Proteomics and Mass Spectrometry

Chris Adams, PhD

Stanford University Mass Spectrometry

http://mass-spec.stanford.edu/

New Additions to the Stanford "Proteomic" Community

Prof. Mike Snyder

Prof. Josh Elias

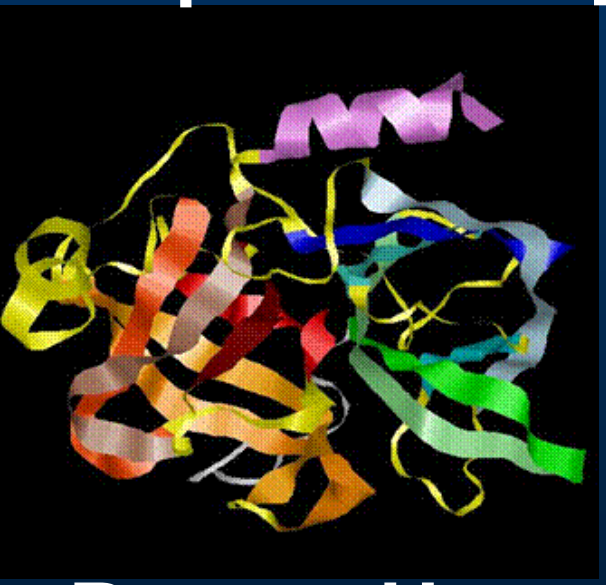
Instrumental Upgrades

Bottom Up and Top Down

MS

Top Down

Protein Centric



LC

Bottom Up

Peptide

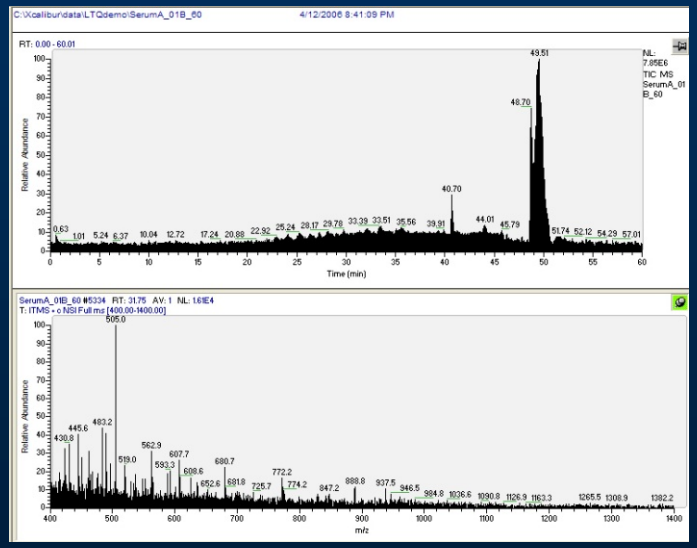
Proteolysis

Centric

Data Analysis and Distribution of Results

MS

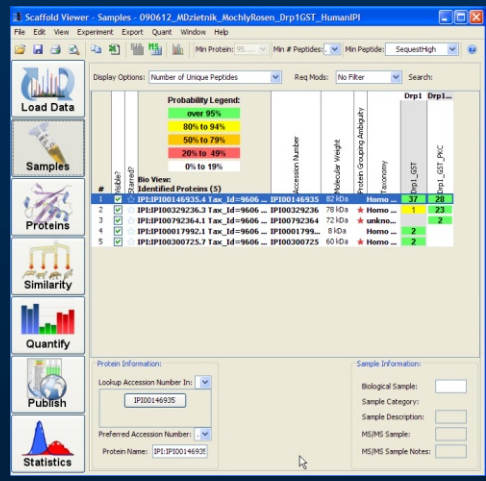
Raw data extract



Database Search

Sequest
Mascot

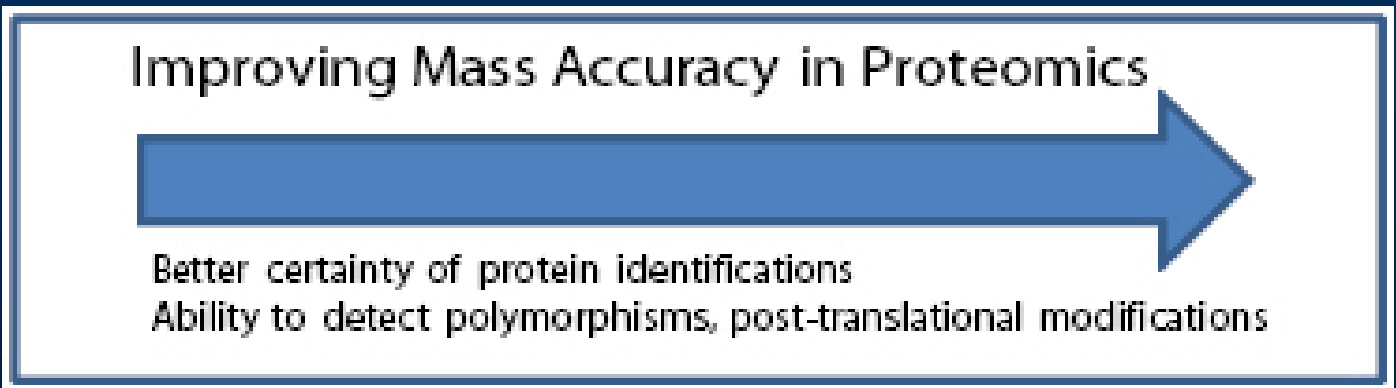
Results



Scaffold

MS Instrumentation: High Mass Accuracy, Resolution, Sensitivity and Speed

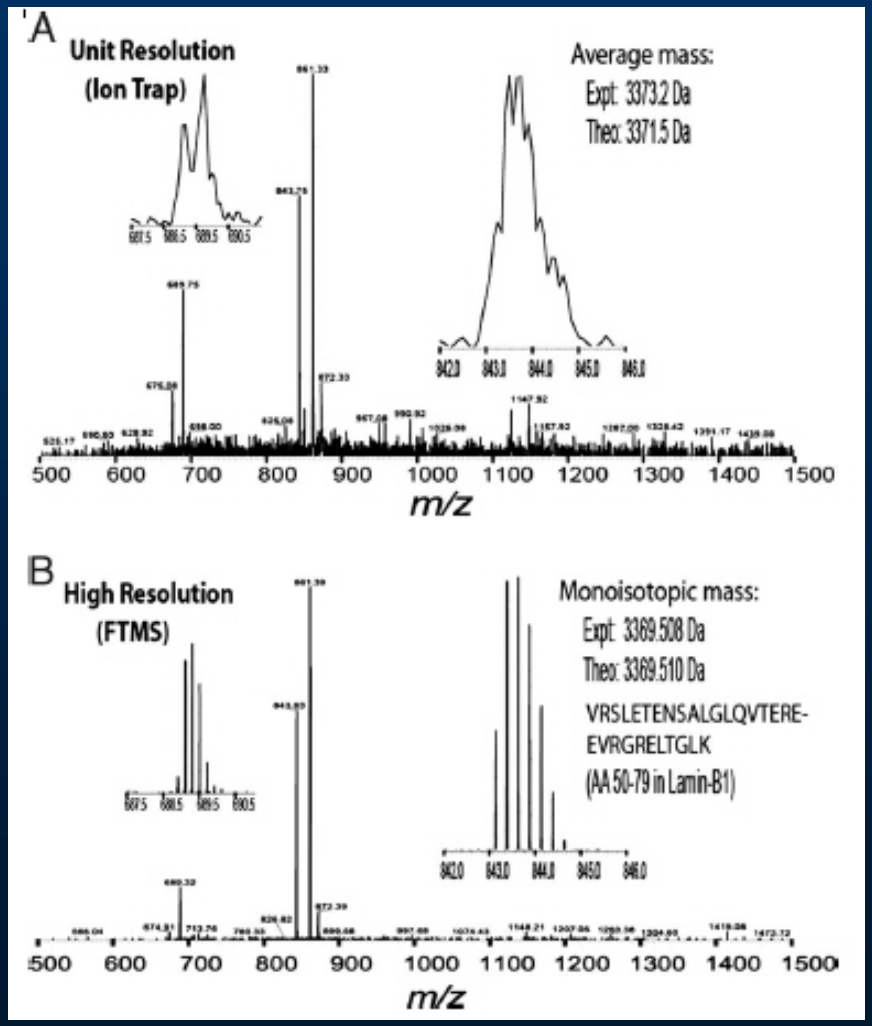
Proteomics 2.0, Precision Proteomics

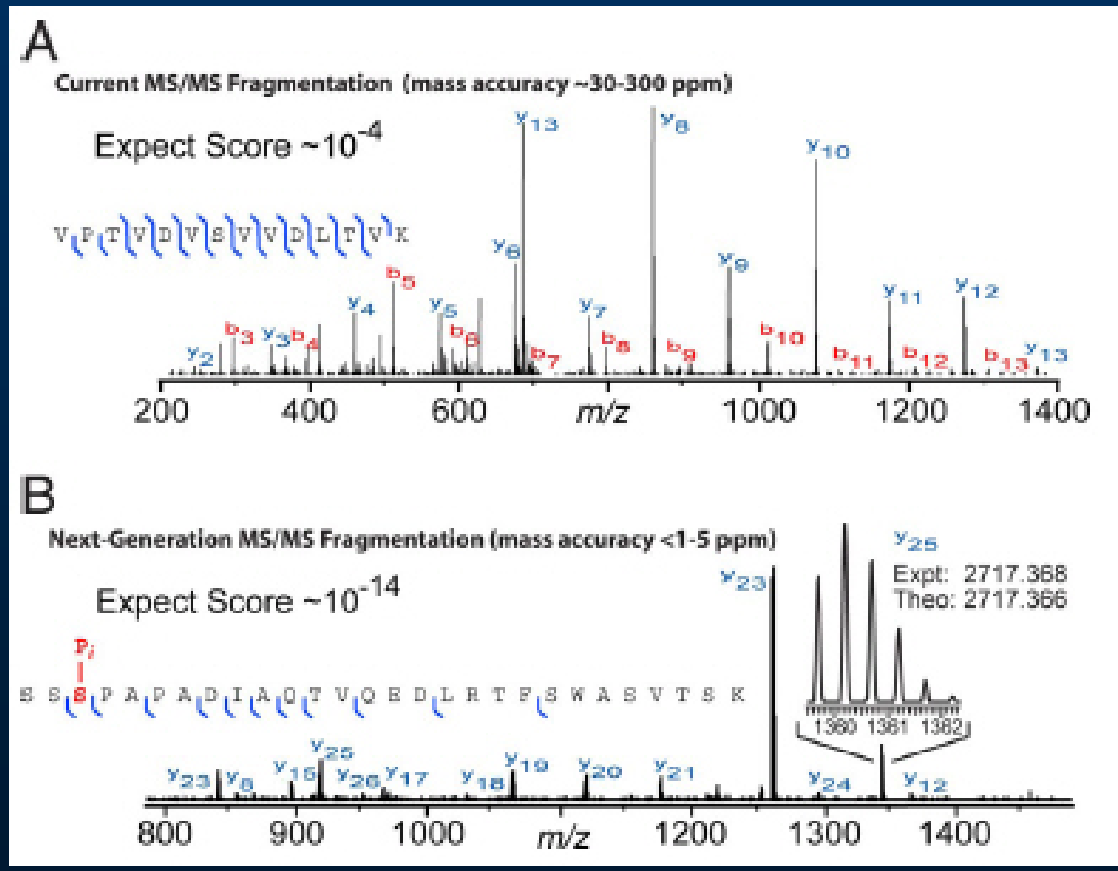


Low Resolution	Medium Resolution	High-Resolution
1 – 0.1 Da accuracy	0.1-0.01 Da accuracy	0.01-0.001 Da accuracy
Ion Traps , Quadrupoles, triple quadrupoles	Time-of-Flight,* hybrids with quadrupoles	FT ICR MS, FT-Orbitraps, hybrids with ion traps

18132–18138 PNAS **November 25, 2008** vol. 105 no. 47

Accurate Mass Measurement Significantly Aids ID and More





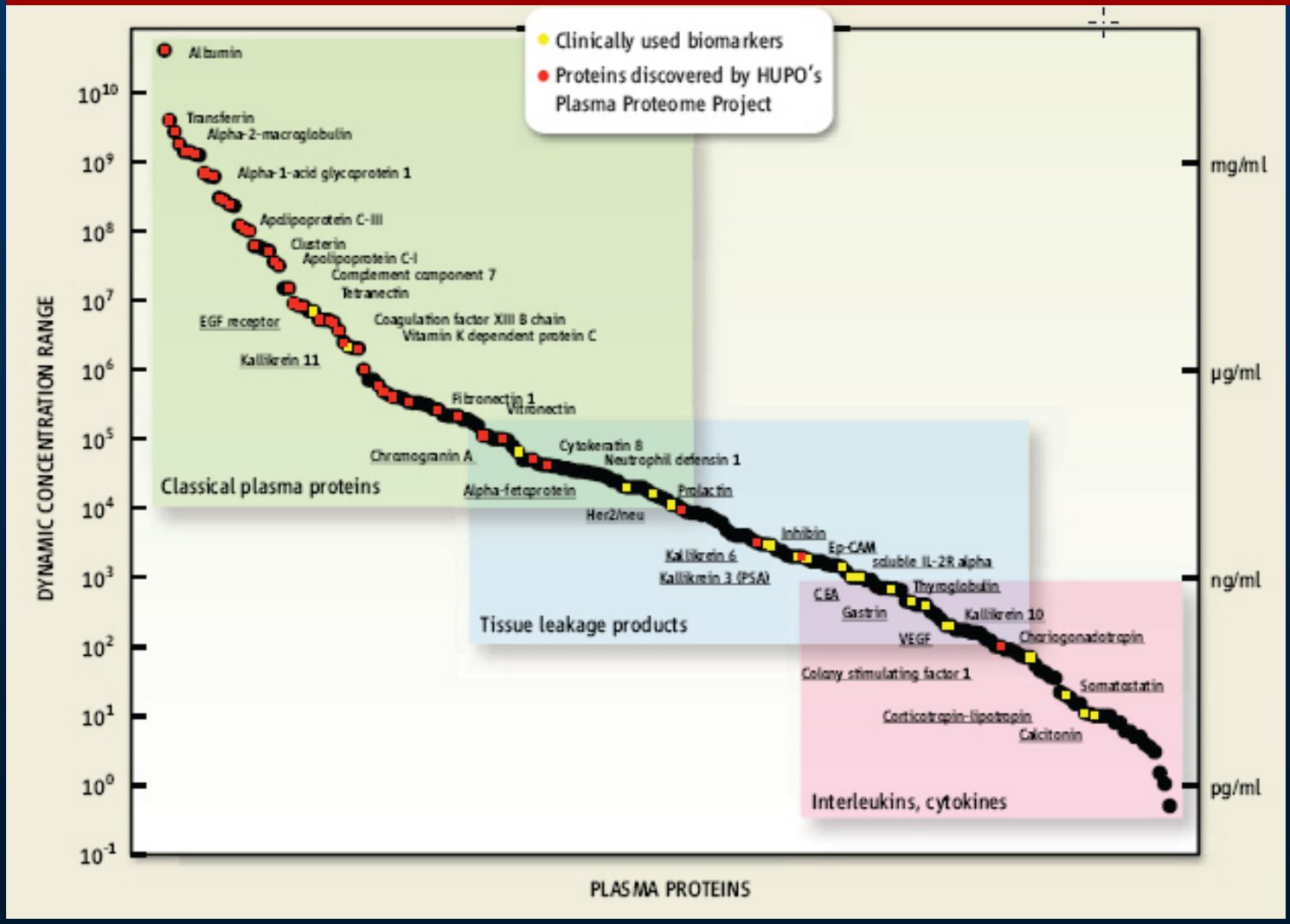
Mass Accuracy: What it Means

1. Anecdotal- report of a single measurement

2. Statistical- accuracy estimated from a statistical distribution of mass errors

3. Max. allowed mass deviation (MMD)- mass accuracy cutoff value when database searching

Need for Sensitivity Over 10 Orders of Magnitude



Human Proteome Organization

26 SEPTEMBER 2008 VOL 321 SCIENCE

Depth of Proteome Coverage

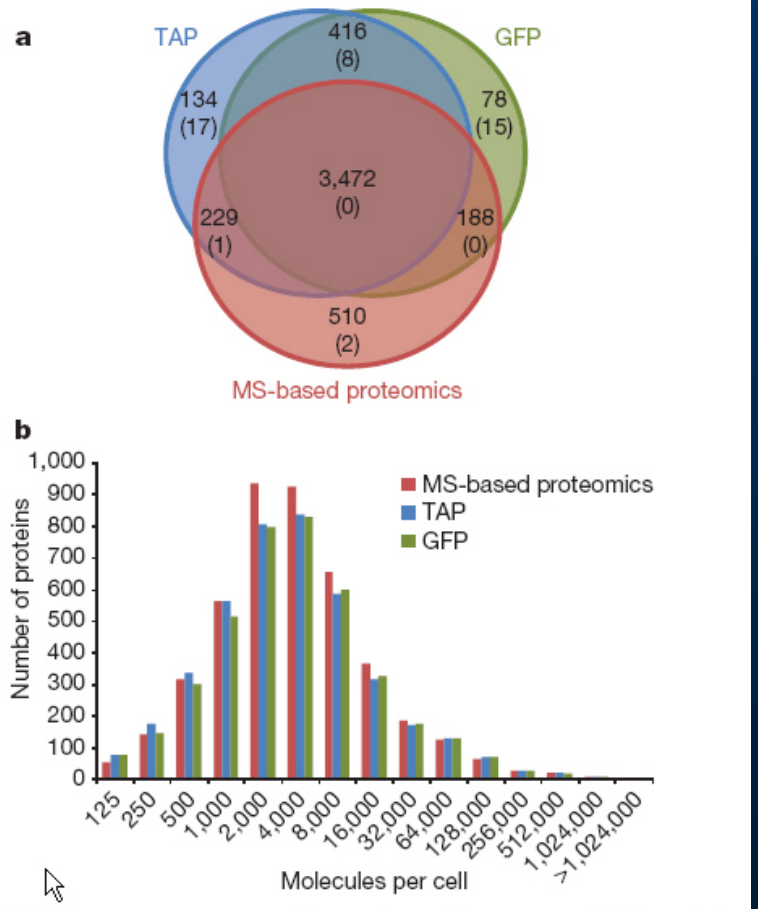


Figure 2 | Proteome coverage. **a**, Comparison of coverage of MS-based proteomics with GFP- and TAP-tagging methods[14,15]. Numbers are the identified proteins by each method and, in parentheses, the number of dubious open reading frames (ORFs). **b**, Identified proteins per copy number bin for MS-based proteomics and the two tagging approaches. Copy numbers were estimated by correlation between summed peptide intensity per protein and the quantitative western blotting data[14] (Methods).

Avg. Sequence coverage 30%

NATURE| Vol 455|30 October 2008

Quantification

- Stable isotope labeling by amino acids cell culture (SILAC)
- Isobaric labeling for relative and absolute quantification (iTRAQ)
- Stable isotope dimethyl labeling
- Label free

Post Translational Modifications (PTM's)

The "Histome"

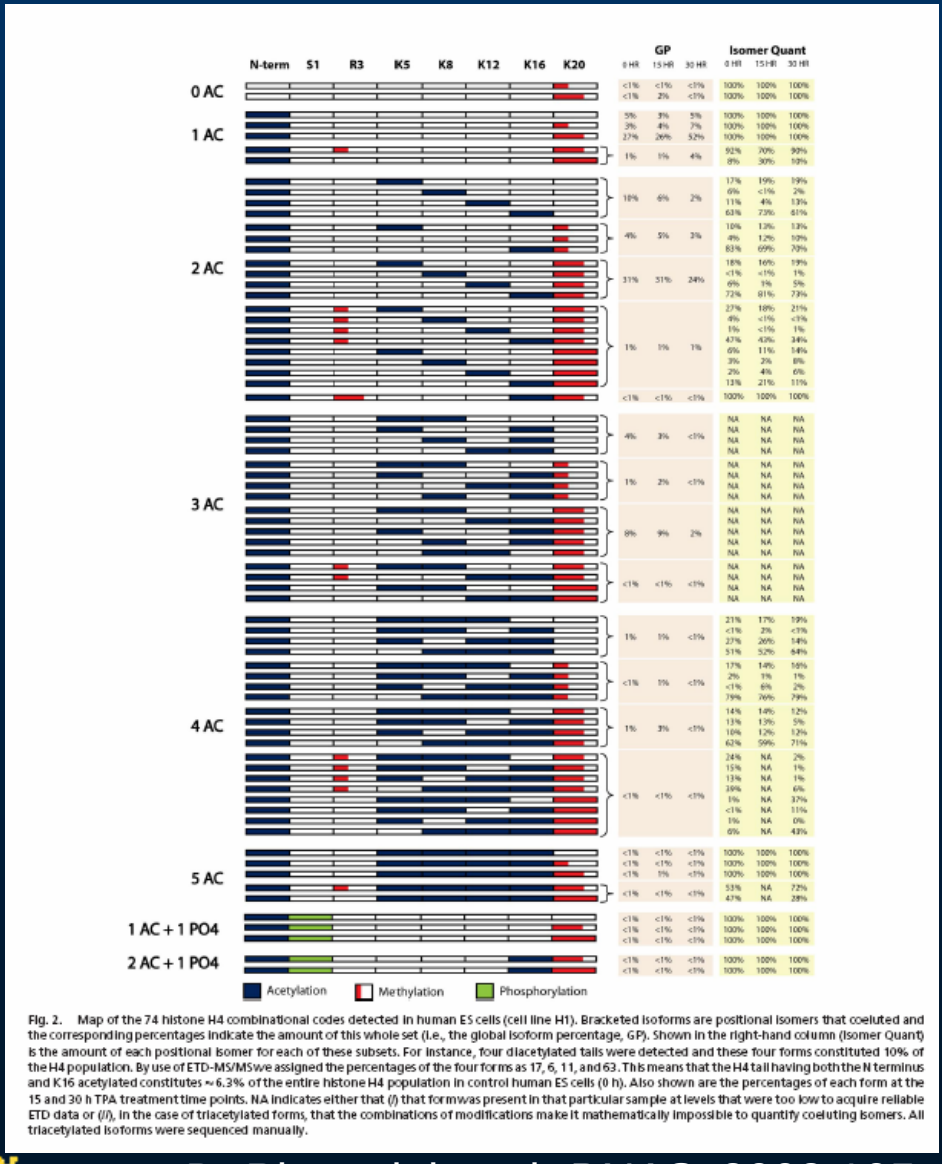


Fig. 2. Map of the 74 histone H4 combinational codes detected in human ES cells (cell line H1). Bracketed isoforms are positional isomers that coeluted and the corresponding percentages indicate the amount of this whole set (i.e., the global isoform percentage, GP). Shown in the right-hand column (Isomer Quant) is the amount of each positional isomer for each of these subsets. For instance, four diacetylated tails were detected and these four forms constituted 10% of the H4 population. By use of ETD-MS/MS we assigned the percentages of the four forms as 17, 6, 11, and 63. This means that the H4 tail having both the N terminus and K16 acetylated constitutes ≈6.3% of the entire histone H4 population in control human ES cells (0 h). Also shown are the percentages of each form at the 15 and 30 h TPA treatment time points. NA indicates either that (/) that form was present in that particular sample at levels that were too low to acquire reliable ETD data or (//), in the case of triacetylated forms, that the combinations of modifications make it mathematically impossible to quantify coeluting isomers. All triacetylated isoforms were sequenced manually.

D. Phanstiel et al, *PNAS*, 2008 105, 11, 4093

Modifications the Need for Enrichment

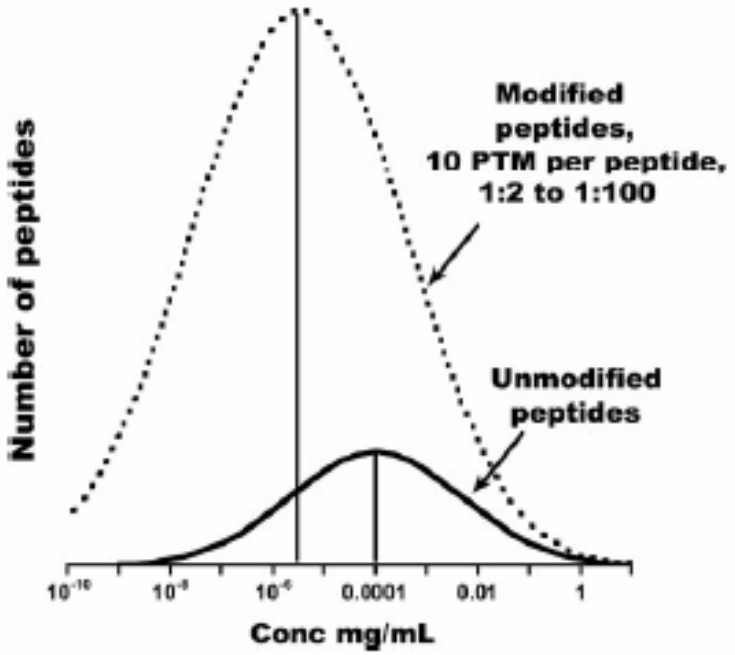


FIG. 4. Theoretical distribution of unmodified tryptic peptide concentrations in a complex biological sample (*solid line*) and the resulting distribution of modified peptide concentrations (*dashed line*) assuming 10 modifications per peptide at a substoichiometric range of 1:2 to 1:100.

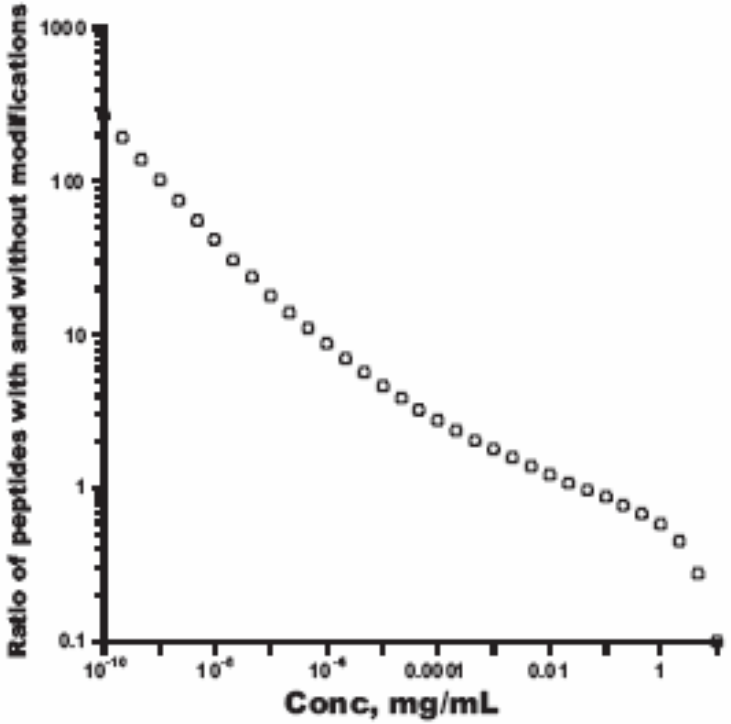
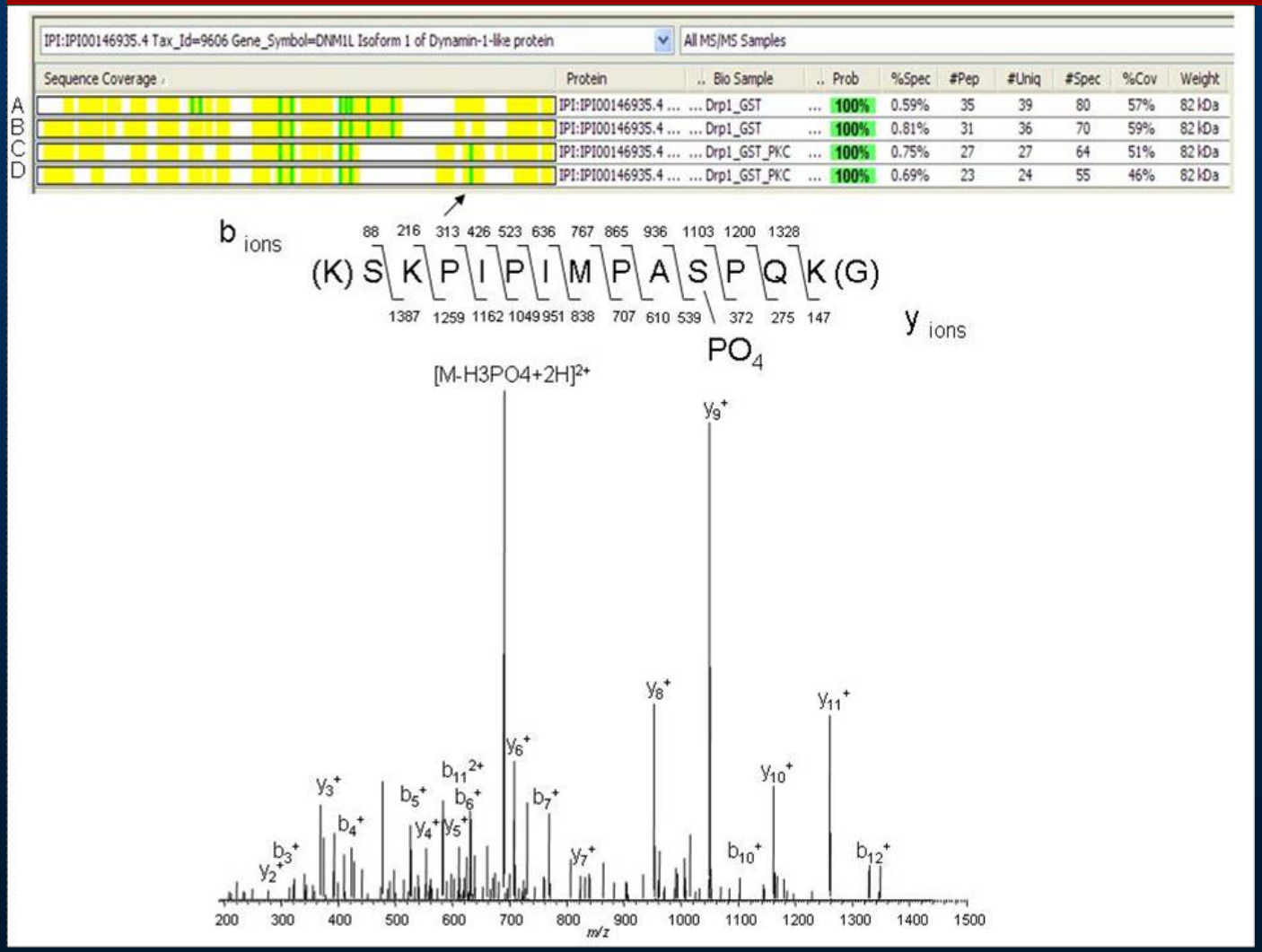


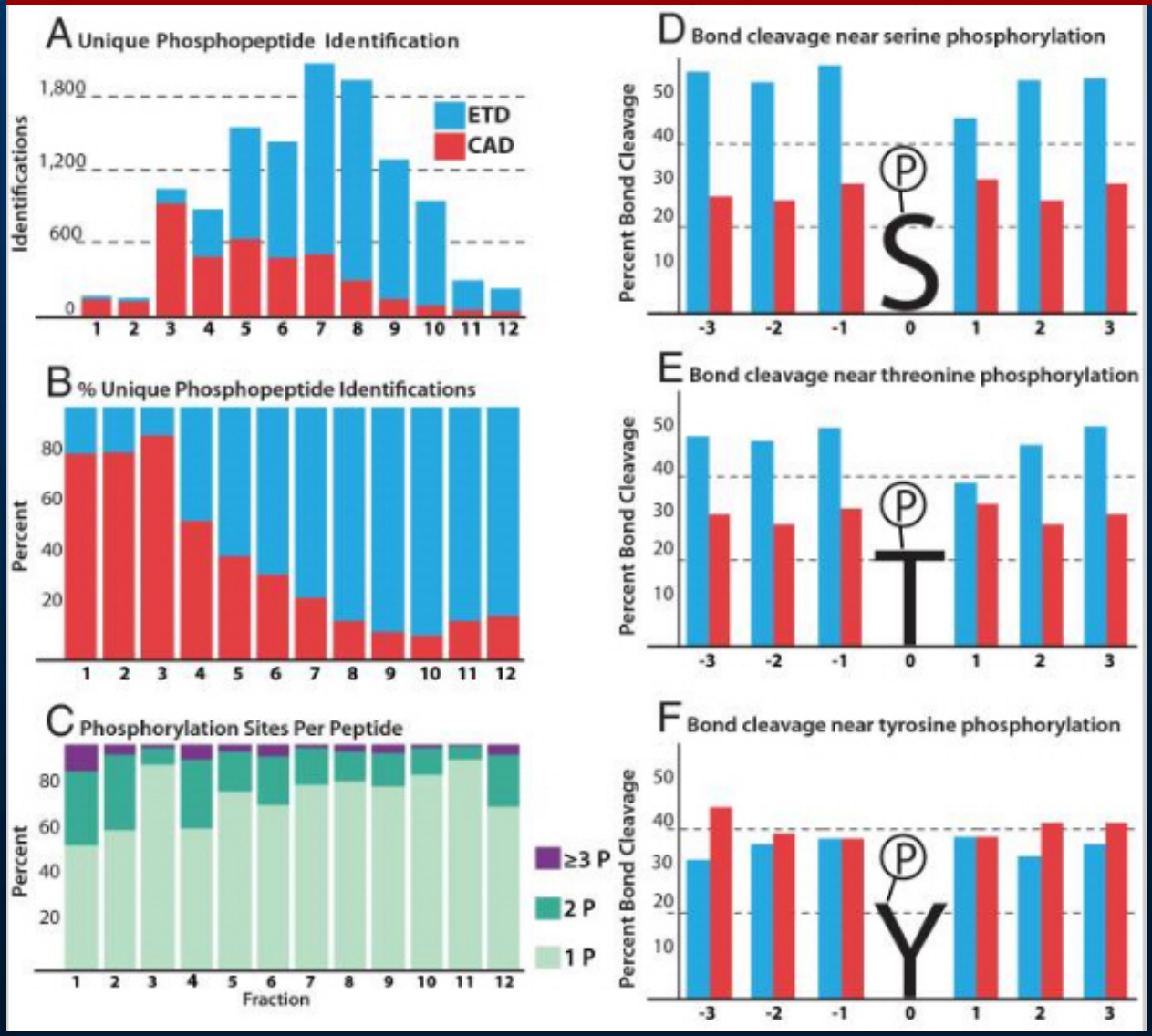
FIG. 5. The average number of modified peptides per single unmodified peptide at a given concentration. The distributions of modified and unmodified peptides are shown in Fig. 4.

M. Nielsen, et al, *Mol. Cell. Prot.*, 2006, 5:2384

Phosphopeptide Analysis



Phosphopeptide Analysis using ETD and/or HCD

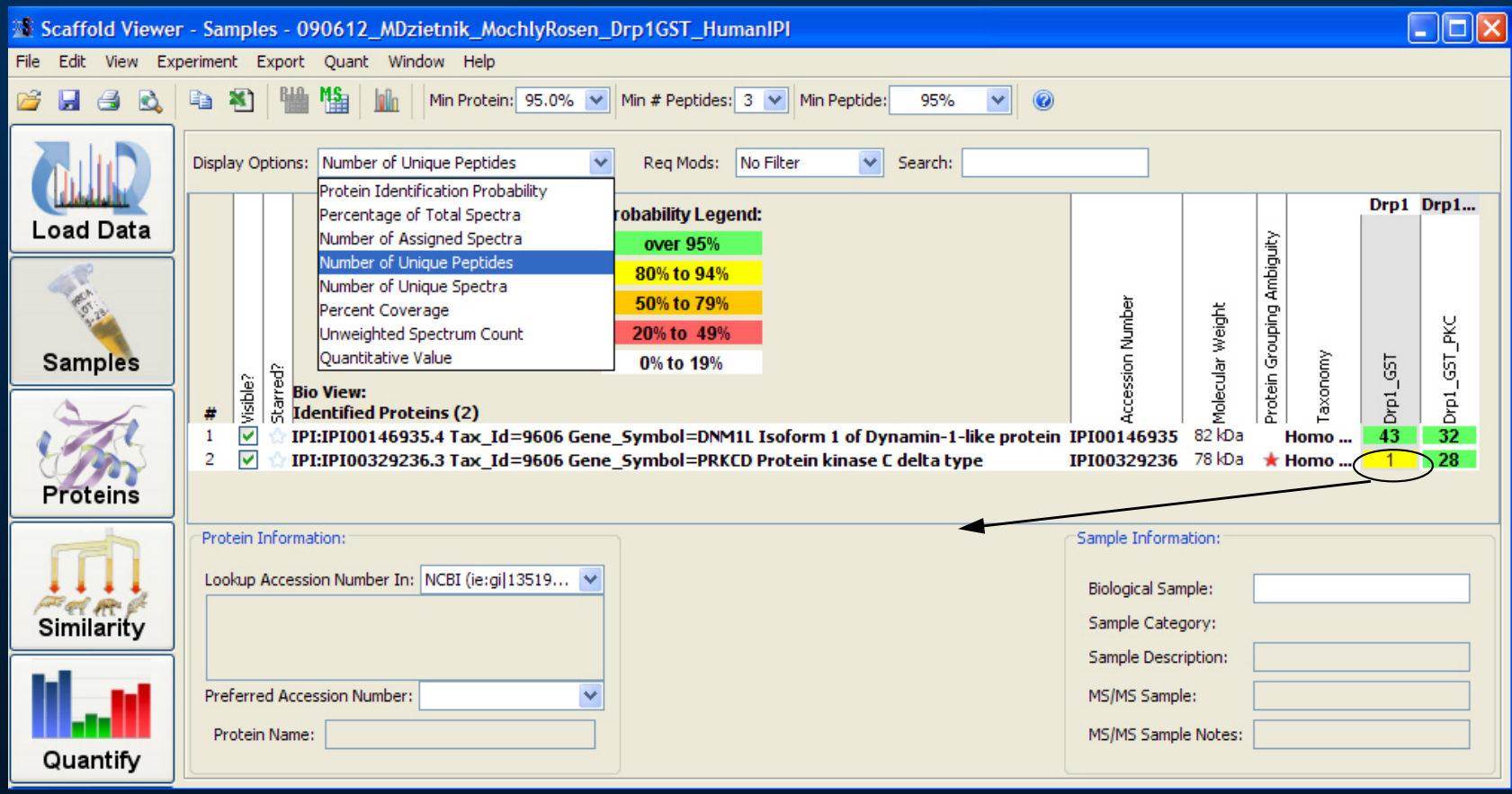


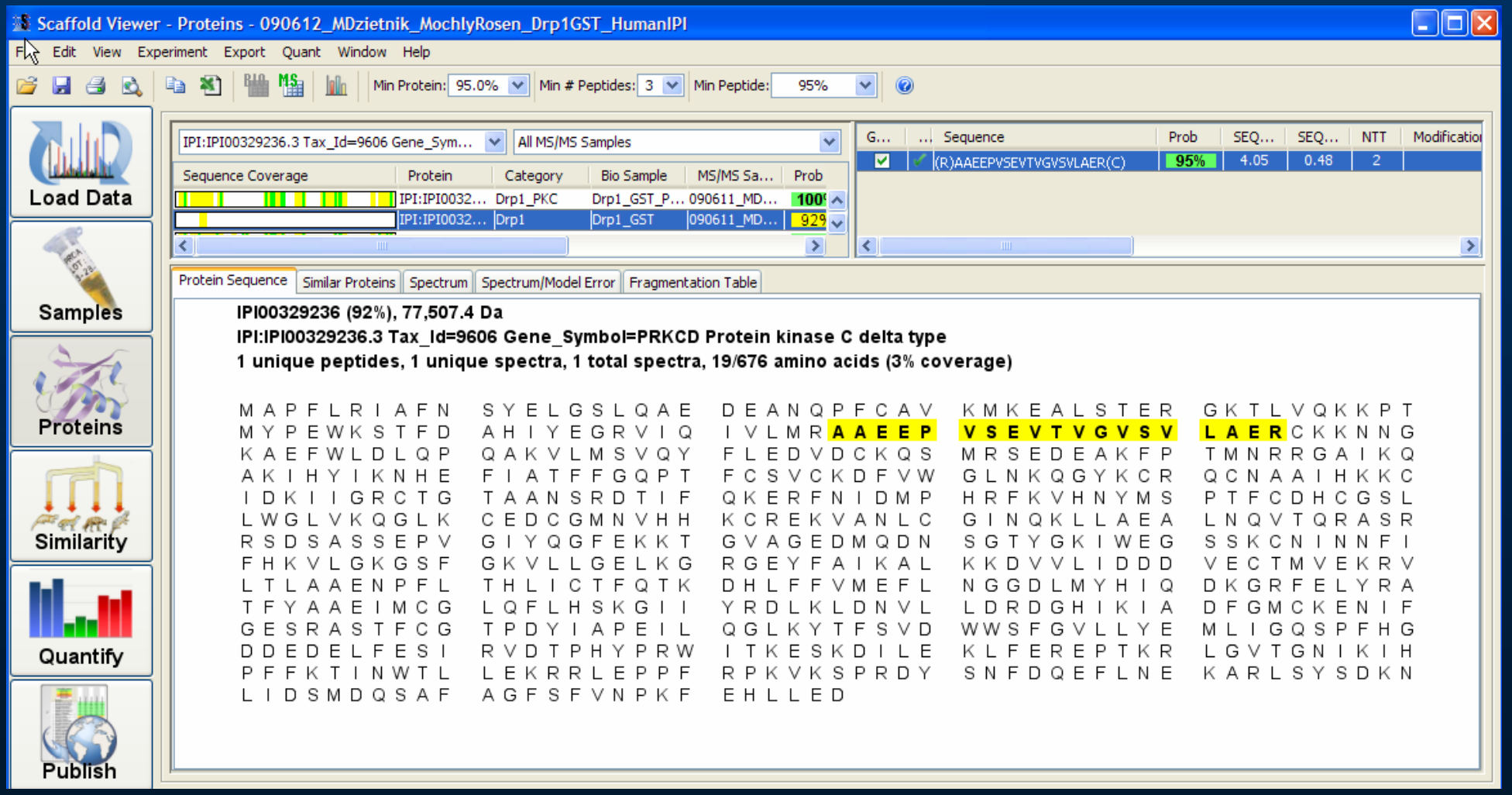
11,955 Phosphopeptides

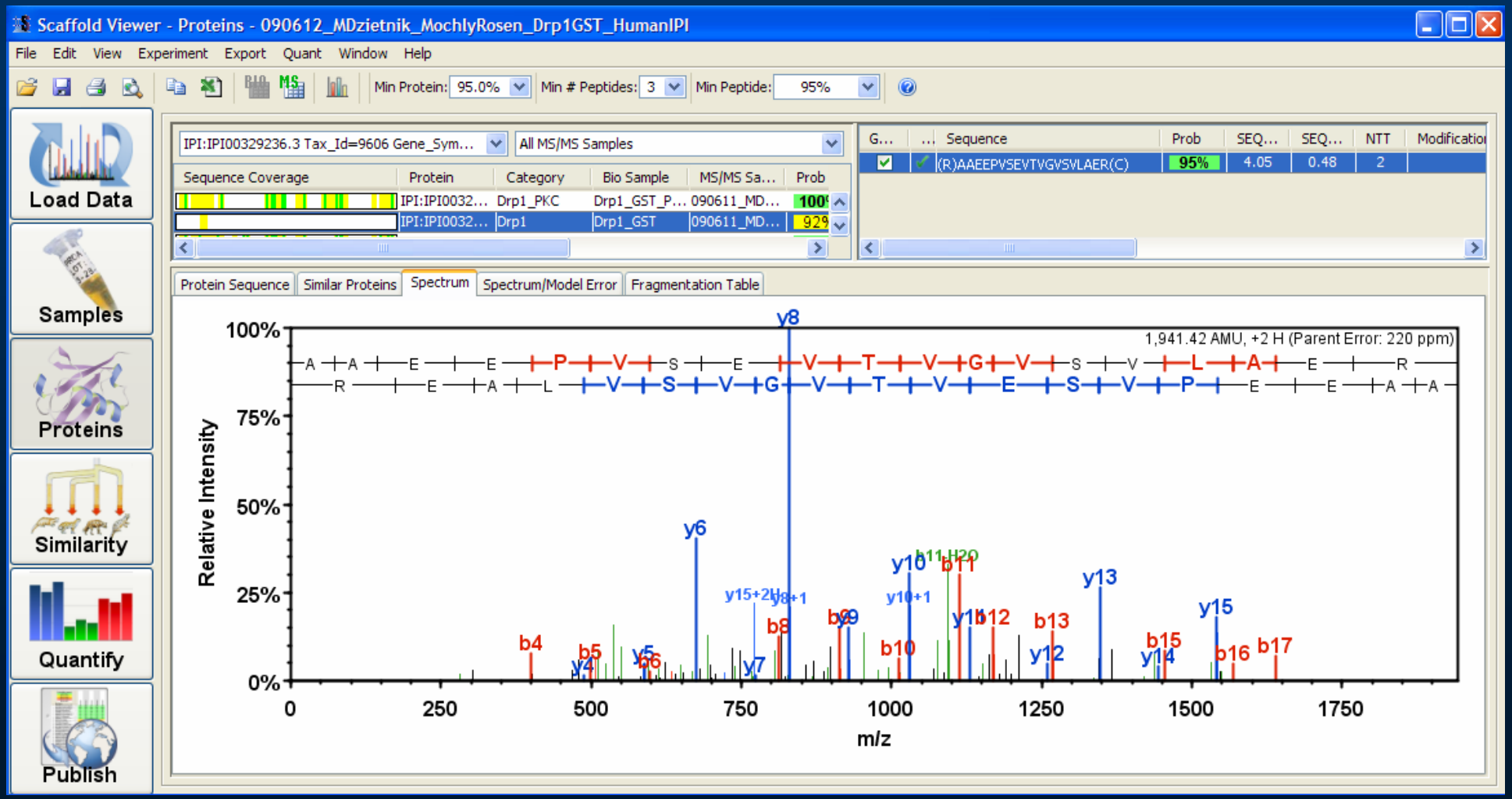
PNAS 2009 January 27; 106(4): 995–1000

Evaluating Data for Accuracy

Identification at the peptide level:
Probability at the protein level:




Scaffold Viewer - Proteins - 090612_MDzietnik_MochlyRosen_Drp1GST_HumanIPI
File Edit View Experiment Export Quant Window Help
Min Protein: 95.0%
Min # Peptides: 3
Min Peptide: 95%
Load Data
Samples
Proteins
Similarity
Quantify
Publish
IPI:IPI00329236.3 Tax_Id=9606 Gene_Sym...
All MS/MS Samples
Sequence Coverage
Protein
Category
Bio Sample
MS/MS Sa...
Prob
IPI:IPI0032...
Drp1_PKC
Drp1_GST_P...
090611_MD...
IPI:IPI0032...
Drp1
Drp1_GST
090611_MD...
G...
Sequence
Prob
SEQ...
SEQ...
NTT
Modificatio
(R)AAEEPVSEVTVGVSVLAER(C)
95%
4.05
0.48
2
Protein Sequence
Similar Proteins
Spectrum
Spectrum/Model Error
Fragmentation Table
IPI00329236 (92%), 77,507.4 Da
IPI:IPI00329236.3 Tax_Id=9606 Gene_Symbol=PRKCD Protein kinase C delta type
1 unique peptides, 1 unique spectra, 1 total spectra, 19/676 amino acids (3% coverage)
M A P F L R I A F N S Y E L G S L Q A E D E A N Q P F C A V K M K E A L S T E R G K T L V Q K K P T
M Y P E W K S T F D A H I Y E G R V I Q I V L M R A A E E P V S E V T V G V S V L A E R C K K N N G
K A E F W L D L Q P Q A K V L M S V Q Y F L E D V D C K Q S M R S E D E A K F P T M N R R G A I K Q
A K I H Y I K N H E F I A T F F G Q P T F C S V C K D F V W G L N K Q G Y K C R Q C N A A I H K K C
I D K I I G R C T G T A A N S R D T I F Q K E R F N I D M P H R F K V H N Y M S P T F C D H C G S L
L W G L V K Q G L K C E D C G M N V H H K C R E K V A N L C G I N Q K L L A E A L N Q V T Q R A S R
R S D S A S S E P V G I Y Q G F E K K T G V A G E D M Q D N S G T Y G K I W E G S S K C N I N N F I
F H K V L G K G S F G K V L L G E L K G R G E Y F A I K A L K K D V V L I D D D V E C T M V E K R V
L T L A A E N P F L T H L I C T F Q T K D H L F F V M E F L N G G D L M Y H I Q D K G R F E L Y R A
T F Y A A E I M C G L Q F L H S K G I I Y R D L K L D N V L L D R D G H I K I A D F G M C K E N I F
G E S R A S T F C G T P D Y I A P E I L Q G L K Y T F S V D W W S F G V L L Y E M L I G Q S P F H G
D D E D E L F E S I R V D T P H Y P R W I T K E S K D I L E K L F E R E P T K R L G V T G N I K I H
P F F K T I N W T L L E K R R L E P P F R P K V K S P R D Y S N F D Q E F L N E K A R L S Y S D K N
L I D S M D Q S A F A G F S F V N P K F E H L L E D



Stanford University
MASS SPECTROMETRY

Filtering Data: FDR/FPR

FDR: False Discovery Rate

For a number *m* of MS/MS spectra (probability within dataset)

FPR: False Positive Rate

A single spectrum

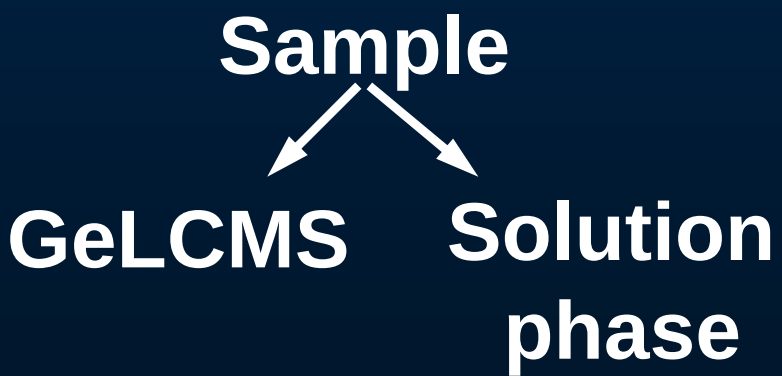
A. Nesvizhskii et al. ***NATURE METHODS*** | VOL.4 NO.10 | OCTOBER 2007 | **795**

Sample Prep: Selective to Desired Outcome

Full characterization of a single molecule- including PTMs?

Global proteome study?

Specific for phosphorylation, acetylation, methylation, ubiquitination…..?

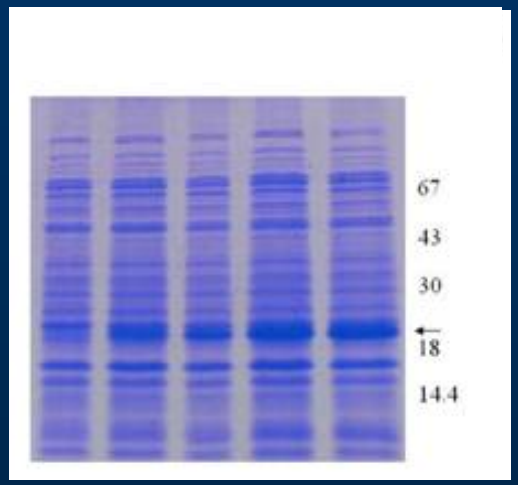


GeLCMS Works Better Than Ever..

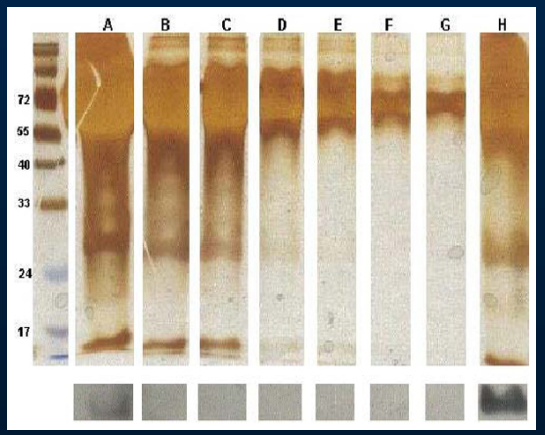
	Gel Region	1*	2	3	4*	5*	6	7	8	9*	Total	Change
Overnight												
	Peptides	135	210	271	108	96	243	163	130	51	**1407**	
	Proteins	23	28	36	18	13	45	36	28	6	**233**	
Pmax_1hr												
	Peptides	210	198	253	253	204	266	149	147	98	**1778**	**>371**
	Proteins	31	28	36	29	24	45	37	31	12	**273**	**>40**

GeLCMS Works Better Than Ever.. But

Coomassie



Silver



Detection Limits

Brilliant Blue 50 ng
Colloidal 10-20 ng

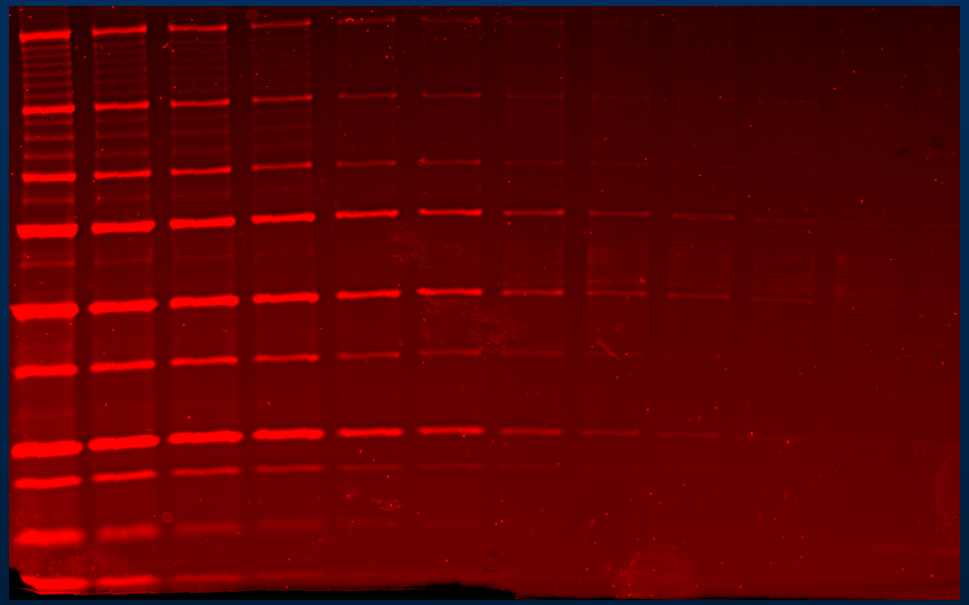
Mass Spec

Compatible* 1-5 ng

***No fixing/staining steps involving formaldehyde/glutaraldehyde**

The Compromise

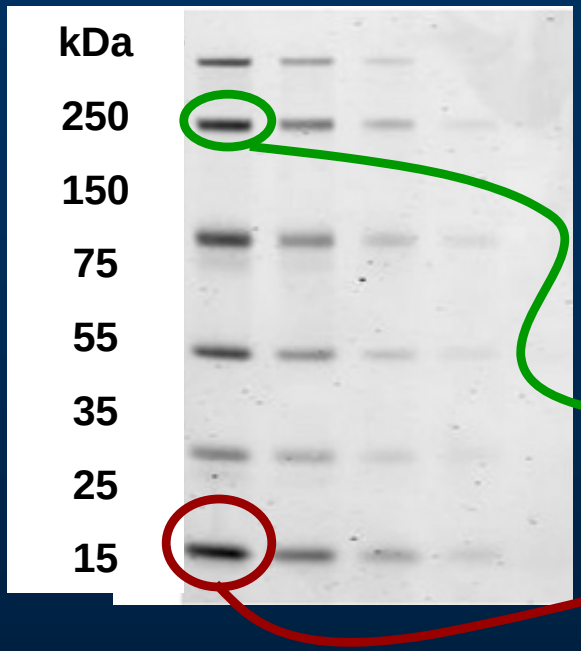
Sypro Ruby



Detection Limit

5-10 ng

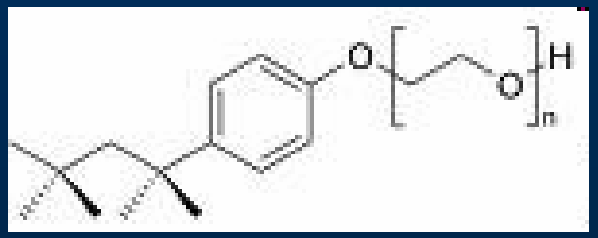
Size Matters ? Why



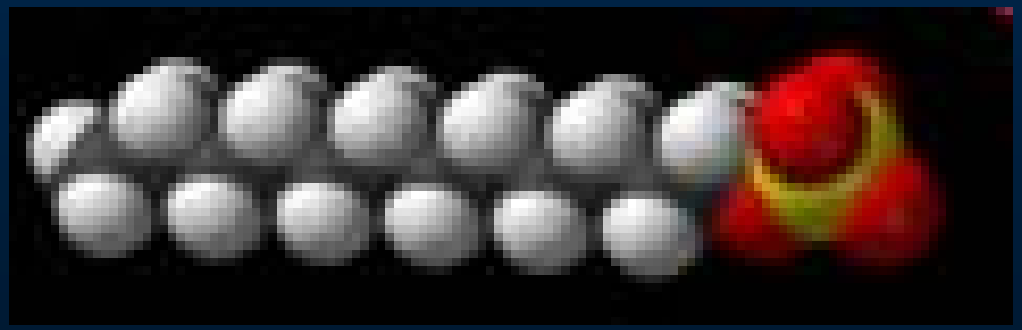
- detection limit of protein staining is on a weight basis
- detection limit of protein with the mass spectrometer is on a molar basis
- higher the molecular weight, at the same mass, the higher the detection limit will be for the mass spectrometer
- 1.0ng of a 15kd protein is 67 fmol, while 1.0ng of a 250kd protein is only 4 fmol.
- Both proteins will have similar stain intensities, but there is 15 times less protein on a molar basis from the 250kd protein.
- Protein stains detect total protein, mass spectrometer detects proteins individually.

In Solution Digests

Most all surfactants and detergents are detrimental



Triton X-100



SDS

Enrichments

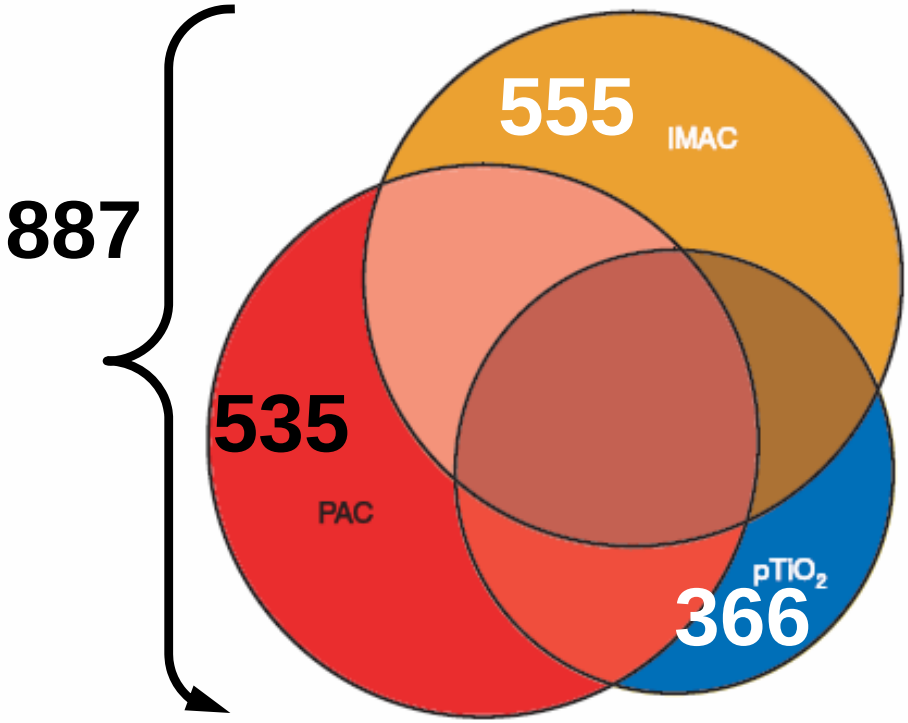


Figure 4 | Overlap between phosphopeptide isolation methods on the level of identified phosphorylation sites. dhbTiO$_2$ is not shown, as 95% of the phosphopeptides identified from the dhbTiO$_2$ samples were also identified in the pTiO$_2$ samples.

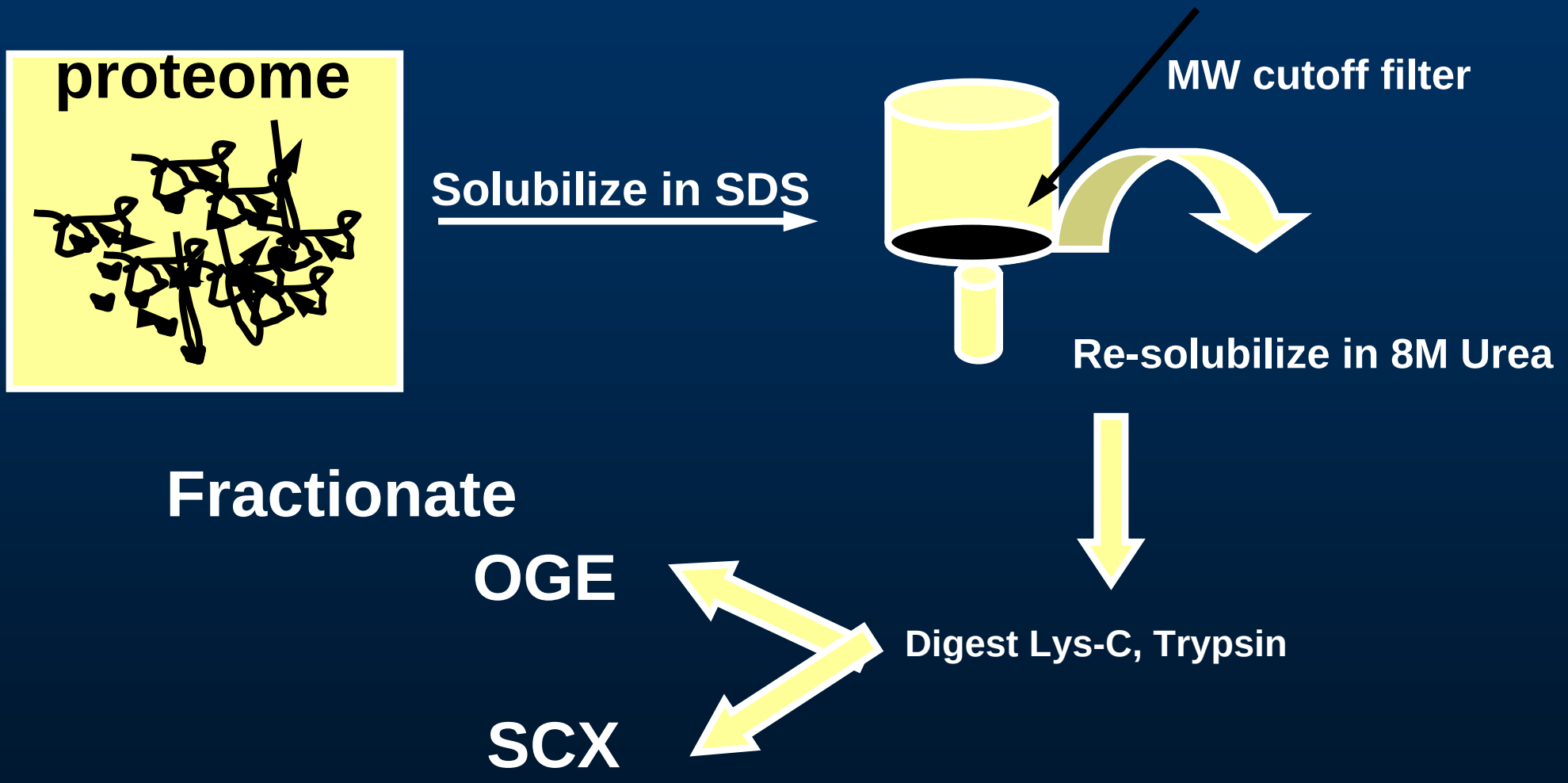
PAC- phosphoramidate chemistry

IMAC- immobilized metal affinity chromatography

pTiO$_2$- pthalic titanium dioxide

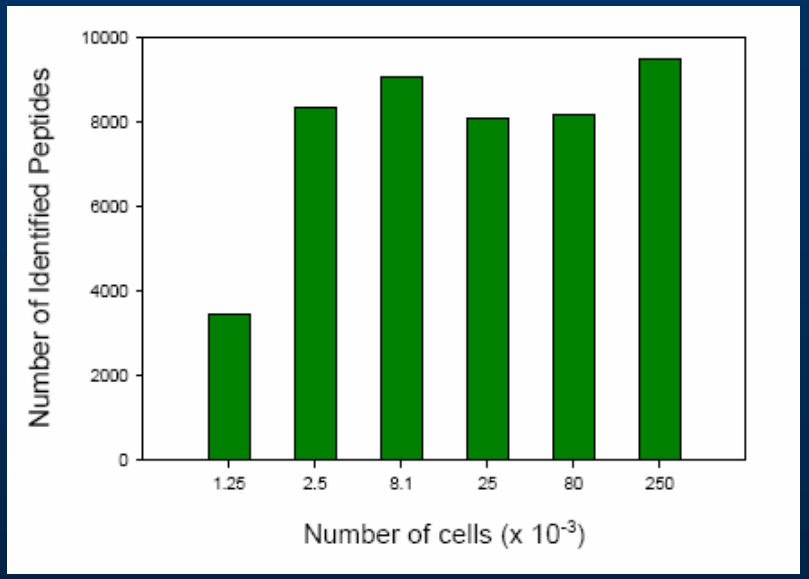
B. Bodenmiller et. al, 234 | VOL.4 NO.3 | MARCH 2007 | *NATURE METHODS*

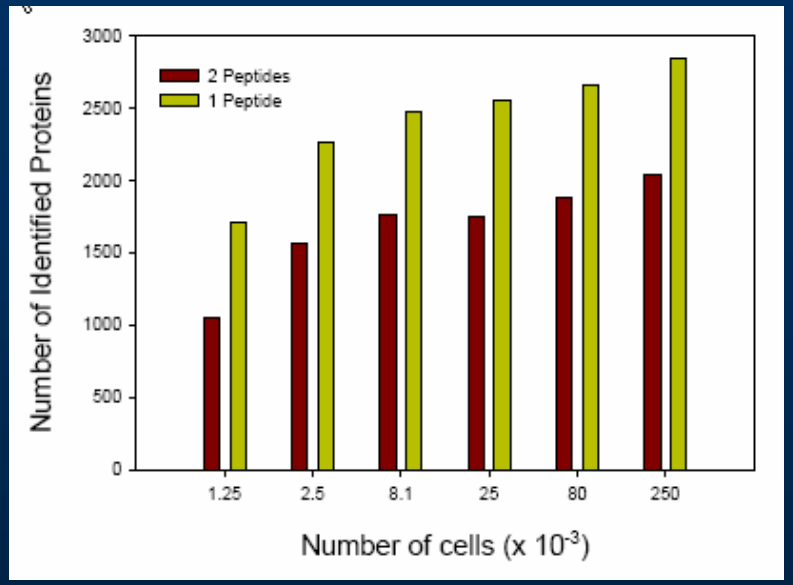
Solution: FASP Solubilizing the Proteome

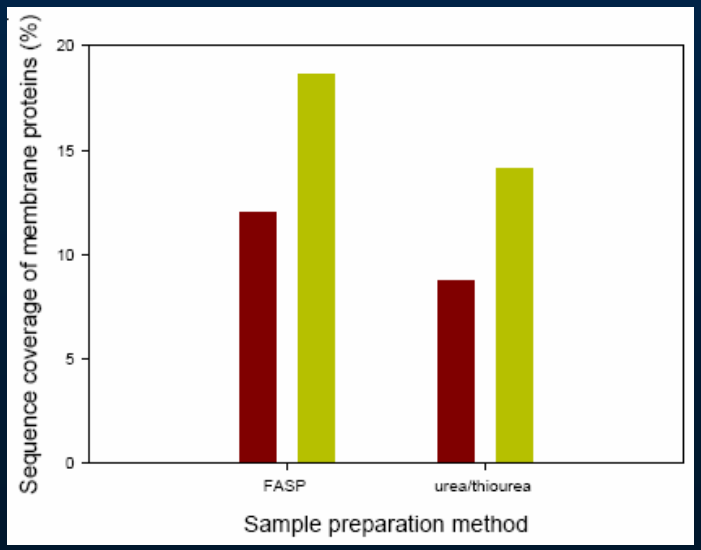


NATURE METHODS | VOL.6 NO.5 | MAY 2009 | 359

Solution: FASP Solubilizing the Proteome Results







NATURE METHODS | VOL.6 NO.5 | MAY 2009 | 359

Acknowledgements

SUMS

- Allis Chien
- Karolina Krasinska
- Pavel Aronov
- Theresa McLaughlin

- Vincent and Stella Coates Foundation