Protein Bioinformatics (260.655)

Lecture 9: Quantitative Proteomics
Tuesday, April 27, 2010

Robert N. Cole, Ph.D.
Mass Spectrometry and Proteomics Facility
Johns Hopkins School of Medicine
371 Broadway Research Bldg
733 N. Broadway St.
Baltimore, MD 21205
Ph: (410) 614-6968
email: rcole@jhmi.edu

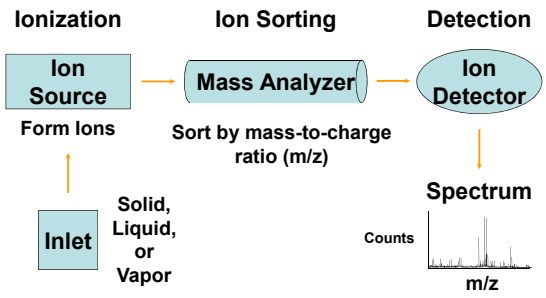
Topics

Protein identification by mass spectrometry

Relative quantification of proteins

Applications

How a Mass Spectrometers Measure Mass



Identifying Proteins by Mass Spectrometry

Three Methods

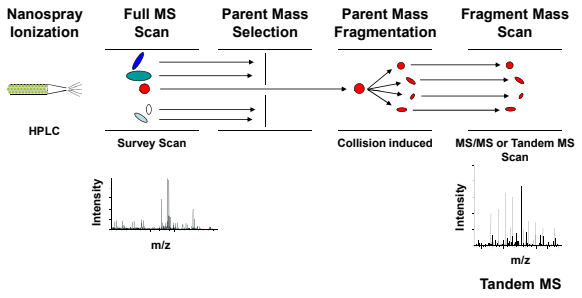
Bottom Up: Proteins identified from **a few peptides**
Know cleavage site, peptide mass and sequence
Do not know proteins actual size and modifications
Problems: Sequence homology, seeing all peptides

Top Down: Proteins identified from **the intact protein**
Know protein's size and whether it is modified
Sequence usually from ends of protein
Problems: Size limitations, Internal sequences

Middle Down: Proteins identified from **large sections of protein**

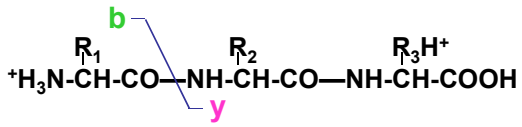
Identifying Proteins from Peptide Sequence

(Many proteins in solution, gel bands or spots)



Tandem MS

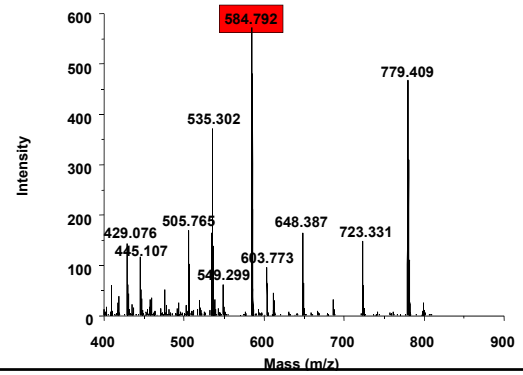
Sequencing Peptides by Fragmentation



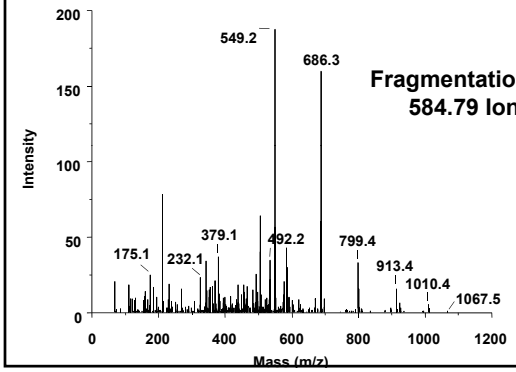
b-ion (b1) $^{+}H_3N\text{-}CH(R_1)\text{-}CO$

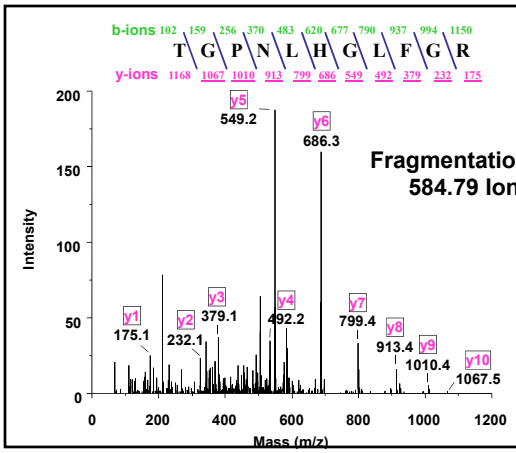
y-ion (y2) $NH\text{-}CH(R_2)\text{-}CO—NH\text{-}CH(R_3H^{+})\text{-}COOH$

Survey (Full MS or MS1) Scan of Peptides



MS/MS (MS2) Scan of Peptide Fragments showing Amino Acid Sequence





Tandem MS (MS/MS)

Uses Peptide Mass and Sequence Tag

Need Only One Peptide Mass with >3 Amino Acids Masses in Sequence (at least two preferred)

High Sample Complexity Tolerated

Protein modifications identified and mapped to an amino acid

High mass accuracy nice, but not required

Search Engines for Protein Identification from MS Data

Summary of Programs

Proteome Software	www.proteomesoftware.com/
ExPASy	expasy.proteome.org.au

Free Programs

ProteinProspector	prospector.ucsf.edu
XProteo	xproteo.com:2698
ProwI	prowl.rockefeller.edu
Mascot	www.matrixscience.com (Free < 300 ions)

Open Source Programs

OMSSA	pubchem.ncbi.nlm.nih.gov/omssa/
X! Tandem	www.thegpm.org/

Commerical Programs

Mascot	www.matrixscience.com
Sequest	fields.scripps.edu/sequest/
Spectrum Mill	www.chem.agilent.com/
Proteolynx	www.waters.com/WatersDivision/

Quantitative Proteomics

Quantifying Individual Proteins in Complex Mixtures (Functional Proteomics)

Quantifying a Proteome Reveals

Changes in protein levels
(*biomarkers, pathways, binding partners*)

Changes in protein modifications
(*structure/function*)

Changes in subcellular localization
(*trafficking*)

Kinetics
(*protein turnover, modification dynamics*)

Quantitative Proteomics Methods

Approach	Methods (Gel or MS based)
Non-Labeling	Gel matching, Densitometry, Spectral Counting, Peak Intensity, Multiple Reaction Monitoring (**MRM**)
Labeling	
Chemical	Difference Gel Electrophoresis (**DIGE**) Isobaric Tags for Relative and Absolute Quantitation (**iTRAQ**) Isotope-Coded Affinity Tags (**ICAT**)
Metabolic	Radiolabeling Stable Isotope Labeling of Amino Acids in Cell Culture (**SILAC**)
Enzymatic	^{18}O-Labeling
Spiking	Absolute Quantification (**AQUA**) Multiple Reaction Monitoring (**MRM**) Quantification Concatamers (**QconCAT**)

Best Approach?

There is NO one best approach.

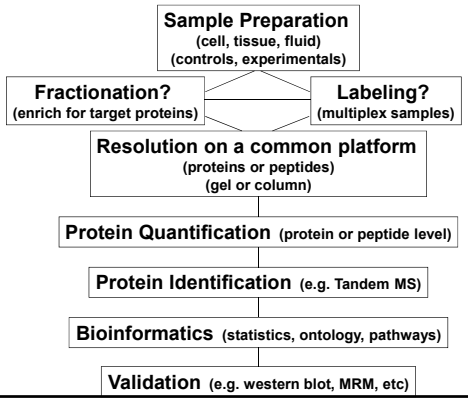
All approaches are:

- Complementary
  - different separation techniques
  - different sets of proteins identified
- Technically challenging
  - sample preparation
  - data acquisition
  - data analysis
- Require fractionation to dig deeper into proteome
  - limited by dynamic range

Best Quantitative Proteomic Experiments

- Defined question (i.e. hypothesis driven)
- Defined system
- Independently measurable phenotype
- Sample preparation
  - Reproducible
  - Scalable
  - Compatible buffer system

Typical Quantitative Proteomic Experiment



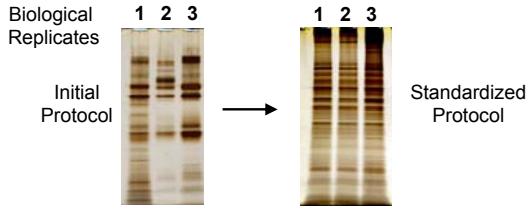
Sample Preparation
(Most Important Step!)

Reproducibility

Protein Amount and Complexity

Buffer Composition Compatibility

Sample Preparation
Must Be Reproducible and Standardized!



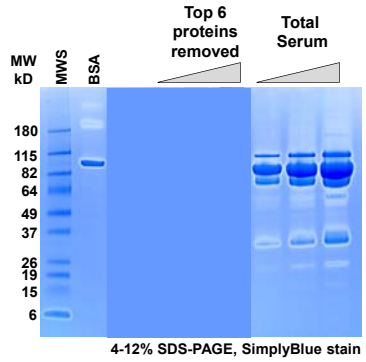
Reduce Sources of Variability:

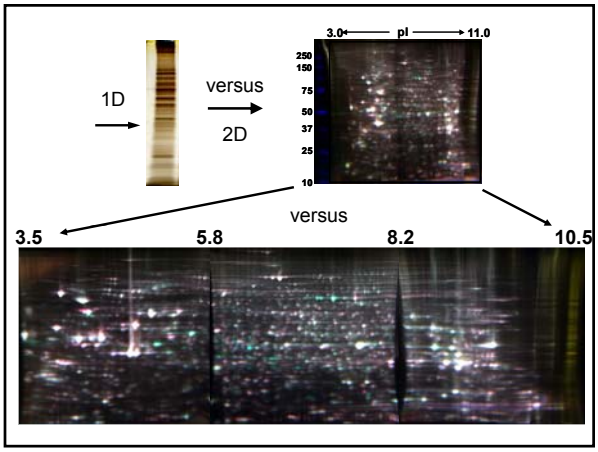
Technical: "good" < "bad hands"
few < many steps

Biological: cells < tissue < body fluids
yeast < nematode < human

Protein Amount and Complexity

Detect Low Abundance Proteins by Fractionating





Quantitative Proteomics Methods

<u>Non-Labeling Methods</u>

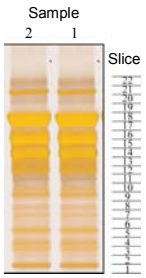
No additions to analysis

Separate analysis of ***<u>Each</u>*** sample

Biological Variability

Technical Variability (Sample Prep and MS analysis)

Non-Labeling – Densitometry -1D Gel



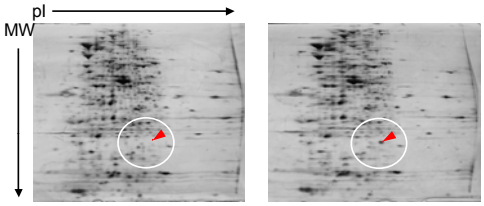
Gel slice number	Name of protein
22	Polymeric immunoglobulin receptor
22	Transferrin
22	Vanin 1
21	1B-glycoprotein
21	Complement component 5
21	hGC-1 (human G-CSF-stimulated clone-1)
21	IgG Fc-binding protein
21	Mac-2-binding protein
18	-1-antichymotrypsin
18	Albumin

Comparing protein bands on same gel
Quantify by densitometry
Often more than one protein per band

Kristiansen et al. Molec Cell Proteomics 3:715–728, 2004.

No Labeling – Densitometry - 2D Gels

Bacterial Strain #1 Bacterial Strain #2

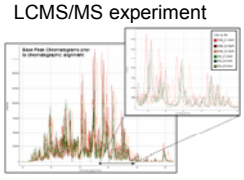


Compare spots from different gels
Quantify by relative spot volume
Gel reproducibility and spot matching critical
Gel warping for matching spots can warp out modification
Often more than one protein per spot

No Labeling – MS Methods

Spectral Counting

Each sample separate LCMS/MS experiment



Compares multiple LCMS/MS experiments

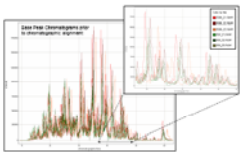
Quantify each protein from:
number of peptides from protein
number of spectra for each peptide

Reproducibility of LCMS/MS system critical

No Labeling – MS Methods

Peak Intensity

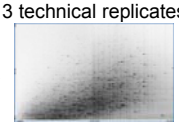
Each sample separate LCMS/MS experiment



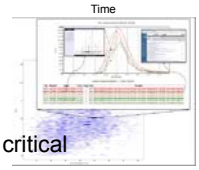
Compares multiple LCMS/MS experiments

Quantify each protein from: intensity of 3 most intense peptides averaged from 3 technical replicates

3D Peak Intensity Map



Annotate "Spots"



Reproducibility of LCMS/MS system critical

Quantitative Proteomics Methods

Non-Labeling Methods

No additions to analysis

Separate analysis of Each sample

Biological and Technical Variability

Labeling Methods

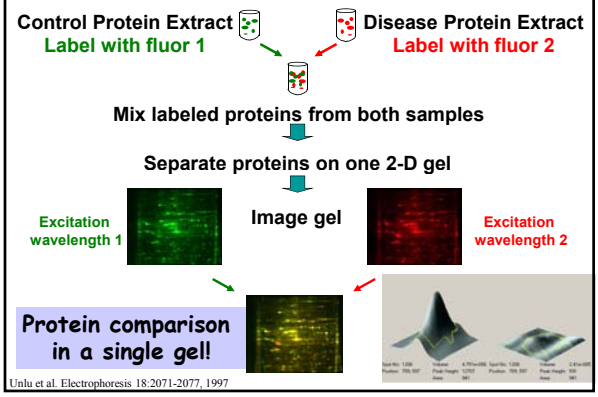
Typically adds steps to analysis

Simultaneous analysis of Many samples (Multiplexing)

Biological Variability, Reduces Technical Variability

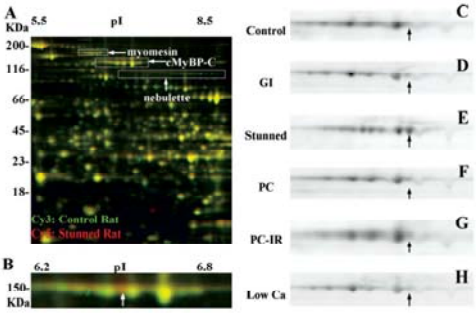
Cuts instrument time (data collection) by 50-75%

Difference Gel Electrophoresis (DIGE)



Unlu et al. Electrophoresis 18:2071-2077, 1997

Novel Phosphorylation of Myofilament Protein Correlates with Myocardial Stunning



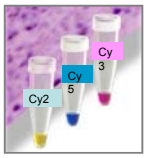
Yuan et al. Proteomics 6:4176-4186, 2006

Three Fluorescent Cy Dyes

Cy2, Cy3, Cy5

ε-amino group of lysine

Matched for Charge and MW



Third Cy Dye used as an Internal Standard

- All possible protein spots overlaid on every gel.
- Simplifies gel to gel matching.
- Each spot has it's own internal standard spot for normalizing across gels.
- Reduces experimental variations.
- Accounts for differences in sample load.

DIGE Analysis

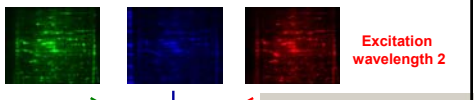
Pooled Samples labeled with Cy2
(on all gels to compare one gel to another)

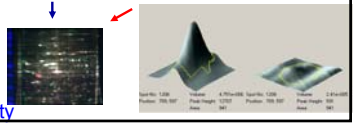
Control Protein Extract
Label with Cy3

Disease Protein Extract
Label with Cy5

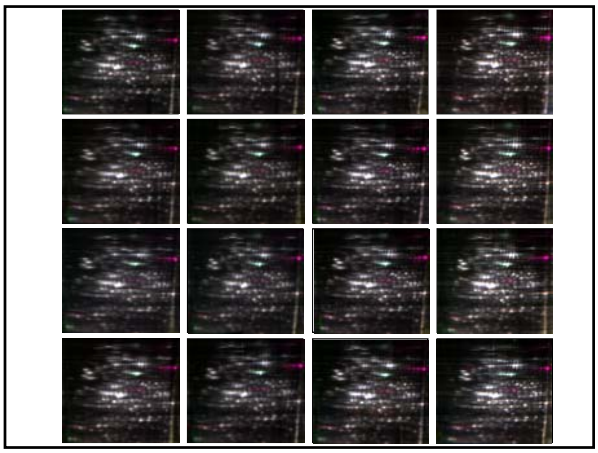
Mix samples

Separate proteins on one 2-D gel





Quantify intact proteins
Cells, tissue, fluids
One additional step
Reduced technical variability



Relative MS Quantitative Methods

Stable Isotope Labeling by Amino Acids in Cell Culture (SILAC)

^{12}C-Arg ^{13}C-Arg

Metabolic labeling
Only for cell culture
No added steps
No technical variability
Standard Deviations 5%

Mix Cells

Isolate proteins and digest with trypsin

Peptide ID and quantify by tandem MS

Quantify mass pairs

Ong et al.
J Proteome Res
2: 173- 181, 2003

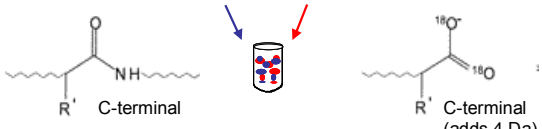
^{18}O-labeling Protocol

Control Protein Extract

Experimental Protein Extract

Trypsin

Trypsin + ^{18}O-water



Enzymatic labeling
Cells, tissue, fluids
No added steps

Mix labeled peptides

Peptide ID and quantify by tandem MS

No MS technical variability
Standard deviations 10-20%
Loss of label to water

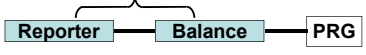
Yao et al. Anal. Chem 73:2836-2842, 2001
Yao et al. J Proteome Res 2:147-152 , 2003

iTRAQ Tags

(Isobaric Tag for Relative and Absolute Quantitation)

Isobaric Tag
(Total mass = 305)

Eight possible tags

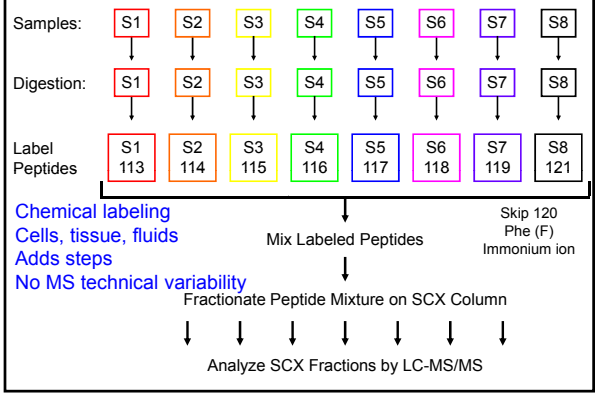


Reporter Mass	Balance Mass
113	192
114	191
115	190
116	189
117	188
118	187
119	186
121	184
Charged	Neutral

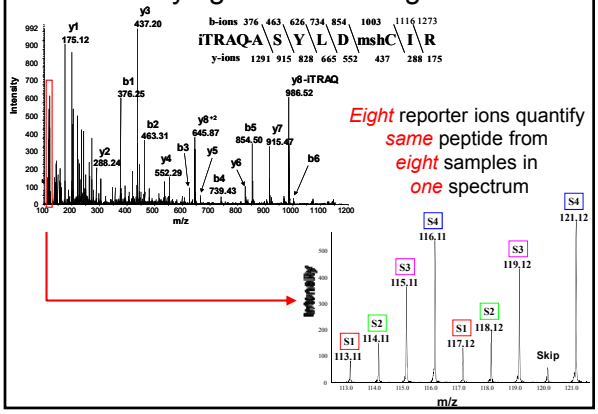
Peptide Reactive Group
(binds to amines)

Ross et al. Mol. Cell. Proteomics 3:1154-1169, 2004
Pierce et al Mol. Cell Proteomics 7:853-863,2007

8-plex iTRAQ Workflow



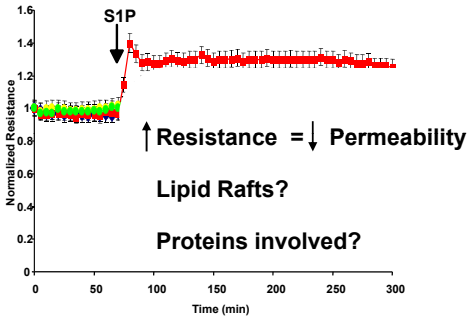
Quantifying Proteins using iTRAQ



Changes in Protein Levels

(***biomarkers and pathways***)

Phospholipid Growth Factor (S1P) on Pulmonary Endothelial Permeability



↑Resistance =↓ Permeability

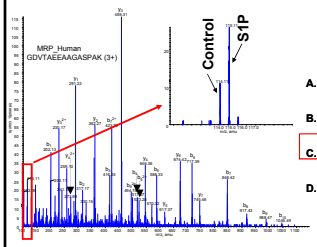
Lipid Rafts?

Proteins involved?

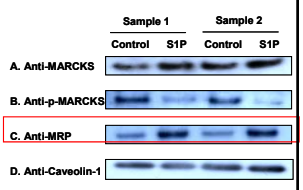
Guo Y et al (2007) Mol Cell Proteomics 6:689-696

iTRAQ Revealed Proteins More Abundant in S1P Stimulated Lipid Rafts

MRP Peptide Fragmentation



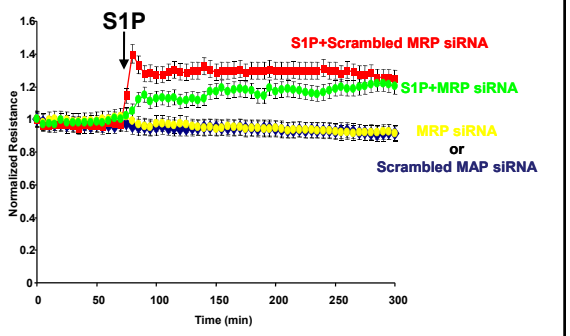
Western Blots Confirm iTRAQ



MRP is up-regulated using both methods

Guo Y et al (2007) Mol Cell Proteomics 6:689-6

MRP siRNA Attenuates S1P Stimulated Endothelial Barrier Enhancement



Guo Y et al (2007) Mol Cell Proteomics 6:689-6

Can more than 8 samples be analyzed using 8-plex iTRAQ?

Yes!

Experimental Design

Pool of all samples: Standard in all iTRAQ experiments

Repeat labeling of at least 1 sample in all iTRAQ experiments

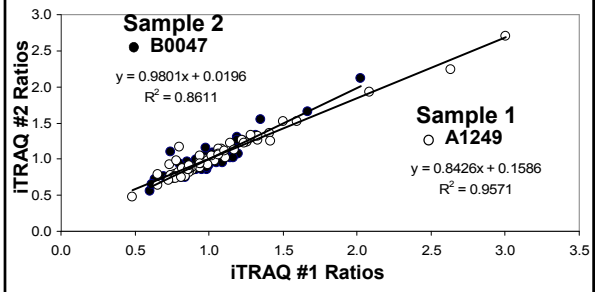
Completely randomize labeling

Expected Result

$$\frac{\text{iTRAQ 1 - Sample 1}}{\text{iTRAQ 1 - Pool}} = \frac{\text{iTRAQ 2 - Sample 1}}{\text{iTRAQ 2 - Pool}}$$

$$\frac{\text{iTRAQ 1 - Sample 2}}{\text{iTRAQ 1 - Pool}} = \frac{\text{iTRAQ 2 - Sample 2}}{\text{iTRAQ 2 - Pool}}$$

Ratios for All Proteins in Sample 1 or Sample 2 Relative to Pool are the Same in iTRAQ 1 and iTRAQ 2



Micronutrient Deficiencies and Health of Undernourished

Child and Maternal Health Problems

Infant/ Child
- Poor growth
- Impaired development
- Disability
- Infection
- Chronic disease
- Childhood Death

Mother
- Obstetric morbidity
- Infection/sepsis
- Anemia
- Death

Nutritional Deficiencies

Micronutrient Status
Vitamin A, zinc, iron, iodine, folate, others

Serum Protein Profiles?
500 samples
75 iTRAQ experiments

Gates Foundation Grant, PI: Keith West, JHSPH

Changes in Protein **M**odifications

(*structure/function*)

Using iTRAQ to Quantify Site Specific Auto-Phosphorylation of the EGF Receptor Kinase

```
   1 MRPSGTAGAALLALLAALCPASRALEEKKVCQGTSNKLTQLGTFEDHFLS
  51 LQRMFNNCEVVLGNLEITYVQRNYDLSFLKTIQEVAGYVLIALNTVERIP
 101 LENLQIIRGNMYYENSYALAVLSNYDANKTGLKELPMRNLQEILHGAVRF
 151 SNNPALCNVESIQWRDIVSSDFLSNMSMDFQNHLGSCQKCDPSCPNGSCW
 201 GAGEENCQKLTKIICAQQCSGRCRGKSPSDCCHNQCAAGCTGPRESDCLV
 251 CRKFRDEATCKDTCPPLMLYNPTTYQMDVNPEGKYSFGATCVKKCPRNYV
 301 VTDHGSCVRACGADSYEMEEDGVRKCKKCEGPCRKVCNGIGIGEFKDSLS
 351 INATNIKHFKNCTSISGDLHILPVAFRGDSFTHTPPLDPQELDILKTVKE
 401 ITGFLLIQAWPENRTDLHAFENLEIIRGRTKQHGQFSLAVVSLNITSLGL
 451 RSLKEISDGDVIISGNKNLCYANTINWKKLFGTSGQKTKIISNRGENSCK
 501 ATGQVCHALCSPEGCWGPEPRDCVSCRNVSRGRECVDKCNLLEGEPREFV
 551 ENSECIQCHPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCVKTCPAGVM
 601 GENNTLVWKYADAGHVCHLCHPNCTYGCTGPGLEGCPTNGPKIPSIATGM
 651 VGALLLLLVVALGIGLFMRRRHIVRKRTLRRLLQERELVEPLTPSGEAPN
 701 QALLRILKETEFKKIKVLGSGAFGTVYKGLWIPEGEKVKIPVAIKELREA
 751 TSPKANKEILDEAYVMASVDNPHVCRLLGICLTSTVQLITQLMPFGCLLD
 801 YVREHKDNIGSQYLLNWCVQIAKGMNYLEDRRLVHRDLAARNVLVKTPQH
851 VKITDFGLAKLLGAEEKEYHAEGGKVPIKWMALESILHRIYTHQSDVWSY
901 GVTVWELMTFGSKPYDGIPASEISSILEKGERLPQPPICTIDVYMIMVKC
951 WMIDADSRPKFRELIIEFSKMARDPQRYLVIQGDERMHLPSPTDSNFYR
```

Tyr[869] Tyr[998]

Qiu et al. 2009 Biochemistry 48: 6624-6632

Experimental Design

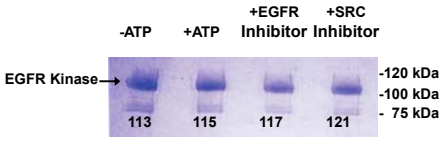
Expressed EGFR Kinase

Incubated +/- ATP and +/- kinase inhibitors

Isolated EGFR Kinase

Resolved EGFR Kinase by SDS-PAGE

In gel digest, iTRAQ label, SCX, LCMS/MS



Qiu et al. 2009 Biochemistry 48: 6624-6632

EGFR Top Hit
>35 peptide IDs with at least 95% confidence
Same amount of EGRF in all gel bands
(Ratios relative to No ATP sample labeled with 113)

					+ATP	+EGFR kinase Inhibitor	+SRC kinase Inhibitor
Rank	Score	%Cov	Name	Species	115:113	117:113	121:113
1	70	65	Epidermal growth factor receptor precursor - Homo sapiens (Human)	HUMAN	1.10	1.15	1.32
2	9	49	Keratin, type II cytoskeletal 1 - Homo sapiens (Human)	HUMAN	0.76	1.00	1.20
3	8	43	Exportin-2 - Homo sapiens (Human)	HUMAN	0.98	0.89	1.31
4	5	36	Exportin-4 - Homo sapiens (Human)	HUMAN	0.98	0.98	1.25
5	4	39	Exportin-1 - Homo sapiens (Human)	HUMAN	0.94	1.00	1.29
6	2	46	Exportin-T - Homo sapiens (Human)	HUMAN	1.07	1.19	1.56
7	2	34	Keratin, type I cytoskeletal 10 - Homo sapiens (Human)	HUMAN	0.70	0.75	1.21
8	2	71	Uncharacterized protein C14orf139 - Homo sapiens (Human)	HUMAN	1.15	0.87	1.12

Qiu et al. 2009 Biochemistry 48: 6624-6632

Phosphotyrosines: **Increase** with ATP or Inhibitor to another kinase, but ***NOT*** with an Inhibitor to EGFR Kinase.
Phosphothreonines: No change
Phosphoserines: No change

Confidence	Sequence	Phosphate Modification	+ATP Alone 115:113	+EGFR Kinase Inhibitor 117:113	+Scr Kinase Inhibitor 121:113
99	ELVEPL**T**PSGEAPNQALLR	**Phospho(T)**@7	1.14	1.01	1.52
99	LLGAEEKE**Y**HAEGGK	**Phospho(Y)**@9	**16.98**	1.71	**14.62**
99	LLGAEEKE**Y**HAEGGKVPIK	**Phospho(Y)**@9	**5.18**	1.27	**3.99**
99	LLGAEEKE**Y**HAEGGKVPIK	**Phospho(Y)**@9	**6.33**	1.27	**4.89**
99	MHLPSPTDSNF**Y**R	**Phospho(Y)**@12	**2.71**	1.63	**2.88**
99	MHLPSPTDSNF**Y**R	**Phospho(Y)**@12	**2.24**	1.15	**2.28**
99	MHLPSPTDSNF**Y**R	**Phospho(Y)**@12	**2.00**	0.78	**1.75**
99	MHLP**S**PTDSNFYR	**Phospho(S)**@5	0.77	0.80	0.98
70	MHLP**S**PTDSNFYR	**Phospho(S)**@5	1.10	1.21	1.74

Qiu et al. 2009 Biochemistry 48: 6624-6632

Kinetics

(*protein turnover, modification dynamics*)

Which peptide is a better substrate?

Expermental Design: **Wade Gibson**

Two time courses (one for each peptide)

Two iTRAQ experiments (one for each time course)

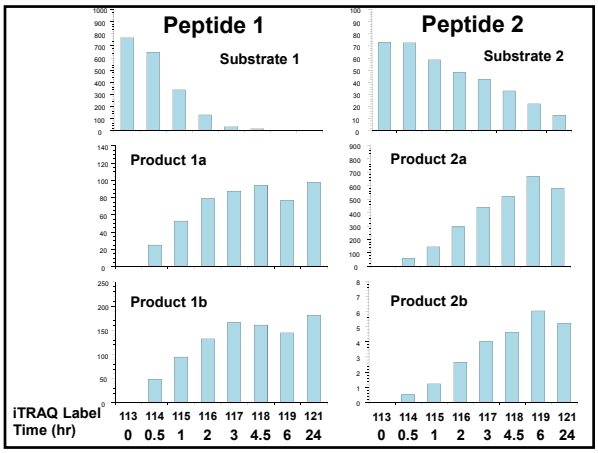
Substrate and products labeled with *same* iTRAQ tag at *each* time point reagent

***Different* iTRAQ label for *different* time points**

Time (hr)	iTRAQ Label
0	113
0.5	114
1	115
2	116
3	117
4.5	118
6	119
24	121

Mix *all* iTRAQ labeled substrates and proteins from *all* time points

Run *one* MS analysis (LCMS/MS) for *each* time course



Software for Identifying and Quantifying Proteins

Label Free

Sieve	www.thermo.com
MSQuant	msquant.alwaysdata.net

Labeling

ProteinPilot	www.absciex.com
Mascot	www.matrixscience.com
Scaffold Q+	www.proteomesoftware.com
Protein Discoverer	www.thermo.com