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**.

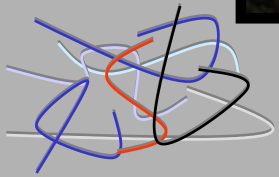
Eugene Kolker, PhD
President & Director
BIATECH
nonprofit research center

Editor-in-Chief
OMICS A Journal of
Integrative Biology

BIATECH (Kolker et al.)

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

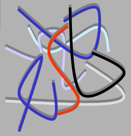
Sample



Proteins

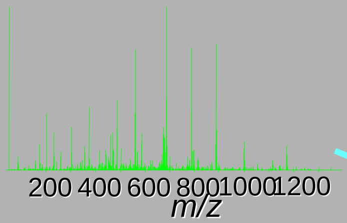
trypsin

Step **1**



Peptides

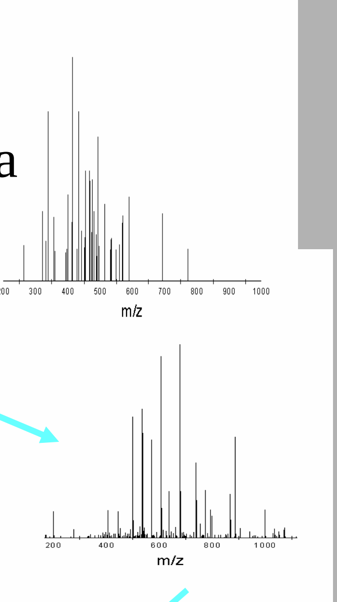
Step **2**



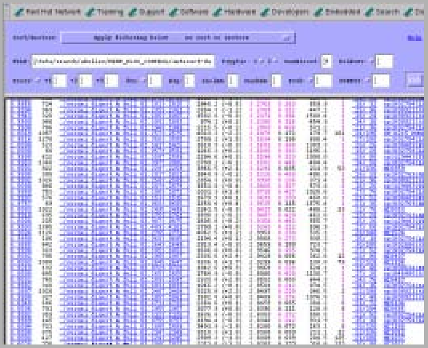
Bad Spectra

Step **3**

Good Spectra



Step **4**



Step **5**

MS/MS Database Search Results

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

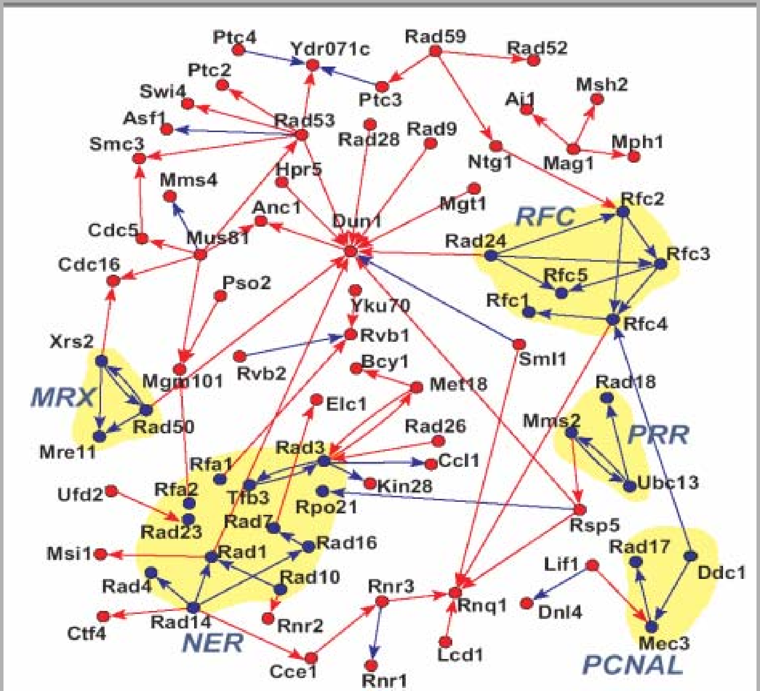
Peptide Probabilities

Step **6**

High Confidence Protein Identifications

BIATECH

Protein-Protein Interaction Maps



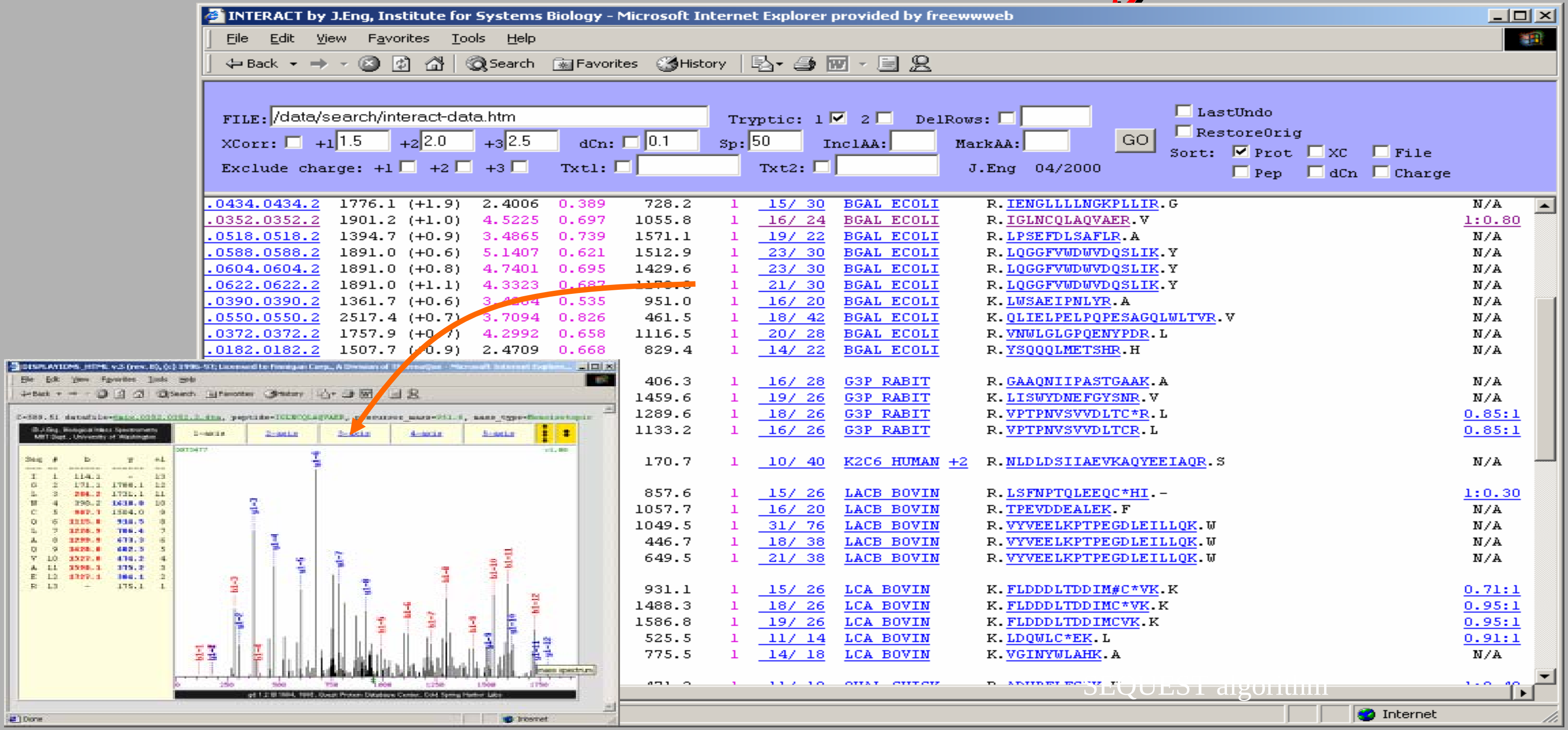
high-throughput mass spectrometric protein complex identification approach

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


INTERACT by J.Eng, Institute for Systems Biology - Microsoft Internet Explorer provided by freewwweb
File Edit View Favorites Tools Help
Back Search Favorites History
GO
Internet

FILE: /data/search/interact-data.htm Tryptic: 1 ☑ 2 □ DelRows: □ □ LastUndo □ RestoreOrig

XCorr: □ +1 1.5 +2 2.0 +3 2.5 dCn: □ 0.1 Sp: 50 InclAA: MarkAA: Sort: ☑ Prot □ XC □ File

Exclude charge: +1 □ +2 □ +3 □ Txt1: □ Txt2: □ J.Eng 04/2000 □ Pep □ dCn □ Charge

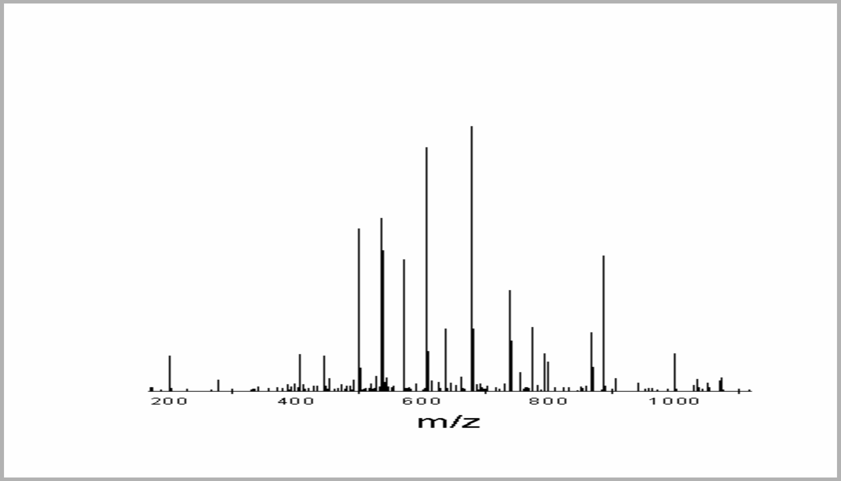
```
.0434.0434.2  1776.1 (+1.9)  2.4006  0.389   728.2   1   15/ 30  BGAL ECOLI      R.IENGLLLLNGKPLLIR.G        N/A
.0352.0352.2  1901.2 (+1.0)  4.5225  0.697  1055.8   1   16/ 24  BGAL ECOLI      R.IGLNCQLAQVAER.V           1:0.80
.0518.0518.2  1394.7 (+0.9)  3.4865  0.739  1571.1   1   19/ 22  BGAL ECOLI      R.LPSEFDLSAFLR.A            N/A
.0588.0588.2  1891.0 (+0.6)  5.1407  0.621  1512.9   1   23/ 30  BGAL ECOLI      R.LQGGFVWDWVDQSLIK.Y        N/A
.0604.0604.2  1891.0 (+0.8)  4.7401  0.695  1429.6   1   23/ 30  BGAL ECOLI      R.LQGGFVWDWVDQSLIK.Y        N/A
.0622.0622.2  1891.0 (+1.1)  4.3323  0.687  [illegible] 1   21/ 30  BGAL ECOLI   R.LQGGFVWDWVDQSLIK.Y        N/A
.0390.0390.2  1361.7 (+0.6)  3.[illegible]  0.535   951.0   1   16/ 20  BGAL ECOLI      K.LWSAEIPNLYR.A             N/A
.0550.0550.2  2517.4 (+0.7)  3.7094  0.826   461.5   1   18/ 42  BGAL ECOLI      K.QLIELPELPQPESAGQLWLTVR.V  N/A
.0372.0372.2  1757.9 (+0.7)  4.2992  0.658  1116.5   1   20/ 28  BGAL ECOLI      R.VNWLGLGPQENYPDR.L         N/A
.0182.0182.2  1507.7 (+0.9)  2.4709  0.668   829.4   1   14/ 22  BGAL ECOLI      R.YSQQQLMETSHR.H            N/A

                                             406.3   1   16/ 28  G3P RABIT       R.GAAQNIIPASTGAAK.A         N/A
                                            1459.6   1   19/ 26  G3P RABIT       K.LISWYDNEFGYSNR.V          N/A
                                            1289.6   1   18/ 26  G3P RABIT       R.VPTPNVSVVDLTC*R.L         0.85:1
                                            1133.2   1   16/ 26  G3P RABIT       R.VPTPNVSVVDLTCR.L          0.85:1

                                             170.7   1   10/ 40  K2C6 HUMAN +2   R.NLDLDSIIAEVKAQYEEIAQR.S   N/A

                                             857.6   1   15/ 26  LACB BOVIN      R.LSFNPTQLEEQC*HI.-         1:0.30
                                            1057.7   1   16/ 20  LACB BOVIN      R.TPEVDDEALEK.F             N/A
                                            1049.5   1   31/ 76  LACB BOVIN      R.VYVEELKPTPEGDLEILLQK.W    N/A
                                             446.7   1   18/ 38  LACB BOVIN      R.VYVEELKPTPEGDLEILLQK.W    N/A
                                             649.5   1   21/ 38  LACB BOVIN      R.VYVEELKPTPEGDLEILLQK.W    N/A

                                             931.1   1   15/ 26  LCA BOVIN       K.FLDDDLTDDIM#C*VK.K        0.71:1
                                            1488.3   1   18/ 26  LCA BOVIN       K.FLDDDLTDDIMC*VK.K         0.95:1
                                            1586.8   1   19/ 26  LCA BOVIN       K.FLDDDLTDDIMCVK.K          0.95:1
                                             525.5   1   11/ 14  LCA BOVIN       K.LDQWLC*EK.L               0.91:1
                                             775.5   1   14/ 18  LCA BOVIN       K.VGINYWLAHK.A              N/A
```

SEQUEST algorithm

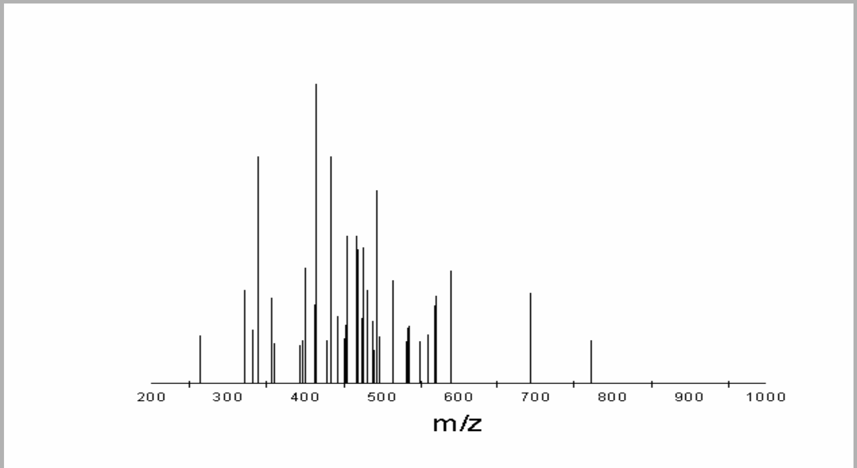
- Thousands of spectra from each experiment, but much of the data are of low quality
- Correct peptide identification or false positive? Requires decision from a human analyst

Quality Assessment of MS is needed

Quality Assessment of MS/MS Spectra

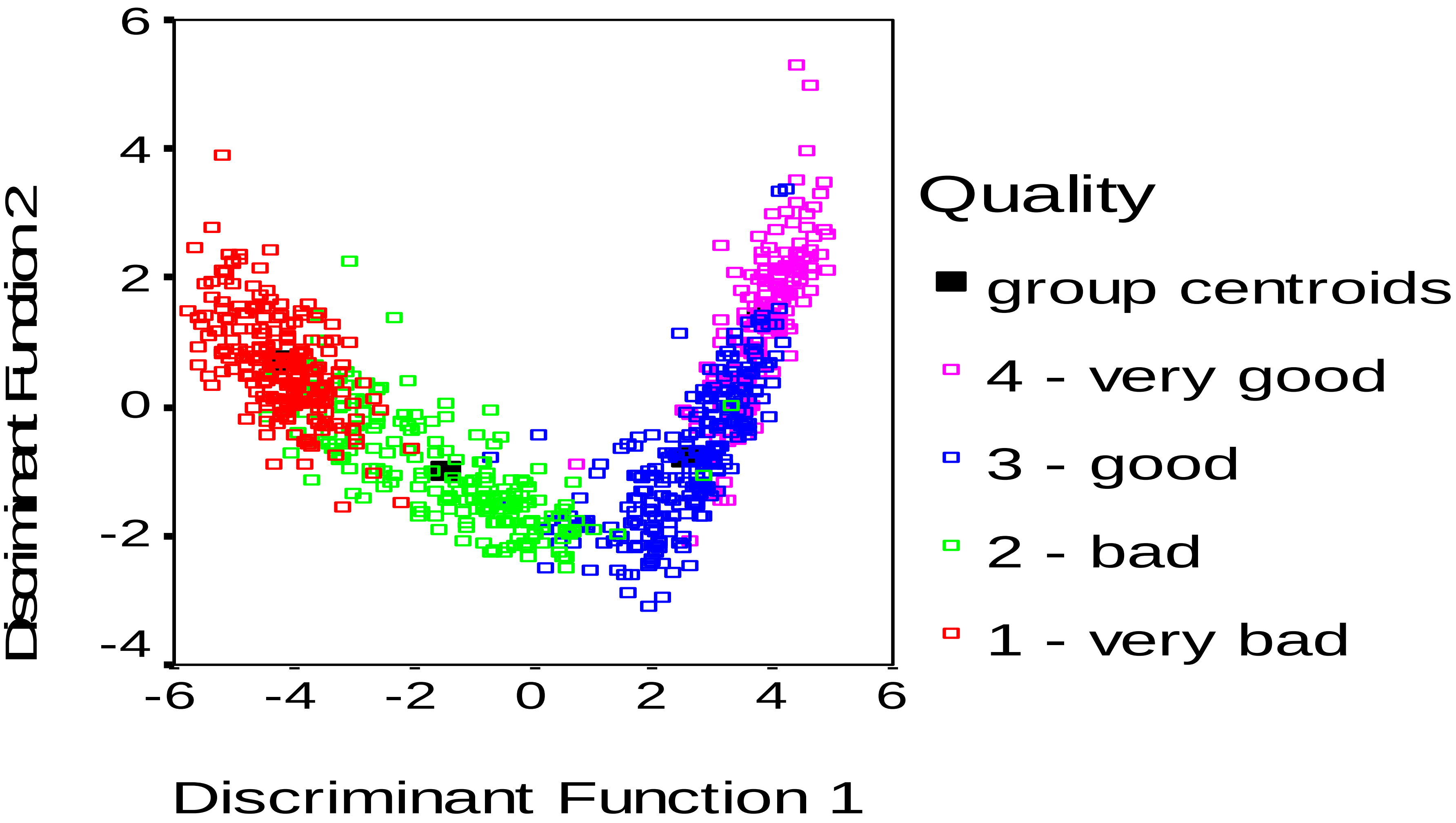


Good Spectrum



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*

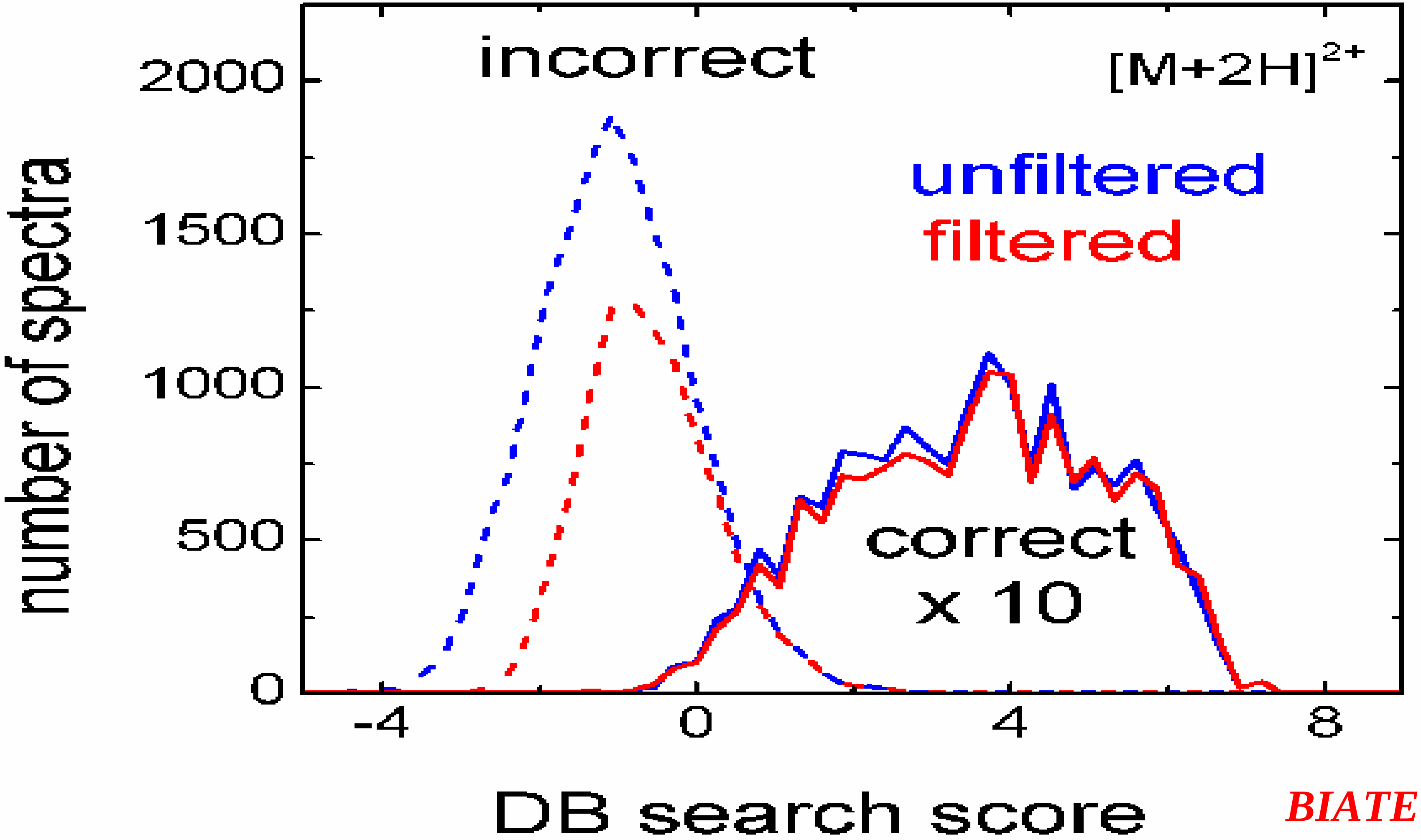
BIATECH

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

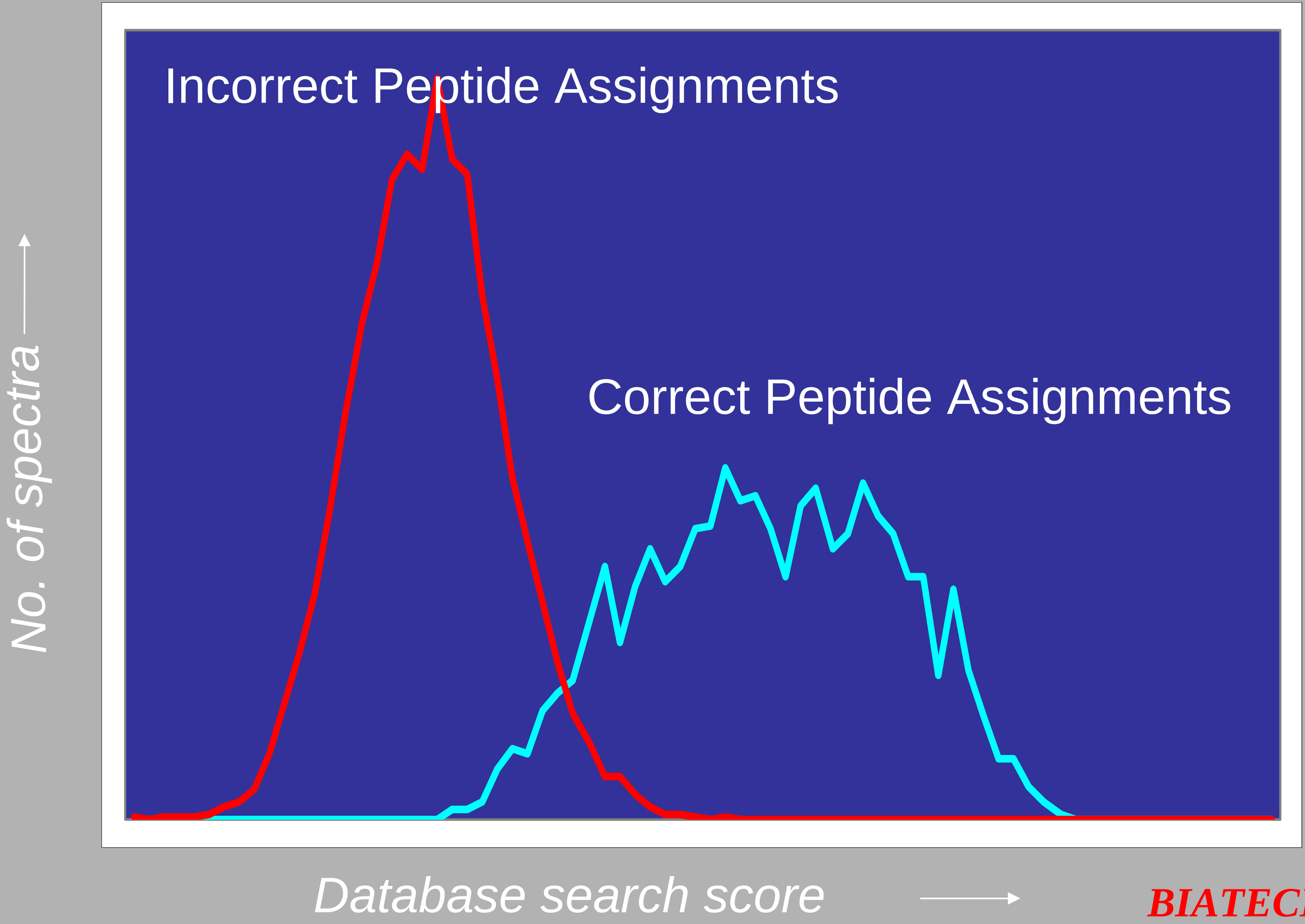
BIATECH

Filtered and Unfiltered Distributions

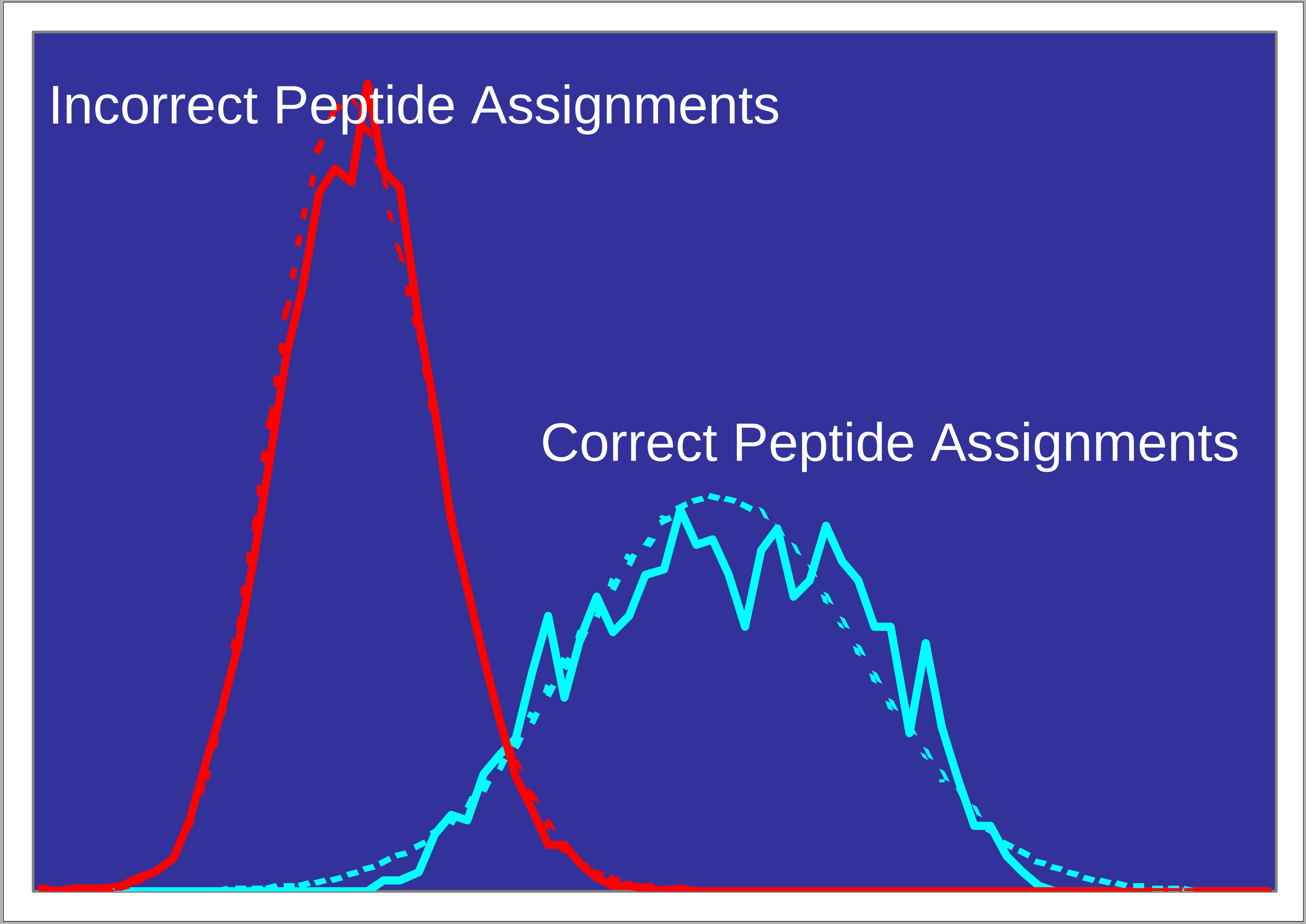


BIATECH

EM Learns Search Score Distributions



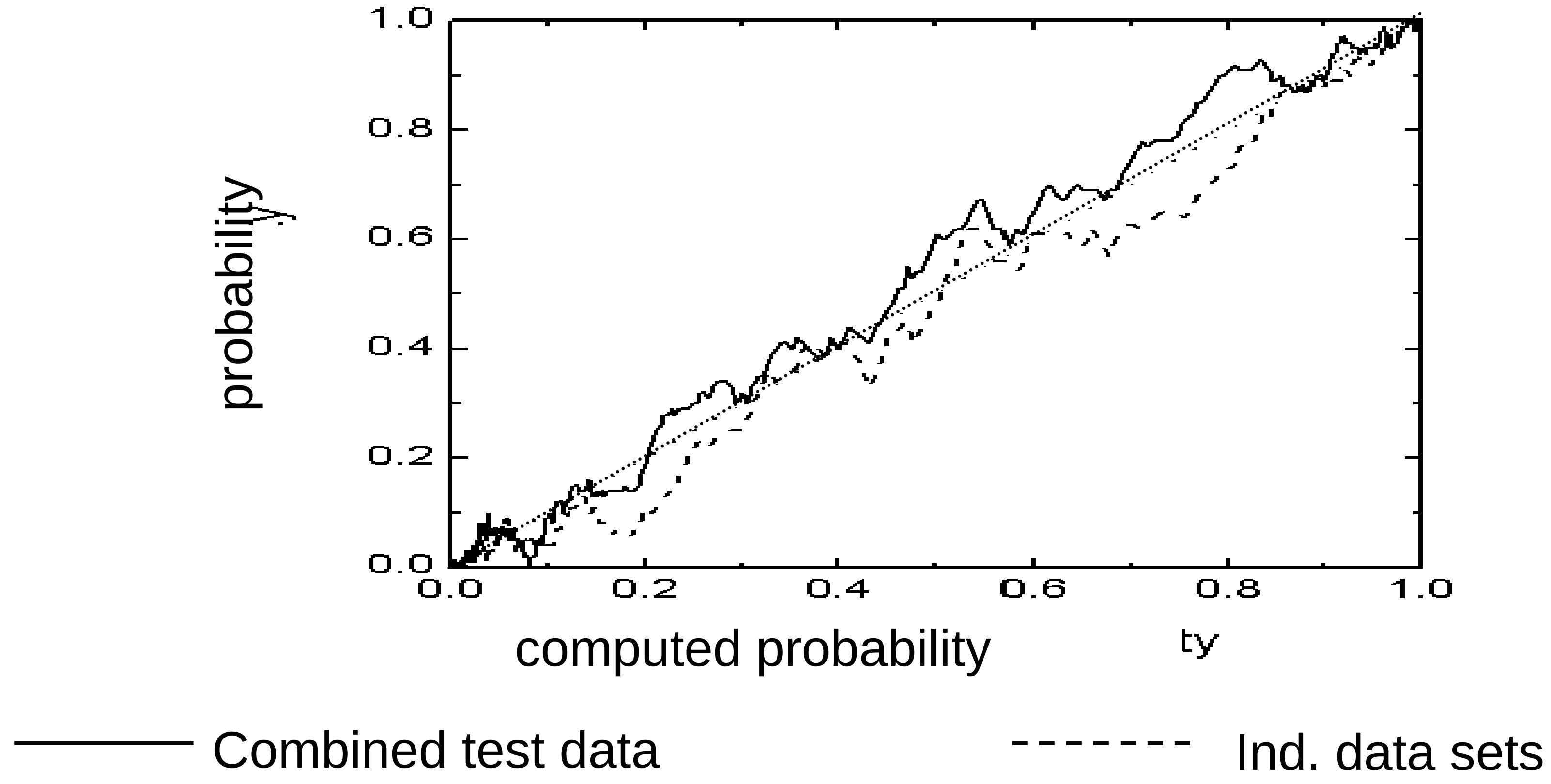
EM Iteration 8



No. of spectra

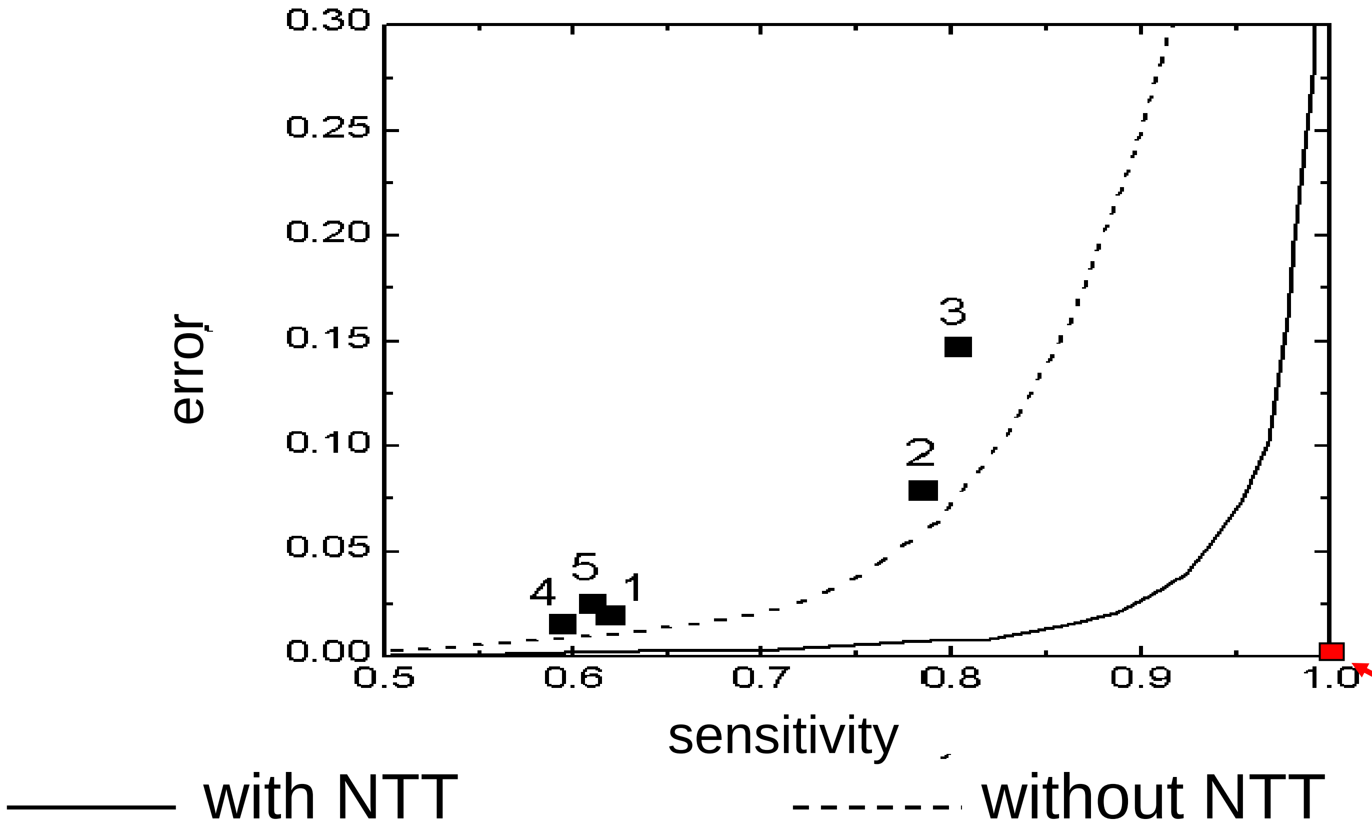
Database search score

Accuracy of Probability Model: Test Set



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

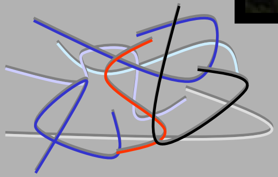


Ideal Spot

Sensitivity: fraction of all correct results passing filter.

Error: fraction of all results passing filter that are incorrect

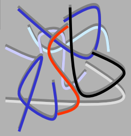
Sample



Proteins

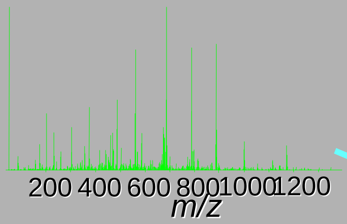
trypsin

Step **1**



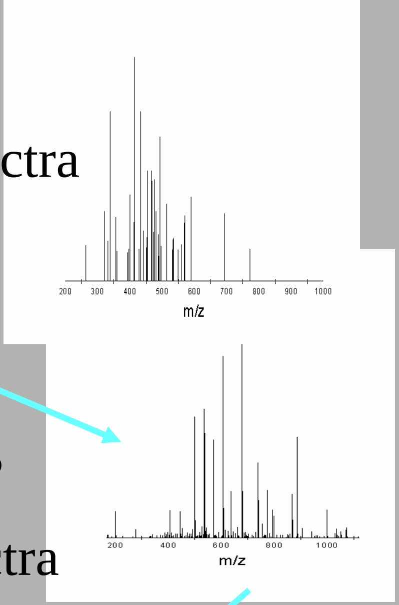
Peptides

Step **2**



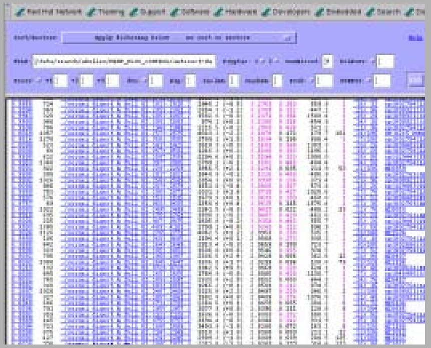
Step **3**

Bad Spectra



Good Spectra

Step **4**



Step **5**

MS/MS Database Search Results

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

Peptide Probabilities

Step **6**

High Confidence Protein Identifications

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** www.biatech.org

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***