

## **BMG 744 Proteomics-Mass Spectrometry**

# **Qualitative burrowing of the proteome**

**Stephen Barnes, PhD**  
[sbarnes@uab.edu](mailto:sbarnes@uab.edu)

1/24/14

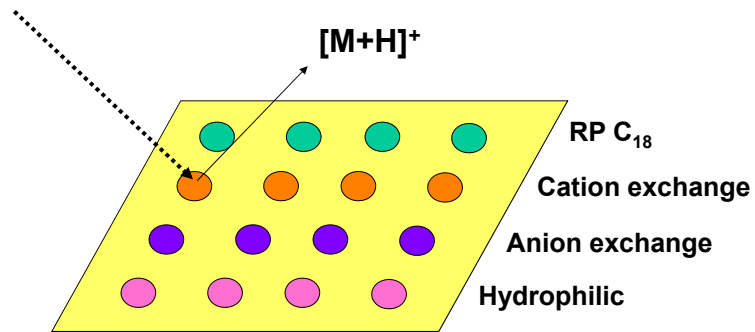
## **Topics**

### **Qualitative changes in the predicted proteome**

- Digging deeper - SELDI/MALDI
- Posttranslational modifications
  - Isolation and characterization
  - Making use of the chemistry of the modification
  - Ubiquitination/SUMOylation
  - Global discovery
  - Targeted discovery

1/24/14

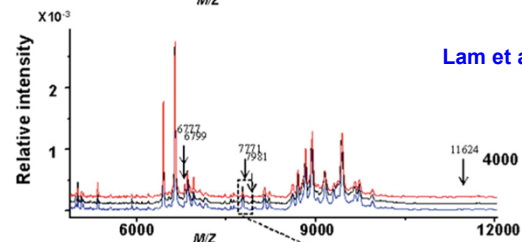
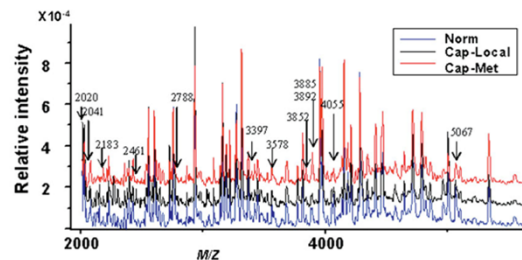
# Surface enhanced laser desorption ionization (SELDI)



SELDI chip (plate) with local chemistry on surface causing selective binding of peptides

1/24/14

## MALDI-TOF analysis of peptides recovered by C<sub>18</sub> extraction of sera from prostate patients, with and without metastases, and controls

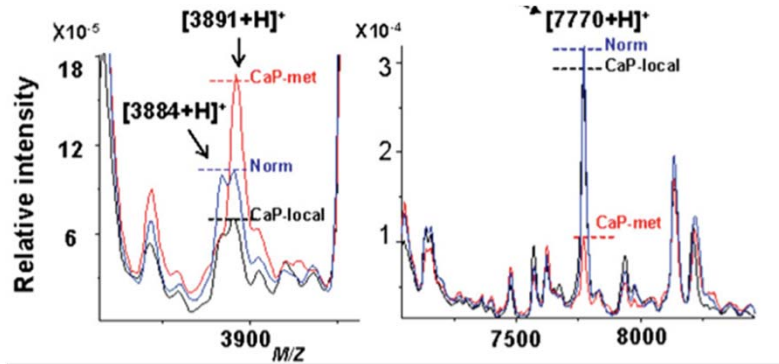


Lam et al., Proteomics 5, 2927

1/24/14

## Identification of a significantly altered peptide in men with metastatic prostate cancer

Lam et al., Proteomics 5, 2927

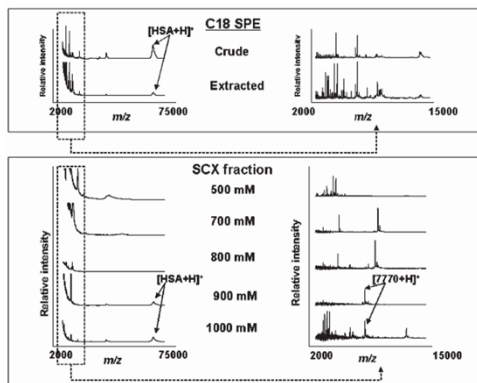


But what is it?

1/24/14

## Purification of prostate peptide by SPE extraction and strong cation exchange

Lam et al., Proteomics 5, 2927



Serum was first extracted with a  $C_{18}$  SPE cartridge

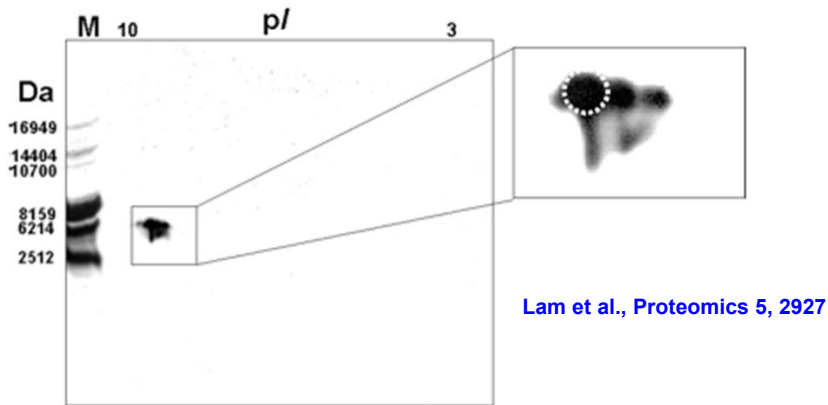
SPE fraction was passed over a  $[H^+]$ -form cation exchange resin. Bound peptides were eluted with increasing step gradients of NaCl

Analyses of fractions were performed by MALDI-TOF MS

The most enriched fraction was contaminated with albumin

1/24/14

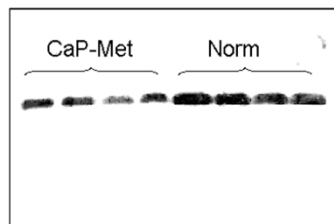
## Purification of prostate peptide needed 2D-electrophoresis



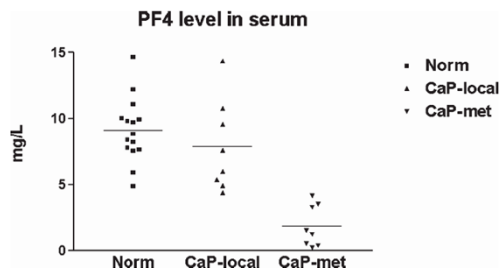
1/24/14

## Non-mass spec verification of prostate peptide

Lam et al., Proteomics 5, 2927



Western blot analysis for PF4



ELISA analysis for PF4

**Moral: proteomics is a serious business that requires multiple dimensions of separation - glib methods don't work**

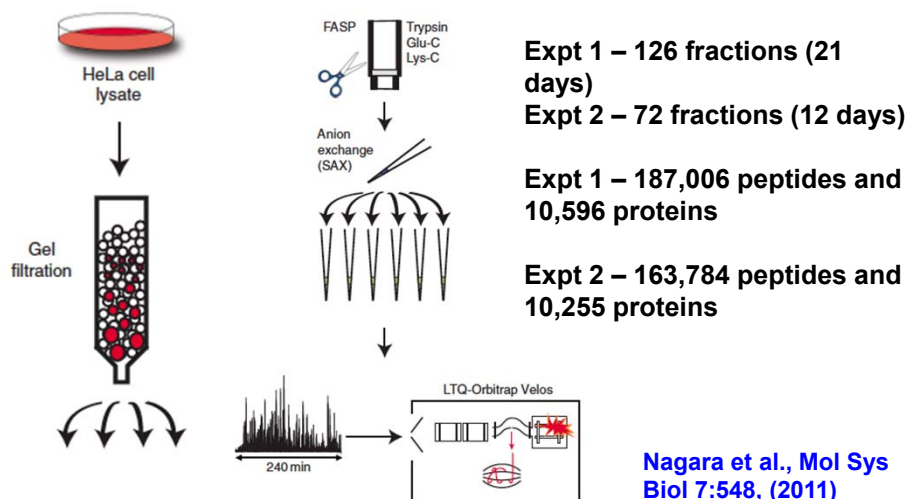
1/24/14

## Challenges in proteomics

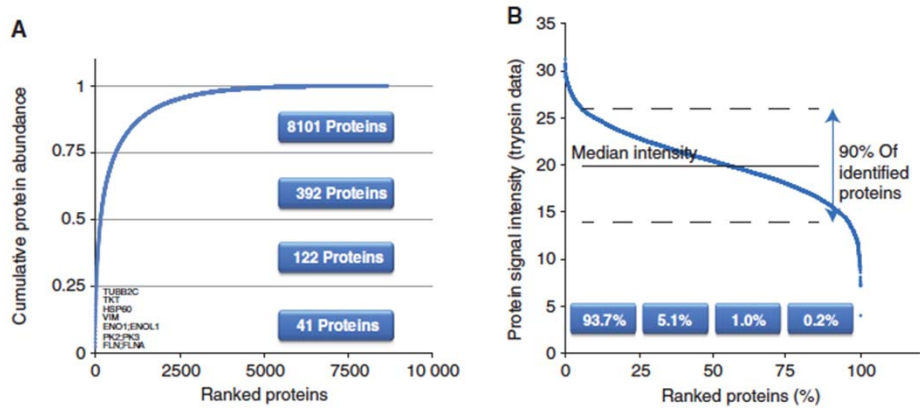
- RNA can be amplified by PCR and converted to DNA (cDNA) - then it has been used to hybridize to DNA sequences on a microarray
  - Proteomics doesn't have an amplification system
- RNA-Seq sequences rather than hybridizes and dispenses with predetermined microarrays
  - Discovers new messages - signals not predicted by conventional gene annotations

1/24/14

## How deep can we go?



## Description of protein abundances

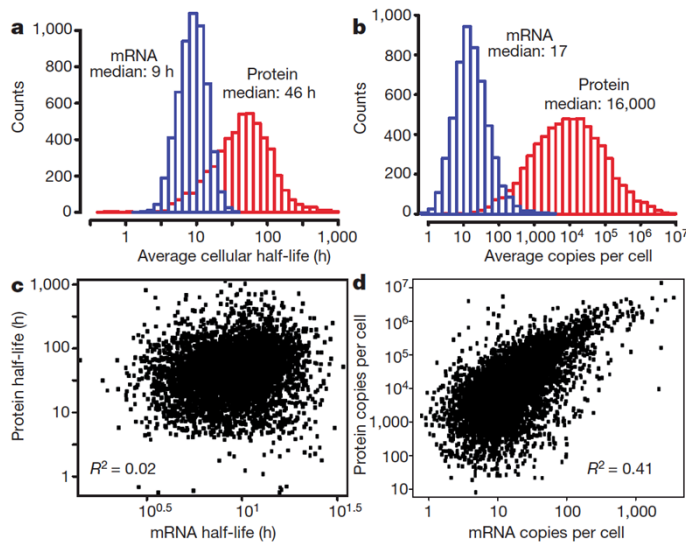


Just 41 proteins accounted for 25% of the combined tryptic peptide ion intensities – 122 accounted for 50% and 392 for 75%.

1/24/14

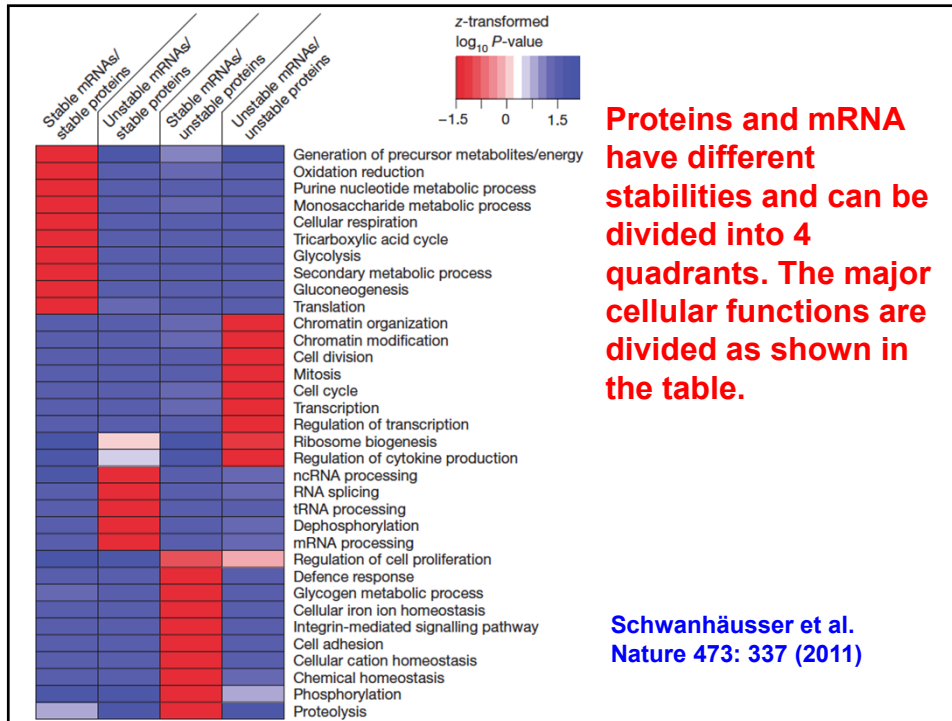
Nagara et al., Mol Sys Biol 7:548, (2011)

## Protein – mRNA correlations



1/24/14

Schwanhäusser et al. Nature 473: 337 (2011)



## Moral about protein discovery in 2014

- It is now possible to detect more proteins than there are separate forms of message
- However, such discovery is still “heroic”
- Proteomics does not have the equivalent of the PCR reaction
  - However, a cell does amplification of message
- Detecting the minor components requires being able to start with enough sample
  - 10 attomole is 6.023 million molecules
  - If present as 10 copies per cell, then minimum # of cells needed is  $0.6 \times 10^5$

1/24/14

## How to dig deeper

- **Sub-fractionate the sample being analyzed to enrich for the lower abundance components**
  - Choose the right biofluid/tissue
  - Organelle isolation
  - Affinity chromatography
  - Antibody pulldowns
  - Abundant protein(s) depletion

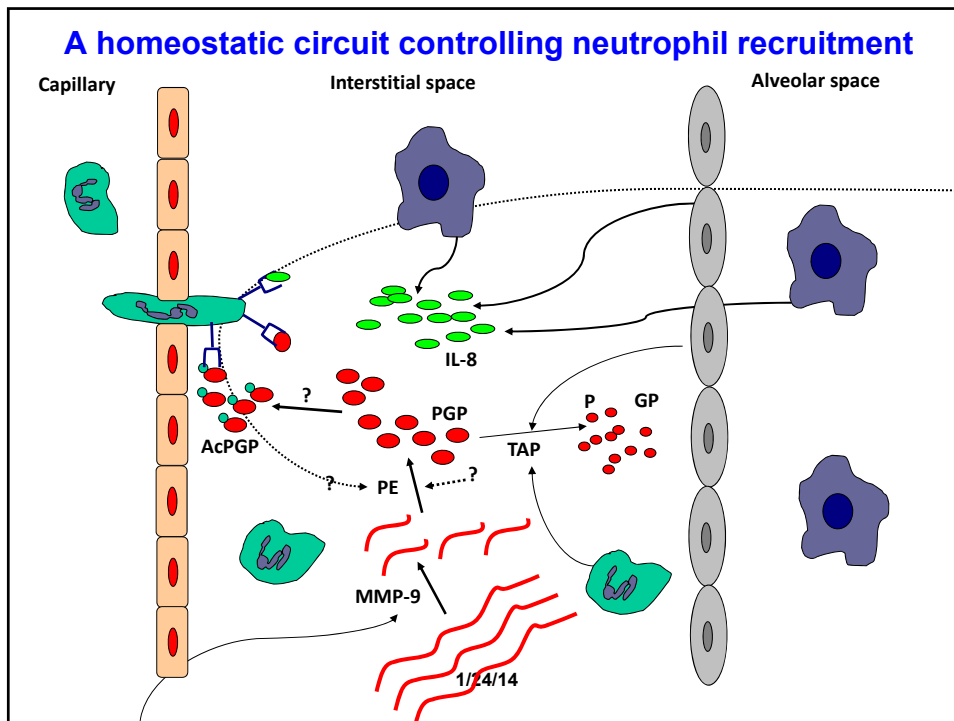
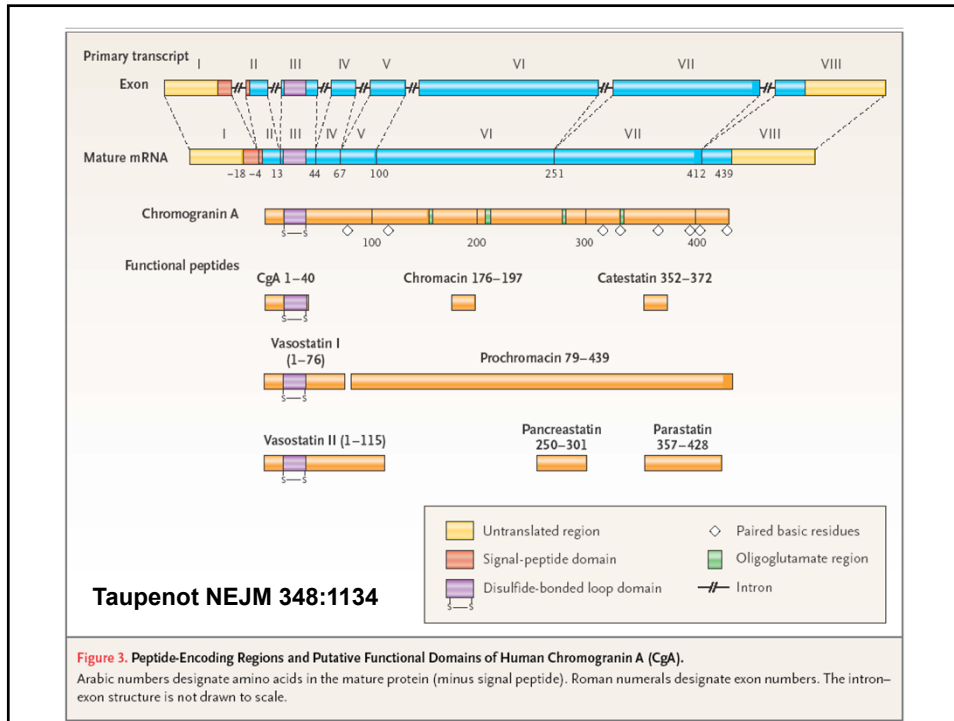
1/24/14

## General classes of modification

- **Biochemical events involving peptide processing**
- **Biochemical events stimulated by enzymes**
- **Chemical events driven by reactive species**
- **Chemical events determined by investigator**

1/24/14

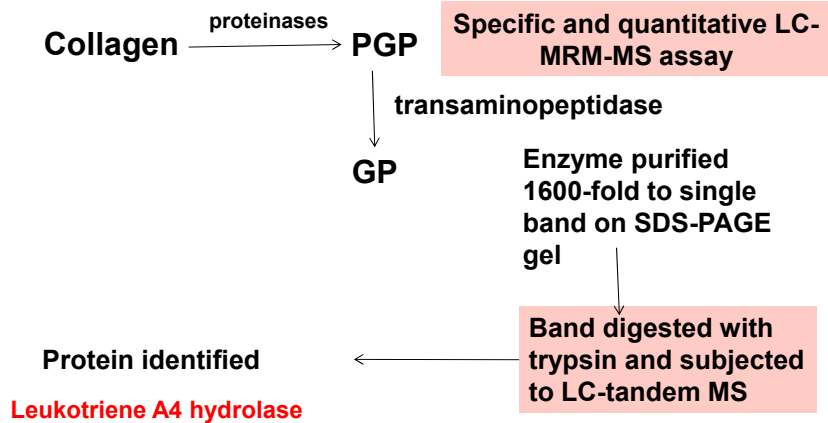




## PGP is a common peptide in human collagen

MFSFVDLRLLLLLAATALLTHGQEEGQVEGQDEDEIPPIITCVQNGRLRYHDRDVKPEPCRI  
 CVCDNGKVLCDDDVICDETKNCPGAEVPEGECCPVC PDGSESP TDQETTGVVEGPKGDTGPR  
 GPRGPAGPPGRDGI PGQPGL **PGPPGPPGPPGPP** PGLGGNFAPQLSYGYDEKSTGGISV **PGP**  
 MGPSGPRGL **PGPPGAPGP** QGFQGP PGEPGEPGASGPMGRGP **PGP** PGKNGDDGEAGKPGR  
 PGERGP **PGP** QGARGLPGTAGLPGMKGHRGFSGLDGAKGDAGPAGPKGEPGSPGENGAPGQ  
 MGPRGLPGERGRPGA **PGP** AGARGNDGATGAAGP **PGP** TGPAGPPGFP GAVGAKGEAGPQGP  
 RGSEGPQGV RGE **PGPPGPA** GAAGPAGNPGADGQP GAKGANGAPGIAGAPGFPGARGPSG  
 QGPGGP **PGP** KNGSGEPGAPGSKGDTGAKGE **PGP** VGVQGP **PGP** AGEEGKR GARGE **PGP** TGL  
**PGP** PGERGGPSRGFP GADGVAGPKGPAGERGS **PGP** AGPKGSPGEAGRPEAGLP GAKGL  
 TGSPGS **PGP** DGKTGP **PGP** AGQDGR **PGPPGP** PGARGQAGVMGF **PGP** KAAGEPGKAGERGV  
**PGP** PGAVGPAGKDGEAGA QGP **PGP** AGPAGERGEQGPAGSPGFQGL **PGP** AGPPGEAGKPGE  
 QGVPGDLGA **PGP** SGARGERGF PGERGVQGP **PGP** AGPRGANGAPGNDGAKGDAGAPGAPGS  
 QGAPLQGM PGERGAAGL **PGPKG** DRGDAGPKGADGSPKDGVRGLTGPIGPPGAPAGPD  
 KGESGSPGAPGTGARGAPGDRGE **PGPPGP** AGFAGPPGADQGP GAKGEPGDAGAKGDAGP  
**PGP** AGPAGP **PGP** IGVNVPAGAKGARGSAGPPGATGFP GAAGRVP **PGPSGNAGP** **PGPPGP**  
 AGKEGGKPRGETGPAGRPGEVGP **PGPPGP** AGEKGS PGADGPAGAPGT **PGP** QGIAGQGV  
 VGLPGQRGERGFPL **PGPS** GEPKQGPSGASGERGP **PGP** MGPPGLAGPPGESGREGAPGA  
 EGSPGRDGS PGAKDRGETGPAGPPGAPGAPGA **PGP** VPAGKSGDRGETGPAGPAGVGP  
 VGARGPAGQGP RGDKGETGEQDGRGIKHRGFSGLQGP **PGP** PGSPGEQGPSGASGPAGP  
 RGP PGSAGAPKDG LNL **PGP** IGP **PGPR** RGTGDAGPVGP **PGPPGPPGPPGPP** PSAGFDFS  
 LPQPPQEK AHDGGRYYRADDANVVRDRDLEVDTT LKSLSQIENIRSPESGRKNPARTCR  
 DLKMCHSDWKS GEYWIDPNQGCNLDAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDK  
 KRHVWFGESMTDGFQFEYGGQSDPADVAIQ LTLFLRLMSTEASQNTIYHCKNSVAYMDQQ  
 TGNLKKALLLQGSNEIEIRAEGNSRFTYSVTVDGCTSHTGAWGKTVIEYKTTKTSRLPII  
 DVAPLDVGPADQEF GFDVGPVCF L

## Mass spec and PGP



(Robert Snelgrove et al., Science 2010)

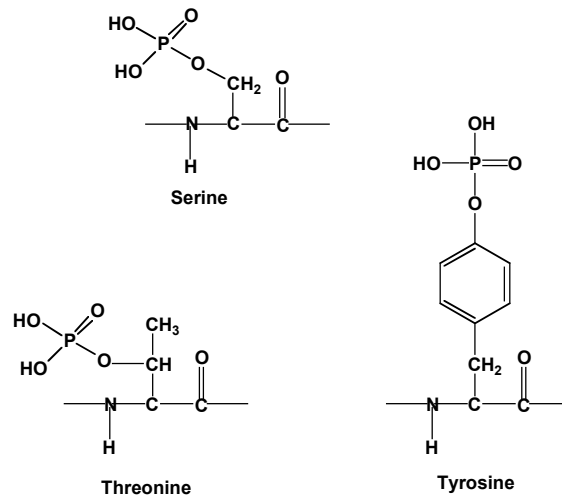
1/24/14

## Enzymatic modifications

- Phosphorylation/dephosphorylation
  - On serine, threonine, tyrosine (also histidine)
- Glycosylation
  - N-glycosylation (asparagine-linked)
  - O-glycosylation (serine-, threonine-linked)
- N-Acetylation/deacetylation
  - On lysines – also higher acyl groups
- N-Methylation/demethylation
  - On lysines - mono-, di- and trimethylation

1/24/14

## Chemistry of phosphorylation



1/24/14

## Mass spectrometry of phosphorylated proteins

- Adds  $\text{H}_3\text{PO}_4$  (+98)
- Eliminates water (-18)
- Net change +80
- if the phosphate ion is released (i.e., in negative ion spectra), it is seen as  $m/z$  79 and/or  $m/z$  63
- Phosphate can be confused with sulfate and bromide (sulfate, 79.9568 Da; phosphate, 79.9663 Da; and  $^{79}\text{Br}$ , 79.9083 Da)

1/24/14

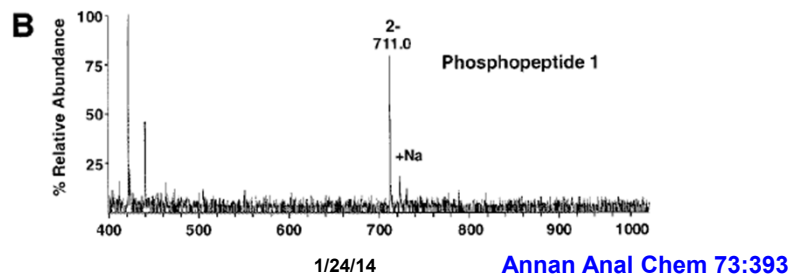
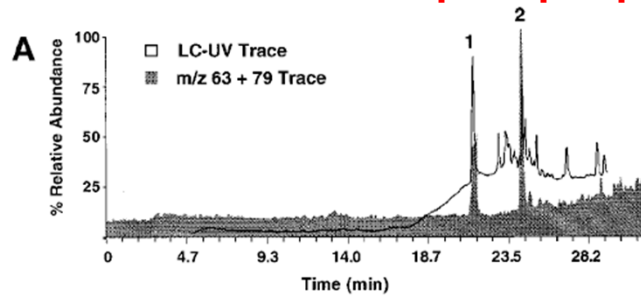
## Finding a phosphate group

Several methods are in use for detection of phosphopeptides

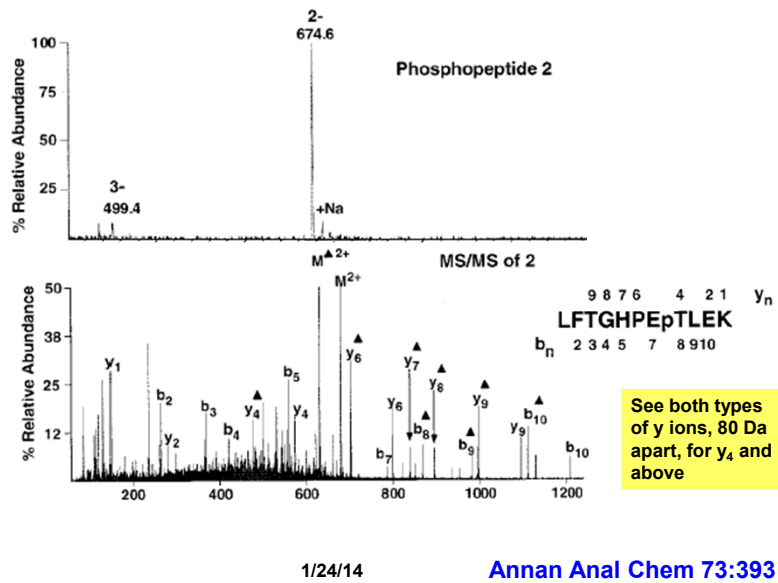
- use of parent ion or neutral loss scanning
- phosphatase sensitivity
- affinity methods for enrichment of phosphopeptides
  - anti-phospho-Ser/Thr/Tyr antibodies
  - metal ion affinity
  - chemical reaction/biotin affinity

1/24/14

## Parent ion scan to detect phosphopeptide



## Phosphopeptide 2

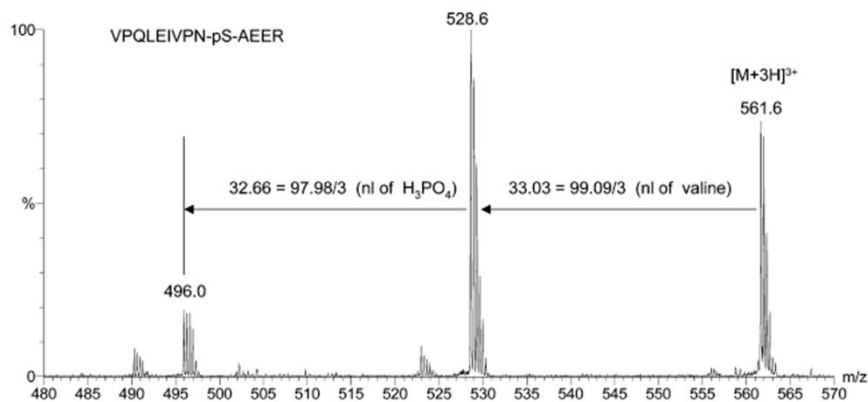


## Neutral loss for phosphopeptides

- Based on the loss of phosphoric acid ( $\text{H}_3\text{PO}_4$ ) - 97.98 Da
- If the peptide is doubly charged, then it's  $m/z$  48.88
- If it's triply charged, then it's  $m/z$  32.66
- Problems can occur if the peptide contains an N-terminal proline (97 Da) or valine (99 Da) and a low resolution instrument is used for the analysis

1/24/14

## Casein peptide - mono- or diphosphate?



Lehmann et al., J Prot Res 7-2866

1/24/14

## Recovery and enhancement of phosphopeptides

- The biggest problem in the detection of phospho-peptides is how to convert the initial sample matrix into a form suitable for mass spectrometry analysis.
  - how to handle minute samples with minimal losses
  - how to recover and detect all the phosphopeptides
  - how to recover and detect the non-phosphorylated proteins to determine the extent of phosphorylation at individual sites

1/24/14

## Antibodies and phosphopeptides

- In this approach, both the phosphorylated and non-phosphorylated forms of a protein may be recovered from the sample matrix
- This can most easily be achieved by immunoprecipitation of the protein with an antibody that recognizes epitope(s) that is(are) in common with both forms

**NOTE** *that as with all immunoprecipitation methods, the best results will be achieved if the antibody is coupled to agarose beads. This allows selective immuno-absorption of the antigen, washing free of contaminating proteins, AND following elution, minimization of the amount of antibody that is in the eluate and therefore would be analyzed by mass spectrometry*

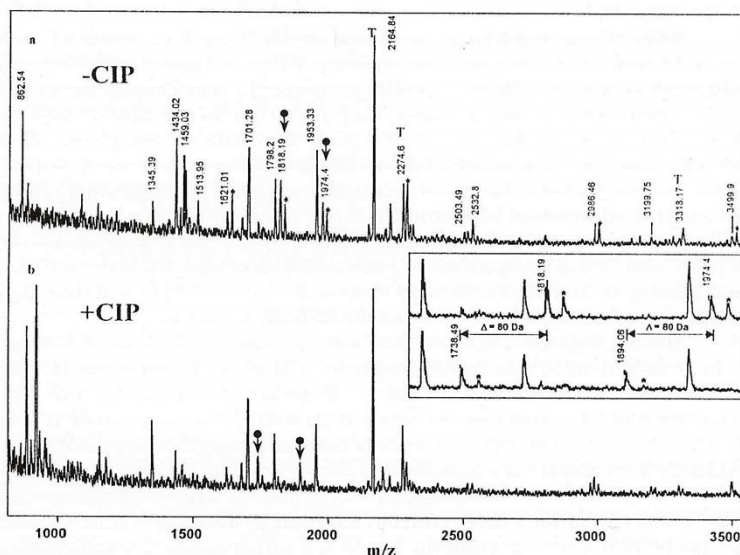
1/24/14

## Detection of phosphopeptides based on their sensitivity to phosphatase

- An alternative source of potentially phosphorylated proteins are individual spots on 2D-IEF/SDS gels. The protein preparation so isolated is either hydrolyzed by trypsin in solution (or in the gel piece) or using solid-phase trypsin
- One portion of the resulting tryptic peptides (in 50% acetonitrile:water) is analyzed by MALDI-TOF-MS. A second portion is diluted into 50 mM  $\text{NH}_4\text{HCO}_3$  buffer and reacted with 0.5 U calf intestinal alkaline phosphatase at 37°C for 30 min. Sample is dried with a SpeedVac, redissolved in 50% acetonitrile:water, and re-analyzed by MALDI-TOF-MS

1/24/14

## Detecting a phosphopeptide with alkaline phosphatase



1/24/14

Zhang et al., 2000  
MS in Biology & Medicine

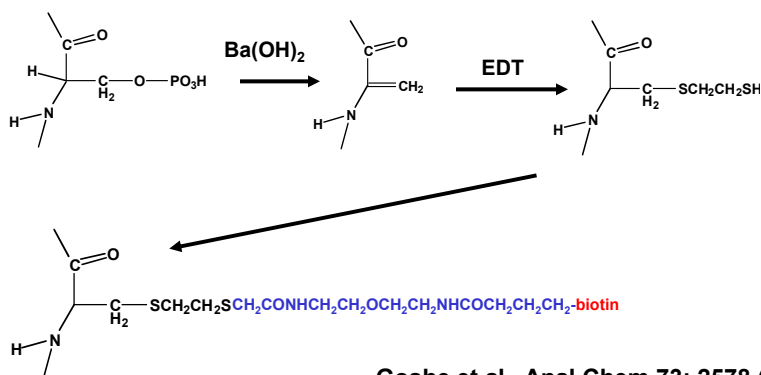


## Selective enhancement of phosphopeptides in tryptic digests

- Immobilized metal affinity chromatography (IMAC). Similar to Ni-affinity resins used in the purification of 6xHis-tagged proteins. The affinity phase can be charged with different metal ions (as their chlorides)
- Fe(III) and Ga(III), and to a lesser extent Zr(IV), were the most effective for the recovery of two synthetic phosphopeptides
- A tryptic digest containing both phosphorylated and non-phosphorylated peptides is passed over the IMAC column at acid pH (pH 2.5-3). The column is washed with 0.1 M acetic acid to remove unbound peptides. Elute with sodium phosphate (have to desalt) or with  $\text{NH}_4\text{OH}$
- Esterification may prevent Asp- or Glu-containing peptides from binding
- $\text{TiO}_2$  is now being used with success

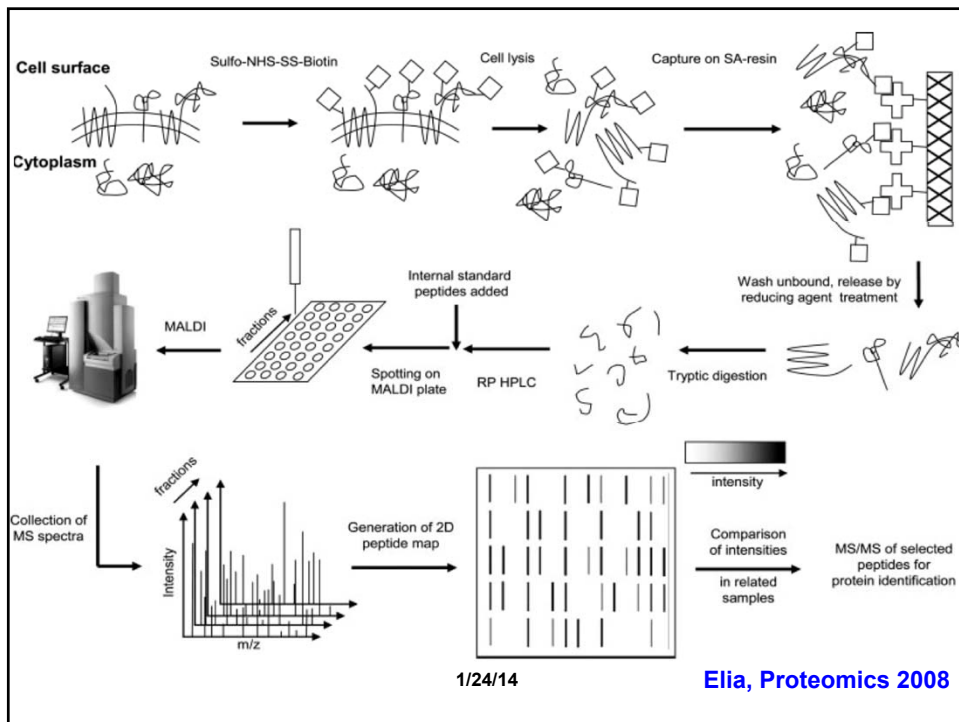
1/24/14

## Selective biotinylation of phospho-groups



Goshe et al., Anal Chem 73: 2578 (2001)  
Adamcyk et al., Rapid Commun Mass Spec 15; 1481 (2001)

1/24/14



## How to identify phosphorylated peaks by searching databases

- The databases you have used earlier in this class have some ability to predict the expected masses for a limited number of posttranslational modifications
  - MASCOT (<http://www.matrixscience.com>)
  - PROTEIN PROSPECTOR (<http://prospector.ucsf.edu/>)
  - Phosphosite is more focused <http://www.phosphosite.org/staticAboutPhosphosite.do>

1/23/12

# How to identify posttranslational modifications

FindMod at

<http://www.expasy.org/tools/findmod/>

It examines mass fingerprinting data for mass differences between empirical and theoretical peptides. If the mass difference corresponds to a known modification, it also makes intelligent guesses as to the site of modification.

1/24/14

# Unimod

- As Matt Renfrow described on January 17<sup>th</sup>, Unimod – ([http://www.unimod.org/modifications\\_list.php](http://www.unimod.org/modifications_list.php)) provides a more detailed and more accurate list of possible modifications

Unimod Logged as Guest Log out Change password Advanced search

Search for: Any field Contains Search Details found: 956 Page 1 of 48 Records Per Page: 20

| Accession #               | PSI-MS Name | Interim name | Description   | Monoisotopic mass | Average mass | Composition                   |
|---------------------------|-------------|--------------|---|-------------------|--------------|-------------------------------|
| <a href="#">View</a> 1287 |             | Arg-loss     | Loss of arginine due to transpeptidation                | -156.101111       | -156.1857    | H(-12) C(-6) N(-4) O(-1)      |
| <a href="#">View</a> 765  |             | Met-loss     | Removal of initiator methionine from protein N-terminus | -131.040485       | -131.1961    | H(-9) C(-5) N(-1) O(-1) S(-1) |
| <a href="#">View</a> 676  |             | Trp->Gly     | Trp->Gly substitution                                   | -129.057849       | -129.1586    | H(-7) C(-9) N(-1)             |
| <a href="#">View</a> 313  | Lys-loss    | -K           | Loss of C-terminal K from Heavy Chain of MAb            | -128.094963       | -128.1723    | H(-12) C(-6) N(-2) O(-1)      |

1/24/14

# FindMod

<http://web.expasy.org/findmod/>

Some of the modifications considered are:

|                                |                             |                             |
|--------------------------------|-----------------------------|-----------------------------|
| acetylation                    | amidation                   | biotinylation               |
| carboxymethyl cysteine         | carboxyamidomethyl cysteine |                             |
| C-mannosylation                | deamidation                 | flavinilylation             |
| farnesylation                  | formylation                 | geranyl-geranylation        |
| $\gamma$ -carboxyglutamic acid | hydroxylation               | lipoylation                 |
| methylation                    | myristoylation              | N-acyl diglyceride          |
| O-GlcNac                       | palmitoylation              | phosphorylation             |
| pyridoxal phosphate            | phospho-pantetheine         | pyrrolidone-carboxylic acid |
| sulfation                      |                             |                             |

Wilkins M.R., Gasteiger E., Gooley A., Herbert B., Molloy M.P., Binz P.A., Ou K., Sanchez J.-C., Bairoch A., Williams K.L., Hochstrasser D.F. *High-throughput Mass Spectrometric Discovery of Protein Post-translational Modifications*. [Journal of Molecular Biology, 289, p. 645-657 \(1999\)](#)

Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A.; *Protein Identification and Analysis Tools on the ExPASy Server*; (In) John M. Walker (ed): [The Proteomics Protocols Handbook. Humana Press \(2005\)](#).

1/24/14

# Site for compilation of PTMs

[https://www.abrf.org/index.cfm/dm.home\\_](https://www.abrf.org/index.cfm/dm.home_)

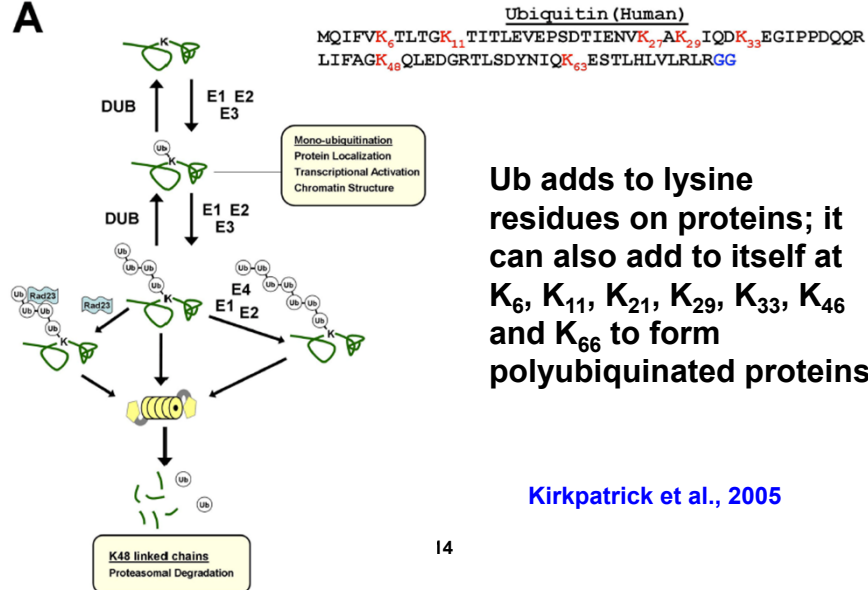
This site was put together by Ken Mitchelhill, Len Packman and friends

Currently ranges from dephospho (-79) to (Hex)3-HexNAc-(dHex)HexNAc (+1,039)

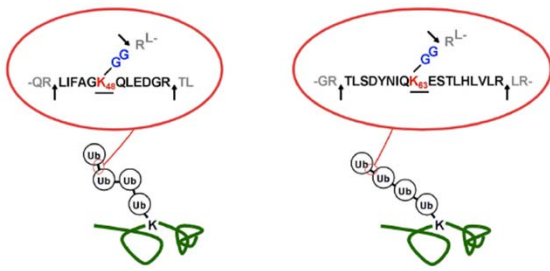
1/24/14

# Ubiquitination

A



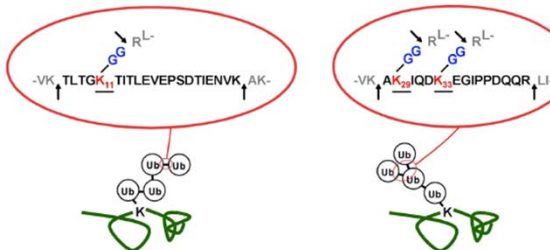
## Sites of attachment of Ub groups



**K48 Chains**

**K63 Chains**

Note that Ub is attached through the side chain of lysine via the two C-terminal glycines of Ubq.

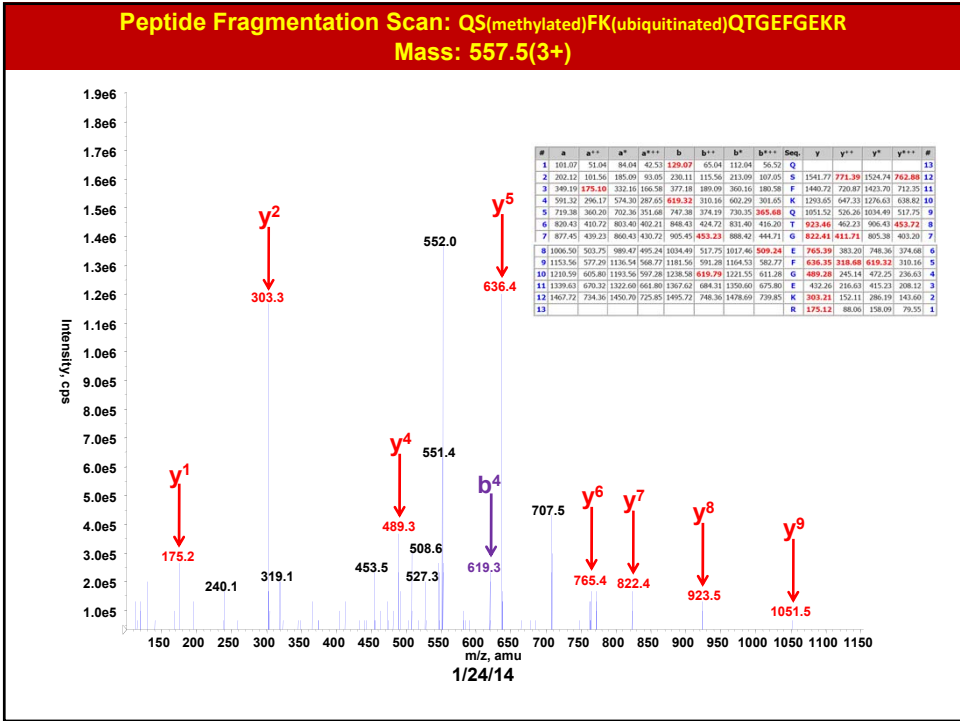


**K11 Chains**

**K29-33 Fork**

The next residue is Arg – therefore, trypsin hydrolysis will yield a peptide modified by the addition of 114 (*m/z* 261)

Kirkpatrick et al., 2005



**SUMO-modification**

- Protein Localization
- Transcriptional Activation
- Chromatin Structure
- Prevent Ubiquitination
- Poly-Sumoylation

**Others**

- Immune Response (ISG15)
- Ubiquitination (Nedd8)
- Morphogenesis (Hub1)

## SUMOylation

There are ubiquitin-like proteins (Ubls), one of which is Small Ubiquitin-like Modifier (SUMO)

Have different sequence, but similar structure. Recognize the consensus motif  $\Psi$ -K-x-D/E for SUMOylation. Connected via a Gly-Gly group to the lysine residue (just like ubiquitin)

<http://sumosp.biocuckoo.org/faq.php>

1/24/14 Kirkpatrick et al., 2005

## Ubiquitin versus SUMO

```
Human Ubiquitin ...GKQLEDGRTLSDYNIQKESTLHLVLRIRGG
Human NEDD8     ...GKQMNDEKTAADYKILGGSVLHLVLRIRGG
Human ISG15     ...GKPLEDQLPLGEYGLKPLSTVFMNLRIRGG
Human SUMO-1   ...GQRIADNHTPKELGMEEDVIEVYQEQGTGG
Human SUMO-2   ...GQPINETDTPAQLEMEDEDIDVFQQQTGG
Human SUMO-3   ...GQPINETDTPAQLEMEDEDIDVFQQQTGG
Drosophila SUMO-3 ...GQPINENDTPTSLEMEEGDTIEVYQEQGTGG
Arabidopsis SMT3 ...GRRLRAEQTPDELEMEDGDEIDAMLHQGTGG
Yeast SMT3     ...GIRIQADQTPEDIDMEDNDIEAHREQIGG
```

**So, trypsin will not cut SUMO sites (unlike for ubiquitin) to generate a Gly-Gly modified Lysine**

1/24/14

Knuesel et al., 2005

## Detecting PTMs

- If we don't have any idea about what the modification(s) is(are), how do we proceed?
- We won't have the modification on MASCOT or Protein Prospector
- No antibodies or convenient affinity phases

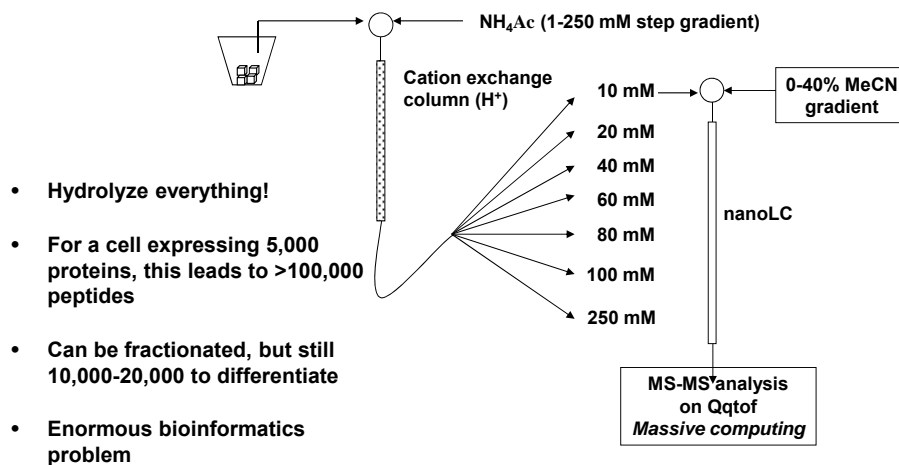
1/24/14

## Can we approach this globally?

- It's asking a lot
  - Too many degrees of freedom
  - Endless modifications
- Should we try?
  - John Yates' group did
  - MacCoss et al. PNAS 99:7900 (2002)

1/24/14

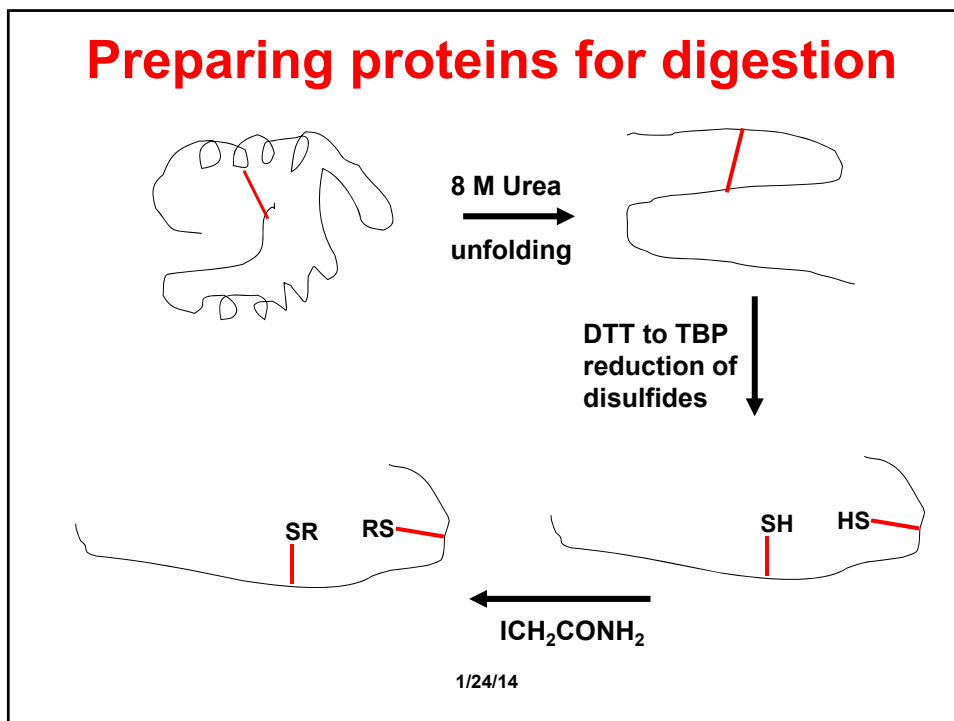
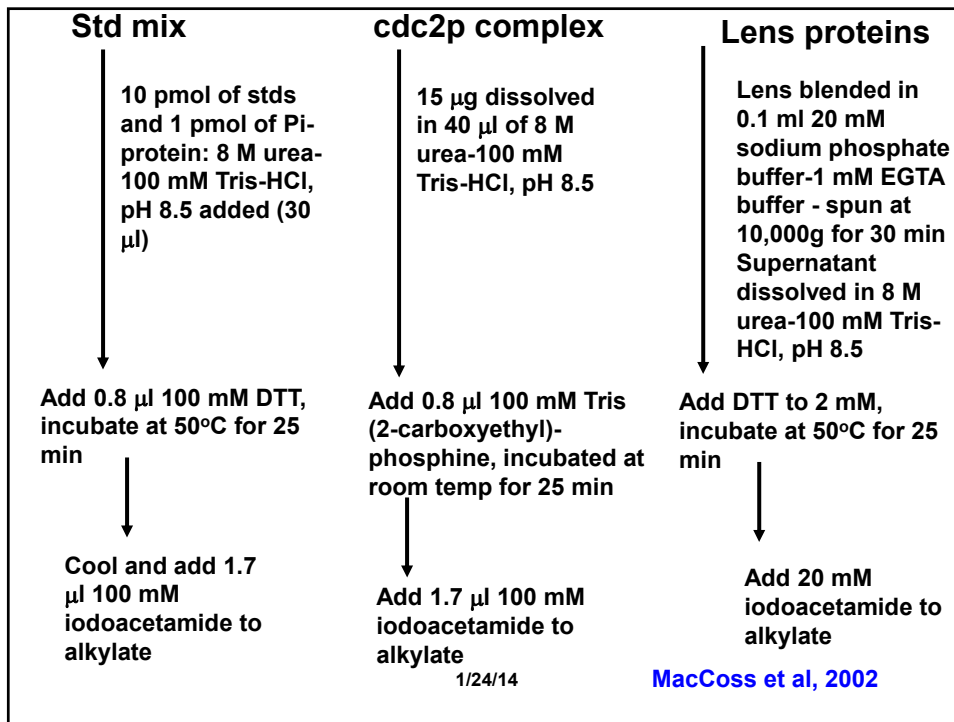
## MudPIT - Multi-dimensional Protein Identification Technology



1/24/14

John Yates





# Hydrolysis procedures

Reduced, carboxymethylated in 8 M urea

Diluted x 3 with  
100 mM Tris-HCl,  
pH 8.5

100 mM CaCl<sub>2</sub>  
added to 1 mM

Roche **trypsin** 1:50,  
incubated 12-24 h  
at 37°C

quenched with  
90% formic acid  
to 4%

Diluted x 3 with 4.8  
M Urea-100 mM Tris-  
HCl, pH 8.5

**subtilisin** added  
1:50, incubated 2-3  
hr at 37°C

quenched with  
90% formic acid  
to 4%

1/24/14

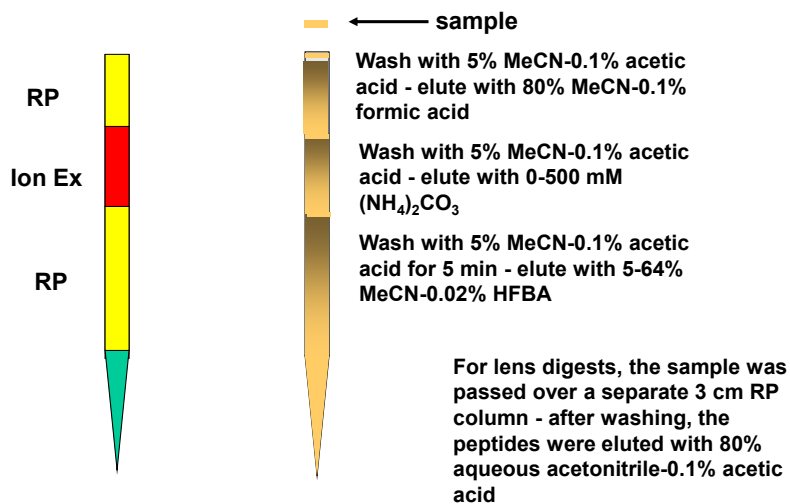
Diluted x 3 with  
100 mM Tris-HCl,  
pH 8.5

incubated with  
**elastase** 1:50 12 h  
at 37°C

quenched with  
90% formic acid  
to 4%

MacCoss et al, 2002

# Elution from a triphasic column



1/24/14

MacCoss et al, 2002

## Automated MS-MS analysis

- Limit analysis to 2<sup>+</sup> or 3<sup>+</sup>-charged peptides
- Delete poor quality spectra
- Identify peptides with 98-Da neutral loss
- Analyze the remaining spectra with SEQUEST (operating on a 31-node beowulf computer cluster) to identify proteins
- Search sequences of identified proteins for PTMs by 80 (STY phosphorylation), 42 (K acetylation), 16 (MWY oxidation) and 14 (K methylation)

1/24/14

MacCoss et al, 2002

## Results

### Protein standard mixture:

- The three digests were combined - 83.7% and 95.4% coverage for glycogen phosphorylase and BSA
- Identified the glycogen phosphorylase phosphorylation site as well as the two known sites (S69/S345) in ovalbumin
- N-acetylation found plus many sites of methionine oxidation (? due to work up or real)
- New sites found - phosphorylation at S237/S241 in ovalbumin and methylation at R652

1/24/14

MacCoss et al, 2002

## Results

### Cdc2p complex:

- 200 proteins - 20 showed >40% sequence coverage
- Expected Y15 and T167 phosphorylation sites on Cdc2p found
- New phosphorylation sites found on cyclin partners Cdc13 and Cig1p
- Multiple methylation sites on Cdc2p

1/24/14

MacCoss et al, 2002

## Results

### Lens proteins:

- Found 272 proteins - 52 had >40% sequence coverage
- 90% are crystallins
- PTMs accumulate over your lifetime
- Used 18-step MudPIT because of complexity
- 73 different PTMs found on the 11 crystallins
- Found the 13 of 18 PTMs previously described in all species
- Found 60 other new PTMs in phosphorylation, oxidation, acetylation and methylation

1/24/14

MacCoss et al, 2002

## Summary

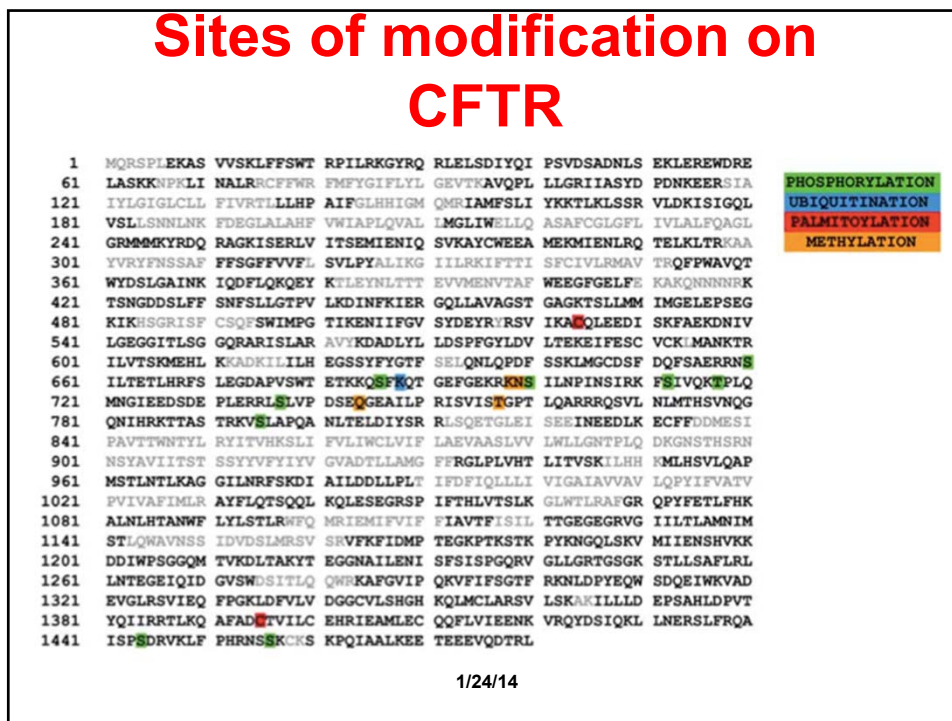
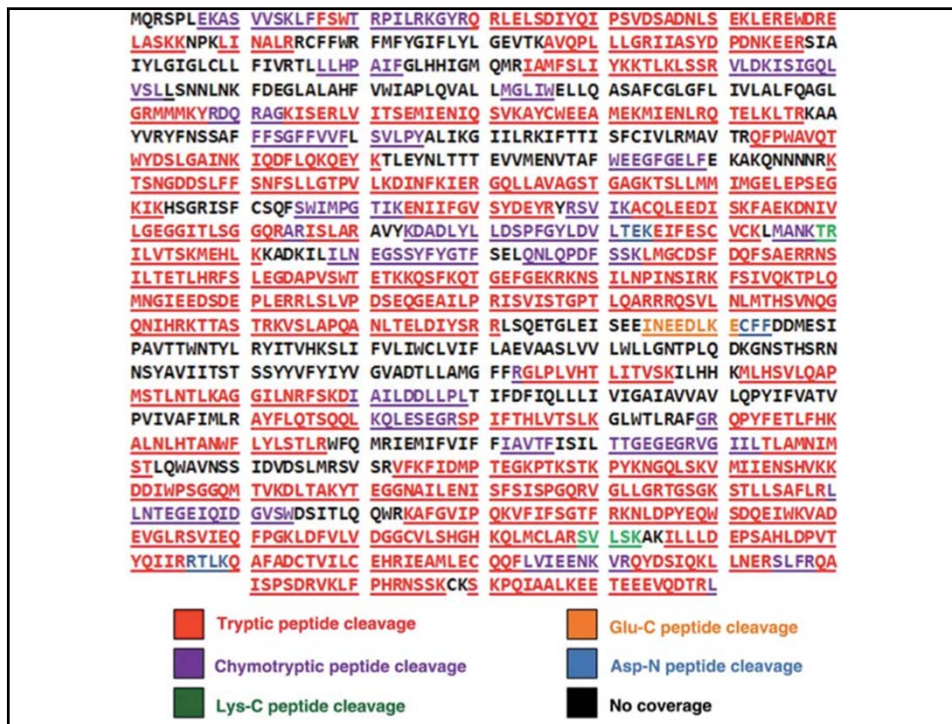
- The proteome is very complex
- Some proteins are in pro-forms and can be found in many, often small, but active fragments
- In addition, proteins are heavily modified
- Modification can account for marked differences in turnover
- Selective fractionation is needed to demonstrate PTMs that are in low abundance

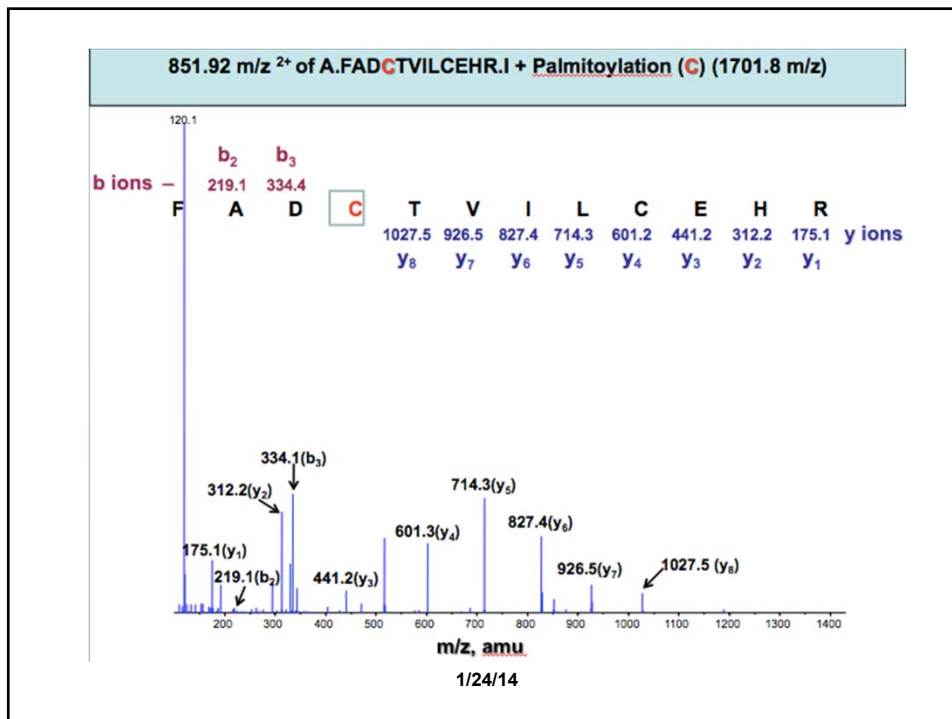
1/24/14

## CFTR modifications

- McClure M, DeLucas LJ, Wilson L, Ray M, Rowe S, Wu X, Dai Q, Hong JS, Sorscher EJ, Kappes JC, Barnes S. Palmitoylation and other post-translational modifications of CFTR with relevance to gating and processing. [\*Protein Eng Des Sel.\* 2012 Jan;25\(1\):7-14.](#)

1/24/14





## Palmitoylation of CFTR

- Palmitoylation of Cys1395 was discovered in a “discovery” method, i.e., fully untargeted
- It opened the question as to (1) was it biologically important and (2) were other cysteine residues palmitoylated?
- This latter question was addressed by using “targeted” discovery using a MRM method on a triple quad mass spectrometer

1/24/14

## Targeted MRM method

- We know the  $m/z$  value of the tryptic peptide that could become palmitoylated
- We add the mass of the palmitoyl group
- We also adjust the values of the b- and y-ions to reflect where the addition occurs

1/24/14

## Ions for the peptide containing Cys524

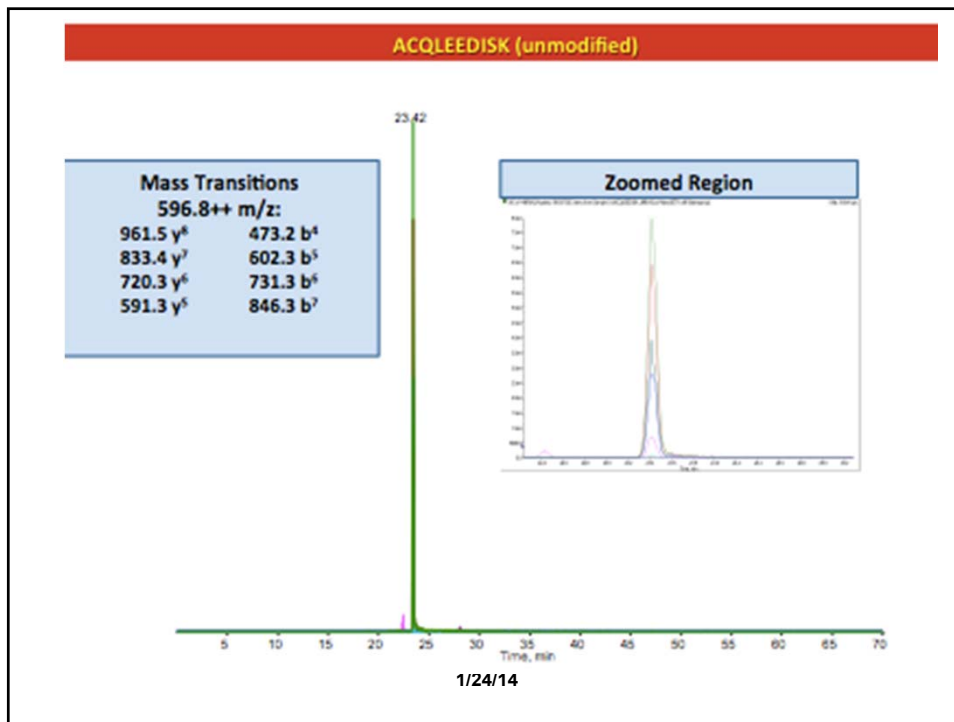
**ACQLEEDISK** ( $m/z$  596.8)

|          |           |          |          |          |          |          |          |          |          |
|----------|-----------|----------|----------|----------|----------|----------|----------|----------|----------|
|          | 232       | 360      | 473      | 602      | 731      | 845      | 958      | 1045     |          |
| <b>A</b> | <b>C*</b> | <b>Q</b> | <b>L</b> | <b>E</b> | <b>E</b> | <b>D</b> | <b>I</b> | <b>S</b> | <b>K</b> |
|          | 971       | 843      | 720      | 591      | 462      | 347      | 234      | 147      |          |

In this case, the Cys has been modified by IOA

1/24/14





## Palmitoylated peptide

**ACQLEEDISK (*m/z* 687.4)**

413 541 654 783 912 1027 1140 1227  
**A C\* Q L E E D I S K**  
 971 843 720 591 462 347 234 147

In this case, the Cys has been modified by palmitate

1/24/14

ACQLEEDISK (palmitoylated)

