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

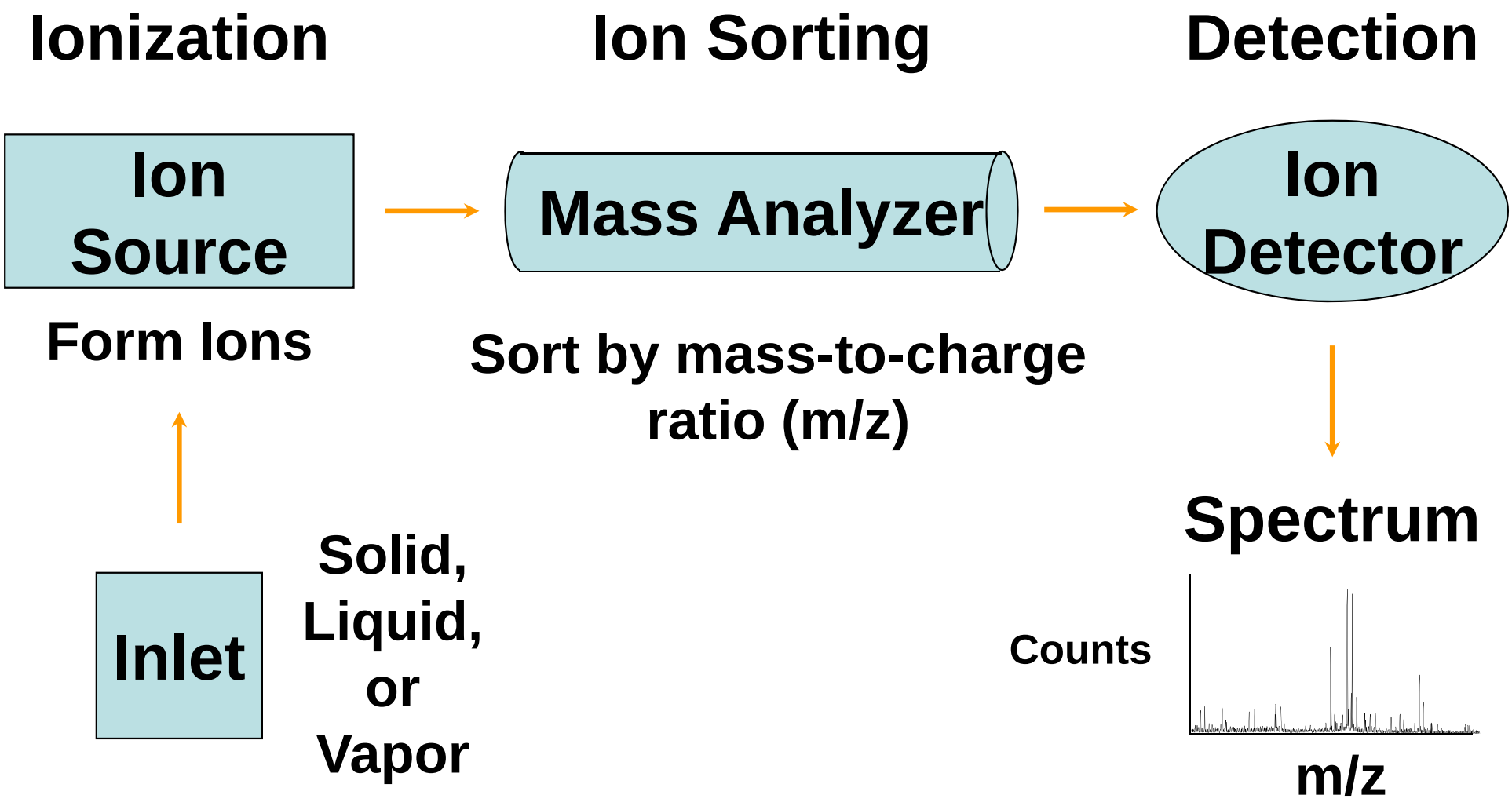
Topics

Protein identification by mass spectrometry

Relative quantification of proteins

Applications

How a Mass Spectrometers Measure Mass



Identifying Proteins by Mass Spectrometry

Three Methods

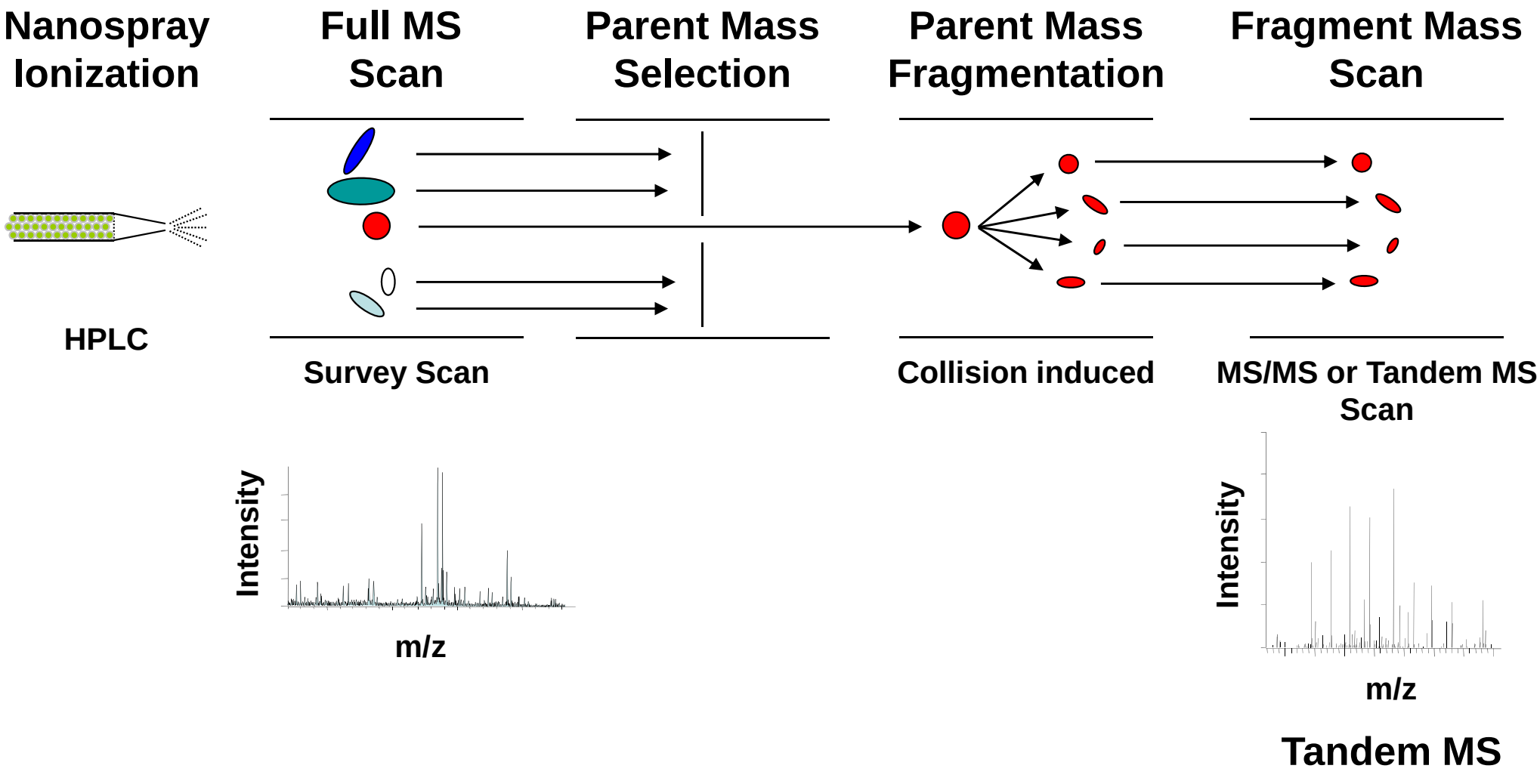
Bottom Up: Proteins identified from **a few peptides**
Know cleavage site, peptide mass and sequence
Do not know proteins actual size and modifications
Problems: Sequence homology, seeing all peptides

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

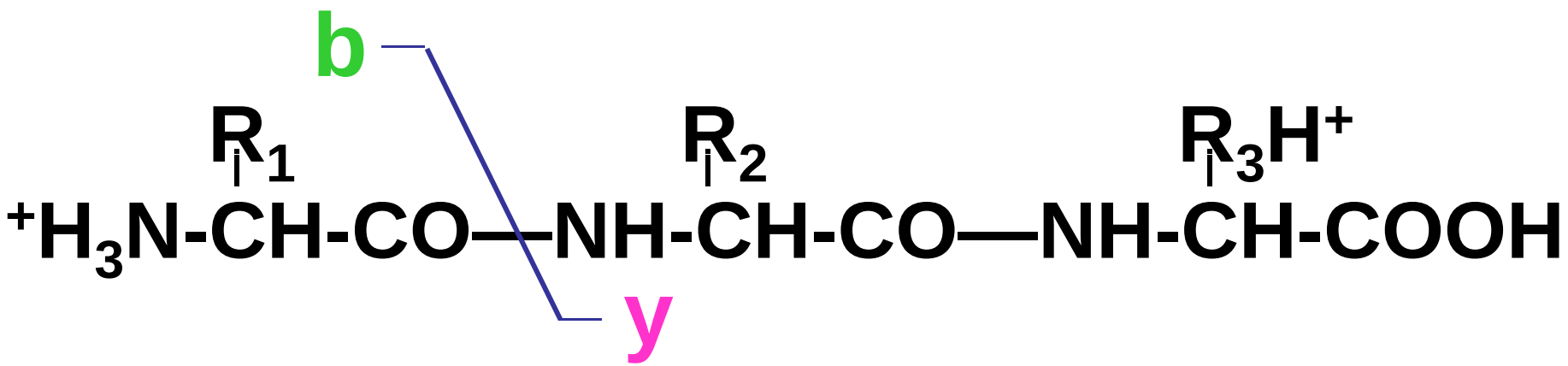
Middle Down: Proteins identified from **large sections of protein**

Identifying Proteins from Peptide Sequence

(Many proteins in solution, gel bands or spots)

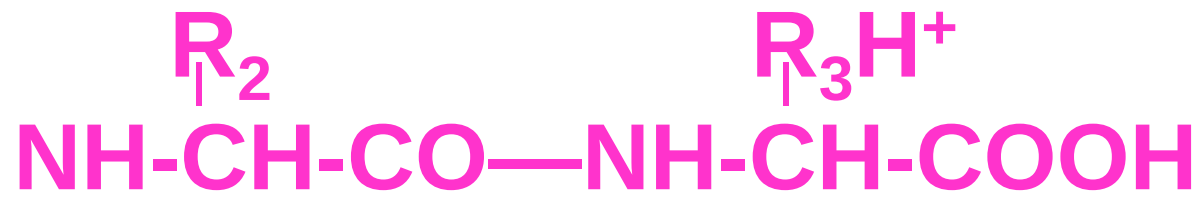


Sequencing Peptides by Fragmentation

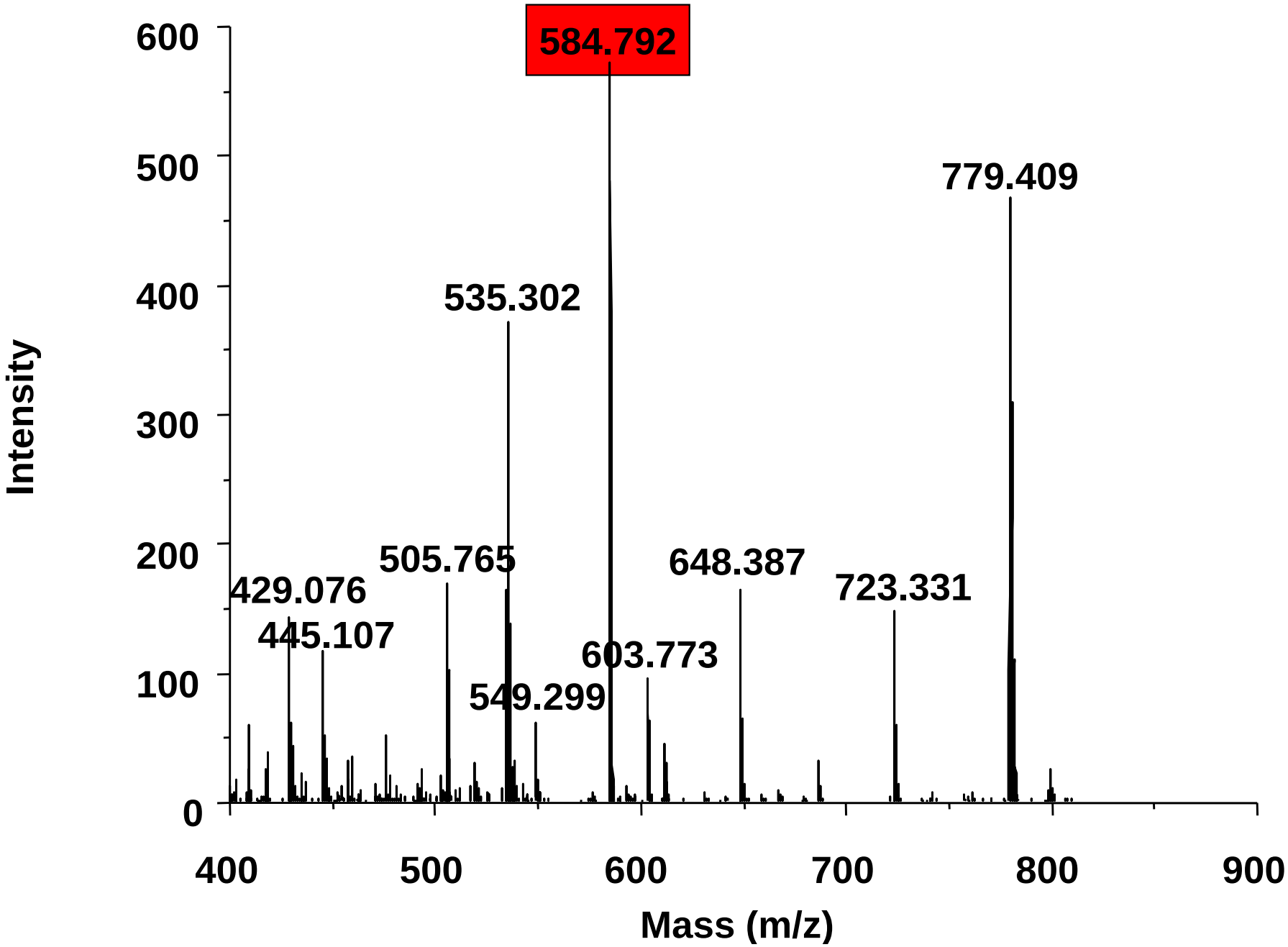


b-ion (b1)

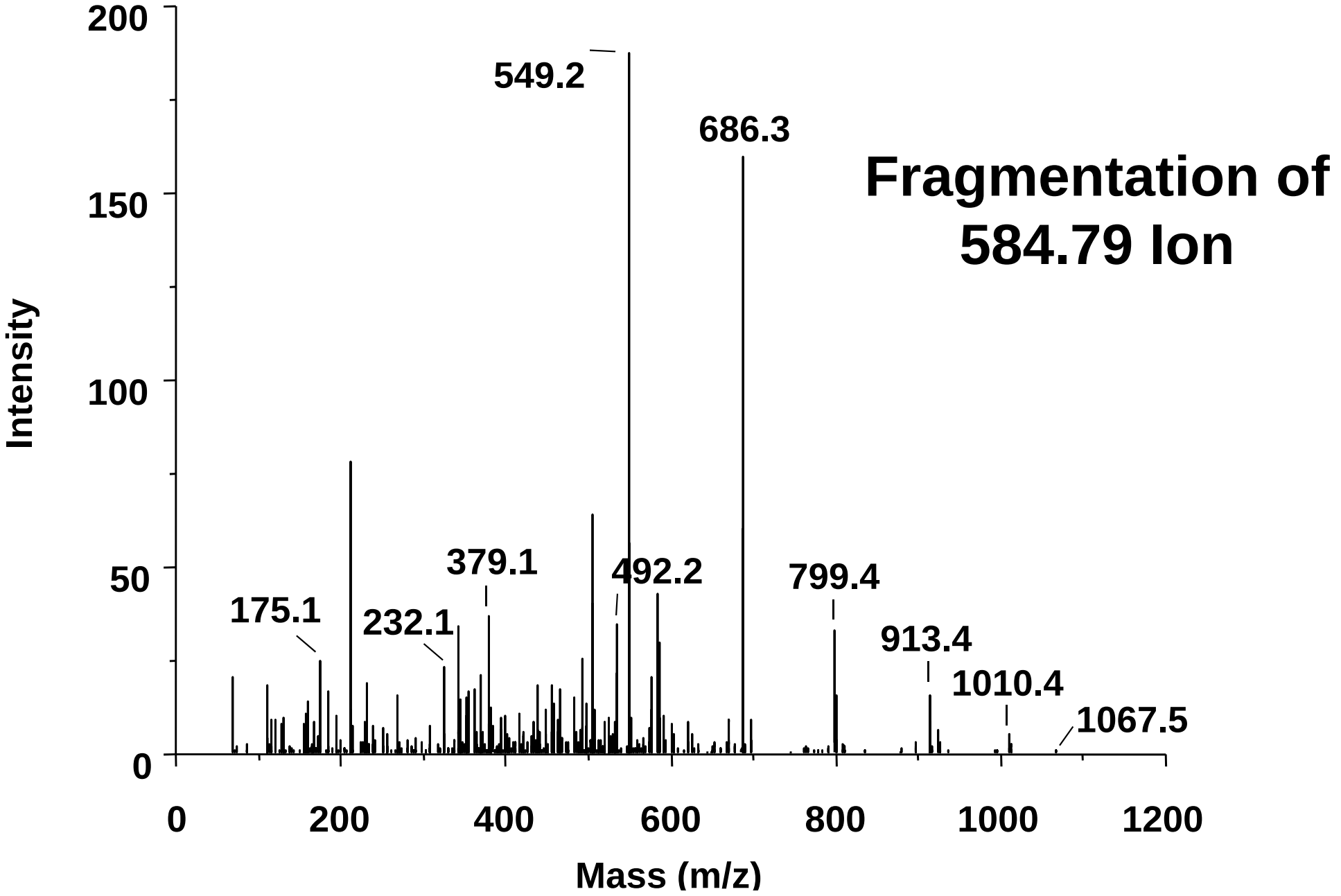
y-ion (y2)

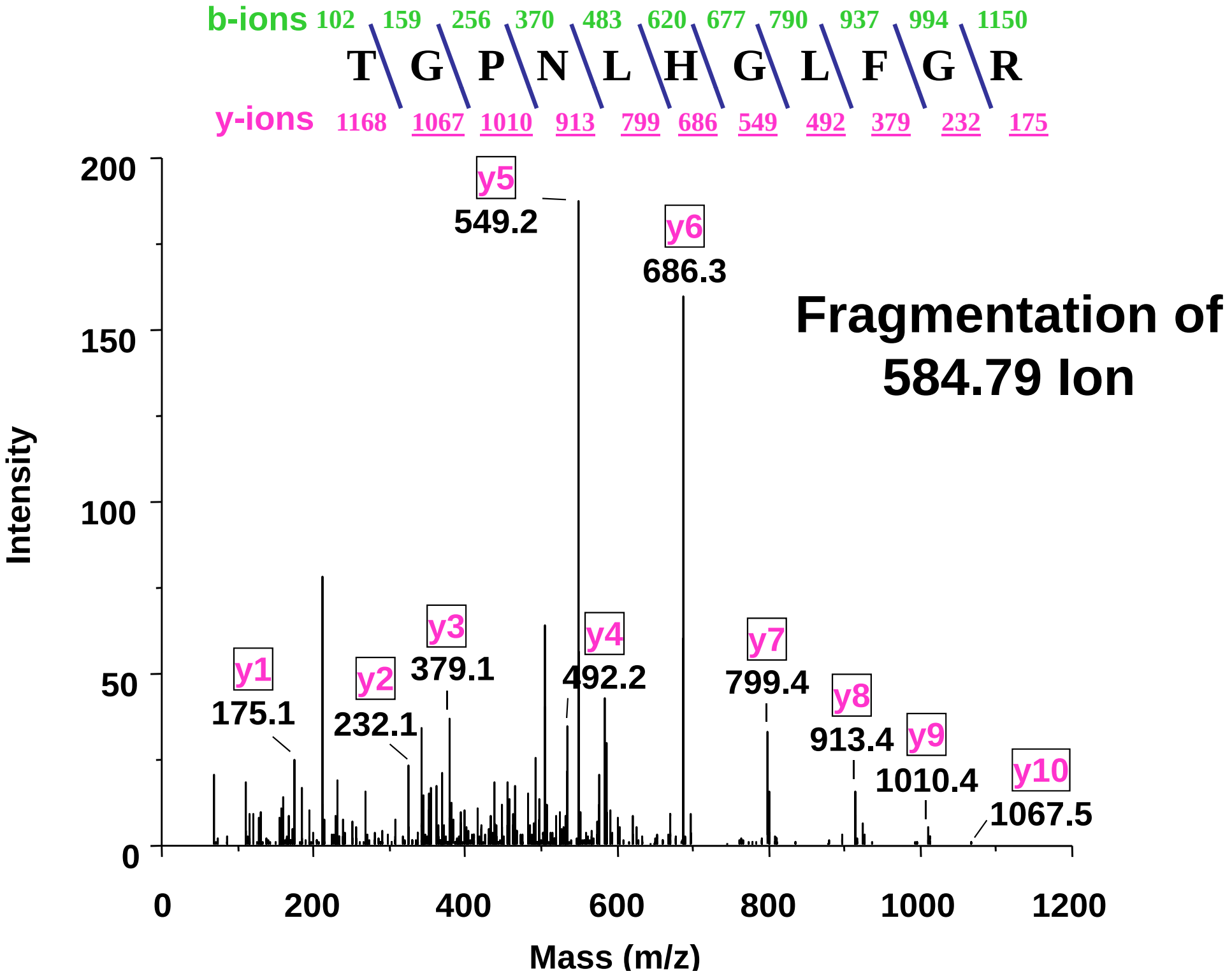


Survey (Full MS or MS1) Scan of Peptides



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




b-ions 102 159 256 370 483 620 677 790 937 994 1150
T G P N L H G L F G R
y-ions 1168 1067 1010 913 799 686 549 492 379 232 175
Fragmentation of 584.79 Ion
y5 549.2
y6 686.3
y1 175.1
y2 232.1
y3 379.1
y4 492.2
y7 799.4
y8 913.4
y9 1010.4
y10 1067.5
Intensity
Mass (m/z)
0
50
100
150
200
0
200
400
600
800
1000
1200

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
Prowl	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

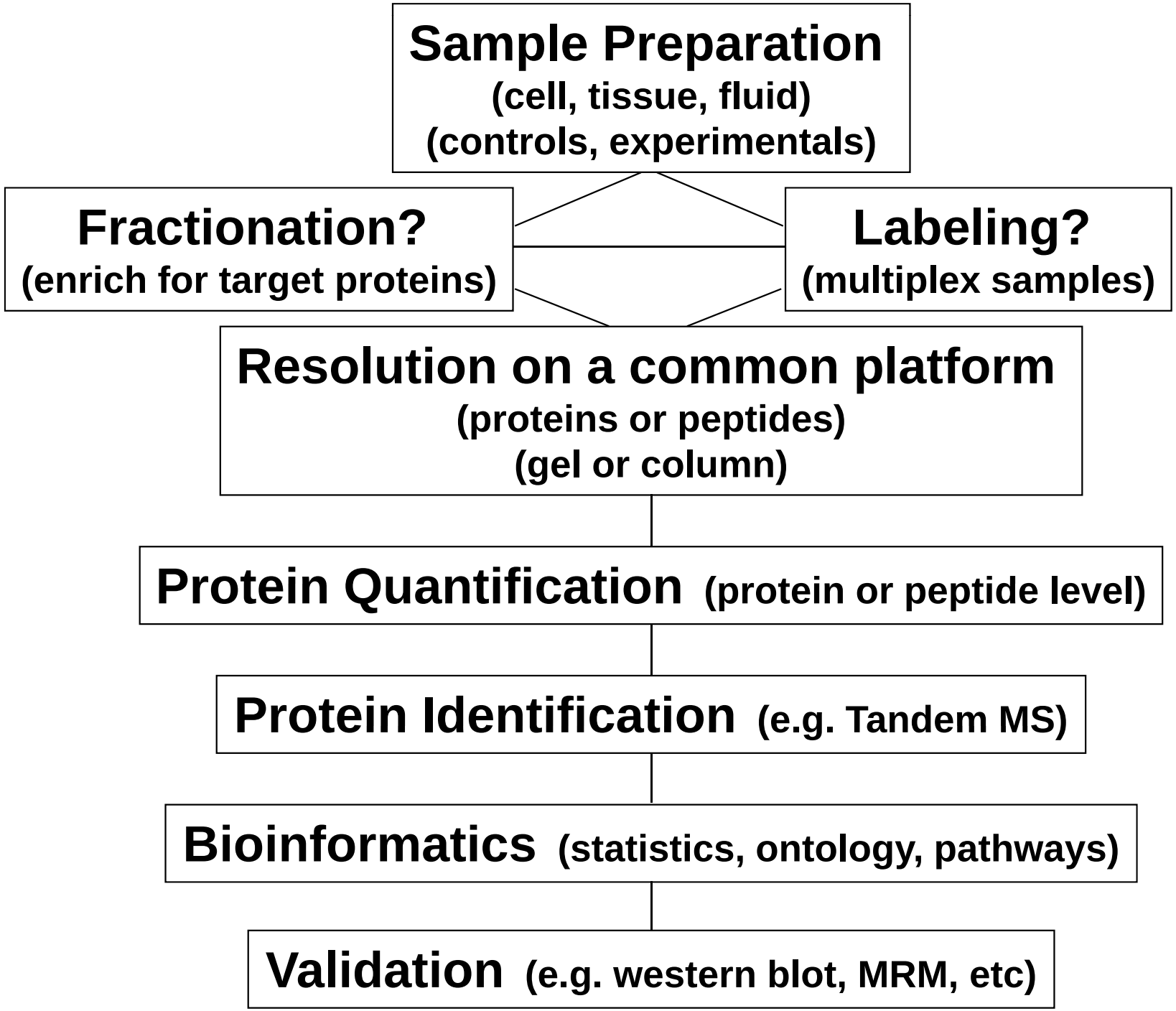
Defined question (i.e. hypothesis driven)

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

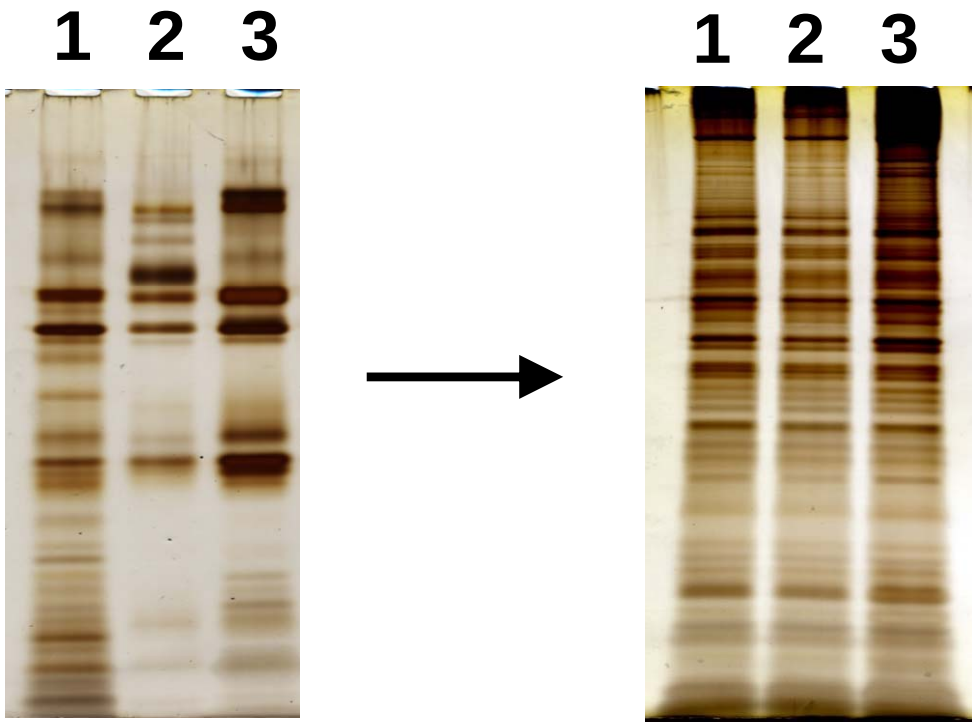
Buffer Composition Compatibility

Sample Preparation

Must Be Reproducible and Standardized!

Biological Replicates

Initial Protocol



Standardized Protocol

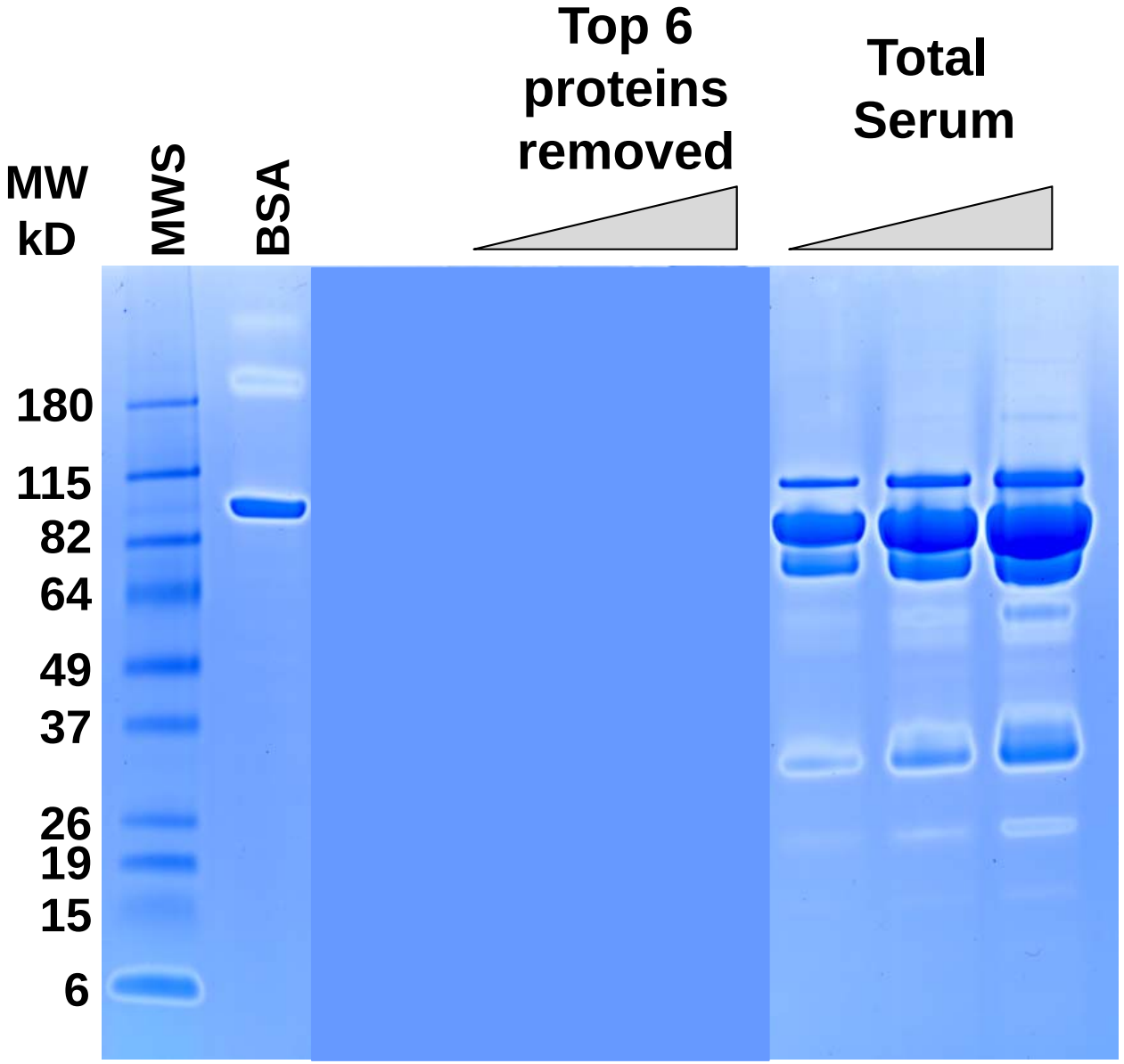
Reduce Sources of Variability:

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

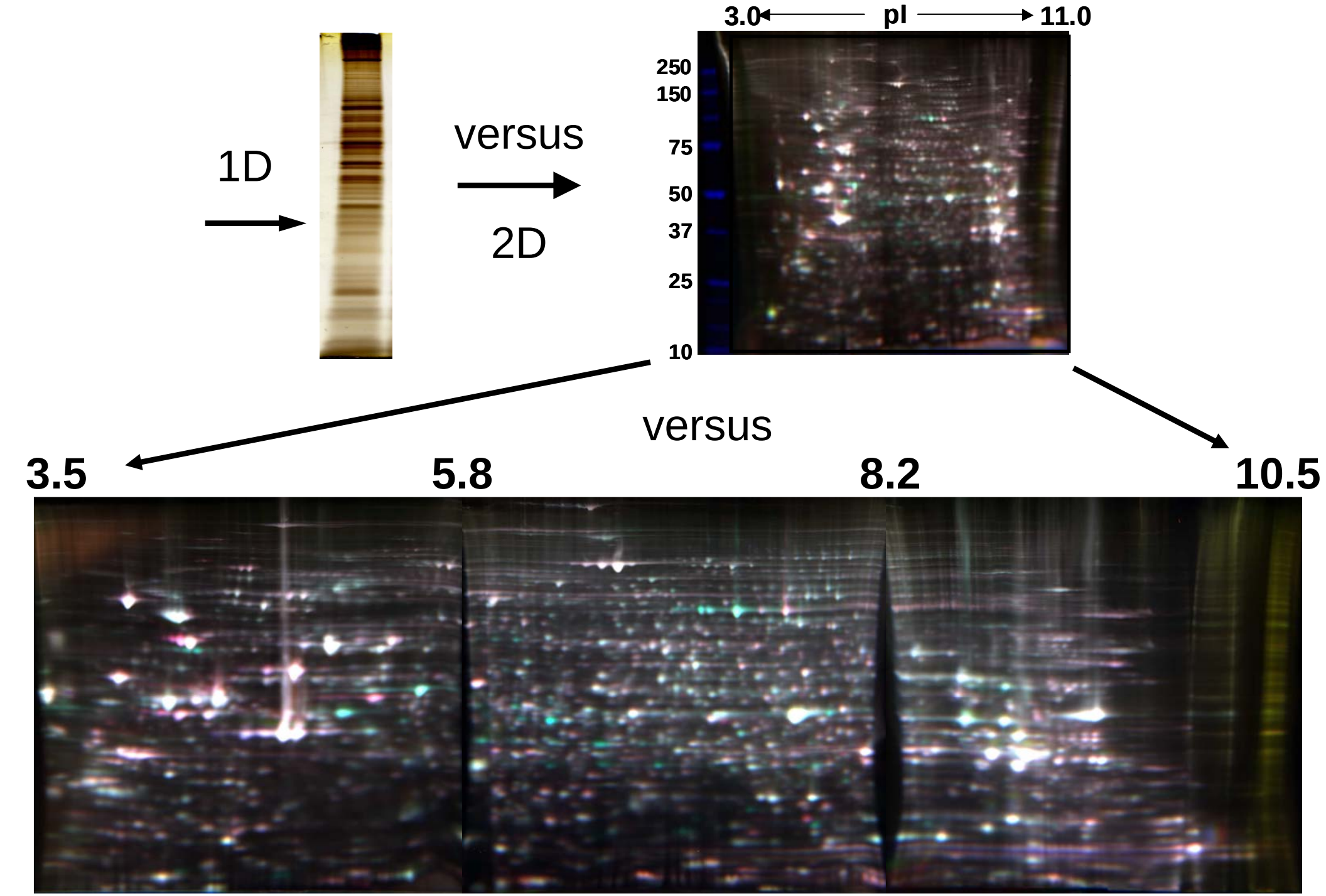
Biological: cells < tissue < body fluids
yeast < nematode < human

Protein Amount and Complexity

Detect Low Abundance Proteins by Fractionating



4-12% SDS-PAGE, SimplyBlue stain


1D
versus
2D
3.0
pI
11.0
250
150
75
50
37
25
10
versus
3.5
5.8
8.2
10.5

Quantitative Proteomics Methods

Non-Labeling Methods

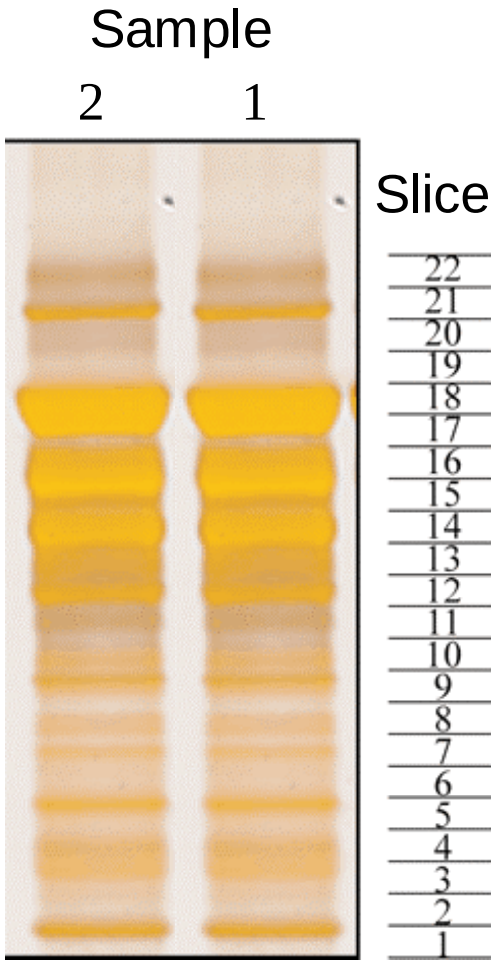
No additions to analysis

Separate analysis of *Each* 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

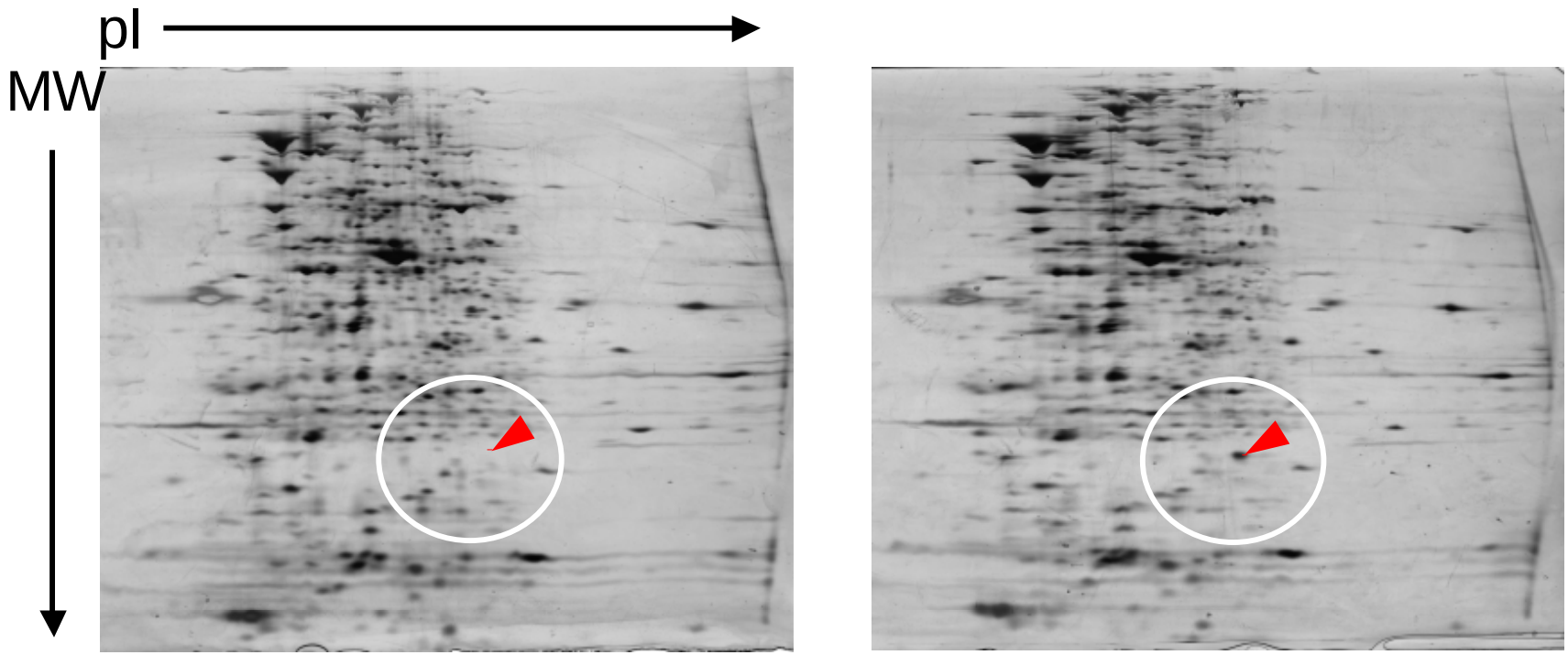
Comparing protein bands on same gel
Quantify by densitometry
Often more than one protein per band

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

No Labeling – Densitometry - 2D Gels

Bacterial Strain #1

Bacterial Strain #2

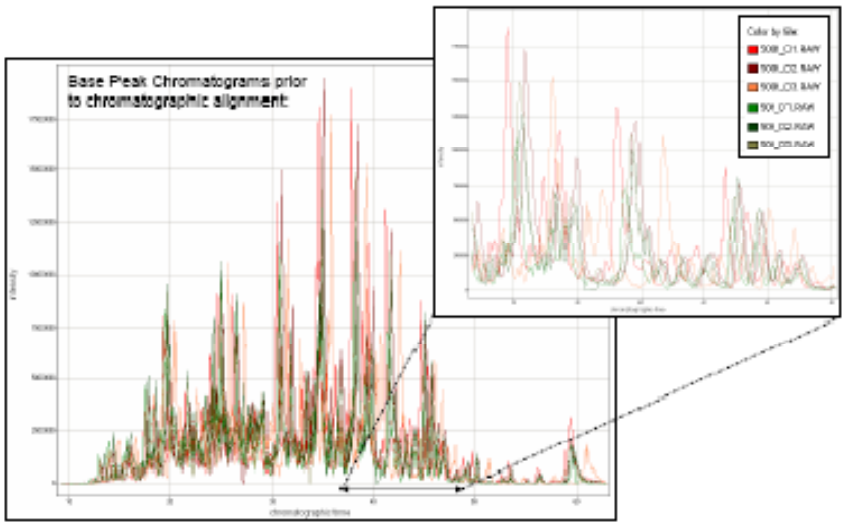


Compare spots from different gels
Quantify by relative spot volume
Gel reproducibility and spot matching critical
Gel warping for matching spots can warp out modification
Often more than one protein per spot

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

1. CH60_HUMAN Mass: 61016 Score: 1225 Queries matched: 31
60 kDa heat shock protein, mitochondrial precursor (Hsp60) (60 kDa chaperonin) (CPN60) (Heat shock

Check to include this hit in error tolerant search

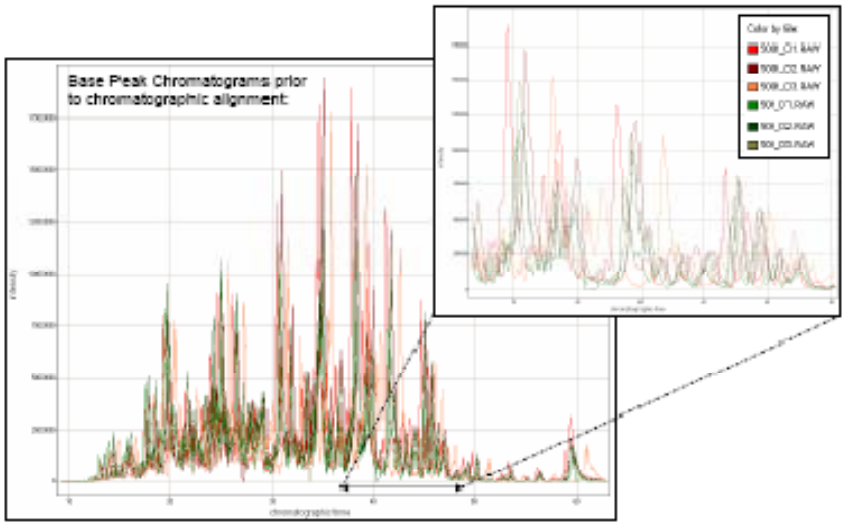
Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
11	417.1822	832.3498	832.3828	-0.0329	0	45	0.016	1	K.APGFGDNR.K
12	422.7433	843.4720	843.5066	-0.0346	0	46	0.017	1	K.VGEVIVTK.D
13	430.7328	859.4510	859.4837	-0.0327	0	36	0.15	1	K.IPAMTIAK.N + Oxidation (M)
15	451.2499	900.4853	900.5280	-0.0428	0	52	0.0039	1	K.LSDGVAVLK.V
16	456.7806	911.5467	911.5804	-0.0337	0	59	0.00056	1	K.VGLQVVAVK.A
21	480.7447	959.4748	959.5036	-0.0288	0	45	0.017	1	R.VTDALNATR.A
24	595.7855	1189.5565	1189.6012	-0.0447	0	(57)	0.0011	1	K.EIGNIISDAMK.K
25	603.7720	1205.5294	1205.5962	-0.0668	0	60	0.00048	1	K.EIGNIISDAMK.K + Oxidation (M)
26	608.3099	1214.6052	1214.6507	-0.0455	0	73	2.2e-05	1	K.NAGVEGSLIVEK.I
27	617.2857	1232.5569	1232.5885	-0.0316	0	81	4e-06	1	K.VGGTSDVEVNEK.K
31	672.8375	1343.6605	1343.7085	-0.0480	0	64	0.00016	1	R.TVIIEQSWGSPK.V
34	714.8884	1427.7623	1427.8058	-0.0435	0	(65)	0.00014	1	R.GVMLAVDAVIAELK.K
35	714.8938	1427.7730	1427.8058	-0.0327	0	(73)	2.1e-05	1	R.GVMLAVDAVIAELK.K
36	722.8849	1443.7552	1443.8007	-0.0455	0	75	1.2e-05	1	R.GVMLAVDAVIAELK.K + Oxidation (M)
37	722.8934	1443.7722	1443.8007	-0.0285	0	(73)	2.2e-05	1	R.GVMLAVDAVIAELK.K + Oxidation (M)
39	752.8643	1503.7141	1503.7490	-0.0349	0	90	4.3e-07	1	K.TLNDELEIIEGMK.F
40	760.8461	1519.6777	1519.7439	-0.0662	0	(89)	4.7e-07	1	K.TLNDELEIIEGMK.F + Oxidation (M)
45	640.3281	1917.9625	1918.0636	-0.1010	0	102	2.1e-08	1	K.ISSIQSIVPALEIANAHR.K
46	960.0327	1918.0509	1918.0636	-0.0127	0	(87)	5.1e-07	1	K.ISSIQSIVPALEIANAHR.K
48	1019.5106	2037.0067	2037.0153	-0.0087	0	52	0.0015	1	R.IQEIIEQLDVTTSEYEK.E
51	1057.0537	2112.0929	2112.1323	-0.0394	0	116	6.8e-10	1	R.ALMLQGVDLLADAVAVTMGPK.G

Reproducibility of LCMS/MS system critical

No Labeling – MS Methods

Peak Intensity

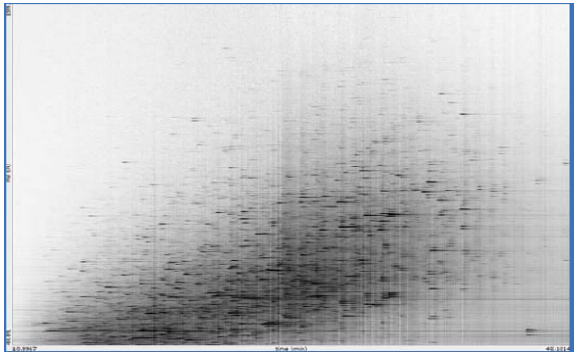
Each sample separate LCMS/MS experiment



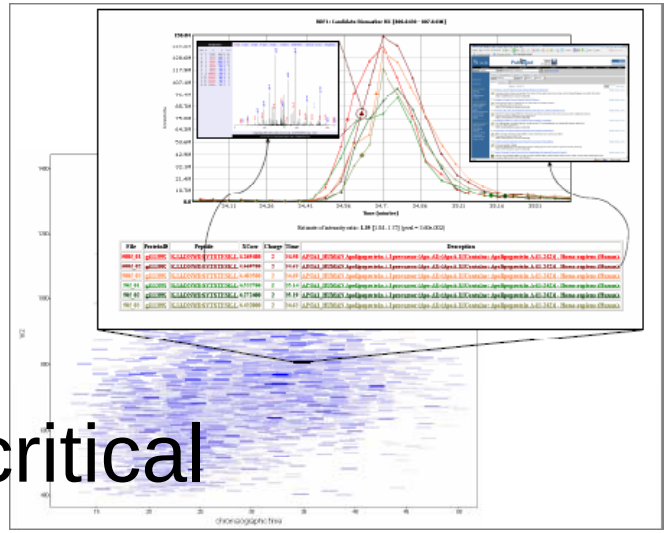
Compares multiple LCMS/MS experiments

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%

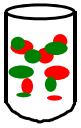
Difference Gel Electrophoresis (DIGE)

Control Protein Extract

Label with fluor 1

Disease Protein Extract

Label with fluor 2



Mix labeled proteins from both samples

Separate proteins on one 2-D gel

Excitation wavelength 1

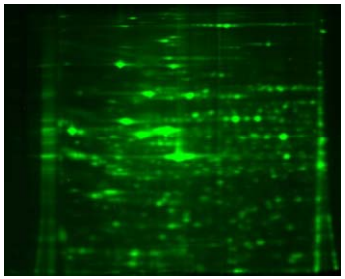
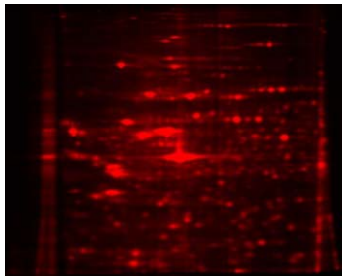
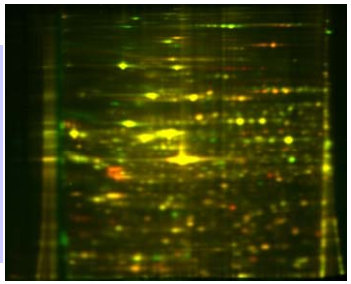


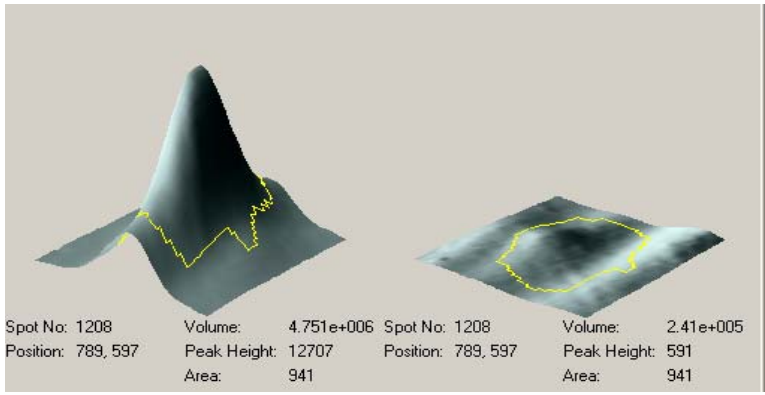
Image gel



Excitation wavelength 2

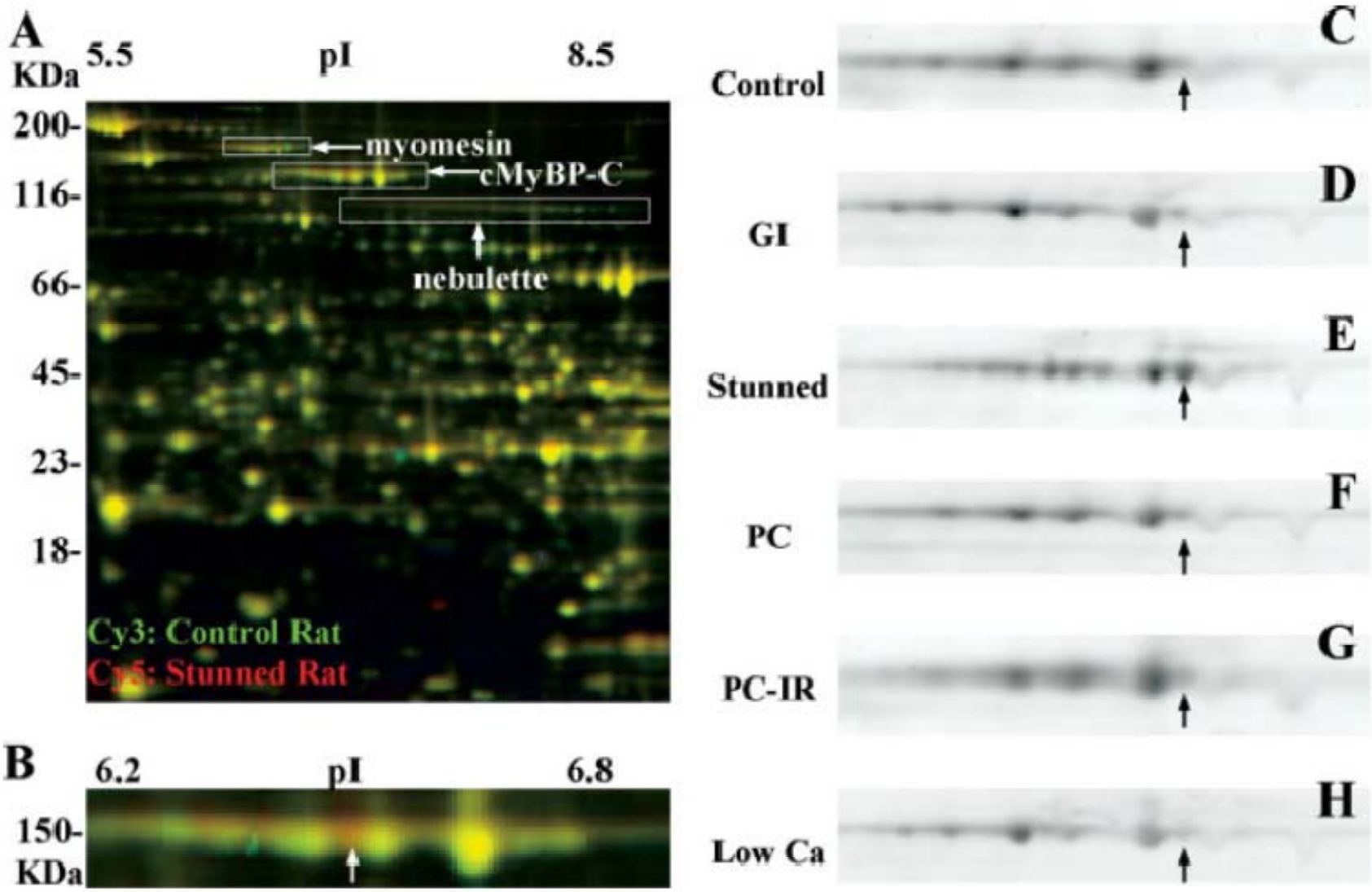


Protein comparison in a single gel!



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

Novel Phosphorylation of Myofilament Protein Correlates with Myocardial Stunning



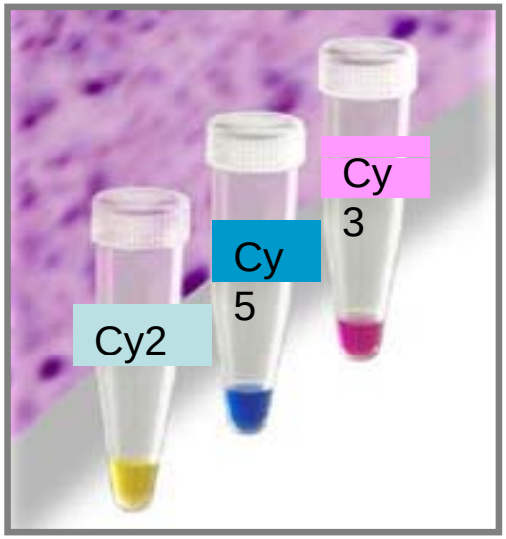
Yuan et al. Proteomics 6:4176-4186, 2006

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.

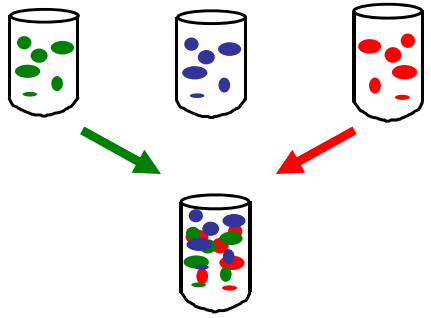
Reduces experimental variations.

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

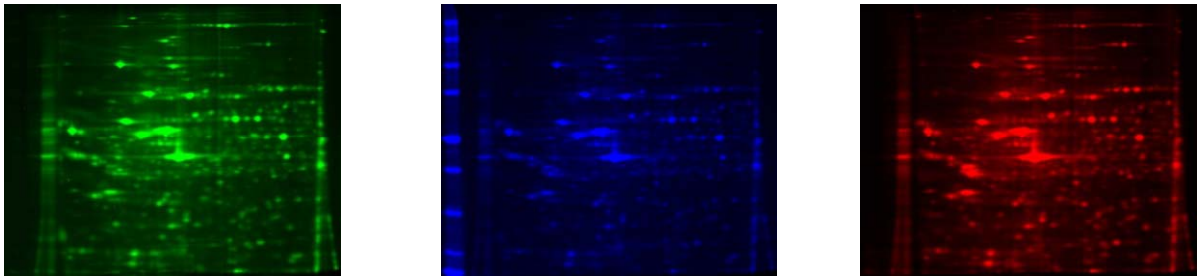


Disease Protein Extract
Label with Cy5

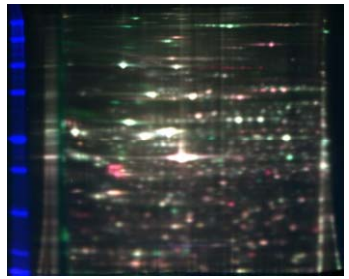
Mix samples

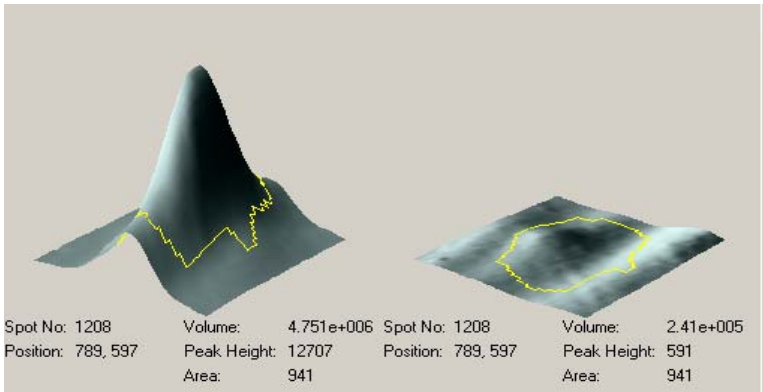
Separate proteins on one 2-D gel

Excitation wavelength 1

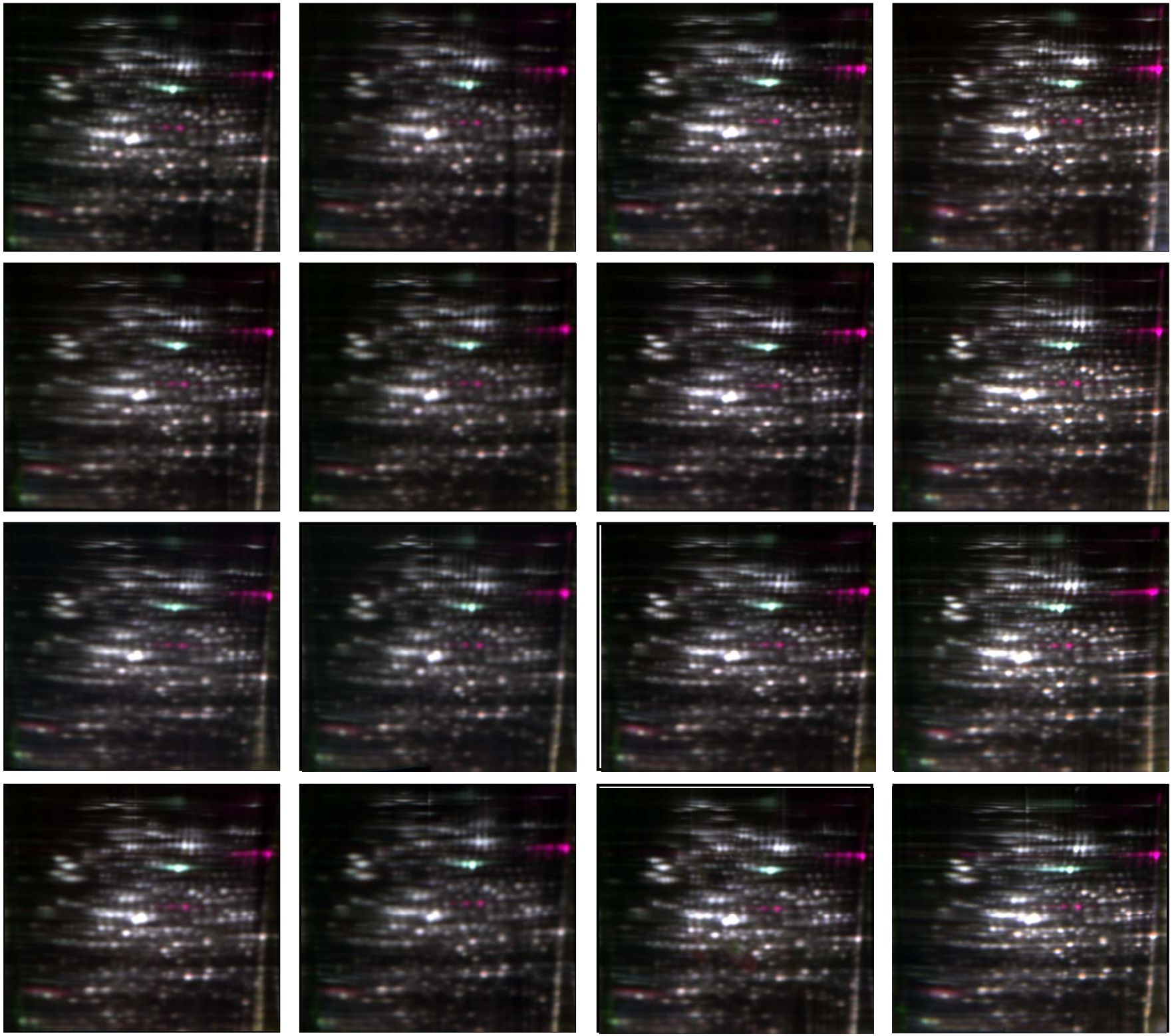


Excitation wavelength 2



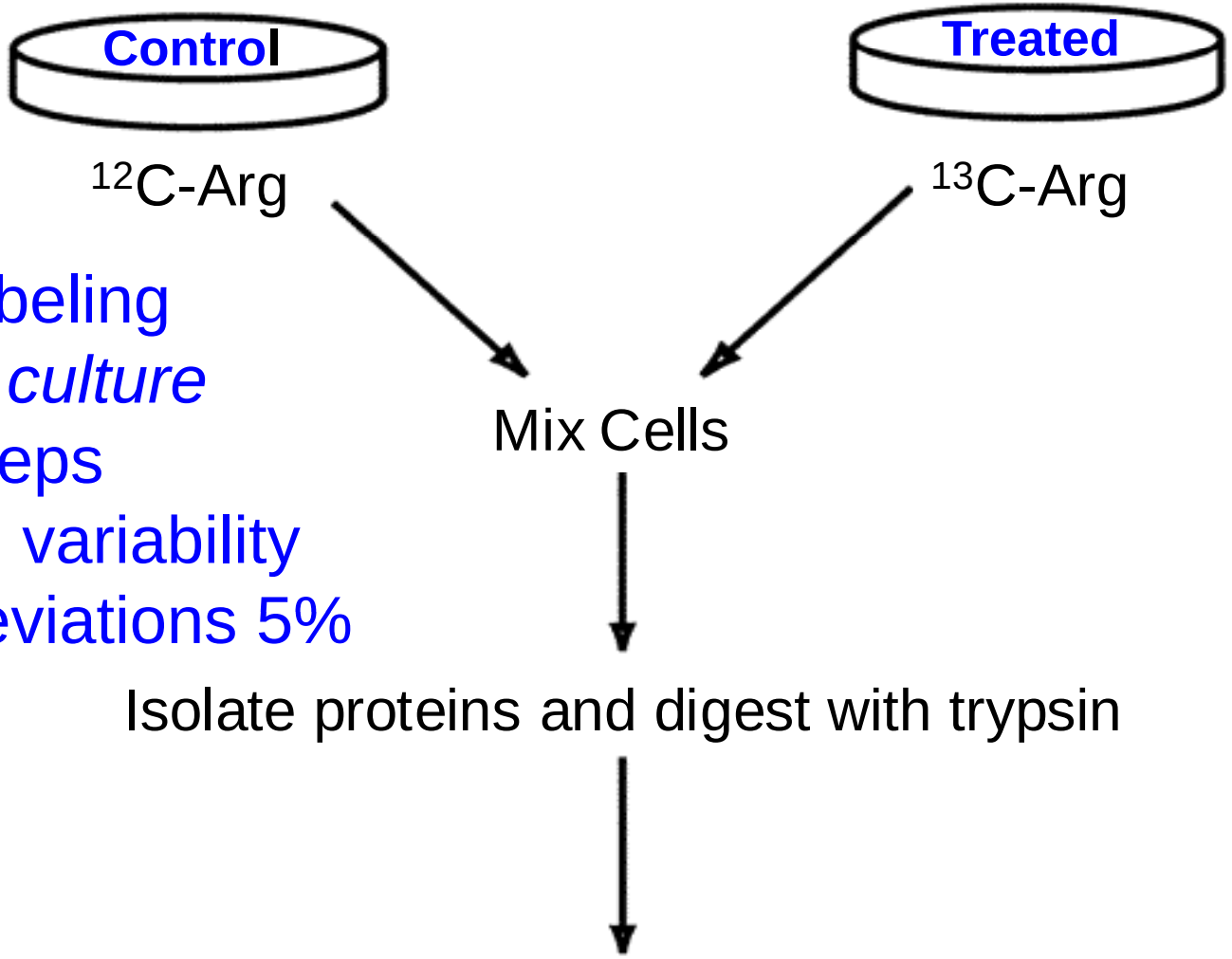


Quantify intact proteins
Cells, tissue, fluids
One additional step
Reduced technical variability



Relative MS Quantitative Methods

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



Metabolic labeling

Only for cell culture

No added steps

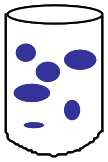
No technical variability

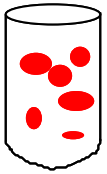
Standard Deviations 5%

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

^{18}O-labeling Protocol

Control
Protein Extract

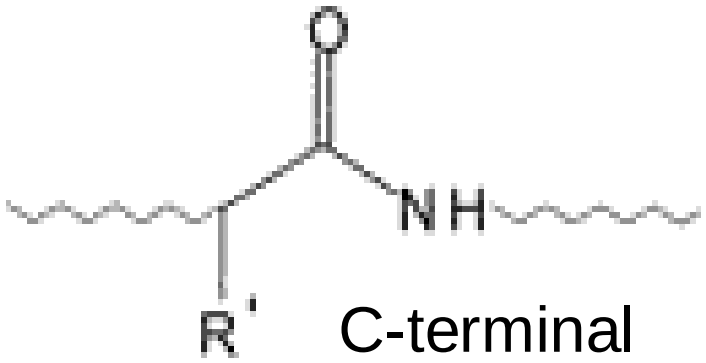




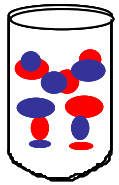
Experimental
Protein Extract

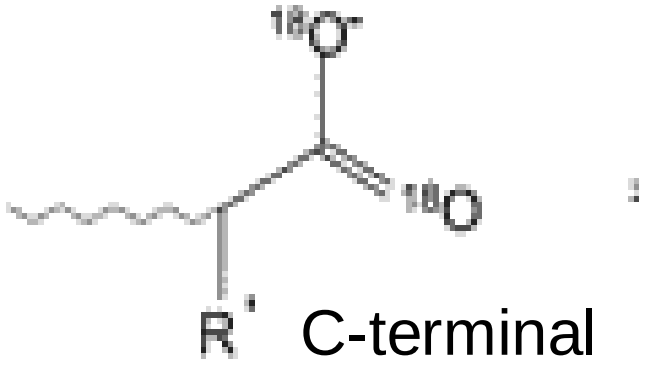
Trypsin

Trypsin + ^{18}O-water



C-terminal





C-terminal
(adds 4 Da)

Enzymatic labeling
Cells, tissue, fluids
No added steps

Mix labeled peptides

Peptide ID and quantify by tandem MS

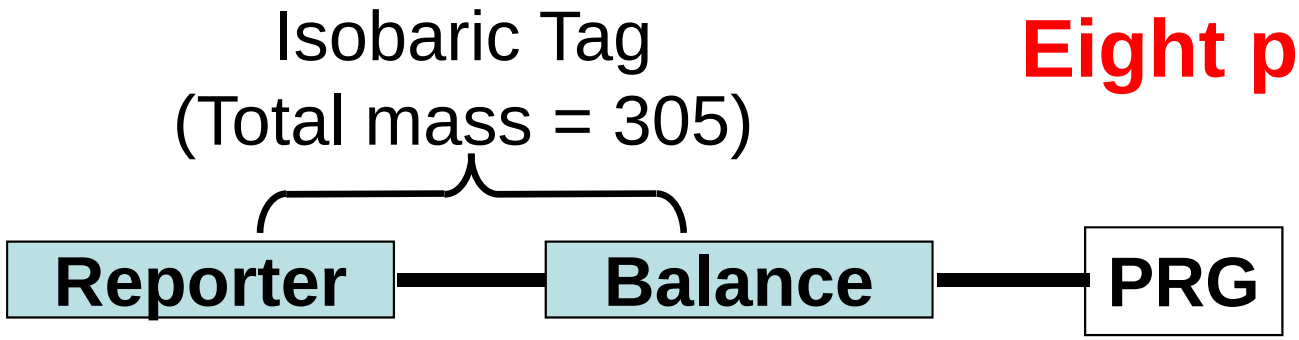
No MS technical variability
Standard deviations 10-20%
Loss of label to water

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)



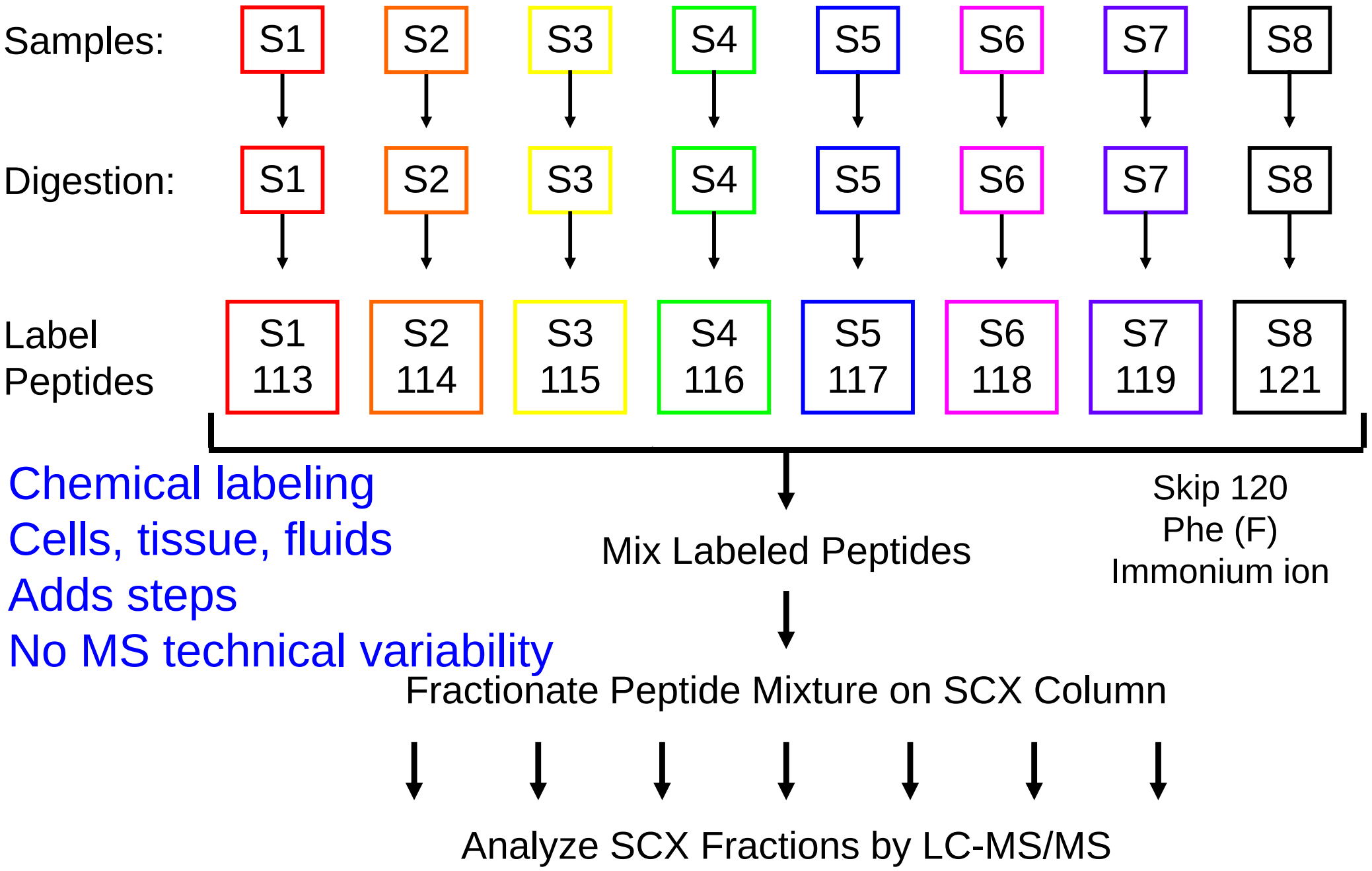
Eight possible tags

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

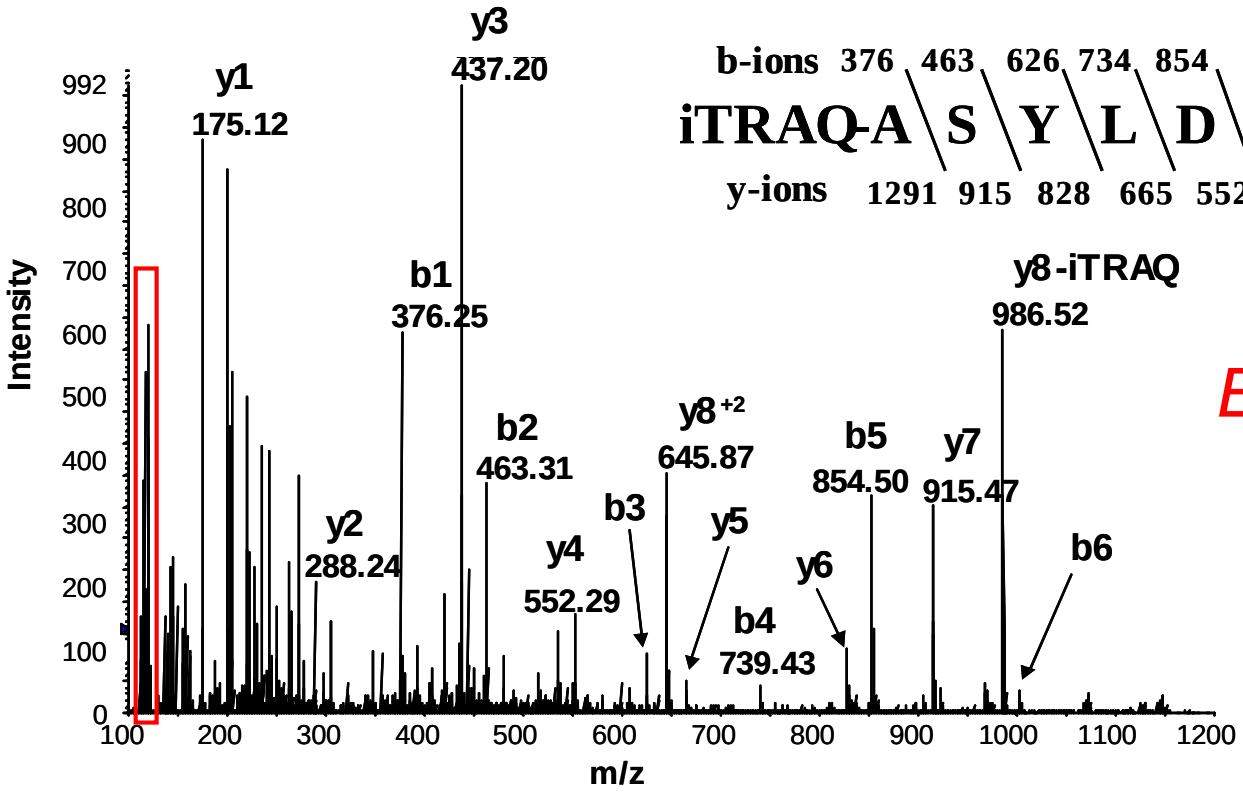
Peptide Reactive Group (binds to amines)

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

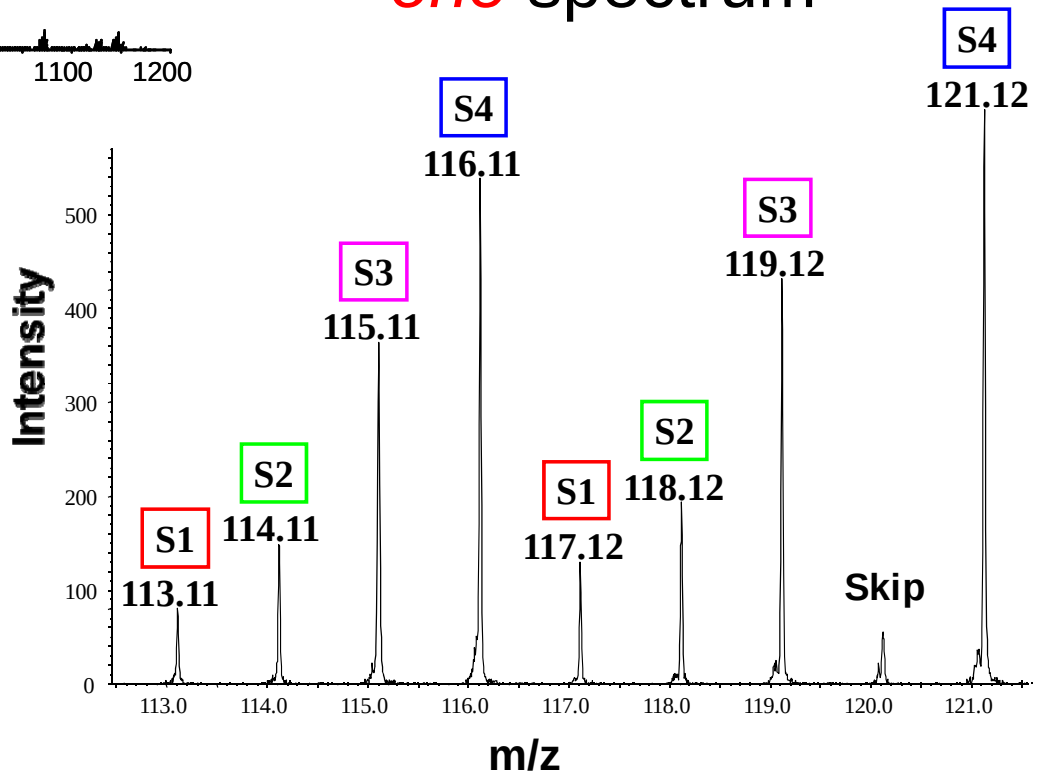
8-plex iTRAQ Workflow



Quantifying Proteins using iTRAQ



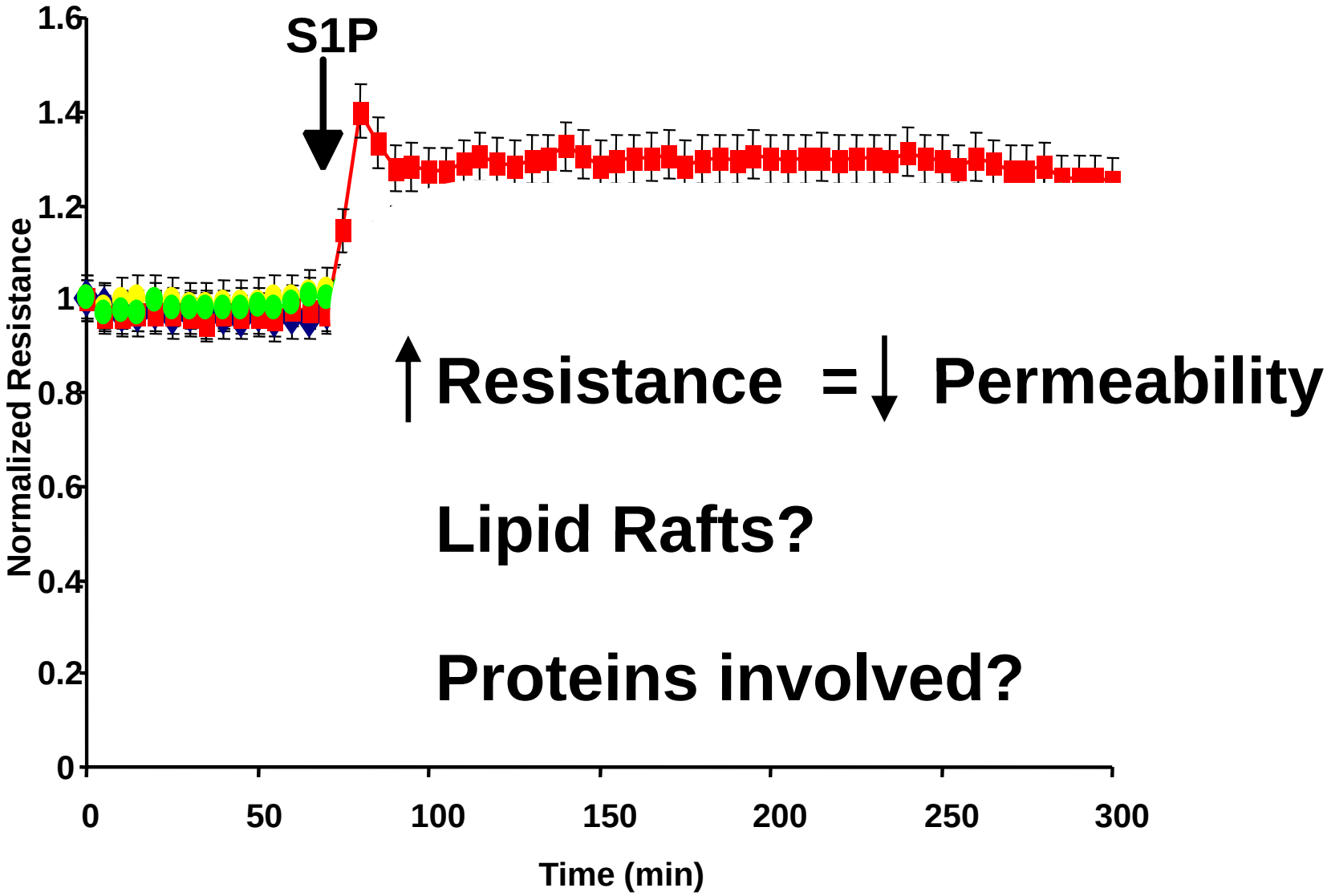
Eight reporter ions quantify *same* peptide from *eight* samples in *one* spectrum



Changes in Protein Levels

(*biomarkers and pathways*)

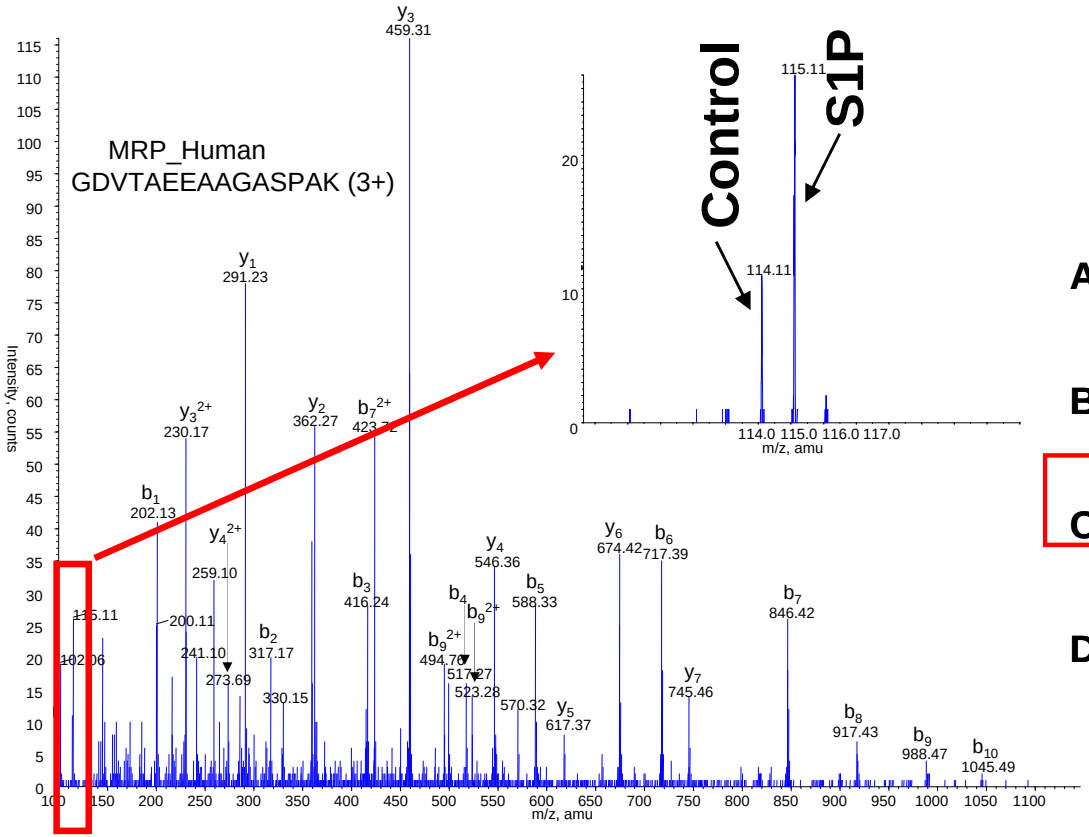
Phospholipid Growth Factor (S1P) on Pulmonary Endothelial Permeability



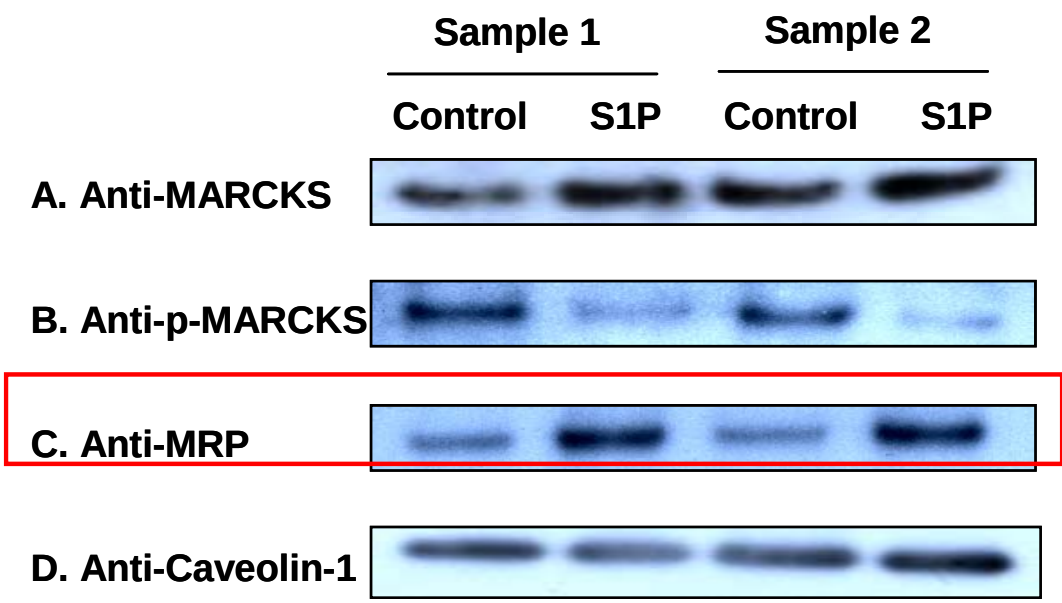
Guo Y et al (2007) Mol Cell Proteomics 6:689-696

iTRAQ Revealed Proteins More Abundant in S1P Stimulated Lipid Rafts

MRP Peptide Fragmentation



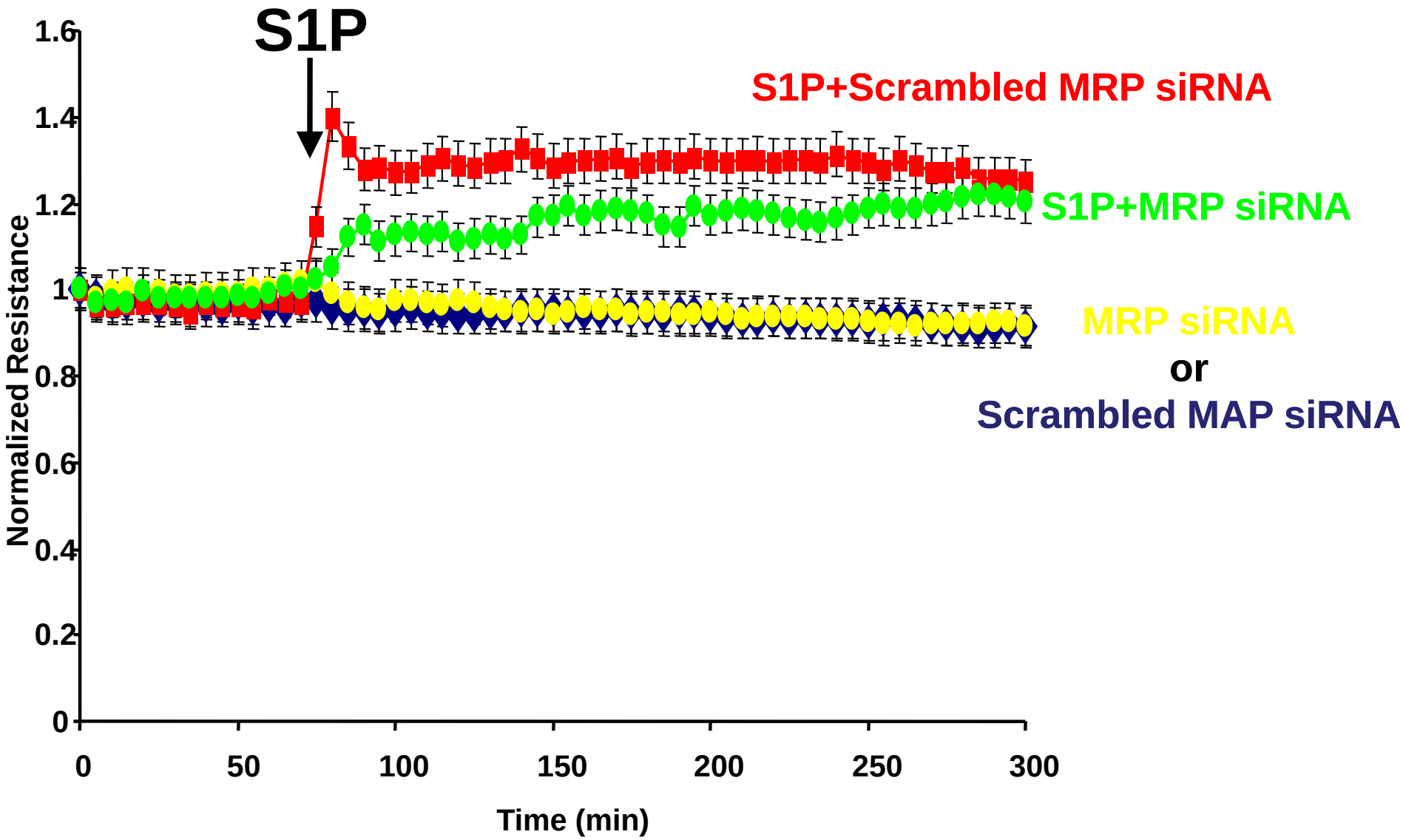
Western Blots Confirm iTRAQ



MRP is up-regulated using both methods

Guo Y et al (2007) Mol Cell Proteomics 6:689-69

MRP siRNA Attenuates S1P Stimulated Endothelial Barrier Enhancement



Guo Y et al (2007) Mol Cell Proteomics 6:689-69

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

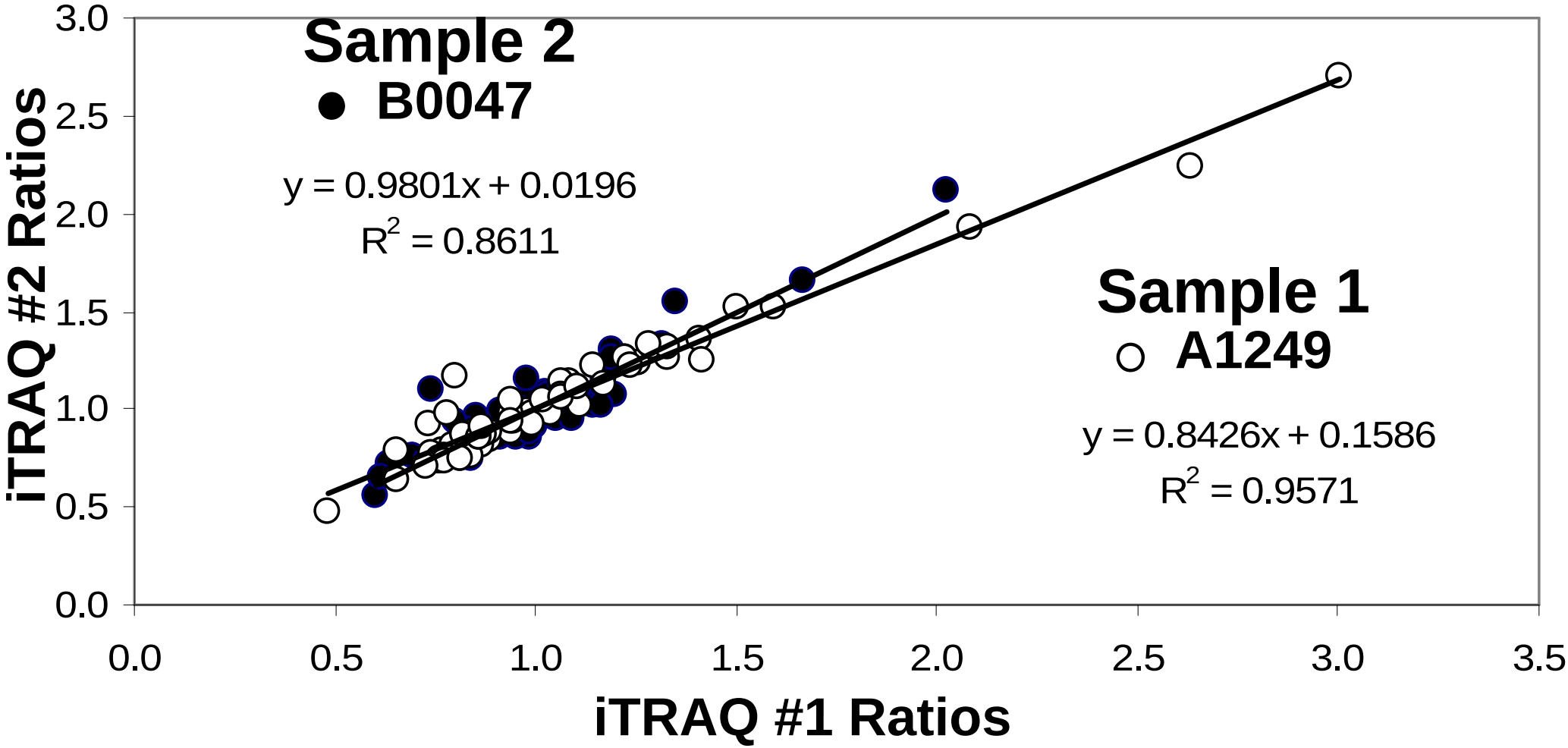
Completely randomize labeling

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

Nutritional Deficiencies

Infant/ Child

- Poor growth
- Impaired development
- Disability
- Infection
- Chronic disease
- Childhood Death

Mother

- Obstetric morbidity
- Infection/sepsis
- Anemia
- Death

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

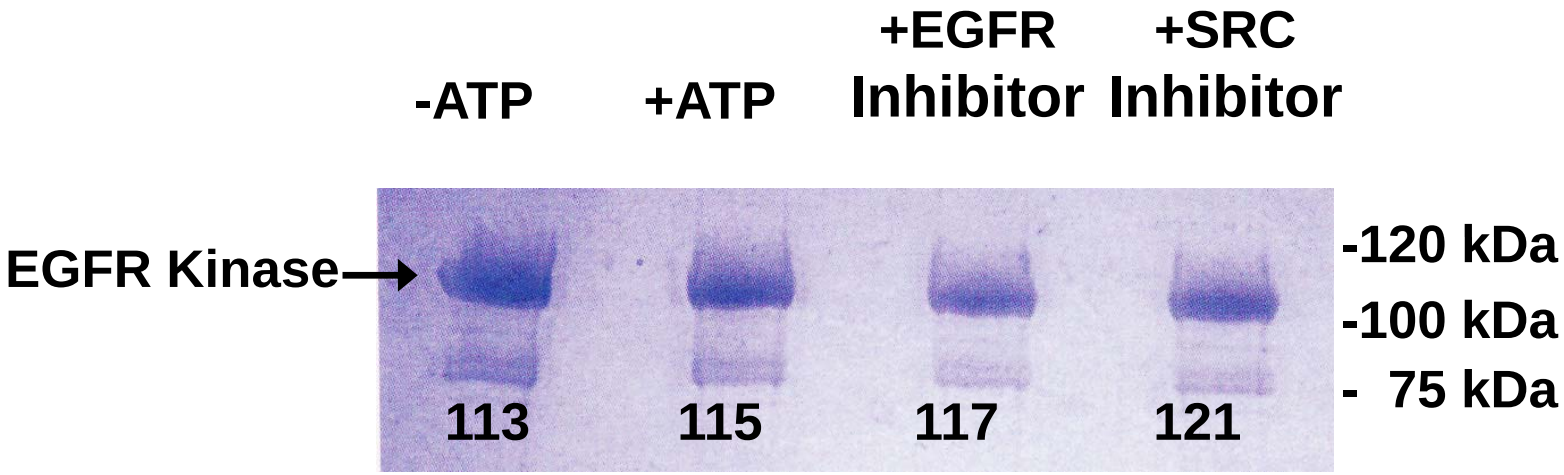
Expressed EGFR Kinase

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)

Rank	Score	%Cov	Name	Species	+ATP 115:113	+EGFR kinase Inhibitor 117:113	+SRC kinase Inhibitor 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)

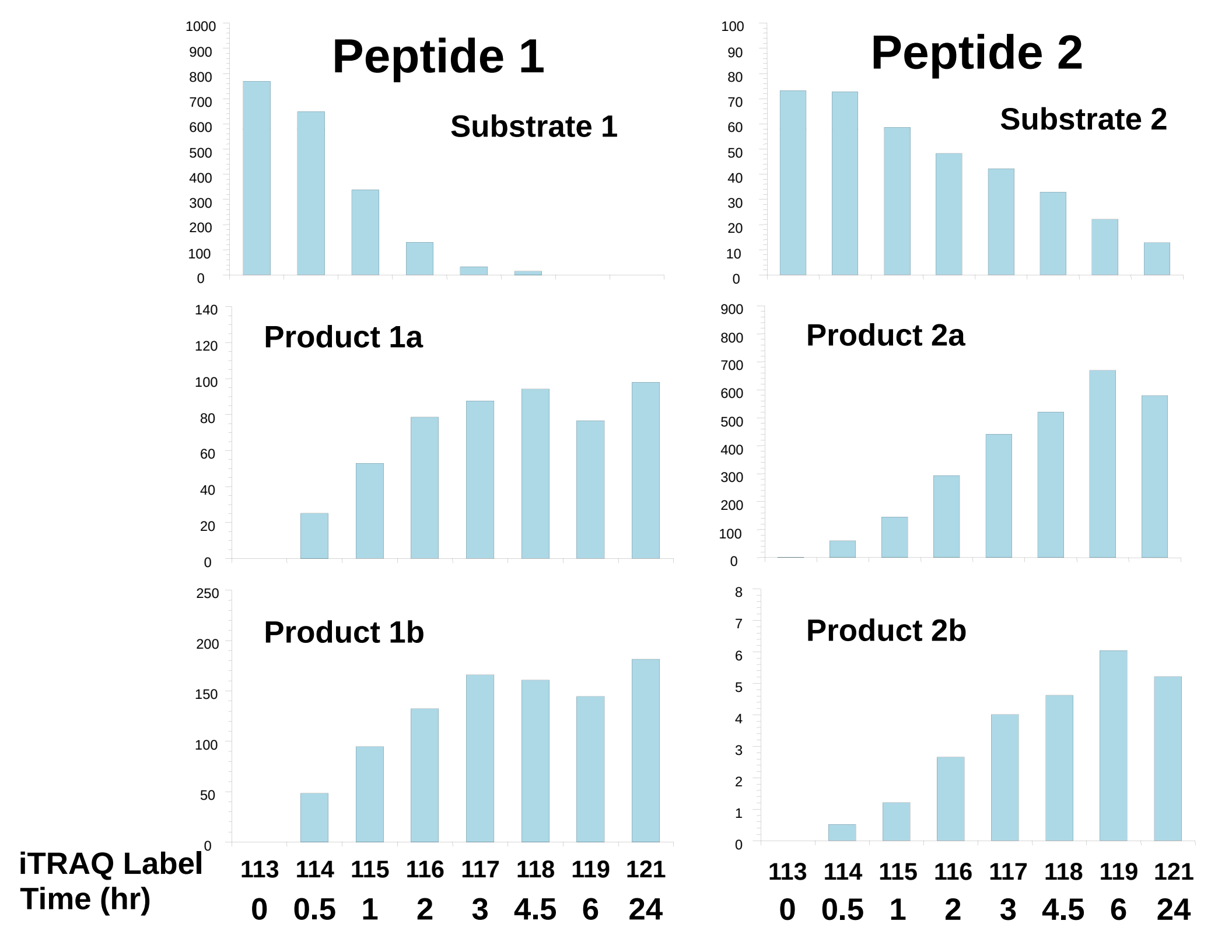
Substrate and products labeled with *same* iTRAQ tag at *each* time point reagent

***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


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

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