Analyzing mass spectrometry data

UWPR Proteomics Workshop

July 10, 2012

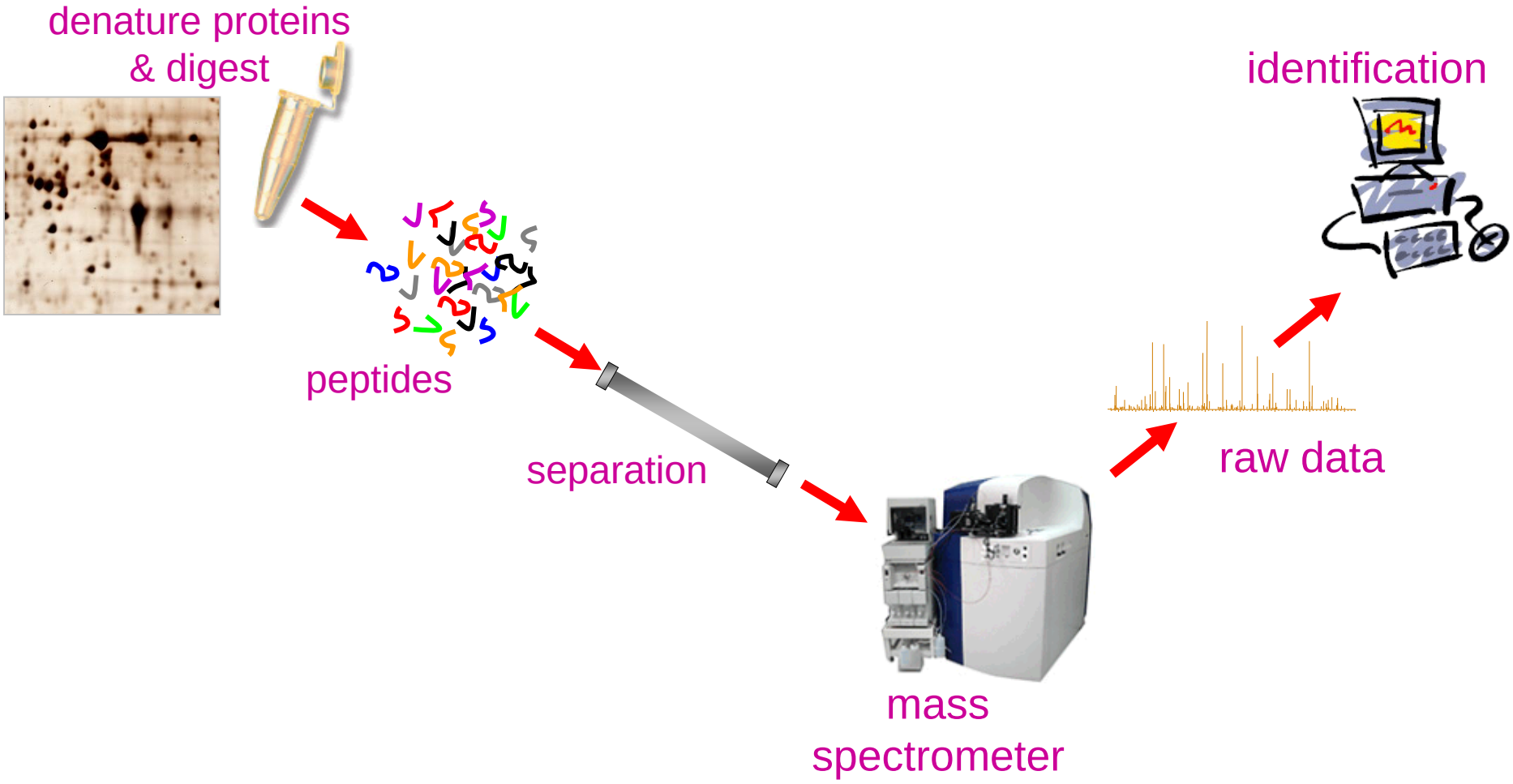
Jimmy Eng

UW Proteomics Resource

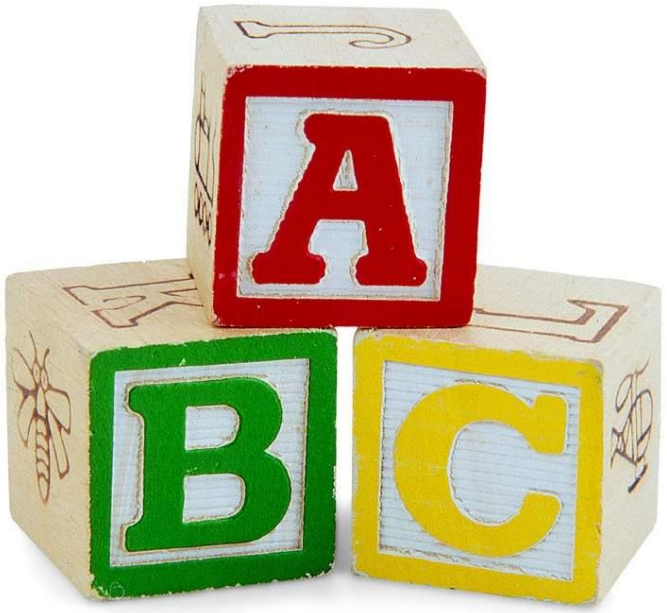
Outline

- background & principles
- software tools
- strengths & limitations
- scoring, expect values
- reasons why searches fail

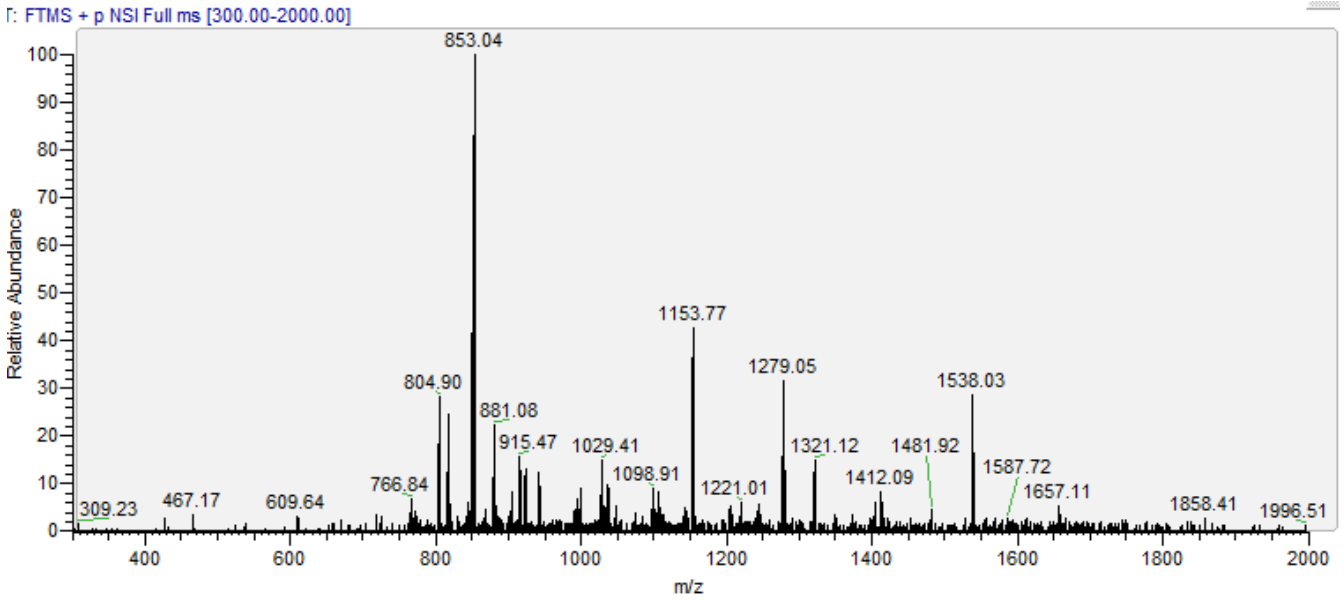
General MS/MS proteomics workflow



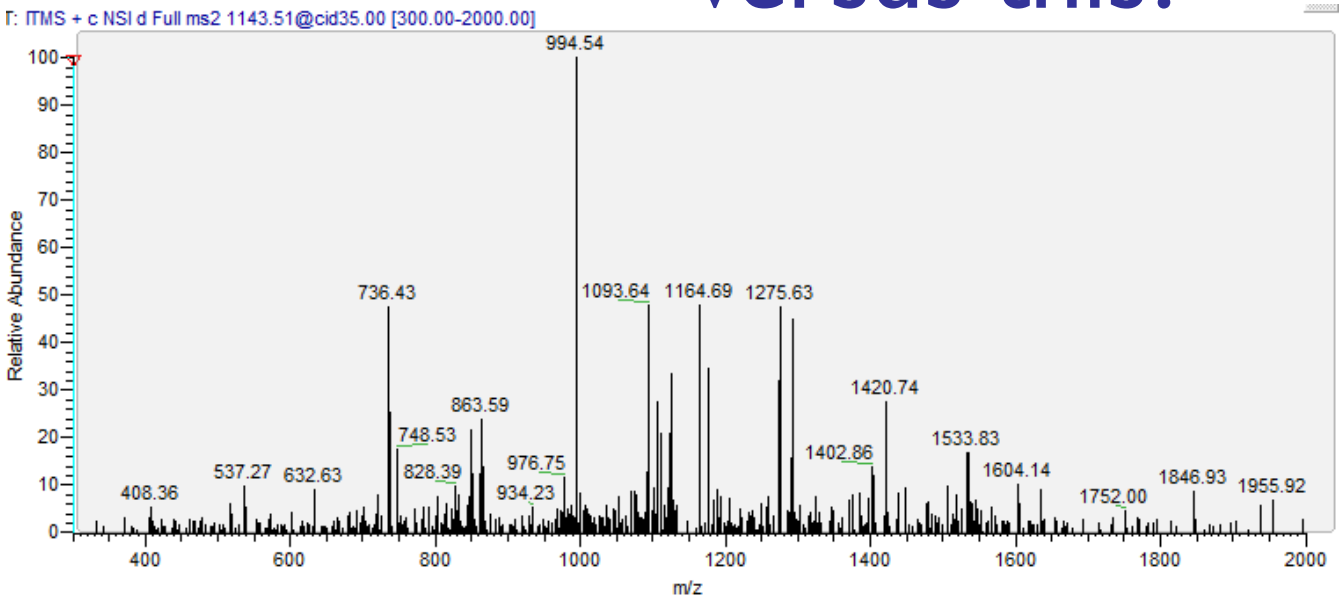
Background and principles



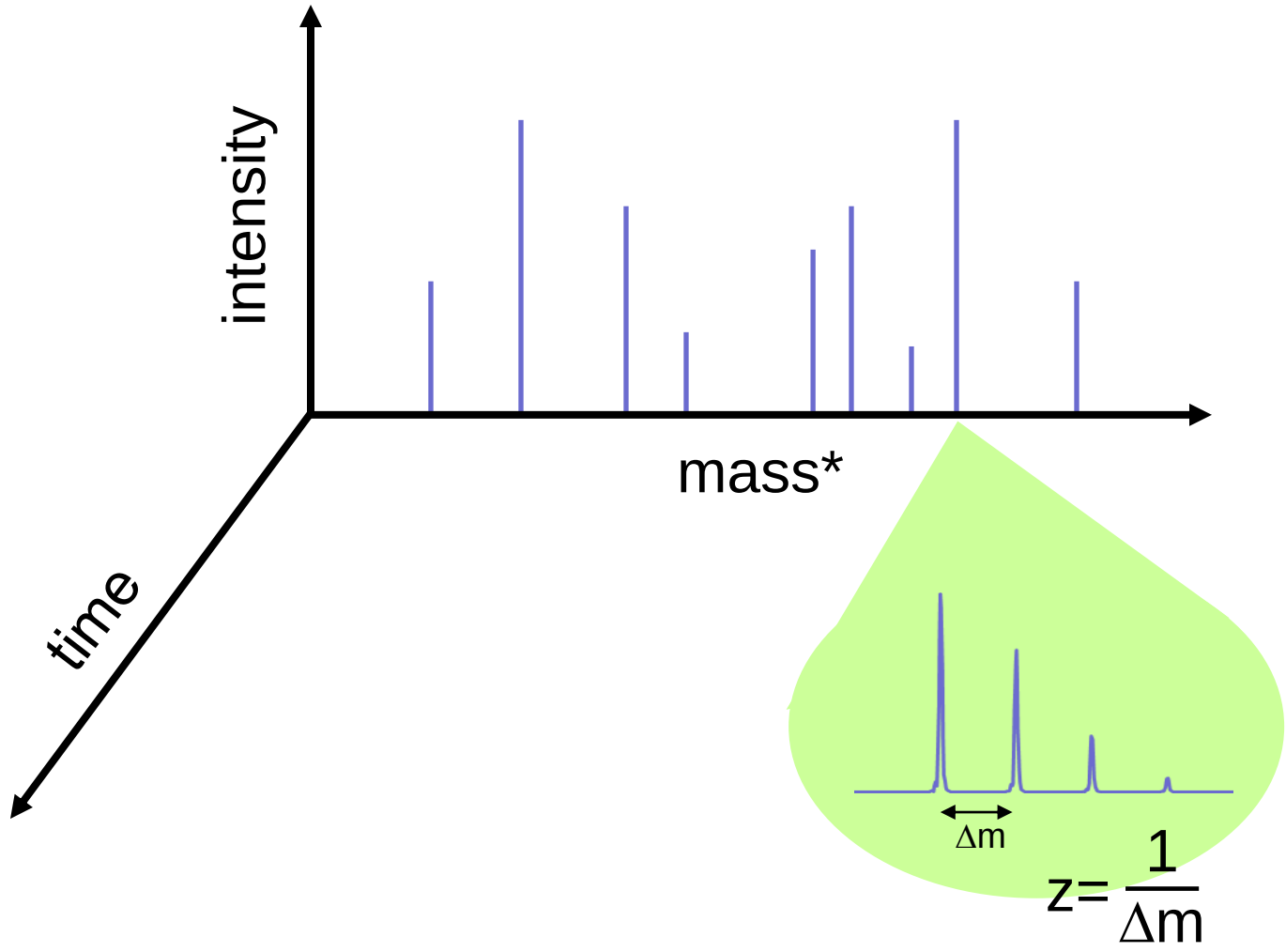
What is this?



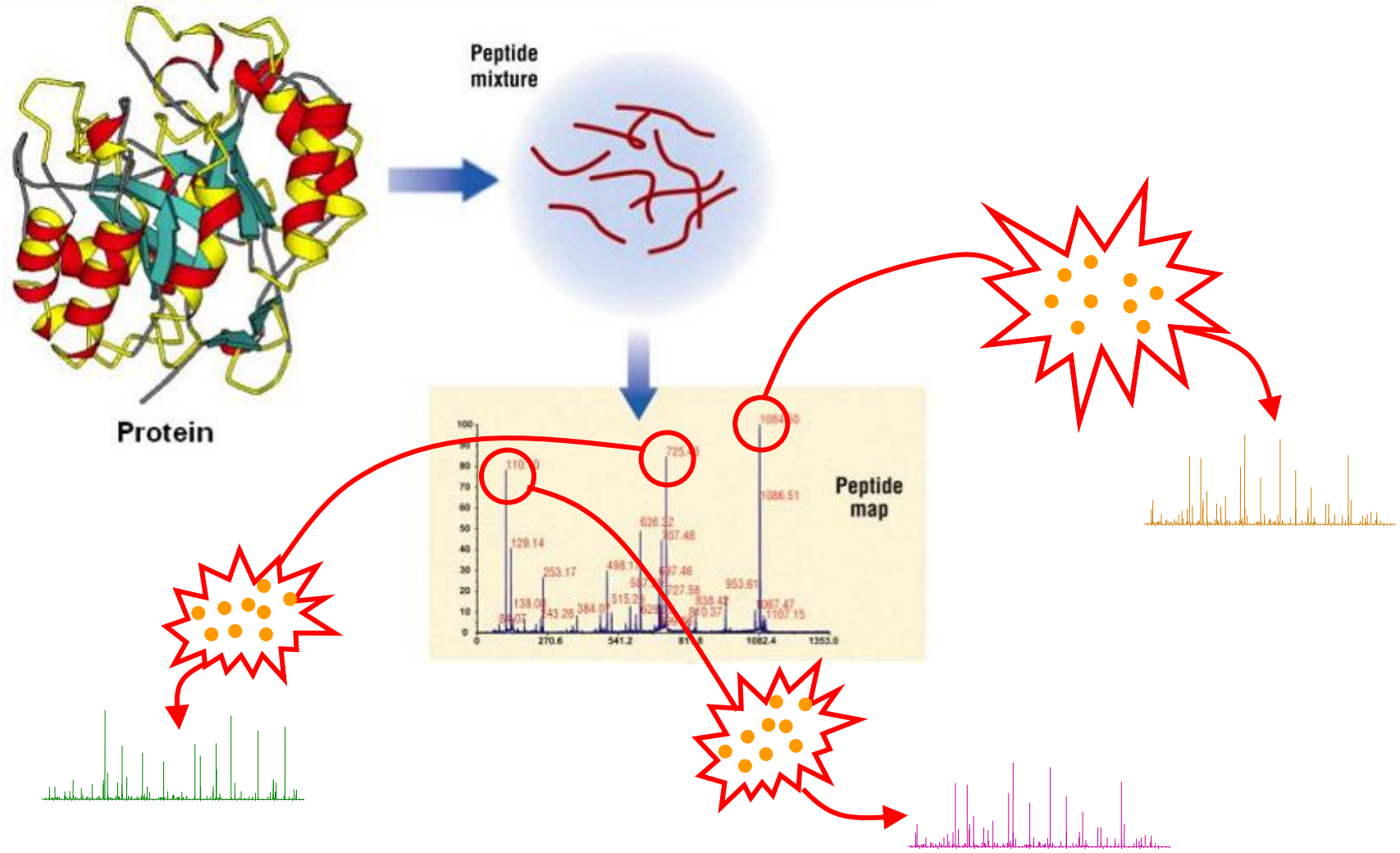
Versus this?



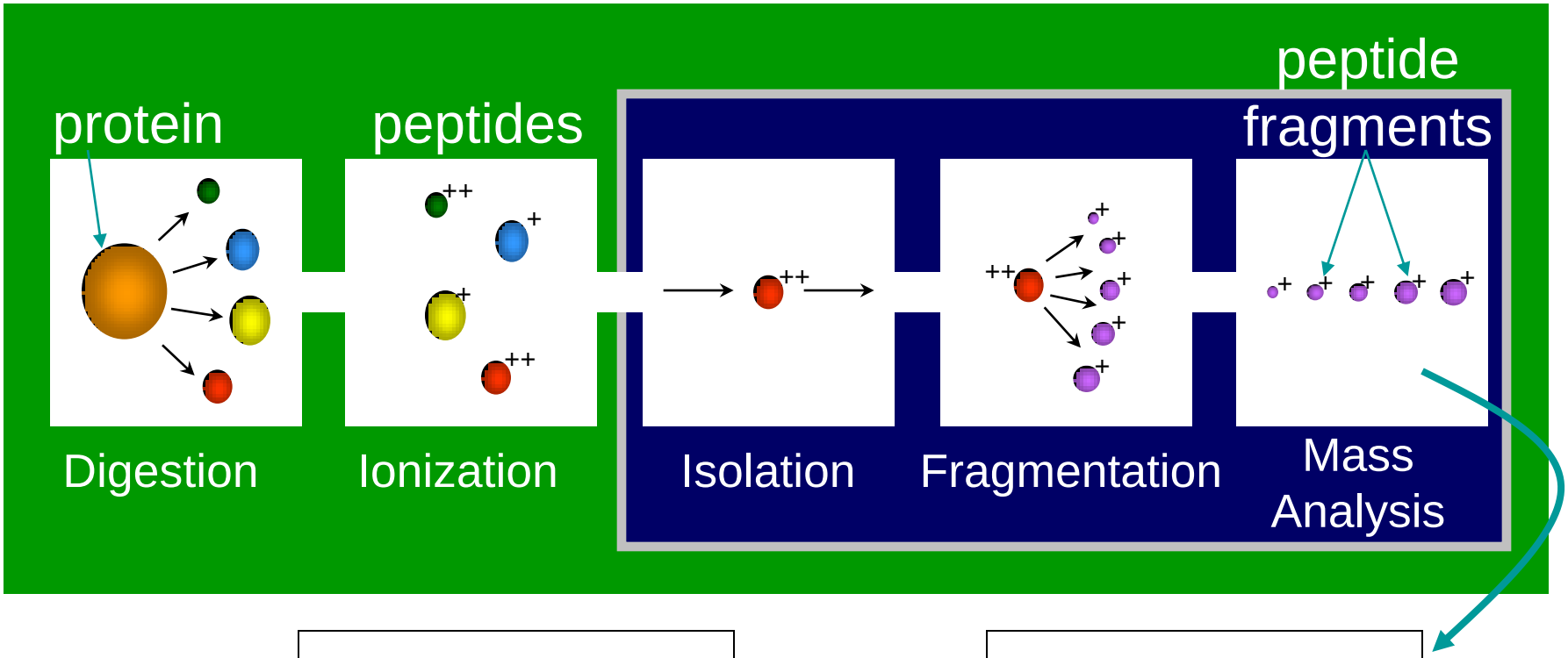
The data

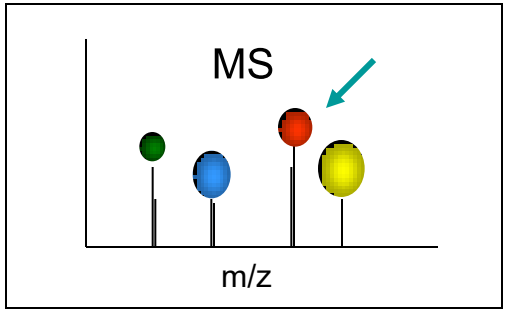


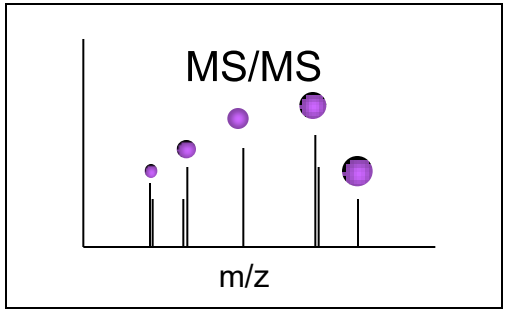
Fragmenting the intact peptides



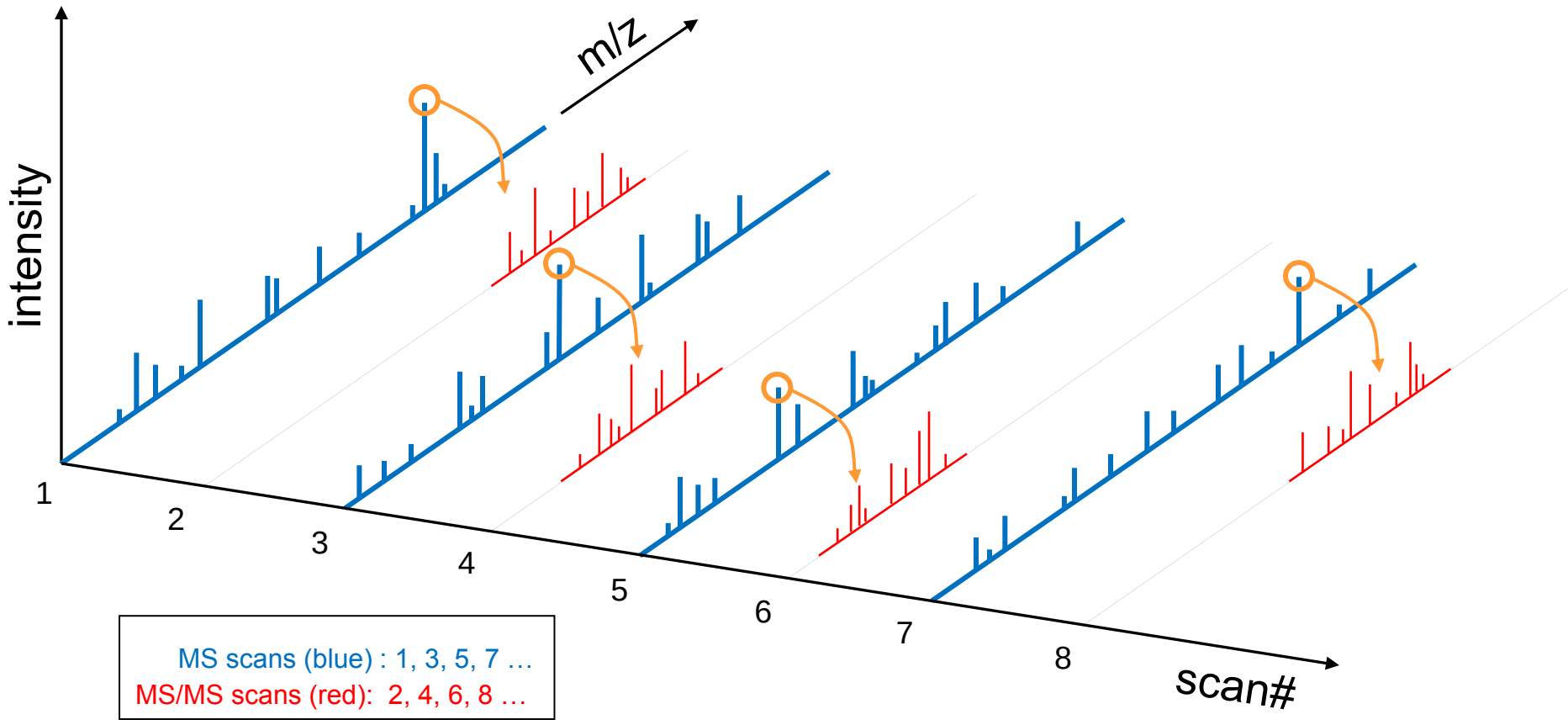
Tandem MS


protein
peptides
peptide fragments
Digestion
Ionization
Isolation
Fragmentation
Mass Analysis

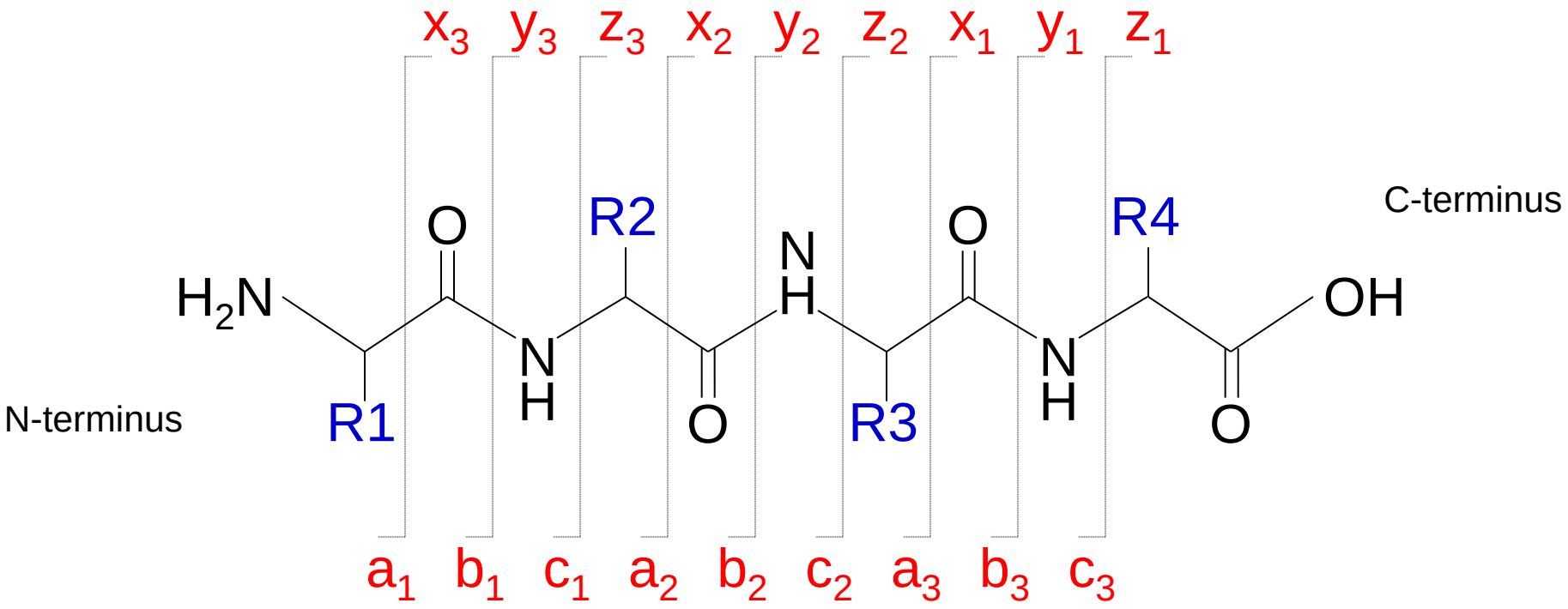

MS
m/z


MS/MS
m/z

MS and MS/MS relationship



Peptide fragmentation



Peptide fragmentation

There are other possible fragment ion types (immonium ions, d-, v- and w-ions) but these are not common in the data we acquire here at UWPR.

For a good overview:
http://www.matrixscience.com/help/fragmentation_help.html

Peptide fragmentation

Peptide fragmentation

A-P-N-D-F-N-L-K

(MH+ 918.5)

b-ions		y-ions
72.0	A ..... P-N-D-F-N-L-K	847.4
169.1	A-P ..... N-D-F-N-L-K	750.4
283.1	A-P-N ..... D-F-N-L-K	636.3
398.2	A-P-N-D ..... F-N-L-K	521.3
545.2	A-P-N-D-F ..... N-L-K	374.2
659.3	A-P-N-D-F-N ..... L-K	260.2
772.4	A-P-N-D-F-N-L ..... K	147.1

b-ions = ΣAA + proton

y-ions = ΣAA + H_2O + proton

monoisotopic masses

UWPR

https://proteomicsresource.washington.edu/masses.php

Apr 24 **ISB TPP course**

TPP software course is being offered by ISB on June 4-8.

Mar 26 **Symposium**

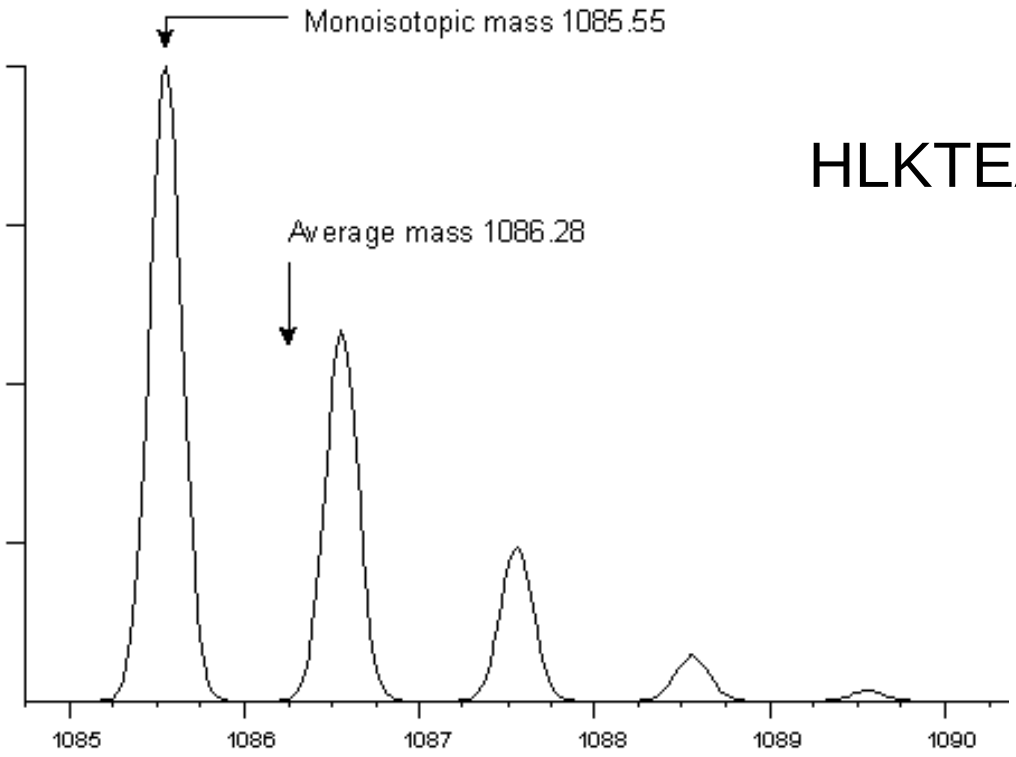
Please join us for a one day Proteomics Symposium at the UW South Lake Union Campus preceeding Vancouver's ASMS.

Click on structure header for larger (viewable!) images.
Click on column headers to sort these tables:

amino acid	code	composition	mono mass	avg mass	structure
glycine	G	C_2H_3NO	57.021463735	57.05132	
alanine	A	C_3H_5NO	71.037113805	71.0779	
serine	S	$C_3H_5NO_2$	87.032028435	87.0773	
proline	P	C_5H_7NO	97.052763875	97.11518	
valine	V	C_5H_9NO	99.068413945	99.13106	
threonine	T	$C_4H_7NO_2$	101.047678505	101.10388	
cysteine	C	C_3H_5NOS	103.009184505	103.1429	
leucine	L	$C_6H_{11}NO$	113.084064015	113.15764	
isoleucine	I	$C_6H_{11}NO$	113.084064015	113.15764	
asparagine	N	$C_4H_6N_2O_2$	114.042927470	114.10264	
aspartic acid	D	$C_4H_5NO_3$	115.026943065	115.0874	
glutamine	Q	$C_5H_8N_2O_2$	128.058577540	128.12922	
lysine	K	$C_6H_{12}N_2O$	128.094963050	128.17228	
glutamic acid	E	$C_5H_7NO_3$	129.042593135	129.11398	
ornithine	O	$C_5H_{12}N_2O_2$	132.089877680	132.16098	
methionine	M	C_5H_9NOS	131.040484645	131.19606	
histidine	H	$C_6H_7N_3O$	137.058911875	137.13928	
phenylalanine	F	C_9H_9NO	147.068413945	147.17386	
arginine	R	$C_6H_{12}N_4O$	156.101111050	156.18568	
tyrosine	Y	$C_9H_9NO_2$	163.063328575	163.17326	
tryptophan	W	$C_{11}H_{10}N_2O$	186.079312980	186.2099	

element	code	mono mass	avg mass
hydrogen	h	1.007825035	1.00794
carbon	c	12.0000000	12.0107
nitrogen	n	14.003074	14.0067
oxygen	o	15.99491463	15.9994
phosphorus	p	30.973762	30.973761
sulphur	s	31.9720707	32.065
proton		1.00727646688	

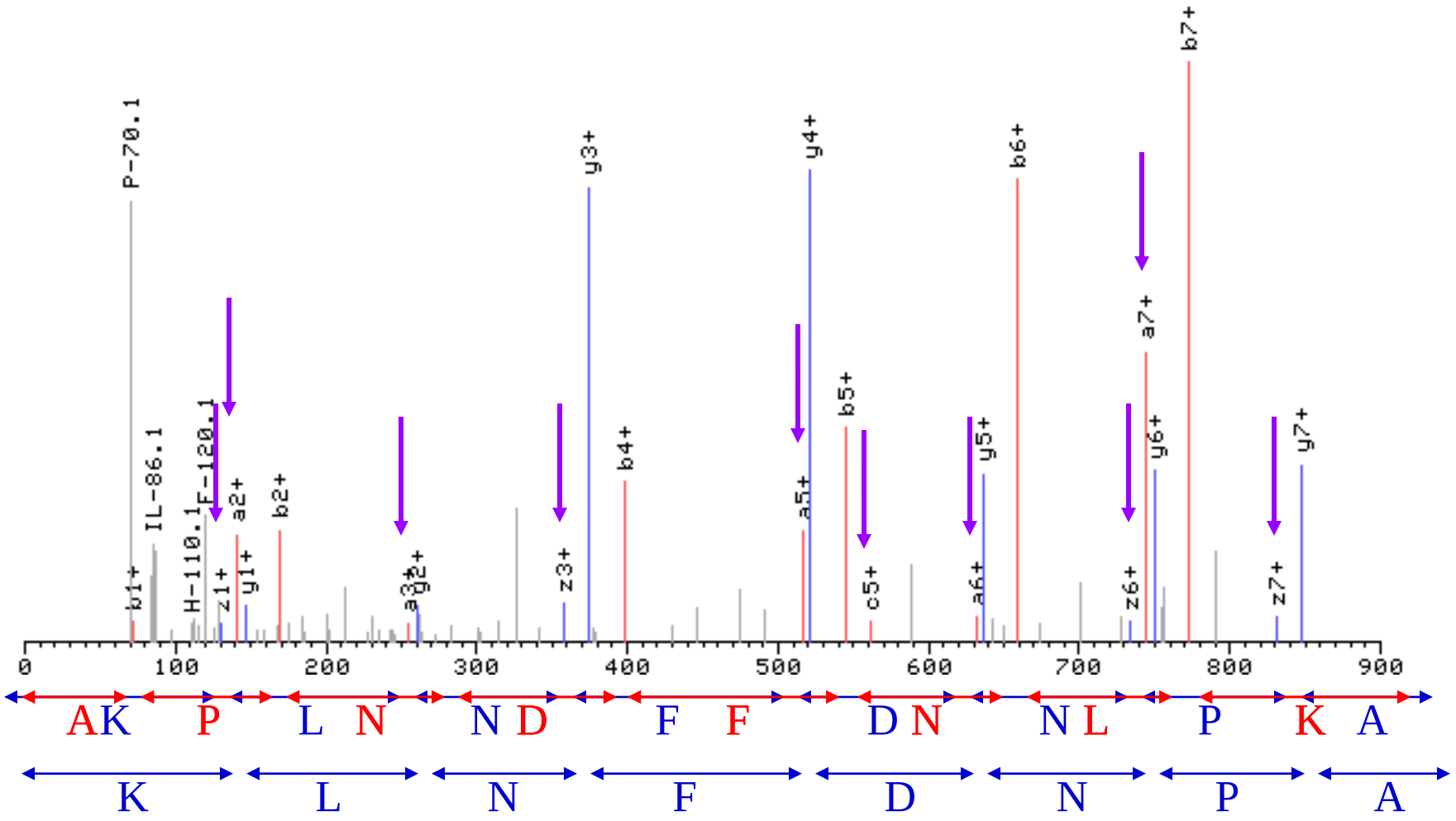
Average vs. monoisotopic mass



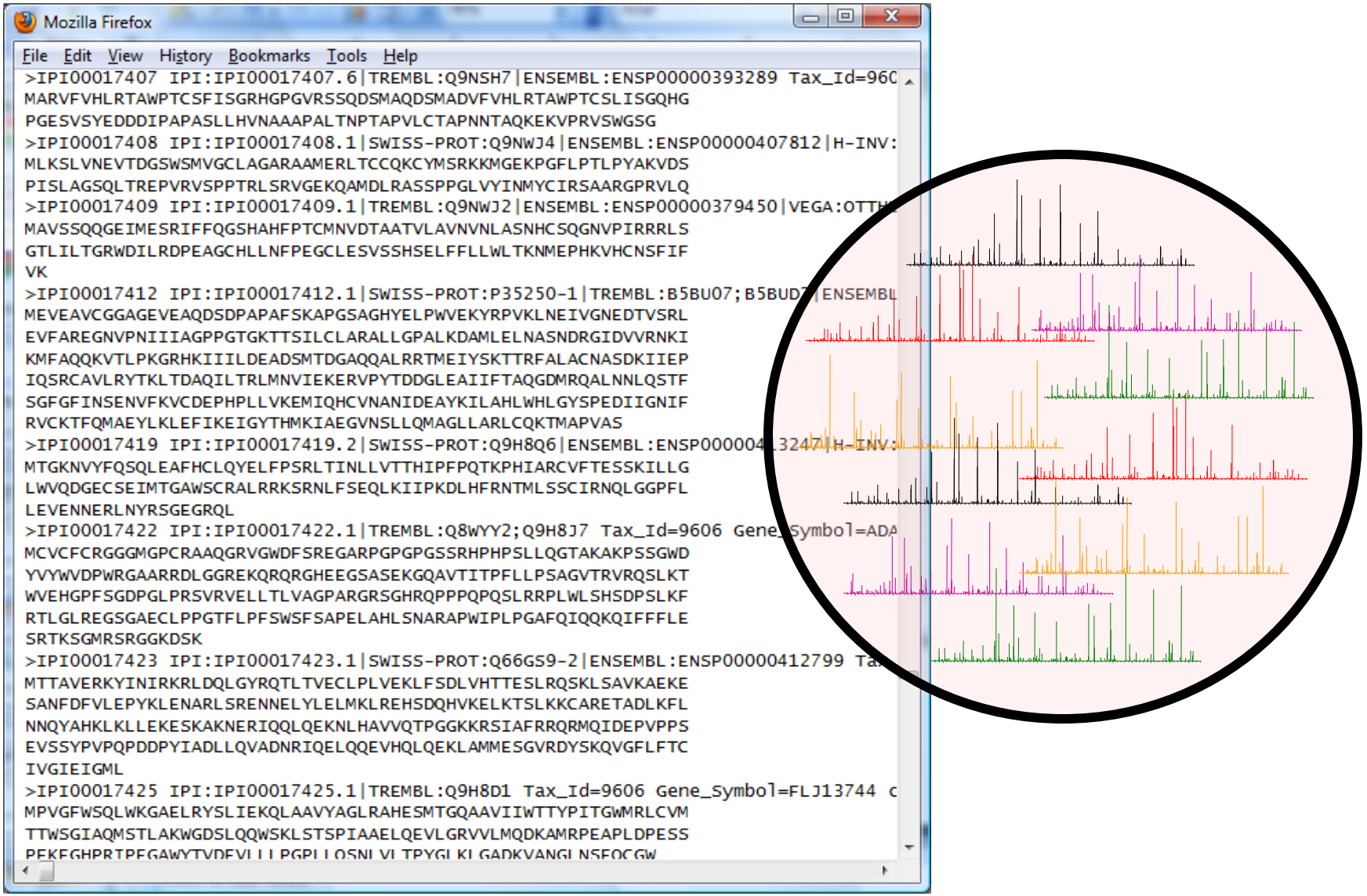
HLKTEAEMK

Higher resolution = narrower peaks

Sequence vs. tandem mass spectrum

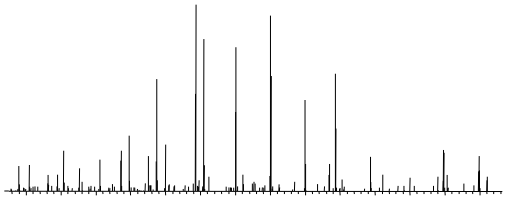


On to MS/MS database searching



Uninterpreted MS/MS database search

raw MS/MS spectra



similarity score

1.00

0.34

0.29

0.28

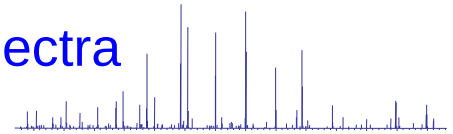
Peptides of same nominal mass

sequence database

```
>SEQUENCE1
CVVRELCPTPEGKDIGES
VDLLKLQWCWENGTLRSL
DCDVVSRDIGSESTEDRA
MEDIK
>SEQUENCE2
DLRSWTVRIDALNHGVKP
HPPNVSVVDLTNRGDVEK
GKKIFVQKCAQCHTVEKG
GKHKT
```

What info do we have from the input?

raw MS/MS spectra



- the spectrum itself
  - hopefully peaks correspond to fragment ions
- precursor m/z
- possibly precursor charge
- the sample origin
- enzyme used to digest the proteins
- the instrument used to acquire the data

what sequence database to search

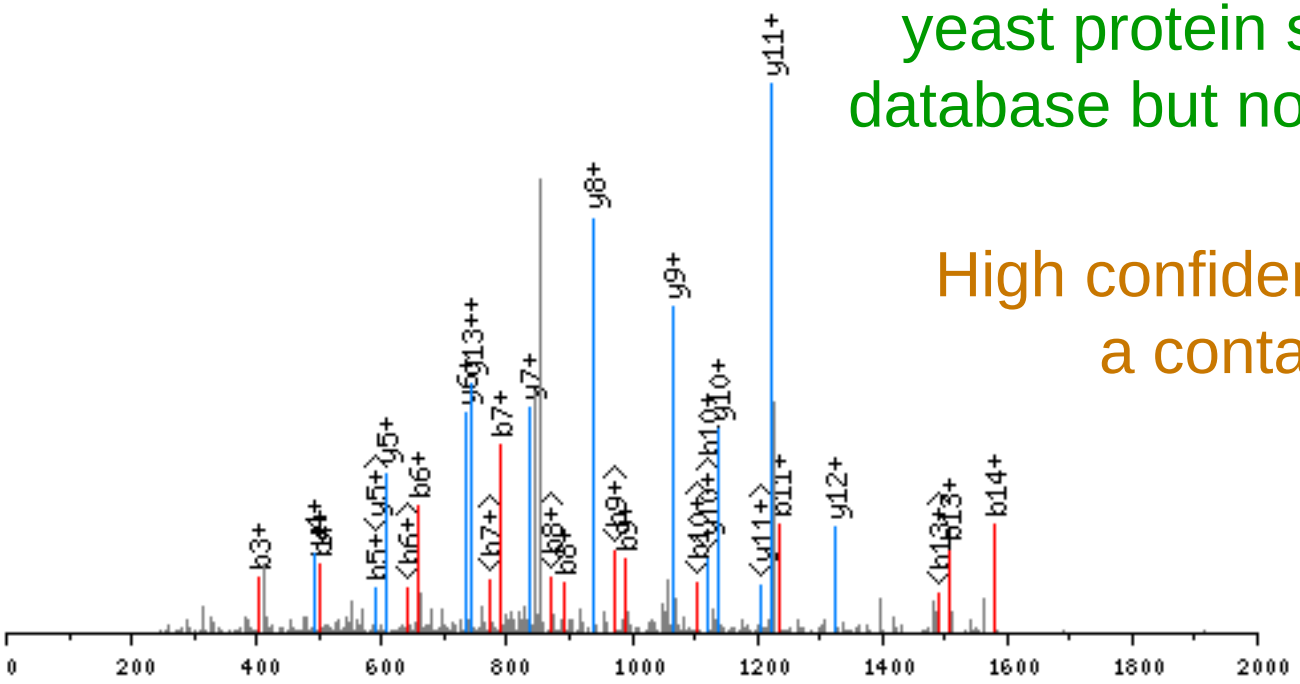
modifications to consider

enzyme specificity for search

mass accuracy & fragmentation settings

Obvious fact to keep in mind

- Sequence must be in database in order to possibly be identified.



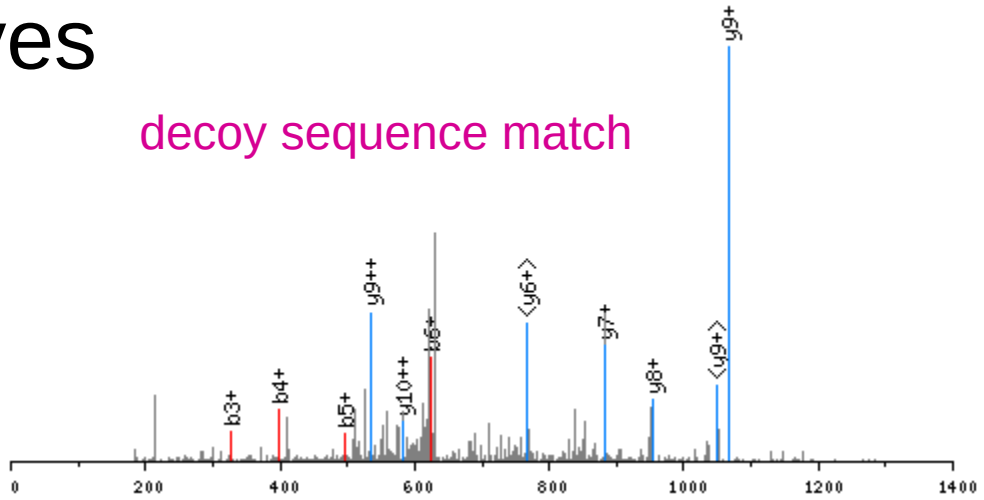
Yeast sample, gel separation, gorgeous spectrum, searched yeast protein sequence database but not identified.

High confidence match from a contaminant protein.

We don't always get black+white answers

- With so much data, it's easy to identify false positives

decoy sequence match



5%??

1% PSM FDR = ??? protein level FDR

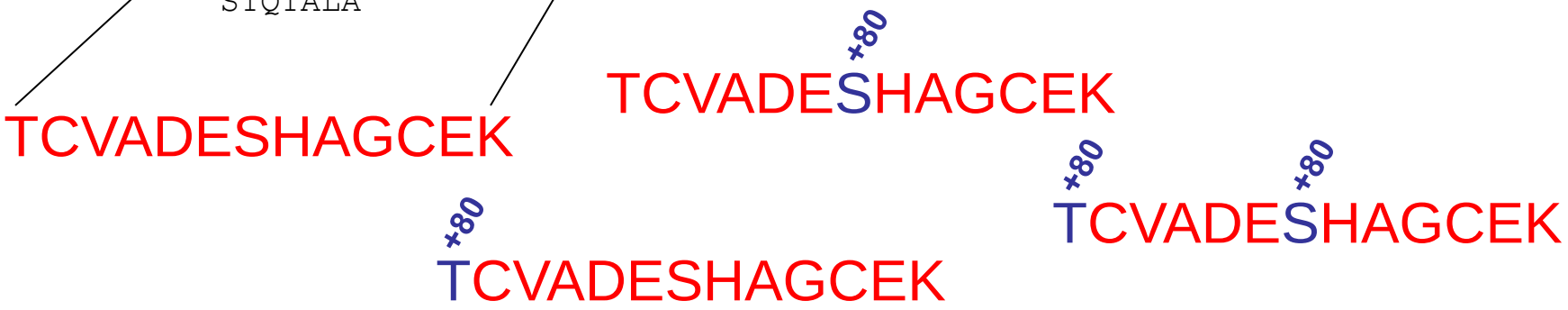
10%??

20%??

Post-translation modifications

- PTMs can be identified

```
>sp|P02769|ALBU_BOVIN Serum albumin OS=Bos taurus
MKWVTFISLLLLFSSAYSRGVFRRDTHKSEIAHRFKDLGEEHFKGLVLIAFSQYLQQCPF
DEHVKLVNELTEFAKTCVADESHAGCEKSLHTLFGDELCKVASLRETYGDMADCCEKQEP
ERNECFLSHKDDSPDLPKLKPDPNTLCDEFKADEKKFWGKYLYEIARRHPYFYAPELLYY
ANKYNGVFQECCQAEDKGACLLPKIETMREKVLASSARQRLRCASIQKFGERALKAWSVA
RLSQKFPKAEFVEVTKLVTDLTKVHKECCHGDLLECADDRADLAKYICDNQDTISSKLKE
CCDKPLLEKSHCIAEVEKDAIPENLPPLTADFAEDKDVCKNYQEAKDAFLGSFLYEYSRR
HPEYAVSVLLRLAKEYEATLEECCAKDDPHACYSTVFDKLKHLVDEPQNLIKQNCDQFEK
LGEYGFQNALIVRYTRKVPQVSTPTLVEVSRSLGKVGTRCCTKPESERMPCTEDYLSLIL
NRLCVLHEKTPVSEKVTKCCTESLVNRRPCFSALTPDETYVPKAFDEKLFTFHADICTLP
DTEKQIKKQTALVELLKHKPKATEEQLKTVMENFVAFVDKCCAADDKEACFAVEGPKLVV
STQTALA
```



Quantification

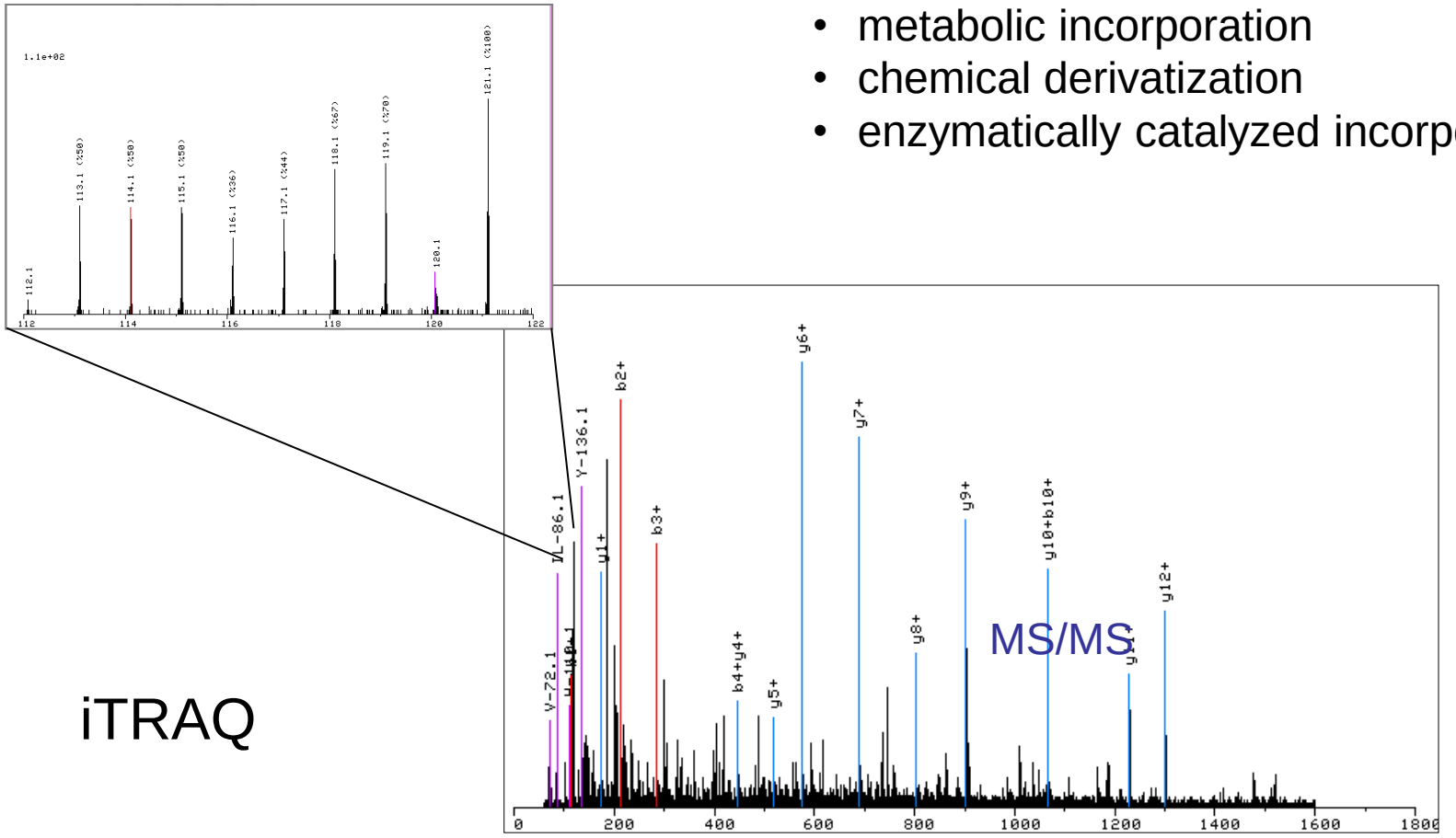
- Many ways of doing it
  - Stable isotope labelling
    - metabolic incorporation
    - chemical derivatization
    - enzymatically catalyzed incorporation
  - Spectral counting
  - Label free
- Targeted vs. untargeted analysis

Quantification

- Data analysis not as straightforward as we might hope for.
- Quantification tools might not be closely coupled to peptide & protein ID tools.

iTRAQ / TMT

- metabolic incorporation
- chemical derivatization
- enzymatically catalyzed incorporation

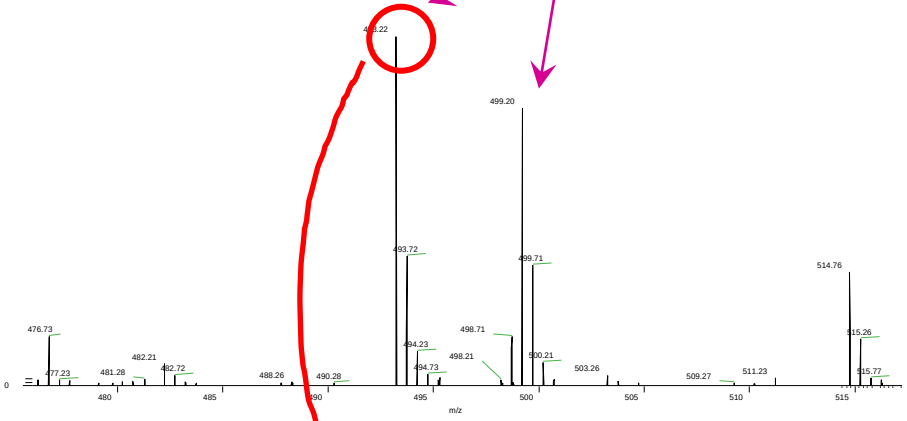


iTRAQ

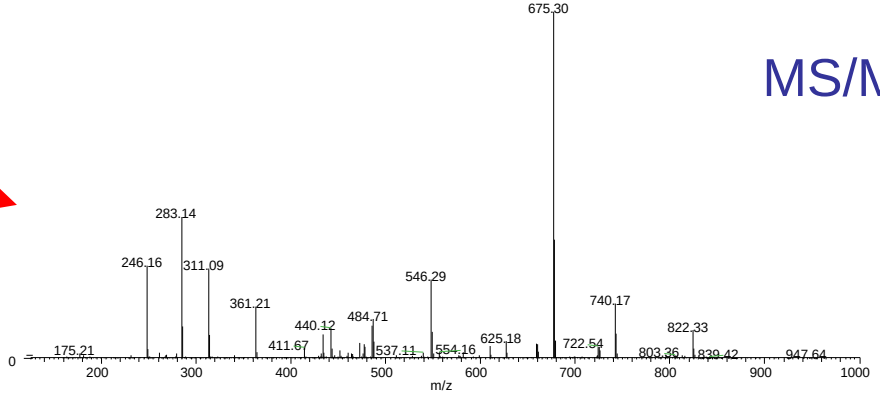
MS/MS

SILAC, ICAT, ^{18}O digestion

MS=ratio

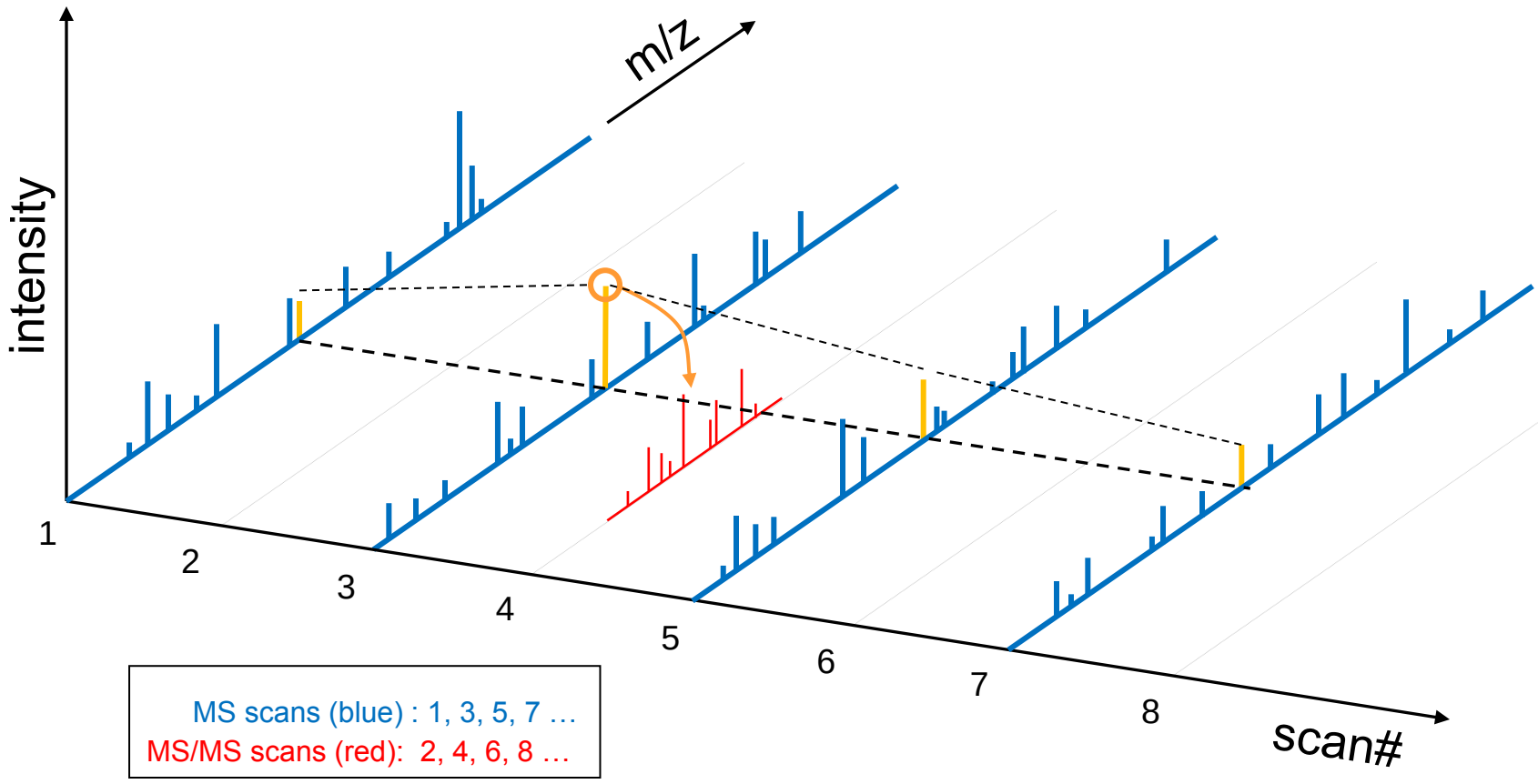


SILAC



MS/MS=ID

MS and MS/MS relationship

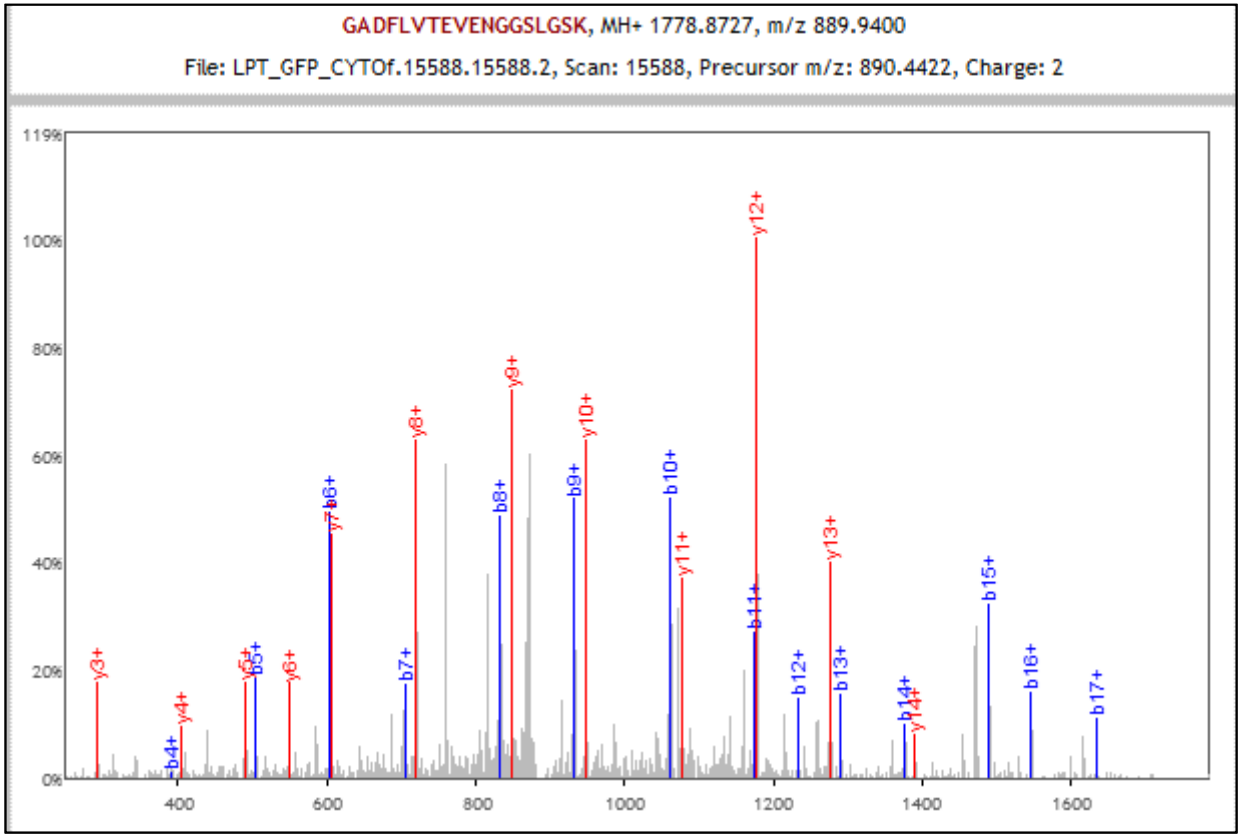


Scoring


HUSKY STADIUM
HUSKIES
0:08
BRUINS
29
19
DOWN
2
TO GO
11
BALL ON
39

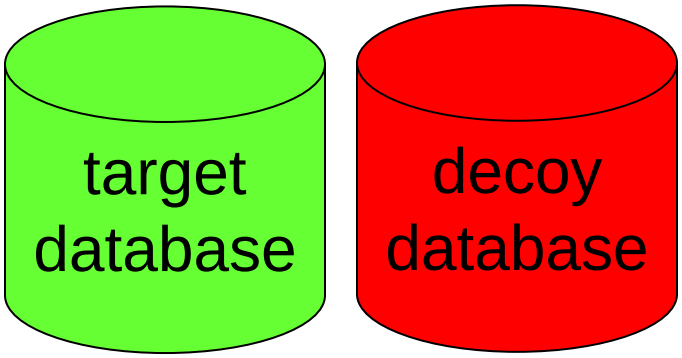
Scoring

How do we know if this is a right or wrong match?



Target-Decoy strategy

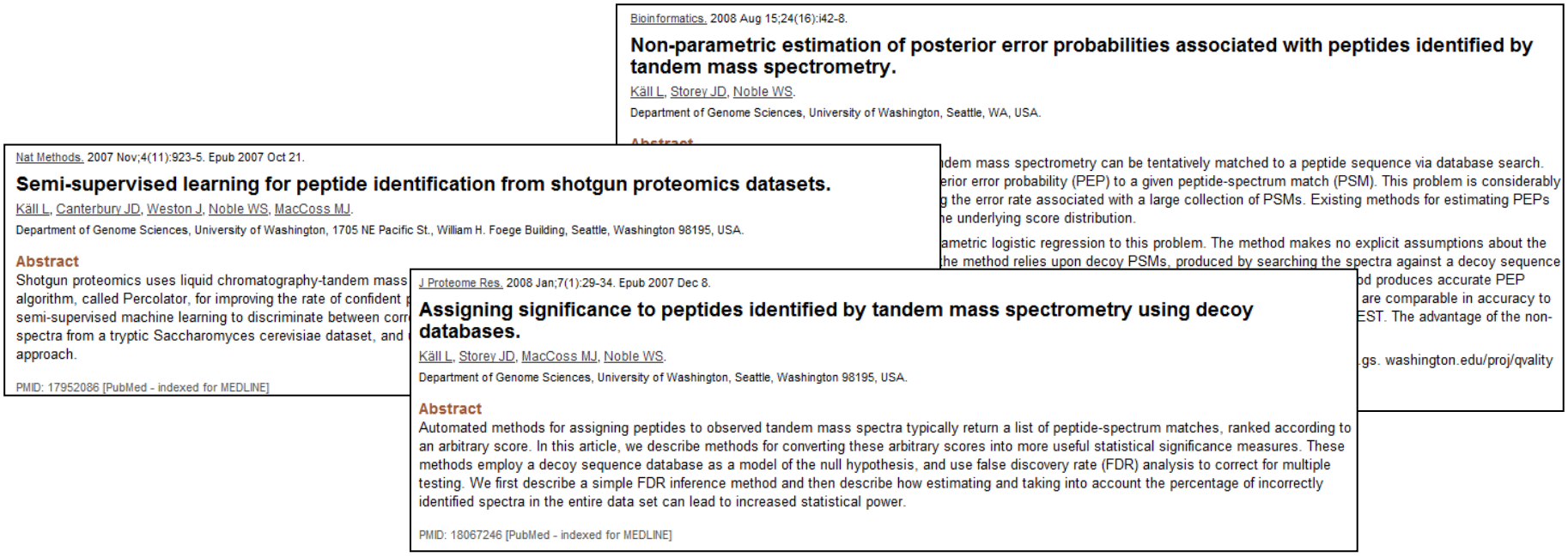
- Search against known wrong “decoy” sequences; use these matches to estimate false discovery rate (FDR).
- FDR = #decoys / #targets



score	sequence
230.0	KAADETWEPFASGK
229.3	AADETWEPFASGK
218.5	TSESGELHGLTTEDK
195.4	TYLFVEGLYKVELDTK
192.3	WAVELDTKSYWK
186.9	QECPLMVKVLDAVR
181.7	LSPYSYSTTALVSSPK
169.3	VLDAVRGSPAANVGVK
165.6	HEFAEVVFTANDSGPR
159.3	DLRSWTAADTAAQIK

More sophisticated algorithms

- Percolator
  - support vector machine
  - requires target+decoy searches
  - Implemented in Mascot



Bioinformatics. 2008 Aug 15;24(16):i42-8.

Non-parametric estimation of posterior error probabilities associated with peptides identified by tandem mass spectrometry.

Käll L, Storey JD, Noble WS.

Department of Genome Sciences, University of Washington, Seattle, WA, USA.

dem mass spectrometry can be tentatively matched to a peptide sequence via database search.
erior error probability (PEP) to a given peptide-spectrum match (PSM). This problem is considerably
g the error rate associated with a large collection of PSMs. Existing methods for estimating PEPs
he underlying score distribution.

ametric logistic regression to this problem. The method makes no explicit assumptions about the
he method relies upon decoy PSMs, produced by searching the spectra against a decoy sequence
od produces accurate PEP
are comparable in accuracy to
EST. The advantage of the non-

gs. washington.edu/proj/qvality

Nat Methods. 2007 Nov;4(11):923-5. Epub 2007 Oct 21.

Semi-supervised learning for peptide identification from shotgun proteomics datasets.

Käll L, Canterbury JD, Weston J, Noble WS, MacCoss MJ.

Department of Genome Sciences, University of Washington, 1705 NE Pacific St., William H. Foege Building, Seattle, Washington 98195, USA.

Abstract

Shotgun proteomics uses liquid chromatography-tandem mass
algorithm, called Percolator, for improving the rate of confident
semi-supervised machine learning to discriminate between corr
spectra from a tryptic Saccharomyces cerevisiae dataset, and
approach.

PMID: 17952086 [PubMed - indexed for MEDLINE]

J Proteome Res. 2008 Jan;7(1):29-34. Epub 2007 Dec 8.

Assigning significance to peptides identified by tandem mass spectrometry using decoy databases.

Käll L, Storey JD, MacCoss MJ, Noble WS.

Department of Genome Sciences, University of Washington, Seattle, Washington 98195, USA.

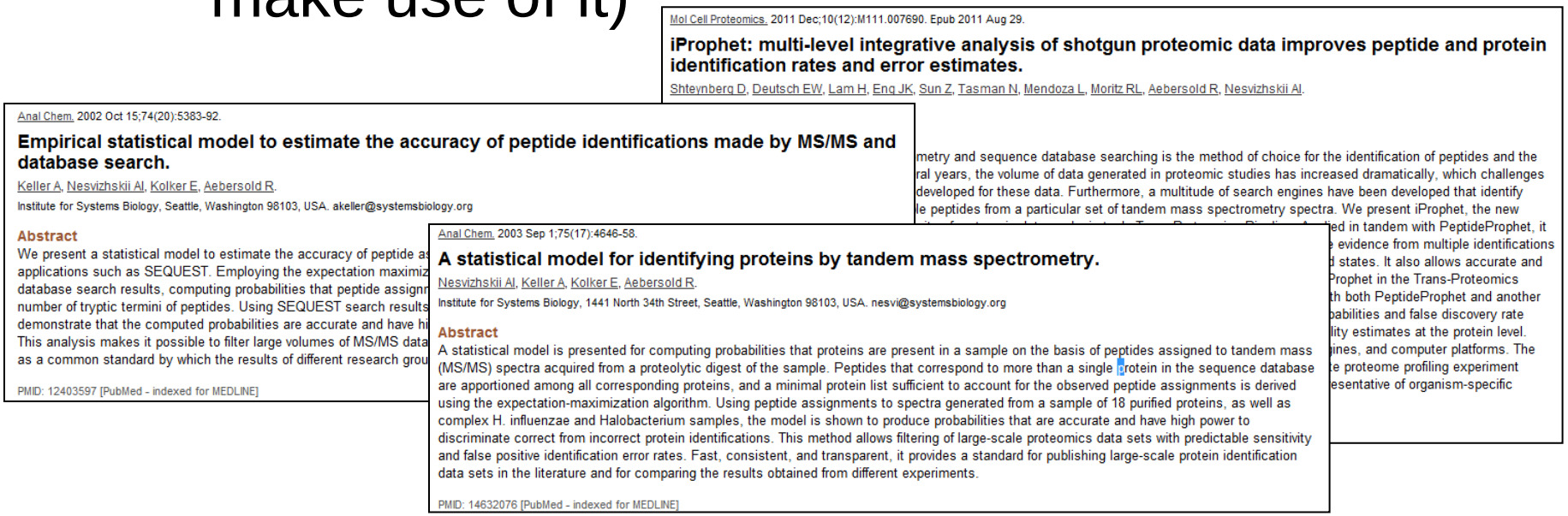
Abstract

Automated methods for assigning peptides to observed tandem mass spectra typically return a list of peptide-spectrum matches, ranked according to an arbitrary score. In this article, we describe methods for converting these arbitrary scores into more useful statistical significance measures. These methods employ a decoy sequence database as a model of the null hypothesis, and use false discovery rate (FDR) analysis to correct for multiple testing. We first describe a simple FDR inference method and then describe how estimating and taking into account the percentage of incorrectly identified spectra in the entire data set can lead to increased statistical power.

PMID: 18067246 [PubMed - indexed for MEDLINE]

More sophisticated algorithms

- TPP's Prophets (Peptide, Protein, i)
  - mixture model analysis, protein grouping, multilevel integrative analysis
  - does not require decoy search (although can make use of it)



Mol Cell Proteomics. 2011 Dec;10(12):M111.007690. Epub 2011 Aug 29.

iProphet: multi-level integrative analysis of shotgun proteomic data improves peptide and protein identification rates and error estimates.

Shteynberg D, Deutsch EW, Lam H, Eng JK, Sun Z, Tasman N, Mendoza L, Moritz RL, Aebersold R, Nesvizhskii AI.

metry and sequence database searching is the method of choice for the identification of peptides and the
ral years, the volume of data generated in proteomic studies has increased dramatically, which challenges
developed for these data. Furthermore, a multitude of search engines have been developed that identify
le peptides from a particular set of tandem mass spectrometry spectra. We present iProphet, the new
ied in tandem with PeptideProphet, it
e evidence from multiple identifications
d states. It also allows accurate and
Prophet in the Trans-Proteomics
th both PeptideProphet and another
babilities and false discovery rate
lity estimates at the protein level.
ines, and computer platforms. The
te proteome profiling experiment
esentative of organism-specific

Anal Chem. 2002 Oct 15;74(20):5383-92.

Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search.

Keller A, Nesvizhskii AI, Kolker E, Aebersold R.

Institute for Systems Biology, Seattle, Washington 98103, USA. akeller@systemsbiology.org

Abstract

We present a statistical model to estimate the accuracy of peptide as
applications such as SEQUEST. Employing the expectation maximiz
database search results, computing probabilities that peptide assignr
number of tryptic termini of peptides. Using SEQUEST search results
demonstrate that the computed probabilities are accurate and have hi
This analysis makes it possible to filter large volumes of MS/MS data
as a common standard by which the results of different research grou

PMID: 12403597 [PubMed - indexed for MEDLINE]

Anal Chem. 2003 Sep 1;75(17):4646-58.

A statistical model for identifying proteins by tandem mass spectrometry.

Nesvizhskii AI, Keller A, Kolker E, Aebersold R.

Institute for Systems Biology, 1441 North 34th Street, Seattle, Washington 98103, USA. nesvi@systemsbiology.org

Abstract

A statistical model is presented for computing probabilities that proteins are present in a sample on the basis of peptides assigned to tandem mass (MS/MS) spectra acquired from a proteolytic digest of the sample. Peptides that correspond to more than a single protein in the sequence database are apportioned among all corresponding proteins, and a minimal protein list sufficient to account for the observed peptide assignments is derived using the expectation-maximization algorithm. Using peptide assignments to spectra generated from a sample of 18 purified proteins, as well as complex H. influenzae and Halobacterium samples, the model is shown to produce probabilities that are accurate and have high power to discriminate correct from incorrect protein identifications. This method allows filtering of large-scale proteomics data sets with predictable sensitivity and false positive identification error rates. Fast, consistent, and transparent, it provides a standard for publishing large-scale protein identification data sets in the literature and for comparing the results obtained from different experiments.

PMID: 14632076 [PubMed - indexed for MEDLINE]

Software tools

Good time for a break

A partial list of MS/MS search tools

- Mascot *
- X!Tandem *
- OMSSA *
- SpectrumMill
- Phenyx
- MS-GF+
- MyriMatch
- ProbID
- SEQUEST *
- Andromeda
- PepSplice
- PepProbe
- RAId_DbS
- ProteinPilot
- ProteinLynx GS
- Crux *

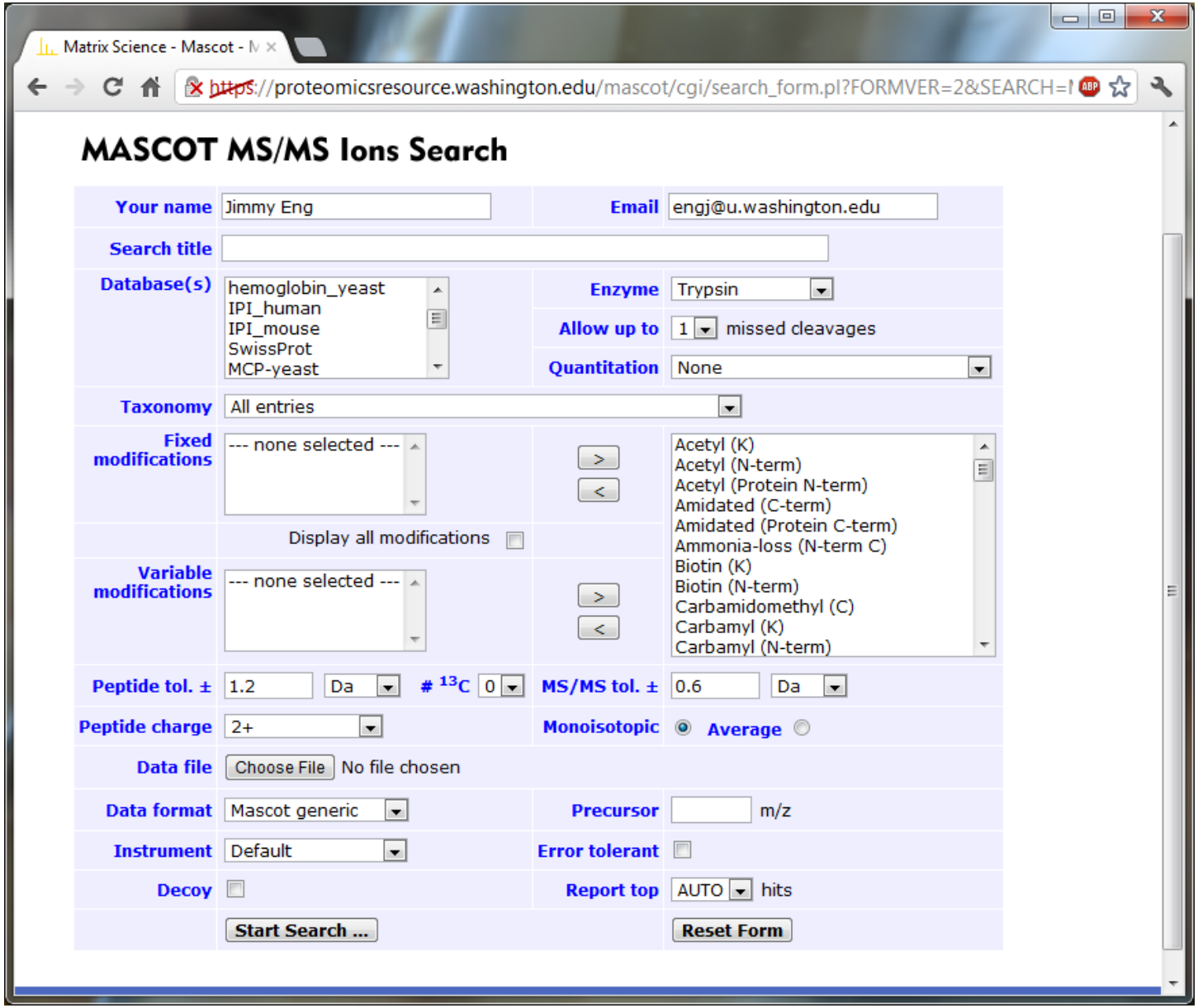
And *many* others …

Mascot

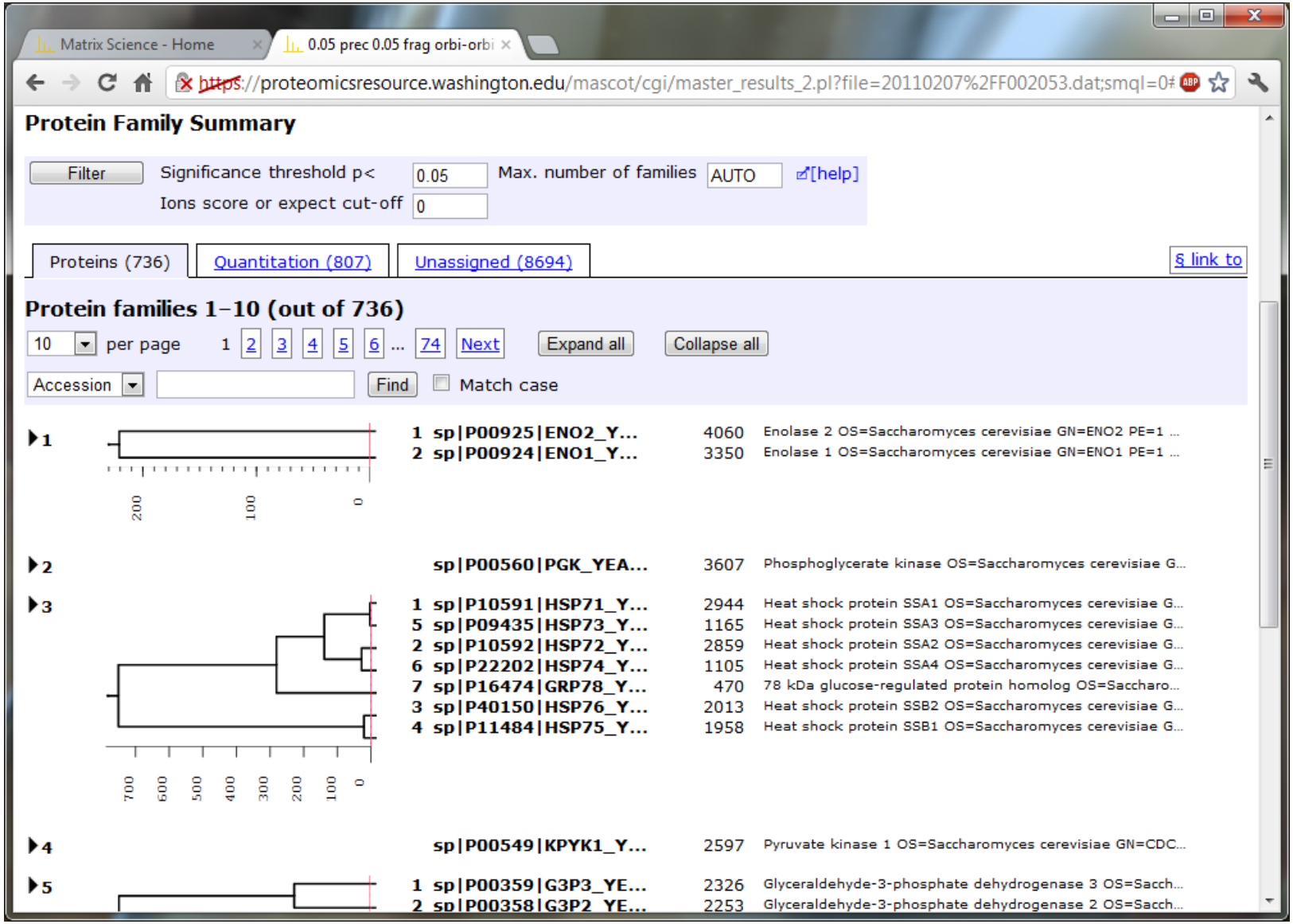
- Widely used search engine
- Probabilistic scoring
- Consistently good identification performance based on various comparisons
- Supported by lots of 3rd party software
- Online version available

www.matrixscience.com

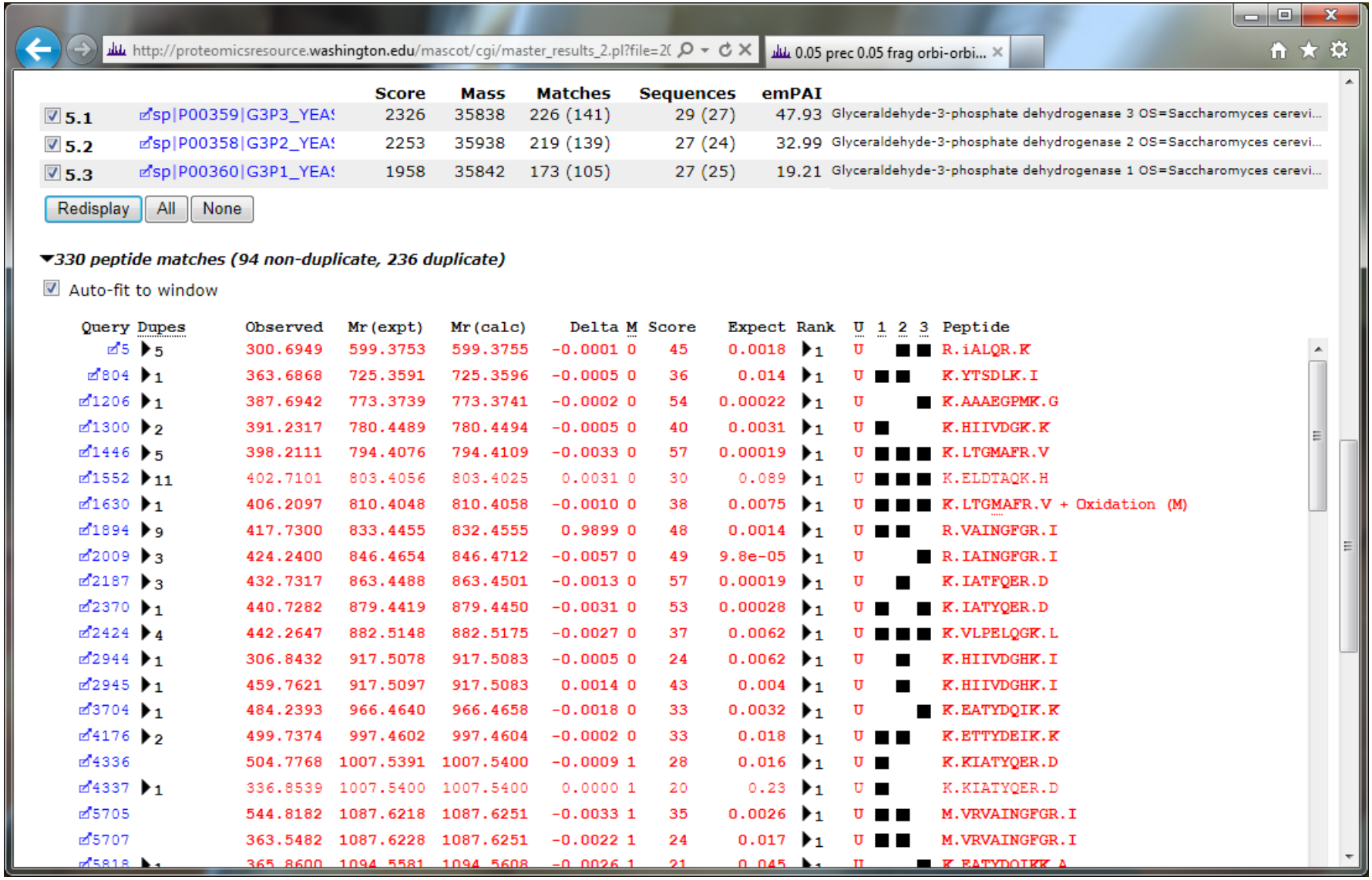
Mascot


Matrix Science - Mascot - M
https://proteomicsresource.washington.edu/mascot/cgi/search_form.pl?FORMVER=2&SEARCH=M
MASCOT MS/MS Ions Search
Your name
Jimmy Eng
Email
engj@u.washington.edu
Search title
Database(s)
hemoglobin_yeast
IPI_human
IPI_mouse
SwissProt
MCP-yeast
Enzyme
Trypsin
Allow up to
1
missed cleavages
Quantitation
None
Taxonomy
All entries
Fixed modifications
--- none selected ---
Display all modifications
Variable modifications
--- none selected ---
Acetyl (K)
Acetyl (N-term)
Acetyl (Protein N-term)
Amidated (C-term)
Amidated (Protein C-term)
Ammonia-loss (N-term C)
Biotin (K)
Biotin (N-term)
Carbamidomethyl (C)
Carbamyl (K)
Carbamyl (N-term)
Peptide tol. ±
1.2
Da
13C
0
MS/MS tol. ±
0.6
Da
Peptide charge
2+
Monoisotopic
Average
Data file
Choose File
No file chosen
Data format
Mascot generic
Precursor
m/z
Instrument
Default
Error tolerant
Decoy
Report top
AUTO
hits
Start Search ...
Reset Form

Mascot


Matrix Science - Home
0.05 prec 0.05 frag orbi-orbi
https://proteomicsresource.washington.edu/mascot/cgi/master_results_2.pl?file=20110207%2FF002053.dat;smql=0#
Protein Family Summary
Filter
Significance threshold p< 0.05
Max. number of families AUTO
[help]
Ions score or expect cut-off 0
Proteins (736)
Quantitation (807)
Unassigned (8694)
§ link to
Protein families 1–10 (out of 736)
10 per page
1 2 3 4 5 6 ... 74 Next
Expand all
Collapse all
Accession
Find
Match case
1
1 sp|P00925|ENO2_Y... 4060 Enolase 2 OS=Saccharomyces cerevisiae GN=ENO2 PE=1 ...
2 sp|P00924|ENO1_Y... 3350 Enolase 1 OS=Saccharomyces cerevisiae GN=ENO1 PE=1 ...
200
100
0
2
sp|P00560|PGK_YEA... 3607 Phosphoglycerate kinase OS=Saccharomyces cerevisiae G...
3
1 sp|P10591|HSP71_Y... 2944 Heat shock protein SSA1 OS=Saccharomyces cerevisiae G...
5 sp|P09435|HSP73_Y... 1165 Heat shock protein SSA3 OS=Saccharomyces cerevisiae G...
2 sp|P10592|HSP72_Y... 2859 Heat shock protein SSA2 OS=Saccharomyces cerevisiae G...
6 sp|P22202|HSP74_Y... 1105 Heat shock protein SSA4 OS=Saccharomyces cerevisiae G...
7 sp|P16474|GRP78_Y... 470 78 kDa glucose-regulated protein homolog OS=Saccharo...
3 sp|P40150|HSP76_Y... 2013 Heat shock protein SSB2 OS=Saccharomyces cerevisiae G...
4 sp|P11484|HSP75_Y... 1958 Heat shock protein SSB1 OS=Saccharomyces cerevisiae G...
700
600
500
400
300
200
100
0
4
sp|P00549|KPYK1_Y... 2597 Pyruvate kinase 1 OS=Saccharomyces cerevisiae GN=CDC...
5
1 sp|P00359|G3P3_YE... 2326 Glyceraldehyde-3-phosphate dehydrogenase 3 OS=Sacch...
2 sp|P00358|G3P2_YE... 2253 Glyceraldehyde-3-phosphate dehydrogenase 2 OS=Sacch...

Mascot



		Score	Mass	Matches	Sequences	emPAI	
☑ 5.1	sp|P00359|G3P3_YEAS	2326	35838	226 (141)	29 (27)	47.93	Glyceraldehyde-3-phosphate dehydrogenase 3 OS=Saccharomyces cerevi...
☑ 5.2	sp|P00358|G3P2_YEAS	2253	35938	219 (139)	27 (24)	32.99	Glyceraldehyde-3-phosphate dehydrogenase 2 OS=Saccharomyces cerevi...
☑ 5.3	sp|P00360|G3P1_YEAS	1958	35842	173 (105)	27 (25)	19.21	Glyceraldehyde-3-phosphate dehydrogenase 1 OS=Saccharomyces cerevi...

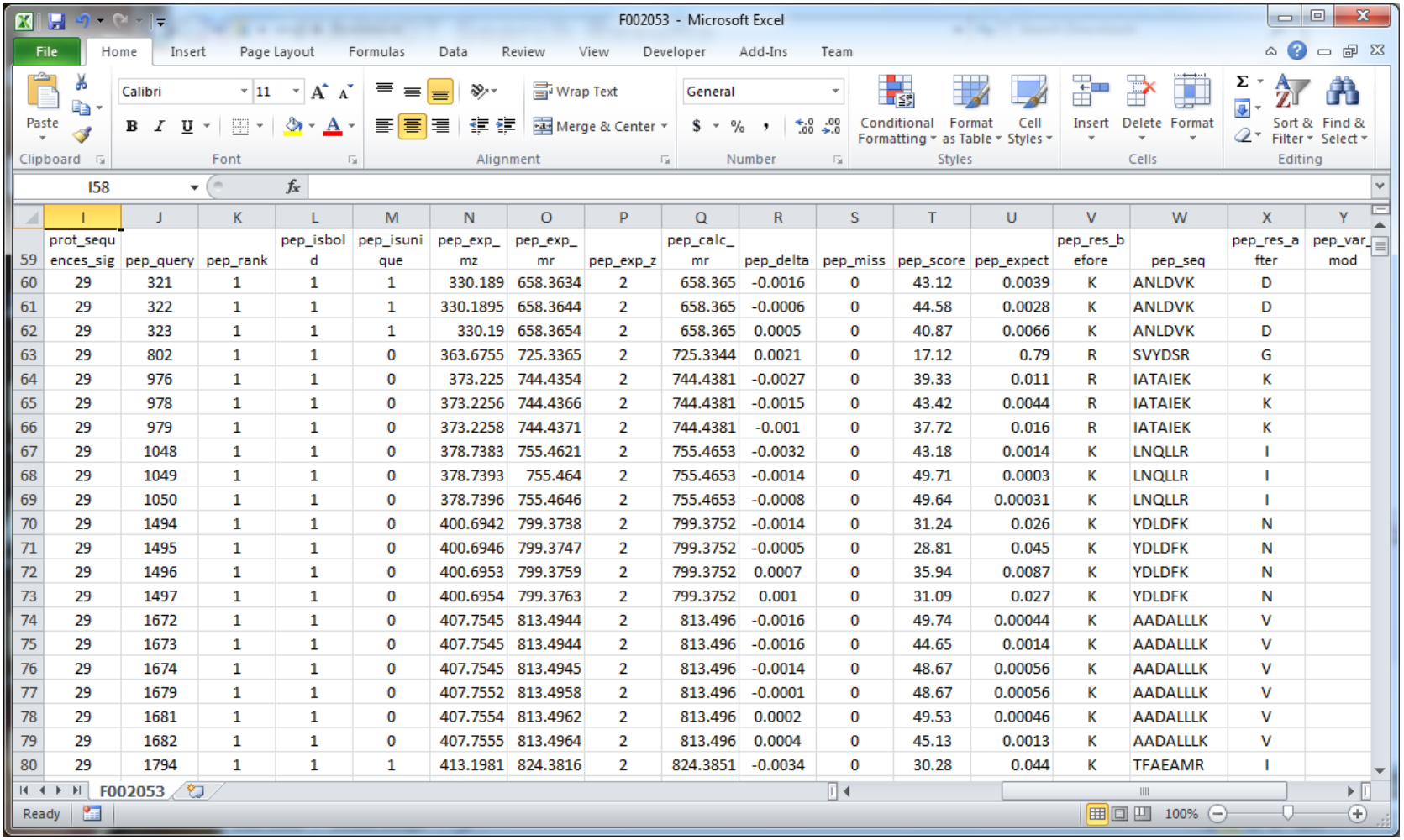
Redisplay | All | None

▼*330 peptide matches (94 non-duplicate, 236 duplicate)*

☑ Auto-fit to window

Query	Dupes	Observed	Mr(expt)	Mr(calc)	Delta	M	Score	Expect	Rank	U	1	2	3	Peptide
5	▶5	300.6949	599.3753	599.3755	-0.0001	0	45	0.0018	▶1	U		■	■	R.iALQR.K
804	▶1	363.6868	725.3591	725.3596	-0.0005	0	36	0.014	▶1	U	■	■		K.YTSDLK.I
1206	▶1	387.6942	773.3739	773.3741	-0.0002	0	54	0.00022	▶1	U			■	K.AAAEGPMK.G
1300	▶2	391.2317	780.4489	780.4494	-0.0005	0	40	0.0031	▶1	U	■			K.HIIVDGK.K
1446	▶5	398.2111	794.4076	794.4109	-0.0033	0	57	0.00019	▶1	U	■	■	■	K.LTGMAFR.V
1552	▶11	402.7101	803.4056	803.4025	0.0031	0	30	0.089	▶1	U	■	■	■	K.ELDTAQK.H
1630	▶1	406.2097	810.4048	810.4058	-0.0010	0	38	0.0075	▶1	U	■	■	■	K.LTGMAFR.V + Oxidation (M)
1894	▶9	417.7300	833.4455	832.4555	0.9899	0	48	0.0014	▶1	U	■	■		R.VAINGFGR.I
2009	▶3	424.2400	846.4654	846.4712	-0.0057	0	49	9.8e-05	▶1	U			■	R.IAINGFGR.I
2187	▶3	432.7317	863.4488	863.4501	-0.0013	0	57	0.00019	▶1	U		■		K.IATFQER.D
2370	▶1	440.7282	879.4419	879.4450	-0.0031	0	53	0.00028	▶1	U	■		■	K.IATYQER.D
2424	▶4	442.2647	882.5148	882.5175	-0.0027	0	37	0.0062	▶1	U	■	■	■	K.VLPELQGK.L
2944	▶1	306.8432	917.5078	917.5083	-0.0005	0	24	0.0062	▶1	U		■		K.HIIVDGHK.I
2945	▶1	459.7621	917.5097	917.5083	0.0014	0	43	0.004	▶1	U		■		K.HIIVDGHK.I
3704	▶1	484.2393	966.4640	966.4658	-0.0018	0	33	0.0032	▶1	U			■	K.EATYDQIK.K
4176	▶2	499.7374	997.4602	997.4604	-0.0002	0	33	0.018	▶1	U	■	■		K.ETTYDEIK.K
4336		504.7768	1007.5391	1007.5400	-0.0009	1	28	0.016	▶1	U	■			K.KIATYQER.D
4337	▶1	336.8539	1007.5400	1007.5400	0.0000	1	20	0.23	▶1	U	■			K.KIATYQER.D
5705		544.8182	1087.6218	1087.6251	-0.0033	1	35	0.0026	▶1	U	■	■		M.VRVAINGFGR.I
5707		363.5482	1087.6228	1087.6251	-0.0022	1	24	0.017	▶1	U	■	■		M.VRVAINGFGR.I
5818	▶1	365.8600	1094.5581	1094.5608	-0.0026	1	21	0.045	▶1	U			■	K.EATYDQIKK.A

Mascot

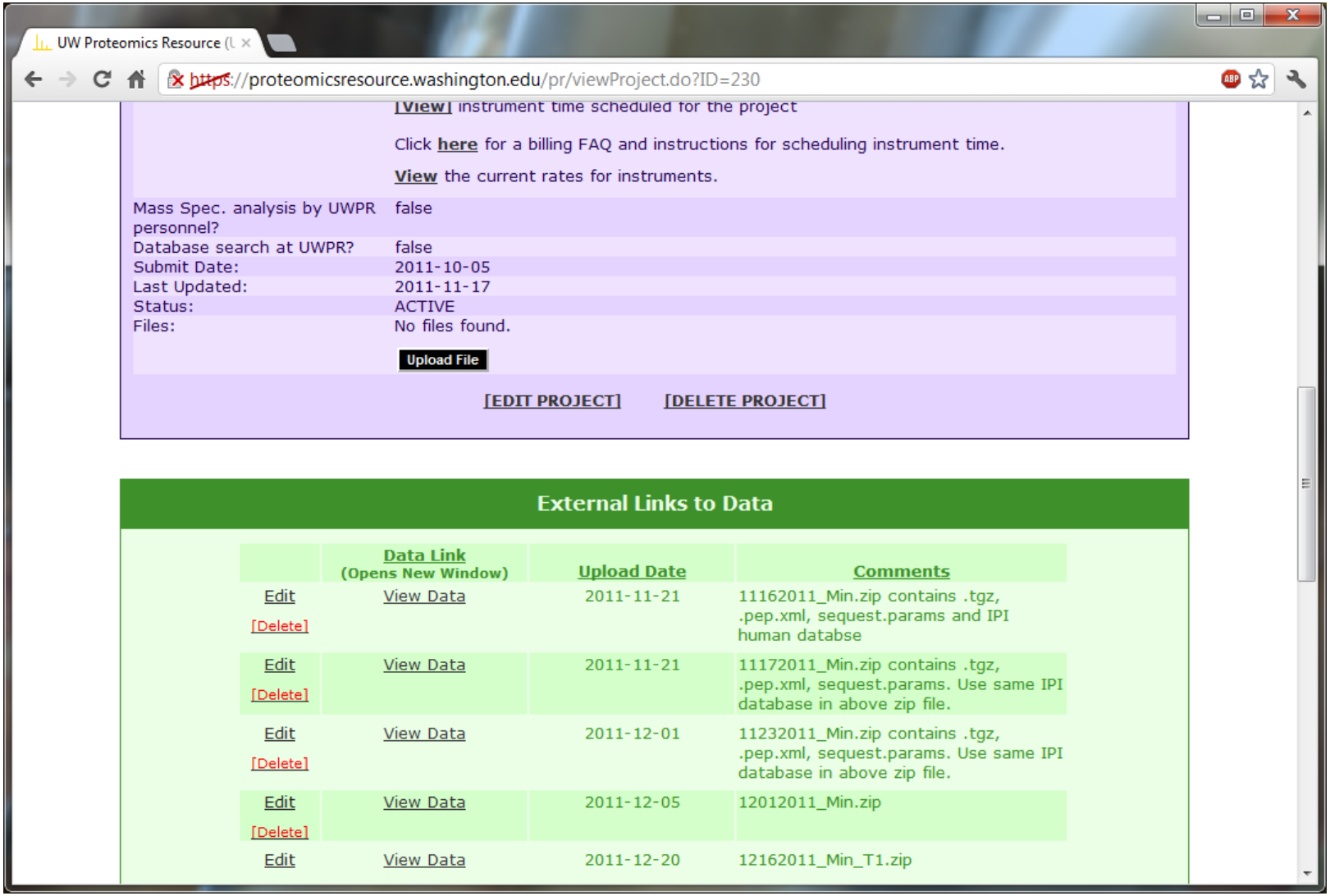


	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y
59	prot_sequences_sig	pep_query	pep_rank	pep_isbold	pep_isunique	pep_exp_mz	pep_exp_mr	pep_exp_z	pep_calc_mr	pep_delta	pep_miss	pep_score	pep_expect	pep_res_before	pep_seq	pep_res_after	pep_var_mod
60	29	321	1	1	1	330.189	658.3634	2	658.365	-0.0016	0	43.12	0.0039	K	ANLDVK	D	
61	29	322	1	1	1	330.1895	658.3644	2	658.365	-0.0006	0	44.58	0.0028	K	ANLDVK	D	
62	29	323	1	1	1	330.19	658.3654	2	658.365	0.0005	0	40.87	0.0066	K	ANLDVK	D	
63	29	802	1	1	0	363.6755	725.3365	2	725.3344	0.0021	0	17.12	0.79	R	SVYDSR	G	
64	29	976	1	1	0	373.225	744.4354	2	744.4381	-0.0027	0	39.33	0.011	R	IATAIEK	K	
65	29	978	1	1	0	373.2256	744.4366	2	744.4381	-0.0015	0	43.42	0.0044	R	IATAIEK	K	
66	29	979	1	1	0	373.2258	744.4371	2	744.4381	-0.001	0	37.72	0.016	R	IATAIEK	K	
67	29	1048	1	1	0	378.7383	755.4621	2	755.4653	-0.0032	0	43.18	0.0014	K	LNQLLR	I	
68	29	1049	1	1	0	378.7393	755.464	2	755.4653	-0.0014	0	49.71	0.0003	K	LNQLLR	I	
69	29	1050	1	1	0	378.7396	755.4646	2	755.4653	-0.0008	0	49.64	0.00031	K	LNQLLR	I	
70	29	1494	1	1	0	400.6942	799.3738	2	799.3752	-0.0014	0	31.24	0.026	K	YDLDFK	N	
71	29	1495	1	1	0	400.6946	799.3747	2	799.3752	-0.0005	0	28.81	0.045	K	YDLDFK	N	
72	29	1496	1	1	0	400.6953	799.3759	2	799.3752	0.0007	0	35.94	0.0087	K	YDLDFK	N	
73	29	1497	1	1	0	400.6954	799.3763	2	799.3752	0.001	0	31.09	0.027	K	YDLDFK	N	
74	29	1672	1	1	0	407.7545	813.4944	2	813.496	-0.0016	0	49.74	0.00044	K	AADALLLK	V	
75	29	1673	1	1	0	407.7545	813.4944	2	813.496	-0.0016	0	44.65	0.0014	K	AADALLLK	V	
76	29	1674	1	1	0	407.7545	813.4945	2	813.496	-0.0014	0	48.67	0.00056	K	AADALLLK	V	
77	29	1679	1	1	0	407.7552	813.4958	2	813.496	-0.0001	0	48.67	0.00056	K	AADALLLK	V	
78	29	1681	1	1	0	407.7554	813.4962	2	813.496	0.0002	0	49.53	0.00046	K	AADALLLK	V	
79	29	1682	1	1	0	407.7555	813.4964	2	813.496	0.0004	0	45.13	0.0013	K	AADALLLK	V	
80	29	1794	1	1	1	413.1981	824.3816	2	824.3851	-0.0034	0	30.28	0.044	K	TFAEAMR	I	

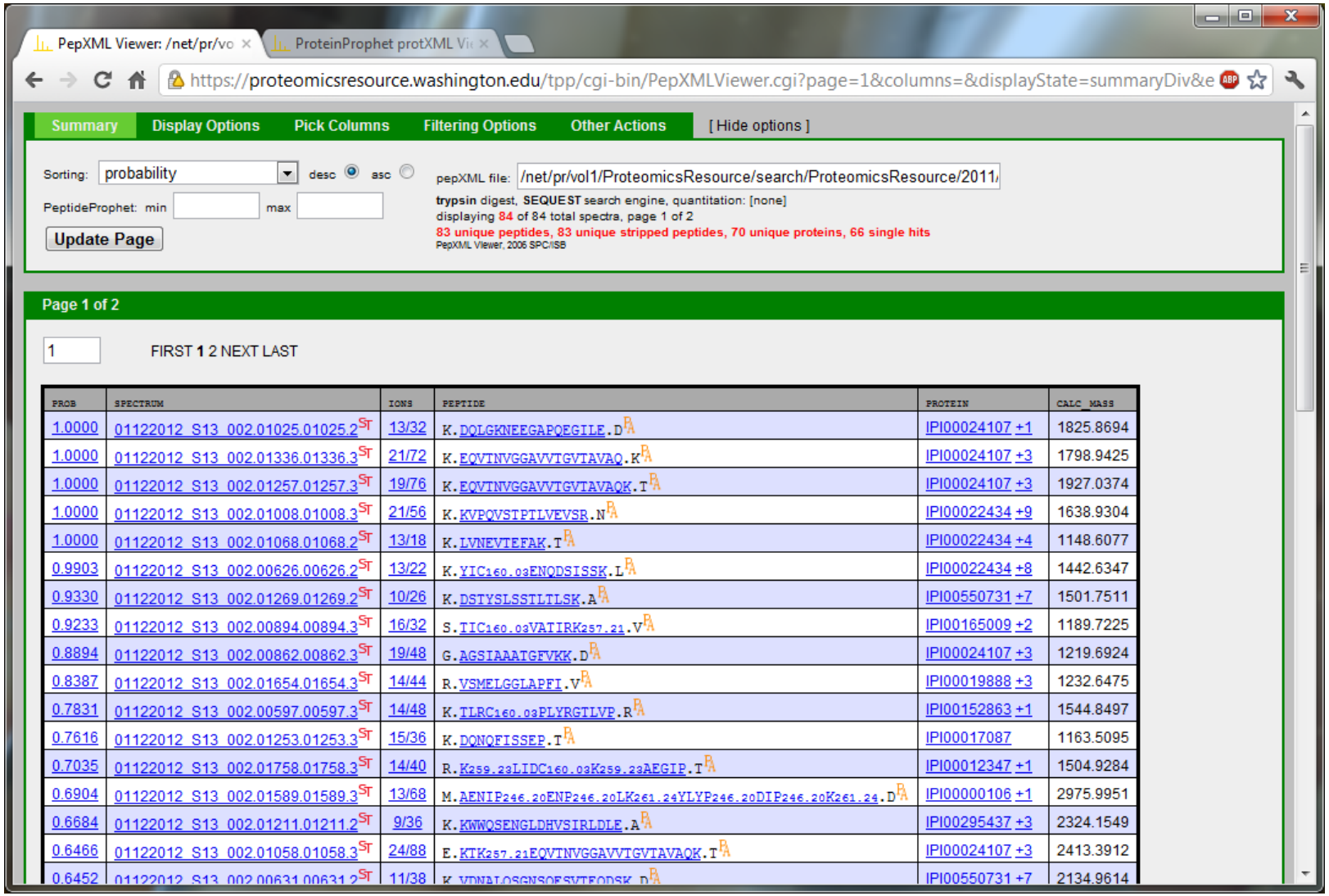
TPP

- Trans-Proteomic Pipeline (TPP) is a suite of open source tools originally developed in the Aebersold group at ISB in ~2003/2004
- Particularly noted for peptide and protein statistical validation

TPP output at UWPR

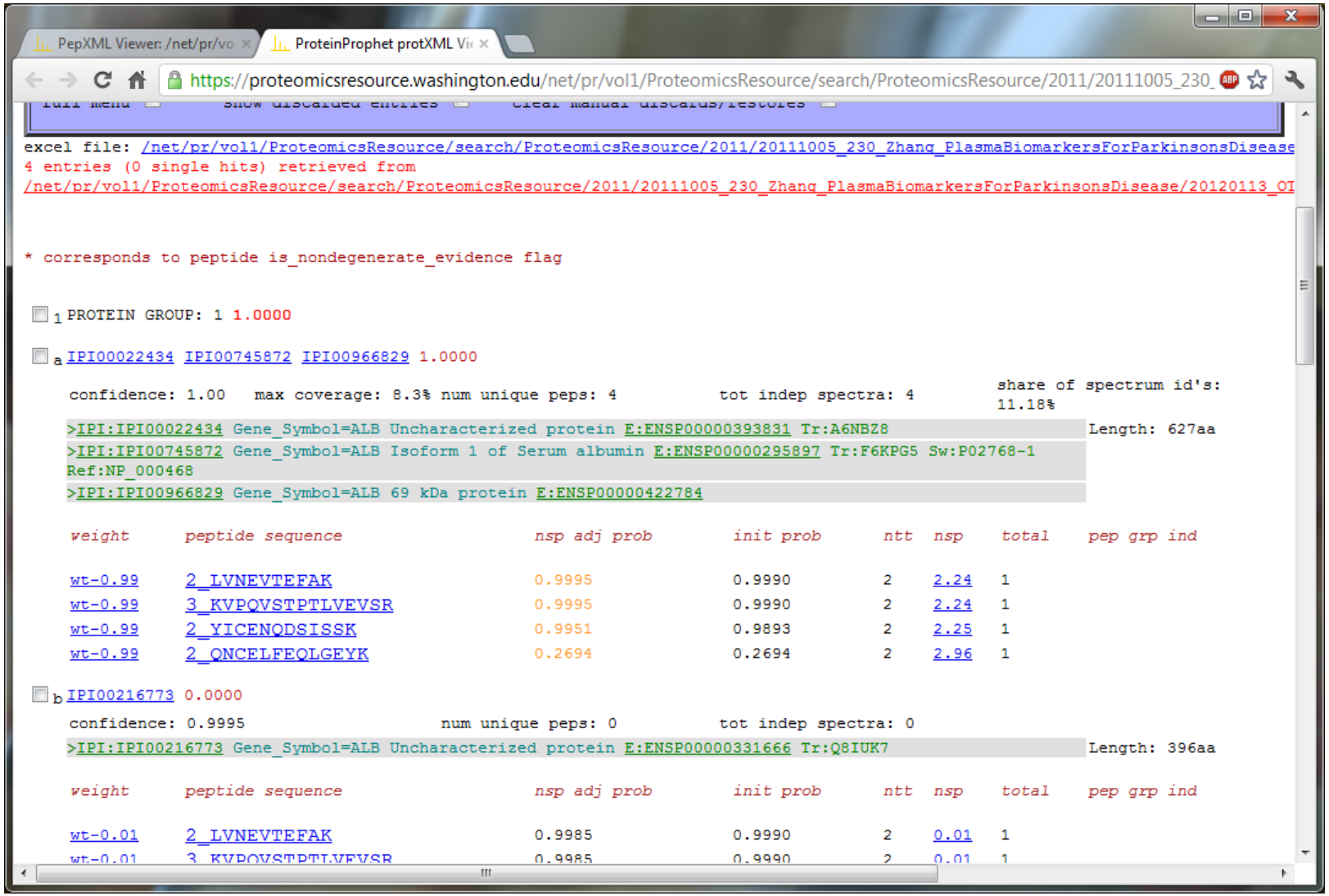


TPP peptide level view



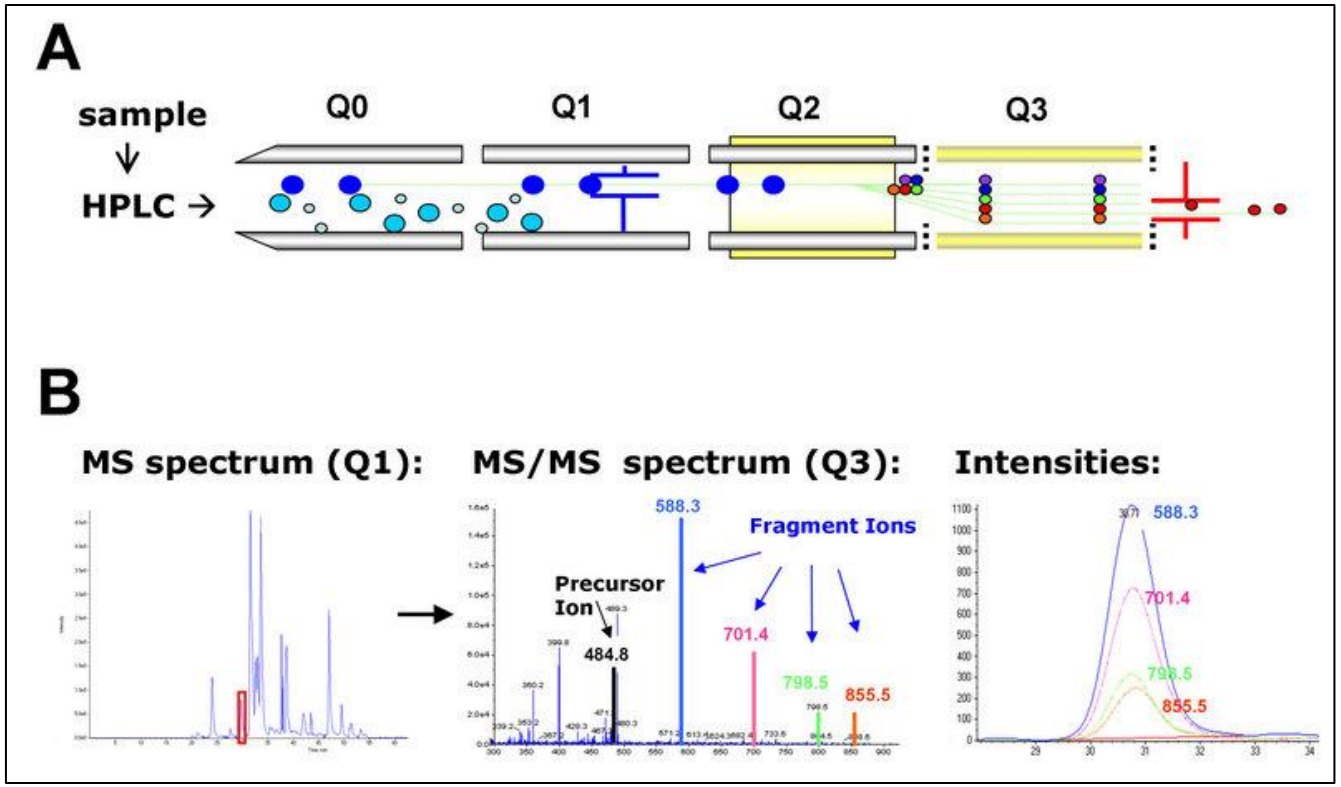
PROB	SPECTRUM	IONS	PEPTIDE	PROTEIN	CALC_MASS
1.0000	01122012_S13_002.01025.01025.2 ST	13/32	K.DQLGKNEEGAPQEGILE.D PA	IPI00024107 +1	1825.8694
1.0000	01122012_S13_002.01336.01336.3 ST	21/72	K.EQVTNVGGAVVTGVTAVAQ.K PA	IPI00024107 +3	1798.9425
1.0000	01122012_S13_002.01257.01257.3 ST	19/76	K.EQVTNVGGAVVTGVTAVAQK.T PA	IPI00024107 +3	1927.0374
1.0000	01122012_S13_002.01008.01008.3 ST	21/56	K.KVPQVSTPTLVEVSR.N PA	IPI00022434 +9	1638.9304
1.0000	01122012_S13_002.01068.01068.2 ST	13/18	K.LVNEVTEFAK.T PA	IPI00022434 +4	1148.6077
0.9903	01122012_S13_002.00626.00626.2 ST	13/22	K.YIC160.03ENQDSISSK.L PA	IPI00022434 +8	1442.6347
0.9330	01122012_S13_002.01269.01269.2 ST	10/26	K.DSTYSLSSTLTLSK.A PA	IPI00550731 +7	1501.7511
0.9233	01122012_S13_002.00894.00894.3 ST	16/32	S.TIC160.03VATIRK257.21.V PA	IPI00165009 +2	1189.7225
0.8894	01122012_S13_002.00862.00862.3 ST	19/48	G.AGSIAAATGFVKK.D PA	IPI00024107 +3	1219.6924
0.8387	01122012_S13_002.01654.01654.3 ST	14/44	R.VSMELGGLAPFI.V PA	IPI00019888 +3	1232.6475
0.7831	01122012_S13_002.00597.00597.3 ST	14/48	K.TLRC160.03PLYRGTLVP.R PA	IPI00152863 +1	1544.8497
0.7616	01122012_S13_002.01253.01253.3 ST	15/36	K.DQNQFISSEP.T PA	IPI00017087	1163.5095
0.7035	01122012_S13_002.01758.01758.3 ST	14/40	R.K259.23LIDC160.03K259.23AEGIP.T PA	IPI00012347 +1	1504.9284
0.6904	01122012_S13_002.01589.01589.3 ST	13/68	M.AENIP246.20ENP246.20LK261.24YLYP246.20DIP246.20K261.24.D PA	IPI00000106 +1	2975.9951
0.6684	01122012_S13_002.01211.01211.2 ST	9/36	K.KWWQSENGLDHVSIRLDLE.A PA	IPI00295437 +3	2324.1549
0.6466	01122012_S13_002.01058.01058.3 ST	24/88	E.KTK257.21EQVTNVGGAVVTGVTAVAQK.T PA	IPI00024107 +3	2413.3912
0.6452	01122012_S13_002.00631.00631.2 ST	11/38	K.VDNALQSGNSQESVTEQDSK.D PA	IPI00550731 +7	2134.9614

TPP protein view



Targeted Proteomics

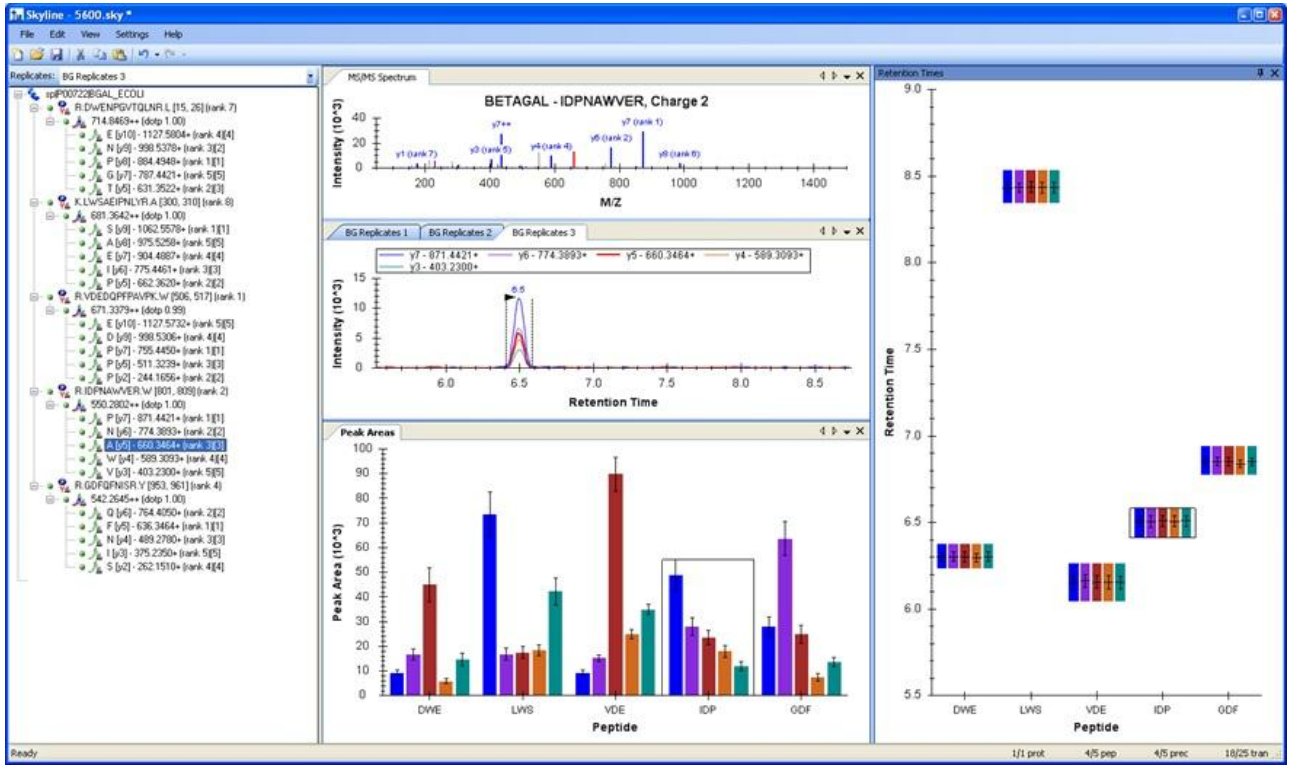
- If you know what you're looking for, you find it more sensitively.



https://www.helmholtz-muenchen.de/?id=12458

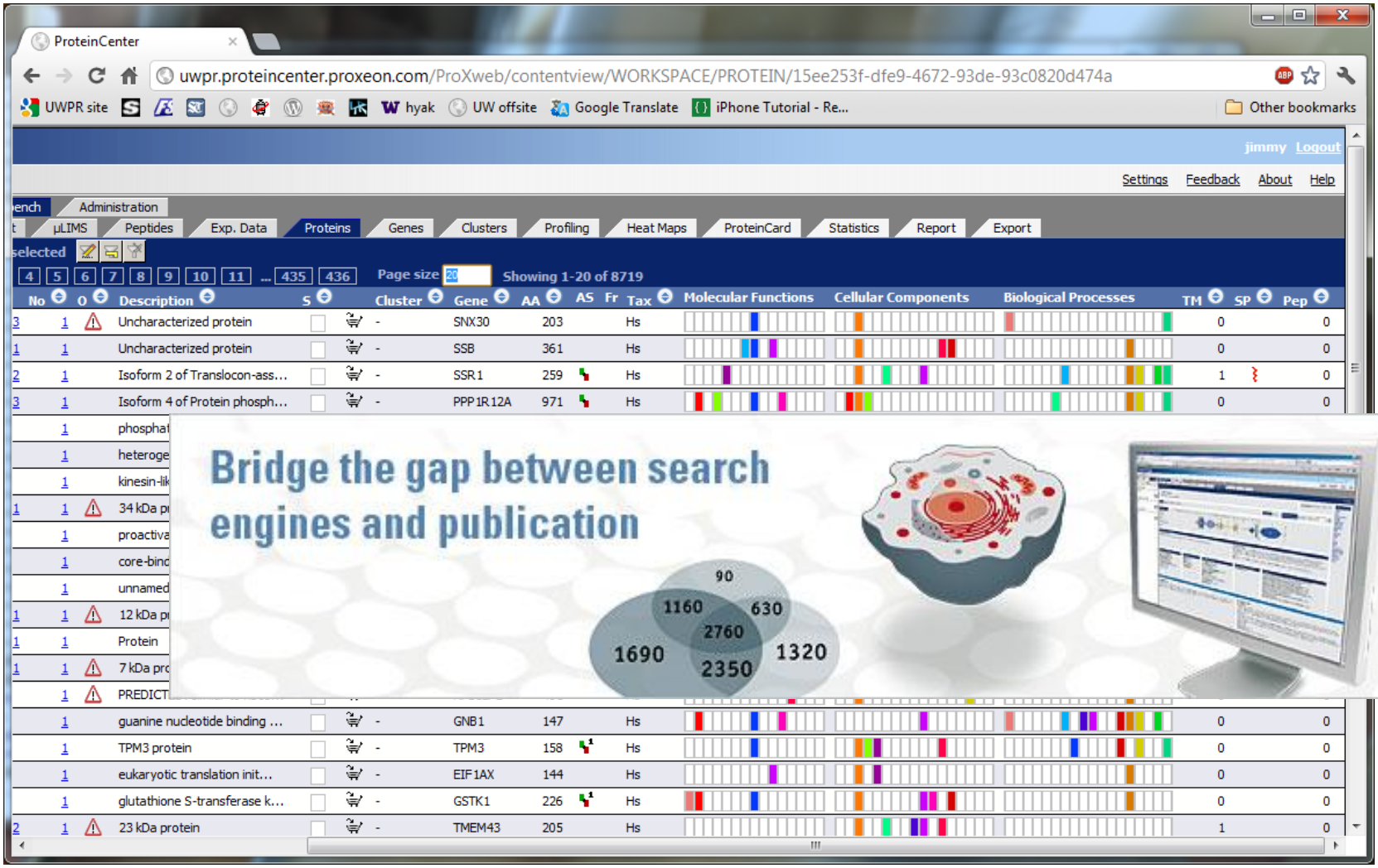
Skyline

- For targeted proteomics & general quantification



Proxeon's ProteinCenter

available at the UWPR



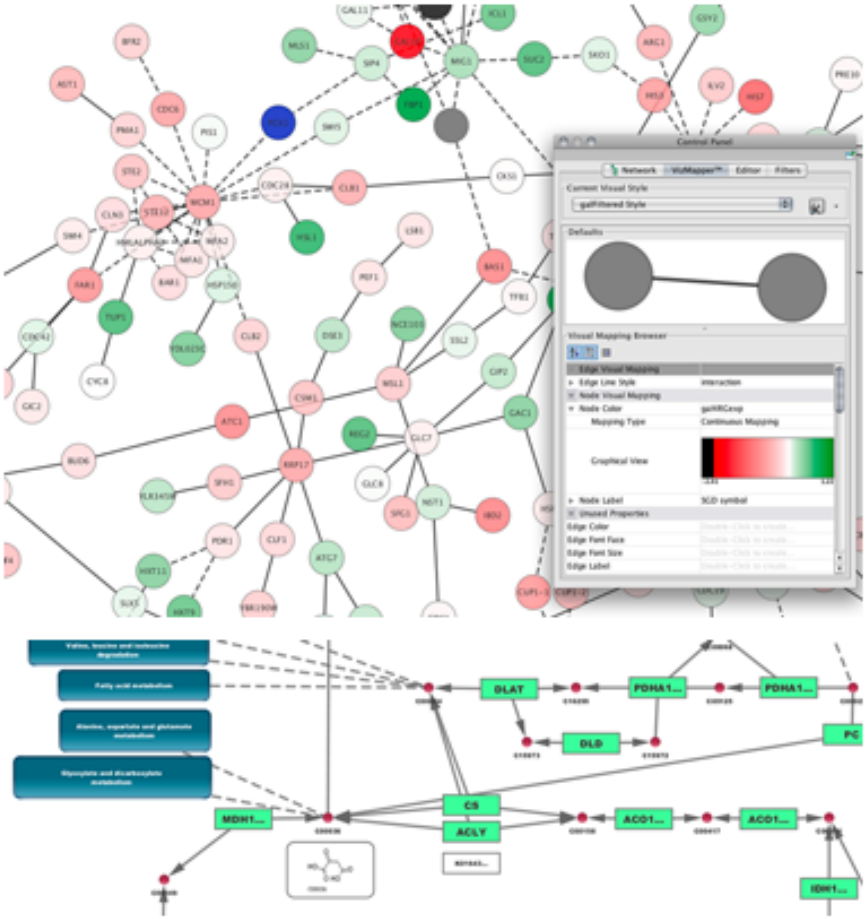
Cytoscape

What Can You Do With Cytoscape?

▾ Biology

Cytoscape supports many use cases in molecular and systems biology, genomics, and proteomics:

- Load molecular and genetic interaction data sets in many formats
- Project and integrate global datasets and functional annotations
- Establish powerful visual mappings across these data
- Perform advanced analysis and modeling using Cytoscape plugins
- Visualize and analyze human-curated pathway datasets such as Reactome or KEGG.



PANTHER

The PANTHER (**P**rotein **AN**alysis **TH**rough Evolutionary **R**elationships) Classification System is a unique resource that **classifies genes by their functions**, using published scientific experimental evidence and evolutionary relationships to predict function even in the absence of direct experimental evidence. Proteins are **classified by expert biologists** according to:

- Gene families and subfamilies, including annotated phylogenetic trees
- Gene Ontology classes: molecular function, biological process, cellular component
- PANTHER Protein Classes
- Pathways, including diagrams