BMB/Bi/Ch 173 – Winter 2017
Problem Set 8.2: Mass Spectrometry– Assigned 3-2-17. Due 3-7-17 by 10:30am
TA: Wen Zhou (201 Kerckoff, office hour: Fri 3/3 5-6pm, Mon 3/6 5-6pm or by appointment
100 points total

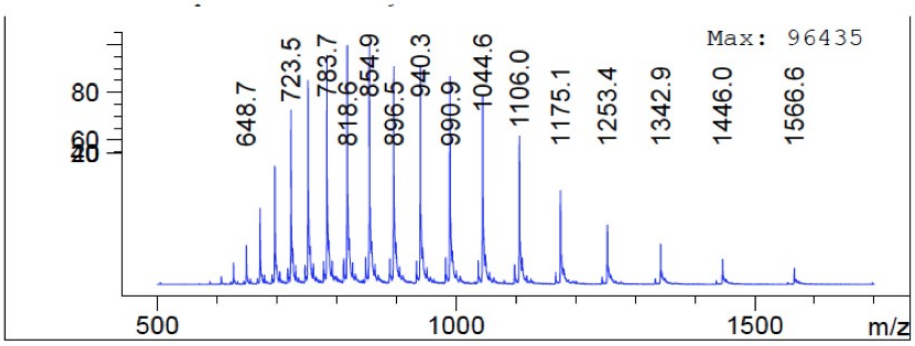
1. **(10 points) Ionization Basics:**

   What is the difference between hard and soft ionization techniques? Describe a scenario in which one would use a soft ionization technique rather than a hard one.

   Hard ionization techniques, such as electron and chemical ionization, result in the analyte being fragmented into ions of multiple varying sizes. Soft ionization techniques, such as MALDI and electrospray ionization, preserve the analyte in one fragment. An example of a scenario in which one would use soft rather than hard techniques would be in the analysis of intact large biomolecules (DNA, proteins, peptides and sugars) or large organic molecules (polymers, dendrimers and other macromolecules), as these analytes would be too fragmented by hard ionization techniques.

2. **(20 points) Mass Spec of Intact Proteins:**

a. (8 Points) Below is the mass spectrum of an intact protein. Why are there multiple peaks in the spectrum if the protein has not been digested? What ionization technique was used?



Although the protein is intact, there are multiple peaks because each peak represents a different charge state of the protein. Because the spectrum shows multiple charge states, the data was collected using electrospray ionization.

b. (12 points) Determine the mass of the protein. Show your work. (Hint: use the formulas: $(M + nH)/n = (m/z)_n$ and $(M + (n-1)H)/(n-1) = (m/z)_{(n-1)}$)

The answer should be around 18,785 Da. Using the equations from the lecture, where H = 1.00784, we can see that we have two equations and two unknowns:

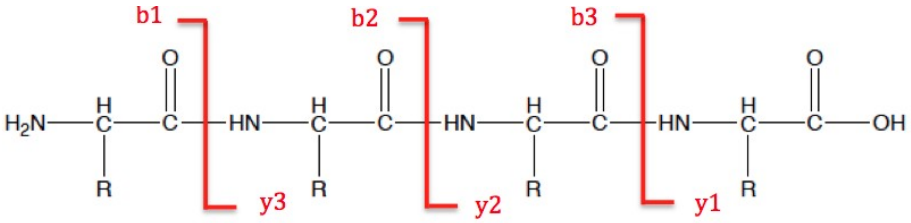
$(M + nH)/n = (m/z)_n$

$(M + (n-1)H)/(n-1) = (m/z)_{(n---1)}$

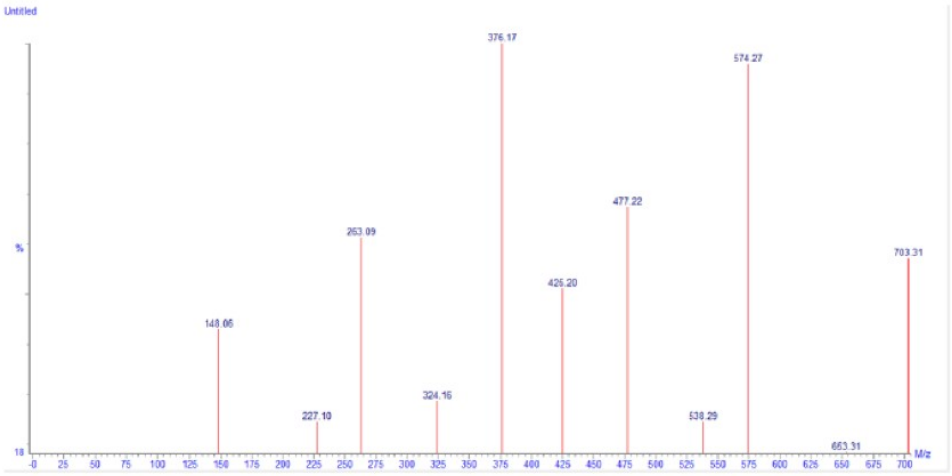
with $(m/z)_n$ and $(m/z)_{(n---1)}$ being the mass---to---charge ratios of two adjacent peaks, M being the mass of the protein, and n being the charge state.

3. (35 points) Peptide Analysis:

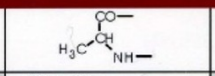
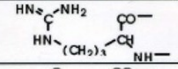
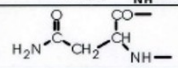
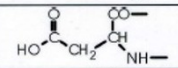
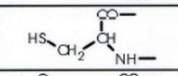
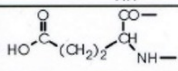
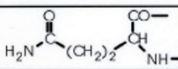
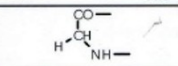
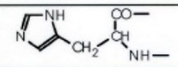
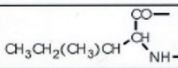
a. (20 points)When peptides are fragmented, they are broken into pieces at their amide bonds. This process forms a series of b-ions (with C- terminal acylium ions) and y-ions (with N-terminal ammonium ions), as shown:



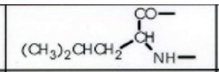
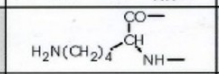
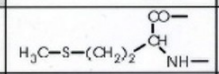
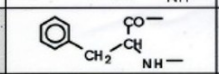
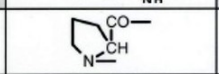
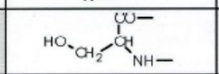
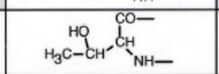
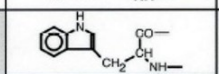
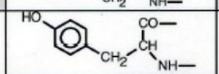
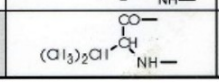
Shown below are the b- and y-ion series of a fragmented peptide ion with a molecular weight of 799.360 Da. The singly charged precursor was measured at 800.3638 Da. Using the spectrum, determine the amino acid sequence of the peptide. On the spectrum, please label which peaks correspond to which b- and y- ions (b1, b2, etc.) and give the sequence of each ion. The residual weights of the amino acids are listed in tables below (you only need to consider the monoisotopic masses). Be sure to show your work. For clarity, the peaks in the spectrum have the following m/z values: 148.06, 227.10, 263.09, 324.16, 376.17, 425.20, 477.22, 538.29, 574.27, 653.31, and 703.31



Residual molecular weights of amino acids

Name (Symbols)	Residue Composition	Residue Structure	Average Mass	Monoisotopic Mass	Sidechain Mass	Immonium Ion Mass (related i
Alanine (Ala or A)	C_3H_5NO		71.0788	71.03711	15	44
Arginine (Arg or R)	$C_6H_{12}N_4O$		156.1876	156.10111	100	129 (112, 87,73,70,59
Asparagine (Asn or N)	$C_4H_6N_2O_2$		114.1039	114.04293	58	87 (70)
Aspartic Acid (Asp or D)	$C_4H_5NO_3$		115.0886	115.02694	59	88
Cysteine (Cys or C)	C_3H_5NOS		103.1448	103.00919	47	76
Glutamic Acid (Glu or E)	$C_5H_7NO_3$		129.1155	129.04259	73	102
Glutamine (Gln or Q)	$C_5H_8N_2O_2$		128.1308	128.05858	72	101 (84,129
Glycine (Gly or G)	C_2H_3NO		57.0520	57.02146	1	30
Histidine (His or H)	$C_6H_7N_3O$		137.1412	137.05891	81	110(166, 138, 121, 82)
Isoleucine (Ile or I)	$C_6H_{11}NO$		113.1595	113.08406	57	86 (72)

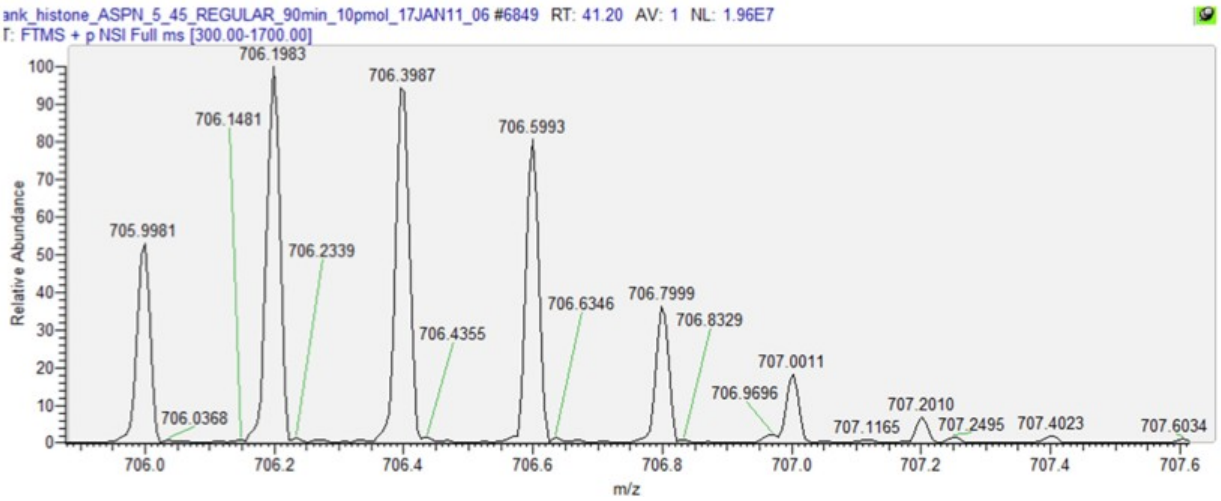
Residual molecular weights of amino acids

Leucine (Leu or L)	$C_6H_{11}NO$		113.1595	113.08406	57	86 (72)
Lysine (Lys or K)	$C_6H_{12}N_2O$		128.1742	128.09496	72	101 (129,112,84
Methionine (Met or M)	C_5H_9NOS		131.1986	131.04049	75	104 (61)
Phenylalanine (Phe or F)	C_9H_9NO		147.1766	147.06841	91	120 (91)
Proline (Pro or P)	C_5H_7NO		97.1167	97.05276	41	70
Serine (Ser or S)	$C_3H_5NO_2$		87.0782	87.03203	31	60
Threonine (Thr or T)	$C_4H_7NO_2$		101.1051	101.04768	45	74
Tryptophan (Trp or W)	$C_{11}H_{10}N_2O$		186.2133	186.07931	130	159
Tyrosine (Tyr or Y)	$C_9H_9NO_2$		163.1760	163.06333	107	136
Valine (Val or V)	C_5H_9NO		99.1326	99.06841	43	72

800.36-703.31=97.05→proline
703.31(y6)-574.27=129.04→Glu
574.27(y5)-477.22=97.05→proline
477.22(y4)-376.17=101.05→Thr
376.17(y3)-263.09=113.08→ Leu/Ile
263.09(y2)-148.06(y1)=115.05→Asp
148.06(y1) →Glu+ -OH + H

227.1(b2)=Pro+Glu
324.16(b3)=Pro+Glu+Pro
425.2(b4)=Pro+Glu+Pro+Thr
538.29(b5)= Pro+Glu+Pro+Thr+Ile(or Leu)
653.31(b6)=Pro+Glu+Pro+Thr+Ile(or Leu)+Asp
thus :
Pro-Glu-Pro-Thr-Leu/Ile-Asp-Glu
The sequence is PEPTIDE or PEPTLDE (I and L have the same molecular weight and can't be distinguished from this spectrum). One might think that the first peak is due to an F ion. However, it's actually an E ion since it is a y-ion and therefore has an ammonium group. You need to remember that the monoisotopic mass is given without the extra H^+ on the nitrogen, as well as the extra –OH required to complete the mass.

b. (15 points) Shown below is part of the MS spectrum of an AspN digested peptide from a thymus calf histone. Note: the small peaks labeled with green lines represent low abundance fragments and can be ignored. What is the charge state of this peptide fragment? What is the molecular weight of this peptide? For your calculation, use 1.00784 as the mass of a proton. Show your work.



These peaks were zoomed in at 1 charge state of ESI spectrum, they are resulted from isotopic masses (each peak differ in mass of exactly proton). Difference between major peaks is ~0.2, so the charge state is 1/0.2 =**5**
The mass is about **3524.9541 Da.**
$M = (m/z)n - n\,H$

4. **(35 points) Mass spec proteomics**. For your thesis, you are studying a protein in *C.elegans*, WoRM1. You are interested how WoRM1 expression affects the *C. elegans* proteome, and you have a knock-out line of WoRM1.
   a. (6 points) Describe two types of additional information you could get from a mass spec proteomics experiment that you could not get from studying the transcriptome (ie the sequences of all of the mRNAs in the cell).

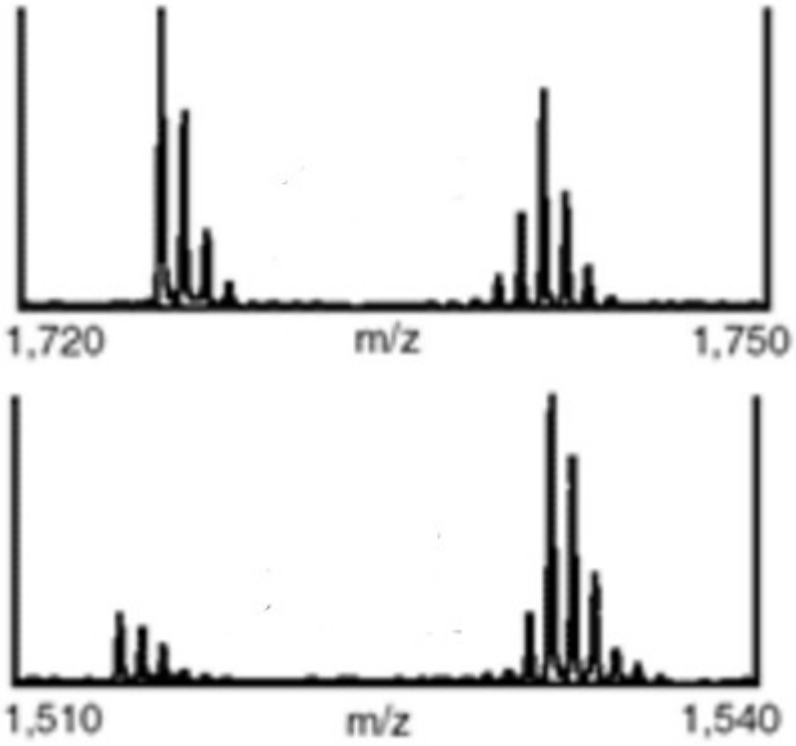
   If you just study the transcriptome, you are only measuring the transcriptional activity and mRNA stability of that gene. Other processes that affect the proteome after transcription include protein translation, folding and stability/degradation rates. Furthermore, mass spec has the ability to give information about post-translational modifications such as lysine methylation or protein cleavage, sub-cellular localization of the protein, and interaction partners of the protein.

   b. (21 points) Because different proteins and peptides behave differently during ionization and detection, peak intensities do not correlate with abundance in a typical mass spec experiment. Therefore, to make a mass spec experiment quantitative, controls using heavy isotopes are employed. One common way to do so is the stable isotope labeling by amino acids in cell culture (SILAC) method. In this type of experiment, one population of cultured cells is grown in media with isotopically-labeled amino acids, producing a "heavy" sample relative to a "light" sample grown in normal media. These two samples are then mixed and enzymatically digested to produce peptides. These peptides are separated via column chromatography and measured by a mass spectrometer. By searching databases, the proteins that produced the peptides by enzymatic cleavage can be identified. The relative measured intensities for the same "light" and "heavy" peptides can then be correlated to fold-changes in the abundance in that protein between the two samples since the same exact peptides are being compared.

      i. (6 points) You are interested in using SILAC to determine the effects of WoRM1 KO on the *C. elegans* proteome. While cells in tissue culture can easily be labeled using cell culture media with heavy isotope amino acids, labeling an entire organism is a bit trickier. How can you create a "heavy" isotope population of *C. elegans*?

         You can create heavy isotope-labeled *C. elegans* by feeding the worms *E. coli* grown on an isotopically-labeled food source, such as ^{15}N. For example, see Krijgsveld et al. *Nature Biotechnology* **21**, 927 - 931 (2003).

ii. (10 points) You label the WT worms with "heavy" and the KO worms with "light." Below are spectra of peptides from two different proteins. Why are there multiple peaks for the "light" and "heavy" samples? Why are the patterns of intensities of these multiple peaks different between "light" and "heavy"?



There are multiple peaks for both light and heavy peptides because there are naturally occurring isotopes that incorporate naturally into the peptides. The patterns are different for light and heavy because the heavy peptides will have varying levels of incorporation of the heavy amino acids in addition to the presence of naturally occurring isotopes.

iii. (5 points) Estimate the relative abundances of light versus heavy peptides in each of the two spectra (ie what is the fold-change). Is each protein up-regulated, down-regulated, or unaffected by the KO of WoRM1?

Right side protein: 1:1 light:heavy; no change/slight upregulated in the KO
Left side protein: 1: 4 light:heavy; down-regulated in the KO

c. (8 points) You get some interesting hits from your initial experiment, and you want to know which of the proteins physically interact with WoRM1. You decide to do a co-immunoprecipitation of WoRM1 and use mass spec to

identify its binding partners. There are no available antibodies to WoRM1, so you decide to introduce an exogenous epitope-tagged version of WoRM1. You plan on co-IPing WoRM1 using an antibody to this tag and identifying the bound proteins by mass spec in the same way you identified the proteins in the SILAC experiment (except without isotopic labeling). Explain two potential problems that could arise from using this method rather than co-IPing endogenous untagged WoRM1.

exogenous epitope tag can change the WoRM1 protein functionality, resulted in changes in binding partners, thus sample being mass spec-ed can contain tagging artifact. Unless one introduce the tag on its original genomic locus, the expression level of tagged protein is difficult to control, and they can also compete with endogenous untagged protein (if didn't knock out endogenous copy) resulted in inability to pulldown interaction partners or perturbation of functional complexes due to inaccurate expression level.