

# Modifications

MASCOT

 MATRIX  
SCIENCE

## Types of Modifications

### Post-translational

- Phosphorylation, acetylation

### Artefacts

- Oxidation, acetylation

### Derivatisation

- Alkylation of cysteine, ICAT, SILAC

### Sequence variants

- Errors, SNP's, other variants.



Modifications are a very important topic in database searching.

In some cases, the main focus of a study is to characterise post translational modifications, which may have biological significance. Phosphorylation would be a good example.

In other cases, the modification may not be of interest in itself, but you need to allow for it in order to get a match. Oxidation during sample preparation would be an example.

And, of course, many methods of quantitation involve modifications containing isotopic labels

Some sequence variants, such as the substitution of one residue by another, are equivalent to modifications, and can be handled in a similar way

Unimod - protein modifications for mass spectrometry

Unimod Logged as unimod Log out Change password Advanced search Help

Search for: Any field Contains Search Details found: 531 Page 14 of 27 Records Per Page: 20

Select/Unselect all Delete selected

	Accession #	PSI-MS Name	Interim name	Description	Monoisotopic mass	Average mass	Composition
Edit Copy View	40	Sulfo	Sulfation	O-Sulfonation	79.956815	80.0632	O(3) S
Edit Copy View	21	Phospho	Phospho	Phosphorylation	79.966331	79.9799	H O(3) P
Edit Copy View	549		Cys->Trp	Cys->Trp substitution	83.070128	83.0670	H(5) C(8) N S(-1)
Edit Copy View	211	NEIAA	NEIAA-d0	N-ethyl iodoacetamide-d0	85.052764	85.1045	H(7) C(4) N O
Edit Copy View	747		Malonyl	Malonylation of C and S residues	86.000394	86.0462	H(2) C(3) O(3)
Edit Copy View	371	HMVK	HMVK86	Michael addition of hydroxymethylvinyl ketone to cysteine	86.036779	86.0892	H(6) C(4) O(2)
Edit Copy View	324	DTBP	DTBP	dimethyl 3,3'-dithiobispropionimidate	87.014270	87.1435	H(5) C(3) N S
Edit Copy View	178	DAET	ser_thr_DAET	phosphorylation to amine thiol	87.050655	87.1866	H(9) C(4) N O(-1) S
Edit Copy View	379	Hypusine	hypusine	hypusine	87.068414	87.1204	H(9) C(4) N O
Edit Copy View	126	Thioacyl	DSP	thioacylation of primary amines (N-term and Lys)	87.998285	88.1283	H(4) C(3) O S
Edit Copy View	185	Label:13C(9)+Phospho	13C9_Phospho_Tyr	C13 label (Phosphotyrosine)	88.996524	88.9138	H C(-9) 13C(9) O(3) P
Edit Copy View	212	NEIAA:2H(5)	NEIAA-d5	N-ethyl iodoacetamide-d5	90.084148	90.1353	H(2) 2H(5) C(4) N O
Edit Copy View	724		O-Methylphosphate	O-Methylphosphorylation	93.981981	94.0065	H(3) C O(3) P

**MASCOT : Modifications** © 2007-2012 Matrix Science **MATRIX SCIENCE**

Comprehensive and accurate information about post translational and chemical modifications is an essential factor in the success of protein identification. In Mascot, we take our list of modifications from Unimod, which is an on-line modifications database.

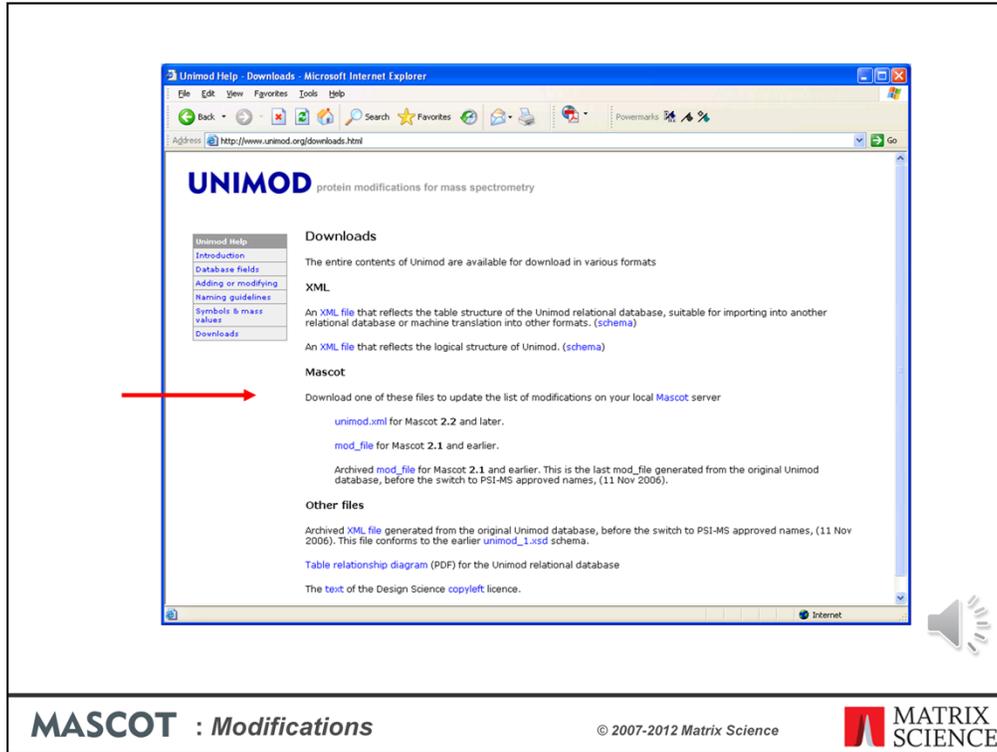
**UNIMOD** protein modifications for mass spectrometry  
 UNIMOD, View record [ Accession # : 56 ]

[Back to list](#)

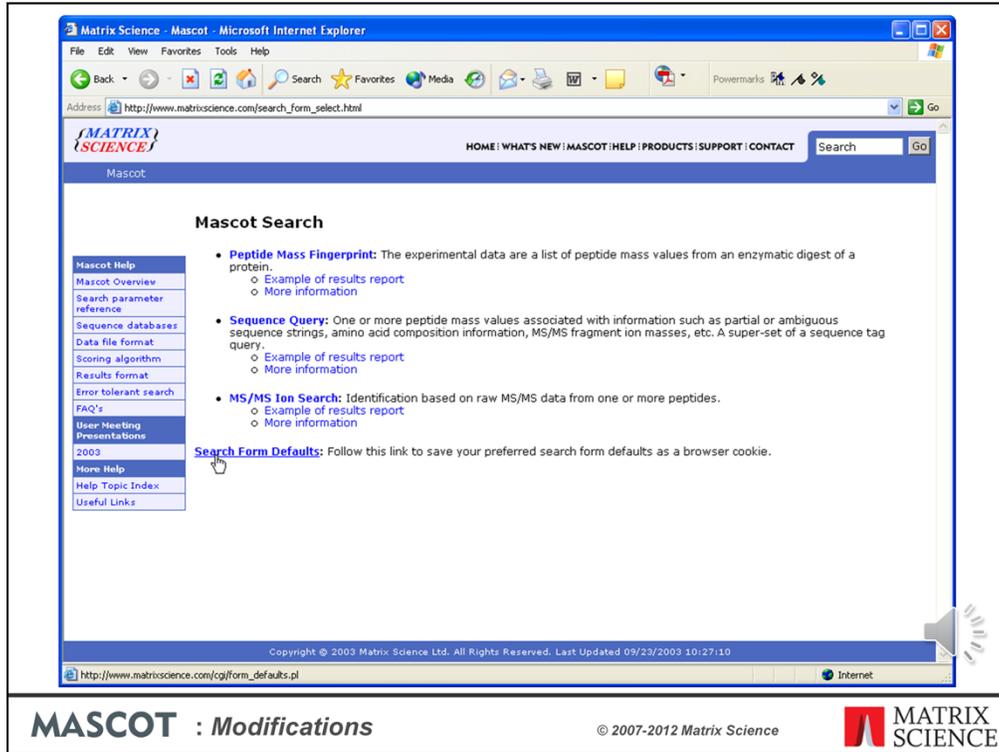
Accession #	56	PSI-MS Name	Acetyl:2H(3)	Interim Name	Acetyl_heavy
Description	Acetate labeling reagent (N-term & K) (heavy form, +3amu)				
Alt. Description	N-trideuteriumacetoxyl				
Composition	H(-1) 2H(3) C(2) O	Monoisotopic	45.029395	Average	45.0552
<b>Specificity Definition 1</b>					
Site	K	Position	Anywhere	Classification	Isotopic label
Hidden	1	Group	1		
<b>Specificity Definition 2</b>					
Site	N-term	Position	Any N-term	Classification	Isotopic label
Hidden	1	Group	2		
<b>Notes and References</b>					
Source	PubMed PMID	Reference	11857757		
Source	PubMed PMID	Reference	11999733		
Source	PubMed PMID	Reference	12175151		
Source	Journal	Reference	Controlling Deuterium isotope effects in comparative proteomics. Zhang, Roujan; Sioma, Cathy S.; Thompson, Robert A.; Xiong, Li; Regnier, Fred E. Department of Chemistry, Purdue University, West Lafayette, IN, USA. Analytical Chemistry (2006), 78(12), 3453-3458.		
Source	Journal	Reference	Global internal standard technology for comparative proteomics. Chakraborty, Asish; Regnier, Fred E. Department of Chemistry, Purdue University, West Lafayette, IN, USA. Journal of Chromatography, A (2002), 949(1-2), 173-184.		
Source	Journal	Reference	Comparative proteomics based on stable isotope labeling and affinity selection. Regnier, Fred E.; Riggs, Larry; Zhang, Roujan; Xiong, Li; Liu, Peiran; Chakraborty, Asish; Seeley, Erin; Sioma, Cathy; Thompson, Robert A. Department of Chemistry, Pu		
Curator	penner	Last Modified	2006-10-16 10:02:50	Verified	Yes

There are other lists of modifications on the web, like DeltaMass on the ABRF web site and RESID from the EBI, but none is as comprehensive as Unimod

Mass values are calculated from empirical chemical formulae, eliminating the most common source of error. Specificities can be defined in ways that are useful in database searching, and there is the option to enter mass-spec specific data, such as neutral loss information. This screen shot shows one of the better annotated entries, I can't pretend that all of them are this detailed. Nevertheless, it is a very useful, public domain resource that beats having to create your own list in an Excel spreadsheet or on the back of an envelope.



If you go to the help page, there is a link to download the contents of Unimod as a Mascot modifications file. This is the easiest way to keep the modifications list on an in-house Mascot server up-to-date



Here is a tip. The default list of modifications displayed in the Mascot search form is a short list, containing only the most common mods. If you want to see the complete list of mods, and you are using Mascot 2.2 or earlier, you need to follow the link at the bottom of the search form selection page

Matrix Science - Mascot - Set Search Defaults - Microsoft Internet Explorer

Address: [http://www.matrixscience.com/cgi/form\\_defaults.pl](http://www.matrixscience.com/cgi/form_defaults.pl)

HOME | WHAT'S NEW | MASCOT | HELP | PRODUCTS | SUPPORT | CONTACT

Mascot > Set Search Defaults

### Set Mascot search form defaults

Database: MSDB

Taxonomy: All entries

Enzyme: Trypsin

Allow up to: 1 missed cleavages

Fixed modifications: AB\_oldest\_ICATd0 (C), AB\_oldest\_ICATd8 (C), Acetyl (K), Acetyl (N-term), Amide (C-term)

Variable modifications: AB\_oldest\_ICATd0 (C), AB\_oldest\_ICATd8 (C), Acetyl (K), Acetyl (N-term), Amide (C-term)

Show all mods.

ICAT  (MS/MS only)

Peptide tol. ± 1.2 Da

MS/MS tol. ± 0.6 Da

Peptide charge: 1+

Monoisotopic:  Average

Data format: Mascot generic (MS/MS only)

**MASCOT : Modifications** © 2007-2012 Matrix Science **MATRIX SCIENCE**

Check the box for Show all mods, then choose Save. This still sets the default state of the checkbox in Mascot 2.3, but we decided to place the checkbox on the search form, so as to make it easier to swap between the short and long lists.

## Be sparing with variable modifications

### Some modifications are worse than others

- Mods that affect a terminus are less of a problem, e.g. Pyro-glu
- Mods that apply to residue(s) with a high fractional abundance and at any position are BIG problem, e.g. Phospho (ST) = 13%

### Use an error tolerant search to pick up uncommon modifications

- Efficient
- Also catch non-specific peptides



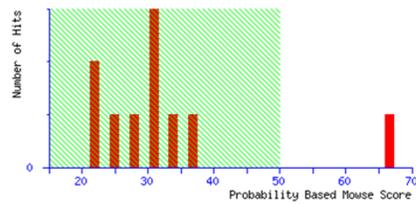
It is extremely important that you do not choose more than the absolute minimum number of variable modification in a search. We talked about this in an earlier presentation, but it is worth repeating.

Variable or differential or non-quantitative modifications are expensive, in the sense that they increase the time taken for a search and reduce its specificity. This is because the software has to permute out all the possible arrangements of modified and unmodified residues that fit to the peptide molecular mass. As more and more modifications are considered, the number of combinations and permutations increases geometrically. The so-called combinatorial explosion.

Some variable modifications are worse than others. Modifications that only apply to a terminus, especially if they only apply when particular residue is at the terminus, like pyro-glu, make little difference to the number of peptides to be tested. The problem modifications are the ones that apply to residues in any position, especially if they apply to multiple residues, like phosphorylation.

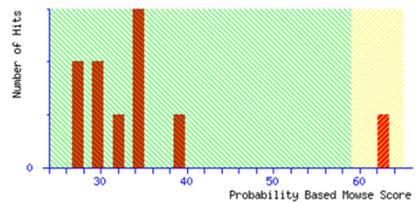
Unless you have enriched the sample in a particular PT-mod, e.g IMAC for phosphopeptides, it is usually not a good idea to try and catch PT-mods in a first pass search. Better to use a second pass search, which we call an error tolerant search, to catch the low abundance mods. We will come back to this later.

## Be sparing with variable modifications



Oxidation (M)

8 sec



Acetyl (K)  
Carbamidomethyl (C)  
Carboxymethyl (C)  
Me-ester (DE)  
Oxidation (M)  
Phospho (ST)  
Phospho (Y)  
Sodiated (DE)

92 sec

**MASCOT** : Modifications

© 2007-2012 Matrix Science



To illustrate this point. This search of a single MS/MS spectrum, using one variable mod, gives a nice, statistically significant match.

If the search is repeated with 8 mods, the match is the same, with an identical score, but now it is barely significant.

All of these mods have effectively increased the size of the database by a factor of 30

What's worse, the search takes over 10 times as long!

So, use variable mods sparingly. You'll get better results and faster.

By the way, the yellow region in the histogram indicates scores above the homology but below the identity thresholds. You will only see these regions highlighted in an MS/MS search report if it is a search of a single spectrum.

## Why is phosphorylation such a challenge?

**Site heterogeneity**

**Poor ionisation efficiency**

**3 fragmentation channels**

- intact fragments
- neutral loss of  $\text{HPO}_3$  (80 Da)
- neutral loss of  $\text{H}_3\text{PO}_4$  (98 Da)

**Can occur at STY - ~16% of residues.**



Of all post-translational modifications, phosphorylation is one of the most interesting and also one of the most difficult. Why is it such a challenge?

**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: 0 | 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

1. [CASP\\_BOVIN](#) Mass: 25091 Score: 88 Matches: 1(1) Sequences: 1(1)  
 Beta-casein OS=Bos taurus GN=CSN2 PE=1 SV=2  
 Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Unique	Peptide
<input checked="" type="checkbox"/>	1031.4000	2060.7854	2060.8212	-0.0357	0	88	1.6e-06	1	U	K.FQSEEQQTDELQDK.I + Phosp

Top scoring peptide matches to query 1  
 Score greater than 39 indicates homology  
 Score greater than 43 indicates identity

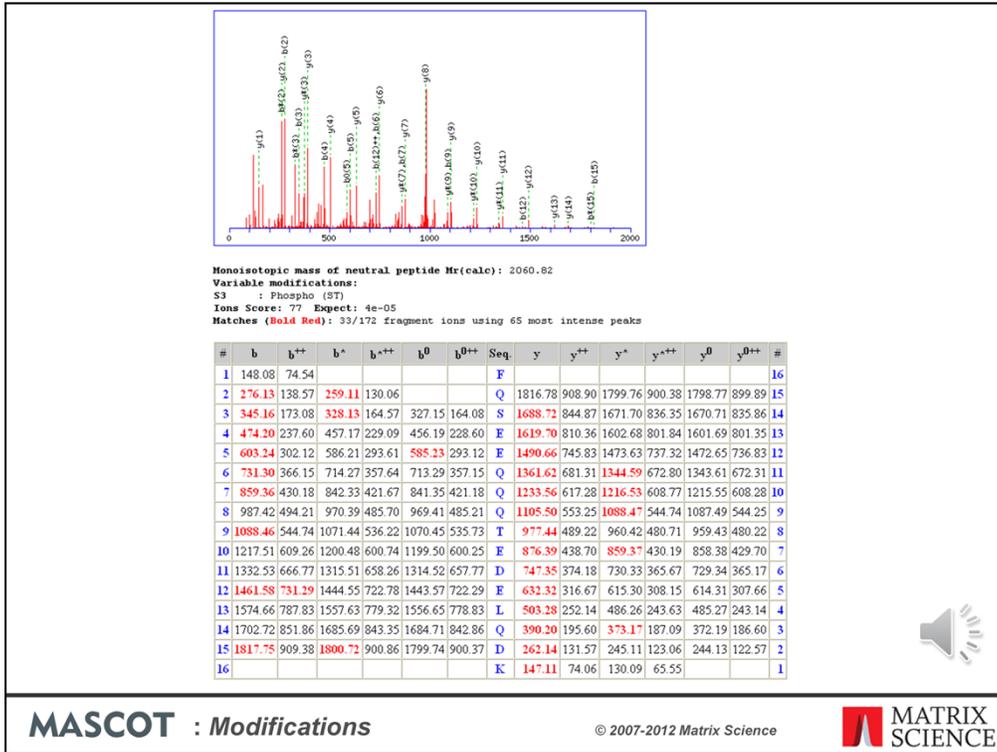
Score	Expect	Delta	Hit	Protein	Peptide
88.3	1.6e-06	-0.0357	1	CASP_BOVIN	K.FQSEEQQTDELQDK.I
28.6	1.5	-0.0357	1	CASP_BOVIN	K.FQSEEQQTDELQDK.I
21.0	8.8	-0.1886			K.CLSLSKQVDFEETIEK.H
15.9	28	-0.0907			K.QMVDKDFPVEPEDEK.G
14.1	42	-0.1713			K.QLAGGEYFLNQEQRK.R
13.6	47	-0.1489			K.IIFLEELYFKDQNEK.S
12.8	58	-0.1366			K.SSSQIPTQPPVTKSPYK.G
12.3	64	-0.2007			K.SLQEGGDLVAEDRLSEK.R
11.9	71	-0.1635			K.YLILCVGELTNERDEK.T
11.5	78	-0.0655			R.QPFLDILLQAKSENK.D

Search Parameters: Nonisotopic

MASCOT : Modifications | © 2007-2012 Matrix Science | MATRIX SCIENCE

Lets look at an example or two.

One of the most common phosphopeptides comes from the milk protein, beta casein. There are two potential phosphorylation sites, S and T, but only one is modified. Because the two sites are widely separated, the two arrangements get very different scores.



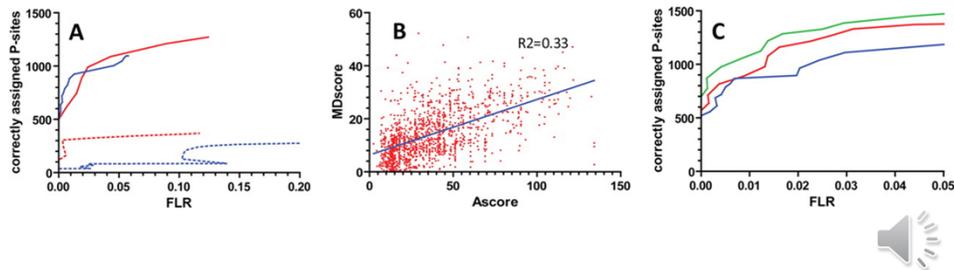
Beautiful spectrum; long run of y ions; move site to T9 and many matches would disappear

## Confident Phosphorylation Site Localization Using the Mascot Delta Score<sup>®</sup>

Mikhail M. Savitski<sup>‡</sup>, Simone Lemeer<sup>§</sup>, Markus Boesche<sup>‡</sup>, Manja Lang<sup>‡</sup>,  
Toby Mathieson<sup>‡</sup>, Marcus Bantscheff<sup>‡||</sup>, and Bernhard Kuster<sup>§||</sup>

*Molecular & Cellular Proteomics* 10.2

10.1074/mcp.M110.003830-1



**MASCOT**

**MATRIX  
SCIENCE**

Mascot 2.4 reports site localisation probabilities using the delta score method published in MCP by Bernhard Kuster's group. They analysed a collection of synthetic analogs of real phosphopeptides and determined what score difference was required to determine the correct site with an error rate of (say) 5%. Because we don't expect everyone to calibrate their data in this way, we have made the calculation slightly more conservative. A score difference of 10 would give approximately 90% probability that the higher scoring arrangement was correct.

Peptide Summary Report (..) | Mascot Search Results: Peptide

www.matrixscience.com/cgi/peptide\_view.pl?file=...%2Fdata%2F20120703%2FFTgmbcEO.dat&query=18&hit=18&index

Error (ppm)

RTS error: 60 ppm

Mass (Da)

NCBI BLAST search of [FQSEEQQOTEDELODK](#)  
 (Parameters: blastp, nr protein database, expect=20000, no filter, PAM30)  
 Other BLAST [web gateways](#)

All matches to this query

Score	M <sub>r</sub> (calc)	Delta	Sequence	Site Analysis
88.3	2060.8212	-0.0357	<a href="#">FQSEEQQOTEDELODK</a>	Phospho S3 100.00%
28.6	2060.8212	-0.0357	<a href="#">FQSEEQQOTEDELODK</a>	Phospho T9 0.00%
21.0	2060.9741	-0.1886	<a href="#">CLSLKQVDLFEETEK</a>	
15.9	2060.8762	-0.0907	<a href="#">QMVVDKDSPHVEPEDEK</a>	
14.1	2060.9568	-0.1713	<a href="#">QLASGEYFLNQEQRQAK</a>	
13.6	2060.9343	-0.1489	<a href="#">ITFLLELYPKDQDNEK</a>	
12.8	2060.9221	-0.1366	<a href="#">SSSQIPTOPPVTKSPYQK</a>	
12.3	2060.9862	-0.2007	<a href="#">SLOEGEGDLSVAEDRLSEK</a>	
11.9	2060.9489	-0.1635	<a href="#">YLILCVGETLNEDSEK</a>	
11.5	2060.8509	-0.0655	<a href="#">QDFLDLILSAKSENTK</a>	

Mascot: <http://www.matrixscience.com/>

**MASCOT : Modifications** © 2007-2012 Matrix Science **MATRIX SCIENCE**

A very large score difference such as the one we were just looking at gives 100% likelihood that the phosphate is on S3.

**Peptide Summary Report**

Format As: Peptide Summary | Help

Significance threshold p < 0.05 | Max. number of hits: 20

Standard scoring:  MudPIT scoring | Ions score or expect cut-off: 0 | 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

1. [KAPCA\\_BOVIN](#) Mass: 40594 Score: 79 Matches: 1(1) Sequences: 1(1)  
 cAMP-dependent protein kinase catalytic subunit alpha OS=Bos taurus GN=PRKACA PE=1 SV=3  
 Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Unique	Peptide
<input checked="" type="checkbox"/>	1107.9039	2213.7933	2214.0683	-0.2750	0	80	8.5e-06	1	U	R.TWTLCGTPEYLAPEILSK.G + Ph

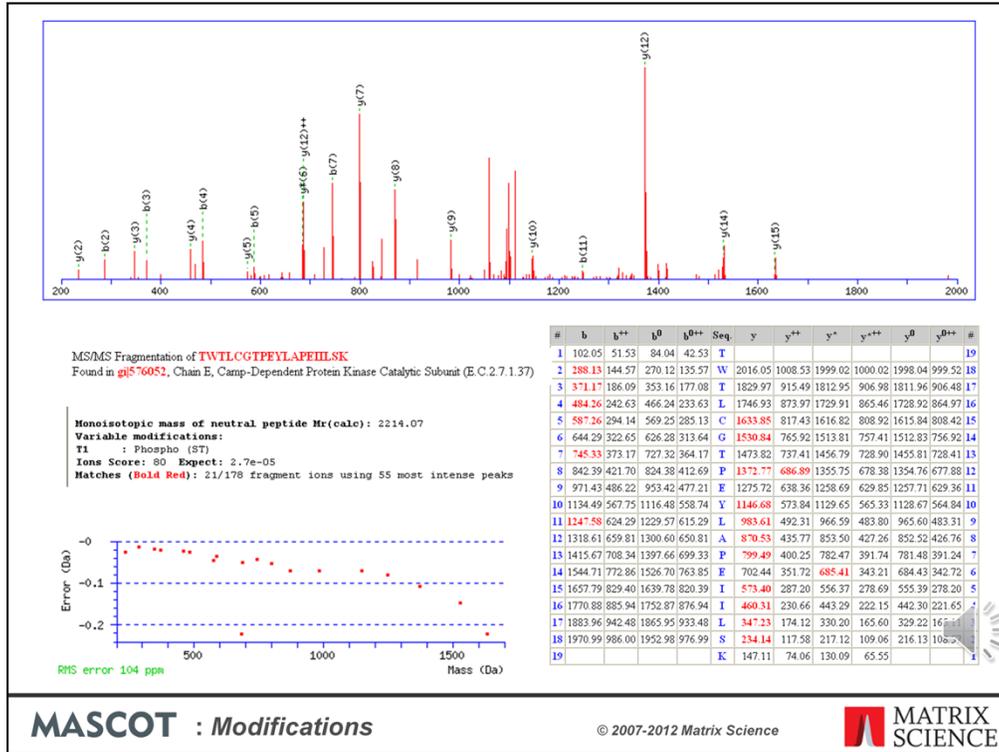
Top scoring peptide matches to query 1  
 Score greater than 30 indicates homology  
 Score greater than 42 indicates identity

Prot	Score	Expect	Delta	Hit	Protein	Peptide	
<a href="#">KAPC</a>	80.4	8.5e-06	-0.2750	1	KAPCA_BOVIN	R.TWTLCGTPEYLAPEILSK.G	PRKACA PE=2 SV=3
<a href="#">cAMP</a>	76.9	1.9e-05	-0.2750	1	KAPCA_BOVIN	R.TWTLCGTPEYLAPEILSK.G	N=PRKACA PE=2 SV=2
<a href="#">cAMP</a>	38.7	0.13	-0.2750	1	KAPCA_BOVIN	R.TWTLCGTPEYLAPEILSK.G	
<a href="#">KAPC</a>	18.0	15	-0.2750	1	KAPCA_BOVIN	R.TWTLCGTPEYLAPEILSK.G	
<a href="#">cAMP</a>	12.6	51	-0.2111	3	GSA_XYLFT	K.GGSGHLTGLIPSSPGVPAELSK.L	CA PE=1 SV=2
<a href="#">KAPC</a>	12.6	51	-0.2111	3	GSA_XYLFT	K.GGSGHLTGLIPSSPGVPAELSK.L	
<a href="#">cAMP</a>	12.6	51	-0.2111	2	GSA_XYLFH	K.GGSGHLTGLIPSSPGVPAELSK.L	ca PE=1 SV=3
<a href="#">cAMP</a>	12.6	51	-0.2111	2	GSA_XYLFH	K.GGSGHLTGLIPSSPGVPAELSK.L	
<a href="#">cAMP</a>	11.9	61	-0.2111	3	GSA_XYLFT	K.GGSGHLTGLIPSSPGVPAELSK.L	PE=1 SV=4
<a href="#">cAMP</a>	11.9	61	-0.2111	3	GSA_XYLFT	K.GGSGHLTGLIPSSPGVPAELSK.L	

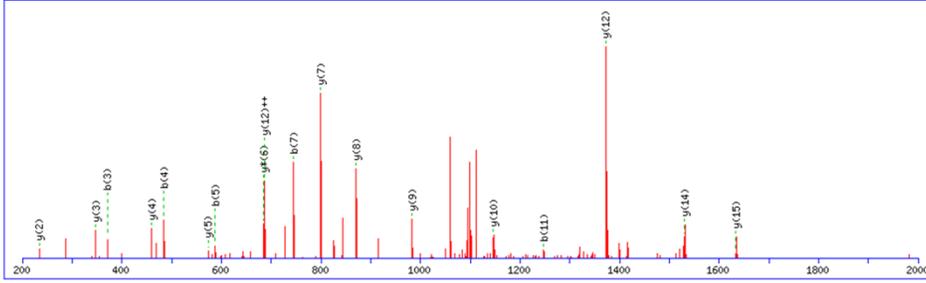
[KAPCA\\_BOVIN](#) Mass: 40594 Score: 79 Matches: 1(1) Sequences: 1(1)

However, casein peptides are unusually easy to analyse. Here is a more typical example of what you can expect to find - a strong match to a phosphopeptide from a protein kinase.

There is little to choose in terms of score between having the phosphate on T1 or T3.

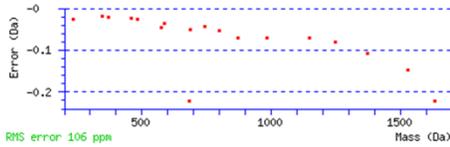


We can see why there is little difference in score between placing the phosphate on T1 or T3. There is just one extra matched peak, and in probability terms, there isn't a huge difference between 20 matches using 55 experimental peaks and 21. However, if you had to choose one or the other, you'd probably go for T1



MS/MS Fragmentation of **TWTLCGTPEYLAPHLK**  
 Found in **gi576052**, Chain E, Camp-Dependent Protein Kinase Catalytic Subunit (E.C.2.7.1.37)

Monoisotopic mass of neutral peptide H<sub>r</sub>(calc): 2214.07  
 Variable modifications:  
 T3 : Phospho (ST)  
 Ions Score: 77 Expect: 6.1e-05  
 Matches (Bold Red): 20/178 fragment ions using 55 most intense peaks



#	b	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	84.04	42.53	66.03	33.52	T							19
2	270.12	135.57	252.11	126.56	W	2034.06	1017.53	2017.04	1009.02	2016.05	1008.53	18
3	<b>371.17</b>	186.09	353.16	177.08	T	1847.98	924.49	1830.96	915.98	1829.97	915.49	17
4	<b>484.26</b>	242.63	466.24	233.63	L	1746.93	873.97	1729.91	865.46	1728.92	864.97	16
5	<b>587.26</b>	294.14	569.25	285.13	C	<b>1633.85</b>	817.43	1616.82	808.92	1615.84	808.42	15
6	644.29	322.65	626.28	313.64	G	<b>1530.84</b>	765.92	1513.81	757.41	1512.83	756.92	14
7	<b>745.33</b>	373.17	727.32	364.17	T	1473.82	737.41	1456.79	728.90	1455.81	728.41	13
8	842.39	421.70	824.38	412.69	P	<b>1372.77</b>	<b>686.89</b>	1355.75	678.38	1354.76	677.88	12
9	971.43	486.22	953.42	477.21	F	1275.72	638.36	1258.69	629.85	1257.71	629.36	11
10	1134.49	567.75	1116.48	558.74	Y	<b>1146.68</b>	573.84	1129.65	565.33	1128.67	564.84	10
11	<b>1247.58</b>	624.29	1229.57	615.29	L	<b>983.61</b>	492.31	966.59	483.80	965.60	483.31	9
12	1318.61	659.81	1300.60	650.81	A	<b>870.53</b>	435.77	853.50	427.26	852.52	426.76	8
13	1415.67	708.34	1397.66	699.33	P	<b>799.49</b>	400.25	782.47	391.74	781.48	391.24	7
14	1544.71	772.86	1526.70	763.85	F	702.44	351.72	<b>685.41</b>	343.21	684.43	342.72	6
15	1657.79	829.40	1639.78	820.39	I	<b>573.40</b>	287.20	556.37	278.69	555.39	278.20	5
16	1770.88	885.94	1752.87	876.94	I	<b>460.31</b>	230.66	443.29	222.15	442.30	221.65	4
17	1883.96	942.48	1865.95	933.48	L	<b>347.23</b>	174.12	330.20	165.60	329.22	165.11	3
18	1970.99	986.00	1952.98	976.99	S	<b>234.14</b>	117.58	217.12	109.06	216.13	108.57	2
19					K	147.11	74.06	130.09	65.55			1

**MASCOT** : Modifications

© 2007-2012 Matrix Science



Peptide Summary Report (..) Mascot Search Results: Pepti...

www.matrixscience.com/cgi/peptide\_view.pl?file=...%2Fdata%2F20120704%2FTGmifew.T.dat&query=18&hit=18&index

RTS error 62 ppm Mass (Da)

Error (ppm)

RTS error 62 ppm Mass (Da)

NCBI BLAST search of [TWTLCGTPEYLAPHLISK](#)  
 (Parameters: blastp, nr protein database, expect=20000, no filter, PAM30)  
 Other BLAST [web gateways](#)

All matches to this query

Score	Mr(calc)	Delta	Sequence	Site Analysis
80.4	2214.0683	-0.2750	<a href="#">TWTLCGTPEYLAPHLISK</a>	Phospho T1 69.17%
76.9	2214.0683	-0.2750	<a href="#">TWTLCGTPEYLAPHLISK</a>	Phospho T3 30.83%
38.7	2214.0683	-0.2750	<a href="#">TWTLCGTPEYLAPHLISK</a>	Phospho T7 0.00%
18.0	2214.0683	-0.2750	<a href="#">TWTLCGTPEYLAPHLISK</a>	Phospho Y10 0.00%
12.6	2214.0044	-0.2111	<a href="#">GGSGMLTLGIPSSPGVPAELSK</a>	
12.6	2214.0044	-0.2111	<a href="#">GGSGMLTLGIPSSPGVPAELSK</a>	
12.6	2214.0044	-0.2111	<a href="#">GGSGMLTLGIPSSPGVPAELSK</a>	
12.6	2214.0044	-0.2111	<a href="#">GGSGMLTLGIPSSPGVPAELSK</a>	
12.6	2214.0044	-0.2111	<a href="#">GGSGMLTLGIPSSPGVPAELSK</a>	
11.9	2214.0044	-0.2111	<a href="#">GGSGMLTLGIPSSPGVPAELSK</a>	
11.9	2214.0044	-0.2111	<a href="#">GGSGMLTLGIPSSPGVPAELSK</a>	

Mascot: <http://www.matrixscience.com/>

**MASCOT : Modifications** © 2007-2012 Matrix Science **MATRIX SCIENCE**

The delta score site analysis suggests 70% probability on T1 and 30% on T3 ... much less clear cut. We can't be confident which site is modified, or whether there is a mixture of both isoforms. But, we can be confident it is not on T7 or Y10 because the score drops dramatically, and these are assigned 0% probability.

Sometimes, it is worth looking at the sequence annotations to see whether these are known phosphorylation sites. If the database sequence doesn't have detailed annotations, you can follow the BLAST link to try and match the peptide to an entry from a better annotated database. In this case, we're searching SwissProt, so we can go straight to the protein view report

The screenshot displays a web browser window with the URL [www.matrixscience.com/cgi/protein\\_view.pl?file=...%2Fdata%2F20120704%2FTGmIfewT.dat&hit=KAPCA\\_BOVIN8.db](http://www.matrixscience.com/cgi/protein_view.pl?file=...%2Fdata%2F20120704%2FTGmIfewT.dat&hit=KAPCA_BOVIN8.db). The page content includes a list of modifications and their frequencies:

FT	MOD_RES	140	140	Phosphoserine (By similarity).
FT	MOD_RES	196	196	Phosphothreonine (By similarity). = T1
FT	MOD_RES	198	198	Phosphothreonine: by PDPK1.
FT	MOD_RES	202	202	Phosphothreonine (By similarity). = T3
FT	MOD_RES	339	339	Phosphoserine.
FT	LIPID	2	2	N-myristoyl glycine.
FT	MUTAGEN	3	3	N->D: No myristoylation.
FT	CONFLICT	202	202	T -> N (in Ref. 4; AA sequence).
FT	CONFLICT	204	204	E -> Q (in Ref. 4; AA sequence).
FT	CONFLICT	206	206	L -> S (in Ref. 4; AA sequence).
FT	CONFLICT	287	287	N -> D (in Ref. 2; AA sequence and 3; AA sequence).
FT	HELIX	16	32	
FT	HELIX	41	43	
FT	STRAND	44	52	
FT	STRAND	54	63	
FT	TURN	64	66	
FT	STRAND	69	76	
FT	HELIX	77	82	
FT	HELIX	86	96	
FT	STRAND	107	112	
FT	STRAND	114	122	
FT	HELIX	129	136	
FT	HELIX	141	160	
FT	HELIX	170	172	
FT	STRAND	173	175	
FT	STRAND	181	183	
FT	HELIX	203	205	
FT	HELIX	208	211	
FT	HELIX	219	234	
FT	HELIX	244	253	
FT	HELIX	264	273	
FT	TURN	286	289	
FT	HELIX	290	293	
FT	HELIX	296	298	
FT	HELIX	303	307	
FT	HELIX	346	348	

At the bottom, the protein sequence is shown with modifications T1 and T3 highlighted in red:

```

SQ SEQUENCE 351 AA: 40620 MW: 5955227D2DEE5D CRC64:
KQIAAAKKG SQEVSKEPL AKAKEDFLKK WENPAQNTAH LDFPERIKTL GTCSPGRVNL
VKHMETGNHY AHKILDGKRV VKLKQIEHTL NEKRILQAVN PFLVLEKFS FKDNENLHYV
MEYVPGGEMF SHLRDGRFS EDNADRYIQ TLTTFEVLHS LDLIVRDLKP ENLLIDQQGY
IQUTDFGFAK RVKGTWTLC GTPHYLAPEI ILSKPYNKAV DWALGVLIY EMAAGYPPFF
ADQPIQITEK IVSGRWKFFS HSDGRRDL RDELVDLTK RFGNLENGVN DIRNHWKVFAT
TDWAIYQQRK VEAFFIPPKK GPGDTSNFDV YEZEIIRVSI NEKCKEKFSE F

```

According to Swissprot, both T1 and T3 are possible phosphorylation sites. If you really needed to know which was the case here, or whether it was a mixture, you'd have to acquire more data. Maybe try a different enzyme or target the incomplete cleavage peptide that includes the preceding KG so as to move the sites towards the centre of the peptide, where you might get stronger b and y fragments

## Site Analysis

- If alternative sites differ by 20 in score, safe-ish to disregard lower one(s)
- If alternative sites have similar scores, you may be able to choose a preferred site by inspection
- Often, you just can't differentiate between closely spaced sites, even with great data.



If you are using Mascot 2.3 or earlier, the delta score calculation is not performed in Peptide View. These are our suggested guidelines when using Mascot for site analysis:

If alternative sites differ by 20 in score, safe-ish to disregard lower one(s)

If alternative sites have similar scores, you may be able to choose one by inspection. But, be careful ... one peak is just one peak

Often, you just can't differentiate between adjacent sites, even with great data.

## Error Tolerant Search

### First pass - simple search of entire database

- Minimal modifications
- Enzyme specificity

### Second pass - exhaustive search of selected protein hits

- Wide range of modifications
- Look for SNPs
- Relax enzyme specificity

### Reference

➤ Creasy, D. M. and Cottrell, J. S., Error tolerant searching of uninterpreted tandem mass spectrometry data, *Proteomics* 2 1426-1434 (2002)



Now, back to the challenge of finding PT modifications. There are many hundreds of modifications in Unimod, yet I've emphasised the importance of using the minimum number of variable modifications in a search. So, how are we supposed to find unusual modifications?

If you are searching uninterpreted MS/MS data, the efficient way to find unusual modifications, as well as variations in the primary sequence, is a two pass search. The first pass search is a simple search of the entire database with minimal modifications. The protein hits found in the first pass search are then selected for an exhaustive second pass search. During this second pass search, we can look for all possible modifications, sequence variants, and non-specific cleavage products.

Because only a handful of entries are being searched, search time is not an issue. It would be extremely difficult to calculate meaningful statistics for the additional matches in an error tolerant search, and we don't report expect values. The evidence for the presence of any particular protein are the matches from the first pass search. The additional matches from the second pass search serve to increase coverage and may discover interesting modifications or SNPs.

## Error Tolerant Search

### Unsuspected chemical & P-T modifications

- Iterate serially through comprehensive list
- All fixed and variable mods retained
- Allow for one additional “unsuspected” modification per peptide



For modifications, an error tolerant search looks for one unsuspected modification per peptide in addition to those mods specified as fixed or variable. This is sufficient because it will be rare to get two unsuspected mods on a single peptide

## Error Tolerant Search

### Primary sequence variants

- Protein database
  - Look for all residue substitutions
  - No attempt to identify single base insertions & deletions because of frame shifts
- Nucleic acid database
  - Look for all single base substitutions, insertions & deletions



The error tolerant search also looks for sequence variants, such as single nucleotide polymorphisms (SNPs) or sequencing errors.

For a protein database, we can't look for the consequences of inserted or deleted bases, because these give rise to frame shifts, and the entire sequence changes from that point on.

## Error Tolerant Search

### The following constraints apply to the standard, first pass search:

1. Enzyme must be fully specific
2. A reduced ceiling on the number of variable modifications, (default is 2, but this can be changed globally in mascot.dat or for a user group in Mascot security)
3. Cannot be combined with an automatic decoy database search
4. Cannot be combined with quantitation
5. Search cannot include error tolerant sequence tag



There are some constraints on the standard, first pass search

## Error Tolerant Search

The screenshot shows the Mascot MS/MS Ions Search web interface. The 'Error tolerant' checkbox is highlighted with a red circle. The interface includes fields for 'Your name', 'Email', 'Search title', 'Database(s)', 'Enzyme', 'Allow up to', 'Quantitation', 'Taxonomy', 'Fixed modifications', 'Variable modifications', 'Peptide tol.', 'MS/MS tol.', 'Peptide charge', 'Data file', 'Data format', 'Instrument', 'Decoy', 'Report top', and 'Hits'. The 'Error tolerant' checkbox is checked and highlighted with a red circle.

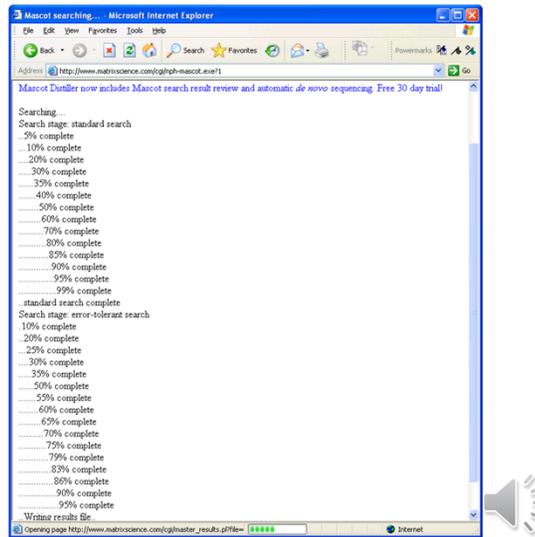
**MASCOT** : Modifications

© 2007-2012 Matrix Science



Otherwise, submitting the search is just like submitting a standard search except that you check the Error Tolerant Checkbox

# Error Tolerant Search



**MASCOT** : *Modifications*

© 2007-2012 Matrix Science



You see two sets of progress reports

Peptide Summary Report (error tolerant example) - Microsoft Internet Explorer

Address: http://www.matrixscience.com/cgi/master\_results.pl?file=.../data20070626/FoGo1e5.dat

Select All Select None Search Selected Error tolerant

1. [AAAS1708](#) Mass: 56371 Score: 782 Queries matched: 27 emPAI: 0.78  
 HUMALPPA NID: - Homo sapiens  
 Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
<input checked="" type="checkbox"/> 27	462.6807	923.3468	923.5116	-0.1649	0	33	16	1	R.FPYVALSK.T
<input checked="" type="checkbox"/> 41	517.1760	1032.3375	1032.5604	-0.2229	0	70	0.0036	3	R.GSSIFGLAPGK.A
<input checked="" type="checkbox"/> 62	564.6804	1127.3463	1127.5764	-0.2301	0	10	2.8e+03	6	R.GFELFVEGGR.I
<input checked="" type="checkbox"/> 65	567.6567	1133.2987	1133.5499	-0.2511	0	44	1.1	1	R.GNEVLSVHNR.A + Oxidation (M)
<input checked="" type="checkbox"/> 86	614.2001	1226.3856	1226.6329	-0.2473	0	28	41	2	K.LGPEIPLAIDR.F + Oxidation (M)
<input checked="" type="checkbox"/> 100	653.2101	1304.4057	1304.6837	-0.2780	0	(87)	5.7e-05	1	K.GHFQIIGLSAAR.F
<input checked="" type="checkbox"/> 124	710.2235	1418.4324	1418.7154	-0.2829	0	95		1	K.GHFQIIGLSAAR.F + Acetyl (N-term); [+72.0211 at N-term G]
<input checked="" type="checkbox"/> 126	726.1806	1450.3465	1450.6477	-0.3011	0	73	0.0012	1	R.HVYSADVPASAR.Q
<input checked="" type="checkbox"/> 133	499.1349	1494.3828	1494.6694	-0.2866	0	92		1	L.DPSLIDFTEAIR.L + 2 Oxidation (M) ←
<input checked="" type="checkbox"/> 145	526.1538	1575.4396	1575.7814	-0.3418	0	(61)		1	R.ALTEIDFDAIER.A + [-48.0000 at F8]
<input checked="" type="checkbox"/> 156	820.7283	1639.4420	1639.7763	-0.3343	0	97	5.1e-06	1	R.ALTEIDFDAIER.A + Oxidation (M)
<input checked="" type="checkbox"/> 165	841.2310	1680.4474	1680.8029	-0.3554	0	(75)		1	R.ALTEIDFDAIER.A + Oxidation (M); [+41.0266 at N-term A]
<input checked="" type="checkbox"/> 170	864.2888	1726.5629	1726.9294	-0.3664	0	44	0.9	1	K.AYTVLLYGHGPPYVLR.D
<input checked="" type="checkbox"/> 176	879.2425	1756.4705	1756.8420	-0.3715	0	83		1	G.IIPVEENHDFVNR.E ←
<input checked="" type="checkbox"/> 204	956.2437	1910.4729	1910.8601	-0.3872	0	29	28	3	R.DSTLPSLIDFTEAIR.L + 2 Oxidation (M)
<input checked="" type="checkbox"/> 208	975.8100	1949.6055	1950.0245	-0.4190	0	85	6.6e-05	1	K.NLIIFLGDQGVSTVAAR.I + Oxidation (M)
<input checked="" type="checkbox"/> 209	976.2340	1950.4534	1950.8555	-0.4021	0	(27)	42	1	K.DGARDVTESESGSPEYR.Q
<input checked="" type="checkbox"/> 211	656.1752	1965.5039	1964.8712	0.6327	0	(72)		1	K.DGARDVTESESGSPEYR.Q + [+14.0157 at T0]
<input checked="" type="checkbox"/> 213	664.5518	1990.6336	1991.0510	-0.4174	0	(58)	4	1	K.NLIIFLGDQGVSTVAAR.I + Oxidation (M); [+41.0266 at N-term N]
<input checked="" type="checkbox"/> 216	1001.2027	2000.3908	2000.8038	-0.4150	0	(65)	0.0069	1	R.HGTFDPEYDYSQGT.R.I + Oxidation (M)
<input checked="" type="checkbox"/> 217	667.8046	2000.3919	2000.8038	-0.4139	0	70	0.002	1	R.HGTFDPEYDYSQGT.R.I + Oxidation (M)
<input checked="" type="checkbox"/> 218	670.1561	2007.4466	2007.8770	-0.4304	0	75		1	K.DGARDVTESESGSPEYR.Q + [+57.0215 at N-term D]
<input checked="" type="checkbox"/> 222	681.0205	2042.4397	2041.8324	0.6073	0	(61)		1	R.HGTFDPEYDYSQGT.R.I + Acetyl (N-term); Oxidation (M); [-0.9840 at E7]
<input checked="" type="checkbox"/> 252	784.5440	2350.6103	2351.1030	-0.4927	0	(69)		1	R.QQSAVPLDEETHAGEDVAVFAR.G + [-17.0265 at N-term Q]
<input checked="" type="checkbox"/> 253	790.2187	2367.6341	2368.1295	-0.4954	0	94	7.6e-06	1	R.QQSAVPLDEETHAGEDVAVFAR.G
<input checked="" type="checkbox"/> 260	809.2208	2424.6406	2425.1510	-0.5104	0	(66)		1	R.QQSAVPLDEETHAGEDVAVFAR.G + [+57.0215 at N-term Q]
<input checked="" type="checkbox"/> 275	920.5078	2758.7415	2759.3502	-0.6167	0	90		1	R.QECCDIATQLISMDIDVILGGR.K + Acetyl (N-term); Oxidation (M); [-0.945 at T1]

MASCOT : Modifications © 2007-2012 Matrix Science MATRIX SCIENCE

And here is the first hit of the results report. The additional matches, found in the error tolerant search, are the ones without expect values. One of these, query 133, is a simple, non-specific peptide with a very good score. There's another example for query 176. The error tolerant search is a much better way of picking up non-specific peptides than searching the entire database with semi-trypsin or no enzyme. We only fail to get such matches in an error tolerant search if there are no matches to the protein in the first pass search. However, you have to ask yourself whether you would believe a protein hit in which the only peptide match was non-specific. I think the answer is no.

## Error Tolerant Search

### To reduce 'junk' matches

- An individual peptide can be semi-specific OR have one unsuspected modification OR have one primary sequence mutation.
- If the mass delta of the modification is less than the smaller of the precursor mass tolerance and the fragment mass tolerance, the modification is rejected. This eliminates modifications that are meaningless given the estimated mass error, like Q->K, in most cases.
- Match must have a score of at least the identity threshold for the same query in the first pass search
- Match must have a score in excess of the highest scoring match to the same query in the first pass search



The matches from an error tolerant search are aggressively filtered to remove junk matches

Peptide Summary Report (error tolerant example) - Microsoft Internet Explorer

Address: http://www.matrixscience.com/cgi/master\_results.pl?file=.../data/20070626/FoGone5.dat

Select All Select None Search Selected Error tolerant

1. AAAS1708 Mass: 56371 Score: 782 Queries matched: 27 emPAI: 0.78  
 HUMALPPA NID: - Homo sapiens  
 Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
27	462.6807	923.3468	923.5116	-0.1649	0	33	16	1	R.FPYVALSK.T
41	517.1760	1032.3375	1032.5604	-0.2229	0	70	0.0036	3	R.GSSIFGLAPGK.A
62	564.6804	1127.3463	1127.5764	-0.2301	0	10	2.8e+03	6	R.GFELFVEGGR.I
65	567.6567	1133.2987	1133.5499	-0.2511	0	44	1.1	1	R.GNEVLSVDR.A + Oxidation (M)
86	614.2001	1226.3856	1226.6329	-0.2473	0	28	41	2	K.LGPEIPLADR.F + Oxidation (M)
100	653.2101	1304.4057	1304.6837	-0.2780	0	(87)	5.7e-05	1	K.GHFQIIGLSAAR.F
124	710.2235	1418.4324	1418.7154	-0.2829	0	95	1	1	K.GHFQIIGLSAAR.F + Acetyl (N-term); [+72.0211 at N-term G]
126	726.1806	1450.3465	1450.6477	-0.3011	0	73	0.0012	1	R.HVYSADVPASAR.Q
133	499.1349	1494.3828	1494.6694	-0.2866	0	92	1	1	L.DPSLIDFTEAALR.L + 2 Oxidation (M)
145	526.1538	1575.4396	1575.7814	-0.3418	0	(61)	1	1	R.ALTEIDFDAIER.A + [-48.0000 at F8]
156	820.7283	1639.4420	1639.7763	-0.3343	0	97	5.1e-06	1	R.ALTEIDFDAIER.A + Oxidation (M)
165	841.2310	1680.4474	1680.8029	-0.3354	0	(75)	1	1	R.ALTEIDFDAIER.A + Oxidation (M); [+41.0266 at N-term A]
170	864.2888	1726.5629	1726.9294	-0.3664	0	44	0.9	1	K.AYTVLLYGHGPPYVLR.D
176	879.2425	1756.4705	1756.8420	-0.3715	0	83	1	1	G.IIPVEENPFVNR.E
204	956.2437	1910.4729	1910.8601	-0.3872	0	29	28	3	R.DSTLPSLIDFTEAALR.L + 2 Oxidation (M)
208	975.8100	1949.6055	1950.0245	-0.4190	0	85	6.6e-05	1	K.DLIIFLGDQGVSTVAAR.I + Oxidation (M)
209	976.2340	1950.4534	1950.8555	-0.4021	0	(27)	42	1	K.DGARDPVTESESGSPEYR.Q
211	656.1752	1965.5039	1964.8712	0.6327	0	(72)	1	1	K.DGARDPVTESESGSPEYR.Q + [+14.0157 at T0]
213	664.5518	1990.6336	1991.0510	-0.4174	0	(58)	4	1	K.DLIIFLGDQGVSTVAAR.I + Oxidation (M); [+41.0266 at N-term N]
216	1001.2027	2000.3908	2000.8038	-0.4150	0	(65)	0.0069	1	R.HGTFDPEYDYSQGT.R.I + Oxidation (M)
217	667.8046	2000.3919	2000.8038	-0.4139	0	70	0.002	1	R.HGTFDPEYDYSQGT.R.I + Oxidation (M)
218	670.1561	2007.4466	2007.8770	-0.4304	0	75	1	1	K.DGARDPVTESESGSPEYR.Q + [+57.0215 at N-term D]
222	681.0205	2042.4397	2041.8324	0.6073	0	(61)	1	1	R.HGTFDPEYDYSQGT.R.I + Acetyl (N-term); Oxidation (M); [-0.9840 at E7]
252	784.5440	2350.6103	2351.1030	-0.4927	0	(69)	1	1	R.QQSAVPLDEETHAGEDVAVFAR.G +
253	790.2187	2367.6341	2368.1295	-0.4954	0	94	7.6e-06	1	R.QQSAVPLDEETHAGEDVAVFAR.G +
260	809.2208	2424.6406	2425.1510	-0.5104	0	(66)	1	1	R.QQSAVPLDEETHAGEDVAVFAR.G +
275	920.5078	2758.7415	2759.3502	-0.6167	0	90	1	1	R.QECCDIATQLISMDIDVILGGR.K

Possible Assignments:  
 Carbamidomethyl (N-term) [+57.0215]  
 Carbamidomethyl (D) [+57.0215]  
 Carboxymethyl (N-term) [+58.0055]

1:AAAS1708 2:AAH09647 3:CAJ15103 4:S12076 5:AAAS1709 8:AAAR8616

MASCOT : Modifications © 2007-2012 Matrix Science MATRIX SCIENCE

Take a look at the match to query 218. The mass tolerance for this search was fairly wide, so the observed mass difference could correspond to either carbamidomethylation or carboxymethylation at the N-terminus. Since this sample was alkylated with iodoacetamide, we would choose carbamidomethylation as the more likely suspect, especially as this brings the error on the precursor mass into line with the general trend, whereas carboxymethylation would give an error of +0.6 Da. The assignment to carbamidomethylation is also very believable, because this is a known artefact of over-alkylation. The same modification is found for query 260.

Peptide Summary Report (error tolerant example) - Microsoft Internet Explorer

Address: http://www.matrixscience.com/cgi/master\_results.pl?file=.../data/20070626/FoGone5.dat

Select All Select None Search Selected Error tolerant

1. AAAS1708 Mass: 56371 Score: 782 Queries matched: 27 emPAI: 0.78  
 HUMALPPA NID: - Homo sapiens  
 Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
<input checked="" type="checkbox"/> 27	462.6807	923.3468	923.5116	-0.1649	0	33	16	1	R.FPYVALSK.T
<input checked="" type="checkbox"/> 41	517.1760	1032.3375	1032.5604	-0.2229	0	70	0.0036	3	R.GSSIFGLAPGK.A
<input checked="" type="checkbox"/> 62	564.6804	1127.3463	1127.5764	-0.2301	0	10	2.8e+03	6	R.GFLFVVEGGR.I
<input checked="" type="checkbox"/> 65	567.6567	1133.2987	1133.5499	-0.2511	0	44	1.1	1	R.GNEVLSVDR.A + Oxidation (M)
<input checked="" type="checkbox"/> 86	614.2001	1226.3856	1226.6329	-0.2473	0	28	41	2	K.LGPEIPLADR.F + Oxidation (M)
<input checked="" type="checkbox"/> 100	653.2101	1304.4057	1304.6837	-0.2780	0	(87)	5.7e-05	1	K.GHFQIIGLSAAR.F
<input checked="" type="checkbox"/> 124	710.2235	1418.4324	1418.7154	-0.2829	0	95	1	1	K.GHFQIIGLSAAR.F + Acetyl (N-term); [+72.0211 at N-term Q]
<input checked="" type="checkbox"/> 126	726.1806	1450.3465	1450.6477	-0.3011	0	73	0.0012	1	R.HVYSADVPASAR.Q
<input checked="" type="checkbox"/> 133	499.1349	1494.3828	1494.6694	-0.2866	0	92	1	1	L.DPSLIDFTEAAR.L + 2 Oxidation (M)
<input checked="" type="checkbox"/> 145	526.1538	1575.4396	1575.7814	-0.3418	0	(61)	1	1	R.ALTEIDFDAIER.A + [-48.0000 at F8]
<input checked="" type="checkbox"/> 156	820.7283	1639.4420	1639.7763	-0.3343	0	97	5.1e-06	1	R.ALTEIDFDAIER.A + Oxidation (M)
<input checked="" type="checkbox"/> 165	841.2310	1680.4474	1680.8029	-0.3554	0	(75)	1	1	R.ALTEIDFDAIER.A + Oxidation (M); [+41.0266 at N-term A]
<input checked="" type="checkbox"/> 170	864.2888	1726.5629	1726.9294	-0.3664	0	44	0.9	1	K.AYTVLLYGHGPGYVLR.D
<input checked="" type="checkbox"/> 176	879.2425	1756.4705	1756.8420	-0.3715	0	83	1	1	G.IIPVEENPFVNR.E
<input checked="" type="checkbox"/> 204	956.2437	1910.4729	1910.8601	-0.3872	0	29	28	3	R.DSTLDPSTLHTEAALR.L + 2 Oxidation (M)
<input checked="" type="checkbox"/> 208	975.8100	1949.6055	1950.0245	-0.4190	0	85	6.6e-05	1	K.MLIIFLGDQGVSTVTAAR.I + Oxidation (M)
<input checked="" type="checkbox"/> 209	976.2340	1950.4534	1950.8555	-0.4021	0	(27)	42	1	K.DGARDPVTESGSPYER.Q
<input checked="" type="checkbox"/> 211	656.1752	1965.5039	1964.8712	0.6327	0	(72)	1	1	K.DGARDPVTESGSPYER.Q + [+14.0157 at T0]
<input checked="" type="checkbox"/> 213	664.5518	1990.6336	1991.0510	-0.4174	0	(58)	4	1	K.MLIIFLGDQGVSTVTAAR.I + Oxidation (M); [+41.0266 at N-term N]
<input checked="" type="checkbox"/> 216	1001.2027	2000.3908	2000.8038	-0.4150	0	(65)	0.0069	1	R.HGTFDPEYDYSQGTQTR.I + Oxidation (M)
<input checked="" type="checkbox"/> 217	667.8046	2000.3919	2000.8038	-0.4139	0	70	0.002	1	R.HGTFDPEYDYSQGTQTR.I + Oxidation (M)
<input checked="" type="checkbox"/> 218	670.1561	2007.4466	2007.8770	-0.4304	0	75	1	1	K.DGARDPVTESGSPYER.Q + [+57.0215 at N-term D]
<input checked="" type="checkbox"/> 222	681.0205	2042.4397	2041.8324	0.6073	0	(61)	1	1	R.HGTFDPEYDYSQGTQTR.I + Acetyl (N-term); Oxidation (M); [-0.9840 at E7]
<input checked="" type="checkbox"/> 252	784.5440	2350.6103	2351.1030	-0.4927	0	(69)	1	1	R.QQSAVPLDETHAGEDVAVTAR.G + [-17.0265 at N-term Q]
<input checked="" type="checkbox"/> 253	790.2187	2367.6341	2368.1295	-0.4954	0	94	7.6e-06	1	R.QQSAVPLDETHAGEDVAVTAR.G
<input checked="" type="checkbox"/> 260	809.2208	2424.6406	2425.1510	-0.5104	0	(66)	1	1	R.QQSAVPLDETHAGEDVAVTAR.G + [+57.0215 at N-term Q]
<input checked="" type="checkbox"/> 275	920.5078	2758.7415	2759.3502	-0.6167	0	90	1	1	R.QECCDIATQLISMDIVILGGR.K + Acetyl (N-term Q); [-17.0265 at N-term Q]

Possible Assignments:  
 Gln->pyro-Glu (N-term Q) [-17.0265]

1:AAAS1708 2:AAH09647 3:CA115103 4:S12076 5:AAAS1709 8:AAAA9816

MASCOT : Modifications © 2007-2012 Matrix Science MATRIX SCIENCE

Another easily believable assignment is pyro-Glu for the match to query 252.

Peptide Summary Report (error tolerant example) - Microsoft Internet Explorer

Address: http://www.matrixscience.com/cgi/master\_resuls.pl?file=.../data20070626/FoGone5.dat

Select All | Select None | Search Selected | Error tolerant

1. AAAS1708 Mass: 56371 Score: 782 Queries matched: 27 emPAI: 0.78  
 HUMALPPA NID: - Homo sapiens  
 Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
<input checked="" type="checkbox"/> 27	462.6807	923.3468	923.5116	-0.1649	0	33	16	1	R.FFYVALSK.Y
<input checked="" type="checkbox"/> 41	517.1760	1032.3375	1032.5604	-0.2229	0	70	0.0036	3	R.GSIFGLAPGK.A
<input checked="" type="checkbox"/> 62	564.6804	1127.3463	1127.5764	-0.2301	0	10	2.8e+03	6	R.GFLFVGGGR.I
<input checked="" type="checkbox"/> 65	567.6567	1133.2987	1133.5499	-0.2511	0	44	1.1	1	R.GHEVLSVDR.A + Oxidation (M)
<input checked="" type="checkbox"/> 86	614.2001	1226.3856	1226.6329	-0.2473	0	28	41	2	K.LGPEIPLADR.F + Oxidation (M)
<input checked="" type="checkbox"/> 100	653.2101	1304.4057	1304.6837	-0.2780	0	(87)	5.7e-05	1	K.GHFYIQLSAAAR.F
<input checked="" type="checkbox"/> 124	710.2235	1418.4324	1418.7154	-0.2829	0	95		1	K.GHFYIQLSAAAR.F + Acetyl (N-term); [+22.0211 at N-term G]
<input checked="" type="checkbox"/> 126	726.1806	1450.3465	1450.6477	-0.3011	0	73	0.0012	1	R.WFYSDADVPASAR.Q
<input checked="" type="checkbox"/> 133	499.1349	1494.3828	1494.6694	-0.2866	0	92		1	L.DPSLIDHTEAALR.L + 2 Oxidation (M)
<input checked="" type="checkbox"/> 145	526.1538	1575.4396	1575.7814	-0.3418	0	(61)		1	R.ALLETIDHDAIER.A + [-48.0000 at F8]
<input checked="" type="checkbox"/> 156	820.7283	1639.4420	1639.7763	-0.3343	0	97	5.1e-06	1	R.ALLETIDHDAIER.A + Oxidation (M)
<input checked="" type="checkbox"/> 165	841.2310	1680.4474	1680.8029	-0.3554	0	(75)		1	R.ALLETIDHDAIER.A + Oxidation (M); [+41.0266 at N-term A]
<input checked="" type="checkbox"/> 170	864.2888	1726.5629	1726.9294	-0.3664	0	44	0.9	1	K.AYTVLLHGDPYVLR.D
<input checked="" type="checkbox"/> 176	879.2425	1756.4705	1756.8420	-0.3715	0	83		1	G.IIPVEEHPFVRR.E
<input checked="" type="checkbox"/> 204	956.2437	1910.4729	1910.8601	-0.3872	0	29	28	3	R.DSTLDPSEHTEAALR.L + 2 Oxidation (M)
<input checked="" type="checkbox"/> 208	975.8100	1949.6055	1950.0245	-0.4190	0	85	6.6e-05	1	K.NLIIFLGDHGVSTVTAAR.I + Oxidation (M)
<input checked="" type="checkbox"/> 209	976.2340	1950.4534	1950.8555	-0.4021	0	(27)	42	1	K.DGARPDVTESESGSPEYR.Q
<input checked="" type="checkbox"/> 211	656.1792	1965.5039	1964.8712	0.6327	0	(72)		1	K.DGARPDVTESESGSPEYR.Q + [+14.0137 at T8]
<input checked="" type="checkbox"/> 213	664.5518	1990.6336	1991.0510	-0.4174	0	(58)	4	4	K.NLIIFLGDHGVSTVTAAR.I + Oxidation (M); [+41.0266 at N-term N]
<input checked="" type="checkbox"/> 216	1001.2027	2000.3908	2000.8058	-0.4150	0	(65)	0.0069	1	R.HGTPDPEYDDYSQGGTR.L + Oxidation (M)
<input checked="" type="checkbox"/> 217	667.8046	2000.3919	2000.8058	-0.4139	0	70	0.002	1	R.HGTPDPEYDDYSQGGTR.L + Oxidation (M)
<input checked="" type="checkbox"/> 218	670.1561	2007.4466	2007.8770	-0.4304	0	75		1	K.DGARPDVTESESGSPEYR.Q + [+57.0215 at T8]
<input checked="" type="checkbox"/> 222	681.8205	2042.4397	2041.8324	0.6073	0	(61)		1	R.HGTPDPEYDDYSQGGTR.L + Acetyl (N-term)
<input checked="" type="checkbox"/> 232	784.5440	2350.6103	2351.1030	-0.4927	0	(69)		1	R.QQSAPVLDDEETHAGEDVAVFAR.G + [-11.0156 at T8]
<input checked="" type="checkbox"/> 233	790.2187	2367.6341	2368.1295	-0.4954	0	94	7.6e-06	1	R.QQSAPVLDDEETHAGEDVAVFAR.G
<input checked="" type="checkbox"/> 240	809.2208	2424.6406	2425.1510	-0.5104	0	(66)		1	R.QQSAPVLDDEETHAGEDVAVFAR.G + [+57.0215 at N-term Q]
<input checked="" type="checkbox"/> 275	920.5878	2758.7415	2759.3582	-0.6167	0	90		1	R.QEGQDIATQLISMIDIVILGGGR.K + Acetyl (N-term); Oxidation (M); [-0.9476 at E7]

1:AAAS1708 2:AAH09647 3:CAJ15103 4:S12076 5:AAAS1709 8:AAA98616

MASCOT : Modifications © 2007-2012 Matrix Science MATRIX SCIENCE

As is methylation ay T8 for query 211

Peptide Summary Report (error tolerant example) - Microsoft Internet Explorer

Address: http://www.matrixscience.com/cgi/master\_results.pl?file=...data20070626/FoGone5.dat

Select All Select None Search Selected Error tolerant

1. AAAS1708 Mass: 56371 Score: 782 Queries matched: 27 emPAI: 0.78  
 HUMALPPA NID: - Homo sapiens  
 Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
27	462.6807	923.3468	923.5116	-0.1649	0	33	16	1	R.FPYVALSK.T
41	517.1760	1032.3375	1032.5604	-0.2229	0	70	0.0036	3	R.GSSIFGLAPGK.A
62	564.6804	1127.3463	1127.5764	-0.2301	0	10	2.8e+03	6	R.GFELFVEGGR.I
65	567.6567	1133.2987	1133.5499	-0.2511	0	44	1.1	1	R.GNEVLSVDR.A + Oxidation (M)
86	614.2001	1226.3856	1226.6329	-0.2473	0	28	41	2	K.LGPEIPLADR.F + Oxidation (M)
100	653.2101	1304.4057	1304.6837	-0.2780	0	(87)	5.7e-05	1	K.GHFQIIGLSAAR.F
124	710.2235	1418.4324	1418.7154	-0.2829	0	95		1	K.GHFQIIGLSAAR.F + Acetyl (N-term); [+72.0211 at N-term G]
126	726.1806	1450.3465	1450.6477	-0.3011	0	73	0.0012	1	R.HVYSADVPASAR.Q
133	499.1349	1494.3828	1494.6694	-0.2866	0	92		1	L.DPSLIDFTEALR.L + 2 Oxidation (M)
145	526.1538	1575.4396	1575.7814	-0.3418	0	(61)		1	R.ALTEIDFDAIER.A + [-48.0000 at F8]
156	820.7283	1639.4420	1639.7763	-0.3343	0	97	5.1e-06	1	R.ALTEIDFDAIER.A + Oxidation (M)
165	841.2310	1680.4474	1680.8029	-0.3554	0	(75)		1	R.ALTEIDFDAIER.A + Oxidation (M)
170	864.2888	1726.5629	1726.9294	-0.3664	0	44	0.9	1	K.AYTVLLVGHGPGYVLR.D
176	879.2425	1756.4705	1756.8420	-0.3715	0	83		1	G.IIPVEENPFVNR.E
204	956.2437	1910.4729	1910.8601	-0.3872	0	29	28	3	R.DSTLDFSLIDFTEALR.L + 2 Oxidation (M)
208	975.8100	1949.6055	1950.0245	-0.4190	0	85	6.6e-05	1	K.NLIIFLGDQGVSTVTAAR.I + Oxidation (M)
209	976.2340	1950.4534	1950.8555	-0.4021	0	(27)	42	1	K.DGARDVTESESGSPEYR.Q
211	656.1752	1965.5039	1964.8712	0.6327	0	(72)		1	K.DGARDVTESESGSPEYR.Q + [+14.0157 at T0]
213	664.5518	1990.6336	1991.0510	-0.4174	0	(58)		4	K.NLIIFLGDQGVSTVTAAR.I + Oxidation (M); [+41.0266 at N-term N]
216	1001.2027	2000.3908	2000.8038	-0.4150	0	(65)	0.0069	1	R.HGTFDPEYDYSQGT.R.I + Oxidation (M)
217	667.8046	2000.3919	2000.8038	-0.4139	0	70	0.002	1	R.HGTFDPEYDYSQGT.R.I + Oxidation (M)
218	670.1561	2007.4466	2007.8770	-0.4304	0	75		1	K.DGARDVTESESGSPEYR.Q + [+57.0215 at N-term D]
222	681.0205	2042.4397	2041.8324	0.6073	0	(61)		1	R.HGTFDPEYDYSQGT.R.I + Acetyl (N-term); Oxidation (M); [-0.9840 at E7]
252	784.5440	2350.6103	2351.1030	-0.4927	0	(69)		1	R.QQSAVPLDETHAGEDVAVFAR.G + [-17.0265 at N-term Q]
253	790.2187	2367.6341	2368.1295	-0.4954	0	94	7.6e-06	1	R.QQSAVPLDETHAGEDVAVFAR.G
260	809.2208	2424.6406	2425.1510	-0.5104	0	(66)		1	R.QQSAVPLDETHAGEDVAVFAR.G + [+57.0215 at N-term Q]
275	920.5078	2758.7415	2759.3502	-0.6167	0	90		1	R.QECCDIATQLISNIDIVILGGR.K + Acetyl (N-term); Oxidation (M); [-0.9271 at ...]

1:AAAS1708 2:AAH09647 3:CA115103 4:S12076 5:AAAS1709 8:AAAR8616

MASCOT : Modifications © 2007-2012 Matrix Science MATRIX SCIENCE

In other cases, the match may be good, but the assignment is not believable. Query 145 is listed with a substitution at F8 causing a loss of 48 Da. This seems unlikely because we have 2 other matches to the same peptide without any substitution. What else could it be? Well, notice that the other two matches are both oxidised at M7. If we suppose this peptide is also oxidised, then the mass shift becomes -64, which is a well-known loss for oxidised methionine, (loss of methanesulfenic acid). This would seem a much more likely explanation for this match.

It is important to understand that the error tolerant search finds new matches by introducing mass shifts at different positions in the database sequences. The match may be very strong, but figuring out a credible assignment can require a bit of detective work.

Peptide Summary Report (error tolerant example) - Microsoft Internet Explorer

Address: http://www.matrixscience.com/cgi/master\_resuls.pl?file=...data20070626/FoG1e5.dat

11. **INTP** Mass: 23978 Score: 454 Queries matched: 16 emPAI: 1.42  
 trypsin (EC 3.4.21.4) (isopropylphosphorylated) - bovine  
 Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
71	577.1685	1152.3225	1152.5663	-0.2438	0	87	4.6e-05	1	K.SSGTSPDVLK.C
74	584.6704	1167.3263	1167.5747	-0.2484	0	90	2.2e-05	1	K.VCHYVSWIK.Q
78	598.1756	1194.3366	1194.5768	-0.2402	0	(69)		1	K.SSGTSPDVLK.C + [+42.0106 at N-term S]
83	606.1852	1210.3559	1210.5717	-0.2158	0	(61)		1	K.SSGTSPDVLK.C + [+58.0055 at N-term S]
94	640.1278	1278.2411	1278.4629	-0.2219	0	(67)		1	K.SSGTSPDVLK.C + [+125.8966 at Y6]
132	745.7224	1489.4302	1489.7348	-0.3046	0	72	0.0017	1	K.LQGVSWGSCAQK.N
229	1081.7685	2161.5224	2162.0491	-0.5267	0	156	5.1e-12	1	R.LGEDHINVEGHEQFISASK.S
230	721.5398	2161.5976	2162.0491	-0.4515	0	(94)	8.2e-06	1	R.LGEDHINVEGHEQFISASK.S
231	721.8998	2162.6775	2162.0491	0.6284	0	(42)	1.5	1	R.LGEDHINVEGHEQFISASK.S
233	729.5354	2185.5845	2186.0240	-0.4395	0	(109)		1	R.LGEDHINVEGHEQFISASK.S + [+23.9748 at N-term L]
234	1094.8114	2187.6082	2188.0284	-0.4201	0	(97)		1	R.LGEDHINVEGHEQFISASK.S + Acetyl (N-term); [-16.0313 at N-term L]
236	1102.8029	2203.5912	2204.0961	-0.5048	0	(102)		1	R.LGEDHINVEGHEQFISASK.S + [+42.0470 at G2]
237	735.5400	2203.5983	2204.0597	-0.4614	0	(67)	0.0043	1	R.LGEDHINVEGHEQFISASK.S + Acetyl (N-term)

Top scoring peptide matches to query 236  
 142: Sum of 0 scans in range 2405 (rt=2102.40) to 2426 (rt=2116.32)

Score	Expect	Delta	Hit	Protein	Peptide
102.4	1.7e-06	-0.5048	11+	INTP	R.LGEDHINVEGHEQFISASK.S
100.8	4e+02	-0.4684	11+	INTP	R.LGEDHINVEGHEQFISASK.S
17.0	7.4e+02	-0.5677			R.VGDPFNPKVTVGPNPQGVK.Y
14.3	1e+03	-0.6187			K.LARGHIVADVLEPOLTHK.I
13.0	1e+03	0.5940			K.GGARVGVIVVCHGEQHEEDK.S
12.9	1e+03	-0.4395			K.VLSCDTPVQSSNLTIFSSK.E
12.8	1.1e+03	0.3580			R.DPNSPKVSAVIVNRGLPK.A
12.3	1.2e+03	-0.5412			R.FTPGNASSAVPSSKTVIAK.D
11.9	1.3e+03	0.4783			H.LVULTRHEDLFRABVHK.T
11.8	1.4e+03	0.2933			R.YPQLPIVGLVFLKPAISASK.T
11.5	1.4e+03	-0.4837			R.QAEVKPEDIDYIXANGSOTK.Q

12. **TR**  
 17.0 4e+02 -0.5677 R.VGDPFNPKVTVGPNPQGVK.Y  
 14.3 7.4e+02 -0.6187 K.LARGHIVADVLEPOLTHK.I  
 13.0 1e+03 0.5940 K.GGARVGVIVVCHGEQHEEDK.S  
 12.9 1e+03 -0.4395 K.VLSCDTPVQSSNLTIFSSK.E  
 12.8 1.1e+03 0.3580 R.DPNSPKVSAVIVNRGLPK.A  
 12.3 1.2e+03 -0.5412 R.FTPGNASSAVPSSKTVIAK.D  
 11.9 1.3e+03 0.4783 H.LVULTRHEDLFRABVHK.T  
 11.8 1.4e+03 0.2933 R.YPQLPIVGLVFLKPAISASK.T  
 11.5 1.4e+03 -0.4837 R.QAEVKPEDIDYIXANGSOTK.Q

Peptide  
 K.SIVHPSYNSH.T  
 K.SSGTSPDVLK.C  
 K.VCHYVSWIK.Q  
 K.SSGTSPDVLK.C + [+42.0106 at N-term S]  
 K.SSGTSPDVLK.C + [+58.0055 at N-term S]  
 K.SSGTSPDVLK.C + [+125.8966 at Y6]  
 R.LGEDHINVEGHEQFISASK.S

11:INTP 12:TRBOTR

MASCOT : Modifications © 2007-2012 Matrix Science MATRIX SCIENCE

You should also look at the other yellow pop-up when trying to decide whether to accept a match or not. In this example, the error tolerant search was able to get a slightly higher score by shifting a modification of +42 Da from the amino terminus to the adjacent glycine. However, as score increase of 2 in 100 is negligible. Much more believable to take the original match from the first pass search, which can be explained as N-terminal acetylation.

## Error Tolerant Search

- Can successfully locate mass differences corresponding to a single unsuspected modification or a single SNP per peptide
- User must decide on best explanation for the observed differences
- Limited to proteins which have at least one good peptide match ... not very useful for (say) MHC peptides.



In summary, an error tolerant search

- Can successfully locate mass differences corresponding to a single unsuspected modification or a single SNP per peptide
- User must decide on best explanation for the observed differences
- Limited to proteins which have at least one good peptide match ... not very useful for (say) MHC peptides