Mascot-based MS Database Searching:
MS Data Analysis & Interpretation

Sheng Zhang

Director, Proteomics & MS Core Facility,
Cornell University

Presented at MS Workshop
October 21, 2009

Contact: sz14@cornell.edu
http://cores.lifesciences.cornell.edu/brcinfo/?f=2

Overview

- **Significance of Bioinformatics tools in Proteomics Studies**
- **Mascot Database Searching Options:**
  - *Methods of database searching*
  - *Practical tips*
  - *Probability based scoring*
  - *Validation*
  - *Quantitation search*
  - *Error tolerance*
- **Interpreting Mascot Results**
  - *Summary reports: filters*
  - *Peptide view report*
  - *Protein view report*
  - *Export search results*

Significances of Bioinformatics Tools

- **Speed! Speed! Speed!**
  - ✓ Automation
- **Higher Confidence**
  - ✓ Statistical algorithms applied automatically
  - ✓ Filters
  - ✓ False Discovery Rate
- **Higher Accuracy/Less Subjective:**
  - ✓ Peak detection from large MS/MS datasets
  - ✓ Large genomic/protein database processing

Database Searching Fundamentals

- Three approaches for using MS data for protein ID

✓ PMF (peptide mass fingerprint)

- matching measured masses of a subset of an enzymatic peptides against theoretically calculated values from database.
- key factors: size of database, mass accuracy and appropriate threshold

✓ Sequence Query

- Peptide masses combined with a piece of sequence, composition and fragment ion data: called sequence tag, require interpretation of spectrum (manual process)

✓ MS/MS Ion Search

- Searching MS/MS data from a single spectrum or complete LC-MS/MS run
- Automated, using software to match lists of fragment ion mass and intensity

Pioneered by John Yates: *J. Am Soc. Mass Spectrom.* 5: 976-989 1994

MS/MS Ions Search Servers & Engines

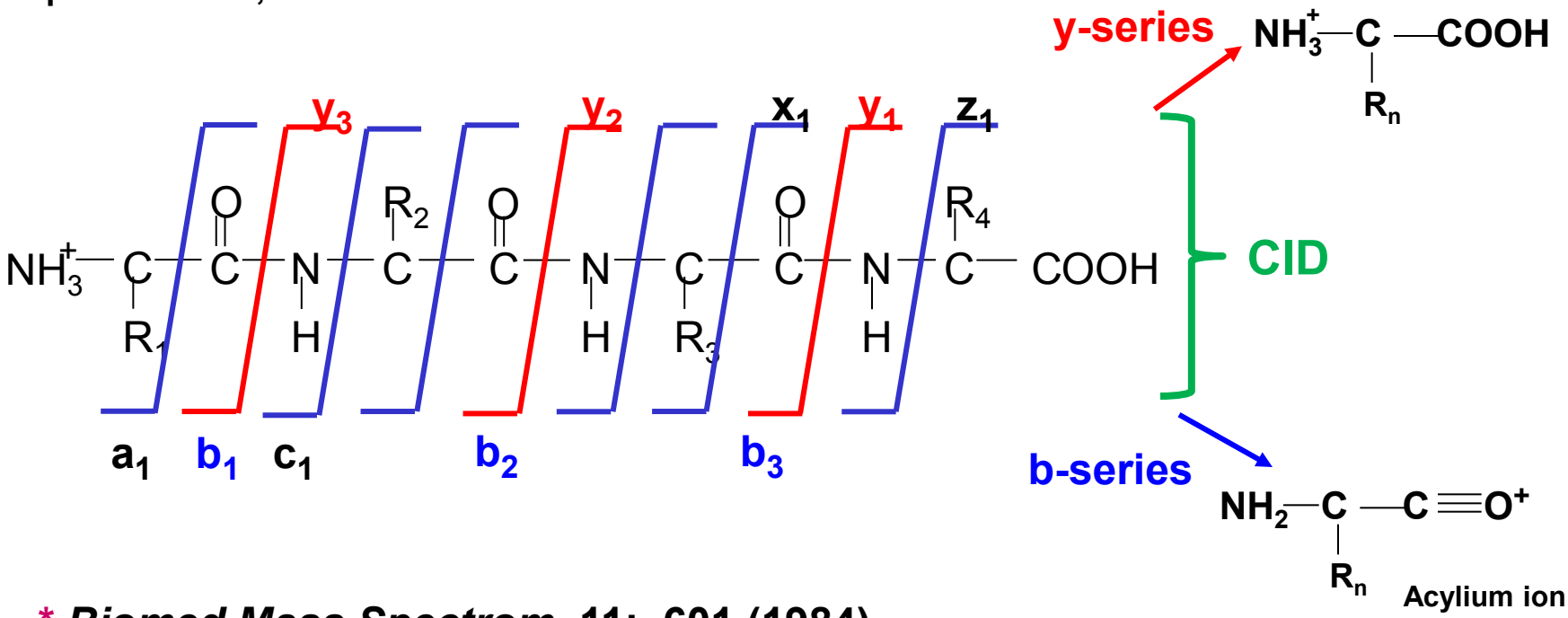
- ✓ Database search engines offered by MS instrument vendors
  - BioWorks (Proteome Discoverer 1.0) from Thermo
  - ProteinPilot 2.1 from Applied Biosystems; GPS Explorer 3.5 for MALDI-TOF/TOF from ABI
  - ProteinLynx Global Server (PLGS) 2.4 from Waters

- ✓ MS/MS Ions Search Servers on the Web
  - **Mascot:**
    http://www.matrixscience.com/search_form_select.html
  - MS-Tag (Protein Prospector):
    http://prospector2.ucsf.edu/prospector/mshome.htm
  - OMSSA:
    http://pubchem.ncbi.nlm.nih.gov/omssa/index.htm
  - PepFrag (Prowl):
    http://prowl.rockefeller.edu/prowl/pepfrag.html
  - X!Tandem (The GPM site):
    http://prowl.rockefeller.edu/tandem/thegpm_tandem.html
  - Phenyx:
    http://www.genebio.com/products/phenyx/
  - Xproteo:
    http://xproteo.com:2698/
  - Peaks:
    http://www.bioinformaticssolutions.com/products/peaks/proteinID.php

Nomenclature for MS/MS Fragmentation of Peptide Ion (How does MS work?)

Bottom-up: CID/CAD#

Top-down: ECD; Middle-down: ETD:



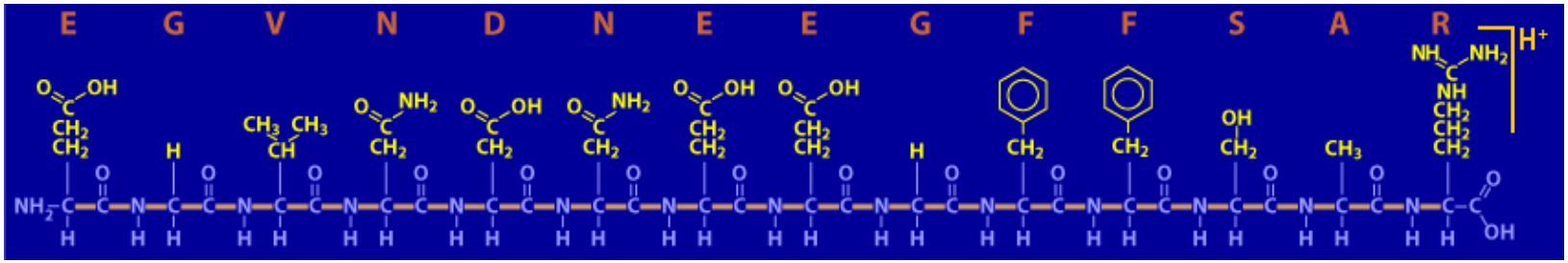
* ***Biomed Mass Spectrom.*** **11: 601 (1984)**

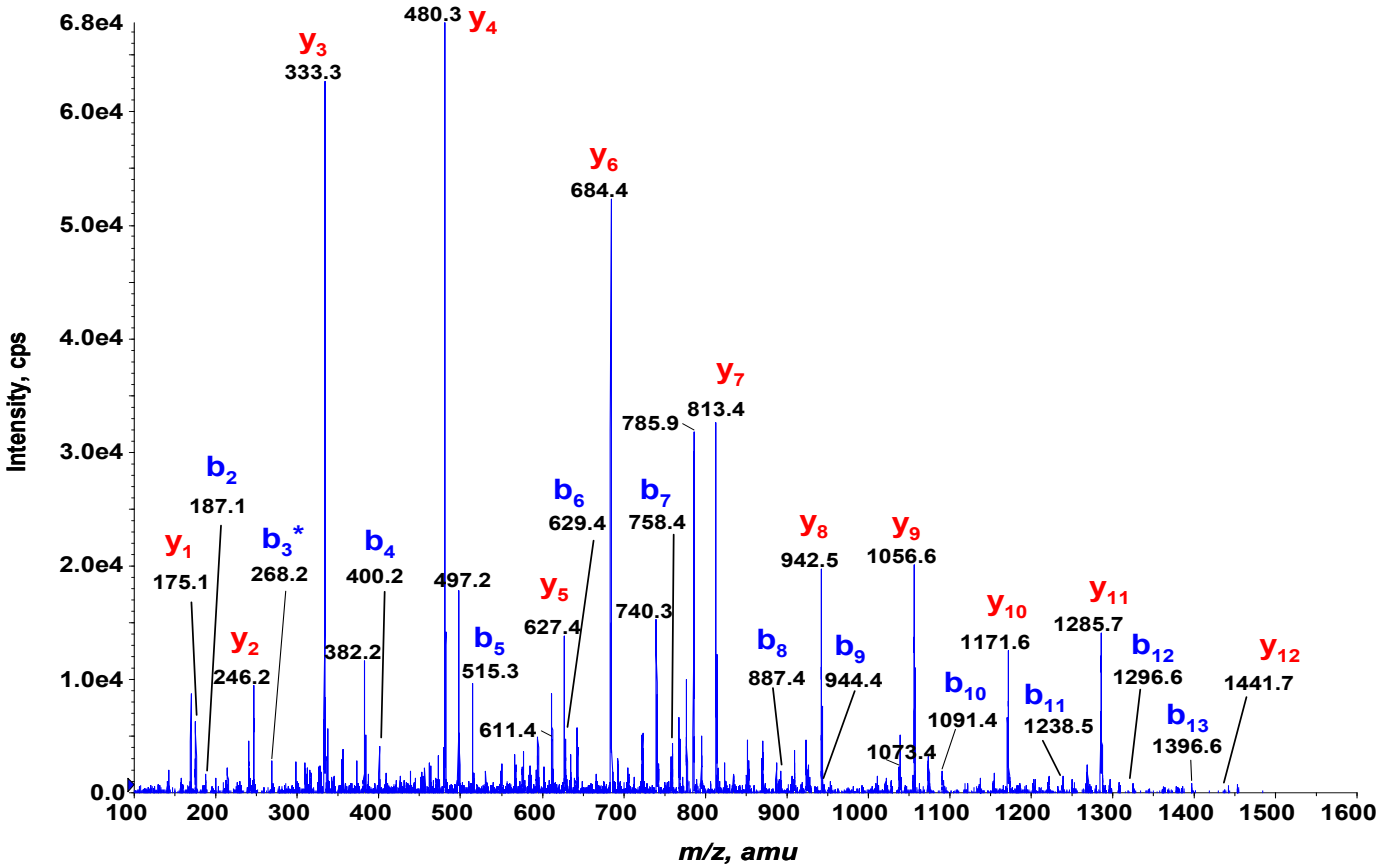
The best paper for interpretation of CID tandem spectra of peptides is:

***Mass Spectrom. Review*** **14: 49-73 (1995)**

ESI Positive CID of Glu-Fibrinopeptide B

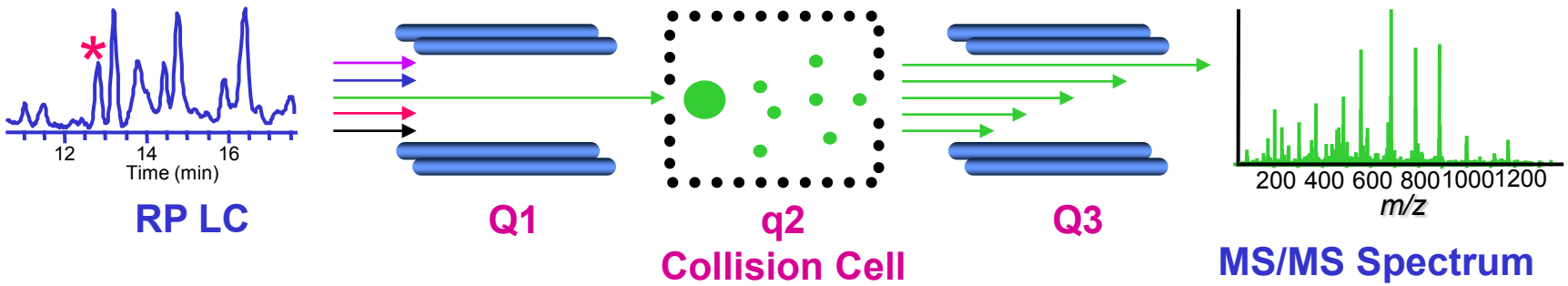
EGVNDNEEGFFSAR

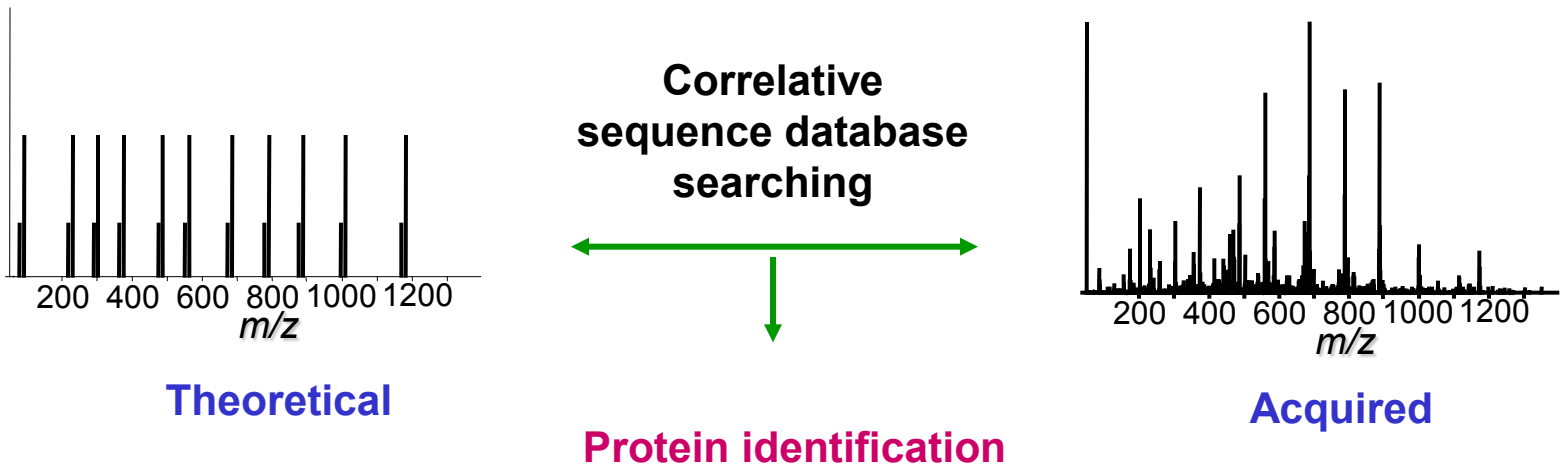




b	y
130.0499	1570.6768
187.0713	1441.6342
286.1397	1384.6128
400.1827	1285.5444
515.2096	1171.5014
629.2525	1056.4745
758.2951	942.4316
887.3377	813.3890
944.3592	684.3464
1091.4276	627.3249
1238.4960	480.2565
1325.5281	333.1881
1396.5652	246.1561
1552.6663	175.1190

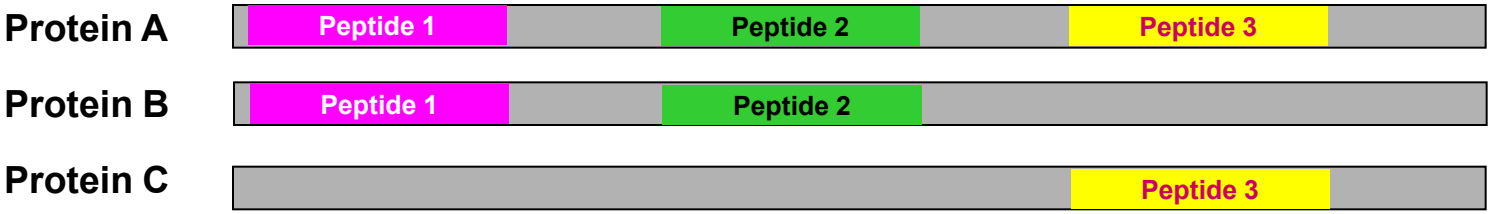
Principle of MS/MS Ion Search





MS/MS Ion Search

- ✓ Easily automated
  - Extremely useful for most proteomics pipelines
- ✓ Can get matches from marginal data
  - Multiple low-confidence score peptides could be accumulated to high confidence probability
- ✓ Search speed is dependent on parameters
  - no enzyme, number of missed cleavage, database size, variable modifications
- ✓ MS/MS is peptide identification
  - Protein identifications by inference
  - Assigning peptide matches to protein hits can be arbitrary: particularly in shotgun analysis



Practical Tips

Mascot > MS/MS Ions Search

MASCOT MS/MS Ions Search

Your name		Email	
Search title			
Database	MSDB		
Taxonomy	All entries		
Enzyme	Trypsin	Allow up to	1 missed cleavages
Fixed modifications	Acetyl (K) Acetyl (N-term) Acetyl (Protein N-term) aldoamine (R) Amidated (C-term)	Variable modifications	Acetyl (K) Acetyl (N-term) Acetyl (Protein N-term) aldoamine (R) Amidated (C-term)
Quantitation	None		
Peptide tol. ±	1.2 Da # ^{13}C 0	MS/MS tol. ±	0.6 Da
Peptide charge	2+	Monoisotopic	Average
Data file	Browse...		
Data format	Mascot generic	Precursor	m/z
Instrument	Default	Error tolerant	
Decoy		Report top	AUTO hits
	Start Search ...		Reset Form

Practical Searching Tips

✓ Select appropriate database and Taxonomy

- Don't cheat yourself: Iteratively adjusting search parameters to get a better score can yield misleading results

✓ Modifications

- Fixed/static modifications cost nothing
- Variable modifications (multiple VM) are expensive
- Use minimum variable modifications

✓ Miss cleavages

- One is generally enough, >1 adds cost of searching time
- Set missed cleavages by inspection of standards

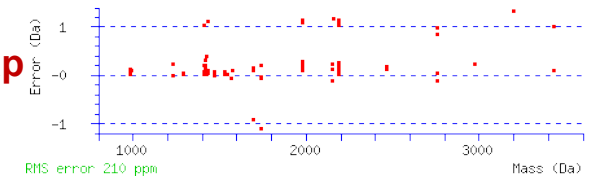
✓ Enzyme

- Understanding cleavage sites of enzyme
- "No enzyme" costs more. Use if strictly necessary

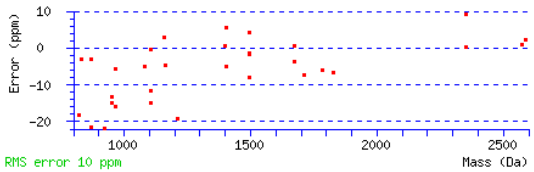
✓ Peptide or MS/MS Tolerance

- Depending on mass accuracy of MS instruments
- Make a reasonable estimate of mass error: 1.2/0.6 Da for Ion Trap; 0.1/0.05 Da for Q-TOF

Ion Trap



Q-TOF

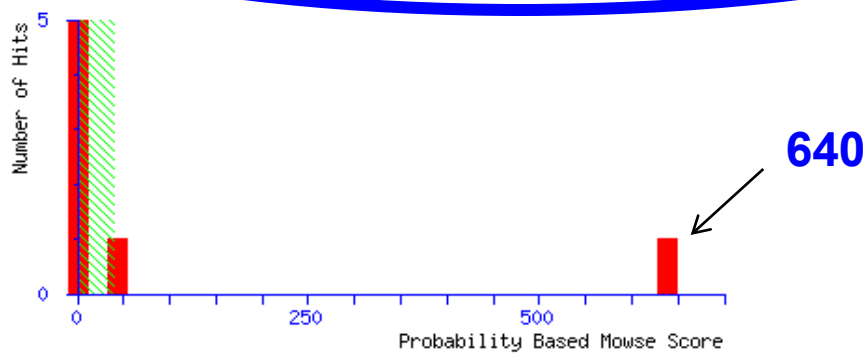


Probability Based Scoring

```
User            : Sheng
Email           : sz14@cornell.edu
Search title    : D:\Analyst Data\Projects\Nano16Jul09\2009_07_16\Data\28Sept09\StdPaVa_11.wiff (sample number 1)
MS data file    : C:\DOCUME~1\mslab\LOCALS~1\Temp\mas142.tmp
Database        : SwissProt 20080728 (392667 sequences; 141217034 residues)
Taxonomy        : Mammalia (mammals) (63164 sequences)
Timestamp       : 30 Sep [illegible]
Protein hits    : ALBU_BOVIN Serum albumin OS=Bos taurus [illegible]ALB PE=1 SV=4
```

Probability Based Mowse Score

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 39 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.



Peptide Summary Report

Format As	Peptide Summary		Help
	Significance threshold p< 0.05	Max. number of hits AUTO	
	Standard scoring ○ MudPIT scoring ◉	Ions score or expect cut-off 30	Show sub-sets 0
	Show pop-ups ◉ Suppress pop-ups ○	Sort unassigned Decreasing Score	Require bold red ☑

Understanding Probability Based Scoring

- ✓ Algorithm is based on probability that the observed match between the experimental data, and mass values calculated from a candidate peptide or protein sequence, is a random event. The real match, which is not a random event, then has a very low probability.
- ✓ Probability based scoring enables standard statistical tests to be applied to results.
- ✓ The calculated probability may be converted into a score
  Ion Score = -10 * Log (P) P = probability. If P = 0.00001, Score = 50
- ✓ The calculated probability may be reported as an expectation value: how often you would expect to get a match this good or better purely by chance; or how hard you try to get a match.
  If match with expectation value: 0.001, meaning one match from 1000 trials, it is definitely not random.
- ✓ Probabilities = multiply the acceptable false positive rate by the number of trials: For an MS/MS search, the number of trials is the number of candidate peptides.

Probability Based Ion Score & Expectation Value

1. ALBU_BOVIN Mass: 69248 Score: 639 Queries matched: 30 emPAI: 2.66

Serum albumin OS=Bos taurus GN=ALB PE=1 SV=4

☐ Check to include this hit in error tolerant search or archive report

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	3	379.7279	757.4413	757.4156	0.0257	0	55	0.0019	1	K.GACLLPK.I + Carbamidomethyl (C)
☑	35	395.2261	788.4376	788.4644	-0.0267	0	49	0.0091	1	K.LVTDLTK.V
☑	146	443.7055	885.3964	885.4080	-0.0116	0	40	0.05	1	K.DDSPDLPK.L
☑	182	461.7628	921.5111	921.4807	0.0303	0	38	0.091	1	K.AEFVEVTK.L
☑	186	464.2850	926.5555	926.4861	0.0694	0	41	0.04	1	K.YLYEIAR.R
☑	211	507.8422	1013.6699	1013.6121	0.0578	0	51	0.0041	1	K.QTALVELLK.H
☑	229	536.7995	1071.5845	1071.5019	0.0827	0	44	0.025	1	K.SHCIAEVEK.D + Carbamidomethyl (C)
☑	238	554.2835	1106.5524	1106.5066	0.0457	0	52	0.0034	1	K.EACFAVEGPK.L + Carbamidomethyl (C)
☑	255	569.7624	1137.5103	1137.4907	0.0196	0	64	0.00021	1	K.CCTESLVNR.R + 2 Carbamidomethyl (C)
☑	257	571.8939	1141.7733	1141.7070	0.0663	1	57	0.001	1	K.KQTALVELLK.H
☑	283	582.1292	1162.2439	1162.6234	-0.3795	0	61	0.00041	1	K.LVNELTEFAK.T
☑	340	417.2403	1248.6990	1248.6139	0.0852	1	33	0.23	1	R.FKDLGEEHFK.G
☑	393	435.7988	1304.3744	1304.7088	-0.3344	0	(54)	0.0018	1	K.HLVDEPQNLIK.Q
☑	394	653.2737	1304.5328	1304.7088	-0.1760	0	(40)	0.047	1	K.HLVDEPQNLIK.Q
☑	395	653.3822	1304.7498	1304.7088	0.0409	0	76	1.3e-05	1	K.HLVDEPQNLIK.Q
☑	459	473.9127	1418.7163	1418.6864	0.0299	0	54	0.0017	1	K.SLHTLFGDELCK.V + Carbamidomethyl (C)
☑	470	480.6425	1438.9057	1438.8045	0.1012	1	58	0.00063	1	R.RHPEYAVSVLLR.L
☑	472	722.3332	1442.6519	1442.6347	0.0172	0	94	1.9e-07	1	K.YICDNQDTISSK.L + Carbamidomethyl (C)
☑	494	751.7968	1501.5791	1501.6065	-0.0273	0	51	0.0036	1	K.EYEATLEECCAK.D + 2 Carbamidomethyl (C)
☑	506	383.9706	1531.8534	1531.7738	0.0796	1	(31)	0.33	1	K.LKECCDKPLLEK.S + 2 Carbamidomethyl (C)
☑	507	511.6304	1531.8693	1531.7738	0.0955	1	52	0.0028	1	K.LKECCDKPLLEK.S + 2 Carbamidomethyl (C)
☑	526	518.9133	1553.7180	1553.6457	0.0724	0	36	0.1	1	K.DDPHACYSTVFDK.L + Carbamidomethyl (C)
☑	549	526.2357	1575.6853	1575.7603	-0.0750	0	48	0.0077	1	K.LKPDPNTLCDEFK.A + Carbamidomethyl (C)
☑	584	547.3371	1638.9896	1638.9305	0.0591	1	86	1.1e-06	1	R.KVPQVSTPTLVEVSR.S
☑	676	874.3116	1746.6086	1746.6978	-0.0892	0	82	2.5e-06	1	K.YNGVFQECCQAEDK.G + 2 Carbamidomethyl (C)
☑	686	583.9370	1748.7892	1748.6553	0.1339	0	40	0.042	1	K.ECCHGDLLECADDR.A + 3 Carbamidomethyl (C)
☑	733	627.6342	1879.8808	1879.9138	-0.0330	0	43	0.019	1	R.RPCFSALTPDETYVPK.A + Carbamidomethyl (C)
☑	742	476.2425	1900.9409	1900.8625	0.0784	1	44	0.013	1	R.NECFLSHKDDSPDLPK.L + Carbamidomethyl (C)
☑	750	643.2715	1926.7925	1926.7910	0.0015	1	66	8.9e-05	1	K.CCAADDKEACFAVEGPK.L + 3 Carbamidomethyl (C)

Validation

- ✓ Repeat the search, using identical search parameters, against a database in which the sequences have been reversed or randomized Elias, J. E., et al., *Nature Methods* 2 667-675 (2005).
- ✓ Reversed entries sufficient for MS/MS with enzyme
- ✓ Randomized entries required for MS/MS without enzyme
- ✓ Direct estimate for FDR present in the results from the real database
- ✓ An excellent validation method for MS/MS searches of large data sets. It is not as useful for a search of a small number of spectra, because the numbers are too small to give an accurate estimate.

	NCBInr	Decoy	False discovery rate
Peptide matches above identity threshold	956	9	0.94 %
Peptide matches above homology or identity threshold	1029	22	2.14 %

Quantitation Searching: emPAI Protocol

- ✓ The **E**xponentially **M**odified **P**rotein **A**bundance **I**ndex (emPAI) offers approximate, label-free, relative quantitation of the proteins in a mixture based on protein coverage by the peptide matches in a database search result.
  Ishihama, Y., et al., *Molecular & Cellular Proteomics 4 1265-1272 (2005)*
- ✓ The information required for emPAI is always present in a search result, and there are no parameter settings, as long as the MS/MS search contains at least 100 spectra.
- ✓ $emPAI = 10^{PAI} - 1$; $PAI = N_{observed} / N_{observable}$; $\% \ molar = emPAI / \Sigma[emPAI] * 100$
- ✓ The count of observed peptides only includes peptide matches with scores at or above the homology threshold
- ✓ A calculated estimate of the number of observable peptides based on the protein mass, the average amino acid composition of the database, and the enzyme specificity via *in silico* digests.

Relative Quantitation: emPAI Protocol

Format As | Select Summary (protein hits) | Help

Significance threshold p< 0.01 | Max. number of hits AUTO

Standard scoring ○ MudPIT scoring ◉ | Ions score or expect cut-off 25 | Show sub-sets 0

Show pop-ups ◉ Suppress pop-ups ○ | Sort unassigned Decreasing Score | Require bold red ☑

1. TRFL_BOVIN Mass: 80002 Score: 404 Queries matched: 34 emPAI: 1.32
Lactotransferrin OS=Bos taurus GN=LTF PE=1 SV=2

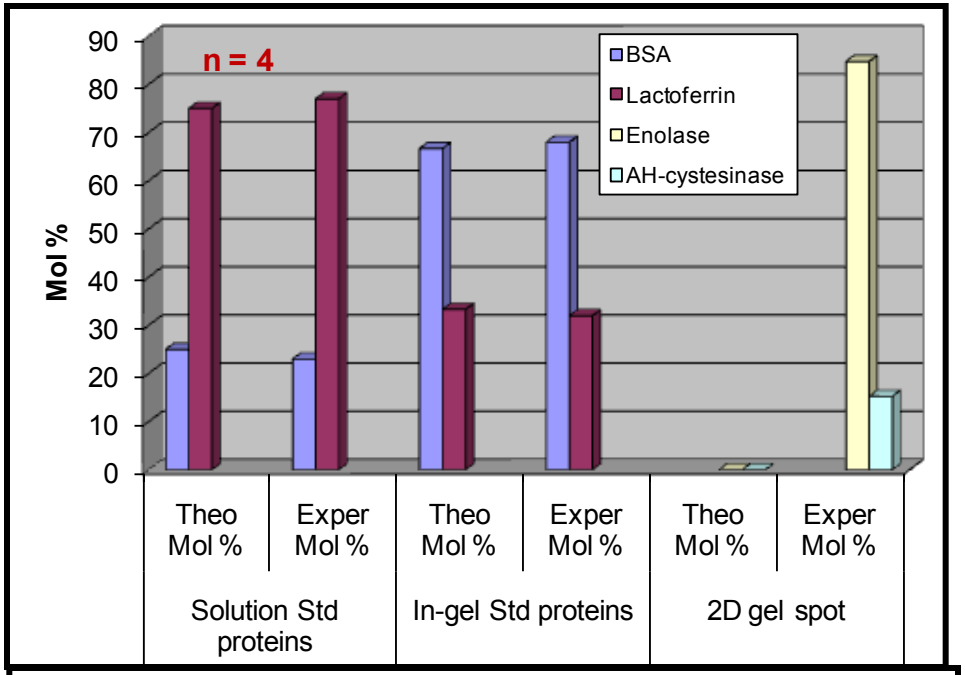
Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
2	374.2787	746.5428	746.3963	0.1465	0	35	0.21	1	R.APVDAFK.E
50	439.8468	877.6790	877.4658	0.2133	0	43	0.025	1	K.DSALGFLR.I
70	474.2959	946.5772	946.4906	0.0866	0	43	0.026	1	R.LLCLDGTR.K 71
84	490.2879	978.5612	978.4957	0.0656	0	37	0.084	1	R.AFALECIR.A
107	537.3104	1072.6062	1072.5699	0.0363	0	60	0.00057	1	K.LGAPSITCVR.R 108
112	540.8253	1079.6360	1079.5611	0.0749	0	57	0.00079	1	K.VDSAVLGSR.Y 110 111

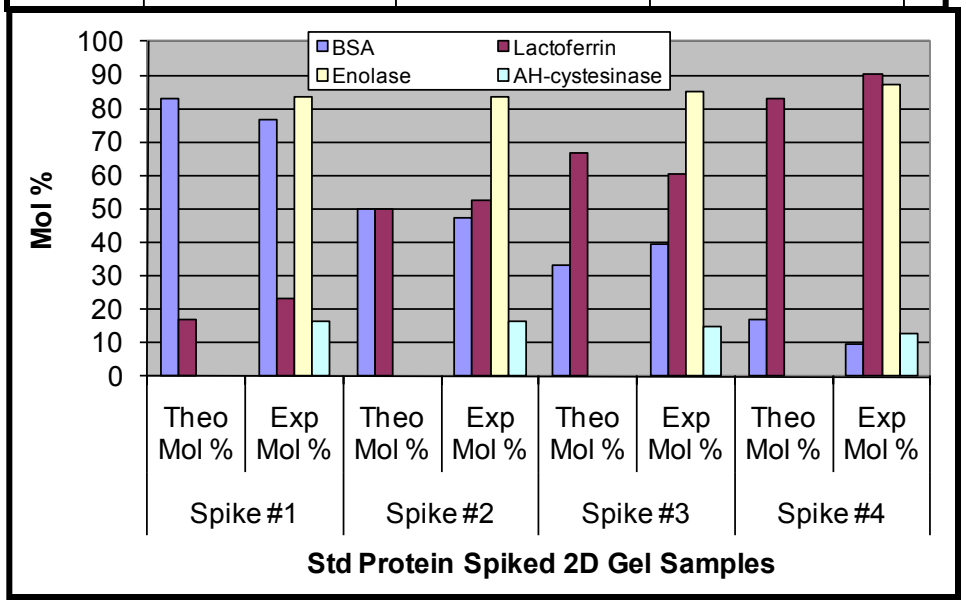
2. ALBU_BOVIN Mass: 71244 Score: 157 Queries matched: 13 emPAI: 0.37
Serum albumin OS=Bos taurus GN=ALB PE=1 SV=4

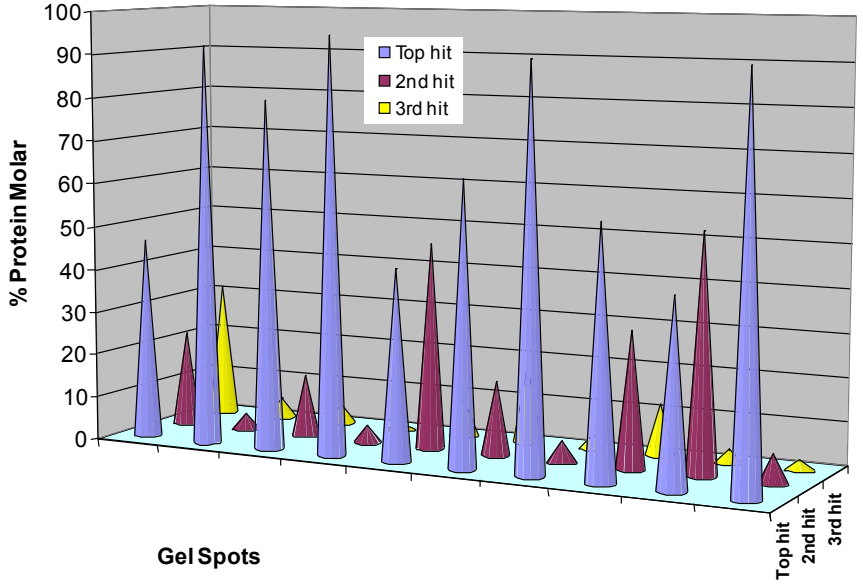
Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
3	379.7842	757.5539	757.4156	0.1383	0	36	0.15	1	K.GACLLPK.I
16	395.3147	788.6148	788.4644	0.1504	0	49	0.0088	1	K.LVTDLTK.V
58	449.7627	897.5109	897.4742	0.0367	0	33	0.21	1	R.LCVLHEK.T
65	462.2817	922.5489	921.4807	1.0682	0	36	0.11	2	K.AEFVEVTK.L
67	464.3096	926.6047	926.4861	0.1185	0	36	0.12	1	K.YLYEIAR.R
95	507.8840	1013.7535	1013.6121	0.1414	0	55	0.0016	1	K.QTALVELLK.H
123	569.7853	1137.5561	1137.4907	0.0654	0	59	0.00057	1	K.CCTESLVNR.R
129	582.3986	1162.7826	1162.6234	0.1593	0	48	0.0063	1	K.LVNELTEFAK.T
152	417.3001	1248.8785	1248.6139	0.2647	1	37	0.088	1	R.FKDLGEEHFK.G
162	653.4246	1304.8346	1304.7088	0.1258	0	48	0.0068	1	K.HLVDEPQNLIK.Q

Proteins	emPAI	% Molar	Expected ratio
Lactotransferrin	1.32	78.1	75%
BSA	0.37	21.9	25%
Σ[emPAI]	**1.69**	**100**	

Quantitation Searching: emPAI Protocol







emPAI is particularly useful for determining the relative concentration of multiple proteins Identified in single samples (e.g. 2D gel spots)

Yang Y. *et al.*, *Electrophoresis 2007 28, 2080-94*

emPAI Analysis for a Mouse Brain Sample

Top 30 Proteins

No.	Prot_acc	Prot_desc	emPAI	Relative Protein Conc. % Molar
3	HBB1_MOUSE	Hemoglobin subunit beta-1 (Hemoglobin beta-1 chain) (Beta-1-globin) (Hemoglobin beta-majc	5.93	**19.859**
15	HBA_MOUSE	Hemoglobin subunit alpha (Hemoglobin alpha chain) (Alpha-globin) - Mus musculus (Mouse)	1.76	**5.894**
30	ATP5E_MOUSE	ATP synthase epsilon chain, mitochondrial (EC 3.6.3.14) - Mus musculus (Mouse)	1.60	**5.358**
4	MDHM_MOUSE	Malate dehydrogenase, mitochondrial precursor (EC 1.1.1.37) - Mus musculus (Mouse)	1.23	**4.119**
1	ATPB_MOUSE	ATP synthase subunit beta, mitochondrial precursor (EC 3.6.3.14) - Mus musculus (Mouse)	1.22	**4.086**
10	ACTB_MOUSE	Actin, cytoplasmic 1 (Beta-actin) - Mus musculus (Mouse)	1.14	**3.818**
29	ATP5I_MOUSE	ATP synthase e chain, mitochondrial (EC 3.6.3.14) - Mus musculus (Mouse)	1.04	**3.483**
2	ATPA_MOUSE	ATP synthase subunit alpha, mitochondrial precursor (EC 3.6.3.14) - Mus musculus (Mouse)	1.01	**3.382**
20	ATP5H_MOUSE	ATP synthase D chain, mitochondrial (EC 3.6.3.14) - Mus musculus (Mouse)	0.94	**3.148**
28	NDUA4_MOUSE	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 4 (EC 1.6.5.3) (EC 1.6.99.3)	0.89	**2.981**
7	G3P_MOUSE	Glyceraldehyde-3-phosphate dehydrogenase (EC 1.2.1.12) (GAPDH) - Mus musculus (Mous	0.86	**2.880**
22	MBP_MOUSE	Myelin basic protein (MBP) (Myelin A1 protein) - Mus musculus (Mouse)	0.78	**2.612**
24	COX5A_MOUSE	Cytochrome c oxidase subunit 5A, mitochondrial precursor (EC 1.9.3.1) (Cytochrome c oxida	0.78	**2.612**
9	TBA2_MOUSE	Tubulin alpha-2 chain (Alpha-tubulin 2) (Alpha-tubulin isotype M-alpha-2) - Mus musculus (Mc	0.77	**2.579**
12	TMOD2_MOUSE	Tropomodulin-2 (Neuronal tropomodulin) (N-Tmod) - Mus musculus (Mouse)	0.76	**2.545**
26	BASP_MOUSE	Brain acid soluble protein 1 (BASP1 protein) (Neuronal axonal membrane protein NAP-22) (2	0.76	**2.545**
18	CALM_MOUSE	Calmodulin (CaM) - Mus musculus (Mouse)	0.73	**2.445**
27	CH10_MOUSE	10 kDa heat shock protein, mitochondrial (Hsp10) (10 kDa chaperonin) (CPN10) - Mus musc	0.73	**2.445**
21	ATPO_MOUSE	ATP synthase O subunit, mitochondrial precursor (EC 3.6.3.14) (Oligomycin sensitivity confe	0.71	**2.378**
5	TBB2C_MOUSE	Tubulin beta-2C chain - Mus musculus (Mouse)	0.67	**2.244**
6	TBB5_MOUSE	Tubulin beta-5 chain - Mus musculus (Mouse)	0.67	**2.244**
8	TBA1_MOUSE	Tubulin alpha-1 chain (Alpha-tubulin 1) (Alpha-tubulin isotype M-alpha-1) - Mus musculus (Mc	0.67	**2.244**
11	CN37_MOUSE	2~,3~-cyclic-nucleotide 3~-phosphodiesterase (EC 3.1.4.37) (CNP) (CNPase) - Mus musculu	0.61	**2.043**
13	KCRU_MOUSE	Creatine kinase, ubiquitous mitochondrial precursor (EC 2.7.3.2) (U-MtCK) (Mia-CK) (Acidic-t	0.61	**2.043**
17	TPIS_MOUSE	Triosephosphate isomerase (EC 5.3.1.1) (TIM) (Triose-phosphate isomerase) - Mus musculus	0.60	**2.009**
23	PRDX5_MOUSE	Peroxiredoxin-5, mitochondrial precursor (EC 1.11.1.15) (Prx-V) (Peroxisomal antioxidant en	0.53	**1.775**
19	ALDOA_MOUSE	Fructose-bisphosphate aldolase A (EC 4.1.2.13) (Muscle-type aldolase) (Aldolase 1) - Mus n	0.50	**1.674**
16	TBA4_MOUSE	Tubulin alpha-4 chain (Alpha-tubulin 4) (Alpha-tubulin isotype M-alpha-4) - Mus musculus (Mc	0.47	**1.574**
25	PHB2_MOUSE	Prohibitin-2 (B-cell receptor-associated protein BAP37) (Repressor of estrogen receptor activ	0.46	**1.541**
14	DPYL2_MOUSE	Dihydropyrimidinase-related protein 2 (DRP-2) (ULIP 2 protein) - Mus musculus (Mouse)	0.43	**1.440**
			29.86	**100**

Searches fail

- General reasons for failing to get a hit:
  - ✓ Poor quality of MS/MS data
  - ✓ Inappropriate parameter settings (enzyme, wrong charge state, underestimated mass measurement error) for the database search
  - ✓ No targeted proteins/peptides in the database
  - ✓ Unsuspected chemical and other modifications
  - ✓ In source fragmentation or high-energy fragmentation

Mascot Quantitation Searching

MASCOT MS/MS Ions Search

Your name		Email	
Search title			
Database	NCBInr		
Taxonomy	All entries		
Enzyme	Trypsin	Allow up to	1 missed cleavages
Fixed modifications	Acetyl (K) Acetyl (N-term) Acetyl (Protein N-term) Amidated (C-term) Amidated (Protein C-term)	Variable modifications	Acetyl (K) Acetyl (N-term) Acetyl (Protein N-term) Amidated (C-term) Amidated (Protein C-term)
Quantitation	None		
Peptide tol. ±		MS/MS tol. ±	0.6 Da
Peptide charge		Monoisotopic	Average
Data file	owse...		
Data format		Precursor	m/z
Instrument		Error tolerant	☐
Decoy		Report top	AUTO hits
			Reset Form

None
iTRAQ 4plex
iTRAQ 8plex
18O corrected multiplex
SILAC K+6 R+6 multiplex
TMT 6plex
ICAT ABI Cleavable [MD]
ICPL duplex pre-digest [MD]
ICPL duplex post-digest [MD]
SILAC K+6 R+10 [MD]
18O corrected [MD]
15N Metabolic [MD]
iTRAQ 4plex-115-median Norm
iTRAQ 4plex-115-no Norm
iTRAQ 4plex-114-no Norm
iTRAQ 4plex-114-Mediam Norm

Quantitation Searching for ITRAQ Labeling

```
Email                    : sz14@cornell.edu
Search title             : iTRAQ Std
MS data file             : X:\huiming\PKL\09OCT09Ecoli\22Jan09_iTRAQ_DDAcs4a.pkl
Database                 : SwissProt 20080728 (392667 sequences; 141217034 residues)
Quantitation             : iTRAQ 4plex   method details
                         : Applied Biosystems iTRAQ(TM) 4-plex reagent
Timestamp                : 14 Oct 2009 at 02:43:30 GMT
Enzyme                   : Trypsin
Fixed modifications      : iTRAQ4plex (K),iTRAQ4plex (N-term),Methylthio (C),Methylthio (C),iTRAQ4plex (N-term),iTRAQ4plex (K)
Variable modifications   : iTRAQ4plex (Y),Oxidation (M)
Mass values              : Monoisotopic
Protein Mass             : Unrestricted
Peptide Mass Tolerance   : ± 0.2 Da
Fragment Mass Tolerance  : ± 0.1 Da
Max Missed Cleavages     : 1
Instrument type          : Default
Number of queries        : 363
Protein hits             :
```

Quantitation based on the relative intensities of fragment peaks at fixed m/z values within an MS/MS spectrum.

115/114	116/114	117/114	
0.048	0.025	2.175	ALBU_BOVIN Serum albumin OS=Bos taurus GN=ALB PE=1 SV=4
---	---	2.052	TRFE_HUMAN Serotransferrin OS=Homo sapiens GN=TF PE=1 SV=2
---	---	2.180	LACB_BOVIN Beta-lactoglobulin OS=Bos taurus GN=LGB PE=1 SV=3
---	---	2.132	LYSC_CHICK Lysozyme C OS=Gallus gallus GN=LYZ PE=1 SV=1
---	---	2.123	LACB_OVIMU Beta-lactoglobulin OS=Ovis orientalis musimon GN=LGB PE=1 SV=1
---	---	2.132	BGAL_ECOLI Beta-galactosidase OS=Escherichia coli (strain K12) GN=lacZ PE
---	---	1.933	TRY1_BOVIN Cationic trypsin (Fragment) OS=Bos taurus PE=1 SV=2
---	---	---	TRFL_HORSE Lactotransferrin (Fragment) OS=Equus caballus GN=LTF PE=1 SV=1

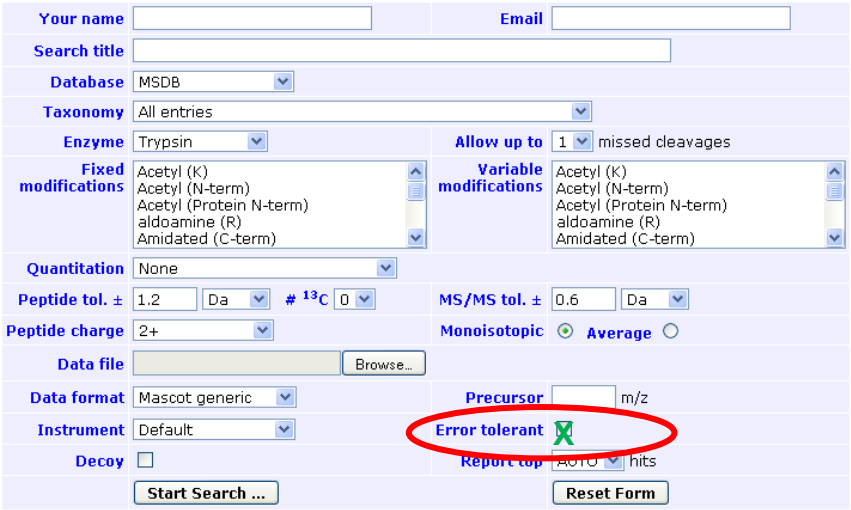
Select Summary Report

Format As	Select Summary (protein hits)		Help
	Significance threshold p< 0.05	Max. number of hits AUTO	
	Standard scoring ○ MudPIT scoring ◉	Ions score or expect cut-off 45	Show sub-sets 0
	Show pop-ups ◉ Suppress pop-ups ○	Sort unassigned Decreasing Score	Require bold red ☑
	Protein ratio type Weighted	Normalisation None	Help
	Report peptide ratios ☑	Outlier removal Automatic	Min. # peptides 2
	Min. precursor charge 1	Peptide threshold At least homology	0.05

Error Tolerance Search

✓ Address unsuccessful hits due to the following three reasons:

- Enzyme non-specificity
- Unsuspected modifications
- No sequence in database



✓ Automatic error tolerance search:

- Normal first-pass search first
- Auto tolerance search as second-pass against the databases entries found above
- Selected enzyme becomes semi-specific
- Miss cleavage increases by 1
- Modifications are tested serially
- AA substitution tested serially
- Only one of modifications/substitutions is allowed per peptide

✓ Manual error tolerance search:

- Normal first-pass search first. Manual tolerance search as second-pass against the results report used as databases entries.
- No enzyme specificity combined with modifications or AA substitutions
- “Junk” matches will be higher than auto error tolerance search

Automatic Error Tolerance Search Results

MASCOT MS/MS Ions Search

Your name	Sheng	Email	sz14@cornell.edu
Search title	Auto error tolerance test		
Database	SwissProt		
Taxonomy	. . . . . Mammalia (mammals)		
Enzyme	Trypsin	Allow up to	1 missed cleavages
Fixed modifications	Acetyl (N-term) Acetyl (Protein N-term) Amidated (C-term) Amidated (Protein C-term) Ammonia-loss (N-term C)	Variable modifications	Biotin (K) Biotin (N-term) Carbamidomethyl (C) Carbamyl (K) Carbamyl (N-term)
Quantitation	None		
Peptide tol. ±	1.5 Da # ^{13}C 0	MS/MS tol. ±	0.6 Da
Peptide charge	2+ and 3+	Monoisotopic	Average
Data file	X:\Sheng\mas504A.tmp Browse...		
Data format	Mascot generic	Precursor	m/z
Instrument	Default	Error tolerant	☑
Decoy	☐	Report top	AUTO hits
	Start Search ...		Reset Form

MATRIX SCIENCE Mascot Search

Licensed to: Cornell University, Ithaca, (1 processor).
Finished uploading search details and file...
Searching....
Search stage: standard search
..10% complete
...15% complete
....20% complete
.....25% complete
.....30% complete
.....40% complete
.....50% complete
.....60% complete
.....70% complete
.....80% complete
.....90% complete
.....100% complete
standard search complete
Search stage: error-tolerant search
.10% complete
..20% complete
...30% complete
....40% complete
.....45% complete
.....50% complete
.....Writing results file.
error-tolerant search complete
392667 sequences and 141217034 residues checked.

Automatic Error Tolerance Search Results

1. ALBU_BOVIN Mass: 69248 Score: 2238 Queries matched: 94 emPAI: 1.30
Serum albumin OS=Bos taurus GN=ALB PE=1 SV=4

☐ Check to include this hit in error tolerant search or archive report

BSA Succinylation

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
7	386.2956	770.5767	771.4313	-0.8545	0	48		1	K.GACLLPK.I + [+71.0371 at C3]
22	395.2537	788.4928	788.4644	0.0284	0	48	0.011	1	K.LVTDLTK.V
75	443.7565	885.4985	885.4080	0.0905	0	54	0.0022	1	K.DDSPDLPK.L
175	507.8259	1013.6373	1013.6121	0.0253	0	(51)	0.0042	1	K.QTALVELLK.H
176	507.8342	1013.6539	1013.6121	0.0418	0	51	0.004	1	K.QTALVELLK.H
177	507.8389	1013.6632	1013.6121	0.0512	0	(50)	0.0052	1	K.QTALVELLK.H
178	507.8480	1013.6815	1013.6121	0.0695	0	(49)	0.0063	1	K.QTALVELLK.H
264	560.3150	1118.6154	1118.5641	0.0512	0	(54)		1	K.LVNELTEFAK.T + [-44.0592 at F8]
267	561.2785	1120.5425	1120.5223	0.0202	0	50		1	K.EACFAVEGPK.L + [+71.0371 at C3]
299	582.2787	1162.5429	1162.6234	-0.0805	0	(56)	0.0011	1	K.LVNELTEFAK.T
300	582.3039	1162.5932	1162.6234	-0.0301	0	(56)	0.0011	1	K.LVNELTEFAK.T
301	582.3169	1162.6192	1162.6234	-0.0041	0	60	0.00048	1	K.LVNELTEFAK.T
302	582.3302	1162.6458	1162.6234	0.0224	0	(59)	0.0006	1	K.LVNELTEFAK.T
305	582.8404	1163.6663	1163.6074	0.0589	0	(57)		1	K.LVNELTEFAK.T + [+0.9840 at N3]
367	621.8409	1241.6672	1241.7231	-0.0559	1	(60)		1	K.KQTALVELLK.H + [+100.0160 at N-term K]
369	621.8558	1241.6970	1241.7231	-0.0261	1	(52)		1	K.KQTALVELLK.H + [+100.0160 at N-term K]
370	621.8588	1241.7030	1241.7231	-0.0201	1	(54)		1	K.KQTALVELLK.H + [+100.0160 at N-term K]
371	621.8637	1241.7129	1241.7231	-0.0102	1	(58)		1	K.KQTALVELLK.H + [+100.0160 at N-term K]
372	621.8781	1241.7416	1241.7231	0.0186	1	62		1	K.KQTALVELLK.H + [+100.0160 at N-term K]
383	418.6065	1252.7976	1252.7027	0.0949	1	50		1	K.LVTDLTKVHK.E + [+100.0160 at C-term K]
436	653.3410	1304.6674	1304.7088	-0.0415	0	(57)	0.001	1	K.HLVDEPQNLIK.Q
437	653.3438	1304.6731	1304.7088	-0.0357	0	(68)	7.7e-05	1	K.HLVDEPQNLIK.Q
442	653.3832	1304.7518	1304.7088	0.0430	0	69	6.4e-05	1	K.HLVDEPQNLIK.Q
443	653.3952	1304.7759	1304.7088	0.0671	0	(53)	0.0025	1	K.HLVDEPQNLIK.Q
444	653.3976	1304.7806	1304.7088	0.0718	0	(54)	0.0017	1	K.HLVDEPQNLIK.Q
445	653.4365	1304.8584	1304.7088	0.1496	0	(49)	0.006	1	K.HLVDEPQNLIK.Q
446	435.9996	1304.9770	1305.6677	-0.6907	0	(52)		1	K.HLVDEPQNLIK.Q + [+0.9589 at I10]
474	671.3431	1340.6716	1340.6799	-0.0083	0	(57)		1	K.TVMENFVAFVDK.C + [-58.0055 at D11]
515	697.8561	1393.6976	1393.7129	-0.0154	1	75		1	K.FPKAEFVEVTK.L + [+100.0160 at K3]
519	700.3382	1398.6619	1398.6853	-0.0235	0	70	4.9e-05	1	K.TVMENFVAFVDK.C

Possible Assignments:

malonyl methylation (K) [+100.0160]
Succinyl (K) [+100.0160]

Manual Error Tolerance Search Steps

1. Regular search

MASCOT MS/MS Ions Search

Your name	Sheng	Email	sz14@cornell.edu
Search title	Auto error tolerance test		
Database	SwissProt		
Taxonomy	. . . . . Mammalia (mammals)		
Enzyme	Trypsin	Allow up to	1 missed cleavages
Fixed modifications	Acetyl (N-term) Acetyl (Protein N-term) Amidated (C-term) Amidated (Protein C-term) Ammonia-loss (N-term C)	Variable modifications	Biotin (K) Biotin (N-term) Carbamidomethyl (C) Carbamyl (K) Carbamyl (N-term)
Quantitation	None		
Peptide tol. ±	1.5 Da # ^{13}C 0	MS/MS tol. ±	0.6 Da
Peptide charge	2+ and 3+	Monoisotopic	Average
Data file	X:\Sheng\mas504A.tmp Browse...		
Data format	Mascot generic	Precursor	m/z
Instrument	Default	Error tolerant	☐
Decoy	☐	Report top	AUTO hits
	Start Search ...		Reset Form

2. Manual error tolerance

Mascot Search Results

```
User          : Sheng
Email         : sz14@cornell.edu
Search title  : Auto error tolerance  test
MS data file  : X:\Sheng\mas504A.tmp
Database      : SwissProt 20080728 (392667 sequences; 141217034 residues)
Taxonomy      : Mammalia (mammals) (63164 sequences)
Timestamp     : 13 Oct 2009 at 19:10:01 GMT
Protein hits  : ALBU_BOVIN Serum albumin OS=Bos taurus GN=ALB PE=1 SV=4
                TRYP_PIG   Trypsin OS=Sus scrofa PE=1 SV=1
                K1C9_HUMAN Keratin, type I cytoskeletal 9 OS=Homo sapiens GN=KRT9 PE=1 SV=2
                K2C1_HUMAN Keratin, type II cytoskeletal 1 OS=Homo sapiens GN=KRT1 PE=1 SV=5
                DHE3_BOVIN Glutamate dehydrogenase 1, mitochondrial OS=Bos taurus GN=GLUD1 P
```

Probability Based Mowse Score

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 39 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.

Peptide Summary Report

3. Perform search

Format As: Peptide Summary — Help

Significance threshold p< 0.05 — Max. number of hits AUTO

Standard scoring ○ MudPIT scoring ◉ — Ions score or expect cut-off 35 — Show sub-sets 0

Show pop-ups ◉ Suppress pop-ups ○ — Sort unassigned Decreasing Score — Require bold red

Select All | Select None | Search Selected | ☑ Error tolerant | Archive Report

1. ALBU_BOVIN Mass: 69248 Score: 2325 Queries matched: 75 emPAI: 1.77
Serum albumin OS=Bos taurus GN=ALB PE=1 SV=4

☑ Check to include this hit in error tolerant search or archive report

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	22	395.2537	788.4928	788.4644	0.0284	0	48	0.011	1	K.LVTDLTK.V
☑	75	443.7565	885.4985	885.4080	0.0905	0	54	0.0022	1	K.DDSPDLPK.L
☑	173	507.8014	1013.5882	1013.6121	-0.0238	0	(48)	0.0087	1	K.QTALVELLK.I
☑	174	507.8206	1013.6266	1013.6121	0.0145	0	(47)	0.0096	1	K.QTALVELLK.I
☑	175	507.8259	1013.6373	1013.6121	0.0253	0	(51)	0.0042	1	K.QTALVELLK.I
☑	176	507.8342	1013.6539	1013.6121	0.0418	0	51	0.004	1	K.QTALVELLK.I
☑	177	507.8389	1013.6632	1013.6121	0.0512	0	(50)	0.0052	1	K.QTALVELLK.I

Manual Error Tolerance Search Steps

MASCOT MS/MS Ions Search

Your name	Sheng	Email	sz14@cornell.edu
Search title	Manual error tolerance test		
Database	SwissProt		
Taxonomy	..... Mammalia (mammals)		
Enzyme	None	Allow up to	1 missed cleavages
Fixed modifications	Acetyl (K) Acetyl (N-term) Acetyl (Protein N-term) Amidated (C-term) Amidated (Protein C-term)	Variable modifications	Oxidation (HW) Oxidation (M) Phospho (ST) Phospho (Y) Propionamide (C)
Quantitation	None		
Peptide tol. ±	1.5 Da # ^{13}C 0	MS/MS tol. ±	0.8 Da
Peptide charge	2+ and 3+	Monoisotopic	Average
Data file	X:\Sheng\mas504A.tmp		
Data format	Mascot generic	Precursor	m/z
Instrument	Default	Error tolerant search	
		Report top	AUTO hits

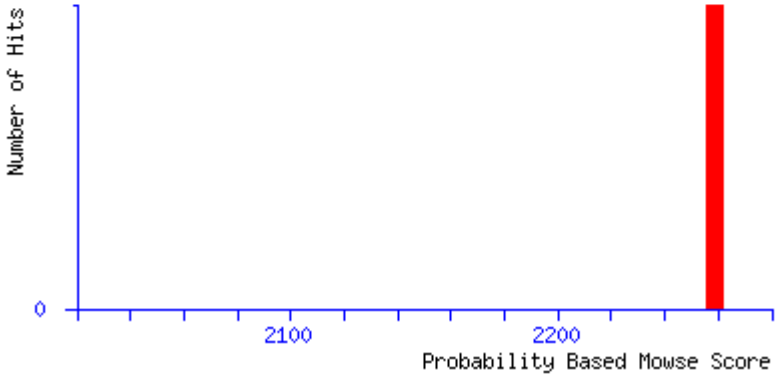
Start Search ... Reset Form

MATRIX SCIENCE Mascot Search Results

```
User            : Sheng
Email           : sz14@cornell.edu
Search title    : Manual error tolerance  test
MS data file    : X:\Sheng\mas504A.tmp
Database        : SwissProt 20080728 (392667 sequences;
Taxonomy        : Mammalia (mammals) (63164 sequences)
                : Error tolerant search of 1 entries
Timestamp       : 13 Oct 2009 at 19:58:21 GMT
Protein hits    : ALBU_BOVIN Serum albumin OS=Bos taurus
```

Probability Based Mowse Score

Ions score is -10*Log(P), where P is the probability that the observed ma
Protein scores are derived from ions scores as a non-probabilistic basis fo



Manual Error Tolerance Search Results

1. ALBU_BOVIN Mass: 69248 Score: 2129 Queries matched: 87

Serum albumin OS=Bos taurus GN=ALB PE=1 SV=4

☑ Check to include this hit in error tolerant search or archive report

BSA Succinylation

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	75	443.7565	885.4985	885.3716	0.1269	0	54		1	K.DDSPDLPK.L + [-0.0364 at C-term K]
☑	176	507.8342	1013.6539	1013.5757	0.0782	0	51		1	K.QTALVELLK.H + [-0.0364 at C-term K]
☑	264	560.3150	1118.6154	1118.4954	0.1200	0	54		1	K.LVNELTEF.A + [+155.0041 at C-term F]
☑	267	561.2785	1120.5425	1119.5747	0.9678	0	63		1	D.KEACFAVEGPK.L + [-58.0055 at E2]
☑	299	582.2787	1162.5429	1162.5870	-0.0441	0	(56)		1	K.LVNELTEFAK.T + [-0.0364 at C-term K]
☑	300	582.3039	1162.5932	1162.5870	0.0062	0	60		1	K.LVNELTEFAK.T + [-0.0364 at C-term K]
☑	301	582.3169	1162.6192	1162.5870	0.0323	0	(60)		1	K.LVNELTEFAK.T + [-0.0364 at C-term K]
☑	302	582.3302	1162.6458	1162.5870	0.0588	0	(59)		1	K.LVNELTEFAK.T + [-0.0364 at C-term K]
☑	305	582.8404	1163.6663	1163.6074	0.0589	0	(57)		1	K.LVNELTEFAK.T + [+0.9840 at N3]
☑	367	621.8409	1241.6672	1241.7231	-0.0559	0	(60)		1	K.KQTALVELLK.H + [+100.0160 at N-term K]
☑	369	621.8558	1241.6970	1241.7231	-0.0261	0	(52)		1	K.KQTALVELLK.H + [+100.0160 at N-term K]
☑	370	621.8588	1241.7030	1241.7231	-0.0201	0	(54)		1	K.KQTALVELLK.H + [+100.0160 at N-term K]
☑	371	621.8637	1241.7129	1241.7231	-0.0102	0	(58)		1	K.KQTALVELLK.H + [+100.0160 at N-term K]
☑	372	621.8781	1241.7416	1241.7231	0.0186	0	62		1	K.KQTALVELLK.H + [+100.0160 at N-term K]
☑	436	653.3410	1304.6674	1304.6725	-0.0051	0	(57)		1	K.HLVDEPQNLIK.Q + [-0.0364 at C-term K]
☑	437	653.3438	1304.6731	1304.6725	0.0007	0	(68)		1	K.HLVDEPQNLIK.Q + [-0.0364 at C-term K]
☑	442	653.3832	1304.7518	1304.6725	0.0793	0	69		1	K.HLVDEPQNLIK.Q + [-0.0364 at C-term K]
☑	443	653.3952	1304.7759	1304.6725	0.1034	0	(53)		1	K.HLVDEPQNLIK.Q + [-0.0364 at C-term K]
☑	444	653.3976	1304.7806	1304.6725	0.1082	0	(54)		1	K.HLVDEPQNLIK.Q + [-0.0364 at C-term K]
☑	446	435.9996	1304.9770	1305.6677	-0.6907	0	(52)		1	K.HLVDEPQNLIK.Q + [+0.9589 at I10]
☑	474	671.3431	1340.6716	1340.6799	-0.0083	0	(57)		1	K.TVMENFVAFVDK.C + [-58.0055 at D11]
☑	510	693.7849	1385.5553	1385.4768	0.0785	0	50		1	K.EYEATLEECC.A + [+197.0453 at E8]
☑	515	697.8561	1393.6976	1393.7129	-0.0154	0	75		1	K.FPKAEFVEVTK.L + [+100.0160 at K3]
☑	519	700.3382	1398.6619	1398.6490	0.0129	0	70		1	K.TVMENFVAFVDK.C + [-0.0364 at C-term K]
☑	528	708.3355	1414.6565	1414.6439	0.0126	0	(52)		1	K.TVMENFVAFVDK.C + Oxidati
☑	551	480.6339	1438.8798	1438.8045	0.0753	0	(63)		1	R.RHPEYAVSVLLR.L
☑	552	480.6396	1438.8970	1438.8045	0.0925	0	(57)		1	R.RHPEYAVSVLLR.L
☑	553	480.6459	1438.9159	1438.8045	0.1114	0	(54)		1	R.RHPEYAVSVLLR.L
☑	554	480.6477	1438.9213	1438.8045	0.1168	0	(58)		1	R.RHPEYAVSVLLR.L
☑	555	480.6489	1438.9250	1438.8045	0.1205	0	(76)		1	R.RHPEYAVSVLLR.L

Possible Assignments:

malonyl methylation (K) [+100.0160]
Succinyl (K) [+100.0160]

Interpreting Mascot Results:

Summary Report & Filters

```
User                    : Celeste Ptak
Email                   : acp49@cornell.edu
Search title            : Y:\MS\Sheng backup\Analyst data since 24Jul06\Projects\2009_07_16Data\09Oct09\XSm005_2
MS data file            : c:\DOCUME~1\mslab\LOCALS~1\Temp\mas50F7.tmp
Database                : SwissProt 20080728 (392667 sequences; 141217034 residues)
Taxonomy                : Mammalia (mammals) (63164 sequences)
Timestamp               : 13 Oct 2009 at 18:20:45 GMT
Enzyme                  : Trypsin
Variable modifications  : Carbamidomethyl (C),Oxidation (M),Succinyl (K)
Mass values             : Monoisotopic
Protein Mass            : Unrestricted
Peptide Mass Tolerance  : ± 1.5 Da
Fragment Mass Tolerance: ± 0.6 Da
Max Missed Cleavages    : 3
Instrument type         : Default
Number of queries       : 1237
Protein hits            : ALBU_BOVIN Serum albumin OS=Bos taurus GN=ALB PE=1 SV=4
                          TRYP_PIG   Trypsin OS=Sus scrofa PE=1 SV=1
                          K1C9_HUMAN Keratin, type I cytoskeletal 9 OS=Homo sapiens GN=KRT9 PE=1 SV=2
                          K2C1_HUMAN Keratin, type II cytoskeletal 1 OS=Homo sapiens GN=KRT1 PE=1 SV=
                          DHE3_BOVIN Glutamate dehydrogenase 1, mitochondrial OS=Bos taurus GN=GLUD1 PE=1 SV=2
```

Search parameters

Proteins identified

Select Summary Report

Format As	Select Summary (protein hits)		Help
	Significance threshold p< 0.05	Max. number of hits AUTO	
	Standard scoring ○ MudPIT scoring ◉	Ions score or expect cut-off 35	Show sub-sets 0
	Show pop-ups ◉ Suppress pop-ups ○	Sort unassigned Decreasing Score	Require bold red ☑

Report filters

```
1.    ALBU_BOVIN    Mass: 69248    Score: 2211    Queries matched: 86    emPAI: 2.49
      Serum albumin OS=Bos taurus GN=ALB PE=1 SV=4
```

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
22	395.2537	788.4928	788.4644	0.0284	0	48	0.024	1	K.LVTDLTK.V
75	443.7565	885.4985	885.4080	0.0905	0	54	0.005	1	K.DDSPDLPK.L

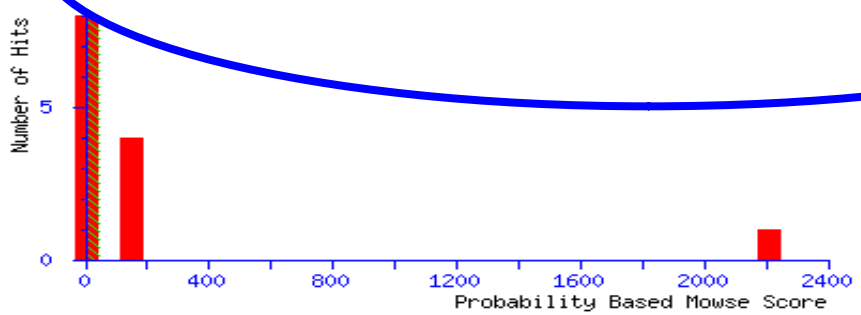
Interpreting Mascot Results:

Peptide Summary

Probability Based Mowse Score

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 44 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.

Probability scores



Peptide Summary Report

Format As | Peptide Summary | Help

Peptide Summary
Select Summary (protein hits)
Select Summary (unassigned)
Export Search Results

Max. number of hits AUTO

Ions score or expect cut-off 35 | Show sub-sets 0

Show pop-ups | Suppress pop-ups | Sort unassigned Decreasing Score | Require bold red

Report filters

Select All | Select None | Search Selected | Error tolerant | Archive Report

1. ALBU_BOVIN Mass: 69248 Score: 2211 Queries matched: 86 emPAI: 2.49
Serum albumin OS=Bos taurus GN=ALB PE=1 SV=4

Check to include this hit in error tolerant search or archive report

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
22	395.2537	788.4928	788.4644	0.0284	0	48	0.024	1	K.LVTDLTK.V
75	443.7565	885.4985	885.4080	0.0905	0	54	0.005	1	K.DDSPDLPK.L

Interpreting Mascot Results:

Peptide Summary

Peptide Summary Report

Format As: Peptide Summary | Help

Significance threshold p< 0.05 | Max. number of hits AUTO

Standard scoring ○ MudPIT scoring ◉ | Ions score or expect cut-off 58 | Show sub-sets 0

Show pop-ups ◉ Suppress pop-ups ○ | Sort unassigned Decreasing Score | Require bold red ☑

Select All | Select None | Search Selected | ☐ Error tolerant | Archive Report

ALBU_BOVIN Mass: 69248 Score: 1525 Queries matched: 37 emPAI: 0.66
Serum albumin OS=Bos taurus GN=ALB PE=1 SV=4
☐ Check to include this hit in error tolerant search or archive report

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	300	582.3169	1162.6192	1162.6234	-0.0041	0	61	0.0013	1	K.LVNELTEFAK.T
☑	301	582.3302	1162.6458	1162.6234	0.0224	0	(60)	0.0016	1	K.LVNELTEFAK.T
☑	366	621.8409	1241.6672	1241.7231	-0.0559	1	58	0.0021	1	K.KQTALVELLK.H + Succinyl (K)
☑	370	621.8637	1241.7129	1241.7231	-0.0102	1	(58)	0.0022	1	K.KQTALVELLK.H + Succinyl (K)
☑	436	653.3438	1304.6731	1304.7088	-0.0357	0	(68)	0.00022	1	K.HLVDEPQNLIK.Q
☑	441	653.3832	1304.7518	1304.7088	0.0430	0	69	0.00019	1	K.HLVDEPQNLIK.Q
☑	513	697.8561	1393.6976	1393.7129	-0.0154	1	74	6.2e-05	1	K.FPKAEFVEVTK.L + Succinyl (K)
☑	514	700.3382	1398.6619	1398.6853	-0.0235	0	68	0.00027	1	K.TVMENFVAFVDK.G
☑	54									
☑	55									
☑	55									
☑	55									
☑	55									
☑	55									
☑	58									
☑	58									
☑	58									

Top scoring peptide matches to query 513
File: XSm005_20.wiff, Sample: XSm005_20 (sample number 1), Elution: 38.13 to 38.19 min, Period: 1, Cycle(s):
Score greater than 44 indicates identity

Score	Expect	Delta	Hit	Protein	Peptide
73.6	6.2e-05	-0.0154	1	ALBU_BOVIN	K.FPKAEFVEVTK.L
28.6	2	0.9363			R.TKAENEFVTVKK.D
22.2	8.6	-0.0273			K.KQMREFRGNTK.N
19.0	18	-0.0702			K.RAAALEAHLDVTK.Q
17.7	24	0.0395			K.ENCDIIIQAMTK.I
17.7	24	0.1083			R.TRNGGGGCESETVQ.-
16.1	35	0.9871			R.MAQVVQVMASAVK.V

Pop-ups show more details

Interpreting Mascot Results:

Protein View

Protein View

Match to: ALBU_BOVIN Score: 2217
Serum albumin OS=Bos taurus GN=ALB PE=1 SV=4
Found in search of c:\DOCUME~1\mslab\LOCALS~1\Temp\mas50F7.tmp

Nominal mass (M_r): 69248; Calculated pI value: 5.82
NCBI BLAST search of ALBU_BOVIN against nr
Unformatted sequence string for pasting into other applications

Taxonomy: Bos taurus

Variable modifications: Carbamidomethyl (C),Oxidation (M),Succinyl (K)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Sequence Coverage: 32%

Matched peptides shown in Bold Red

```
  1 MKWVTFISLL LLFSSAYSRG VFRRDTHKSE IAHRFKDLGE EHFKGLVLIA
 51 FSQYLQQCPF DEHVKLVNEL TEFAKTCVAD ESHAGCEKSL HTLFGDELCK
101 VASLRETYGD MADCCEKQEP ERNECFLSHK DDSPDLPKLK PDPNTLCDEF
151 KADEKKFWGK YLYEIARRHP YFYAPELLYY ANKYNGVFQE CCQAEDKGAC
201 LLPKIETMRE KVLASSARQR LRCASIQKFG ERALKAWSVA RLSQKFPKAE
251 FVEVTKLVTD LTKVHKECCH GDLLECADDR ADLAKYICDN QDTISSKLKE
301 CCDKPLLEKS HCIAEVEKDA IPENLPPLTA DFAEDKDVCK NYQEAKDAFL
351 GSFLYEYSRR HPEYAVSVLL RLAKEYEATL EECCAKDDPH ACYSTVFDKL
401 KHLVDEPQNL IKQNCDQFEK LGEYGFQNAL IVRYTRKVPQ VSTPTLVEVS
451 RSLGKVGTRC CTKPESERMP CTEDYLSLIL NRLCVLHEKT PVSEKVTKCC
501 TESLVNRRPC FSALTPDETY VPKAFDEKLF TFHADICTLP DTEKQIKKQT
551 ALVELLKHKP KATEEQLKTV MENFVAFVDK CCAADDKEAC FAVEGPKLVV
601 STQTALA
```

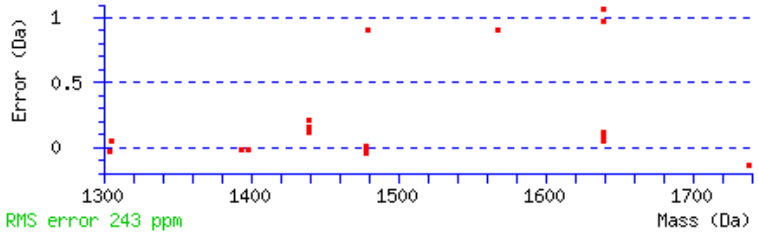
Show predicted peptides also

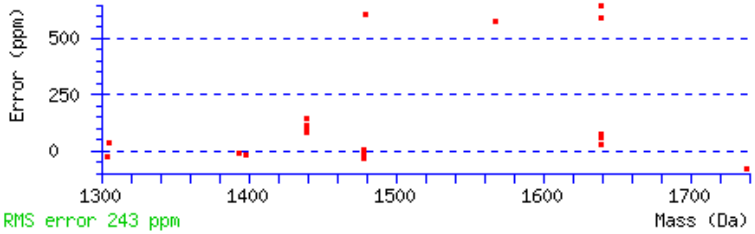
Interpreting Mascot Results:

Protein View

Sort Peptides By ◉ Residue Number ○ Increasing Mass ○ Decreasing Mass

Start - End	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Sequence
246 - 256	697.8561	1393.6976	1393.7129	-0.0154	1	K.FPKAEFVEVTK.L Succinyl (K) (Ions score 74)
347 - 359	784.8267	1567.6389	1566.7354	0.9034	0	K.DAFLGSFLYEYSR.R (Ions score [illegible])
360 - 371	480.6489	1438.9250	1438.8045	0.1205	1	R.RHPEYAVSVLLR.L (Ions score 73)
360 - 371	480.6511	1438.9314	1438.8045	0.1269	1	R.RHPEYAVSVLLR.L (Ions score 68)
360 - 371	480.6610	1438.9610	1438.8045	0.1566	1	R.RHPEYAVSVLLR.L (Ions score 81)
360 - 371	480.6781	1439.0124	1438.8045	0.2079	1	R.RHPEYAVSVLLR.L (Ions score 72)
402 - 412	653.3438	1304.6731	1304.7088	-0.0357	0	K.HLVDEPQNLIK.Q (Ions score 68)
402 - 412	653.3832	1304.7518	1304.7088	0.0430	0	K.HLVDEPQNLIK.Q (Ions score 69)
421 - 433	740.3784	1478.7422	1478.7881	-0.0459	0	K.LGEYGFQNALIVR.Y (Ions score 95)
421 - 433	740.3809	1478.7472	1478.7881	-0.0410	0	K.LGEYGFQNALIVR.Y (Ions score 86)
421 - 433	740.3816	1478.7486	1478.7881	-0.0396	0	K.LGEYGFQNALIVR.Y (Ions score 105)
421 - 433	740.3851	1478.7557	1478.7881	-0.0324	0	K.LGEYGFQNALIVR.Y (Ions score 99)
421 - 433	740.3978	1478.7811	1478.7881	-0.0071	0	K.LGEYGFQNALIVR.Y (Ions score 89)
421 - 433	740.4003	1478.7861	1478.7881	-0.0020	0	K.LGEYGFQNALIVR.Y (Ions score 112)
421 - 433	740.4018	1478.7891	1478.7881	0.0009	0	K.LGEYGFQNALIVR.Y (Ions score 87)
421 - 433	740.4040	1478.7935	1478.7881	0.0054	0	K.LGEYGFQNALIVR.Y (Ions score 105)
421 - 433	740.4044	1478.7943	1478.7881	0.0062	0	K.LGEYGFQNALIVR.Y (Ions score 99)
421 - 433	494.2381	1479.6925	1478.7881	0.9044	0	K.LGEYGFQNALIVR.Y (Ions score 80)
437 - 451	547.3323	1638.9750	1638.9305	0.0445	1	R.KVPQVSTPTLVEVSR.S (Ions score 89)
437 - 451	547.3490	1639.0251	1638.9305	0.0946	1	R.KVPQVSTPTLVEVSR.S (Ions score 71)
437 - 451	547.3529	1639.0369	1638.9305	0.1064	1	R.KVPQVSTPTLVEVSR.S (Ions score 73)
437 - 451	547.3568	1639.0486	1638.9305	0.1181	1	R.KVPQVSTPTLVEVSR.S (Ions score 77)
437 - 451	547.6403	1639.8992	1638.9305	0.9687	1	R.KVPQVSTPTLVEVSR.S (Ions score 68)
437 - 451	547.6712	1639.9917	1638.9305	1.0612	1	R.KVPQVSTPTLVEVSR.S (Ions score 75)
437 - 451	870.4118	1738.8091	1738.9465	-0.1374	1	R.KVPQVSTPTLVEVSR.S Succinyl (K) (Ions score 69)
569 - 580	700.3382	1398.6619	1398.6853	-0.0235	0	K.TVMENFVAFVDK.C (Ions score 68)





Interpreting Mascot Results:

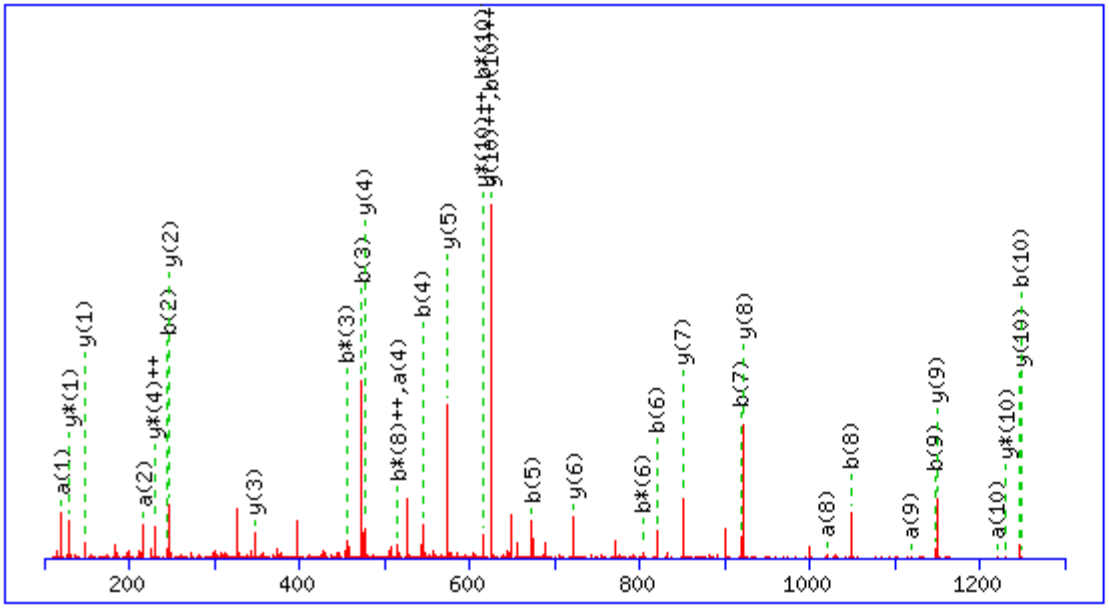
Peptide View

Peptide View

MS/MS Fragmentation of **FPKAEFVEVTK**
Found in **ALBU_BOVIN**, Serum albumin OS=Bos taurus GN=ALB PE=1 SV=4

Match to Query 513: 1393.697564 from(697.856058,2+)
Title: File: XSm005_20.wiff, Sample: XSm005_20 (sample number 1), Elution: 38.13 to 38.19 min, Period: 1, Cycle(s): 465 (Experiment 3), Data file c:\DOCUME~1\mslab\LOCALS~1\Temp\mas50F7.tmp

Click mouse within plot area to zoom in by factor of two about that point
Or, Plot from 100 to 1300 Da Full range



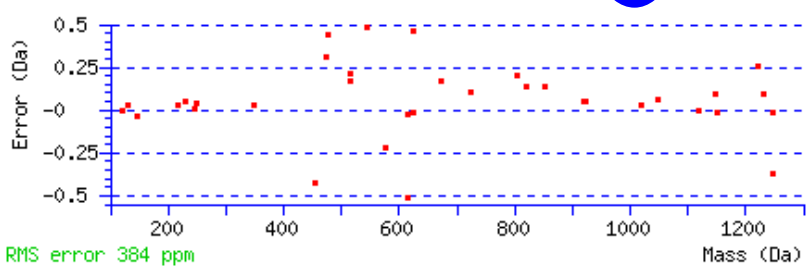
Interpreting Mascot Results:

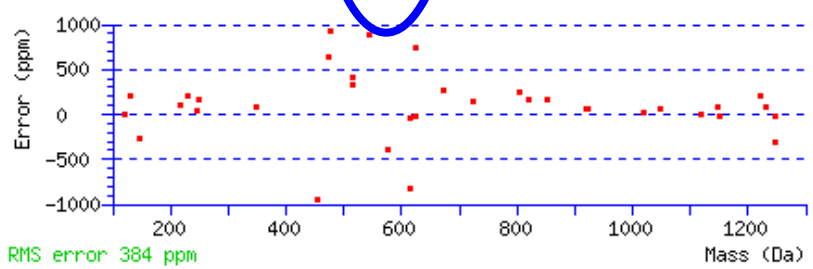
Peptide View

Monoisotopic mass of neutral peptide Mr(calc): 1393.7129
Variable modifications:
K3 : Succinyl (K)
Ions Score: 74 Expect: 6.2e-05
Matches (Bold Red): 35/112 fragment ions using 50 most intense peaks

Probability scores

#	a	a++	a*	a*++	b	b++	b*	b*++	Seq.	y	y++	y*	y*++	#
1	**120.0808**	60.5440			148.0757	74.5415			F					11
2	**217.1335**	109.0704			**245.1285**	123.0679			P	**1247.6518**	**624.3295**	**1230.6252**	**615.8163**	10
3	445.2445	223.1259	428.2180	214.6125	**473.2395**	237.1234	**456.2129**	228.6101	K	**1150.5990**	575.8032	1133.5725	567.2899	9
4	**516.2817**	258.6445	499.2551	250.1312	**544.2766**	272.6419	527.2500	264.1286	A	**922.4880**	461.7477	905.4615	453.2344	8
5	645.3242	323.1658	628.2977	314.6525	**673.3192**	337.1632	656.2926	328.6499	E	**851.4509**	426.2291	834.4244	417.7158	7
6	792.3927	396.7000	775.3661	388.1867	**820.3876**	410.6974	**803.3610**	402.1842	F	**722.4083**	361.7078	705.3818	353.1945	6
7	891.4611	446.2342	874.4345	437.7209	**919.4560**	460.2316	902.4294	451.7184	V	**575.3399**	288.1736	558.3134	279.6603	5
8	**1020.5037**	510.7555	1003.4771	502.2422	**1048.4986**	524.7529	1031.4720	**516.2397**	E	**476.2715**	238.6394	459.2449	**230.1261**	4
9	**1119.5721**	560.2897	1102.5455	551.7764	**1147.5670**	574.2871	1130.5404	565.7739	V	**347.2289**	174.1181	330.2023	165.6048	3
10	**1220.6198**	610.8135	1203.5932	602.3002	**1248.6147**	**624.8110**	1231.5881	**616.2977**	T	**248.1605**	124.5839	231.1339	116.0706	2
11									K	**147.1128**	74.0600	**130.0863**	65.5468	1





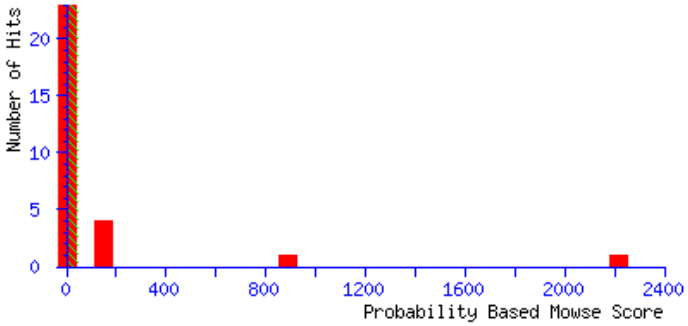
Interpreting Mascot Results:

Export search results

```
Timestamp        : 13 Oct 2009 at 18:20:45 GMT
Protein hits     : ALBU_BOVIN Serum albumin OS=Bos taurus GN=ALB PE=1 SV=4
                   ALBU_SHEEP Serum albumin OS=Ovis aries GN=ALB PE=2 SV=1
                   TRYP_PIG   Trypsin OS=Sus scrofa PE=1 SV=1
                   K1C9_HUMAN Keratin, type I cytoskeletal 9 OS=Homo sapiens GN=KRT9 PE=1 SV=2
                   K2C1_HUMAN Keratin, type II cytoskeletal 1 OS=Homo sapiens GN=KRT1 PE=1 SV=5
                   DHE3_BOVIN Glutamate dehydrogenase 1, mitochondrial OS=Bos taurus GN=GLUD1 PE=1 SV=2
```

Probability Based Mowse Score

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 44 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.



Peptide Summary Report

Format As	Export Search Results		Help
	Significance threshold p< 0.05	Max. number of hits AUTO	
	Standard scoring ○ MudPIT scoring ◉	Ions score or expect cut-off 30	Show sub-sets 0
	Show pop-ups ◉ Suppress pop-ups ○	Sort unassigned Decreasing Score	Require bold red ☑

Interpreting Mascot Results:

Export search results

Export search results

Help

Export format	CSV
Significance threshold p<	0.05
Max. number of hits	AUTO
Protein scoring	Standard ○ MudPIT ◉
Ions score cut-off	30
Include same-set protein hits (additional proteins that span the same set of peptides)	☐
Include sub-set protein hits (additional proteins that span a sub-set of peptides)	0
Require bold red	☑

Search Information ☑

Header	☑
Variable mod. info.	☑
Search parameters	☑
Format parameters	☑
Residue masses	☐

Protein Hit Information ☑

Score	☑
Description*	☑
Mass (Da)*	☑
Number of queries matched	☑
Percent coverage**	☑
Length in residues**	☐
pI**	☑
Taxonomy**	☐
Taxonomy ID**	☐
Protein sequence**	☐
emPAI	☑

* Occasionally requires information to be retrieved from external utilities, which can be

** Always requires information to be retrieved from external utilities, which can be slow

Query Level Information ☐

Query title	☐
seq(), comp(), tag(), etc.	☐
Query level search parameters	☐
MS/MS Peak lists	☐
Raw peptide match data	☐

Export search results

Peptide Match Information ☑

Experimental Mr (Da)	☑
Experimental charge	☑
Calculated Mr (Da)	☑
Mass error (Da)	☑
Start	☐
End	☐
Number of missed cleavages	☑
Score	☑
Homology threshold	☐
Identity threshold	☐
Expectation value	☑
Sequence	☑
Frame number	☐
Variable Modifications	☑
Number of fragment ion matches	☐
Query title	☑
Unassigned queries (peptide matches not assigned to protein hits)	☐

Interpreting Mascot Results:

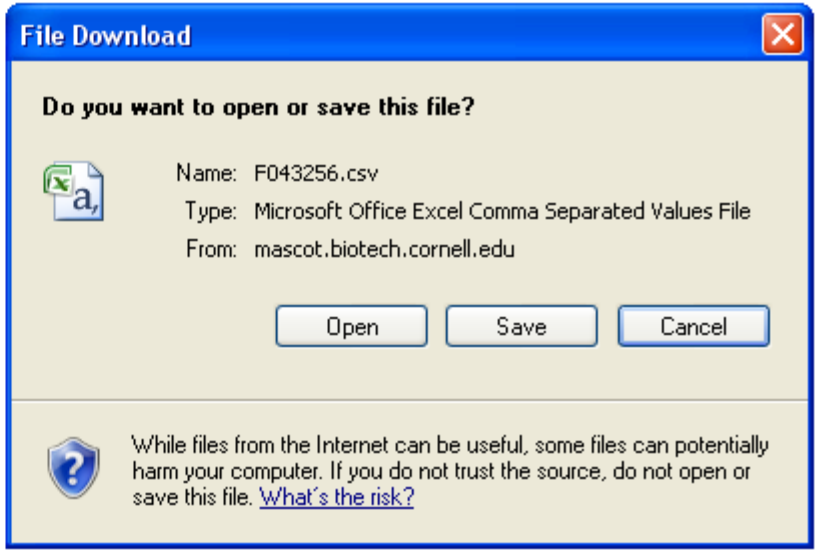
Export search results

Score	☑
Homology threshold	☐
Identity threshold	☐
Expectation value	☑
Sequence	☑
Frame number	☐
Variable Modifications	☑
Number of fragment ion matches	☐
Query title	☑
Unassigned queries (peptide matches not assigned to protein hits)	☐

Query Level Information ☐

Query title	☐
seq(), comp(), tag(), etc.	☐
Query level search parameters	☐
MS/MS Peak lists	☐
Raw peptide match data	☐

Export search results



Interpreting Mascot Results:

Export search results:

Search title	Y:\MS\Sheng backup\Analyst data since 24Jul06\Projects\2009_07_16Data\09Oct09\XSm005_20.wiff (sample number 1)
Timestamp	2009-10-13T18:20:45Z
User	Celeste Ptak
Email	acp49@cornell.edu
Report URI	http://mascot.biotech.cornell.edu/mascot/cgi/master_results.pl?file=../data/20091013/F043256.dat
Peak list data path	c:\DOCUME~1\mslab\LOCALS~1\Temp\mas50F7.tmp
Peak list format	Mascot generic
Search type	MIS
Mascot version	2.2.04
Database	SwissProt
Fasta file	SwissProt_20080728.fasta
Total sequences	392667
Total residues	141217034
Sequences after taxonomy filter	63164
Number of queries	1237

Variable modifications ----------

Identifier	Name	Delta	Neutral loss(es)	
1	Carbamidomethyl (C)	57.02147	0	
2	Oxidation (M)	15.99492	63.99829	0
3	Succinyl (K)	100.016	0	

Search Parameters ----------

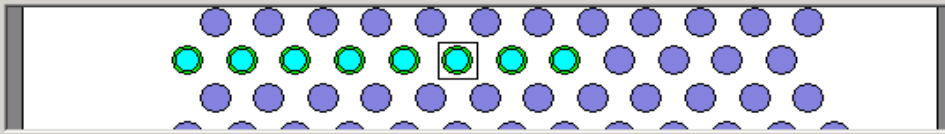
Taxonomy filter	. . . . . Mammalia (mammals)
Enzyme	Trypsin
Maximum Missed Cleavages	3
Fixed modifications	
Quantitation method	None
Variable modifications	Carbamidomethyl (C),Oxidation (M),Succinyl (K)
Peptide Mass Tolerance	1.5
Peptide Mass Tolerance Units	Da
Fragment Mass Tolerance	0.6

Mascot-based GPS Explorer Database Search for MALDI-TOF/TOF Data

Beta-Galac

Results Browser

MALDI PLATES | SOURCE PLATES | METRIC PLOT | METRIC DATA | MS/MS SUMMARY



PROTEINS IN CURRENT SPOT | PROTEIN SUMMARY

Proteins - Plate: 1 Spot: C11

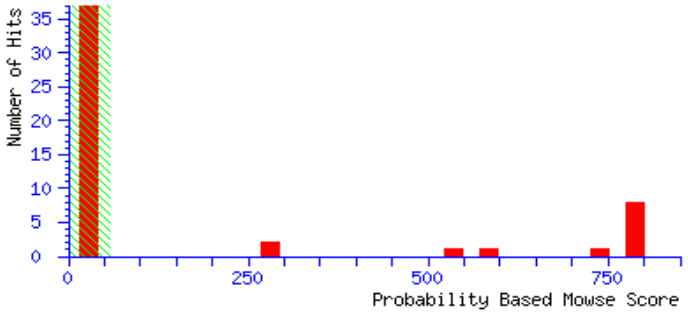
Rank	Protein Name	Accession Number	Protein MW	Protein PI	Peptide Count	Protein Score	Protein Score C.I. %	Total Ion Score	Best Ion Score C.I. %	Best Ion Score	Total Ion Score C.I. %
1	unnamed protein product [Escherichia coli]	gi|1197203	116278.1	5.2800	37	788	100.000	487	100.000	97	100.000
2	Chain D, E. Coli (Lacz) Beta-Galactosidase (E537q	gi|17943282	116311.0	5.3000	36	775	100.000	487	100.000	97	100.000

Mascot Search Results

```
User            :
Email           :
Search title    : SampleSetID: 2882, AnalysisID: 5898, MaldiWellID: 78894, SpectrumID: 159720, Path=\Yong\sensitivity tes
Database        : NCBInr 20050405 (2419798 sequences; 819541351 residues)
Taxonomy        : Escherichia coli (28251 sequences)
Timestamp       : 22 Apr 2005 at 17:48:56 GMT
Top Score       : 788 for gi|17943265, Chain H, E. Coli (Lacz) Beta-Galactosidase In Complex With 2-F- Lactose. Chains A-
```

Probability Based Mowse Score

Score is -10*Log(P), where P is the probability that the observed match is a random event.
Protein scores greater than 57 are significant (p<0.05).



Acknowledgments

- **IT Service Facility:** Maintenance of Mascot Server
  - James VanEe
  - John Flaherty
- **Computational Biology Service Unit:** Database Info
  - Qi Sun
- **Proteomics & MS Facility:**
  - Celeste Ptak
  - Huiming Yan
  - Robert Sherwood
- **USDA-ARS Proteomics Center:**
  - Ted Thannhauser
  - Yong Yang
  - Kevin Howe

Take-home Messages

- ✓ Describe how MS/MS Ion Search works
- ✓ List most common reasons for failed searches
- ✓ Describe the significance of False Discovery Rate determination
- ✓ Describe the circumstances required for error tolerance search