Proteomics:
Can it really tell what's going on?
Part I:

Helen Kim, Ph.D.

Department of Pharmacology & Toxicology
2-DE Proteomics Lab; Comprehensive Cancer Center Mass Spectrometry Shared Facility

University of Alabama at Birmingham

How can proteomics help in free radical biology research?

- **Separation and identification of proteins**
- **Need to identify posttranslational modifications**
  - *Nitration, chlorination, bromination, 4-HNE, carbonyls*
  - *Mostly analyzed by GC-MS after hydrolysis to nitrotyrosine, chlorotyrosine, bromotyrosine, etc.*
  - *Current methods result in loss of information as to which amino acid residue is modified*

What is Proteomics?

- **Genome : the complement of genes that defines a biological system;**
- **Proteome : the complement of proteins encoded by a genome**
- **Current definition of a proteome--**

  the complement of proteins that defines a biological system, or compartment, such as the mitochondrial proteome.

Thus, PROTEOMICS = *study of patterns of protein expression or modification* in a biological system.

Types of Proteomics Technologies

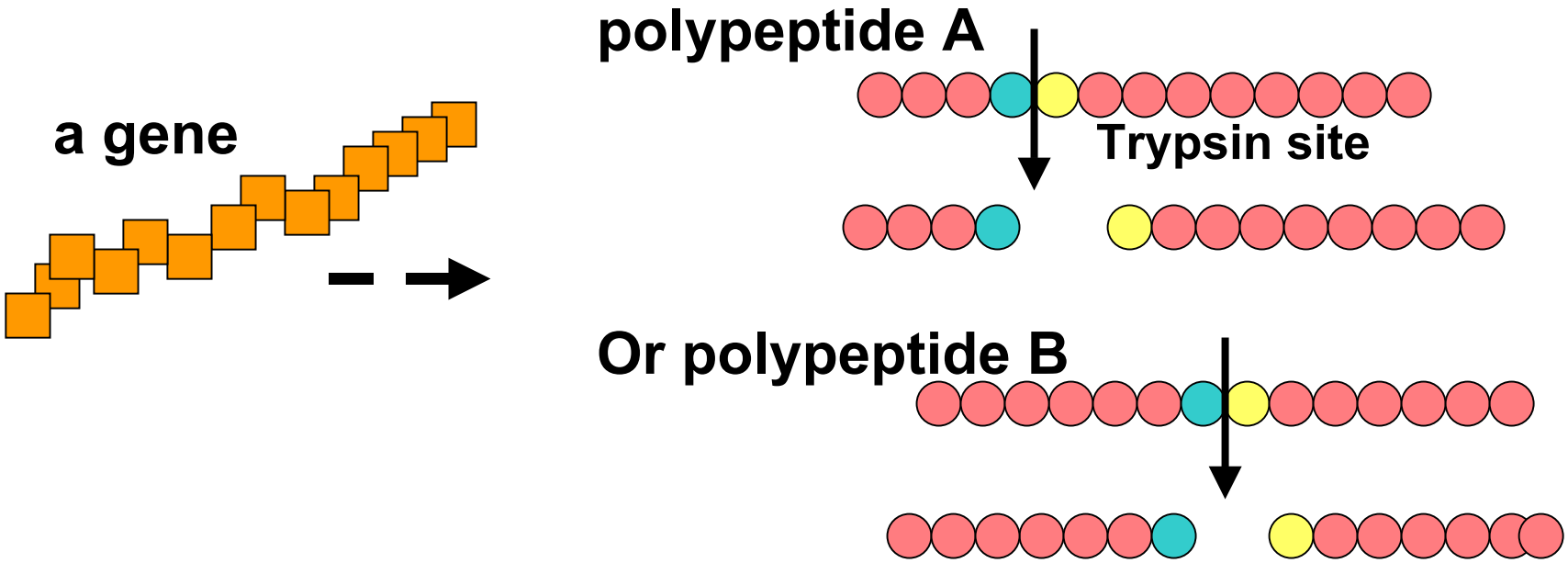
I. 2-dimensional electrophoresis (2-DE) & mass spectrometry

II. Liquid chromatography & mass spectrometry (LC-MS or LC/LC MS/MS)

III. "Chip" technology:

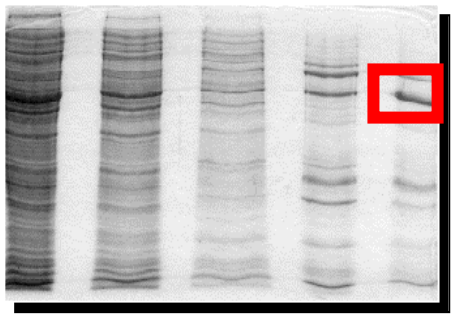
2-D array of recombinant polypeptides or antibodies on a single microscope slide; the entire chip is probed with a labelled "ligand" (protein, lipid, drug)

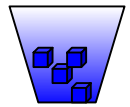
Proteolysis generates sets of peptides that are a “fingerprint” for that polypeptide



So, polypeptides of identical mass can have nonidentical tryptic “fingerprints.”

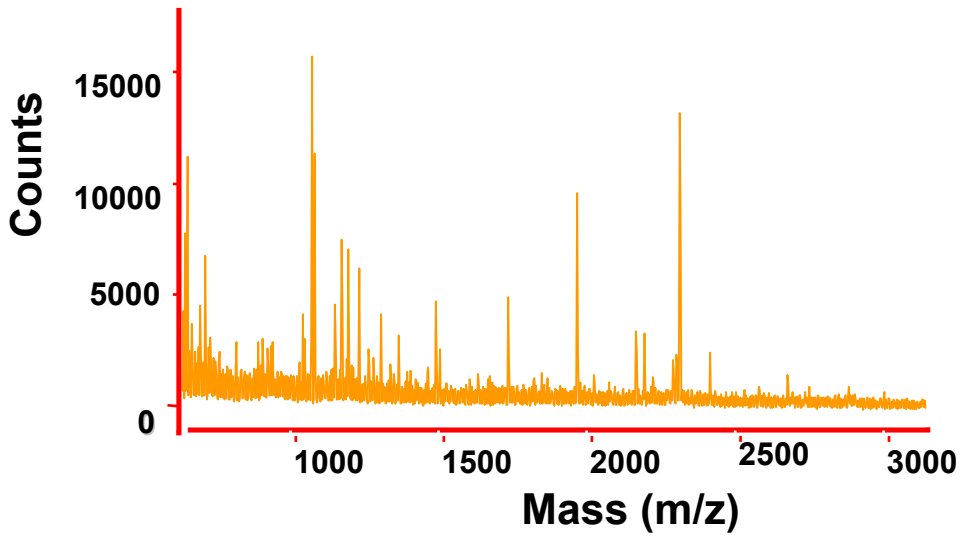
REVIEW: basics of protein mass spectrometry: From a band on a gel to a "tryptic peptide fingerprint"





trypsin

MALDI-TOF mass spectrometry



So, why do we need "proteomics"?

- When you don't have a clue;
- When you have only a very small clue;
- When you knock out a gene (and hence a protein) that you're convinced is essential for life, and the animal pees as usual.

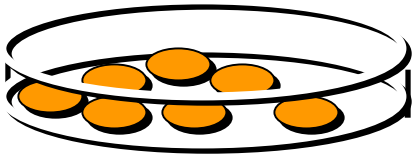
➡ When you know that protein modifications are involved, but

1. don't know the proteins or

2. don't know what the modifications are

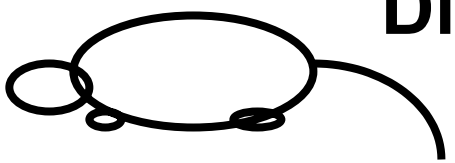
The Elements of 2-Dimensional Electrophoresis

Biological sample preparation:



Harvest, rinse, and pellet the cells;

or



Dissect out tissue, organ, or fluids;

Homogenize/lyse in buffer that unfolds the proteins w/o adding or disturbing the charges;
Clarify by centrifugation;
Protein assay;
Separate the polypeptides in two dimensions.

Parameters that affect detection of changes in a proteome by 2-DE

- **Experimental conditions**
  - **Conc of stimulus; Time point of cell harvest;**
- **Homogenization buffer conditions;**
- **Electrophoresis parameters:**
  - **pH range**
  - **Acrylamide gradient or straight %**
- **Use of pre-exisiting data/reagents:**
  - **Subcellular location of protein of interest known?**
  - **Are antibodies available?**

What 2-D electrophoresis involves:

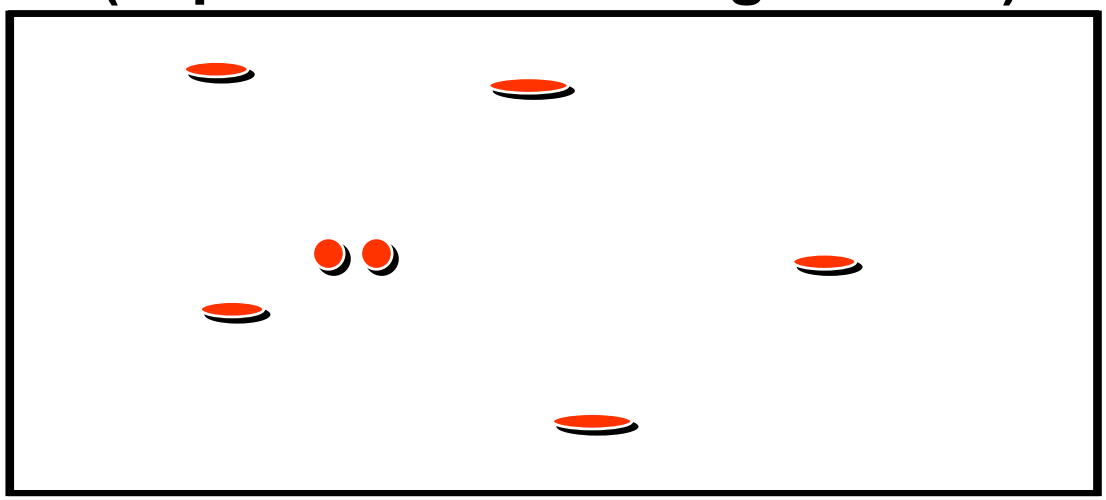
- 1st dimension: **Isoelectric focussing**

(separation according to charge)

pH 3 **pH 10**

- 2nd dimension: **(SDS)-PAGE**

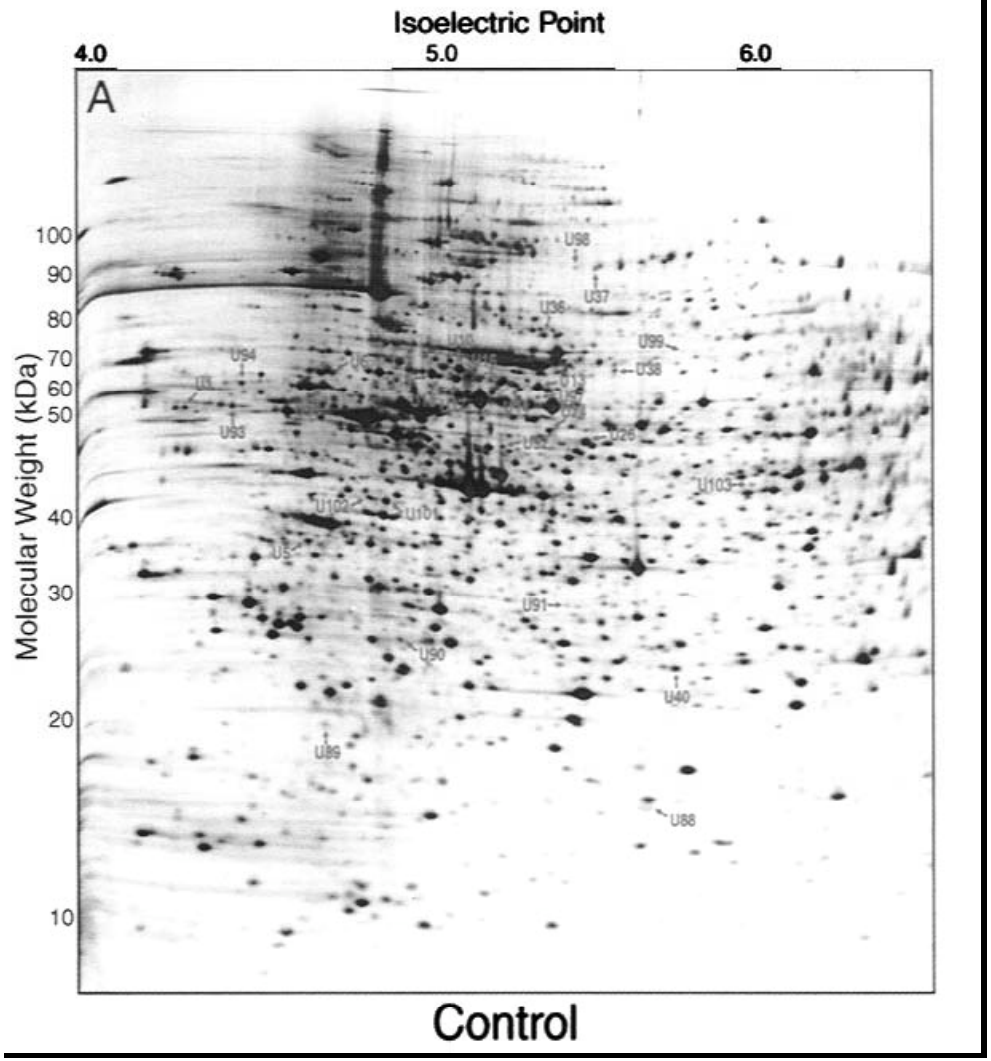
(separation according to size)



High m.w.

Low m.w.

A real 2-D gel



Find this and other 2-D gels at http:www.expasy.org

Lewis et al., [2000] Molec. Cell, 6)

(from Natalie Ahn's lab)

From this set of gels, 25 novel phosphorylations were identified as part of the MAP-kinase signalling pathway.

How to predict or interpret 2-D gel data:

I. Existing databases and weblinks:

www.expasy.org

helpful links:

proteomics tools

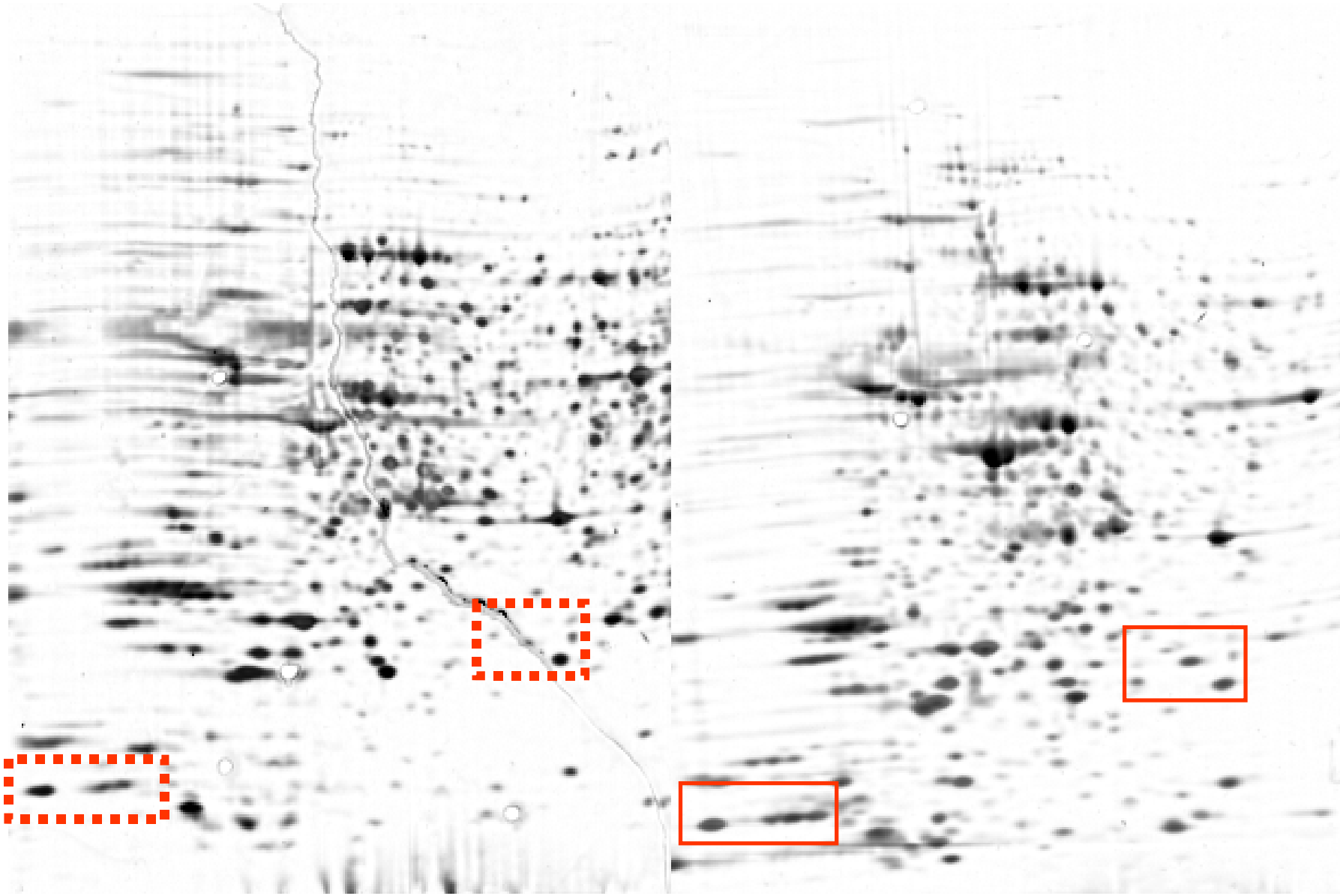
II. Keep up with the literature:

Electrophoresis

Proteomics

Molecular & Cellular Proteomics

Silver-stain detection of soy-induced protein changes

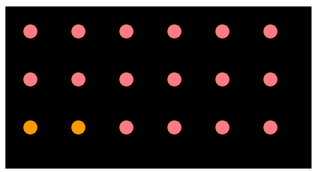


Results of MALDI-TOF mass spectrometry analysis

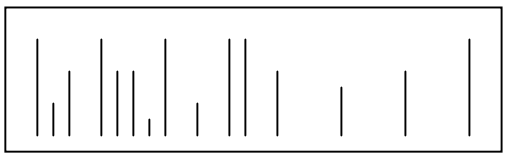
Brain proteins modulated by soy isoflavones:

	Change
Enolase	→
Dihydropyrimidinase-related protein-2	↓
Hipp. Chol.Neurostim.Pept. Precursor	↑

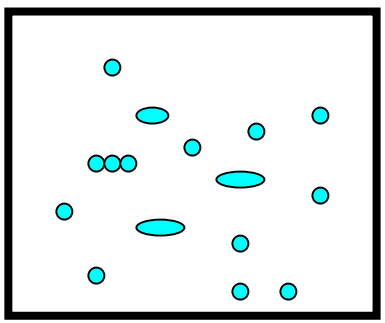
When to use which technology?



Chip: To identify all possible targets of a drug, or ligand, or protein; HIGH THROUGHPUT



LC/LC MS/MS: To catalogue a new Proteome; HIGH THROUGHPUT



2-D gels (& MS):

- **Use in conjunction with chip data;**
- **Is the only method that shows the relative levels of expression of proteins;**
- **Is the only method that readily indicates posttranslational modifications;**
- ***Least high throughput, but most informative?***

Goals of part 2 of this tutorial

- **To appreciate the two major protein/peptide ionization methods and how they can be connected to your research problem**
  - ***Electrospray ionization (ESI)***
  - ***Matrix-assisted laser desorption ionization (MALDI)***

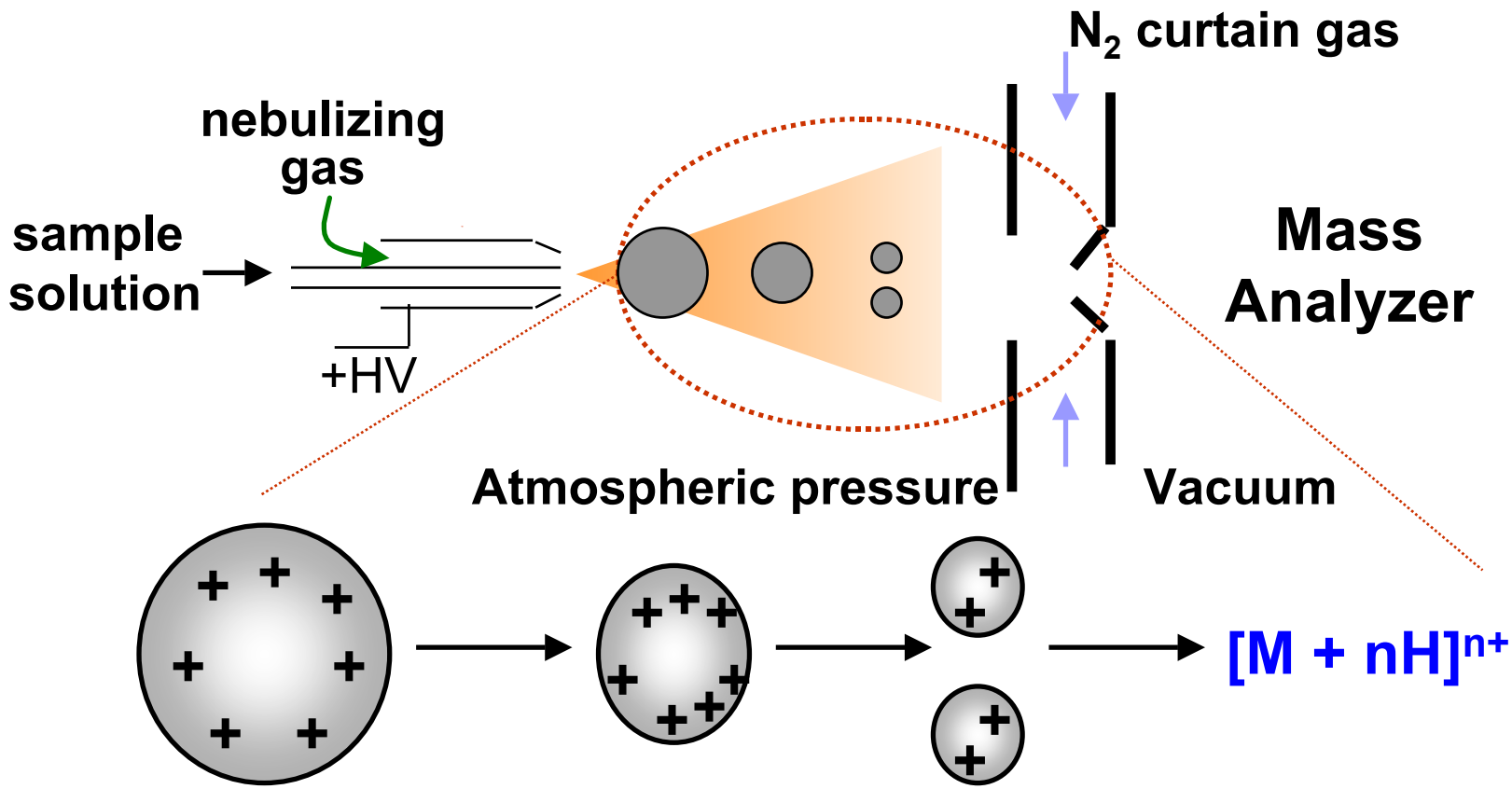
- **How to identify a protein, determine the sequence of a peptide and the site(s) of its modification**
  - ***Database searching***
  - ***Collision-induced dissociation and MS-MS analysis***

1990's revolution in protein sequencing by mass spectrometry

Due to several major factors

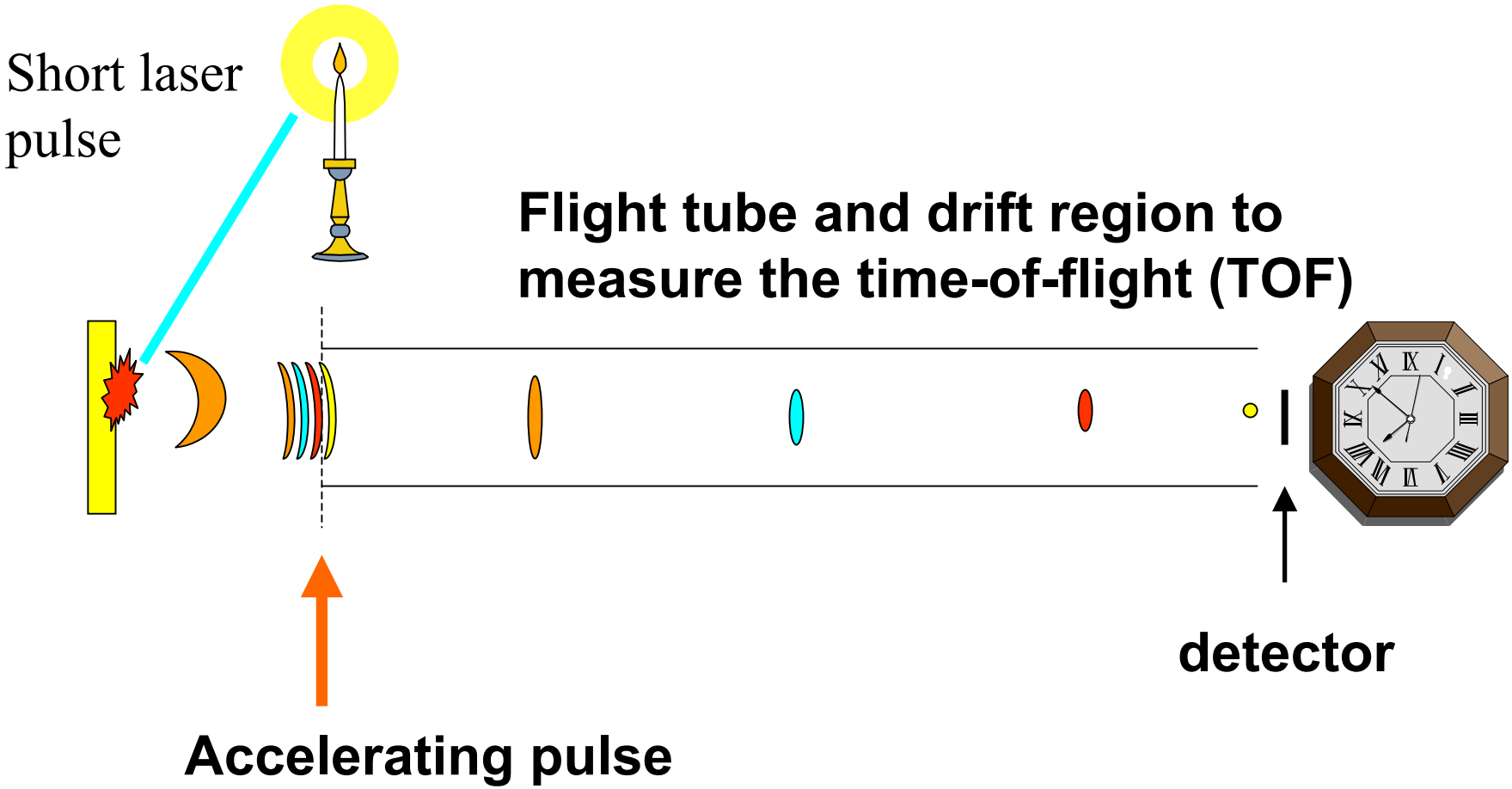
- ***The development of protein/peptide mass spectrometry ionization methods***
- ***The cataloging of many genomes***
- ***Speed and sensitivity of mass spectrometry***

Electrospray Ionization (ESI)

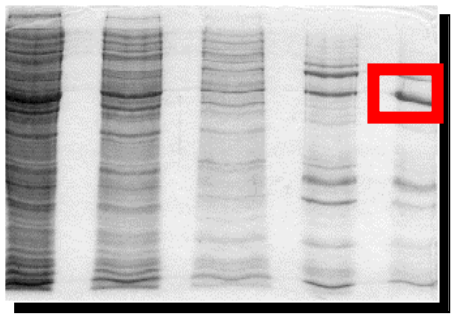


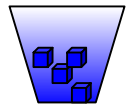
1. Solvent evaporation
2. Coulombic repulsion

Matrix-Assisted Laser Desorption Ionization (MALDI)



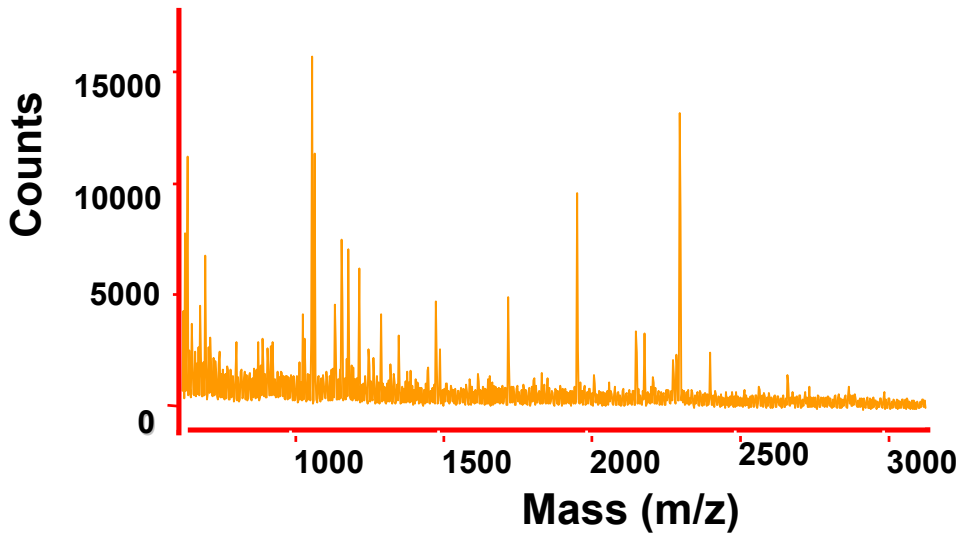
MALDI-TOF analysis of a “tryptic peptide fingerprint” generated from a spot on a gel



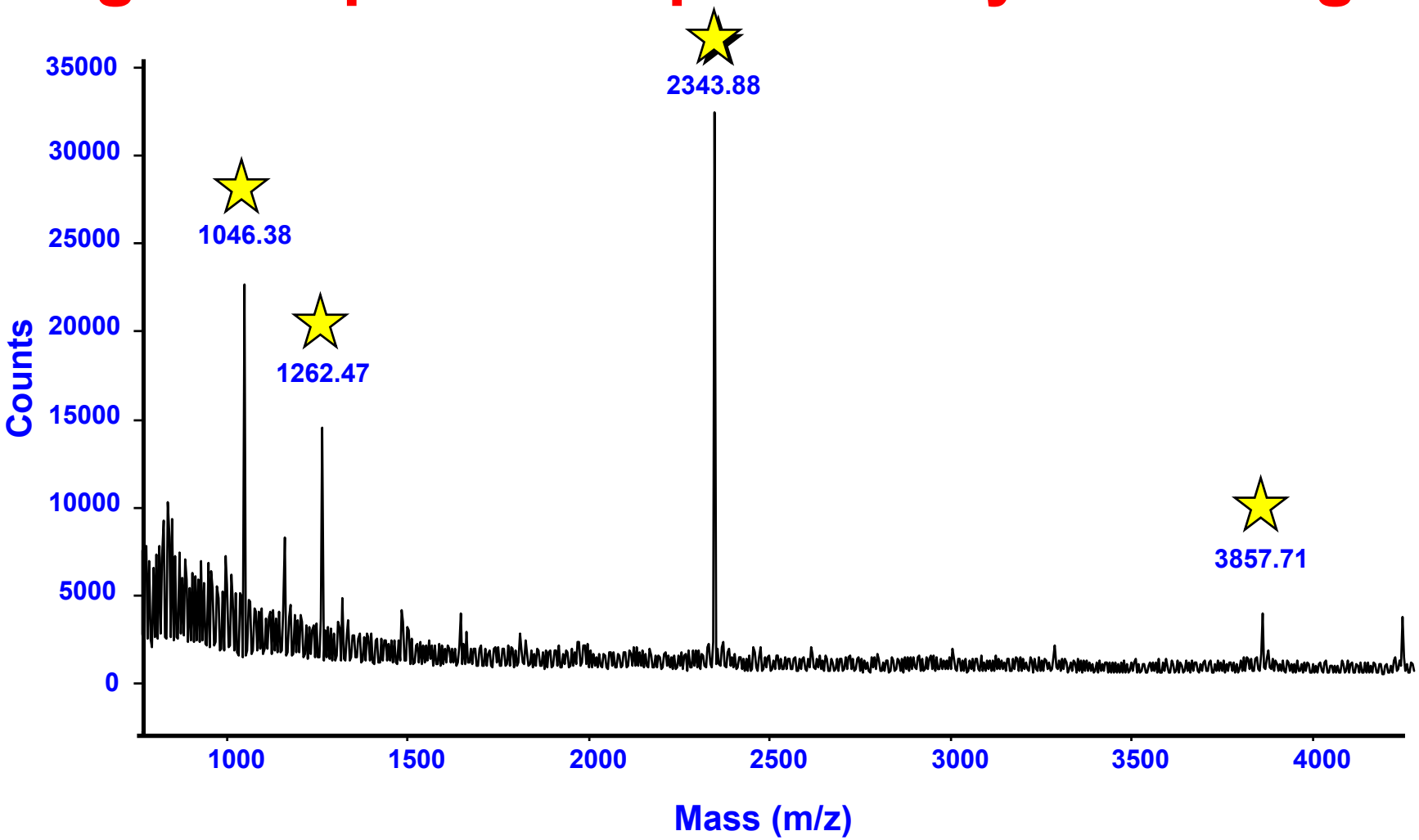


trypsin

MALDI-TOF mass spectrometry



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



From Proteins to Sequence Tags

- If each protein (average 500 residues) had a cleavage site every 10 residues, then about 1.5-3.0 million peptides describe the expressed products of the human genome
- Each peptide has a molecular weight value that is its individual sequence tag
- Any modification will increase the peptide's molecular weight, e.g., a nitro group adds 45

Websites for identifying proteins from peptides

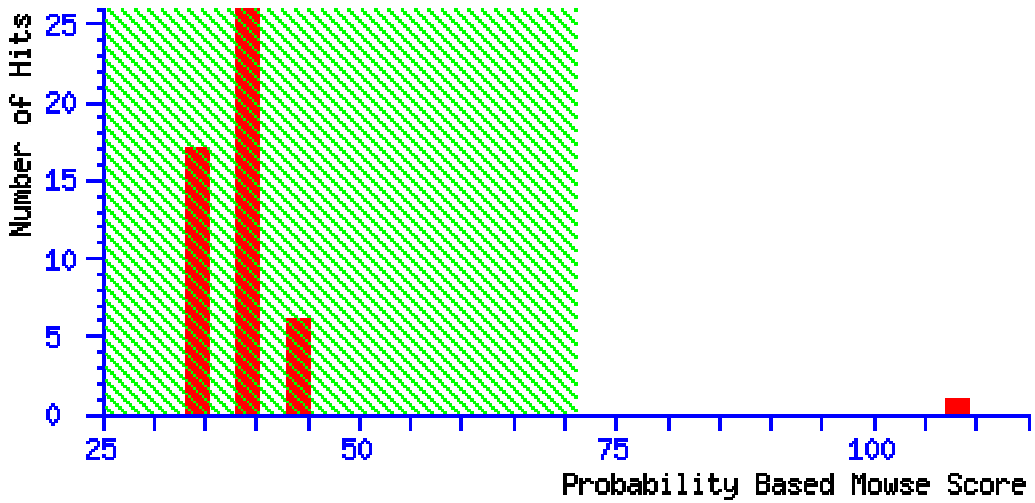
- **Analyze the peaks detected by MALDI or electrospray at the following:**

http://www.matrixscience.com (MASCOT)
http://Prowl.rockefeller.edu (PROWL)
http://donatello.ucsf.edu/ (PROTEIN PROSPECTOR)

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

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

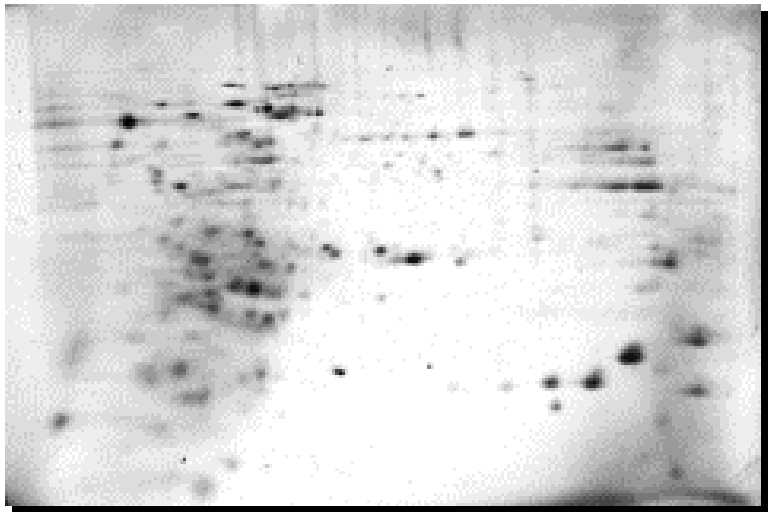
Nominal mass of protein (Mr): 20840

```
  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

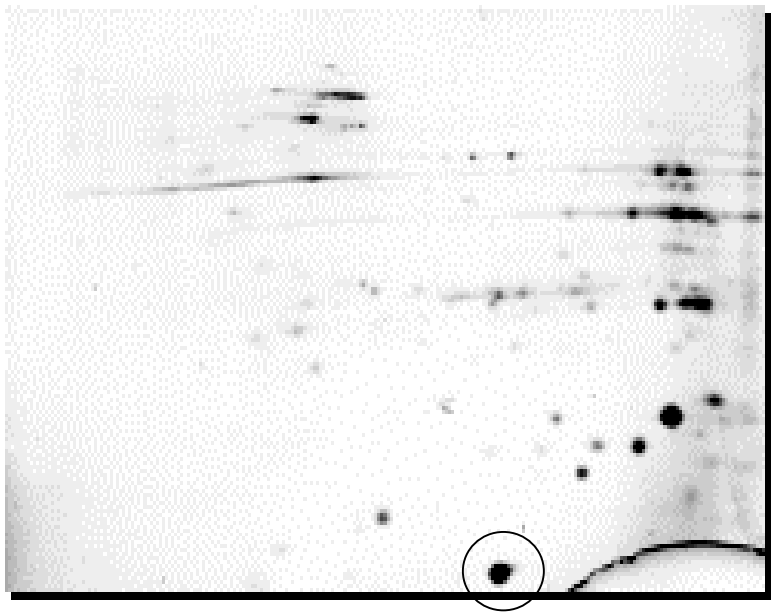
```
132-140 FNVEVVAIR
  6- 16 DLVVSLAYQVR
 58- 78 FDVAVGANDAYGQYDENLVQR
 96-131 FLAETDQGPVPVEITAVEDDHVVVDGNHMLAGQNLK
```

Human leukemic cell line (HL-60)



DMSO-induced differentiation

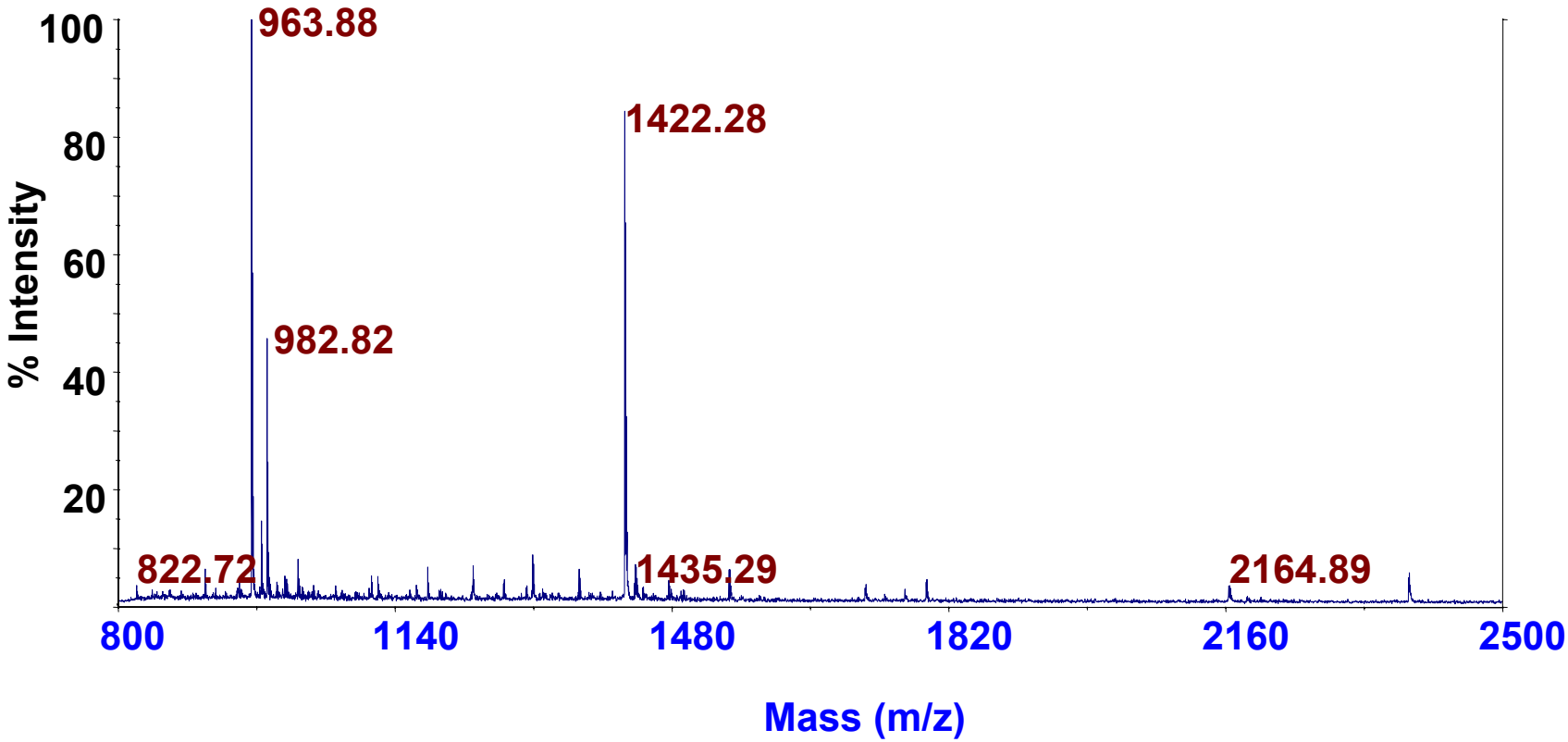
Differentiated neutrophil-like cells



New spot

D'Alessandro

MALDI-TOF analysis of trypsinized spot from 2D-IEF/SDS-PAGE analysis of DMSO-differentiated HL-60 cells



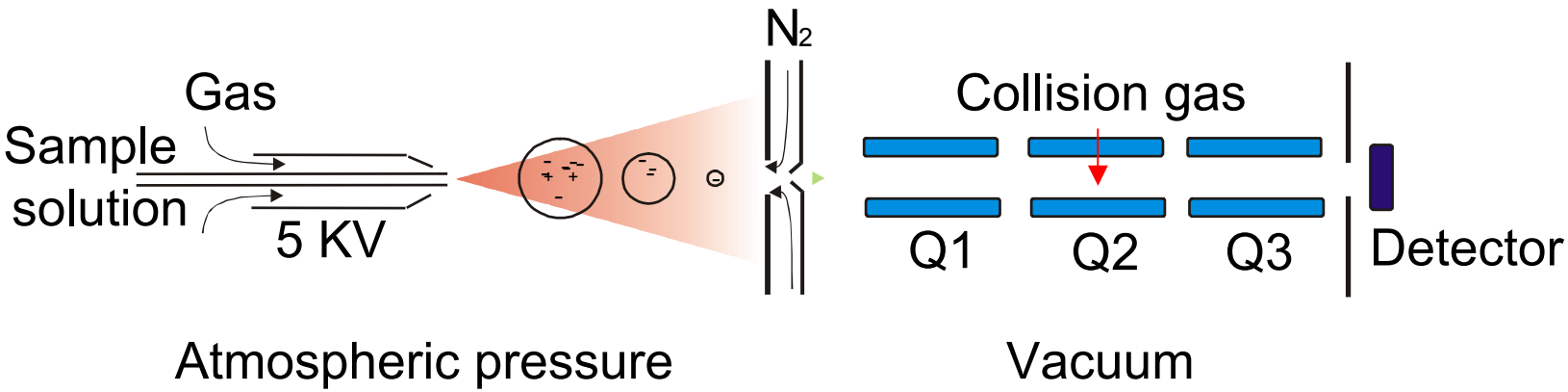
DMSO-treated HL-60 cell spot analysis

- DMSO-induced spot was shown to be S-100 (or calgranulin A) - a calcium binding protein
- BLAST showed that the sequence of S-100 is shared by migratory inhibitory factor related protein 8
- Two of these entries have a pdb entry (4-letter alphanumeric descriptor) - this means there is a molecular structure available

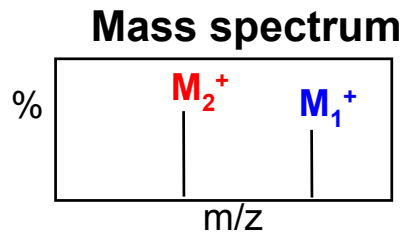
Options in proteomics analysis

- **Methods for protein separation and identification**
  - ***2D-electrophoresis of proteins***
  - **Reverse phase nanoLC-MSMS of peptides**
  - **Ion exchange/reverse phase LC-LC-ESI-MSMS**
  - **Isotope-coded affinity tagging LC-ESI-MSMS**
  - **CE- or reverse phase nanoLC/MALDI-TOF-MS**
- **Radical methods on the horizon**

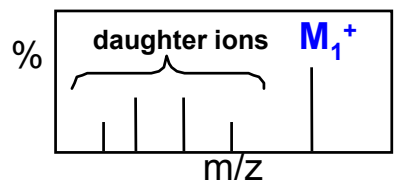
LC-ESI-MS Analysis



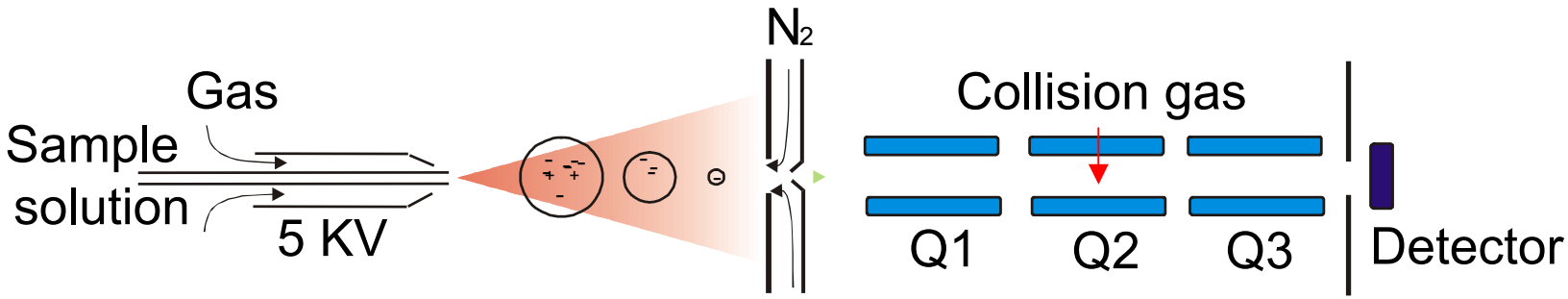
MS M_1^+ M_2^+ **Q1** → **Q2 and Q3 set to allow ions pass**



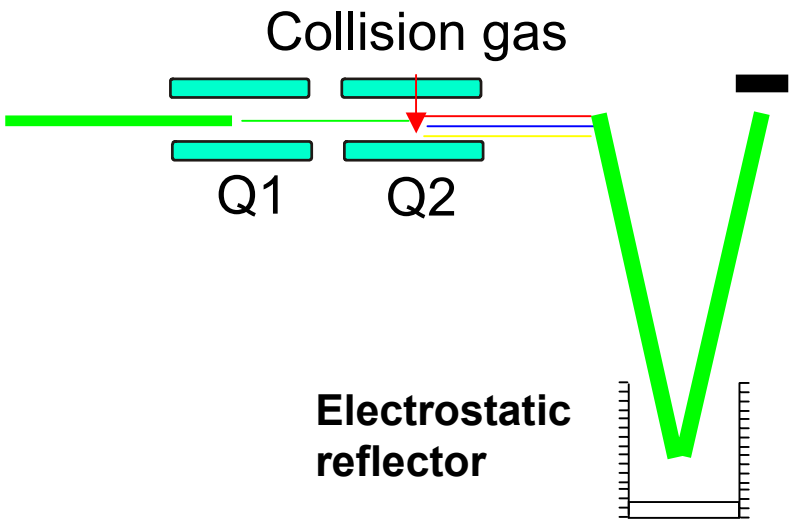
MS 2 M_1^+ M_2^+ **Q1** → M_1^+ → **Q2 fragmentation** (**Ar** (collision gas)) → **daughter fragment ions** → **Q3**



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

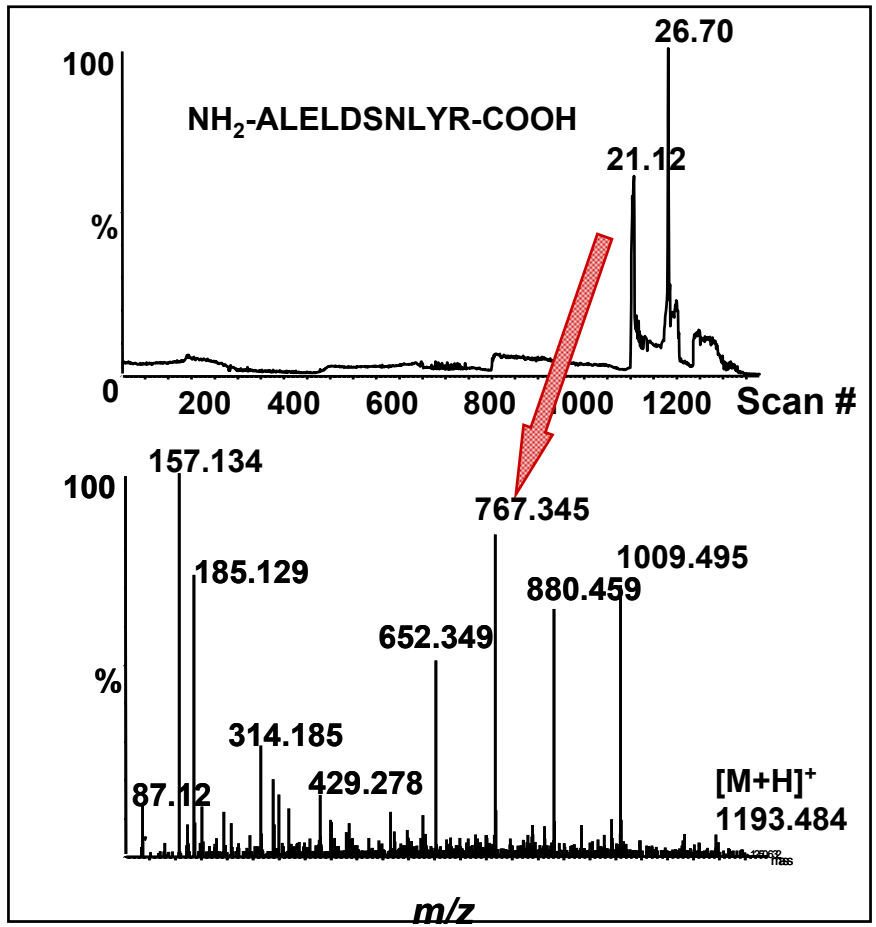
Also gives far greater mass accuracy - from 1000 ppm on the triple quad to 5-10 ppm on the Q-tof

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

Reverse phase nanoLC-MSMS

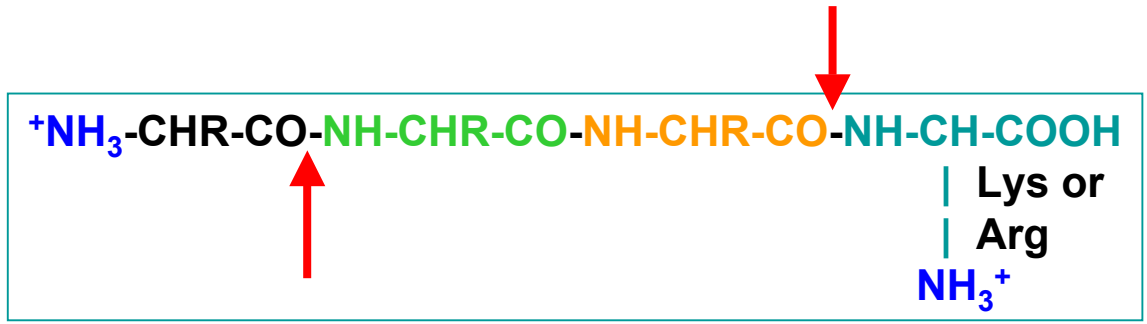
- Peptides are separated on an acetonitrile gradient using columns with i.d.s of 0.05-0.30 mm. These operate at 200-2000 nL/min

- Peptides are introduced by electrospray and analyzed on a Qqtof. Ions are selected by a quadrupole filter, collision-dissociated and analyzed by time-of-flight (accuracy 5-10 ppm)

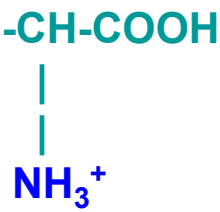


Fragmentation of peptide ions

Tryptic peptides are charged at each end

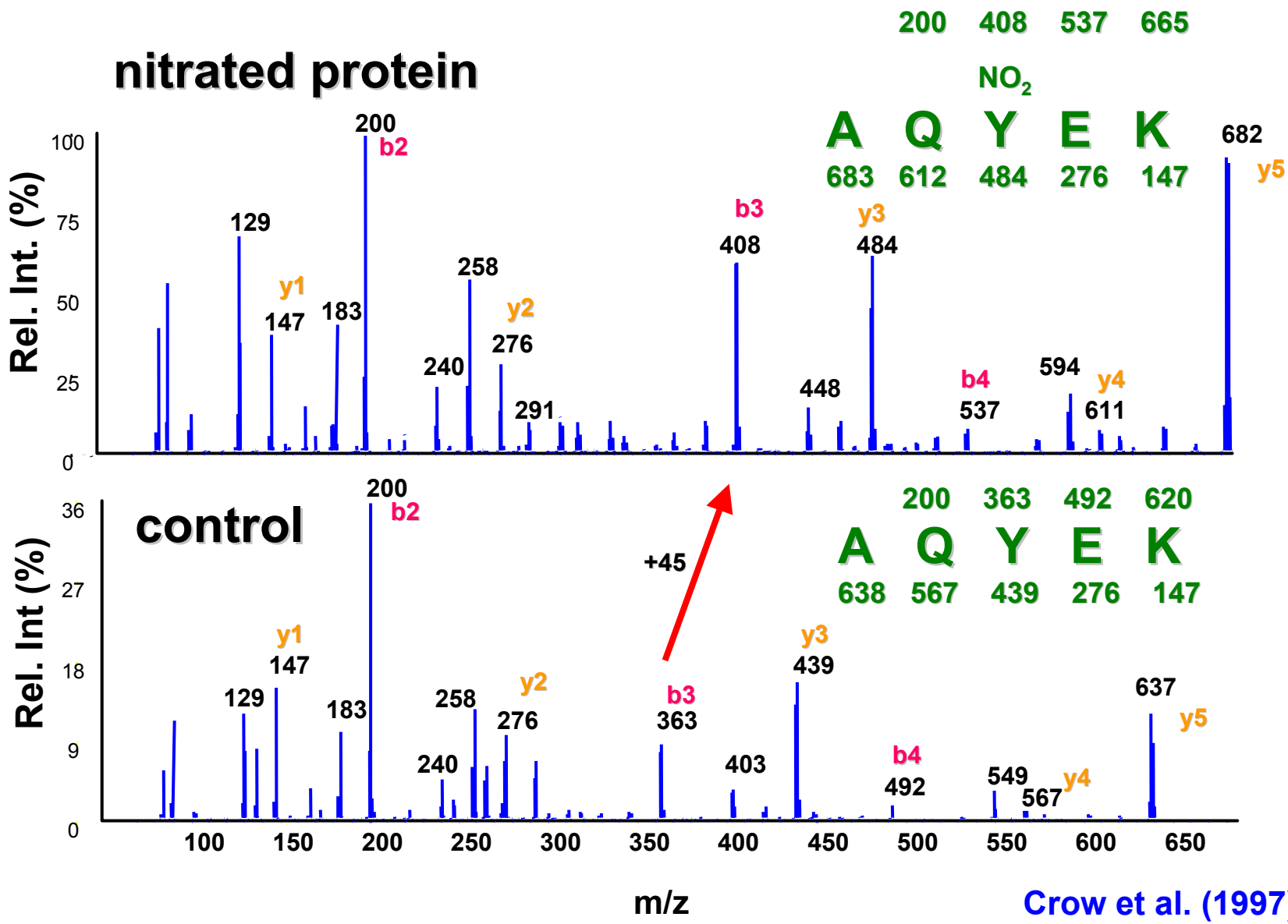


b1 ion



y1 ion

Site-specific nitration of a tyrosine-containing peptide using CID MS-MS spectra



Crow et al. (1997)

Limitations of proteomics

- Unlike its mRNA counterpart, proteomics doesn't have a PCR equivalent
- It's limited by Avogadro's number - 1 fmol is 6 x 10^8 molecular ions
- If a cell has a 100 copies of a protein, then at a minimum you need to have 6 x 10^6 cells - in reality, you need much more

So, where do we go next?

- Which are the players in the proteome of your system?
- How and where is your favorite protein modified?
- How does the modification affect the biochemical and biological properties of this protein?
- Do the other proteins care?

Acknowledgements

Kim Lab:
Lisa Chaves
Patti Hall
Robert Mills
Kiran Varma

Barnes' Lab
Brenda Boersma
Mindan Sfakianos
Tracy D'Alessandro
Amanda Foxwell
Steven Burke
Michelle Smith

Colleagues:
Virginia M.Y.Lee (U. Penn.)
Lester I. Binder (Northwestern U.
Karen Hsiao (U. Minn.)
Tom Clarkson (Wake Forest)
Y.L. Pan (Wake Forest)
Peter Prevelige (UAB)
Joe Beckman (Oregon St.)

Proteomics & MS Facility:
Tivanka DeSilva
Lori Coward
Marion Kirk
Scott Isbell
Ramu Vempati

Support:
NCI
NIDDK
NIH-NCCAM-ODS
United Soybean Board
UAB-HSF-GEF