

Modifications

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

File Edit View History Bookmarks Tools Help

Unimod

www.unimod.org/modifications_list.php?goto=28

UNIMOD protein modifications for mass spectrometry

Unimod Logged as richardj Log out Change password Advanced search

Add new Search for: Any field Contains Show all Search Details found: 1507 Page 28 of 76 Records Per Page: 20

Select/Unselect all Delete selected

	Accession #	PSI-MS Name	Interim name	Description	Monoisotopic mass	Average mass	Composition
Copy View	200	Ethanedithiol	EDT	EDT	75.980527	76.1838	H(4) C(2) O(-1) S(2)
Copy View	303	DeStreak	DeStreak	Cysteine mercaptoethanol	75.998285	76.1176	H(4) C(2) O S
Copy View	208	Delta:H(4)C(6)	Acrolein76	Acrolein addition +76	76.031300	76.0960	H(4) C(6)
Copy View	653		Ser->Tyr	Ser->Tyr substitution	76.031300	76.0960	H(4) C(6)
Copy View	1045		Ala->Phe	Ala->Phe substitution	76.031300	76.0960	H(4) C(6)
Copy View	340	Bromo	bromo	bromination	77.910511	78.8961	H(-1) Br
Copy View	728		Methylphosphonate	Methylphosphorylation	77.987066	78.0071	H(3) C O(2) P
Copy View	316	DimethylpyrroleAdduct	pyrrole	2,5-dimethylpyrrole	78.046950	78.1118	H(6) C(6)
Copy View	423	Delta:Se(1)	selenyl	selenyl	79.916520	78.9600	Se
Copy View	40	Sulfo	Sulfation	O-Sulfonation	79.956815	80.0632	O(3) S
Copy View	21	Phospho	Phospho	Phosphorylation	79.966331	79.9799	H O(3) P
Copy View	1927		Delta:H(4)C(S)O(1)	methylglyoxal-derived argpyrimidine	80.026215	80.0847	H(4) C(5) O
Copy View	1104		Gly->His	Gly->His substitution	80.037448	80.0880	H(4) C(4) N(2)
Copy View	837		Arg->Npo	Arginine replacement by Nitroprymidyl ornithine	80.985078	81.0297	H(-1) C(3) N O(2)
Copy View	2000		Xlink:SDA	NHS-Diazirine crosslinker	82.041865	82.1005	H(6) C(5) O
Copy View	549		Cys->Trp	Cys->Trp substitution	83.070128	83.0670	H(5) C(8) N S(-1)
Copy View	1211		Thr->Trp	Thr->Trp substitution	85.031634	85.1060	H(3) C(7) N O(-1)
Copy View	211	NEIAA	NEIAA-d0	N-ethyl iodoacetamide-d0	85.052764	85.1045	H(7) C(4) N O
Copy View	1886		Xlink:BuIrBu[85]	Bu fragment of BuIrBu crosslinker	85.052764	85.1045	H(7) C(4) N O
Copy View	1052		Ala->Arg	Ala->Arg substitution	85.063997	85.1078	H(7) C(3) N(3)

First : Previous [21 22 23 24 25 26 27 28 29 30] Next : Last

MASCOT : Modifications

© 2007-2022 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.

The screenshot displays the Unimod website interface in a web browser. The URL is www.unimod.org/modifications_view.php?editid1=56. The page title is "UNIMOD protein modifications for mass spectrometry". Below the title, it says "Unimod, View record [Accession #: 56]".

Key information from the record:

- 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

Below this, there are seven "Specificity Definition" sections, each with a "Site" and a "Position":

- Specificity Definition 1:** Site: K, Position: Anywhere, Classification: Isotopic label
- Specificity Definition 2:** Site: N-term, Position: Any N-term, Classification: Isotopic label
- Specificity Definition 3:** Site: H, Position: Anywhere, Classification: Isotopic label
- Specificity Definition 4:** Site: S, Position: Anywhere, Classification: Isotopic label
- Specificity Definition 5:** Site: T, Position: Anywhere, Classification: Isotopic label
- Specificity Definition 6:** Site: Y, Position: Anywhere, Classification: Isotopic label
- Specificity Definition 7:** Site: N-term, Position: Protein N-term, Classification: Isotopic label

At the bottom, there is a "Notes and References" section with a table of references:

Source	PubMed PMID	Reference
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, Roujian; Sioma, Cathy S.; Thompson, Robert A.; Xiong, Li; Ragnies, Fred E.. Department of Chemistry, Purdue University, West Lafayette, IN, USA. Analytical Chemistry 72
Source	Journal	Reference Global internal standard technology for comparative proteomics. Chakraborty, Ashish; Ragnies, 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. Ragnies, Fred E.; Riggs, Larry; Zhang, Roujian; Xiong, Li; Liu, Peiran; Chakraborty, Ashish; Siesley, Erin; Sioma, Cathy; Thompson, Robert A. Department of Chemistry, IN
Curator	penner	Last Modified 2011-11-21 10:07:03
		Verified Yes

At the bottom of the page, there is a footer with the text "MASCOT : Modifications", "© 2007-2022 Matrix Science", and the "MATRIX SCIENCE" logo.

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.

Mascot Configuration: Modifications

Displaying 1358/1358

Visibility:

☐ Short list

☐ Long list

☐ Mixed

☐ Not listed

Error tolerant:

☐ Yes

☐ No

☐ Mixed

Classifications: [clear](#)

-

Post-translational

Co-translational

Pre-translational

Chemical derivative

Artefact

Source:

☐ Unimod

☐ Edited Unimod

☐ Local

Apply to selected: 0

[Include in short list](#)

[Include in long list](#)

[Include in error tolerant](#)

[Exclude from error tolerant](#)

[Delete](#)

Modifications

<input type="checkbox"/> Title	Monoisotopic	Average	Composition	Source	Visibility	Err Tol
<input type="checkbox"/> 15N-oxobutanoic	-18.023584	-18.0239	H(-3) 15N(-1)	Unimod	long	yes Copy Print
<input type="checkbox"/> 2-dimethylsuccinyl	144.042259	144.1253	H(8) C(6) O(4)	Unimod	long	yes Copy Print
<input type="checkbox"/> 2-monomethylsuccinyl	130.026609	130.0987	H(6) C(5) O(4)	Unimod	long	yes Copy Print
<input type="checkbox"/> 2-nitrobenzyl	135.032028	135.1201	H(5) C(7) N O(2)	Unimod	long	yes Copy Print
<input type="checkbox"/> 2-succinyl	116.010959	116.0722	H(4) C(4) O(4)	Unimod	long	yes Copy Print
<input type="checkbox"/> 2HPG	282.052824	282.2476	H(10) C(16) O(5)	Unimod	long	yes Copy Print
<input type="checkbox"/> 3-deoxyglucosone	144.042259	144.1253	H(8) C(6) O(4)	Unimod	long	yes Copy Print
<input type="checkbox"/> 3-phosphoglyceryl	167.982375	168.0420	H(5) C(3) O(6) P	Unimod	long	yes Copy Print
<input type="checkbox"/> 3sulfo	183.983029	184.1693	H(4) C(7) O(4) S	Unimod	long	yes Copy Print
<input type="checkbox"/> 4-ONE	154.099380	154.2063	H(14) C(9) O(2)	Unimod	long	yes Copy Print
<input type="checkbox"/> 4-ONE+Delta:H(-2)O(-1)	136.088815	136.1910	H(12) C(9) O	Unimod	long	yes Copy Print
<input type="checkbox"/> 4AcAllylGal	372.142033	372.3671	H(24) C(17) O(9)	Unimod	long	yes Copy Print
<input type="checkbox"/> a-type-ion	-46.005479	-46.0254	H(-2) C(-1) O(-2)	Unimod	long	yes Copy Print
<input type="checkbox"/> AccQTag	170.048013	170.1674	H(6) C(10) N(2) O	Unimod	long	yes Copy Print
<input type="checkbox"/> Acetyl	42.010565	42.0367	H(2) C(2) O	Unimod	mixed	yes Copy Print
<input type="checkbox"/> Acetyl:13C(2)	44.017274	44.0220	H(2) 13C(2) O	Unimod	long	yes Copy Print
<input type="checkbox"/> Acetyl:2H(3)	45.029395	45.0552	H(-1) 2H(3) C(2) O	Unimod	long	yes Copy Print
<input type="checkbox"/> Acetyldeoxyhypusine	97.089149	97.1582	H(11) C(6) N	Unimod	long	yes Copy Print
<input type="checkbox"/> Acetylhypusine	113.084064	113.1576	H(11) C(6) N O	Unimod	long	yes Copy Print
<input type="checkbox"/> ADP-Ribosyl	541.061110	541.3005	H(13) C(10) N(5) O(9) P(2)	Pent Unimod	long	yes Copy Print

Page 1/68 Go to page [<<](#) [>>](#) Page size

[Add new modification](#) [Main menu](#) [Check Unimod](#)

MASCOT

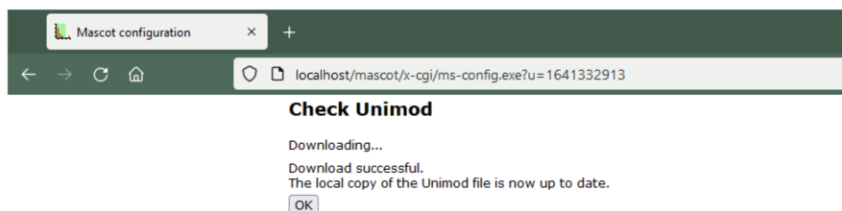
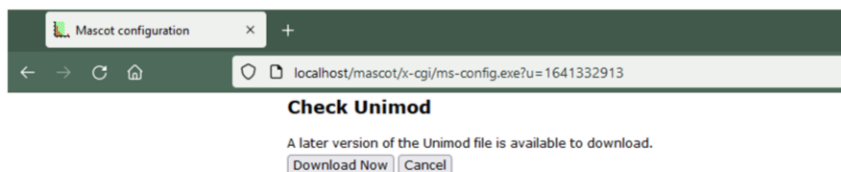
: Modifications

© 2007-2022 Matrix Science

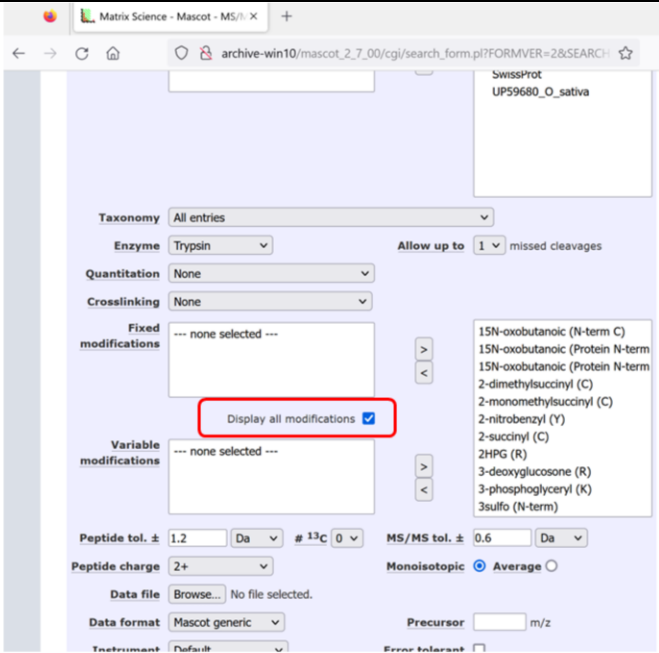


MATRIX
SCIENCE

If you go to Mascot Server Modification editor, there is a link to check to see if there is an updated unimod file.



If there is a newer version available, click on the “Download Now” button and “OK” once the download is complete. This is the easiest way to keep the modifications list on an in-house Mascot server up-to-date. Note that updating the Unimod modifications does not affect or change your local modifications.



Matrix Science - Mascot - MS/MS X

archive-win10/mascot_2_7_00/cgi/search_form.pl?FORMVER=2&SEARCH=

SwissProt
UP59680_O_sativa

Taxonomy: All entries

Enzyme: Trypsin

Allow up to: 1 missed cleavages

Quantitation: None

Crosslinking: None

Fixed modifications: --- none selected ---

Variable modifications: --- none selected ---

Display all modifications ☒

15N-oxobutanoic (N-term C)
15N-oxobutanoic (Protein N-term)
15N-oxobutanoic (Protein N-term)
2-dimethylsuccinyl (C)
2-monomethylsuccinyl (C)
2-nitrobenzyl (Y)
2-succinyl (C)
2HPG (R)
3-deoxyglucosone (R)
3-phosphoglyceryl (K)
3sulfo (N-term)

Peptide tol. \pm 1.2 Da # ^{13}C 0 MS/MS tol. \pm 0.6 Da

Peptide charge: 2+

Data file: Browse... No file selected.

Data format: Mascot generic

Instrument: Default

Precursor: m/z

Monoisotopic ☒ Average ☐

Error tolerant ☐

MASCOT : Modifications

© 2007-2022 Matrix Science

MATRIX SCIENCE

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, check the Display all modifications box.

Variable Modification Permutation

KKKSTKKSTKSKSK

Acetyl (K), Phospho (ST)

- 1 x Acetyl (K) : 8 arrangements
- 1 x Phospho (ST) : 6 arrangements
- 2 x Acetyl (K) : 28 arrangements
- 2 x Acetyl (K) + 1 x Phospho (ST) : 168 arrangements
- ... and so on ...

MASCOT

: *Modifications*

© 2007-2022 Matrix Science

 **MATRIX
SCIENCE**

Now let's define what it is the software is having to do when looking for modified sites.

Let's consider this slightly unusual peptide comprised of 8 Lysine, 4 serine and 2 threonine residues as an example. If we carried out a search with Lysine acetylation and Phospho Serine & Threonine:

If the precursor mass and tolerance allowed for a single Lysine acetylation, there are just 8 possible arrangements of this.

Likewise, if a single phosphorylation was possible, there are just 6 possible arrangements. However, the number of possible arrangements increases rapidly as we consider more modifications and modifiable sites.

If we need two lysine residues to be acetylated to match the precursor, we now have 28 possible arrangements.

Add in a single phosphorylation and we now have the possible 28 acetylation sites combined with 6 possible phosphorylation sites for a grand total of 168 possible arrangements.

This is the so call combinatorial explosion, and it's one reason why some searches with a large number of frequent modifications can take a long time

Modification permutation in Mascot 2.6 or earlier

No upper limit on no. modified sites

Permutation has built in limits

Number of arrangements < 16

- All are tried

Greater than 16 'sliding window' used

- Testing all possibilities would be too slow

Here's how variable modification permutation works in Mascot 2.6 or earlier.

There is no upper limit on the number of modified sites per peptide. However, permutation of the modifications options does have built in limits.

If the maximum number of arrangements for a peptide is less than 16, then all possible permutations are tested for matching by Mascot.

However, if there are more than 16 arrangements, then a second approach is automatically used, where a sliding window is applied to the peptide. This is to prevent the search from getting too slow and taking too long.

Problems with 2.6 approach

Less than 16 possible permutations:

- No issues, all possibilities tested

More than 16 possible permutations:

- Tends to cluster modifications on adjacent modifiable sites
- Often stops before 16 different permutations tested

In general, the approach works well. However, it isn't without its limitations.

If a peptide has less than 16 possible variable modification permutations, then there are no issues as all possibilities are tested. It's in the cases where the peptide has more than 16 possible permutations that issues can arise.

The sliding window method tends to cluster modifications on adjacent modifiable sites, and it will often stop before 16 different permutations have been tested.

Modification iterator in Mascot 2.7

Single, consistent, permutation method

- No switching between methods

Controlled by 3 user definable parameters:

- **MaxPepNumVarMods**
Max no. of different variable modifications per peptide
- **MaxPepNumModifiedSites**
Max no. of modified residues per peptide
- **MaxPepModArrangements**
Max no. of arrangements of an individual varmod composition

In Mascot 2.7 we've taken a different approach. We use a single, consistent permutation method – there's no switching between different methods. The new permutation iterator samples arrangements using a uniformly random scheme. The operation of this is controlled by 3 user definable settings.

MaxPepNumVarMods – this specifies the maximum number of different variable modifications which can be applied to a peptide

MaxPepNumModifiedSites – this specifies the maximum number of residues which can be modified on a peptide

MaxPepModArrangements – "this specifies the maximum number of arrangements of an individual variable modification composition to test"

Modification iteration in Mascot 2.7

Defaults chosen to give similar speed and depth of search to Mascot 2.6 or earlier

- MaxPepNumVarMods 3
- MaxPepNumModifiedSites 5
- MaxPepModArrangements 64

Two main cases for changing the defaults

- Decrease limits to reduce search time
- Increase limits to improve site analysis, or if you're looking at a highly modified protein

Mascot 2.7 ships with the following default values for these parameters:

MaxPepNumVarMods 3, MaxPepNumModifiedSites 5 and MaxPepModArrangements 64 – these have been chosen to give a similar speed and depth of search to Mascot 2.6 or earlier. So for most searches with variable modifications, you won't see major differences in the results if you repeat an old search on Mascot 2.7. Assuming you're using the same database release of course.

There are two main cases where you might want to change these defaults. Decreasing any of these values will reduce the search space, as fewer arrangements will be tested. This will decrease the search time – so if you're looking at a sample which is not highly modified and where definitive site analysis is not the aim of the study, you may wish to decrease some of these values.

However, if site analysis is important, or if you're looking at a highly modified protein such as Histone, then you may need to increase these limits in order to gain accurate modification localisation results.

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, 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.

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

▼Sensitivity and FDR (reversed protein sequences)

	SwissProt	Decoy	FDR	
Protein family members	28	1	3.57%	
PSMs	above	homology	84	1 1.19% Adjust to 1%

Decoy results are available in [the decoy report](#).

Oxidation (M)

4 sec

▼Sensitivity and FDR (reversed protein sequences)

	SwissProt	Decoy	FDR	
Protein family members	26	1	3.85%	
PSMs	above	homology	71	1 1.41% Adjust to 1%

Decoy results are available in [the decoy report](#).

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

253 sec

MASCOT : Modifications

© 2007-2022 Matrix Science

**MATRIX
SCIENCE**

To illustrate this point. This search of the error tolerant example data from Mascot help, using one variable mod, results in 84 statically significant matches.

If the search is repeated with 7 variable mods, the individual matches have identical scores, but the significance threshold is higher and there are fewer matches overall.

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

What's worse, the search takes over 50 times as long!

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

Why is phosphorylation such a challenge?

Site heterogeneity

Poor ionisation efficiency

3 fragmentation channels

- intact fragments
- neutral loss of HPO_3 (80 Da)
- neutral loss of H_3PO_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?

The screenshot displays the Matrix Science Mascot web interface. The browser address bar shows the URL: https://www.matrixscience.com/cgi/master_results_2.pl?file=.%2Fdata%2F2. The interface includes a search filter section with options for significance threshold (p < 0.05), target FDR, display of non-sig. matches, preferred taxonomy (All entries), max. number of families (AUTO), FDR type (PSM), min. number of sig. unique sequences (1), and dendrograms cut at (0). Below this, the 'Sensitivity and FDR (reversed protein sequences)' section shows 'Proteins (1)' and a 'Report Builder' link. The main results section, 'Protein family 1 (out of 1)', shows a search for 'CASB_BOVIN' with a score of 88, mass of 25091, and 1 match. The results are displayed in a table with columns: Query Dupes, Observed, Mr (expt), Mr (calc), Delta M, Score, Expect, Rank, U, and Peptide. The top result is 'K.PQSEKQQQTDEELQK.I + Phospho (ST)' with a score of 88 and an expectation of 6.1e-07. The table also shows other peptides with lower scores and expectations.

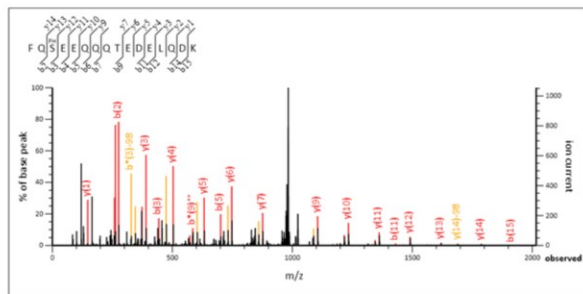
MASCOT : Modifications

© 2007-2022 Matrix Science

MATRIX SCIENCE

Let's 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.



Label all possible matches ☐ Label matches used for scoring *

Monoisotopic mass of neutral peptide Mr(calc): 2060.8212
 Variable modifications:
 83 + Phospho (87), with neutral losses 0.0000 (shown in table), 97.9769
 Ions Score: 88 Expect: 1.7e-06
 Matches: 39/262 Fragment Ions using 56 most intense peaks (help)

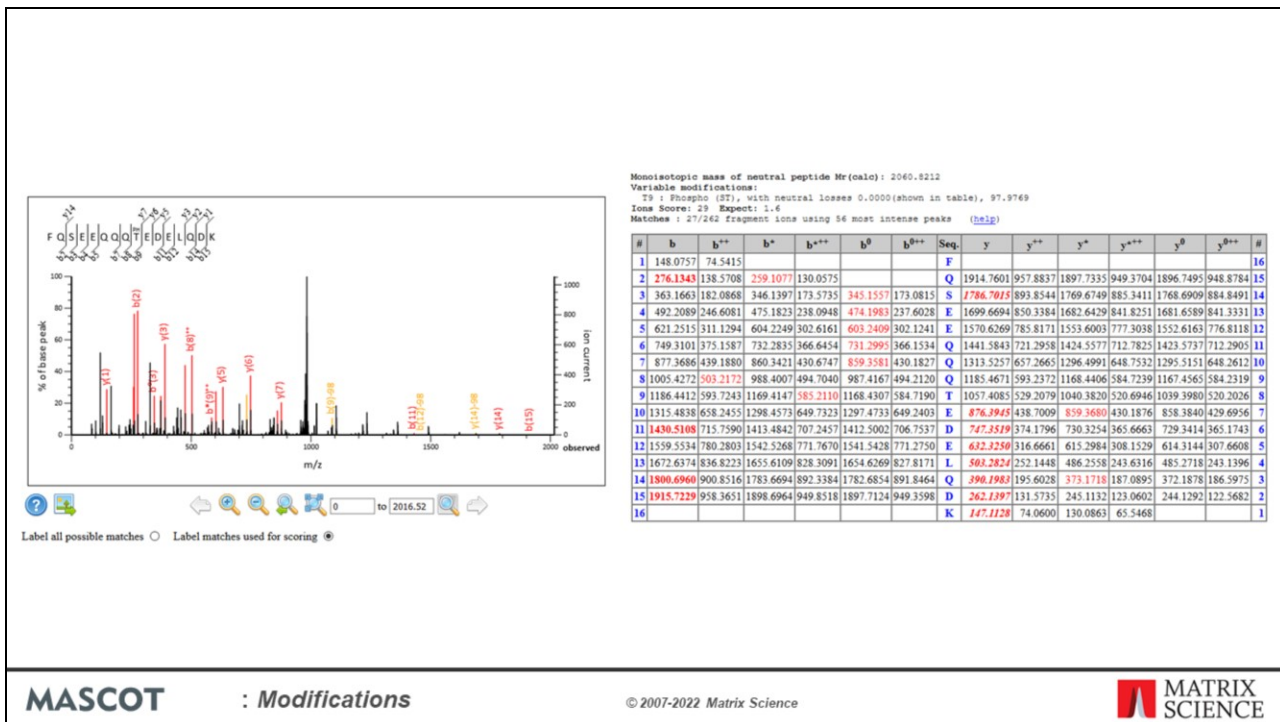
#	b	b ⁺⁺	b ⁺	b ⁺⁺	b ⁰	b ⁰⁺⁺	Seq.	y	y ⁺⁺	y ⁺	y ⁺⁺	y ⁰	y ⁰⁺⁺	#
1	148.0757	74.5415					F							16
2	276.1343	138.5708	259.1077	130.0575			Q	1914.7601	957.8837	1897.7335	949.3704	1896.7495	948.8784	15
3	443.1326	222.0700	426.1061	213.5567	425.1221	213.0647	S	1786.7015	893.8544	1769.6749	885.3411	1768.6909	884.8491	14
4	572.1757	286.5912	555.1487	278.0780	554.1647	277.5860	E	1618.7031	810.3552	1602.6766	801.8419	1601.6926	801.3499	13
5	701.2178	351.1125	684.1913	342.5993	683.2072	342.1073	E	1490.6605	745.8339	1473.6340	737.3206	1472.6500	736.8286	12
6	829.2764	415.1418	812.2498	406.6286	811.2658	406.1366	Q	1361.6179	681.3126	1344.5914	672.7993	1343.6074	672.3073	11
7	957.3350	479.1711	940.3084	470.6578	939.3244	470.1655	Q	1223.5594	617.2833	1216.5328	608.7700	1215.5485	608.2780	10
8	1085.3935	543.2004	1068.3670	534.6871	1067.3830	534.1951	Q	1105.5008	553.2540	1088.4742	544.7408	1087.4902	544.2487	9
9	1186.4412	593.7243	1169.4147	585.2110	1168.4307	584.7190	T	977.4422	489.2247	960.4156	480.7115	959.4316	480.2195	8
10	1315.4838	658.2455	1298.4573	649.7323	1297.4733	649.2403	E	876.3945	438.7009	859.3680	430.1876	858.3840	429.6956	7
11	1430.5108	715.7590	1413.4842	707.2457	1412.5002	706.7537	D	747.3519	374.1796	730.3254	365.6663	729.3414	365.1743	6
12	1559.5534	780.2803	1542.5268	771.7670	1541.5428	771.2750	E	632.3250	316.6661	615.2984	308.1529	614.3144	307.6605	5
13	1672.6374	836.8223	1655.6109	828.3091	1654.6269	827.8171	L	503.2824	252.1448	486.2558	243.6316	485.2718	243.1396	4
14	1800.6960	900.8516	1783.6694	892.3384	1782.6854	891.8464	Q	390.1983	195.6028	373.1718	187.0895	372.1878	186.5975	3
15	1915.7229	958.3651	1898.6964	949.8518	1897.7124	949.3598	D	262.1397	131.5735	245.1132	123.0602	244.1292	122.5682	2
16							K	147.1128	74.0600	130.0863	65.5468			1

MASCOT : Modifications

© 2007-2022 Matrix Science

MATRIX SCIENCE

Beautiful spectrum; long run of y ions



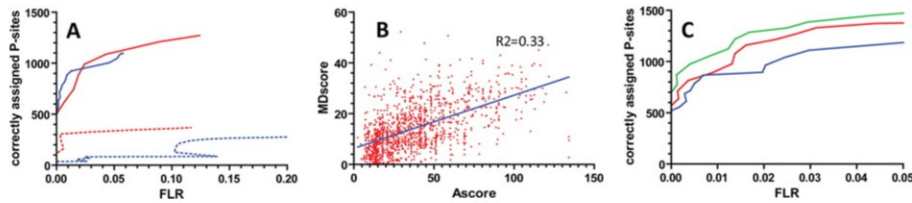
Move site to T9 and many matches would disappear

Confident Phosphorylation Site Localization Using the Mascot Delta Score[®]

Mikhail M. Savitski[‡], Simone Lemeer[§], Markus Boesche[‡], Manja Lang[‡],
Toby Mathieson[‡], Marcus Bantscheff^{‡||}, and Bernhard Kuster^{§||}

Molecular & Cellular Proteomics 10.2

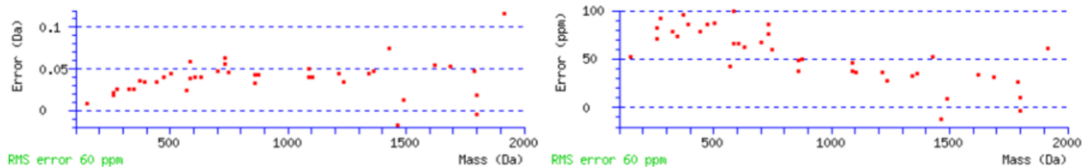
10.1074/mcp.M110.003830-1



MASCOT

**MATRIX
SCIENCE**

Mascot 2.4 reports site localisation probabilities using the delta score method published in MCP by Bernard 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.



NCBI **BLAST** search of [FQSEEQQTDELODK](#)
 (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
88.3	2060.8212	-0.0357	FQSEEQQTDELODK	Phospho S3 100.00%
28.6	2060.8212	-0.0357	FQSEEQQTDELODK	Phospho T9 0.00%
21.0	2060.9741	-0.1886	CLSLSKQVDLFEETIEK	
15.9	2060.8762	-0.0907	QMVVDKDSPHVEPEDEK	
14.5	2060.6855	0.1000	EENEEEQDDDEQSEK	
14.1	2060.9568	-0.1713	QLASGEYFLNQEQKQAK	
13.6	2060.9343	-0.1489	ITFLEELYPKDQDNEK	
12.4	2060.9862	-0.2007	SLQEGEGDLSVAEDRLSEK	
11.9	2060.9489	-0.1635	YLILCVGETLNERDSEK	
11.6	2060.9820	-0.1965	YDSFPRSDIVTVVIGADK	

MASCOT : *Modifications*

© 2007-2022 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.

Protein Family Summary

Format: Significance threshold p< 0.05 Max. number of families 20 [\[help\]](#)
 Target FDR (overrides sig. threshold) (not set) FDR type PSM
 Display non-sig. matches ☐ Min. number of sig. unique sequences 1
 Dendrograms cut at 0
 Preferred taxonomy All entries

Sensitivity and FDR (reversed protein sequences)

Proteins (1) [Report Builder](#) [\[permalink\]](#)

Protein family 1 (out of 1)

10 per page 1 Expand all Collapse all

Accession contains Find Clear

1 KAPCA_BOVIN 80 cAMP-dependent protein kinase catalytic subunit alpha OS=Bos taurus OX=9913 GN=PRKACA PE=1 SV=3


1.1 [KAPCA_BOVIN](#) Score 80 Mass 40594 Matches 1 (1) Sequences 1 (1) cAMP-dependent protein kinase catalytic subunit alpha OS=Bos taurus OX=9913 GN=PRKACA PE=1 SV=3

17 same sets of KAPCA_BOVIN

1 peptide matches (1 non-duplicate, 0 duplicate)

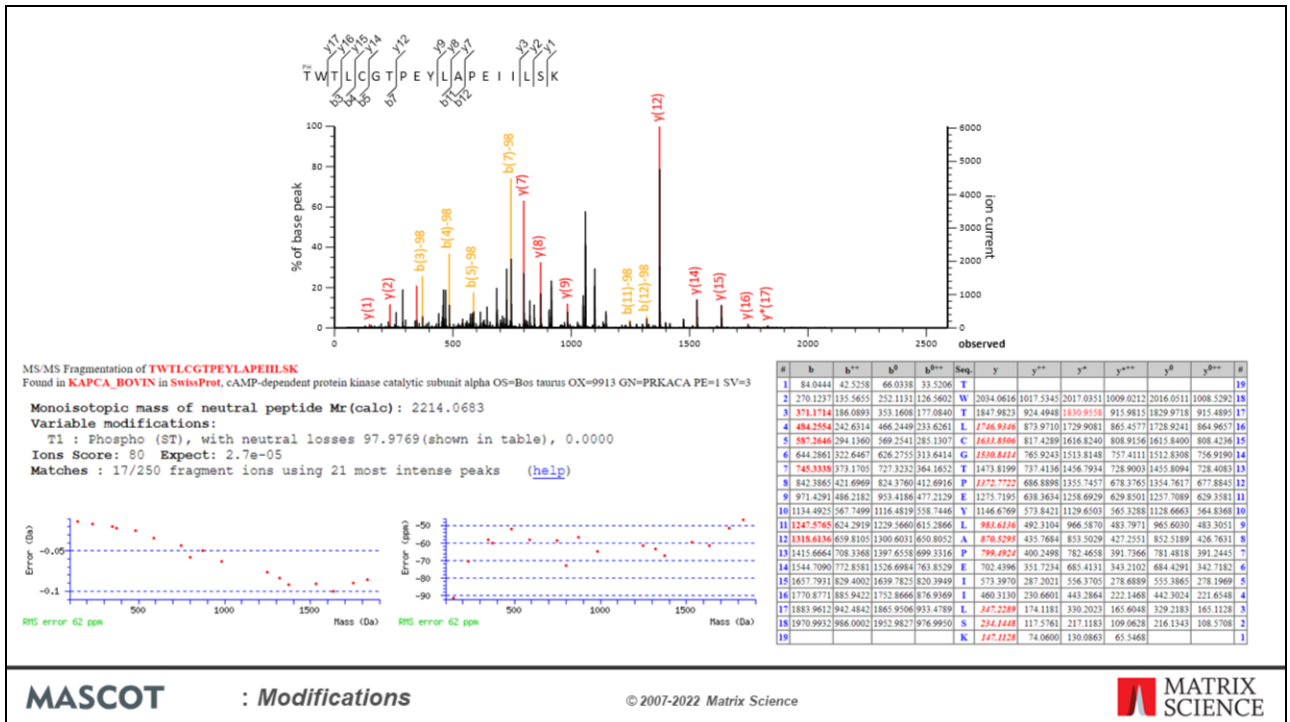
☒ Auto-fit to window

Query	Dupes	Observed	Mr (exp.)	Mr (calc)	Delta M	Score	Expect	Rank	U	Peptide
1		1107.9039	2213.7933	2214.0683	-0.2750	80	5.1e-07	1	U	R.TWTLGOTPEYLAEEILK.G + Phospho (ST)
					-0.2750	77	1.1e-06	2		R.TWTLGOTPEYLAEEILK.G + Phospho (ST)
					-0.2750	39	0.0076	3		R.TWTLGOTPEYLAEEILK.G + Phospho (ST)
					-0.2750	18	0.08	4		R.TWTLGOTPEYLAEEILK.G + Phospho (ST)
					-0.2111	13	3.1	5		GGSMGLGIPSPQVFAELK + 2 Phospho (ST)
					-0.2111	13	3.1	5		GGSMGLGIPSPQVFAELK + 2 Phospho (ST)
					-0.2111	13	3.1	5		GGSMGLGIPSPQVFAELK + 2 Phospho (ST)
					-0.2111	12	3.6	9		GGSMGLGIPSPQVFAELK + 2 Phospho (ST)
					-0.2111	12	3.6	9		GGSMGLGIPSPQVFAELK + 2 Phospho (ST)

MASCOT : Modifications © 2007-2022 Matrix Science 

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.



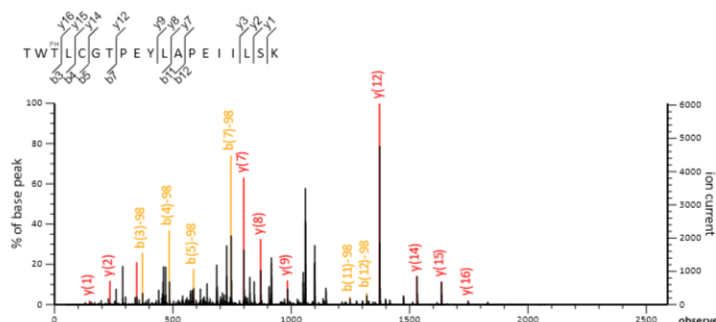
MASCOT

: **Modifications**

© 2007-2022 Matrix Science

**MATRIX
SCIENCE**

We can see why there is little difference in score between placing the phosphate on T1 or T3.



MS/MS Fragmentation of **TWTLCTGTP EYL APE I I L S K**

Found in **KAPCA_BOVIN** in **SwissProt**, cAMP-dependent protein kinase catalytic subunit alpha OS=Bos taurus OX=9913 GN=PRKACA PE=1 SV=3

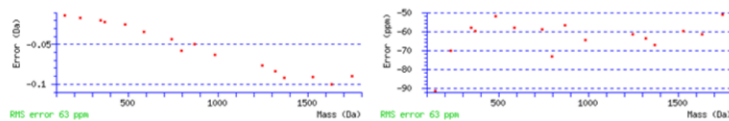
Monoisotopic mass of neutral peptide Mr(calc): 2214.0683

Variable modifications:

T3: Phospho (ST), with neutral losses 97.9769 (shown in table), 0.0000

Ions Score: 77 Expect: 6e-05

Matches : 16/254 fragment ions using 21 most intense peaks ([help](#))



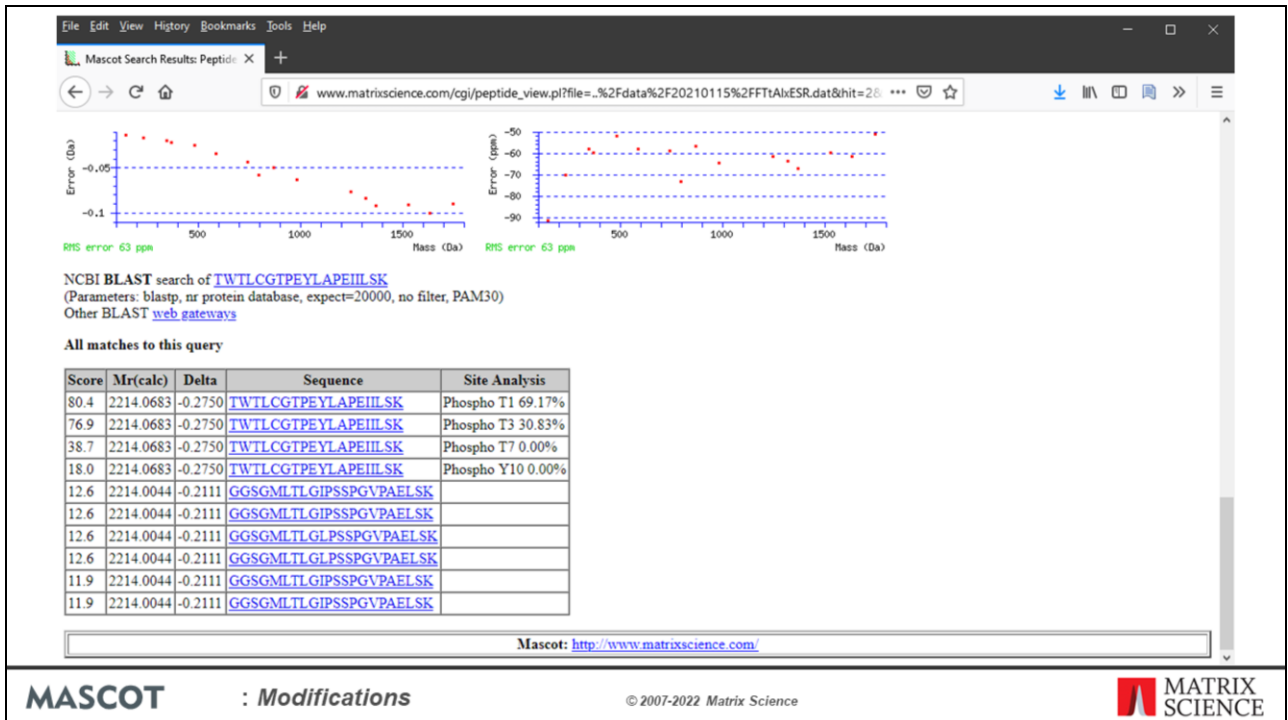
#	b	b ⁺⁺	b ⁰	b ⁰⁺⁺	Seq.	y	y ⁺⁺	y ⁰	y ⁰⁺⁺	#		
1	102.0550	51.5311	84.0444	42.5258	T	2016.0511	1008.5292	1999.0245	1000.0159	1998.0405	999.5239	18
2	288.1343	144.5708	270.1237	135.5653	W	1829.9718	915.4895	1812.9452	906.9762	1811.9612	906.4842	17
3	371.1714	186.0893	353.1608	177.0840	T	1746.9346	873.9710	1729.9081	865.4577	1728.9241	864.9657	16
4	484.2554	242.6314	466.2449	233.6261	L	1633.8768	817.4289	1616.8240	808.9136	1615.8400	808.4236	15
5	587.3048	294.1260	569.2541	285.1207	C	1538.8414	769.9243	1513.8148	757.4113	1512.8308	756.9190	14
6	644.2863	322.6467	626.2755	313.6414	G	1473.8199	737.4136	1456.7934	728.9003	1455.8094	728.4083	13
7	745.3338	373.1705	727.3232	364.1653	T	1372.7222	686.8889	1355.7457	678.3765	1354.7617	677.8845	12
8	842.3865	421.6969	824.3760	412.6916	E	1275.7195	638.3634	1258.6929	629.8501	1257.7089	629.3581	11
9	971.4291	486.2182	953.4186	477.2129	E	1146.6769	573.8421	1129.6503	565.3288	1128.6663	564.8368	10
10	1134.4925	567.7499	1116.4819	558.7446	V	983.6136	492.3104	966.5870	483.7971	965.6030	483.3051	9
11	1247.5765	624.2919	1229.5660	615.2866	L	878.5295	439.2648	853.5029	427.2551	852.5189	426.7631	8
12	1318.6136	659.8105	1300.6031	650.8052	A	799.0927	400.2498	782.4658	391.7366	781.4818	391.2445	7
13	1415.6664	708.3368	1397.6538	699.3316	P	702.4396	351.2234	685.4131	343.2102	684.4291	342.7182	6
14	1548.7090	772.8381	1528.6984	763.8529	E	573.3970	287.2021	556.3705	278.6889	555.3865	278.1969	5
15	1657.7933	829.4002	1639.7825	820.3949	I	460.3130	230.6601	443.2864	222.1468	442.3024	221.6548	4
16	1770.8771	885.9432	1752.8666	876.9369	I	347.2289	174.1181	330.2023	165.6048	329.2183	165.1128	3
17	1883.9612	942.4842	1865.9506	933.4789	L	234.1448	117.5761	217.1183	109.0628	216.1343	108.5708	2
18	1970.9932	986.0002	1952.9827	976.9950	S	147.1128	74.0600	130.0863	65.5468			1
19					K							

MASCOT : Modifications

© 2007-2022 Matrix Science

MATRIX
SCIENCE

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



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

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.

MASCOT

: *Modifications*

© 2007-2022 Matrix Science

 **MATRIX
SCIENCE**

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

Your name: Lou Scene Email: leu@res.edu

Search title: Automatic error tolerant search example

Database(s): SwissProt (AA) > <
 Spectral library (SL)
 PRIDE_Contaminants
 Amino acid (AA)
 SARS-CoV-2
 UP59680_O_sativa

Taxonomy: Mammalia (mammals)
 Enzyme: Trypsin/P
 Allow up to: 2 missed cleavages
 Quantitation: None
 Crosslinking: None
 Fixed modifications: Carbamidomethyl (C)
 Display all modifications ☐
 Variable modifications: Acetyl (N-term)
 Oxidation (M)
 Acetyl (K)
 Acetyl (Protein N-term)
 Amidated (C-term)
 Amidated (Protein C-term)
 Ammonia-loss (N-term C)
 Carbamidomethyl (N-term)
 Carbamyl (K)
 Carbamyl (N-term)
 Carboxymethyl (C)
 Cation:Na (C-term)
 Cation:Na (DE)

Peptide tol. \pm 0.8 Da \neq ^{13}C 0 MS/MS tol. \pm 0.4 Da
 Peptide charge: 1+
 Monoisotopic ☐ Average ☒
 Data file: C:\Distiller test data\Masslynx\qtof20953.mgf
 Data format: Mascot generic
 Instrument: ESI-QUAD-TOF
 Decoy: ☐
 Precursor: m/z
 Error tolerant ☒
 Report top: AUTO hits
 Start Search ... Reset Form

MASCOT : Modifications

© 2007-2022 Matrix Science

MATRIX
SCIENCE

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

► Search parameters

► Score distribution

▼ Modification statistics for all protein families

Modification	Delta	Type	Site	Total matches
Oxidation	15.994915	variable	M	10
Carbamidomethyl	57.021464	fixed	C	8
Carbamidomethyl	57.021464	ET	M	6
Acetyl	42.010565	variable	N-term	6
Non-specific cleavage		ET	-	5
Carbamidomethyl	57.021464	ET	N-term	2
Gly	57.021464	ET	K	2
Lys->CamCys	31.935685	ET	K	1
Label:15N(1)	0.997035	ET	V	1
Deamidated	0.984016	ET	N	1
Hydroxamic_acid	15.010899	ET	D	1
Hex(1)HexNAc(1)Kdn(1)Sulf(1)	695.157878	ET	T	1
Dethiomethyl	-48.003371	ET	M	1
Delta:H(2)C(2)	26.015643	ET	N-term	1
Succinyl	100.01604	ET	N-term	1
Methyl	14.01565	ET	E	1
Sulfo	79.956815	ET	C	1
Iodo	125.896648	ET	Y	1
Carboxymethyl	58.005481	ET	N-term	1
Gly	57.021464	ET	S	1
DiLeu4plex118	145.140471	ET	K	1
dHex(1)Hex(3)	632.216379	ET	T	1
Acetyl	42.010565	ET	S	1
Cys->Dha	-33.987721	ET	C	1
Xle->His	23.974848	ET	L	1
Gln->pyro-Glu	-17.026532	ET	N-term	1
Methylamine	13.031634	ET	T	1
Asp->Asn	-0.984016	ET	D	1
Methyl+Deamidated	14.999666	ET	N	1
Diethylphosphothione	152.006087	ET	K	1
Propyl	42.04695	ET	E	1

► Legend

MASCOT

: Modifications

© 2007-2022 Matrix Science



At the top of the protein family report there is an expandable section that summarises the modifications. You can see counts of the number of queries matching fixed and variable modifications used in the first pass of the search and then for the results of the Error Tolerant search.

Mascot database search: Error | Automatic error tolerant search: +

archive-win10/mascot_2_7_00/master_results_2.pl?file=%2Fdata%2F981130.dat_ign 80%

Accession: contains: Find Clear

▼2

1 PPB1_HUMAN
3 PPB1_HUMAN
2 PPB1_HUMAN

487 Alkaline phosphatase, placental type OS=Homo sapiens OX=9606 GN=ALPP PE=1 SV=2
69 Intestinal-type alkaline phosphatase OS=Homo sapiens OX=9606 GN=ALPI PE=1 SV=2
343 Alkaline phosphatase, germ cell type OS=Homo sapiens OX=9606 GN=ALPG PE=2 SV=4

Threshold (0): 0 Cut

	Score	Mass	Matches	Sequences
2.1 PPB1_HUMAN	487	58259	31 (12)	16 (11) Alkaline phosphatase, placental type OS=Homo sapiens OX=9606 GN=ALPP PE=1 SV=2
2.2 PPB1_HUMAN	343	57626	25 (9)	12 (9) Alkaline phosphatase, germ cell type OS=Homo sapiens OX=9606 GN=ALPG PE=2 SV=4
2.3 PPB1_HUMAN	69	57119	8 (2)	7 (2) Intestinal-type alkaline phosphatase OS=Homo sapiens OX=9606 GN=ALPI PE=1 SV=2

Redisplay All None

▼34 peptide matches (34 non-duplicate, 0 duplicate)
Auto-fit to window

Query Dups	Observed	Mr (expt)	Mr (calc)	Delta M	Score	Expect	Rank	U	1	2	3	Peptide
d11	517.1760	1032.3375	1032.5604	-0.2229	70	1.1e-05	1	1	1	1	1	R. GSSIFGLAPQK.A
d14	532.1837	1042.3528	1042.5710	-0.2181	60	0.00016	1	1	1	1	1	R. GSSIFGLAPQK.A
d13	545.6818	1089.3491	1089.5819	-0.2327	54		1	1	1	1	1	R. GSSIFGLAPQK.A + [+57.0215 at 82]
d15	547.4564	1133.3987	1133.5439	-0.2511	45	0.007	1	1	1	1	1	R. GSSIFGLAPQK.A + Oxidation (M)
d16	614.9051	1226.3854	1226.6339	-0.2473	38	0.039	1	1	1	1	1	R. LQWFLAPQK.F + Oxidation (M)
d100	653.2101	1304.4057	1304.6037	-0.2780	87	3.6e-07	1	1	1	1	1	R. GNPFTGLAQAAR.F
d124	710.2235	1418.4324	1418.7154	-0.2829	92		1	1	1	1	1	R. ANPFTGLAQAAR.F + [+100.0160 at N-term]
d126	726.1806	1450.3465	1450.6477	-0.3011	69	7e-06	1	1	1	1	1	R. NYSADYVQAAR.Q
d133	499.1349	1494.3828	1494.6694	-0.2866	89		1	1	1	1	1	R. DFLSKNTAQAAR.L + 2 Oxidation (M) ←
d136	754.6864	1507.3582	1507.6691	-0.3109	43		1	1	1	1	1	R. NYSADYVQAAR.Q + [+57.0215 at N-term]
d145	526.1538	1575.4394	1575.7780	-0.3384	81		1	1	1	1	1	R. ALYETIMFQDAER.A + [-48.0034 at M7]
d156	820.7083	1639.4420	1639.7763	-0.3343	104	6.3e-09	1	1	1	1	1	R. ALYETIMFQDAER.A + Oxidation (M)
d158	552.5010	1654.4812	1654.8315	-0.3503	54		1	1	1	1	1	R. NYDNGATATATLCQVE.G + [-33.9877 at C18]
d162	836.2372	1670.4598	1670.8052	-0.3454	88		1	1	1	1	1	R. VLPADKSNFAPWQ.Q
d165	841.2310	1680.4474	1680.8029	-0.3554	93		1	1	1	1	1	R. ALYETIMFQDAER.A + [+57.0215 at M7]
d170	864.2888	1726.5629	1726.9294	-0.3664	43	0.0036	1	1	1	1	1	R. AYTVLLYQGFVPLK.D
d175	586.4951	1756.4635	1756.8420	-0.3786	48		1	1	1	1	1	R. IIPVEKDFQWQK ←
d176	879.2425	1756.4705	1756.8420	-0.3715	83		1	1	1	1	1	R. IIPVEKDFQWQK
d179	593.4854	1777.4285	1777.7764	-0.3478	46		1	1	1	1	1	R. NYDNGATATATLCQVE.G + [+31.9357 at C-term K]
d208	975.8100	1945.4055	1950.0245	-0.4190	86	7.7e-08	1	1	1	1	1	R. NLLIFGLDGGGVVQDAAR.I + Oxidation (M)
d209	876.3360	1968.4334	1968.8628	-0.4301	57		1	1	1	1	1	R. NLLIFGLDGGGVVQDAAR.I

MASCOT : Modifications © 2007-2022 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 queries 175 and 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

Mascot database search Error: X Automatic error tolerant search: X

archive-win10/mascot_2_7_00/cgi/master_results_2.pl?file=.%2Fdata%2F981130.dat_ignoresions: F ☆

Accession contains Find Clear

2.2 PPB_HUMAN 69 57119 8 (2) 7 (2) Intestinal-type alkaline phosphatase OS=Homo sapiens OX=9606 GN=ALPI PE=1 SV=2

Redisplay All None

▼ 34 peptide matches (34 non-duplicate, 0 duplicate)

Auto-fit to window

Query	Dupes	Observed	Mr (expt)	Mr (calc)	Delta M	Score	Expect	Rank	U	1	2	3	Peptide
41		517.1760	1032.3375	1032.5604	-0.2229	70	1.1e-05	1					R.GSSIFGLAPK.A
46		532.1837	1062.3528	1062.5710	-0.2181	60	0.00016	1					R.GSSIFGLAPK.A
53		545.6818	1089.3491	1089.5819	-0.2327	54		1					R.GSSIFGLAPK.A + [+57.0215 at S2]
65		567.6566	1133.2987	1133.5499	-0.2511	45	0.007	1					R.GNEVLSVMNR.A + Oxidation (M)
66		614.2001	1226.3856	1226.6329	-0.2473	28	0.039	1					K.LQPEIPLAMDR.F + Oxidation (M)
100		653.2101	1304.4057	1304.6837	-0.2780	87	3.6e-07	1					K.GNFTIGLSAAR.F
124		710.2235	1418.4324	1418.7154	-0.2829	92		1					K.ANFTIGLSAAR.F + [+100.0160 at N-term]
126		726.1806	1450.3465	1450.6477	-0.3011	69	7e-06	1					R.WNYSDADYFASAR.Q
133		499.1349	1494.3828	1494.6694	-0.2866	89		1					L.DPSLMEHTAALR.L + 2 Oxidation (M)
136		754.6864	1507.3592	1507.6691	-0.3109	43		1					R.WNYSDADYFASAR.Q + [+57.0215 at N-term]
145		526.1538	1575.4396	1575.7780	-0.3384	81		1					R.ALTIMFDC
156		820.7283	1639.4420	1639.7763	-0.3343	106	6.3e-09	1					R.ALTIMFDC
158		552.5010	1654.4812	1654.8315	-0.3503	54		1					K.RVPSGATAT
162		836.2372	1670.4598	1670.8052	-0.3454	88		1					G.VIPAEKENPA
165		841.2310	1680.4474	1680.8029	-0.3554	93		1					R.ALTIMFDDAAR.A + [+57.0215 at M7]
170		864.2888	1726.5629	1726.9294	-0.3664	43	0.0036	1					K.AYTVLLYGGPGTVLK.D
175		586.4951	1756.4635	1756.8420	-0.3786	48		1					G.IIPVEENPQPMNR
176		879.2425	1756.4705	1756.8420	-0.3715	83		1					G.IIPVEENPQPMNR
179		593.4834	1777.4285	1777.7764	-0.3478	46		1					K.RVPSGATATAYLCQVK.G + [+31.9357 at C-term K]
208		975.8100	1949.6055	1950.0245	-0.4190	86	7.7e-08	1					K.NLIIFLDGNGVSTVTAAR.I + Oxidation (M)
209		976.2340	1950.4534	1950.8555	-0.4021	1	0.024	1					K.DGAPQVTESESGSYR.Q
211		656.1752	1965.5039	1966.8712	0.6327	1	68						K.DGAPQVTESESGSYR.Q + [+14.0156 at R11]
213		664.5518	1990.6336	1991.0510	-0.4174	65		1					K.NLIIFLDGNGVSTVTAAR.I + [+57.0215 at M10]
214		665.1736	1992.4991	1992.9132	-0.4141	54		1					R.DSTLDPSLMEHTAALR.L + 2 [+57.0215 at M9,M11]

Possible assignments:
 Carbamidomethyl (N-term) [+57.0215]
 Carboxymethyl (N-term) [+58.0055]
 Delta-H(6)C(3)O(1) (Protein N-term) [+58.0419]

MASCOT : Modifications

© 2007-2022 Matrix Science

MATRIX SCIENCE

Take a look at the match to query 136. 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 can be seen in this screen shot for three other queries

Mascot database search: Error X

Auto10 error tolerant search: X

archive-win10/mascot_2_7_00/cgi/master_results_2.pl?file=%2Fdata%2FF981130.dat_ignoreions: 0

Find

Clear

Accession

contains

Find

Clear

Auto-fit to window

Query Dups	Observed	Mr (expt)	Mr (calc)	Delta M	Score	Expect	Rank	U	1	2	3	Peptide
d126	726.1806	1450.3465	1450.6477	-0.3011	0	69	7e-06	1	■	■	■	R.WNYSDADVPSAR.Q
d133	499.1349	1494.3828	1494.6694	-0.2866	0	89		1	■	■	■	L.DPSLMEMTSALR.L + 2 Oxidation (M)
d136	754.6864	1507.3582	1507.6691	-0.3109	0	43		1	■	■	■	R.WNYSDADVPSAR.Q + [+57.0215 at N-term]
d145	526.1538	1575.4396	1575.7780	-0.3384	0	81		1	■	■	■	R.ALTETIMFDGAIER.A + [-48.0034 at M7]
d156	820.7283	1639.4420	1639.7763	-0.3343	0	106	6.3e-09	1	■	■	■	R.ALTETIMFDGAIER.A + Oxidation (M)
d158	552.5010	1654.4812	1654.8315	-0.3503	0	54		1	■	■	■	K.HVDSGATATAYLCQVK.G + [-33.9877 at C14]
d162	836.2372	1670.4598	1670.8052	-0.3454	0	88		1	U	■	■	G.VIPAEENPAPNRR.Q
d165	841.2310	1680.4474	1680.8029	-0.3554	0	93		1	■	■	■	R.ALTETIMFDGAIER.A + [+57.0215 at M7]
d170	864.2888	1726.5629	1726.9294	-0.3664	0	43	0.0036	1	■	■	■	K.ATTVLLYNGPOVLK.D
d175	586.4951	1756.4635	1756.8420	-0.3786	0	48		1	■	■	■	G.IIPVEENPFPNRR
d176	879.2425	1756.4705	1756.8420	-0.3715	0	83		1	■	■	■	G.IIPVEENPFPNRR
d179	593.4834	1777.4285	1777.7764	-0.3478	0	46		1	■	■	■	K.HVDSGATATAYLCQVK.G + [+31.9357 at C-term K]
d208	975.8100	1949.6055	1950.0245	-0.4190	0	86	7.7e-08	1	■	■	■	K.NLIIFLDGDMGVSTVTAAR.I + Oxidation (M)
d209	976.2340	1950.4534	1950.8555	-0.4021	1	27	0.024	1	■	■	■	K.DGARFQVTESESGSPETR.Q
d211	656.1752	1965.5039	1964.8712	0.6327	1	68		1	■	■	■	K.DGARFQVTESESGSPETR.Q + [+14.0156 at M11]
d213	664.5518	1990.6336	1991.0510	-0.4174	0	65		1	■	■	■	K.NLIIFLDGDMGVSTVTAAR.I + [+57.0215 at M10]
d214	665.1736	1992.4991	1992.9132	-0.4141	0	54		1	U	■	■	R.DSTLDPSLMENTAALR.L + 2 [+57.0215 at M9,M11]
d216	1001.2027	2000.3908	2000.8058	-0.4150	0	65	3.4e-05	1	U	■	■	R.MGTFQPEYFPDYSQQGTR.L + Oxidation (M)
d217	667.8046	2000.3919	2000.8058	-0.4139	0	73	8.9e-07	1	U	■	■	R.MGTFQPEYFPDYSQQGTR.L + Oxidation (M)
d218	670.1561	2007.4466	2007.8770	-0.4304	1	75		1	■	■	■	K.DGARFQVTESESGSPETR.Q + Acetyl (N-term) [+15.0109 at N-term S]
d222	681.8205	2042.4397	2041.8324	0.6073	0	79		1	U	■	■	R.MGTFQPEYFPDYSQQGTR.L + [+57.0215 at N-term M]
d245	766.2128	2295.6165	2296.1084	-0.4919	0	55	3.3e-05	1	U	■	■	R.QQSAVPLDGETHAGEDVAVFAR.G
d252	784.5440	2350.6103	2351.1030	-0.4927	0	68		1	U	■	■	R.QQSAVPLDGETHAGEDVAVFAR.G + [-17.0265 at N-term]
d253	790.2186	2367.6341	2368.1295	-0.4954	0	93	1e-08	1	U	■	■	R.QQSAVPLDGETHAGEDVAVI
d260	809.2208	2424.6406	2425.1510	-0.5104	0	66		1	U	■	■	R.QQSAVPLDGETHAGEDVAVI
d274	914.9160	2741.7263	2741.2306	0.4956	0	44		1	■	■	■	R.QREGCDIATQLINMDIDVILGGGR.K + Oxidation (M) [+79.9588 at C4]
d275	920.5878	2758.7415	2759.3218	-0.5804	0	126		1	■	■	■	R.QREGCDIATQLINMDIDVILGGGR.K + [+57.0215 at M15]

1


VEC MIIMAN

767

Search results contain 1000000 results. You can view the first 10000 results. You can view the first 10000 results.

MASCOT : Modifications

© 2007-2022 Matrix Science

MATRIX SCIENCE

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

Mascot database search: Error

Automatic error tolerant search

archive-win10/mascot_2_7_00/cgi/master_results_2.pl?file=%2Fdata%2F981130.dat_ignoreions:

Accession

contains

Find

Clear

Auto-fit to window

Query Dups	Observed	Mr (expt)	Mr (calc)	Delta M	Score	Expect	Rank	U	1	2	3	Peptide
#41	517.1760	1032.3375	1032.5604	-0.2229	0	70	1.1e-05	1	■	■	■	R.GSSIFGLAPGK.A
#46	532.1837	1062.3528	1062.5710	-0.2181	0	60	0.0016	1	U	■	■	R.GSSIFGLAPGK.A
#53	545.6818	1089.3491	1089.5819	-0.2327	0	54		1	■	■	■	R.GSSIFGLAPGK.A + [+57.0215 at S2]
#65	567.6566	1133.2987	1133.5499	-0.2511	0	45	0.007	1	■	■	■	R.GNEVISVMNR.A + Oxidation (M)
#66	614.2001	1226.3856	1226.6329	-0.2473	0	28	0.039	1	U	■	■	K.LGPEIPLAMDR.F + Oxidation (M)
#100	653.2101	1304.4057	1304.6837	-0.2780	0	87	3.6e-07	1	■	■	■	K.GNFQTIGLSAAAR.F
#124	710.2235	1418.4324	1418.7154	-0.2829	0	92		1	U	■	■	K.ANFQTIGLSAAAR.F + [+100.0160 at N-term]
#126	726.1806	1450.3465	1450.6477	-0.3011	0	69	7e-06	1	■	■	■	R.NNYSDADVPASAR.Q
#133	499.1349	1494.3828	1494.6694	-0.2866	0	89		1	■	■	■	L.DPSLMENTEALR.L + 2 Oxidation (M)
#136	754.6864	1507.3582	1507.6691	-0.3109	0	43		1	■	■	■	R.NNYSDADVPASAR.Q + [+57.0215 at N-term]
#145	526.1538	1575.4396	1575.7780	-0.3384	0	81		1	■	■	■	R.ALTTETIMFDDAER.A + [-48.0034 at M7]
#156	820.7283	1639.4420	1639.7763	-0.3343	0	106	6.3e-09	1	■	■	■	R.ALTTETIMFDDAER.A + Oxidation
#158	552.5010	1654.4812	1654.8315	-0.3503	0	54		1	■	■	■	K.HVFDGGATATAYLCGVK.G + [-33.6
#162	836.2372	1670.4598	1670.8052	-0.3454	0	88		1	U	■	■	G.VIPAEKENPAPWNR.Q
#165	841.2310	1680.4474	1680.8029	-0.3554	0	93		1	■	■	■	R.ALTTETIMFDDAER.A + [+57.0215 at M7]
#170	866.2888	1726.5629	1726.9294	-0.3664	0	43	0.0036	1	■	■	■	K.ATVLLYNGSPQVLK.D
#175	586.4951	1756.4635	1756.8420	-0.3786	0	48		1	■	■	■	G.IIPVEKENPQWNR
#176	879.2425	1756.4705	1756.8420	-0.3715	0	83		1	■	■	■	G.IIPVEKENPQWNR
#179	593.4834	1777.4285	1777.7764	-0.3478	0	46		1	■	■	■	K.HVFDGGATATAYLCGVK.G + [+31.9357 at C-term K]
#208	975.8100	1949.6055	1950.0245	-0.4190	0	86	7.7e-08	1	■	■	■	K.NLIIFLGDGMGVSTVTAAR.I + Oxidation (M)
#209	976.2340	1950.4534	1950.8555	-0.4021	1	27	0.024	1	■	■	■	K.DGARDVTESESGSPYR.Q
#211	656.1752	1965.5039	1964.8712	0.6327	1	68		1	■	■	■	K.DGARDVTESESGSPYR.Q + [+14.0156 at M11]
#213	664.5518	1990.6336	1991.0510	-0.4174	0	65		1	■	■	■	K.NLIIFLGDGMGVSTVTAAR.I + [+57.0215 at M10]
#214	665.1736	1992.4991	1992.9132	-0.4141	0	54		1	U	■	■	R.DSTLDPSLMENTEALR.L + 2 [+57.0215 at M9,M11]
#216	1001.2027	2000.3908	2000.8058	-0.4150	0	65	3.4e-05	1	U	■	■	R.MOTFDPPEYFDQYSGQOTR.L + Oxidation (M)
#217	667.8046	2000.3919	2000.8058	-0.4139	0	73	8.9e-07	1	U	■	■	R.MOTFDPPEYFDQYSGQOTR.L + Oxidation (M)
#218	670.1561	2007.4466	2007.8770	-0.4304	1	75		1	■	■	■	K.DGARDVTESESGSPYR.Q + Acetyl (N-term): [+15.0109 at N-term D
#222	681.8205	2042.4397	2041.8324	0.6073	0	79		1	■	■	■	R.MOTFDPPEYFDQYSGQOTR.L + [+57.0215 at N-term M]

Possible assignments:
Dethiomethyl (M) [-48.0034]

VEC MIMAN

Timeline contains Minus Year DC Minus contains PG-BASE GC-VECI SE-L-CL-3

MASCOT

: Modifications

© 2007-2022 Matrix Science

MATRIX
SCIENCE

Possible assignments:
Dethiomethyl (M) [-48.0034]

Query 145 is an interesting case. There are 2 other matches to the same peptide, oxidised at M7 and carbamidomethylation at M7. The minus 48 modification occurs when the methionine is oxidised and then loses the side chain as methanesulfenic acid.



MASCOT : Modifications

© 2007-2022 Matrix Science



You should also look at the other matches to the same query when trying to decide whether to accept a match or not. For this query the top two matches are essentially to the same sequence. The error tolerant match is to a peptide that has undergone deamidation of aspartic acid to asparagine. However, the original match from the first pass search, which is a match to the unmodified peptide with a slightly lower score and two less ions, has a precursor measurement error of -0.5Da compared to +0.5Da for the modified match, that is in line with the majority of matches to more abundant proteins. It is likely that both forms of the peptide are present in the sample and this is in fact a chimeric spectra.

Automatic error tolerant search: x

archive-win10/mascot_2.7.00/master_results_2.pl?file=%2Fdata%2F981130.dat_ignoreon: ...

Accession contains Find Clear

1 subset or intersection (1 subset protein in total)

7 RPN2_HUMAN 113 Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 2 OS=Homo sapiens

7.1 RPN2_HUMAN Score 113 Mass 69355 Matches 3 (2) Sequences 3 (2) Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 2 OS=Homo sapiens OX=9606 GN=RPN2 PE=1 ...

2 same sets of RPN2_HUMAN

3 peptide matches (3 non-duplicate, 0 duplicate)

Auto-fit to window

Query Dups	Observed	Mr (expt)	Mr (calc)	Delta M	Score	Expect	Rank	U	Peptide
530	539.6746	1077.3347	1077.5818	-0.2471	0	60	9.2e-05	1	U R.YIANTVELR.V
139	766.7314	1531.4482	1531.7518	-0.3036	0	86	3.6e-07	1	U K.TQGEVVFVAEPDNK.N
246	766.7314	2297.1723	2297.2318	-0.0594	1	68		1	U R.LINQKTOGEVVFVAEPDNK.N * [+145.1405 at K5]

1 subset or intersection (3 subset proteins in total)

8 BASI_BOVIN 104 Basigin (Fragment) OS=Bos taurus OX=9913 GN=BSG PE=2 SV=1

9 KPVM_HUMAN 92 Pyruvate kinase PKM OS=Homo sapiens OX=9606 GN=PKM PE=1 SV=4

10 CBPM_HUMAN 83 Carboxypeptidase M OS=Homo sapiens OX=9606 GN=CPM PE=1 SV=2

10 per page 1 2 3 Next Expand all Collapse all

Not what you expected? Try the peptide summary.

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

Possible assignments:
 DiLeu4plex118 (K) [+145.1405]
 DiLeu4plex (K) [+145.1322]
 DiLeu4plex117 (K) [+145.1283]
 DiLeu4plex115 (K) [+145.1200]
 CAmthiopropionyl (K) [+145.0197]

MASCOT : Modifications

© 2007-2022 Matrix Science

MATRIX SCIENCE

I do not have a good solution for this match. The modifications are either from DiLeu labelled quantitation experiments or a side product in crosslinking while the sample will have been exposed to neither of these two scenarios. Most likely the +145 at the C terminal is a combination of modifications in the C-terminal region but I don't know exactly what.

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.

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