

Abstract

E. Kolker, BIATECH

“Interdisciplinary Study of *Shewanella oneidensis* MR-1’s Metabolism & Metal Reduction”

Since our project became part of the *Shewanella Federation*, we focused our work mostly on analysis of different types of data produced by global high-throughput technologies to characterize gene and protein expression as well as getting a better understanding of the cellular metabolism. Specifically, first year activities include development of:

- new labeling technique for quantitative proteomics, so called methyl esterification labeling approach, complementary to currently available methods;
- new algorithm for *de novo* protein sequencing;
- new statistical model for spectral analysis of arbitrary shape data;
- one of the **first** analyses of the transcriptome of the entire microorganism;
- new approach to predict operon structures and transcripts within untranslated regions;
- the **first** control protein experimental mixtures with known physico-chemical characteristics for high-throughput proteomics experiments;
- the **first** statistical models for peptide and protein identifications for high-throughput proteomics analysis;
- *Shewanella* metabolic capability experiments with minimal media on aerobically & anaerobically grown cells and transformation experiments.

Several collaborations have been established within the *Shewanella Federation* with **PNNL**, **USC**, **ORNL**, and **MSU**. The first year of this project, supported by DOE’s Offices of Biological and Environmental Research and Advanced Scientific Computing Research, also resulted in 6 published papers.

This is a joint work of **A. Keller**, **A. Nesvizhskii**, **A. Picone**, **B. Tjaden**, **D. Goodlett**, **S. Purvine**, **S. Stolyar**, and **T. Cherny** done at **BIATECH** and **ISB**.

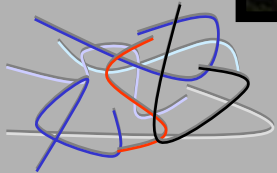
Eugene Kolker, PhD
President & Director
BIATECH
nonprofit research center

Editor-in-Chief
*OMICS A Journal of
Integrative Biology*

BIATECH (Kolker et al.)

- Sequence and data analysis
- Statistical models
- Quality assessments for HT analyses

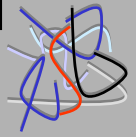
Sample



Proteins

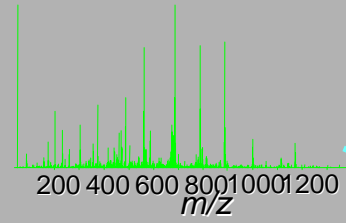
trypsin

Step 1

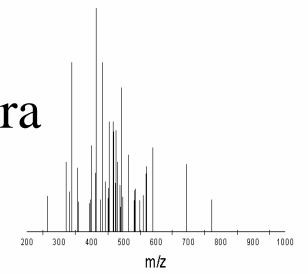


Peptides

Step 2



Bad Spectra



Step 3

Good Spectra

Step 4

Step 5

MS/MS Database Search Results

Step 6

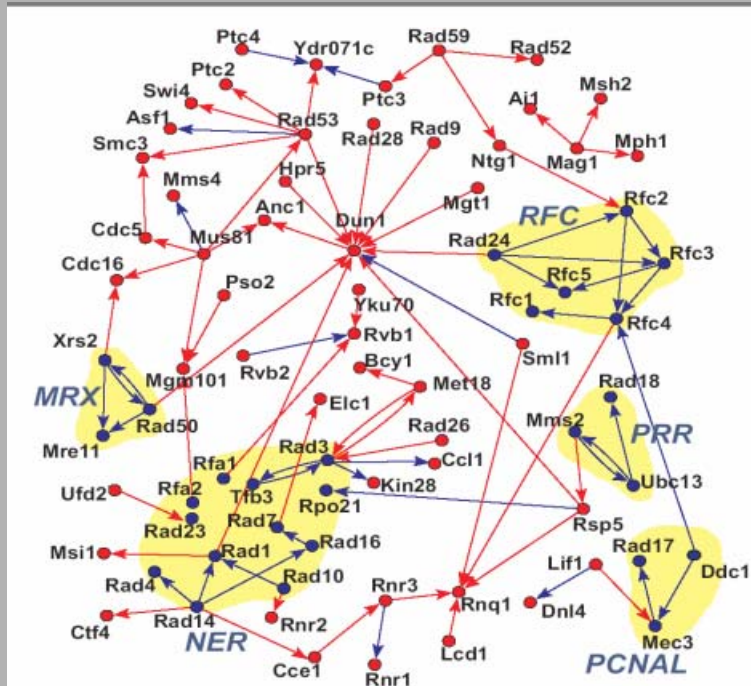
High Confidence Protein Identifications

Peptide	Prob
Peptide 1	0.999
Peptide 2	0.500
Peptide 3	0.750
Peptide 4	0.001

Peptide Probabilities

BIATECH

Protein-Protein Interaction Maps



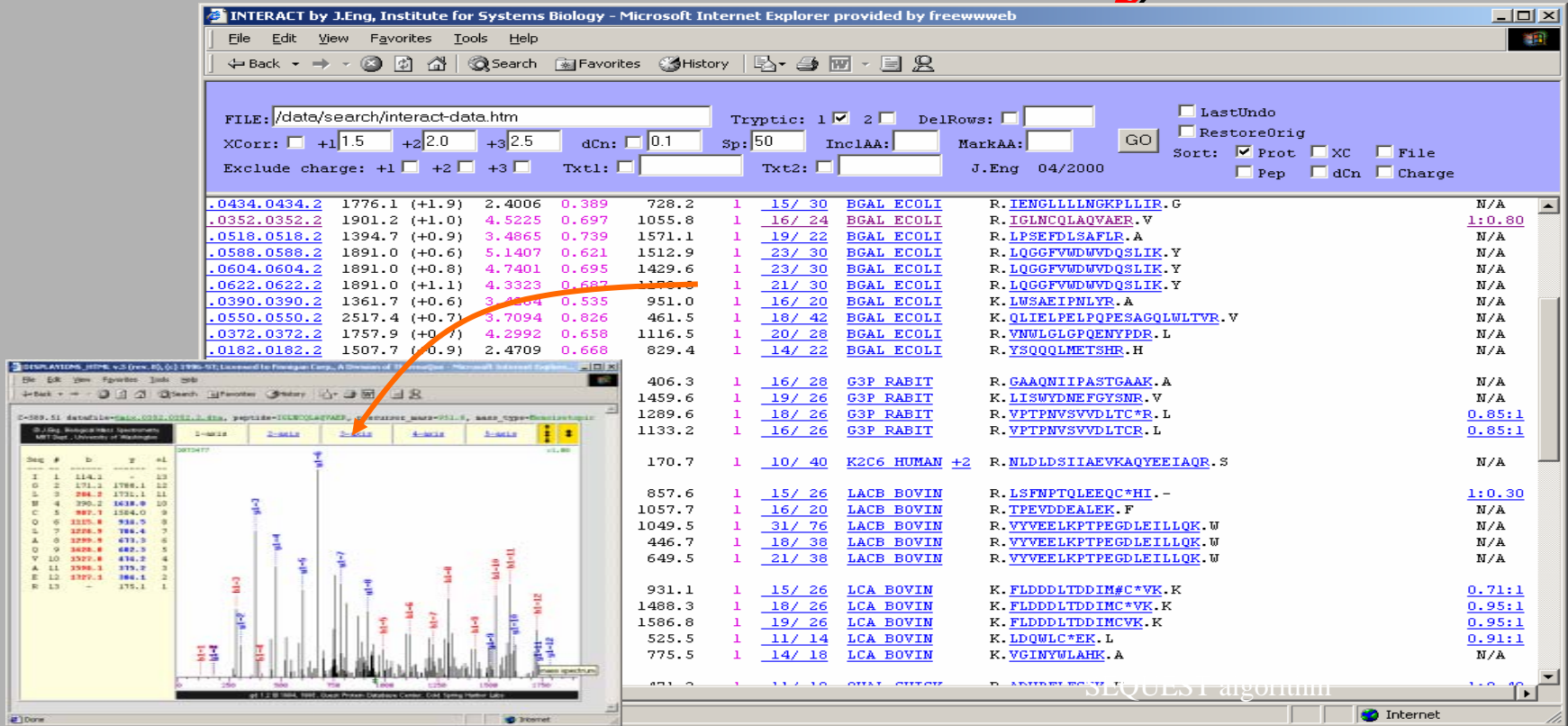
high-throughput mass
spectrometric protein
complex identification
approach

940,000 MS/MS spectra
35,000 peptide identifications
8,118 potential interactions

Y. Ho *et al.*, Nature 415, 180 (2002), “Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry”

No attempt to estimate confidence levels of protein identifications

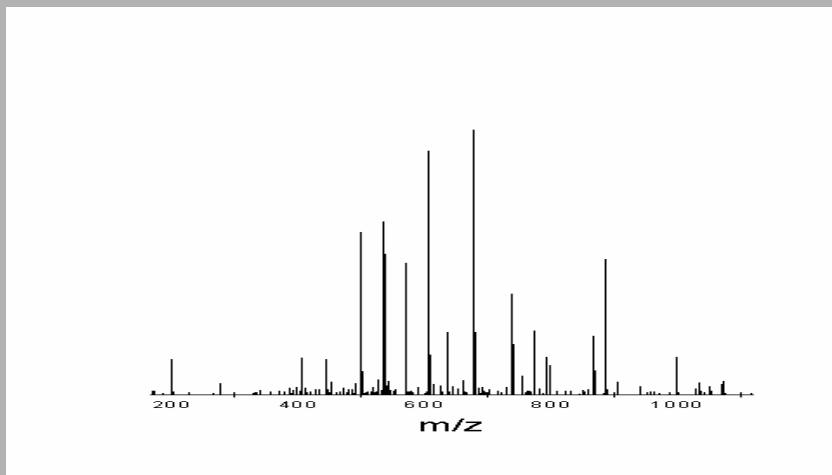
MS/MS Data Analysis



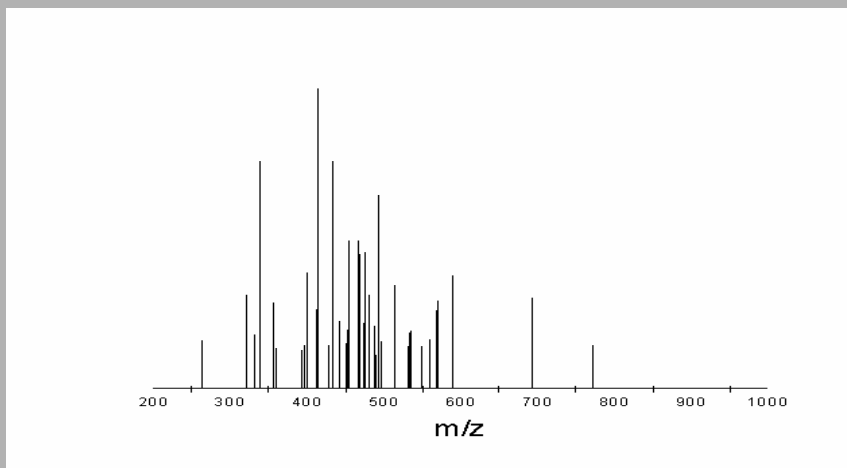
- Thousands of spectra from each experiment, but much of the data are of low quality
- Correct peptide identification or false positive? Requires decision from a human analyst

Quality Assessment of MS is needed

Quality Assessment of MS/MS Spectra

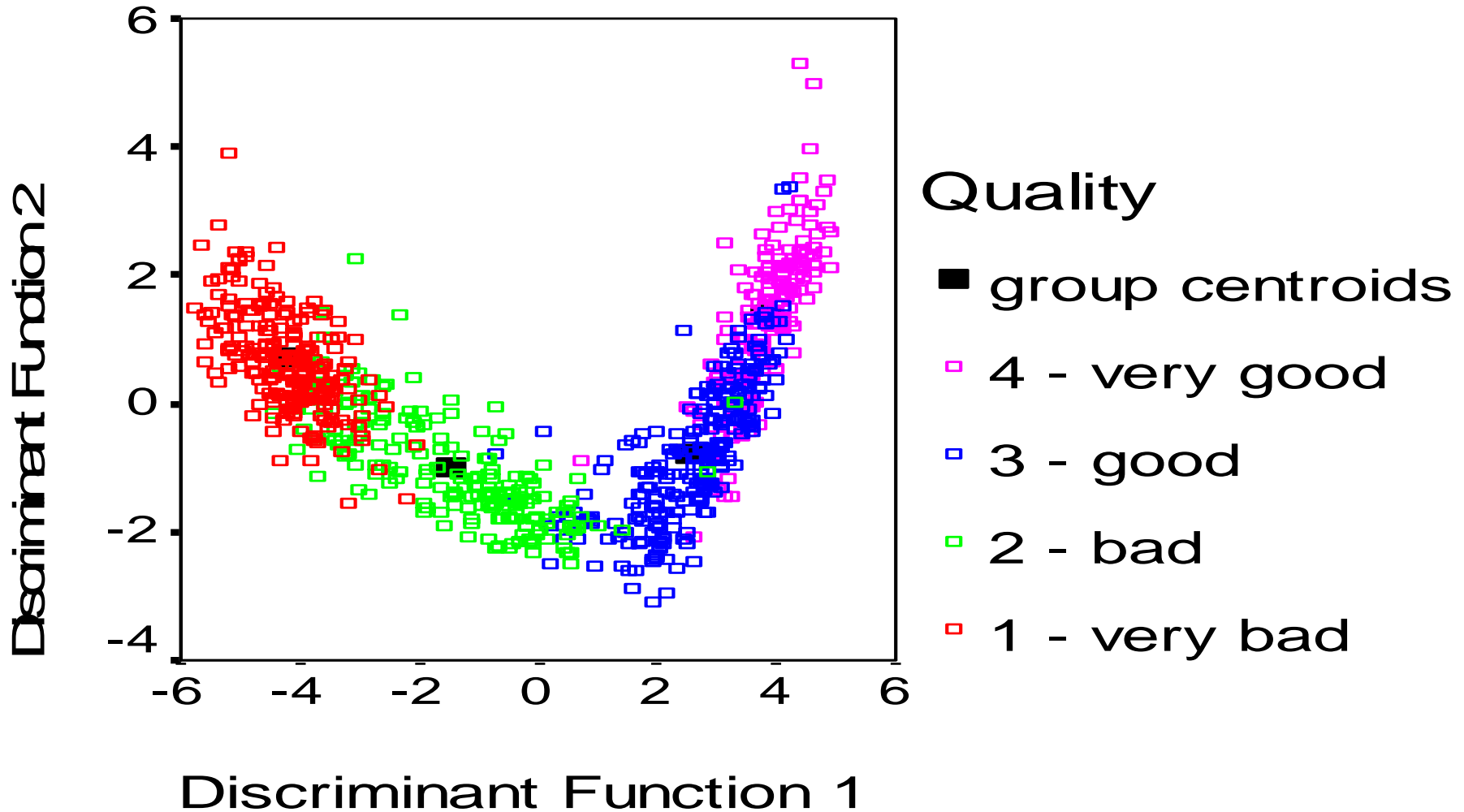


Good Spectrum



Bad Spectrum

Spectrum Quality Clustering



Training set (manually assigned quality): ~ 1,000 spectra,
HI, LC/MS/MS

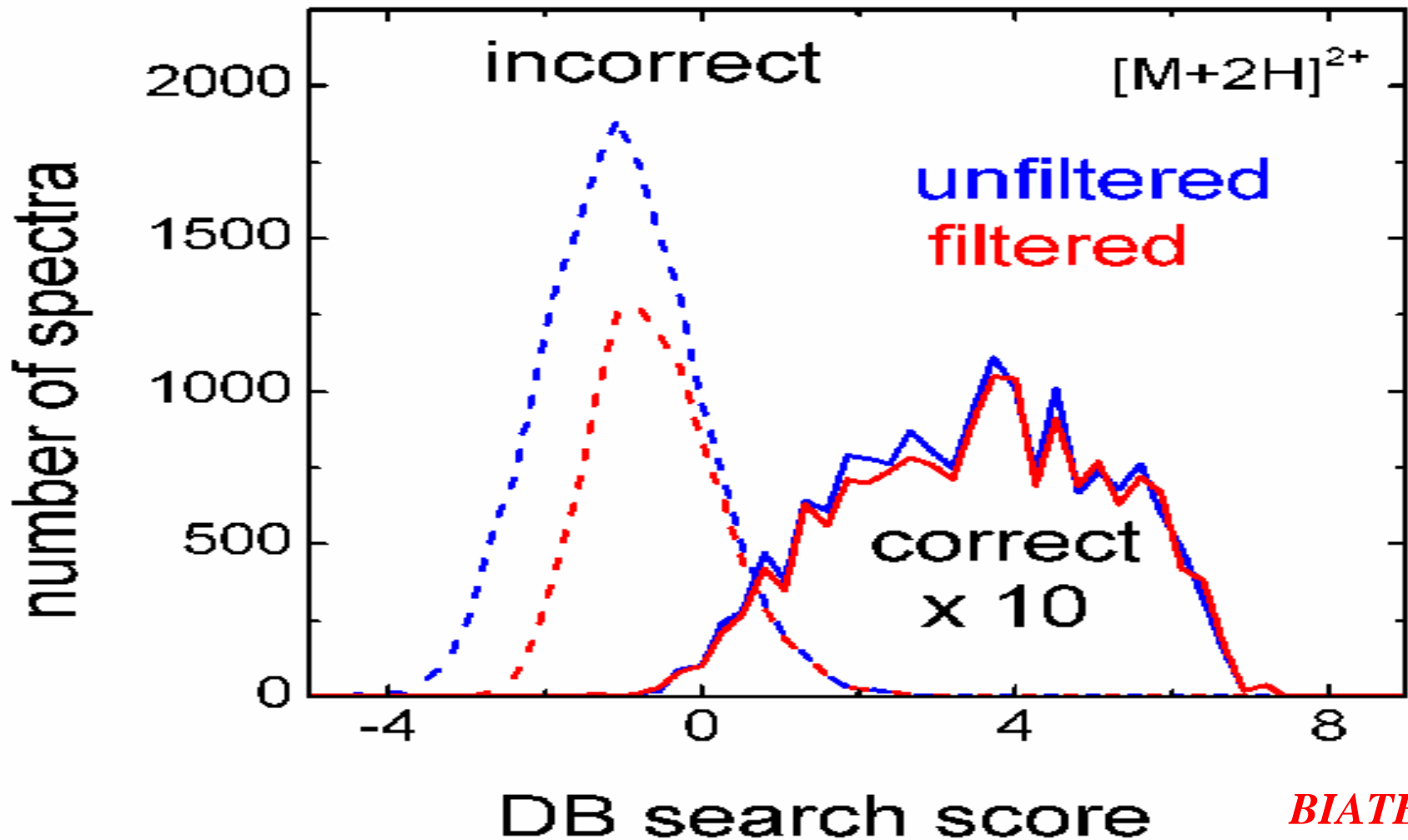
How to Mimic Complex Samples & Develop Statistical Models of Peptide & Protein Identifications?

- **20** selected, purified proteins
- Different concentrations
- **1,000:1** dynamic range
- Different database searches
- *To Build Statistical Models*

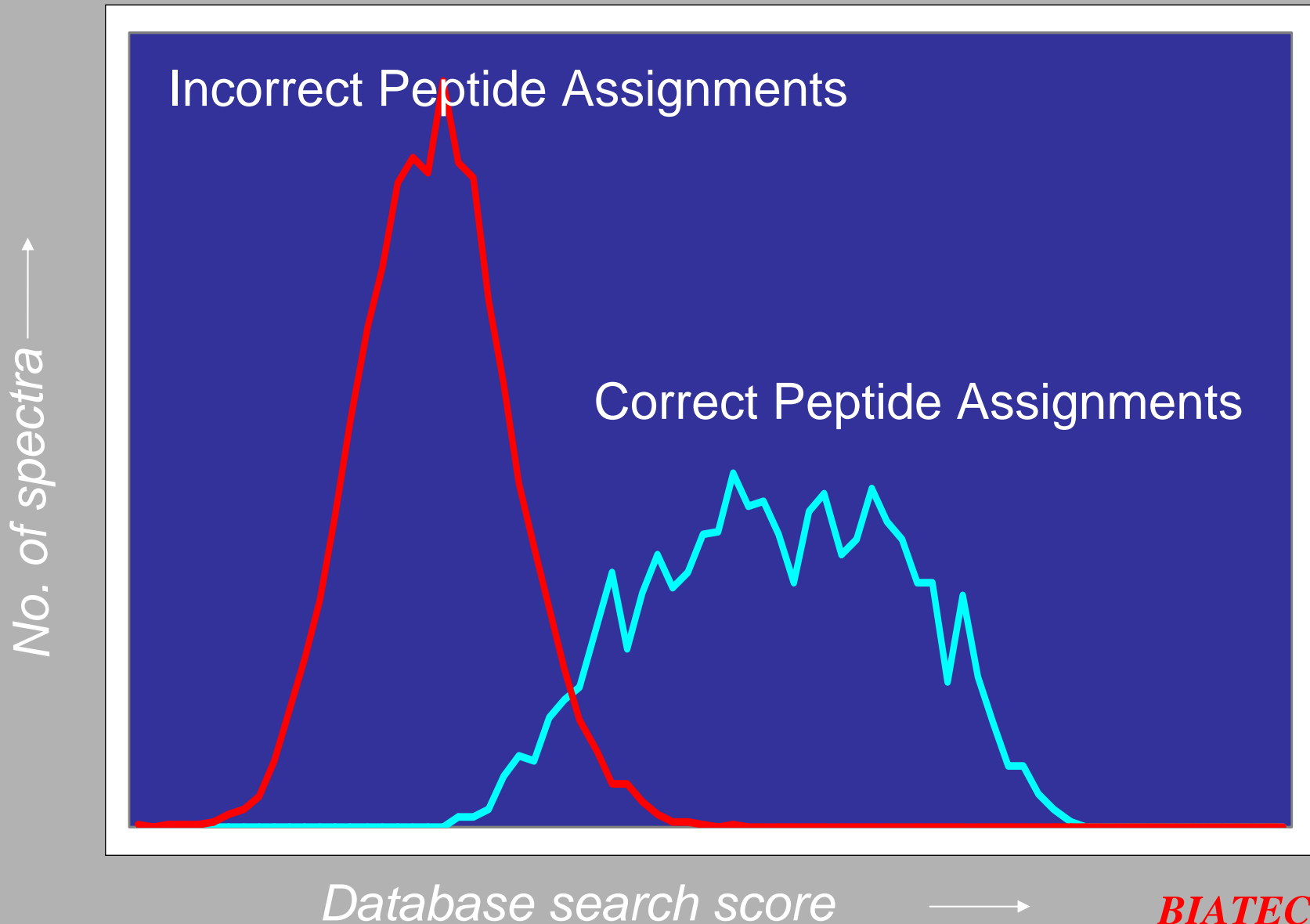
Control Protein Mixture

	Protein Name	MW(Daltons)	Conc.(nM)	CL
1.	Bovine beta-casein	25,107	100	100
2.	Bovine carbonic anhydrase	28,980	100	100
3.	Bovine cytochrome c	11,572	40	63
4.	Bovine beta-lactoglobulin	19,883	20	100
5.	Bovine alpha-lactalbumin	16,246	10	0
6.	Bovine serum albumin	69,293	40	100
7.	Chicken ovalbumin	42,750	0.4	0
8.	Bovine serotransferrin	77,753	10	100
9.	Rabbit GAPDH	35,688	2	100
10.	Rabbit glycogen phosphorylase	97,158	1	100
11.	EC beta-galactosidase	116,351	0.4	18
12.	Bovine gamma-actin	41,661	0.2	0
13.	Bovine catalase	57,585	2	100
14.	Rabbit myosin	241,852	0.2	0
15.	EC alkaline phosphatase	49,438	20	100
16.	Horse myoglobin	16,951	4	0
17.	B. lich. alpha amylase	66,924	4	0
18.	S. cer. mannose-6-phosphate isomerase	48,057	1	0

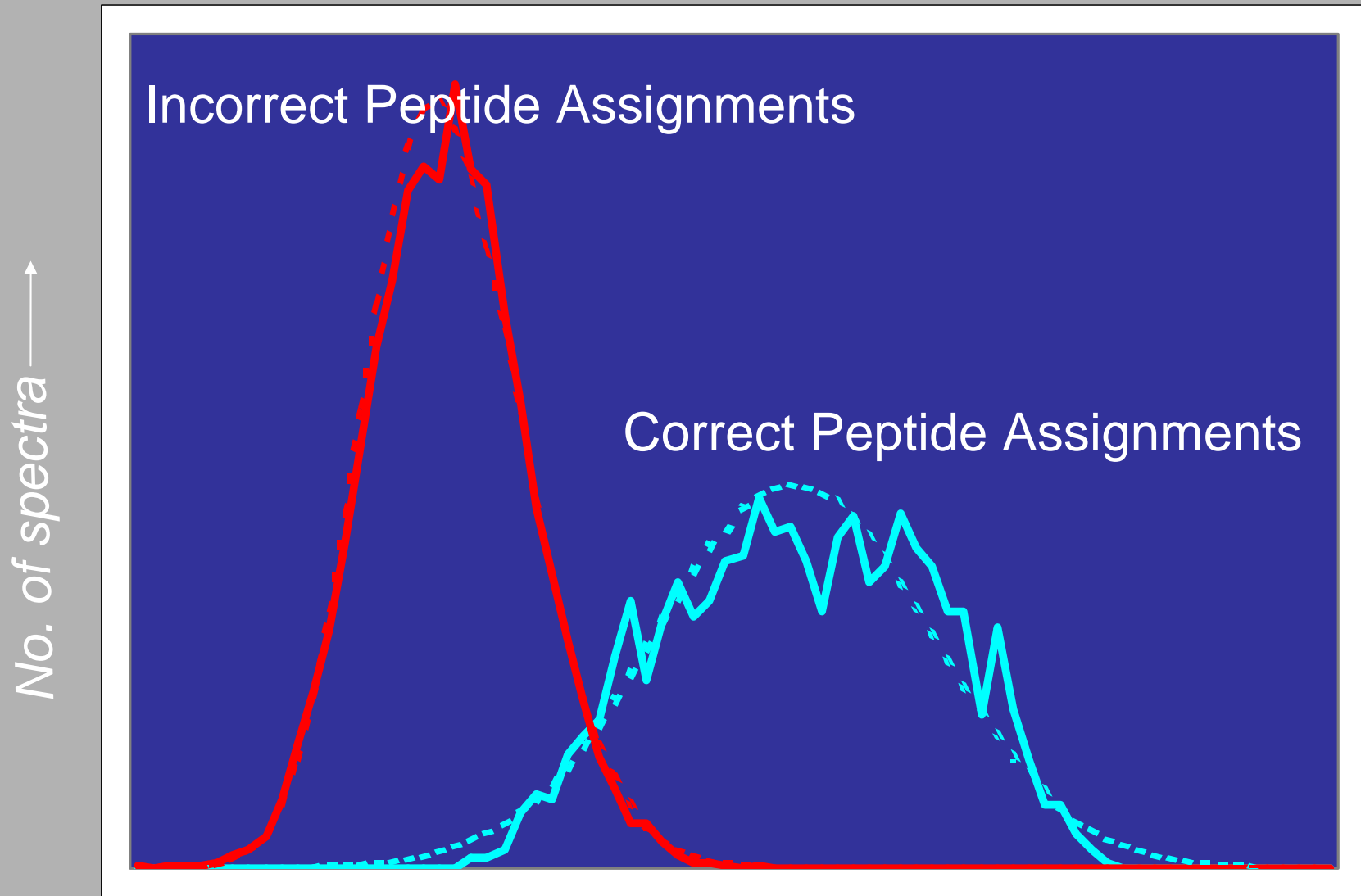
Filtered and Unfiltered Distributions



EM Learns Search Score Distributions

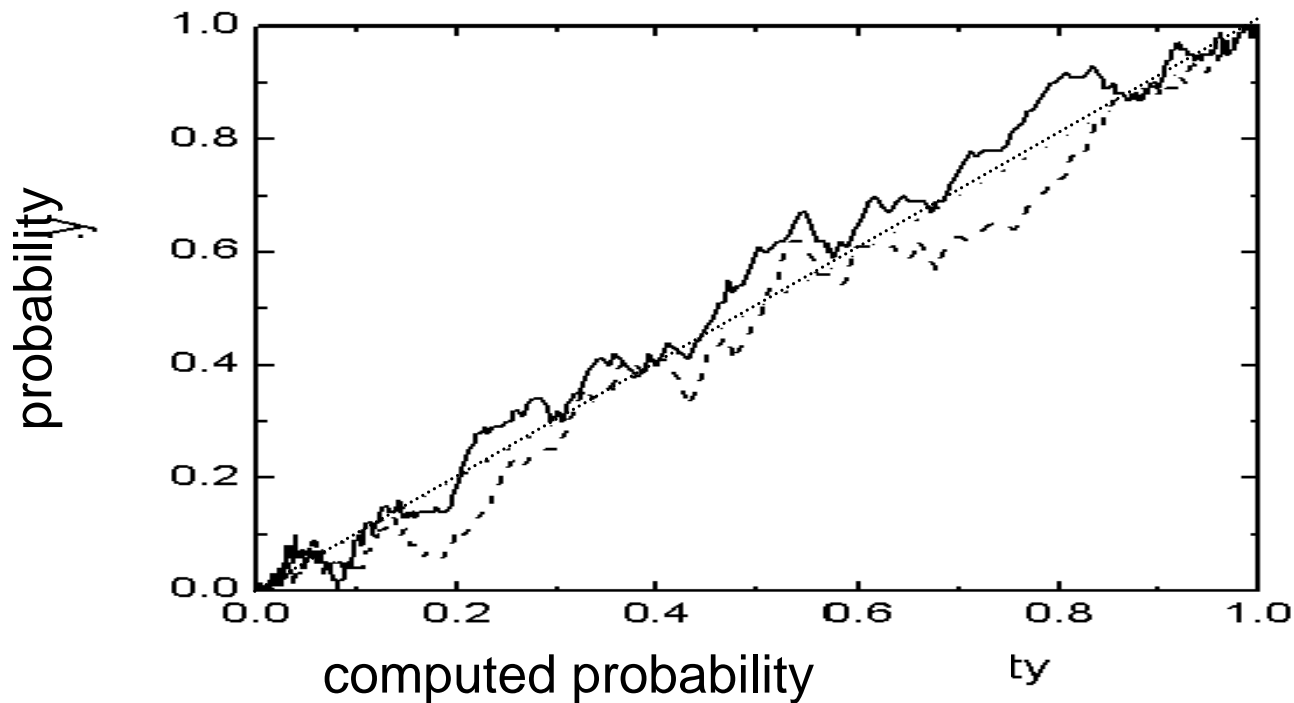


EM Iteration 8



Database search score →

Accuracy of Probability Model: Test Set

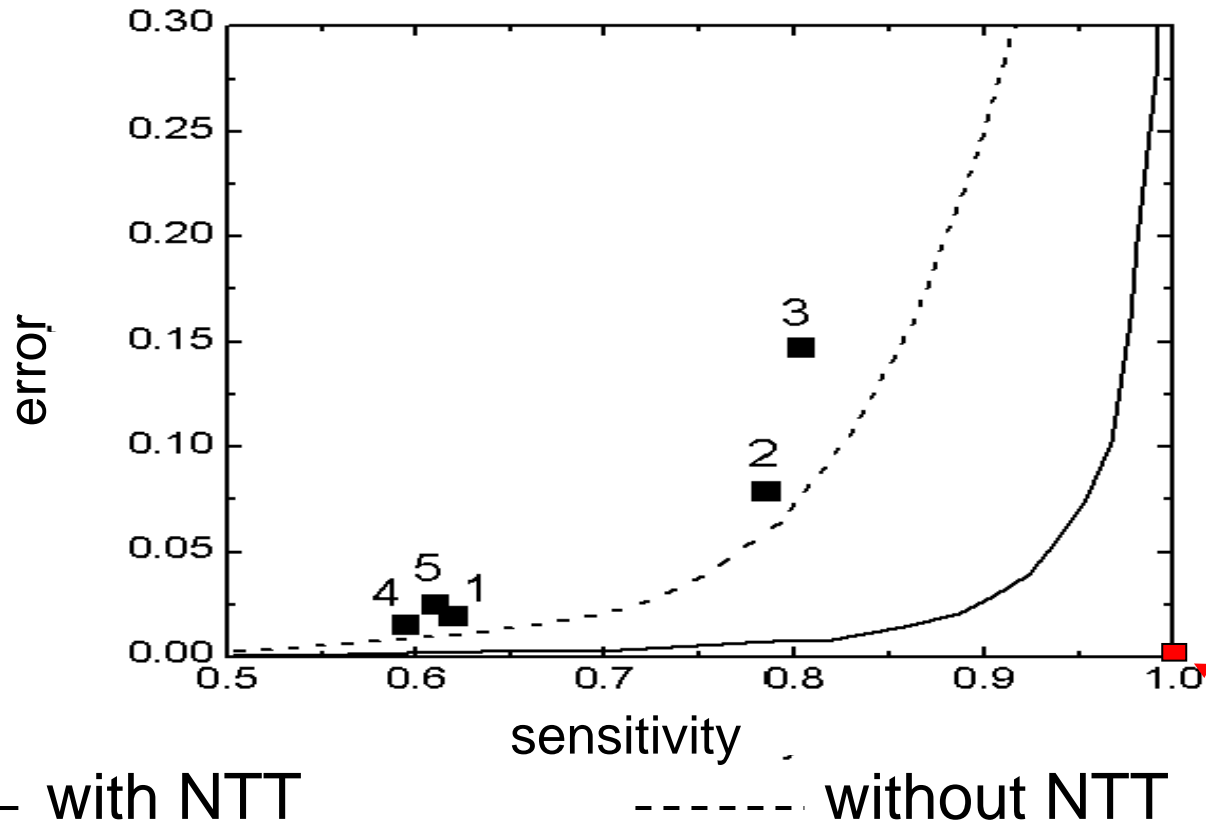


———— Combined test data

----- Ind. data sets

Test data: SEQUEST results of known validity; ~36,000 MS/MS spectra generated from the control protein mixture (22 LC/MS/MS runs)

Discriminating Power of Computed Probabilities: Test Data Set



—— with NTT

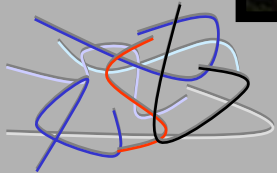
----- without NTT

Sensitivity: fraction of all correct results passing filter.

Error: fraction of all results passing filter that are incorrect

Ideal Spot

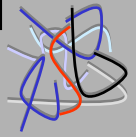
Sample



Proteins

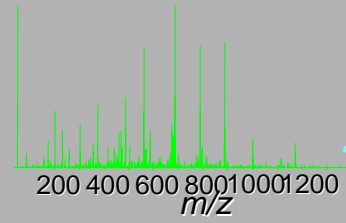
trypsin

Step 1

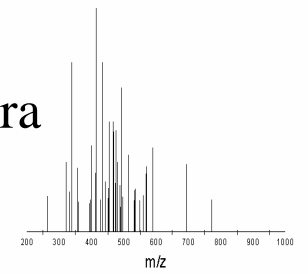


Peptides

Step 2



Bad Spectra



Step 3

Good Spectra

Step 4

Step 5

MS/MS Database Search Results

Step 6

High Confidence Protein Identifications

Peptide	Prob
Peptide 1	0.999
Peptide 2	0.500
Peptide 3	0.750
Peptide 4	0.001

Peptide Probabilities

BIATECH

Future Needs - SF

- **Data Integration**
- **Modeling**
- **WGA**

- **Sequencing of *Shewanella* strains**
- **Controls, standards, and quality assessments for sample preps, HT and data analyses**
- **GtL coordination/updates**

ATTENTION:

***OMICS J Integr Biol* : Integrative Microbiology, 2003 (GtL) issue**
and *ASM, May 2003, DC* : Systems Microbiology

Shewanella oneidensis

S. oneidensis is the abbreviated name of the bacterium *Shewanella oneidensis*, which according to the definitive text, which categorizes bacteria [Bergey's Manual](#), belongs to the gram negative gamma-subgroup (as [E. coli](#) and [H. influenzae](#)) Alteromonadales, genus XII *Shewanella*.

The name *oneidensis* comes from the name of *Lake Oneida* where from our collaborator and friend Ken Nealson first isolated and characterized *S. oneidensis* fifteen years ago. *S. oneidensis* is at the very top of the priority list of the [US Department of Energy](#), because of its unique ability to reduce heavy metals like uranium, degrade organic wastes, and sequester a range of toxic metals.

Environments in places like Hanford or Chernobyl can be significantly improved if we would understand *Shewanella* better.

We are still not there...

More Information on *S. oneidensis*:

- [DOE's information on *S. oneidensis*](#)
- [DOE's Genomes to Life](#)
- [Shewanella Federation Web site](#)
- [Shewanella Genome Annotation \(02/03/03\)](#)

PIs of *Shewanella* Federation:

Eugene Kolker, *BIATECH*

James Fredrickson, *Pacific Northwest National Laboratory*

James Tiedje, *Michigan State University*

Jizhong Zhou, *Oak Ridge National Laboratory*

Kenneath Nealson, *University of Southern California*

