FT-ICR for Proteomics

Ljiljana Paša-Tolić

Environmental Molecular Sciences Laboratory
Pacific Northwest National Laboratory

Environmental Molecular Sciences Laboratory: A National Scientific User Facility

The William R. Wiley Environmental Molecular Sciences Laboratory, a U.S. Department of Energy national scientific user facility located at Pacific Northwest National Laboratory in Richland, Washington provides integrated experimental and computational resources for discovery and technological innovation in the environmental molecular sciences to support the needs of DOE and the nation.

Science Themes

Instruments are available to users, generally at **no cost** to the user, for research in four science theme areas:

- Atmospheric aerosol chemistry
- Biological interactions and dynamics
- Geochemistry/biogeochemistry and subsurface science
- Science of interfacial phenomena.

Capabilities

EMSL houses several state-of-the art mass spectrometry, nuclear magnetic resonance, computational, and laser, optical and imaging capabilities under one roof. Instruments include:

- A 12-tesla mass spectrometer, one of a handful in existence.
- One of the world's largest (900 MHz) wide bore NMR spectrometers.
- Supercomputing resources and associated premier computational chemistry software.

For more information, see:

WWW.EMSL.PNL.GOV

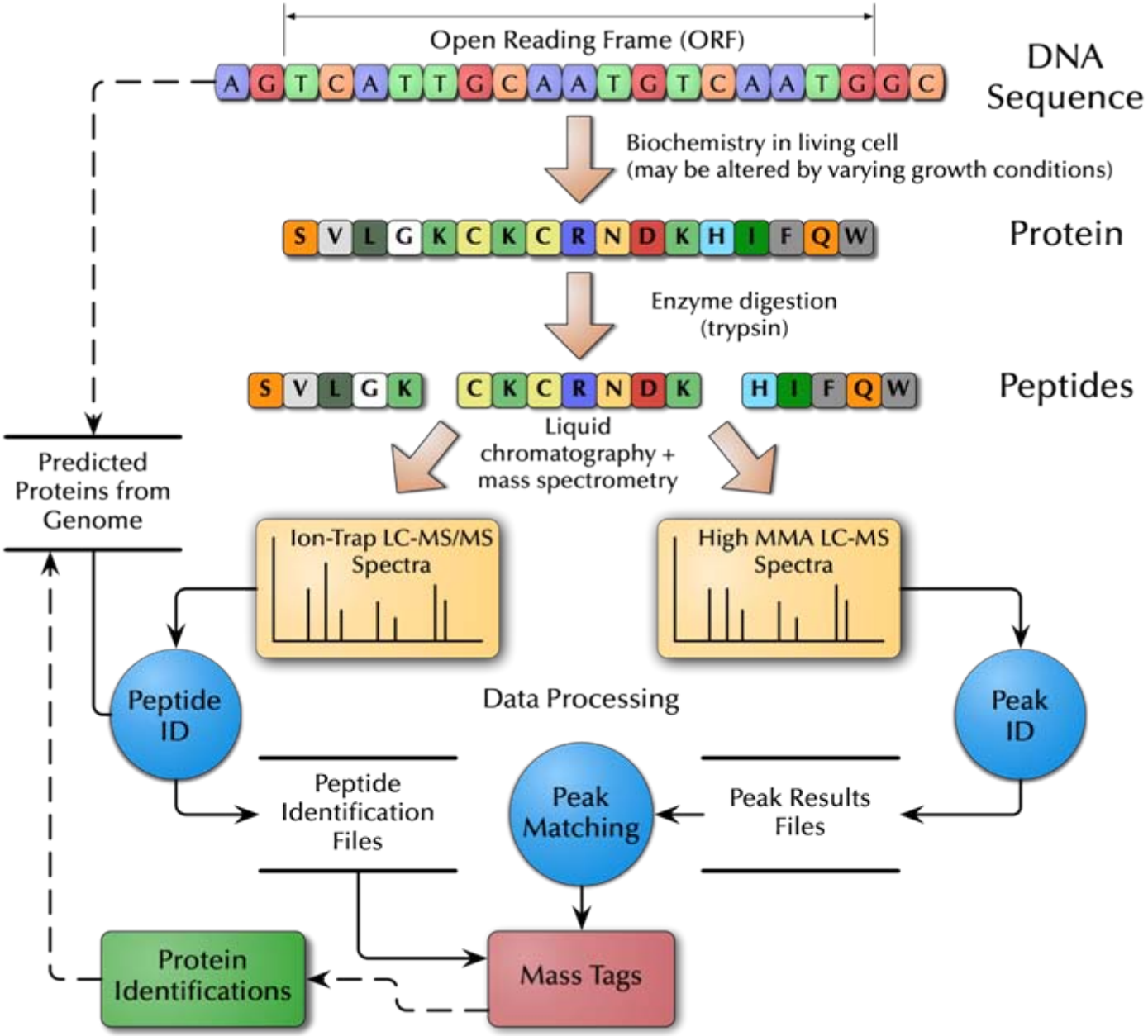
Proteomics:
The Good, the Bad, and the Ugly

- Studies the real functional molecules of the cell & enables important advances in our understanding of biological systems

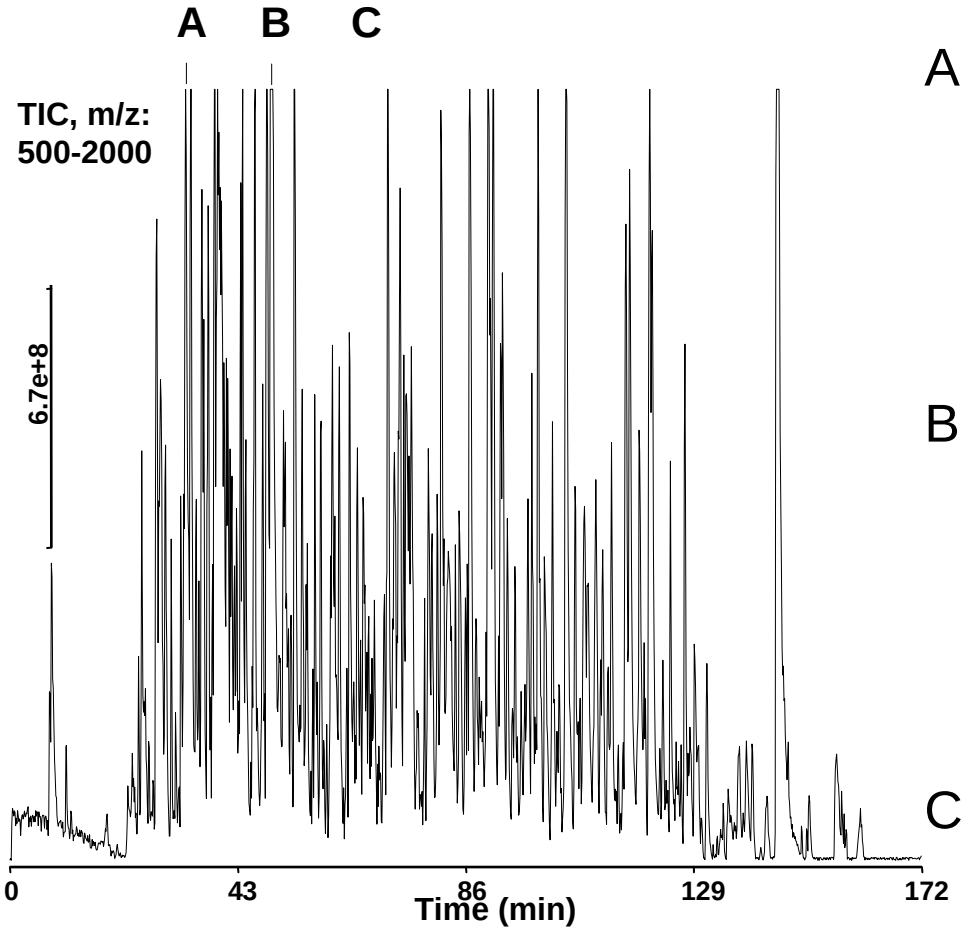
- Limited depth
  - sample complexity and/or heterogeneity
  - limit of detection (the lowest number of molecules that can be detected)
  - dynamic range (ability to detect a low abundant species in the presence of other more-abundant species)
  - analysis throughput

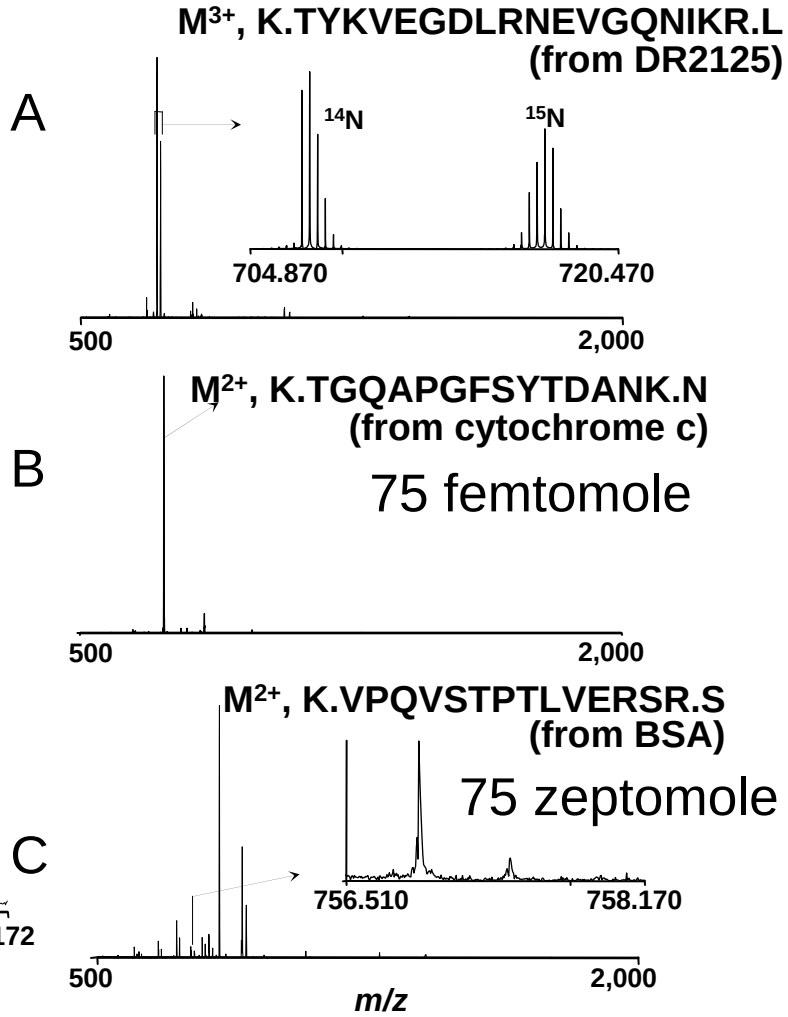
- Misidentifications
- Protein isoforms, post-translational modifications (PTMs) and covariation between PTMs on single protein

Accurate Mass and Time Tag Approach

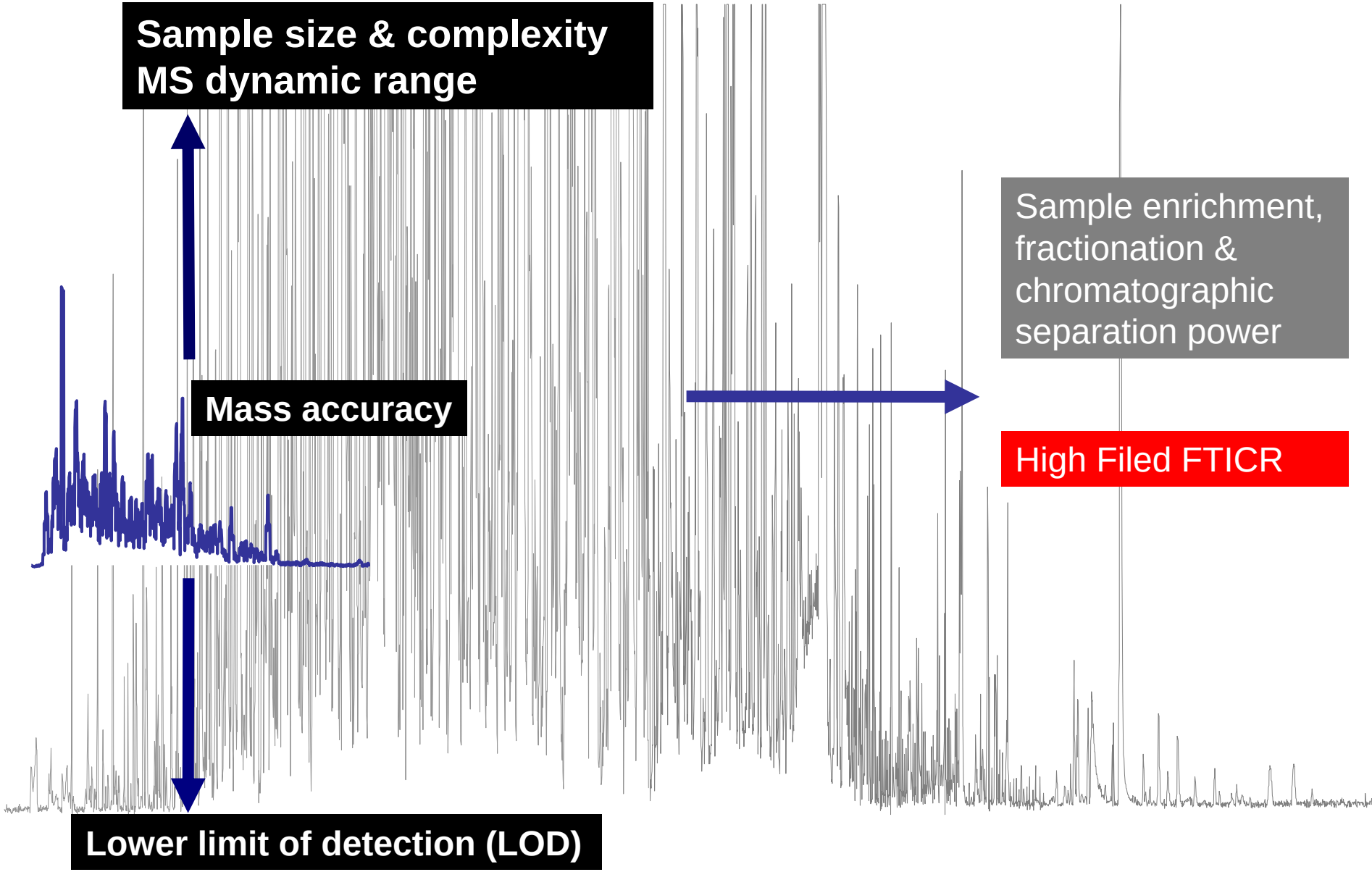


Dynamic range: 10^6
Detection limit: 100 zeptomoles
MMA 1-5 ppm

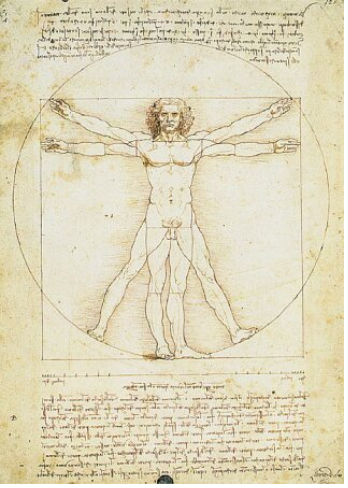




What Limits Proteome Measurement Coverage?



Proteome Complexity & Mass Measurement Accuracy

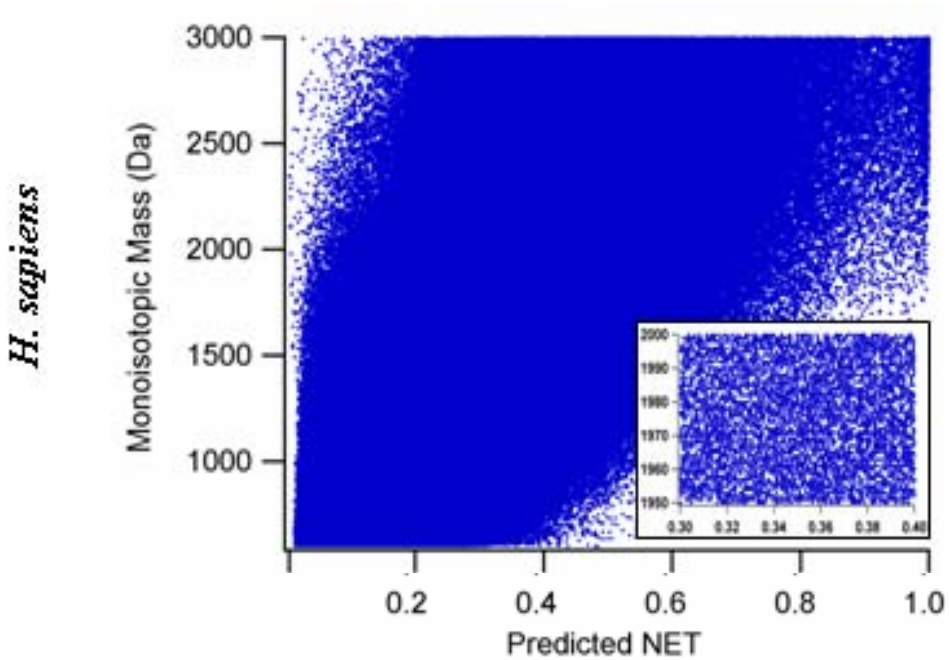


The human organism:

- ~ 3 billion nucleotides
- ~ 30,000 genes coding for
- ~ 100,000-300,000 transcripts
- ~ 1-2 million proteins
- ~ 60 trillion cells of
- ~ 300 cell types in
- ~14,000 distinguishable

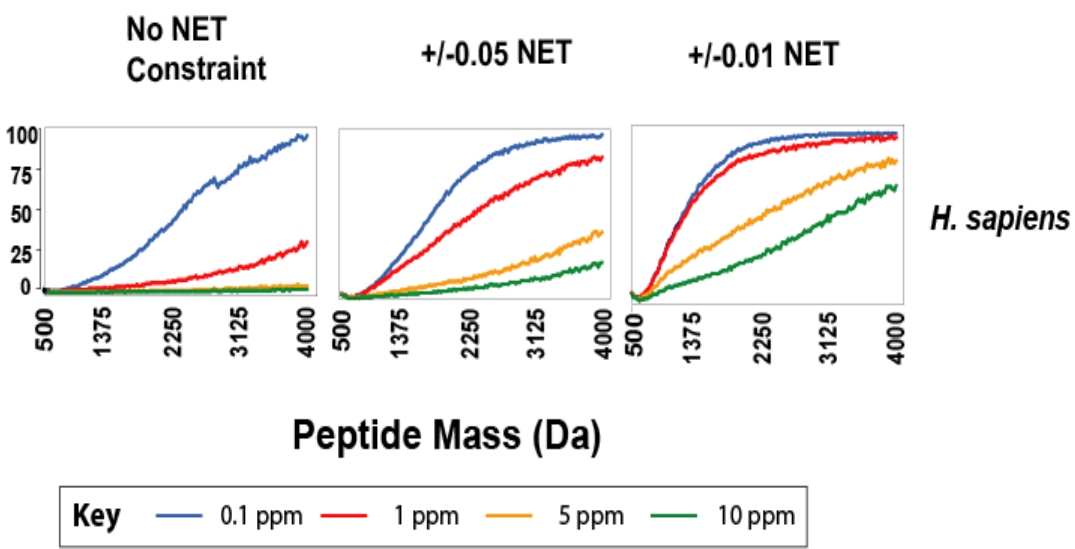
morphological structures

~1.7 million tryptic peptides (allowing 1 missed cleavage) from human



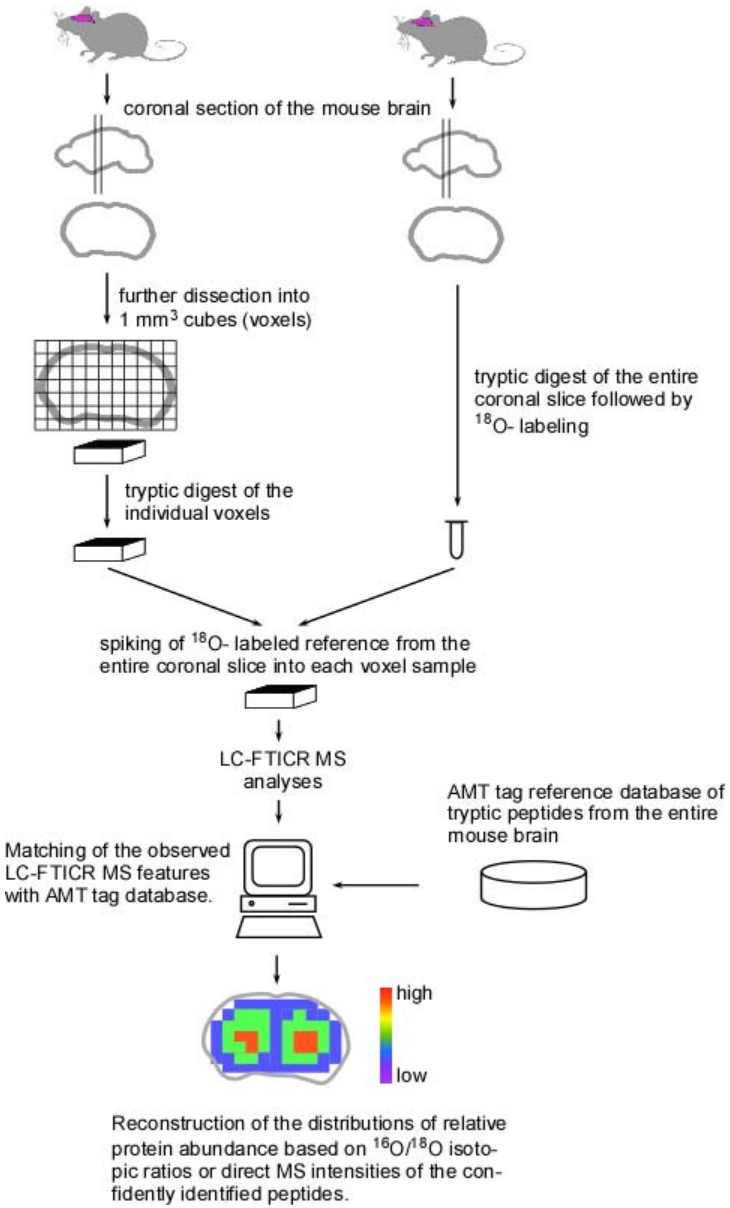
MMA <1 ppm

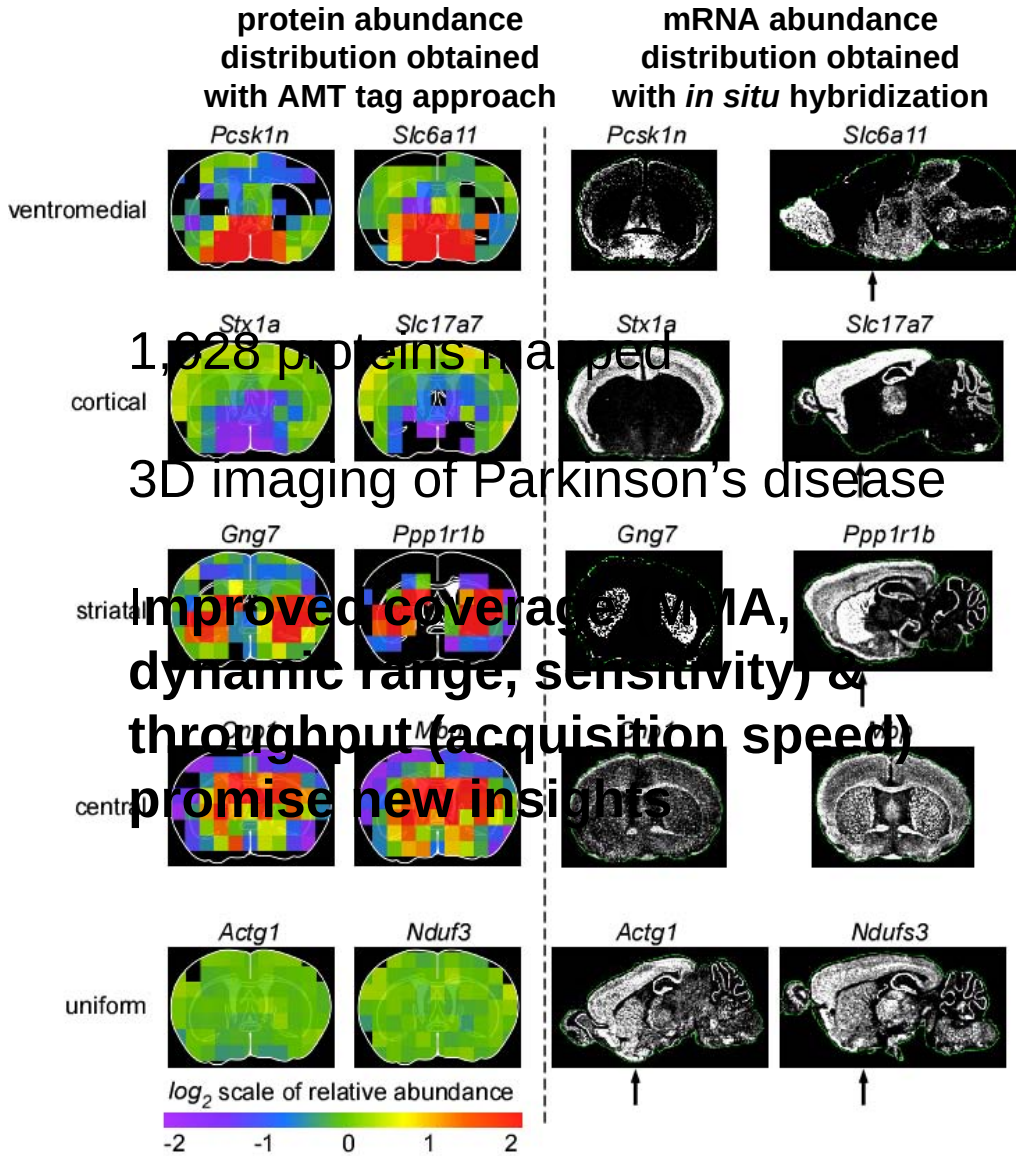
NET <1%



Norbeck at al., JASMS, 2005, 16, 1239–1249

AMT tags for spatial mapping of protein abundances in the mouse brain





1,928 proteins mapped

3D imaging of Parkinson's disease

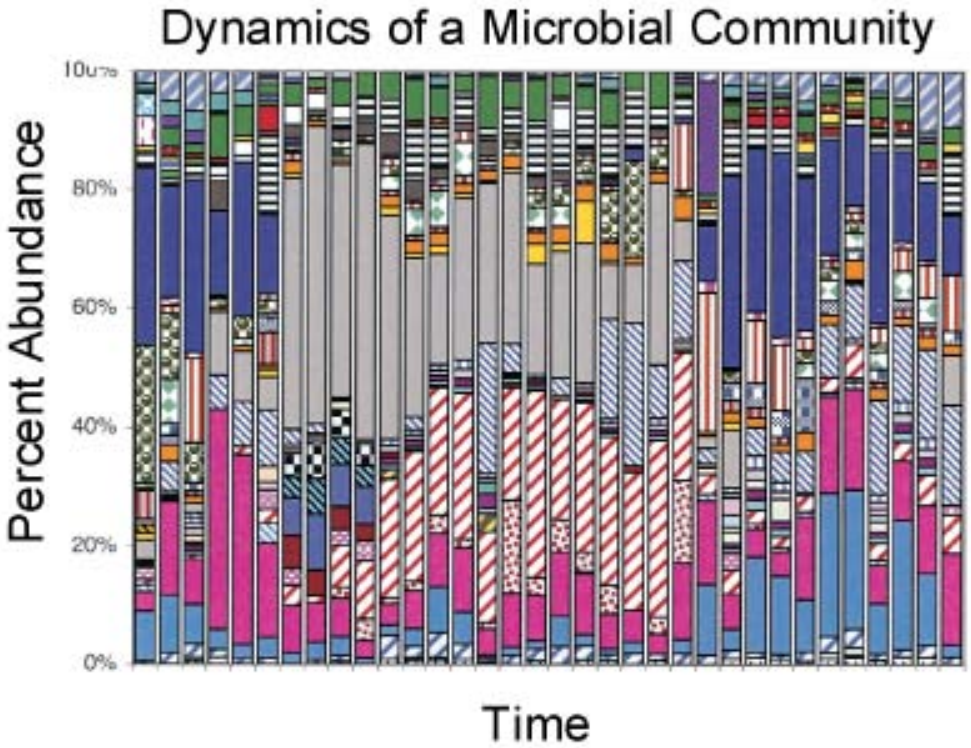
Improved coverage (MA, dynamic range, sensitivity) & throughput (acquisition speed) promise new insights

VA Petyuk et al., Genome Res. 2007 17(3):328-36.

DJ Smith, Molecular & Medical Pharmacology
UCLA School of Medicine

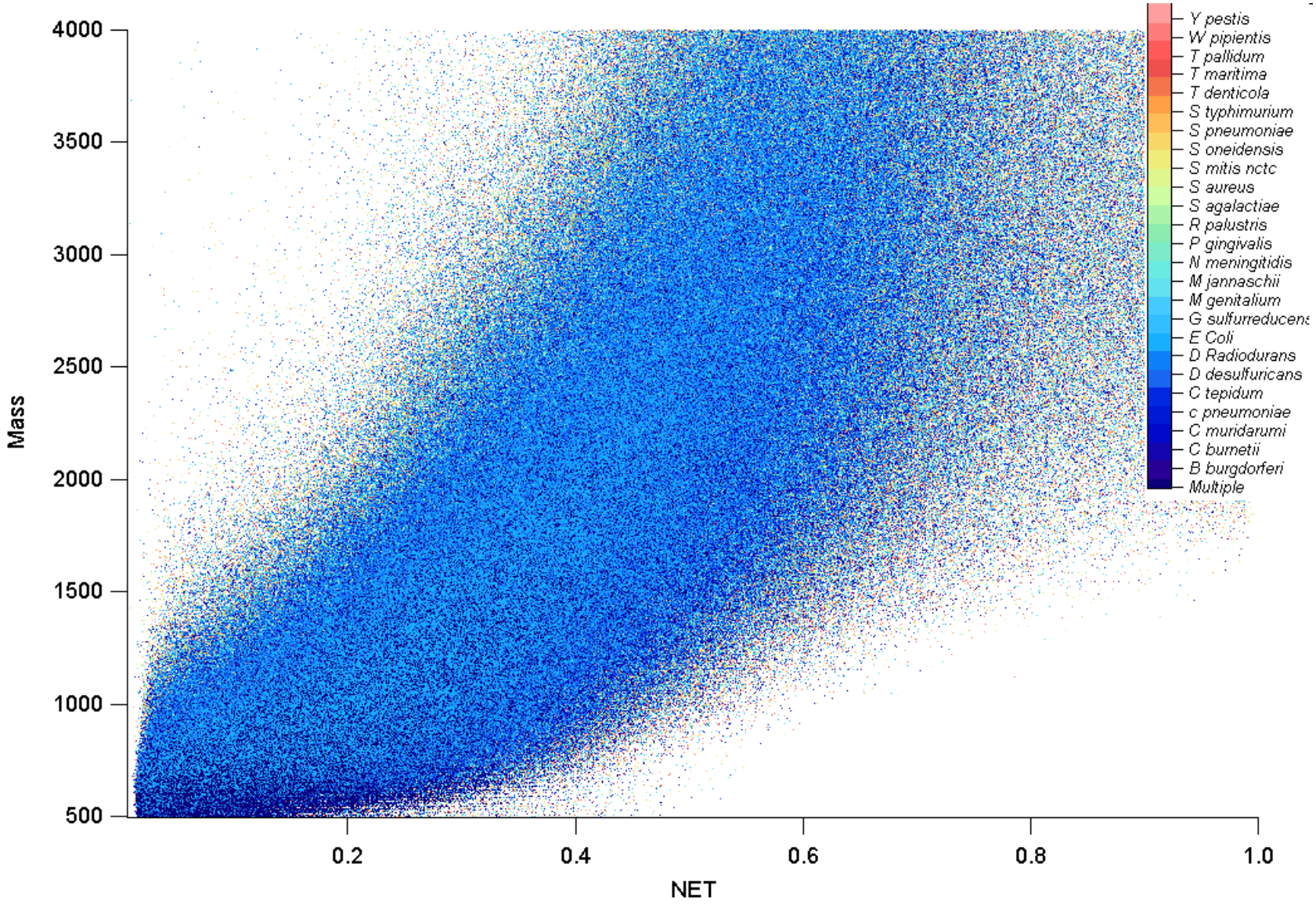
Bacterial Community

- Capable of radical impacts on human health and the health of our environment (evolution, disease, corrosion, degradation, biorem
- 99% o with many being
- Enviro ction & evolve unders
- Multipl ynamics (popul
- **Speci** differentiate one pe peting inform
- **Sensi** dances within a community versus the dynamic range of instrument detection



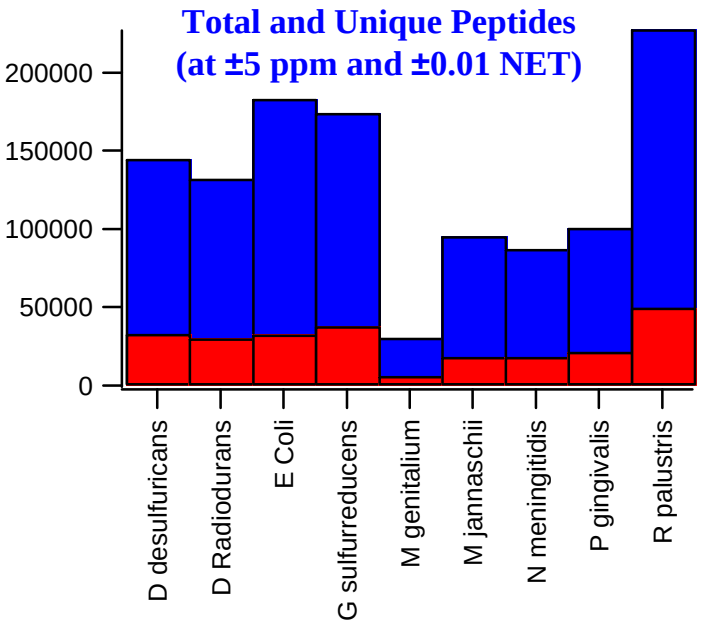
Dynamics of terminal restriction fragments (bacteria) within a phenol fed batch reactor microbial community. Source: Ayala-del Rio, H., Callister, S.J., et al. 2004. AEM 70(8):4950-4960.

~2.8 million tryptic peptides (allowing 1 missed cleavage) from 25 microbes



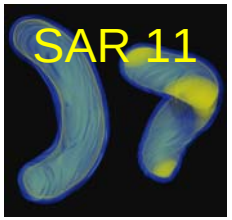
MMA & NET Requirements

Organism	Number of Amino Acids	Total *in Silico* peptides	5 ppm and 0.01 NET Percent Unique	1 ppm and 0.01 NET Percent Unique
D desulfuricans	1,008,695	144,870	23%	48%
D Radiodurans	963,879	132,105	23%	47%
E Coli	1,358,990	183,168	18%	37%
G sulfurreducens	1,139,133	174,184	22%	46%
M genitalium	174,919	30,593	20%	41%
M jannaschii	505,037	95,610	19%	41%
N meningitidis	593,191	87,226	21%	45%
P gingivalis	646,417	100,845	22%	46%
R palustris	1,586,736	227,623	22%	47%
B burgdorferi	432,403	73,276	21%	41%
C burnetii	569,561	89,802	22%	46%
C muridarumi	324,399	48,907	21%	45%
c pneumoniae	365,118	54,314	22%	46%
C tepidum	629,322	98,661	21%	46%
S agalactiae	618,957	94,699	21%	42%
S aureus	772,978	115,414	22%	45%
S mitis nctc	584,835	78,896	13%	27%
S oneidensis	1,422,029	182,728	23%	48%
S pneumoniae	597,559	87,297	14%	29%
S typhimurium	1,423,622	193,846	18%	38%
T denticola	864,153	144,704	21%	45%
T maritima	581,312	102,984	20%	43%
T pallidum	350,744	56,462	21%	45%
W pipientis	335,882	56,523	21%	43%
Y pestis	1,320,997	170,589	22%	46%
	Total:	2,825,326		



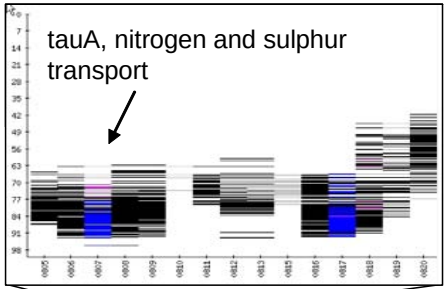
As the number of organisms grow, the fewer the number of differentiating peptides!

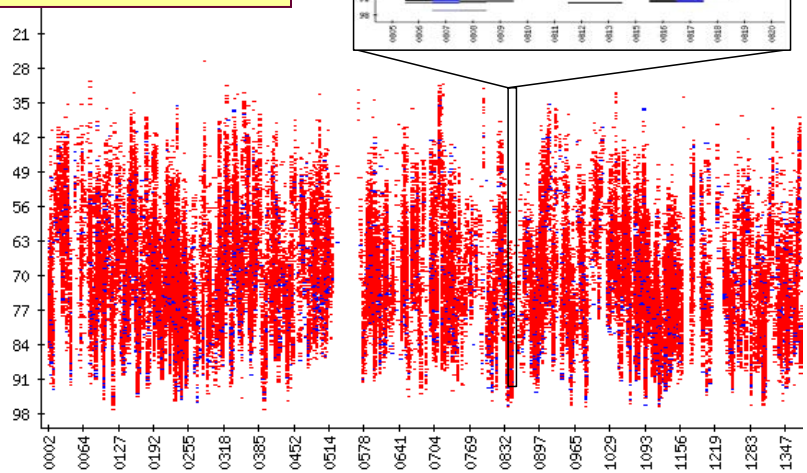
Community profiling: current status

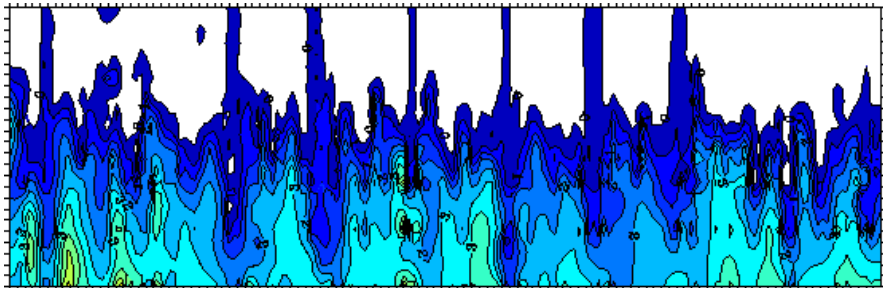


Pelagiobacter ubique **accounts for ~25% of the bacteria in the oceans and may play a large role in carbon sequestration.**

Syntig plots show the natural variation of genetic profiles, and proteomic data are overlaid to show expression.







Proteomic data collected over time will be mapped and correlated to known biological cycling.

Stephen Giovannoni
Douglas Barofsky
Sarah Sowell
Larry Wilhelm

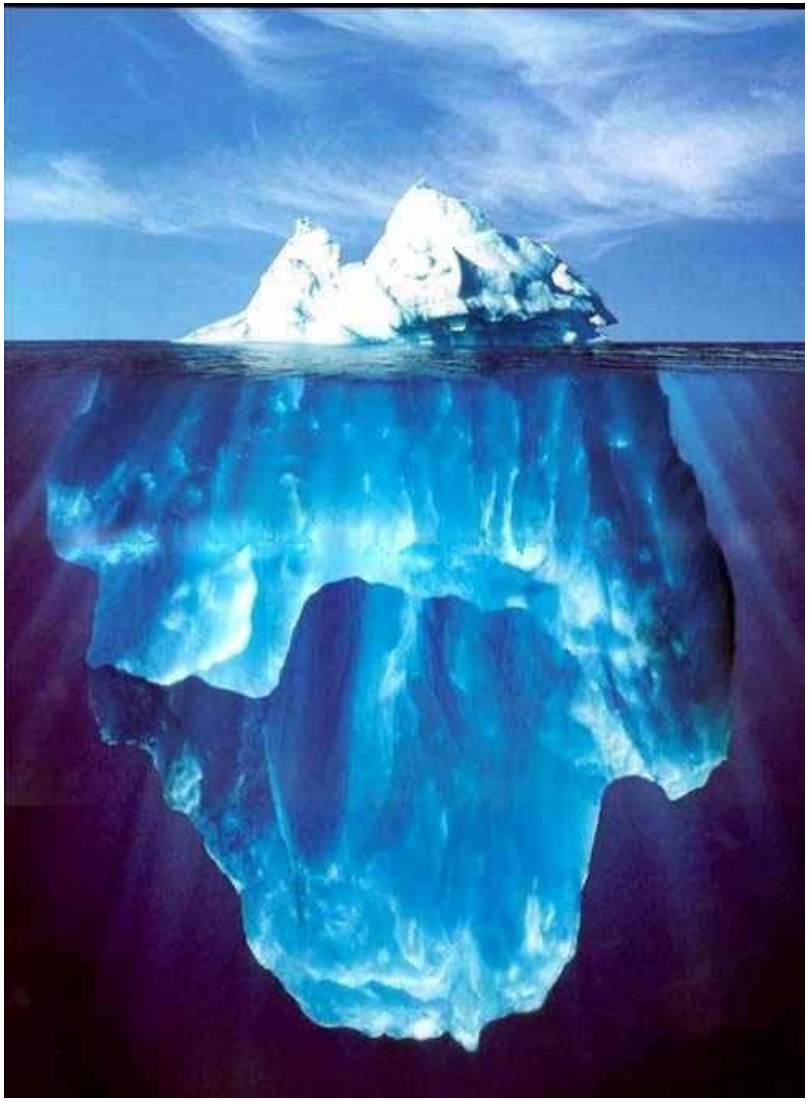
Oregon State University OSU
Department of Microbiology

High field FTICR for improved specificity & selectivity (i.e. more unique peptide identifiers)

Dynamic range &
Limit of detection

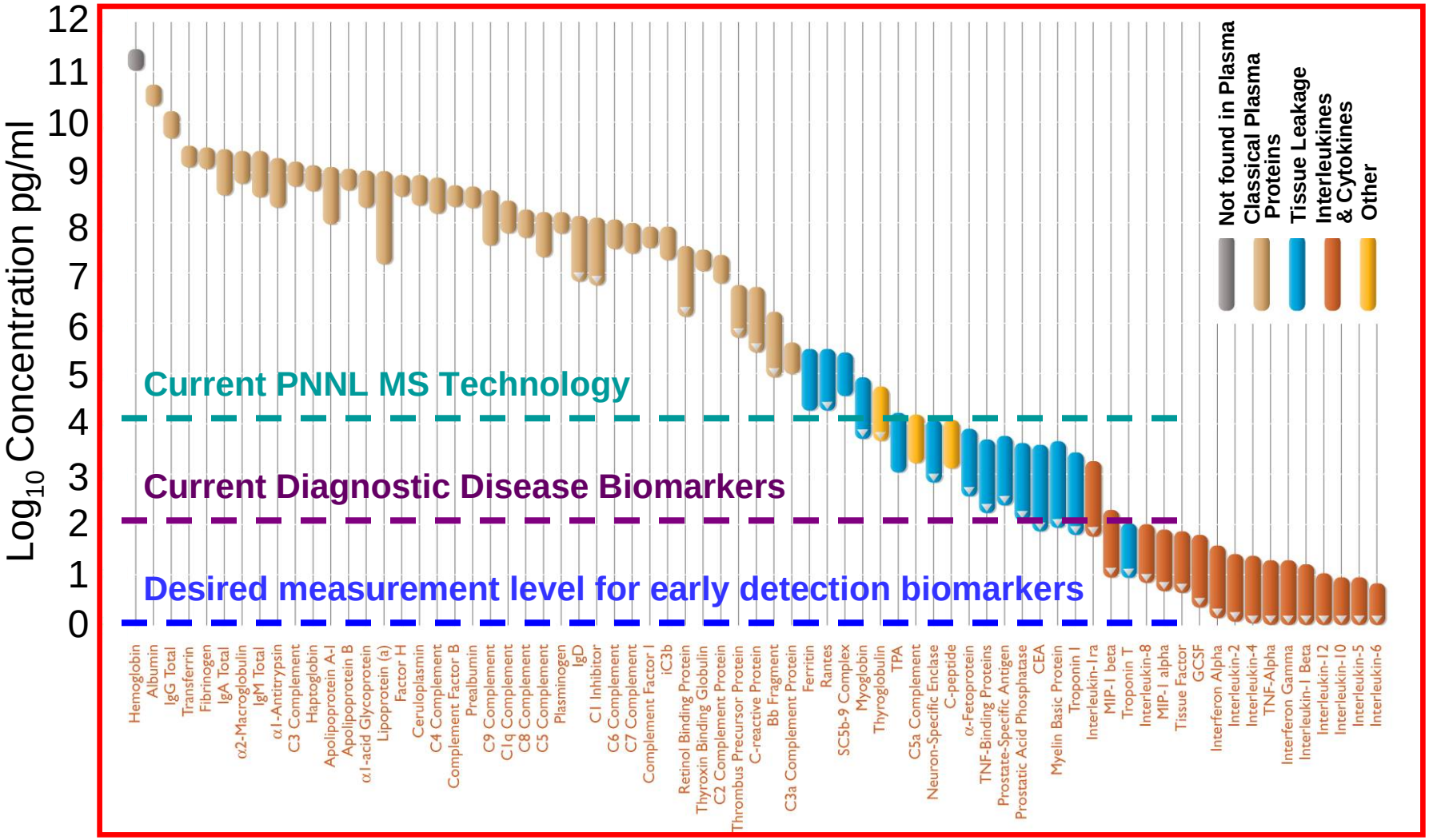
Proteomics Iceberg Paradigm:

Low # of highly abundant proteins, but high protein amount



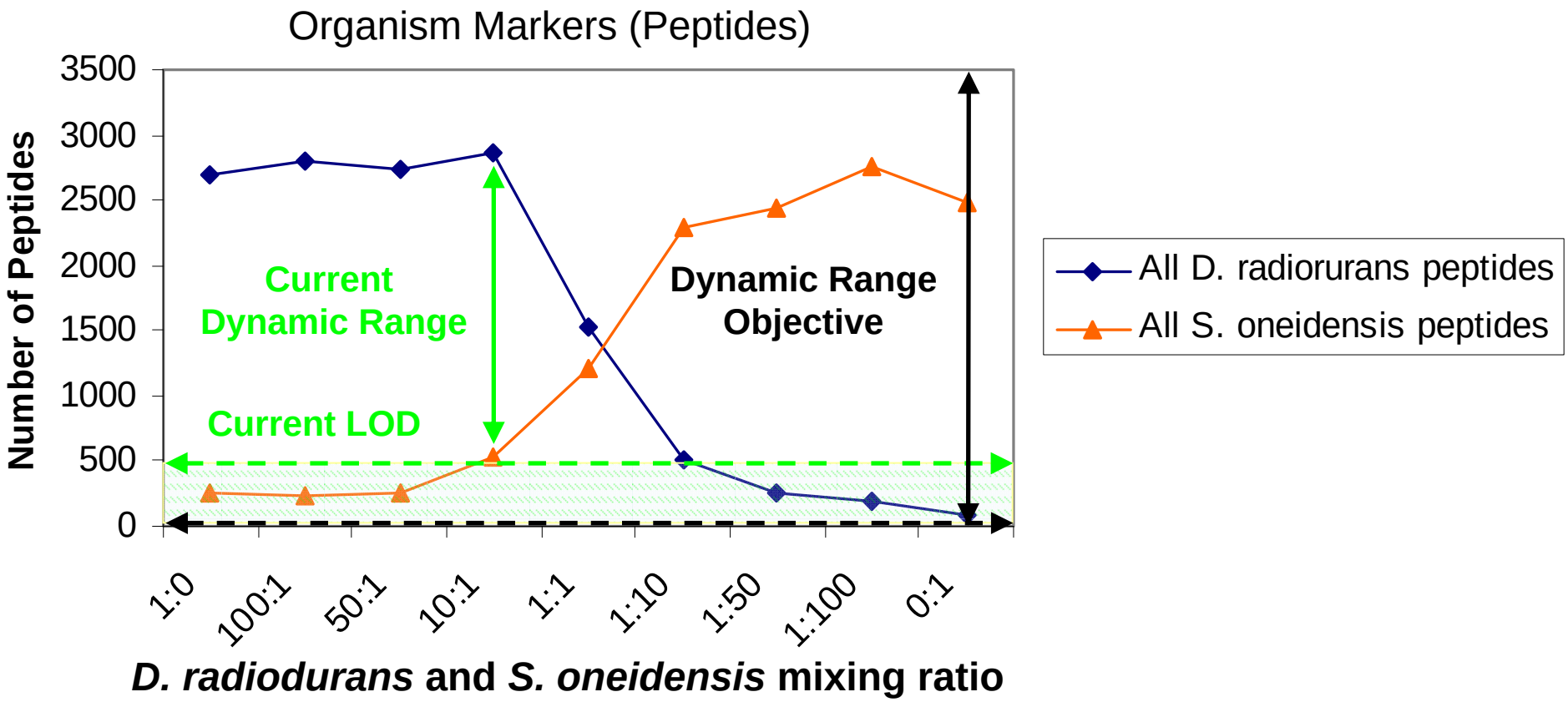
High # of low abundant proteins, but low protein amount

>10^{10} Dynamic Range of Protein Expression in Human Plasma



Anderson, NL & Anderson, NG (2002) *Mol. Cell Proteomics* **1** (11):845-867

Probing the metaproteome of a bacterial community is limited by dynamic range and sensitivity of mass measurements

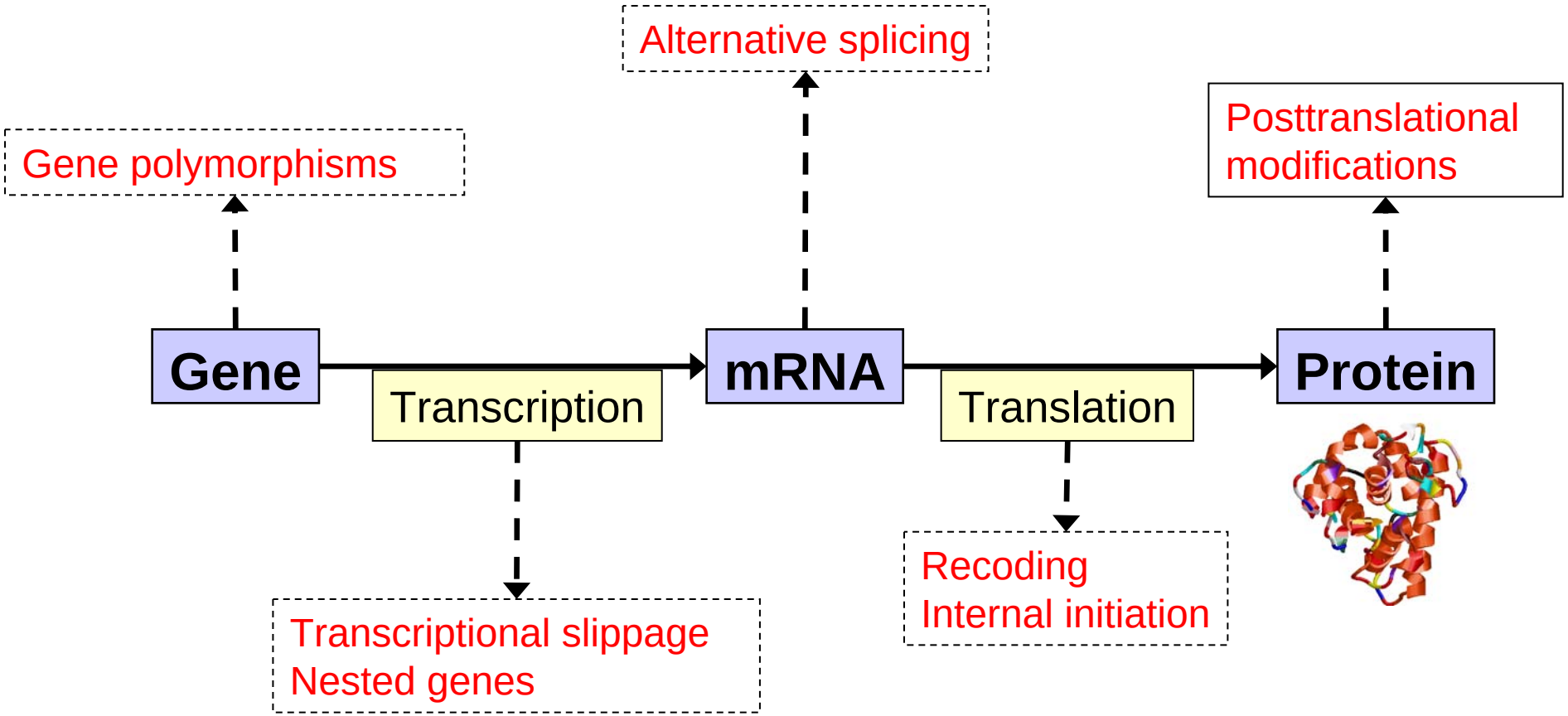


AD Norbeck at al., J. Microbiol Meth. 67(3):473-86 (2006).

LOD Objective: Single cell detection
Dynamic Range Objective: 100-fold

Protein Isoforms & PTMs: Top-Down Proteomics

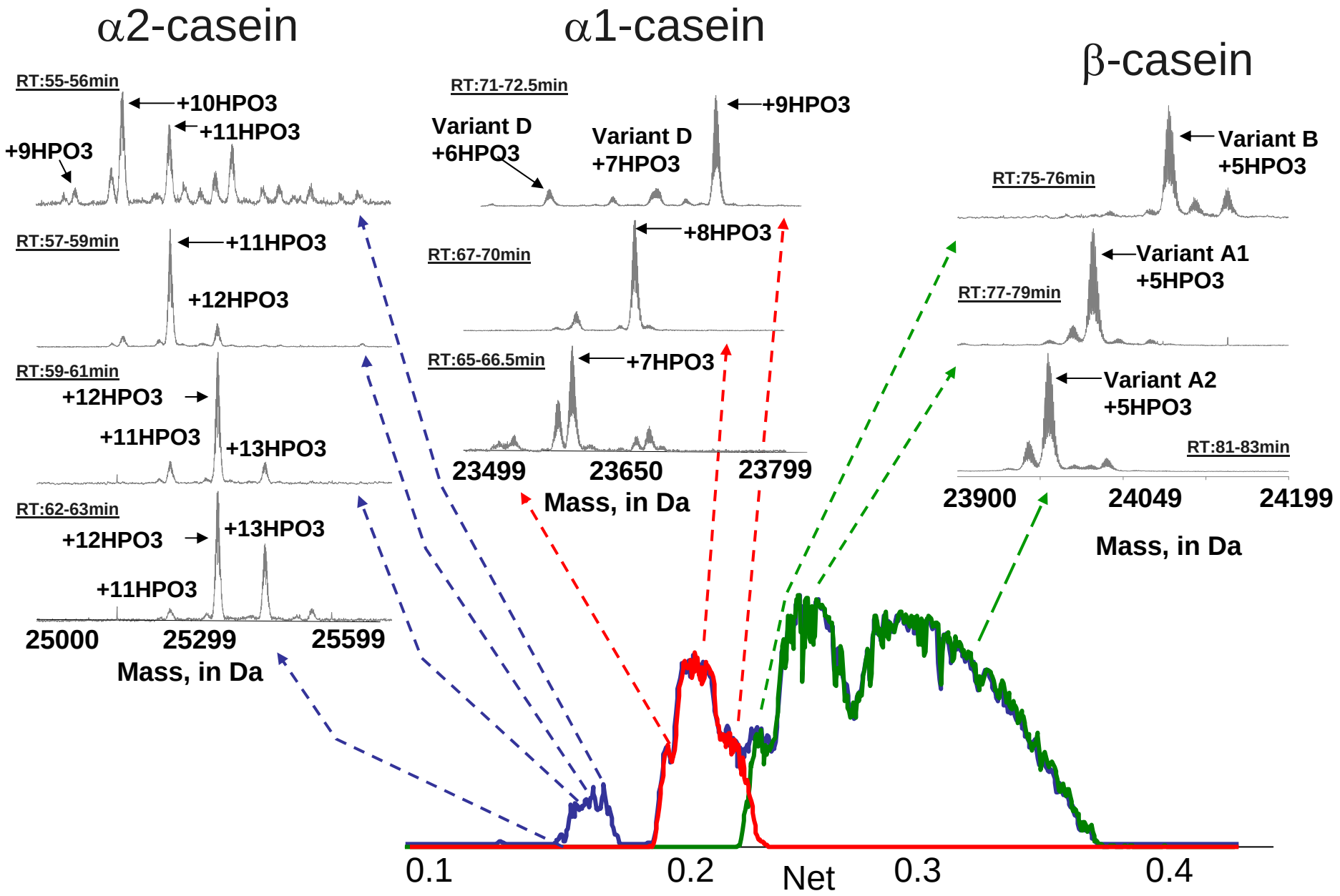
Looking at intact proteins: the rationale



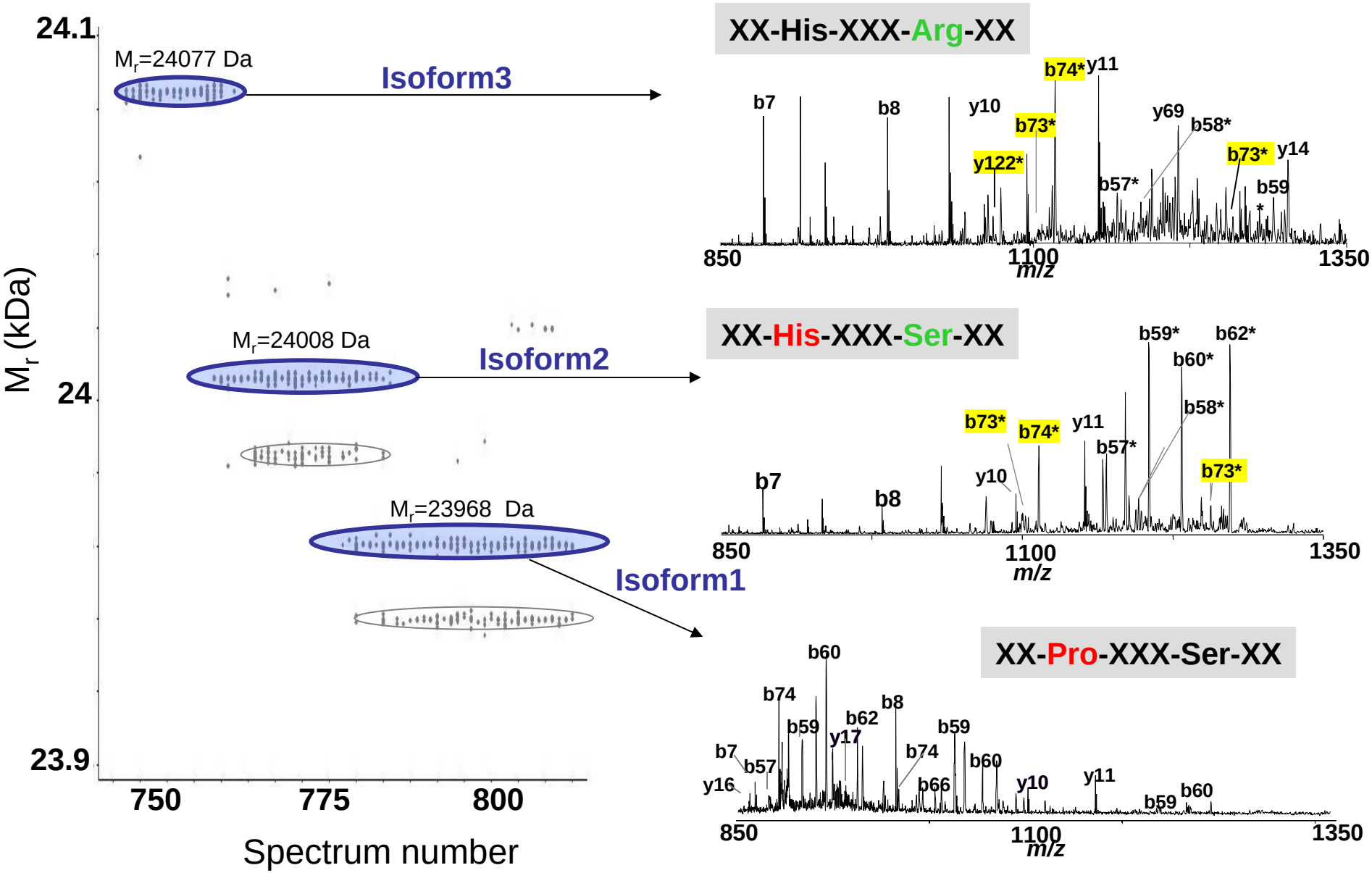
38,016 protein products possible from a single gene in Drosophila

~ 2.5 splice variants; 3 PTMs; 8 distinctive states = 680, 000 protein form possible from ~25,000 genes in Human

Structural characterization of casein isoforms using integrated approach

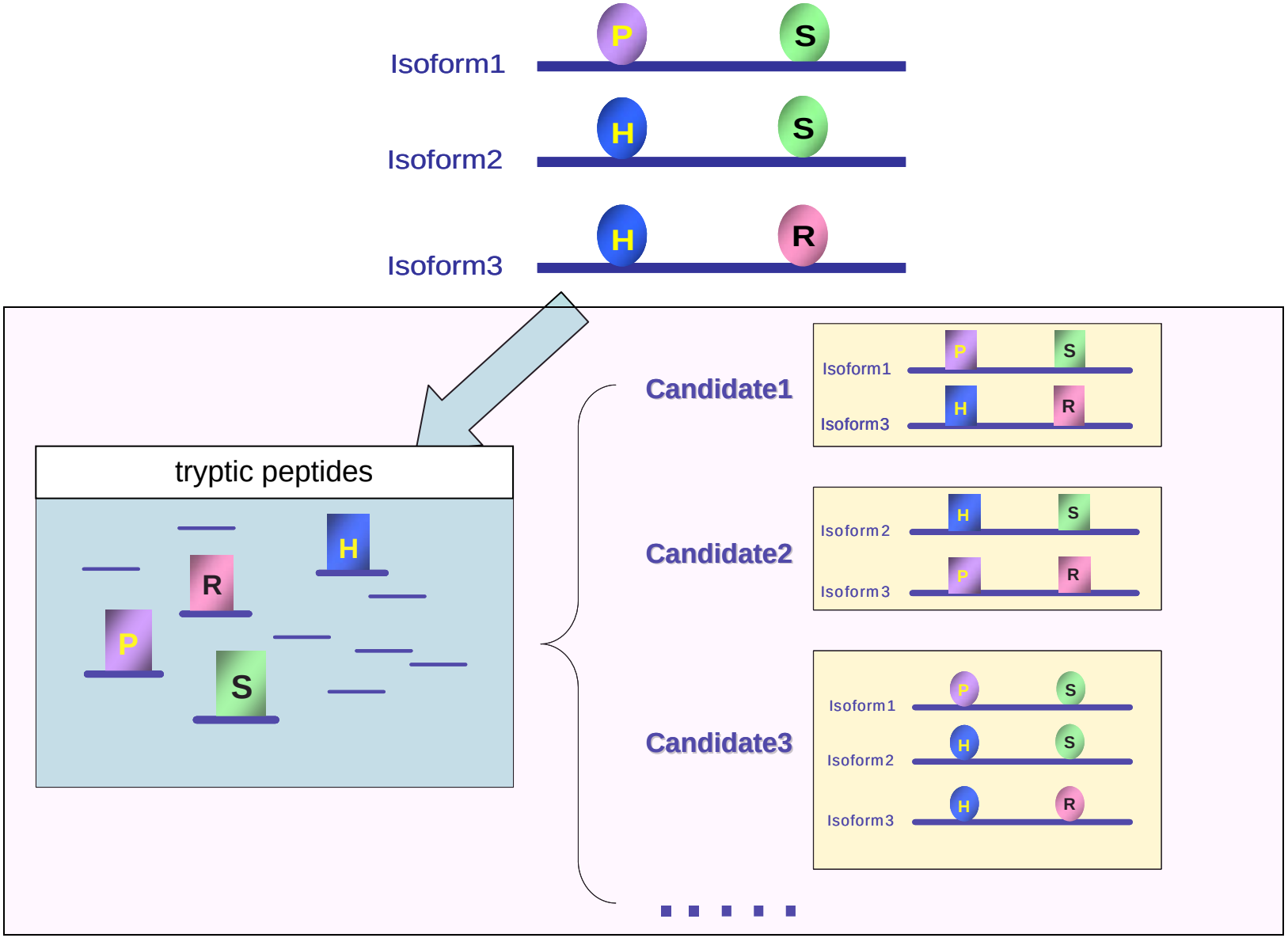


β-casein isoforms from top-down



Putting Humpty Together Again

β-casein isoforms: top-down vs. bottom-up



Intact Proteins

SPECIAL FEATURE METHODS TO WATCH

Top-down mass spectrometry

Top-down mass spectrometry offers the ability to sequence intact proteins—post-translational modifications and all—but is not yet a high-throughput method.

As the well-established 'bottom-up' mass spectrometry–based approach continues its success in high-throughput proteomics, an emerging approach known as 'top-down' is beginning to make headlines, especially for the analysis of post-translational modifications (PTMs).

PTMs often occur in different combinations on individual proteins, and understanding these combinations is crucial for understanding biological regulation, such as for 'cracking' the histone code. In a bottom-up experiment, the proteomic mixture is digested into short peptides before analysis, so information about the correlated relationships of different PTMs is lost. Compounding the problem is the fact that many PTMs are unstable under typical mass spectrometry conditions, and that the mass spectrometer does not detect every last peptide, so minor PTMs occurring on a small percentage of proteins are often not observed.

In the top-down approach, intact proteins are introduced into the mass spectrometer, so important information about combinatorial PTMs is retained. In recent years, highly efficient fragmentation methods have been developed (such as electron capture dissociation and electron transfer dissociation) that are particularly good at preserving labile PTMs. The mass range of top-down has been extended to proteins as large as 229 kDa (*Science* **314**, 109–112; 2006), and increasingly larger numbers of intact proteins can be detected in a single analysis. Yet top-down is still mainly a technique for analyzing single purified proteins.

Currently, larger sample quantities are required and the analysis time is longer than for a bottom-up experiment, precluding high-throughput analyses. New methods are needed for efficient protein separation, and robust computational tools for assigning protein identities and PTMs from top-down data are also lacking.

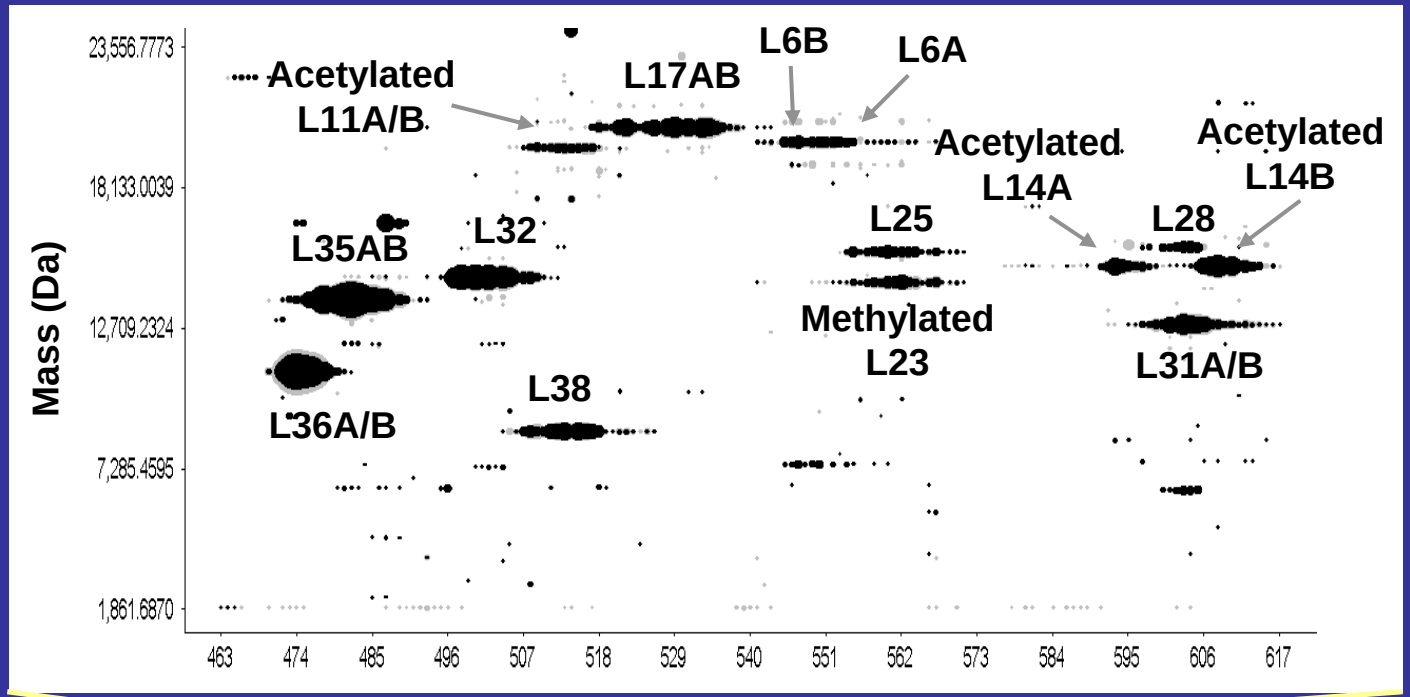
So stay on the lookout for new methods driving the limits of top-down mass spectrometry. Perhaps one day this approach will be the method of choice for investigating the biological importance of combinatorial PTMs.

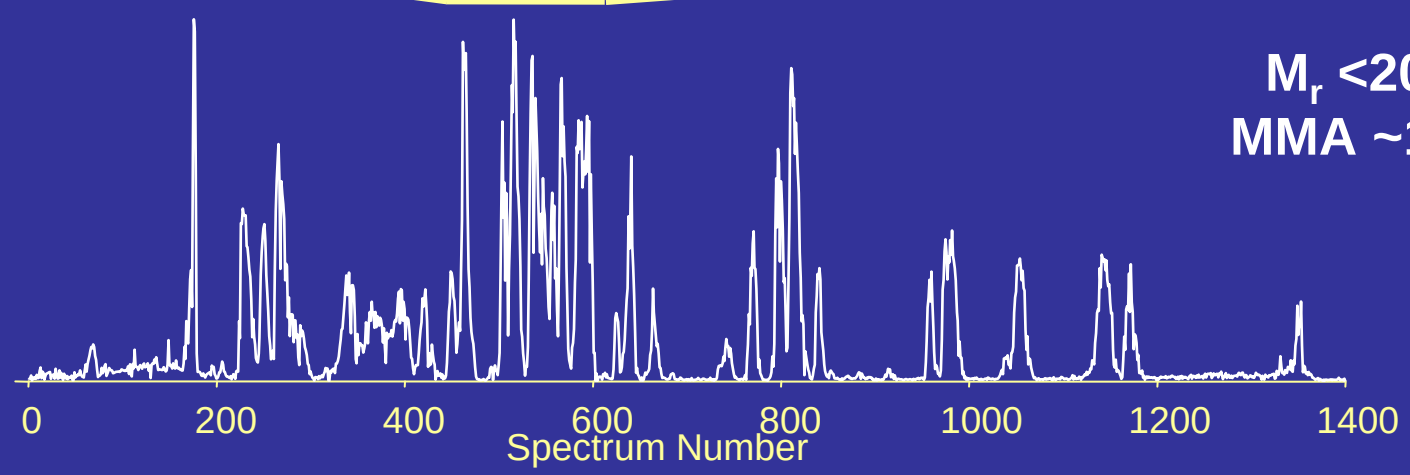
Allison Doerr

Deciphering the histone code could be aided tremendously by high-throughput top-down mass spectrometry.

© 2008 Nature Publishing Group http://www.nature.com/naturemethods

Analysis of Intact Proteins of the Yeast Large Ribosomal Subunit using RPLC-FTICR MS operating at 7T

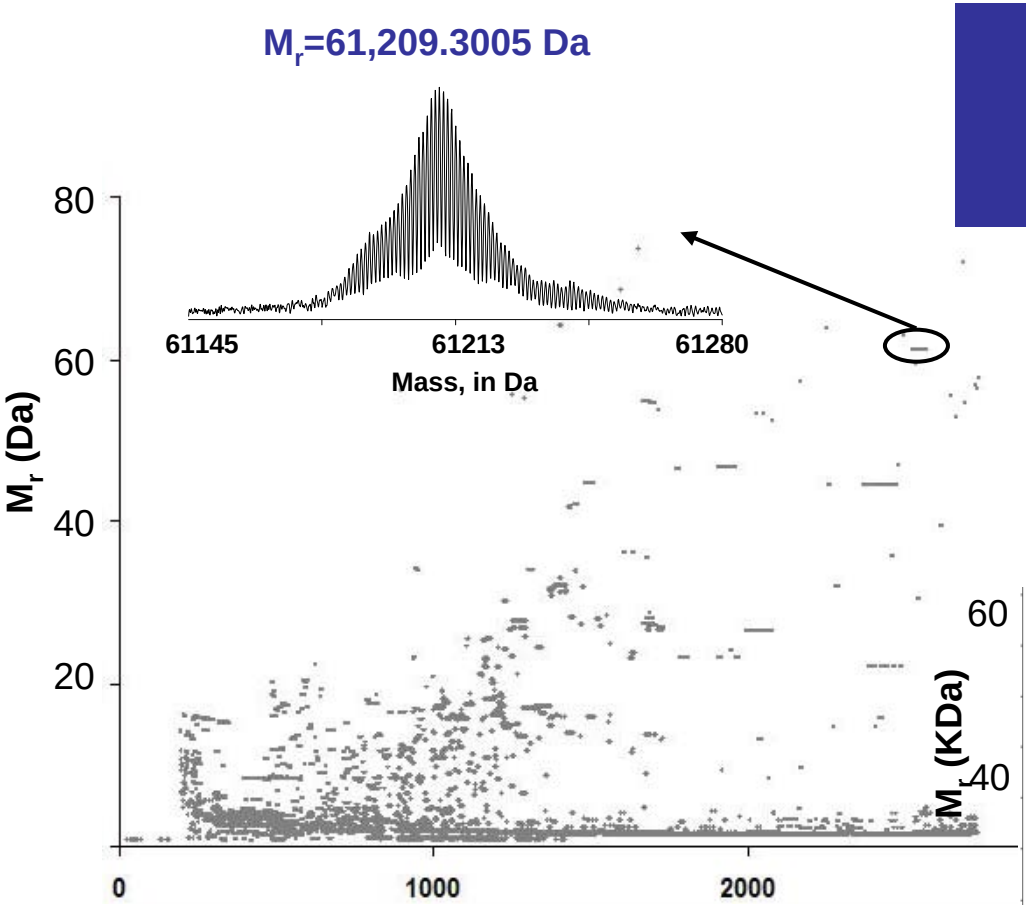


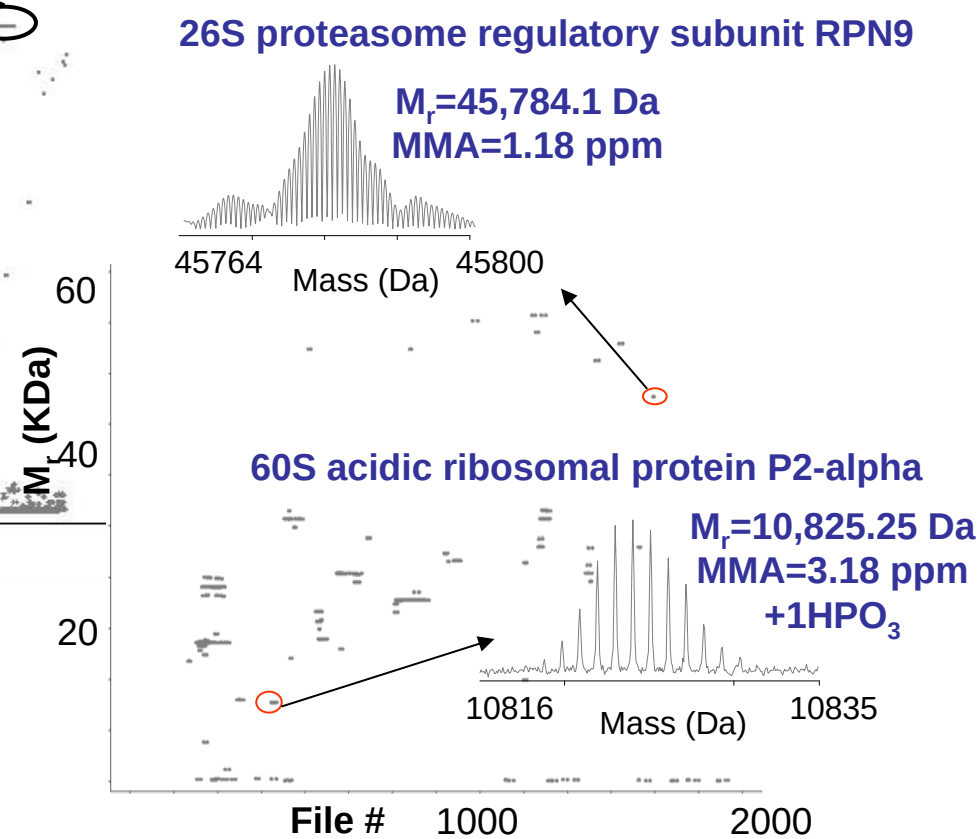


M_r <20 kDa
MMA ~10ppm

S.-W. Lee et al., *PNAS* 99 (2002) 2076-2084

Yeast lysate & proteasome at 12T



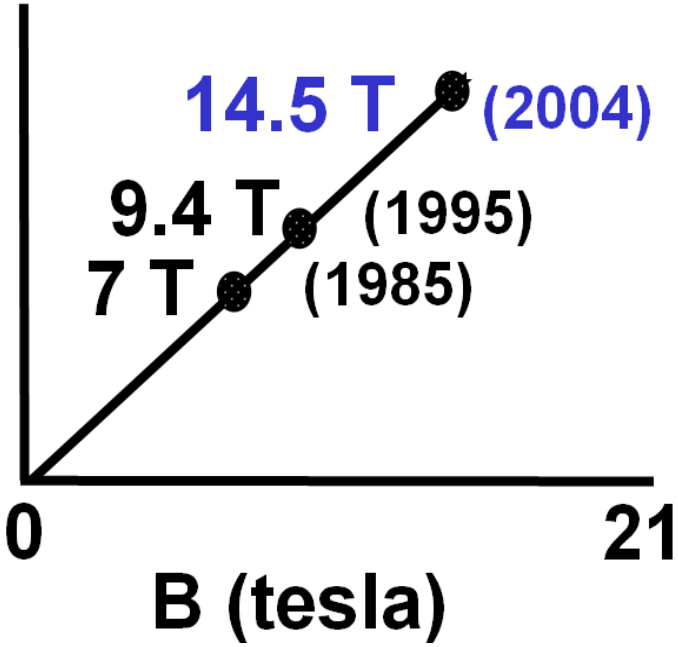


>1,000 proteins
5 kDa < M_r <70 kDa
MMA~1-5 ppm

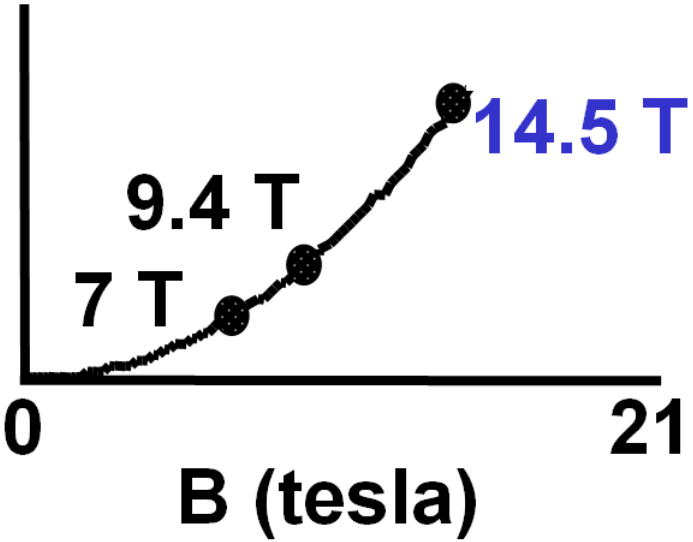
High-Field FT-ICR MS will enable intact protein AMT Tag Proteomics

The Science Case for High Field FTICR Proteomics

Mass resolving power
Scan Speed (LC-MS)
Highest Non-Coalesced Mass



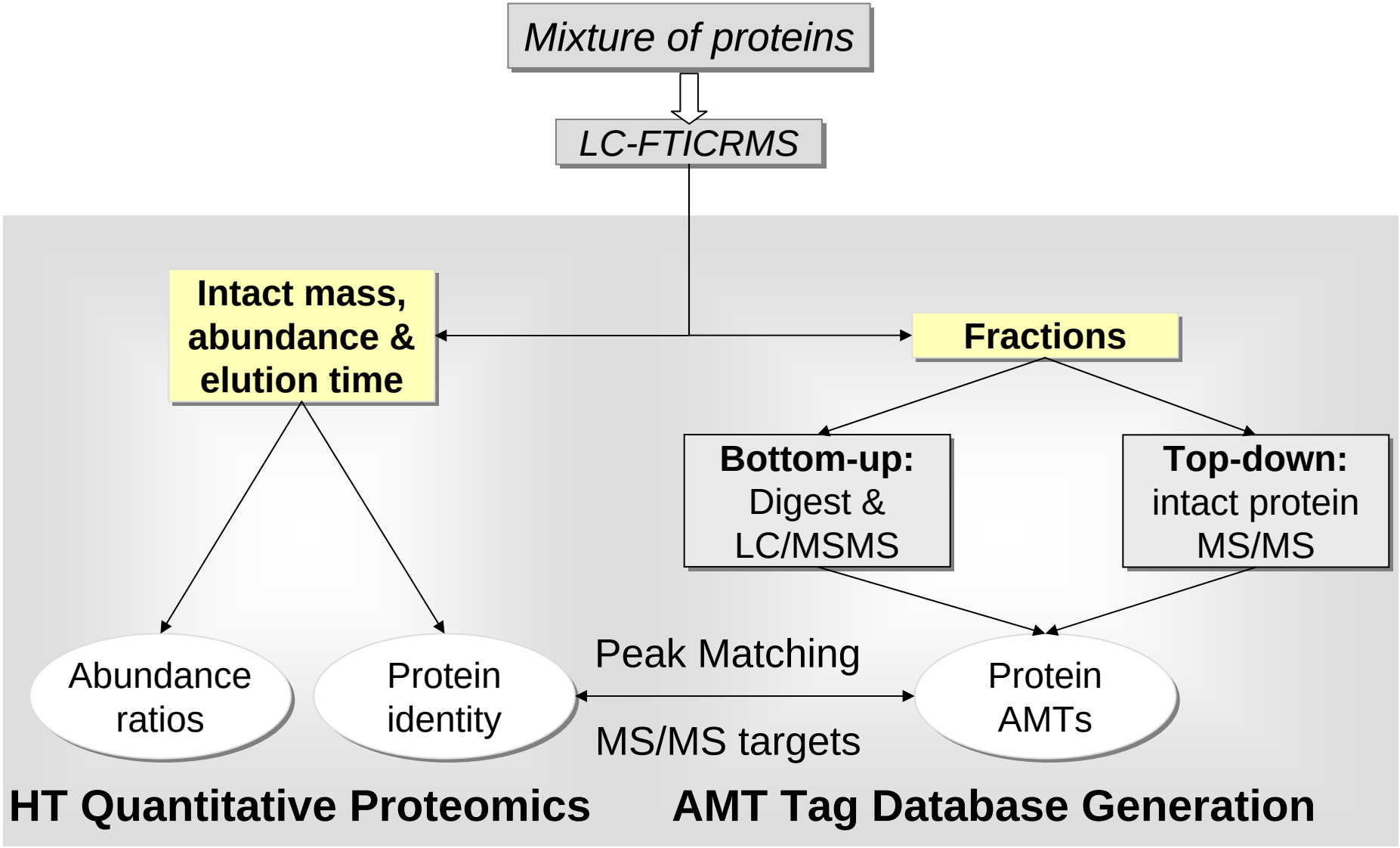
Number of Ions
Dynamic Range
Upper Mass Limit
Ion Energy
Ion Trapping Period



"Experiments that can be performed only with heroic effort at low field become routine at high field!"

Acknowledgment

All present and past members of
the EMSL-HPMS Facility &
Bio-separations & Mass Spectrometry group
at PNNL

$ DOE & NIH $

High-Field FT-ICR MS will enable intact protein AMT Tag Proteomics