

# PeptideProphet: Validation of Peptide Assignments to MS/MS Spectra

---

Andrew Keller

# Outline

---

- Need to validate peptide assignments to MS/MS spectra
- Statistical approach to validation
- Running PeptideProphet software
- Interpreting results of PeptideProphet
- Exercises

# Most search results are wrong

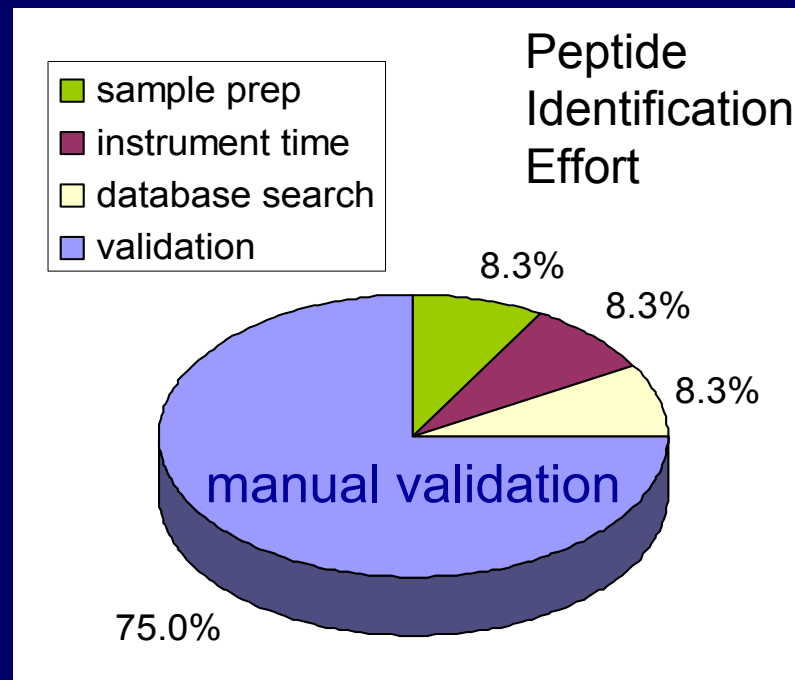
---

- $[M+2H]^{2+}/[M+3H]^{3+}$  uncertainty (LCQ)
- Non-peptide noise
- Incomplete database  
    *e.g.* post-translational modifications
- Limitation of database search algorithm

# Validation of peptide assignments

In the past, a majority of analysis time was devoted to identifying the minority of correct search results from the majority of incorrect results

Required manual judgment



# Results of 50 Spectrum Test

---

## Consistency among 'Experts':

Of 50 search results

9 had < 67% 'publishable', 'borderline', or 'not pub'

## Consistency of Individual 'Experts':

Of 10 duplicated search results, on average

0.4 were assessed 'publishable'/'not publishable'

2 were assessed inconsistently

The true validity of the search results are known.

## Accuracy of 'Experts':

of 511 total 'publishable': 95% correct

of 102 total 'borderline': 49% correct

of 387 total 'not publishable': 14% correct

*Even 'Experts' are not dependable!*

# Need for objective criteria

---

Manual scrutiny of search results is not practical for large datasets common to high throughput proteomics

As an alternative to relying on human judgment, many research groups employ search scores and properties of the assigned peptides to discriminate between correct and incorrect results

# Traditional filtering criteria

---

Each SEQUEST search result has a:

Xcorr, dCn, Sp, NTT (number of tryptic termini)

Accept all results that satisfy:

$[M+2H]^{2+}$ :  $Xcorr \geq 2$ ,  $dCn \geq 0.1$ ,  $Sp \leq 50$

$[M+3H]^{3+}$ :  $Xcorr \geq 2.5$ ,  $dCn \geq 0.1$ ,  $Sp \leq 50$

# Traditional filtering criteria

---

Each SEQUEST search result has a:

Xcorr, dCn, Sp, NTT (number of tryptic termini)

Accept all results that satisfy:

$[M+2H]^{2+}$ : Xcorr  $\geq 2$ , dCn  $\geq 0.1$ , Sp  $\leq 50$ , NTT  $\geq 1$

$[M+3H]^{3+}$ : Xcorr  $\geq 2.5$ , dCn  $\geq 0.1$ , Sp  $\leq 50$ , NTT  $\geq 1$



# Traditional filtering criteria

---

Each SEQUEST search result has a:

Xcorr, dCn, Sp, NTT (number of tryptic termini)

Accept all results that satisfy:

[M+2H]<sup>2+</sup>: Xcorr  $\geq 2$ , dCn  $\geq 0.1$ , Sp  $\leq 50$  (NTT  $\geq 1$ )

[M+3H]<sup>3+</sup>: Xcorr  $\geq 2.5$ , dCn  $\geq 0.1$ , Sp  $\leq 50$  (NTT  $\geq 1$ )

[M+2H]<sup>2+</sup>: Xcorr  $\geq 2$ , dCn  $\geq 0.1$ , Sp  $\leq 50$

[M+3H]<sup>3+</sup>: Xcorr  $\geq 2$ , dCn  $\geq 0.1$ , Sp  $\leq 50$

# Traditional filtering criteria

---

Each SEQUEST search result has a:

Xcorr, dCn, Sp, NTT (number of tryptic termini)

Accept all results that satisfy:

[M+2H]<sup>2+</sup>: Xcorr  $\geq$  2, dCn  $\geq$  0.1, Sp  $\leq$  50 (NTT  $\geq$  1)

[M+3H]<sup>3+</sup>: Xcorr  $\geq$  2.5, dCn  $\geq$  0.1, Sp  $\leq$  50 (NTT  $\geq$  1)

[M+2H]<sup>2+</sup>: Xcorr  $\geq$  2, dCn  $\geq$  0.1, Sp  $\leq$  50, NTT  $\geq$  1

[M+3H]<sup>3+</sup>: Xcorr  $\geq$  2, dCn  $\geq$  0.1, Sp  $\leq$  50, NTT  $\geq$  1

# Traditional filtering criteria

---

Each SEQUEST search result has a:

Xcorr, dCn, Sp, NTT (number of tryptic termini)

Accept all results that satisfy:

[M+2H]<sup>2+</sup>: Xcorr  $\geq$  2, dCn  $\geq$  0.1, Sp  $\leq$  50 (NTT  $\geq$  1)

[M+3H]<sup>3+</sup>: Xcorr  $\geq$  2.5, dCn  $\geq$  0.1, Sp  $\leq$  50 (NTT  $\geq$  1)

[M+2H]<sup>2+</sup>: Xcorr  $\geq$  2, dCn  $\geq$  0.1, Sp  $\leq$  50 (NTT  $\geq$  1)

[M+3H]<sup>3+</sup>: Xcorr  $\geq$  2, dCn  $\geq$  0.1, Sp  $\leq$  50 (NTT  $\geq$  1)

# Problems with traditional filtering

---

- Different research groups use different thresholds
- Combines scores in unsatisfactory manner:  
What if Xcorr is just below its threshold, but dCn is far above?
- Divides data into correct and incorrect- no in between
- Unknown error rates (fraction of data passing filter that are incorrect)
- Unknown sensitivity (fraction of correct results passing filter)
- Appropriate threshold may depend on database, mass spectrometer type, sample, etc.

# Statistical Approach

---

Use search scores and properties of the assigned peptides to compute a probability that each search result is correct

Desirable model properties:

- Accurate
- High power to discriminate correct and incorrect results
- Robust

# Training dataset

---

Want dataset of SEQUEST search results for which the true correct and incorrect peptide assignments are known

Sample of 18 control proteins (bovine, yeast, bacterial)

Collect ~40,000 MS/MS spectra, and search using SEQUEST vs. a Drosophila database appended with sequences of 18 control proteins and common sample contaminants

# Training dataset

1	01.1261.1263.2	1384.7	(-0.7)	3.5634	0.167	928.8	1	17/ 22	sp P02666 CASB_BOVIN	F.LLYQEPVLGPVR.G
1	01.1707.1711.3	2535.7	(-0.0)	3.5606	0.370	527.5	1	27/ 88	SW:AMY_BACLI	K.GTSQADVGYGAYDLYDLGEFHQK.G
1	01.1881.1887.2	2052.4	(+0.5)	3.5302	0.343	555.5	1	20/ 34	sp P02754 LACB_BOVIN	Y.VEELKPTPEGDLLEILLQK.W
1	01.2479.2481.3	2254.5	(+0.4)	3.4933	0.303	1177.8	1	29/ 84	sp P00921 CAH2_BOVIN	K.YGDFGTAAQQPDGLAVGVFLK.V
1	01.2443.2443.3	4014.4	(-1.1)	3.4880	0.155	354.1	3	26/144	GP:AE003454_32	M.SLLSNKSTKVRSDSPQNRKVATVQEH
1	01.2377.2381.2	2854.3	(+0.1)	3.4856	0.225	294.7	1	17/ 48	sp P00921 CAH2_BOVIN	R.TLNFNAEGPELLMLANWRPAQPLK.I
1	01.2495.2499.3	2241.4	(-0.1)	3.4792	0.060	873.8	3	27/ 76	GP:AE003810_10	S.MSQPKSKTEFATDNQNGNRS.S
1	01.1405.1413.3	1610.9	(+0.7)	3.4700	0.146	717.2	1	24/ 52	sp P00489 PHS2_RABIT	K.VHINPNSLFDVQVK.R
1	01.0893.0893.3	2100.2	(+0.7)	3.4473	0.135	956.4	1	27/ 68	sp P00921 CAH2_BOVIN	R.MVNNGHSFNVEYDSDQK.A
1	01.0673.0675.3	2116.5	(+0.5)	3.4145	0.009	596.9	82	24/ 72	GP:AE003549_12	I.LPIFSPSEKVKQGPSHRIT.V
1	01.1203.1205.3	2848.9	(+0.7)	3.4106	0.186	273.5	82	25/ 96	GP:AE003824_13	H.VKTFEVNONGEAKQDEESKEALQE.
1	01.2027.2027.2	2278.5	(+0.2)	3.4070	0.352	732.6	1	19/ 40	sp P46406 G3P_RABIT	K.WGDAAGAEVYVVESTGVFTTMEK.A
1	01.2573.2575.3	4009.4	(+1.7)	3.3559	0.085	240.4	271	23/152	GP:AE003671_25	A.PNGDLYAQPMKSSSRNSLIPRPTKPA
1	01.0813.0817.2	963.1	(+0.1)	3.3490	0.131	877.3	3	12/ 14	sp P00432 CATA_BOVIN	K.DAQLFIQK.K
1	01.1111.1113.2	1467.5	(-0.5)	3.3379	0.384	811.4	1	17/ 22	sp Q29443 TRFE_BOVIN	K.TYDSYLGDDTVR.A
1	01.1981.1985.2	2224.6	(-0.0)	3.3017	0.361	745.5	1	20/ 38	sp P02666 CASB_BOVIN	K.IHPFAQTQSLVYVFPFGPIP.N
1	01.0485.0487.2	1013.1	(-0.1)	3.2952	0.254	1456.6	1	17/ 18	sp P00921 CAH2_BOVIN	K.VGDANPALQK.V
1	01.1419.1423.3	1690.9	(+0.6)	3.2929	0.233	952.0	3	25/ 52	sp P00489 PHS2_RABIT	K.ARPEFTLPVHFYGR.V
1	01.1751.1753.3	1890.1	(+1.0)	3.2805	0.116	635.8	1	21/ 56	sp P02769 ALBU_BOVIN	R.HPYFYAPELLYYANK.Y
1	01.1813.1815.2	1945.2	(-0.5)	3.2763	0.242	559.9	1	16/ 32	sp P02754 LACB_BOVIN	R.VYVEELKPTPEGDLLEIL.L
1	01.2645.2647.3	3338.7	(-0.2)	3.2529	0.156	220.6	8	25/108	GP:AE003473_27	L.KKTKPKPLTQQVTEETPHEEIIKES
1	01.2359.2361.3	2854.3	(-1.2)	3.2389	0.273	447.1	2	25/ 96	sp P00921 CAH2_BOVIN	R.TLNFNAEGPELLMLANWRPAQPLK.I
1	01.3057.3065.3	3843.3	(-1.2)	3.2233	0.212	365.0	3	26/132	SWN:VAF2_DROME	T.VDAHNLAVPTVLEIPSKQHPYDSRDR
1	01.2477.2485.3	2307.6	(+0.9)	3.2190	0.042	323.6	125	25/ 84	GP:AE003820_20	Q.KLVNQGQSPQAQKTSPPAQN.Q
1	01.1063.1065.2	1582.8	(-0.0)	3.2171	0.340	1508.4	1	18/ 24	sp P00921 CAH2_BOVIN	K.YAAELHLVHWNK.Y
1	01.1081.1083.2	974.1	(+0.2)	3.2106	0.163	454.9	2	14/ 16	sp P00921 CAH2_BOVIN	K.VLDALDSIK.T
1	01.1365.1369.3	1761.1	(-0.4)	3.2103	0.035	1195.9	17	23/ 56	GP:AE003798_9	T.SKGSRKMNINELLRN.E
1	01.1949.1949.2	2215.5	(-0.2)	3.2068	0.249	531.2	1	21/ 36	sp P02754 LACB_BOVIN	V.YVEELKPTPEGDLLEILLQK.W
1	01.2503.2509.3	2853.2	(+1.9)	3.1960	0.030	265.4	183	22/108	GP:AE003678_44	R.KASNACGSPFFVSSAESCPLKADL
1	01.0617.0623.3	2022.3	(-0.6)	3.1955	0.214	338.5	48	23/ 64	GP:M33019_1	G.KQLXDGRITSDYNIQKE.S
1	01.3503.3505.3	2709.1	(+1.5)	3.1834	0.207	686.4	1	27/100	sp P02754 LACB_BOVIN	K.VAGTWYSLAMAASDSLDDAQSAPLR
1	01.1791.1793.3	2325.6	(+1.9)	3.1792	0.160	742.5	1	25/ 80	GP:AE003576_11	T.IGFAWATYASGFVLGQSOPPH.R
1	01.1095.1097.3	3124.7	(-2.3)	3.1791	0.015	419.2	219	21/108	GP:AE003473_27	Q.NVQKEIRIIPTSAVETSMNVIKVKQ
1	01.1417.1425.2	1610.9	(-0.3)	3.1601	0.142	1389.3	1	20/ 26	sp P00489 PHS2_RABIT	K.VHINPNSLFDVQVK.R
1	01.0579.0581.2	1246.3	(+0.2)	3.1600	0.169	1196.1	1	18/ 20	sp P02754 LACB_BOVIN	R.TPEVDDEALEK.F
1	01.1587.1589.3	2204.6	(+0.3)	3.1594	0.240	1098.0	1	28/ 80	sp Q29443 TRFE_BOVIN	K.IMKGEADAMSLDGGYLYIAGK.C
1	01.3977.3985.2	2709.1	(-0.1)	3.1502	0.270	404.4	1	14/ 50	sp P02754 LACB_BOVIN	K.VAGTWYSLAMAASDSLDDAQSAPLR

Peptides  
corresponding to  
drosophila  
proteins are  
incorrect

Peptides  
corresponding to  
18 control  
proteins or  
contaminants are  
correct\*

# Combine multiple SEQUEST scores into single discriminant score F

---

Want to combine together Xcorr, dCn, and Sp in a linear manner to produce a new score, F, that maximally separates the correct and incorrect search results in the training dataset:

$$F = c_0 + c_1 Xcorr + c_2 dCn + c_3 Sp$$

Actually first transform Xcorr and Sp:

[M+2H]<sup>2+</sup>:  $Xcorr' = \log(Xcorr) / (\log 2 * \text{peplength})$

[M+3H]<sup>3+</sup>:  $Xcorr' = \log(Xcorr) / (\log 4 * \text{peplength})$

$$Sp' = \log Sp$$



# Derive Discriminant Function

---

Derive F for each precursor ion charge separately:

$$F = c_0 + c_1 X_{\text{corr}}' + c_2 dCn + c_3 Sp'$$

For  $[M+3H]^{3+}$  search results in training dataset,

$$c_0 = -2.0$$

$$c_1 = 10.68$$

$$c_2 = 11.26$$

$$c_3 = -0.2$$

$$F = -2.0 + 10.68 * X_{\text{corr}}' + 11.26 * dCn - 0.2 * Sp'$$

# Compute Discriminant Score

---

$$F = -2.0 + 10.68 * X_{\text{corr}}' + 11.26 * dCn - 0.2 * Sp'$$

Example:

Peptide **K.ARPEFTLPVHFYGR.V**

$X_{\text{corr}} = 3.29$

$dCn = 0.233$

$Sp = 3$

Precursor Ion Charge = 3

Peplength = 14

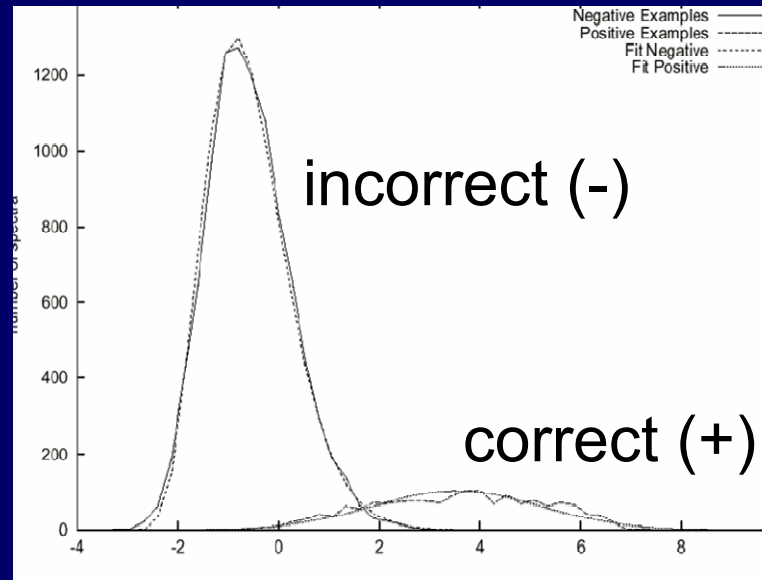
$X_{\text{corr}}' = \log(3.29) / (\log 56) = 0.296$

$Sp' = \log(3) = 1.09$

$$F = -2.0 + 10.68 * 0.296 + 11.26 * 0.233 + \\ -0.2 * 1.09 = 3.56$$

# Discriminant Score Distributions

*no  
of  
spectra*

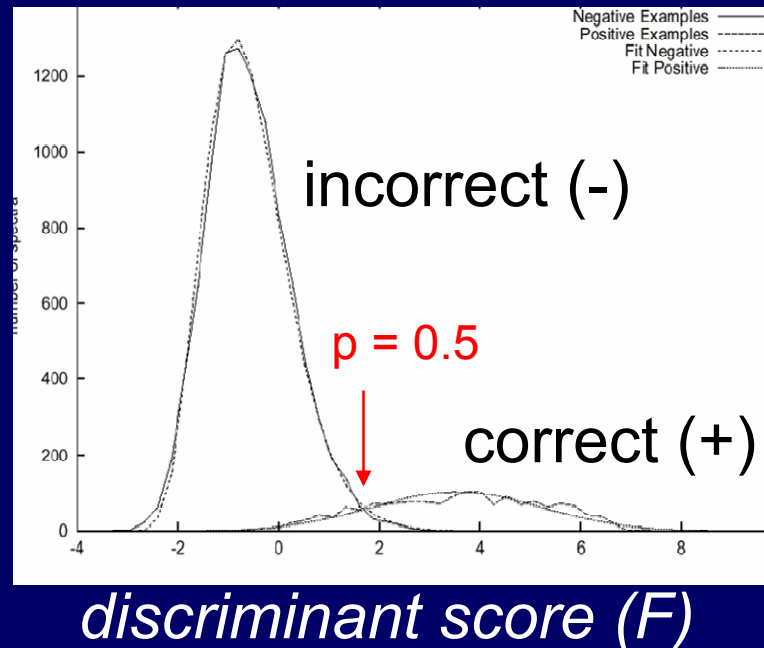


*discriminant score (F)*

Training dataset  $[M+2H]^{2+}$  spectra

# Computing probabilities from discriminant score distributions

*no  
of  
spectra*

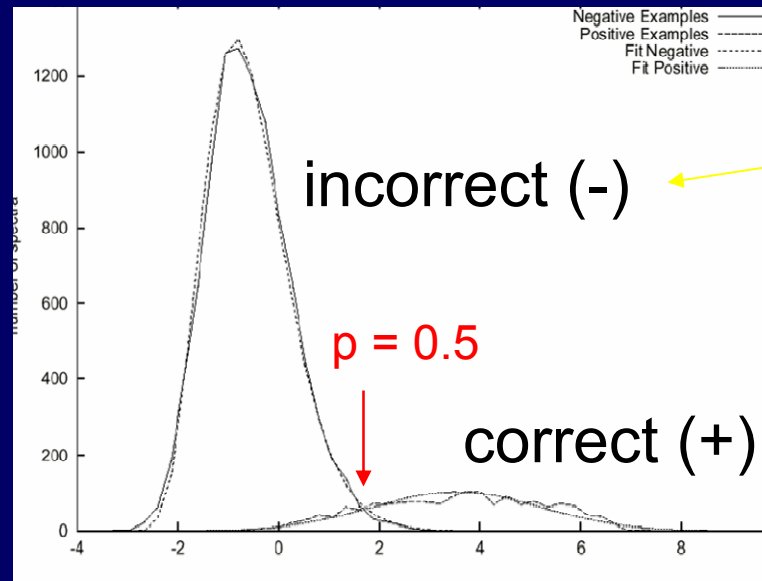


Probability of being correct, given discriminant score  $F_{\text{obs}}$ , is:

$$p = \frac{\text{Number of correct search results with } F_{\text{obs}}}{\text{Total number of search results with } F_{\text{obs}}}$$

# Computing probabilities from discriminant score distributions

*no  
of  
spectra*



*discriminant score (F)*

Model Incorrect  
results as  
**Gamma  
Distribution**

Model Correct  
results as  
**Normal  
Distribution**

Probability of being correct, given discriminant score  $F_{\text{obs}}$ , is:

$$p = \frac{\text{Normal}_{\mu,\sigma}(F_{\text{obs}}) * \text{Total correct}}{\text{Normal}_{\mu,\sigma}(F_{\text{obs}}) * \text{Total correct} + \text{Gamma}_{\alpha,\beta,\text{zero}}(F_{\text{obs}}) * \text{Total incorrect}}$$

# Employing peptide properties

---

Properties of the assigned peptides, in addition to search scores, are useful information for distinguishing correct and incorrect results

For example in unconstrained SEQUEST searches for MS/MS spectra collected from trypsinized samples, a majority of correct assigned peptides have 2 tryptic termini (preceded by K,R), whereas a majority of incorrect assigned peptides have 0 tryptic termini

# Number of tryptic termini (NTT)

---

NTT can equal 0, 1, or 2:

G.HVEQLDSSS.D NTT = 0

K.HVEQLDSSS.D NTT = 1

G.HVEQLDSSR.D NTT = 1

K.HVEQLDSSR.D NTT = 2

# Number of tryptic termini (NTT)

---

For the same value of  $F$ , assigned peptides with *higher* NTT values are *more* likely to be correct

Example: training dataset

**Correct:** 0.03 NTT=0, 0.28 NTT=1, **0.69** NTT=2

**Incorrect:** 0.80 NTT=0, 0.19 NTT=1, **0.01** NTT=2

Probability of being correct, given discriminant score  $F_{\text{obs}}$  with NTT=2 is:

$$p = \frac{\text{Normal}_{\mu,\sigma}(F_{\text{obs}}) * \text{Total corr} * 0.69}{\text{Normal}_{\mu,\sigma}(F_{\text{obs}}) * \text{Total corr} * 0.69 + \text{Gamma}_{\alpha,\beta,\text{zero}}(F_{\text{obs}}) * \text{Total incorr} * 0.01}$$

$F_{\text{obs}}$ :  $p = 0.5$  without NTT becomes  $p=0.99$  using NTT



# Number of tryptic termini (NTT)

---

For the same value of  $F$ , assigned peptides with *lower* NTT values are *less* likely to be correct

Example: training dataset

**Correct:** 0.03 NTT=0, 0.28 NTT=1, 0.69 NTT=2

**Incorrect:** 0.80 NTT=0, 0.19 NTT=1, 0.01 NTT=2

Probability of being correct, given discriminant score  $F_{\text{obs}}$  with NTT=0 is:

$$p = \frac{\text{Normal}_{\mu,\sigma}(F_{\text{obs}}) * \text{Total corr} * 0.03}{\text{Normal}_{\mu,\sigma}(F_{\text{obs}}) * \text{Total corr} * 0.03 + \text{Gamma}_{\alpha,\beta,\text{zero}}(F_{\text{obs}}) * \text{Total incorr} * 0.80}$$

$F_{\text{obs}}$ :  $p = 0.5$  without NTT becomes  $p=.04$  using NTT

# Additional peptide properties

---

Number of missed tryptic cleavages (NMC)

Mass difference between precursor ion and peptide

Presence of light or heavy cysteine (ICAT)

Presence of N-glyc motif (N-glycosylation capture)

Calculated pI (FFE)

Incorporate similar to NTT above, assuming independence of peptide properties and search scores among correct and incorrect results

# Computed Probabilities

---

Given training dataset distributions of F, NTT, NMC, Massdiff, ICAT, N-glyc, and pl among correct and incorrect search results,...

...then the probability of any search result with  $F_{\text{obs}}$ ,  $NTT_{\text{obs}}$ ,  $NMC_{\text{obs}}$ ,  $Massdiff_{\text{obs}}$ ,  $ICAT_{\text{obs}}$ ,  $N\text{-glyc}_{\text{obs}}$ , and  $pl_{\text{obs}}$  can be computed as described above, with terms for each piece of information

Accurate

Discriminating

# Robust Model

---

One cannot rely on the **training dataset** distributions of F, NTT, NMC, Massdiff, ICAT, N-glyc, and pl among correct and incorrect search results

These distributions are expected to vary depending on:

- Database used for search
- Mass spectrometer
- Spectrum quality
- Sample preparation and purity

# EM Algorithm

---

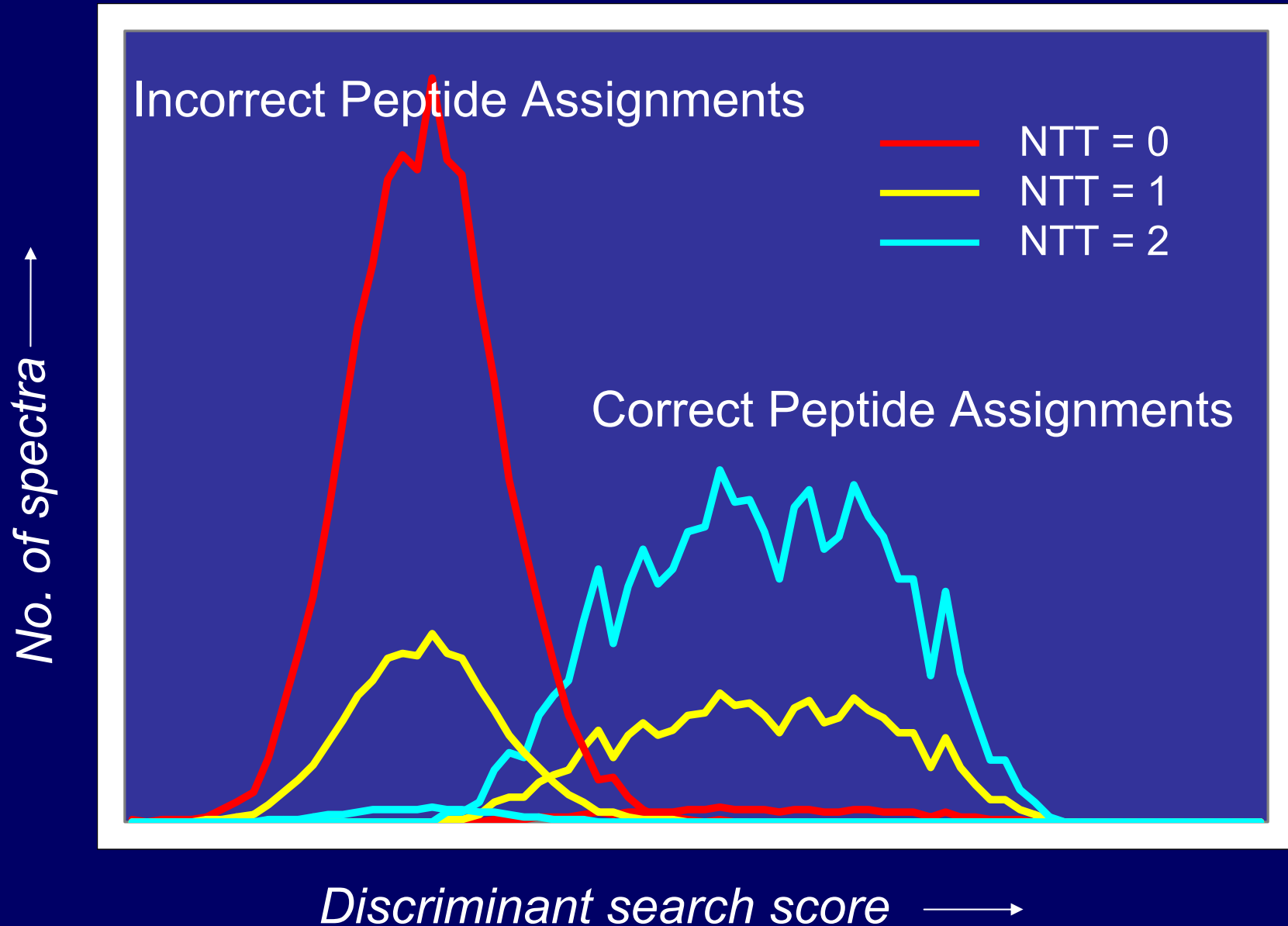
PeptideProphet learns the distributions of  $F$  and peptide properties among correct and incorrect search results in each dataset

It then uses the learned distributions to compute probabilities that each search result is correct

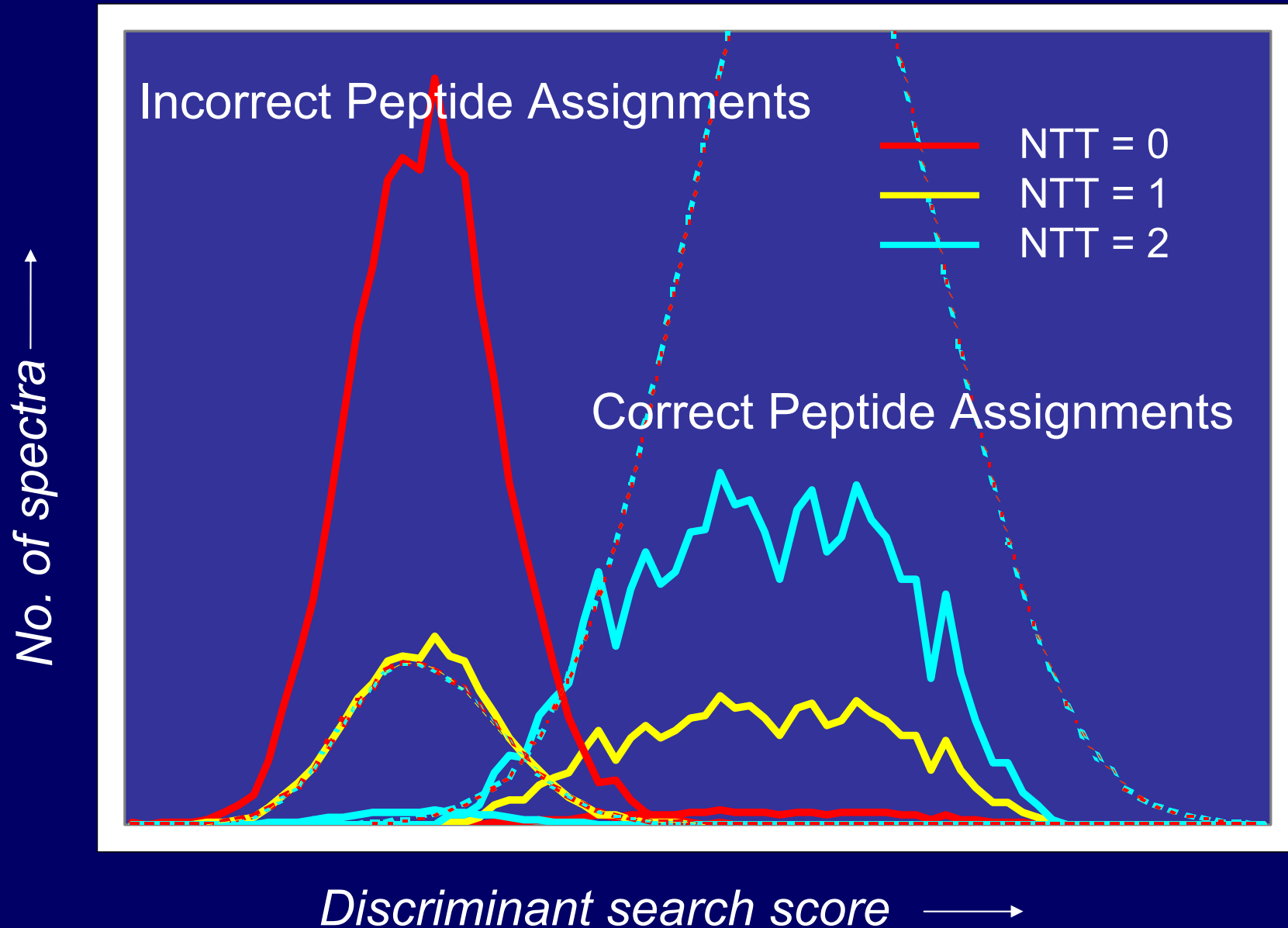
EM algorithm: unsupervised learning method that ***iteratively*** estimates the distributions given probabilities that each search result is correct, and then computes those probabilities given the distributions

Initial settings help guide algorithm to good solution

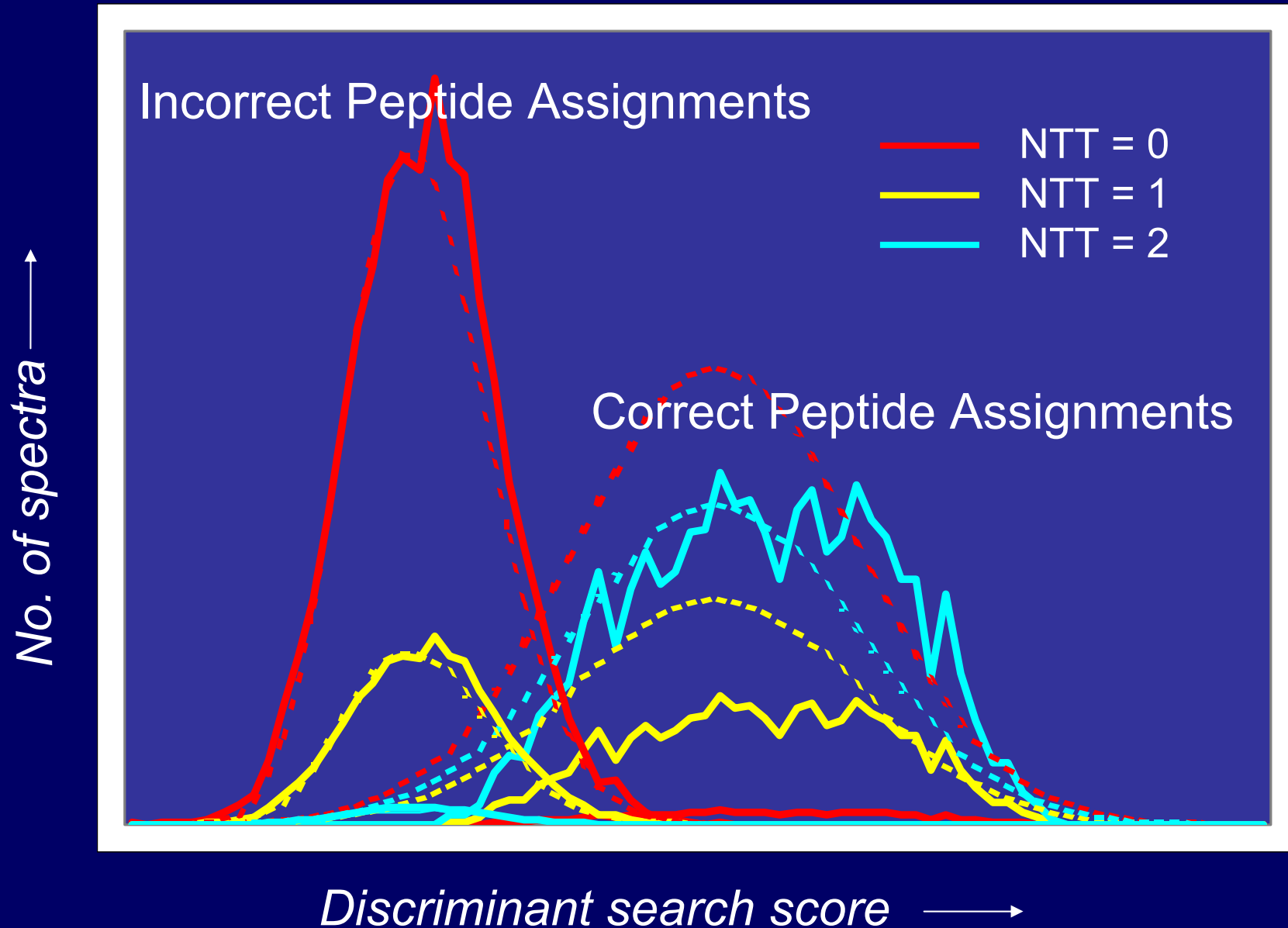
# E-M Algorithm learns test data score distributions



# E-M Iteration 0

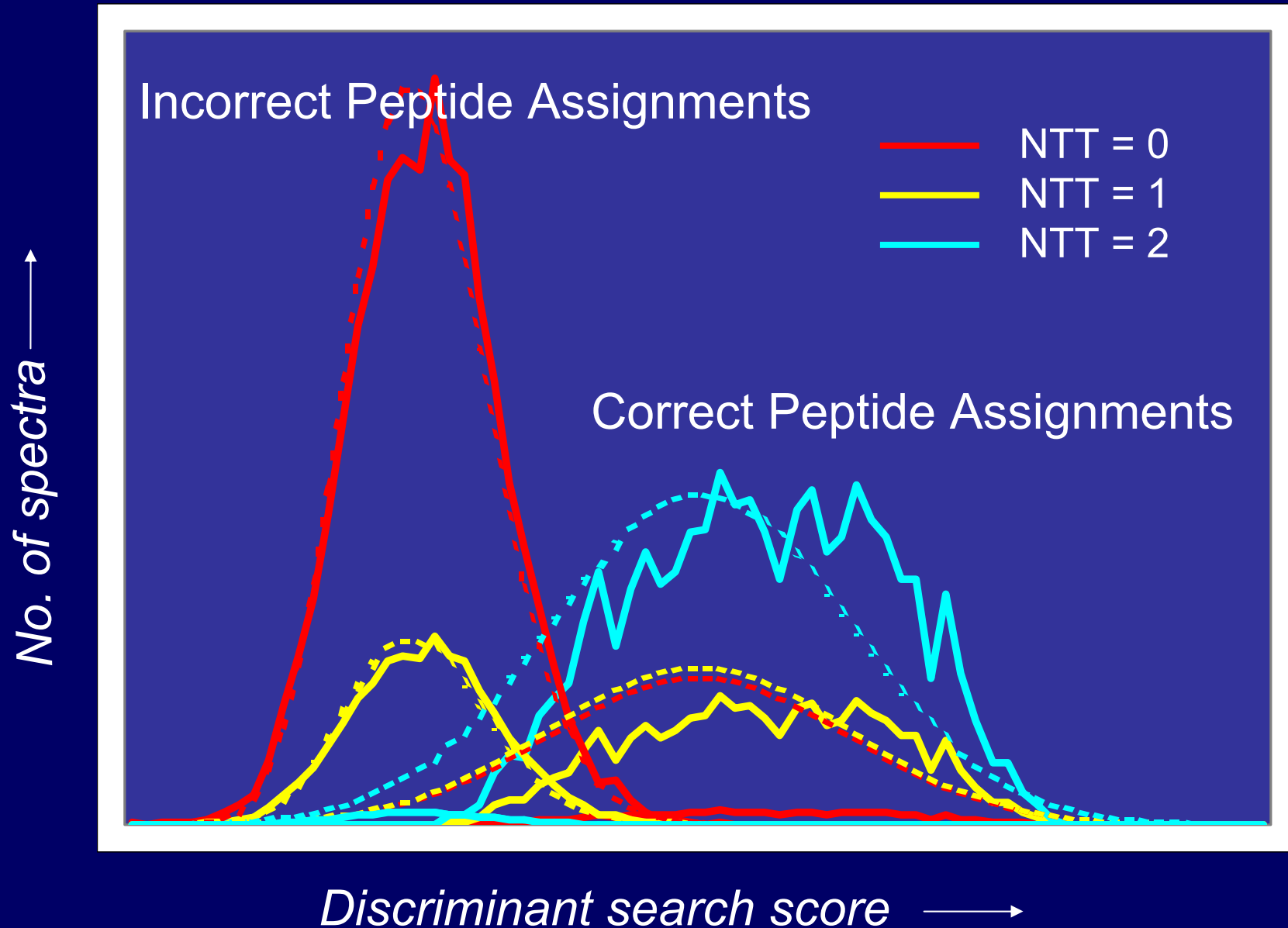


# E-M Iteration 1

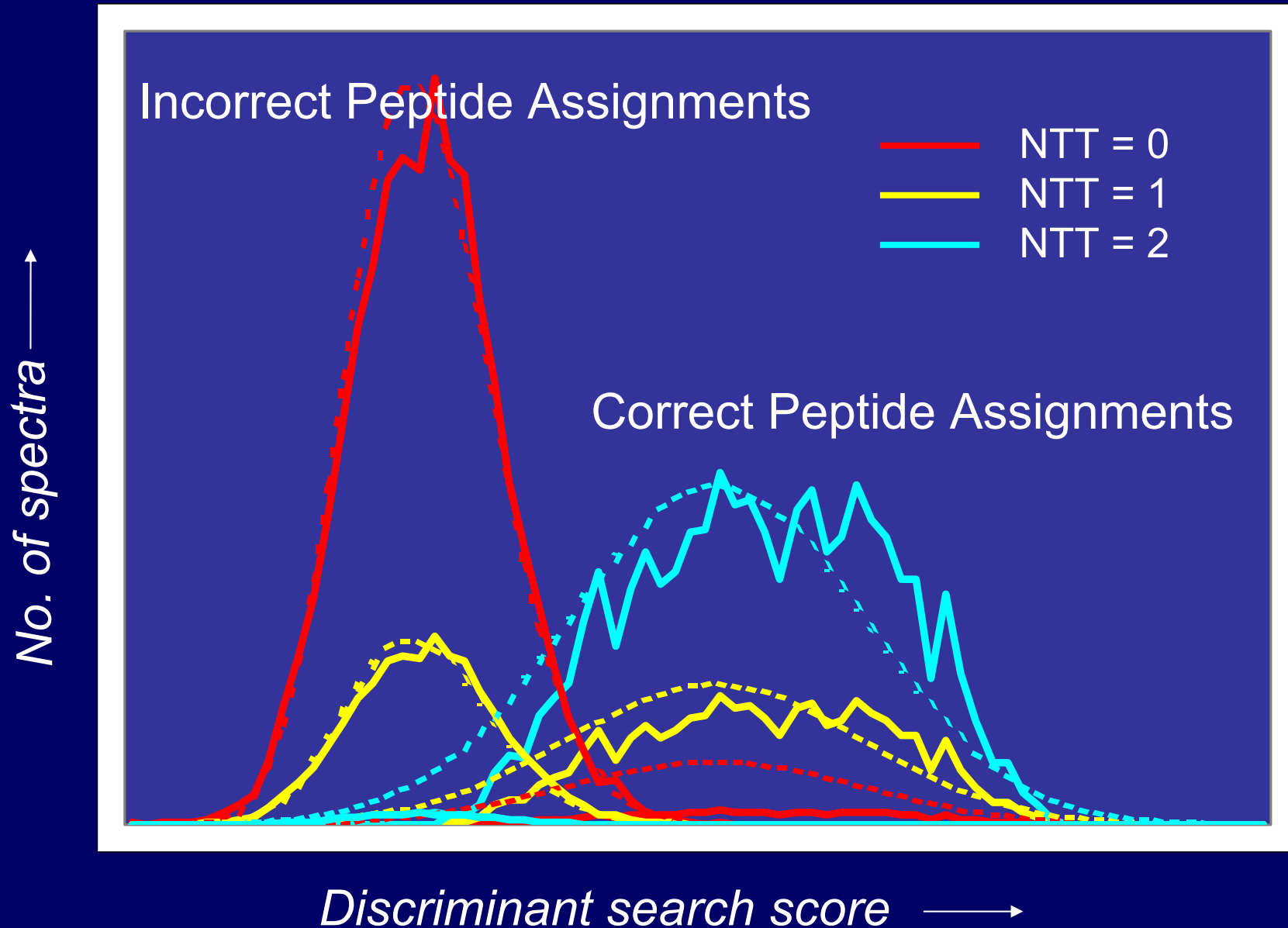




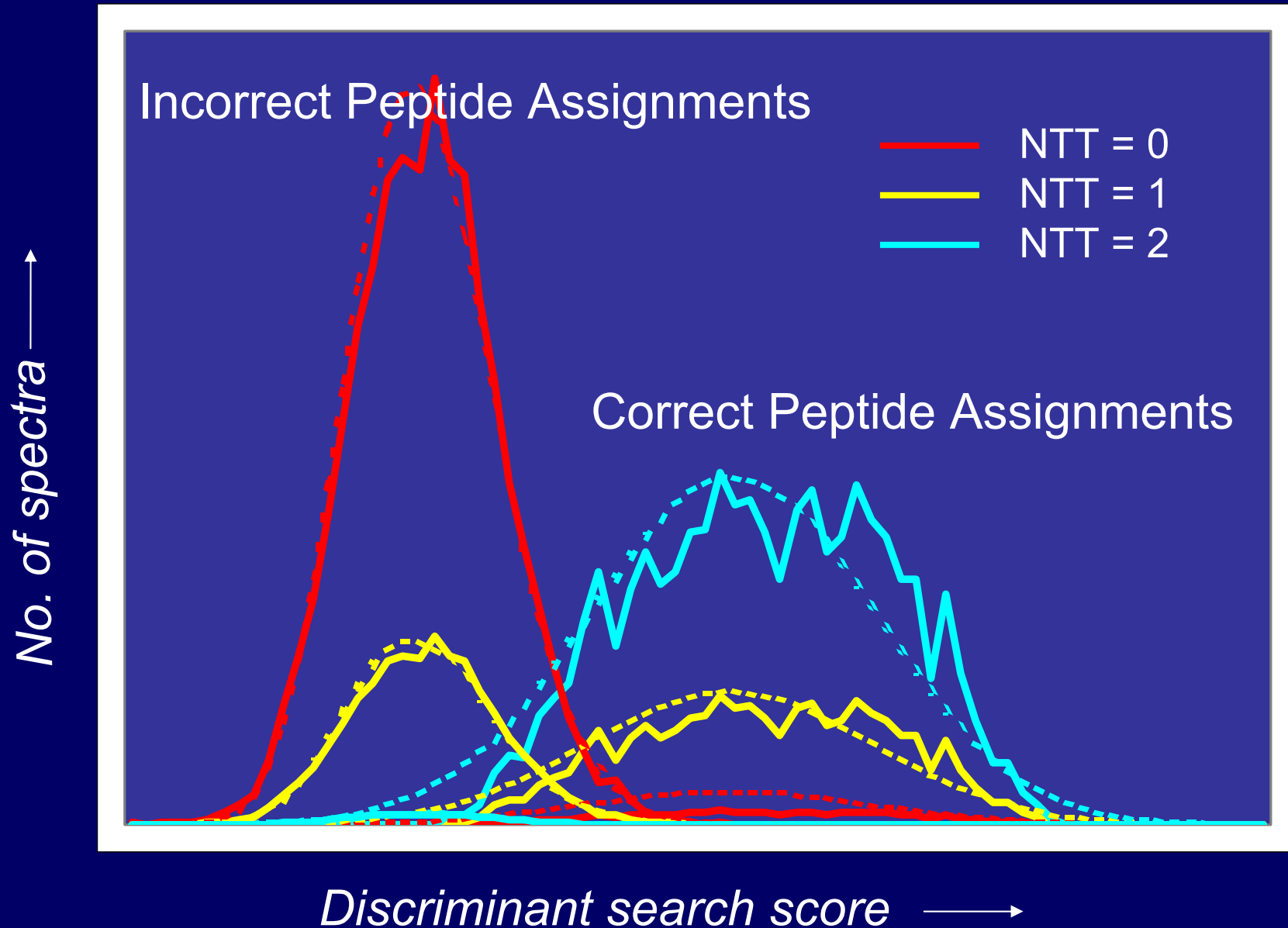
# E-M Iteration 2



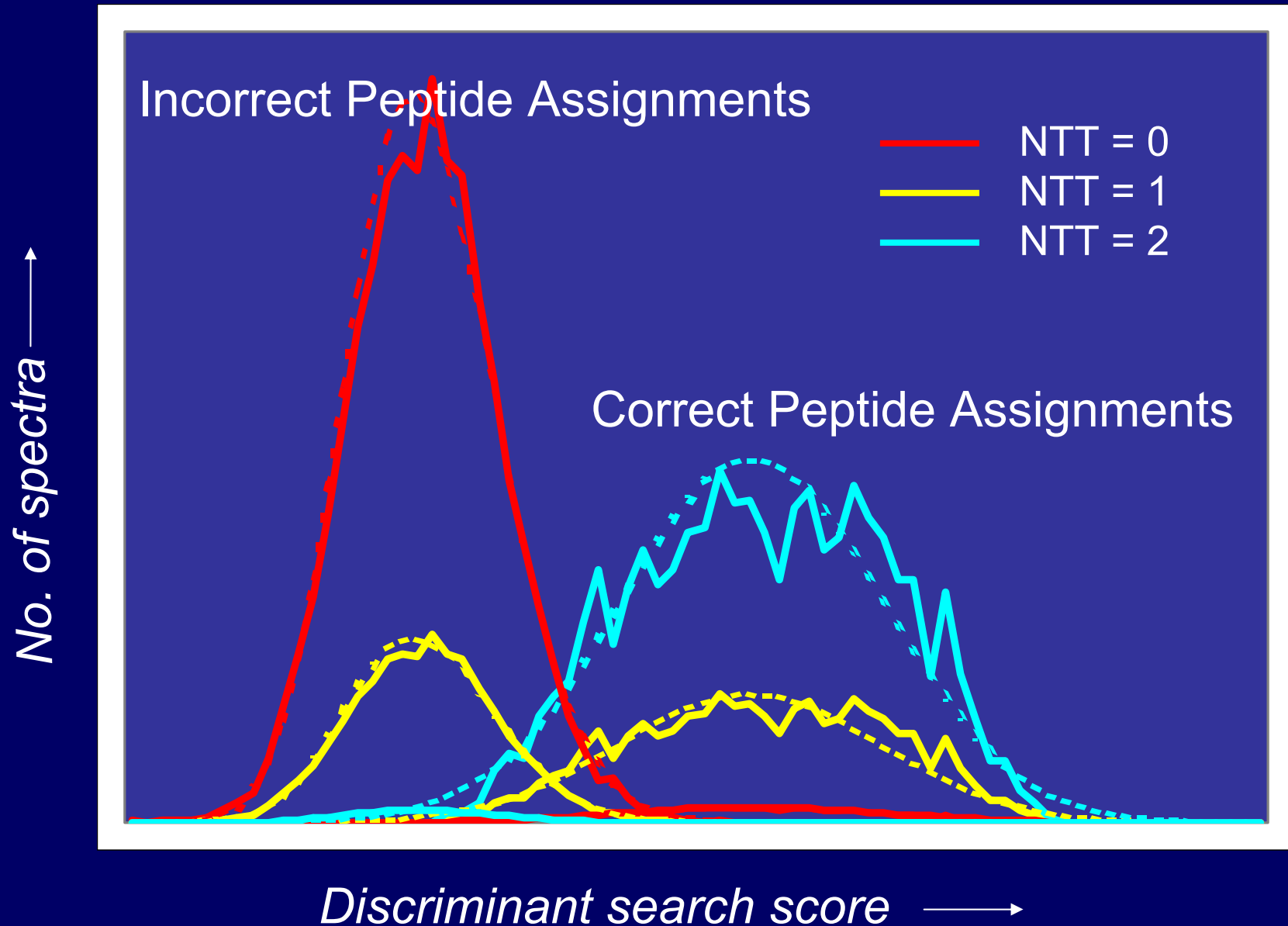
# E-M Iteration 3



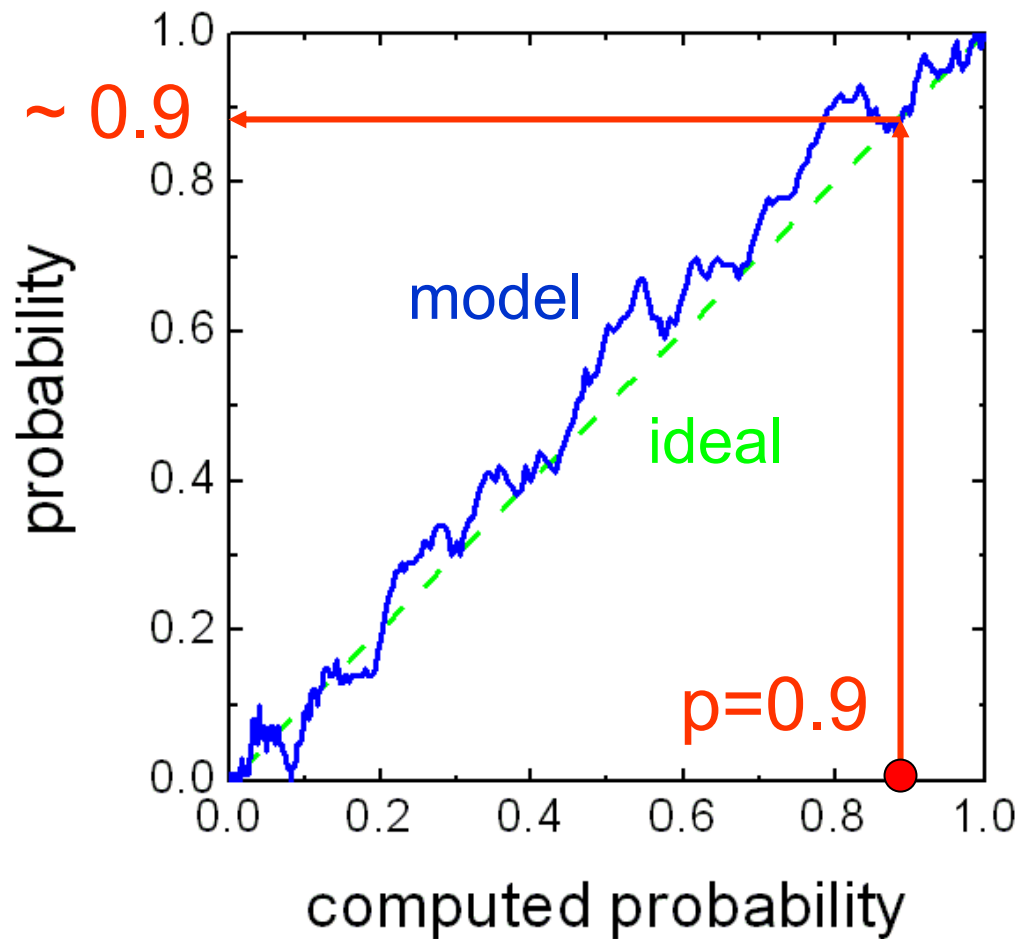
# E-M Iteration 4



# E-M Iteration 7



# Accuracy of the Model



test data: A. Keller *et al.* OMICS 6, 207 (2002)

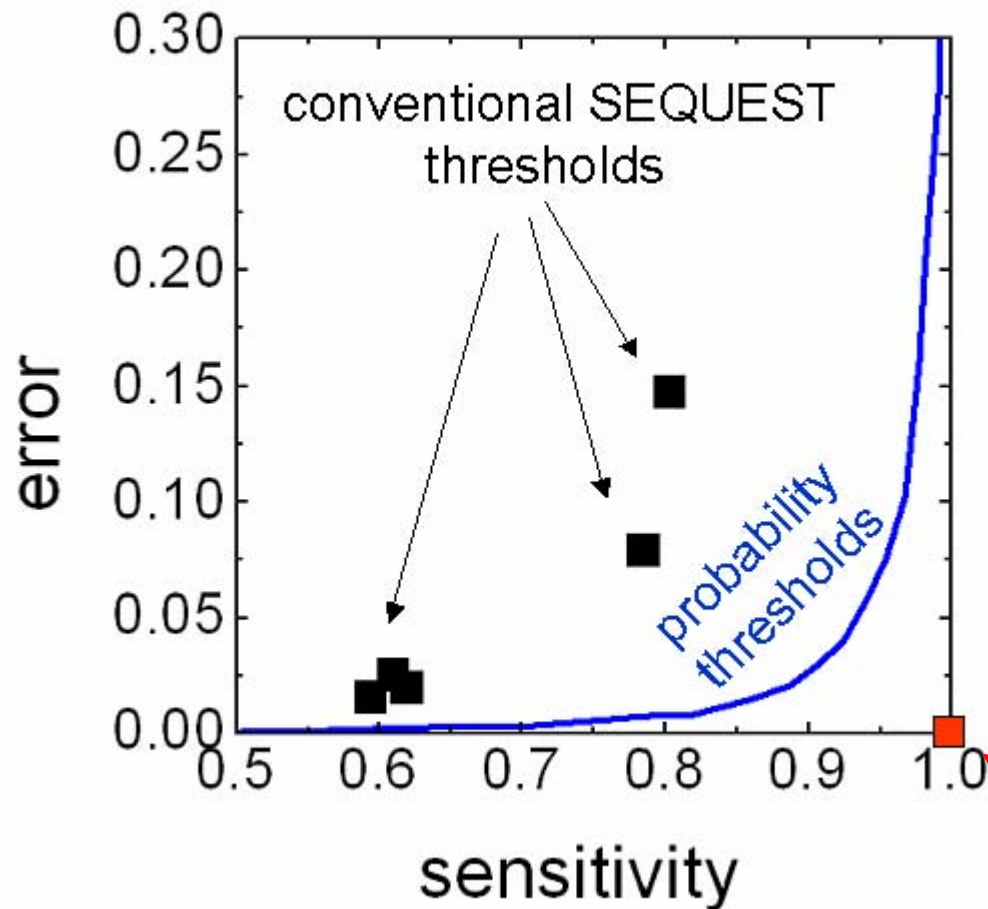
100 spectra with  
computed  $p \sim 0.9$

90% of them (90)  
should be correct

observed  
probability is  
around 0.9

Model is accurate

# Discriminating Power of Computed Probabilities



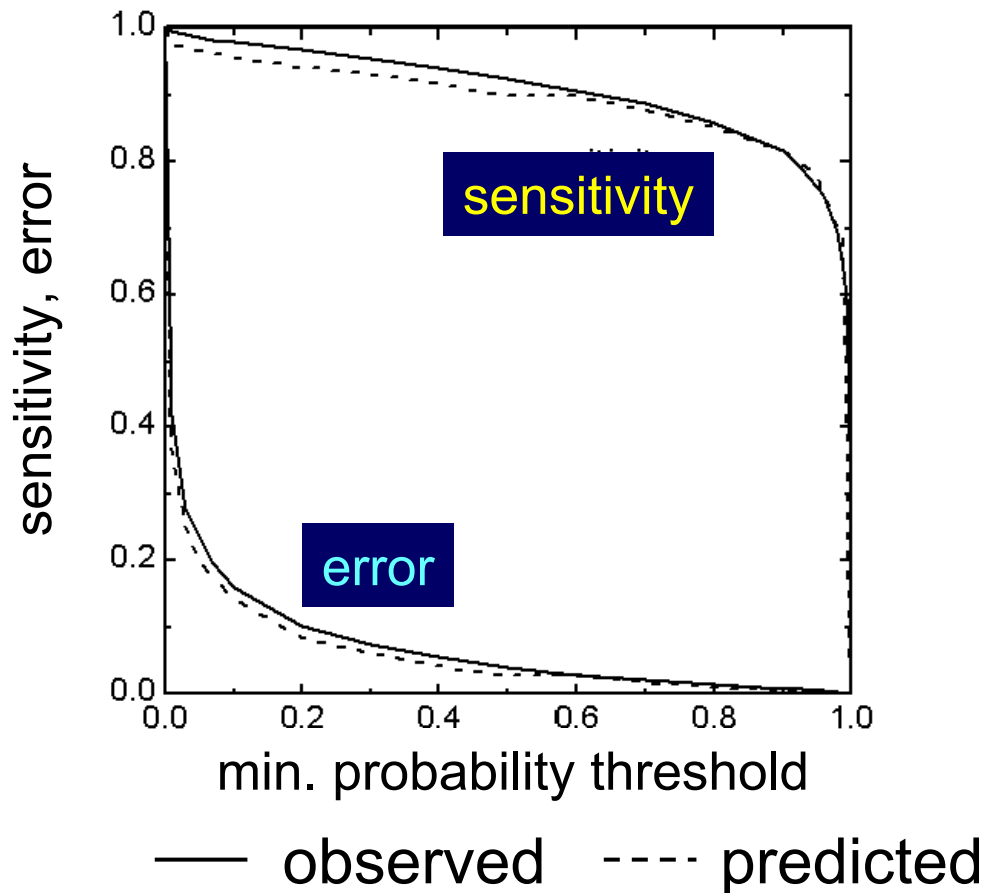
**Sensitivity:**  
fraction of all  
correct results  
passing filter

**Error:**  
fraction of all  
results passing  
filter that are  
incorrect

**Ideal Spot**

test data: A. Keller *et al.* OMICS 6, 207 (2002)

# Discriminating Power of Computed Probabilities



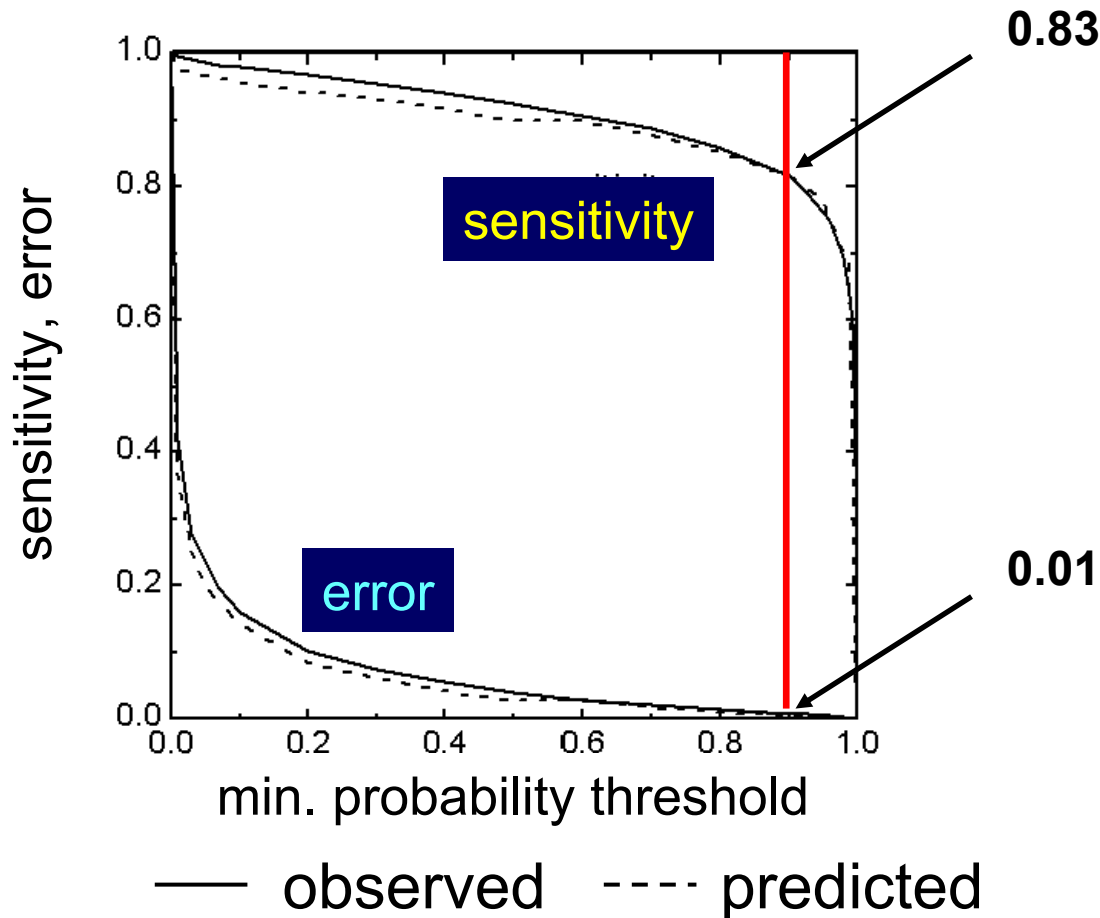
***Sensitivity:***  
fraction of all  
correct results  
passing filter

***Error:***  
fraction of all  
results passing  
filter that are  
incorrect

test data: A. Keller *et al.* OMICS 6, 207 (2002)

# Discriminating Power :

## Example $p \geq 0.9$



**Sensitivity:**  
fraction of all  
correct results  
passing filter

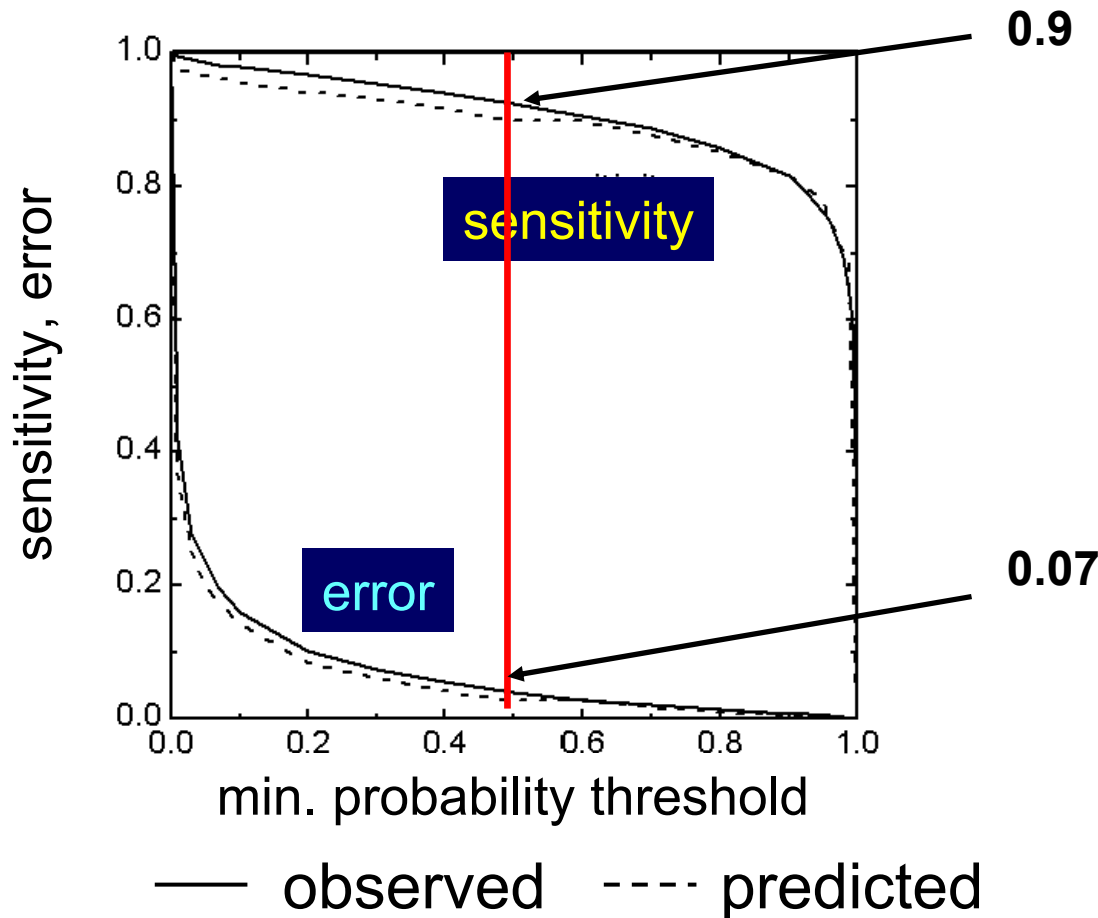
**Error:**  
fraction of all  
results passing  
filter that are  
incorrect

test data: A. Keller *et al.* OMICS 6, 207 (2002)



# Discriminating Power :

## Example $p \geq 0.5$



**Sensitivity:**  
fraction of all  
correct results  
passing filter

**Error:**  
fraction of all  
results passing  
filter that are  
incorrect

test data: A. Keller *et al.* OMICS 6, 207 (2002)

# Can experts discriminate better than model?

---

20 of the test spectra were assigned probabilities close to 0.5 (in complete test dataset)

Of those, 10 were correct:

on average:

51% 'publishable', 24% 'borderline', 25% 'not pub'

and 10 were incorrect:

on average:

11% 'publishable', 16% 'borderline', 74% 'not pub'

# Getting started with PeptideProphet

---

Input: SEQUEST summary html files (`file.html`)

Interact merges files together into `interact.htm`, then PeptideProphet runs model, computes probabilities, and writes probabilities as first column

Combine together runs that are similar (sample, database, search constraints, mass spectrometer)

# Getting started with PeptideProphet

File Edit View Favorites Tools Help

Address [http://localhost/isb-bin/interact\\_web.pl](http://localhost/isb-bin/interact_web.pl)

INSTITUTE FOR Systems Biology

## ISB Interact/PeptideProphet - Web Interface

**Specify Parameters**

Please specify the Interact and (optional) PeptideProphet parameters. Then click on 'Run' to process the files.

**1. Specify Input/Output Parameters**

Add File:  Browse... Update

File List:

C:\inetpub\wwwroot\class\day2\Unconstrained\haloICAT2\_30.html

Location for output files: C:\inetpub\wwwroot\class\day2\Unconstrained

Rename output files: noicat

**2. Specify Interact Parameters**

**WARNING: Changing these options if PeptideProphet is also run may invalidate the resulting probabilities.**

Required residues:

DeltaCn cutoff:  +1 XCorr cutoff:

Number of Tryptic Termini:  +2 XCorr cutoff:

Sp Rank Cutoff:  +3 XCorr cutoff:

ISB Interact/PeptideProphet - Web Interface - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Address [http://localhost/isb-bin/interact\\_web.pl?monitor=1&result=http://localhost/class/day2/Unconstrained/interact-noicat.htm](http://localhost/isb-bin/interact_web.pl?monitor=1&result=http://localhost/class/day2/Unconstrained/interact-noicat.htm)

INSTITUTE FOR Systems Biology

## ISB Interact/PeptideProphet - Web Interface

**Command Finished**

Results URL: <http://localhost/class/day2/Unconstrained/interact-noicat.htm>

Results Path: C:\inetpub\wwwroot\class\day2\Unconstrained\interact-noicat.htm

**Run ProteinProphet**

# PeptideProphet Results

INTERACT by J.Eng. Institute for Systems Biology - Mozilla {Build ID: 2002051319}

Sort/Restore:   [Help](#)

FILE:  Tryptic: ☐ 1 ☐ 2 ☐ MaxMissed:  DelRows: ☐ ☐

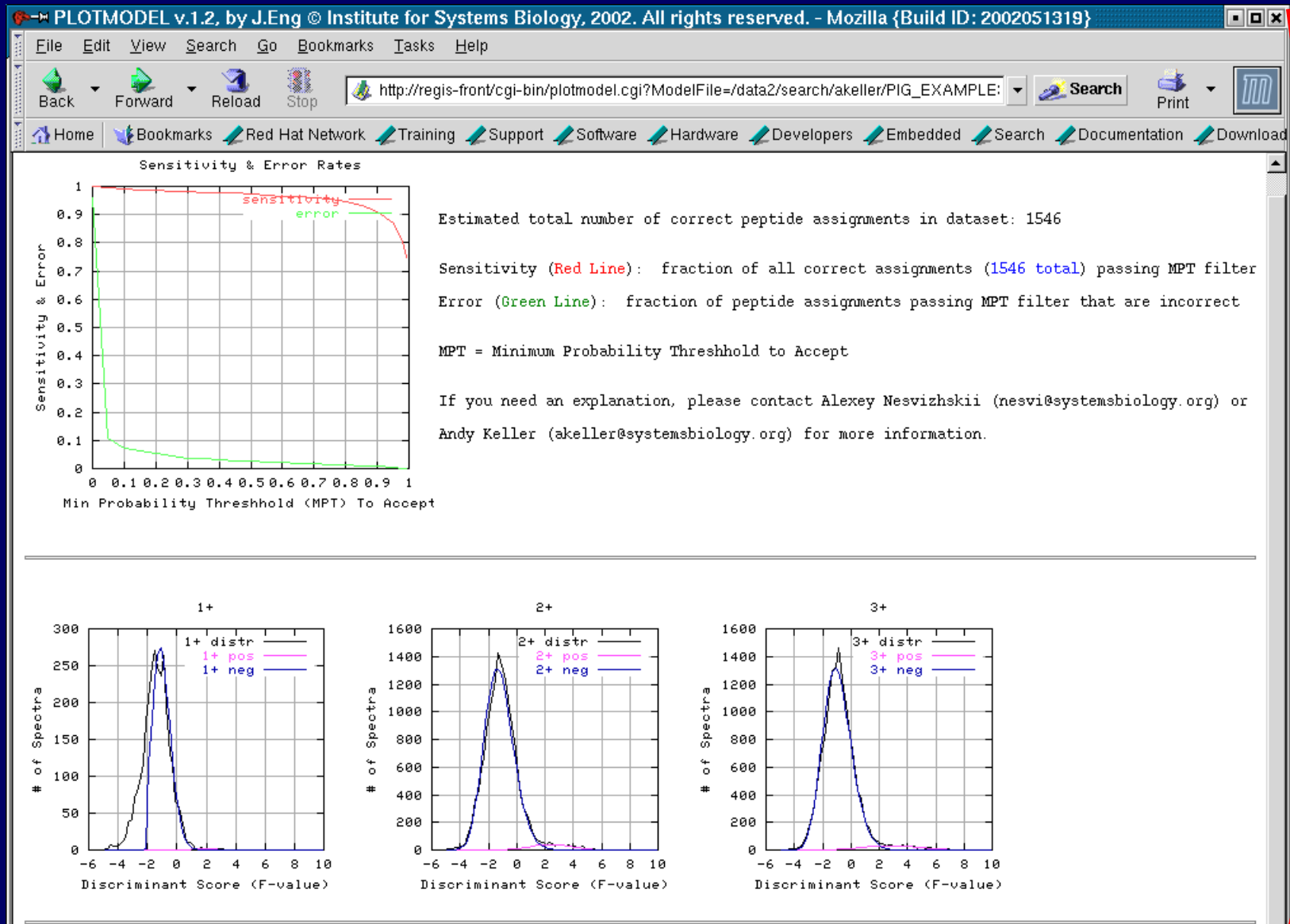
XCorr: ☐ +1 ☐ +2 ☐ +3 ☐ dCn: ☐ ☐ RSp: ☐ ☐ InclAA: ☐ MarkAA: ☐ NxS/T: ☐ Prob: ☐ ☐ XPRESS: ☐ ☐

Exclude charge: +1 ☐ +2 ☐ +3 ☐ Txt1:  Txt2:  J.Eng 04/2000

1.0000	366	/raft1011.1689.1689.3	2905.9	(-0.8)	3.9132	0.239	884.9	1	28/ 84	SW:POR2_HUMAN	+4	K. WNTDNLGTEIAIEDQICQGLK. L
1.0000	346	/raft1011.1665.1665.3	2335.3	(+0.4)	3.7540	0.337	1398.5	1	30/ 72	GP:AF380932_1	+3	R. GILAVLDEACSSAGTTIDR. I
0.9988	364	/raft1011.1687.1687.3	2913.9	(-0.4)	3.6121	0.068	597.8	1	27/ 84	SW:POR2_HUMAN	+4	K. WNTDNLGTEIAIEDQIC+QGLK. L
0.9992	400	/raft1011.1753.1753.2	2454.4	(-0.4)	3.4908	0.161	1049.6	1	20/ 38	SW:NUBM_HUMAN	+4	R. GAGAYICGEETALIESIEGK. Q
1.0000	395	/raft1011.1745.1745.3	2945.0	(+0.5)	3.1644	0.327	573.0	1	27/ 88	SW:ACTB_HUMAN	+8	K. LC*YVALDFEQEMATAASSSSLEK. S
1.0000	336	/raft1011.1645.1645.3	2251.3	(+0.4)	3.1525	0.293	700.5	1	25/ 64	SWN:TBA2_HUMAN	+13	R. AVCMLSNITAIAEAWAR. L
0.9999	384	/raft1011.1729.1729.3	2802.9	(+0.4)	3.0756	0.206	740.7	1	25/ 88	SW:GBB1_HUMAN	+6	R. VSCLGVTDGMAVATGSWDSFLK. I
1.0000	313	/raft1011.1593.1593.2	2744.7	(-0.9)	3.0040	0.245	445.5	1	14/ 38	SW:MLRM_HUMAN	+2	R. NAFAC+PDEATGTIQEDYLR. E
0.9993	382	/raft1011.1725.1725.3	2810.9	(+0.5)	3.0039	0.129	832.0	1	26/ 88	SW:GBB1_HUMAN	+6	R. VSC+LGVTDGMAVATGSWDSFLK. I
1.0000	396	/raft1011.1749.1749.2	2462.4	(+0.2)	3.0024	0.258	826.3	1	20/ 38	SW:NUBM_HUMAN	+4	R. GAGAYIC+GEETALIESIEGK. Q
1.0000	350	/raft1011.1671.1671.2	2335.3	(-1.1)	2.8016	0.338	1135.7	1	20/ 36	GP:AF380932_1	+3	R. GILAVLDEACSSAGTTIDR. I
0.9999	334	/raft1011.1641.1641.3	2259.3	(-0.2)	2.7139	0.211	672.7	1	23/ 64	SWN:TBA2_HUMAN	+13	R. AVC+MLSNITAIAEAWAR. L
0.9999	347	/raft1011.1667.1667.2	2343.3	(-0.5)	2.6325	0.215	607.7	1	16/ 36	GP:AF380932_1	+3	R. GILAVLDEAC+SSAGTTIDR. I
0.9666	430	/raft1011.1811.1811.3	2461.7	(+0.3)	1.6820	0.177	162.0	197	16/ 72	SWN:SEK1_HUMAN	+3	R. FVPFAAANAANCINIPLMR. Q
0.2196	354	/raft1011.1677.1677.2	1560.9	(+0.7)	1.6171	0.108	336.8	1	14/ 26	GP:AF486833_1		R. M#ASSM#KQVNPPLPK. V
0.9988	275	/raft1213.1757.1757.3	2937.0	(-0.7)	2.6890	0.183	562.6	1	24/ 88	SW:ACTB_HUMAN	+8	K. LCYVALDFEQEMATAASSSSLEK. S
0.9957	273	/raft1213.1755.1755.3	2945.0	(-0.4)	2.4975	0.096	658.8	1	26/ 88	SW:ACTB_HUMAN	+8	K. LC*YVALDFEQEMATAASSSSLEK. S
0.1629	446	/raft1213.2131.2131.1	709.8	(-0.1)	1.4822	0.249	256.6	29	8/ 12	SW:IRX2_HUMAN	+1	R. LQGPPTP. A
1.0000	442	/raft1415.1992.1992.2	2510.8	(-1.4)	3.7934	0.348	1175.7	1	22/ 44	SW:CY1_HUMAN	+3	A. PPIYTDVLEFDDGTPTATMSQIAK. D
0.9997	340	/raft1415.1752.1752.3	2937.0	(+1.5)	3.5180	0.130	765.5	1	27/ 88	SW:ACTB_HUMAN	+8	K. LCYVALDFEQEMATAASSSSLEK. S
1.0000	338	/raft1415.1746.1746.3	2945.0	(+0.4)	2.6882	0.278	774.5	1	27/ 88	SW:ACTB_HUMAN	+8	K. LC*YVALDFEQEMATAASSSSLEK. S
0.0669	70	/raft1415.0930.0930.2	1048.3	(+0.9)	1.1702	0.119	266.2	5	8/ 14	SW:VU8_HSV6U	+2	R. KHYIAM#LR. E
0.0584	503	/raft1415.2136.2136.2	2268.2	(-0.9)	1.1592	0.258	91.4	2	7/ 34	GP:AF194659_1		Y. CYSTDSSGNHRVVGGGTK. L
0.8050	734	/raft1617.1975.1975.2	2510.8	(-0.8)	3.0485	0.306	623.3	1	17/ 44	SW:CY1_HUMAN	+3	A. PPIYTDVLEFDDGTPTATMSQIAK. D
0.9496	8	/raft1617.0020.0020.3	2953.0	(+0.7)	2.0280	0.025	638.6	1	25/ 88	SW:ACTB_HUMAN	+8	K. LCYVALDFEQEMATAASSSSLEK. S
0.9191	12	/raft1617.0024.0024.3	2961.0	(-1.0)	1.9067	0.076	563.0	1	23/ 88	SW:ACTB_HUMAN	+8	K. LC*YVALDFEQEMATAASSSSLEK. S
0.0920	209	/raft1617.1191.1191.2	1104.3	(+0.4)	1.4424	0.035	202.1	57	8/ 16	GP:AF039913_1	+1	K. ASSFEYLR. H
0.9999	1101	/raft1819.2294.2294.3	2937.0	(-0.4)	3.5105	0.217	590.2	9	25/ 88	SW:ACTB_HUMAN	+8	K. LCYVALDFEQEMATAASSSSLEK. S
0.9994	471	/raft1819.1450.1450.2	2419.4	(+0.8)	3.3857	0.316	508.6	1	16/ 36	SW:ATP0_HUMAN		V. PCTVTSASPLEEATLSLK. T
0.9999	462	/raft1819.1440.1440.2	2704.7	(+1.5)	3.3215	0.264	648.6	1	19/ 42	SW:ATP0_HUMAN		R. GEVPCTVTSASPLEEATLSLK. T

Document: Done (22.815 secs)

# PeptideProphet Results: Model Summary



INTERAC

Sort/Restore

FILE: /data2/s

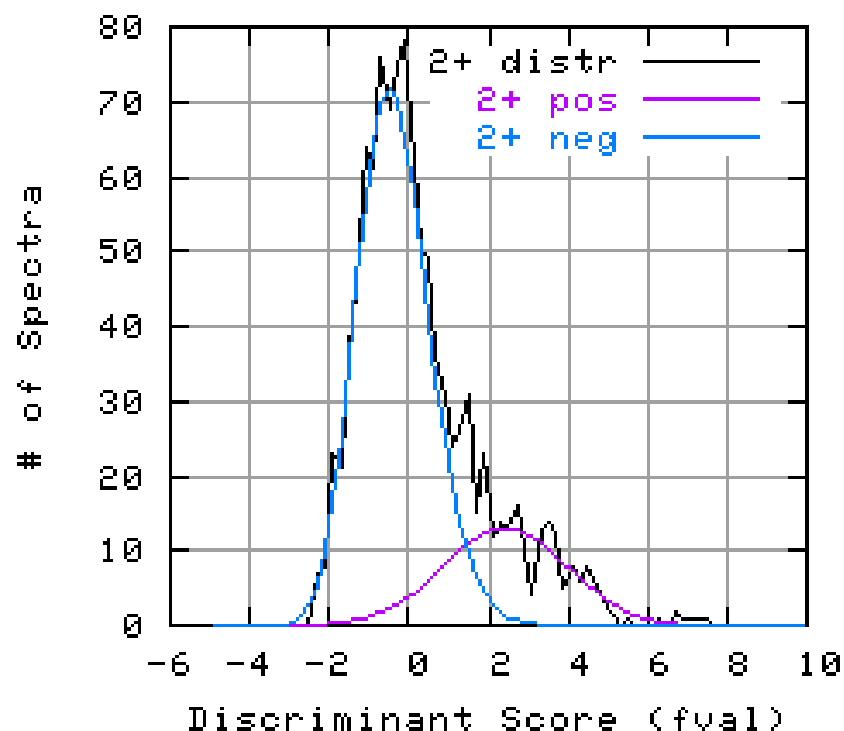
XCorr: ☐ +1

Exclude char

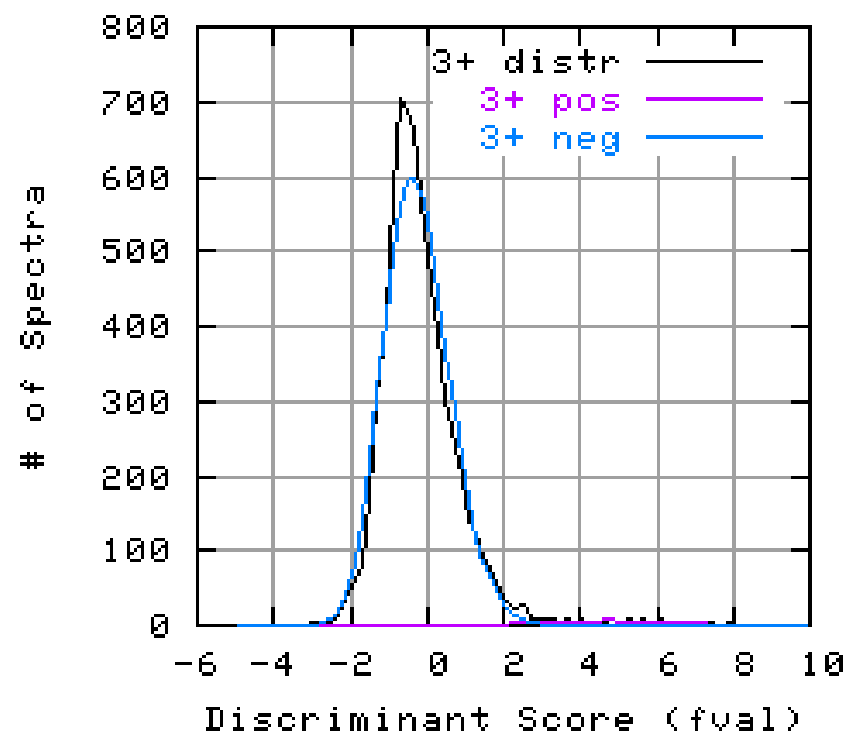
1.0000	366
1.0000	346
0.9988	364
0.9992	400
1.0000	395
1.0000	336
0.9999	384
1.0000	313
0.9993	382
1.0000	396
1.0000	350
0.9999	334
0.9999	347
0.9666	430
0.9136	354
0.9988	275
0.9957	273
0.1629	446
1.0000	442
0.9997	340
1.0000	338
0.0669	70
0.0584	503
0.8050	734
0.9496	8
0.9191	12
0.0920	209
0.9999	100

# Reasonable Learned Discriminant Score Distributions

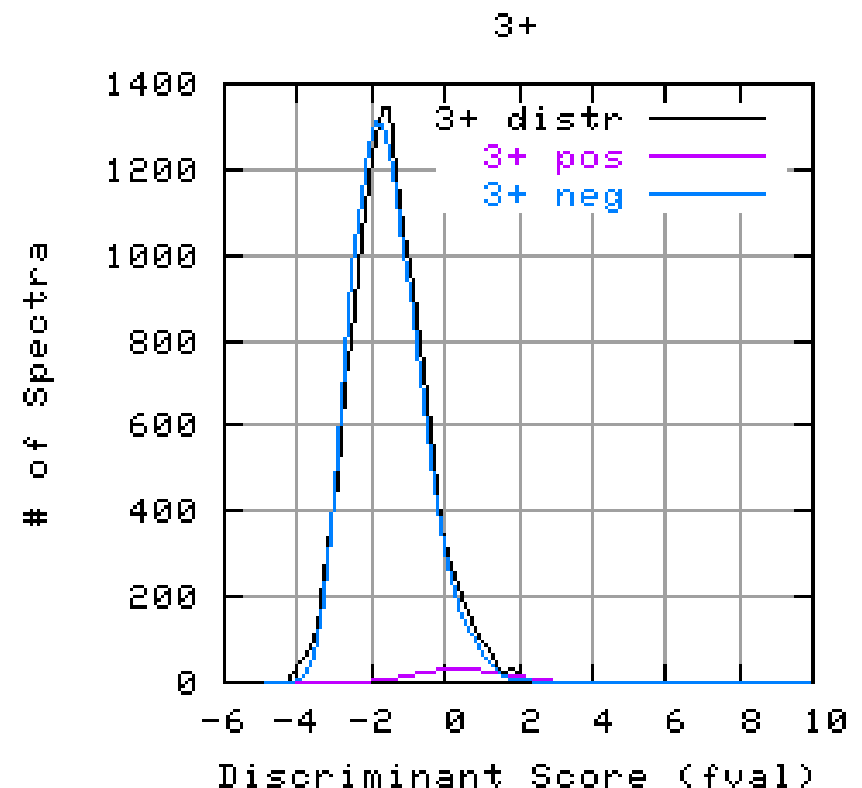
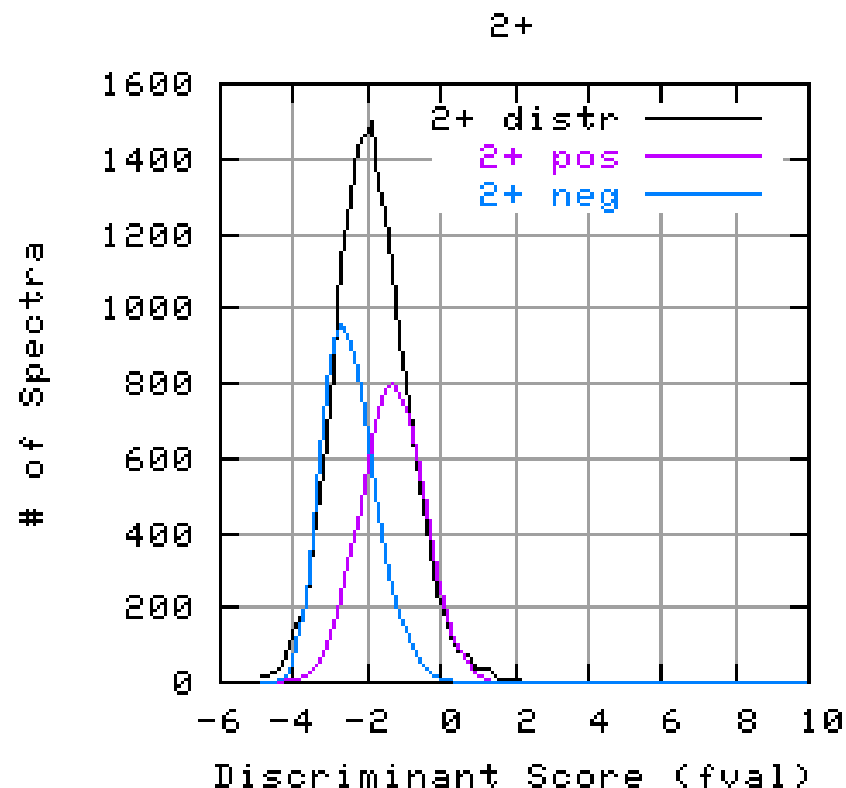
2+



3+



# *Suspicious* Looking Learned Discriminant Score Distributions





# PeptideProphet Results: Model Summary

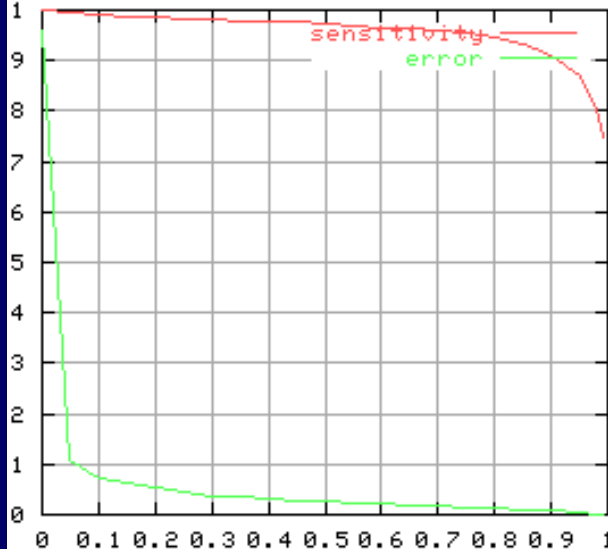
OTMODEL v.1.2, by J.Eng © Institute for Systems Biology, 2002. All rights reserved. - Mozilla {Build ID: 2002051319}

Edit View Search Go Bookmarks Tasks Help

Forward Reload Stop [http://regis-front/cgi-bin/plotmodel.cgi?ModelFile=/data2/search/akeller/PIG\\_EXAMPLE:](http://regis-front/cgi-bin/plotmodel.cgi?ModelFile=/data2/search/akeller/PIG_EXAMPLE:) Search

me Bookmarks Red Hat Network Training Support Software Hardware Developers Embedded Search Doc

Sensitivity & Error Rates



in Probability Threshold (MPT) To Accept

Estimated total number of correct peptide assignments in dataset: 1546

Sensitivity (Red Line): fraction of all correct assignments (1546 total) p

Error (Green Line): fraction of peptide assignments passing MPT filter tha

MPT = Minimum Probability Threshold to Accept

If you need an explanation, please contact Alexey Nesvizhskii (nesvi@system

Andy Keller (akeller@systemsbiology.org) for more information.

1+

2+

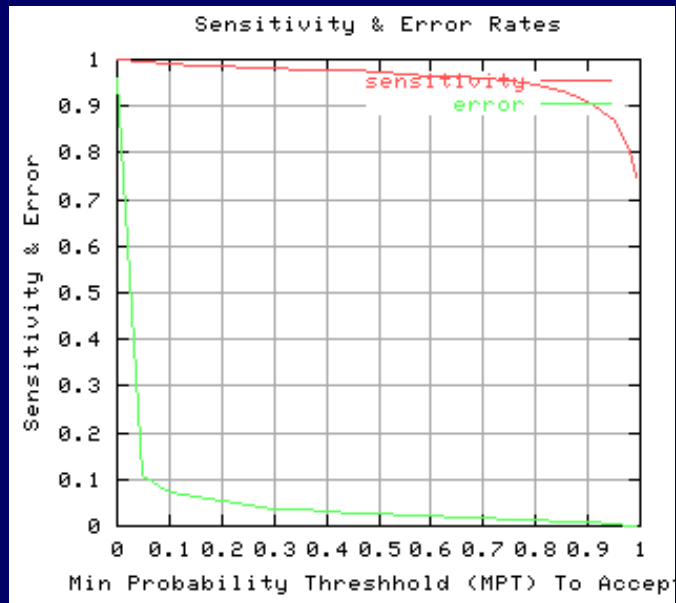
3+

1+ distr

2+ distr

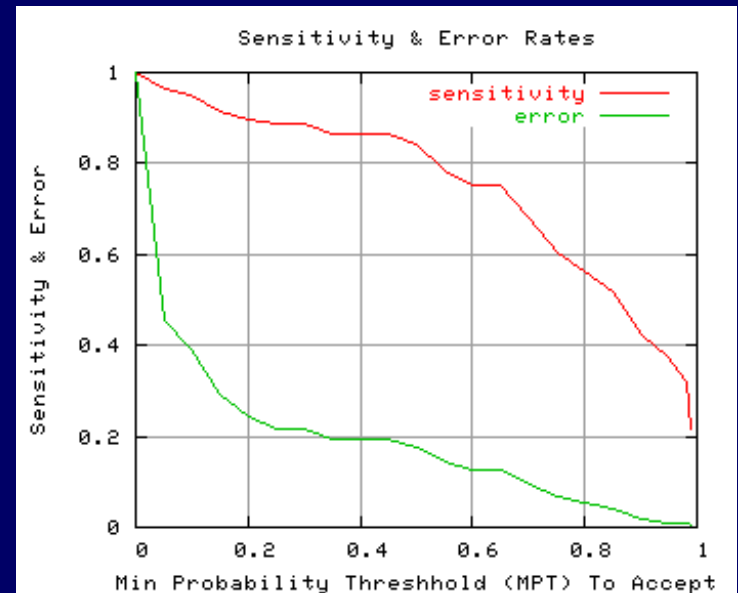
3+ distr

# PeptideProphet Results: Model Summary

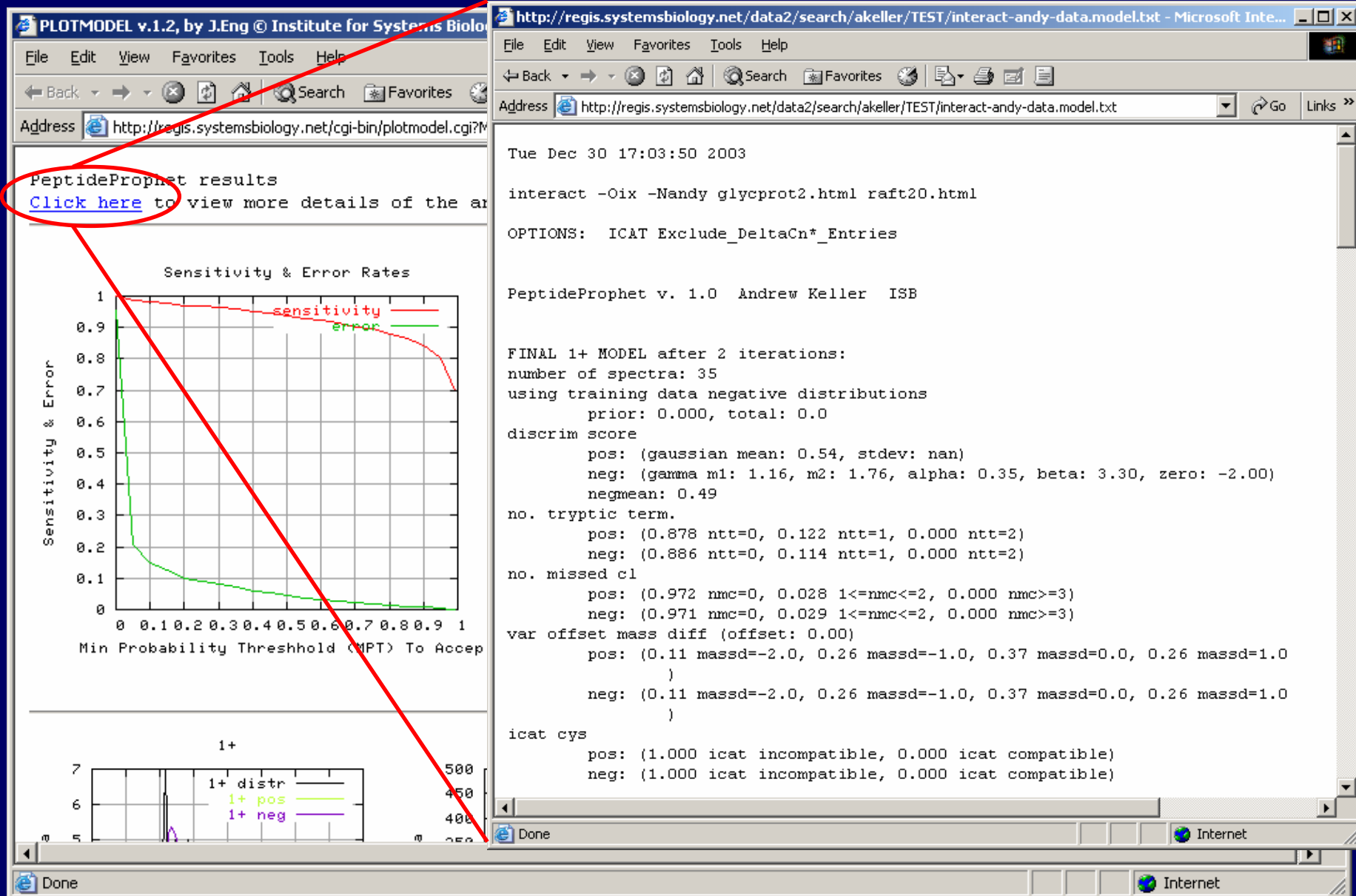


Good

Not so good



# PeptideProphet Results: Model Summary



# PeptideProphet Results: Model Summary

PeptideProphet results

[Click here](#) to view more details of the analysis

## Sensitivity & Error Rates

FINAL 2+ MODEL after 11 iterations:

number of spectra: 4518

using no. tolerable tryptic term. [ntt] 0 data as pseudonegatives

prior: 0.057, est. total no. correct: 258.2

SEQUEST discrim score [fval]

pos: (gaussian mean: 3.27, stdev: 1.36)

neg: (gamma m1: 4.78, m2: 23.54, alpha: 0.14, beta: 35.31, zero: -5.17)

negmean: -0.79

no. tolerable tryptic term. [ntt]

pos: (0.011 ntt=0, 0.136 ntt=1, 0.853 ntt=2)

neg: (0.752 ntt=0, 0.229 ntt=1, 0.019 ntt=2)

no. missed enz. cleavages [nmc]

pos: (0.775 nmc=0, 0.224 1<=nmc<=2, 0.001 nmc>=3)

neg: (0.415 nmc=0, 0.486 1<=nmc<=2, 0.099 nmc>=3)

var offset mass diff [massd] (offset: -0.60)

pos: (0.00 massd=-2.0, 0.04 massd=-1.0, 0.53 massd=0.0, 0.33 massd=1.0,  
0.05 massd=2.0, 0.04 massd=3.0, 0.00 massd=4.0)

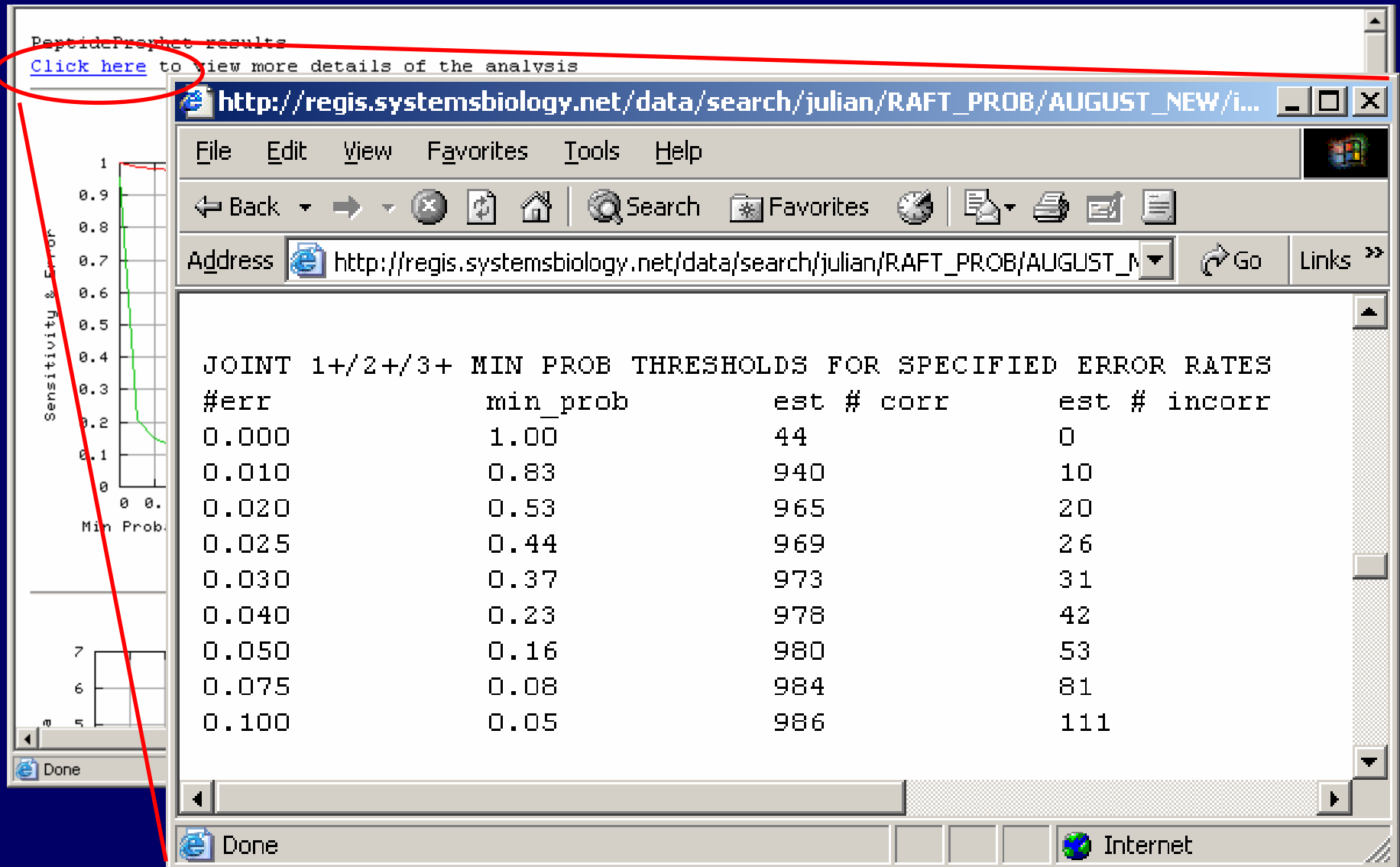
neg: (0.12 massd=-2.0, 0.13 massd=-1.0, 0.15 massd=0.0, 0.18 massd=1.0,  
0.20 massd=2.0, 0.22 massd=3.0, 0.01 massd=4.0)

icat cys [icat]

pos: (0.022 icat=0 (incompatible), 0.978 icat=1 (compatible))

neg: (0.927 icat=0 (incompatible), 0.073 icat=1 (compatible))

# PeptideProphet Results: Predicted Numbers of Correct and Incorrect Peptides



# PeptideProphet $[M+2H]^{2+}$ vs $[M+3H]^{3+}$ Precursor Ions

INTERACT by J.Eng, Institute for Systems Biology - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites

Address <http://regis-front/data2/search/akeller/COURSE/MARCELLO/interact.htm>

Sort/Restore: Sort by dta filename GO Help

FILE: /data2/search/akeller/COURSE/MARCELLO/interact- Tryptic: ☐ 1 ☐ 2 MaxMissed: 9 DelRows:

XCorr: ☐ +1 ☐ +2 ☐ +3 dCn: ☐ RSp: ☐ InclAA: ☐ MarkAA: ☐ NxS/T: ☐

Exclude charge: +1 ☐ +2 ☐ +3 ☐ Txt1: ☐ Txt2: ☐ J.Eng 04/2000

857.	<a href="#">0.9999</a>	860	<a href="#">./012403_08.1448.1448.3</a>	1995.0	(+2.5)	3.1945	<a href="#">0.243</a>	663.6	1
858.	<a href="#">0.9999</a>	861	<a href="#">./012403_08.1449.1449.2</a>	1995.0	(+0.9)	2.8027	0.189	447.7	1
859.	<a href="#">0.9991</a>	869	<a href="#">./012403_08.1455.1455.2</a>	1793.9	(+1.7)	2.8058	<a href="#">0.271</a>	443.6	1
860.	<a href="#">0.9965</a>	887	<a href="#">./012403_08.1470.1470.1</a>	1341.7	(+0.1)	2.0992	<a href="#">0.332</a>	218.6	1
861.	<a href="#">0.9809</a>	890	<a href="#">./012403_08.1473.1473.1</a>	1270.7	(+0.1)	2.1856	0.195	473.7	1
862.	<a href="#">0.9916</a>	891	<a href="#">./012403_08.1474.1474.2</a>	1341.7	(+1.1)	3.5879	<a href="#">0.455</a>	1623.8	1
863.	<a href="#">0.4887</a>	895	<a href="#">./012403_08.1478.1478.2</a>	1270.7	(+2.9)	3.3294	<a href="#">0.230</a>	966.9	2
864.	<a href="#">0.5113</a>	896	<a href="#">./012403_08.1478.1478.3</a>	1908.0	(+1.9)	3.6315	<a href="#">0.375</a>	660.1	2
865.	<a href="#">0.0876</a>	902	<a href="#">./012403_08.1483.1483.2</a>	1332.7	(-0.3)	1.6835	0.094	161.9	446
866.	<a href="#">0.9998</a>	905	<a href="#">./012403_08.1485.1485.2</a>	1995.0	(+0.5)	2.7644	<a href="#">0.281</a>	421.9	1
867.	<a href="#">0.5804</a>	909	<a href="#">./012403_08.1488.1488.2</a>	2017.0	(+1.0)	2.2738	0.090	455.6	1
868.	<a href="#">0.0742</a>	913	<a href="#">./012403_08.1490.1490.2</a>	1792.9	(+1.1)	3.0163	<a href="#">0.263*</a>	798.4	1
869.	<a href="#">0.6946</a>	930	<a href="#">./012403_08.1508.1508.2</a>	1389.7	(+1.1)	1.6639	0.170	194.0	73
870.	<a href="#">1.0000</a>	934	<a href="#">./012403_08.1511.1511.2</a>	1341.7	(+1.9)	3.7209	<a href="#">0.439</a>	924.2	2
871.	<a href="#">0.0122</a>	936	<a href="#">./012403_08.1512.1512.1</a>	1270.7	(+0.3)	1.7862	0.182	282.2	1

# PeptideProphet Results: Incomplete Analysis

INTERACT by J.Eng. Institute for Systems Biology - Mozilla {Build ID: 2002051319}

Sort/Restore: Sort by charge state  [Help](#)

FILE: /data2/search/akeller/PIG\_EXAMPLES/interact-example2-data Tryptic: ☐ 1 ☐ 2 MaxMissed: 9 DelRows: ☐

XCorr: ☐ +1 ☐ +2 ☐ +3 dCn: ☐ RSp: ☐ InclAA: ☐ MarkAA: ☐ NxS/T: ☐ Prob: ☐ XPRESS: ☐

Exclude charge: +1 ☐ +2 ☐ +3 ☐ Txt1:  Txt2:  J.Eng 04/2000

983.	0.0025	1698	./raft2021.4038.4038.2	1962.6	(+1.6)	1.1066	0.002	200.3	4	7/ 16	GP:P
984.	0.0004	1700	./raft2021.4042.4042.2	1587.8	(+1.1)	1.5229	0.112	332.3	2	10/ 24	GP:P
985.	0.0000	1702	./raft2021.4044.4044.2	1456.4	(+1.6)	1.2226	0.053	214.1	457	9/ 26	GP:U
986.	0.0000	1704	./raft2021.4046.4046.2	1476.7	(+1.1)	1.1634	0.028	294.7	130	9/ 22	SW:2
987.	0.0001	1706	./raft2021.4048.4048.2	1455.6	(-1.2)	1.4622	0.008	233.7	5	9/ 22	PIR2
988.	0.0000	1708	./raft2021.4050.4050.2	1464.5	(+1.4)	1.2318	0.107*	424.8	68	11/ 28	GP:P
989.	0.0010	1711	./raft2021.4066.4066.2	1447.5	(-1.3)	1.3330	0.130	286.9	38	10/ 26	PIR2
990.	0.0000	1714	./raft2021.4072.4072.2	1338.4	(+0.0)	0.9468	0.086	54.1	290	5/ 22	GP:P
991.	0.0000	1716	./raft2021.4076.4076.2	1958.9	(-0.5)	1.1636	0.063	226.5	42	9/ 28	GP:E
992.	0.0001	1718	./raft2021.4082.4082.2	1964.8	(-0.6)	1.4058	0.096	230.9	212	10/ 32	SW:N
993.	0.0000	1720	./raft2021.4088.4088.2	1553.8	(+0.5)	0.9884	0.002	161.5	30	7/ 26	GP:P
994.	0.0000	1722	./raft2021.4090.4090.2	1382.5	(+0.3)	1.0221	0.051	108.6	341	7/ 28	GP:P
995.	0.0000	1724	./raft2021.4092.4092.2	1456.6	(-1.9)	1.7071	0.156	355.8	162	11/ 28	GP:P
996.	0.0002	1726	./raft2021.4094.4094.2	1457.6	(+1.2)	1.4689	0.246	492.6	13	11/ 24	SW:E
997.	0.0000	1729	./raft2021.4102.4102.2	1477.6	(+1.4)	1.2024	0.023	323.3	40	9/ 22	GP:P
998.	0.0004	1731	./raft2021.4114.4114.2	1595.8	(-0.9)	1.1586	0.101	328.6	45	10/ 24	GP:P
999.	0.0001	1734	./raft2021.4118.4118.2	1339.6	(-1.0)	1.0846	0.122	106.9	1	6/ 20	GP:P
1000.	0.0000	1736	./raft2021.4124.4124.2	1959.4	(-1.3)	1.0980	0.115	113.5	182	7/ 36	GP:N
1001.	0.0000	1738	./raft2021.4126.4126.2	1964.2	(-1.2)	0.9374	0.070	144.0	53	9/ 38	SW:I
1002.	0.0000	1740	./raft2021.4146.4146.2	1457.3	(+0.6)	1.1638	0.110	267.4	195	7/ 16	GP:P
1003.	0.0000	1743	./raft2021.4152.4152.2	1482.7	(-0.4)	0.9447	0.009	102.8	379	6/ 24	GP:P
1004.	0.0000	1745	./raft2021.4162.4162.2	1320.6	(+1.8)	0.8804	0.011	52.1	207	4/ 18	SW:X
1005.	0.0000	1748	./raft2021.4170.4170.2	1962.2	(+1.8)	0.9944	0.046	86.1	369	6/ 30	PIR2
1006.	0.0000	1750	./raft2021.4172.4172.2	1310.6	(+0.1)	1.0680	0.072	63.8	19	5/ 22	GP:P
1007.	0.0000	1752	./raft2021.4176.4176.2	1957.2	(+0.7)	1.0973	0.125	92.3	94	7/ 36	SW:F
1008.	0	46	./raft2021.0760.0760.3	1614.0	(-1.3)	1.0205	0.025	83.9	290	9/ 52	SW:M
1009.	-3	165	./raft2021.1030.1030.3	3800.5	(-0.5)	4.5183	0.320	1105.8	1	37/132	SW:M
1010.	-3	503	./raft2021.1424.1424.3	2144.2	(+1.0)	4.1322	0.079	1630.8	11	28/ 60	SW:F
1011.	-3	709	./raft2021.1668.1668.3	2291.3	(+1.3)	5.3741	0.269	2079.7	1	30/ 68	GP:P
1012.	0	1622	./raft2021.3844.3844.3	1987.0	(+1.7)	1.0929	0.017	45.3	300	7/ 56	GP:P
1013.	0	1660	./raft2021.3942.3942.3	2737.9	(-1.3)	1.0746	0.074	112.1	5	10/ 76	GP:U

Document: Done (6.436 secs)

# Sort Data by Computed Probability

INTERACT by J.Eng, Institute for Systems Biology - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites

Address <http://regis-front/data2/search/akeller/COURSE/MARCELLO/interact.htm>

Sort/Restore: Apply filtering below ... no sort or restore  [Help](#)

FILE: /data2/see

XCorr: ☐ +1

Exclude charge

MaxMissed: 9 DelRows:

MarkAA: NxS/T:

J.Eng 04/2000

- Apply filtering below ... no sort or restore
- Restore original dataset (RestoreOrig)
- Restore previous dataset (LastUndo)
- Sort by protein
- Sort by protein but exclude single hits
- Sort by protein and only keep single hits
- Sort by peptide
- Sort by probability score**
- Sort by XCorr
- Sort by dCn
- Sort by dta filename

<a href="#">1.0000</a>	268	<a href="#">0.193</a>	1681.6	1	<a href="#">19/ 24</a>
<a href="#">0.1496</a>	607	<a href="#">0.038</a>	242.6	1	<a href="#">11/ 18</a>
<a href="#">0.0511</a>	1637	<a href="#">0.103</a>	302.7	1	<a href="#">13/ 26</a>
<a href="#">0.8367</a>	2499	<a href="#">0.038</a>	172.2	162	<a href="#">11/ 32</a>
<a href="#">0.9981</a>	677	<a href="#">0.152</a>	397.8	3	<a href="#">8/ 10</a>
<a href="#">0.1550</a>	445	<a href="#">0.030</a>	201.7	4	<a href="#">12/ 22</a>
<a href="#">0.8900</a>	302	<a href="#">0.036</a>	294.8	20	<a href="#">9/ 12</a>
<a href="#">0.9683</a>	1682	<a href="#">0.072</a>	409.0	13	<a href="#">9/ 14</a>
<a href="#">0.9946</a>	807	<a href="#">0.097</a>	674.3	1	<a href="#">11/ 12</a>
<a href="#">0.9972</a>	2493	<a href="#">0.259</a>	82.4	112	<a href="#">10/ 20</a>
<a href="#">0.9802</a>	2308	<a href="#">0.049</a>	173.5	1	<a href="#">11/ 22</a>
<a href="#">0.1766</a>	2847	<a href="#">0.029</a>	301.7	3	<a href="#">20/ 68</a>
<a href="#">0.0522</a>	754	<a href="#">0.194</a>	230.0	12	<a href="#">9/ 14</a>



# Some Options for Interact

Rename Output File (e.g. to `interact-noicat.htm`):

The screenshot shows the 'Specify Parameters' section of the Interact web interface. On the left is a navigation menu with links like 'Local Home', 'Class Exercises', 'Tool Interfaces', 'Intro page', 'RAW to mzXML', 'Sequest', 'Interact/PeptideProphet', 'ASAPRatio', 'ProteinProphet', 'My Session [OFF]', 'Start session', 'Resources', 'SPC', 'SPC Tools', 'Sashimi', 'PeptideProphet', and 'ProteinProphet'. The main content area is titled 'Specify Parameters' and contains the instruction: 'Please specify the Interact and (optional) PeptideProphet parameters. Then click on process the files.' Below this is a section '1. Specify Input/Output Parameters' with an 'Add File' input field, a 'Browse...' button, and an 'Update' button. The 'File List' shows a file path: 'C:\Inetpub\wwwroot\class\day2\Unconstrained\haloICAT2\_30.html'. The 'Location for output files' is set to 'C:\Inetpub\wwwroot\class\day2\Unconstrained'. The 'Rename output files' field is circled in red and contains the text 'noicat'. Below this is a section '2. Specify Interact Parameters' with a red warning: 'WARNING: Changing these options if PeptideProphet is also run may invalidate the results'.

**Specify Parameters**

Please specify the Interact and (optional) PeptideProphet parameters. Then click on process the files.

**1. Specify Input/Output Parameters**

Add File:

File List:

C:\Inetpub\wwwroot\class\day2\Unconstrained\haloICAT2\_30.html

Location for output files:

Rename output files:

**2. Specify Interact Parameters**

**WARNING:** Changing these options if PeptideProphet is also run may invalidate the results

# Some Options for Interact

- no PeptideProphet analysis
- alternative minimum probability
- sample enzyme other than trypsin

ISB Interact/PeptideProphet - Web Interface - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites Media

Address [http://localhost/isb-bin/interact\\_web.pl](http://localhost/isb-bin/interact_web.pl)

### 3. Specify PeptideProphet Parameters

Include PeptideProphet in the Analysis: ☒

Probability Threshold:  (Minimum PeptideProphet probability of results to display.)

Enzyme Options:

- ☒ Trypsin Enzyme (default)
- ☐ Chymotrypsin Enzyme
- ☐ AspN Enzyme
- ☐ Triple Mix (Chymotrypsin, AspN, Trypsin)
- ☐ GluC Enzyme
- ☐ GluC Bicarb Enzyme
- ☐ Nonspecific Cleavage
- ☐ CNBr Cleavage

# Use of Supplemental Discriminating Information

Use additional discriminating information, including ICAT or N-glyc, when relevant

*PeptideProphet automatically uses ICAT information when it thinks appropriate. Nevertheless, you can explicitly set whether or not ICAT information is utilized*

Use N-glycosylation motif information for computing probabilities ☐

**4. Advanced PeptideProphet Parameters**

**WARNING: These are advanced PeptideProphet options. You might want to leave these unchanged unless you are an expert user.**

DeltaCn\* Options: ☒ Set DeltaCn\* values to 0 (default)  
☐ Leave DeltaCn\* values alone (do not set to 0)  
☐ Exclude DeltaCn\* entries (set Probability to 0)

Autodetect ICAT cysteine information for computing probabilities ☐  
Ignore ICAT cysteine information for computing probabilities ☐  
Use ICAT cysteine information for computing probabilities ☒

The screenshot shows the PeptideProphet parameter interface. A red circle highlights the checkbox for 'Use N-glycosylation motif information for computing probabilities'. Another red circle highlights the 'Autodetect ICAT cysteine information for computing probabilities' section, which includes three radio button options: 'Autodetect ICAT cysteine information for computing probabilities' (unselected), 'Ignore ICAT cysteine information for computing probabilities' (unselected), and 'Use ICAT cysteine information for computing probabilities' (selected).

# DeltaCn\* Example

Exclude charge: +1 <input type="checkbox"/> +2 <input type="checkbox"/> +3 <input type="checkbox"/> Txt1: <input type="text"/> Txt2: <input type="text"/> J.Eng 04/2000									
<a href="#">0.9999</a>	91	<a href="#">./sergei digest A full 01.0607.0609.2</a>	1983.0	(-0.4)	4.6602	<a href="#">0.358</a>	3224.6	1	
<a href="#">0.9994</a>	625	<a href="#">./sergei digest A full 01.1523.1525.3</a>	2046.3	(-0.7)	4.6539	<a href="#">0.307</a>	1901.4	1	
<a href="#">1.0000</a>	117	<a href="#">./sergei digest A full 01.0663.0665.3</a>	2100.2	(+0.0)	4.5901	<a href="#">0.286</a>	1627.3	1	
<a href="#">1.0000</a>	712	<a href="#">./sergei digest A full 01.1675.1681.2</a>	1832.0	(-0.3)	4.5772	<a href="#">0.393</a>	1148.8	1	
<a href="#">1.0000</a>	651	<a href="#">./sergei digest A full 01.1569.1571.3</a>	3106.3	(+0.4)	4.5737	<a href="#">0.362</a>	537.5	1	
<a href="#">0.9981</a>	942	<a href="#">./sergei digest A full 01.2085.2089.3</a>	2314.7	(+1.4)	4.5707	<a href="#">0.299</a>	1348.2	1	
<a href="#">0.9997</a>	1045	<a href="#">./sergei digest A full 01.2277.2279.3</a>	2804.0	(-0.5)	4.5012	<a href="#">0.352</a>	1236.7	1	
<a href="#">0.8970</a>	861	<a href="#">./sergei digest A full 01.1951.1953.2</a>	2314.7	(-0.4)	4.4874	<a href="#">0.348</a>	482.4	1	
<a href="#">0.9810</a>	626	<a href="#">./sergei digest A full 01.1529.1531.2</a>	1480.7	(-0.5)	4.4192	<a href="#">0.315</a>	1731.6	1	
<a href="#">1.0000</a>	176	<a href="#">./sergei digest A full 01.0767.0767.2</a>	1983.0	(+0.7)	4.4109	<a href="#">0.414</a>	1966.2	1	
<a href="#">0.9998</a>	328	<a href="#">./sergei digest A full 01.1009.1011.3</a>	1440.7	(+0.8)	4.4102	<a href="#">0.302</a>	2269.3	1	
<a href="#">0.9999</a>	1150	<a href="#">./sergei digest A full 01.2487.2489.2</a>	2219.5	(-0.5)	4.3932	<a href="#">0.334</a>	1350.2	1	
<a href="#">0.9505</a>	940	<a href="#">./sergei digest A full 01.2083.2087.3</a>	2508.8	(-0.3)	4.3518	<a href="#">0.375</a>	1166.5	1	
<a href="#">0.9900</a>	340	<a href="#">./sergei digest A full 01.1027.1029.3</a>	1640.9	(-0.4)	4.3039	<a href="#">0.368*</a>	1451.7	1	
<a href="#">0.9770</a>	766	<a href="#">./sergei digest A full 01.1783.1785.3</a>	2119.5	(+1.2)	4.2653	<a href="#">0.356*</a>	1147.4	2	
<a href="#">0.9982</a>	863	<a href="#">./sergei digest A full 01.1955.1957.2</a>	2052.4	(+0.4)	4.2588	<a href="#">0.382</a>	453.4	1	
<a href="#">0.9996</a>	738	<a href="#">./sergei digest A full 01.1727.1729.2</a>	1531.8	(-0.1)	4.2559	<a href="#">0.427</a>	1676.8	1	
<a href="#">0.9999</a>	168	<a href="#">./sergei digest A full 01.0753.0753.2</a>	1983.0	(+0.8)	4.2504	<a href="#">0.398</a>	2025.6	1	

#	Rank/Sp	(M+H)+	deltCn	XCorr	Sp	Ions	Reference	Peptide
1.	1 / 2	2119.4870	0.0000	4.2653	1147.4	<a href="#">30/72</a>	<a href="#">SW:PHS3_HUMAN</a>	+2 K. <a href="#">VAIQLNDTHPALSIPELMR</a> .I
2.	2 / 1	2119.4870	0.0283	4.1445	1150.6	<a href="#">29/72</a>	<a href="#">sp P00489 PHS2_RABIT</a>	+2 K. <a href="#">VAIQLNDTHPSLAPELMR</a> .V
3.	3 / 158	2119.4894	0.3564	2.7454	377.2	<a href="#">22/76</a>	<a href="#">SWN:PRKD_HUMAN</a>	+4 A. <a href="#">VPSAGRLRLFALHASQFSTCL</a> .L
4.	4 / 461	2119.5767	0.3948	2.5813	318.8	<a href="#">19/68</a>	<a href="#">GP:AY027526_1</a>	S. <a href="#">VAKLLHPQLTCRLLELRD</a> .I

# DeltaCn\* Options

There are three ways asterisked deltas can be treated by PeptideProphet:

1. Penalize (the default option, sets asterisked deltas to 0)
2. Leave alone (suitable for the context of homologues)
3. Exclude (the most conservative, assigns probability 0)

## 4. Advanced PeptideProphet Parameters

**WARNING:** These are advanced PeptideProphet options. You might want to leave these unchanged unless you are an expert user.

- DeltaCn\* Options:
- ☒ Set DeltaCn\* values to 0 (default)
  - ☐ Leave DeltaCn\* values alone (do not set to 0)
  - ☐ Exclude DeltaCn\* entries (set Probability to 0)

# Ongoing Developments for PeptideProphet

---

Optimize for various additional mass spectrometers

New discriminant function

Adapt to additional methods for assigning peptides to tandem mass spectra

Mascot ✓

COMET ✓

X!Tandem

Others

# Pep3D mzXML Data Viewer

Xiao-jun Li

1. Main features of Pep3D
2. Evaluating sample quality
3. Evaluating LC-ESI-MS/MS performance

# 1. Main Features of Pep3D

## Pep3D Image for LC-ESI-MS Data

This program displays LC-ESI-MS data stored in mzXML format in a Pep3D image.  
Developed by Dr. Xiao-jun Li at [Institute for Systems Biology](#).  
Data: October 8, 2002

Specify parameters here:

- Full path of [interact-data.htm](#) or [.mzXML](#) file:  
Example: /data2/search/xiaojun/ASAPRatio\_demo/interact-1vs1-data.htm  
Or: /data2/search/dan/17\_mix\_folder/\*.mzXML

Data in "interact-data.htm": ☒ original ☐ filtered

- M/Z range:
- Elution time range (in min):
- Intensity range:
- Display peptides:
- Score type:
- Mapping function:
- Image type:

Generate Pep3D image

Save as

Pep3D.htm

interact

Sort/Restore:

FILE:

XCorr: ☐ +1 ☐ +2 ☐ +3 dCn: ☐ RSp:

Exclude charge: +1 ☐ +2 ☐ +3 ☐ Txt1: ☐

Generate Pep3D image

Save as

Pep3D.htm

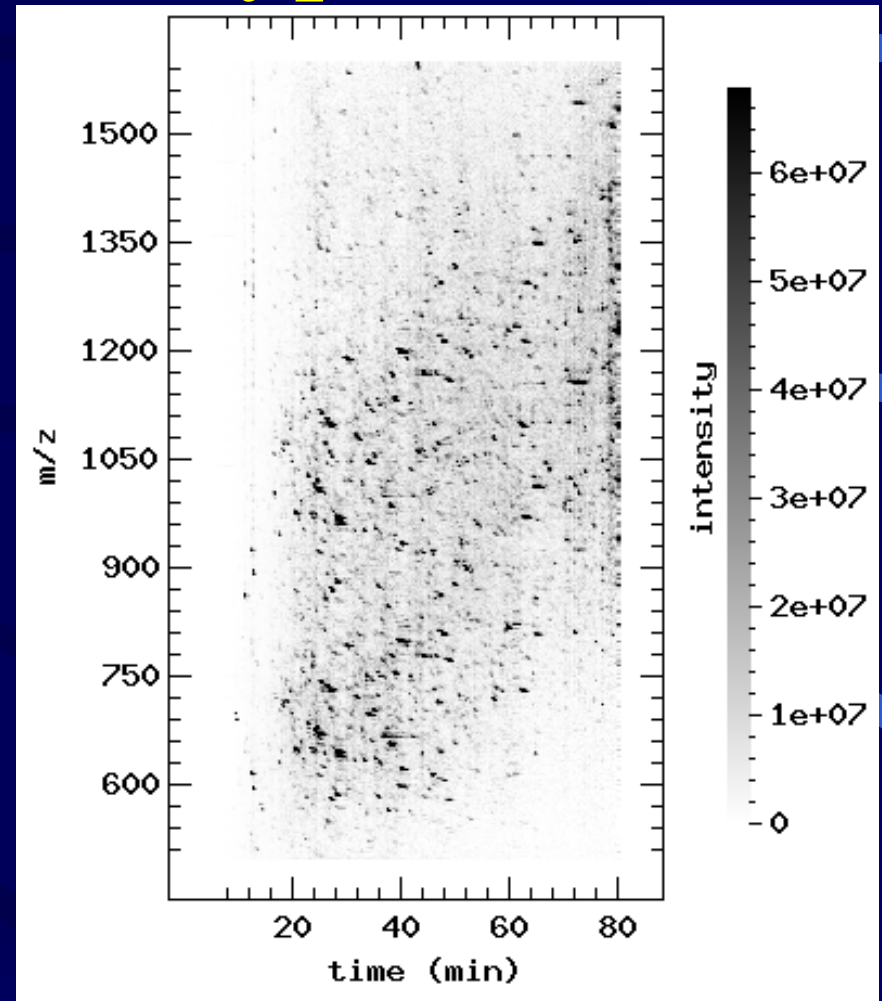
<a href="#">0.1597</a>	168	<a href="#">./p200.0774.0774.2</a>	930.0 (-2.1)
<a href="#">0.7852</a>	178	<a href="#">./p40.0803.0803.2</a>	1403.5 (+0.4)

- Intensity range:
- Display peptides:
- Score type:
- Mapping function:
- Image type:

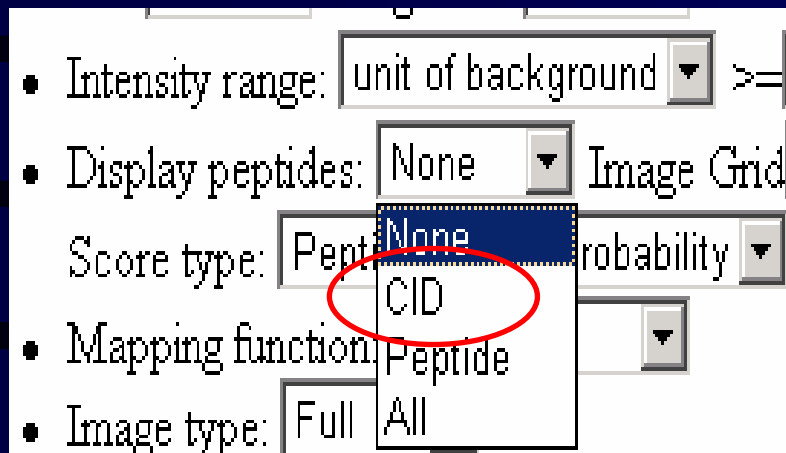


# Pep3D Images: Type A

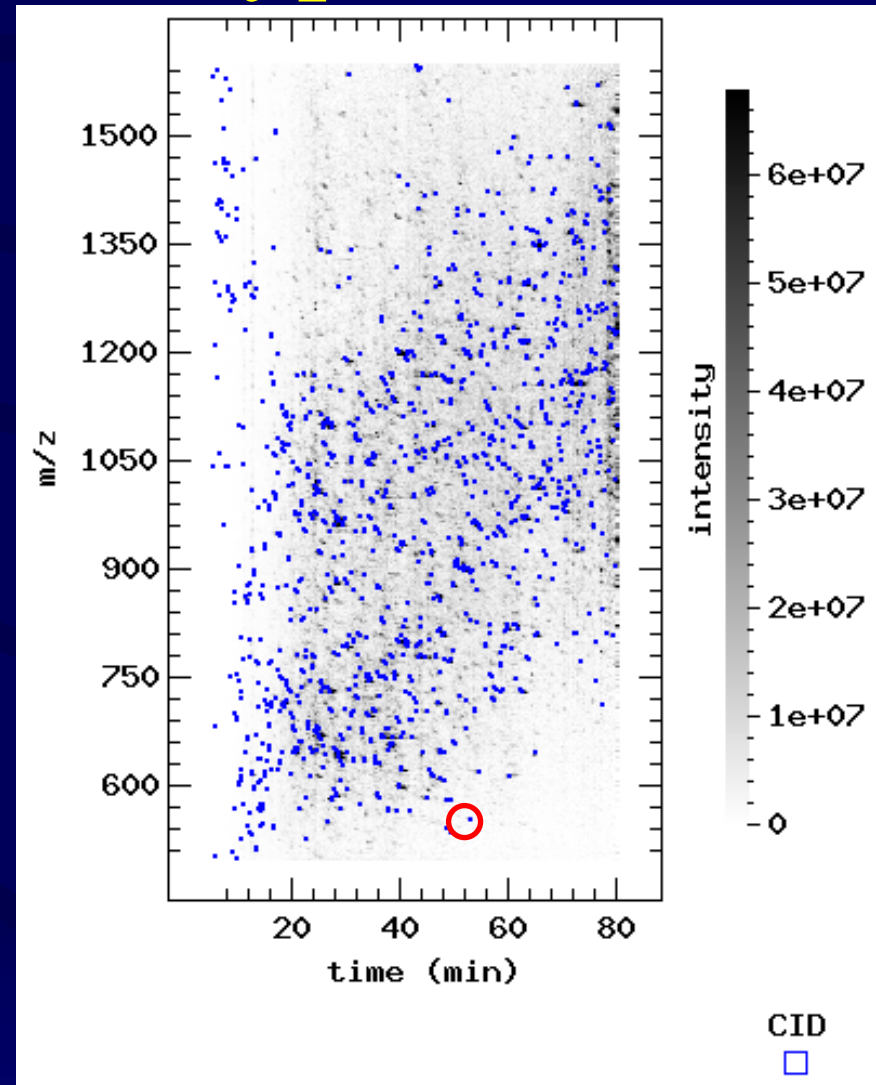
- Intensity range: unit of background ▾ >=
- Display peptides: None ▾ Image Grid
- Score type: Peptide ▾ Probability ▾
- Mapping function: Peptide ▾
- Image type: Full ▾ All



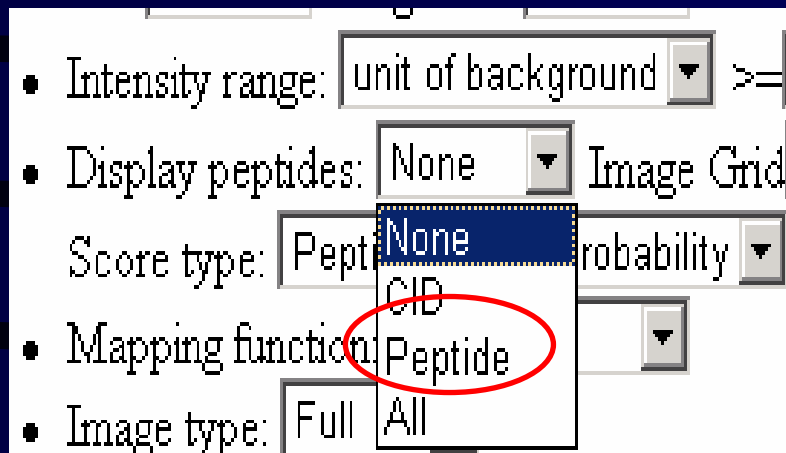
# Pep3D Images: Type B



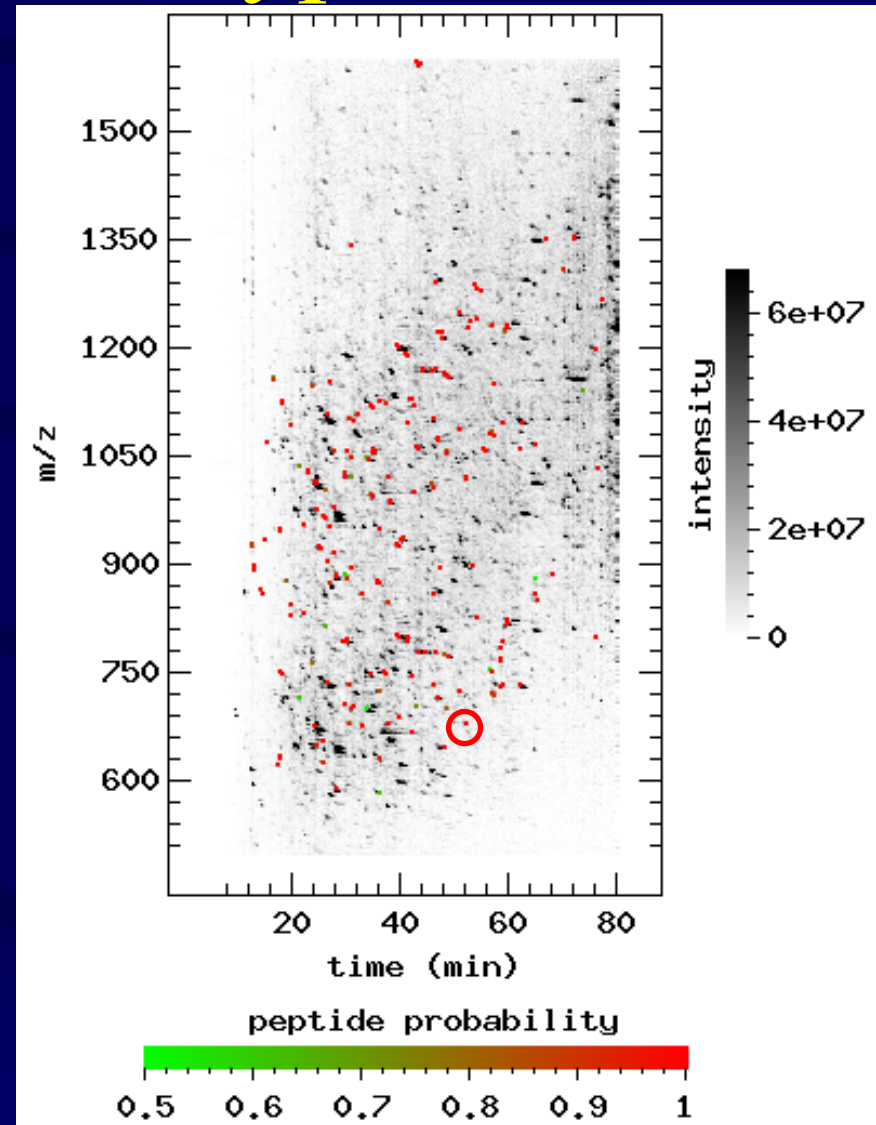
CID (1198)



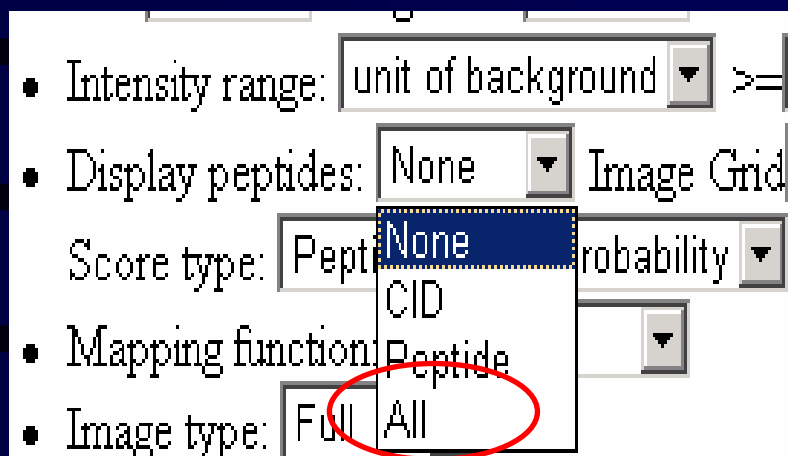
# Pep3D Images: Type C



peptide (221)

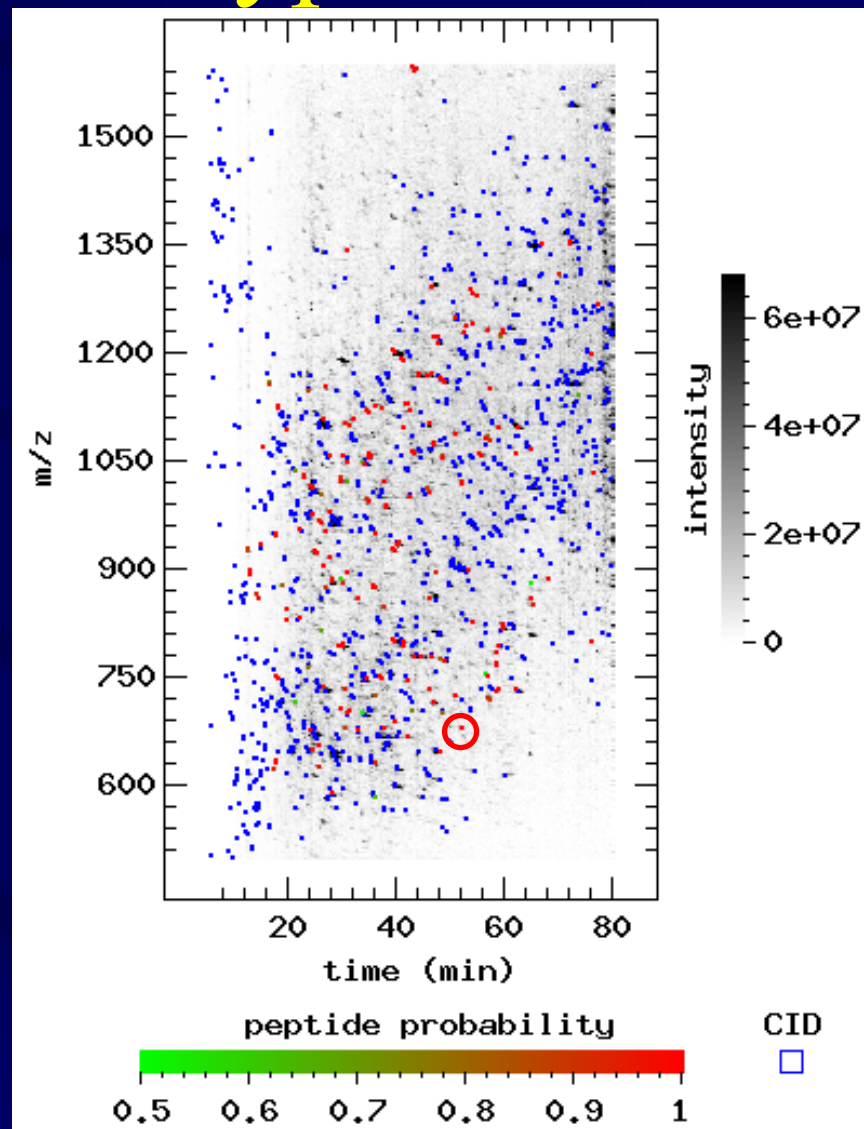


# Pep3D Images: Type D

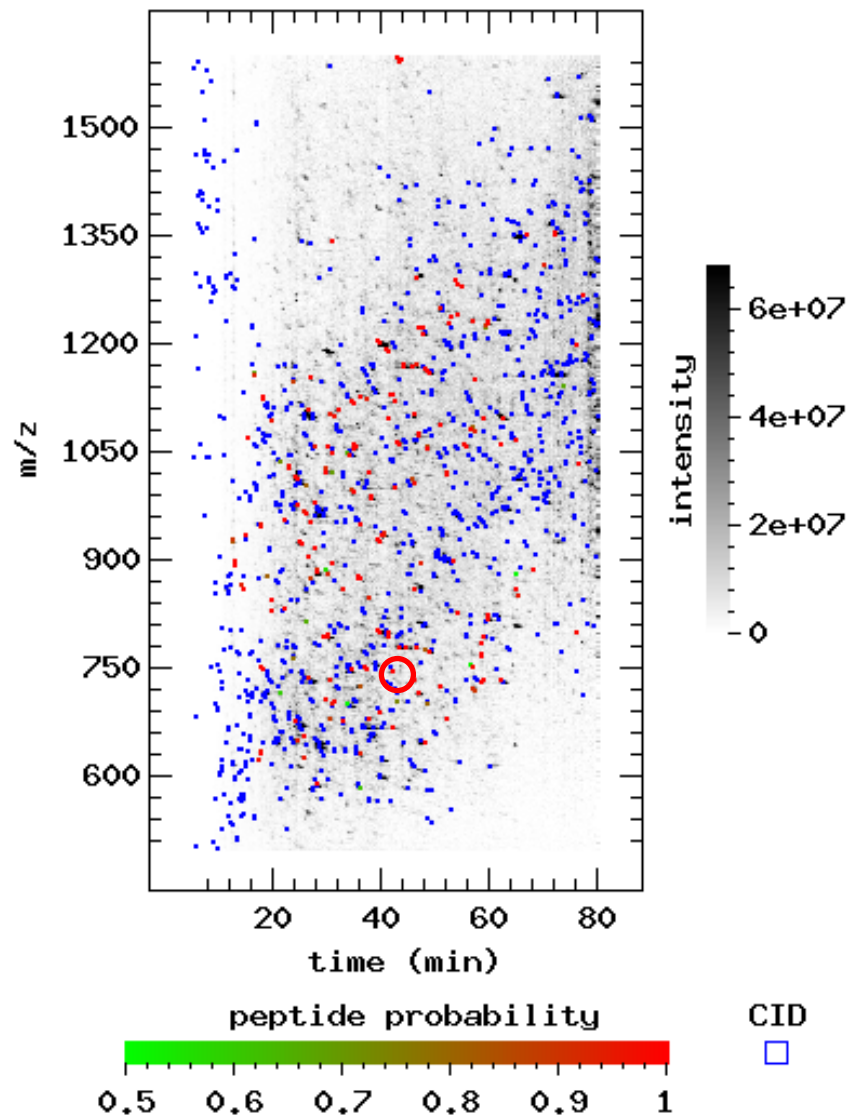


peptide/CID

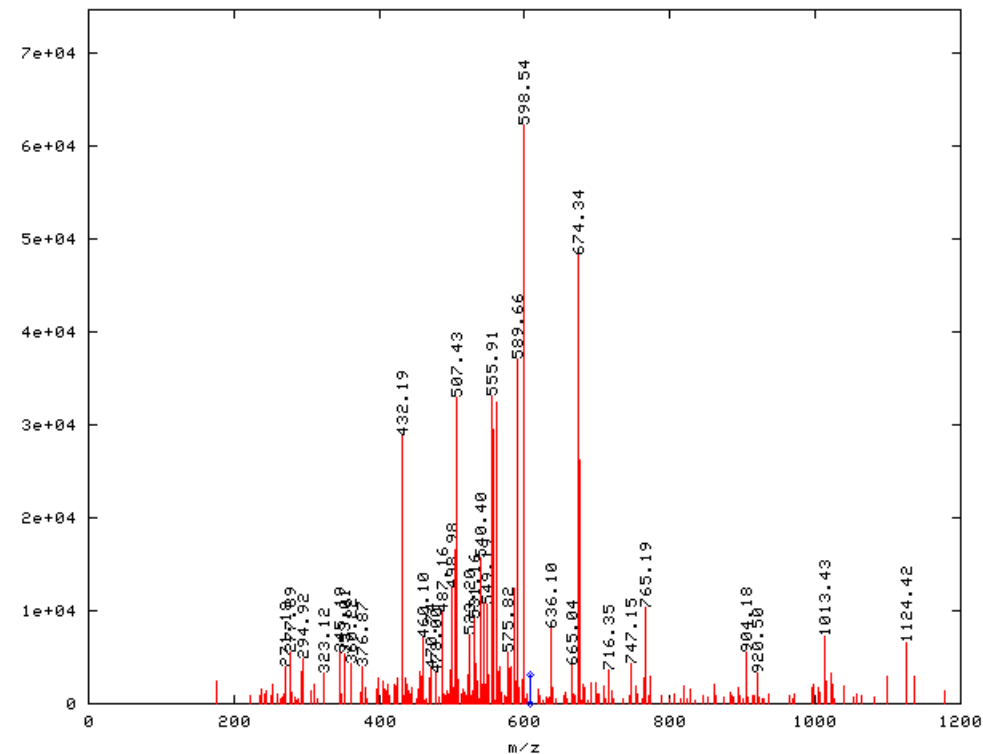
(221/1196 = 0.184783)



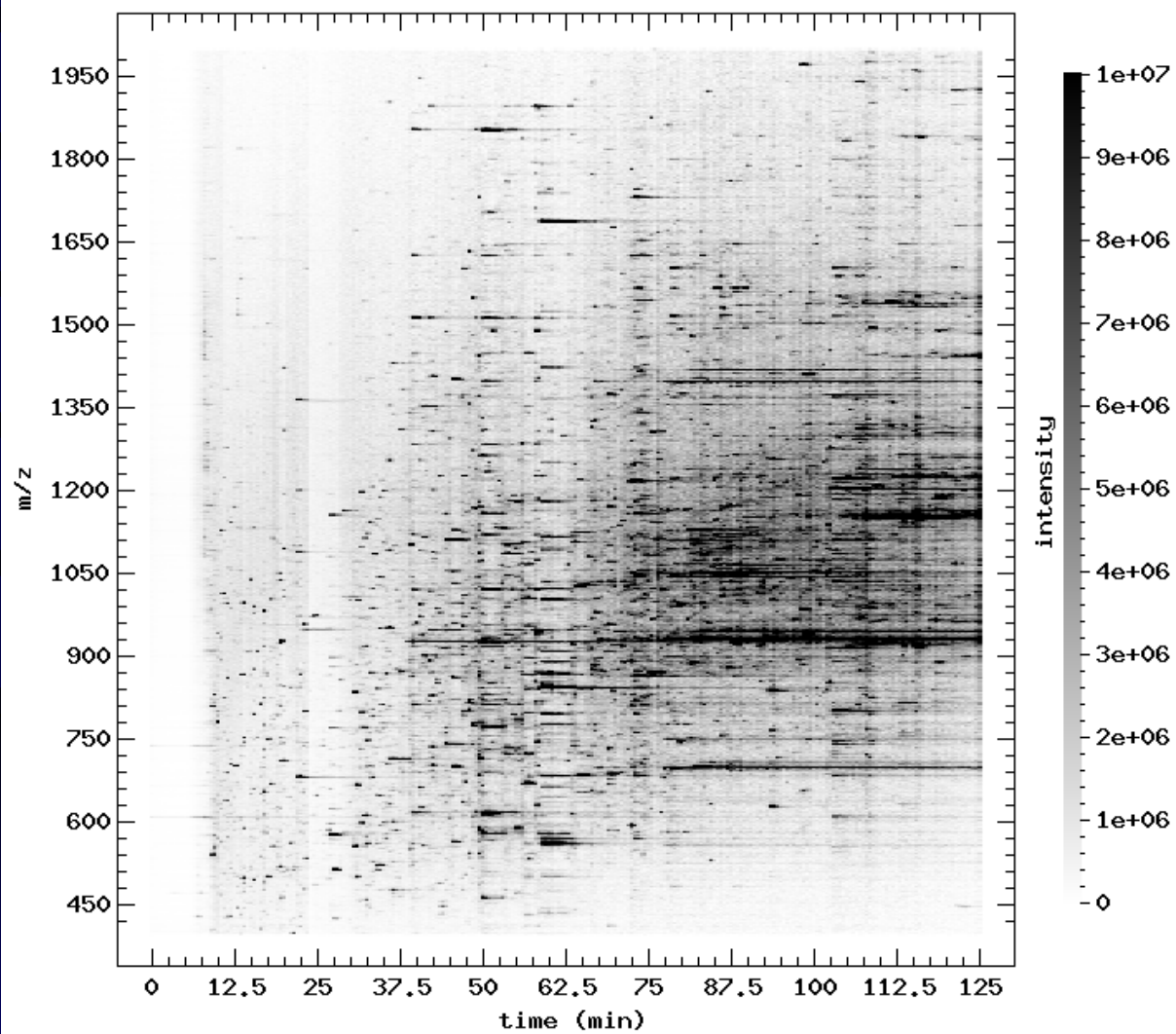
# Display CID Spectrum



- scan number: 783  
m/z: 607.578003  
time: 40.491000
- ID:  
[0.9157](#) 155 [/p40.0783.0783.2](#) 1214.3 (-0.1) 1.6765 0.144 244.8 64 [8/ 16](#) [IP100289535](#) R. [VTC](#)  
[0.713](#) (56.292%) [0.845580](#) 1

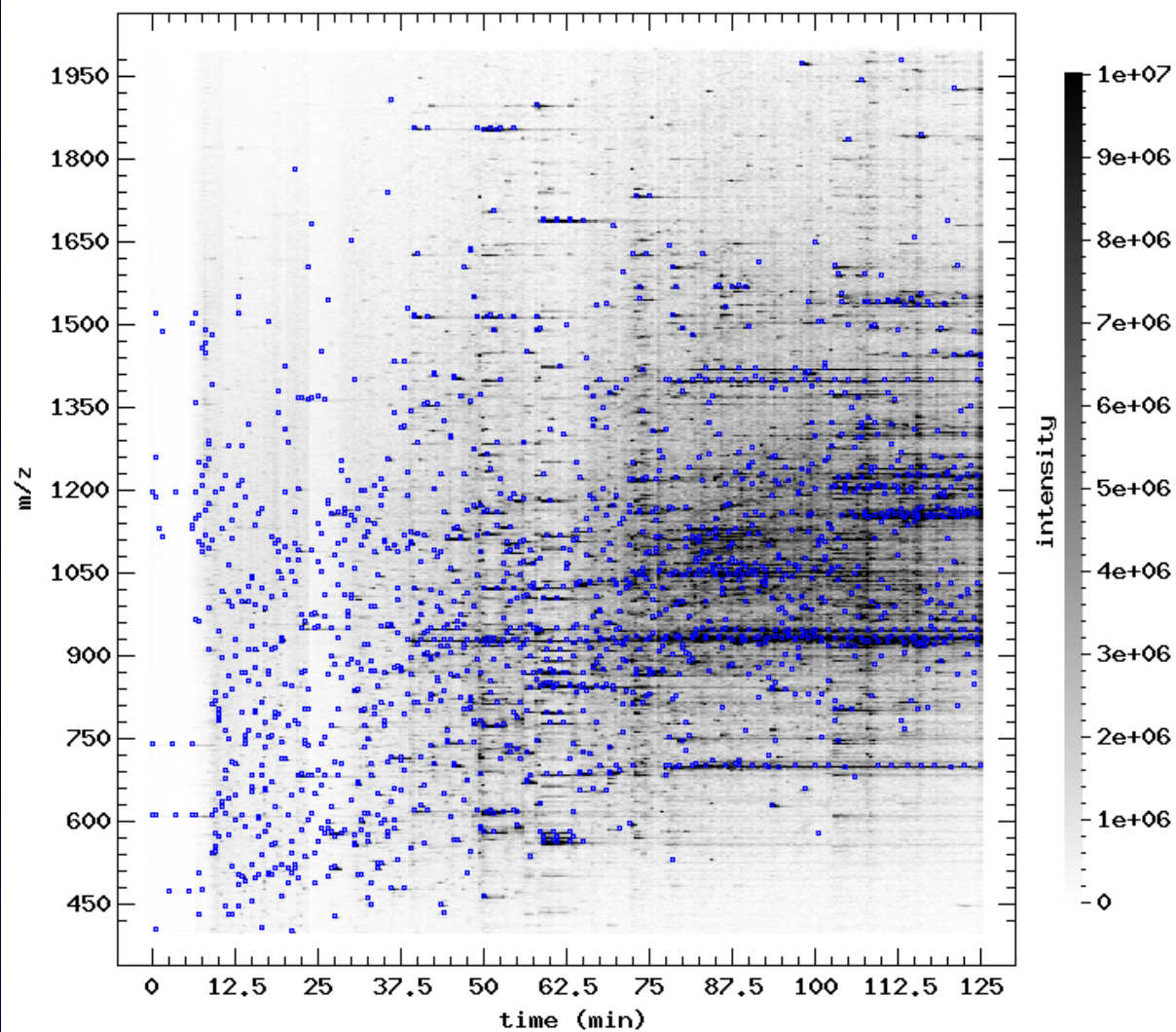


Export to dta | charge: ☐ +1 ☐ +2 ☐ +3 ☐ +4



Features: 2720



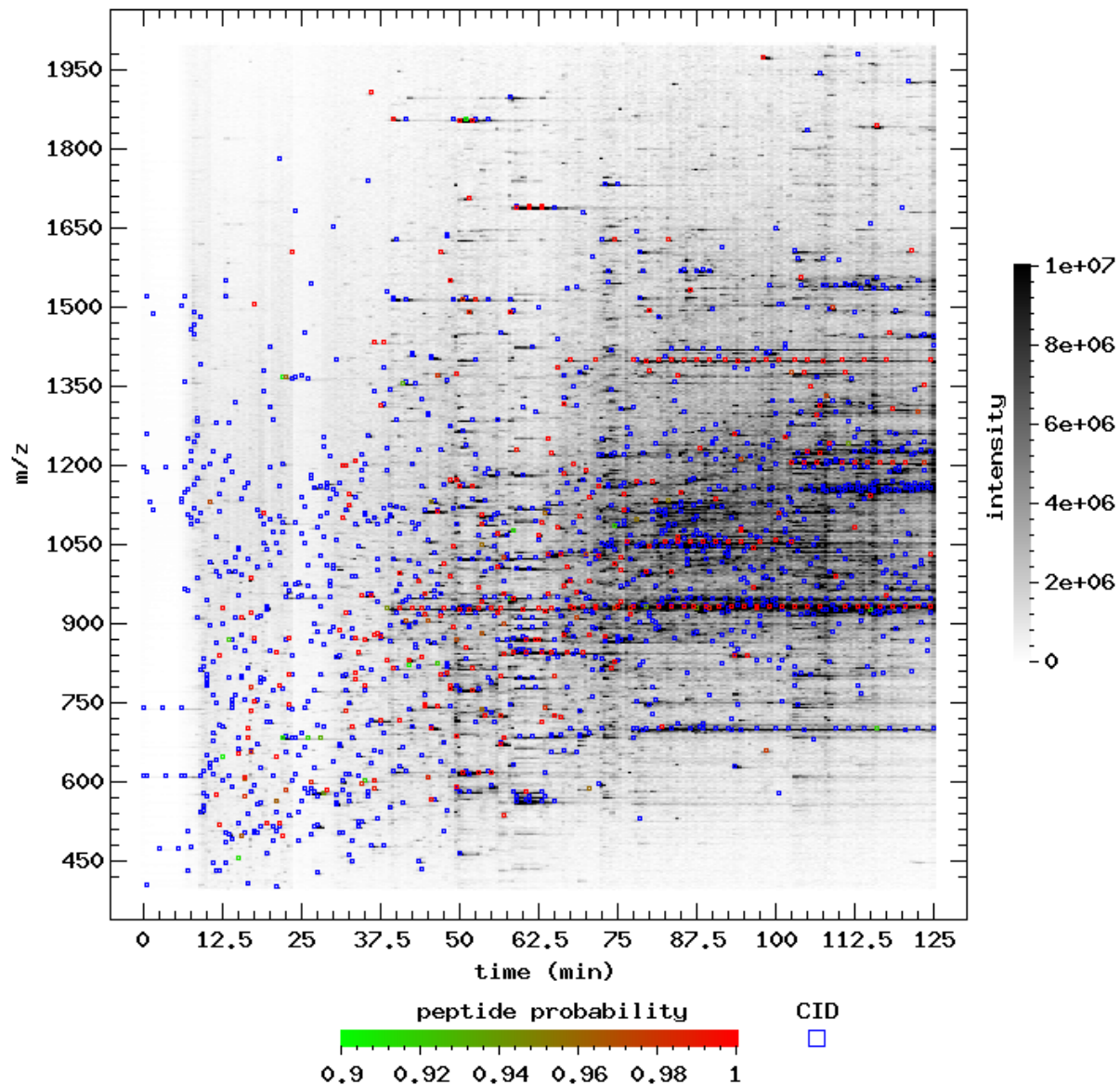


CID



Features: 2720

CIDs: 1633



Features: 2720

CIDs: 1633

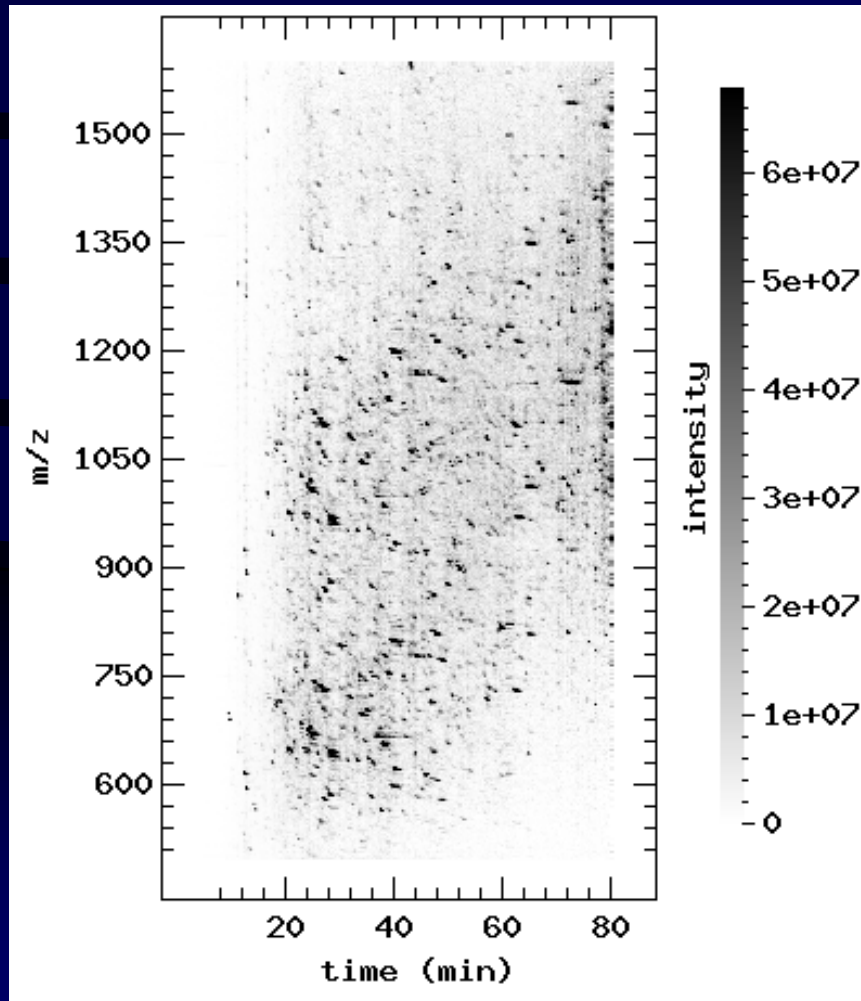
IDs: 363

ID/CID: 22%

ID/feature: 13%



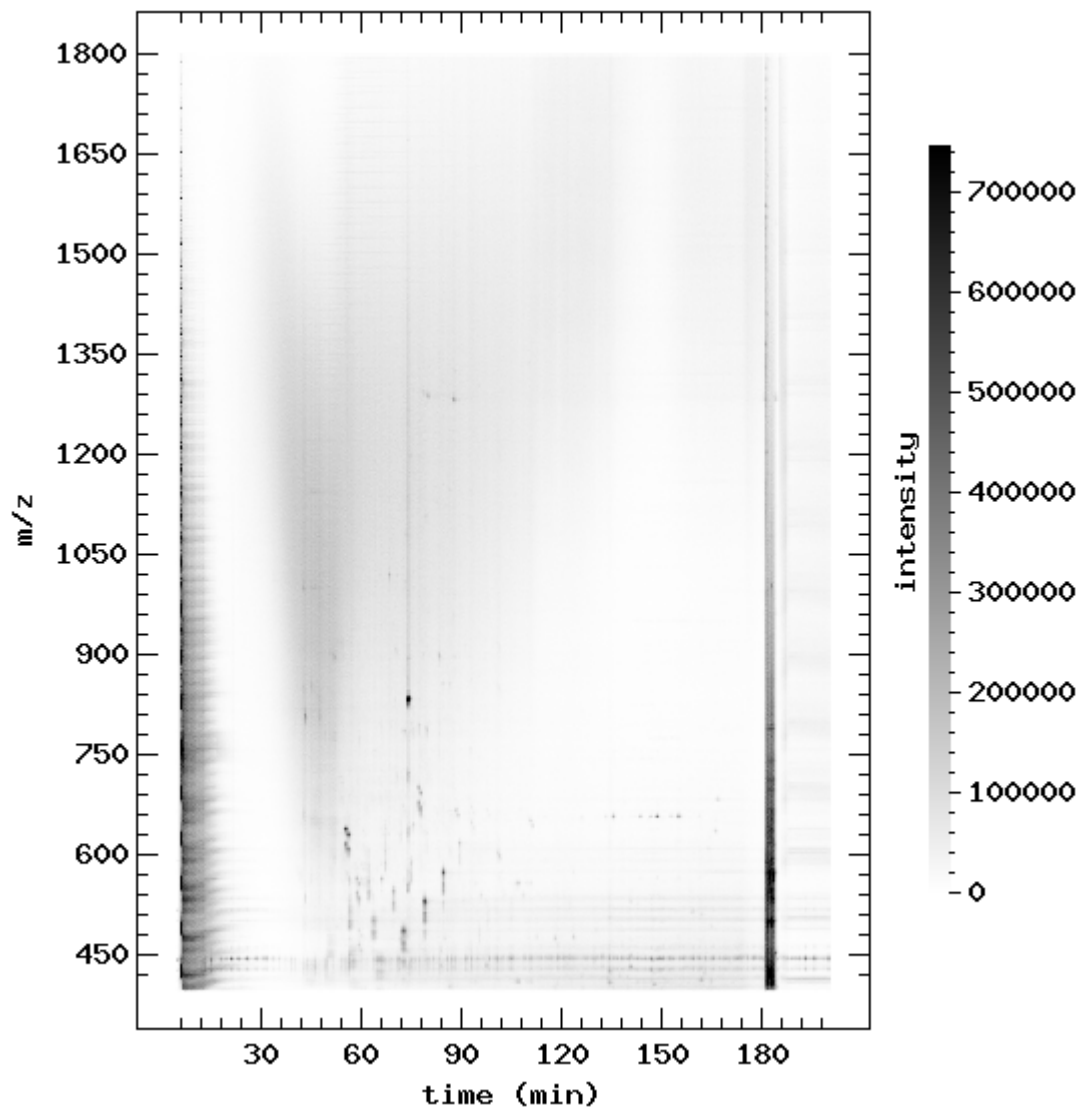
## 2. Evaluating Sample Quality



Good sample:

Plenty well-localized spots  
without any particular  
large-scale pattern

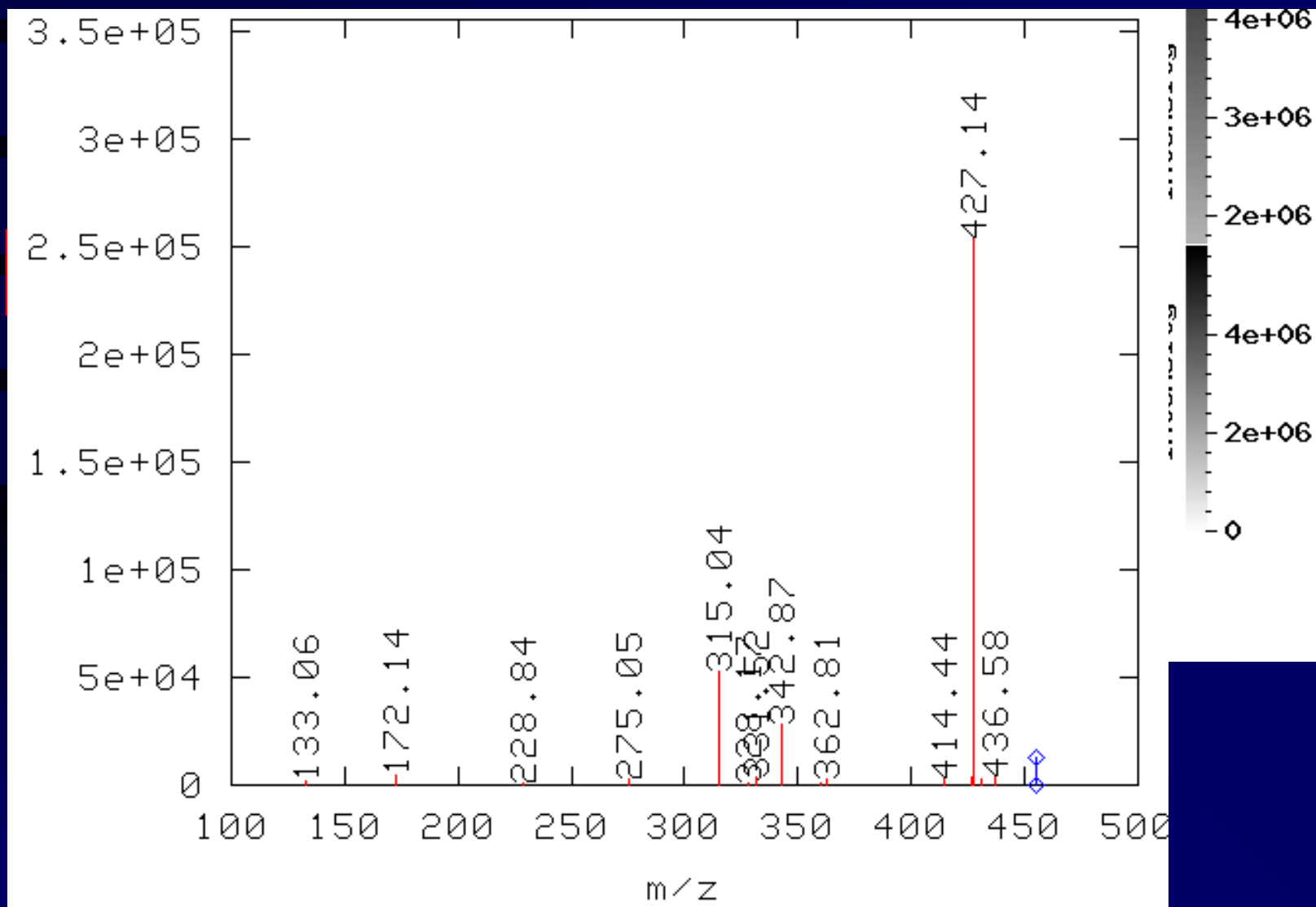
# Empty Sample



peptide/CID (0/311 = 0)

- Very few localized spots
- Mainly background noise
- Distinguishable from no-spray

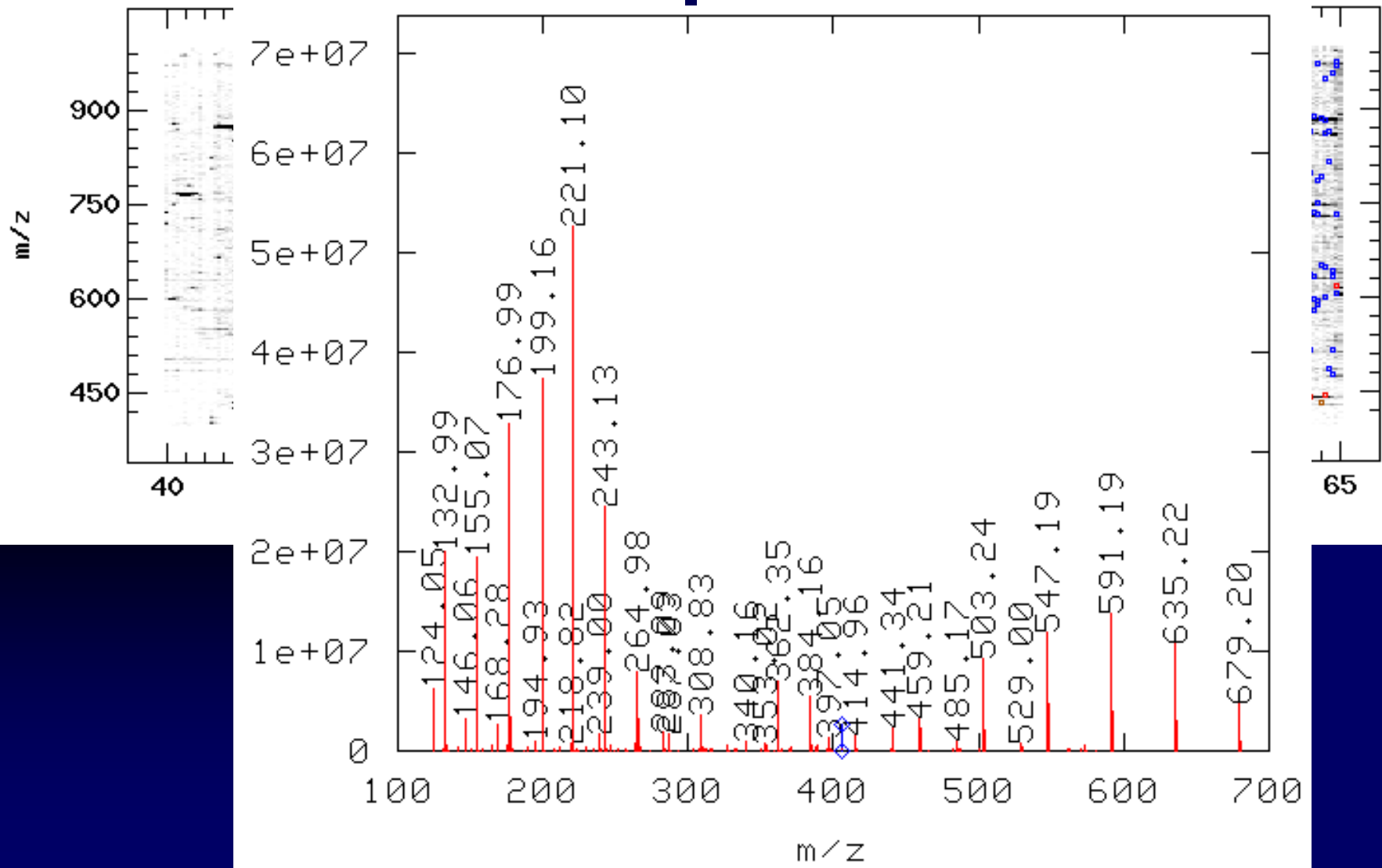
# Chemical Contamination



# Features of Chemical Contamination

- Long horizontal streaks
- Low  $m/z$  values
- Singly charged ions
- Many wasteful CID attempts
- Can be put on CID exclusive list

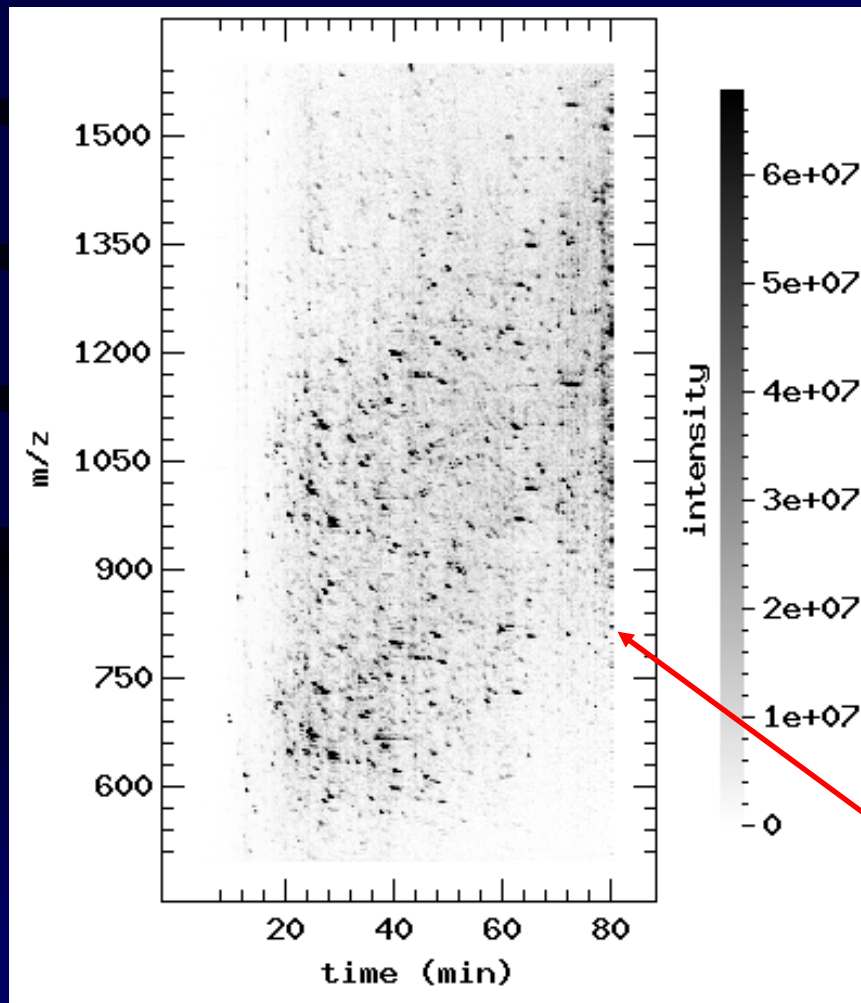
# Polymer Contamination



# Features of Polymer Contamination

- Localized spots running off-diagonally
- Equal distance in  $m/z$
- Almost equal distance in time
- May be ionized in multiple charge states
- Many wasteful CID attempts
- Eliminated by better washing steps?

# 3. Evaluating LC-ESI-MS/MS performance



Good performance:

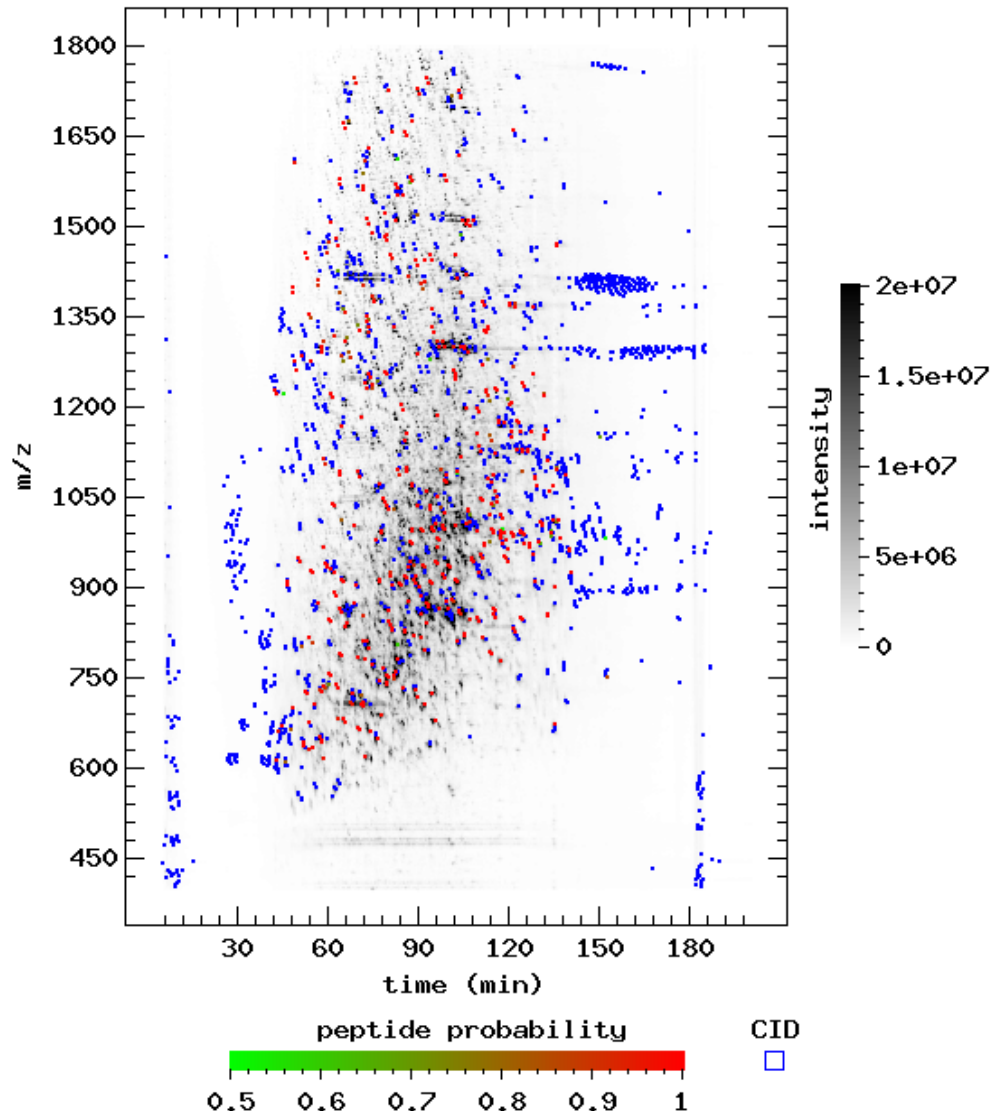
- Peptide ions evenly distributed
- Smooth background
- Majority of intensive ions fragmented

End too soon

# Insufficient Sample Separation

peptide/CID

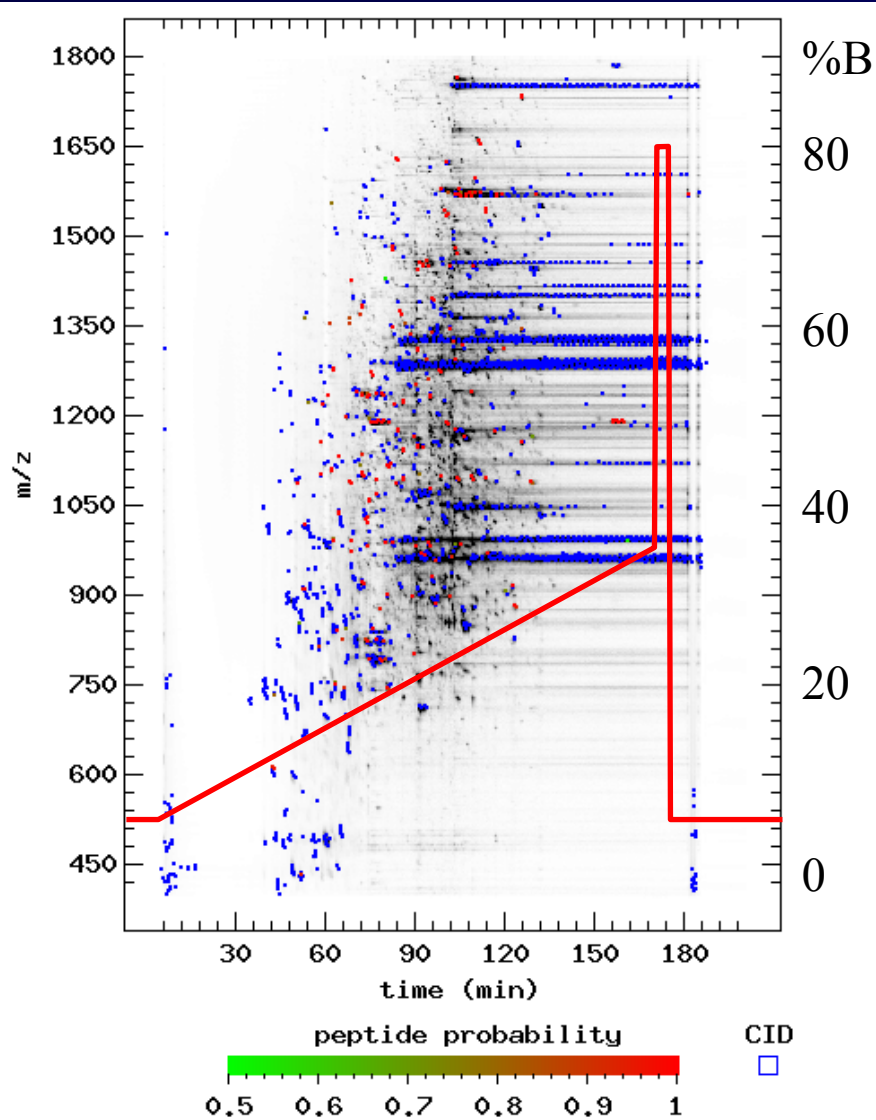
$(527/2150 = 0.245116)$



- Large proportion of intensive ions not fragmented
- Insufficient SCX separation
- Non-optimal RP separation

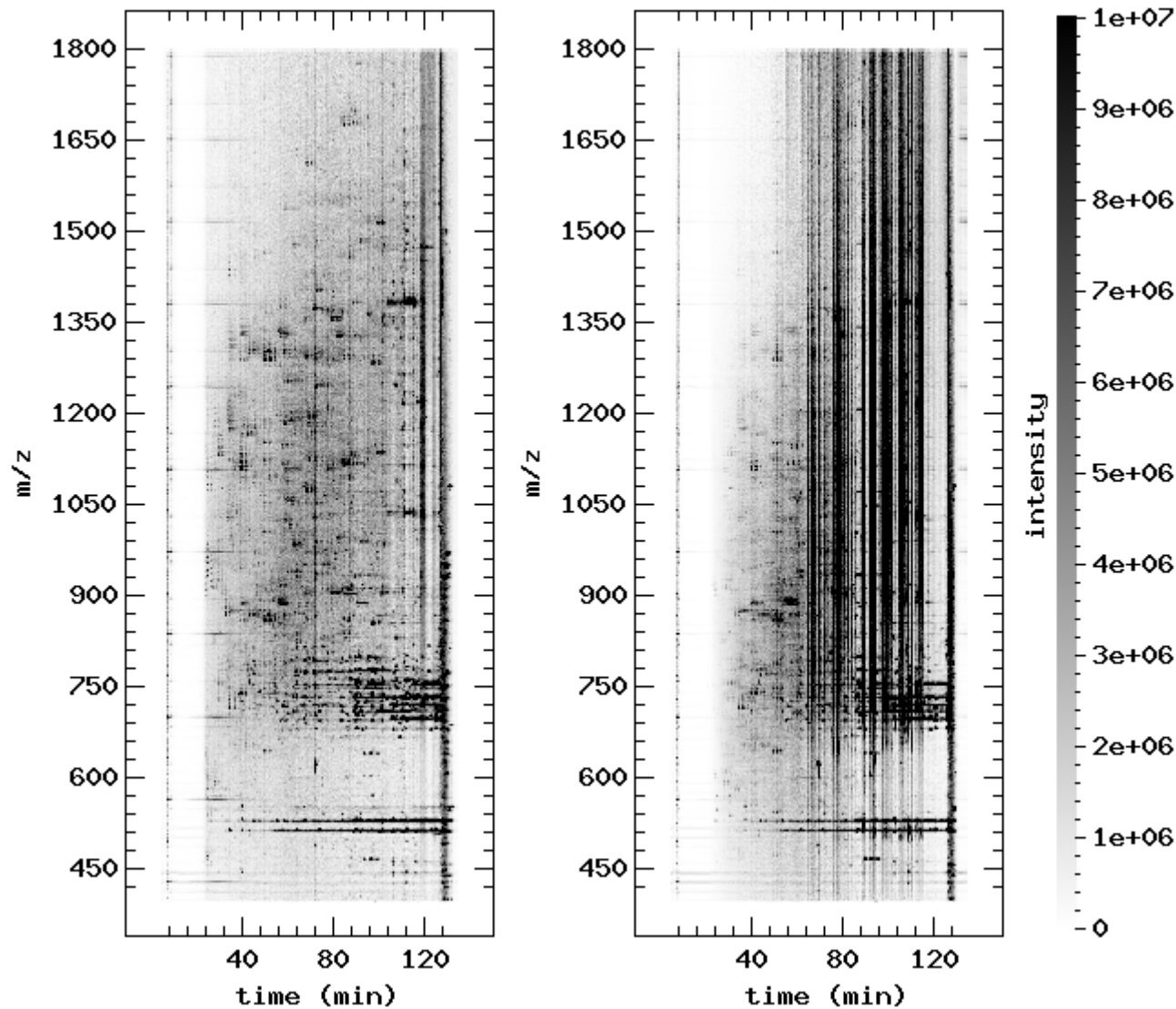


# Non-Optimal RP Gradient



- 1<sup>st</sup> ID: 42 min
- Effective range: 60-120 mins
- Horizontal streaks
- Larger slope at beginning
- Slower slope in middle
- Higher %B at end

# Bad RP Column



Same sample,  
same system,  
different columns

peptide/CID

(354/3004 = 0.12)

(224/2760 = 0.08)

37% less IDs

Quantification also  
suffers

# Summary

- Pep3D can be used to evaluate sample quality and LC-ESI-MS/MS performance
- Other applications possible
- Suggestion: Use complex standard sample to check system performance

# Exercises with PeptideProphet

---

- Accuracy of computed probabilities
- Utility of conventional SEQUEST score thresholds and PeptideProphet analysis
- Model results for ICAT data analyzed with and without ICAT information
- Model results for unconstrained vs. tryptic constrained search results

# Exercise Datasets

---

Many of the exercises utilize SEQUEST search results generated from datasets for which true results are known:

- HaloICAT: ICAT halobacterium sample searched against a **halo\_plus\_human** protein database

PeptideProphet run on these datasets automatically **colors all correct corresponding proteins red !**