Targeted
Metabolomics &
Proteomics
Laboratory

Quantification in Mass Spectrometry: the who, the what and the how much…

Stephen Barnes, PhD

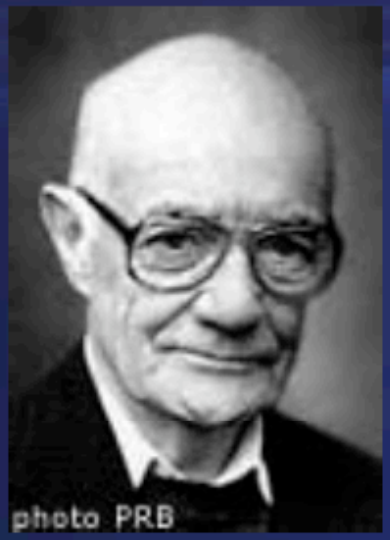
Targeted Metabolomics and Proteomics Laboratory

Synopsis

- Quantification during discovery
  - Spectral counting
  - ICAT (isotope-coded affinity tag)
  - iTRAQ
  - SILAC
- Discovery starting points
- Multiple reaction ion monitoring (MRM)
  - Advantages and limitations
  - Importance of bioinformatics and computational support

Nobel Prize in Chemistry- 2002

For getting proteins and peptides into the gas phase

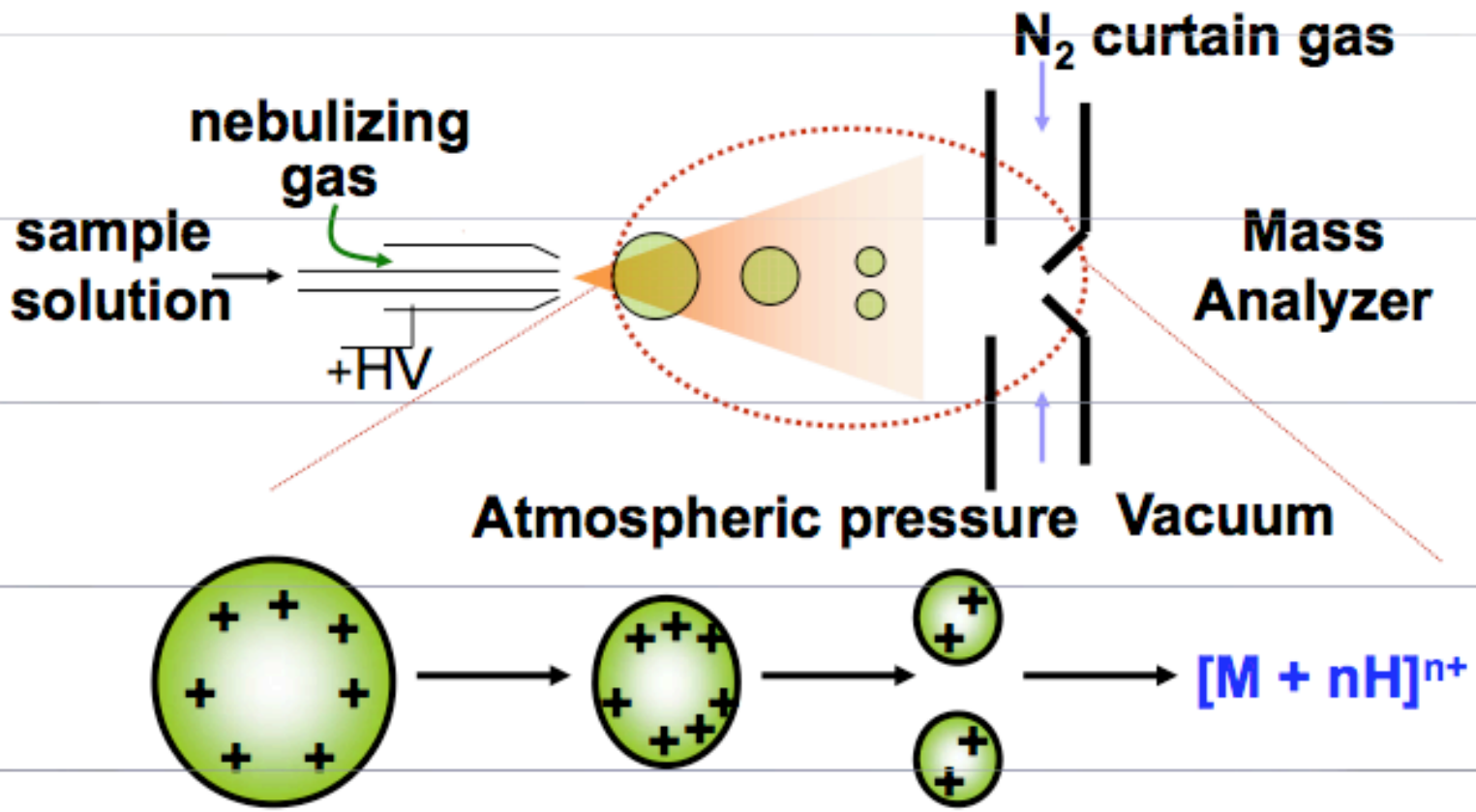


John Fenn

Koichi Tanaka

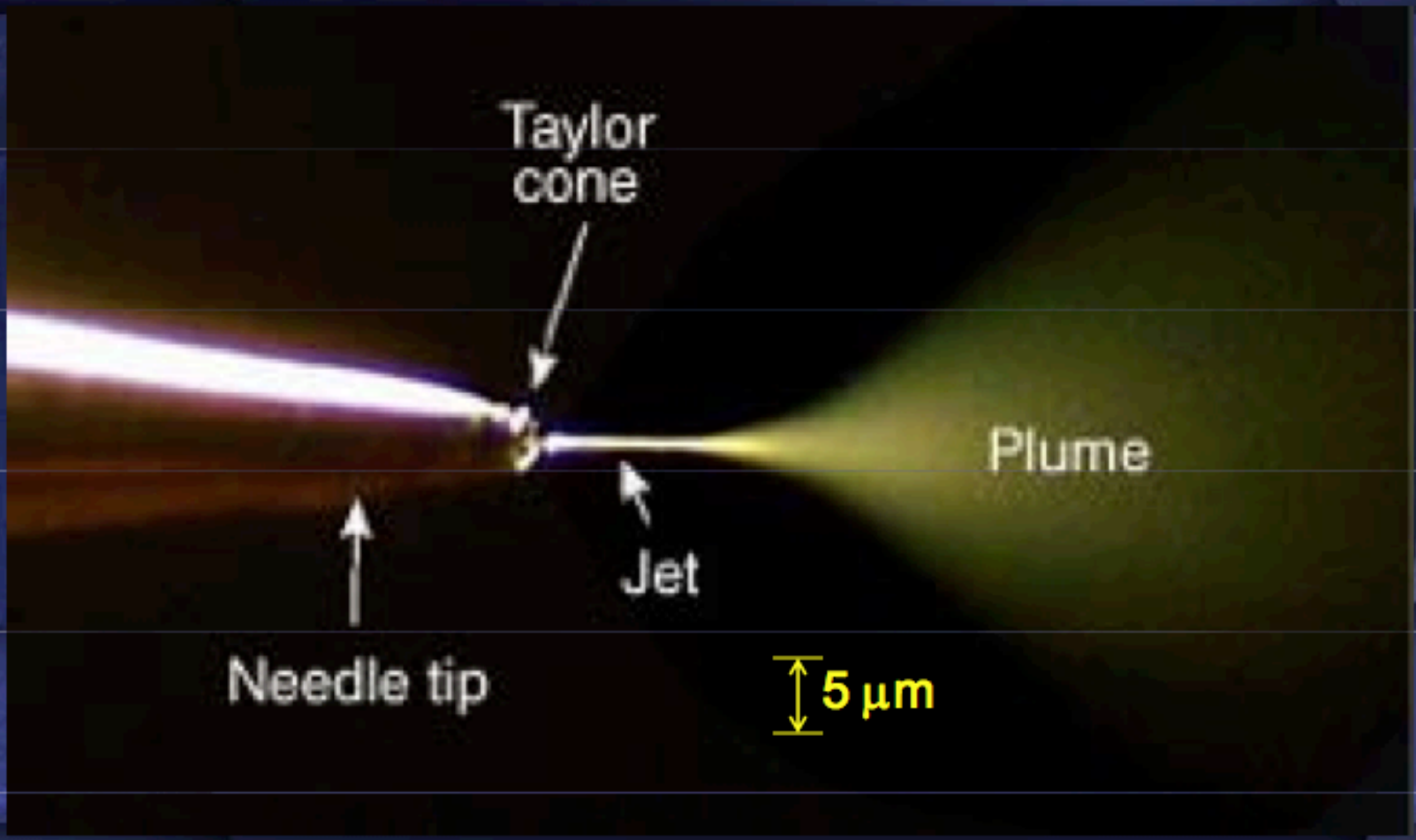
"***for the development of methods for identification and structure analyses of biological macromolecules***"
and
"***for their development of soft desorption ionisation methods for mass spectrometric analyses of biological macromolecules***"

Electrospray Ionization (ESI)

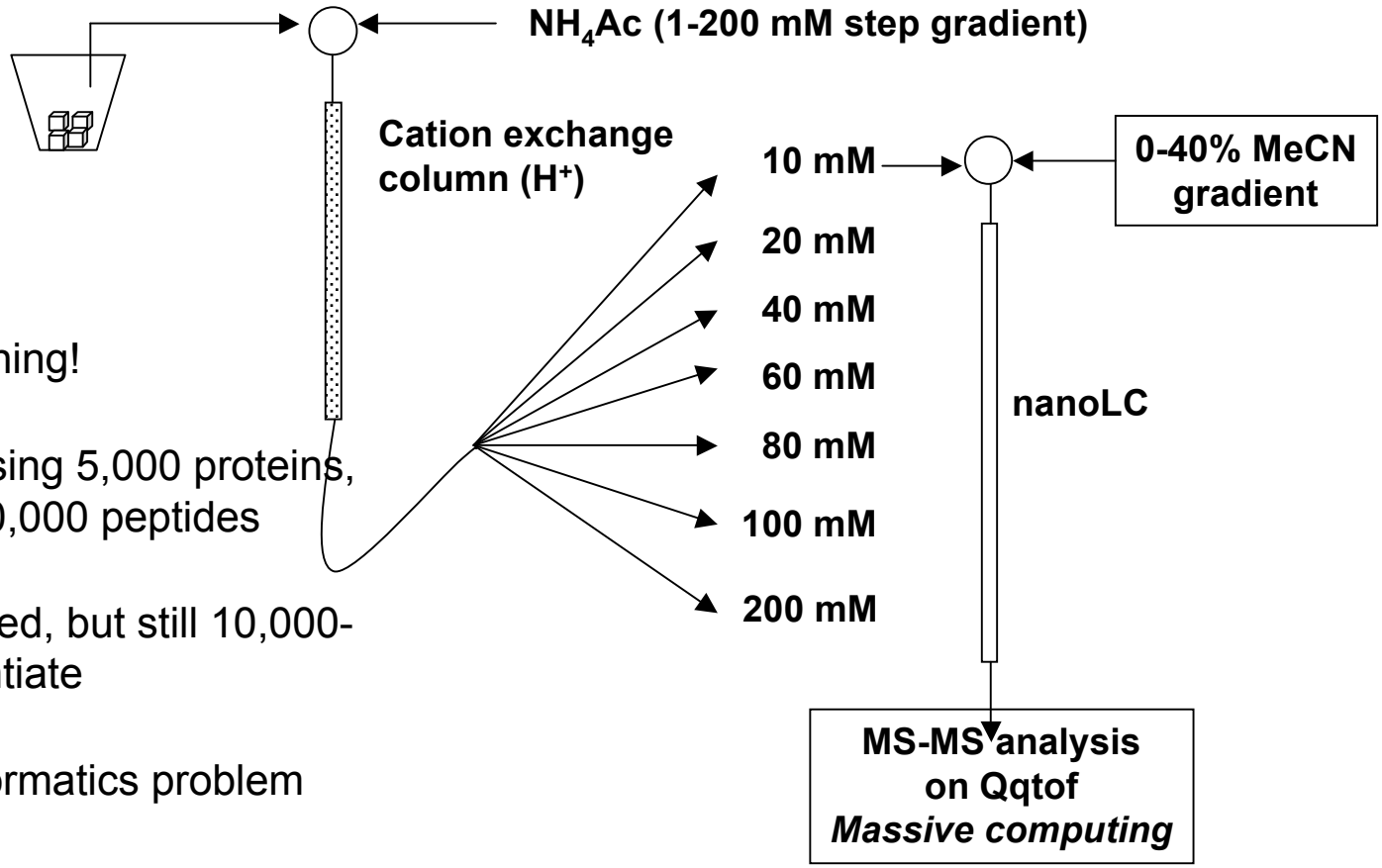


1. Solvent evaporation
2. Coulombic repulsion

Electrospray Ionization (ESI)



MUDPIT - MUlti-Dimensional Protein Identification Technology



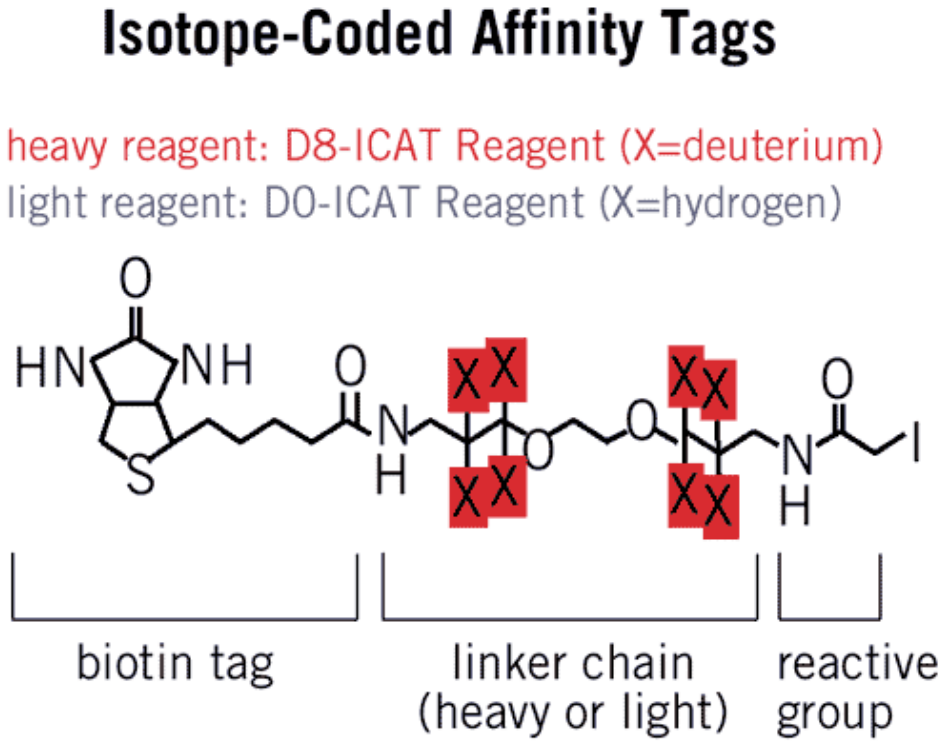
- Hydrolyze everything!
- For a cell expressing 5,000 proteins, this leads to >100,000 peptides
- Can be fractionated, but still 10,000-20,000 to differentiate
- Enormous bioinformatics problem

John Yates

Spectral counting

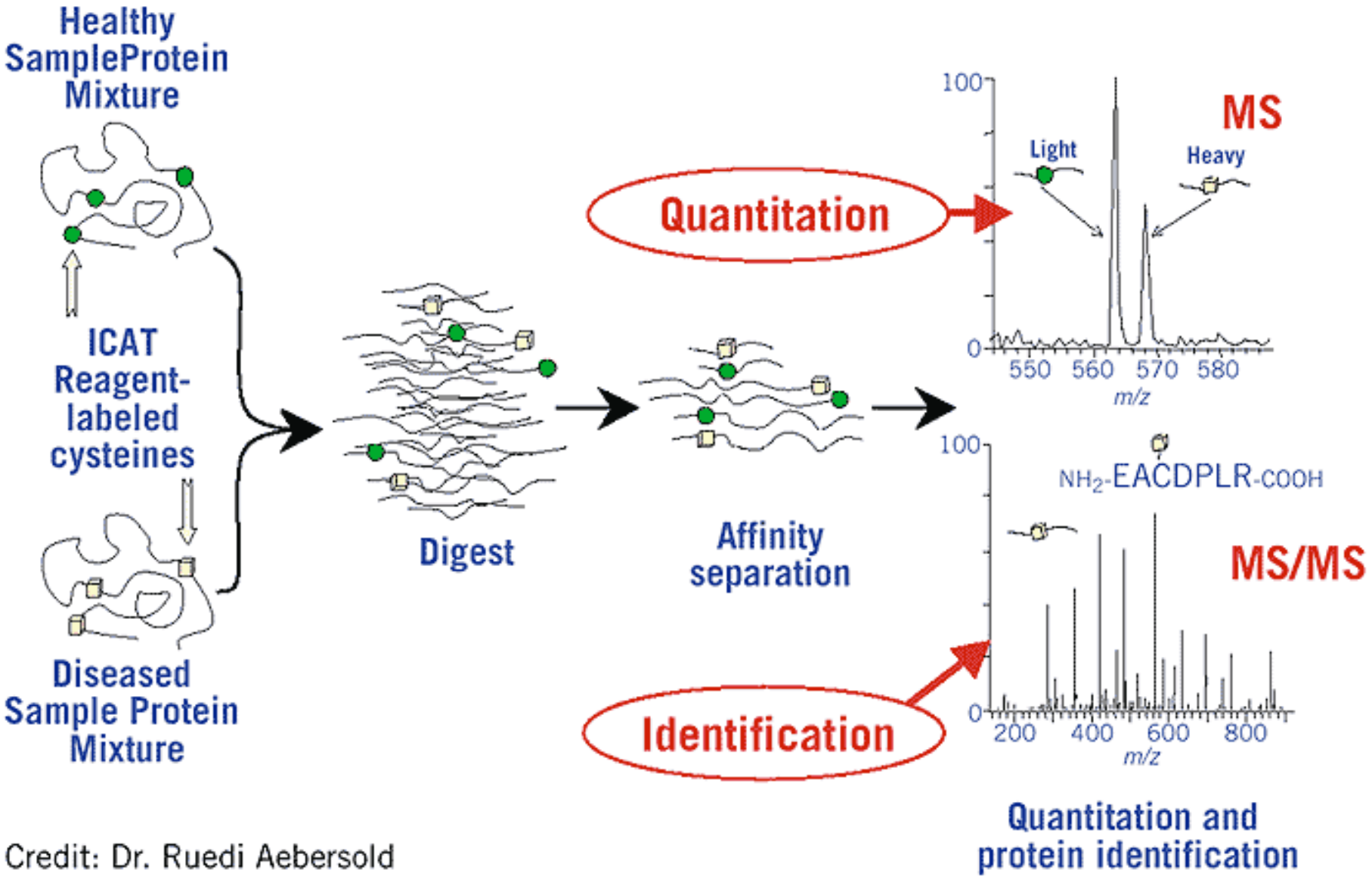
- *Spectral counting* depends on how many peptides from a protein are detected in a tandem MS experiment
- A completely digested protein (1 pmol) should produce 1 pmol of each of its constituent peptides
- However, the molar ionization intensity for these peptides varies at least 100-fold
- A low abundance protein will have only a few recognizable peptides
- It's a robust method that doesn't require special reagents and is part of a discovery experiment

Isotope-coded affinity technology



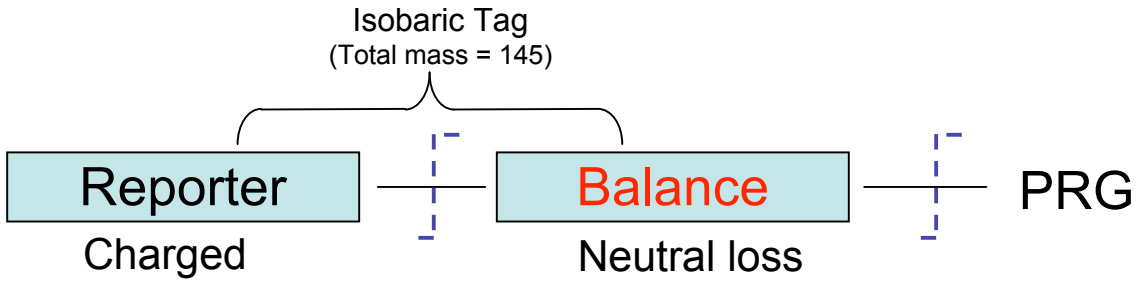
This reagent reacts with cysteine-containing proteins (80-85% of proteome)

Labeling can be replacement of hydrogens (X) with deuterium, or better to exchange ^{12}C with ^{13}C in the linker region (this avoids chromatography issues)



Credit: Dr. Ruedi Aebersold
Institute for Systems Biology, Seattle, WA

iTRAQ™ Reagent Design

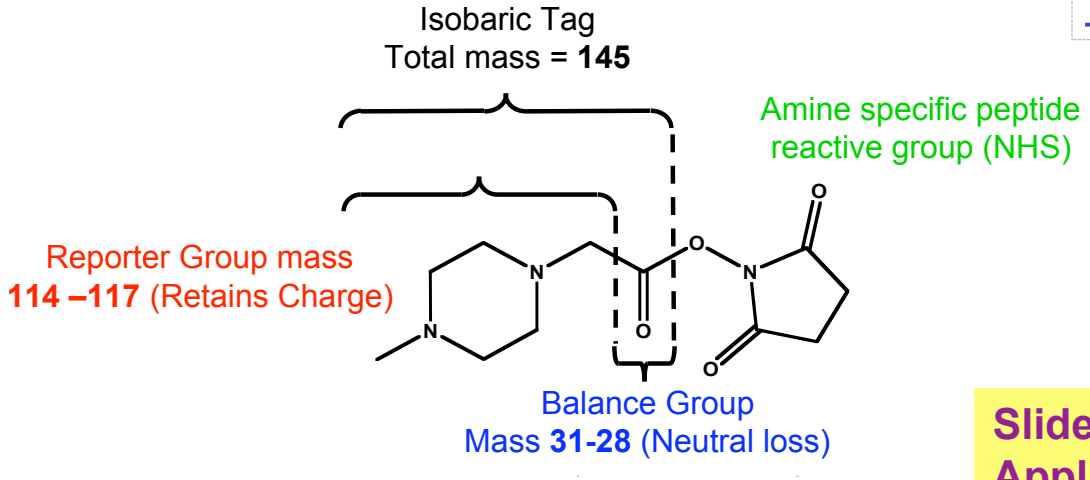


- Gives strong signature ion in MS/MS
- Gives good b- and y-ion series
- Maintains charge state
- Maintains ionization efficiency of peptide
- Signature ion masses lie in quiet region

- Balance changes in concert with reporter mass to maintain total mass of 145
- Neutral loss in MS/MS

Amine specific

= MS/MS Fragmentation Site

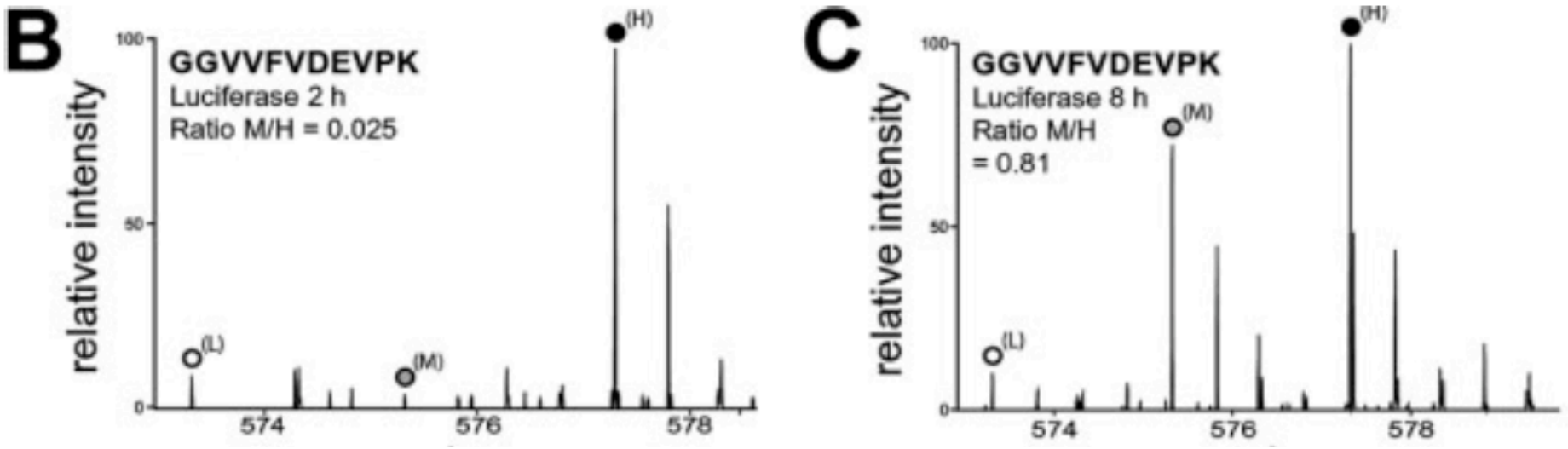


Slide provided by Applied Biosystems

SILAC and protein quantification

- Stable isotope labeling with amino acids in cell culture (SILAC)
- See http://www.silac.org
- Amino acids such as leucine, lysine or tyrosine are added as light (all ^{12}C) or heavy forms (all ^{13}C)
  - e.g., $^{13}C_6$-lysine, $^{13}C_9$-tyrosine
- These amino acids are incorporated into proteins and increase masses of peptides
- Samples are mixed and digested with proteases
- The intensities of the peptide pairs (M, M+6) compared

Pulsed SILAC



Proteomics 2009, 9, 205–209

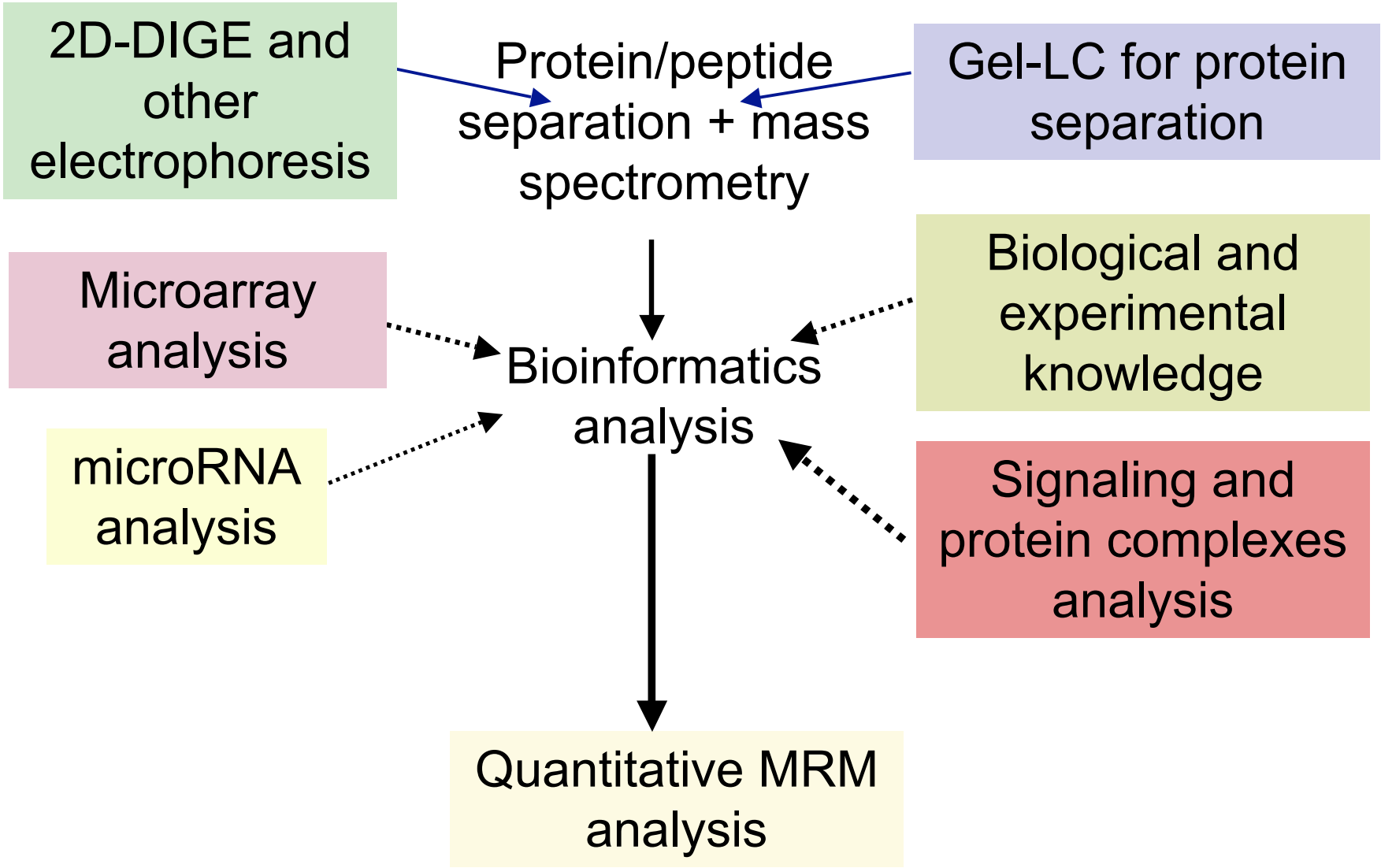
Limitations of quantitative methods

- ***Spectral counting***: **depends on the number of peptides produced from a protein which are recognizable -**
  - **however, tandem MS is a stochastic process**
  - **can measure 2-fold differences**
- ***ICAT***: **the proteins that have cysteines and ones that are free and not oxidized are limited**
  - **the reagent is reacted with the intact proteins**
- ***iTRAQ***: **reacts with lysines (more abundant than cysteines)**
  - **the large number of steps that are involved leads to variable losses**
  - **Also, the label is added after the hydrolysis**
- ***SILAC:*** **requires good chromatography**
  - **best for cells - too expensive for animal studies**

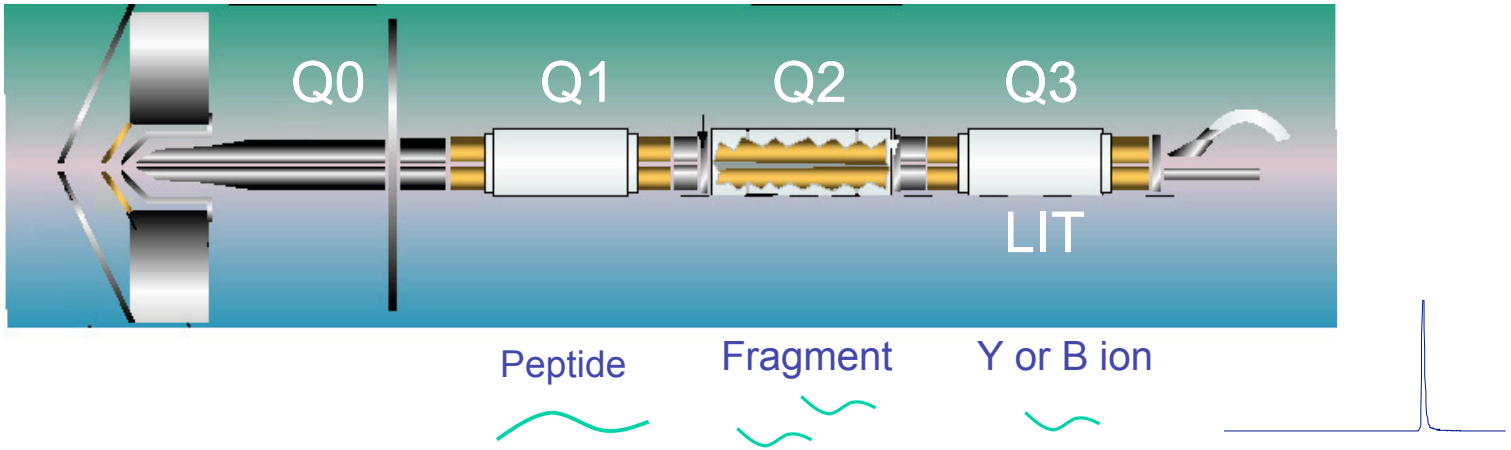
On the way to absolute quantification

- **All forms of proteomics and mass spectrometry are a compromise**
- **At best only a small part of the proteome can be observed due to sample solubility, complexity and concentration (dynamic range of 9-10 orders of magnitude), as well as instrument detection bias**
- **If we "know" what we want to measure, then we can focus our attention on a limited, but if carefully selected, important group of proteins and determine them at much higher sensitivity**

Workflow for generation of proteomics data



Multiple Reaction Monitoring

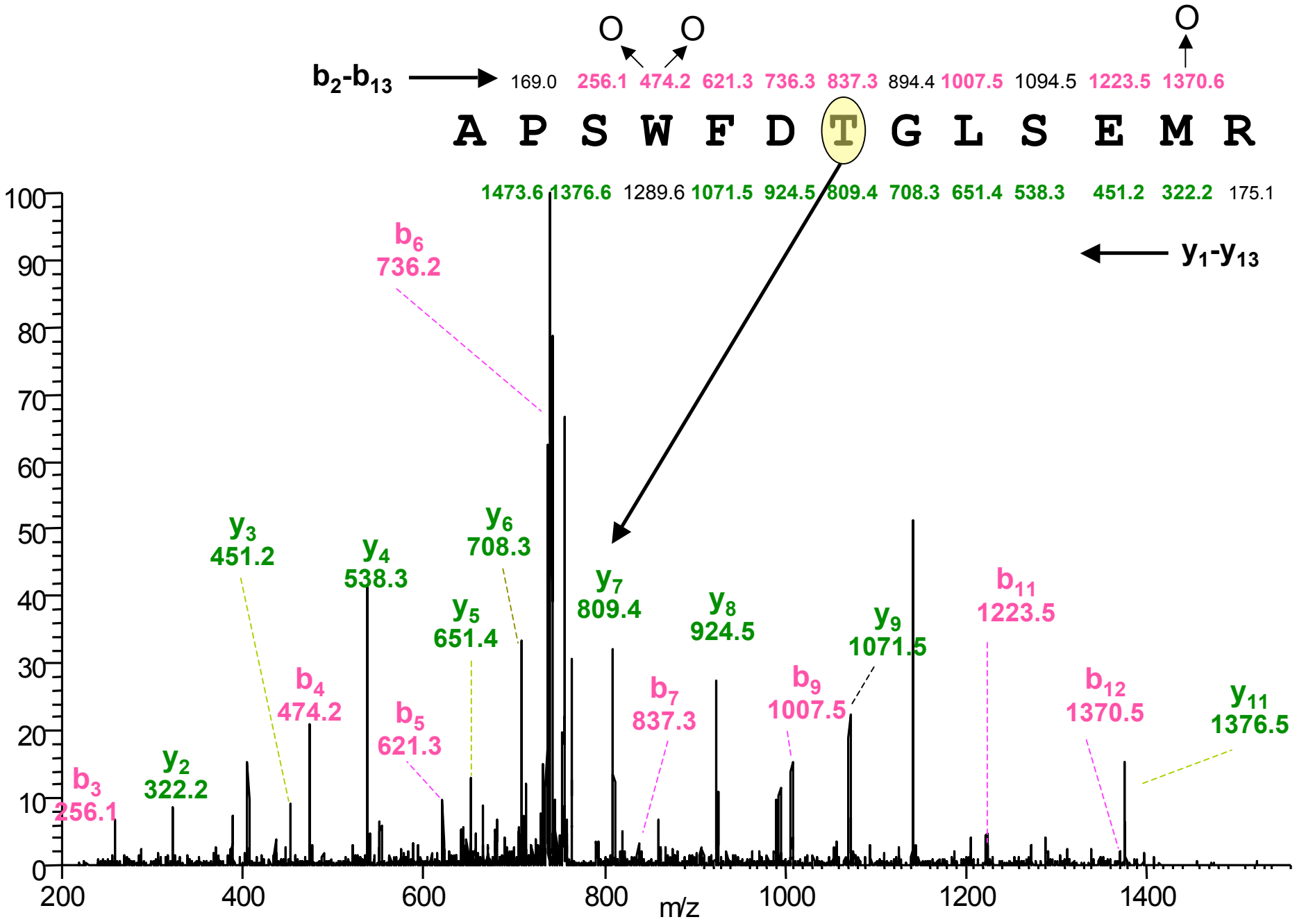


- **Highest specificity and sensitivity for detecting components in a complex mixture**
- **Largest linear dynamic range for quantitation**
- **Well accepted as the MS technique for quantification (small molecule world)**
- **Triple quadrupole or Q TRAP® system mass spectrometers**

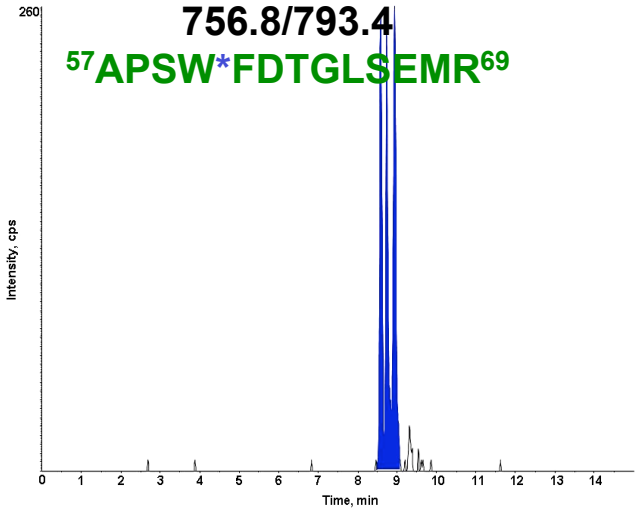
Quantification of peptides

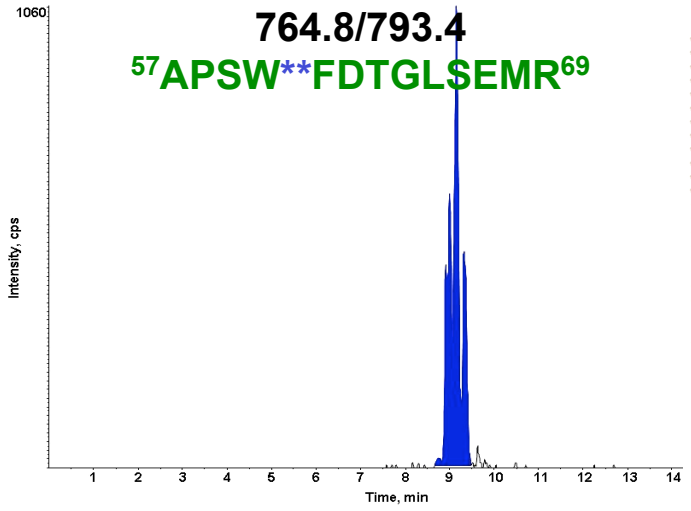
- **MRM-MS collects data points to describe a peak**
- **Enough ions can be collected in 10-20 msec, enabling a switch to a different parent ion/fragment ion combination ("channel") for the next 10-20 msec, etc.**
- **When the list of peptide measurements has been completed, then we go back to the top of the list for the next cycle (multiple reaction ion monitoring - MRM)**
- **The data points (10 per peak) are used to recreate the peak and hence measure the area underneath it.**

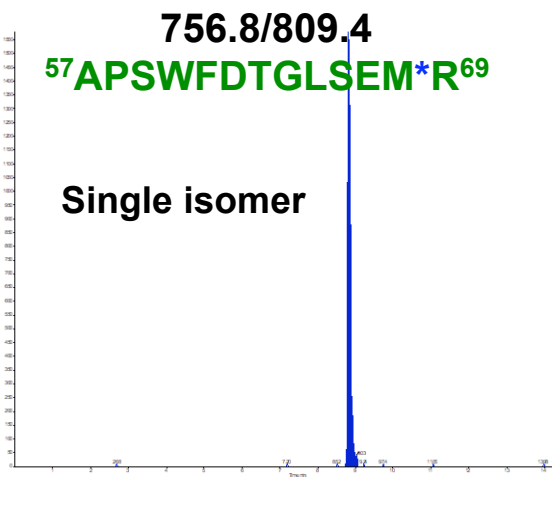
Oxidized tryptic peptide aa57-69 from hαB-crystallin

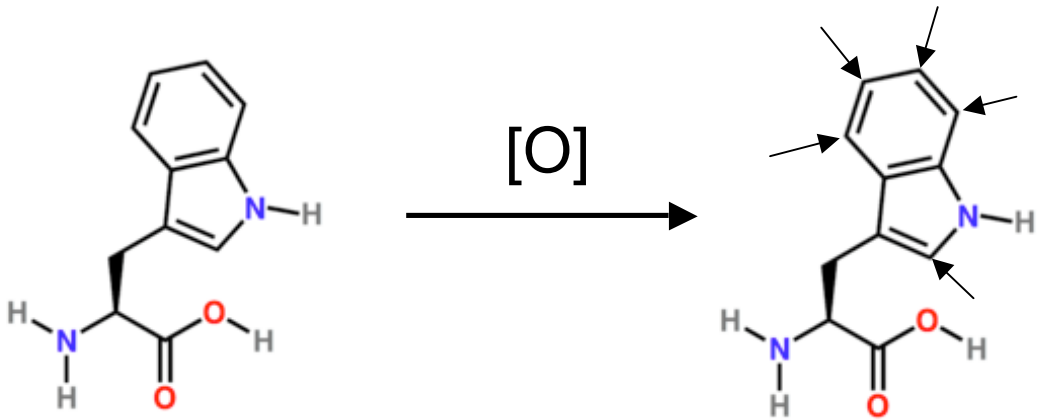


nanoLC-MRM-MS analysis reveals that UVA light oxidation of tryptophan occurs at several sites







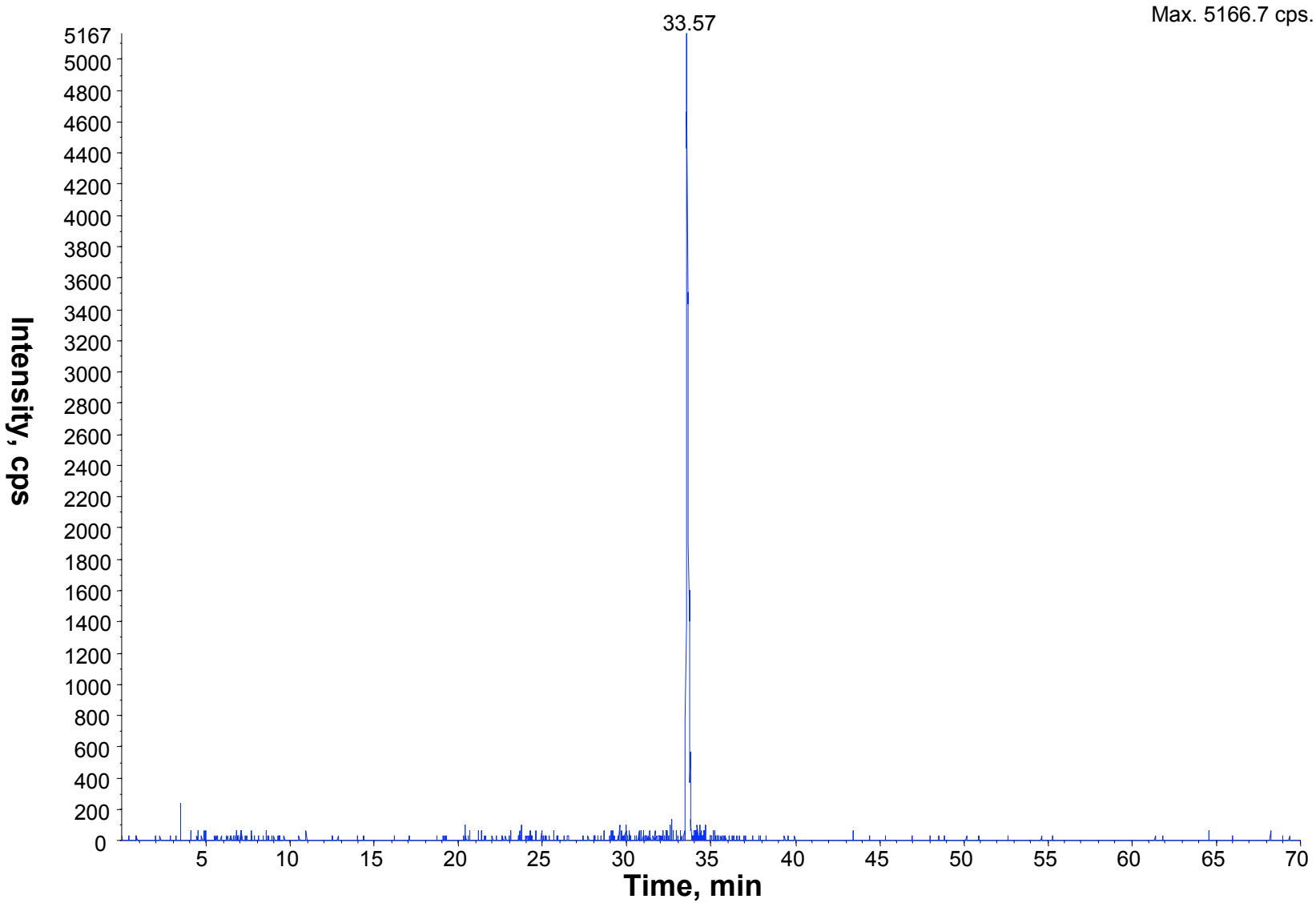


Addition of one O leads to five possible isomers

Addition of two O leads to ten possible isomers

Cystin peptide detected in whole kidney digest by MRM-MS

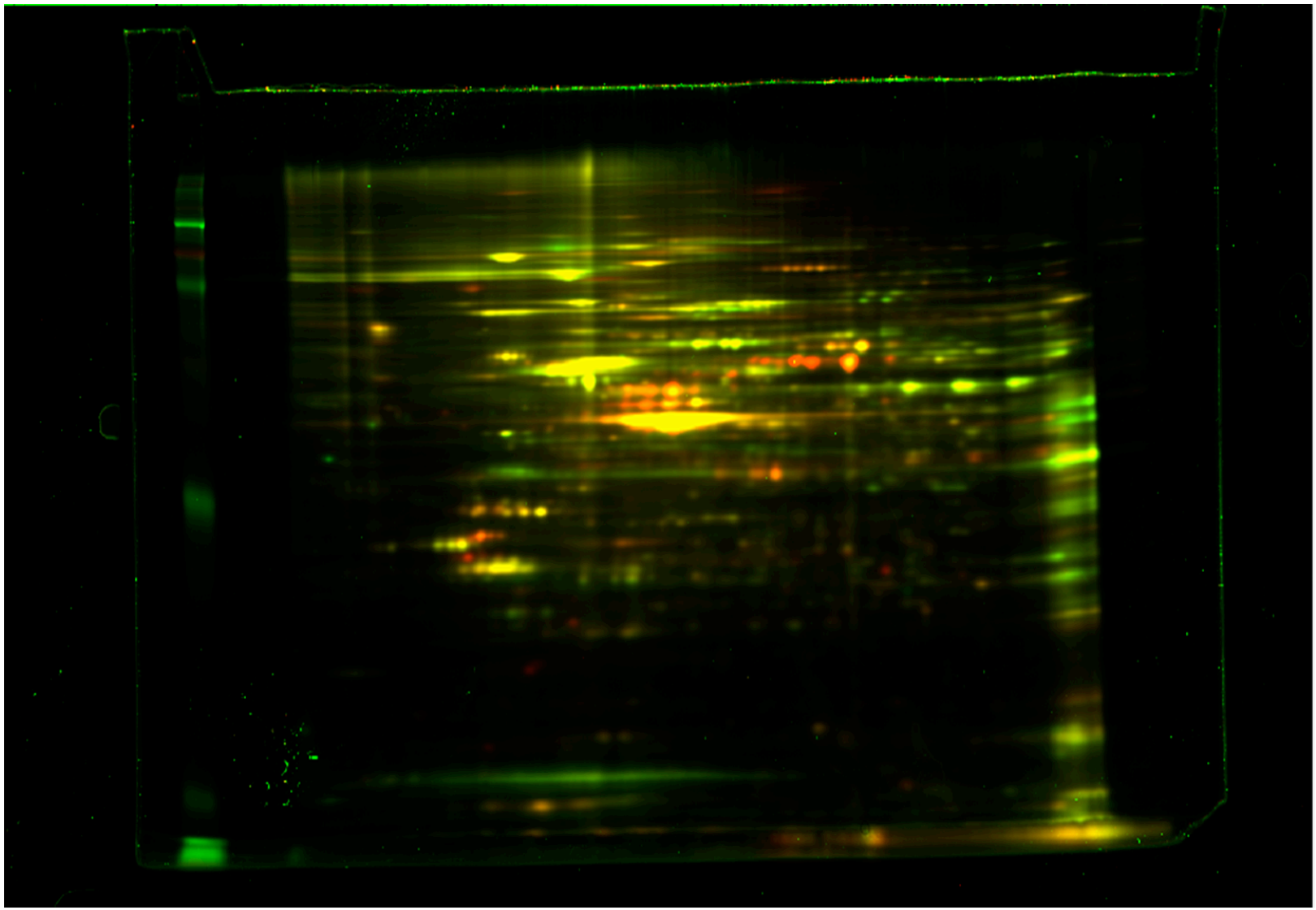
LAPAVSPEK MRM Parent ion: 456.26++ *m/z* Daughter Ion: 559.3 *m/z* (y^5)



Choosing the precursor and fragment ions for MRM

- **Without preliminary data**
  - **Use the Sciex software program MIDAS - it interrogates a library of known peptides and selects the predicted most intense ions**
  - **Our experience is that it is a poor predictor**
- **With preliminary data**
  - **Inspect the peptide MSMS spectra that we have previously obtained to identify proteins**
  - **These spectra are "guaranteed" to work since they are obtained on the same instrument; however, the Qtrap and the triple quad collision parameters are not the same**

DIGE of two cell protein extracts



Red: cell 1; Green: cell 2

Partial list of protein identities of spots indicated to be significantly different between cell 1 and cell 2 (rescued with *Polaris*)

Spot #	Protein Identification	MOWSE score	Accession number	mw (obs) kDa	mw (pred) kDa	pI (obs)	pI (pred)	T-test	ratio R/C
83	No identificaiton	N/A	N/A	250.00	N/A	5.75	N/A	1.60E-03	-2.03
211	gelsolin, cytosolic	49	gi|90508	101.00	81.00	5.85	5.53	2.16E-05	-2.02
212	gelsolin, cytosolic	203	gi|90508	100.00	81.00	6.01	5.53	7.98E-05	-2.05
404	Keratin 8	502	gi|76779293	54.00	55.00	5.75	5.71	7.30E-07	-2.52
406	Keratin 8	995	gi|76779293	50.00	55.00	5.75	5.71	4.17E-06	-5.5
412	Keratin 8	988	gi|76779293	52.00	55.00	5.83	5.71	6.02E-07	-3.77
468	Keratin 18 Keratin 20	320 269	gi|6754482 gi|21592285	43.00	47.00 49.00	5.31	5.31 5.31	1.10E-04	-2.47
476	hypothetical protein LOC100037084 (Keratin 18)	678	gi|147899980	45.00	47.00	5.32	5.22	4.69E-06	-3.24
477	hypothetical protein LOC100037084 (Keratin 18)	1251	gi|147899980	47.00	48.00	5.40	5.22	1.65E-08	-4.44
490	eukaryotic translation initiation factor 4A isoform 1 hypothetical protein LOC100037084 (Keratin 18) Keratin 20	471 384 344	gi|4503529 gi|147899980 gi|21592285	43.00	46.00 48.00 49.00	5.60	 5.22 5.31	3.40E-02	-3.47

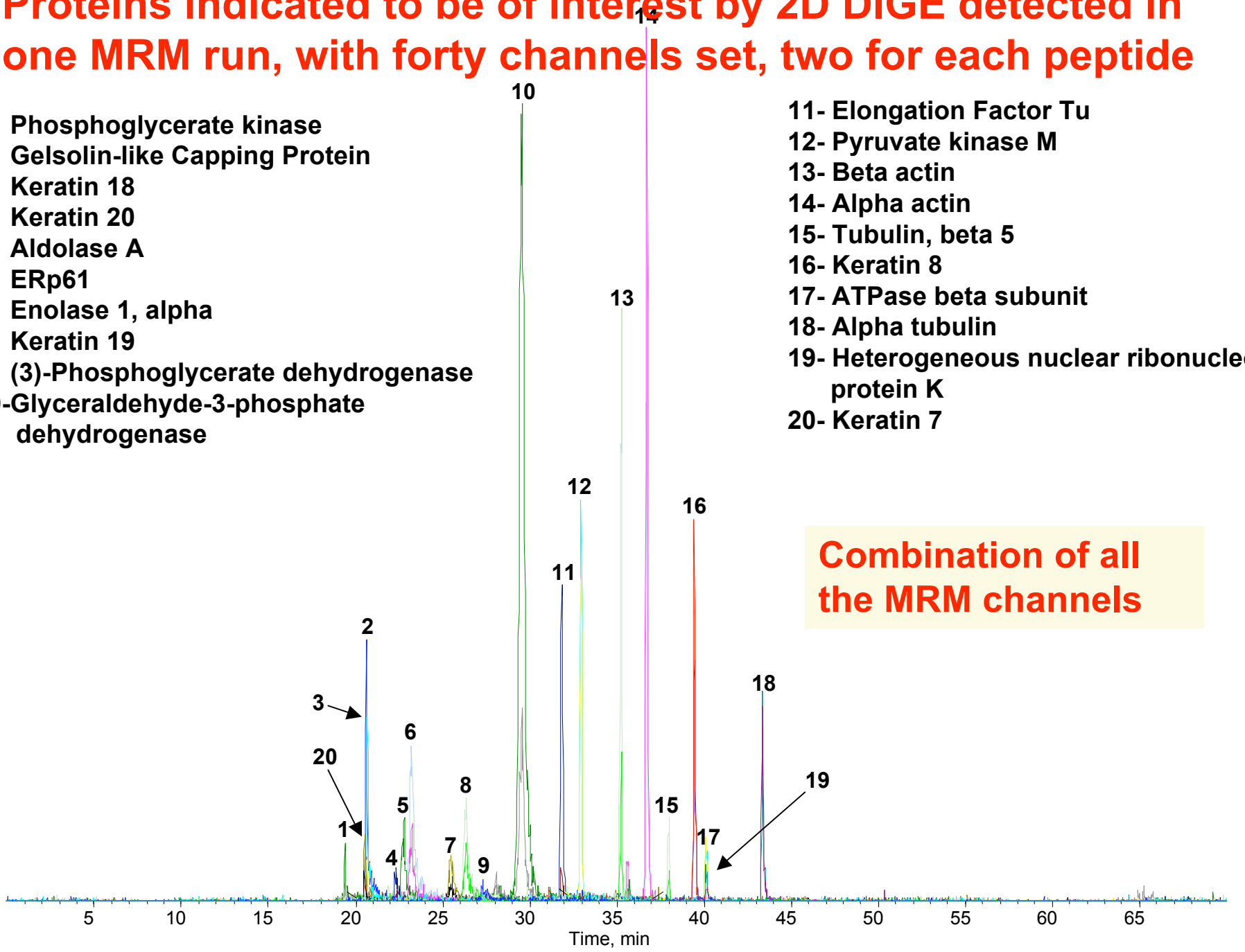
Being selective - pragmatically choosing the MRM ions from existing tandem spectra

MRM Transitions for PKD Proteins R3/C3				
Protein	Parent m/z	Daughter Ion		Sequence
3-phosphoglycerate dehydrogenase (9)	566.04	645.4	y6	VTADVINAAEK
		744.43	y7	VTADVINAAEK
Aldolase A (5)	566.9	563.3	y5	ALANSLACQGK
		763.4	y7	ALANSLACQGK
Alpha actin (14)	896.04	901.5	y8	SYELPDGQVITIGNER
		1086.6	y10	SYELPDGQVITIGNER
Alpha tubulin (18)	744.6	834.5	y8	LISQIVSSITASLR
		933.5	y9	LISQIVSSITASLR
ATPase beta subunit (17)	720.3	885.5	y7	VALTGLTVAEYFR
		1055.6	y9	VALTGLTVAEYFR
Beta actin (13)	473.3	531.3	y5	AVFPSIVGR
		628.4	y6	AVFPSIVGR
Elongation factor Tu (11)	513.3	685.4	y7	IGGIGTVPVGR
		912.5	y10	IGGIGTVPVGR
Enolase 1, alpha non-neuron (7)	572.3	511.3	y4	IGAEVYHNLK
		674.4	y5	IGAEVYHNLK

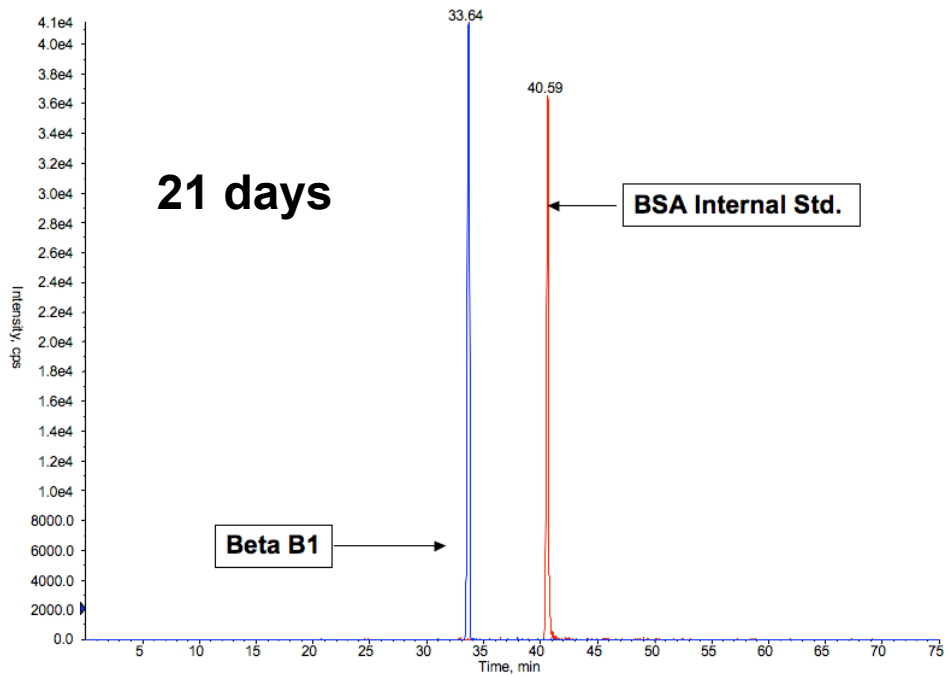
Proteins indicated to be of interest by 2D DIGE detected in one MRM run, with forty channels set, two for each peptide

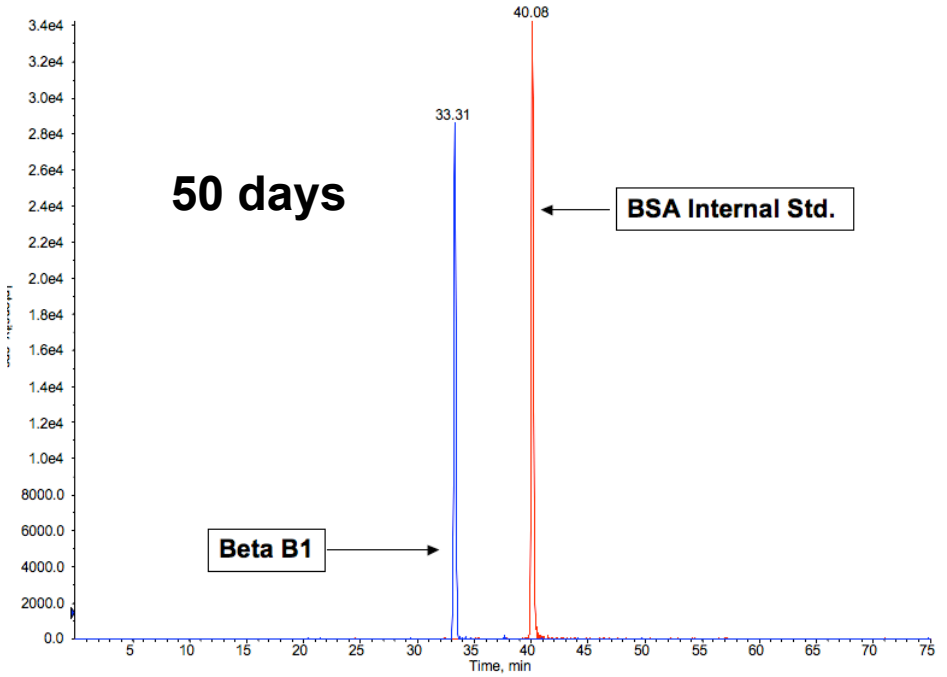
1- Phosphoglycerate kinase
2- Gelsolin-like Capping Protein
3- Keratin 18
4- Keratin 20
5- Aldolase A
6- ERp61
7- Enolase 1, alpha
8- Keratin 19
9- (3)-Phosphoglycerate dehydrogenase
10-Glyceraldehyde-3-phosphate dehydrogenase

11- Elongation Factor Tu
12- Pyruvate kinase M
13- Beta actin
14- Alpha actin
15- Tubulin, beta 5
16- Keratin 8
17- ATPase beta subunit
18- Alpha tubulin
19- Heterogeneous nuclear ribonucleo protein K
20- Keratin 7

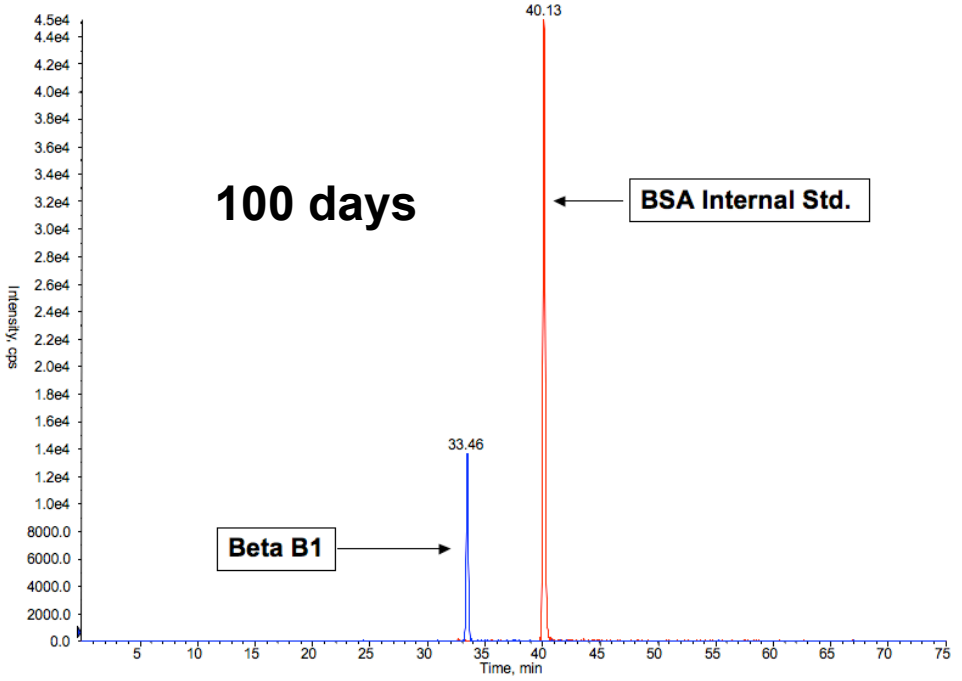


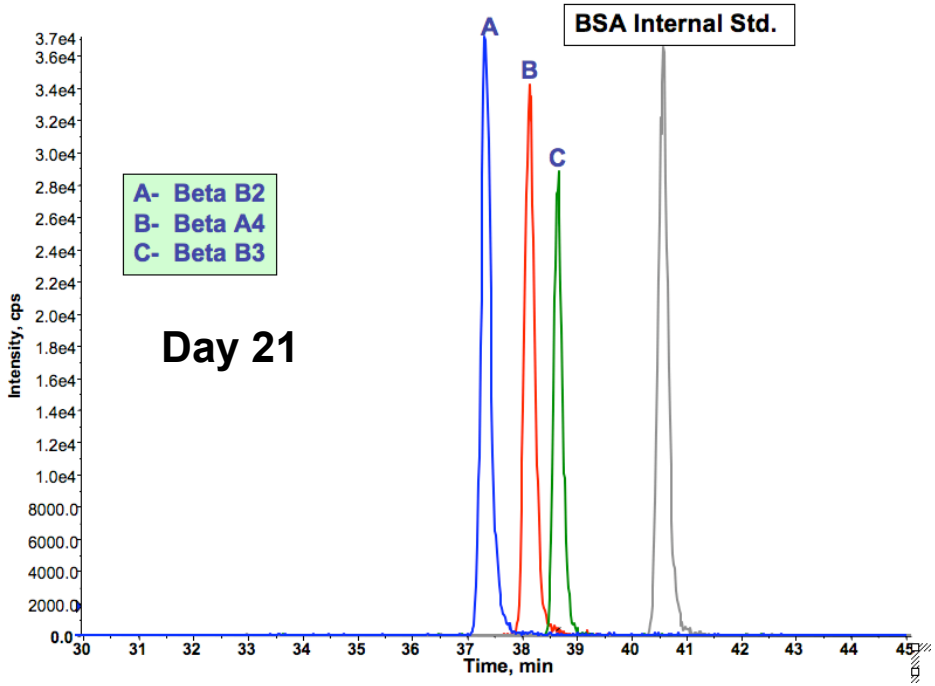
Combination of all the MRM channels

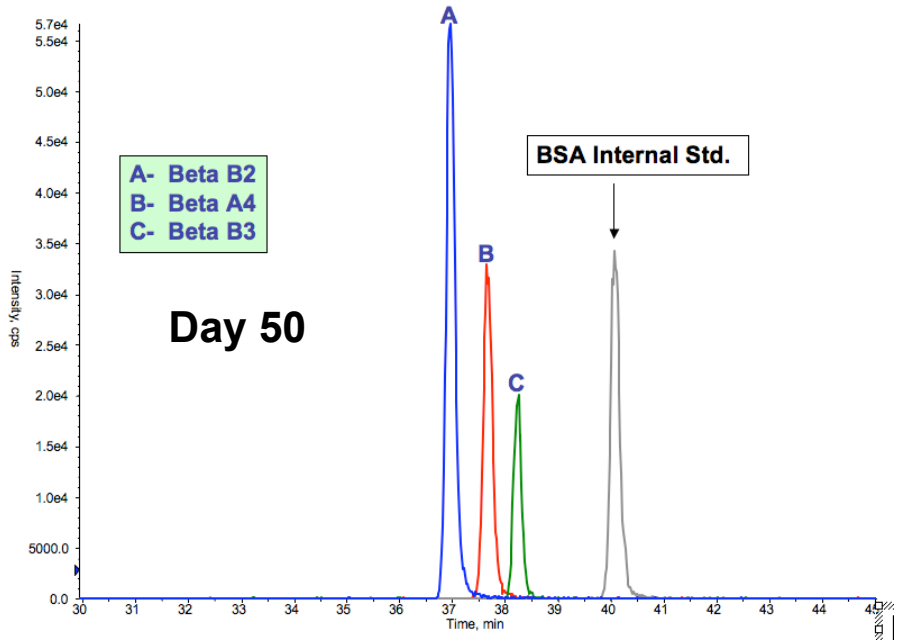




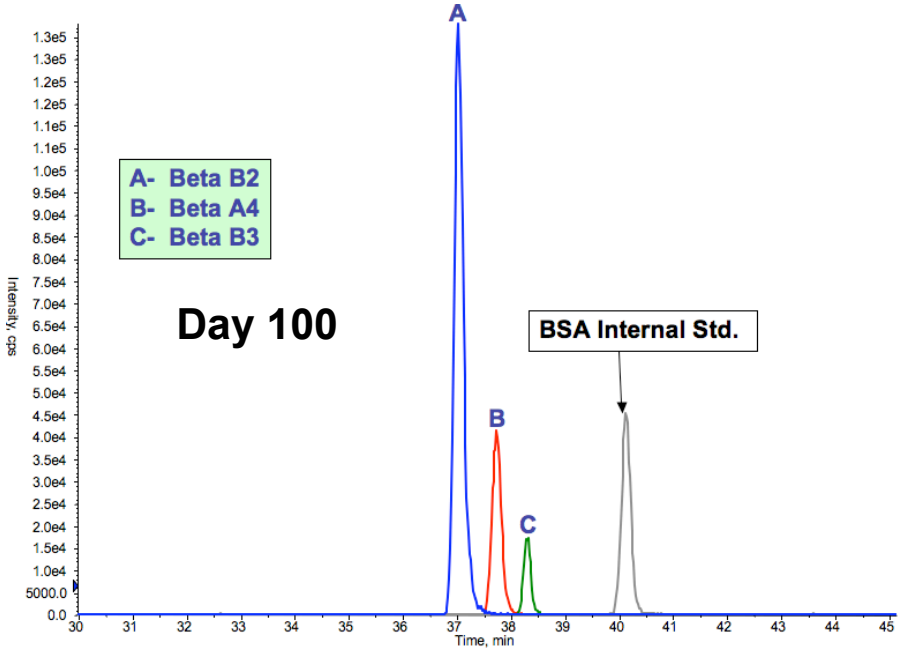
Decreasing amount of beta B1 crystallin in the lens with the age of the rat - note the BSA peptide as the internal standard



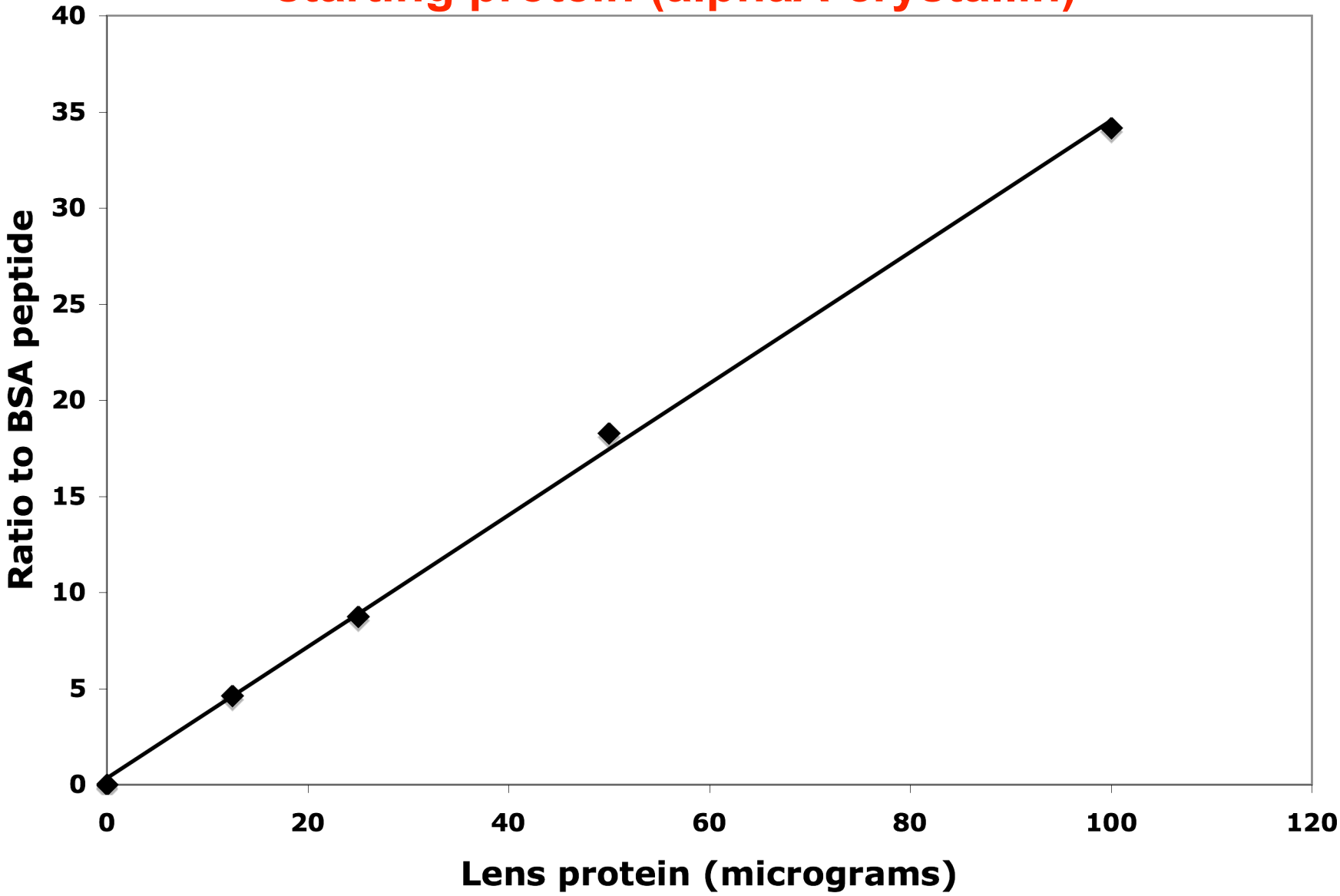




Increasing amount of beta B2 crystallin in the lens with the age of the rat - note the BSA peptide as the internal standard



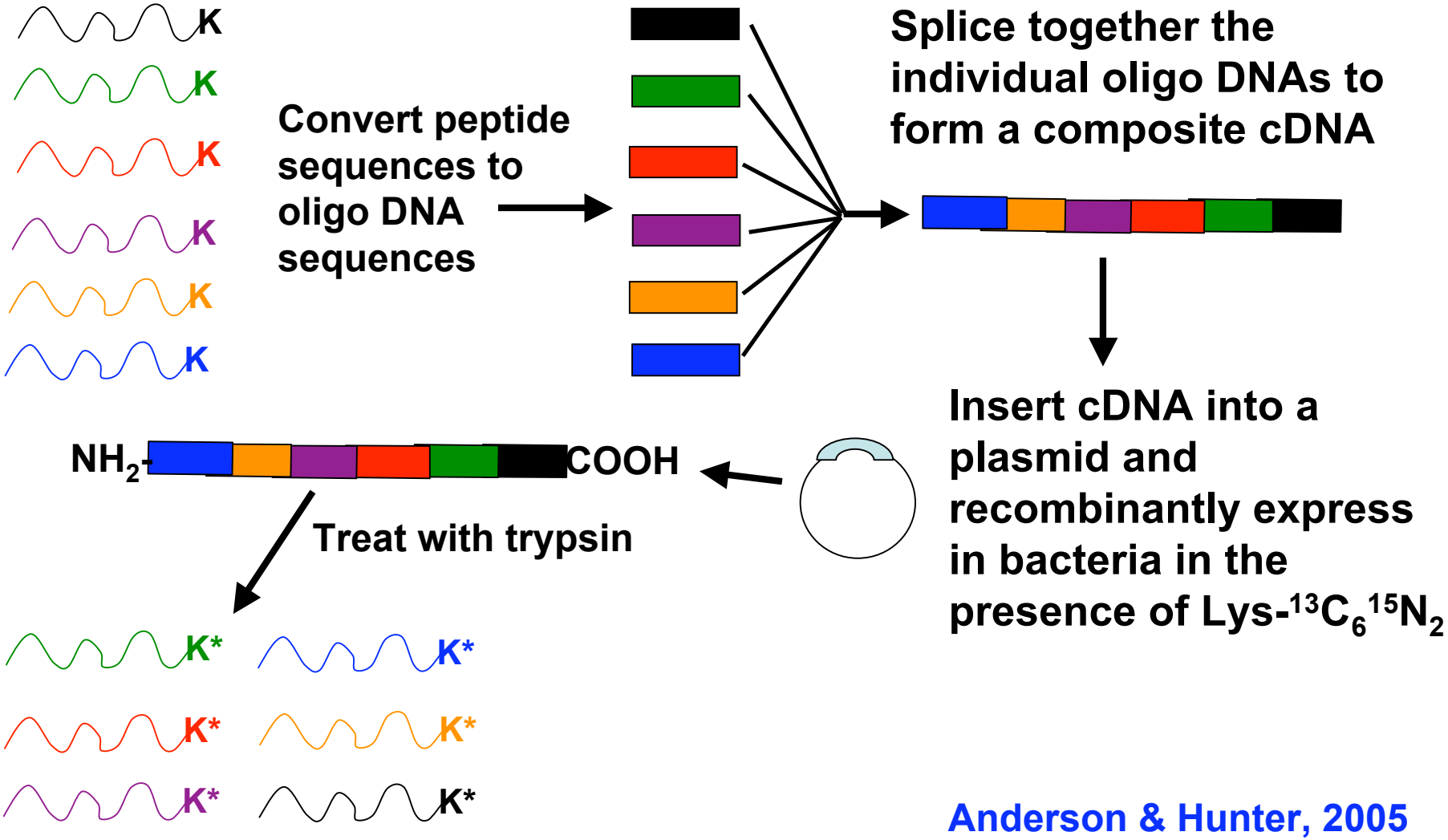
Linearity of the area response for the peptide from a starting protein (alphaA-crystallin)



Selecting an internal standard

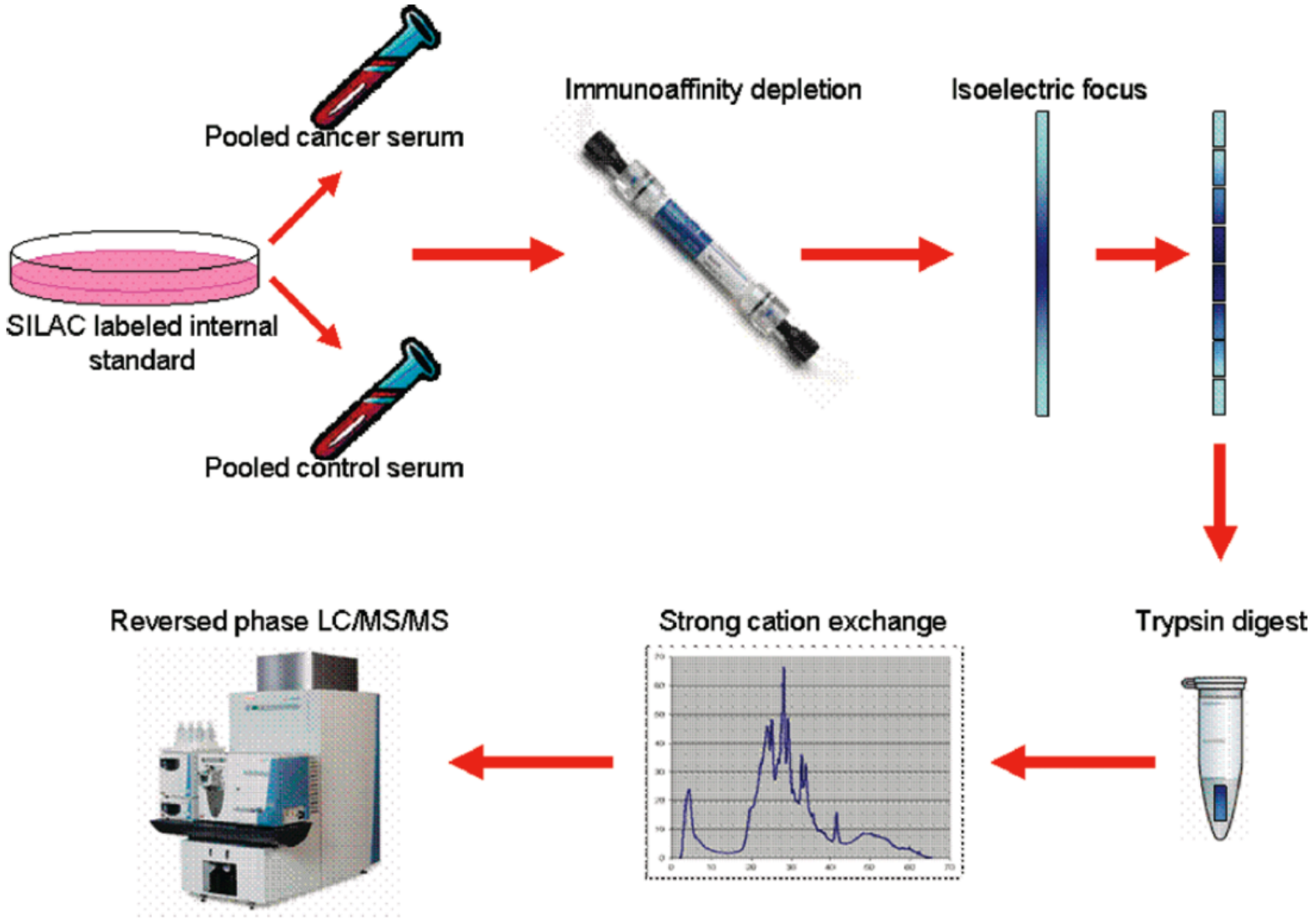
- A protein that is not present in the sample of interest can be
  - e.g., in rat lens, add BSA and measure a unique peptide
- A synthetic ^{13}C-labeled peptide for each protein of interest
  - Very elegant , but expensive ($1,000 per peptide)
  - Does not assess hydrolysis efficacy
- Concatenation (joining the peptides as their oligonucleotide analogs, creating a macro cDNA and expressing the hybrid protein)
  - Very elegant, sophisticated use of molecular biology
- SILAP labeling - creates proteome-wide, stable isotope labeled proteins

Concatenation - making ^{13}C-labeled peptide internal standards



Anderson & Hunter, 2005

SILAC use as an internal standard



Journal of Proteome Research 2009, 8, 1565–1576

Stable isotope amino acids

http://www.isotope.com

Amino acid	Amount (g)	Price ($)
• L-Arg $^{13}C_6$	0.1, 0.5	650, 1880
• L-Arg $^{13}C_6$ $^{15}N_4$	0.1, 0.5	855, 2445
• L-Lys $^{13}C_6$	0.1, 0.25	443, 1129
• L-Lys $^{13}C_6$ $^{15}N_2$	0.1, 0.25	345, 580
• L-Tyr $^{13}C_9$	0.1, 0.25	635, 1240
• L-Tyr $^{13}C_9$ ^{15}N	0.1, 0.25	275, 410

These are all 97-99% pure. Prices for 99% or better are not listed, but will be higher.

Summary

- Quantitative proteomics has arrived
- Specific hypotheses can now be tested with high precision
- Limitations lie in having enough bioinformatics resources to “boil” approaches down
- Other limitations are in the basic biochemistry of proteolysis

Many thanks to

- Landon Wilson
- Ray Moore
- Sai Sai Dong
- Richie Herring

- David Stella
- Michael Heaven
- Kyle Floyd
- Elizabeth Gulsby

- Chiquito Crasto, PhD
- Helen Kim, PhD
- Matt Renfrow, PhD
- Om Srivastava, PhD

UAB Centers support

- Acute Kidney Injury Center
- Botanicals
- CNGI
- Polycystic Kidney Disease
- Skin Disease Research
- (Cancer Center)

Federal support

- R21 AT004621
- S10 RR17261
- S10 RR19231

Protein Name:	Molecular Weight	Assay Peptides	MRM Scans Needed	Swiss Pro	Accession:
Eukaryotic translation initiation factor 4E	25053	5	1	P63073	AAH10759
eukaryotic translation initiation factor 4E binding protein 2	12898	1	1	P70445	NP_034254
vascular endothelial growth factor A	15963	1	1	Q00731	AAM55477
hypoxia-inducible factor 1 alpha	91873	16	5	Q61221	AAC52730
Akt1 protein	23290	7	2	P31750	AAI15584
eukaryotic translation initiation factor 4, gamma 1	176078	39	8	Q6NZJ6	NP_666053
Eukaryotic translation initiation factor 4A	46154	12	3	Q5F2A7	BAA25075
rhodopsin	39070	3	1	P15409	NP_663358
Rac GTPase-activating protein 1	70159	19	4	Q9WVM1	NP_036155
protein kinase C, alpha	76825	8	2	P20444	NP_035231
eukaryotic translation initiation factor 4B	68840	16	4	Q8BGD9	NP_663600
G protein beta subunit-like	35851	4	1	Q9DCJ1	NP_064372
Mammalian target of rapamycin	288739	59	10	Q9JLN9	Q9JLN9
raptor	149499	31	7	Q8K4Q0	NP_083174
protein phosphatase 2a	35609	7	2	P63330	NP_062284
FK506 binding protein 1a	11923	0	0	P26883	AAH04671
RAS-homolog enriched in brain	20452	7	2	Q921J2	NP_444305
Rapamycin-insensitive companion of mTOR	191571	51	11	Q6QI06	Q6QI06
ribosomal protein S6	28681	4	1	P62754	NP_033122
eukaryotic translation initiation factor 3, subunit 10 (theta)	161938	44	9	P23116	NP_034253
ribosomal protein S6 kinase, polypeptide 1 isoform 1	59146	12	3	Q505N6	NP_001107806
Tuberous sclerosis 2	194100	50	10	Q7TT21	AAH52449
tuberous sclerosis 1	128250	34	7	Q9EP53	CAM22296
DNA-damage-inducible transcript 4	24871	4	1	Q9D3F7	NP_083359
ribosomal protein S6 kinase polypeptide 1	82847	23	5	A3KGL7	NP_033123
3-phosphoinositide dependent protein kinase-1 isoform A	63759	17	4	Q9Z2A0	NP_035192
LKB1	49627	10	2	Q9WTK7	BAA76749
insulin receptor substrate 1	130585	23	5	P35569	NP_034700
Harvey rat sarcoma virus oncogene 1 isoform 1	21298	6	2	Q76LV5	NP_032310
mitogen-activated protein kinase (p42)	41276	7	2	P63085	CAA41548
diacylglycerol kinase, gamma	88524	12	3	Q91WG7	NP_619591
Dystroglycan	70564	19	4	Q62165	CAA60031
pallidin	19683	3	1	Q9R0C0	NP_062762
pyruvate carboxylase	129686	31	7	Q05920	NP_032823
phospholipase D1	119589	31	7	Q9Z280	AAC84041