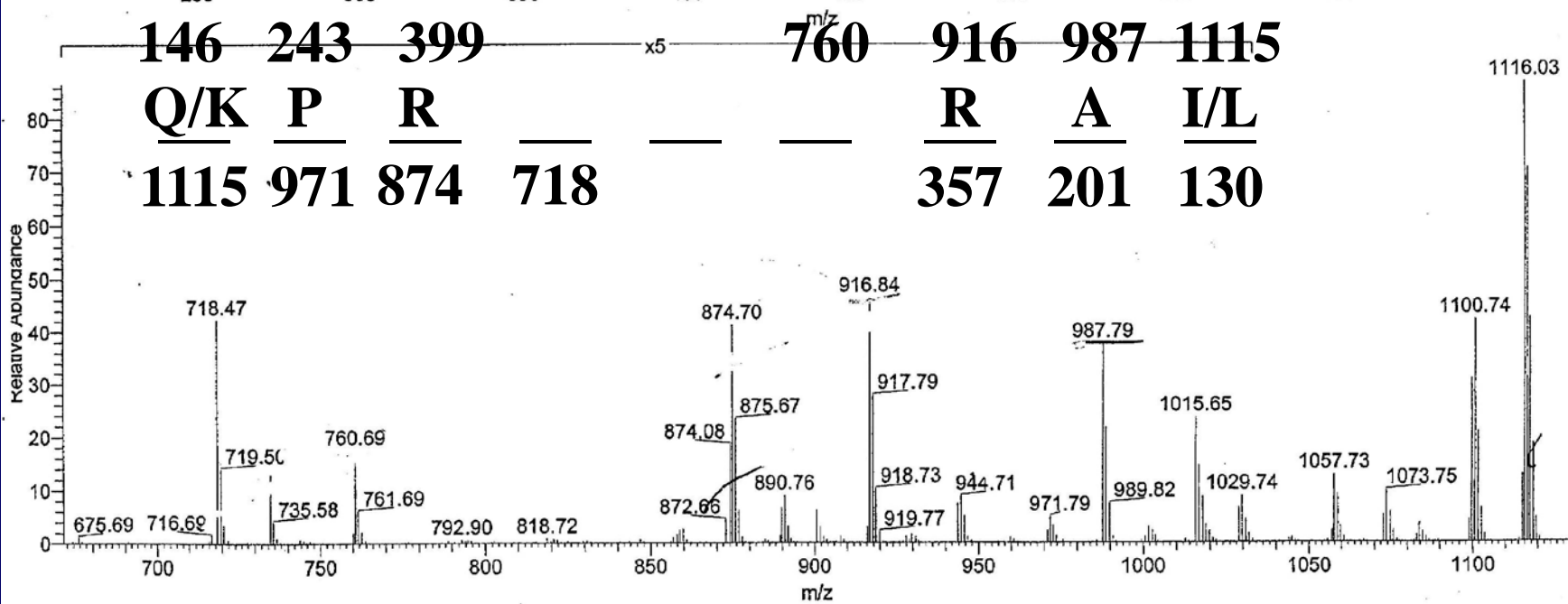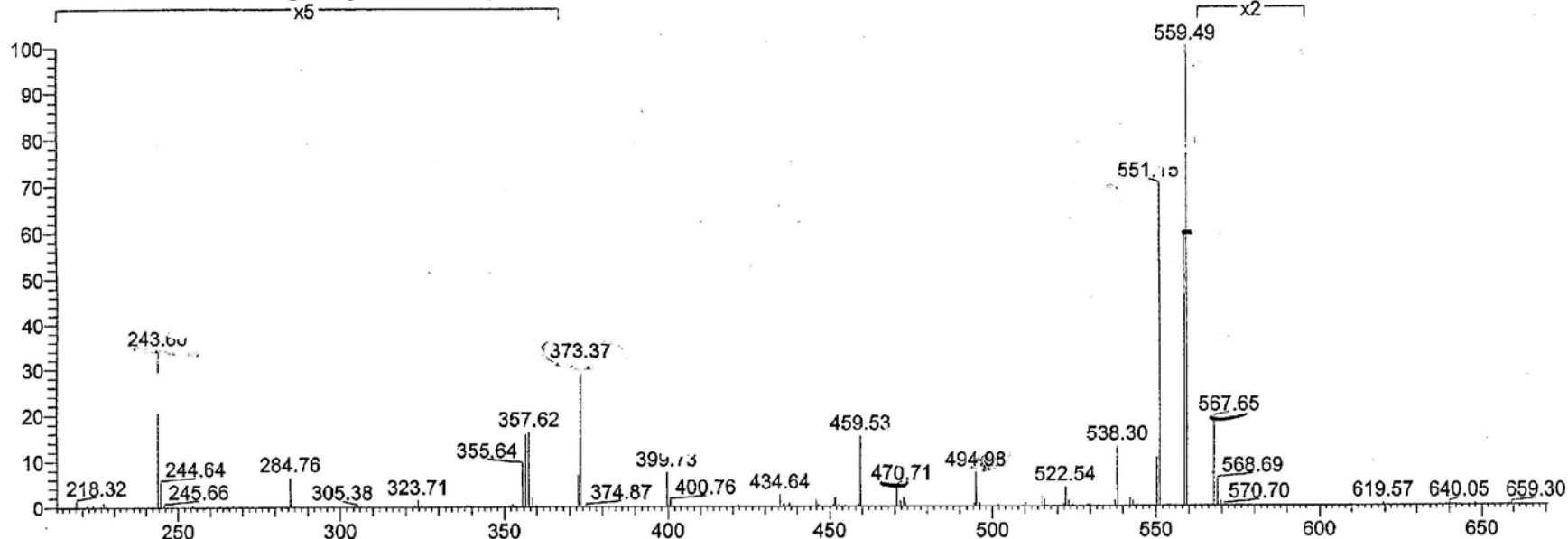mjsJYB7IMAC_030304 #2144 RT: 26.21 AV: 1 NL: 2.12E6
: ITMS + c NSI d Full ms2 372.85@0.00 [90.00-1130.00]

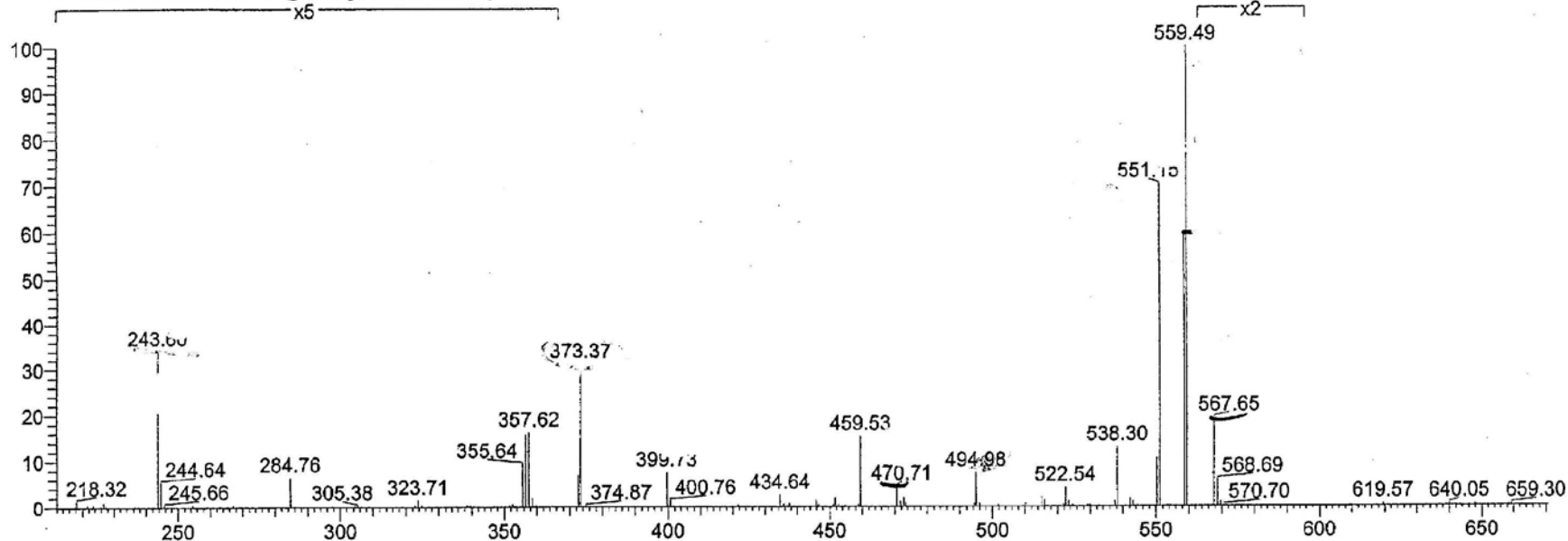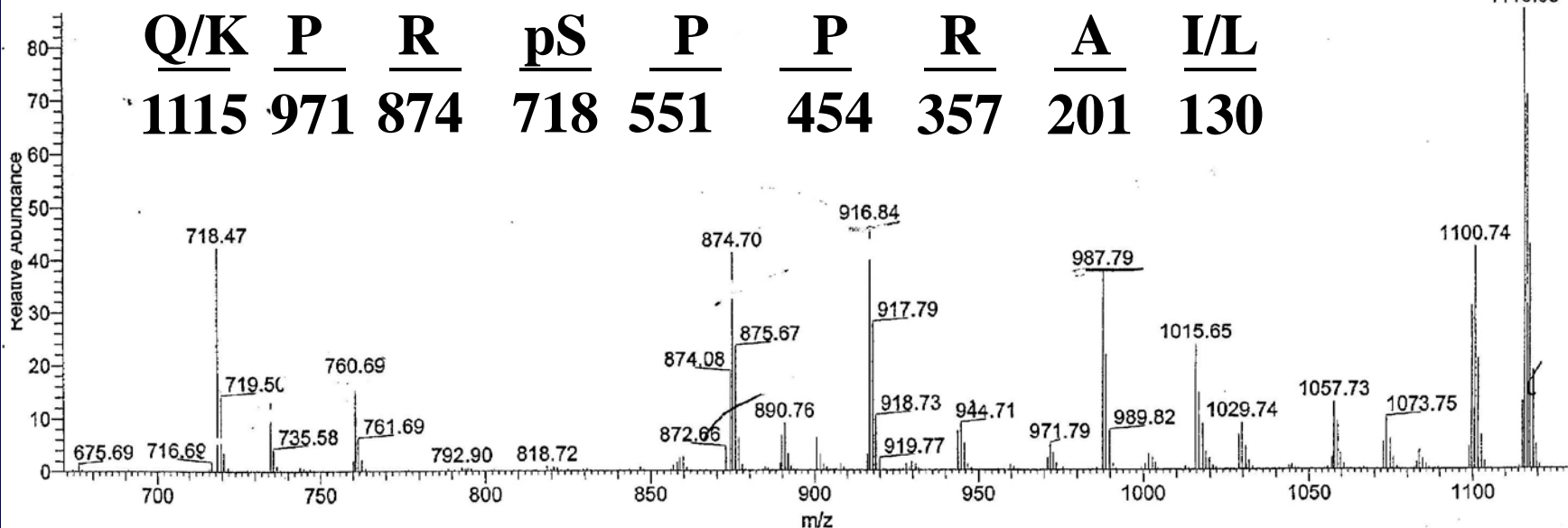| 146	243	399	566	663	760	916	987	1115
Q/K	P	R	pS	P	P	R	A	I/L
1115	971	874	718	551	454	357	201	130

Proteomics Informatics –
Protein identification III:
de novo sequencing (Week 6)