



You will remember from the introduction, that sequence queries are searches where mass information is combined with amino acid sequence or composition information



The best known example is a sequence tag search, where a few residues of amino acid sequence are interpreted from the MS/MS spectrum.



You can enter sequence tags, and other types of query, into the sequence query form.

Remember that all the search parameters, including enzyme specificity, modifications, and precursor charge, still apply to this type of search.

Mascot will look for a match between the tag and the ion series specified by the instrument type. Note that Mascot will only try to match the tag against ion series formed by a single backbone cleavage, and maybe a neutral loss, like y or  $b^*$  or  $y^{++}$ . It won't try to match against side chain cleavage fragments, like d, v, w or internal fragments.



Unless you have high energy fragmentation, and are able to distinguish L from I by side chain cleavage fragments, then this tag is wrong. It should be I or L.

Ambiguity in a tag is fine as long as it is recognised and spelt out. Most times, you won't know whether a residue is Q or K. F is almost identical to oxidised M. If the peaks are weak, are you sure you have a mass difference of R, or could it be VG and the intermediate peak is missing?



If the sequence is in the database, it is easier and safer to perform an MS/MS search of the peak list. In this sense, the standard sequence tag is obsolete.

The error tolerant tag, which can find a match when there is an unsuspected modification or a small difference in the sequence, is very powerful and very useful.

Imagine we had an unmodified peptide of MH+ 1314.7 and we interpreted a tag of TISP in the b+ series between peaks at 614.3 and 911.5.

What happens if there is a modification or SNP that increases mass by 100 Da?

If the mod is on the N-term side of the tag, all the masses shift up by 100. However, if it is on the C-term side, only the peptide mass changes.

If the tag was in the y ion series, the reverse would be observed



The error tolerant tag allows for this. In effect, it allows the peptide mass to vary and allows the tag to float. However, the tag must stay attached to one end or the other. Either both fragment ion masses are unchanged or both fragment ion masses shift by the same amount as the precursor.

This causes a huge loss of specificity, so we cannot allow etag searches with very wide peptide mass tolerance (> 1% or > 10 Da) or with no enzyme specificity. The enzyme specificity in an etag search is never fully specific, in any case, because one end of the peptide can just extend until it finds a cleavage point.



Tags can be entered with the high mass fragment on the left or the right. These two tags are identical

Mascot allows multiple tags in a single query. That is, you can call multiple tags from a single MS/MS spectrum. Tags are scored probabilistically. If one tag is wrong, you can still get a good match from the tags that are correct.

If one tag in a query is an etag than all the tags for that query are treated as etags, (not all tags in the search, just in the query)

Finally, you cannot mix ions qualifiers with tag or etag qualifiers. It would just be too complicated.



A lot of people call tags using a calculator and a table of mass values. An alternative is to use Mascot Distiller. Here is a short movie to illustrate

-Maximise the window

-Choose a likely looking peak, such as 987.384

-Right click to start a tag

-Click on any arrow to extend the tag

-In general, I'll just go for the biggest peak

-Stop when it starts to look tricky

-Here's the tag

-Do a Mascot search of the peak list to see what the answer should have been. Here's one I prepared earlier.

-Whoops! Got it wrong, should have been GE, not W. I'll stick to the day job.



Alternatively, you can automate the process entirely by using the de novo algorithm. Here's a nice spectrum where the Mascot database search has failed to find a match If we right click the peak list and choose de novo ...



We get a reasonably high scoring solution, but with a fair amount of uncertainty



Right click the solution and choose Mascot search from the context menu. Note that we have already toggled the tag type to error tolerant

Kascot Distiller - qtof20953*							
Ele Edit View Processing Analysis Ioo	ls <u>W</u> indows <u>H</u> elp						
i 📸 🚔 🔁   X 🗈 🛍 🝓 👀 🞫 😨	Mascot Search (t61-j						
Peak Lists						<u>^</u>	
VIILEGILLLTNPK(3+),	MASCOT	S					
LASQTAILLRGKHK(3+),	MASCUI	Sequence Query					
LIINFIENPARIR(3+), 9			Tree H		_		
THIMOAVGTENME(3+), 6	r our name		Email	1			
IIETGGNDVYVAER(3+),	Search title	Raft - 8					
LLKTDNGPAYTSQK(3+),	Database(s)	MSIPI human	Enzyme	Trynsin/P			
IIQVHFSPLPGDGR(3+),		MSIPI_mouse					~
ia m/z 754.686, 2+ 147: Scar		NCBInr 💼	Allow up to	2 🗾 missed cleavages		10	
Summed Scans		Trembl 💌	Quantitation	None	-	r i i i i i i i i i i i i i i i i i i i	
Denovo Solutions (Tryps	Tawanamu	All optrios		-			
-Men G-INGIVDADSVIPCIVCI	Taxonomy						
-% -MINGIVIHDISYIPGIIKG	Fixed modifications	Carbamidomethyl (C)	>	Acetyl (K) Acetyl (Ni torm)	~		
- TING IVDADSY [ PG ] [ KG ]	modifications		-	Acetyl (Protein N-term)			
-[NG][NG]VDADS[Gf]1[			-	Amidated (C-term)			
-MERNJWDSY[PG][KG],		Display all modifications		Amidated (Protein C-term)			
- 🏍 G-YSDADV[HC][KG], 65	the state is a	Display an mouncacions		Biotin (K)			
- 🎭 G-[NG]V[WD]SY[PG][KG	Variable modifications	Oxidation (M)	2	Biotin (N-term)			
- See GNGT[NG]VEGDSY[PG][K	mouncations		-	Carbamyl (N-term)			
.‰ G-YSDADV[YYR], 63			-	Carboxymethyl (C)	~		
Ascot Search Results	Pentide tol +		MS/MS tol +	0.300			
VEEYEMQIVVNR(2+), 10	repeace con 1		113) 113 COL 1	10.000 100 11			
VEFGEIDLINK(2+),	Peptide charge	Mr 💌	Monoisotopic	Average C			
A NAVED CHEFT (2+)		BEGIN IONS		^			
FLWFAMGFLVAAAR(2+)		TITLE=147: Sum of 9 scans in rai	nge 2217 (rt=3	2.8655, f=2, i=520) to			
GGATALMPYDEVLR(2+)	Query	2241 (rt=33.0972, f=2, i=528)					
EEVLEGEHPODAB(2+), 4		PEPMASS=1508.3654					
DLENKEKTGGFDR(2+), 4		FTAG=501.20154.VDA.786.2873	4	×			
- A KALEGTDMEEIKK(2+), 2	Instrument	ESI-TRAD					
📝 RRLMTASTAYPIK(2+), 2	anstrument.						
ģ∰ m/z 530.675, 2+ 148: Scm	Decoy		Report top	AUTO 🗾 hits			
🖬 🏠 Summed Scans		Start Search		Reset Form			
E Mascot Search Results							
TLLDIDNTR(2+), 45							
TIDOWEDK(2+) 10		Copyright © 2008 Mi	atrix Science <u>Ltd.</u>	All Rights Reserved.		~	
TTLIDNSED(2+) 17							
<						Cancel	
Acquisition Precursors Peak Lists Searches							m/
Ready						1678.4	417 - 0.284
						1	
MASCOT . C		Overies				SM .	ATRIX
MAJUUI : 30	equence	Queries		© 2007-2010 Matrix \$	Science	1.80	TENCE
	-						

Distiller populates the query field with the tags taken from the non-ambiguous parts of the de novo solution. We submit the search ...



And back comes the result. Note that the results from this most recent search have replaced the original database search. You can switch back to the previous results by selecting them on the searches tab.

This match looks promising. If we right click and choose to view the full Mascot report in a browser ...



We can see a good illustration of my earlier point about the enzyme being almost semispecific in an etag search. The peptide can just extend until it finds a cleavage point and then hypothesise a modification that causes a loss of mass to bring the peptide mass back into line.

If we click on the hyperlink to see the peptide view ...

) м	ascot Searcl	h Results: P	eptide View	- Moz	illa Firefox				
Eile	Edit View	Higtory Boo	okmarks Iools	; <u>H</u> el	lp .				
	)> - C	: × 🏠	http://	t61-jsc	/mascot/cgi/pepi	tide_view.pl?file	=/data/2010	D421/F016234.dat&query=18hit=18index=PPB1_HUMAN&px=1&section=5&ave_thre 🏠 🔹 🔣 🕤 Google	
Su	ugar 🐛 pubw	ww1 🐛 pubr	www1 Status 🖸	Curr	ency Converter	🗋 MS Bu	gs & FAQ 阻	fwiki 🗋 Family report 😓 2009_ASMS_Fal_Wor 🔃 MascotImproveReport	
•	Peptide Summa	ry Report (Rafi	t - 8) 🛛 🔀	-	Mascot Search	n Results: Pep	tide V 🔀 🗍	*	
r M SC	ATRIX	Masc	ot Sea	arc	h Res	ults			
Pej	ptide Viev	w							
IS ou Aat tag	MS Fragme nd in PPB1 ich to Query (794.22814	entation of N _HUMAN, 1: 1507.35 ,ADS,521.1	Alkaline pho 8124 from(1 15782)	VPAS ospha 508.1	SAR tase, placent: 365400,1+) (	al type OS=] etag(501.20	Homo sapie: 154,VDA,7:	15 GN=ALPP PE=1 SV=2 36.28734) etag(786.28734,DSY,1151.38063) etag(600.28103,DAD,901.30219)	
on ix ns	oisotopic ed modifi uspected n s Score:	mass of cations: modificat 88 Expec	neutral pe Carbamidor ion: 56.71 t: 2e-005	u=3 eptic methy 104 I (1	ie Mr(calc yl (C) (ap Da, locate help)	): 1450.64 ply to spe d in the r	241 (1=33) 177 scified re region N-1	vyrz, (, (, z-) esidues or termini only) erm to W2	
#	b	b*	b <sup>0</sup>	Seq.	у	y*	y <sup>0</sup>	#	
1	115.0502	98.0237		N				13	
2	301.1295	284.1030		W	1337.6121	1320.5855	1319.6015	12	
3	464.1928	447.1663		Y	1151.5327	1134.5062	1133.5222	11	
4	551.2249	534.1983	533.2143	S	988.4694	971.4429	970.4588	10	
5	666.2518	649.2253	648.2413	D	901.4374	884.4108	883.4268	9	
6	737.2889	720.2624	719.2784	Α	786.4104	769.3839	768.3999	8	
7	852.3159	835.2893	834.3053	D	715.3733	698.3468	697.3628	7	
8	951.3843	934.3577	933.3737	v	600.3464	583.3198	582.3358	6	
9	1048.4371	1031.4105	1030.4265	P	501.2780	484.2514	483.2674	5	
0	1119.4742	1102.4476	1101.4636	Α	404.2252	387.1987	386.2146	4	
ıı	1206.5062	1189.4796	1188.4956	S	333.1881	316.1615	315.1775	3	
2	1277.5433	1260.5168	1259.5327	Α	246.1561	229.1295		2	
F	Find:			Prev	vious 🖌 Highliq	, pht <u>a</u> l 🔲 Mat			
one									
									_
	AC					~		(MATR	IX

The match was obtained by placing a modification delta of +57 Da on the N-term residue. This is almost certainly carbamidomethylation, which can derivatise amino groups if the conditions aren't right. This was why the original database search failed to get a match and this is why the error tolerant tag is so useful



If you want to get as many identifications as possible, as efficiently as possible, you might come up with a strategy similar to this.

seq()			
•Like a tag •Most likely 1234 sec	, but without fragment n y, from non-MS sequenci g(n-AC[DHK]) seq(c-HI) seq(*-	nass information ng, e.g. Edman GF) 4 it is a filtor	
•seq() is no		, it is a filler	
•seq() is no	Meaning	Example	1
•seq() is no Prefix b-	Meaning N->C sequence	Example seq(b-DEFG)	9
•seq() is no Prefix b- y-	Meaning N->C sequence C->N sequence	Example seq(b-DEFG) seq(y-GFED)	-
•seq() is no Prefix b- y- *-	Meaning   N->C sequence   C->N sequence   Orientation unknown	Example seq(b-DEFG) seq(y-GFED) seq(*-DEFG)	-
•seq() is no Prefix b- y- *- n-	Meaning   N->C sequence   C->N sequence   Orientation unknown   N terminal sequence	Example seq(b-DEFG) seq(y-GFED) seq(*-DEFG) seq(n-ACDE)	-
•seq() is no Prefix b- y- *- n- c-	Meaning   N->C sequence   C->N sequence   Orientation unknown   N terminal sequence   C terminal sequence	Example seq(b-DEFG) seq(y-GFED) seq(*-DEFG) seq(n-ACDE) seq(c-FGHI)	
•seq() is no prefix b- y- *- n- c-	Meaning   N->C sequence   C->N sequence   Orientation unknown   N terminal sequence   C terminal sequence	Example seq(b-DEFG) seq(y-GFED) seq(*-DEFG) seq(n-ACDE) seq(c-FGHI)	

Besides tag and etag, Mascot supports a number of other sequence qualifiers. One of these is seq()

Note that seq() is a filter. It must be correct or there will be no match



The other important one is comp(). This would be useful in an ICAT search. Note that comp() is a filter. It must be correct or there will be no match



As always, there is more information in the Mascot help pages. These references are a good starting point if you are interested in learning more about the potential of combining mass and sequence information.