

Why is phosphorylation such a challenge?

- Site heterogeneity
- Poor ionisation efficiency
- 3 fragmentation channels
 - intact fragments
 - neutral loss of HPO₃ (80 Da)
 - neutral loss of H_3PO_4 (98 Da)
- Can occur at STY ~16% of residues.

ASMS 2004

{MATRIX} {SCIENCEJ

ABRF Discussion List - Micro	soft Internet Explorer	
<u>Eile E</u> dit <u>V</u> iew F <u>a</u> vorites <u>T</u> ool	s <u>H</u> elp	
🌀 Back 🝷 🐑 🔺 🛃	🏠 🔎 Search 🤺 Favorites 🜒 Media 🥝 🍰 🐼 🔹 🧾 🕏	• Powermarks 🌃 👍 %
Address 🕘 http://www.abrf.org/inde	x.cfm/list.msg/43593	🗸 🄁 @
Research • Technology Communication • Education	ssociation Tiomolecular Esource Facilities	Home ABRF Sponsors Contact Us Help 🛆
Search	ABRF Web-Based Discussion [Information] [Archives Home]	1 Forum
About ABRF	I	[Message Index/Search] [Join Group] [Sign In]
News and Announcements	PE-LC/MS	
Communications	CLast Maccare by Throad: LC/MS from can francisco sucan	
Research Groups	Next Message by Thread: BC/MS from Nick Morrice	
Open Research Studies	[Thread Index]	Browse from here by [Date] [Author]
Committees	Date: Mon. 19 May 2003 12:58:08 -0400	·· <u>···</u> ·
Recerchin	From: "Jeffrey A Kowalak, Ph.D." <ikowalak@intra.nimh.nih.gov></ikowalak@intra.nimh.nih.gov>	
Sponsoramp	To: "ABRF Discussion List" < ABRF@list.abrf.org>	
Flyout Menu Off	Subject: RE: LC/MS	
	Susan,	Search for phosphorylation,
	Anecdotally, phosphopeptides seem to have lower ionization efficiency	May 13 2004 - 656
	than non-phosphorylated peptides. Coupled with sub-stoichiometric levels	
	of phosphorylation, phosphopeptide analysis can be quite challenging. A	messages
	quick literature search will reveal that there are almost as many papers	
	describing methods of phosphopeptide characterization as there are	
	papers describing the biological importance of phosphorylation.	
	Please see the following reference as an example of one of the best examples - "A multidimensional electrospray MS-based approach to	.
é		Internet

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

ຢ່າມ	niMod: Basic Data View - Microsoft Internet Explorer								
Eile	Edit View Favorites Iools Help					<i></i>			
G	Back 🔹 🕥 - 💌 😰 🏠 🔎 Search 🛧 Favorites 🜒 Media 🧭	🔗 - 🍓 🗖 🗖) 🔁	Power	marks 🕅 🔏 🌂				
Addre	ss 🕘 http://www.unimod.org/cgi/unimod.cgi?records_per_page=25&columns_to_view=full_n	ame&columns_to_view=co	de_name&columns	s_to_view=ma	ono_mass&columns_to_view=	avge 💙 🛃 Go			
	UNIMOD protein modifications for mass spectrometry								
	View All Records Add Record		<u>Help</u>	Home Op	tions Advanced Search	Logon			
			Search	۱ <u> </u>	Go				
	Result Set				· ···				
	Modification Phosphorylation	Short name Phospho	79.966331	Average 79.9799	Composition H O(3) P	Details			
	Phosphorylation without neutral loss	Phospho-NL	79.966331	79.9799	H O(3) P	ŏ			
	phosphorylation to amine thiol	ser_thr_DAET	87.050655	87.1866	H(9) C(4) N O(-1) S	õ l			
	thioacylation of primary amines (N-term and Lys)	DSP	87.998285	88.1283	H(4) C(3) O S	õ			
	C13 label (Phosphotyrosine)	13C9_Phospho_Tyr	88.996524	88.9138	H C(-9) C13(9) O(3)	٠ ا			
	Acrolein addition +94	Acrolein94	94.041865	94.1112	H(6) C(6) O	e			
	N-isopropylcarboxamidomethyl	NIPCAM	99.068414	99.1311	H(9) C(5) N O	٠ ا			
	Succinic anhydride labeling reagent light form (N-term & K)	Suc_anh_light	100.016044	100.0728	H(4) C(4) O(3)	۲			
	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)	۲			
	Succinic anhydride labeling reagent, heavy form (+4amu, 4H2), N-term &	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	e			
	Acrolein addition +112	Acrolein112	112.052430	112.1265	H(8) C(6) O(2)	۲			
	ubiquitinylation residue	GlyGly	114.042927	114.1026	H(6) C(4) N(2) O(2)	۲			
	Pyridyl	Pyridyl	119.037114	119.1207	H(5) C(7) N O	e			
	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				
🙆 Do	ne				🥑 Intern	et			

How is phosphorylation handled in Mascot?

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

🗿 UniMod: Record Details - Microso	oft Internet Explorer						
File Edit View Favorites Tools Help							
😮 Back 🔹 🐑 🕐 🗶 🚱 🎧 🔑 Search 🦿 Favorites 😻 Media 🤣 😥 - 🌦 🔄 🦞 Powernarks 🏠 🛧 🛠							
Address 🕘 http://www.unimod.org/cgi/uni	mod.cgi?sort_field1=mono_mas	s&sort_field2=full_name&first_re	cord_to_display=0&records_per_p	age=25&session_id=	ST2QVTWUEQXULGMT&recor 🛩 🛃 Go		
UNIMOD	protein modifications f	or mass spectrometry					
View All Records Add Record			Held	Home Options	Advanced Search Logon		
					60		
			Sear	rch			
Descend Ratella							
	21	Chort name	Phoenbo				
Accession #	21 Deconhomilation	snort name	Phospho				
Composition		Monoisotonic	70 066221	Auerado	79.9799		
Specificity Definition 1	110(3) F	Monoisocopic	79.900331	Average	73.3733		
Site	s	Position	Anywhere				
Neutral Loss	H(3) O(4) P	Monoisotonic	97,976896	Average	97 9952		
Classification	Post-translational		371370030	monugo	5713302		
Comment							
Specificity Definition 2	-						
Site	Т	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	unations of all	Look Modified	2002 10 04 154	52.04			
Gurator	uriimoa	Last Modified	2002-10-04 15:	53:04	Email Change Request		
					Email Change Request		
ど Done					Internet		

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

🗿 Matrix Science - Masco	ot - Set Search Defaults - Microsoft Internet Expl	orer		
File Edit View Favorites	Tools Help			
🌏 Back 🔹 🕥 🕤 🖹	💈 🏠 🔎 Search 🤺 Favorites 🌒 Media	🛛 🖉 - 🌺 🖃 📮 📢	Powermarks 👫 🔺 🛠	
Address 🍓 http://www.matri:	<pre>cscience.com/cgi/form_defaults.pl</pre>			🛩 🄁 Go
(MATRIX)			_	^
(SCIENCE)		HOME : WHAT'S NEW : MASCOT : HELP : PRO	DUCTS SUPPORT CONTACT	earch <u>Go</u>
Mascot > Set Sea	arch Defaults			
Set Mascot	search form defaults			
Database	MSDB			
Taxonomy	All entries	×		
Enzyme	Trypsin			
Allow up to	1 v missed cleavages			
Fixed modifications	AB_old_ICATd0 (C) AB_old_ICATd8 (C) Acetyl (K) Acetyl (K) Acetyl (N-term) Mide (C-term)			
Variable modifications	AB_old_ICATd0 (C) AB_old_ICATd8 (C) Acetyl (K) Acetyl (N-term) Amide (C-term)			
Show all mods.				
ICAT	(MS/MS only)			
Peptide tol. ±	1.2 Da 💌			
MS/MS tol. ±	0.6 Da 💙			
Peptide charge	1+ 💌			
Monoisotopic	● Average ○			
Data format	Mascot generic 🛛 🛛 (MS/MS only)			~
ē				Internet

And check the box for Show all mods

rix Science - Masco Edit View Eavorites	t - MS/MS lons Search - Microsoft Inte Tools Help	ernet Explorer			
		A		F4 - 4	<u>.</u>
ack 🔹 🜍 🖞 🞽	Search X Favorites	Media 🥹		Powermarks 🕅 🕕	*
🗧 🕘 http://www.matrix	science.com/cgi/search_form.pl?FORMVER=28	SEARCH=MIS			✓
		ном			Search
Marces	in a Carante	100		13.5011 OKT CONTACT	
Mascot > MS/MS I	ons Search				
MASCOT	MS/MS Ions Search				
Your name	190	Email	icottrell@matrivscience.com		
Course title	,550	Lindi	Josed engineericstericsteri		
Search due					
Database	MSDB 💌				
Taxonomy	All entries		*		
Enzyme	Trypsin 💌	Allow up to	1 🚩 missed cleavages		
Fixed modifications	AB_old_ICATd0 (C)	Variable modifications	Phospho-NL (D) Phospho-NL (H)	~	
	Acetyl (K)		Phospho-NL (S)	_	
	Amide (C-term)		Probiotinhydrazide (P)	~	
Protein mass	kDa	ICAT			
Peptide tol. ±	1.2 Da 💌	MS/MS tol. ±	0.6 Da 💌		
Peptide charge	1+ 👻	Monoisotopic	⊙ Average ○		
Data file	Brow	wse			
Data format	Mascot generic 🔽	Precursor	m/z		
Instrument	Default				
Overview		Report top	20 V hits		
	Start Search		Reset Form		

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

Peptide Summary Report (/data/20040513/FamTlxHm.dat) - Microsoft Inter	rnet Explorer	
File Edit View Favorites Tools Help		
🌀 Back 🔻 🕥 - 💌 🗟 🏠 🔎 Search 📌 Favorites 🜒 Media 🤗	🔗 · 🍓 📄 🧾 👘 rowermarks 🕅 🔺 🕉	
Address 🗃 http://www.matrixscience.com/cgi/master_results.pl?file=/data/20040513/FamTlxHm.	dat	🗸 芛 Go
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 🛛	Show sub-sets	
Show pop-ups 💿 Suppress pop-ups 🔿 Sort unassigned De	ecreasing Score 🛛 🖌 Require bold red 🗔	
Select All Select None Search Selected Error tolerar	nt	=
1. <u>KBBOA2</u> Mass: 25091 Score: 77 Que	ries matched: 1	
Check to include this bit in error tolerent search		
Check of Include Chis hit in error colerant search		
Query Observed Mr(expt) Mr(calc) Delta Miss Score	Expect Rank Peptide	
✓ 1031.40 2060.78 2060.82 -0.04 0 77	6.2e-05 1 FQSEEQQQTEDELQDK + Phospho (ST)	
V		
Top scoring peptide matches to query 1 Prot Score greater than 35 indicates homology		
Q95L Score greater than 47 indicates identity	ries matched: 1	
Beta Status bar shows all hits for this peptide		
JC13 Score Delta Mit Protein Pentide	ries matched: 1	
beta 76.6 -0.04 1 KBBOA2 FQSEEQQQTEDELQDK		
AAB2 22.5 -0.04 1 KBBOA2 FQSEEQQQTEDELQDK	ries matched: 1	
5672 18.1 -0.19 2 G84453 CLSLSKQVDLFEETIEK	ning makelind, 1	
BOWC 13.7 -0.17 4 Q9VPU8 QLASGEYFLNQEQKQAK	ries matcheu: 1	
10.7 -0.20 5 069721 YDSFPRSDIVTVVIGADK	ries matched: 1	
BOVC 9.9 -0.17 7 OTRLUG NMETILDENDDERLACT		
A329 9.6 -0.12 8 AB1213 KTLTFFGVDPEDENNQK	ries matched: 1	
hata 9.1 -0.20 9 T45584 FQSISCSVEGVLFVLEPK		~
E 1:KBBOA2	🥑 Internet	

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.



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	
File Edit View Favorites Tools Help	
🔇 Back 🔹 🐑 - 💌 😰 🏠 🔎 Search 🤺 Favorites 🜒 Media 🚱 🍃 🍚 💭 😨 -	Powermarks 👫 🔌 🎋
Address 🕘 http://www.matrixscience.com/cgi/master_results.pl?file=/data/20040511/FamuCaEn.dat	💌 🄁 Go
Standard scoring MudPIT scoring Ions score cut-off Show sub-sets Show pop-ups Souppress pop-ups Sort unassigned Decreasing Score Require bold re-	a .
Select All Select None Search Selected Error tolerant 1. P02662-00-00 Mass: 24570 Score: 102 Queries matched: 1	
(CAS1_BOVIN) Splice isoform Displayed; Variant Displayed; Conflict Displayed; from Check to include this hit in error tolerant search	PO2662 Alpha-S1 casein precursor
Query Observed Mr(expt) Mr(calc) Delta Miss Score Expect Rank Peptide Image: style="text-align: center;">Image: style="text-align: center;"/>Image: style="text-align	+ Phospho (ST)
Top scoring peptide matches to query 1 Prote Score greater than 43 indicates identity Prote Score greater than 43 indicates identity Prote Score greater than 43 indicates identity Prote Score greater than 43 indicates identity Prote Score greater than 43 indicates identity Prote Score greater than 43 indicates identity Prote Score greater than 43 indicates identity Prote Score greater than 43 indicates identity Prote Score greater than 43 indicates identity Prote Score greater than 43 indicates identity Prote Score greater than 43 indicates identity Prote Score greater than 43 indicates identity Prote Score greater Prote Prote Prote Prote Prote Prote Prote Prote Prote Prote Prote	P02662 klpha-S1 casein precursor P02662 klpha-S1 casein precursor P02662 klpha-S1 casein precursor EC 3.1.1) (HSL) R + Phospho+PL (S); Phospho+PL (T)
3. <u>P29475-03-00-00</u> Mass: 43868 Score: 26 Queries matched: 1	×
1:P02662-00-00	🖉 🖉 Internet

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 ...

ABRF Discussion List - Micros	soft Internet Explorer	
<u>Eile E</u> dit <u>V</u> iew F <u>a</u> vorites <u>T</u> ools	; Help	1
🔇 Back 👻 🐑 👻 🛃 (🏠 🔎 Search 🤺 Favorites 🜒 Media 🤣 🍰 📓 🐑 📃 😨 🔭 Powermarks 🎊 🔺 %	
Address 🕘 http://www.abrf.org/index	x.cfm/list.msg/43675	> Go
Research Technology	ssociation iomolecular source Facilities (2) White Pages 22 Log In 22 Join ABRF	ip ^
Communication • Education		-
Search 60	ABRF Web-Based Discussion Forum [Information] [Archives Home]	=
About ABBE	[<u>Message Index/Search</u>] [<u>Join Group</u>] [<u>Sign In</u>]	
News and Announcements Communications Research Groups	Re: LC/MS < <u>Last Message</u> by Thread: Re: LC/MS from Svante Resj=?ISO-8839-1?B79g==?= > <u>Next Message</u> by Thread: Re: LC/MS from Willy Bienvenut	
Committees	[Thread Index] Browse from here by [Date] [Author	1
Reference	Date: Tue, 20 May 2003 20:59:32 +0100	
Sponsorship	From: "Nick Morrice" < <u>nick morrice@virgin net</u> >	
Flyout Menu Off	To: "ABKF Discussion List" < <u>ABKF@list abrt org</u> > Subject Re: LC/MS	
	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 wi ll not find from a phosphoprotein derived from in vitro and/or in vivo sourc es. The literature is littered with methods devoted to the analysis of casein phosphorylation sites. However these methods are rarely cited in papers reporting novel sites in biochemical journals. Try either buying or obtaining a known phosphoprotein of say 40-80kDa (common MWt for phosphoproteins), and analyse the tryptic digest of this	×
é	Internet	

If we go the 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 Mitchelhill

So, no more phosphopeptides from casein in this presentation!

Peptide Summary Report (/data/20040513/FamTlbae.dat) - Microsoft Internet Experimentation	cplorer 📃 🗖 🔀					
File Edit View Favorites Tools Help	N					
🌀 Back 🔻 🕥 - 😠 🛃 🏠 🔎 Search 🤺 Favorites 🜒 Media 🔗 🔗	🕹 📄 🥠 😨 * Powermarks 🎼 🔺 🕉					
Address 🗃 http://www.matrixscience.com/cgi/master_results.pl?file=/data/20040513/FamTlbae.dat	💙 🄁 Go					
	<u> </u>					
1. <u>g1 59281</u> Mass: 39323 Score: 79 Queries	matched: 1 catalutic chain - nic (fracmant)					
Check to include this hit in error telerent coord	catalytic chain - pig (liagment)					
Check to include this hit in eitor tolerant search						
Query Observed Mr(expt) Mr(calc) Delta Miss Score Exc	pect Rank Peptide					
✓ 1 1107.90 2213.79 2214.07 -0.28 0 80 2.7e	e-05 1 TWTLCGTPEYLAPEIILSK + Phospho (ST)					
Top scoring peptide matches to query 1						
Prote Score greater than 30 indicates homology						
gi 89 Score greater than 47 indicates identity	natched: 1					
prote Status bar shows all hits for this peptide	calytic chain - pig (fragment)					
gi 10	natched: 1					
prote Score Delta Ait Protein Peptide	sea hare					
gi 10 76.9 -0.28 1 gi 89281 TWTLCGTPEYLAPEIILSK	natched: 1					
prote 37.7 -0.28 1 gi 89281 TWTLCGTPEYLAPEIILSK	nain - California sea hare					
gi 12 18.0 -0.28 1 gi 89281 TWTLCGTPE <u>y</u> LAPEIILSK	natched: 1					
CAMP- 12.9 0.74 2 gi 31224299 <u>T</u> WTLCGTPEYLAPEIIQAK	C-alpha)					
gi 12 9 0.74 2 gi 31224299 TWTLCGTPEYLAPEIIQAK	natched: 1					
CAMP- 0.5 -0.16 2 cillocators TATVESDITEDITEDITED	C-alpha)					
gilla 7.7 -0.06 5 gill3470885 STYGTGCFALLNTGSDLVB	natched: 1					
CAMP- 7.7 -0.06 5 gi 13470885 STYGTGCFALLNTGSDLVR	C-beta-2)					
gi 20	natched: 1					
cAMP-dependent protein kinase catalvtic subunit						
gil230462 Mass: 40414 Score: 79 Oueries	matched: 1					
Chain E. c-AMP-Dependent Protein Kinase (E.C.2.7.1.37) (cA	PK) (Catalvtic Subunit)					
gil284054 Mass: 23939 Score: 79 Oueries	matched: 1					
protein kinase (EC 2.7.1.37), cAMP-dependent, alpha cataly	tic chain, short splice form - human (fragment)					
cril349816 Mass: 40398 Score: 79 Oueries matched: 1						
Chain E, c-AMP-Dependent Protein Kinase (E.C.2.7.1.37) (cAPK) (Catalytic Subunit) ""alpha"" Isoenzyme Mutant With Ser						
cild76500 Mace: 41142 Score: 70 Overies	matched: 1					
🕘 1:gi 89281	🥥 Internet					

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



🗿 Ma	ascot Sear	ch Resu	lts: Pepti	de View -	Mici	osoft Inte	rnet Expl	orer											
File	Edit Viev	v Favor	ites Tools	Help															.
G	Back 🔹 🌘	Ð - [× 2	6	Sear	ch ☆ Fa	vorites 🍕	Media 🚽	Ø	• 🎍 🛙	2								
Addre	ss 🙋 http	//www.ma	atrixscience.	com/cgi/pe	ptide_	view.pl?file=.	./data/2004	10517/Famro	oaYn.dat&qu	uery=18hit=	18index=	gi%7	c892818px=1	*	🔁 Go	Links	🛱 SnagIt	B	• 🖈
8	842.39	421.70	824.38	412.69	Р	1372.77	686.89	1355.75	678.38	1354.76	677.88	12							^
9	971.43	486.22	953.42	477.21	E	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	Α	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	Е	702.44	351.72	685.41	343.21	684.43	342.72	6							
15	1657.79	829.40	1639.78	820.39	Ι	573.40	287.20	556.37	278.69	555.39	278.20	5							
16	1770.88	885.94	1752.87	876.94	Ι	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					к	147.11	74.06	130.09	65.55			1							
G.	•	· · ·	7.5																
9 2.	-0.1																		
Erro							· •												
-	0.2																		
	1	50	0		1000		1500												
RMS	error 104	ppm					Mas	ss (Da)											
NC	BI BLAS	T searci	h of <u>TWT</u>	LCGTPI	EYL	APEIILSK													
(Pa	rameters:	blastp, n	r protein	database,	exp	∜_} =20000	, no filter	, PAM30)										
Oth	Other BLAST web gateways																		
	Marcat Marchane anticipation of a set																		
							Ivia			init in a cicil									
🙆 htt	p://www.nc	bi.nlm.nih.	gov/blast/Bl	ast.cgi?ALI	GNMEN	JTS=50&ALIG	SNMENT_VIE	EW=Pairwise	&AUTO_FO	RMAT=Semi	iauto&CLIE	NT=	web&DATABASE=r	nr&D			Internet		

Follow the link to run a Blast search of this peptide

🕘 NCBI Blast - Micro	soft Internet Explorer	
File Edit View Fax	rorites Tools Help	27
🕝 Back 🔹 🕑 👻	🖹 😰 🏠 🔎 Search 👷 Favorites 🔇 Media 🚱 🗟 - چ 👿 - 📙	
Address 🕘 http://www	ncbi.nlm.nih.gov/blast/Blast.cgi?ALIGNMENTS=508ALIGNMENT_VIEW=Pairwise8AUTO_FORMAT=Semiauto8cCLIENT=web8dDATAE 🔽 🛃 Go 🛛 Links 🗰 Snaglt	🖻 🔁 •
	Protein—protein BLAST Protein Translations Retrieve results for an RD	~
Search		
Set subsequence	From: To:	
<u>Choose database</u>	nr 💌 nr	
Do CD-Search	pet ^k S pet ^k S	
Now:	env_nr v: reserved of reserver	
Options	for advanced blasting	
Limit by entrez query	or select from. All organisms	
<u>Composition-</u> based statistics		
		~
e	Internet 🥶 Internet	

Choose Swiss-Prot as the database, because we want a database with good annotations $% \mathcal{A} = \mathcal{A} = \mathcal{A}$



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

NiceProt View of Swiss-Prot: P2261	2 - Microsoft Internet Explorer	
<u>File E</u> dit <u>V</u> iew F <u>a</u> vorites <u>T</u> ools <u>H</u> elp		A
🌀 Back 🝷 🕥 🔹 😰 🏠 🍃	🛇 Search 🤺 Favorites 🜒 Media 🕢 🍰 🏹 🐨 🖕	
Address 🕘 http://us.expasy.org/cgi-bin/nicepr	bt.pl?KAPG_HUMAN	💌 🄁 Go Links 🛱 Snagit 🗃 🍖 🔹
Features		·
Feature table viewer		
Key From To Length	Description	
INTI_MEI 0 0	The second second second second second second second	
	N-myristoyi giycine (<i>By Similarity</i>).	
ND BIND 49 57 9	ATD (Pro cimilarity)	
BINDING 72 72	AIF (by Similarity).	
ACT SITE 166 166	Proton accentor (By similarity).	
MOD RES 197 197	Phosphothreonine (by autocatalysis) (By similarity).	
MOD RES 338 338	Phosphoserine (by autocatalysis) (By similarity).	
CONFLICT 267 267	$H \rightarrow D$ (in Ref. <u>1</u>).	
CONFLICT 344 344	$A \rightarrow P$ (in Ref. 3).	
Sequence information		
Length: 350 AA Molecular wei	ht: 40303 Da CRC64: 4CA40198369B8D3B [This is a chec	ksum on the sequence]
10 20	30 40 50 60	
GNAPAKEDIE QEESVNEFLA KARGDF.	IKW GWFRQWIRDD DUFERERIEG MGDFGRVMEV	
70 80	90 100 110 120	
KNUEIGGNIK AKIENKOKVV KMKQVE.	IILN EKKILQAIDT FFLYKLQFSF KDNSYLYLYM	E
130 140	150 160 170 180	
FWURCCENES DI ODUCEESE DULCE		
EIVPGGENES REQRVGRESE PHACEY.	IRQV VERVQIERSE DEIRKDEKPE NEEIDQQGYE	
190 200	210 220 230 240	
QUIDIGFARR UNGNIBLEG IPETER	LII LERUTINKAVD WWALGVLITE NAVGPPPFTA	
ê		🔮 Internet

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 Ex	plorer 🗧 🗖 🔀
File Edit View Favorites Tools Help	N
🔇 Back 🔻 🕥 - 💌 🗟 🏠 🔎 Search 📌 Favorites 🜒 Media 🚱 🔗	See See See Powermarks Min 🔺 🛠
Address 🕘 http://www.matrixscience.com/cgi/master_results.pl?file=/data/20040513/FamTibae.dat	💌 🄁 Go
Error tolerant	<u>^</u>
1. gil89281 Mass: 39323 Score: 79 Overies	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 Exp	ect Rank Peptide
✓ 1 1107.90 2213.79 2214.07 -0.28 0 80 2.7e	-05 1 TWTLCGTPEYLAPEIILSK + Phospho (ST)
Prote Score greater than 30 indicates homology	
gi 89 Score greater than 47 indicates identity	natched: 1
prote Status bar shows all hits for this peptide	talytic chain - pig (fragment)
<u>gi 10</u>	natched: 1
prote Score Delta Hit Protein Peptide	sea hare
gi 10 76.9 -0.28 1 $gi 89281$ TWTLCGTPEYLAPEIILSK	natched: 1
prote 37.7 -0.28 1 gi 89281 TWTLCGTPEYLAPEIILSK	hain - California sea hare
<u>gi 12</u> 18.0 -0.28 1 gi 89281 TWTLCGTPE <u>Y</u> LAPEIILSK	natched: 1
CAMP- 12.9 0.74 2 g1 31224299 TWTLCGTPEYLAPEIIQAK	C-alpha)
$\frac{gi 12}{10.1} = \frac{10.74}{10.1} = 10$	natched: 1
CAMP- 9.5 -0.16 3 gi 7021083 TATVSSPLTSPTSPSTLSLK	C-alpha)
gi 12 7.7 -0.06 5 gi 13470885 STYGTGCFALLNTGSDLVR	natched: 1
CAMP- 7.7 -0.06 5 gi 13470885 S <u>T</u> YGTGCFALLN <u>TGS</u> DLVR	A C-beta-2)
<u>gi 20</u>	natched: 1
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) (cA)	PK) (Catalytic Subunit)
gi 284054 Mass: 23939 Score: 79 Queries	matched: 1
protein kinase (EC 2.7.1.37), cAMP-dependent, alpha cataly	tic 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) (cA)	PK) (Catalytic Subunit) ""alpha"" Isoenzyme Mutant With Ser
Cil476500 Mace: 41142 Score: 70 Omerice	matched: 1
🕘 1:gi 89281	🔮 Internet 🤢

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 I-rat) - Microsoft Internet Explorer	×
File Edit View Favorites Tools Help	1
🕝 Back 🔻 🕥 - 🖹 📓 🏠 🔎 Search 📌 Favorites 🜒 Media 🍪 🍛 - 🌺 🚍 🛄 😨 - Powermarks K 🔺 🛠	
Address 🕘 http://www.matrixscience.com/cgi/master_results.pl?file=/data/20040511/FamuIa5e.dat 🗸 🎦	50
	^
Significance threshold p< 0.05 Max. number of hits AUTO	
Standard scoring 💿 MudPIT scoring 🔿 Ions score cut-off 🛛 Show sub-sets 🗆	
Show pop-ups 💿 Suppress pop-ups 🔿 Sort unassigned Decreasing Score 🛛 Require bold red 🗌	
Select All Select None Search Selected True tolerant	
1. <u>P27703</u> Mass: 41249 Score: 64 Queries matched: 1	
(NK01_NOUSE) Nitogen-activated protein kinase 1 (EC 2.7.1.37) (Extracellular signal-regulated kinase 2) (ERK-2) (Nit	с
Check to include this hit in error tolerant search	
Uuery Ubserved Mr(expt) Mr(calc) Delta Miss Score Expect Kank Peptide	
Top scoring peptide matches to query 1	
Score greater than 36 indicates identity	
Search Par Scatus par shows all hits for this peptide	
Type of sea Score Delta Hit Protein Peptide	
Enzyme 67.9 0.80 1 P27703 VADPDHDHTGFLTEYVATR	
Variable mo 44.4 0.80 1 P27703 VADPPHDBHTGFLTEVVATR (S), Phospho-NL (T)	
PROTEIN MAR 1315 0.100 1 PETTOD VARPENDINGLIEVVAR	
Frament Ma 12.6 -0.97 SDYENDECWSALESFR	
Max Missed 12.6 -0.97 SDYENDDECWSALESFR	
Instrument 11.4 0.80 1 P27703 VADPDHDHTGFLTEYVATR	
Number of q 10.5 0.80 1 P27703 VADPDHDHTGFLTEYVATR	
7.6 0.74 VINF <u>Y</u> AGANQSHNVTCVGKK	
w matrixscience.com/	
	-

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

Ho Va V1 Io	ini	(1)5- (2)9- ++(0)9- 	ic s odif	(E)1-	••(2) ⁶	*:(5) *:(5) 600 tral pc 3.4e-05	(9) ntr((9) 000	**(\$1)5. (2)5. (6)07. **(91)5. 100 is hr(cal	6) 5 0 12 c): 2223	(11) ⁵	8	007-1-3(13)			C Mon Var T13 Ion	(1)n. (2)a ++(2)a •+(2)a isotop isotop isotop s Score	(C) (C) (C) (C) (C) (C) (C) (C) (C) (C)	*(L) ^{f,} (8) ^{f,} *(0) *(0) *(1) *(0) *(1) *(0) *(1) *(0) *(1) *(0) *(1) *(0) *(1) *(0) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1) *(1)	(2)0 ⁶⁷ ++(21)3 ++(21)3 600 ral per	ano	**(52).5.(20) **(6)(0) **(6)(0) **(6)(0) **(6)(0) **(6)(0) **(6)(0) **(6)(0) **(6)(0) **(6)(0) **(6)(0) **(6)(0) **(6)(0) **(6)(0) **(6)(0) **(6)(0) **(6)(0) **(6)(0) **(6)(0) **(6)(0) **(6)(0) **(6)(0) **(6)(0) **(6)(0) **(6)(0) **(6)(0) **(6)(0) **(6)(0) **(6)(0) **(6)(0) **(6)(0) **(6)(0) **(6)(0) **(6)(0) **(6)(0) **(6)(0) **(6)(0) **(6)(0)(0) **(6)(0)(0) **(6)(0)(0)(0)(0)(0)(0)(0)(0)(0)(0)(0)(0)(0)	»; 12); 222:	(11)/i	5- 	(FT)A-7-00		
На	ch	:# (<mark>8</mark>	old	Red): 27/1	74 frag	ment	ions us	ing 50 s	most int	ense pe	aks		.	Hat	ches (old Red): 24/1	4 frage	oent	lons usi	ng 50 i	nost int	ense pe	aks		
**		b	b	**	P0	P0++	Seq	ŀУ	y**	y*	y***	y ⁰	y0++		#	b	b**	Pn	P0++	Seq.	у	y**	У*	y***	yu	y0++	#
	1	00.08	5	0.54			V V	0104.00	10/0.05	0107.00	1054.44	2102.00	1052.05	19		171.11	96.06			V A	2124.00	1062.05	2107.00	1054.44	2106.90	1052.05	19
H	1	/1.11 06 14		0.00	520.12	124.6	2 D	2124.90	1002.95	2107.88	1019.44	2106.89	1033.95	18	÷	286.14	143.57	269.13	134 57	n	2053.97	1002.95	2036.84	1012.95	2035.95	1019.43	17
H	2	02.19	10	2.10	265.10	192.0	0 P	1020 04	069.91	1921.91	961.41	1020 92	960.92	16	1 di	383.19	192.10	365.18	183.09	P	1938.84	969.92	1921.81	961.41	1920.83	960.92	16
	2	00.00	24	2.10	490.21	240.6	7 F	1930.04	909.92	1921.01	012.00	1920.85	900.92	16		498.22	249.61	480.21	240.61	D.	1841 79	921.40	1824 76	912.85	1823 77	912 39	15
H	4	26.00	21	0.14	617.97	240.0	4 H	1796.76	962.00	1209.70	916.00	1209.25	054.00	14	6	635.28	318 14	617.27	309.14	н	1726.76	863.88	1709.73	855.37	1708 75	854.88	14
	7	50.20	37	5.66	732.20	366.6	5 D	1589.70	795.35	1572.67	796.94	1571.69	786.35	13	7	750.31	375.66	732.29	366.65	D	1589.70	795.35	1572.67	786.84	1571.69	786.35	13
		87.36	44	4 19	869 35	435.11	8 1	1474.67	737.84	1457.65	729.33	1456.66	728.83	12	8	887.36	444 19	869.35	435.18	н	1474.67	737.84	1457.65	729 33	1456.66	728.83	12
	9	88.41	49	4.71	970.40	485 7	0 7	1337,61	669.31	1320.59	660.80	1319.60	660.31	11	9	988.41	494.71	970.40	485.70	T	1337.61	669.31	1320.59	660.80	1319.60	660.31	11
	10	45.43	52	3.22	1027.42	514.2	2 0	1236.57	618.79	1219.54	610.27	1218.56	609.78	10	10	1045.43	523.22	1027.42	514.22	G	1236.57	618.79	1219.54	610.27	1218.56	609.78	10
	11	92.50	59	6.75	1174.49	587.7	5 F	1179.54	590.28	1162.52	581.76	1161.53	581.27	9	11	1192.50	596.75	1174.49	587.75	F	1179.54	590.28	1162.52	581.76	1161.53	581.27	9
17	13	05.59	65	3.30	1287.58	644.2	9 L	1032.48	516.74	1015.45	508.23	1014.47	507.74	8	12	1305.59	653.30	1287.58	644.29	L	1032.48	516.74	1015.45	508.23	1014.47	507.74	8
13	14	06.63	70	3.82	1388.62	694.8	2 T	919.39	460.20	902.37	451.69	901.38	451.19	7	13	1486.60	743.80	1468.59	734.80	T	919.39	460.20	902.37	451.69	901.38	451.19	7
14	15	35.68	76	8.34	1517.67	759.3	4 E	\$18.34	409.68	801.32	401.16	800.33	400.67	6	14	1615.64	808.32	1597.63	799.32	E	738.38	369.69	721.35	361.18	720.37	360.69	6
15	17	78.71	88	9.86	1760.70	880.8	5 Y	689.30	345.15	672.28	336.64	671.29	336.15	5	15	1778.71	889.86	1760.70	880.85	Y	609.34	305.17	592.31	296.66	591.32	296.17	5
10	18	77.77	93	9.39	1859.76	930.3	9 V	446.27	223.64	429.25	215.13	428.26	214.63	4	16	1877.77	939.39	1859.76	930.39	V	446.27	223.64	429.25	215.13	428.26	214.63	4
1	19	48.81	97	4.91	1930.80	965.9	0 A	347.20	174.11	330.18	165.59	329.19	165.10	3	17	1948.81	974.91	1930.80	965.90	A	347.20	174.11	330.18	165.55	329.19	165.10	3
							Ċ.																				
		A	S	M	IS 2	20	04	1																	ATI TEN	CE)	

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
File Edit View Favorites Tools Help
🔇 Back 🔹 🕥 - 📓 🛃 🏠 🔎 Search 👷 Favorites 🜒 Media 🚱 🍙 - 🌺 📄 🗾 🛱 - Powermarks Ma 🔺 %
Address 🕘 http://www.matrixsdence.com/cgi/master_results.pl?file=/data/20040513/FamTleET.dat 🔍 💽 G
Standard scoring MudPIT scoring Ions score cut-off Show sub-sets Show pop-ups Suppress pop-ups Soft unassigned Decreasing Score Require bold red
Select All Select None Select All Error tolerant 1. gi 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 I 1038.00 2073.98 2073.81 0.17 0 69 0.00026 1 TYVDPFTYEDPNQAVR + 2 Phospho (X)
Top scoring peptide matches to query 1 Prote Score greater than 45 indicates identity gil3 Status bar shows all hits for this peptide simi simi
Galactic Dotation Dotation
2. <u>gi 23</u> growt 6.5 0.13 2 gi 21645679 QEDSYITTESLTTTAVR 6.5 0.13 2 gi 21645679 QEDSYITTESLTTTAVR 6.0 0.07 3 gi 4758794 YTLTETPLLHTAQEAAR Check
Query Observed Mr(expt) Mr(calc) Delta Miss Score Expect Rank Peptide <u>1</u> 1038.00 2073.98 2073.85 0.13 0 6 4.4e+02 8 QEDSYITESLTTTAVR + Phospho (ST); Phospho (Y)
▲ 1:g (4758280

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	
Elle Edit View Favorites Iools Help	
🕜 Back · 📀 · 🖹 🗟 🏠 🔎 Search 👷 Favorites 🜒 Media 🤣 🍰 - 🍃 🕞 🕞 Powermarks 🕅 🔺 %	
Address 🕘 http://www.matrixscience.com/cgi/protein_view.pl?file=/data/20040513/FamTleET.dat8ht=gi%7c5853748px=18protscore=66.80246753856978_mudpit=1000	🔁 Go
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P	~
Sequence Coverage: 1%	
Natched peptides shown in Bold Red	
1 MAGIFYFILF SFLFGICDAV TGSRVYPANE VTLLDSRSVQ GELGWIASPL	
51 EGGWEEVSIM DEKNTPIRTY QVCNVMEASQ NNWLRTDWIT REGAQRVYIE	-
101 IKFILRDCNS LPGVMGTCKE TFALYYYESD NDKERFIRES QFKTDTIAA	-
131 DESFIQUDIG DRINKLINIEI RDVGPLSKKG FYLAFQUVGR CILLVSVRVF 201 VUVCDI TUDNI LACEDITITA ADTESILVEVD GSCUNNEREV DUDUWYGAD	
251 GEWLYPICNC LCNAGHERON GECOACKIGY YKALSTDAK AKCPPESYSV	_
301 WEGATSCTCD RGFFRADNDA ASMPCTRPPS APLNLISNVN ETSVNLEWSS	
351 PONTGGRODI SYNVVCKKCG AGDPSKCRPC GSGVHYTPQQ NGLKTTRVSI	
401 TDLLAHTNYT FEIWAVNGVS KYNPSPDQSV SVTVTTNQAA PSSIALVQAK	
451 EVTRYSVALA WLEPDARNGV ILEYEVKYYE KDONERSYRI VRTARNIDI	
301 ROLARED STOF FROM RATE AND TO DESCRIPTION OF ADDRESS TO COMPANY TO DE	
601 TYEDPNOAVE FFAKEIDASC IKIEKVIGVE FFGEVCSGRL KVPGKREICV	
651 AIKTLKAGYT DKORRDFLSE ASIMGOFDHP NIIHLEGVVT KCKPVMIITE	
701 YMENGSLDAF LRKNDGRFTV IQLVGMLRGI GSGMKYLSDM SYVHRDLAAR YOUZ - NIGN IEVEI OF CONTIGENC	e
751 NILVNSNLVC KVSDFGMSRV LEDDPEAAYT TRGGKIPIRW TAPEAIAYRK Y596 more likely than T595	
801 FTSASDVUSY GIVNUEVNSY GERPYUDNEN ODVIKAIEEG VIEVPPPNDCP	
831 TALHQLERDC WQKEKSDRPK FGQIVNHDDK LINNPNSLKR IGSESSRPNI	
951 LARIGITATT HONKILSSVO AMETOMOONI GRWPV	_
Show predicted peptides also	
Sort Peptides By Residue Number Increasing Mass Decreasing Mass 	
Start - End Observed Mr(expt) Mr(calc) Delta Miss Sequence 595 - 610 1038.00 2073.98 2073.81 0.17 0 TYVDPFTYEDPNQAVR 2 Phospho (Y) (<u>Ions score 69</u>)	
(er	
U.18	
	~
	-
Sinternet	

The numbering in the protein is T595 and Y596. Again, if we hop over to Expasy $% \left(T_{1}^{2}\right) =0$

🗿 NiceProt View of Swiss-Prot: Q03137 - Microsoft Internet Explorer 🗧	
Elle Edit View Favorites Iools Help	.
🚱 Back + 🕑 - 🖹 🗟 🏠 🔎 Search 🌟 Favorites 🜒 Media 🚱 🔗 🍇 🖬 + 🛄 🔁 Powermarks 隆 🔌	
Address 🗃 http://ca.expasy.org/cgi-bin/niceprot.pi?EPA4_MOUSE	→ Go
CHAIN <u>20 986</u> 967 Ephrin type-A receptor 4.	^
DOMAIN <u>20 547</u> 528 Extracellular (Potential).	
TRANSMEM <u>548 569</u> 22 Potential.	
DOMAIN <u>570 986</u> 417 Cytoplasmic (<i>Potential</i>).	
DOMAIN <u>191 325</u> 135 Cys-rich. Y602 - high level of confidence	
DOMAIN <u>328 431</u> 104 Fibronectin type-III 1.	
DOMAIN 441 532 92 Fibronectin type-III 2. Y596 MORE likely than 1595	
DOMAIN <u>621 882</u> 262 Protein kinase.	
DOMAIN <u>911 975</u> 65 SAM.	
SITE <u>984 986</u> 3 PDZ-binding motif (Potential).	
NP_BIND <u>627 635</u> 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 <u>602</u> Phosphotyrosine (by autocatalysis).	
MOD_RES <u>779 779</u> Phosphotyrosine (by autocatalysis) (<i>Potential</i>).	
MOD_RES <u>928 928</u> Phosphotyrosine (by autocatalysis) (<i>Potential</i>).	
CARBOHYD 235 235 N-linked (GlcNAc) (Potential).	
CARBOHYD <u>340 340</u> N-linked (GlcNAc) (Potential).	
CARBOHYD 408 408 N-linked (GlcNAc) (Potential).	
CARBOHYD <u>423 423</u> N-linked (GlcNAc) (Potential).	
VARSPLIC <u>783 832</u> Missing (in <u>isoform Short</u>). VSP_002998	
Sequence information	
Length: 986 AA [This is the length of the Molecular weight: 109801 Da [This is the MW of the CRC64: D16AD8B85668C80E [This is a checksun	1
unprocessed precursor] on the sequence]	
10 20 30 40 50 60	
RATIFIELD SPECIEDAR ISSNIPARE VIELDSKSVQ GELGUTASE EGGUELVSTR	
70 80 90 100 110 120	
	~
I DENNIFIERT OVUNVERASO INNIGETDIELE REGAUNVER. EKFERNOUNS EPOVMETCEK	

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.

🕈 Peptide Summary Report (/data/20040511/FamuCfES.dat) - Microsoft Internet Explorer	
Elle Edit View Favorites Iools Help	
🕝 Back 🔻 🕥 🕆 🖹 📓 🏠 🔎 Search 👷 Favorites 🜒 Media 🚱 🔗 🌺 📄 🛄 😨 * Powermarks 🎼 🔺 %	
Agdress 🕘 http://www.matrixscience.com/cgi/master_results.pl?file=/data/20040511/FamuCfES.dat	🖌 🄁 Go
Significance threshold p< 0.05 Max. number of hits 20	~
Standard scoring 💿 MudPIT scoring 🔿 Ions	
Show pop-ups 💿 Suppress pop-ups 🔿 Sor 🔒 -0.02	
Select All Select None Search Selected	
1. g1 92 686 Mass: 6057 Score 0.05	
protein-tyrosine kinase (EC 2.7.1.112) tyr 200 400 600 800 1000 1200 1400	
Check to include this hit in error tolerar PMS error 44 ppm Mass (Da)	
Query Observed Mr(expt) Mr(calc) Delta Miss Score Expect Rank Peptide	
✓ 1 780.30 1558.58 1558.66 -0.07 U 63 0.00076 1 VLEDDPEAAYTTR + Phospho (Y)	
Top scoring pettide matches to guery 1	
Prote Score greater than 17 indicates homology	
gil91 Score greater than 44 indicates identity peries matched: 1	
prote Status bar shows all hits for this peptide (fragment)	
gilla Score Delta Hit Protein Pentide	
Ephr 63.1 -0.07 1 gij92686 VLEDDPEAAYTTR kinase receptor ETK1) (MEK4)	
<u>gi 11</u> 60.6 -0.07 1 gi 92686 VLEDDPEANTTR peries matched: 1	
Ephr 60.6 -0.07 1 gi 92686 VLEDDPEAAYT <u>T</u> R kinase receptor ETK1) (HEK) (HEK4)	
$g_{1} _{42}$ 8.6 0.07 2 $g_{1} _{37360416}$ <u>SOSNLQGLDDSR</u> peries matched: 1	
prote 7.1 -0.03 3 g114195007 AVIGYNPTITEK (fragment)	
gil41 6.1 -0.03 3 gil4150007 AVTGYRDPTTE heries matched: 1	
prote 6.1 -0.03 3 gi14195007 AVTORDPYTEK (fragment)	
gil54 6.0 -0.03 3 gil14195007 AVTGYRDPYTEK heries matched: 1	
Ephr: 2.4 0.07 2 gi 37360416 <u>SQSNLQGLDDS</u> R kinase receptor SEK) (MPK-3)	
gi 63 ueries matched: 1	
receptor tyrosine kinase - rat	
c	>
1:g]92686	.:

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





🙆 Pe	ptide Sum	ımary Re	port (data/200	40514/Famerisa	. dat) - Microso	ft Internet E	xplorer						
File	Edit View	Favorit	es Tools	Help										
()	Back 🔹 🌔) - [1	6	Search	orites 📢 Media	@	• 🎍 🖻		e -	Powermarks	16 1/2	6	
Addres	s 🕘 http:,	//www.mal	rixscience	.com/cgi/ma	ster_results.pl?file=.	./data/20040514/F	amerisa.dat							🖌 🔁 📀
		Signir	icance tr	resnoia p	< 0.00	IVIAX. numbe	er or mits 20							^
		Stand	lard scot	ing 💿 N	fudPIT scoring(🔵 Ions score c	ut-off 0		ŝ	Show sub-sets				
		Show	pop-up	s 💿 Sup	press pop-ups (🔘 Sort unassig	ned Decrea	asing Score	✓ I	Require bold rea	L 🔲			
Se	elect All	Sel	ect None		Search Selected	Error	tolerant							
1.	<u>gi 1</u>	6078640	2	Ma	ss: 71879	Score: 101	Querie	s matched	: 1					
	simi	lar to	protei	n kinas	e [Bacillus :	subtilis]								
	Chec:	k to in	nclude	this hi	t in error to	olerant sear	ch							
	Query	7 Obse	rved	Mr(expt) Mr(calc)	Delta Miss	Score E	xpect Rani	k Pept	ide				
	Image: A state of the state	282	5.20	2824.1	9 2824.19	0.01 0	109 1.0	6e-07 1	SSTT	ITHINSVLGSVI	YLSPEQAR	+ 3 Phos	spho (ST)	_
	9	<u> </u>												
2.	gi 2	Top so Score	oring greate	peptide r than	matches to 31 indicates	query 1 homology			: 1					
	prob	Score	greate	r than	54 indicates	identity								
	Chec	Status	s bar s	hows al	l hits for t	his peptide								
	0	Score	Delt	a Hit	Protein	Peptide			- D					
	Query	109.0	0.0	1 1	gi 16078640	SST <u>T</u> I <u>T</u> H <u>T</u> NSV	LGSVHYLSP	EQAR	FELF	PSKARGAATGFT	TLVLSAANL	IV + Pho	spho (ST)	
	1	103.0	0.0	1 1	gi 16078640	SSTTITHTNSV	LGSVHYLSP	EQAR			10100man			
		90.7	0.0	1 1	gi 16078640	SSTTITHINSV	LGSVHYLSP	FOIR						
	Prot	80.7	0.0	1 1	gi 16078640	SSTTITHTNSV	LGSVHYLSP	EQAR						
	gi 1	50.4	0.0	1 1	gi 16078640	SSTTITHTNSV	lgsvhyl <u>s</u> p	EQAR	: 1					
	simi	29.8	0.0	1 1	gi 16078640	SSTTITHINSV	lg <u>s</u> vhylsp	EQAR						
		11.3	0.7	0 2	gi 2280500	PELFPSKARGA	ATGFTTLVL	SAANL IV	_					
з.	<u>gi 1</u>	10.5	-0.1	4 3	gi[2200300 gi[16077729	KEIKDILAYLR	LVSNPDDDI	SET	: #	<pre># arrange</pre>	ements	s of 3	/ 9 = 84	F
	simi			·						U				
	Chec	k co n	ic ruue	chis hi	c in error c	Jierant Sear	cn.							
	Query	7 Obse	rved	fr(expt) Mr(calc)	Delta Miss	Score E	xpect Ranl	k Pept	ide				~
ど 1:gi	16078640												🥝 Internet	

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	v of Swiss-Pro	t: 034507	- Microsoft Internet	Explorer					
<u>E</u> ile <u>E</u> dit ⊻iew	F <u>a</u> vorites <u>T</u> o	ols <u>H</u> elp							A.
🌀 Back 🝷 🌔	- \star 💈) 🏠 🔎	Search 📌 Favorites	🔇 Media	🕝 🔗 🎍	.	•	Powermarks 👫 \land 🎋	
Address 🙆 http://	ca.expasy.org/c	gi-bin/niceprot.	.pl?O34507						💌 🄁 Go
PRESAGE		03450	<u>/</u> .						<u>^</u>
DIP		03450	<u>7</u> .						
ModBase		<u>03450</u>	<u>7</u> .						
SMR		03450	7; 9653AB5CFBAA	17900.					
SWISS-2DPA	.GE	Get reg	non on 2D PAGE.						
UmRef		View c	luster of proteins with	n at least <u>50%</u>	<u>6</u> / <u>90%</u> identity.				
Keywords									
Hypothetical	protein; <u>Tra</u>	<u>nsferase; S</u>	Serine/threonine-pr	otein kinase	; <u>ATP-binding</u> ;	<u>Repeat; Comple</u>	ete proteom	<u>e</u> .	
Features									
Peat	ire table view	<u>er</u>			Ŧ	Feature alia	<u>mer</u>		
Кеу	From To	Length l	Description						
DOMAIN	<u>11 271</u>	261 1	Protein kinase.						
DOMAIN	356 424	69 1	PASTA 1.						
DOMAIN	425 492	68 1	PASTA Z.						
NR BIND	17 25	- 10	PADIA J. ATR (Presimilari	***)					
BINDING	40 40	<i>.</i>	AIP (<i>by similari</i> ATP (<i>by similari</i>	cy). tw					
ACT SITE	134 134		By similarity.	·37).					
Sequence infe	rmation								
Length: 648 A		cular weight	: 71866 Da	CRC64-9	653AB5CEBA	47900 [This is a	checksum or	the sequence]	
10	20	contar weight	30 40	50	60 60	4/200 [1103 13 2	CHECKSGIII OI	and sequence]	
1			1 1	1					
MLIGKRISGR	YQILRVIGGG	GMANVYLA	ED IILDREVAIK II	RFDYANDN	EFIRRFRREA				
70	80		90 100	110	120				
1	I		1 1	1	l.				
QSASSLDHPN	IVSIYDLGEE	DDIYYIVM	IEY VEGMTLKEYI TA	NGPLHPKE .	ALNIMEQIVS				
130	140	1	.50 160	170	180				
	1		1 1	1	1				
e									Internet

If we follow through the Blast search route and hop over to Expasy, 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 - Microsoft Internet Explorer		
Eile Edit View Favorites Iools Help		
🕝 Back 🔹 🕥 - 🖹 🖉 🐔 🔎 Search 👷 Favorites 🗬 Media 🤣 🍰 📓 🛀 🧾	Powerm	arks 🌃 🔺 🎋
Address 🚳 http://www.cbs.dtu.dk/cgi-bin/nph-webface?jobid=netphos,40AB4085061B98D9&opt=none		🗸 🄁 🕫
CENTERFO RBIOLOGI CALSEQU ENCEANA LYSIS CBS		
648 Sequence MLICKRISCRYCILEVICGCGHANVYLAEDIILDREVAIKILEFDYANDNEFIRFFREAQSASSLDHPNIVSIYDLGEE DDIYYIVHEVVECHTLKEVITANCPLHPKEALNIHEQIVSAIAHAHGNQIVHEDIKPENILIDHKONIKVTDFGIATALS STITIHTNSVLCSVHYLSPEQARCGLATKKSDIVALGIVLFELLTGRIPFDCESAVSIALKHLQAETBAKKNMPSVPCS VENILKATAKDPFHRYETAEDMEADIKTAFDADRLNEKRFTIQEDEEMTKAIPIIKDELAKAAGEKEAEVTTAQENKT KNNCKRKKWPWLLTICLVITAGILAVTVFPSLFHENOVKIPVSGORVEYRAGLLEKEGLQVDSEVLEISDEL NVKTDFRADTTVKEGATVILKSTCKAKTEIGDVTGGTVDQAKKALKDQGNHVTNNEVNDENNGTVIDGPSAGTELV PSEDQVKLVTSIGPEDITALDLTYSKEASGYLENGLUVLEKEAYSDPFGQVVCKPAAGTAVCPONEVEVTSLC PEKFAKTVKEKVKIPVEPENEGDELQVQIAVDDADHSISDTYEEFKIKEPTERTIELKIEPCQKCYVQHVNNKVVSYK TIEVPKDE STS.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.Y.S.S.Y.S.S.Y.S.Y.S.S.Y.S.S.Y.S.Y.S.Y.S.S.Y.S.S.Y.S.S.Y.S.S.Y.S.Y.S.Y.S.S.Y	80 160 240 320 400 480 560 640 720 80 160 240 320 400 480 560 640 720	
Phosphorylation sites predicted: Ser: 16 Thr: 13 Tyr: 7		
Serine predictions Name Pos Context Score Pred		
		Internet

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.



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