# Peptide ion fragmentation in mass spectrometry

**Stephen Barnes**
sbarnes@uab.edu
**4-7117**



# Where we are so far

- We've discussed the nature of the problem, how we might attack it and what we believe in
- Matt Renfrow has told you
  - how (remarkably) we get peptide and protein molecular ions into gas phase
  - the importance of isotopes in mass spectrometry
  - how we measure the *m/z* values of the ions
- We also talked about:
  - how to measure the molecular weight of a protein
  - How to fragment a protein into smaller pieces to get a peptide mass fingerprint and hence "identify" it

# Lecture goals

- Value of fragmentation in determining structure
- How peptides fragment
  - Interpreting the tandem mass spectrum
- Automating identification of peptides from their fragment ions
  - pros and cons
- Controlling fragmentation
  - Choice of ionization and fragmentation methods

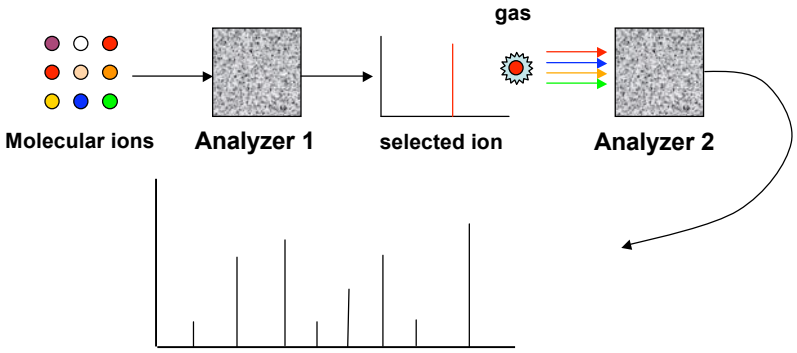
S. Barnes 1/16/09

# Why ion fragmentation provides useful information

- Compounds can have the same empirical formula, i.e., the same molecular weight or *m/z*, but be different chemically.
- Breaking them into parts (fragmenting them) helps to identify what they are.
- Each of the following peptides gives rise to the same $[M+2H]^{2+}$ ion
  - $NH_2$VFAQHLK-COOH  $NH_2$VAFQHLK-COOH
  - $NH_2$VFQHALK-COOH  $NH_2$VHLAFQK-COOH
- In proteomics we want to distinguish these peptides

S. Barnes 1/16/09

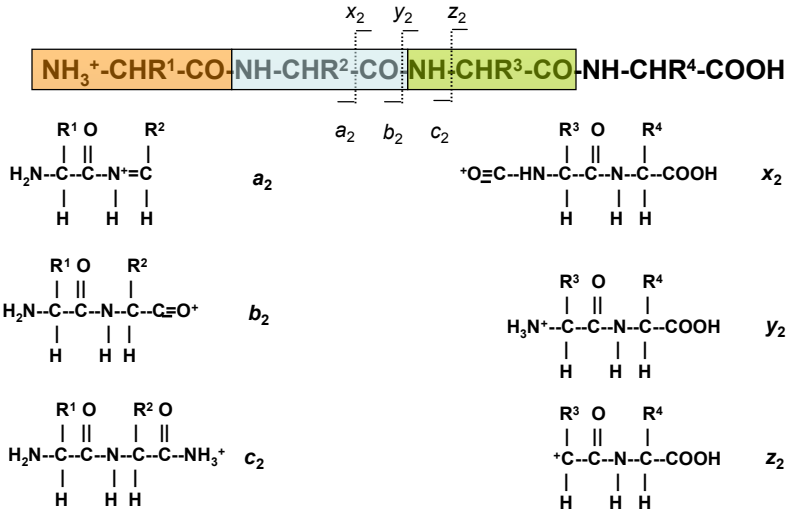
# What is MS-MS (tandem mass spectrometry)?




**Collision-induced dissociation**

S. Barnes 1/16/09

# Fragmenting a peptide




Adapted from http://www.matrixscience.com/help/fragmentation_help.html

S. Barnes 1/16/09

# Calculating expected *b*- and *y*-ion fragments

| | | | |
|---|---|---|---|
| Alanine | 71.037 | Leucine | 113.084 |
| Arginine | 156.101 | Lysine | 128.094 |
| Asparagine | 114.043 | Methionine | 131.040 |
| Aspartic acid | 115.027 | Phenylalanine | 147.068 |
| Cysteine | 103.009 | Proline | 97.053 |
| Glutamic acid | 129.043 | Serine | 87.032 |
| Glutamine | 128.058 | Threonine | 101.048 |
| Glycine | 57.021 | Tryptophan | 186.079 |
| Histidine | 137.059 | Tyrosine | 163.063 |
| Isoleucine | 113.084 | Valine | 99.068 |

$b_n$ = [residue masses + 1] - these come from the N-terminus
$y_n$ = [residue masses + $H_2O$ + 1] = these come from the C-terminus

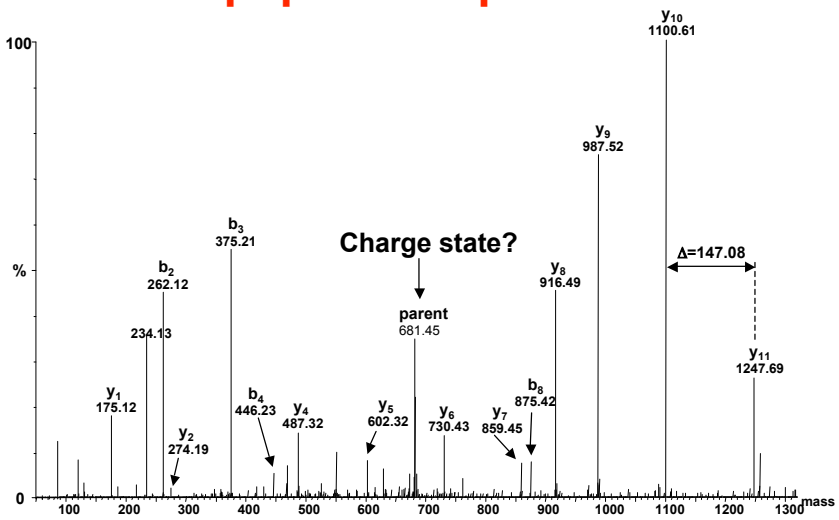
**ADGTWLEVR**

$b_3$ = ADG, 71.04+115.03+57.02+1= 244.09

$y_3$ = EVR, 129.04+99.07+156.10+18+1= 403.21

**S. Barnes 1/16/09**

# Identification of daughter ions and peptide sequence




**S. Barnes 1/16/09**

# What's in a peptide MSMS spectrum?

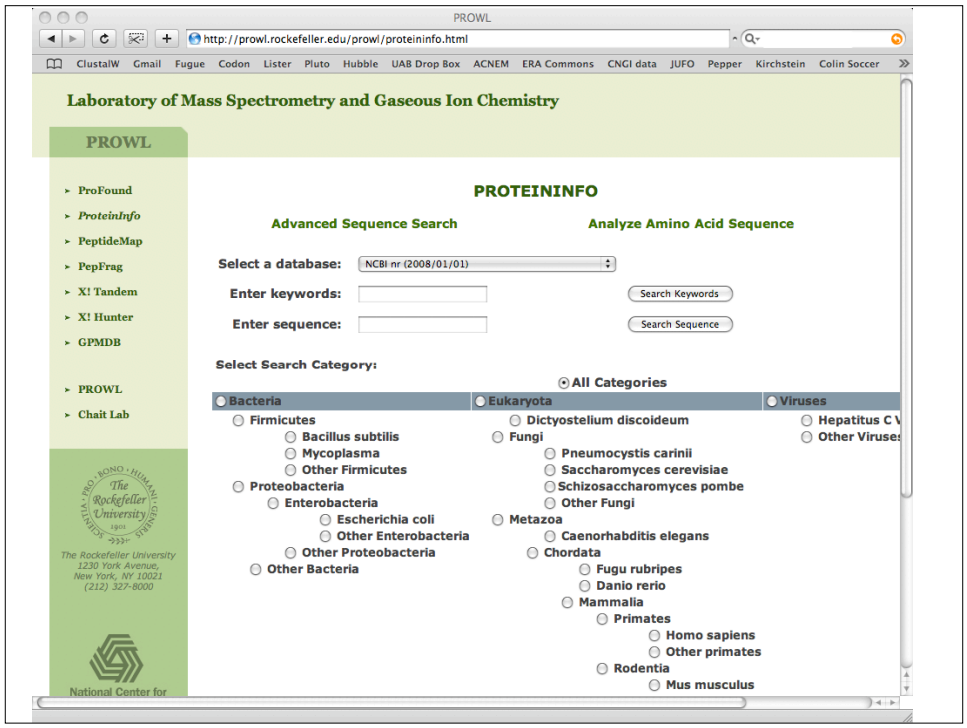
- In most cases, some, but rarely all, of the theoretic *b*- and *y*-ions are observed
- Besides *b*- and *y*-ions, other types of fragmentation can occur to form $a_n$ and $x_n$ ions, as well as also losing CO, $NH_3$ and $H_2O$ groups
- Internal cleavage reactions can occur at acidic (Asp - Glu) residue sites

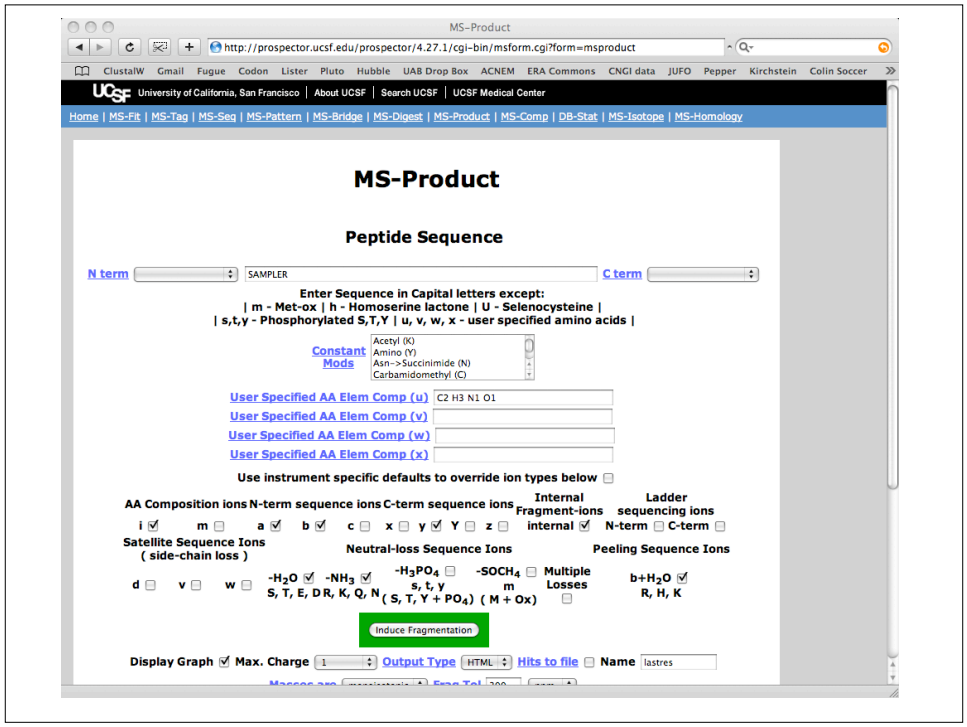
S. Barnes 1/16/09

# Identifying a peptide by de novo sequencing

- Take the partial sequence that can be identified manually and submit it to PROWL (http://prowl.rockefeller.edu/) - click on PROTEININFO and enter sequence - select all species
- Use suggested sequences to fill in the gaps and then check all theoretical ions using MS-Product at http://prospector.ucsf.edu/prospector/4.27.1/cgi-bin/msform.cgi?form=msproduct

S. Barnes 1/16/09


PROWL
http://prowl.rockefeller.edu/prowl/proteininfo.html
ClustalW Gmail Fugue Codon Lister Pluto Hubble UAB Drop Box ACNEM ERA Commons CNCI data JUFO Pepper Kirchstein Colin Soccer
Laboratory of Mass Spectrometry and Gaseous Ion Chemistry
PROWL
ProFound
ProteinInfo
PeptideMap
PepFrag
X! Tandem
X! Hunter
GPMDB
PROWL
Chait Lab
The Rockefeller University
1230 York Avenue,
New York, NY 10021
(212) 327-8000
National Center for
PROTEININFO
Advanced Sequence Search
Analyze Amino Acid Sequence
Select a database:
NCBI nr (2008/01/01)
Enter keywords:
Search Keywords
Enter sequence:
Search Sequence
Select Search Category:
All Categories
Bacteria
Eukaryota
Viruses
Firmicutes
Bacillus subtilis
Mycoplasma
Other Firmicutes
Proteobacteria
Enterobacteria
Escherichia coli
Other Enterobacteria
Other Proteobacteria
Other Bacteria
Dictyostelium discoideum
Fungi
Pneumocystis carinii
Saccharomyces cerevisiae
Schizosaccharomyces pombe
Other Fungi
Metazoa
Caenorhabditis elegans
Chordata
Fugu rubripes
Danio rerio
Mammalia
Primates
Homo sapiens
Other primates
Rodentia
Mus musculus
Hepatitus C
Other Viruses

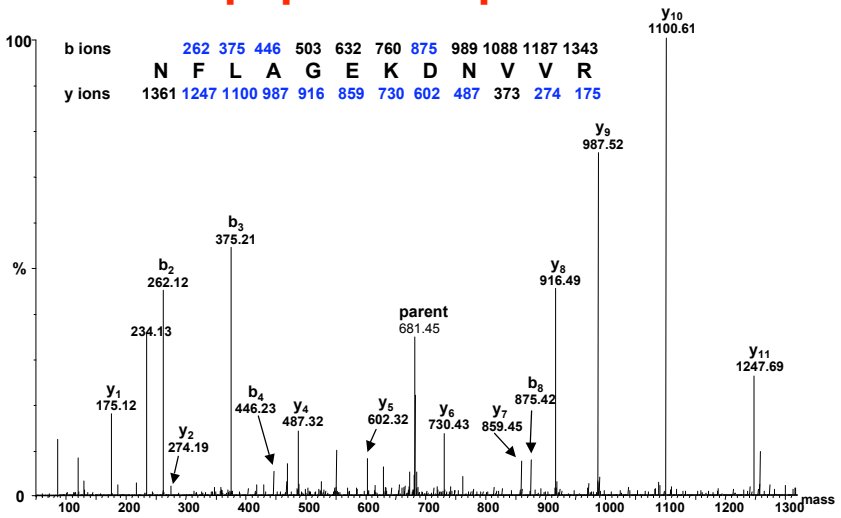

MS-Product
http://prospector.ucsf.edu/prospector/4.27.1/cgi-bin/msform.cgi?form=msproduct
ClustalW Gmail Fugue Codon Lister Pluto Hubble UAB Drop Box ACNEM ERA Commons CNCI data JUFO Pepper Kirchstein Colin Soccer
UCSF University of California, San Francisco | About UCSF | Search UCSF | UCSF Medical Center
Home | MS-Fit | MS-Tag | MS-Seq | MS-Pattern | MS-Bridge | MS-Digest | MS-Product | MS-Comp | DB-Stat | MS-Isotope | MS-Homology
MS-Product
Peptide Sequence
N term
SAMPLER
C term
Enter Sequence in Capital letters except:
| m - Met-ox | h - Homoserine lactone | U - Selenocysteine |
| s,t,y - Phosphorylated S,T,Y | u, v, w, x - user specified amino acids |
Constant Mods
Acetyl (K)
Amino (Y)
Asn->Succinimide (N)
Carbamidomethyl (C)
User Specified AA Elem Comp (u)
C2 H3 N1 O1
User Specified AA Elem Comp (v)
User Specified AA Elem Comp (w)
User Specified AA Elem Comp (x)
Use instrument specific defaults to override ion types below
AA Composition ions
N-term sequence ions
C-term sequence ions
Internal Fragment-ions
Ladder sequencing ions
i m a b c x y Y z
internal N-term C-term
Satellite Sequence Ions ( side-chain loss )
Neutral-loss Sequence Ions
Peeling Sequence Ions
d v w
-H2O S, T, E, D
-NH3 R, K, Q, N
-H3PO4 s, t, y ( S, T, Y + PO4)
-SOCH4 m ( M + Ox)
Multiple Losses
b+H2O R, H, K
Induce Fragmentation
Display Graph
Max. Charge
Output Type
HTML
Hits to file
Name
lastres

# Other ions observed in CID peptide fragmentation

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Immonium and Related Ions | | | | | | | | | | | | |
| | 87.06 | 120.08 | 86.10 | --- | --- | 102.05 | 84.08<br>101.11<br>129.10 | 88.04 | 87.06 | 72.08 | 72.08 | 70.07<br>87.09<br>100.09<br>112.09 |
| N-terminal ions | | | | | | | | | | | | |
| a-$NH_3$ ions | --- | 217.10 | 330.18 | 401.22 | 458.24 | 587.28 | 715.38 | 830.40 | 944.45 | 1043.52 | 1142.58 | --- |
| a ions | --- | 234.12 | 347.21 | 418.24 | 475.27 | 604.31 | 732.40 | 847.43 | 961.47 | 1060.54 | 1159.61 | --- |
| b-$NH_3$ ions | --- | 245.09 | 358.18 | 429.21 | 486.23 | 615.28 | 743.37 | 858.40 | 972.44 | 1071.51 | 1170.58 | --- |
| b-$H_2O$ ions | --- | --- | --- | --- | --- | 614.29 | 742.39 | 857.42 | 971.46 | 1070.53 | 1169.59 | --- |
| b ions | --- | 262.12 | 375.20 | 446.24 | 503.26 | 632.30 | 760.40 | 875.43 | 989.47 | 1088.54 | 1187.61 | --- |
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| H - | N | F | L | A | G | E | K | D | N | V | V | R |
| y ions | --- | 1247.67 | 1100.61 | 987.52 | 916.48 | 859.46 | 730.42 | 602.33 | 487.30 | 373.26 | 274.19 | 175.12 |
| y-$NH_3$ ions | --- | 1230.65 | 1083.58 | 970.50 | 899.46 | 842.44 | 713.39 | 585.30 | 470.27 | 356.23 | 257.16 | 158.09 |
| y-$H_2O$ ions | --- | 1229.66 | 1082.60 | 969.51 | 898.47 | 841.45 | 712.41 | 584.32- | --- | --- | -- | --- |

S. Barnes 1/16/09

# Identification of daughter ions and peptide sequence




S. Barnes 1/16/09

# Towards automated MSMS sequencing

- The 2D-LC-ESI-MSMS method (MuDPIT) generates 50,000+ MSMS spectra for each sample
- If it takes 15 min to hand interpret one MS-MS spectrum, then it would take 12,500 hours to complete the analysis. For someone working 8 hours/day and a five-day week, this would be about 6 years!
- Using SEQUEST and MASCOT, methods were developed to use computer-driven approaches to analyze MSMS data

S. Barnes 1/16/09

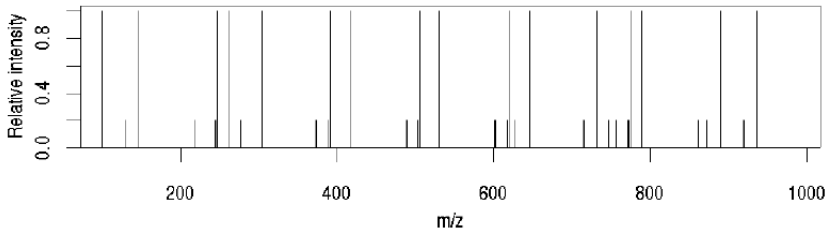
# Issues in MS-MS experiment

- At any one moment, several peptides may be co-eluting
- Data-dependent operation:
  - The most intense peptide molecular ion is selected first (must exceed an initial threshold value)
  - A 2-3 Da window is used (to maximize the signal)
  - The ion must be in $2^+$ or $3^+$ state
  - Since the ion trap scan of the fragment ions takes ~ 1 sec, only the most intense ions will be measured
  - However, can use an exclusion list on a subsequent run to study minor ions


S. Barnes 1/16/09

# The SEQUEST approach

- Each observed MSMS spectrum has a corresponding molecular ion $[M+nH]^{n+}$. For ion trap data, ions are selected from the known or virtual proteome that are within 1 Da. These are then "fragmented" *in silico* to produce *b*- and *y*-ions and less abundant fragment ions.




- The cross correlations of the observed MSMS spectrum to each of the virtual MSMS spectra are calculated. The peptides are scored and the one having the highest score is deemed to be "identified".

S. Barnes 1/16/09

# What has SEQUEST provided to proteomics?

- Initially, it seemed an awful lot! Typically, the "identified" proteins covered most of known biochemistry, so they satisfied everybody
- But the method obviously has limitations. There is redundancy - each protein yields multiple peptides
- The number of unique proteins is much less than the observed peptides
- Critically, it was missing controls

S. Barnes 1/16/09

# SEQUEST sequencing

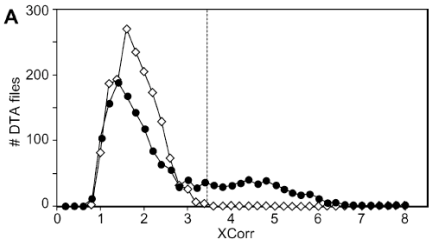
- Use of SEQUEST requires considerable computing power - if there are 500 possible peptides to compare, then examination of 50,000+ spectra would require 25 million correlations
- Data analysis is typically carried out using computer clusters to accelerate the analysis

S. Barnes 1/16/09

# More haste, less speed?

- Post analysis, the masses of the peptides triggering MS-MS are used to create a set of virtual peptides with masses within ± 1 Da
- Predicted MS-MS are compared to the observed and the best fit is reported as a hit
- The abundance of these hits are plotted in the figure as closed circles

Resing et al. (2004)




However, if the sequences of the peptides within ± 1 Da are reversed *in silico* and their predicted MS-MS compared to the observed spectra, a similar histogram is obtained (open circles), but without the right side tail

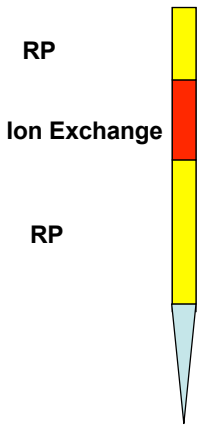
A forced fit to a set of data will always come up with a match, but not necessarily the truth

# How to improve MUDPIT

- Reproducible column engineering
  - Tandem columns, each built to separate, but high specifications
  - Columns on a chip
- More careful selection of the parent ion
  - Accurate measurement of the peptide's mass will eliminate many false peptides
  - Accurate measurement of peptide fragments' masses
- Greater stringency in assessing score cutoff

S. Barnes 1/16/09

# Engineering of a MuDPIT column




Packing right to end of the tip - no dead volume


Wash with 5% MeCN-0.1% acetic acid - elute with 80% MeCN-0.1% formic acid

Wash with 5% MeCN-0.1% acetic acid - elute with 0-500 mM $(NH_4)_2CO_3$

Wash with 5% MeCN-0.1% acetic acid for 5 min - elute with 5-64% MeCN-0.02% HFBA

Problem is that it is very difficult to reproducibly pack these columns - typically only used once

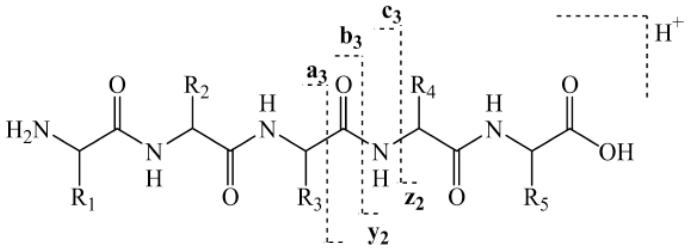
S. Barnes 1/16/09

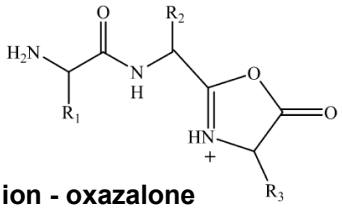
# Mass accuracy limits the ions to consider

| | Theoretical Mass | Delta [ppm] | Delta [mmu] | RDB | Composition |
|---|---|---|---|---|---|
| 1 ppm (4) | 516.76671 | 0.0 | 0.0 | 21.0 | $C_{49}$ $H_{71}$ $O_{12}$ $N_{13}$ |
| | 516.76647 | 0.5 | 0.2 | 15.0 | $C_{49}$ $H_{79}$ $O_{11}$ $N_{9}$ $S_{2}$ |
| | 516.76638 | 0.6 | 0.3 | 12.0 | $C_{41}$ $H_{75}$ $O_{14}$ $N_{15}$ $S_{1}$ |
| | 516.76705 | -0.7 | -0.3 | 11.5 | $C_{43}$ $H_{77}$ $O_{15}$ $N_{12}$ $S_{1}$ |
| 2 ppm (10) | 516.76604 | 1.3 | 0.7 | 16.0 | $C_{48}$ $H_{75}$ $O_{16}$ $N_{9}$ |
| | 516.76738 | -1.3 | -0.7 | 20.5 | $C_{51}$ $H_{73}$ $O_{13}$ $N_{10}$ |
| | 516.76604 | 1.3 | 0.7 | 21.5 | $C_{47}$ $H_{69}$ $O_{11}$ $N_{16}$ |
| | 516.76580 | 1.8 | 0.9 | 15.5 | $C_{47}$ $H_{77}$ $O_{10}$ $N_{12}$ $S_{2}$ |
| | 516.76772 | -2.0 | -1.0 | 16.5 | $C_{44}$ $H_{73}$ $O_{11}$ $N_{16}$ $S_{1}$ |
| | 516.76773 | -2.0 | -1.0 | 11.0 | $C_{45}$ $H_{73}$ $O_{16}$ $N_{8}$ $S_{1}$ |
| 5 ppm (23) | 516.76805 | -2.6 | -1.3 | 25.5 | $C_{52}$ $H_{69}$ $O_{9}$ $N_{14}$ |
| | 516.76537 | 2.6 | 1.3 | 16.5 | $C_{46}$ $H_{73}$ $O_{15}$ $N_{12}$ |
| | 516.76807 | -2.6 | -1.4 | 7.0 | $C_{38}$ $H_{79}$ $O_{14}$ $N_{15}$ $S_{2}$ |
| | 516.76513 | 3.0 | 1.6 | 10.5 | $C_{46}$ $H_{81}$ $O_{14}$ $N_{8}$ $S_{2}$ |
| | 516.76513 | 3.1 | 1.6 | 16.0 | $C_{45}$ $H_{75}$ $O_{9}$ $N_{15}$ $S_{2}$ |
| | 516.76839 | -3.3 | -1.7 | 16.0 | $C_{46}$ $H_{75}$ $O_{12}$ $N_{13}$ $S_{1}$ |
| | 516.76479 | 3.7 | 1.9 | 20.0 | $C_{52}$ $H_{75}$ $O_{11}$ $N_{9}$ $S_{1}$ |
| | 516.76872 | -3.9 | -2.0 | 25.0 | $C_{54}$ $H_{71}$ $O_{10}$ $N_{11}$ |
| | 516.76470 | 3.9 | 2.0 | 17.0 | $C_{44}$ $H_{71}$ $O_{14}$ $N_{15}$ |
| | 516.76874 | -3.9 | -2.0 | 6.5 | $C_{40}$ $H_{81}$ $O_{15}$ $N_{12}$ $S_{2}$ |
| | 516.76446 | 4.3 | 2.2 | 11.0 | $C_{44}$ $H_{79}$ $O_{13}$ $N_{11}$ $S_{2}$ |
| | 516.76897 | -4.4 | -2.3 | 12.5 | $C_{40}$ $H_{73}$ $O_{16}$ $N_{16}$ |
| | 516.76907 | -4.6 | -2.4 | 15.5 | $C_{48}$ $H_{77}$ $O_{13}$ $N_{10}$ $S_{1}$ |

S. Barnes 1/16/09

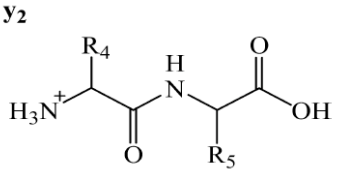
# Let's take a closer look at fragmentation



**$b_3$ ion - oxazalone**

no $b_1$ ion



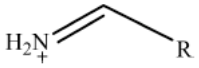

S. Barnes 1/16/09

Wysocki et al. 2005

# Other amino acid fragment ions

*m/z* values of common immonium ions

| Immonium ion (*m/z*) | Amino acid residue | Major (M) or minor (m) peak |
|---|---|---|
| 60.04 | S | M |
| 70.07 | R or P | M |
| 72.08 | V | M |
| 73.00 | R | m |
| 74.06 | T | M |
| 84.08 | K or Q | M |
| 86.1 | I or L | M |
| 87.09 | N or R | M |
| 88.04 | D | M |
| 100.09 | R | m |
| 101.11 | K or Q | M |
| 102.06 | E | M |
| 104.05 | M | M |
| 110.07 | H | M |
| 112.09 | R | M |
| 120.08 | F | M |
| 126.06 | P | M |
| 129.1 | K or Q | m |
| 136.08 | Y | M |
| 138.07 | H | m |
| 159.09 | W | M |




S. Barnes 1/16/09

Wysocki et al. 2005

# Detecting posttranslational modifications (PTMs) by MS

- A key issue is that the energy of ionization or the collisional process should not exceed the dissociational energy of the PTM
- MALDI-TOF MS with a $N_2$ laser causes fragmentation of a nitrated tyrosine residue
  - Use ESI to make the molecular ion
  - Go to another laser wavelength (YAG laser at 355 nm or IR)
- O-glucosyl and phospho groups fragment more easily than the peptide to which they are attached
  - Use electron capture dissociation

S. Barnes 1/16/09

# Types of fragmentation (1)

- Collision-induced dissociation (CID)
  - Also called CAD (collision-activated dissociation)
  - Multiply charged peptide ions are isolated by an *m/z* based filter
  - Selected ions are accelerated into a field of inert gas (He, $N_2$, Ar, Xe) at moderate pressure
  - The energy gained in collision events increases vibrational and stretching modes of the peptide backbone (and anything attached to it!)
  - The increased motion of the energized peptide causes breaks that occur typically at the peptide bond
    - Side chain groups can also be broken, some times more easily than the peptide chain

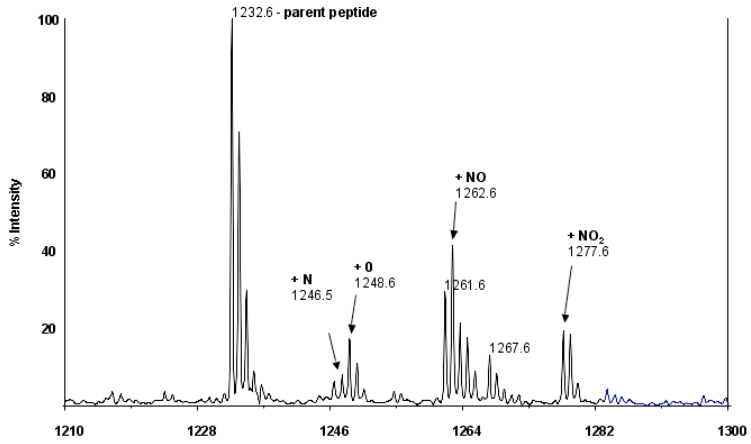
S. Barnes 1/16/09

# Types of fragmentation (2) IRMPD

- InfraRed Multi-Photon Dissociation
  - Used in FT-ICR instruments where a vacuum better than 1 x $10^{-10}$ torr is necessary for the analysis of peptide ions
  - The infra-red radiation is delivered by an IR laser operating at 10.6 microns
  - No gas is involved
  - In this case, the fragmentation is induced in the ICR cell
  - Effects are essentially equivalent to CID

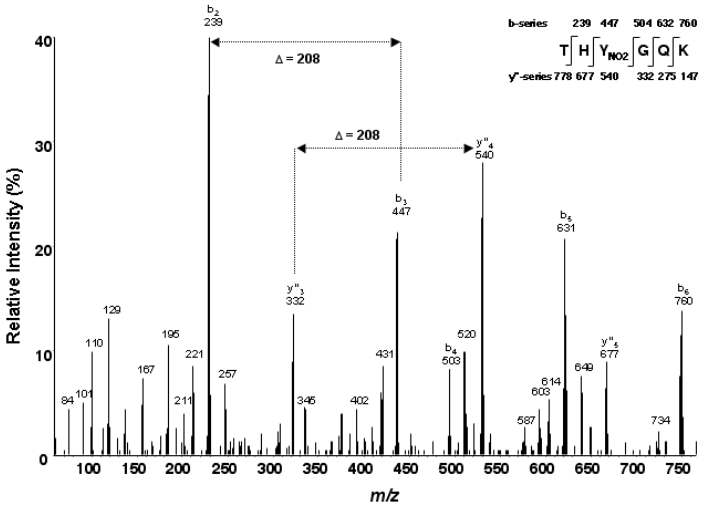
S. Barnes 1/16/09

# Fragmentation of nitrated peptides in MALDI-TOF experiment




S. Barnes 1/16/09

# ESI-tandem MS of a nitrated peptide




S. Barnes 1/16/09

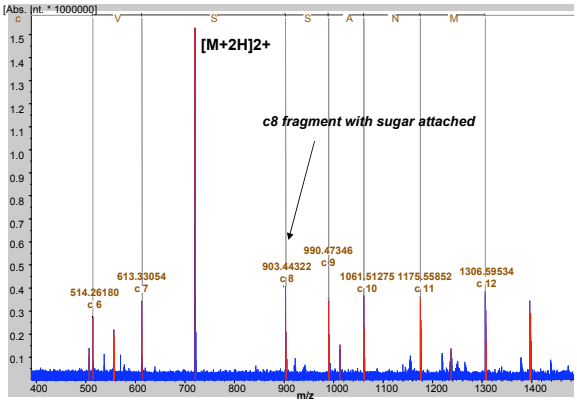
# Types of fragmentation (3) ECD

- **E**lectron **C**apture **D**issociation
  - Used in an ICR cell of an FT-MS instrument
  - Low energy electrons interact with the multiply charged peptide and are absorbed
  - They disturb bonding of the peptide backbone and cleave it without altering the side chain
  - Yields *c*- and *z*-ions
  - MS-MS spectra often very clean
  - In conjunction with an IR laser, ECD can fragment whole proteins (top-down)

S. Barnes 1/16/09

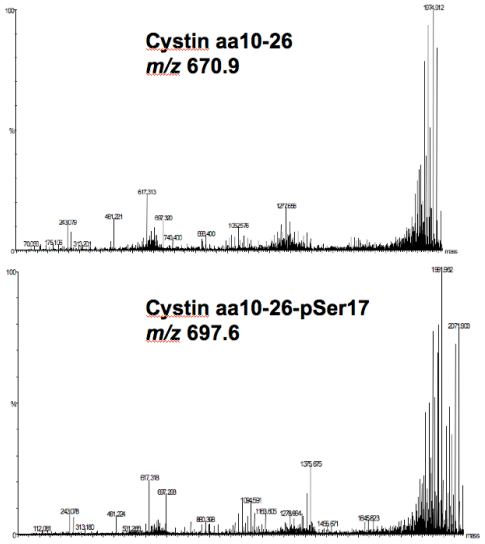
## Sequencing O-GlcNAc peptides by ECD FT-ICR-MS

**Casein kinase II - AGGSTPVSSANMMSG**




S. Barnes 1/16/09

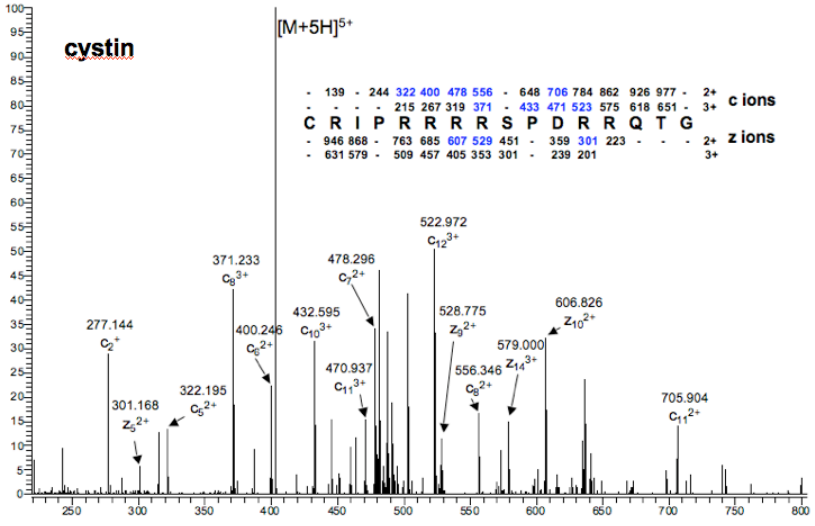
# CID spectra of Arg-rich peptide




Furry spectra due to -$NH_3$ losses from Arg residues

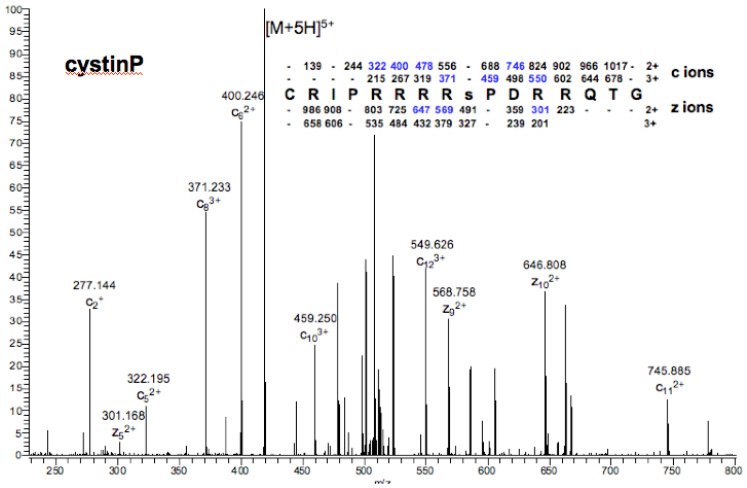
Spectra are uninterpretable

# ECD spectra of cystin peptide




S. Barnes 1/16/09

# ECD spectra of phosphorylated cystin peptide



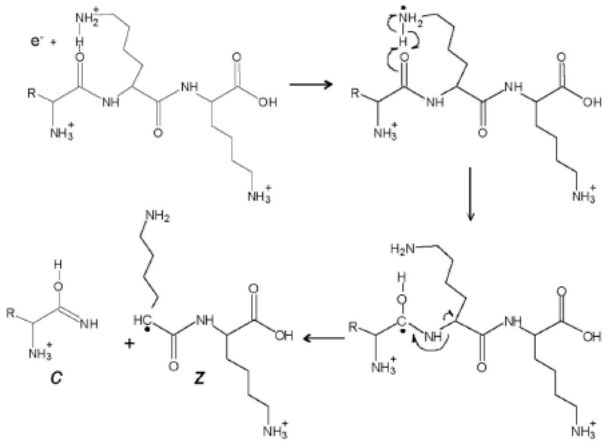

S. Barnes 1/16/09

# Types of fragmentation (4) ETD

- **Electron Transfer Dissociation**
  - **The electron is provided by an electron donating chemical species, a radical anion (azobenzene, fluoranthene) directly infused as a reagent gas, or from their precursors introduced by ESI - 9-anthracenecarboxylic acid, 2-fluoro-5-iodobenzoic acid, and 2-(fluoranthene-8-carbonyl)benzoic acid)**

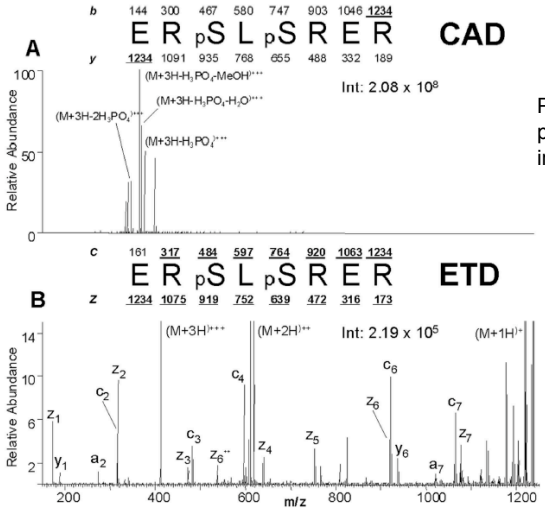
S. Barnes 1/16/09

# Electron transfer dissociation




S. Barnes 1/16/09

# CAD versus ETD for phosphopeptide

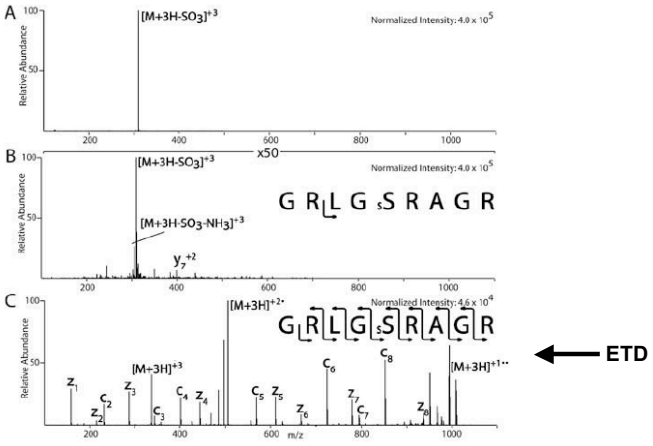



Phosphopeptide loses phosphate, but no sequence information is obtained

Rich series of *c* and *z* ions

S. Barnes 1/16/09

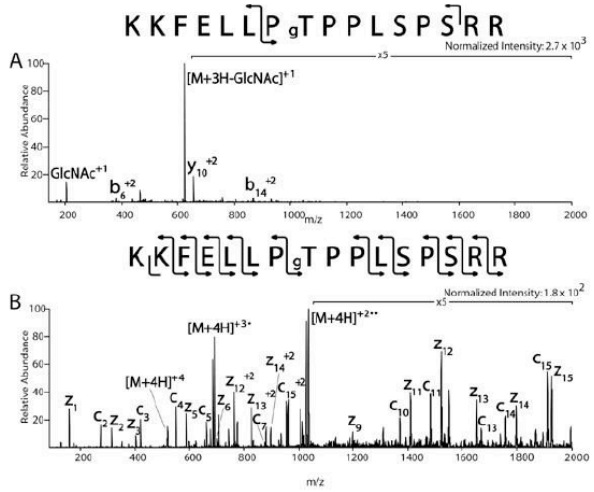
# ETD better for sulfonated peptides




ETD

S. Barnes 1/16/09

# ETD and *O*-glycosylation

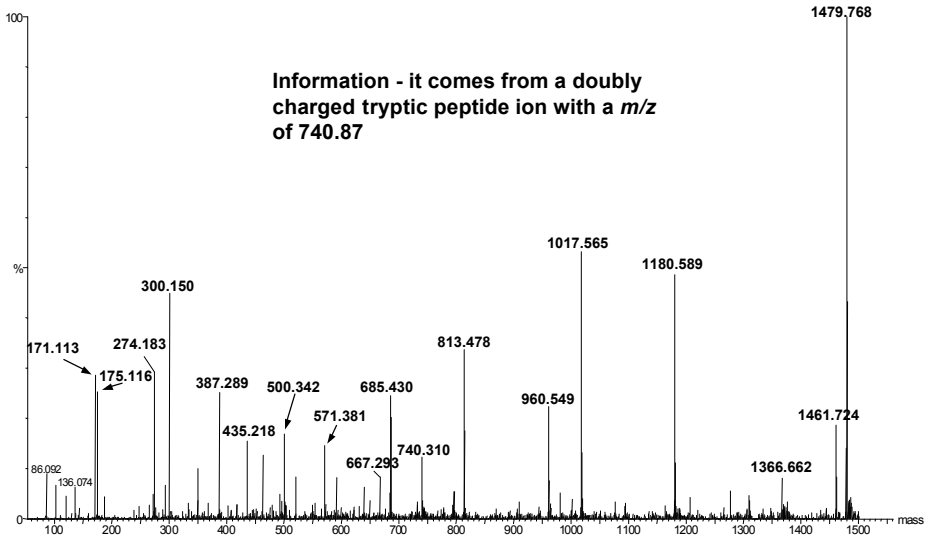



CID

ETD

S. Barnes 1/16/09

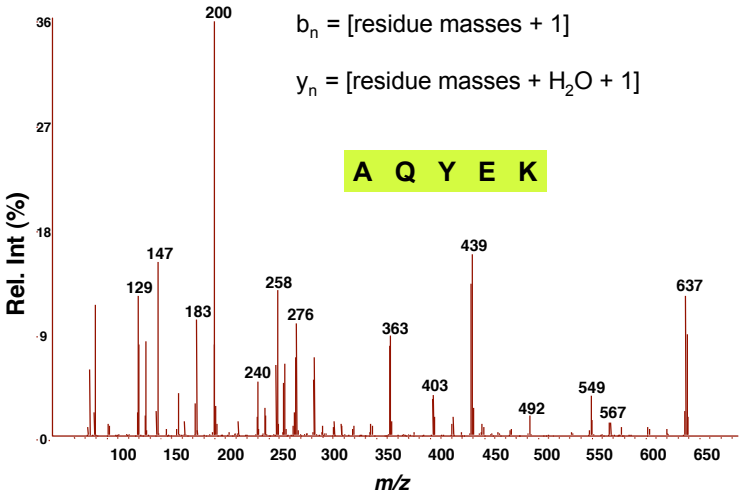
# A tandem mass spectrum - what is the peptide? (Homework due Jan 27)




S. Barnes 1/16/09

# Fragment ions of a small 5-mer peptide

## Homework - write down the masses of the b and y ions




S. Barnes 1/16/09