

***Error Tolerant Searching  
of Uninterpreted MS/MS  
Data***

*{MATRIX}  
{SCIENCE}*

Mascot Search Results - Microsoft Internet Explorer

Address [http://deli5000/mascot/cgi/master\\_results.pl?file=..data/20020522/F593324.dat](http://deli5000/mascot/cgi/master_results.pl?file=..data/20020522/F593324.dat)

**Mascot Search Results**

User : JSC  
 Email : jcottrell@matrixscience.com  
 Search title : Annexin first pass  
 MS data file : C:\Projects\Echhaust\ms\_data\qtof20953.mgf  
 Database : MSDB 20020415 (851746 sequences; 265326103 residues)  
 Timestamp : 22 May 2002 at 15:35:36 GMT

Significant hits:

<a href="#">AAA51708</a>	HUMALPPA NID: - Homo sapiens
<a href="#">Q96DB7</a>	HYPOTHETICAL 58.0 KDA PROTEIN.- Homo sapiens (Human).
<a href="#">Q16727</a>	PLACENTAL-LIKE ALKALINE PHOSPHATASE PRECURSOR (EC 3.1.3.1).- Homo sapiens (Human).
<a href="#">TVHUY5</a>	protein-tyrosine kinase (EC 2.7.1.112) yes-1 - human
<a href="#">AA641947</a>	AF304164 NID: - Homo sapiens
<a href="#">216D</a>	trypsin (EC 3.4.21.4) precursor (diisopropylphosphorylated) - bovine
<a href="#">1HTP</a>	trypsin (EC 3.4.21.4) (isopropylphosphorylated) - bovine
<a href="#">YES AVISY</a>	Tyrosine-protein kinase transforming protein YES (EC 2.7.1.112).- Avian sarcoma virus (str.
<a href="#">Q9UBR3</a>	INSULIN RECEPTOR SUBSTRATE PROTEIN OF 53 KDA (A SHORTER FORM).- Homo sapiens (Human).
<a href="#">I51593</a>	protein-tyrosine kinase (EC 2.7.1.112) yes - Xiphophorus helleri
<a href="#">A49114</a>	protein-tyrosine kinase (EC 2.7.1.112) fyn - Pacific electric ray
<a href="#">KPY1 HUMAN</a>	Pyruvate kinase, M1 isozyme (EC 2.7.1.40) (Pyruvate kinase muscle isozyme) (Cytosolic thyr-
<a href="#">I37984</a>	Keratin 9, type I, cytoskeletal - human
<a href="#">B26168</a>	ribophorin II precursor - human
<a href="#">Q9PVU9</a>	SRC-LIKE B (FRAGMENT).- Lethenteron reissneri.
<a href="#">TVHULY</a>	protein-tyrosine kinase (EC 2.7.1.112) lyn, splice form A - human
<a href="#">PAHUI</a>	alkaline phosphatase (EC 3.1.3.1) precursor, intestinal - human
<a href="#">Q9JL10</a>	NEURONAL C-SRC TYROSINE-SPECIFIC PROTEIN KINASE.- Rattus norvegicus (Rat).
<a href="#">Q9DK66</a>	SRC-FAMILY TYROSINE KINASE SCK.- Salmo salar (Atlantic salmon).
<a href="#">A46506</a>	leukocyte activation antigen M6 - human
<a href="#">A44861</a>	keratin, 67K type II epidermal - human
<a href="#">Q92957</a>	SRC TYROSINE KINASE.- Rous sarcoma virus.
<a href="#">Q95357</a>	PUTATIVE G PROTEIN-COUPLED RECEPTOR (CDNA FLJ10899 FIS, CLONE NT2RP5003506) (RETINOIC ACID
<a href="#">Q9H2K9</a>	CARBOXYPEPTIDASE M.- Homo sapiens (Human).
<a href="#">Q9UHR4</a>	INSULIN RECEPTOR TYROSINE KINASE SUBSTRATE.- Homo sapiens (Human).
<a href="#">I61771</a>	keratin 6f, type II - human
<a href="#">AAD45866</a>	AF099011 NID: - Homo sapiens
<a href="#">AAD05191</a>	MUSKTEP2A NID: - Mus musculus
<a href="#">Q9D2K8</a>	0 DAY NEONATE HEAD CDNA, RIKEN FULL-LENGTH ENRICHED LIBRARY, CLONE:4833436C19, FULL INSERT

Done Local intranet

In any search of a large LC-MS/MS dataset

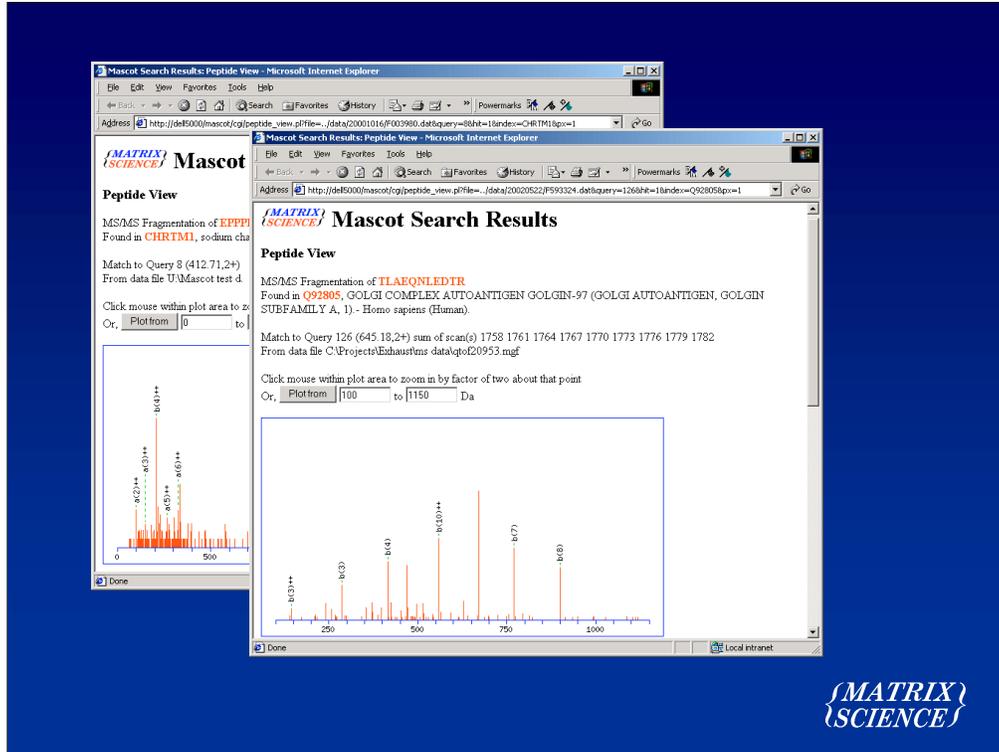
Mascot Search Results - Microsoft Internet Explorer

Address: http://del5000/mascot/cgi/master\_results.pl?file=../data/20001016/F003980.dat

Unassigned queries: (no details means no match)

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Rank	Peptide
<a href="#">20</a>	491.66	981.31	980.44	0.87	0	35	1	QMVVCCNK + Carbamidomethyl (C)
<a href="#">16</a>	464.68	927.35	926.41	0.94	0	29	1	FCNLGHR + Oxidation (M); Propionamide (C)
<a href="#">44</a>	604.15	1206.29	1205.65	0.63	0	24	1	EEGDYVLVK + 3 Methyl ester (DE)
<a href="#">4</a>	787.30	786.29	785.46	0.82	0	23	1	QVLAEVK
<a href="#">9</a>	825.34	824.33	824.51	-0.18	0	22	1	IYVPEIR
<a href="#">6</a>	403.66	805.29	804.41	0.89	0	21	1	MDLSPVK + Oxidation (M)
<a href="#">77</a>	711.73	1421.45	1420.77	0.67	2	20	1	FDAKLHLEKLR + Acetyl (N-term); Oxidation (M)
<a href="#">58</a>	659.17	1316.32	1315.61	0.71	0	20	1	WPLDFPDNK + Acetyl (N-term); 2 Methyl ester (DE)
<a href="#">19</a>	491.20	980.38	980.54	-0.15	0	20	1	LINSELMK + Oxidation (M)
<a href="#">24</a>	515.19	1028.36	1028.45	-0.09	0	18	1	GQDITGDDR + Acetyl (N-term); Methyl ester (DE)
<a href="#">86</a>	761.71	1521.41	1521.82	-0.41	0	18	1	FPAPFFVCLFSIK + Carbamidomethyl (C)
<a href="#">55</a>	642.70	1283.39	1283.70	-0.31	1	18	1	SPRYTHAIVAR + Acetyl (N-term)
<a href="#">167</a>	803.47	3209.84	3209.46	0.38	0	18	1	ERPDDGQSYEEQAAAAGGGGSAHQSGANEVK + 5 Methyl ester (DE);
<a href="#">114</a>	894.30	1786.59	1786.97	-0.37	0	17	1	TVDEVVSTGTDLIIVGR + Methyl ester (DE)
<a href="#">14</a>	431.65	861.29	860.37	0.92	0	17	1	EYNYYK + Pyro-glu (N-term E)
<a href="#">113</a>	893.28	1784.54	1784.96	-0.41	2	17	1	KRLNHPSPVFELEK + Acetyl (N-term)
<a href="#">62</a>	670.20	1338.38	1337.57	0.81	1	17	1	KDGSQCPCQCK + Carbamidomethyl (C); Methyl ester (DE); P
<a href="#">137</a>	739.52	2215.53	2215.97	-0.43	1	17	1	YWFYSTKCYFFIDMK + Carbamidomethyl (C)
<a href="#">151</a>	833.57	2497.69	2498.33	-0.64	2	16	1	MTVRIVSNVAVMALISGADDNVKR + Acetyl (N-term); Methyl ester
<a href="#">26</a>	521.20	1040.38	1040.41	-0.03	0	14	1	SIFDCGDDK + Acetyl (N-term)
<a href="#">39</a>	581.66	1161.31	1161.57	-0.26	0	14	1	QQLLENFEK + Methyl ester (DE); Pyro-glu (N-term Q)
<a href="#">35</a>	554.18	1106.35	1105.55	0.80	0	14	1	ESLCSVLDK + Acetyl (N-term); Carbamidomethyl (C); Methyl
<a href="#">136</a>	731.51	2191.52	2190.94	0.58	2	13	1	ECDLTRADEAERDAEAGR + Acetyl (N-term); 2 Methyl ester (I
<a href="#">165</a>	788.69	3150.71	3150.41	0.30	1	13	1	EDLRMLCICSVDPGCTDIDDLHCR + Acetyl (N-term); Carbamidom
<a href="#">61</a>	446.81	1337.42	1336.65	0.77	1	13	1	QGRVTWEEYR + Methyl ester (DE)
<a href="#">89</a>	764.71	1527.40	1526.87	0.54	0	13	1	STQLATLVLSLSIPR + Acetyl (N-term)
<a href="#">64</a>	453.17	1356.48	1355.73	0.75	2	13	1	SKVEVPAVRER
<a href="#">111</a>	592.87	1775.60	1775.76	-0.16	1	13	1	EREISDDAEPEDEK + Methyl ester (DE)
<a href="#">154</a>	855.90	2564.66	2565.17	-0.51	0	13	1	LCVWIDFPPDMSTAAASSSELEK + Carbamidomethyl (C)

There are always a number of spectra which get poor scores, or even no match at all.



Sometimes, this is because the spectra are empty, or contain little more than noise. However, some of the spectra may contain clear sequence ion ladders at good signal to noise

## Why do we fail to get a match?

### Things that shouldn't happen:

- Incorrect determination of precursor charge
- Underestimated mass measurement error

### Things that often happen:

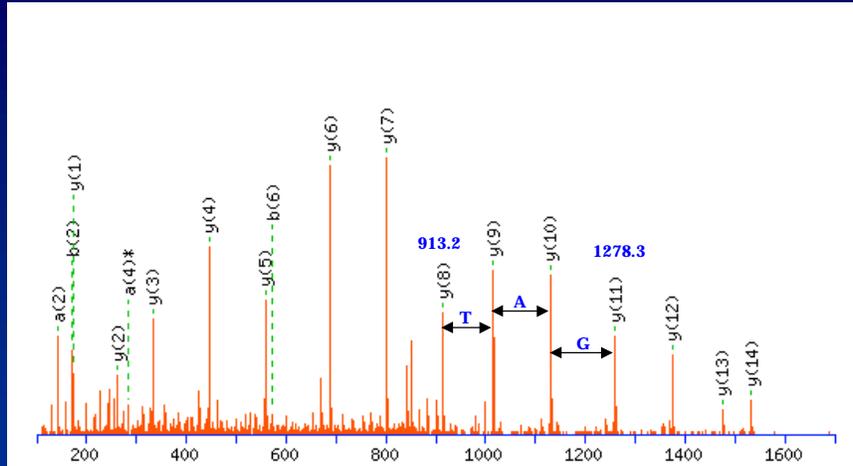
- Enzyme non-specificity
- Unsuspected chemical & post-translational modifications
- Peptide sequence not in the database.

*{MATRIX}*  
*{SCIENCE}*

Why do the good quality spectra fail to match? Some problems are easily remedied. For example, the precursor charge may have been called incorrectly, or the mass tolerance estimate may be over-optimistic.

Enzyme non-specificity is very common, and can be addressed with a no-enzyme search, but this is very time consuming.

The two other causes of failure to match are more difficult to deal with: Unsuspected modifications and variations in the primary sequence.



Mann, M., Wilm, M., *Anal Chem* 1994, 66, 4390-9

{MATRIX}  
{SCIENCE}

The best tool available for finding matches when there are unsuspected modifications or variations in the primary sequence is the error tolerant sequence tag, developed at EMBL. The standard tag combines an interpreted tag with the flanking fragment ion masses, the peptide mass, and the enzyme specificity.

By relaxing one or more of these constraints, the tag can accommodate enzyme non-specificity and / or unexpected mass differences to one side or the other of the tag.

## Error Tolerant Sequence Tag

- Rapid search times
- Requires good quality data
- Only a filter: no score
- Results need interpretation.

*{MATRIX}*  
*{SCIENCE}*

The error tolerant sequence tag is a very fast search, but the data quality has to be high enough to interpret a reliable tag. Also, the error tolerant sequence tag is less specific, and it becomes more likely that multiple hits will be found.

Since there is no score to indicate which hit is the best match, and whether the best match is significant, the user has to reconcile the spectrum to each of the candidate sequences, and decide which match to accept.

To address these limitations, we decided to investigate an error tolerant approach using Mascot to search uninterpreted data

## Error Tolerant Search of Uninterpreted MS/MS Data

### Enzyme non-specificity

- 'no-enzyme' search

### *In Silico* digestion of NCBI nr (850,000 entries)

- 6000 tryptic limit peptides 1500 Da  $\pm$  0.5
- 2,400,000 non-specific peptides 1500 Da  $\pm$  0.5

*{MATRIX}*  
*{SCIENCE}*

Searching a complete sequence database with no enzyme specificity takes much longer than the same search with (say) tryptic specificity. This is because there are between 100 and 1000 times as many peptides to be tested.

## Error Tolerant Search of Uninterpreted MS/MS Data

Unsuspected chemical & P-T modifications

- Iterate through comprehensive list.

*{MATRIX}*  
*{SCIENCE}*

For unsuspected modifications, it would be nice to try all possible mass values at all residue locations. Unfortunately, this doesn't work because all specificity is lost. We have chosen to iterate through a very comprehensive list of modifications

# Chemical & P-T Modifications

Name	Mono	Average	Site(s)	Short name	Hidden	Notes	C	H	N	O	P	S	Na	F	S
PEO Biotin	414.19370	414.519	C	PEO Biotin (C)			18	30	4	5	1				
Phosphorylation	79.96633	79.980	S,T	Phospho (ST)			1	3	1						
Phosphorylation Neutral loss	97.97690	97.995	NeutralLoss	Phospho (ST)			3	4	1						
Phosphorylation	79.96633	79.980	Y	Phospho (Y)			1	3	1						
Phosphorylation with neutral loss of H3PO4	-18.01057	-18.015	S,T	PhosphoNL (ST)	hidden		-2	-1							
Acrylamide adduct (propionamide)	71.03712	71.079	C	Propionamide (C)			3	5	1	1					
Pyridyl	119.03712	119.123	K	Pyridyl (K)	hidden		7	5	1	1					
Pyridyl	119.03712	119.123	Nterm	Pyridyl (N-term)	hidden		7	5	1	1					
Pyro-glu from E	-18.01057	-18.015	ResiduesNterm	Pyro-glu (N-term E)			-2	-1							
Pyro-glu from Q	-17.02655	-17.030	ResiduesNterm	Pyro-glu (N-term Q)			-3	-1							
SMA	127.06333	127.143	K	SMA (K)			6	9	1	2					
SMA	127.06333	127.143	Nterm	SMA (N-term)			6	9	1	2					
Sodium adduct	21.98194	21.982	Cterm	Sodated (C-term)			-1						1		
Sodium adduct	21.98194	21.982	E	Sodated (E)			-1							1	
S-pyridylethylolation	105.05785	105.139	C	S-pyridylethyl (C)			7	7	1						
Sulphone	31.98983	31.999	M	Sulphone (M)						2					
Citrullination	0.98402	0.985	R	Citrullination	hidden		-1	-1	1						
Methylation	14.01565	14.027	C,K,R,H,D,E,N	Methylation (res)	hidden		1	2							
Methylation	14.01565	14.027	ResiduesNterm	Methylation (term)	hidden		1	2							
Hydroxylation	15.99492	15.999	P,K,D,N	Hydroxylation	hidden				1						
di-Methylation	28.03130	28.054	C,K,R,H,D,E,N	di-Methylation (res)	hidden		2	4							
di-Methylation	28.03130	28.054	ResiduesNterm	di-Methylation (term)	hidden		2	4							
tri-Methylation	42.04695	42.080	C,K,R,H,D,E,N	tri-Methylation (res)	hidden		3	6							
tri-Methylation	42.04695	42.080	ResiduesNterm	tri-Methylation (term)	hidden		3	6							
Gamma-carboxylation	43.98983	44.010	D,E	Gamma-carboxylation	hidden		1		2						
Beta-methylthiolation	45.98772	46.087	D	Beta-methylthiolation	hidden		1	2					1		
Sulfation	79.95682	80.058	Y	Sulfation	hidden		1	3	1						
Phosphorylation	79.96633	79.980	H,C,D	Phosphorylation	hidden		1	3	1						
C-Mannosylation	162.05283	162.142	W	C-Mannosylation	hidden		6	10	5						
Glycation	162.05283	162.142	N,T,K	Glycation (res)	hidden		6	10	5						
Glycation	162.05283	162.142	Nterm	Glycation (term)	hidden		6	10	5						
Lipoyl	168.03286	168.302	K	Lipoyl	hidden		8	12	1						
O-GlcNac	203.07938	203.194	S,T,N	O-GlcNac	hidden		8	13	1	5					
Farnesylation	204.18781	204.355	C	Farnesylation	hidden		15	24							
Myristoylation	210.19837	210.359	K	Myristoylation	hidden		14	26	1						
Myristoylation	210.19837	210.359	ResiduesNterm	Myristoylation (term)	hidden		14	26	1						

*(MATRIX)  
(SCIENCE)*

The longer the list, the better. Unfortunately, there is no standard database of modifications that is suitable for MS use, so we had to compile this list ourselves

## Error Tolerant Search of Uninterpreted MS/MS Data

### Peptide sequence not in the database

- single base substitutions - Yes
- single base insertions & deletions - No (frame shift)
- multiple base substitutions, insertions & deletions - No (*de novo*).

*{MATRIX}*  
*{SCIENCE}*

Virtually all of the protein database entries come from translations of DNA or RNA. Hence, for variations in the primary sequence, it is essential to consider changes at the NA level.

We choose to allow all possible single NA substitutions, but not insertions or deletions, because these give rise to frame shifts.

We also draw the line at multiple variations in a single peptide. These must be handled by *de novo* sequencing

# Error Tolerant Search of Uninterpreted MS/MS Data

## Single base substitutions:

- Back translate AA to codons
- Perform all possible single base substitutions
- Translate resulting codons to AA.

*{MATRIX}*  
*{SCIENCE}*

The procedure to generate a substitution matrix for all possible single base substitutions

# Primary Sequence Variations

	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
A	1			1			1	1							1	1	1			1
R		1			1	1			1	1	1	1	1			1	1	1	1	
N			1							1	1		1				1	1		1
D	1		1	1			1	1	1											1
C		1			1				1						1		1		1	1
Q		1				1			1			1	1			1				
E	1			1		1	1		1				1							1
G	1	1		1	1		1	1									1		1	1
H		1	1	1			1					1				1				1
I		1	1							1		1	1	1			1	1		1
L		1			1				1	1	1			1	1	1		1		1
K		1	1			1	1			1		1						1		1
M			1							1	1	1	1					1		1
F					1					1	1			1						1
P	1	1					1			1					1	1		1		1
S	1	1	1		1			1		1	1			1	1	1	1	1	1	1
T	1	1	1							1		1	1		1	1	1			1
W		1			1			1			1							1		
Y			1	1	1				1						1		1			1
V	1			1			1	1		1	1		1	1						1

*{MATRIX}*  
*{SCIENCE}*

This matrix is considerably sparser than all possible residue substitutions

# Error Tolerant Search of Uninterpreted MS/MS Data

## Procedure

- Perform standard search
- Select one or more protein hits
- With no enzyme specificity, iterate through extensive list of modifications + substitution matrix.

*{MATRIX}*  
*{SCIENCE}*

The error tolerant search is a second pass search

Mascot Search Results - Microsoft Internet Explorer

Address: http://deli5000/mascot/cgi/master\_results.pl?file=../data/20020522/F593324.dat

6. [2TGD](#) Mass: 23177 Total score: 571 Peptides matched: 11  
 trypsin (EC 3.4.21.4) precursor (diisopropylphosphorylated) - bovine  
 Check to include this hit in error tolerant search or archive report

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Rank	Peptide
<input checked="" type="checkbox"/> <a href="#">6</a>	403.13	804.25	804.41	-0.16	0	65	1	SAASLNSR
<input checked="" type="checkbox"/> <a href="#">90</a>	577.17	1152.32	1152.57	-0.25	0	87	1	SSGTSYDPVLK
<input checked="" type="checkbox"/> <a href="#">93</a>	584.67	1167.33	1167.57	-0.25	0	97	1	VCHFVSWIK + Carbamidomethyl (C)
<input checked="" type="checkbox"/> <a href="#">100</a>	598.18	1194.34	1194.58	-0.24	0	(68)	1	SSGTSYDPVLK + Acetyl (N-term)
<input checked="" type="checkbox"/> <a href="#">173</a>	745.72	1489.43	1489.73	-0.31	0	78	1	LQGIVSWGSGCAQK + Carbamidomethyl (C)
<input checked="" type="checkbox"/> <a href="#">195</a>	805.17	1608.32	1608.65	-0.33	0	79	1	DSCQEDSGGPPVCSGK + 2 Carbamidomethyl (C)
<input checked="" type="checkbox"/> <a href="#">289</a>	1081.77	2161.52	2162.05	-0.53	0	138	1	LGEDHINVVVEGNEQFISASK
<input checked="" type="checkbox"/> <a href="#">290</a>	721.88	2162.60	2162.05	0.56	0	(87)	1	LGEDHINVVVEGNEQFISASK
<input checked="" type="checkbox"/> <a href="#">296</a>	1102.80	2203.59	2204.06	-0.47	0	(102)	1	LGEDHINVVVEGNEQFISASK + Acetyl (N-term)
<input checked="" type="checkbox"/> <a href="#">297</a>	735.54	2203.60	2204.06	-0.46	0	(46)	1	LGEDHINVVVEGNEQFISASK + Acetyl (N-term)
<input type="checkbox"/> <a href="#">305</a>	758.55	2272.63	2272.15	0.48	0	29	2	SIVHPSYNSNTLHNDIMLIK

Proteins matching the same set of peptides:

[1TAWA](#) Mass: 23276 Total score: 571 Peptides matched: 11  
 trypsin (EC 3.4.21.4), chain A - bovine

[BAA07516](#) Mass: 25408 Total score: 571 Peptides matched: 11  
 BOVPCP NID: - Bos taurus

[1AAU](#) Mass: 23290 Total score: 571 Peptides matched: 11  
 trypsin (EC 3.4.21.4) - bovine

[TRBOTB](#) Mass: 23978 Total score: 571 Peptides matched: 11  
 trypsin (EC 3.4.21.4) precursor - bovine

[1MCUE](#) Mass: 23203 Total score: 571 Peptides matched: 11  
 trypsin (EC 3.4.21.4), chain E - bovine

[1TGSZ](#) Mass: 23533 Total score: 571 Peptides matched: 11  
 trypsin (EC 3.4.21.4) precursor (with pancreatic secretory trypsin inhibitor), chain Z - bovine

[2TPIZ](#) Mass: 23021 Total score: 571 Peptides matched: 11  
 trypsin (EC 3.4.21.4) precursor (with basic proteinase inhibitor, Ile-Val, 2.4 M magnesium sulfate), chain Z - bovine

Here we see the matches to trypsin autolysis products from a standard search of an LC-MS/MS dataset from the analysis of a human cell lysate.

Mascot Search Results - Microsoft Internet Explorer

Address: http://dell5000/mascot/cgi/new\_master\_results.pl?file=../data/20020522/F593383.dat

3. **2TGD** Mass: 23177 Total score: 482 Peptides matched: 17  
 trypsin (EC 3.4.21.4) precursor (diisopropylphosphorylated) - bovine  
 Check to include this hit in error tolerant search or archive report

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss Score	Rank	Peptide
<a href="#">6</a>	403.13	804.25	804.41	-0.16	0	65	1 SAASLNSR
<a href="#">15</a>	431.63	861.24	861.43	-0.19	0 (51)	1	1 SAASLNSR + Carbamidomethyl (N-term) [+57.05]
<a href="#">74</a>	559.15	1116.29	1116.52	-0.23	0	56	1 SIVHPSYNSN
<a href="#">90</a>	577.17	1152.32	1152.57	-0.25	0	87	1 SSGTSYDPDLK
<a href="#">93</a>	584.67	1167.33	1167.57	-0.25	0	97	1 VCHYVSWIK + Carbamidomethyl (C)
<a href="#">100</a>	598.18	1194.34	1194.58	-0.24	0 (68)	1	1 SSGTSYDPDLK + Acetyl (N-term)
<a href="#">107</a>	606.19	1210.35	1210.55	-0.20	0 (57)	1	1 SSGTSYDPDLK + Acetyl (N-term); S->C [+16.06]
<a href="#">177</a>	745.72	1489.43	1489.73	-0.31	0	84	1 LQGIYSWGGCAQK + Carbamidomethyl (KR) [+57.05]

Top scoring peptide matches to query 107  
 (sum of scan(s) 1650 1653 1656 1659 1662 1665 1668 1671 1674 )  
 Status bar shows all hits for this peptide

Score	Delta	Hit	Protein	Peptide
57.4	-0.20	3	2TGD	SSGTSYDPDLK
57.4	-0.20	3	2TGD	SSGTSYDPDLK
44.2	0.77	3	2TGD	SSGTSYDPDLK
36.4	-0.23			KTSTFPGDV
34.0	-0.22	3	2TGD	SSGTSYDPDLK
28.5	0.84			GGGGGGGGSSGGGRGS
28.4	0.84			CGSGGGGGSSGGGRGS
26.1	-0.12			CGSGGGGGHGSYCSGS
21.2	-0.23			GARSMSGSGTIV
17.5	-0.15			SGYRGGGFSSGS

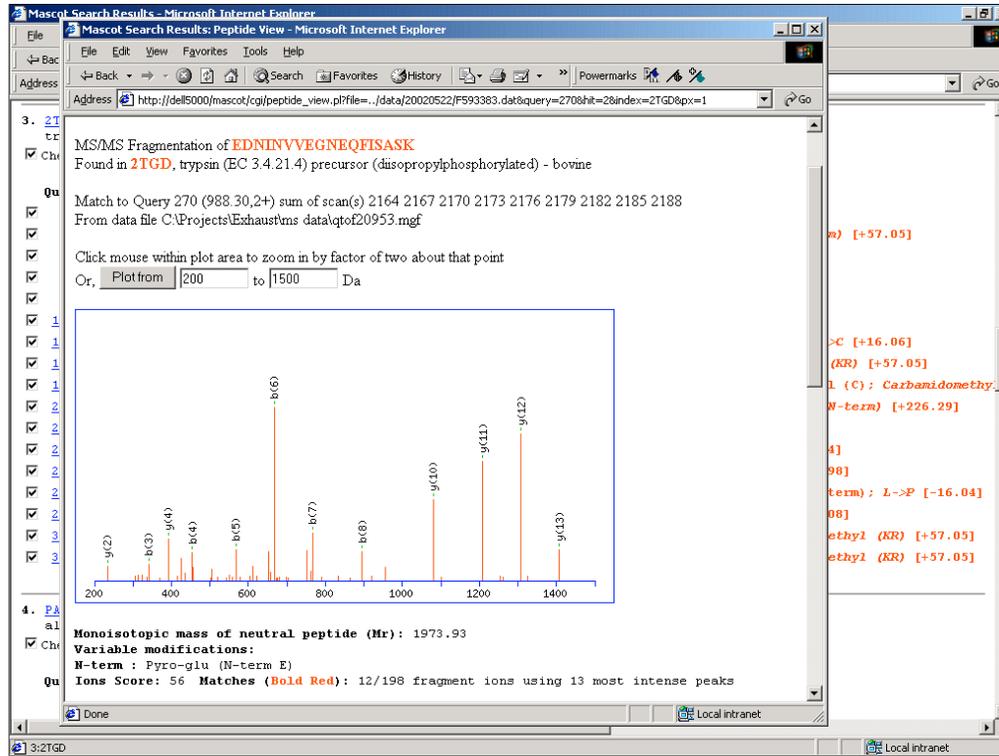
4. **PAH**  
 alk  
 Check to include this hit in error tolerant search or archive report

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss Score	Rank	Peptide
<a href="#">52</a>	517.18	1032.34	1032.56	-0.22	0	67	2 GSSIFGLAPSK + S->G [-30.03]

And here are the matches from the second pass, error tolerant search. The number of matches has increased from 10 to 17. Only those matches that were as good as or better than the original search are reported.

Query 74 is a simple, non-specific cleavage product

Query 107 shows acetylation at the N-terminus plus a mass increase of 16 at one of the N-term serines. It is not easy to find a reasonable mechanism to account for either an increase of 16 or an increase of 58 (42+16) at the N-term of this peptide



Query 270 is assigned to NINNVVEGNEQFISASK with a biotinylated N-term. However, this sample was not biotinylated, so this assignment cannot be correct. In fact, we get exactly the same mass increment by moving the cleavage point two residues further towards the N-terminus and adding a pyro-glu modification. This is a good illustration of how difficult it is for the report to always give the correct assignment

Mascot Search Results - Microsoft Internet Explorer

Address: http://deli5000/mascot/cgi/new\_master\_results.pl?file=.../data/20020522/F593383.dat

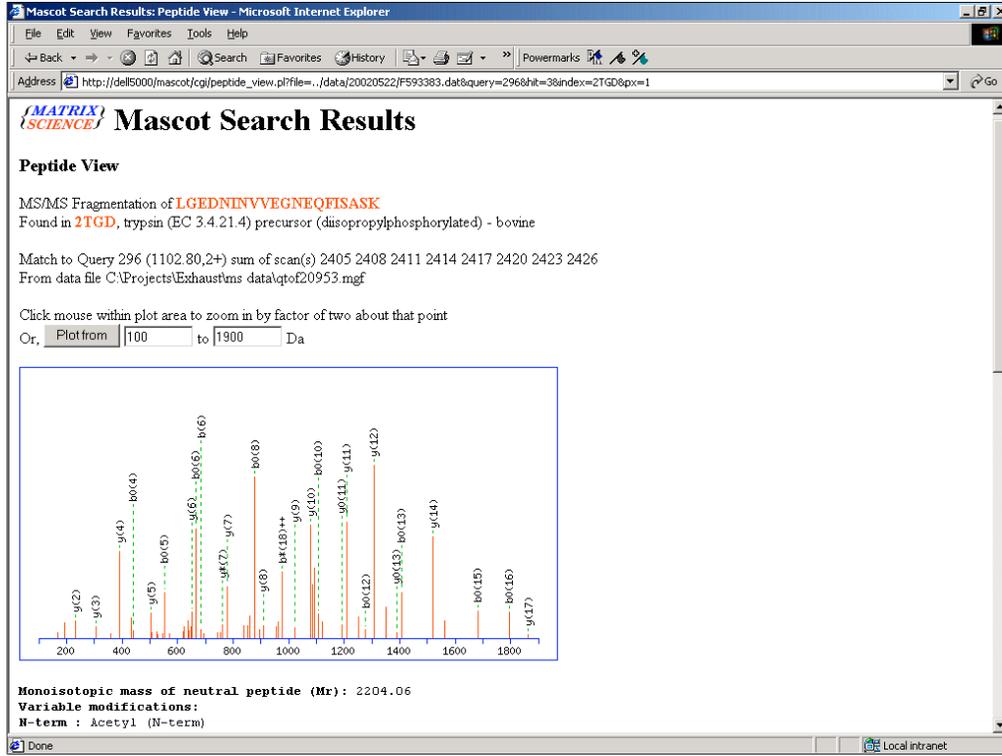
3. [2TGD](#) Mass: 23177 Total score: 482 Peptides matched: 17  
 trypsin (EC 3.4.21.4) precursor (diisopropylphosphorylated) - bovine  
 Check to include this hit in error tolerant search or archive report

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss Score	Rank	Peptide
<a href="#">6</a>	403.13	804.25	804.41	-0.16	0	65	SAASLNSR
<a href="#">15</a>	431.63	861.24	861.43	-0.19	0 (51)	1	SAASLNSR + Carbamidomethyl (N-term) [+57.05]
<a href="#">74</a>	559.15	1116.29	1116.52	-0.23	0	56	SIVHPSYNSN
<a href="#">90</a>	577.17	1152.32	1152.57	-0.25	0	87	SSGTSYDPDLK
<a href="#">93</a>	584.67	1167.33	1167.57	-0.25	0	97	VCHYVSWIK + Carbamidomethyl (C)
<a href="#">100</a>	598.18	1194.34	1194.58	-0.24	0 (68)	1	SSGTSYDPDLK + Acetyl (N-term)
<a href="#">107</a>	606.19	1210.35	1210.55	-0.20	0 (57)	1	SSGTSYDPDLK + Acetyl (N-term); S->C [+16.06]
<a href="#">173</a>	745.72	1489.43	1489.73	-0.31	0	84	LQGIVSWGSGCAQK + Carbamidomethyl (KR) [+57.05]
<a href="#">195</a>	805.17	1608.32	1608.65	-0.33	0	79	DSCQGDSGGPPVCSGK + Carbamidomethyl (C); Carbamidomethyl (KR) [+57.05]
<a href="#">270</a>	988.30	1974.58	1973.95	0.63	0	57	NINVVEGNEQFISASK + Biotinylated (N-term) [+226.29]
<a href="#">289</a>	1081.77	2161.52	2162.05	-0.53	0	138	LGEDNINVVEGNEQFISASK
<a href="#">290</a>	721.88	2162.60	2162.09	0.52	0 (92)	1	LGEDNINVVEGNEQFISASK + Q->K [+0.04]
<a href="#">293</a>	729.53	2185.58	2186.02	-0.44	0 (92)	1	LGEDNINVVEGNEQFISASK + L->H [+23.98]
<a href="#">294</a>	1094.81	2187.61	2188.03	-0.42	0 (94)	1	LGEDNINVVEGNEQFISASK + Acetyl (N-term); L->P [-16.04]
<a href="#">296</a>	1102.80	2203.59	2204.10	-0.50	0 (102)	1	LGEDNINVVEGNEQFISASK + G->V [+42.08]
<a href="#">300</a>	1110.30	2218.58	2219.07	-0.49	0 (105)	1	LGEDNINVVEGNEQFISASK + Carbamidomethyl (KR) [+57.05]
<a href="#">301</a>	740.54	2218.58	2219.07	-0.49	0 (93)	1	LGEDNINVVEGNEQFISASK + Carbamidomethyl (KR) [+57.05]

4. [PAHVI](#) Mass: 56776 Total score: 342 Peptides matched: 9  
 alkaline phosphatase (EC 3.1.3.1) precursor, intestinal - human  
 Check to include this hit in error tolerant search or archive report

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss Score	Rank	Peptide
<a href="#">52</a>	517.18	1032.34	1032.56	-0.22	0	67	GSSIFGLAPSK + S->G [-30.03]

There are several matches to the same peptide LGEDNINVVEGNEQFISASK with a variety of modifications. The first, Q->K, is identical to no modification. Although the match to Query 296 is shown as G->V, we would prefer to assign the mass change of 42 Da to N-term acetylation.



Looking at the details for this match

Mascot Search Results: Peptide View - Microsoft Internet Explorer

Address: http://deli5000/mascot/cgi/peptide\_view.pl?file=.../data/20020522/F593383.dat&query=296&hit=3&index=21GD&px=1

Monoisotopic mass of neutral peptide (Mr): 2204.06  
 Variable modifications:  
 N-term : Acetyl (N-term)  
 Ions Score: 102 Matches (Bold Red): 30/214 fragment ions using 43 most intense peaks

#	b	b <sup>++</sup>	b <sup>+</sup>	b <sup>+++</sup>	b <sup>0</sup>	b <sup>0++</sup>	Seq.	y	y <sup>++</sup>	y <sup>+</sup>	y <sup>+++</sup>	y <sup>0</sup>	y <sup>0++</sup>	#
1	156.10	78.56					L							20
2	213.12	107.07					G	2049.97	1025.49	2032.95	1016.98	2031.96	1016.49	19
3	342.17	171.59			324.16	162.58	E	1992.95	996.98	1975.92	988.47	1974.94	987.97	18
4	457.19	229.10			<b>439.18</b>	220.10	D	<b>1863.91</b>	932.46	1846.88	923.95	1845.90	923.45	17
5	571.24	286.12	554.21	277.61	<b>553.23</b>	277.12	N	1748.88	874.94	1731.86	866.43	1730.87	865.94	16
6	<b>684.32</b>	342.66	667.29	334.15	<b>666.31</b>	333.66	I	1634.84	817.92	1617.81	809.41	1616.83	808.92	15
7	798.36	399.69	781.34	391.17	<b>780.35</b>	390.68	N	<b>1521.75</b>	761.38	1504.73	752.87	1503.74	752.38	14
8	897.43	449.22	880.41	440.71	<b>879.42</b>	440.21	V	<b>1407.71</b>	704.36	1390.69	695.85	<b>1389.70</b>	695.35	13
9	996.50	498.75	979.47	490.24	<b>978.49</b>	489.75	V	<b>1308.64</b>	654.83	1291.62	646.31	1290.63	645.82	12
10	1125.54	563.28	1108.52	554.76	<b>1107.53</b>	554.27	E	<b>1209.58</b>	605.29	1192.55	596.78	<b>1191.56</b>	596.29	11
11	1182.56	591.79	1165.54	583.27	1164.55	582.78	G	<b>1080.53</b>	540.77	1063.51	532.26	1062.52	531.76	10
12	1296.61	648.81	1279.58	640.29	<b>1278.60</b>	639.80	N	<b>1023.51</b>	512.26	1006.48	503.75	1005.50	503.25	9
13	1425.65	713.33	1408.62	704.82	<b>1407.64</b>	704.32	E	<b>909.47</b>	455.24	892.44	446.72	891.46	446.23	8
14	1553.71	777.36	1536.68	768.84	1535.70	768.35	Q	<b>780.43</b>	390.72	<b>763.40</b>	382.20	762.42	381.71	7
15	1700.78	850.89	1683.75	842.38	<b>1682.77</b>	841.89	F	<b>652.37</b>	326.69	635.34	318.17	634.36	317.68	6
16	1813.86	907.43	1796.83	898.92	<b>1795.85</b>	898.43	I	<b>505.30</b>	253.15	488.27	244.64	487.29	244.15	5
17	1900.89	950.95	1883.87	942.44	1882.88	941.95	S	<b>392.21</b>	196.61	375.19	188.10	374.20	187.61	4
18	1971.93	986.47	1954.90	<b>977.96</b>	1953.92	977.46	A	<b>305.18</b>	153.10	288.16	144.58	287.17	144.09	3
19	2058.96	1029.98	2041.94	1021.47	2040.95	1020.98	S	<b>234.15</b>	117.58	217.12	109.06	216.13	108.57	2
20							K	147.11	74.06	130.09	65.55			1

We see that the fragment ions have a quantitative neutral loss of water. Is this behaviour well known for acetylated Leu?

Mascot Search Results - Microsoft Internet Explorer

Address: http://deli5000/mascot/cgi/new\_master\_results.pl?file=.../data/20020522/F593383.dat

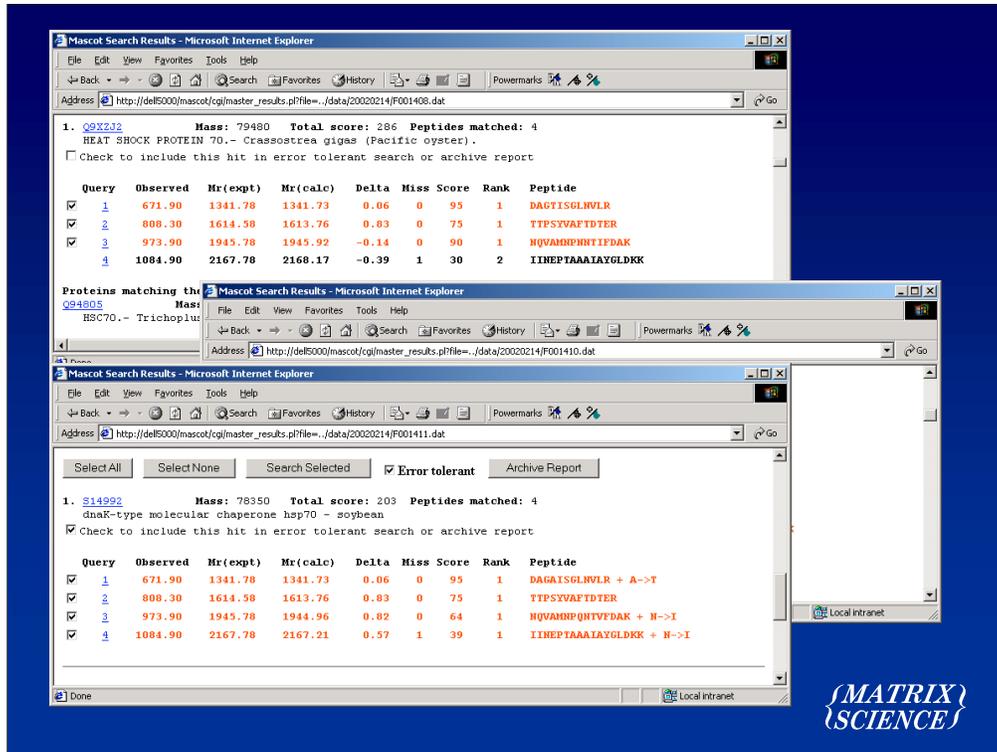
3. [2TGD](#) Mass: 23177 Total score: 482 Peptides matched: 17  
 trypsin (EC 3.4.21.4) precursor (diisopropylphosphorylated) - bovine  
 Check to include this hit in error tolerant search or archive report

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss Score	Rank	Peptide
<a href="#">6</a>	403.13	804.25	804.41	-0.16	0	65	SAASLNSR
<a href="#">15</a>	431.63	861.24	861.43	-0.19	0 (51)	1	SAASLNSR + Carbamidomethyl (N-term) [+57.05]
<a href="#">74</a>	559.15	1116.29	1116.52	-0.23	0	56	SIVHPSYNSN
<a href="#">90</a>	577.17	1152.32	1152.57	-0.25	0	87	SSGTSYDPDLK
<a href="#">93</a>	584.67	1167.33	1167.57	-0.25	0	97	VCHYVSWIK + Carbamidomethyl (C)
<a href="#">100</a>	598.18	1194.34	1194.58	-0.24	0 (68)	1	SSGTSYDPDLK + Acetyl (N-term)
<a href="#">107</a>	606.19	1210.35	1210.55	-0.20	0 (57)	1	SSGTSYDPDLK + Acetyl (N-term); S->C [+16.06]
<a href="#">173</a>	745.72	1489.43	1489.73	-0.31	0	84	LQGIVSWGSGCAQK + Carbamidomethyl (KR) [+57.05]
<a href="#">195</a>	805.17	1608.32	1608.65	-0.33	0	79	DSCQGDSSGPPVCSGK + Carbamidomethyl (C); Carbamidomethyl (KR) [+57.05]
<a href="#">270</a>	988.30	1974.58	1973.95	0.63	0	57	NIMVVEGNEQFISASK + Biotinylated (N-term) [+226.29]
<a href="#">289</a>	1081.77	2161.52	2162.05	-0.53	0	138	LGEDNIMVVEGNEQFISASK
<a href="#">290</a>	721.88	2162.60	2162.09	0.52	0 (92)	1	LGEDNIMVVEGNEQFISASK + Q->K [+0.04]
<a href="#">293</a>	729.53	2185.58	2186.02	-0.44	0 (92)	1	LGEDNIMVVEGNEQFISASK + L->H [+23.98]
<a href="#">294</a>	1094.81	2187.61	2188.03	-0.42	0 (94)	1	LGEDNIMVVEGNEQFISASK + Acetyl (N-term); L->P [-16.04]
<a href="#">296</a>	1102.80	2203.59	2204.10	-0.50	0 (102)	1	LGEDNIMVVEGNEQFISASK + G->V [+42.08]
<a href="#">300</a>	1110.30	2218.58	2219.07	-0.49	0 (105)	1	LGEDNIMVVEGNEQFISASK + Carbamidomethyl (KR) [+57.05]
<a href="#">301</a>	740.54	2218.58	2219.07	-0.49	0 (93)	1	LGEDNIMVVEGNEQFISASK + Carbamidomethyl (KR) [+57.05]

4. [PAHVI](#) Mass: 56776 Total score: 342 Peptides matched: 9  
 alkaline phosphatase (EC 3.1.3.1) precursor, intestinal - human  
 Check to include this hit in error tolerant search or archive report

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss Score	Rank	Peptide
<a href="#">52</a>	517.18	1032.34	1032.56	-0.22	0	67	GSSIFGLAPSK + S->G [-30.03]

This suggests that a possible assignment of +24 Da for query 293 is prompt loss of water:  $42-18=24$ ???



*MATRIX*  
*SCIENCE*

Finally, a slightly artificial illustration of finding a point mutation in the primary sequence. We can lose the match to query 1 by changing the taxonomy to 'green plants', which excludes both entries containing the correct peptide. If we now do an error tolerant search, the correct match is recovered via the substitution A->T

## Error Tolerant Search of Uninterpreted MS/MS Data

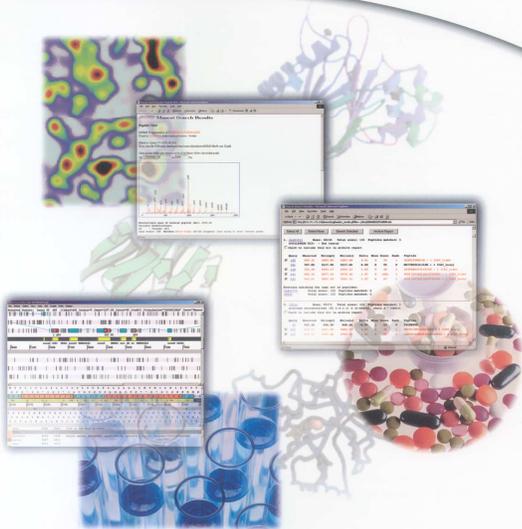
- Successfully locate mass differences
- Difficult to report correct assignment
- Limited to proteins which have at least one good peptide match ... not very useful for (say) MHC peptides
- Next step: hybrid of Sequence Tag & MS/MS Ions Search?

*{MATRIX}*  
*{SCIENCE}*

To summarise, the error tolerant search of uninterpreted MS/MS data is a powerful way of finding additional peptide matches. However, it is very difficult to provide a chemically credible assignment for many of the observed mass differences. Some expertise in mass spectrometry and protein chemistry is required to review and correct the reported assignments.

# MASCOT®

Take the guesswork out of protein identification...



*MATRIX*  
*SCIENCE*  
www.matrixscience.com