GCB/CIS 535
Fall 2004

Home work 6 (part 2): Mass Spectrum Analysis

Due in class, Wednesday, Dec 8, 2004

Tools:

1. PeptideCutter:
http://us.expasy.org/tools/peptidecutter/
2. Protein identification by PepFrag:
http://prowl.rockefeller.edu/PROWL/pepfragch.html
3. Entrez-NCBI:
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi

Questions:

We used Trypsin to digest a certain unknown protein sample and got the following mass spectrometric data. Here we are going to try to identify the peptide and the corresponding protein.

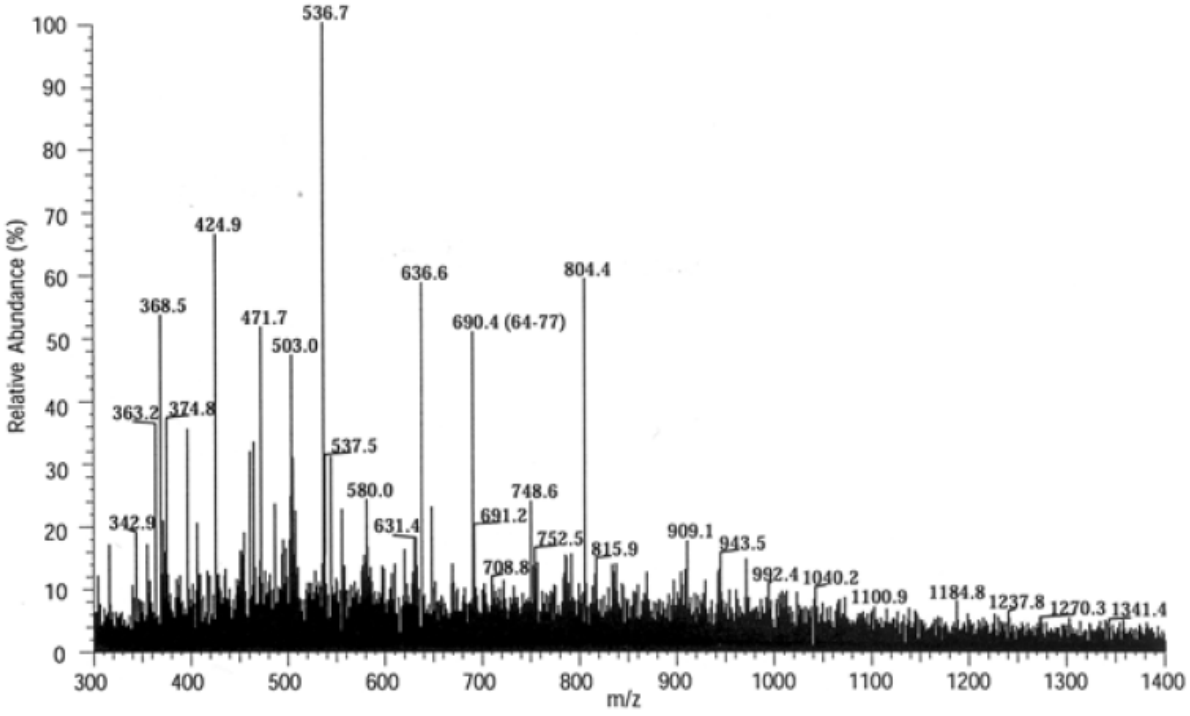

Fig 1. Full Scan

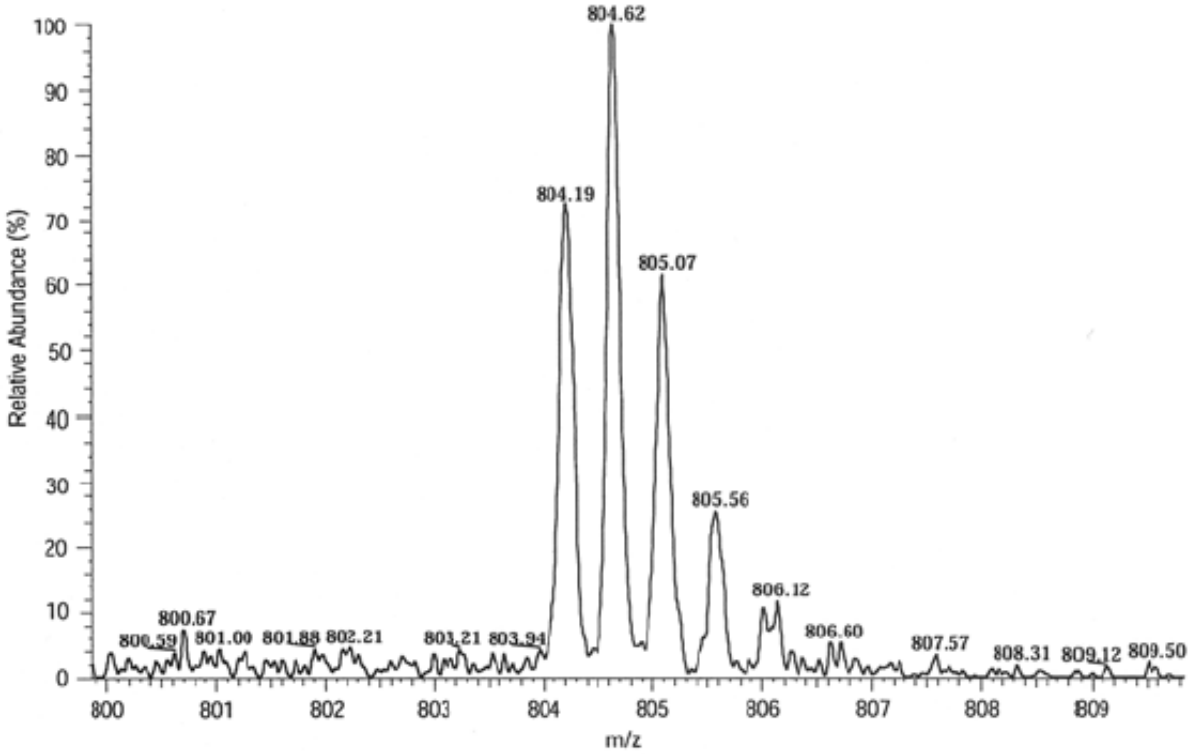

Fig 2. Zoom Scan

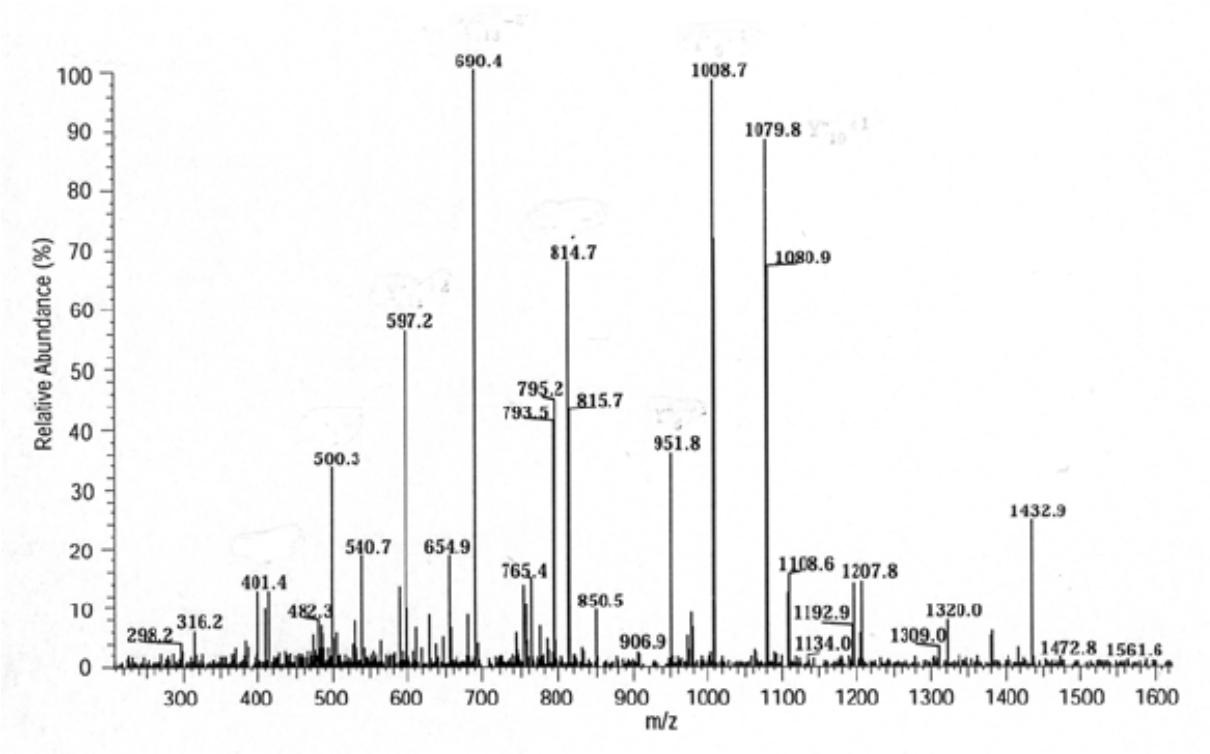

Fig 3. MS/MS spectrum

1. (3 pts) In the Zoom Scan, the center peaks (804.19, 804.62, 805.07, 805.56) with relatively high abundance correspond to peak 804.4 in the full scan. How are these peaks formed? What does the difference in m/z value of these center peaks tell you about the charge state of the peptide samples? Based on this, what is the average molecular mass (in Daltons) of the peptides corresponding to peak 804.4?

2. (2 pts) We then generated the tandem MS/MS spectrum of this peptide. We know that most MS/MS fragment ions are single charged, so the m/z value equals the mass. Based on the mass difference between neighboring peaks and the molecular weights for each amino acids given in the appendix, try to deduce 3 amino acids that are connected to each other in the peptide. For instance, if you have 4 neighboring peaks (831.4, 994.6, 1081.5, 1178.4), you would get YSP (Tyr-Ser-Pro). And the sequence tag representation of the data would be (831.4)YSP(1178.4). Note, not all the peaks in the spectrum are informative in this way, since they may not be the b-type and y-type ions we are interested in. Report your result as a sequence tag.

3. (6 pts) (UPDATED) Use the PepFraq tool given above, try to identify the peptide with the mass calculated in question 1 and the sequence tag inferred in question 2. Use all default values except for the followings:

   --Maximum number of cleavage sites not cleaved in a peptide = 1
   --Use "Average mass"
   --Input your calculated mass from question 1; allow +/- 2 (Dalton) errors; choose "**M**" to indicate that the mass is for neutral peptide.
   --Input the Fragment ion masses separated by space you used in question 2 to estimate the sequence tag. Following the example given in question 2, you would input " 831.4 994.6 1081.5 1178.4). Allow for +/- 1 (Dalton) errors.
   --For the first round, select ONLY the **y" type** of ions for search. For the second round, select ONLY the **b type** ions for search.
   --Input the amino acids you estimated in question 2 after "**Contains** the following amino acids:". Following the example given in question 2, you would input "YSP".

   What peptides and proteins are returned when you search by **y" type** ions?
   What peptides and proteins are returned when you search by **b type** ions?
   What peptides and proteins are returned when you search by both **b type** and **y" type** ions and **without** telling the program what amino acids are contained in the peptide?

4. (4 pts) (UPDATED) Retrieve the protein sequence from the search results using **y" type** ions only. If you got more than one peptide or protein from the result, for each different peptide (only look at the highlighted and underlined part, which is the peptide having the mass calculated in question 1), choose only one

protein. The program will give you the protein id number (the number following "**gi**"). Use Entrez-NCBI to query for protein with that id. Then Use PeptideCutter to calculate the theoretical cutting sites of Trypsin digestion (with more than 80% probability) and the resulting peptides. List the peptides for which you could identify the corresponding peaks in the Full Scan. What percentage of the protein is covered by these peptides?

Appendix: Individual properties and images of amino acids[a]

Properties and images (name: NIST WebBook, three letter code: GIF, one letter code: VRML)

amino acid			mass	surface [b]	volume [c]	pK_a [d]	pI [e]	solubility [e]	density [e]
Alanine	ALA	A	71.09	115	88.6	-	6.107	16.65	1.401
Arginine	ARG	R	156.19	225	173.4	~12	10.76	15	1.1
Aspartic Acid	ASP	D	114.11	150	111.1	4.5	2.98	0.778	1.66
Asparagine	ASN	N	115.09	160	114.1	-	-	3.53	1.54
Cysteine	CYS	C	103.15	135	108.5	9.1-9.5	5.02	very high	-
Glutamic Acid	GLU	E	129.12	190	138.4	4.6	3.08	0.864	1.460
Glutamine	GLN	Q	128.14	180	143.8	-	-	2.5	-
Glycine	GLY	G	57.05	75	60.1	-	6.064	24.99	1.607
Histidine	HIS	H	137.14	195	153.2	6.2	7.64	4.19	-
Isoleucine	ILE	I	113.16	175	166.7	-	6.038	4.117	-
Leucine	LEU	L	113.16	170	166.7	-	6.036	2.426	1.191
Lysine	LYS	K	128.17	200	168.6	10.4	9.47	very high	-
Methionine	MET	M	131.19	185	162.9	-	5.74	3.381	1.340
Phenylalanine	PHE	F	147.18	210	189.9	-	5.91	2.965	-
Proline	PRO	P	97.12	145	112.7	-	6.3	162.3	-
Serine	SER	S	87.08	115	89.0	-	5.68	5.023	1.537
Threonine	THR	T	101.11	140	116.1	-	-	very high	-
Tryptophan	TRP	W	186.12	255	227.8	-	5.88	1.136	-
Tyrosine	TYR	Y	163.18	230	193.6	9.7	5.63	0.0453	1.456
Valine	VAL	V	99.14	155	140.0	-	6.002	8.85	1.230

[a] mass [dalton], surface [$Å^2$], volume [$Å^3$], pK_a [side chain], pI [at 25°C], solubility [g/100g, 25°C], density [crystal density, g/ml],
name: information from NIST Chemistry WebBook, three letter code: GIF, one letter code: VRML

[b] C.Chothia, J. Mol. Biol., 105(1975)1-14

[c] A.A. Zamyatin, Prog. Biophys. Mol. Biol., 24(1972)107-123

[d] C. Tanford, Adv. Prot. Chem., 17(1962)69-165

[e] The Merck Index, Merck & Co. Inc., Nahway, N.J., 11(1989); CRC Handbook of Chem.& Phys., Cleveland, Ohio, 58(1977)