Introduction to Mass Spectrometry Based Proteomics

Christopher Mason
Mayo Proteomics Research Center

What is Proteomics?

- A *proteome* is the set of all proteins expressed by an organism, organ, tissue, etc, at a given time.
- Parallel to Genome = all genes expressed in an organism, etc...
- One goal: Find a protein or proteins that indicates disease.

Motivation

- Protein Biomarkers
  - Proteins are most directly responsible for biological function.
  - RNA expression ≠ Protein expression.
  - Post Translational Modifications (phosphorylation, glycoslyation) not observable at DNA/RNA level.
- Many diseases while treatable at early stages are still fatal at advanced stages.

DNA ⟶ RNA ⟶ Protein

Motivation

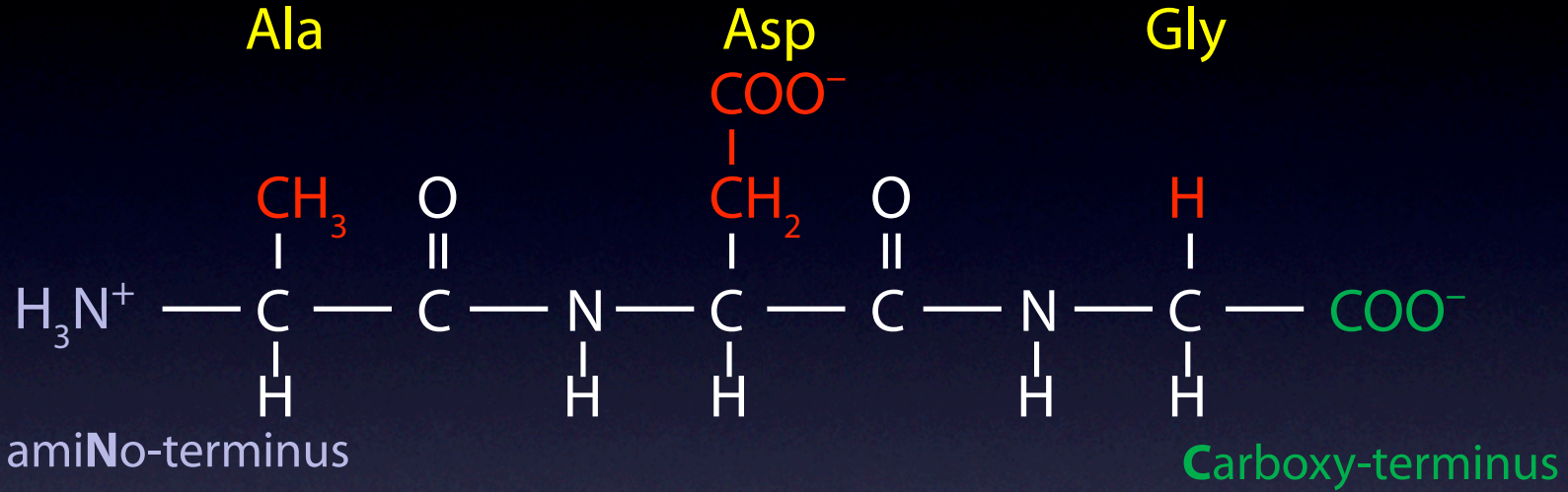
- Protein Biomarkers
  - Proteins are most directly responsible for biological function.
  - RNA expression $\not\approx^{*}$ Protein expression.
  - Post Translational Modifications (phosphorylation, glycoslyation) not observable at DNA/RNA level.
- Many diseases while treatable at early stages are still fatal at advanced stages.

DNA ⟶ RNA ⟶ Protein

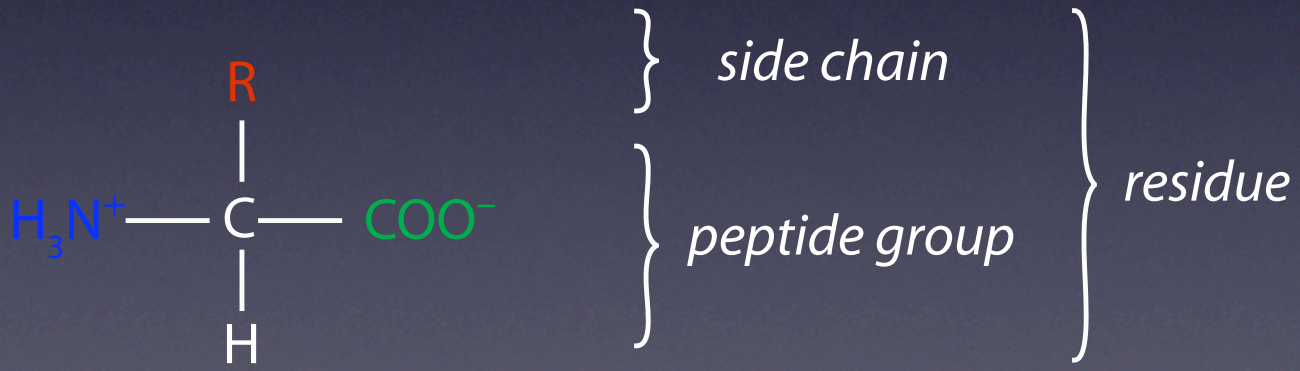
* Not necessarily equals: Anderson, L and Seilhamer J. Electrophoresis. 1997 Mar-Apr;18(3-4):533-7.

What is a protein?

- Proteins are *hetero-polymers*, meaning they are made of repeating chemical units with different composition.



- Repeating unit is an *amino acid*.



20 possible side chains have different chemical properties.

Where might we look for proteins?

Blood

Serum

Plasma

“Proximal” Fluids

Bile

Cerebrospinal Fluid

Synovial Fluid

Feces

Urine

Tissue

Colon

Liver

Pancreas

Kidney

Heart

Prostate

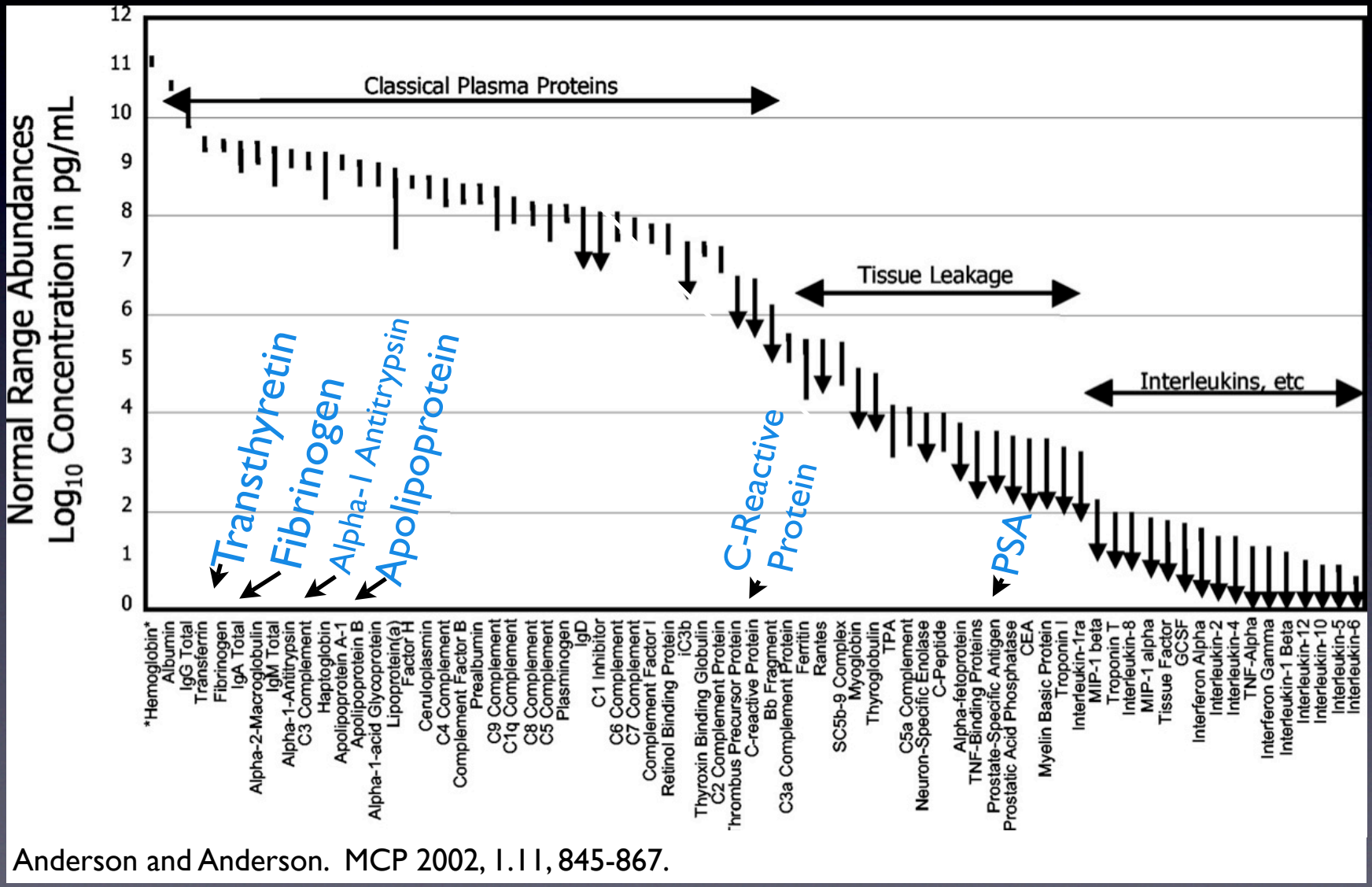
Others...

- Where shouldn’t we look?

Challenges of Blood

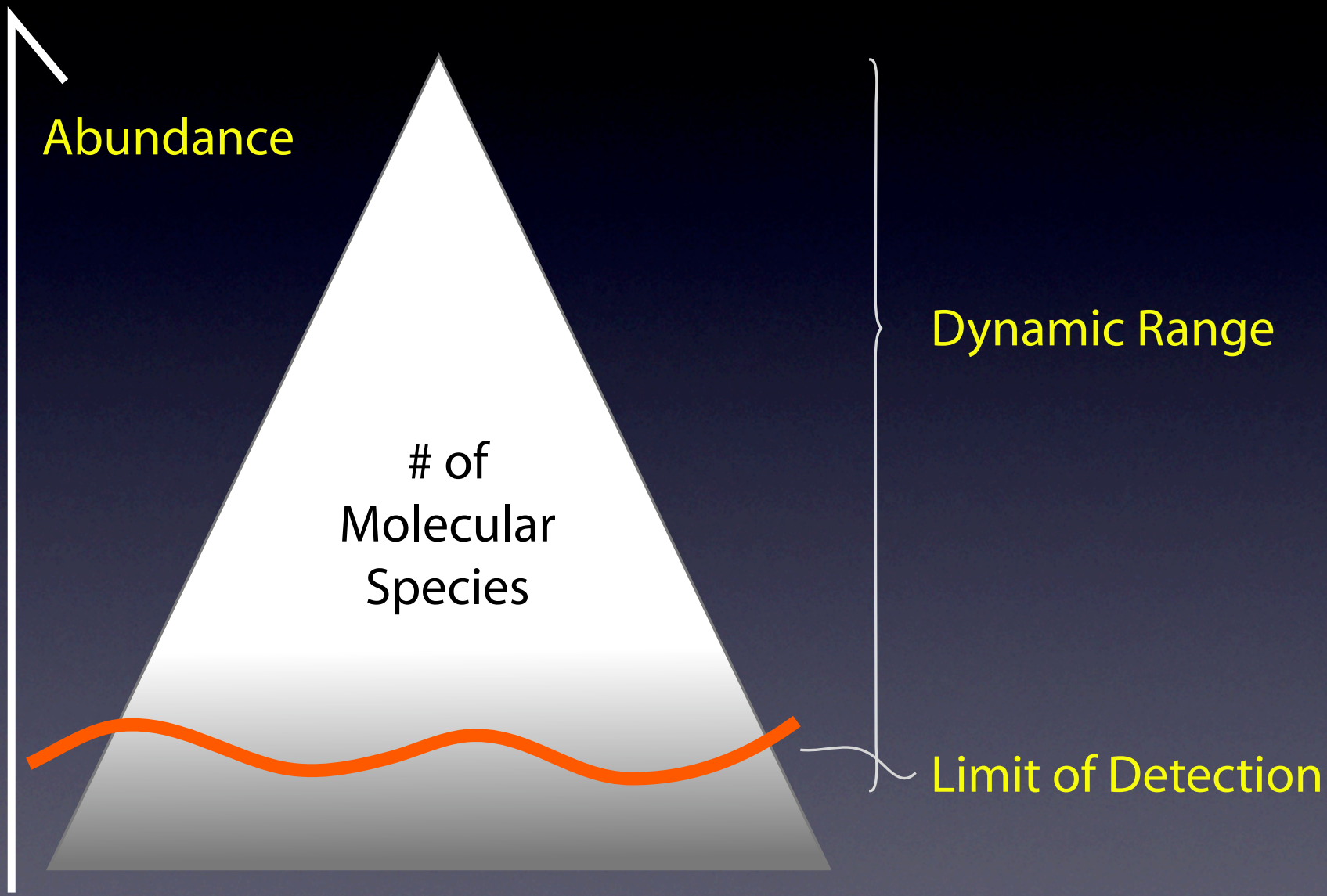
- Millions of molecular forms
- $>10^{10}$ dynamic range
- Extremely diverse chemically
- A few, highly abundant proteins

Dynamic Range of Human Plasma



Anderson and Anderson. MCP 2002, 1.11, 845-867.

Detectability / Sampling

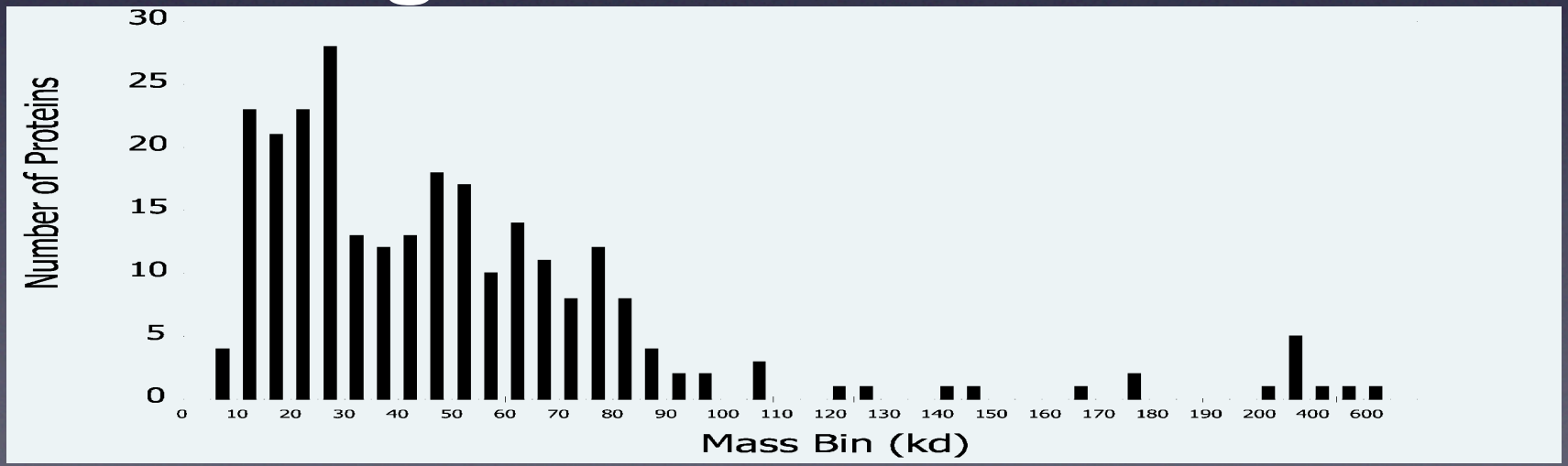


Protein Species in Plasma

"True" Plasma Genes:
~500 genes x 20 glycosylated forms x 5 different "sizes" — 50,000

Tissue leakage: ~35,000 genes x 10 variants (splicing, PTMs, cleavage, etc.) — 350,000

Immunoglobulins — 10,000,000



Anderson and Anderson. MCP 2002, 1.11, 845-867.

What can we do?

- Spread Proteins Out
  - In Time (Time-of-Flight Mass Spec, Chromatography)
  - In Space (2D-Gel)
  - In Frequency (Fourier Transform Mass Spec)
- Throw Proteins Away
  - Filter (Molecular Weight Cut Off)
  - Retain (ELISA)
  - Focus (Glycoproteins)
- Break Proteins Apart
  - Digestion (Trypsin)
  - Fragmentation (MS^n)

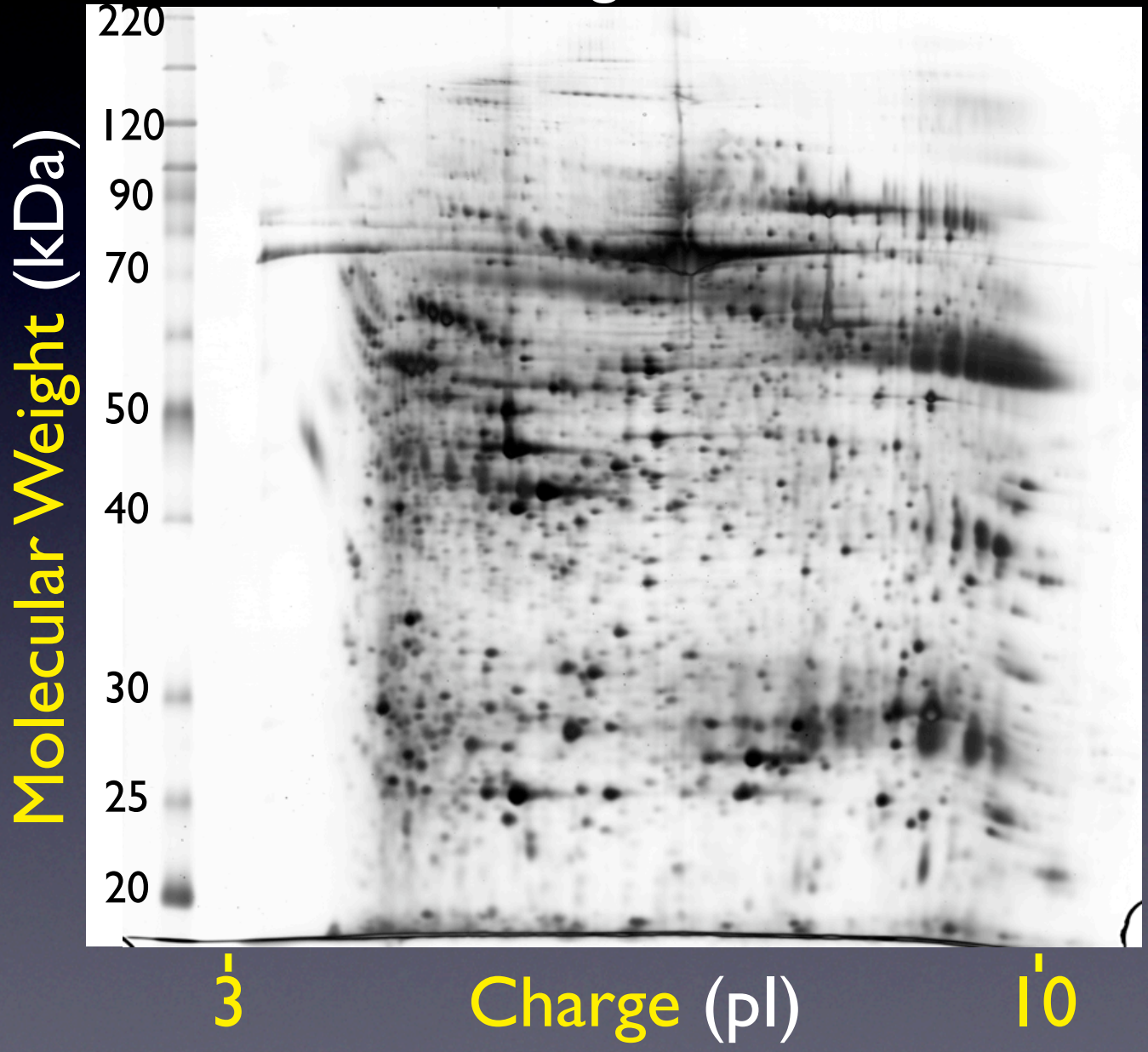
2D Gel

Intact Proteins

Poly-Acrylamide Gel Electrophoresis (PAGE) separates proteins by approx molecular weight.

Iso-Electric Focusing (IEF) separates proteins by charge (pI).

1,268 spots (proteins)



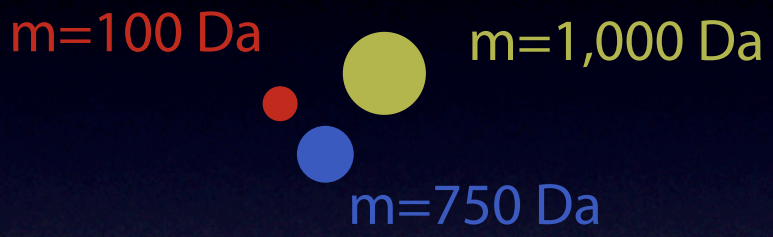
What is a Mass Spectrometer?

A *mass spectrometer* measures the mass-to-charge ratios (m/z) of molecules.

Molecules

What is a Mass Spectrometer?

Molecular Weight

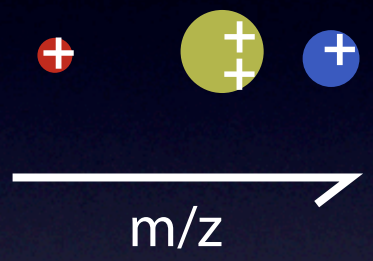


Molecules

What is a Mass Spectrometer?

Apply charge
(Ionization)

What is a Mass Spectrometer?

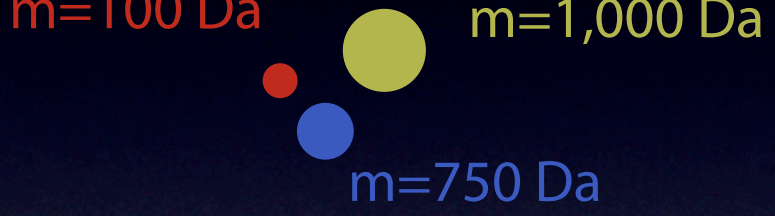


Separate based on
mass-to-charge ratio
(m/z)

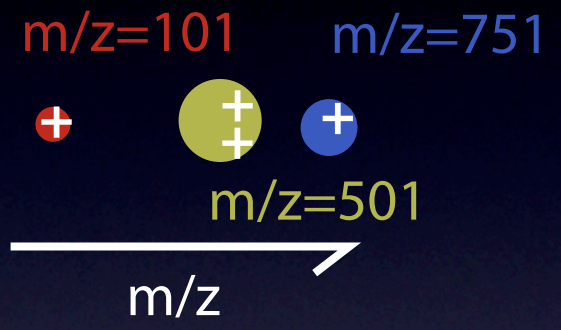
(Mass Analysis)

What is a Mass Spectrometer?

Molecular Weight



Mass-to-Charge

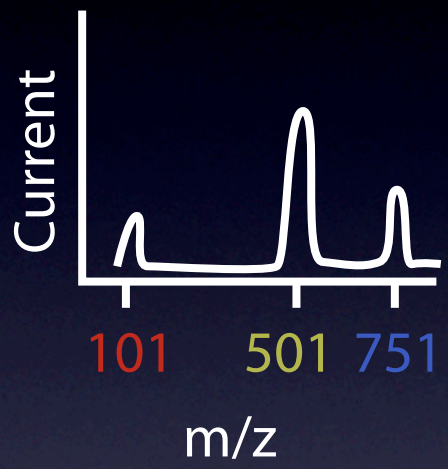


Molecules

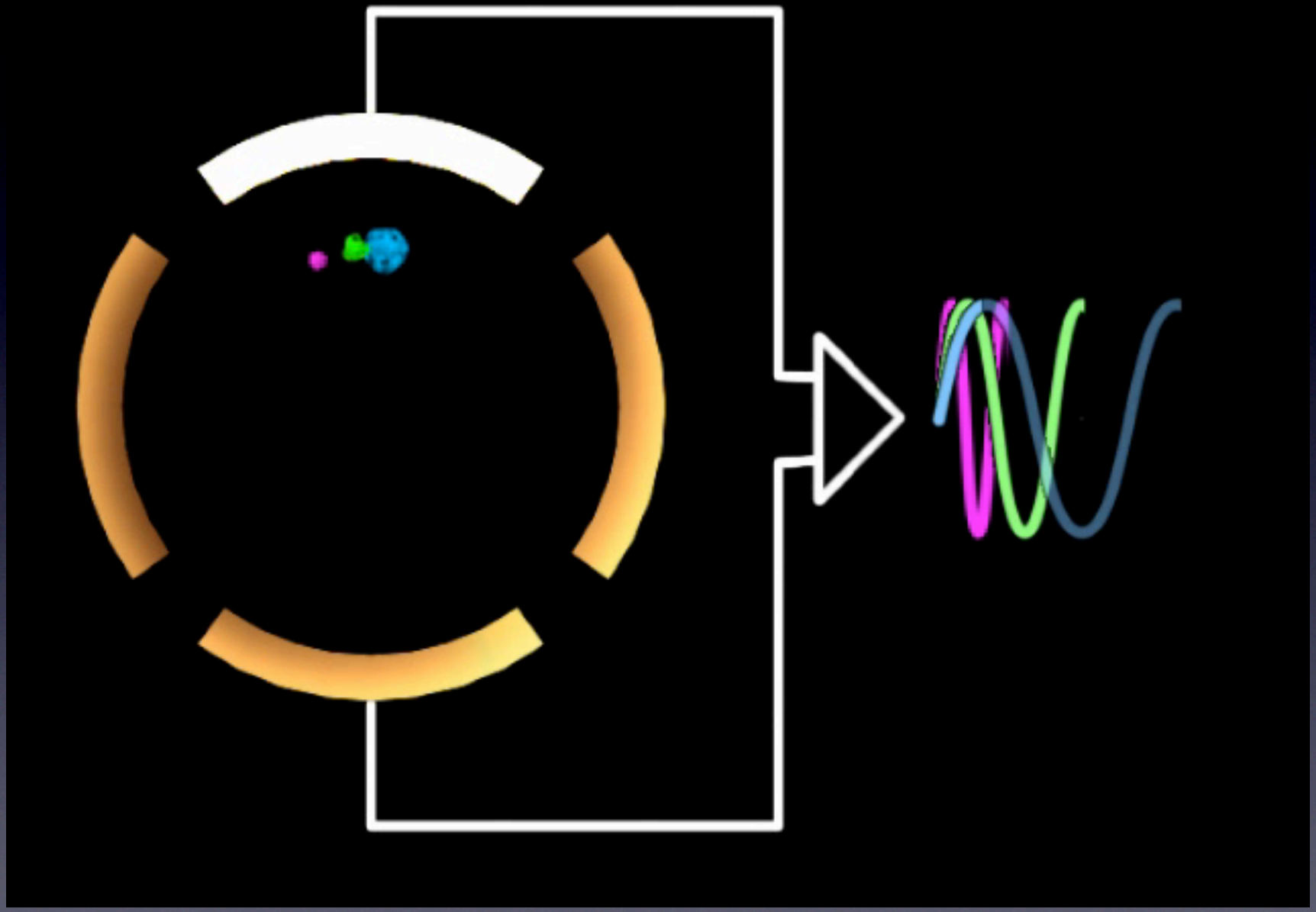
Separate based on mass-to-charge ratio (m/z)

(Mass Analysis)

What is a Mass Spectrometer?



Convert these ions into current that varies in time/ frequency. (Detection)

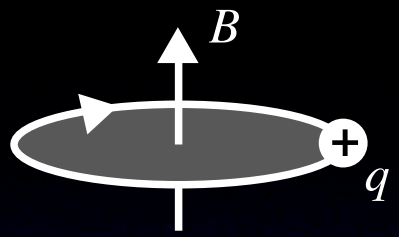


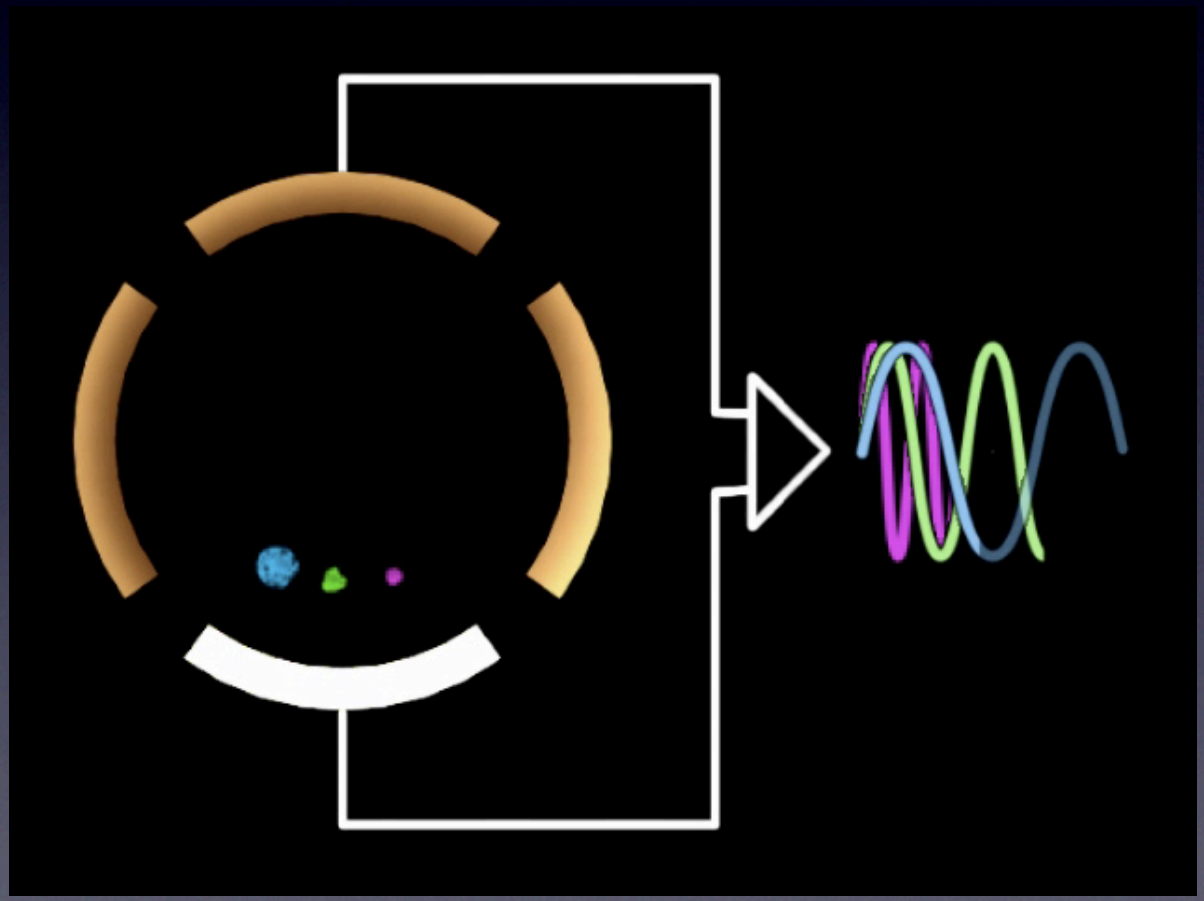
Muddiman, Hawkridge

Ion Cyclotron Resonance

Basic Principle:

Ions rotate about a magnetic field with frequency ~ 1 / m/z.





Important Relationships:

Calibration

$$\frac{m}{z} = \frac{A}{f} + \frac{B}{f^2}$$

Cyclotron Frequency

$$\omega_c = \frac{qB_0}{m}$$

Resolving Power

$$R \sim \frac{qB_0T_{\mathrm{acq'n}}}{m}$$

Muddiman, Hawkridge

Marshall, Hendrickson IJMS 215 (2002) 59-75

Time of Flight

Basic Principle: ion flight time proportional to m/z.

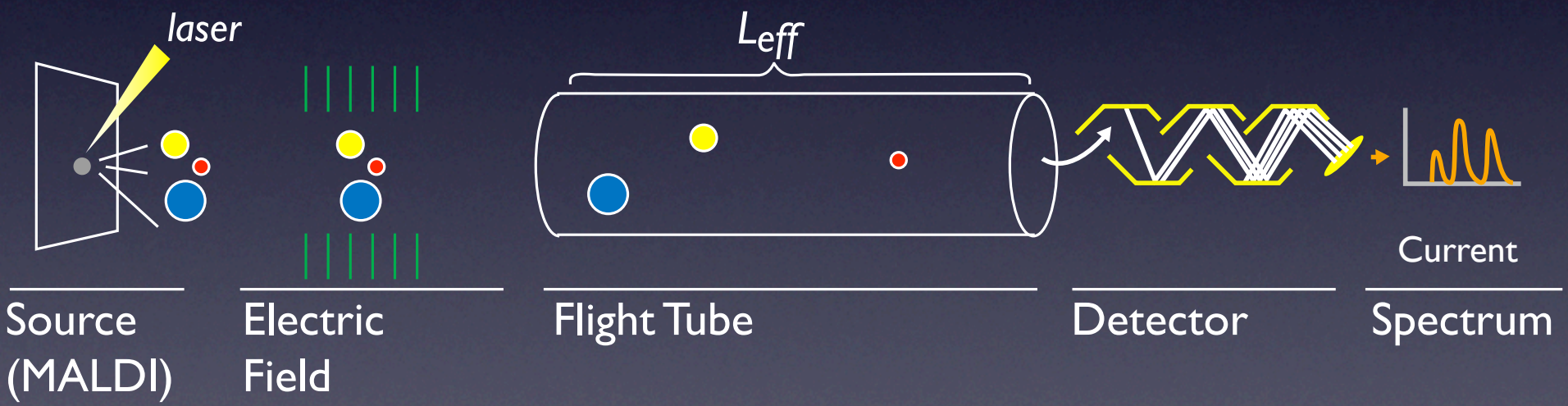
Important Relationships:

Calibration

$$\frac{m}{z} = At^2$$

Resolving Power

$$R \sim \frac{L_{\text{eff}}}{2\Delta z}$$

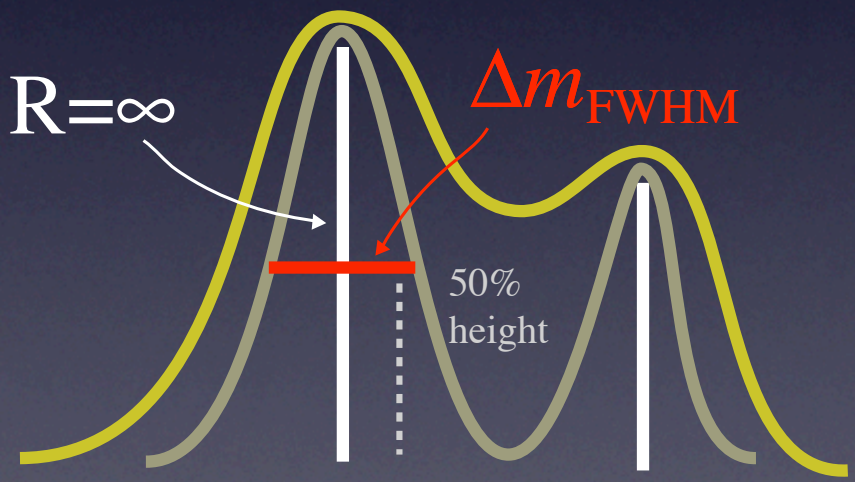


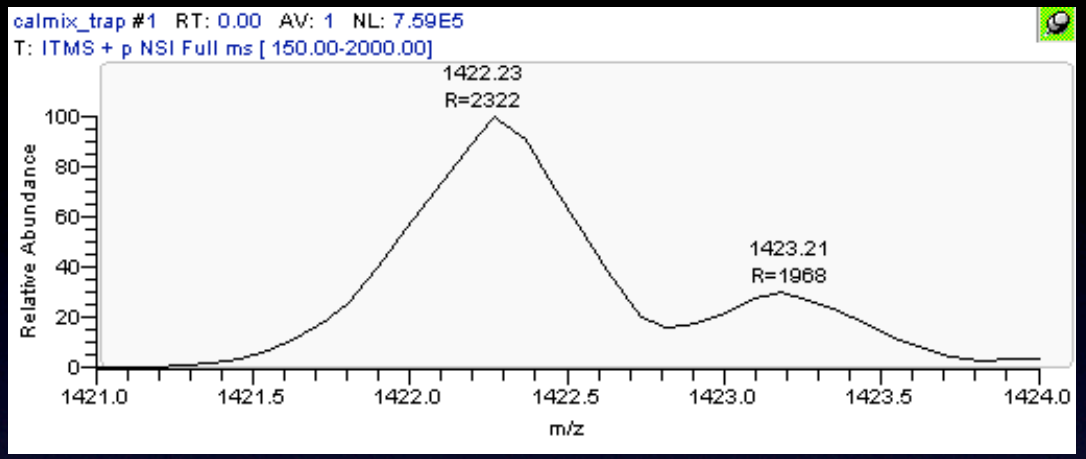
Chernuschevich et al. J. Mass Spectrom. 2001; 36: 849-865.
Guilhaus. J. Mass Spectrom. 1995; 30: 1519-1532.

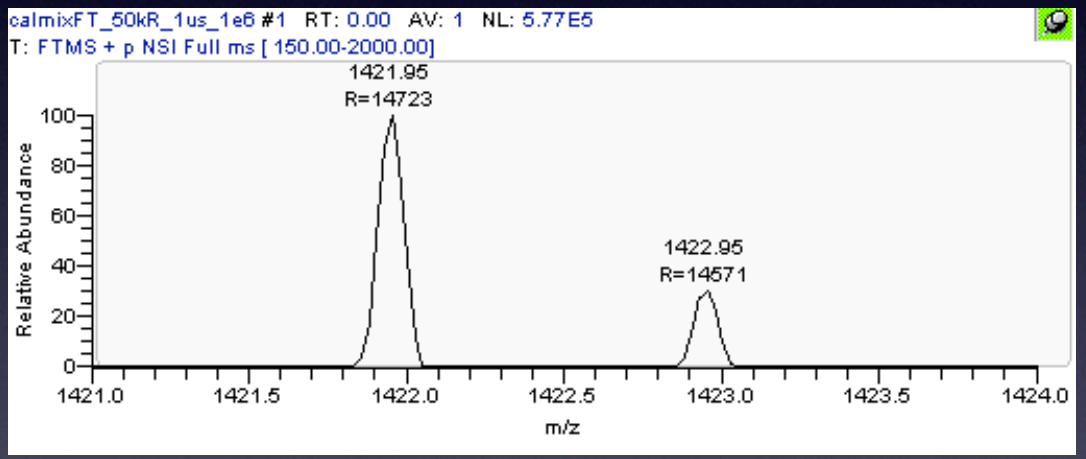
Resolution / Resolving Power

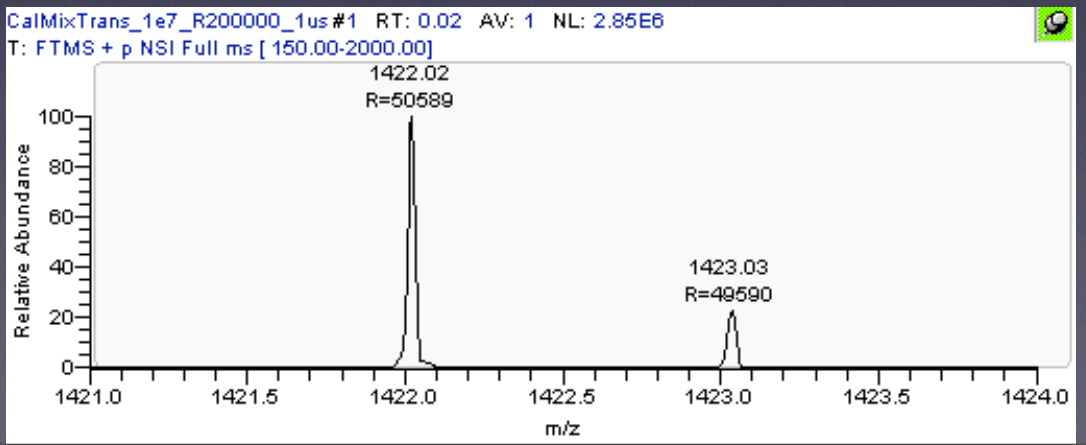
A measure of an instrument's ability to discriminate or "resolve" peaks adjacent in mass.

$$R = \frac{m}{\Delta m_{\mathrm{FWHM}}}$$

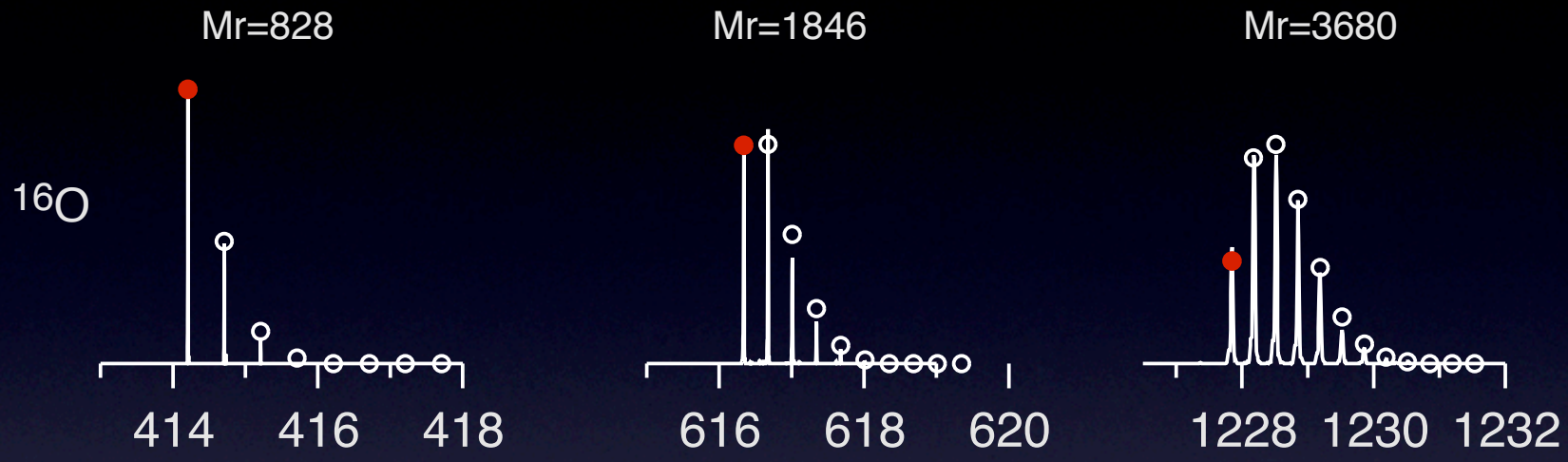








Importance of Resolving Power: Isotopes



Isotopes

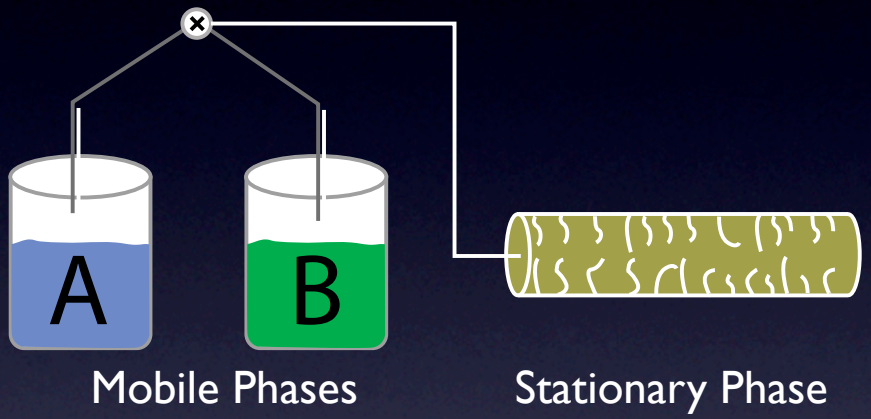
(differing # of neutrons)

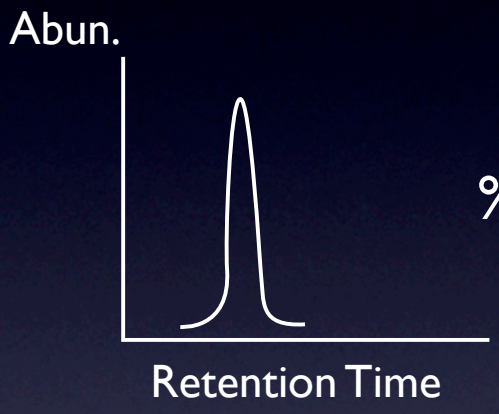
spacing ~$1/z$

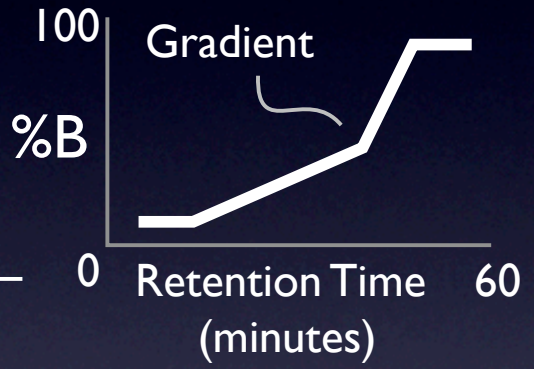
^{1}H	99.9885%	^{16}O	99.7570%
^{2}H	0.0115%	^{17}O	0.038%
		^{18}O	0.205%
^{12}C	98.93%		
^{13}C	1.07%	^{32}S	94.93%
		^{33}S	0.76%
^{14}N	99.995%	^{34}S	4.29%
^{15}N	0.005%	^{36}S	0.02%

Chromatography

Chromatography separates molecules in time based on their chemical properties.

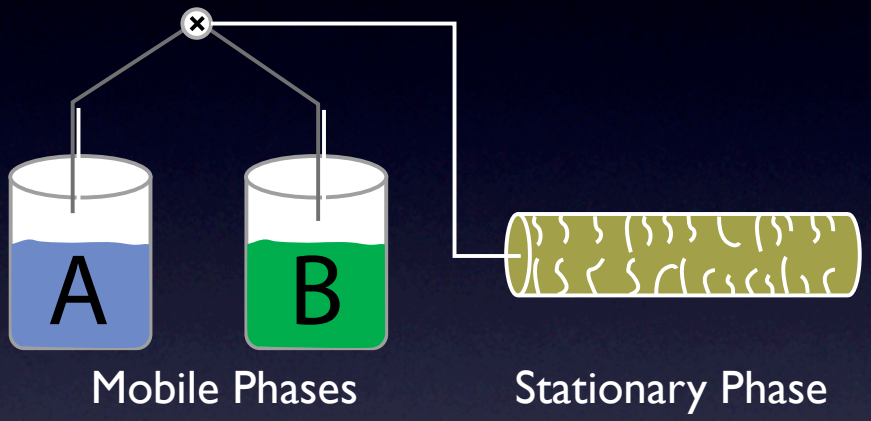


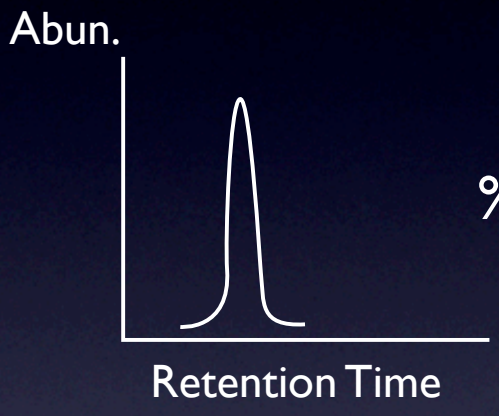


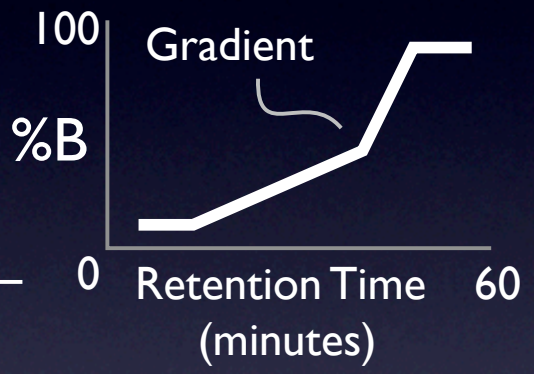


Chromatography

Chromatography separates molecules in time based on their chemical properties.



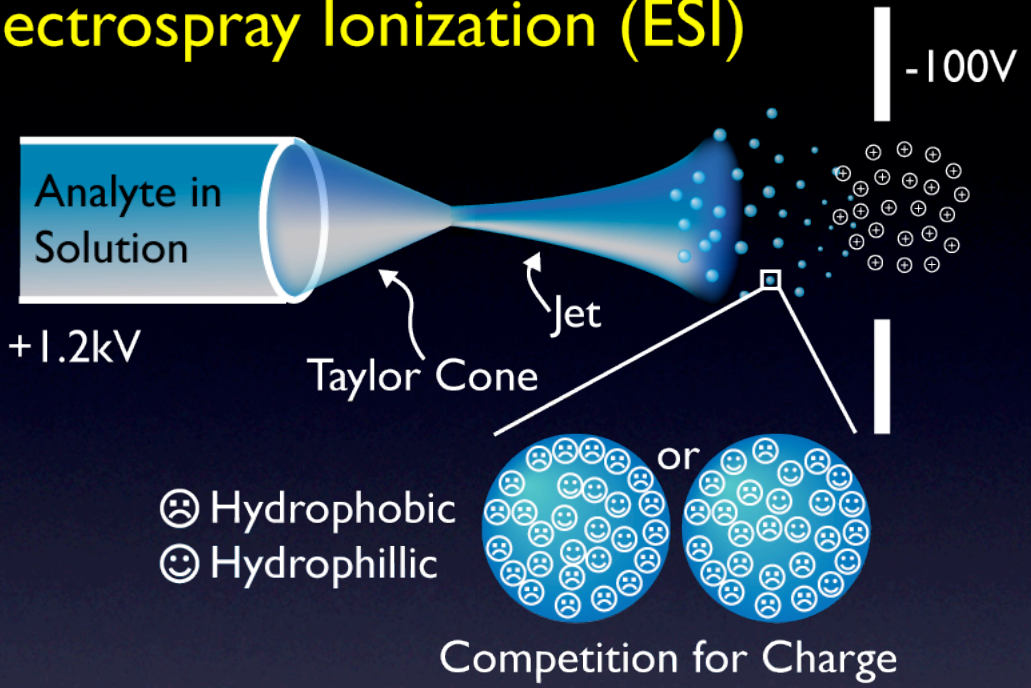




	Chemical Property
Reverse Phase	Water-dislike (Hydrophobicity)
Strong Cation Exchange (SCX)	Charge

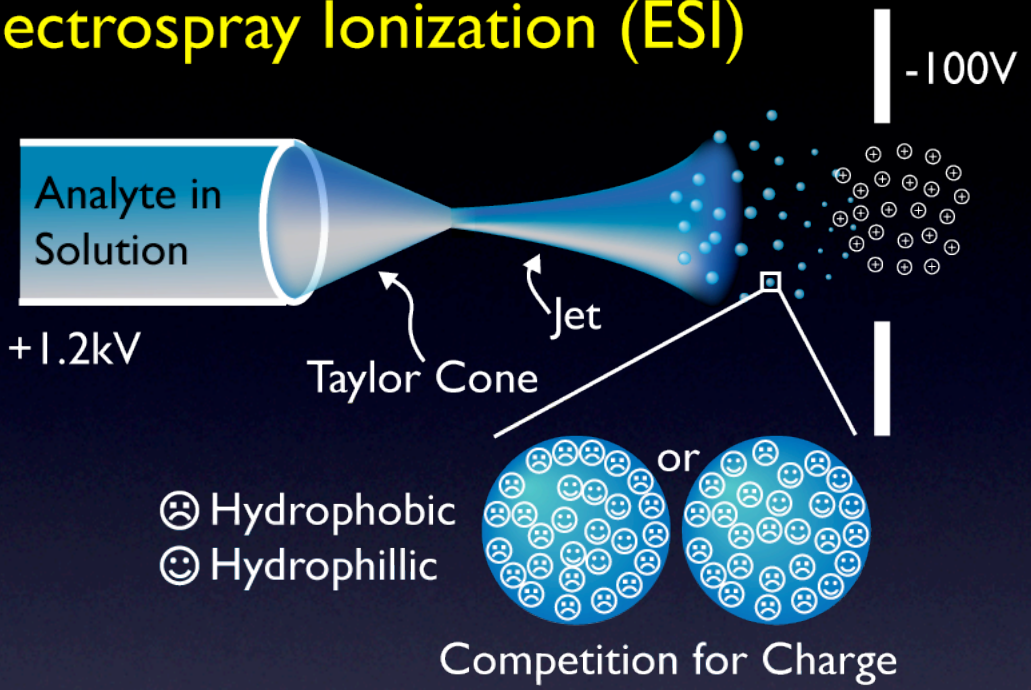
Ionization

Electrospray Ionization (ESI)



Ionization

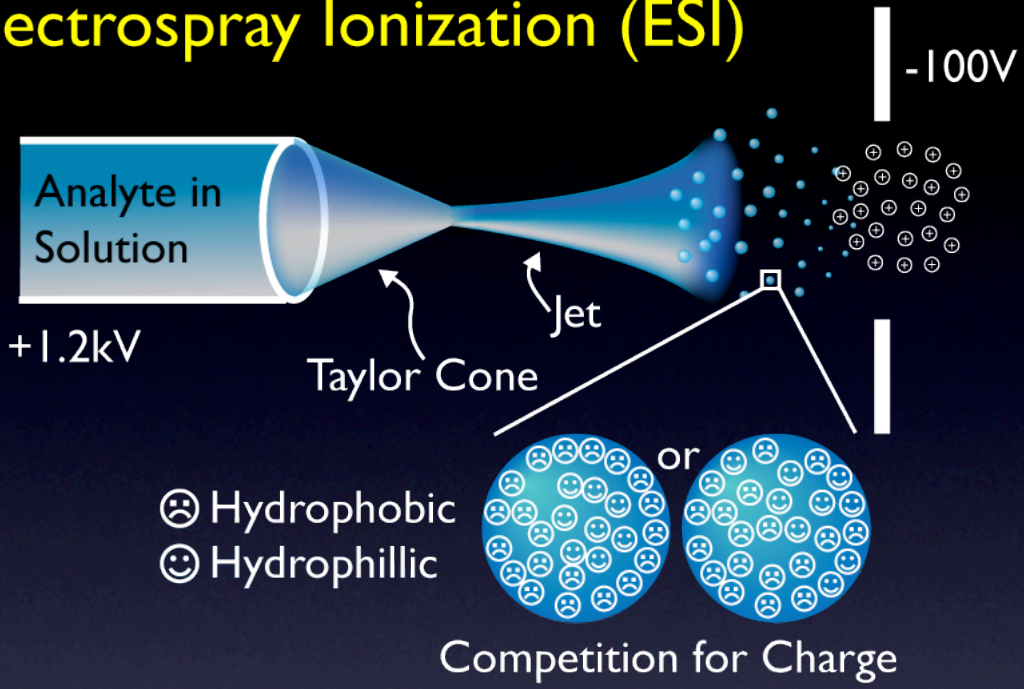
Electrospray Ionization (ESI)



- Multiply charged species
- Solution phase
- Fenn et al: Nobel Prize

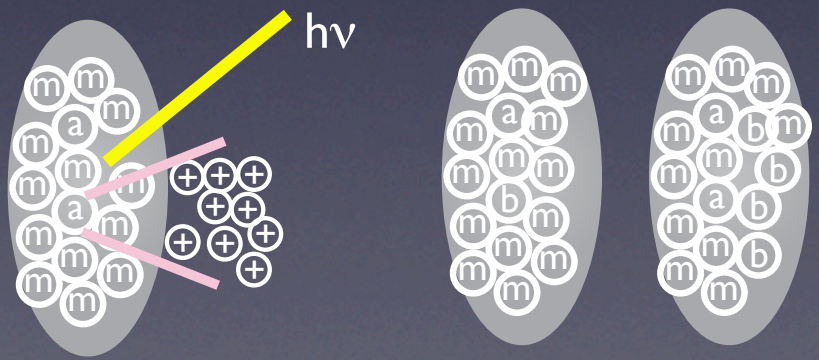
Ionization

Electrospray Ionization (ESI)



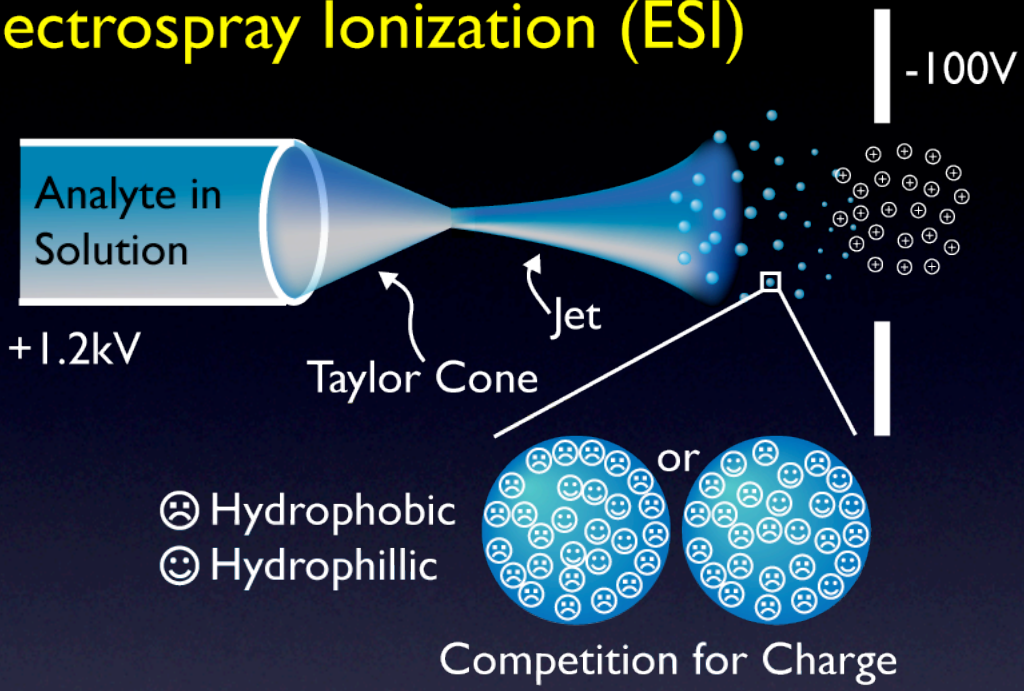
- Multiply charged species
- Solution phase
- Fenn et al: Nobel Prize

Matrix Assisted Laser Desorption Ionization (MALDI)



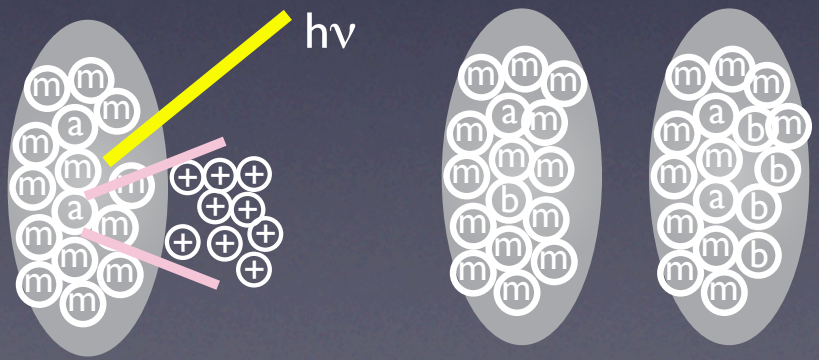
Ionization

Electrospray Ionization (ESI)



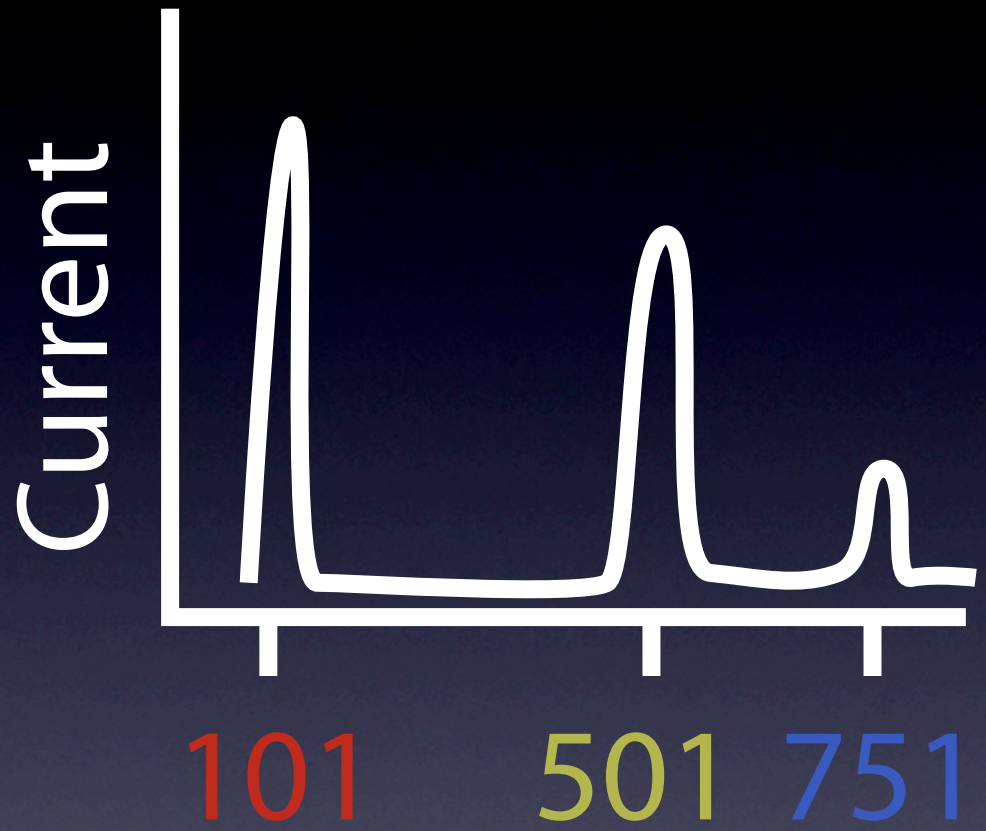
- Multiply charged species
- Solution phase
- Fenn et al: Nobel Prize

Matrix Assisted Laser Desorption Ionization (MALDI)



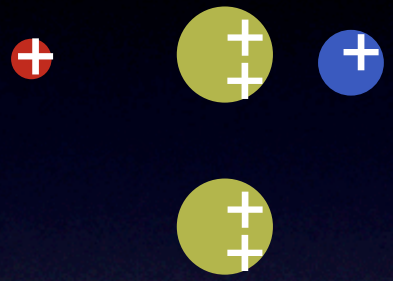
- Mostly singly charged species
- Solid phase
- Tanaka et al: Nobel Prize

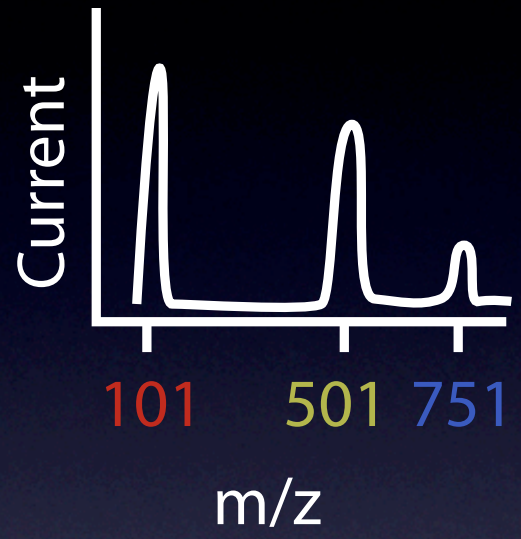
Mass Spec Semi-quantitative



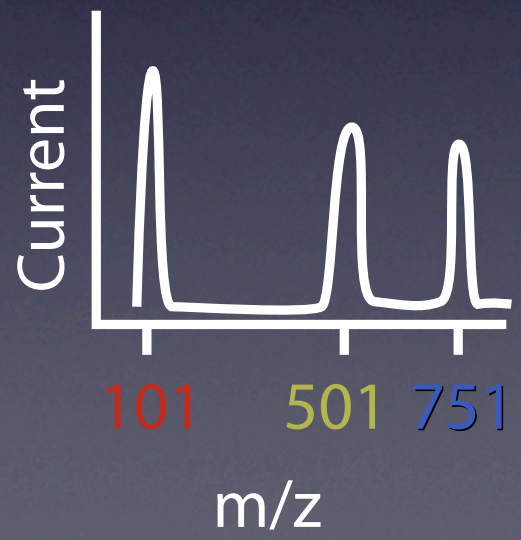
Relative Quantification

Sample 1

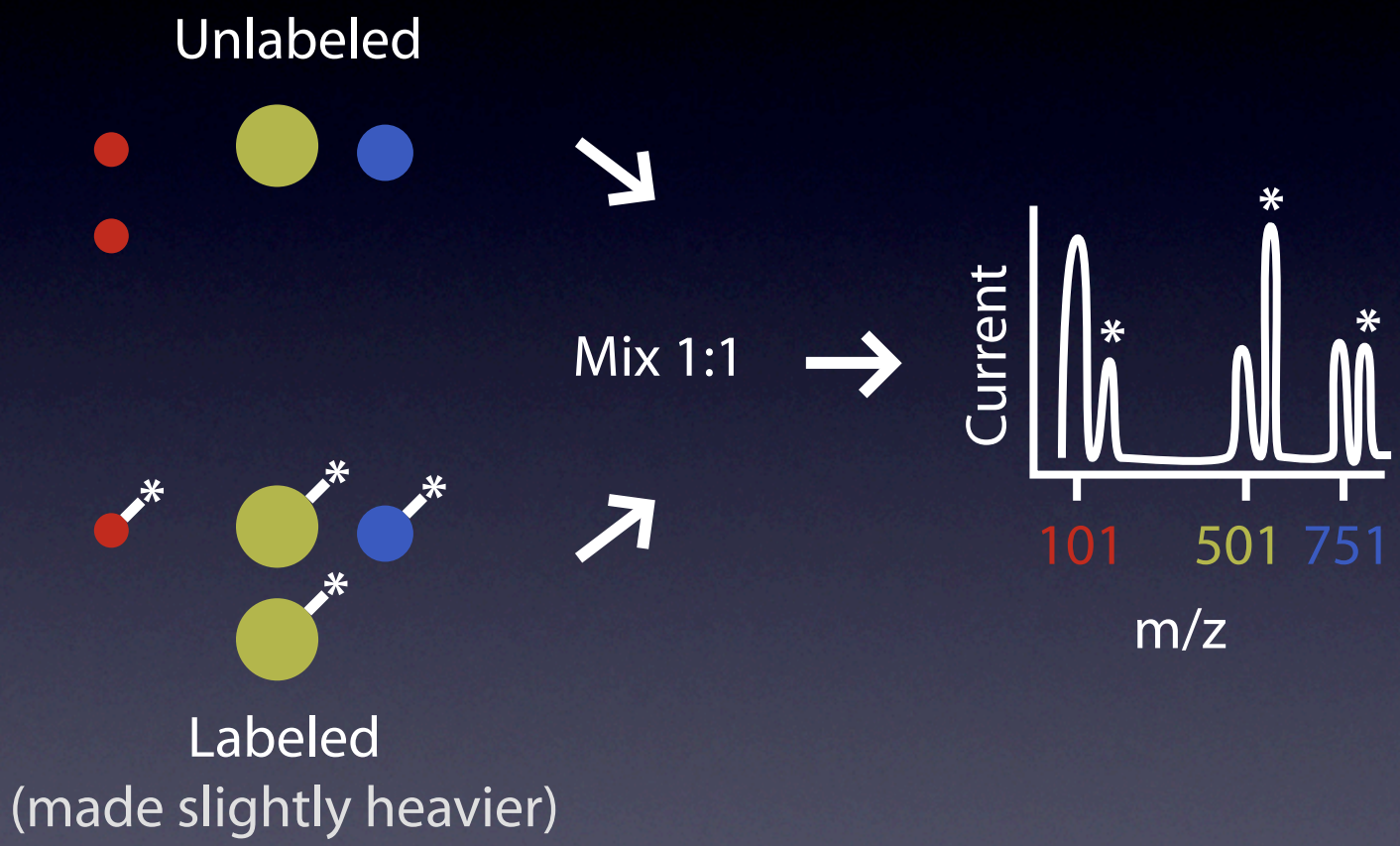




Sample 2

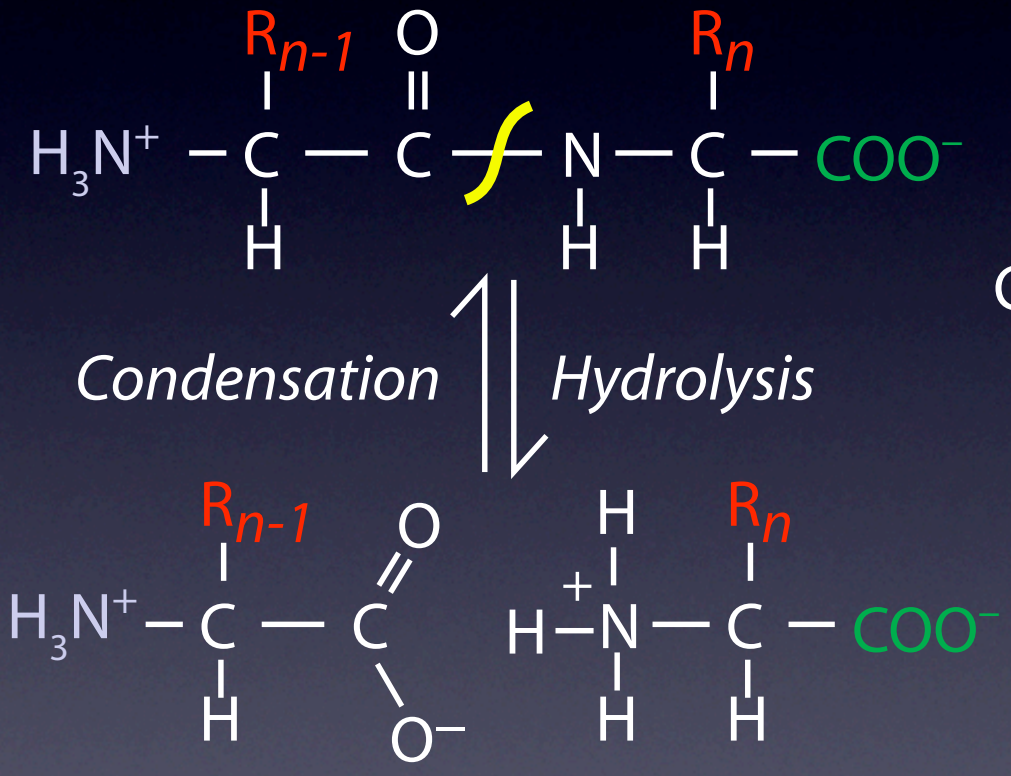


Labeling


Unlabeled
Mix 1:1
Labeled
(made slightly heavier)
Current
101
501
751
m/z

Break proteins apart

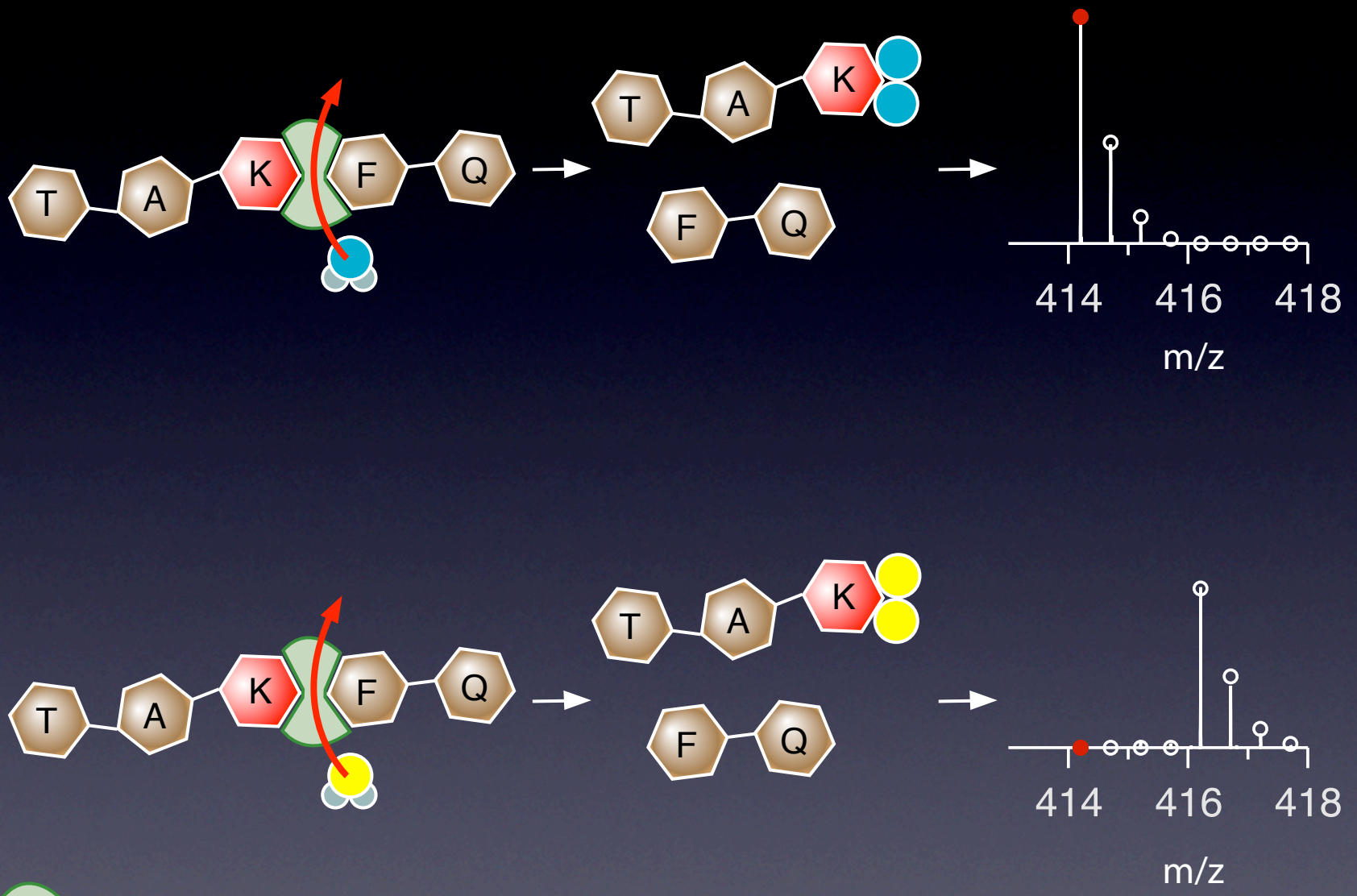
- An *endoprotease* breaks proteins into smaller *peptides* at specific residues.



Common endoproteases

	R_{n-1}	R_n
Trypsin	K, R	!P
Chymotrypsin	F, W, Y	!P
Pepsin	L, F, W, Y	!P
Endopep. V8	E	
Lys-C	K	

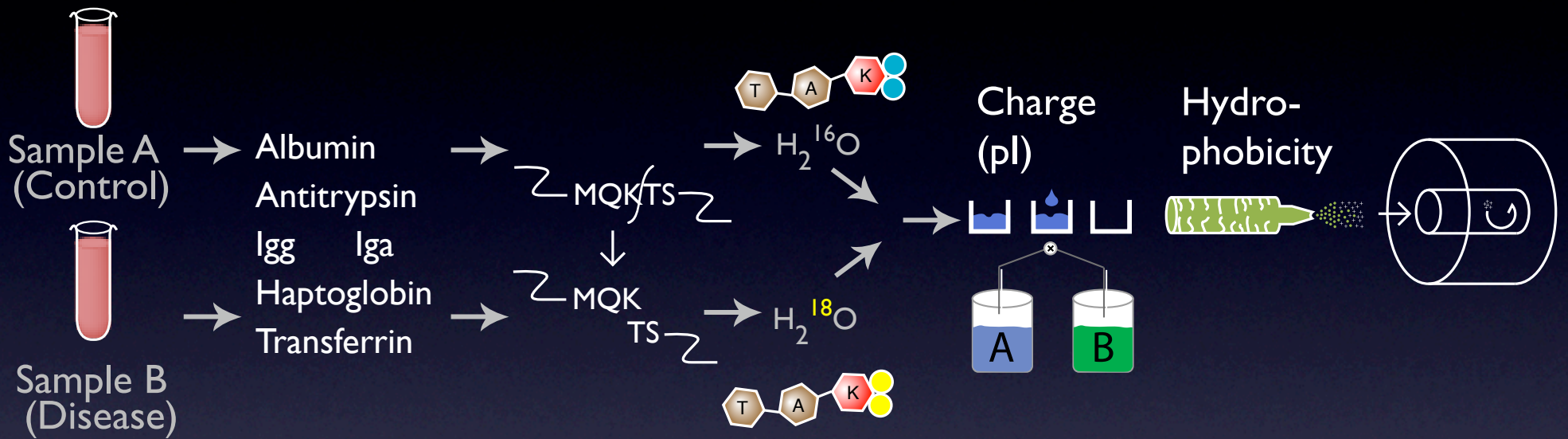
^{18}O Labeling



Alphabet soup

Labeling	*Separation*	*Ionization*	*Mass Analysis*
(None)	2D Gel	ESI	TOF
^{18}O	Reverse-Phase	MALDI	Quadrupole
ICAT	SCX	EI/CI	FT-ICR
iTRAQ	IEF	DESI	Orbitrap
SILAC	...	FAB	...
...		...	

Experimental Overview



Blood Plasma

Deplete Top 6 Proteins

Trypsin Digestion

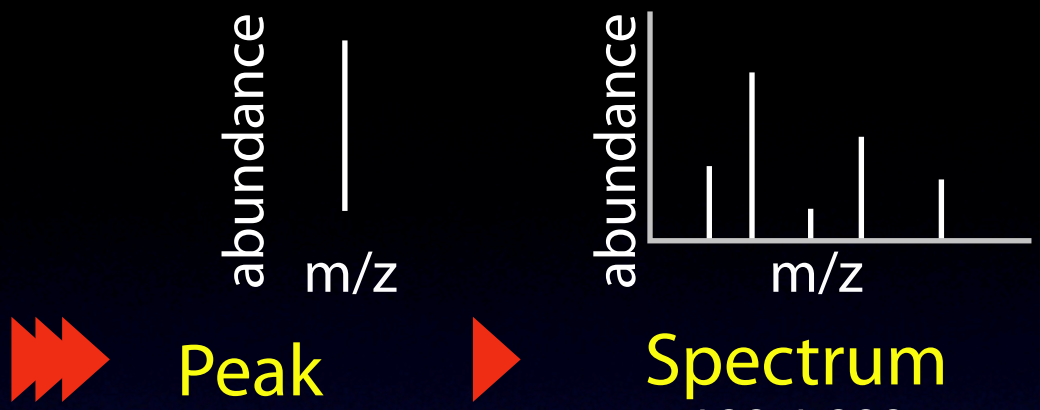
^{18}O Labeling

Strong Cation Exchange Chromato-graphy

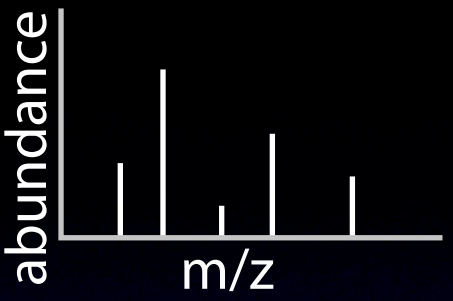
Reverse Phase Chromato-graphy

FT-ICR/ Oribtrap Mass Spectro-metry


abundance
m/z
Peak

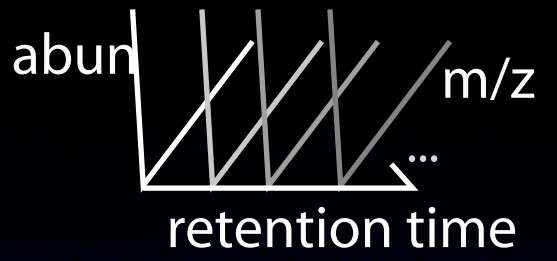

abundance
m/z
Peak
abundance
m/z
Spectrum
100-1,000
peaks/spectrum

Peak



Spectrum

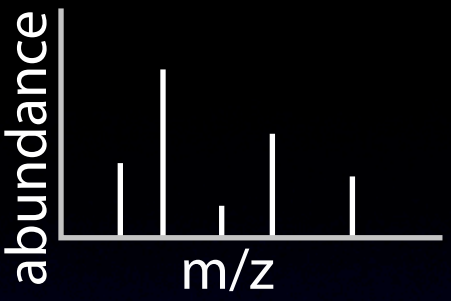
100-1,000
peaks/spectrum



RP Chromatogram

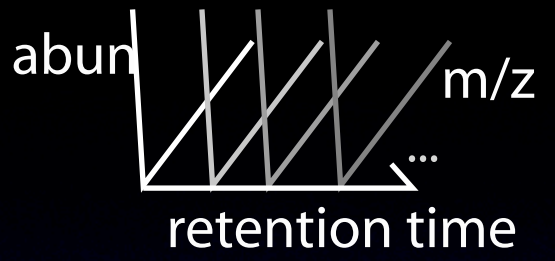
~1,500
spectra/chromatogram

Peak



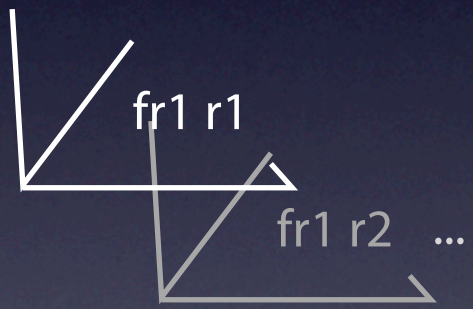
Spectrum

100-1,000
peaks/spectrum



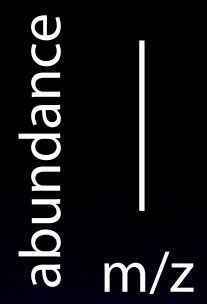
RP Chromatogram

~1,500
spectra/chromatogram

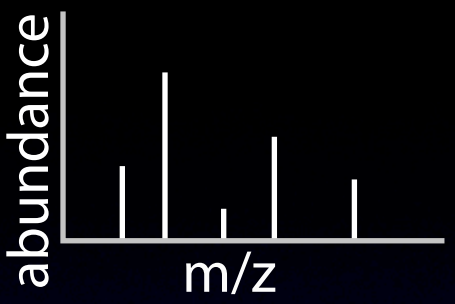


SCX Fraction

1-3 Chromatograms/
SCX Fraction
(technical replicates)

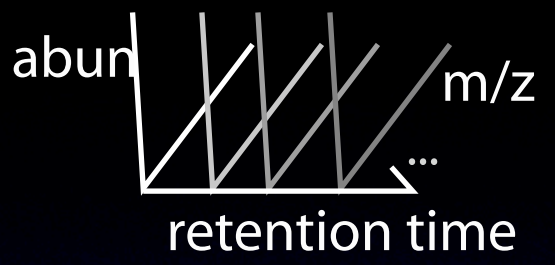


Peak



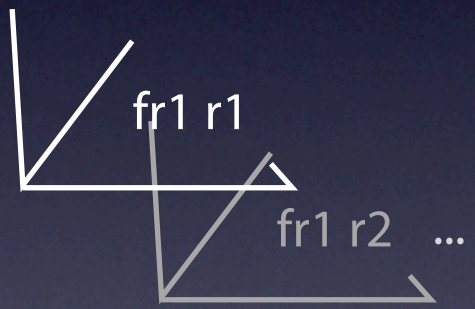
Spectrum

100-1,000
peaks/spectrum



RP Chromatogram

~1,500
spectra/chromatogram

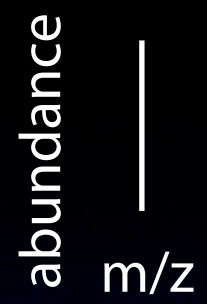


SCX Fraction

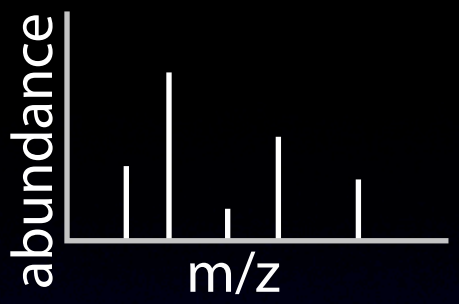
1-3 Chromatograms/
SCX Fraction
(technical replicates)

Patient

30 SCX Frxns/
Patient

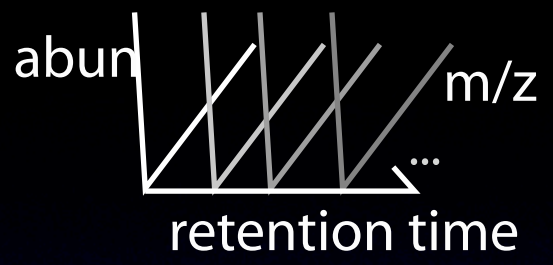


Peak



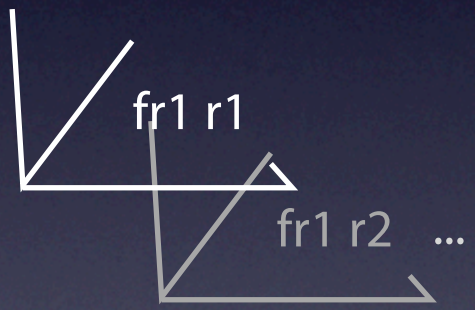
Spectrum

100-1,000
peaks/spectrum



RP Chromatogram

~1,500
spectra/chromatogram

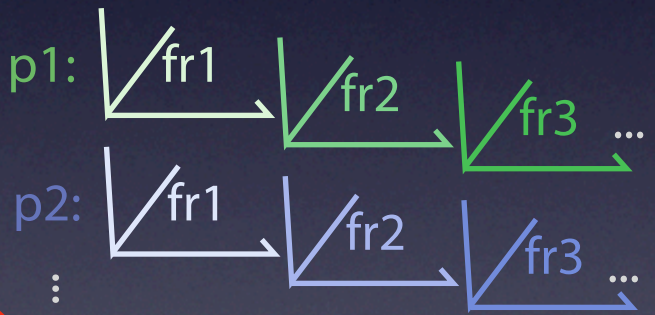


SCX Fraction

1-3 Chromatograms/
SCX Fraction
(technical replicates)

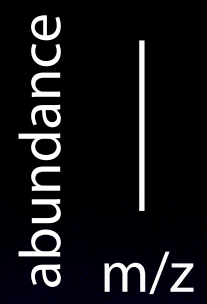
Patient

30 SCX Frxns/
Patient

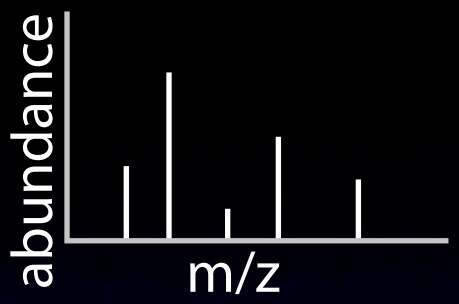


Study

?? Patients/ study
(12 best so far)



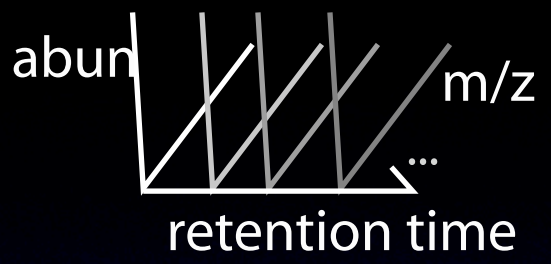
Peak



Spectrum

100-1,000
peaks/spectrum

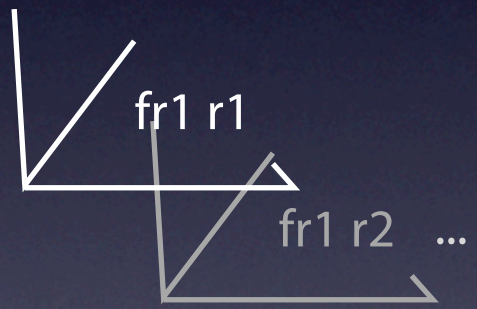
1-2 sec/spectrum



RP Chromatogram

~1,500
spectra/chromatogram

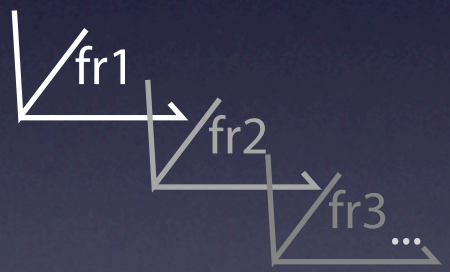
60-90 min/chromatogram



SCX Fraction

1-3 Chromatograms/
SCX Fraction
(technical replicates)

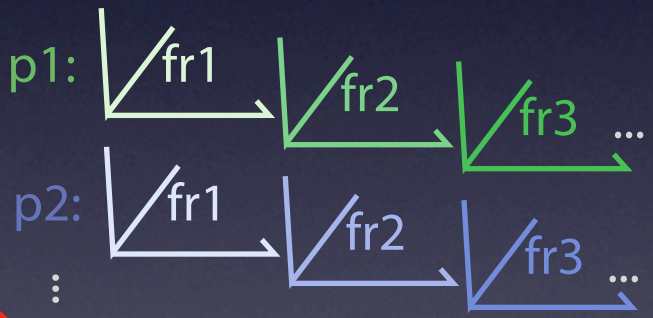
60-270 min/fraction



Patient

30 SCX Frxns/
Patient

30-135 hours/patient



Study

?? Patients/ study
(12 best so far)

weeks-months-...

Data Size Per Patient Sample

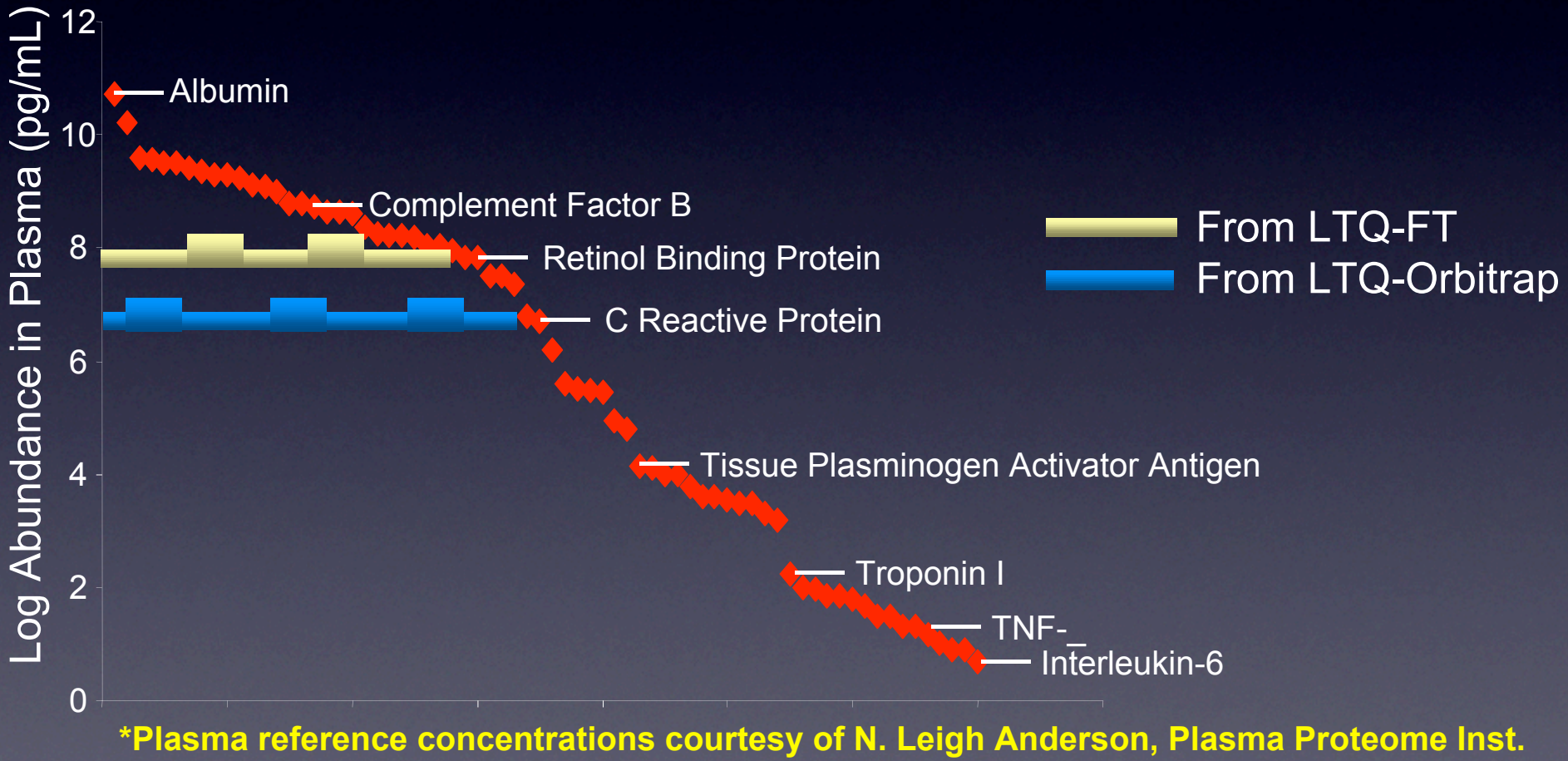
	LTQ-FT
Peaks	~9M
Spectra	57k / 228k
Fractions	30
Total Raw Size	~1.5GB
Analysis Size (Compressed, Transient)	>20GB

Data Set 1: LTQ-FT, 4 samples
(Scaffold 99% Protein Probability)
(20 ppm precursor mass tolerance)

- 119800 MS/MS spectra
- 164 Proteins with ≥ 2 peptides
- 206 Proteins with ≥ 1 peptide

Data Set 2: LTQ-Orbitrap, 1 sample
(Scaffold 99% Protein Probability)
(20 ppm precursor mass tolerance)

- 270700 MS/MS spectra
- 437 Proteins with ≥ 2 peptides
- 862 Proteins with ≥ 1 peptide

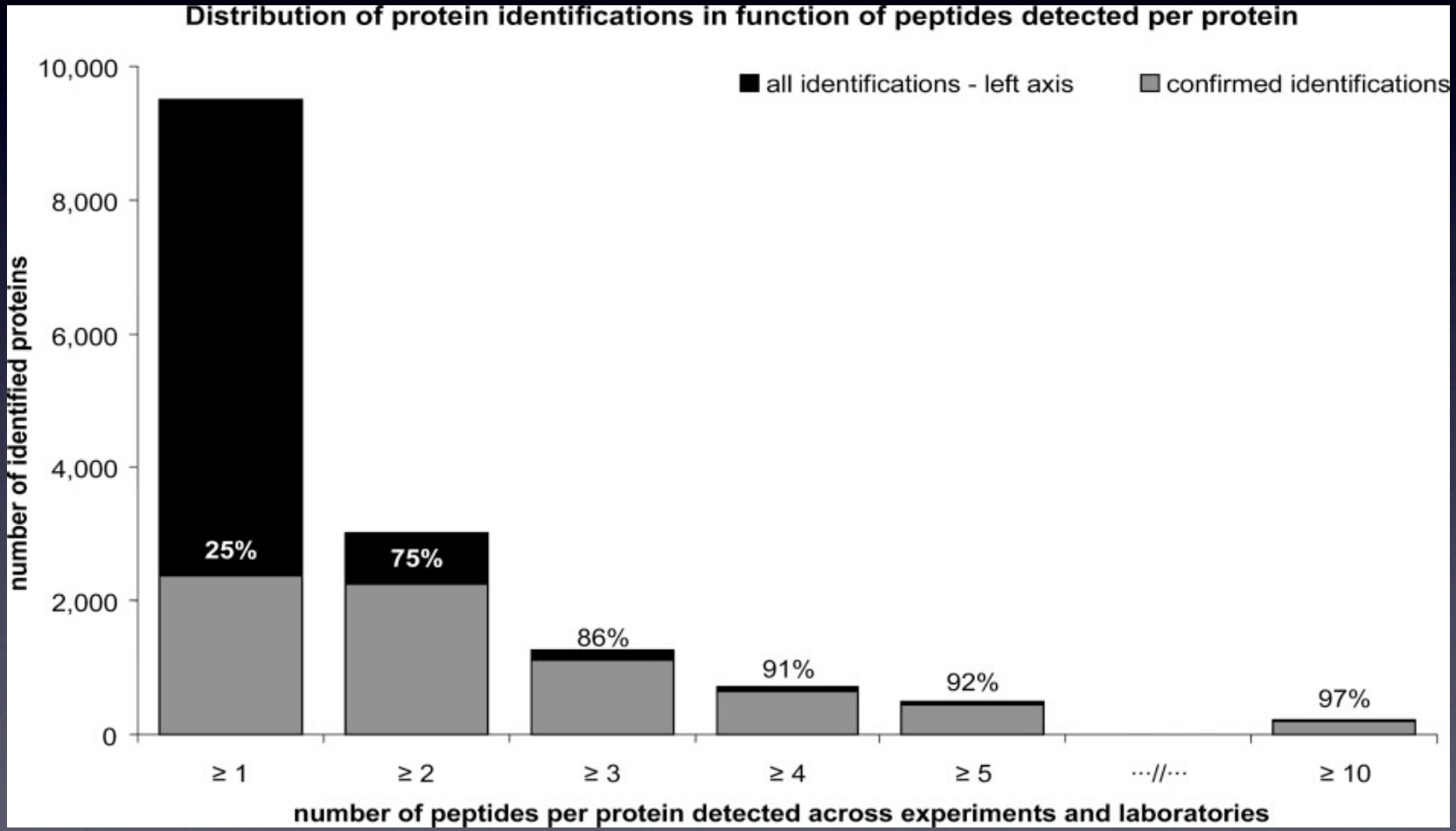


*Plasma reference concentrations courtesy of N. Leigh Anderson, Plasma Proteome Inst.

Kenneth Johnson et al. *ASMS* 2006

How are we doing?

- HUPO: 3,020 total proteins with at least two peptides from a consortium of 18 laboratories.



G. S. Omenn et al. *Proteomics* 2005, 5, 3226–3245

Acknowledgments

- MPRC: Kenneth Johnson, H Robert Bergen, Rudi Chiarito.
- Biostatistics: Terry Therneau, Jeanette Eckel-Passow, Ann Oberg, Douglas Mahoney.
- CTSA/BIC: K. Sreekumaran Nair, Yan Asmann.
- Cancer Center: Janet Olson.
- David Muddiman (NCSU).
- Funding: National Institute of Diabetes and Digestive and Kidney Diseases, (R01DK 70179-2), the National Cancer Institute (R33 CA105295 to D.C.M. and R25 CA92049 to J.E.E.P.), the Lustgarten Foundation for Pancreatic Cancer Research, the James M. Kemper Biomarker Research Fund, the Gordon C. and Elizabeth W. Gilroy Fellowship in Proteomic Research, and the Mayo Clinic College of Medicine.