Qualitative Proteomics

(how to obtain high-confidence high-throughput protein identification!)

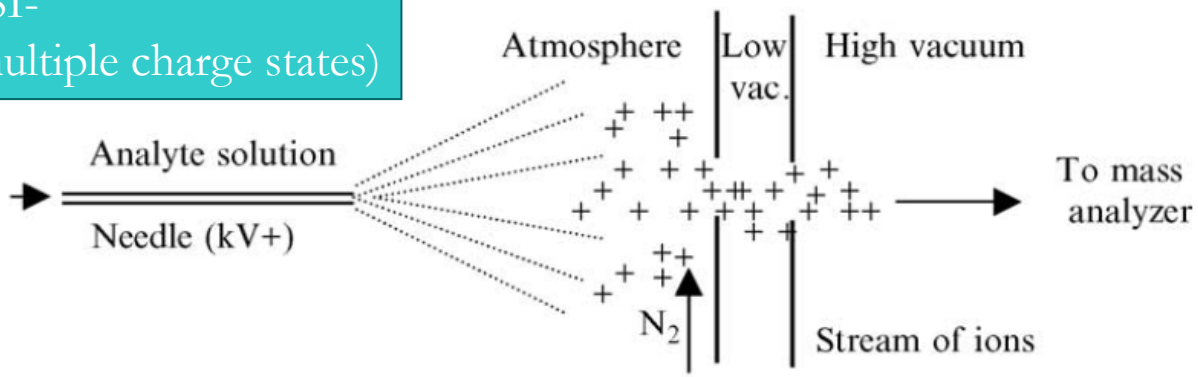
James A. Mobley, Ph.D.
Director of Research in Urology
Associate Director of Mass Spectrometry
(contact: mobleyja@uab.edu)

The "Birth" of Proteomics

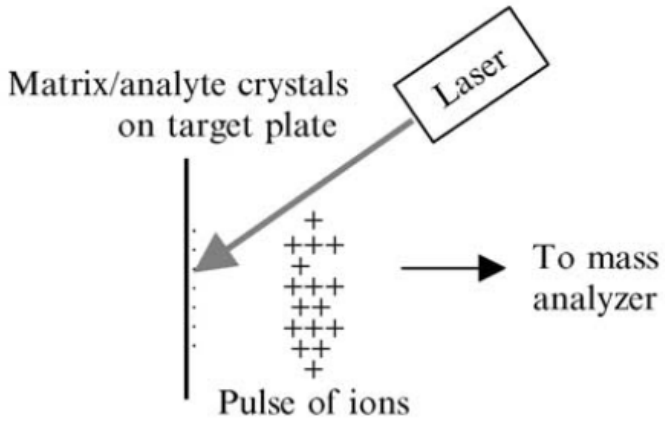
- Proteomics was not "born" as many will say, but was indeed "ignited" with the "discovery" of "*very* soft ionization" techniques.
- For those in the field at the time, *this was "huge"*, but still didn't really take of until the mid to late 90's!!
- However, prior to this …. ..both "hard and soft ionization" processes were well established! These include, electron ionization (EI), chemical ionization (CI), fast atom bombardment (FAB), liquid secondary ion MS (LSIMS), and plasma desorption MS (PDMS).
- The newer techniques were initiated by……and "still include" matrix associated desorption ionization (MALDI)- [Hillenkamp, et. al. 1988] and electrospray ionization (ESI)- [Fenn et. al. 1989] based sources.

Soft Ionization Techniques

ESI-
(multiple charge states)



MALDI-
(singly charged ions)



Baldwin M.A., Mass spectrometers for the analysis of biomolecules, Methods in Enzymology, Vol 402

Protein Characterization

- So….There are just three simple points to remember in the business of protein identification/ characterization.
  - 1) Generation of a peptide-based mass spectra.
    - [MS^1 only]; peptide mass fingerprinting [PMF] (protein must be nearly pure)
    - [MS^2]; sequence data (high purity is not necessarily required)
  - 2) Analysis.
    - Matching algorithms based on *in-silico* digestion (MS^1) & fragmentation (MS^2) of known genes
    - *Denovo* Sequencing (MS^2)
  - 3) Validation.
    - Immuno-affinity techniques (Western analysis, Immuno-histochemistry, ELISA, Luminex, etc).
    - Mass Spectrometry (standard heavy isotope tags)

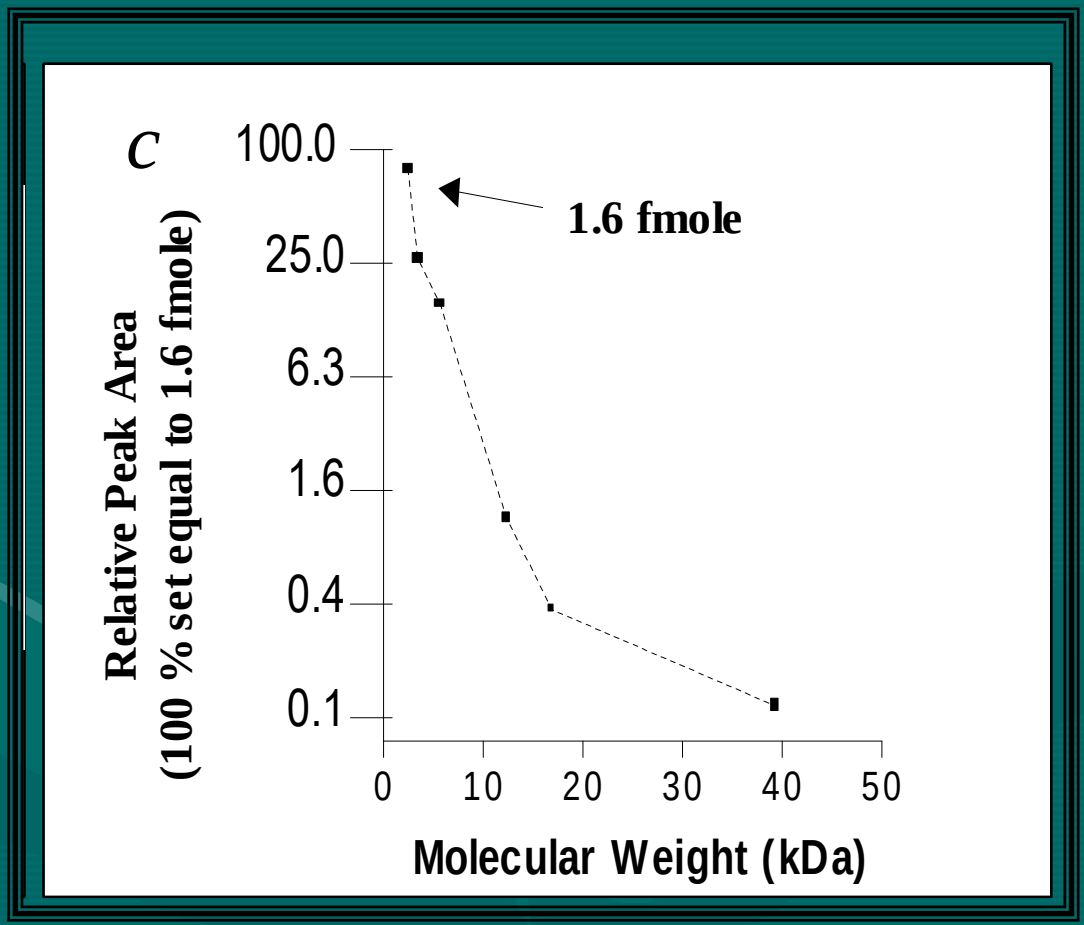
Proteins to Peptides….

- Even today, we are highly limited by decreased detection, resolving power, and poor fragmentation of "whole" proteins!
- Therefore, we "digest" proteins to peptides prior to MS analysis.
- Many chemical and enzymatic techniques have been published; however, trypsin remains the most commonly utilized enzyme for use in proteomics!
- This enzyme cleaves at arginine and lysine, yielding peptides that are easily detected and fragmented in the most common mass analyzers today.
- Keeping in mind that utilizing multiple digestion procedures carried out on the same sample can be very complementing!

http://donatello.ucsf.edu/
A lot of information here……
Take a look at Protein Prospector…….

Sensitivity is Inversely Proportional to Mass

(MALDI-ToF Example)



Student Lab Example!

So….digestion is necessary, but does increase complexity!

Common Enzymes:

Trypsin*	K-X, R-X
Chymotrypsin*	X-L, -F, -Y, -W
Lys-C*	K-X
Arg-C*	R-X
Asp-N	X-D
Glu-C*	E-X

Common Chemicals:

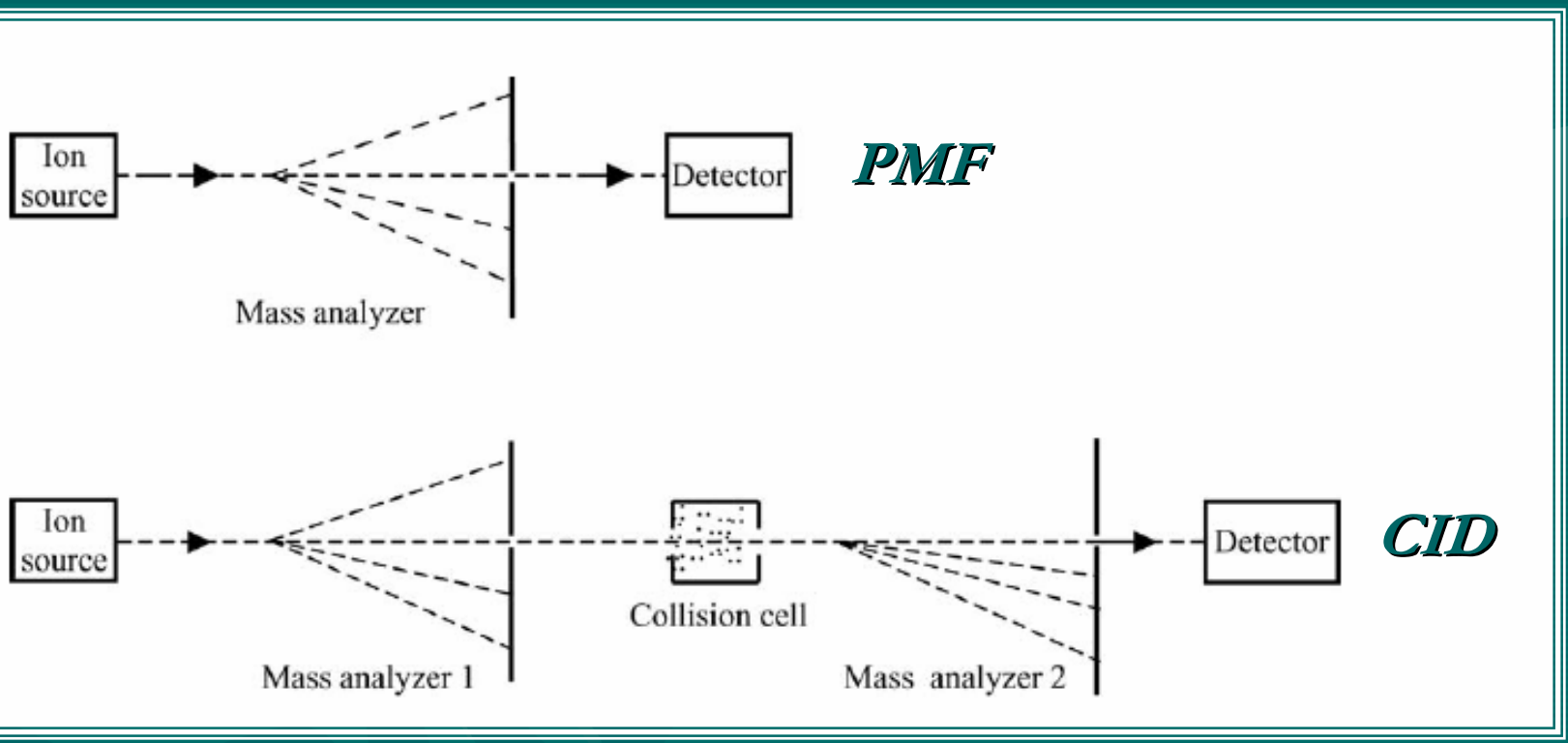
Cyanogen Bromide	X-M
HCL	X-X

* a tailing proline inhibits digestion

Concept of PMF [MS^1]

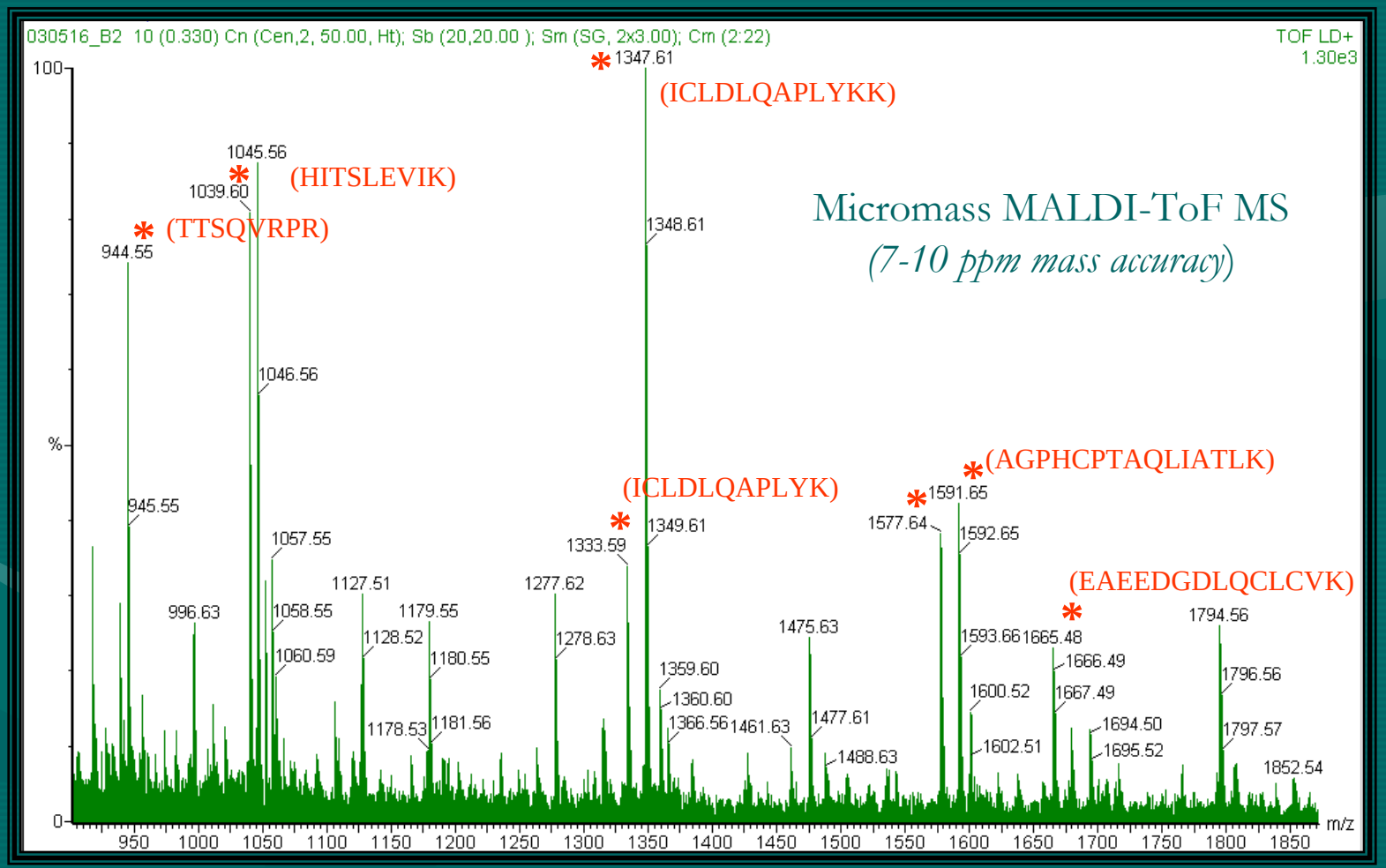
- MS^1 approaches – spectra containing peptide parent molecules only!
- This type of unambiguous protein ID is referred to as "peptide mass fingerprinting" (PMF) introduced ~1990.
- In this case, no sequence information is generated, but it is very sensitive when very little sample is availible!
- The downfall is that the sample must be very pure! Highly complementing for 2D PAGE work.
- However, high mass accuracy is a must as well!
- Overall, these days……unless absolutely necessary…… PMF should not be used!
- There are simply too many matches possible with this technique even with access to high resolution instruments.

MS^1 and/or MS^2



Baldwin M.A., Mass spectrometers for the analysis of biomolecules, Methods in Enzymology, Vol 402

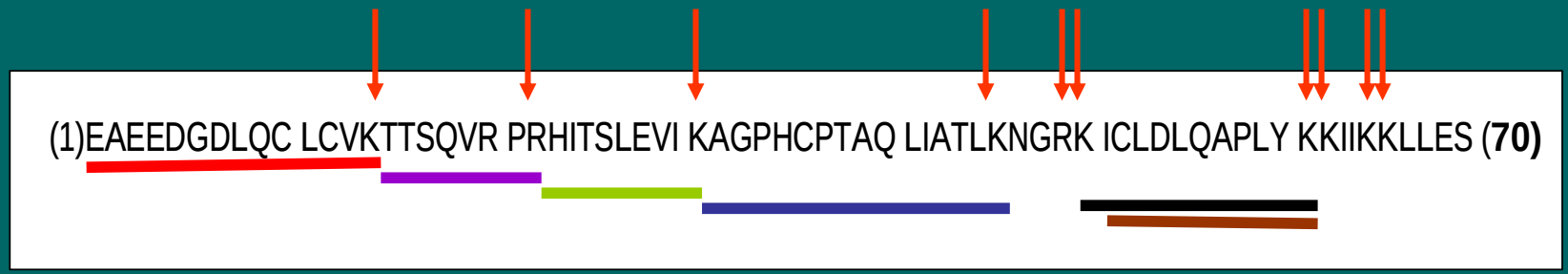
Example of [MS^1] From an In-Gel Digest



Nearly Complete Coverage! (Single Protein Matched by Mascot)

PMF Match!

Predicted and determined mass for CXC4 (~ 8 kDa Protein)



Measured m/z	Theoretical Mass (mi)	Peptide Sequence	Modi.	MC	Δm (Da)	Start	End
944.5486	944.5278	TTSQVRPR		0	+ 0.021	15	22
1039.5997	1039.6152	HITSLEVIK		0	-0.016	23	31
1333.5878	1333.7190	ICLDLQAPLYK	CAM-C	0	-0.131	51	61
1347.6028	1347.7346	KICLDLQAPLYK	Acry	1	-0.132	50	61
1461.6180	1461.8139	ICLDLQAPLYK	CAM-C	0	-0.196	51	61
1577.6407	1577.8474	AGPHCPTAQLIATLK	CAM-C	0	-0.207	32	46
1591.6497	1591.8474	AGPHCPTAQLIATLK	Acry	0	-0.198	32	46
1665.4830	1665.710	EAEEDGDLQCLCVK		0	-0.227	1	14

Protein Fragmentation [MS^2]

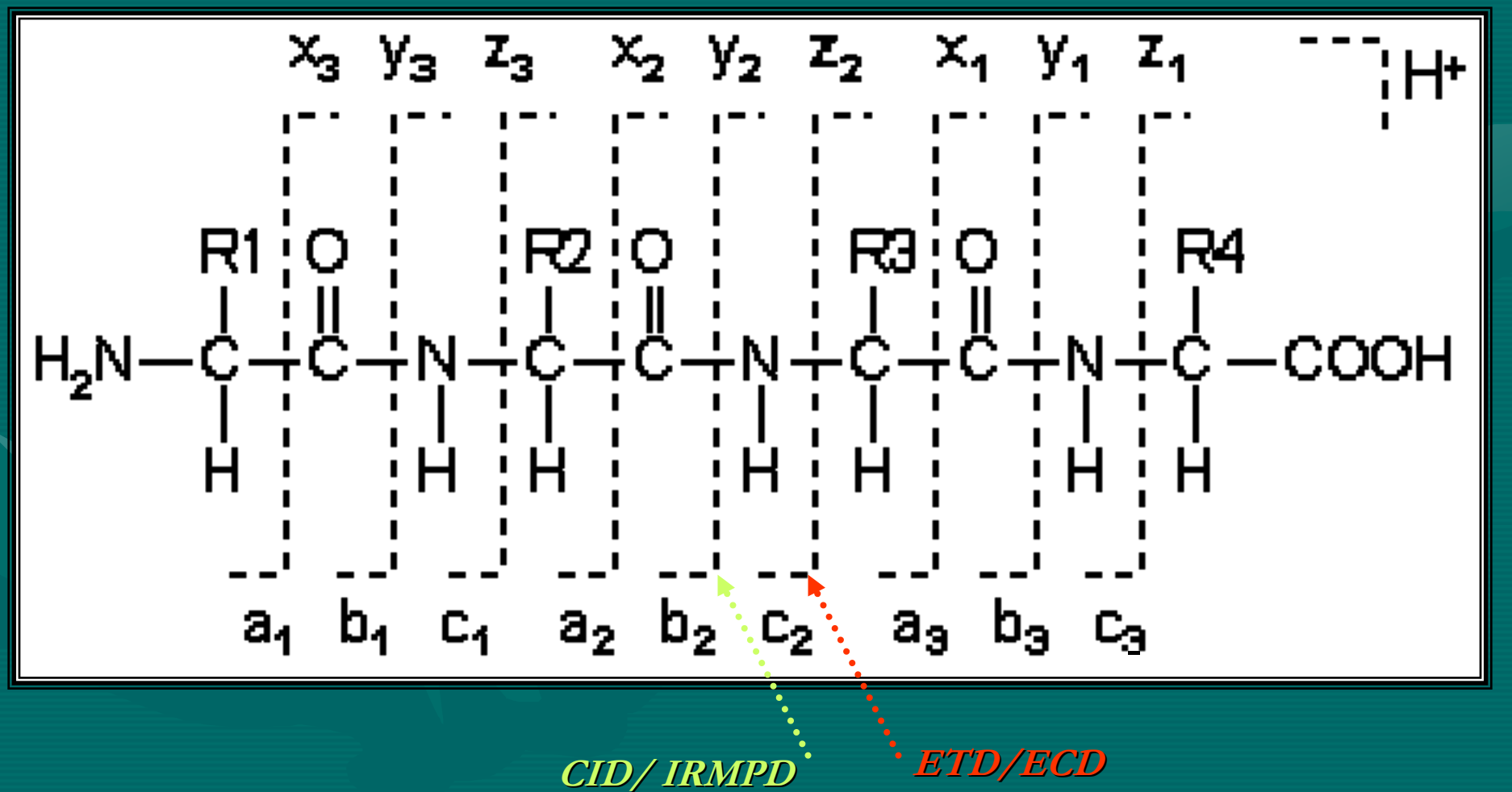
- Both ESI and MALDI based tandem instruments are common in most core settings, each with a combination of mass analyzers.
- Each source and combination of mass analyzers have their selective advantages worthy of a second talk!
- Fragmentation is generally similar, primarily with the generation of either…….
  - b and y ions; collision induced decay (CID) & infrared multiphoton dissociation (IRMPD)
  - z and c ions; electron transfer dissociation (ETD) & electron capture dissociation (ECD)

ESI or MALDI?
Quad, ToF, Ion Trap, and…or FT?

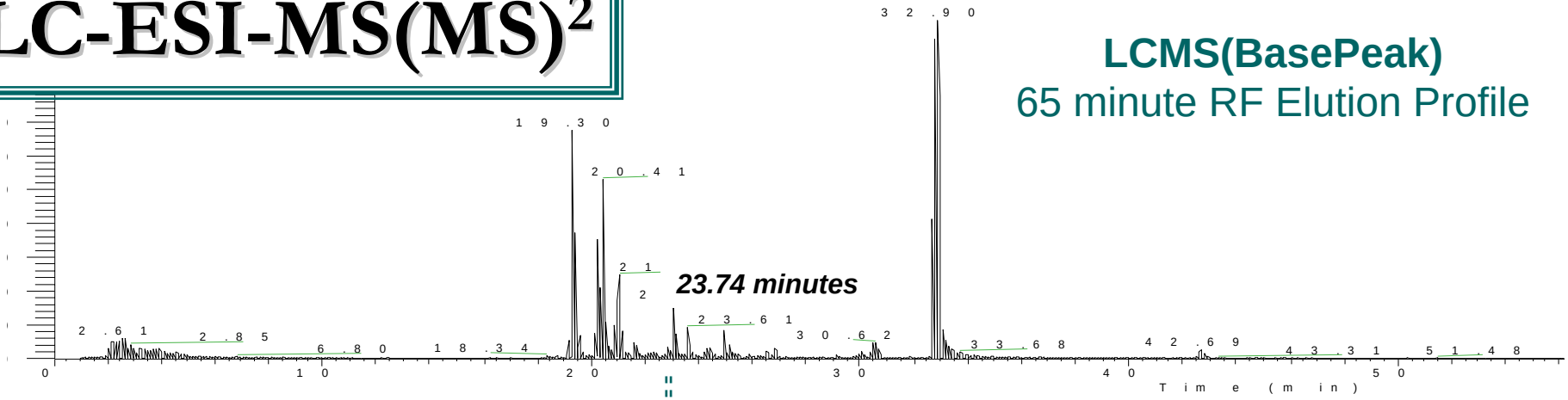
	IT-LIT	Q-Q-ToF	ToF-ToF	FT-ICR	Q-Q-Q	QQ-LIT
Mass accuracy	Low	Good	Good	Excellent	Medium	Medium
Resolving power	Low	Good	High	Very high	Low	Low
Sensitivity (LOD)	Good		High	Medium	High	High
Dynamic range	Low	Medium	Medium	Medium	High	High
ESI	✓	✓		✓	✓	✓
MALDI	(✓)	(✓)	✓			
MS/MS capabilities	✓	✓	✓	✓	✓	✓
Additional capabilities	Seq. MS/MS			Precursor, Neutral loss, MRM		
Identification	++	++	++	+++	+	+
Quantification	+	+++	++	++	+++	+++
Throughput	+++	++	+++	++	++	++
Detection of modifications	+	+	+	+		+++

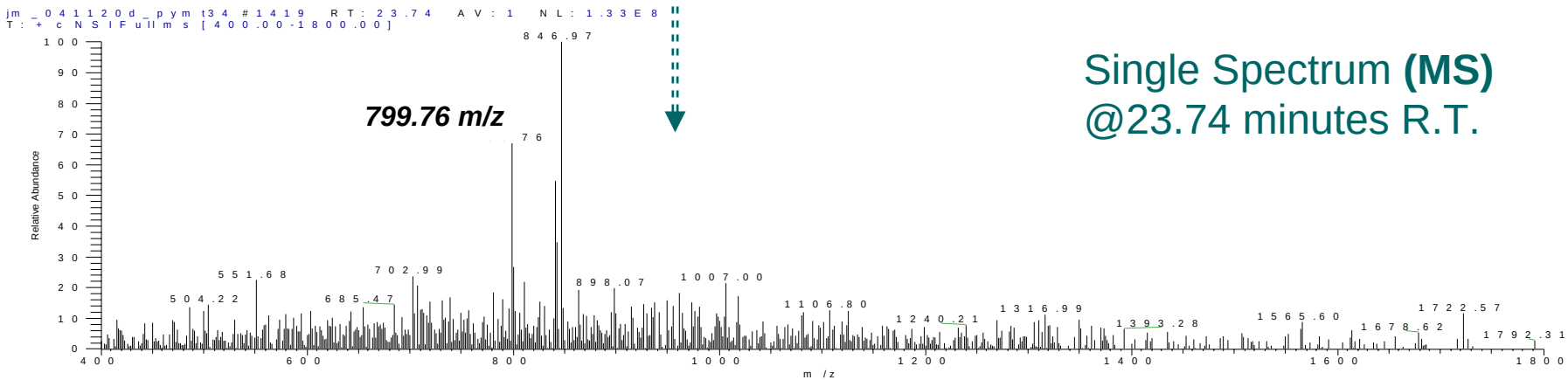
Domon & Aebersold, Science, V312, 14 April, 2006

Peptide Fragmentation



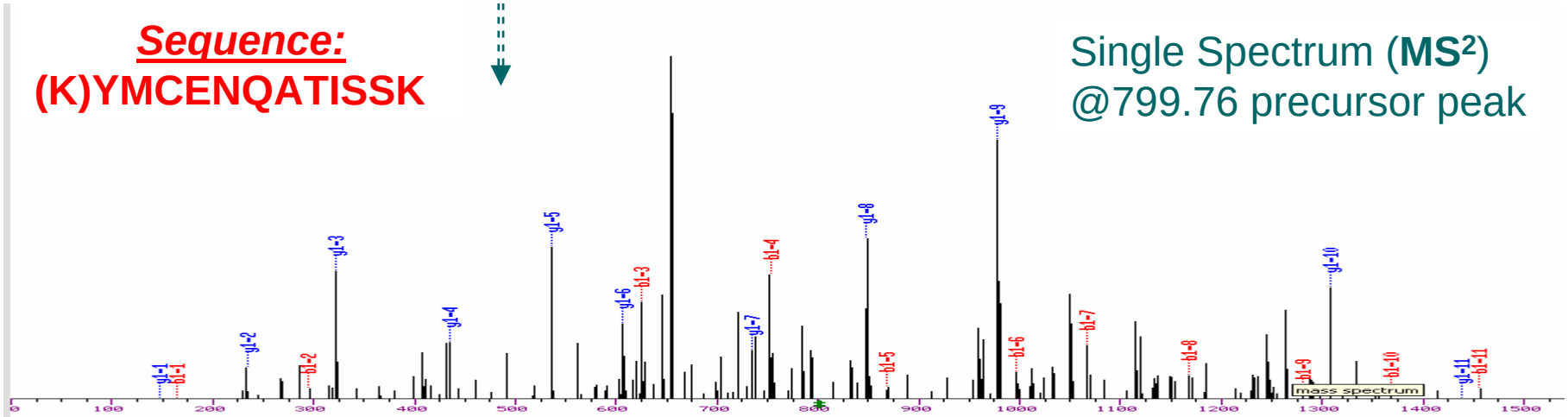
LC-ESI-MS(MS)^2



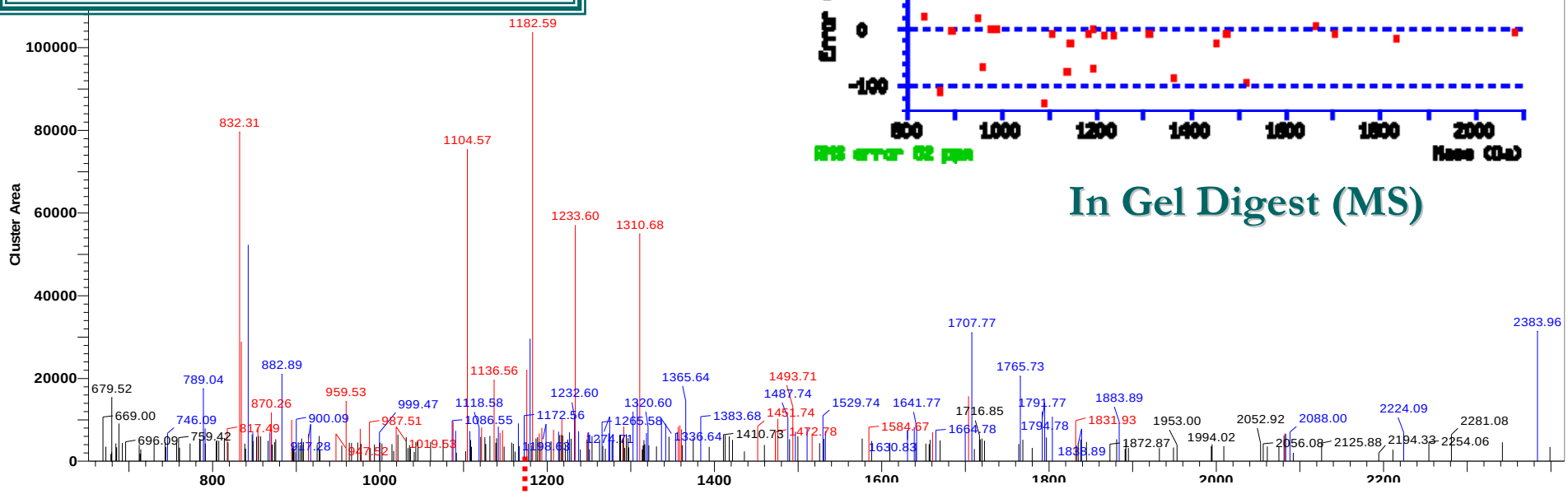


Sequence:

(K)YMCENQATISSK

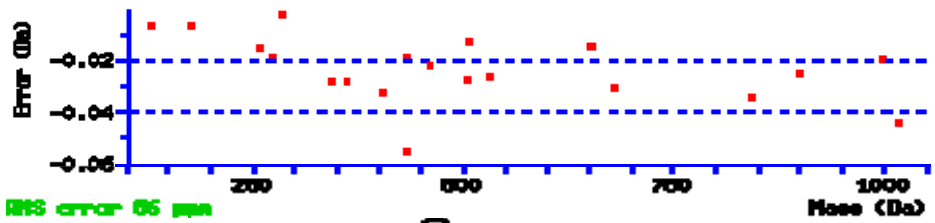


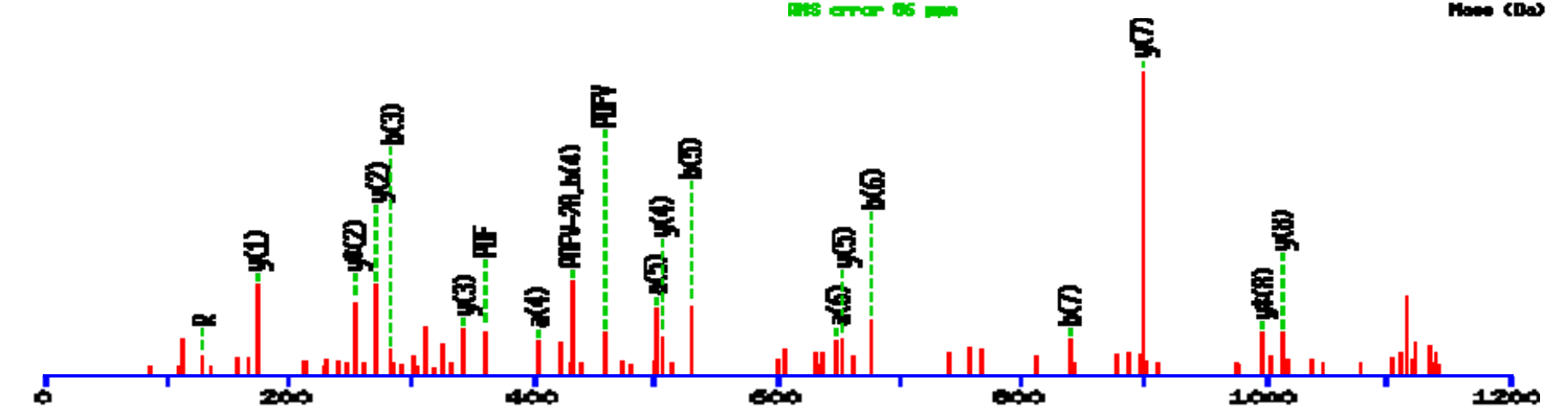
MALDI-MS(MS)^2



In Gel Digest (MS)

MS/MS (1182 ion)

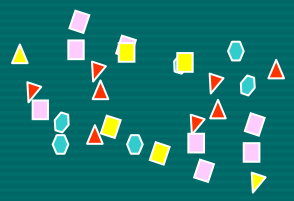




Directed or Non-Directed Proteomics?

(the road to global proteomics!)

Complex Protein Mixture



Chemical or Enzymatic Digestion

Multidimensional Separations Affinity, CaP-LC

MS/MS Analysis

Computational Time Extensive

Multidimensional Separations Affinity, 2D, HPLC

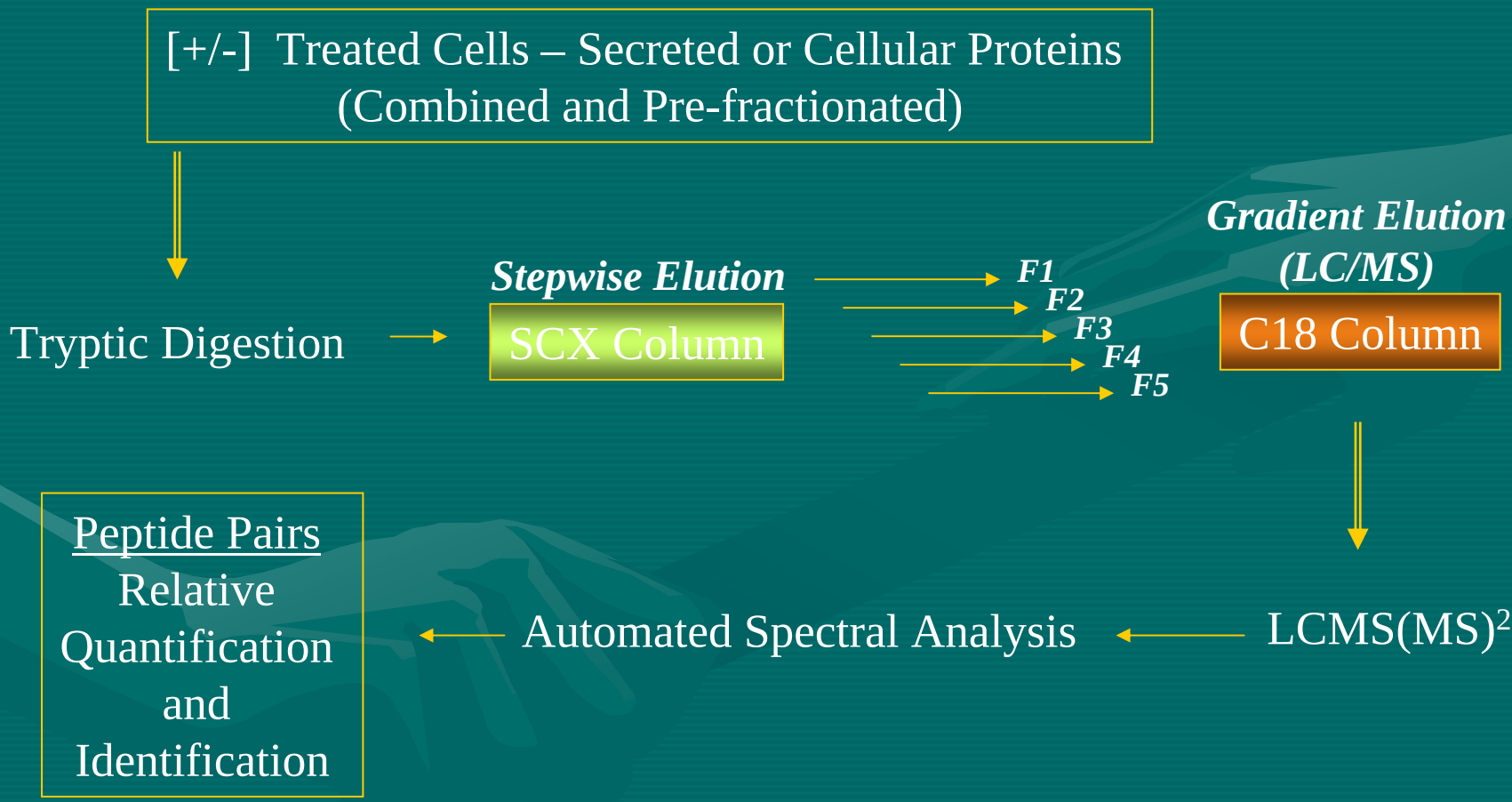
Pure Protein

Chemical or Enzymatic Digestion

MS Analysis Computational Time Minimum

Protein ID

Multidimensional Protein Identification Technologies (MuDPIT)



*MudPIT developed in **John Yate's** Lab, Scripps Research Institute*

Data Analysis

- A standard 1D LC-ESI run may have as many as 4,000-6,000 MS files!!
- A MuDPIT run may contain 25,000-60,000 files!!
- While LC-MALDI runs generate far fewer data files, they still contain too-much data to analyze by hand!
- Therefore, automated data analysis is required!!

Data Analysis

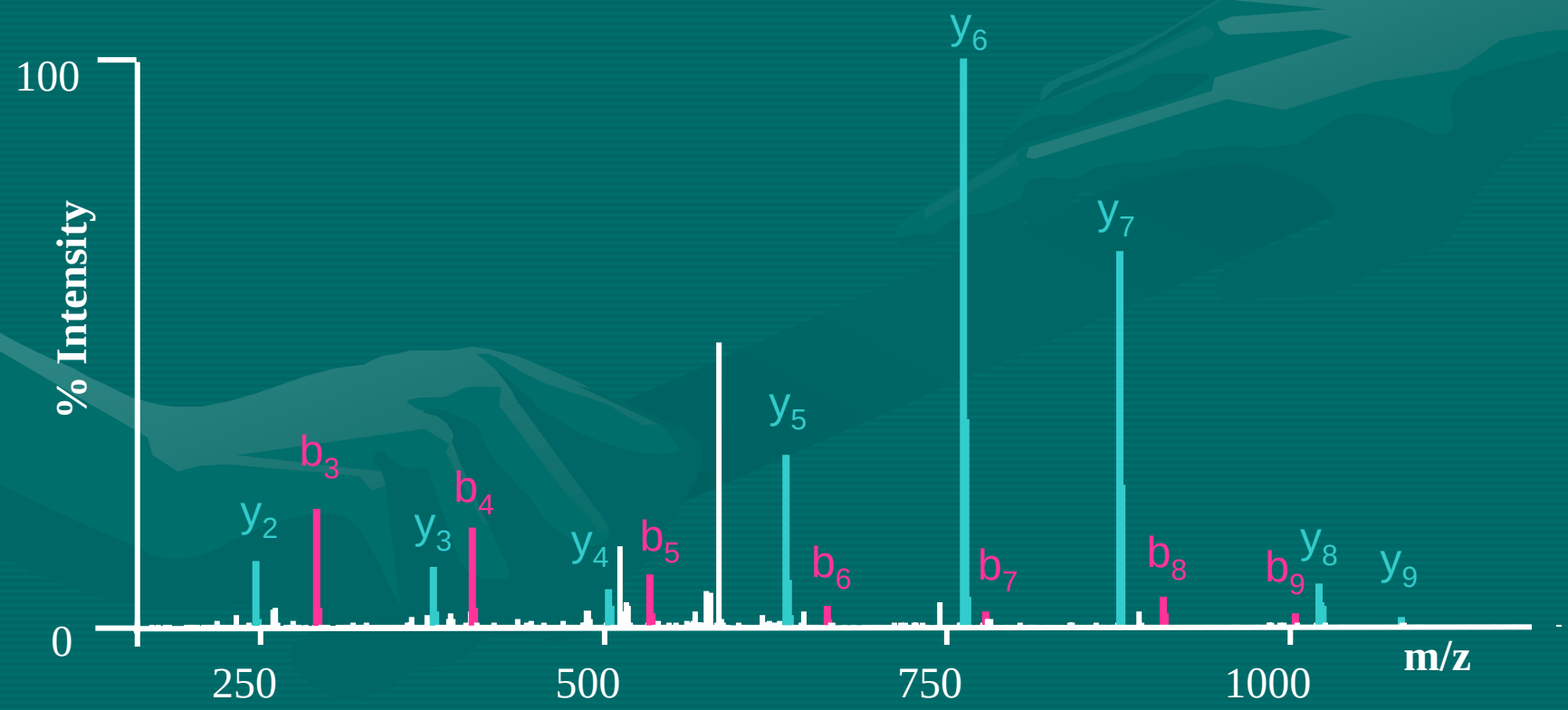
- Common Matching Algorithms;
  - *Sequest, MASCOT, XTandem*
- Automated Denovo Sequence Tools;
  - *Peaks, Rapid Denovo, DenovoX, Mascot Distiller, PepNovo, others…..*
- Statistical Software
  - *Scaffold, Protein Profit, Finnigan, others…*
- Standardizing the Field!
  - *Trans Proteomic Pipeline (Sashimi Project; mzXML based universal software package)*

Matching Algorithms

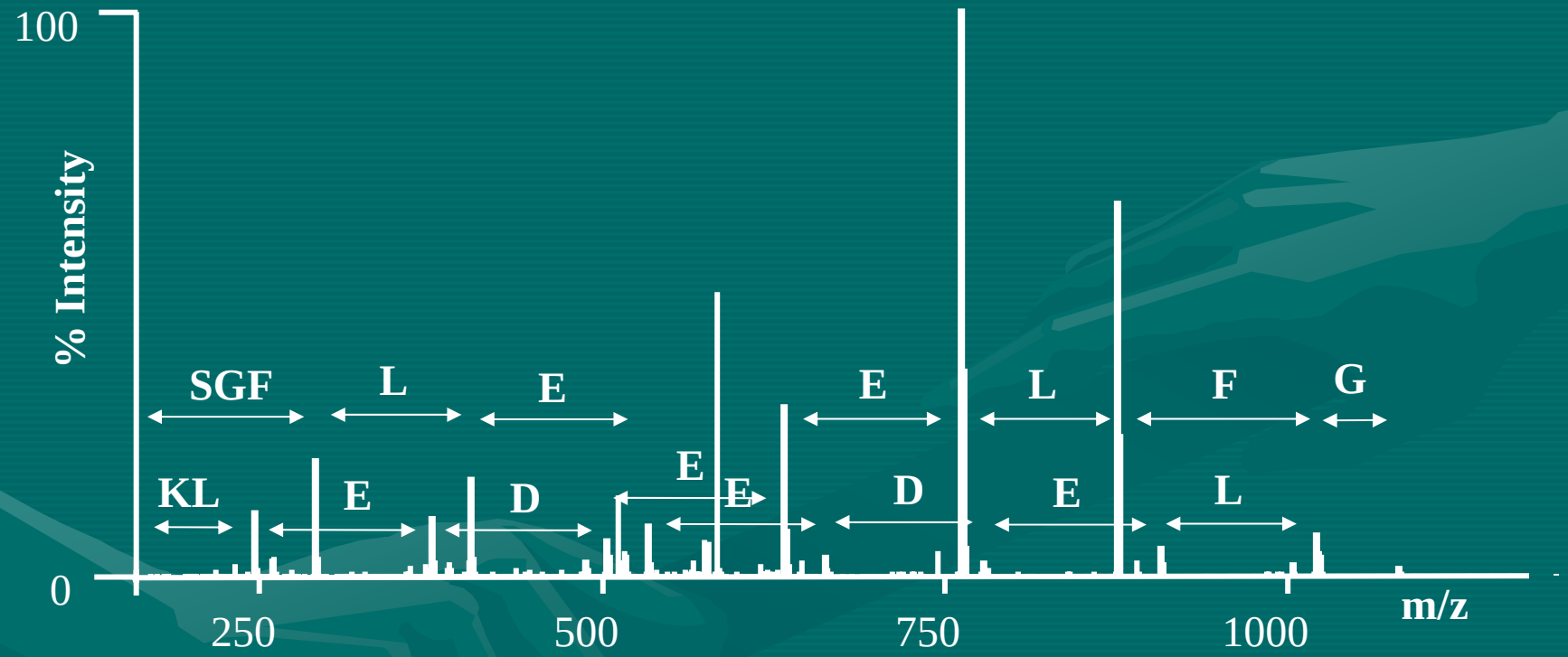
- All matching algorithms are based on generating a score based on "closeness of fit" between the peptides measured in the mass spectrometer and the *in-silico* digestion of known genes or proteins in a database.
  - The two most commonly used databases include: NCBI-NR and Uniprot

Peptide Fragmentation

88	145	292	405	534	663	778	907	1020	1166	b ions
S	G	F	L	E	E	D	E	L	K	
1166	1080	1022	875	762	633	504	389	260	147	y ions

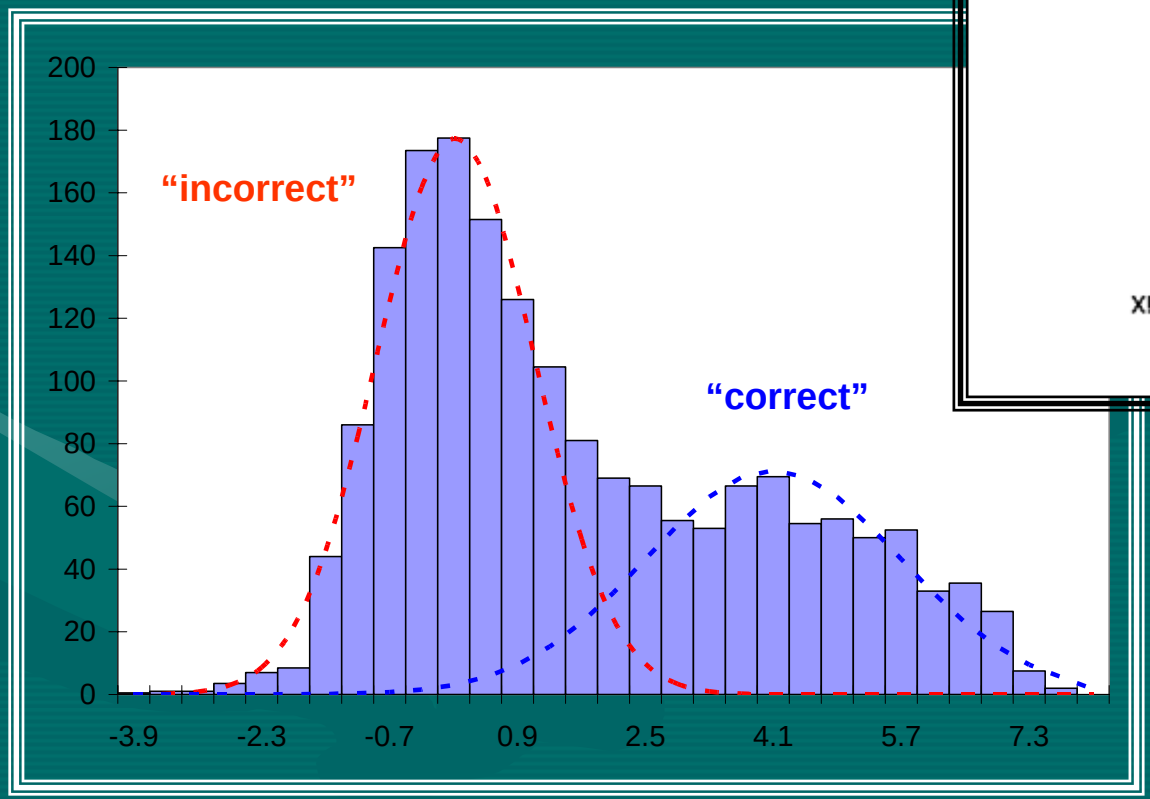


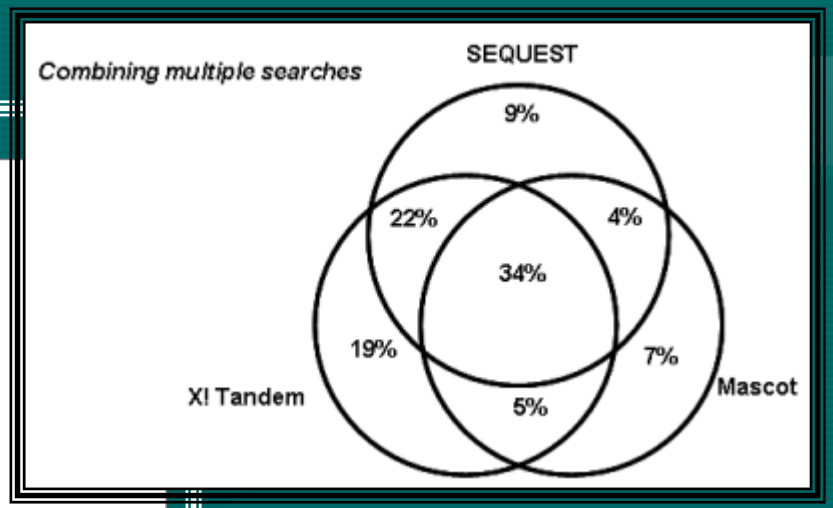
De Novo Interpretation



Generating a Universal Score

Mixture of distributions

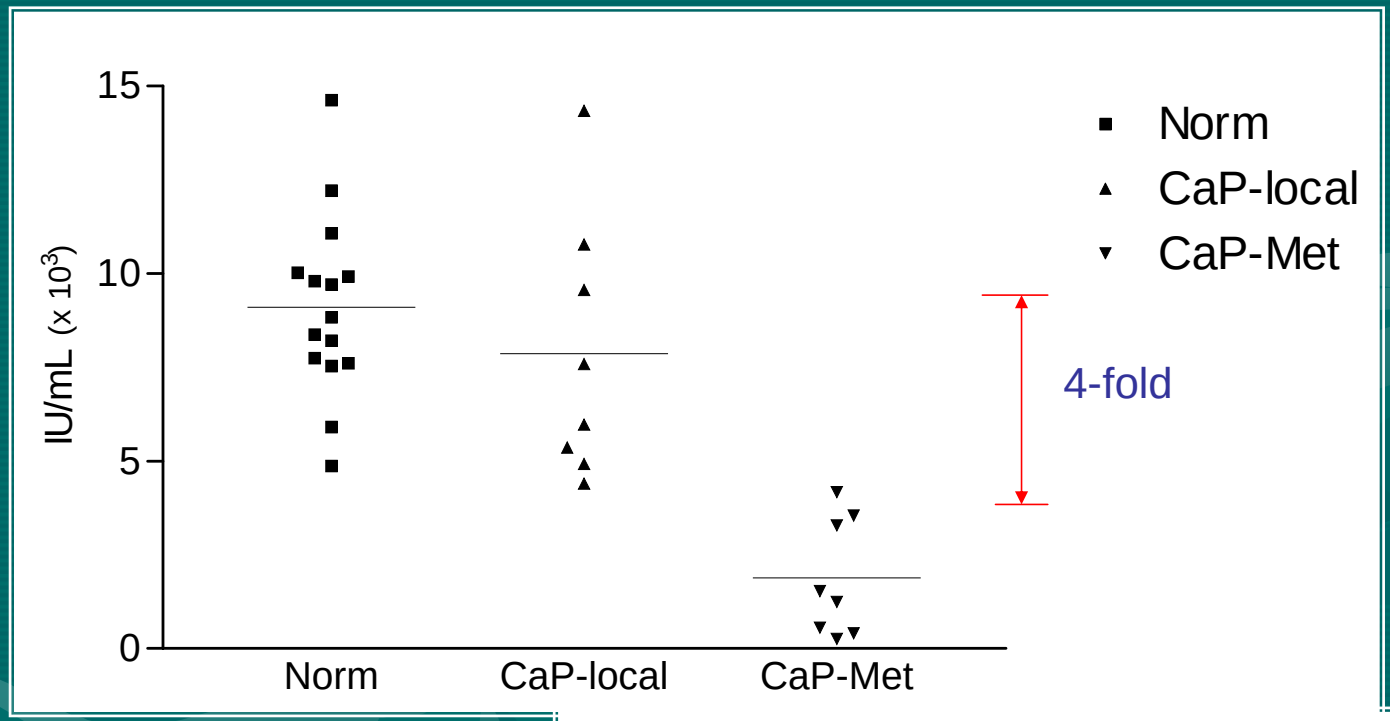




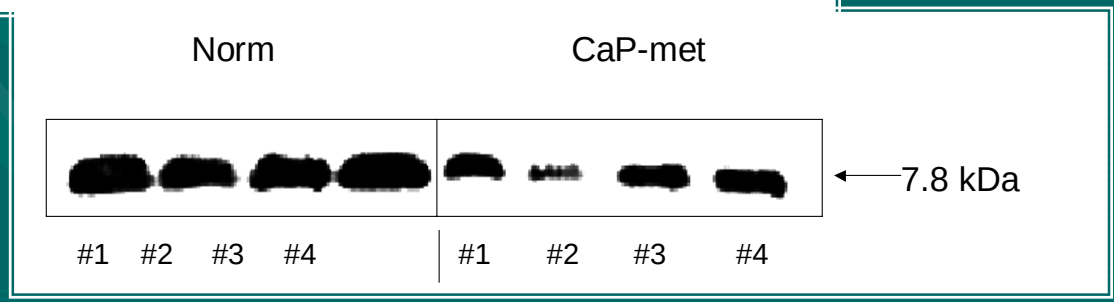
So….We Have ID's, Now What?

- Validation…….
- Validation……..
- Validation……

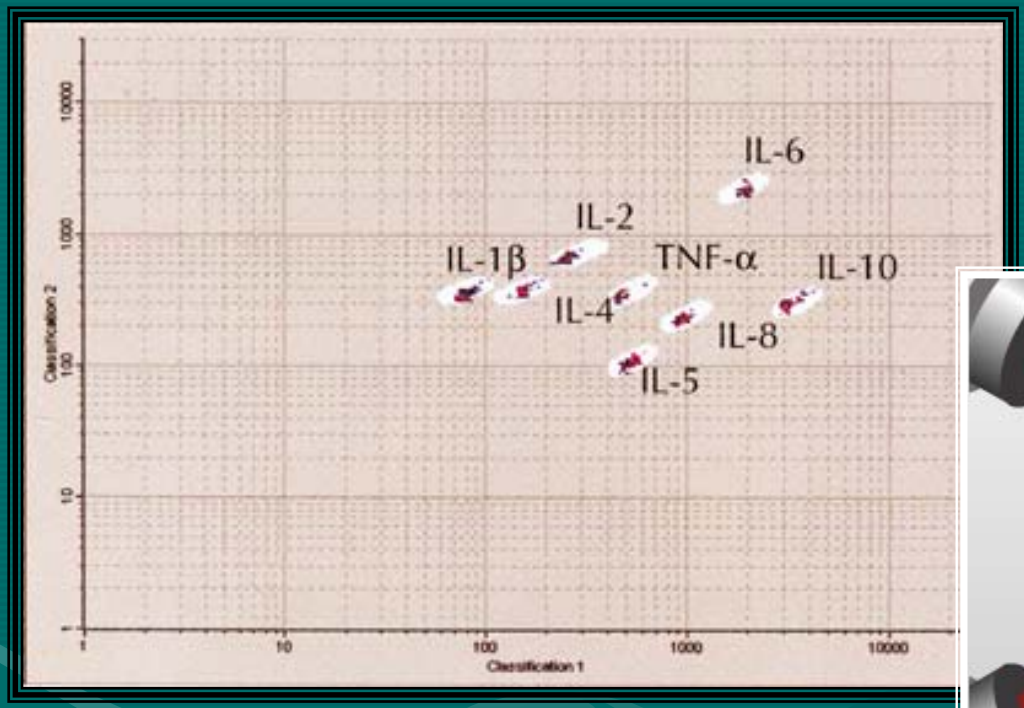
(CXC4 – Western & ELISA)

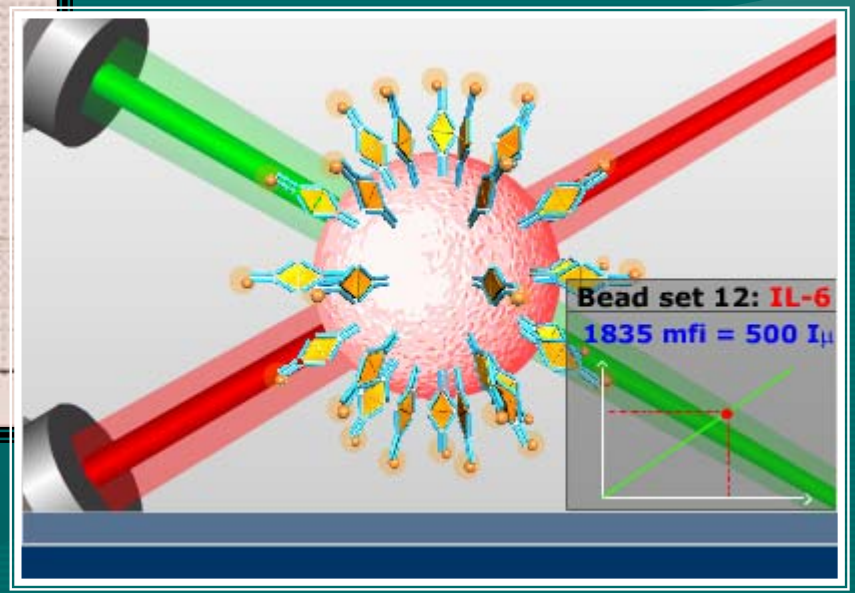


Western Blot CXC4 →



HTP Quantitative Validation





Multiplex Bead Assay for Cytokines

The highlighted area represent populations of fluorescent beads, distinctively labeled, and carrying capture antibodies for sandwich assay of different cytokines.

All detection antibodies carry the same fluorophore, which is read in a third channel to quantify sample cytokine concentration

Summary

- Whether or not you do the MS work yourself……..
  - Know the specifics……!
  - Know the limitations…..!
- Sample Prep is always important…..but which instrument you have access to is also important!
- Similarly important….how is your data analyzed??
  - Denovo?
  - Matching Algorithm??
  - Mix of techniques??
  - What are the cut off points and does it make sense?
- Keep in mind that validation must be part of your workflow!!
- So, make sure you have confidence in your choices before going forward!

Useful Links!

- i-mass.com
- spectroscopynow.com
- expasy.ch/tools
- cprmap.com
- psidev.sourceforge.net
- prospector.ucsf.edu
- jeolusa.com/ms/docs/ionize.html
- asms.org (become a member!)
- hupo.org
- matrixscience.com
- proteomecenter.org/software.php

ionsource.com
bruker.com
thermo.com
appliedbiosystems.com
shimadzu.com
luminexcorp.com


Questions??