

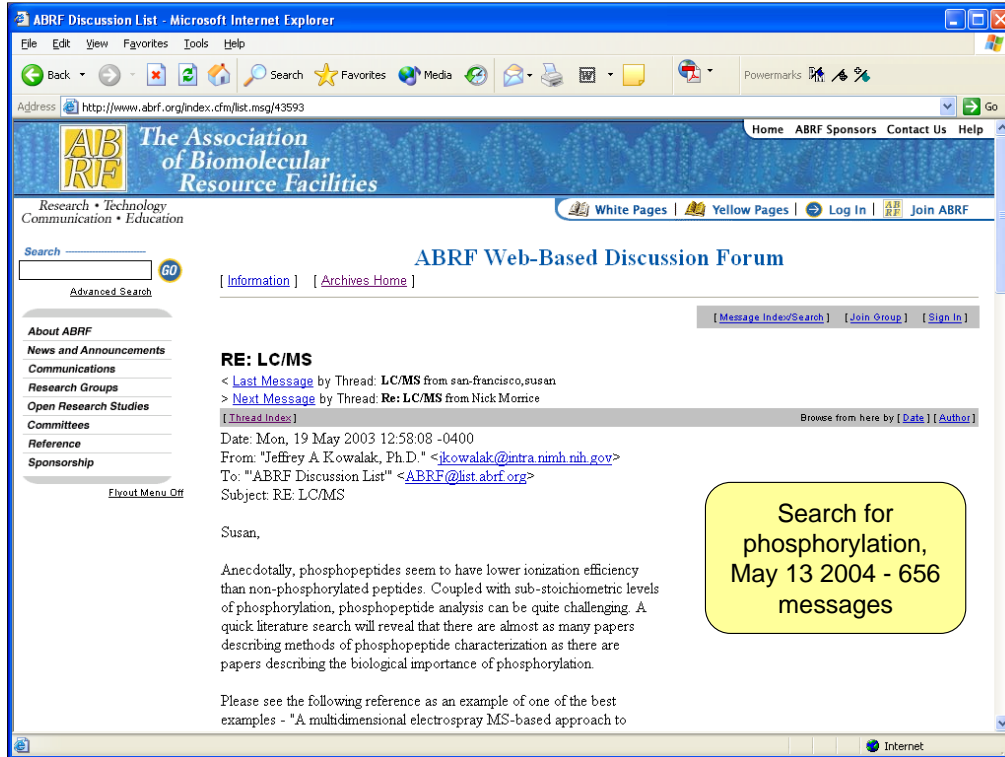
Phosphorylation Site Analysis

ASMS 2004

**{MATRIX}
{SCIENCE}**

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.



One of the best sources of information on phosphorylation is the ABRF email discussion group. When I checked a few days ago, there were 656 messages on this topic

UniMod: Basic Data View - Microsoft Internet Explorer

Address: http://www.unimod.org/cgi/unimod.cgi?records_per_page=25&columns_to_view=full_name&columns_to_view=code_name&columns_to_view=mono_mass&columns_to_view=avge

UNIMOD protein modifications for mass spectrometry

View All Records | Add Record

Help | Home | Options | Advanced Search | Logout

Search: Go

Result Set

Modification	Short name	Monoisotopic	Average	Composition	Details
Phosphorylation	Phospho	79.966331	79.9799	H O(3) P	ⓘ
Phosphorylation without neutral loss	Phospho-NL	79.966331	79.9799	H O(3) P	ⓘ
phosphorylation to amine thiol	ser_thr_DAET	87.050655	87.1866	H(9) C(4) N O(-1) S	ⓘ
thioacylation of primary amines (N-term and Lys)	DSP	87.998285	88.1283	H(4) C(3) O S	ⓘ
C13 label (Phosphotyrosine)	13C9_Phospho_Tyr	88.996524	88.9138	H C(-9) C13(9) O(3) P	ⓘ
Acrolein addition +94	Acrolein94	94.041865	94.1112	H(6) C(6) O	ⓘ
N-isopropylcarboxamidomethyl	NIPCAM	99.068414	99.1311	H(9) C(5) N O	ⓘ
Succinic anhydride labeling reagent light form (N-term & K)	Suc_anh_light	100.016044	100.0728	H(4) C(4) O(3)	ⓘ
labeling reagent light form (N-term & K)	benzoyl	104.026215	104.1061	H(4) C(7) O	ⓘ
Succinic anhydride labeling reagent, heavy form (+4amu, 4C13), N-term & K	Suc_anh+4C13	104.029463	104.0434	H(4) C13(4) O(3)	ⓘ
Succinic anhydride labeling reagent, heavy form (+4amu, 4H2), N-term & K	Suc_anh+4H2	104.041151	104.0974	H2(4) C(4) O(3)	ⓘ
S-pyridylethylation	S-pyridylethyl	105.057849	105.1372	H(7) C(7) N	ⓘ
Acrolein addition +112	Acrolein112	112.052430	112.1265	H(8) C(6) O(2)	ⓘ
ubiquitinylation residue	GlyGly	114.042927	114.1026	H(6) C(4) N(2) O(2)	ⓘ
Pyridyl	Pyridyl	119.037114	119.1207	H(5) C(7) N O	ⓘ
N-ethylmaleimide on cysteines	NEM	125.047679	125.1253	H(7) C(6) N O(2)	ⓘ
Iodination	Iodination	125.896648	125.8965	H(-1) I	ⓘ

How is phosphorylation handled in Mascot?

Whether you use the public web site or whether you have Mascot in-house, we recommend using the Unimod web site to browse and update the list of modifications

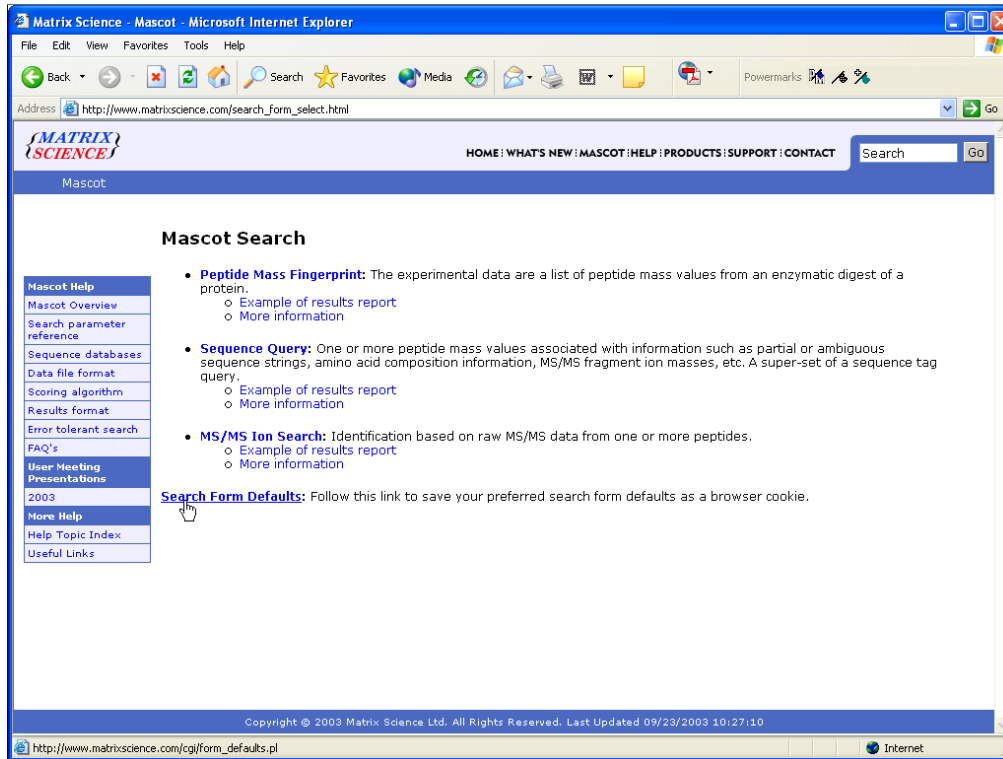
The screenshot shows the UniMod website interface in a Microsoft Internet Explorer browser window. The page title is "UniMod: Record Details - Microsoft Internet Explorer". The address bar shows the URL: http://www.unimod.org/cgi/unimod.cgi?sort_field1=mono_mass&sort_field2=full_name&first_record_to_display=0&records_per_page=25&session_id=ST2QYTWUEQXJLGMT&recor. The page content includes a search bar, navigation links (View All Records, Add Record, Help, Home, Options, Advanced Search, Logout), and a detailed table of modification information.

Record Details			
Accession #	21	Short name	Phospho
Modification	Phosphorylation		
Composition	H O(3) P	Monoisotopic	79.966331
		Average	79.9799
Specificity Definition 1			
Site	S	Position	Anywhere
Neutral Loss	H(3) O(4) P	Monoisotopic	97.976896
		Average	97.9952
Classification	Post-translational		
Comment			
Specificity Definition 2			
Site	T	Position	Anywhere
Neutral Loss	H(3) O(4) P	Monoisotopic	97.976896
		Average	97.9952
Classification	Post-translational		
Comment			
Specificity Definition 3			
Site	Y	Position	Anywhere
Neutral Loss		Monoisotopic	
		Average	
Classification	Post-translational		
Comment			
Notes and References			
Notes			
Curator	unimod	Last Modified	2002-10-04 15:53:04

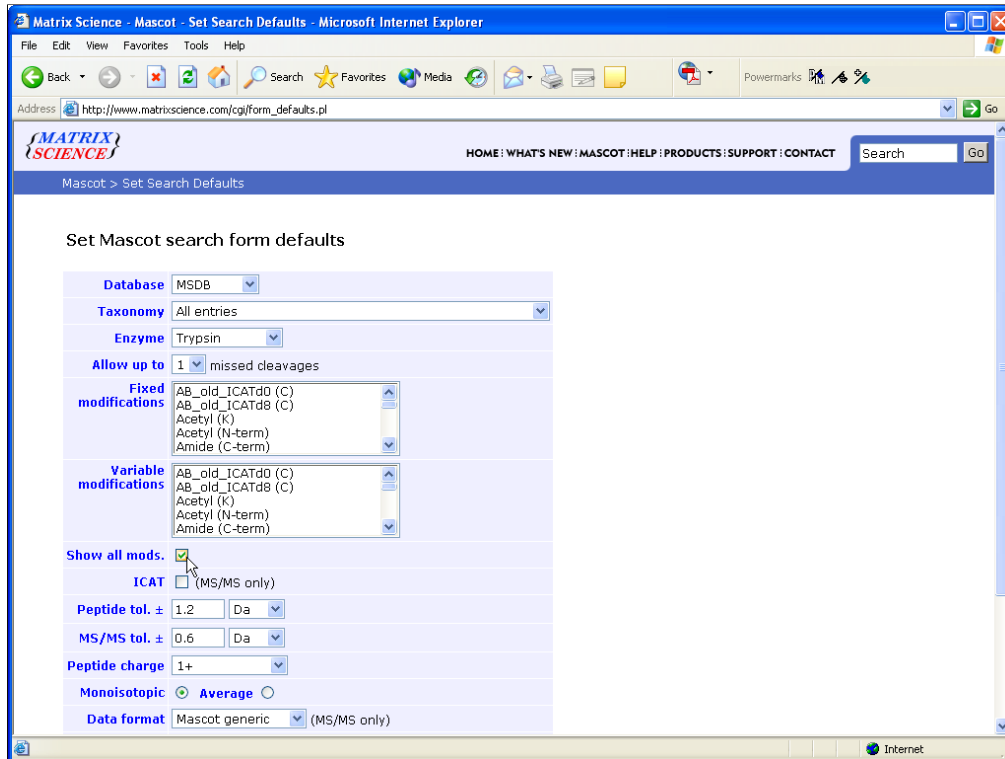
At the bottom right of the table, there is a link: [Email Change Request](#).

The standard modifications, that appear in the short, default list, assume that serine and threonine predominantly lose phosphate as a neutral loss of 98 Da, while tyrosine fragment ions stay intact

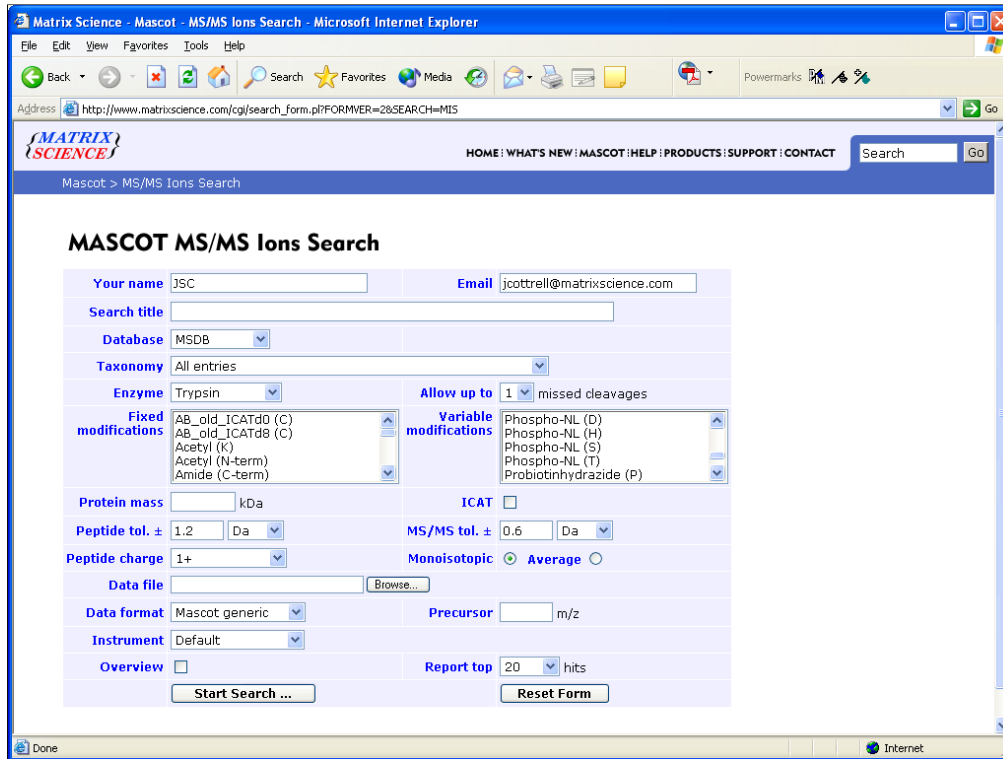
However, several other possibilities are also defined. For example, serine and threonine without any neutral loss. You don't see these listed on the short, default list of mods



If you want to see the complete list of mods, you just need to choose this option in the search form defaults. Follow the link at the bottom of the search form selection page



And check the box for Show all mods



Then, you'll see all of the mods listed in the search form.

Peptide Summary Report (../data/20040513/FamTlxHm.dat) - Microsoft Internet Explorer

Address: http://www.matrixscience.com/cgi/master_results.pl?file=../data/20040513/FamTlxHm.dat

Peptide Summary Report

Format As: Peptide Summary [Help](#)

Significance threshold p < 0.05 Max. number of hits 20

Standard scoring MudPIT scoring Ions score cut-off 0 Show sub-sets

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

Select All Select None Search Selected Error tolerant

1. [KBBOA2](#) Mass: 25091 Score: 77 Queries matched: 1
 beta-casein precursor - bovine
 Check to include this hit in error tolerant search

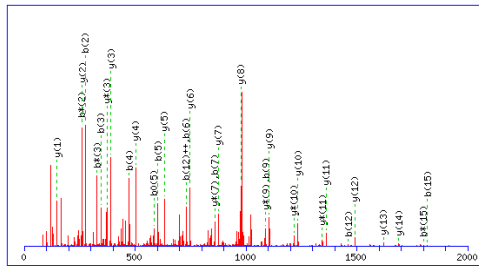
Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
<input checked="" type="checkbox"/>	1031.40	2060.78	2060.82	-0.04	0	77	6.2e-05	1	FOSEEQQTDELDQK + Phospho (ST)

Top scoring peptide matches to query 1
 Score greater than 35 indicates homology
 Score greater than 47 indicates identity
 Status bar shows all hits for this peptide

Prot	Score	Delta	Hit	Protein	Peptide	ries matched:
beta	76.6	-0.04	1	KBBOA2	FOSEEQQTDELDQK	1
Q95L	22.5	-0.04	1	KBBOA2	FOSEEQQTDELDQK	1
S672	18.1	-0.19	2	G84453	CLSLSRQVDLFEETIEK	1
AAA3	17.9	-0.21	3	Q7QU69	LYLDIKLIYPEETDEK	1
BOVC	13.7	-0.17	4	Q9VPUS	QLASGEYFLNQEORQAK	1
BOVC	10.7	-0.20	5	O69721	YDSFFRSDIVTVVIGADK	1
AAA3	9.9	-0.23	6	Q7VZ27	TSPDATVVVSFPGGDEARTR	1
BOVC	9.9	-0.17	7	Q7RLU6	NMELILDENDDKRELADGK	1
A329	9.6	-0.12	8	AB1213	KTLTFFGVDPEDENNQK	1
beta	9.1	-0.20	9	T45584	FQISCSVEGVLFVLEPK	1

1:KBBOA2

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, there is no ambiguity, even if the spectrum is not the greatest.



Monoisotopic mass of neutral peptide Mr(calc): 2060.82
 Variable modifications:
 S3 : Phospho (ST)
 Ions Score: 77 Expect: 4e-05
 Matches (Bold Red): 33/172 fragment ions using 65 most intense peaks

#	b	b ⁺⁺	b ⁺	b ⁺⁺⁺	b ⁰	b ⁰⁺⁺	Seq	y	y ⁺⁺	y ⁺	y ⁺⁺⁺	y ⁰	y ⁰⁺⁺	#
1	148.08	74.54					F							16
2	276.13	138.57	259.11	130.06			Q	1816.78	908.90	1799.76	900.38	1798.77	899.89	15
3	345.16	173.08	328.13	164.57	327.15	164.08	S	1688.72	844.87	1671.70	836.35	1670.71	835.86	14
4	474.20	237.60	457.17	229.09	456.19	228.60	E	1619.70	810.36	1602.68	801.84	1601.69	801.35	13
5	603.24	302.12	586.21	293.61	585.23	293.12	E	1490.66	745.83	1473.63	737.32	1472.65	736.83	12
6	731.30	366.15	714.27	357.64	713.29	357.15	Q	1361.62	681.31	1344.59	672.80	1343.61	672.31	11
7	859.36	430.18	842.33	421.67	841.35	421.18	Q	1233.56	617.28	1216.53	608.77	1215.55	608.28	10
8	987.42	494.21	970.39	485.70	969.41	485.21	Q	1105.50	553.25	1088.47	544.74	1087.49	544.25	9
9	1088.46	544.74	1071.44	536.22	1070.45	535.73	T	977.44	489.22	960.42	480.71	959.43	480.22	8
10	1217.51	609.26	1200.48	600.74	1199.50	600.25	E	876.39	438.70	859.37	430.19	858.38	429.70	7
11	1332.53	666.77	1315.51	658.26	1314.52	657.77	D	747.35	374.18	730.33	365.67	729.34	365.17	6
12	1461.58	731.29	1444.55	722.78	1443.57	722.29	E	632.32	316.67	615.30	308.15	614.31	307.66	5
13	1574.66	787.83	1557.63	779.32	1556.65	778.83	L	503.28	252.14	486.26	243.63	485.27	243.14	4
14	1702.72	851.86	1685.69	843.35	1684.71	842.86	Q	390.20	195.60	373.17	187.09	372.19	186.60	3
15	1817.75	909.38	1800.72	900.86	1799.74	900.37	D	262.14	131.57	245.11	123.06	244.13	122.57	2
16							K	147.11	74.06	130.09	65.55			1

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

Peptide Summary Report (../data/20040511/FamuCaEn.dat) - Microsoft Internet Explorer

Address: http://www.matrixscience.com/cgi/master_results.pl?file=../data/20040511/FamuCaEn.dat

Standard scoring MudPIT scoring Ions score cut-off: 0 Show sub-sets
 Show pop-ups Suppress pop-ups Sort unassigned: Decreasing Score Require bold red

Select All Select None Search Selected Error tolerant

1. [P02662-00-00-00](#) Mass: 24570 Score: 102 Queries matched: 1
 (CAS1_BOVIN) Splice isoform Displayed; Variant Displayed; Conflict Displayed; from P02662 Alpha-S1 casein precursor
 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"/>	1660.80	1659.79	1659.79	0.01	0	102	7.1e-08	1	VPQLEIVPNSAEER + Phospho (ST)

Top scoring peptide matches to query 1
 Score greater than 43 indicates identity
 Status bar shows all hits for this peptide

Protein	Score	Delta	Hit	Protein	Peptide	Matched
(CAS1)	102.3	0.01	1	P02662-00-00-00	VPQLEIVPNSAEER	Conflict Displayed; from P02662 Alpha-S1 casein precursor
(CAS1)	32.3	-0.03	2	Q05469	NETGLSVAASSLFTSGR	Conflict Displayed; from P02662 Alpha-S1 casein precursor
(CAS1)	28.9	0.04	3	P29475-03-00-00	EIEPVLISLLTSQGR	Conflict Displayed; from P02662 Alpha-S1 casein precursor
(CAS1)	21.7	0.04	3	P29475-03-00-00	EIEPVLISLLTSQGR	Conflict Displayed; from P02662 Alpha-S1 casein precursor
(LIP3)	20.2	-0.07	4	Q86UR5-01-00-00	VFKTSAQVVEGAVEER	Conflict Displayed; from P02662 Alpha-S1 casein precursor
(LIP3)	20.0	0.05	5	Q9XT50-00-00-00	VQYNPSLVSPGALR	Conflict Displayed; from P02662 Alpha-S1 casein precursor
(LIP3)	19.3	-0.03	2	Q05469	NETGLSVAASSLFTSGR	Conflict Displayed; from P02662 Alpha-S1 casein precursor
(LIP3)	16.5	-0.06	6	Q96P48-00-00-00	RSVAAFIADPLLLR	Conflict Displayed; from P02662 Alpha-S1 casein precursor
(LIP3)	16.5	-0.06	6	Q96P48-00-00-00	RSVAAFIADPLLLR	Conflict Displayed; from P02662 Alpha-S1 casein precursor
(LIP3)	16.5	-0.06	6	Q96P48-00-00-00	RSVAAFIADPLLLR	Conflict Displayed; from P02662 Alpha-S1 casein precursor

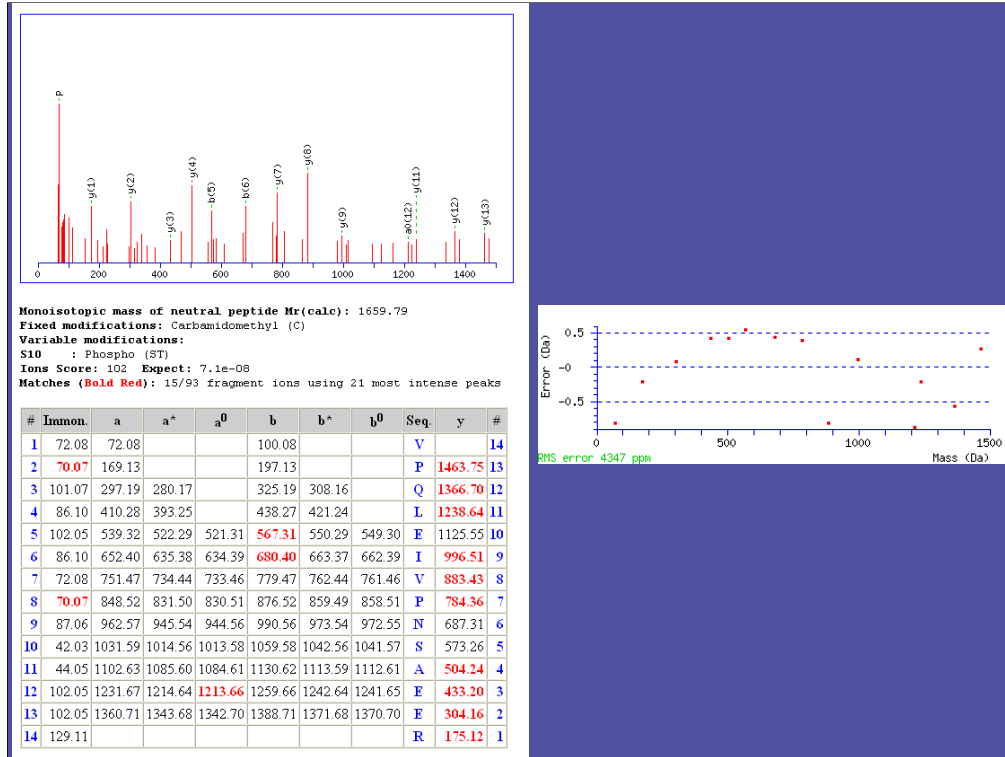
2. [Q05469](#) Mass: 43868 Score: 26 Queries matched: 1
 (LIP3) none sensitive lipase (EC 3.1.1.-) (HSL)
 Check

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
<input type="checkbox"/>	1660.80	1659.79	1659.82	-0.03	0	32	0.72	2	NETGLSVAASSLFTSGR + Phospho+PL (S); Phospho+PL (T)

3. [P29475-03-00-00](#) Mass: 43868 Score: 26 Queries matched: 1

1: P02662-00-00-00

Also, when there is only one site, can get very clear cut matches. Here for example, is a peptide from alpha casein. There is only one potential site and look at the difference in score.



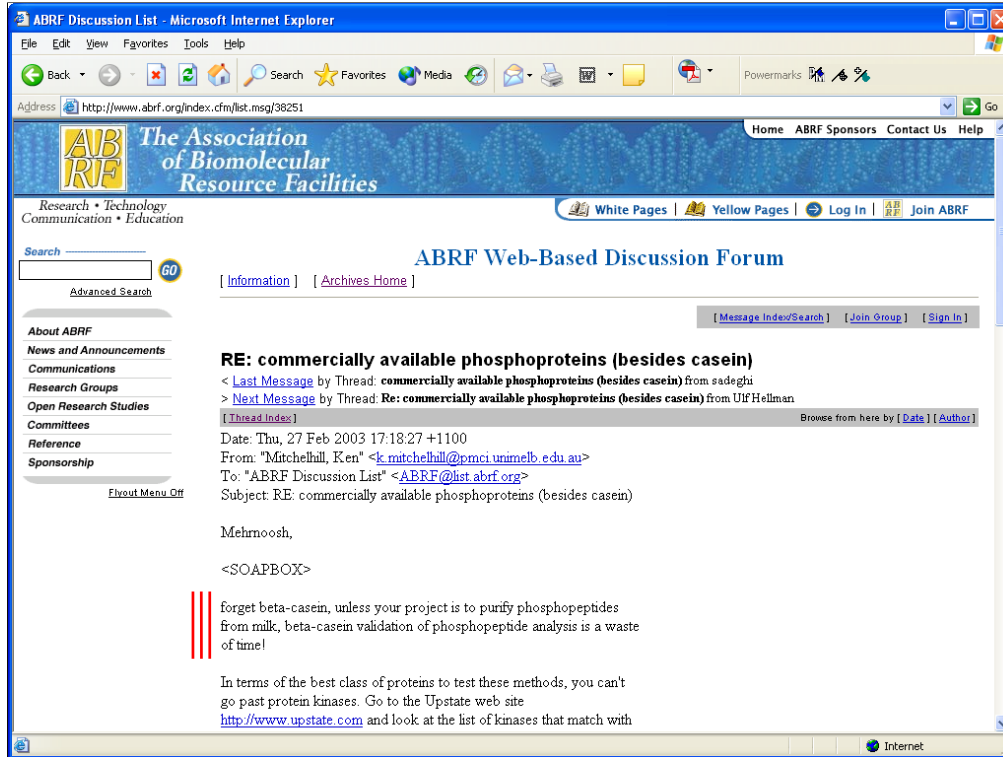
What may surprise you is that the spectrum doesn't look as pretty as the last one. This is MALDI PSD data. In this particular case, the mass accuracy is not so good, +/- 0.5 Da, but Mascot still gets a great match because the sequence coverage is good.

However, phosphopeptides from caseins are notoriously easy ...



If we go to the ABRF archives, we'll find comments like these from experts in the field:

“Casein will give you a false sense of success with any ms method. It is a low molecular weight , highly phosphorylated protein that is available in gram quantities from any supplier. Basically all the criteria that you will not find from a phosphoprotein derived from in vitro and/or in vivo sources.” - Nick Morrice



Or: “forget beta-casein, unless your project is to purify phosphopeptides from milk, beta-casein validation of phosphopeptide analysis is a waste of time!” - Ken Mitchellhill

So, no more phosphopeptides from casein in this presentation!

Peptide Summary Report (../data/20040513/FamTlbae.dat) - Microsoft Internet Explorer

Address: http://www.matrixscience.com/cgi/master_results.pl?file=../data/20040513/FamTlbae.dat

1. [gi|89281](#) Mass: 39323 Score: 79 Queries matched: 1
 protein kinase (EC 2.7.1.37), cAMP-dependent, type alpha, catalytic chain - pig (fragment)
 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"/>	1107.90	2213.79	2214.07	-0.28	0	80	2.7e-05	1	TWFLCGTPEYLAPEIILSK + Phospho (ST)

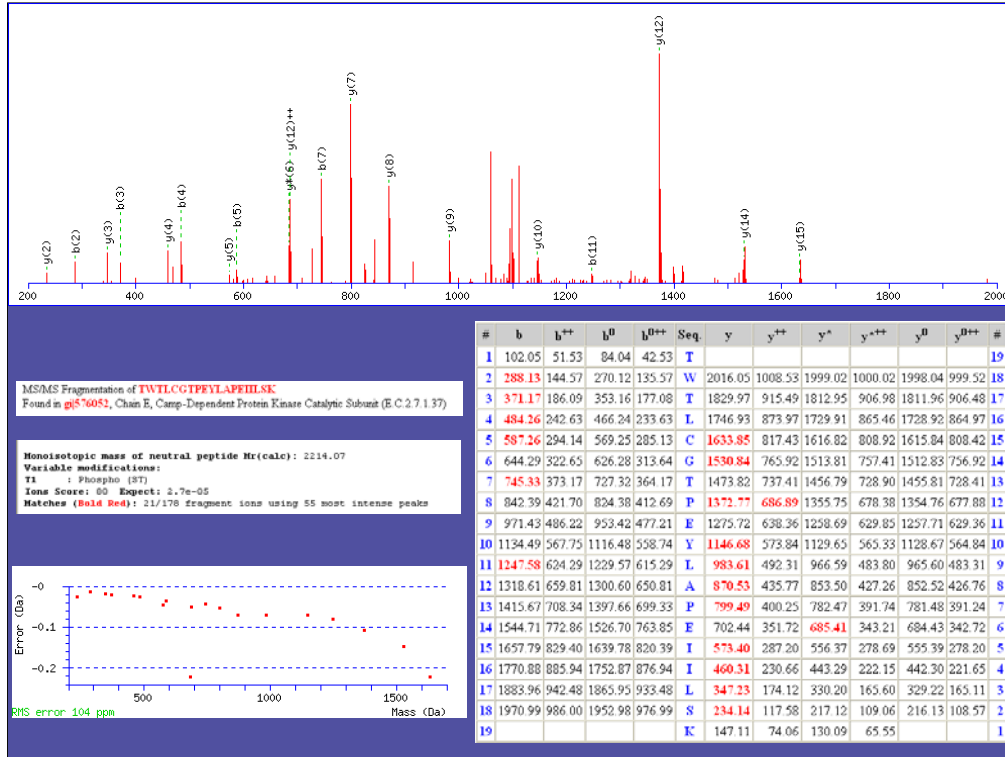
Top scoring peptide matches to query 1
 Score greater than 30 indicates homology
 Score greater than 47 indicates identity
 Status bar shows all hits for this peptide

Score	Delta	Hit	Protein	Peptide
80.4	-0.28	1	gi 89281	TWFLCGTPEYLAPEIILSK
76.9	-0.28	1	gi 89281	TWFLCGTPEYLAPEIILSK
37.7	-0.28	1	gi 89281	TWFLCGTPEYLAPEIILSK
18.0	-0.28	1	gi 89281	TWFLCGTPEYLAPEIILSK
12.9	0.74	2	gi 31224299	TWFLCGTPEYLAPEIIQAK
12.9	0.74	2	gi 31224299	TWFLCGTPEYLAPEIIQAK
10.1	0.71	4	gi 30689682	ALSTRPDLPSIYCOELSK
9.5	-0.16	3	gi 7021083	TATVSSPLTSPSPSTLSLK
7.7	-0.06	5	gi 13470865	STYGTGCFALLNTGSDLVR
7.7	-0.06	5	gi 13470865	STYGTGCFALLNTGSDLVR

cAMP-dependent protein kinase catalytic subunit
[gi|230462](#) Mass: 40414 Score: 79 Queries matched: 1
 Chain E, c-AMP-Dependent Protein Kinase (E.C.2.7.1.37) (cAPK) (Catalytic Subunit)
[gi|284054](#) Mass: 23939 Score: 79 Queries matched: 1
 protein kinase (EC 2.7.1.37), cAMP-dependent, alpha catalytic chain, short splice form - human (fragment)
[gi|349816](#) Mass: 40398 Score: 79 Queries matched: 1
 Chain E, c-AMP-Dependent Protein Kinase (E.C.2.7.1.37) (cAPK) (Catalytic Subunit) "alpha" Isoenzyme Mutant With Ser
[gi|476590](#) Mass: 41142 Score: 79 Queries matched: 1

Here is a more realistic example - a strong match to a phosphopeptide from cAMP-dependent protein kinase.

This is very typical and reproducible ... there is little to choose in terms of score between having the phosphate on T1 or T3. But, we can be very confident it is not on T7 or Y10 because the score drops dramatically



If we compare the two top hits carefully, we can see that there is just one peak difference.

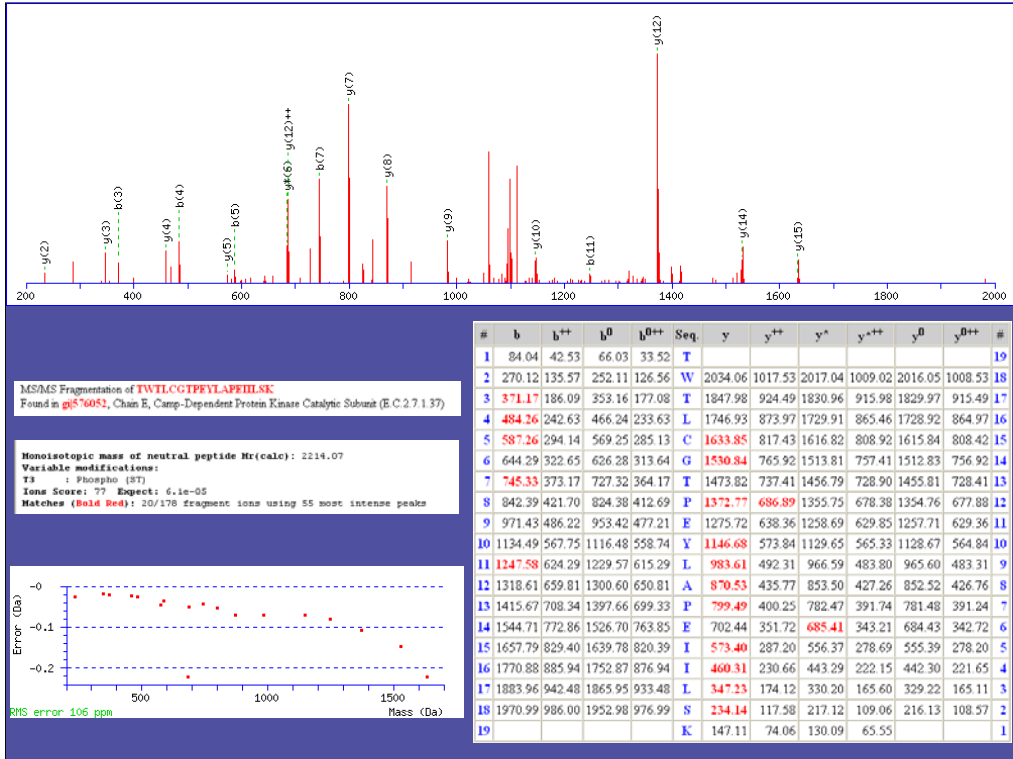
Unfortunately, in this case, the scoring comes mainly from the y ions, so this additional peak has little effect on the score

This highlights the difference between finding a match and verifying it. Mascot is doing a good job of pulling the correct peptide out of all the molecular weight matches in the database.

And, if we accept that this is the correct peptide sequence, and look for evidence to prefer one phosphorylation site over the other, then we immediately focus on this one peak.

However, there are many unmatched peaks in the spectrum. Also, there is always the possibility of a random match. For example, looking at the mass errors, we might guess that y*(6) is wrong.

If we want to follow through this particular example, we go to the bottom of peptide view



Mascot Search Results: Peptide View - Microsoft Internet Explorer

Address: http://www.matrixscience.com/cgi/peptide_view.pl?file=.../data/20040517/FanroaYn.dat&query=1&hit=1&index=g%7c89281&px=1

8	842.39	421.70	824.38	412.69	P	1372.77	686.89	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	1146.68	573.84	1129.65	565.33	1128.67	564.84	10
11	1247.58	624.29	1229.57	615.29	L	983.61	492.31	966.59	483.80	965.60	483.31	9
12	1318.61	659.81	1300.60	650.81	A	870.53	435.77	853.50	427.26	852.52	426.76	8
13	1415.67	708.34	1397.66	699.33	P	799.49	400.25	782.47	391.74	781.48	391.24	7
14	1544.71	772.86	1526.70	763.85	E	702.44	351.72	685.41	343.21	684.43	342.72	6
15	1657.79	829.40	1639.78	820.39	I	573.40	287.20	556.37	278.69	555.39	278.20	5
16	1770.88	885.94	1752.87	876.94	I	460.31	230.66	443.29	222.15	442.30	221.65	4
17	1883.96	942.48	1865.95	933.48	L	347.23	174.12	330.20	165.60	329.22	165.11	3
18	1970.99	986.00	1952.98	976.99	S	234.14	117.58	217.12	109.06	216.13	108.57	2
19					K	147.11	74.06	130.09	65.55			1

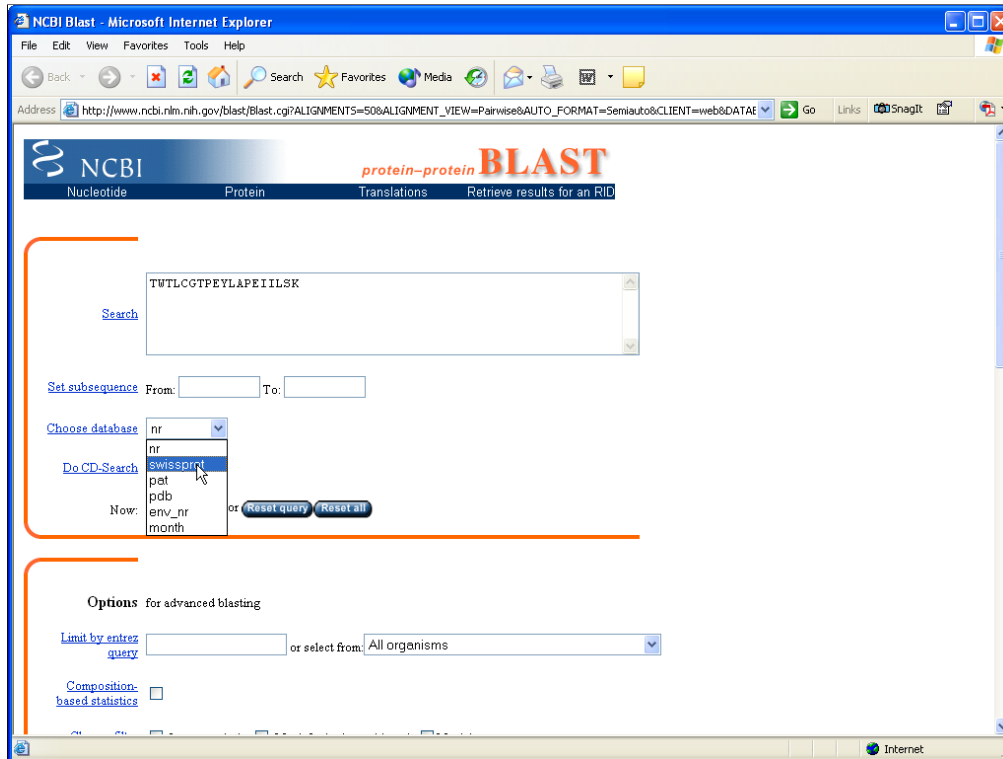
RMS error: 104 ppm

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

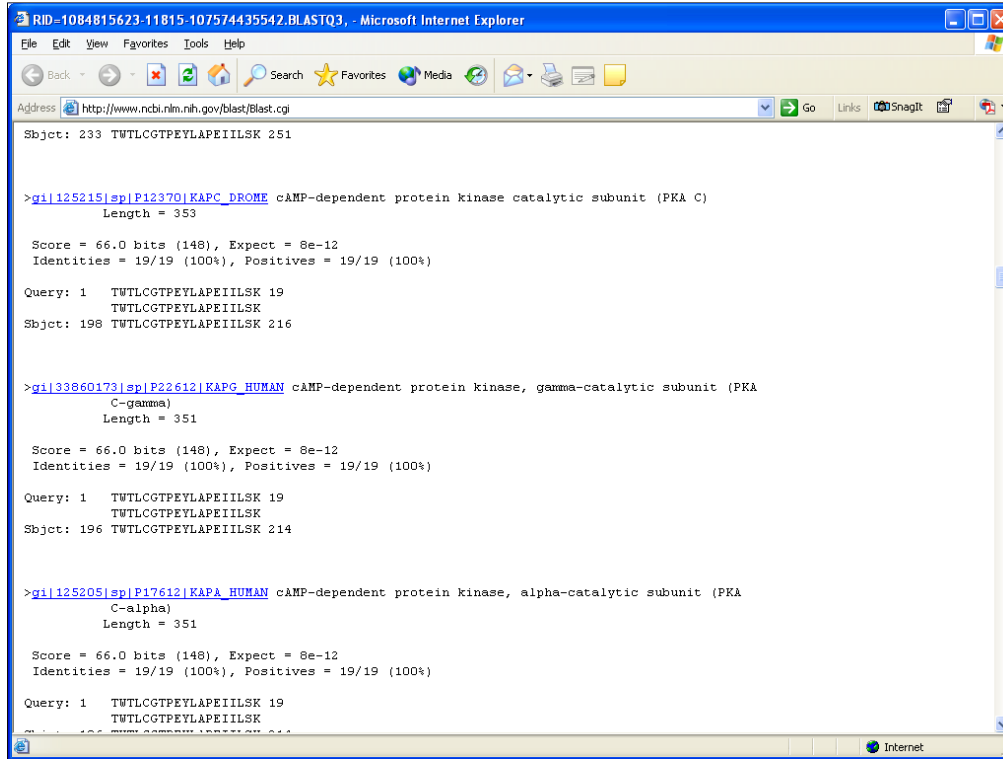
Mascot: <http://www.matrixscience.com/>

Address: http://www.ncbi.nlm.nih.gov/blast/Blast.cgi?ALIGNMENTS=50&ALIGNMENT_VIEW=Pairwise&AUTO_FORMAT=Semiauto&CLIENT=web&DATABASE=nr&D

Follow the link to run a Blast search of this peptide



Choose Swiss-Prot as the database, because we want a database with good annotations



Lots of identity matches to this very common protein. Choose the relevant species, in this case human. Make a mental note of the offset of the peptide in the protein, 196

Features

Key	From	To	Length	Description
INIT_MET	0	0		
LIPID	1	1		N-myristoyl glycine (By similarity).
DOMAIN	43	297	255	Protein kinase.
NP_BIND	49	57	9	ATP (By similarity).
BINDING	72	72		ATP (By similarity).
ACT_SITE	166	166		Proton acceptor (By similarity).
MOD_RES	197	197		Phosphothreonine (by autocatalysis) (By similarity).
MOD_RES	338	338		Phosphoserine (by autocatalysis) (By similarity).
CONFLICT	267	267		H -> D (in Ref. 1).
CONFLICT	344	344		A -> P (in Ref. 2).

Sequence information

Length: 350 AA Molecular weight: 40303 Da CRC64: 4CA40198369B8D3B [This is a checksum on the sequence]

10	20	30	40	50	60
GNAPAKKDE	QEEVNIFLA	KARGDFLYRW	GNPAQNTASS	DQFERLRTLQ	MGSFGRVMLV
70	80	90	100	110	120
RHQETGGHYA	MKILNKQKVV	KNKQVEHILN	EKRILQAIQF	PFLVKLQFSF	KDNSYLVLVM
130	140	150	160	170	180
EYVPGGEMFS	RLQRVGRFSE	PHACFYAAQV	VLAVQYLHSL	DLIHRDLRPE	NLLIDQQGYL
190	200	210	220	230	240
QVIDFGFAKR	VKGR	TWTLCG	TPEYLAPEII	LSR	GYNKAVD
					WVALGVLIVE
					MAVGFPFFFYA

And hop over to ExPASy to see the full text for this entry. Here is the peptide. The ExPASy numbering has it starting at 195 because the initiator methionine is not present.

But, what is this? According to Swissprot, the phosphate is on T3, not T1! So, either Swiss-Prot is wrong or the extra match in the b series, which looked so convincing is spurious. I've no idea which. But, this does illustrate how easy it is to over-interpret noisy MS data.

Peptide Summary Report (../data/20040513/FamTlbae.dat) - Microsoft Internet Explorer

Address: http://www.matrixscience.com/cgi/master_results.pl?file=../data/20040513/FamTlbae.dat

1. [gi|89281](#) Mass: 39323 Score: 79 Queries matched: 1
 protein kinase (EC 2.7.1.37), cAMP-dependent, type alpha, catalytic chain - pig (fragment)

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"/>	1107.90	2213.79	2214.07	-0.28	0	80	2.7e-05	1	TWTLGGTPEYLAPEIILSK + Phospho (ST)

Top scoring peptide matches to query 1
 Score greater than 30 indicates homology
 Score greater than 47 indicates identity
 Status bar shows all hits for this peptide

Score	Delta	Hit	Protein	Peptide
80.4	-0.28	1	gi 89281	TWTLGGTPEYLAPEIILSK
76.9	-0.28	1	gi 89281	TWTLGGTPEYLAPEIILSK
37.7	-0.28	1	gi 89281	TWTLGGTPEYLAPEIILSK
18.0	-0.28	1	gi 89281	TWTLGGTPEYLAPEIILSK
12.9	0.74	2	gi 31224299	TWTLGGTPEYLAPEIIQAK
12.9	0.74	2	gi 31224299	TWTLGGTPEYLAPEIIQAK
10.1	0.71	4	gi 30689682	ALSTRPDLPSIYQQLSK
9.5	-0.16	3	gi 7021083	TATVSSPLTSPSPSTLSLK
7.7	-0.06	5	gi 13470865	STYGTGCFALLNTGSDLVR
7.7	-0.06	5	gi 13470865	STYGTGCFALLNTGSDLVR

cAMP-dependent protein kinase catalytic subunit

[gi|230462](#) Mass: 40414 Score: 79 Queries matched: 1
 Chain E, c-AMP-Dependent Protein Kinase (E.C.2.7.1.37) (cAPK) (Catalytic Subunit)

[gi|284054](#) Mass: 23939 Score: 79 Queries matched: 1
 protein kinase (EC 2.7.1.37), cAMP-dependent, alpha catalytic chain, short splice form - human (fragment)

[gi|349816](#) Mass: 40398 Score: 79 Queries matched: 1
 Chain E, c-AMP-Dependent Protein Kinase (E.C.2.7.1.37) (cAPK) (Catalytic Subunit) "alpha" Isoenzyme Mutant With Ser

[gi|476590](#) Mass: 41142 Score: 79 Queries matched: 1

The Mascot score reminds us that there is little to choose between T1 and T3. All we can say with confidence is that the phosphate is on one or the other ...or maybe there is a mixture of both forms?

Peptide Summary Report (cps l-rat) - Microsoft Internet Explorer

Address: http://www.matrixscience.com/cgi/master_results.pl?file=../data/20040511/FamulaSe.dat

Significance threshold p < 0.05 Max. number of hits AUTO

Standard scoring MudPIT scoring Ions score cut-off 0 Show sub-sets

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

Select All Select None Search Selected Error tolerant

1. [P27703](#) Mass: 41249 Score: 64 Queries matched: 1
 (MK01_MOUSE) Mitogen-activated protein kinase 1 (EC 2.7.1.37) (Extracellular signal-regulated kinase 2) (ERK-2) (Mit

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"/>	742.26	2223.77	2222.96	0.80	0	68	3.4e-05	1	VADPDHDTGFLTEYVATR + Phospho (Y)

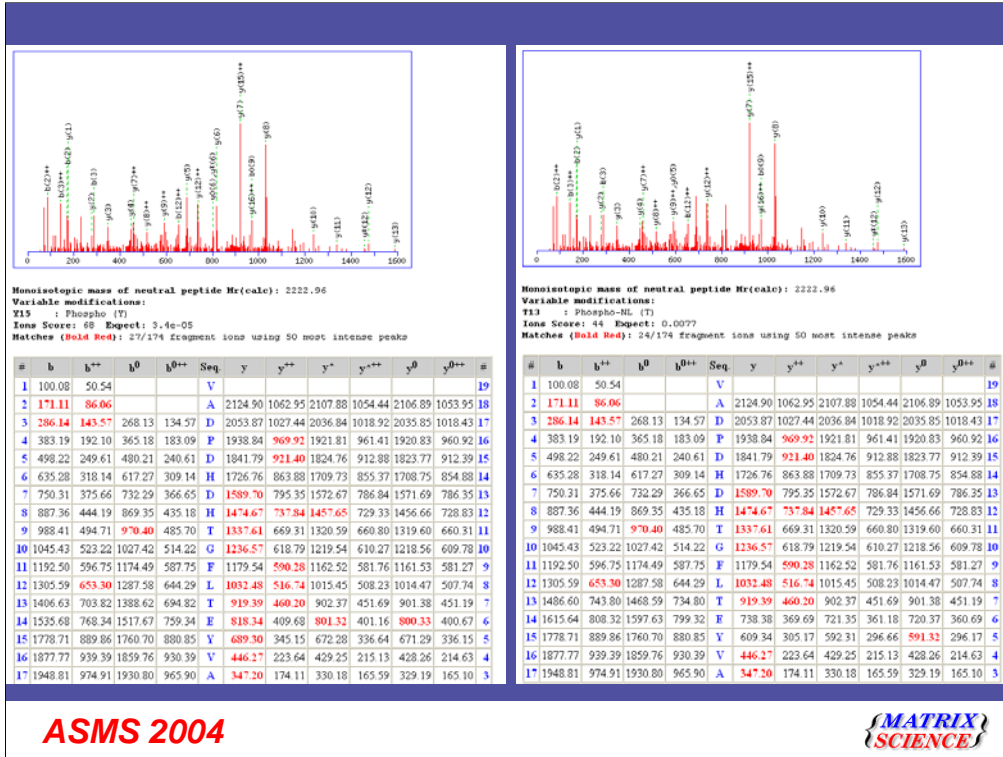
Top scoring peptide matches to query 1
 Score greater than 36 indicates identity
 Status bar shows all hits for this peptide

Search Parameters

Type of search	Score	Delta	Hit	Protein	Peptide
Enzyme	67.9	0.80	1	P27703	VADPDHDTGFLTEYVATR
Variable modification	44.4	0.80	1	P27703	VADPDHDTGFLTEYVATR
Mass values	38.5	0.80	1	P27703	VADPDHDTGFLTEYVATR
Protein Mass	19.8	0.80	1	P27703	VADPDHDTGFLTEYVATR
Peptide Mass	13.5	0.80	1	P27703	VADPDHDTGFLTEYVATR
Fragment Mass	12.6	-0.97			SDYENDDECWSALESFR
Max Missed	12.6	-0.97			SDYENDDECWSALESFR
Instrument	11.4	0.80	1	P27703	VADPDHDTGFLTEYVATR
Number of peptides	10.5	0.80	1	P27703	VADPDHDTGFLTEYVATR
	7.6	0.74			VINFYAGANQSMNVTCVGKK

1:P27703

Here is an example where a similar difference, moving the site by 2 residues, causes a much bigger change in the score

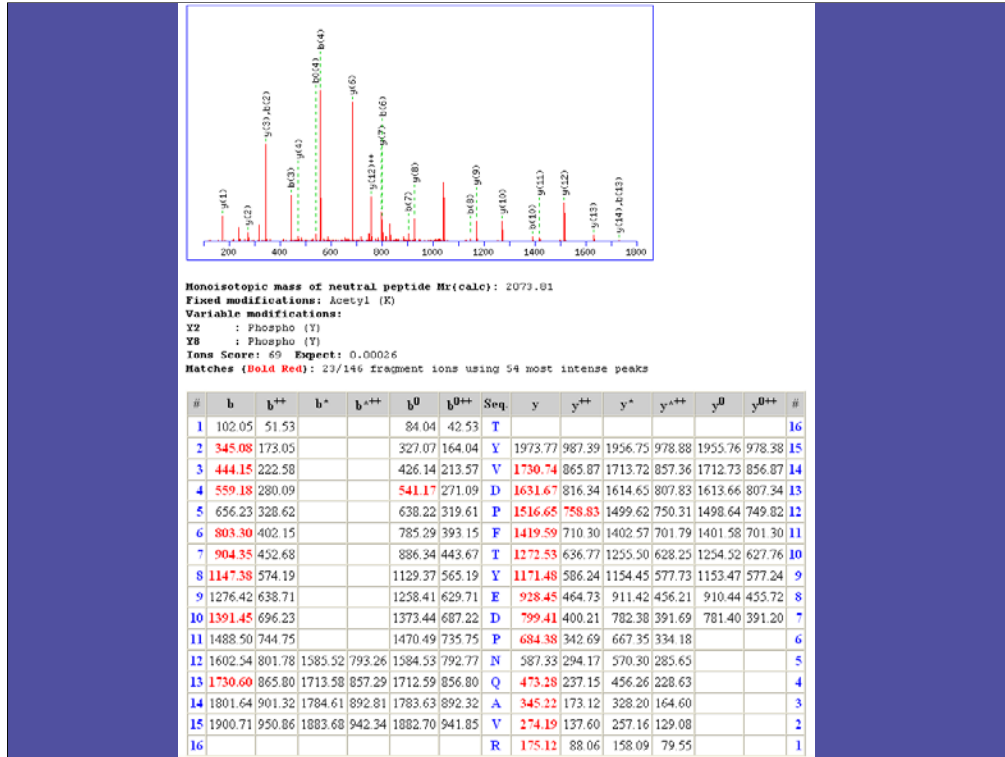


ASMS 2004



This is because there are two additional peaks, rather than one, and they show up in the y series, which is where the score is coming from.

Additional evidence that the phosphate is on Y15 comes from the lack of a strong neutral loss peak for the molecular ion. However, this does not contribute to the Mascot score.



A slightly more complicated example. Two phosphates and three potential sites.

Peptide Summary Report (.../data/20040513/FamTleET.dat) - Microsoft Internet Explorer

Address: http://www.matrixscience.com/cgi/master_results.pl?file=.../data/20040513/FamTleET.dat

Standard scoring MudPIT scoring Ions score cut-off: 0 Show sub-sets
 Show pop-ups Suppress pop-ups Sort unassigned: Decreasing Score Require bold red

Select All Select None Search Selected Error tolerant

1. [q1|4758280](#) Mass: 111974 Score: 67 Queries matched: 1
 ephrin receptor EphA4; ephrin type-A receptor 4; TYRO1 protein tyrosine kinase; tyrosine-protein kinase receptor SEK;
 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"/>	1038.00	2073.98	2073.81	0.17	0	69	0.00026	1	TYVDPFTYEDPNQAVR + 2 Phospho (Y)

Top scoring peptide matches to query 1
 Score greater than 45 indicates identity
 Status bar shows all hits for this peptide

Score	Delta	Hit	Protein	Peptide
68.7	0.17	1	g1 4758280	TYVDPFTYEDPNQAVR
60.6	0.17	1	g1 4758280	TYVDPFTYEDPNQAVR
27.5	0.17	1	g1 4758280	TYVDPFTYEDPNQAVR
14.5	0.17	1	g1 4758280	TYVDPFTYEDPNQAVR
14.5	0.17	1	g1 4758280	TYVDPFTYEDPNQAVR
10.9	0.17	1	g1 4758280	TYVDPFTYEDPNQAVR
7.4	0.07	3	g1 4758794	YTLTEPLLHTAQEAAR

2. [q1|21645679](#) Mass: 111974 Score: 67 Queries matched: 1
 QEDSYITTESLTTAVR + Phospho (ST); Phospho (Y)
 Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
1	1038.00	2073.98	2073.85	0.13	0	6	4.4e+02	8	QEDSYITTESLTTAVR + Phospho (ST); Phospho (Y)

1:q1|4758280

Y9 - high level of confidence
 Y2 more likely than T1

Looking at the scores, we see a big score drop when Y9 is not carrying a phosphate, so we definitely want to put one there. However, not so clear cut whether the other phosphate is on T1 or Y2

Mascot Search Results: Protein View - Microsoft Internet Explorer

Address: http://www.matrixscience.com/cgi/protein_view.pl?file=.../data/20040513/FamTitE1.dat&h=gl%7c5853748px=1&prtscore=66.80246753856978_mudpit=1000

Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Sequence Coverage: 1%

Matched peptides shown in **Bold Red**

```

1  MAGIFYF ILF  SFLFGICDAV  TGSRVYPANE  VTLLDSRSVQ  GELGWIASPL
51  EGGHEEVSIM  DEKNTPIRTY  QVCNVMEASQ  NNWLRDWDIT  REGAQRVYIE
101 IKFTLRDCNS  LPGVMGTCKE  TFNLVYYESD  NDKERFIRES  QFGKTDITAA
151 DESFTQVDIG  DRIMKLNTEI  RDVGPLSKKG  FYLAFQDVGA  CIALVSVRVF
201 YKKPLTVRN  LAQFPDITIG  ADTSSLVEVR  GSCVNNSEK  DVPKMYCGAD
251 GEMLVPIGNC  LCNAGHEEQN  GECOACKIGY  YKALSTDASC  AKCPHYSYSV
301 WEGATCTCD  RGFFRADNDA  ASMPCTRPPS  APLNLISNVN  ETSVNLWESS
351 PONTGGRDID  SYNVCCKKCG  AGDPSKCRPC  GSGVHYTPOQ  NGLKTRVSI
401 TDLLAHTNYT  FEIWAIVGVS  KYNFSPDQSV  SVTVTNGQAA  PSSIALLVOAK
451 EYTRYSVALA  WLEPDRPNGV  ILEYEVKYYE  KDQNSRYRI  VRTAARNTDI
501 KGLNPLTSYV  FHVRRATAAG  YGDFSEPLEV  TINTVPSRLI  GDGANSTVLL
551 VSVSGSVLVV  VILIAAFVIS  RRRSKYSKAK  QEADEEKHLN  QGVRTYVDFP
601 TYEDDPNQAVR  EFAKEIDASC  IKIEKVIGVG  EFGVCSGRL  KVPKREICV
651 AIKTLKAGYT  DKQRDFLSE  ASIMGQFDHP  NIIHLEGVVT  KCKPVMIIIE
701 YMENSLDAF  LRKRDGRFTV  IQLVGMLRGI  GSGHKYLSDH  SYVHRDLAAR
751 NILVNSNLVC  KVSDFGMSRV  LEDDPEAAYT  TRGGKIPIRM  TAFEALAYRK
801 FTSASDVMSY  GIVMWEVMSY  GERPYWMSN  QDVKAIEEG  YRLPPPMDCP
851 IALHQLMLDC  WQKERSDRPK  FGGIVNMLDK  LIRNPNSLKR  TGSSESRPNT
901 ALLDPSSEPF  SAVVSVGDWL  QAIKMDRYED  NFTAAGYITL  EAVVHMSQDD
951 LARIGITAIT  HQNKILSSVQ  AMRTQMOMH  GRMVPV

```

Y602 - high level of confidence
Y596 more likely than T595

Show predicted peptides also

Sort Peptides By: Residue Number Increasing Mass Decreasing Mass

Start	End	Observed	Mr(expt)	Mr(calcd)	Delta	Miss	Sequence
595	610	1038.00	2073.98	2073.61	0.17	0	TYVDFP TYEDDPNQAVR 2 Phospho (Y) (Ions score 69)

Error (Da): 0.18, 0.17

The numbering in the protein is T595 and Y596. Again, if we hop over to Expsy

NiceProt View of Swiss-Prot: Q03137 - Microsoft Internet Explorer

Address: http://ca.expasy.org/cgi-bin/niceprot.plPEPA4_MOUSE

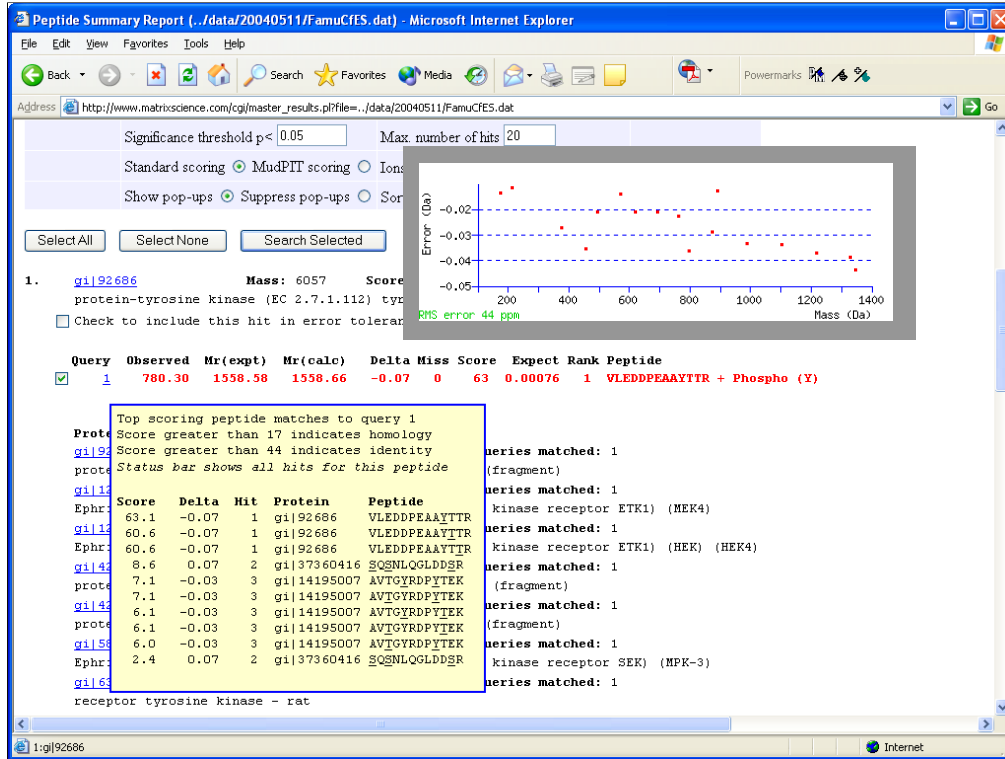
CHAIN 20 986 967 Ephrin type-A receptor 4.
 DOMAIN 20 547 528 Extracellular (Potential).
 TRANSMEM 548 569 22 Potential.
 DOMAIN 570 986 417 Cytoplasmic (Potential).
 DOMAIN 191 325 135 Cys-rich.
 DOMAIN 328 431 104 Fibronectin type-III 1.
 DOMAIN 441 532 92 Fibronectin type-III 2.
 DOMAIN 621 862 262 Protein kinase.
 DOMAIN 911 975 65 SAM.
 SITE 984 986 3 PDZ-binding motif (Potential).
 NP_BIND 627 635 9 ATP (By similarity).
 BINDING 653 653 ATP (By similarity).
 ACT_SITE 746 746 By similarity.
 MOD_RES 596 596 Phosphotyrosine (by autocatalysis).
 MOD_RES 602 602 Phosphotyrosine (by autocatalysis).
 MOD_RES 779 779 Phosphotyrosine (by autocatalysis) (Potential).
 MOD_RES 928 928 Phosphotyrosine (by autocatalysis) (Potential).
 CARBOHYD 235 235 N-linked (GlcNAc...) (Potential).
 CARBOHYD 340 340 N-linked (GlcNAc...) (Potential).
 CARBOHYD 408 408 N-linked (GlcNAc...) (Potential).
 CARBOHYD 423 423 N-linked (GlcNAc...) (Potential).
 VARSPLIC 783 832 Missing (in isoform Short). VSP_002998

Sequence information

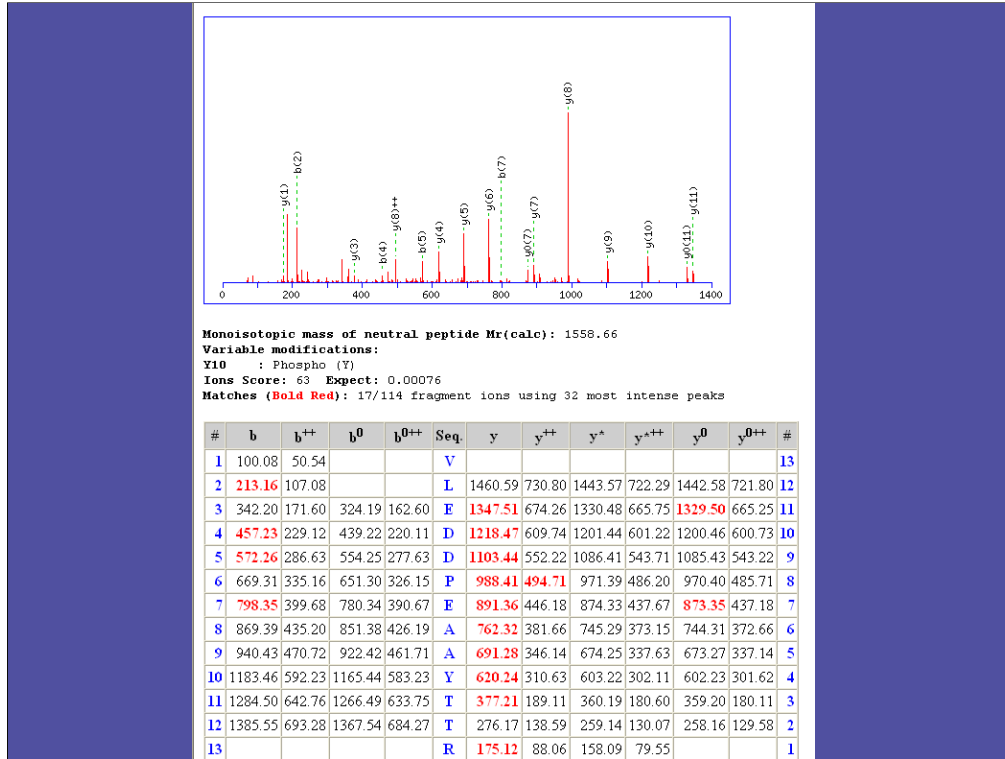
Length: 986 AA [This is the length of the unprocessed precursor] Molecular weight: 109801 Da [This is the MW of the unprocessed precursor] CRC64: D16AD8B85668C80E [This is a checksum on the sequence]

10	20	30	40	50	60
MAGIFYPILF	SFLFGICDAV	TGSRVYPANE	VTLIDRSVQ	GELGWIASPL	EGGWEEVSIM
70	80	90	100	110	120
DEKNTPTRY	OVCMVMEASQ	NNMLRTDMIT	REGAORVYIE	IKFTLRDCNS	LPGVMTCKE

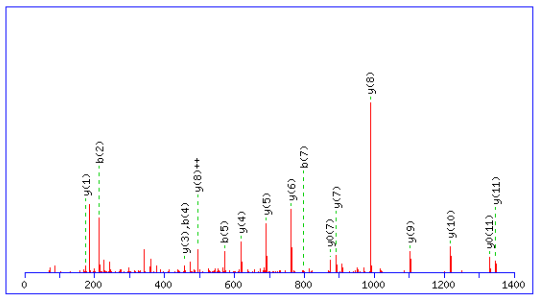
We find that, this time, we agree. Again, the lack of a neutral loss peak from the molecular ion suggests Y rather than T. However, the Mascot score difference is less than 10, which can come down to just a single extra peak being matched. In the absence of additional evidence, this would be a shaky assignment. My subjective feeling, not very scientific, but based on looking at a range of examples, is that you can be reasonably confident in an assignment when the score difference is 20 or more.



Here's a really difficult example. Great mass accuracy, good sequence coverage, but almost nothing to choose in score between the top 3 matches

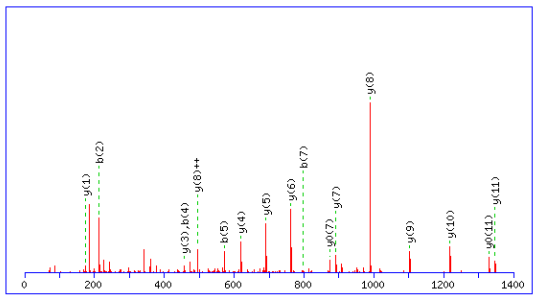


If we flick between these matches, you can see that there is no evidence in the score to support preferring one of these arrangements over the other two. The absence of a neutral loss peak from the molecular ion would indicate a preference for Y10



Monoisotopic mass of neutral peptide Mr(calc): 1558.66
 Variable modifications:
 T11 : Phospho-NL (T)
 Ions Score: 61 Expect: 0.0013
 Hatches (Bold Red): 17/114 fragment ions using 32 most intense peaks

#	b	b ⁺⁺	b ⁰	b ⁰⁺⁺	Seq.	y	y ⁺⁺	y ⁺	y ⁺⁺⁺	y ⁰	y ⁰⁺⁺	#
1	100.08	50.54			V							13
2	213.16	107.08			L	1460.59	730.80	1443.57	722.29	1442.58	721.80	12
3	342.20	171.60	324.19	162.60	E	1347.51	674.26	1330.48	665.75	1329.50	665.25	11
4	457.23	229.12	439.22	220.11	D	1218.47	609.74	1201.44	601.22	1200.46	600.73	10
5	572.26	286.63	554.25	277.63	D	1103.44	552.22	1086.41	543.71	1085.43	543.22	9
6	669.31	335.16	651.30	326.15	P	988.41	494.71	971.39	486.20	970.40	485.71	8
7	798.35	399.68	780.34	390.67	E	891.36	446.18	874.33	437.67	873.35	437.18	7
8	869.39	435.20	851.38	426.19	A	762.32	381.66	745.29	373.15	744.31	372.66	6
9	940.43	470.72	922.42	461.71	A	691.28	346.14	674.25	337.63	673.27	337.14	5
10	1103.49	552.25	1085.48	543.24	Y	620.24	310.63	603.22	302.11	602.23	301.62	4
11	1284.50	642.76	1266.49	633.75	T	457.18	229.09	440.15	220.58	439.17	220.09	3
12	1385.55	693.28	1367.54	684.27	T	276.17	138.59	259.14	130.07	258.16	129.58	2
13					R	175.12	88.06	158.09	79.55			1



Monoisotopic mass of neutral peptide Mr(calc): 1558.66
 Variable modifications:
 T12 : Phospho-NL (T)
 Ions Score: 61 Expect: 0.0013
 Hatches (Bold Red): 17/114 fragment ions using 32 most intense peaks

#	b	b ⁺⁺	b ⁰	b ⁰⁺⁺	Seq.	y	y ⁺⁺	y ⁺	y ⁺⁺⁺	y ⁰	y ⁰⁺⁺	#
1	100.08	50.54			V							13
2	213.16	107.08			L	1460.59	730.80	1443.57	722.29	1442.58	721.80	12
3	342.20	171.60	324.19	162.60	E	1347.51	674.26	1330.48	665.75	1329.50	665.25	11
4	457.23	229.12	439.22	220.11	D	1218.47	609.74	1201.44	601.22	1200.46	600.73	10
5	572.26	286.63	554.25	277.63	D	1103.44	552.22	1086.41	543.71	1085.43	543.22	9
6	669.31	335.16	651.30	326.15	P	988.41	494.71	971.39	486.20	970.40	485.71	8
7	798.35	399.68	780.34	390.67	E	891.36	446.18	874.33	437.67	873.35	437.18	7
8	869.39	435.20	851.38	426.19	A	762.32	381.66	745.29	373.15	744.31	372.66	6
9	940.43	470.72	922.42	461.71	A	691.28	346.14	674.25	337.63	673.27	337.14	5
10	1103.49	552.25	1085.48	543.24	Y	620.24	310.63	603.22	302.11	602.23	301.62	4
11	1204.54	602.77	1186.53	593.77	T	457.18	229.09	440.15	220.58	439.17	220.09	3
12	1385.55	693.28	1367.54	684.27	T	356.13	178.57	339.11	170.06	338.12	169.56	2
13					R	175.12	88.06	158.09	79.55			1

Peptide Summary Report (../data/20040514/Famerisa.dat) - Microsoft Internet Explorer

Address: http://www.matrixscience.com/cgi/master_results.pl?file=../data/20040514/Famerisa.dat

Standard scoring MudPIT scoring Ions score cut-off: 0 Show sub-sets
 Show pop-ups Suppress pop-ups Sort unassigned: Decreasing Score Require bold red

Select All Select None Search Selected Error tolerant

1. [g1|16078640](#) Mass: 71879 Score: 101 Queries matched: 1
 similar to protein kinase [Bacillus subtilis]
 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"/>	1	2825.20	2824.19	2824.19	0.01	0	109	1.6e-07	1	SSTIITHINSVLGSHVHLSPEQAR + 3 Phospho (ST)

2. [g1|2](#)
 Top scoring peptide matches to query 1
 Score greater than 31 indicates homology
 Score greater than 54 indicates identity
 Check to include this hit in error tolerant search

Query	Score	Delta	Hit	Protein	Peptide	Peptide
1	109.0	0.01	1	g1 16078640	SSTIITHINSVLGSHVHLSPEQAR	SSTIITHINSVLGSHVHLSPEQAR
2	103.0	0.01	1	g1 16078640	SSTIITHINSVLGSHVHLSPEQAR	PELFPSKARGAATGFTTLVLSAANLIV + Phospho (ST)
3	96.7	0.01	1	g1 16078640	SSTIITHINSVLGSHVHLSPEQAR	
4	92.2	0.01	1	g1 16078640	SSTIITHINSVLGSHVHLSPEQAR	
5	80.7	0.01	1	g1 16078640	SSTIITHINSVLGSHVHLSPEQAR	
6	50.4	0.01	1	g1 16078640	SSTIITHINSVLGSHVHLSPEQAR	
7	29.8	0.01	1	g1 16078640	SSTIITHINSVLGSHVHLSPEQAR	
8	11.3	0.70	2	g1 2280500	PELFPSKARGAATGFTTLVLSAANLIV	
9	11.3	0.70	2	g1 2280500	PELFPSKARGAATGFTTLVLSAANLIV	
10	10.5	-0.14	3	g1 16077729	KEIKDILAYLRVSNPDDDISFT	

arrangements of 3 / 9 = 84

How about this? Three phosphates and 9 possible sites.

This is a nice illustration of why searches with lots of modifications take a long time. For this peptide, there are 84 possible arrangements of the phosphates. And, of course, many of the other peptides that have to be tested will be even worse. Even though they don't show up in the top 10, Mascot still had to work its way through all the different arrangements.

Looking at the scores, I would feel very confident that the phosphorylation is all down towards the N terminus. You can see how the score drops when we try to push phosphate up to S10 or beyond. However, it would be a brave (or foolish) person who claimed they could pin it down more precisely.

NiceProt View of Swiss-Prot: O34507 - Microsoft Internet Explorer

Address: <http://ca.expasy.org/cgi-bin/niceprot.p?O34507>

PPRESAGE: [O34507](#)
 DIP: [O34507](#)
 ModBase: [O34507](#)
 SMR: [O34507](#), 9653AB5CFBAA7900.
 SWISS-2DPAGE: [Get region on 2D PAGE](#)
 UniRef: View cluster of proteins with at least [50%](#) / [90%](#) identity.

Keywords
 Hypothetical protein; Transferase; Serine/threonine-protein kinase; ATP-binding; Repeat; Complete proteome.

Features
[Feature table viewer](#) [Feature aligner](#)

Key	From	To	Length	Description
DOMAIN	11	271	261	Protein kinase.
DOMAIN	356	424	69	PASTA 1.
DOMAIN	425	492	68	PASTA 2.
DOMAIN	493	559	67	PASTA 3.
NP_BIND	17	25	9	ATP (By similarity).
BINDING	40	40		ATP (By similarity).
ACT_SITE	134	134		By similarity.

Sequence information
 Length: 648 AA Molecular weight: 71866 Da CRC64: 9653AB5CFBAA7900 [This is a checksum on the sequence]

10	20	30	40	50	60
MLIGKRISGR	YQILRVIGGG	GHANVYLAED	IILDREVAIK	ILRFDYANDN	EFIRRFREA
70	80	90	100	110	120
QSASSLDHPN	IVSIYDLGEE	DDIYYIVMEY	VEGHTLKEYI	TANGPLHPKE	ALNIMEQIVS
130	140	150	160	170	180

If we follow through the Blast search route and hop over to Expsy, this illustrates another reality of phosphopeptide analysis. The Swissprot entry has no information on phosphorylation for this particular protein.

We can try entering the protein sequence into NetPhos, a popular tool for predicting phosphorylation sites

NetPhos 2.0 Server - prediction results
Technical University of Denmark

648 Sequence

```

MLIGKRISGRYQILRVIGGGHMANVYLAEDIILDREVAIKILRFDYANDNEFIRRFREAQSSSLDHPNIVSIYDLGEE 80
DDIYYIVMEYVEGNTLKEYITANGPLHPKEALNIMEQIVSAIAHAHQIVHRDIKPHNILLDHMGNIRVDFGIATALS 160
STTIITHNSVLGSDVHYLSPEQARGGLATKKSDIYALGIVLFEELLTGRIPFDGESAVSIALKHLQAETPSAKRWNPSPVQS 240
VENIILKATAKDPFHRVETAEDMEADIKTAFDADRLNEKRFITQDEEMTKAIPIIKDEELAKAAGEKEAEVTTAQENKT 320
KRNKGRKRKWPVVLLTICLVFITAGILAVTVFPSLFPMPKDVKIPDVSQMEYKAAAGLLEKGLQVDSEVLEISDEKIEEGL 400
MVKTDPKADTTVKEGATVTLKSTGKARTEIGDVTGGTVDAQKALKDQGFNHVTVNEVNDERKAGTVIDQNPAGTELV 480
PSEDQVKLTVSIGPEDITLRDLKTSKEAASGYLEDNGLKLVKEAYSDDVPEGGQVVKQKPAAGTAVKPGNEVEVTFSLG 560
PEKPKAKTVKERVKIPYEPENEGDELQVQIAVDADHISISDTYEEFKIKEPTERTIELKIEPGQKGYQVMVNNKVVSYK 640
TIEYPKDE 720
.....S.....Y.....S.S..... 80
..YY.....T.....S..... 160
.....S.....T.....S..... 240
.....T.....T.....T.....T..... 320
.....S.....Y.....S..... 400
.....TT.....Y.S.....T..... 480
.S.....T.....S.....S.....YS.....T..... 560
.....T.....S.S.TY.....S..... 640
..... 720

```

Phosphorylation sites predicted: Ser: 16 Thr: 13 Tyr: 7

Serine predictions

Name	Pos	Context	Score	Pred
		V		

Unfortunately, we get a prediction that doesn't fit at all well with the MS data. The MS evidence puts the 3 phosphates close to the N terminus. It is possible that one could be at S169, but more likely not. And definitely not at S178. So, we're none the wiser in this particular case.

Conclusions

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

ASMS 2004

**MATRIX
SCIENCE**

I had hoped to present some clean statistics on how often Mascot called phosphorylation sites correctly. However, I quickly found that, in most cases, it was not possible to find independent evidence on which to base such judgements. It would be a very interesting study ... but far more work than we are able to undertake.

So, my subjective conclusions are as follows.

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.

Thank you for your attention