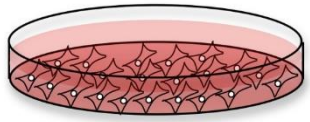
Fundamentals
Data Analysis Workshop

Ben Collins
Queen's University Belfast | School of Biological Sciences
BIO8206 – Introduction to Biological Mass Spectrometry

Experimental workflow for today's data

HeLa cells



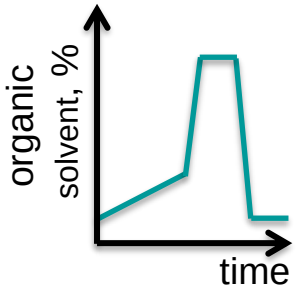
Protein extraction

Proteins

Protein digestion

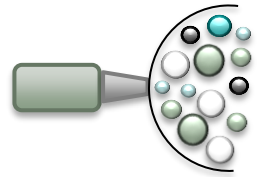
Peptides

Peptide separation



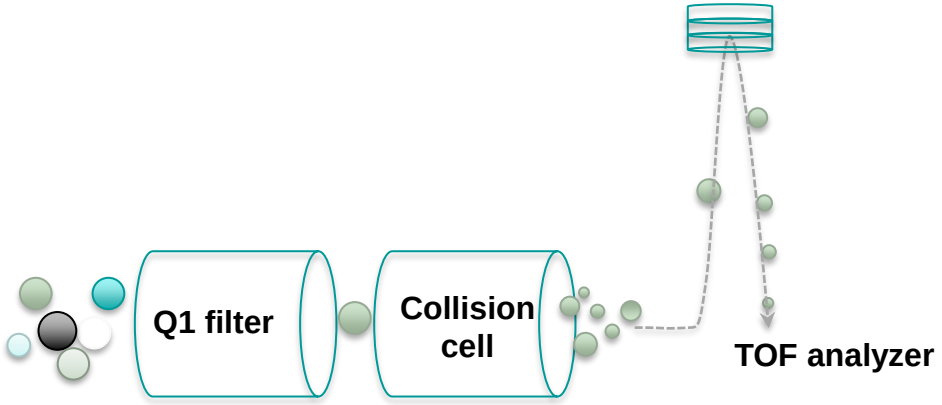
HPLC

Ionization



ESI

Mass spectrometry



mass spectrometer

MS1 – mass (m/z) of peptides

Isolation + fragmentation

MS2 – mass (m/z) of fragments

2 topics for workshop – also related to assignment!

- Calculating mass from m/z
  - We measure m/z but we want to know mass!
  - Review charge states and isotope distributions
- Interpreting fragmentation (or MS2) spectra
  - Using peptide as example – but generalizable to other analytes
  - Manual annotation of amino acid sequence of peptide from MS2 spectrum

Recap on isotopic distributions and charge stages

- 2 separate issues to review:
  1. Elemental isotopic abundances and MS
  2. Using isotope distributions to determine charge state

1. In nature for given elements we find different isotopic abundances of various elements because or different numbers of neutrons
   - We are primarily concerned with carbon
   - 98.9 % of the carbon in nature is ^{12}C (6 protons + 6 neutrons in nucleus)
   - 1.1 % is ^{13}C (6 protons + 7 neutrons in nucleus)
   - 1 part per trillion is ^{14}C (6 protons + 8 neutrons in nucleus)
     - Unstable with half life of 5,730 years → radio carbon dating

Some elemental isotope abundances

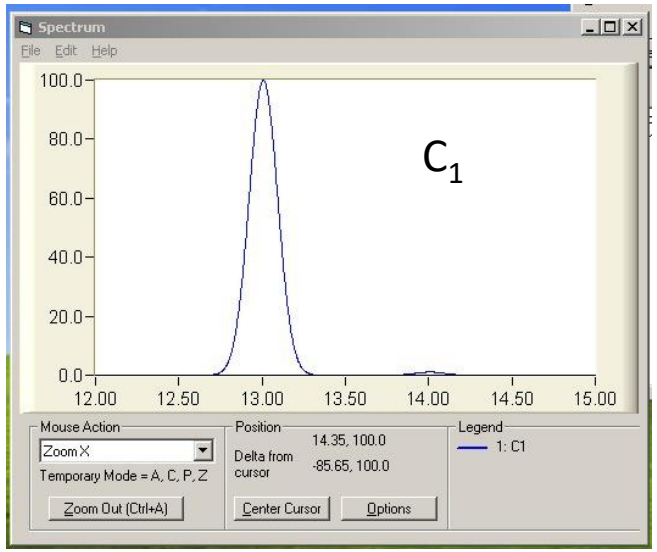
Atom	Mass	Rel. Abund.
Hydrogen	1.008 2.001	99.985 0.015
Carbon	12.000 13.003	98.90 1.10
Nitrogen	14.003 15.000	99.63 0.37
Oxygen	15.995 17.999	99.76 0.20
Sulfur	31.972 33.968	95.02 4.21

Isotope Distributions

Mw of carbon is 12 Daltons

Question:
What is the natural abundance of the carbon 13 isotope?

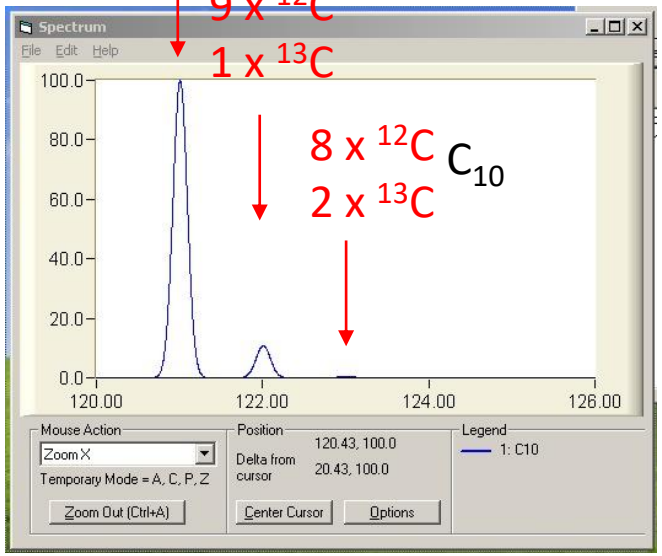
Answer: ~1.1 %

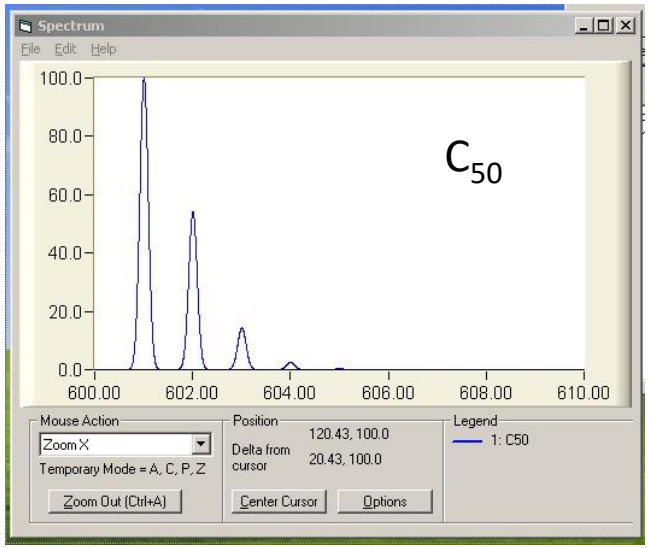


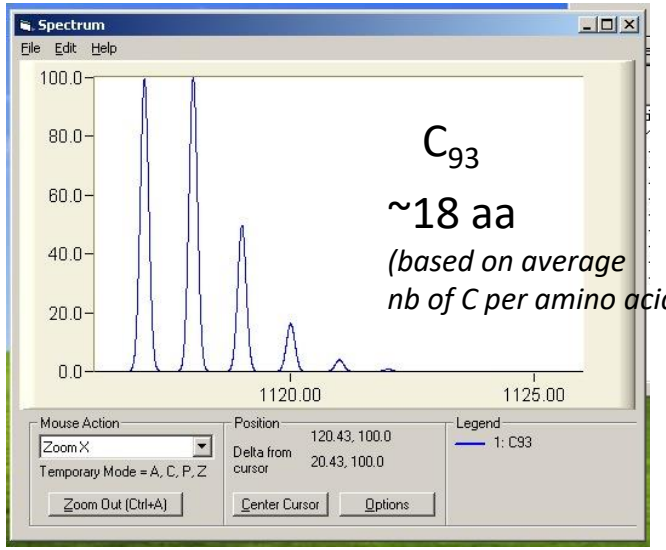
10 x ^{12}C

9 x ^{12}C
1 x ^{13}C

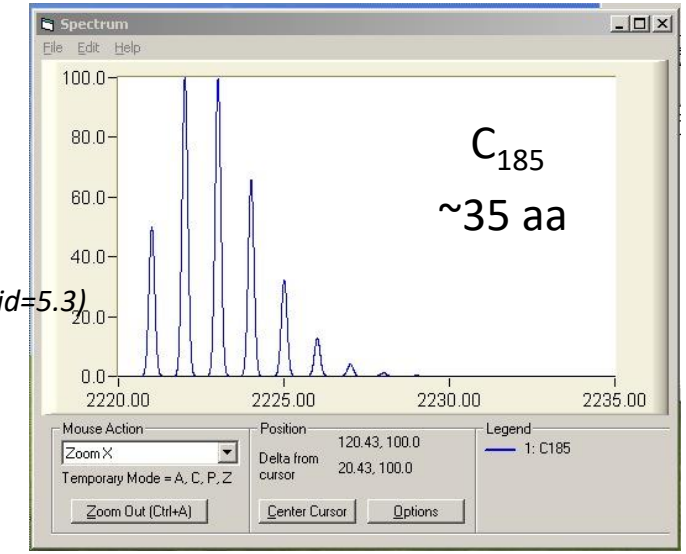
8 x ^{12}C
2 x ^{13}C

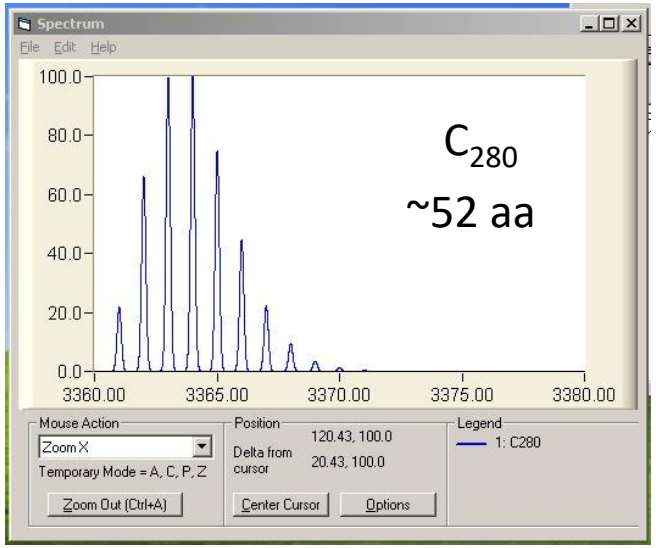






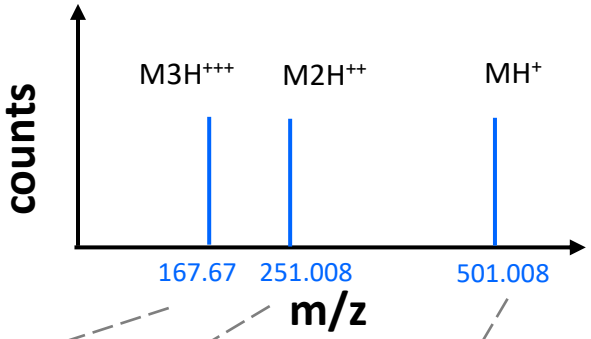
(based on average nb of C per amino acid=5.3)



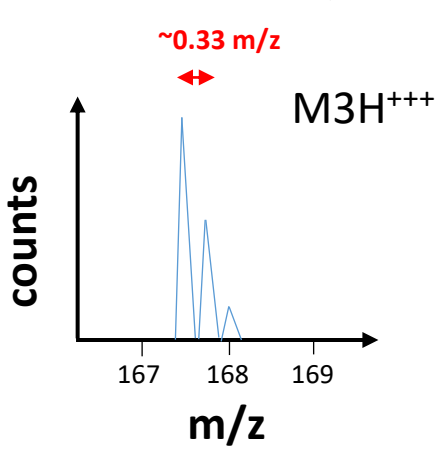


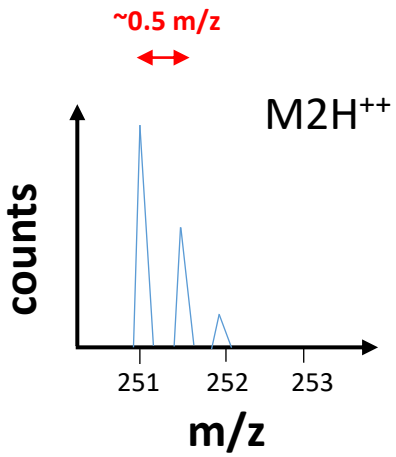
Do we actually measure mass? No, mass to charge ratio

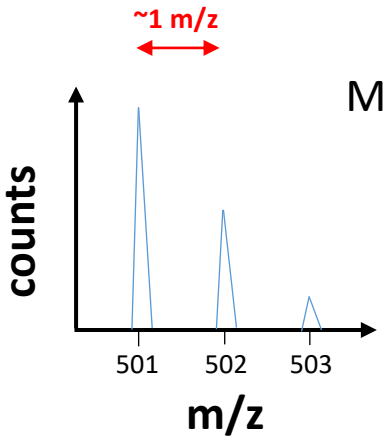
	mass/charge	mass (inc. protons)	neutral mass
• MH^+	501.008	501.008	500
• $M2H^{++}$	251.008	502.016	500
• $M3H^{+++}$	167.674	503.024	500



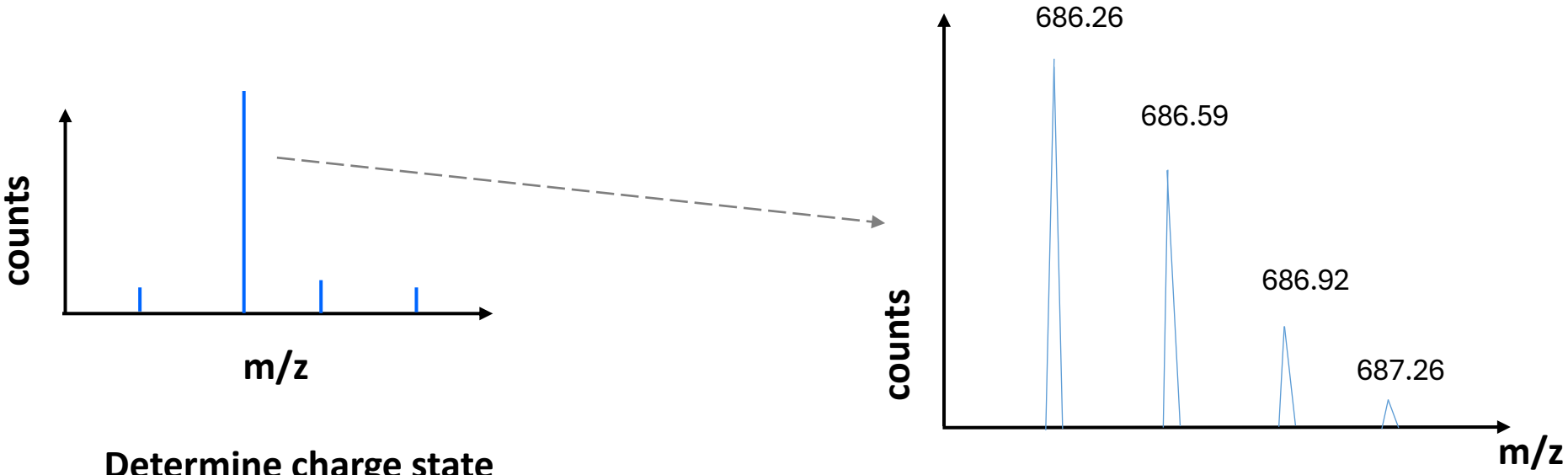
To calculate the neutral mass from the measured m/z we need to know the charge state
We use the existence of the isotopic distribution as a trick to figure out the charge state!







So, if we want to calculate neutral mass from m/z…
Worked example:



Determine charge state
spacing between istopalogues = ~0.33
=> charge state = 1 / ~0.33 = 3+

Calculate mass (with protons)
Monoisotopic mass x charge state
686.26 x 3 = 2,058.78

Calculate neutral mass
Subtract mass of protons
2,058.78 – (3 x 1.008) = **2,055.756 Da**

Note: assume positive mode electrospray so we expect ionizing protons

2 topics for workshop – also related to assignment!

- Calculating mass from m/z
  - We measure m/z but we want to know mass!
  - Review charge states and isotope distribubtions

- Interpreting fragmentation (or MS2) spectra
  - Using peptide as example – but generalizable to other analytes
  - Manual annotation of amino acid sequence of peptide from MS2 spectrum

Fragmenting a Peptide

A-P-N-D-F-N-L-K

(MH+ 919)

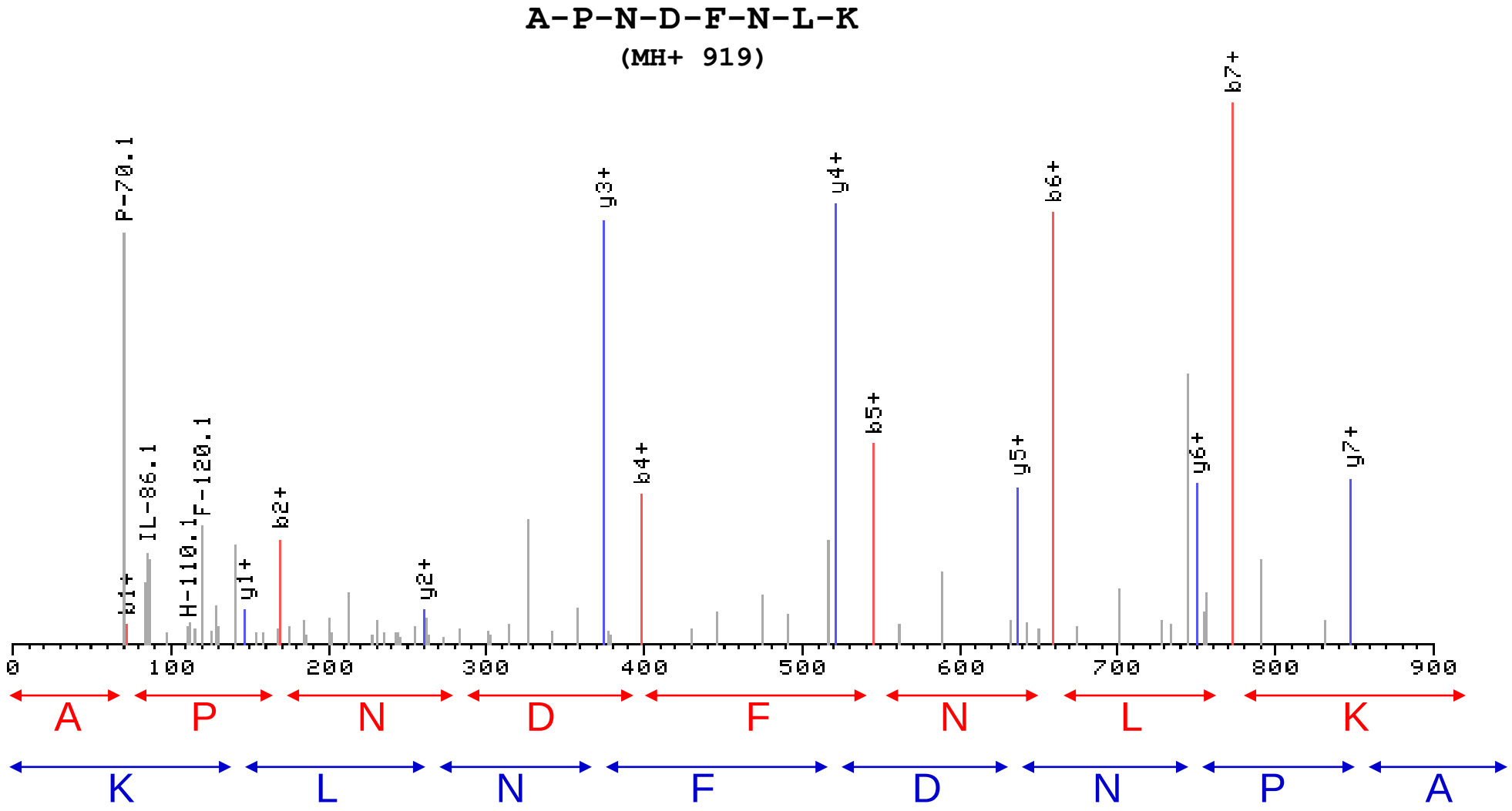
B-ions		Y-ions
72.0	A ..... P-N-D-F-N-L-K	847.4
169.1	A-P ..... N-D-F-N-L-K	750.4
283.1	A-P-N ..... D-F-N-L-K	636.3
398.2	A-P-N-D ..... F-N-L-K	521.3
545.2	A-P-N-D-F ..... N-L-K	374.2
659.3	A-P-N-D-F-N ..... L-K	260.2
772.4	A-P-N-D-F-N-L ..... K	147.1

$b\text{-ions} = \Sigma AA + H^+$

$y\text{-ions} = \Sigma AA + H_2O + H^+$

monoisotopic masses

Sequence & MS2 Spectrum – b and y ions



Amino Acid Masses

Code	TLC	Formula	Monoisotopic	Average
A	Ala	C_3H_5ON	71.03711	71.0788
R	Arg	$C_6H_{12}ON_4$	156.10111	156.1875
N	Asn	$C_4H_6O_2N_2$	114.04293	114.1038
D	Asp	$C_4H_5O_3N$	115.02694	115.0886
C	Cys	C_3H_5ONS	103.00919	103.1388
E	Glu	$C_5H_7O_3N$	129.04259	129.1155
Q	Gln	$C_5H_8O_2N_2$	128.05858	128.1307
G	Gly	C_2H_3ON	57.02146	57.0519
H	His	$C_6H_7ON_3$	137.05891	137.1411
I	Ile	$C_6H_{11}ON$	113.08406	113.1594
L	Leu	$C_6H_{11}ON$	113.08406	113.1594
K	Lys	$C_6H_{12}ON_2$	128.09496	128.1741
M	Met	C_5H_9ONS	131.04049	131.1926
F	Phe	C_9H_9ON	147.06841	147.1766
P	Pro	C_5H_7ON	97.05276	97.1167
S	Ser	$C_3H_5O_2N$	87.03203	87.0782
T	Thr	$C_4H_7O_2N$	101.04768	101.1051
W	Trp	$C_{11}H_{10}ON_2$	186.07931	186.2132
Y	Tyr	$C_9H_9O_2N$	163.06333	163.176
V	Val	C_5H_9ON	99.06841	99.1326

Procedure

1. Convert the precursor m/z to single charged (e.g. $M2H^{++}$ → MH^{+})
   - (If everything is singly charged it makes the math easier….)

2. Locate the peak with the biggest m/z value (i.e right most peak) and subtract that value from the precursor m/z

3. Check this value against the amino acid table (monoisotopic)
   - if match then this should be one of the terminal AA residues (go to step 4)
   - If no match then move onto the next biggest m/z value and try again
   - Note: error of ~0.01-0.02 m/z is still a match, but the error should not be larger than this
   - Hint: m/z gap should be ~57-186 to match an AA residue

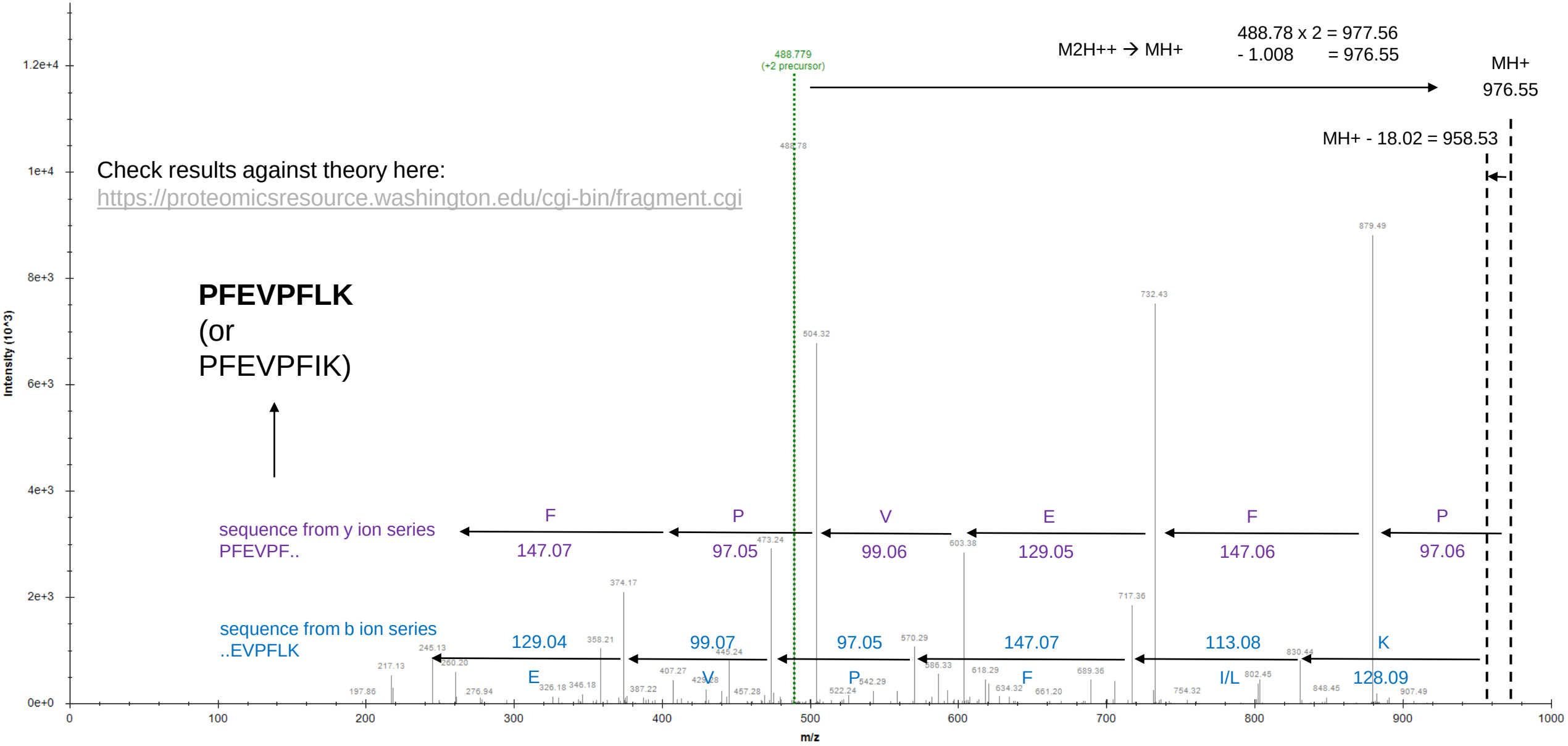
4. Find the next biggest m/z peak and subtract it from the last peak you assigned, and then keep repeating this procedure until you get to lowest m/z peaks and can't find more AA residues
   - This will be the y ion series series

5. Start again from the precursor MH^{+} calculated in step 1 and subtract 18.02 (H_2O). Then look for another gap that fits an AA residue mass and work your way down again looking for gaps that correspond to AA residues as in step 4
   - This will be the b ion series
   - Hint: in peptides produced by trypsin cleavage the first AA residue you find should be K or R

6. Reverse the order of the b series AA residues and align the 2 sequences!

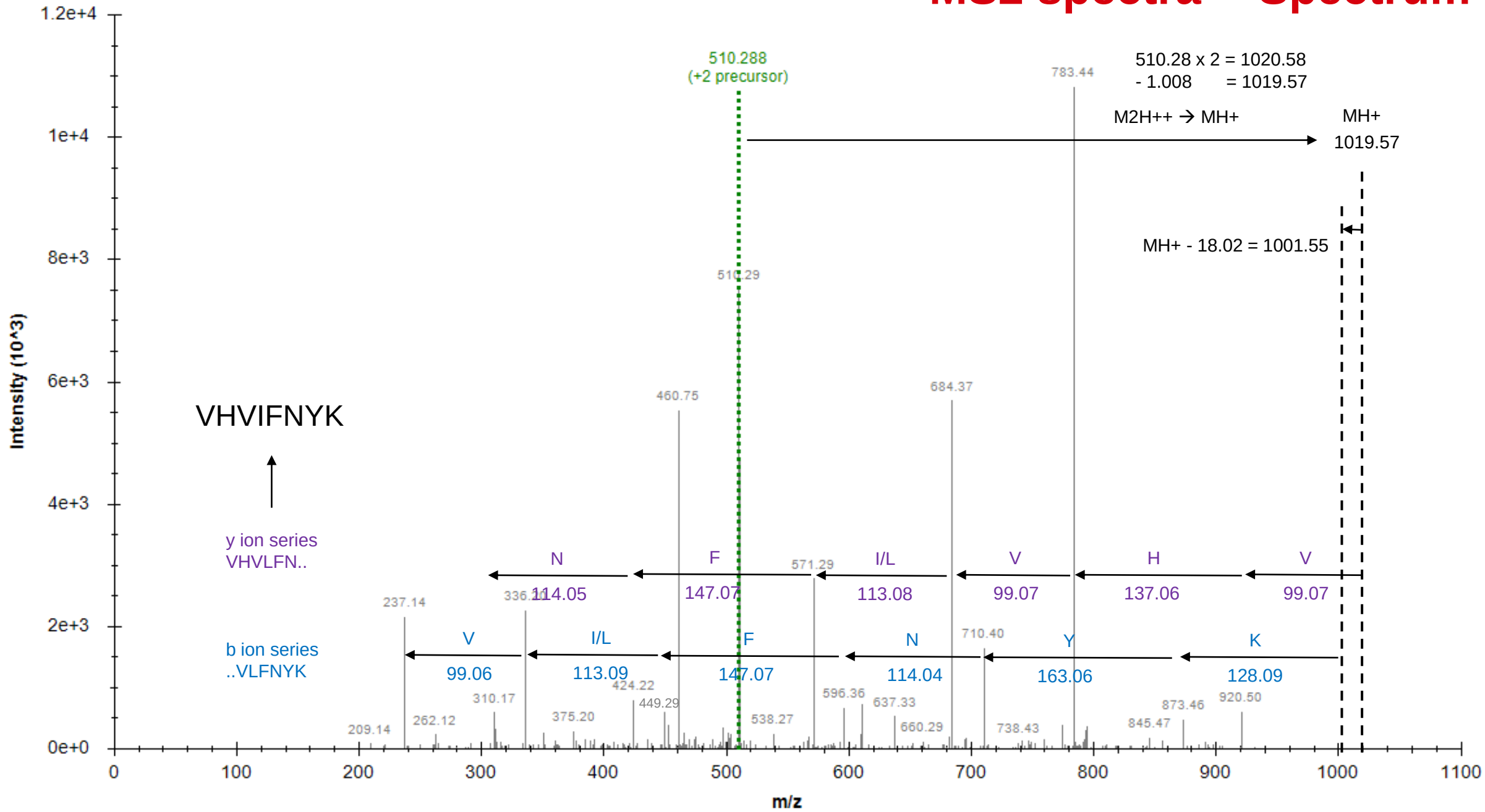
For a deeper tutorial see http://ionsource.com/tutorial/DeNovo/introduction.htm

MS2 spectra -- Spectrum 1



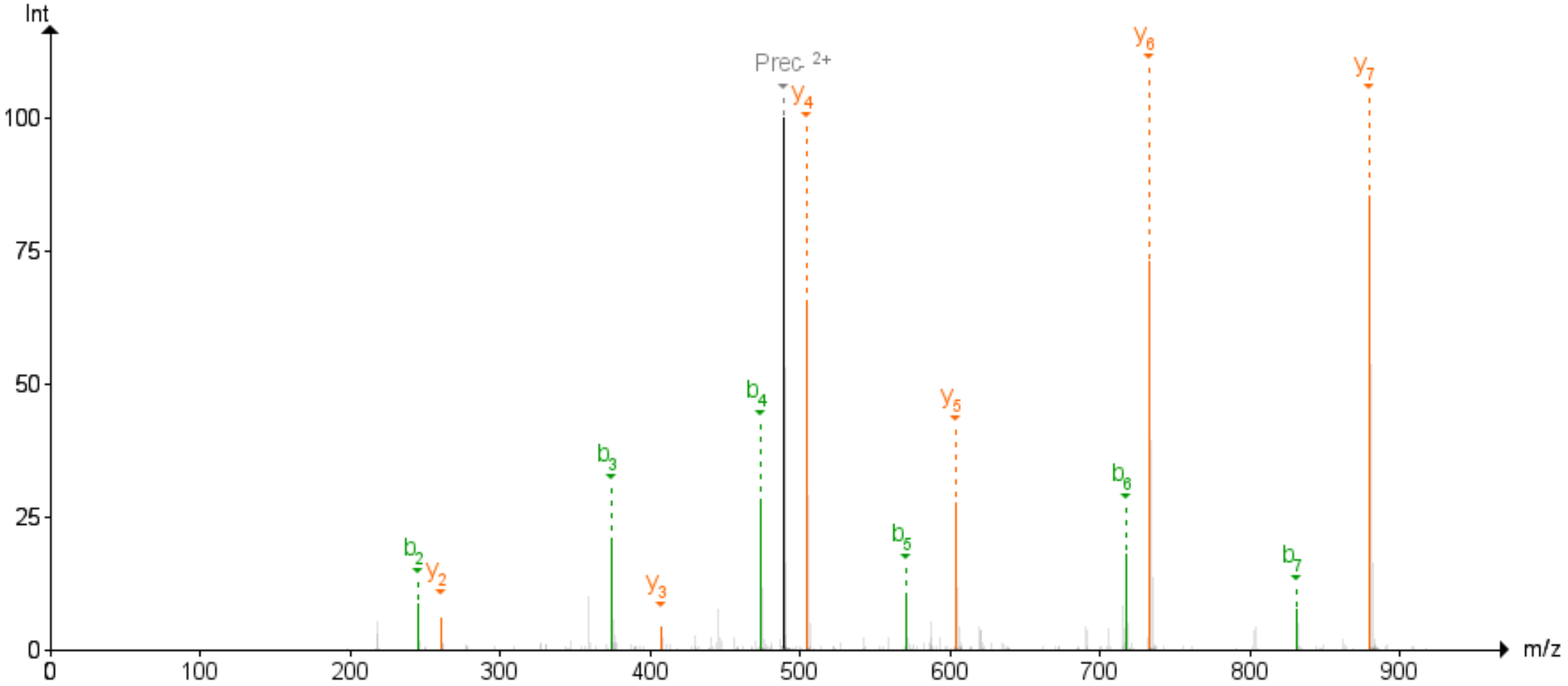
A210622_bcc_hela_200ng_dda_1538.60512.60512.2_PFEVPFLK

MS2 spectra -- Spectrum 2



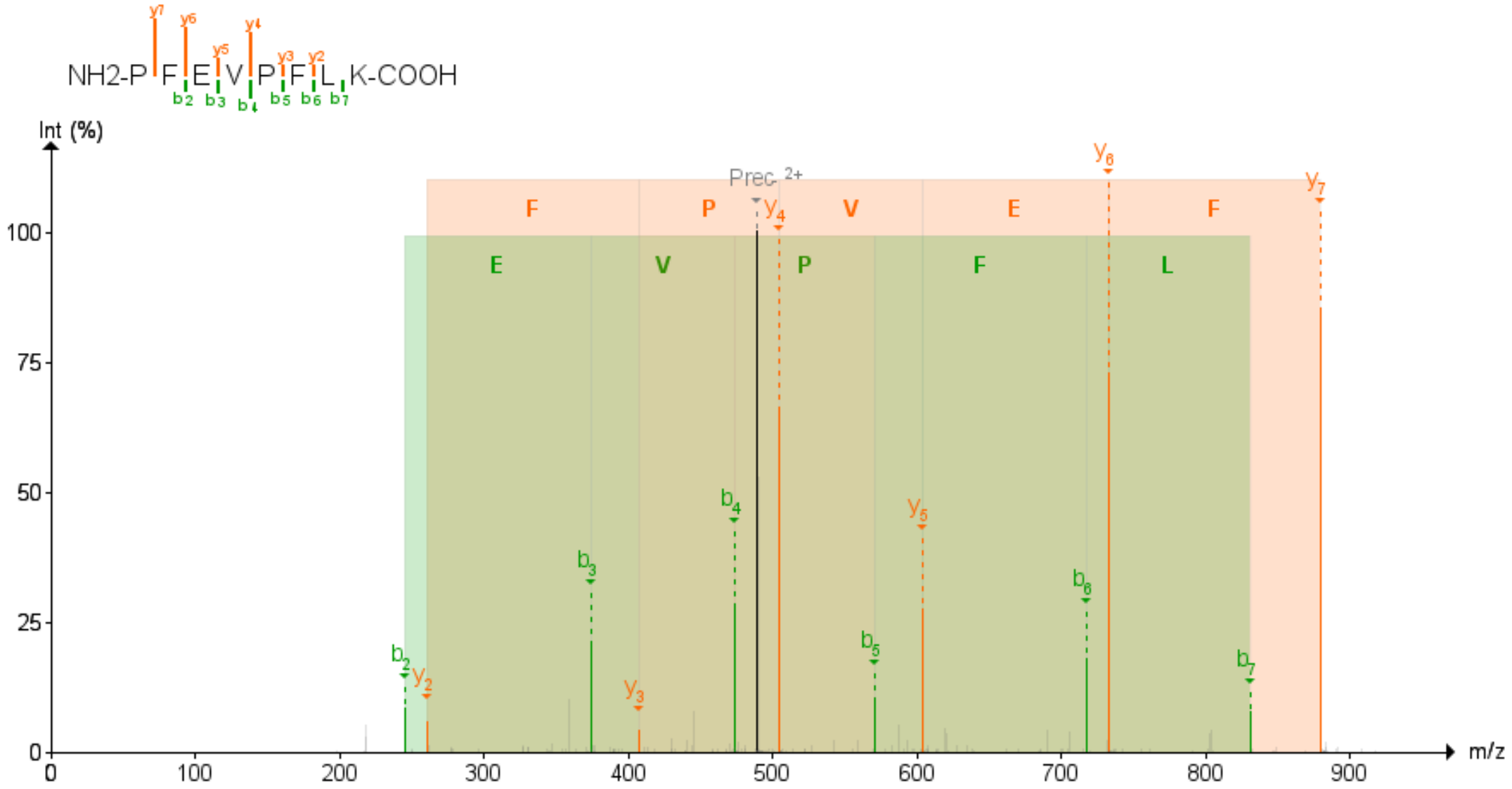
A210622_bcc_hela_200ng_dda_1538.21410.21410.2_VHVIFNYK

Computer annotation – spectrum 1



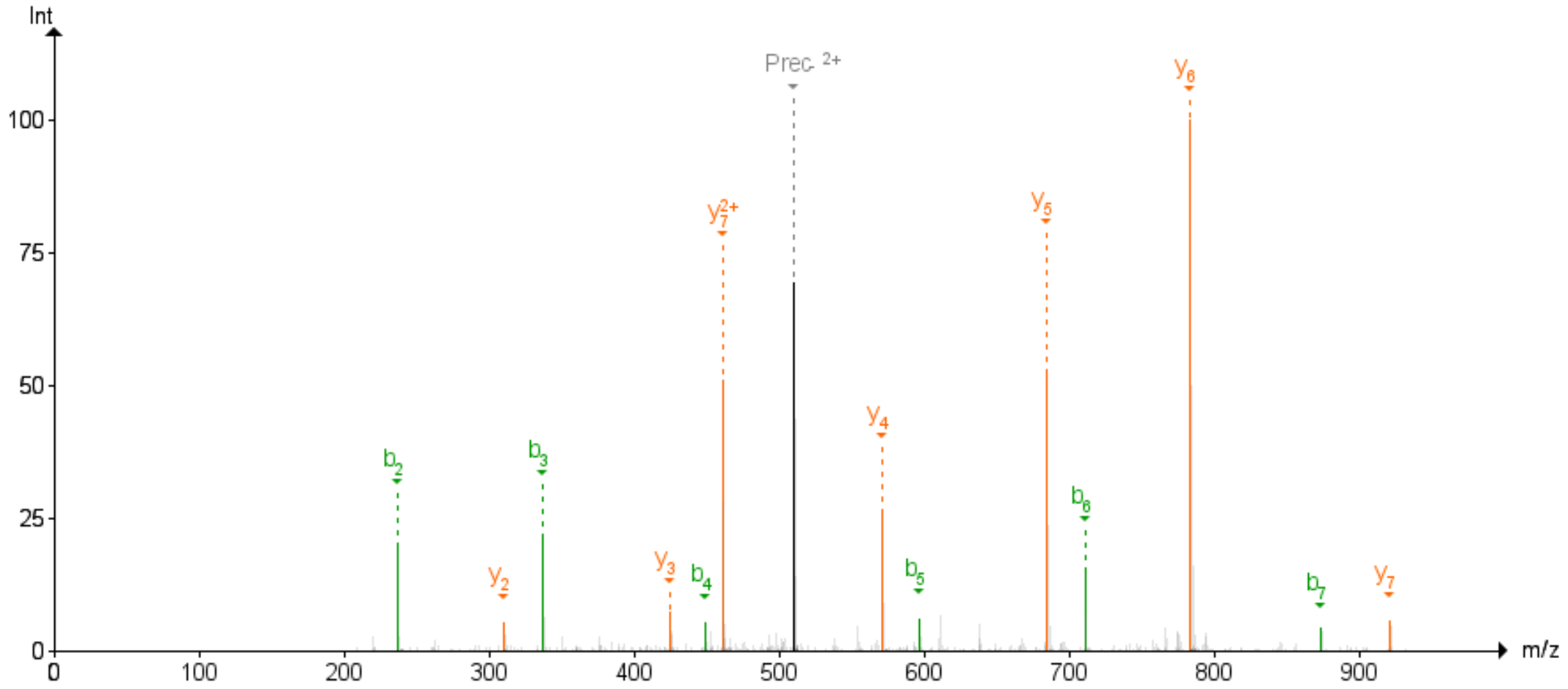
A210622_bcc_hela_200ng_dda_1538.60512.60512.2_PFEVPFLK

Computer annotation – spectrum 1



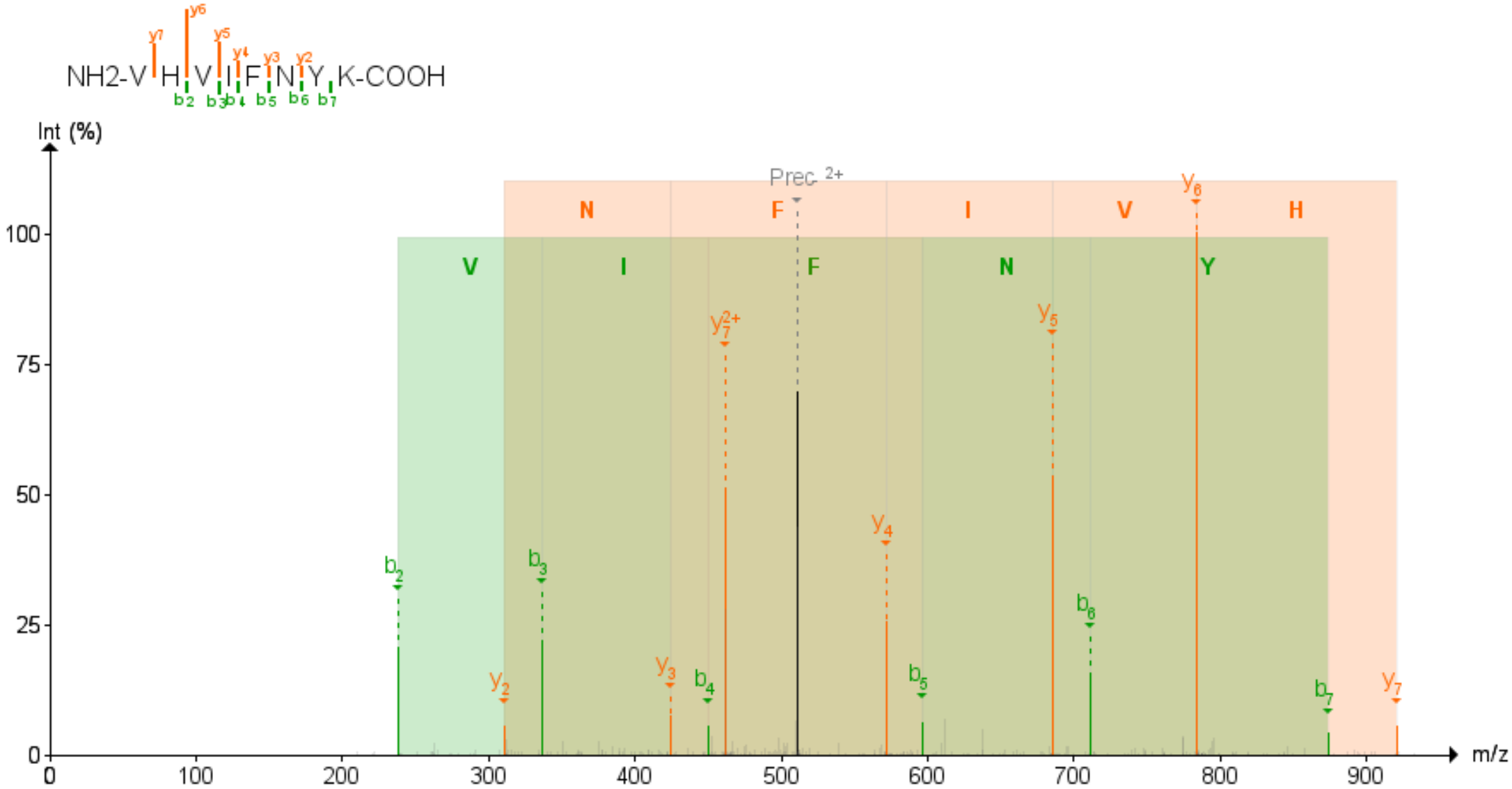
A210622_bcc_hela_200ng_dda_1538.60512.60512.2_PFEVPFLK

Computer annotation – spectrum 2



A210622_bcc_hela_200ng_dda_1538.21410.21410.2_VHVIFNYK

Computer annotation – spectrum 2



A210622_bcc_hela_200ng_dda_1538.21410.21410.2_VHVIFNYK