Mass spectra of peptides and proteins - and LC analysis of proteomes

Stephen Barnes, PhD

4-7117

sbarnes@uab.edu

Overview

- **MALDI-TOF MS**
  - **Protein modifications**
  - **Peptide mass fingerprinting**
- **Electrospray MS**
  - **Analysis of intact proteins**
  - **Molecular weight calculations**
  - **Max Entropy for MW estimation**
- **Peptide analysis**
  - **Purity - ESI-MS is a revelation**
- **Integration of MS with LC and CE**
  - **Multidimensional LC of peptides**
- **Tandem MS**
  - **Identifying peptide amino acid sequences**

Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry (MALDI-TOF MS)

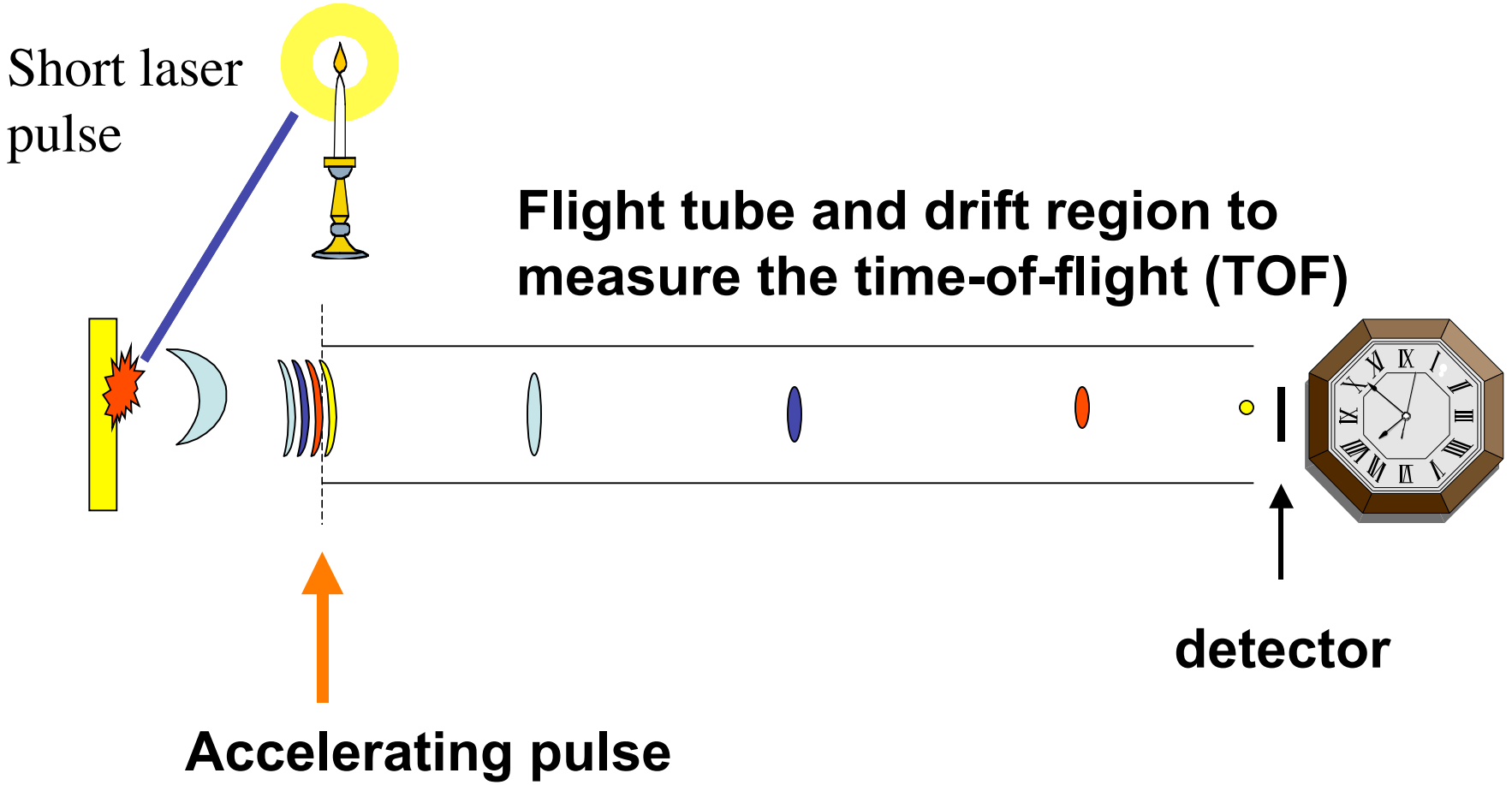
- **Advantages of MALDI-TOF**
  - More tolerant to common buffers than ESI
  - High degree of sensitivity, moderate mass accuracy, and mass resolution
  - High mass compounds, i.e. proteins, PEG…

- **Common Applications of MALDI-TOF**
  - Masses of large proteins and other compounds
  - Enzymatic digestion profiles of proteins to establish their identity
  - Peptide sequencing (TOF-TOF)

Factors from conventional experiments that impact MALDI-TOF analysis

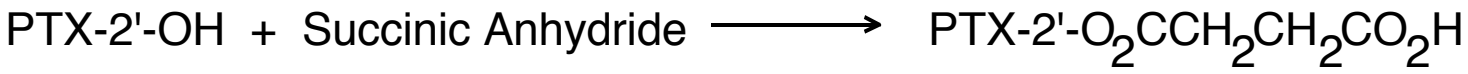
- **Tolerance of buffers/chemicals used in sample preparation**
  - **NaCl up to 150 mM**
  - **Urea up to 2-3 M (carbamoylation can occur!)**
  - **Guanidinium-HCl up to 2 M**
- **Tolerance of detergents**
  - **SDS up to 0.05%**
- **Staining Protocols**
  - **Whole proteins form adducts with Coomassie**
  - **Silver staining modifies selected peptides**

Matrix-Assisted Laser Desorption Ionization (MALDI)



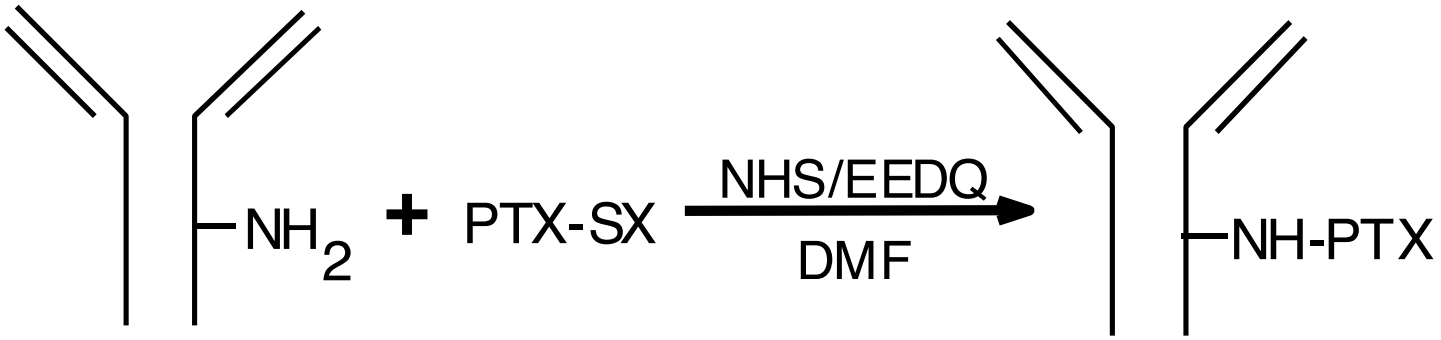
Chemically modifying an antibody

Scheme 1



PTX-SX

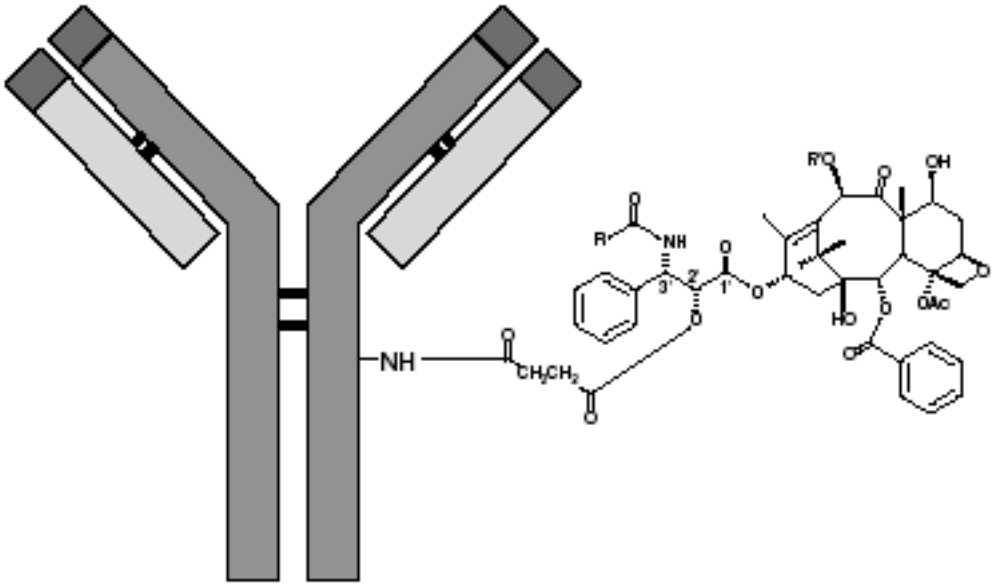
Scheme 2



MAb **PTXMAb**

Ahmad Safavy

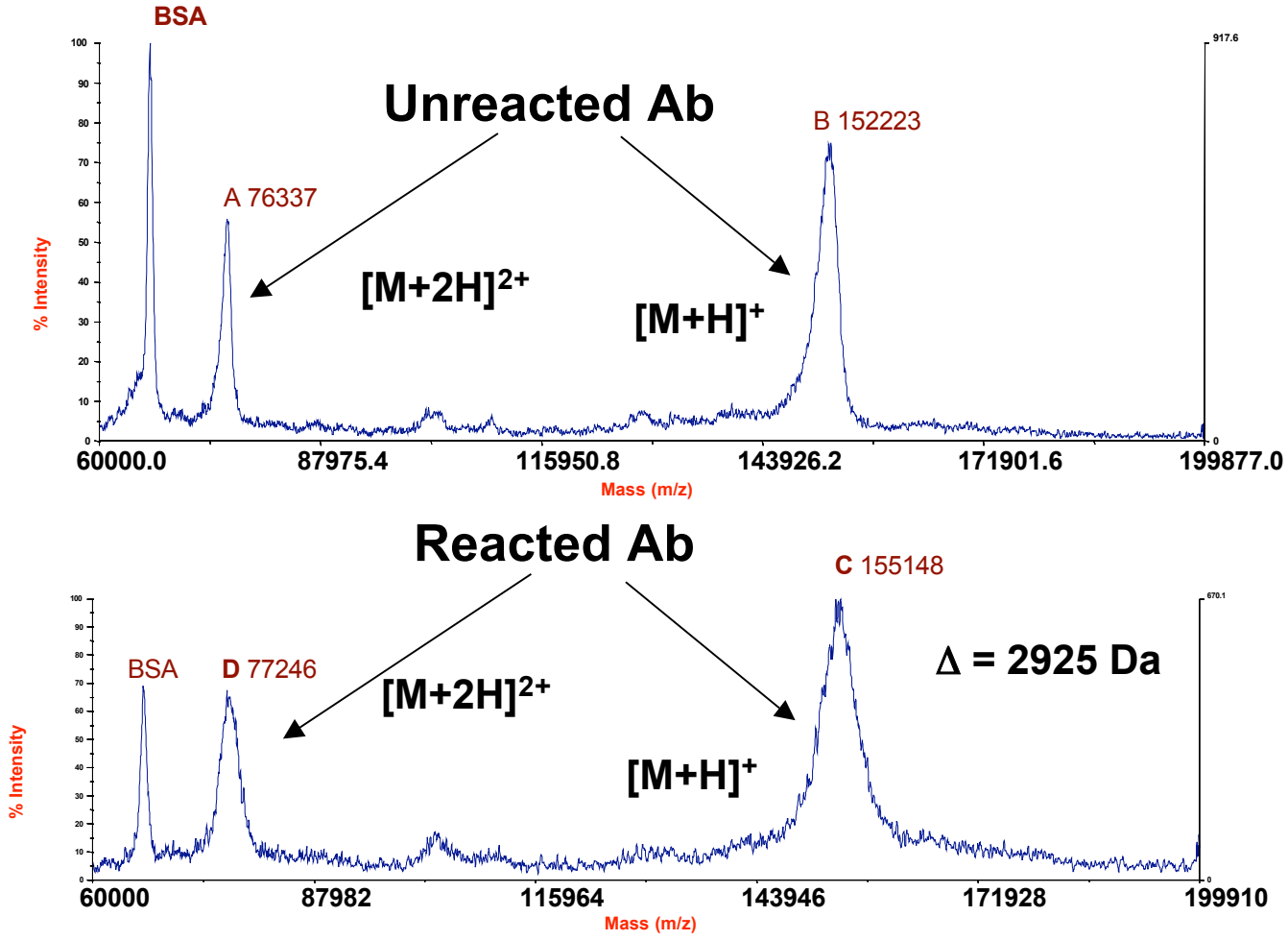
Structure of modified antibody



PTX-MAb

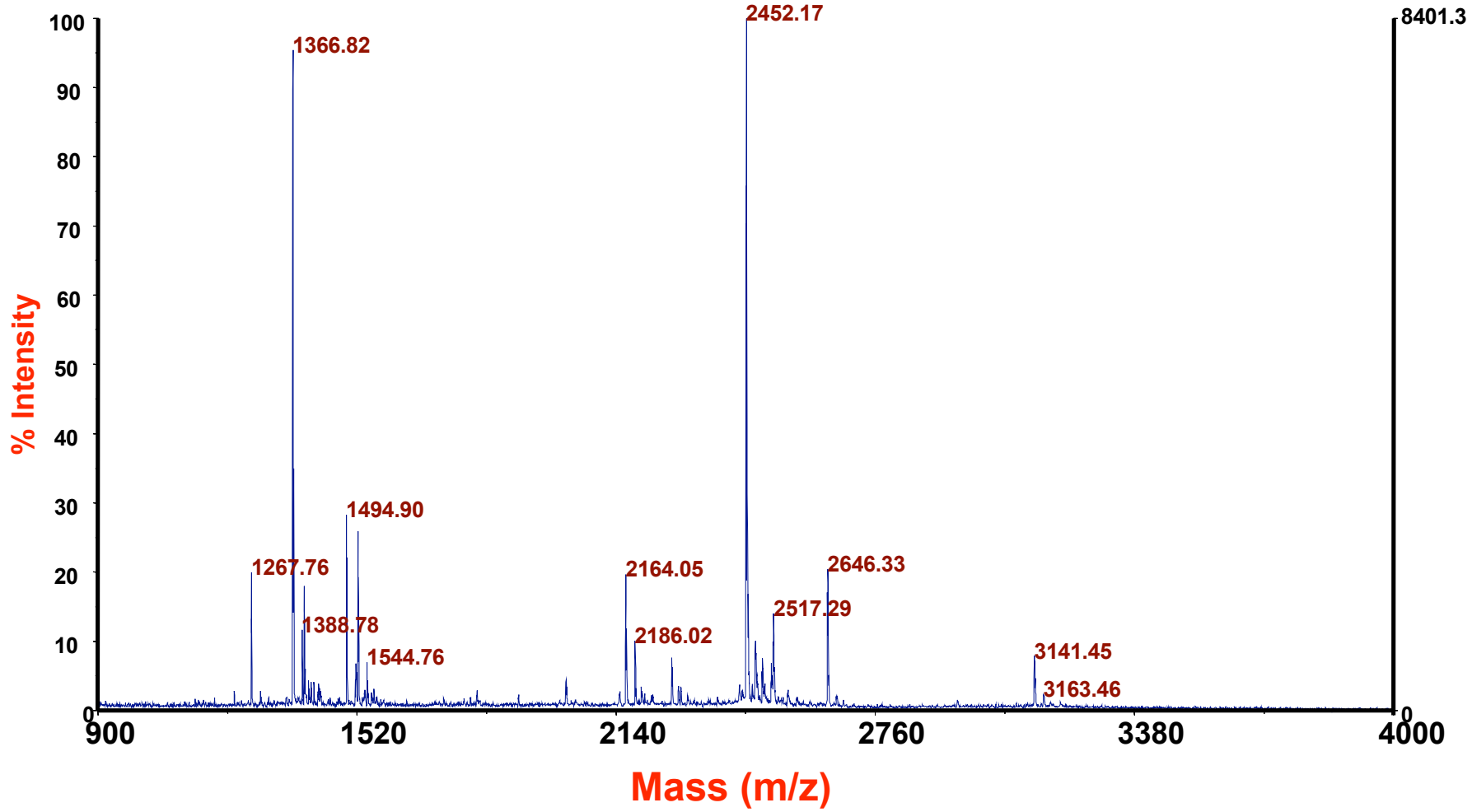
Ahmad Safavy

Modification of an antibody by MALDI-TOF

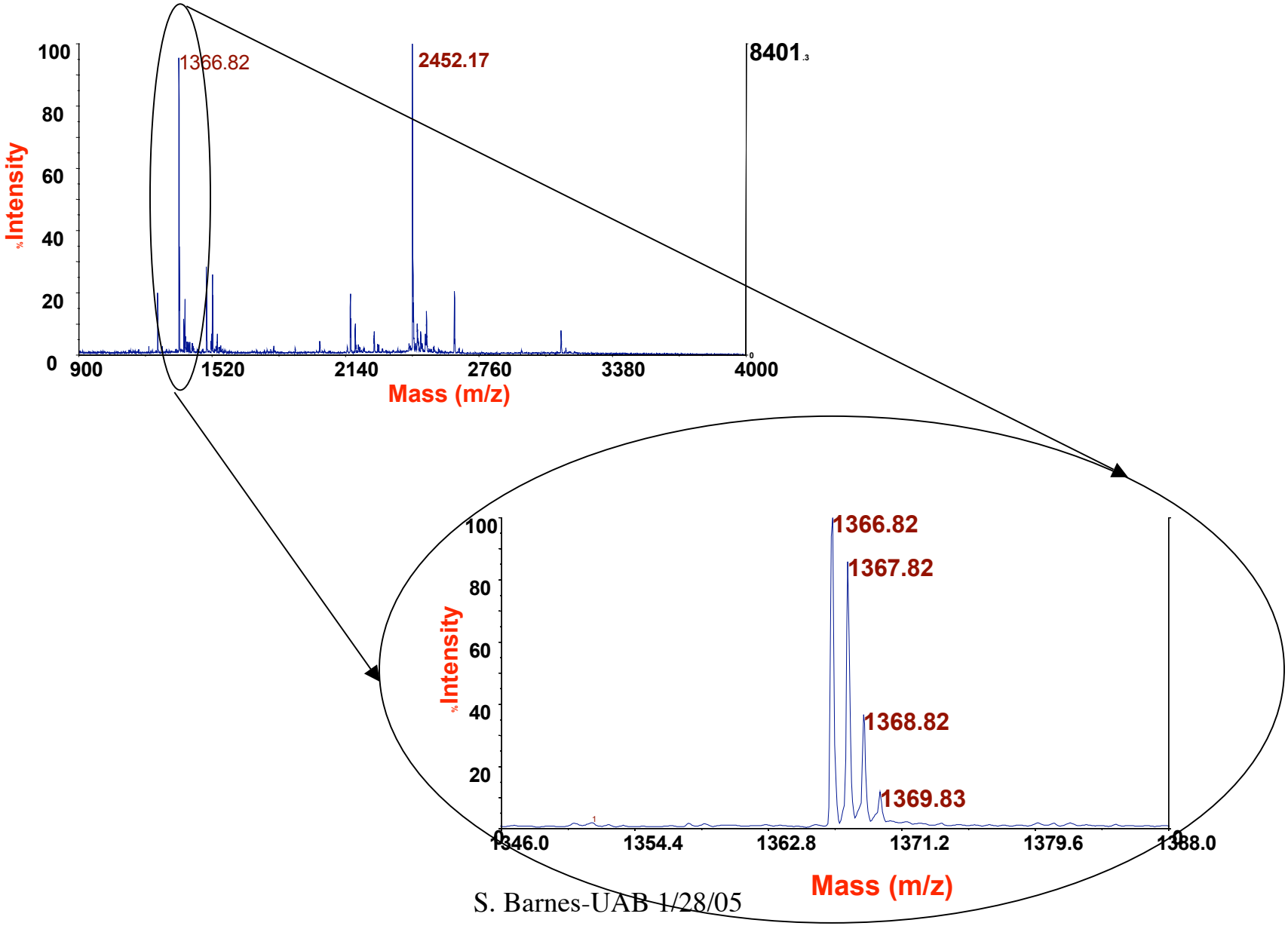


Ahmad Safavy

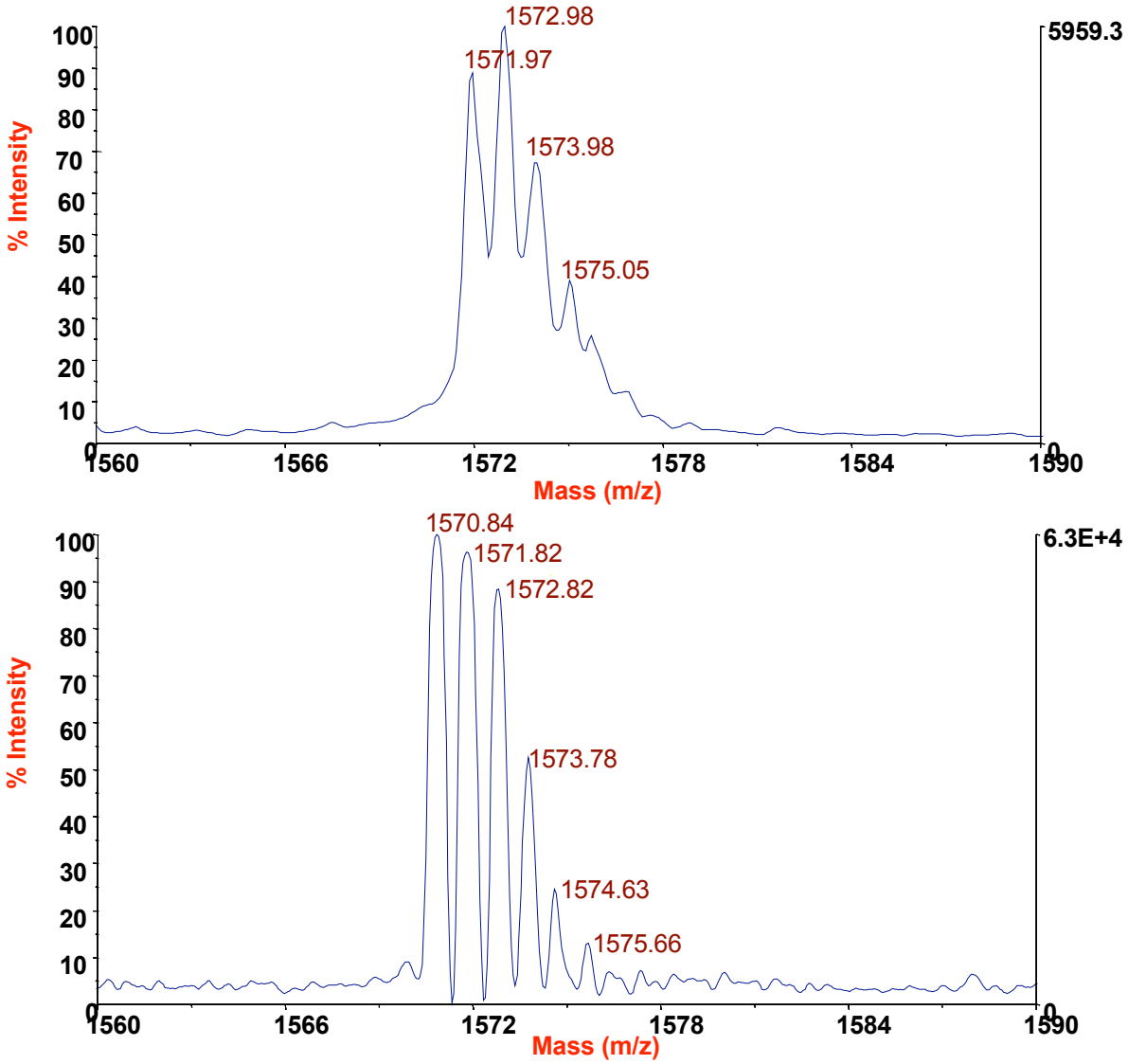
A mass spectrum of peptides from a tryptic digest



Isotope profile of individual peptide ion



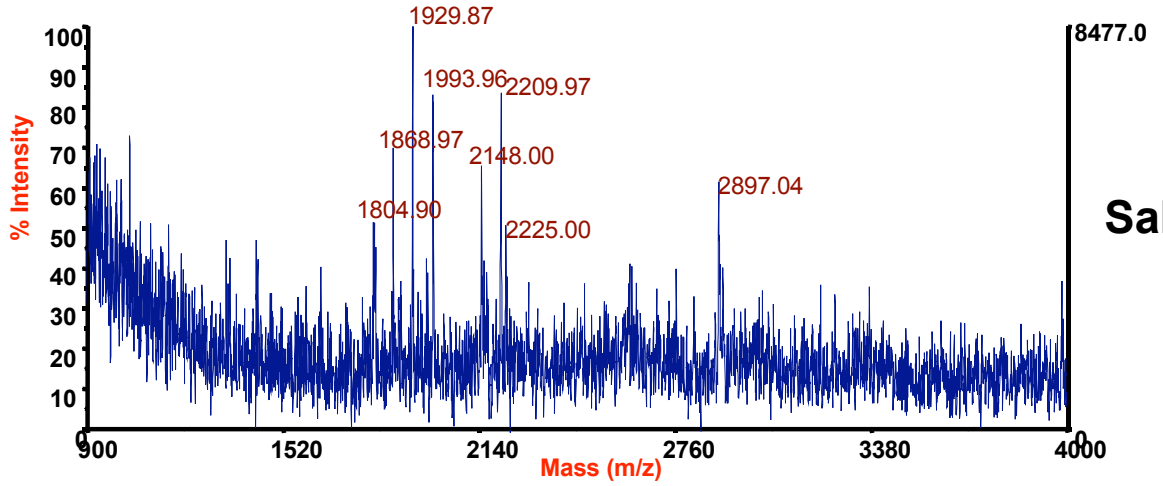
Increased sensitivity in reflector vs. linear mode



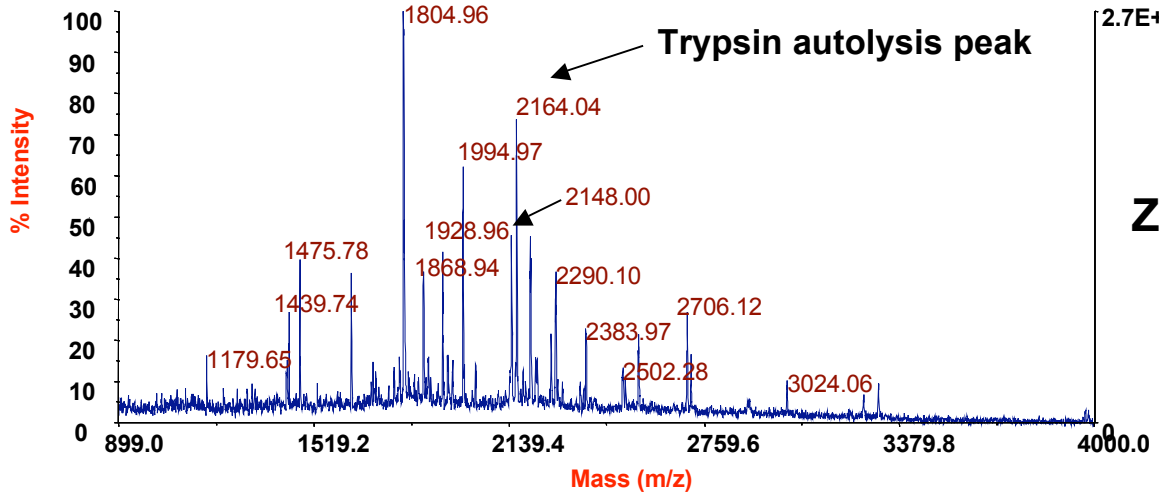
Linear Mode

Reflector Mode

Benefit of removing salt from tryptic digest



Salt Contamination



ZipTip (desalting)

Protein analysis by MALDI 2004

Eppendorf tube

destain

Speed-Vac

trypsin

1:20

Incubate

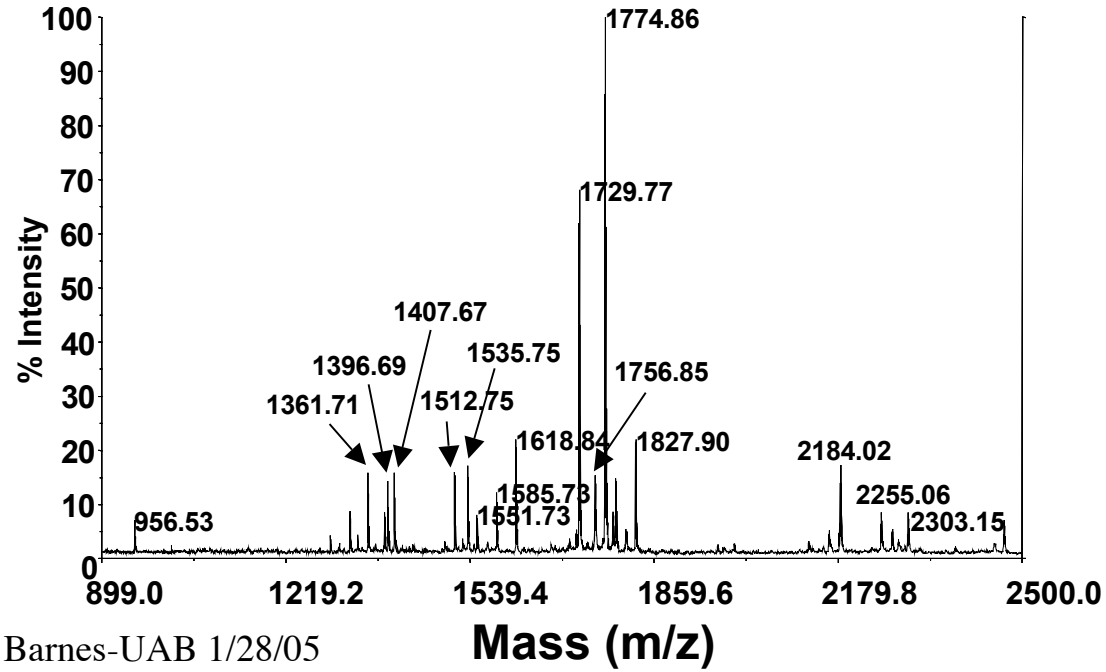
overnight

Water Bath 37°C

Peptide extraction

MALDI plate

Desalting Ziptip



Peptide mass fingerprinting

- **This method has been developed because of the availability of predicted protein sequences from genome sequencing**
- **Proteins do not have to have been previously sequenced - only that the open reading frame in the gene is known - the rest is a virtual exercise in the hands of statisticians, bioinformaticists and computers**

Choice of peptidase

- Analogous to DNA restriction enzymes
- Tryptic peptide fingerprinting may identify several highly related protein candidates (e.g., actins)
- Inspection of the sequences may reveal that there is a difference at one residue that distinguishes between two candidates.
- If for instance it is a glutamate, then use of Glu-C or V8-protease may enable the two proteins to be correctly identified
- INSPECT sequences carefully

Proteolytic enzymes used to hydrolyze proteins

The choice of enzyme largely depends on the nature of the amino acid sequence and the specific issue that is being addressed

- **Trypsin -** ***cleaves at arginine and lysine residues***
- **Chymotrypsin -** ***cleaves hydrophobic residues***
- **Arg-C -** ***cleaves at arginine residues***
- **Glu-C -** ***cleaves at glutamic acid residues***
- **Lys-C -** ***cleaves at lysine residues***
- **V8-protease -** ***cleaves at glutamic acid residues***
- **Pepsin -** ***cleaves randomly, but at acid pH***

See **http://www.abrf.org/JBT/1998/September98/sep98m_r.html**

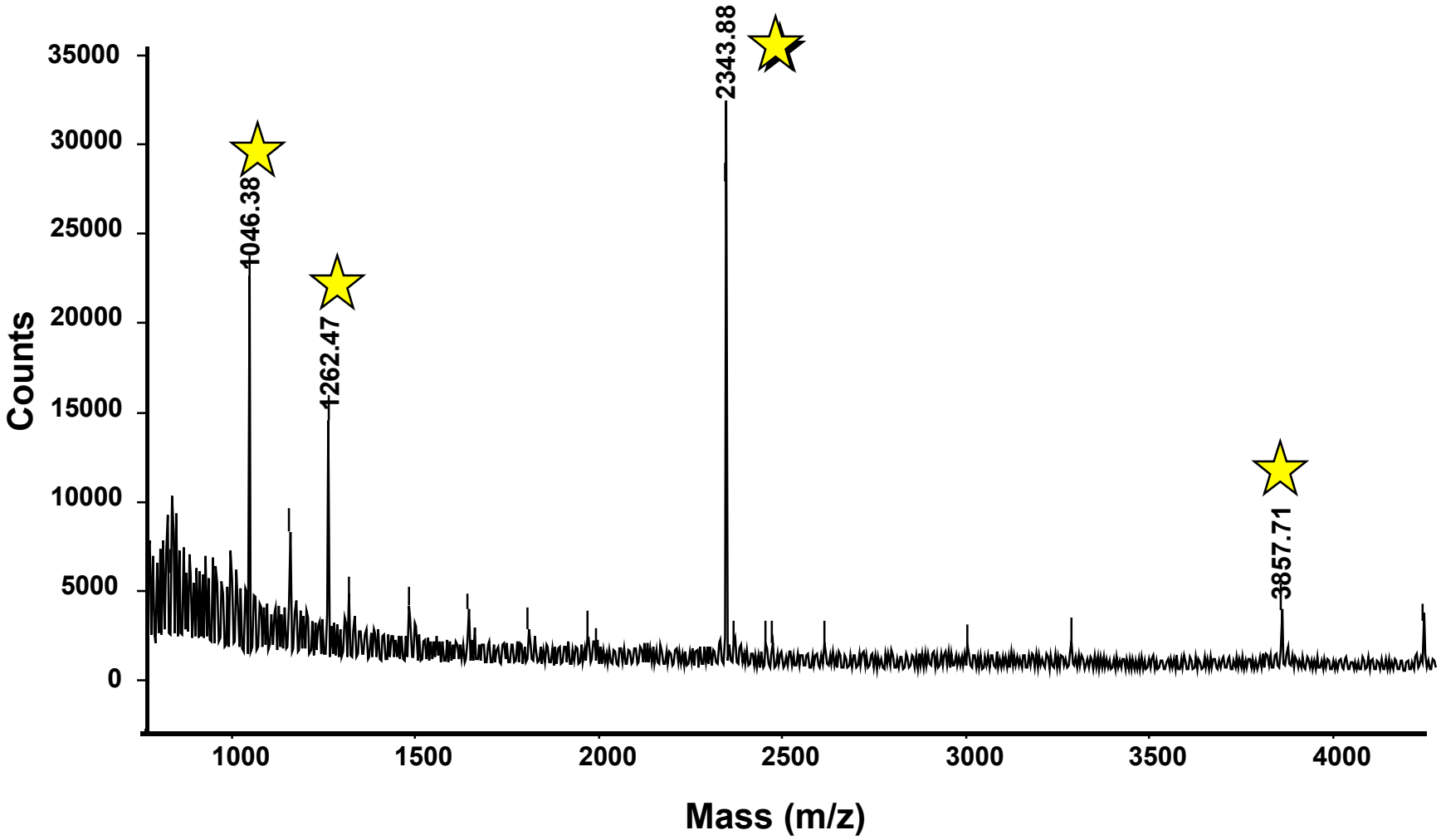
Searching databases with peptide masses to identify proteins

Best site is at www.matrixscience.com

The program (MASCOT) can search the OWL or NCBI databases using a set of tryptic peptide masses, or the fragment ions (specified or unspecified) of peptides

Presents the expected set of tryptic peptides for each matched protein

MALDI-TOF mass spectrum of tryptic digest of p22 band purified by 6xHis-tag

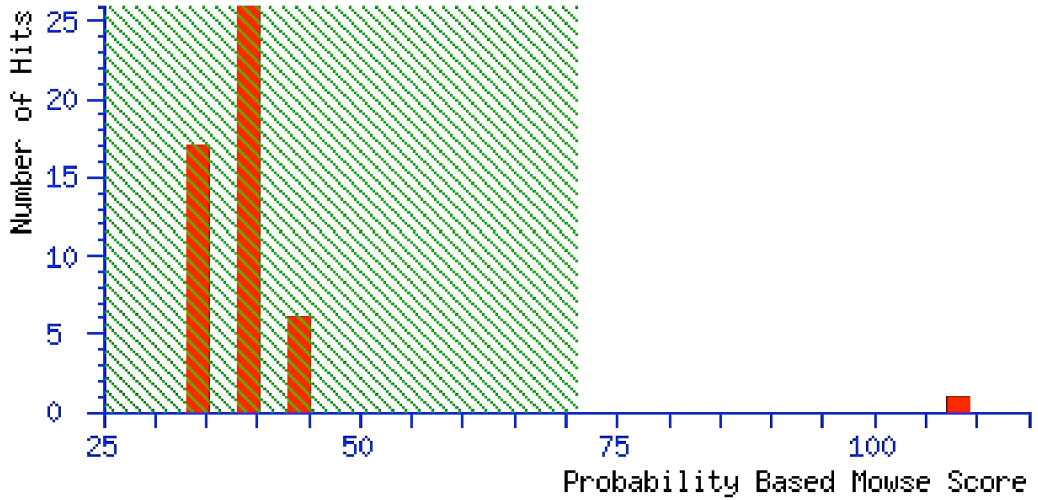


S. Barnes-UAB 1/28/05

Probability Based Mowse Score

Score is -10*Log(P), where P is the probability that the observed match is a random event.

Protein scores greater than 71 are significant (p<0.05).



Accession	Mass	Score	Description
1. gi|548939	20840	108	FKBP-TYPE PEPTIDYL-PROLYL CIS-TRANS ISOMERASE SLYD (PPIASE) (ROTAMA
2. gi|13384624	46931	45	myocyte enhancer factor 2C [Mus musculus]
3. gi|5257384	43424	44	(AF137308) phytochrome B [Lolium perenne]
4. gi|4505147	50305	44	MADS box transcription enhancer factor 2, polypeptide C (myocyte enhan
5. gi|1515365	44552	43	(U52596) nucleocapsid protein [Avian infectious bronchitis virus]
6. gi|6093850	49443	42	PRESENILIN 2 (PS-2)
7. gi|15225198	47999	42	hypothetical protein [Arabidopsis thaliana]
8. gi|113854	58376	41	NITROGENASE IRON-IRON PROTEIN ALPHA CHAIN (NITROGENASE COMPONENT I)
9. gi|13928425	13831	40	(AB040419) envelope protein [Bovine immunodeficiency virus]
10. gi|4389228	56064	40	Chain Z, Crystal Structure Of The Complex Between Escherichia Coli Glycerol

MASCOT SEARCH SUMMARY

1. gi|548939 Mass: 20840 Score: 108
FKBP-TYPE PEPTIDYL-PROLYL CIS-TRANS ISOMERASE SLYD (PPIASE) (ROTAMA

Observed	Mr(expt)	Mr(calc)	Delta	Start	End	Miss	Peptide
1046.38	1045.37	1045.59	-0.22	132 -	140	0	FNVEVVAIR
1262.47	1261.46	1261.70	-0.24	6 -	16	0	DLVVSLAYQVR
2343.88	2342.87	2343.08	-0.20	58 -	78	0	FDVAVGANDAYGQYDENLVQR
3857.71	3856.70	3856.89	-0.19	96 -	131	0	FLAETDQGPVPVEITAVEDDHVVVDGNHMLAGQNLK

2. gi|13384624 Mass: 46931 Score: 45
myocyte enhancer factor 2C [Mus musculus]

Observed	Mr(expt)	Mr(calc)	Delta	Start	End	Miss	Peptide
1046.38	1045.37	1045.50	-0.13	263 -	271	0	NTMPSVNQR
3857.71	3856.70	3856.76	-0.06	178 -	218	0	NSMSPGVTHRPPSAGNTGGLMGGDLTSGAGTSAGNGYGNPR

No match to: 1262.47, 2343.88

3. gi|5257384 Mass: 43424 Score: 44
(AF137308) phytochrome B [Lolium perenne]

Observed	Mr(expt)	Mr(calc)	Delta	Start	End	Miss	Peptide
1046.38	1045.37	1045.54	-0.17	380 -	389	0	GIDELSSVAR
3857.71	3856.70	3856.72	-0.02	86 -	122	0	SPHGCHAQYMANMGSIASLVMAVIISSGGEDEHNMGR

No match to: 1262.47, 2343.88

4. gi|4505147 Mass: 50305 Score: 44
MADS box transcription enhancer factor 2, polypeptide C (myocyte enhan

Observed	Mr(expt)	Mr(calc)	Delta	Start	End	Miss	Peptide
1046.38	1045.37	1045.50	-0.13	265 -	273	0	NTMPSVNQR
3857.71	3856.70	3856.76	-0.06	180 -	220	0	NSMSPGVTHRPPSAGNTGGLMGGDLTSGAGTSAGNGYGNPR

No match to: 1262.47, 2343.88

E. coli: FKBP-TYPE PEPTIDYL-PROLYL CIS-TRANS ISOMERASE

Nominal mass of protein (Mr): 20840

```
  1 MKVAKDLVVS LAYQVRTEDG VLVDESPVSA PLDYLHGHGS
 41 LISGLETALE GHEVGDKFDV AVGANDAYGQ YDENLVQRVP
 81 KDVFMGVDEL QVGMRFLAET DQGPVPVEIT AVEDDHVVVD
121 GNHMLAGQNL KFNVEVVAIR EATEEELAHG HVHGAHDHHH
161 DHDHDGCCGG HGHDHGHEHG GEGCCGGKGN GGCGCH
```

Tryptic fragments detected by MALDI-TOF-MS

```
132-140 FNVEVVAIR
  6- 16 DLVVSLAYQVR
 58- 78 FDVAVGANDAYGQYDENLVQR
 96-131 FLAETDQGPVPVEITAVEDDHVVVDGNHMLAGQNLK
```

Other web sites for peptide analysis

- **http://prowl.rockefeller.edu/**
  - **Choose ProFound**
- **http://prospector.ucsf.edu/**
  - **Choose MS-fit**

Further information on identified protein

- **Take the protein identifier number:**
  - **For this example it is gi|548939**
  - **Go to http://www.ncbi.nlm.nih.gov**
  - **Under Entrez, paste in the gi number**
  - **A link to the protein will appear**
  - **Click on Blink - this is similar to BLAST, but better**
  - **Select 3D-structures on this page to get Protein Data Base record(s) of crystal structure data of the nearest protein - this will yield 1IX5**
  - **Go to Structure (top of web page) and enter 1IX5 and click on its icon on the next page**
  - **To view a 3D-image of the protein, first download Cn3D from the NCBI site**

Examples for homework (due Feb 8th)

- **Identify the following proteins from these MALDI ions (corrected for isotope effects):**
  - **910.46, 1350.81, 1515.66, 1632.87, 1800.92, 1853.96 (human)**
  - **965.46, 998.56, 1001.45, 1068.48, 1581.81, 1677.96 (rat)**
  - **937.52, 972.49, 1049.56, 1209.64, 1508.70, 1844.89 (mouse)**
- **Set the number of tryptic cuts to 0 and try varying the mass accuracy from 0.02 to 1.0 Da. How does this alter the MOWSE score?**

How to represent the mass of compound?

- **At low mass resolution (where the isotope peaks cannot be resolved) what is observed is the *average mass***

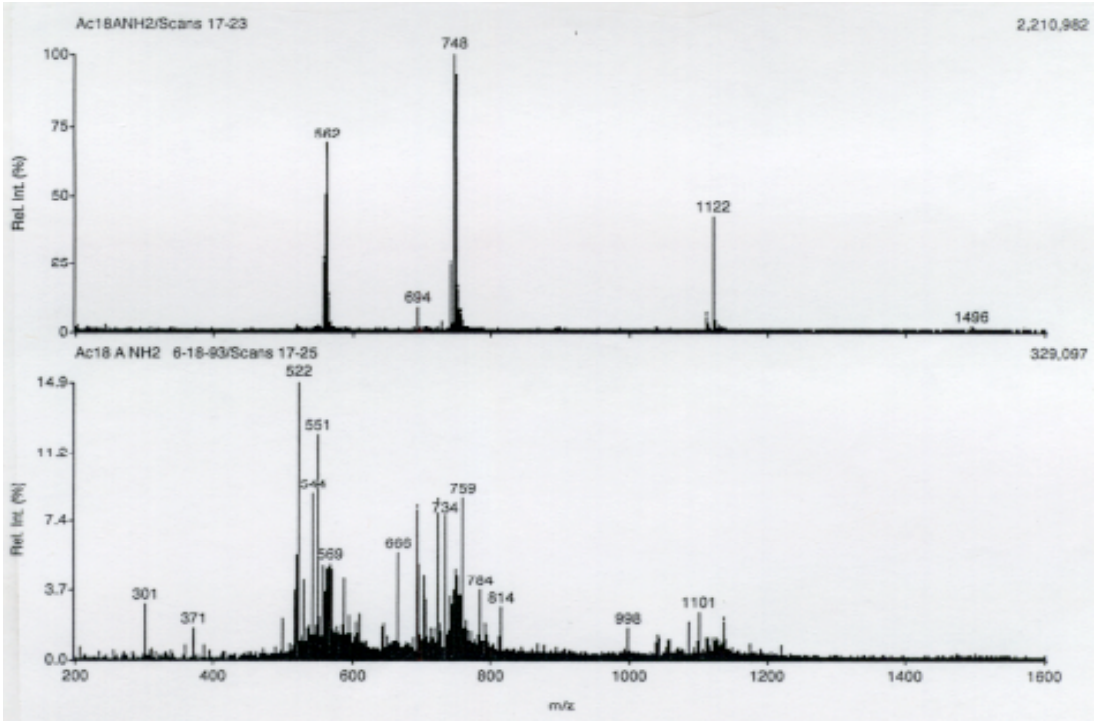
- **At high resolution where the isotopic peaks are fully resolved, then we can determine the *monoisotopic mass* for each one**

Take home question (due Feb 4th)

1. **What is the monoisotopic mass of human myoglobin?**
   - **Hint: workout the empirical formula of hMyoglobin - its sequence can be obtained from record P02144 at http://www.ExPasy.org**

2. **What is the molecular weight of the most abundant species of human myoglobin?**
   - **Hint: assume that the abundance of ^{13}C is 1.00% of total carbon atoms - do not worry about ^{2}H or other isotopes**

ESI-MS and purity of peptides

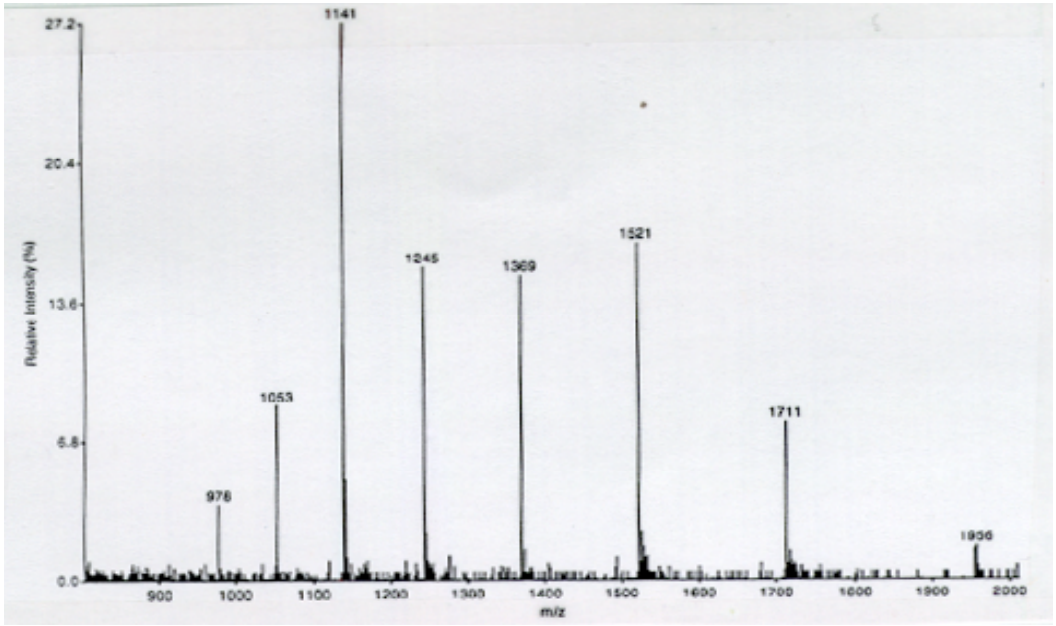


Guarantees of purity based on observation of "a single peak by reverse-phase HPLC" and by "it gave the correct sequence when analyzed by Edman degradation" are hollow.The lower spectrum was of a "pure" HPLC peak. The method of purification was amended and the upper spectrum was obtained

Ionizing proteins and peptides

- $^{+}H_3NCHR_1CO(NHCHR_nCO)_nNHCHR_2COOH$ is the ion that's found in dilute acid solution
- If there are internal basic residues, then the ions will be of the form $[M+nH]^{n+}$, where n = 1, 2, etc.
- A tryptic peptide will have a N-terminal amino group and an amino group from Arg or Lys
  - If the peptide has a mol. wt. of 1000 Da, then the singly charged ion will have a *m/z* of 1001, whereas the doubly charged ion has a *m/z* of 501

ESI mass spectrum of ribonuclease



Cumulative MW estimate = 13,680.29

SD = 2.94

Peak (m/z)	Intensity	Charge (est.)	Mol. Wt. (Est.)
978.00	7,778	14.00000	13,677.89
1,053.00	18,532	13.02656	13,675.90
1,141.00	59,087	11.95446	13,679.91
1,245.00	33,275	10.96146	13,683.91
1,369.00	32,390	10.03219	13,679.92
1,521.00	35,668	8.99995	13,679.93
1,711.00	16,624	7.99996	13,679.94
1,956.00	3,333	6.97955	13,684.94

Calculation of molecular weights and ion states

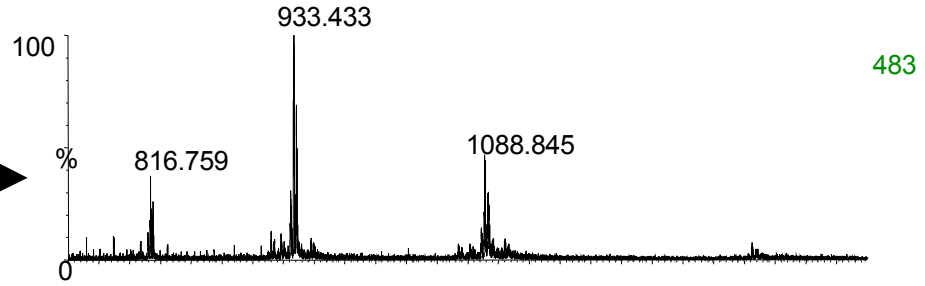
- **For two ions in a series for a peptide of molecular weight M, the lower *m/z* value (x) will be for the n+1 ion state and the larger *m/z* value (y) will be for the n+ ion state.**
  - **(1) $(M+n)/n = y$**
  - **(2) $(M+n+1)/(n+1) = x$**
- **Hence**
  - **(3) $M+n = ny$ and $M = ny-n$**
  - **(4) $M+n+1 = (n+1)x$ and $M = (n+1)x-(n+1)$**
- **Hence**
  - **$ny-n = (n+1)x - (n+1)$**
  - **$ny-n-xn+n = x-1$**
  - **$n(y-x) = x-1$**
  - **$n = (x-1)/(y-x)$**
- **The value of n can then be substituted in equation (1) to obtain the molecular weight of the peptide**

Deconvolution of MS data

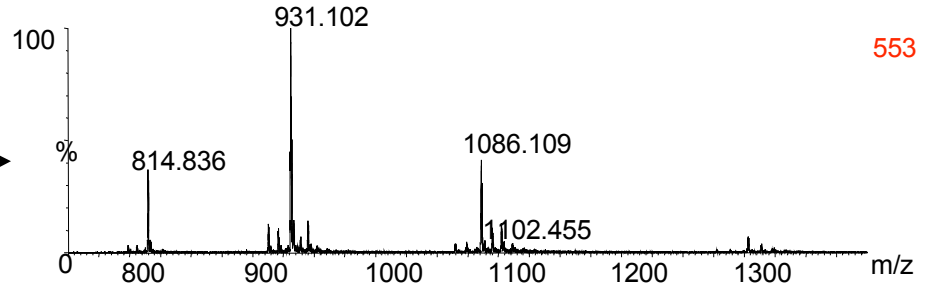
- When several proteins are present, then their multiply charged ion clusters overlap

- Can this be overcome? - yes, use the MaxEntropy program provided by Micromass

Oxidized Aprotinin ESI mass spectrum



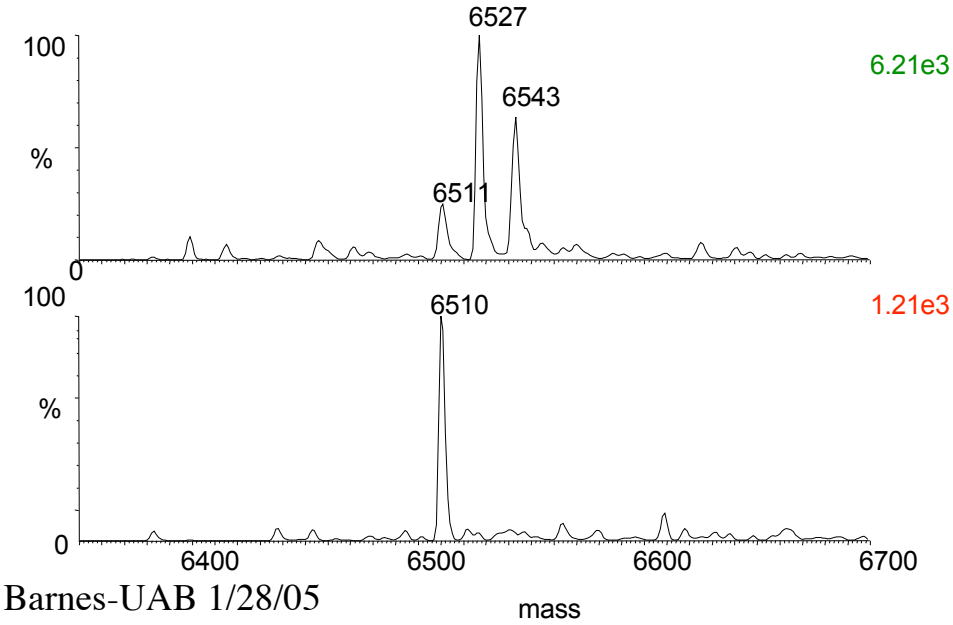
Control Aprotinin ESI mass spectrum



Deconvoluted mass spectra

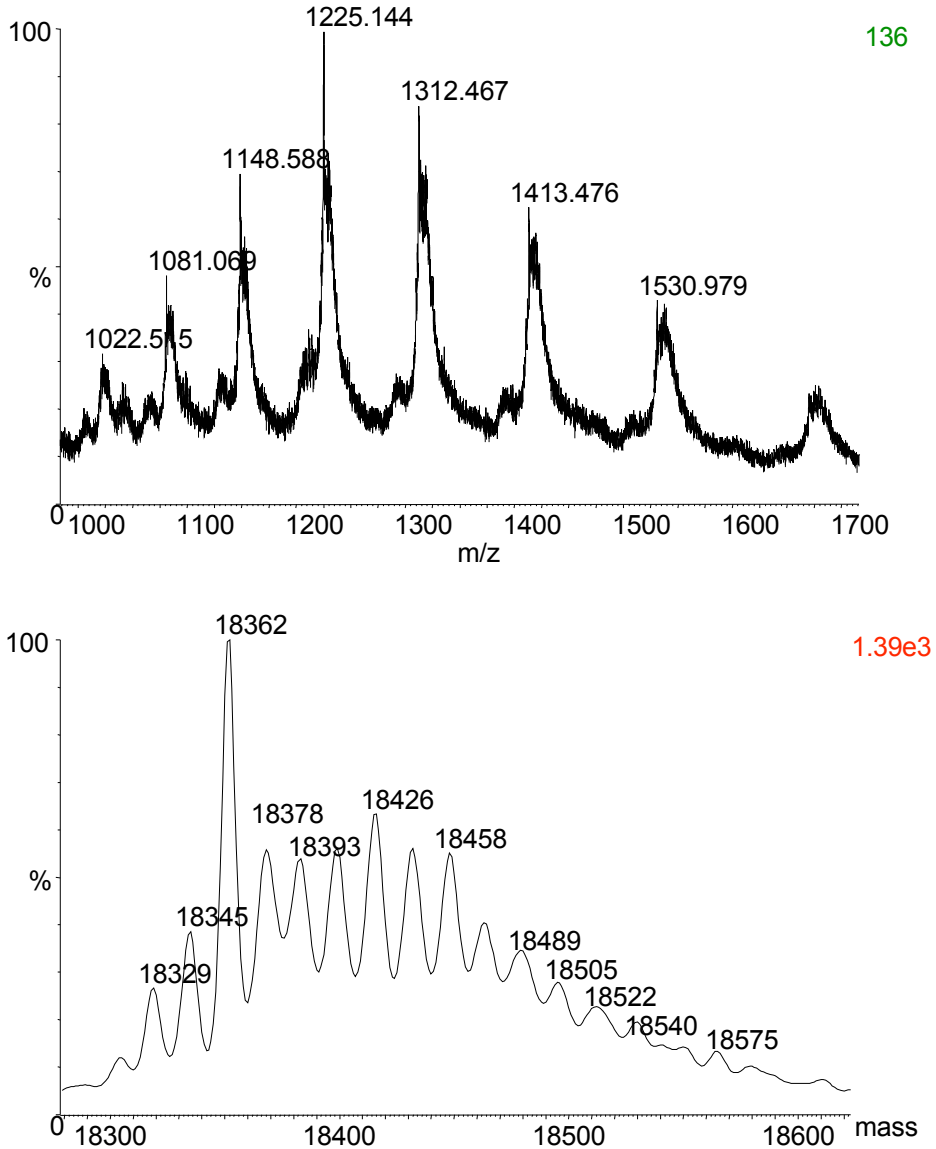
Oxidized

Control



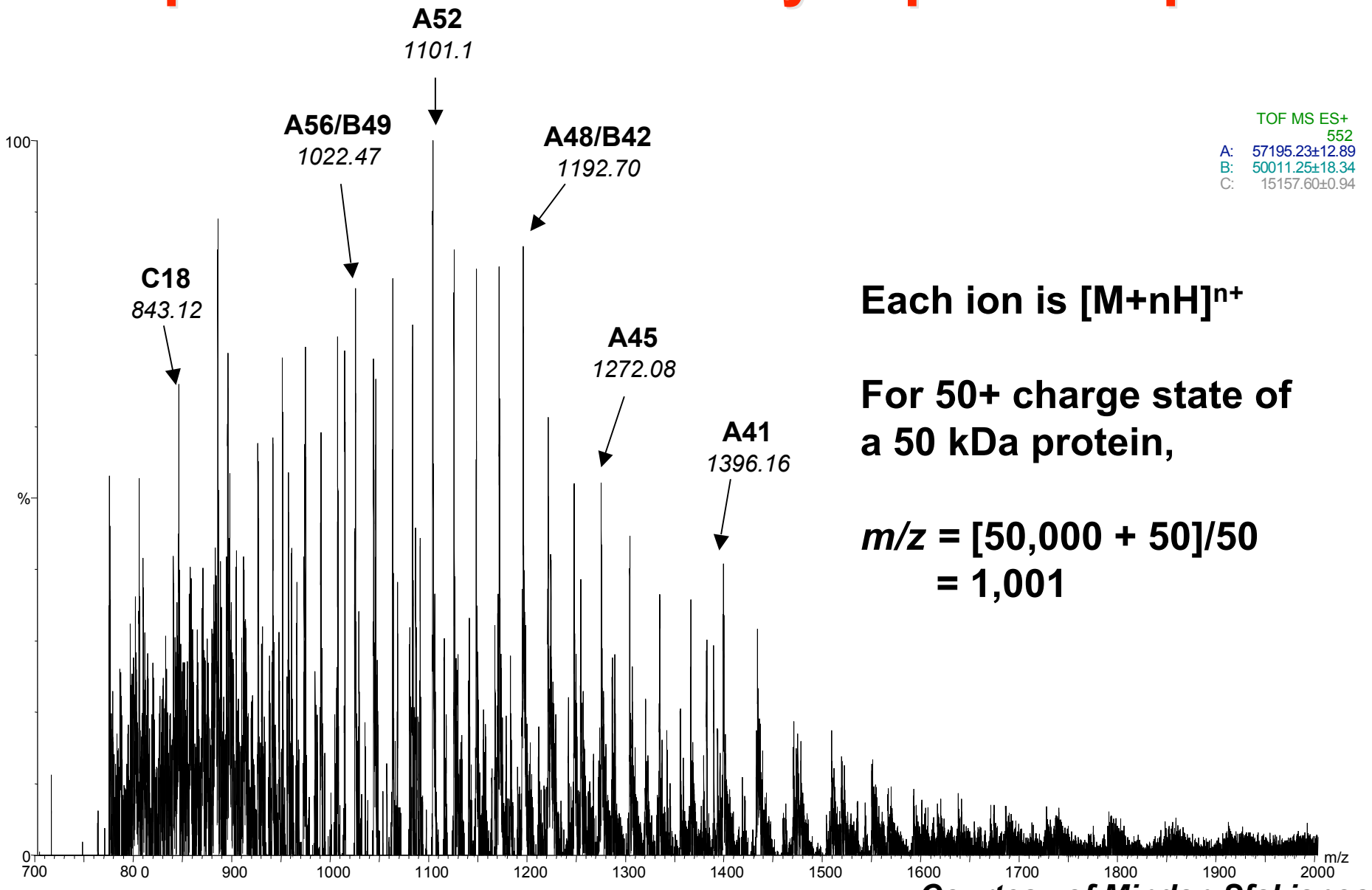
Junlong Shao

Deconvolution of oxidized forms of β-lactoglobulin



Junlong Shao

ESI spectrum of bacterially expressed protein



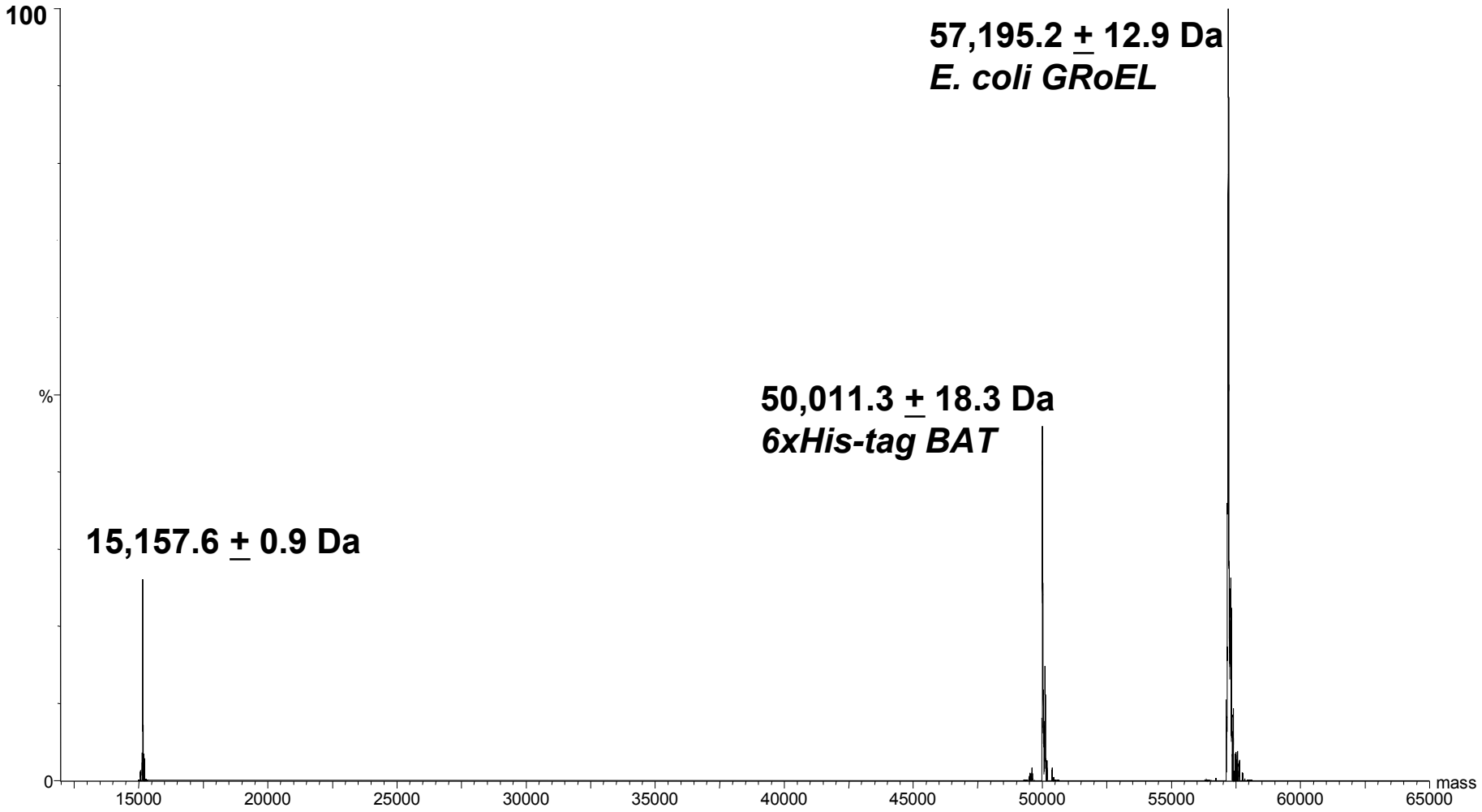
Each ion is $[M+nH]^{n+}$

For 50+ charge state of a 50 kDa protein,

***m/z* = [50,000 + 50]/50**
= 1,001

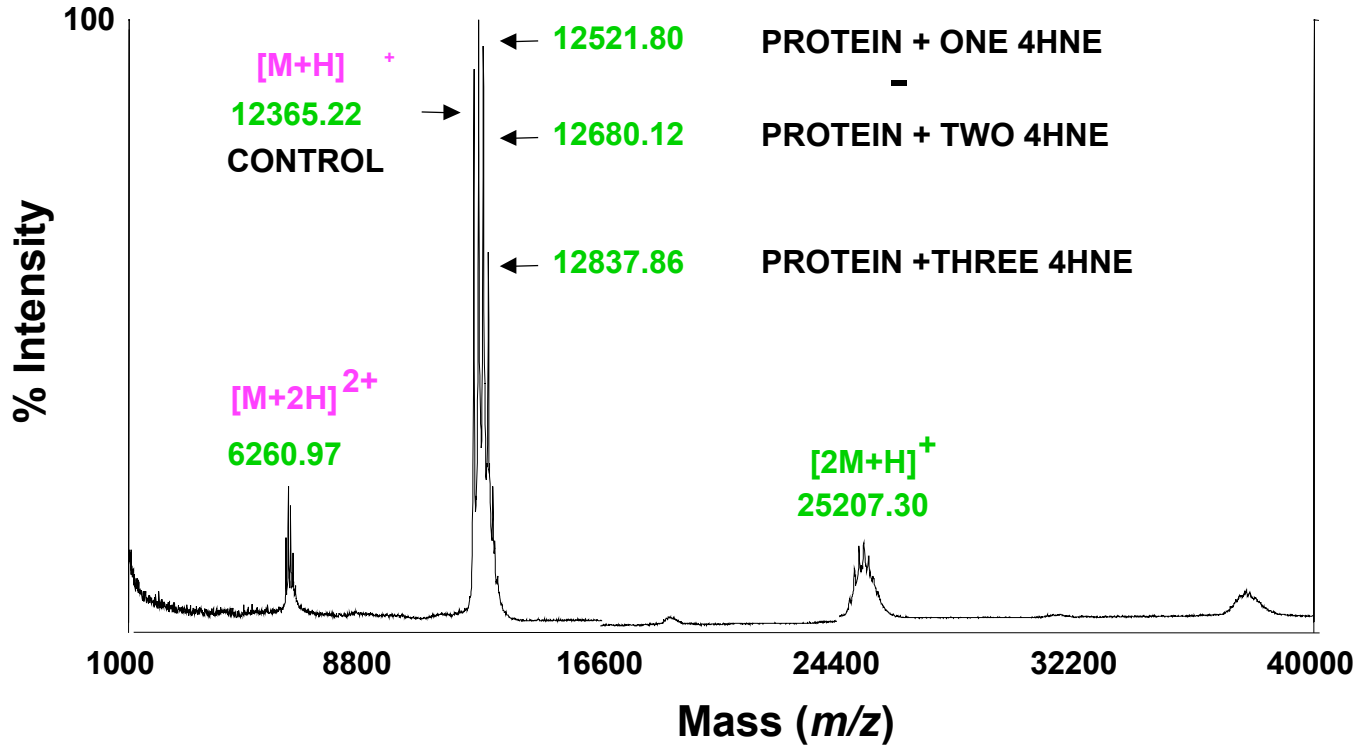
Courtesy of Mindan Sfakianos

MaxEnt deconvolution of MWs



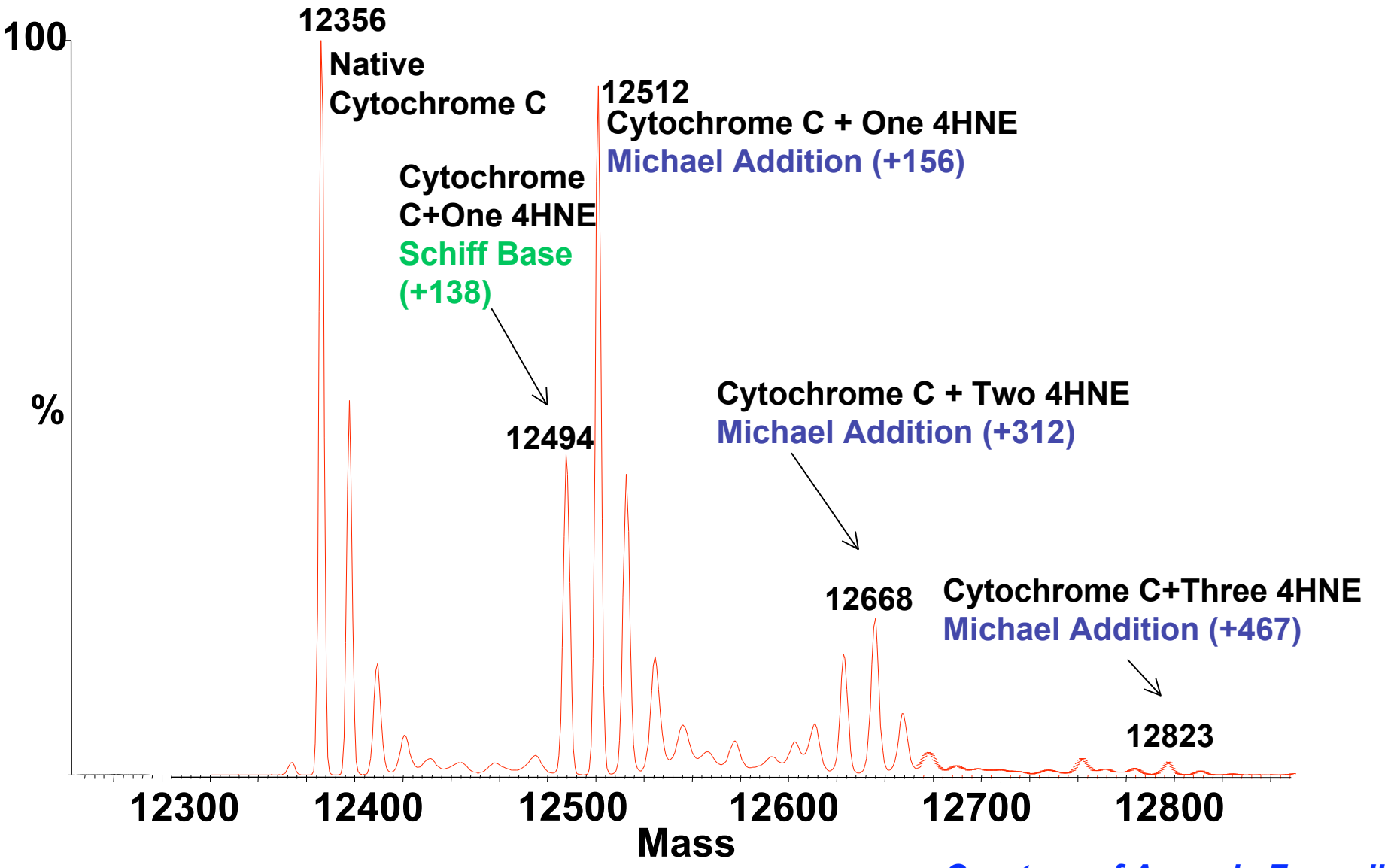
Courtesy of Mindan Sfakianos

Cytochrome C Modified by HNE
MALDI-TOF Mass Spectrum



MALDI spectra usually contain only the **molecular ion $[M+H]^+$**. This is an advantage since it maximizes the signal, but is a disadvantage in that it gives no clues as to structure.

ESI-MS of 4HNE-Modified Cytochrome C



Courtesy of Amanda Foxwell

Summary of determining MW by ESI

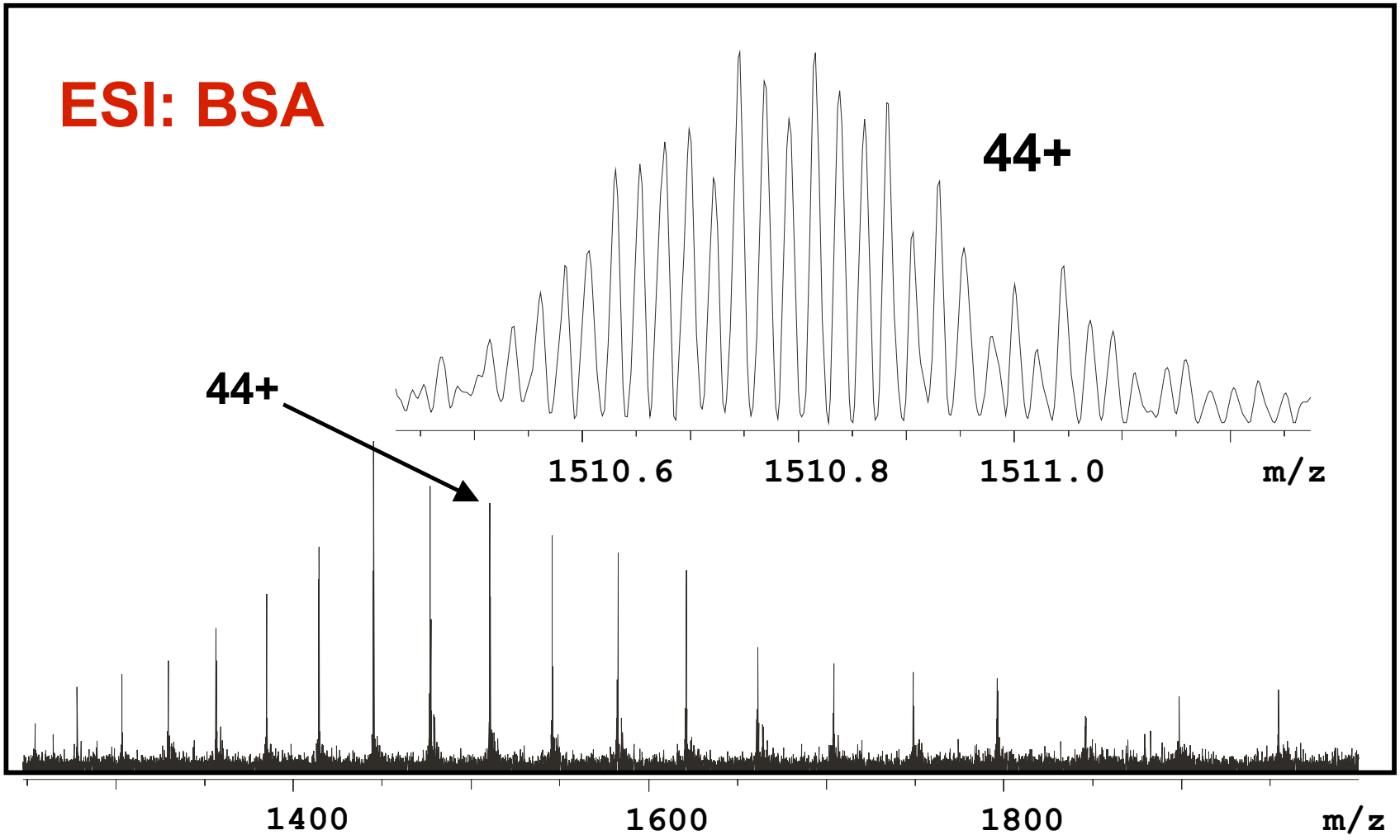
- **The multiple charge states of a protein allow:**
  - **Mol Wt of large proteins to be estimated**
  - **accurate estimation of mol wt (super SDS-PAGE gel)**
- **Important to remember that the protein sample must be free of salt**
  - **Typically, a sample is cleaned up on a short reverse-phase column prior to electrospray**
  - **Alternative, use ammonium acetate as buffer**

Use of FT-MS in ESI of proteins

- **The very high resolving power of FT-MS enables a direct measure of charge state of an individual ion since each peptide or polypeptide will have several/many isotope peaks**

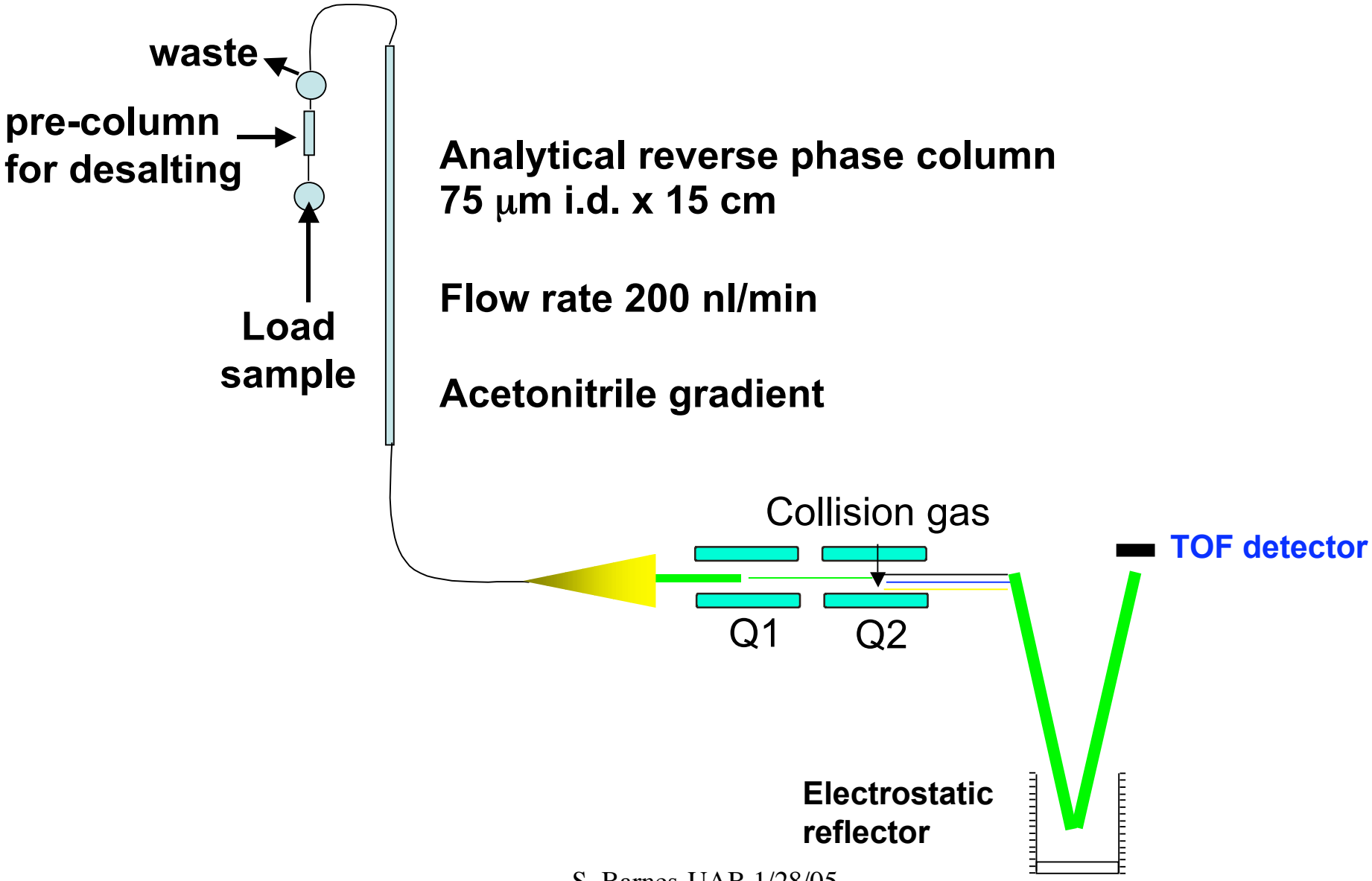
- **The distance in Da between successive isotope peaks of a multiply charged ion is the reciprocal of the number of charges**

Bovine Serum Albumin (66 kDa)
4.7 T Actively Shielded Magnet

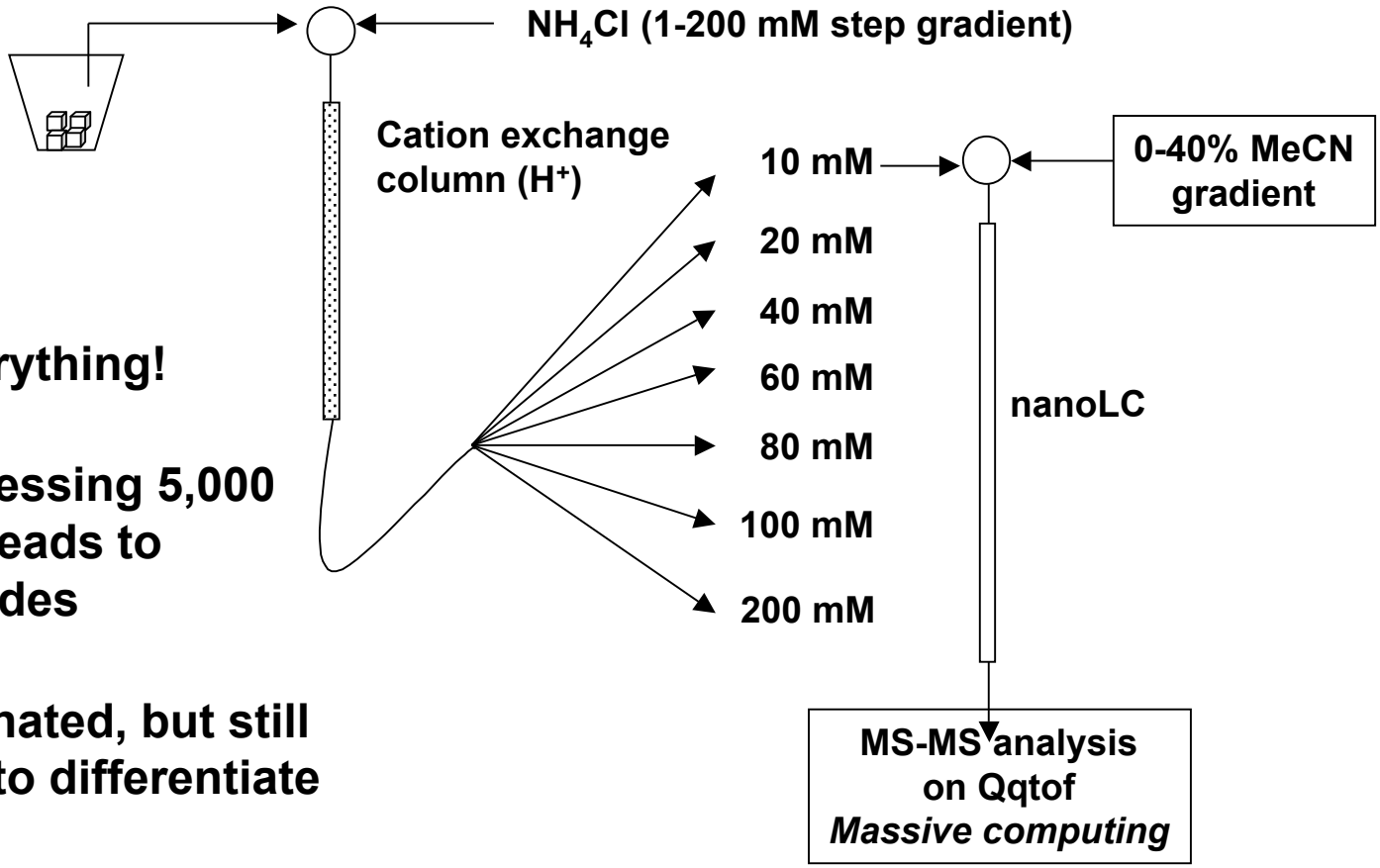


Bruker Daltonics

LC-MS of peptide mixtures



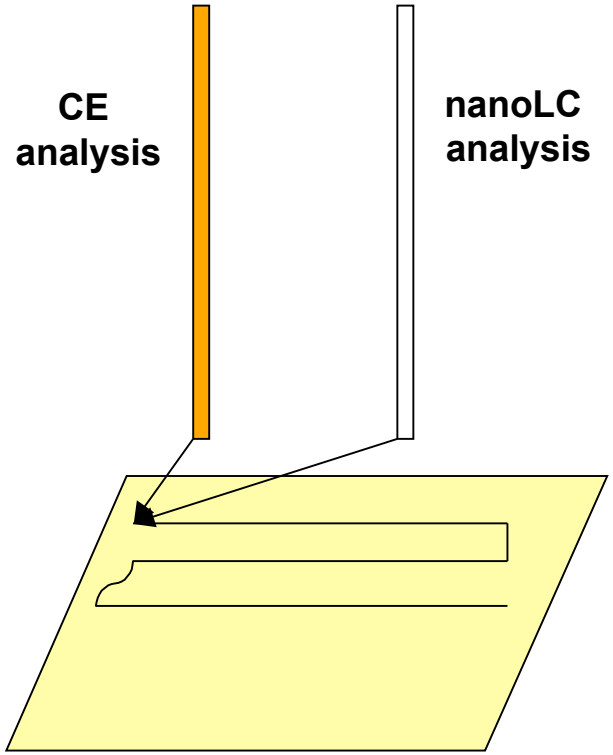
MUDPIT - MUlti-Dimensional Protein Identification Technology



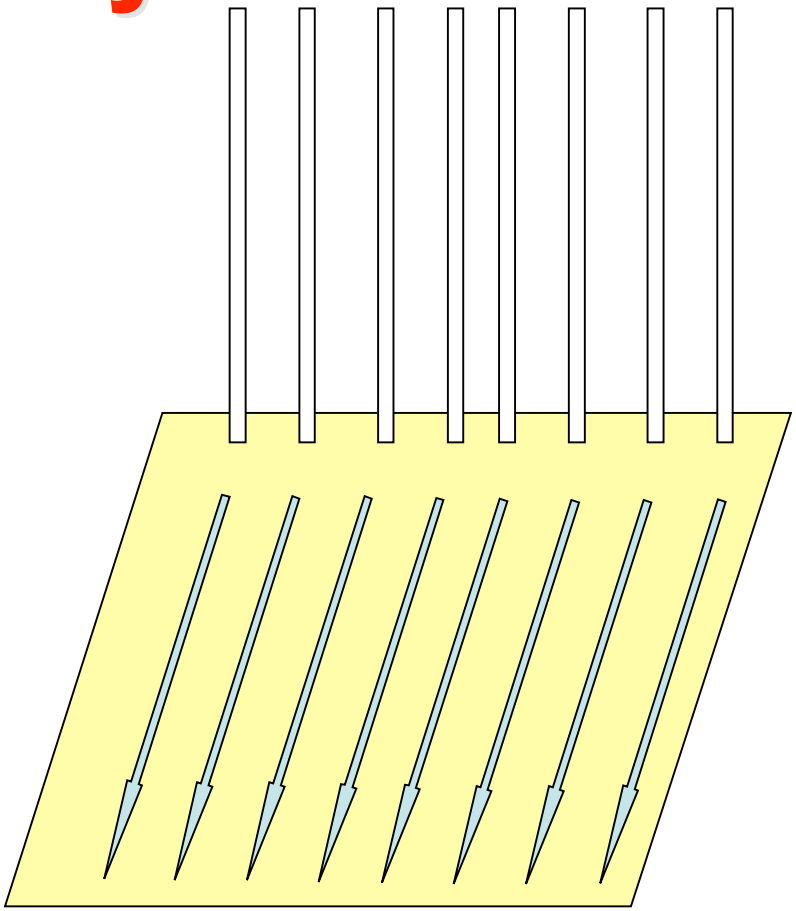
- Hydrolyze everything!
- For a cell expressing 5,000 proteins, this leads to >100,000 peptides
- Can be fractionated, but still 10,000-20,000 to differentiate
- Enormous bioinformatics problem

John Yates

Connecting CE and LC to MALDI analysis



Creates 20 mm wide tracks that can be scanned by MALDI laser for MS analysis



Parallel capture of effluents of 8 nanoLC separations on Mylar - can be scanned simultaneously by fast laser

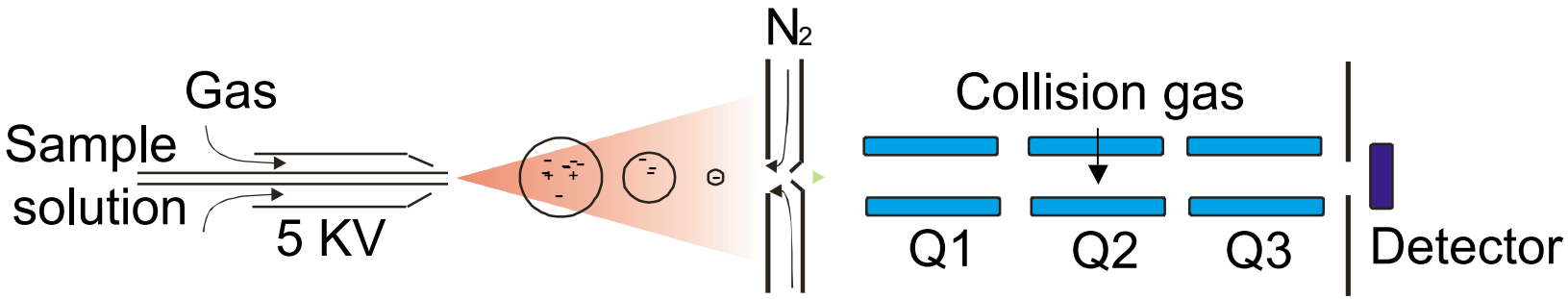
Pros/Cons of laying down LC or EC separations on matrix plate

- Allows off-line analysis both in real time and then in a retrospective mode
- MALDI-TOF analysis is very fast
- Can do TOF-TOF MS-MS analysis
- BUT what happens chemically on the acidic environment on the surface of the plate during storage?
- Also, can the laser beam cause chemical changes?

Sequencing of peptides

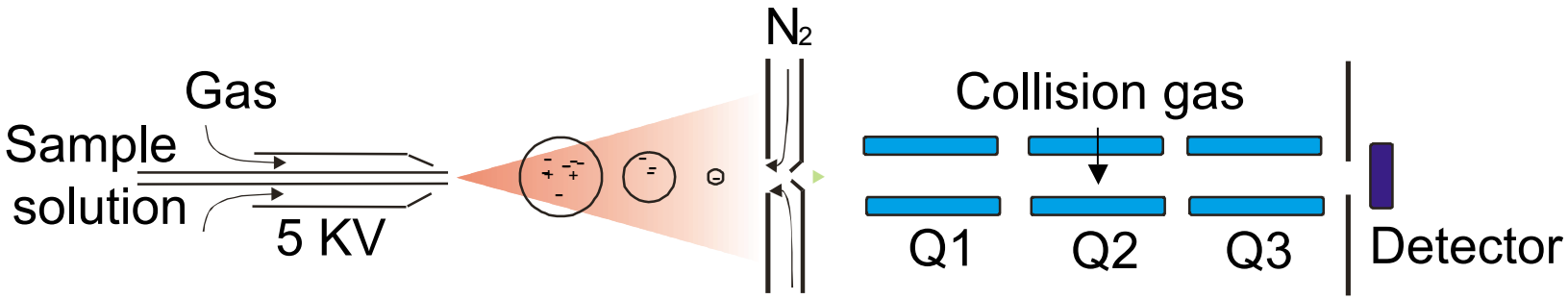
- **Using tandem mass spectrometry in a triple quadrupole, Q-tof, or ion trap instrument, the parent ion is first selected in the first quadrupole**
- **The parent ion is collided with argon gas and it breaks into fragments (daughter ions)**
- **By identifying the daughter ions, the peptide amino acid sequence is inferred**

Tandem mass spectrometry on a triple quadrupole instrument

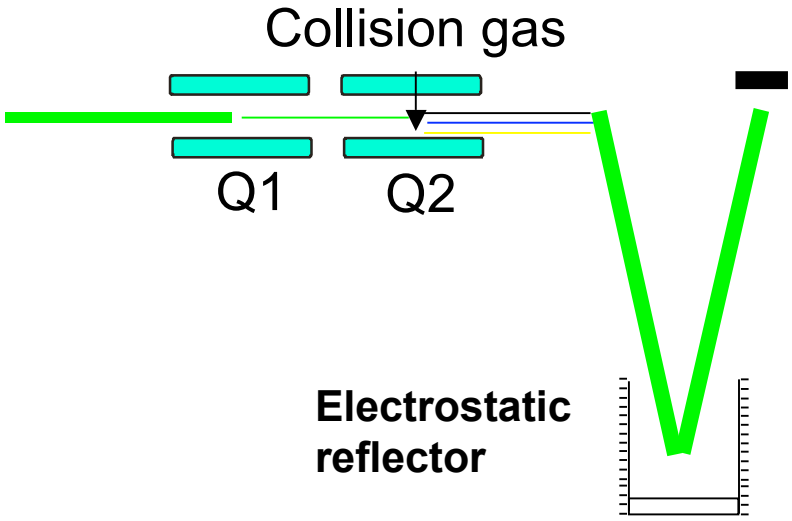


- **Daughter ion spectra**
- **Parent ion spectra**
- **Multiple reaction ion scanning**

Triple quad versus Q-tof and sensitivity



The quadrupole analyzer (Q3) is slow and insensitive - it's a filter - thus throws away large amounts of data

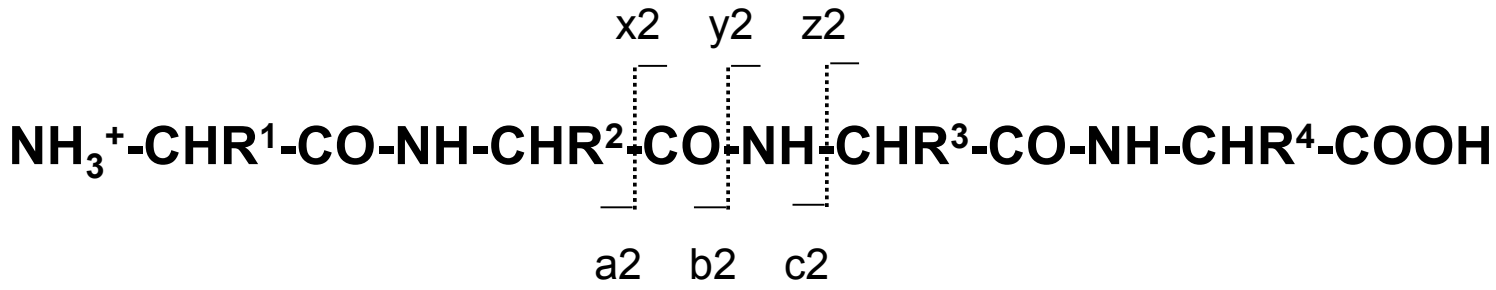


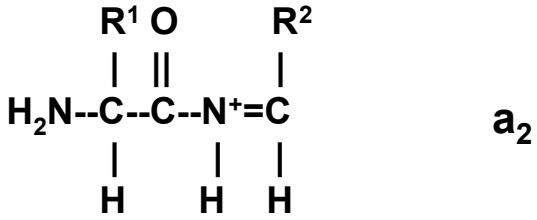
TOF detector collects all ions generated and yields fmol rather than pmol sensitivity

Also gives far greater mass accuracy - from 1000 ppm on the triple quad to <20 ppm on the Q-tof

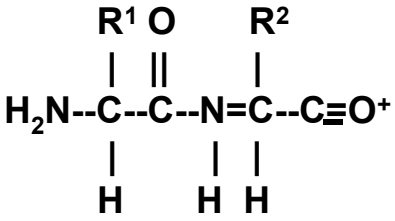
Crucially important for automated interpretation of MS-MS spectra to yield amino acid sequence

Fragmenting a peptide

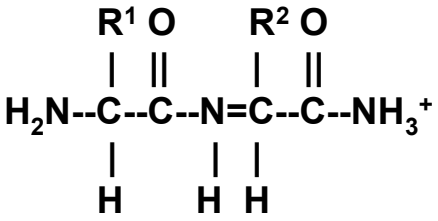




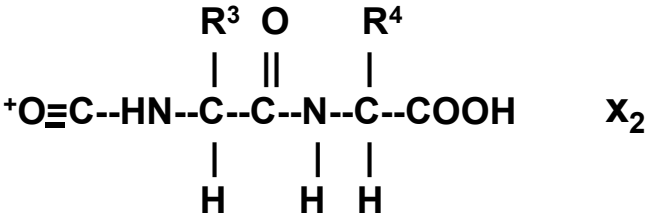
a_2



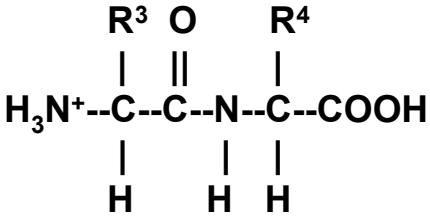
b_2



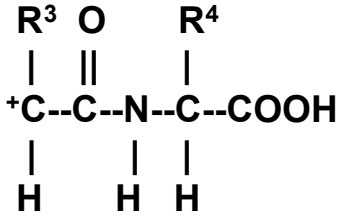
c_2



x_2



y_2



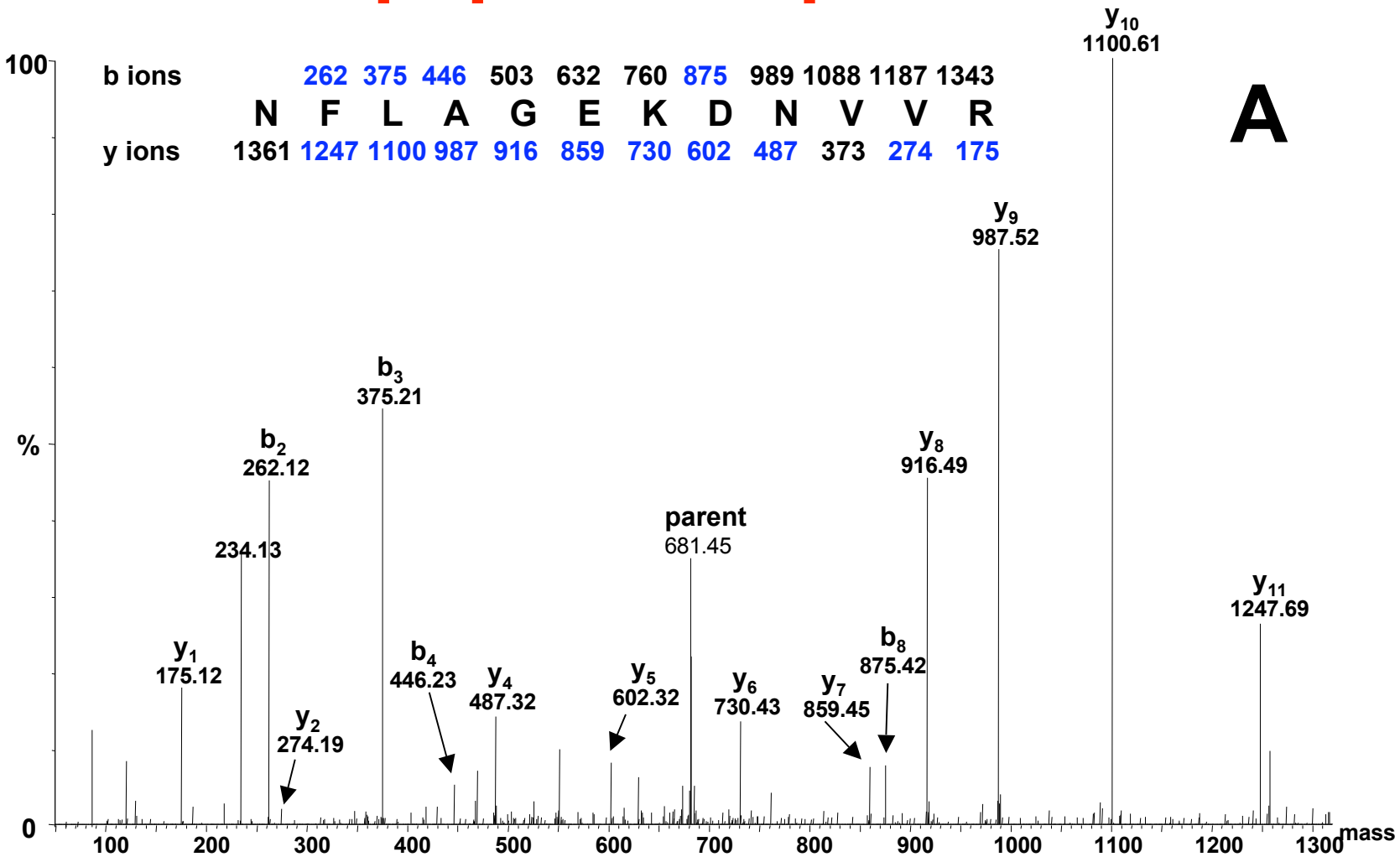
z_2

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

Amino acid residues masses

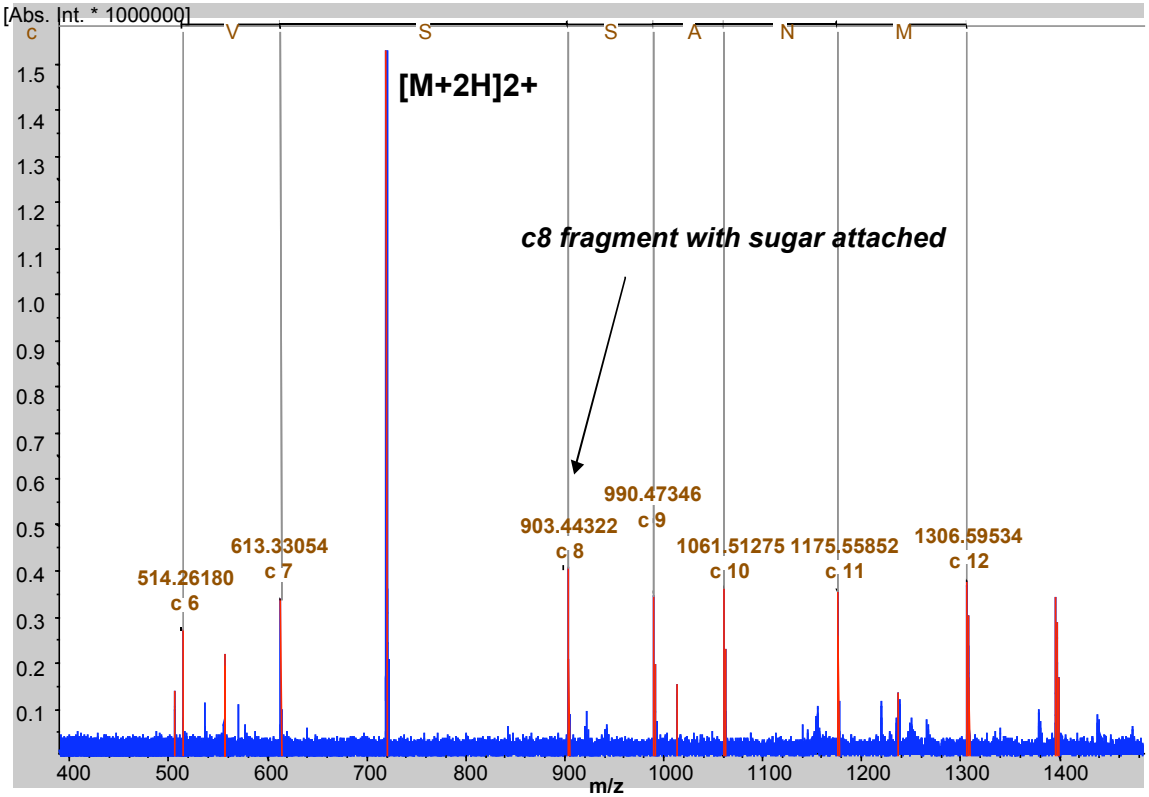
Alanine	71.037	Leucine	113.084
Arginine	156.101	Lysine	128.094
Asparagine	114.043	Methionine	131.040
Aspartic acid	115.027	Phenylalanine	147.068
Cysteine	103.009	Proline	97.053
Glutamic acid	129.043	Serine	87.032
Glutamine	128.058	Threonine	101.048
Glycine	57.021	Tryptophan	186.079
Histidine	137.059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

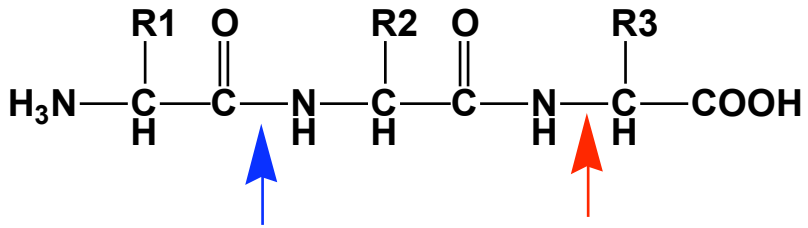
Identification of daughter ions and peptide sequence



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

Casein kinase II - AGGSTPVSSANMMSG

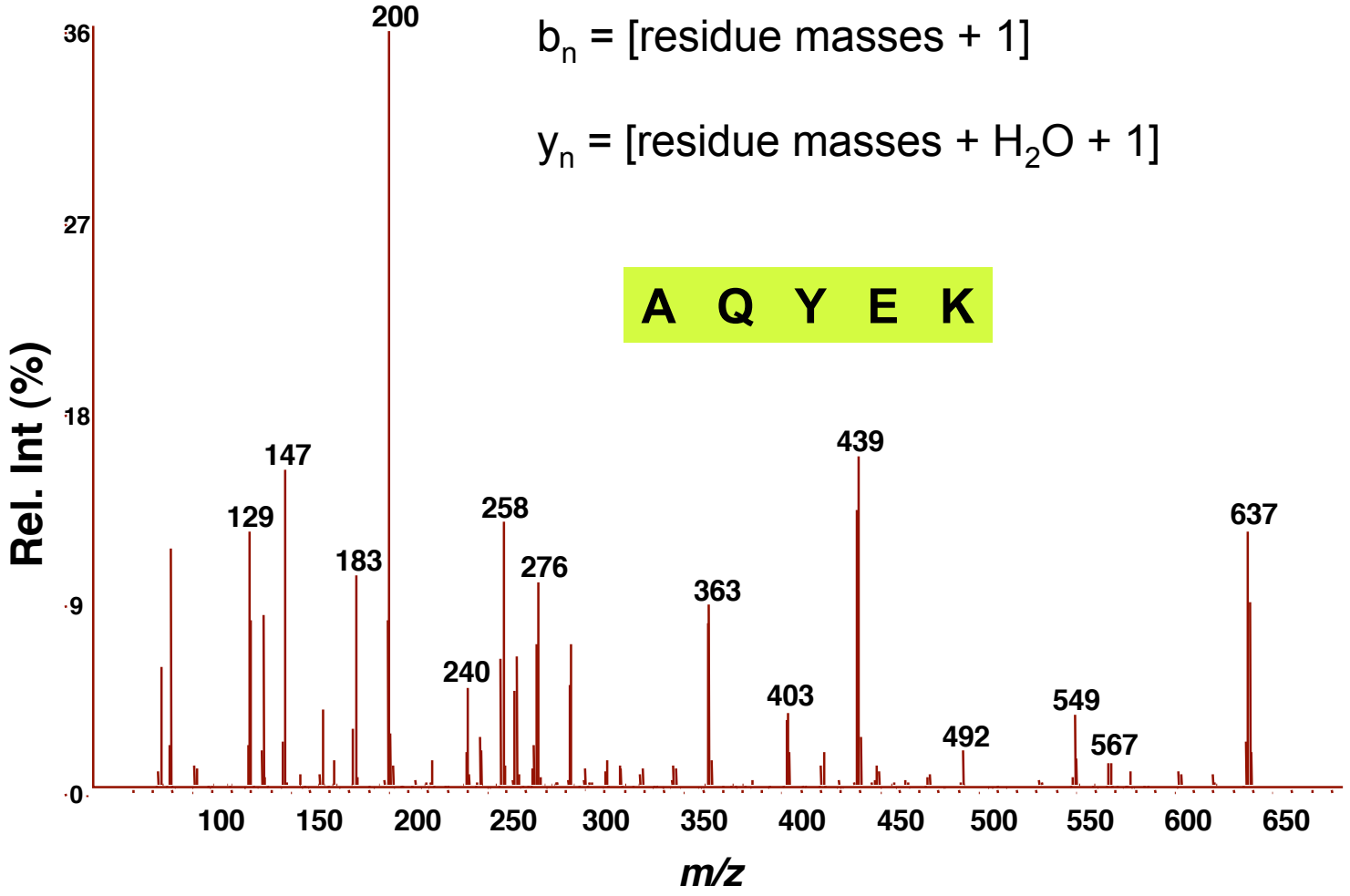




b ion cleavage **c ion cleavage**

Fragment ions of a small 5-mer peptide

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



b_n = [residue masses + 1]

y_n = [residue masses + H_2O + 1]

A Q Y E K