Protein Bioinformatics (260.655)

Lecture 15: Quantitative Proteomics

Tuesday, May 5, 2008

Robert N. Cole, Ph.D.
Mass Spectrometry/Proteomics Facility at
Johns Hopkins School of Medicine
371 Broadway Research Bldg
733 N. Broadway St.
Baltimore, MD 21205
Ph: (410) 614-6968/Fax: (410) 614-7566
email: rcole@jhmi.edu

Progression of Protein Analysis by Mass Spectrometry

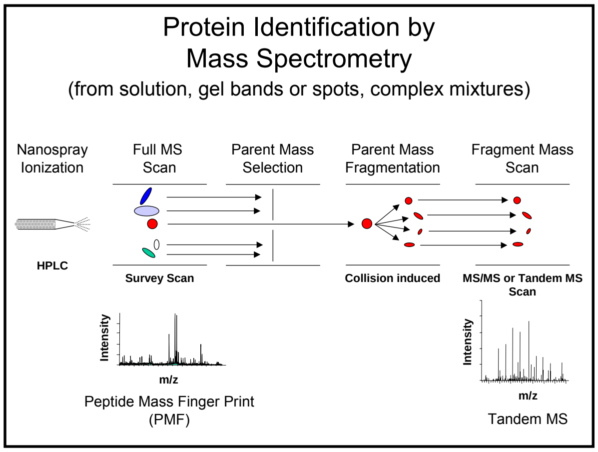
Analysis	>5 years ago	Now
De novo sequencing	Hard	Still Hard
Identify isolated protein	Easy	Routine
Identify protein mixtures	Hard	Routine
Quantify proteins	Hard	Routine?
Identify/Map Modifications	Hard	Not Easy
Quantify PTMs	Hard	Still Hard

Topics

Protein identification by mass spectrometry

Sample preparation

Relative quantification of proteins

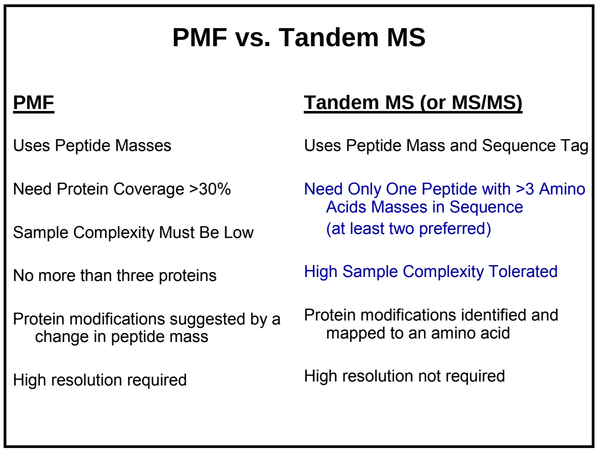


Protein Identification by Mass Spectrometry

(from solution, gel bands or spots, complex mixtures)

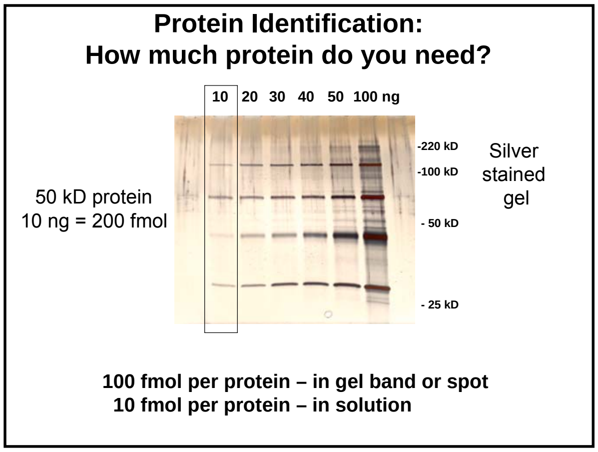
Peptide Mass Finger Print (PMF)

Tandem MS



PMF vs. Tandem MS

PMF	**Tandem MS (or MS/MS)**
Uses Peptide Masses	Uses Peptide Mass and Sequence Tag
Need Protein Coverage >30%	Need Only One Peptide with >3 Amino Acids Masses in Sequence (at least two preferred)
Sample Complexity Must Be Low	
No more than three proteins	High Sample Complexity Tolerated
Protein modifications suggested by a change in peptide mass	Protein modifications identified and mapped to an amino acid
High resolution required	High resolution not required



Protein Identification: How much protein do you need?

50 kD protein
10 ng = 200 fmol

Silver stained gel

100 fmol per protein – in gel band or spot
10 fmol per protein – in solution

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
Prowl	prowl.rockefeller.edu
Mascot	www.matrixscience.com (Free up to 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/

Next Step: Quantify Individual Proteins in Complex Mixtures (Functional Proteomics)

Identify proteins expressed at different levels in diseases, mutant cell lines, during developmental stages, etc.

Determine stoichiometry of proteins in protein complexes

Determine protein subcellular localization

Identify protein binding partners

Follow protein trafficking

Map metabolic pathways

…many, many others

Quantitative Proteomics Methods

Approach	Labeling	Method
2D Gel Based	None	Gel matching
	Metabolic	Radiolabeling
	Chemical	Difference Gel Electrophoresis (**DIGE**)
MS Based	None	Spectral Counting, Peak Intensity
	Spiking	Absolute Quantification (AQUA)
	Metabolic	Stable Isotope Labeling of Amino Acids in Cell Culture (SILAC)
	Enzymatic	^{18}O-Labeling
	Chemical	Isotope-Coded Affinity Tags (ICAT), Isobaric Tags for Relative and Absolute Quantitation (**iTRAQ**)

Quantitative Proteomics Strategies

Sample Preparation	(sample complexity)
Fractionation?	(low abundance proteins)
Protein/Peptide Label?	
Fluorescent tags	(Cy dyes)
Stable isotopes	(^{13}C, ^{15}N, ^{18}O)
Mass tags	(Signature mass)
Protein/Peptide Separation?	
Gel Electrophoresis	(MW, pI)
Chromatography	(MW, charge, affinity)
Both	
Quantification	(comparison)
Protein Identification	(by mass spectrometry)

Sample Preparation

Most important step!

Reproducibility essential!

Variability in sample prep:

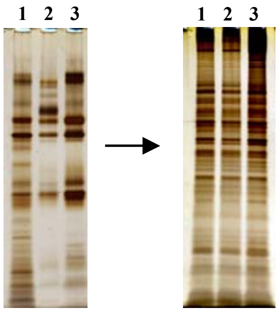
Technical: “good” < “bad hands”
few < many steps

Biological: cells < tissue < serum
yeast < mouse < human

Sample Preparation Must Be Reproducible and Standardized!

Biological Replicates

Initial Protocol



Standardized Protocol

Sample Preparation

Most important step!

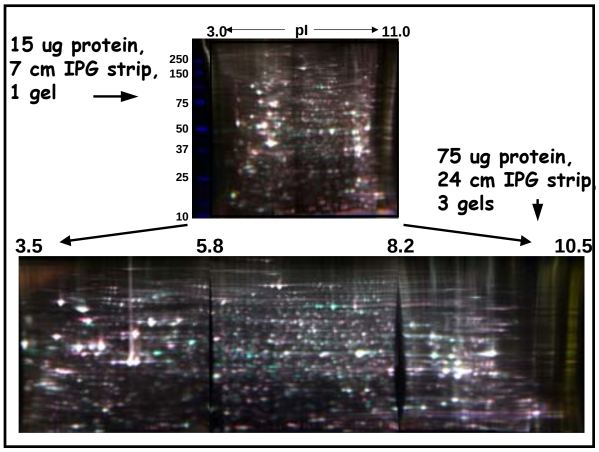
Reproducibility essential!

Variability in sample prep:

- Technical: "good" < "bad hands"
  - few < many steps
- Biological: cells < tissue < serum
  - yeast < mouse < human

Detecting low abundance proteins requires

- Fractionation (subcellular or protein)
- More protein

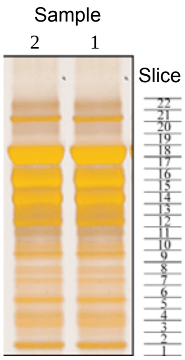


15 ug protein,
7 cm IPG strip,
1 gel

75 ug protein,
24 cm IPG strip
3 gels

Gel Based Techniques

1D Gel Analysis



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
17	Amylase, pancreatic, -2A
17	Amylase, salivary, -1A
17	Catalase

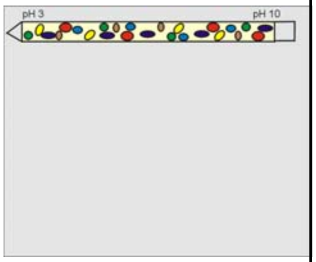
Kristiansen et al. Molec Cell Proteomics 3:715–728, 2004.

1D Gel Analysis

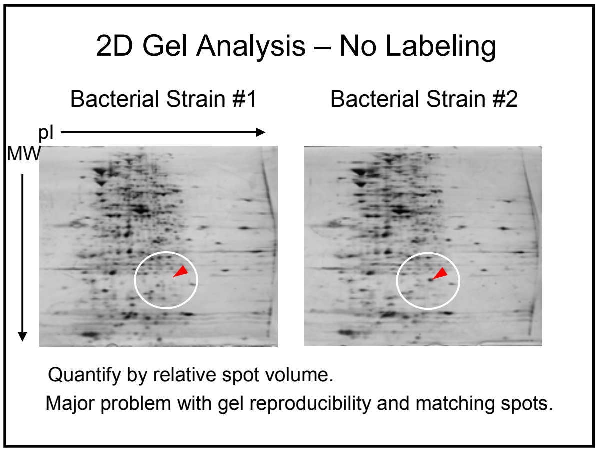
Strengths:	Limitations:
• Simple	• Time consuming
• In expensive sample prep	• Not Quantitative
• Membrane proteins	• Limited resolution
• High MW proteins (>100 kDa)	• Low MW Proteins (<7kDa)

Principle of 2-D Gel Electrophoresis

First dimension:
according to the isoelectric point
(isoelectric focusing)



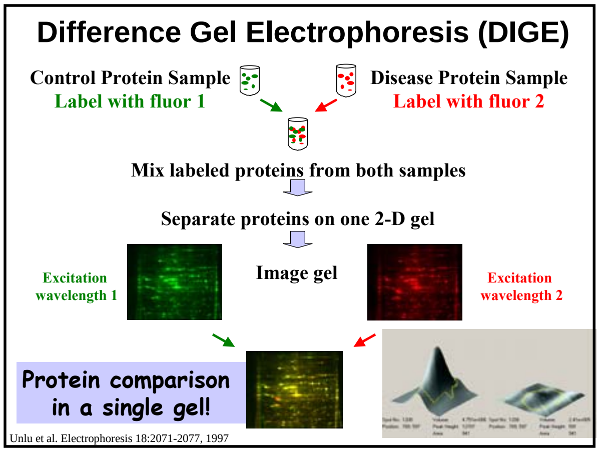
Second dimension:
separation according to molecular weight
(SDS gel electrophoresis)



2D Gel Analysis – No Labeling

Quantify by relative spot volume.
Major problem with gel reproducibility and matching spots.

Gel Based Techniques
with
Fluorescent Tags



Difference Gel Electrophoresis (DIGE)

Control Protein Sample
Label with fluor 1

Disease Protein Sample
Label with fluor 2

Mix labeled proteins from both samples

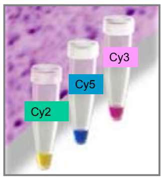
Separate proteins on one 2-D gel

Image gel

Protein comparison in a single gel!

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

Three Fluorescent Cy Dyes



Cy2, Cy3, Cy5

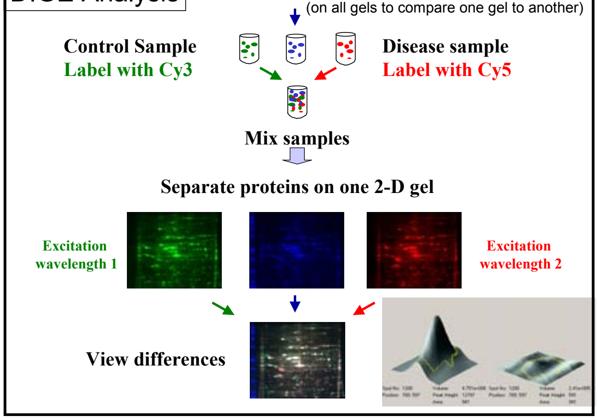
ε-amino group of lysine

Matched for charge (carries +1 charge)

Matched for MW (~450 Da)

Only 3-5% of proteins are labeled (minimal labeling)

DIGE Analysis



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

Control Sample
Label with Cy3

Disease sample
Label with Cy5

Mix samples

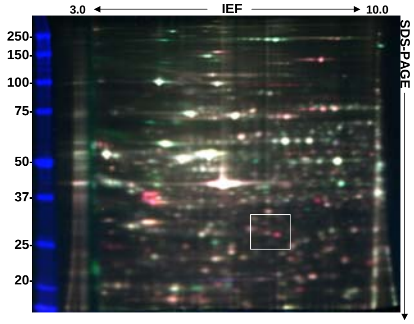
Separate proteins on one 2-D gel

Excitation wavelength 1

Excitation wavelength 2

View differences

Differential Protein Expression Between Two Glial Cell Lines



Third Cy Dye (Cy2) Labeling as an Internal Standard

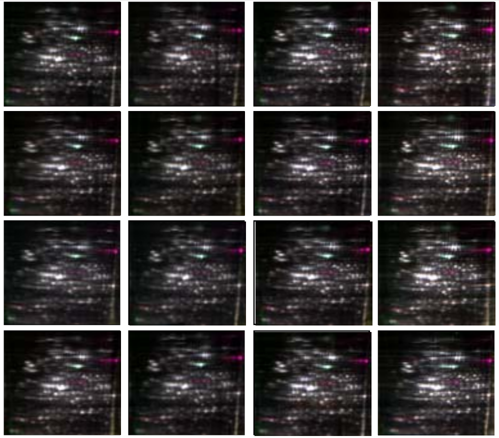
All possible protein spots overlaid on every gel.

Simplifies gel to gel matching.

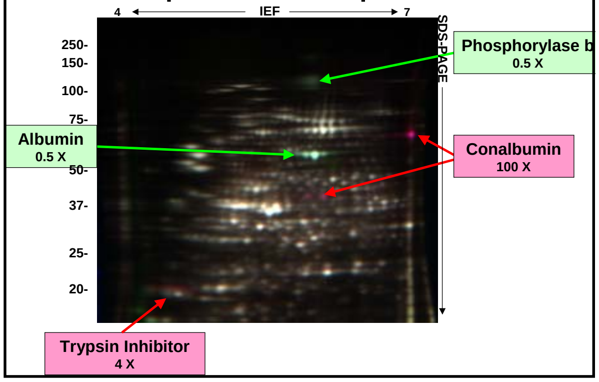
Each spot has it’s own internal standard spot for normalizing across gels.

Reduces experimental variations.

Accounts for differences in sample load.



DIGE Analysis of *Ecoli* proteins spiked with four proteins



Relative Abundances of Four Proteins in *Ecoli* Lysates are within 10% of Predicted Values

Spiked Proteins	Expected Ratio	Observed Ratio	**Error**	t-test
Albumin	0.5	0.51	**3%**	2.7E-06
Conalbumin	100	89.67	**10%**	1.6E-14
Phosphorylase B	0.5	0.46	**8%**	1.3E-10
Trypsin inhibitor	4	3.84	**4%**	0.00014

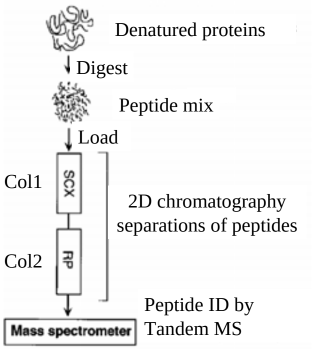
Strengths/Limitations of 2D-DIGE Gels

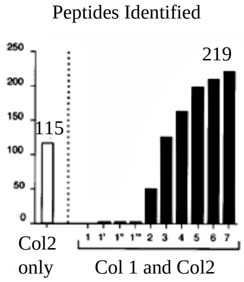
Strengths	Limitations
Superb resolving power	<7,000 and >130,000 Da
Quantitative comparisons of ~2000 proteins on one gel	Strongly hydrophobic Highly basic proteins
Simplified comparison between gels	Difficult to automate
Highly reproducible	Multiple proteins per spot
Detect post-translational modifications	Quantify but not identify, still need MS analysis

Mass Spectrometry Based Techniques

Multi-dimensional Chromatography

(Shotgun or MuDPIT)





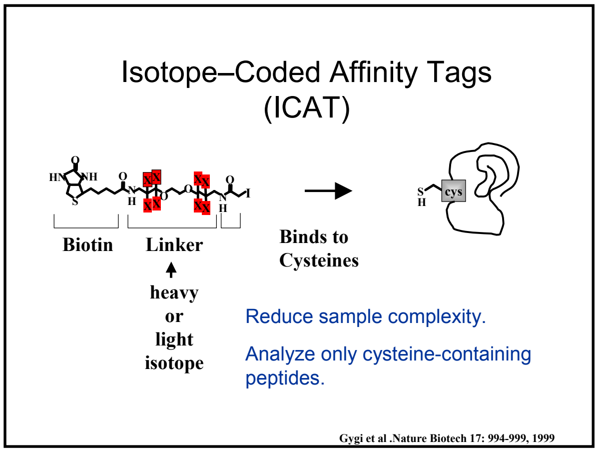
Link et al. Nat. Biotech 17:676-682, 1999

Summary of Multi-Dimensional Chromatography

Strengths:	Limitations:
• Reduces sample complexity	• Complex setup
• Low abundance proteins	• No faster than 2D gels
• Membrane proteins	• Computer intensive
• All MW Proteins	• Quantitative?
• Flexible (different column combinations)	• Reproducibility
• Can automate	

Mass Spectrometry Based Techniques

with
Stable Isotopes or Mass Tags

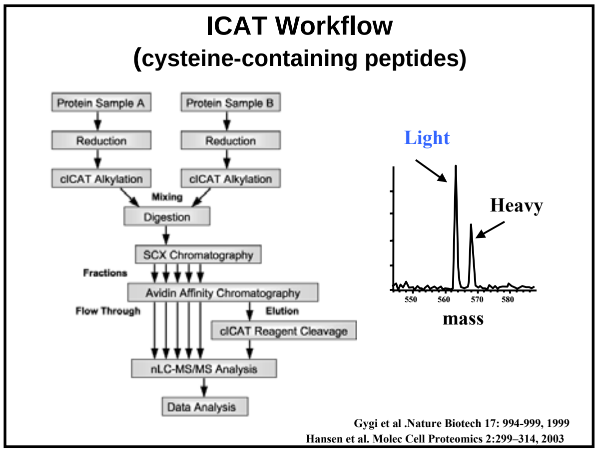


Isotope–Coded Affinity Tags (ICAT)

Reduce sample complexity.

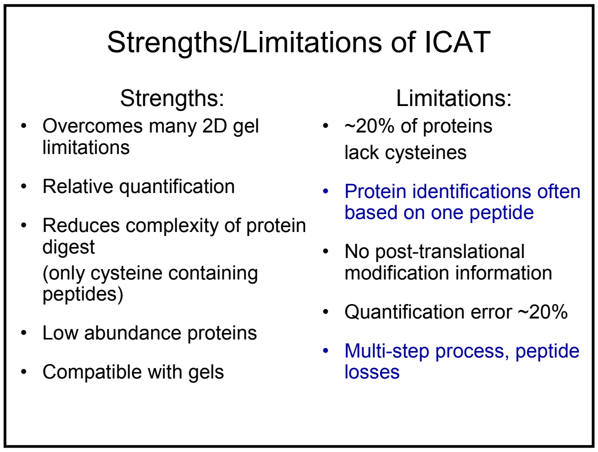
Analyze only cysteine-containing peptides.

Gygi et al .Nature Biotech 17: 994-999, 1999



ICAT Workflow
(cysteine-containing peptides)

Gygi et al .Nature Biotech 17: 994-999, 1999
Hansen et al. Molec Cell Proteomics 2:299–314, 2003



Strengths/Limitations of ICAT

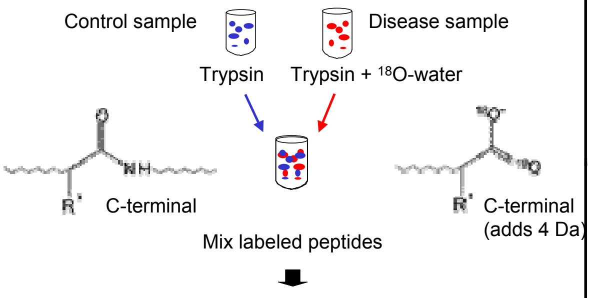
Strengths:

- Overcomes many 2D gel limitations
- Relative quantification
- Reduces complexity of protein digest
  (only cysteine containing peptides)
- Low abundance proteins
- Compatible with gels

Limitations:

- ~20% of proteins lack cysteines
- Protein identifications often based on one peptide
- No post-translational modification information
- Quantification error ~20%
- Multi-step process, peptide losses

^{18}O-labeling Protocol


Control sample

Disease sample

Trypsin

Trypsin + ^{18}O-water

C-terminal

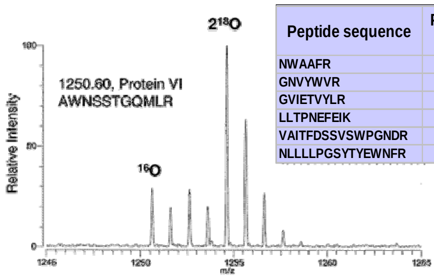
C-terminal
(adds 4 Da)

Mix labeled peptides

Analyze by mass spectrometry

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

Quantify by ^{18}O to ^{16}O peptide ratios



Peptide sequence	Peak intensity ratio ± error	
NWAAFR	2.5 ± 0.1	(4%)
GNVYWVR	3.4 ± 0.7	(20%)
GVIETVYLR	3.0 ± 0.3	(10%)
LLTPNEFEIK	3.1 ± 0.5	(19%)
VAITFDSSVSWPGNDR	2.9 ± 0.4	(14%)
NLLLLPGSYTYEWNFR	2.8 ± 0.2	(7%)

Yao et al. Anal. Chem 73:2836-2842, 2001

Strengths/Limitations of ^{18}O-Labeling

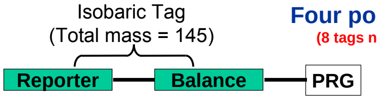
Strengths:

- No extra steps
- Relative quantification using multiple peptides from same protein
- Post-translational modification information
- ^{18}O labeling kit (Prolytica)

Limitations:

- No reduction in sample complexity
- Only 4 amu mass increase
- ^{18}O exchange with water
- Quantification error ~20%
- No good software yet

iTRAQ Tags
(Isobaric Tag for Relative and Absolute Quantitation)



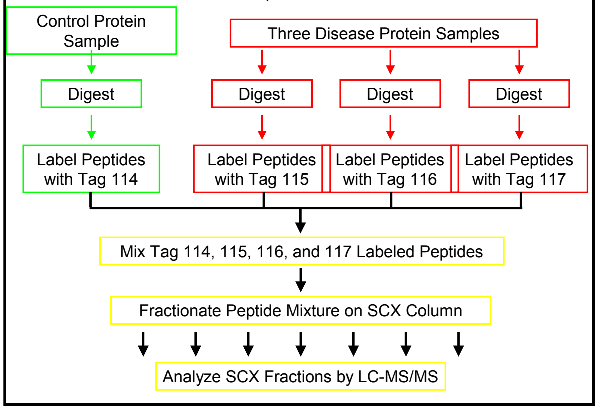
Four possible tags
(8 tags now available)

Reporter Mass	Balance Mass
114	31
115	30
116	29
117	28

Peptide Reactive Group
(binds to amines)

Ross et al. Mol. Cell. Proteomics 3:1154-1169, 2004

iTRAQ Workflow 1



Control Protein Sample → Digest → Label Peptides with Tag 114

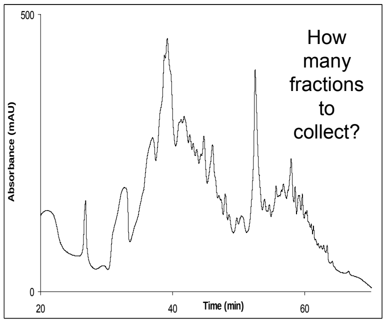
Three Disease Protein Samples → Digest → Label Peptides with Tag 115 / Label Peptides with Tag 116 / Label Peptides with Tag 117

Mix Tag 114, 115, 116, and 117 Labeled Peptides

Fractionate Peptide Mixture on SCX Column

Analyze SCX Fractions by LC-MS/MS

Strong Cation Exchange (SCX) Chromatography
Removes labeling reagents and buffers
Reduces sample complexity

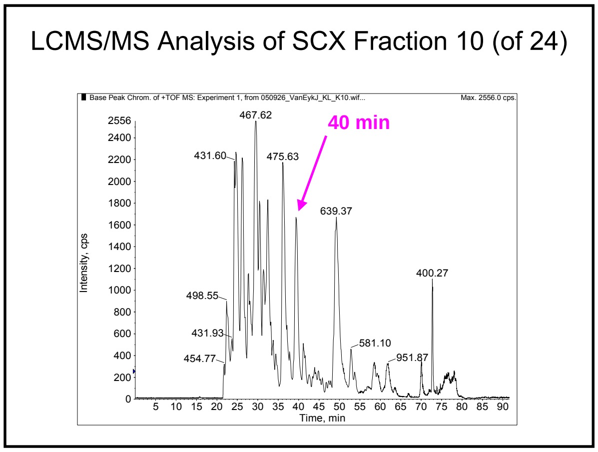


How many fractions to collect?

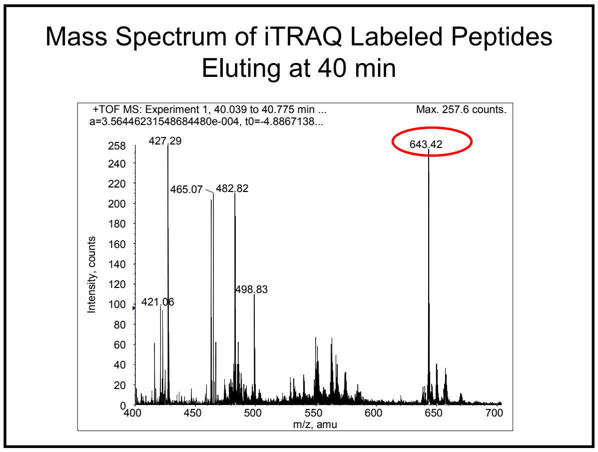
Depends on

sample complexity

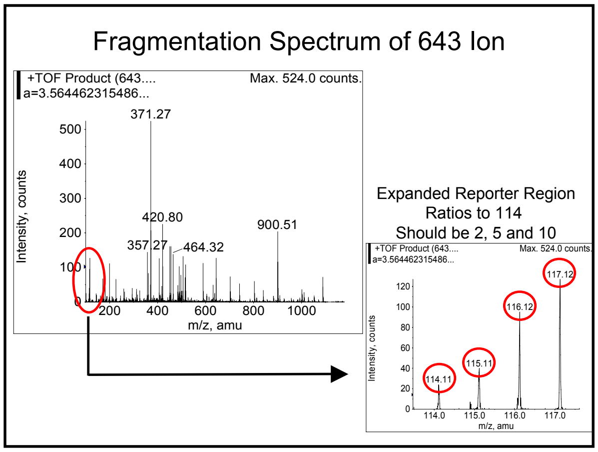
speed of mass spectrometer



LCMS/MS Analysis of SCX Fraction 10 (of 24)



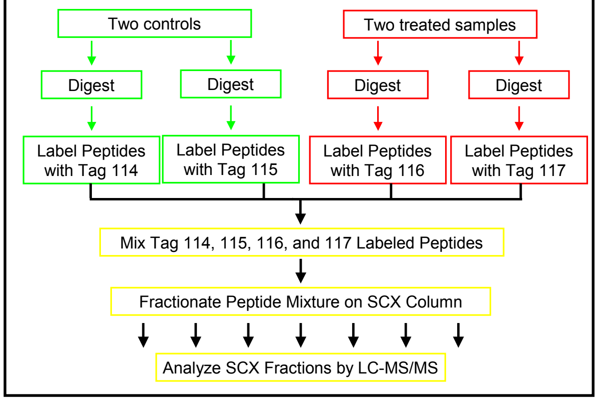
Mass Spectrum of iTRAQ Labeled Peptides Eluting at 40 min



Fragmentation Spectrum of 643 Ion

Expanded Reporter Region
Ratios to 114
Should be 2, 5 and 10

iTRAQ Workflow 2 (preferred)



Two controls

Two treated samples

Digest

Digest

Digest

Digest

Label Peptides with Tag 114

Label Peptides with Tag 115

Label Peptides with Tag 116

Label Peptides with Tag 117

Mix Tag 114, 115, 116, and 117 Labeled Peptides

Fractionate Peptide Mixture on SCX Column

Analyze SCX Fractions by LC-MS/MS

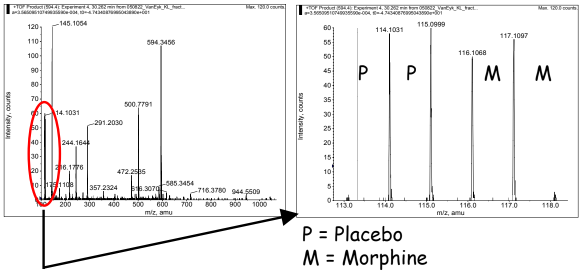
Protein: ATP SYNTHASE BETA CHAIN, MITOCHONDRIAL PRECURSOR

Peptide: **VLDSGAPIJ**

No Change

Fragmentation Spectrum

Expanded Reporter Region



P = Placebo
M = Morphine

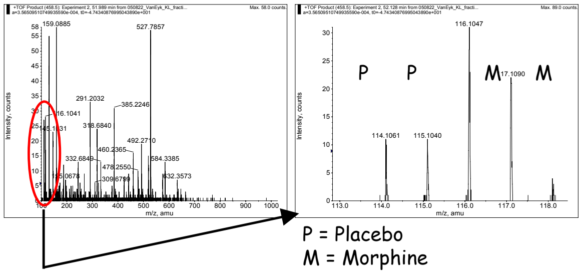
Protein: Cytochrome c oxidase subunit 1

Peptide: **VFSWLATLHGGNIJ**

Up Regulated!

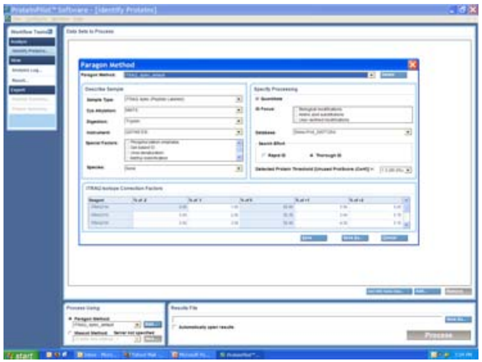
Fragmentation Spectrum

Expanded Reporter Region

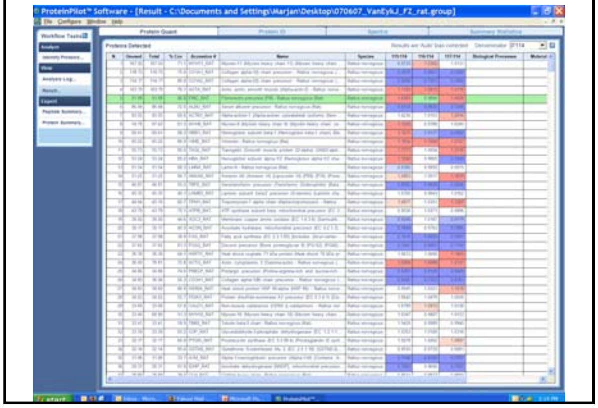


P = Placebo
M = Morphine

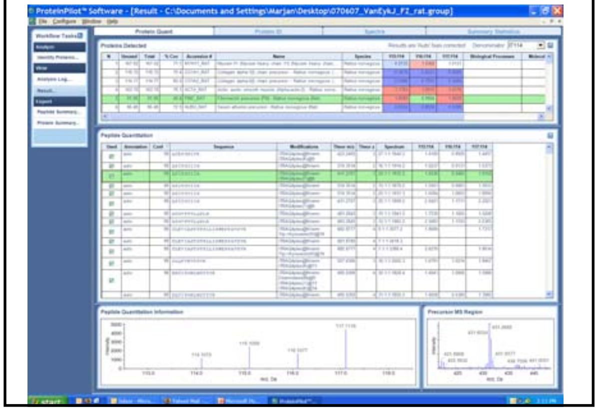
ProteinPilot
software for identifying and quantifying of proteins
based on iTRAQ chemistry



Protein Lists



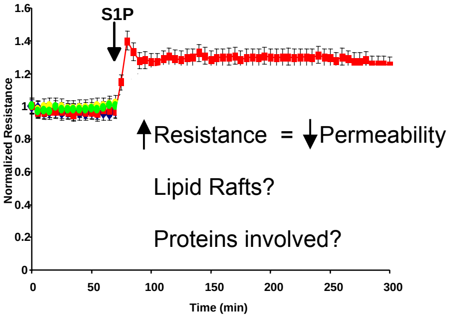
Quantification based on multiple ratios of same peptide



View Spectrum

ProteinPilot™ Software - [Result - C:\Documents and Settings\Marjan\Desktop\070607_VanEykJ_F2_rat.group]

Phospholipid Growth Factor (S1P) on Pulmonary Endothelial Permeability



↑Resistance = ↓Permeability

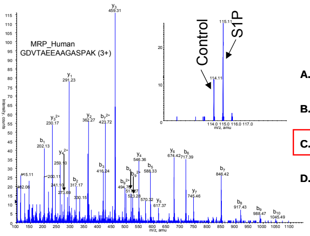
Lipid Rafts?

Proteins involved?

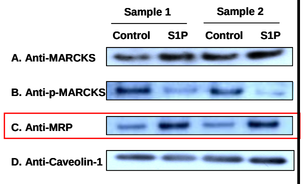
Guo Y et al (2007) Mol Cell Proteomics 6:689-696

What is next?
Confirmation by an independent technique

iTRAQ Results

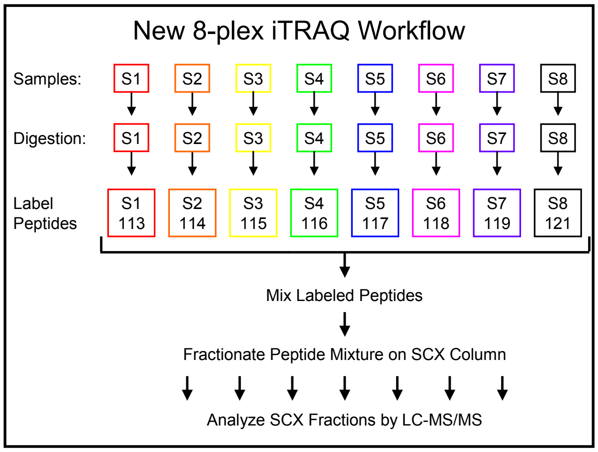


Western blots

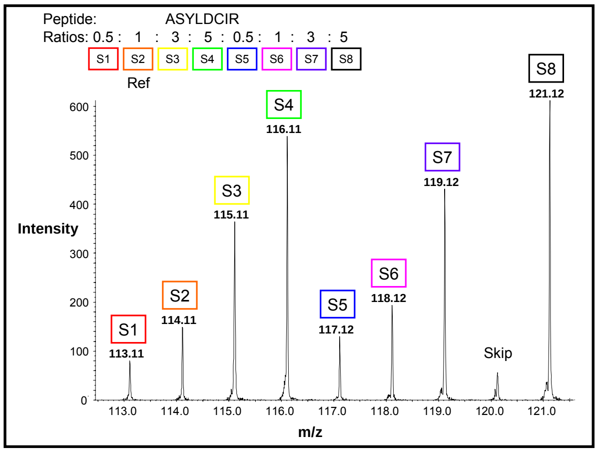


MRP is up-regulated using both methods

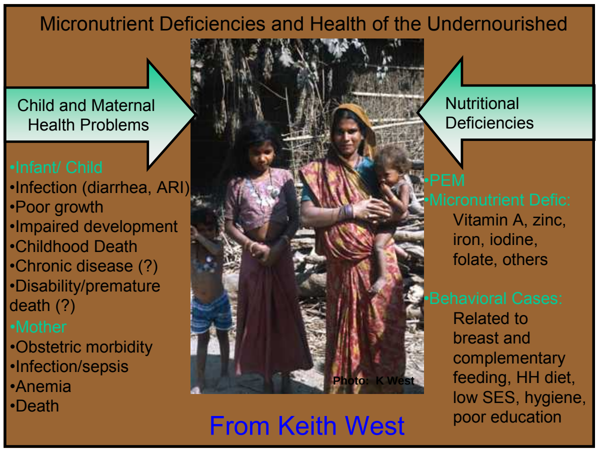
Guo Y et al (2007) Mol Cell Proteomics 6:689-696



New 8-plex iTRAQ Workflow



Peptide: ASYLDCIR
Ratios: 0.5 : 1 : 3 : 5 : 0.5 : 1 : 3 : 5



Micronutrient Deficiencies and Health of the Undernourished

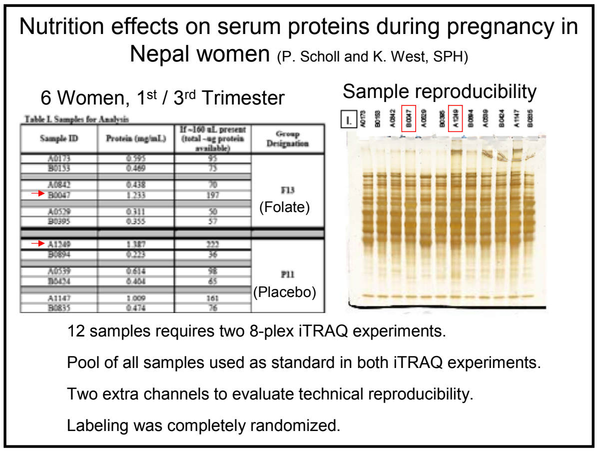
Child and Maternal Health Problems

- •Infant/ Child
- •Infection (diarrhea, ARI)
- •Poor growth
- •Impaired development
- •Childhood Death
- •Chronic disease (?)
- •Disability/premature death (?)
- •Mother
- •Obstetric morbidity
- •Infection/sepsis
- •Anemia
- •Death

Nutritional Deficiencies

- •PEM
- •Micronutrient Defic: Vitamin A, zinc, iron, iodine, folate, others
- •Behavioral Cases: Related to breast and complementary feeding, HH diet, low SES, hygiene, poor education

From Keith West



Nutrition effects on serum proteins during pregnancy in Nepal women (P. Scholl and K. West, SPH)

6 Women, 1^{st} / 3^{rd} Trimester

Table I. Samples for Analysis

Sample ID	Protein (mg/mL)	If ~160 uL present (total ~ug protein available)	Group Designation
A0173	0.595	95	F13 (Folate)
B0153	0.469	75	
A0842	0.438	70	
→ B0047	1.233	197	
A0529	0.311	50	
B0395	0.355	57	
→ A1249	1.387	222	P11 (Placebo)
B0894	0.223	36	
A0539	0.614	98	
B0424	0.404	65	
A1147	1.009	161	
B0835	0.474	76	

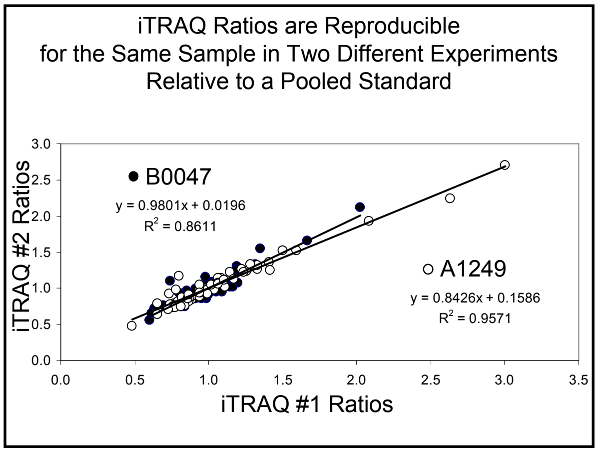
Sample reproducibility

12 samples requires two 8-plex iTRAQ experiments.

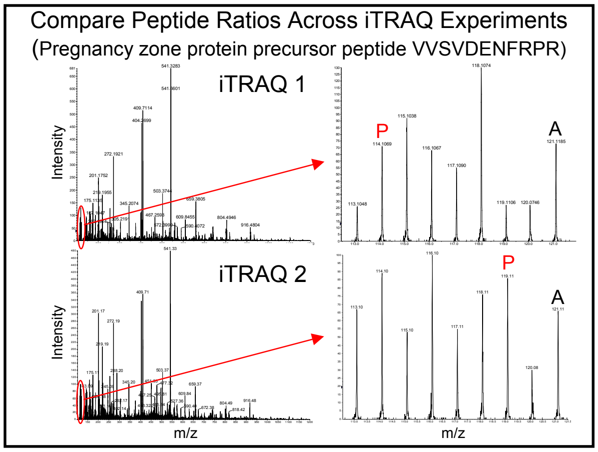
Pool of all samples used as standard in both iTRAQ experiments.

Two extra channels to evaluate technical reproducibility.

Labeling was completely randomized.



iTRAQ Ratios are Reproducible for the Same Sample in Two Different Experiments Relative to a Pooled Standard



Compare Peptide Ratios Across iTRAQ Experiments (Pregnancy zone protein precursor peptide VVSVDENFRPR)

iTRAQ vs. DIGE

iTRAQ	**DIGE**
Compares up to 8 samples in one exp.	Compares up to 3 samples on one gel
Compare multiple experiments by including same standard sample in all experiments	Compare multiple experiments by including same standard sample in all experiments
Labels N-terminus + primary amines	Labels N-terminus + primary amines
Complex samples (can fractionate with SCX chromatography after labeling)	Complex samples (can fractionate before labeling)
Quantify from fragmenting peptides (MS/MS) with protein identification	Quantify from intact proteins (Different fluorescence emission)
PTM analysis only if detect modified peptide	PTM analysis only if it changes the protein's pI or MW

Best Approach?

There is NO one best approach

Gel based techniques - Total Protein (isoforms, PTMs)

MS based techniques - Peptides (membrane proteins, large or small proteins)

All approaches:
Complementary
Technically challenging
Require fractionation to dig deeper into proteome

Quantitative Proteomics Experiments

Projects should meet the following requirements:

- Defined question, hypothesis driven
- Defined system
- Independently measurable phenotype
- Reproducible sample preparation
- Samples prepared to minimize protein degradation or contamination
- Buffers must be mass spectrometry or 2D Gel compatible or removed by buffer exchange

Additional Information

How to prepare your samples for DIGE?

Decide on what you want to compare (consult bioinformatist).

Demonstrate reproducibility of sample preparation by analyzing three separate preparation on an SDS-PAGE gel.

Need 15-500 µg proteins at 1-5 ug/ul.

No thiols (like DTT, β-mercaptoethanol).

No primary amines (including ammonium acetate or AEBSF, but TRIS buffers ok).

No salts (<10 mM), HEPES or PPA buffers

Detergents: 0.1% SDS or 0.5% Triton X-100 ok, but not NP-40.

Denaturants: urea and guanidine HCl ok.

TCA/acetone precipitation to remove interfering substances and concentrate sample. Immediately resuspend in 20ul of 7M urea, 2M thiourea, 4% CHAPS, 30mM Tris-HCl, pH 8.8. (Urea containing buffers must NEVER be heated above room temperature)

How to prepare your samples for iTRAQ?

Set up a meeting to discuss your project (410-614-6968).

Demonstrate reproducibility of sample preparation by analyzing three separate preparation on an SDS-PAGE gel.

Need 10-100 µg proteins at 5 ug/ul.

No thiols (like DTT, β-mercaptoethanol).

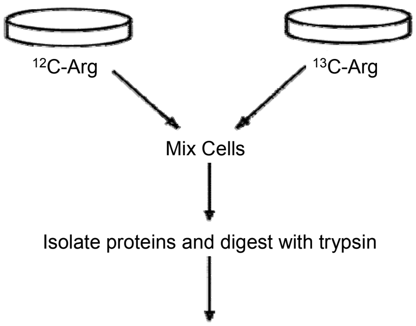
No primary amines (including ammonium acetate, TRIS buffers, AEBSF, etc).

Detergents: 0.1% SDS or 0.5% Triton X-100 ok.

Denaturants: 1M urea or 1M guanidine HCl ok.

TCA/acetone precipitation to remove interfering substances and concentrate sample. Immediately resuspend in 20ul of 0.5M TEAB (triethylammonium bicarbonate)

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



^{12}C-Arg

^{13}C-Arg

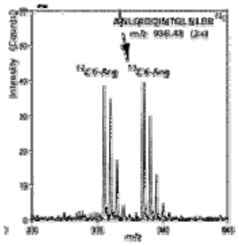
Mix Cells

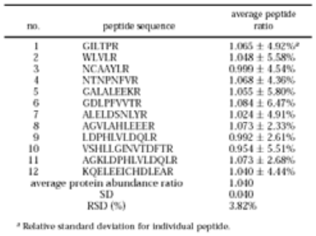
Isolate proteins and digest with trypsin

Peptide ID and quantify by tandem MS

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

Quantify by ^{13}C to ^{12}C-Arg-labeled peptide ratios





no.	peptide sequence	average peptide ratio
1	GILTPR	1.065 ± 4.92%^{a}
2	WLVLR	1.048 ± 5.58%
3	NCAAYLR	0.999 ± 4.54%
4	NTNPNFVR	1.068 ± 4.36%
5	GALALEEKR	1.055 ± 5.80%
6	GDLPFVVTR	1.084 ± 6.47%
7	ALELDSNLYR	1.024 ± 4.91%
8	AGVLAHLEEER	1.073 ± 2.33%
9	LDPHLVLDQLR	0.992 ± 2.61%
10	VSHLLGINVTDFTR	0.954 ± 5.51%
11	AGKLDPHLVLDQLR	1.073 ± 2.68%
12	KQELEEICHDLEAR	1.040 ± 4.44%
	average protein abundance ratio	1.040
	SD	0.040
	RSD (%)	3.82%

^{a} Relative standard deviation for individual peptide.

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

Strengths/Limitations of SILAC

Strengths:

- No extra steps
- Relative quantification using multiple peptides from same protein
- Post-translational modification information
- Mass increase of up to 9 amu with ^{13}C and ^{15}N Arg

Limitations:

- No reduction in sample complexity
- Expensive
- Culture only
- Cannot use for tissue or isolated proteins

iTRAQ versus ICAT

iTRAQ

- Compare up to 4 samples
- Labels amines
- Protein ID and quantification based on many peptides
- PTM information
- Peptide based
- Quantify from MS/MS
- Sample complexity not reduced
- LC compatible

ICAT

- Compare up to 2 samples
- Labels cysteines
- Protein ID and quantification based on few peptides
- No PTM information
- Protein based
- Quantify from MS
- Sample complexity reduced
- Gels and LC compatible

iTRAQ vs. ^{18}O-Labeling

iTRAQ	**^{18}O-Labeling**
Compares up to 4 samples	Compares 2 samples
Labels N-terminus + primary amines	Labels C-terminus of peptides
Complex samples (can fractionate with SCX chromatography after labeling)	Sample complexity limited (due to ^{18}O exchange with water over time)
Quantify from fragmenting peptides (MS/MS)	Quantify from intact peptides (MS)
Adds one step to protein identification	No steps added to protein identification
Analysis Software available	No Analysis Software available

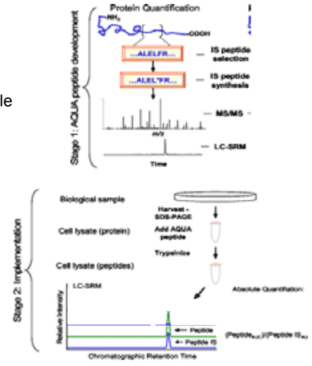
Absolute Quantification (AQUA) Scheme

Select a peptide from the protein.

Synthesize of this peptide using stable isotopes (e.g., ^{13}C, ^{15}N, etc.) at a single amino acid.

Spike protein sample with synthetic peptide.

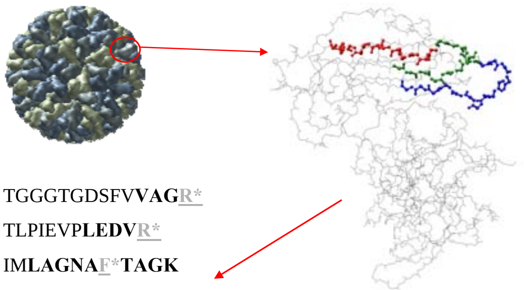
Compare abundance of synthetic peptide with the native peptide.



Gerber et al. PNAS 100: 6940-6945, 2003

In silico Analyses for Selecting Peptides

Norovirus Coat Protein (VLP)



TGGGTGDSFVVAGR*

TLPIEVPLEDVR*

IMLAGNAF*TAGK

Δ = 10 Da

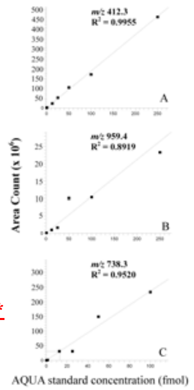
180 copies per virus

Standard Curve for AQUA Peptides

TLPIEVPLEDVR*

IMLAGNAF*TAGK

TGGGTGDSFVVAGR*



R* addes 12 DA

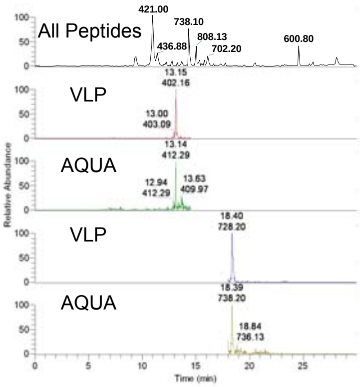
E. coli spiked with VLP and AQUA

15 µg *E. coli*

100 fmol VLPs

20 fmol AQUA peptides

Calculated recovery:
79 ± 14.6 %



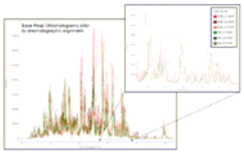
Mass Spectrometry Based Techniques

No Labeling

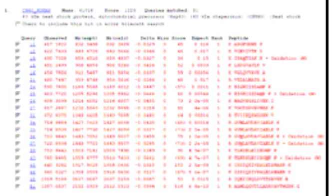
Spectral Counting

Each sample separate LCMS/MS experiment

Each protein in an LCMS/MS experiment,
Count number of times a peptide is detected
Count number of peptides from protein

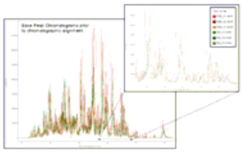


Compares multiple LCMS/MS experiments

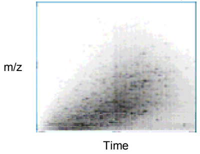


Peak Intensity

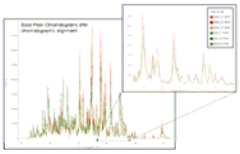
1-Overlay Base Peak Chromatograms



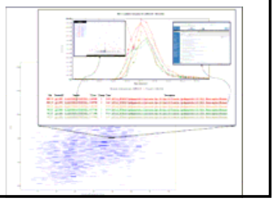
3-Construct 3D Peak Intensity Map



2-Align Chromatograms



4-Annotation of “Spots”



Strengths/Limitations of Non-Labeling Methods

Strengths:	Limitations:
• No limit on number of samples	• No limit on number of samples
• No added steps in protein identification	• Instrument intensive
	• Reproducibility over many experiments LC system Ionization Detection