CW#1 (due April 23rd @17:00, 2024)

Instructions:

i. This assignment consists of 50% of the CW marks of BIO312;

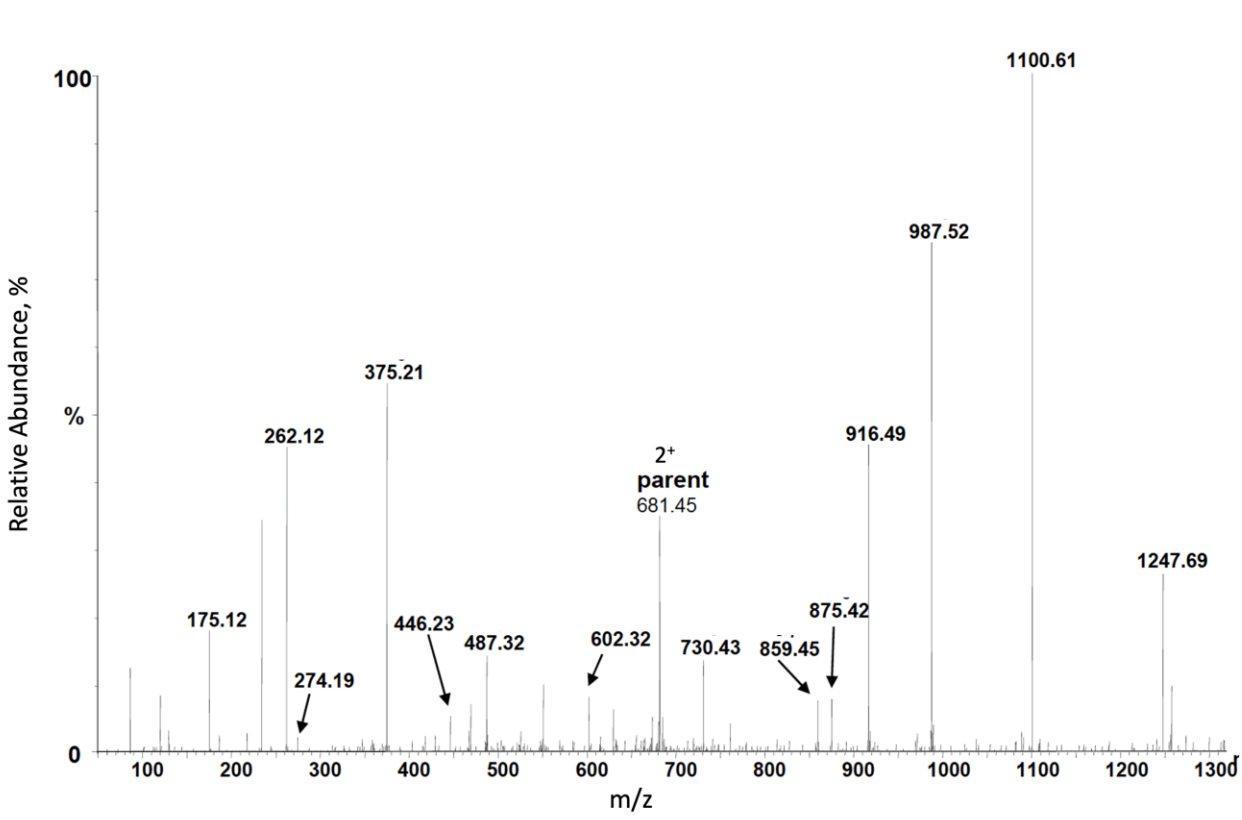
ii. You can refer to any materials, but please complete the coursework by yourself.

iii. Please submit your work in Word or PDF format;

iv. Marking: 50 marks for answering Q1-Q3.

---

1. **[15 marks]** Determine the full sequence of the **12-residue** peptide. Can you assign the sequence with this data alone? If not, briefly explain why. Please list the key steps of the de novo sequencing.

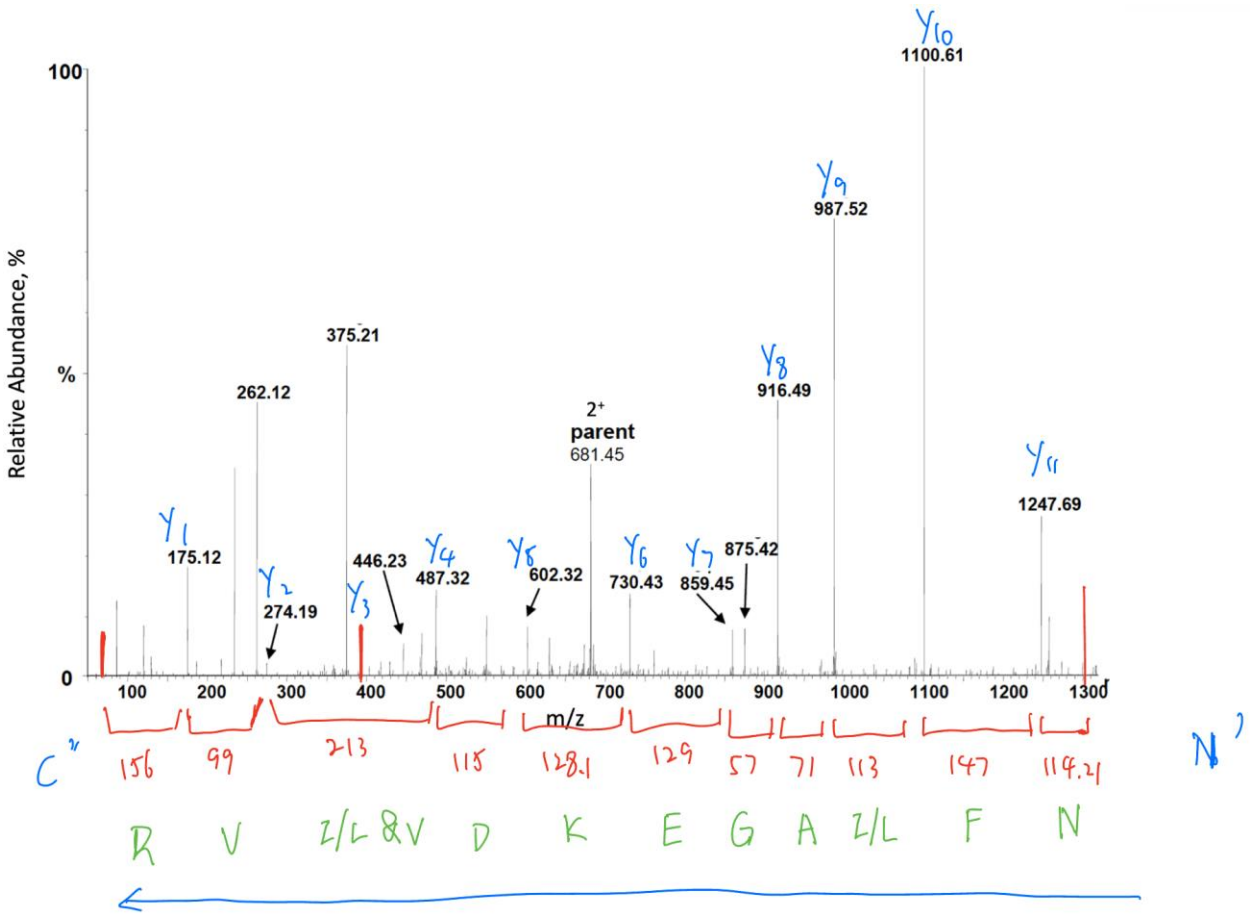


(1) (681.45*2)-2-1247.69+1=114.21, which is the mass of N. So, 1247.69 m/z represents an y_{n-1} ion and N is the N terminus residue.

(2) Δm/z = 1247.69–1100.61=147.08, which is the mass of F.

(3) The iterative operation process is shown in the figure below:

(4) 175.12-19=156.12 m/z is probably the y_1 ion with 1 residue. Because it is an y ion, so the mass of the residue = y_1 – 19 = 156.12, which is the R.



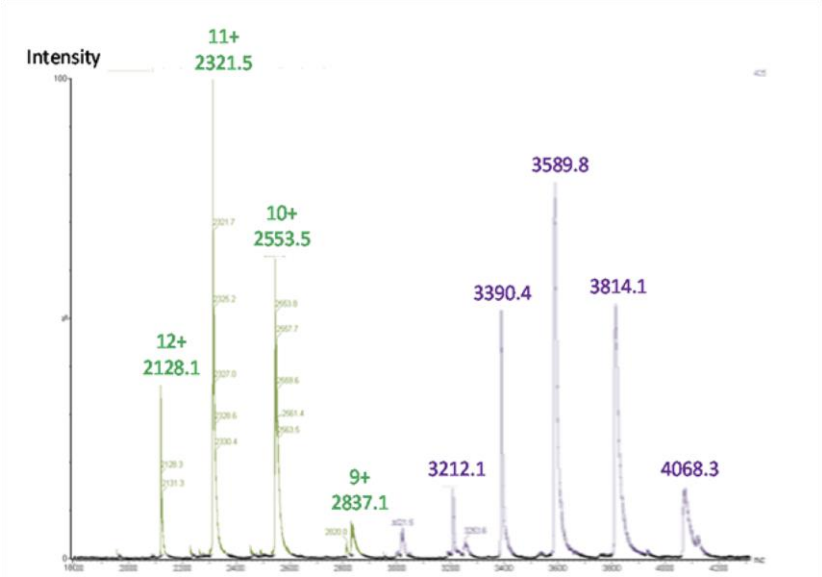
From the results of the difference between the peaks calculated in the mass spectrum above, we can obtain the sequence of this peptide chain: N F I/L A G E K D V I/L V R.

De Novo Sequencing Steps:

- Calculate the mass of peptides.
- Find the b-ions and y-ions based on the MS spectrum peaks.
- Try to set y_1/b_1 residue or y_{n-1}/b_{n-1} before sequencing.
- Here are the key points:
  - Mass of b-ions = Σ (residue masses) + 1 (H^+)
  - Mass of y-ions = Σ (residue masses) + 19 (OH + H + H^+)
  - M - y_{n-1} ion + 1 = mass of 1st residue on N terminus
  - M- b_{n-1} ion - 17 = mass of 1st residue on C terminus
  - Mass of a-ions = mass of b-ions – 28 (CO)
  - Ser-, Thr-, Asp- and Glu-containing ions generate neutral molecular loss of water (-18).
  - Asn-, Gln-, Lys-, Arg-containing ions generate neutral molecular loss of ammonia (-17).
  - A complementary b-y ion pair can be observed in multiply charged ions spectra.
    - For this b-y ion pair, the sum of their subscripts is equal to the total number of amino acid residues in the unknown peptide.

- If there are some modification, we need to use formula: $M_{residue}+M_{modification}$

2. **[10 marks]** Figure below shows a mass spectrum of protein A (green) and a noncovalent protein complex AB (purple).
   (i) Which ionization source has been used? Briefly describe it.
   (ii) Calculate the molecular weight of the protein and the protein complex from the mass spectrum.



(1) The ionization source used in this mass spectrum is likely Electrospray Ionization (ESI). ESI is a technique commonly employed in mass spectrometry to generate ions from large biomolecules (such as proteins) in solution. The sample (protein A and protein complex AB) is first volatilized into gas-phase molecules and then apply a high-voltage electrospray to the sample creating an aerosol of charged droplets. After that, these charged droplets undergo desolvation resulting in the formation of gas-phase ions. Finally, the molecular ions ($M^{+\cdot}$) are then introduced into the mass analyzer for further analysis.

(2) W(A) = (m/z) × charge = 2128.1*12 + 2321.5*11 + 2553.5*10 + 2837.1*9 = 102142.6 Da
W(AB) = (m/z) × charge = 3390.4*12 + 3589.8*11 + 3814.1*10 + 4068.3*9 = 154928.3 Da

3. **[25 marks]** An investigator found two single nucleotide polymorphisms (SNPs) on a protein but not sure whether the corresponding mutated protein(s) is also expressed. He decided to isolate this protein using an immunoprecipitation approach and hope to pull down both the WT and the mutant(s) (if it's expressed at all) together. To confirm the mutant(s) was indeed expressed, he then submitted the sample to a proteomics core lab looking for help.

If you are a scientist working in the core lab and asked to take on this project, how would you proceed? Please propose necessary experimental steps to find out whether the mutant(s) was expressed.

Here is the amino acid sequence information for the WT protein and the two potential mutation sites (red, underlined and bold).

MILVDLEPGTMDSVRSGPFGQIFRPDNFVFGQSGAGNNWAKGHYTEGAELVDSV
LDVVRKEAESCDCLQGFQLTHSLGGGTGSGMGTLLISKIREEYPDRIMNTFSVVP
SPKVSDTVVEPYNATLSVHQLVENT**D**ETYCIDNEALYDICFRTLRLTTPTYGDLN
HLVSGTMECVTTCLRFPGQLNADLRKLAVNMVPFPRLHFFMPGFAPLTSRGSQQ
YRALTVPDLTQQVFDAKNMMAACDPRHGRYLTVAAVFRGRMSMKEVDEQML
NVQNKNSSYFVEWIPNNV**K**TAVCDIPPRGLKMAVTFIGNSTAIQELFKRISEQFTA
MFRRKAFLHWYTGEGMDEMEFTEAESNMNDLVSEYQQYQDATAEEEEDFGEE
AEEEA

(potential mutations suggested by the SNP analysis: "D" to "N" and "K" to "Q")

First, obtain a protein sample containing both the wild-type and mutant forms. Then, perform immunoprecipitation (IP) to selectively pull down the proteins of interest. Use an antibody specific to the protein sequence for IP. The IP will capture both the wild-type and mutant proteins if they are expressed. Next, separate the immunoprecipitated proteins by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and transfer them onto a membrane. Probe the membrane with antibodies specific to the mutant sites. If the mutant proteins are expressed, they will appear as additional bands on the blot. Additionally, perform liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis to identify and quantify the peptides corresponding to the wild-type and mutant sequences. Finally, if the mutant peptides are detected which means this confirms the mutant proteins expression.