Biomolecular Mass Spectrometry

- Lipids
  - Not different than other organic small molecules
- Carbohydrates
  - Polymers of monosaccharides linked via glycosidic bonds (acetals/ ketals) many different combinations- very interesting no time
- Nucleic acids
  - Polymers of nucleotide linked via phosphodiester bonds- MS not generally used….why?
- Proteins/ Peptides
  - Remainder of this lecture

Biomolecular Mass Spectrometry

- Say you have a purified protein/ peptide sample and you need to know what it is
  - How do you go about identifying it?
    - Mass Spectrometry is the modern way to solve this problem
    - Proteins and peptides are easily ionized using modern sources
    - A little biochemistry knowledge makes this a tractable problem (with the help of some pioneers)

Biomolecular Mass Spectrometry

- Mass Spectrometry Based-Sequencing
  - Mass Analyzers
    - Separates compounds based on mass/ charge (m/z)
      - A compound with mw = 100 g/mol that have a +1 charge will be seen at m/z 101, why? mw + 1 (mass of adding H^+)/ charge of 1
      - The same compound with a +2 charge will be seen at m/z 51, why? mw + 2 (mass of adding 2 H^+)/ charge of 2
    - Two possible approaches
      - Mass Fingerprinting (one mass [M+H]+) used to ID (ADGCY same mass as YCGDA same mass as CGDAY, etc)
      - Peptide Sequencing (select [M+H]+) fragment into many smaller masses and look at the pieces to ID (MUCH more confidence)

Biomolecular Mass Spectrometry

- Mass Spectrometry Based-Sequencing
  - Source (ionize the samples)
    - Peptides analyzed in acidic aqueous solutions (allows positive ions to be formed (singly and doubly charged most common), this makes fragmentation along the backbone relatively easy and straightforward
    - This happens either by applying heat and high voltage (1-4 kV) (ESI) or by using laser light and an organic matrix molecule to transfer energy to the peptide (MALDI)

Biomolecular Mass Spectrometry

- So we just determine the overall mass of entire proteins, right? (Mass fingerprinting)
  - What are the issues with this approach?
  - How about determining the mass of pieces (peptides) of proteins?
- In most cases proteins are identified based on peptide fragments (Sequencing)
  - MS after specific enzymatic digestion
  - Two main approaches

Biomolecular Mass Spectrometry

- What is the general structure of an amino acid?
- What is the general structure of a polypeptide?
- What is the convention for writing/ drawing peptide sequences?

Biomolecular Mass Spectrometry

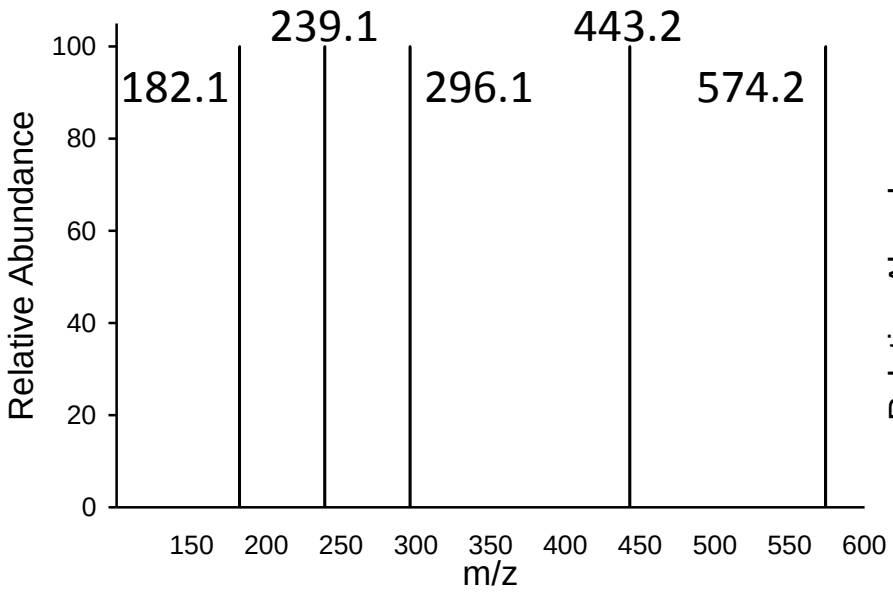
Amino Acid	Abbreviation	Residue Mass (Mono, Da)	Amino Acid Mass (Mono, Da)	Immonium Ion Mass (Nom., Da)	Side-chain Mass (Nom., Da)	Neutral Loss (Nom., Da)
Alanine	A	71.0371	89.04767	44	15	--
Arginine	R	156.1011	174.1117	129	100	17
Asparagine	N	114.0429	132.0535	87	58	17
Aspartic Acid	D	115.0269	133.0375	88	59	18
Cysteine	C	103.0092	121.0198	76	47	34(92)
Glutamic Acid	E	129.0426	147.0532	102	73	18
Gluatamine	Q	128.0586	146.0692	101	72	17
Glycine	G	57.0215	75.03207	30	--	--
Histidine	H	137.0589	155.0695	110	81	--
Isoleucine	I	113.0841	131.0947	86	57	--
Leucine	L	113.0841	131.0947	86	57	--
Lysine	K	128.0950	146.1056	101	72	17
Methionine	M	131.0405	149.0511	104	75	48
Phenylalanine	F	147.0684	165.079	120	91	--
Proline	P	97.0528	115.0634	70	--	--
Serine	S	87.0320	105.0426	60	31	18
Threonine	T	101.0477	119.0583	74	45	18
Tryptophan	W	186.0793	204.0899	159	130	--
Tyrosine	Y	163.0633	181.0739	136	107	--
Valine	V	99.0684	117.079	72	43	--

Biomolecular Mass Spectrometry

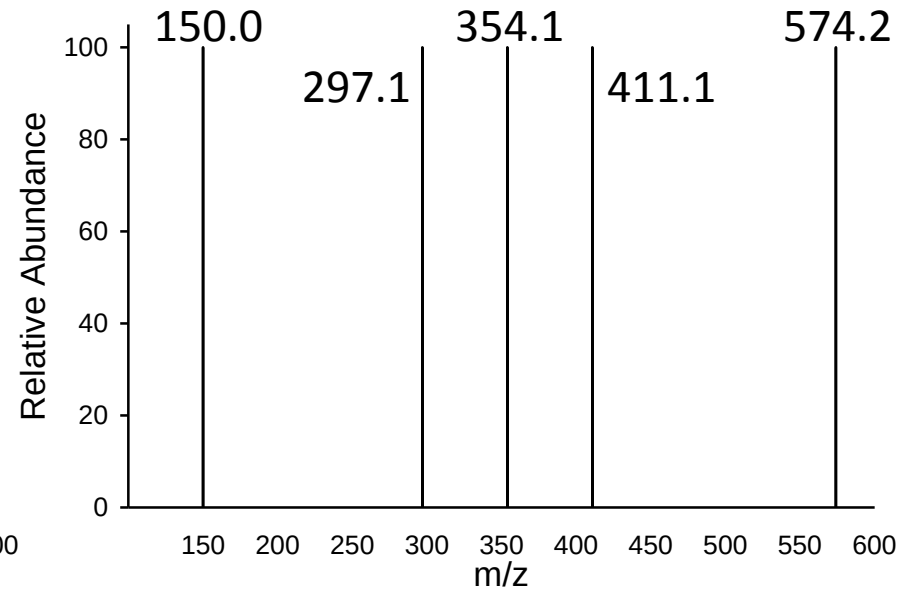
- Peptide Ladder Sequencing
  - Relatively small, pure peptide
  - Incompletely treat with peptidase
    - Amino peptidase (cleaves 1 amino acid at a time from the N-terminus)
    - Carboxy peptidase (cleaves 1 amino acid at a time from the C-terminus)
    - Results in a nested set of peptides differing by the mass of each residue
    - Not the same as sequence determination by MS/MS

Biomolecular Mass Spectrometry

Carboxypeptidase digestion



Aminopeptidase digestion



Biomolecular Mass Spectrometry

- Aminopeptidase digestion of peptide U yields ions at m/z 132, 245, 359, 474, 561, 690, and 876 what is the sequence of this peptide?

- Carboxypeptidase digestion of peptide U2 yields ions at m/z 90, 221, 350, 497, 644 what is the sequence of this peptide?

Biomolecular Mass Spectrometry

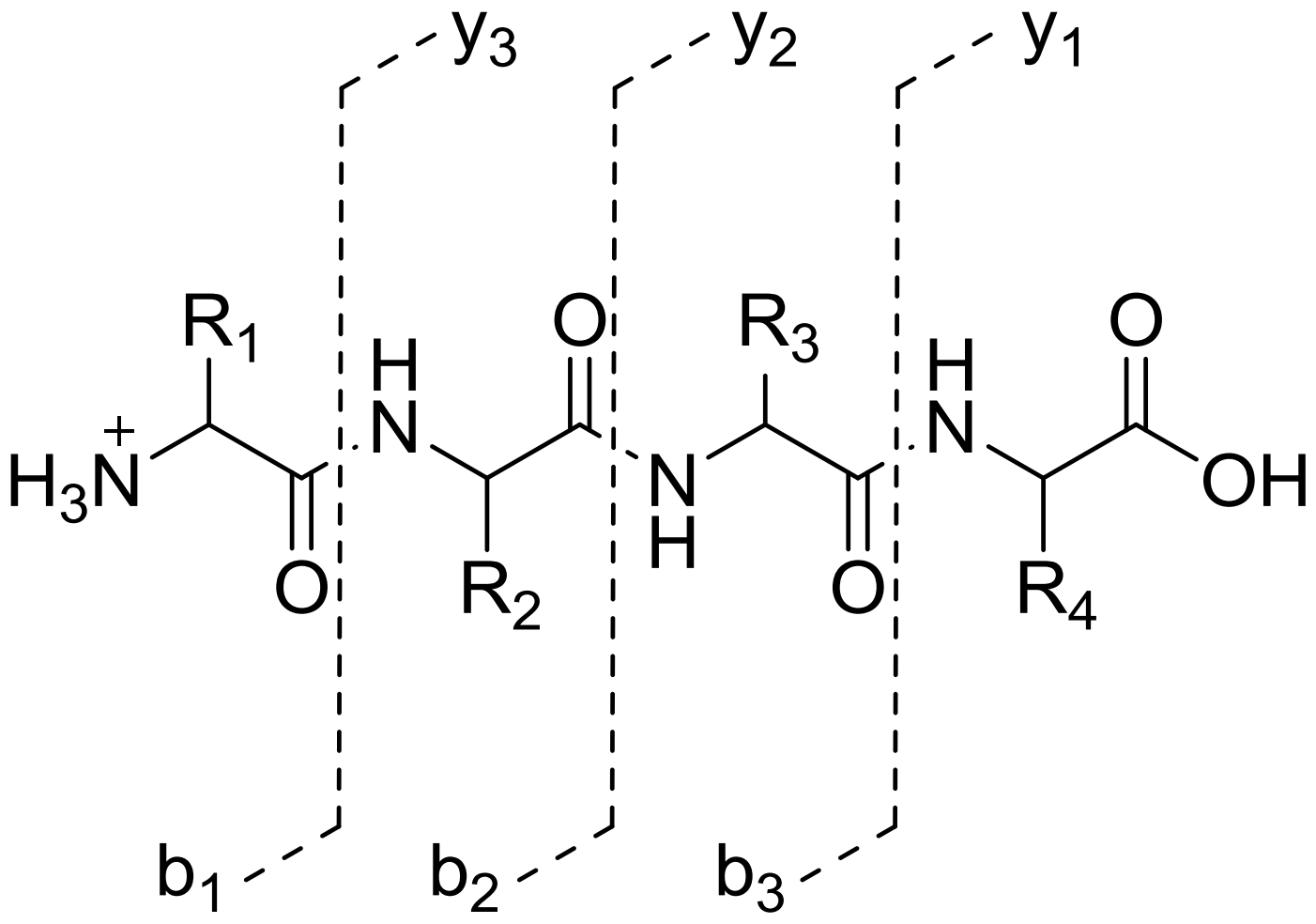
- Limitations of peptide laddering?
  - Pure sample a must
  - Has to be relatively small (<2000 Da)
  - Multiply charged fragments cause a problem
  - Incomplete digestion required
    - Too much problematic
    - Too little problematic

Biomolecular Mass Spectrometry

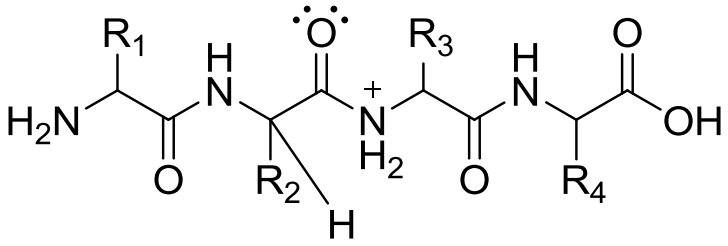
- How do we deal with complex mixtures of proteins/ peptides?
- Or unknown samples?
- Or samples that are likely to be greater than ~2000 Da?

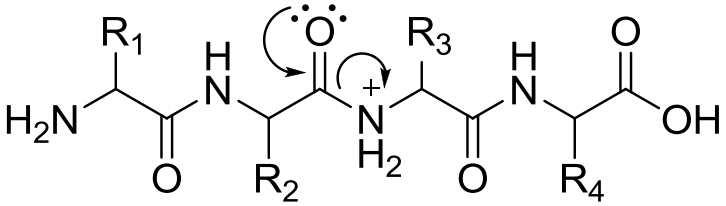
- Collect sequence information for tandem mass spectrometry experiments

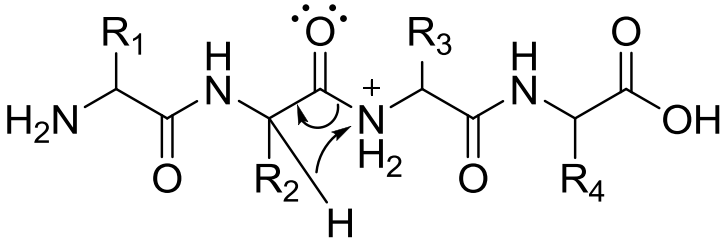
Biomolecular Mass Spectrometry

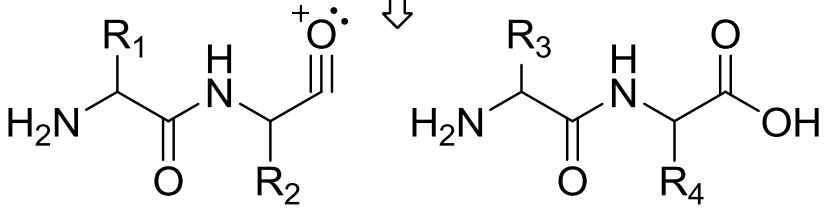


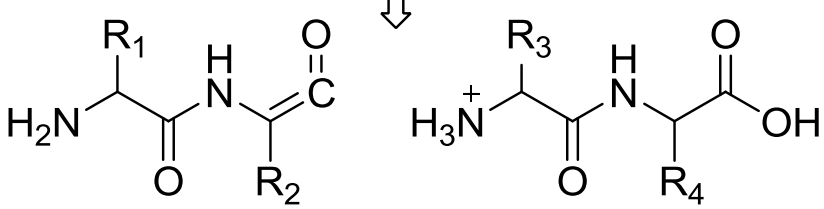
Biomolecular Mass Spectrometry











b_2

y_2

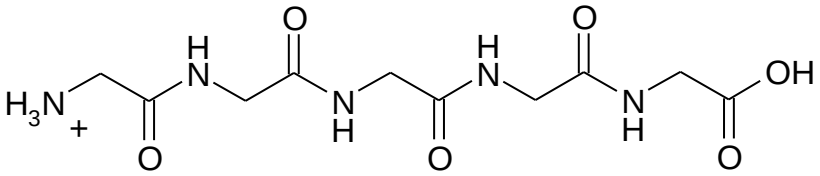
Biomolecular Mass Spectrometry

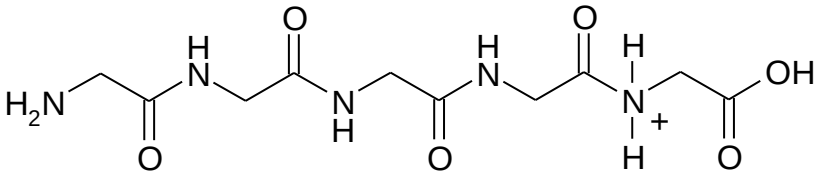
- If you know the sequence of your protein/ peptide then interpreting MS/MS data is relatively straightforward
- Based on the sequence you can calculate appropriate fragmentations and compare them to the experimental data
- For example (actually four examples)…

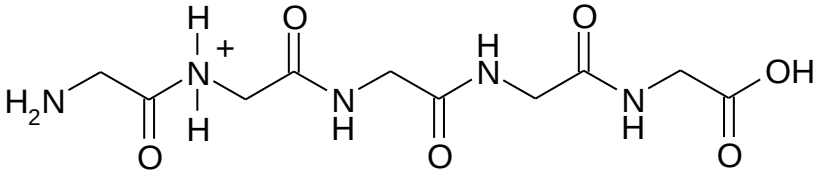
Biomolecular Mass Spectrometry

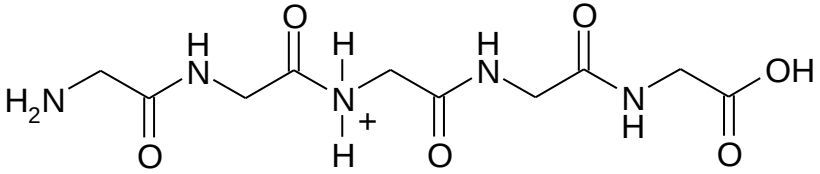
- Draw the structure of penta-glycine (uncharged)
- Identify all the b- and y-series cleavages
- Draw all the possible protonated species for this structure
- Calculate the m/z for all b- and y-series ions (nominal masses)

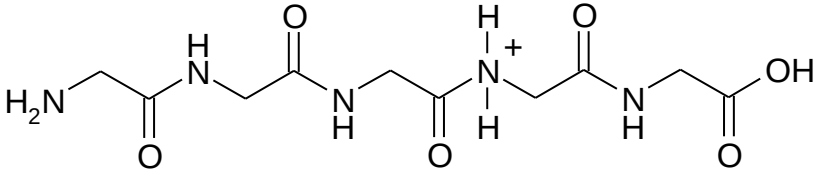
Biomolecular Mass Spectrometry





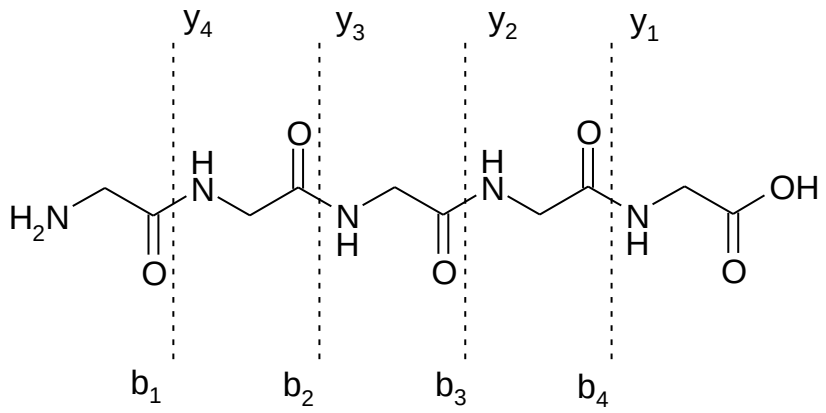






$[M+H]^+ = 304$

247 190 133 76



58 115 172 229

Biomolecular Mass Spectrometry

- Draw the structure of GYNKE under acidic LC-MS conditions (i.e. H_20:ACN:H_3COOH, 50/50/0.1 v/v/v)
- Calculate the molecular weight (nominal) and anticipated full scan m/z for this peptide
- Show the mechanism for the formation of the b_3- and y_4-ion of this peptide
- Calculate all b- and y-series ions resulting from the low energy CID of this peptide (nominal masses)

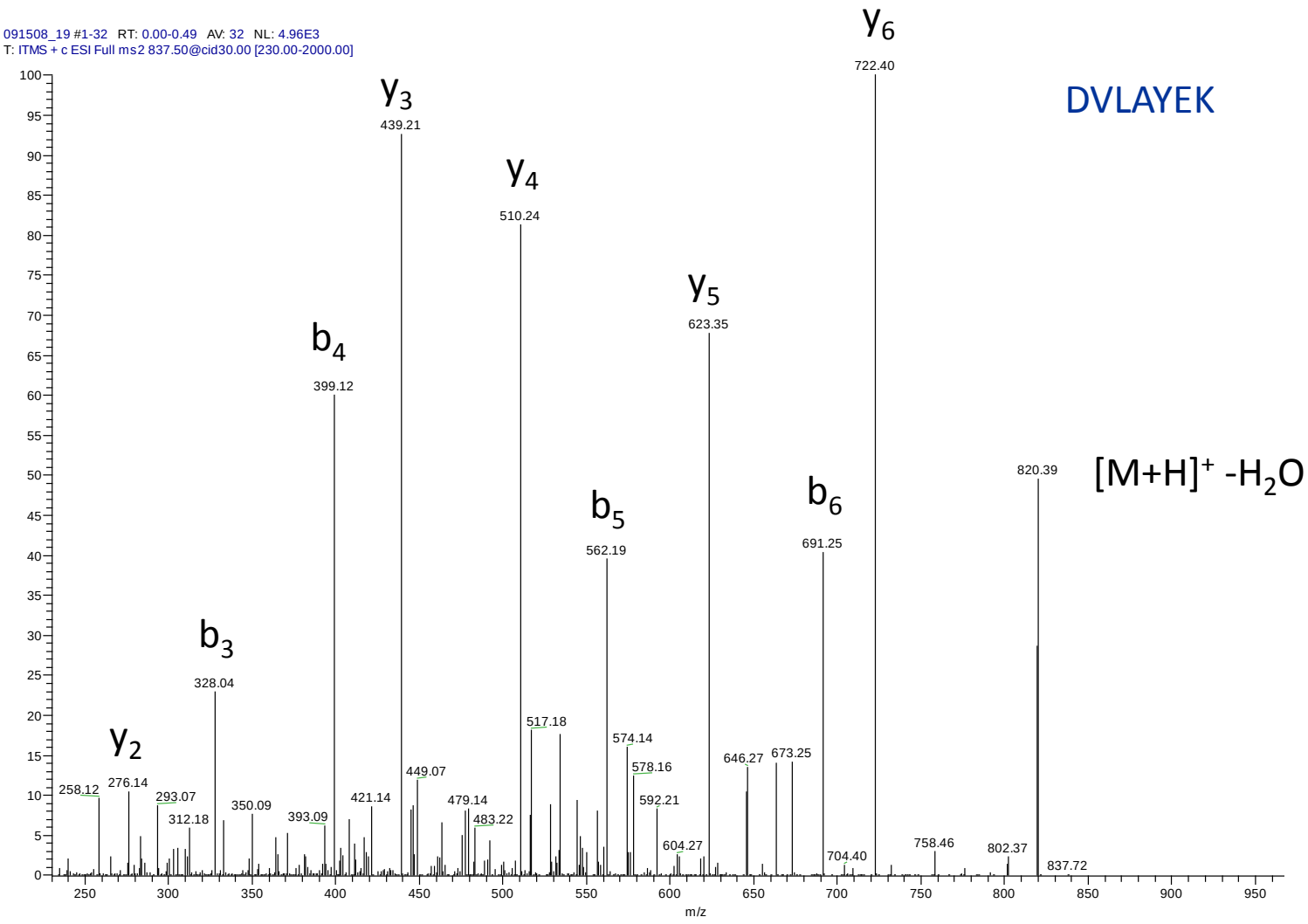
Biomolecular Mass Spectrometry

- Draw the structure of TQSTFK under acidic LC-MS conditions (i.e. H_20:ACN:H_3COOH, 50/50/0.1 v/v/v)
- Calculate the molecular weight (nominal) and anticipated full scan m/z for this peptide
- Show the mechanism for the formation of the b_5- and y_3-ion of this peptide
- Calculate all b- and y-series ions resulting from the low energy CID of this peptide (nominal masses)

Biomolecular Mass Spectrometry

- Draw the structure of VLRIST under acidic LC-MS conditions (i.e. H_20:ACN:H_3COOH, 50/50/0.1 v/v/v)
- Calculate the molecular weight (nominal) and anticipated full scan m/z for this peptide
- Show the mechanism for the formation of the a_2- and y_2-ion of this peptide
- Calculate all b- and y-series ions resulting from the low energy CID of this peptide (nominal masses)

Biomolecular Mass Spectrometry



Biomolecular Mass Spectrometry

- What if you are conducting LC-MS/MS analysis of peptide digests and you do not know their sequences?
- How do you determine sequence information from these data?

- De novo peptide sequencing or determining sequence information without prior knowledge

Biomolecular Mass Spectrometry

- Modern MS software packages will do de novo sequence analysis via the comparison of MS/MS data against known protein database information
- Proteins are first digested *in silico* to generate appropriate fragments (tryptic, etc), then b- and y-series ions are calculated
- Comparisons are made and correlations scores are calculated

Biomolecular Mass Spectrometry

- Many times important data will require manual verification of the “automated analysis”
- To prevent the data reduction and evaluation being thought of as a “black box” you need to understand how to do de novo sequencing yourselves
- This website walks you thorugh the steps at your own pace
  - http://www.ionsource.com/tutorial/DeNovo/low_mass_match.htm

Biomolecular Mass Spectrometry

- De novo sequencing rules (MS/MS according to Dass)

1. Look for immonium ions, this provides a hint to the residues present (but no their relative positions) this is problematic for ion trap data (why?)
2. Ignore loss of neutrals (H_2O, NH_3, CO_2, etc at high m/z
3. Look for mass pairs differing by 28 Da (loss of CO) b- and a-series

Biomolecular Mass Spectrometry

- De novo sequencing rules (MS/MS according to Dass)

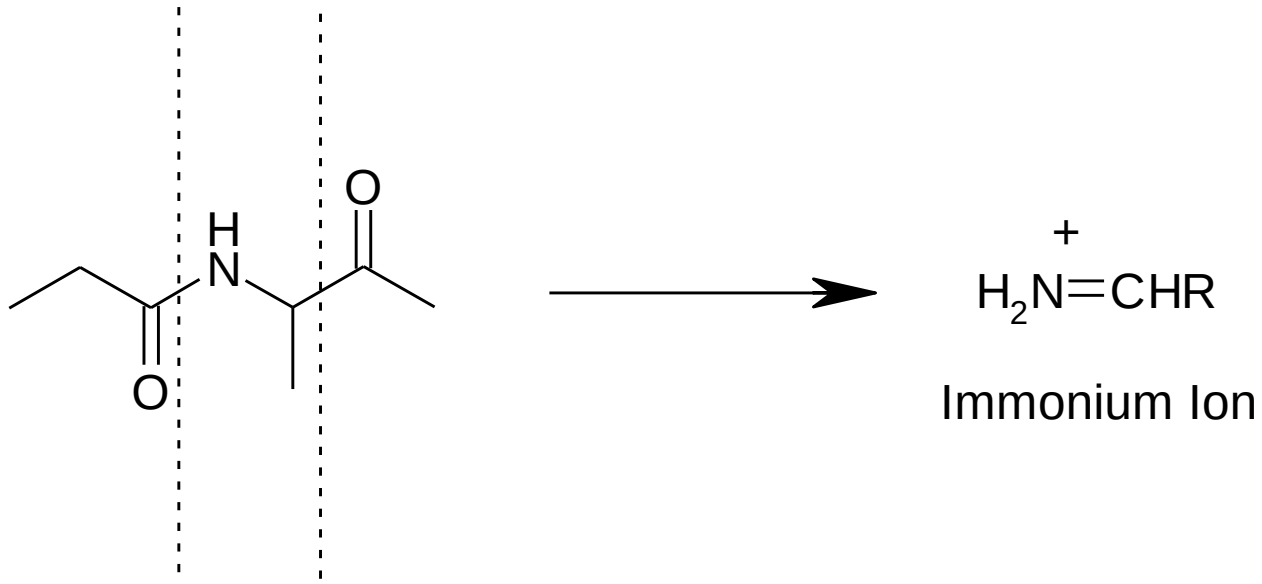
4. Look for b_2 (b_1 not seen) this will allow determination of the first two N-terminal residues (XY or YX) be careful about redundant masses and those that are equivalent to single residue masses, determination of b_2 allows assignment of y_{n-2}
5. Once an ion is assigned try to determine the next in sequence by adding/ subtracting residue masses of the common amino acids

Biomolecular Mass Spectrometry

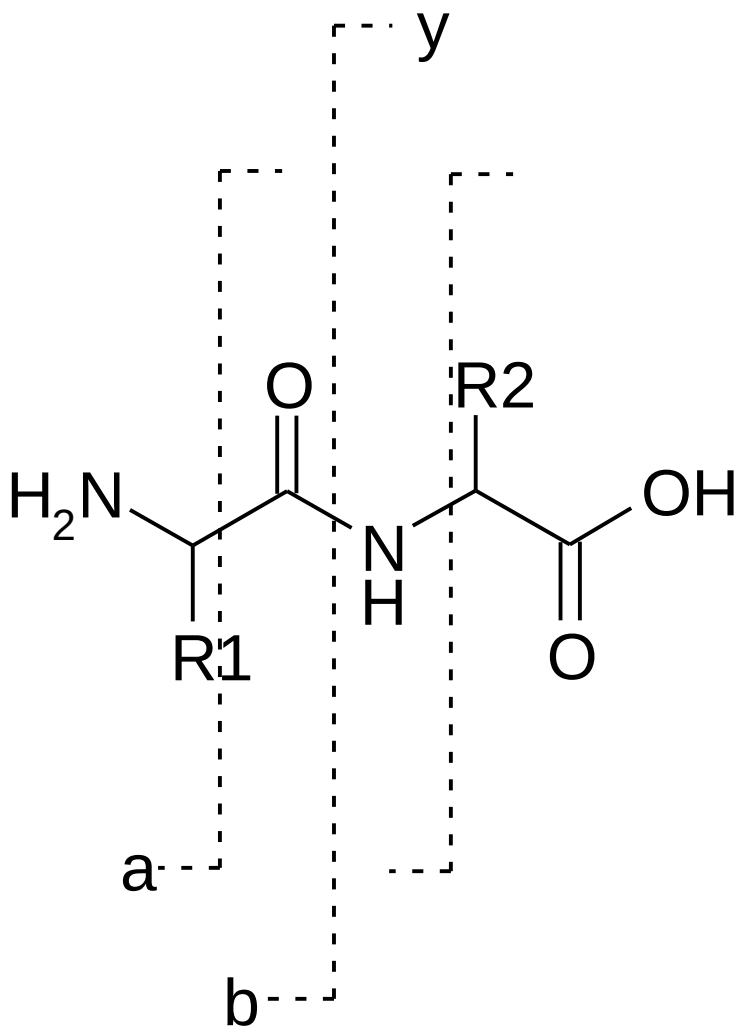
- De novo sequencing rules (MS/MS according to Dass)

6. Try to determine y_{n-1}, once b_2 is known only two possibilities can get you from $[M+H]^+$ to y_{n-1} (either X or Y)
7. Once the y-series is identified determine the b-series ions (b + y) = $[M+H]^+ + 1$
8. If you know you have a tryptic peptide look for y_1 at 147 m/z (K) or 175 m/z (R)
9. Using y_1 calculate and identify b_{n-1} and then follow the b-series as describe for the y-series

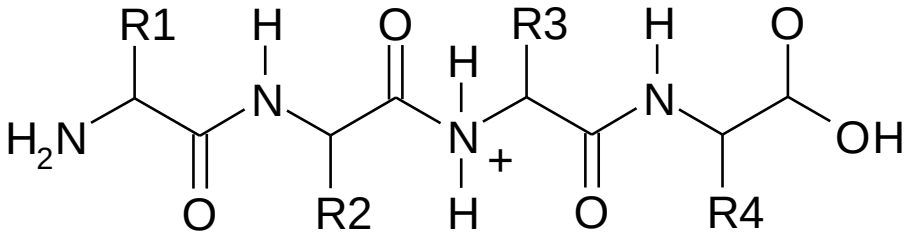
Biomolecular Mass Spectrometry

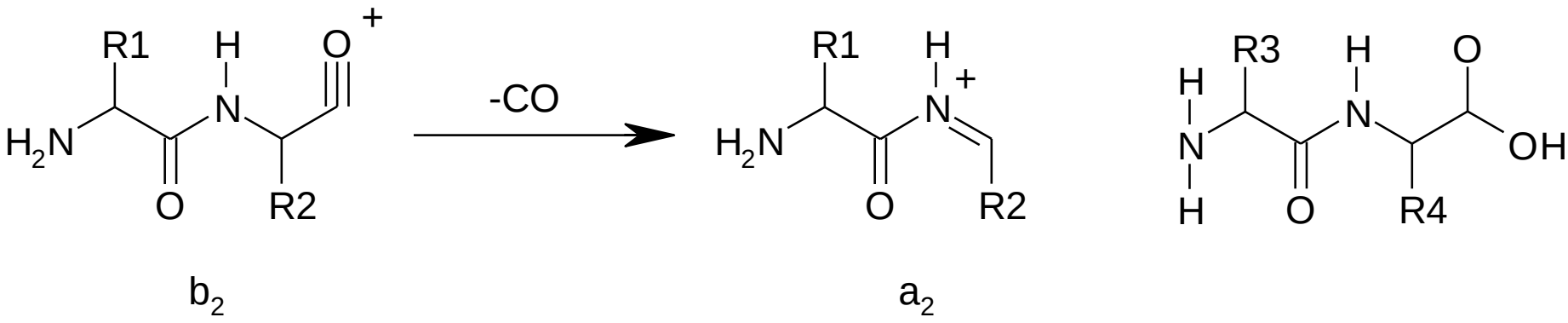


Biomolecular Mass Spectrometry



Biomolecular Mass Spectrometry





b_2 a_2

Biomolecular Mass Spectrometry

Masses of b_2 ions for all possible unmodified dipeptides

	G	A	S	P	V	T	C	I/L	N	D	K/Q	E	M	H	F	R	Y	W
G	**115**																	
A	**129**	**143**																
S	**145**	159	175															
P	**155**	**169**	185	195														
V	**157**	171	187	**197**	199													
T	159	173	189	199	201	203												
C	**161**	175	**191**	201	203	205	**207**											
I/L	171	185	201	**211**	213	215	217	227										
N	**172**	186	**202**	**212**	214	216	218	228	229									
D	173	187	203	213	215	217	219	229	230	**231**								
K/Q	186	**200**	216	226	228	230	**232**	**242**	243	244	**257**							
E	187	201	217	227	229	231	233	243	244	245	**258**	**259**						
M	189	203	219	229	231	233	**235**	245	**246**	247	260	**261**	**263**					
H	195	**209**	225	235	**237**	**239**	**241**	251	**252**	**253**	**266**	267	**269**	**275**				
F	205	219	235	245	247	**249**	251	261	**262**	263	**276**	277	279	285	295			
R	214	228	244	**254**	**256**	258	260	**270**	**271**	**272**	285	286	288	**294**	**304**	**313**		
Y	**221**	235	251	261	263	**265**	267	277	**278**	279	**292**	**293**	295	301	**311**	**320**	**327**	
W	244	258	**274**	**284**	286	288	**290**	**300**	**301**	**302**	**315**	**316**	**318**	**324**	**334**	**343**	**350**	**373**

Biomolecular Mass Spectrometry

Amino Acid	Abbreviation	Residue Mass (Mono, Da)	Amino Acid Mass (Mono, Da)	Immonium Ion Mass (Nom., Da)	Side-chain Mass (Nom., Da)	Neutral Loss (Nom., Da)
Alanine	A	71.0371	89.04767	44	15	--
Arginine	R	156.1011	174.1117	129	100	17
Asparagine	N	114.0429	132.0535	87	58	17
Aspartic Acid	D	115.0269	133.0375	88	59	18
Cysteine	C	103.0092	121.0198	76	47	34(92)
Glutamic Acid	E	129.0426	147.0532	102	73	18
Gluatamine	Q	128.0586	146.0692	101	72	17
Glycine	G	57.0215	75.03207	30	--	--
Histidine	H	137.0589	155.0695	110	81	--
Isoleucine	I	113.0841	131.0947	86	57	--
Leucine	L	113.0841	131.0947	86	57	--
Lysine	K	128.0950	146.1056	101	72	17
Methionine	M	131.0405	149.0511	104	75	48
Phenylalanine	F	147.0684	165.079	120	91	--
Proline	P	97.0528	115.0634	70	--	--
Serine	S	87.0320	105.0426	60	31	18
Threonine	T	101.0477	119.0583	74	45	18
Tryptophan	W	186.0793	204.0899	159	130	--
Tyrosine	Y	163.0633	181.0739	136	107	--
Valine	V	99.0684	117.079	72	43	--

Biomolecular Mass Spectrometry

- Example 1 (Synder problem 9.3)

- Tryptic peptide with $[M+2H]2^{+}$ = 786.0 m/z
- Prominent m/z include, 175.1, 246.0, 333.2, 480.3, 627.3, 684.3, 813.3, 942.4, 1056.5, 1171.5, 1285.5

- What is the sequence?

Biomolecular Mass Spectrometry

- Example 2 (Snyder problem 9.1)

- Tryptic peptide with $[M+2H]2^{+}$ = 697.4 m/z
- Prominent m/z include, 147.2, 195.2, 260.4, 294.3, 373.5, 395.4, 444.6, 494.6, 501.6, 607.7, 614.8, 697.4, 708.8, 779.9, 787.0, 893.1, 900.2, 950.1, 999.3, 1021.2, 1100.4, 1134.4, 1199.5, 1256.6, 1393.7
- What is the sequence?

Biomolecular Mass Spectrometry

- Example 3 (Dass, example 8.5)