AMT Tag Proteomics

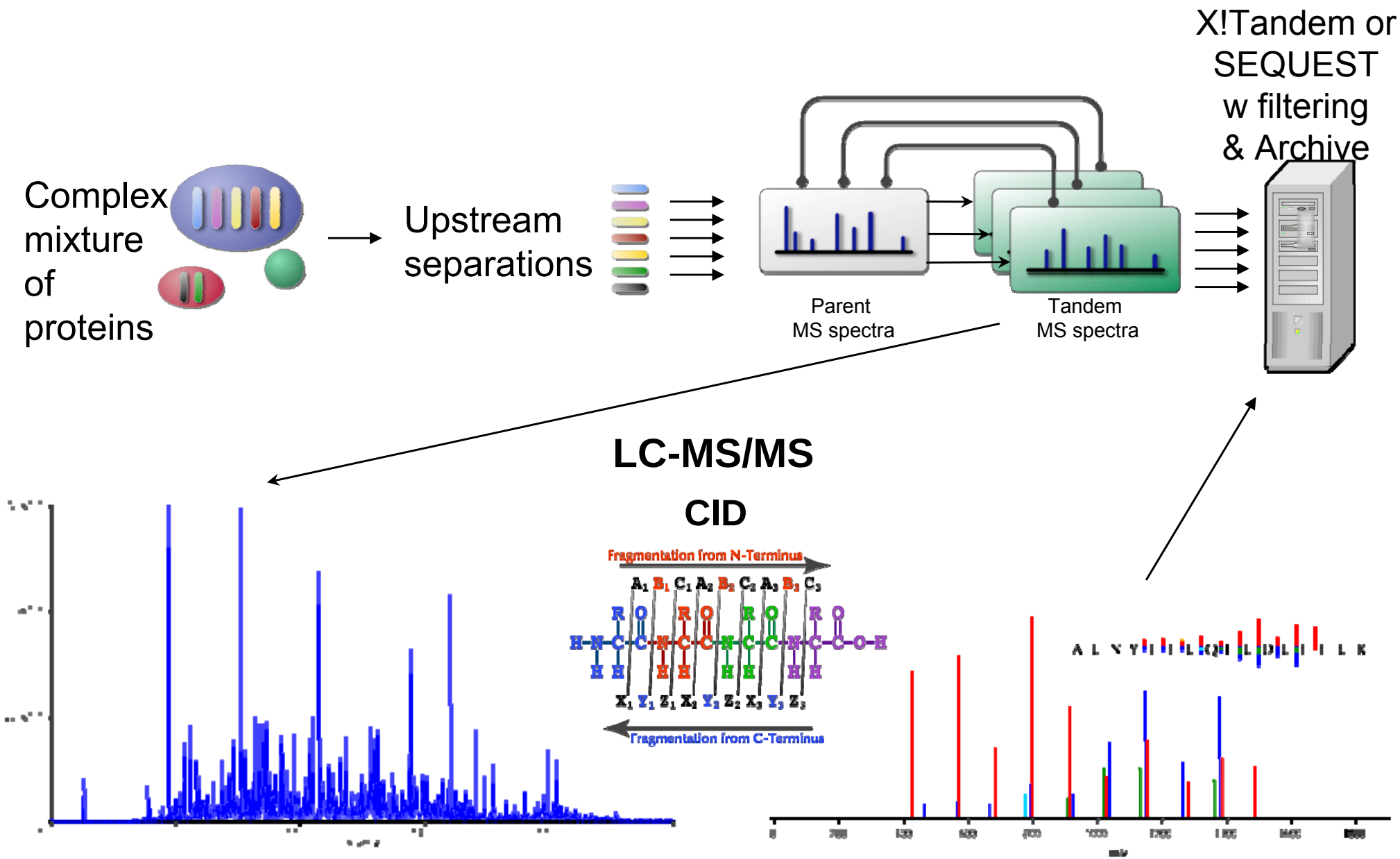
Data Extraction and Analysis for LC-MS Based Proteomics

Presentation prepared by
Deep Jaitly and Matt Monroe
Pacific Northwest National Laboratory, Richland, WA 99354

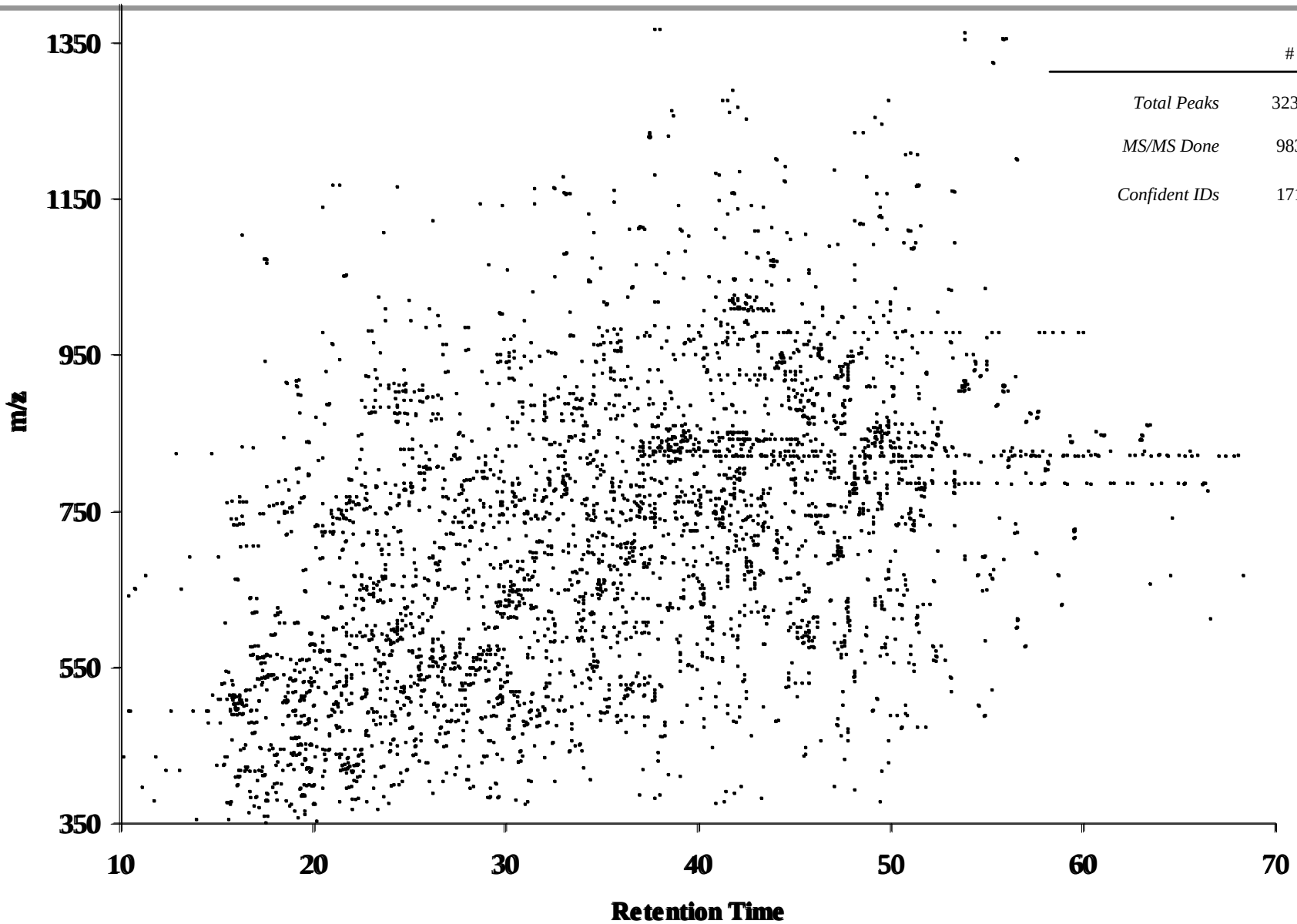
Outline

- Introduction
  - Pipelines
  - Data and Tools Availability
- Feature discovery in LC-MS datasets
  - Feature discovery in individual spectra
  - Feature definition over elution time
- Identifying LC-MS Features using an AMT tag DB

Shotgun or MuDPIT Proteomics



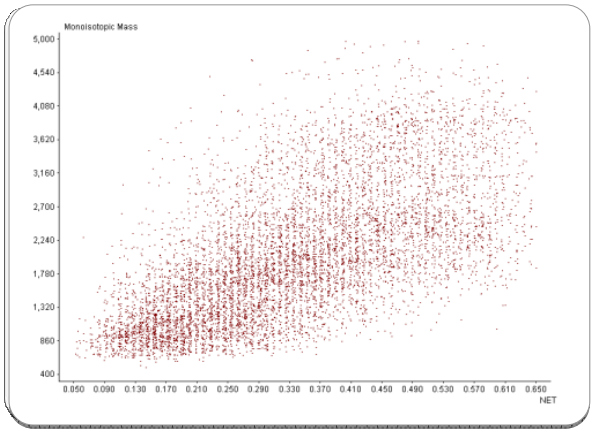
LCMS Information Funnel



	#	%
Total Peaks	3230	--
MS/MS Done	983	30%
Confident IDs	171	5%

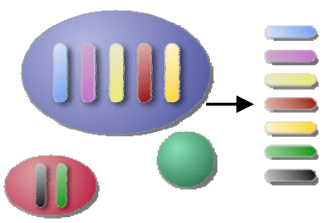
1. Detected Peaks 2. Selected for MS/MS 3. Confident peptide ID

Accurate Mass and Time Tag Approach

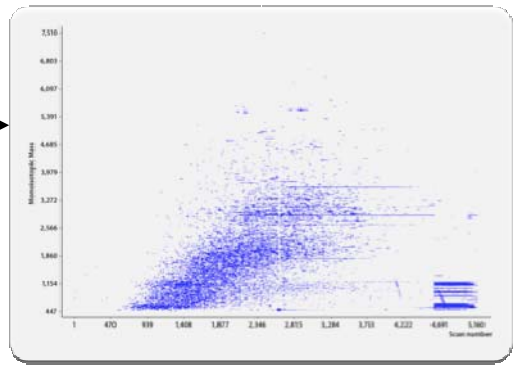


SEQUEST and/or
X!Tandem Results
•Filtering
•Calculate Exact Mass
•Normalize Observed
Elution Time

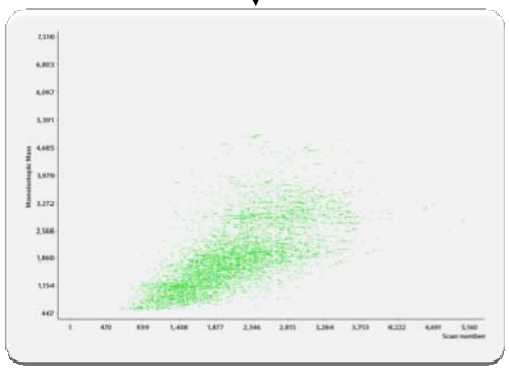
High-throughput LC-FTICR-MS Analysis (AMT) tag



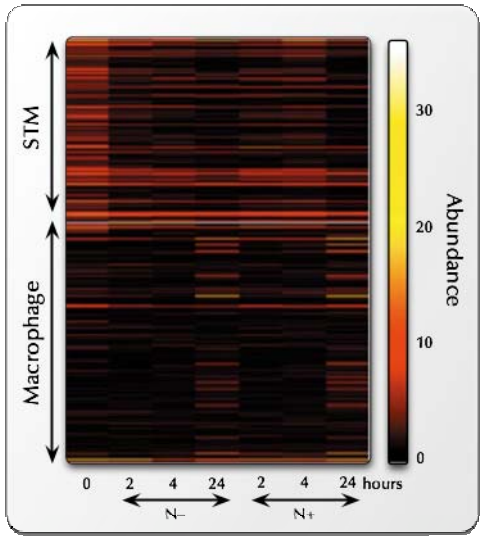
Complex samples



μLC- FTICR-MS



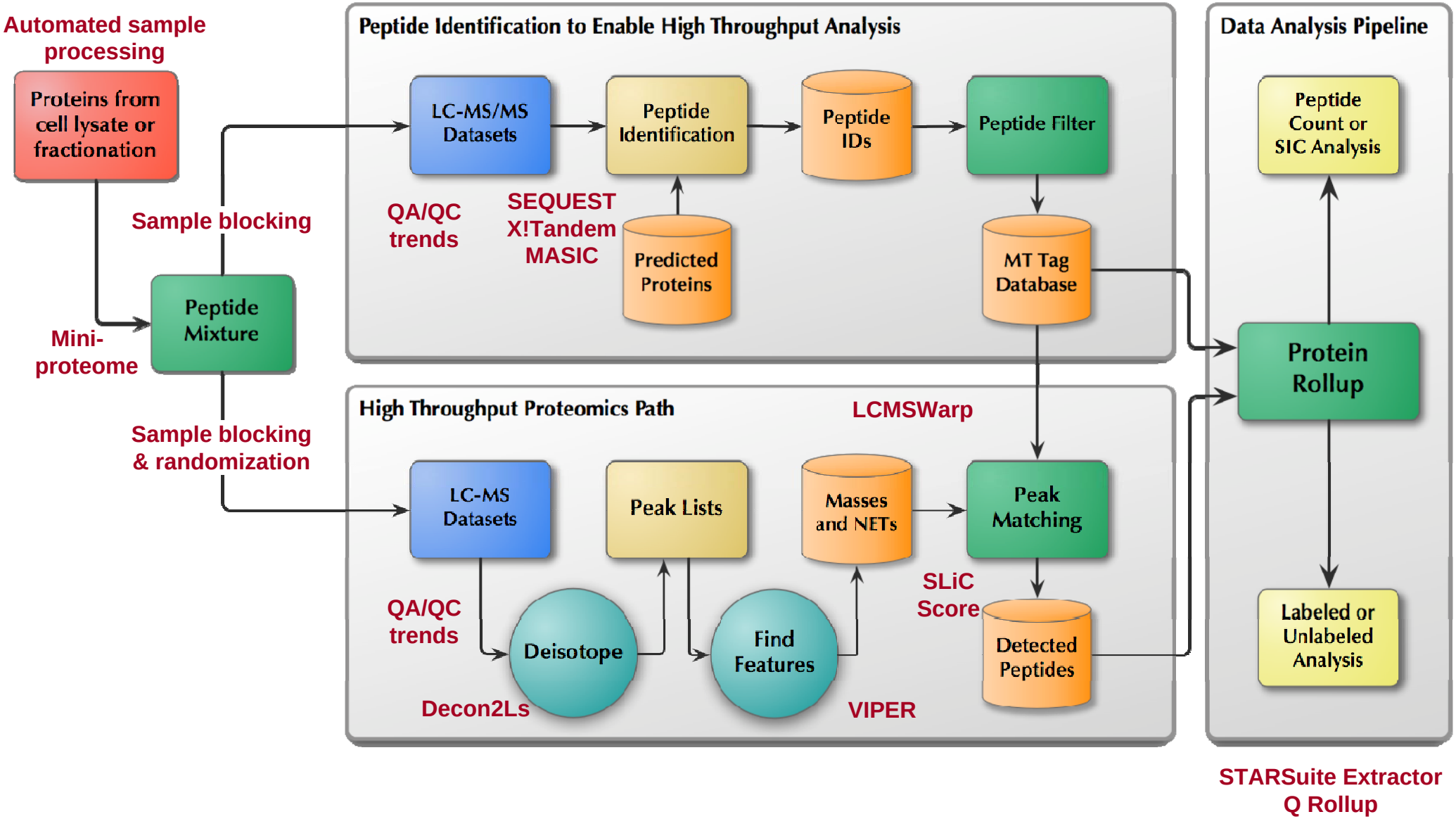
Peak-Matched Results



Compare Abundances across Multiple Proteomes

Shi, Adkins, et. al., *J. Bio. Chem*. **2006**, 29131-29140.

Accurate Mass and Time (AMT) Tag Data Processing Pipeline



PRISM: G.R. Kiebel et. al. *Proteomics* **2006**, *6*, 1783-1790.

Example Data for the AMT tag Pipeline Demo

- Salmonella typhimurium, LC-MS/MS
  - Grown in **LB** (**Luria-Bertani**) up to log phase
  - Soluble portion of cell lysis
  - “Mini-AMT tag” database, composed of 25 SCX fractions analyzed by LC-MS/MS
  - Mass and time tag database composed from searches using X!Tandem (Log E_Value ≤ -2)
  - Linear alignment of datasets for AMT tag database
- LC-MS
  - Different sample, grown and prepared in the same conditions
  - LC-FTICR-MS analysis (11T FTICR)
  - Non-linear alignment and peak matching to the database

Software & Data

AMT tag Pipeline Software

PROTEOMICS
National Center for Research Resources

http://ncrr.pnl.gov/

Salmonella typhimurium data resource

Resource Center for
Biodefense Proteomics Research

http://www.proteomicsresource.org/

Funding for Tool Development

- NIH
  - National Center for Research Resources
  - National Institute of Allergy and Infectious Diseases
  - National Cancer Institute
  - National Institute of General Medical Sciences
  - National Institute of Diabetes & Digestive & Kidney Diseases
- DOE Office of Biological and Environmental Research

Other Excellent Software Resources

- http://www.ms-utils.org/ (Magnus Palmblad)
- http://open-ms.sourceforge.net/index.php (European consortium)
- http://tools.proteomecenter.org/SpecArray.php (ISB)
- http://fiehnlab.ucdavis.edu/staff/kind/Metabolomics/Peak_Alignment/ (Tobias Kind with Oliver Fiehn)
- http://www.proteomecommons.org/tools.jsp (Phil Andrews and Jayson Falkner)
- https://proteomics.fhcrc.org/CPAS/Project/home/begin.view (CPAS)
- http://arep.med.harvard.edu/MapQuant/ (MapQuant)

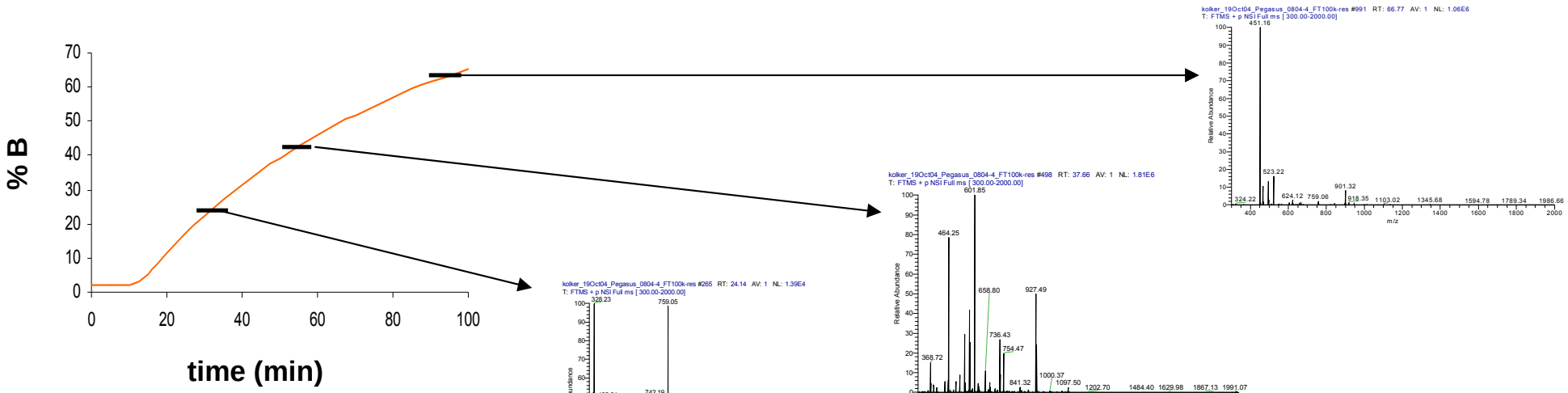
Outline

- ✓Introduction
- Feature discovery in LC-MS datasets
  - Feature discovery in individual spectra
  - Feature definition over elution time
- Identifying LC-MS Features using an AMT tag DB

LC-MS data

- Complex peptide mixture on a column is separated by liquid chromatography over a period of time
- Changing composition of the mobile phase causes different peptides to elute at different times
- The components eluting from a column is sampled continuously by sequential mass spectra

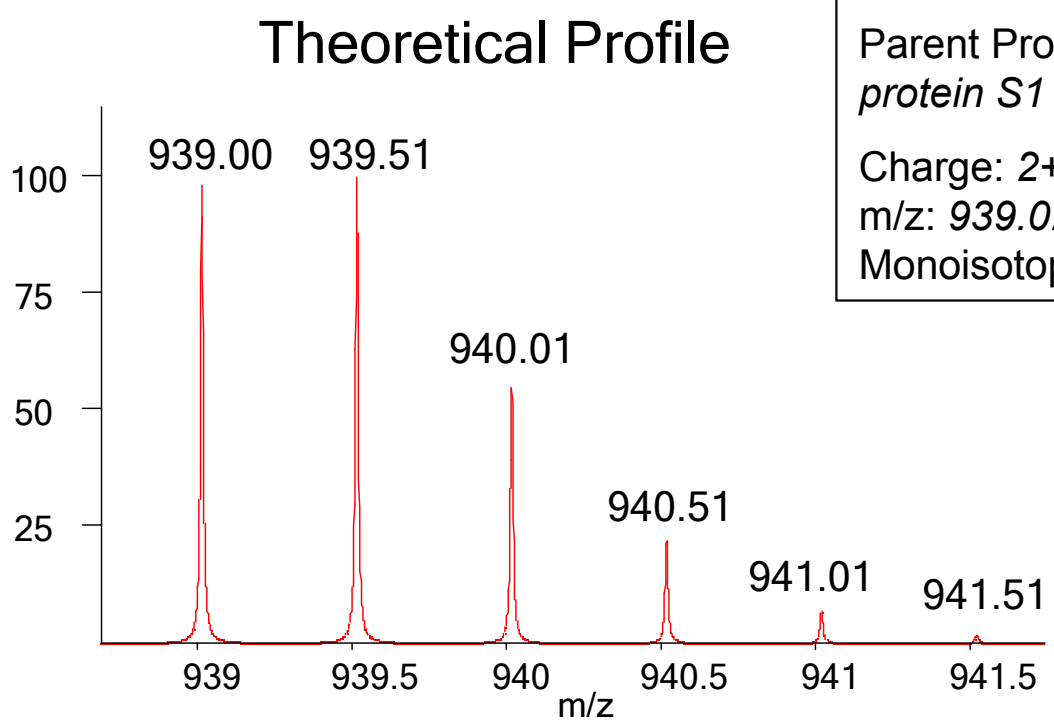
Mass spectra capture the changing composition of peptides eluting from the column



QC Standards
(12 protein digest)

Structure of LC-MS Data

- Each compound is observed as an <u>isotopic pattern</u> in a mass spectrum which is dependent on its chemical composition, charge and resolution of instrument



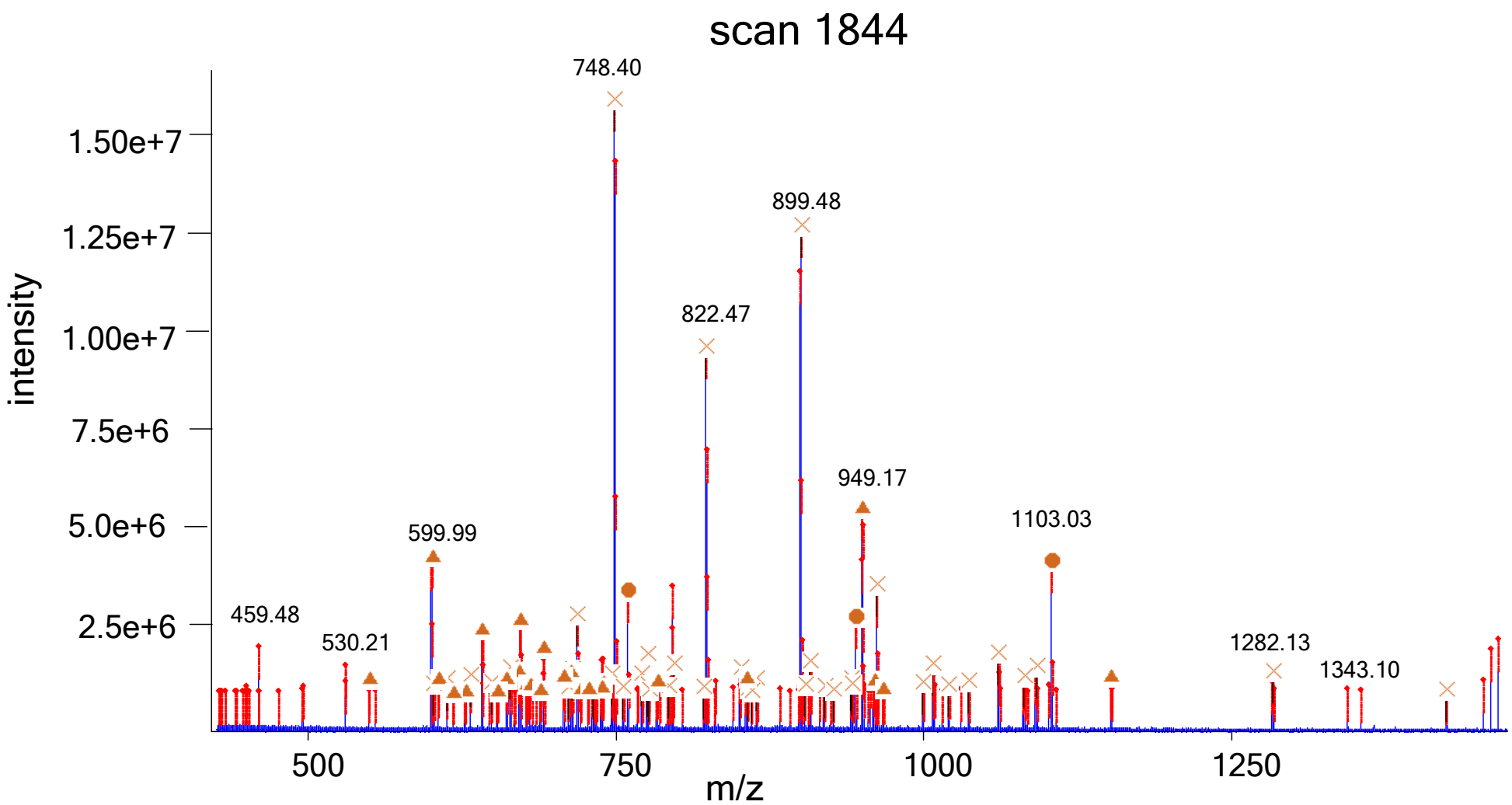
Peptide: *VKHPSEIVNVGDEINVK*

Parent Protein: *gi|16759851 30S ribosomal protein S1*

Charge: *2+*
m/z: *939.0203*
Monoisotopic Mass: *1876.0054 Da*

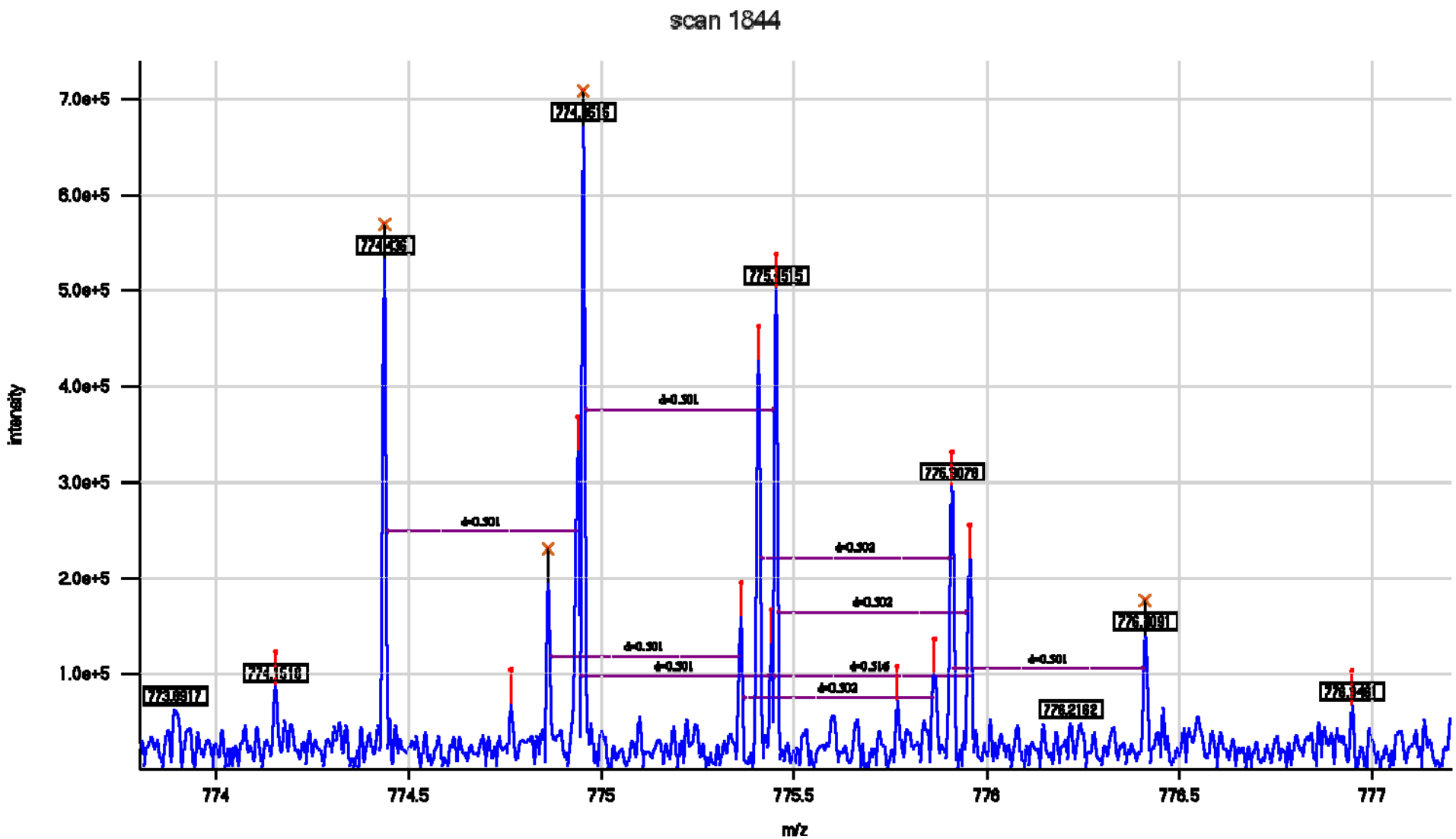
Structure of LC-MS Data

- A mass spectrum of a complex mixture contains overlaid distributions of several different compounds



Structure of LC-MS Data

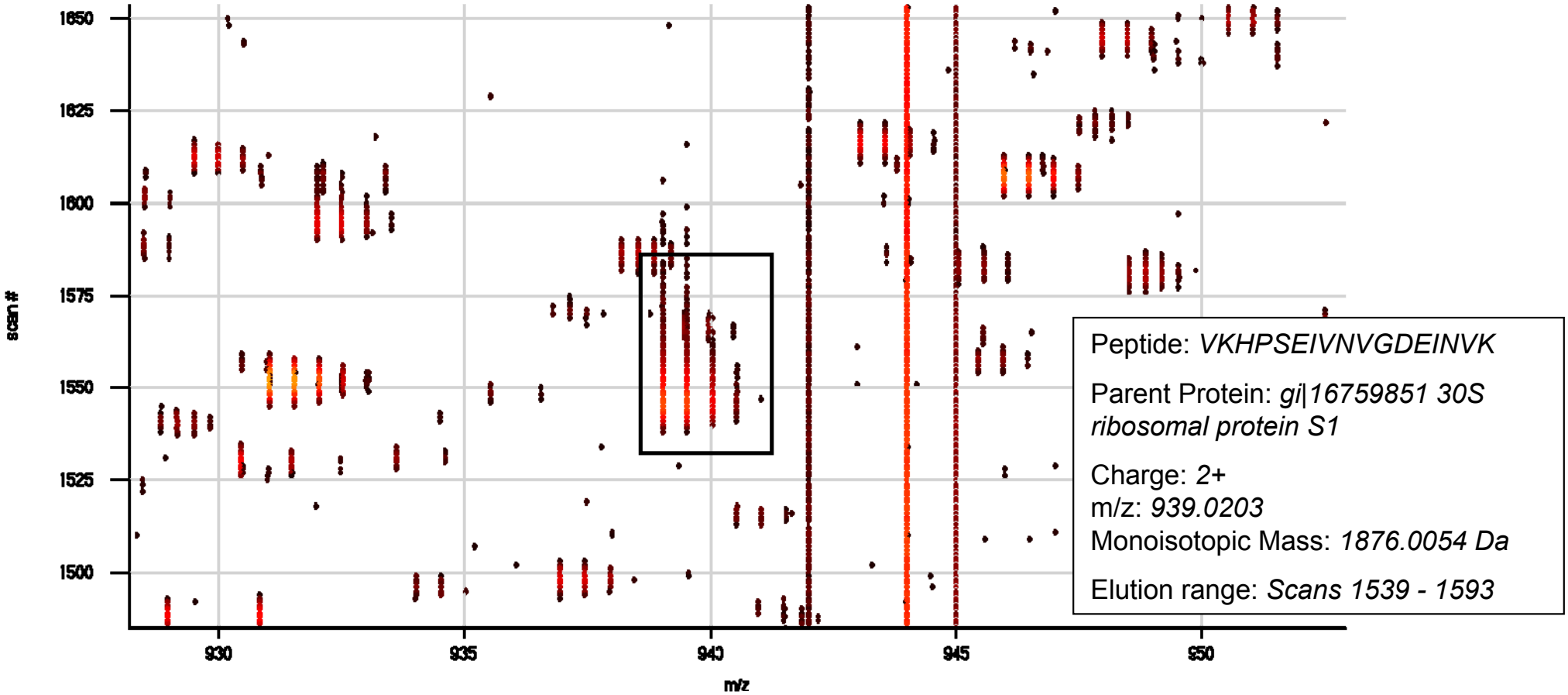
- A mass spectrum of a complex mixture contains overlaid distributions of several different compounds.



Structure of LC-MS Data

- With LC as the first dimension, each compound is observed over multiple spectra, showing a three-dimensional pattern of m/z, elution time and abundance

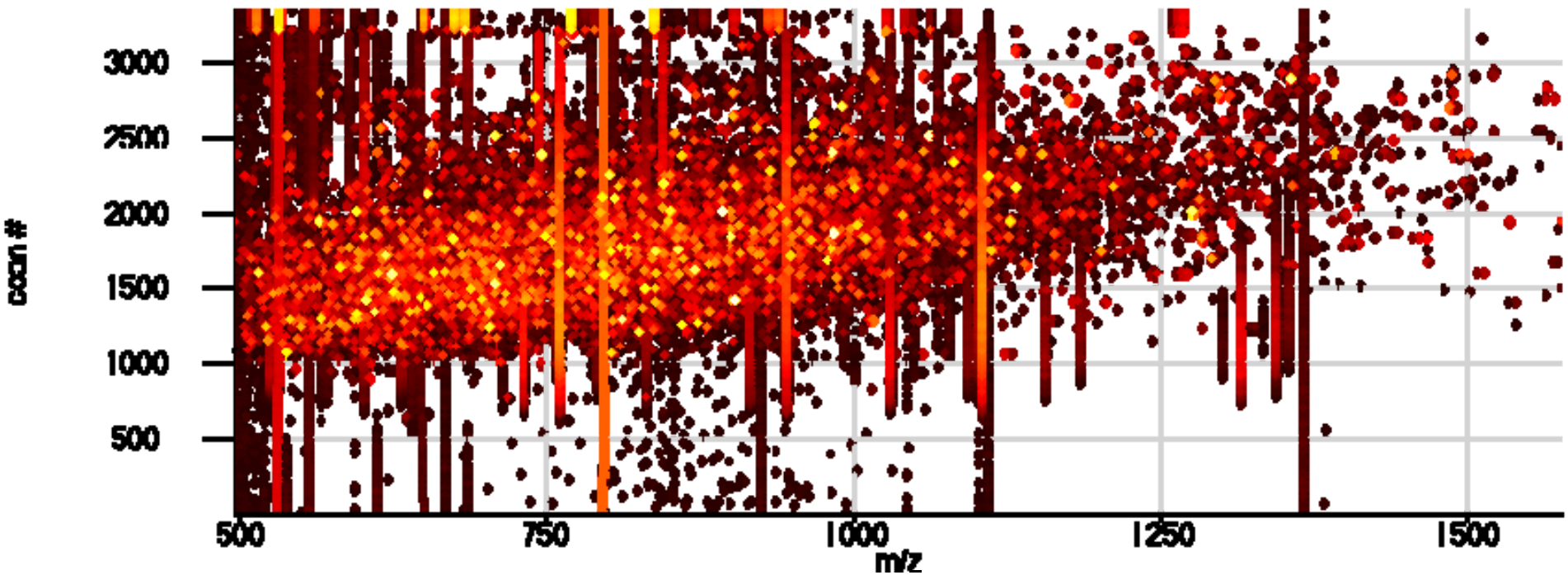
Salmonella typhimurium dataset



Feature Discovery in LC-MS data

- Goal: Infer *(mass, elution time, intensity)* of compounds that are present in data obtained from an LC-MS dataset
  - Since their identities are unknown, the compounds are more appropriately termed features to refer to the idea that these are inferred from a three dimensional pattern

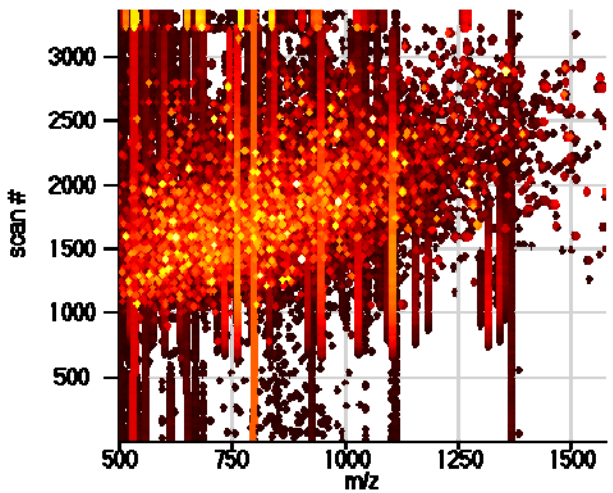
2D view of an LC-MS analysis of Salmonella typhimurium

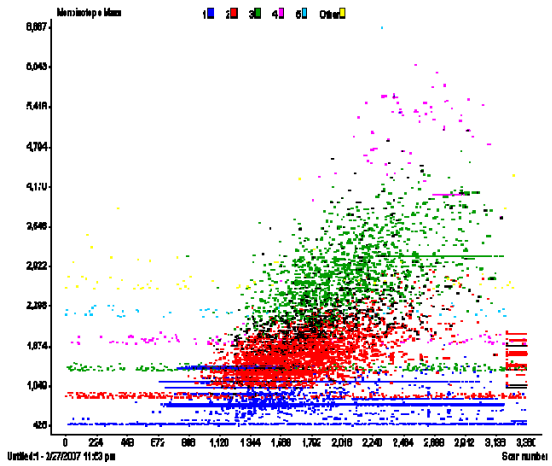


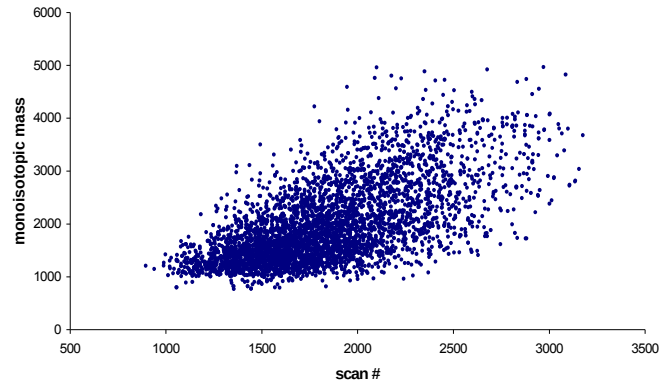
Feature Discovery in LC-MS data

- Sequential process of finding features in each mass spectrum is followed by grouping of features over multiple spectra together

2D views of an LC-MS dataset in different stages of processing





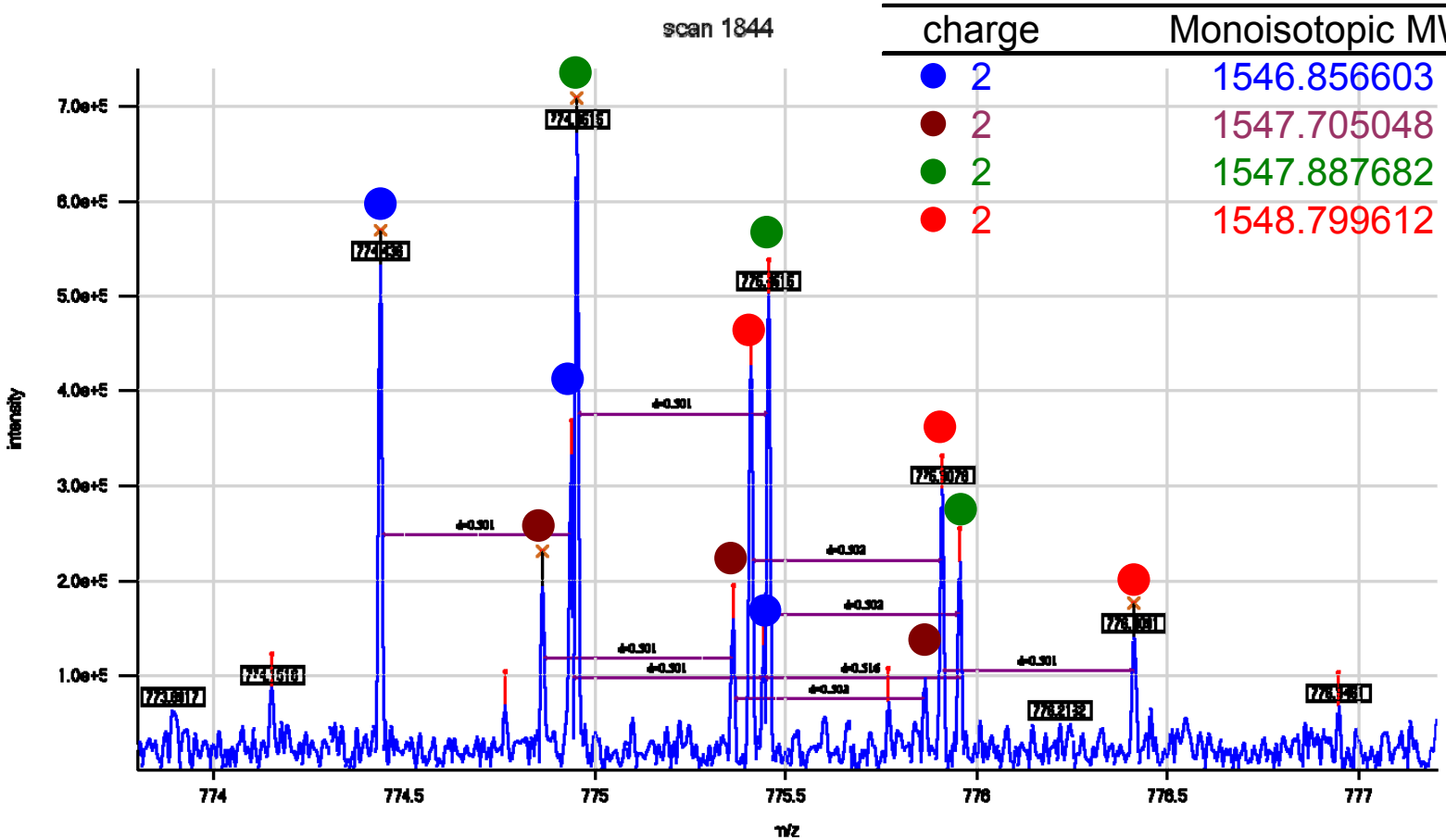


raw data → deisotoping → Collapsed monoisotopic features in all spectra → Elution profile discovery → LC-MS features

Feature discovery in individual spectra

- Deisotoping
  - Process of converting a mass spectrum (*m/z, intensity*) into a list of species (*mass, abundance, charge*)

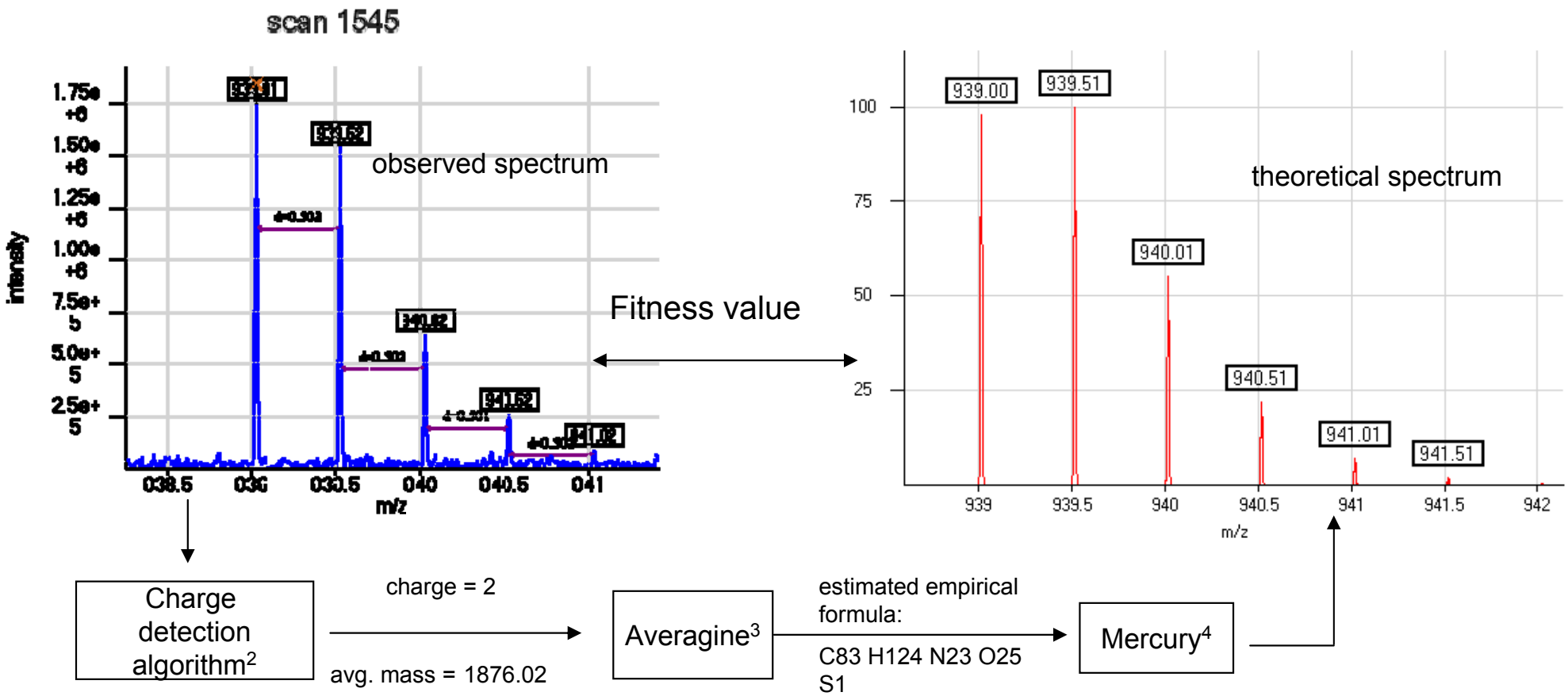
Deisotoping a mass spectrum of 4 overlapping species



charge	Monoisotopic MW	abundance
2	1546.856603	533467
2	1547.705048	194607
2	1547.887682	671947
2	1548.799612	426939

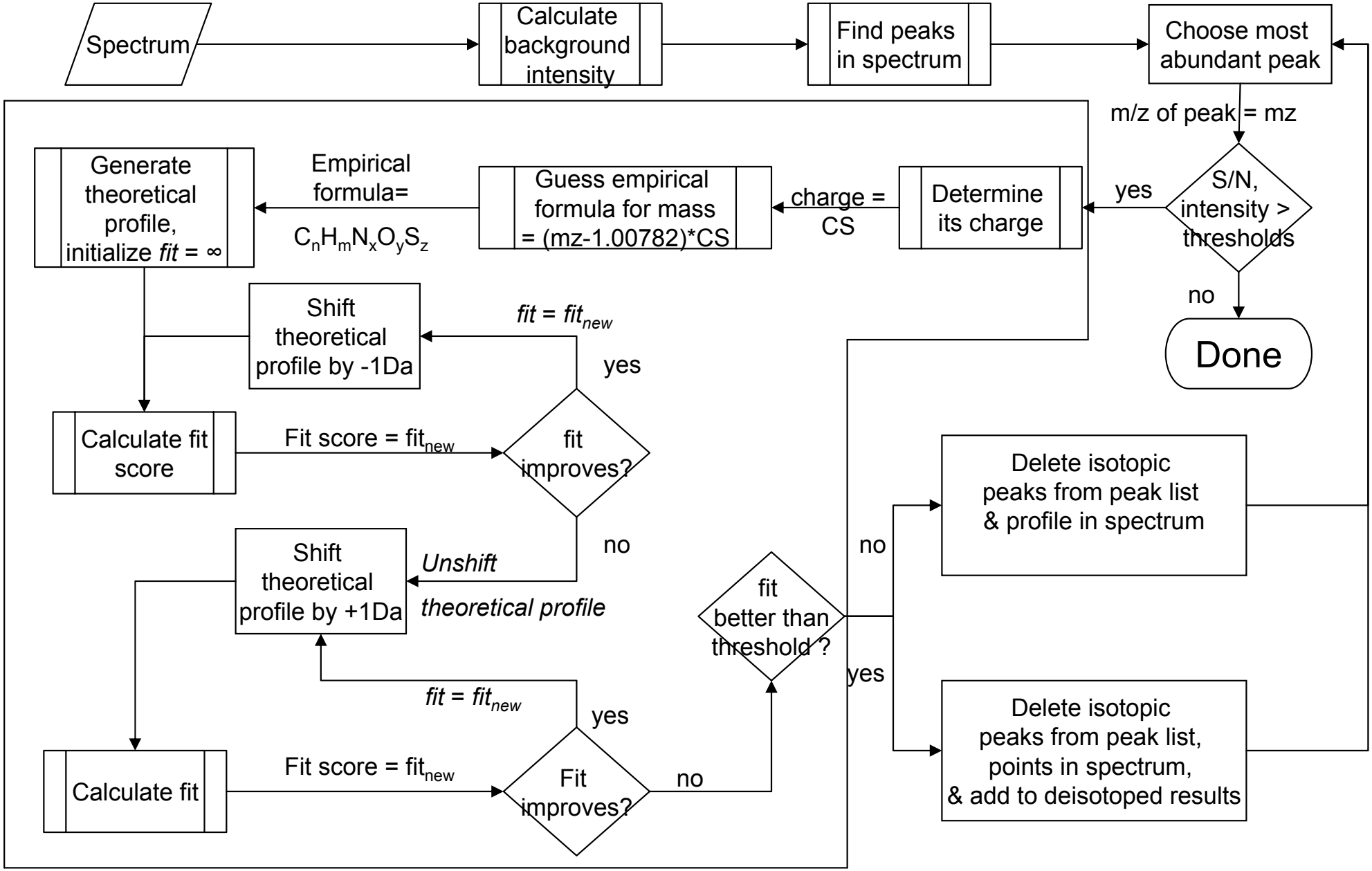
Deisotoping routine for a peak

- Algorithm to detect peptides in a complex spectrum



1. Horn, D.M., Zubarev, R.A., McLafferty, F.W. Automated Reduction and Interpretation of High Resolution Electrospray Mass Spectra of Large Molecules. *J. Am. Soc. Mass Spectrom.* **2000**, *11*, 320-332.
2. Senko, M. W.; Beu, S. C.; McLafferty, F. W. Automated assignment of charge states from resolved isotopic peaks for multiplycharged ions. *J. Am. Soc. Mass Spectrom.* **1995,** *6*, 52–56.
3. Senko, M. W.; Beu, S. C.; McLafferty, F. W. Determination of monoisotopic masses and ion populations for large biomolecules from resolved isotopic distributions. *J. Am. Soc. Mass Spectrom.* **1995,** *6*, 229–233.
4. Rockwood, A. L.; Van Orden, S. L.; Smith, R. D. Rapid Calculation of Isotope Distributions. *Anal. Chem.* **1995,** *67,* 2699–2704.

Deisotoping entire spectrum – Modification of THRASH[1]



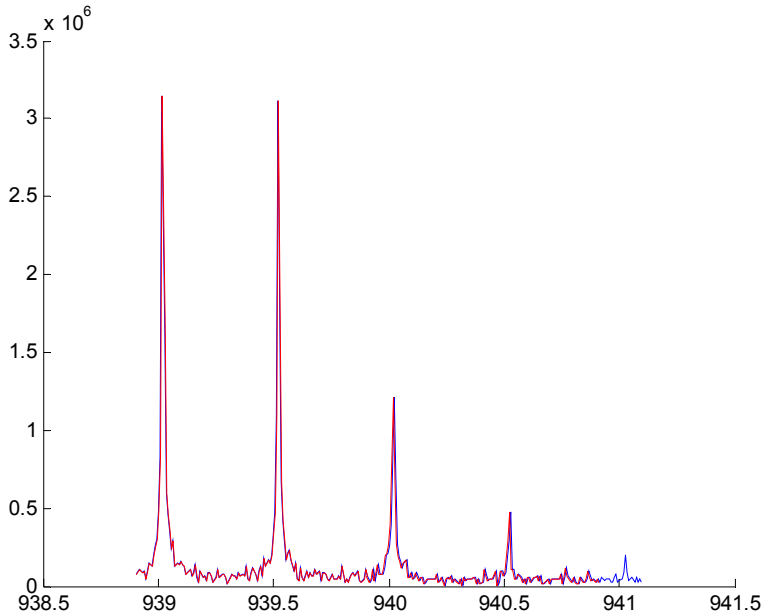
1. Horn, D.M., Zubarev, R.A., McLafferty, F.W. Automated Reduction and Interpretation of High Resolution Electrospray Mass Spectra of Large Molecules. *J. Am. Soc. Mass Spectrom.* **2000**, *11*, 320-332.

Modified THRASH Routine

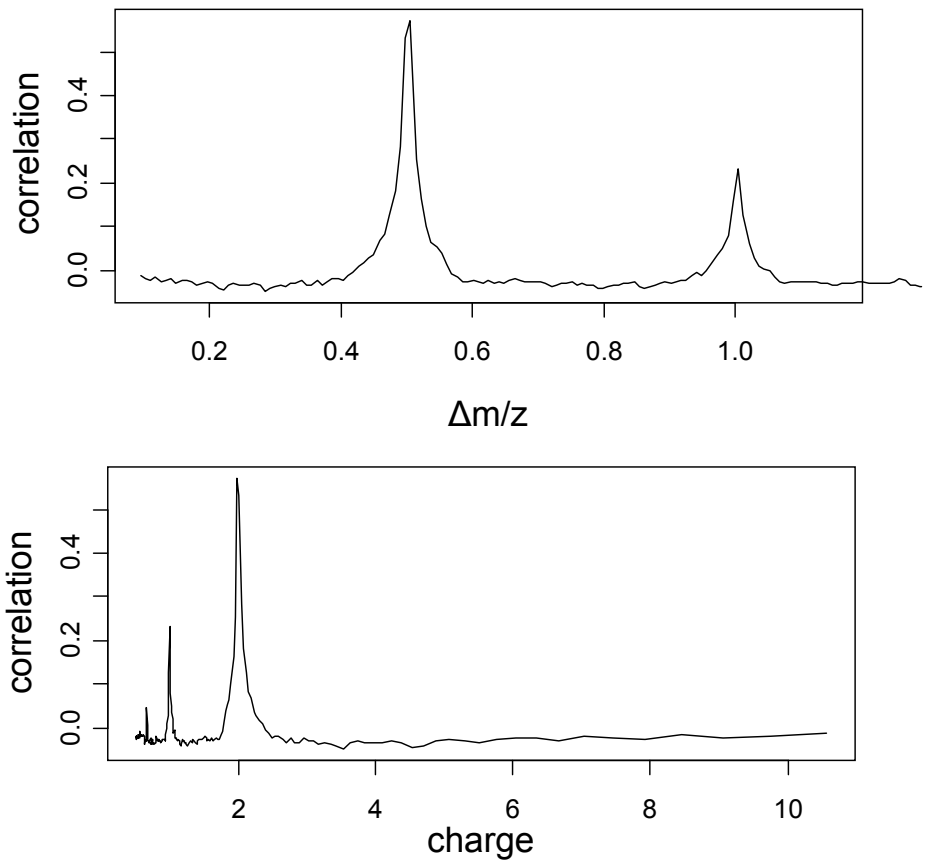
- Algorithm to detect peptides in a complex spectrum
  1. Discover all peaks in a spectrum above a specified S/N and keep in unprocessed list
  2. Select most abundant peak still unprocessed
  3. Compute charge for peak using charge detection algorithm
  4. Compute average mass from observed m/z and predicted charge value
  5. Use “Averagine” algorithm to guess empirical formula based on mass and average composition of peptides in database
  6. Use Mercury algorithm to generate theoretical spectrum from the predicted empirical formula, charge of peak, and resolution of peak
  7. Calculate fitness value for similarity between theoretical and observed spectrum
  8. Perform “THRASHING” by overlaying theoretical and observed spectra after applying “isotopic” one dalton shift to the theoretical spectrum. Keep best fit
  9. If successful fit was observed, delete isotopic peaks and associated profile of height above specified threshold of most intense ions using the theoretical pattern as template. Otherwise remove current peak only from list of unprocessed peaks.
  10. While unprocessed peaks remain, repeat steps starting at step 2.

Charge Detection Algorithm

- Patterson (Autocorrelation) algorithm to detect charge of a peak in a complex spectrum



$$P(\Delta mz) = \Sigma I(mz_i) * I(mz_i + \Delta mz)$$



1. Zhang, Z; Marshall, A.G. A Universal Algorithm for Fast and Automated Charge State Deconvolution of Electrospray Mass-to-Charge Ratio Spectra. *J. Am. Soc. Mass Spectrom.* **1998**, *9*, 225-233.
2. Senko, M. W.; Beu, S. C.; McLafferty, F. W. Automated assignment of charge states from resolved isotopic peaks for multiplycharged ions. *J. Am. Soc. Mass Spectrom.* **1995,** *6*, 52–56.
3. Labowsky, M; Whitehouse, C.; Fenn, J.B. *Rapid Commun. Mass Spectrom.* **1993**, *7*, 71-84.
4. Reinhold, B.B.; Reinhold, V.N. Electrospray Ionization Mass Spectrometry: Deconvolution by an Entropy-Based Algorithm. *J. Am. Soc. Mass Spectrom.* **1992,** *3*, 207-215.
5. Mann, M.; Meng, C.K.; Fenn, J.B. Interpreting Mass Spectra of Multiply Charged Ions. *Analytical Chemistry.* Aug. 1, **1989,** *61*, 1702-1708.

Averagine Algorithm

- Algorithm to guess an average empirical formula for a given mass
  - Uses average composition of all peptides in peptide database as the empirical formula for all peptides
  - Protein database Averagine formula: $C_{4.9384}\ H_{7.7583}\ N_{1.3577}\ O_{1.4773}\ S_{0.0417}$, Mass = 111.1254
  - Average Mass of 1877.025 would give a multiplier of 1877.025/111.1254

Element	Copies	Atomicity
C	4.9834*1877.025/111.1254	84
H		Remainder = 112
N	1.3577*1877.025/111.1254	23
O	1.4773*1877.025/111.1254	25
S	0.0417*1877.025/111.1254	1

Empirical formula used for theoretical profile = $C_{83}\ H_{112}\ N_{23}\ O_{25}\ S_{1}$

1. Senko, M. W.; Beu, S. C.; McLafferty, F. W. Determination of monoisotopic masses and ion populations for large biomolecules from resolved isotopic distributions. J. Am. Soc. Mass Spectrom. **1995**, *6*, 229–233.

Theoretical Isotopic Profile

- Mercury algorithm to generate a theoretical profile for a compound
  - Treat each element's isotopic distribution as a sum of delta (δ) functions

Element	Isotope distribution Function
Carbon	0.98893 δ(m-12) + 0.01107 δ(m-13.00336)
Hydrogen	0.99985 δ(m-1.007825) + 0.00015 δ(m-2.014102)
Nitrogen	0.996337 δ(m-14.00307) + 0.003663 δ(m-15.00011)
Oxygen	0.99759 δ(m-15.99491) + 0.000374 δ(m-16.99913) + 0.002036 δ(m-17.99916)
Sulphur	0.9502 δ(m-31.97207) + 0.0075 δ(m-32.97145)+ 0.0421 δ(m-33.96786) + 0.0002 δ(m-35.96708)

Relative isotope abundance

Isotope Mass

1. Rockwood, A. L.; Van Orden, S. L.; Smith, R. D. Rapid Calculation of Isotope Distributions. *Anal. Chem.* **1995**, *67*, 2699–2704.
2. Kubinyi, H. Calculation of isotope distributions in mass spectrometry. A trivial solution for a non-trivial problem. *Analytica Chemica Acta*. **1991**, *247*, 107-119.

Theoretical Isotopic Profile

- Mercury algorithm to generate a theoretical profile for a compound
  - Treat each element's isotopic distribution as a sum of delta (δ) functions
  - Convert distribution function into frequency domain: delta functions convert to simple exponential functions

Element	Frequency Spectrum Function ($f_{Elem}(\mu)$)
Carbon	$0.98893\ e^{12(i2\pi)\mu} + 0.01107\ e^{13.00336(i2\pi)\mu}$
Hydrogen	$0.99985\ e^{1.007825(i2\pi)\mu} + 0.00015\ e^{2.014102(i2\pi)\mu}$
Nitrogen	$0.996337\ e^{14.00307(i2\pi)\mu} + 0.003663\ e^{15.00011(i2\pi)\mu}$
Oxygen	$0.99759\ e^{15.99491(i2\pi)\mu} + 0.000374\ e^{16.99913(i2\pi)\mu} + 0.002036\ e^{17.99916\ (i2\pi)\mu}$
Sulphur	$0.9502\ e^{31.97207\ (i2\pi)\mu} + 0.0075\ e^{32.97145\ (i2\pi)\mu} + 0.0421\ e^{33.96786(i2\pi)\mu} + 0.0002\ e^{35.96708(i2\pi)\mu}$

Theoretical Isotopic Profile

- Mercury algorithm to generate a theoretical profile for a compound
  - Treat each element's isotopic distribution as a sum of delta (δ) functions
  - Convert distribution function into frequency domain: delta functions convert to simple exponential functions.
  - Calculate the isotopic profile for a compound from the convolution of isotopic distributions of individual atoms and the imposition of a peak shape reflecting resolution of instrument
  - Compute convolution using multiplication in the frequency domain and by applying a Fourier transform

For the empirical formula $C_nH_mN_xO_yS_z$

$$F(m) = FT\,[s(\mu)\, f_C(\mu)^n\, f_H(\mu)^m\, f_N(\mu)^x\, f_O(\mu)^y\, f_S(\mu)^z]$$

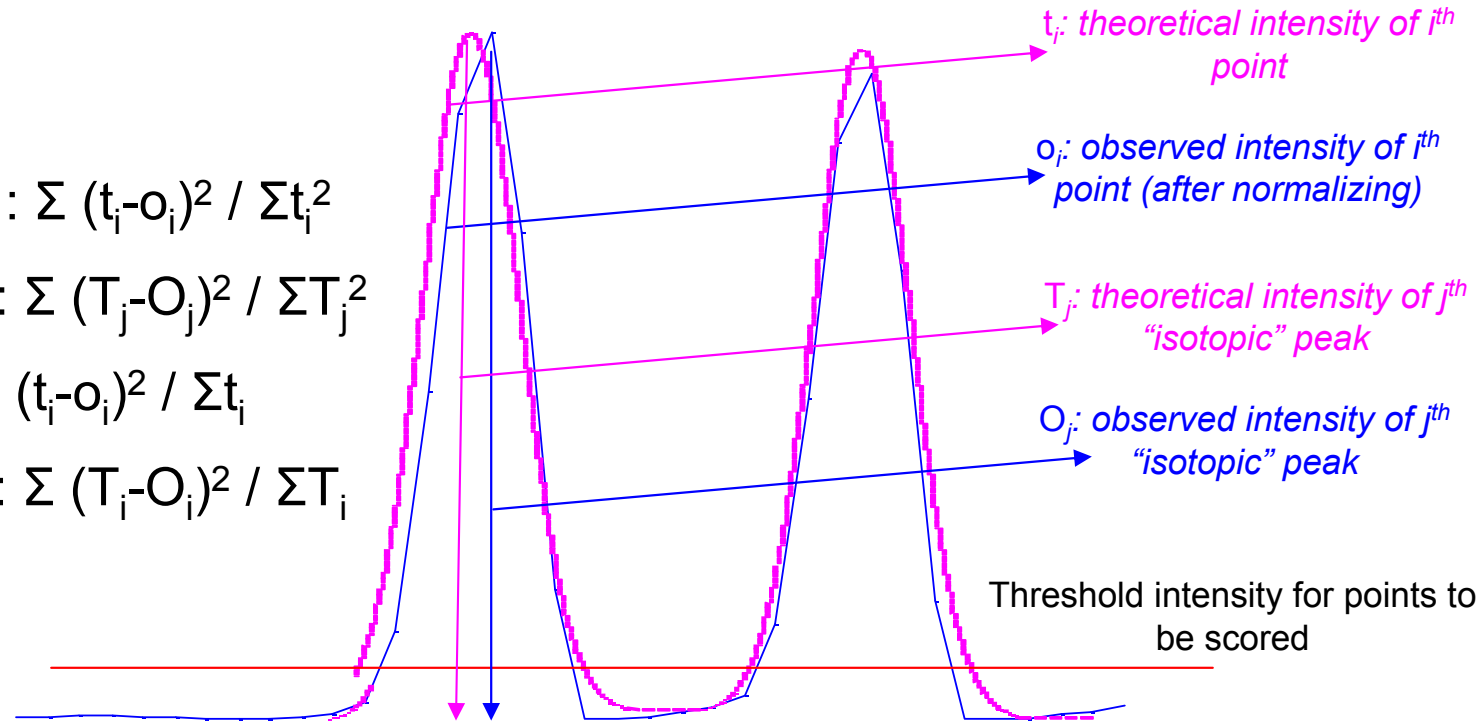
Inverse transform of shape function

Frequency spectra of elements

Fit Functions

- Fit functions to quantitate quality of match between theoretical and observed profiles

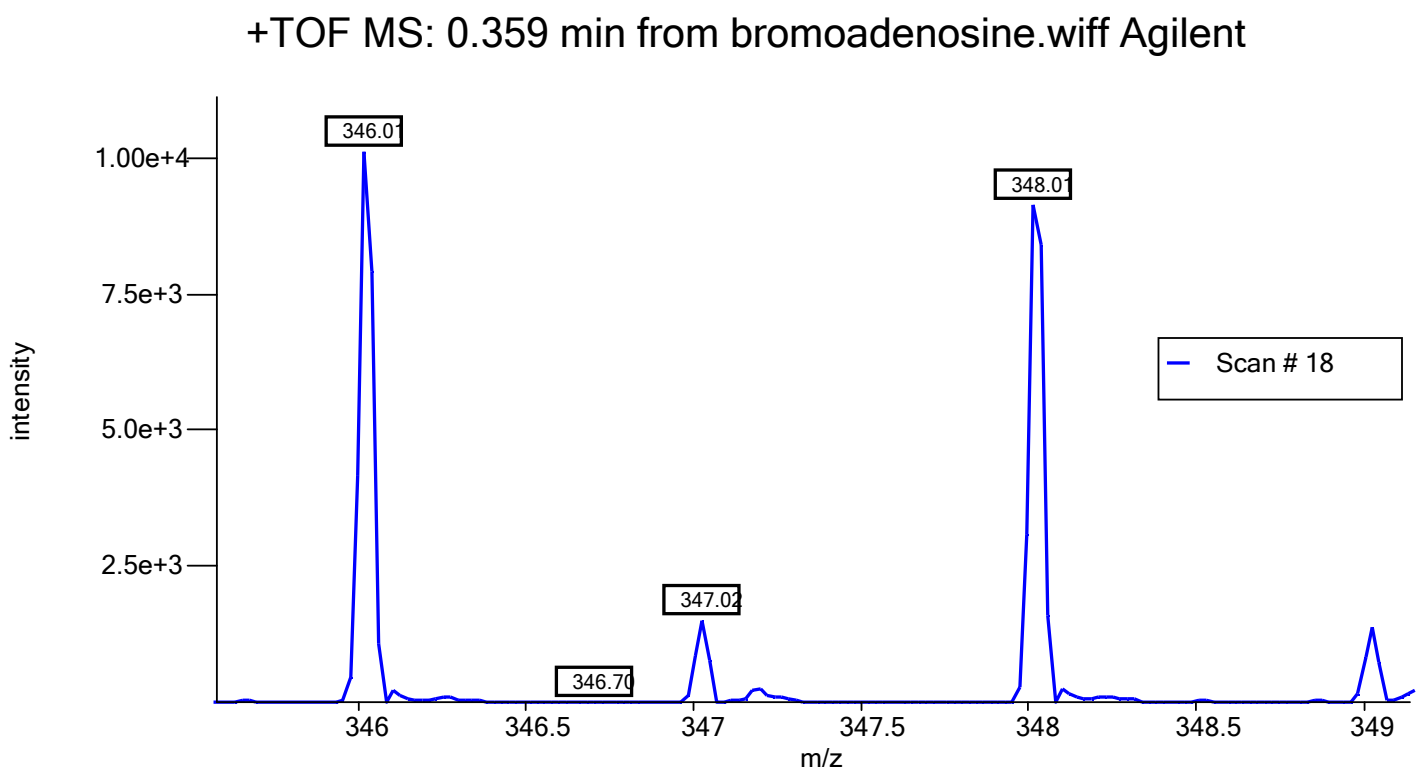
- Least square area[1]: $\Sigma\ (t_i\text{-}o_i)^2\ /\ \Sigma t_i^2$
- Least square peak: $\Sigma\ (T_j\text{-}O_j)^2\ /\ \Sigma T_j^2$
- Chi-square area: $\Sigma\ (t_i\text{-}o_i)^2\ /\ \Sigma t_i$
- Chi-square peaks[2]: $\Sigma\ (T_i\text{-}O_i)^2\ /\ \Sigma T_i$



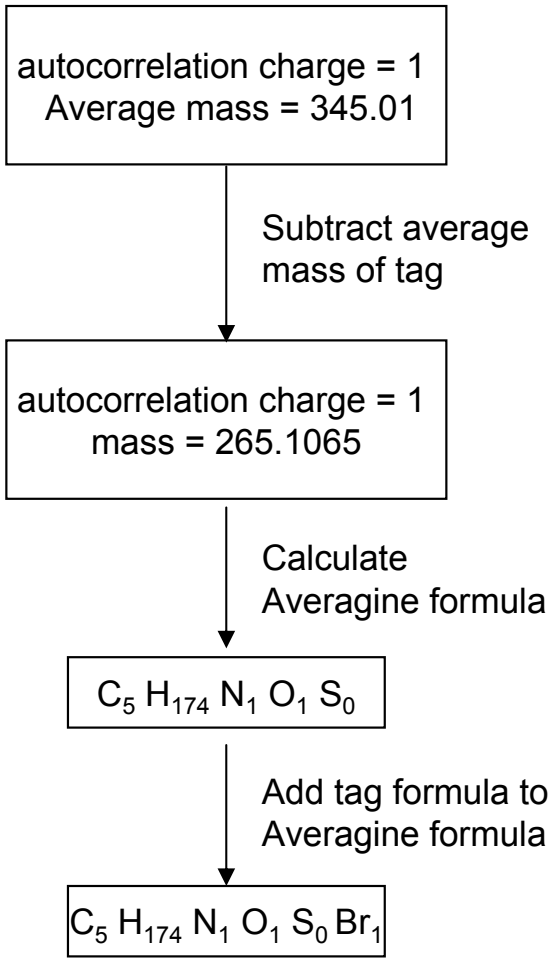
1. Horn, D.M., Zubarev, R.A., McLafferty, F.W. Automated Reduction and Interpretation of High Resolution Electrospray Mass Spectra of Large Molecules. *J. Am. Soc. Mass Spectrom.* **2000**, *11*, 320-332.
2. Senko, M. W.; Beu, S. C.; McLafferty, F. W. Determination of monoisotopic masses and ion populations for large biomolecules from resolved isotopic distributions. *J. Am. Soc. Mass Spectrom.* **1995,** *6*, 229–233.

Chemical Labeling with Tags

- Specify a static tag to be applied to Averagine formula
  - Changes the Averagine formula generated

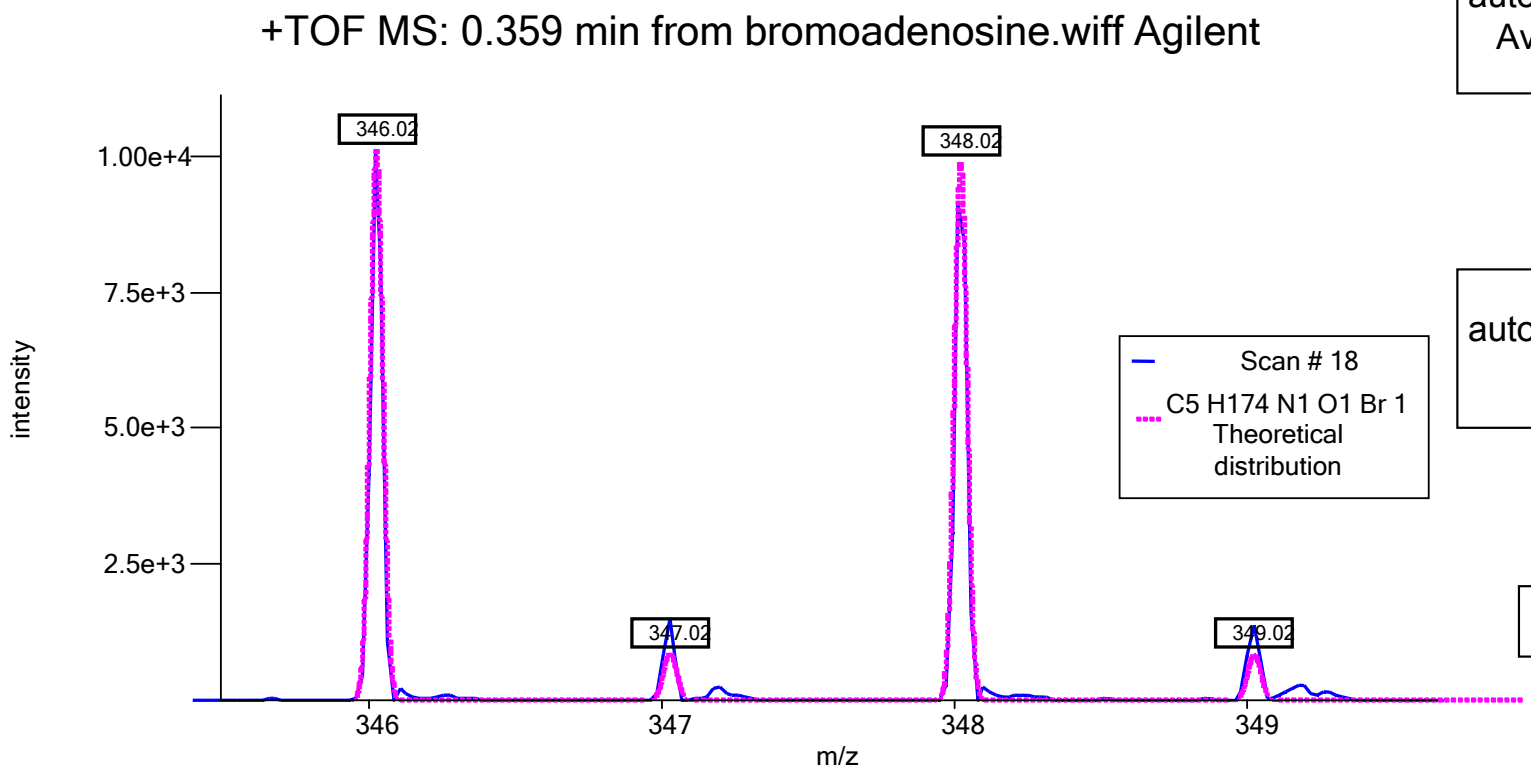


Interesting profile because of the isotopic distribution of bromine (51% 78.91833 , 49% 80.91629)



Chemical Labeling with Tags

- Specify a static tag to be applied to Averagine formula
  - Changes the Averagine formula generated



autocorrelation charge = 1
Average mass = 345.01

Subtract average mass of tag

autocorrelation charge = 1
mass = 265.1065

Calculate Averagine formula

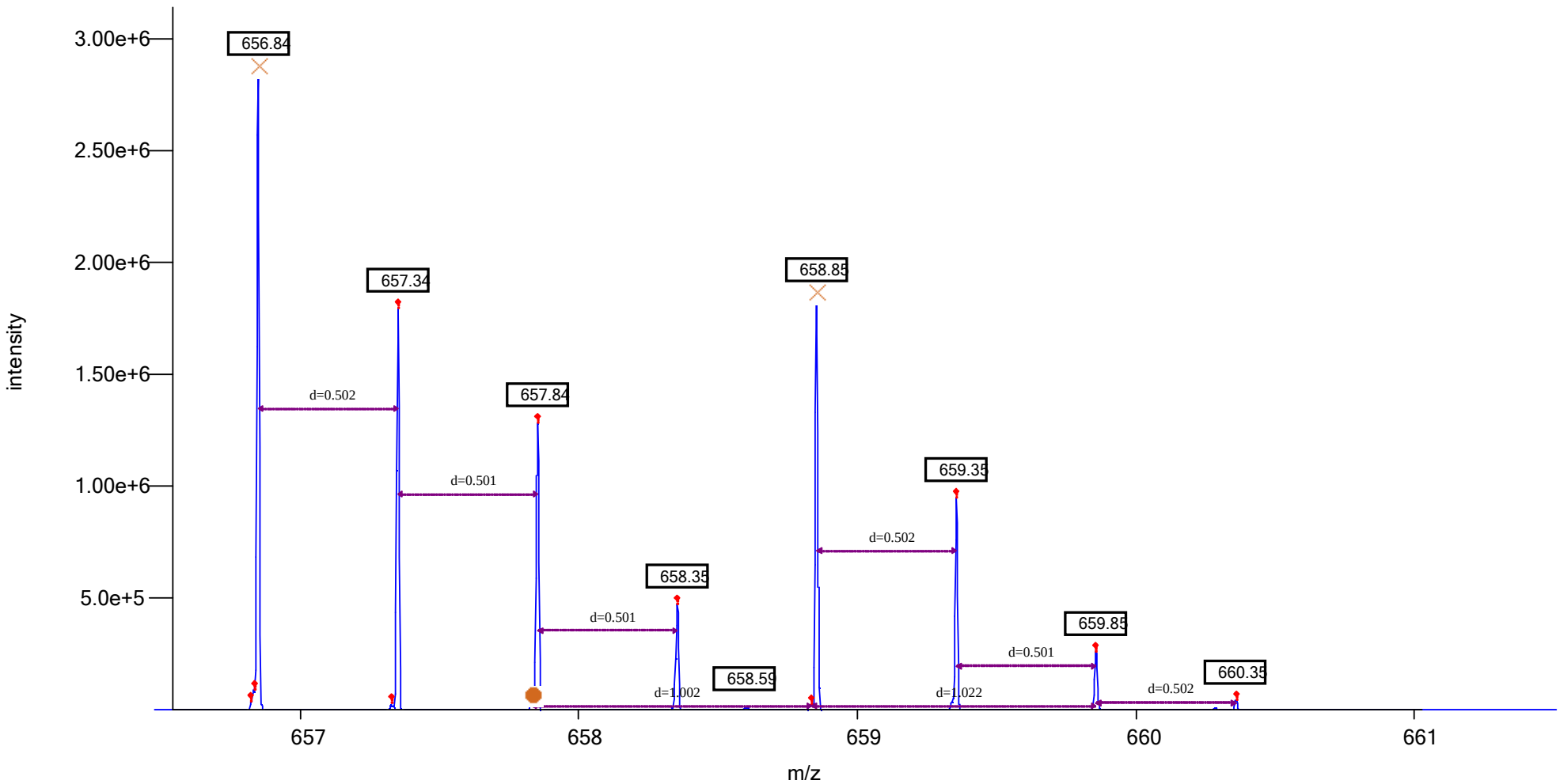
$C_5 H_{174} N_1 O_1 S_0$

Add tag formula to Averagine formula

$C_5 H_{174} N_1 O_1 S_0 Br_1$

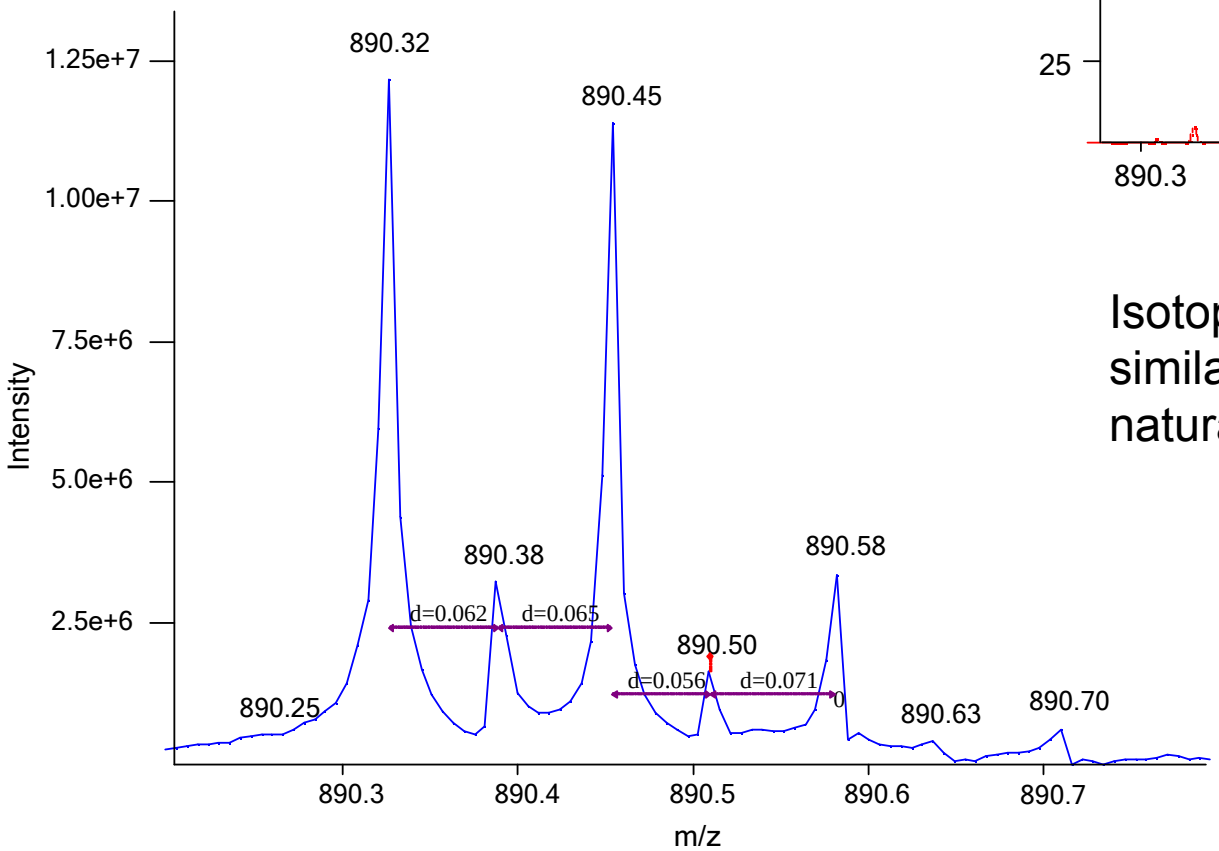
$^{16}O/^{18}O$ Mixtures

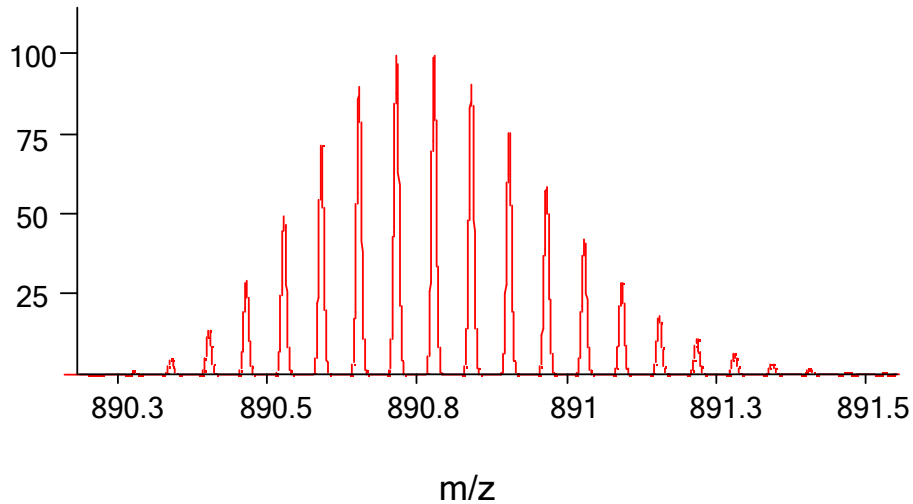
- Overlapping isotope patterns separated by 4 Da
  - If peaks exist 4 Da before current peak, those are processed first, and only the first four isotopic peaks are removed



Isotopic Composition

- Changing natural abundances
  - ^{13}C, ^{15}N depleted media – isotopic composition of atoms is different from those found in nature. Distribution of isotopes of Sulfur predominates the distribution shown below





Isotopic distribution of peptide with similar mass and charge (16+), but with natural isotopic distribution of atoms

Isotopic Composition

- Changing natural abundances

Transform Options

- Peak Picking
- Horn Transform
- Isotope Distribution
  - Averagine
  - Isotopic Composition
- Miscellaneous Options
- FTICR Preprocessing Options

Compostion

Atomicity	Element Nam	Element Symbol	Isotope Mass	Isotope Percentage
1	Hydrogen	H	1.007825	0.99985
1	Hydrogen	H	2.014102	0.00015
2	Helium	He	3.01603	1E-06
2	Helium	He	4.0026	0.999999
3	Lithium	Li	6.015121	0.075
3	Lithium	Li	7.016003	0.925
4	Berellium	Be	9.012182	1
5	Boron	B	10.01294	0.199
5	Boron	B	11.00931	0.801
6	Carbon	C	12	0.98893
6	Carbon	C	13.00336	0.01107
7	Nitrogen	N	14.00307	0.996337
7	Nitrogen	N	15.00011	0.003663
8	Oxygen	O	15.99491	0.99759

Load | Save | Save As

Helpful Tips

Controls Isotopic Composition

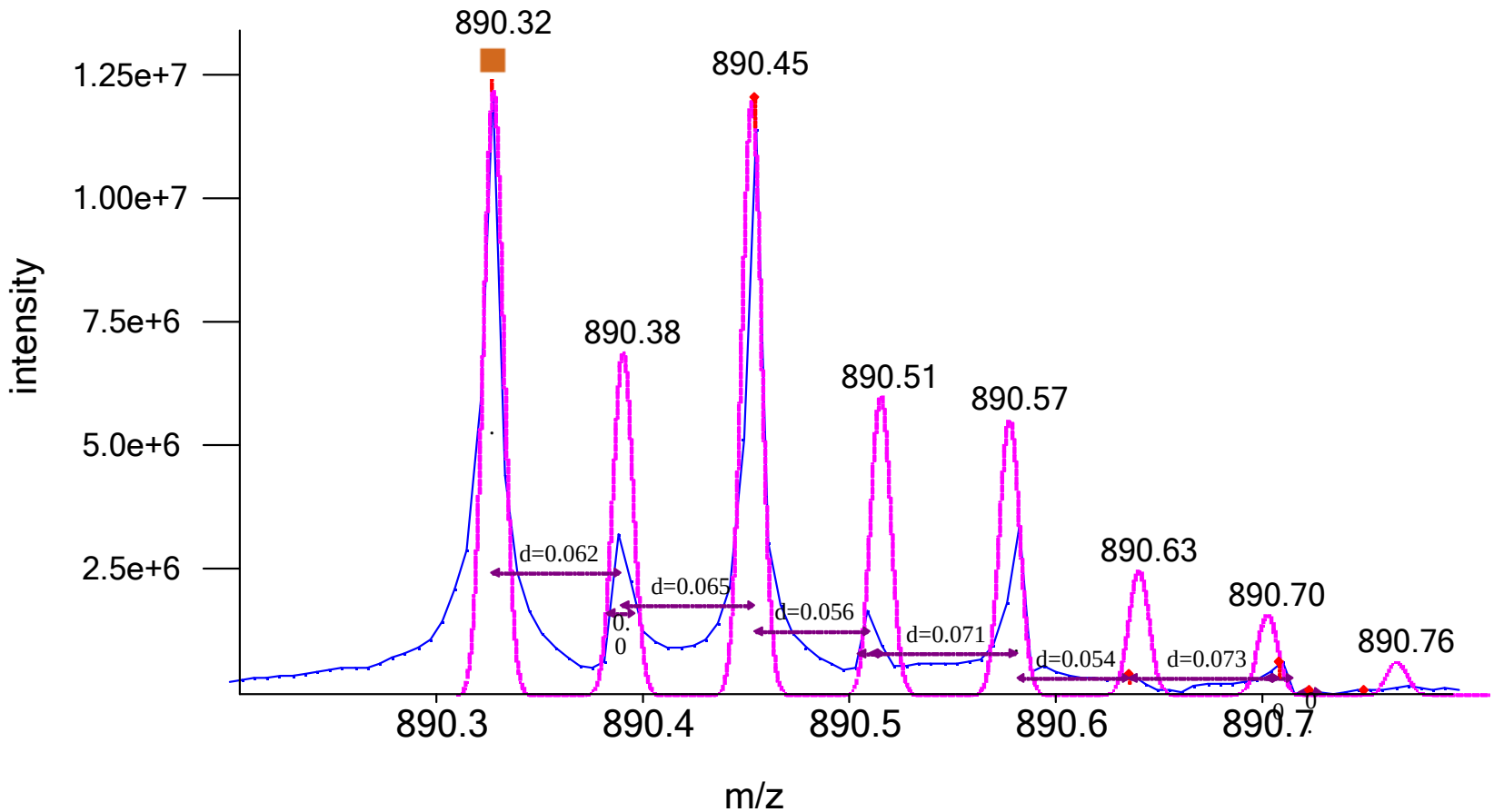
- Loads in the isotopic composition in the form of a .xml document

- Values can be edited in the grid and saved in original format or as new .xml

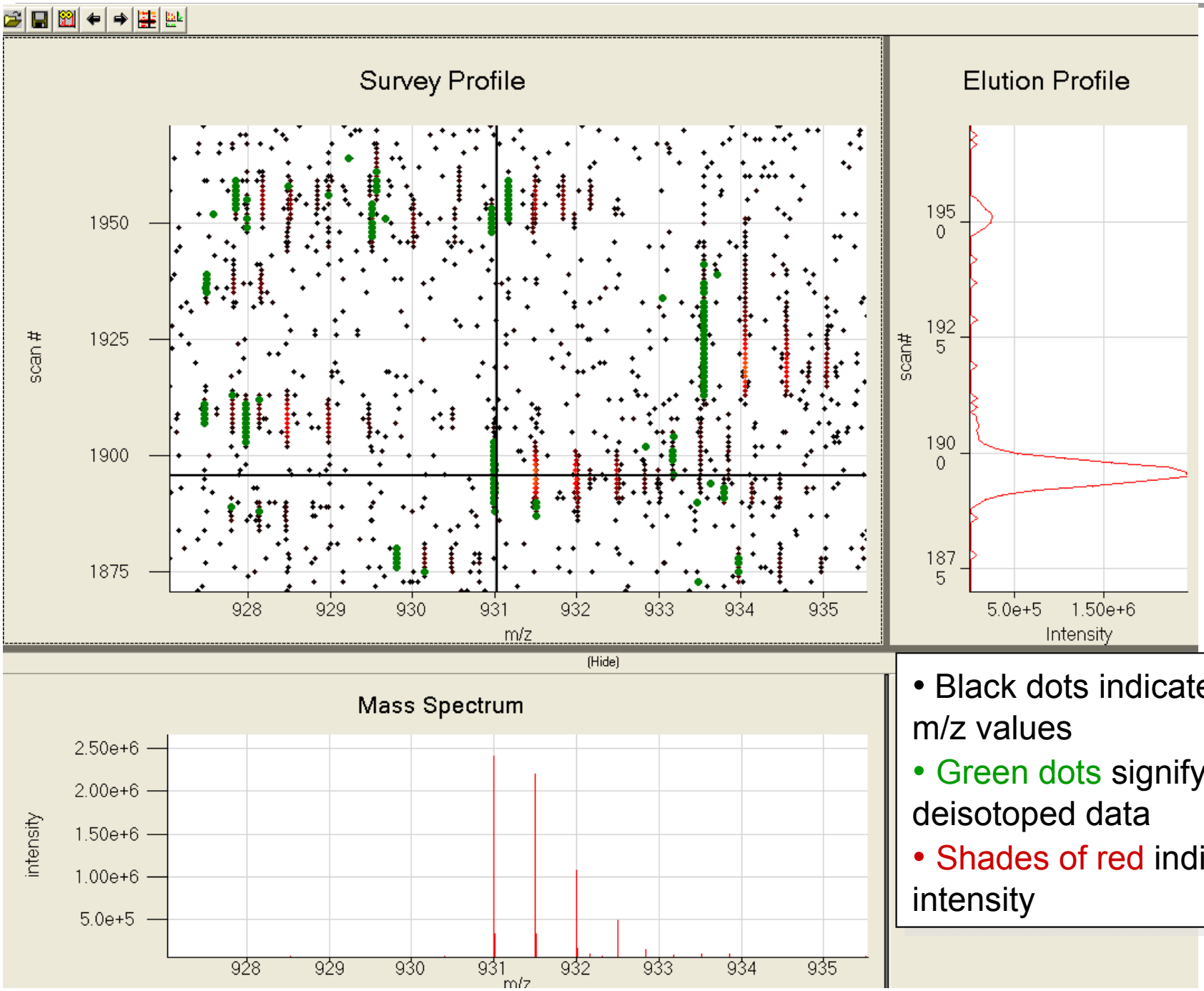
Load Parameters | Save Parameters | OK | Cancel

Isotopic Composition

- Changing natural abundances
  - Changing $^{12}C/^{13}C$, $^{14}N/^{15}N$ isotopic abundances from those in nature to approriate ones results in a better fit
  - As shown, estimated isotopic abundances were still not perfect



LC-MS Feature Discovery



- Black dots indicate individual m/z values
- Green dots signify successfully deisotoped data
- Shades of red indicate data intensity

Outline

- ✓ Introduction
- ✓ Feature discovery in LC-MS datasets
  - ✓ Feature discovery in individual spectra
  - Feature definition over elution time
- Identifying LC-MS Features using an AMT tag DB

Feature definition over elution time

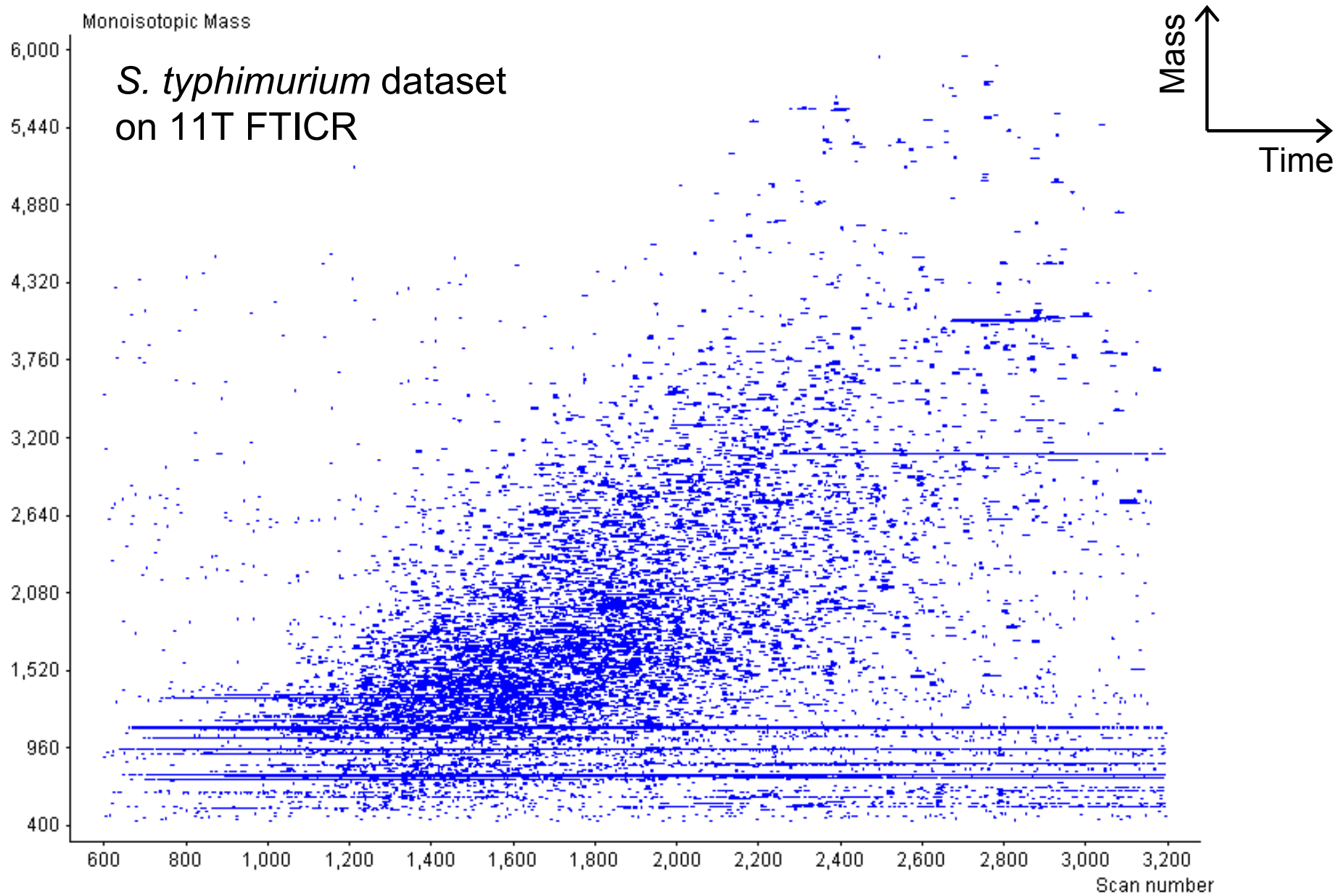
- Deisotoping collapses original data into data lists

scan num	charge	abundance	mz	fit	average mw	monoiso mw	most abu. mw	fwhm	signal noise
1500	1	2772933	759.0649	0.0716	758.5222	758.0576	758.0576	0.0106	718.83
1500	1	2614913	1103.033	0.1111	1102.698	1102.026	1102.026	0.0222	74.04
1500	1	2422829	864.4919	0.0156	864.0073	863.4846	863.4846	0.0137	74.75
1500	2	2297822	563.3253	0.012	1125.322	1124.636	1124.636	0.006	77.94
1500	1	1213607	943.9815	0.1025	943.5518	942.9742	942.9742	0.0165	120.36
1500	3	988761	675.0246	0.02	2023.375	2022.052	2023.0549	0.0086	79.22
1500	2	734070	688.392	0.0384	1375.694	1374.77	1374.7695	0.009	92.09
1500	2	663954	642.3243	0.0253	1283.417	1282.634	1282.6341	0.0076	109.01
1500	1	661477	730.1117	0.024	729.5461	729.1045	729.1045	0.0096	39.06
1500	2	630657	689.3645	0.0446	1377.64	1376.715	1376.7145	0.0088	57.52
1500	2	569896	591.8343	0.0198	1182.379	1181.654	1181.6541	0.0065	111.2
1500	2	503993	757.8854	0.0706	1513.762	1512.753	1512.7533	0.0105	80.4
1500	2	451007	936.9389	0.0296	1873.091	1871.863	1872.8662	0.0156	46.74

- Goal: Given series of deisotoped mass spectra, group related data across elution time
  - Look for repeated monoisotopic mass values in sequential spectra, allowing for missing data
  - Can also look for expected chromatographic peak shape

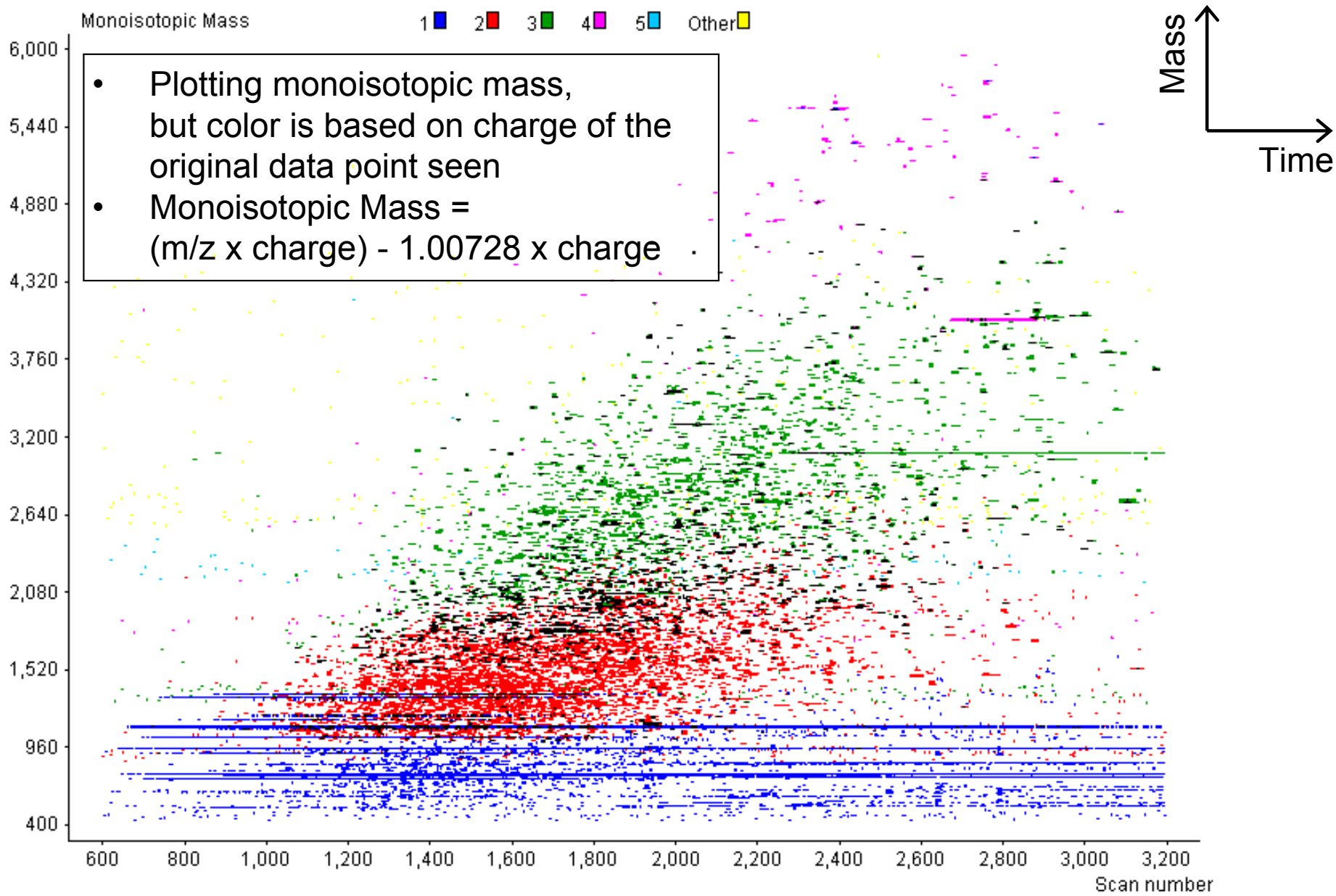
Feature definition over elution time

- Can visualize deisotoped data in two-dimensions



Feature definition over elution time

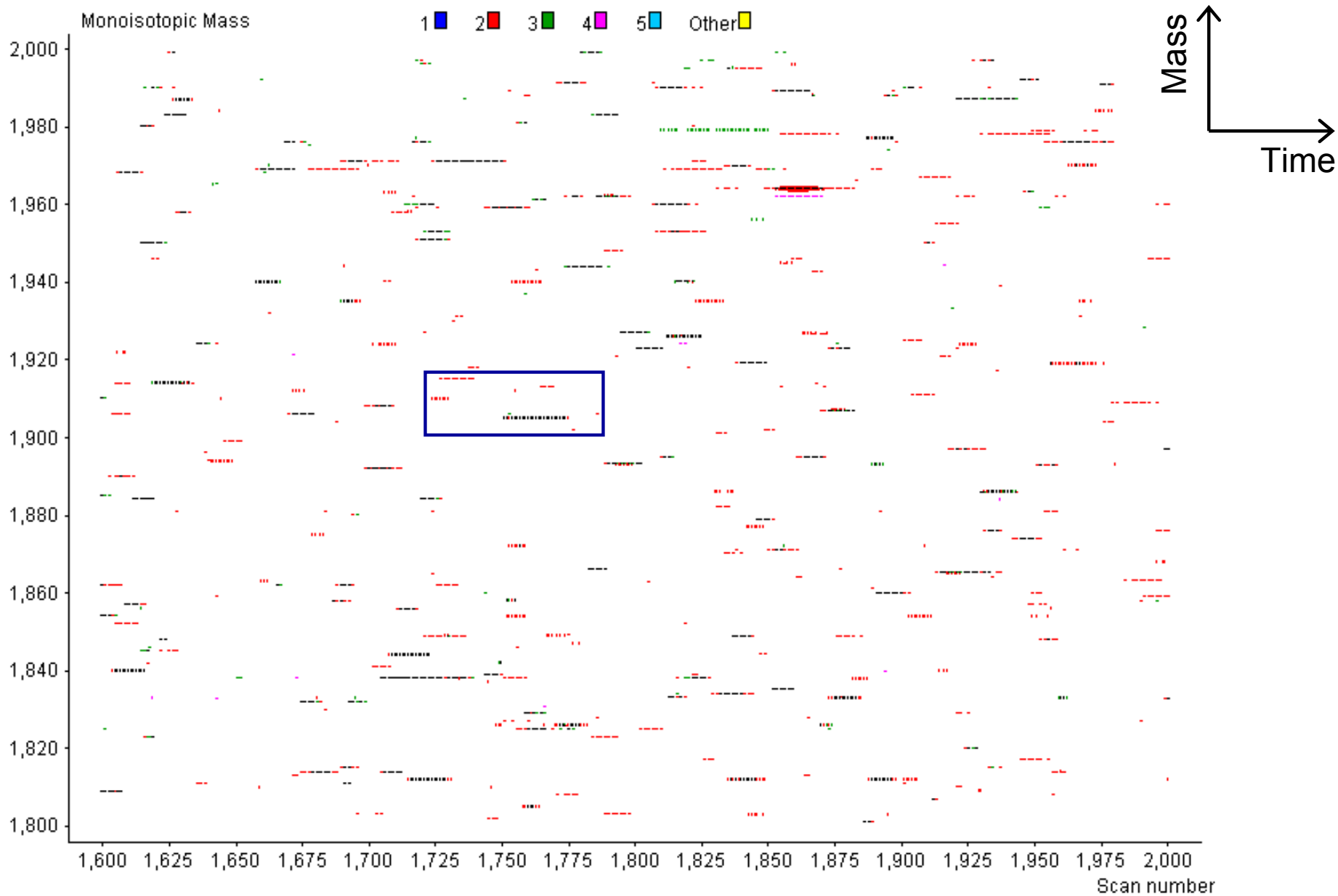
- Charge state view



- Plotting monoisotopic mass, but color is based on charge of the original data point seen
- Monoisotopic Mass = (m/z x charge) - 1.00728 x charge

Feature definition over elution time

- Zoom-in view of species

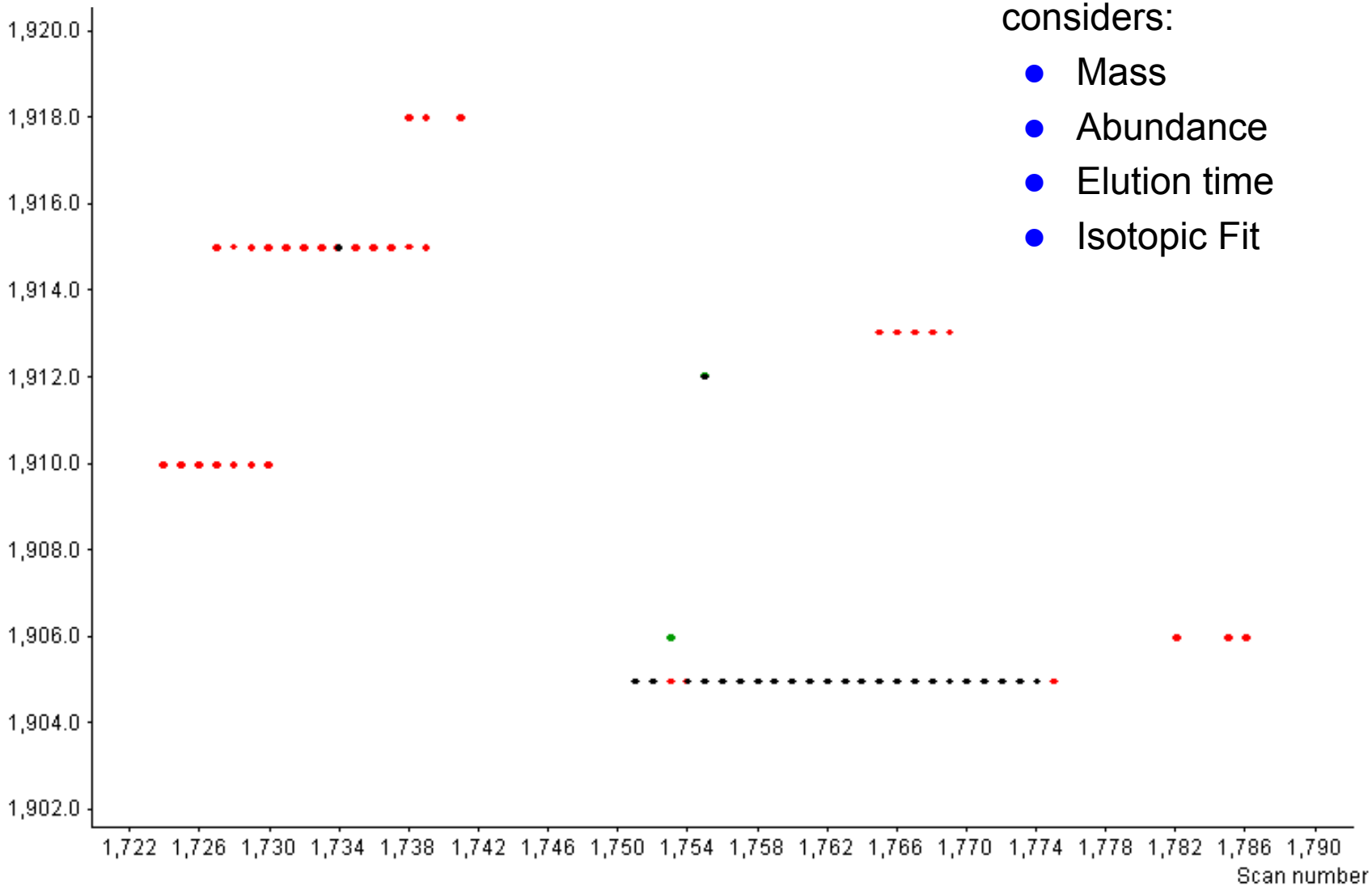


Feature definition over elution time

- Same species in multiple spectra need to be grouped together

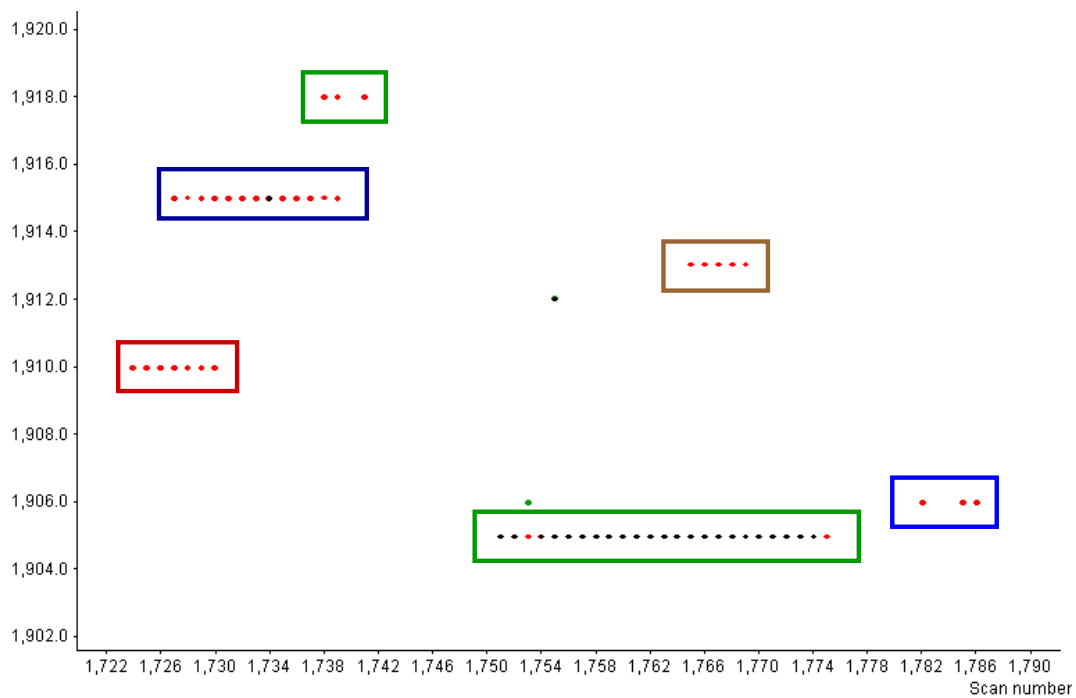
Related peaks found using a weighted Euclidean distance; considers:

- Mass
- Abundance
- Elution time
- Isotopic Fit



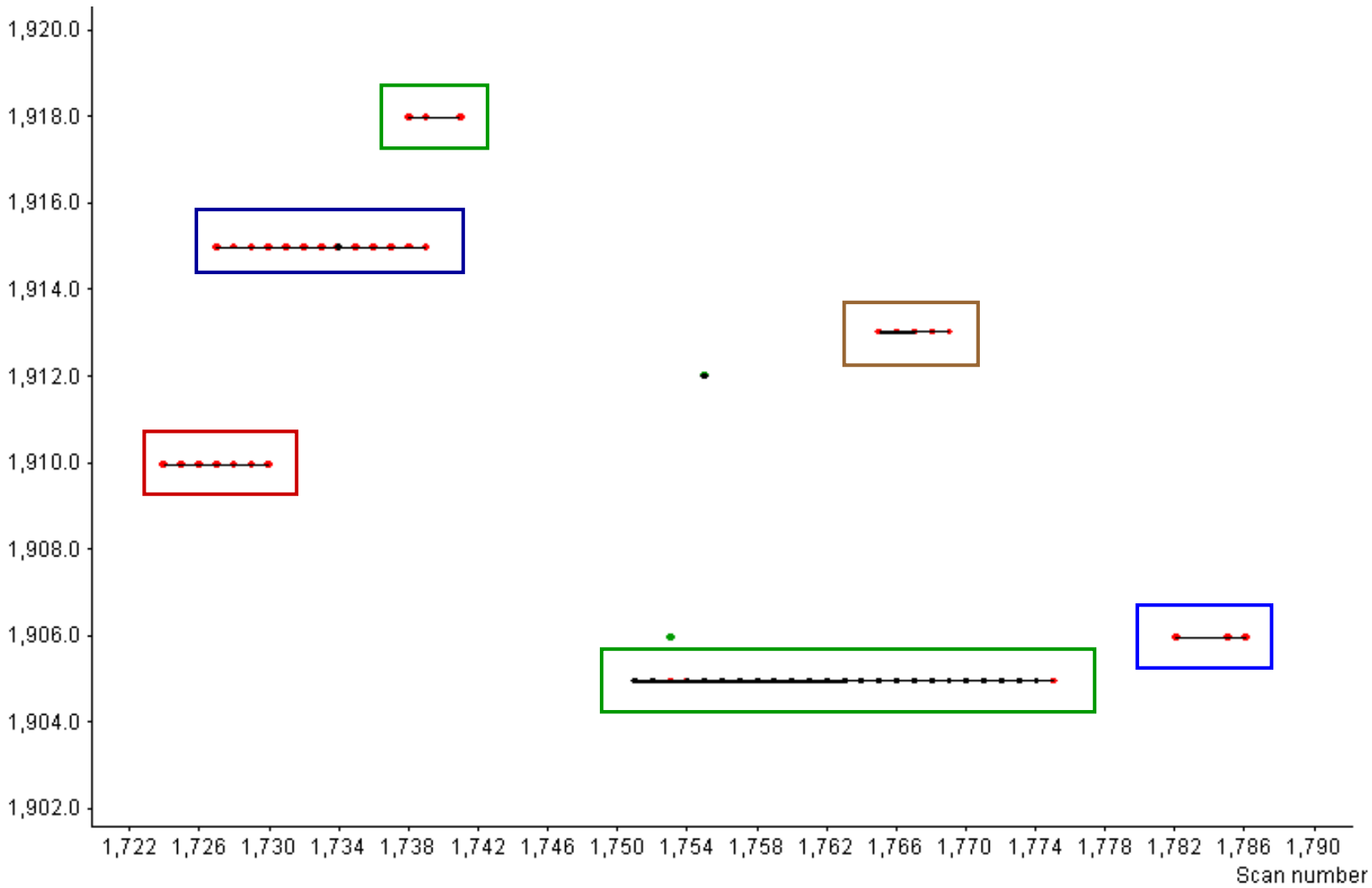
Feature definition over elution time

- Grouping uses single linkage clustering
  - Form connections between data points in n-dimensions
  - Compute the Euclidean distance between two points
    - distance = Sqrt { $[weight_{mass} \times (mass_a - mass_b)]^2 +$
      $[weight_{abu} \times (LogAbu_a - LogAbu_b)]^2 +$
      $[weight_{ET} \times (ET_a - ET_b)]^2 +$
      $[weight_{fit} \times (fit_a - fit_b)]^2$ }
  - If distance < threshold, combine points together



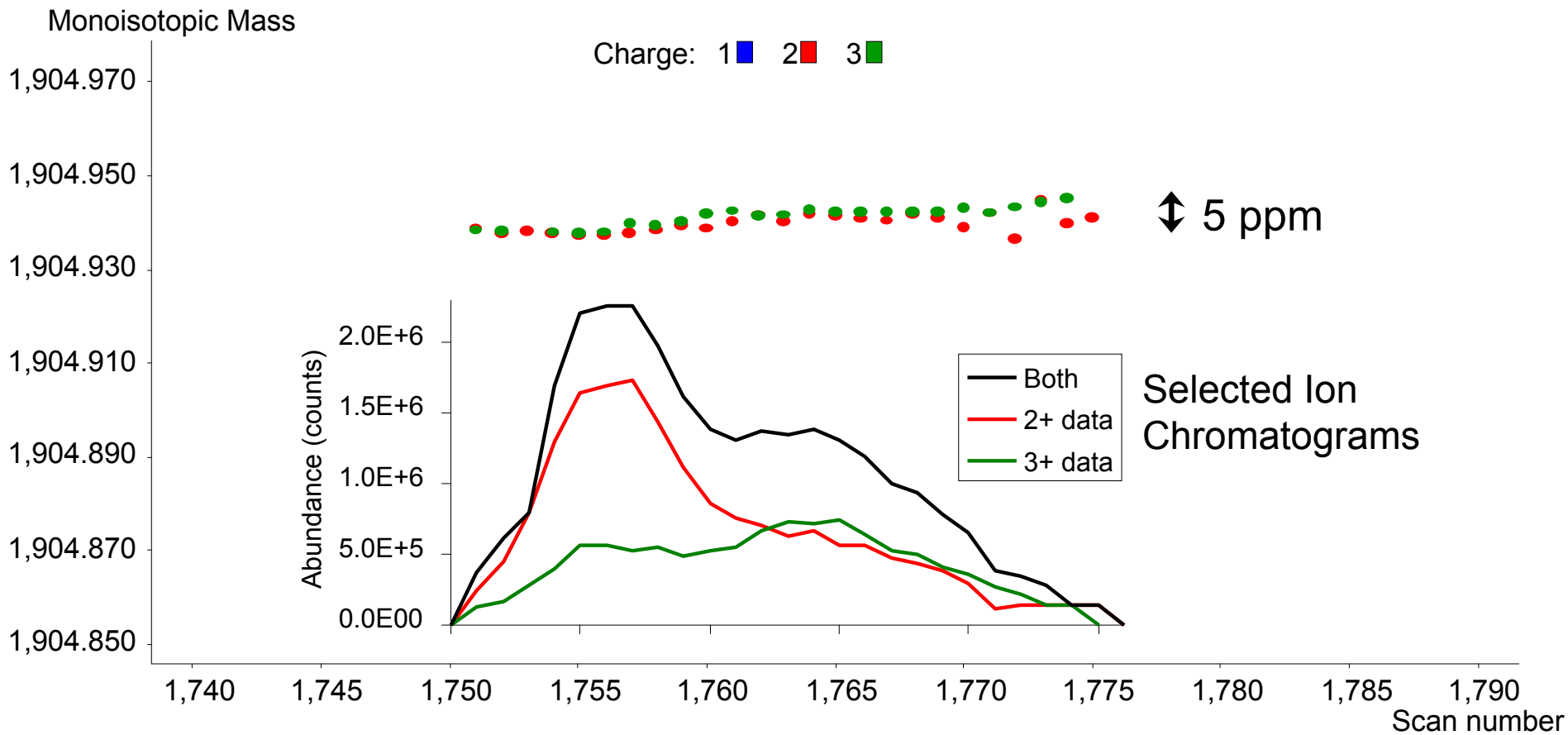
Feature definition over elution time

- Determine 6 separate groups
  - Typically require 2 or 3 points per group



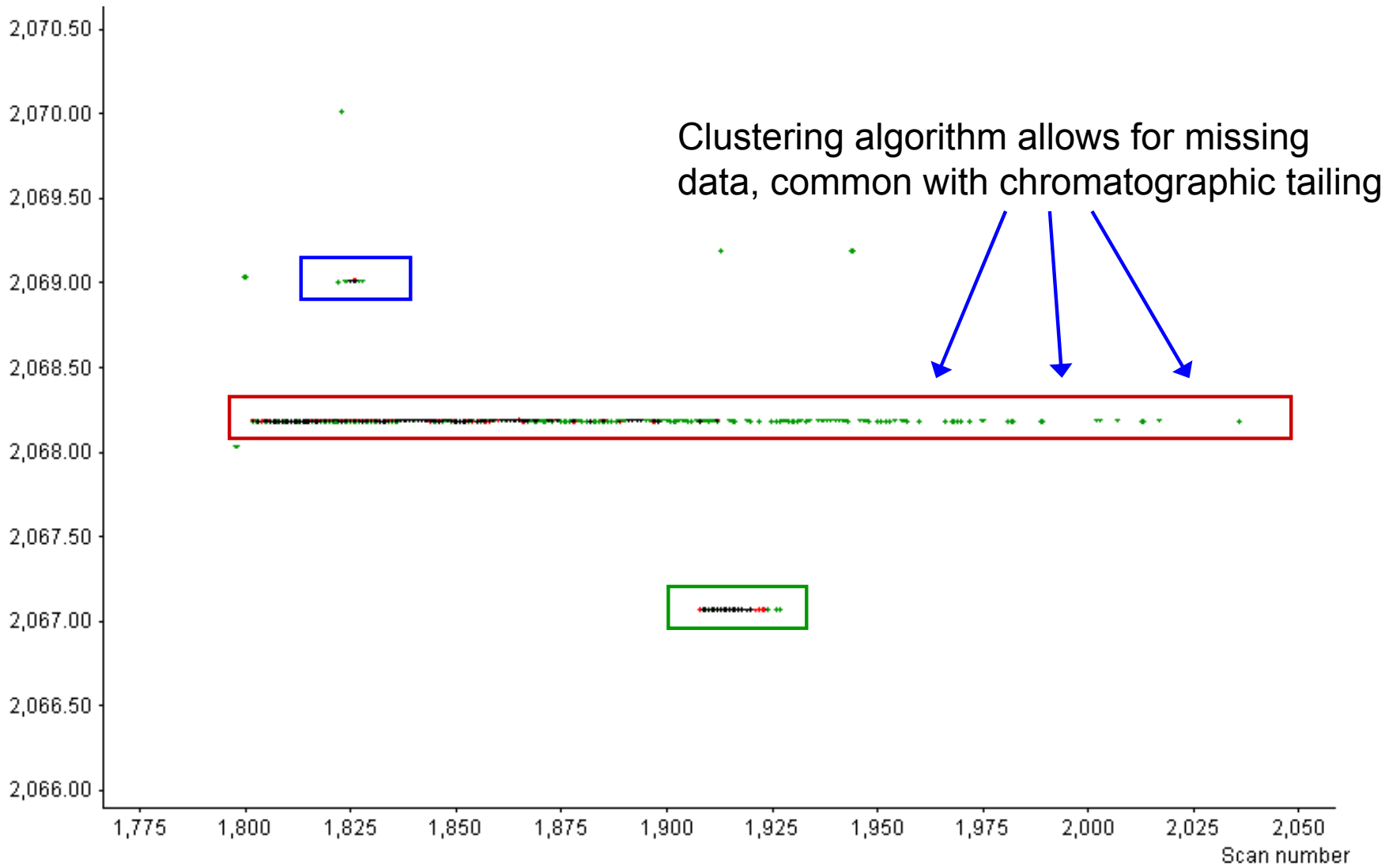
Feature definition over elution time

- Feature detail
  - Median Mass: 1904.9399 Da (more tolerant to outliers than average)
  - Elution Time: Scan 1757 (0.363 NET)
  - Abundance: 1.7×10^7 counts (area under 2+ SIC)
    - See both 2+ and 3+ data
    - Stats typically come from the most abundant charge state



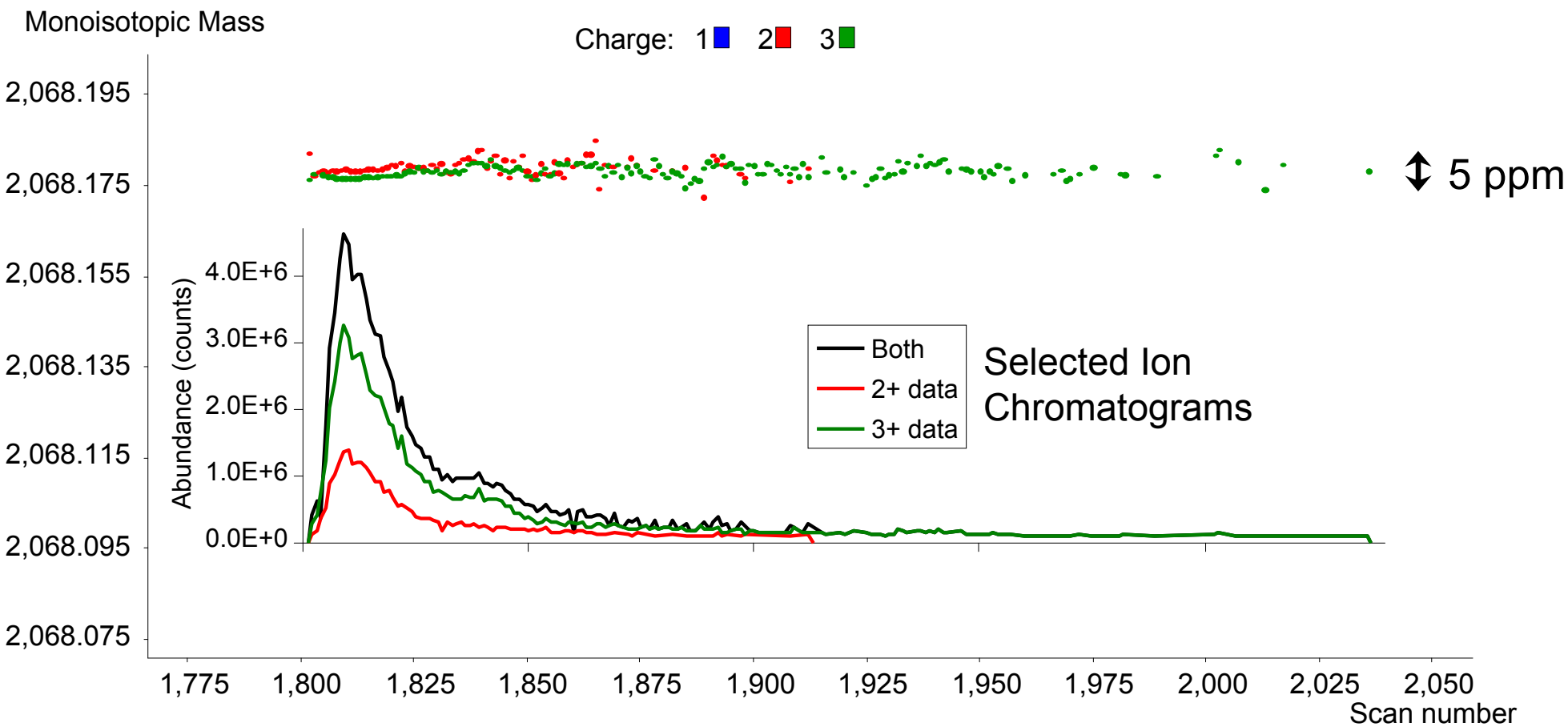
Feature definition over elution time

- Second example
  - LC-MS feature eluting over 7.5 minutes



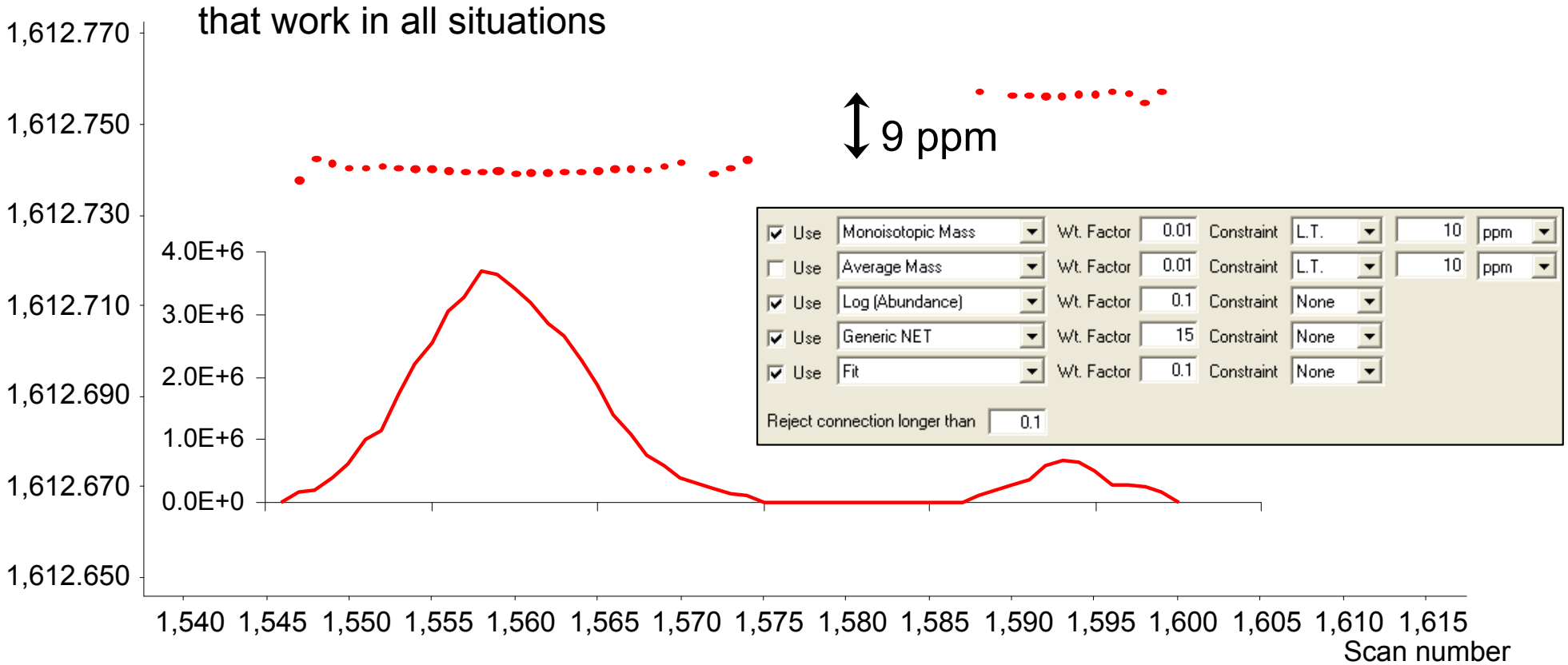
Feature definition over elution time

- Second example, feature detail
  - Median Mass: 2068.1781 Da
  - Elution Time: Scan 1809 (0.380 NET)
  - Abundance: 8.7×10^7 counts (area under 3+ SIC)
    - This example has primarily 3+ data; previous had even mix of 2+ and 3+ data



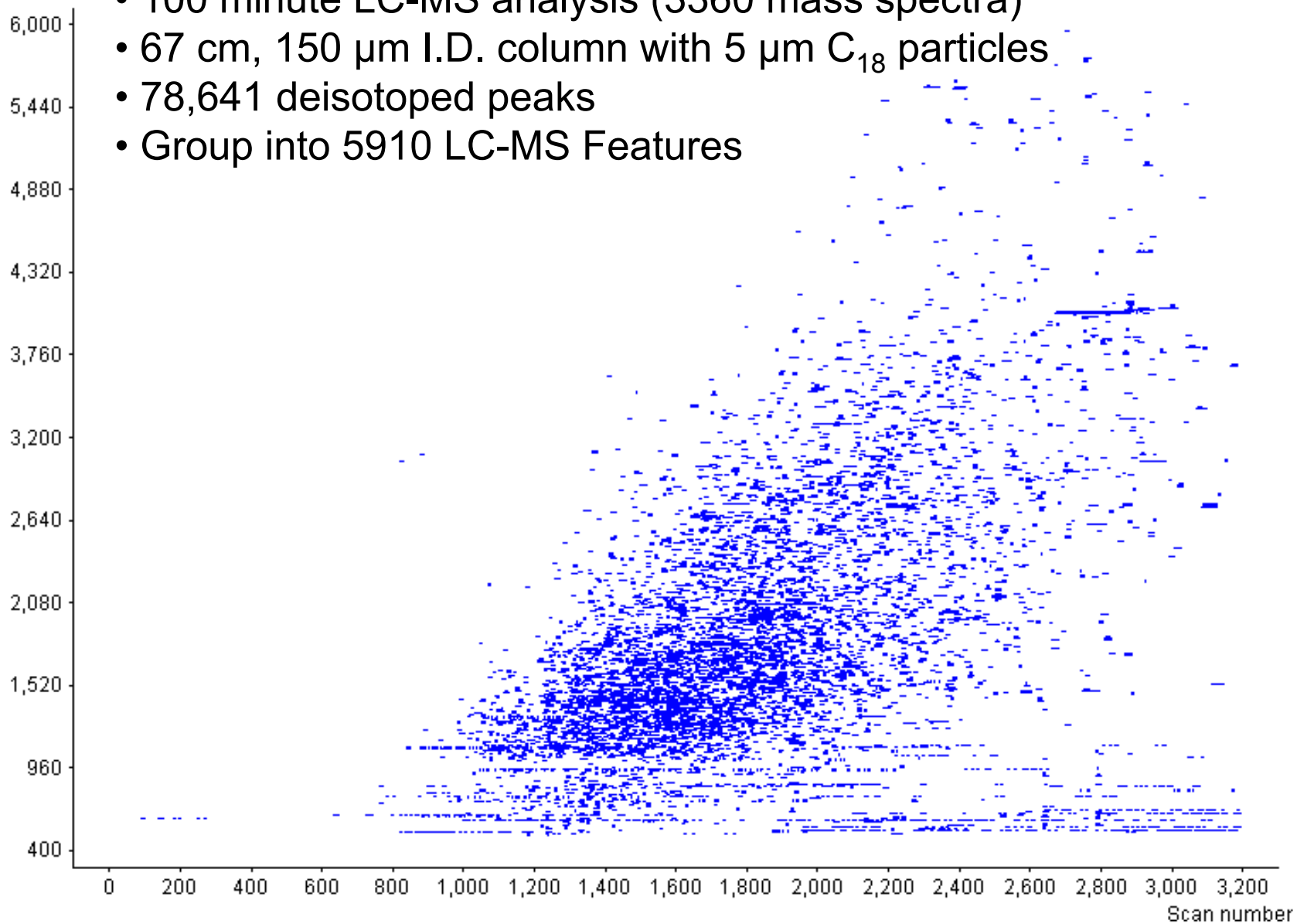
Feature definition over elution time

- Refining the features
  - Require data spans at least 3 spectra
  - Exclude grouped feature if it is too long (e.g. ≥ 15% of dataset)
  - Sometimes the Euclidean distance results in undesirable clustering
    - Split if elution profile indicates two or more entities with a mass difference ≥ threshold (e.g. 4 ppm)
    - Necessary since hard to define clustering weights and distance constraints that work in all situations



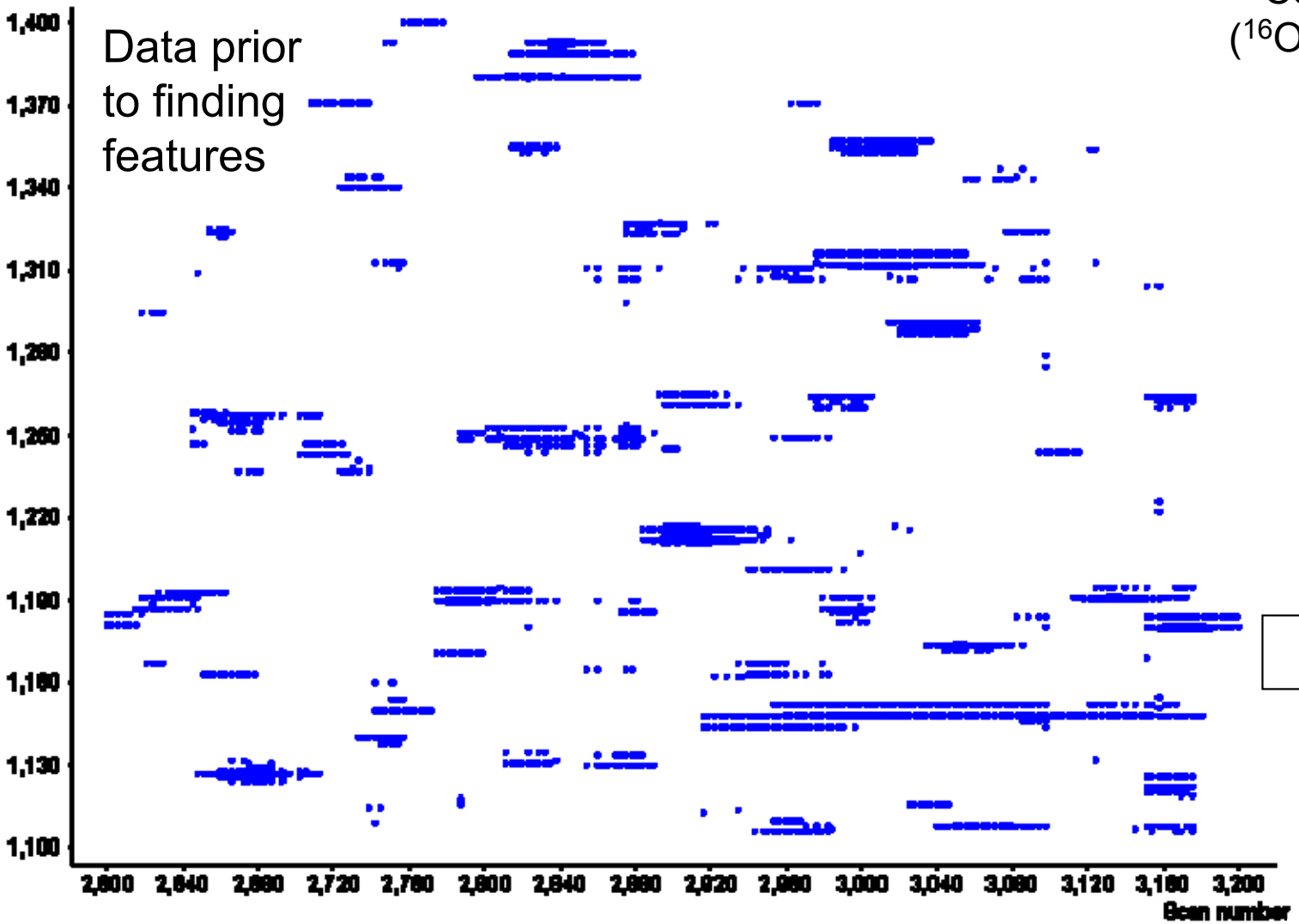
Feature definition over elution time

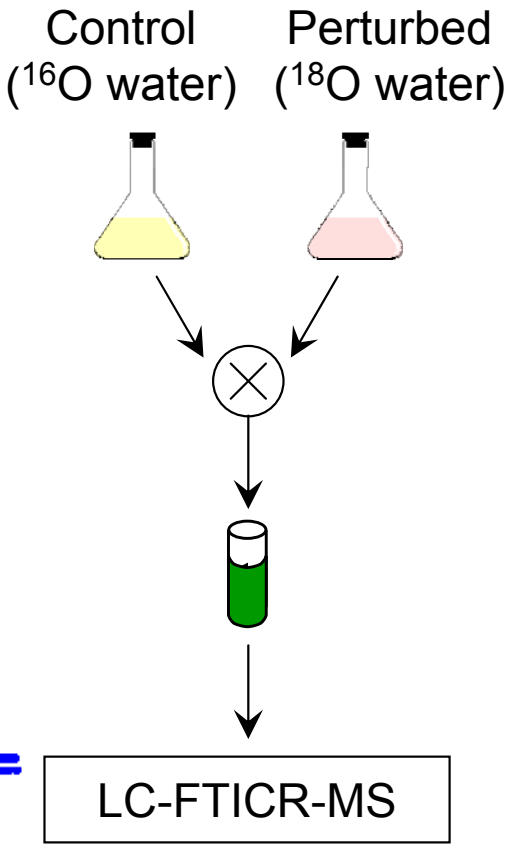
- Example: *S. typhimurium* dataset on 11T FTICR
  - 100 minute LC-MS analysis (3360 mass spectra)
  - 67 cm, 150 µm I.D. column with 5 µm C_{18} particles
  - 78,641 deisotoped peaks
  - Group into 5910 LC-MS Features



Isotopic Pairs Processing

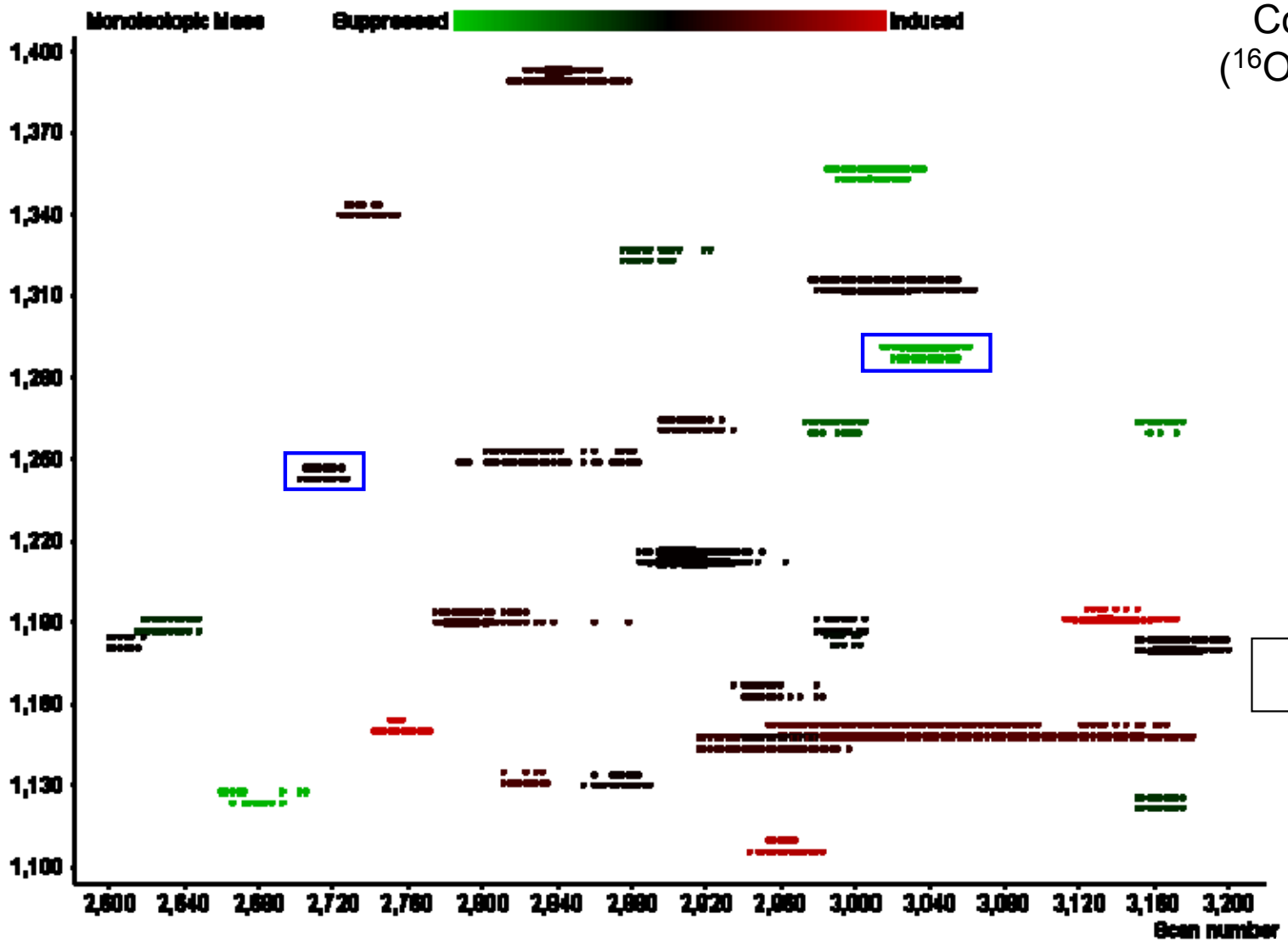
- Paired features typically have identical sequences, with and without an isotopic label
  - e.g. $^{16}O/^{18}O$ pairs or $^{14}N/^{15}N$ pairs

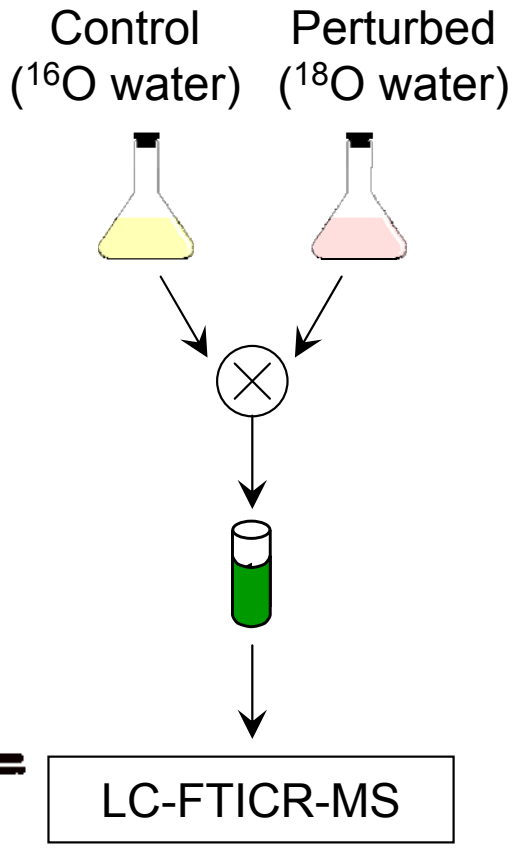




Isotopic Pairs Processing

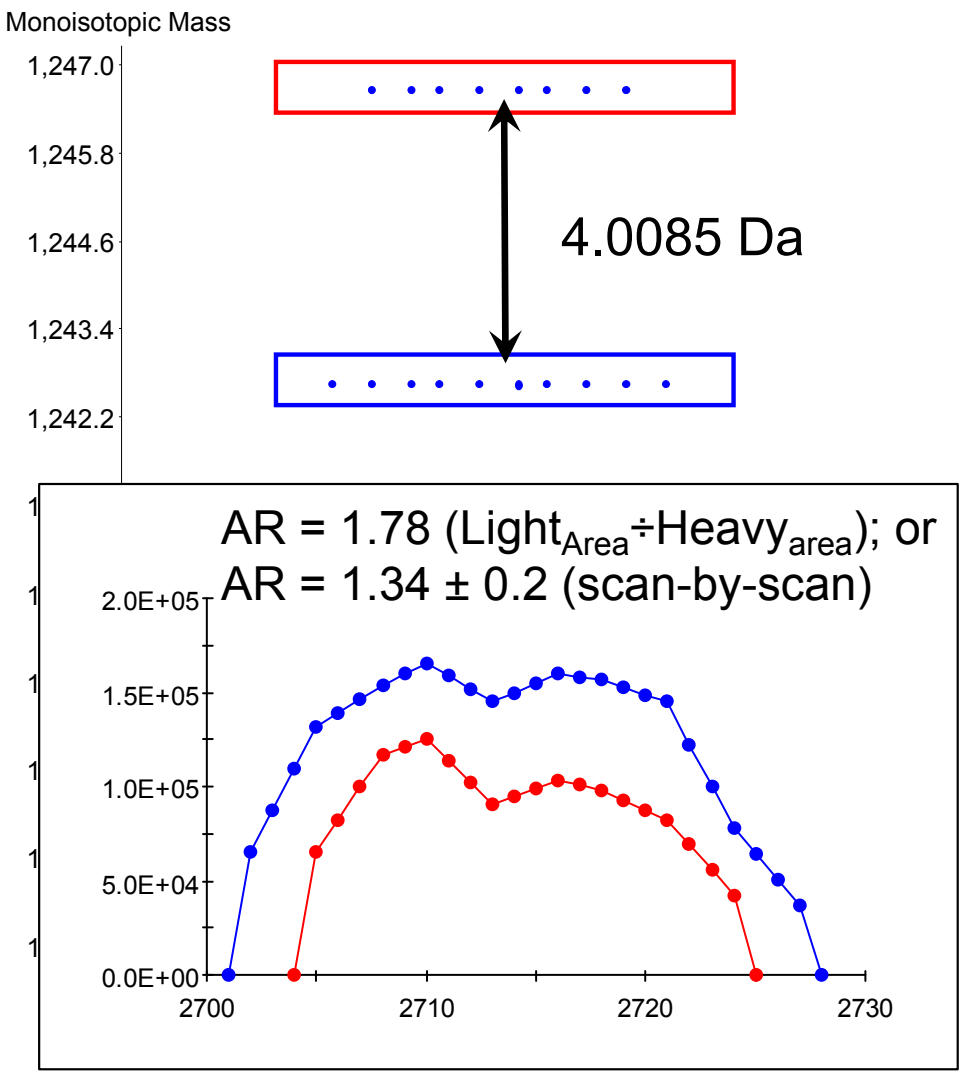
- Data after finding paired features
  - 4 Da pair spacing due to incorporation of two ^{18}O atoms

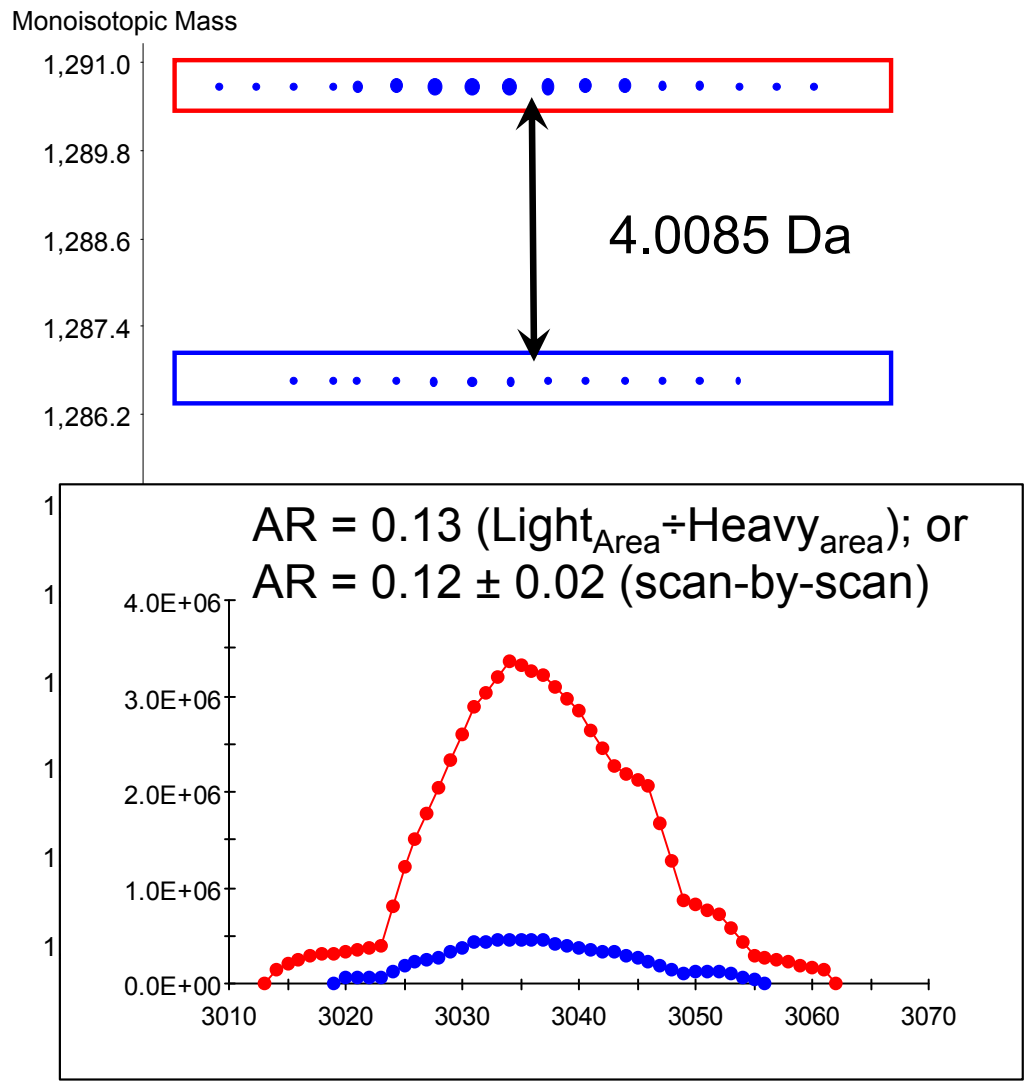




Isotopic Pairs Processing

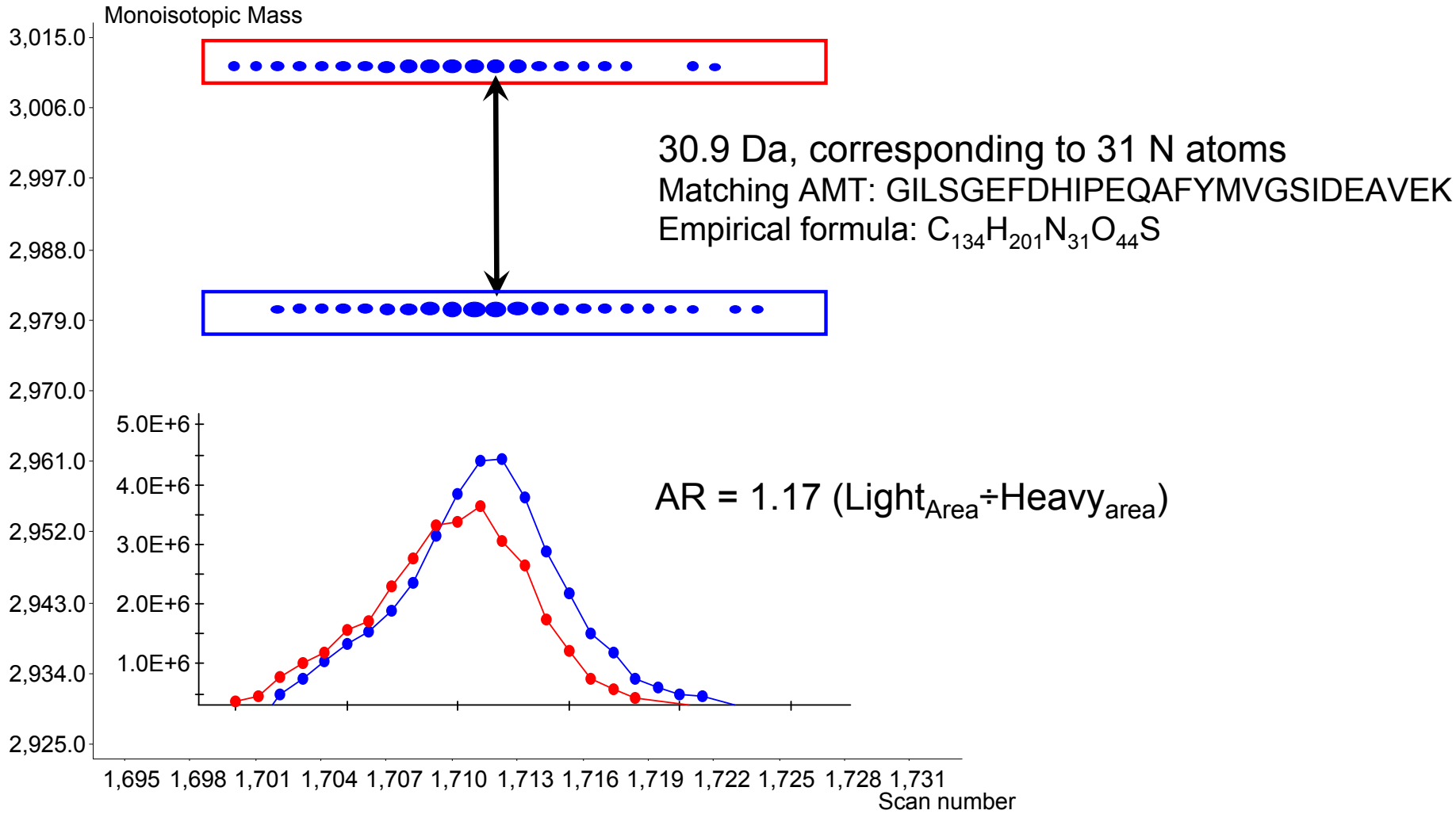
- Paired feature example: $^{16}O/^{18}O$ data
  - Compute AR using ratio of areas, or
  - Compute AR scan-by-scan, then average AR values (members must co-elute)





Isotopic Pairs Processing

- Paired feature example: $^{14}N/^{15}N$ data
  - Pair members often do not co-elute
  - Use bulk area ratio, or re-align pair members then compute AR scan-by-scan



Feature definition over elution time

- Numerous options for clustering data to form LC-MS features and for finding paired features

LC-MS Feature (UMC) Ion Networks

1. Find Connections | 2. Edit/Filter Connections | 3. Define LC-MS Features using Connections

Definition Scope
All Data Points
Current View

Metric Type: Euclidean
Set to Old Defaults | Set to Defaults

☑ Use	Monoisotopic Mass	Wt. Factor	0.01	Constraint	L.T.	10	ppm
☐ Use	Average Mass	Wt. Factor	0.01	Constraint	L.T.	10	ppm
☑ Use	Log (Abundance)	Wt. Factor	0.1	Constraint	None		
☑ Use	Generic NET	Wt. Factor	15	Constraint	None		
☑ Use	Fit	Wt. Factor	0.1	Constraint	None		

Reject connection longer than 0.1

LC-MS Feature (UMC) Ion Networks

1. Find Connections | 2. Edit/Filter Connections | 3. Define LC-MS Features using Connections

LC-MS Feature Stats
Class Representative: Highest Abundance
Class Abundance: Sum of Class Abu.
Class Mass: Class Median
Most Abu Charge State Group Type: Highest Abu Sum
☑ Use most abundant charge state group stats for class stats
☐ Make single member classes
FeatureDraw Type: Actual LC-MS Feature

Auto-Refine Options | Split Features Options | Adv Class Stats

☐ Remove low intensity classes 30 %
☐ Remove high intensity classes 30 %
☑ Remove cls. with less than 3 scans
☐ Remove cls. with length over 400 scans
☑ Remove cls. with length over 15 % all scans
Percent max abu for gauging width 33 %
☑ Test feature length using scan range — Minimum member count: 3

☑ Interpolate abundances across gaps
Maximum size of gap to interpolate: 4

Set to Defaults
Report
Find LC-MS Features

LC-MS Feature Delta Pairing Analysis

Function Report

Find Pairs | Set to Defaults

Delta Mass Options
Delta: 4.0085
☐ Calculate N14/N15 Min/Max Deltas from class molecular mass
Min Deltas: 1 Max Deltas: 1
Delta count step size: 1
Set to N14/N15 | Set to C12/C13 | Set to O16/O18 | Set to Deuterium

Tolerance Options
Pair Tolerance: 0.02 Da Scan Tolerance:
☑ Require pair-classes overlap at feature edges 15
Note: Even if two LC-MS Features do not overlap at the edges, if one feature is completely within a second feature, then pairing is allowed; to prevent this, enable overlap at LC-MS Feature apexes
☑ Require pair-classes overlap at feature apexes 15

Inclusion/Exclusion Options
ER Inclusion Range: -5 to 5
☑ Ambiguous pairs exclusion keeps most confident pair

Pair Search and ER Calculation Options
☑ Require matching charge states for pair
☑ Use identical charge states for expression ratio
☑ Average ER's for all charge states — Weight by Abu
☑ Compute ER Scan by Scan
☑ Enable I-Report ER computation
☑ Remove outlier ER values using Grubb's test (95% conf.)
☑ Repeatedly remove outliers
☑ Use symmetric ERs
Minimum final data point count 3

Outline

- ✓ Introduction
- ✓ Feature discovery in LC-MS datasets
  - ✓ Feature discovery in individual spectra
  - ✓ Feature definition over elution time
- Identifying LC-MS Features using an AMT tag DB

Assembling an AMT tag DB

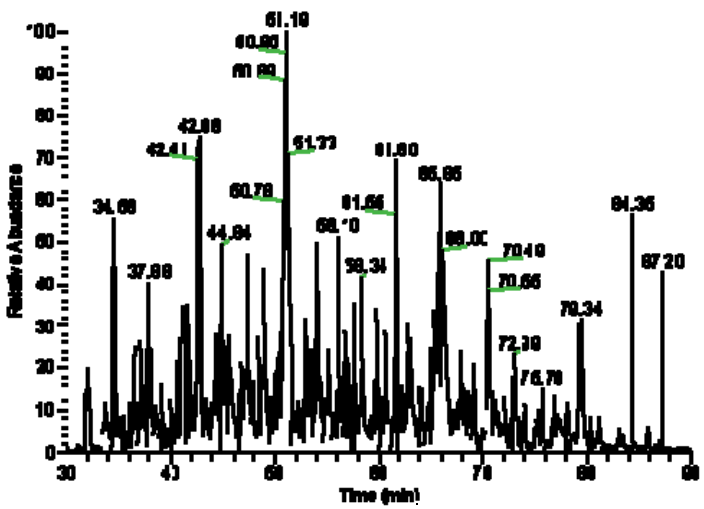
- Accurate Mass and Time (AMT) tag
  - Unique peptide sequence whose monoisotopic mass and normalized elution time are accurately known
  - AMT tags also track any modified residues in peptide
- AMT tag DB
  - Collection of AMT tags
- AMT tag approach articles
  - R.D. Smith et. al. *Proteomics* **2002**, *2*, 513-523.
  - J.D.S. Zimmer, M.E. Monroe et. al., *Mass Spec. Reviews* **2006**, *25*, 450-482.
  - L. Shi, J.N. Adkins, et. al., *J. of Biological Chem.* **2006**, *281*, 29131-29140.

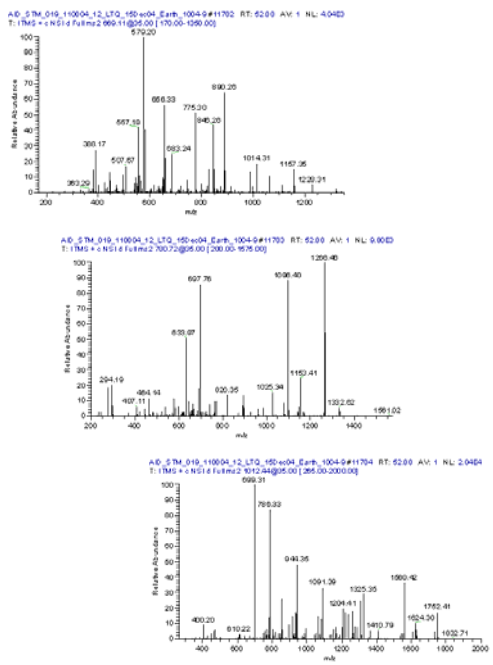
Assembling an AMT tag DB

- What can we use an AMT tag DB for?
  - Query LC-MS/MS data to answer questions
    - How many distinct peptides were observed passing filter criteria?
    - Which peptides were observed most often by LC-MS/MS?
    - How many proteins had 2 or more partially or fully tryptic peptides?
  - Correlate LC-MS features to the AMT tags
    - Analyze multiple, related samples by LC-MS using a high mass accuracy mass spectrometer
      - e.g. Time course study, 5 data points with 3 points per sample
    - Characterize the LC-MS features
      - Deisotope to obtain monoisotopic mass and charge
      - Cluster in time dimension to obtain abundance information
    - Match to AMT tags to identify peptides
      - Align in mass and time dimensions
      - Match mass and time of LC-MS features to mass and time of AMT tags

Assembling an AMT tag DB

- Characterizing AMT tags
  - Analyze samples by LC-MS/MS
    - 10 minute to 180 minute LC separations
    - Obtain 1000's of MS/MS fragmentation spectra for each sample
  - Analyze spectra using SEQUEST, X!Tandem, etc.
    - SEQUEST: http://www.thermo.com/bioworks/
    - X!Tandem: http://www.thegpm.org/TANDEM/index.html
      R. Craig and R.C. Beavis, *Bioinformatics* **2004**, *20*, 1466-1467.
  - Collate results

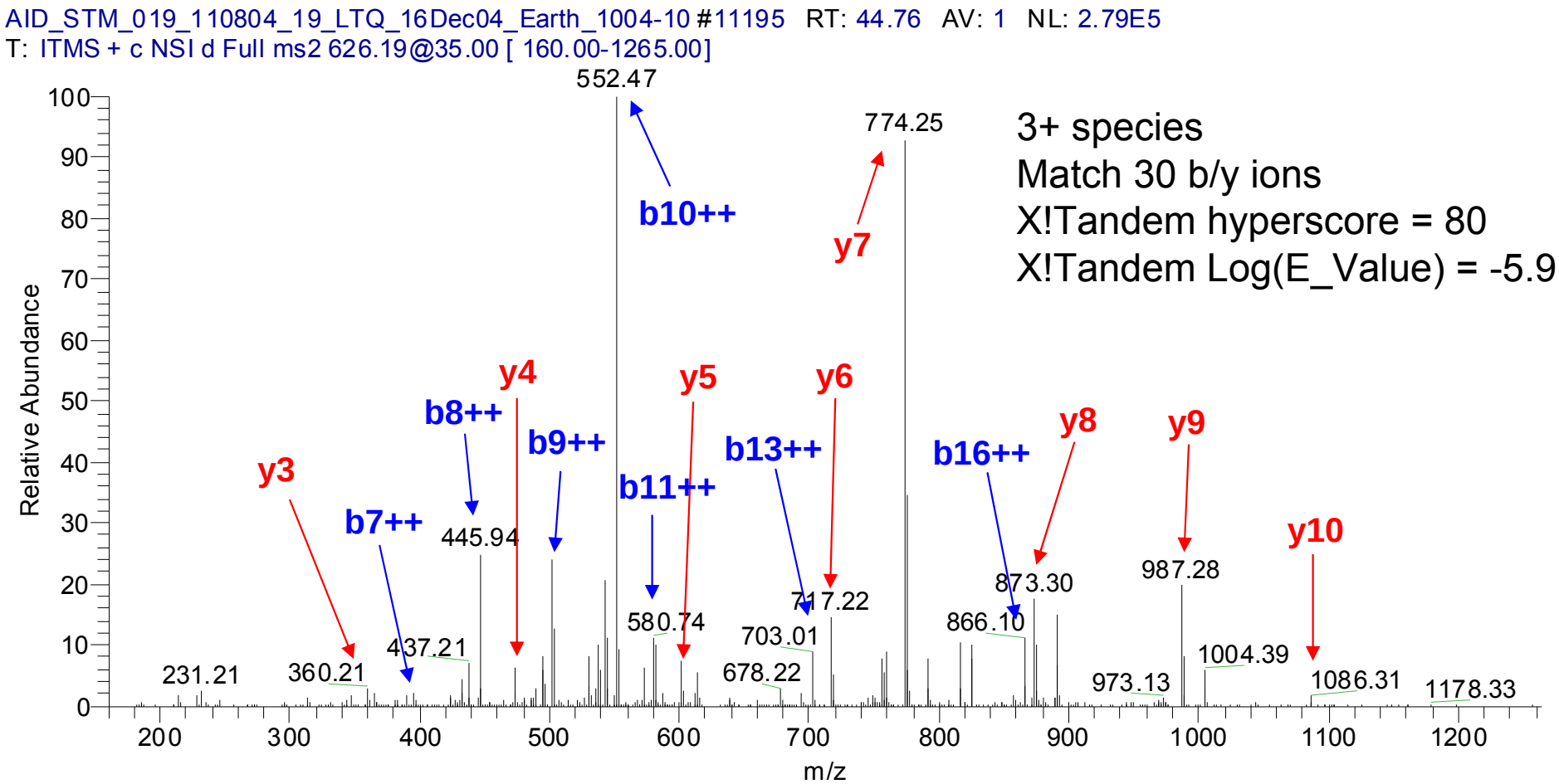




List of peptide and protein matches

Assembling an AMT tag DB

- AMT tag example
  - R.VKHPSEIVNVGDEINVK.V
  - Observed in scan 11195 of dataset #19 in an SCX fractionation series



Assembling an AMT tag DB

- AMT tag example
  - R.VKHPSEIVNVGDEINVK.V
  - Observed in scan 11195 of dataset #19 in an SCX fractionation series

#	Immon.	b	b++	Seq.	y	y++	#
1	72.08			V	1877.01	939.01	17
2	101.11	228.17		K	1777.94	889.48	16
3	110.07	365.23		H	1649.85	825.43	15
4	70.07	462.28		P	1512.79	756.90	14
5	60.04	549.31		S	1415.74	708.37	13
6	102.06	678.36	339.68	E	1328.71	664.86	12
7	86.10	791.44	396.22	I	1199.66	600.33	11
8	72.08	890.51	445.76	V	1086.58	543.79	10
9	87.06	1004.55	502.78	N	987.51	494.26	9
10	72.08	1103.62	552.31	V	873.47	437.24	8
11	30.03	1160.64	580.83	G	774.40	387.70	7
12	88.04	1275.67	638.34	D	717.38	359.19	6
13	102.06	1404.71	702.86	E	602.35	301.68	5
14	86.10	1517.80	759.40	I	473.31		4
15	87.06	1631.84	816.42	N	360.22		3
16	72.08	1730.91	865.96	V	246.18		2
17	101.11			K	147.11		1

3+ species
Match 30 b/y ions
X!Tandem hyperscore = 80
X!Tandem Log(E_Value) = -5.9

Assembling an AMT tag DB

- Align related datasets using elution times of observed peptides
  - One option: utilize NET prediction algorithm to create theoretical dataset to align against
    - NET prediction uses position and ordering of amino acid residues to predict normalized elution time

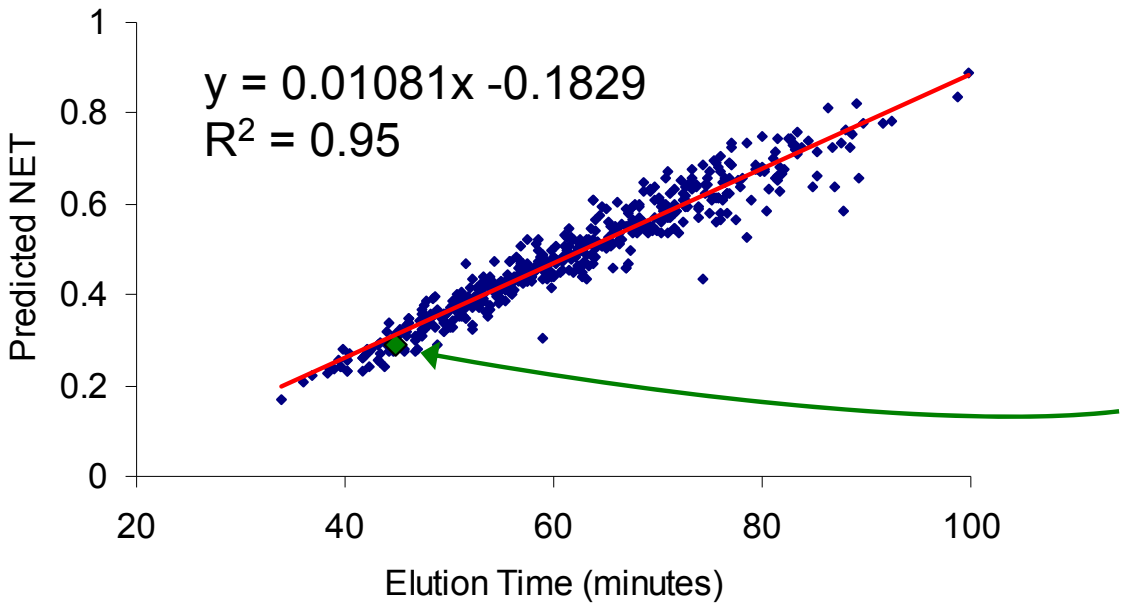
Peptide	X!Tandem Log (E_Value)	Elution Time	Predicted NET
R.AARPAKYSYVDENGETK.T	-6.1	33.958	0.167
R.LVHGEEGLVAAKR.I	-8.8	36.915	0.224
R.GIIKVGEEVEIVGIK.E	-8.2	53.003	0.415
K.RFNDDGPILFIHTGGAPALFAYHPHV.-	-7.3	62.583	0.519
K.KTGVLAQVQEALKGLDVR.E	-11.6	62.803	0.438
R.KVAAQIPNGSTLFIDIGTTPEAVAHALLGHSNLR.I	-8.9	73.961	0.589
R.TFAISPGHMNQLRAESIPEAVIAGASALVLTSYLVR.C	-6.5	88.043	0.764

K. Petritis, L.J. Kangas, et al., *Analytical Chemistry* **2003**, *75*, 1039-1048.
K. Petritis, L.J. Kangas, et al., *Analytical Chemistry* **2006**, *78*, 5026-5039.

Assembling an AMT tag DB

- Align related datasets using elution times of observed peptides
  - One option: utilize NET prediction algorithm to create theoretical dataset to align against
    - NET prediction uses position and ordering of amino acid residues to predict normalized elution time
  - Alignment yields NET values based on observed elution times
    - Observed NET = Slope×(Observed Elution Time) + Intercept

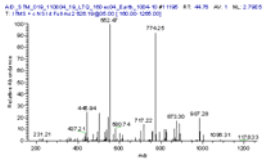
Example: 506 unique peptides used for alignment; Log(E_Value) ≤ -6

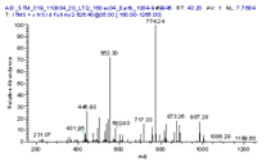


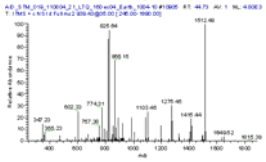
VKHPSEIVNVGDEINVK
Elution time: 44.923 minutes
Predicted NET: 0.292
Observed NET: 0.303

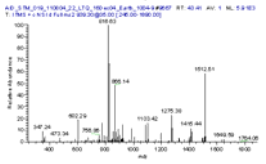
Assembling an AMT tag DB

- AMT tag example
  - R.VKHPSEIVNVGDEINVK.V
  - Observed in 7 (of 25) LC-MS/MS datasets in the SCX fractionation series

Analysis 1, scan 11195  → 3+, hyperscore 80, Obs. NET 0.303

Analysis 2, scan 9945  → 3+, hyperscore 69, Obs. NET 0.298

Analysis 3, scan 10905  → 2+, hyperscore 74, Obs. NET 0.301

Analysis 4, scan 9667  → 2+, hyperscore 77, Obs. NET 0.302

⋮ ⋮

Compute monoisotopic mass: 1876.0053 Da
Average Normalized Elution Time: 0.3021 (StDev 0.0021)

Assembling an AMT tag DB

- Mass and Time Tag Database
  - Repository for AMT tags
    - Mass, elution time, modified residues, and supporting information for each AMT tag
  - Allows samples of unknown composition to be matched quickly and efficiently, without needing to perform tandem MS
  - Assembled by analyzing a control set of samples, cataloging each peptide identification until subsequent analyses no longer provide new identifications

MT Tag ID	Peptide	LC-MS/MS Obs. Count	Calculated Monoisotopic Mass	Average Observed NET	Observed NET StDev
1662039	MTGRELKPHDR	1	1338.6826	0.143	0.000
17683899	SSALNTLTNQK	3	1175.6146	0.235	0.005
36609588	HRDLLGATNP…TLR	5	1960.0602	0.379	0.002
36715875	WVKVDGWDN…FER	11	2590.2815	0.459	0.011
36843675	MYGHLKGEVA…QER	8	2533.2304	0.557	0.005

Assembling an AMT tag DB

- Mini AMT tag DB
  - Database constructed from a relatively small number of datasets
  - e.g. 25 SCX fractionation samples from S. typhimurium, each analyzed by LC-MS/MS and then by X!Tandem
  - Protein database: S_typhimurium_LT2_2004-09-19
    - 4550 proteins and 1.4 million residues

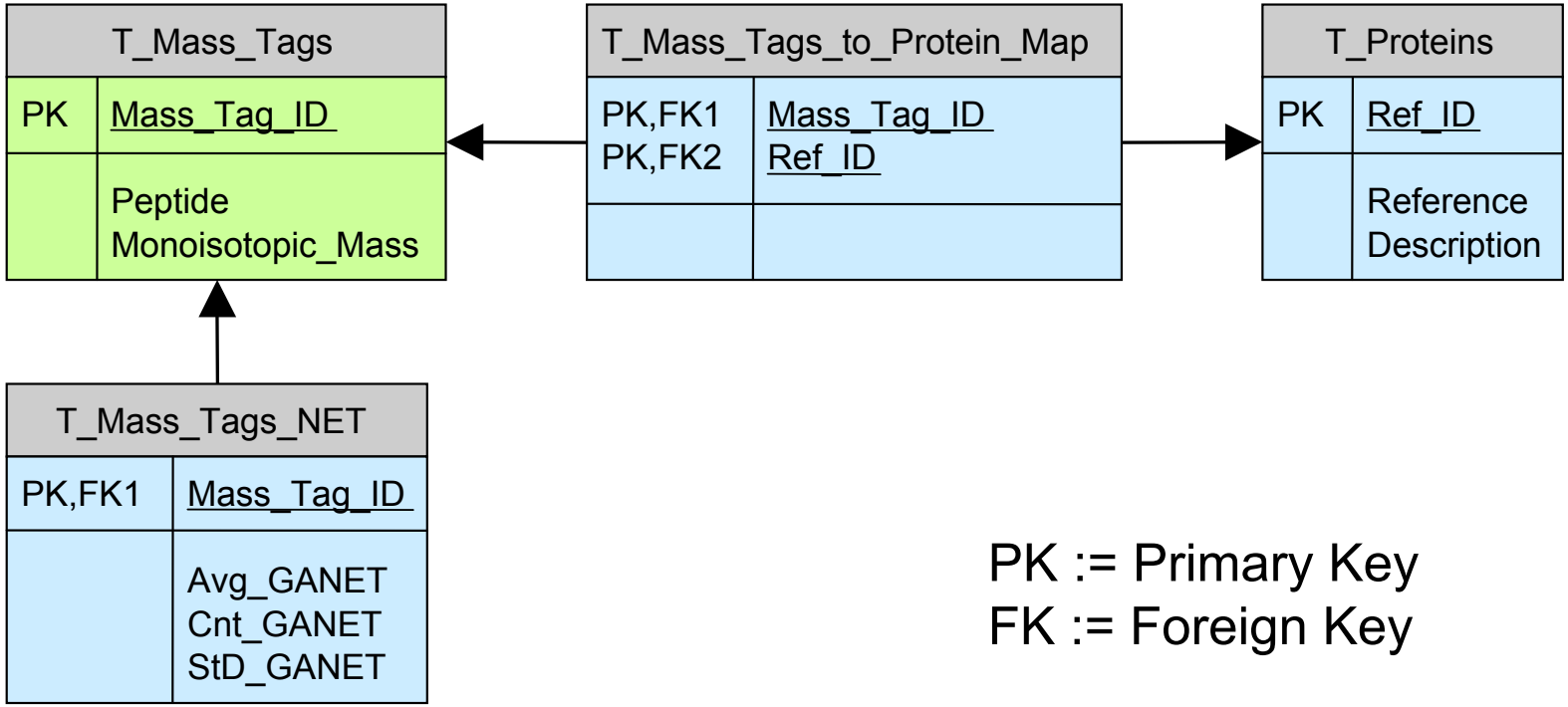
```
>STM1834 putative YebN family transport protein (yebN) {Salmonella typhimurium LT2}
MFAGGSDVFNGYPGQDVVMHFTATVLLAFGMSMDAFAASIGKGATLHKPKFSEALRTGLI
FGAVETLTPLIGWGLGILASKFVLEWNHWIAFVLLIFLGGRMIIEGIRGGSDEDETPLRR
HSFWLLVTTAIATSLDAMAVGVGLAFLQVNIIATALAIGCATLIMSTLGMMIGRFIGPML
GKRAEILGGVVLIGIGVQILWTHFHG
>STM1835 23S rRNA m1G745 methyltransferase (rrmA) {Salmonella typhimurium LT2}
MSFTCPLCHQPLTQINNSVICPQRHQFDVAKEGYINLLPVQHKRSRDPGDSAEMMQARRA
FLDAGHYQPLRDAVINLLRERLDQSATAILDIGCGEGYYTHAFAEALPGVTTFGLDVAKT
AIKAAAKRYSQVKFCVASSHRLPFADASMDAVIRIYAPCKAQELARVVKPGGWVVTATPG
PHHLMELKGLIYDEVRLHAPYTEQLDGFTLQQSTRLAYHMQLTAEAAVALLQMTPFAWRA
RPDVWEQLAASAGLSCQTDFNLHLWQRNR
```

Assembling an AMT tag DB

- Database Relationships
  - Minimum information required:
    - Single table with Mass and NET

T_Mass_Tags	
PK	Mass_Tag_ID
	Peptide Monoisotopic_Mass NET

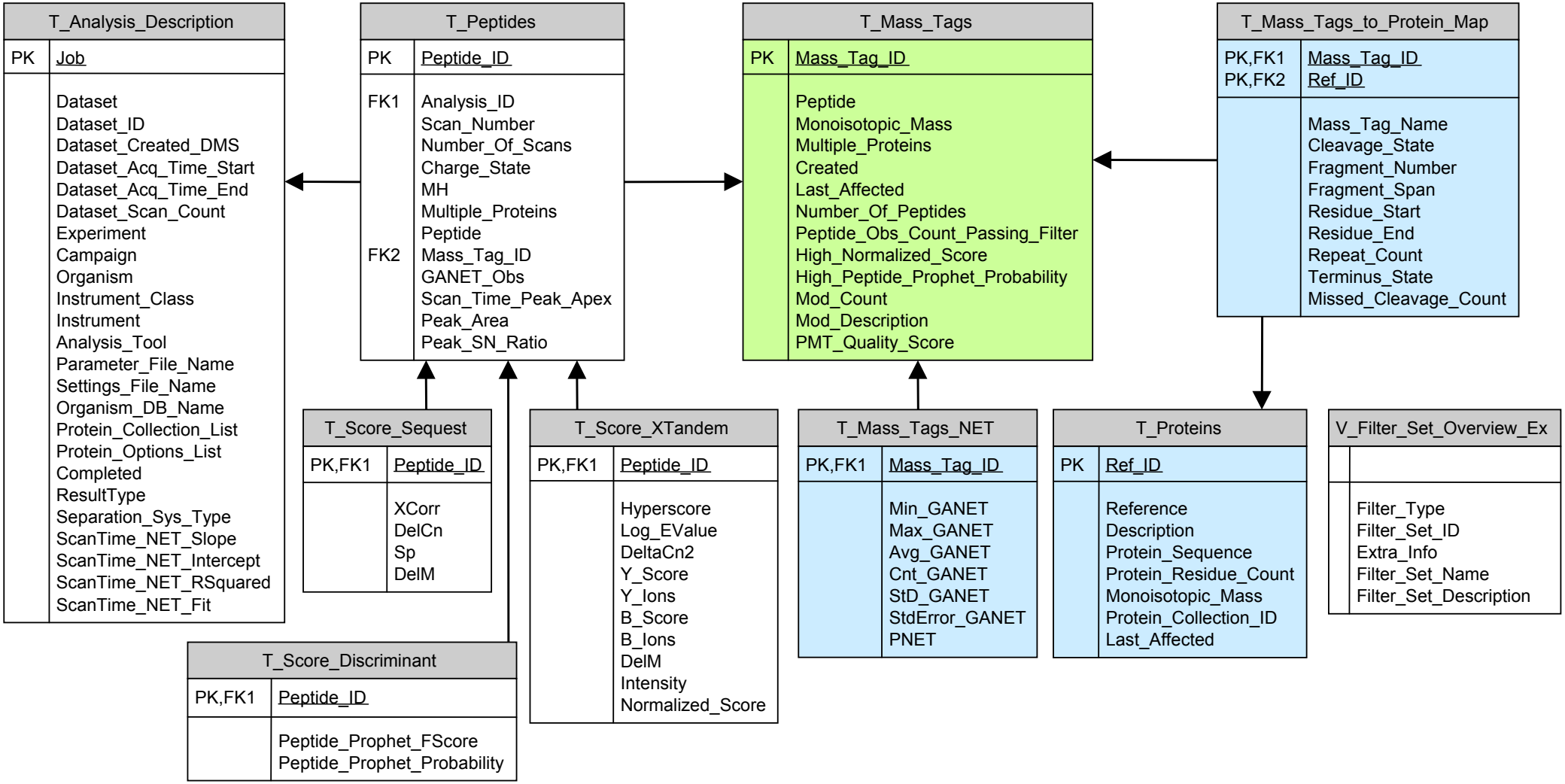
  - Expanded schema:



PK := Primary Key
FK := Foreign Key

Assembling an AMT tag DB

- Microsoft Access DB Relationships
  - Full schema to track individual peptide observations



Assembling an AMT tag DB

- Example data

T_Mass_Tags

Mass_Tag_ID	Peptide	Monoisotopic_Mass
24847	VKHPSEIVNVGDEINVK	1876.00533

T_Mass_Tags_NET

Mass_Tag_ID	Avg_GANET	Cnt_GANET	StD_GANET
24847	0.3021	7	2.11E-03

T_Peptides

Peptide_ID	Peptide	Mass Tag ID	Job	Scan Number	Charge State
53428	R.VKHPSEIVNVGDEINVK.V	24847	206386	11195	3
57461	R.VKHPSEIVNVGDEINVK.V	24847	206387	9945	3
61511	R.VKHPSEIVNVGDEINVK.V	24847	206388	10905	2
65386	R.VKHPSEIVNVGDEINVK.V	24847	206389	9667	2
69081	R.VKHPSEIVNVGDEINVK.V	24847	206390	9118	2
72556	R.VKHPSEIVNVGDEINVK.V	24847	206391	9159	2
76263	R.VKHPSEIVNVGDEINVK.V	24847	206392	9421	2

T_Score_XTandem

Peptide_ID	Hyperscore	Log(E_Value)
53428	80.2	-5.89
57461	69.2	-4.92
61511	74	-12.85
65386	77.2	-12.80
69081	69	-12.82
72556	78	-13.77
76263	60.3	-11.27

Assembling an AMT tag DB

- Processing steps

Thermo-Finnigan LTQ .Raw files

↓

MS/MS spectra files

↓

X!Tandem Results

↓

Tab delimited text files

↓

Summarized result files

↓

Microsoft Access DB

Convert to .Dta using Extract_MSn.exe. Concatenate .Dta files into _Dta.txt file using Perl script. Improved application (under development): Decon_MSn.exe

Process _Dta.txt files with X!Tandem
(round 1 partially tryptic; round 2 dynamic oxidized methionine)

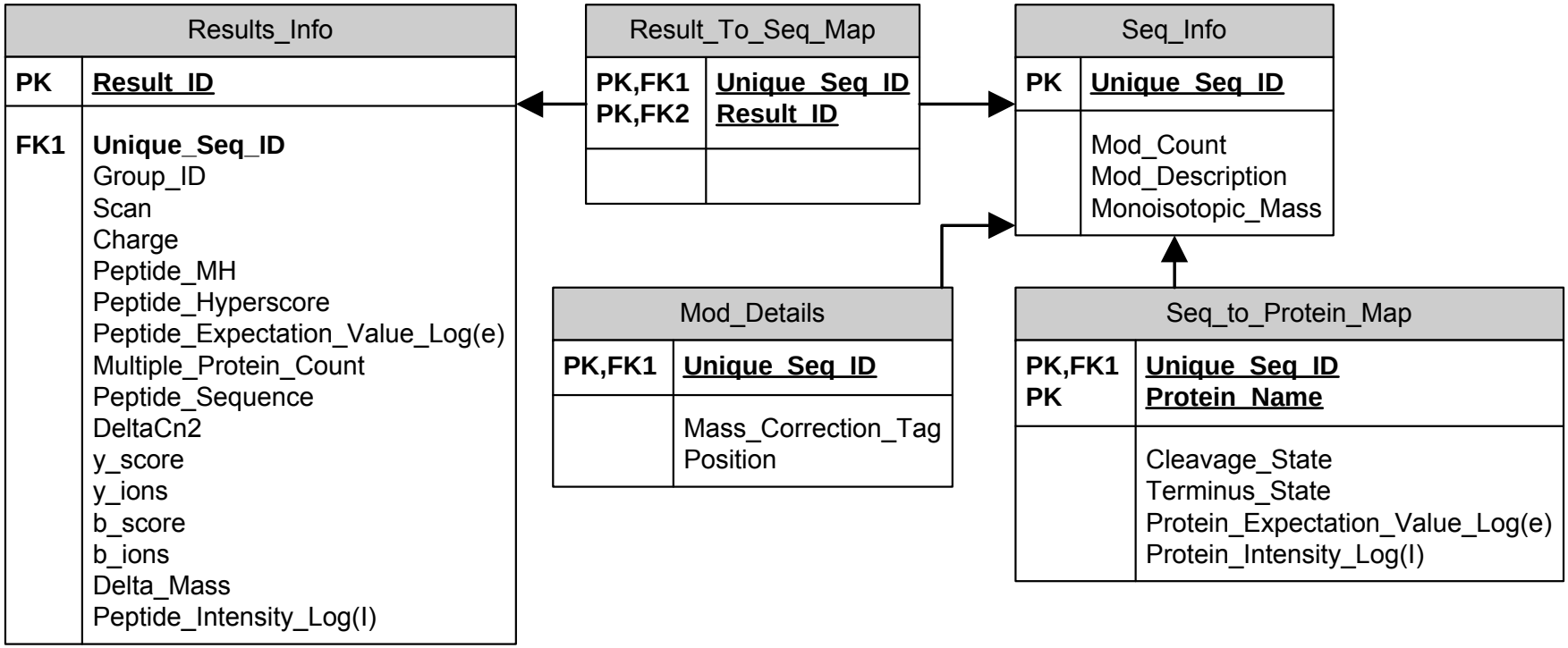
Convert X!Tandem .XML files to tab-delimited files using the Peptide Hit Results Processor application

Align datasets using MTDB Creator application

Load into database using MTDB Creator

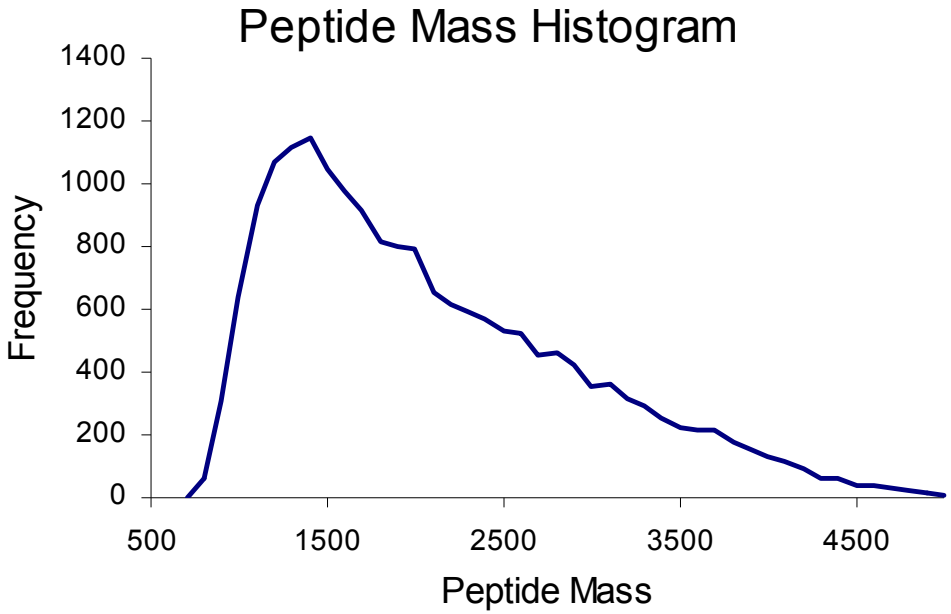
Assembling an AMT tag DB

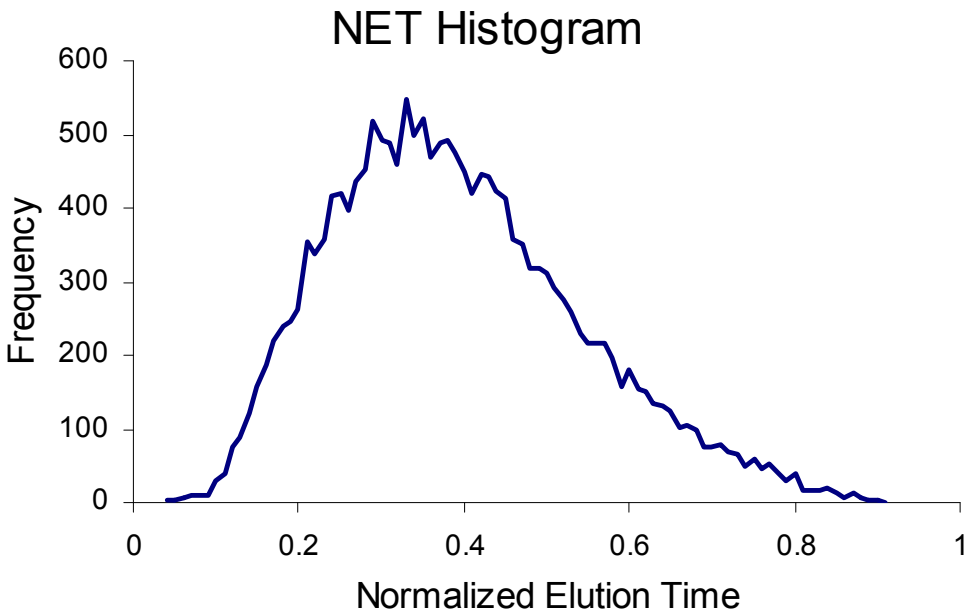
- PHRP Relationships

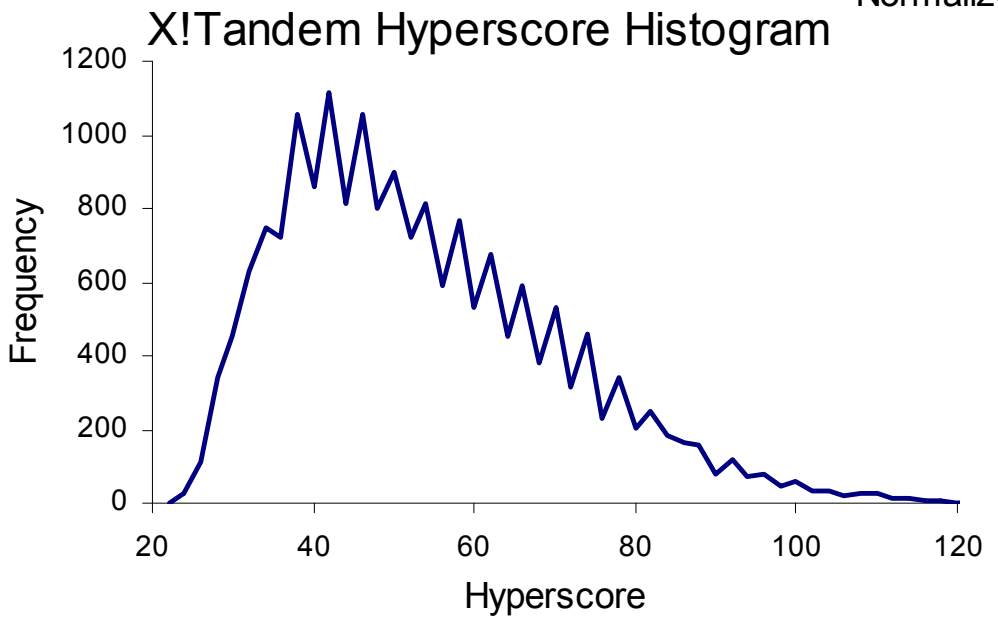


Assembling an AMT tag DB

- Database histograms – filtered on Log(E_Value) ≤ -2

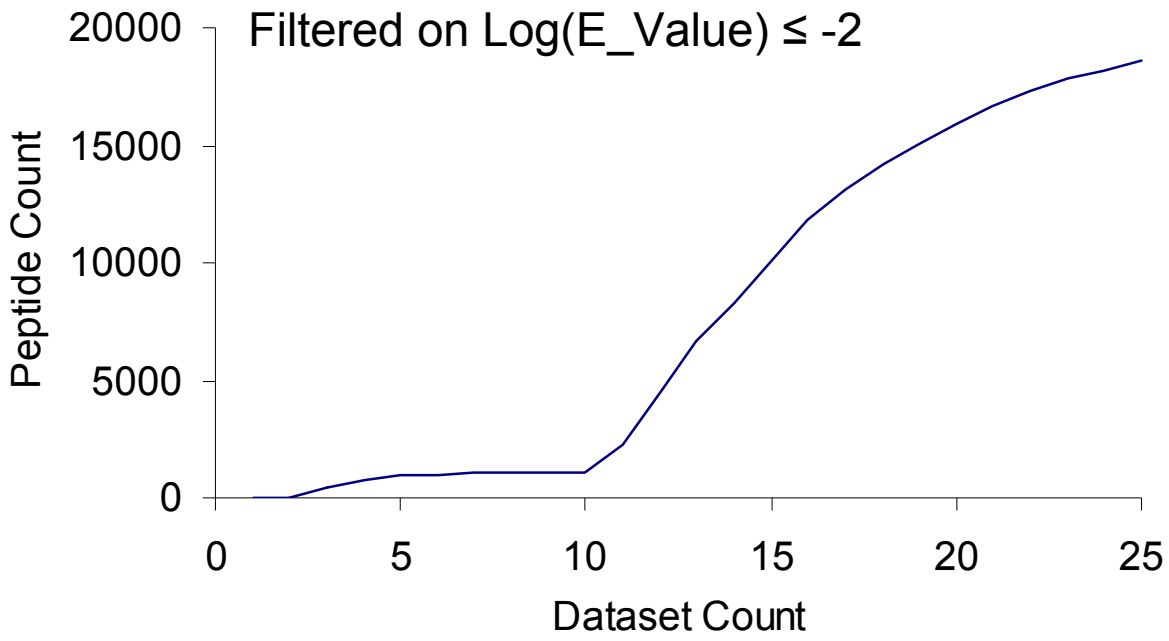




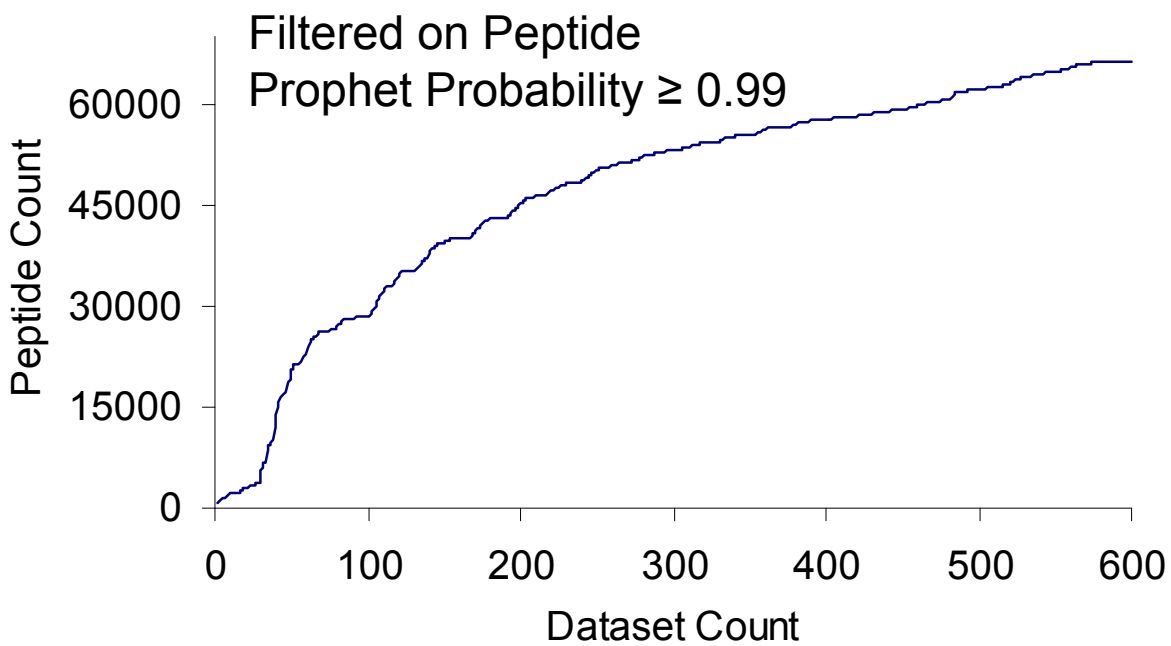


AMT Tag DB Growth Trend

- Trend for Mini AMT Tag DB
  - 25 SCX fractionation datasets of a single growth condition

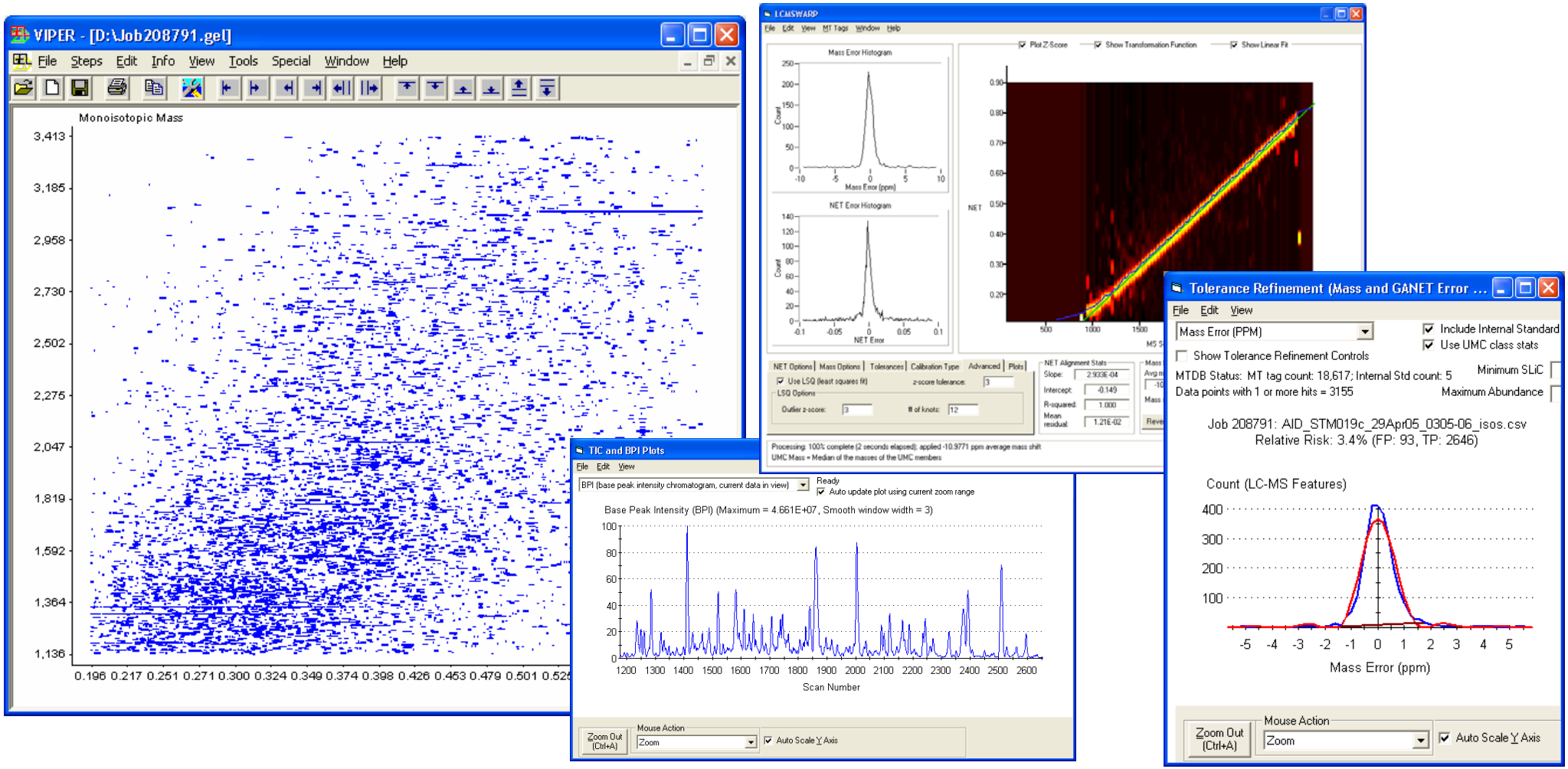


- Trend for Mature AMT Tag DB
  - 521 different samples from ~25 different conditions
  - Slope of curve decreases as more datasets added and fewer new peptides are seen



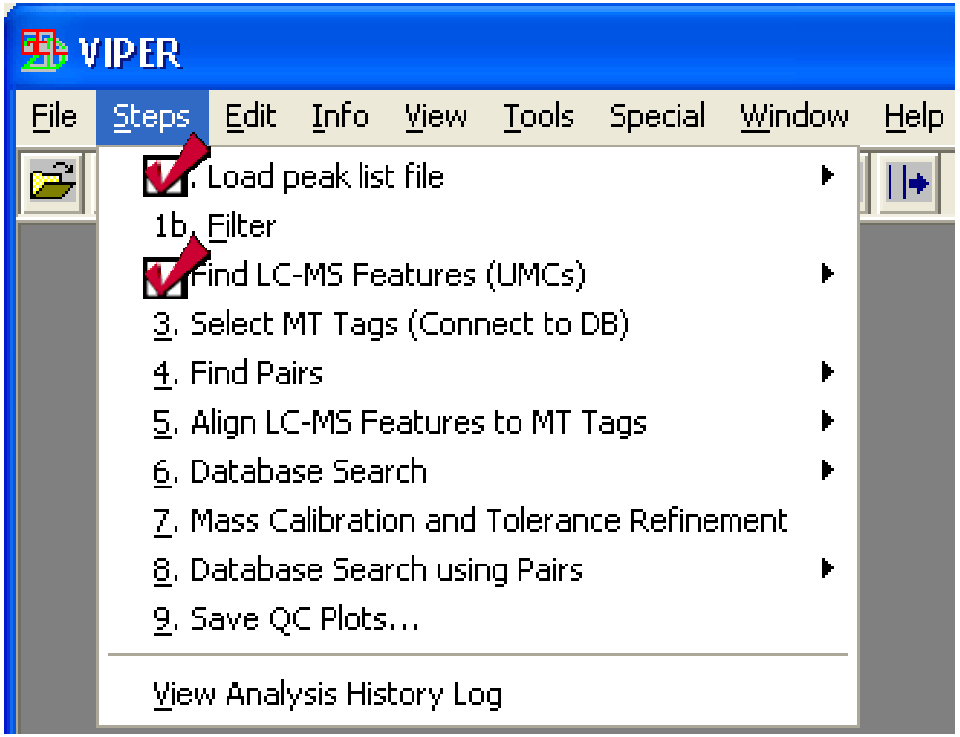
Identifying LC-MS Features

- VIPER software
  - Visualize and find features in LC-MS data
  - Match features to peptides (AMT tags)
  - Graphical User Interface and automated analysis mode



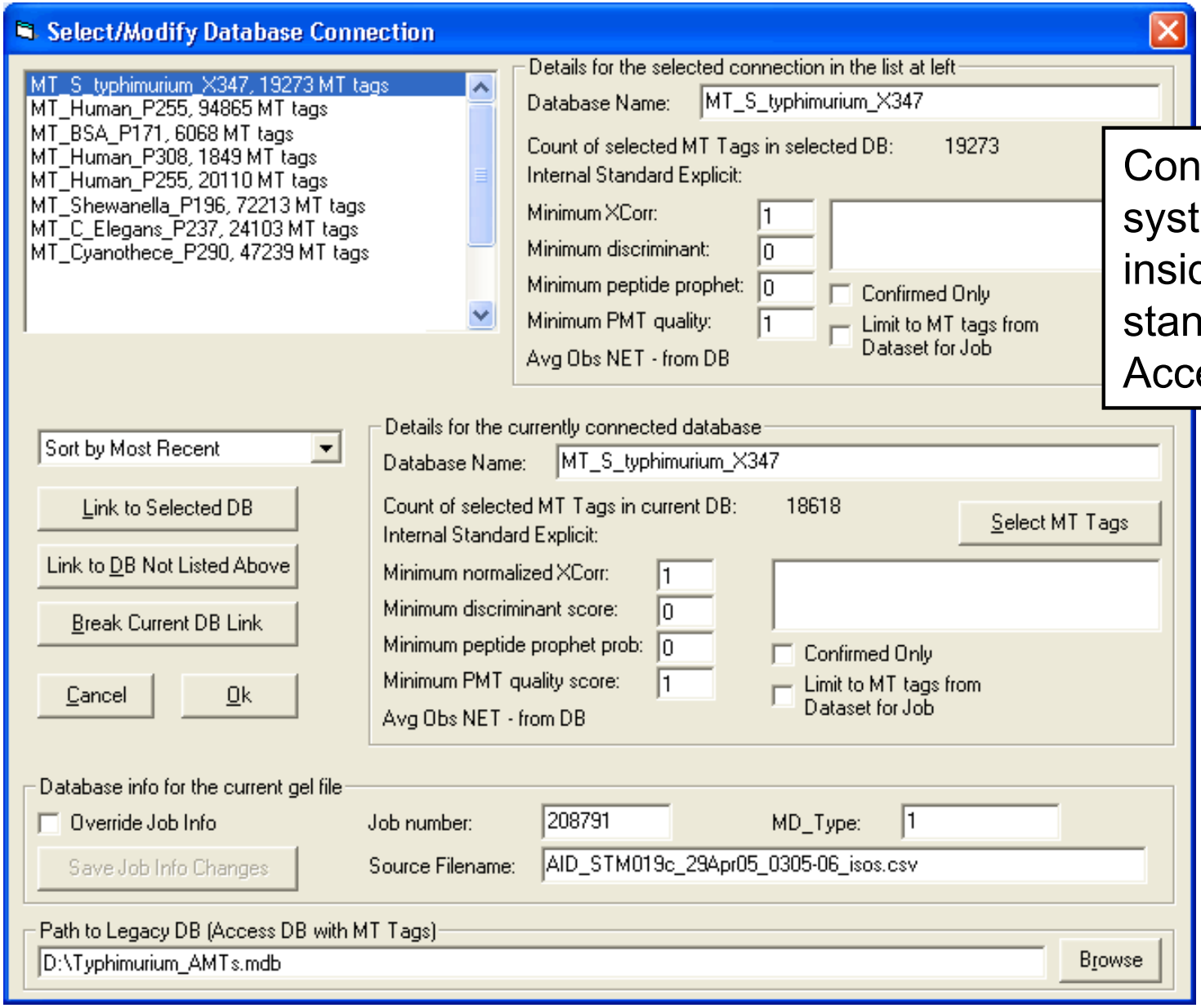
Identifying LC-MS Features

- Peak Matching Steps
  - ☑ Load LC-MS peak lists from Decon2LS
  - ☑ Filter data
  - ☑ Feature definition over elution time
  - Select AMT tags to match against
  - Optionally, find paired features (e.g. $^{16}O/^{18}O$ pairs)
  - Align LC-MS features to AMT tags using LCMSWarp
  - Broad AMT tag DB search
  - Search tolerance refinement
  - Final AMT tag DB search
  - Report results



Identifying LC-MS Features

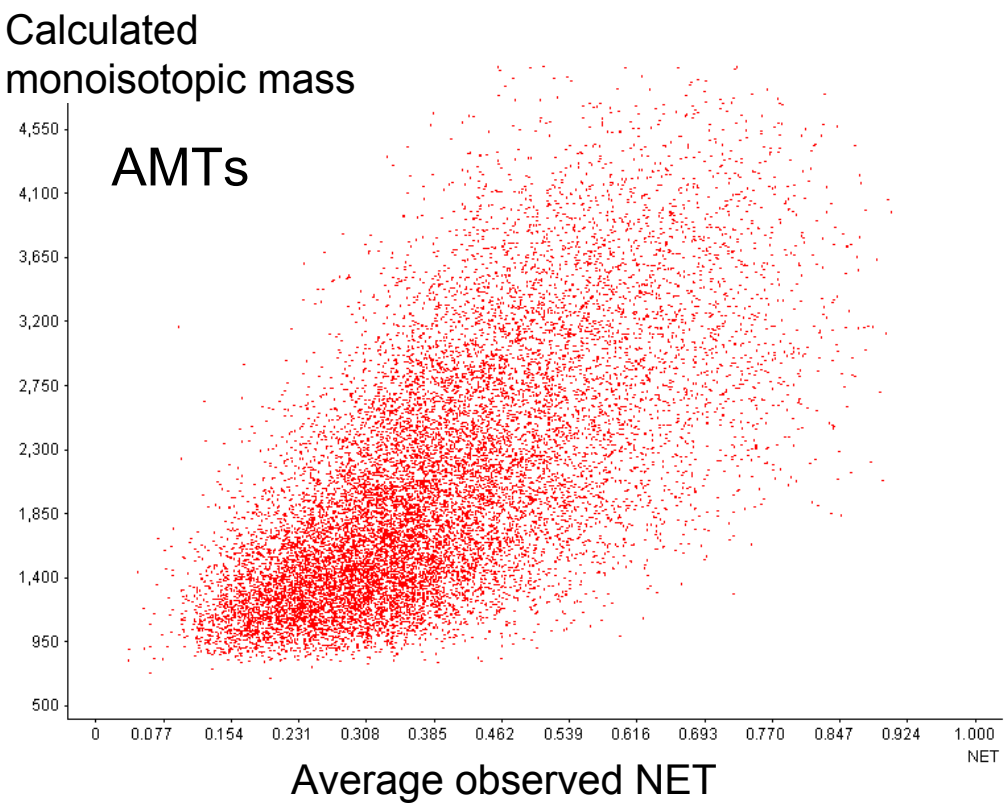
- AMT Tag database selection

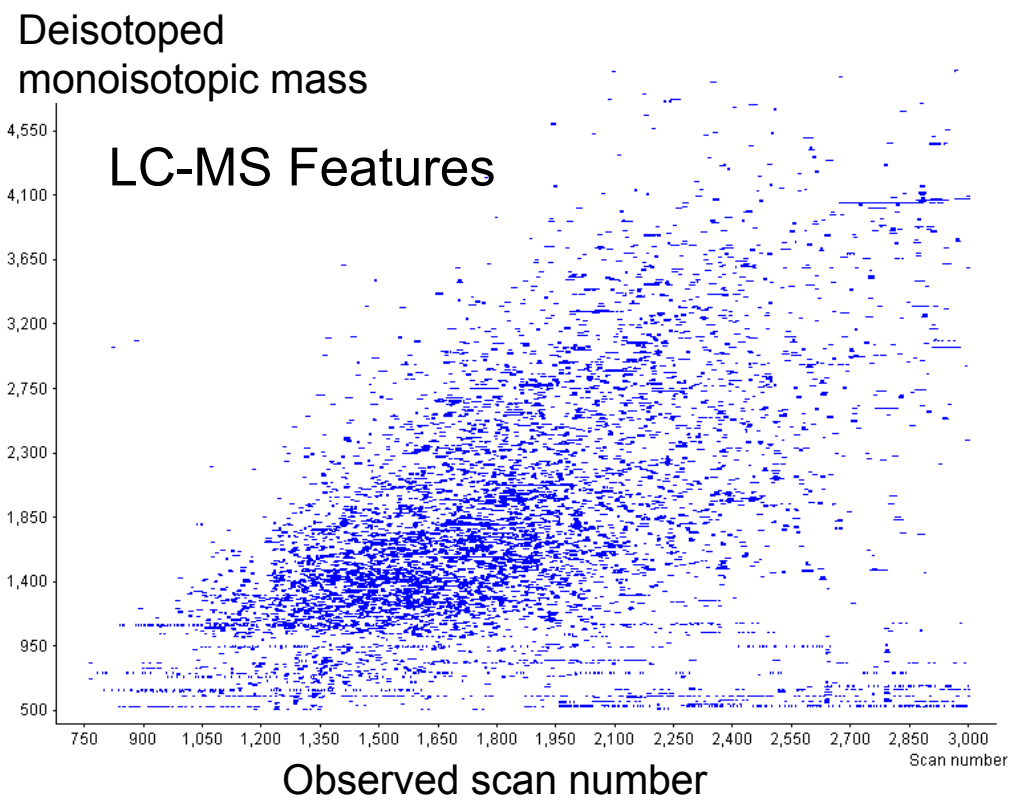


Connect to mass tag system (MTS) if inside PNNL or use standalone Microsoft Access DB

Alignment using LCMSWarp

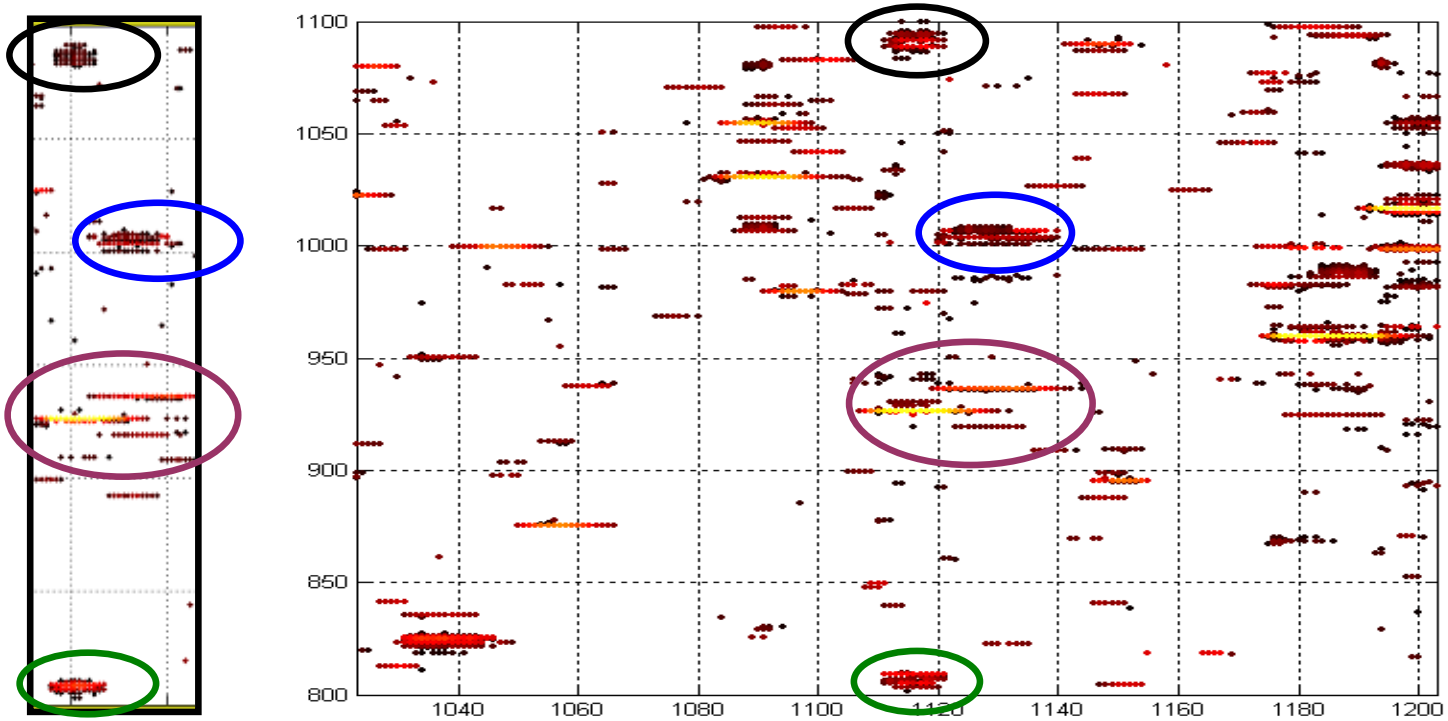
- Align scan number (i.e. elution time) of features to NETs of peptides in given AMT tag database
  - Match mass and NET of AMT tags to mass and scan number of MS features
  - Use LCMSWarp algorithm to find optimal alignment to give the most matches

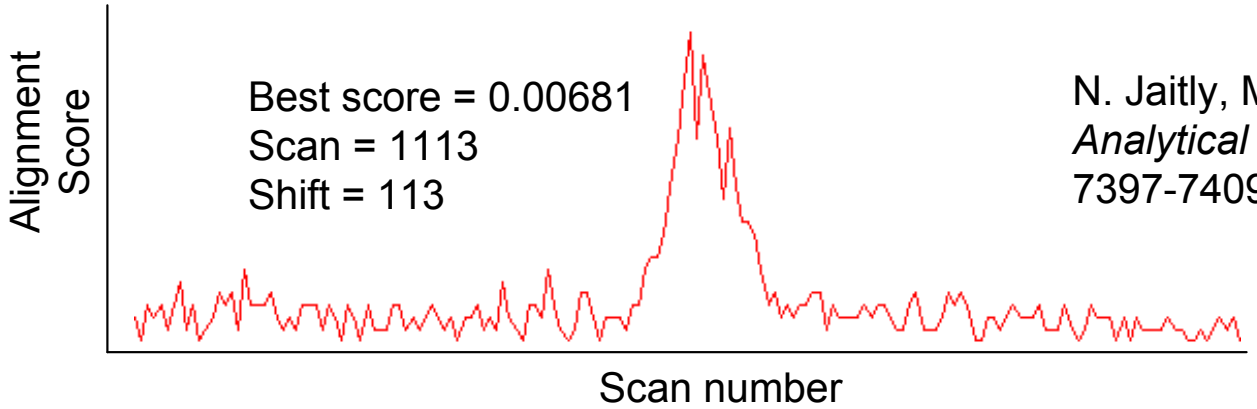




Alignment using LCMSWarp

- LCMSWarp computes a similarity score from conserved local mass and retention time patterns

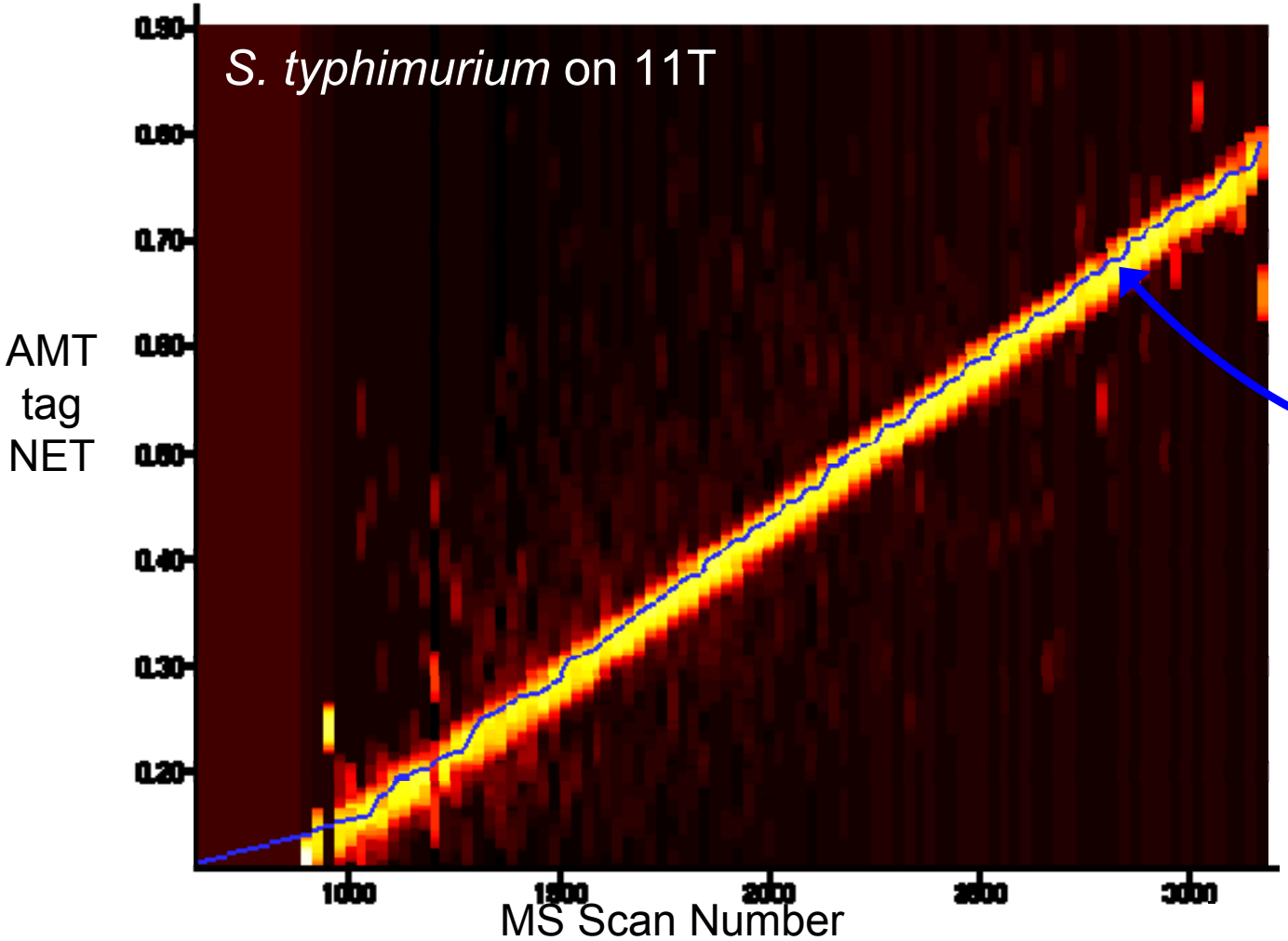




N. Jaitly, M.E. Monroe et. al., *Analytical Chemistry* **2006**, *78*, 7397-7409.

Alignment using LCMSWarp

- Similarity scores between LC-MS features and AMT tags are used to generate a score graph of similarity
- Best alignment is found using a dynamic programming algorithm that determines the transformation function with maximum likelihood



Heatmap of similarity score between LC-MS features and AMT tags (z-score representation)

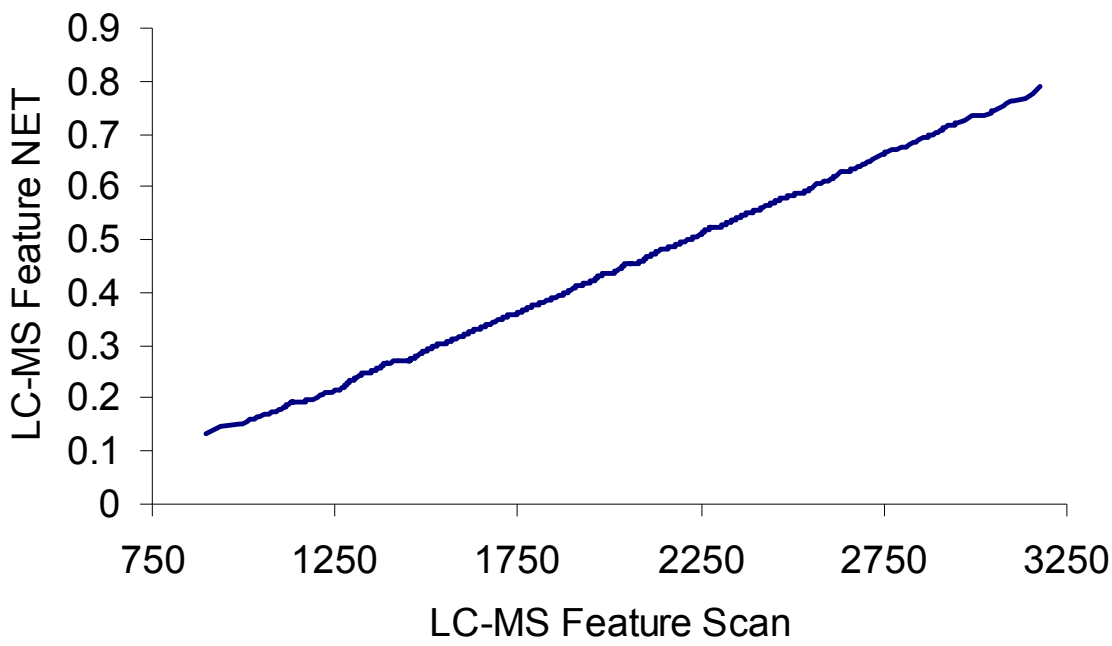
Alignment Function

N. Jaitly, M.E. Monroe et. al., *Analytical Chemistry* **2006**, *78*, 7397-7409.

Alignment using LCMSWarp

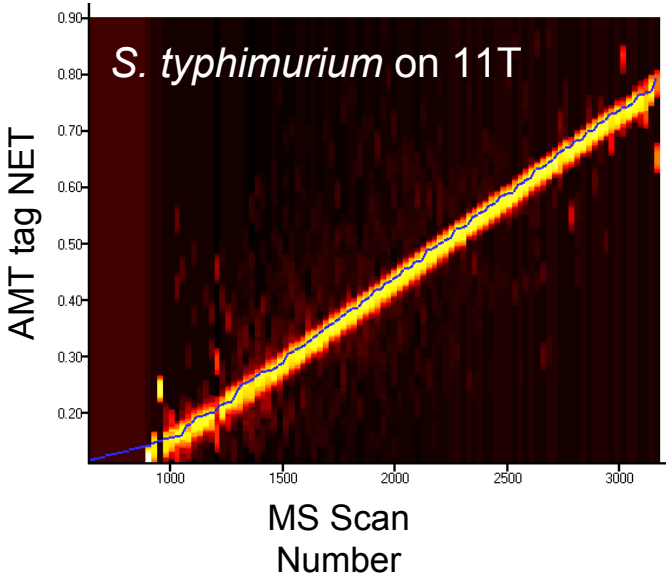
- Transformation function is used to convert from scan number to NET
  - Features centered at same scan number get the same obs. NET value
  - When matching LC-MS features to AMTs, we will search +/- a NET tolerance, which effectively allows for LC-MS features to shift around a little in elution time

LC-MS Feature Scan	Matching AMT tag NET	LC-MS Feature NET
1011	0.1519	0.1569
1019	0.1626	0.1589
1019	0.1507	0.1589
1021	0.1653	0.1594
1027	0.1509	0.1609
1037	0.1519	0.1633
1042	0.183	0.1645
1055	0.1652	0.1677
1056	0.1862	0.1679
1056	0.1697	0.1679
1056	0.1682	0.1679

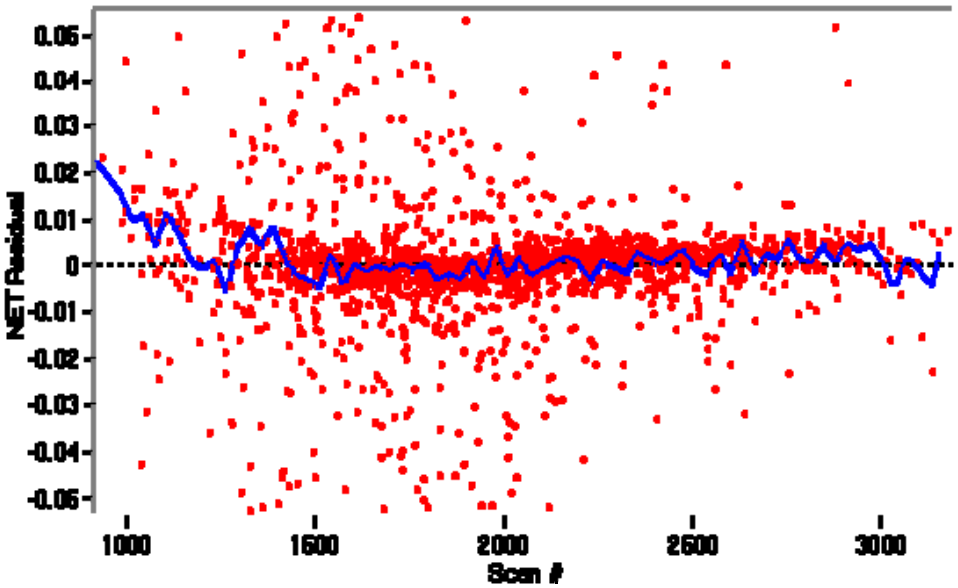


Alignment using LCMSWarp

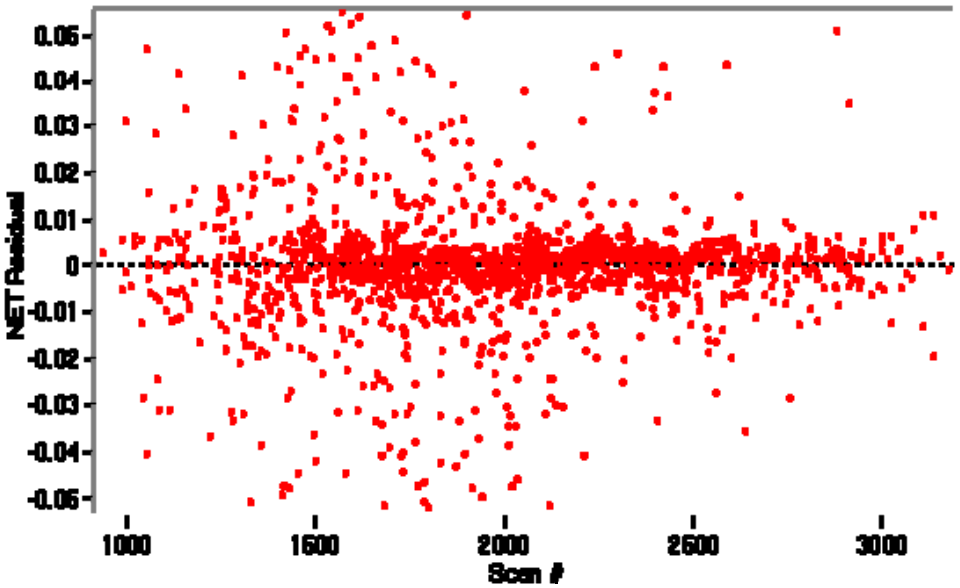
- NET Residual Plots
  - Difference between NET of LC-MS feature and NET of matching AMT tag
    - Indicates quality of alignment between features and AMT tags
  - This data shows nearly linear alignment between features and AMTs, but the algorithm can easily account for non-linear trends



NET Residuals if a linear mapping is used

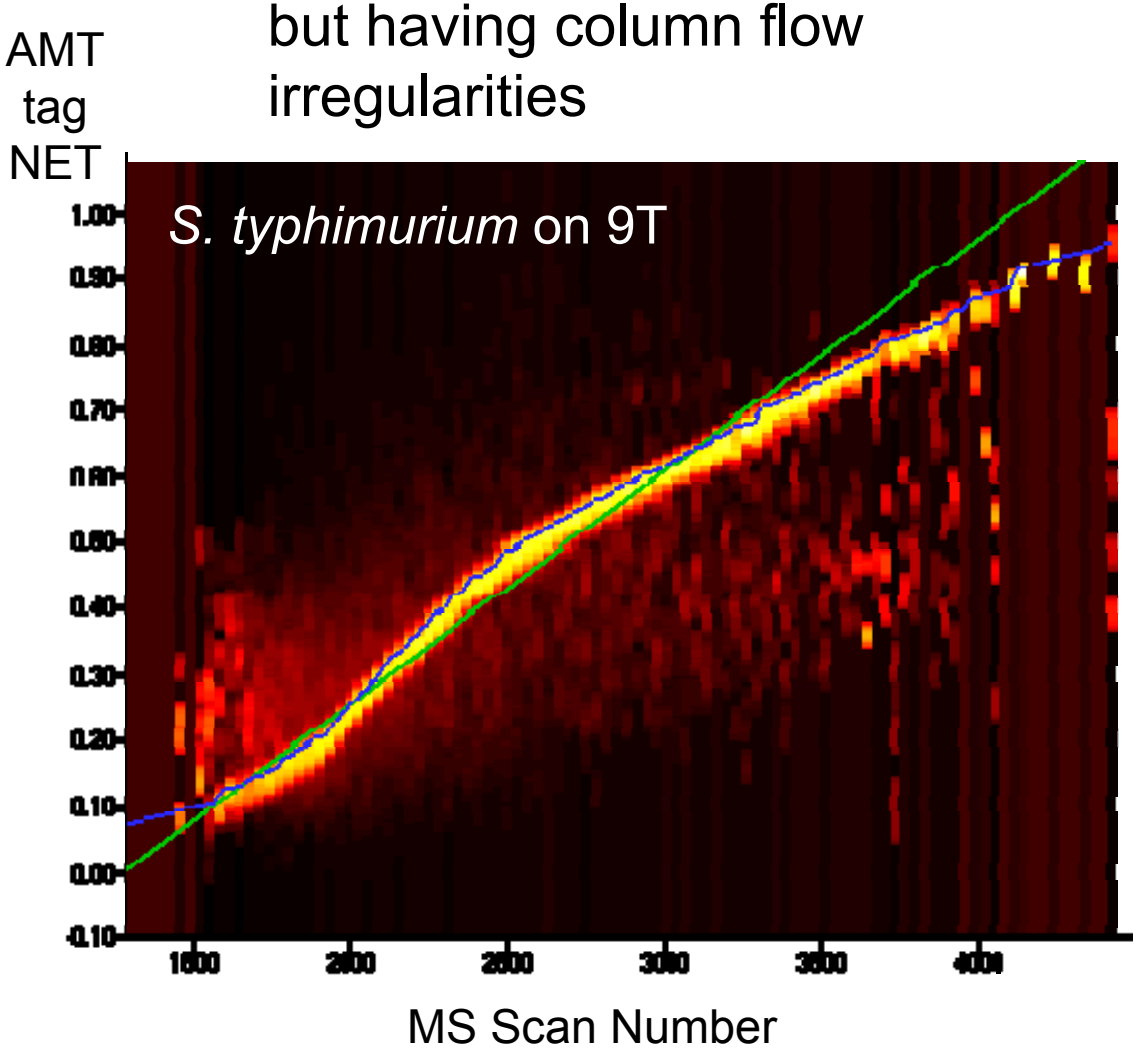


NET Residuals after LCMSWarp

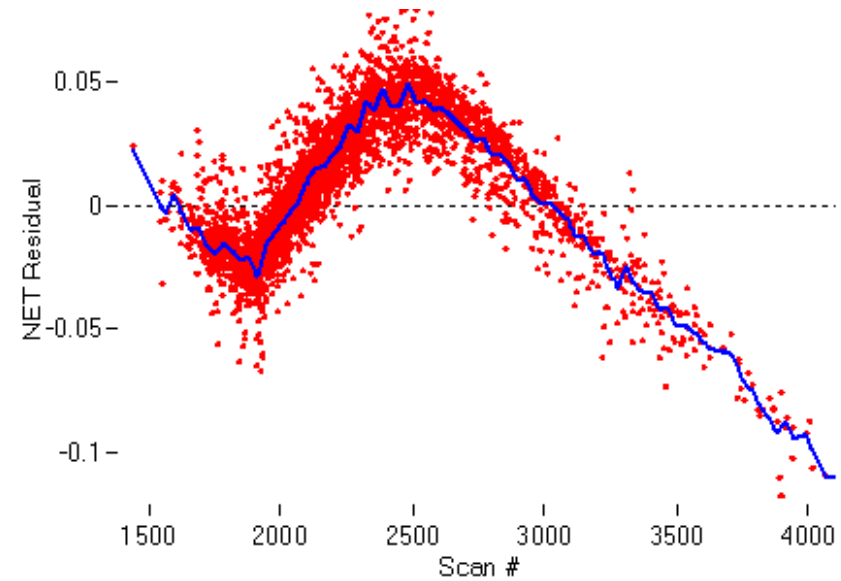


Alignment using LCMSWarp

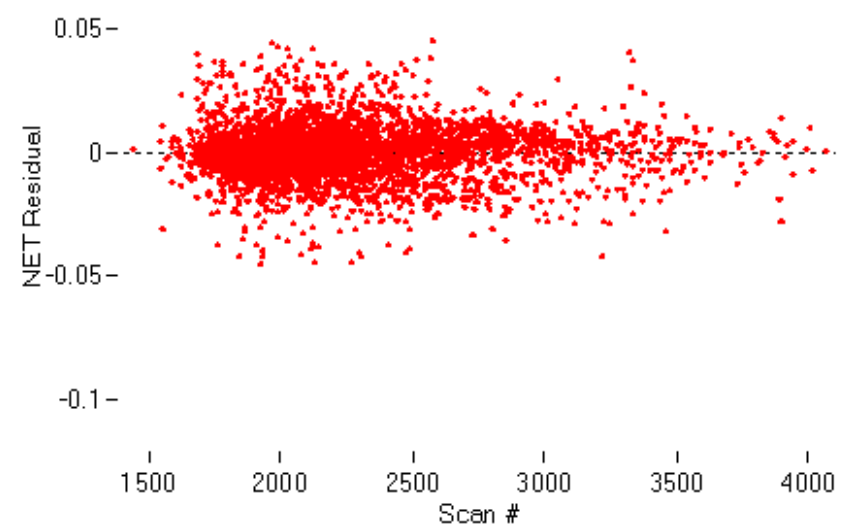
- Non-linear alignment example #1
  - Identical LC separation system, but having column flow irregularities



NET Residuals if a linear mapping is used

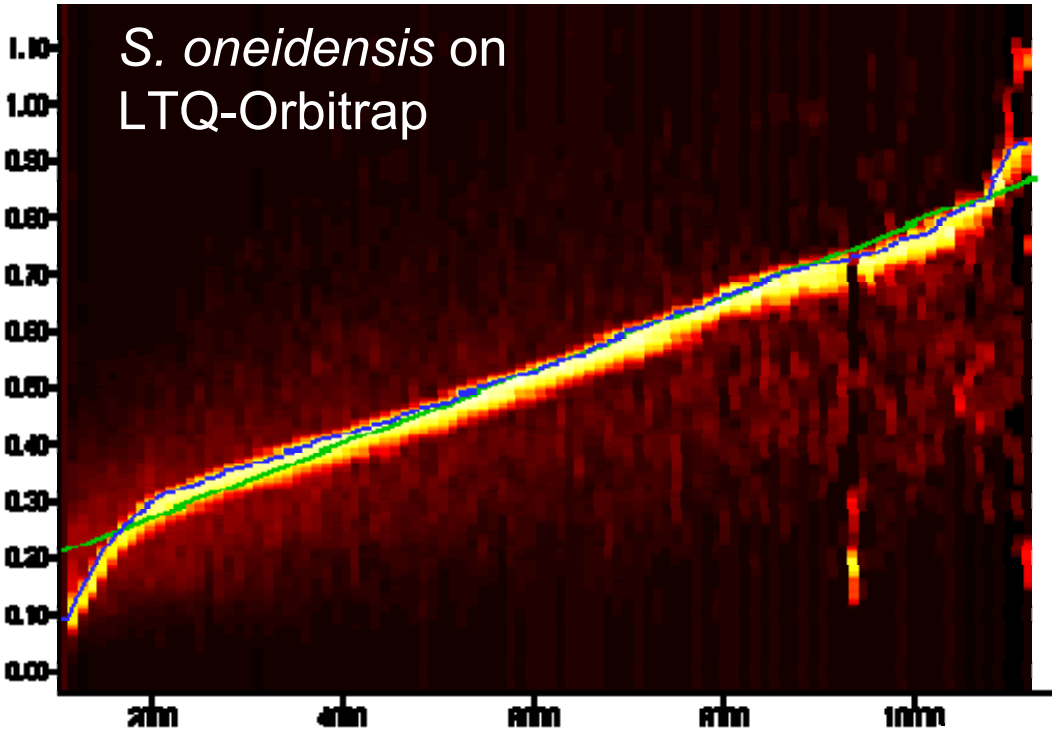


NET Residuals after LCMSWarp

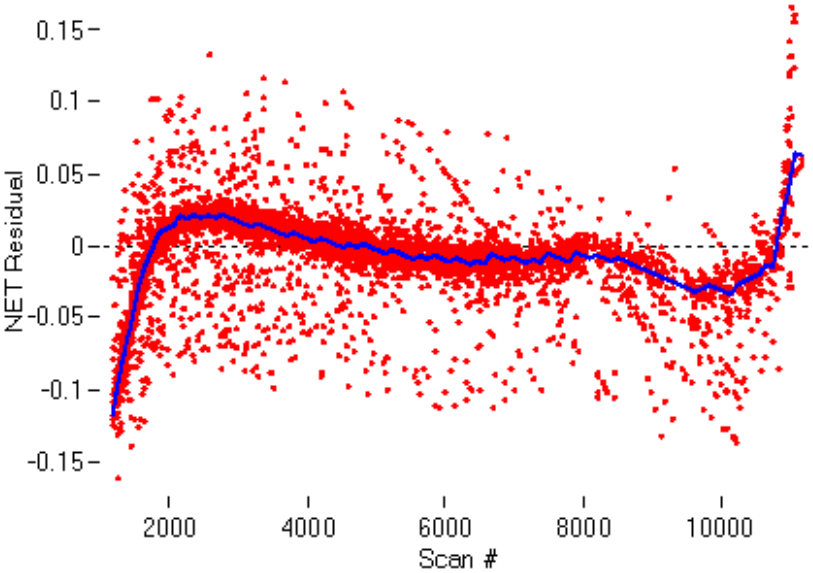


Alignment using LCMSWarp

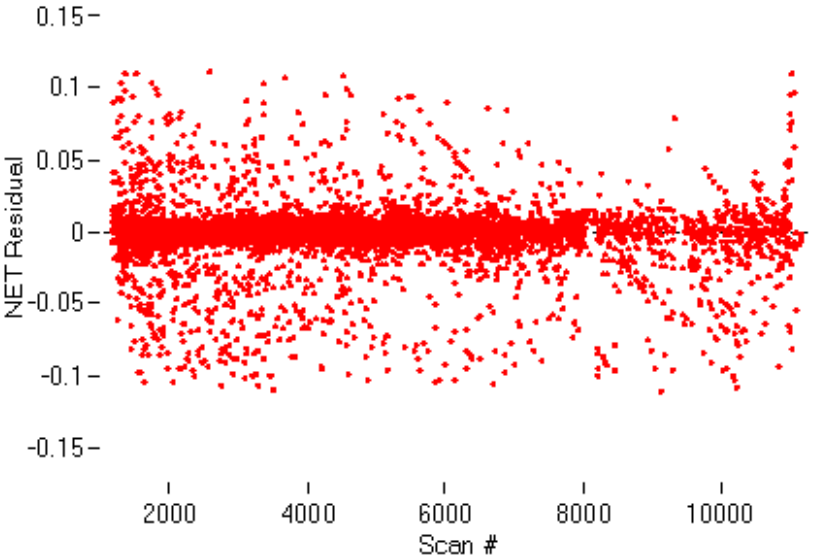
- Non-linear alignment example #2
  - AMT Tag DB from C_{18} LC-MS/MS analyses using ISCO-based LC (**exponential dilution gradient**)
  - LC-MS analysis used C_{18} LC-MS via Agilent **linear gradient** pump



NET Residuals if a linear mapping is used

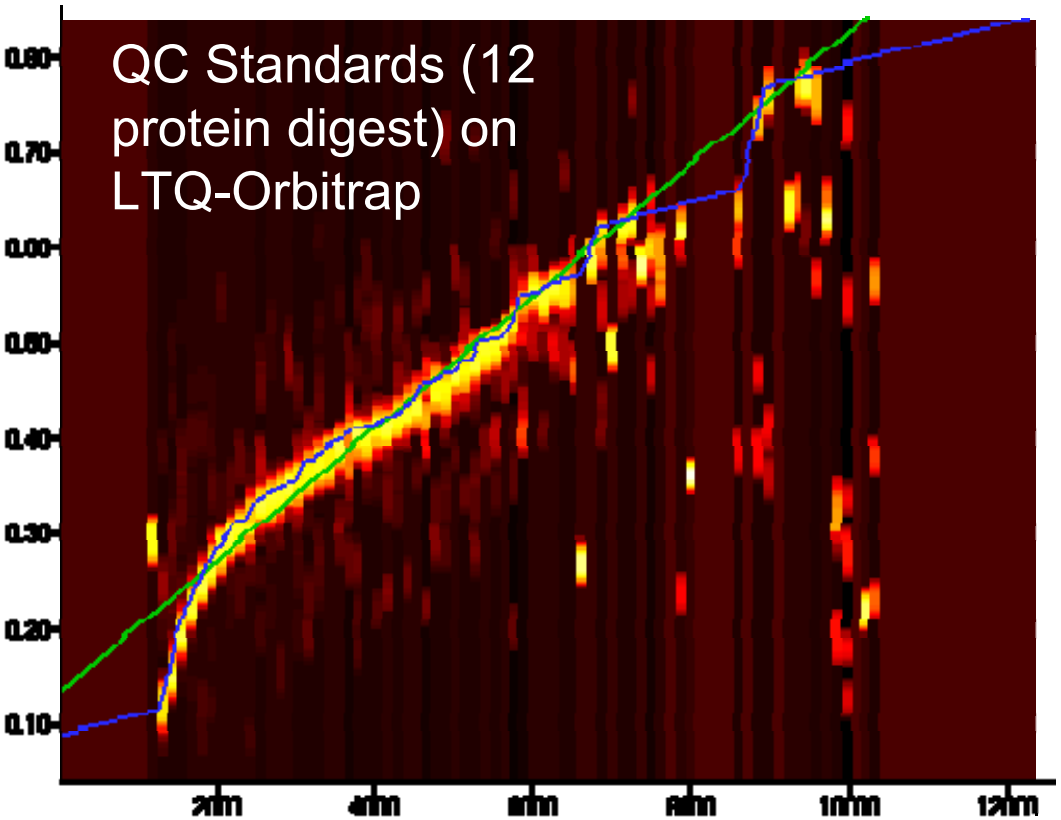


NET Residuals after LCMSWarp

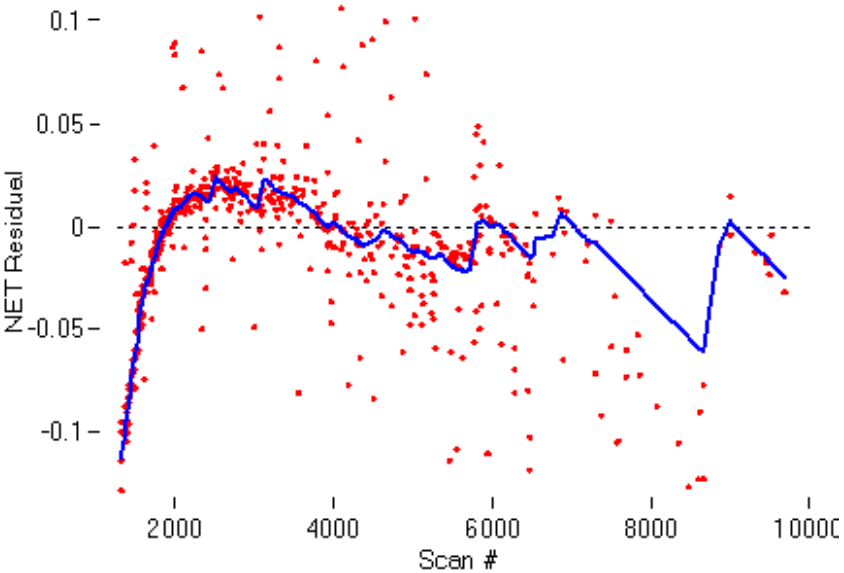


Alignment using LCMSWarp

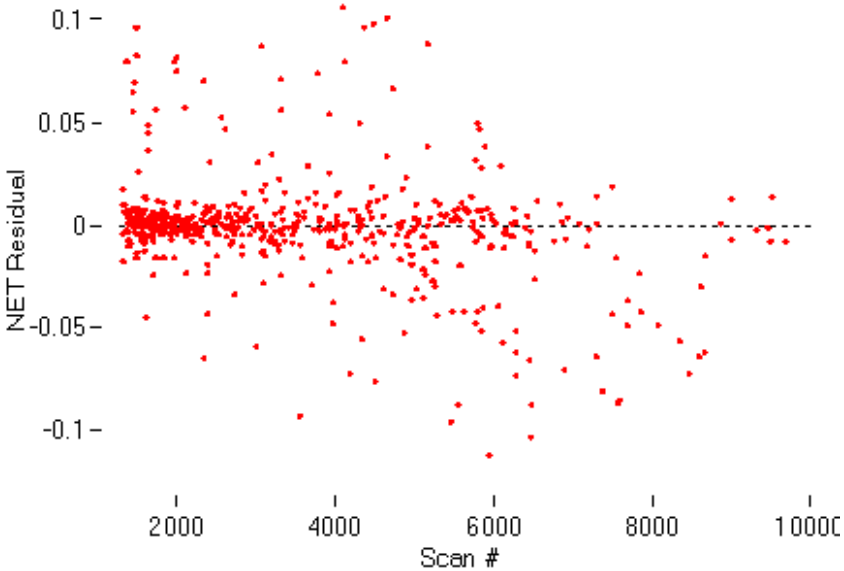
- Non-linear alignment example #3
  - AMT Tag DB from C_{18} LC-MS/MS analyses using ISCO-based LC
  - LC-MS analysis used C_{18} LC-MS via Agilent linear gradient pump



NET Residuals if a linear mapping is used

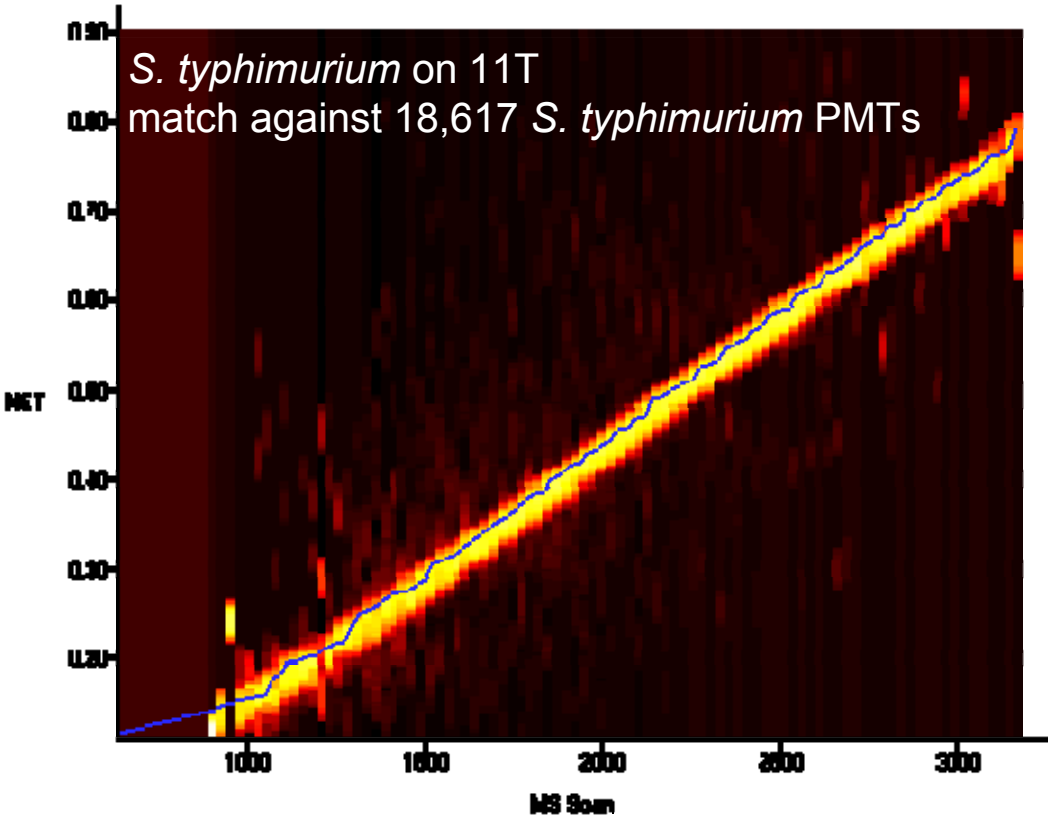


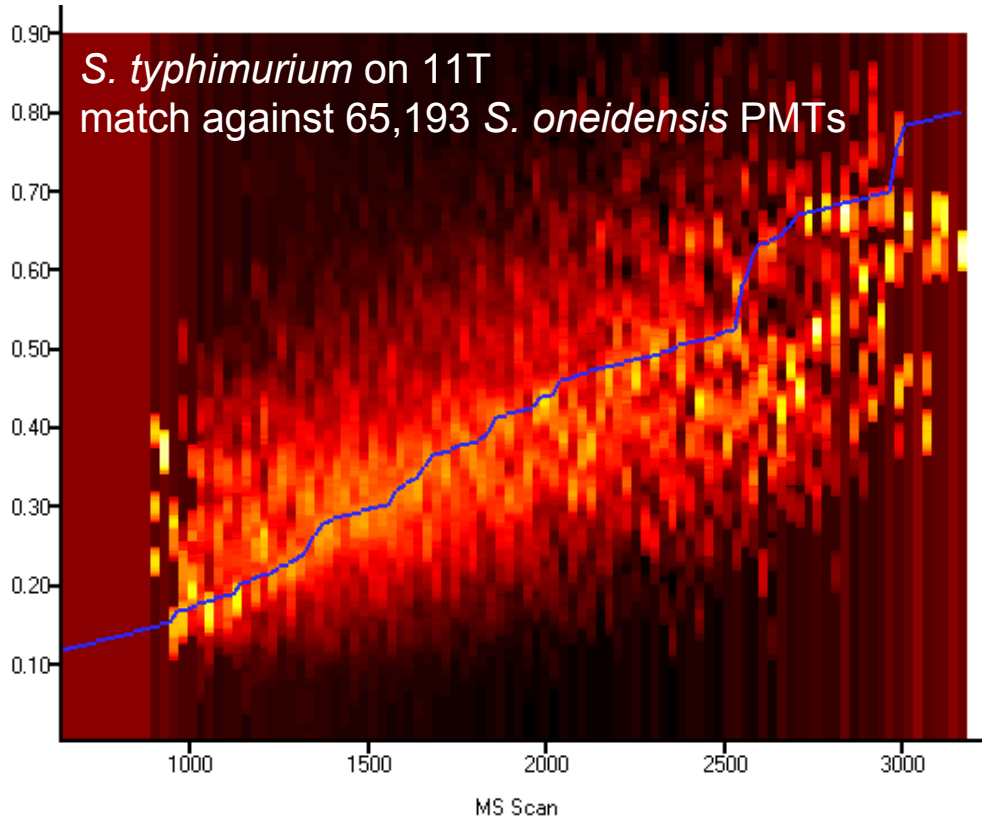
NET Residuals after LCMSWarp



Alignment using LCMSWarp

- LCMSWarp Features
  - Fast and robust
    - Previous method used least-squares regression, iterating through a large range of guesses (slow and often gave poor alignment)
  - Requires that a reasonable number of LC-MS features match the AMT Tag DB





Alignment using LCMSWarp

- In addition to aligning data in time, we can also recalibrate the masses of the LC-MS features
  - Possible because mass and time values are available for both LC-MS features and AMT tags
- Two options for mass re-calibration
  - Bulk linear correction
  - Piece-wise correction via LCMSWarp
- Visualize mass differences using mass error histogram or mass residual plot

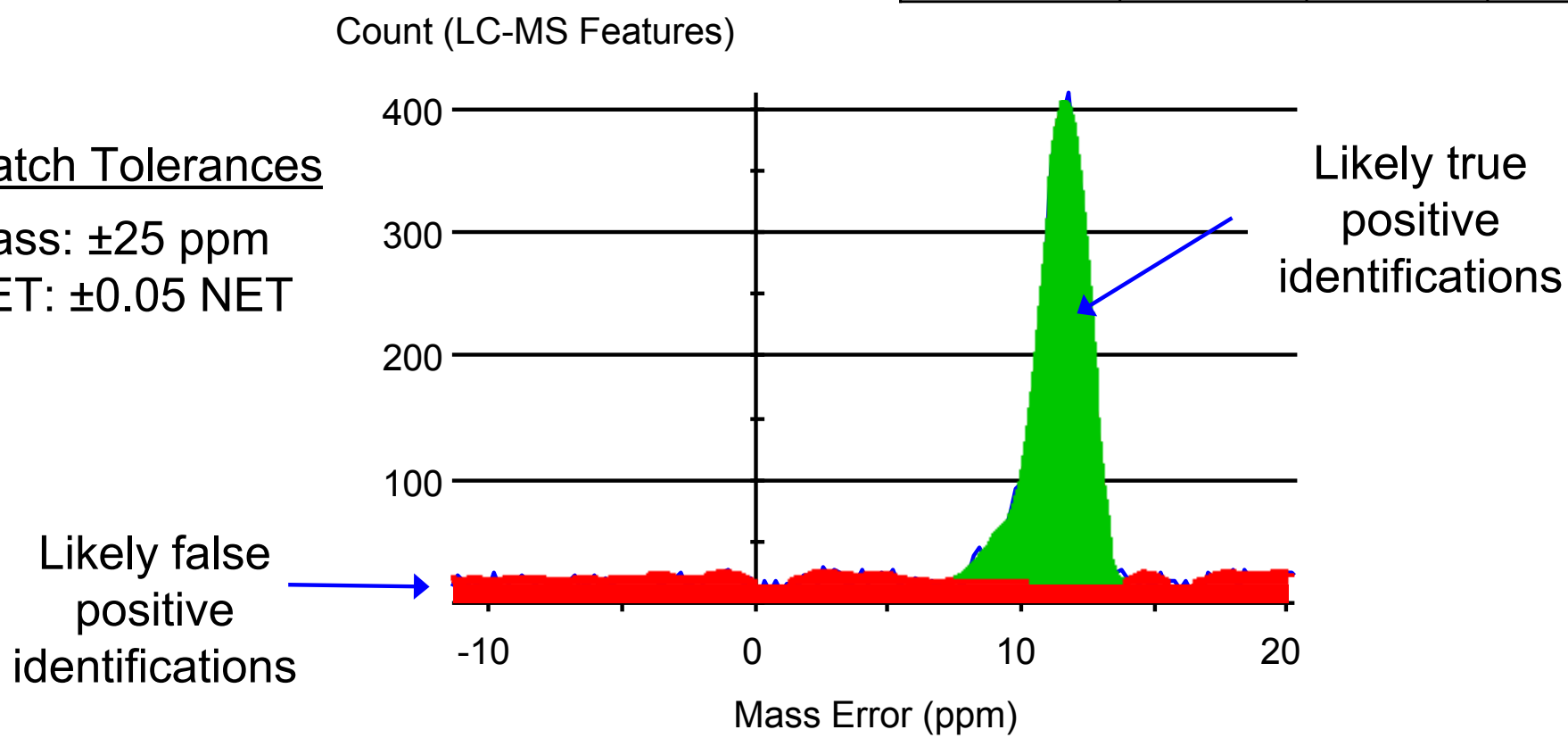
Mass Error Histogram

- List of binned mass error values
  - Difference between feature's mass and matching AMT tag's mass
  - Bin values to generate a histogram
  - Typically observe background false positive level

LC-MS Feature Mass (Da)	AMT Tag Mass (Da)	Delta Mass (Da)	Mass Error (ppm)
1570.9005	1570.883	0.01745	11.1
1571.74325	1571.726	0.01770	11.3
1571.8498	1571.831	0.01912	12.2
1571.9107	1571.892	0.01848	11.8
1573.8381	1573.832	0.00569	3.6

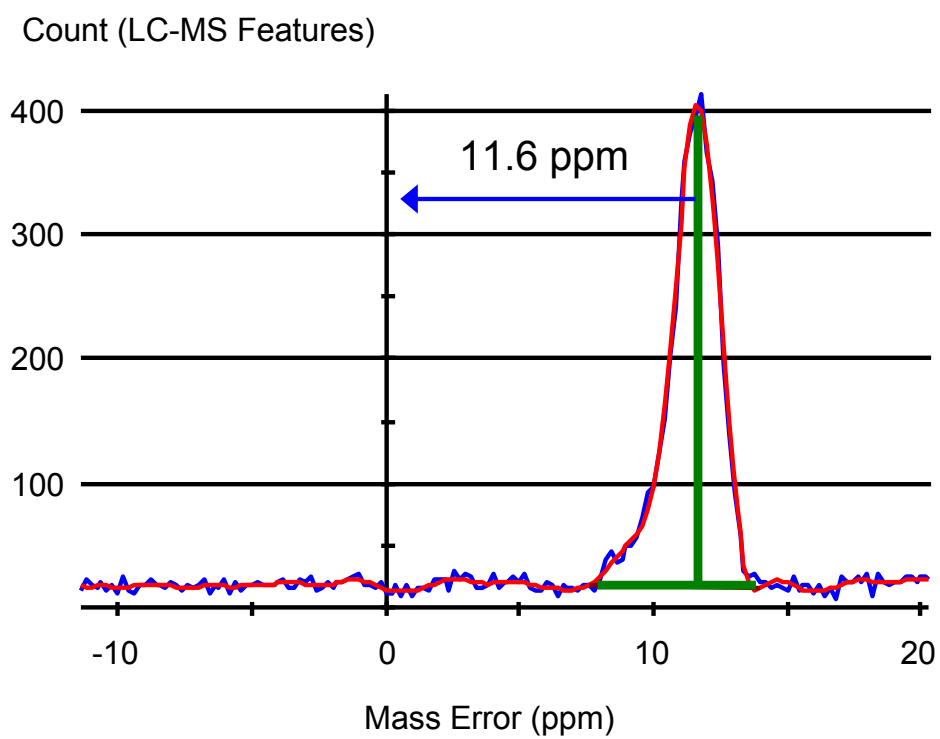
<u>Match Tolerances</u>

Mass: ±25 ppm
NET: ±0.05 NET



Mass Calibration

- Option 1: Bulk linear correction
  - Use location of peak in mass error histogram to adjust masses of all features
  - Shift by ppm mass; absolute shift amount increases as monoisotopic mass increases



Peak Center of mass: 11.6 ppm
Peak Width: 2 ppm at 60% of max
Peak Height: 404 counts/bin
Noise level: 19 counts/bin

Shift all masses -11.6 ppm:

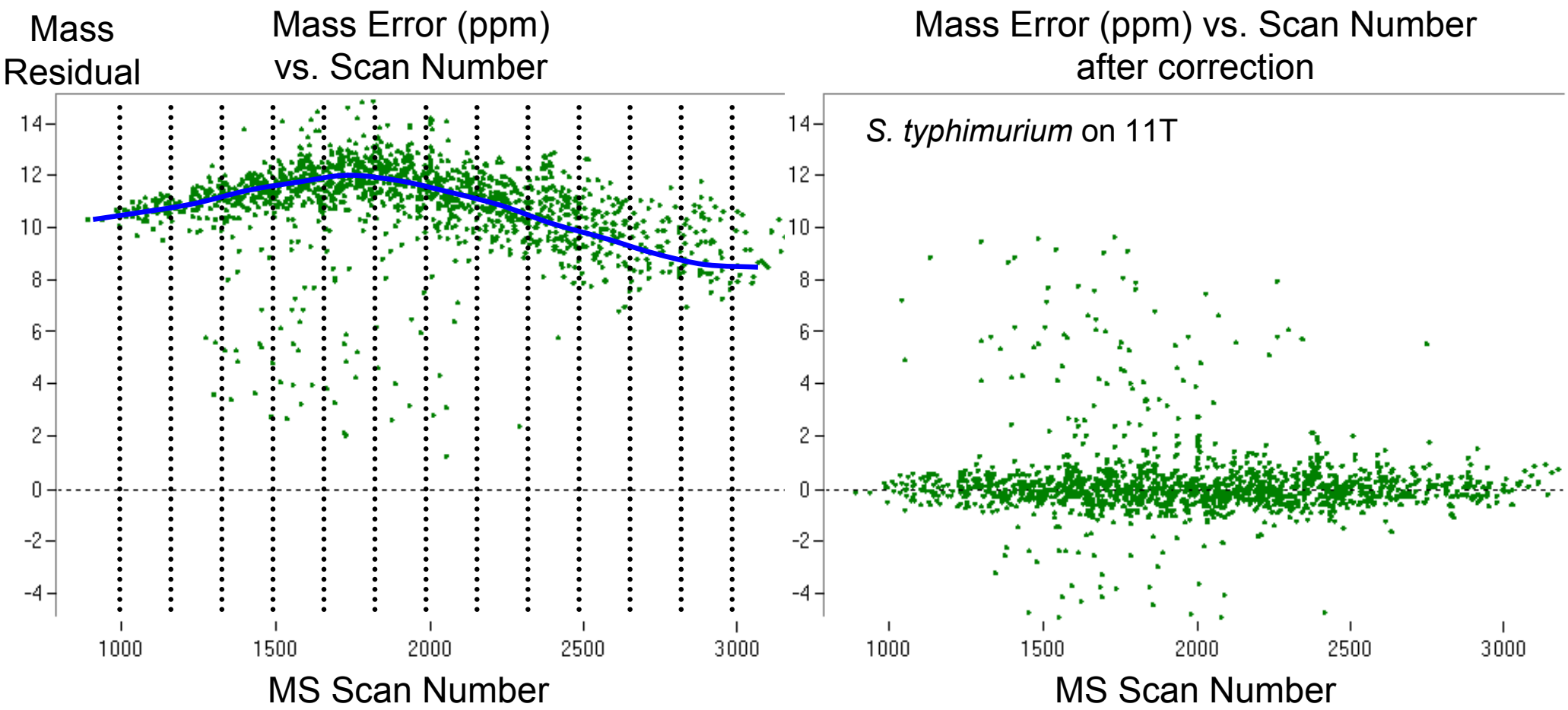
$$\Delta_{mass} = -11.6\text{ppm} \times \frac{\text{mass}_{old}}{1\text{x}10^6\ \text{ppm/Da}}$$

For 1+ feature at 1570.9005 Da,
Δ_{mass} = -0.0182 Da

For 3+ feature at 2919.4658 Da,
Δ_{mass} = -0.0339 Da

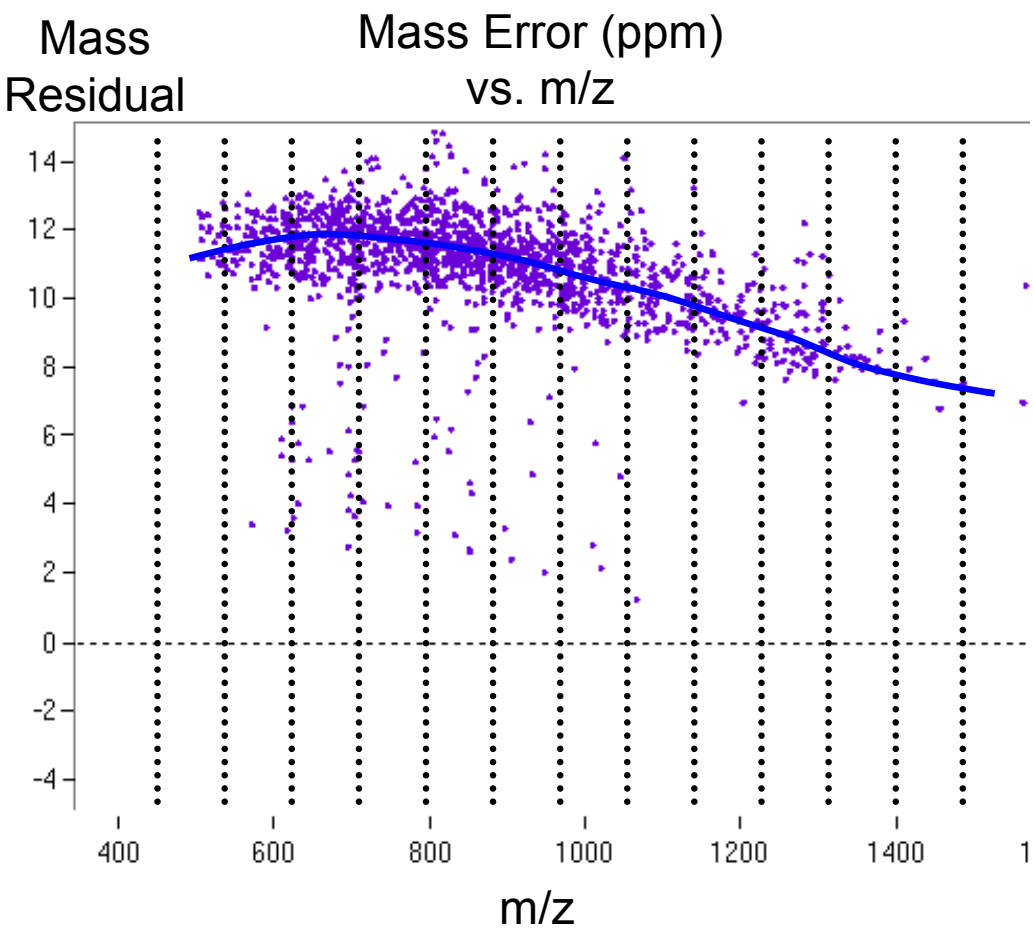
Mass Calibration

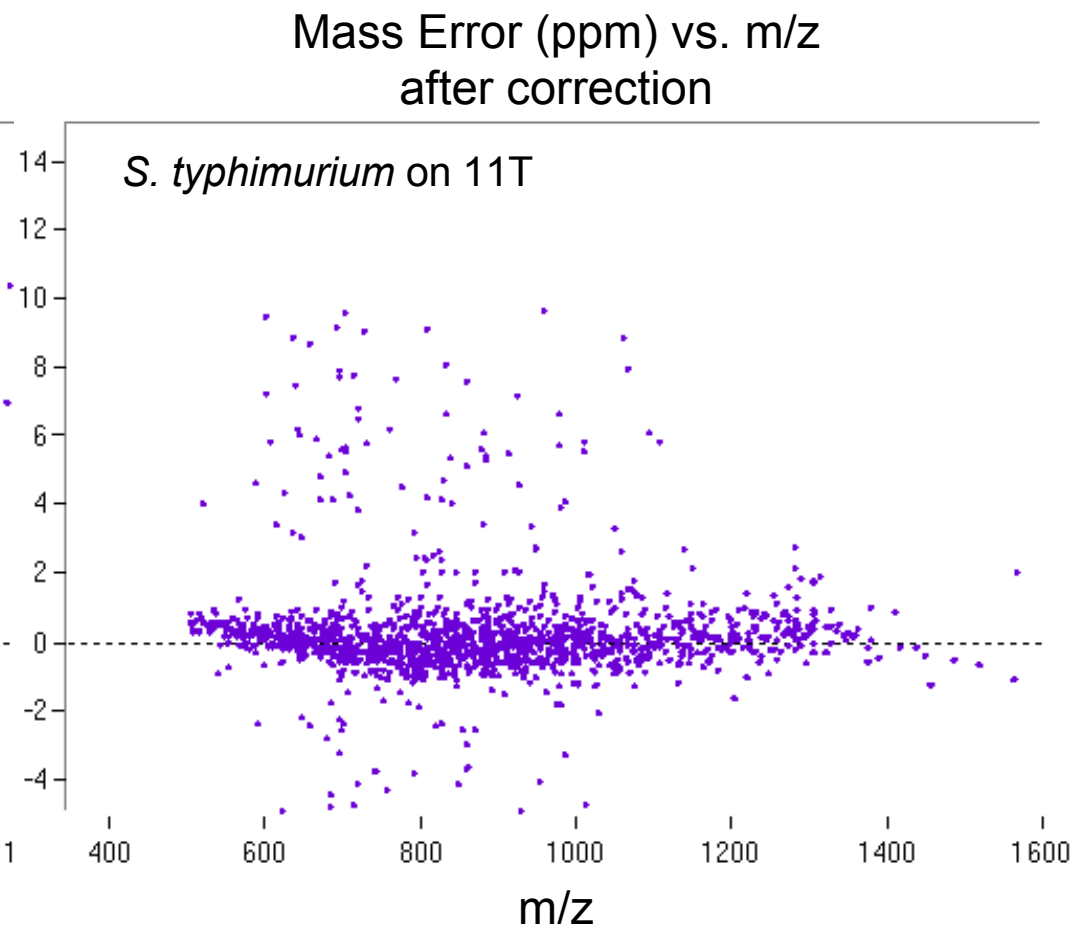
- Option 2: Piece-wise correction via LCMSWarp
  - Examine sections of the data to determine a custom mass shift for each section
  - One option is to divide into time sections



Mass Calibration

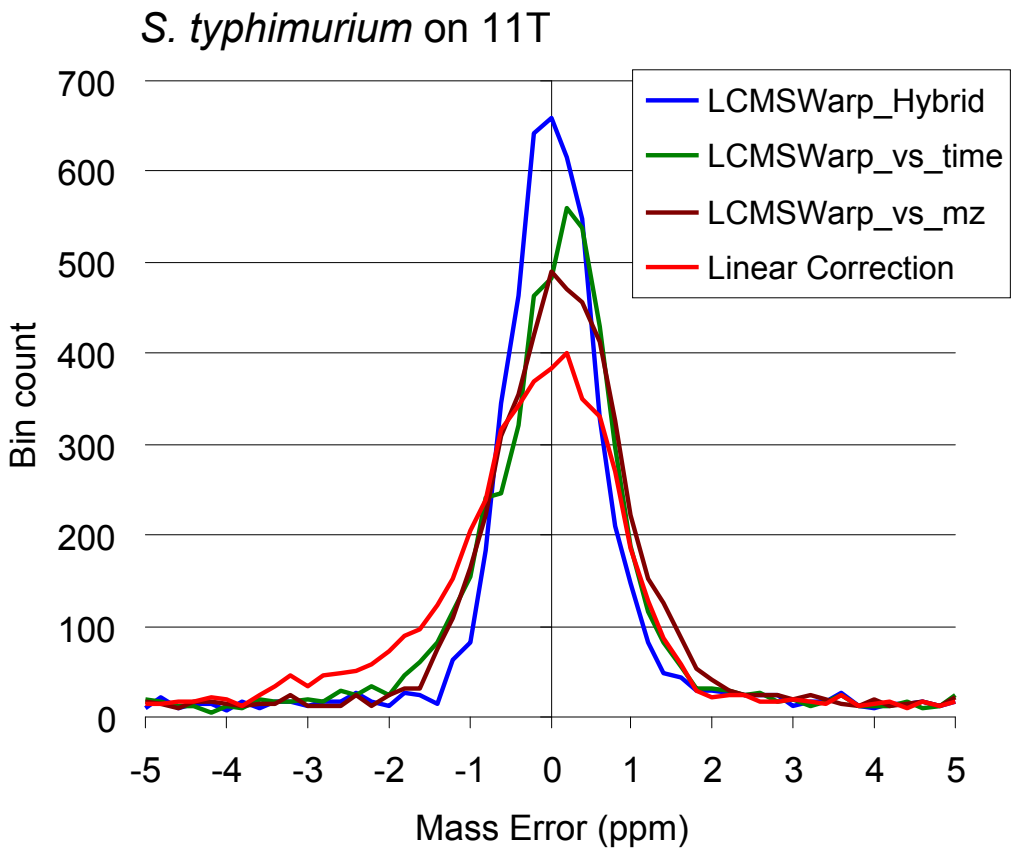
- Option 2: Piece-wise correction via LCMSWarp
  - Second option is to divide into m/z sections
  - LCMSWarp utilizes a hybrid correction based on both mass error vs. time and mass error vs. m/z

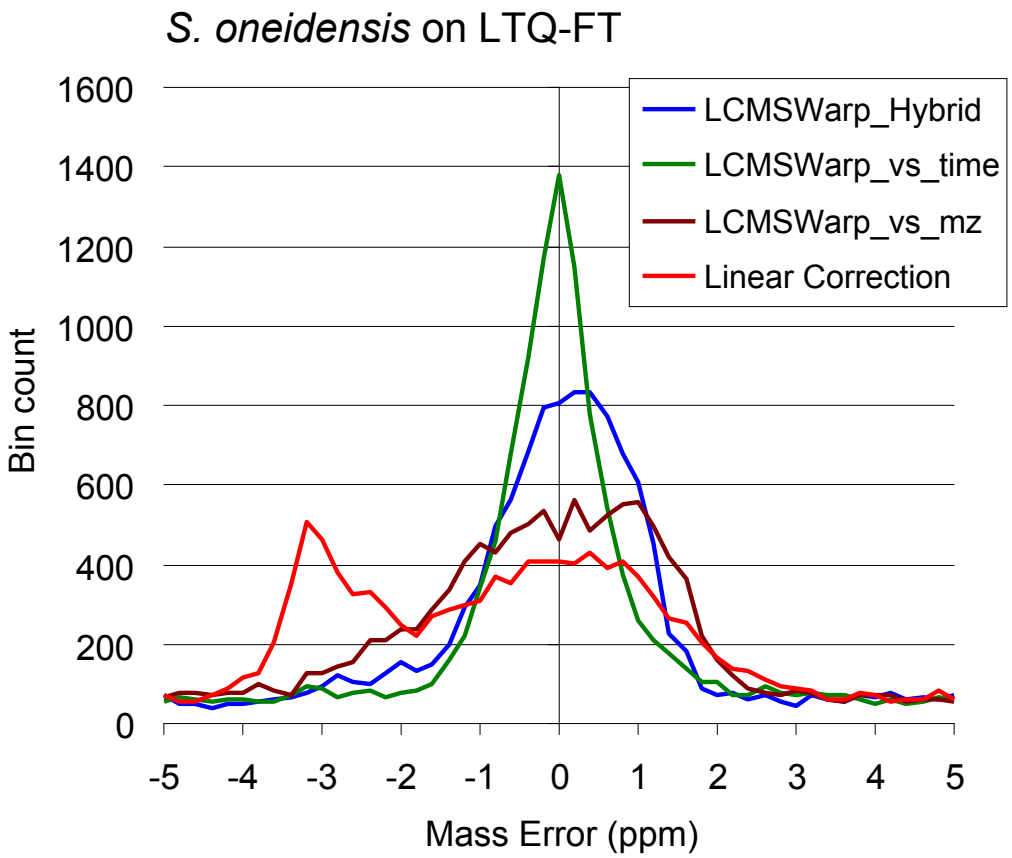




Mass Calibration

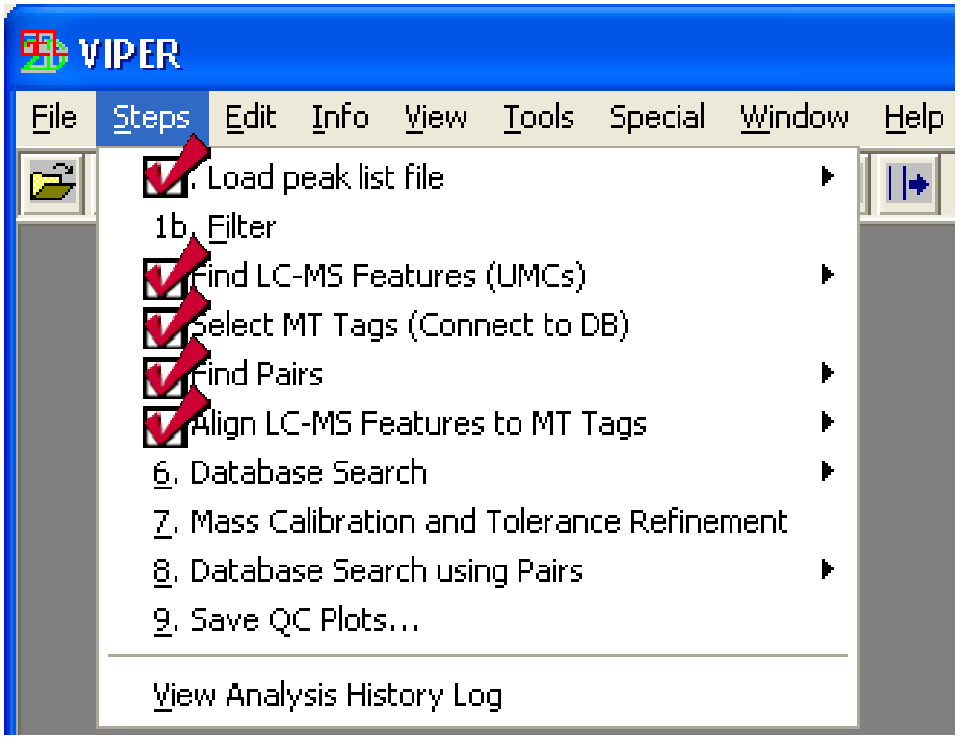
- Comparison of the three methods
  - Mass error histogram gets taller, narrower, and more symmetric
    - Linear → Mass error vs. m/z → Mass error vs. time → Hybrid
  - Not all datasets show the same trends, but Hybrid mass recalibration is generally superior





Identifying LC-MS Features

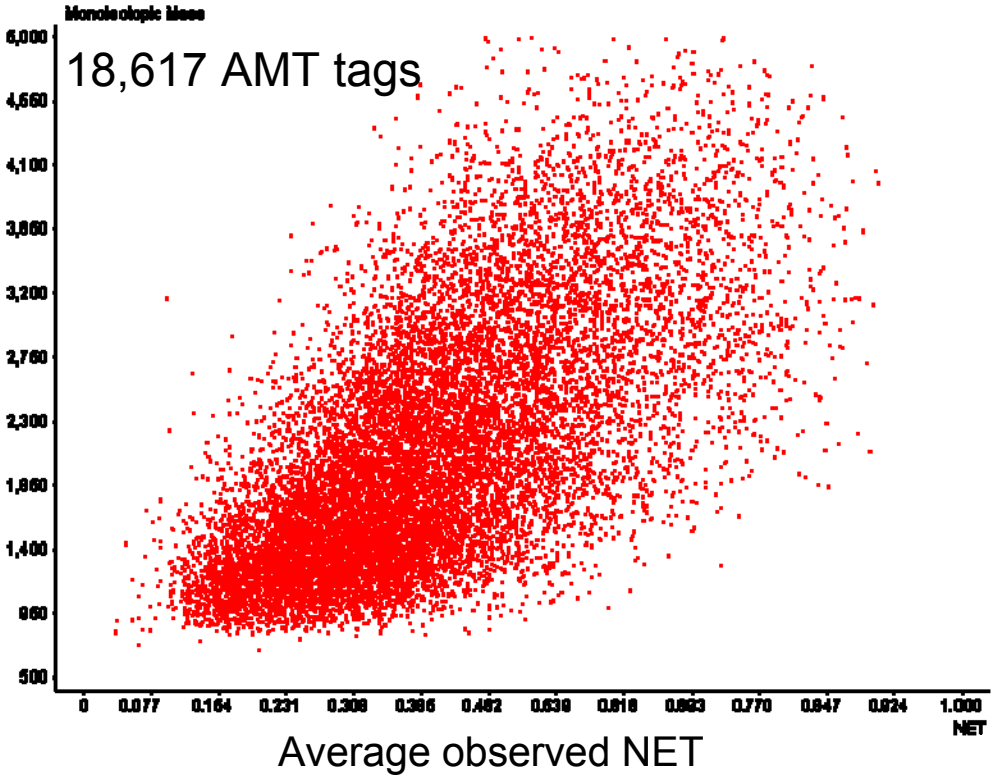
- Peak Matching Steps
  - ☑ Load LC-MS peak lists from Decon2LS
  - ☑ Filter data
  - ☑ Feature definition over elution time
  - ☑ Select AMT tags to match against
  - ☑ Optionally, find paired features (e.g. $^{16}O/^{18}O$ pairs)
  - ☑ Align LC-MS features to AMT tags using LCMSWarp
  - Broad AMT tag DB search
  - Search tolerance refinement
  - Final AMT tag DB search
  - Report results



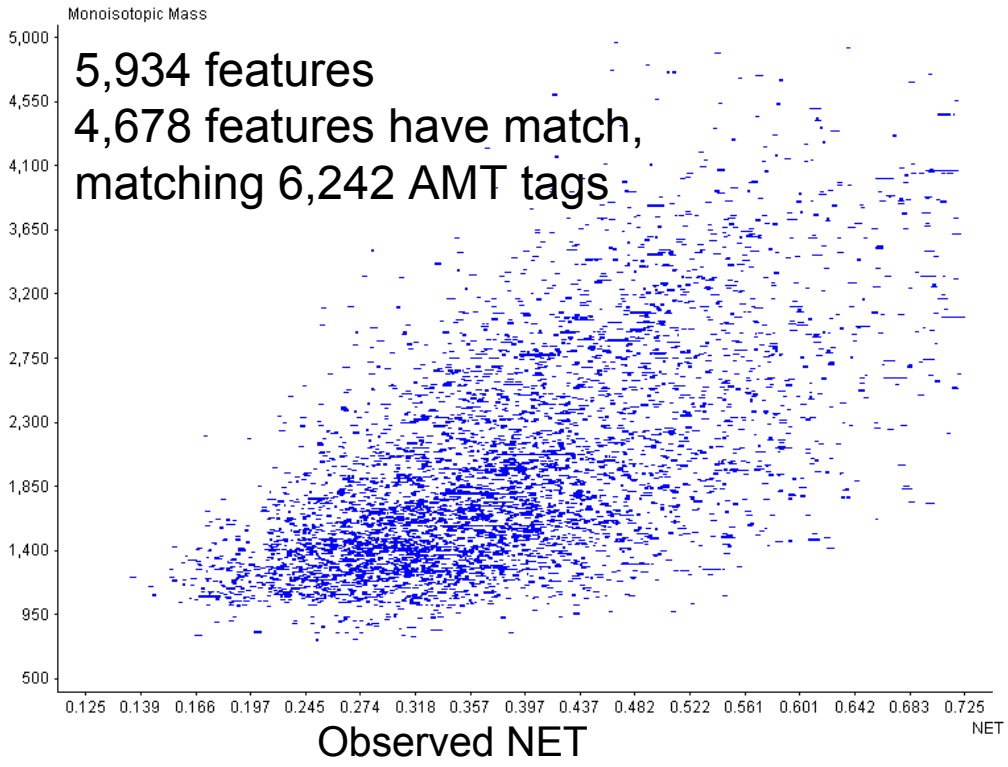
Identifying LC-MS Features

- Match Features to LC-MS/MS IDs
- *S. typhimurium* DB, from 25 LC-MS/MS analyses
  - 18,617 AMT tags, all fully or partially tryptic
  - Look for AMT tags within a broad mass range, e.g., ±25 ppm and ±0.05 NET of each feature

S. typhimurium AMT Tag Database

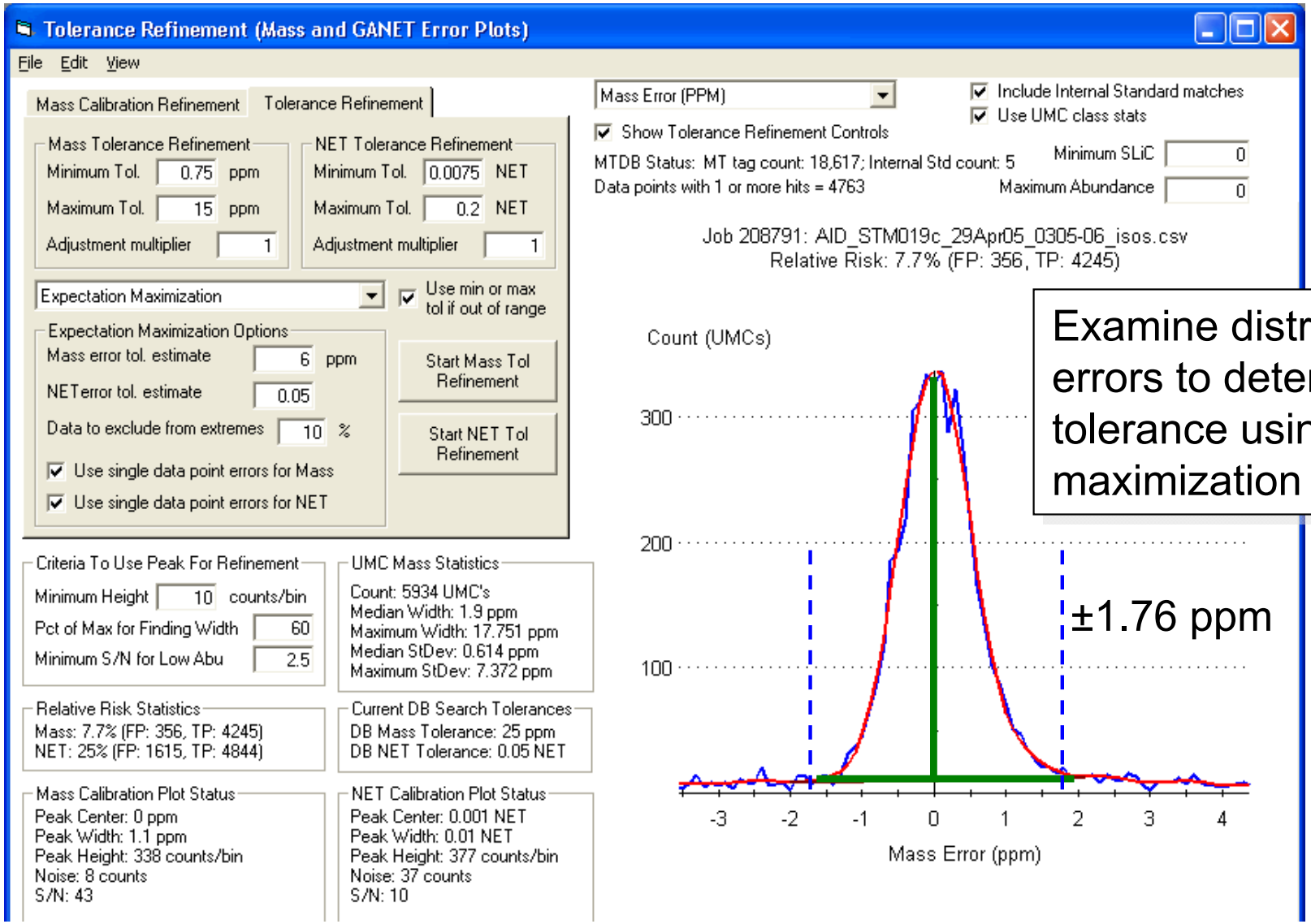


S. typhimurium on 11T FTICR



Search tolerance refinement

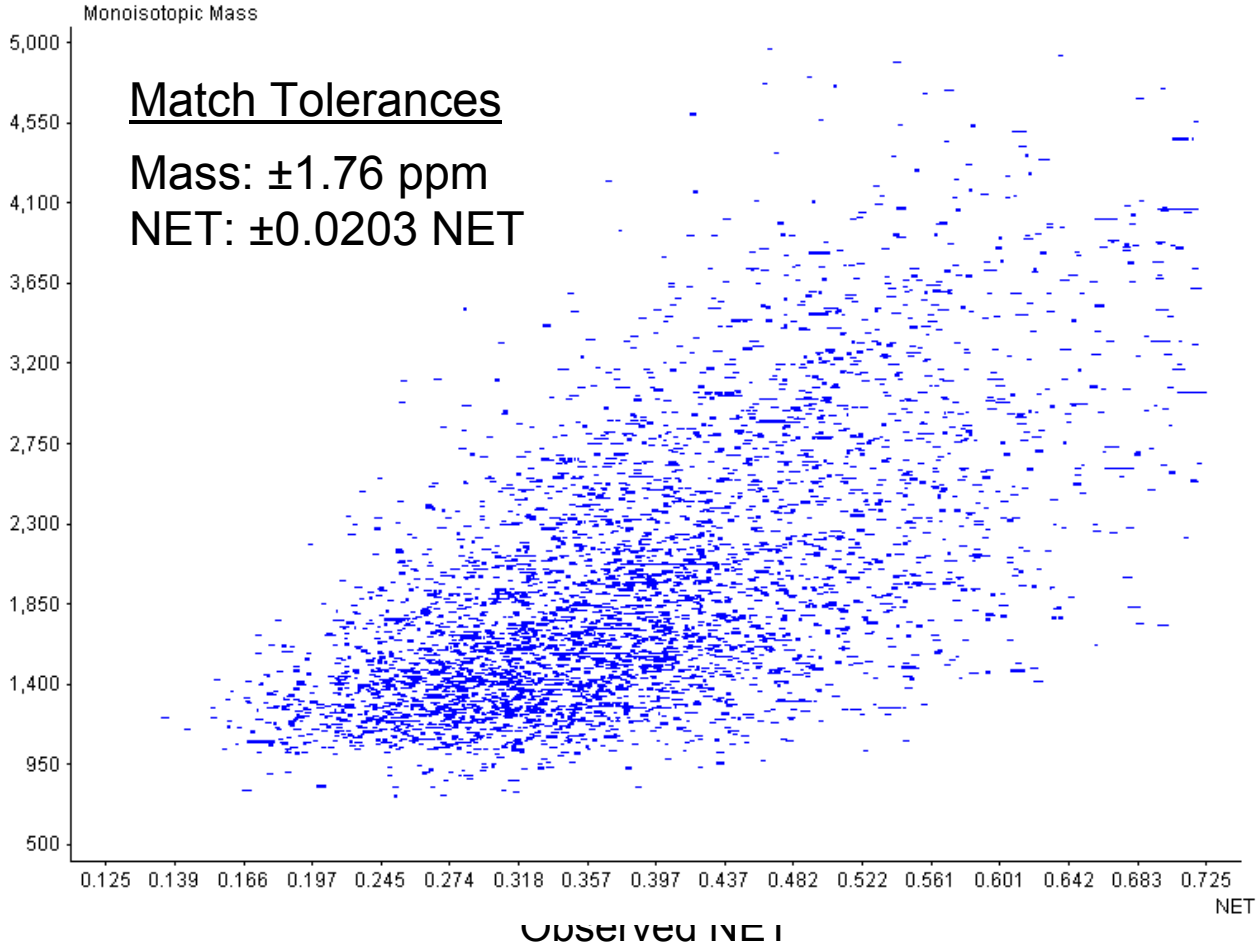
- Can use mass error and NET error histograms to determine optimal search tolerances



Examine distribution of errors to determine optimal tolerance using expectation maximization algorithm

Identifying LC-MS Features

- Repeat search with final search tolerances
  - 5,934 features
  - 3,866 features with matches
  - 3,958 out of 18,617 AMT tags matched using ±1.76 ppm



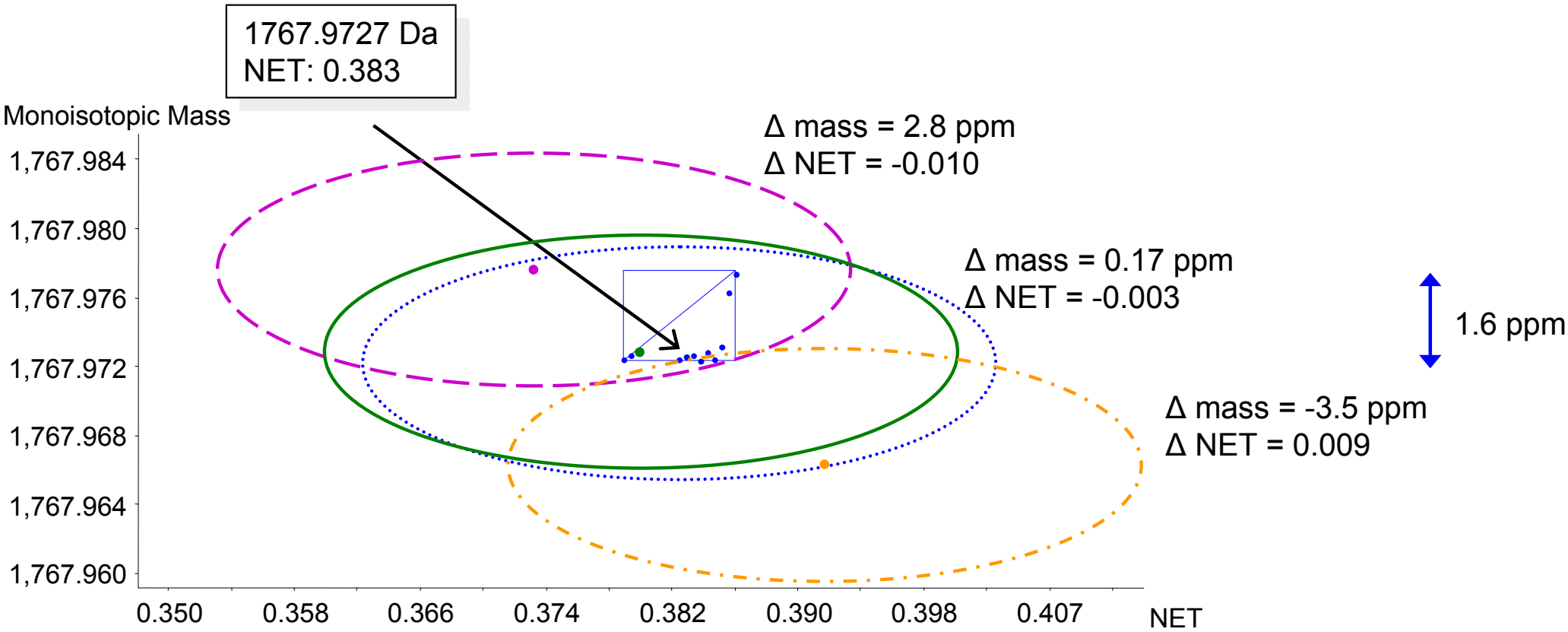
Identifying LC-MS Features

- Given feature can match more than one AMT tag
  - Need measure of ambiguity

Match Tolerances

Mass: ±4 ppm
NET: ±0.02 NET

AMT Tag ID	Peptide	Mass (Da)	NET
35896216	T.RALMQLDEALRPSLR.S	1767.9777	0.373
105490	K.DLETIVGLQTDAPLKR.A	1767.9730	0.380
36259992	R.SIGIAPDVLICRGDRAI.P	1767.9664	0.392



Identifying LC-MS Features

$$d_{ij}^2 = \frac{(m_i - \mu_{mj})^2}{\sigma_{mj}^2} + \frac{(t_i - \mu_{tj})^2}{\sigma_{tj}^2}$$

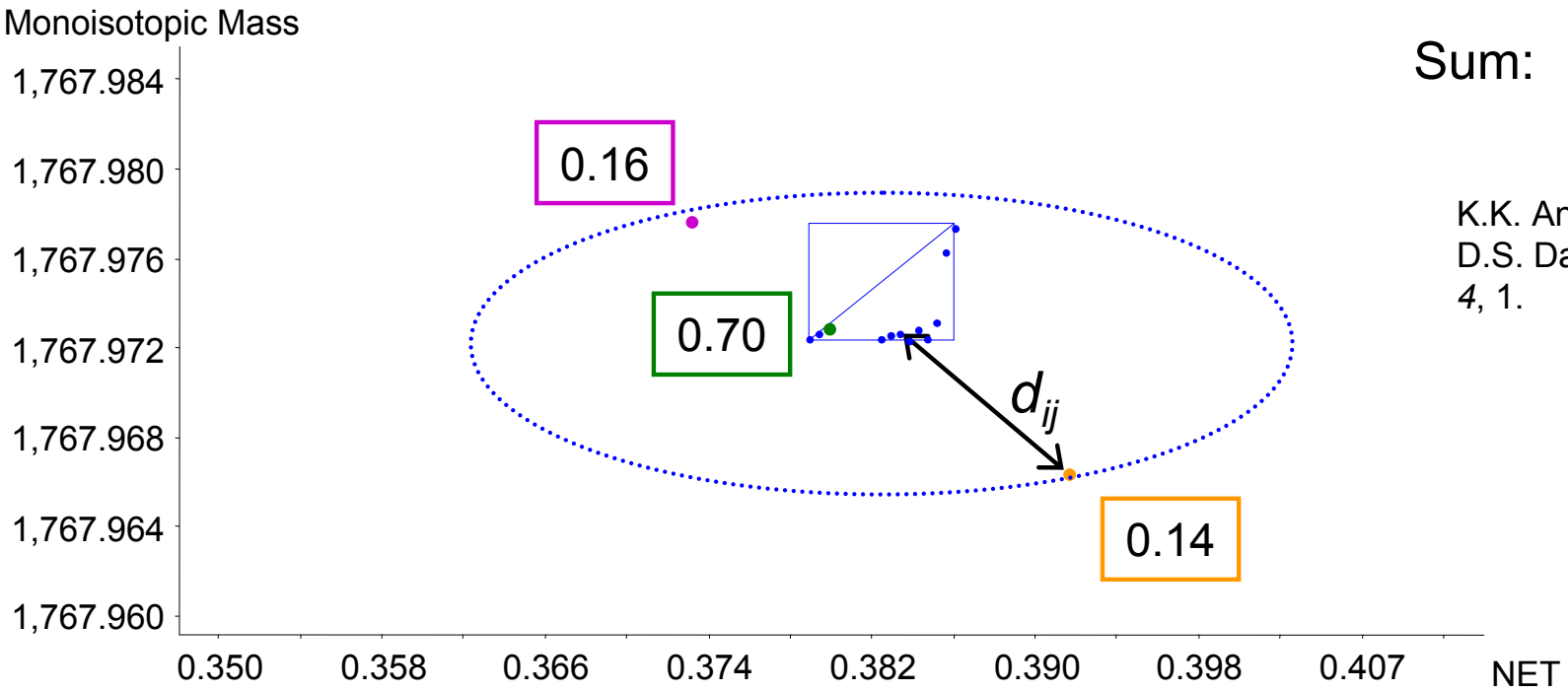
σ_{mj} = 4 ppm, σ_{tj} = 0.025

$$p_{ij} = \frac{(\sigma_{mj}\sigma_{tj})^{-1}\exp(-d_{ij}^2/2)}{\left(\sum_{k=1}^{N}(\sigma_{mk}\sigma_{tk})^{-1}\exp(-d_{ik}^2/2)\right)}$$

Match Tolerances

Mass: ±4 ppm
NET: ±0.02 NET

AMT Tag ID	Mass (Da)	NET	d_{ij}^2	Numerator	p_{ij}
35896216	1767.9777	0.373	3.012	6273.3	0.16
105490	1767.9730	0.380	0.090	27042.5	0.70
36259992	1767.9664	0.392	3.267	5521.4	0.14
			Sum:	38837.2	

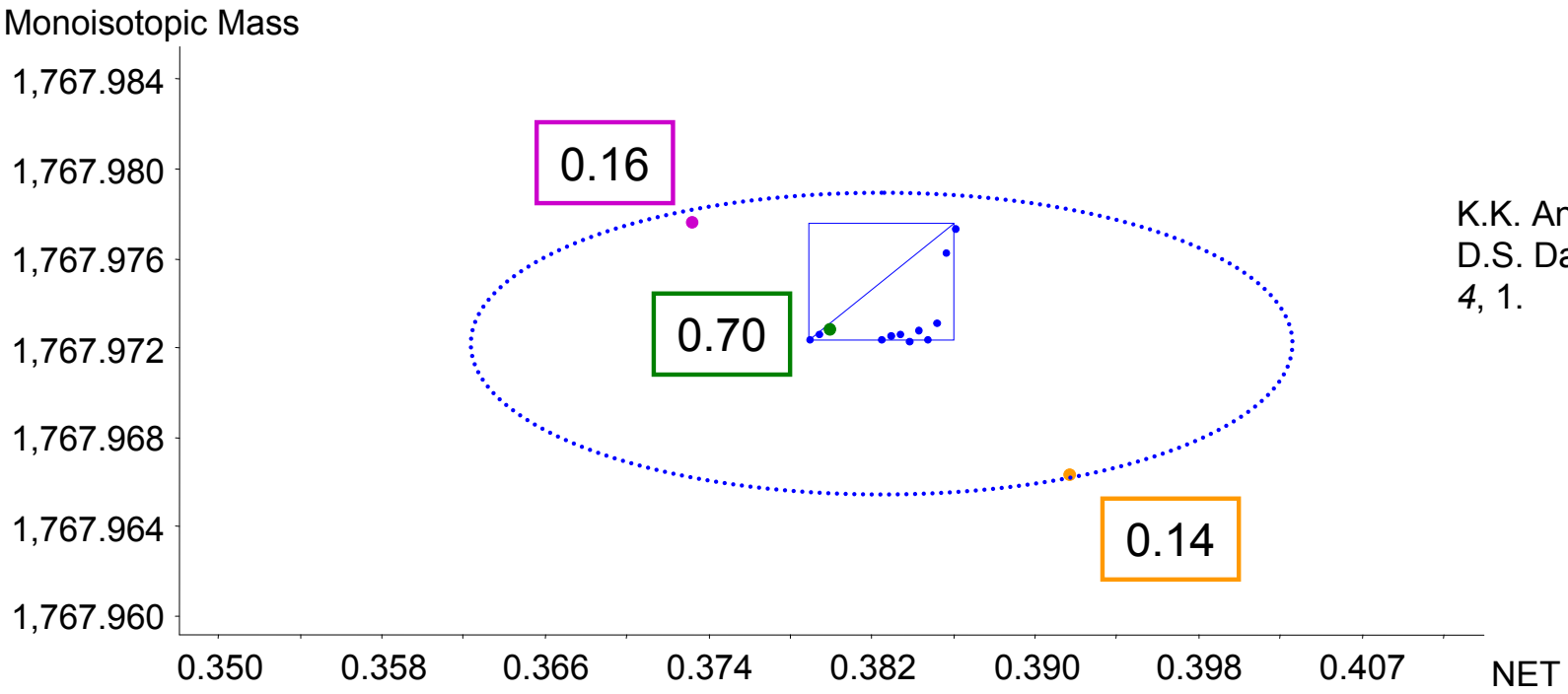


K.K. Anderson, M.E. Monroe, and D.S. Daly. *Proteome Science* **2006**, *4*, 1.

Identifying LC-MS Features

- SLiC: Spatially Localized Confidence Score
  - Measures uniqueness of match

AMT Tag ID	Peptide	Mass (Da)	NET	SLiC Score	Average XCorr	Avg Disc Score
35896216	T.RALMQLDEALRPSLR.S	1767.9777	0.373	0.16	3.13	0.61
105490	K.DLETIVGLQTDAPLKR.A	1767.9730	0.380	0.70	3.68	0.97
36259992	R.SIGIAPDVLICRGDRAI.P	1767.9664	0.392	0.14	2.15	0.06

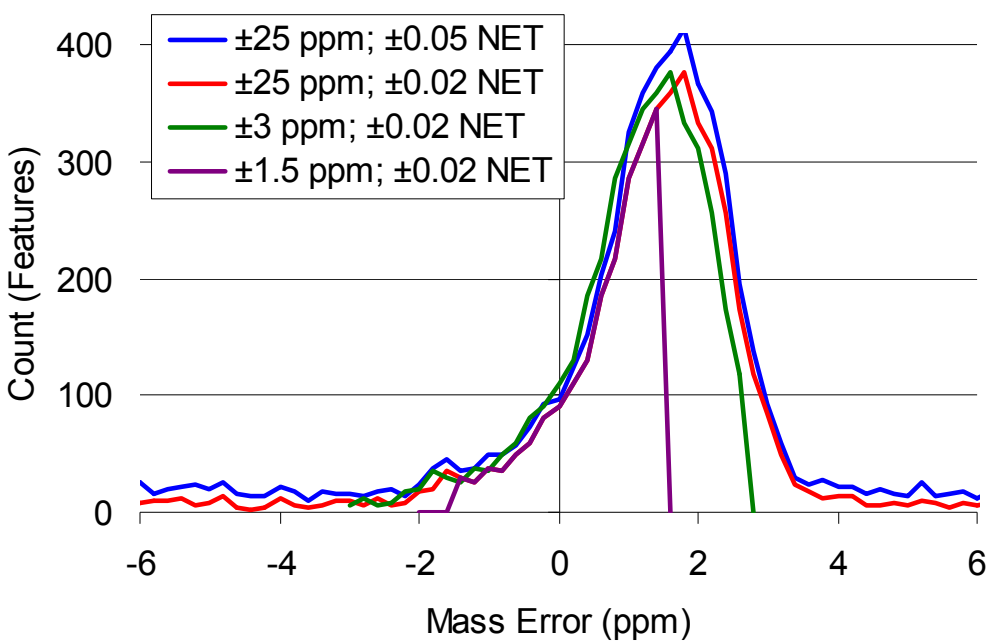


K.K. Anderson, M.E. Monroe, and D.S. Daly. *Proteome Science* **2006**, *4*, 1.

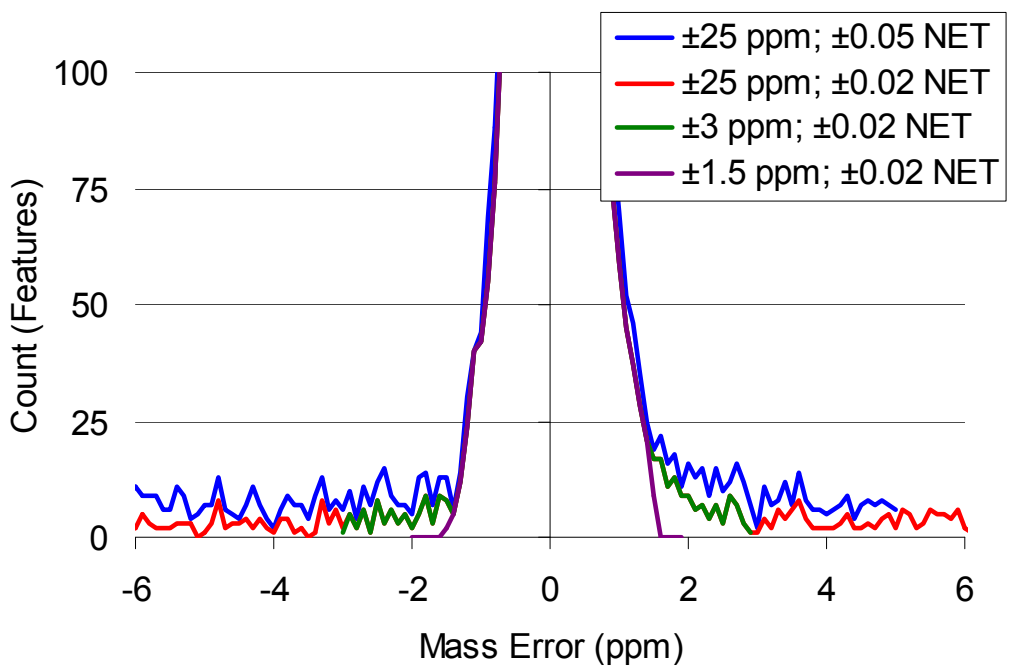
Search tolerance refinement

- Effect of search tolerances on Mass Error histogram
  - If mass error plot not centered at 0, then narrow mass windows exclude valid data
  - Decreasing mass and/or NET tolerance reduces background false positive level

Mass error histograms with linear mass correction



Mass error histograms with LCMSWarp mass correction

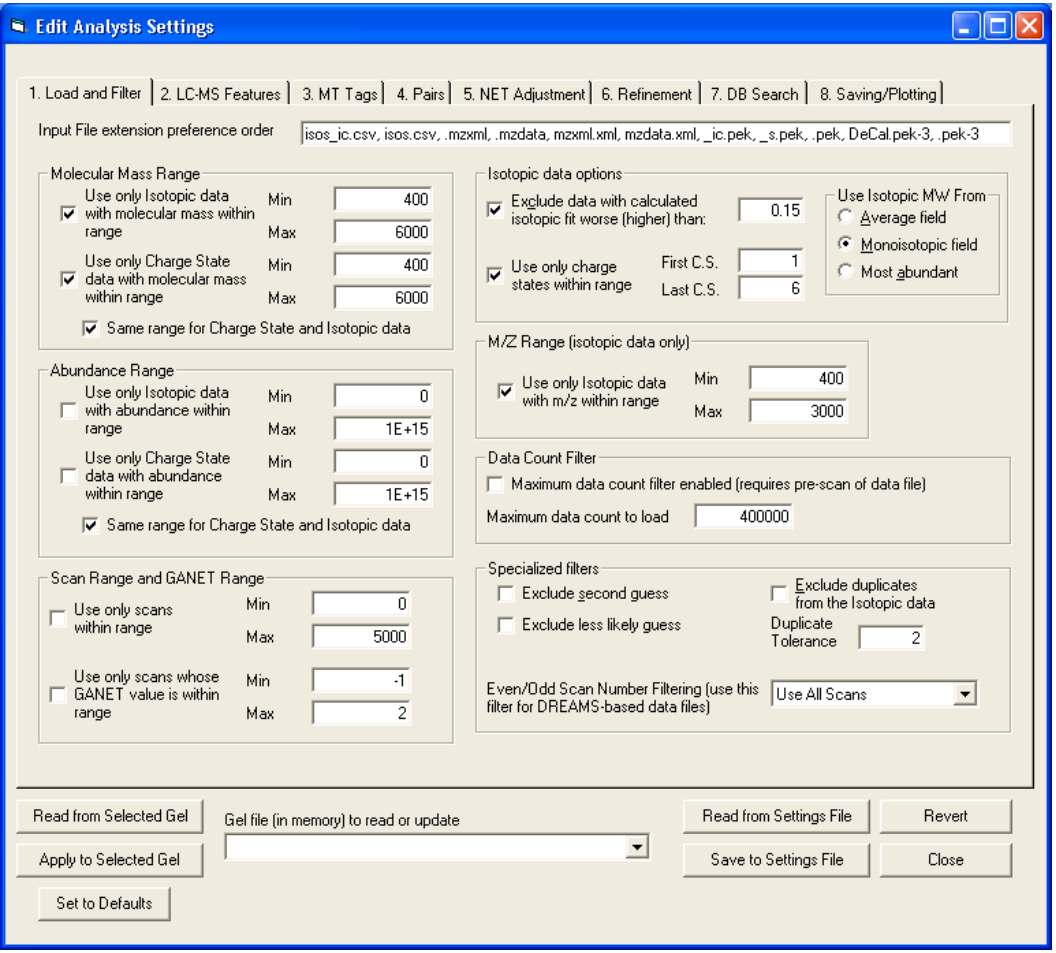


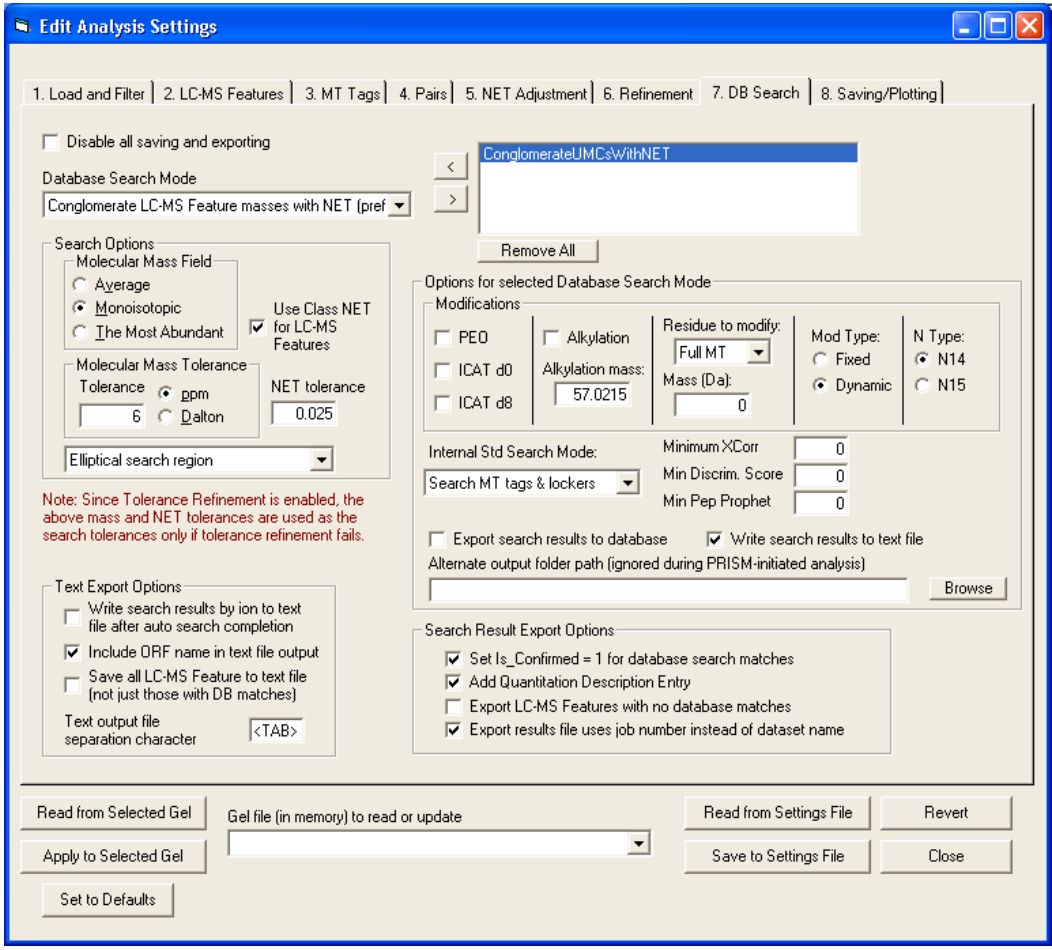
Identifying LC-MS Features

- Peak Matching Steps
  - ☑ Load LC-MS peak lists from Decon2LS
  - ☑ Filter data
  - ☑ Feature definition over elution time
  - ☑ Optionally, find paired features (e.g. $^{16}O/^{18}O$ pairs)
  - ☑ Align LC-MS features to AMT tags using LCMSWarp
  - ☑ Broad AMT tag DB search
    - ±25 ppm and ±0.05 NET
  - ☑ Search tolerance refinement
  - ☑ Final AMT tag DB search
    - e.g. ±1.8 ppm and ±0.02 NET
  - Report results

Automated Peak Matching

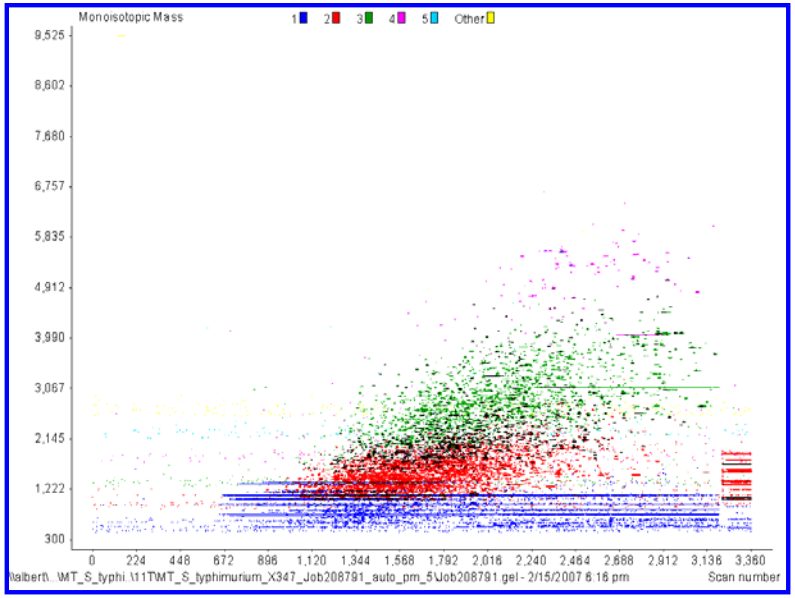
- Automated processing using VIPER
  - Processing steps and parameters defined in .Ini file
    - Separate .Ini file for $^{14}N/^{15}N$ pairs and $^{16}O/^{18}O$ pairs



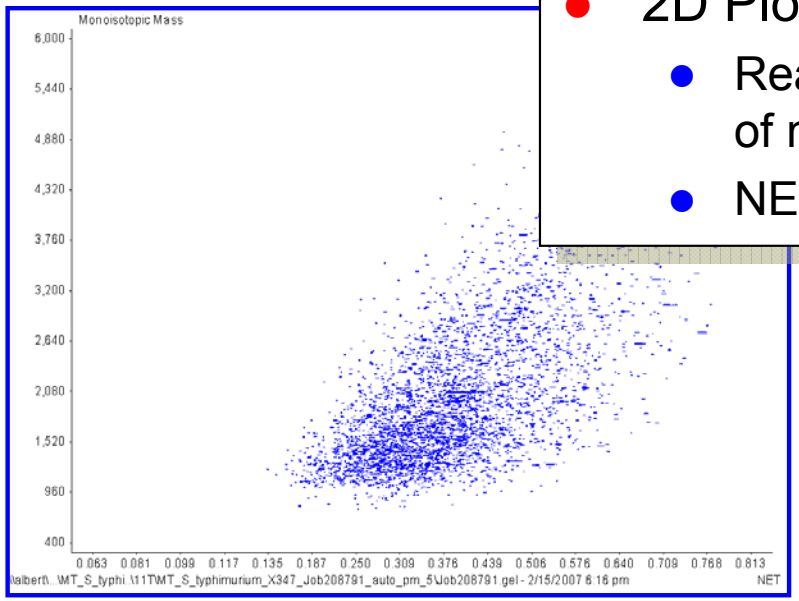


Peak Matching Results

- Browsable result folders for visual QC of each dataset
  - *S. typhimurium* on 11T FTICR

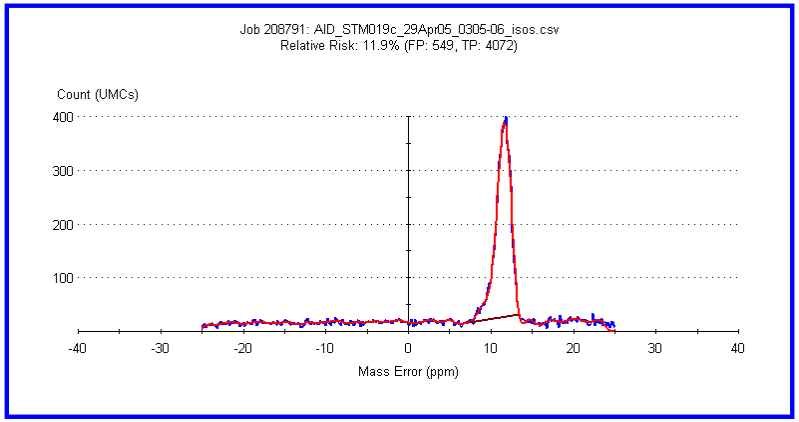


Data Searched

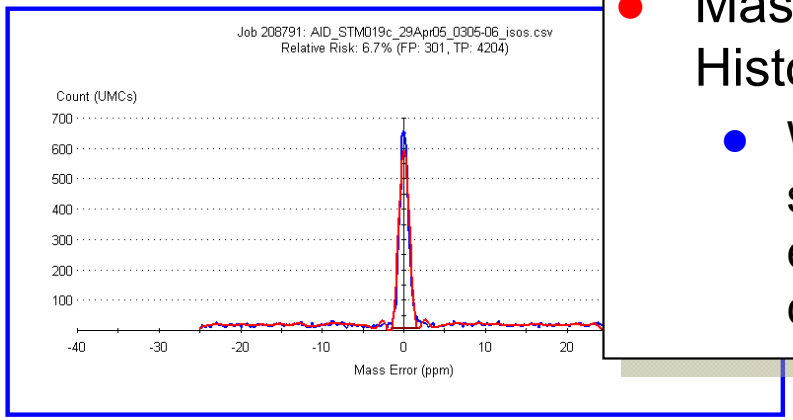


Data With Matches

- 2D Plot Metrics
  - Reasonable number of matches
  - NET range ≈ 0 to 1



Mass Errors Before Refinement

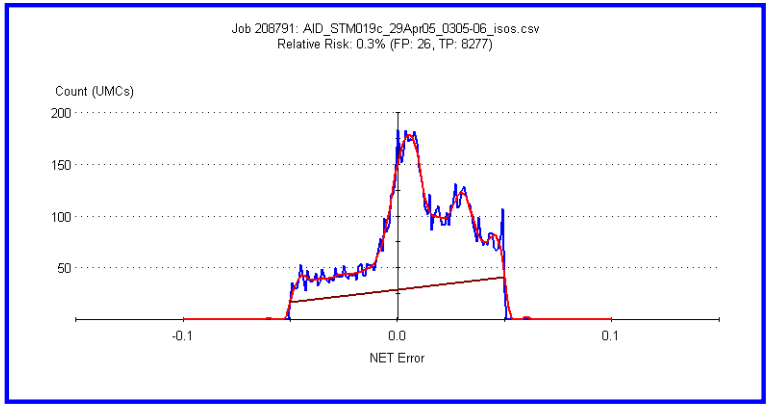


Mass Errors After Refinement

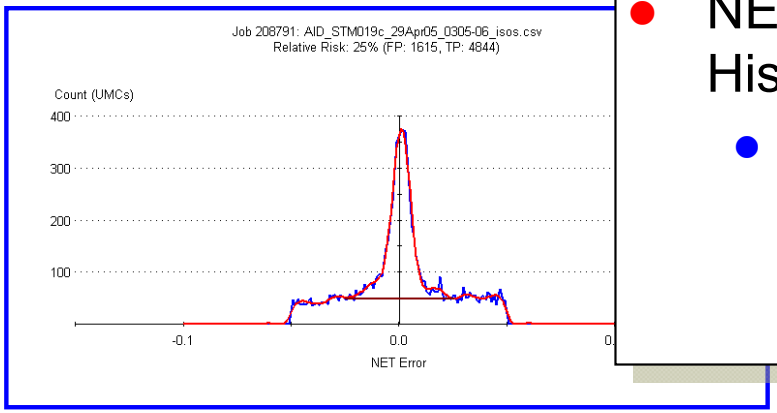
- Mass Error Histogram Metrics
  - Well defined, symmetric mass error peak centered at 0 ppm

Peak Matching Results

- Browsable result folders for visual QC of each dataset
  - *S. typhimurium* on 11T FTICR

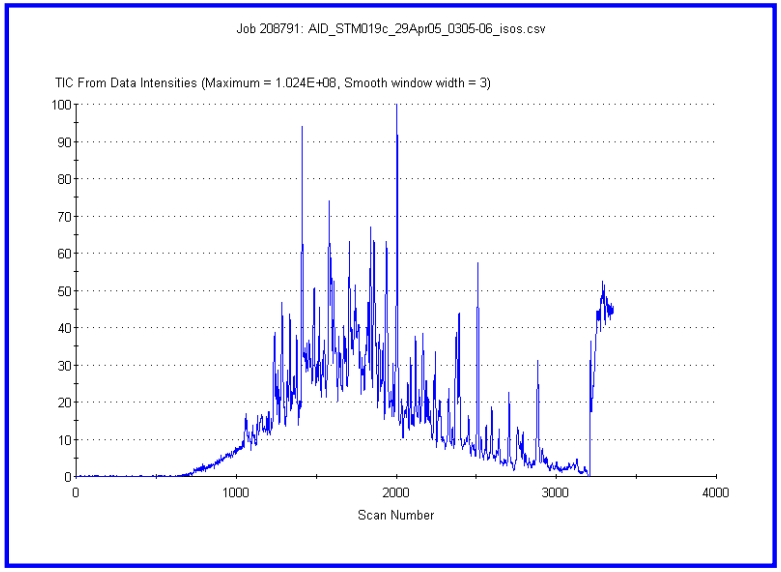


NET Errors Before Refinement

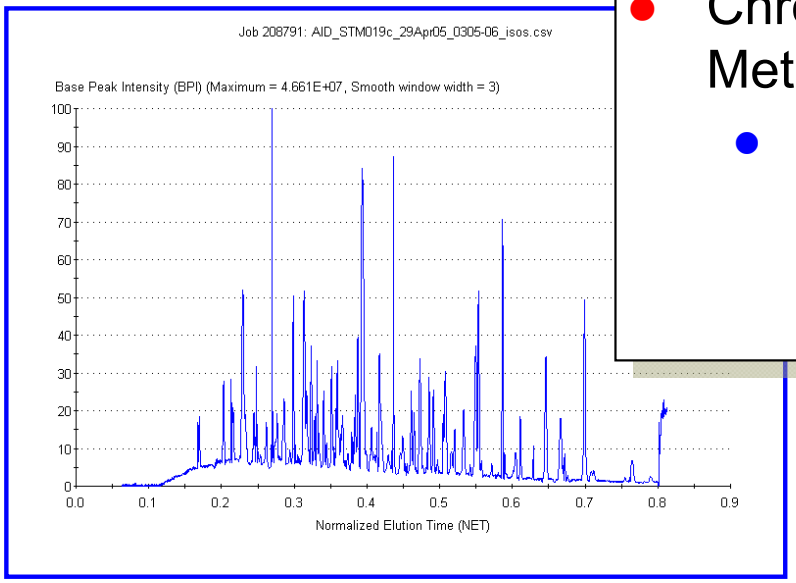


NET Errors After Refinement

- NET Error Histogram Metrics
  - Well defined, symmetric NET error peak centered at 0



Total Ion Chromatogram (TIC)

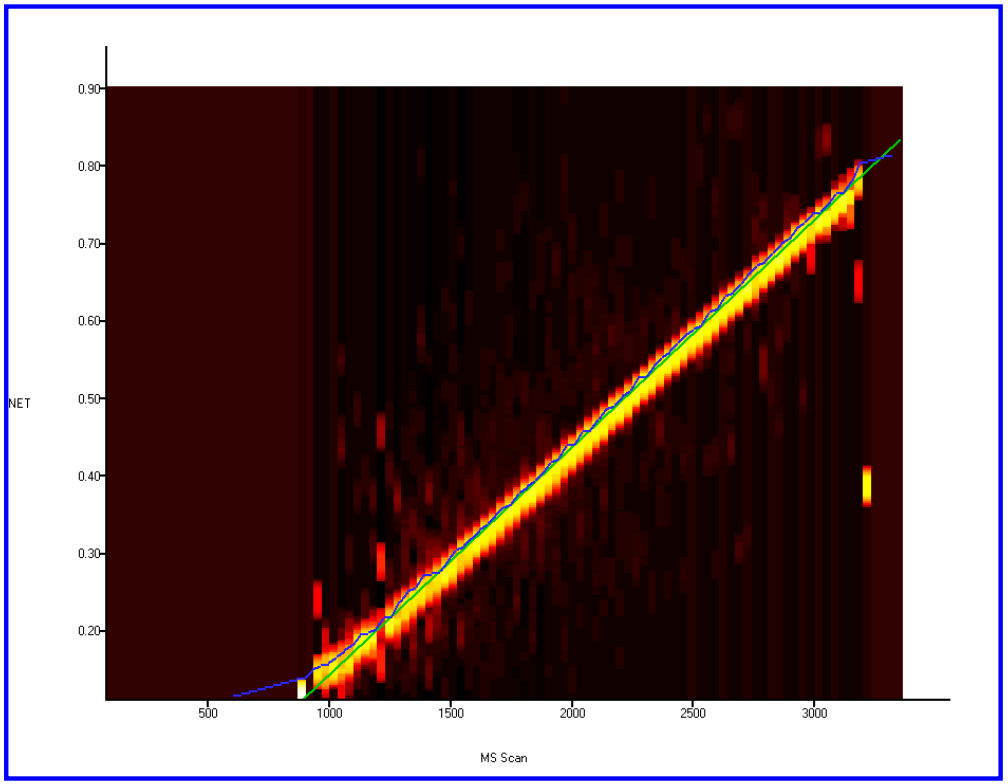


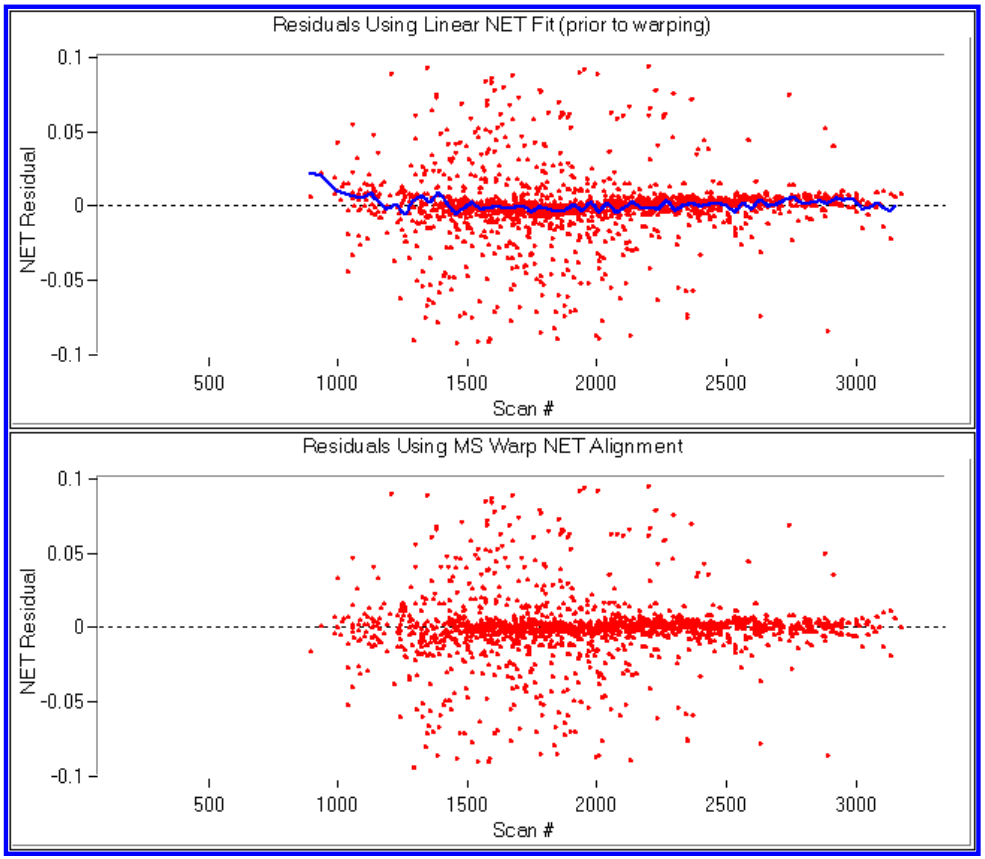
Base Peak Intensity (BPI) Chromatogram

- Chromatogram Metrics
  - Narrow peaks evenly distributed throughout separation window

Peak Matching Results

- Browsable result folders for visual QC of each dataset
  - *S. typhimurium* on 11T FTICR



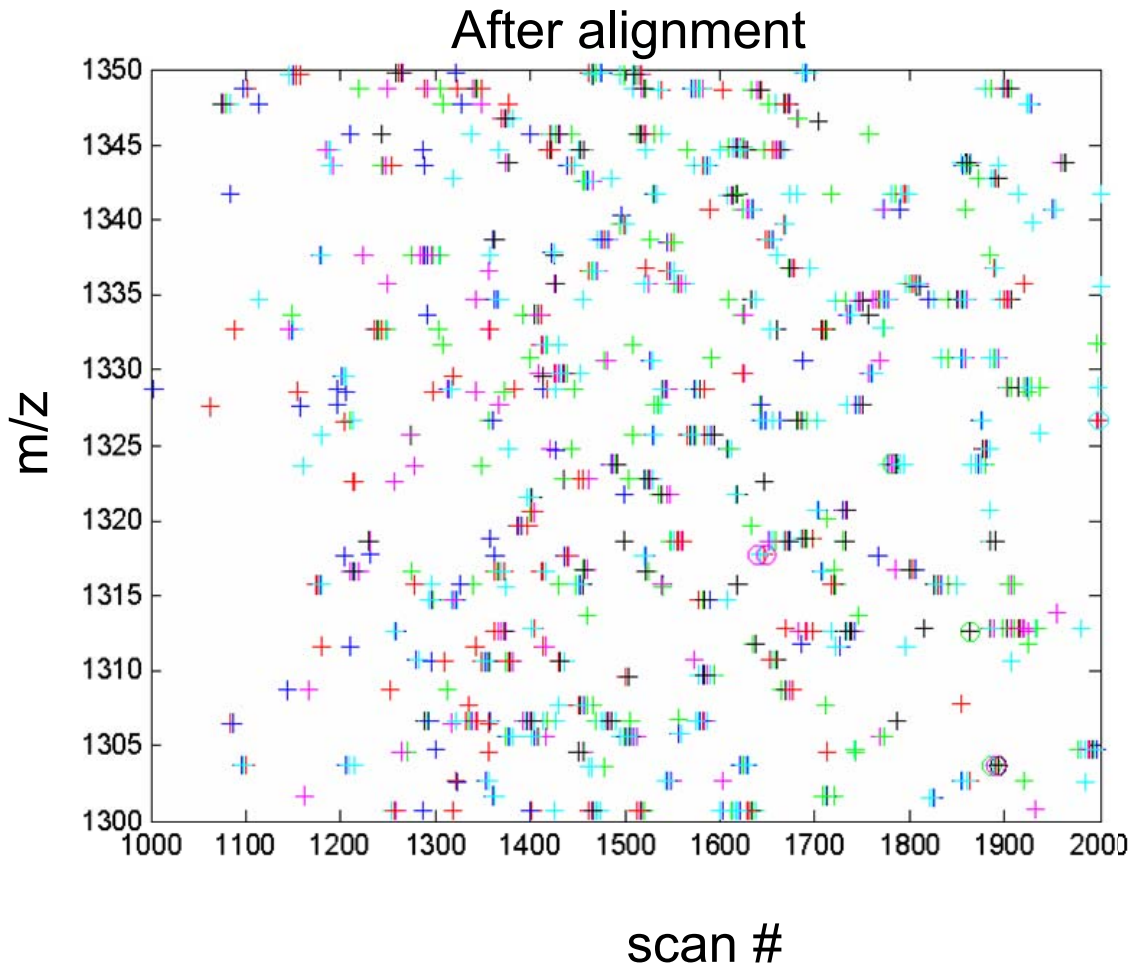


- NET Alignment Surface Metrics
  - Should show a smooth, bright yellow, diagonal line

- NET Alignment Residual Metrics
  - Data after recalibration should be narrowly distributed around zero

Peak Matching Results

- What about the unmatched LC-MS features?
  - Could align LC-MS features across datasets
  - Find the unmatched ones that show interesting trends
  - Generate list of the mass and elution times for the interesting features
  - Re-analyze the sample to perform targeted LC-MS/MS
  - Alignment example for 36 datasets using prototype software tool



LC-MS Feature Discovery

- Similar approaches and software tools: High Res LC-MS
  - SpecArray (Pep3D, mzXML2dat, PepList, PepMatch, PepArray)
    - X.-J. Li, et. al. *Mol Cell Proteomics* **2005**, *4*, 1328-1340.
  - msInspect
    - M. Bellew et. al. *Bioinformatics* **2006**, *22*, 1902-1909.
  - PEPPeR
    - J. Jaffe et.al. *Mol. Cell. Proteomics* **2006**, *5*, 1927-1941.
  - XCMS (for Metabolite profiling)
    - C.A. Smith et. al. *Analytical Chemistry* **2006**, *78*, 779-787.
  - Surromed label-free quantitation software (MassView)
    - W. Wang et al. *Analytical Chemistry* **2003**, *75*, 4818-4826.

LC-MS Feature Discovery

- Similar approaches and software tools: Low Res LC-MS
  - Signal maps software
    - A. Prakash et. al. *Mol. Cell Proteomics* **2006**, *5*, 423-432.
  - Informatics platform for global proteomic profiling using LC-MS
    - D. Radulovic, et al. *Mol. Cell. Proteomics* **2004**, *3*, 984-997.
  - Computational Proteomics Analysis System (CPAS)
    - A. Rauch et. al. *J. Proteome Research* **2006,** *5*, 112-121.

Outline

- ✓ Introduction
- ✓ Feature discovery in LC-MS datasets
  - ✓ Feature discovery in individual spectra
  - ✓ Feature definition over elution time
- ✓ Identifying LC-MS Features using an AMT tag DB