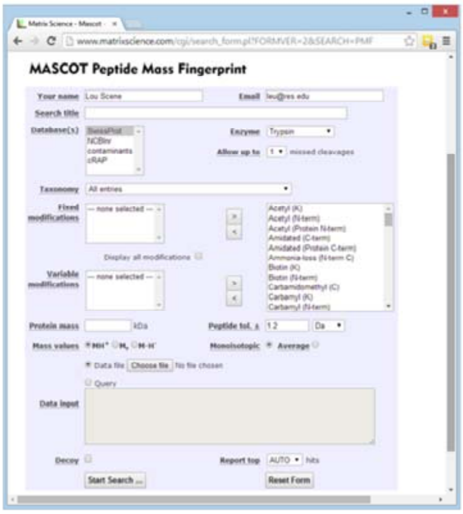
Search Parameters

MASCOT

MATRIX SCIENCE

Search Parameters



www.matrixscience.com/cgi/search_form.pl?FORMVER=2&SEARCH=PMF

MASCOT Peptide Mass Fingerprint

Your name: Lou Scene

Email: lou@res.edu

Search title:

Database(s): SwissProt, NCBInr, contaminants, cRAP

Enzyme: Trypsin

Allow up to: 1 missed cleavages

Taxonomy: All entries

Fixed modifications: --- none selected ---

Display all modifications □

Variable modifications: --- none selected ---

Acetyl (K), Acetyl (N-term), Acetyl (Protein N-term), Amidated (C-term), Amidated (Protein C-term), Ammonia-loss (N-term C), Biotin (K), Biotin (N-term), Carbamidomethyl (C), Carbamyl (K), Carbamyl (N-term)

Protein mass: kDa

Peptide tol. ±: 1.2 Da

Mass values: ● MH⁺ ○ Mr ○ M-H⁻

Monoisotopic ● Average ○

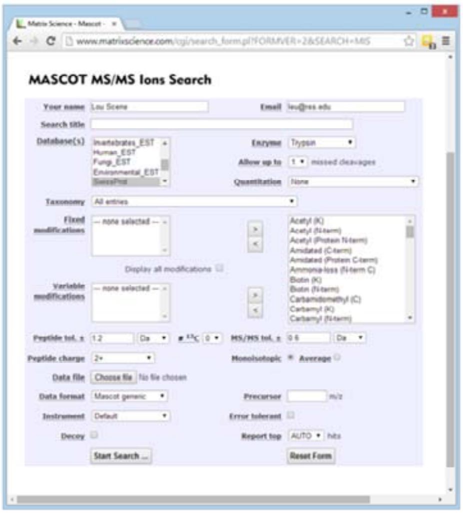
● Data file [Choose file] No file chosen

○ Query

Data input:

Decoy □

Report top: AUTO hits



www.matrixscience.com/cgi/search_form.pl?FORMVER=2&SEARCH=MIS

MASCOT MS/MS Ions Search

Your name: Lou Scene

Email: lou@res.edu

Search title:

Database(s): Invertebrates_EST, Human_EST, Fungi_EST, Environmental_EST, SwissProt

Enzyme: Trypsin

Allow up to: 1 missed cleavages

Quantitation: None

Taxonomy: All entries

Fixed modifications: --- none selected ---

Display all modifications □

Variable modifications: --- none selected ---

Acetyl (K), Acetyl (N-term), Acetyl (Protein N-term), Amidated (C-term), Amidated (Protein C-term), Ammonia-loss (N-term C), Biotin (K), Biotin (N-term), Carbamidomethyl (C), Carbamyl (K), Carbamyl (N-term)

Peptide tol. ±: 1.2 Da # ^{13}C: 0

MS/MS tol. ±: 0.6 Da

Peptide charge: 2+

Monoisotopic ● Average ○

Data file: [Choose file] No file chosen

Data format: Mascot generic

Precursor: m/z

Instrument: Default

Error tolerant □

Decoy □

Report top: AUTO hits

MASCOT : *Search Parameters*

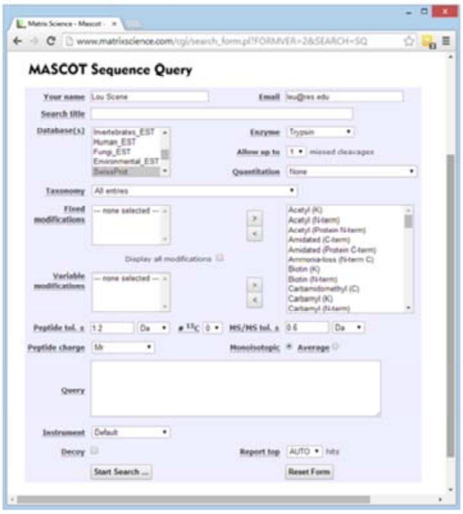
© 2007-2014 *Matrix Science*

MATRIX SCIENCE

In this presentation, we will describe each of the Mascot search parameters.

If you submit a search from a web browser, you have a choice of three different search forms. All three forms submit to the same search engine, but they have been optimised for three different types of search. The form for a peptide mass fingerprint is shown on the left, and the form for a search of uninterpreted MS/MS data on the right. Most of the controls are common to both.

Search Parameters



www.matrixscience.com/cgi/search_form.pl?FORMVER=2&SEARCH=SQ

MASCOT Sequence Query

Your name: Lou Scene

Email: lou@res.edu

Search title:

Database(s): SwissProt

Enzyme: Trypsin

Allow up to: 1 missed cleavages

Quantitation: None

Taxonomy: All entries

Fixed modifications: --- none selected ---

Display all modifications □

Variable modifications: --- none selected ---

Peptide tol. ±: 1.2 Da

^{13}C: 0

MS/MS tol. ±: 0.6 Da

Peptide charge: Mr

Monoisotopic ● Average ○

Query:

Instrument: Default

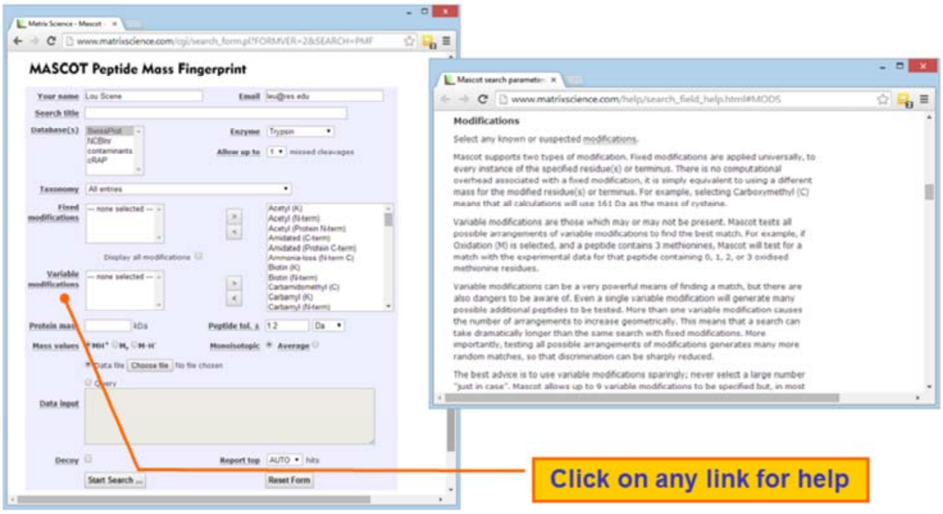
Decoy □

Report top: AUTO hits

The third form is for a sequence query, such as a sequence tag search. The controls on this form are very similar to those on the MS/MS form. The main difference is that we have a text area to type in the queries, rather than a data file upload control.

Help

PMF✓ SQ✓ MS/MS✓



Matrix Science - Mascot

www.matrixscience.com/cgi/search_form.pl?FORMVER=2&SEARCH=PMF

MASCOT Peptide Mass Fingerprint

Your name	Lou Scene	Email	lou@res.edu
Search title			
Database(s)	SwissProt NCBInr contaminants cRAP	Enzyme	Trypsin
		Allow up to	1 missed cleavages
Taxonomy	All entries		
Fixed modifications	-- none selected --		Acetyl (K) Acetyl (N-term) Acetyl (Protein N-term) Amidated (C-term) Amidated (Protein C-term) Ammonia-loss (N-term C) Biotin (K) Biotin (N-term) Carbamidomethyl (C) Carbamyl (K) Carbamyl (N-term)
	Display all modifications □		
Variable modifications	-- none selected --		
Protein mass	kDa	Peptide tol. ±	1.2 Da
Mass values	● MH⁺ ○ Mr ○ M-H⁻	Monoisotopic	● Average ○
	● Data file Choose file No file chosen ○ Query		
Data input			
Decoy	□	Report top	AUTO hits

Start Search ... Reset Form

Mascot search parameter

www.matrixscience.com/help/search_field_help.html#MODS

Modifications

Select any known or suspected modifications.

Mascot supports two types of modification. Fixed modifications are applied universally, to every instance of the specified residue(s) or terminus. There is no computational overhead associated with a fixed modification, it is simply equivalent to using a different mass for the modified residue(s) or terminus. For example, selecting Carboxymethyl (C) means that all calculations will use 161 Da as the mass of cysteine.

Variable modifications are those which may or may not be present. Mascot tests all possible arrangements of variable modifications to find the best match. For example, if Oxidation (M) is selected, and a peptide contains 3 methionines, Mascot will test for a match with the experimental data for that peptide containing 0, 1, 2, or 3 oxidised methionine residues.

Variable modifications can be a very powerful means of finding a match, but there are also dangers to be aware of. Even a single variable modification will generate many possible additional peptides to be tested. More than one variable modification causes the number of arrangements to increase geometrically. This means that a search can take dramatically longer than the same search with fixed modifications. More importantly, testing all possible arrangements of modifications generates many more random matches, so that discrimination can be sharply reduced.

The best advice is to use variable modifications sparingly; never select a large number "just in case". Mascot allows up to 9 variable modifications to be specified but, in most

Click on any link for help

MASCOT : Search Parameters

© 2007-2014 Matrix Science

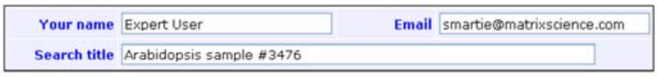
MATRIX SCIENCE

At the top of each slide, there is a key to show which search parameter applies to which type of search.

The labels on the search form are hyperlinks. Just click on them to get detailed help

User details and title

PMF✓ SQ✓ MS/MS✓



Your name	Expert User	Email	smartie@matrixscience.com
Search title	Arabidopsis sample #3476		

- Search form will ‘remember’ user name and email address in cookie
- If Mascot security is enabled, then this information taken from user database
- Email address used for sending results
- Search title is shown in the report, and can help locate a search in the search log.

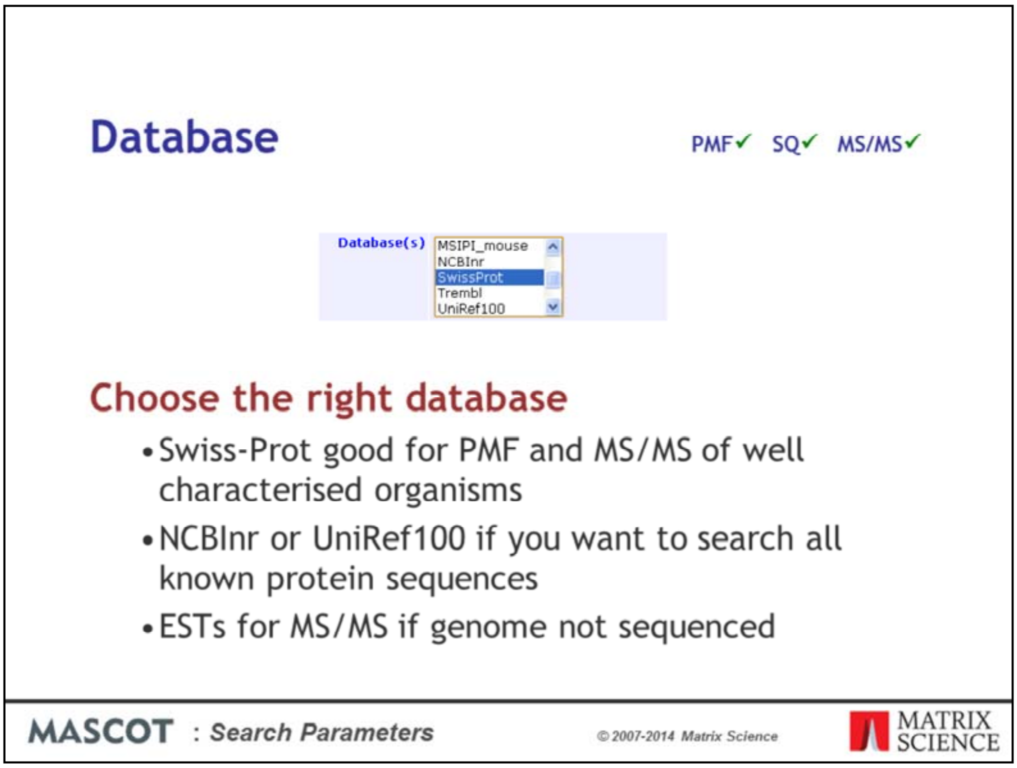
MASCOT : *Search Parameters* © 2007-2014 Matrix Science MATRIX SCIENCE

At the top of the form are a couple of fields for user information. The name and email are saved as a browser cookie when a search is submitted, so you don’t need to complete them every time.

If you have an in-house server, and Mascot security is enabled, these fields will be populated automatically with the details of the user who is logged in

When you use the Matrix Science public web site, you have to supply a name and email address. This is to allow the results of a search to be returned by email. Usually, search results are returned promptly to your browser window. However, if your connection to the web site is broken before the search is complete, they will be emailed to the supplied address. If you have an in-house server, you can enable this if you wish. It is turned off by default

The search title is free text. You don’t have to enter anything. However, it is a good idea to fill in all of these fields, because it makes it much easier to find your old search results in the search log.



Database

PMF✓ SQ✓ MS/MS✓

Database(s)

- MSIPI_mouse
- NCBInr
- SwissProt
- Trembl
- UniRef100

Choose the right database

- Swiss-Prot good for PMF and MS/MS of well characterised organisms
- NCBInr or UniRef100 if you want to search all known protein sequences
- ESTs for MS/MS if genome not sequenced

MASCOT : *Search Parameters*

MATRIX SCIENCE

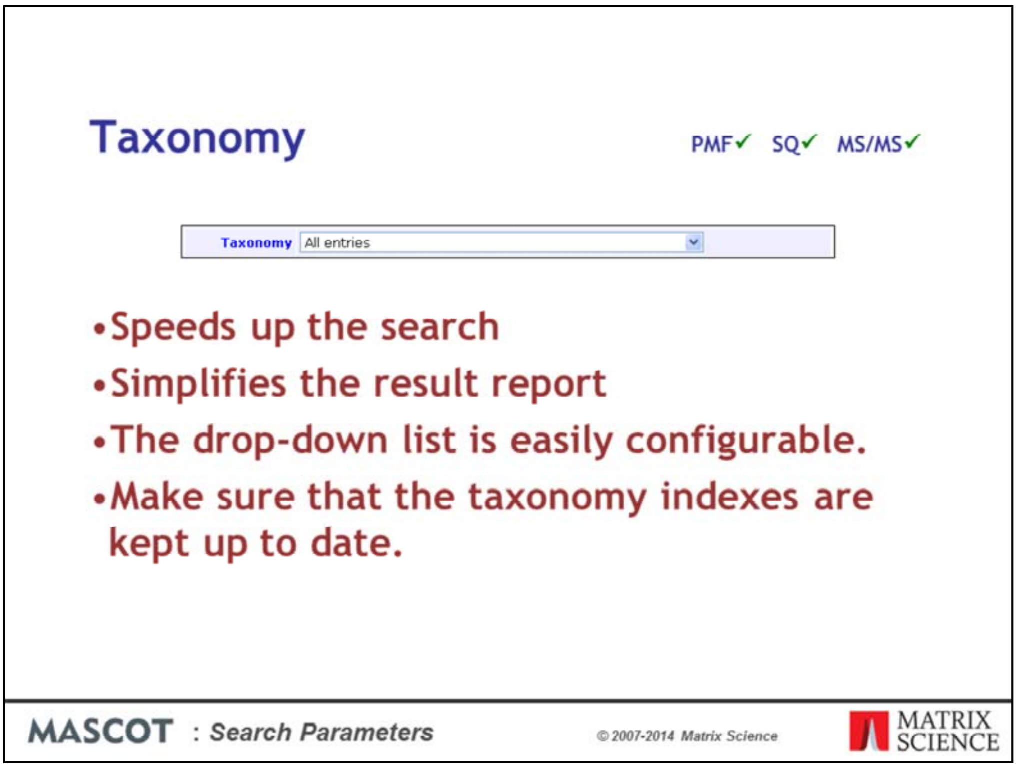
Choosing the right database is so important that there will be a complete presentation on this topic.

Very briefly, for a peptide mass fingerprint, search a comprehensive, non-redundant database, like SwissProt. If the data are any good, it won’t matter if one or two mass values fail to find matches. The advantage of searching a small database is that the search is fast and the reports are concise.

For MS/MS of a well characterised organism, such as human or mouse or yeast, SwissProt is still a good choice. In other cases, search a comprehensive, non-identical database, where every single peptide is explicitly represented, such as NCBInr or UniRef100.

If the genome of your organism of interest has not been sequenced, it won’t be represented in the protein databases, but there may be lots of Expressed Sequence Tags (ESTs). Not advisable for PMF, because many sequences correspond to protein fragments.

In Mascot 2.3 and later, you can select multiple databases for a search. This is particularly useful when you want to search a single organism database and include the sequences of common contaminants, such as BSA and trypsin. One restriction is that you cannot mix AA and DNA databases.



Taxonomy

PMF✓ SQ✓ MS/MS✓

Taxonomy: All entries

- Speeds up the search
- Simplifies the result report
- The drop-down list is easily configurable.
- Make sure that the taxonomy indexes are kept up to date.

MASCOT : *Search Parameters*

© 2007-2014 Matrix Science

MATRIX SCIENCE

If a database contains taxonomy information, we can use this to restrict the search to entries for a particular organism or family. This speeds up the search because, in effect, it makes the database smaller.

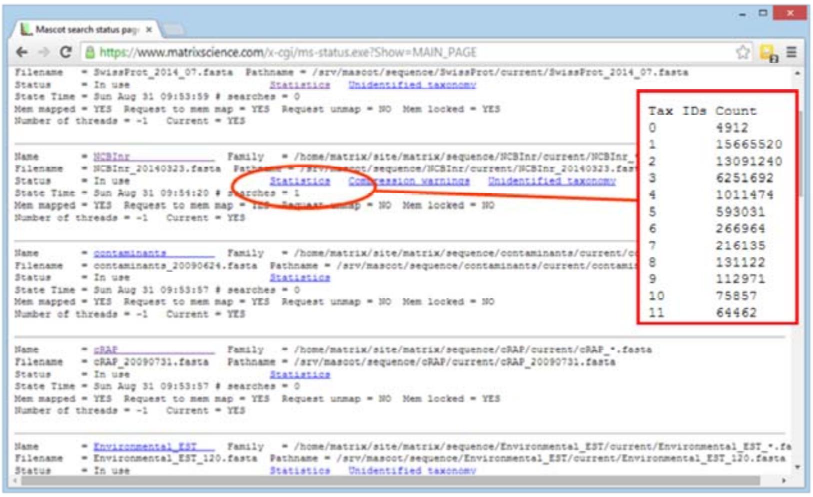
Limiting the taxonomy simplifies the result report, because you don’t see all the homologous proteins from other species.

The drop down list in the search form is configurable. If you are working on a particular organism, you can easily add this to the list

It is important that the taxonomy is as accurate as possible, which means keeping the indexes up to date

Taxonomy

PMF✓ SQ✓ MS/MS✓



Mascot search status page ×
https://www.matrixscience.com/x-cgi/ms-status.exe?Show=MAIN_PAGE

```
Filename   = SwissProt_2014_07.fasta  Pathname = /srv/mascot/sequence/SwissProt/current/SwissProt_2014_07.fasta
Status     = In use                   Statistics   Unidentified taxonomy
State Time = Sun Aug 31 09:53:59 # searches = 0
Mem mapped = YES  Request to mem map = YES  Request unmap = NO  Mem locked = YES
Number of threads = -1   Current = YES

Name       = NCBInr                Family   = /home/matrix/site/matrix/sequence/NCBInr/current/NCBInr_
Filename   = NCBInr_20140323.fasta  Pathname = /srv/mascot/sequence/NCBInr/current/NCBInr_20140323.fas
Status     = In use                   Statistics   Compression warnings   Unidentified taxonomy
State Time = Sun Aug 31 09:54:20 # searches = 1
Mem mapped = YES  Request to mem map = YES  Request unmap = NO  Mem locked = NO
Number of threads = -1   Current = YES

Name       = contaminants          Family   = /home/matrix/site/matrix/sequence/contaminants/current/c
Filename   = contaminants_20090624.fasta  Pathname = /srv/mascot/sequence/contaminants/current/contami
Status     = In use                   Statistics
State Time = Sun Aug 31 09:53:57 # searches = 0
Mem mapped = YES  Request to mem map = YES  Request unmap = NO  Mem locked = NO
Number of threads = -1   Current = YES

Name       = cRAP                  Family   = /home/matrix/site/matrix/sequence/cRAP/current/cRAP_*.fasta
Filename   = cRAP_20090731.fasta   Pathname = /srv/mascot/sequence/cRAP/current/cRAP_20090731.fasta
Status     = In use                   Statistics
State Time = Sun Aug 31 09:53:57 # searches = 0
Mem mapped = YES  Request to mem map = YES  Request unmap = NO  Mem locked = YES
Number of threads = -1   Current = YES

Name       = Environmental_EST     Family   = /home/matrix/site/matrix/sequence/Environmental_EST/current/Environmental_EST_*.fa
Filename   = Environmental_EST_120.fasta  Pathname = /srv/mascot/sequence/Environmental_EST/current/Environmental_EST_120.fasta
Status     = In use                   Statistics   Unidentified taxonomy
```

```
Tax IDs Count
0         4912
1         15665520
2         13091240
3         6251692
4         1011474
5         593031
6         266964
7         216135
8         131122
9         112971
10        75857
11        64462
```

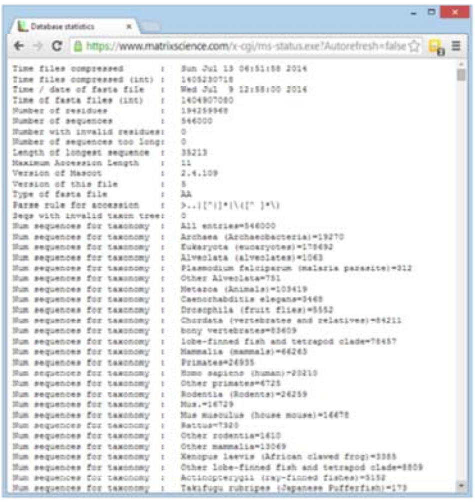
From time to time, its a good idea to go to the database status page and check the stats file for each database. The stats file contains lots of useful information, like whether entries contain illegal characters or whether an entry is too long.

It also tells you how good your taxonomy is. Here are the numbers for the nr database on our web site in March 2014. There are 38 million entries, and 4912 have no taxonomy. In other words, 99.99% of the entries have a taxonomy assigned. If you look at your stats file and see that (say) 10% of the entries have no taxonomy, that's 10% of the entries that are going to be missed whenever you do a search with taxonomy specified.

Taxonomy

PMF✓ SQ✓ MS/MS✓

In most cases, if the correct protein is not in the database, you'd like to see the closest match … whatever the species


Database statistics
https://www.matrixscience.com/x-cgi/ms-status.exe?Autorefresh=false

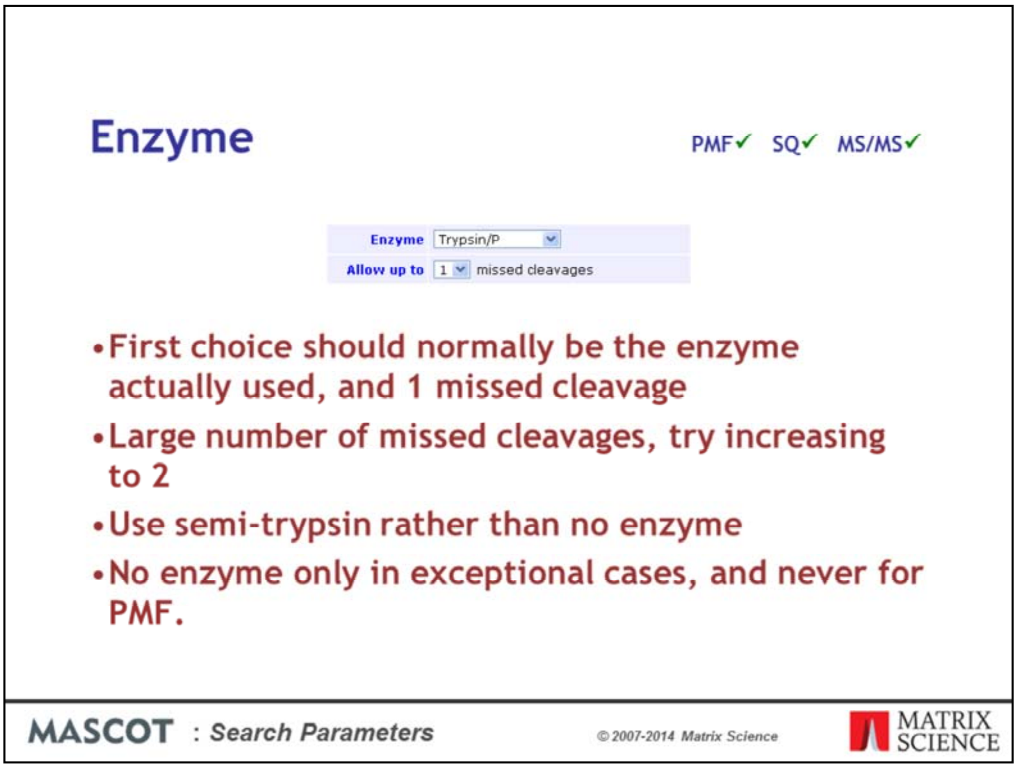
```
Time files compressed        :   Sun Jul 13 06:51:58 2014
Time files compressed (int)  :   1405230718
Time / date of fasta file    :   Wed Jul  9 12:58:00 2014
Time of fasta files (int)    :   1404907080
Number of residues           :   194259968
Number of sequences          :   546000
Number with invalid residues:    0
Number of sequences too long:    0
Length of longest sequence   :   35213
Maximum Accession Length     :   11
Version of Mascot            :   2.4.109
Version of this file         :   5
Type of fasta file           :   AA
Parse rule for accession     :   >..|[^|]*|\([^ ]*\)
Seqs with invalid taxon tree:    0
Num sequences for taxonomy   :   All entries=546000
Num sequences for taxonomy   :   Archaea (Archaeobacteria)=19270
Num sequences for taxonomy   :   Eukaryota (eucaryotes)=178692
Num sequences for taxonomy   :   Alveolata (alveolates)=1063
Num sequences for taxonomy   :   Plasmodium falciparum (malaria parasite)=312
Num sequences for taxonomy   :   Other Alveolata=751
Num sequences for taxonomy   :   Metazoa (Animals)=103419
Num sequences for taxonomy   :   Caenorhabditis elegans=3468
Num sequences for taxonomy   :   Drosophila (fruit flies)=5552
Num sequences for taxonomy   :   Chordata (vertebrates and relatives)=84211
Num sequences for taxonomy   :   bony vertebrates=83609
Num sequences for taxonomy   :   lobe-finned fish and tetrapod clade=78457
Num sequences for taxonomy   :   Mammalia (mammals)=66263
Num sequences for taxonomy   :   Primates=26935
Num sequences for taxonomy   :   Homo sapiens (human)=20210
Num sequences for taxonomy   :   Other primates=6725
Num sequences for taxonomy   :   Rodentia (Rodents)=26259
Num sequences for taxonomy   :   Mus.=16729
Num sequences for taxonomy   :   Mus musculus (house mouse)=16678
Num sequences for taxonomy   :   Rattus=7920
Num sequences for taxonomy   :   Other rodentia=1610
Num sequences for taxonomy   :   Other mammalia=13069
Num sequences for taxonomy   :   Xenopus laevis (African clawed frog)=3385
Num sequences for taxonomy   :   Other lobe-finned fish and tetrapod clade=8809
Num sequences for taxonomy   :   Actinopterygii (ray-finned fishes)=5152
Num sequences for taxonomy   :   Takifugu rubripes (Japanese Pufferfish)=173
```

A word of warning. Don't specify a very narrow taxonomy in a search.

Think carefully about what you are trying to achieve when you do this.

If the correct protein from the correct species is not in the database, wouldn't you want to see a good match to a protein from a similar species?

This is especially important for poorly represented species. For example, look at these numbers for the Swiss-Prot 2014_07: half a million entries; 26 thousand entries for rodents, but only 1600 are not either mouse or rat. So, even if you are studying hamster or porcupine, you don't want to choose 'Other rodentia'.



Enzyme

PMF✓ SQ✓ MS/MS✓

Enzyme Trypsin/P
Allow up to 1 missed cleavages

- First choice should normally be the enzyme actually used, and 1 missed cleavage
- Large number of missed cleavages, try increasing to 2
- Use semi-trypsin rather than no enzyme
- No enzyme only in exceptional cases, and never for PMF.

MASCOT : Search Parameters © 2007-2014 Matrix Science MATRIX SCIENCE

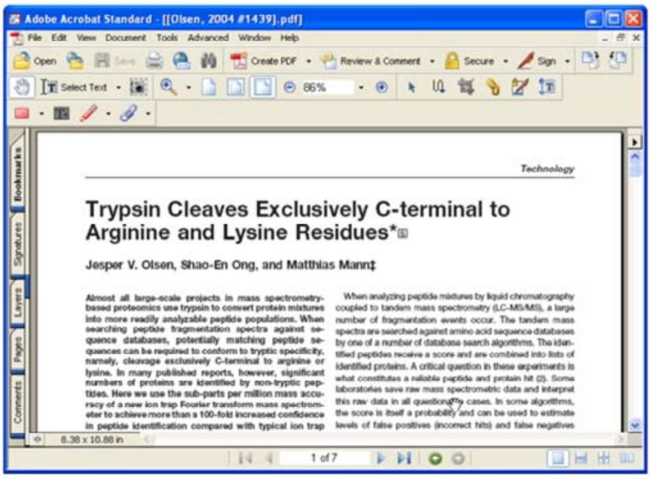
All the search forms have a drop down list for choosing an enzyme. If your peptides come from an enzyme digest, you need to know what the enzyme was and then choose it from the list.

Setting the number of allowed missed cleavage sites to zero simulates a limit digest. If you are confident that your digest is perfect, with no partial fragments present, this will give maximum discrimination and the highest score for a peptide mass fingerprint.

If experience shows that your digest mixtures usually include some partials, that is, peptides with missed cleavage sites, you should choose a setting of 1, or maybe 2 missed cleavage sites. Don't specify a higher number without good reason, because each additional level of missed cleavages increases the number of calculated peptide masses to be matched against the experimental data. In other words, the missed cleavage parameter should set by looking at some successful search results to see how complete your digests really are.

Enzyme

PMF✓ SQ✓ MS/MS✓



Technology

Trypsin Cleaves Exclusively C-terminal to Arginine and Lysine Residues*

Jesper V. Olsen, Shao-En Ong, and Matthias Mann‡

Almost all large-scale projects in mass spectrometry-based proteomics use trypsin to convert protein mixtures into more readily analyzable peptide populations. When searching peptide fragmentation spectra against sequence databases, potentially matching peptide sequences can be required to conform to tryptic specificity, namely, cleavage exclusively C-terminal to arginine or lysine. In many published reports, however, significant numbers of proteins are identified by non-tryptic peptides. Here we use the sub-parts per million mass accuracy of a new ion trap Fourier transform mass spectrometer to achieve more than a 100-fold increased confidence in peptide identification compared with typical ion trap

When analyzing peptide mixtures by liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS), a large number of fragmentation events occur. The tandem mass spectra are searched against amino acid sequence databases by one of a number of database search algorithms. The identified peptides receive a score and are combined into lists of identified proteins. A critical question in these experiments is what constitutes a reliable peptide and protein hit (2). Some laboratories save raw mass spectrometric data and interpret this raw data in all questionable cases. In some algorithms, the score is itself a probability and can be used to estimate levels of false positives (incorrect hits) and false negatives

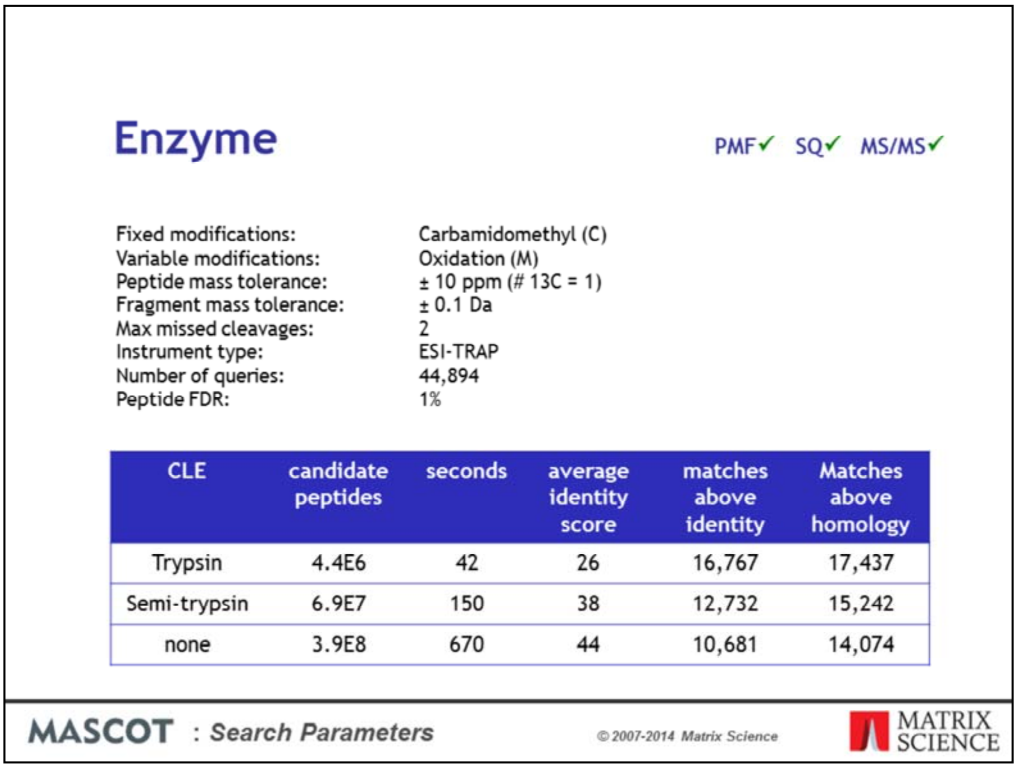
Olsen, J. V., Ong, S.-E. and Mann, M., Mol. and Cellular Proteomics, 3, 608-14 (2004)

Although some people like to perform searches without enzyme specificity, and then gain confidence that a match is correct if the match is tryptic, this isn’t a good idea. If there is evidence for a lot of non-specific cleavage, then a semi-specific enzyme allows one end of the peptide to be non-specific, but not both. Only abandon enzyme specificity if you have no other choice, such as when searching endogenous peptides.

You cannot perform a no-enzyme peptide mass fingerprint. It simply won’t work, even if you have good mass accuracy

There is some controversy over the level of non-specific peptides that can be expected in a tryptic digest. Our experience is that the levels of non-specific peptides are very low, less than 3%, unless there is something seriously wrong with the trypsin or the protocol.

Why do we advise so strongly against no-enzyme searches?



Enzyme

PMF✓ SQ✓ MS/MS✓

Fixed modifications:	Carbamidomethyl (C)
Variable modifications:	Oxidation (M)
Peptide mass tolerance:	± 10 ppm (# 13C = 1)
Fragment mass tolerance:	± 0.1 Da
Max missed cleavages:	2
Instrument type:	ESI-TRAP
Number of queries:	44,894
Peptide FDR:	1%

CLE	candidate peptides	seconds	average identity score	matches above identity	Matches above homology
Trypsin	4.4E6	42	26	16,767	17,437
Semi-trypsin	6.9E7	150	38	12,732	15,242
none	3.9E8	670	44	10,681	14,074

MASCOT : *Search Parameters* © 2007-2014 Matrix Science MATRIX SCIENCE

Here are some numbers for an Orbitrap dataset when we search using strict trypsin, semi-specific trypsin, and no enzyme specificity

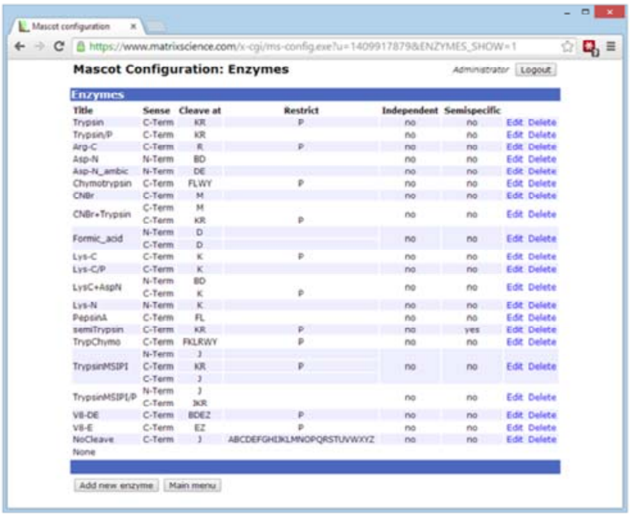
As you can see, the no enzyme search takes a lot longer and we get fewer reliable matches.

The reason is simple, the search space for a no-enzyme search is much, much larger than for a tryptic search. This means that the thresholds are higher and we lose marginal matches. Unless you have a high level of non-specific peptides, you lose more than you gain.

So, doing a no-enzyme search in Mascot is not a good idea unless there is a very high level of non-specific peptides. Semi-trypsin will be a better choice if the peptides came from a tryptic digest but there is a high level of non-specific cleavage. Only use no enzyme if the peptides are not the products of an enzyme digest, e.g. MHC peptides or endogenous peptides.

Enzyme

PMF✓ SQ✓ MS/MS✓


Mascot configuration
https://www.matrixscience.com/x-cgi/ms-config.exe?u=1409917879&ENZYMES_SHOW=1

Mascot Configuration: Enzymes

Enzymes

Title	Sense	Cleave at	Restrict	Independent	Semispecific	
Trypsin	C-Term	KR	P	no	no	Edit Delete
Trypsin/P	C-Term	KR		no	no	Edit Delete
Arg-C	C-Term	R	P	no	no	Edit Delete
Asp-N	N-Term	BD		no	no	Edit Delete
Asp-N_ambic	N-Term	DE		no	no	Edit Delete
Chymotrypsin	C-Term	FLWY	P	no	no	Edit Delete
CNBr	C-Term	M		no	no	Edit Delete
CNBr+Trypsin	C-Term	M		no	no	Edit Delete
	C-Term	KR	P			
Formic_acid	N-Term	D		no	no	Edit Delete
	C-Term	D				
Lys-C	C-Term	K	P	no	no	Edit Delete
Lys-C/P	C-Term	K		no	no	Edit Delete
LysC+AspN	N-Term	BD		no	no	Edit Delete
	C-Term	K	P			
Lys-N	N-Term	K		no	no	Edit Delete
PepsinA	C-Term	FL		no	no	Edit Delete
semiTrypsin	C-Term	KR	P	no	yes	Edit Delete
TrypChymo	C-Term	FKLRWY	P	no	no	Edit Delete
TrypsinMSIPI	N-Term	J		no	no	Edit Delete
	C-Term	KR	P			
	C-Term	J				
TrypsinMSIPI/P	N-Term	J		no	no	Edit Delete
	C-Term	JKR				
V8-DE	C-Term	BDEZ	P	no	no	Edit Delete
V8-E	C-Term	EZ	P	no	no	Edit Delete
NoCleave	C-Term	J	ABCDEFGHIJKLMNOPQRSTUVWXYZ	no	no	Edit Delete
None						

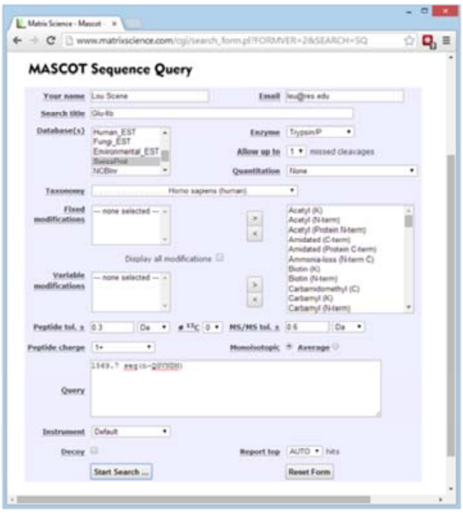
The list of enzymes is user configurable. Standard entries are described in the help. If you wish, you can modify the definitions or create new ones using the configuration editor.

Mascot supports two categories of mixed enzyme definitions. An independent mixed enzyme is used where multiple sample aliquots have been digested separately, and the digests combined for analysis. This means that the sample could contain (say) tryptic peptides and Asp-N peptides, but no peptides that are tryptic at one end and Asp-N at the other. The second category simulates a single sample aliquot being digested simultaneously or serially by more than one cleavage agent. For example CNBr followed by trypsin.

Remember that enzyme type None simulates cleavage at every peptide bond. For top down searches, where you don't want any cleavage, choose NoCleave.

Enzyme

PMF✓ SQ✓ MS/MS✓



Matrix Science - Mascot

www.matrixscience.com/cgi/search_form.pl?FORMVER=2&SEARCH=SQ

MASCOT Sequence Query

Your name	Lou Scene	Email	lou@res.edu
Search title	Glu-fib		
Database(s)	Human_EST Fungi_EST Environmental_EST SwissProt NCBInr	Enzyme	Trypsin/P
		Allow up to	1 missed cleavages
		Quantitation	None
Taxonomy	Homo sapiens (human)		
Fixed modifications	--- none selected ---		Acetyl (K) Acetyl (N-term) Acetyl (Protein N-term) Amidated (C-term) Amidated (Protein C-term) Ammonia-loss (N-term C) Biotin (K) Biotin (N-term) Carbamidomethyl (C) Carbamyl (K) Carbamyl (N-term)
	Display all modifications ☐		
Variable modifications	--- none selected ---		
Peptide tol. ±	0.3 Da	# ^{13}C 0	MS/MS tol. ± 0.6 Da
Peptide charge	1+	Monoisotopic ● Average ○	
Query	1569.7 seq(n-[illegible])		
Instrument	Default		
Decoy	☐	Report top	AUTO hits

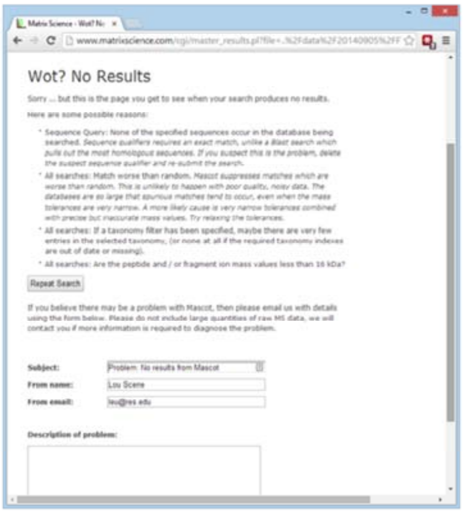
Remember that enzyme specificity also applies to Sequence Queries

Quite often, we receive a support email along the lines of "Mascot is broken. I did a search for this peptide and I know its in the database but Mascot failed to find it“

For example, here's a search for glu-fib, a very common sequencing standard. The mass is correct and the sequence is correct. But, when we do a search of SwissProt …

Enzyme

PMF✓ SQ✓ MS/MS✓



Matrix Science - Wot? No
www.matrixscience.com/cgi/master_results.pl?file=../%2Fdata%2F20140905%2FF

Wot? No Results

Sorry ... but this is the page you get to see when your search produces no results.

Here are some possible reasons:

- Sequence Query: None of the specified sequences occur in the database being searched. *Sequence qualifiers requires an exact match, unlike a Blast search which pulls out the most homologous sequences. If you suspect this is the problem, delete the suspect sequence qualifier and re-submit the search.*
- All searches: Match worse than random. *Mascot suppresses matches which are worse than random. This is unlikely to happen with poor quality, noisy data. The databases are so large that spurious matches tend to occur, even when the mass tolerances are very narrow. A more likely cause is very narrow tolerances combined with precise but inaccurate mass values. Try relaxing the tolerances.*
- All searches: If a taxonomy filter has been specified, maybe there are very few entries in the selected taxonomy, (or none at all if the required taxonomy indexes are out of date or missing).
- All searches: Are the peptide and / or fragment ion mass values less than 16 kDa?

If you believe there may be a problem with Mascot, then please email us with details using the form below. Please do not include large quantities of raw MS data, we will contact you if more information is required to diagnose the problem.

Subject:	Problem: No results from Mascot
From name:	Lou Scene
From email:	lou@res.edu

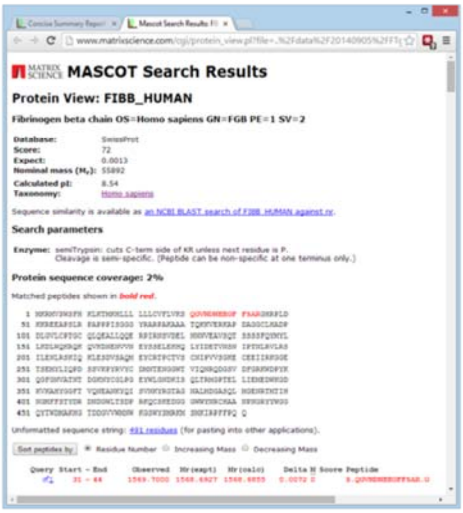
Description of problem:

No results!

Why?

Enzyme

PMF✓ SQ✓ MS/MS✓



Concise Summary Report
Mascot Search Results
www.matrixscience.com/cgi/protein_view.pl?file=../data%2F20140905%2FFT

MASCOT Search Results

Protein View: FIBB_HUMAN

Fibrinogen beta chain OS=Homo sapiens GN=FGB PE=1 SV=2

Database:	SwissProt
Score:	72
Expect:	0.0013
Nominal mass (M_r):	55892
Calculated pI:	8.54
Taxonomy:	Homo sapiens

Sequence similarity is available as an NCBI BLAST search of FIBB_HUMAN against nr.

Search parameters

Enzyme: semiTrypsin: cuts C-term side of KR unless next residue is P.
Cleavage is semi-specific. (Peptide can be non-specific at one terminus only.)

Protein sequence coverage: 2%

Matched peptides shown in ***bold red***.

```
  1 MKRMVSWSFH KLKTMKHLLL LLLCVFLVKS QGVNDNEEGF FSARGHRPLD
 51 KKREEAPSLR PAPPPISGGG YRARPAKAAA TQKKVERKAP DAGGCLHADP
101 DLGVLCPTGC QLQEALLQQE RPIRNSVDEL NNNVEAVSQT SSSSFQYMYL
151 LKDLWQKRQK QVKDNENVVN EYSSELEKHQ LYIDETVNSN IPTNLRVLRS
201 ILENLRSKIQ KLESDVSAQM EYCRTPCTVS CNIPVVSGKE CEEIIRKGGE
251 TSEMYLIQPD SSVKPYRVYC DMNTENGGWT VIQNRQDGSV DFGRKWDPYK
301 QGFGNVATNT DGKNYCGLPG EYWLGNDKIS QLTRMGPTEL LIEMEDWKGD
351 KVKAHYGGFT VQNEANKYQI SVNKYRGTAG NALMDGASQL MGENRTMTIH
401 NGMFFSTYDR DNDGWLTSDP RKQCSKEDGG GWWYNRCHAA NPNGRYYWGG
451 QYTWDMAKHG TDDGVVWMNW KGSWYSMRKM SMKIRPFFPQ Q
```

Unformatted sequence string: 491 residues (for pasting into other applications).

Sort peptides by ● Residue Number ○ Increasing Mass ○ Decreasing Mass

```
Query  Start - End    Observed  Mr(expt)  Mr(calc)    Delta M Score Peptide
          31 - 44    1569.7000 1568.6927 1568.6855  0.0072 0       S.QGVNDNEEGFFSAR.G
```

MASCOT : *Search Parameters*

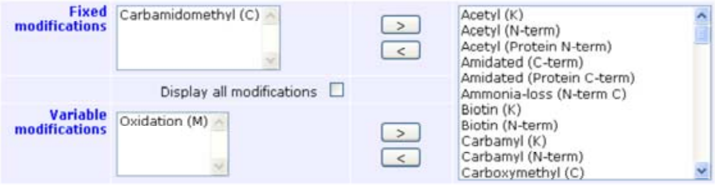
© 2007-2014 Matrix Science

MATRIX SCIENCE

Because glu-fib in SwissProt is not a tryptic peptide. The N-terminus is created by a post-translational cleavage after serine. If you now go back to the search form and select semi-trypsin or enzyme type none, you'll get the match.

Modifications

PMF✓ SQ✓ MS/MS✓



Fixed modifications: Carbamidomethyl (C)

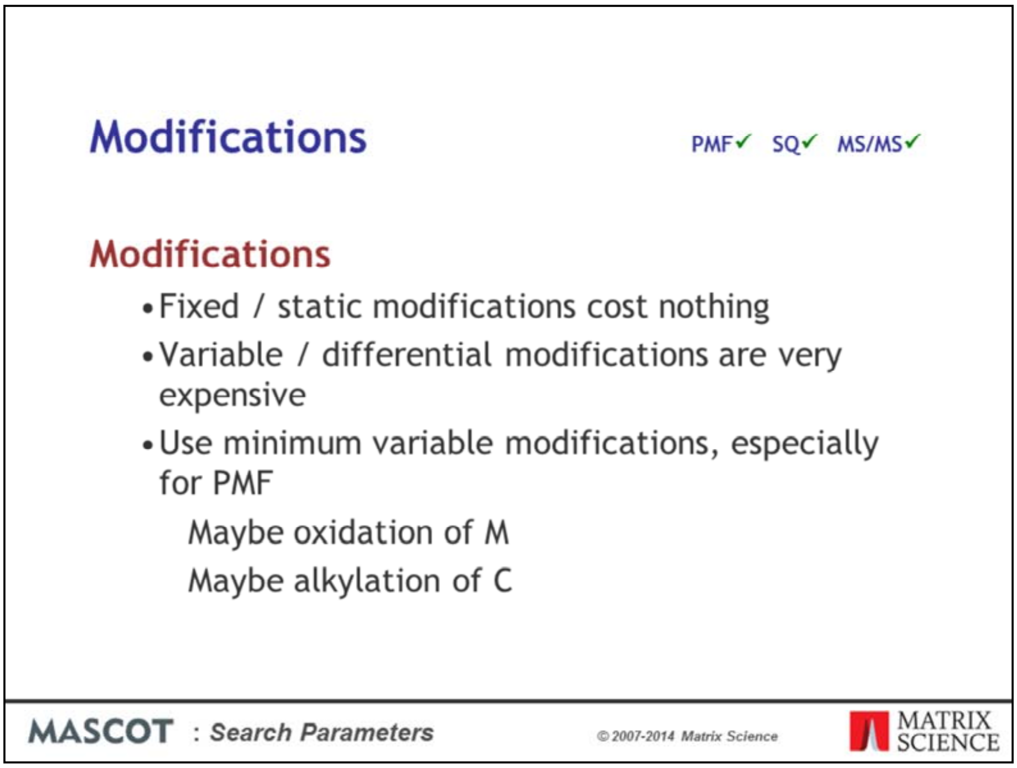
Display all modifications ☐

Variable modifications: Oxidation (M)

- Acetyl (K)
- Acetyl (N-term)
- Acetyl (Protein N-term)
- Amidated (C-term)
- Amidated (Protein C-term)
- Ammonia-loss (N-term C)
- Biotin (K)
- Biotin (N-term)
- Carbamyl (K)
- Carbamyl (N-term)
- Carboxymethyl (C)

- Get details of current modifications, download updates, and define new entries at http://www.unimod.org
- User definable with an in-house Mascot installation

This screen shot shows how modifications are displayed in the search form in Mascot 2.3 and later. If you are using an earlier version, there are just two list boxes, one for fixed modifications and one for variable. In the current arrangement, you move modifications from the single list on the right to and from the lists on the left. This makes it easier to see at a glance what has been selected for the search. If the checkbox labelled ‘Display all modifications’ is clear, as shown here, you get a relatively short list of the most common modifications. If you check the box, a much longer list is available. You can keep your list of modifications up-to-date by downloading the latest information from Unimod. If you have a modification which you don’t want to share with others, you can add it to the local configuration file. We’ll describe how to go about doing this in detail in the Mascot Server Administration talk.



Modifications

PMF✓ SQ✓ MS/MS✓

Modifications

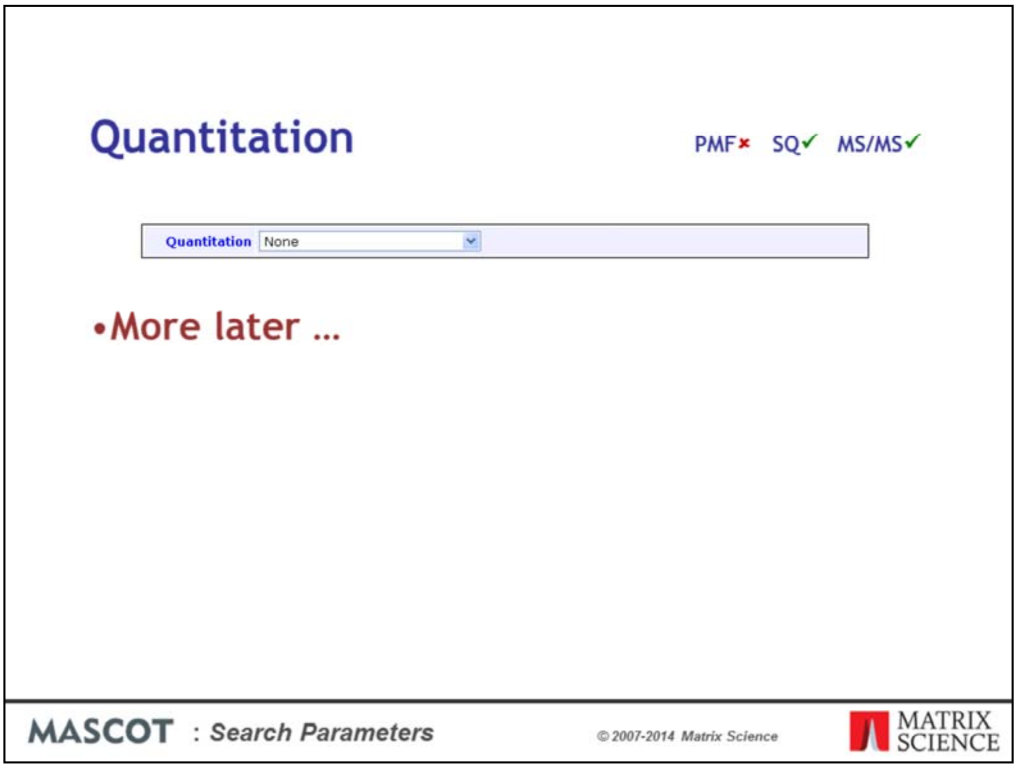
- Fixed / static modifications cost nothing
- Variable / differential modifications are very expensive
- Use minimum variable modifications, especially for PMF
  - Maybe oxidation of M
  - Maybe alkylation of C

MASCOT : *Search Parameters* © 2007-2014 Matrix Science MATRIX SCIENCE

Modifications in database searching are handled in two ways. First, there are the fixed or static or quantitative modifications. An example would be a the efficient alkylation of cysteine. Since all cysteines are modified, this is effectively just a change in the mass of cysteine. It carries no penalty in terms of search speed or specificity.

In contrast, most post-translational modifications do not apply to all instances of a residue. For example, phosphorylation might affect just one serine in a peptide containing many serines. These variable or differential or non-quantitative modifications are expensive in the sense that they increase the time taken for a search and reduce its specificity. This is because the software has to permute out all the possible arrangements of modified and unmodified residues that fit to the peptide molecular mass. As more and more modifications are considered, the number of combinations and permutations increases geometrically. A so-called combinatorial explosion.

Hence, it is very important to be as sparing as possible with variable modifications. Especially in a peptide mass fingerprint, where the increase in the number of calculated peptides quickly makes it impossible to find a statistically significant match.



Quantitation

PMF✗ SQ✓ MS/MS✓

Quantitation: None

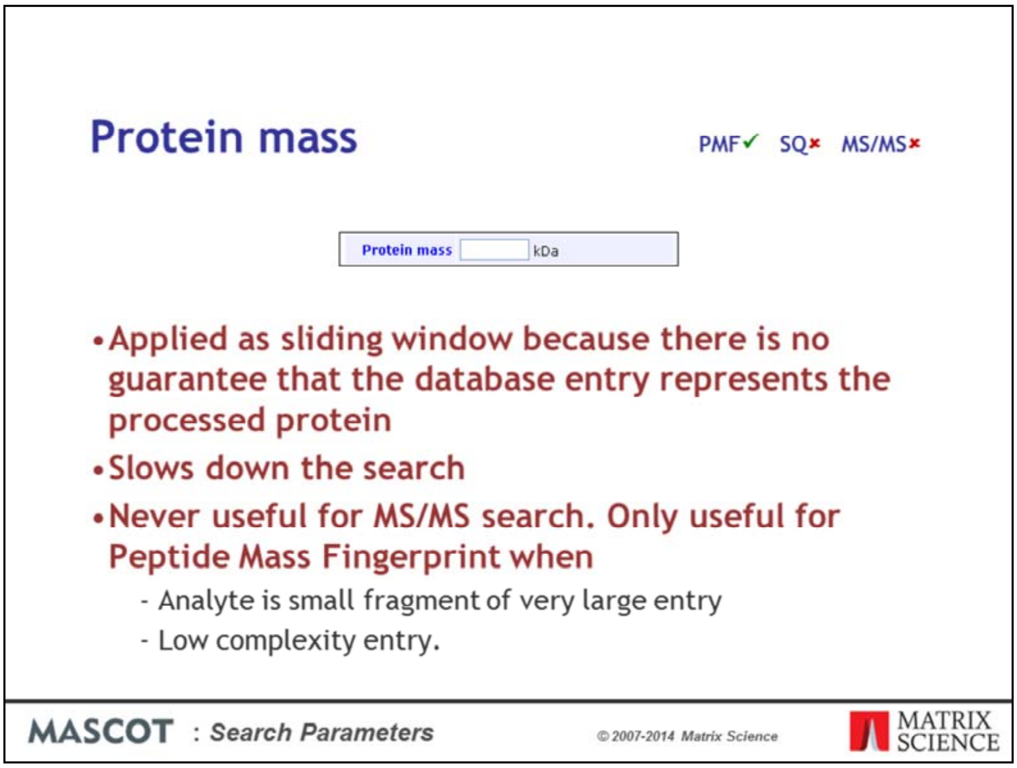
- More later …

MASCOT : Search Parameters

© 2007-2014 Matrix Science

MATRIX SCIENCE

Quantitation is the subject of a separate presentation.



Protein mass

PMF✓ SQ✗ MS/MS✗

Protein mass ▭ kDa

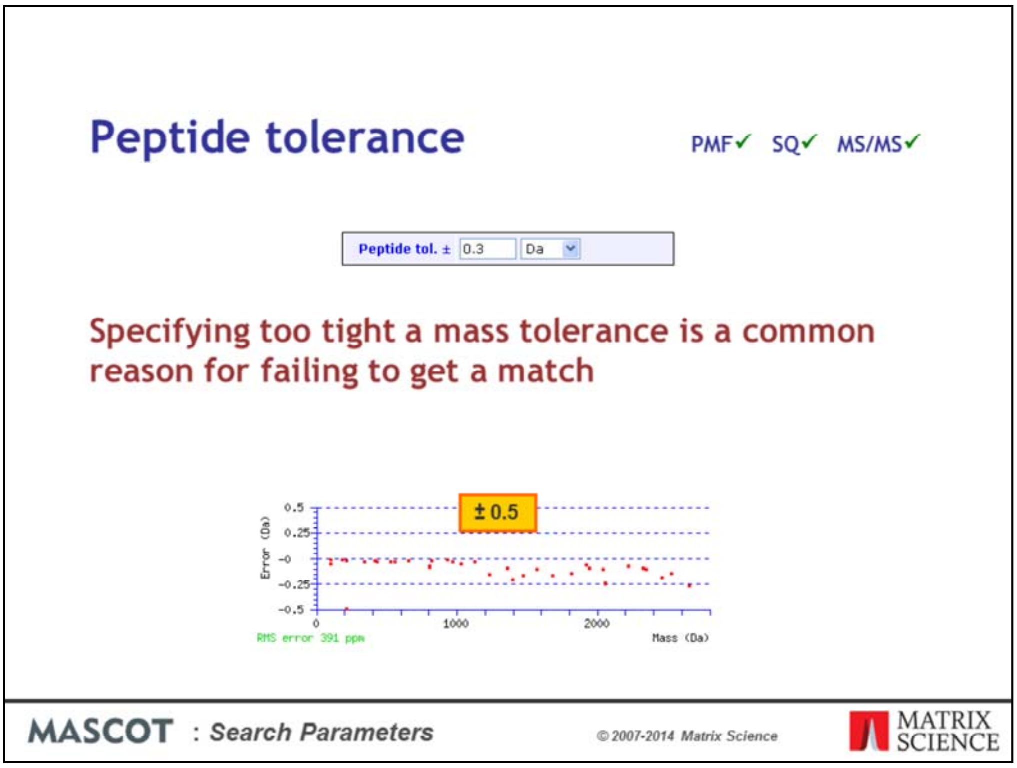
- Applied as sliding window because there is no guarantee that the database entry represents the processed protein
- Slows down the search
- Never useful for MS/MS search. Only useful for Peptide Mass Fingerprint when
  - Analyte is small fragment of very large entry
  - Low complexity entry.

The protein mass is the mass of the intact protein in kDa applied as a sliding window. That is, the mass of the contiguous stretch of sequence which contains all of the matched peptide mass values. This will generally be less than the mass of the entire sequence entry. Consequently, if you specify a value for the protein mass, this acts only as a ceiling. Not only will you see smaller proteins on the hit list, you will also see larger ones, but all of the reported matches will be within a stretch of sequence less than or equal to the specified mass.

If this field is left blank, there is no restriction on protein mass

Specifying a protein mass will slow down the search a little.

Its hard to find examples where this parameter is useful. We include it mainly because many people requested it. It could give a better score if the analyte was small fragment of very large entry, or a low complexity protein. But, you can’t know this in advance, so our general recommendation is to leave the protein mass open



Peptide tolerance

PMF✓ SQ✓ MS/MS✓

Peptide tol. ± 0.3 Da

Specifying too tight a mass tolerance is a common reason for failing to get a match

MASCOT : *Search Parameters*

© 2007-2014 Matrix Science

MATRIX SCIENCE

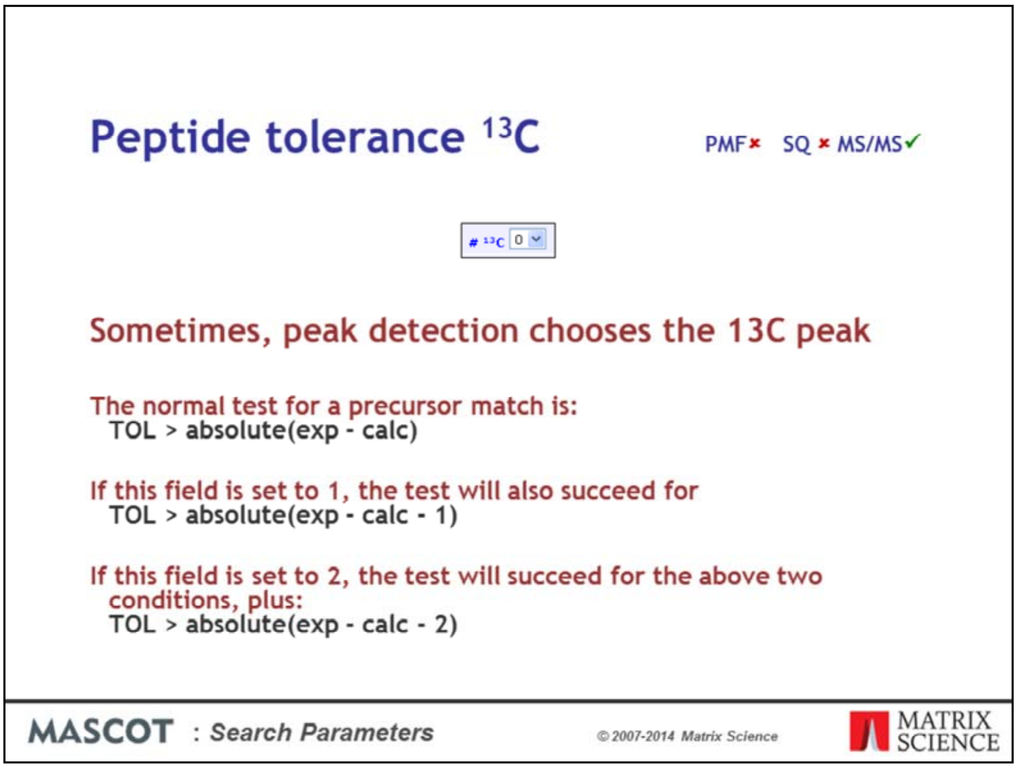
This is the error window on experimental peptide mass values, not the error window for MS/MS fragment ion mass values, which is set using the MS/MS tol. ± parameter.

Units can be selected from: percentage, milli-mass units, parts per million, or Daltons.

Specifying too tight a tolerance is a very common reason for failing to get a match.

Making an estimate of the mass accuracy doesn't have to be a guessing game. Protein View includes a graph of the mass errors for intact peptides. Just search a strong standard and look at the error graph. You'll normally see some kind of trend. Add on a safety margin and this is your error estimate. If you see something that looks like this, a mass tolerance of +/- 0.5 Da is about right. It gives some safety margin. Remember that there will always be the odd outlier, like the data point at the lower left. It is the general trend and distribution of the majority of the data points that is important.

For a peptide mass fingerprint, the score depends on the peptide tolerance. In an MS/MS search, this parameter has no effect on the ions score. However, it does affect the search time. The larger the tolerance, the longer the search will take.



Peptide tolerance ^{13}C

PMF✗ SQ✗ MS/MS✓

Sometimes, peak detection chooses the 13C peak

The normal test for a precursor match is:
TOL > absolute(exp - calc)

If this field is set to 1, the test will also succeed for
TOL > absolute(exp - calc - 1)

If this field is set to 2, the test will succeed for the above two conditions, plus:
TOL > absolute(exp - calc - 2)

MASCOT : *Search Parameters*

© 2007-2014 Matrix Science

MATRIX SCIENCE

Sometimes, peak detection chooses the 13C peak rather than the 12C. In extreme cases, it may pick the 13C2 peak. The normal test for a precursor match is:

TOL > absolute(exp - calc)

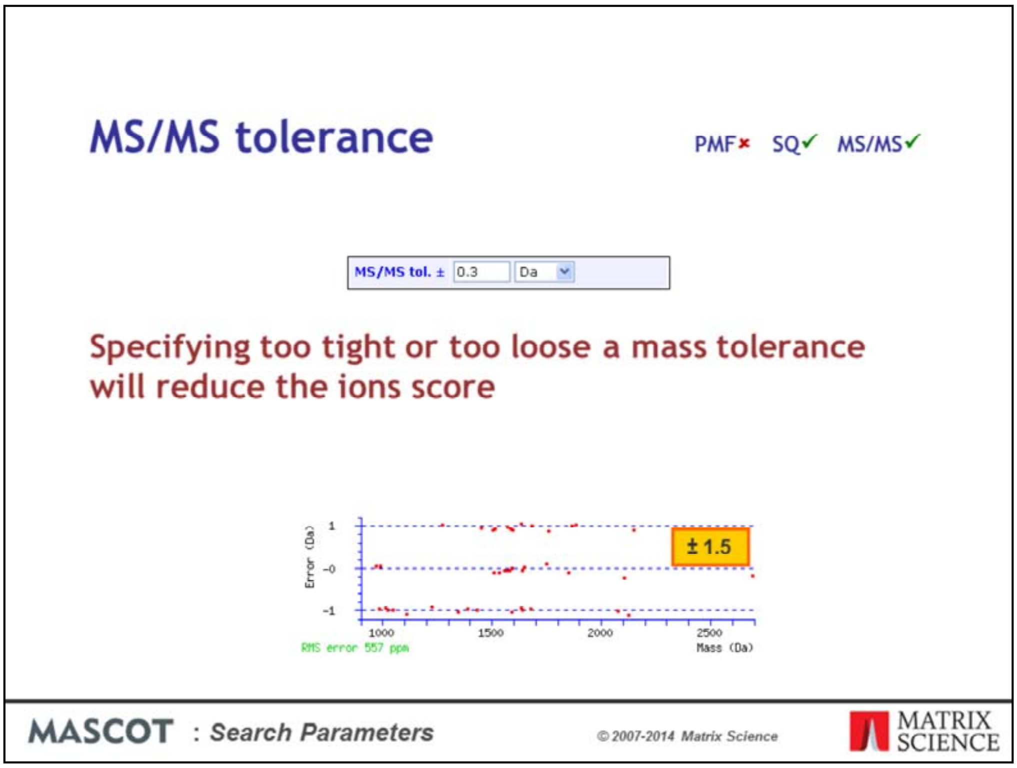
Assuming the mass values and tolerance are in Da, if this field is set to 1, the test will also succeed for

TOL > absolute(exp - calc - 1)

If this field is set to 2, the test will succeed for the above two conditions, plus:

TOL > absolute(exp - calc - 2)

This means that you can use a tight mass tolerance and still get a match to a 13C peak. If you are using a very high accuracy instrument, note that the precise shifts are the carbon isotope spacings of 1.00335 and 2.00670, rather than 1 and 2.



MS/MS tolerance

PMF✗ SQ✓ MS/MS✓

MS/MS tol. ± 0.3 Da

Specifying too tight or too loose a mass tolerance will reduce the ions score

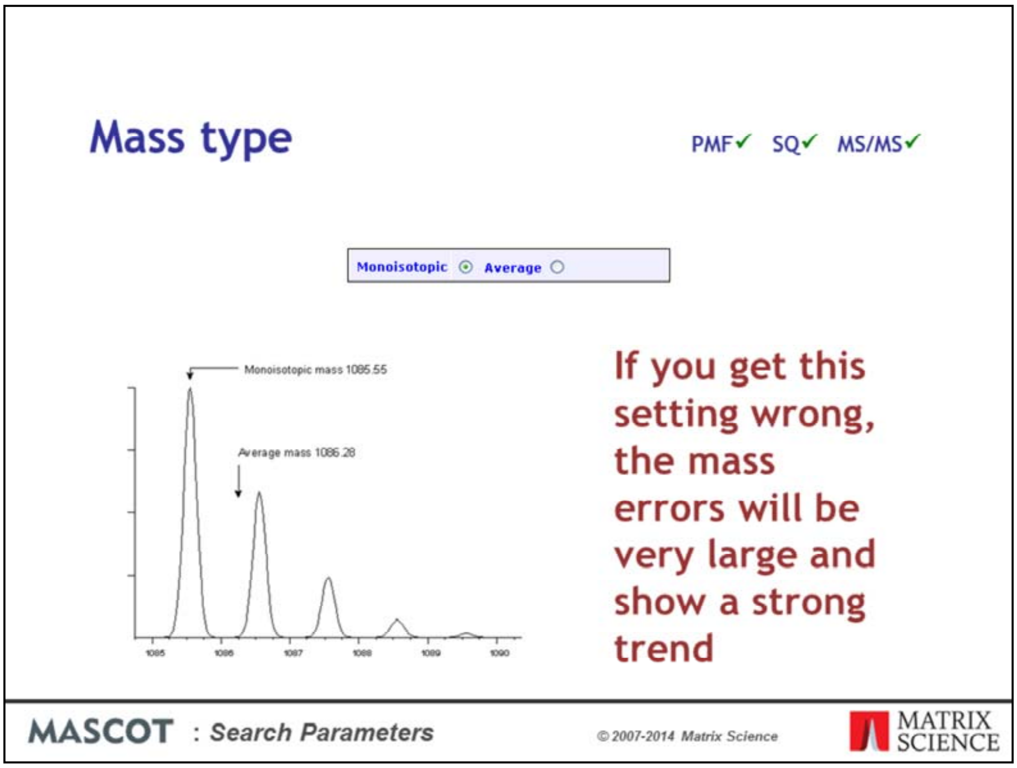
MASCOT : Search Parameters © 2007-2014 Matrix Science MATRIX SCIENCE

This is the error window on MS/MS fragment mass values.

Units can be milli-mass units, Daltons, or ppm (Mascot 2.5 and later).

Specifying too tight or too loose a mass tolerance will reduce the ions score. Peptide View includes a graph of the mass errors for fragment ions.

Here, the mass tolerance is much too high. A more appropriate tolerance might be +/- 0.3. Having a tolerance which is much too high can sometimes lead to artefacts and false positives



Mass type

PMF✓ SQ✓ MS/MS✓

Monoisotopic ● Average ○

If you get this setting wrong, the mass errors will be very large and show a strong trend

MASCOT : Search Parameters

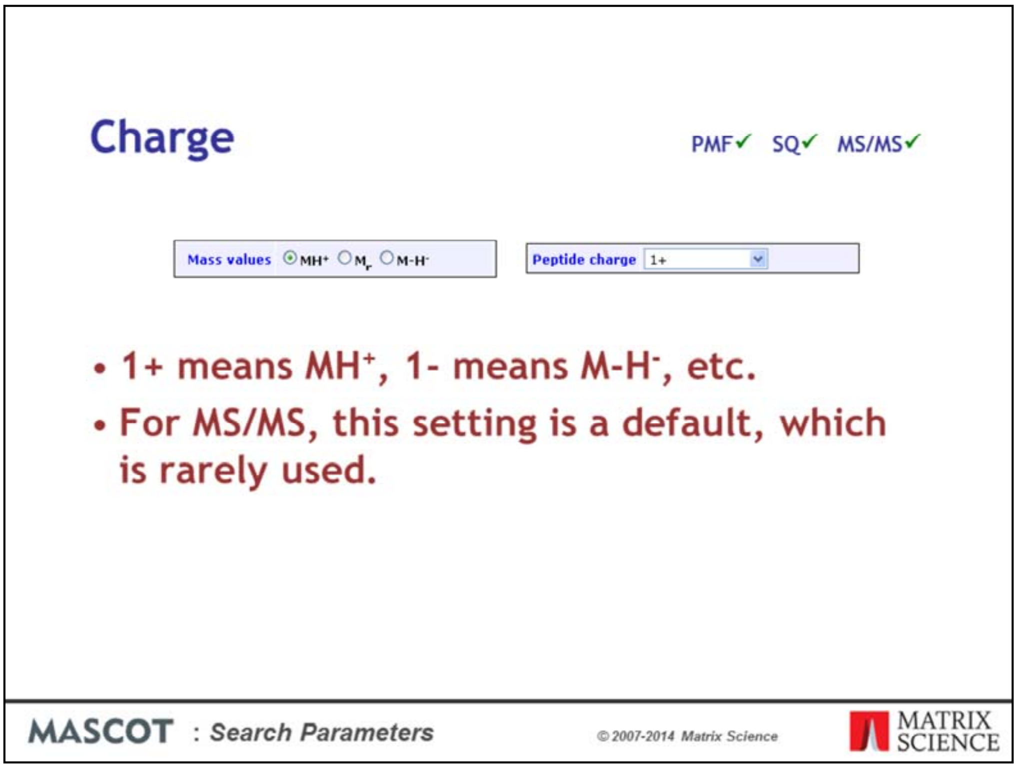
© 2007-2014 Matrix Science

MATRIX SCIENCE

Mass type specifies whether the experimental mass values are average or monoisotopic. Monoisotopic mass is the mass of the peptide where all atoms are the most abundant natural isotopes of their elements, e.g. Carbon 12, Nitrogen 14, Hydrogen 1, etc. In most cases, this is the first peak of the natural isotope distribution. Average mass is the chemical mass, which is the centre of gravity of the isotope distribution.

In Mascot, you cannot mix the two, and have (say) average precursors and monoisotopic fragments.

Most modern instruments produce monoisotopic mass values. You will only have an average mass if the entire isotope distribution has been centroided into a single peak, which usually implies very low resolution. If you get this setting wrong, the mass errors will be very large and show a strong trend, because the difference between an average and a monoisotopic mass for peptides and proteins is approximately 0.06%.



Charge

PMF✓ SQ✓ MS/MS✓

Mass values ⊙ MH^+ ○ M_r ○ $M\text{-}H^-$

Peptide charge 1+

- 1+ means MH^+, 1- means $M\text{-}H^-$, etc.
- For MS/MS, this setting is a default, which is rarely used.

MASCOT : *Search Parameters* © 2007-2014 Matrix Science MATRIX SCIENCE

These fields are used to specify the peptide charge state. The radio buttons are from the peptide mass fingerprint form. The drop down list is used on the sequence query and MS/MS forms.

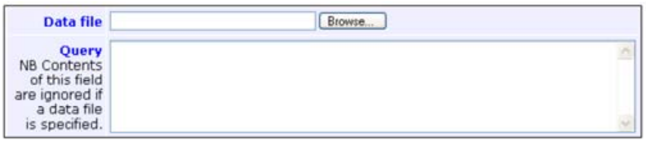
The notation "1+", "2+", etc. is used to save space and because some HTML form fields do not support the use of superscripts and subscripts. "1+" always means MH+, "1-" always means M-H-, etc.

For MALDI-PSD, the precursor peptides will generally be MH+, so the charge state should be set to "1+“

For an MS/MS search, the value specified here is a default. Most peak lists always specify a charge state, so default is never used.

Data (PMF)

PMF✓ SQ✓ MS/MS✗



Data file

Browse...

Query
NB Contents of this field are ignored if a data file is specified.

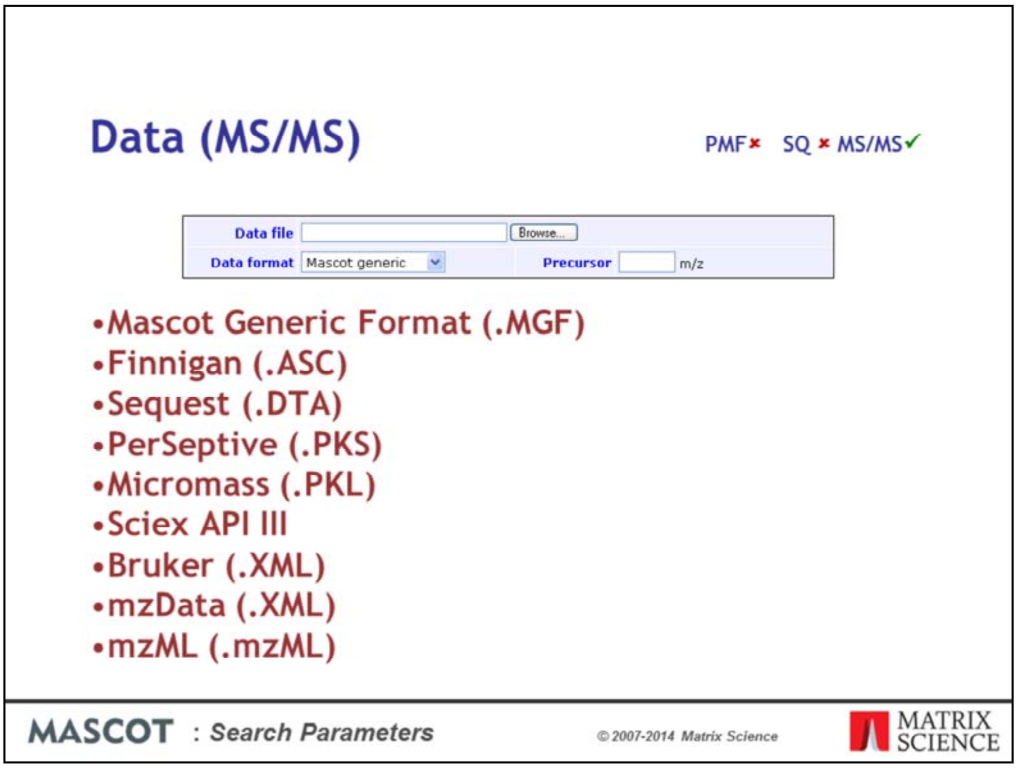
- Mass [intensity] [additional text]
- Applied Biosystems Data Explorer (.pkm)
- Bruker Analysis AutoXecute Data Report
- Bruker XML
- mzData (1.05)
- mzML

The contents of the query window on the peptide mass fingerprint form are only used when no data file has been specified.

The data format for a peptide mass fingerprint is auto detected. It can be a simple list of mass values, one per line. If a second values is present, it is assumed to be intensity. Any further values on the same line are ignored

Mascot also supports other peak list formats, as listed.

mzData is the standard interchange format sponsored by the HUPO Proteomics Standards Initiative working group



Data (MS/MS)

PMF✗ SQ ✗ MS/MS✓

Data file [ ] Browse...

Data format: Mascot generic

Precursor [ ] m/z

- Mascot Generic Format (.MGF)
- Finnigan (.ASC)
- Sequest (.DTA)
- PerSeptive (.PKS)
- Micromass (.PKL)
- Sciex API III
- Bruker (.XML)
- mzData (.XML)
- mzML (.mzML)

MASCOT : *Search Parameters* © 2007-2014 Matrix Science MATRIX SCIENCE

Data for MS/MS ion searches must be supplied as an ASCII file in one of these supported formats. The format cannot be auto-detected, and must be specified using the drop down list.

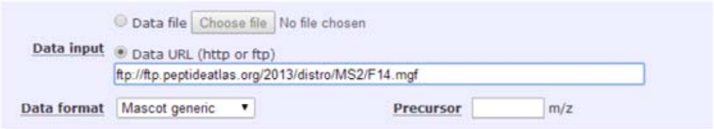
Certain data file formats, SCIEX API III, PerSeptive (.PKS), and Bruker (.XML), do not include m/z information for the precursor peptide. For these formats only, the Precursor field is used to specify the m/z value of the parent peptide.

A data file may include embedded search parameters. Most embedded parameters can only appear once, at the head of the data file. In a Mascot generic format file, a few parameters can appear within an MS/MS dataset. See the Data File Format help page for further details

If there is a conflict between the values of the embedded parameters and values entered into search form fields, the embedded parameters always take precedence. The search form fields are essentially defaults for values missing from the data file.

Data URL

PMF ✗ SQ ✗ MS/MS ✓



○ Data file Choose file No file chosen

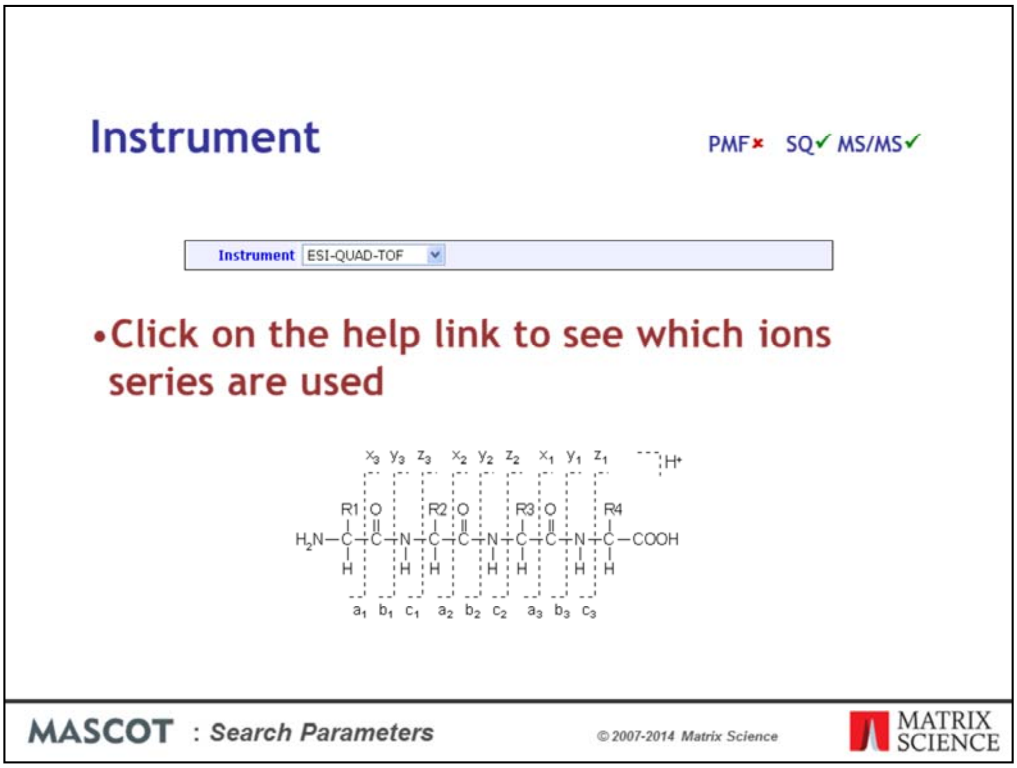
Data input ● Data URL (http or ftp)

ftp://ftp.peptideatlas.org/2013/distro/MS2/F14.mgf

Data format: Mascot generic ▾

Precursor: m/z

In Mascot 2.5 and later, if security is enabled, it is possible to specify a URL to the peak list. This means that the peak list file doesn't have to be downloaded to the client PC then uploaded to the Mascot server, which is useful for very large peak lists or when the client network connection is slow.



Instrument

PMF✗ SQ✓ MS/MS✓

Instrument ESI-QUAD-TOF

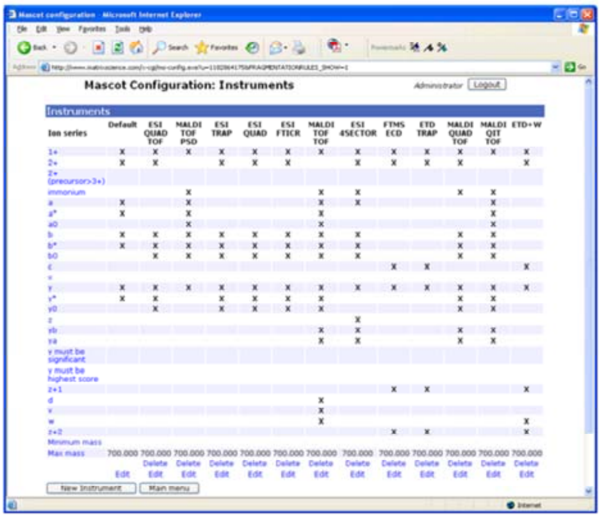
- Click on the help link to see which ions series are used

x3 y3 z3 x2 y2 z2 x1 y1 z1
H+
R1 O R2 O R3 O R4
H2N—C—C—N—C—C—N—C—C—N—C—COOH
H H H H H H H
a1 b1 c1 a2 b2 c2 a3 b3 c3

For an MS/MS Ions Search, choose the description which best matches the type of instrument used to acquire the data. This setting determines which fragment ion series will be used for scoring, according to the following table.

Instrument

PMF✗ SQ✓ MS/MS✓


Mascot configuration - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites
Administrator Logout

Mascot Configuration: Instruments

Instruments

Ion series	Default	ESI QUAD TOF	MALDI TOF PSD	ESI TRAP	ESI QUAD	ESI FTICR	MALDI TOF TOF	ESI 4SECTOR	FTMS ECD	ETD TRAP	MALDI QUAD TOF	MALDI QIT TOF	ETD+W
1+	X	X	X	X	X	X	X	X	X	X	X	X	X
2+	X	X		X	X	X		X	X	X	X		X
2+ (precursor>3+)													
immonium			X				X	X			X	X	
a	X		X				X	X				X	
a*	X		X				X					X	
a0			X				X					X	
b	X	X	X	X	X	X	X	X			X	X	
b*	X	X	X	X	X	X	X	X			X	X	
b0		X	X	X	X	X	X	X			X	X	
c									X	X			X
x													
y	X	X	X	X	X	X	X	X	X	X	X	X	X
y*	X	X		X	X	X	X				X	X	
y0		X		X	X	X	X				X	X	
z								X					
yb							X	X			X	X	
ya							X	X			X	X	
y must be significant													
y must be highest score													
z+1									X	X			X
d							X						
v							X						
w							X						X
z+2									X	X			X
Minimum mass													
Max mass	700.000	700.000	700.000	700.000	700.000	700.000	700.000	700.000	700.000	700.000	700.000	700.000	700.000
		Delete	Delete	Delete	Delete	Delete	Delete	Delete	Delete	Delete	Delete	Delete	Delete
	Edit	Edit	Edit	Edit	Edit	Edit	Edit	Edit	Edit	Edit	Edit	Edit	Edit

"Default" corresponds to the configuration used in Mascot version 1.7 and earlier.

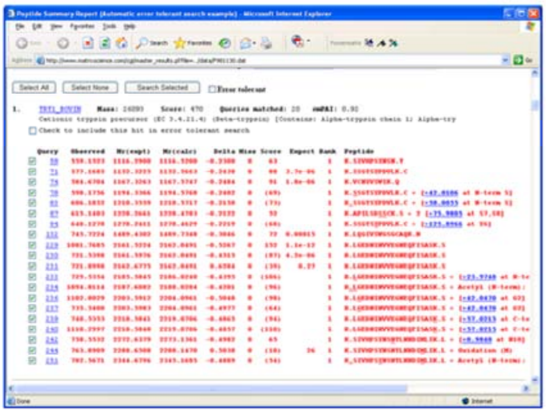
Many of the instruments are very similar.

You can modify instrument settings or create new ones using the configuration editor. In this screenshot, the right hand column is an experiment to see how the addition of w ions affects ETD matching

Error Tolerant

PMF ✗ SQ ✗ MS/MS ✓

Error tolerant ☐


Peptide Summary Report (Automatic error tolerant search example) - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Search Favorites
Select All
Select None
Search Selected
Done
Internet

☐ Error tolerant

1. TRY1_BOVIN **Mass:** 26093 **Score:** 470 **Queries matched:** 20 **emPAI:** 0.92
Cationic trypsin precursor (EC 3.4.21.4) (Beta-trypsin) [Contains: Alpha-trypsin chain 1; Alpha-try

☐ Check to include this hit in error tolerant search

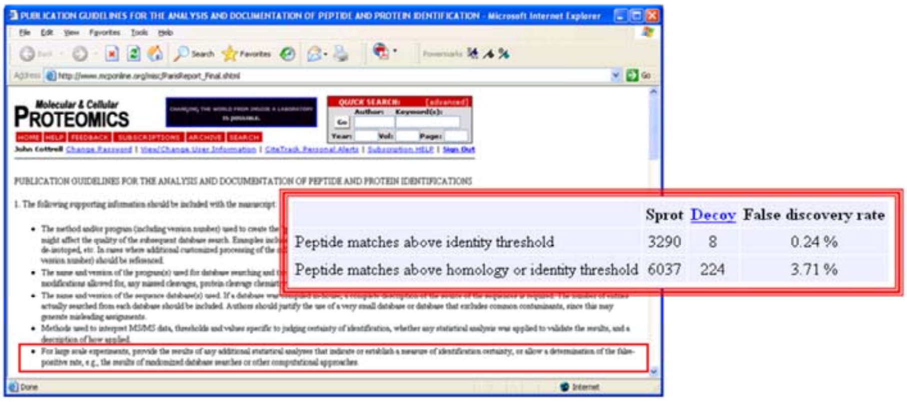
	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	59	559.1523	1116.2900	1116.5200	-0.2300	0	63		1	K.SIVHPSYNSR.T
☑	71	577.1689	1152.3233	1152.5663	-0.2430	0	88	3.7e-06	1	K.SSGTSYPDVLK.C
☑	74	584.6704	1167.3263	1167.5747	-0.2484	0	91	1.8e-06	1	K.VCNYVSWIK.Q
☑	78	598.1756	1194.3366	1194.5768	-0.2402	0	(69)		1	K.SSGTSYPDVLK.C + [+42.0106 at N-term S]
☑	81	606.1853	1210.3559	1210.5717	-0.2158	0	(73)		1	K.SSGTSYPDVLK.C + [+58.0055 at N-term S]
☑	87	615.1403	1228.2661	1228.4783	-0.2122	0	52		1	K.APILSDSSCK.S + ? [+75.9885 at S7,S8]
☑	94	640.1278	1278.2411	1278.4629	-0.2219	0	(68)		1	K.SSGTSYPDVLK.C + [+125.8966 at Y6]
☑	112	745.7224	1489.4302	1489.7348	-0.3046	0	72	0.00015	1	K.LQGIVSWGSGCAQK.N
☑	229	1081.7685	2161.5224	2162.0491	-0.5267	0	132	1.1e-12	1	R.LGEDNINVVEGNEQFISASK.S
☑	230	721.5398	2161.5976	2162.0491	-0.4515	0	(87)	4.3e-06	1	R.LGEDNINVVEGNEQFISASK.S
☑	231	721.8998	2162.6775	2162.0491	0.6284	0	(35)	0.27	1	R.LGEDNINVVEGNEQFISASK.S
☑	233	729.5354	2185.5845	2186.0240	-0.4395	0	(106)		1	R.LGEDNINVVEGNEQFISASK.S + [+23.9748 at N-te
☑	234	1094.8114	2187.6082	2188.0284	-0.4201	0	(96)		1	R.LGEDNINVVEGNEQFISASK.S + Acetyl (N-term);
☑	236	1102.8029	2203.5912	2204.0961	-0.5048	0	(98)		1	R.LGEDNINVVEGNEQFISASK.S + [+42.0470 at G2]
☑	237	735.5400	2203.5983	2204.0961	-0.4977	0	(64)		1	R.LGEDNINVVEGNEQFISASK.S + [+42.0470 at G2]
☑	239	740.5353	2218.5841	2219.0706	-0.4865	0	(94)		1	R.LGEDNINVVEGNEQFISASK.S + [+57.0215 at C-te
☑	240	1110.2997	2218.5849	2219.0706	-0.4857	0	(110)		1	R.LGEDNINVVEGNEQFISASK.S + [+57.0215 at C-te
☑	242	758.5532	2272.6379	2273.1361	-0.4982	0	65		1	K.SIVHPSYNSNTLNNDIMLIK.L + [+0.9848 at N10]
☑	244	763.8909	2288.6508	2288.1470	0.5038	0	(18)	26	1	K.SIVHPSYNSNTLNNDIMLIK.L + Oxidation (M)
☑	251	782.5671	2344.6796	2345.1685	-0.4889	0	(56)		1	K.SIVHPSYNSNTLNNDIMLIK.L + Acetyl (N-term);

If you have MS/MS data, and are interested in finding post-translational modifications, you can perform an error tolerant search by checking this box on the search form. This is a much more efficient way to discover unusual modifications, as well as non-specific peptides and sequence variants. More about this in a later presentation.

Decoy

PMF✓ SQ✓ MS/MS✓

Decoy □



PUBLICATION GUIDELINES FOR THE ANALYSIS AND DOCUMENTATION OF PEPTIDE AND PROTEIN IDENTIFICATION - Microsoft Internet Explorer
File Edit View Favorites Tools Help
http://www.mcponline.org/misc/ParisReport_Final.shtml
Molecular & Cellular PROTEOMICS
QUICK SEARCH: [advanced]
HOME HELP FEEDBACK SUBSCRIPTIONS ARCHIVE SEARCH
John Cottrell Change Password | View/Change User Information | CiteTrack Personal Alerts | Subscription HELP | Sign Out
Done
Internet

PUBLICATION GUIDELINES FOR THE ANALYSIS AND DOCUMENTATION OF PEPTIDE AND PROTEIN IDENTIFICATIONS

1. The following supporting information should be included with the manuscript:

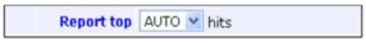
- The method and/or program (including version number) used to create the might affect the quality of the subsequent database search. Examples incl de-isotoped, etc. In cases where additional customized processing of the version number) should be referenced.
- The name and version of the program(s) used for database searching and th modifications allowed for, any missed cleavages, protein cleavage chemist
- The name and version of the sequence database(s) used. If a database was compiled in-house, a complete description of the source of the sequences is required. The number of entries actually searched from each database should be included. Authors should justify the use of a very small database or database that excludes common contaminants, since this may generate misleading assignments.
- Methods used to interpret MS/MS data, thresholds and values specific to judging certainty of identification, whether any statistical analysis was applied to validate the results, and a description of how applied.
- For large scale experiments, provide the results of any additional statistical analyses that indicate or establish a measure of identification certainty, or allow a determination of the false-positive rate, e.g., the results of randomized database searches or other computational approaches.

	Sprot	Decoy	False discovery rate
Peptide matches above identity threshold	3290	8	0.24 %
Peptide matches above homology or identity threshold	6037	224	3.71 %

The decoy checkbox enables you to validate the false discovery rate according to the approach recommended in the Molecular & Cellular Proteomics Guidelines for Publication: “For large scale experiments, provide the results of any additional statistical analyses that indicate or establish a measure of identification certainty, or allow a determination of the false-positive rate, e.g., the results of randomized database searches or other computational approaches”

Report

PMF✓ SQ✓ MS/MS✓



Report top AUTO hits

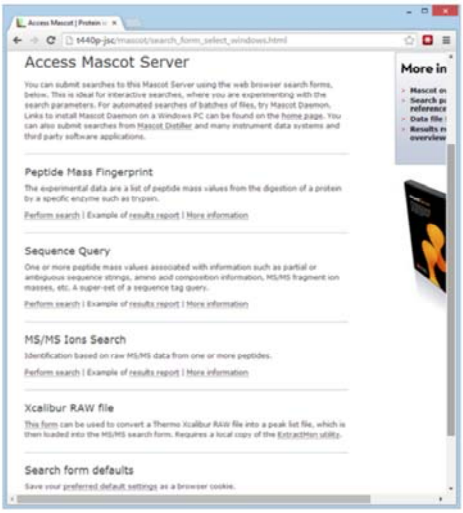
Report top should normally be set to auto.

MASCOT : Search Parameters © 2007-2014 Matrix Science MATRIX SCIENCE

REPORT determines the *maximum* number of hits displayed in a search results report. Choose AUTO to display only protein hits with significant scores. In a protein summary report, one additional hit is reported after the cutoff at the significant score. This is to ensure that the report shows the highest scoring hit, even though it is not significant.

Setting defaults

PMF✓ SQ✓ MS/MS✓



Access Mascot | Protein ... ×

t440p-jsc/mascot/search_form_select_windows.html

Access Mascot Server

You can submit searches to this Mascot Server using the web browser search forms, below. This is ideal for interactive searches, where you are experimenting with the search parameters. For automated searches of batches of files, try Mascot Daemon. Links to install Mascot Daemon on a Windows PC can be found on the home page. You can also submit searches from Mascot Distiller and many instrument data systems and third party software applications.

Peptide Mass Fingerprint

The experimental data are a list of peptide mass values from the digestion of a protein by a specific enzyme such as trypsin.

Perform search | Example of results report | More information

Sequence Query

One or more peptide mass values associated with information such as partial or ambiguous sequence strings, amino acid composition information, MS/MS fragment ion masses, etc. A super-set of a sequence tag query.

Perform search | Example of results report | More information

MS/MS Ions Search

Identification based on raw MS/MS data from one or more peptides.

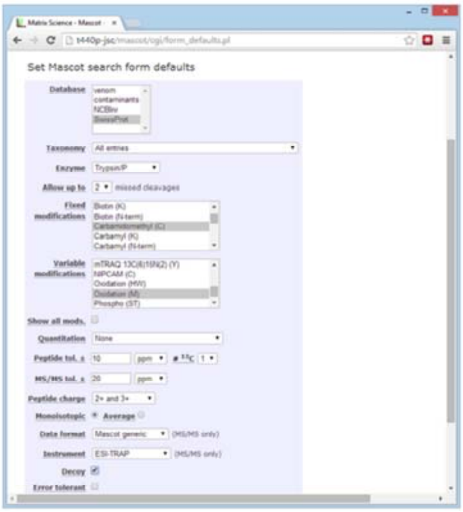
Perform search | Example of results report | More information

Xcalibur RAW file

This form can be used to convert a Thermo Xcalibur RAW file into a peak list file, which is then loaded into the MS/MS search form. Requires a local copy of the ExtractMsn utility.

Search form defaults

Save your preferred default settings as a browser cookie.



Matrix Science - Mascot ×

t440p-jsc/mascot/cgi/form_defaults.pl

Set Mascot search form defaults

Database	venom contaminants NCBInr SwissProt
Taxonomy	All entries
Enzyme	Trypsin/P
Allow up to	2 missed cleavages
Fixed modifications	Biotin (K) Biotin (N-term) Carbamidomethyl (C) Carbamyl (K) Carbamyl (N-term)
Variable modifications	mTRAQ 13C(6)15N(2) (Y) NIPCAM (C) Oxidation (HW) Oxidation (M) Phospho (ST)
Show all mods.	□
Quantitation	None
Peptide tol. ±	10 ppm # ^{13}C 1
MS/MS tol. ±	20 ppm
Peptide charge	2+ and 3+
Monoisotopic	● Average ○
Data format	Mascot generic (MS/MS only)
Instrument	ESI-TRAP (MS/MS only)
Decoy	☑
Error tolerant	□

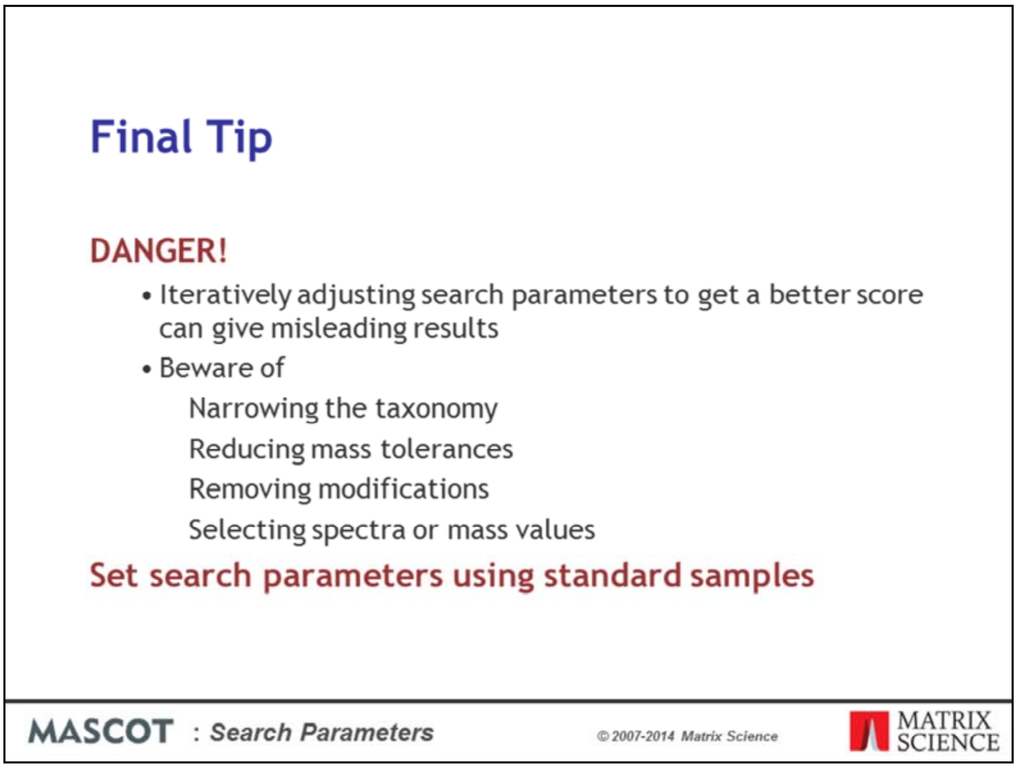
MASCOT : *Search Parameters*

© 2007-2014 *Matrix Science*

MATRIX SCIENCE

You can choose your own defaults for the search forms. Look for the link at the bottom of the search form selection page

When you save the defaults, they are saved as a browser cookie. If you go to a different PC, or switch to a different browser, you'll need to repeat this step



Final Tip

DANGER!

- Iteratively adjusting search parameters to get a better score can give misleading results
- Beware of

  Narrowing the taxonomy

  Reducing mass tolerances

  Removing modifications

  Selecting spectra or mass values

Set search parameters using standard samples

MASCOT : *Search Parameters* © 2007-2014 Matrix Science MATRIX SCIENCE

A final word of advice: It is easy to distort the search results without realising.

Basically, it is risky to adjust the search parameters interactively to get a better score for an unknown.

For example, you search the complete database and don’t get a significant match. However, a very interesting looking protein is near the top of the list, surrounded by some others that are clearly wrong. You change the taxonomy filter so as to exclude the “wrong” proteins. Sorry, but this is cheating.

Search parameters should be set using standards. Broadening the search if you get a negative result is usually OK, but not narrowing the search.