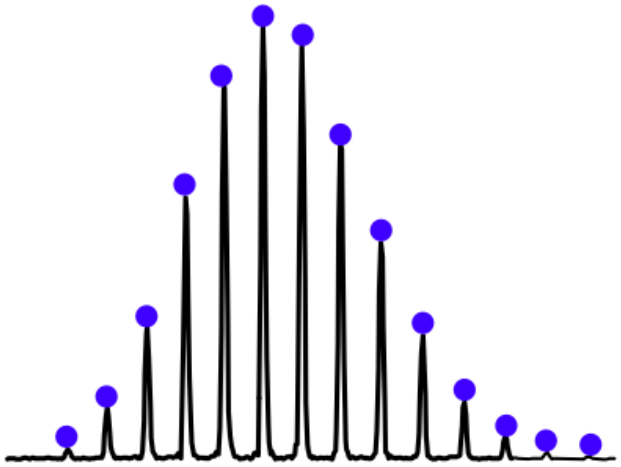
Taking a Top Down Approach to the Human Proteome

NORTHWESTERN UNIVERSITY



CBC

Kelleher Research Team

Northwestern University

The Chicago Biomedical Consortium

AACC Conference, San Jose, CA

April 18th, 2012

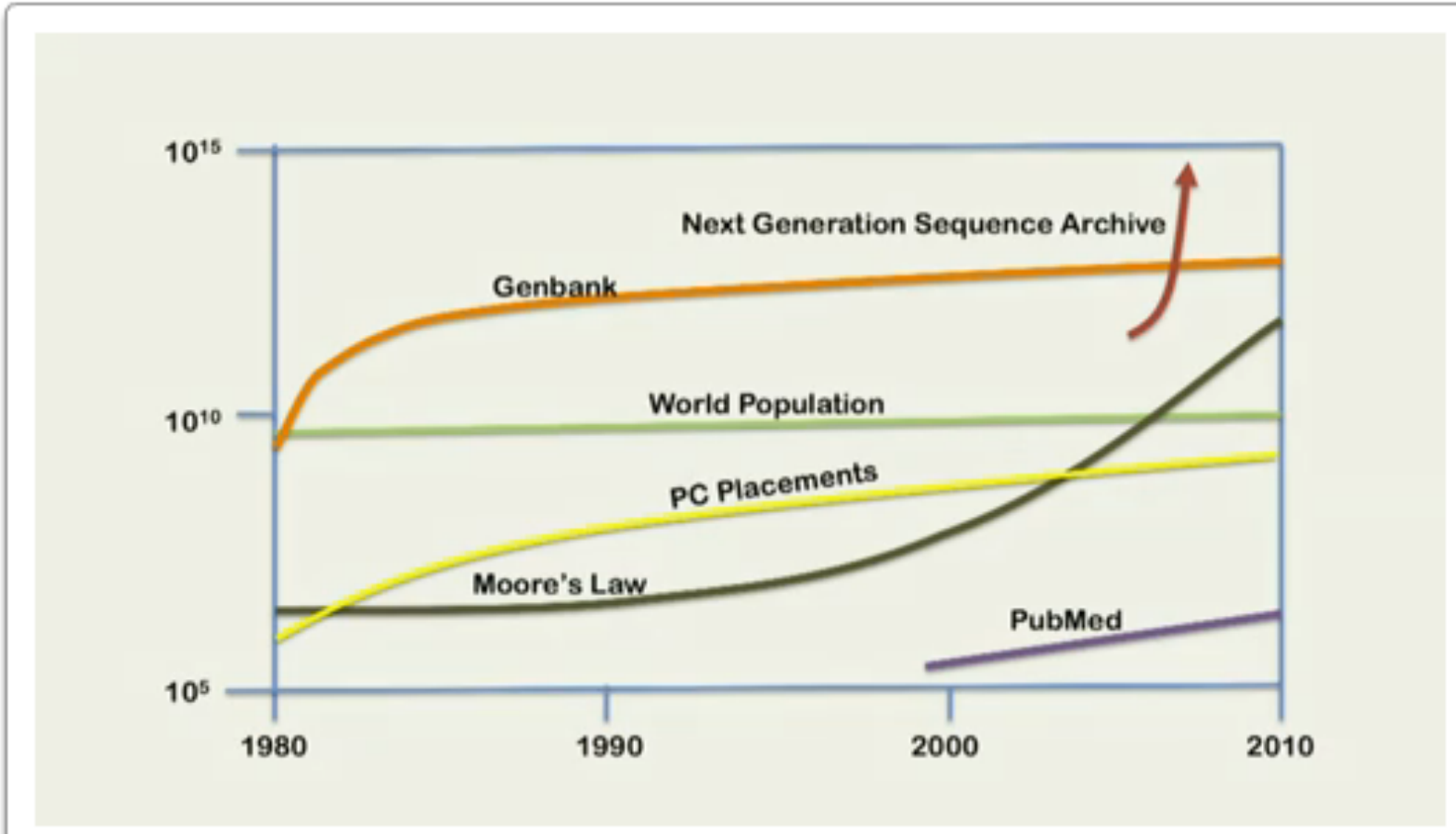
Executive Summary

- The Genomics Revolution: A Perspective
- Proteomics and the Age of Protein Isoforms
  - The Evolution of Top Down Technology (Brief)
  - Top Down Proteomics in High Throughput Mode
- Formation of a Consortium for Top Down Proteomics
- On The Role of Top Down MS For Translational Proteomics

The pace of development in genomics is breathtaking

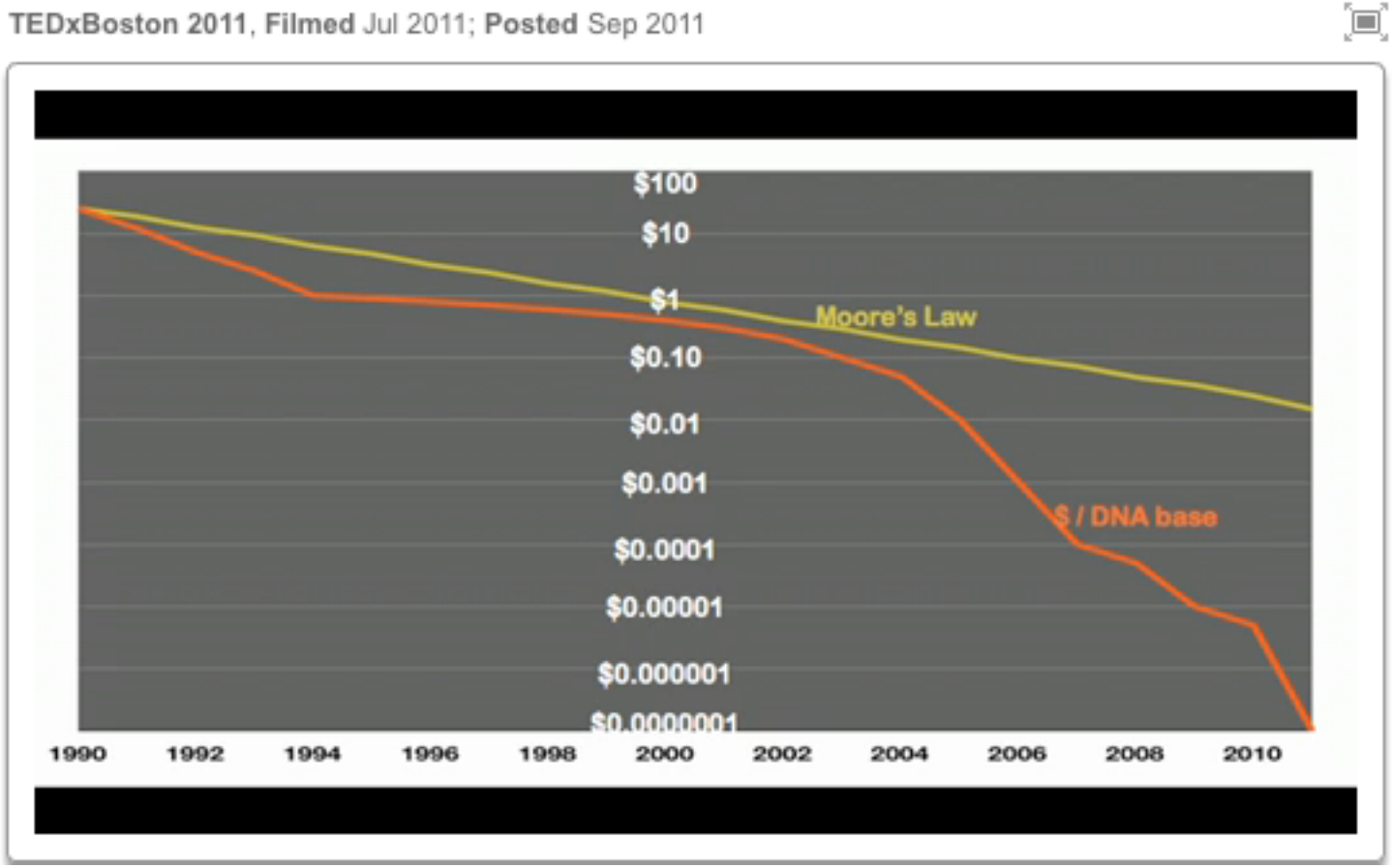
Richard Resnick: Welcome to the genomic revolution

TEDxBoston 2011, Filmed Jul 2011; Posted Sep 2011



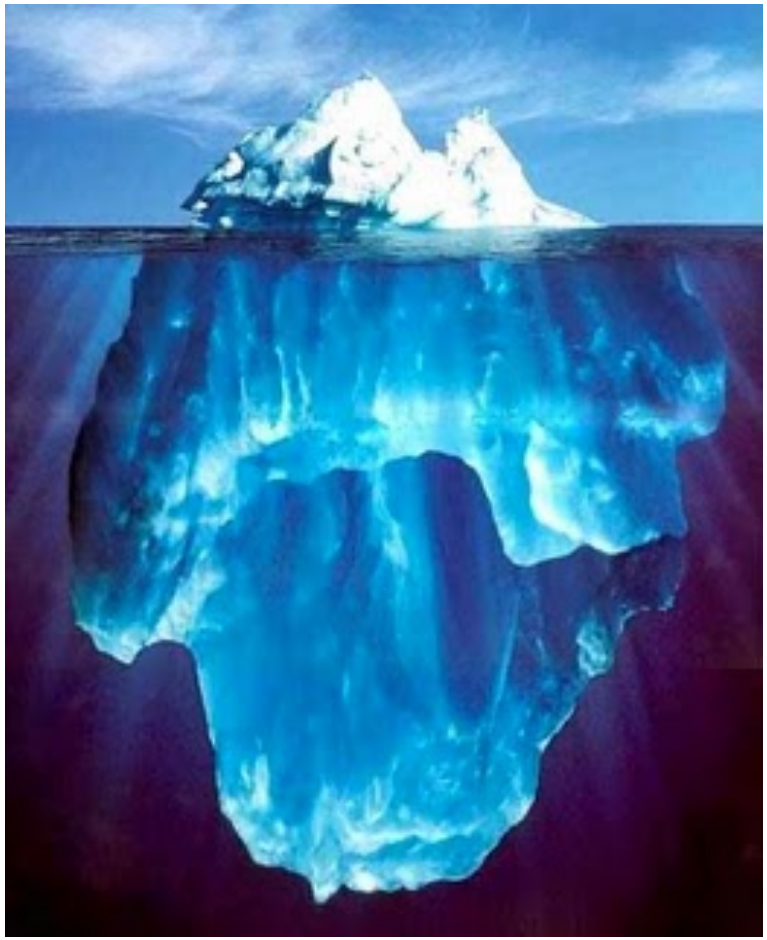
The pace of development in genomics is breathtaking

TEDxBoston 2011, Filmed Jul 2011; Posted Sep 2011

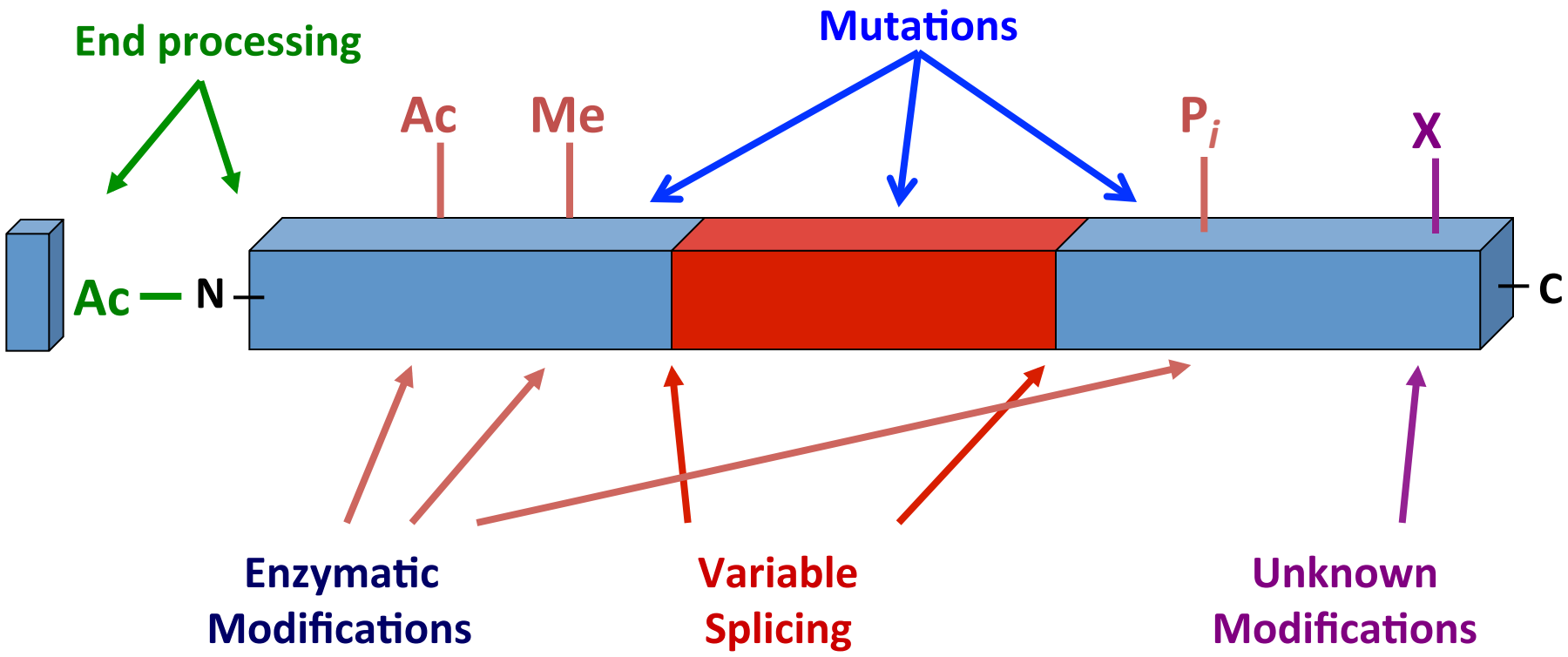


From a TedX Seminar - Richard Resnick

Abundance range of protein molecules spans ~1 million fold, and there is no way to amplify them



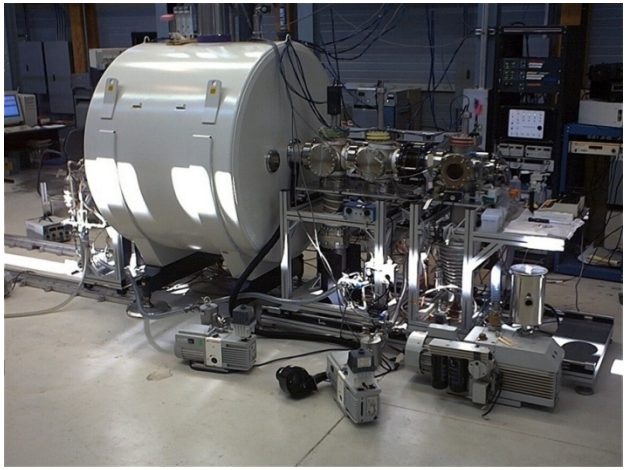
Origins of Complexity in the Human Proteome:
The Age of Protein Isoforms



Key Concept: sources of protein variability result in a large, but finite number of protein forms (~250,000), resulting in a measurement challenge.

Progression of “Top Down” and FT Mass Spectrometry

1999 → 2011



...

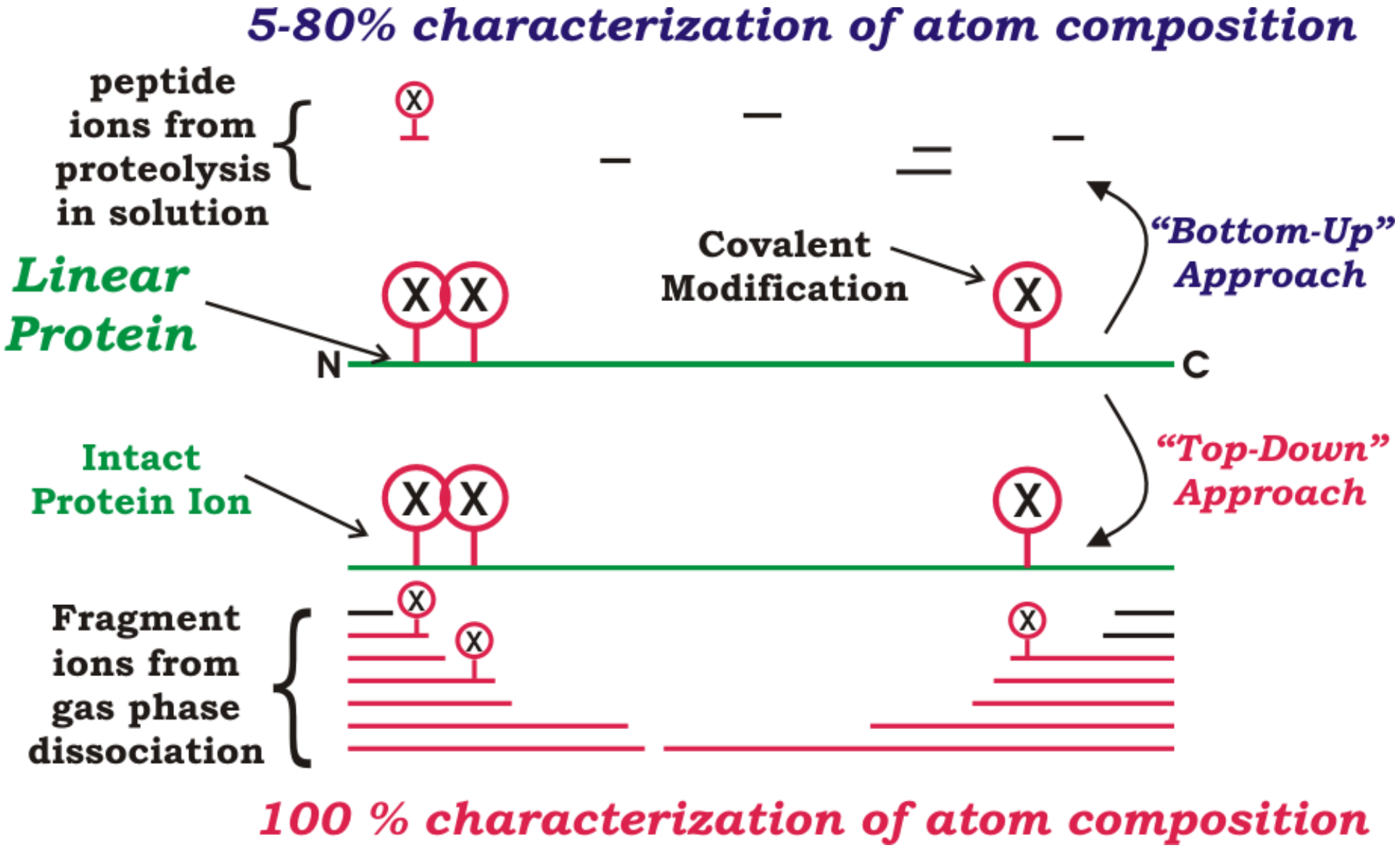
General Strategies of Protein Analysis by Mass Spectrometry

- Top Down (intact protein)

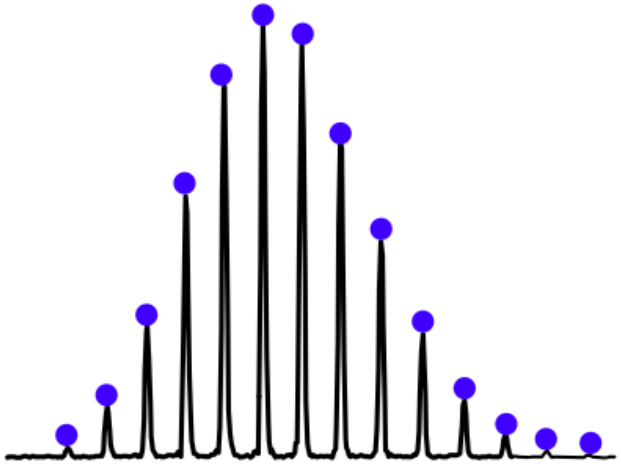
- Middle Down (endogenous biomarkers, larger peptides >3 kDa)

- Bottom Up (tryptic peptides; classical)

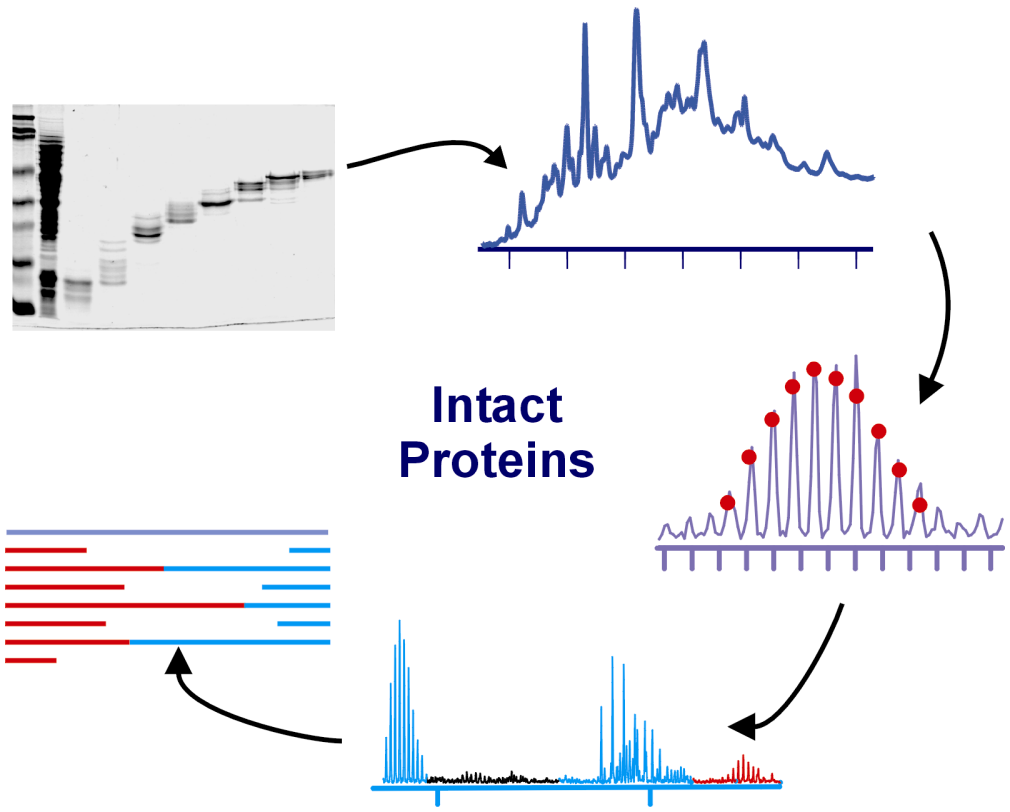
Top Down vs. Bottom Up Protein Characterization



Evolution of Technology Development for Top Down Proteomics

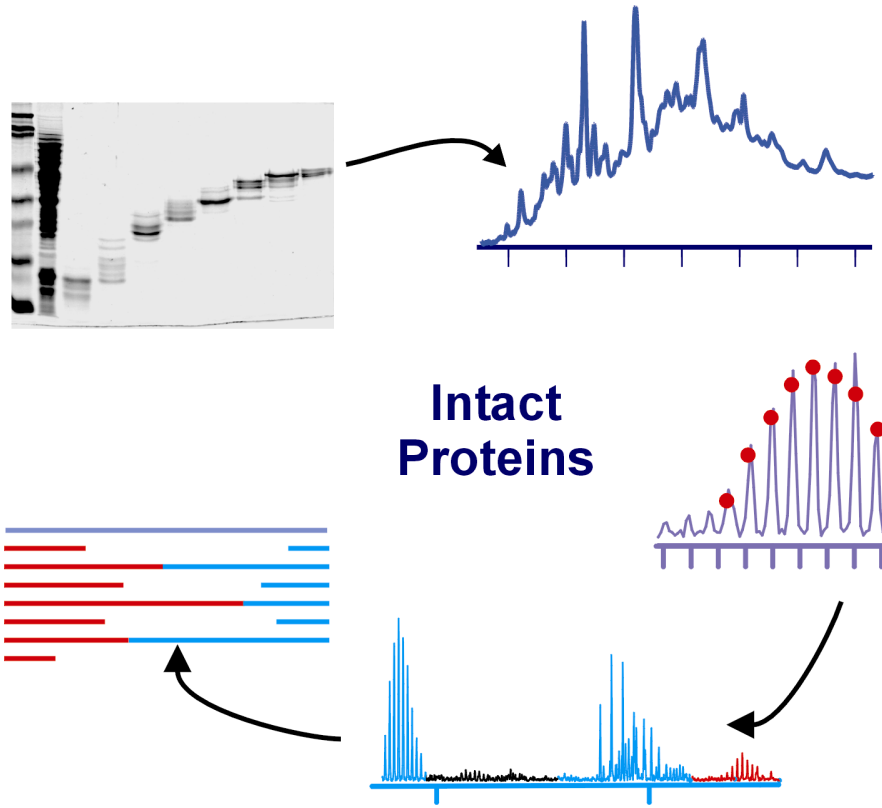


Development of A General Measurement Platform for Intact, Modified Proteins



Development of A General Measurement Platform for Intact, Modified Proteins

1. "Front End" Separations



2. Automated Instrumentation

3. "Back End" Data Processing and Informatics

Chemistry of Life Processes Institute

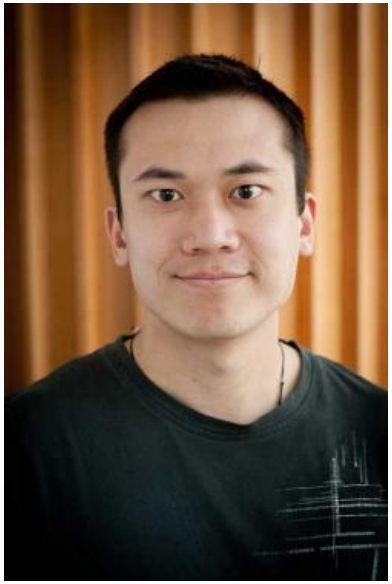
Science Without Boundaries
Discoveries Without Limits

Silverman Hall @ Northwestern

Kelleher Research Group

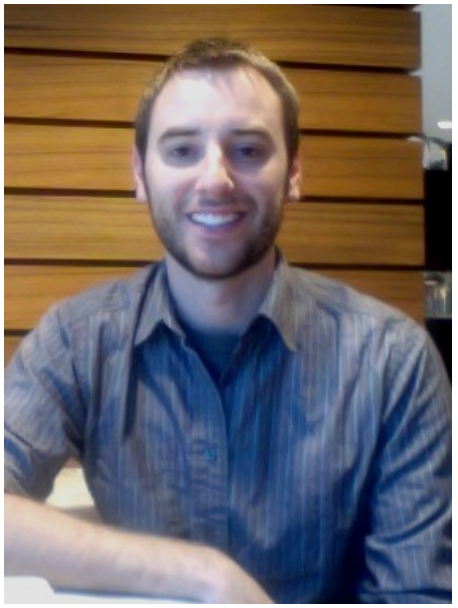
Proteomics Center @ Northwestern

Dr. Paul Thomas,
Associate Director



Dr. John Tran,
Director of Separations Development

Dr. Phil Compton,
Director of Instrumentation



Ryan Fellers,
Senior Software Engineer

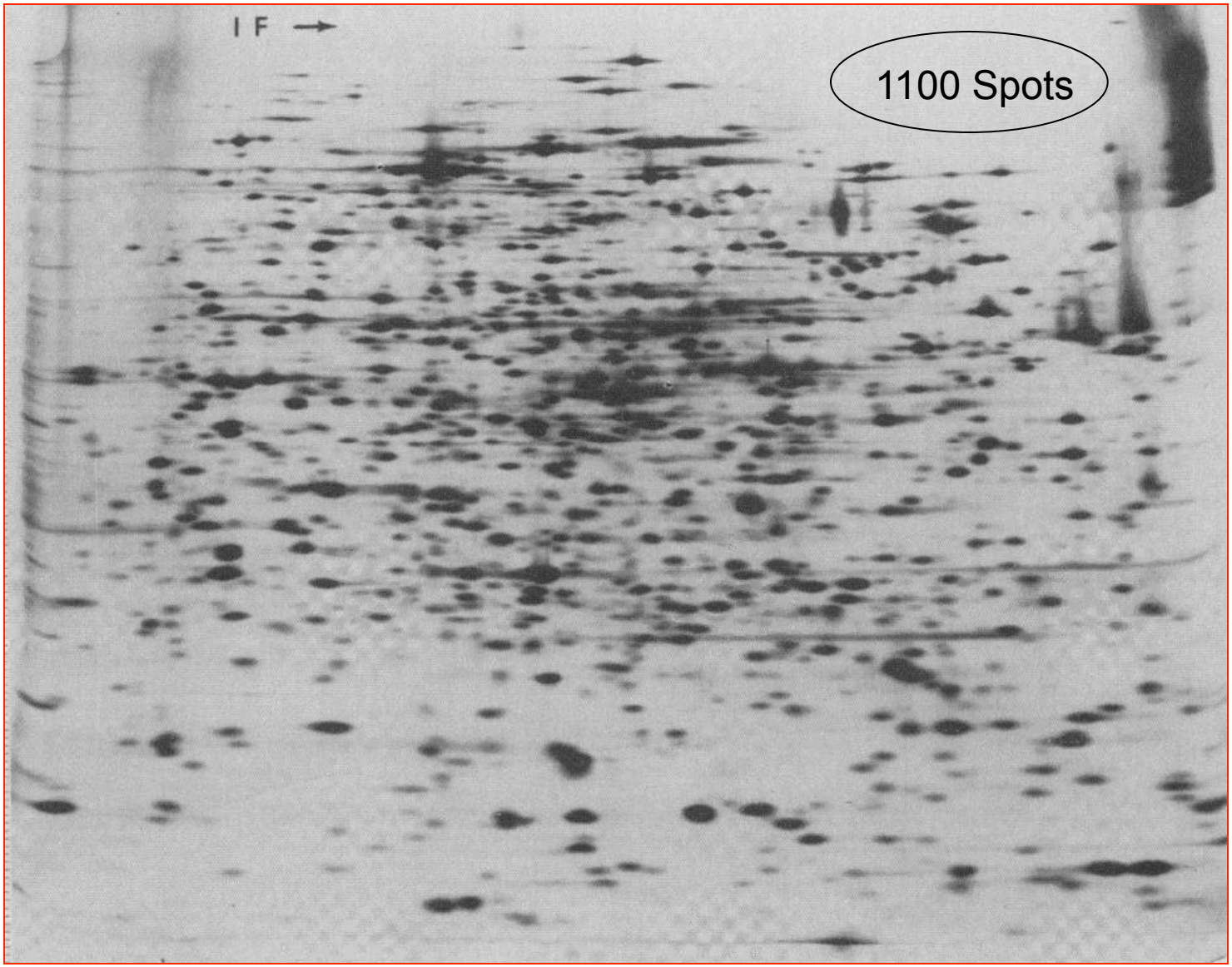
Mapping Intact Protein Isoforms in Discovery Mode Using Top Down Proteomics

John C. Tran, Leonid Zamdborg, Ji Eun Lee, Adam D. Catherman, Kenneth R. Durbin, Dorothy R. Ahlf, Jeremiah D. Tipton, Adaikkalam Vellaichamy, John F. Kellie, Mingxi Li, Cong Wu, Steve M. Sweet, Bryan P. Early, Nertila Siuti, Philip D. Compton, Paul M. Thomas, and Neil L. Kelleher

Published on-line in *Nature,* October 31st, 2011

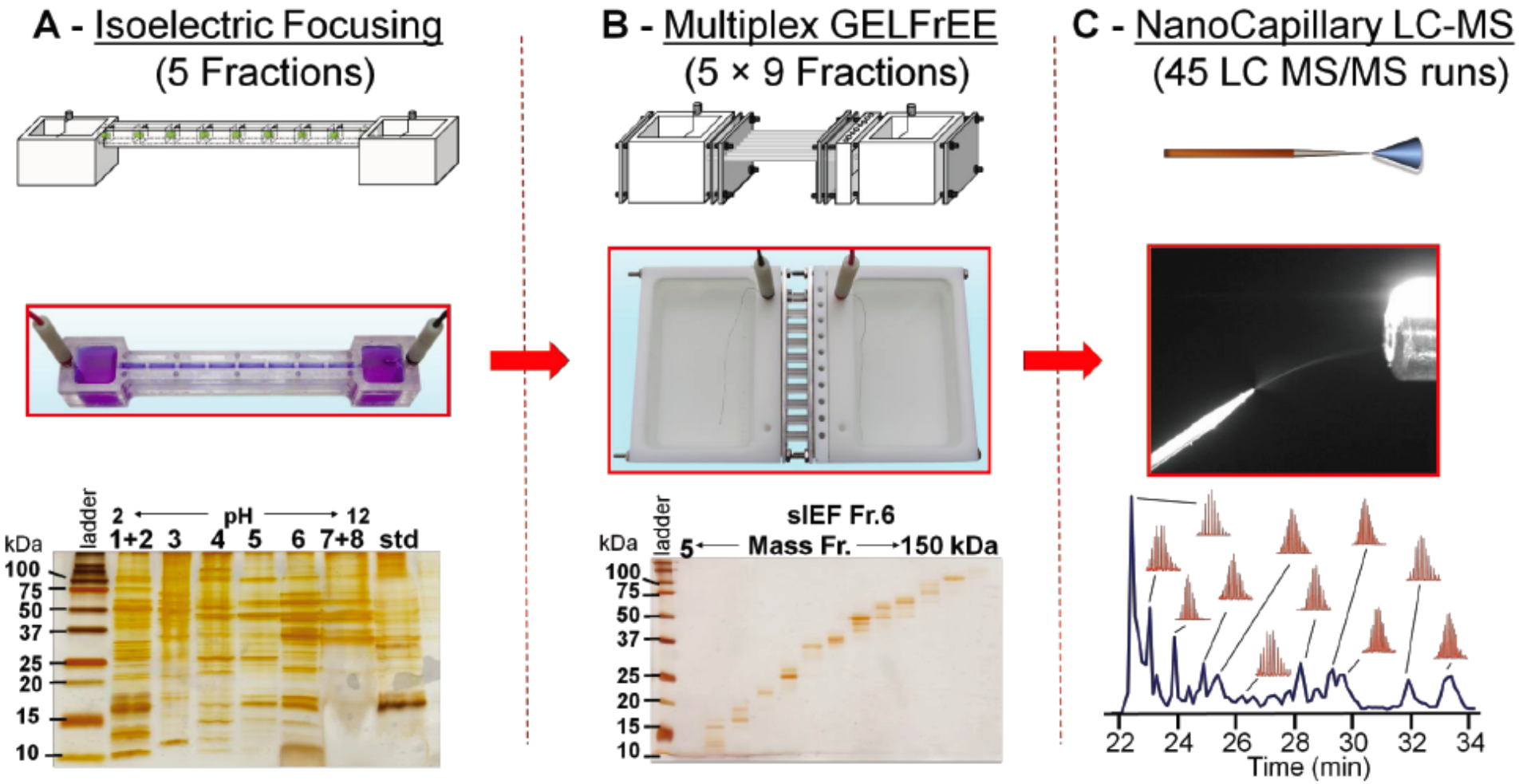
NORTHWESTERN
UNIVERSITY

High Resolution Separation of Intact Proteins



O' Farrell, P. H. *J. Biol. Chem.* 1975, *250*, 4007-4021.

4-Dimensional Fractionation of Intact Proteins (0.5-1 mg Loading)



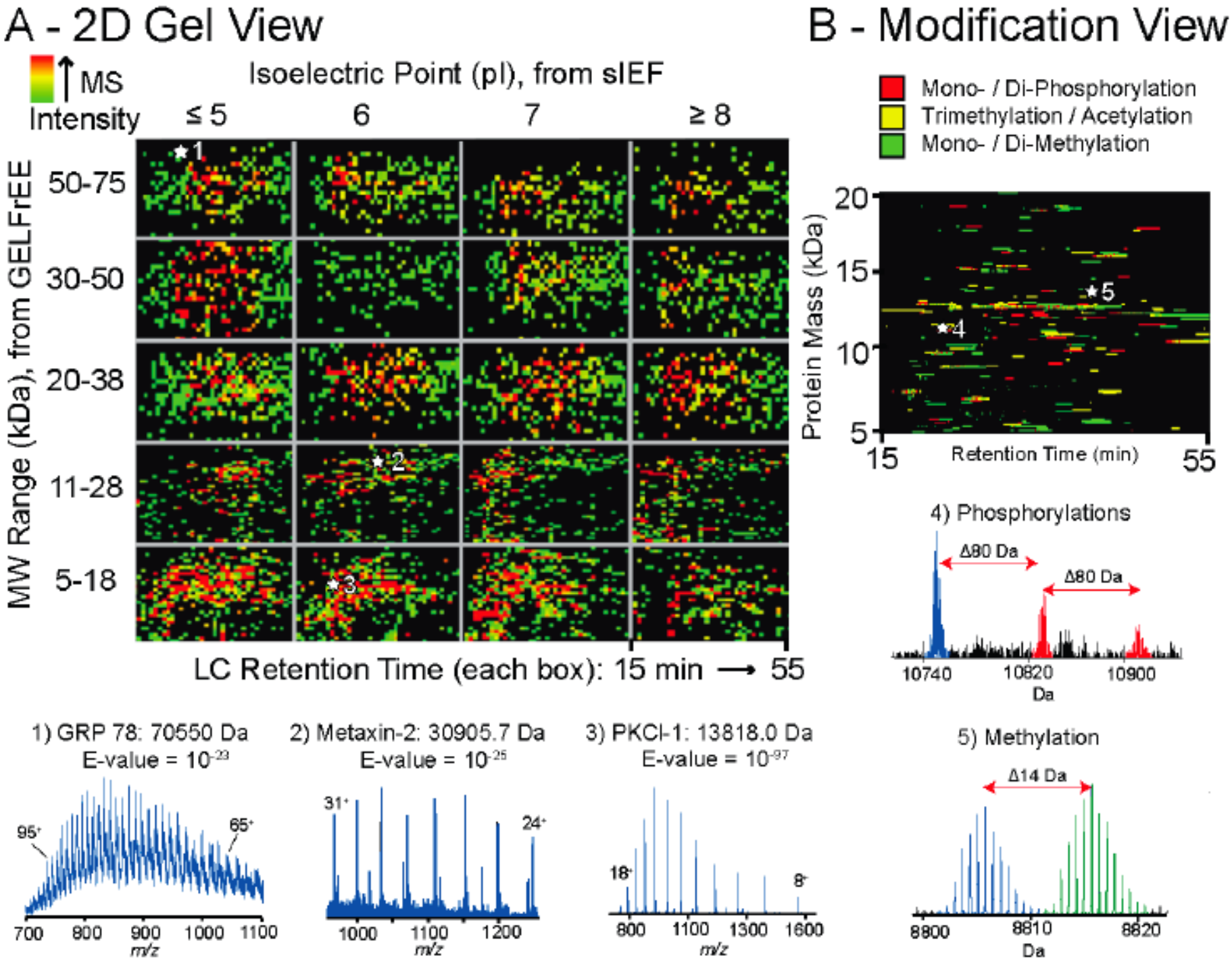
Mapping intact protein isoforms in discovery mode using top-down proteomics
Tran, J. *et al*., *Nature.* **2011**, 480, 254–258.

Instruments Used for Top Down Proteomics

Orbitrap Elite

Custom 12T LTQ-FT Ultra

- Protein Identification
- Protein Characterization
- Quantitative Top Down

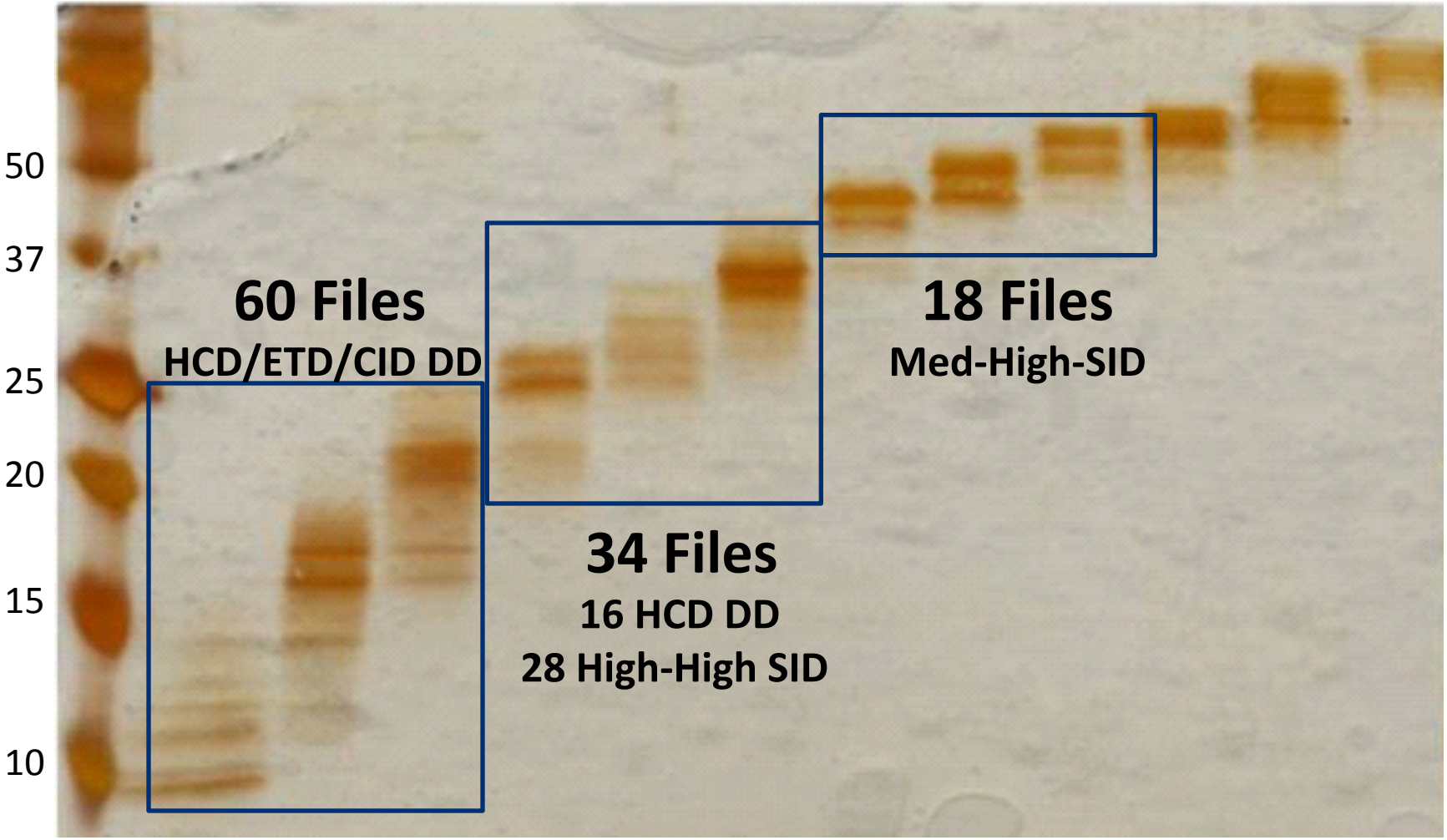

A - 2D Gel View
MS Intensity
Isoelectric Point (pI), from sIEF
≤ 5
6
7
≥ 8
MW Range (kDa), from GELFrEE
50-75
30-50
20-38
11-28
5-18
LC Retention Time (each box): 15 min → 55
B - Modification View
Mono- / Di-Phosphorylation
Trimethylation / Acetylation
Mono- / Di-Methylation
Protein Mass (kDa)
Retention Time (min)
4) Phosphorylations
Δ80 Da
Δ80 Da
Da
1) GRP 78: 70550 Da
E-value = 10-23
m/z
2) Metaxin-2: 30905.7 Da
E-value = 10-25
m/z
3) PKCI-1: 13818.0 Da
E-value = 10-97
m/z
5) Methylation
Δ14 Da
Da

An Example Results Set…

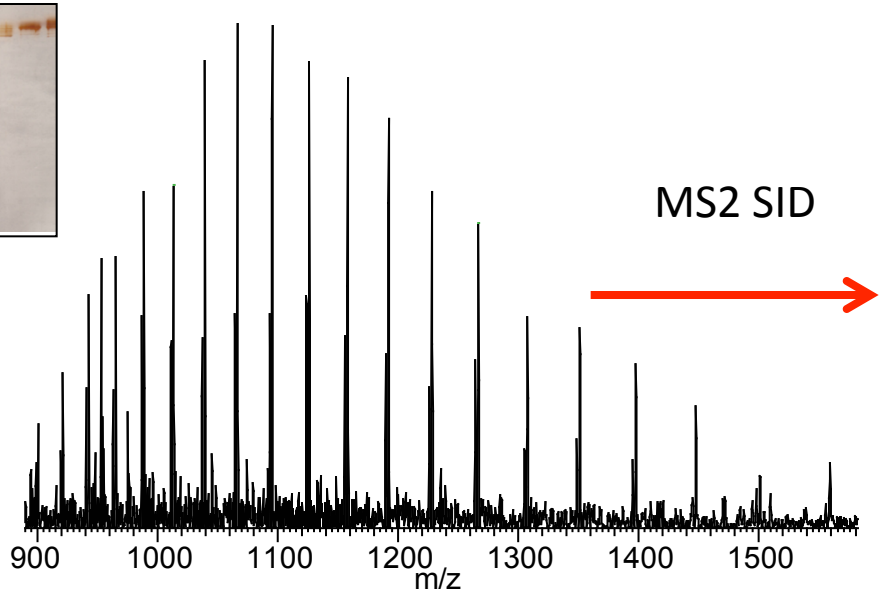
	# unique base sequences	# of isoforms	# of genes	# of transcripts
H2A.O	**1**	**(2)**	**[1]**	**<1>**
H4	**1**	**(14)**	**[1]**	**<1>**
Ubiquitin	**1**	**(2)**	**[4]**	**<4>**
GAL-P	**1**	**(3)**	**[1]**	**<2>**
HMGA1	**1**	**(10)**	**[1]**	**<2>**
	● ● ●	● ● ●	● ● ●	● ● ●
Σ	**1500**	**(4000)**	**[1550]**	**<1700>**

Simplified Data Acquisition and Processing for Top Down: *Reduction to Practice*

“Gelfree” Separation of H1299 Cell Extract

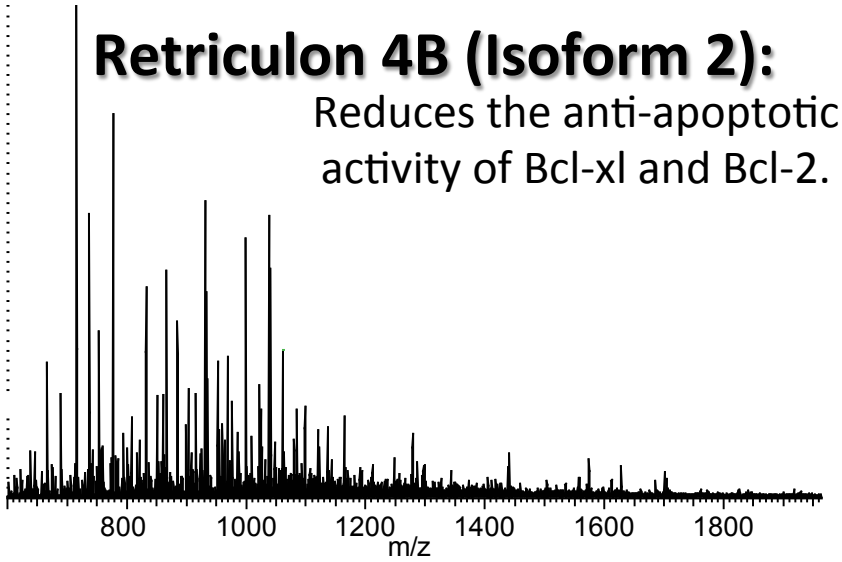


Example Identification at 40 kDa (FT/FT)



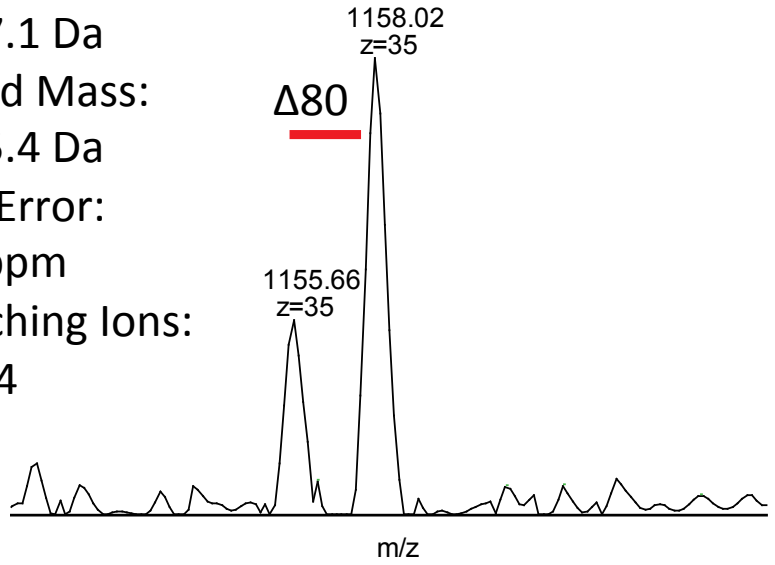
Retriculon 4B (Isoform 2):

Reduces the anti-apoptotic activity of Bcl-xl and Bcl-2.

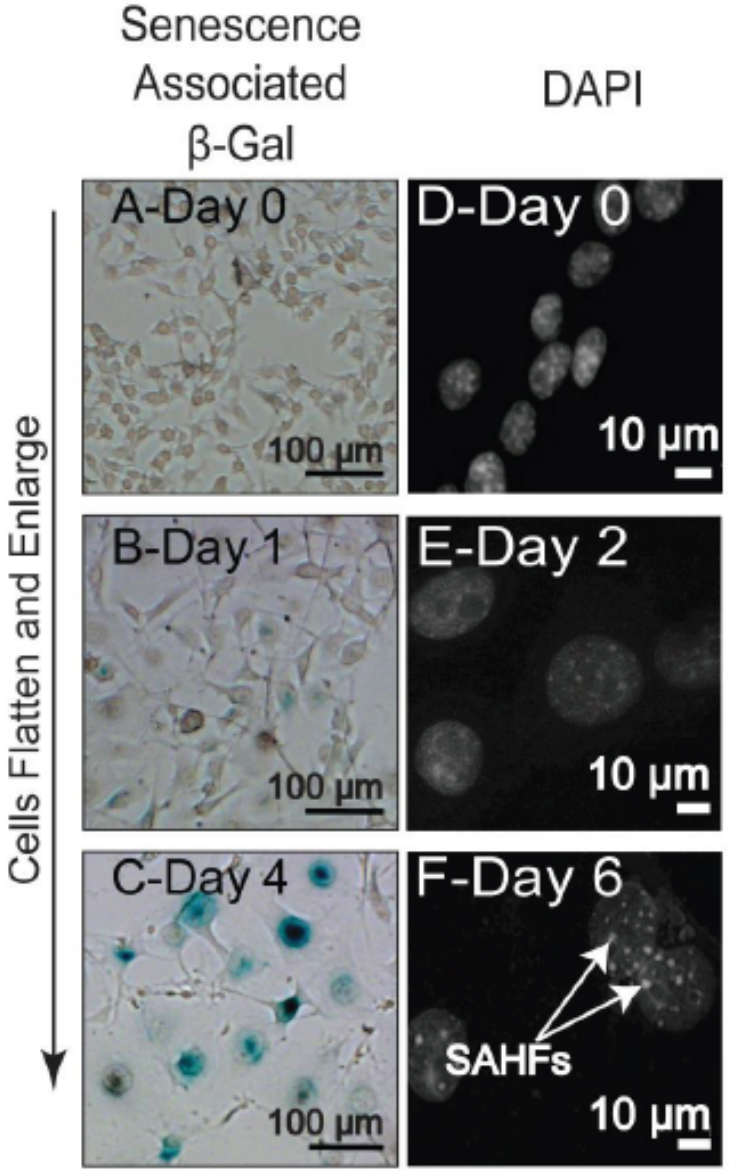


M-E-D-L-D-Q-S-P-L-V-S-S-S-D-S-P-P-R-P-Q-P-A-F-K-Y-Q-F-V-R-E-P-
E-D-E-E-E-E-E-E-E-E-E-E-D-E-D-E-D-L-E-E-L-E-V-L-E-R-K-P-A-A-G-L
-S-A-A-P-V-P-T-A-P-A-A-G-A-P-L-M-D-F-G-N-D-F-V-P-P-A-P-R-G-P-L-
P-A-A-P-P-V-A-P-E-R-Q-P-S-W-D-P-S-P-V-S-S-T-V-P-A-P-S-P-L-S-A-A
-A-V-S-P-S-K-L-P-E-D-D-E-P-P-A-R-P-P-P-P-P-P-A-S-V-S-P-Q-A-E-P-
V-W-T-P-P-A-P-A-P-A-A-P-P-S-T-P-A-A-P-K-R-R-G-S-S-G-S-V-V-V-D-L
-L-Y-W-R-D-I-K-K-T-G-V-V-F-G-A-S-L-F-L-L-L-S-L-T-V-F-S-I-V-S-V-
T-A-Y-I-A-L-A-L-L-S-V-T-I-S-F-R-I-Y-K-G-V-I-Q-A-I-Q-K-S-D-E-G-H
-P-F-R-A-Y-L-E-S-E-V-A-I-S-E-E-L-V-Q-K-Y-S-N-S-A-L-G-H-V-N-C-T-
I-K-E-L-R-R-L-F-L-V-D-D-L-V-D-S-L-K-F-A-V-L-M-W-V-F-T-Y-V-G-A-L
-F-N-G-L-T-L-L-I-L-A-L-I-S-L-F-S-V-P-V-I-Y-E-R-H-Q-A-Q-I-D-H-Y-
L-G-L-A-N-K-N-V-K-D-A-M-A-K-I-Q-A-K-I-P-G-L-K-R-K-A-E-

Theoretical Mass:
40497.1 Da
Observed Mass:
40495.4 Da
Mass Error:
42 ppm
Total Matching Ions:
34

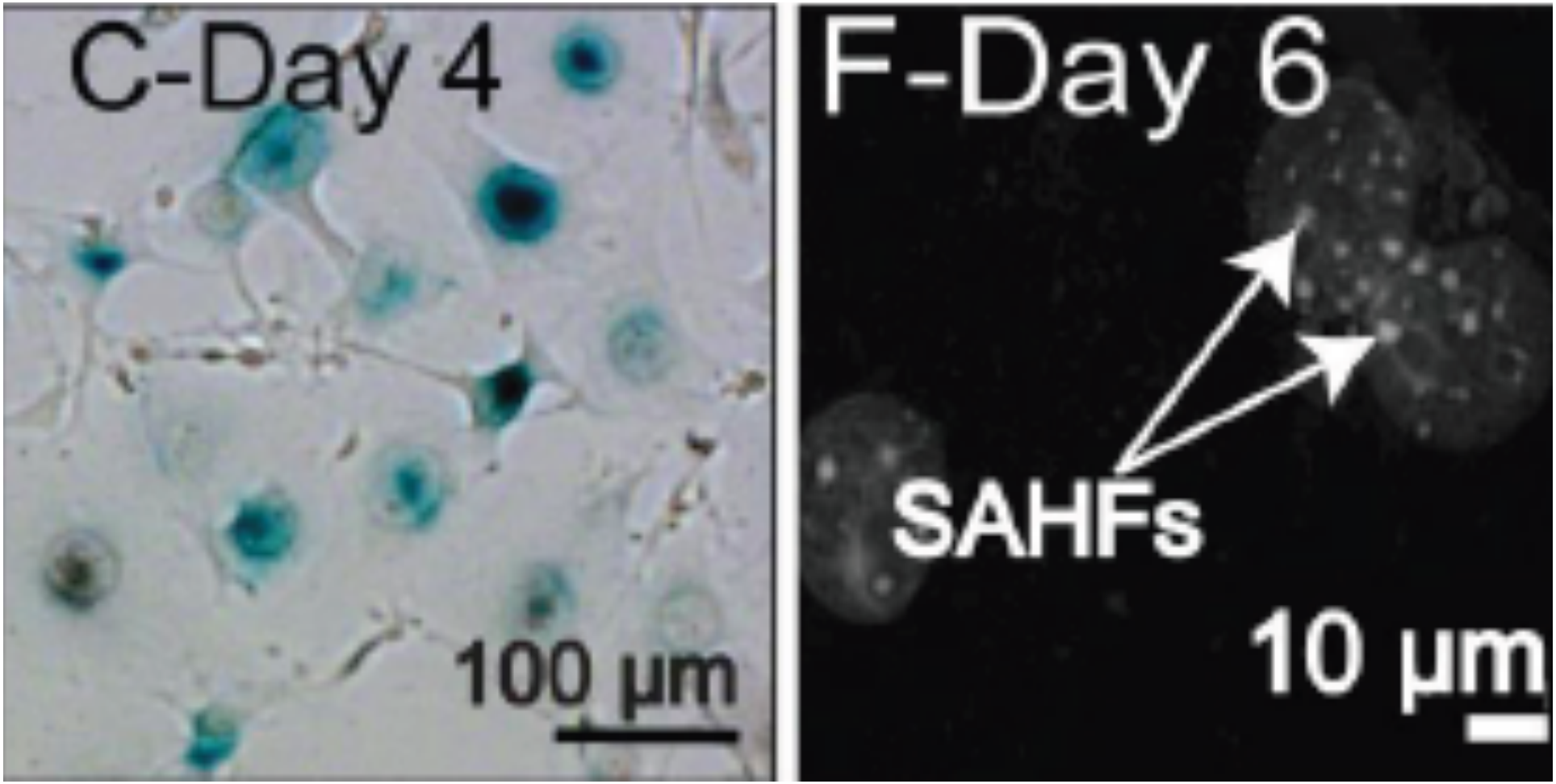


Stress-Induced Senescence in B16F10 Melanoma Cells



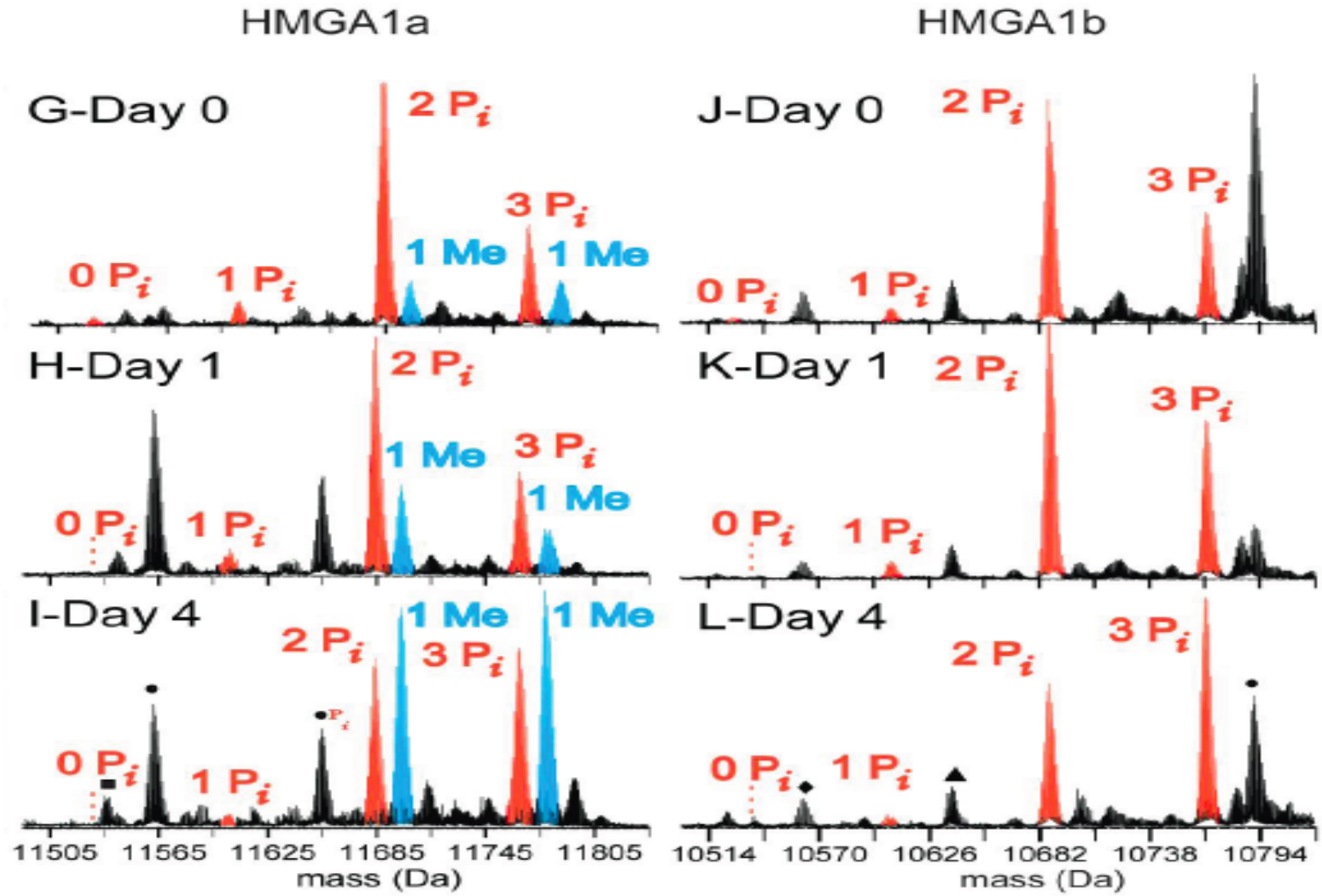
Ji Eun Lee

Chemotherapy-Induced Senescence in B16F10 Melanoma Cells



Ji Eun Lee

A Specific Example of the Isoform Challenge



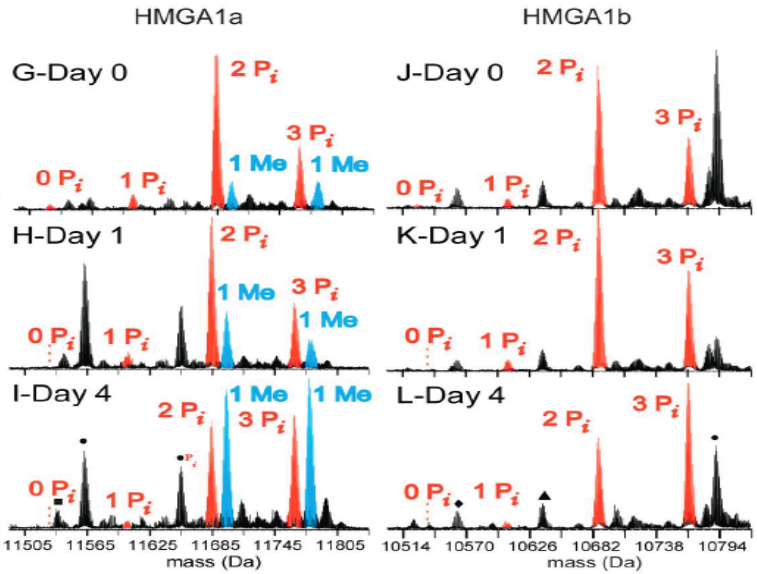
Ji Eun Lee

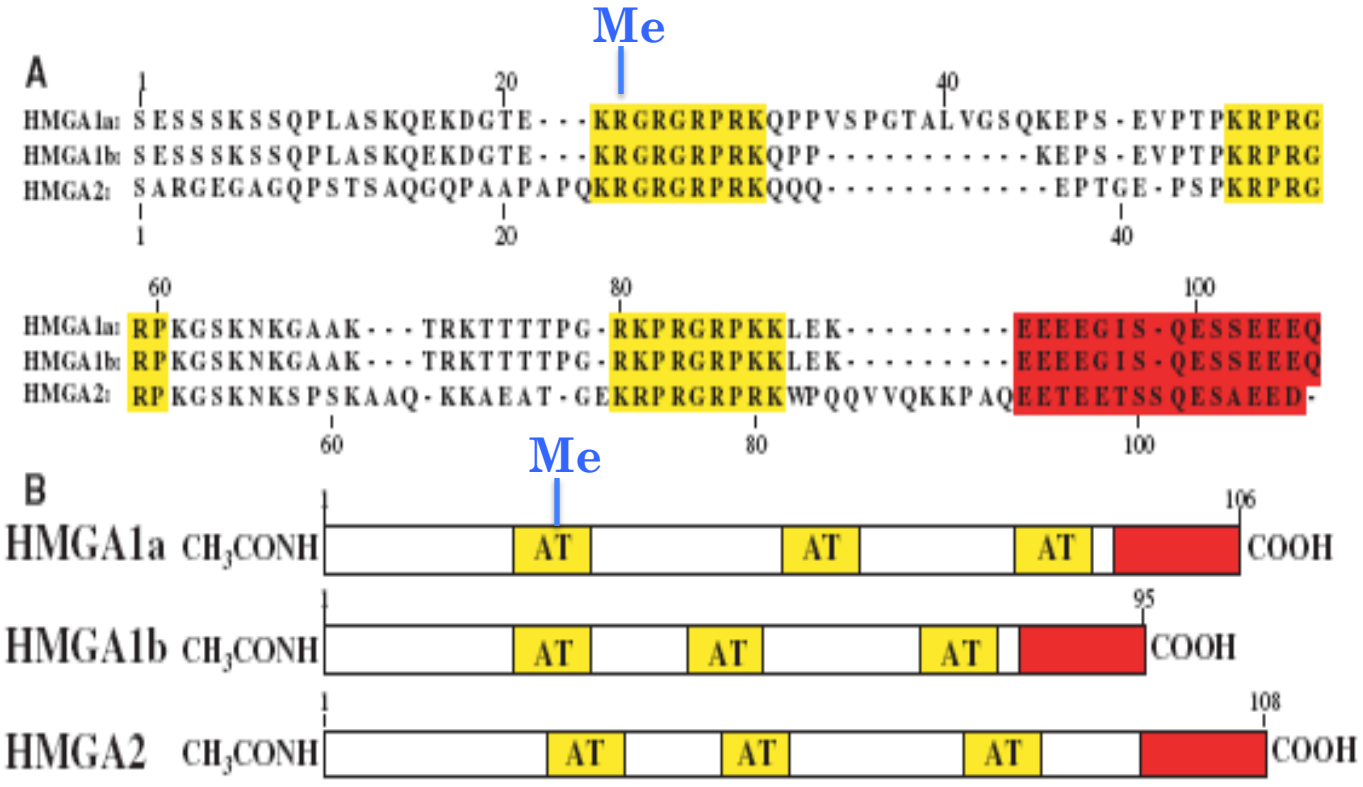
HMGA Family Members

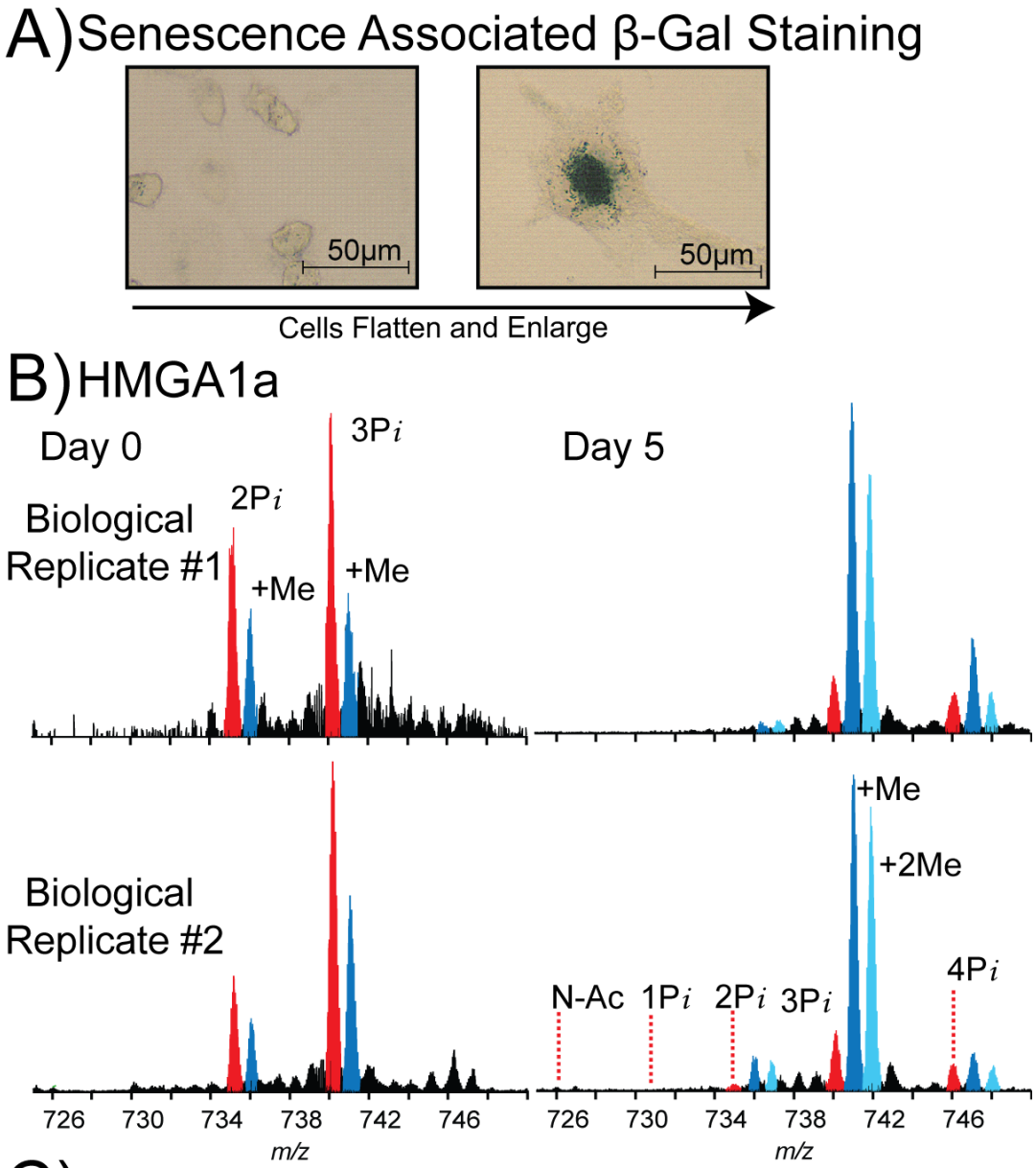
HMGA1a is encoded by gene HMG1

HMGA1b is a result of alternative splicing of HMG1 mRNA

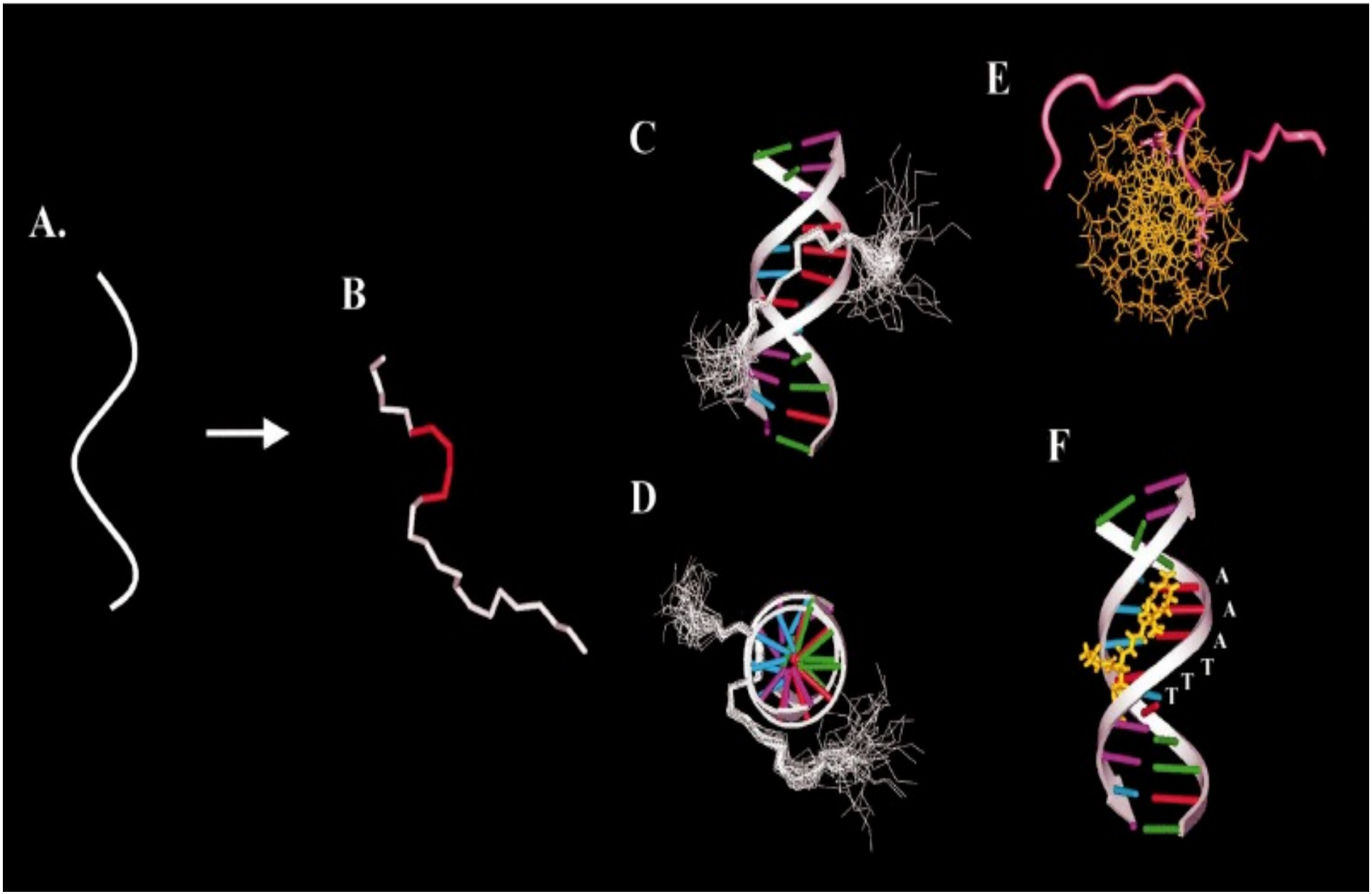
HMG2 gene encodes for the HMGA2 protein






A) Senescence Associated β-Gal Staining
50μm
50μm
Cells Flatten and Enlarge
B) HMGA1a
Day 0
Day 5
Biological Replicate #1
Biological Replicate #2
3Pi
2Pi
+Me
+Me
+Me
+2Me
4Pi
N-Ac
1Pi
2Pi
3Pi
726 730 734 738 742 746
m/z
726 730 734 738 742 746
m/z
C) Graphical Fragment Map of HMGA1a
S-E-S-S-S-K-S-S-Q-P-L-A-S-K-Q-E-K-D-G-T-E-K-R-G-R-G-R-P-R-K-Q-P-P-V-S-P-G-T-A-L
-V-G-S-Q-K-E-P-S-E-V-P-T-P-K-R-P-R-G-R-P-K-G-S-K-N-K-G-A-A-K-T-R-K-T-T-T-T-P-G-
R-K-P-R-G-R-P-K-K-L-E-K-E-E-E-E-G-I-S-Q-E-S-S-E-E-E-Q-

Structure of the second and third DNA binding domains of HMGA1a in complex with a synthetic duplex oligonucleotide



Conclusions

- High-throughput Top Down Proteomics is becoming a reality. 5-50 kDa now quite routine.

- 50-200 kDa under development and will be accessed in the next few years. Targeted analysis in this size regime now readily feasible.

- Proteomics will become a stronger complement to genomics in the new age of molecular medicine.