

Profiling a large number of peptides in complex mixtures using LC/ESI-MS Peptide Identification

Martin McIntosh

PI: Computational Proteomics
Laboratory

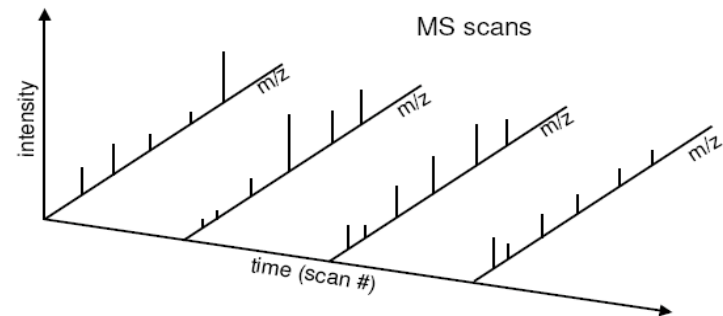
Fred Hutchinson Cancer Research
Center

Proteomics.fhcrc.org

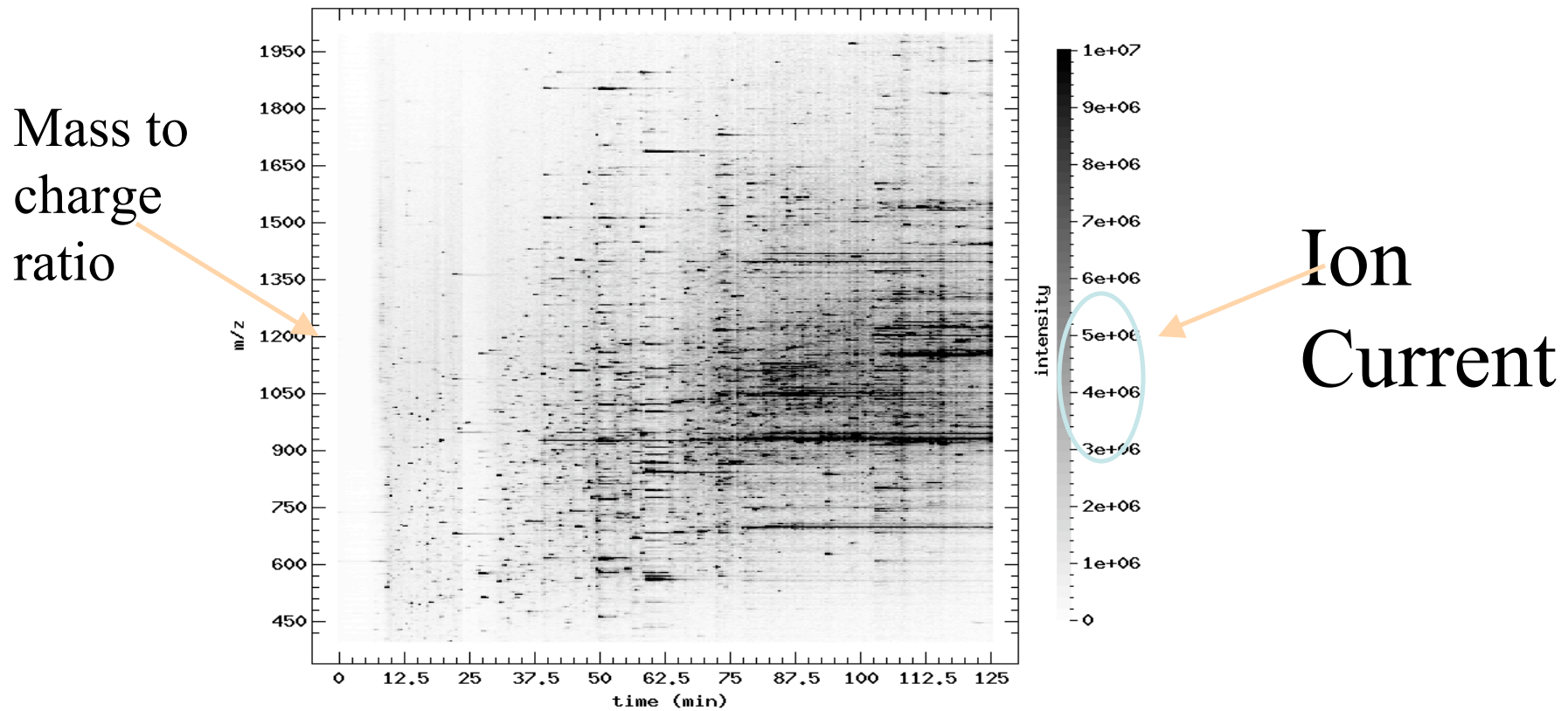
LC/ESI-TOF

- ESI (electro spray ionization):
 - Proteins liquid charged gas-phase ions.
- TOF (time of flight) mass spec:
 - Estimates **mass to charge ratio** of ions by their time of flight.
- LC: Liquid chromatography:
 - A procedure for on-line separation of complex mixture.

Mass vs. Intensity vs. Time

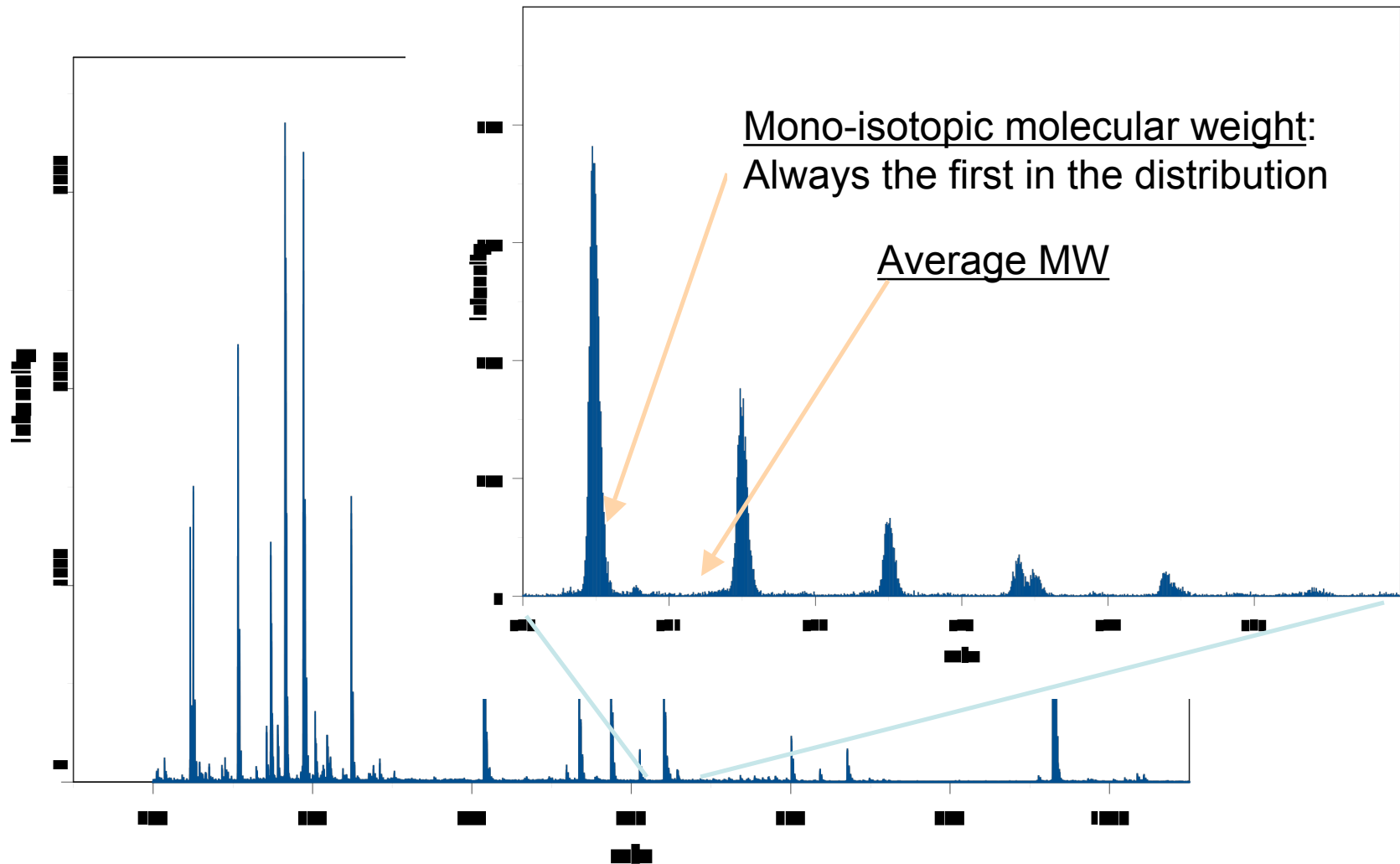


LC/ESI-TOF Measurement



Time of MS measurement

Data reduction: Isotopes in TOF-MS

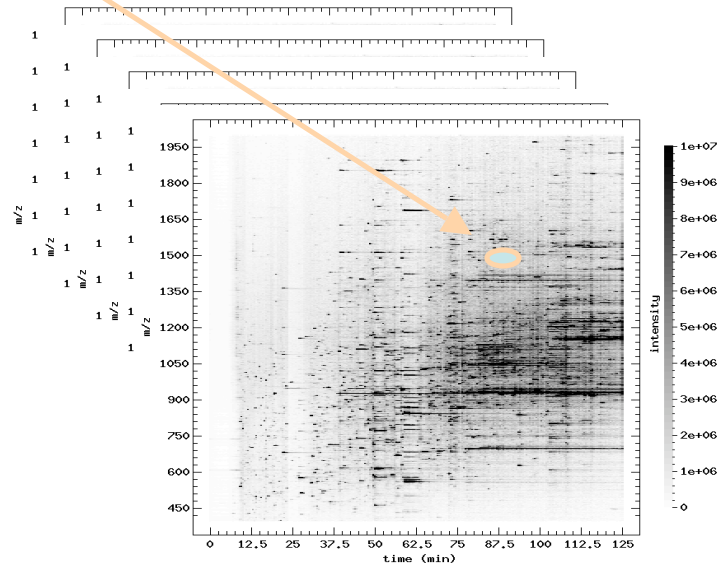
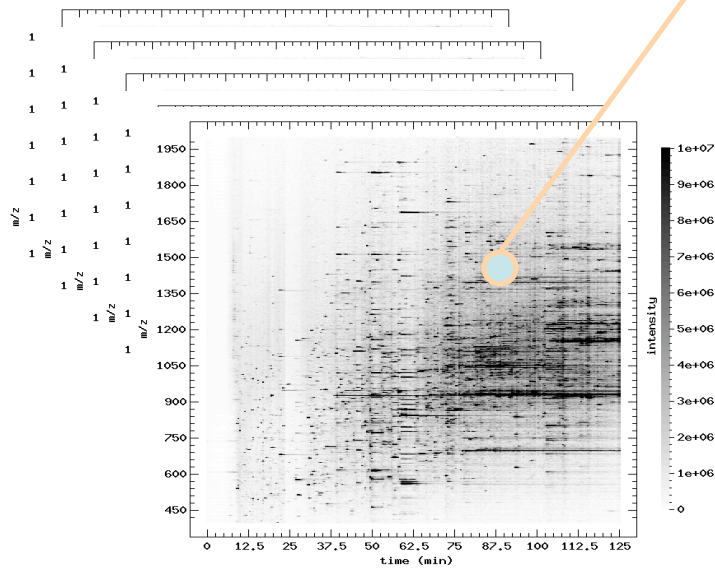


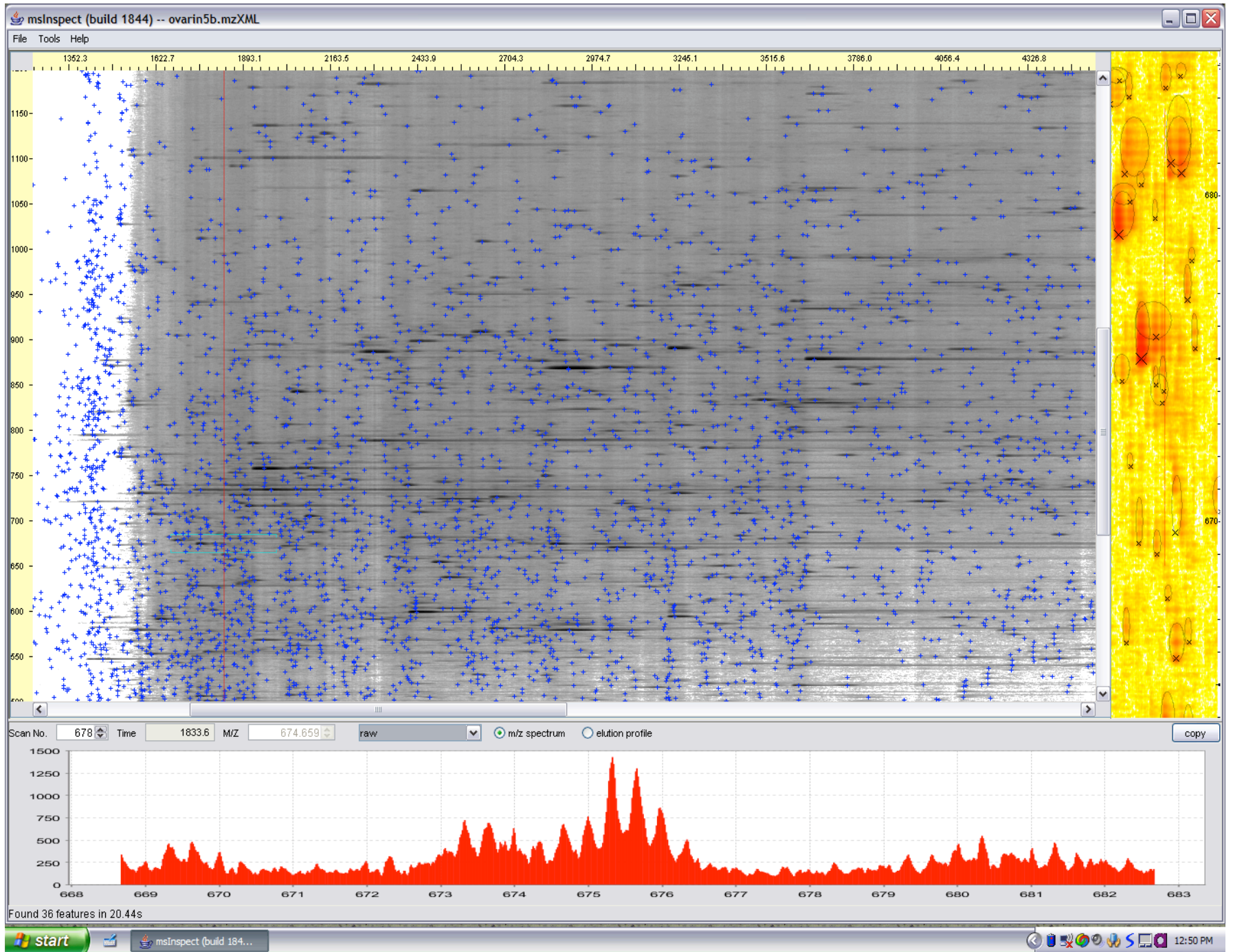
Outline of msInspect

Tool for peptide/protein signature
analysis

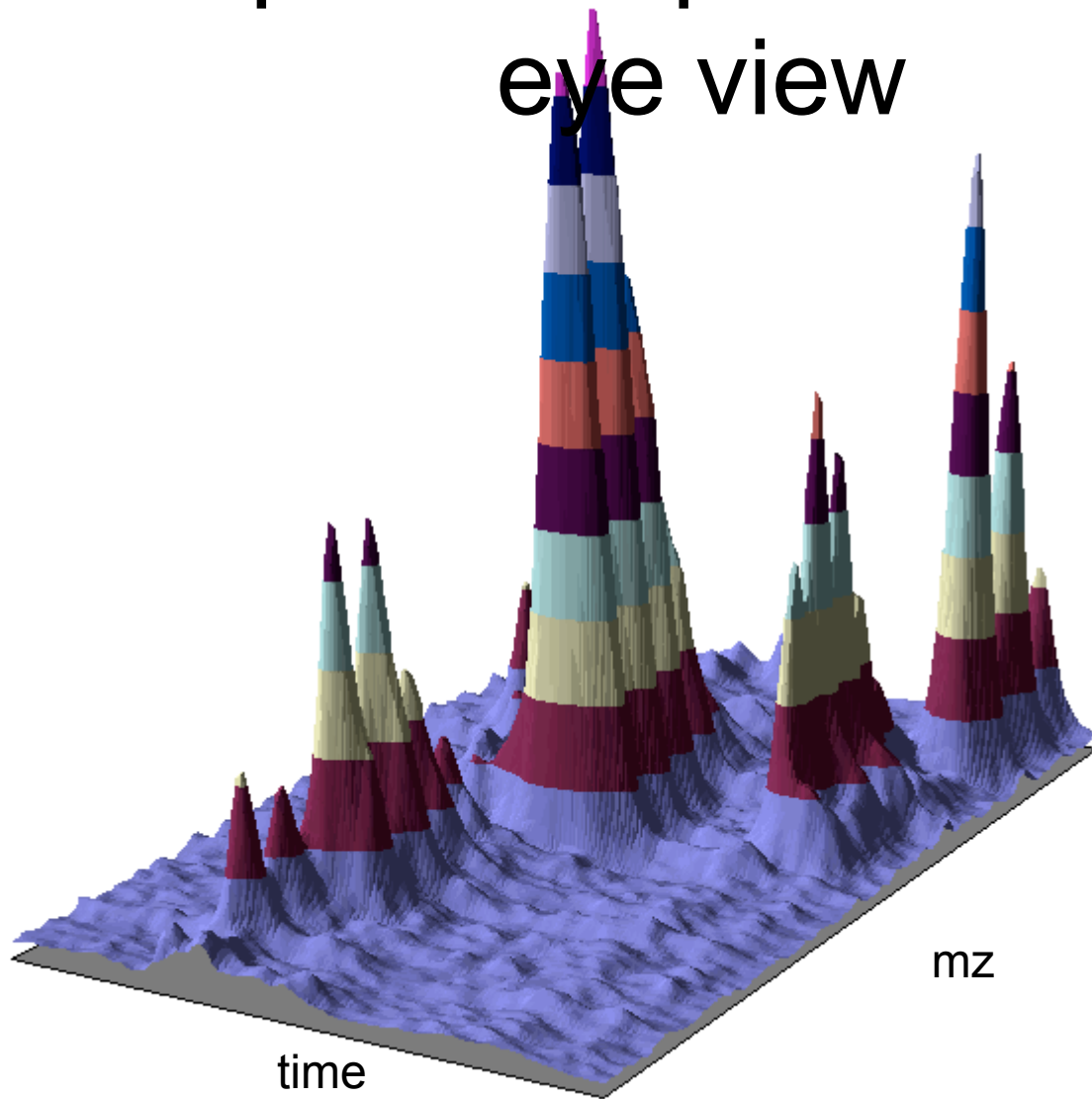
Biomarker Discovery

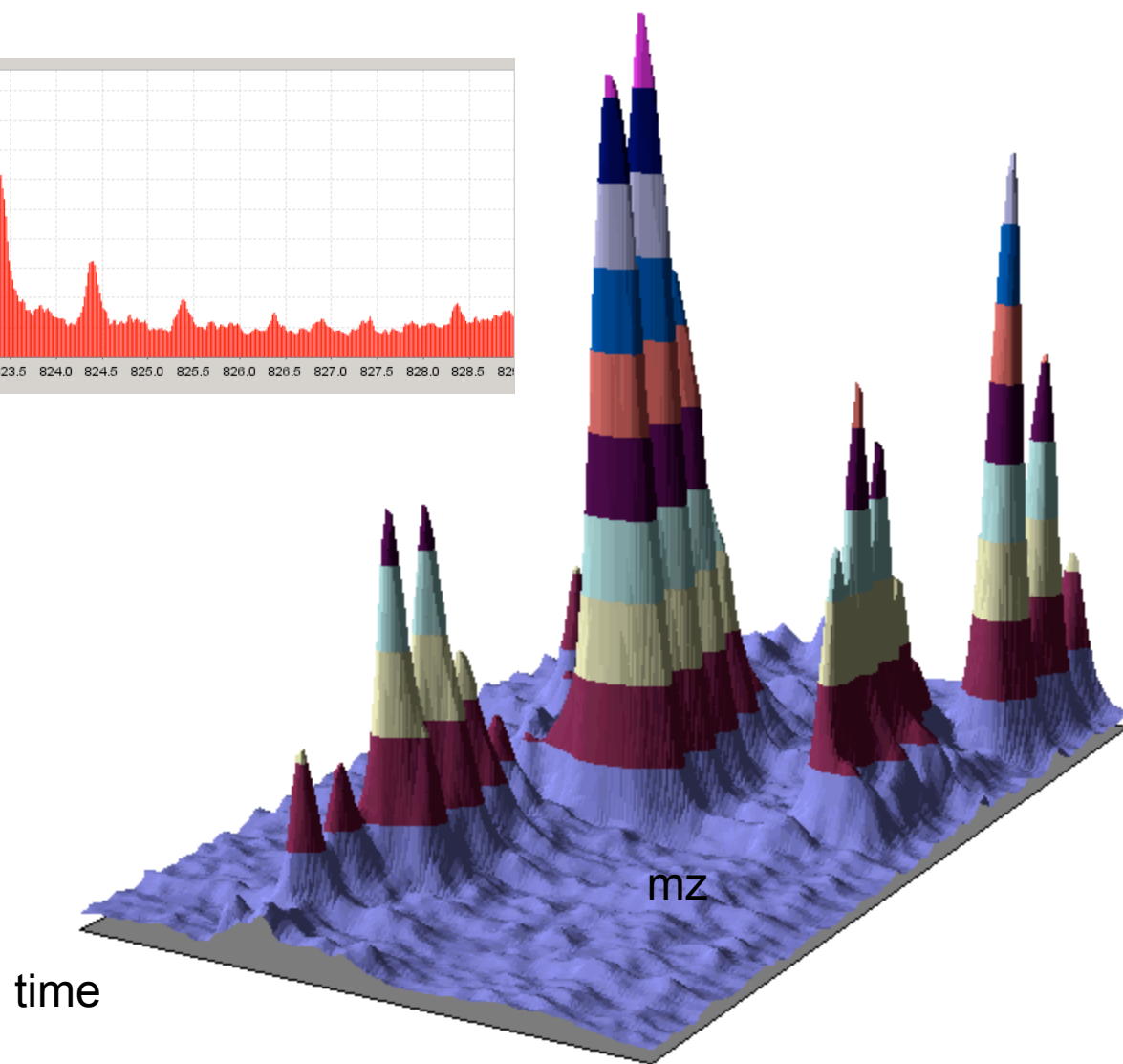
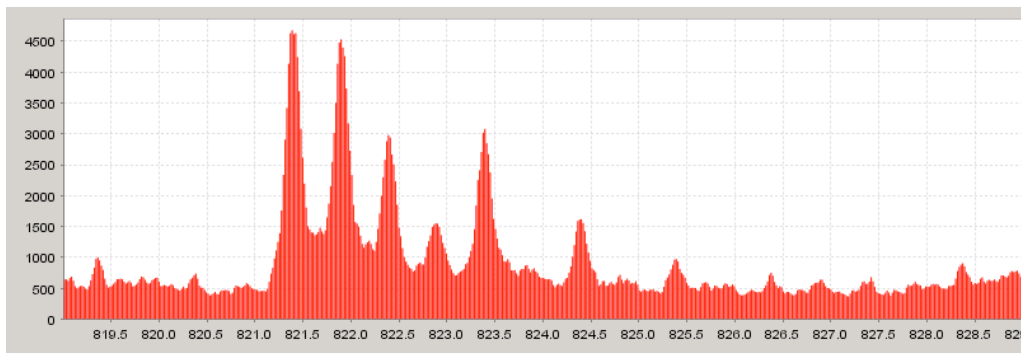
Two groups of samples; identify peptides/proteins that differentiate the groups.



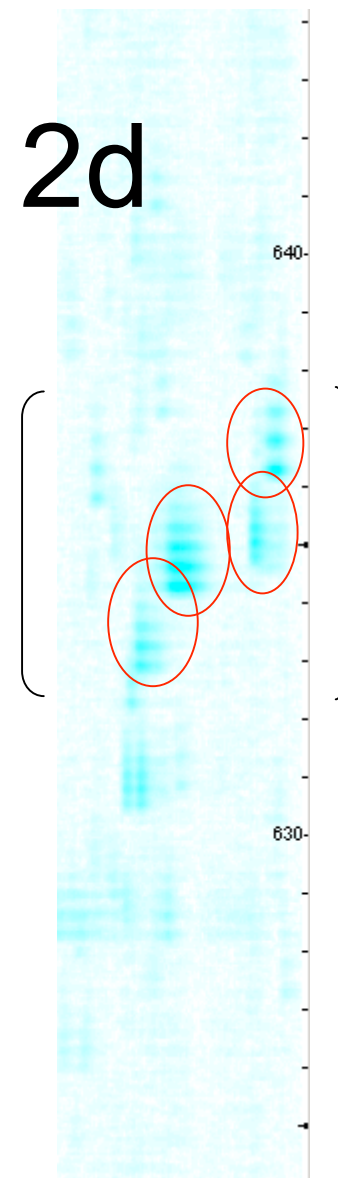
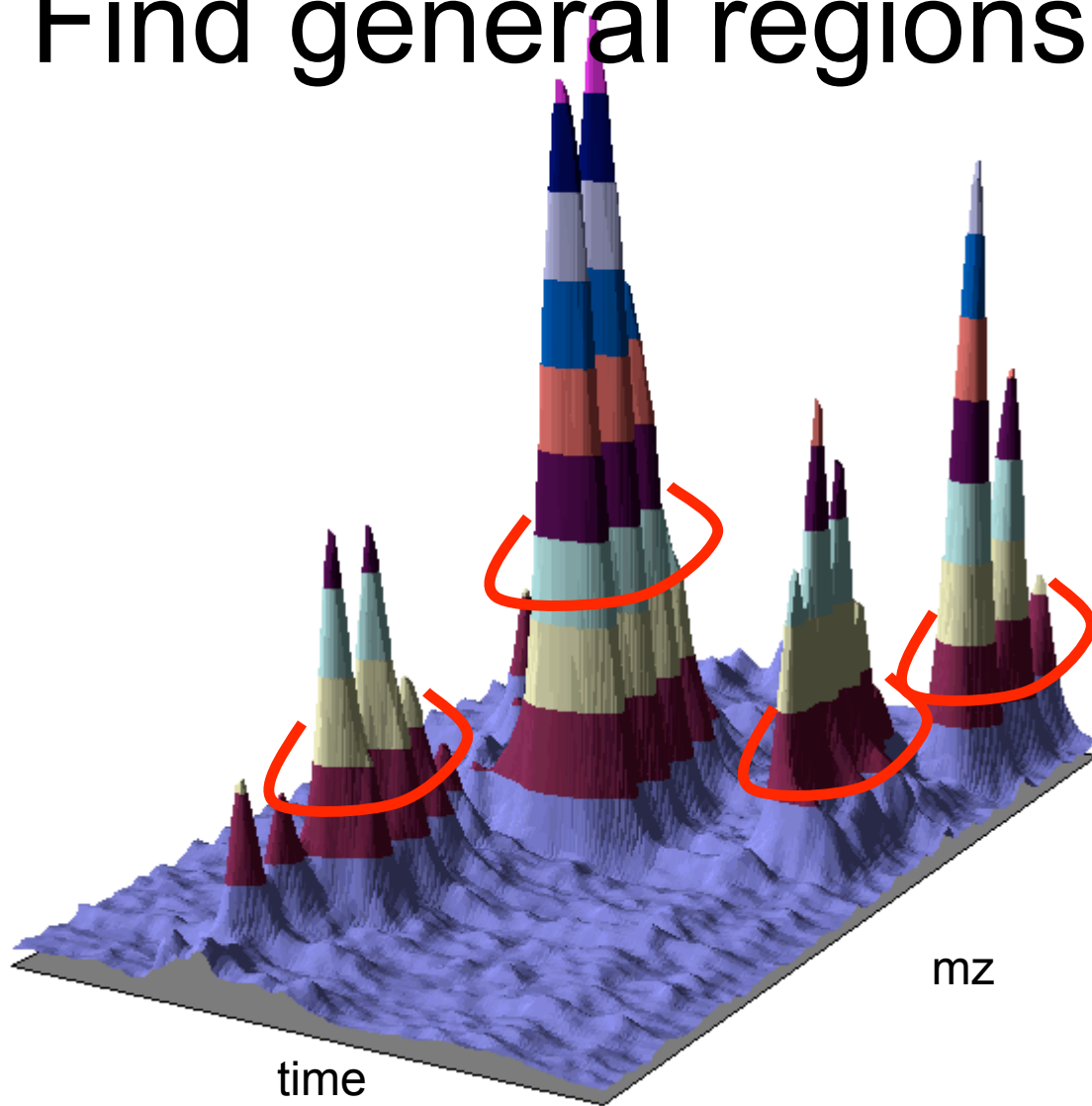


Complex sample with quarks eye view

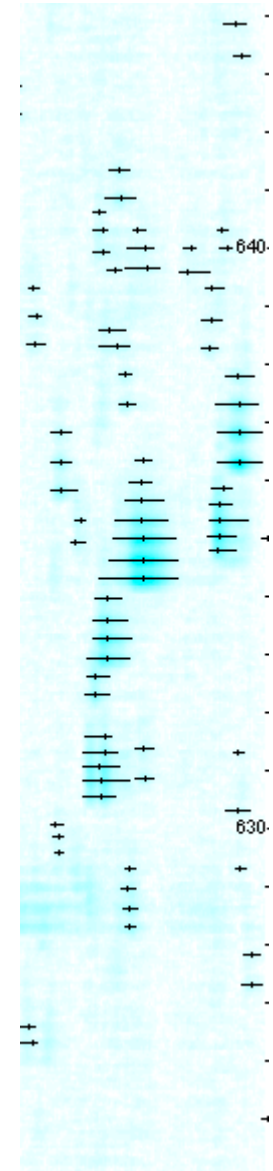
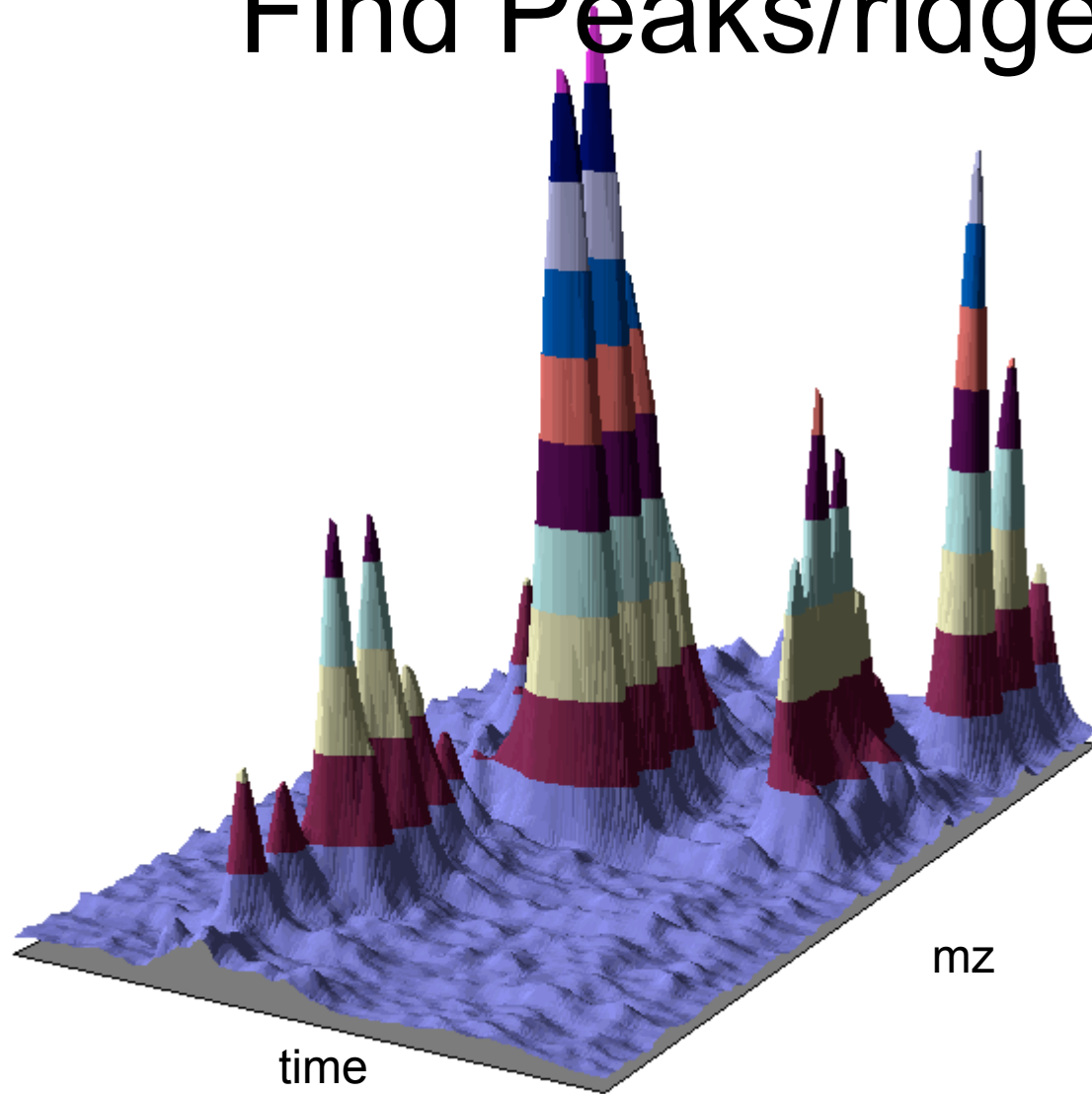




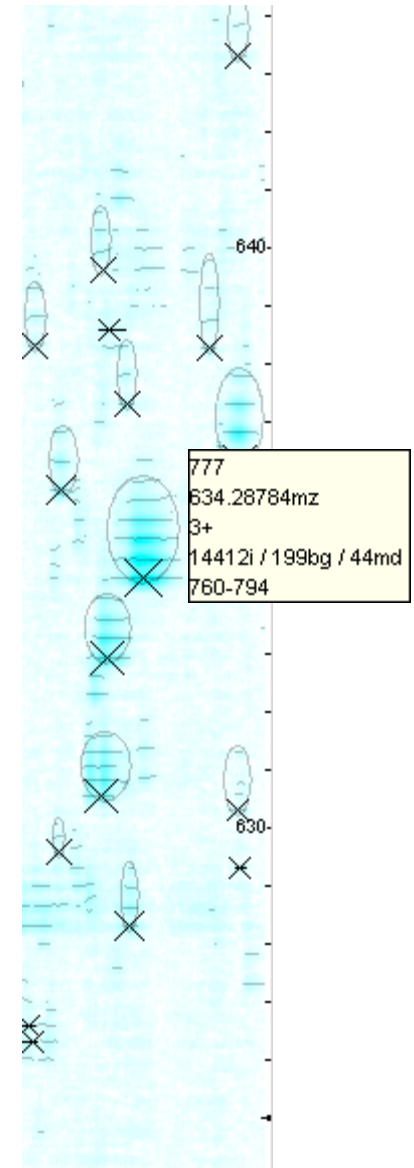
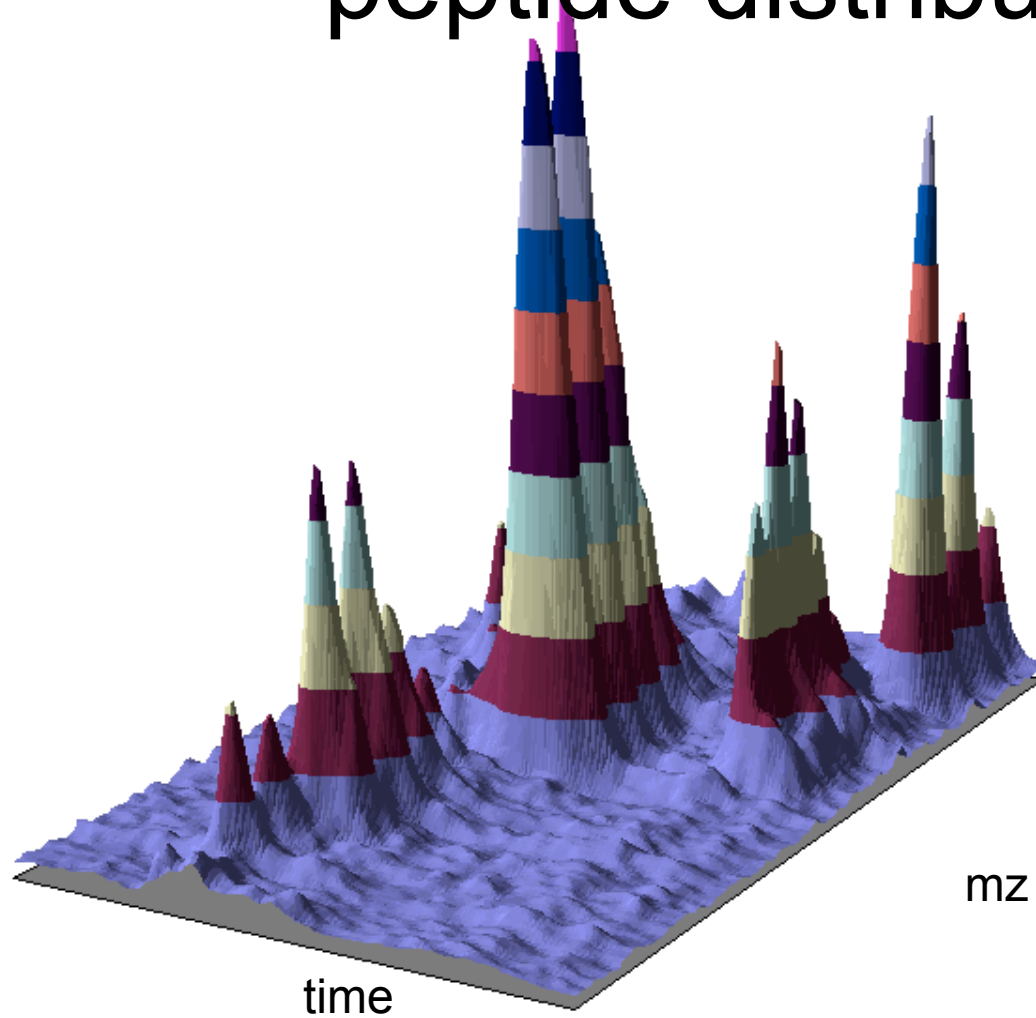
Find general regions in 2d



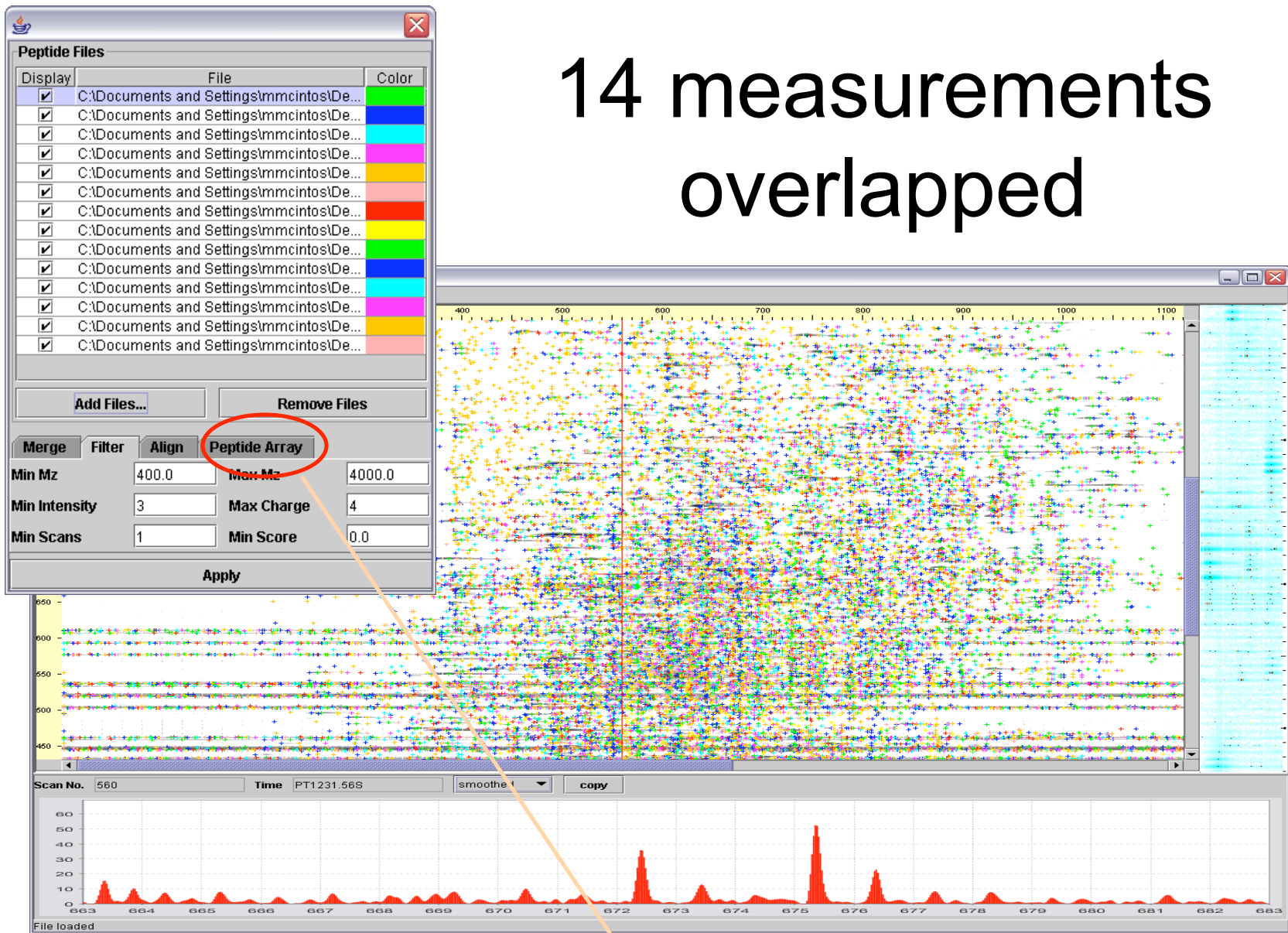
Find Peaks/ridges



Features are ridges whose peaks align in time and match a peptide distribution

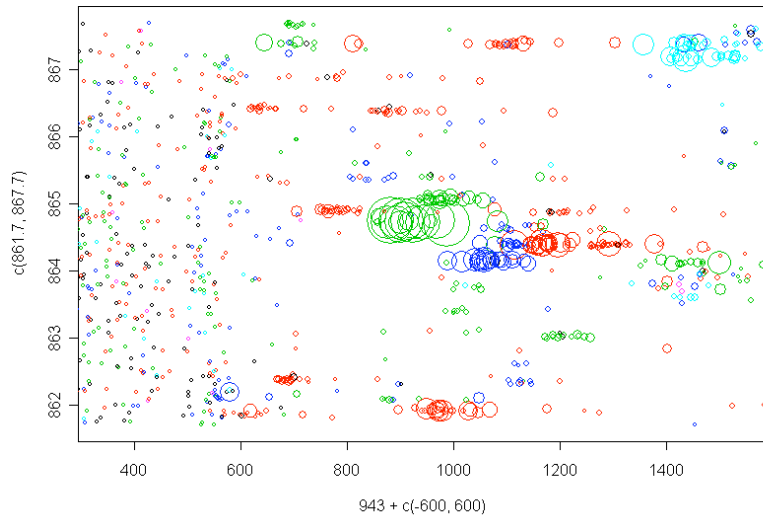


14 measurements overlapped

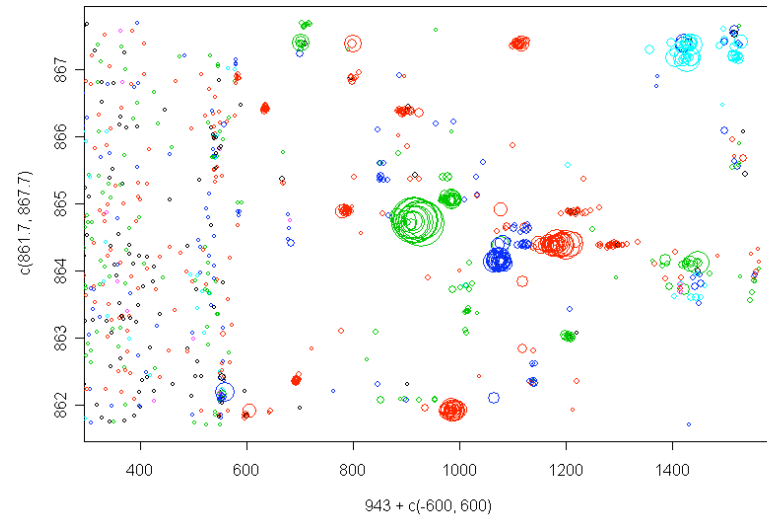


Peptide array button

Time warping/mapping

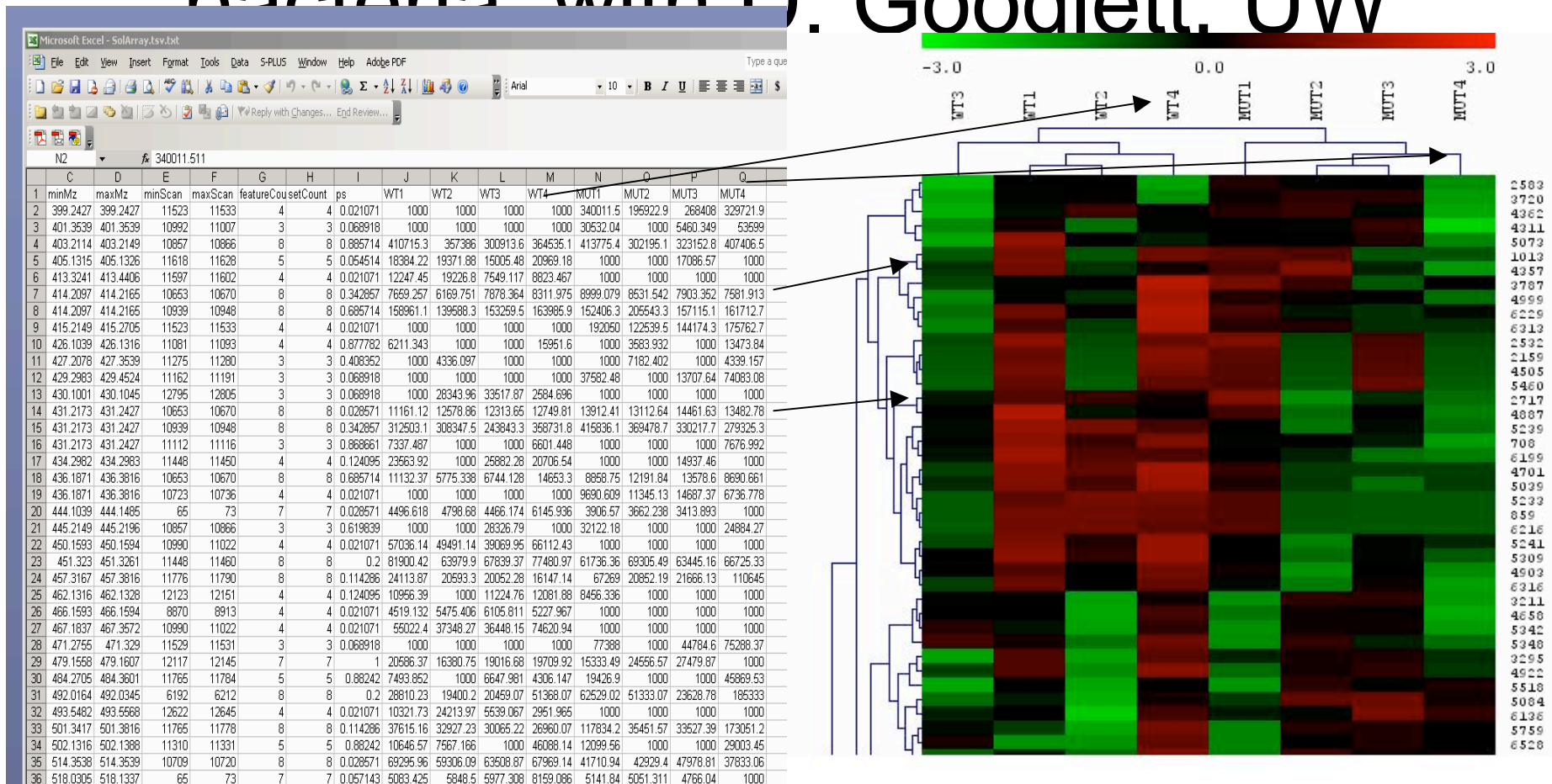


Before mapping



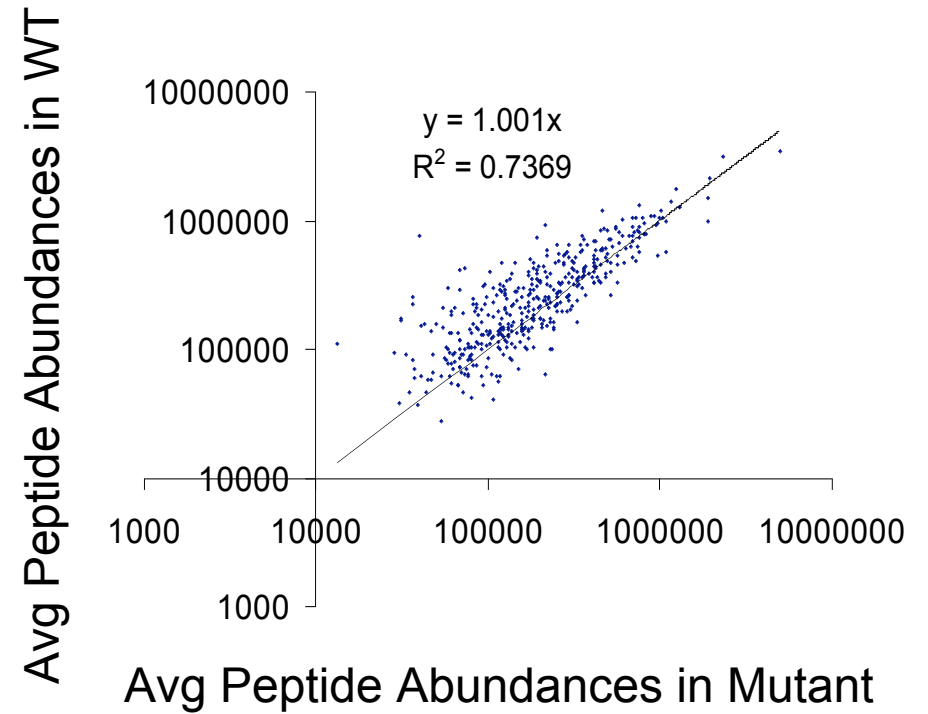
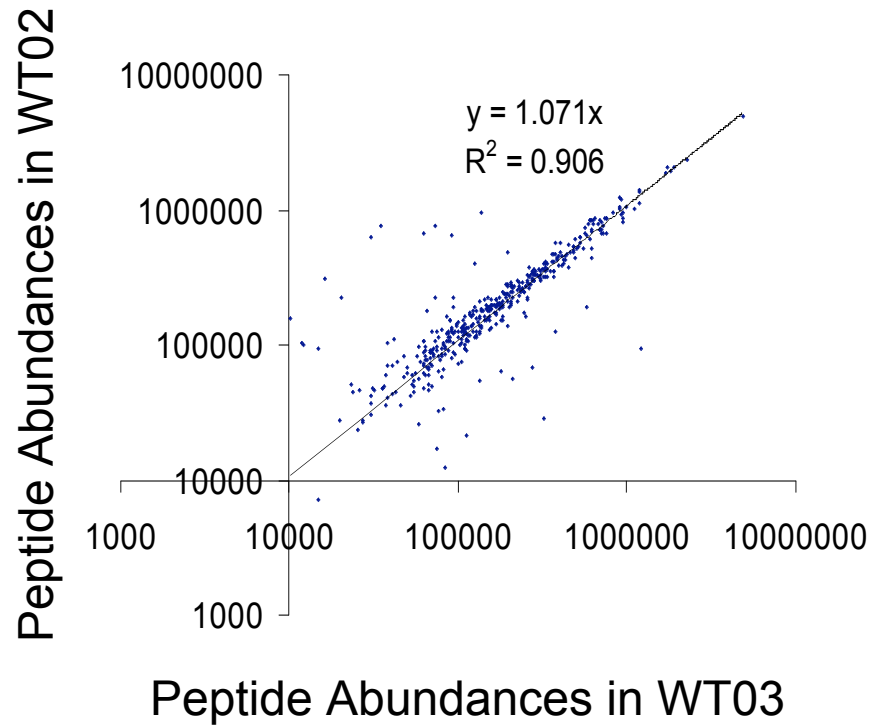
After mapping

Example: Wildtype/mutant bacteria: with D. Goodlett, UW

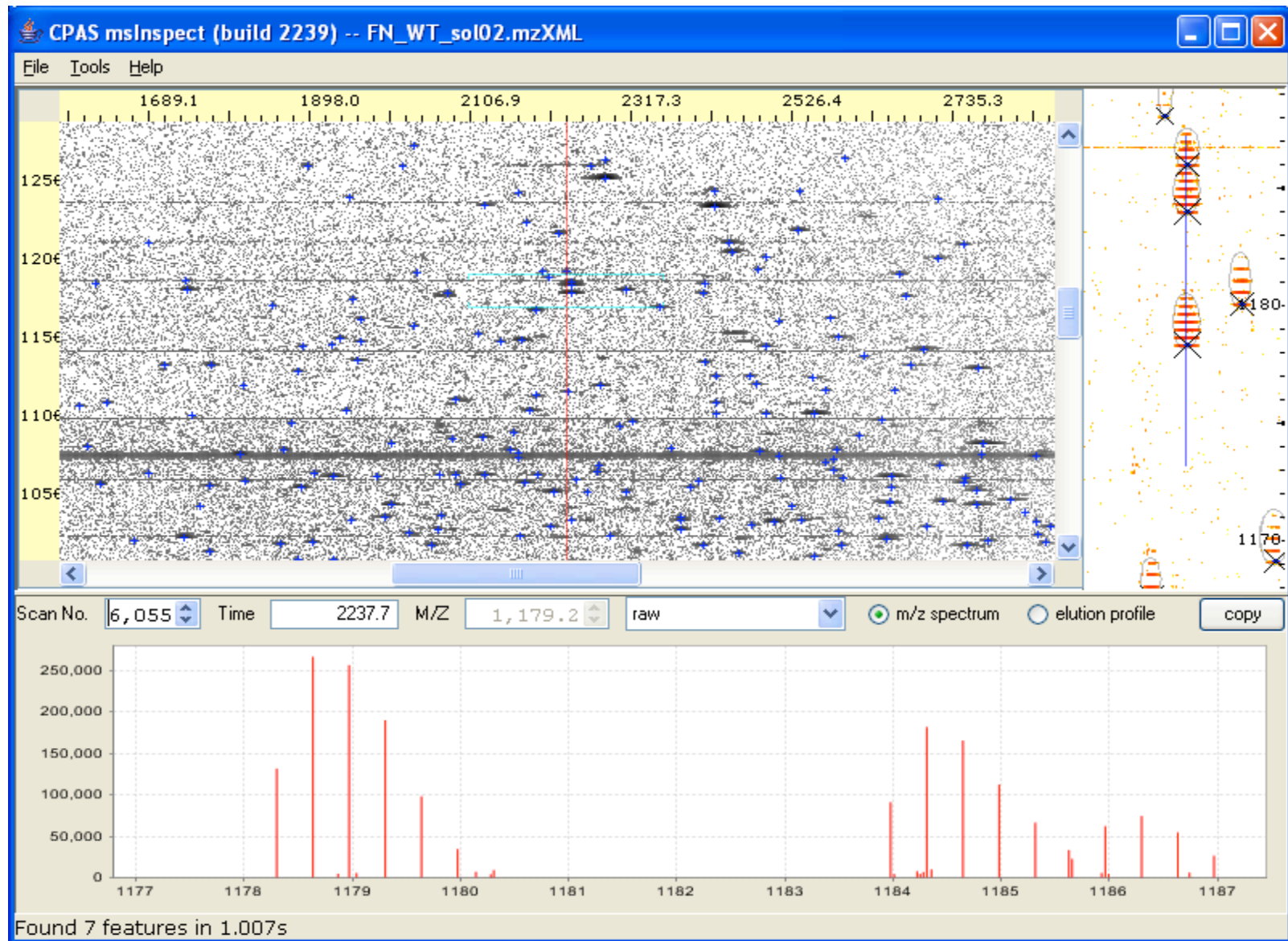


Membrane bound proteins clustering results. Data Source: D. Goodlett

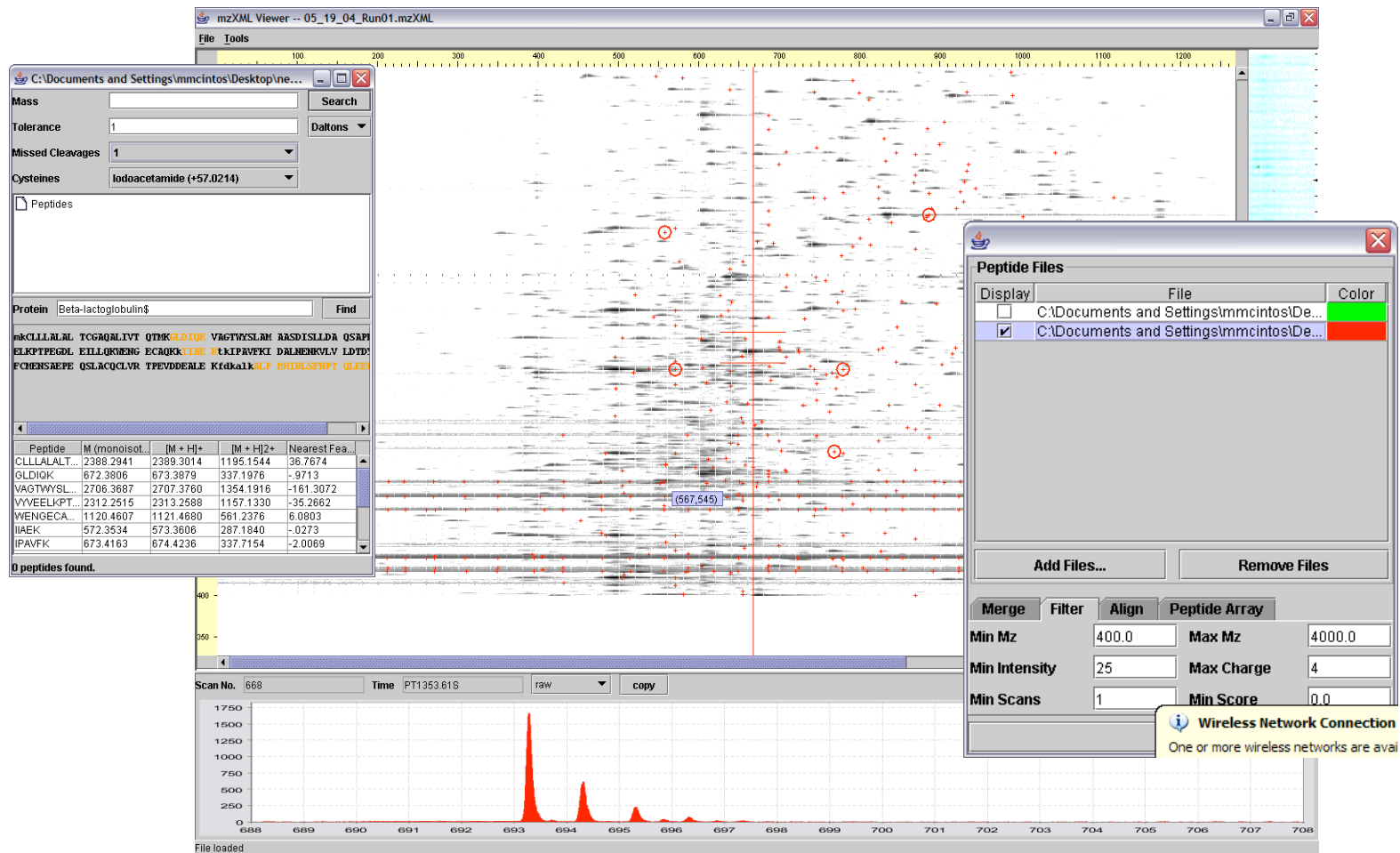
Scatter Plot of Peptide Abundances



msInspect for De-isotoping & De-convolution of LC-MS Data



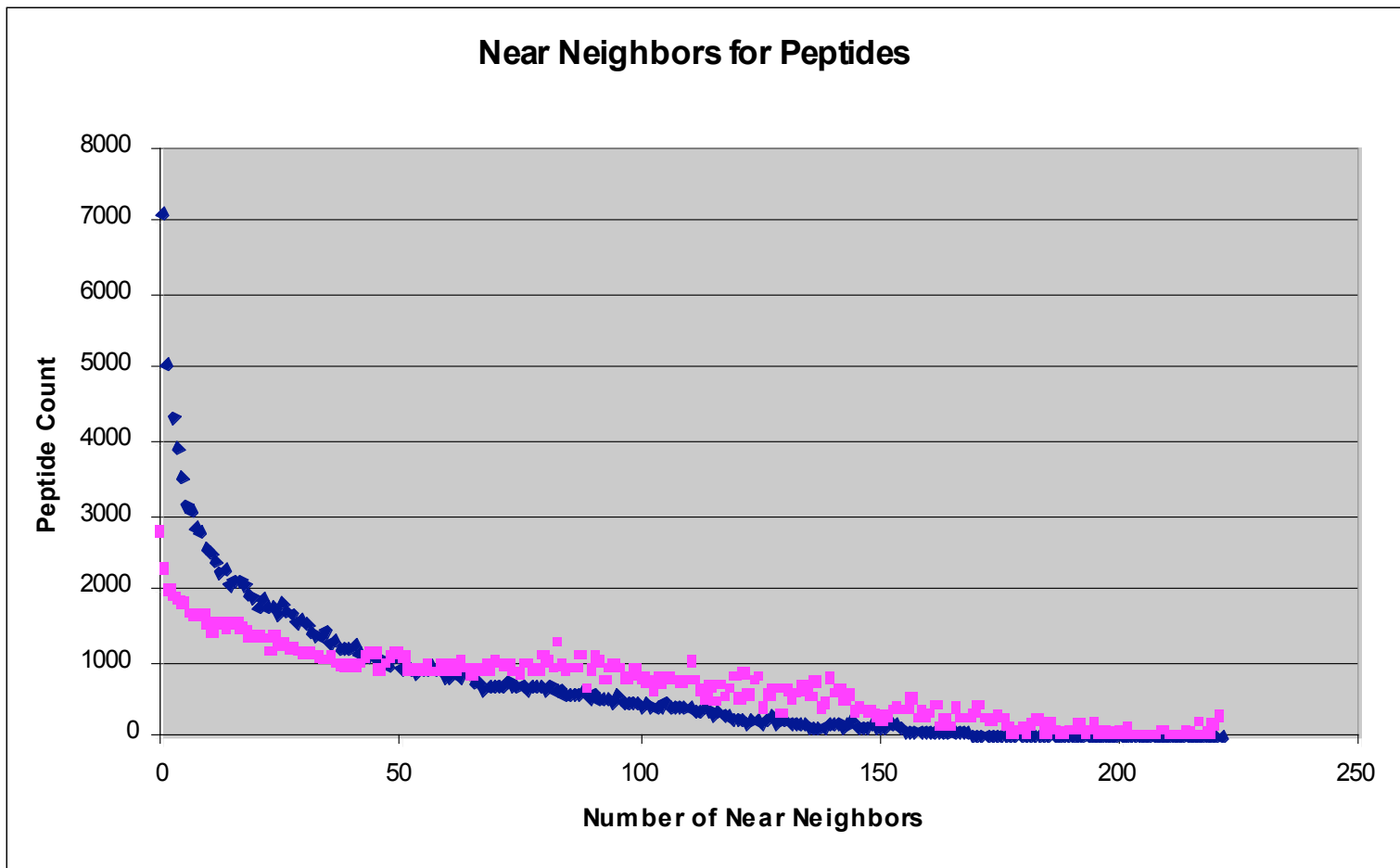
Other features: Protein Fingerprinting (AMT-like)



Also: command-line access to all functions

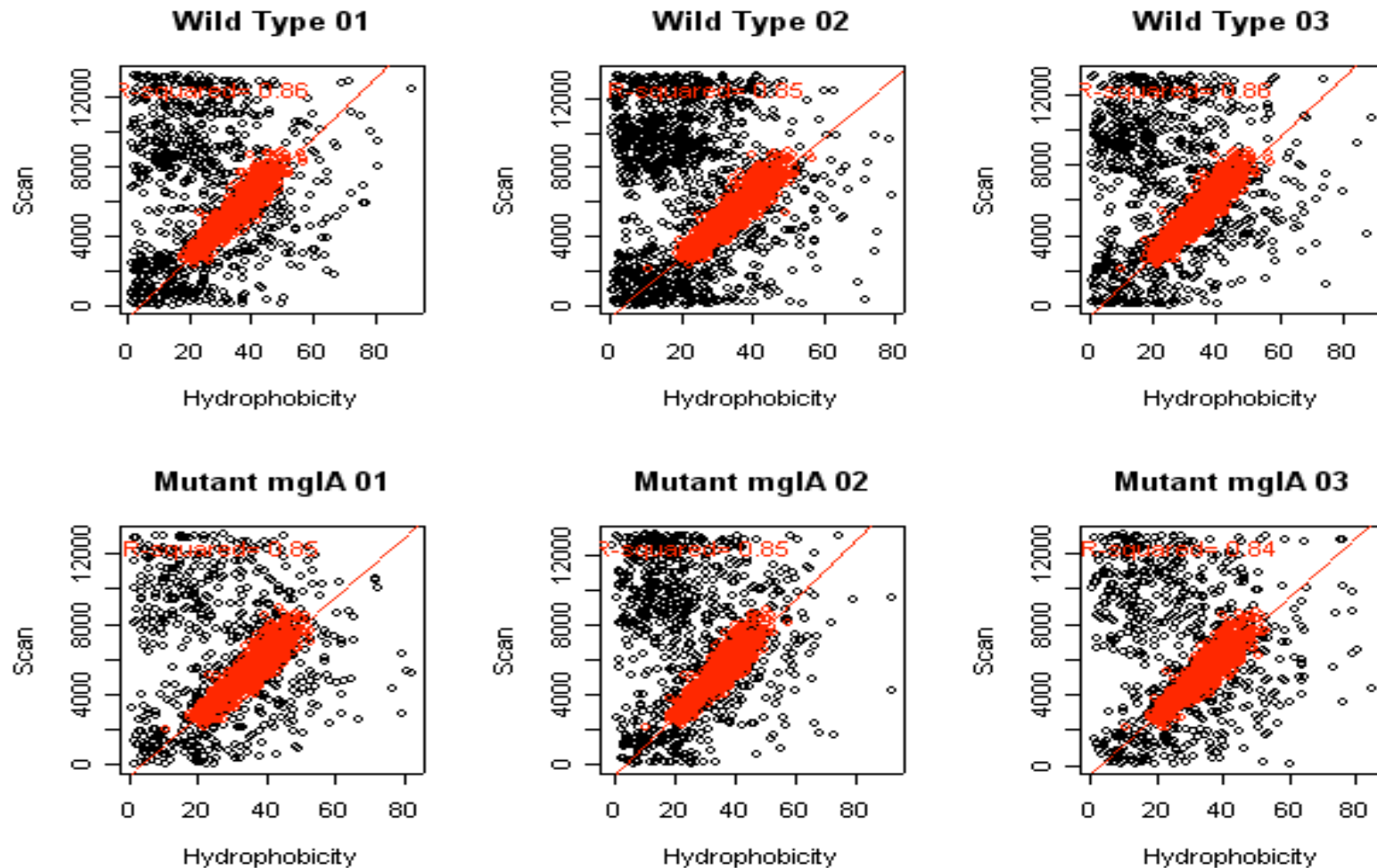
Future developments, omissions

- Develop:
 - Flexible use of standard annotations and ontologies.
 - Inclusion of quantitation (both MS and MS/MS).
 - Integration of MS/MS development.
 - Pluggable into toolkits (PRIME).
- Omissions (not really):
 - We are not a general LIMS system or data repository: general lims system is not our mission.
- Other extensions:
 - Transcript array, flow cytometry, other high throughput informatics systems



Number of unique peptide: 152,000

Scatter Plot of LC Scan Number of Peptides and Their Predicted Hydrophobicity

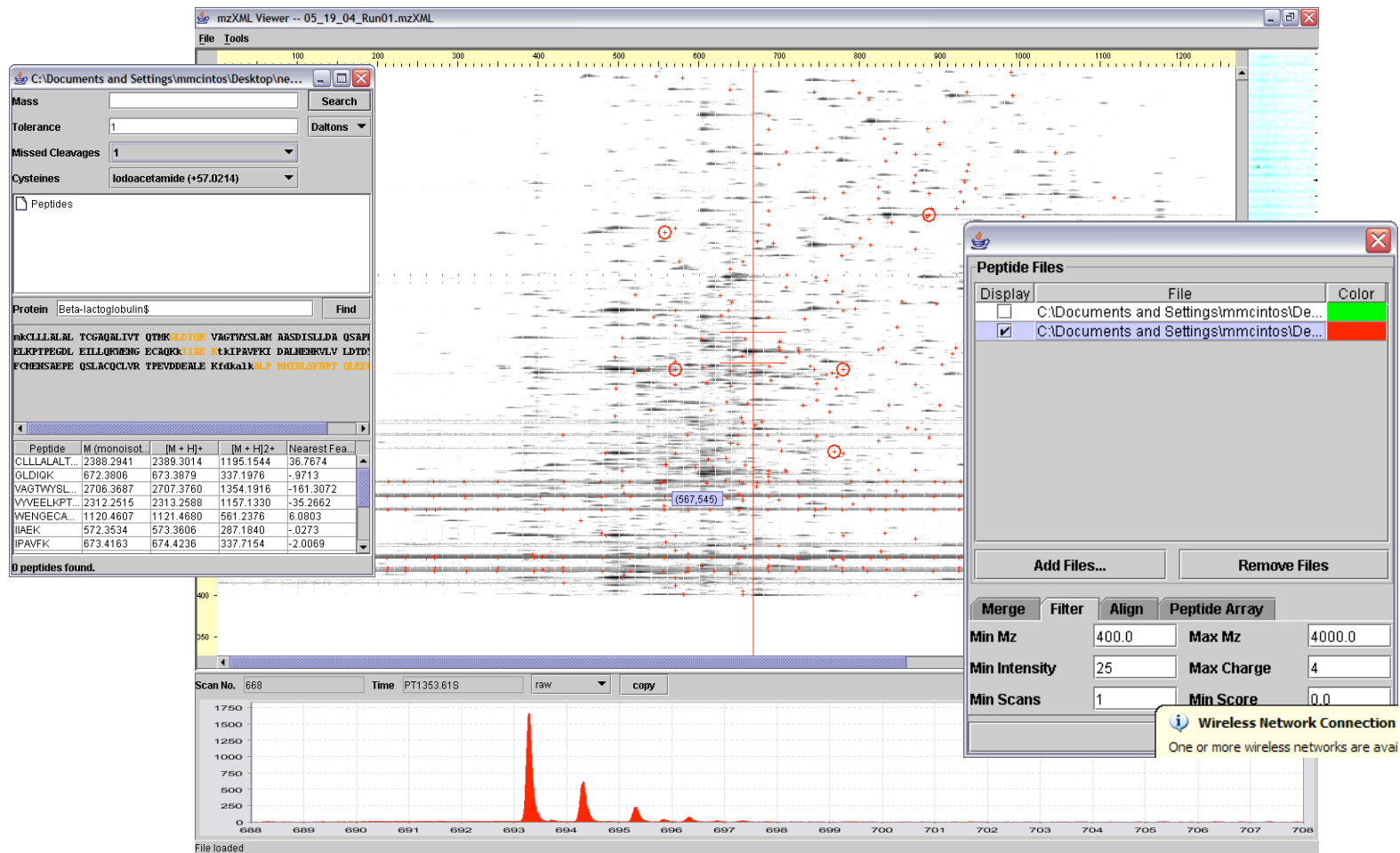


- Peptide with Pprophet score ≥ 0.99 , peptide number: 1148 ± 31
- peptide with $0.01 \leq P_{\text{prophet}} < 0.99$, peptide number: 1163 ± 157

*Amount of sample loaded onto LC column: 0.05 ug protein; ** LC: Michrom Bioresource

** LC: Michrom Bioresource PG4 nano HPLC system, LC column, 10 cm, 75 id, flow rate: 150-200 nl/min

Other features: Protein Fingerprinting (AMT-like)



Also: command-line access to all functions

Peptide Abundance Determined by msInspect

scan	time	mz	mass	intensity	charge	kl	peaks	start	end	totalIntensity
154	100.256	659.0834	658.0761	7183.608	1	0.425834	2	120	212	210031.5
67	68.5333	593.1668	592.1594	6540.017	1	0.163552	3	32	235	287823.2
73	70.6071	667.1667	666.1594	5037.392	1	0.222513	3	32	159	180346
67	68.5333	519.1389	518.1317	4467.056	1	0.288747	3	32	171	176820.6
148	98.1354	646.0555	645.0483	4197.537	1	0.379578	2	114	206	112336.5
67	68.5333	445.1111	444.1039	3951.4	1	0.215707	3	32	171	136702.3
73	70.6071	613.9687	1838.978	3334.867	3	0.586656	3	32	126	82650.3
67	68.5333	741.1945	740.1872	3239.066	1	0.52242	2	32	109	80949.08
79	72.7207	537.1703	1608.395	3221.449	3	0.476398	3	32	148	102143.3

CrossTab of Peptide Abundances of WT and MT Replicates

poson	peptide	WT01_totalInt	wt02_totalInt	wt03_totalInt	MT01_totalInt	MT02_totalInt	MT03_totalInt
PSN053337.2	K.VPIILLTAR.D	322976.28	401819.28	386688.1	195903.19	207294.11	189517.47
PSN068297.2	K.VLGSAVNPVLR.E	112213.2	123895.69	155476.66	117655.445	112587.086	128400.61
PSN052494.2	K.DAVEGTPFTVK.E	625551.25	802901.94	850028.4	907869.2	1177234.8	933233.8
PSN081995.2	K.VGDETLMMR.E	276533.25	350451.47	365765.78	220662.05	235295.53	237956.19
PSN062705.2	K.YYDVDNIIR.V	173483.11	206028.3	202683.25	181279.12	184550.95	191678.1
PSN008878.2	R.VAEILNSSDIK.E	612478.94	716321.7	811235.3	570521.5	658416.6	617378.9
PSN029186.2	R.VAAEILEAVEGR.G	275632.16	312599.25	289406.9	508151.16	442852.06	508085.8
PSN049482.2	K.AVLINIFGGIVR.C	405312.25	581600	545744.1	201339.31	194845.53	199554.02
PSN061458.2	K.TIAAEINTALTK.V	244437.27	251067.7	260582.53	491006.8	462756.38	539630.7
PSN015644.2	R.ATGNVIESYTVK.G	111380.94	93088.46	140143.75	75926.44	87363.91	74446.49
PSN014802.2	K.FVDEVVGGVVPK.E	506127.62	521274.5	492680.22	544588.25	507870.38	554112.1
PSN071732.2	R.HQVVNDLPLGR.N	480489.66	531496.1	565245.06	442455.56	484584.28	486951.62

Proteins with Expression Level Changed in MT vs. WT

poson	Pfm dom names	peptid eCoun t	fold_ mea n	std
PSN070648.2	ATP-grasp domain; Carbamoyl-phosphate synthase L chain, ATP binding domain; D-ala D-ala ligase C-terminus; Prokaryotic glutathione synthetase, ATP-grasp domain; Mur ligase family, catalytic domain; Mur ligase family, glutamate ligase domain	7	7.24	3.20
PSN081776.2	RNA polymerase Rpb1, domain 1; RNA polymerase Rpb1, domain 2; RNA polymerase Rpb1, domain 3; RNA polymerase Rpb1, domain 4; RNA polymerase Rpb1, domain 5;	5	5.51	3.15
PSN020578.2	Dephospho-CoA kinase; Shikimate kinase	4	5.06	1.53
PSN066538.2	Anion-transporting ATPase; 4Fe-4S iron sulfur cluster binding proteins, NifH/frxC family; CobQ/CobB/MinD/ParA nucleotide binding domain;	4	4.32	1.11
PSN023663.2	Carbamoyl-phosphate synthase L chain, N-terminal domain; Carbamoyl-phosphate synthase L chain, ATP binding domain; D-ala D-ala ligase C-terminus; Biotin carboxylase C-terminal domain	3	3.10	0.72
PSN055143.2	short chain dehydrogenase;	3	2.57	1.80
PSN070648.2	ATP-grasp domain; Carbamoyl-phosphate synthase L chain, ATP binding domain; D-ala D-ala ligase C-terminus; Prokaryotic glutathione synthetase, ATP-grasp domain; Mur ligase family, catalytic domain; Mur ligase family, glutamate ligase domain	3	2.49	0.41
PSN006542.2	Ribosomal protein S4/S9 N-terminal domain; S4 domain	2	2.27	0.18
PSN028215.2	Aconitase family (aconitate hydratase); Aconitase C-terminal domain	4	1.97	0.24
PSN021333.2	Histidine kinase-, DNA gyrase B-, and HSP90-like ATPase; Hsp90 protein	3	0.52	0.09
PSN071155.2	Dehydratase family;	2	0.49	0.28
PSN046096.2	Clp amino terminal domain; Clp amino terminal domain; Sigma-54 interaction domain; NACHT domain; ATPase family associated with various cellular activities (AAA); Torsin; ATPase family associated with various cellular activities (AAA);	2	0.48	0.11
PSN083679.2	Phosphoglycerate kinase;	2	0.42	0.09
PSN083678.2	Dihydrodipicolinate reductase, N-terminus; Glyceraldehyde 3-phosphate dehydrogenase, NAD binding domain; Glyceraldehyde 3-phosphate dehydrogenase, C-terminal domain;	5	0.42	0.09
PSN085065.2	Aminotransferase class I and II;	4	0.25	0.13
PSN070070.2	Peptidase family M13; Peptidase family M13	5	0.23	0.04
PSN062623.2	Cytosol aminopeptidase family, catalytic domain;	5	0.19	0.07

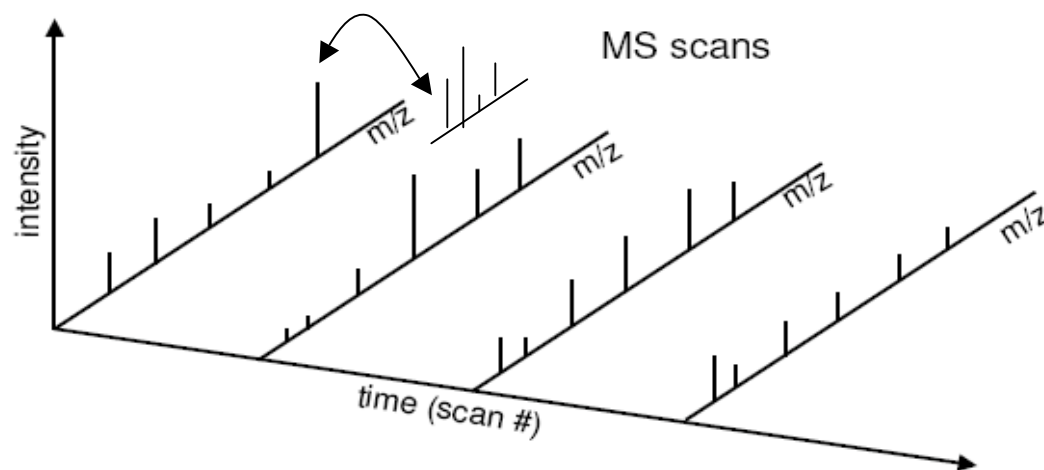
CPAS

Tandem Mass Spectrometry

LC/ESI-TOF (MS/MS)

- ESI (electro spray ionization):
 - Proteins liquid charged gas-phase ions.
- TOF (time of flight) mass spec:
 - Estimates **mass to charge ratio** of ions by their time of flight.
- LC: Liquid chromatography:
 - A procedure for on-line separation of complex mixture.

Mass vs. Intensity vs. Time



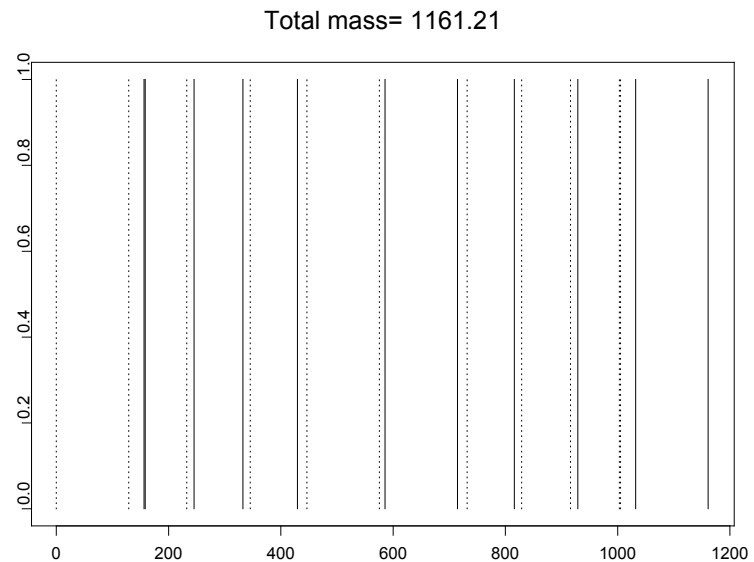
Mass of protein is determined by its alphabet

Name	Letter code	3-letter code	Monoisotopic Mass	Average Mass
Alanine	A	Ala	71.03711	71.0788
Arginine	R	Arg	156.10111	156.1875
Asparagine	N	Asn	114.04293	114.1038
Asp. Acid	D	Asp	115.02694	115.0886
Cysteine	C	Cys	103.00919	103.1388
Glut. Acid	E	Glu	129.04259	129.1155
Glutamine	Q	Gln	128.05858	128.1307
Glycine	G	Gly	57.02146	57.0519
Histidine	H	His	137.05891	137.1411
Isoleucine	I	Ile	113.08406	113.1594
Leucine	L	Leu	113.08406	113.1594
Lysine	K	Lys	128.09496	128.1741
Methionine	M	Met	131.04049	131.1926
Phenylalanine	F	Phe	147.06841	147.1766
Proline	P	Pro	97.05276	97.1167
Serine	S	Ser	87.03203	87.0782
Threonine	T	Thr	101.04768	101.1051
Tryptophan	W	Trp	186.07931	186.2132
Tyrosine	Y	Tyr	163.06333	163.1760
Valine	V	Val	99.06841	99.1326

- Proteins are made up of 20 “letters” (two are the same)
- Think of a protein mass as the sum total of 19 letters.

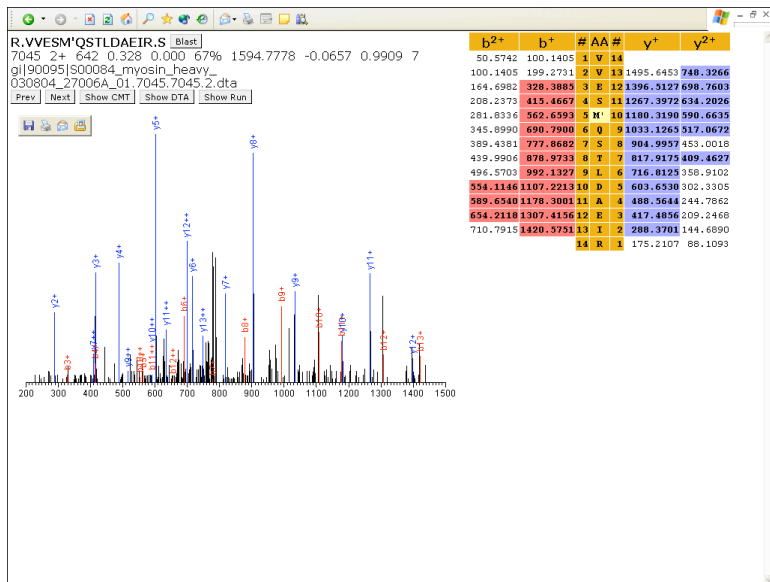
Peptide Example

- b ioconducor
- bi oconductor
- Bio conductor
-
- Biocondu ctor
-
- Bioconducto r
- Bioconductor



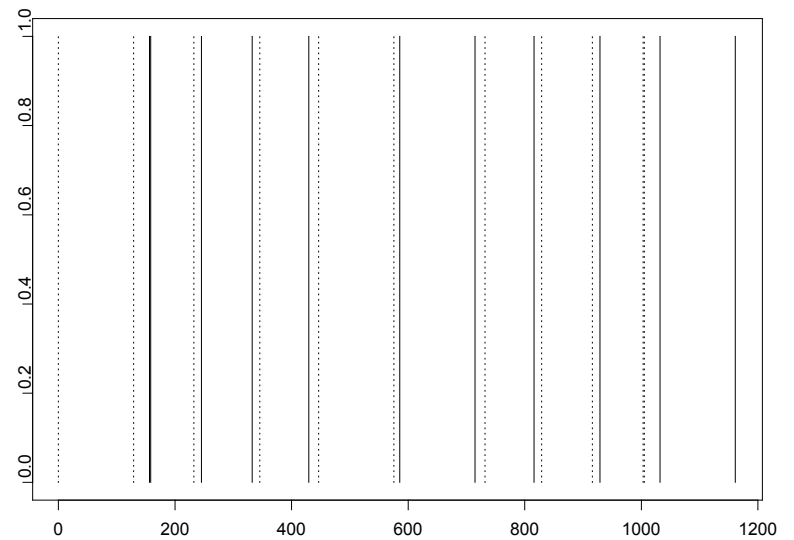
Dan-leibler

Peptide Example

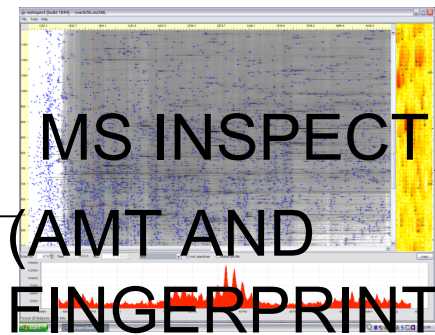
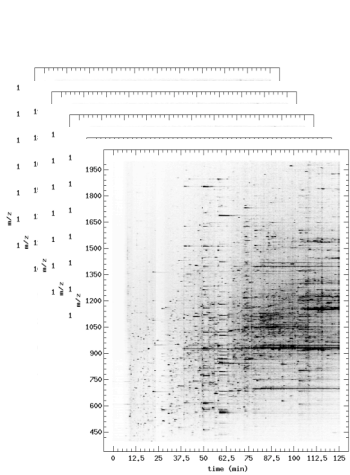


X
t
a
n
d
e
m

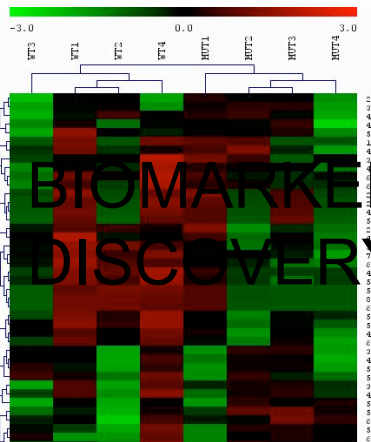
Total mass= 1161.21



LC MS/MS → mzDATA
mxXML

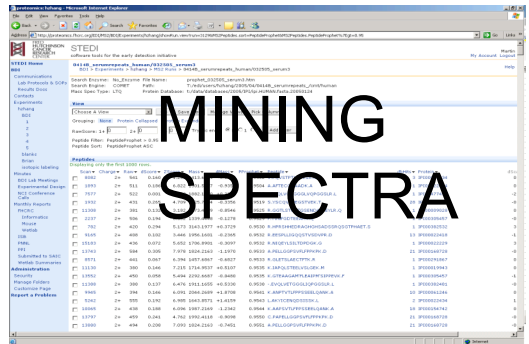


Search Engine
pepXML



Peptide array table. A table with columns for protein names (e.g., WT3, WT4, WT5, WT6, WT7, WT8, WT9) and rows for various peptides. The table contains numerical values representing peptide abundance for each protein.

QUANTITATE
Peptide array



Decisions at every step annotated and can be changed at any time.

Sample of External Tools Currently Integrated.

- ISB Suite of Tools (Aebersold Lab)
 - Protein Prophet
 - Peptide Prophet.
- X!Tandem development;
 - Ron Beavis and others.
- Elution time prediction algorithm
 - Krokhin, R. Craig,(2002).
- CPL Tools (msInspect) Proteomics Consortia and CPAS
- NCI Proteomics Initiative: 6 Laboratories.
- Breast Cancer Biomarker Initiative: 6 Laboratories.

CPL Leaders

- PI:
 - Martin McIntosh
- Staff Scientists
 - Jimmy Eng
 - Ruihua Fang
- Platform Development.
 - Mark Igra
 - Mat Bellew
 - Brendan Maclean
 - Adam Rauch
- Bioinformatics
 - Matt Fitzgibons
 - Chenwei Lin
- Signal processing/analysis
 - Pei Wang
 - Li Hsu
 - Charles Kooperberg

Application history: Sequest, Bio conductor, EndNote, Visual Basic, Visual C++, Excel, Access, Sequel

Financial supporter for infrastructures

- NCI (proteomics initiative), and caBIG
- LabKey:
 - Private company founded by FHCRC and CPL.
 - Develop user support for open source platform.
- Canary Fund:
 - Not-for profit group
 - dedicated to early detection.
 - Supporting methods for user adoption.