High Throughput proteomics applied to microbes and microbial communities

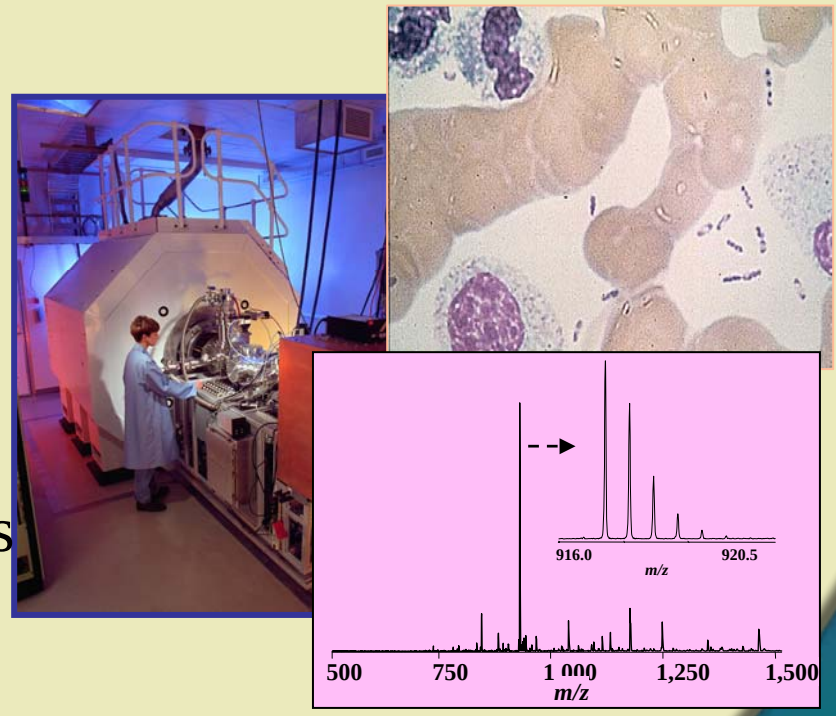
Mary S. Lipton, Ph.D.

Senior Research Scientist

Biological Separations and Mass Spectrometry

Biological Sciences Division

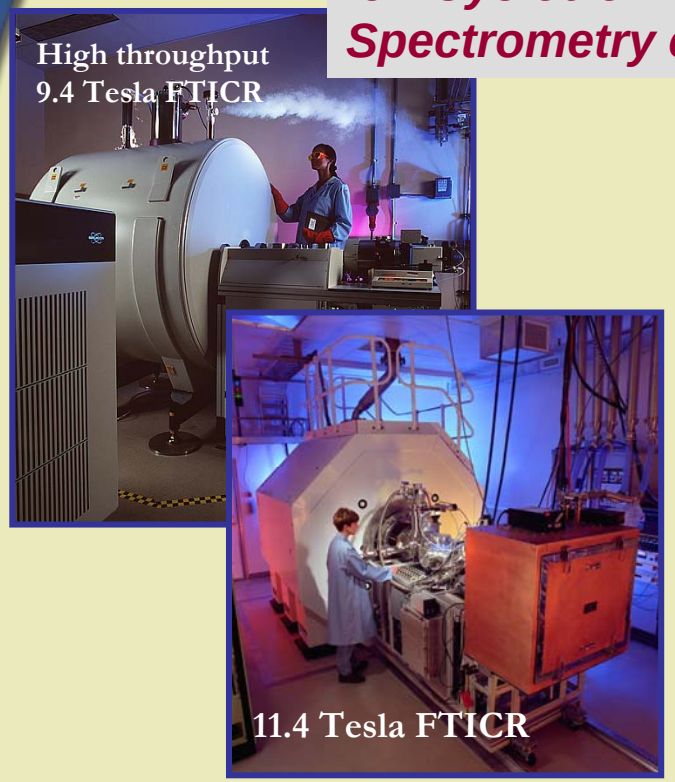
Pacific Northwest National Laboratory



Proteomics

Global proteomics: simultaneous characterization of all the proteins expressed in a cell.

Unique *High Field Fourier Transform Ion Cyclotron Resonance Mass Spectrometry capability*



General Principle

Peptide identification based on

- MS/MS fragmentation pattern
- very high measured mass accuracy
- elution profile

As Practiced

Proteomes are more complex to completely characterize by one parameter alone

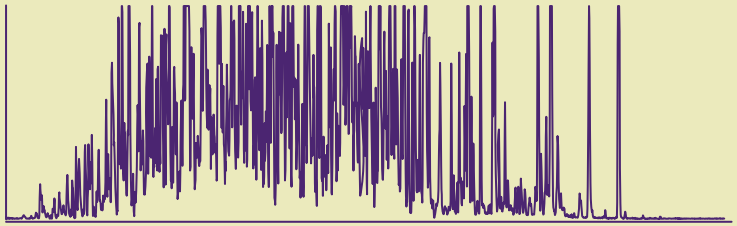
Therefore, we have 2 part process
-First practicing traditional proteomics
-Second using FTICR-MS

PNNL global proteomics capability has provided the most comprehensive map of the proteome of microbial systems to date.

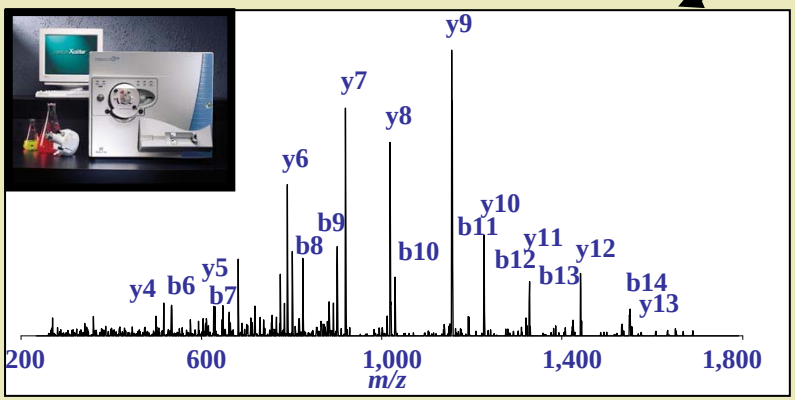
Protein Identification – The Accurate Mass and Time Tag (AMT) Approach Stage 1

Peptide Complex

SCX fractionation



Ultra High Pressure HPLC Peptide Separation



Structural Information from Finnigan LCQ

MS/MS—Putative Mass Tags

Mass & Elution Time calculated

Peptide Mass and Time Tag Database

Serves as a look up table

Reason for using the LTQ ion trap

MS/MS Peptide Identification

- "Cheap", user friendly, robust way to obtain MS/MS
- Provides structural information
- Relatively Poor dynamic range and resolution

Filtered 488 LC-MS/MS runs

Peptide seen >3 times	Xcorr >1.9	
Peptides seen 1 or 2X	Xcorr >=1.9	1+
	Xcorr 2.2	2+
	Xcorr 3.5	3+

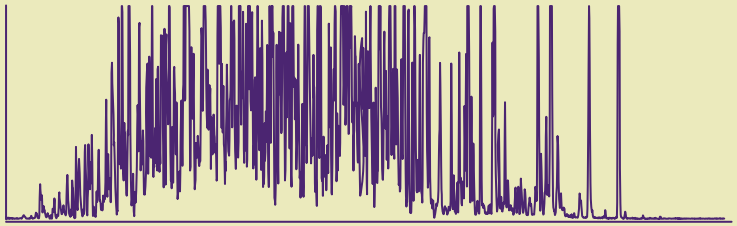
All peptides are partial or fully tryptic

24800 Putative Mass Tags

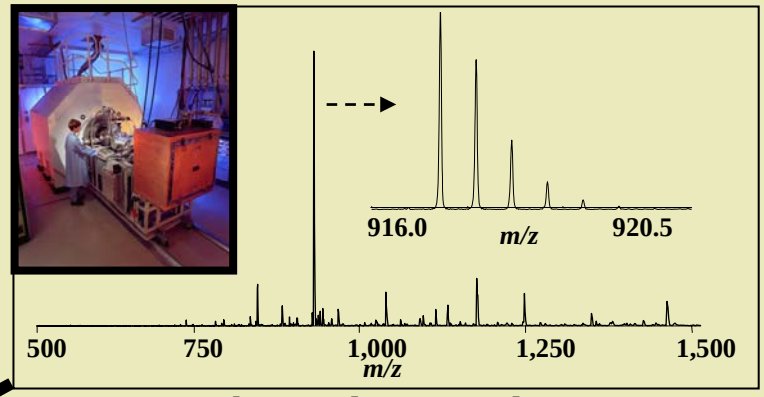
Replaced by Discriminant score

Protein Identification – The Accurate Mass and Time Tag (AMT) Approach Stage 2

Peptide Complex



Ultra High Pressure HPLC Peptide Separation



High Resolution and Mass Measurement Accuracy **FT-ICR MS**

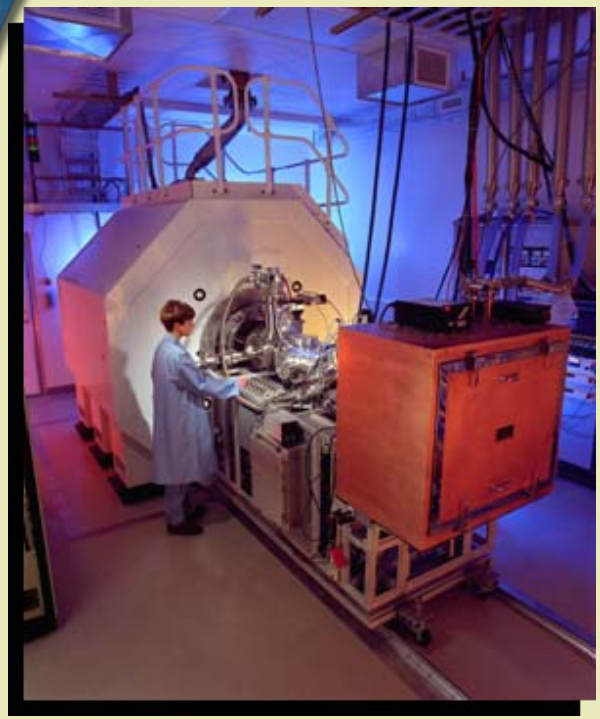
Mass & Elution Time Matched

Peptide Mass and Time Tag Database

Serves as a look up table

Validated Accurate Mass Tag

Reason for using the FT-ICR MS



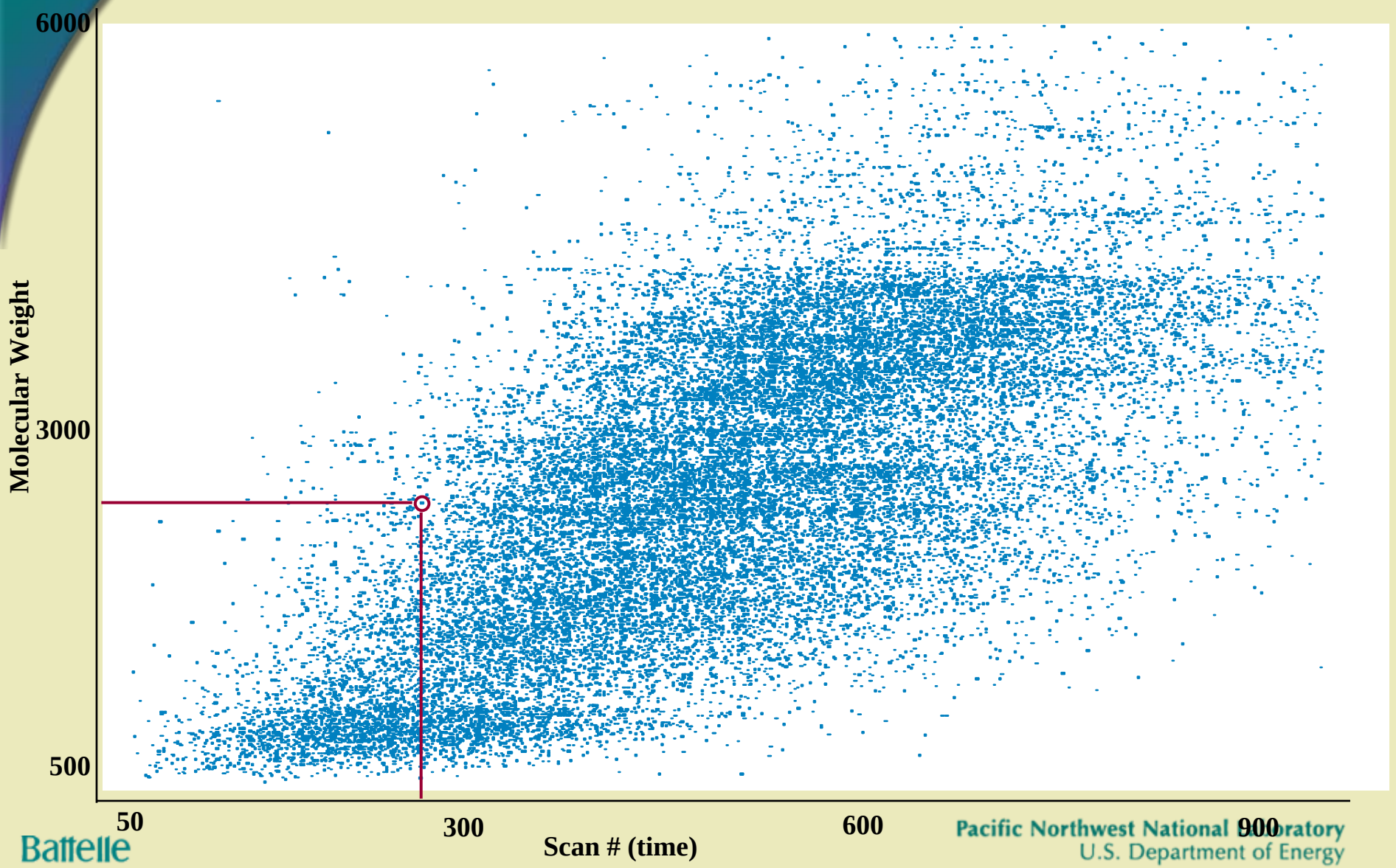
Parent Peptide MS

1)Peptide Validation

2)Abundance Value

- Provides high mass measurement accuracy
- High sensitivity/High resolution
- Large Dynamic Range
- Reproducible Peak Intensities

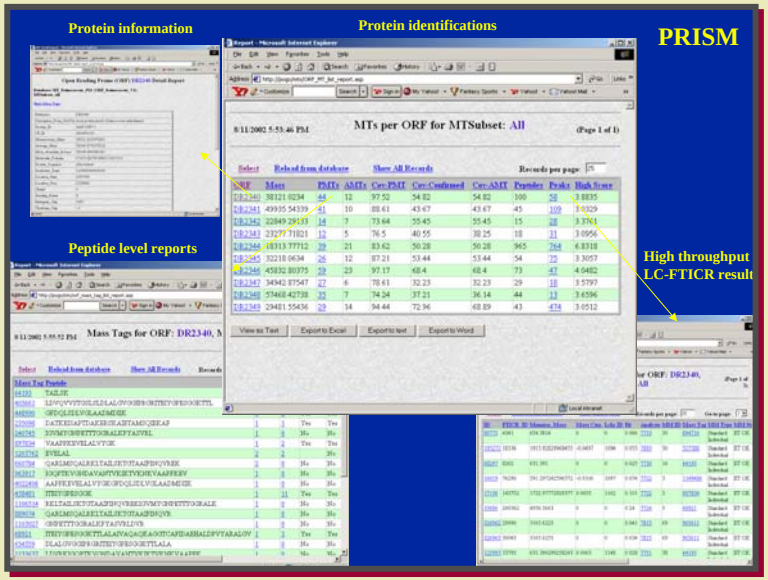
2-D Representation of AMTs



Enabling High Throughput Proteomics

FT-MS Proteome Express

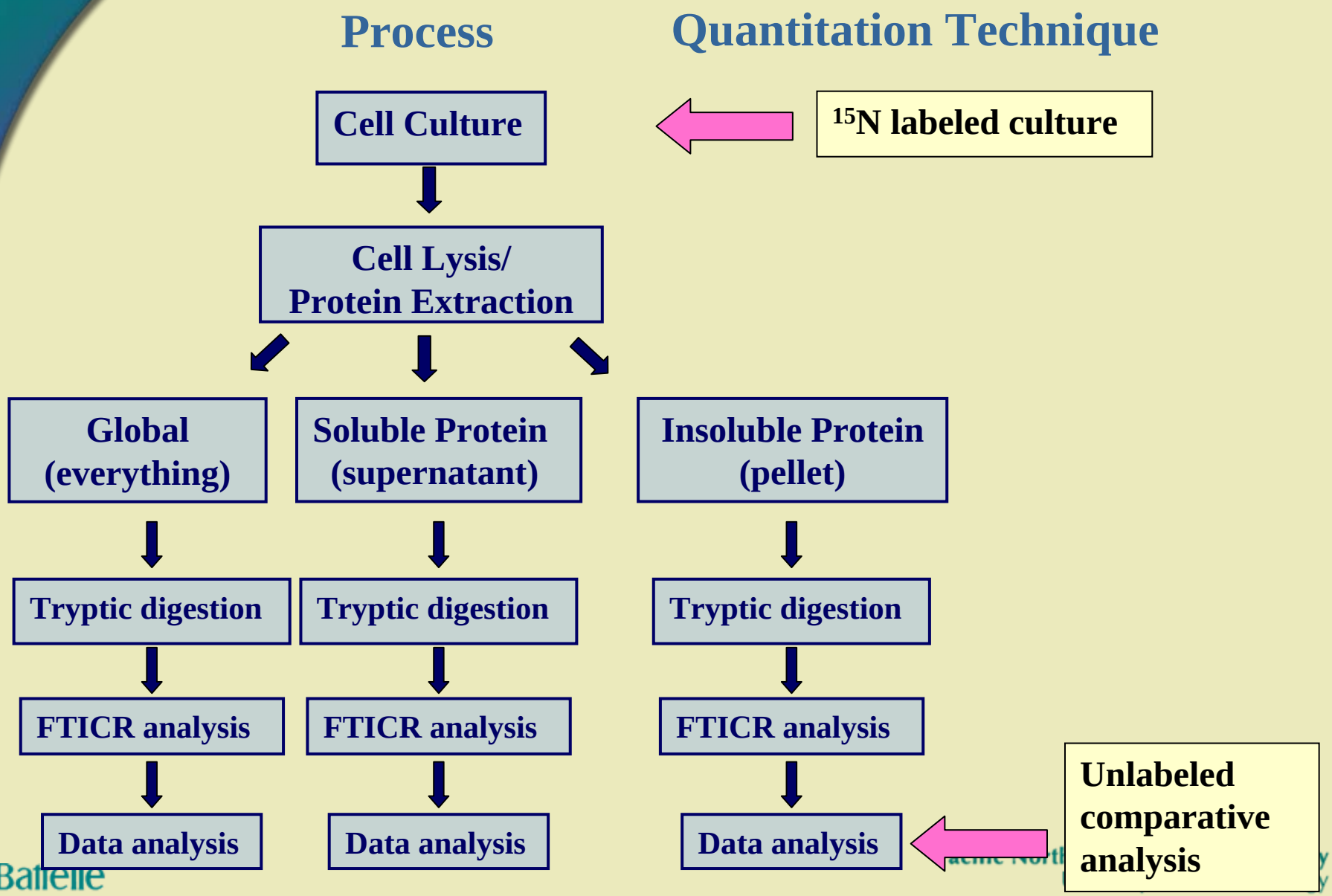
- Automated 24/7 operation of the capillary LC-FTICR instrumentation increases both throughput and data quality.
- Specificity provided by high resolution separations and mass measurements enables analysis of complex systems, *e.g.*, microbial communities.



Data Throughput Management (PRISM)

- Automation decreases analysis turnaround time.
- Easy-to-use web interface improves data accessibility.

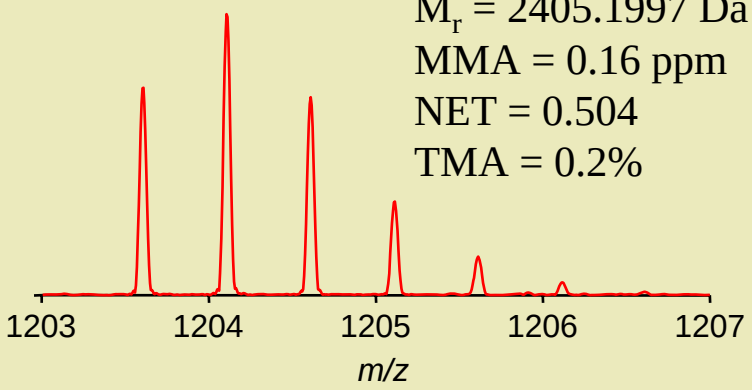
Single Microbe Characterization
Quantitative/ Comparative Proteomics

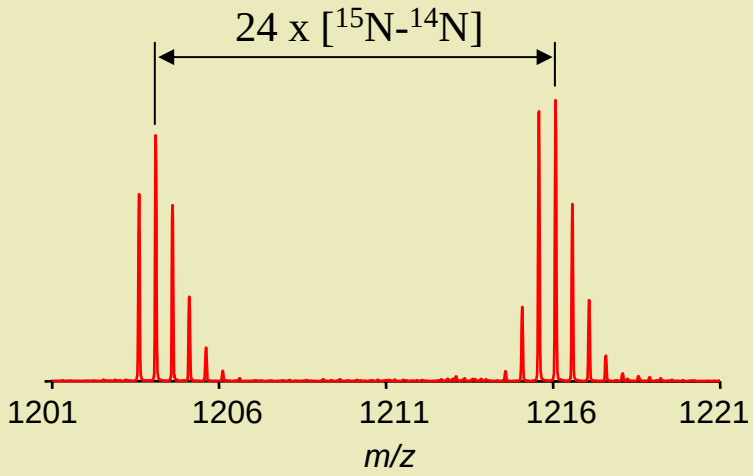


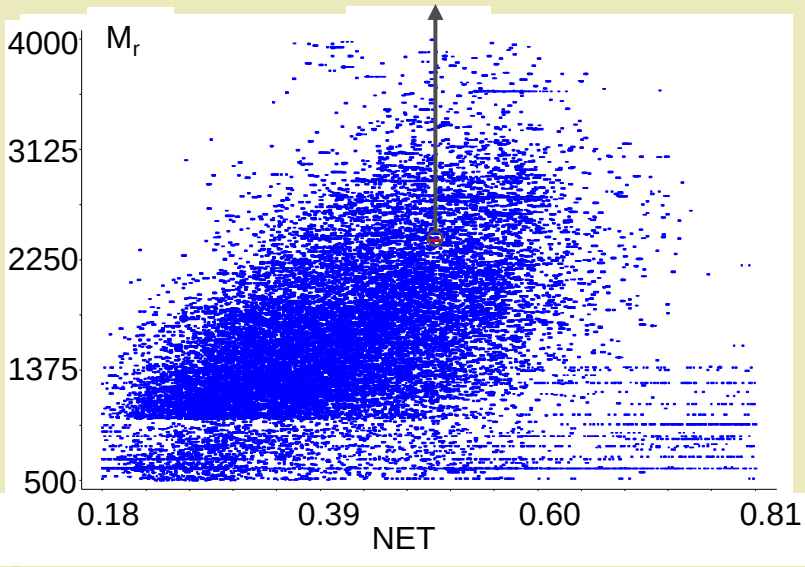
Single Microbe Characterization
Quantitation by Stable Isotope Labeling

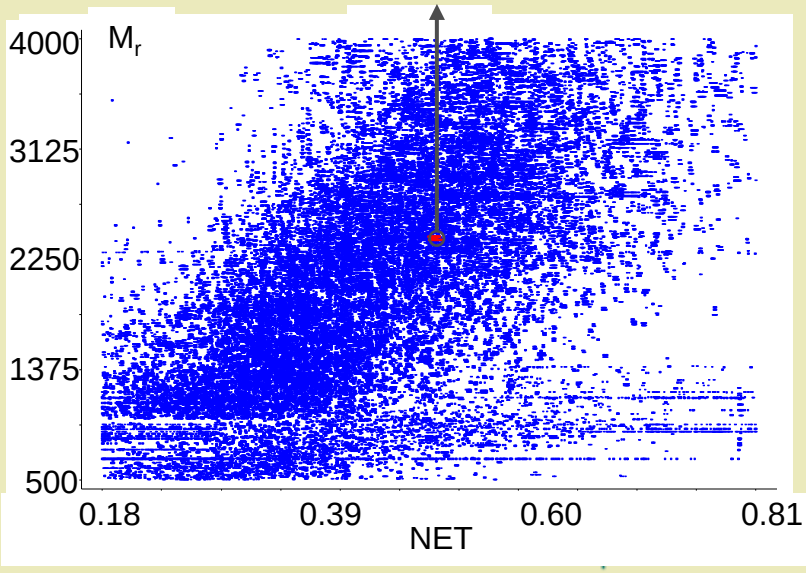
AMT #1146 [ORF #02142]
translation elongation factor Tu (tufB)
ELLSEYDFPGDDLPVIQGSALK
Theor. M_r = 2,405.2001 Da
NET = 0.502

M_r = 2405.1997 Da
MMA = 0.16 ppm
NET = 0.504
TMA = 0.2%





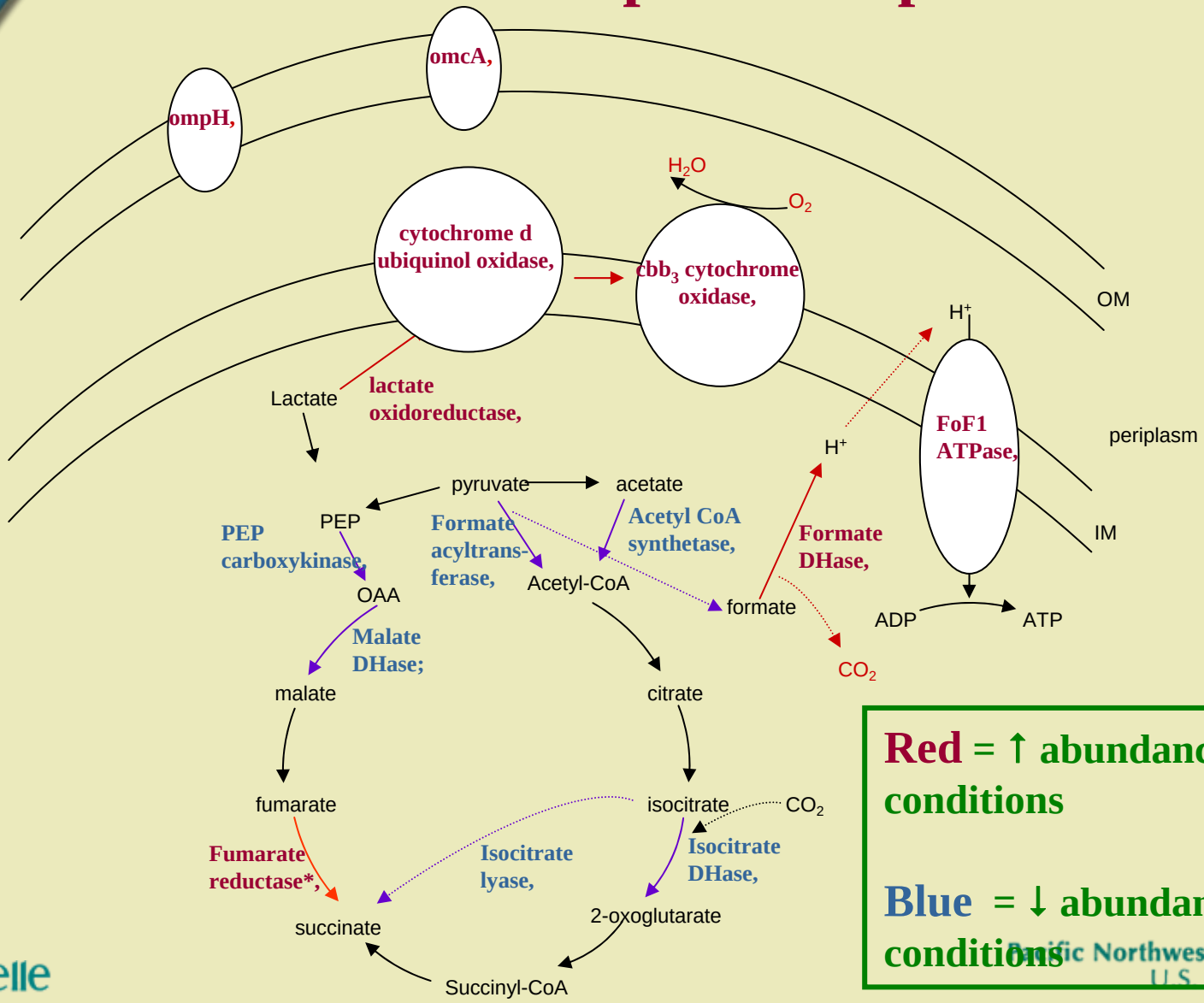




$^{14}N/^{15}N$ comparison of *S. oneidensis* aerobic vs. sub-oxic growth

	Description_From_FASTA
	tRNA 2-selenouridine synthase, ybbB
	type IV pilus biogenesis protein PilB
	hypothetical protein
	fumarate reductase flavoprotein subunit precursor
	chloride channel
	sigma-E factor regulatory protein RseC
	isocitrate lyase
	glutathione-regulated potassium-efflux system protein KefB putativ
SO1679	transcriptional regulator MerR family
SO1755	phosphoglucomutase/phosphomannomutase family protein
SO1776	outer membrane protein precursor MtrB
SO1777	decaheme cytochrome c MtrA
SO1778	decaheme cytochrome c
SO1779	decaheme cytochrome c
SO2345	glyceraldehyde 3-phosphate dehydrogenase
SO2361	cytochrome c oxidase cbb3-type, subunit III
SO2602	conserved hypothetical protein
SO2743	acetyl-coenzyme A synthetase
SO2767	asparagine synthetase B glutamine-hydrolyzing
SO2893	conserved hypothetical protein
SO3213	flagellar biosynthesis protein FlhA
SO3765	conserved hypothetical protein
SO3770	conserved hypothetical protein TIGR00153
SO3802	ABC transporter ATP-binding protein
SO3833	peptide chain release factor 1
SO4068	hypothetical protein
SO4312	adenylate cyclase CyaA putative
SO4408	virulence regulator BipA
SO4410	glutamine synthetase type I

Schematic of the TCA cycle and Electron Transport chain illustrating the changes in protein expression

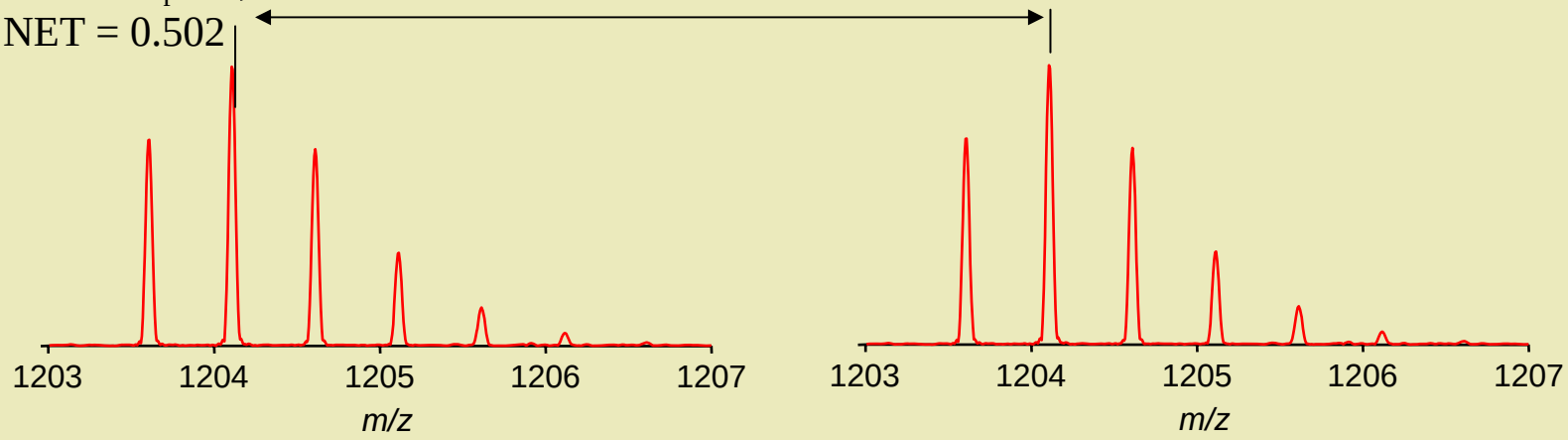


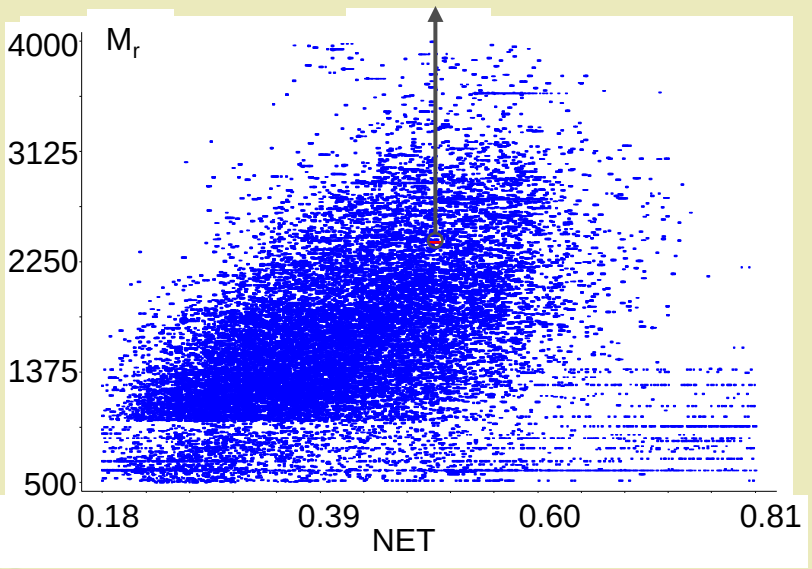
Red = ↑ abundance under suboxic conditions

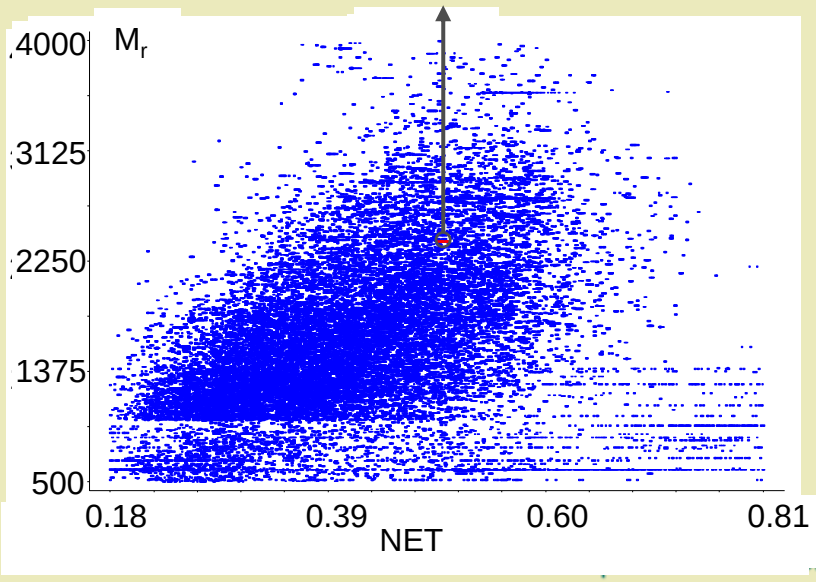
Blue = ↓ abundance under aerobic conditions

Single Microbe Characterization
Quantitation by Absolute Peak Intensity

AMT #1146 [ORF #02142]
translation elongation factor Tu (tufB)
ELLSEYDFPGDDLPVIQGSALK
Theor. M_r = 2,405.2001 Da
NET = 0.502

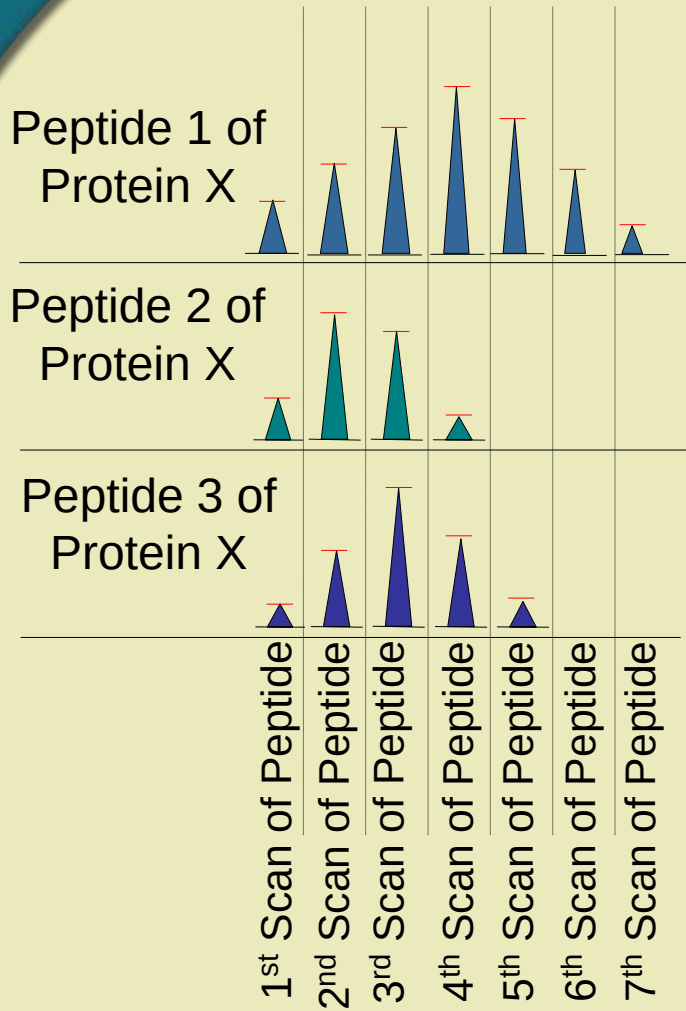






Battell

Single Microbe Characterization
Protein Abundance Value Calculation



Summed Peak Intensity for Peptide 1

Summed Peak Intensity for Peptide 2

Summed Peak Intensity for Peptide 3

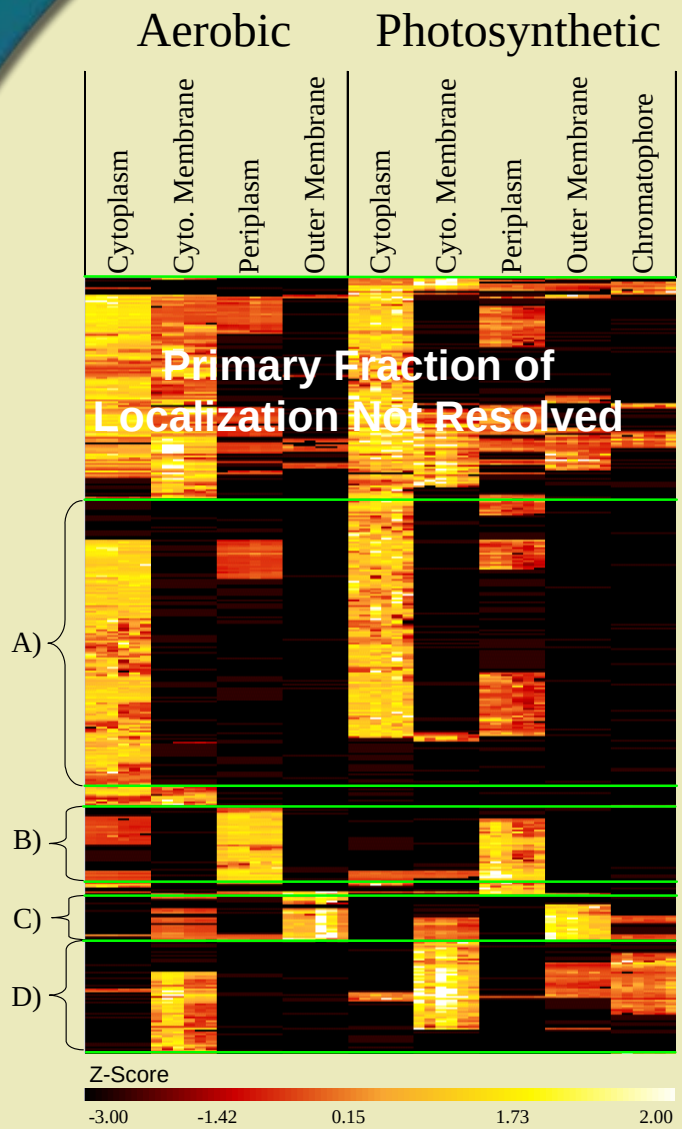
Average of all Summed Peak Intensities for all Peptides of Protein X

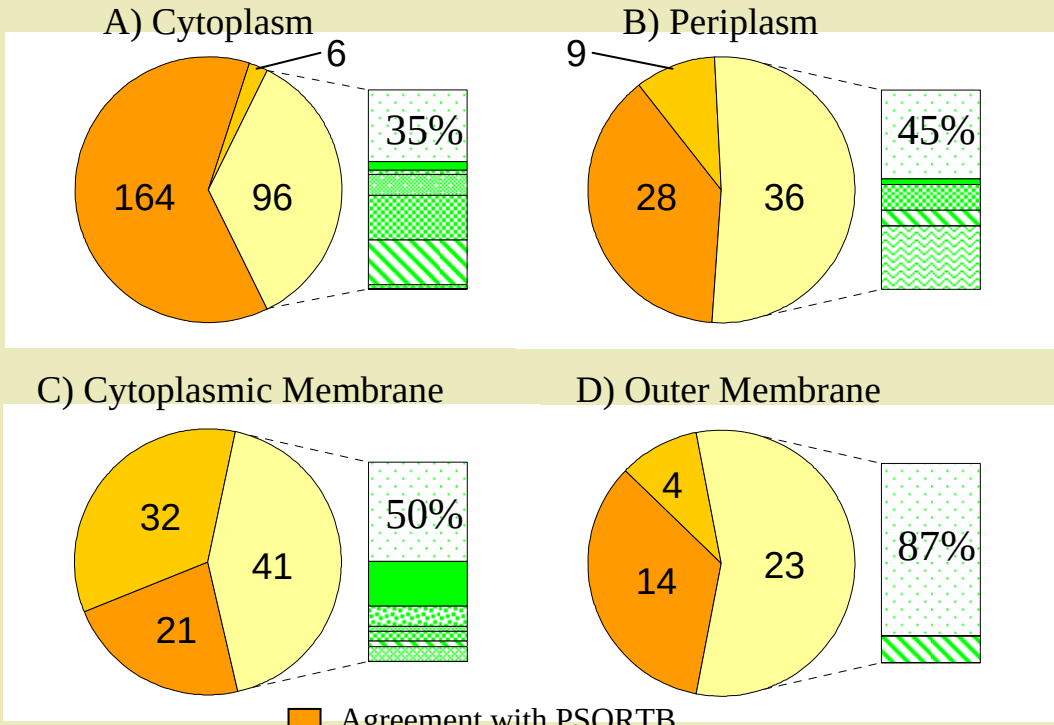
Abundance Value for Protein X

$$Zscore = \frac{Value - Xbar}{StDev}$$

- **Zscore = significance of each abundance value**
- **Value = the individual rolled-up protein abundance**
- **Xbar = average of a protein across all conditions and measurements**
- **StDev =standard deviation of a protein across all measurements**

Localization of *Rhodobacter* Proteins





Microbial Community Complexity

Ability to characterize one organism in a simple community

			Organism-specific Peptides			
Organism	Proteins	Total in silico Peptides	5 ppm & 5% NET Percent Unique	**5 ppm & 1% NET* Unique Peptides**	**5 ppm & 1% NET* Percent Unique**	1 ppm & 1% NET Percent Unique
D. radiodurans	3117	207,863	6%	**44,728**	**22%**	36%
G. metallireducens	4767	309,911	5%	**61,238**	**20%**	33%
G. sulfurreducens	3449	284,034	5%	**53,094**	**19%**	32%
S. oneidensis	4798	299,601	6%	**64,538**	**22%**	36%
Total	**16,131**	**666,861**	**9%**	**223,598**	**34%**	**57%**

TABLE 1 . Protein and peptide counts for a simple microbial community of 4 organisms of importance to bioremediation projects. The values indicate the number of peptides that are distinctive to a single organism within a mixture of the four organisms. The asterisk indicates the level used for feature matching of the MS data from the simple community to the mass tag database.

Microbial Community Complexity

Ability to characterize one organism in a complex community

Organism	Amino Acids	Proteins	In *silico* Peptides	Proteins w/ > 2 peptides
B. burgdorferi	432,403	1637	68,845	1633
C. *burnetii*	569,561	2051	84462	2046
C. *muridarumi*	*324,399*	*916*	*45,880*	*916*
C. *pneumoniae*	*365,118*	*1116*	*51,051*	*1115*
C.tepidum	*629,322*	*2241*	*92970*	*2230*
D. desulfuricans	*1,008,695*	*3007*	*137,441*	*3005*
D. radiodurans	*963,879*	*3117*	*125,640*	*3112*
E. coli	1,358,990	4289	174,328	4283
G.metallireducens	1,309,450	4767	183,448	4685
G. sulfurreducens	1,139,133	3449	165,509	3445
J. jannaschii	505,037	1780	89,926	1778
N. meningitidis	593,191	2047	82,033	2042
P. *gingivalis*	646,417	1916	95,150	1912
R. palustris	1,586,736	4833	216,946	4828
S. agalacticae	618,957	2115	89,116	2111
S. aureus	772,978	2676	109,044	2669
S. mitisnctc	584,835	2148	74,320	2136
S. oneidensis	1,422,029	4798	173,719	4711
S. pneumoniae	597,559	2098	82,352	2084
S. typhimurium	1,423,622	4550	184,652	4539
T. denticola	864,153	2756	136,903	2754
T. maritima	581,312	1842	97,215	1840
T. pallidum	350,744	1030	53,116	1029
W. pepientis	335,882	1176	52,897	1176
Y. pestis	1,320,982	4286	162,250	4267
	Total	2,829,213		

TABLE 2. Protein and peptide counts for an artificial, complex microbial community of 25 organisms, in alphabetical order by organism. Peptides are fully tryptic, with the allowance of 1 missed cleavage site.

Microbial Community Complexity

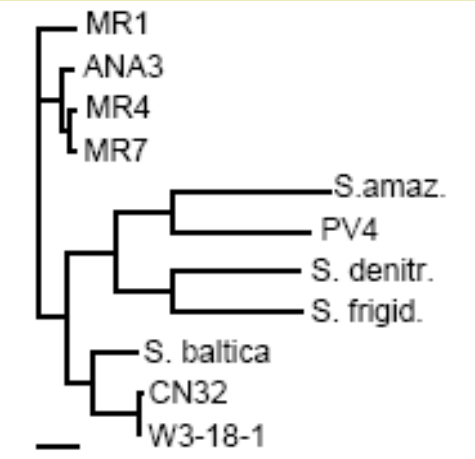
Advances in instrumentation allows more complete organism characterization

Organism	6ppm/5%NET		6ppm/1%NET		1ppm/1%NET	
B. burgdorferi	2%	22%	9%	78%	30%	97%
C. burnetii	2%	17%	9%	72%	33%	96%
C. muridarumi	2%	22%	9%	80%	33%	98%
C. pneumoniae	2%	19%	10%	78%	34%	97%
C.tepidum	2%	16%	9%	71%	33%	96%
D. desulfuricans	2%	21%	10%	79%	35%	99%
D. radiodurans	2%	19%	10%	76%	34%	99%
E. coli	1%	17%	8%	75%	27%	98%
G.metallireducens	2%	19%	9%	68%	32%	93%
G. sulfurreducens	2%	22%	9%	77%	31%	98%
J. jannaschii	2%	21%	8%	79%	30%	99%
N. meningitidis	2%	15%	9%	73%	33%	97%
P. gingivalis	2%	22%	9%	78%	33%	97%
R. palustris	2%	18%	9%	76%	33%	98%
S. agalacticae	2%	18%	9%	74%	31%	97%
S. aureus	2%	19%	10%	70%	33%	97%
S. mitisnctc	1%	16%	6%	69%	20%	95%
S. oneidensis	2%	16%	9%	70%	35%	96%
S. pneumoniae	1%	16%	6%	71%	21%	96%
S. typhimurium	1%	16%	8%	75%	28%	98%
T. denticola	2%	23%	9%	80%	33%	98%
T. maritima	2%	23%	9%	82%	31%	99%
T. pallidum	2%	21%	9%	79%	33%	98%
W. pepientis	2%	18%	9%	73%	32%	97%
Y. pestis	2%	16%	9%	75%	33%	97%

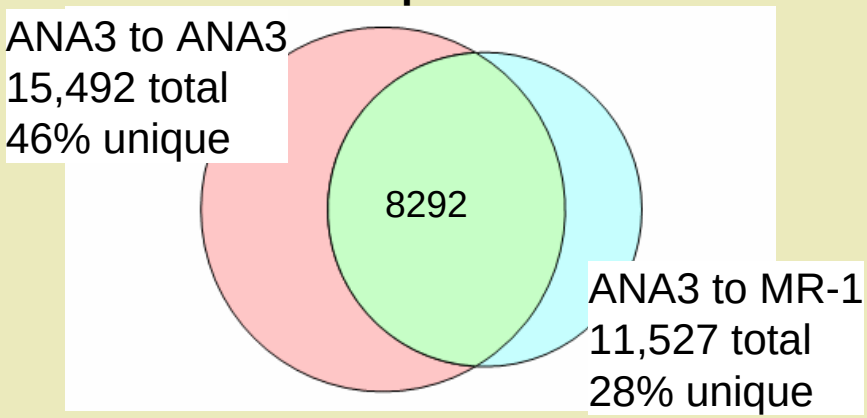
% unique
Peptide
Protein ≥ peptides

Table 3. Percentage of uniquely-identifiable peptides within an artificial complex community, calculated at two mass and NET tolerance levels.

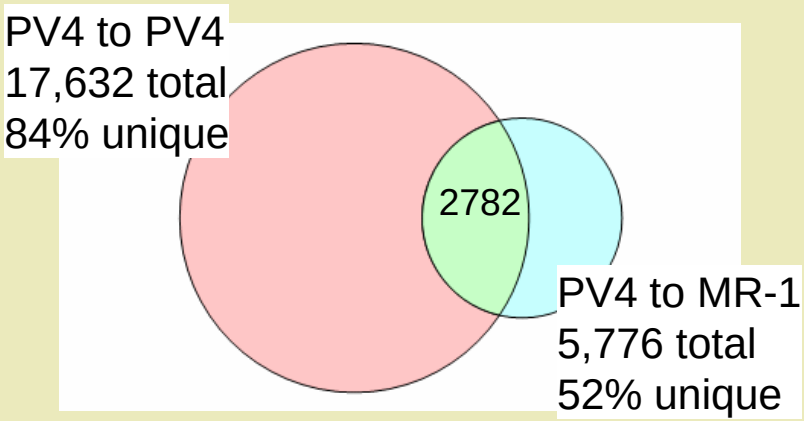
Using the Genome of a Near Neighbor to Characterized the Unsequenced Organism



Peptides



Peptides



Strain	MR1	ANA3	PV4
ANA3	1202	1577	----
PV4	792	----	1599

Further considerations

Speciation in a given environment

In silico predictions of proteomic characterization were assuming equal distribution of microbes. What happens with real world communities with a dynamic range of microbes

Quantitiative tools provide insight for microbial activities

- **Proteomics will tell you what proteins are there**
- **Can't tell you if the proteins but not are active**
- **Need to combine with metabolomics**

Acknowledgements

Pacific Northwest National Laboratory:

Proteomics

Ms.Kim K. Hixson

Dr. Stephen Callister

Sam Purvine Dr.

Angela Norbeck

Dr Richard Smith

Dr. Joshua Adkins

Dr. Matthew Monroe

Ron Moore

Shewanella Biology

Dr. Jim Fredrickson

Dr. Margie Romine

Dr. Alex Beliaev

Dr. Yuri Gorby

Dr. Dwayne Elias

Proteomics Infrastructure Support

Instrumentation Development Team

Informatics Development Team

Instrumentation Operations Support

Rhodobacter Biology

Dr. Sam Kaplan (University of Houston Medical Branch)

Dr. Tim Donohue (University of Wisconsin Madison)

DOE- Office of Biological and Environmental Research