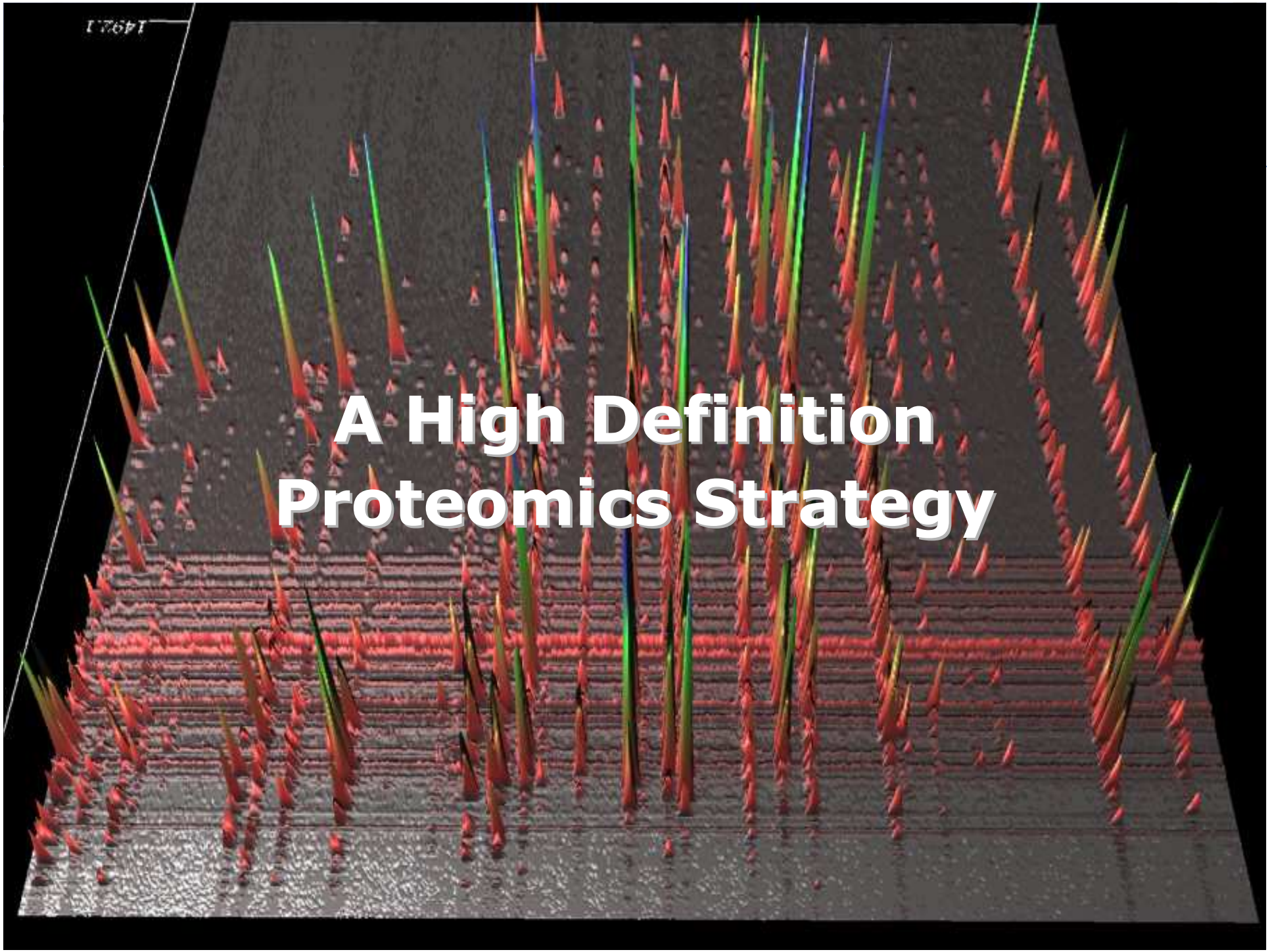

A High Definition
Proteomics Strategy

OVERVIEW

…a High Definition (LC/MS) Proteomics Strategy

- **Data Acquisition**
  - Sample Complexity
  - LC/MS/MS Band-Width
- **Protein Identification**
  - Comprehensive Ion Accounting
- **Protein Quantification**
  - Relative quantification
  - Absolute Quantification
- **Intact Protein Analysis**
  - Protein-Protein Complexes
  - …"Interactomics"

The Next 10 Years of Proteomics

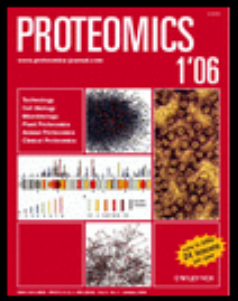
…require stringent data generation & analysis

Waters
THE SCIENCE OF WHAT'S POSSIBLE.™

" …the degree of stringency required in proteomic data generation and analysis appears to have been underestimated…

…As a result, there are likely to be numerous published findings that are of questionable quality…"

M.R. Wilkins *et al.* (2006)
Guidelines for the Next 10 years of Proteomics.
Proteomics, 6, 4-8



Considerations in Discovery Proteomics

Experimental Design

- Biological Context / Diversity
- Complexity / Fractionation
- Copy No. / Dynamic Range
- ID / QUAN Strategy
- Informatics / Statistics

Analysis

- IEF / 1D Gel / 2D Gel
- 1D / 2D HPLC On/Off-Line
- Protein Load Optimization
- MALDI / ESI
- Data Interpretation

Data

- Protein ID (FDR)
- Sequence Coverage
- PTM Survey
- Replication Rate (n>1)
- Quantification (CV)

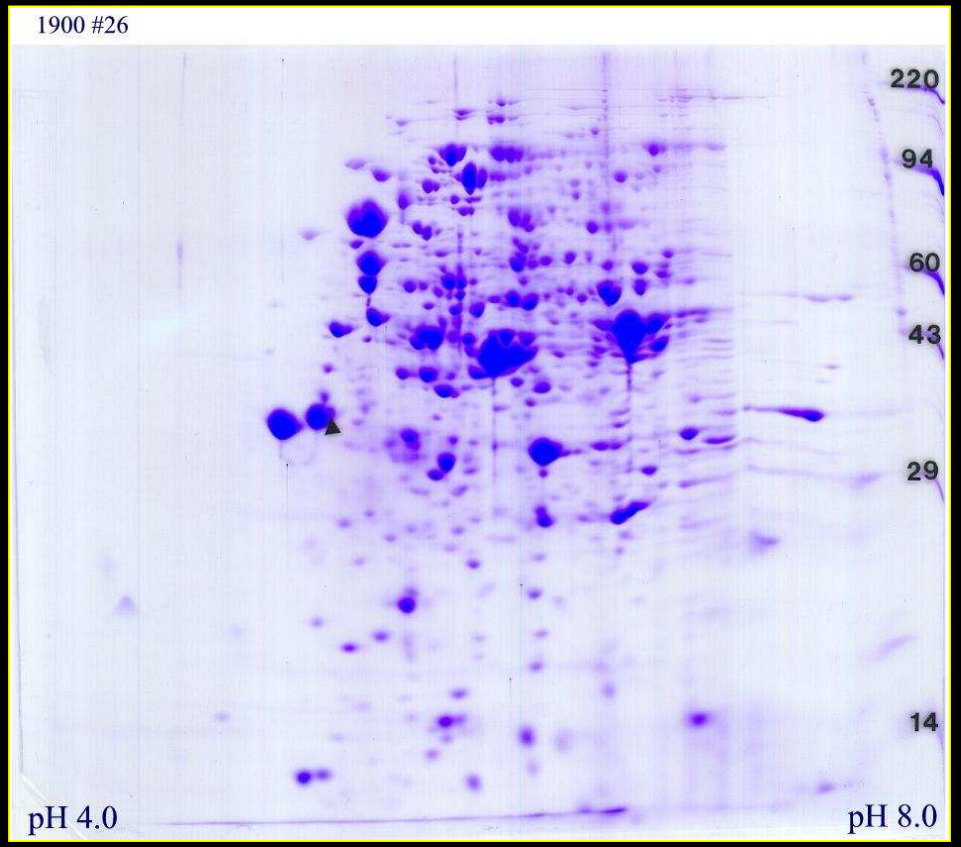
2D PAGE Gel Analysis

In general for complex samples such as cell lysates run on Kendric Labs Standard sized gel, for Coomassie blue or Sypro ruby staining we suggest a protein load of:

> 100 µg

Coomassie blue stain gives a reasonably dark spot at 1 µg and a discernable one at:

100 ng



Escherichia coli.
2DE conditions: standard size gel, **200 µg** protein loaded, pH 4-8 ampholines, 10% slab gel, Coomassie blue stain.

www.kendricklabs.com/bacteria&yeast.htm#E.%20coli

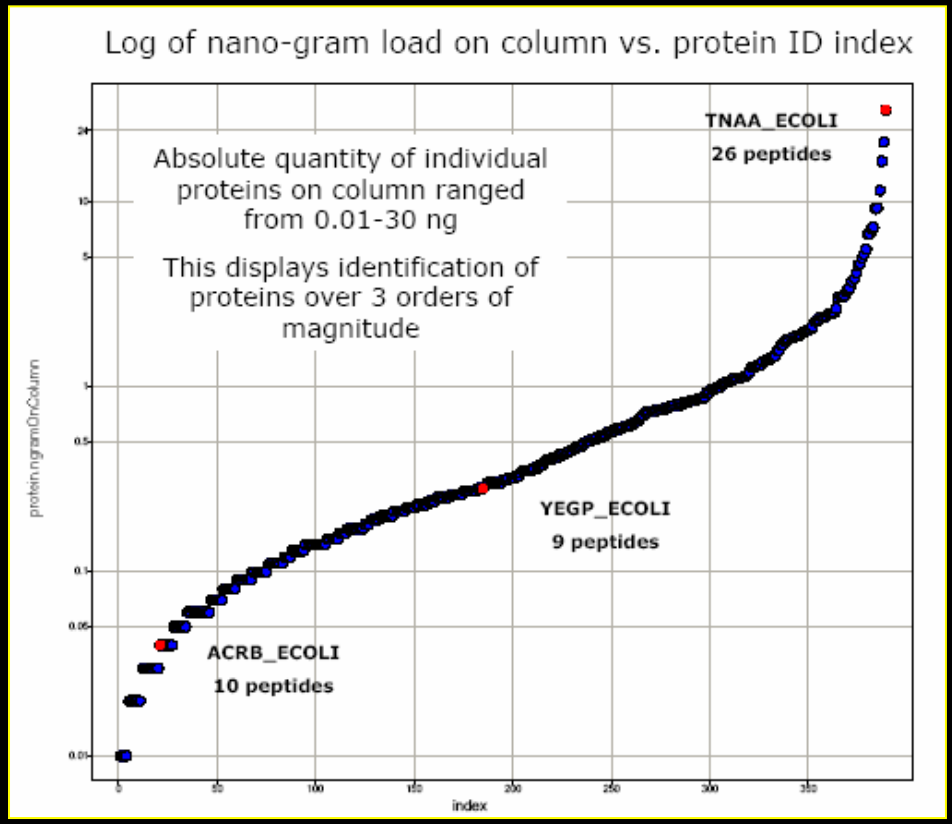
LC/MS Analysis

For LC/MS analysis of complex samples such as cell lysates run on Waters 75µm internal diameter UPLC columns we require an optimum protein load of:

500ng

The absolute quantity of proteins identified and quantified typically range between 30ng and:

10pg

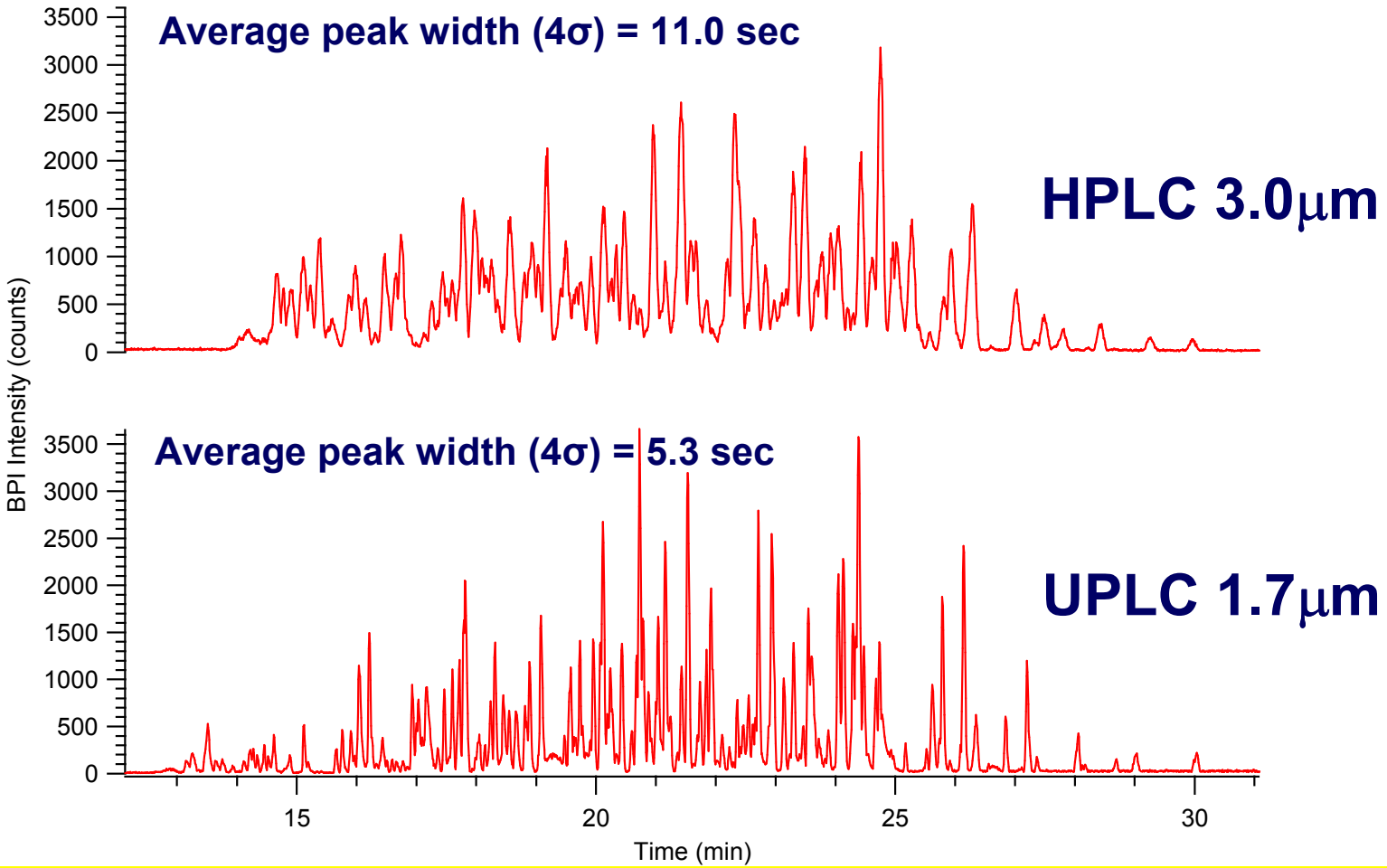


Escherichia coli.
Standard digestion and on-line UPLC/MS^E analytical protocol, 500 ng protein loaded.

Internal Standard: 75fmol Phosphorylase B

Ultra Performance LC

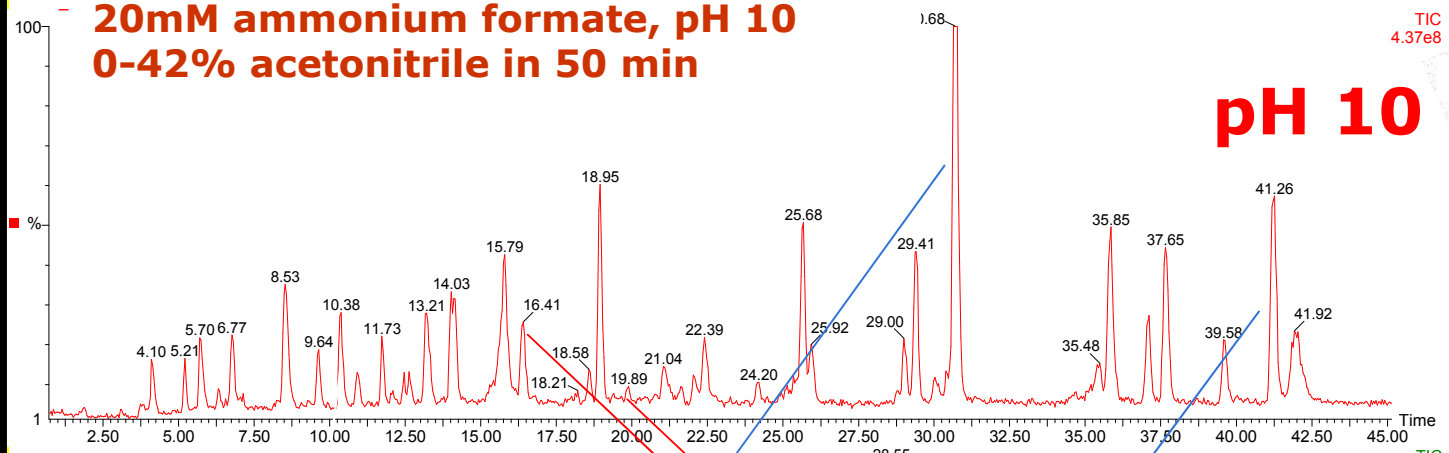
...increases Sensitivity, Resolution (& Speed)



5%B to 55%B in 30 minutes, 0.5 sec/spectra.
5 protein MassPREP™ digest standard mixture, 20 fmol each of enolase, phosphorylase b, hemoglobin, ADH, BSA

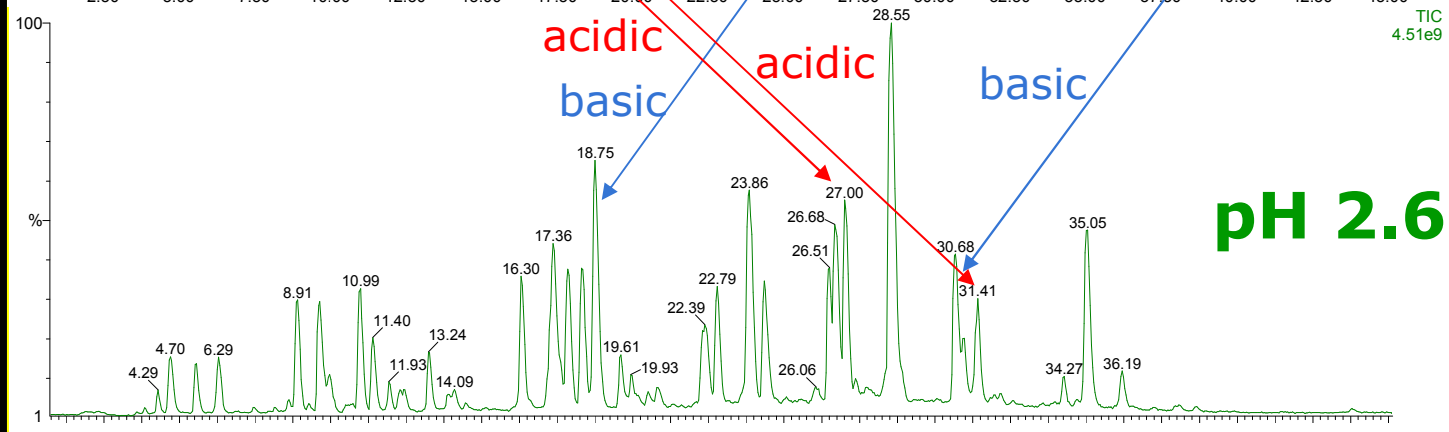
2D RP/RP UPLC (High pH/Low pH)
...maximizing peak capacity

20mM ammonium formate, pH 10
0-42% acetonitrile in 50 min

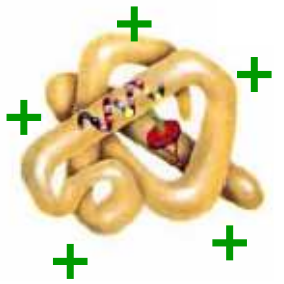


pH 10

acidic
basic
acidic
basic



pH 2.6

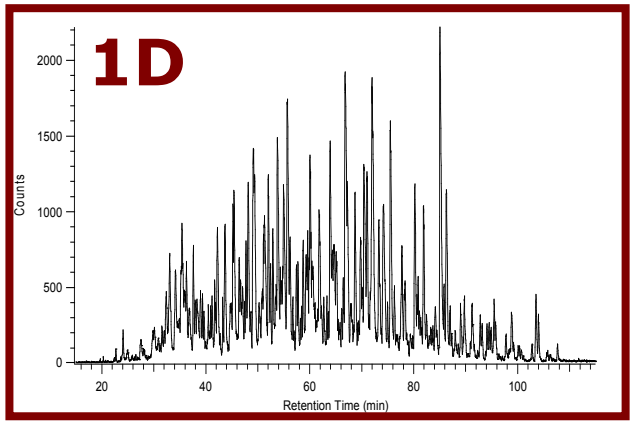


0.2% Formic acid, pH 2.6 0-42% acetonitrile in 50 min

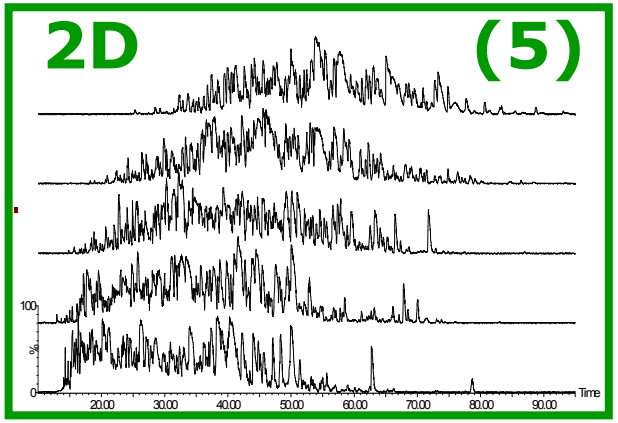
Gilar M. et al, J. Sep. Sci. 2005, 28, 1694-1703

2D RP/RP UPLC Peak Capacity

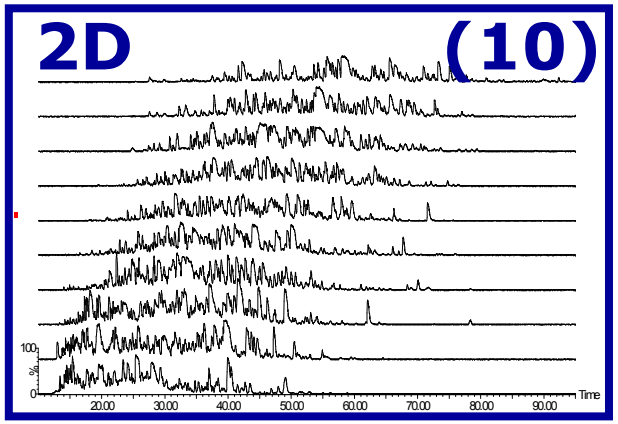
...Comparison of 1D PR & 2D RP/RP Separations



365 Proteins
2874 Peptides



695 Proteins
7961 Peptides

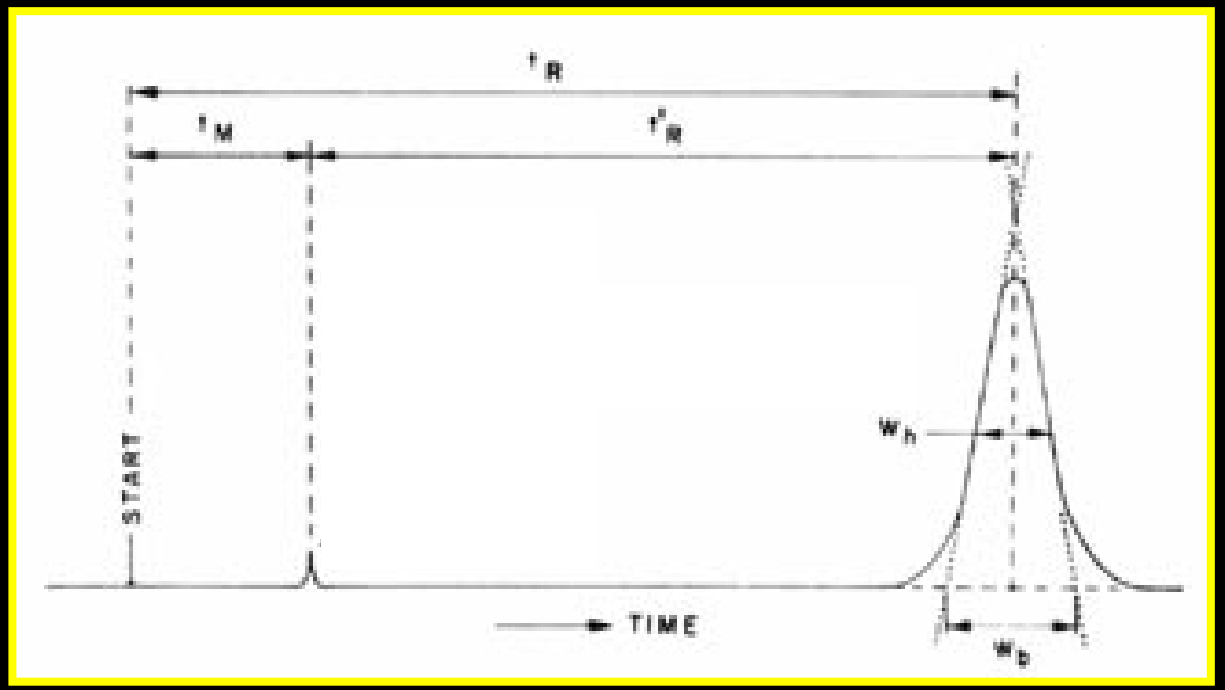


778 Proteins
9415 Peptides

massPREP Ecoli standard
IDENTITY^{E} Search Engine
3 Replicates (ID ≥2 / 3)

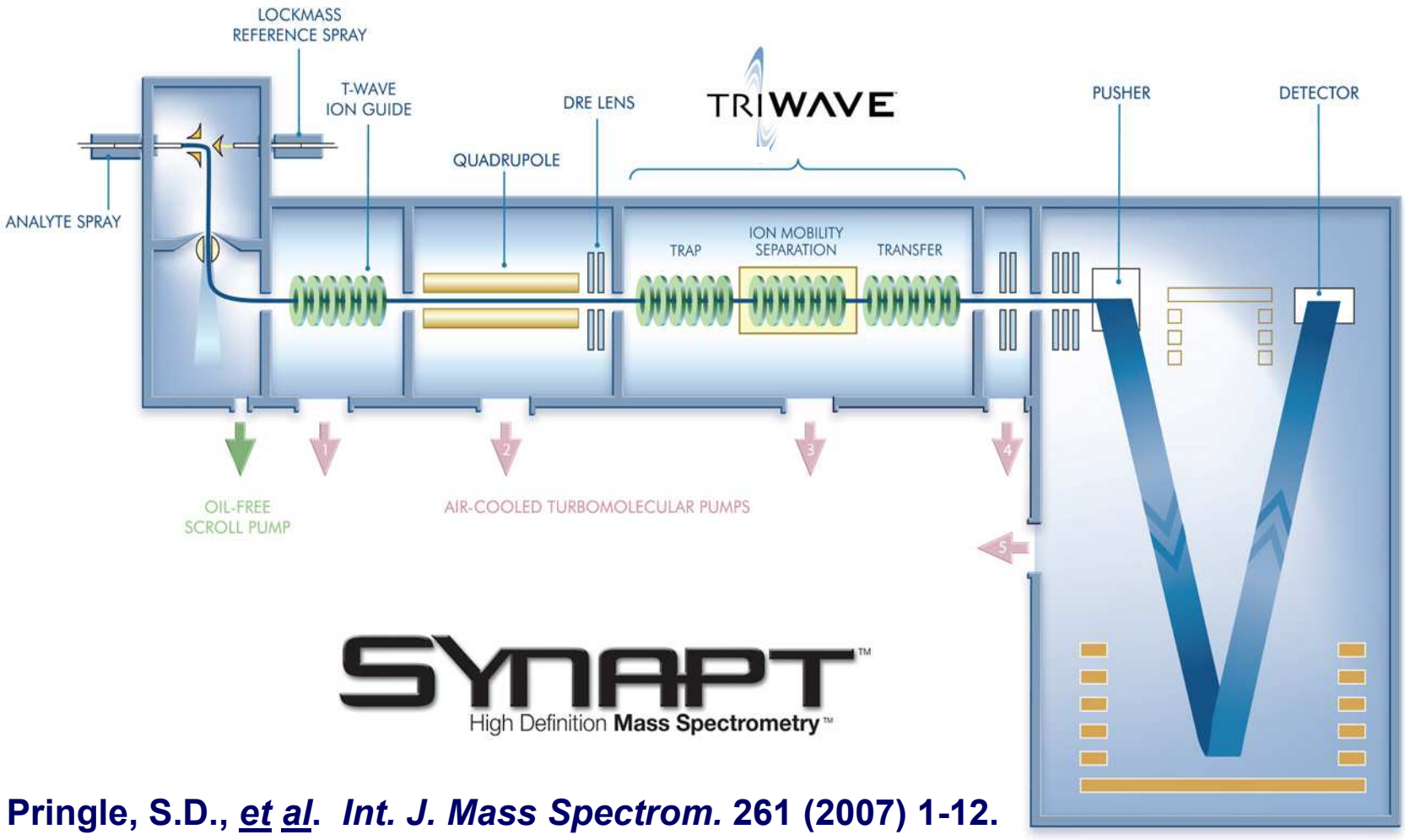
No. of Data Points Per Peak ?

- **15 to 20 points across a chromatographic peak are required for good quantification.**
  - **If you have fewer points you will not be able to describe the peak adequately and may lose information (e.g. peak top)**
  - **Reproducibility is negatively affected with fewer points and you will observe RSD's increasing to unacceptable values.**



http://www.ionsource.com/tutorial/msquan/tips.htm

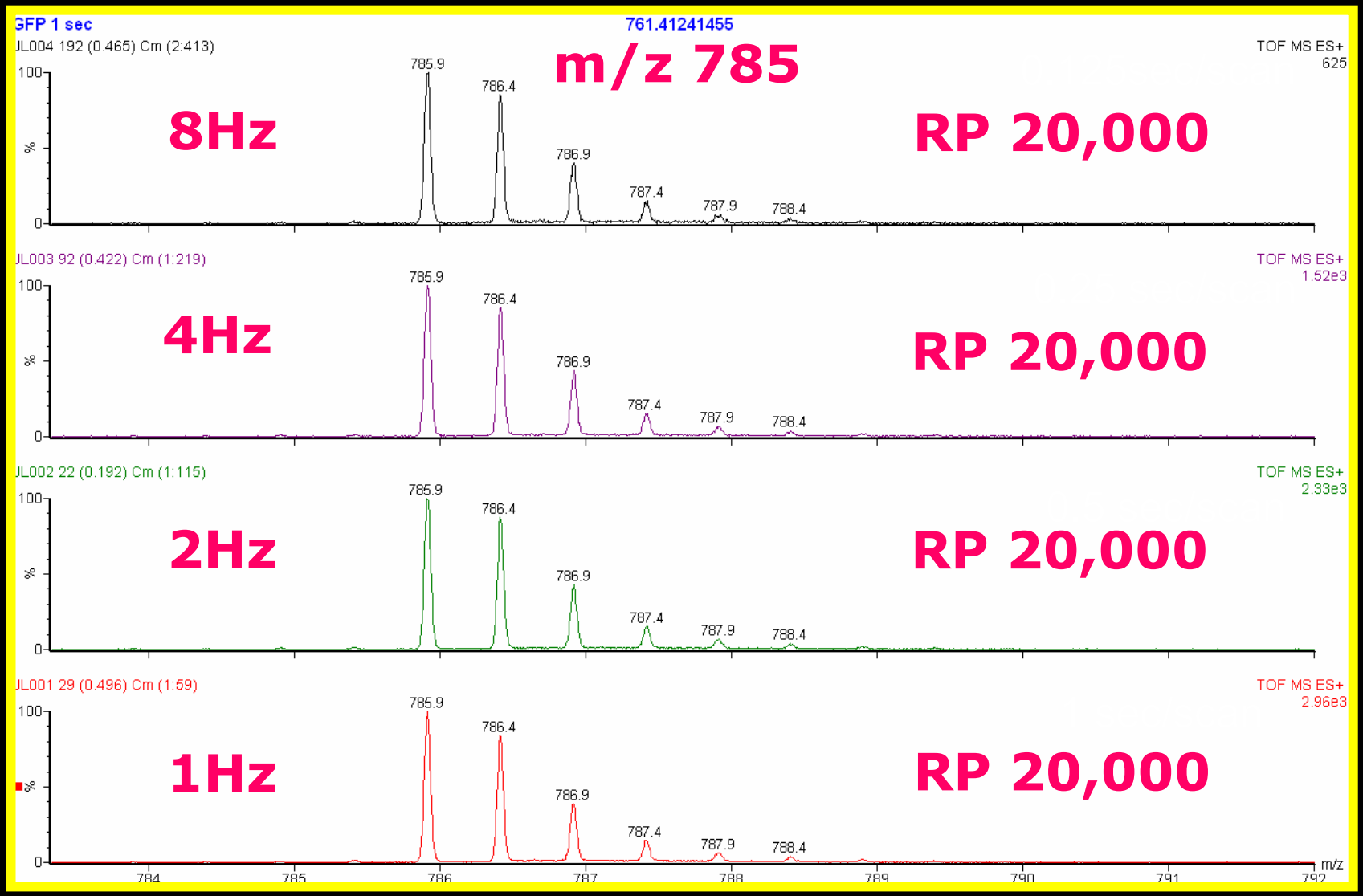
High Resolution & Exact Mass MS/MS



Pringle, S.D., *et al*. *Int. J. Mass Spectrom.* 261 (2007) 1-12.

Dynamic Mass Resolution of oa-TOF MS

...Resolution is Independent of speed



Dynamic Mass Resolution...

m/z 785

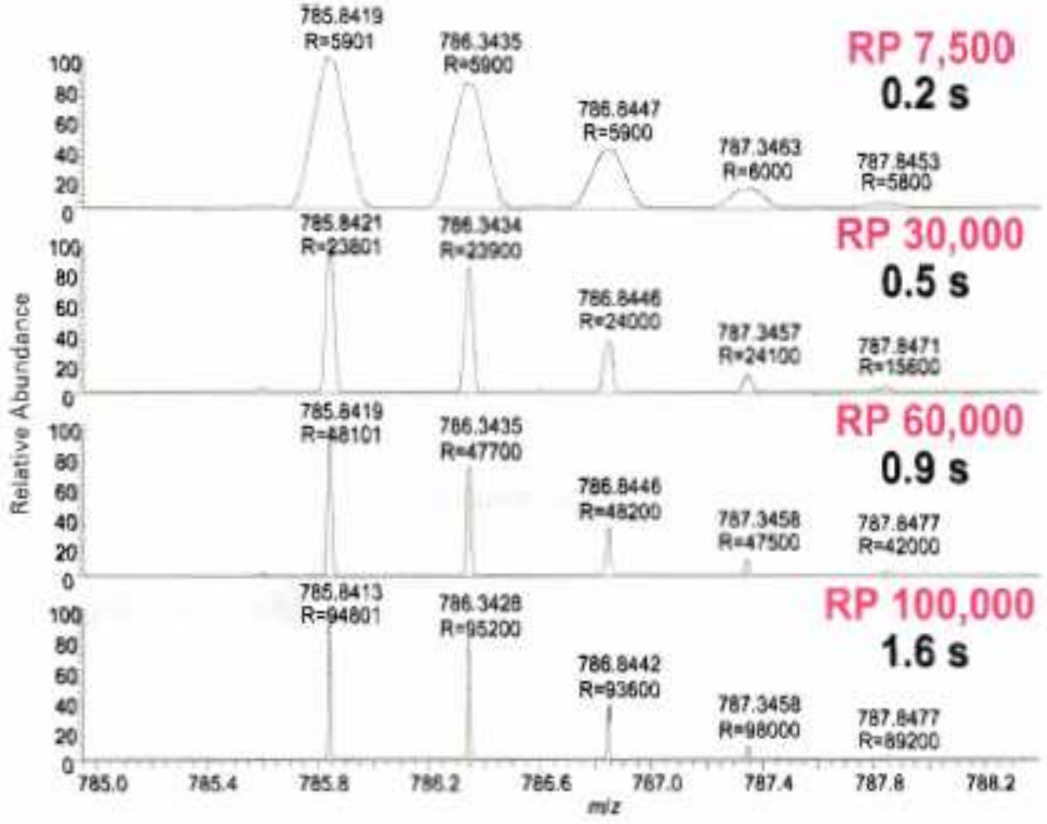


Figure 4. Resolving power of the orbitrap mass spectrometer achieved for a doubly charged species of EGVNDNEEGFFSAR peptide within specified times. Resolving power is referenced to *m/z* 400.

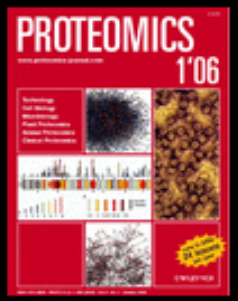
Orbitrap Mass Analyser Overview and Applications in Proteomics.

Michaela Scigelova & Alexander Makarov. Practical Proteomics 1-2/2006

Undersampling
...analytical incompleteness in "classic" LC/MS/MS

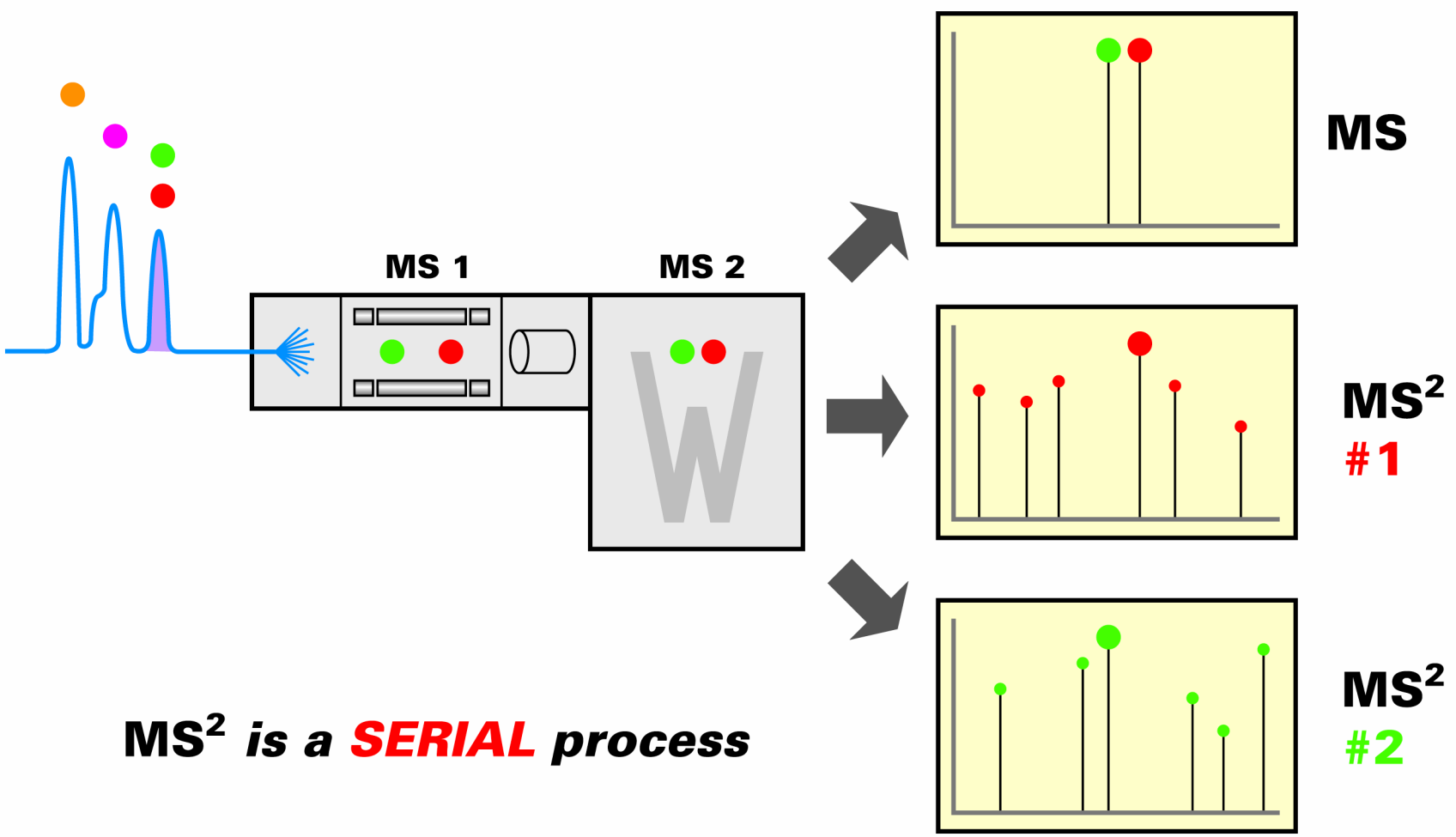
"Analytical incompleteness refers to a phenomenon where a technique used for the analysis of complex mixtures of peptides may only yield information for a fraction of relevant peptides in any single run."

M.R. Wilkins *et al.* (2006)
Guidelines for the Next 10 years of Proteomics.
Proteomics, 6, 4-8



"Classic" LC/MS/MS

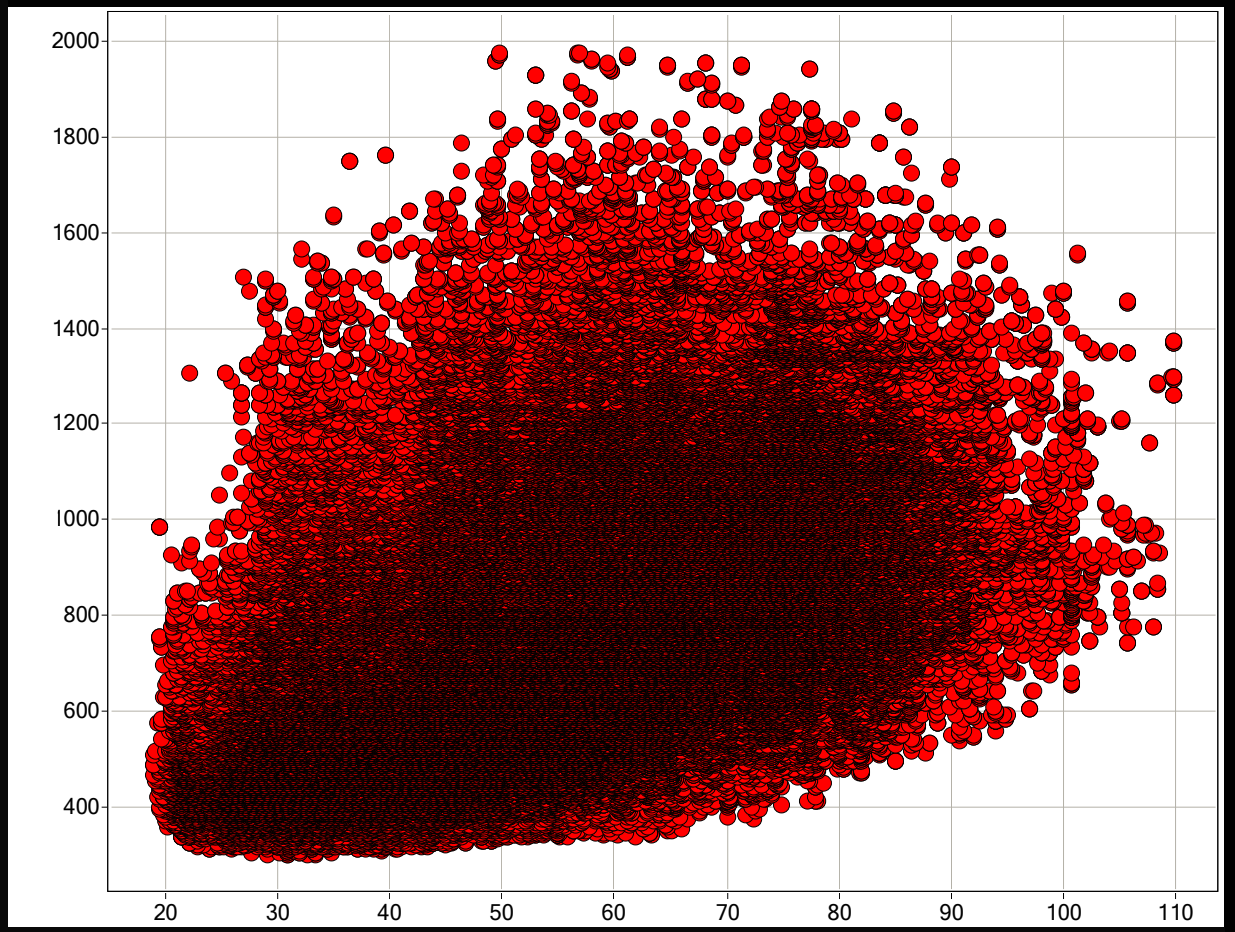
...involves stochastic selection of precursor ions



MS² *is a SERIAL process*

Complex Discovery Samples

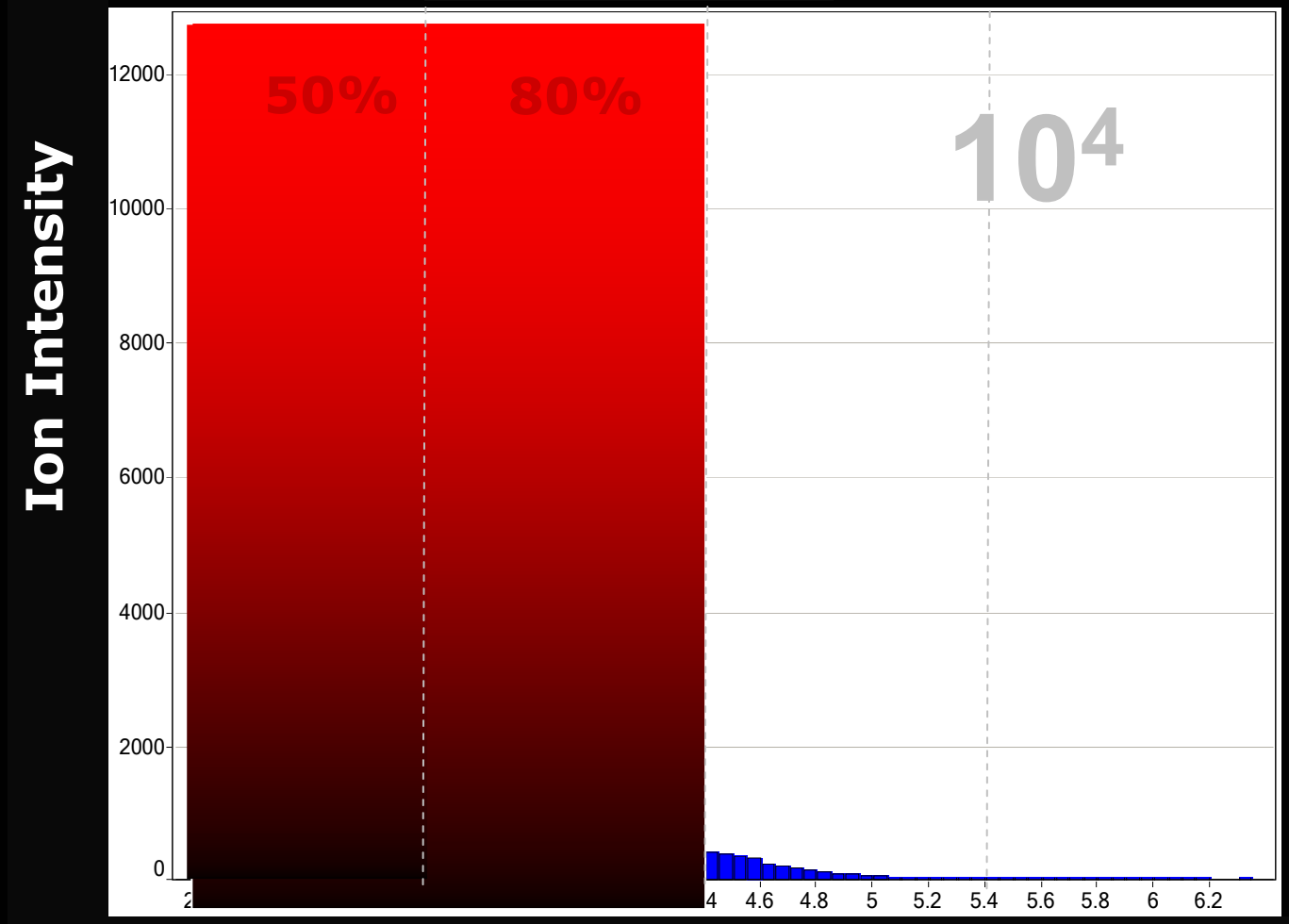
...require higher sampling efficiency/bandwidth



Digested Whole Cell Lysate of *E. Coli* : 164,533 Ion Detections

Complex Discovery Samples

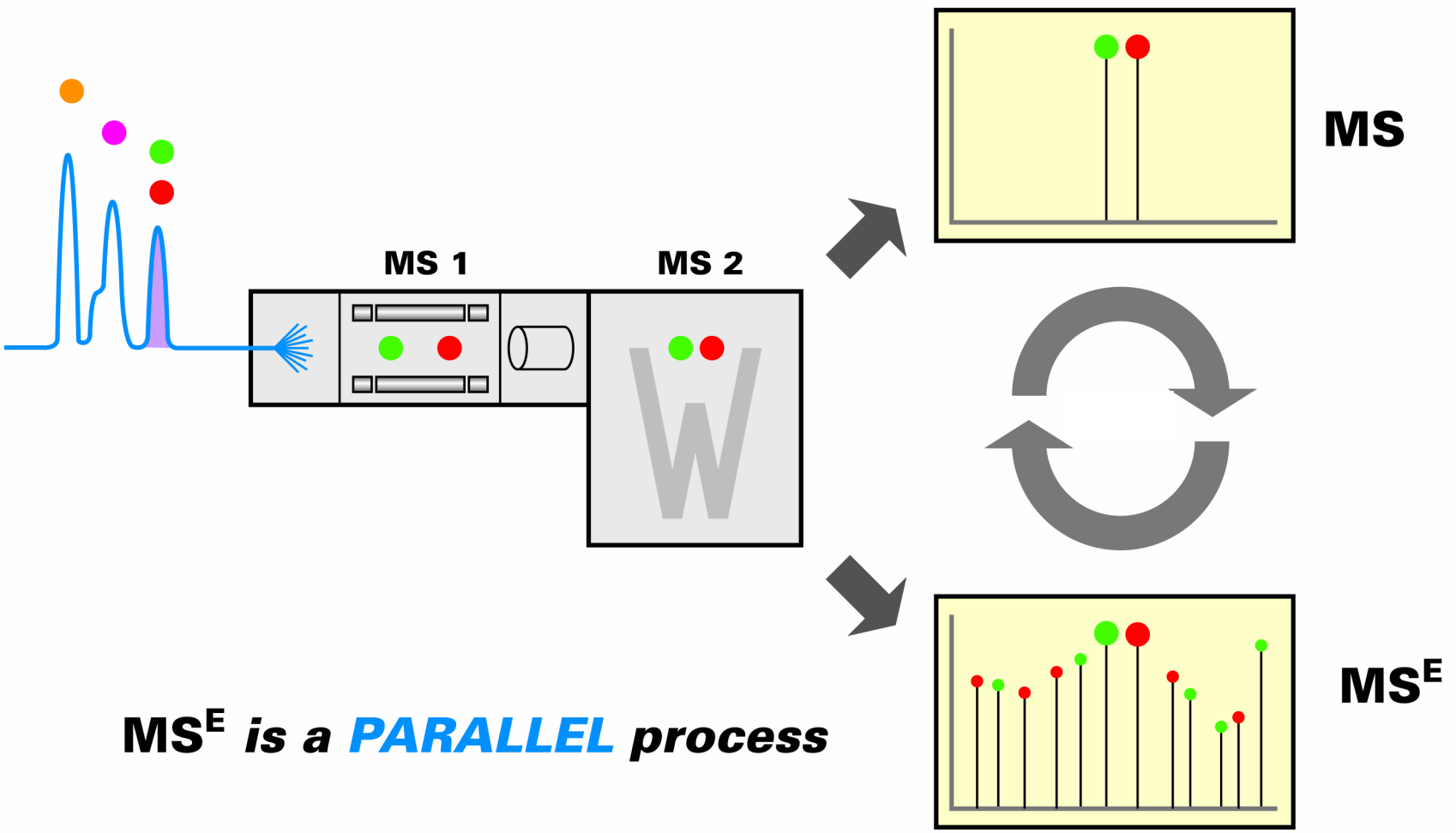
...90% of the information is in the grass (≤ 1% intensity)



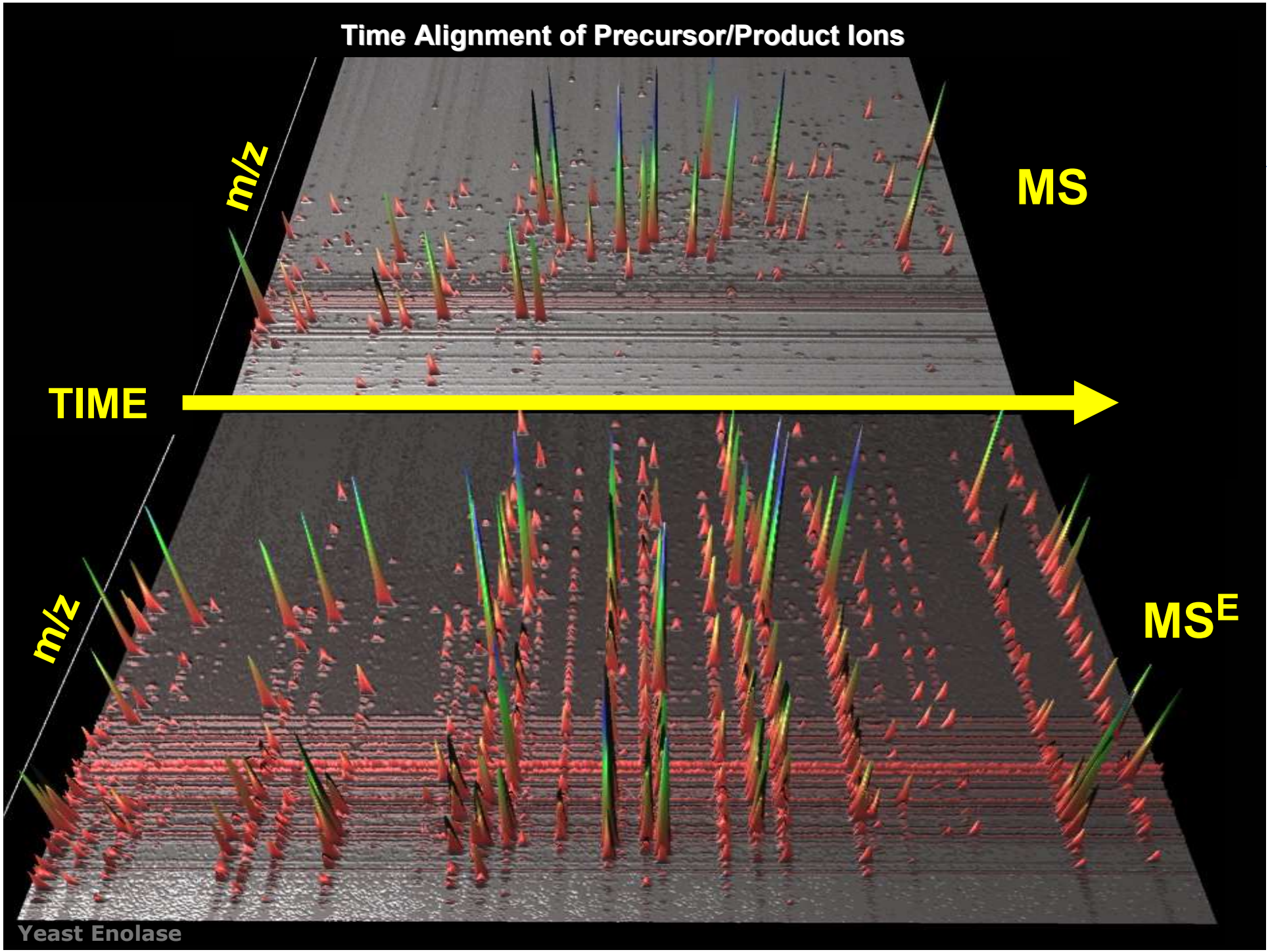
Digested Whole Cell Lysate of *E. Coli* : 164,533 Ion Detections

UPLC/MS^E Data Acquisition

...enables predictable/reproducible peptide sampling

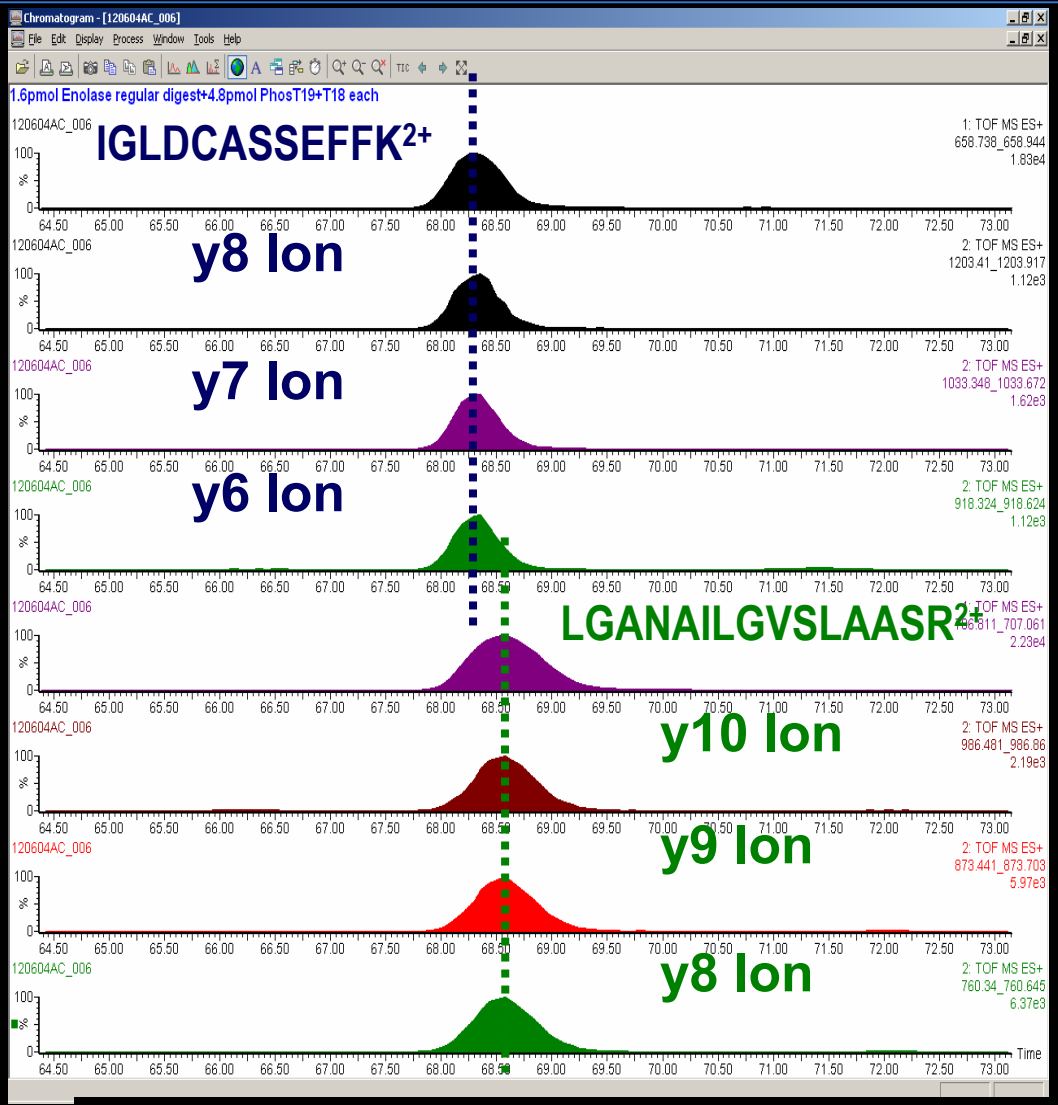


MS^E ***is a PARALLEL process***

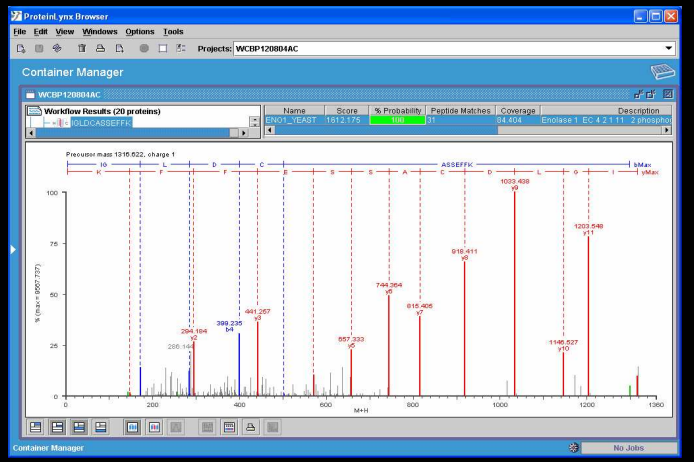

Time Alignment of Precursor/Product Ions
m/z
MS
TIME
m/z
MS^E
Yeast Enolase

Label-Free UPLC/MS^E

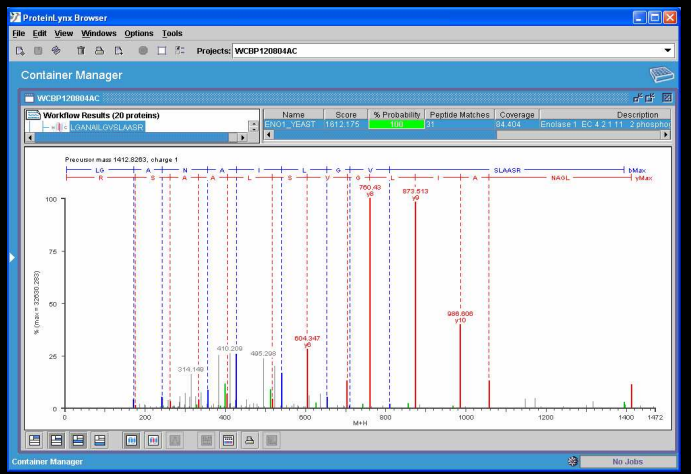
...Time Alignment: Molecular & Fragment Ions



(K)IGLDCASSEFFK(D) Rt= 68.36



(K)LGANAILGVSLAASR(A) Rt= 68.57

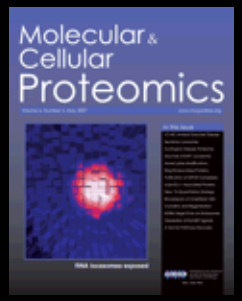


The two peptides exhibit different peak apex times and peak profiles

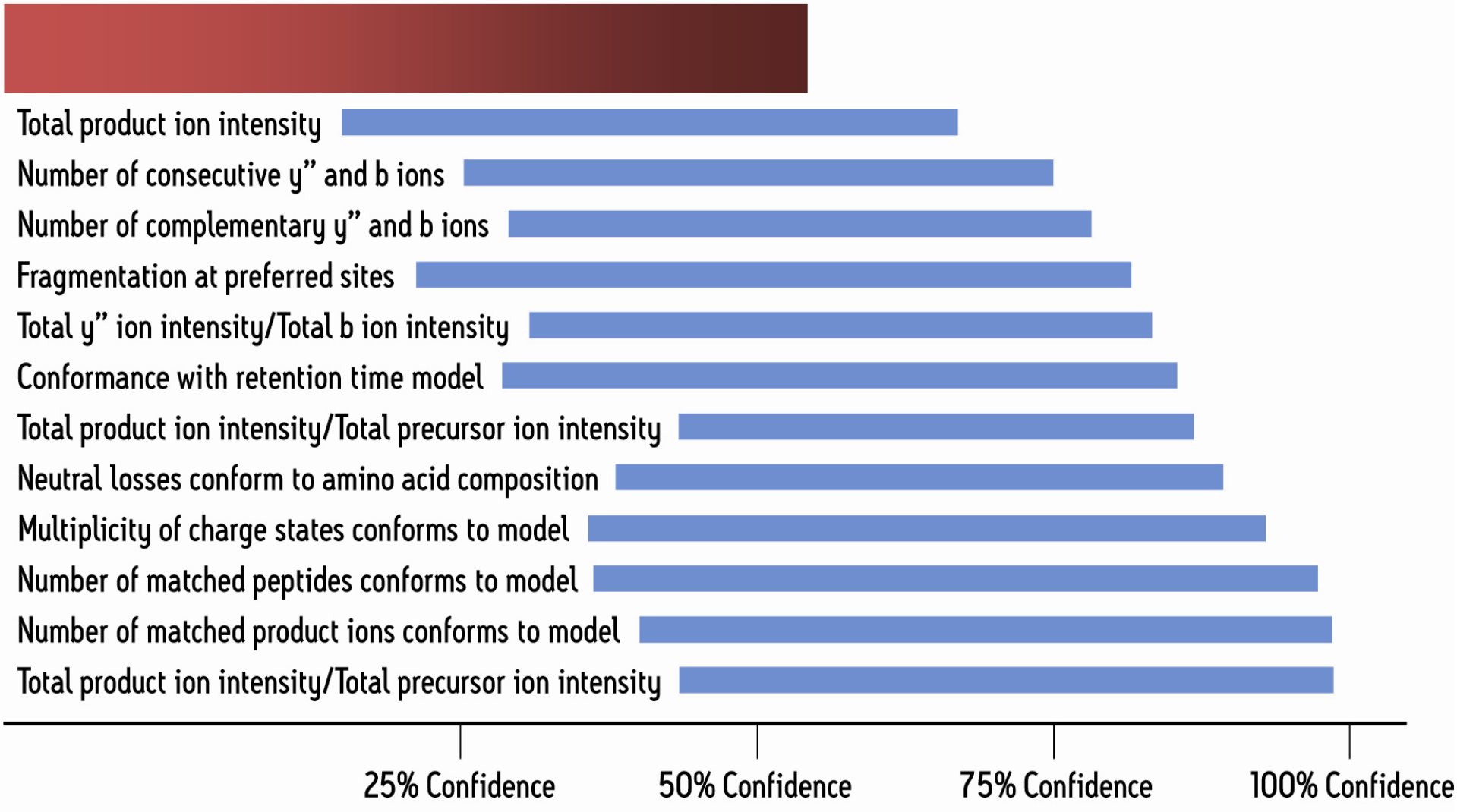
Stringent Data Analysis
...“one hit wonders” and false-positive identification

“…the risk of a false-positive protein assignment is greater when only a single peptide is used to identify a protein...”

Steven Carr *et al*. (2004)
The Need for Guidelines in Publication of Peptide and Protein Identification Data.
***Mol. Cell. Proteomics, 3.6,* 531-533**

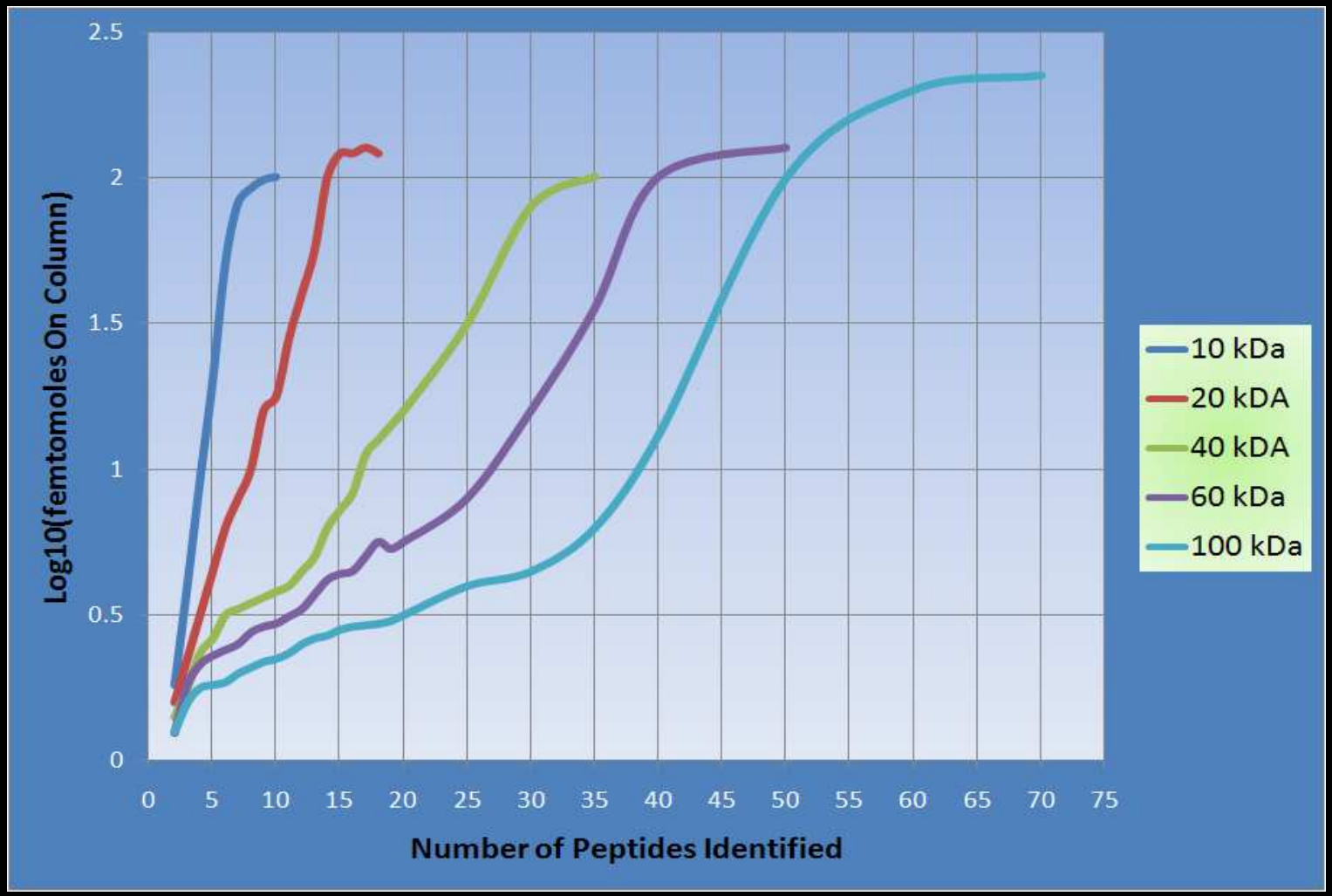


IDENTITY^{E} Search Engine
...Comprehensive Peptide Ion Accounting



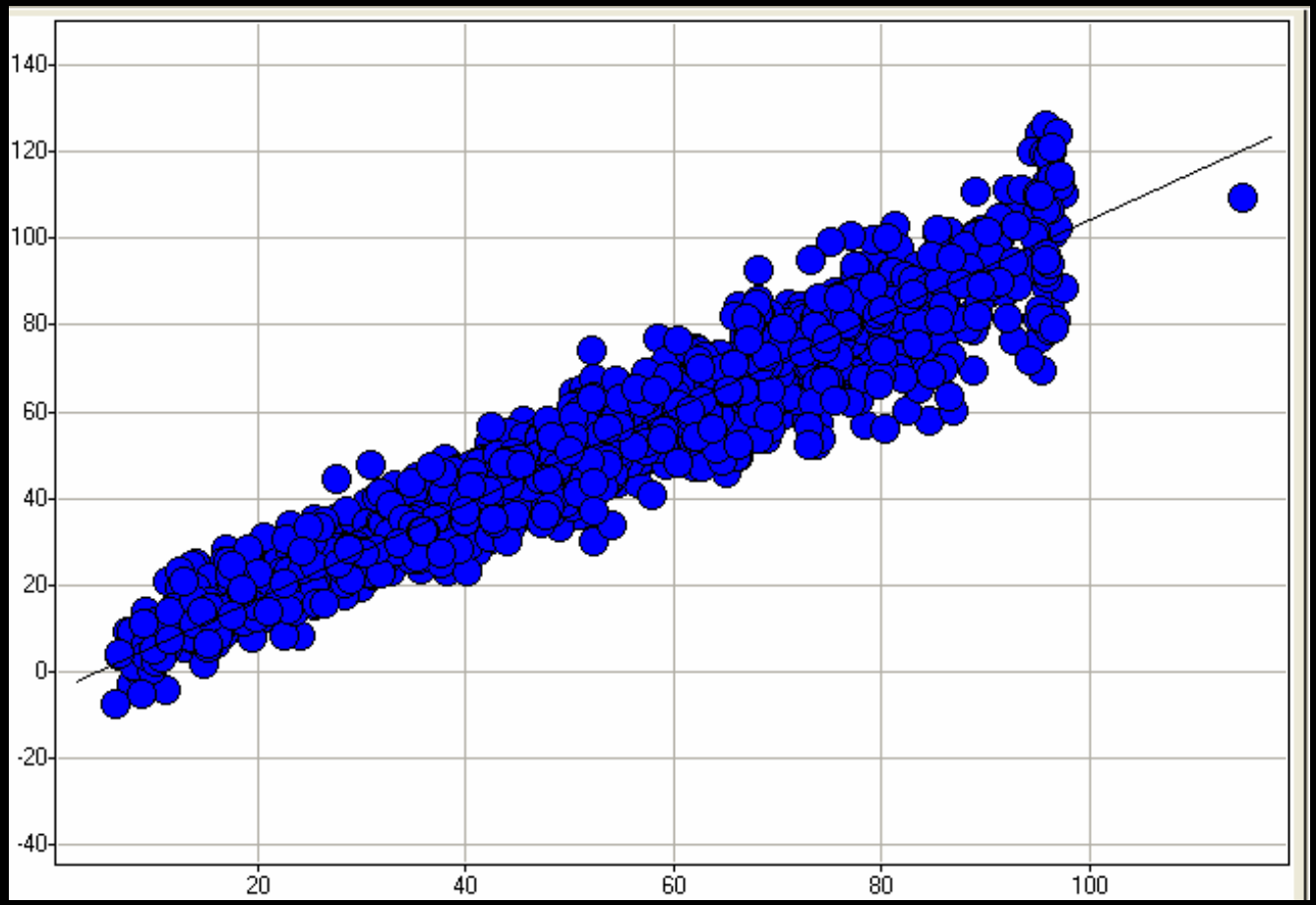
Number of "Identifiable" Peptides

...proportional to protein mass & on-column loading



#3 **The sum product ion intensity (Σ b+y") for a peptide provides the means to predict the total # of detectable peptides.**

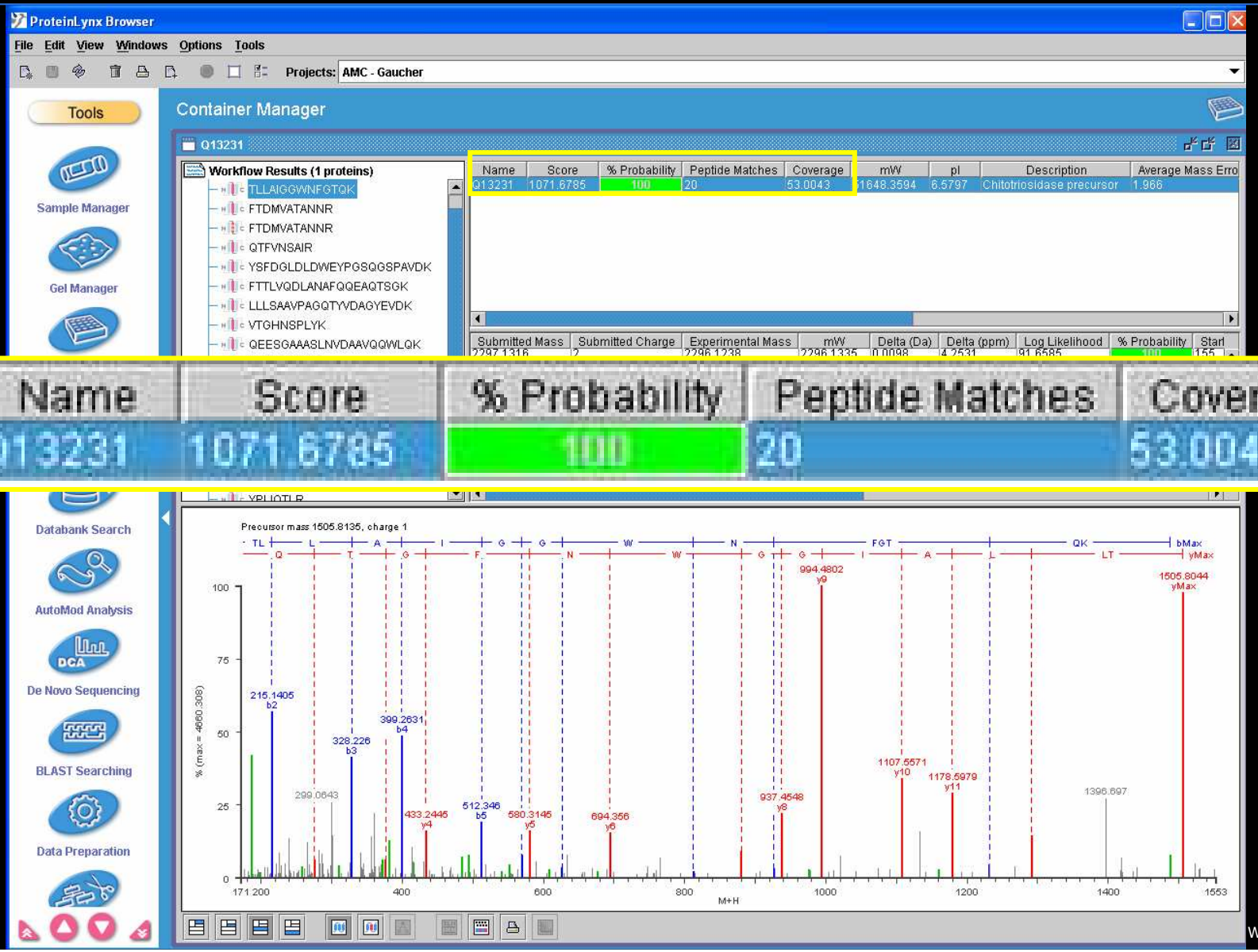
UPLC Retention Time



#8

2,700 Peptides (*E. Coli*)

Multi-Peptide Identifications
...with high sequence coverage



n = 3

Controlling False Discovery

Sample Digest	Database	Proteins identified
E.coli	*E.coli*	411
E.coli	*B.malayi*	2
E.coli	*B.subtilus*	5
E.coli	*M.bovis*	3
E.coli	*P.aeruginosa*	6
E.coli	*S.cerevisiae*	10
E.coli	*Wolbachia*	8

Identity^{E} analysis of a cytosolic fraction of *E. Coli* spiked with 5 standards. 411 proteins were identified with an *E. Coli* specific dB. Matches to proteins in 6 additional dBs are tabulated. Each dB was appended with a random (decoy) dB, to assess the false discovery rate: ≤ 1.5% in each case.

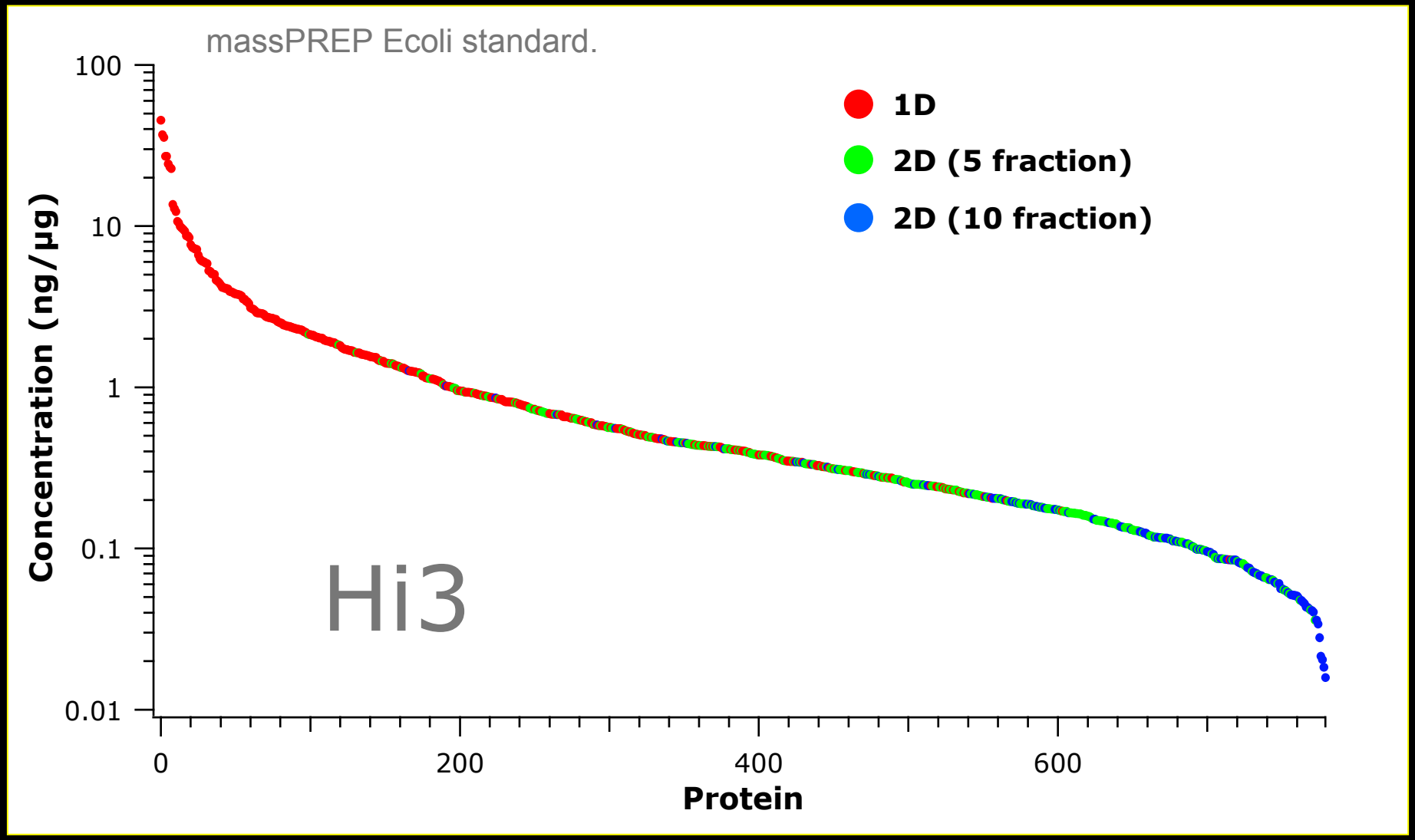
Hi3 Absolute Protein Quantification

- Absolute Protein Quantification conventionally requires co-determination of a unique peptide (for each protein) with its corresponding stable isotope labelled internal standard
  - e.g. "Protein-AQUA" Peptide Standards (SIGMA) [1]

- We have developed a novel Label-Free UPLC/MS^E protocol for **Hi3 Absolute Protein Quantification** [2]
  - Based on an observation that averaged signal response for the most abundant tryptic peptides (n $\geq$3) may be quantitative [3]
  - Under LC-MS^E conditions the average signal response (ion counts per pmole) is constant +/- 10%
  - A known protein standard therefore enables the concentration of all well-characterised proteins in a mixture to be estimated

1 Gygi SP *et al* *Cell* 2001, Dec 14, 107: 715-726
2 Silva JC *et al* *Mol Cell Proteomics* 2006; 5: 144 – 156
3 Mann M *et al* *Nature Reviews Mol Cell Biology* 2004; 5: 699-711

UPLC/MS^E Dynamic Range



Significance of Column Loading

...Number of proteins identified

Hi3

75 µm Column
- 250 to 500 ng of digest

100 µm Column
- 440 to 880 ng of digest

150 µm Column
- 1 to 2 µg of digest

300 µm Column
- 4 to 8 µg of digest

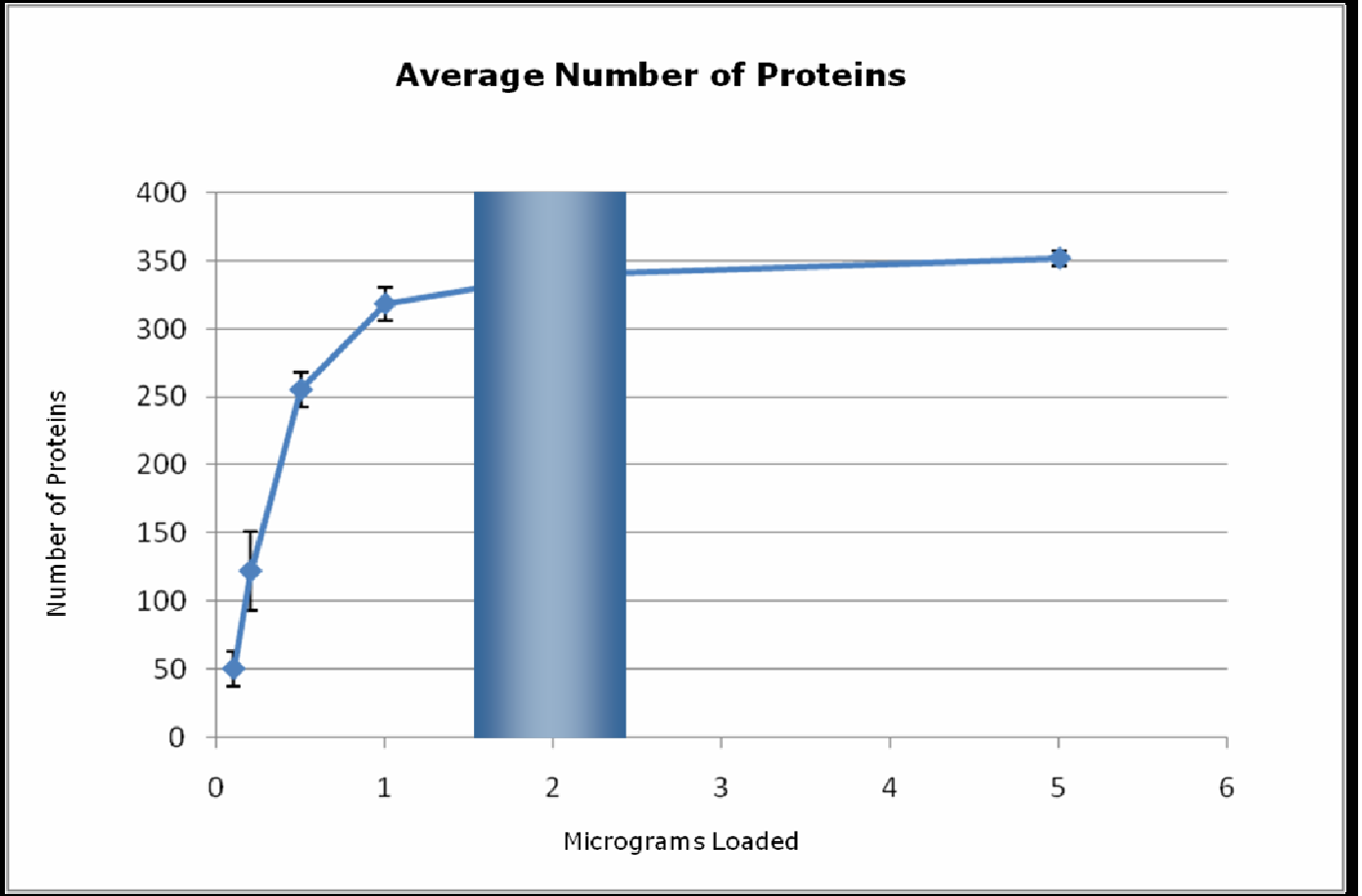
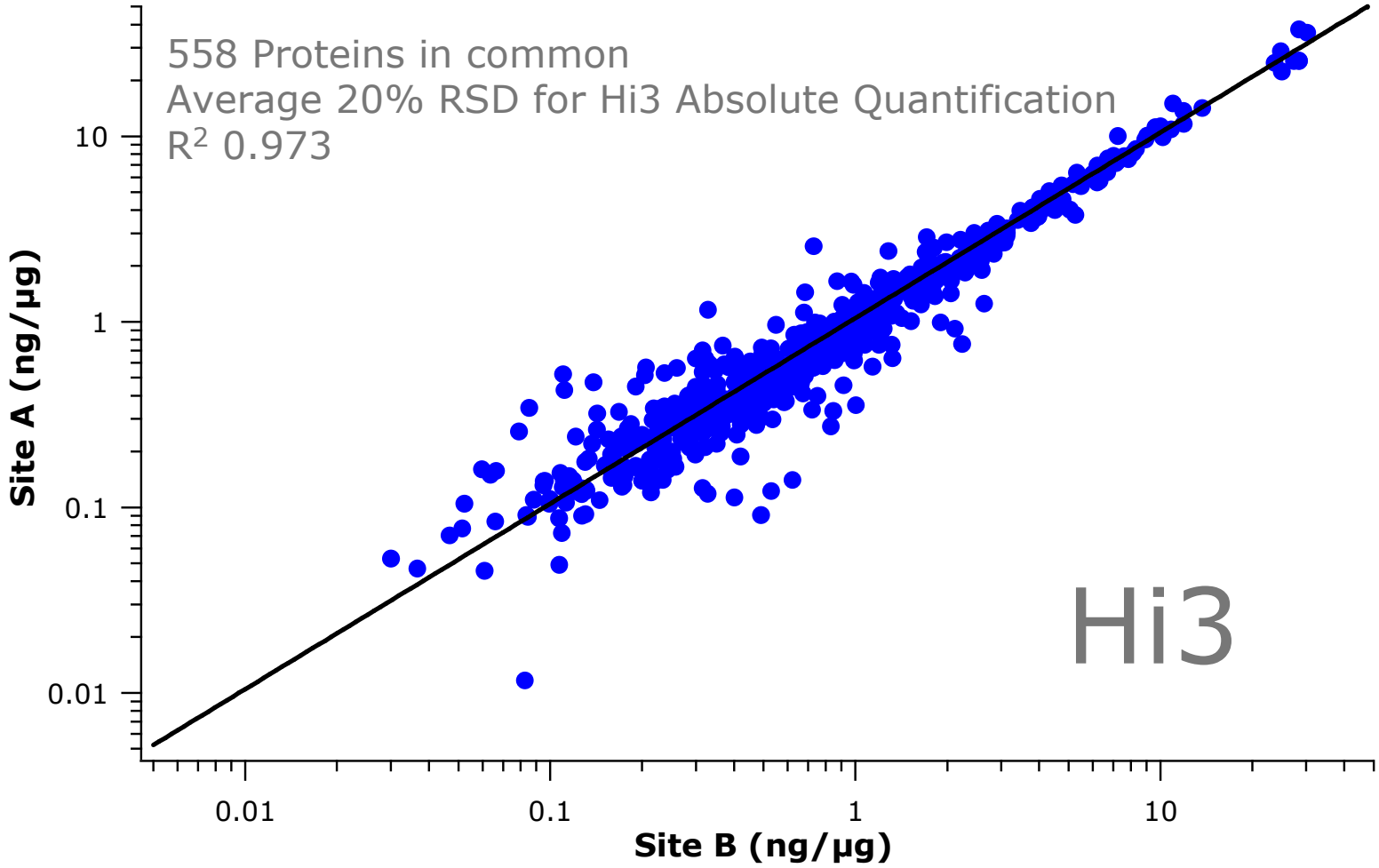


Figure 1. Number of E. coli *proteins (average of 3 injections) identified as a function of quantity of digest on-column.*

2D RP/RP UPLC (High pH/Low pH)

...Hi3 Absolute Quantification Intra-Lab Reproducibility



massPREP Ecoli standard analyzed with identical 2D (5 fraction) protocols at 2 locations.

Relative Protein Quantification

- Established HPLC/MS methods for Relative Protein Quantification include:
  - Stable isotope labelling (*e.g.* ICAT, GIST, SILAC)
  - Isobaric labelling (*e.g.* I-TRAQ)
  - Label-Free (*a.k.a.* ion current intensity)
  - Spectral Counting

- A recent ABRF quantification study involving both Label-Free and isotope/isobaric labelling approaches indicate that the Label-Free methods performed at least as well as the conventional labelled methods [1]

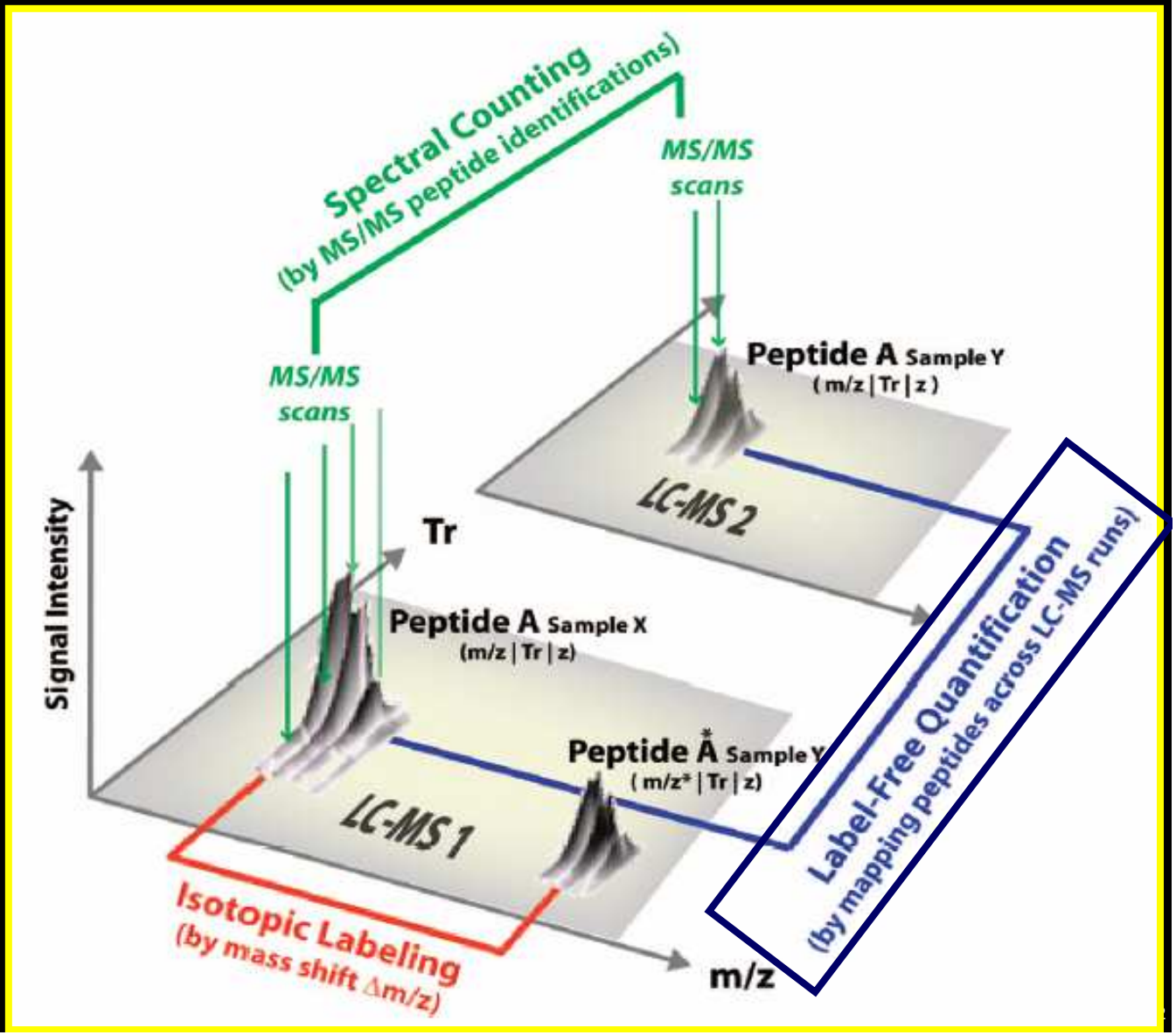
- We have developed a novel, high bandwidth, UPLC/MS^E method for **Label-Free Relative Protein Quantification** [2]
  - Waters Expression^E High Definition Proteomics System

1 ABRF PRG: 2006 Relative Protein Quantification Study. www.abrf.org
2 Quantitative Proteomic Analysis by Accurate Mass Retention Time Pairs. JC Silva *et al*. *Anal.Chem.* 2005, 77, 2187-2200

Relative Quantification Strategies

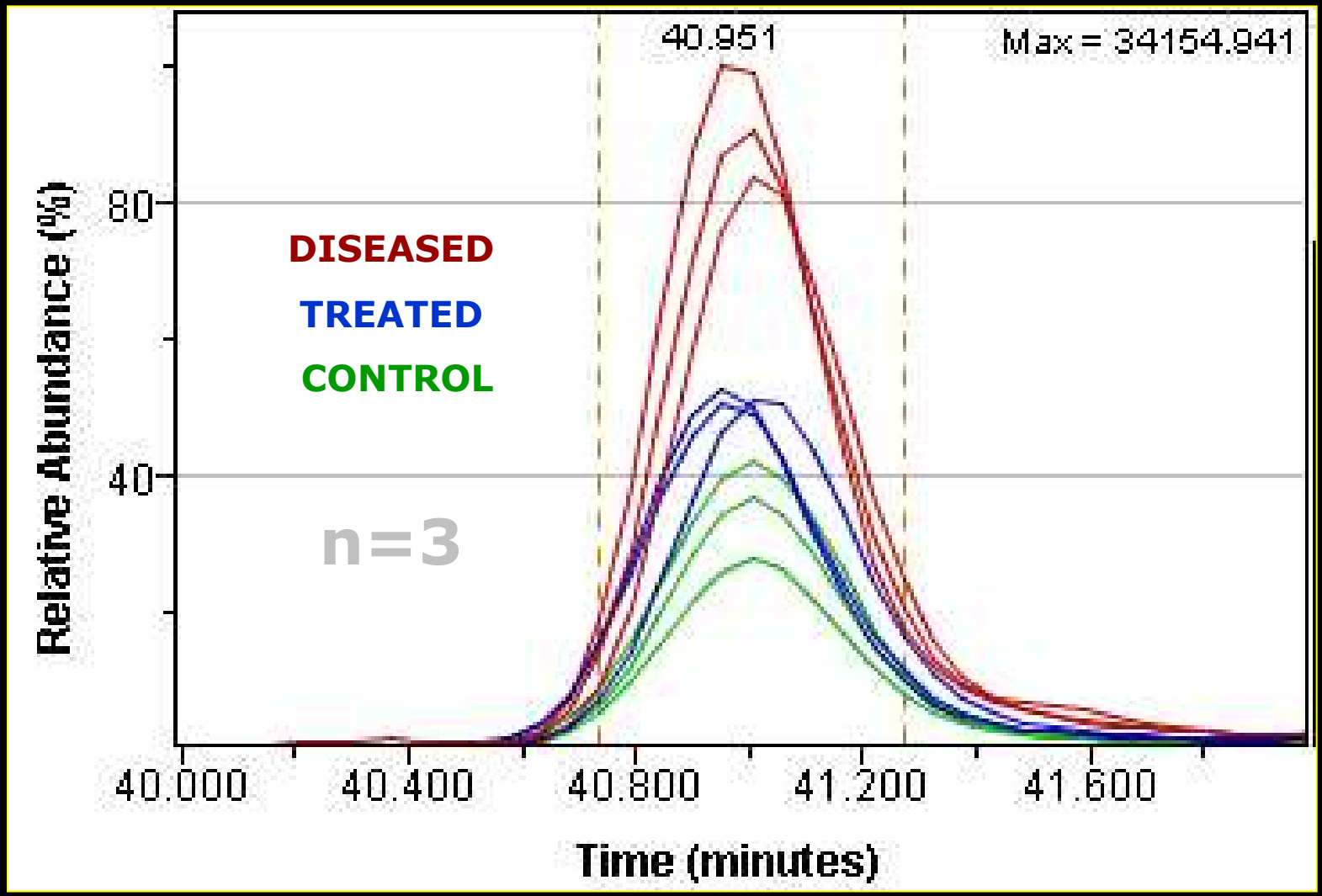
An Assessment of Software Solutions for the Analysis of Mass Spectrometry Based Quantitative Proteomics Data.

Lukas N. Mueller, *et al*.
J. Proteome Res.; *2008*, 7, 51-61



Label-Free Relative Quantification

...tryptic peptide ratios are compared (n=3)



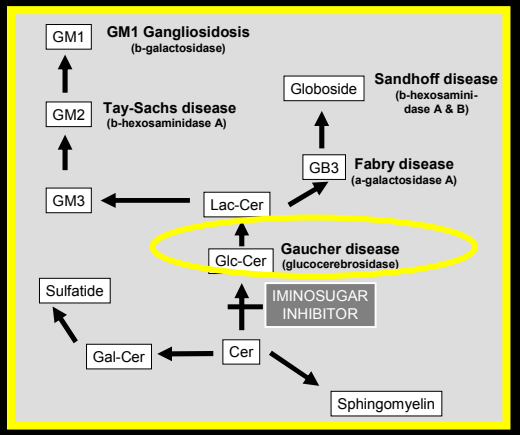
Label-Free Relative Quantification

PROTEIN	TRUE FOLD CHANGE	EXPRESSIONE FOLD CHANGE	EXPRESSIONE FOLD CHANGE % ERROR
Bovine Serum Albumin	0.00	0.0	0.0 %
Lactoperoxidase	0.00	+0.06	6.0 %
Ribonuclease	0.00	-0.06	6.0 %
Peroxidase C1A	0.00	+0.05	5.0 %
Casein	+4.00	+3.89	2.75%
Catalase	+5.00	+4.83	3.4 %
Carbonic Anhydrase	-3.22	-3.22	0.0 %
Glycogen Phosphorylase	-76.9	-66.6	13.4 %
			Average = 4.6 %

The fold change differences across two samples containing eight well characterized proteins in known amounts was determined using Waters ExpressionE Label-Free relative quantification protocol with an average error of 4.6%. These data compare very favorably with a recent 20-laboratory study of identical samples, employing all of the common commercial isotopic/isobaric labelling techniques, which exhibited an average error of 46%.

Gaucher Disease
...Lysosomal Storage Disorder

- **A Proto-Typical Lysosomal Storage Disease**
  - **Inherited deficiency in the activity of Glucocerebrosidase**
    - **Autosomal recessive genetic defect**
  - **Resulting in the accumulation of Glucocerebroside in tissue macrophages (Gaucher cells)**
  - **Prominent clinical symptoms including:**
    - **Hepatosplenomegaly**
    - **Bone deformities**
    - **Pancytopenia**
    - **Neurological abnormalities**
- **Enzyme Replacement Therapy**
  - **Glucocerebrosisdase (e.g. Ceredase)**

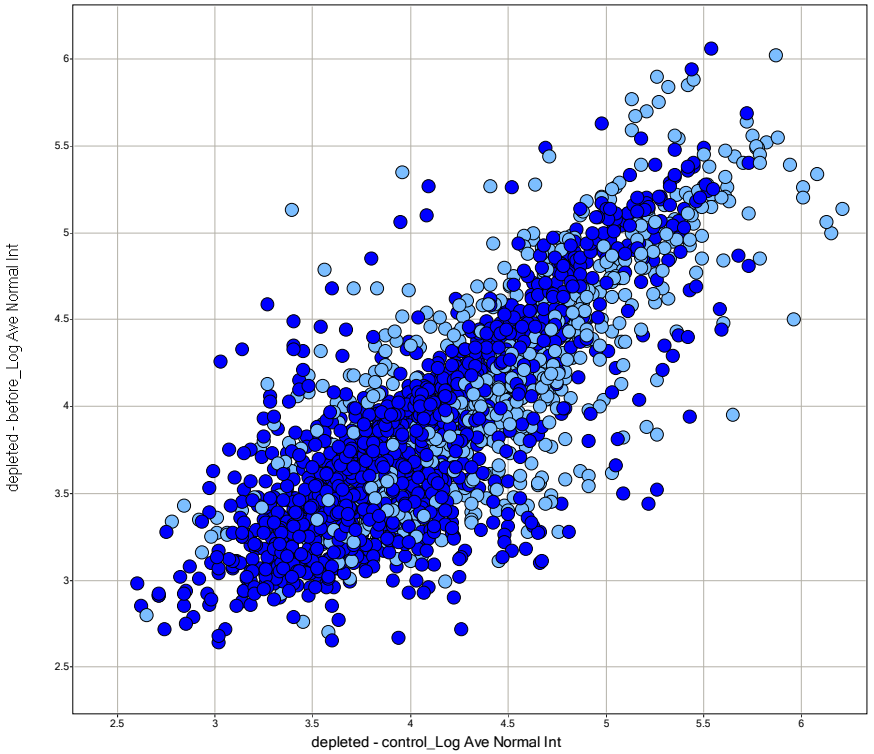


Gaucher Disease Study
...Serum sample preparation

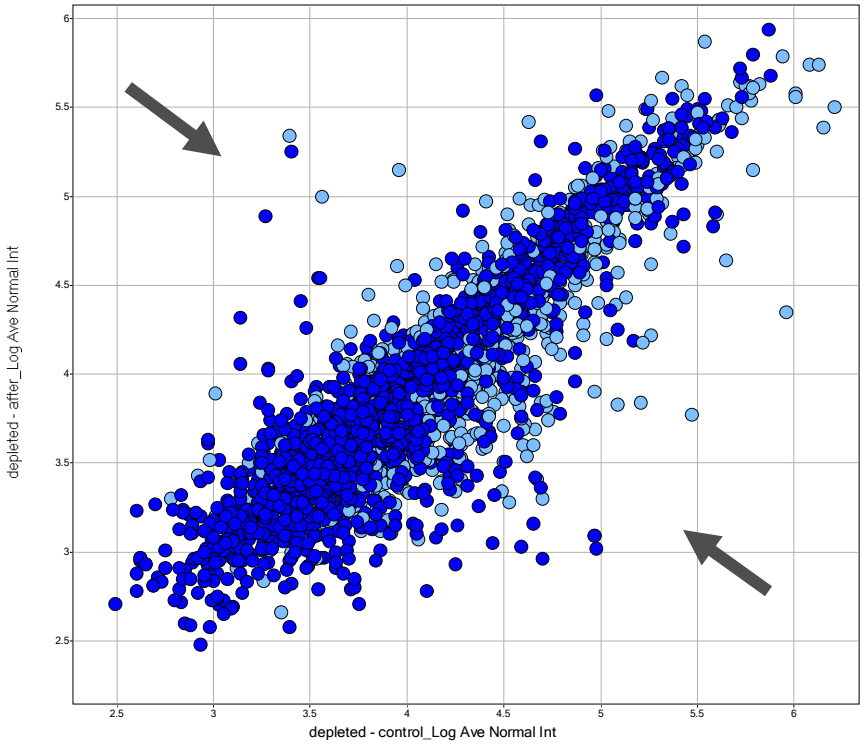
- **Depletion (Agilent MARS)**
  - 20 µL serum + 80 µL buffer A
  - Filtration 0.22 µm filter @ 13,000 rpm/5min
  - Affinity separation followed by buffer exchange with 50 mM NH_4HCO_3
  - Total volume ~ 80 µL
- **Tryptic digestion**
  - 10 µL 0.1% RapiGest (15 min @ 80 ºC)
  - 5 µL 100 mM DTT (30 min @ 60 ºC)
  - 5 µL 200 mM IAA (30 min @ ambient/dark)
  - 20 µL 0.5 µg/ µL trypsin (overnight @ 37 ºC)
  - 4 µL 6M HCl
  - Total volume ~ 124 µL
- **UPLC/MS^E analysis**
  - 1:10 dilution with 0.1% aqueous formic acid solution
  - 1:1 dilution with 100 fmol Enolase/µL Internal Standard

Log/Log Intensity Plots
...treatment effect at the peptide level

3 Conditions / 3 Replicate injections per condition
Control + Patient Before Treatment + Patient After Treatment



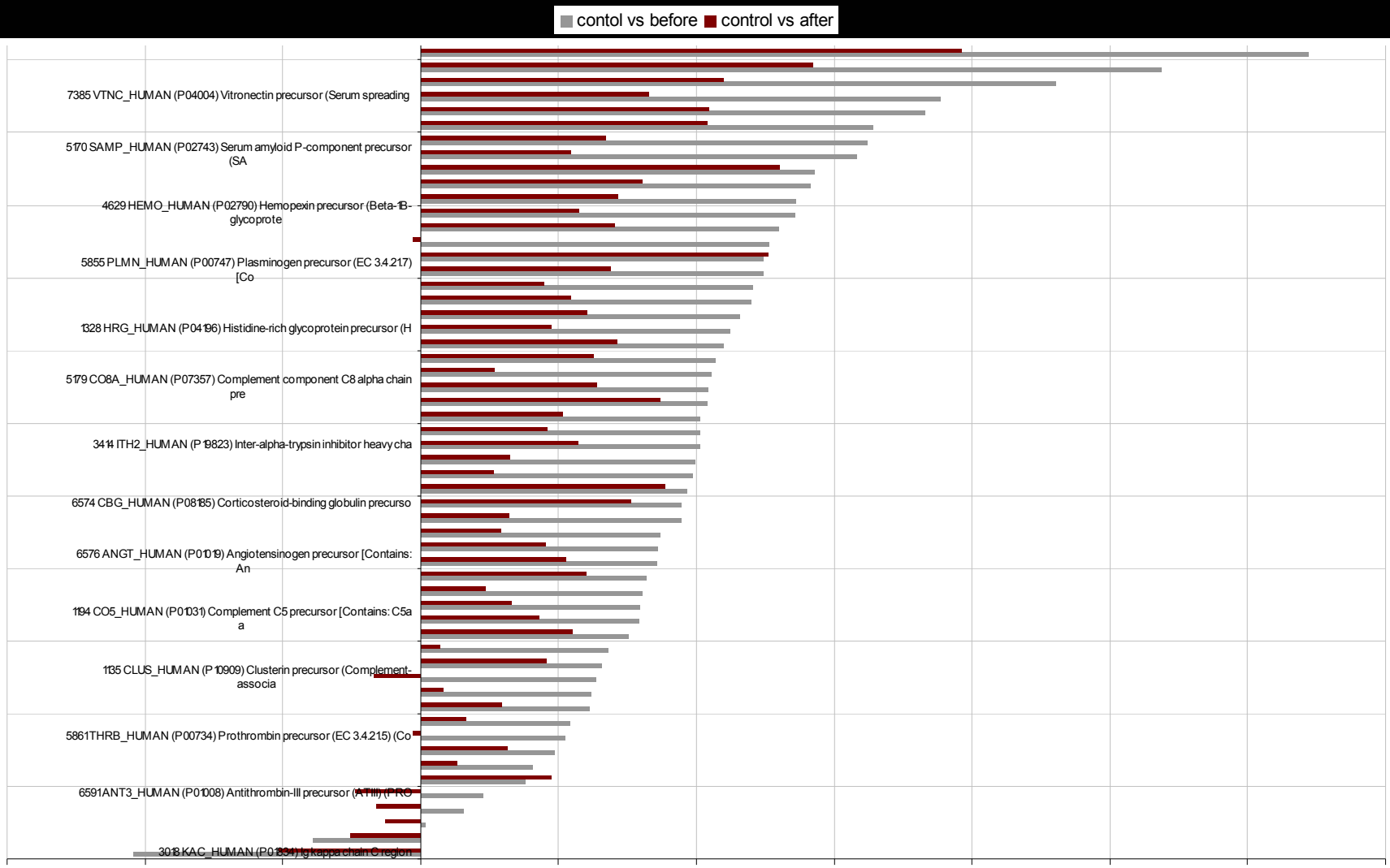
Before Treatment versus Control (depleted)



After Treatment versus Control (depleted)

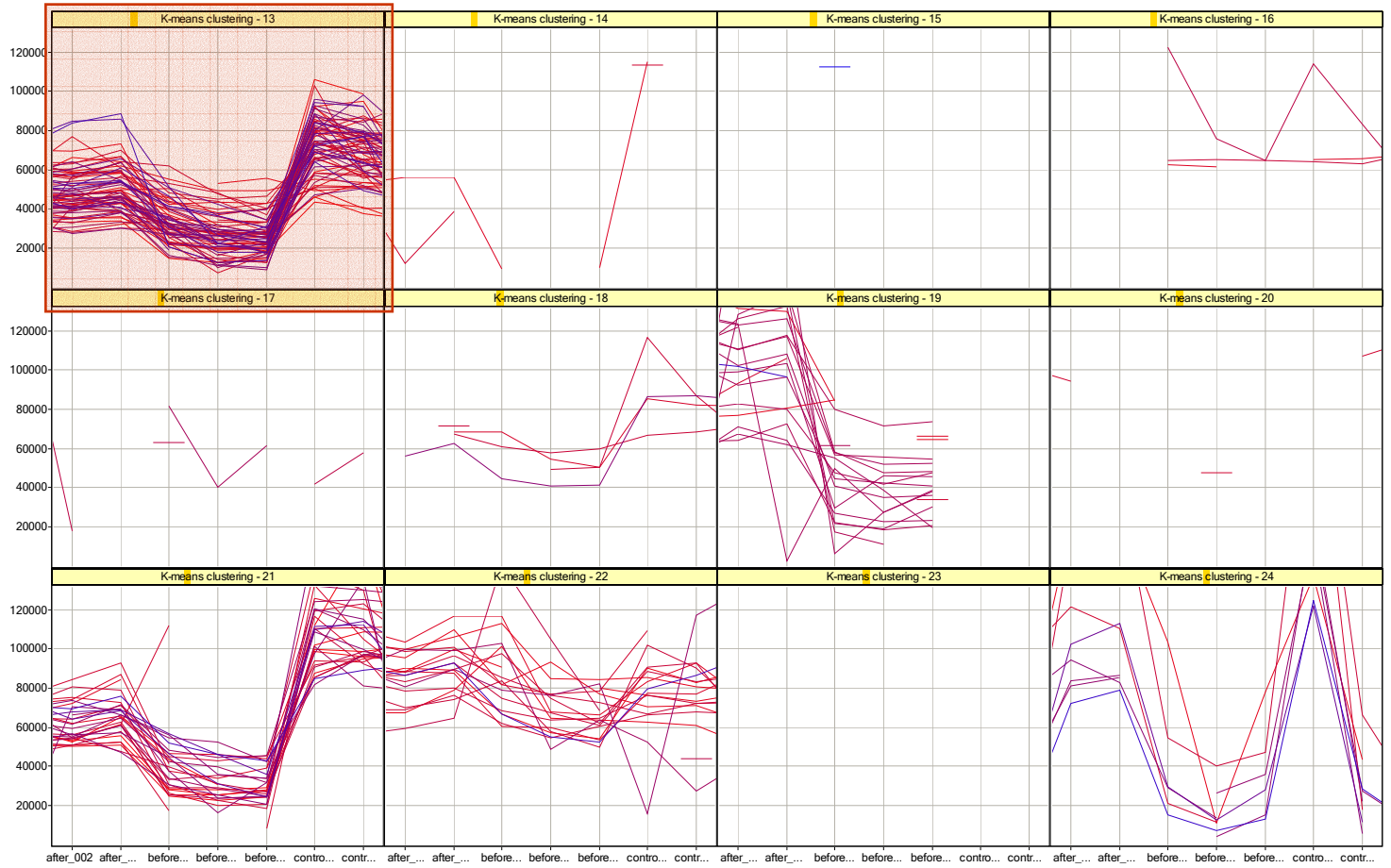
Acknowledgement: Hans Aerts *et al*. Academic Medical Center, University of Amsterdam

Protein ID Regulation Results
...treatment effect at the protein level



Acknowledgement: Hans Aerts *et al*. Academic Medical Center, University of Amsterdam

Clustering Analysis
...peptide intensity profiling by K-means clustering

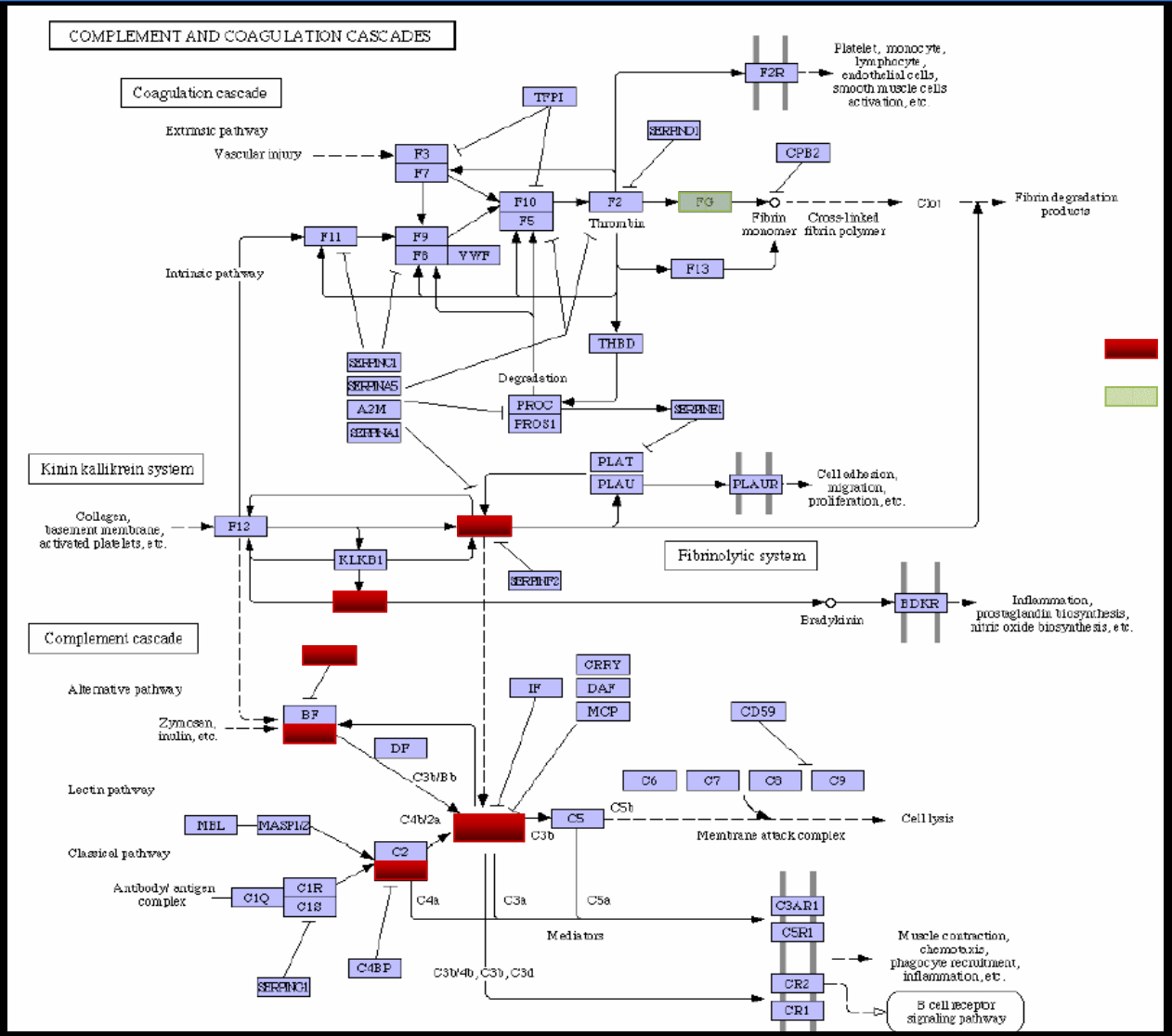


"grouping" of peptides/proteins with similar regulation profiles

Acknowledgement: Hans Aerts *et al*. Academic Medical Center, University of Amsterdam

Complement & Coagulation Cascades

...clustering reveals plausible protein relationships



K-means clustering

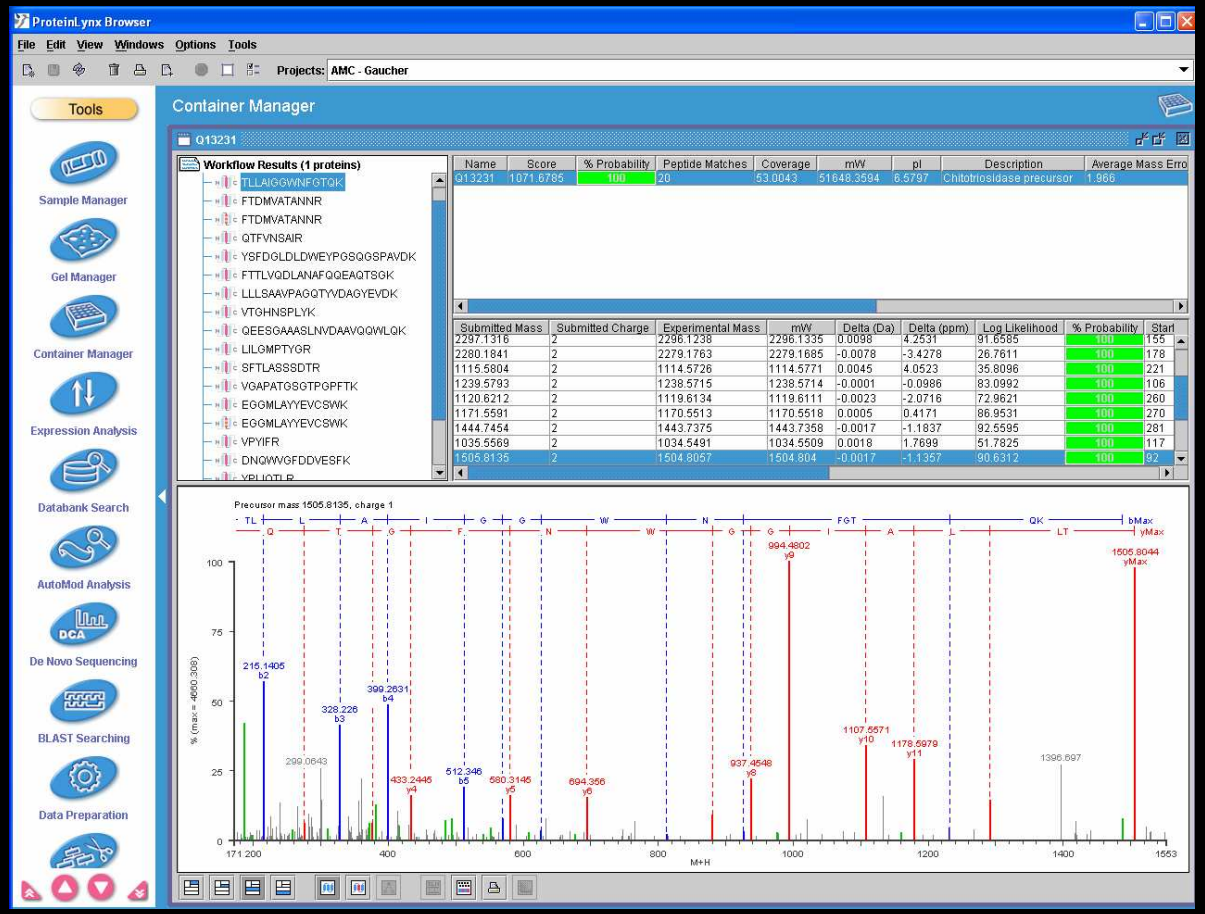
hierarchical clustering

Acknowledgement: Hans Aerts *et al*. Academic Medical Center, University of Amsterdam

Chitotriosidase
...quantification of a known Gaucher Biomarker

Chitotriosidase is a known indicator of Gaucher cells

Chitotriosidase shows a large activity increase **(~100x)** in the serum of symptomatic patients



Acknowledgement: Hans Aerts *et al*. Academic Medical Center, University of Amsterdam

Chitotriosidase

...Ion tracking: 3 replicates for each of 3 conditions

				Replication Rate		
$[M+H]^+$	St. dev. (mDa)	ppm	Average Rt	Before Treatment	After Treatment	Control
1003.5934	0.0016	1.6	42.35	3	0	0
1120.6200	0.0020	1.8	47.07	3	0	0
1171.5692	0.0037	3.2	28.57	3	0	0
1444.7467	0.0024	1.7	30.22	3	0	0
2297.1311	0.0042	1.8	73.76	3	0	0

Chitotriosidase is known to show a large activity increase **(~100x)** in the serum of symptomatic patients

Acknowledgement: Hans Aerts *et al*. Academic Medical Center, University of Amsterdam

Absolute Quantification: UPLC/MS^E
...Gaucher disease before treatment (depleted serum)

- **[Chitotriosidase]$_{MEASURED}$ = 1.60 fmol/µL in 0.5 µg**
  - **Serum activity (UPLC/MS^E) = 39,500 nmol/ml/h**
  - **Serum activity (biomolecular assay)[1] = 31,800 nmol/ml/h**

$$\frac{\sum_{i=1}^{3} \text{peptide intensity Chitotriosidase}}{\sum_{i=1}^{3} \text{peptide intensity Enolase}} \cdot [50\ \text{fmol}/\mu\text{L}]$$

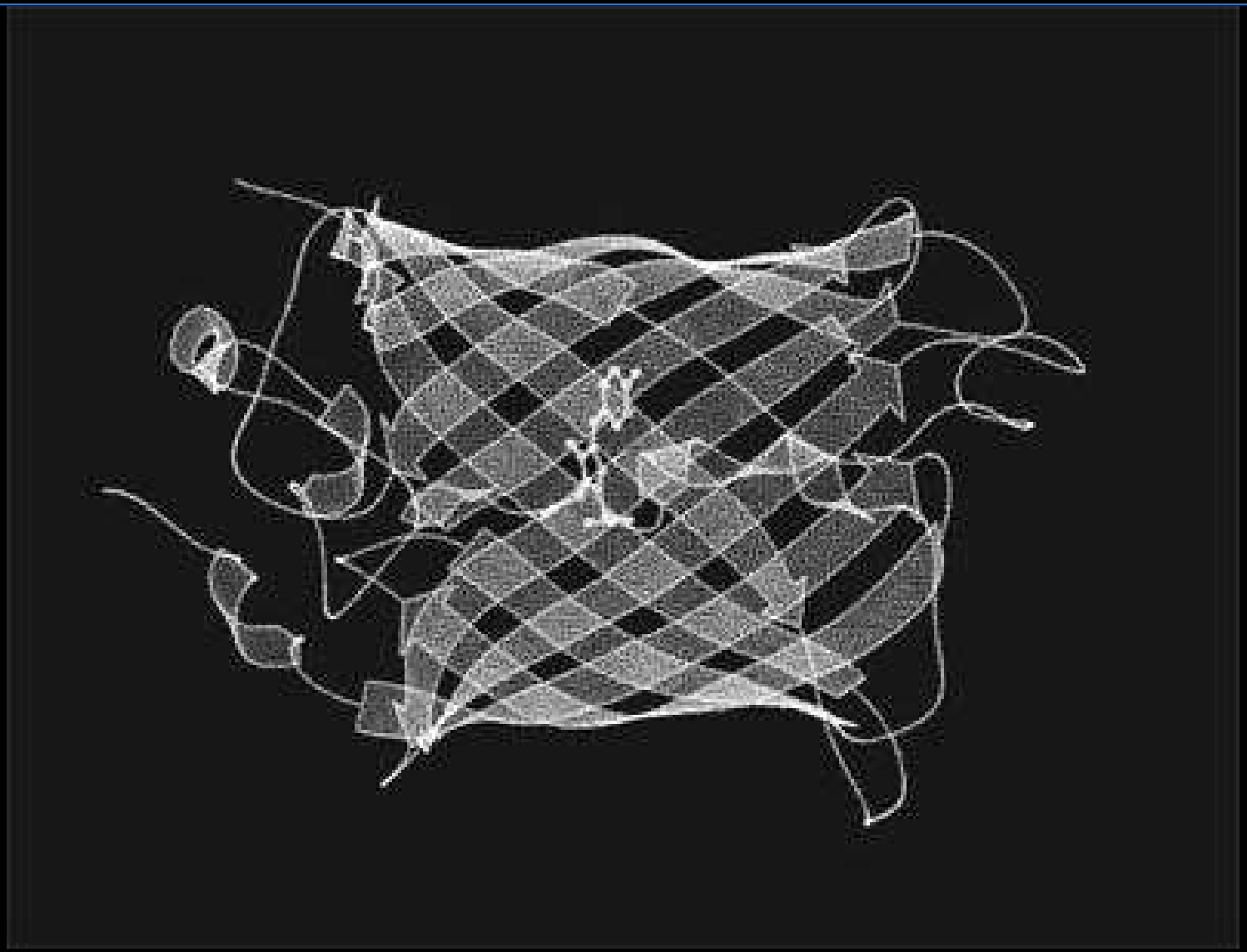
Enzyme Substrate Assay:

1. **Aerts, J.M.F.G., *et al*. Deficiecient Activity of Glycocerebrosidase in Urine from Patients with Type 1 Gaucher Disease. *Clin. Chem. Acta* 158, 155-164. (1986)**

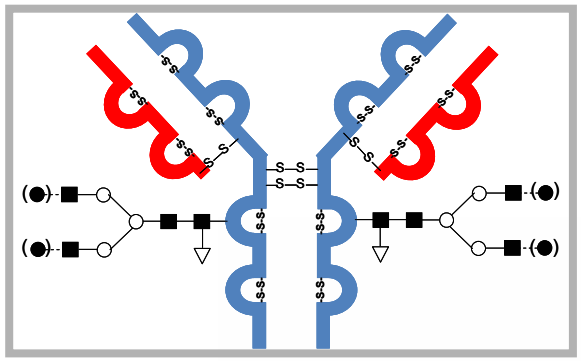
Summary: Gaucher Study

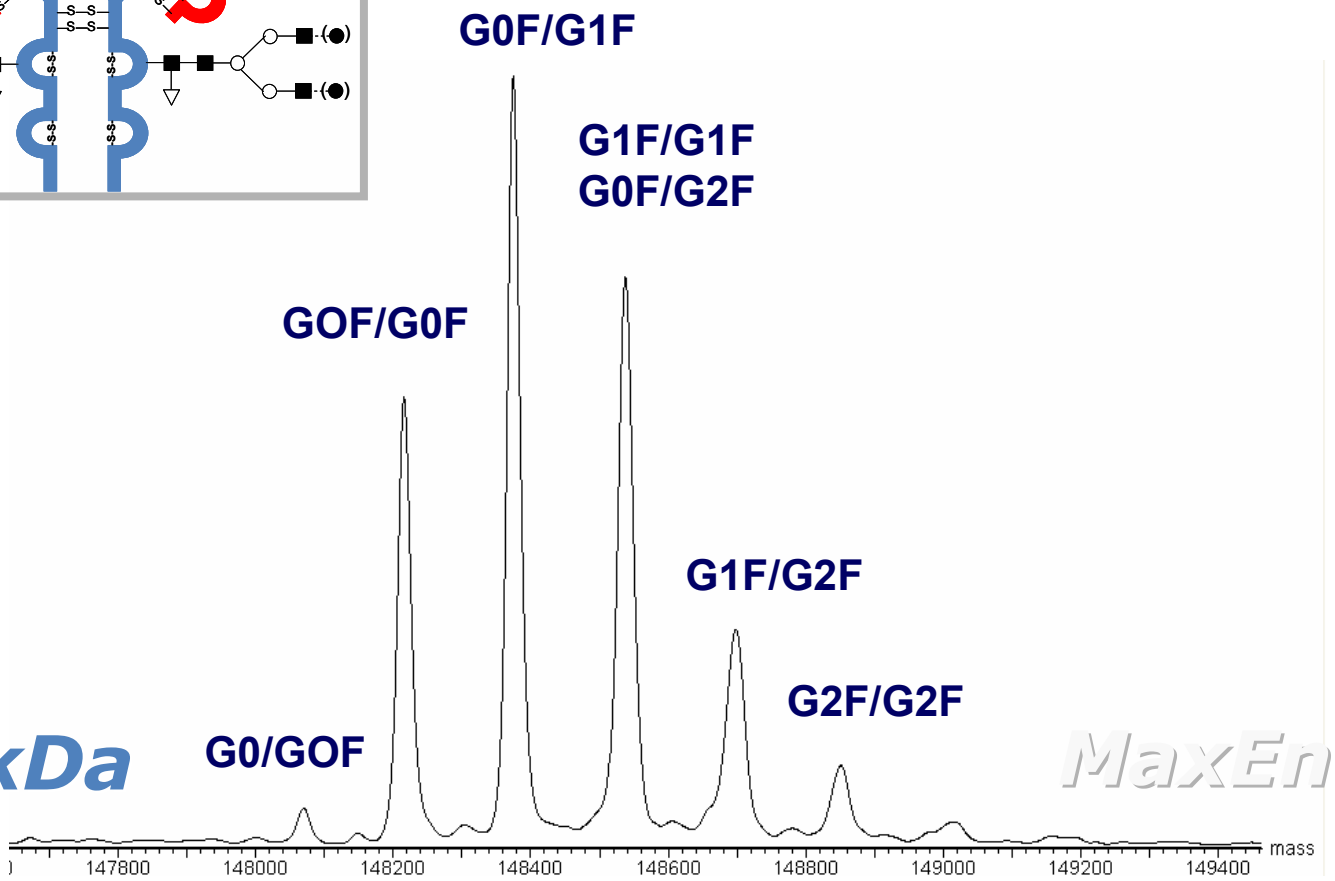
- UPLC/MS^E provides a high bandwidth alternative to conventional LC/MS/MS for quantitative and qualitative profiling of complex digest mixtures
- EXPRESSION^E (Label-Free) relative quantification results demonstrated that post-treatment Gaucher patient protein profiles show conformity to the control (healthy) profile
- Clustering techniques applied to EXPRESSION^E Proteomics data reveal biologically plausible protein relationships
- Hi3 absolute quantification of Chitotriosidase (pre-treatment) by UPLC/MS^E is consistent with published enzyme substrate assay data

Intact Protein Analysis *...Interactomics ?*



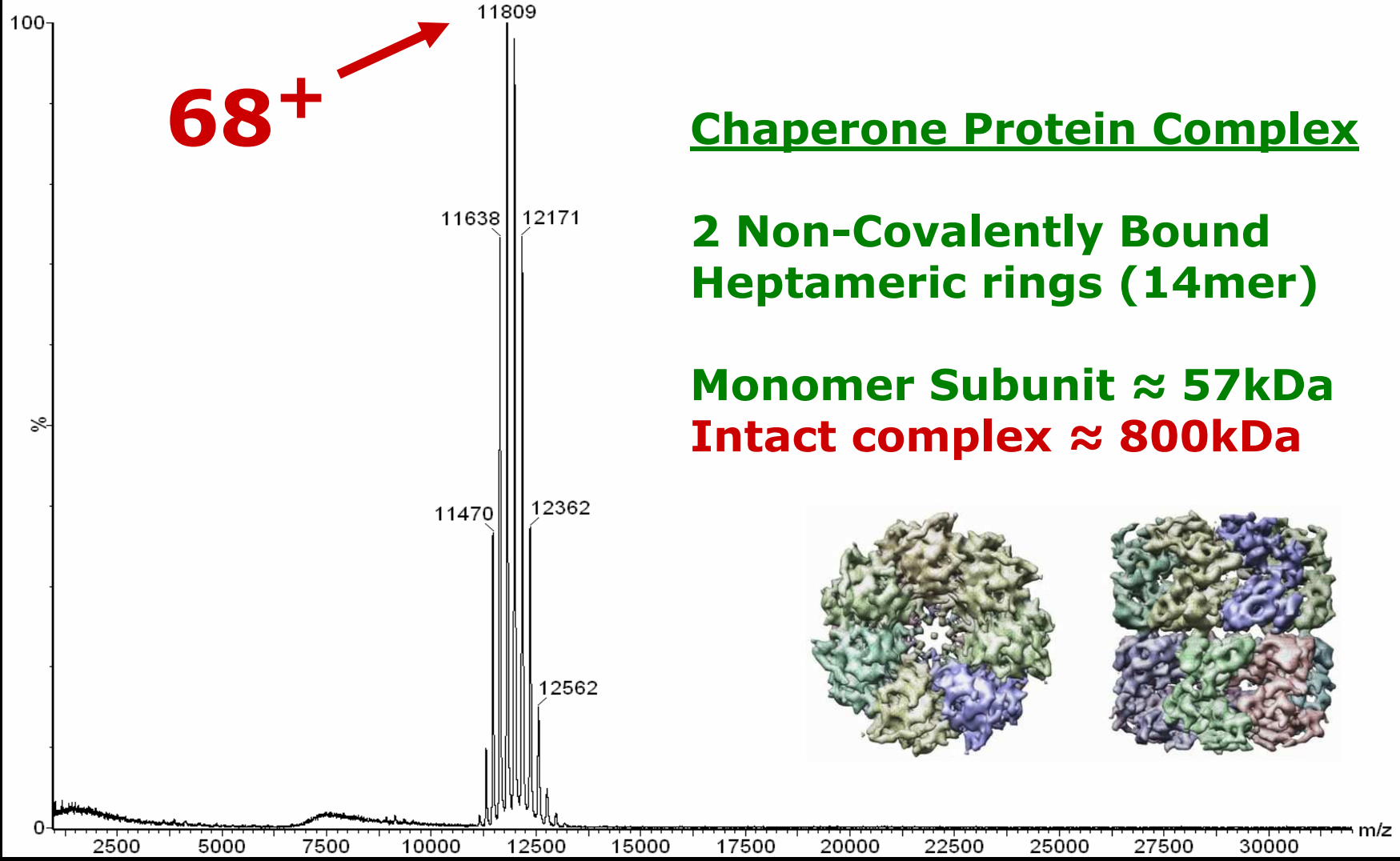
Intact Protein Analysis: IgG
...Deconvoluted Spectrum (Z=0)





148 kDa

Non-Covalent Protein Complex Analysis
...Electrospray MS of GroEL

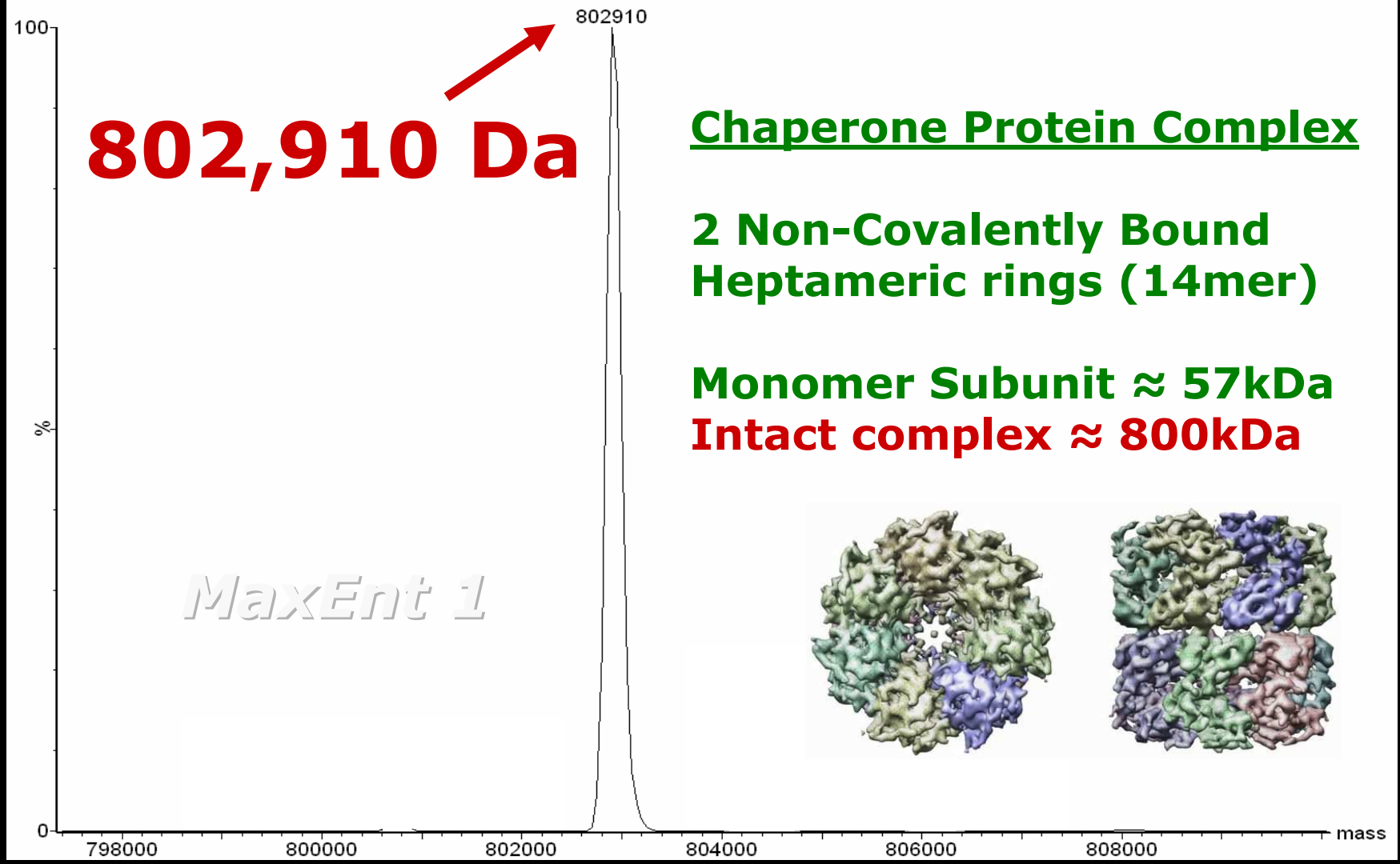


Chaperone Protein Complex

2 Non-Covalently Bound Heptameric rings (14mer)

Monomer Subunit ≈ 57kDa
Intact complex ≈ 800kDa

Non-Covalent Protein Complex Analysis
...Electrospray MS of GroEL



802,910 Da

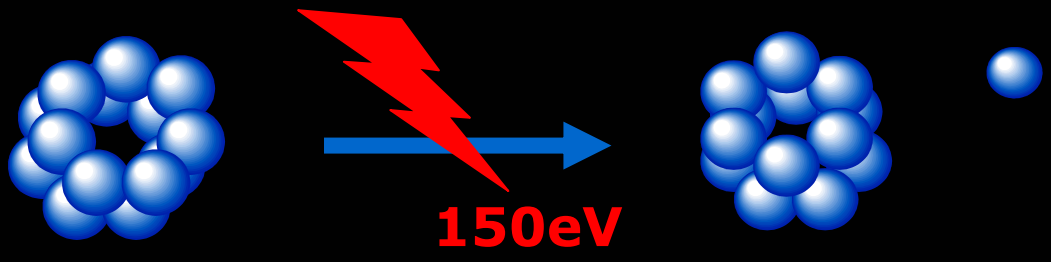
Chaperone Protein Complex

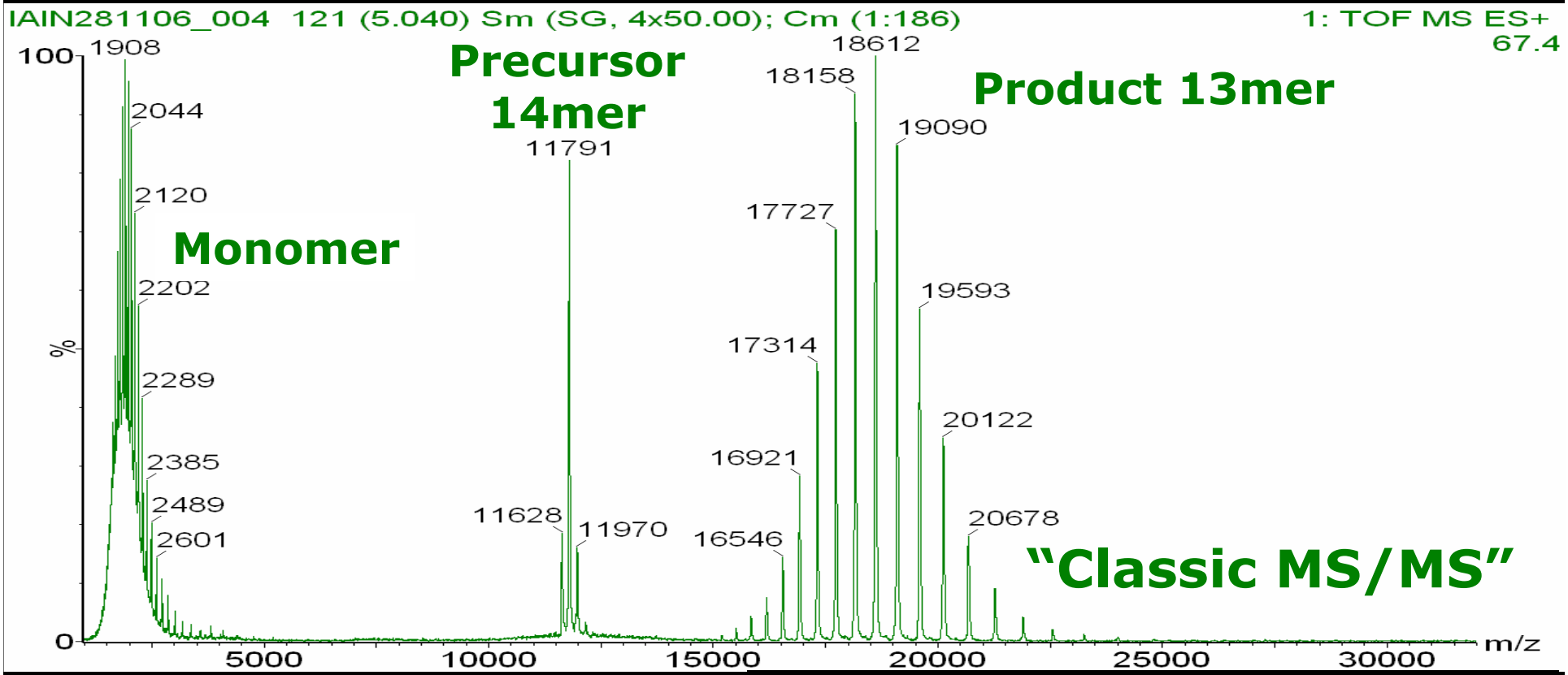
2 Non-Covalently Bound Heptameric rings (14mer)

Monomer Subunit ≈ 57kDa
Intact complex ≈ 800kDa

Non-Covalent Protein Analysis

...MS/MS: CID Fragmentation of GroEL $[M+68H]^{68+}$

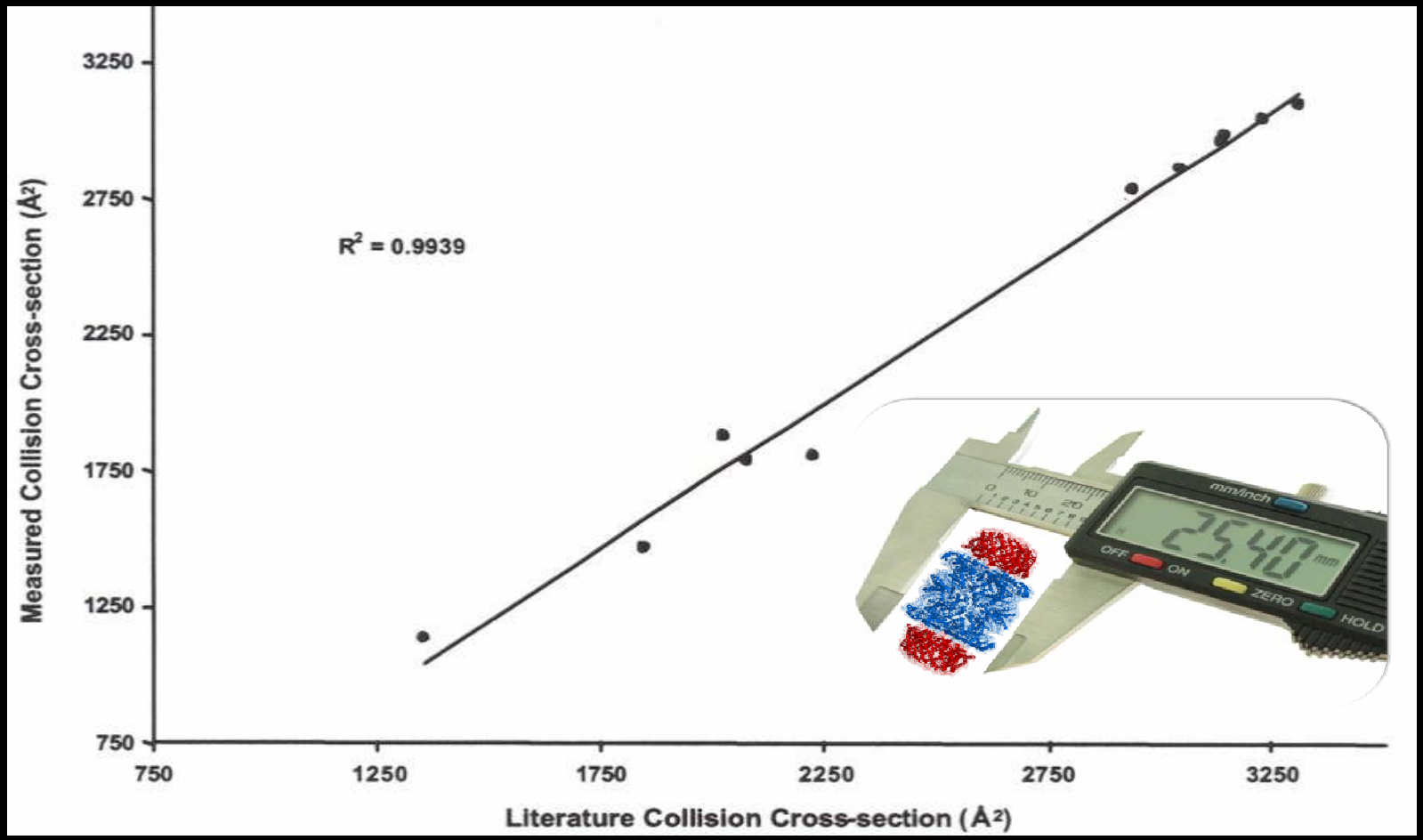




Acknowledgement: Albert Heck, Utrecht University

Measuring Protein Collision Cross Section

…IMS correlates with literature measurements



Brandon T. Ruotolo, *et al.*, Science, vol 310, 9th December 2005, 1658-1660. [Supporting On-Line Material @ *ScienceExpress*]

Acknowledgements

Scott Berger
Catalin Doneanu
Craig Dorschel
Barry Dyson
Scott Geromanos
Marc Gorenstein
Dan Golick
Jim Langridge
Guo-Zhong Li
Therese McKenna
Jeff Silva
Hans Vissers

www.waters.com